A promoter characterisation of the Ets transcription factor gene, *Ese-2/Elf-5*

Kathy Ann Speed

2002

A promoter characterisation of the Ets transcription factor gene, *Ese-2/Elf-5*

Kathy Ann Speed

A thesis submitted
in partial fulfilment of
the requirements for the degree of
Doctor of Philosophy

Department of Biological Sciences
School of Science and Technology
The University of Waikato

2002

Abstract

E26 transformation specific (Ets) proteins are a versatile class of transcription factors. They act both up and downstream of signal transduction cascades, are causative in cancer generation and are able to control the expression of other transcription factors. Thus they play an important role in cellular growth, proliferation and differentiation. The *ets* gene family is characterised by an 85 amino acid DNA binding domain and currently consists of around 30 members. They have been shown to be proto-oncogenes and it is this feature that has elicited much research into their regulation and function in both normal and neoplasic tissue. The high level of conservation within the ETS domain of the family identifies all Ets proteins as sequence-specific DNA-binding proteins. Many have been identified as having a role in the regulation of gene expression. The retention of many *ets* genes throughout evolution implies there would be few redundant genes (Thomas, 1993) and therefore each gene would have a unique biological role. Because of the common DNA binding domain, the regulation of the *ets* genes must be complex to provide specificity of action throughout the family (Laudet *et al.*, 1999).

A cDNA sequence was isolated from mammary gland tissue and characterised as belonging to the *Ets* family of transcription factors. The gene was designated *BrETS* for Breast ETS. Recent publications have described the gene as *Elf*-5 (Zhou *et al.*, 1998) and *ESE*-2 (Oettgen *et al.*, 1999). *BrEts* showed a restricted pattern of spatial expression, being strongly expressed in the mammary gland. Its expression in the mammary gland also exhibits a tightly controlled temporal pattern, which implicates it in mammary processes such as development and differentiation. As *BrETS* is highly expressed in the mammary gland (where intense extracellular remodelling occurs), and has been localised to the epithelial tissue, it may play a role in the regulation of matrix metalloproteinases or their inhibitors.

Genomic DNA spanning the Br*ETS* locus was isolated in this study. It was found to encompass approximately 3kb of upstream sequence, the first two exons and intron 1. Multiple transcription start sites have been identified, indicative of a TATA-less, tissue-specific gene. Deletion constructs of the proximal promoter (850bp) were examined in cell culture. The mammary gland derived cell lines HC11 and Comma1D were used, in conjunction with a chloramphenicol acetyltransferase reporter system. A region of sequence between -70 and -200bp was identified as having up-regulated transcriptional activity above an empty expression vector in cell culture. Analysis of more distal promoter constructs resulted in decreased activity, to a level representative of basal activity. The first intron (6kb) was also examined *in vitro* as a complete construct and as smaller fragments. Low levels of reporter activity were seen with these constructs in cell culture.

DNaseI hypersensitive analysis on the immediate promoter region revealed the presence of a hypersensitive site within the first intron. This implies a regulatory element of potential importance is located here, though further research is required to determine the element(s) and significance.

The identification of regulatory regions may allow for the determination of tissue-specific elements and regions that control the expression of the gene. This information may be later extrapolated and manipulated *in vivo* as a means of controlling the expression of the Br*ETS* gene.

Acknowledgements

I would like to acknowledge the support of my supervisors who have been involved throughout the research and writing of this PhD thesis. First, Dick Wilkins, who has lent major support to the technical problems encountered throughout the research, as well as having made a large contribution toward the manuscript preparation. Dick has also encouraged me in scientific discussion to further my knowledge gained in the laboratory. Second, Phil L'Huillier, who provided motivation, support and encouragement early on, and who was always approachable in the lab. Third, Goetz Laible, who inherited the supervision of my thesis as the research drew to an end and has been especially helpful with the manuscript preparation.

I would also like to thank Brigid, Angela, Brett, Helen and Peter, who have given advice and general technical support. Also thanks to Ben for help with statistics, and Megan, for help with tissue culture. Thanks to my team of proof-readers who came through under pressure in the final days, Brigid, Leesa and Tracie, your efforts were fantastic.

A big thanks to all the members of the Ruakura social sports teams, and to everyone involved with organising and participating in social events around campus. These were the things that kept me going.

Special thanks to all my friends who have been around throughout my time at Ruakura - Brigid, Angela, Megan and Ben, and everyone who has shared tea breaks, gym classes and sports events. Thanks also to those who have been a positive influence – particularly Chrissy, Dave, Meagan, James, David and Ann, Kris and Warrick, and Chook.

A huge thank you to my whole family. Mum and Dad, I am so grateful for your never ending love and support.

Finally, to Dave, whose love and belief carried me through the final days.

Table of Contents

List of Figures

Page No.

Chapter 1. Introduction and literature review

Chapter 3. Identification and analysis of the murine *Ese*-2 5' region

Chapter 4. Human genome analysis and sequence alignments

Chapter 5. Promoter and enhancer *in vitro* activity

List of Tables

Page No.

Chapter 1. Introduction and literature review

Chapter 2. Materials and Methods

Chapter 4. Human genome analysis and sequence alignments

Chapter 5. Promoter and enhancer *in vitro* activity

Abbreviations

Systeme Internationale (d'Unites) abbreviations are used for units and standard notations for chemical elements, formulae and chemical abbreviations are used in the text. Other abbreviations used are listed below.

°C	degrees Celcius
aa	amino acid
amp	ampicillin resistance gene
APS	ammonium persulfate
β-Gal	beta galactosidase
BME	beta- mercapto-ethanol
bp	base pairs
BSA	Bovine serum albumin
CAT	chloramphenicol acetyltransferase
cDNA	complementary deoxyribonucleic acid
cpm	counts per minute
DEPC	diethyl pyrocarbonate
DHS	DNase hypersensitive assay
DNA	deoxyribonucleic acid
dNTP	deoxy nucleotide triphosphate
DTT	dithiothreitol
EDTA	disodium ethylenediaminetetra-acetic acid $2H_20$
EGF	epidermal growth factor
ELISA	enzyme linked immuno sorbent assay
EtBR	ethidium bromide
FCS	fetal calf serum
hr(s)	hour(s)
Kan	kanamycin
kb	kilobase(s)
LB	Luria broth
LF2000	Lipofectamine 2000
Min	minute(s)
MOPS	3-(N-morpholine)-propane-sulfonic acid $2H_2O$

MW	molecular weight
nt	nucleotide(s)
OD	optical density
PBS	phosphate buffered saline
PCR	polymerase chain reaction
RNA	ribonucleic acid
rpm	revolutions per minute
SDS	sodium dodecyl sulphate
sec	second(s)
SSC	saline sodium citrate
TB	transformation buffer
TCA	trichloro acetic acid
Tdt	terminal transferase
TE	Tris-EDTA
TEMED	N,N,N',N',-tetramethylethylenediamine
tk	thymidine kinase
TLC	thin layer chromotography
Tris-Cl	Tris solution, pH adjusted with HC1
U	units of enzyme
UV	ultraviolet
w/v	weight per volume
w/w	weight per weight
xg	times g force

Abbreviations for commercial suppliers

Amersham	Amersham Pharmacia Biotech, Dominion Rd, Auckland, NZ
Biolab	Biolab Scientific Ltd, North Shore, Auckland, NZ
BioRad	BioRad Laboratories PTY Ltd, Auckland, NZ
CLONTECH	BD Biosciences Clontech UK, Oxford, UK
Invitrogen	Invitrogen NZ Ltd, Penrose, Auckland, NZ
Pierce	Pierce Biotechnology, distributed by Global Sciences and Technology, Auckland, NZ

Progenz	Progenz Ltd, Luckens Point, Auckland, NZ
Promega	Dade Behring Diagnostics Ltd (Promega), Auckland, NZ
Roche	Roche Diagnostics NZ Ltd, Mt Wellington, Auckland, NZ
Sigma	Sigma Aldrich NZ, Penrose, Auckland, NZ
Stratagene	Stratagene, distributed by Global Sciences and Technology, Auckland, NZ
USB Chemicals	United States Biochemical Corporation, distributed by Amersham

Abbreviations for Construct Number

Construct Description	Construct Number
pBLCAT2 no promoter	0 (negative)
pBLCAT2 + 39bp proximal 5'	1
pBLCAT2 + 76bp proximal 5'	2
pBLCAT2 + 200bp proximal 5'	3
pBLCAT2 + 500bp proximal 5'	4
pBLCAT2 + 850bp proximal 5'	5
pBLCAT2 + 1300bp distal 5'	6
pBLCAT2 + 1700bp distal 5'	7
pBLCAT2 + 2600bp distal 5'	8
pBLCAT2 + 2600bp distal 5' + 6300bp Intron 1	9
pBLCAT2 + 500bp 5' proximal + 2400bp Intron 1(5')	10
pBLCAT2 + 500bp proximal 5' + 2500bp Intron 1 (mid)	11
pBLCAT2 + 500bp proximal 5' + 1400bp Intron 1 (3')	12

Chapter 1. Introduction and literature review

1.0 Background

This investigation characterises some features of the *Ese*-2/*Elf*-5 gene promoter. The gene will be referred to as *Ese*-2 throughout this manuscript, as it is more closely related to members of the *Ese* subfamily than the *Elf* subfamily, despite the locus link classification.

Ese-2 is a member of the *Ets* family of transcription factors, which are characterised by an 85 amino acid DNA binding domain. Both the human and mouse *Ese*-2 genes were described by Zhou *et al.*, in 1998 (named ELF-5) and subsequently in 1999 by Oettgen *et al.* *ESE*-2 was characterised as an epithelium specific factor closely related to *ESE*-1, though functionally distinct (Oettgen *et al.*, 1999).

Zhou *et al.*, (1998) and Oettgen *et al.*, (1999) have shown that *Ese*-2 expression is restricted to cells of epithelial origin, including mammary gland, salivary gland, kidney, lung, prostate, colon, ovary, pancreas, liver, placenta, and the gastrointestinal tract.

The starting point for the present investigation was a cDNA sequence that was isolated from mammary gland tissue and characterised as belonging to the *Ets* family of transcription factors. The gene was designated *BrETS* for Breast ETS.

This chapter discusses the previous research and information known regarding the *ets* transcription factor family, including physical characteristics, binding properties and regulation with respect to the specificity required to meet a high level of control. As *BrETS* is highly expressed in the mammary gland (where intense extracellular remodelling occurs), and has been localised to the epithelial tissue, it may well play a role in the regulation of matrix metalloproteinases or their inhibitors. Some background on Ets

factors acting in the context of the extracellular matrix is given. In the later section of this chapter, a brief overview of the aims and content of the study are given, together with the previous research already completed in the area.

1.1 The Ets family of transcription factors

Transcription factors are proteins that bind to specific sequence motifs on the DNA and influence the rate of initiation of transcription as well as the amount of transcription that occurs from that gene.

The *ets* family of transcription factors was named after the first member was described in 1983 (Leprince *et al.*, 1983; Nunn *et al.*, 1983). v-*ets* (E26 transformation specific or E-twenty six specific) was discovered to be a cellular sequence that was transduced by the avian retrovirus E26. E26 is a fusion hybrid consisting of the 5' region of the retroviral *gag* gene, myb^{E} and the *ets* sequence. The cellular counterpart of v-*ets* (c-*ets*) is transcribed under normal conditions in chicken cells as a major 7.5kb polyadenylated RNA and the sequence is conserved through evolution (Leprince *et al.*, 1983; Nunn *et al.*, 1983).

Following the discovery of v-*ets*, many other *ets* genes were identified from phylogenetically divergent species and the family now comprises upward of 30 members. These genes are listed in Table 1.1.

The oncogenic nature of the family was confirmed with experiments *in vitro* and *in vivo* using the *Ets*-2 gene. c-*Ets*-2 was shown to be oncogenic and mitogenic when expressed in NIH3T3 cells. Its over-expression results in stimulated cell proliferation and abolishment of a serum requirement. Cells transfected with c-*ets*-2 form colonies in a semi-solid medium and induce tumours in nude mice (Seth *et al.*, 1989).

1.1.1 The ETS domain

The conserved region that identifies the family is known as the ETS-domain and is responsible for the DNA binding of this group of proteins. Through a comparison of the DNA binding properties of the murine ets-1 proto-oncogene, murine PU.1 and *Drosophila* E74, Karim *et al.* discovered a sequence specific region of 85 amino acids which was highly conserved. This was termed the ETS domain (Karim *et al.*, 1990). At the same time a role for the ETS protein domain was discovered as it was found to bind DNA in a sequence specific manner (Gunther *et al.*, 1990; Klemsz *et al.*, 1990; Urness and Thummel, 1990). Fragments of the conserved domain from PU.1 and Ets-1 proteins retained the DNA binding capability (Gunther *et al.*, 1990; Klemsz *et al.*, 1990).

For many years, the ETS domain was believed to be a novel DNA binding motif due to its lack of sequence homology to any other well characterised motif (Johnson and McKnight, 1989; Mitchell and Tijan, 1989). It contains no cysteines or histidines positioned for stabilising a zinc finger structure, neither are the α helical regions spaced appropriately for a helix-turn-helix configuration (Murre *et al.*, 1989). Basic amino acids are found in the C-terminus of the ETS domain but have no similarity with leucine zipper/basic region proteins (Landschulz *et al.*, 1989) or helix-loop-helix proteins (Murre *et al.*, 1989). However, the ETS domain does contain a region of three highly conserved tryptophans separated by 17-20 amino acids and a basic region (Janknecht and Nordheim, 1993). This spacing resembles a region found in the myb proteins, which function in DNA binding (Anton and Frampton, 1988). PU.1 is an exception to the three-tryptophan pattern having one replaced by a tyrosine residue. The tryptophan side chain was thought to intercalate into the DNA of Ets-1 and rotate 180° in relation to the ETS site. However, more recent evidence suggests that there is no tryptophan intercalation (Graves and Gillespie, 1996). In fact, it was the elucidation of the 3D

structure of Fli-1 (Liang, 1994a), Ets-1 (Donaldson *et al.*, 1994; Werner *et al.*, 1995) and PU.1/Spi-1 (Kodandapani, 1996) that demonstrated the ETS domain belonged to the superfamily of helix-turn-helix proteins. Analysis of the binding domain whilst targets are either bound or un-bound has shown the domain has 3 α-helices and a 4 stranded antiparallel β-sheet (Sharrocks *et al.*, 1997). Major groove contacts with the DNA binding site GGAA/T are made by a loop formed between the β3-β4 turn and between α-helices 2 and 3 (Kodandapani, 1996). The ETS domain is so highly conserved throughout the family that this mode of binding is likely to be identical in all the family proteins. The ETS domain is found to be in one of three locations within the protein. Near the carboxyl terminus, the amino terminus or centrally located, though it is not known if its position is of biological significance.

1.1.2 Phylogeny of the Ets transcription factor family

Ets genes are widely dispersed throughout metazoans ranging from chordates, arthropods and higher mammals to sponges, ctenophores and flatworms. So far, no examples have been found in plant, fungi or protozoans and none in the genome of the budding yeast, *Saccharomyces cerevisiae.* The origin of *ets* genes may pre-date the diversification of metazoans, as indicated by the discovery of *ets* members in sponges and coelenterates (Degnan *et al.*, 1993). Previous phylogenetic studies hypothesise a period of gene duplication in early metazoan evolution and an extended period of conservation within the family implicated by a relatively small number of genes (Lautenberger *et al.*, 1992; Laudet *et al.*, 1993; Shenk and Steele, 1993; Graves and Petersen, 1998; Laudet *et al.*, 1999).

The strong conservation of the ETS domain which spans 84-90 amino acids is the defining feature of the ets family members, and is important both structurally and functionally (Donaldson *et al.*, 1994; Liang *et al.*, 1994b; Werner *et al.*, 1995; Donaldson *et al.*,

1996; Kodandapani, 1996). The discovery of the ETS domain winged-helix-turn-helix (WHTH) structure with the resolution of the 3D structure presents a possible evolutionary grouping in this superfamily of transcriptional regulators. This structure retains high similarity to the helix-turn-helix (HTH) of the *E.coli* catabolite gene activator protein, the HNF3/fork head family and the heat shock factor family (Donaldson *et al.*, 1994; Liang, 1994a; Donaldson *et al.*, 1996; Kodandapani, 1996).

A highly conserved N-terminal region is found in some Ets family members, referred to as the Pointed domain (Klambt, 1993). This domain is conserved in Ets-1, Ets-2, GABPα, Erg, Fli-1, Tel, Yan and Ese members and it is proposed to play a role in protein-protein interactions and specifically for Tel in its self-association (Carroll *et al.*, 1996; Golub *et al.*, 1996; McLean *et al.*, 1996; Jousset *et al.*, 1997).

Table 1.1 List of *ets* genes.

Ets gene name(s)	Subfamily
ERG	ERG
FLI-1/ERG-B	ERG
ETS-3	ERG
FEV	ERG
ETS-6	ERG
ERF	ERF
PE-1/ETV-3	ERF
ELG	GABP
GABPα/E4TF1-60	GABP
ETS-1	ETS
ETS-2	ETS
POINTED	ETS
ER71/ETV2	ER71
ERP/NET	ELK
SAP-1	ELK
ELK-1	ELK
ER81/ETV1	PEA3
ERM	PEA3
PEA3/E1AF/ETV4	PEA3
ETS-4	DETS4
PDEF	DETS4
MEF	ELF
NERF	ELF
ELF-1	ELF
E74	ELF
TEL/ETV6	TEL
TEL-2	TEL
YAN	YAN

ESE-1/ELF-3/ESX/ jen/ERT	ESE
ESE-3/EHF	ESE
ESE-2/ELF-5/BrETS	ESE
PU.1/Spi-1	SPI
SpiB	SPI

In support of the protein-protein interaction, the more conserved core of Pointed has been aligned with sequences in other non-Ets proteins including a region in the Polycomb protein family and in cytoplasmic proteins involved in yeast sexual differentiation (Alkema *et al.*, 1996; Gunster *et al.*, 1997; Graves and Petersen, 1998). The conserved domain is called SEP, for yeast sterility, Ets-related, polycomb proteins.

Several recent phylogenetic analyses have been performed, mostly resulting in a consensus alignment and grouping of the *ets* genes (Laudet *et al.*, 1999; Kas *et al.*, 2000). Laudet *et al.* used several methods to align amino acid sequences, and a distance tree was constructed with the neighbour-joining method and tested by 'bootstrap' replications. This tree displays high similarity to all other attempts at phylogenetic relationships and includes 13 subgroups within the family (ETS, ER71, GABP, PEA3, ERG, ERF, ELK, DETS4, ELF, ESE, TEL, YAN, SP-1). The phylogenetic tree derived by Kas *et al.* (2000) reveals the same 13 subgroups, though with some variation in nearest neighbours, likely to be due to the different methods used to construct the trees. This was generated by alignment using the Clustal W program of the ETS domain (Thompson *et al.*, 1994) and is shown in Figure 1.1. The same group presents an alignment of genes retaining the conserved Pointed domain, as shown in Figure 1.2.

The divergence between *Ets-1* and *Ets-2* in vertebrates came before the split between amphibians and other vertebrates (Degnan *et al.*, 1993). A second series of duplication events seems to have occurred specifically for the vertebrate lineage, giving rise to paralogues within each group *Ets-1/Ets-2, PEA3/ERM/ER81, Erg/Fli-1, Erf/Pe-1, Elk/Sap-1/Net, Elf-1/Mef/Nerf, Spi-1/Spi-B.* These block duplications are supported by some striking chromosomal locations, such as *Ets-1* and *Fli-1* neighbouring on chromosome 11 q23.24, and *Ets-2/Erg* neighbouring on chromosome 21 at q22 (Kasahara *et al.*, 1996; Katsanis *et al.*, 1996).

In the tree constructed by Laudet *et al.*, distant relations of the WHTH family provided a root for the tree, which was located between *Spi-1* and others as previously predicted (Laudet *et al.*, 1999). Evolutionary rates were measured for each gene between species and the results indicated that evolution was comparable with that of other transcription factors such as the nuclear receptors (Laudet, 1997).

1.1.3 Ets DNA binding sites

Most Ets proteins bind as monomers to DNA sequences, except for GABP, which is dimeric, containing a purine rich core GGAA/T (Karim *et al.*, 1990; Urness and Thummel, 1990; Hagman and Grosschedl, 1992; Janknecht and Nordheim, 1992; Nye *et al.*, 1992; Woods *et al.*, 1992). The sequences that flank the core target modulate the binding specificity and affinity, often containing 9-15 bases of DNA that make contact with the protein (Graves and Petersen, 1998). Experimental evidence for the core GGAA binding site initially came from methylation interference analysis which determined that the two guanine residues act as contact points for both Ets-1 and PU.1 protein binding (Gunther *et al.*, 1990; Klemsz *et al.*, 1990). Also, a mutant core sequence of CCAA was not able to bind PU.1 (Klemsz *et al.*, 1990).

DNA mutagenesis studies have identified the subgroups of sequences around the core site. Significant overlap is seen within a 9bp region, though a major difference is seen in some proteins (e.g. Elk-1, Fli-1, Elf-1, E74) which will only tolerate an A at position GGA<u>A</u>, whereas others like Ets-1, GABPα, SAP-1a can have either an A or T (Shore and Sharrocks, 1995). Protein-protein interactions are likely to change the specificity and affinity of Ets binding by causing preferential binding of one Ets protein over another through a strengthened interaction.

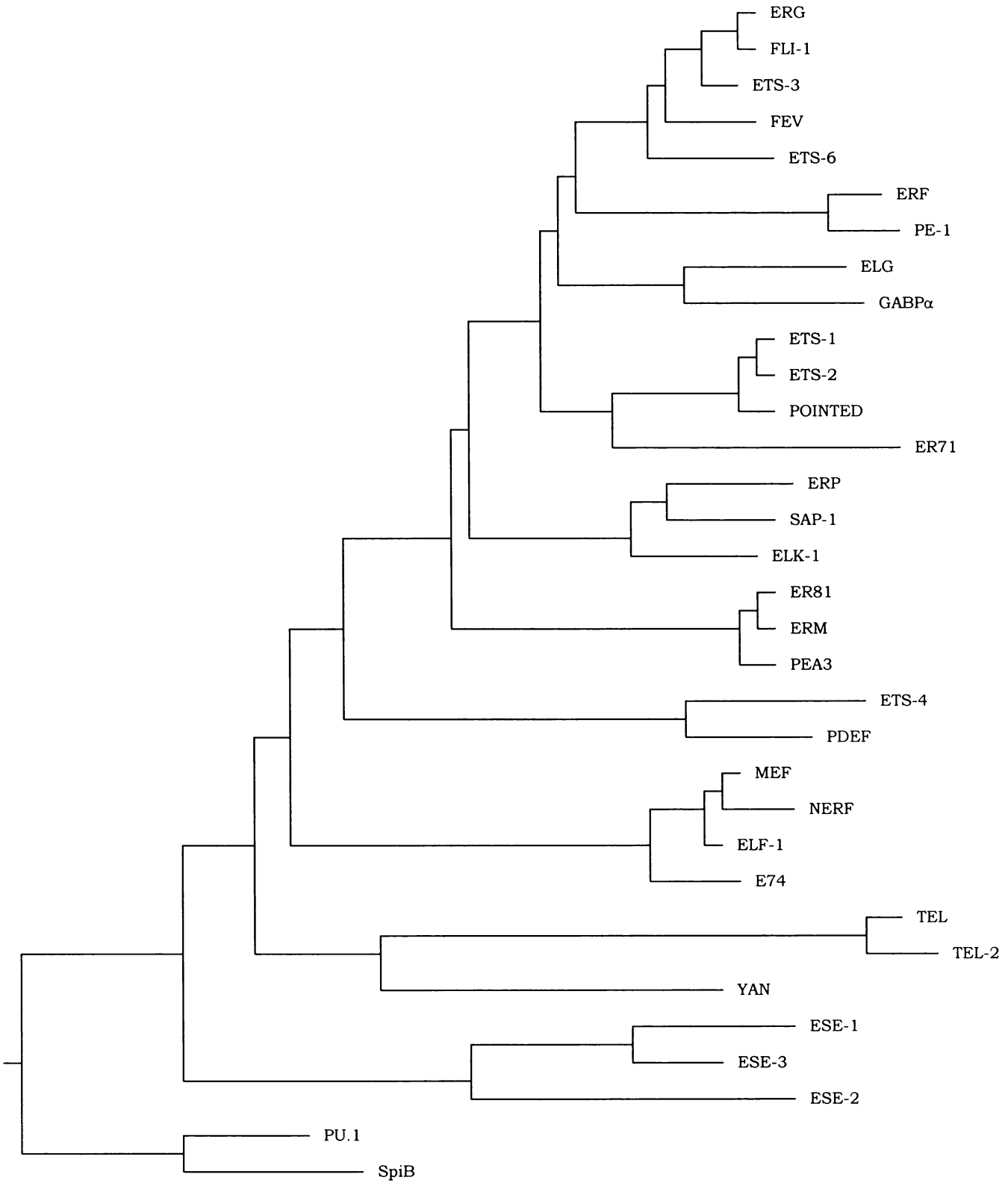

ERG
FLI-1
ETS-3
FEV
ETS-6
ERF
PE-1
ELG
GABPα
ETS-1
ETS-2
POINTED
ER71
ERP
SAP-1
ELK-1
ER81
ERM
PEA3
ETS-4
PDEF
MEF
NERF
ELF-1
E74
TEL
TEL-2
YAN
ESE-1
ESE-3
ESE-2
PU.1
SpiB

Figure 1.1 Phylogeny tree of the Ets family of transcription factors Adapted from Kas *et al.*, 2000. All known members of the *Ets* gene family are shown of human and *Drosophila* origin. Branch lengths indicate the relative similarity of the ETS domain between members.

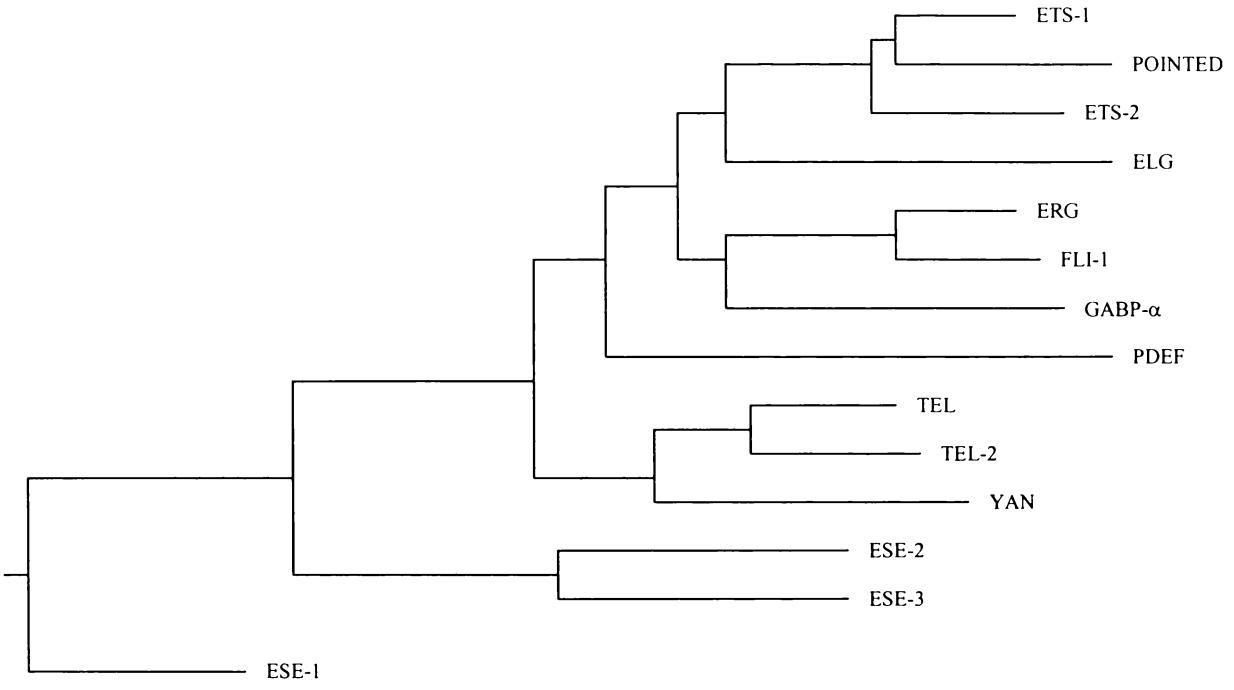


Figure 1.2 Phylogeny tree of the Ets factors containing the Pointed domain Adapted from Kas *et al.*, 2000. An alignment of the Pointed domain, including 12 mammalian and 2 *Drosophila* factors. Branch lengths indicate relative similarity.

Elk-1 and SAP-1 for example both bind the GGAT sequence of the c-fos serum response element only in conjunction with the DNA bound serum response factor (Hipskind *et al.*, 1991; Dalton and Treisman, 1992). Another example of binding affinity is GABPα, which will bind DNA more strongly in the presence of GABPβ (Thompson *et al.*, 1991). The different binding specificity of *ets* genes may reflect an important mechanism by which the factors regulate different genes expression patterns (Wasylyk *et al.*, 1992).

1.1.4 Regulation of DNA binding

A primary level of regulation by transcription factors occurs via sequence specific DNA binding. Furthermore, regulation of the binding is also critical. Binding is known to occur through the GGAA site as already discussed, and the ETS domain may play a role in such binding specificity. However, several studies have indicated that regions outside the ETS domain are important for regulating DNA binding.

Ets-1 is the most widely studied example of the Ets factors and now is known to have three regions that have a negative influence on DNA binding. First, a central region of approximately 70 amino acids (Lim *et al.*, 1992), a second 90 amino acid region upstream of the ETS domain (Wasylyk *et al.*, 1992) and finally a region in the carboxy terminus (Hagman and Grosschedl, 1992; Lim *et al.*, 1992).

In 1992, Lim *et al.* published their findings from a study of the oncogenes v-*myb* and v-*ets*, which are fused with the viral *gag* gene in E26 avian retrovirus. The cellular precursor c-*ets*-1 gives rise to v-*ets* which undergoes two point mutations and substitution of the C-terminal 13 amino acids by a different 16 amino acid sequence. The DNA binding domain was thought to be present at the C-terminus, so a comparison of the DNA binding between v-*ets* and c-*ets* was of interest. The polyoma PEA3 element was used as a known Ets binding site (EBS) and results indicated weak binding of

c-*ets*-1 but strong binding of v-*ets*. Also, fragments of c-*ets*-1 showed strong DNA binding suggesting that the full-length molecule has a repressed binding domain. v-*ets* counterparts of the c-*ets* deletion fragments differing by only a few mutations retained DNA binding activity. The regions of repression were mapped to the C-terminus end and a middle region. Truncations deleting these regions resulted in the loss of the DNA binding capability. The C-terminus in v-*ets* is altered compared to that of c-*ets* and retains high DNA binding activity (Lim *et al.*, 1992).

Lim *et al.* (1992) proposed a mechanism for the repressor action of the C-terminus and central repressor domains on DNA binding. Both domains contact each other masking the intervening ETS domain and preventing its contact. Models propose that an inhibitory molecule (two α-helices N-terminal to the ETS domain and a single C-terminal α-helix) alters the DNA binding potential of Ets-1. The C and N-terminal α-helices function cooperatively and a conformational change of one N-terminal α-helix allows relief of the intramolecular inhibition (Petersen *et al.*, 1995; Jonsen *et al.*, 1996; Skalicky *et al.*, 1996).

Hagman and Grosschedl (1992) provided further evidence for the importance of the C-terminal amino acids in DNA binding regulation (Hagman and Grosschedl, 1992). They reported that Ets-1 and PU.1 both bind the *mb-1* promoter (B-lymphocyte specific *mb-1* gene) *in vitro* though with different efficiency. The cause of this is the presence of an inhibitory C-terminal domain in Ets-1 and Ets-2, which is absent in PU.1 and v-Ets. Substituting the C-terminus of the v-Ets protein with Ets-1 C-terminus repressed DNA binding. This was not the case for the substitution with the PU.1 C-terminus (Hagman and Grosschedl, 1992), which may be due to a lack of the central repressor domain in PU.1. The C-terminal inhibitory domain of Ets-1 was predicted to interfere directly with DNA binding and not homo-dimerisation. Repressor domains of DNA binding are therefore not present in PU.1 because the deletion of either carboxy or amino sequences did not result in

enhanced DNA binding (Hagman and Grosschedl, 1992; Wasylyk *et al.*, 1992). This mechanism of regulation may be specific to certain factors only.

Wasylyk *et al.* (1992) examined the DNA binding of deletion mutants of the c-*ets*-1 (p68) gene (chicken Ets1 p68 gene). They found that binding was abolished when an amino terminal region was removed and that this region in Ets-1 and Ets-2 is able to inhibit DNA binding.

1.1.5 Ets target genes

Well over 200 Ets target genes have been documented (Sementchenko and Watson, 2000). These target genes are involved in cell proliferation, differentiation, development, hematopoiesis, apoptosis, metastasis, tissue remodelling, angiogenesis and transformation. The various Ets target gene interactions are an important area of investigation.

Early Ets-1 targets were discovered in the long terminal repeats of the human T-cell leukemia virus 1 (HTLV-1) and Moloney murine sarcoma virus (Gunther *et al.*, 1990; Gegonne *et al.*, 1993). GABPα and GABPβ were shown to interact with the Herpes Simplex Virus and the Adenovirus E4 promoter region (LaMarco *et al.*, 1991; Watanabe *et al.*, 1993). These target sites provided an early indication that *ets* genes might play a role in transformation and tumour associated processes.

More recently Ets factors have been found to bind to sites not predicted by their *in vitro* consensus sequences (Sementchenko and Watson, 2000). The J-chain transcription factor NF-JB (found to be identical to PU.1) is an example of this. It contains an AAGAAA core sequence rather than the GGA core (Shin and Koshland, 1993). It has been found to bind sequences such as the macrophage scavenger receptor at AGAGAAGT (Moulton *et al.*, 1994) and also the IL-1β at GCAGAAGT (Buras *et al.*, 1995; Kominato *et al.*, 1995). It therefore seems that target gene

specificity and affinity to the EBS is controlled by cis-elements, as demonstrated by the ability of many Ets factors to act in synergy with other transcription factors. Another example is the Ets-1 cooperation with CREB and AML1 non-consensus binding in the promoter activation of the human T-cell receptor beta chain (Halle *et al.*, 1997). Halle *et al.* (1997) propose that the low affinity binding may mediate cooperative concentration dependent regulation of a promoter and that high affinity binding is involved in constitutive activation. Ets factors may also be in competition with other transcription factors. Fli-1 for example, acts as a repressor of retinoblastoma (Rb) (Tamir *et al.*, 1999), as an activator for the tenascin-C (TN-C) promoter and a repressor of collagen promoter activity (Shirasaki *et al.*, 1999). Ets-1 however is capable of activating both TN-C and collagen promoter activities (Shirasaki *et al.*, 1999).

Other Ets factors interact specifically with matrix type promoters including the collagenase promoter (MMP-1), (by Erg and Ets-2) and the stromelysin (MMP-3) promoter. Erg was seen to inhibit the stromelysin promoter, while Ets-2 stimulated both stromelysin and collagen (Buttice *et al.*, 1996). A synergistic interaction is also observed when stromelysin is activated by the CBP/p300 protein and Ets-2 (Jayaraman *et al.*, 1999). Erg however was not able to synergise with CBP/p300, which may explain the different activity. The matrix type promoters are discussed in further detail later.

1.1.6 Regulation of function via protein-protein interactions

The regulation of gene transcription involves combinatorial interactions of transcription factors. Many protein-DNA and protein-protein interactions are able to influence promoter activity. While many transcription factor families bind DNA as hetero or homodimers most Ets proteins have been observed to bind DNA as monomers (Liang, 1994a; Werner *et al.*, 1995; Kodandapani, 1996).

The transcriptional activity of Ets proteins is more often modulated by their synergistic interaction with other factors than with themselves. Synergistic control is a mechanism by which tissue-specific gene expression may be controlled by transcription factors unique to each cell lineage or differentiation state (Li *et al.*, 2000). This is likely to be important in the case of the ets family of transcription factors, contributing to the complexity of their regulation of gene expression. This is exemplified in the interaction of Ets-1 with the tissue restricted AML1 transcription factor, which is regulated in a cell dependent manner (Kim *et al.*, 1999).

Another example of synergistic interaction is the transcriptional activation of the Human T cell lymphotrophic virus type 1 (HTLV-1) LTR by Tax-1, that is controlled by cooperation between NFKB, Ets-1 and Tax-1 (Gitlin *et al.*, 1993; Seeler *et al.*, 1993). Tax/Ets cooperation affects the parathyroid hormone related protein (PTHrP) P2 promoter. The co-operation is performed through a quaternary interaction of Tax, Ets-1 and Sp-1 with the DNA. Tax appears to facilitate the Ets-1/Sp-1 interaction (Dittmer *et al.*, 1997).

In other situations, Ets members are recruited to bipartite binding sites, forming ternary complexes as is the case for Elk-1 and SAP-1. The proteins Elk-1 and SAP-1 are recruited by the serum response factor (SRF) to the serum response element (SRE) to stably contact the EBS within the SRE. Sequences C-terminal to the ETS domain are capable of repressing DNA binding (Dalton and Treisman, 1992; Rao and Reddy, 1992).

An example of heterodimerisation in the Ets family is the GABP complex. This is made up of two of each GABPα and β subunits, when bound to a target such as Herpes simplex virus I immediate early genes (Thompson *et al.*, 1991). Binding is highly dependent on both subunits being present, as GABPα is able to weakly bind on its own but in complex with GABPβ it forms stable DNA contacts. GABPβ is unable to bind on its own yet it makes the

physical contact with DNA once in complex formation (LaMarco *et al.*, 1991; Thompson *et al.*, 1991; Brown and McKnight, 1992).

Oligomerisation of other Ets proteins may also be important, as demonstrated by Elk-1, which self-associates on the c-*fos* SRE (Gille *et al.*, 1996). Intermolecular association of two Elk-1 proteins and two SRF molecules occurs through the C-terminus at the phosphorylation site. It is therefore possible that oligomerisation is more important than previously thought, considering that there are many repeats of multiple EBS on promoters and enhancers (Gitlin *et al.*, 1991; Wasylyk *et al.*, 1991; Seth *et al.*, 1993; Coffer *et al.*, 1994; Seth *et al.*, 1994; Villena *et al.*, 1994).

The Pointed domain is a conserved region found in a subfamily of *ets* factors made up of a monomeric 5-helix bundle, which is implicated in homo-dimerisation (Slupsky *et al.*, 1998). Homo-dimerisation of Pointed leads to a 'permanent signal activity' in signal transduction of the leukemic cells (Golub *et al.*, 1994; Papadopoulous *et al.*, 1995; Romana *et al.*, 1995; Golub *et al.*, 1996; Jousset *et al.*, 1997; Lacronique *et al.*, 1997; Peeters *et al.*, 1997; Ho *et al.*, 1999). ERG is part of this subfamily and is capable of forming homo-dimers or hetero-dimers with other Ets proteins including Fli-1, Ets-2, ER81 and PU.1 via either the Pointed or ETS domains (Carrere *et al.*, 1998).

PU.1 is also known to act as an oncoprotein in erythroid precursors, as its activation causes erythroleukaemia in mice, through its interference with erythroid differentiation, this is thought to occur via its interaction with GATA-1, a zinc finger transcription factor. PU.1 represses GATA-1 transcriptional activation by binding to the zinc finger region of GATA-1, known to bind DNA (Martin and Orkin, 1990). The transactivation domain of PU.1 is also necessary for the repression, though it is not involved in binding GATA-1. A possible mechanism for the repression involves the PU.1 ETS domain mediating the binding to the GATA-1 finger region and the transactivation domain either recruiting other factors or inhibiting activities required for transcriptional

activation. This model is supported by evidence of the interaction of PU.1 transactivation domain with both retinoblastoma protein (Rb) and TATA binding protein (TBP) (Hagemeier *et al.*, 1993). The protein-protein interactions that bind Rb to a promoter is implicated in transcriptional repression (Weintraub *et al.*, 1995).

Ap-1 is a transcription factor that responds to a large range of stimuli making it an important signalling molecule (Wisdom, 1999). Adjacent Ets and Ap-1 binding sites are found in many promoter elements (Wasylyk *et al.*, 1993; Westermarck and Kahari, 1999), making their cooperative effect important for cytokine gene expression (Gottschalk *et al.*, 1993; Wang *et al.*, 1994), MMP expression (Gutman and Wasylyk, 1990; Crawford et al., 1996; Jayaraman *et al.*, 1999) and viral gene expression (Wasylyk *et al.*, 1990; Nothias *et al.*, 1993). Ap-1 and Ets-1 appear to act synergistically as their separate over-expression results in only 20 times transcriptional activation but jointly resulted in 250 times activation at the polyoma enhancer (Wasylyk *et al.*, 1990). Transcriptional activation of the HTLV-1 LTR was also synergistically affected by Ets-1 and the ubiquitous factor Sp-1. Further, it was shown that Ets-1 binds through protein-protein contacts with Sp-1 before contacting the DNA (Gegonne *et al.*, 1993).

A variety of stimuli can induce Ets-1 and Ets-2. For example, the human *Ets-1* promoter may be induced by serum, Ap-1 and Ets-1. *Ets-2* is stimulated by serum, bacterial lipopolysaccharide or protein kinase C (Bhat *et al.*, 1987; Boulukos *et al.*, 1990; Majerus *et al.*, 1992). Ets-1 and Ets-2 may also be phosphorylated, as is the case for Ets-1 in thymocytes undergoing mitogenic stimulation by calcium (Pognonec *et al.*, 1988). An implication of this is that Ets-1 has a role in early T-cell activation. It is also hyper-phosphorylated during mitosis, when transcription is often blocked, suggesting a negative regulation of transcriptional activity.

Competition between ets family members is also likely to play a role in transcriptional activation due to the conservation of the ETS

domain. The synergistic transactivation of MMP-1 is achieved through ERG and Ap-1 interaction, while ERG represses transcriptional activation of MMP-1 by Ets-2. Repression in this case is dependent on the ERG-Ets2 interaction (Buttice *et al.*, 1996; Basuyaux *et al.*, 1997).

As for Ap-1, adjacent binding sites for NF-KB and Ets-1 have been observed in many genes. NFKB is another ubiquitous transcription factor, often having a role in immunity, inflammatory and stress responses (Gilmore, 1999), and binding sites are found in lymphoid genes including IL-2, IL-2 receptor (John *et al.*, 1995), IL-3 (Gottschalk *et al.*, 1993), viral enhancer genes including the HIV1 enhancer (Seth *et al.*, 1993) and signalling molecule genes. Functional cooperation between multiple Ets proteins such as Ets-1, Elf-1 and PU.1 with NF-KB is required for transcriptional activation of HIV I, HIV II (Bassuk *et al.*, 1997) and others.

1.1.7 Ets transcription factors and functionality

Transcriptional Repressor Activity of Ets

To date, most Ets protein studies have focused on transcriptional activation. However, recent reports identify several Ets members with transcriptional repressor activity (O'Neill *et al.*, 1994; Sgouras *et al.*, 1995; Maira *et al.*, 1996; Lopez *et al.*, 1999). The functional roles of the repressor members appear to be much like those of activators. For example, in cell proliferation and differentiation and possibly in regulating this balance. An identical extracellular signal can result in quite different responses, depending on cell type and development stage. Because of the cell dependent response, it is important for correct function that *ets* genes execute the proper signal responses.

Yan has the ability to act as an inhibitor of differentiation in several developmental processes, preventing an incorrect response of neuronal and non-neuronal cell types to extracellular stimuli

(Rebay and Rubin, 1995). It negatively regulates the *sevenless* EGF and FGF receptor pathways as a downstream factor. *Yan* expression in the *Drosophila* developing eye is restricted to undifferentiated cells, being downregulated upon differentiation. A similar expression pattern is seen in other developmental stages.

A further role of Yan may be in inhibiting cells entering the cell cycle without the appropriate signal. Consistent with this is the observation of increased apoptosis at the morphogenetic furrow of the eye disk, effected by an activated Yan, which blocks entry to G1 on receiving a proliferation signal (Rogge *et al.*, 1995).

Cell culture transfection assays have demonstrated the ability of Yan to antagonise Pointed-P2 activity, an Ets family member involved in transactivation (O'Neill *et al.*, 1994; Treier *et al.*, 1995). This antagonism of Pointed proteins and D-Jun activity in developmental systems is usually seen to occur in response to ERK or JNK protein kinase activation (O'Neill *et al.*, 1994; Riesgo-Escovar *et al.*, 1994; Treier *et al.*, 1995). As yet no Yan target genes are known and the mechanism of repression remains unidentified. It is proposed that Yan is phosphorylated by MAPK and degraded in the cytoplasm (Rebay and Rubin, 1995) allowing for activation of its normally downregulated targets, possibly via other *ets* transcription factors such as *Pointed* (Klambt, 1993; Scholz *et al.*, 1993).

Ets-2 repressor factor (ERF) is ubiquitously expressed and forms a subfamily with PE-1 (Klemsz *et al.*, 1994). Its strong transcriptional repressor activity is derived from a C-terminal domain, not found in PE-1. It was identified through its interaction with the H1 element of Ets-2, which it was able to regulate (Sgouras *et al.*, 1995). ERF activity is dependent on the Ras/ERK pathway, as its phosphorylation is required for export of ERF from the nucleus. ERF is normally localised in the cytoplasm, but migrates to the nucleus when ERK is inactivated (Le Gallic *et al.*, 1999). ERF is always phosphorylated prior to the G1 phase of the cell cycle. Sgouras *et al.* proposed that ERF represses transcription of a set of

genes normally activated on G1 phase entry. Hybrids of ERF either with other *ets* genes or with ERF itself are able to inhibit tumourgenic phenotypes associated with abnormal *ets* expression, as for oncogenic transformation of E26, and Fli-1 transformation of EWS in Ewing sarcoma. Therefore, such abnormal expression of ERF repressed genes by overexpressed transactivating Ets proteins may be the cause of tumourgenic activity of many Ets members (Sgouras *et al.*, 1995).

Genes involved in cell proliferation including *Ets-2*, *fos*, *JunB* and *Rb* all contain EBS in their promoter regions. They may be ERF targets, being released from repression when ERF is phosphorylated under mitogenic stimulation and cell proliferation. Protein phosphorylation of ERF and other Ets members may be highly important in regulating gene transcription of G1 phase entry (Sgouras *et al.*, 1995). The ubiquitous expression of ERF suggests it may play a universal role in regulating cell cycle machinery (Mavrothalassitis and Ghysdael, 2000).

TEL was identified by its chromosomal translocation in some chronic myelomonocytic leukemias (Golub *et al.*, 1994). A knockout mouse model (by homologous recombination) has proven TEL to be essential in development, as its deletion resulted in embryo lethality at day E10.5-11.5 (Wang *et al.*, 1997). The DNA binding affinity of TEL is lower than that of an isolated ETS domain, suggesting intramolecular repression is occurring in TEL. TEL and its family group TEL-B/TEL-2 share the B-domain (as for Ets-1 and Fli-1) where it modulates transcriptional activation. Data published to date suggest that TEL mediated repression would depend on interactions with co-repressors, likely to be involved in histone deacteylation. However, this may not be the sole mechanism of repression (Bartel *et al.*, 2000).

Ets function in carcinogenesis

Many Ets family members are known to play important roles in development and progression of cancer. Several mechanisms of Ets involvement in cancers have been elucidated and will be described in further detail.

Fli-1 was first identified as a site for retroviral integration in Friend virus-induced erythroleukemias (Ben-David *et al.*, 1991). *Fli*-1 is now recognised as having a broader role in cell proliferation, differentiation and survival. Genetic mutation of *Fli*-1 often occurs during malignant transformation and progression of certain tumours. Human and mouse *Fli*-1 encode two proteins, p51 and p48, and the gene is localised near the *Ets*-1 gene on mouse chromosome 9 and human chromosome 11. This implies that a common ancestral gene duplication event gave rise to the two *ets* genes (Ben-David *et al.*, 1991; Watson *et al.*, 1992). Fli-1 has 81% homology with Erg-2 of the same subgroup, both containing helix-loop-helix structures in N- and C-terminal ETS domains (Rao *et al.*, 1993). A Fli-1 specific region and the N-terminal ETS domain contain sequences for transcriptional activation. A C-terminus region is also involved in transcriptional activation. *In vitro* studies have shown that this C-terminal region may contain both activation and repression activities. When this region was deleted, an increase in reporter gene transcription was observed (Rao *et al.*, 1993). Human and mouse *Fli*-1 genes show high homology within the promoter region (Barbeau *et al.*, 1996; Dhulipala *et al.*, 1998). Starck *et al.* localised the region of promoter activity and found it contains one GATA and two EBSs (Starck *et al.*, 1999). Mutations in either of the EBS abolished promoter activity and subsequently Spi-1/PU.1 was shown to bind here in the activation of transcription in erythroleukemic cells. Importantly, sequences within intron 1 of *Fli*-1 containing GATA and EBS are capable of increasing transcription from the minimal promoter, and the

factors binding to these regions were identified as Spi-1 and GATA-1 (Barbeau *et al.*, 1999).

Fli-1 is involved in a translocation event associated with the paediatric form of the tumor Ewings Sarcoma. This disease is characterised by fusion of the C-terminal Fli-1 and N-terminal region of an RNA binding protein of the Ewing Sarcoma gene (EWS) (Delattre *et al.*, 1992). Both Fli-1 and EWS/Fli-1 proteins recognise identical DNA binding sequences and retain DNA binding specificity through the C-terminal ETS binding domain. However, the transactivation potential of the EWS/Fli-1 oncogene is much greater than wildtype Fli-1 (May *et al.*, 1993; Bailley *et al.*, 1994; Mao *et al.*, 1994). Point mutations in the Fli-1 C-terminal ETS domain of the oncoprotein abolish DNA binding, but not oncogenic function. From this study, it was concluded that transactivation of the oncoprotein is distinct from its DNA binding capability (Jaishankar *et al.*, 1999).

The EWS/Fli-1 fusion protein exists in two forms, 60% of fusion products are type 1, which include exons 1-7 of EWS and 6-9 of Fli-1. Type 2 also includes the fifth exon of Fli-1. Comparison of the two proteins demonstrates a greater transactivation potential of type 2 yet similar DNA binding potential and associates type 2 patients with a worse prognosis (Lin *et al.*, 1999). Protein interactions with Fli-1 wildtype and EWS/Fli-1 revealed an RNA polymerase II subunit called hsRPB7 binds EWS/Fli-1 but not wildtype. This interaction could be responsible for enhanced promoter activity of EWS/Fli-1 and hence its greater transactivation potential than Fli-1 (Peterman *et al.*, 1998).

The ErbB2 (HER2/neu) receptor tyrosine kinase is overexpressed in human breast and other epithelial cancers as a result of activation of the *Erb*B2 oncogene. Many *in vitro* transfection experiments have described positive and negative regulatory elements on the *Erb*B2 promoter (Hudson *et al.*, 1990; Suen and Hung, 1990; Hollywood and Hurst, 1993; Chen and Gill, 1994;

Grooteclaes *et al.*, 1994; Mizuguchi *et al.*, 1994; Scott *et al.*, 1994; Bosher *et al.*, 1995; Bosher *et al.*, 1996; Chen and Gill, 1996; Chen *et al.*, 1997; Raziuddin *et al.*, 1997). The proximal promoter (220bp) contains the most positive acting regulatory elements including a conserved CCAAT box, non-conserved TATAA box and a conserved EBS adjacent to a polypurine polypyrimidine mirror repeat. Most notably in this region was a single DNase I hypersensitive chromatin site spanning the mirror repeat and EBS (Scott *et al.*, 1994). Mutation of the EBS severely reduces *Erb*B2 promoter activity (Scott *et al.*, 1994; Xing *et al.*, 2000). Many different *ets* transcription factors are expressed in human breast cancer cells including *GABPα, Elk*-1, *Ets*-2, *PEA*3, *ERM, ER81, NET, Elf*-1, *ESX/Elf*-3, *Elf*-5. Of these, *PEA*3, *Elf*-3 and *Elf*-1 have been examined for potential transactivation roles of *Erb*B2 (Benz *et al.*, 1997; Chang *et al.*, 1997; Scott *et al.*, 2000). Transient transfection of *Elf*-1 was shown to stimulate *ErbB2* promoter activity in some breast cancer cell lines in an Ets-specific manner. Also, endogenous Elf-1 protein readily binds to the EBS of the *Erb*B2 promoter, however Elf-1 endogenous mRNA and protein levels are not correlated with co-expression of ErbB2 in tumour and cell lines. PEA3 and Elf-3 do show correlation of mRNA and protein to that of *Erb*B2 (Scott *et al.*, 2000). Ets recruitment to the *Erb*B2 EBS is likely to be cell context dependent and rely on other promoter regulating factors. Triplex formation at the mirror repeat element can inhibit the ETS protein binding at the EBS and consequently prevent *Erb*B2 transcription (Ebbinghaus *et al.*, 1993; Noonberg *et al.*, 1994; Porumb *et al.*, 1996; Ebbinghaus *et al.*, 1999). As well as the negative regulation of the mirror-repeat element, the EBS may control tissue-specific regulation (Chang *et al.*, 1997) and development specific regulation via the EBS and cooperative interactions with other elements such as AP2 (Bosher *et al.*, 1995; Bosher *et al.*, 1996).

PTHrP is expressed in most normal tissues and in aggressive tumours with a malignant metastatic phenotype, particularly of the

breast, pituitary, ovary and prostate (Dittmer *et al.*, 1994; Dittmer and Nordheim, 1998). Ets-1 is likely to be involved in regulation of PTHrP and indeed a relationship between Ets-1 and Sp-1 has been established. Sites for both transcription factors are found in the human PTHrP P2 promoter and the murine PTHrP promoter (Karperien *et al.*, 1997). PTHrP expression as well as Ets-1 expression is induced in endothelial and vascular smooth muscle cells by mitogens (Dittmer and Nordheim, 1998). It therefore follows that Ets-1 is acting upstream of PTHrP, regulating its expression in many tissues and tumours.

Ets function in development

Knockout mice provide a model for studying the role of *Ets*-1 in T and B-cell processes (Bories *et al.*, 1995; Muthusamy *et al.*, 1995). Ets-1 appears to be crucial for the survival and maturation of these cell lineages yet it is not required for their development. Regulation of *Ets*-1 during T cell activation is both transcriptionally and post-translationally controlled (Pognonec *et al.*, 1988; Bhat *et al.*, 1989). Once T cells are activated, a transcriptionally active form of *Ets*-1 is inactivated, which is linked with T cell apoptosis (Pognonec *et al.*, 1988; Bhat *et al.*, 1990; Rabault and Ghysdael, 1994; Bories *et al.*, 1995; Muthusamy *et al.*, 1995).

Ets-1 has other roles, including cell adhesion and endothelial organisation during angiogenesis (Mattot *et al.*, 2000), vascular re-organisation (Grevin *et al.*, 1993; Kola *et al.*, 1993). It appears to not be crucial in embryonic vascularisation as indicated by normal blood vessel development in *Ets*-1 knockout mice (Bories *et al.*, 1995; Muthusamy *et al.*, 1995; Lelievre *et al.*, 2000). However, this may be attributed to functional redundancy with other *ets* genes such as *Erg* and *Fli*-1, which have similar patterns of expression.

Knockout mice models of the *Fli*-1 gene have given insight to its role in hematopoietic development and maintenance (Spyropoulos *et al.*, 2000). It is clearly a key factor in megakaryocyte maturation, as knockout mice have severely reduced megakaryocyte and

derived platelets (Hart *et al.*, 2000; Spyropoulos *et al.*, 2000). It is able to trans-activate the GATA-1 gene, previously shown to be important in erythrocyte maturation (Hromas and Klemsz, 1994). In lymphoid lineages it may be important in the survival and proliferation of a T cell progenitor (Melet *et al.*, 1996), being highly expressed in quiescent T cells but not during T cell activation (Anderson *et al.*, 1999). Transgenic mice over-expressing *Fli-1* show a hyper-responsiveness to mitogenic signals and reduced apoptosis (Zhang *et al.*, 1995).

Ets function in tissue and organ development

The extracellular matrix (ECM) is an important feature of cell structure, providing support as well as an interaction for cell function. The ECM is synthesised, deposited and degraded throughout embryogenesis, the female reproductive cycle, angiogenesis and wound repair, making the balance of synthesis and degradation crucial for normal cell function. When this balance is upset, pathological conditions may arise and *ets* factors have been implicated in this as they play an important role in the transcriptional activation of ECM degrading targets such as serine proteases, matrix metalloproteinases and their inhibitors (TIMPS) (Trojanowska, 2000). *Ets* factors may also be involved more widely, including the targeting of matrix proteins like tenascin, collagen, fibronectin and others.

Tissue remodelling is an important part of wound healing which implicates a role for Ets factors via matrix metalloproteinases (MMPs). Some MMPs, for example urokinase plasminogen activator (uPA) and tissue-type plasminogen activator (tPA), are induced in this process (Yang *et al.*, 1997). Migrating keratinocytes express collagenase 1 (MMP-1), stromelysin-2 (MMP-10) and gelatinase (MMP-9), uPA, tPA and are involved in re-epithelisation and are downregulated at termination of this process (Parks, 1999). Dermal cells also express a range of MMPs, and expression is continued

through repair to scar formation. A different subset of MMPs is expressed by immune cells in the inflammamatory response, indicating specificity of expression for the different processes (Trojanowska, 2000).

1.1.8 Matrix Metalloproteinases and the Extracellular Matrix

Matrix metalloproteinases (MMPs) are a family of proteolytic enzymes responsible for the degradation of the extracellular matrix (ECM). They are usually grouped according to substrate specificity and include the collagenases, gelatinases and the stromelysins. The expression of the MMP genes is usually tissue-specific, except for gelatinase A, which is widely expressed. Of particular interest are stromelysin-1 and collagenase, which are expressed primarily in stromal fibroblasts. The type of collagenase discussed here is the interstitial form in humans (MMP-1), which does not have a murine homologue, but murine collagenase-3 (MMP-3) is expressed in a similar manner to that of human MMP-1.

The ability of MMPs to breakdown the matrix proteins implicates them in tumourigenesis, as structural barriers of basement membranes and stromal ECM must first be weakened for tumour formation (Stracke *et al.*, 1994). Most MMP promoters contain what is referred to as the 'oncogene-responsive-unit', which is also found in many other genes promoter and enhancer regions (Gutman and Wasylyk, 1991). It consists of an AP-1 element which binds dimeric combinations of the Fos and Jun family of oncoproteins and also one or more PEA3 elements capable of binding members of the Ets family of oncoproteins. Specific interactions between these cis-acting elements may therefore be important in controlling positive and negative regulation of MMP tissue restricted patterns in both normal and neoplastic tissues (Crawford and Matrisian, 1996). Indeed, mutations in either the AP-1 or PEA3 elements of interstitial collagenase promoter reduce basal activity and

responsiveness of the promoter to external stimuli (Gutman and Wasylyk, 1990; Wasylyk *et al.*, 1990; Westermarck *et al.*, 1997).

Evidence for MMPs role in tumourigenesis has come from several MMP knockout mice. When MMP-7 was removed, a reduction in intestinal tumours was observed (Wilson *et al.*, 1997), MMP-2 deficient mice exhibited a reduction in angiogenesis and tumour progression (Itoh *et al.*, 1998), and MMP-11 knockout mice show reduced tumourigenesis when chemical mutagens are administered (Masson *et al.*, 1998).

The role of ets in this tumourigenesis is apparent through the co-expression of ets with matrix degrading proteases often seen in invasive tumours, whilst it is notably absent from non-invasive or benign tumours (Calmels *et al.*, 1995). Ets-1 coexpression with collagenase 1 (MMP-1), urokinase-type plasminogen activator (uPA) and stromelysin (MMP-3) is known to occur in lung carcinomas (Wernert *et al.*, 1994; Bolon *et al.*, 1995) and skin angiosarcomas (Naito *et al.*, 1994). Also, the role of Ets-1 in regulating MMPs in tumours is evidenced by a single nucleotide polymorphism in the MMP-1 promoter, creating an EBS that increases Ets-1 binding and enhances promoter activity role (Rutter *et al.*, 1998).

E1A-F, the human homologue of PEA3, was found to activate MMP-1, MMP-3 and MMP-9 and was also found to be involved in the induction of tumourgenesis in MCF-7 cells (Kaya *et al.*, 1996). Further, Ets-1, Ets-2 and PEA3 overexpression are all found to increase activity of MMP-1,-3 and –9 promoters (Buttice *et al.*, 1996; Gum *et al.*, 1996; Westermarck *et al.*, 1997). A further study using antisense oligonucleotides to E1A-F in a squamous cell carcinoma resulted in the inhibition of both MMP-9 expression and invasiveness (Hida *et al.*, 1997; Hanzawa *et al.*, 2000).

The synergistic interaction between Ets and AP-1 factors has been well documented. Specifically in the case of MMPs, the Ets factors Ets-1, Fli-1 and PU.1 are known to differentially regulate AP-1 dependent MMP-1 promoter activation (Westermarck *et al.*, 1997). In this study, the effects of each Ets factor were observed in

NIH3T3 fibroblasts. Ets-1 was shown to increase the activity of MMP-1 alone and greatly increase activity in conjunction with c-Jun and JunB. In contrast, neither PU.1 nor Fli-1 had any effect alone, while Fli-1 was able to enhance the effect of c-Jun. However, PU.1 inhibited MMP-1 induction by both c-Jun and JunB. MMP-1 5' deletion constructs showed that in the case of PU.1 inhibiton, the AP-1 site was necessary (Westermarck *et al.*, 1997). In another study, the Ets factor Erg was shown to interact with Jun/Fos and potentiate the induction of MMP-1 promoter activity (Angel and Karin, 1991). Further to this, Buttice *et al.* showed that Ets-2 activates both MMP-1 and MMP-3, while Erg only induces MMP-1. Erg inhibits this induction of MMP-3 by Ets-2. Erg was shown to be recruited by Fos/Jun in the transcriptional activation of MMP-1 but not MMP-3 (Buttice *et al.*, 1996).

This AP-1 dependent regulation, which seems to be different for different Ets members reflects the importance of promoter specific sequences and protein-protein interactions. Another example of the specific sequence requirement is seen in the case of CREB binding protein (CBP) and p300, which bind to Ets-1 (Yang *et al.*, 1998). CBP/p300 cooperates with Ets-1 and Ets-2 but not PEA3 or Erg2 in activation of MMP-3 promoter (Buttice *et al.*, 1996; Jayaraman *et al.*, 1999).

An analogous situation to the MMP control by Ets is seen for the TIMP-1 promoter, which contains an EBS adjacent to an AP-1 site (Edwards *et al.*, 1992; Logan *et al.*, 1996). C-Ets-1 and AP-1 complex was capable of activating transcription of TIMP-1 in an F9 embryonic carcinoma cell line (Logan *et al.*, 1996). Interestingly, EHF (Ets homologous factor) appears to have tumour suppressor activity as it represses Ets-2 stimulation of MMP-1 and MMP-3 promoters (Kleinbaum *et al.*, 1999).

Tumour studies have shown expression of Ets factors are not aligned with the ECM enzyme. Ets-1 for example is localised to endothelial and fibroblast stromal cells (Wernert *et al.*, 1994) as opposed to the tumour cell. Stromal fibroblasts may therefore play

a role in the synthesis of matrix proteases thereby supporting tumour growth (Trojanowska, 2000).

Antisense Ets-1 oligonucleotides were transfected to a glioma cell line in which Ets-1 and uPA expression was inhibited and glioma cell migration and invasion was stopped (Kitange *et al.*, 1999). This suggests an involvement of Ets-1 in induction of protease enzymes and hence tumourigenesis. Further studies supporting this include the stable expression of the Ets-1 DNA binding domain in mouse epithelial mammary cells of both normal and cancerous lines, which inhibited expression of uPA and reduced cell migration and invasion in both cell types (Delannoy-Courdent *et al.*, 1998).

The data in support of the role of Ets factors in the function both of the normal and neoplastic mammary gland is overwhelming. Normal ets expression in the mammary gland contributes to its development as early as primary epithelial bud formation in the embryo and Ets-1 is detected later in epithelial duct extremities. Ets-1 expression also coincides with invasiveness (Delannoy-Courdent *et al.*, 1998). During tubulogenesis in normal and invasive cells, Ets-1 is co-expressed with plasminogen (uPA). In invasive cells, Ets-1 activates uPA leading to proteolytic events to degrade ECM and hence promote tumour metastasis (Delannoy-Courdent *et al.*, 1996). The PEA3 subfamily shows a high level of expression in the developing embryo mammary gland and during tubular morphogenesis later in mammary development (Chotteau-Lelievre *et al.*, 1997). As already mentioned, E1A-F (the human form of PEA3) was able to upregulate the transcription of MMP1, 3 and 9 (Higashino *et al.*, 1995), and this conferred an invasive phenotype on human breast cancer cells (Kaya *et al.*, 1996). These observations suggest that PEA3 members may take part in epithelial-mesenchymal interactions by regulating MMP gene expression and in turn the remodelling of the ECM (Chotteau-Lelievre *et al.*, 1997). Abberant forms of PEA3 are also shown to be associated with breast cancer in both mice (Trimble *et al.*, 1993;

Chotteau-Lelievre *et al.*, 1997) and humans (Baert *et al.*, 1997; Xing *et al.*, 2000). Trimble *et al.* (1993) observed high PEA3 RNA expression in mouse mammary tumours, but low expression in surrounding mammary epithelium. Also, when mammary tumours metastasised to the lung, the lung tumour showed an over-expression of PEA3 that was not detected in normal lung tissue (Trimble *et al.*, 1993).

Ets-2 studies reveal it has transforming potential in the mammary epithelia (Galang *et al.*, 1996). Over-expression of Neu is a feature identified in approximately 20% of breast tumours. Galang *et al.* show that expression of a mutant form of Neu (mutated to become constitutively active) causes transcriptional activation of Ets, Ap-1 or NFKB dependent reporter genes. This activation is mediated by the Ras signalling pathway. The Neu mutant mediates Ets-2 transactivation activity by phosphorylating a specific threonine residue. Ets activation was also shown to be required for Neu-mediated cellular transformation (Galang *et al.*, 1996). Similar experiments have shown that Ets-2 mediates the regulation of anchorage dependent growth and cellular invasiveness of neoplastic mammary epithelial cells (Sapi *et al.*, 1998). Also, a deleted *Ets*-2 gene reduced mammary gland tumour growth in mice carrying the polyoma virus middle T oncogene (Neznanor *et al.*, 1999). The reduction in size of the tumour was correlated with a more differentiated state of early hyperplastic growths. Ets-2 was thought to regulate the progression of these tumours (Neznanor *et al.*, 1999).

Tissue Inhibitors of Metalloproteinases

MMP activity can be specifically inhibited by tissue inhibitors of metalloproteinases (TIMPs) that bind to the zinc binding site of MMPs. The regulation of MMPs important in tissue remodelling via TIMPs may also involve the use of Ets factors. It is also likely that TIMPs themselves are regulated in part by Ets factors.

There are four members of this family, TIMP-1, -2, -3, -4. TIMPS are potentially responsible for changes in the morphology and growth stimulation of many cells (Gomez *et al.*, 1997). Specifically, TIMP-1 and –3 are anti-angiogenic (Gomez *et al.*, 1997) and TIMP-2 is involved in MMP-2 activation (Murphy and Knauper, 1997). The roles of TIMPs that have been documented are somewhat contradictory, which may be a result of several independent roles, including inhibition of MMPs and also involvement in the cell-surface-targeted MMP activation cascades (Gomez *et al.*, 1997). TIMP-3 can induce apoptosis in both normal and malignant cells (Bian *et al.*, 1996; Ahonen *et al.*, 1998; Baker *et al.*, 1998), implicating a further role in the survival of malignant cells. Structurally, the *TIMP*-1 promoters of human, mouse and rat homologues all lack a TATA box but contain a conserved 22bp SRE at 75bp upstream of the major transcription start site (Gewart *et al.*, 1987; Campbell *et al.*, 1991; Bugno *et al.*, 1995; Clark *et al.*, 1997; Bahr *et al.*, 1999). The SRE is critical for promoter responsiveness to growth factors, cytokines and viruses. It contains elements for AP-1, STATs and PEA3 transcription factors. The AP-1 site is critical for both basal and inducible transcription (Campbell *et al.*, 1991; Edwards *et al.*, 1992; Bugno *et al.*, 1995; Logan *et al.*, 1996; Clark *et al.*, 1997).

1.2 *BrETS*, *Elf*-5 and *Ese*-2

1.2.1 *BrETS*

The initial discovery of the *BrETS* gene resulted from a screen of bovine mammary cDNA clones (L'Huillier *et al.*, unpublished). A partial cDNA sequence was isolated and sequenced and a subsequent blast search revealed homology to the Ets family of proteins. Later, the mouse homologue was identified and the cDNA sequence elucidated (L'Huillier *et al.*, unpublished). Northern blot analysis demonstrated that two transcripts were expressed in

murine mammary cells, of approximately 1.5kb and 2.5kb, as seen in Figure 1.3. The transcripts were later cloned and sequenced (L'Huillier *et al.*, unpublished). Characterisation of the two transcripts revealed alternate polyadenylation sites, which were shown to be responsible for the 1.5 and 2.5 kb mRNA species (Langley, 1999).

Analysis of the murine cDNA revealed an open reading frame of 765 bp representing a 255 amino acid protein. The ETS domain was found to be located in the C-terminus at positions 161-242 and a conserved B domain was found at the N-terminus at positions 47-118 amino acids. The B domain may be associated with protein-protein interactions with the serum response factor, as in Elk-1 (Ling *et al.*, 1997).

The expression of *BrETS* was studied in a variety of tissues. Of these, high expression was seen in the mammary gland (see Figure 1.3) with less expression detected in the salivary gland (L'Huillier *et al.*, unpublished). No expression was detected in other tissues studied including liver, lung, kidney, spleen, heart or brain. This pattern of expression was later found to be more widespread (Zhou *et al.*, 1998), possibly due to a weakly hybridising probe and the use of total RNA in the experiments by L'Huillier *et al.*

With the highly specific expression in the mammary gland, the temporal expression pattern was then examined in the same tissue (L'Huillier *et al.*, unpublished). Temporal analysis of mammary gland tissues from virgin, pregnant, lactating and involuted mice revealed strong expression of BrETS from day 6 of pregnancy, throughout pregnancy and lactation with a sharp decline at day 2 of involution, as indicated in Figure 1.4.

1.2.2 *Elf*-5

In 1998, Zhou *et al.* published their findings on a novel ETS factor, which they termed *Elf*-5 of the Ets-like-factor subfamily (Zhou *et al.*, 1998). The gene was identified in both mouse and human, and

appears to be identical to *BrETS*. Its Human chromosomal position was localised to 11p13-15, a region that frequently undergoes loss of heterozygosity in carcinomas of the heart, kidney and prostate. The expression pattern of poly(A)+ mRNA *Elf*-5 was shown to be highly specific to epithelial tissues and was detected in mouse tissues including the lung, kidney, stomach, ovary, tongue, bladder and both 2-day and 10-day pregnant mammary gland. It was not evident in the liver, heart, small intestine, spleen, thymus, pancreas, skeletal muscle, colon or fat samples. Interestingly, a differential pattern of the two transcripts was observed in some tissues in the mouse, with the 2.5kb transcript predominantly expressed in the neonatal and embryonic lung, kidney and adult ovary. The 1.5kb transcript was more abundantly present in the adult tongue and developmental stages of the stomach.

In the human adult, the 2.5kb transcript was expressed more strongly in the kidney and prostate whilst being low in the placenta and lung. Of several human cancer lines tested, only T47D (a breast tumour cell line) expressed Elf-5, which may be due to re-arranged alleles.

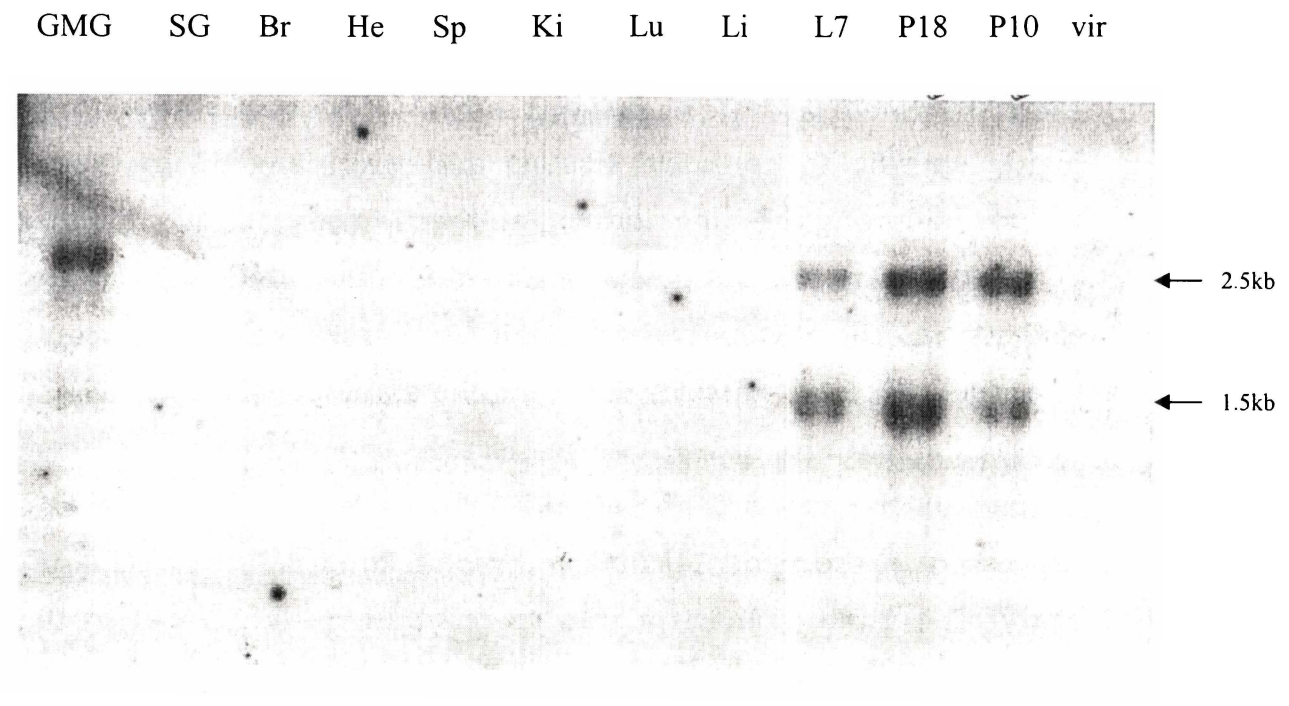


Figure 1.3 Spatial expression pattern of *BrETS*. Tissues are from mouse if unspecified. Two isoforms are indicated by arrows at the right of the diagram. Abbreviations are as follows: GMG/ goat mammary gland, SG/salivary gland, Br/brain, He/ heart, Sp/ spleen, Ki/ kidney, Lu/ lung, Li/ liver, L7/ 7 day lactating mammary gland, P18/ 18 day pregnant mammary gland, P10/ 10 day pregnant mammary gland, vir/ virgin mammary gland. (L'Huillier *et al.*, unpublished)

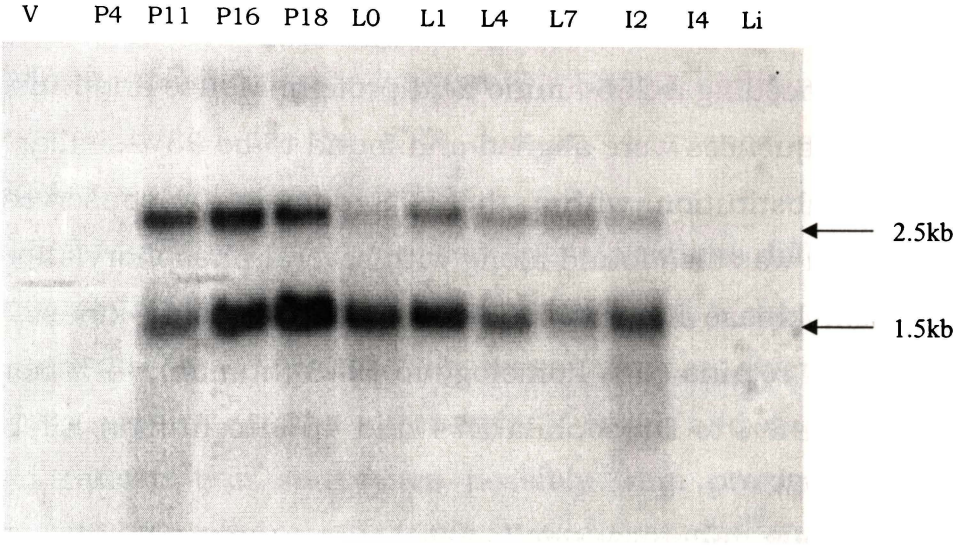


Figure 1.4 Temporal expression pattern of murine *BrETS*. *BrETS* expression was examined by Northern blot analysis, throughout the stages of mouse mammary gland development. Abbreviations: V/virgin, P4/pregnant day 4, L0/lactating day 0, I2/involuted day2, Li/liver. The two transcript sizes are shown to the right of the figure. (L'Huillier *et al.*, unpublished)

Murine *Elf*-5 cDNA was found to be 1437bp in size, encoding an ORF of 759 bp corresponding to a 253 amino acid protein. The two transcript sizes were given as 2224bp and 1528bp. The human *ELF*-5 cDNA isolated from a lung library revealed one transcript only, encoding a 255 amino acid protein. Mouse and human amino acid sequences were aligned and found to be 95% similar with only one substitution within the ETS domain. A conserved Pointed domain was also found along with several phosphorylation sites for casein kinase II, protein kinase C and tyrosine kinase. The ETS domain retains 67% homology to Elf-3 (human), 49% homology to NERF, 48% to Drosophila E74 and 46% to human Elf-1 (Zhou *et al.*, 1998).

Phylogeny relationships show that *hELF*-5 and *mElf*-5 are orthologous genes (traced back to a speciation event) and that *ELF*-3 and *ELF*-5 are paralogous (traced back to gene duplication). A recent study of *Elf*-3 and *Elf*-5 examined the ability of these genes to control Whey Acidic Protein (WAP) transcription via a MAF-1 (Ets-1) site and an adjacent Ets-2 site (Thomas *et al.*, 2000). *Elf*-3 is expressed in the virgin mammary gland and in early pregnancy when cells are proliferating and declines in late pregnancy when epithelial cells terminally differentiate and produce milk. Its expression is low throughout lactation but increases again at involution when apoptosis of the epithelia occurs. Its expression is restricted to epithelial cells in the ducts and lobular structures of the mammary gland. Both Elf-3 and Elf-5 were able to activate a WAP promoter construct in mammary cell lines, though Elf-3 is more potent than Elf-5. Both were shown to be independent of the lactogenic hormones. Point mutations of either of the *Ets*-1 or -2 sites inactivated the transactivation capability. Mutations in both Ets sites simultaneously had only a slightly greater effect, implying a cooperative interaction between Ets factors.

1.2.3 *ESE-2*

Simultaneous to the *Elf*-5 publication another group Oettgen *et al.*, (1999), published findings of the novel *ets* factor human *ESE*-2 of the epithelium specific factor subfamily. ESE-2 contains 65% homology with ESE-1 in the ETS domain. Sequence analysis revealed *ESE*-2 is identical to *ELF*-5.

Human *ESE*-2 was found to express two splice products differing in the 5' sequence and the cDNA lengths were found to be 2451bp (ESE-2a) and 2317bp (ESE-2b). The reason for the divergent 5' sequence was predicted to be alternative 5' exons and implicated alternative transcription start sites possibly from proximal and distal promoters (Oettgen *et al.*, 1999). Polyadenylation sites were found along with an ATTTA motif thought to be involved in mRNA turnover (Savant-Bhonsale and Cleveland, 1992). In the human, no transcript was found homologous to the mouse 1.5kb transcript.

In the human adult a highly restricted pattern of expression of *hESE*-2 was found. This included tissues with high epithelial cell content such as kidney, prostate, small intestine, colon, ovary, pancreas, liver and placenta, and throughout the gastrointestinal tract. The highest expression was found in the salivary gland followed by the mammary gland, fetal kidney and trachea. Human ESE-2 retains the C-terminal ETS domain and N-terminal Pointed domain. Differential expression of the two transcripts was observed with the kidney only expressing *ESE*-2a and the prostate expressing higher levels of *ESE*-2b than *ESE*-2a.

Both *ESE*-1 and *ESE*-2 are expressed in differentiated but not undifferentiated keratinocytes. *In vitro* analysis shows *ESE*-2 is induced later during differentiation than *ESE*-1, which may allow for temporally different activation of different epithelium specific genes. ESE-1 and –2 regulate the two prostate specific genes PSA and PMSA oppositely. ESE-1 represses PSA while ESE-2 activates it. The opposite occurs for PSMA activation. ESE-2 is likely to be critical in the regulation of PSP (a salivary gland gene) with a

different effect from ESE-1, which is also expressed in the salivary gland.

Examination of the DNA binding properties of ESE-2 reveals it has an N-terminal regulatory domain that inhibits DNA binding and this is not present in ESE-1 (Oettgen *et al.*, 1999).

1.3 Objectives of this research

In this project the overall objective was to characterise the promoter region of the mouse *Ese*-2 gene. Specifically, this involved the mapping of genomic clones from a mouse 129 library and the determination of the orientation of each clone. Appropriate 5' regions were sub-cloned, sequenced and characterised. Computer analyses were performed to identify potential transcription factor binding sites and to locate any other typical promoter features such as a TATA box. The transcription start site(s) were then identified using primer extension and the sequence verified by 5' RACE. Subsequent to this initial analysis a functional characterisation of the *Ese*-2 promoter was performed. The approach involved the use of deletion constructs of the proximal 5' region attached to the reporter construct Chloramphenicol acetyltransferase gene (CAT). These were transfected both stably and transiently into cell lines including mouse mammary lines HC11, Comma1D and a human breast tumour line T47D. Further to this, any regions of apparent importance were searched further with the goal of identifying specific sites of regulation.

The localisation of 5' promoter and regulatory elements would aid in the understanding of how the gene is regulated in the context of the mammary gland. This would also be useful in allowing the use of site directed mutagenesis to artificially control the regulation of the gene.

Chapter 2. Materials and Methods

2.1 Materials

2.1.1 Growth Media

(i) Luria broth

Suspend 25g of LB Base in 1l of distilled water and autoclave for 15min at 121°C. Millers Luria Broth Base was purchased from Invitrogen.

(ii) Luria agar

Suspend 37g of Millers LB agar (Invitrogen) in 1l of distilled water. Heat to boiling with stirring to dissolve and autoclave for 15min at 121°C. Cool to 50°C and pour 20ml each into sterile petri dishes.

(iii) Terrific broth

To 37g of Terrific Broth base (Invitrogen) add 1l of distilled water. Autoclave for 15min at 121°C.

(iv) S.O.C. media

Dissolve 2g of bactotryptone, 0.5g of yeast extract, 1ml of 1M NaCl and 0.25ml of 1M KCl in 97ml of distilled water. Once dissolved, auotclave and cool. Prepare a stock solution of 2M Mg^{2+} from 1M $MgCl_2.6H_2O$ and 1M $MgSO_4.7H_2O$ and filter sterilise. To prepare media, make up to 20mM Mg^{2+} by adding 1ml of the 2M Mg^{2+} stock solution. Finally add 1ml of a 2M glucose stock solution to a final

concentration of 20mM. Filter the media through a 0.2µm filter unit.

(v) DMEM-F12 media

Dissolve one packet of DMEM-F12 media (Invitrogen) in 1l of distilled water. Add 1.2g sodium bicarbonate and stir until completely mixed. Adjust the pH to 7.1 using 5M NaOH and stir for a further 20min. Filter the media in a sterile hood using a vacuum filter unit. Store at 4°C.
For HC11 and Comma1D cell media, supplement 180ml of the above stock media by adding 8% fetal calf serum (FCS), insulin (1mg/ml), epidermal growth factor (EGF) (10ng/ml), bovine serum albumin (BSA) (300µg/ml) and antibiotics (penicillin/streptomycin 10mg/ml). Store the media at 4°C.

For T47D cells, prepare MEM-α by adding one packet of powdered media (Invitrogen) to one l of distilled water and 2.2g of sodium bicarbonate. Stir until dissolved and filter sterilise. Supplement 180ml of media with 20ml FCS, 2µg/ml insulin, and 1µg/ml hydrocortisone.

2.1.2 Bacterial strains

Cell strains used in this study are listed in Table 2.1. Glycerol stocks of cells were stored in 50% glycerol at -80°C.

Table 2.1 Bacterial strains

Species	Strain	Genotype	Source
E.coli	STBL2	F-	Invitrogen
E.coli	XL1-Blue	MRF'	Stratagene

2.1.3 Eukaryotic cell lines

Cell strains used for cell culture experiments are listed in Table 2.2. Stocks of cells were stored in 10% DMSO/25% FCS/ 65% media in liquid nitrogen.

Table 2.2 Eukaryotic lines

Species	Cell line	Originating tissue	Source
Mouse	Comma1D	Mammary gland, mid-pregnant	J. Demmer, AgResearch
Mouse	HC11	Derivative of Comma1D	J. Demmer, AgResearch
Human	T47D	Breast cancer cell line	P. L'Huillier, AgResearch

2.1.4 Plasmids

Plasmids used in this study are listed in Table 2.3. Plasmid DNA was stored in water or TE buffer at -20°C.

Table 2.3 Plasmids and bacteriophage

Plasmid	Description	Source
pBLCAT2	4.5kb, Amp^r, CAT, SV40 poly A, HSV tk promoter	Promega
pBluescript KS+	2958bp, Amp^r, T3 and T7 RNA pol initiation sites, LacZ	Stratagene
pBluescript SK+	2958bp, Amp^r, T3 and T7 RNA pol initiation sites, LacZ	Stratagene

pCH110	7128bp, Amp^r, LacZ, SV40 early promoter	Promega
pEGFP-N1	4.7kb, Kan^r/Neo^r, EGFP, SV40 poly A	CLONTECH
pGEM3	2967bp, Amp^r, SP6 and T7 RNA pol initiation sites	Promega
pGEM-TEasy	3015bp, Amp^r, lacZ, T7 and SP6 RNA pol initiation sites	Promega
Lambda Phage FIX II Library	*red* and *gam* genes, T3 and T7 pol initiation sites	Stratagene

2.1.5 Oligonucleotides

The oligonucleotides used in this work were purchased from Invitrogen. The sequences of the oligonucleotides are listed in Table 2.4. Start position is given relative to the major transcription start site at +1 and the second column describes the method the oligonucleotide was used in.

Table 2.4 Oligonucleotide sequences

Oligonucleotide, start position for *Ese*-2	Method	Sequence 5' – 3'
1.11(ATG), +59	Southern blot analysis	ACCTAGCCACCACTTGTCTT CACGGTGATG
1.11(1), -22	Southern blot analysis	CTACAGGTGCCTTTATTTCT
1.11(2), +176 cDNA	Southern blot analysis	ATCTGTTCAGCAATG

5' seq., +1	Sequence, and Southern blot analysis	GGTCCGCTGGTGCTGGG
AAP, N/A	5' RACE	GGCCACGCGTCGACTAGTAC GGGIIGGGIIGGGIIG
Exon2 ATG mod. *Bgl*II, +6136	PCR	GTTACGGAGATCTACAACACC CTGGAATATC
Exon2 ATG mod., +6136	PCR	GTTACGGACTCGAGCAACAC CCTGGAATATCAGAG
Exon2 ATG mod.2, +6136	PCR	GTTACGGAGCTCGAGATCAC CCTGGAATATCAGAG
Far 5' #104, -2561	PCR	GATCATCCAAACATCCATT
Far 5' #146, -2519	PCR	GATCAAATGTGATGCTTACGC TTAGAG
Far 5' #30, -2635	PCR	TTCCTGCCCTTCCCACTT
Far 5' #50, -2615	PCR	CAGCTCATCTTCACCCACAG
Far 5' #84, -2581	PCR	TTCCCATAGCTTAGTGTTTG
Far 5' *Sac*I, -2665	PCR	GAGCTCAGTTCTTCCCCCACC ACCC
Far 5' *Sph*I mod., -2528	PCR	TCCGTCTTTGTTACTTTGTGAA TC
Far 5' *Xba*I, -1805	PCR	CAGTCAATCTCTAGAACCCA
Forward +1, -59	Sequence analysis	AACGGACCTGTCTGTAGG
GSP1, +442 cDNA	5' RACE	TACAGGTACTCCCCACAGA
GSP2, +366 cDNA	5' RACE	GCTGCAGGCCGGTGATGTTG AAGTGA
Intron1 2700,+2926	Sequence analysis	CTTCTGCCATTGAGA
Intron1 4200,+4181	Sequence analysis	GATGATGCTATACCC

Intron1 5500,+5433	Sequence analysis	CACAGAATGCTACAG
Intron1 6000,+6016	Sequence analysis	GTGTAAAGCACTTGA
Intron1 770, +855	Sequence analysis	CCAAAGACTCAGACC
PE 35, +46	Primer extension	CCTTCTGTCTGGACCTAGCCA CCACTTGTCTTCAC

2.1.6 Chemicals, enzymes and radioisotopes

(i) Chemicals

Chemicals used in this study were purchased from one of the following suppliers:
Sigma Chemical Co., St Louis, MO, USA
Roche Diagnostics NZ Ltd, PO Box 62-089, Akl
Invitrogen NZ Ltd, PO Box 12-502, Akl
Progenz Ltd, PO Box 180-029, Akl
Amersham Pharmacia Biotech, PO Box 56-364, Akl
Biolab Scientific Ltd, Private Bag, Akl
Bio-Rad Laboratories PTY Ltd, PO Box 300-571, Akl
Global Science and Technology Ltd, PO Box 101-253, Akl

(ii) Enzymes

Restriction enzymes were purchased from Roche Diagnostics Ltd. Other enzymes were purchased from the suppliers as listed in Table 2.5.

Table 2.5 Enzymes

Enzyme	Source
DNA Polymerase I (Klenow fragment)	Invitrogen
DNaseI	Roche
Elongase PCR system	Invitrogen
Expand PCR system	Roche
Ribonuclease A	Invitrogen
RNase H	Roche
RNase T1	Invitrogen
Superscript II RNase H-RT	Invitrogen
T4 DNA ligase	Roche
T4 DNA ligase (pGEM-Teasy)	Promega
T4 DNA polymerase	Roche
T4 polynucleotide kinase	Invitrogen
T7 RNA polymerase	Promega
Taq DNA polymerase	Roche
Terminal transferase	Invitrogen

(iii) Radioisotopes

The radioisotopes used in this study were purchased from Amersham, and are listed in Table 2.6.

Table 2.6 Radioisotopes

Radioisotope	Specific activity	Use
[γ-^{32}P]-dATP	3000 Ci/mmol	Labelling oligonucleotides
[α-^{32}P]-dCTP	3000 Ci/mmol	Labelling cDNA by random priming
^{14}C-Chloramphenicol	50µCi	CAT detection

2.1.7 Solutions

The solutions used in the study are listed in Table 2.7.

Table 2.7 Solutions

Solution	Composition
10X MOPS	41.8% MOPS 50mM NaOAc 10mM EDTA
Alternate DNA Loading dye 6X	0. 25% bromophenol blue 0.25% xylene cyanol 30% glycerol
Dispase	9.6mg/ml neural protease from Bacillus polymyxa (Roche Diag.)
DNA denaturing solution	1.5M NaCl 0.5M NaOH
DNA loading dye 10X	15% ficoll 0.25% bromophenol blue 0.25% xylene cyanol 1M EDTA pH 8.0 1% SDS
DNase STOP buffer	20mM Tris-Cl pH8 10mM EDTA 1% w/v SDS 600mM NaCl 400µg/ml Proteinase K

Formaldehyde loading dye (2X)	50% deionised formamide 20% deionised formaldehyde 10% MOPS (10X) 40μg/ml ethidium bromide 0.02% bromophenol blue 0.02% xylene cyanol 1mM EDTA pH 8.0 5% glycerol
Formamide loading dye (STOP dye)	95% deionised formamide 20mM EDTA pH 8.0 0.05% xylene cyanol 0.05% bromophenol blue
Hybridisation buffer (Church and Gilbert)	7% SDS 1mM EDTA 0.25 M Na_2HPO_4 pH 7.2
Plasmid DNA extraction: Solution I	 50mM glucose 10mM EDTA 25mM Tris-Cl pH 8.0
RSB buffer	100mM Tris-Cl pH7.4 10mM NaCl 3mM $MgCl_2$
RSB buffer with NP-40	100mM Tris-Cl pH7.4 10mM NaCl 3mM $MgCl_2$ 0.02% Nonidet P-40
SM buffer	50mM Tris-Cl pH7.5 100mM NaCl 8mM $MgSO_4$ 0.01% gelatin

Solution A (cell culture)	1.8g/l glucose 7.6g/l NaCl 0.22g/l KCl 0.14g/l Na_2HPO_4 7.15g/l Hepes
Solution II	0.2mM NaOH 1% SDS
Solution III	3M potassium acetate 2M acetic acid
Solution Z	PBS 3.6µl/ml BME
SSC (20X)	3.0M NaCl 0.3M sodium citrate pH 7.0
TAE buffer	40mM Tris-acetate 2mM EDTA pH 8.0
TB buffer	10mM Pipes 55mM $MnCl_2$ 15mM $CaCl_2$ 250mM KCL
TBE buffer	80mM Tris-borate 8mM EDTA pH 8.0
TCA (100%)	500g trichloroacetic acid 227ml H_2O
TE buffer	10mM Tris-Cl 1mM EDTA pH 8.0
Trypan Blue	PBS 0.16% Trypan blue
Trypsin/EDTA	0.25% Trypsin 10mM EDTA 5mg Phenol Red 1X TBS

Urea 47%	94g urea 200ml H_2O

2.2 Methods

2.2.1 Bacterial growth

(i) Agar plates

Transformed cells were spread onto LB agar medium containing the appropriate antibiotic. Plates were grown for 12-16hrs at 37°C until colonies were approximately 1mm in diameter.

(ii) Liquid cultures

Using a sterile toothpick or loop, a single colony was transferred to a 3ml (minipreparation) or 100ml (maxipreparation) LB or TB medium in a 2l flask with the appropriate antibiotic. Cultures were incubated at 37°C on a rotary shaker (300rpm) until an OD_{600} of 0.4 was reached.

2.2.2 Preparation and transformation of chemically competent cells

Preparation of XL1-Blue cells followed the method described by Inoue *et al.* (1990).

XL1-Blue cells were grown overnight on LB agar plates with the appropriate antibiotic(s), ten colonies of 2-3mm diameter were used to inoculate 250ml Terrific Broth culture in a 1l flask. This was grown at 18°C with shaking (225rpm) to an OD_{600} of 0.6. The cells were then chilled on ice for 10min, centrifuged at 2500xg for 10min at 4°C, and the pellet resuspended in 80ml of ice cold TB. The

suspension was chilled on ice for a further 10min, centrifuged for 10min at 2500xg 4°C and resuspended in 20ml of TB plus 1.5ml DMSO (7%). After another 10min incubation on ice, the cells were aliquoted and frozen in liquid nitrogen. Competent cells were stored at -80°C.

For heat shock transformation of XL1-Blue MRF' cells, plasmid DNA or ligation mixture (1-20ng DNA) was added to 80µl of competent cells. Ligations containing low melting point agarose were diluted to <0.02% agarose before addition of competent cells. The cells and DNA were incubated on ice for 30min and then heat shocked at 42°C for 1min. Cells were snap cooled on ice and 500µl of LB was added. The transformed cells were then incubated at 37°C for 60min before plating onto LB agar and antibiotic, and were grown overnight at 37°C.

For transformation of STBL2 cells (Invitrogen), the procedure is the same except the heat shock is performed in a 15ml falcon tube, at 42°C for 25sec. The cells were incubated in S.O.C. media at 30°C prior to plating. They were grown overnight at 30°C for 16hrs.

2.2.3 Preparation and transformation of competent cells for electroporation

Electrocompetent E.coli XL1-Blue-MRF' cells were prepared according to the protocol given by Sambrook *et al.* (1989).

This involved the innoculation of 5ml of LB media by a single colony of XL1-Blue-MRF' cells. This was grown overnight at 37°C with shaking at 250rpm. 2.5ml of the overnight culture was then used to inoculate 500ml of LB media in a sterile 2l flask. This was allowed to incubate at 37°C, shaking at 250rpm until an OD_{600} of 0.5 – 0.6 was reached. The culture was cooled in an ice bath for 15min prior to transferral to two 250ml centrifuge bottles. The cells were pelleted by centrifugation at 4200xg for 15min at 4°C, in the

Sorvall GSA rotor. Once the supernatant had been removed each pellet was resuspended in 250ml of ice-cold distilled water. The centrifugation and resuspension steps were repeated, with a final centrifugation pelleting the cells ready for use. Cells for immediate use were resuspended in an equal volume of ice-cold water. For storage of electrocompetent cells, the pellet was resuspended in ice cold 10% glycerol and water solution. This suspension was aliquoted into 700µl eppendorf tubes and snap frozen in a dry-ice/ethanol bath. The cells were stored at -80°C.

STBL2 cells for electroporation were prepared in the same way.

Electroporation of XL1-Blue-MRF' cells was performed according to the Gene Pulser Electroprotocol using the Pulse controller (Bio-Rad). 10-100ng of ligation mixture in a volume of 1-2.5µl was added to a 1.5ml eppendorf tube already containing 50µl of electrocompetent cells. Once the contents were mixed, they were transferred to a clean electroporation cuvette (0.2cm). The cuvette was placed in the electroporation chamber and the pulse was applied at 2.5kV, 25µF, 200ohms. Immediately following the pulse, 1ml of S.O.C media was added, the suspension transferred to a 1.5ml eppendorf tube and incubated at 37°C for 1hr, shaking at 225rpm. The cells were aliquoted onto LB agar with ampicillin and incubated at 37°C for approximately 15hrs.

The procedure for STBL2 cell electroporation was the same with the exception that cells were incubated at 30°C prior to and proceeding plating.

2.2.4 Isolation of nucleic acid

(i) Small scale extraction of plasmid DNA

Bacteria containing the plasmid were grown overnight in a 3ml liquid culture with antibiotics that was inoculated with a single colony from an agar plate. The plasmid DNA was recovered from

half this culture by alkaline lysis (Sambrook *et al.*, 1989), the minipreparation method by Berghammer and Auer, 1993, or the Qiagen proprietary method 'concert minipreparation'.

(a) Alkaline lysis

Bacteria were recovered by centrifugation at 12000xg for 30sec. The bacterial pellet was resuspended in 100µl of Solution I, and lysed with the addition of 200µl Solution II. After 5min incubation on ice, 150µl of Solution III was added to precipitate proteins and chromosomal DNA. The preparation was centrifuged for 15min at 10000xg to pellet protein and chromosomal DNA. Plasmid DNA was removed as the supernatant and precipitated in ethanol.

(b) Minipreparation by Berghammer and Auer, 1993.

The bacterial pellet was resuspended in 80µl of lysis buffer and incubated at room temperature for 5 min. The suspension was then boiled for 60sec and immediately snap-cooled on ice. Centrifugation at maximum speed in a microcentrifuge for 15min removes the cell debris and genomic DNA so that the plasmid DNA remains in the supernatant.

(ii) Large scale extraction of plasmid DNA

Bacteria were grown overnight at 37°C as a 100ml culture. The plasmid DNA was extracted either by alkaline lysis or by Qiagen ion exchange column (Biolab).

(a) Alkaline lysis

Crude extraction was performed as per small scale preparation in section 2.2.4(i)(a). The plasmid DNA was then resuspended in 5ml

H_2O and 5ml LiCl was added, with incubation at 4°C overnight. The RNA was pelleted by centrifugation at 10000xg for 10min in the Sorvall SS34. DNA was recovered by precipitation with ethanol and resuspended in 5ml H_2O. Remaining RNA was removed by incubating the DNA with RNase A added to a final concentration of 0.02µg/µl. This was then extracted twice with phenol / chloroform (1:1), once with chloroform, and precipitated in ethanol. Recovered DNA was resuspended in water or TE and stored at -20°C.

(b) Qiagen purification

Purification of plasmid DNA was performed using the Qiagen Maxi Kit protocol. Cells were lysed and DNA was loaded onto a pre-equilibrated Qiagen-tip by gravity flow. Cell debris and RNA was washed through before elution of the DNA. DNA was precipitated in 0.7volume of isopropanol, washed twice in 70% ethanol and resuspended in DEPC treated water. This method was used for DNA preparations required for cell culture transfections.

(iii) RNA extraction by TriZOL

Cultured cells were washed twice in PBS and lysed in the plastic culture dish by the addition of TriZOL (1.5ml per 10cm dish). Genomic DNA was sheared by pipeting the lysate with a 1ml eppendorf pipette several times. The lysate was transferred to a 2ml tube and incubated at room temperature for 5min. Following the addition of 200µl chloroform, the tube was shaken for 15sec and incubated for 3min at room temperature. The lysate was then centrifuged in a microfuge at 12000xg for 15min and the aqueous phase separated to a new tube. This was precipitated with one volume of isopropanol and the RNA pelleted at 14000xg for 10min in a microfuge. The RNA pellet was washed twice with 70% ethanol

and allowed to air day for 10min. It was finally resuspended in 50-100µl of DEPC treated water.
Extracted RNA was analysed by Northern blot hybridisation.

2.2.5 Purification of nucleic acids

(i) Phenol/Chloroform extraction

For all protein removal, nucleic acids were extracted twice with 1 volume of Tris-saturated phenol:chloroform, (1:1). Phenol was removed by an extraction with chloroform. Nucleic acids were then precipitated in 0.3M NaOAc and 2 volumes of ethanol at -20°C before recovery by centrifugation (10000xg for 15min). Ethanol was removed and the pellet washed thoroughly twice in 70% ethanol, dried under vacuum and resuspended in water or TE.

(ii) GFX spin column extraction

Purification by GFX spin column (Amersham Pharmacia Biotech) was performed as per the kit instructions. Briefly, this involved dissolving the reaction mixture or agarose sample in 500µl of capture buffer. For agarose, the sample was heated to 65°C until completely dissolved. This was applied to the column and centrifuged at 10000xg in a microcentrifuge for 30sec. 500µl of wash buffer was then added to the column, and centrifuged again at 10000xg for 30sec. The column was removed to an eppendorf tube, and 50µl of TE or water applied and allowed to incubate for 1min at room temperature. The column was finally centrifuged at 10000xg for 1min to elute the nucleic acid.

(iii) Lambda DNA extraction

Genomic DNA cloned into the Lambda Fix II library (Stratagene) was extracted using Lambda SORB (Promega) or WIZARD lambda preps DNA purification system (Promega).

(a) λSORB

DNA was extracted from phage lysates stored at 4°C. In a centrifuge, 10ml of lysate and 100µl of λSORB were mixed and rotated at room temperature for 30mins. The suspension was then centrifuged at 14000xg in the Sorvall SS34 rotor for 15min. Supernatant was removed and the pellet resuspended in 1ml of SM buffer. This was transferred to a 1.5ml eppendorf tube and centrifuged in a microfuge at 15000xg for 2min. This SM buffer was removed and centrifugation was then repeated. The pellet was resuspended in 250µl of 10mM Tris-Cl pH7.8 and 250µl of 10mM EDTA. The suspension was heated to 67°C for 5min, centrifuged again as above and the supernatant collected. The supernatant was extracted twice with phenol/chloroform/isoamyl alcohol (24:24:1), once with chloroform/isoamyl alcohol (24:1) and precipitated in ethanol and ammonium acetate.

(b) WIZARD λ preps DNA extraction system

40µl of nuclease mixture was added per 10ml of lysate and incubated at 37°C for 15min. 4ml of phage precipitant was then added and allowed to incubate on ice for 30min. This was then centrifuged at 10000xg in a microfuge for 10min and the supernatant removed. The pellet was resuspended in 500µl of phage buffer and the centrifugation repeated. The supernatant was collected, mixed with 1ml of purification resin and pipeted into the

WIZARD minicolumn syringe barrel. The resin and lysate mixture was gently forced through the column using a syringe plunger. Following this, the column was washed by pushing 2ml of 80% isopropanol through. The minicolumn was then transferred to an eppendorf tube and centrifuged for 10000xg for 2min to dry the resin. 100µl of water preheated to 80°C was applied to the column and immediately centrifuged at 10000xg for 20sec to elute the DNA. Purified lambda DNA was stored at -20°C.

2.2.6 Electrophoresis

(i) Agarose gel electrophoresis of DNA

Electrophoresis was performed according to Sambrook *et al.* (1989). Gels were made containing 1-2.5% agarose (SeaKem) dependent on the size and separation of DNA required. They were made in either 1X TAE or 1X TBE, with 300ng/ml ethidium bromide (EtBr). Electrophoresis was carried out in Horizon or Owl electrophoresis tanks, in a buffer of either 1X TAE or 1X TBE. Prior to loading, DNA samples and markers were mixed with glycerol loading dye. A voltage of between 50-100V was applied and separation achieved in 1-3hrs. Fractionated DNA was visualised on a UV transilluminator screen or UV gel doc (Bio-Rad) and photographed onto Polaroid or thermal paper respectively.

(ii) Agarose gel electrophoresis of RNA

Electrophoresis was carried out as described by Sambrook *et al.* (1989).

Gels contained 1.6% agarose, 1X MOPS and 0.66M formaldehyde. For Northern blot analysis, total RNA was quantified by spectrophometry at OD_{260} and 10µg loaded after incubation at 65°C for 5min with one volume of formaldehyde loading dye.

Electrophoresis was performed at 100V for 3-8hrs at 4°C. Dyes containing ethidium bromide allowed visualisation by UV.

(iii) Denaturing polyacrylamide gel electrophoresis

Gels contained 4-6% w/v acrylamide (19:1) or (29:1) w/w acrylamide:N,N'methylene-bis-acrylamide, 7M urea and 1X TBE buffer. Polymerisation was initiated with the addition of APS to 0.1% and TEMED to 0.04%. The nucleic acid to be analysed was mixed with one volume of formamide loading dye and heat denatured at 65°C for 5min.

Polyacrylamide sequencing gels were run on Gibco BRL sequencing apparatus Model SA. The gel dimensions were 20cm x 90cm x 0.15cm. Prior to loading samples, the gel was pre-run for one hr at 65W. Samples were loaded through a shark-tooth comb and electrophoresed for 3-5hrs at 65W. Gels were transferred to 3MM Whatman paper, vacuum dried and visualised by autoradiography.

Polyacrylamide gels of 4-6% were used for sequencing and identifying primer extension products. Both mini gel apparatus (10 x 8 x .15cm) and large apparatus (20 x 20 x .15cm) were used. Samples were loaded into preformed wells after heat denaturation with loading dye. Electrophoresis was performed at 100V for 1-8hrs. Gels were transferred to 3MM and dried before exposure to autoradiograph film.

2.2.7 Radiolabelled nucleic acids

(i) Radiolabelled cDNA probes

DNA to be labelled was excised from agarose after electrophoresis and extracted by GFX spin column into a final volume of 50µl of distilled water. 25ng of DNA was denatured by heating to 100°C for 5 min in a water bath before addition to the Rediprime random

priming mix (Amersham). 2.5µl of [α-^{32}P]-dCTP was added and mixed. The reaction was initiated on transferral to a 37°C incubator and allowed to proceed for at least 20min. Labelling was terminated by the addition of 5µl of 0.2M EDTA. Radioactive incorporation was measured for each probe by washing a sample in trichloroacetic acid to remove unincorporated nucleotides. Prior to hybridisation, the probe was denatured with 0.2 volumes of 4M NaOH for 5min.

(ii) T4 polynucleotide kinase labelling

Oligonucleotides to be end-labelled were diluted to a final concentration of 1pmol/µl. 5µl of the oligonucleotide was added to a 1.5ml eppendorf tube containing 9.5µl of distilled water. 5µl of 5X exchange buffer was added along with 5µl of [γ-32P]dATP (10µCi/µl, 3000 Ci/mmol). The reaction was initiated with the addition of 5U of T4 polynucleotide kinase and incubated at 37°C for 10min. The reaction was terminated with the addition of EDTA to 5mM. Incorporation was determined by TCA precipitation and scintillation counting. Unincorporated nucleotides were removed using a spin column (ProbeQuant column, Amersham).

(iii) Terminal transferase labelling of oligonucleotides

15pmol of oligonucleotide was end-labelled by terminal transferase (Tdt). The reaction took place in an eppendorf tube containing distilled water, 4µl of 5X Tdt buffer, 3µl of 5mM $CoCl_2$, 4µl of [α^{32}P]dCTP, the oligonucleotide and 1µl of Tdt. The labelling rection was allowed to proceed for a minimum of 20min at 37°C.

(iv) Removal of unincorporated nucleotides

Purification of DNA labelling reactions was performed using ProbeQuant G-50 micro columns (Amersham). The sample to be purified was applied to a column of resuspended resin without disturbing the resin bed. This was centrifuged at 735xg for 2min and the purified sample collected in the bottom of the support tube.

2.2.8 Enzyme Reactions

(i) Restriction enzyme digestions

Plasmid DNA was digested with the appropriate restriction enzyme at an excess concentration of 2-3U per μg. Digests included 1X buffer as specified by Roche Diagnostics. Digests were performed in a volume of distilled water such that the enzyme concentration was less than 10%. All digests were carried out in a water bath at 37°C for at least 2hrs. Reactions were terminated by agarose gel electrophoresis or GFX purification.

(ii) Dephosphorylation

Plasmid DNA was digested with the appropriate restriction enzyme as above, and purified by GFX spin column. The eluted DNA was mixed with 1X dephosphorylation buffer and 1U of calf alkaline phosphatase (Roche). The reaction was incubated at 30°C for 30min followed by incubation at 50°C for 30min. Dephosphorylation was stopped by incubation on ice and purification by GFX column.

(iii) Klenow Fragment end filling

The Klenow fragment of DNA polymerase I was used for the blunt ending of restriction fragments used in blunt ended cloning, as described in section 2.2.9(i). DNA was digested with the appropriate restriction enzyme and the fragment required was purified from agarose gel, or straight from the digest by GFX spin column. The eluate was approximately 50µl, to which 6µl of 10X L buffer (Roche) was added. 0.6µl of 10mM dNTPs were added and 2µl of klenow enzyme. The reaction was allowed to proceed at 37°C for 30mins. Reaction products were then purified again by GFX spin column.

(iv) Taq polymerase tailing reaction

The PCR product to be tailed was purified by GFX spin column. Up to 7µl of PCR product (100-500ng) was added to an eppendorf containing 0.2mM dATP and 1µl of 10X Taq reaction buffer. 5U of Taq DNA polymerase (Roche) was added and the total volume of the reaction made up to 10µl with distilled water. The reaction proceeded at 70°C for 15-30min. 1-2µl of this reaction was used in a ligation reaction.

2.2.9 Ligation reactions

(i) Cloning blunt ended fragments

Both vector DNA and the fragment of DNA to be cloned, were digested with the appropriate restriction enzyme. If necessary, these were separated by agarose gel electrophoresis and the fragment excised. Excised DNA was purified by GFX spin column and blunt ended using the klenow fragment of DNA polymerase I

as described already. Blunt end vector DNA was then dephosphorylated using calf alkaline phosphotase as described, to reduce the chance of self-ligation. Blunt end and dephosphorylated products were purified again and the concentration calculated by UV spectrometry at OD_{280}/OD_{260}. Control ligations were performed with vector only, containing vector DNA, 5X buffer and water. Also a reaction including vector DNA, buffer and 1U of T4 DNA ligase was used to test the possibility of self-ligation occurring. Ligation reactions were performed in the presence of 5X buffer, 1U of T4 DNA ligase with a 3 or 6 fold molar excess of insert over vector. All reactions were incubated at 16°C overnight.

(ii) Cloning using cohesive termini

Vector DNA was digested with restriction enzyme(s) the same as or compatible to those used in the digestion of DNA to be cloned. If this was a single enzyme, then the vector DNA was dephosphorylated following restriction enzyme digestion. Ligation reactions were performed in 5X buffer with 1U of T4 DNA ligase (Roche) and a 3 or 6 fold molar excess of insert DNA over vector DNA. The ligation reactions were incubated at 16°C for a minimum of 4hrs.

(iii) Cloning using a T-overhang vector

For PCR products that did not have an existing or manufactured restriction enzyme site at the required position, a T-tail was attached for cloning. This method of cloning used the pGEM-Teasy vector (Promega). PCR performed with Taq (Invitrogen) or Elongase (including Taq and Pyrococcus species GB-D thermostable DNA polymerises, Roche) generates an addititional A nucleotide on the ends of the product. Where a proof reading enzyme was used the A-tail was added in a separate reaction. The ligation reaction

incorporated 50-300ng of PCR product, rapid ligation buffer, 50ng of pGEM-Teasy vector and 3U of T4 DNA ligase. The total reaction was performed in a 10µl volume made up with distilled water. Incubation of the reaction was either at room temperature for 1hr or overnight at 4°C for low efficiency reactions.

2.2.10 DNA Sequencing

Manual Sequencing was performed according to the Sanger-dideoxy mediated chain termination method (as described by Sanger *et al.* (1977)).

DNA was sequenced from the T7 primer using Sequenase (USB Chemicals). Denaturation of 5µg of double-standed DNA was performed in 10µl of 0.2M NaOH for 10min. The DNA was then precipitated in ethanol and 0.3M NaOAc, washed in 70% ethanol and resuspended in 7µl of distilled water for annealing. 0.5pmol of T7 primer was added to the DNA along with Sequenase reaction buffer (20 mM Tris-Cl pH 7.5, 50mM $MgCl_2$, 250mM NaCl). This mix was heated to 65°C in a heating block and allowed to cool to 30°C. Reaction mixes containing 3µM dATP, dGTP, dTTP, 1µl of 0.1M DTT, 0.5µl [α-^{32}P]-dCTP and 2U of Sequenase were added to the annealing mix and chain extension was allowed to proceed for 5min at room temperature. Four termination mixes (15µM ddNTP, 150µM dNTP) were prepared in eppendorf tubes and pre-warmed to 37°C in a heating block. 3.5µl of the reaction mixture was added to each termination reaction and incubated for a further 5min at 37°C. The reaction was stopped by the addition of an equal volume of STOP dye and samples were heat denatured at 65°C for 5min. Electrophoresis was on 4% polyacrylamide/7M urea denaturing gels as described in section 2.2.6(iii).

Automated DNA Sequencing was performed at the Waikato Sequencing Facility, University of Waikato.

2.2.11 Southern blot hybridisation

DNA electrophoresis was as described in section 2.2.6(i), except no EtBr was added to the gel. After electrophoresis, gels were stained with a solution of water and EtBr (300ng/ml) and photographed under UV illumination. The gel was then washed with water for 15min. Prior to transfer the gel was washed in two successive amounts of denaturing solution. The DNA was transferred to Hybond-N^+ membrane (Amersham) by capillary transfer in denaturing solution, as described by Sambrook *et al.* (1989). Transfer was allowed to proceed for 16 hrs. The nylon membrane was rinsed in 2X SSC prior to crosslinking in a UV Stratalinker. Membranes were prehybridised in Hybaid bottles for at least 1hr at 60°C in Church and Gilbert hybridisation buffer (7% SDS, 0.001M EDTA, 0.25M $Na_2HPO_4.2H_2O$). Radiolabelled cDNA or oligonucleotide probe was added after denaturation with a fresh aliquot of hybridisation buffer. Hybridisation was performed in a Hybaid oven at 60°C for 16hrs.

When hybridisation was complete, membranes were washed in 2X SSC / 0.5% SDS for 1 min at 55-65°C with gentle rocking. This was followed by a wash in 1X SSC / 0.5% SDS for 15min. If required, a third wash in 0.1X SSC / 0.1%SDS was performed.

Alternately, hybridisation was performed using the high efficiency hybridisation system (Progenz). Prehybridisation was performed using the 10X washing and prehybridisation solution (HS 114) diluted to 1X. Prehybridisation was carried out in either a rotating hybaid bottle or on the bench in a container at room temperature. The prehybridisation solution was removed after 20min and replaced with hybridisation buffer (HS 114F). Double stranded probe was denatured before adding to the hybridisation buffer and membrane. Hybridisation was performed in a hybaid oven at 55-65°C for 16hrs. Washing of the membrane was performed at room

temperature using the wash and prehybridisation solution. The washing cycle was repeated five times, consisting of gentle shaking of the membrane in 1X wash and prehybridisation solution for 7-10min at room temperature.

Membranes were sealed in a plastic bag and visualised by autoradiography.

Prior to a membrane being re-probed it was stripped in boiling 0.1% SDS until cooled. This was repeated and then the membrane was washed in 2X SSC. Autoradiography confirmed the stripping process.

2.2.12 Northern blot hybridisation

RNA electrophoresis was as described in section 2.2.6 (ii). After electrophoresis, gels were washed for 10-20 min in 10X SSC to remove formaldehyde and the RNA transferred to Hybond-N membrane (Amersham) by capillary transfer in 10X SSC, as described by Sambrook *et al.* (1989). Transfer was allowed to proceed for 16hrs. The nylon membrane was rinsed in 2X SSC prior to crosslinking in the UV Stratalinker. Transfer to the nylon membrane could be visualised under UV due to the addition of EtBr to the RNA samples.

Membranes were prehybridised in Hybaid bottles for at least 1hr at 60°C in Church and Gilbert hybridisation buffer (7% SDS, 0.001M EDTA, 0.25M $Na_2HPO_4.2H_2O$). Radiolabelled cDNA probe was added after denaturation with a fresh aliquot of hybridisation buffer. Hybridisation was performed in a Hybaid oven at 60°C for 16hrs.

Washing of the membranes was performed in the same manner as for Southern blot hybridisation. Membranes were sealed in a plastic bag and visualised by autoradiography.

2.2.13 5' RACE

Initial analysis by 5' RACE was performed on early lactating (1 day or 4 day) mouse mammary gland RNA and also a sample of polyadenylated 4 day lactating RNA. Oligonucleotides included gene specific primer 1 (GSP1) for first strand cDNA synthesis and a nested gene specific primer 2 (GSP2) annealing 5' to GSP1, which was employed in the final PCR amplification reaction. For oligonucleotide sequences, refer to section 2.1.5. The position of GSP1 was approximately 350nt 3' of the putative transcription start site so that a reasonably short product would be generated but still be long enough to extract by column. 2.5pmol of GSP1 and 1µg of RNA were denatured at 70°C for 10min. First strand cDNA synthesis was completed with Superscript II RNase H-RT (200U) (Invitrogen) at 42°C for 30min in the presence of 10x Platinum PCR buffer (Invitrogen), ($MgCl_2$ (2.5mM), dNTPS (0.4mM), DTT (10mM)). The reaction was terminated by incubation at 70°C for 15min and RNA was removed by the addition of RNase T1 (3000U) (Invitrogen) and RNase H (3U)(Roche) for 30min at 37°C. Purification of cDNA was accomplished using a GFX spin coloumn (Amersham). Eluted DNA was then 'C' tailed with Tdt (25U) (Roche) in 10mM Tris HCl pH 8.4, 25mM KCl, 15mM $CoCl_2$, 200µM dCTP. Before addition of enzyme the reaction was denatured at 94°C for 3min and snap cooled on ice. Following addition of enzyme the reaction was allowed to proceed at 37°C for 10min and stopped by incubation at 65°C for 10min. The tailed cDNA was then finally amplified in a PCR reaction using oligonucleotides GSP2 at 400nM and abridged anchor primer (AAP) at 400nM. PCR was performed in a Perkin Elmer thermal cycler in the presence of 1.5mM $MgCl_2$, 200µM dNTPs and 1U of Taq DNA polymerase (Roche). Initial denaturation was at 94°C for 2min followed by 35 cycles of denaturation (94°C, 30sec), annealing (50°C, 30sec) and extension (72°C, 60sec). A final elongation step was performed at 72°C for 5mins. PCR products

were analysed on a 1% agarose gel, visualised by EtBr and UV illumination.

2.2.14 Primer extension

Primer extension reactions were performed with 1-100µg of RNA from 16 day lactating or 18 day pregnant mice. 10pmol of oligonucleotide (PE35) homologous to position +46 to +81, was labelled by T4 polynucleotide kinase (Invitrogen) as described already. The labelled oligonucleotide was run on an 8% polyacylamide gel and the band excised and eluted in 10mM Tris-Cl pH8.0. 100µg of RNA was annealed with 7.5 x 10^4cpm PE35 in hybridisation buffer (1.5M KCl, 0.1M Tris-Cl, 10mM EDTA) at 45°C for 2hrs after a 5min denaturation at 75°C. All reaction components (30mM Tris-Cl pH 8.3, 15mM $MgCl_2$, 8.2mM DTT, 0.22mg Actinomycin D, 0.22mM dNTPs) were preincubated for 1min at 42°C before the addition of 25U of Superscript II RNase H (Invitrogen). Extension was carried out at 42°C for 1hr at which time 2µg of RNase A was added to degrade the template. The extension product was extracted in phenol/chloroform twice, chloroform once and precipitated in ethanol and salt. The sample was loaded onto a 4% polyacrylamide gel after the addition of 2X STOP dye and heat denaturation at 65°C for 5min. The products were visualised after 1 day by autoradiography and the size of the products was determined by comparison with sequence reactions from a subclone (5' SacI 4.9) using the same primer (PE35), run alongside the extension reaction.

2.2.15 PCR amplification

(i) PCR using Taq DNA polymerase for short products (<5kb)

A master mix was prepared in a sterile eppendorf tube on ice, using filter pipette tips and freshly autoclaved distilled water. A standard reaction consisted of 1X PCR amplification buffer plus $MgCl_2$, 20mM of each dNTP, 100pmol of each forward and reverse primer, template DNA (0.1-2µg) and water to a final volume of 100µl. The reaction was transferred to individual thin-walled PCR tubes and inserted into a Perkin-Elmer thermocycler for amplification. The first denaturation step was performed at 94°C for 5min. Whilst hot, 0.5µl of Taq DNA polymerase (5U/µl) was added. Typically, the amplification cycle consisted of 94°C, 30sec denaturation; 50°C, 30sec annealing; 72°C, 1min extension for 30-35 cycles, with a final extension of 5min at 72°C. Products were visualised on an agarose gel under UV light.

(ii) 'Elongase' PCR for long products (Invitrogen)

PCR reactions were prepared on ice using distilled water, filter pipette tips and fresh autoclaved eppendorf tubes. A master mix was prepared containing 200µM of each dNTP, 200nM of each forward and reverse primers, approximately 100ng of genomic DNA template or 25pg of other DNA template. This was made to a final volume of 20µl per reaction, with water and aliquoted into individual PCR tubes. A second master mix was prepared, containing a total of 10µl of A and B buffer depending on required $MgCl_2$ concentration, 1µl of elongase enzyme mix and water to 30µl. For example, 1.5mM $MgCl_2$ requires 5µl of each A and B buffers. This was aliquoted into the tubes containing master mix one, mixed and centrifuged briefly to collect contents at the bottom of the tube. Amplification was performed in a Perkin-Elmer

thermocycler, beginning with denaturation at 94°C for 30sec, then 35 cycles of denaturation (94°C, 30sec), annealing (50-60°C, 30sec) and extension (68°C, 60sec per kb of target).

(iii) 'Expand High Fidelity' PCR for long products (Roche)

The Expand system utilises Taq DNA polymerase and Pwo DNA polymerase. PCR set up was performed on ice, using sterile equipment. The first master mix consisted of 200µM of each dNTP, 300nM of each primer, 0.1-0.75µg of template DNA and water to make a final volume of 25µl. The second master mix was made up of 10X Expand buffer (including $MgCl_2$) and 0.75µl of enzyme mix. Water was added to master mix two to a final volume of 25µl . 25µl of each mix was pipeted together into PCR tubes. The PCR was performed in a Perkin-Elmer thermocycler and followed the cycle given above in (ii). A final prolonged extension incubation was given following the 35 cycles of 7min at 72°C.

2.2.16 Cell culture

(i) Storage and retrieval of cells

For storage of cells, cells were harvested as described below, but resuspended in freezing mix following centrifugation. Freezing mix contains 10% DMSO, 25% FCS and 65% 1X media. The resuspended cells were aliquoted in 1ml amounts in 1.5ml cryovials and frozen at -80°C in an isopropanol container. They were transferred after 24 hrs to liquid nitrogen for long-term storage.

Cells were retrieved from liquid nitrogen by incubating the cryovial at 37°C until defrosted. The volume was then doubled with the addition of FCS, mixed and incubated at 37°C for 10min. The cells were transferred to a 10cm dish and the volume increased to 10ml

using media supplemented with growth factors and allowed to incubate overnight at 37°C. They were then lifted from the plate and split to continue growing.

(ii) Dividing and counting cells

Once cells reach 80-100% confluence, they were divided to allow them to continue growing. Media was removed from cells and they were washed twice with 10ml of PBS. HC11 cells were lifted with the addition of 2ml trypsin/EDTA and incubated at 37°C for 5min. Comma1D cells were lifted using 0.5ml of dispase diluted in 1.5ml of Solution A and incubated at 37°C for 5min. The Comma1D cells were scraped off the plate using a cell scraper. Cells were transferred to a 10ml falcon tube and centrifuged at 500xg for 5min. The supernatant was removed and cells resuspended in 2-5ml of fresh media. An aliquot of cells (30µl) was diluted in 30µl of Trypan blue solution and pipetted onto a counting chamber. Cells were counted under a microscope and calculated per ml of media resuspended in.

(iii) Transfection

Cells to be transfected were plated to a density of $0.75x10^5$ (cells/well) for HC11 in DMEM-F12 supplemented with growth factors, and $5x10^5$ for Comma1D cells also in supplemented DMEM-F12. The numbers given are for a single well of a 6 well plate and the cells were left to re-attach overnight.

(a) Lipofectamine 2000 (Invitrogen)

Transfection was performed with 1µg of construct DNA and 0.2µg of pCH110 control plasmid. The DNA was mixed and added to Lipofectamine 2000 (LF2000) (3µl) previously diluted in 50µl of

DMEM-F12 media. This was allowed to complex at room temperature for 20min. The DNA-LF2000 complex was then added to each well and gently mixed. Cells were left for 48hrs at 37°C and 5% CO_2 without changing the media.

(b) Fugene6 (Roche)

Optimisation of DNA amounts for transfection by Fugene6 reagent resulted in the use of 1µg DNA construct with 0.2µg PCH110 plasmid. These were mixed and added to a diluted volume of Fugene6 (3µl) in 97µl of serum free DMEM-F12 and allowed to complex at room temperature for 15min. The mixture was added drop-wise to the well of cells, mixed and incubated at 37°C with 5% CO_2 for 48hrs.

(c) Addition of Hormones

Cells to be tested with the addition of several hormones or growth factors were transfected in the same way as given above for Fugene6. Twelve hrs after transfection, the hormone or factor was added and the cells were allowed to continue growing for a further 48 hrs. The hormones added were Progesterone (10^{-7}M), and β-Estradiol (10^{-7}M).

(d) Serum dilution series

Cells were transfected as already described with 1µg molar ratio of each construct, and pCH110. Transfected cells were incubated overnight in media containing 8% FCS. The media was removed and replaced with media containing either 0.1,1,5,10 or 20% FCS. The cells were then allowed to replicate for 48hrs before harvest and assay. Each construct was transfected in duplicate for each serum dilution.

(e) Stable transfection

HC11 cells were grown to confluence in a 10cm plate, lifted and counted. Cells were plated for transfection at $0.75x10^5$ for HC11. Two wells of a 6-well plate were prepared for each construct to be transfected and the cells left overnight to re-attach. The cells were given fresh media and transfected by the LF2000 procedure as described, with the exception that 1µg of linearised construct DNA was transfected with 100ng of a puromycin cassette plasmid instead of pCH110. Cells were grown overnight at 37°C and the following day they were lifted from the 6 well plate and each well re-plated on a 10cm plate in fresh media.

Prior to the cells being treated with puromycin, a test experiment was performed using a dilution series of puromycin to determine the minimum amount required to kill all cells. Cells were plated into six wells and given a varying amount of puromycin (0.2, 0.5,1,2,3,5 µg/ml). 2µg/ml was the amount required for all cells to die after 48hrs incubation.

Cells were allowed to grow for 48hrs at 37°C before the first addition of puromycin for selection of positive cells. Puromycin was added to a final concentration of 2µg/ml into normal media. Cells were incubated at 37°C with sufficient changes of media and puromycin until they reached 80-90% confluence. Cells were then trypsinized (HC11) from plates and several aliquots were frozen, while others were allowed to grow on for harvest and assay.

2.2.17 Cell culture assays

Cell culture assays were performed on the extract obtained from transfected cells after 48 hrs growth. Media was removed from cells and the cells were washed 3 times in ice-cold PBS, which was removed between washes. After the last wash was removed cells were lysed in 1ml of 1X lysis buffer (either ELISA kit lysis buffer, or thin layer chromatography (TLC) CAT reporter lysis buffer). The cells were allowed to incubate at room temperature in the buffer for 15min (TLC) or 30min (ELISA). For cells harvested in the TLC buffer, they were scraped from the plate and the cells and lysis buffer pipetted into a 1.5ml eppendorf tube. Cells harvested in ELISA buffer were not scraped but the extract pipetted from the plate into an eppendorf tube. All extracts were immediately put on ice for 10min and then vortexed and centrifuged for 3min at 10000xg. Extract was kept on ice for assays to be run immediately, or stored at 80°C.

(i) BCA protein assay kit (Pierce)

In a microwell plate (96 wells), 25µl of each standard Bovine Serum Albumin (BSA) was pipetted into the first row. The standard BSA was diluted from 2000µg/ml stock to 0,20,50,100,200,500,1000 and 2000µg/ml. 25µl of each unknown sample was added to the next row(s) of wells. The working reagent was freshly prepared by mixing 50 parts of BCA reagent A with 1 part of BCA reagent B. 200µl of this was pipetted into each well and mixed. The covered plate was incubated at 37°C for 30min and absorbance read at OD_{595}nm on an ELISA plate reader.

(ii) β-Galactosidase ONPG assay

In a microwell plate, 150µl of solution Z, 30µl of ONPG (4mg/ml) and 40µl of unknown sample were mixed. For standard β-galactosidase (β-Gal), a dilution of stock β-Gal (1000U/ml) was prepared between 0-2U/ml with 0.25M Tris-Cl pH7.8. The plate was covered and mixed, and incubated at 37°C for 30min. The absorbance was read at 405nm on an ELISA plate reader.

(iii) ELISA CAT assay (Roche)

CAT enzyme standards were prepared by diluting CAT enzyme stock solution in sample buffer at concentrations of 0-1ng/ml. 200µl of each standard was pipetted into a well on the microtitre plate supplied with the kit. 200µl of each sample was also pipetted into wells on the plate, covered and incubated at 37°C for 2 hrs. The solution was removed and the wells washed 5 times with 250µl of washing buffer for 30sec each. Anti-CAT-DIG working dilution (0.2µg/ml) was prepared and 200µl added to each well. The plate was covered and incubated at 37°C for 1 hr. This solution was removed and the plate washed 5 times in washing buffer. 200µl of anti-DIG-POD working dilution (150mU/ml) was added to each well and incubated again at 37°C for 1 hr. Once the solution was removed, the wells were washed 5 times in washing buffer. POD substrate was mixed with substrate enhancer (1mg/ml) and mixed at room temperature for 30mins before adding 200µl of this to each well. The plate was incubated at room temperature for 10-40min until colour had developed sufficiently to be read at OD_{405}nm.

(iv) CAT enzyme assay by thin layer chromatography (Promega)

Cell lysate to be tested was heated for 10min at 65°C. 100µl of this was pipetted into a clean 1.5ml eppendorf tube and 5µl of n-

butyryl CoA (5mg/ml) and 17µl water added. Lastly, 3µl of ^{14}C-chloramphenicol (0.05Ci/mmol) was added, the contents mixed and incubated at 37°C for 5 hrs. The reaction was terminated with the addition of 500µl of ethyl acetate. The samples were vortexed for 1min and centrifuged for 3min at 10000xg. The upper organic phase was transferred to a new tube and the samples speedvac dried until no liquid remained. The sample was resuspended in 20µl of ethyl acetate and 10µl of each was spotted onto a silica plate. The chromatograph was run in a pre-equilibrated tank containing chloroform/methanol (97:3) until the solvent was three quarters up the plate. It was then dried at room temperature, wrapped in plastic wrap and exposed to autoradiographic film for 1-5 days.

The auotradiograph of the TLC CAT result was aligned with the original TLC plate and the bands were excised and scintillation counted to quantify the amount of CAT. Duplicates of each construct were averaged and then standardised against the negative construct (0).

(v) DNase I Hypersensitive assay

Cells to be harvested for DNase assay were grown to confluence in a 15cm plate in the appropriate media. HC11 cells were lifted from the plate by treatment with trypsin and Comma1D by dispase treatment. Cells were pipetted immediately into 50ml falcon tubes and placed on ice. They were centrifuged at 500xg for 10min at 4°C and washed twice in cold PBS. The pellet was resuspended in 5 volumes of RSB buffer including NP-40. This was incubated on ice for 10min and centrifuged at 700xg for 10min. The pellet was washed twice in RSB buffer (without NP-40) to obtain an OD260 of 0.4 for a 1/100 dilution. 100µl aliquots of this suspension were then incubated with DNaseI (0-50µg/ml, Roche) for 10mins at 37°C. The reactions were stopped with the addition of DNase STOP

buffer. These were left to incubate at 37°C overnight on a rotator. The DNA was then extracted twice with phenol/chloroform, once with chloroform and then treated with RNase A (0.1mg/ml) and RNase T1 (1500U/ml) for 2hrs at 37°C. The extraction was repeated and DNA precipitated in ethanol and salt. The DNaseI products were digested with a 10 fold excess of restriction enzyme and analysed by Southern blotting.

Chapter 3. Identification and analysis of the murine *Ese*-2 5' region

3.1 Introduction

Promoter regions of the *ets* genes have not been widely studied. Most analysis concentrates around the open reading frame although in some cases the presence of several isoforms of a gene requires a brief study of alternative 5' exons, multiple promoters and start sites. The few *ets* gene promoter regions that have been studied are discussed at the end of this chapter. Some sequence analyses of human *ESE*-2 has previously been done (Oettgen *et al.*, 1999b), and these identified two isoforms for the gene detected in human kidney and prostate. Rapid Amplification of cDNA Ends (RACE), performed using the 'Ready-for-RACE clontech cDNA' kit from Invitrogen, showed that *ESE*-2a differs from *ESE*-2b at the 5' end as a result of alternative 5' exons (Oettgen *et al.*, 1999b). Such alternate exons are presumably due to two different transcription start sites derived from a proximal and distal promoter. This is in contrast to the mouse sequence, which has two isoforms as a result of alternative polyadenylation (Langley, 1999).

Promoter analysis usually involves the cloning and sequencing of 5' flanking regions so that specific regions involved in gene regulation can be characterised. The regulation of *Ese*-2 was of interest and in order to locate regions of importance for *Ese*-2 transcriptional regulation, the promoter has been characterised.

This chapter deals with the initial characterisation of the proximal promoter region of murine *Ese*-2. Prior to the investigation reported in this thesis, genomic and cDNA murine libraries had been screened to isolate clones containing the full-length coding sequence and upstream sequence. Individual clones were inserted into lambda bacteriophage (Stratagene) by L'Huillier *et al.*, (unpublished). The two cDNA transcripts had been sequenced from

this data and some initial restriction mapping had been performed on several of the genomic clones isolated (L'Huillier *et al.*, unpublished). However, the orientation had not been determined. These clones were further analysed in the current investigation and restriction maps were generated for those clones that contained some 5' sequence according to Southern blot analysis with a 5' probe. Fragments were isolated from the 5' regions and were subcloned and mapped extensively to allow smaller clones to be sequenced.

The availability of the draft mouse genomic sequence shortly before this publication allowed a greater level of information to be mined. Sophisticated promoter prediction software has been utilised and results are reported here, together with experimental data. This experimental information came from sequence analyses by primer extension and 5' RACE methods, to identify transcription start sites. Potential regulatory sites were later identified by computer analysis of consensus transcription factor searches, and is discussed later in the chapter.

3.2 Results

3.2.1 Reference material for Chapter 3

Several genomic clones had been previously isolated from a mouse 129 genomic library. One of these was discovered to contain upstream *Ese*-2 gene sequence (λ3.2), and was used for further cloning. A subclone from this (5' Sac-I 4.9) was used for promoter analysis. A number of oligonucleotides were produced from sequence generated from this clone, and were used throughout this chapter. Sequence information is given in Chapter 2, Materials and Methods. Figure 3.1 is a representation of the genomic clone λ3.2, and includes the subclone 5' Sac-I 4.9kb as well as oligonucleotide positions and transcription start sites (as determined later in the chapter).

3.2.2 Analysis of the Lambda genomic clones

Clones from the 129 mouse genomic library that were known to contain some 5' sequence based on hybridisation to a 5' cDNA probe (*Pst*-I 400bp, predominantly hybridising to exon 2) included λ2.1, λ2.4, λ3.1, λ3.2. Some overlap was found between these clones, reducing the number that would need full mapping to identify a suitable 5' region. Clones λ3.1 and λ3.2 have significant overlap, and as a result only one was mapped (λ3.1). The alignment of the lambda clones is depicted in Figure 3.2.

In order to identify the region of the immediate 5'UTR, restriction maps were generated to allow the alignment of probed regions. The initial mapping was done using a series of restriction enzymes with single and double digests. Preliminary digestion was performed with the enzymes *Bam*HI, *Xba*I, *Hin*dIII, *Sac*I, *Bgl*II and *Xho*I, as seen in Figure 3.3(A). Combinations of two of these enzymes gave information for the ordering of fragments. Where one enzyme gave the same size fragment as in a double digest, it was assumed that the fragment was un-digested by the second enzyme. An initial alignment of the fragments was generated by probing with a single oligonucleotide (1.11(2) for example, as seen in Figure 3.3(B)), so that bands containing a common sequence could be identified in each digest. Where these enzymes did not give enough information, further digests were performed using enzymes such as *Asp*718, *Cla*I, *Stu*I, *Eco*RI, *Pst*I, *Hin*dII, and *Dra*I.

The overall orientation of each clone was determined by performing a Southern analysis using oligonucleotide probes. Each gel of digested DNA was transferred by the Southern technique onto Hybond N^+ nylon membrane (Amersham, as described in Chapter 2) and later hybridised in Church and Gilbert solution in the presence of a labelled oligonucleotide [γ^{32}P]-dATP (Amersham Life Sciences). Oligonucleotides used for orientating the clones were based on the cDNA sequence, including 5'seq, 1.11(ATG), 1.11(1), 1.11(2) in 5' to 3' order respectively so that use of a 5'

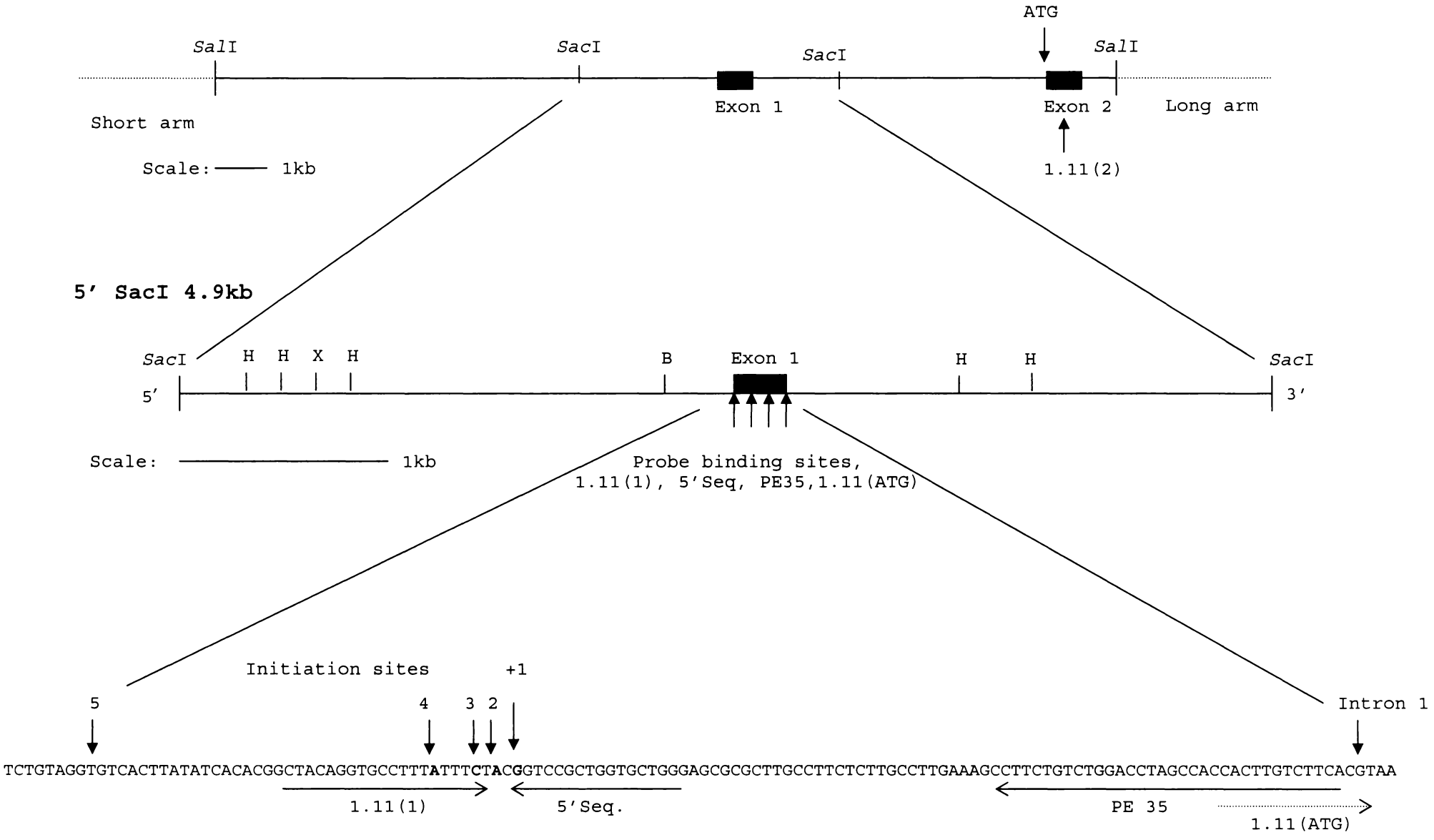

ATG
SalI
SacI
SacI
SalI
Exon 1
Exon 2
Long arm
Short arm
Scale: —— 1kb
1.11(2)
5′ SacI 4.9kb
SacI
H H X H
B
Exon 1
H H
SacI
5′
3′
Scale: —————— 1kb
Probe binding sites,
1.11(1), 5′Seq, PE35,1.11(ATG)
Initiation sites
+1
5
4
3 2
Intron 1
TCTGTAGGTGTCACTTATATCACACGGCTACAGGTGCCTTTATTTCTACGGTCCGCTGGTGCTGGGAGCGCGCTTGCCTTCTCTTGCCTTGAAAGCCTTCTGTCTGGACCTAGCCACCACTTGTCTTCACGTAA
1.11(1)
5′Seq.
PE 35
1.11(ATG)

Figure 3.1 Genomic clone λ3.2 and subclone 5' SacI 4.9kb. The uppermost line represents the 16kb λ3.2 genomic clone containing exon 1 and 2, intron 1 and 5' upstream sequence. This clone contains the sequence for the 5' oligonucleotides shown for 5' SacI 4.9kb, as well as the 1.11(2) oligonucleotide (in exon 2). The two *Sac*I sites are marked. This fragment was subcloned and is depicted by the second horizontal line. Restriction sites B (*Bam*HI) X (*Xba*I) and H (*Hin*dIII) are shown above the line, which contain the oligonucleotide sequences for 5'Seq., 1.11(ATG), 1.11(1) and PE 35. The exon 1 sequence is shown at the bottom of the figure. The first intron boundary is marked, along with the initiation sites determined by primer extension. The major transcription site is labelled +1, and minor sites are 2, 3, 4 and 5. Oligonucleotide locations are given below the sequence.

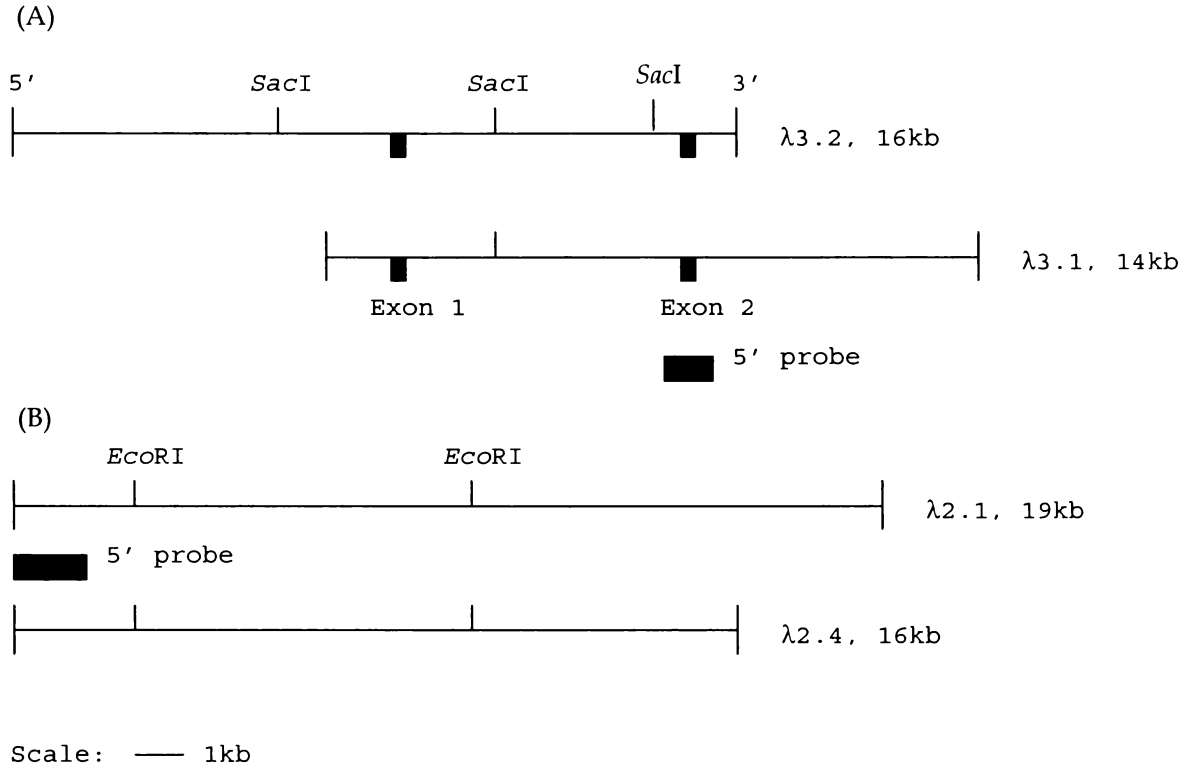


Figure 3.2 Lambda genomic clones containing overlapping sequence (A) λ3.1 and λ3.2 contain approximately 10kb of overlapping sequence, λ3.2 extends further 5'. Neither contain exon 3. The 5' probe (*Pst*I 400bp) from bovine cDNA predominantly hybridised to exon 2, and also partially to exon 1. (B) λ2.1 and λ2.4 genomic clones contain 19 and 16kb of sequence respectively. λ2.4 is completely duplicated by λ2.1. The orientation of these clones was not determined. However absence of overlap with λ3.2 restriction sites suggests they extend further 3', the 5' probe potentially hybridises to exon 3 which is not present in λ3.1/λ3.2.

oligonucleotide compared to 1.11(2) meant the overall orientation of the clone could be determined.

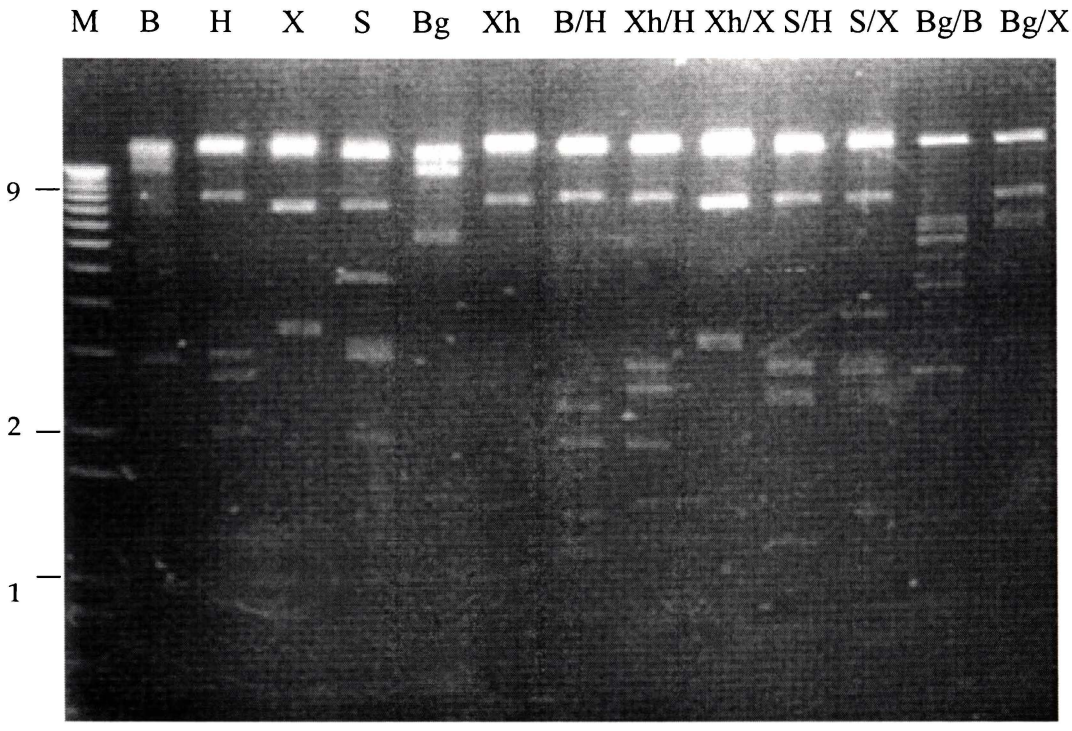
3.2.3 Subcloning and sequencing of the 5' UTR

The clone λ3.2 contained approximately 9kb of 5' DNA sequence plus approximately 6kb of the gene including exons 1 and 2 and intron 1 (depicted in Figure 3.1). The exact position of the first exon was unknown, so a fragment spanning the approximate area was subcloned for further analysis. This encompassed 4.9kb from the *Sac*I sites and was cloned into pBSIIKS (Stratagene). The resulting clone (5' SacI 4.9kb) was confirmed to be accurate by digestion with enzymes digesting at known sites (*Hin*dIII, *Bam*HI, *Xba*I). Further, a Southern blot of these digests was hybridised in successive rounds (stripped between hybridisation) by oligonucleotides 5'seq and 1.11(ATG) which bound at the *Bam*HI-*Hin*dIII 1.3kb fragment and 1.11(2), which did not hybridise. This suggests the ATG translation start codon is indeed present in this subclone, along with an intron that separates the ATG containing exon from the sequence encoding the 1.11(2) oligonucleotide. This can be seen from the probe binding sites shown in Figure 3.1.

The clone 5' SacI 4.9 was then further subcloned into fragments for sequencing, again into the vector pBSIIKS (Stratagene). Fragments were required to be less than 1.5kb in length to allow complete sequence verification from both ends, with some overlap. The subclones of 5' SacI 4.9 are indicated in Figure 3.4, and the complete sequence is given in Appendix 1.

Either the T3/T7 or the forward/reverse combination of primers was used for sequencing from each end of each clone, by automated sequencing either at the Waikato or Massey University DNA sequencing facilities. The location of the first exon-intron boundary was noted by comparison with the cDNA sequence, shown in Figure 3.1. This indicated that the ATG codon would indeed be in exon 2.

(A) Fractionation of DNA restriction enzyme digestion products by agarose gel electrophoresis



(B) Autoradiograph of a Southern blot, hybridised with the 1.11(2) oligonucleotide

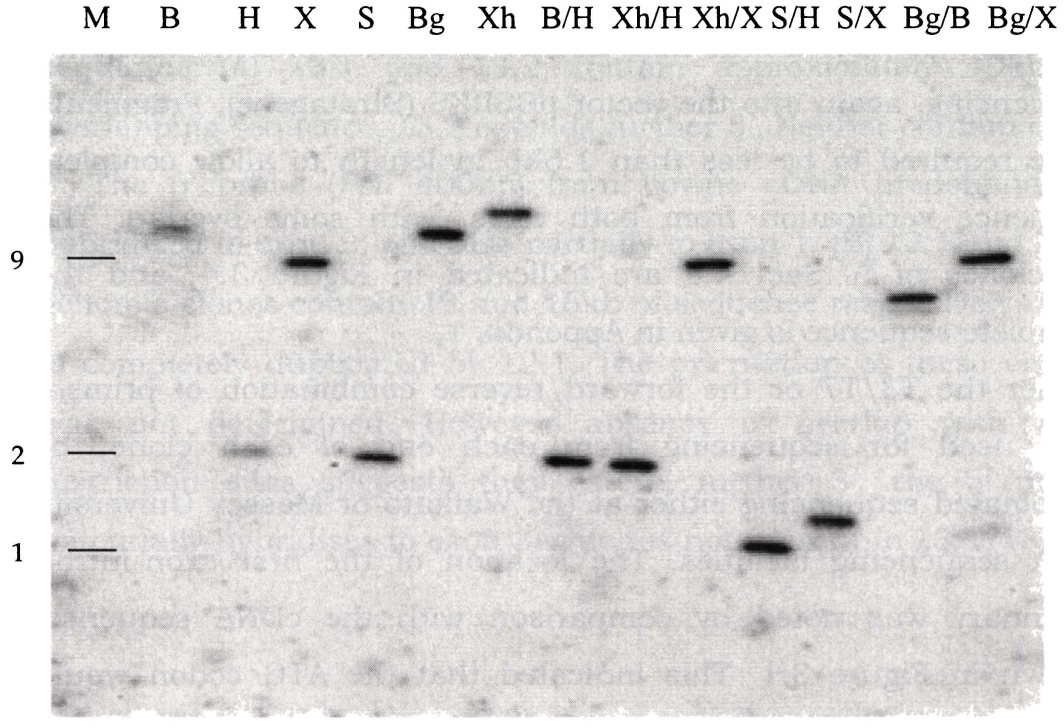


Figure 3.3 Preliminary mapping of the λ3.2 genomic clone (A) An agarose gel photographed under UV light. The gel shows λ3.2 DNA digested with restriction enzymes as indicated above the photo. Abbreviations are as follows: M- 1kb Marker, B- *Bam*HI, H- *Hin*dIII, X- *Xba*I, S- *Sac*I, Bg- *Bgl*II, Xh- *Xho*I. When two enzymes were used, both abbreviations are given for that lane. (B) Southern blot analysis of the agarose gel from part (A), probed with the 1.11(2) oligonucleotide to allow the alignment of restriction fragments. Marker sizes are indicated to the left of the marker lane, in kb.

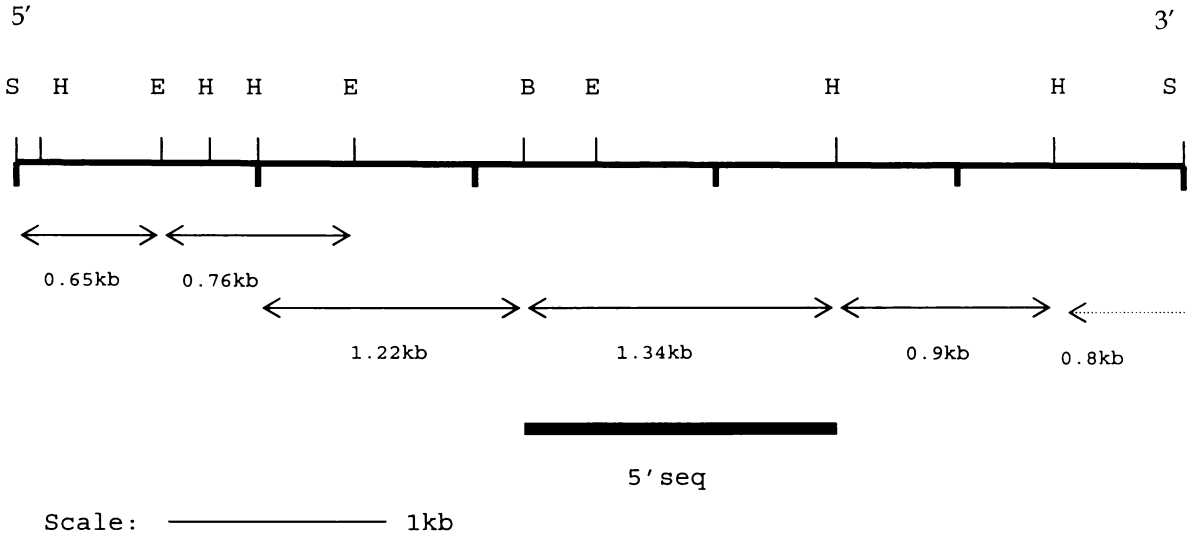


Figure 3.4 Restriction map depicting the clone 5' SacI 4.9
Restriction sites are as follows: S - *Sac*I, H - *Hin*dIII, E - *Eco*RI, B - *Bam*HI. Subclones for sequencing are indicated by the lines ending with arrows and their approximate size is shown below. The dashed line represents sequence generated from a vector priming site (T7). The location of the 5'seq oligonucleotide hybridisation is shown by a bold line below the subclones.

3.2.4 Identification of multiple transcription start sites

The experimental identification of a transcription start site or the 5' termini of mRNA can be achieved by several methods. These include primer extension, ribonuclease protection assay and 5' RACE. Each has limitations, so to overcome these and achieve a reliable result two complementary methods were used as verification. Several software packages are available for the prediction of transcription start sites and promoters. A brief study using these programs was conducted and is presented in Chapter 4.

The location of transcription start sites for the murine *Ese*-2 promoter was first investigated using primer extension.

Primer extension is a widely used method for the determination of initiation sites. Reverse transcription is initiated from a radiolabelled primer, producing cDNA products, the size of which is measured by rate of migration through a sequencing polyacrylamide gel. The nucleotide lengths of the cDNA products are estimated so that the initiation site of the RNA can be identified on the genomic sequence. In cases of multiple transcription start sites, the yield of each product is proportional to the strength of initiation from that site.

Initial primer extension experiments using mouse pregnant mammary gland total RNA as a template and PE 35 as the primer did not show any products (data not shown). The concentration of the RNA was later increased 10 fold to 100µg, which was found to give five products, which could possibly represent five different initiation sites. A total of six separate experiments confirmed the presence of these same product sizes. Three different preparations of RNA were used in these experiments. One of these autoradiographs is shown in Figure 3.5. Clearly the major transcription start site is the distinct G nucleotide (labelled 1 in Figure 3.5), where as any distinction between the other

transcription start sites is less clear. The position of this product at nucleotide G is therefore labelled as the major transcription start site, and is given the position number +1 in all sequence references here-in. Four less abundant products represent minor transcription start sites, and are at positions -2, -4, -8 and -41.

Primer extension does not confirm the sequence of the product, only the length. 5' RACE is useful because the actual sequence is obtained. This ensures that the correct gene is being analysed and also that there are no further 5' exons.

The first step in 5' RACE involves the use of a gene specific primer that is homologous to sequences downstream of the predicted initiation site (Frohman, 1993). First strand cDNA synthesis is performed from this primer. The cDNA then has a homopolymeric tail added to the 3' end. The modified cDNA is then amplified by PCR using a nested gene specific primer and a novel deoxyinosine-containing anchor primer that anneals to the homopolymeric tail. The products of the PCR reaction can then be cloned and sequenced. Several control reactions were performed due to an initially high background, possibly resulting from non-specific binding of the anchor primer. Single oligonucleotide primer reactions were also run to ensure that neither primer was binding non-specifically (seen in Figure 3.6(A)).

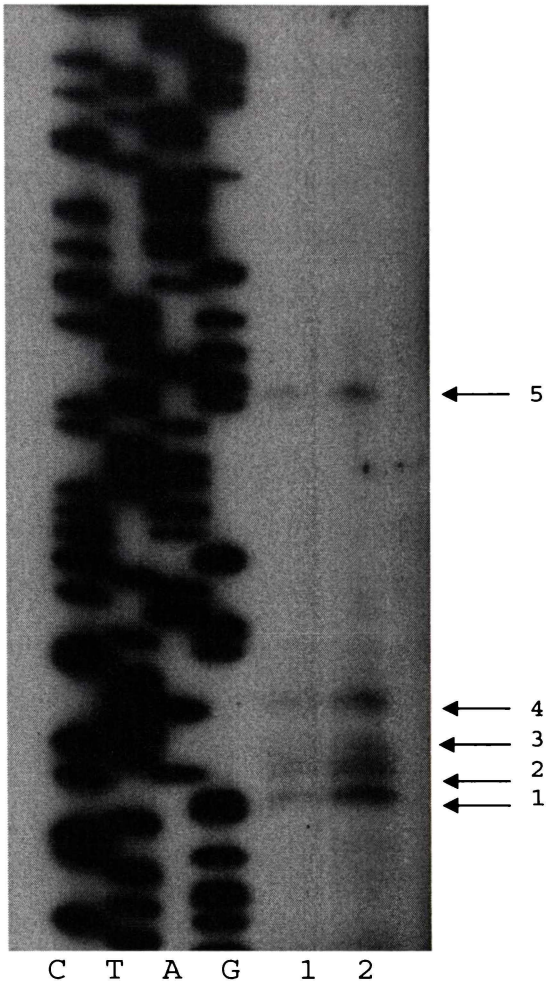
This procedure was performed on three different RNA samples from mouse mammary gland. Thirty-eight clones were picked and sequenced manually to determine their length. The majority of clones were clustered about 93bp upstream of the ATG as is shown in Figure 3.6. The Figure shows an example of the sequenced clones, where the stretch of 'G' nucleotides is the anchor primer, and the length of the clone is taken from the nucleotide prior to the first G.

Of the clones sequenced, eighteen were from a one day lactating mouse mammary gland. Eleven ended within seven nucleotides of the major site, and seven others near the second cluster of sites further 5'. Fourteen clones derived from the polyadenylated RNA

sample (mouse mammary gland RNA from a day 17 pregnant animal) were sequenced, revealing an equal split in clones terminating at both sites.

5' RACE results show a scattered pattern of lengths of mRNA transcripts. This suggests a secondary structure problem that is not seen in the primer extension. RACE has confirmed the sequence of the 5' untranslated mRNA, and the location of the sites from RACE are all in near proximity to those found by primer extension.

(A)

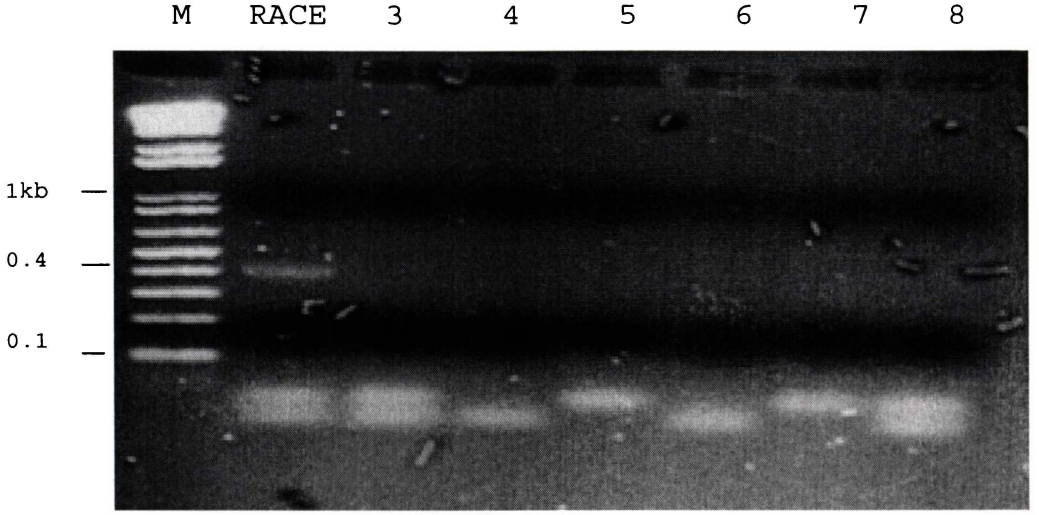


(B)

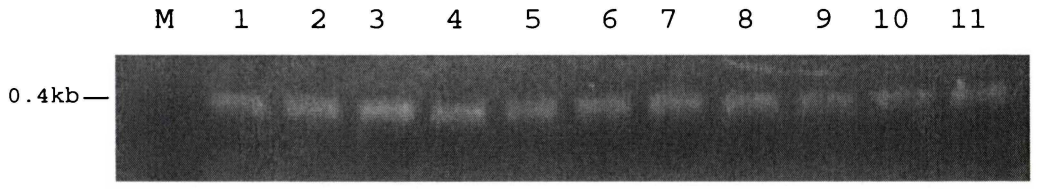
```
                                                    5
                                                    ↓
C C A A A C G G A C C T G T C T G T A G G T G T C -27

A C T T A T A T C A C A C G G C T A C A G G T G C -12

        4       3   2   1
        ↓       ↓   ↓   ↓
C T T T A T T T C T A C G G T C C G C T G G T G C +13

T G G G A G C G C G C T T G C C T T C T C T T G C +38

C T T G A A A G C C T T C T G T C T G G A C C T A +63

G C C A C C A C T T G T C T T C A C G G T G A T G
```

Figure 3.5 Primer Extension results (A) Two primer extension reactions in lanes 1 (10ug RNA) and 2 (100ug RNA) run alongside a sequencing reaction from the same primer (PE 35) on genomic DNA. Arrows indicate the primer extension products, labelled 1 to 5. (B) Sequence of Ese-2 cDNA (includes ATG, and also some 5' genomic sequence), arrows indicate positions of products, and hence start sites.

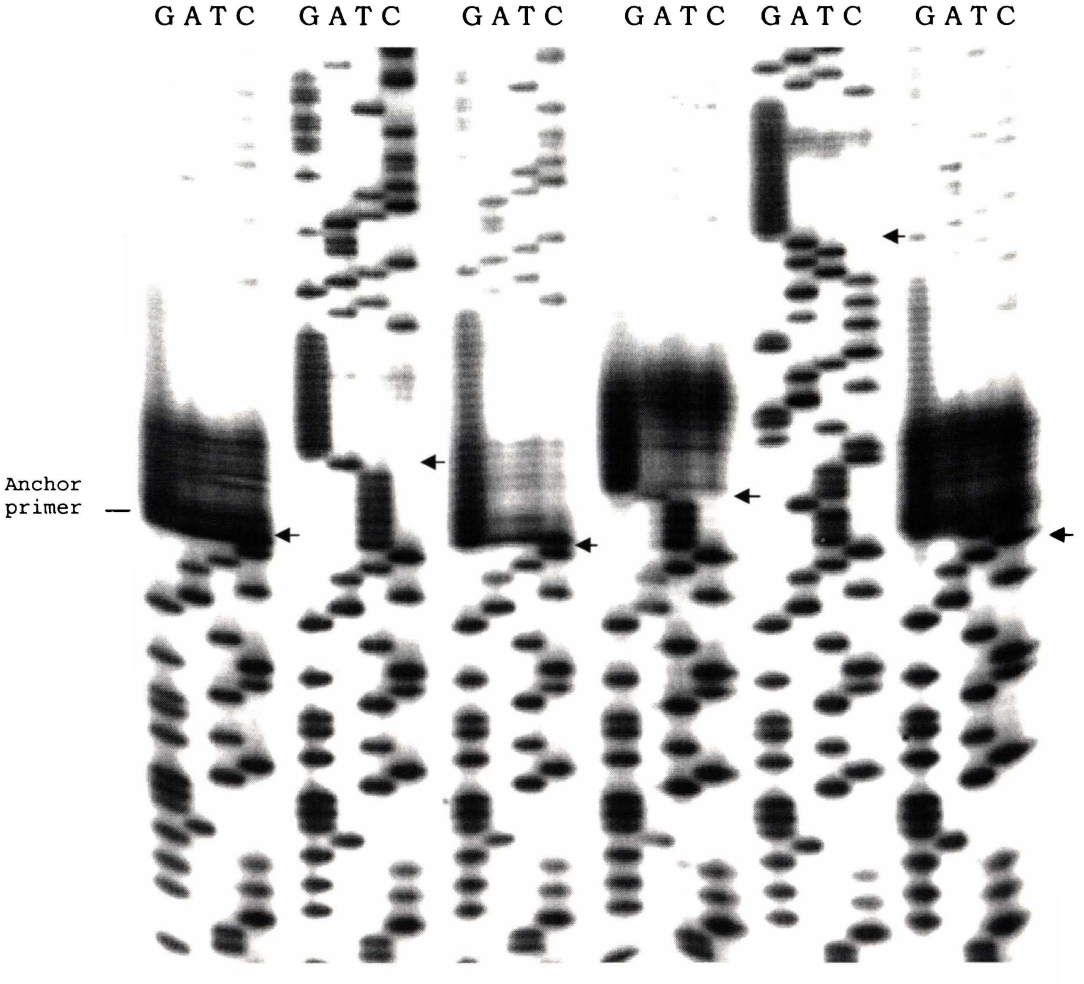
(A) 5' RACE product and controls on Agarose



(B) 5' RACE subclones



(C) Sequenced clones



(D) Sequence of 5' UTR and lengths of clones

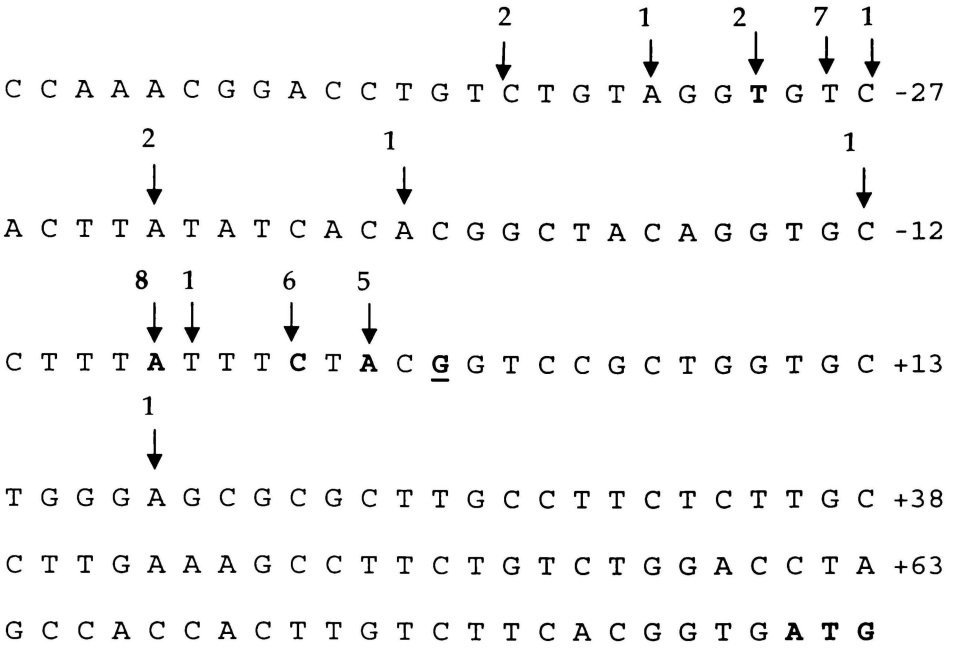


Figure 3.6 5' RACE experimental results (A) 5' RACE PCR product and control reactions fractionated on a 1% agarose gel. Lane M is a 1kb marker, with actual sizes shown on the left (in kb). The RACE lane is the PCR reaction containing two primers, template cDNA that had been tailed by terminal transferase. Lanes 3, 4, 5, 6, 7 and 8 all contain negative control reactions. (B) Cloned products of the 5' RACE that have been digested to release the approximately 400bp product. (C) Manual sequencing of the 5' RACE clones, with the nucleotide above the lane. Shown are six clones sequenced on the acrylamide gel. Arrows indicate the nucleotide at the 5' end of the transcript. (D) Locations of 5' clone ends, shown in reference to the actual cDNA sequence, with the ATG translation start codon at the 3' end of the sequence shown. The nucleotides marked in bold are the sites identified by primer extension and the major transcription start site is underlined. The numbers above the arrows refer to the number of RACE clones mapping to each product size.

3.2.5 Computer predicted analysis of potential regulatory regions

Computer predicted analyses of promoter sequences have previously been performed on many genes to identify potential binding sites, based on consensus sequences, for known transcription factors. However, the identification of a potential binding site does not mean that it is functional.

A computer analysis was performed to search for the presence of potential regulatory elements in the *Ese*-2 promoter and first intron sequence, using the programs TFSEARCH (Akiyama, 1995) and MatInspector V2.2 (Quandt *et al.*, 1995).

As is often observed in promoters with multiple transcription start sites, there is no typical TATA box present in *Ese*-2 within at least 1.5kb upstream of the major transcription start site (Crawford *et al.*, 1999). Outside of this region there are consensus TATA sequences, but they would not be expected to operate in cis, as typical boxes operate at –25bp from the transcription start site (Wiley *et al.*, 1992). However, a degenerated TATA box may be operational. At –25bp a DNA sequence, which has 5 of 6 nucleotides conserved with the consensus sequence, is present.

No CCAAT box was identified by the computer search in the immediate 5' region, though several consensus CCAAT elements do lie further 5' (-557, -667, -835bp). Typically, a functional CCAAT box would be present closer than this to the TATA-box, though neither CCAAT nor TATA are necessary features of a promoter.

Analysis of the proximal 2.5kb upstream region from the start site revealed a multitude of potential sites. Several hundred consensus binding sites were identified. Many would appear to be inconsequential and deciding which elements are relevant is difficult without functional analysis. The use of promoter recognition programs may help determine those elements of importance, and is performed in Chapter 4. In this study, the

elements discussed are those that have been previously found in other mammary gland genes and ets transcription factors. These are shown in Figure 3.7, which shows only the first 1017bp upstream, as no unique sites were found further 5'. The consensus sites include activator protein-1 and -2 (Ap-1, Ap-2) (Briggs *et al.*, 1986; Faisst and Meyer, 1992; Lee *et al.*, 1991; Lee *et al.*, 1987; Yamauchi *et al.*, 1996) at more than 10 and 5 positions respectively, GATA elements in multiple locations (Faisst and Meyer, 1992; Whyatt *et al.*, 1993;Aird *et al.*, 1994; Kawana *et al.*, 1995;), three octamer 1 binding sites (Oct-1) (Sturm *et al.*, 1988; Roth *et al.*, 1991; Faisst and Meyer, 1992;) with 100% consensus sequence, and two Signal Transducer and Activator of Transcription (STAT) sites with 100% sequence homology to the consensus sequence. One serum response factor (SRF) consensus site was identified in the 5' promoter region, which is known to be important in combination with Ets factors. In addition, potential regulatory sites that occur in the proximal 5' region to the major start site include two GC-box elements all within 1kb of this site, two YY1 (Yin and Yang-1) elements at -701 and -763, three Sp-1 sites (Briggs *et al.*, 1986) located between -100 and -900bp, and also two NF-KB elements with 100% identity to the consensus at positions -618 and -669. c/EBP (CCAAT/Enhancer binding protein) consensus elements are found in at least 7 locations 5' of +1. Their potential importance is related to a known influence of c/EBPβ on cell fate determination during mammary gland development (Seagraves *et al.*, 2000). Lastly, of the elements recognised were multiple ETS binding sites. In particular, an Elf-1 site at -2.5kb and multiple c-Ets-1 sites (Faisst and Meyer, 1992; Tymms and Kola, 1994).

Following the analysis of the proximal and 5' promoter regions, the first intron was examined for consensus binding sites that may play enhancer roles in the transcriptional regulation of *Ese*-2. The first intron sequence contained hundreds of consensus binding site

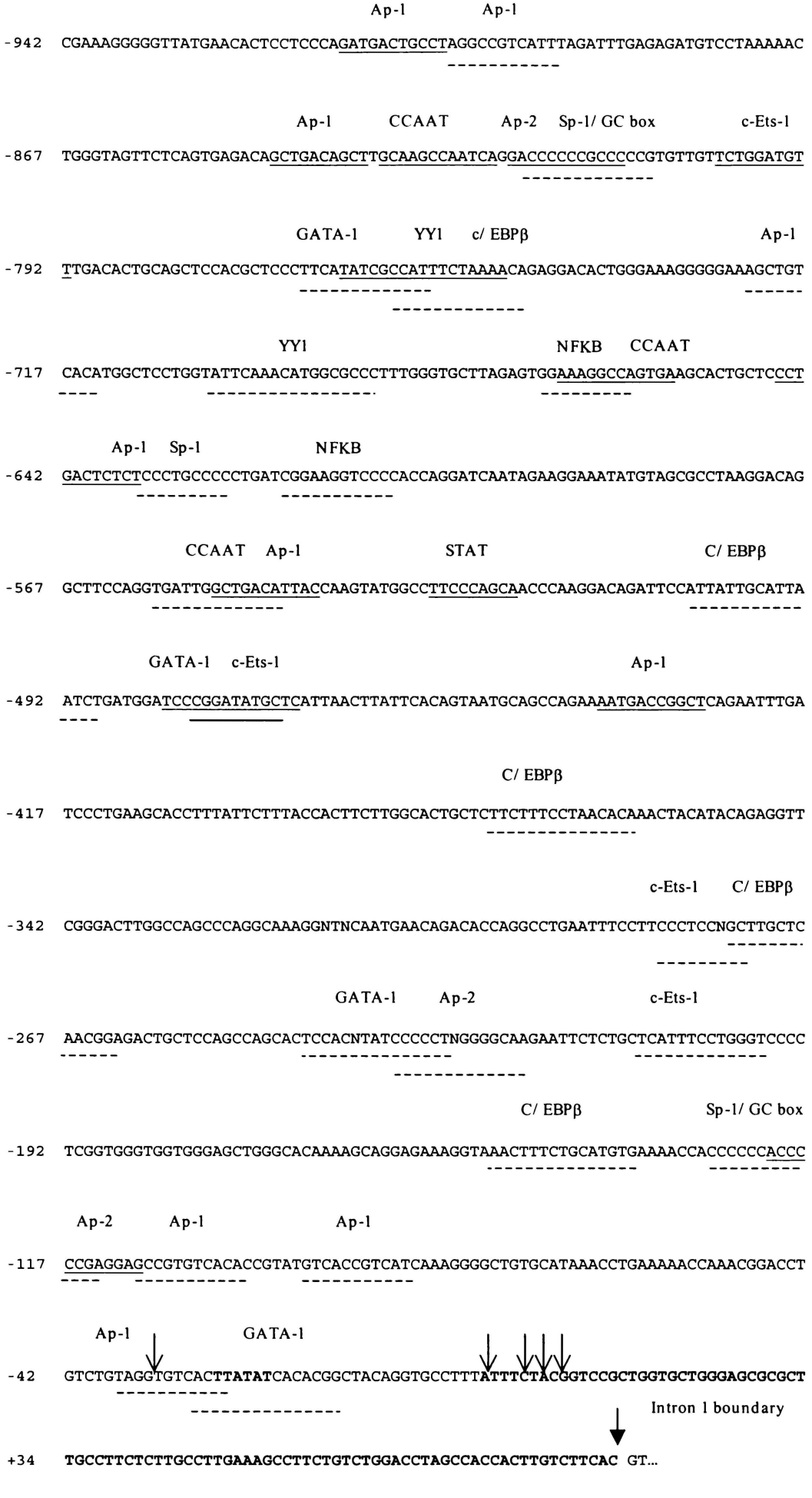

Ap-1 Ap-1
-942 CGAAAGGGGGTTATGAACACTCCTCCCAGATGACTGCCTAGGCCGTCATTTAGATTTGAGAGATGTCCTAAAAAC
Ap-1 CCAAT Ap-2 Sp-1/ GC box c-Ets-1
-867 TGGGTAGTTCTCAGTGAGACAGCTGACAGCTTGCAAGCCAATCAGGACCCCCCGCCCCCGTGTTGTTCTGGATGT
GATA-1 YY1 c/ EBPβ Ap-1
-792 TTGACACTGCAGCTCCACGCTCCCTTCATATCGCCATTTCTAAAACAGAGGACACTGGGAAAGGGGGAAAGCTGT
YY1 NFKB CCAAT
-717 CACATGGCTCCTGGTATTCAAACATGGCGCCCTTTGGGTGCTTAGAGTGGAAAGGCCAGTGAAGCACTGCTCCCT
Ap-1 Sp-1 NFKB
-642 GACTCTCTCCCTGCCCCCTGATCGGAAGGTCCCCACCAGGATCAATAGAAGGAAATATGTAGCGCCTAAGGACAG
CCAAT Ap-1 STAT C/ EBPβ
-567 GCTTCCAGGTGATTGGCTGACATTACCAAGTATGGCCTTCCCAGCAACCCAAGGACAGATTCCATTATTGCATTA
GATA-1 c-Ets-1 Ap-1
-492 ATCTGATGGATCCCGGATATGCTCATTAACTTATTCACAGTAATGCAGCCAGAAAATGACCGGCTCAGAATTTGA
C/ EBPβ
-417 TCCCTGAAGCACCTTTATTCTTTACCACTTCTTGGCACTGCTCTTCTTTCCTAACACAAACTACATACAGAGGTT
c-Ets-1 C/ EBPβ
-342 CGGGACTTGGCCAGCCCAGGCAAAGGNTNCAATGAACAGACACCAGGCCTGAATTTCCTTCCCTCCNGCTTGCTC
GATA-1 Ap-2 c-Ets-1
-267 AACGGAGACTGCTCCAGCCAGCACTCCACNTATCCCCCTNGGGGCAAGAATTCTCTGCTCATTTCCTGGGTCCCC
C/ EBPβ Sp-1/ GC box
-192 TCGGTGGGTGGTGGGAGCTGGGCACAAAAGCAGGAGAAAGGTAAACTTTCTGCATGTGAAAACCACCCCCCACCC
Ap-2 Ap-1 Ap-1
-117 CCGAGGAGCCGTGTCACACCGTATGTCACCGTCATCAAAGGGGCTGTGCATAAACCTGAAAAACCAAACGGACCT
Ap-1 GATA-1
-42 GTCTGTAGGTGTCACTTATATCACACGGCTACAGGTGCCTTTATTTCTACGGTCCGCTGGTGCTGGGAGCGCGCT
Intron 1 boundary
+34 TGCCTTCTCTTGCCTTGAAAGCCTTCTGTCTGGACCTAGCCACCACTTGTCTTCAC GT...

Figure 3.7 Sequence of murine *Ese*-2 gene promoter The proximal sequence is shown from –1017bp relative to the major transcription start site at +1, as determined by primer extension. The sequence includes exon 1 (bold), a putative TATA box also in bold and the transcription start sites marked with open arrows. Potential regulatory factor binding sites are underlined with solid lines for sense orientation and broken lines for antisense. Factors are named above the sequence.

sequences. Those that have been identified in other specific mammary gland genes and *ets* genes are mentioned here. In particular there were several AP-1, GATA and C/EBPβ sites. Ets sites were limited to one Elf-1, one Elk-1 and four c-Ets-1. The Elf-1 site was in the first 1kb of intron sequence, and the Elk-1 at approximately 2kb. Two NF-KB sites were identified, both in the 3' end of the intron. Only one GC- box was found with the computer search. The importance of these sites cannot be predicted from this method of analysis, but may become useful when deletion studies are being performed.

Repeat sequences were detected using RepeatMasker (Smit and Green, 2000). In the 5' promoter region an alternating pyrimidine-purine repeat (TG) was located, along with three SINE (short interspersed nucleotide elements) elements. One of these was a B1 element, typical of a rodent repetitive sequence, corresponding to the human Alu repeat sequence (Bois *et al.*, 1998). In the first intron, several simple repeats were located (TTTTTG), (CA), (TGGGGG), (GA). Also, three MIR (mammalian wide interspersed repeats) (Smit and Riggs, 1995) and two low-complexity AT rich stretches and a MER58 element were found in the first intron.

The regulatory elements discussed above are only a sample of those detected in the analysis and are those that are consistent with other *ets* genes regulatory sequences. Another use for this data is in the comparison of the mouse *Ese*-2 sequence with that of the human, recently published in the human genome project. Regions of homology between the species in the promoter and other regulatory regions give clues as to the location of important regulatory regions. Conserved regions between species often suggest a slower rate of evolution because they are more important than those areas evolving quickly. Alignments of mouse and human promoter and intron 1 sequences are performed in the following chapter.

3.3 Discussion

This chapter describes the preliminary identification and characterisation of the *Ese*-2 promoter region. Firstly, this involved the localisation of a genomic clone containing a suitable sequence of DNA oriented 5' to the cDNA start. This was subsequently subcloned and sequenced so that regions involved in *Ese*-2 regulation could be analysed.

In order to identify the actual promoter region, primer extension analysis was performed. A total of five transcription start sites were identified by repeated primer extension experiments. The most abundant band was labelled as the major transcription start site and numbered as +1. The sequence of these start sites were confirmed by 5' RACE. 5' RACE results showed two main clusters of products, each ranging in size over 8-10 nucleotides. Both clusters were grouped around the region of the nucleotides identified by primer extension as transcription start sites. 5' RACE resulted in a total of thirteen different product sizes. The scattering effect seen by this method could be due to secondary structure formation that prevents full reverse transcription. This is more likely in 5'RACE than primer extension due to the length of the product. Also, three products of 5' RACE were longer than the longest primer extension product. This was unusual and may imply that a more distal initiation site is present that was not detected by primer extension or possibly a distal exon. Alternatively, a contamination of genomic DNA could explain the longer products.

Limitations of primer extension include the low assay temperature (37°C), which could permit RNA to form secondary structures that might interfere with movement of the reverse transcriptase along the template RNA. It is also the preferred temperature for RNase degradation. In order to overcome this, primers were designed in close proximity to the predicted site (based on early 5' RACE

attempts) so that extension products would be short and secondary structure minimised.

The RACE technique requires the use of a PCR reaction, which introduces a degree of variability. Certain populations may be amplified at a higher rate due to length, annealing of primers, which results in a biased representation of the starting template. 5' RACE was therefore performed with different RNA samples, including polyadenylated RNA and different states of development of the mammary gland. RACE also requires a minimised secondary structure, as this may effect the reverse transcription by inhibiting the enzyme from reading along the template. This may have been the cause of the RACE results in this study, and was thought to be critical in the identification of start sites for *Fli*-1. *Fli*-1 is an *ets* gene, found to contain two major initiation sites. However eighteen shorter products of 5' RACE were also seen (Barbeau *et al.*, 1996). The ends of these clones were non-randomly distributed over 387bp of sequence, not corresponding to any previously published mouse *Fli*-1 cDNA 5' ends. The shorter products were located in two main clusters, the more 3' cluster corresponding to the 3' end of the first exon. Barbeau *et al.*, explain this phenomenon of 5' RACE as an artifact due to a highly stable hairpin structure predicted to be formed by the first exon. This secondary structure may interfere with the reverse transcriptase step in 5' RACE causing smaller products that are made due to incomplete reverse transcriptase synthesis.

Perhaps the best-characterised *ets* promoters are those of *Ets*-1 and *Ets*–2. *ETS*-1 transcription initiation sites were identified by primer extension and S1 nuclease protection analysis (Jorcyk *et al.*, 1991). A total of fourteen different sites were located, spanning a region of 137bp by S1 nuclease mapping. Primer extension also identified multiple sites, though the 3' sites were more favoured. This was explained as a result of inefficient procession of the reverse transcriptase to upstream regions (Jorcyk *et al.*, 1991). This phenomenon was also observed here with *Ese*-2 primer

extension. The smaller bands were slightly favoured, though it is not known if this is an artifact of reverse transcriptase inefficiency or not.

Mavrothalassitis *et al.* (1990) determined *ETS*-2 transcription initiation sites in a similar manner. Again, multiple (8) major sites were found, as well as many minor sites by both S1 nuclease and primer extension. To determine the effect of any secondary structure, an *in vitro* synthesised RNA template of the same region was included in the analysis. No extra bands were seen with this template, indicating no secondary structure problems (Mavrothalassitis *et al.*, 1990).

More recently a method for determining the transcription start site has become available, that involves detecting RNA transcripts that have a 5' cap. This is now the preferred method to determine the RNA 5' end (Suzuki *et al.*, 2001).

Both *ETS*-1 and -2 lack a typical TATA box. They are found to contain consensus sites for regulatory elements including SRE-like, GC-rich, EBS and Ap-2 elements (Mavrothalassitis *et al.*, 1990; Jorcyk *et al.*, 1991). In contrast to these promoters, the *Elk*-1 promoter was shown to contain a single transcription start site and a typical TATA box. Potential binding sites for transcription factors included Sp-1, Ap-1, cAMP response element binding protein and the CCAAT binding protein, as well as binding sites for Egr-1 and SRF/TCF (Lehmann *et al.*, 1999).

Interestingly, the *Ese*-2 major start site lies approximately 25bp downstream of a degenerate TATA box element. The consensus sequence of the TATA box (TATAAA) is found to be altered in the Ese-2 sequence (TTATAT). The presence of multiple start sites within a promoter is a feature seen in common with TATA-less promoters. The functionality of the *Ese*-2 potential TATA box is not known at this point.

A TATA-box at –25bp relative to the transcription start site is found in less than 50% of proximal promoters (Crawford *et al.*, 1999). In such promoters it is bound by TFIID, an essential element for all

mRNA transcription (Roeder, 1991; Smale, 1994). In TATA-less promoters, other initiator sequences are important for transcription and TFIID component binding including Sp-1 binding sites or interactions of binding sites within the proximal promoter (Martinez *et al.*, 1994; Smale, 1994; Emami *et al.*, 1995; Ernst and Smale, 1995b; Weis and Reinberg, 1997). The proximal promoter of a gene contains many DNA sequences that bind transcription factors (TFIID and others), accurately position RNA pol II for transcription initiation and also direct the rate of polymerase initiation (Roeder, 1991; Smale, 1994; Ernst and Smale, 1995a; Yean and Grall, 1997). This does not take into account the effect of distant enhancers or chromatin structure, so overall conclusions of regulation within the natural context of the whole genome are difficult to support.

The use of computer programs to generate information on potential promoters and transcription start sites is becoming more robust. Such programs are more sophisticated than using simple consensus binding sites as they incorporate information on distances and sequences from known promoters. The use of a transcription factor search engine is of some use and may help validate an experimental result, or suggest other locations are possibly not detected experimentally.

The consensus binding site search for potential regulatory transcription factors has identified a complex range of possible target sites. This process must be treated with caution and results taken as an indication only that a certain element may be able to bind and not that it necessarily does. Analysis often results in the identification of hundreds of consensus sequences, which would be difficult and time consuming to confirm. Though the threshold for mismatches can be altered in such searches, factors are known to tolerate some variance in binding sequence, so again care must be taken.

The transcription factor search was used on the first intron as well as the 5' promoter region. The first intron of *Ese*-2 is very large,

approximately 6kb, which has been previously documented in other genes as potentially harbouring enhancer elements (Aronow *et al.*, 1989). Often enhancer elements operating from a distance cause greater or no change (due to redundancy) to regulation than is seen in proximal elements (Crawford *et al.*, 1999). Functional redundancy in regulatory regions may be widespread for enhancer or tissue-specific regulatory regions in particular as seen in the *Adh Drosophila* locus (Fang and Brennan, 1992; McKenzie *et al.*, 1994) and the mammalian *globin* gene (Gumucio *et al.*, 1993). *Ese*-2 is tissue specific and its tightly regulated pattern of expression may be due to enhancer elements acting from a distance.

An in depth discussion of the potentially important transcription factors is given in the main discussion, Chapter 6.

Thus far, in comparison to members of the Ets transcription factor family, *Ese*-2 has many features in common. Of those *ets* genes that have regulatory regions described, a majority appear to lack TATA and CCAAT consensus sequences, as is the case for *Ese*-2. This is common for house-keeping, growth related genes and oncogenes, the latter of these describing many ets factors. Both *ETS*-1 and *ETS*-2 human genes lack these sequences and again implicate other sequences as being critical in the replacement of their function within the proximal promoter. The TATA box normally defines the exact point of transcription, but it is also neither necessary nor sufficient for such initiation alone.

Of potentially important elements, Sp-1, Ap-1, Ap-2, GC-elements all appear in *ETS*-1 and *ETS*-2, and also *Fli*-1. Two GC elements in the human *ETS*-1 promoter are located near a transcription start site and are postulated to play a role in transcription initation as shown in other multiple transcription start site and TATA-less promoters (Ishii *et al.*, 1985; Miyamoto *et al.*, 1988). *Fli*-1 also contains adjacent GATA and EBS, which has been described earlier as having a potentially important cooperative role (Barbeau *et al.*, 1999). Likewise, *Elf*-3 has a partial overlap between EBS and NFKB (Oettgen *et al.*, 1999a).

The importance of repeat sequences is somewhat less clear in the promoter than in other regions of the gene. Jorcyk *et al.* (1991) speculate that a polypurine-polypyrimidine stretch may be important in *ETS*-1 regulation. Certainly in other promoters, such a repeat is able to form a triple helix which is sensitive to S1 nuclease and able to bind nuclear proteins implicating they are involved in regulation (Gilmour *et al.*, 1989).

Functional analysis of these regions is required to confirm the involvement of any such factors in the regulation of *Ese*-2. The location of such sites provides a suitable basis for an experimental study of the features involved in *Ese*-2 regulation.

Chapter 4. Human genome analysis and sequence alignments

4.1 Introduction

With the recent publication of the draft human genome, a whole new raft of information is available for investigating features of genes. When this data is combined with other genomes, comparisons can be made which increase the capability to investigate the structure, function and evolution of genes. The more recent release of the mouse genome in particular provides another level of information.

The human genome is much larger than other genomes recently sequenced. It is thirty times larger than the worm and fly, and two hundred and fifty times greater than yeast. However, it does not have a proportional increase in the number of genes and indeed estimates are now predicting 25-35000 genes (Ewing and Green, 2000). This means there is alot of non-coding DNA. Of this, 46% is made up of repeat DNA, which may or may not have a function. The remainder includes promoters, transcription regulatory sequences and other unknown features. The increase in size of the human genome from other genomes brings with it an increase in complexity, and makes the location of individual genes more difficult through the decrease in signal to noise ratio.

Computational strategies, based on finding gene characteristics such as splice sites and coding regions, often lead to false positives. More reliable predictions are made through the use of expressed sequence tags (ESTs) and complementary-DNA homology comparisons to other genomes, because exons are more highly conserved than introns.

The human genome has been compiled by the systematic sequencing of several thousand (10,000-20,000) Bacterial artificial

chromosome clones (Bac), each of 100-200kb, which come from known chromosomal locations. The sequenced fragments are assembled based on overlaps and then have a unique identifier assigned (accession number) by the high-throughput genome sequence division of the international nucleotide database (GenBank/EMBL/DDBJ). The overlap of the Bac clones is aligned and redundancy removed to create a contiguous DNA sequence for each chromosome.

Many errors exist in the current data set and caution must be exercised with any analysis. Fragmentation of genes can cause misleading results, often predicting more than one gene present. Assembly errors may duplicate genes or give the wrong orientation. Also, the actual presence of a gene does not mean a protein will be expressed, and therefore expression studies must complement such information to confirm a protein.

Wolfsberg, McEntyre and Schuler (2001) outline two main uses of the human genome sequence. Firstly, using a known DNA sequence or protein from a different source, the human sequence can be compared to find an apparent homologue. Indeed this approach has been used in this chapter as the mouse *Ese-2* gene has a human homologue *ESE-2* (Oettgen *et al.*, 1999b; Zhou *et al.*, 1998). Secondly, a group of proteins can be identified that contain a conserved sequence or domain. For example, the ETS domain is highly conserved in this family of transcription factors and a search of this highlights the individual family members in the human genome. Such paralogues are often the result of gene duplication, where a conserved region is found in a set of genes (Wolfsberg *et al.*, 2001).

Variations in the DNA sequence can give information on regions of importance. Commonly utilised are single nucleotide polymorphisms (SNPs) of which 1.5 million have been documented to date. When these occur in coding regions, amino acid substitutions may occur, which may or may not be functionally important.

This chapter deals with the data gained from the published human and mouse genome working draft sequences, and utilises tools for its manipulation (National Centre for Biotechnology Information-NCBI, and UCSC Human Genome Project Working Draft). Firstly, the basic sequence references are discussed followed by a more detailed study of the clones representing *ESE-2* and its predicted gene structure. Exon-intron boundaries are identified, and different isoforms are discussed. Promoter prediction programs have been employed and offer a comparison to the experimental approach. The chapter focuses on the comparison between mouse and human (and bovine) sequence data for *Ese*-2. In particular, the chromosomal location (syntenic gene block), mRNA alignment and EST database information. Blast comparison is also performed to allow a comparison to other genes having some similarity.

The first publication of the *Ese*-2 sequence was for the mouse (Acc. No. AF 049702) and the human (Acc. No. AF 049703) (Zhou *et al.*, 1998). The human sequence was isolated from a human lung cDNA library and was mapped to chromosome location 11p15-p13, which is known to undergo loss of heterozygosity in some cancers (Zhou *et al.*, 1998). The human sequence was later recognised as having alternate 5' exons, predicted to result from two different promoters (Oettgen *et al.*, 1999b). *ESE*-2 was labelled as either the *ESE*-2a (Acc. No. AF 115402) or *ESE*-2b (Acc. No. AF 115403) isoform (Oettgen *et al.*, 1999b). *ESE*-2a encodes a protein of 265 amino acids with a molecular weight of 31.3kD, and contains a 10 amino acid extension at the 5' end to that of *ESE*-2b. *ESE*-2b therefore encodes a 255 amino acid protein of 30.1kD. *ESE*-2a and 2b both contain a longer 3' UTR than that published for *ELF*-5 (Zhou *et al.*, 1998), thought this may be due to a lack of 3' RACE data. *ESE*-2b was equivalent to that published by Zhou *et al.* as human *ELF*-5.

One ATG initiation codon exists in frame, with multiple termination codons in frame upstream of this for *ESE*-2a (no upstream termination codons present for *ESE*-2b). Oettgen *et al.*, 1999 also

report both isoforms as having a poly(A) tract preceded by a consensus polyadenylation signal (ATAAA). An ATTTA motif after the polyA site has been identified, which is associated with rapid mRNA turnover (Savant-Bhonsale and Cleveland, 1992).

4.2 Results

4.2.1 *Ese*-2 sequence data

All analyses completed in this investigation used the human genome draft data released in August, 2001.
Analysis of the murine *Ese*-2 has not revealed any alternate 5' exons or promoters, though it is known that alternative polyadenylation results in two murine mRNA species (Langley, 1999). These are 1.5 and 2.5kb mRNA species that differ in the 3' region, by a truncation in the shorter isoform. *Elf*-5 appears to lack exon 4 (shown in Figure 4.1). This is an anomaly of the UCSC genome browser, as when the sequence is BLAT searched, exon 4 is present. However, this may also be a function of exon skipping that is observed in several EST clones (discussed later).

4.2.2 Locuslink and chromosome structure

Analysis of the human genome draft sequence identifies a single gene match for *ESE*-2 on chromosome 11 band position p13-15, as originally identified by Zhou *et al.*, 1998. The ordered BAC clone that the sequence was constructed from is RP4-594L9, which overlaps clone RP1-53C18. The location of the gene (coding sequence and introns) in the working draft sequence of human chromosome 11 is 36972718 to 37006302 as seen in Figure 4.1. The gene is on the (-) strand, and runs from right to left in the diagram below. Human chromosome 11 working draft segment contains the complete gene sequence from partial exon 1b (*ESE*-2b)

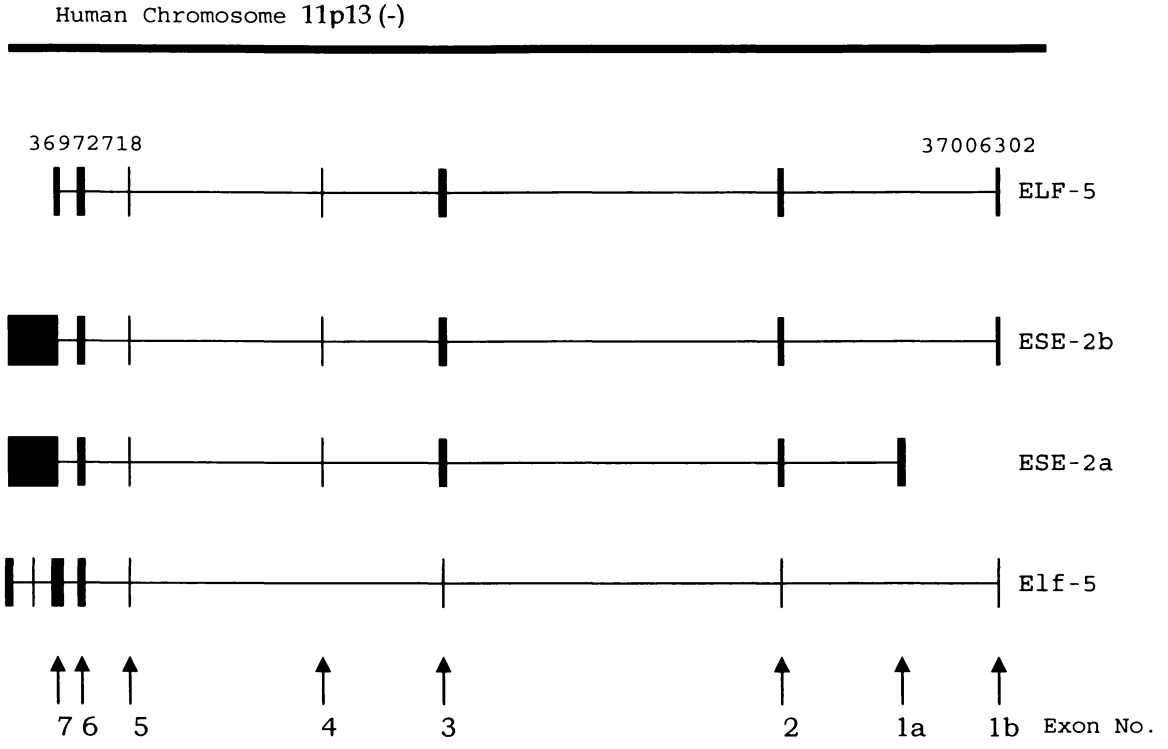
to partial exon 7 and spans 33585bp DNA. The alignment of the genomic DNA and the exon structure is shown in Figure 4.1. Intron boundaries conform to the gt-ag consensus splice donor and acceptor junctions.

4.2.3 Comparison of Mouse and Human *ESE-2*

(a) Chromosomal location

Human *ESE*-2 is located on chromosome 11p13-15, flanked by the catalase (*CAT*) gene and *ESE*-3 *(EHF).* The murine *Ese*-2 gene is located on chromosome 2, in a region that shows close conservation with human chromosome 11, also flanked by the murine Catalase gene (*Cas1*) and *Ese*-3. Figure 4.2 shows the Human Map View of Chromosome 11, focused around the *ESE*-2 locus. Table 4.1 outlines the syntenic block of genes from chromosome 11 of the human, that are grouped in the same order on chromosome 2 of the mouse. *ESE*-3/*EHF*, located at 11p12 retains the most homology with *ESE*-2 in the *Ets* transcription factor family. As discussed in Chapter 1, the *ets* genes appear to have been duplicated in groups of two, always from the same chromosomal region. This would appear to be the case for *ESE*-2 and *ESE*-3 on chromosome 11 of the human and chromosome 2 of the mouse.

(A)



(B)

Exon and Intron positions of *ESE-2* on chromosome 11.

Start nt	End nt	Exon/Intron No.	Length (bp)
37006302	37006197	Partial exon 1b (ESE2b)	105 bp
37004320	37004064	Partial exon 1a (ESE2a)	256 bp
37006196	36998273	Intron 1 gt-ag	5789a/7923b bp
36998272	36998147	Exon 2	125 bp
36998146	36986230	Intron 2 gt-ag	11916 bp
36986229	36985995	Exon 3	234 bp
36985994	36982575	Intron 3 gt-ag	3419 bp
36982574	36982523	Exon 4	51 bp
36982522	36975013	Intron 4 gt-ag	7509 bp
36975012	36974943	Exon 5	69 bp
36974942	36973479	Intron 5 gt-ag	1463 bp
36973478	36973282	Exon 6	196 bp
36973281	36972824	Intron 6 gt-ag	457 bp
36972823	36972718	Partial exon 7	131 bp

Figure 4.1 Alignment of published Human and Mouse *Ese*-2 sequences (A) Genomic DNA showing exon and intron positions on human chromosome 11. The genomic DNA spanning the exons of *ELF*-5 starts at 37006302 on the negative strand, and finishes at 36972718. The sequence reads from right to left, with the untranslated exon 1 starting at the right of the diagram (due to the gene being on the antisense strand). *ELF*-5 refers to the human sequence published by Zhou *et al.*, 1998, and *ESE*-2a and 2b are human sequences published by Oettgen *et al.*, 1999. *Elf*-5 is the mouse homologue published by Zhou *et al.*, 1998. Exon numbers are given below the sequence data. (B) The table represents the coordinates of the exons and introns in the template genomic DNA, relative to the position on chromosome 11. The gene is present on the negative strand, so numbers decrease from 5' to 3'. Comparison of the mRNA sequences submitted to the genome sequence shows that all 6 introns follow the consensus [gt-ag] rule.

Table 4.1 Syntenic Block of Chromosome 11 (human) vs Chromosome 2 (mouse).

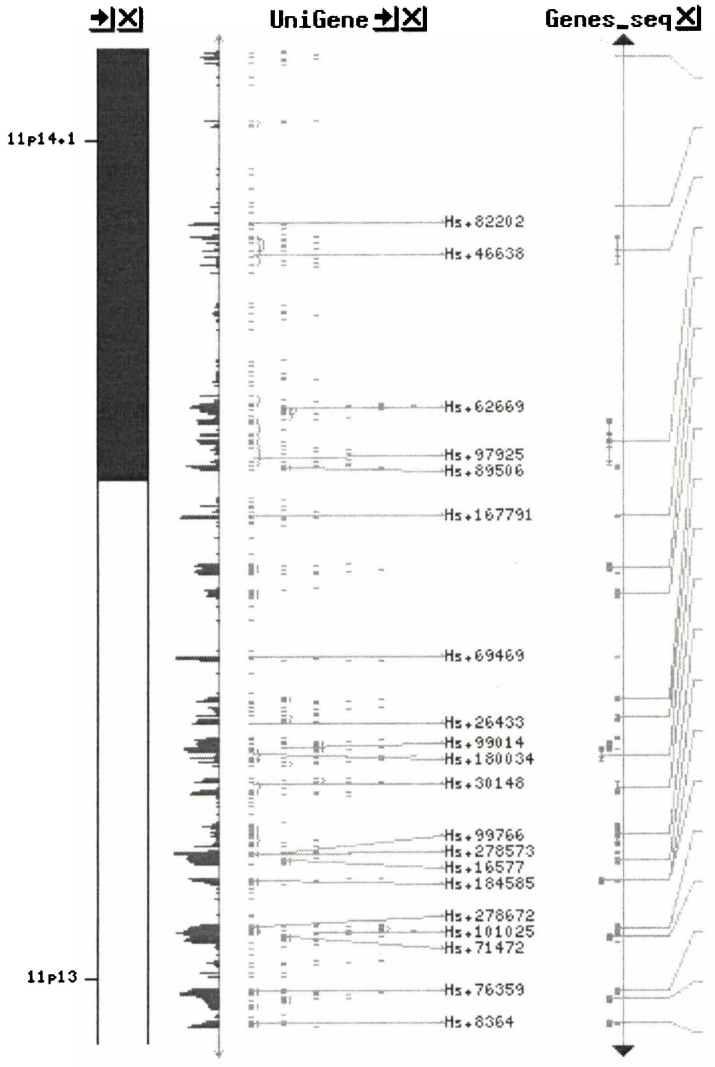
Chr 11	Human gene	Mouse gene	Chr 2	Gene name
11p13	*FSHB*	*Fshb*	60cM	Follicle stimulating hormone, Beta
11p13	*PAX6*	*Pax6*	58cM	Paired box homeotic gene-6
11p13	*RCN1*	*Rcn1*	58cM	Reticulocalbin1
11p13	*WT1*	*Wt1*		Wilms Tumor
11p13	*WIT1*	*Wit1*		Wilms Tumor associated protein
11p13	*HIPK3*	*Hipk3*		Homeodomain interacting protein kinase3
11p13	*CD59*	*Cd59*	55cM	CD59 antigen
11p13	*LMO2*	*Lmo2*		LIM domain only 2
11p13	*M11S1*	*Gpiap-pending*		Membrane component surface marker 1
11p13	*CAT*	*Cas1*	57cM	Catalase
11p13-p15	*ELF5/ ESE-2*	*Elf5/ Ese-2*		Epithelium specific factor 5
11p12	*EHF/ ESE-3*	*Ehf/ Ese-3*		Epithelium specific factor 3
11p13	*CD44*	*Cd44*	56cM	CD44 antigen
11p13	*RAG1*	*Rag1*	56cM	Recombination activating gene-1

(b) Comparison of *Ese*-2 mRNA and genomic sequences

The mouse genomic sequence data released in August 2001 was found to have 97% identity to that isolated in this investigation (not shown). Such differences are may be accounted for by sequencing and assembly errors. The mismatches that occur between sources of sequence mostly comprise gaps in longer runs of a single nucleotide. The results generated in this study were particularly variable for the automated sequence calling at distances over 400bp from the primer.

Sequence comparison of the mouse *Ese*-2 gene (AF_049702) shows 87% homology within the coding sequence to the human *ESE*-2. Comparison to the bovine cDNA sequence shown in Figure 4.3, reveals a homology of 88% to mouse and 92% to human (coding sequence only). The predicted amino acid sequence of human and mouse reveals a homology of 95% and only one amino acid difference in the ETS domain. Interestingly, there is good homology between the bovine 5'UTR and human *ESE*-2a 5' UTR (78%) seen in Figure 4.4. In comparison *ESE*-2b has significant homology with the mouse 5'UTR.

Genomic sequence homology between the mouse and human is outlined in Table 4.3. The 5' promoter region and the first intron were of interest, as these areas were studied for enhancer properties. Regions of homology may indicate an important regulatory sequence due to conservation of such sequences through evolution. A pairwise alignment was performed between the mouse *Ese*-2 genomic sequence from the sequence data obtained in this investigation, extending from 2.6kb upstream of the transcription start site through the first intron, and the human working draft sequence of chromosome 11 to include 5' of the *ESE*-2 gene. Regions of high homology are shown in the table. Briefly there are two short regions of high homology in the extended 5' region, which mapped to 5' upstream regions of the human counterpart.



symb	orient.	links		cyto.	full name
FSHB	↓	sv	ev	11p13	follicle stimulating hormone, beta polypeptide
RCN1	↓	sv	ev	11p13	reticulocalbin 1, EF-hand calcium binding domain
WT1	↑	sv	ev	11p13	Wilms tumor 1
HIPK3	↓	sv	ev	11p13	homeodomain-interacting protein kinase 3
LMO2	↑	sv	ev	11p13	LIM domain only 2 (rhombotin-like 1)
M11S1	↓	sv	ev	11p13	membrane component, chromosome 11, surface marker 1
CAT	↓	sv	ev	11p13	catalase
ELF5	↑	sv	ev	11p13-p15	E74-like factor 5 (ets domain transcription factor)
EHF	↓	sv	ev	11p12	ets homologous factor

Figure 4.2 Chromosomal location of *ESE*-2 (*ELF*5) on human chromosome 11 The diagram represents a detailed segment of chromosome 11 around the ESE-2 locus, showing selected genes only. Data were generated from the UCSC Human genome browser.


Human ESE-2b GTCTTTATTTCCACTGCACGCTGGTGCTGGGAGCGC-CT-GCCTTCT
Mouse Ese-2 CACGGCTACAGGTGCCTTTATTTCTACAGTCCGCTGGTGCTGGGAGCGCGCTTGCCTTCT
BovineEse-2 CATGGACACGACACCTGTGACTGCCGCCTCTGACTCGTG-TGAGGGCAAG-AAGCC--CA
Human CTTGCCTTGAAAGCCTC-CTCTTTGGACCTAGCCACCGCTGCCCT-CACGGTAATGTTGG
Mouse CTTGCCTTGAAAGCCTT-CTGTCTGGACCTAGCCACCACTTGTCTTCACGGTGATGTTGG +92
Bovine CAGGCC-TGTATCTCTGACTGTGTGGACCCTTCCGGTGCACG-CTTG--GGTGATGTTGG
Human ACTCGGTGACACACAGCACCTTCCTGCCTAATGCATCCTTCTGCGATCCCCTGATGTCGT
Mouse ACTCCGTAACCCATAGCACCTTCCTGCCCAACGCATCCTTCTGTGACCCCCTGATGCCTT +152
Bovine ACTCAGTGACACACAGCACCTTCCTGCCCAACACGTCCTTCTGCGACCCCCTGATGTCGT
Human GGACTGATCTGTTCAGCAATGAAGAGTACTACCCTGCCTTTGAGCATCAGACAGCCTGTG
Mouse GGACCGATCTGTTCAGCAATGAAGACTACTACCCTGCCTTTGAGCATCAGACAGCCTGTG +212
Bovine GGACTGACCTGTTCAGCAATGAAGAGTACTACCCTGCCTTTGAGCATCAGACAGCTTGCG
Human ACTCATACTGGACATCAGTCCACCCTGAATACTGGACTAAGCGCCATGTGTGGGAGTGGC
Mouse ATTCCTACTGGACATCAGTGCACCCTGAATACTGGACCAAGCGCCATGTCTGGGAATGGC +272
Bovine ACTCCTACTGGACATCCGTCCACCCCGAATACTGGACGAAGCGCCACGTCTGGGAATGGC
Human TCCAGTTCTGCTGCGACCAGTACAAGTTGGACACCAATTGCATCTCCTTCTGCAACTTCA
Mouse TCCAATTCTGCTGTGACCAGTACAAGCTTGATGCCAACTGCATCTCCTTCTGTCACTTCA +332
Bovine TCCAGTTCTGCTGTGACCAGTACAAGCTGGACGCCAACTGCATCTCTTTCTGCCATTTCA
Human ACATCAGTGGCCTGCAGCTGTGCAGCATGACACAGGAGGAGTTCGTCGAGGCAGCTGGCC
Mouse ACATCAGCGGCCTGCAGCTCTGCAGCATGACGCAGGAGGAGTTCATTGAGGCAGCCGGCA +392
Bovine ACATCAGTGGCCTGCAGCTGTGCGGCATGACACAGGAGGAGTTCATGGAGGCGGCCGGCG
Human TCTGCGGCGAGTACCTGTACTTCATCCTCCAGAACATCCGCACACAAGGTTACTCCTTTT
Mouse TCTGTGGGGAGTACCTGTACTTCATTCTCCAGAACATTCGCTCGCAAGGTTACTCCTTTT +452
Bovine TCTGTGGGGAGTATCTGTACTTTATCCTCCAGAGCATCCGCTCACAAGGTTACTCCTTTT
Human TTAATGACGC TGAAGAAGC-AAGGCCACCATCAAAGACTATGCTGATTCCAACTGCTTG
Mouse TCAATGATGCTGAAGAGACCAAGACTGGCATCAAAGACTATGCTGATTCCAGT TGCTTG +512
Bovine TTAATGATCCTGATGAGACCAAGGCCACCCTCAAAGACTATGCTGATTCCAGT TGCTTG
Human AAAACAAGTGGCATCAAAAGTCAAGACTGTCACAGTCATAGTAGAACAAGCCTCCAAAGT
Mouse AAAACAAGTGGCATCAAGAGTCAAGACTGTCACAGCC------GAACAAGCCTCCAAAGT +572
Bovine AAAACAAGTGGCATCAAAAGTCAAGACTGTCACAGTCATAGTCGAACAAGCCTCCAGAGT
Human TCTCATCTATGGGAATTTGTACGAGACCTGCTTCTATCTCCTGAAGAAAACTGTGGCATT
Mouse TCTCACCTGTGGGAATTTGTCAGAGACTTGCTGCTGTCCCCTGAAGAGAACTGTGGCATC +632
Bovine TCTCATCTATGGGAATTTGTGCGAGATCTGCTTCTATCTCCTGAGGAAAACTGCGGCATT
Human CTGGAATGGGAAGATAGGGAACAAGGAATTTTTCGGGTGGTTAAATCGGAAGCCCTGGCA
Mouse CTGGAATGGGAAGACAGGGAGCAGGGCATTTTCCGAGTGGTTAAGTCAGAAGCCCTGGCA +692
Bovine CTGGAATGGGAAGATAGGGAACAAGGTATTTTTCGGGTGGTTAAATCAGAAGCCCTGGCG
Human AAGATGTGGGGACAAAGGAAGAAAAATGACAGAATGACATATGAAAAGTTGAGCAGAGCC
Mouse AAGATGTGGGGACAAAGGAAGAAGAATGACAGGATGACGTACGAGAAGCTGAGCCGAGCC +752
Bovine AAGATGTGGGGACAAAGGAAGAAAAATGACAGAATGACGTACGAAAAGCTGAGCAGAGCT
Human CTGAGATACTACTATAAAACAGGAATTTTGGAGCGGGTTGACCGAAGGTTAGTGTAC
Mouse CTGAGATACTACTATAAAACGAGAATTCTGGAGCGGGTTGACCGGAGGTTAGTGTAC +809
Bovine CTGAGGTACTACTATAAAACCGGAATTTTGGAACGGGTTGACCGAAGATTAGTGTAC

Figure 4.3 Partial cDNA sequence comparison of *ESE*-2 from human, mouse and cow The coding nucleotide sequence of *Ese*-2 to 837bp is shown. The human sequence corresponds to the *ESE*-2b sequence (AF_115403) and mouse corresponds to *Ese*-2 (AF_049702). The bovine DNA was isolated during the initial investigation of *Ese*-2 (unpublished results), and the complete known sequence is shown. The ATG initiation and termination codons are in bold and the mouse transcription start site (bold and underlined) is numbered as +1. Underlined sequences in the main sequence body represent the Pointed domain (first section) and the ETS domain in the second underlined section. Mismatches are in blue type.

Figure 4.4 Bovine 5'UTR aligned with human *ESE*-2a 5'UTR

Sequences are shown only partially where the alignment occurs. The best-fit alignment including the aligned sequence shown is 78%. The major difference is in the length of the human exon, though the bovine may extend further 5'. Also, the location of the ATG is different. The ATG initiation codon is shown in bold. Mismatches are shown in blue. Human and bovine sequences are numbered with respect to the ATG codon at +1.

The region immediately 5' of the first exon of the mouse also showed high homology, possibly representing the proximal promoter. The first exon of the mouse is very similar to that of *ESE*-2b first exon, and also a small region 3' of the exon has homology. In the first intron of the mouse there are 3 regions that are similar to the human. One of these is a part of the *ESE*-2a first exon.

Some, but not all of the sequence similarities, span potential transcription factor binding sites as indicated in Table 4.3.

4.2.4 Promoter prediction

Computer based analysis of DNA sequence can be extended from simple consensus site transcription factor searches to sophisticated algorithms designed to predict transcription start sites and promoter regions. The basis for such programs varies, though a combination of transcription factor sites, locus distances and non-transcription factor sequence similarities are combined to generate a high level of information. Several such programs are available via the world wide web, including the Sanger Center Gene Finder program, which uses an algorithm based in combining characteristics describing functional motifs and oligonucleotide composition of these (Prestridge, 1995). Further programs exist such as 'Genscan" based on the prediction of exons, to locate the transcription start site (Burge and Karlin, 1997). Also, 'GRAIL' experimental gene discovery, which uses a neural net searching for specific elements such as TATA, CAAT and GC boxes (Hyatt *et al.*, 2000). A more sophisticated program was developed by the Berkeley Drosophila Genome Project called 'Genie" (Reese and Eeckman, 1995). Genie is based on generalised hidden Markov model principles. This program was used to predict transcription start sites of *Ese*-2, as outlined in Table 4.2. The search was limited by the amount of sequence entered, 2.6kb 5', exons 1 and 2, intron 1 determined from this investigation, and the score was

restricted to those promoters returning a value above 0.8, where 1.0 is a perfect match for a transcription start site.

One predicted site agrees with the experimental data, but most striking is the number of other sites that are predicted to exist. The probability of other start sites and promoter regions may be high, though their use in the biological system examined in this investigation is not supported by the experimental analysis of mRNA transcripts. Other promoters may be used infrequently giving rise to few alternate mRNA transcripts. These may not be detected due to their very low expression, but may be important in tissues where *Ese*-2 has not been found thus far. Within the first intron of *Ese*-2, the region of homology with exon 1a of the human *ESE*-2a gene is not detected as a putative promoter. However, the homology is not complete between the mouse and human in this sequence, which may account for the promoter being present in the human but not the mouse.

Table 4.2 Promoter prediction of the *Ese*-2 sequence

Start	End	Score	Promoter sequence and transcription start site (bold)
779	829	1.00	CAGGACACTTTTTAAACTCCCGCCTAAGGGCTGGAGAGATGGTGCTGTGG
1201	1251	1.00	ACAAAACATGTTTTAAAAATGCCTAGTGATACAGGGTCTCATTGTGTGGA
1269	1319	0.85	AACTCACTGTGTAAATCAACCTGGCCTCAAACCCACAGATATCTGCCGCC
1772	1822	0.88	TGAGAGATGTCCTAAAAACTGGGTAGTTCTCAGTGAGACAGCTGACAGCT
1895	1945	0.92	ATCGCCATTTCTAAAACAGAGGACACTGGGAAAGGGGGAAAGCTGTCACA
2622	2672	0.80	AGGTGTCACTTATATCACACGGCTACAGGTGCCTTTATTTCTACGGTCCG
4209	4259	0.99	AGCCACCCTTCTTAAAGAGCCCGAAGGGTGGGAAAGTAACACCCCCTATC
4807	4857	0.93	TCAGAAAGGAGATAAAAAGTTGCTGGAGGCTCTCAGTTCCAGGCCTGCCC
4889	4939	0.89	TCTCACCAGTAAAAAAAAGAGCTCATTTGATTCCACCACCTTAGGGTGAT
5048	5098	0.84	CATGGGAGTATAGGATCCCAGCTCAGTTGTCCAGCTGAGCAGGAGAGGGA
5900	5950	0.99	TCACTGCTTTCATAAAAGACCTCCCCACGCCTCCCCCCCCAACCCCCACC
6021	6071	0.87	CTAATGGGTATAATTTGCCATAAAAGGAATTATCGGAGGGAACCAAATCC
6298	6348	0.86	CCCATAGACTTTAAAGACCTGGAGACTGCATTCTGTTCTCTTCCATTTGG
6822	6872	0.99	AGTTTCCTTATATAAATGTGCCTGGGATGATGCTATACCCTTCAGATAGG

4.2.5 Expressed Sequence Tag information

Expressed sequence tags (ESTs) provide a useful tool for searching for anomalies within genes. They give sequence information directly from cDNA sequence, representing the 'coding' portion of the genome. Sequence data from ESTs is collated with the human genome draft sequence (and other genome sequence data) and also separately in a database of ESTs alone (TIGR). From the 'UCSC genome browser' internet tool, the set of ESTs aligning to a specific gene can be observed. In this study, human chromosome 11p13 at position 36972252-37006397 was displayed and the ESTs aligned.
An alignment of the available murine ESTs at the 5' end of *Ese*-2 predicts a start position between nucleotide -11 and +12 with respect to the +1 site determined in this study. This correlates well with the predicted major transcription start site identified by primer extension. It also reinforces the likelihood of multiple start sites as seen experimentally. The discrepancy of approximately 20 nucleotides for the exact start site may be accounted for by the inaccuracy of sequencing near the 5' end of EST clones.
When the full set of available ESTs were aligned to the *ESE*-2 gene, some anomalies could be observed (Figure 4.5).
Firstly, there appears to be other splicing events occurring in the human gene after exon 3 and exon 5 (Human ESTs that have been spliced). Such splicing events appear to be a common feature of the human genome, as discussed later. However, in this instance the splicing is represented by a small number of EST sequences only and their importance should not be overemphasised.
Secondly, exon 4 is not present in 21 of 51 mouse EST clones that would be expected to encompass this exon. The high percentage from a large number of clones is likely to represent a real feature of the gene. Skipping of exon 4 has not been reported in published reports of the gene.

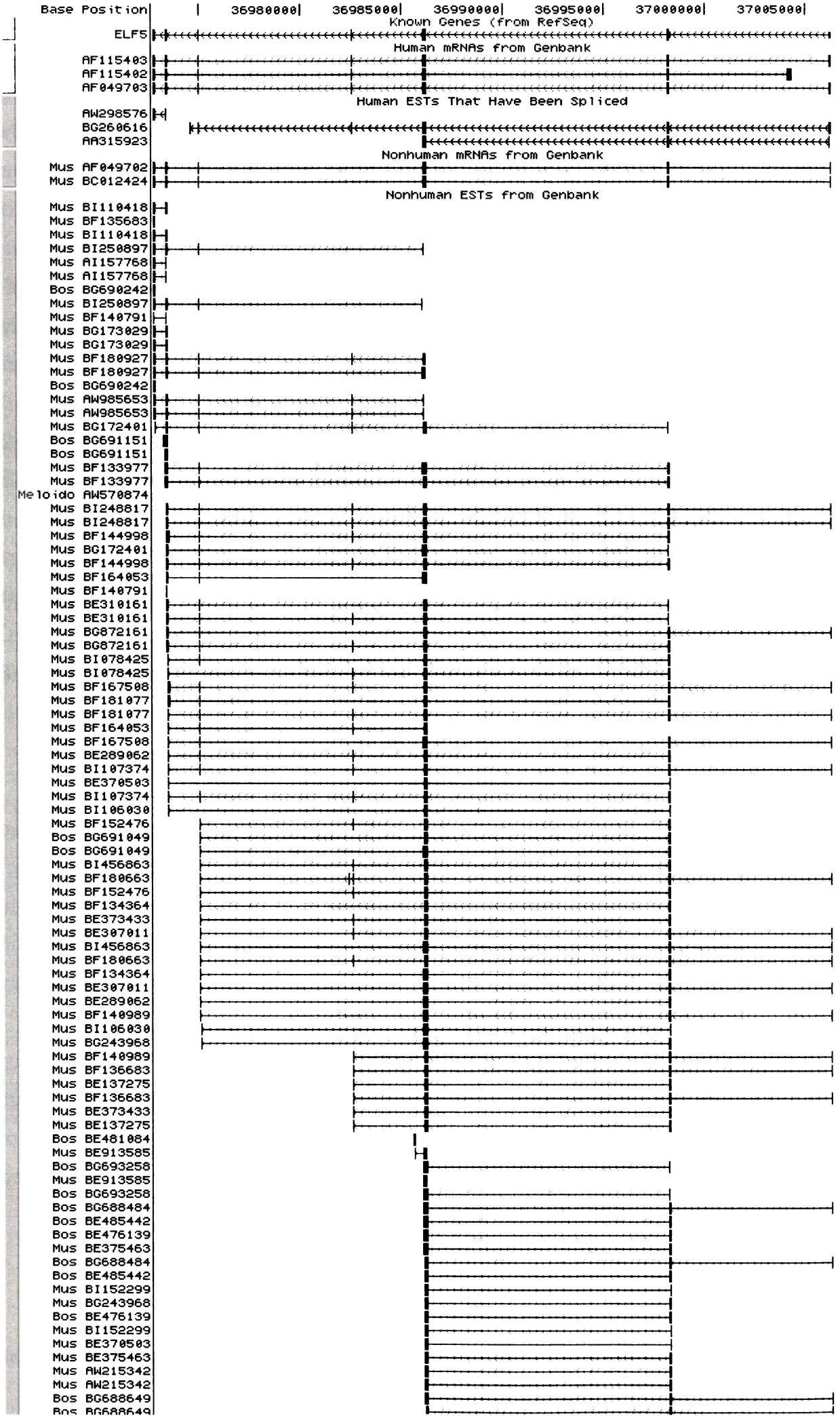

Base Position
36980000
36985000
36990000
36995000
37000000
37005000
Known Genes (from RefSeq)
ELF5
Human mRNAs from Genbank
AF115403
AF115402
AF049703
Human ESTs That Have Been Spliced
AW298576
BG260616
AA315923
Nonhuman mRNAs from Genbank
Mus AF049702
Mus BC012424
Nonhuman ESTs from Genbank
Mus BI110418
Mus BF135683
Mus BI110418
Mus BI250897
Mus AI157768
Mus AI157768
Bos BG690242
Mus BI250897
Mus BF140791
Mus BG173029
Mus BG173029
Mus BF180927
Mus BF180927
Bos BG690242
Mus AW985653
Mus AW985653
Mus BG172401
Bos BG691151
Bos BG691151
Mus BF133977
Mus BF133977
Meloido AW570874
Mus BI248817
Mus BI248817
Mus BF144998
Mus BG172401
Mus BF144998
Mus BF164053
Mus BF140791
Mus BE310161
Mus BE310161
Mus BG872161
Mus BG872161
Mus BI078425
Mus BI078425
Mus BF167508
Mus BF181077
Mus BF181077
Mus BF164053
Mus BF167508
Mus BE289062
Mus BI107374
Mus BE370503
Mus BI107374
Mus BI106030
Mus BF152476
Bos BG691049
Bos BG691049
Mus BI456863
Mus BF180663
Mus BF152476
Mus BF134364
Mus BE373433
Mus BE307011
Mus BI456863
Mus BF180663
Mus BF134364
Mus BE307011
Mus BE289062
Mus BF140989
Mus BI106030
Mus BG243968
Mus BF140989
Mus BF136683
Mus BE137275
Mus BF136683
Mus BE373433
Mus BE137275
Bos BE481084
Mus BE913585
Bos BG693258
Mus BE913585
Bos BG693258
Bos BG688484
Bos BE485442
Bos BE476139
Mus BE375463
Bos BG688484
Bos BE485442
Mus BI152299
Mus BG243968
Bos BE476139
Mus BI152299
Mus BE370503
Mus BE375463
Mus AW215342
Mus AW215342
Bos BG688649
Bos BG688649

Figure 4.5 UCSC Genome Browser from the August 6, 2001 freeze on chromosome 11, position 36972252-37006397
Selected tracks are shown, labelled in the centre of the diagram above the sequence. The Genbank accession numbers of sequences are listed to the left of the figure. The gene sequence runs from right to left as it is on the negative chromosome 11 strand. Band position refers to the nucleotide numbering on chromosome 11. Exon skipping can be seen for exon 4 under the 'Nonhuman ESTs from Genbank', where it is present and absent in some sequences. Two EST sequences appear to have an exon skipping event for exon 5 also.

4.2.6 Blast Searches against *Ese-2*

Following the comparison of the *ESE*-2 human and mouse sequences, a genome wide Blast search was performed to identify any regions of similarity with genes other than *ESE*-2, and with other organisms.

Most obvious was the homology to other *ets* genes, due to the ETS domain being so highly conserved. The ETS domain of ESE-2 is highly similar to ESE-1 at the amino acid level (67%), and 65% similar to that of ESE-3. ESE-1 and ESE-3 retain 84% identity (Kas *et al.*, 2000). ESE-2 has an ETS domain that retains only moderate amino acid homology with other family members. The highest is 49% to human NERF, 48% to Drosophila ETS4 and E74A, and 46% to human ELF-1 (Zhou *et al.*, 1998).

Outside of the ETS domain, some other genes showed apparent homology. In the 5' region of *Ese*-2 a sequence of 64bp matched with 90% homology to the cDNA sequence of the mouse *dystroglycan* 1 (*DAG*1) gene (Figure 4.6). The same region was also homologous (90%) to the mouse upstream region of the promoter of the *uroplasminogen activator* (*uPA*) gene. *DAG*1 is a laminin-binding component that connects the subsarcolemmal cytoskeleton with the extracellular matrix. *DAG*1 is a candidate gene for autosomal recessive muscular dystrophy, as its repression results in a loss of the connection between extracellular matrix and sarcolemma making muscle fibers susceptible to necrosis (Henry *et al.*, 2001). The role of many *ets* genes and possibly *Ese*-2 in extracellular matrix regulation makes the sequence homology of interest. The sequence may contain sites for common factors to bind that are involved in a common function for both genes.

Similarly, *uPA* is involved in extracellular proteolysis, important in tissue remodelling and cell migration, which is important for mammary gland involution (Degen *et al.*, 1987). The potential role of *Ese*-2 here again implies some common role of the homologous sequence.

Table 4.3 Genomic DNA alignment of human and mouse for the 5' promoter and intron 1. Each row indicates a region of homology. The sequence given is that for the mouse, with any transcription factor binding sites in bold, and named in the last column, in order. Mouse sequence numbering is relative to the transcription start site at +1. Human numbering is relative to the *ESE*-2a first exon at +1, with *ESE*-2b first exon at –1982 to –1878.

Ese-2/ ESE-2	Start site	Mouse sequence of conserved region	End site	% ID	Binding site
Mouse	-2484	TCTTTT**GTTTTTGTTTTATAT**TAATCATCCCTTTTAACGACT GGTCCAGAAGCCT	-2430	87%	Elf-1
Human	-4026		-3972		
Mouse	-1159	ATTTTTCCAT**GATGACACATGAC**TCCTCACATGAACTGCCTT TTGTTTCTTCACGCCGTTCGGGTGTTTGGATTTGAGTGTTTG TGCCTATGCCCAGTGTTCCCAAGA**AAAACCTAACCAACCA**GTC	-1033	81%	Ap1 C/EBP
Human	-2899		-2773		
Mouse	-177	GAGCTGGGCACAAAAGCAGGAGAAAGGTA**AACTTTCTGCATG TGAAAACCA**	-127	92%	C/EBPβ
Human	-2142		-2092		
Mouse	-113	AGGA**GCCGTGTCACA**CCGTAT**GTCACCGTCATC**AAAGGGGCT GTGCATAAACCTGAAAAACCAAACGGACCTGTCTGT**AGGTGT CACTTATATCACACGG**CTACAGGTGCCTTTATTTCTACGGTC CGCTGGTGCTGGGAGCGCGCTTGCCTTCTCTTGCCTTGAAAG	+76	92%	Ap1, Ap1, Ap1,
Human	-2077	CCTTCTGTCTGGACCTAGCCACC	-1889		GATA1
Mouse	+97	CTTTTCTTTGCTCTTTTGAATGAATTGAAAGGGGTGTCGGGG CTCAGAG	+145	89%	n/a
Human	-1869		-1821		
Mouse	+2000	GAATTTTCCTCCTCTGAC**AAATTGGGCAAGGCTGG** GTAGTTTGC	+2043	88%	C/EBPβ
Human	-29		+14		
Mouse	+3137	CACACCTAGTGCATGCTGGCTAAGTAGCCTGGCCTTTTGTGG TGAGCTTAGACTTAATAAC	+3198	87%	n/a
Human	+1390		+1450		
Mouse	+3359	AAACCTAATGGGTATAATTT**GCCATAAAAGG**AATTATCGGAG GGAACCAAATCCATCATTGTGTGTAGCGCTGAAGAGATTAAA GATCAAG	+3449	82%	TATA
Human	+1586		+1676		

Ese-2	TTGCTGCTCTTGCAGAGGACCCAGGTTCAATTCCCAGCACCCACGTGGCAGCTCACAACC -1548
DAG1	TTGCTGCTCTTCCAGAGGACCCAGGTTCAATTCTCAGCACCCACATGGTGGCTCACAACT +838
uPA	TCTTCCAGAGAAGCCAGGTTCAATTCCCAGCACCCACATGGCAGCTCACAACT -487

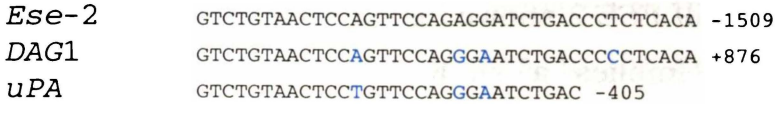


Figure 4.6 Sequence alignments between *Ese-2*, *dystroglycan 1* (*DAG1*) and *urokinase plasminogen activator* (*uPA*) *DAG1* is homologous from positions 778 to 876 of the cDNA. *uPA* is homologous in the 5' promoter region, 487 to 405. Mismatches are shown in blue relative to the mouse sequence, where +1 is the major transcription start site.

4.3 Discussion

This chapter presents analyses based on current genome databases. Sequences which had been submitted to GenBank already spanned the cDNA for *ESE*-2, though genomic sequence was unavailable. The sequencing of the genome has amended this and now a complete set of data is available, albeit draft only. For the purposes of this research, the human genome data is very useful. In particular it facilitated a comparison of the known mouse genomic sequence in the 5' and first intron, which were studied here in detail.

The coding sequence of *ESE*-2 showed good homology with murine *Ese*-2 (87%), which implies a similar function is maintained between species. Interestingly, both the mouse and human gene is expressed as two distinct isoforms. For the mouse, these differ in the use of polyadenylation signals, making a difference of 1kb mRNA size. In the human, alternative 5' splicing results in a difference of 10 amino acids seen as an extension in *ESE*-2a over *ESE*-2b. The whole of the first exon of *ESE*-2b is different from that of *ESE*-2a suggesting separate promoter usage (Oettgen *et al.*, 1999b). However, it must be recognised that this investigation dealt with cDNA obtained from murine mammary and that of the human *ESE*-2 investigation was from prostate. The tissue of origin of the sequence may play a role in the isoforms that were detected. Indeed, possibly two isoforms equivalent to the human were detected by Northern analysis in the Comma1D cell line. These may be at very low levels and expressed at higher levels in tissues not examined here. The EST sequences referred to in the database were mostly generated from normal or tumourigenic mammary gland.

When these sequences were compared to the known bovine sequence, it was discovered that the 5'UTR of bovine was very similar to the *ESE*-2a 5'UTR. In contrast, the mouse 5'UTR is similar to that of *ESE*-2b. Also, a section (43bp) of the *ESE*-2a exon

was found to have homology to sequence within the first intron of the mouse gene. This raises the possibility of an alternate promoter and first exon in the mouse that has not yet been detected. It is not known whether the bovine *Ese*-2 has two isoforms, as Northern data is not available.

The presence of the two isoforms in both human and mouse supports the hypothesis that the isoforms have slightly divergent functions. Indeed, in both the mouse and human, a slightly different pattern of expression is observed for each isoform, supporting the hypothesis mentioned earlier. For example, *Ese*-2 (2.5kb) is expressed in the mouse in neonatal and embryonic lung and kidney, and adult ovary, compared to the *Ese*-2 (1.5kb), which is not expressed in these tissues. In the human, the kidney was found to express only *ESE*-2a, whereas the prostate expressed both, but more *ESE*-2b than *ESE*-2a (Oettgen *et al.*, 1999b).

The sequence of the *ESE*-2 gene highlights its homology to the *ets* family of transcription factors through the ETS domain. This implies *Ese*-2 is a transcription factor, most closely associated with *ESE*-1 and *ESE*-3. ESE-1,2, and 3 are all epithelially expressed genes. The subgroup is phylogenetically grouped as discussed in Chapter 1. A predicted function of *ESE*-2 is an involvement in cell proliferation and migration. Evidence for this is from its homology with *ESE*-1, another gene showing epithelial restricted expression, known to be differentially regulated during differentiation of some epithelial cells (Andreoli *et al.*, 1997).

The second domain that retains homology to the Ets transcription factors is the Pointed domain, also found in ESE-1 and ESE-3 and some other Ets proteins. It is very similar to the SAM domain, which is involved in homo- and hetero-dimerisation. It is therefore a possibility that the role of the Pointed domain is similar in the *ets* genes.

The chromosomal location of *ESE*-2 (11p13-15) has been reported to frequently undergo Loss of Heterozygosity (LOH) in certain cancers (Reeve *et al.*, 1989; Weinberg, 1991; Fong *et al.*, 1994;

Shipman *et al.*, 1998). Studies of this region have focused on Wilms Tumor patients. The specific region of LOH in Wilms Tumor has been mapped to chromosome 11p15 loci, and is thought to extend to include the Wilms Tumor, aniridia, genitourinary malformations and mental retardation (WAGR) locus at 11p13 (Reeve *et al.*, 1989). Indeed a survey of 48 samples of primary non-small cell lung cancer (NSCLC) allelic deletions on chromosome 11 were localised to 11p13 catalase (*CAT1*) and beta FSH (*FSHB*), and the other at band 11p15 corresponding to the *IGF II* locus (Ludwig *et al.*, 1991).

With *ESE*-2 and *ESE*-3 both located in close proximity to catalase and the region of LOH, it may be that their absence is also significant in the cancers resulting from LOH. *ESE-3* is a known transcription repressor of some Ets/Ap-1 responsive genes such as interstitial collagenase gene (*MMP*-1) and it is absent from epithelial carcinomas of tissues where it is normally expressed (Tugores *et al.*, 2001). *MMP*-1 is a matrix metalloproteinase gene involved in the degradation of the extracellular matrix in processes such as mammary gland involution. Its abnormal regulation has been implicated in tumor formation (Bolon *et al.*, 1995). Therefore, the loss of an *ets* gene such as *ESE*-2 or *ESE*-3 may at least partially be responsible for the loss of regulation and hence tumor formation.

Also within the syntenic block of genes of chromosome 11 (and chromosome 2 of the mouse) is the Paired Box Homeotic Gene-6 (*PAX6*). Mutations to this gene have been implicated as the genetic cause of aniridia. Some of these mutations also encompass chromosome deletions of the *WT1* locus and increase the risk of Wilms Tumor (Gronskov *et al.*, 2001). It appears that many genes in this chromosomal region play a role in tumor and disease formation through LOH or other mutations.

The mouse EST alignment analysis revealed an exon skipping event that appears to occur in nearly 50% of all the ESTs published in this region. ESTs for *Ese*-2 are seen to often lack exon

4. Also, the published *Ese*-2 sequence is also lacking exon 4 when aligned by the UCSC genome browser.

Alternative transcription is predicted to occur in at least 30-40% of human genes. This includes alternative splicing, polyadenylation and transcription start sites, based on EST studies (Hide *et al.*, 2001). Of the features responsible for altering the protein coding sequence of a gene, exon skipping is the most frequent (Mironov *et al.*, 1999). The use of such alternative transcripts has been confirmed in developmental, tissue-specific and disease-specific processes (Zacharias *et al.*, 1995; Dufour *et al.*, 1998; Jiang and Wu, 1999; Lambert de Rouvroit *et al.*, 1999; Lim *et al.*, 1999; Unsworth *et al.*, 1999; Kawahara *et al.*, 2000; Mercatante and Kote, 2000; Strehler and Zacharias, 2001). Such diversity in the processing of transcripts can enhance the diversity of protein products available within the genome. This is particularly important as the exon skipping events often have a significant impact on the protein. This has been demonstrated for many genetic disorders and tumor formations (Birrell *et al.*, 2001; Manson and Huxley, 2001; Sironi *et al.*, 2001).

The functional implications of exon skipping of exon 4 in *Ese*-2 have not been studied. However, the loss of exon 4 is predicted to result in a frame shift, likely to cause major changes in the protein product.

The mRNA sequence alignment for mouse, human and bovine revealed a high degree of similarity. Although the bovine mRNA may not be complete, the coding region is of a similar length to both human and mouse. The mouse and human differ widely in the 3' UTR and the two mouse isoforms also differ widely in this region as mentioned earlier. Genomic sequence alignments revealed several interesting homologous regions. With respect to the human alternate first exons, *ESE*-2b exon 1 was very similar to that of the mouse. *ESE*-2a first exon showed some homology to approximately 40bp of the mouse first intron at 2kb. The 5' region of mouse and human had several regions of similarity. In the more

distal 5' region, an alignment of 50bp spanned an *Elf*-1 consensus binding site, and a larger alignment of 120bp encompassed conserved Ap1 and C/EBP binding sites. These conserved regions may be more important than unconserved regions in the regulation of transcription of the gene.

Of particular note was a homologous region in the proximal 5' promoter of the *ESE*-2b gene. The proximal promoter alignment spanned several Ap1 transcription factor consensus binding sites, as well as a potential GATA-1 site, which suggest their importance as a result of their conservation. Several short alignments throughout the intron of the mouse were also located though these didn't appear to contain many potential transcription factor binding sites.

A 5' upstream region of the mouse sequence also showed high homology to both the *uPA* and *DAG*1 genes. As explained earlier, the area of homology was to the coding region of *DAG*1 and the upstream promoter of *uPA*. Both these genes, along with *ESE*-2 may have some common ancestry, also suggested by their similar predicted functions. The possible functional relationship between these genes is striking. Plasminogen activators are involved in extracellular proteolysis such as tissue remodelling and cell migration during processes such as mammary gland involution and tumor cell metastasis (Degen *et al.*, 1987). Dystroglycan is expressed in cell types directly adjacent to basement membranes (including epithelia) and its function is in the formation of basement membranes as an attachment between extracellular matrix and cytoskeleton (Henry *et al.*, 2001). Abnormalities in both structure and function of basement membranes are involved in metastatic disease, thus it is likely that *DAG*1 will also play a role here. Reduced expression of *DAG*1 has been observed in both prostate and breast tumours, suggesting this is a feature of epithelial neoplasia, leading to abnormal cell and extracellular matrix interactions (Henry *et al.*, 2001).

ESE-2 is an epithelial expressed gene, and information to date implicates it having a role in the regulation of the extracellular matrix degradation. These functional similarities seem to suggest that the sequence homology may be of significance. For example, it may harbour sequences to direct expression to epithelia, extracellular matrix or their regulators.

Functional analysis of these regions of homology is required to ascertain their biological significance. The use of deletion constructs in cell culture may give valuable evidence to the regulation of *Ese-2*.

Chapter 5. Investigation of *Ese-2* sequences for promoter activity *in vitro*

5.1 Introduction

Usually, a promoter is a region of DNA that contains all the response elements required for the complete transcriptional regulation of a gene. In practice the promoter requires additional, non-contiguous regions of DNA, encompassing elements such as enhancers for its correct function.

In the first instance, the promoter is often referred to as the proximal promoter, which is an arbitrary amount of sequence immediately 5' of the first exon, usually extending 300-800bp 5' of the transcription initiation site. It is normally able to drive a basal level of transcription of the gene, without the need for additional elements such as enhancers.

Theoretical characteristics of the promoter region of *Ese-2* have already been discussed in Chapter 3 and 4. Multiple programs exist that locate the most probable promoter and transcription start sites based on algorithms of promoter elements and their locations from known sequences (Prestridge, 1995; Reese and Eeckman, 1995; Burge and Karlin, 1997; Hyatt *et al.*, 2000). With respect to the activity of the promoter, functional predictions based on such computer analyses, potential sites of interest or comparisons to other related genes are not at all reliable. In order to gain insight into the functional regulation of *Ese-2*, the next step was to test the activity in a cell culture system. Ideally, the use of an *in vivo* model such as a mouse with inserted transgene constructs would give the best functional representation. However, with time restraints and other impracticalities, the *in vitro* system is often the next choice for examining promoter activity. Once some initial regions and factors have been pinpointed from such an *in*

vitro analysis, then a mouse model could be used to confirm the activity.

In this study, a model system was set up using two cell lines derived from the mouse mid-pregnant mammary gland. One of these (Comma1D) expresses the endogenous *Ese*-2 gene *in vitro*, while the other (HC11) does not (Figure 5.1) (see also Langley, 1999). The Northern blot shown in Figure 5.1 highlights the existence of endogenous *Ese*-2 mRNA transcripts in the Comma1D cell line, as well as in mouse mammary gland tissue (17 day pregnant). It is notably absent from the HC11 cell line (at least at this level of sensitivity). Both the 1.5 and 2.5kb bands are detected, though the smaller band is much weaker. The discrepancy in size between the Comma1D and mammary gland band may be due to the presence of salt in some samples. There may also be two bands present around 2.5kb, which may correspond to 2.4 and 2.6kb as seen in the human form of *ESE*-2 (Oettgen *et al.*, 1999b). This is discussed in more detail at the end of the chapter.

For the purposes of this study, approximately 850bp of DNA 5' to the major transcription start site is referred to as the proximal promoter and was studied first. This length of sequence is purely arbitrary, allowing for a convenient location for deletion constructs to be designed based on restriction sites and potential binding sites of interest.

Following the proximal promoter study, enhancer regions were searched for. Potential sites for enhancers are often further 5' and in some cases in the first intron. Both of these possibilities were examined.

In keeping with the proximal promoter and enhancer theory, *Elk*-1 (an Ets transcription factor) was recently identified as having good proximal activity *in vitro* within -43/+63bp of the major initiation site, showing a 10 fold increase in activity over a vector only control (Lehmann *et al.*, 1999). This activity increased over 2-fold when the promoter construct was extended to 480bp. The addition of 700bp of 5' promoter sequence to the first intron (679bp) showed

a further 2-fold increase in activity. This represents an overall increase of 40 fold above the vector alone. Elements in the *Elk*-1 promoter include a TATA box at –25, which corresponds to a region of DNase 1 hypersensitive chromatin. A second DNase 1 hypersensitive region was isolated in the first intron, which may be the location of important factors responsible for increasing the *Elk*-1 *in vitro* activity.

Another example is that of *ETS*-2, another *ets* transcription factor member. Its promoter activity was studied in HeLa cells using a series of deletion constructs (Mavrothalassitis *et al.*, 1990). The 5' boundary for the proximal promoter and maximal activity was found to be 159bp from the major start site. Sequences further upstream between 890-1460bp were found to be dispensable for *ETS*-2 promoter activity. Multiple elements exist between 1460-1599bp, which may play a role in the regulation, seen as a loss in activity when this region is removed.

This chapter describes studies of first, the *Ese*-2 proximal promoter, second, the 5' region, which may harbour enhancers and finally the first intron. Many problems were encountered with all stages of this analysis, from the cloning to the transfection. Finally the analysis of these results will be discussed.

In addition to these transient assays, a stable transfection was performed and results from that experiment are detailed. In an attempt to increase or modulate promoter activity in some way, the effect of the hormones β-estradiol and progesterone were examined. This was analysed in relation to the expression of chloramphenicol acetyltransferase (CAT) by several of the *Ese*-2 promoter constructs. These hormones were chosen as they have a similar expression pattern to *Ese*-2 in the developing mouse mammary gland. These results are reported, along with a serum induction experiment at the end of the chapter. Finally to supplement the results given in the chapter, a DNase hypersensitive assay is discussed with some preliminary results that support an enhancer region in the first intron.

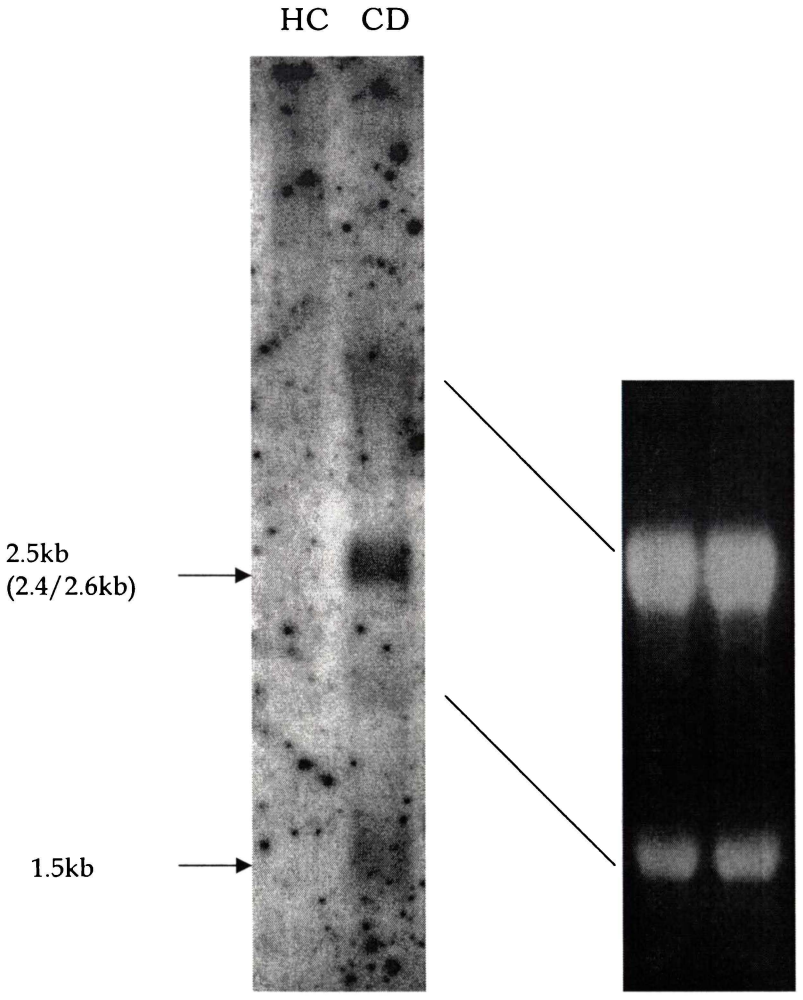


Figure 5.1 Expression of *Ese*-2 in HC11 and Comma1D cells
Northern blot analysis of RNA from HC11 (HC) and Comma1D (CD) cells. The 2.5 and 1.5 kb *Ese*-2 fragments are marked to the left of the diagram. The ethidium bromide stained gel showing the 18 and 28s rRNA is to the right (in the same order) of the Northern blot (not to scale).

5.2 Results

5.2.1 *Ese*-2 proximal promoter analysis

(a) Construct design

In order to examine whether or not the *Ese*-2 proximal promoter of approximately 850bp contains a functional promoter, constructs of the gene were tested in a cell culture system. Five deletion constructs of this region were generated to ascertain the minimal sequence necessary for transcriptional activity. The promoter fragments were inserted upstream of the chloramphenicol acetyltransferase-coding region (CAT) in a modified pBLCAT2 (Promega) vector. pBLCAT2 normally contains a minimal thymidine kinase (tk) promoter, which was replaced in the construction of the deletion plasmids by *Ese*-2 promoter sequence. The vector also contains ampicillin resistance and the SV40 polyadenylation sequence.

Deletion fragments were based on the presence of suitable restriction sites as well as the location of potential transcription factor binding sites. The 3' end of each fragment was a *Bss*HII site, which falls midway through the first untranslated exon, 19bp after the major transcription start site. To simplify cloning, the *Bss*HII site and another identical site further 5' were cleaved and the resulting 2kb fragment cloned into pBluescript II KS (Statagene). This allowed an *Xho*I digest to free the 3' end such that it would make a compatible ligation with the pBLCAT2 vector, when also cut with *Xho*I.

The longest fragment included in this analysis was about 850bp with a *Pvu*II restriction site at the 5' end, leaving a blunt end (construct 5) (Figure 5.2). It was excised from the pBluescript II KS vector using *Xho*I and *Pvu*II and inserted into a pBLCAT2 vector that had been previously digested with *Xba*I, Klenow blunt ended and *Xho*I digested. The digestion of the pBLCAT2 vector results in

the removal of the minimal tk promoter sequence. Further deletions were made with *Bam*HI at approximately 500bp (construct 4) and *Eco*RI at 200bp (construct 3). These fragments were released in a similar fashion using *Bam*HI-*Xho*I and *Eco*RI-*Xho*I respectively. They were then cloned into the pBLCAT2 vector that had been digested with the same enzyme pairs. Two short deletions were also included subsequently to identify a non-active promoter. These were made by PCR amplification. Firstly a 76bp construct was generated by PCR amplification from a 5' primer 'Forward +1' (sequence given in Chapter 2) that anneals at –50bp in relation to +1, and 3' sequence primer (5' Seq.) that anneals at position +1. This was subcloned into the pGEM-TEasy (Promega) vector and subsequently into the pBLCAT2 expression vector (construct 2). Later a 39bp PCR extension product was amplified from primers 1.11(1) and 5' sequence (encompassing –22bp to +17bp) and excluded the first GATA-1 site. Again this PCR fragment was cloned into pGEM-TEasy initially and then pBLCAT2 (construct 1). The constructs are depicted in Figure 5.2.

Finally, a negative control plasmid was constructed by digesting the pBLCAT2 vector with *Bgl*II and *Bam*HI, and religating these compatible restriction sites. This deleted the minimal tk promoter and hence is an empty expression vector containing no promoter sequence (construct 0).

All constructs were confirmed after cloning to be correct by full length sequencing across both ligation boundaries.

The potential transcription factor binding sites were discussed in Chapter 3. For the proximal promoter, the longest construct (5) encompasses the majority of the sites shown in Figure 3.7. These include YY1, GATA-1, NFKB, STAT, Ets, Ap-1 and others. As the constructs are shortened, more sites are lost so that construct (4) does not contain the YY1 or NFKB sites. For construct (3), the sites included in the sequence are Ets, C/EBPβ, GATA-1, a single Sp1-GC box, and multiple Ap-1 elements. Construct (2) only contains the first Ap-1 and GATA-1 sites.

(b) Transfection of the constructs

To obtain the best *in vitro* representation of the *in vivo* system, several cell lines were used. First, Comma1D, a mouse mammary gland cell line from a mid-pregnant animal was studied (Danielson *et al.*, 1984). Comma1D was found to express *Ese-2* mRNA, and was therefore considered a good model system. Second, a derived cell line from Comma1D, called HC11 was tested (Ball *et al.*, 1988). HC11 was isolated from the Comma1D mouse mammary epithelial line as a single cell clone that had high transfection efficiency and retained the ability to be induced by the lactogenic hormones. HC11 had the added advantage of growing well and being induced on a plastic surface (Ball *et al.*, 1988). Comma1D also appears as a mixed population of two cell types when viewed under magnification. This was the best cell culture model available, though these lines do not produce a full spectra of milk proteins. Analysis by Langley (1999) and in this study found HC11 did not express endogenous *Ese*-2 (Figure 5.1). The reason for this is unknown, though may be due to the lack of a transcription factor or a sequence mutation in the promoter or other critical location, or aberrant RNA processing for example. HC11 cells replicate more quickly than Comma1D cells, show higher transfection efficiency and appear more robust under cell culture conditions. The differences between HC11 and Comma1D did not appear to affect the major activities of the constructs, though some differences were seen and are discussed later. Together, these two lines provided a suitable model system for examining the *in vitro* activity of the *Ese*-2 promoter constructs. Cell lines were used with passage number below 30 to ensure the characteristics of the cell line were unaltered. Further to these lines, a human breast tumor cell line was also used, T47D (Keydar *et al.*, 1979). However, transfection of this line using either Fugene6 (Roche) or Lipofectamine 2000 (Invitrogen) failed to give detectable levels of transfection and this experiment was curtailed.

The deletion constructs were co-transfected with a control reporter plasmid pCH110 (Promega), which allowed measurement of β-galactosidase activity that could be used to normalise differing transfection efficiencies, explained further in section 5.2.1c. Initially, transient transfections were performed using Lipofectamine 2000 (LF2000)(Invitrogen). The exact protocol is given in Chapter 2. LF2000 is a mixture of lipids and other components that are complexed to the DNA to be transfected. The cells are more susceptible to taking DNA up when it is packaged into highly condensed particles. Moderate success was achieved with this transfection system, though better results were acheived using Fugene6 (Roche). LF2000 often resulted in cytotoxicity, killing up to 80% of plated cells (data not shown). The method was successful in the tansfection of constructs for stable integration, as detailed in section 5.2.10.

Fugene6 is another proprietary blend of lipids and other components that also increases transfection efficiency. The procedure of transfection is very similar to that of LF2000 as described in Chapter 2. DNA and Fugene6 ratios were optimised in several preliminary experiments to obtain the consistently highest level of transfection. No signs of cytotoxicity were seen with Fugene6, however the efficiency of transfection did vary between experiments and in some cases single results or whole experiments were discarded when variance was great (data not shown).

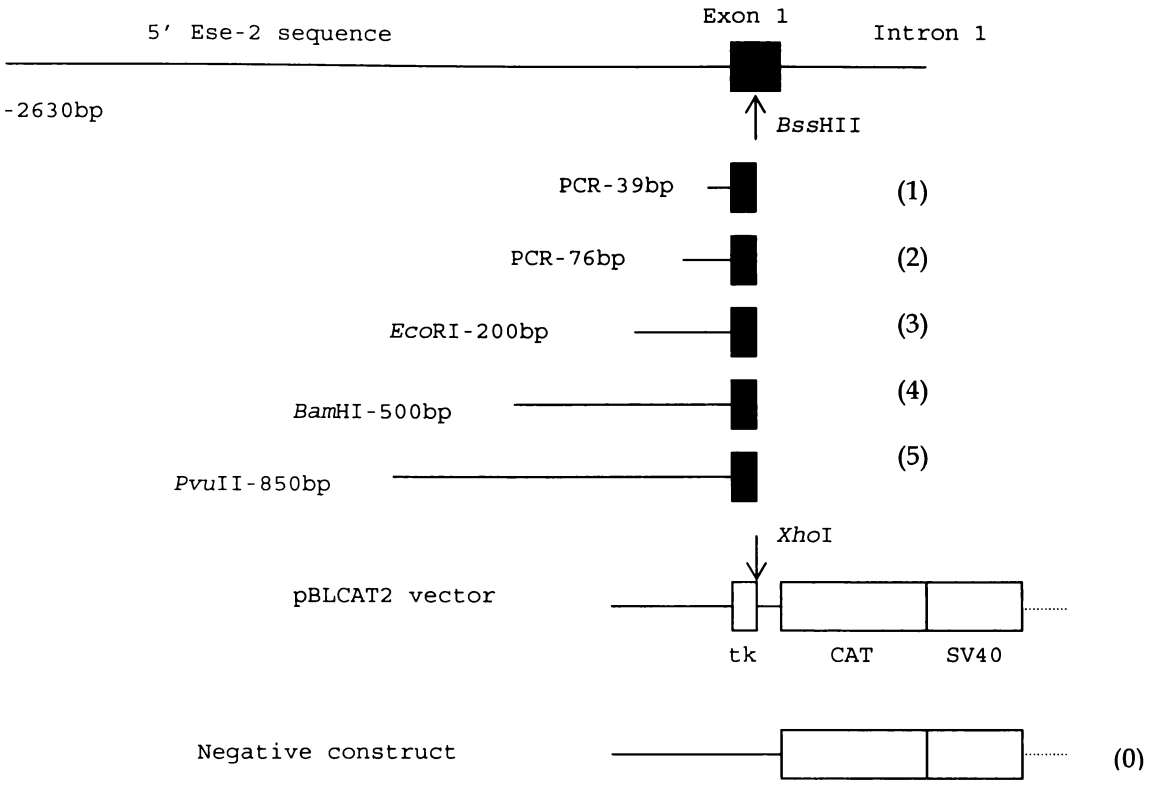


Figure 5.2 Diagram of deletion constructs (0),(1),(2),(3),(4) and (5) The *Ese*-2 5' region is represented in the line above the constructs. Exon 1 and the *Bss*HII restriction site are marked. The constructs extend in the 5' direction and include 30bp of exon 1 (untranslated), and are not drawn to scale. The pBLCAT2 vector is outlined below the constructs, showing the tk promoter, which is replaced in the constructs at the *Xho*I site and the CAT and SV40 polyadenylation coding sequences. The negative construct (0) is shown below, with no tk promoter sequence.

(c) Control and CAT assays

Transiently transfected cells were harvested after 48 hours in a lysis buffer compatible with the type of assay to be performed. CAT was measured either by the ELISA (Roche) protocol or the traditional ^{14}C thin layer chromatography method (TLC). Initially the ELISA method was used, but after multiple repeat experiments results were characterised as having low activity, and variances between duplicates and experiments were seen to be high. To ascertain if this was an artefact of the assay system and not a real result, the thin layer chromatography method was performed. This method allowed the visual detection of activity and was found to give more reproducible results between experiments and within duplicates. Once this method had been established, it was used for the remainder of the experiments. At a later time, the ELISA protocol was repeated and found to give similar data to that of TLC, suggesting the earlier unreproducible results were an artefact of the experimental procedure.

Once the cells had been extracted into lysis buffer, control assays were performed. A BCA protein determination assay (Pierce) allowed the protein content per well to be standardised, so that any large deviances would be corrected for or with extreme cases discarded, and hence each well would represent the same amount of protein (and cells). This assay usually resulted in similar protein levels indicating consistent cell death or other disturbance in cell growth. There were no trends for any construct to have consistently higher protein levels. This indicates the results were due to CAT differences and not cell numbers. The use of a control with no transfected DNA also ensured protein was due to cell number alone.

A β–Galactosidase (β-Gal) assay was also performed, using the substrate ONPG in a solution of BME and PBS (see Chapter 2 for methodology). The yellow end product was quantified by a microtiter plate reader at 405nm. This assay was used as a control

for transfection efficiency. In many cases, differences in transfection efficiency were seen and in cases where differences were great, the experiment was discarded. For example, if the control without DNA gave a reading equal to or above one of the promoter constructs, then this result would be omitted. In some cases the whole experiment was discarded if the overall β-Gal transfection efficiency was lower than a non-transfected control (data not shown).

The result from the protein assay was used to standardise the amount of extract used in the CAT assay. The β-Gal transfection efficiency control was corrected for after quantification of CAT. Both CAT methods are given in Chapter 2, and the principles are explained below. The CAT ELISA protocol is based on pre-bound antibodies (to CAT) on a microtiter plate, which bind any CAT enzyme from the cell extract. A digoxigenin labelled antibody to CAT is then bound, followed by an antibody to this complex, which is conjugated with peroxidase. The substrate for peroxidase is added, resulting in its cleavage by any bound enzyme and yielding a coloured end-product that can be measured by an ELISA plate reader at 405nm. The assay is in kit form from Roche Diagnostics, with preparation involving the dilution of lyophilised antibodies and solutions.

In comparison to ELISA, the radioactive CAT assay (Promega) quantifies CAT in the cell extract by transferring the n-Butyryl group from the cofactor (n-Butyryl Coenzyme A) to ^{14}C-chloramphenicol. This butyrylated radioactive product runs with a faster Rf value along a thin layer chromatography plate. The converted product can then be quantified by cutting the product from the plate and scintillation counting. This method gave an accurate quantification unaffected by saturation values often seen in overexposed autoradiographs analysed by densitometry. An example of an autoradiograph of a completed thin layer chromatography plate is shown in Figure 5.3(A).

Table 5.1 CAT assays of HC11 cells transfected with the proximal promoter constructs (experiment H-6), in duplicate indicated as a/b.

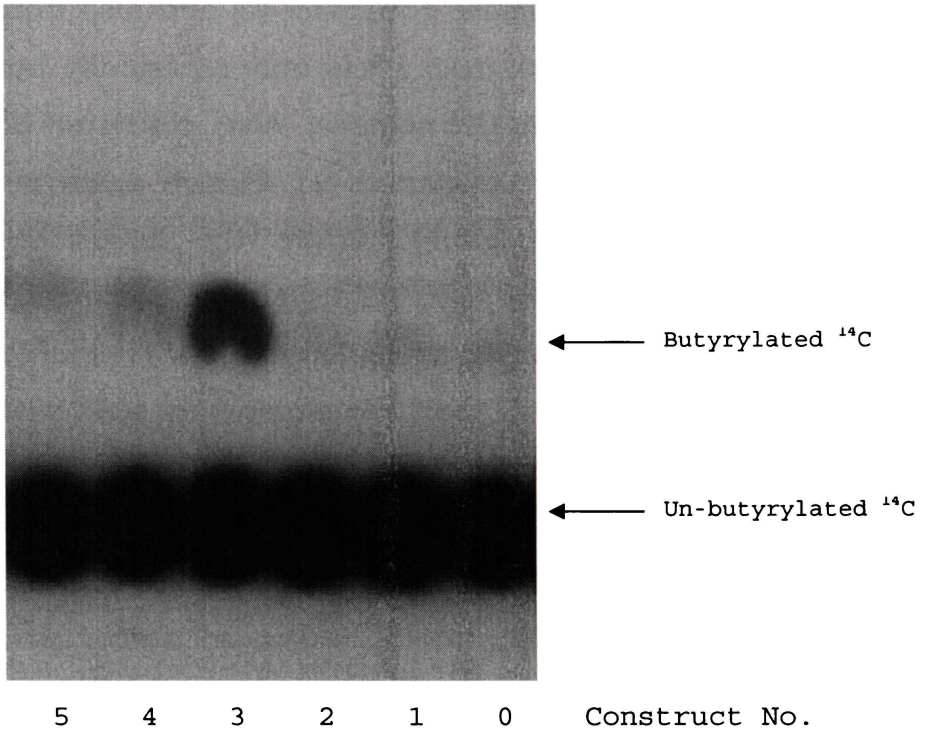
Construct	BCA data	β-Gal data	^{14}C count	Corrected ^{14}C	Average ^{14}C(a&b)	Fold induction
5 a	0.244	0.155	219	327	393	3.0
5 b	0.247	0.149	320	459		
4 a	0.252	0.136	352	461	391	3.0
4 b	0.301	0.134	249	321		
3 a	0.253	0.156	2160	3240	2946	22.2
3 b	0.238	0.15	1840	2653		
2 a	0.306	0.16	127	195	178	1.3
2 b	0.199	0.151	111	161		
1 a	0.269	0.117	156	175	162	1.2
1 b	0.238	0.116	133	149		
0 a	0.268	0.107	148	152	132	1
0 b	0.249	0.111	106	113		
No DNA a	0.275	0.088	N/a	N/a	N/a	N/a
No DNA b	0.298	0.092	N/a	N/a		

From the TLC experimental results depicted in Figure 5.3, the highest activity is seen with construct (3), at about 22 times greater than the negative background level. This is significantly (95% level) different from the negative construct (0) and also from all other deletion constructs. Construct (1) is only marginally but not statistically significantly above the negative. Also, construct (2) shows only a slight increase over construct (1), though again not significant compared to the negative. While the lengthened proximal sequence of construct (3) gives increased activity, a large decrease in activity is seen with the longer constructs (4 and 5). Even so, a 5-6 fold level of activity over the negative is seen with these two constructs, which is significant at the 95% level. (Statistical data given in Appendix 3)

Differences between the TLC and ELISA methods of detecting CAT activity are small. For both the trends were the same, with construct (3) showing a clear increase in activity. A difference between methods is seen in the actual level of induction relative to the negative control for each construct. For constructs (4) and (5), both methods show an approximate three-fold increase in activity over the negative construct (0). Similarly for the constructs (1) and (2), the induction in activity is only 1-2 fold greater than the negative control when measured by either ELISA or TLC. The main discrepancy was for construct (3), though the result is still a clear induction for this construct compared to the other constructs, shown by both methods of detection. These discrepancies were not evaluated to be significant given the limitations of the methods of analysis.

The results shown for one experiment are representative of the accumulated data set of several separate experiments shown in Figure 5.4 and are discussed further as accumulated data. The accumulated data set is shown as individual experiments superimposed on the same graph. This was the chosen method rather than calculating an average over all the data, allowing for individual variances between experiments to be evaluated.

(A) CAT TLC results (H-6)



(B) ELISA results (H-6)

Construct	B-Gal	ELISA result	Corrected CAT	Averaged a/b	Fold induction over negative
5a	.155	0.454	0.66	0.656	2.6
5b	.149	0.470	0.65		
4a	.136	0.577	0.73	0.724	2.8
4b	.134	0.571	0.72		
3a	.156	1.648	2.4	2.368	9.3
3b	.15	1.665	2.33		
2a	.16	0.281	0.42	0.408	1.6
2b	.151	0.281	0.40		
1a	.117	0.255	0.28	0.283	1.1
1b	.116	0.264	0.29		
0a	.107	0.257	0.26	0.255	1.0
0b	.111	0.244	0.25		

(C)

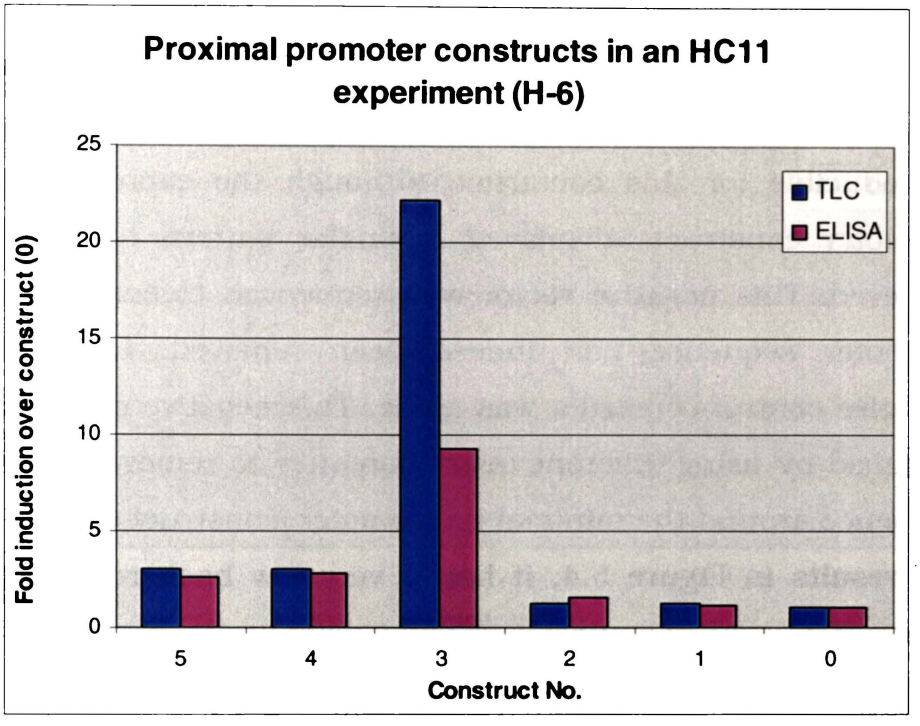


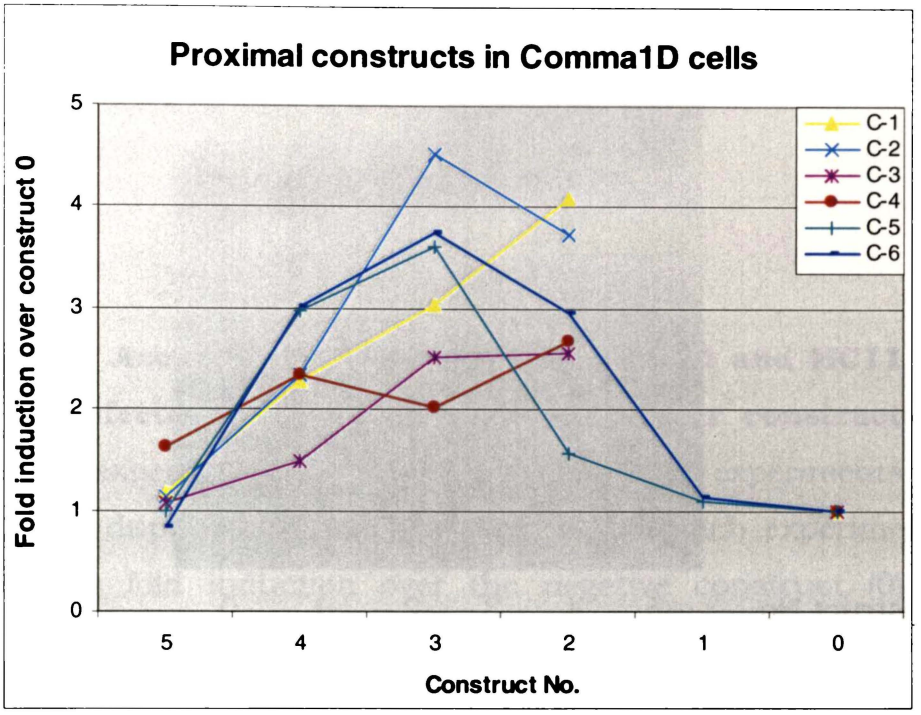
Figure 5.3 Analysis of a single HC11 experiment (raw data given in Table 5.1). (A) Autoradiograph of the thin layer chromatography plate showing ^{14}C chloramphenicol results. The un-butyrylated chloramphenicol is at the bottom with the butyrylated form running faster up the plate. The solvent direction is from bottom to top. (B) Tabulated results from the ELISA method of the same experiment. (C) Graphical representation of both ELISA and TLC results for each construct after corrections have been made and duplicates have been averaged. Results are shown as fold induction over the negative control (construct 0).

In a single experiment (Figure 5.3C for HC11) the difference between the negative control and the promoter constructs is clear. However, over the course of the experimentation, two different negative control plasmids (both with the minimal tk promoter deleted), were found to give varying results. In initial experiments the first negative construct gave a high CAT background activity (approximately equal to that of construct 5). Data was not reproducible for this construct, although the expression of the promoter constructs displayed a similar pattern to that finally achieved. This negative vector was sequenced to ensure that the promoter sequence had indeed been removed. Later, a new negative control construct was made. This negative construct was designed by using different restriction sites to remove more of the sequence around the minimal tk promoter (construct 0). As seen in the results in Figure 5.4, it had a very low background activity. Where the negative control gave some activity, the lowest activity of any other construct was used as the baseline. (Construct 1 gave activity similar to the new negative construct and was therefore a suitable negative control).

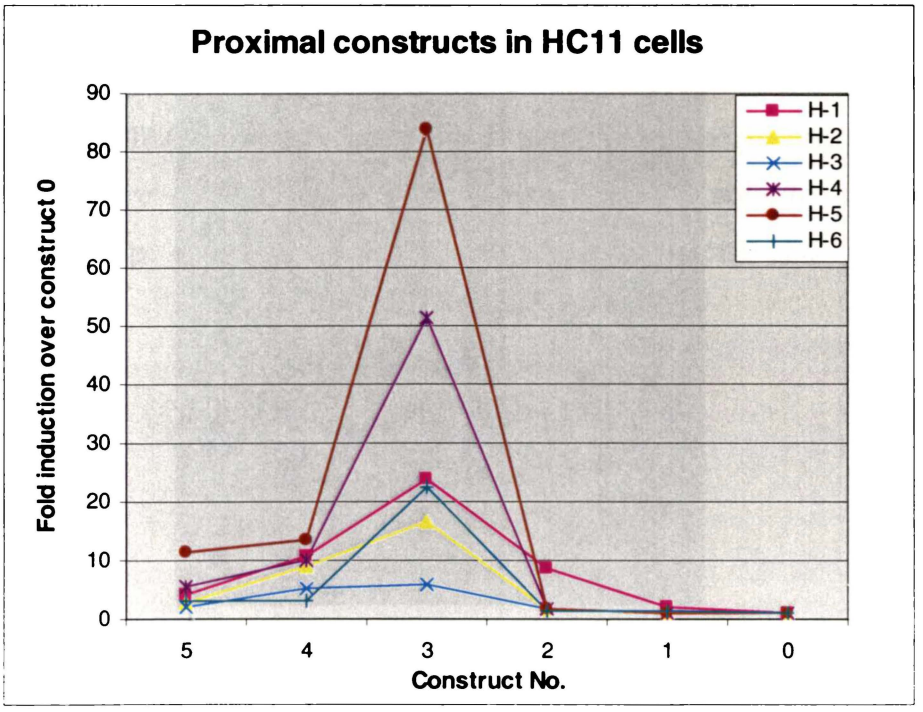
The complete set of data from the Comma1D line represents six experiments, though for construct (1) it is representative of only two. This was due to a bacterial infection causing cell death in several experiments. A new preparation of DNA was made and used for the remainder of the experiments. The trend seen in Figure 5.4A indicates that construct (1) was not sufficient for proximal promoter activity. However, the addition of 30bp to construct (2) gives near maximal activity. The three mid-sized constructs (2, 3, 4) all show high activity relative to construct (1) and the negative control (0), while a decline is seen in the extended construct (5).

The data set representing the HC11 cell line shows a similar trend for highest activity in construct (3). Two of the six experiments show very high activity for this construct compared to construct

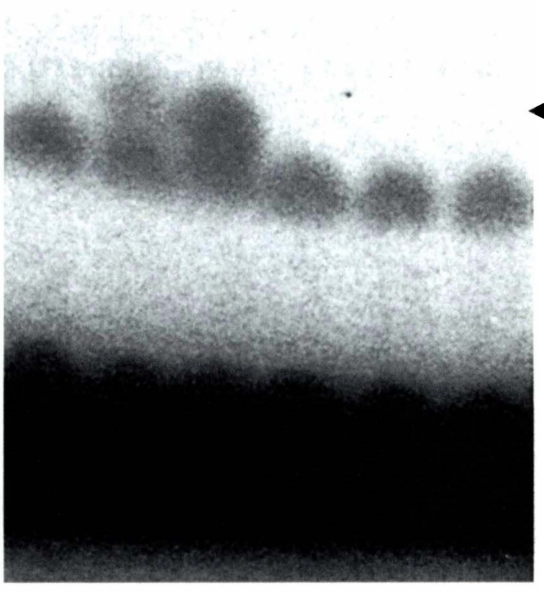
(A)


Proximal constructs in Comma1D cells
Fold induction over construct 0
0
1
2
3
4
5
5
4
3
2
1
0
Construct No.
C-1
C-2
C-3
C-4
C-5
C-6

(B)

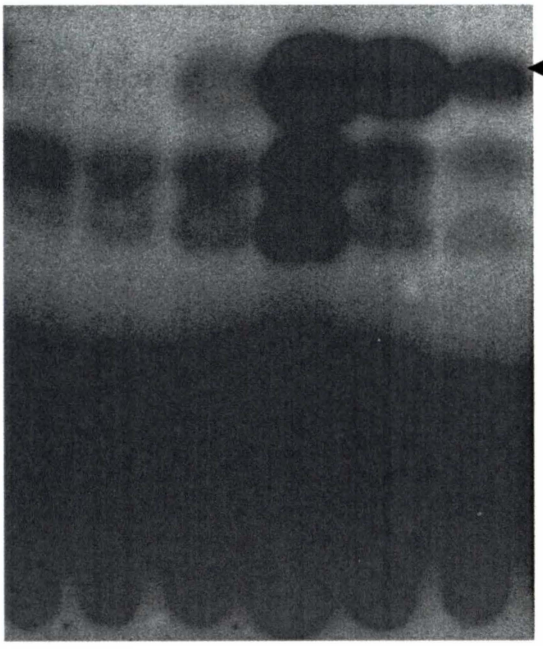

Proximal constructs in HC11 cells
Fold induction over construct 0
0
10
20
30
40
50
60
70
80
90
5
4
3
2
1
0
Construct No.
H-1
H-2
H-3
H-4
H-5
H-6

(C) Comma1D experiment C-6



Construct No. 5 4 3 2 1 0

(D) HC11 experiment H-2



Construct No. 5 4 3 2 1 0

Figure 5.4 Accumulated data from Comma1D and HC11 cell lines transfected with the proximal promoter constructs (A) Comma1D experimental results from six separate experiments (C1-C6), where duplicate samples were averaged. Each experiment is graphed as fold induction over the negative construct (0). (B) Graphical representation of the results from six experiments performed in HC11 cells (H1-H6), transfected with the proximal promoter deletion constructs. Duplicate samples were averaged and the individual experiments plotted. All results were measured by TLC CAT assay. Lines connecting values on the graphs are not representative of intermediate values but are an indication of each data set. (C) TLC plate result from Comma1D experiment C-6. The constructs are listed below the plate, and the solvent ran from bottom to top. (D) TLC plate result of experiment H-2 in HC11 cells.

(0), while the level of activity falls sharply in both shorter and longer constructs.

The overall level of induction between Comma1D and HC11 cell lines differs dramatically. Comma1D cells respond slowly to transfection, showing very low levels of DNA uptake. This cell line also has a longer doubling time, resulting in lower numbers of cells expressing the transfected construct. Within the HC11 line, the condition of the cell population appears to vary between experiments. In experiments H-4 and H-5 the induction of construct (3) was much greater than seen in other experiments (50-80 fold compared to 5-20 fold). Whether or not the high level of induction is a gross overestimation of activity is difficult to explain as all conditions of the experiment were kept as constant as possible. The major difference was most likely a reflection of the status of the cells on the day of the experiment.

The next part of the chapter sets out the findings from studying the distal 5' promoter region. Sequences further beyond the proximal promoter may contain enhancer elements that may be capable of increasing activity to account for the biological levels observed *in vivo*.

5.2.2 Regulation of *Ese*-2 promoter activity including the distal 5' promoter region

(a) Construct design

In order to examine sequences outside the proximal promoter, a further three deletion constructs were generated. From the original clone SacI4.9kb (see Figure 3.1) a total of 2.6kb of 5' sequence was isolated. Again, the new constructs were based on the location of suitable restriction sites that divided the sequence into appropriate sizes. An *Xba*I site in the pBluescript II KS vector was used to generate the full 2.6kb construct (construct 8, shown in Figure 5.5). This was cloned into the pBLCAT2 expression vector at sites

*Xho*I-*Xba*I. Similarly, the next construct of 1.7kb total length (construct 7) was excised from another *Xba*I site within the promoter sequence. The third construct was generated from an *Eco*RV digest, which gave a construct size of 1.3kb (construct 6). This was cloned into the expression vector that had been digested with *Xba*I, Klenow blunt ended and digested with *Xho*I. Cloning of these constructs was relatively straight- forward and they were confirmed to be correct by sequencing from each end across the ligation boundaries. Two preparations of each construct were made and purified by Qiagen maxiprep column to ensure their purity.

(b) Transfection of the constructs

Transient transfections were performed for the distal constructs in the same manner as for the proximal constructs. Both HC11 and Comma1D cell lines were transfected with Fugene6 as the transfection agent for all experiments. A summary of the results is given in Figure 5.6 representing the data from a set of seven (Comma1D) or six (HC11) experiments, with each construct transfected in duplicate and the duplicates averaged.

When averages were calculated from the seven Comma1D experiments of distal *Ese*-2 constructs, no significant difference (95% level) was detected between any construct and the negative construct. For the six HC11 experiments, averaged results showed a significant difference at the 95% level, for construct (6) and (7) above the negative construct.

For HC11 experiments, a 7-10 fold average increase was seen for all distal constructs over the negative, though the error between experiments is high. Again in the HC11 experiments, two separate experiments show much higher levels of activity than other experiments. These differences are most likely due to differences in the cell culture from day to day.

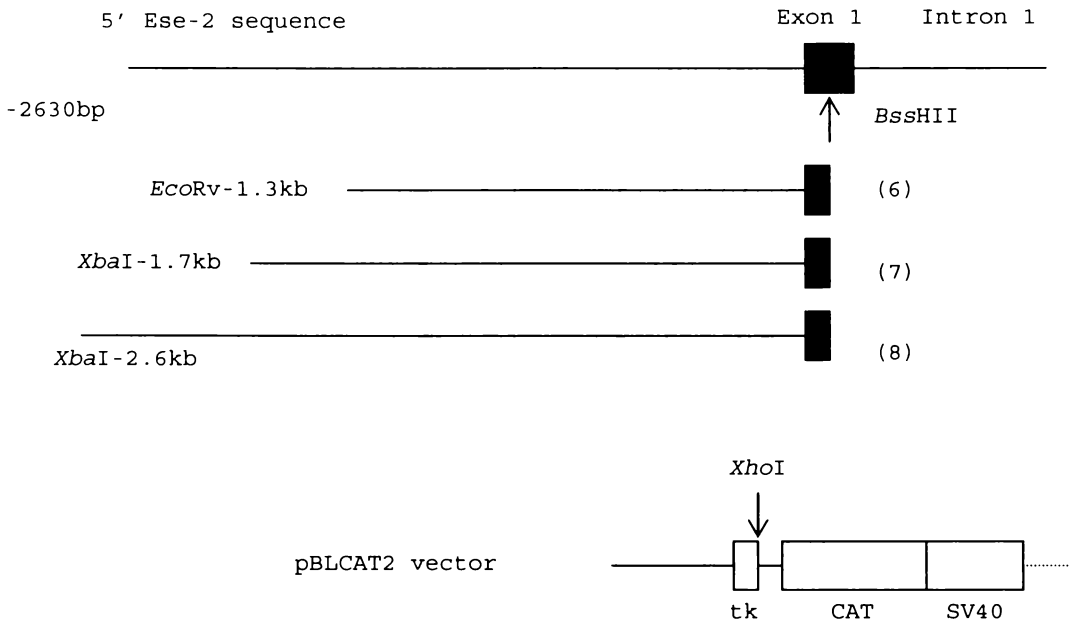
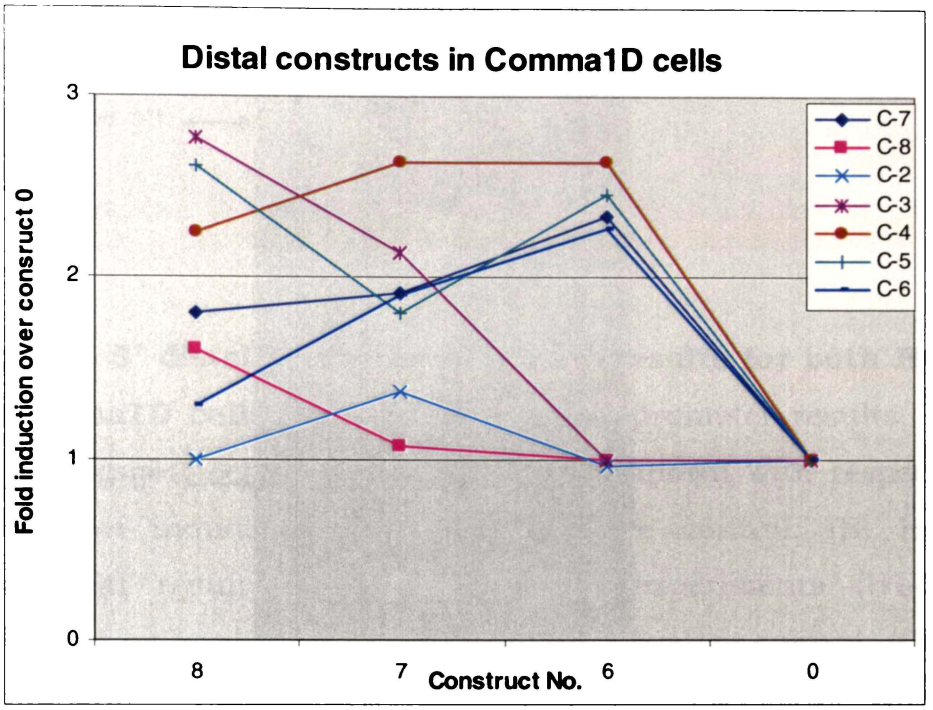
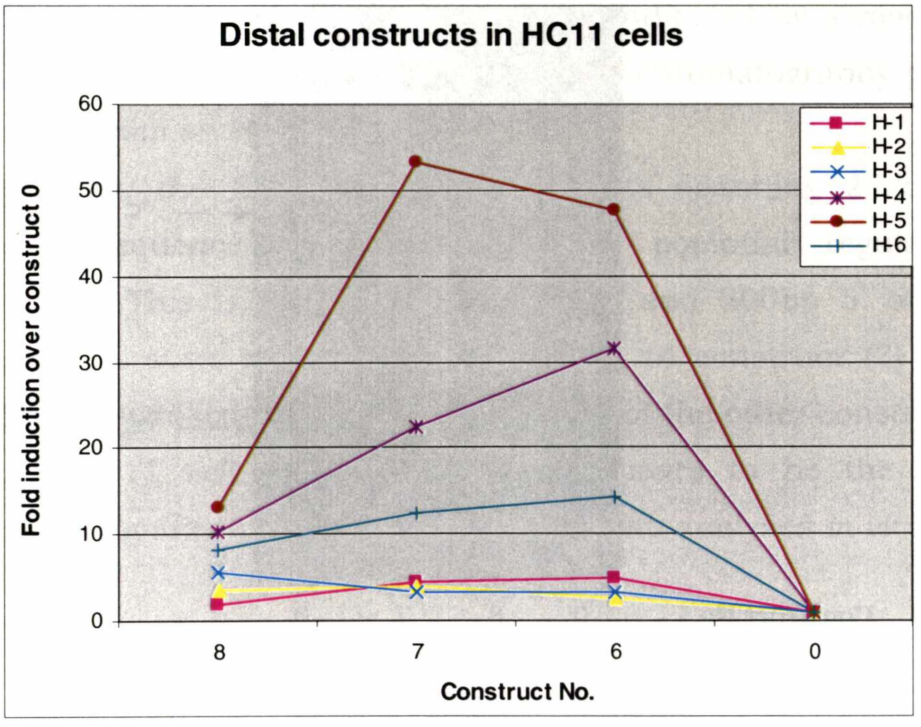


Figure 5.5 Deletion constructs of the distal 5' promoter The full length of 5' sequence is depicted above the constructs, with exon 1 represented by the black box, and the cloning site of *Bss*HII indicated by the arrow. The constructs are not drawn to scale. Below the constuct representation is an outline of the expression vector, the arrow points to the location of the *Xho*I site used in cloning. Each construct is numbered to the right of exon 1.

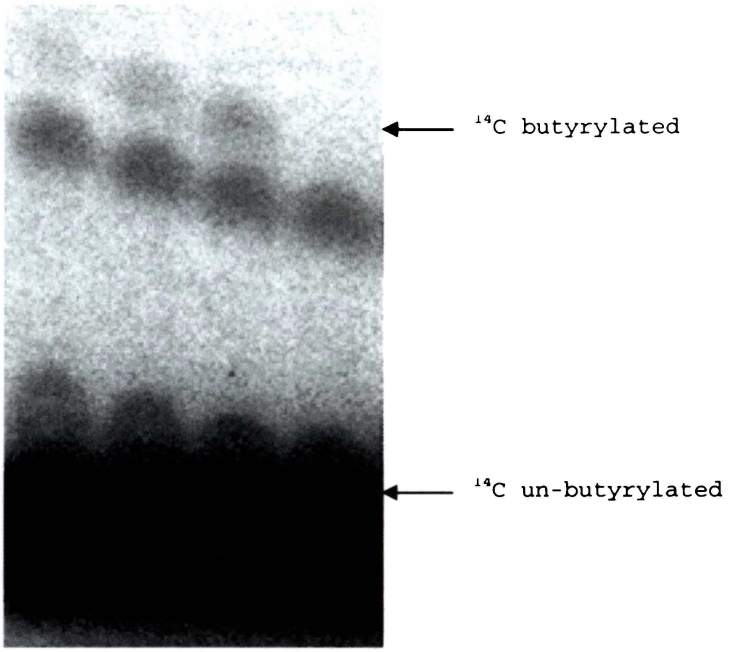
(A)


Distal constructs in Comma1D cells
Fold induction over construct 0
3
2
1
0
8
7
6
0
Construct No.
C-7
C-8
C-2
C-3
C-4
C-5
C-6

(B)

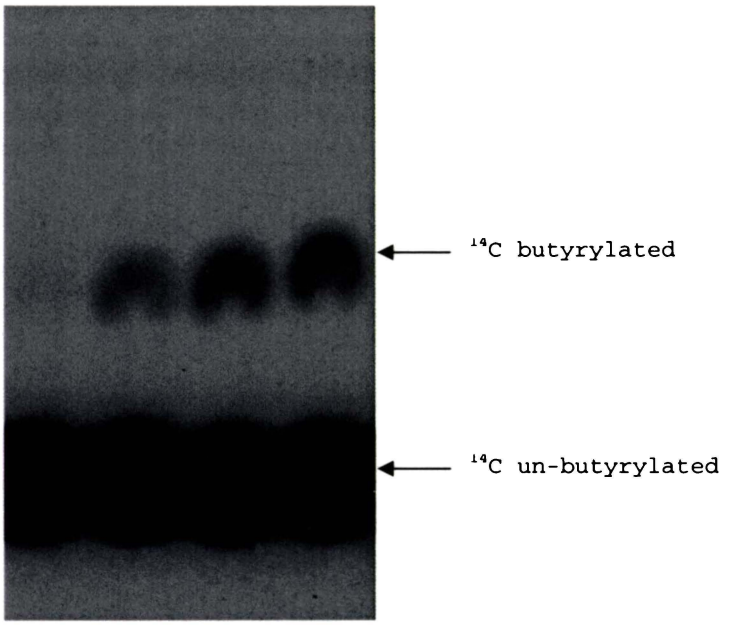

Distal constructs in HC11 cells
Fold induction over construct 0
60
50
40
30
20
10
0
8
7
6
0
Construct No.
H-1
H-2
H-3
H-4
H-5
H-6

(C) Comma1D experiment C-6



Construct No. 8 7 6 0

(D) HC11 experiment H-6



Construct No. 0 8 7 6

Figure 5.6 5' distal promoter construct results for both HC11 and Comma1D cells (A) A graph of distal promoter results from seven Comma1D cell experiments (C2-C8), shown with respect to the level of induction above the negative control. (B) HC11 experimental results from six separate experiments (H1-H6), represented as level of induction over the negative control. (C) An autoradiograph of a thin layer chromatography plate with samples from experiment C6. Lanes are marked below the autoradiograph. The solvent direction is from bottom to top. The butyrylated and un-butyrylated chloramphenicol forms are indicated by arrows on the right. (D) Autoradiograph of a thin layer chromatography plate of samples from an HC11 experiment (H6).

In summary, the *Ese*-2 deletion constructs spanning 2.6kb of promoter sequence reveal a region of potential regulatory importance. This is located between 76bp and 200bp 5' of the transcription start site, where the activity of construct (3) was found to be significantly different from each of the other constructs (for the HC11 cell line). This region appears to be the only potentially significant location within the 2.6kb analysed *in vitro*.

5.2.3 Regulation within the first intron of *Ese*-2

To look for further regulatory elements, the first intron of *Ese*-2 was studied. Other studies have shown Intron 1 as a potential site of regulation, including for the Ets factor *Elk*-1 (Lehmann *et al.*, 1999), the mouse *connexin*40 gene (Seul *et al.*, 1997) and the α-*fetoprotein* gene (Scohy *et al.*, 2000).

(a) Construct design

The ninth construct was generated to contain the entire first intron, the first exon and 5' promoter sequence as far as 2.6kb, where construct (8) finished. It contains a total of 9kb of sequence. It also contains a mutation in the ATG start codon in the beginning of exon 2, to ensure that translation is initiated from the ATG codon present in the pBLCAT2 expression vector. Primers were designed at each end of the region for the construct (see Figure 5.8). The 3' primer contained a mutation at the ATG, changing it to TTG (primer exon2ATG*mod.), and an engineered *Xho*I site for cloning. The forward primer (Far 5' *Sac*I) encompassed a *Sac*I restriction site for cloning. PCR was performed as described in Chapter 2 using the Elongase system (Roche) for amplifying long products. This amplification did not initially work even after manipulation of conditions including magnesium concentration, annealing temperature, extension time and template concentration. Seven more primers were generated at the 5' end (Far5' -#30, -#50, -#84, -#104, -#146, Far5' *Sph*I mod., Far5' *Xba*I) to achieve better annealing. Moderate success was achieved and a small proportion of a 9kb product was excised from an agarose gel from multiple PCR reactions that were accumulated.

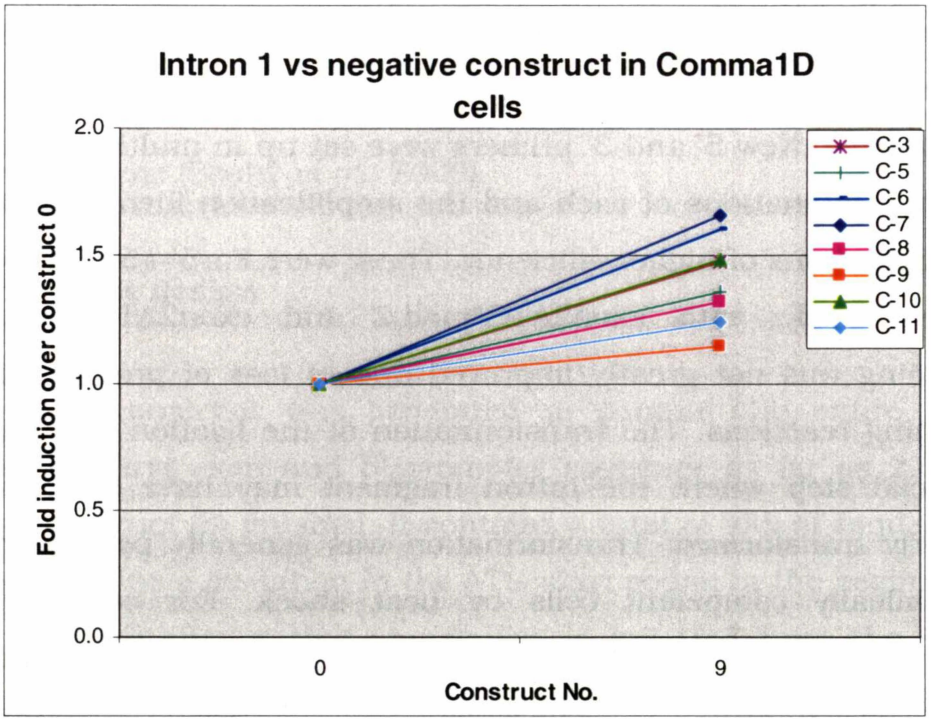
Attempts to clone this product were unsuccessful. Cloning directly into the CAT expression vector by digesting the PCR product with *Xho*I was limited by the very low concentration of digested DNA recovered. Cloning into the pGEM-TEasy vector (Promega) was also

limited by the tailing reaction required to generate enough product with suitable "A tail" for cloning. Further primers were generated at the 3' end to try and increase the yield of product. These were exon2ATG*mod.2 and exon2ATG*mod.*Bgl*II. The first contained a different location for the *Xho*I site and the second contained a *Bgl*II restriction site, which may digest better at the end of the fragment than *Xho*I. New 5' and 3' primers were set up in multiple reactions with combinations of each and the amplification identified several combinations of higher efficiency. These were Far5'-#51, -#104 and -*Sph*I mod., with exon2ATG*mod.2 and exon2ATG*mod.*Bgl*II. Cloning was not greatly improved due to loss of product in pre-cloning reactions. The transformation of the ligation was another crucial step where the intron fragment may have been lost or poorly transformed. Transformation was generally performed into chemically competent cells by heat shock. For construct (9), electroporation of the DNA was performed to enhance transformation efficiency. This resulted in many more clones, though still none of those selected and tested contained the correct sequence. Finally, using STBL2 cells (Invitrogen) transfected by heat shock, success was achieved. Many PCR reactions were accumulated with post-amplification modification and a very large amount ligated into the pGEM-TEasy vector. From pGEM-TEasy, the 9kb product was digested with *Sph*I and *Bgl*II and cloned into the pBLCAT2 expression vector.

(b) Transfection of the construct

Several attempts were made to transfect construct (9) into HC11 and Comma1D cells in the same way as for the other constructs. An equi-molar amount of DNA was co-transfected with pCH110 using an optimised volume of Fugene6. No transfection was observed as evidenced by baseline βGal levels, and CAT activity in these experiments. However, other constructs in the same experiment showed high βGal and CAT activities (data not shown).

(A)


Intron 1 vs negative construct in Comma1D cells
Fold induction over construct 0
2.0
1.5
1.0
0.5
0.0
0
9
Construct No.
C-3
C-5
C-6
C-7
C-8
C-9
C-10
C-11

(B)

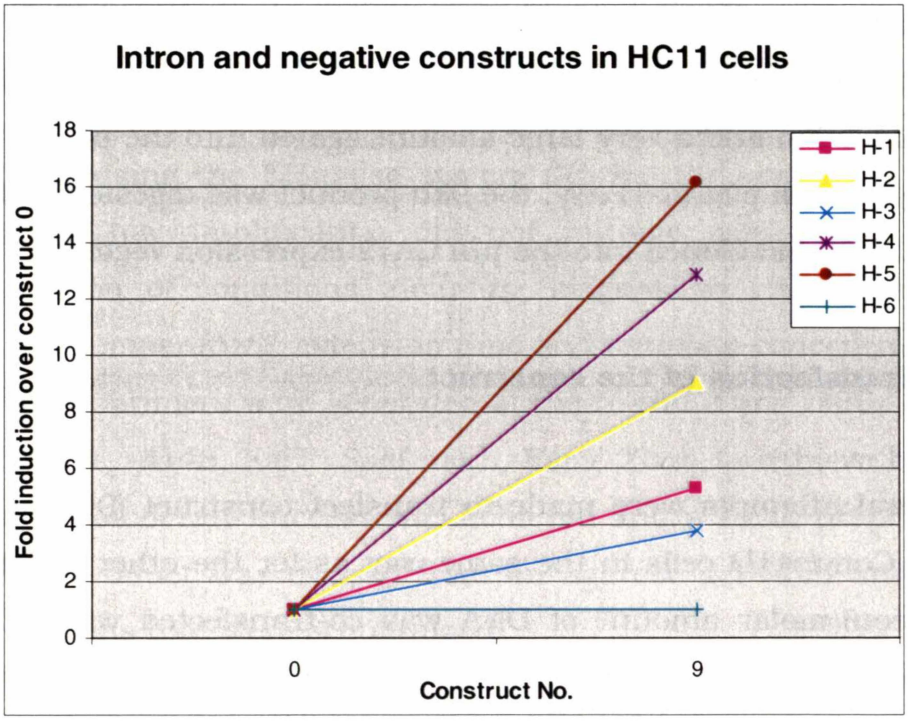

Intron and negative constructs in HC11 cells
Fold induction over construct 0
18
16
14
12
10
8
6
4
2
0
0
9
Construct No.
H-1
H-2
H-3
H-4
H-5
H-6

Figure 5.7 Intron 1 and negative construct results in HC11 and Comma1D cells (A) Graphical representation of eight separate experiments of Comma1D cells transfected with construct (9) (C3,5-11). (B) Graphical analysis of six HC11 cell line experiments (H1-H6) transfected with construct (9) and (0). All results are shown as relative to the negative construct. Coloured lines are used as a graphical device purely to differentiate data.

The preparation of new plasmids did reduce this problem and it appeared that one of these preparations contained some contaminant that inhibited both the construct and pCH110 transfection. Using the new preparations, measurable transfection efficiency was obtained, though not comparable to that of the proximal and distal 5' constructs. Construct (9) was much larger in size than other constructs, being 13kb in total, whilst construct (8) was only 7kb. This size difference was considered to be a major factor contributing to the low transfection of the construct. Normalising for the construct (9) CAT activity in reference to the transfection efficiency was not reliable as the pCH110 transfection may also have been affected. To avoid this, a volume of "junk" DNA was added to the shorter constructs to make the total concentration of DNA equal between transfections. Figure 5.7 shows eight Comma1D transfected lines and six HC11 lines. The activity from construct (9) containing the first intron, did not show any increased activity over the level seen with other constructs. In Comma1D cells the activity only marginally increased above the proximal promoter constructs. In the HC11 cells the activity of construct (9) ranged between a single and 16 fold increase over the negative construct. The large difference between experiments reflects the poor transfection, and other technical difficulties with the system. The size of construct (9) (13kb total) may have reduced the apparent activity of the construct, based on its low transfection efficiency. This is examined in the next section through the use of smaller constructs containing part of intron 1. This approach varies the plasmid size, intron size and potentially the RNA processing of the transcript.

5.2.4 Reduced intron 1 constructs

(a) Construct design

To reduce the size of the constructs, but retain some of the intron 1 and proximal promoter, three constructs were generated. Each of these contained 500bp of proximal promoter sequence, equivalent to that of construct (4). They also contained between 1.5 and 3kb of intron 1 sequence directly attached to the proximal sequence, preceding the partial CAT gene in pBLCAT2. The constructs are depicted in Figure 5.8. Construct (10) contains the most 5' sequence from intron 1, and is continuous genomic sequence. Construct (11) was generated from 2.5kb of intron 1 sequence and the 500bp of proximal promoter, containing a gap in the genomic sequence. Similarly, construct (12) is not continuous, but contains the most 3' 2kb of intron 1 and approximately 15bp of exon 2 with a mutated ATG codon. Construct (10) was generated by releasing a *Bam*HI 3kb fragment from construct (9) and re-ligating into pBLCAT2 digested with *Bam*HI and *Bgl*II (to remove the minimal tk promoter). The generation of this construct was not straight-forward, appearing to be hindered by sequence within the 5' region of the first intron. Using STBL2 (Invitrogen) cells and high ratios of insert to vector in the ligation, a single clone was correct from sampling over one hundred.

Construct (11) was generated by subcloning the *Bam*HI-*Hin*dIII fragment into the pEGFP-N1 vector (CLONTECH) and later releasing it from this vector using the restriction enzymes *Sal*I and *Xho*I. This allowed cohesive termini to re-ligate into construct (4) that had been linearised with *Xho*I. The orientation of this clone was random, but was later checked by sequence analysis to confirm the correct orientation was achieved. The 3' intron 1 fragment of 2kb was also subcloned into another vector (pGEM-3), which allowed for its release at *Sac*I and *Xho*I sites to be cloned in the same manner as construct (11) into the CAT expression vector.

(b) Transfection of the constructs

Constructs (10), (11) and (12) were transfected into HC11 and Comma1D cells as described in Chapter 2. Each cell line was transfected four times in duplicate with each construct, including the negative (construct 0). The results were corrected for βGal activity and averaged between duplicates. They are shown graphically in Figure 5.9.

Overall very low levels of activity were observed for these constructs. The trend seen for both HC11 and Comma1D cells showed that the activity generated by the complete intron and promoter (construct 9) was mostly accounted for by construct (12) (3' intron region). Again construct (9) did not transfect well, so that activity may be much higher than measured.

Due to the shortening of the constructs, splice sites have been deleted. This may or may not be significant with respect to the DNA processing of the transcript. Construct (10) in particular has a splice donor site but no acceptor, which may cause aberrant splicing. Aberrant splicing is less likely to occur in construct (11) which has no splice sites, or construct (12), which has a splice acceptor but no donor. Analysis of the RNA would resolve this, by showing the size of the spliced products. The potential for aberrant splicing means interpretation of the results must be made with caution.

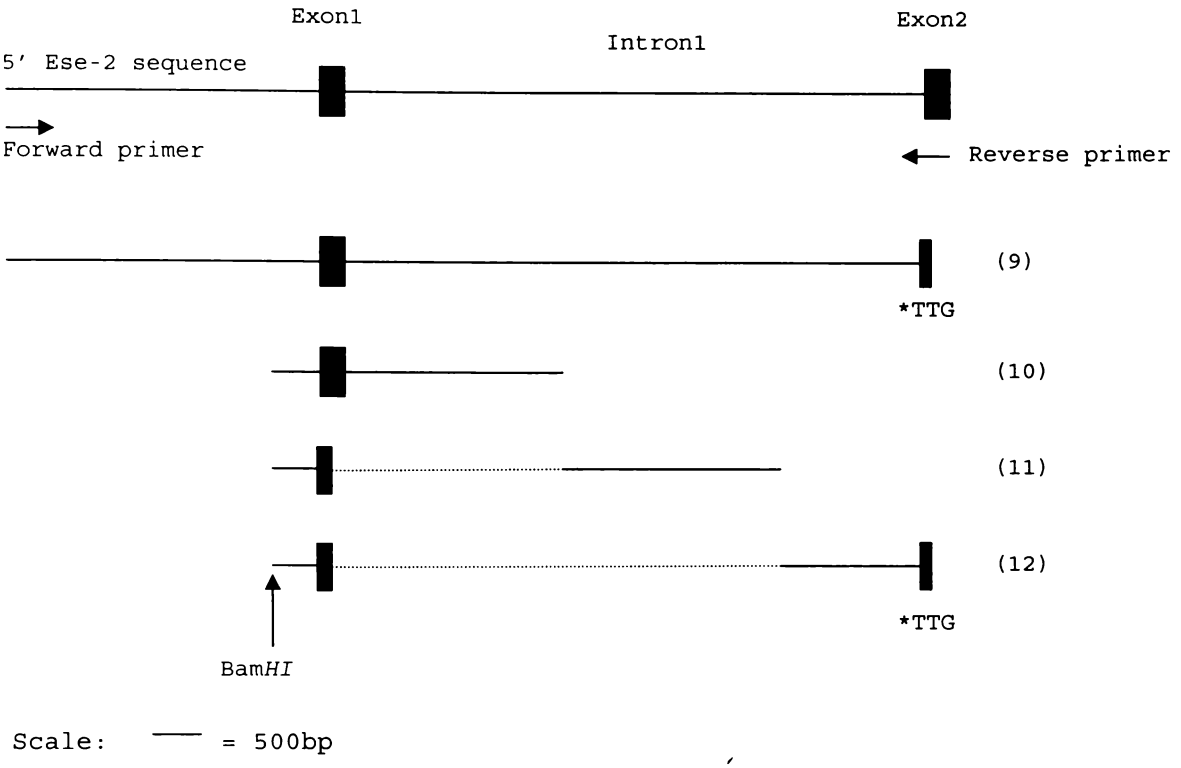


Figure 5.8 Deletion constructs (9), (10), (11) and (12) The full length genomic sequence is depicted at the top of the diagram by a line, with exons 1, 2 and intron 1 marked. Construct (9) encompasses the entire 5' region of 2.6kb and the 9kb region of intron 1, with a mutation of ATG to TTG at the start codon. Constructs (10), (11) and (12) contain 500bp of proximal promoter sequence and a segment of the intron. Construct (10) contains a splice donor site only, construct (11) contains no splice sites and construct (12) contains a splice acceptor site only. The dotted line represents deleted sequences. Each construct then has the CAT sequence immediately 3' followed by the remainder of the expression vector sequence. The size of the exons is not to scale.

5.2.5 Hormone supplemented cell culture

In separate experiments using selected promoter and intron deletion constructs, transfected cells were treated with progesterone or estradiol. The effect of the hormone was studied in relation to the activity of the *Ese*-2 construct. These two hormones were chosen for their known role in mammary processes (Mepham, 1987). Such hormones may play a secondary role in the regulation of *Ese*-2 as it is found at elevated levels in the mammary gland at stages of development co-inciding with the hormone levels.

Cells were transfected as usual then treated with 10^{-7}M progesterone or estradiol. The treatment was given 24 hours after transfection and cells were harvested 48 hours after this. Both HC11 cells and Comma1D cells were included in experiments. Each line was transfected at least three times with the constructs (0), (4), (8), (9) and (12). These constructs were selected as a representative sample from all the constructs. Construct (9) contained the complete intron, and construct (12) had shown some activity as part of the intron sequence. Construct (4) contained the 5' sequence used in the intron constructs and construct (8) was the largest 5' construct.

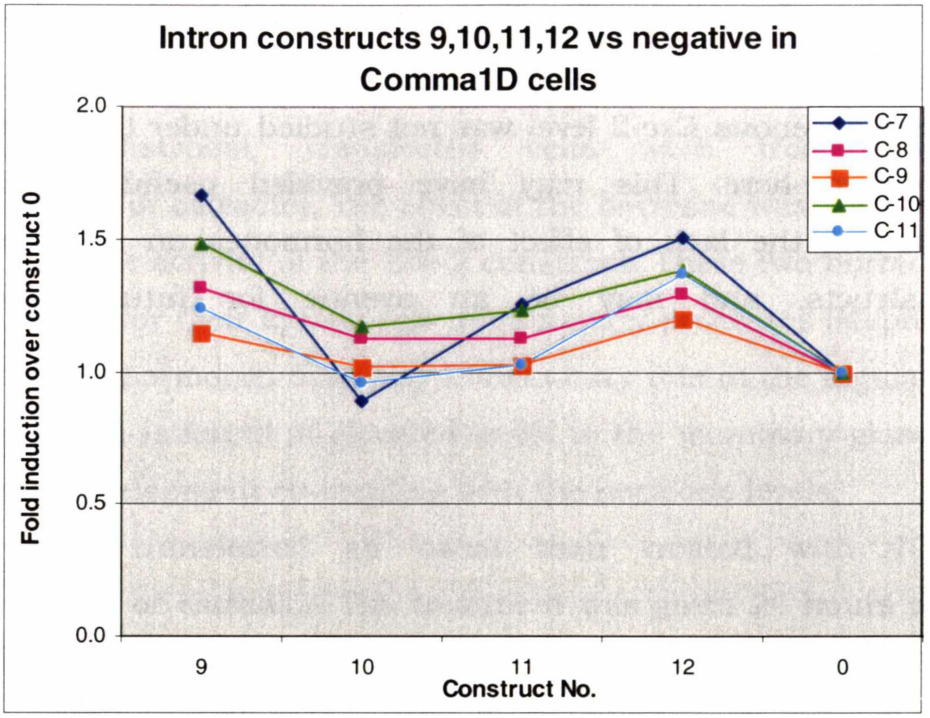
The results of one experiment from HC11 cells are shown in Figure 5.10. Results from the Comma1D cell line experiments did not show any difference in CAT activity for each construct with or without hormone. Between the constructs, the differences were the same as already observed.

Results from the progesterone treated cells (Figure 5.10A) show that the addition of progesterone may have had some effect on construct (12). All other constructs had a level of activity between untreated and treated cells. In a repeat experiment (data not shown) the level of activity was higher overall, but there was no difference between treated and untreated cells, even for construct 12.

In the estradiol treated cells, no difference was observed between treated and control cells. This was verified in a repeat of the experiment (not shown).

The endogenous *Ese*-2 level was not studied under the conditions described here. This may have provided useful information regarding the lack of effect of the hormones on the promoter constructs, and may be an avenue for future research.

(A)


Intron constructs 9,10,11,12 vs negative in Comma1D cells
Fold induction over construct 0
2.0
1.5
1.0
0.5
0.0
9
10
11
12
0
Construct No.
C-7
C-8
C-9
C-10
C-11

(B)

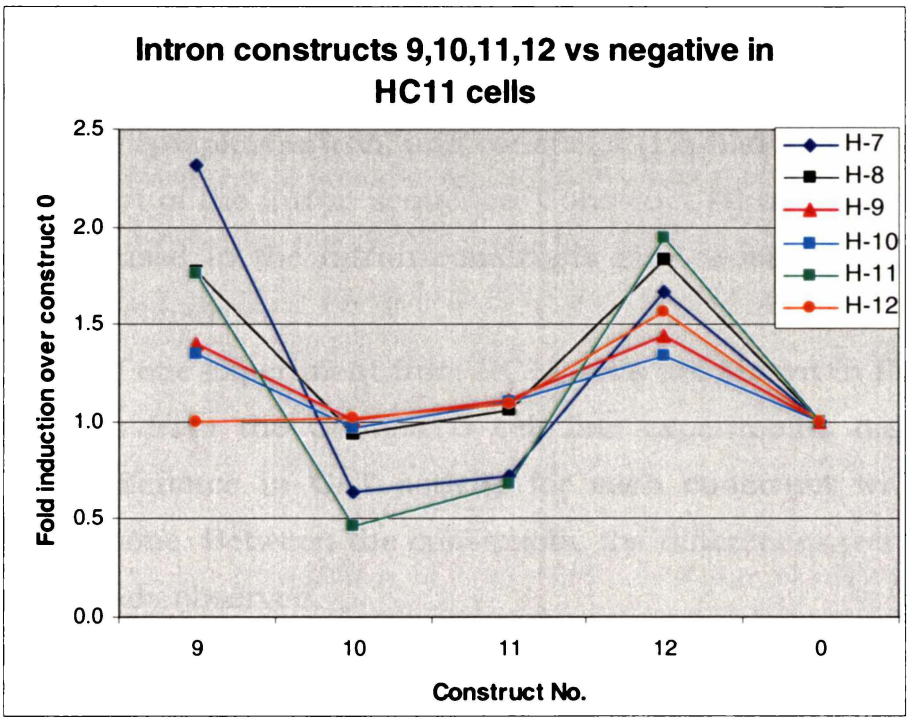
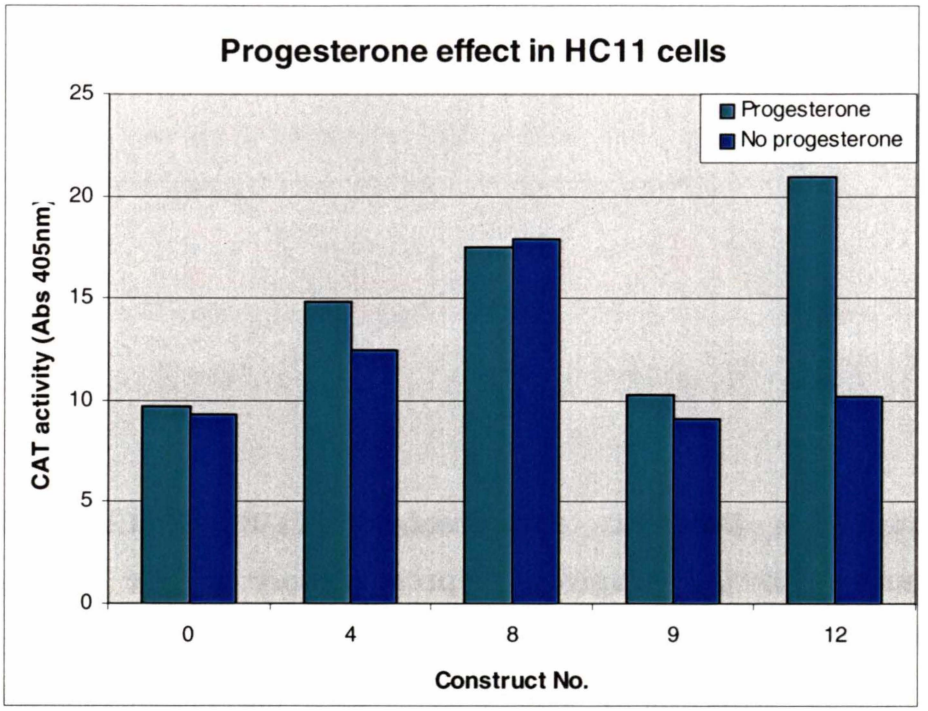

Intron constructs 9,10,11,12 vs negative in HC11 cells
Fold induction over construct 0
2.5
2.0
1.5
1.0
0.5
0.0
9
10
11
12
0
Construct No.
H-7
H-8
H-9
H-10
H-11
H-12

Figure 5.9 Deletion constructs (9),(10),(11),(12) and the negative (0) (A) Graphical representation of five experiments performed in duplicate in Comma1D cells (C7-C11), with constructs (9),(10),(11) and (12). Each experiment is graphed relative to the amount of induction above the negative construct, which has been normalised to 1. (B) HC11 cell line experiments representing six separate experiments (H7-H12), where duplicate samples have been averaged. The results are calculated as fold induction above the negative construct.

(A)


Progesterone effect in HC11 cells
Progesterone
No progesterone
CAT activity (Abs 405nm)
25
20
15
10
5
0
0
4
8
9
12
Construct No.

(B)

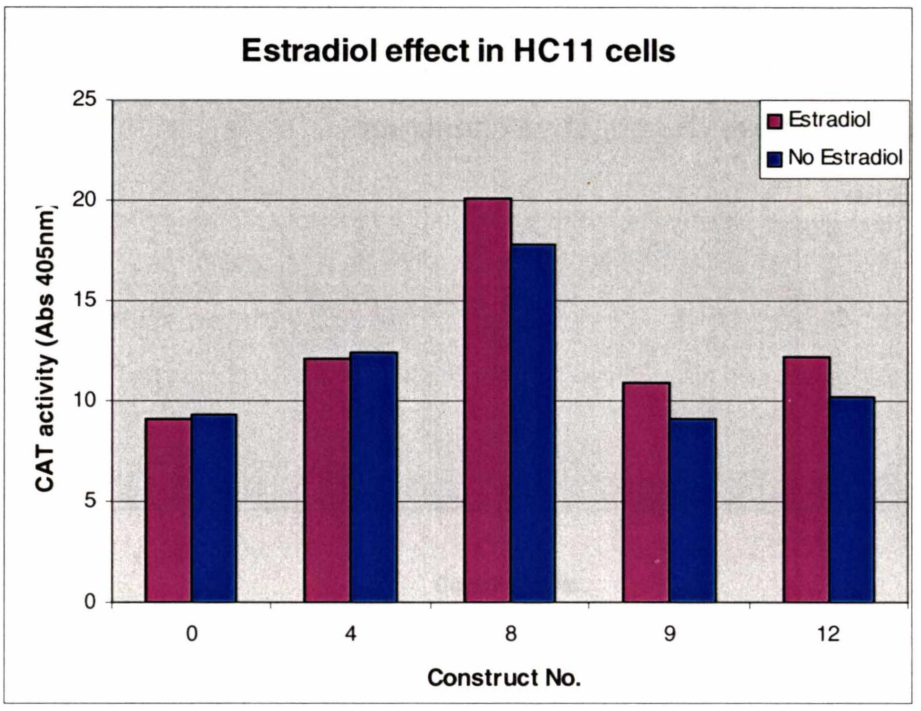

Estradiol effect in HC11 cells
Estradiol
No Estradiol
CAT activity (Abs 405nm)
25
20
15
10
5
0
0
4
8
9
12
Construct No.

Figure 5.10 Hormone treatment experiments in HC11 cells (A) HC11 cells transfected with constructs (4), (8), (9), (12) and (0) followed by treatment of 10^{-7}M Progesterone. The graph shows an averaged figure for each construct representing 2 duplicate experiments. Each construct was also assayed without hormone (blue bars). (B) Estradiol (10^{-7}M) treated HC11 cells. Experimental conditions were the same as described for (A).

5.2.6 Effect of serum on *Ese*-2 deletion construct activity

Several deletion constructs were tested in cell culture for their ability to respond to serum. As for the hormone treatments, only five constructs were tested for simplicity. These were constructs (0), (4), (8), (9) and (12). HC11 cells were transfected with each construct under normal serum conditions (8% FCS) and 16-20 hours later the media was removed and replaced with a series of dilutions of media plus serum. These ranged from 0.1% to 20% serum. The cells were harvested after 48 hours with the serum dilutions and assayed. Results from one experiment are shown in Figure 5.11.

The protein levels of each sample show that cell density is lower in those cells treated with lower percentages of serum. This was also reflected in transfection efficiency. Where there were fewer cells present, the βGal activity was lower. These factors were accounted for after the CAT activity was measured. Overall, four sets of stably transfected cell lines were analysed and when averaged the data showed no response to serum was occurring in the *Ese*-2 deletion constructs as examined *in vitro*.

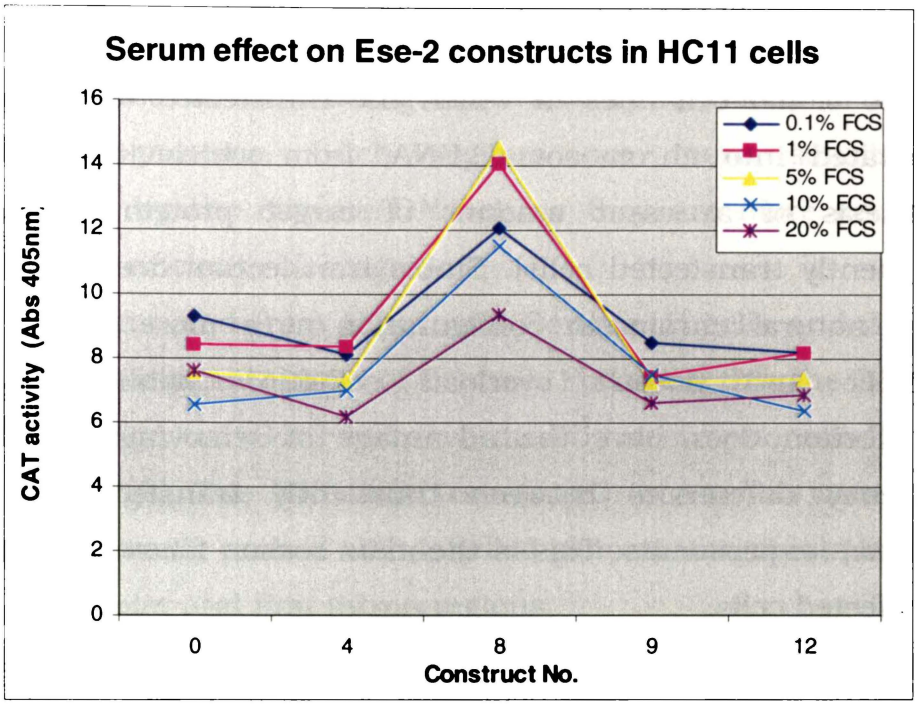


Figure 5.11 Effect of serum on *Ese*-2 deletion constructs in HC11 cells The data shown represents one of four experiments where each construct was tested in duplicate and the results averaged. Each line indicates the percentage of FCS in the media. Constructs tested were negative (0), (4), (8), (9) and (12).

5.2.7 Stable transfection of deletion constructs

Differences have been previously reported between stable and transient transfection on the activity of constructs tested in cell culture (Berland and Chasin, 1988). Stably transfected cells give rise to clonal cell lines in which the transfected target gene is integrated into chromosomal DNA, from which it directs the synthesis of a lesser amount of target protein (relative to transiently transfected cells). Stable transfection does result in a more natural chromatin configuration at a more normal copy number and doesn't overload cellular machinery. Stable transfection does have the advantage of removing transfection efficiency differences between transiently transfected cells of different experiments. This is the main reason for studying stably transfected cells.

To extend the results obtained with the transient system used in the previous experiments, a set of stably transfected cells was also produced. The method for the transfection is given in Chapter 2. Stable transfection was performed for constructs (8-2) and (0) in HC11 cells. Five sub-lines from each construct were grown. The CAT activity results from one duplicate experiment are shown in Figure 5.12.

Results from this experiment are consistent with those from transient experiments. High activity was observed in construct (3), which was reduced when the sequence was extended to construct (5). Due to technical difficulties with construct (4), no data is available for that construct in Figure 5.12.

5.2.8 DNaseI hypersensitive analysis

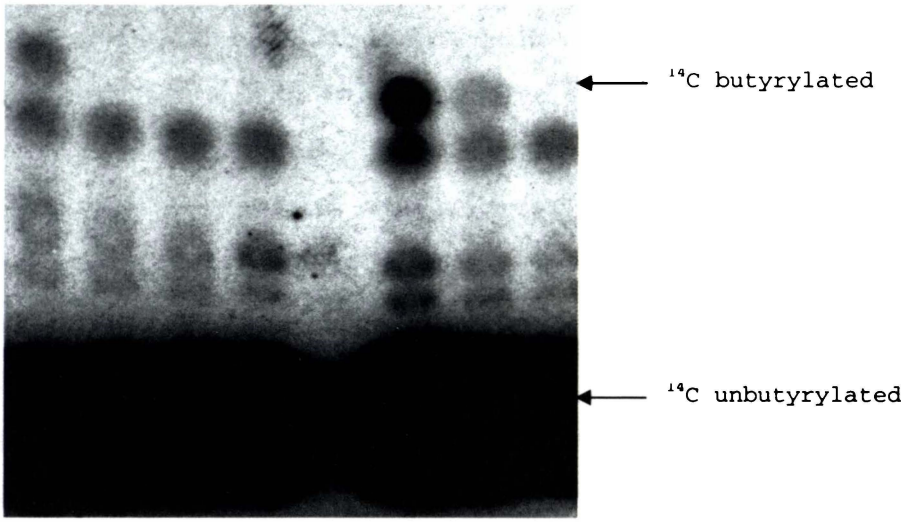
As another approach to search for important regulatory regions, DNaseI hypersensitive analysis (DHS) was performed. DNaseI hypersensitive sites represent regions where nucleosomes have been displaced by the binding of a transcription factor or by DNA unwinding proteins. Such regions are more sensitive to degradation by the enzyme DNaseI, as they are more accessible and open to nuclease cleavage.

DNA was extracted from HC11 and Comma1D cells and digested with various concentrations of DNase1. Following this the DNA was restriction digested (*Sac*I) and separated by electrophoresis through an agarose gel. The DNA was transferred to a nylon membrane, and subsequently probed with labelled genomic fragments from the *Ese*-2 promoter and first intron regions.

Difficulties with both the recovery of DNA and degradation of DNA, resulting from contaminating DNases limited the quality of the results. Figure 5.13 shows one example of DHS on HC11 and Comma1D cells. The probe spanned approximately 850bp of the first intron, which hybridised to a 4.8kb *Sac*I fragment. Increasing amounts of DNase1 revealed the generation of a 1.6kb DNase1 digestion product.

The same pattern was seen in both HC11 and Comma1D cells. The size of the product and probe indicate that the DHS region falls in the 5' region of the first intron, as explained in Figure 5.13B. Only one fragment was generated even at higher DNase I concentrations, suggesting only one hypersensitive site lies in the region of this probe. Multiple probes were hybridised to several membranes, though most resulted in non-specific binding to regions such as the ETS domain or repetitive sequences, and were consequently of no use (data not shown).

(A)



Construct No. 8 7 6 5 4 3 2 0

(B)

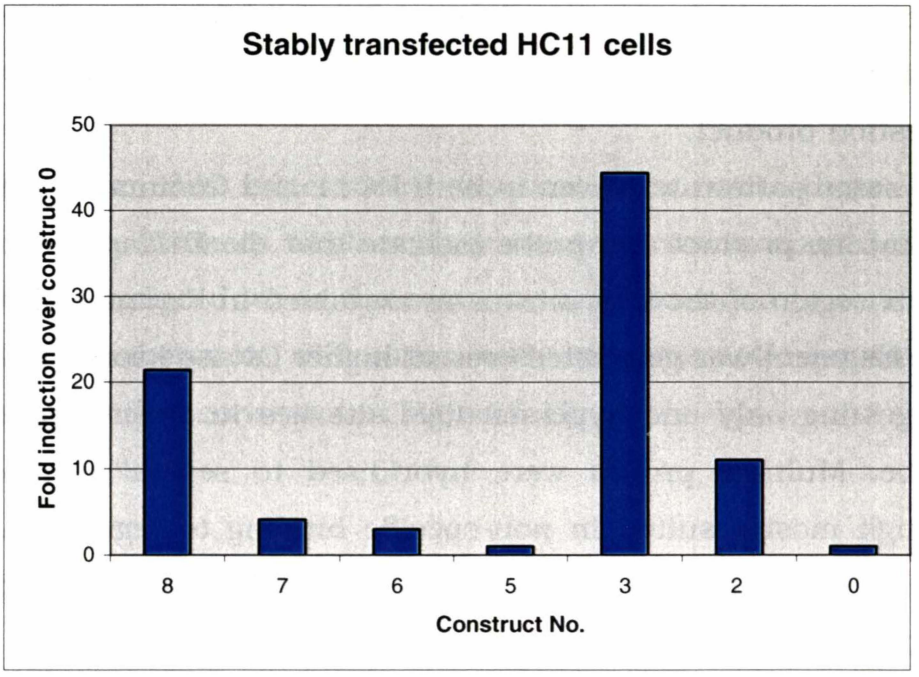
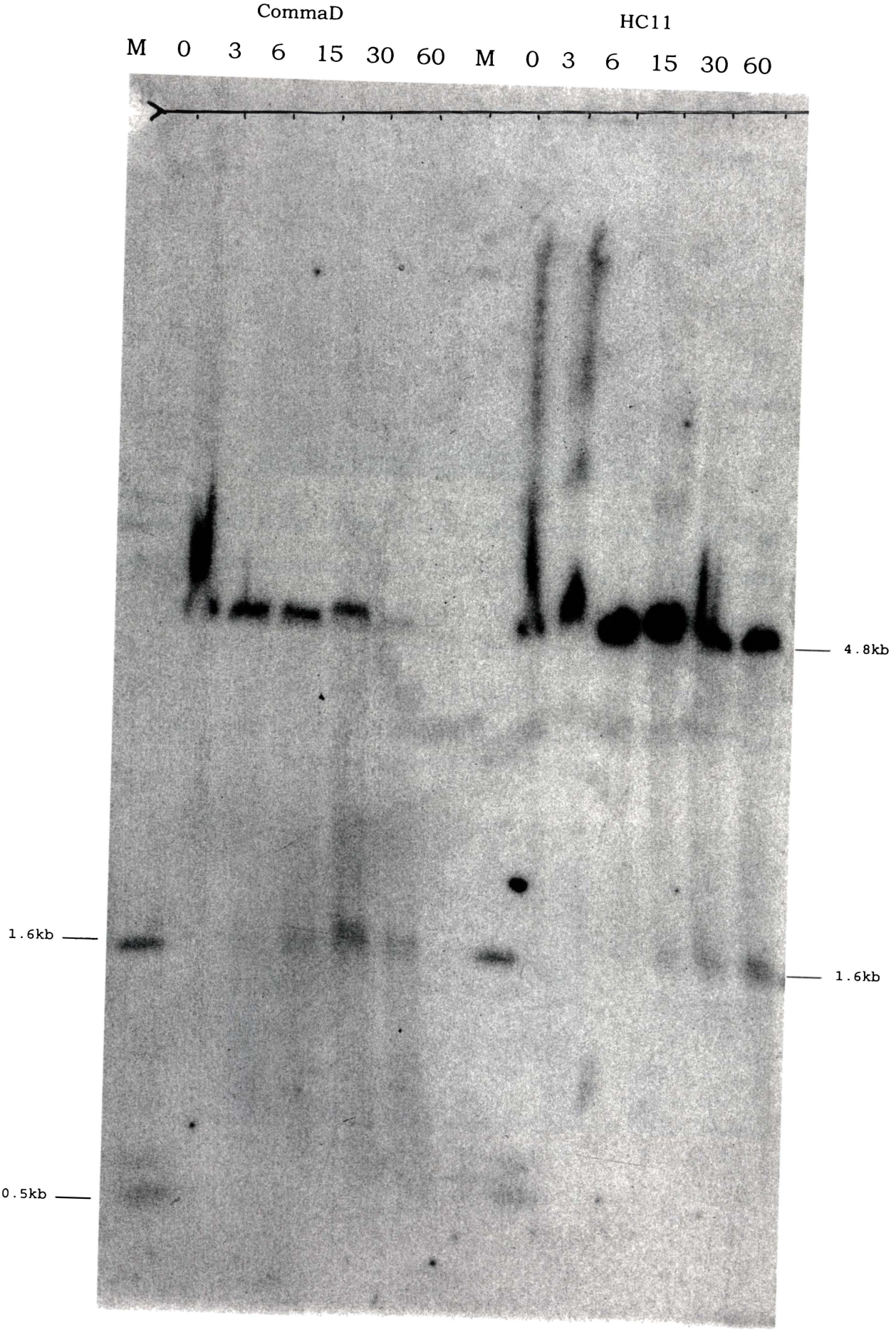


Figure 5.12 Stably transfected HC11 cells with *Ese*-2 deletion constructs (A) TLC plate visualised by autoradiography of different stably transfected cell lines, numbered below the figure. (B) Graphical analysis of the fold induction of each construct above the level of the negative construct. An experimental error occurred in the recovery of construct (4) hence it is not represented in the graph.

(A)



(B)

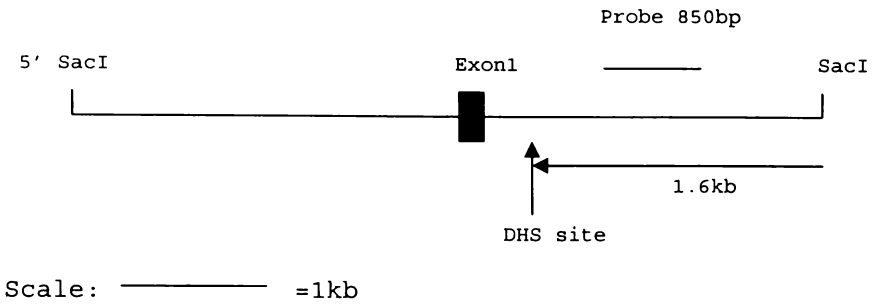


Figure 5.13 DNase hypersensitive analysis of HC11 and Comma1D DNA (A) Autoradiograph of DHS showing both Comma1D (left set of data) and HC11 DNA. The lanes are labelled above the diagram. M, 1kb marker labelled to the left of the autoradiograph; 0-60, DNase1 concentration (µg/ml). Each sample of DNase treated DNA was digested with *Sac*I restriction enzyme prior to electrophoresis. Sizes of products are shown to the right. (B) Schematic diagram showing the position of the identified DHS site in the 5' region of the first intron. The position of the probe is indicated above the sequence, exon 1 is marked by the black box.

5.3 Discussion

This chapter describes an investigation of the promoter and first intron region from the *Ese*-2 gene, for *in vitro* transcriptional regulatory activity. The proximal promoter (850bp of immediate 5' sequence) was predicted to function at a basal level of transcription whilst the distal promoter and first intron could harbour enhancer type elements. Enhancers and promoters affect transcription differently, with enhancers being responsible for increasing the probability of forming transcription complexes and hence increasing the rate of transcription. The promoter affects the half-life of the transcription complex and the rate of re-initiation of transcription (Yean and Grall, 1997). The biological importance of the promoter is therefore dependent on other such elements, including enhancers.

The use of the chloramphenicol acetyltransferase reporter system allowed a simple assay (either ELISA or TLC) to determine any *in vitro* promoter activity. The choice of the *in vitro* model was made over an *in vivo* option (transgenic mouse for example) due to time constraints and simplicity. The CAT gene was used as a reporter as no endogenous CAT activity is present in eukaryotic cells, and several methods have been described for its detection. The use of an *in vivo* model is limited in the time it may take to produce suitable animals with the desired transgene. This is reflected in the cost and examining larger numbers of constructs *in vivo* would be expensive. Once important elements have been located by an *in vitro* model, they could then be further investigated *in vivo* to get an idea of their biological relevance. As the *in vitro* system does not fully represent the biological situation, the results cannot be as accurate and many factors must be considered in such a system. The surrounding tissues and factors may play a role that cannot be replicated in a cell culture environment.

The cell system used in this investigation included two lines from mouse mid pregnant mammary gland tissue. Comma1D cells are

responsive to the lactogenic hormones and are therefore considered to be a good representation of the murine mammary gland situation (Danielson *et al.*, 1984). HC11 cell line is derived from the Comma1D cell line, also expressing the endogenous β-casein gene when induced *in vitro* (Ball *et al.*, 1988). The HC11 line has the advantage of growing well on plastic culture dishes, while the Comma1D line requires a support. Interestingly, endogenous *Ese*-2 mRNA transcripts can be detected by Northen blot analysis of total RNA in the Comma1D line, but not in the HC11 line. This may reflect a much lower level of expression in the HC11 line, as no expression was detected in other tissues such as the kidney or lung where *Ese*-2 expression has previously been seen (Zhou *et al.*, 1998). It may also reflect aberrant RNA processing, where the transcribed RNA is incorrectly processed. As seen in Figure 5.1, the Comma1D RNA expresses two transcripts of approximately 2.5kb. This is of interest as the human *ESE*-2 transcript is also detected as two bands (2.6 and 2.4kb) (Oettgen *et al.*, 1999b). These were determined to have alternate promoter usage resulting in different first exons, causing the discrepancy in size. Such a possibility may also exist in Comma1D cells, though this was not further investigated. Neither has it been reported in mouse tissues. The possibility also exists for alternate promoter usage for the *Ese*-2 gene in the mouse, as transcripts may be present in low levels and in tissues not yet examined.

Within the proximal promoter deletion constructs there was a difference in activity of approximately 3.5-fold for the mid-sized proximal constructs (1-5) in Comma1D cells over the negative construct. The average difference for construct (3) compared to the negative construct (0) was calculated to be statistically significant to the 95% confidence level. The definite increase in activity in both HC11 and Comma1D cells indicates the presence of possible regulatory elements of importance existing within this region. As seen in Chapter 3, (Figure 2.7) a GATA-1, an AP-1 element and the putative TATA box are all located in this region and could account

for the increase in activity. Similarly, the decrease in activity seen between construct (3) and (5) may be the result of negative elements acting here. The decline in activity between these two constructs is statistically significant and many potential transcription factor elements are possible candidates here.

The average of data from the proximal constructs from HC11 (Figure 5.4B) experiments shows a significant change in activity within the proximal promoter region. Construct (3) showed a 20-fold activity increase over the negative. In this context, a strong positively acting element appears to be located between construct (2) and (3). Several candidates for this element were identified by computer analysis including Ap-1, Ap-2, Sp-i/GC-box, C/EBPβ and Ets as shown in Figure 3.7. No activity was seen for constructs (1) and (2) compared to the negative. Following the sharp increase in activity of construct (3), activity appeared to drop again with the extended constructs (4) and (5). The decline in activity when the sequence was extended 200bp upstream from construct 3 would suggest the presence of a strong negatively acting element located in this region.

Analysis of each cell line here has shown some differences. Overall the scale of difference is marked, with HC11 cells line showing up to an 80 fold increase over the negative control, but only a 6 fold increase for Comma1D cell experiments. This difference in the level of activity is reflected in the remainder of the results. The difference between individual experiments may be due to the varied status of the cells on each experimental day.

Most other proximal construct results are consistent between Comma1D and HC11. Construct (2) indicates some discrepancy between HC11 and Comma1D, where it is has a higher relative activity in Comma1D. That is, construct (2) shows near equal activity to construct (3) in Comma1D cells. In this case, the level of activity is then increased further 5' at construct (3).

Overall, the results of the proximal promoter constructs expressed *in vitro* show a significant level of activity was seen over the

negative control. Whilst it is difficult to compare this level of activity to the *in vivo* biological level it may partially account for the increase seen in the biological mRNA expression as seen during pregnancy and lactation of the mouse (unpublished observations).

The promoter constructs were extended further 5' to encompass a maximum of 2.6kb 5' of the +1 transcription start site. These were divided into three separate constructs. Activity in Comma1D and HC11 cells was found to remain constant between these three constructs, with some small changes in activity seen. Comma1D cell experiments showed an approximate 2-fold increase in activity over the negative. This was above the level of construct (5), but below the high activity of construct (3). In HC11 cells, the overall induction of the distal constructs ranged between experiments but averaged around a 10-20 fold induction. The level of activity between the three distal constructs did not vary greatly on average. This suggests there are no important regulatory elements in the regions between them. However, the level of activity is increased significantly from 0.85kb (construct 5) to 1.3kb (construct 6) which indicates some factor(s) is acting to overcome the repression seen following construct 3 at 200bp. The level of induction is significant only for HC11 cells and not Comma1D. HC11 experiments H4 and H5 showed much higher levels of induction in most distal and some proximal constructs. This may be a feature of the cell condition on the day of experimentation, though the trend of results is retained.

It is known that a large first intron can harbour sites capable of influencing regulation of transcription. *Elk*-1 is another ets transcription factor that has been shown to have strong positive regulatory sequences in the first intron (Lehmann *et al.*, 1999). A similar approach was used in the current investigation to study the promoter activity, which, for *Elk*-1, was found to be increased 2-fold when the first intron was included, over that of the proximal promoter activity. In the *Elk*-1 study, the intron was only 700bp and the longest construct a total of 1400bp. This may have been a

factor in the difference in results. The cell line analysed (RK-13) may also have been a more authentic replica of the *in vivo* situation, having strong *Elk*-1 endogenous activity.

Ese-2 contains a large first intron, which was examined firstly as a complete sequence including 2.6kb of 5' sequence and the entire intron of 6kb. This analysis resulted in only a small induction in transcription within the Comma1D system. In HC11 cell culture the induction was seen to range between 1 and 16 fold. This range appeared to be due to a lack of transfection efficiency. Several different preparations were generated with little improvement in cell uptake. A likely reason for this may be the large size of the construct. Size limitations in transfection efficiency have been reported previously (Smith and Morris, 1991) (Hofman *et al.*, 2000), where an inverse relationship was found between transfection and size of construct.

One possible complication in the experiments, that was not fully appreciated at the time, concerned aberrant splicing due to inadvertent removal of donor and/or acceptor sites in constructs. Clearly, the best way of checking for such complications would be to measure the size of the RNA transcripts produced from the transfected construct. Moreover, given the large size of some constructs it is possible that various cryptic splice signals could complicate *in vitro* experiments of this nature, even in the presence of authentic splice sites.

To determine if size was a limiting factor in the transfection and hence activity of the intron, three smaller constructs were made to span 500bp of proximal promoter and 2-3kb each of the intron. Transfection was not found to increase as expected with the reduction in size of the constructs. Sequences within the intron could possible interfere with transfection; anecdotal reports suggest that cloning problems can be experienced with some sequences. Sequence analysis of this region did not reveal repeat or secondary structure that may be responsible for such difficulties. The activity was measured and found to be low overall,

with a 1-2 fold induction over the negative. Construct (12) (3' intron region) appeared to have activity approximately equal to that of the total intron (construct 9) in these experiments and may therefore contain regulatory elements. The analysis has not eliminated the intron from containing important regions as technical difficulties could not be fully overcome.

In the mammary gland, hormone levels fluctuate with the state of development. Specifically, progesterone is strongly expressed from the beginning of pregnancy and abruptly declines at parturition. Estrogen has a role linked to progesterone, though its expression increases gradually following pregnancy (Mepham, 1987; Hennighausen and Robinson, 1998;).

The influence of progesterone on female mammalian physiology is well known (Clarke and Sutherland, 1990). Both progesterone and estrogen steroid hormones are found to control growth, differentiation and function of reproductive tissues and others via the control of gene expression. They also are known to influence the growth of tumours, especially of the breast and uterine tissues. As *Ese*-2 is highly up-regulated in the mammary gland at the start of pregnancy the effect of these two hormones was examined *in vitro* on the *Ese*-2 constructs. The concentration of progesterone and estradiol was experimented with to achieve a level suitable for the number of cells, without causing cell death. After several experiments it was clear that the *in vitro* system using a range of constructs was not responding to either hormone. A possible *in vivo* relationship between the hormone and *Ese*-2 not detected *in vitro* may still be possible.

Following this analysis, the *Ese*-2 deletion constructs were examined in cell culture for a response to serum stimulation. Serum is important during cellular activities including tissue remodelling, by stimulating collagens and other proteins (Ihn *et al.*, 2001). A response to serum was observed for *Ets*-1 in HeLa cells (Majerus *et al.*, 1992) and increases collagen gene expression via Sp1 phosphorylation (Ihn *et al.*, 2001). The method for this

analysis was similar to that described by Ihn *et al.*, (2001), but did not result in any significant change in activity other than increasing general cell growth.

The transfer and activity of foreign genes into cells is dependent on a number of parameters of the system. These include the type of promoter/enhancer sequence, the level of expression and the type of expression (transient or stable). Transient expression in cells is a simpler procedure for obtaining a brief period of high expression. Stable expression is more laborious and results in an overall lower detectable level per cell transfected. The two types of expression have also been reported to differ in the regulation or expression of a transfected foreign gene (Berland and Chasin, 1988).

All transfections performed in this study were by the transient method. To ensure the accuracy of the transient method, a set of stably transfected cells was generated. The harvested and assayed cells from these lines showed a similar pattern of activity for the tested constructs, confirming the transient method was a suitable substitute in this study.

Overall, the use of the *in vitro* cell culture system may have limited the expression of the CAT reporter by the constructs compared to an *in vivo* model. Factors influencing the activity may include the transcription factors available within the cell system, signals lacking from neighbouring tissues or signalling pathways. Generally the cellular context plays an important role in the overall expression of a gene.

Many studies have been carried out where an *in vitro* system was used to gain information. Results of such studies vary greatly, between no expression being observed, as in the *Fli*-1 *ets* gene, where 600bp of 5' proximal promoter did not increase activity in erythroid cell lines (Barbeau *et al.*, 1996). Moderate levels of expression have also been observed. The β-casein gene promoter contains positive and negative elements between -221 and -170bp. A deletion of -344 to -183 results in a 50 fold increase in basal activity of the reporter gene by the rat β-casein promoter construct.

This is probably due to the disruption of an activator element (Groner and Gouilleux, 1995). Levels of expression several hundred fold greater than a negative control were seen with promoter constructs of the GABRB3 gene, associated with imprinted genetic disorders of the brain (Kirkness and Fraser, 1993).

As an alternative method for defining important regulatory regions within the *Ese*-2 gene, DNase-I hypersensitive analysis was performed. DNase-I cleaves at exposed regions on the chromatin in a higher frequency than at other locations. As transcriptionally active regions are usually more exposed during transcription processes, DHS can identify regions of importance.

Several problems were experienced with this technique, including low yield of DNA, degradation by DNases and non-specific binding of probes. Only one probe was successfully used to detect products of DHS (*Pst*I-*Dra*I 850bp). This probe was hybridised to reveal different preparations of DNA, digested with different restrictions enzymes and confirmed the location of a single DHS site. Due to the technical problems, other DHS sites cannot be ruled out. The presence of the DHS site in the 5' region of the intron does not correlate with the deletion construct results. Construct (10) (5' intron) did not show high activity, though this may not be representative of the *in vivo* situation.

Chapter 6. Final Discussion

6.1 Introduction

Gene expression is regulated by a complex combination of transcription factors binding to cis-regulatory elements in target genes. Elucidation of the underlying molecular mechanisms is necessary to better understand the complexities of transcriptional regulation.

Ets proteins make up a transcription factor family which is known to be important in the regulation of transcription events involved in cell proliferation, differentiation, development, transformation and apoptosis. Determination of the regulatory mechanisms that control these factors would assist an understanding of the regulation pattern of their downstream targets.

Neither the function nor the mode of regulation of the *Ese*-2 transcription factor is known.

This study was aimed at determining the important regulatory regions of the *Ese*-2 gene, in particular, the elements responsible for the specific temporal and spatial features of the gene. Expression of *Ese*-2 is restricted to epithelial tissues (Oettgen *et al.*, 1999b) and it displays an interesting pattern of temporal expression during the development of the mammary gland through pregnancy and lactation (unpublished observations by L'Huillier *et al.*). This temporal and spatial pattern of expression implies that an important *Ese*-2 function may be in mammary gland tissue remodelling. It follows that knowledge of the regulation of *Ese*-2 may help determine its role in such processes.

Gene expression is most often controlled at the level of transcriptional regulation (de la Brousse and McKnight, 1993) but post-transcriptional and translational levels may also contribute. However, clues to temporal and spatial control are more likely to be found at the level of gene transcription, through cis-acting DNA binding elements. To date, studies of members of the *ets*

transcription factor family have focussed on transcriptional control and indicate that a complex array of factors operate at this level.

In the twenty years since the v-*ets* gene was identified, over 30 other ETS domain genes have been isolated. The extent of homology varies greatly in regions outside the ETS domain, making comparisons between important regulatory regions of multiple *ets* genes difficult. Individually, only a few *ets* genes have been analysed with the aim of identifying important regulatory elements. Of particular relevance to the current study was the characterisation of the human *Elk*-1 promoter by Lehmann *et al.*, in 1999. *In vitro* cell culture assays in RK13 cells showed that approximately 1kb of 5' UTR sequence had a 10-fold greater activity over an empty expression vector. This reached a maximum of a 23-fold increase with 480bp of 5' sequence. Interestingly, when 700bp of 5' sequence was combined with the complete first intron (another 679bp), activity was increased a further 2-fold. This data was supported by DNase hypersensitivity analyses that revealed two hypersensitive sites in this region. One spanned the TATA box in the proximal 5'UTR and the other, a site in the first intron contained binding elements for factors Egr-1 and SRF/TCF. Little is known of the function of Elk-1 although, as it is a ternary complex factor, it is likely to play a role in effecting a response to a certain stimulus, such as in differentiating monocytes (Lehmann *et al.*, 1999). Thus, the potential for a modest degree of regulation within the first intron is shown by the *Elk*-1 study. In contrast to the Elk-1 study, the current investigation did not show a similar enhancement for *Ese*-2. Technical problems were encountered during analysis of *Ese*-2, and these may account for the overall lack of induction of the CAT activity by the promoter constructs. Conversely, the lack of activity may be a true reflection of the *in vivo* situation, with the main areas of regulation sited outside of those studied.

6.2 Preliminary *Ese*-2 promoter analysis

Initial studies into the regulatory mechanisms of gene expression must first begin with the localisation and characterisation of the promoter DNA sequence. Thus, the first step in the current study was to isolate genomic DNA spanning the first exon and intron, and 2.6kb of 5' sequence. Analysis of the nucleotide sequence did not reveal consensus TATA or CAAT boxes. However, these elements are not necessary for, and do not necessarily define, a promoter. Indeed, several other elements have been implicated as having a transcription initiating role in TATA-less promoters, including YY1 (Seto *et al.*, 1991; Usheva and Shenk, 1994) Sp1, Ap1, ATF or TEF1 (Gilinger and Alwine, 1993). Promoters without any such elements also exist, and are thought to initiate from as yet unidentified upstream elements (Novina and Roy, 1996).

In the current study, the confirmation of the location of the *Ese*-2 gene promoter came from primer extension analysis that identified multiple transcription start sites. The core promoter (approximately 50bp), sited adjacent to the transcription start site is usually involved in temporal regulation, as in the *Drosophila* Alcohol dehydrogenase gene (Adh). Adh is transcribed from two different promoters (proximal and distal) controlled by discrete initiator (Inr) elements. Promoter usage represents different stages of development (Hansen and Tjian, 1995). Interestingly, the human *ESE*-2 gene is transcribed from two core promoters to produce slightly different transcripts. This may or may not have some relationship to its temporal (or spatial) regulation. Alternate promoters have not been identified for murine *Ese*-2, though they may be active at low levels in some tissues.

The core promoter is not necessarily the only point of control for temporal or spatial expression, though these factors are most often independent of enhancer type activity.

Crawford *et al.* (1999) examined a population of TATA-less promoters of *Ldh-B* genes and concluded that these regions did

bind transcription factors *in vivo,* which effected a change in transcription rates. Sequence variation in the population was found to be due to directional selection. The variation effected a phenotypic change, resulting from a transcription rate change from the proximal promoter. They concluded, on the basis of this combined molecular and evolutionary analysis, that the biological importance of the proximal promoter is relevant in the context of the whole genome.

Sp-1 like sequences have previously been reported to affect transcription of proximal promoters by altering interactions among different transcription factors (Bigger *et al.*, 1997). Specifically, in TATA-less promoters, Sp-1 binding has a critical role in transcription initiation (Zenzie-Gregory *et al.*, 1993; Zhang *et al.*, 1994; Colgan and Manley, 1995; Block *et al.*, 1996; Chen *et al.*, 1997a). The interactions between transcription factor IID (TFIID), Sp-1 proteins and other transcription factors usually effect subtle changes in expression (Wiley *et al.*, 1992; Azizkhan *et al.*, 1993; Al-Asadi *et al.*, 1995; Yean and Grall, 1997). Much larger variations in expression (around 10 fold or complete on/off control) are caused by upstream enhancers or tissue-specific regulatory elements, and these are important in determining patterns of temporal and spatial gene expression (Guarente and Bermingham-McDonogh, 1992; Patel, 1994; Ernst and Smale, 1995a; Ernst and Smale, 1995b).

In the context of the *Ese-2* proximal promoter, the lack of a TATA box may be compensated for by the presence of Sp-1 and other sites. Indeed, studies have shown that the –25bp region is important for the initiation of transcription via binding of TFIID- whether it be a TATA-less or TATA-box containing promoter (Wiley *et al.*, 1992). The major difference is the affinity of the –25bp region for TFIID, and in TATA-less cases, the support of nearby sequence elements is necessary for the functional binding of TFIID.

The most proximal Sp-1/GC-box element in *Ese*-2 occurs at approximately -120bp, with two other Sp-1 sites found within the first 1kb of the 5' upstream region.

As for other *ets* genes, *Fli*-1 lacks typical TATA and CCAAT boxes, but contains sequences for elements such as Ap-1, Ap-2, EBS, GATA and GC-rich stretches (Barbeau *et al.*, 1996). Transcription of *Fli*-1 is initiated at multiple sites, as is predicted for a TATA-less promoter. Like *Fli*-1, *ETS*-1 lacks TATA and CCAAT elements whilst containing many Ap-1, Ap-4 and Sp-1 sites (Jorcyk *et al.*, 1991). On the other hand, human *ELF*-3 contains consensus sequences for TATA and CCAAT, and sites for Ap-1, EBS and NFκB (Oettgen *et al.*, 1999a). Interestingly, the mouse *Elf*-3 sequence shows neither the TATA nor Ap-1 sites, in contrast to human *ELF*-3.

In some cases the function of a TATA box may be replaced by a GATA motif. Originally, the GATA element was identified as an upstream cis-regulatory site in erythroid specific promoters (Evans *et al.*, 1988). To date, four GATA factors have been isolated, with distinct patterns of expression (Ho *et al.*, 1991; Lee *et al.*, 1991; Arceci *et al.*, 1993). The presence of a GATA element in the core promoter appears to play an inhibitory role in the interaction of both the basal transcription machinery and the promoter, and inhibits the initiation of transcription from the PF4 promoter (Aird *et al.*, 1994). However, upstream GATA elements are capable of functioning as transcriptional activators, and several cases of a dual interaction, dependent on the distance between the GATA motif and TATA-binding-protein (TBP) binding site, have been reported (Ohkuma *et al.*, 1990; Dostatni *et al.*, 1991; Kaufman and Rio, 1991). Aird *et al.* (1994) suggest that a GATA motif must be "sufficiently upstream of the TBP binding site to preclude steric interference of pre-initiation complex formation".

Interestingly, a GATA core motif is present in the antisense direction at approximately –30bp upstream from the major transcription start site of *Ese*-2. Whether this element (or others further upstream) are functional remains to be determined. In a

study of the *ets* gene *Fli*-1, a GATA/EBS dual element was identified in exon 1, shown to bind GATA-1 and Spi-1, and had enhancer activity (Barbeau *et al.*, 1999).

The presence of multiple c/EBPβ (CCAAT/enhancer binding protein) sites may prove important, as this factor has been reported to play a role in the development and differentiation of epithelial cells in the mammary gland. *Ese*-2 contains at least five c/EBPβ consensus sequences in the first 1kb of 5' upstream sequence, with more further upstream.

The c/EBP proteins are known to have broad regulatory functions and the isoforms (α, β, δ, ε) are expressed in overlapping patterns (Cao *et al.*, 1991; Williams *et al.*, 1991). Cell culture studies reveal that both α and β are strong regulators of cell proliferation and differentiation (Johnson and Williams, 1994). Of particular note is the developmentally regulated pattern of expression of α and β in the mammary gland. C/EBPα is most highly expressed in virgin mouse mammary gland and at a slightly lower level during pregnancy. On the other hand, c/EBPβ levels are lowest in the virgin state and increase throughout pregnancy (Robinson *et al.*, 1998). C/EBPδ mRNA expression is normally low but increases dramatically at involution (Gigliotti and DeWille, 1998) and correlates with the growth arrest of mammary epithelial cells in culture (O'Rourke *et al.*, 1997). Use of a mouse model deficient in the transcription factor c/EBPβ has indicated a critical role for c/EBPβ in both the proliferation and morphogenesis of epithelial cells in the mammary gland, and also for the differentiation of these cells to secretory cells during pregnancy (Robinson *et al.*, 1998). The *Ese*-2 mRNA temporal expression pattern is most dramatically regulated at involution with a sudden decrease in levels. As *Ese*-2 is known to be epithelial specific, with a possible role in tissue remodelling, the potential exists that one or more of the c/EBP isoforms may play a role in the regulation of *Ese*-2.

Robinson *et al.* (1998) also hypothesise that c/EBPβ is required for epithelial cell recognition of hormonal signals, and the appropriate

responses to extracellular signals (Robinson *et al.*, 1998). This observation would support a role for c/EBPβ in the regulation of *Ese-2*.

A recent publication on the transcription factor NFκB and its role in mammary epithelial apoptosis supports a relationship with *Ese-2* (Clarkson *et al.*, 2000). In the mouse mammary gland a major re-organisation of gene expression occurs during the transition from lactation to involution. This dramatic example of developmentally regulated epithelial apoptosis is correlated with changes in the expression of a number of transcription factors including Ap-1, STATs, Ets isoforms, c/EBP isoforms, NF-1 and c-myc (Strange *et al.*, 1992; Feng *et al.*, 1995; Furlong *et al.*, 1996; Li *et al.*, 1997; Gigliotti and DeWille, 1998; Neve *et al.*, 1998; Chapman *et al.*, 1999). Involution is characterised by an increase in apoptotic events in the luminal epithelial layer of the lobulo-alveolar compartment following weaning (Strange *et al.*, 1992; Lund *et al.*, 1996). This proceeds with the supporting basement membrane being proteolytically degraded and the gland remodelled to resemble its pre-pregnant state (Strange *et al.*, 1992; Lund *et al.*, 1996). Such maintenance and survival of the secretory epithelia is dependent on extracellular signals from the basement membrane, cell adhesion molecules, lactogenic hormones and other factors (Feng *et al.*, 1995; Boudreau *et al.*, 1996; Lund *et al.*, 1996; Li *et al.*, 1997; Streuli and Gilmore, 1999). Mammary gland involution is divided into two stages. First, local stimuli result in activation of STAT3 (Liu *et al.*, 1996; Philp *et al.*, 1996), Bcl-X_L and Bax (Heermeier *et al.*, 1996), down-regulation of milk protein genes and STAT5 (Liu *et al.*, 1996; Philp *et al.*, 1996), and an increase in cells undergoing apoptosis (Strange *et al.*, 1992; Lund *et al.*, 1996; Li *et al.*, 1997). Second, in the following 48 hours, increases in metalloproteinase gene expression together with the degradation of basement membrane and remodelling of the gland are observed (Strange *et al.*, 1992; Lund *et al.*, 1996).

NFκB up-regulation is coincident with the increase in apoptotic cells in the mammary gland and the decrease of lactogenic hormones (Strange *et al.*, 1992; Quarrie *et al.*, 1996; Travers *et al.*, 1996; Li *et al.*, 1997). Its expression is effectively zero during lactation, becomes evident 1 hour after suckling ceases, and reaches a maximum level at 3 days of involution. At 3 days of involution an increase in DNA-protein mobile complexes is observed, correlating to the increase in metalloproteinase activity and gland remodelling.

NFκB subunits have been found in breast tumour and transformed mammary cell lines. Due to its role in negative regulation of apoptosis, its aberrant expression may contribute to mammary epithelial carcinogenesis through incorrect regulation of apoptosis in both developing and regressing mammary gland (Dejardin *et al.*, 1995; Nakshatri *et al.*, 1997; Sovak *et al.*, 1997).

Ese-2 expression correlates negatively with both NFκB expression and the metalloproteinases involved in the degradation of mammary epithelia. Theoretically, there may be a relationship between the two transcription factors. Specifically, negative regulation of the *Ese*-2 gene by NFκB at involution may play a role. Sequence analysis has revealed the presence of two consensus NFκB sites within the proximal promoter region of *Ese*-2 and further sites within the first intron.

The combined effects of factors working cooperatively also play an important role in the regulation of gene expression. Ap-1 and Ets are assumed to interact in the regulation of gene expression, as binding sites are often juxtaposed in promoter sequence. Also, proteins are able to act synergistically at a multimerised viral enhancer site (Wasylyk *et al.*, 1990) and do so to regulate the *TIMP*-1 gene (Logan *et al.*, 1996). The *Transglutaminase*-3 (*TGM*3) gene is regulated by Ets-1 and Sp-1 proteins in cooperation (Lee *et al.*, 1996). Since the identification of sites for Ets-1 in proximity to sites for Sp-1 and Ap-1 in the *extracellular matrix protein*-1 gene

(*Ecm*1), it has been hypothesised that these factors are important for its control (Smits *et al.*, 1999).

Another example of this cooperative function of factors is documented for the 92-kDa type IV *collagenase* gene (*MMP*9), which is important for extracellular matrix degradation. A PEA3/Ets site is located 7bp upstream of an AP-1 consensus sequence and mutation of either results in near complete loss of stimulation of this promoter by ras (Gum *et al.*, 1996). These sites are not predicted to be sufficient for optimal expression of the gene with other identified sites such as Sp-1, GT-box, Ap-1 and NFκB being implicated. When any of these sites throughout the promoter (600bp) were individually mutated, they all caused loss of gene expression.

Futher, the *urokinase* promoter, which is also stimulated by ras, required the closely spaced PEA3/Ets and Ap-1 motifs for induction by the GTP binding protein (Lengyel *et al.*, 1995).

Ese-2 contains closely spaced sites for Ets and Sp-1, Ap-2, GATA, c/EBP and Ap-1. Given the wealth of data on cooperative interactions between such factors, it is likely that similar interactions are important for Ese-2, and function may relate to the matrix metalloproteinases or tissue inhibitors of matrix metalloproteinases.

6.3 Human and mouse genome and EST analysis

The publication of the draft human (and mouse) genome has given researchers another tool for studying gene structure, and provides vast quantities of information for use in the various branches of molecular biology. Not only does it pertain to the human, but it also combines information from other species. This is particularly useful in comparison studies of the genomic structure of orthologues. The release of the completed mouse genome draft sequence has further increased the depth of knowledge that can be gained.

Prior to the human genome sequence release, only the cDNA sequences of the human and mouse *Ese-2* genes were known. The present study has provided several kilobases of mouse genomic sequence, and, in Chapter 4, comparisons have been made with that of the human. This comparison identifies several small regions of strong homology between the mouse and human, while the majority of genomic DNA is highly divergent. It was hypothesised that the regions of close homology (>80%) might be important regulatory locations. However, the deletion constructs analysed in the present investigation did not reveal this to be the case. The regions of homology spanned by the constructs did not show marked levels of enhanced or depressed CAT expression. Again, this may reflect a technical failing of the system and not be representative of the *in vivo* situation.

The chromosomal location of human *ESE*-2 shows a group of adjacent genes that are similar to the syntenic block of the mouse chromosome 2 surrounding *Ese*-2.

Alternative splicing of the human *ESE*-2 gene has been reported in the form of two isoforms produced by alternative promoter usage. The *ESE*-2a isoform contains 10 more amino acids at the N-terminus than *ESE*-2b. This is thought to be species specific, as in the homologous gene of the mouse, the two transcripts that have been identified to date, contain C-terminal but not N-terminal differences. Alignment of the first untranslated exon of each human isoform with the first mouse exon (which is also untranslated), reveals sequence homology between the mouse and the 2b isoform. Interestingly, the 2a human isoform has a striking similarity to the bovine 5' UTR, suggesting an evolutionary pathway for the two isoforms. The two known isoforms of the mouse *Ese*-2 gene represent alternative polyadenylation products (Langley, 1999). Northern analysis of total RNA from the Comma1D cell line revealed two possible isoforms around the 2.5kb size that may correspond to the human isoforms and implicate alternative promoter usage.

Alternative splicing is an important mechanism involved in generating functionally distinct products from the same gene (Maniatis, 1991; Smith *et al.*, 1989). It adds another level of regulation to the expression and function of genes and is thought to be important in a number of cellular processes. Alternative splicing actually encompasses a wide range of events such as alternate polyadenylation, pre-mRNA splicing, promoter usage, transcription start site usage, exon skipping and other events. A study of alternative splicing in the human genome (using the TIGR human gene index) showed 35% of genes had some form of alternative splicing (Mironov *et al.*, 1999). This is likely to be an underestimation limited by the incompleteness of current data. The role of the two isoforms of *Ese*-2 in the human and mouse are not yet known, though differences in expression patterns suggest they will have functionally distinct roles (Zhou *et al.*, 1998; Oettgen *et al.*, 1999b).

One key feature discovered during the analysis of the murine EST database sequences was the apparent exon skipping of exon 4. Such an event could be functionally significant, resulting in a separate protein domain present or absent in certain instances. Indeed, the removal of exon 4 from the cDNA sequence would result in a frame shift, causing a quite different protein to be translated.

The skipping of murine *Ese*-2 exon 4 has not previously been reported which implies it is a rare event, or occurs in temporal or spatial regulatory patterns that have not yet been studied. However, the event occurred in almost 50% of the deposited EST sequences, suggesting that it may indeed have an important role. This has potential as a future avenue of research, with regard to the regulation of splicing and its functional significance.

Loss of heterozygosity (LOH) and other chromosomal events, such as translocations and inversions, are key factors involved in the development of many tumors and diseases. With respect to the *ets* transcription factor family, Ewings sarcoma deserves special

mention (discussed in Chapter 1). LOH is a frequent event at the locus of *ESE*-2 (chromosome 11p13-15), which spans the *WT*1 (Wilms tumor) gene at 11p15. Many tumors are known to involve rearrangements in this region of chromosome 11. *ESE*-2 is flanked by *EHF*, another *ets* factor. As *ets* genes are known proto-oncogenes, frequent LOH in this region is consistent with them having a role in tumorigenesis.

The significance of alternative splicing, especially the apparent exon skipping in *Ese*-2, may provide useful information on the gene's regulation and spatio-temporal expression. Knowledge of the involvement of *Ese*-2 in tumorigenesis and an understanding of its transformation to an oncogene may prove useful in future gene modulation therapies that focus on this chromosomal location.

6.4 Regulation of *Ese-2*

The *in vitro* cell culture experiments performed in this study resulted in a modest enhancement of activity above the negative control, showing that the promoter sequence of 200bp was active. The level of activity varied greatly between experiments, from a minimal 5-fold induction of construct (3) above the shortest promoter construct (1), to a maximum of an 80-fold induction, on a different experimental day (HC11 cells). Such discrepancies between results may be due to variations in culture conditions and epithelial cell properties on different days. When the data was averaged, it gave an approximate 20-fold induction of construct (3) above construct (1) and the negative construct (0). While the level of induction was much lower in Comma1D cells (around 5-fold), construct (3) still showed a significant level of induction above constructs (1) and (0).

Enhancement levels of distal promoter constructs also varied between experimental days, fluctuating between 5 and 50-fold levels of activity (for constructs (6) and (7)), above the negative construct in HC11 cells. Again the level of enhancement was

greatly reduced in Comma1D cells. It was hard to interpret the data due to the fluctuations between experiments, but it does appear that the *Ese*-2 promoter constructs are active and regulated in this system.

The constructs encompassing the intron and proximal 500bp of promoter sequence showed similar fluctuations between experimental days. The overall level of activity of the complete intron and promoter sequence (construct (9)) was low compared to construct (3). This may indeed be a true reflection of the regulation by the intron, though technical problems (as discussed later) were more prevalent in these experiments.

The promoter region of a gene is an arbitrary length of DNA sequence that is critical for the initiation and correct temporal and spatial expression of that gene. It usually consists of 200-1000bp 5' of the transcription initiation site, though it does vary greatly amongst genes, so as to encompass distant enhancer elements. In higher eukaryotes, genes are transcribed by a polymerase II dependent promoter, characterised by many regulatory elements where the order and spacing is critical. Intervening sequences between these elements are not necessarily conserved and appear to have no obvious function. The identification of regulatory elements based on consensus sequence alignment does not provide information as to their biological importance. To this end experimental evidence must be gained.

Confirmation of a functional promoter is most often provided by an *in vitro* system using a reporter gene driven by a segment of potential promoter sequence, transfected into an appropriate cell line. The *in vitro* system, while not ideal usually provides a rough parallel to the *in vivo* situation such that an approximate delineation of important regions can be gained. A vast number of technical factors need to be considered when designing transfer and expression experiments for studying a gene in eukaryotic cells. Many approaches have been developed, though optimisation for each study is usually necessary. The level of gene expression will

depend on the number of copies per cell and the number of cells containing the transferred gene. Other factors influencing levels of gene expression include the rate of transcription and the stability of the mRNA.

The cell culture system used in the analysis of the *Ese*-2 gene was a mouse mammary cell line derived from a mid-pregnant mouse. This line (Comma1D) expresses the beta-casein milk protein when induced by the lactogenic hormones and grown on matrigel (Danielson *et al.*, 1984). A derivative of this line (HC11), which expresses beta-casein under the same hormonal influence but has the advantage of good growth on plastic, was previously generated (Ball *et al.*, 1988). Interestingly, Comma1D was found to constitutively express endogenous *Ese*-2 mRNA while HC11 did not.

Based on the *Ese*-2 endogenous expression data, it was expected that only transfected Comma1D cells would express the construct while HC11 cells would act as a negative control. In the current experiments, this was not found to be the case, as the HC11 line actually showed higher expression of the reporter gene under the control of the *Ese*-2 promoter. The lack of endogenous expression in the HC11 line possibly could be the result of a mutated regulatory element, aberrant RNA, interference in the transcript processing or some such phenomenon. However, it is more likely that the enhanced expression of the reporter in HC11 was due to greater transfection efficiency, with a higher percentage of cells taking up the foreign DNA. It has not been investigated whether other non-mammary cell lines express endogenous *Ese*-2. Such cell lines may be of future use. Indeed, a cell line that was more receptive to transfection regardless of endogenous *Ese*-2 expression would be of great value in future research.

Other factors that limit the transfection efficiency would also influence the reporter expression. One of the most well documented factors is the size of the transfected plasmid, with small DNA molecules showing higher transfection efficiency than

those more than a few kb in sequence (Labat-Moleur *et al.*, 1996; Xu and Szoka, 1996; Darquet *et al.*, 1997; Kreiss *et al.*, 1999). Baker *et al.*, (1997) suggest that the ultimate expression of larger molecules is not the problem, rather, some intracellular phenomenon interferes with the efficiency of actual uptake of larger molecules. In keeping with this intracellular phenomenon, Kreiss *et al.*, (1999) suggest that a large DNA size may affect the mechanism of DNA release from cationic lipids or may retard the intracellular migration of DNA from the cytoplasm to the nucleus. There may also be a sieving effect on the DNA molecules during migration into the nucleus. Kreiss *et al.*, (1999) also report an effect due to the topology of the transfected DNA molecules.

There does appear to be a size effect on the uptake of the largest construct used in this study, as evidenced by a reduction in the co-transfected reporter gene activity. The smallest plasmid constructs (up to 1.5kb of inserted sequence) gave the highest transfection efficiency. The largest construct (construct (10)) showed very low levels of transfection efficiency, though an attempt to reduce the size of this construct (by generating three smaller constructs each spanning part of the intron), did not appear to lessen the effect of plasmid size on transfection efficiency. However, the size of the truncated constructs still remained at 7.5kb, much larger than the proximal constructs (largest of those was construct (5) at 5kb). There may be other factors involved, though the true activity of the intron region must first be proven.

There is much literature on the effect of construct size, and all shows that similar problems occur when larger constructs are used. For example, decreased expression of the mouse *renin* gene (total of 8.5kb) was observed in comparison to a promoter-less control (Ekker *et al.*, 1989). It was hypothesised that the reason for this was that non-specific initiation of transcription at upstream locations result in decreased CAT protein expression, simply due to the length of inserted sequences. Thus, the longer constructs resulted in more non-specific initiation. This may also be the case

for *Ese*-2, though for promoter lengths of less than 1.5kb this was not observed.

Interference has also been observed between co-transfected control plasmids (used for measuring transfection efficiency), and the test plasmid (Hofman *et al.*, 2000). Inhibition of the test plasmid appears to be inversely related to the length of the insert construct in the control plasmid, with maximal inhibition occuring with an empty expression vector. Conversely, the use of the empty expression vector for standardising transfected DNA can also be misleading, as it may result in apparent co-activation for co-transfected control constructs (Hofman *et al.*, 2000). These interference problems are thought to be due, at least in part, to the total amount of DNA transfected exceeding the transcription capacity of the cellular machinery, which results in competition between the co-tranfected plasmid and the promoter plasmid.

Again, this explanation may have some importance for the system used to study *Ese*-2, which included an empty expression vector and a co-transfected control plasmid. Indeed, the pCH110 vector (and others) have been reported to have a bias when used to standardise transfection efficiency (Farr and Roman, 1992; Bergeron *et al.*, 1995). In a series of experiments performed by Bergeron *et al.*, (1995) constructs of the *Fli*-1 gene were co-transfected with different control reporter plasmids. Significant variations were found in the control reporter expression that did not reflect differences in transfection efficiency. The mechanism involved was not studied, but was suggested to involve differences between DNA sequences within the co-transfected plasmid (Bergeron *et al.*, 1995).

The method of transfection also plays an important role in the overall expression of the promoter construct. The current study has been based on the use of two liposome reagents. Alternative methods for transfection have been developed and are constantly being researched, especially for gene therapy purposes. Alternative methods employ, amongst others, $CaPO_4$, cationic polymers,

electroporation and liposomal transfection. Another agent, chloroquine, acts by interfering with lysosomal degradation, and allows DNA to be released from an endosome into the cytoplasm (Hasan *et al.*, 1991; Wolfert and Seymour, 1998). Peptides have also been successfully used, and a variety of groups are utilised to link the DNA to the peptide, including lysine residues, endosomolytic peptides, or RGD peptides (Gottschalk *et al.*, 1995; Wilke *et al.*, 1996; Zanta *et al.*, 1999). The use of adenoviruses is one of the most efficient methods of gene delivery, particularly for long constructs. This method has also been used in conjunction with peptide delivery (Campeau *et al.*, 2001).

Limitations of the cell culture system were not the only hurdle to overcome in the present investigation. Difficulties in cloning the first intron region hindered the analysis. PCR amplification of the desired fragment was achieved after much time was spent in the optimisation of conditions. The subsequent cloning of this fragment was also difficult. Deletions and translocations of sequences often occurred and resulted in plasmid constructs that had to be discarded. The problematic sequence was narrowed down to a specific region that was repeatedly deleted, though no features known to cause this such as repeat sequences or secondary structure were identified.

From the *in vitro* cell culture analysis of the promoter deletion and the first intron constructs of *Ese*-2, several conclusions can be drawn regarding important regulatory regions. The most significant was the increased activity between the 70 and 200bp upstream region. A 20-30 fold enhancement was detected in this region, though the element(s) responsible was not determined. This would fit with the theory of the proximal promoter being responsible for initiation of the transcription complex and its longevity. The increase in activity was somewhat lower than hypothesised when compared with the biological activity seen in the pregnant mouse mammary gland. The identification of a region containing important regulatory sequences capable of such temporal and

spatial activity was not found in this study. The reasons for this may be the insensitivity and fluctuations in the *in vitro* system, which made it impossible to consistently pick up small increases (2-fold or less) in activity.

The possibility of alternate promoter involvement in the mouse mammary gland regulation of *Ese*-2 has been discussed. However, if a promoter did exist in the first intron of the mouse *Ese*-2 gene, equivalent to the human *ESE*-2a promoter, then transcription of the constructs in the cell system would be altered. In this case, the construct (9) CAT activity could well be interrupted by the use of an out of frame start site. Transcripts could code for an early termination codon, or simply exhaust the cellular machinery. Such a phenomenon would result in low CAT expression.

The level of activity seen here may possibly only reflect the promoter responding to the basal cell machinery and not the enhancing effects of mammary specific factors. Although both Comma1D and HC11 are mammary derived cells it is possible, as discussed earlier, that they respond no better to *Ese*-2 constructs than would a range of other cultured cell types. Moreover, cellular interactions with other cells, tissues, or the extracellular environment are likely to play an important role in the expression of *Ese*-2. Such factors have not been studied in this investigation, as a somewhat more sophisticated *in vitro* system, for example involving an extracellular matrix such as Matrigel, would be required. This limitation may prove critical in the determination of important regulatory mechanisms for *Ese*-2. The use of an *in vivo* model would provide useful information regarding the role of the cellular environment.

The *in vitro* regulatory activity of *Ese*-2 was only studied at the transcriptional level in this investigation. As yet, the regulation occurring at other levels is undetermined and should be investigated in future research. Nuclear run-on assays, for example, could be performed to determine if mRNA accumulation in the mammary gland during pregnancy is a direct result of high

transcription rates, or reflects a high level of mRNA stability. Multiple levels of regulation may be important to provide the necessary spatial and temporal control of *Ese*-2.

In the current study, the constructs were tested for a response to progesterone, estradiol and serum but showed no change in activity. However, in reporting these results, it must be noted that endogenous *Ese*-2 activity was not studied under these conditions. Serum is an important factor influencing the growth of cells during processes such as injury, wound repair and tissue-remodelling. It stimulates collagen production and other such proteins important in these processes (Narayanan and Page, 1987). As previously discussed, collagen, other MMPs and their inhibitors (TIMPS) may be downstream targets of *ets* genes. Indeed *in vitro,* serum has been shown to stimulate the human *ETS*-1 promoter, potentially mediated by AP-1 (Majerus *et al.*, 1992).

Indeed, a study by Lyer *et al.*, (1999) of primary fibroblasts in culture under serum starvation and stimulatory conditions, revealed serum dependent expression of many genes known to be involved with the extracellular matrix and wound repair (Lyer *et al.*, 1999). The transcriptional response to serum stimulation by genes involved in processes such as proliferation, angiogenesis, tissue remodelling and other processes was striking. Although *ets* genes were not studied by Lyer *et al.*, it may be reasonable to expect that their expression would also be increased in a similar manner, due to their known involvement in the processes described above. Based on those reports, the serum responsiveness of the *Ese*-2 promoter elements was tested. No response was observed, though serum responsive elements may not have been present in the region studied. A study of the endogenous mRNA *Ese*-2 level under conditions of serum starvation and activation may be useful in the future.

The basis for examining the responsiveness of *Ese*-2 constructs to progesterone and estradiol was the observation that similar temporal patterns of elevated expression of these hormones and

Ese-2 occurred in the mammary gland. Additionally, the hormones are known to effect epithelial development of the mammary gland, a role that may also involve *Ese*-2 (Dulbecco, 1986).

The pattern of expression of *Ese*-2 does not appear to correlate with other known mammary gland factors such as STAT5 or lactogenic hormones though this does not preclude them from having some role in regulating *Ese*-2.

Stable, rather than transient, expression of a reporter gene in cell culture is known to influence the level of expression (Berland and Chasin, 1988). To this end, a set of stably transfected cell lines with various *Ese*-2 constructs was produced for assay in parallel with transiently transfected constructs. In the constructs studied over a set of repeated experiments no significant difference in expression was observed between stable and transient transfection.

Several researchers have investigated similar deletion constructs of promoters of other *ets* genes in cell culture. *Elk*-1 in particular shows a modest (2-fold), but significant, intronic regulatory activity above that of the proximal promoter alone (Lehmann *et al.*, 1999). Oettgen *et al.*, (1999) showed that *Elf*-3/*Ese*-1, which is closely related to *Ese*-2, has a strong promoter activity in two epithelial lines using 1.5kb of promoter. The main promoter activity gave, depending on the cell line, 200 or 700-fold stimulation of transcription above an empty expression vector control (Oettgen *et al.*, 1999a). The elements responsible for this activity were not determined. In contrast to *Elf*-3, the *Fli*-1 promoter gave no detectable activity with 600bp of the mouse 5' flanking region, exon1 or the 5' end of intron 1. Similarly, 1.5kb of the human *Fli*-1 5' flanking region was also inactive in transfected cells. These examples demonstrate that there must be several modes of regulation for *ets* genes.

6.5 Future Work

An important step in future *Ese*-2 research would be *in situ* analysis. Such studies would provide information on the exact location of expression within specific tissues and throughout development and other processes. Further, immunohistochemistry would show the distribution of the *Ese*-2 protein product. Similarly, the use of Western blots could provide a useful tool for protein analysis. The use of antibodies has been limited in such investigations by cross-reactivity amongst multiple ETS proteins (unpublished observations, L'Huillier *et al.*), and a specific monoclonal, rather than polyclonal, reagent would need to be produced.

Further research into the transcriptional regulation of the *Ese*-2 promoter could be undertaken using a different approach. The *in vitro* cell culture system using Comma1D and HC11 cells has been exhausted as a possibility, though other cell lines expressing high endogenous levels of *Ese*-2 may represent the *in vivo* situation more closely. The research should not be limited to such lines only, as *Ese*-2 mRNA expression has been detected in other tissues (Zhou *et al.*, 1998; Oettgen *et al.*, 1999). Further, a cell line that is easily transfected may be equally as informative as using a poorly transfecting cell line that, theoretically, is expected to express Ese-2 at high levels. The use of a cell line with an invasive phenotype might also be an avenue for future study, as *Ese*-2 expression is likely to be increased under such conditions. Alternatives to the cell culture *in vitro* system include the use of an animal model such as mouse. Indeed, considering the time spent in obtaining the cell culture *in vitro* data, it would have been reasonable to utilise a mouse model! The use of deletion constructs in a transgenic mouse would potentially provide more representative results on the region of regulation. The use of a mouse model would allow different stages of development to be studied as well as multiple tissues. As *Ese*-2 is expressed in many tissues containing epithelial cells,

information could be obtained regarding its regulation in potentially different cellular processes involved in the function of the kidney, lung and mammary gland for example. The mammary gland model system studied in this investigation, represented by cell lines derived from this tissue, could be analysed at the various stages of development at which *Ese*-2 expression is known to be elevated rather than depressed. These include early pregnancy where it is up-regulated and early involution where it is down-regulated.

An alternative and promising method of analysis involves DNase 1 hypersensitive (DHS) mapping. Preliminary analysis identified a possible site of interest within the first intron. More intensive investigations using alternative probes to detect other hypersensitive regions in the gene would be worth pursuing. DHS has successfully been used in many other studies as an approach to identify important regulatory regions (Luo *et al.*, 1999; McArthur *et al.*, 2001; Li *et al.*, 2002). It is often followed by protein binding studies to identify the specific factors involved.

Of particular advance to the current study is the fact that DHS sites were discovered in the *cystic fibrosis transmembrane conductance regulator* gene (*CFTR*), which also has a tight temporal and spatial pattern of expression. The hypersensitive sites of importance were found to be in extreme and diverse locations, including introns 1, 20 and 21, all of which had enhancer activity (Phylactides *et al.*, 2002). Another study, that of the mouse *monocyte chemoattractant protein*-1 (*MCP*-1) by DHS, identified sites between –24kb and +13kb (Wagner *et al.*, 2001). A comprehensive chromatin analysis of *Ese*-2 would also likely reveal regulatory sites at greater distances.

6.6 Potential function of *Ese*-2

This investigation has focussed on the regulation of *Ese*-2 in mammary epithelial cells in culture. Observations show that *Ese*-2

mRNA expression is up-regulated from early pregnancy and throughout lactation and then declines in involution, though this was not conclusively demonstrated to be a result of transcriptional regulation (L'Huillier *et al.*, unpublished observation). As tissue remodelling of the mammary gland is an important process for the correct development of the gland through these stages, it seems likely that *Ese*-2 may have a role in this.

Ets genes are known to play roles in many diverse cellular processes, including cell proliferation, differentiation, apoptosis, development, hematopoiesis, tissue remodelling and angiogenesis (Sementchenko and Watson, 2000). The variety of functions performed by this group of transcription factors suggests that Ese-2 functions will not be restricted to the mammary gland. Its expression has been detected in many epithelial containing tissues and, as such, is likely to play a role in other processes. Knockout studies of *ets* genes such as *PU.*1, *Ets*-1, *Spi*-B, *Fli*-1, and *Tel* all show phenotypes related to defective hematopoiesis (Bories *et al.*, 1995; Muthusamy *et al.*, 1995; Iwama *et al.*, 1998; Scott *et al.*, 1997; Su *et al.*, 1997; Wang *et al.*, 1997; Hart *et al.*, 2000; Spyropoulos *et al.*, 2000). Many *ets* genes also have a role in angiogenesis, in particular Tel-1, in which a null mutation in mice resulted in embryonic lethality with the histology revealing vascular defects (Wang *et al.*, 1997). Other studies have also shown *ets* genes to be involved in development, including an *Ets*-2 mutant mouse with defects in embryonic trophectoderm and a knockout *Ets*-2 mouse that is embryonic lethal at day 8.5 (Yamamoto *et al.*, 1998). Distinct from other mouse *ets* knockouts, the *Er*81 mouse has a phenotype which indicates neurogenesis defects (Arber *et al.*, 2000). A large number of the *ets* genes that have been studied by gene disruption, appear to have distinct roles in the regulation of hematopoiesis. *PU.*1 for example, has a strong effect on the development of multiple hematopoietic lineages, *Fli*-1 has most impact on the development of specialised compartments within

myeloid lineages, *Ets*-1 within lymphoid lineages and *Spi*-B appears only to affect mature B cell function (Bartel *et al.*, 2000).

In the case of *Ese*-2, if it has a role early in development, as seen with most other *ets* genes studied thus far, the potential roles in epithelial tissues in later processes including the remodelling of the mammary gland might only be analysable using conditional tissue specific knockouts.

Ets genes are widely involved in tissue remodelling, whether it be in wound repair, bone and cartilage formation or mammary gland development. Crucial to this process is the extracellular matrix (ECM), which creates a physical barrier that must be penetrated for terminal end bud formation and lateral branching (Williams and Daniel, 1983). The ECM is maintained by the interaction of matrix metalloproteinases (MMPs) and tissue inhibitors of matrix metalloproteinases (TIMPS)(Gomez *et al.*, 1997). TIMPS are involved in the regulation of MMPs and thus are implicated in regulating the development of the mammary gland and the ECM. Fata *et al.*, (1999) propose that TIMPs may control the rapid physical expansion of mammary ducts by limiting matrix degradation. TIMPS may limit cellular proliferation, which may be the basis for their ability to suppress tumor growth (Khokha, 1994; Montgomery *et al.*, 1994; Kruger *et al.*, 1997). MMPs are often found up-regulated in epithelial cancers and are considered to be essential in tumor angiogenesis, invasion and metastasis as they can efficiently degrade the ECM, which is a barrier to tumor growth (Coussens and Werb, 1996).

Some *MMP* and *TIMP* genes have been shown to be Ets responsive. Thus, *MMP*-1 (interstitial *collagenase*) gene expression is up-regulated *in vitro* by ETS-1 (Westermarck *et al.*, 1997; Behrens *et al.*, 2001) and *MMP*-1, -3, and -9 are activated by E1A-F (Kaya *et al.*, 1996), ETS-1, ETS-2 and PEA3 (Buttice *et al.*, 1996; Gum *et al.*, 1996; Westermarck *et al.*, 1997). *Urokinase type plasminogen activator* is another *MMP* also transactivated by ETS-1, and is linked to lung cancer invasion (Bolon *et al.*, 1995). The *TIMP*-1 gene

contains transcriptionally responsive elements in both its promoter and first intron regions that bind Sp-1, Sp-3 and an Ets-related factor (Dean *et al.*, 2000). PEA3 also has a role in basal expression of *TIMP*-1 (Clark *et al.*, 1997).

All *TIMP* genes (1-4) are expressed predominantly in luminal epithelial cells (Fata *et al.*, 1999). The pattern of expression of *Ese*-2 fits well with *TIMP*-1, which has expression limited to high epithelial cell proliferation. Indeed, the yet un-identified Ets factor that is involved in *TIMP*-1 regulation may be Ese-2 or a closely related family member. Thus, a very profitable area for future research could be the interaction between Ese-2 proteins and matrix metalloproteinase genes.

Appendices

Appendix 1.

Sequence of 5' *Sac*I 4.9kb

Sequence begins at -2665. Exon 1 is bold and underlined.

GAGCTCAGTTCTTCCCCCACCACCCCTGGTTCCTGCCCTTCCCACTTNCA -2616
CAGCTCATCTTCACCCACAGGCATAGTGAAGCTTTCCCATAGCTTAGTGT -2566
TTGAGATCATCCAAACATCCATTCTAATCAAATTAAGATCAAATGTGATG -2516
CTTACGCTTAGAGTGCAAGTTCTCTTTTGTTTTTGTTTTATATTAATCAT -2466
CCCTTTTAACGACTGGTCCAGAAGCCTCTAATTCCCCAGGCAGTCACCCA -2416
GAGCTGCTACATTATGTACTCAGGACCACAAAGCAAAACAGGACCACTGT -2366
CATCCACATCGGCTGGTAGATTACTTTCCACTCTCAAACTCGGGTCATAG -2316
AAAAAAATGCAAATTCTCCAAAACTGCTTTGGTTTTCTAAAAACATATTG -2266
TTTTCTAAAGACAGTAGCAGAATAGCCCGTTTTTCTTGGACACTTTTCCT -2216
TGGGAAGACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGCTCGCGCGC -2166
ACGTGTGCGCACGTGCGTGTGGTTCAGAAACTGGTAACGAGGTGGGACAA -2116
AGCTCAGCAGCAGAGAAAATCATGTGCAGGACAGAGGAGTCCCCCAAAAA -2066
ACAGGCTTCAACAGGACAAGTCACAGATGTTAAAACTTCTGTGAGGACAT -2016
CTGAATTCGGTTATTTCCAGATTCTGGTCCTGTGCAGATACTTTTGTTCA -1966
GACGAAATTTTACCAAGAAGATTTAGAGAAAGCTTTGGGGCTCTTCTTGA -1916
ATCCTCTTCTAAACTCTCGACCTTCAAACAGGACACTTTTTAAACTCCCG -1866
CCTAAGGGCTGGAGAGATGGTGCTGTGGATAAAGCACTTCCCTCCCCAAG -1816
CCTTGAGTTCAGTCAATCTCTAGAACCCAAATTAAAGAAAAAGCCAGGTG -1766
TAGCAGGGCAGGCCTGCAATCCCAGGACTAGAGAGGCAGAGTCAGGAGGA -1716
GCCTGGAAGCTTGTAACCCAGCTAGTGTAGCTAAATCAGGCTCAGCAATA -1666
GACTGTGCCAAAAGCAAAGAGGAGAGGCTGGAGAGATGGCTCAAAGTTTG -1616
CTGCTCTTGCAGAGGACCCAGGTTCAATTCCCAGCACCCACGTGGCAGCT -1566
CACAACCGTCTGTAACTCCAGTTCCAGAGGATCTGACCCTCTCACATACA -1516
GGATACAAGTAGGCAGAATACCAGTATGCATTAAAAAAAAAAAACAAAAA -1466
ACAAAACATGTTTTAAAAATGCCTAGTGATACAGGGTCTCATTGTGTGGA -1416
TCACCCTGGGTAGCCTAAAACTCACTGTGTAAATCAACCTGGCCTCAAAC -1366
CCACAGATATCTGCCGCCTGAGTGTTGGGATTAAAGGCACTGTTAAATCC -1316
AACAAATTGATTAAATAAATAAATNAATAACTTAATCCCGTTTTCTCCTA -1266
TTCTGTTGGGAGAATTCTGACCCAGAAAAGCTGGCTCTCCCACCTAAGGT -1216
CCCAGAGCCAAGCCAGAGCCTCAACAGGCCACTTATCTTGCCCCGACTAT -1166
TTTTCCATGATGACACATGACTCCTCACATGAACTGCCTTTTGTTTCTTC -1116
ACGCCGTTCGGGTGTTTGGATTTGAGTGTTTGTGCCTATGCCCAGTGTTC -1066
CCAAGAAAACCTAACCAACCAGTCAATCTCTGCTTTCCACTTCAAGATCA -1016
CAGAGTCTGCAGTTTGAGCACCTTGGTGTCTTGGCTGCCTGAGATTGAGA -966
GAGGAACGGAACCCACGAAAGGGGGTTATGAACACTCCTCCCAGATGACT -916
GCCTAGGCCGTCATTTAGATTTGAGAGATGTCCTAAAAACTGGGTAGTTC -866
TCAGTGAGACAGCTGACAGCTTGCAAGCCAATCAGGACCCCCCGCCCCCG -816
TGTTGTTCTGGATGTTTGACACTGCAGCTCCACGCTCCCTTCATATCGCC -766
ATTTCTAAAACAGAGGACACTGGGAAAGGGGGAAAGCTGTCACATGGCTC -716
CTGGTATTCAAACATGGCGCCCTTTGGGTGCTTAGAGTGGAAAGGCCAGT -666
GAAGCACTGCTCCCTGACTCTCTCCCTGCCCCCTGATCGGAAGGTCCCCA -616
CCAGGATCAATAGAAGGAAATATGTAGCGCCTAAGGACAGGCTTCCAGGT -566
GATTGGCTGACATTACCAAGTATGGCCTTCCCAGCAACCCAAGGACAGAT -516
TCCATTATTGCATTAATCTGATGGATCCCGGATATGCTCATTAACTTATT -466
CACAGTAATGCAGCCAGAAAATGACCGGCTCAGAATTTGATCCCTGAAGC -416
ACCTTTATTCTTTACCACTTCTTGGCACTGCTCTTCTTTCCTAACACAAA -366
CTACATACAGAGGTTCGGGACTTGGCCAGCCCAGGCAAAGGNTNCAATGA -316
ACAGACACCAGGCCTGAATTTCCTTCCCTCCNGCTTGCTCAACGGAGACT -266
GCTCCAGCCAGCACTCCACNTATCCCCCTNGGGGCAAGAATTCTCTGCTC -216
ATTTCCTGGGTCCCCTCGGTGGGTGGNGGGAGCTGGGCACAAAAGCAGGA -166
GAAAGGTAAACTTTCTGCATGTGAAAACCACCCCCCACCCCCGAGGAGCC -116
GTGTCACACCGTATGTCACCGTCATCAAAGGGGCTGTGCATAAACCTGAA -66
AAACCAAACGGACCTGTCTGTAGGTGTCACTTATATCACACGGCTACAGG -16
TGCCTTTATTTCTAC**AGTCCGCTGGTGCTGGGAGCGCGCTTGCCTTCTCT** +35
TGCCTTGAAAGCCTTCTGTCTGGACCTAGCCACCACTTGTCTTCACGTAA +85

GCAGCTTTTCTTTGCTCTTTTGAATGAATTGAAAGGGGTGTCGGGGCTCA +135
GAGACTTAGCTGGGGCCTGGGACTGAGTCTCTGTTTTCAGCTTCCACGGC +185
AGCCCCTCCGTGAGGAAGCCAGAGTGCCTCCTCCGGGGTGTTGACTCAGG +235
TTTGACGNTTGCGGAGAGATTTTTTTTTGTCTCTCCGGAAGGGGCACCCA +285
CCCTNTTACCGGCCCCACTTCTNTGGTCACTAACAAGCAGNTGAGAGAAT +335
GTGTACACGGCGCNTAAATCTCTGCNCATCCTTTAGCTGCTTTAAGGAGG +385
GAGCCGGATTCTGCACGGCCCCCAACCCTGGCTTCACCCTTCAGTTTTTT +435
GAAGGGCNTGCTCACCAAGCTGCCCTTNTTGATACAGAGAATGGTATTCT +485
TACTGTTTAAGGCTTTTTTTGTTTTTATTTTTGTTTTTTTTAAGGAGTTTG +535
TACAAAATAGTTTAAAGGAATCTCATTTGGCCCTCCCACAGCCCCTTAGG +585
ACAGAGTGCTTCCAGAAGAGAAGTGGCTAGGGTTTTTTTCCACTGTATTC +635
TCAAACCTTATCAGTCTGTATGCGTTCAGGGGTTTGTGCGCATTTGCCCC +685
CTACTGGCAGCAACTGGAAACACATGCTCTCCCGCATCTGCGTTAAGGTA +735
GGAGAATTTAAGACAGAATAGGGGTCCTAGATCATTATGCACTTCGTCTC +785
CATTTTACAGATGGGGCAACTGAGGCACAAGAAGGCCCGGTAATTTATAG +835
AGATGGTTTTGTGTGGTAGTCCAAAGACTCAGACCCAGGCTGTTACACGT +885
GCTCATCCATACGGAAAGCTTCATCCTCCTGCCTTCATTCTCTCGGCAAG +935
GGTCTTTCTATGATTCAGTGTATTGGGGACAGCAGCTTTTACAGTTTGGT +985
TGCCTCCTCTTCTCAAAAGTCTTCAGAGTGAAACTTTTTAGCTACTCAAT +1035
CAATGTTCACTGATTTTATTAAATCCTCAGGACGCTCAGCGGACTGTGTG +1085
TTGAGATGTCCATTTTAAGGATTCGGAAGTAGGTTATCAGAGATGGTCAT +1135
GAGTACTTACAGATTAGCAGACCAGGATGGGGAAAAAAATCAAAACTTTT +1185
AGCACGCCTGCAAGAAAGTCAGAGGTTGCAGCAGGCTAGGCAGGAATTGT +1235
GCCCTCCGTCGCTTGCTTTGTGACTACCGCAAACAACACTCATTCTCTGC +1285
GTCTTTAACCTAAAACTTGTGATCTTCCGATCGAAGCCCACAGCTTCCTT +1335
CTATGATGTTTTAGAAAAATATAGCTTAAAAAAAAAAAAACCNAACAGATG +1385
AAGTCCTAAATGAGCTGAAGAAAAATGAAGTATTGCTCAATAAGGCACTG +1435
GCGAGTTGGCTCCCACAATGCAGTGGTTGCTTTCAGAACAAATACTCAAG +1485
AACCCAACCCACAAGCCCTCCACTGCTGCAAGTGTTGAAACCCCACTGAG +1535
CCCTCTGCAGCCACCCTTCTTAAAGAGCCCGAAGGGTGGGAAAGTAACAC +1585
CCCCTATCCGTGAGGCTCTGCCGCAGTGGGCATTCAGGTGTCGTGGAGCG +1635
TTAGGAAAGAAAAAGTGTTTCTCTCCTCAACGGCACTGCCAAGATGATAA +1685
AACTGTGGGAGGATTTTCACTGTCTCACTTTCCTCATTTTGCACTGGTTT +1735
GGCTTAGCTCAAAGACGTTTTGACAGAGAAATAGCCAATGCCTGTGAACA +1785
GACATTCCGAAGATAGAAACAGGAAGCTTGCCTCAAGTTTGAGGCTAGCC +1835
TGATCTGCACTGTAGGTCCATACATGAGAAAAGGTGAAGGGGAGGAAAGA +1885
AATGGAAAGGAAACAGGGGAAATAGCCCAAGTATAGATGCTTAGTTGGGT +1935
AGATATTTGGTCACATCCCCAAAATGACCTTGAGGATCTCGGCTTGGGTC +1985
AAAGGAAGAATTTTCCTCCTCTGACAAATTGGGCAAGGCTGGGTAGTTTG +2035
CACCGTCACCATTATTTCTGCCATTGGTTTGTAAGTGGGACAAAAGCCTA +2085
TAGCCCTGGCTCTTTGGCTGGGTGGGCCCTATTGTCTGCGTGGTATGTGA +2135
GCTTGCTCAGAAAGGAGATAAAAAGTTGCTGGAGGCTCTCAGTTCCAGGC +2185
CTGCCCTGACTCACTCCAAAGGTACCTATAATCGACCTTCTCACCAGTAA +2235
AAAAAAGAGCTC

Appendix 2.

Sequence of *Ese*-2 encompassing exon 1, intron 1 and exon 2.

Exon 1 and 2 underlined in bold type.

AGTCCGCTGGTGCTGGGAGCGCGCTTGCCTTCTCTTGCCTTGAAAGCCTT +50
CTGTCTGGACCTAGCCACCACTTGTCTTCACGTAAGCAGCTTTTCTTTGC +100
TCTTTTGAATGAATTGAAAGGGGTGTCGGGGCTCAGAGACTTAGCTGGGG +150
CCTGGGACTGAGTCTCTGTTTTCAGCTTCCACGGCAGCCCCTCCGTGAGG +200
AAGCCAGAGTGCCTCCTCCGGGGTGTTGACTCAGGTTTGACGNTTGCGGA +250
GAGATTTTTTTTTGTCTCTCCGGAAGGGGCACCCACCCTNTTACCGGCCC +300
CACTTCTNTGGTCACTAACAAGCAGNTGAGAGAATGTGTACACGGCGCNT +350
AAATCTCTGCNCATCCTTTAGCTGCTTTAAGGAGGGAGCCGGATTCTGCA +400
CGGCCCCCAACCCTGGCTTCACCCTTCAGTTTTTTGAAGGGCNTGCTCAC +450
CAAGCTGCCCTTNTTGATACAGAGAATGGTATTCTTACTGTTTAAGGCTT +500
TTTTTGTTTTTATTTTTGTTTTTTTAAGGAGTTTGTACAAAATAGTTTAA +550
AGGAATCTCATTTGGCCCTCCCACAGCCCCTTAGGACAGAGTGCTTCCAG +600
AAGAGAAGTGGCTAGGGTTTTTTTCCACTGTATTCTCAAACCTTATCAGT +650
CTGTATGCGTTCAGGGGTTTGTGCGCATTTGCCCCCTACTGGCAGCAACT +700
GGAAACACATGCTCTCCCGCATCTGCGTTAAGGTAGGAGAATTTAAGACA +750
GAATAGGGGTCCTAGATCATTATGCACTTCGTCTCCATTTTACAGATGGG +800
GCAACTGAGGCACAAGAAGGCCCGGTAATTTATAGAGATGGTTTTGTGTG +850
GTAGTCCAAAGACTCAGACCCAGGCTGTTACACGTGCTCATCCATACGGA +900
AAGCTTCATCCTCCTGCCTTCATTCTCTCGGCAAGGGTCTTTCTATGATT +950
CAGTGTATTGGGGACAGCAGCTTTTACAGTTTGGTTGCCTCCTCTTCTCA +1000
AAAGTCTTCAGAGTGAAACTTTTTAGCTACTCAATCAATGTTCACTGATT +1050
TTATTAAATCCTCAGGACGCTCAGCGGACTGTGTGTTGAGATGTCCATTT +1100
TAAGGATTCGGAAGTAGGTTATCAGAGATGGTCATGAGTACTTACAGATT +1150
AGCAGACCAGGATGGGGAAAAAAATCAAAACTTTTAGCACGCCTGCAAGA +1200
AAGTCAGAGGTTGCAGCAGGCTAGGCAGGAATTGTGCCCTCCGTCGCTTG +1250
CTTTGTGACTACCGCAAACAACACTCATTCTCTGCGTCTTTAACCTAAAA +1300
CTTGTGATCTTCCGATCGAAGCCCACAGCTTCCTTCTATGATGTTTTAGA +1350
AAAATATAGCTTAAAAAAAAAAAACCNAACAGATGAAGTCCTAAATGAGC +1400
TGAAGAAAAATGAAGTATTGCTCAATAAGGCACTGGCGAGTTGGCTCCCA +1450
CAATGCAGTGGTTGCTTTCAGAACAAATACTCAAGAACCCAACCCACAAG +1500
CCCTCCACTGCTGCAAGTGTTGAAACCCCACTGAGCCCTCTGCAGCCACC +1550
CTTCTTAAAGAGCCCGAAGGGTGGGAAAGTAACACCCCCTATCCGTGAGG +1600
CTCTGCCGCAGTGGGCATTCAGGTGTCGTGGAGCGTTAGGAAAGAAAAAG +1650
TGTTTCTCTCCTCAACGGCACTGCCAAGATGATAAAACTGTGGGAGGATT +1700
TTCACTGTCTCACTTTCCTCATTTTGCACTGGTTTGGCTTAGCTCAAAGA +1750
CGTTTTGACAGAGAAATAGCCAATGCCTGTGAACAGACATTCCGAAGATA +1800
GAAACAGGAAGCTTGCCTCAAGTTTGAGGCTAGCCTGATCTGCACTGTAG +1850
GTCCATACATGAGAAAAGGTGAAGGGGAGGAAAGAAATGGAAAGGAAACA +1900
GGGGAAATAGCCCAAGTATAGATGCTTAGTTGGGTAGATATTTGGTCACA +1950
TCCCCAAAATGACCTTGAGGATCTCGGCTTGGGTCAAAGGAAGAATTTTC +2000
CTCCTCTGACAAATTGGGCAAGGCTGGGTAGTTTGCACCGTCACCATTAT +2050
TTCTGCCATTGGTTTGTAAGTGGGACAAAAGCCTATAGCCCTGGCTCTTT +2100
GGCTGGGTGGGCCCTATTGTCTGCGTGGTATGTGAGCTTGCTCAGAAAGG +2150
AGATAAAAAGTTGCTGGAGGCTCTCAGTTCCAGGCCTGCCCTGACTCACT +2200
CCAAAGGTACCTATAATCGACCTTCTCACCAGTAAAAAAAAGAGCTCATT +2250
TGATTCCACCACCTTAGGGTGATCAGACCAATCTTTTTCGTACCGATCAT +2300
GCCTTCCCTATGAGGTGGCCAGGGTGATAATTAATCTGGGGGAAACTGCT +2350
TCCATCTGTTGTTGGTCTATTGAAAGCCAGGGCATGGGAGTATAGGATCC +2400
CAGCTCAGTTGTCCAGCTGAGCAGGAGAGGGAAGGAGGAGGAGGGGAGCA +2450
CAGGACTTTCTGGGGCTACAGGTTTAATCAGAGGTCGTACTGGTTATATT +2500
TTGAGTATTTGGTGACTTTACCTGACCAAAGTGCAAATGTTATCAATGTG +2550
GCTACACGTGTGTACATGAGTACAGGTTTGGAAGCGTTTGGCTGATCTAA +2600
TGAGAGGCTGCAGAAAAATCACAGAGCCAGGGCTGACCTTGAATGCGTCT +2650
CCTAGTGTTTTCCTGTCATGGGTGAAGAAGTGAGGTTTAGAGAGATGAAA +2700
GAGGCAGGATCATCTCTCCATGTTCAGGGCTCTTTCCATAATGACAAGTC +2750

AGAGGATAGCTAGACTAGCCGGCCCAGGAACCTTCTAATATCCCTCTTTT +2800
GGTTCTGTGCACTGCTAGGCCTACCTAACCTGGTTCTTTTCTGCCTTGTT +2850
TTCATTATCAACATACACACACACACACACACACACACACACACACACAC +2900
ACACACACACACACACAAACCAGAGCCTTCTGCCATTGAGAACAGCTGTG +2950
CAAACCCAGANCTGGTCTTTGGATGTTGCTGGTCCAGAGCCTTCACTGAG +3000
AAGAGGGCACCACACTGATTATCCATCTACATTAGAAAACCAGCCAAAAG +3050
TGCCCACTTATTGCAAGCACCTTCCAAGGCCTAGAGAGAAGGTGTTGCTA +3100
AACACAAAGCAGCTTTATTCCTGAGCTTGCACACCTAGTGCATGCTGGCT +3150
AAGTAGCCTGGCCTTTTGTGGTGAGCTTAGACTTAATAACGCCTCCTCTT +3200
ATGACAATAGCAAAACAAAAAAAGAAAGAAATGTTCACTGCTTTCATAAA +3250
AGACCTCCCCACGCCTCCCCCCCCAACCCCCACCACGAAGAAGCAAGAAA +3300
TAATTATGGAAGGAAAACAGCATGGTAAATCCCAGAGAGAGCCCTCTATG +3350
CAAACCTAATGGGTATAATTTGCCATAAAAGGAATTATCGGAGGGAACCA +3400
AATCCATCATTGTGTGTAGCGCTGAAGAGATTAAAGATCAAGAAGGTGTG +3450
CTGGGGTGGGGGTGGGGGTGGGGGTGGGGGTGTGGAATCCTTAAAGGAGG +3500
AGAACTCCCTACTGATAAGATCCAGTTCAATTTTACCAAATAAGTTTCTT +3550
GTTAAATATGAAAGGTCTTCTATTGTTAGTCATGGATGAGAAAGGGCAAA +3600
GTCAAGGTTAGGGTTAATCTTTGGAGGGTGCCCCCATAGACTTTAAAGAC +3650
CTGGAGACTGCATTCTGTTCTCTTCCATTTGGGGCCCACCTCTCCTTTCT +3700
CCATATGCAGTCATATAGAGTCATTGAAGGTCATATCAGTGGTGTTCTTA +3750
GTGTCCAGCCCAAAAAGTATGAGACTCTTTCCCATCCTCTCTTCACACTT +3800
AAAAGACTTTTCAGAAGTCGAGTTTAGAATAAGCGCCATGTACAGTCTTA +3850
GTCCTCCTTGCCGACATTCCCACCCCGTACAGATTCAAGACATCAAGCTG +3900
GCCCATCCATGGGAAAATTAGGGCCCAGACTCAGGAAGCCTGGACTTCTG +3950
ATTGATGGGCTCCCAGGTGAGATCCATGGCTATGATAACTGCTGGGCCCT +4000
TTAACCCTATGGGGTATCAAGAAAAAGGCTTCCTAACTGGGACCTGGGGA +4050
ATAGACAGGATTCTGTGTGTTATACCCAGCATGCAGTTCTTCTGCACACC +4100
GAATGGTCATCATGATGGTAGATTTGTGAGCATCTCGGAAGATCCCTGTA +4150
ACTTTTAGTTTCCTTATATAAATGTGCCTGGGATGATGCTATACCCTTCA +4200
GATAGGAAAGCTAACCAGACCATATGGGAACTTAGGCCTCCCTCCAGCTT +4250
AACCTTGCTGTCCATTGGCTGTTGACCTTGAGCAAGTGACACAGATTCCC +4300
TGGNTCTCCATTTCCTTCCCTATCAAATGACAGAGTGCATTTCCAGCCCT +4350
GTACTACAGTACTTGCAAGGCAGGAGGTAATTCTGCCCAGAGGTTGGCAA +4400
ATCCCTCAAAGGGCCAGGTTGTAAATATTTTACGATTTGCTGGCCAAGAG +4450
ACAAAAATCAAGGATGCTAAGTAAACATGTATTTCGAAATCTAAAAAGAT +4500
TTCTGGTTTCATGATAAATAAAGAAAGGAGCTATGGGCTGGACTGGCTTC +4550
GGGTTGAAGATAACTCCTGAGTTAGGTGGTGCATTTTAATGTACATTAAA +4600
AGAATACTAGCCTTCCATTTATAATGATAATATAAAGTTTATCCATTAAG +4650
ATGCATTCAAACAAAAATAAAGAAATGGCTATTTAAAATTTTTTAATTTA +4700
TTGTATTTAGAGAAGCAAATGGATTTAGTAAAATTACCTTGGTGACAAGC +4750
AAAAGAAAAAAAAAAAACAGTGTTCAGAGGACACTGTACCAAGTGATGGAG +4800
AAGCTTTTGAGCTTTCCAGGTTCCATCTGAGAATGTCCCTGGACGTGGAA +4850
TCTCATCTCTGCCCTGGTATGGTATTTGCCAGTGCAATGATTTTGAGCTG +4900
CTGTCTCCTTCCTGTGAACTGAGTTTGAACCAGCTTATCTCACTAACTTT +4950
ATTTCAAGCATTAAAGAAAATGATGCAAACAGATTTTAAAAAATAATTTA +5000
AAACCTTTCAAACATCACACACACACAGAGAGAGAGAGAGAGAGAGAATG +5050
AAGCACAGAAGCCGTGTTTGGGTGTGTGGGAAGCAGCAAAGCTGAGGTTA +5100
GCGGGCACAACAGTCTTCTGCCAGTCTCAGCCAGACACCTTTGCTTAACC +5150
ACTCTCCTACCACCACGATGTGCCGGGGCAGCAGAGTGCAGCAGTTAGGA +5200
TCAGGAACTTTCCAAGCTGATCCACCTGAGCTGCCACTCAAACCCCATCT +5250
CCTGTCAGCGGGGTTGTCTTGTTCTTGCTACCTGGAGTCCCAGAGCCCCC +5300
ACCATCACCATCCTTCATCCATAAAACAGAACCTTTCCCAGCATCCATCA +5350
GAGTGGTAGCTCTGAGGAAAGTACTAGAGGACGTANCATGGTACTACCCT +5400
CTCCACACTCCTGGAGCACGGTATACTGGAACACACAGAATGCTACAGTG +5450
TCAGGAGTGTTAGACAGTATACATCCTGACGAAACCGAGCTCCTGGGAGG +5500
ACTCANGATNGGTCCTAGCATCTTTAATATNGGCTGCGGGGTATCATGTA +5550
AGCGTTCACTAAATGGAATAAATTACCTACAGATGTAGCATGCTGTTATA +5600
TATGTGTGCATTCATATGCACATATCTAATGTATACAAGGCTTTACAACA +5650
TCAGTTCAGGGCTGGTTTTGCTGGAAAATAAGATTGTGGCAAACACATGC +5700
ACATGCTAACACATAGAAACTTCCAGAGCTGTCTGATTTCAGCATTTTAG +5750
GATAGAGTCACACAGGGCAGTGCTTGGCTTGTTTTCCCAAGAGCACTTTT +5800
AAATATTATATGTCAGAGTTCAGAGGGTCAGACACAAGCAAAGAGAAAAG +5850
TCAATTAAGAGACAGGACTGCACAATCCCAGTCTCCATCCTCAAACCCCA +5900
GCTCTGAAACTTTCTTGTTATGTTACTTTGGGAACTGTCTCAGCATCTCT +5950

GGCTCGGGGTACTCATAGGAGCAATGGGAATAAACAGGGTGTGGTGAGAA +6000
CTGTAGAATACCGTCTGTGTAAAGCACTTGATGAAAGCGCCTACCTGCTG +6050
CCCTTACTAAACGATAAGGGTCATCCTCTCAGGAGTTCTACACCTGGAAG +6100
CTCCTGGAAACTTTTGTTCCAGATGTTCTTACAGCCCCTCTGATATTCCA +6150
GG**GTGATGTTGGACTCCGTAACCCATAGCACCTTCCTGCCCAACGCATCC** +6200
TTCTGTGACCCCCTGATGCCTTGGACCGATCTGTTCAGCAATGAAGACTA +6250
CTACCCTGCCTTTGAGCATCAGACAG

Appendix 3.

Statistical analysis of cell culture results

Formula for determining the least significant difference (LSD)

$|\mu_1 - \mu_2| \geq t_{\alpha/2,\, d.f} \times \sqrt{MSE \times 2/N}$

$|\mu_1 - \mu_2| \geq LSD$

Use a one-way ANOVA table to determine the MSE and d.f. N is the number of replicates for each factor. Use 't' tables to determine the 't' value at the desired level of confidence i.e 0.05 (95%).

Comma1D Proximal constructs

Construct No.	No. Expts (n)	Mean	Std Dev.
0	6	39.39	17.26
1	4	30.92	9.07
2	6	89.25	52.99
3	6	117.5	37.68
4	6	92.84	29.69
5	6	47.36	28.23

LSD = 39.52

Constructs (2), (3) and (4) are significantly different from the negative construct (0) at the 5% level.

Comma1D Distal constructs

Construct No.	No. Expts (n)	Mean	Std Dev.
0	7	62.10	62.97
6	7	90.00	60.04
7	7	93.90	58.27
8	7	107.19	98.26

LSD = 79.2

No construct is significantly different from the negative construct.

Comma1D Intron construct

Construct No.	No. Expts (n)	Mean	Std Dev.
0	8	17.54	13.14
9	8	25.63	20.64

LSD = 18.55

The Intron construct (9) is not significantly different to the negative construct (0).

Comma1D Intron fragment constructs

Construct No.	No. Expts (n)	Mean	Std Dev.
0	5	8.60	1.54
9	5	11.93	3.67
10	5	8.80	1.13
11	5	9.80	2.40
12	5	11.68	2.80

LSD = 3.27

Construct (9) is significantly different from the negative construct (0).

HC11 Proximal and Distal constructs

Construct No.	No. Expts (n)	Mean	Std Dev.
0	6	58.9	37.9
1	6	63.0	49.7
2	6	125.2	67.0
3	6	1718.5	1211.1
4	6	404.8	104.9
5	6	238.5	132.1
6	6	924.5	857.7
7	6	850.8	777.5
8	6	433.1	363.2
9	6	363.5	233.0

LSD = 638.8

Constructs (3), (6) and (7) are significantly different from the negative control at the 5% level. All constructs are significantly different from construct (3) at the 5% level.

HC11 Intron fragment constructs

Construct No.	No. Expts (n)	Mean	Std Dev.
0	6	7.95	1.75
9	6	12.96	5.43
10	6	6.78	2.51
11	6	7.73	2.57
12	6	13.00	3.81

LSD = 3.27

Constructs (9) and (12) are significantly different from the negative construct (0).

References

Ahonen, M., Baker, A. H. and Kahari, V. M. (1998) Adenovirus-mediated gene delivery of tissue inhibitor of metalloproteinases-3 inhibits invasion and induces apoptosis in melanoma cells. *Cancer Res.* **58**, 2310-2315

Aird, W. C., Parvin, J. D., Sharp, P. A. and Rosenberg, R. D. (1994) The interaction of GATA-binding proteins and basal transcription factors with GATA box-containing core promoters. A model of tissue-specific gene expression. *J. Biol. Chem.* **269**, 883-889

Akiyama, Y. (1995) TFSEARCH. Searching transcription factor binding sites. *http://www.rwcp.or.jp/papia/*

Al-Asadi, R., Yi, E. C. and Merchant, J. L. (1995) Sp1 affinity for GC-rich elements correlates with ornithine decarboxylase promoter activity. *Biochem. Biophys. Res. Commun.* **214**, 324-330

Alkema, M. J., Bronk, M., Verhoeven, E., Otte, A., van't Veer, L. J., Berns, A. and van Lohuizen, M. (1996) Identification of Bmi1-interacting proteins as constituents of a multimeric mammalian Polycomb complex. *Genes Dev.* **11**, 226-240

Anderson, K. L., Smith, K. A., Perkin, H., Hermanson, G., Anderson, C.-G., Jolly, D. J., Maki, R. A. and Torbett, B. E. (1999) PU.1 and the granulocyte and macrophage colony stimulating factor receptors play distinct roles in late-stage myeloid cell differentiation. *Blood* **94**, 2310-2318

Andreoli, J. M., Jang, S. I., Chung, E., Coticchia, C. M., Steinert, P. M. and Markova, N. G. (1997) The expression of a novel, epithelium-specific ets transcription factor is restricted to the most differentiated layers in the epidermis. *Nucleic Acids Res.* **25**, 4287-1295

Angel, P. and Karin, M. (1991) The role of Jun, Fos and the AP-1 complex in cell-proliferation and transformation. *Biochim. Biophys. Acta* **1072**, 129-157

Anton, I. A. and Frampton, J. (1988) Tryptophans in myb proteins. *Nature* **336**, 719

Arber, S., Ladle, D. R., Lin, J. H., Frank, E. and Jessell, T. M. (2000) ETS gene Er81 controls the formation of functional connections between group Ia sensory afferents and motor neurons. *Cell* **101**, 485-498

Arceci, R. J., King, A. A., Simon, M. C., Orkin, S. H. and Wilson, D. B. (1993) Mouse GATA-4: a retinoic acid-inducible GATA-binding transcription factor expressed in endodermally derived tissues and heart. *Mol. Cell. Biol.* **13**, 2235-2246

Aronow, B., Lattier, D., Silbiger, R., Dusing, M., Hutton, J., Jones, G., Stock, J., McNeish, J., Potter, S., Witte, D. et al. (1989) Evidence for a complex regulatory array in the first intron of the human adenosine deaminase gene. *Genes Dev.* **3**, 1384-1400

Azizkhan, J. C., Jensen, D. E., Pierce, A. J. and Wade, M. (1993) Transcription from TATA-less promoters: dihydrofolate reductase as a model. *Crit. Rev. Eukaryot. Gene Expr.* **3**, 229-254

Baert, J. L., Monte, D., Musgrove, E. A., Albagli, O., Sutherland, R. L. and de Launoit, Y. (1997) Expression of the PEA3 group of ets-related transcription factors in human breast-cancer cells. *Int. J. Cancer* **70**, 590-597

Bahr, M. J., Vincent, K. J., Arthur, M. J. P., Fowler, A. V., Smart, D. E., Wright, M. C., Clark, I. M., Benyon, R. C., Iredale, J. P. and Mann, D. A. (1999) Control of the tissue inhibitor of metalloproteinases-1 promoter in culture-activated rat hepatic stellate cells: Regulation by activator protein-1 DNA binding proteins. *Hepatology* **29**, 839-848

Bailley, R. A., Bosselut, R., Zucman, J., Cormier, F., Delattre, O., Roussel, M., Thomas, G. and Ghysdael, J. (1994) DNA-binding and transcriptional activation properties of the *EWS-FLI-1* fusion protein resulting from the t(11;22) translocation in Ewing sarcoma. *Mol. Cell. Biol.* **14**, 3230-3241

Baker, A. H., Zaltsman, A. B., George, S. J. and Newby, A. C. (1998) Divergent effects of tissue inhibitor of metalloproteinase-1, -2, or -3 overexpression on rat vascular smooth muscle cell invasion, proliferation, and death in vitro. TIMP-3 promotes apoptosis. *J. Clin. Invest.* **101**, 1478-1487

Ball, R. K., Friis, R. R., Schoenenberger, C. A., Doppler, W. and Groner, B. (1988) Prolactin regulation of β-casein gene expression and of a cytosolic 120-kd protein in a cloned mouse mammary epithelial cell line. *EMBO J.* **7**, 2089-2095

Barbeau, B., Bergeron, D., Beaulieu, M., Nadjem, Z. and Rassart, E. (1996) Characterisation of the human and mouse Fli-1 promoter regions. *Biochim. Biophys. Acta* **1307**, 220-232

Barbeau, B., Barat, C., Bergeron, D. and Rassart, E. (1999) The GATA-1 and Spi-1 transcriptional factors bind to a GATA/EBS dual element in the Fli-1 exon 1. *Oncogene* **18**, 5535-5545

Bartel, F. O., Higuchi, T. and Spyropoulos, D. D. (2000) Mouse models in the study of the Ets family of transcription factors. *Oncogene* **19**, 6443-6454

Bassuk, A. G., Anandappa, R. T. and Leiden, J. M. (1997) Physical interactions between Ets and NFkB/NFAT proteins play an important role in their cooperative activation of the human immunodeficiency virus enhancer in T cells. *J. Virol.* **71**, 3563-3573

Basuyaux, J. P., Ferreira, E., Stehelin, D. and Buttice, G. (1997) The Ets transcription factors interact with each other and with the c-Fos/c-Jun complex via deistinct protein domains in a DNA-dependent and independent manner. *J. Biol. Chem.* **272**, 26188-26195

Behrens, P., Rother, M., Wellmann, A., Krishler, J. and Wernert, N. (2001) The Ets-1 transcription factor is up-regulated together with MMP-1 and MMP-9 in the stroma of pre-invasive breast cancer. *J. Pathol.* **194**, 43-50

Ben-David, Y., Giddens, E. B., Letwin, K. and Berstein, A. (1991) Erythroleukemia induction by friend murine leukemia virus: insertional activation of a new member of the ets gene family, *Fli-1*, closely linked to *c-ets-1*. *Genes Dev.* **5**, 908-918

Benz, C. C., O'Hagen, R. C., Richter, B., Scott, G. K., Chang, C. H., Xiong, X., Chew, K., Ljung, B. M., Edgerton, S., Thor, A. et al. (1997) HER2/Neu and the Ets transcription activator PEA3 are coordinately upregulated in human breast cancer. *Oncogene* **15**, 1513-1525

Bergeron, D., Barbeau, B., Leger, C. and Rassart, E. (1995) Experimental bias in the evaluation of the cellular trnasient expression in DNA co-transfection experiments. *Cell. Mol. Biol. Res.* **41**, 155-159

Berghammer, H. and Auer, B. (1993) Easypreps: Fast and easy plasmid minipreparation for analysis of recombinant clones in *E.coli. Biotechniques* **14**, 526-528

Berland, R. and Chasin, L. A. (1988) The rat albumin gene promoter is appropriately regulated in transient but not in stable transfections. *Nucleic Acids Res.* **16**, 11573-11590

Bhat, N. K., Fisher, R. J., Fujiwara, S., Ascione, R. and Papas, T. S. (1987) Temporal and tissue-specific expression of mouse *ets* genes. *Biochemistry* **84**, 3161-3165

Bhat, N. K., Komschlies, K. L., Fujiwara, S., Fisher, R. J., Mathieson, B. J., Gregorio, T. A., Young, H. A., Kasik, J. W., Ozato, K. and Papas, T. S. (1989) Expression of ets genes in mouse thymocyte subsets and T cells. *J. Immunol.* **142**, 672-678

Bhat, N. K., Thompson, C. B., Lindsten, T., June, C. H., Fujiwara, S., Koizumi, S., Fisher, R. J. and Papas, T. S. (1990) Reciprocal expression of human *ETS-1* and *ETS-2* genes during T-cell activation: regulatory role for the proto-oncogene *ETS-1*. *Proc. Natl. Acad. Sci. USA* **87**, 3723-3727

Bian, J., Wang, Y., Smith, M. R., Kim, H., Jacobs, C., Jackman, J., Kung, H.-F., Colburn, N. H. and Sun, Y. (1996) Suppression of *in vivo* tumor growth and induction of suspension cell death by tissue inhibitor of metalloproteinases (TIMP)-3. *Carcinogenesis* **17**, 1805-1811

Bigger, C. B., Melnikova, I. N. and Gardner, P. D. (1997) Sp1 and Sp3 regulate expression of neuronal nicotinic acetylcholine receptor B4 subunit gene. *J. Biol. Chem.* **272**, 25979-25982

Birrell, G. W., Ramsay, J. R., Tung, J. J. and Lavin, M. F. (2001) Exon skipping in the ATM gene in normal individuals: the effect of blood sample storage on RT-PCR analysis. *Hum. Mutat.* **17**, 75-76

Block, K. L., Shou, Y. P. and Poncz, M. (1996) An Ets/Sp1 interaction in the 5' flanking region of the megakaryocyte-specific alpha-Iib gene appears to stabilize Sp1 binding and is essential for expression of this TATA-less gene. *Blood* **88**, 2071-2080

Bois, P., Williamson, J., Brown, J., Dubrova, Y. E. and Jeffreys, A. J. (1998) A novel unstable mouse VNTR family expanded from SINE B1 elements. *Genomics* **49**, 122-128

Bolon, I., Gouyer, V., Devousassoux, M., Wernert, N., Moro, D., Brambilla, C. and Brambilla, E. (1995) Expression of c-*ets*-1, collagenase 1 and urokinase-type plasminogen activator genes in lung carcinomas. *Am. J. Pathol.* **147**, 1298-1310

Bories, J. C., Willerford, D. M., Grevin, D., Davidson, L., Camus, A., Martin, P., Stehelin, D. and Alt, F. W. (1995) Increased T-cell apoptosis and terminal B-cell differentiation induced by inactivation of the Ets-1 proto-oncogene. *Nature* **377**, 635-638

Bosher, J. M., Williams, T. and Hurst, H. C. (1995) The developmentally regulated transcription factor AP-2 is involved in c-*erbB-2* overexpression in human mammary carcinoma. *Proc. Natl. Acad. Sci. USA* **92**, 744-747

Bosher, J. M., Totty, N. F., Hsuan, J. J., Williams, T. and Hurst, H. C. (1996) A family of AP-2 proteins regulates c-*erb*B-2 expression in mammary carcinoma. *Oncogene* **13**, 1701-1707

Boudreau, N., Werb, Z. and Bissell, M. J. (1996) Suppression of apoptosis by basement membrane requires three-dimensional tissue organisation and withdrawal from the cell cycle. *Proc. Natl. Acad. Sci. USA* **94**, 3425-3430

Boulukos, K. E., Pogninec, P., Sariban, E., Bailly, M., Lagrou, C. and Ghysdael, J. (1990) Rapid and transient expression of Ets2 in mature macrophages following stimulation with cMGF, LPS, and PKC activators. *Genes Dev.* **4**, 401-409

Briggs, M. R., Kadonaga, J. T., Bell, S. P. and Tijan, R. (1986) Purification and biochemical characterisation of the promoter-specific transcription factor Sp1. *Science* **234**, 47-52

Brown, T. A. and McKnight, S. L. (1992) Specificities of protein-protein and protein-DNA interaction of GABPα and two newly defined *ets*-related proteins. *Genes Dev.* **6**, 2502-2512

Bugno, M., Graeve, L., Gatsios, P., Koj, A., Heinrich, P. C., Travis, J. and Kordula, T. (1995) Identification of the interleukin-6/oncostatin M response element in the rat tissue inhibitor of metalloproteinases-1 (TIMP-1) promoter. *Nucleic Acids Res.* **23**, 5041-5047

Buras, J. A., Reenstra, W. R. and Fenton, M. J. (1995) NFκA, a factor required for maximal interleukin-1β gene expression is identical to the *ets* family member PU.1. Evidence for structural alteration following LPS activation. *Mol. Immunol.* **32**, 541-554

Burge, C. and Karlin, S. (1997) Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**, 78-94

Buttice, G., Duterque-Coquillaud, M., Basuyaux, J. P., Carrere, S., Kurkinen, M. and Stehelin, D. (1996) Erg, an Ets-family member, differentially regulates human collagenase1 (MMP1) and stromelysin1 (MMP3) gene expression by physically interacting with the Fos/Jun complex. *Oncogene* **13**, 2297-2306

Calmels, T. P., Mattot, V., Wernert, N., Vandenbunder, B. and Stehelin, D. (1995) Invasive tumors induce c-ets1 transcription factor expression in adjacent stroma. *Biol. Cell* **84**, 53-61

Campbell, C. E., Flenniken, A. M., Skup, D. and Williams, B. R. G. (1991)
Identification of a serum- and phorbol ester-responsive element in the murine tissue inhibitor of metalloproteinase gene. *J. Biol. Chem.* **266**, 7199-7206

Campeau, P., Chapdelaine, P., Seigneurin-Venin, S., Massie, B. and Tremblay, J. P. (2001) Transfection of large plasmids in primary human myoblasts. *Gene Ther.* **8**, 1387-1394

Cao, Z., Umek, R. M. and McKnight, S. L. (1991) Regulated expression of three
C/EBP isoforms during adipose conversion of 3T3-L1 cells. *Genes Dev.* **5**, 1538-1552

Carrere, S., Verger, A., Flourens, A., Stehelin, D. and Duterque-Coquillaud, M. (1998) Erg proteins, transcription factors of the Ets family, form homo, heterodimers and ternary complexes *via* two distinct domains. *Oncogene* **16**, 3261-3268

Carroll, M., Tomasson, M. H., Barker, G. F., Golub, T. R. and Gilliland, D. G. (1996) The TEL/platelet-derived growth factor β receptor (PDGFβR) fusion in chronic myelomonocytic leukemia is a transforming protein that self-associates and activates PDGF□R kinase-dependent signaling pathways. *Proc. Natl. Acad. Sci. USA* **93**, 14845-14850

Chang, C.-H., Scott, G. K., Kuo, W.-L., Xiong, X., Suzdaltseva, Y., Park, J. W., Sayre, P., Erny, K., Collins, C., Gray, J. W. et al. (1997) ESX: a structurally unique *Ets* overexpressed early during human breast tumorgenesis. *Oncogene* **14**, 1617-1622

Chapman, R. S., Lourenco, P. C., Tonner, E., Flint, D. J., Selbert, S., Takeda, K., Akira, S., Clarke, A. R. and Watson, C. J. (1999) Suppression of epithelial apoptosis and delayed mammary gland involution in mice with a conditional knockout of Stat3. *Genes Dev.* **13**, 2604-2617

Chen, Y. and Gill, G. (1994) Positive and negative regulatory elements in the human erbB-2 gene promoter. *Oncogene* **9**, 2269-2276

Chen, Y. and Gill, G. N. (1996) A heterodimeric nuclear protein complex binds two palindromic sequences in the proximal enhancer of the human *erbB-2* gene. *J. Biol. Chem.* **271**, 5183-5188

Chen, Y., Fischer, W. H. and Gill, G. N. (1997) Regulation of the *ERBB-2* promoter by RBPJκ and NOTCH. *J. Biol. Chem.* **272**, 14110-14114

Chen, S., Supakar, P. C., Vellanoweth, R. L., Song, C. S., Chatterjee, B. and Roy, A. K. (1997) Functional role of a conformationally flexible homopurine/homopyrimidine domain of the androgen receptor gene promoter interacting with Sp1 and a pyrimidine single strand DNA-binding protein. *Mol. Endocrinol.* **11**, 3-15

Chotteau-Lelievre, A., Desbiens, X., Pelczar, H., Defossez, P.-A. and de Launoit, Y. (1997) Differential expression patterns of the PEA3 group transcription factors through murine embryonic development. *Oncogene* **15**, 937-952

Clark, A., Krekoski, C., Bou, S., Chapman, K. and Edwards, D. (1997) Increased gelatinase A (MMP-2) and gelatinase B (MMP-9) activities in human brain after focal ischemia. *Neurosci. Lett.* **238**, 53-56

Clarke, C. L. and Sutherland, R. L. (1990) Progestin regulation of cellular proliferation. *Endo. Rev.* **11**, 266-301

Clarkson, R. W. E., Heeley, J. L., Chapman, R., Aillet, F., Hay, R. T., Wyllie, A. and Watson, C. J. (2000) NFκB inhibits apoptosis in murine mammary epithelia. *J. Biol. Chem.* **275**, 12737-12742

Coffer, P., de Jonge, M., Mettouchi, A., Binetruy, B., Ghysdael, J. and Kruijer, W. (1994) junB promoter regulation: Ras mediated transactivation by c-Ets-1 and c-Ets-2. *Oncogene* **9**, 911-921

Colgan, J. and Manley, J. L. (1995) Cooperation between core promoter elements influences transcriptional activity *in vivo*. *Proc. Natl. Acad. Sci. USA* **92**, 1955-1959

Coussens, L. M. and Werb, Z. (1996) Matrix metalloproteinases and the development of cancer. *Chem. Biol.* **3**, 895-904

Crawford, H. C. and Matrisian, L. M. (1996) Mechanisms controlling the transcription of matrix metalloproteinase genes in normal and neoplastic cells. *Enzyme Protein* **49**, 20-37

Crawford, D. L., Segal, J. A. and Barnett, J. L. (1999) Evolutionary analysis of TATA-less proximal promoter function. *Mol. Biol. Evol.* **16**, 194-207

Dalton, S. and Treisman, R. (1992) Characterisation of SAP-1, a protein recruited by serum response factor to the c-*fos* serum response element. *Cell* **68**, 597-612

Danielson, K. G., Oborn, C. J., Durban, E. M., Butel, J. S. and Medina, D. (1984) Epithelial mouse mammary cell line exhibiting normal morphogenesis *in vivo* and functional differentiation *in vitro*. *Proc. Soc. Exp. Biol. Med.* **81**, 3756-3760

Darquet, A., Cameron, B., Wils, P., Scherman, D. and Crouzet, J. (1997) A new DNA vehicle for nonviral gene delivery: supercoiled minicircle. *Gene Ther.* **4**, 1341-1349

de la Brousse, F. C. and McKnight, S. L. (1993) Glimpses of allostery in the control of eukaryotic gene expression. *Trends Genet.* **9**, 151-154

Dean, G., Young, D. A., Edwards, D. R. and Clark, I. M. (2000) The human tissue inhibitor of metalloproteinases (*TIMP*)-1 gene contains repressive elements within the promoter and intron 1. *J. Biol. Chem.* **275**, 32664-32671

Degen, S. J. F., Heckel, J. L., Reich, E. and Degen, J. L. (1987) The murine urokinase-type plasminogen activator gene. *Biochemistry* **26**, 8270-8279

Degnan, B. M., Degnan, S. M., Naganuma, T. and Morse, D. E. (1993) The *ets* multigene family is conserved throughout the Metazoa. *Nucleic Acids Res.* **21**, 3479-3484

Dejardin, E., Bonizzi, G., Bellahcene, A., Castronovo, V., Merville, M. P. and Bours, V. (1995) Highly-expressed p100/p52 (NFKB2) sequesters other NF-κB-related proteins in the cytoplasm of human breast cancer cells. *Oncogene* **11**, 1835-1841

Delannoy-Courdent, A., Fauquette, W., Bourhis, X. F. D.-L., Boilly, B., Vandenbunder, B. and Desbiens, X. (1996) Expression of *c-ETS-1* and *uPA* genes is associated with mammary epithelial cell tubulogenesis or neoplastic scattering. *Int. J. Dev. Biol.* **40**, 1097-1108

Delannoy-Courdent, A., Mattot, V., Fauquette, F. W., Pollet, I., Calmels, T., Vercamer, C., Boilly, B., Vandenbunder, B. and Desbiens, X. (1998) The expression of an Ets1 transcription factor lacking its activation domain decreases uPA proteolytic activity and cell motility, and impairs normal tubulogenesis and cancerous scattering in mammary epithelial cells. *J. Cell Sci.* **111**, 1521-1534

Delattre, O., Zucman, J., Plougastel, B., Desmaze, C., Melot, T., Peter, M., Kovar, H., Joubert, I., de Jong, P., Rouleau, G. et al. (1992) Gene fusion with an *ETS* DNA-binding domain caused by chromosome translocation in human tumours. *Nature* **359**, 162-165

Dhulipala, P. D. K., Lee, L., Rao, V. N. and Reddy, E. S. P. (1998) Fli-1b is generated by usage of differential splicing and alternative promoter. *Oncogene* **17**, 1149-1157

Dittmer, J., Gegonne, A., Gitlin, S. D., Ghysdael, J. and Brady, J. N. (1994) Regulation of parathyroid hormone-related protein (PTHrP) gene expression. *J. Biol. Chem.* **269**, 21428-21434

Dittmer, J., Pise-Masison, C. A., Clemens, K. E., Choi, K.-S. and Brady, J. N. (1997) Interaction of Human T-cell Lymphotrophic Virus Type I Tax, Ets1, and Sp1 in transactivation of the PTHrP P2 promoter. *J. Biol. Chem.* **272**, 4953-4958

Dittmer, J. and Nordheim, A. (1998) Ets transcription factors and human disease. *Biochim. Biophys. Acta* **1377**, F1-F11

Donaldson, L. W., Petersen, J. M., Graves, B. J. and McIntosh, L. P. (1994) Secondary structure of the ETS domain places murine Ets-1 in the superfamily of winged helix-turn-helix DNA-binding proteins. *Biochemistry* **33**, 13509-13516

Donaldson, L. W., Petersen, J. M., Graves, B. J. and McIntosh, L. P. (1996) Solution structure of the ETS domain from murine Ets-1: a winged helix-turn-helix DNA binding motif. *EMBO J.* **15**, 125-134

Dostatni, N., Lambert, P. F., Sousa, R., Ham, J., Howley, P. M. and Yaniv, M. (1991) The functional BPV-1 E2 *trans*-activating protein can act as a repressor by preventing formation of the initiation complex. *Genes Dev.* **5**, 1657-1671

Dufour, C., Weinberger, R. P., Schevzov, G., Jeffery, P. L. and Gunning, P. (1998) Splicing of two internal and four carboxy-terminal alternative exons in nonmuscle tropomyosin 5 pre-mRNA is independently regulated during development. *J. Biol. Chem.* **273**, 18547-18555

Dulbecco, R. (1986) A turning point in cancer research: sequencing the human genome. *Science* **231**, 1055-1056

Ebbinghaus, S. W., Gee, J. E., Rodu, B., Mayfield, C. A., Sanders, G. and Miller, D. M. (1993) Triplex formation inhibits HER-2/neu transcription *In Vitro*. *J. Clin. Invest.* **92**, 2433-2439

Ebbinghaus, S. W., Fortinberry, H. and Gamper, H. B. J. (1999) Inhibition of transcription elongation in the HER-2/neu coding sequence by triplex-directed covalent modification of the template strand. *Biochemistry* **38**, 619-628

Edwards, D. E., Rocheleau, H., Sharma, R. R., Wills, A. J., Cowie, A., Hassell, J. A. and Heath, J. K. (1992) Involvement of AP-1 and PEA3 binding sites in the regulation of murine tissue inhibitor of metalloproteinases-1 (TIMP-1) transcription. *Biochim. Biophys. Acta* **1992**, 41-55

Ekker, M., Sola, C. and Rougeon, F. (1989) The activity of the mouse renin promoter in cells that do not normally produce renin is dependent upon the presence of a functional enhancer. *FEBS Lett.* **255**, 241-247

Emami, K. H., Navarre, W. W. and Smale, S. T. (1995) Core promoter specificities of the Sp1 and VP16 transcriptional activation domains. *Mol. Cell. Biol.* **15**, 5906-5916

Ernst, P. and Smale, S. T. (1995a) Combinatorial regulation of transcription II: the immunoglobulin mu heavy chain gene. *Immunity* **2**, 427-438

Ernst, P. and Smale, S. T. (1995b) Combinatorial regulation of transcription I: general aspects of transcriptional control. *Immunity* **2**, 311-319

Evans, T., Reitman, M. and Felsenfeld, G. (1988) An erythrocyte-specific DNA-binding factor recognises a regulatory sequence common to all chicken globin genes. *Proc. Natl. Acad. Sci. USA* **85**, 5976-5980

Ewing, B. and Green, P. (2000) Analysis of expressed sequence tags indicates 35,000 human genes. *Nat. Genet.* **25**, 232-234

Faisst, S. and Meyer, S. (1992) Compilation of vertebrate-encoded transcription factors. *Nucleic Acids Res.* **20**, 3-26

Fang, X. M. and Brennan, M. D. (1992) Mulitple cis-acting sequences contribute to evolved regulatory variation for Drosophila Adh genes. *Genetics* **131**

Farr, A. and Roman, A. (1992) A pitful of using a second plasmid to determine transfection efficiency. *Nucleic Acids Res.* **20**, 920

Fata, J. E., Leco, K. J., Moorehead, R. A., Martin, D. C. and Khokha, R. (1999) *Timp-1* is important for epithelial proliferation and branching morphogenesis during mouse mammary development. *Dev. Biol.* **211**, 238-254

Feng, Z. W., Marti, A., Jehn, B., Altermatt, H. J., Chicazia, G. and Jaggi, R. (1995) Glucocorticoid and progesterone inhibit involution and programmed cell death in the mouse mammary gland. *J. Cell Biol.* **131**

Fong, K. M., Zimmerman, P. V. and Smith, P. J. (1994) Correlation of loss of heterozygosity at 11p with tumour progression and survival in non-small cell lung cancer. *Genes Chromosomes Cancer* **10**, 183-189

Frantz, S. A., Thiara, A. S., Lodwick, D., Ng, L. L., Eperon, I. C. and Samani, N. J. (1999) Exon repetition in mRNA. *Proc. Natl. Acad. Sci. USA* **96**, 5400-5405

Frohman, M. A. (1993) Rapid amplification of complementary DNA ends for generation of full-length complementary DNAs: Thermal RACE. *Meth. Enzymology* **218**, 340-356

Furlong, E. E. M., Keon, N. K., Thornton, F. D., Rein, T. and Martin, F. (1996) Expression of a 74-kDa nuclear factor 1 (NF1) protein is induced in mouse mammary gland involution. Involution-enhanced occupation of a twin NF1 binding element in the testosterone-repressed prostate message-2/clusterin promoter. *J. Biol. Chem.* **271**, 29688-29697

Galang, C. K., Garcia-Ramirez, J. J., Solski, P. A., Westwick, J. K., Der, C. J., Neznanov, N. N., Oshima, R. G. and Hauser, C. A. (1996) Oncogenic Neu/ErbB-2 increases Ets, AP-1 and NF-□B-dependent gene expression, and inhibiting Ets activation bloks Neu-mediated cellular transformation. *J. Biol. Chem.* **271**, 7992-7998

Gegonne, A., Bosselut, R., Bailly, R.-A. and Ghysdael, J. (1993) Synergistic activation of the HTLV1 LTR Ets-responsive region by transcription factors Ets1 and Sp1. *EMBO J.* **12**, 1169-1178

Gewart, D. R., Coulombe, B., Castelino, M., Skup, D. and Williams, B. R. G. (1987) Characterisation and expression of a murine gene homologous to human EPA/TIMP: a virus-induced gene in the mouse. *EMBO J.* **6**, 651-657

Gigliotti, A. P. and DeWille, J. W. (1998) Lactation status influences expression of CCAAT/enhancer binding protein isoform mRNA in the mouse mammary gland. *J. Cell. Physiol.* **174**, 232-239

Gilinger, G. and Alwine, J. C. (1993) Transcriptional activation by simian virus 40 large T antigen: requirements for simple promoter structures containing either TATA or initiator elements with variable upstream factor binding sites. *J. Virol.* **67**, 6682-6688

Gille, H., Kortenjann, M., Strahl, T. and Shaw, P. E. (1996) Phosphorylation-dependent formation of a quaternary complex at the c-fos SRE. *Mol. Cell. Biol.* **16**, 1094-1102

Gilmore, T. D. (1999) The Rel/NF-κB signal transduction pathway: introduction. *Oncogene* **18**, 6842-6844

Gilmour, D. S., Thomas, G. H. and Elgin, S. C. (1989) Drosophila nuclear proteins bind to regions of alternating C and T residues in gene promoters. *Science* **245**, 1487-1490

Gitlin, S. D., Bosselut, R., Gegonne, A., Ghysdael, J. and Brady, J. N. (1991) Sequence-specific interaction of the Ets1 protein with the long terminal repeat of the human T-lymphotropic virus type I. *J. Virol.* **65**, 5513-5523

Gitlin, S. D., Dittmer, J., Shin, R. C. and Brady, J. N. (1993) Transcriptional activation of the human T-lymphotrophic virus type I long terminal repeat by functional interaction of Tax1 and Ets1. *J. Virol.* **67**, 7307-7316

Golub, T. R., Barker, G. F., Lovett, M. and Gilliland, D. G. (1994) Fusion of PDGF receptor β to a novel *ets*-like gene, *tel*, in chronic myelomonocytic leukemia with t(5;) chromosomal translocation. *Cell* **77**, 307-316

Golub, T. R., Goga, A., Barker, G. F., Afar, D. E. H., McLaughlin, J., Bohlander, S. K., Rowley, J. D., Witte, O. N. and Gilliland, D. G. (1996) Oligomerisation of the ABL tyrosine kinase by the Ets protein TEL in human leukemia. *Mol. Cell. Biol.* **16**, 4107-4116

Gomez, D. E., Alonso, D. F., Yoshiji, H. and Thorgeirsson, U. P. (1997) Tissue inhibitors of metalloproteinases: sturcture, regulation and biological functions. *Eur. J. Cell Biol.*

Gottschalk, L. R., Giannola, D. M. and Emerson, S. G. (1993) Molecular regulation of the human IL-3 gene: Inducible T cell-restricted expression requires intact AP-1 and Elf-1 nuclear protein binding sites. *J. Exp. Med.* **178**, 1681-1692

Gottschalk, S., Tweten, R. K., Smith, L. C. and Woo, S. L. (1995) Efficient gene delivery and expression in mammalian cells using DNA coupled with perfringolysin O. *Gene Ther.* **2**, 498-503

Graves, B. J. and Gillespie, M. E. (1996) DNA binding by the ETS domain. *Nature* **384**, 322

Graves, B. J. and Petersen, J. M. (1998) Specificity within the *ets* family of transcription factors. *Adv. Cancer Res.* **75**, 1-55

Grevin, D., Chen, J. H., Raes, M.-B., Stehelin, D., Vandenbunder, B. and Desbiens, X. (1993) Involvement of the proto-oncogene c-ets 1 and the urokinase plasminogen activator during mouse implantation and placentation. *Int. J. Dev. Biol.* **37**, 519-529

Groner, B. and Gouilleux, F. (1995) Prolactin-mediated gene activation in mammary epithelial cells. *Curr. Opin. Gen. Dev.* **5**, 587-594

Gronskov, K., Olsen, J. H., Sand, A., Pedersen, W., Carlsen, N., Bak Jylling, A. M., Lyngbye, T., Brondum-Nielsen, K. and

Rosenberg, T. (2001) Population-based risk estimates of Wilms tumor in sporadic aniridia. *Hum. Genet.* **109**, 11-18

Grooteclaes, M., Pasleau, F., Dijkmans, H., Berzi, P., Albert, A. and Winkler-Gol, R. (1994) The 6-kilobase c-erbB2 promoter contains positive and negative regulatory elements functional in human mammary cell lines. *Cancer Res.* **54**, 4193-4199

Guarente, L. and Bermingham-McDonogh, O. (1992) Conservation and evolution of transcriptional mechanisms in eukaryotes. *Trends Genet.* **8**, 27-32

Gum, R., Lengyel, E., Juarez, J., Chen, J. H., Sato, H., Seiki, M. and Boyd, D. (1996) Stimulation of 92-kDa gelatinase B promoter activity by *ras* is mitogen-activated protein kinase kinase 1-independent and requires multiple transcription factor binding sites including closely spaced PEA3/*ets* and AP-1 sequences. *J. Biol. Chem.* **271**, 10672-10680

Gumucio, D. L., Shelton, D. A., Bailey, W. J., Slightom, J. L. and Goodman, M. (1993) Phylogenetic footprinting reveals unexpected complexity in trans factor binding upstream from the epsilon-globin gene. *Proc. Natl. Acad. Sci. USA* **90**, 6018-6022

Gunster, M. J., Satijin, D. P., Hamer, K. M., der Blaauwen, J. L., de Bruijn, D., Alkema, M. J., van Lohuizen, M., van Driel, R. and Otte, A. P. (1997) Identification and characterization of interactions between the vertebrate polycomb-group protein BMI1 and human homologs of polyhomeotic. *Mol. Cell. Biol.* **17**, 2326-2335

Gunther, C. V., Nye, J. A., Bryner, R. S. and Graves, B. J. (1990) Sequence-specific DNA binding of the proto-oncoprotein ets-1 defines a transcriptional activator sequence within the long terminal repeat of the Moloney murine sarcoma virus. *Genes Dev.* **4**, 667-679

Gutman, A. and Wasylyk, B. (1990) The collagenase gene promoter contains a TPA and oncogene-responsive unit encompassing the PEA3 and AP-1 binding sites. *EMBO J.* **9**, 2241-2246

Gutman, A. and Wasylyk, B. (1991) Nuclear targets for transcription regulation by oncogenes. *Trends Genet.* **7**, 49-54

Hagemeier, C., Bannister, A. J., Cook, A. and Kouzarides, T. (1993) The activation domain of transcription factor PU.1 binds the retinoblastoma (RB) protein and the transcription fatcor TFIID *in vitro* : RB shows sequence similarity to TFIID and TFIIB. *Proc. Natl. Acad. Sci. USA* **90**, 1580-1584

Hagman, J. and Grosschedl, R. (1992) An inhibitory carboxyl-terminal domain in Ets-1 and Ets-2 mediates differential binding of ETS family factors to promoter sequences of the *mb-1* gene. *Proc. Natl. Acad. Sci. USA* **89**, 8889-8893

Halle, J. P., Haus-Seuffert, P., Woltering, C., Stelzer, G. and Meisterernst, M. (1997) A conserved tissue-specific structure at a human T-cell receptor beta-chain core promoter. *Mol. Cell. Biol.* **17**, 4220-4229

Hansen, S. K. and Tjian, R. (1995) TAFs and TFIIA mediate differential utilization of the tandem Adh promoters. *Cell* **82**, 565-575

Hanzawa, M., Shindoh, M., Higashino, F., Yasuda, M., Inoue, N., Hida, K., Ono, M., Kohgo, T., Nakamura, M., Notasi, K. et al. (2000) Hepatocyte growth factor upregulates E1AF that induces oral squamous cell carcinoma cell invasion by activating matrix metalloproteinase genes. *Carcinogenesis* **21**, 1079-1085

Hart, A., Melet, F., Grossfeld, P., Chien, K., Jones, C., Tunnacliffe, A., Favier, R. and Bernstein, A. (2000) Fli-1 is required for murine vascular and megakaryocytic development and is hemizyously deleted in patients with thrombocytopenia. *Immunity* **13**, 167-177

Hasan, M. T., Subbaroyan, R. and Chang, T. Y. (1991) High efficiency stable gene transfection using chloroquine-treated Chinese hamster ovary cells. *Somat. Cell Mol. Genet.* **17**, 513-517

Heermeier, K., Benedict, M., Li, M., Furth, L., Nunez, G. and Hennighausen, L. (1996) Bax and Bcl-xs are induced at the onset of apoptosis in involuting mammary epithelial cells. *Mec. Dev.* **56**, 197-207

Hennighausen, L. and Robinson, G. W. (1998) Think globally, act locally: the making of a mouse mammary gland. *Genes Dev.* **12**, 449-455

Henry, M. D., Cohen, M. B. and Campbell, K. P. (2001) Reduced expression of dystroglycan in breast and prostate cancer. *Hum. Pathol.* **32**, 791-795

Hida, K., Shindoh, M., Yasuda, M., Hanzawa, M., Funaoka, K., Kohgo, T., Amemiya, A., Totsuka, Y., Yoshida, K. and Fujinaga, K. (1997) Antisense E1AF transfection restrains oral cancer invasion by reducing matrix metalloproteinase activities. *Am. J. Pathol.* **150**, 2125-2132

Hide, W. A., Babenko, V. N., van Heusden, P. A., Seoighe, C. and Kelso, J. F. (2001) The contribution of exon skipping events of chromosome 22 to protien coding diversity. *Genome Res.* **11**, 1848-1853

Higashino, F., Yoshida, K., Noumi, T., Seiki, M. and Fujinaga, K. (1995) Ets-related protein E1A-F can activate three different matrix metalloproteinase gene promoters. *Oncogene* **10**, 1461-1463

Hipskind, R. A., Rao, V. N., Mueller, C. G. F., Reddy, E. S. P. and Nordheim, A. (1991) Ets-related protein Elk-1 is homologous to the c-*fos* regulatory factor p62^{TCF}. *Nature* **354**, 531-534

Ho, I. C., Vorhees, P., Marin, N., Oakley, B. K., Tsai, S. F., Orkin, S. H. and Leiden, J. M. (1991) Human GATA-3: a lineage-restricted transcription factor that regulates the expression of the T cell receptor alpha gene. *EMBO J.* **10**, 1187-1192

Ho, J. M.-Y., Beattie, B. K., Squire, J. A., Frank, D. A. and Barber, D. L. (1999) Fusion of the *ets* transcription factor TEL to Jak2 results in constitutive Jak-Stat signaling. *Blood* **93**, 4354-4364

Hofman, K., Swinnen, J. V., Claessens, F., Verhoeven, G. and Heyns, W. (2000) Apparent coactivation due to interference of expression constructs with nuclear receptor expression. *Mol. Cell. Endocrinol.* **168**, 21-29

Hollywood, D. P. and Hurst, H. C. (1993) A novel transcription factor, OB2-1, is required for overexpression of the prot-oncogene c-*erbB-2* in mammary tumor lines. *EMBO J.* **12**, 2369-2375

Hromas, R. and Klemsz, M. (1994) The ETS oncogene family in development, proliferation and neoplasia. *Int. J. Hematol.* **59**, 257-265

Hudson, L. G., Ertl, A. P. and Gill, G. N. (1990) Structure and inducible regulation of the human c-*erb-B2/neu* promoter. *J. Biol. Chem.* **265**, 4389-4393

Hyatt, I. D., Snoddy, J., Schmoyer, D., Chen, G., Fischer, K., Parang, M., Vokler, I., Petrov, S., Locascio, P., Olman, V. et al. (2000) Improved analysis and annotation tools for whole-genome computational annotation and analysis: GRAIL-EXP genome analysis toolkit and related analysis tools. In *Genome Sequencing & Biology Meeting*, (ed.

Ihn, H., Ihn, Y. and Trojanowska, M. (2001) Spl phosphorylation induced by serum stimulates the human alpha2(I) collagen gene expression. *J. Invest. Dermatol.* **117**, 301-308

Inoue, H., Nojima, H. and Okayama, H. (1990) High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **96**, 23-28

Ishii, S., Xu, Y. H., Stratton, R. H., Roe, B. A., Merlino, G. T. and Pastan, I. (1985) Characterisation and sequence of the promoter region of the human epidermal growth factor receptor gene. *Proc. Natl. Acad. Sci. USA* **82**, 4920-4924

Itoh, T., Tanioka, M., Yoshida, H., Yoshioka, T., Hishimoto, H. and Itohara, S. (1998) Reduced angiogenesis and tumor progression in gelatinase A-deficient mice. *Cancer Res.* **58**, 1048-1051

Iwama, A., Zhang, P., Darlington, G. J., McKercher, S. R., Maki, R. and Tenen, D. G. (1998) Use of RDA analysis of knockout mice to identify myeloid genes regulated in vivo by PU.1 and C/EBPalpha. *Nucleic Acids Res.* **26**, 3034-3043

Jaishankar, S., Zhang, J., Roussel, M. F. and Baker, S. J. (1999) Transforming activity of EWS/FLI is not strictly dependent upon DNA-binding activity. *Oncogene* **18**, 5592-2297

Janknecht, R. and Nordheim, A. (1992) Elk-1 protein domains required for direct and SRF-assisted DNA-binding. *Nucleic Acids Res.* **20**

Janknecht, R. and Nordheim, A. (1993) Gene regulation by Ets proteins. *Biochim. Biophys. Acta* **1155**, 346-356

Jayaraman, G., Srinivas, R., Duggan, C., Ferreira, E., Swaminathan, S., Somasundaram, K., Williams, J., Hauser, C., Kurikinen, M., Dhar, R. et al. (1999) p300/cAMP-responsive element-binding protein interactions with Ets-1 and Ets-2 in the transcriptional activation of the Human Stromelysin promoter. *J. Biol. Chem.* **274**, 17342-17352

Jiang, Z. H. and Wu, J. Y. (1999) Alternative splicing and programmed cell death. *Proc. Soc. Exp. Biol. Med.* **220**, 64-72

John, S., Reeves, R. B., Lin, J.-X., Child, R., Leiden, J. M., Thompson, C. B. and Leonard, W. J. (1995) Regulation of cell-tyoe-specific interleukin-2 receptor α-chain gene expression: Potential role of physical interactions between Elf-1, HMG-I(Y), and NF-κB family proteins. *Mol. Cell. Biol.* **15**, 1786-1796

Johnson, P. F. and McKnight, S. L. (1989) Eukaryotic transcriptional regulatory proteins. *Ann. Rev. Biochem.* **58**, 799-839

Johnson, P. F. and Williams, S. C. (1994) CCAAT/enhancer binding (C/EBP) proteins. In *Liver gene expression.*, (ed. M. Yaniv and F. Tronche), pp. 231-258. Texas: R G Landes, Austin

Jonsen, M. D., Petersen, J. M., Xu, Q. and Graves, B. J. (1996) Characterisation of the cooperative function of inhibitory sequences in Ets-1. *Mol. Cell. Biol.* **16**, 2065-2073

Jorcyk, C. L., Watson, D. K., Mavrothalassitis, G. J. and Papas, T. S. (1991) The human *ETS-1* gene: genomic structure, promoter characterisation and alternative splicing. *Oncogene* **6**, 523-532

Jousset, C., Carron, C., Boureux, A., Tran-Quang, C., Oury, C., Dusanter-Fourt, I., Charon, M., Levin, J., Bernard, O. and Ghysdael, J. (1997) A domain of TEL conserved in a subset of ETS proteins defines a specific oligomerisation interface essential to the mitogenic properties of the TEL-PDGFRβ oncoprotein. *EMBO J.* **16**, 69-82

Karim, F. D., Urness, L. D. and Thummel, C. S. (1990) The ETS domain: a new DNA-binding motif that recognises a purine-rich core DNA sequence. *Genes Dev.* **4**, 1451-1453

Karperien, M., Farih-Sips, H., Lowik, C. W. G. M., de-Laat, S. W., Boonstra, J. and Defize, L. H. K. (1997) Expression of the parathyroid hormone-related peptide gene in retinoic acid-induced differentiation: involvement of ETS and Sp1. *Mol. Endocrinol.* **11**, 1435-1448

Kas, K., Finger, E., Grall, F., Gu, X., Akbarali, Y., Bolax, J., Weiss, A., Oettgen, P. and Libermann, T. A. (2000) ESE-3, a novel member of an epithelium-specific Ets transcription factor subfamily, demonstrates different target gene specificity from ESE-1. *J. Biol. Chem.* **275**, 2986-2998

Kasahara, M., Hayashi, M., Tanaka, K., Inoko, H., Sugaya, K., Ikemura, T. and Ishibashi, T. (1996) Chromosomal localisation of the proteasome *Z* subunit gene reveals an ancient chromosomal duplication involving the major histocompatibility complex. *Proc. Natl. Acad. Sci. USA* **93**, 9096-9101

Katsanis, N., Fitzgibbon, J. and Fisher, E. M. (1996) Paralogy mapping: identification of a region in the human MHC triplicated onto human chromosomes 1 and 9 allows the prediction and isolation of novel PBX and NOTCH loci. *Genomics* **35**, 101-108

Kaufman, P. D. and Rio, D. C. (1991) *Drosophila P*-element transposase is a transcriptional repressor *in vitro*. *Proc. Natl. Acad. Sci. USA* **88**, 2613-2617

Kawahara, H., Kasahara, M., Nishiyama, A., Ohsumi, K., Goto, T., Kishimoto, T., Saeki, Y., Yokosawa, H., Shimbara, N., Murata, S. et al. (2000) Developmentally regulated, alternative splicing of the Rpn10 gene generates multiple forms of 26S proteasomes. *EMBO J.* **19**, 4144-4153

Kawana, M., Lee, M. E., Quertermous, E. E. and Quertermous, T. (1995) Cooperative interaction of GATA-2 and Ap1 regulates transcription of the endothelin-1 gene. *Mol. Cell. Biol.* **15**, 4225-4231

Kaya, M., Yoshida, K., Higashino, F., Mitaka, T., Ishii, S. and Fujinaga, K. (1996) A single *ets*-related transcription factor, E1AF, confers invasive phenotype on human cancer cells. *Oncogene* **12**, 221-227

Keydar, I., Chen, L., Karby, S., Weiss, F. R., Delarea, J., Radu, M., Chaitcik, S. and Brenner, H. J. (1979) Establishment and characterisation of a cell line of human breast carcinoma origin. *Eur. J. Cancer* **15**, 659-670

Khokha, R. (1994) Suppression of the tumorgenic and metastatic abilities of murine B16-F10 melanoma cells *in vivo* by the overexpression of the tissue inhibitor of metalloproteinases-1. *J. Natl. Cancer Inst.* **86**, 299-304

Kim, W. Y., Sieweke, M., Ogawa, E., Wee, H.-J., Engelmeier, U., Graf, T. and Ito, Y. (1999) Mutual activation of Ets-1 and AML1 DNA binding by direct interaction of their autoinhibitory domains. *EMBO J.* **18**, 1609-1620

Kirkness, E. F. and Fraser, C. M. (1993) A strong promoter element is located between alternative exons of a gene encoding the Human □-aminobutyric acid-type A receptor □3 subunit (GABRB3). *J. Biol. Chem.* **268**, 4420-4428

Kitange, G., Shibata, S., Tokunaga, Y., Yagi, N., Yasunaga, A., Kishikawa, M. and Naito, S. (1999) Ets-1 transcription factor-mediated urokinase-type plasminogen activator expression and invasion in glioma cells stimulated by serum and basic fibrolblast growth factors. *Lab. Invest.* **79**, 407-415

Klambt, C. (1993) The *Drosophila* gene *pointed* encodes two ETS-like proteins which are invloved in the development of the middle glial cells. *Development* **117**, 163-176

Kleinbaum, L. A., Duggan, C., Ferreira, E., Coffey, G. P., Buttice, G. and Burton, F. H. (1999) Human chromosomal localization, tissue/tumor expression, and regulatory function of the ets family gene EHF. *Biochem. Biophys. Res. Commun.* **264**, 119-126

Klemsz, M. J., McKercher, S. R., Celada, A., Van Beveren, C. and Maki, R. A. (1990) The macrophage and B cell specific transcription factor PU.1 is realted to the *ets* oncogene. *Cell* **61**, 113-124

Klemsz, M., Hromas, R., Raskind, W., Bruno, E. and Hoffman, R. (1994) PE-1, a novel ETS oncogene family member, localizes to chromosome 1q21-q23. *Genomics* **20**, 291-294

Kodandapani, R. (1996) A new pattern for helix-turn-helix recognition revealed by the PU.1 ETS-domain-DNA complex. *Nature* **380**, 456-460

Kola, I., Brookes, S., Green, A. R., Garber, A. R., Tymms, M., Papas, T. S. and Seth, A. (1993) The Ets1 transcription factor is widely expressed during murine embryo development and is associated with mesodermal cells involved in morphogenetic processes such as organ formation. *Proc. Natl. Acad. Sci. USA* **90**, 7588-7592

Kominato, Y., Galson, D. L., Waterman, W. R., Webb, A. C. and Auron, P. E. (1995) Monocyte expression of the human prointerleukin 1β gene (*IL1B*) is dependent on promoter sequences which bind the hematopoietic transcription factor Spi-1/PU.1. *Mol. Cell. Biol.* **15**, 58-68

Kreiss, P., Cameron, B., Rangara, R., Mailhe, P., Augerre-Charriol, O., Airiau, M., Scherman, D., Crouzet, J. and Pitard, B. (1999) Plasmid DNA size does not affect the physiochemical properties of lipoplexes but modulates gene transfer efficiency. *Nucleic Acids Res.* **27**, 3792-3798

Kruger, A., Fata, J. E. and Khokha, R. (1997) Altered tumor growth and metastasis of a T-cell lymphoma in *Timp-1* transgenic mice. *Blood* **90**, 1993-2000

Labat-Moleur, F., Steffan, A., Brisson, C., Perron, H., Feugeas, O., Furstenberger, P., Oberling, F., Brambilla, E. and Behr, J. (1996) An electron microscopy study into the mechanism of gene transfer with lipopolyamines. *Gene Ther.* **3**, 1010-1017

Lacronique, V., Boureux, A., Della-Valle, V., Poirel, H., Tran-Quang, C., Mauchauffe, M., Berthou, C., Lessard, M., Berger, R., Ghysdael, J. et al. (1997) A TEL-JAK2 fusion protein with constitutive kinase activity in human leukemia. *Science* **278**, 1309-1312

LaMarco, K., Thompson, C. C., Byers, B. P., Walton, E. M. and McKnight, S. L. (1991) Identification of Ets- and Notch-related subunits in GA binding protein. *Science* **253**, 789-792

Lambert de Rouvroit, C., Bernier, B., Royaux, I., de Bergeyck, V. and Goffinet, A. M. (1999) Evolutionary conserved, alternative splicing of reelin during brain development. *Exp. Neurol.* **156**, 229-238

Landschulz, W. H., Johnson, P. F. and McKnight, S. L. (1989) The DNA-binding domain of the rat liver nuclear protein C/EBP is bipartite. *Science* **243**, 395-397

Langley, B. (1999) Alternative polyadenylation and stability of mRNAs encoded by a novel, mammary specific ETS gene, BrETS. *Biological Sciences, University of Waikato*

Laudet, V., Niel, C., Duterque-Coquillaud, M., Leprince, D. and Stehelin, D. (1993) Evolution of the ets gene family. *Biochem. Biophys. Res. Commun.* **190**, 8-14

Laudet, V. (1997) Evolution of the nuclear receptor superfamily: early diversification from an ancestral orphan receptor. *J. Mol. Endocrinol.* **19**, 207-226

Laudet, V., Hanni, C., Stehelin, D. and Duterque-Coquillaud, M. (1999) Molecular phylogeny of the ETS gene family. *Oncogene* **18**, 1351-1359

Lautenberger, J. A., Burdett, L. A., Gunnell, M. A., Qi, S., Watson, D. K., O'Brien, S. J. and Papas, T. S. (1992) Genomic dispersal of the ets gene family during metazoan evolution. *Oncogene* **7**, 1713-1719

Le Gallic, L., Sgouras, D., Beal, G. J. and Mavrothalassitis, G. (1999) Transcriptional repressor ERF is a Ras/Mitogen-activated protein kinase target that regulates cellular proliferation. *Mol. Cell. Biol.* **19**, 4121-4133

Lee, W., Mitchell, P. and Tjian, R. (1987) Purified transcription factor Ap-1 interacts with TPA-inducible enhancer elements. *Cell* **49**, 741-752

Lee, M. E., Dhadly, M. S., Temizer, D. H., Clifford, J. A., Yoshizumi, M. and Quertermous, T. (1991) Regulation of endothelin-1 gene expression by Fos and Jun. *J. Biol. Chem.* **266**, 19034-19039

Lee, J. H., Jang, S. I., Yang, J. M., Markova, N. G. and Steinert, P. M. (1996) The proximal promoter of the human transglutaminase 3 gene. Stratified squamous epithelial-specific expression in cultured cells is mediated by binding of Sp1 and ets transcription factors to a proximal promoter element. *J. Biol. Chem.* **271**, 4561-4568

Lehmann, U., Brocke, P., Dittmer, J. and Nordheim, A. (1999) Characterisation of the human *elk-1* promoter. *J. Biol. Chem.* **274**, 1736-1744

Lelievre, E., Lionneton, F., Soncin, F. and Vandenbunder, B. (2000) The Ets family contains transcriptional activators and reperssors involved in angiogenesis. *Int. J. Biochem. Cell Biol.* **33**, 391-407

Lengyel, E., Stepp, E., Gum, R. and Boyd, D. (1995) Involvement of a mitogen-activated protein kinase signaling pathway in the regulation of urokinase promoter activity by c-Ha-ras. *J. Biol. Chem.* **270**, 23007-23012

Leprince, D., Gegonne, A., Coll, J., de-Taisne, C., Schneeberger, A., Lagrou, C. and Stehelin, D. (1983) A putative second cell-derived oncogene of the avian leukaemia retrovirus E26. *Nature* **306**, 395-397

Li, M. L., Liu, X. W., Robinson, G., BarPeled, U., Wagner, K. U., Young, W. S., Hennighausen, L. and Furth, P. A. (1997) Mammary-derived signals activate programmed cell death during the first stage of mammary gland involution. *Proc. Natl. Acad. Sci. USA* **94**, 3425-3430

Li, R., Pei, H. and Watson, D. K. (2000) Regulation of Ets function by protein-protein interactions. *Oncogene* **19**, 6514-6523

Li, Q., Zhang, Han, H., Rohde, A. and Stamatoyannopoulos, G. (2002) Evidence that DNase I hypersensitive site 5 of the human beta-globin locus control region functions as a chromosomal insulator in transgenic mice. *Nucleic Acids Res.* **30**, 2484-2491

Liang, H. (1994a) Solution structure of the *ets* domain of Fli-1 when bound to DNA. *Nature Struct. Biol.* **1**, 871-876

Liang, H., Olejniczak, E. T., Mao, X., Nettesheim, D. G., Yu, L., Thompson, C. B. and Fesik, S. W. (1994b) The secondary structure of the ets domain of human Fli-1 resembles that of the helix-turn-helix DNA-binding motif of the Escherichia coli catabolite gene activator protein. *Proc. Natl. Acad. Sci. USA* **91**, 11655-11659

Lim, F., Kraut, N., Frampton, J. and Graf, T. (1992) DNA binding by c-Ets-1, but not v-Ets, is repressed by an intramolecular mechanism. *EMBO J.* **11**, 643-652

Lim, S., Nasibitt, S., Yoon, J., Hwang, J. I., Suh, P. G., Sheng, M. and Kim, E. (1999) Characteristion of the Shank family of synaptic proteins. Multiple genes, alternative splicing, and differential expression in brain and development. *J. Biol. Chem.* **274**, 29510-29518

Lin, P. P., Brody, R. I., Hamelin, A. C., Bradner, J. E., Healey, J. H. and Ladanyi, M. (1999) Differential transactivation by alternative EWS-FLI1 fusion proteins correlates with clinical heterogeneity in Ewing's sarcoma. *Cancer Res.* **59**, 1428-1432

Ling, Y., Lakey, J. H., Roberts, C. E. and Sharrocks, A. D. (1997) Molecular characterisation of the B-box protein-protein interaction motif of the ETS-domain transcription factor Elk-1. *EMBO J.* **16**, 2431-2440

Liu, X. W., Robinson, G. W. and Hennighausen, L. (1996) Activation of Stat5a and Stat5b by tyrosine phosphorylation is tightly linked to mammary gland differentiation. *Mol. Endocrinol.* **10**, 1496-1506

Logan, S. K., Garabedian, M. J., Campbell, C. E. and Werb, Z. (1996) Synergistic transcriptional activation of the tissue inhibitor of metalloproteinases-1 promoter via functional interaction of AP-1 and Ets-1 transcription factors. *J. Biol. Chem.* **271**, 774-782

Lopez, R. G., Carron, C., Oury, C., Gardellin, P., Bernard, O. and Ghysdael, J. (1999) TEL is a sequence-specific transcriptional repressor. *J. Biol. Chem.* **274**, 30132-30138

Ludwig, C. U., Raefle, G., Dalquen, P., Stulz, P., Stahel, R. and Obrecht, J. P. (1991) Allelic loss on the short arm of chromosome 11 in non-small-cell lung cancer. *Int. J. Cancer* **49**, 661-665

Lund, L. R., Romer, J., Thomasset, N., Solberg, H., Pyke, C., Bissell, M. J., Dano, K. and Werb, Z. (1996) Two distinct phases of apoptosis in mammary gland involution: proteinase-independent and -dependent pathways. *Development* **122**, 181-193

Luo, D., Guerin, E., Ludwig, M., Stoll, I., Basset, P. and Anglard, P. (1999) Transcriptional induction of stromelysin-3 in mesodermal cells is mediated by an upstream CCAAT/enhancer-binding protein element associated with a DNase I-hypersensitive site. *J. Biol. Chem.* **274**, 37177-37185

Lyer, V. R., Eisen, M. B., Ross, D. T., Schuler, G., Moore, T., Lee, J. C. F., Trent, J. M., Staudt, L. M., Hudson, J., Boguski, M. S. et al. (1999) The transcriptional program in the response of human fibroblasts to serum. *Science* **283**, 83-87

Maira, S.-M., Wurtz, J.-M. and Wasylyk, B. (1996) Net (ERP/SAP2), one of the Ras-inducible TCFs, has a novel inhibitory domain with resemblance to the helix-loop-helix motif. *EMBO J.* **15**, 5849-5865

Majerus, M.-A., Bibollet-Ruche, F., Telliez, J.-B., Wasylyk, B. and Bailleul, B. (1992) Serum, AP-1 and Ets-1 stimulate the human ets-1 promoter. *Nucleic Acids Res.* **20**, 2699-2703

Maniatis, T. (1991) Mechanisms of alternative pre-mRNA splicing. *Science* **251**, 33-34

Manson, A. and Huxley, C. (2001) Skipping of exon 9 of human CFTR in YAC-transgenic mice. *Genomics* **77**, 127-134

Mao, X., Miesfeldt, S., Yang, H., Leiden, J. M. and Thompson, C. B. (1994) The FLI-1 and chimeric EWS-FLI-1 oncoproteins display similar DNA binding specificities. *J. Biol. Chem.* **269**, 18216-18222

Martin, D. I. and Orkin, S. H. (1990) Transcriptional activation and DNA binding by the erythroid factor GF-1/NF-E1/Eryf1. *Genes Dev.* **4**, 1886-1898

Martinez, E., Chiang, C. M., Ge, H. and Roeder, R. G. (1994) TATA-bindsing protein-associated factor(s) in TFIID function through the initiator to direct basal transcription from a TATA-less class II promoter. *EMBO J.* **13**, 3115-3126

Masson, R., Lefebvre, O., Noel, A., Fahime, M. E., Chenard, M. P., Wendling, C., Kebers, F., LeMeur, M., Dierich, A., Foidart, J. M. et al. (1998) In vivo evidence that the stromelysin-3 metalloproteinase contributes in a paracrine manner to epithelial cell malignancy. *J. Cell Biol.* **140**, 1535-1541

Mattot, V., Vercamer, C., Soncin, F., Calmels, T., Huguet, C., Fafeur, V. and Vandenbunder, B. (2000) Constitutive expression of the DNA-binding domain of Ets1 increases endothelial cell adhesion and stimulates their organization into capillary-like structures. *Oncogene* **19**, 762-772

Mavrothalassitis, G. J., Watson, D. K. and Papas, T. S. (1990) Molecular and functional characterisation of the promoter of *ETS2*, the human c-*ets-2* gene. *Proc. Natl. Acad. Sci. USA* **87**, 1047-1051

Mavrothalassitis, G. J. and Ghysdael, J. (2000) Proteins of the ETS family with transcriptional repressor activity. *Oncogene* **19**, 6524-6532

May, W. A., Lessnick, S. L., Braun, B. S., Klemsz, M., Lewis, B. C., Lunsford, L. B., Hromas, R. and Denny, C. T. (1993) The Ewing's sarcoma *EWS/FLI-1* fusion gene encodes a more potent transcriptional activator and is a more powerful transforming gene than *FLI-1*. *Mol. Cell. Biol.* **13**, 7393-7398

McArthur, M., Gerum, S. and Stamatoyannopoulos, G. (2001) Quantification of DNaseI-sensitivity by real-time PCR: quantitative analysis of DNaseI-hypersensitivity of the mouse beta-globin LCR. *J. Mol. Biol.* **313**, 27-34

McKenzie, R. W., Hu, J. and Brennan, M. D. (1994) Redundant cis-acting elements control expression of the Drosophila affinidisjuncta Adh gene in the larval fat body. *Nucleic Acids Res.* **22**, 1257-1264

McLean, T. W., Ringold, S., Neuberg, D., Stegmaier, K., Tantravahi, R., Ritz, J., Koeffler, H. P., Takeuchi, S., Janssen, J. W. G., Seriu, T. et al. (1996) TEL/AML-1 dimerises and is associated with a favourable outcome in childhood acute lymphoblastic leukemia. *Blood* **88**, 4252-4258

Melet, F., Motro, B., Rossi, D. J., Zhang, L. and Bernstein, A. (1996) Generation of a novel Fli-1 protein by gene targeting leads to a defect in thymus development and a delay in friend virus-induced erythroleukemia. *Mol. Cell. Biol.* **16**, 2708-2718

Mepham, T. B. (1987) Physiology of lactation. England: Butler and Tanner Ltd

Mercatante, D. and Kote, R. (2000) Modification of alternative splicing pathways as a potential approach to chemotherapy. *Pharmacol. Ther.* **85**, 237-243

Mironov, A. A., Fickett, J.W., Gelfand, M.S. (1999) Frequent alternative splicing of human genes. *Genome Res.* **9**, 1288-1293

Mitchell and Tijan. (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* **245**, 371-378

Miyamoto, M., Fujita, T., Kimura, Y., Maruyama, M., Harada, H., Sudo, Y., Miyata, T. and Taniguchi, T. (1988) Regulated expression of a gene encoding a nuclear factor, IRF-1, that specifically binds to IFN-beta gene regulatory elements. *Cell* **54**, 903-913

Mizuguchi, G., Kanei-Ishii, C., Sawazaki, T., Horikoshi, M., Roeder, R. G., Yamamoto, T. and Ishii, S. (1994) Independent control of transcription initiation from two sites by an initiator-like element and TATA box in the human c-*erbB-2* promoter. *FEBS Lett.* **348**, 80-88

Montgomery, A. M., Mueller, B. M., Reisfeld, R. A., Taylor, S. M. and DeClerck, Y. A. (1994) Effect of tissue inhibitor of the matrix metalloproteinases-2 expression on the growth and spontaneous metastasis of a human melanoma cell line. *Cancer Res.* **54**, 5467-5473

Moulton, K. S., Semple, K., Wu, H. and Glass, C. K. (1994) Cell-specific expression of the macrophage scavenger receptor gene is dependent on PU.1 and a composite AP-1/ets motif. *Mol. Cell. Biol.* **14**, 4408-4418

Murphy, G. and Knauper, V. (1997) Relating matrix metalloproteinase structure to function: why the "hemopexin" domain? *Matrix Biol.* **15**, 511-518

Murre, c., McCaw, P. S. and Baltimore, D. (1989) A new DNA binding and dimerisation motif in immunoglobulin enhance binding, daughterless, MyoD, and myc proteins. *Cell* **56**, 777-783

Muthusamy, N., Barton, K. and Leiden, J. M. (1995) Defective activation and survival of T cells lacking the Ets-1 transcription factor. *Nature* **377**, 639-642

Naito, K., Kanbayashi, N., Nakajima, S., Murdi, T., Arakawa, K., Nishimura, S. and Okuyama, A. (1994) Inhibition of growth of human tumor cells in nude mice by a metalloproteinase inhibitor. *Int. J. Cancer* **58**, 730-735

Nakshatri, H., BhatNakshatri, P., Martin, D. A., Goulet, R. J. and Sledge, G. W. (1997) Constitutive activation of NF-κB during progression of breast cancer to hormone-independent growth. *Mol. Cell. Biol.* **17**, 3629-3639

Narayanan, A. S. and Page, R. C. (1987) Serum regulation of collagen biosynthesis in human diploid fibroblasts. *Biochem. Biophys. Res. Commun.* **145**, 639-645

Neve, R., Chang, C. H., Scott, G. K., Wong, A., Friis, R. R., Hynes, N. E. and Benz, C. C. (1998) The epithelium-specific Ets transcription factor ESX is associated with mammary gland development and involution. *FASEB J.* **12**, 1541-1550

Neznanor, N., Man, A. K., Yamamoto, H., Hauser, C. A., Cardiff, R. D. and Oshima, R. G. (1999) A single targeted Ets2 allele restricts development of mammary tumours in transgenic mice. *Cancer Res.* **59**, 4242-4246

Noonberg, S. B., Scott, G. K., Hunt, C. A., Hogan, M. E. and Benz, C. C. (1994) Inhibition of transcription factor binding to the HER2 promoter by triplex-forming oligodeoxyribonucleotides. *Gene* **149**, 123-126

Nothias, J.-Y., Weinmann, R., Blangy, D. and Melin, F. (1993) Analysis of transcription factors binding to the duplicated PEA1 and PEA3 sites that are required for polyomavirus mutant expression in PCC4 embryonic carcinoma cells. *J. Virol.* **67**, 3036-3047

Novina, C. D. and Roy, A. L. (1996) Core promoters and transcriptional control. *Trends Genet.* **12**, 351-355

Nunn, M. F., Seeburg, P. H., Moscovici, C. and Duesberg, P. H. (1983) Tripartite structure of the avian erythroblastosis virus E26 transforming gene. *Nature* **306**, 391-395

Nye, J. A., Petersen, J. M., Gunther, C. V., Jonsen, M. D. and Graves, B. J. (1992) Interaction of murine Ets-1 with GGA-binding sites establishes the ETS domain as a new DNA-binding motif. *Genes Dev.* **6**, 975-990

O'Neill, E. M., Rebay, I., Tijan, R. and Rubin, G. M. (1994) The activities of two Ets-related transcription factors required for Drosophila eye development are modulated by the Ras/MAPK pathway. *Cell* **78**, 137-147

O'Rourke, J., Yuan, R. and DeWille, J. (1997) CCAAT/enhancer binding protein-delta (C/EBP-delta) is induced in growth-arrested mouse mammary epithelial cells. *J. Biol. Chem.* **272**, 6291-6296

Oettgen, P., Kas, K., Dube, A., Gu, X., Grall, F., Thamrongsak, U., Akbarali, Y., E, F., Bolax, J., Endress, G. et al. (1999) Characterisation of ESE-2, a novel ESE-1 related Ets trascription factor that is related to glandular epithelium and differentiated keratinocytes. *J. Biol. Chem.* **274**, 29439-29452

Oettgen, P., Barcinski, M., Boltax, J., Stolt, P., Akbarali, Y. and Libermann, T. A. (1999) Genomic organisation of the human ELF3 (ESE-1/ESX) gene, a member of the *ets* transcription factor family, and identification of a functional promoter. *Genomics* **55**, 358-362

Ohkuma, Y., Horikoshi, M., Roeder, R. G. and Desplan, C. (1990) Engrailed, a homeodomain protein, can repress *in vitro* transcription by competition with the TATA box-binding protein transcription factor IID. *Proc. Natl. Acad. Sci. USA* **37**, 2289-2293

Papadopoulous, P., Ridge, S. A., Boucher, C. A., Stocking, C. and Wiedemann, L. M. (1995) The novel activation of *ABL* by fusion to an *ets*-related gene, *TEL. Cancer Res.* **55**, 34-38

Parks, W. C. (1999) Matrix metalloproteinases in repair. *Wound Repair Regen.* **7**, 423-432

Patel, N. H. (1994) The evolution of arthropod segmentation: insights from comparisons of gene expression patterns. *Development* **Suppl.**, 201-207

Peeters, P., Raynaud, S. D., Cools, J., Wlodarska, I., Grosgeorge, J., Philip, P., Monpoux, F., Van Rompaey, L., Baens, M., Van den Berghe, H. et al. (1997) Fusion of *TEL*, the ETS-variant gene 6 (*ETV6*), to the receptor-associated kinase *JAK2* as a result of t(9;12) in a lymphoid and t(9;15;12) in a myeloid leukemia. *Blood* **90**, 2535-2540

Peterman, R., Mossier, B. M., Aryee, D. N. T., Khazak, V., Golemis, E. A. and Kovar, H. (1998) Oncogenic EWS-Fli1 interacts with hsRPB7, a subunit of human RNA polymerase II. *Oncogene* **17**, 603-610

Petersen, J. M., Skalicky, J. J., Donaldson, L. W., McIntosh, L. P., Alber, T. and Graves, B. J. (1995) Modulation of transcription factor Ets-1 DNA binding: DNA-induced unfolding of an α helix. *Science* **269**, 1866-1869

Philp, J. A., Burdon, T. G. and Watson, C. J. (1996) Differential activation of STATs 3 and 5 during mammary gland development. *FEBS Lett.* **396**, 77-80

Phylactides, M., Rowntree, R., Nuthall, H., Ussery, D., Wheeler, A. and Harris, A. (2002) Evaluation of potential regulatory elements identified as DNase I hypersensitive sites in the CFTR gene. *Eur. J. Biochem.* **269**, 553-559

Pognonec, P., Boulukos, K. E., Gesquiere, J. C., Stehelin, D. and Ghysdael, J. (1988) Mitogenic stimulation of thymoctes results in the calcium-dependent phosphorylation of c-*ets*-1 proteins. *EMBO J.* **7**, 977-983

Porumb, H., Gousset, H., Letellier, R., Salle, V., Briane, D., Vassy, J., Amor-Gueret, M., Israel, L. and Taillander, E. (1996) Temporary *ex vivo* inhibition of the expression of the human oncogene *HER2* (*NEU*) by a triple helix-forming oligonucleotide. *Cancer Res.* **56**, 515-522

Prestridge, D. S. (1995) Predicting Pol II promoter sequences using transcription factor binding sites. *J. Mol. Biol.* **249**, 923-932

Quandt, K., Frech, K., Karas, H., Wingender, E. and Werner, T. (1995) MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nuceotide sequence data. *Nucleic Acids Res.* **23**, 4878-4884

Quarrie, L. H., Addey, C. V. P. and Wilde, C. J. (1996) Programmed cell death during mammary tissue involution induced by weaning, litter removal, and milk stasis. *J. Cell. Physiol.* **168**, 559-569

Rabault, B. and Ghysdael, J. (1994) Calcium-induced phosphorylation of ETS1 inhibits its specific DNA binding activity. *J. Biol. Chem.* **269**, 28143-28151

Rao, V. N. and Reddy, S. P. (1992) A divergent *ets*-related protein, Elk-1, recognises similar c-*ets*-1 proto-oncogene target sequences and acts as a transcriptional activator. *Oncogene* **7**, 2335-2340

Rao, V. N., Ohno, T., Prasad, D. D. K., Bhattacharya, G. and Reddy, E. S. P. (1993) Analysis of the DNA-binding and transcriptional activation functions of human *Fli*-1 protein. *Oncogene* **8**, 2167-2173

Raziuddin, A., Court, D., Sarkar, F. H., Liu, Y.-L., Kung, H.-F. and Raziuddin, R. (1997) A c-*erbB-2* promoter-specific nuclear matrix protein from human breast tumor tissues mediates NF-□B DNA binding activity. *J. Biol. Chem.* **272**, 15715-15720

Rebay, I. and Rubin, G. M. (1995) Yan functions as a general inhibitor of differentiation and is negatively regulated by activation of the Ras1/MAPK pathway. *Cell* **81**, 857-866

Reese, M. G. and Eeckman, F. H. (1995) Novel neural netwrok algorithms for improved eukaryotic promoter site recognition. In *The seventh international genome sequencing and analysis conference.*, (ed. Hilton Head Island, South Carolina

Reeve, A. E., Sih, S. A., Raizis, A. M. and Feinberg, A. P. (1989) Loss of allelic heterozygosity at a second locus of chromosome 11 in sporadic Wilms' tumor cells. *Mol. Cell. Biol.* **9**, 1799-1803

Riesgo-Escovar, R., Woodard, C. and Carlson, J. R. (1994) Olfactory physiology in the *Drosophila* maxillary palp requires the visual system gene *rdgB*. *Comp. Phsiol. A* **175**, 687-693

Robinson, G. W., Johnson, P. F., Hennighausen, L. and Sterneck, E. (1998) The C/EBPβ transcription factor regulates epithelial cell proliferation and differentiation in the mammary gland. *Genes Dev.* **12**, 1907-1916

Roeder, R. G. (1991) The complexities of eukayotic transcription initiation: regulatioin of preinitiation complex assembly. *Trends Biochem. Sci.* **16**, 402-408

Rogge, R., Green, P. J., Urano, J., Horn-Saban, S., Mlodzik, M., Shilo, B.-Z., Hartenstein, V. and Banerjee, U. (1995) The role of yan in mediating the choice between cell division and differentiation. *Development* **121**, 3947-3958

Romana, S. P., Mauchauffe, M., Le Coniat, M., Chumakov, I., Le Paslier, D., Berger, R. and Bernard, O. A. (1995) The t(12;21) of acute lymphoblastic leukemia results in a tel-AML1 gene fusion. *Blood* **85**, 3662-3670

Roth, J. H., Das, G. C. and Piatigorsky, J. (1991) Chicken □B1-crystallin gene expression:presence of conserved functional polyomavirus enhancer-like and octamer binding-like promoter elements found in non-lens genes. *Mol. Cell. Biol.* **11**, 1488-1499

Rutter, J. L., Mitchell, T. I., Buttice, G., Meyers, J., Gusella, J. F., Ozelius, L. J. and Brinckerhoff, C. E. (1998) A single nucleotide polymorphism in the matrix metalloproteinase-1 promoter creates an Ets binding site and augments transcription. *Cancer Res.* **58**, 5321-5325

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

Sanger, F., Niclen, S. and Coulson, A. R. (1977) DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467

Sapi, E., Flick, M. B., Rodov, S. and Kacinski, B. M. (1998) Ets-2 transdominant mutant abolishes anchorage-independent growth and macrophage colony-stimulating factor-stimulated invasion by BT20 breast carcinoma cells. *Cancer Res.* **58**, 1027-1033

Savant-Bhonsale, S. and Cleveland, D. W. (1992) Evidence for instability of mRNAs containing AUUUA motifs mediated through translation-dependent assembly of a > 20S degradation complex. *Genes Dev.* **6**, 1927-1939

Scholz, H., Deatrick, J., Klaes, A. and Klambt, C. (1993) Genetic dissection of *pointed*, a Drosophila gene encoging two ETS-related proteins. *Genetics* **135**, 455-468

Scohy, S., Gabant, P., Szpirer, C. and Szpirer, J. (2000) Identification of an enhancer and an alternative promoter in the first intron of the α-fetoprotein gene. *Nucleic Acids Res.* **28**, 3743-3751

Scott, G. K., Daniel, J. C., Xiong, X., Maki, R. A., Kabat, D. and Benz, C. C. (1994) Binding of an ETS-related protein within the DNase I hypersensitive site of the HER2/*neu* promoter in human breast cancer cells. *J. Biol. Chem.* **269**, 19848-19858

Scott, E. W., Fisher, R. C., Olson, M. C., Kehrli, E. W., Simon, M. C. and Singh, H. (1997) PU.1 functions in a cell-autonomous manner to control the differentiation of multipotential lymphoid-myeloid progenitors. *Immunity* **6**, 437-447

Scott, G. K., Chang, C.-H., Erny, K. M., Xu, F., Fredericks, W. J., Rauscher, F. J. I., Thor, A. D. and Benz, C. C. (2000) Ets regulation of the *erb*B2 promoter. *Oncogene* **19**, 6490-6502

Seagraves, T. N., Lydon, J. P., Hovey, R. C., Vonderhaar, B. K. and Rosen, J. M. (2000) C/EBPβ (CCAAT/Enhancer binding protein) controls cell fate determination during mammary gland development. *Mol. Endocrinol.* **14**, 359-368

Seeler, J. S., Muchardt, C., Podar, M. and Gaynor, R. B. (1993) Regulatory elements involved in Tax-mediated transactivation of the HTLV-1 LTR. *Virology* **196**, 442-450

Sementchenko, V. I. and Watson, D. K. (2000) Ets target genes: past, present and future. *Oncogene* **19**, 6533-6548

Seth, A., Watson, D. K., Blair, D. G. and Papas, T. S. (1989) c-*ets-2* protooncogene has mitogenic and oncogenic activity. *Proc. Natl. Acad. Sci. USA* **86**, 7833-7837

Seth, A., Hodge, D. R., Thompson, D. M., Robinson, L., Panayiotakis, A., Watson, D. K. and Papas, T. S. (1993) ETS family proteins activate transcription from HIV-1 long terminal repeat. *AIDS Res. Hum. Retroviruses* **9**, 1017-1023

Seth, A., Robinson, L., Panayiotakis, A., Thompson, D. M., Hodge, D. R., Zhang, X. K., Watson, D. K., Ozato, K. and Papas, T. S. (1994) The EndoA enhancer contains multiple ETS binding site repeats and is regulated by ETS proteins. *Oncogene* **9**, 469-477

Seto, E., Shi, Y. and Shenk, T. (1991) YY1 is an initiator sequence-binding protein that directs and activates transcription in vitro. *Nature* **354**, 241-245

Seul, K. H., Tadros, P. N. and Beyer, E. C. (1997) Mouse connexin40: Gene structure and promoter analysis. *Genomics* **46**, 120-126

Sgouras, D. N., Athanasiou, M. A., Beal, G. J. J., Fischer, R. J., Blair, D. G. and Mavrothalasson, G. J. (1995) ERF: an ETS domain protein with strong transcriptional repressor activity, can supress *ets*-associated tumorigenesis and is regulated by phosphorylation during cell cycle and mitogenic stimulation. *EMBO J.* **14**, 4781-4793

Sharrocks, A., Brown, A., Ling, Y. and Yates, P. (1997) The ETS-domain transcription factor family. *Int. J. Biochem. Cell Biol.* **29**, 1370-1387

Shenk, M. A. and Steele, R. E. (1993) A molecular snapshot of the metazoan 'Eve'. *Trends Biochem. Sci.* **18**, 459-463

Shin, M. K. and Koshland, M. E. (1993) Ets-related protein PU.1 regulates expression of the immounoglobulin J-chain gene through a novel ets-binding element. *Genes Dev.* **7**, 2006-2015

Shipman, R., Schraml, P., Colombi, M. and Ludwig, C. U. (1998) Allelic deletion at chromosome 11p13 defines a tumour suppressor region between the catalase gene and D11S935 in human non-small cell lung carcinoma. *Int. J. Oncol.* **12**, 107-111

Shirasaki, F., Makhluf, H. A., LeRoy, C., Watson, D. K. and Trojanowska. (1999) Ets transcription factors cooperate with Sp1 to activate the human Tenascin- promoter. *Oncogene* **18**, 7755-7764

Shore, P. and Sharrocks, A. D. (1995) The ETS-domain transcription factors Elk-1 and SAP-1 exhibit differential DNA binding specificities. *Nucleic Acids Res.* **23**, 4698-4706

Sironi, M., Corti, S., Locatelli, F., Cagliani, R. and Comi, G. P. (2001) A novel splice site mutation (3157+1G>T) in the dystrophin gene causing total exon skipping and DMD phenotype. *Hum. Mutat.* **17**, 239

Skalicky, J. J., Donaldson, L. W., Petersen, J. M., Graves, B. J. and McIntosh, L. P. (1996) Structural coupling of the inhibitory regions flanking the ETD domain of murine Ets-1. *Prot. Science* **5**, 296-309

Slupsky, C. M., Gentile, L. N. and McIntosh, L. P. (1998) Assigning the NMR spectra of aromatic amino acids in proteins: analysis of two Ets pointed domains. *Biochem. Cell. Biol.* **76**, 379-390

Smale, S. T. (1994) Core promoter architecture for eukaryotic protein-coding genes. In *Transcription: Mechanisms and regulation.*, (ed. R. C. Conaway and J. W. Conaway), pp. 63-81. New York: Raven Press

Smit, A. F. A. and Riggs, A. D. (1995) MIRs are classic, tRNA-derived SINEs that amplified before the mammalian radiation. *Nucleic Acids Res.* **23**, 98-102

Smit, A. F. A. and Green, P. (2000) Repeatmasker. *http://repeatmasker.genome.washington.edu/*

Smith, C. W. J., Patton, J. G. and Nadal-Ginard, B. (1989) Alternative splicing control of gene expression. *Ann. Rev. Genet.* **23**, 527

Smith, D. L. and Morris, B. J. (1991) Transient expression analyses of DNA extending 2.4 kb upstream of the human renin gene. *Mol. Cell. Endocrinol.* **80**, 139-146

Smits, P., Bhalerao, J. and Merregaert, J. (1999) Molecular cloning and characterisation of the mouse Ecm1 gene and its 5' regulatory sequences. *Gene* **226**, 253-261

Sovak, M. A., Bellas, R. E., Kim, D. W., Zanieski, G. J., Rogers, A. E., Traish, A. M. and Sonenshein, G. E. (1997) Aberrant Nuclear Factor-κB/Rel expression and the pathogenesis of breast cancer. *J. Clin. Invest.* **100**, 2952-2960

Spyropoulos, D. D., Pharr, P. N., Lavenburg, K. R., Jackers, P., Papas, T. S., Ogawa, M. and Watson, D. K. (2000) Hemorrhage, impaired hematopoiesis and lethality in mouse embryos carrying a targeted disruption of the *Fli1* transcription factor. *Mol. Cell. Biol.* **20**, 5643-5652

Starck, J., Doubeikovski, A., Sarrazin, S., Gonnet, C., Rao, G., Skoultchi, A., Godet, J., Dusanter-Fourt, I. and Morle, F. (1999) *Spi-1/PU.1* is a positive regulator of the *Fli-1* gene involved iin inhibition of erythroid differentiation in friend erythroleukemic cell lines. *Mol. Cell. Biol.* **19**, 121-135

Stracke, M. L., Murata, J., Aznavoorian, S. and Liotta, L. A. (1994) The role of the extracellular matrix in tumor cell metastasis. *In Vivo* **8**, 49-58

Strange, R., Li, F., Saurer, S., Burkhardt, A. and Friis, R. R. (1992) Apoptotic cell death and tissue remodelling during mouse mammary gland involution. *Development* **115**, 49-58

Strehler, E. E. and Zacharias, D. A. (2001) Role of alternative splicing in generating isoform diversity among plasma membrane calcium pumps. *Physiol. Rev.* **81**, 21-50

Streuli, C. H. and Gilmore, A. P. (1999) Adhesion-mediated signaling in the regulation of mammary epithelial cell survival. *J. Mamm. Gland Biol. Neoplasia* **4**, 183-191

Sturm, R. A., Das, G. and Herr, W. (1988) The ubiquitous octamer-binding protein Oct-1 contains a POU domain with a homeo box subdomain. *Genes Dev.* **2**, 1582-1599

Su, G. H., Chen, H.-M., Muthusamy, N., Garrett-Sinha, L. A., Baunoch, D., Tenen, D. G. and Simon, M. C. (1997) Defective B cell receptor-mediated responses in mice lacking the Ets protein, Spi-B. *EMBO J.* **16**, 7118-7129

Suen, T.-C. and Hung, M.-C. (1990) Mulitple *cis*- and *trans*- acting elements involved in regulation of the *neu* gene. *Mol. Cell. Biol.* **10**, 6306-6315

Suzuki, Y., Tsunoda, T., Sese, J., Taira, H., Mizushima-Sugano, J., Hata, H., Ota, T., Isogai, T., Tanaka, T., Nakamura, Y. et al. (2001) Identification and characterisation of the potential promoter regions of 1031 kinds of human genes. *Genome Res.* **11**, 677-684

Tamir, A., Howard, J., Higgins, R. R., Li, Y.-J., Berger, L., Zacksenhaus, E., Reis, M. and Ben-David, Y. (1999) Fli-1 an Ets-related transcription factor, regulates erythropoietin-induced erythroid proliferation and differentiation: Evidence for direct transcriptional repression of the *Rb* gene during differentiation. *Mol. Cell. Biol.* **19**, 4452-4464

Thomas, J. H. (1993) Thinking about genetic redundancy. *Trends Genet.* **9**, 395-399

Thomas, R. S., Ng, A. N. Y., Zhou, J., Tymms, M. J., Doppler, W. and Kola, I. (2000) The Elf group of Ets-related transcription factors. *Adv. Exp. Med. Biol.* **480**, 123-128

Thompson, C. C., Brown, T. A. and McKnight, S. L. (1991) Convergence of Ets- and Notch-related structural motifs in a heteromeric DNA binding complex. *Science* **253**, 762-768

Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) Clustal W: improving the sensitivity of progressive mulitple sequence alignmant through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673-4680

Travers, M. T., Barber, M. C., Tonner, E., Quarrie, L., Wilde, C. J. and Flint, D. J. (1996) The role of prolactin and growth hormone in the regulation of casein gene expression and mammary cell survival: Relationships to milk synthesis and secretion. *Endocrinology* **137**, 1530-1539

Treier, M., Bohmann, D. and Mlodzik, M. (1995) JUN cooperates with the ETS domain protein Pointed to induce photoreceptor R7 fate in the Drosophila eye. *Cell* **93**, 753-760

Trimble, M. S., Xin, J. H., Guy, C. T., Muller, W. J. and Hassell, J. A. (1993) PEA3 is overexpressed in mouse metastatic mammary adenocarcinomas. *Oncogene* **8**, 3037-3042

Trojanowska, M. (2000) Ets factors and regulation of the extracellular matrix. *Oncogene* **19**, 6464-6471

Tugores, A., Le, J., Sorokina, I., Snijders, A. J., Duyao, M., Reddy, P. S., Carlee, L., Ronshaugen, M., Mushegian, A., Watanaskul, T. et al. (2001) The epithelium-specific ETS protein EHF/ESE-3 is a context-dependent transcriptional repressor downstream of MAPK signaling cascades. *J. Biol. Chem.* **276**, 20397-20406

Tymms, M. J. and Kola, I. (1994) Regulation of gene expression by transcription factors Ets-1 and Ets-2. *Mol. Reprod. Dev.* **39**, 208-214

Unsworth, B. R., Hayman, G. T., Carroll, A. and Lelkes, P. I. (1999) Tissue-specific alternative mRNA splicing of phenylethanolamine N-methyltransferase (PNMT) during development by intron retention. *Int. J. Dev. Neurosci.* **17**, 45-55

Urness, L. D. and Thummel, C. S. (1990) Molecular interactions within the ecdysone of E74 transcription during *Drosophila* development. *Cell*

Usheva, A. and Shenk, T. (1994) TATA-binding protein-independent initiation: YY1, TFIIB, and RNA polymerase II direct basal transcription on supercoiled template DNA. *Cell* **76**, 1115-1121

Villena, J. A., Martin, I., Vinas, O., Cormand, B., Iglesias, R., Mampel, T., Giralt, M. and Villarroya, F. (1994) ETS transcription factors regulate the expression of the gene for the human mitochondrial ATP synthase beta-subunit. *J. Biol. Chem.* **269**, 32649-23654

Wagner, K., Dendorfer, U., Chilla, S., Schlondorff, D. and Luckow, B. (2001) Identification of new regulatory sequences far upstream of the mouse monocyte chemoattractant protein-1 gene. *Genomics* **78**, 113-123

Wang, C.-Y., Bassuk, A. G., Boise, L. H., Thompson, C. B., Bravo, R. and Leiden, J. M. (1994) Activation of the Granulocyte-Macrophage colony-stimulating factor promoter in T cells requires cooperative binding of Elf-1 and Ap-1 transcription factors. *Mol. Cell. Biol.* **14**, 1153-1159

Wang, L. C., Kuo, F., Fujiwara, Y., Gilliland, D. G., Golub, T. R. and Orkin, S. H. (1997) Yok sac angiogenic defect and intra-embryonic apaptosis in mice lacking the Ets-related factor TEL. *EMBO J.* **16**, 4374-4383

Wasylyk, B., Wasylyk, C., Flores, P., Begue, A., Leprince, D. and Stehelin, D. (1990) The c-ets proto-oncogenes encode transcription factors that cooperate with c-Fos and c-Jun for transcriptional activation. *Nature* **346**, 191-193

Wasylyk, C., Gutman, A., Nicholson, R. and Wasylyk, B. (1991) The c-Ets oncoprotein activates the stromelysin promoter through the same elements as several non-nuclear oncoproteins. *EMBO J.* **10**, 1127-1134

Wasylyk, C., Kerckaert, J.-P. and Wasylyk, B. (1992) A novel modulator domain of Ets transcriptional factors. *Genes Dev.* **6**, 965-974

Wasylyk, B., Hahn, S. L. and Giovane, A. (1993) The Ets family of transcription factors. *Eur. J. Biochem.* **211**, 7-18

Watanabe, H., Sawada, J. I., Yana, K. I., Yamaguchi, K., Goto, M. and Handa, H. (1993) cDNA cloning of transcription factor E4TF1 subunits with Ets and notch motifs. *Mol. Cell. Biol.* **13**, 1385-1391

Watson, D. K., Smyth, F. E., Thompson, D. M., Cheng, J. Q., Testa, J. R., Papas, T. S. and Seth, A. (1992) The *ERG*B/*FLI*-1 gene: Isolation and characterisation of a new member of the family of human *ETS* transcription factors. *Cell Growth Differ.* **3**, 705-713

Weinberg, R. A. (1991) Tumor suppressor genes. *Science* **254**, 1138-1146

Weintraub, J. S., Chow, K. N. B., Luo, R. X., Zhnag, S. H., He, S. and Dean, D. C. (1995) Mechanism of active transcriptional repression by the retinoblastoma protein. *Nature* **375**, 812-815

Weis, L. and Reinberg, D. (1997) Accurate postioning of RNA polymerase II on a natural TATA-less promoter is independent of TATA-binding-protein-associated factors and initiator-binding proteins. *Mol. Cell. Biol.* **17**, 2973-2984

Weissman, B. E., Saxon, P. J., Pasquale, S. R., Jones, G. R., Geiser, A. G. and Stanbridge, E. J. (1987) Introduction of a normal human chromosome 11 into a Wilms' tumor cell line controls its tumorgenic expression. *Science* **236**, 175-180

Werner, M. H., Clore, G. M., Fisher, C. L., Fisher, R. J., Trinh, L., Shiloach, J. and Gronenborn, A. M. (1995) The solution structure of the human ETS1-DNA complex reveals a novel mode of binding and true side chain intercalation. *Cell* **83**, 761-771

Wernert, N., Gilles, F., Fafeur, V., Bouali, F., Raes, M. B., Pyke, C., Dupressoir, T., Seitz, G., Vandenbunder, B. and Stehelin, D. (1994) Stromal expression of c-Ets1 transcription factor correlates with tumor invasion. *Cancer Res.* **54**, 5683-5688

Westermarck, J., Seth, A. and Kahari, V. M. (1997) Differential regulation of interstitial collagenase (MMP-1) gene expression by ETS transcription factors. *Oncogene* **14**, 2651-2660

Westermarck, J. and Kahari, V.-M. (1999) Regulation of matrix metalloproteinase expression in tumor invasion. *FASEB J.* **13**, 781-792

Whyatt, D. J., deBoer, E. and Grosveld, F. (1993) The two zinc finger-like domains of GATA-1 have different DNA binding specificities. *EMBO J.* **12**, 4993-5005

Wiley, S. R., Kraus, R. J. and Mertz, J. E. (1992) Functional binding of the TATA box binding component of transcription factor TFIID to the -30 region of TATA-less promoters. *Proc. Natl. Acad. Sci. USA* **89**, 5814-5818

Wilke, M., Fortunati, E., van den Broek, M., Hoogeveen, A. T. and Scholte, B. J. (1996) Efficiency of a peptide-based gene delivery system depends on mitotic activity. *Gene Ther.* **3**, 1133-1142

Williams, J. M. and Daniel, C. W. (1983) Mammary ductal elongation: Differentiation of myoepithelium and basal lamina during branching morphogenesis. *Dev. Biol.* **97**, 274-290

Williams, S. C., Cantwell, C. A. and Johnson, P. F. (1991) A family of C/EBP-related proteins capable of forming covalently linked leucine zipper dimers *in vitro*. *Genes Dev.* **5**, 1553-1567

Wilson, C. L., Heppner, K. J., Labosky, P. A., Hogan, B. L. and Matrisian, L. M. (1997) Intestinal tumorigenesis is suppressed in mice lacking the metalloproteinase matrilysin. *Proc. Natl. Acad. Sci. USA* **94**, 1402-1407

Wisdom, R. (1999) AP-1: One switch for many signals. *Exp. Cell Res.* **253**, 180-185

Wolfert, M. A. and Seymour, L. W. (1998) Chloroquine and amphipathic peptide helices show synergistic transfection *in vitro*. *Gene Ther.* **5**, 409-414

Wolfsberg, T. G., McEntyre, J. and Schuler, G. D. (2001) Guide to the draft human genome. *Nature* **409**, 824-826

Woods, D. B., Ghysdael, J. and Owen, M. J. (1992) Identification of nucleotide preferences in DNA sequences recognised specifically by c-Ets-1 protein. *Nucleic Acids Res.* **20**, 699-704

Xing, X., Wang, S. C., Xia, W., Zou, Y., Shao, R., Kwong, K. Y., Yu, Z., Zhang, S., Miller, S., Huang, L. et al. (2000) The Ets protein PEA3 supresses *HER-2/neu* overexpression and inhibits tumorigenesis. *Nat. Med.* **6**, 189-195

Xu, Y. and Szoka, F. J. (1996) Mechanism of DNA release from cationic liposome/DNA complexes used in cell transfection. *Biochemistry* **35**, 5616-5623

Yamamoto, H., Flannery, M. L., Kupriyanov, S., Pearce, J., McKercher, S. R., Henkel, G. W., Maki, R. A., Werb, Z. and Oshima, R. G. (1998) Defective trophoblast function in mice with a targeted mutation of Ets2. *Genes Dev.* **12**, 1315-1326

Yamauchi, M., Ogata, Y., Kim, R. H., Li, J. J., Freedman, L. P. and Sodek, J. (1996) AP-1 regulation of the rat bone ialoprotein gene transcription is mediated through a TPA response element within a glucocorticod response unit in the gene promoter. *Matrix Biol.* **15**, 119-130

Yang, J. N., Allan, E. H., Anderson, G. I., Martin, T. J. and Minkin, C. (1997) Plasminogen activator system in osteoclasts. *J. Bone Miner. Res.* **12**, 761-768

Yang, C., Shapiro, L. H., Rivera, M., Kumar, A. and K, B. P. (1998) A role for CREB binding protein and p300 transactivational coactivators in Ets-1 transactivation functions. *Mol. Cell. Biol.* **18**, 2218-2229

Yean, D. and Grall, J. (1997) Transcription reinitiation rate: a special role for the TATA box. *Mol. Cell. Biol.* **17**, 3809-3816

Zacharias, D. A., Dalrymple, S. J. and Strehler, E. E. (1995) Transcript distribution of plasma membrane Ca^{2+} pump isoforms and slice variants in the human brain. *Brain Res. Mol. Brain Res.* **28**, 263-272

Zanta, M. A., Belguise-Valladier, P. and Behr, J. P. (1999) Gene delivery: a single nuclear localization signal peptide is sufficient to carry DNA to the cell nucleus. *Proc. Natl. Acad. Sci. USA* **96**, 91-96

Zenzie-Gregory, B., Khachi, A., Garraway, I. P. and Smale, S. T. (1993) Mechanism of initiator-mediated transcription: evidence for a functional interaction between the TATA-binding protein and DNA in the absence of a specific recognition sequence. *Mol. Cell. Biol.* **13**, 3841-3849

Zhang, D. E., Hetherington, C. J., Tan, S., Dziennis, S. E., Gonzalez, D. A., Chen, H. M. and Tenen, D. G. (1994) Sp1 is a critical factor for the monocytic specific expression of human CD14. *J. Biol. Chem.* **269**, 11425-11434

Zhang, L., Eddy, A., Teng, Y.-T., Fritzler, M., Kluppel, M., Melet, F. and Bernstein, A. (1995) An immunological renal disease in transgenic mice that overexpresses *Fli-1*, a member of the *ets* family of transcription factor genes. *Mol. Cell. Biol.* **15**, 6961-6970

Zhou, J., Ng, A. Y. N., Tymms, M. J., Jermiin, L. S., Seth, A. K., Thomas, R. S. and Kola, I. (1998) A novel transcription factor, ELF5, belongs to the ELF subfamily of ETS genes and maps to human chromosome 11p13-15, a region subject to LOH and rearrangement in human carcinoma cell lines. *Oncogene* **17**, 2719-2732