Molecular characterisation of murine *Nfe2l1*

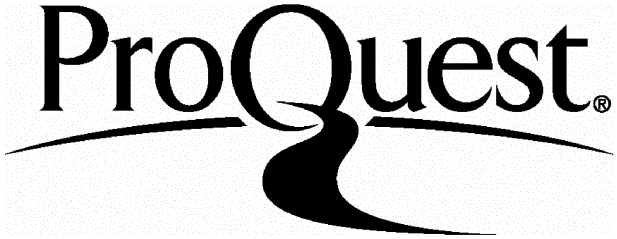
by

Judith Mary McKie

A thesis submitted for the degree Doctor of Philosophy

University of London
July 1998

Molecular Medicine Unit
Institute of Child Health
University College London
30 Guilford Street
London WC1N 1EH



For Jane and Pad

The Moving Finger writes; and, having writ,
Moves on: nor all thy Piety nor Wit
Shall lure it back to cancel half a line,
Nor all thy tears wash out a Word of it.
Rubáiyát of Omar Kayyám

ABSTRACT

The conservation of some developmental processes amongst species as diverse as *Drosophila*, *C.elegans*, man and mouse has been one of the most exciting scientific discoveries in recent years. The initial premise for the thesis was the isolation of a mouse cDNA fragment, 8dl, which had striking similarity to a Drosophila protein, CNC. *cnc* encodes a bZIP transcription factor and is thought to be involved in head specification. The initial aims were therefore to establish if the murine homologue (latterly known as *Nfe2l1*) of CNC had been isolated, and what, if any, was its role in murine development.

The full length sequence of *Nfe2l1* was determined and its chromosomal localisation in both mouse and man identified. The expression pattern of *Nfe2l1* throughout murine development was studied, and although it was ubiquitous throughout the developmental stages studied, specific sites of developmental upregulation could be identified. *Nfe2l1* is unlikely to be the murine homologue of CNC, but rather, one member of a family of CNC-related proteins that form a subclass of the bZIP transcription factor family. The evidence from both the work presented in this thesis and the published studies suggests that *Nfe2l1* is likely to have a role in murine development.

In addition, a novel human family member, *NFM*, has been identified by database screening with NFE2L1. Several cDNAs have been isolated and partially characterised by sequence and northern analyses, and *NFM*-positive human PACs have permitted FISH localisation studies.

ACKNOWLEDGEMENTS

I would like to thank Prof. Peter J. Scambler for allowing me to join his group, and for all his support and encouragement throughout the years. I would also like thank Dr. Baljinder Mankoo and Prof. Bob Williamson for the initial impetus that introduced me to the Biochemistry department at St. Mary's Hospital Medical School.

In particular at St. Mary's, I'd like to mention the other half of our group, Isik and Lizzy, the Friedreich's Ataxia group, and Joy, Moya, Chrissie, Tony, for making my time at St. Mary's so enjoyable. In more recent years, I'd like to thank the rest of Pete's group for all their fun, friendship and technical advice especially Steph, Shalan, Sara, Roy, Katie, Marion, and more recently Paola, Robert, Matt, Roser and Hannah. Outside the lab, I would like to mention my fellow scientists and friends, Eduardo and Mel, Kit and Cath, Sarah, Adam, Phil and Pete, for the beers, chat and fun times. Far flung friends have also been tremendous, so thanks to Lizzie, Helen, Robert, Fran, Maggie, Lawrence, Aileen, Allan and Peter.

A special thanks to the lunch crew who have been made to suffer with me with the writing-up: thanks Helen, Nora and Catherine, you are great! I am also grateful to Catherine for her advice on developmental biology, Hilary for her advice on FISH, and Nora, my collaborator on NFM.

As ever, I am deeply indebted to Karen, Hilary and the Boys for always being there for me, at both good and bad times, thanks guys. Finally I would like to thank my Mum for putting up with a perpetually grumbly daughter, and I hope that she enjoys reading her copy!

ABBREVIATIONS

A ..... adenosine
aa ..... amino acid
A_{260} ..... optical absorbance at 260nm
ATP ..... adenosine triphosphate
BAC ..... bacterial artificial chromosome
BCIP 5-bromo-4-chloro-3-indolyl-phosphate
bp ..... base pair
BSA ..... bovine serum albumin
C ..... cytosine
cDNA ..... complementary DNA
Ci ..... Curie
cM ..... centimorgan
cps ..... counts per second
DAPI ..... 4',6-diamino-2phenylindole
DIG ..... digoxygenin
DMSO ..... dimethyl sulphoxide
DNA ..... deoxyribonucleic acid
Dnase ..... deoxyribonuclease
dNTP ..... deoxynucleotide triphosphate
dpc ..... days post coitum
DTT ..... dithiothreitol
EDTA ..... ethylenediamine-tetra acetic acid
ES ..... embryonic stem
EST ..... expressed sequence tag
FISH ..... fluorescence in situ hybridisation
G ..... guanosine
g ..... gram
g ..... acceleration due to gravity
h ..... hour
HGMP-RC ..... Human Genome Mapping Project-Resource Centre
hncDNA ..... heterogeneous nuclear cDNA
ICRF-RLDB ..... Imperial Cancer Research Fund- Reference Library Database
IMAGE ..... integrated molecular analysis of genomes and their expression

IPTG..... isopropyl-β-D-thiogalactoside
kb ..... kilobase
l ..... litre
µ ..... micro
m ..... milli
M ..... molar
Mb ..... megabase
min..... minute
mol ..... moles
n ..... nano
NBT..... 4-nitro blue tetrazolium chloride
NCBI ..... National Center for Biotechnology Information
nuc..... nucleotide position
ORF..... open reading frame
p ..... chromosome short arm
PAC..... P1 artificial chromosome
PBS..... phosphate buffered saline
PCR..... polymerase chain reaction
pmol ..... picomole
psi ..... pounds per square inch
q..... chromosome long arm
RACE ..... rapid amplification of cDNA ends
RE ..... restriction endonuclease
RNA..... ribonucleic acid
rpm ..... revolutions per minute
RT ..... reverse transcriptase
RT-PCR..... reverse transcription PCR
s..... second
SDS..... sodium dodecyl sulphate
SSC ..... saline sodium citrate
T ..... thymidine
TE ..... tris-EDTA buffer
THC..... Tentative Human Consensus
TIGR..... The Integrated Genome Resource
U..... units
UTP..... uridine triphosphate
UV..... ultraviolet
V..... volt

v / v ..... volume per volume

w / v..... weight per volume

X-gal 5-bromo-4-chloro-3-indolyl-β-D-galactoside

YAC ..... yeast artificial chromosome

MMU ..... Mus musculus

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 INTRODUCTION

1.1 FOREWORD

The phenomenon of development, namely the ordered progressive change of a fertilised egg to a complex individual, has fascinated mankind for centuries. The modern concept of embryology has evolved in the last 150 years or so with the emergence of many theories, models and concepts, often derived from the study of spontaneous or induced variations of normal development. Technological advances in experimental biology and in particular molecular genetics in the last twenty years has allowed the practical testing of many developmental hypotheses, linking the classical approach of studying phenotype to understand the underlying genetic basis to the reverse approach of molecular genetics. One significant outcome is the unification of the field of developmental biology for diverse biological systems, with the realisation that many general developmental processes are conserved within the animal kingdom.

The developmental biology of a large number of organisms is currently being investigated. These include Xenopus, mouse, chick, zebrafish, *C.elegans* and *D. melanogaster.* Each system has its own particular virtue, for example, the chick and Xenopus are the classic experimental embryological systems and are particularly suitable for micromanipulation (Dawid and Sargent, 1988; Klymkowsky and Karnovsky, 1994; Stern, 1994; Dupin *et al.*, 1998). In *C.elegans* the precise fate of each cell can be mapped, and this allows the study of multicellular processes at the level of individual cells (Kenyon, 1988; Hodgkin *et al.*, 1995). Zebrafish, a vertebrate model system, is amenable to both experimental embryology and thorough genetic analysis (Driever, 1994). *Drosophila melanogaster* is particularly amenable to genetic manipulation and analysis, much more so than mice: this has led to a rapid understanding of developmental mechanisms in the fly (Cooley *et al.*, 1988; Rubin, 1988; Gould *et al.*, 1990; St Johnston and Nusslein-Volhard, 1992). The discovery that some basic genetic pathways which establish the body plan have been conserved across species as diverse as insects and mammals has created an extremely important and powerful entry point for identifying murine developmental control genes: genes known to be important in fly development can be investigated to see whether their role is conserved in the mouse.

1.2 THE MOUSE AS A MODEL SYSTEM FOR DEVELOPMENTAL BIOLOGY

The mouse has long been the chosen mammal for genetic analysis for numerous reasons: its short gestation period, large litter sizes and its rapid generation time, the availability of inbred strains, the ability to perform controlled matings and the occurrence of mutants, either spontaneous, or physically or chemically induced.

High-resolution genetic linkage maps are now available for the mouse, covering four types of loci: mutations causing phenotypic variation, isozyme loci, cloned genes, and highly polymorphic anonymous DNA segments (Chapman and Nadeau, 1992; Copeland *et al.*, 1993; Dietrich *et al.*, 1996; Hunter *et al.*, 1996; McCarthy *et al.*, 1997). The generation of comprehensive murine genetic linkage maps has been greatly facilitated by the use of interspecific crosses that exploit the genetic diversity inherent among wild mouse species (Avner *et al.*, 1988; Breen *et al.*, 1994). Continuous refinement of the comparative map between human and mouse has accelerated both genetic analysis and the evolutionary study of genome organisation (Copeland *et al.*, 1993; O'Brien *et al.*, 1993; Carver and Stubbs, 1997; Lyons *et al.*, 1997).

The mouse has been used as an animal model for numerous human genetic diseases such as anaemias (Shehee *et al.*, 1993), polygenic diseases such as insulin dependent diabetes mellitus (Todd *et al.*, 1991), cancers such as intestinal neoplasia (Moser *et al.*, 1995) and birth abnormalities such as neural tube defects (Copp *et al.*, 1990).

The ability to introduce new genetic information into the germ line has advanced the mouse as a model for mammalian developmental biology (Capecchi, 1989; Thompson *et al.*, 1989; Ramírez-Solis *et al.*, 1993). Transgenic and embryonic stem (ES) cell technologies have allowed numerous novel functional analyses to be undertaken: (1) the creation of targeted germ-line gain-of-function and loss-of-function mutants; (2) ectopic gene expression; and (3) the analysis of *cis* and *trans* gene regulatory systems. Other methods for gene-silencing include antisense techniques (Augustine, 1997; Kirby *et al.*, 1997) and dominant negative studies (Krylov *et al.*, 1997; Celli *et al.*, 1998). Advances in techniques for studying spatial gene expression patterns in developing embryos have complemented these new technologies and include whole mount *in situ* hybridisation (Wilkinson and Nieto, 1993; Hauptmann and Gerster, 1994); *lacZ* transgenics to study endogenous expression, identify gene regulatory regions and cell marker studies (Smeyne *et al.*, 1992b; Knittel *et al.*, 1995; Serbedzija and McMahon, 1997); and refinement of antibody staining techniques

(Gitelman, 1997). Recent progress has also been made in producing conditional mutations and is discussed later (section 4.5.2.1).

Therefore the mouse has numerous advantages as a model for mammalian development, notwithstanding the limitations inherent to any model system. The conservation of some developmental processes amongst vertebrates suggests that discoveries in the mouse should give helpful indicative data for the human condition.

1.2.1 Strategies for identifying murine developmental control genes

A common theme in studying development is that by examining the abnormal or mutant phenotype, information on the underlying normal mechanism can be obtained. For the mouse this has been one of the main approaches, with hundreds of mouse mutants and their abnormal phenotypes being documented in the last century. These variants represent spontaneous mutations, as well as induced: the latter can be caused by chemicals, radiation and more recently using various insertional mutagenesis techniques (Flaherty, 1998). The main strategies for isolating murine developmental control genes are outlined below.

1.2.1.1 Spontaneous and induced mouse mutants

Spontaneous mouse mutants have been collected for over a century, initially by mouse fanciers and latterly by specialist laboratories such as The Jackson Laboratory (http://www.jax.org). To increase the number of mouse mutants available for study, large-scale mutagenesis experiments have been initiated using mutagens such as ionising radiation and chemicals (Flaherty, 1998). For example, chemicals such as ethylnitrosurea (ENU) act on stem cell spermatogonia, and can be used optimally to produce small lesions such as point mutations: a large scale mutagenesis programme is currently being undertaken using this approach at the MRC Mammalian Genetics Unit, Harwell, UK (http://www.mgc.har.mrc.ac.uk/mutabase). Some of the approaches for the genetic analysis of these mutations are outlined below.

1.2.1.2 Positional cloning of classical mouse mutations

A positional cloning strategy locates the mutant gene purely on the basis of its chromosomal map position (Collins, 1992, 1995). This involves linkage analysis, chromosomal walking from nearby genetic markers, identification of all transcripts in the candidate region and analysis of each of these for causative mutations in animals

with the abnormal phenotype. Developmental mutations cloned by this method include Brachyury (*T*) on chromosome 17 (Herrmann *et al.*, 1990) and short ear (*se*) on chromosome 9 (Kingsley *et al.*, 1992), caused by defects in the genes encoding T product and bone morphogenetic protein 5 (BMP5) respectively. This laborious approach has been greatly facilitated by the dramatic increase in marker density on current murine genetic and comparative maps (O'Brien *et al.*, 1993; Dietrich *et al.*, 1996), the availability of large insert libraries (Ioannou *et al.*, 1994; Monaco and Larin, 1994) and progress in methods of gene detection (Monaco, 1994) and mutation analysis (Mashal and Sklar, 1996). More recently, and undoubtedly in the future, direct sequencing of large genomic regions will also make a significant contribution.

1.2.1.3 Candidate gene strategy for cloning classical mouse mutations

More commonly, classical mouse mutations are cloned by the candidate gene approach: when a new gene is mapped to a particular chromosomal region, the phenotype of mutants in the vicinity is examined. If a phenotype is consistent with that predicted by a defect in the mapped gene, a mutation screen is undertaken. Mutations cloned in this fashion include dwarf (*dw*) on chromosome 16 (Li *et al.*, 1990; Camper *et al.*, 1990) and splotch (*Sp*) on chromosome 1 (Epstein *et al.*, 1991) caused by defects in the pituitary transcription factor, *Pit-1* and *Pax3* genes respectively. The rate at which mutations are cloned by this method, or in combination with a positional approach (Ballabio, 1993), should increase in parallel with the expansion of the gene-based map of the mouse genome.

1.2.1.4 Random insertional mutagenesis

The application of random mutagenesis to dissect out developmental pathways has been used very successfully in organisms such as *Drosophila* (Cooley *et al.*, 1988; Gellon *et al.*, 1997; Rorth *et al.*, 1998) or *Caenorhabditis elegans* (Plasterk, 1992; Jansen *et al.*, 1997). The principle behind such an approach has also been adopted for studying mammalian development (Friedrich and Soriano, 1993). The mouse mutant *limb deformity* (*ld*), and its affected gene *formin*, were identified fortuitously by the random insertion of a transgene causing a phenotypic mutation (Woychik *et al.*, 1990). There are many different approaches that can be used, some of which are discussed briefly below.

Retroviral vectors can be used as insertional mutagens. One advantage of retroviral infection is that it does not cause genome rearrangement. Simple retroviral vectors based on Moloney murine leukemia virus (MoMLV) have been used to infect mouse

embryos (both preimplantation *in vitro* and postimplantation *in utero*), with random insertion events causing mutation: the integration of the Moloney leukemia virus 13 (*Mov 13*) into the procollagen type I alpha 1 gene on chromosome 11 caused a recessive lethal mutation (Jaenisch *et al.*, 1983; Lohler *et al.*, 1984).

Other approaches have used "trapping" strategies to identify enhancers, promoters and genes that are developmentally regulated (Friedrich and Soriano, 1993). The transgene includes a reporter gene, commonly the *E.coli lacZ* gene encoding β-galactosidase, so that insertions into developmentally regulated genes can be identified: the addition of a chromogenic substrate, X-gal for β-galactosidase, defines the reporter gene's spatial-temporal expression pattern and by association, that of the trapped gene. This is independent of producing a mutant phenotype: breeding the transgenic mice to homozygosity may uncover a recessive mutant phenotype. In particular, promoter traps (or gene traps) have been widely used, and detect insertions into transcribed loci, against a background of random insertions (Gossler *et al.*, 1989; Friedrich and Soriano, 1991; Skarnes *et al.*, 1992; von Melchner *et al.*, 1992). This has been achieved using various approaches, but the common requirement is that the reporter gene is activated only if the trap integrates into the correct position to be transcribed under control of a genomic locus. Unlike enhancer trap vectors, the reporter gene lacks any transcriptional control elements. The route of choice for introducing promoter trap vectors is via ES cells. A novel gene, *jumonji*, which is thought to be required for the normal development of the neural tube was cloned by this approach (Takeuchi *et al.*, 1995). Transcriptional enhancer factor 1 (*TEF-1*) was disrupted with the retroviral gene trap vector, ROSAβ-geo, and revealed that TEF-1 is important in the later stages of cardiogenesis (Chen *et al.*, 1994).

Variations on the basic promoter trap design have been developed (Skarnes *et al.*, 1995), including a recent report of a high-throughput mutagenesis approach in ES cells with two significant improvements: it permitted automated sequence identification by 3'RACE of the coding region of the trapped gene, and the insertion of the gene trap was independent of the expression status of the trapped gene (Zambrowicz *et al.*, 1998). This approach, and other large scale gene trap strategies, are generating a resource of ES cell clones, with sequence-tagged mutations, which should greatly facilitate the functional characterisation of novel genes (Hicks *et al.*, 1997; Townley *et al.*, 1997).

1.2.1.5 *Targeted mutagenesis*

It is possible to introduce targeted mutations to specific loci in ES cells, which can then be used to contribute to a chimaera, and ultimately the murine germ line. The mutations are introduced by homologous recombination between the targeting vector and its homologous chromosomal target, an approach known as gene targeting. The first germ line transmission was reported by Thompson *et al.* (1989), at the hypoxanthine-guanine phosphoribosyltransferase gene (HPRT) and numerous mouse mutants have been produced since then that identify genes important for normal development (Brandon *et al.*, 1995; Copp, 1995). This technique has become almost a prerequisite for the functional analysis of a new gene, and should implicitly identify genes which are important in murine development. The genetic basis of classical mouse mutants has also been elucidated fortuitously by this approach: mice homozygous for targeted mutations in *Wnt*-1 were found to have a phenotype similar to that of the recessive mutation *swaying* (*sw*) (McMahon and Bradley, 1990; Thomas & Capecchi, 1990). The mutant *swaying* is characterised by ataxia and hypertonia attributed to malformation of anterior regions of the cerebellum. Also *Wnt*-1 was localised close to *sw* on mouse chromosome 15. Subsequently *sw* was shown to be due to a single base pair deletion in the proto-oncogene *Wnt*-1 (Thomas *et al.*, 1991).

There are now many variants in the ES cell mutagenesis approach including *in vivid* site-directed mutagenesis (Hasty *et al.*, 1991), deletion of large chromosomal regions (Ramirez-Solis *et al.*, 1995; Li *et al.*, 1996; Lewandoski and Martin, 1997; You *et al.*, 1997) and conditional targeting (Kuhn *et al.*, 1995), as well as combinations of these methods (Justice *et al.*, 1997; Shibata *et al.*, 1997). The refinement of targeting methods is allowing for a more thorough and subtle functional analysis than that achieved by a null mutation for a gene (Chen *et al.*, 1998; Meyers *et al.*, 1998).

1.2.1.6 *Cross-species conservation*

Murine developmental control genes have been isolated by utilising information from a more genetically accessible organism such as *Drosophila melanogaster* or *Caenorhabditis elegans.* Key developmental control genes of *Drosophila* have been cloned and characterised (St.Johnston and Nusslein-Volhard, 1992; Reichert and Boyan, 1997) in mouse, which has helped to confirm the belief that gene duplication and divergence play a primary role in evolution. Many developmental control genes can be grouped into families by the presence of common functional domains. These domains have been found to be conserved within gene families of diverse members of the animal kingdom. Sequences encoding these conserved motifs can be used as molecular probes to

isolate new family members from different species by low stringency screening of genomic and cDNA libraries. Three examples of conserved domains important for development are the homeobox (McGinnis *et al.*, 1984c; Scott and Weiner, 1984), the paired-box (Bopp *et al.*, 1986) and the POU-box (Herr *et al.*, 1988). Members of all the families characterised by these domains have been isolated from a wide variety of species using a low stringency screening approach (McGinnis *et al.*, 1984b; Deutsch *et al.*, 1988), and examples from the homeobox and paired-box will be discussed in detail in the next section.

The above approach can be successful, especially if the conserved motif is relatively large, as in the case of the Hox gene family. However drawbacks include spurious hybridisations, random homologies and the laborious isolation and analysis of potential novel family members. This has prompted the development of an alternative approach based on the polymerase chain reaction (PCR) using degenerate oligonucleotides as primers: the problem of differential codon usage has been overcome by designing mixtures of primers that encompass all possibilities for a conserved amino acid motif. These degenerate primers are used to amplify potentially related sequences, using genomic DNA or cDNA as template. The resulting product can be sequenced directly, or subcloned and sequenced to determine if it is related: it is therefore useful to have some degree of conservation in the intervening sequence. The product can be labelled and used directly as a probe for high stringency screening of genomic DNA and cDNA libraries, or it can extended using the 5' and 3' Rapid Amplification of cDNA ends (RACE) PCR protocol (Frohman *et al.*, 1988). Numerous developmentally regulated gene families have been expanded using this approach such as the *Wnt* -related family (Gavin *et al.*, 1990) and the *hedgehog* (*hh*)-related family (Chang *et al.*, 1994).

1.3 THE CONSERVED ROLE OF TRANSCRIPTION FACTORS IN DEVELOPMENT

There are numerous different gene families that have been found to be conserved between species in development, at least at the structural level. These include zinc-finger genes (Chowdhury *et al.*, 1988; Pabo and Sauer, 1992) and growth factors such as fibroblast growth factors (FGFs) (Mason, 1994). Two well-characterised and conserved transcription factor families that are important in murine development are described below.

1.3.1 HOM-C/Hox family

The HOM-C/Hox family has become the paradigm for conservation of developmental processes. It has been extensively studied in diverse organisms such as nematodes, arthropods and vertebrates. In 1894, Bateson described a set of homeotic mutations in *Drosophila*, whereby one body part was replaced by another which was normally located elsewhere. Most of the genes responsible map to two complexes, Antennapedia (ANT-C)and Bithorax (BX-C), collectively termed the Homeotic complex, or HOM-C. The genes in the HOM-C complexes were found to have a number of strikingly similar features, both structurally and functionally: the most important of these is the homeobox which is a 183 bp domain that encodes a DNA binding domain of 61 amino acids (McGinnis *et al.*, 1984c, Scott and Weiner, 1984). There are mammalian homologues of the HOM-C, termed the Hox complexes, which are located in four chromosomal clusters (MMU2, 6, 11 and 15; McGinnis *et al.*, 1984a,b; McGinnis and Krumlauf, 1992). The homeodomain has been found in many developmentally regulated genes and different classes have been identified (Scott *et al.*, 1989; Gehring *et al.*, 1994). Similarities within the homeodomain and immediately flanking sequences have allowed various paralogous groups to be defined amongst the Hox clusters.

Hox genes are thought to be involved in specifying anteroposterior positional identities in the developing vertebrate embryo. The correlation between the physical order of the genes along the chromosome and their expression/function along the anteroposterior axis of the embryo is very distinctive, and was first noted by Lewis *et al.* (1978) in the *Drosophila* BX-C, and referred to as colinearity. The colinearity phenomenon is both spatial and temporal: the more 3' the position of a gene within the cluster, the

earlier and more anteriorly it is expressed, and the more sensitive it is to retinoic acid whereas the more 5' the position of a gene within the cluster, the later and more posteriorly it is expressed, and the less sensitive it is to retinoic acid. The conservation of colinearity between *Drosophila* and mammalian homeobox clusters was crucial in establishing a homologous relationship at both the structural and functional levels. It seems likely that it is related to a common molecular mechanism for providing a co-ordinated system of axial signals involved in generating different regional identities.

The striking conservation in structure begs the question, does the conservation also extend on a functional level, and if so, how far? Naturally occurring/easily generated mutations that help to define specific functions rarely occur in human and mouse, unlike *Drosophila*. With the advent of gene targeting technology, both gain-of-function and loss-of-function mutations for the same locus can be generated, and mutant crosses undertaken to overcome any functional redundancy effects. This has yielded valuable functional information and some examples are outlined in the table below (Krumlauf, 1994; Brandon *et al.*, 1995). Of note also is the recent implication of Hox gene mutations in human genetic diseases such as synpolydactyly and hand-foot-genital syndrome (Muragaki *et al.*, 1996; Mortlock *et al.*, 1996; Mortlock and Innis, 1997; Goodman *et al.*, 1997).

Gene	Loss of function	Gain of function	Phenotype	References
Hoxa-1	+		Perinatal lethal; hindbrain reorganisation; cranial nerve and inner ear defects	Chisaka *et al.*, 1992; Lufkin *et al.*, 1991; Carpenter *et al.*, 1993; Doll'e *et al.*, 1993
Hoxa-2	+		Perinatal lethal; homeotic transformation of rostral head	Gendron-Maguire *et al.*, 1993; Rijli *et al.*, 1993
Hoxb-5	+		Rostral shift in shoulder girdle; homeotic transformation of vertebrae C6 through T1	Rancourt *et al.*, 1995
Hoxc-8	+		Neonatal lethality; skeletal transformations	Le Mouellic *et al.*, 1992
Hoxc-8		+	anterior transformation	Pollock *et al.*, 1992

Table 1.1 Transgenic analysis of Hox gene function

To date, the elucidated mechanisms for the establishment of the Hox gene expression pattern are not conserved between *Drosophila* (Gellon *et al.*, 1998) and vertebrates (Krumlauf, 1994; Nonchev *et al.*, 1996; Manzanares *et al.*, 1997). However, the maintenance of Hox gene expression and function have cross-species correlates: the Polycomb group (Pc-G) and trithorax (trx-G) gene families assume the stable

maintenance of their expression domains (Kennison 1993; Paro, 1993; Simon, 1995; van der Lugt *et al.*, 1996), and the Exd/Pbx family of proteins assist Hox proteins to achieve specificity (van Dijk and Murre, 1994; van Dijk *et al.*, 1995; Popperl *et al.*, 1995). The downstream targets of the *Drosophila* Hox genes are now being investigated (Graba *et al.*, 1997) and the insights gained may assist in identifying targets in vertebrates, if similar conservations of sequence and function have occurred.

1.3.2 The PAX family

1.3.2.1 General features of the PAX family

The DNA sequence characteristic of the PAX family was first recognised in the *Drosophila* pair-rule gene, *paired* (*prd*) (Frigerio *et al.*, 1986). This DNA binding motif, termed the paired-box (Bopp *et al.*, 1986), consists of approximately 384 nucleotides, encoding a paired domain of 128 amino acids. The three dimensional structure of the paired domain has been predicted by computer analysis to contain three α-helices (Xu *et al.*, 1995), unlike the helix-turn-helix or helix-loop-helix motifs. Another feature common to many of the PAX family is a paired-type homeodomain, similar to other classic homeodomains in its 3D structure, a helix-turn-helix motif, but distinct from them at the nucleotide level. A third highly conserved motif, common to some of the PAX family, is an octapetide (Burri *et al.*, 1989), which may be involved in transcriptional repression (Lechner and Dressler, 1996). The combination of the three highly conserved domains (especially the type of paired-box motif), as well as the spatial/temporal expression pattern of each gene, allows vertebrate PAX family members, as in *Drosophila* , to be loosely placed into paralogous groups (Table 1.2). Similarities also exist in the functional roles that PAX members may have in development in both *Drosophila* and vertebrates, and within their paralogous groups, as all PAX family members are thought to have a crucial role in specification during embryonic development (Gruss and Walther, 1992; Strachan and Read, 1994). Nine *PAX* genes have been described in both mouse and man (*Pax1* to *Pax 9*, and *PAX1* to *PAX9*, respectively) (Walther *et al.*, 1991; Strachan and Read, 1994), and are localised randomly throughout the genome. Loss-of-function studies suggest that there is less functional redundancy than in the vertebrate HOX family, which may account for the occurrence of numerous pathological mutations in *PAX* genes (Table 1.2), whilst fewer and milder mutations have been found for HOX genes (Goff and Tabin, 1996).

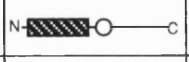
Pax genes	Basic structure: Paired box / OP / Homeo-box	Localisation: Mouse	Localisation: Human	Mouse mutant: Natural	Mouse mutant: Targeted	Human syndrome	References
Group I		2	20p11	Undulated	Not published	Spina bifida	Balling *et al.*, 1988; Hol *et al.*, 1996
Pax1		12	14q12-q13	Not known	Not published	Not known	
Pax9							
Group II		19	10q25	Pax2 1Neu	Yes	Renal coloboma syndrome	Urbanek *et al.*, 1994; Sanyanusin *et al.*, 1995; Torres *et al.*, 1995; Favor *et al.*, 1996
Pax2		2	2q12-q14	Not known	Not known	Not known	
Pax8		4	9p13	Not known	Yes	Not known	
Pax5							
Group III		1	2q35	Splotch	Not known	Waardenberg syndrome I, III	section 1.3.2.3; Mansouri *et al.*, 1996
Pax3		4	1p36.2	Not known	Yes	Not known	
Pax7							
Group IV		6	7q32	Not known	Yes	Not known	Sosa-Pineda *et al.*,1997; section 1.3.2.2
Pax4		2	11p13	Small eye	Not known	Aniridia Peter's anomaly	
Pax6							

Table 1.2 Summary of vertebrate Pax gene family
Reproduced from Dahl et al., 1997

Two members of the PAX family, with naturally occurring mouse and human mutations, are described in more detail to illustrate the conserved role of the PAX family in development, and in particular, organogenesis.

1.3.2.2 The conserved role of PAX6 in eye development

Mutations in *PAX6* cause aniridia (Jordan *et al.*, 1992) and Peter's anomaly (Hanson *et al.*, 1994) in humans, and underlie the genetic defect in the classical mouse mutant *Small eye* (*Sey*) (Hill *et al.*, 1991). The majority of aniridia cases are familial, with autosomal dominant inheritance, and the *Sey* mutation is semi-dominant. Overexpression of *Pax6* in transgenic mice causes a similar phenotype to *Sey* (Schedl *et al.*, 1996). Together, this suggests that PAX6 is extremely dosage-sensitive, a finding common to some other PAX proteins (Sanyanusin *et al.*, 1995; Read and Newton, 1997). In mice, *Pax6* is expressed in discrete regions of the developing brain from 8 dpc onwards, including the lens and nasal placodes (Walther and Gruss, 1991; Grindley *et al*, 1995). The *Drosophila* homologue of PAX6 is *eyeless*, identified in a mutant of the same name: *eyeless* homozygotes have a reduction or complete absence of eyes (Quiring *et al.*, 1994). *eyeless* is expressed in a similar pattern to its vertebrate homologue, *Pax6*, and both vertebrate and fly *Pax6* genes are expressed at least transiently in all tissues of the developing eye. Targeted misexpression of *eyeless* in *Drosophila* induced ectopic eyes on legs, wings and antennae, as did the misexpression of murine *Pax6* in *Drosophila* (Halder *et al.*, 1995). These experiments highlight both the strong degree of functional cross-species conservation for *Pax6*, and its essential role in eye development.

1.3.2.3 PAX3 , MITF and Waardenberg syndrome

Waardenberg syndrome (WS) is an autosomal dominant auditory-pigmentary disorder, that has four phenotypic subclassifications (Read and Newton, 1997). Mutations in *PAX3* are responsible for WS Type I (WS1) and Type III (WS3), causing loss of function and haploinsufficiency in WS1. The phenotypic variation apparent in WS1 probably depends upon the precise PAX3 dosage (Read and Newton, 1997), but there is no clear genotype/phenotype correlation suggesting that modifier loci may be involved (Asher *et al.*, 1996). The auditory-pigmentary syndromes are caused by absence of melanocytes, and for the majority of WS the defect is associated with disorders of the neural crest, the precursor of melanocytes. Similarly, other structures affected in WS such as limb muscles and enteric ganglia are neural crest derivatives. WS mutations are found throughout the protein, and analysis of missense mutations suggest that mutations in each binding domain can influence the binding activity of the other unaffected DNA binding domain (Fortin *et al.*, 1997).

Splotch (*Sp*) mice are a murine model for neural tube defects (Copp *et al.*, 1990), and there are a number of variant alleles. The *splotch* mutation is semi-dominant and homozygotes (for alleles Sp, Sp^{1H}, Sp^{2H}) have a number of abnormalities including exencephaly, meningocele, spina bifida and defects in the heart, limb musculature and neural crest cell-derived spinal ganglia (Copp *et al.*, 1990; Franz, 1993). All heterozygotes are characterised by white spotting of the abdomen, tail, and feet, possibly due to the absence of melanocyte migration to these regions. Analysis of the Sp^{2H} allele identified a deletion in the *Pax3* gene, in the region encoding the paired homeodomain, creating a truncated protein (Epstein *et al.*, 1991). *Pax3* is expressed from day 8 to day 17 of mouse development, mainly in developing neural tissue, and including neural crest cells (Goulding *et al.*, 1991). It is also found in the segmented mesoderm and limb buds, the latter site potentially explaining the limb abnormalities seen in WS3 and *splotch* mice (Bober *et al.*, 1994). Interestingly, the tyrosine kinase receptor, c-*met*, which is required for murine limb muscle development (Bladt *et al.*, 1995), has been shown to have markedly reduced expression in *Splotch* embryos (Epstein *et al.*, 1996). PAX3 was also shown to bind and transactivate the human c-*MET* promoter, and so PAX3 may directly regulate c-*MET* expression during limb muscle development (Epstein *et al.*, 1996).

Another potential target for PAX3 is the bHLH-Zip (basic helix-loop-helix-zipper) gene *MITF* (**mi**cropthalmia **t**ranscription **f**actor) (Tachibana *et al.*, 1994). *MITF* is the human homologue of *mi*, the gene mutated in the classical mouse mutant *micropthalmia* (*mi*) (Hodgkinson *et al.*, 1993; Hughes *et al.*, 1993). *mi/MITF* are

thought to be critical for melanocyte differentiation (Tachibana *et al.*, 1996; Opdecamp *et al.*, 1997) and mutations in *MITF* are the underlying genetic defect in a proportion of WS2 families (Tassabehji *et al.*, 1994). Evidence for a potential pathway for some of the biological effects of *PAX3* mutations in WS1 has been reported recently: PAX3 can transactivate the *MITF* promoter, whereas WS1-type mutant PAX3 proteins fail to do so. The inability of WS1-type mutant PAX3 to activate *MITF* gene expression could result in a failure in melanocyte development, with the concomitant biological sequelae of hypopigmentation and hearing impairment seen in WS1 (Watanabe *et al.*, 1998).

1.4 THE BASIC-LEUCINE ZIPPER (bZIP) FAMILY

The basic-leucine zipper (bZIP) family is another example of a transcription factor family (Pabo and Sauer, 1992) that has a structural DNA binding motif conserved across species as diverse as fungi, plants, flies and vertebrates (Hurst, 1995). The bZIP family of transcription factors bind as dimers to specific DNA recognition sites within the proximal promoter and enhancer regions of transcribed genes and thus regulate their expression. In general, they are activators of transcription, although a few such as the *Drosophila giant* and mammalian E4BP4 act as repressors (Kraut and Levine, 1991; Cowell *et al.*, 1992). However transcriptional activators can also have a repressive effect (Hsu *et al.*, 1994; Gius *et al.*, 1990) and the precise effect of transcription factor binding *in vivo* will depend on many variables such as phosphorylation, protein isoform and dimerisation partner. Many bZIP proteins are cell type specific or developmentally regulated, and are involved in developmental and differentiative processes.

1.4.1 The structural basis of the bZIP family

The term "leucine zipper" was coined by McKnight and colleagues in 1988, to describe a DNA-binding motif that was to common to four proteins: the transcription factors, yeast GCN4 and the CCAAT/enhancer binding protein (C/EBP) and the oncogenes, FOS and JUN (Landschulz *et al.*, 1988). The motif spans 55-65 amino acids and consists of two distinct regions: a basic DNA-binding domain and an adjacent "leucine zipper" region, responsible for dimerisation of the protein.

1.4.1.1 The basic DNA-binding domain

The more conserved basic DNA-binding domain (covering approximately 24 amino acids) is towards the amino-terminus of the protein and consists of many arginines and lysines, as well as other residues conserved throughout the family or particular subfamilies. The precise DNA binding site recognised is determined principally by the combined basic regions of the protein dimers, but it is influenced by other factors such as stability of the dimer complex and protein phosphorylation (Hurst, 1995). The DNA binding site shows dyad symmetry, and may be thought of as two half sites, each of which is contacted by the basic region of one of the dimer subunits. Each bZIP protein has its own preferential binding site, but can bind to a range of sites. In fact, as long as

one half site is close to the optimal binding sequence, the other half site might vary quite widely. With the advent of *in vitro* random-binding site selection assays, a list of allowable binding sites for each bZIP protein has been determined, but the biological relevance of this information still has to be determined. The basic region forms an α-helix that contacts the major groove of the DNA (Ellenberger, 1994).

1.4.1.2 The leucine zipper dimerisation region

The leucine zipper region contributes to DNA binding specificity by determining which subunits will form stable dimers, and by appropriately positioning the basic region helices over the binding site. The 30-40 amino acid leucine zipper region consists of a heptad repeat of leucines, between 3 and 6 in number. The periodicity of the leucines suggested to Landschulz *et al.*(1988) an α-helical arrangement, with seven amino acids every two helical turns, creating a hydrophobic face of leucines which would allow monomers to dimerise, with the leucines interdigitating like a zipper. The helix of GCN4 is represented below using a standard format, permitting a description of the relative position of the other amino acids within the zipper region (Figure 1.1). Position **d** is predominantly occupied by leucines, and residues at both positions **a** and **d** are usually hydrophobic, although small apolar residues are acceptable (Kouzarides and Ziff, 1989). Polar residues at either position do occur but are destabilising (Alber, 1992), the extent to which is dependent on the precise position and on the choice of dimerisation partner (Hu *et al.*, 1990; van Heeckeren *et al.*, 1992). For example, many bZIP proteins interrupt this hydrophobic repeat at position **a** with a polar asparagine towards the centre of the zipper (Hurst, 1995). Charged residues usually occur at positions **e** and **g**: electrostatic interactions between the opposing **e** and **g** residues (**e'** and **g**; **e** and **g'**) are thought to contribute to the stability (or instability) of the dimer (Vinson *et al.*, 1993). For example, the selective heterodimerisation of FOS and JUN can be accounted for in part by charged residues at positions **e** and **g** (Schuermann *et al.*, 1991; O'Shea *et al.*, 1992). In summary, residues at positions **a**, **d**, **e** and **g** are important for the stability and specificity of the dimer and although individual residues are not as conserved as for some other DNA binding motifs, the typical sequence arrangements described can help to determine the authenticity of any potential leucine zipper domain in a novel polypeptide. The significance of the residues at positions **b**, **c** and **f** is at present unknown.

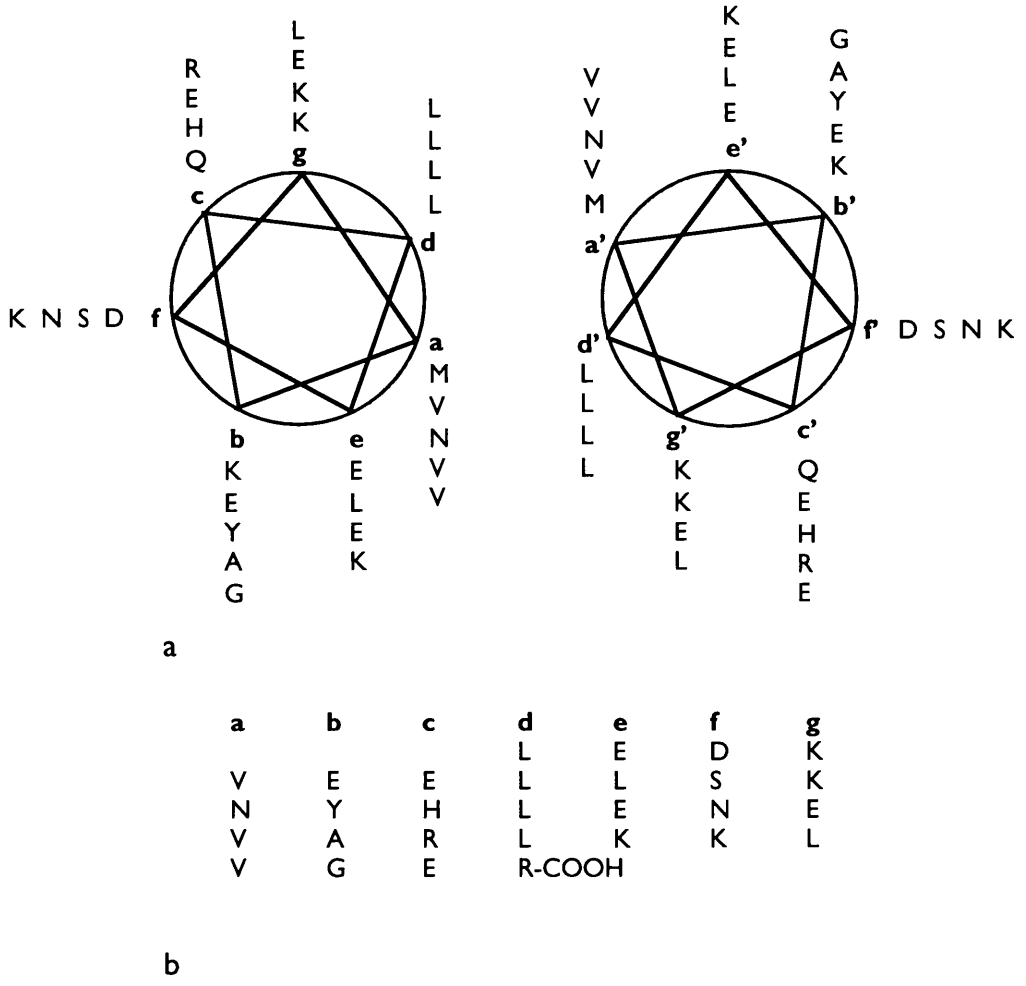


Figure 1.1 Helical wheel of GCN4 zipper region
(a) Helical wheel representation of the GCN4 zipper region (b). The heptad repeat of leucines is at position d. The position of the residues is shown relative to a potential interacting GCN4 monomer. Positions e and g are thought to be important for stabilisation (or destabilisation) of the dimer, and therefore affect dimer specificity.

Biochemical studies suggested that the leucine zipper would form two parallel α-helices in a coiled-coil arrangement (O'Shea *et al.*, 1989), and this was confirmed upon determination of a high resolution structure of the GCN4 leucine zipper (O'Shea *et al.*, 1991). The crystal structure of GCN4 in complex with two DNA sites has also been determined (Ellenberger *et al.*, 1992; Konig and Richmond, 1993), and it is thought that many of the conclusions about structure will be true for the rest of the bZIP proteins (Kerrpola and Curran, 1991). The leucine zipper region has been shown to form amphipathic α-helices, in a parallel coiled-coil arrangement, with the overall appearance being that of an α-helical fork, similar to the proposed "scissors-grips" model (Vinson *et al.*, 1989).

1.4.1.3 Other common features of bZIP proteins

Other domains outwith the bZIP region are responsible for transcriptional activation (or repression), and can vary widely in structure (Hurst, 1995). The most common post-translational modification to bZIP proteins is phosphorylation (Hurst, 1995): probably all of these proteins are phosphoproteins, and often the phosphorylation sites co-localise to protein domains involved in transcriptional activation, causing either activation or repression of transcriptional activity (Sutherland *et al.*, 1992; Chen *et al.*, 1993; Bannister *et al.*, 1994). Alternatively the DNA binding of the protein can be altered (Lin *et al.*, 1992)

Notwithstanding the lack of widespread sequence similarity within the bZIP family of transcription factors, some members of this group can be placed into subgroups: this classification is decided mainly upon sequence similarities, both within and outside the bZIP domain. The families defined by two of the founding genes, *fos* and *jun* are described briefly below.

1.4.2 The Fos and Jun families

The mammalian bZIP subfamilies of FOS and JUN have been studied intensively. Both FOS and JUN are nuclear proto-oncogenes and have several related family members: together in a number of combinations, these can heterodimerise, and form the AP-1 binding activity (Curran and Franza, 1988; Rauscher *et al.*, 1988). The AP-1 binding site is found in the enhancer and promoter regions of a great number and variety of genes, suggesting many roles for FOS/JUN heterodimers in differentiation, development, and the cell cycle (Angel and Karin, 1991; Karin *et al.*, 1997). They are all classed as immediate early response genes (IE), as they are rapidly induced in mitogen-stimulated quiescent cells, with the exception of JunD.

1.4.2.1 The Fos family

The founding member, *c-fos* is the cellular homologue of the oncogene *v-fos*, isolated from the FBR and FBJ murine osteogenic sarcoma viruses (MSVs) (Curran *et al.*, 1983).The Fos family, comprising *c-fos*, *fosB*, *fra1* and *fra2*, share sequence similarity mainly in the bZIP domain, and will only bind DNA specifically as heterodimers. Their heterodimer partners belong to the Jun family, and also the Maf, ATF2 and ATF4 families. Their inability to homodimerise is due substantially to

residues in the leucine zipper motif: of particular importance are the residues at positions **e** and **g** (Figure 1.1), which are potentially involved in interhelical electrostatic interactions (Vinson *et al.*, 1993). In FOS, the majority of these residues are acidic, leaving a net charge of -5 on each monomer: thus homodimers are discouraged by electrostatic repulsion. Mutagenesis studies have confirmed this notion, as replacement of the first or second glutamic acid residue to a lysine, allows the modified FOS to bind DNA, presumably as a dimer (Nicklin and Casari, 1991). In contrast JUN has a slight positive charge which helps to stabilise FOS/JUN heterodimers. It has also been shown that sequences outwith the leucine zipper can affect the choice of partner for example, mutation of a histidine to an alanine residue seven bases distal to the last leucine results in decreased dimerisation (Cohen and Curran, 1990).

As mentioned previously there is substantial *in vitro* evidence that *fos* may be involved in a variety of biological processes (Angel and Karin, 1991), and its expression pattern suggests potential developmental roles in bone formation (Dony and Gruss, 1987) and neurogenesis (Caubet, 1989; Smeyne *et al.*, 1992a). Further *in vivo* evidence for a specific function in skeletal tissue has accumulated: overexpression of *c-fos* in transgenic mice caused abnormalities in bone and thymic epithelium (Ruther *et al.*, 1988, 1989). The bone abnormalities were osteosarcomas of a similar phenotype to that generated following injection into rodents of *v-fos* containing FBJ- and FBR-MSVs (Ward and Young, 1976). Chimaeric mice produced using ES cells overexpressing *c-Fos* also suffered from tumours, but these were chondrosarcomas (Wang *et al.*, 1991): the type of tissue affected may reflect a difference in timing of the overexpression of the transgene. Transgenic mice overexpressing *fos* were crossed with mice, either heterozygotes or homozygotes for a null *Fos* allele, and the offspring had dramatically reduced frequency of osteosarcoma formation and time of onset, suggesting a critical level of FOS and AP-1 activity for osteoblast transformation (Grigoriadis *et al.*, 1995).

After inactivating the endogenous *Fos* gene (Wang *et al.*, 1992; Johnson *et al.*, 1992), it was noted that although *Fos* was not essential for embryonic development, the *Fos* -/- homozygote mice developed osteopetrosis, which is described as an increase in the number of tissue macrophages and a lymphopenia secondary to the bone disease. Osteopetroses are disorders of bone remodelling characterised by impaired osteoclast function, resulting in a net increase in skeletal mass. These diseases are heterogeneous in origin, but in the case of the *Fos* knockout, it was found to be due to a complete absence of functional osteoclasts and their immediate precursors (Grigoriadis *et al.*, 1994). Interestingly, osteoclasts and macrophages are thought to have a common origin

in haematopoietic stem cells (Suda *et al.*, 1992) whereas osteoblasts and chondroblasts differentiate from earlier multipotential mesenchymal stem cells (Grigoriadis *et al.*, 1988). The transgenic data showed that osteoblasts and chondroblasts were highly susceptible to transformation by elevated FOS levels, but the knockout model showed them to be independent of the presence of FOS: the latter could be explained by another Fos-related family member being able to substitute for FOS. In contrast, *Fos* expression is essential for osteoclast differentiation, and the increase in tissue macrophages suggest a role for FOS as a negative regulator in macrophage differentiation.

In *Drosophila*, a homologue has been isolated, dFRA (Fos related antigen), which can homodimerise and bind specifically to DNA, and may be involved in neurogenesis (Perkins *et al.*, 1988, 1990). It was recently shown that dFRA and dJRA (Jun related antigen; section 1.4.2.2) heterodimerise during development to form an essential activity for dorsal closure during mid-embryogenesis, by regulating *decapentaplegic* (*dpp*) expression (Riesgo-Escovar and Hafen, 1997; Zeitlinger *et al.*, 1997). DFRA expression in the lateral epithelium is itself dependent upon DPP activity. DFRA also has a jun-independent function in mediating the expression of DPP target genes in early embryogenesis, possibly acting as a homodimer. The interplay of DPP, DFRA and DJRA has parallels in vertebrate signalling, involving AP-1 and TGFβ1(Kim *et al.*, 1990; Woessner, 1991), and is a good example of an evolutionary conserved regulatory pathway.

1.4.2.2 The Jun family

The founding member, *c-jun* is the cellular homologue of *v-jun*, the transforming oncogene of the avian sarcoma virus 17 (Maki *et al.*, 1987). The family (*c-jun*, *junB* and *junD*) are related by a region of sequence similarity over the C-terminal halves of the genes, including the bZIP domain. The jun family can bind DNA as homodimers, though JUNB homodimers are less stable than C-JUN and JUND homodimers, which are themselves in turn less stable than the FOS/JUN heterodimers. Expression studies suggested a developmental role for *c-jun* and *junB* (Wilkinson *et al.*, 1989b). Knockout experiments have shown *c-jun* to be essential for normal mouse development: mice embryos lacking *c-jun* die mid- to late-gestation and show impaired hepatogenesis (Hilberg *et al.*, 1993; Johnson *et al.*, 1993).

In *Drosophila*, a homologue has been isolated called dJRA (Jun related antigen) (Zhang *et al.*, 1990), that is constitutively expressed throughout development. However it is involved in at least two specific developmental pathways. Firstly, it is thought to act

along with the ETS domain protein, POINTED, on a common set of target genes to induce photoreceptor R7 fate in the developing *Drosophila* eye (Treier *et al.*, 1995). Secondly it is involved in dorsal closure as mentioned above (Riesgo-Escovar and Hafen, 1997; Zeitlinger *et al.*, 1997).

1.4.3 bZIP proteins in development

A few examples of bZIP proteins which are important in development and relevant to this thesis are outlined in the following sections.

1.4.3.1 Drosophila cnc

The *Drosophila* gene, *cnc* (**c**ap'**n**'**c**ollar), was originally identified as an abundantly transcribed locus in a chromosome walk (Mohler *et al.*, 1991). *cnc* encodes a bZIP transcription factor, CNC, with two notable features: a C-terminal glutamine-rich region that may act as an activation domain; and a bZIP motif which, upon sequence analysis, suggests that CNC functions as an obligate heterodimer. *cnc* is expressed during early embryogenesis in two sites: in an anterior dorsal cap and in a narrow stripe caudal to the cap, defined as the labral and mandibular segments respectively. The restricted expression of *cnc* largely persists throughout embryogenesis, and suggested a role for CNC in head specification (Mohler *et al.*, 1991). This was later confirmed by the isolation and characterisation of *cnc* mutants (Mohler *et al.*, 1995): in the mandibular segment, a classical homeotic effect was evident, with missing mandibular structures being replaced with duplicate maxillary structures. Labral structures were also absent in *cnc* mutant larvae: the labral primordium fused with the oesophageal primordium to contribute to formation of the oesophagus. Therefore *Drosophila cnc* appeared to be acting as a segment-specific selector gene controlling the identity of two cephalic segments.

1.4.3.2 The Maf family

The Maf family is a subclass of the bZIP family of transcription factors, whose founding member, v-*maf*, was originally discovered as the transduced transforming component of avian musculoaponeurotic fibrosarcoma virus, AS42 (Nishizawa *et al.*, 1989). The cellular homologue of v-*maf*, c-*maf*, was subsequently identified (Kataoka *et al.*, 1993), and established an extended family of related genes which all share a well conserved bZIP motif. The large Maf proteins c-MAF, MAFB (Kataoka *et al.*, 1994b), NRL (**n**eural **r**etina **l**eucine zipper) (Swaroop *et al.*, 1992) and L-MAF

(Ogino and Yasuda, 1998) all contain an acidic activation domain, as well as an N-terminal extended region of homology adjacent to the basic DNA-binding domain (Kerrpola and Curran, 1994). The small Maf proteins, MAFK (Fujiwara *et al.*, 1993; Andrews *et al.*, 1993b; Igarashi *et al.*, 1995), MAFF (Fujiwara *et al.*, 1993), and MAFG (Kataoka *et al.*, 1995; Blank *et al.*, 1997; Marini *et al.*, 1997) all lack an activation domain. Maf proteins can form homodimers, and intra- and inter-family heterodimers. They have been shown to interact with FOS and JUN, and other bZIP proteins (Igarashi *et al.*, 1994; Kerppola and Curran, 1994; Kataoka *et al.*, 1995; Motohashi *et al.*, 1997), as well as other transcription factors such as c-MYB and ETS1 (Sieweke *et al.*, 1996; Hedge *et al.*, 1998). Many members of the Maf family have been implicated in the regulation of lineage-specific gene expression, and a few examples involving the large Maf proteins are described below. The small Maf family are discussed in more detail in section 4.1.5.1.

I MafB

The *kreisler* (*kr*) mouse mutant has inner ear abnormalities, a consequence of abnormal hindbrain segmentation (Frohman *et al.*, 1993; McKay *et al.*, 1994). Mutation of *MafB* is the underlying genetic defect in *kr* mice (Cordes and Barsh, 1994). *Hoxb-3* is a direct target of MAFB, acting in combination with a possible Ets-related protein, to direct rhombomere (r) 5-specific expression of *Hoxb-3* in the hindbrain (Manzanares *et al.*, 1997). Also MAFB has been shown to interact with ETS-1 and act repressively in order to inhibit erythroid differentiation in chickens (Sieweke *et al.*, 1996). The same researchers have also demonstrated that the *kr* mouse is not a null *MafB* mutant, but a tissue-specific deficit (Eichmann *et al.*, 1997). Another large Maf protein, c-MAF is also involved in lineage-specific regulation (Hegde *et al.*, 1998).

II L-maf *and* NRL

Chicken *L-maf* was recently isolated and characterised: L-MAF is capable of inducing lens-differentiation through direct binding to lens-specific genes such as *αA-crystallin* (Ogino and Yasuda, 1998). Its expression is initiated in the lens placode and is restricted to lens cells. It is most similar to chicken *MafB*, and is distinct from human and mouse *NRL*: no chicken ortholog for *NRL* has been isolated. *NRL* also has a restricted expression pattern during murine embryogenesis, suggestive of a role in neuronal differentiation, and is restricted to adult human and mouse retina (Swaroop *et al.*, 1992; Hsieh *et al.*, 1996; Liu *et al.*, 1996). This pattern is consistent with a role in regulation of retinal function and it can positively regulate rhodopsin, a photo-receptor cell specific gene (Kumar *et al.*, 1996; Rehemtulla *et al.*, 1996). A recent

study of a lens-specific guinea pig promoter suggested that NRL might act as a co-factor or interactor with PAX6 to regulate precise tissue-specific gene expression (Sharon-Friling *et al.*, 1998).

1.5 ISOLATION OF 8dI

A murine cDNA was isolated by Dr. K. A. Johnstone during the development of a novel method to isolate species-specific expressed sequences from somatic cell hybrids (Johnstone, 1995). Species-specific repeats are abundant in heterogeneous nuclear (hn) RNA and hncDNA libraries have been exploited as a method of isolating human expressed sequences. The technique was based upon a novel reverse transcriptase-polymerase chain reaction (RT-PCR) approach using hncDNA synthesised from somatic cell hybrid hn RNA, and was designed to be both species- and coding sequence-specific.

Briefly, hncDNA was synthesised with exon-specific 5' splice-site oligonucleotides and the products were amplified in a PCR between the 5' splice site and a human-specific Alu oligonucleotide (Fuentes *et al.*, 1997). The hybrid cell line WAI7 (Raziuddin *et al.*, 1984), used as the source of poly(A)$^+$ RNA, contained three copies of human chromsome 21 on a mouse A9 fibroblast background. The initial attempt at generating a species-specific hncDNA library resulted in the majority of isolated clones being murine single-copy sequences as determined by sequencing and Southern analysis: the human clones obtained contained human repeats. The murine products appeared to have arisen by amplification of cDNAs between splice site oligonucleotides and were cloned by internal sites for *Eco* RI or *Sst* II.

One of the murine clones, 8dI, was an 800 bp fragment which upon preliminary sequencing through the 'Alu' end was found to have an internal *Eco* RI site, and a plus strand ORF: the translation had 80% similarity with a *Drosophila* protein known as CNC over 63 amino acids (Mohler *et al.*, 1991; section 1.4.3.1). The similarity with CNC was in the region of a basic leucine zipper (bZIP) motif. The reading frame in 8dI remained open after the region of similarity to CNC. The initial NBRF database comparison for 8dI and CNC is shown Figure 1.2.

```
> SW:CNC_DROME P20482 SEGMENTATION PROTEIN CAP'N'COLLAR.
          Length = 533

Score = 81.0 bits (207), Expect = 1e-16
Identities = 43/63 (68%), Positives = 51/63 (80%)

8dI    1   FNELLSKYQLSEAQLSLIRDIRRRGKNKMAAQNCRKRKLDTILNLERDVEDLQRDKARLL 60
           FNE LSKY LSE QLSLIRDIRRRGKNK+AAQNCRKRKLD IL LE +V  + + K +L
CNC    329 FNERLSKYDLSENQLSLIRDIRRRGKNKVAAQNCRKRKLDQILTLEDEVNAVVKRKTQLN 388

8dI    61  REK 63
           +++
CNC    389 QDR 391
```

Figure 1.2 BLAST analysis of 8dI
Initial sequence data from 8dI (259 bp) was analysed using BLAST (Altschul et al., 1990) against the NBRF and SwissProt protein databases. The highest hit was with Drosophila CNC, and lower significant scores with jun-related (Ryder et al., 1988) and fos-related (Franza et al., 1987) proteins.

The 8dI clone was used to screen 500,000 plaques of a lambda bacteriophage 8.5 dpc mouse embryo cDNA library. At a stringency of 0.5x SSC, 0.1% SDS, 8dI identified 6 tertiary positives. These were confirmed by hybridisation of 8dI to *Eco* RI digests of phage clones (Johnstone, 1995). The restriction endonuclease digest and hybridisation was confirmed subsequently by myself (section 3.1).

1.6 AIMS OF THESIS

The isolation of the murine clone, 8dI has identified a potential murine homologue for the Drosophila gene, *cnc*, which has been named *Nfe2l1*. As discussed earlier, *cnc* is expressed in a segmental pattern in anterior structures during early *Drosophila* development, suggesting a role in specification of head structures. The aims of the thesis are to determine if 8dI is indeed the murine homologue of *Drosophila cnc* , and to characterise its potential role in mouse development. This shall include a detailed examination of *Nfe2l1* mRNA expression pattern during murine development. The localisation of *Nfe2l1* in both mouse and human genomes shall be determined, and its potential role in any mouse and human pathologies considered.

CHAPTER 2 MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 Reagents

These were typically either of AnalaR grade from British Drug Houses (BDH, Poole, Dorset) or of molecular biology grade from Sigma Chemical Company. Phenol and sodium hydroxide were from Fisons. Agar, yeast extract and tryptone were from Difco Laboratories. Dextran sulphate, Ficoll 400, oligo-(dT) cellulose, deoxyribonucleotides, random hexanucleotides and Sephadex G-50 were from Pharmacia LKB Biotechnology. Deionised formamide was from International Biotechnologies Incorporated (IBI). DIG-11-UTP, BCIP, and NBT were from Boehringer Mannheim. Commercial competent *E.coli* cells were from Betheseda Research Laboratories (BRL).

2.1.2 Enzymes

Restriction endonucleases and T4 DNA ligase were from BRL. The large fragment of *E.coli* DNA polymerase I (Klenow) was from NBL Gene Sciences Ltd. Proteinase K and calf intestinal alkaline phosphatase were from Boehringer Mannheim. Ribonuclease A was from Sigma Chemical Company. *Taq* DNA polymerase was from Bioline. T3 and T7 RNA polymerases were from Stratagene Ltd..

2.1.3 Kits

BioNick kit was supplied by BRL. Qiagen kit for maxi DNA preps was supplied by Qiagen Inc.. PCR-Script™ Amp cloning kit was from Stratagene®. Sequenase® Version 2.0 kit used for sequencing was supplied by United States Biochemical. 5'-Ready RACE kit was supplied by Clontech.

2.1.4 Radioisotopes

[α-^{32}P] dCTP (3000Ci/mmol) was from ICN Biomedicals Ltd. [α-^{35}S] dATP

(400Ci/mmol), [α-^{35}S] UTP (>1000 Ci/mmol) and [α-^{32}P] UTP (800 Ci/mmol) were from Amersham International plc.

2.1.5 Electrophoresis and blotting materials

Agarose was from IBI, low melting point (LMP) agarose was from BRL and NuSieve agarose was from FMC Marine Colloids. Hybond-N nylon membrane was from Amersham International plc, Gene Screen *Plus*® membrane from NEN® Research Products, PALL Biodyne B and 3MM filter paper from Whatman Ltd.. The 1kb ladder, *Hind*III digested lambda DNA, and the 0.24-9.5kb RNA ladder used as size markers were from BRL. Band sizes were estimated using the computer program Gel version 1.01.

2.1.6 Photography and autoradiography

Polaroid 667 was used for photographing agarose gels on the UV transilluminator. Kodak Biomax MR film was used for ^{35}S exposures, and X-ograph blue film from X-ograph Imaging Systems for other autoradiography. K5 emulsion (Ilford) was used in the ^{35}S slide exposure. Ektachrome 64-T (Kodak) was used for photography of whole-mount embryos and sections. T-MAX 100 (Kodak) was used for dark field photography of ^{35}S tissue sections.

2.1.7 Mice

Strain CD-1 mice were supplied by Charles River UK Limited.

2.1.8 DNA probes

The mouse *ß-actin* probe was a gift from Dr. Aviva Symes. The *Krox 20* probe was a gift from Dr. J. Whiting.

2.1.9 Oligonucleotides

Initially, oligonucleotides were synthesised by Mark Fagan on a Pharmacia LKB Gene Assembler Plus. Then they were purchased from Oswel DNA Services and Genosys. Latterly they were supplied by Paul Rutland using an Applied Biosystems 381A DNA Synthesizer.

Primer	Sequence (5' to 3')	Nucleotide position*	Experimental use
FAM1A	TTICCIGTIGAIGAITTTAATGAA C T C C	2207- 2230	Degenerate PCR
FAM1B	TTICCIGTIGAIGAITTTAATGAG C T C C	2207- 2230	Degenerate PCR
FAM2	ICGCTTICGACAATTCTGIGCIGC TT TG G T	2332- 2309R	Degenerate PCR
SP1	TATTCTGGCAGTATCTG	3439- 3455	*Nfe2l1* 3' sequence
SP2	CAATCTTGGTCAGCTCT	4114- 4130	*Nfe2l1* 3' sequence
CNC5	AGGTAAGTATCCACGTC	498-482R	*Nfe2l1* 5' sequence, 5' RACE
CNC5B	ACAACACAAGGGCTGAG	274- 258R	*Nfe2l1* 5' sequence, 5' RACE nested primer
DCNC.FOR	GGTTGACAACAGCACTAGC	98-116	Isolation of *Drosophila cnc*
DCNC.REV	AGCGACAACTGGTTCTCGC	1127- 1109R	Isolation of *Drosophila cnc*
UK1.FOR	AGCCTTCTGCTCTCCAAT	2- 20	10.5UK1 CLONE PCR
UK1.REV	GGATAGGGATGCTTCAAA	250- 231R	10.5UK1 CLONE PCR
NFMMAP1	CAAGATCACACTTGTGGGCAATC	325- 347	Mapping of NFM by PCR
NFMMAP2	AAGGATGGCTGCCAAGATGG	447- 428R	Mapping of NFM by PCR

Table 2.1 Oligonucleotides used in primer walking and 5'RACE
** Nucleotide position refers to Nfe2l1 sequence (Figure 3.10), R= reverse primer*
I= Inosine

2.1.10 Vectors and libraries

M13 sequencing vectors mp18 and mp19 were purchased from Pharmacia LKB Technology. pBluescript SK+and PCRscript SK+ (supplied as a component in the PCRscript cloning kit) were from Stratagene Ltd. The libraries used are described below (Table 2.2).

LIBRARY	VECTOR	HOST	SOURCE
cDNA			
8.5 day mouse embryo	λgt10	POP101	B. Hogan
10.5 day mouse embryo	ISH*lox*	ER1647	Novagen
Drosophila gridded			J. Hoheisel
Human placental	λgt10	LE392	ATCC
Genomic			
ICRF Chromosome 17 cosmid library	Lawrist 4	DH5α	D. Nizetic
Human PAC library	pCYPAC2N	DH10B	HGMP

Table 2.2 Details of libraries (if known)

2.1.11 Bacterial strains

DH5α: F^-, *f*80*lacZ*DM15, Δ(*lacZ*YA-argF), U169, *deoR*, *recA*1, *endA*1, *hsdR*17(r_K^-, m_K^+) *supE*44, I^-, *thi*-1, *gyrA*96, *relA*1.

ER1647: tet^R, str^R F^-, λ^-, *trp*-31, *his*-1, *rspL*104(StrR), *fhuA* 2Δ (*lacz*)r1, *supE*44, *xyl-7*, *mtl*-2, *metB*1, *recD* 1014, *mcrA* 1272::Tn*10*, Δ(*mcrB*$^-$*hsdRMS*$^-$ *mrr*$^-$)2::Tn*10*.

JM101: *supE*, *thi*-1, Δ(*lac-proAB*), [F'*traD*36, *proAB*, *lacI*^{q}ZDM15].

POP101:

XL1-Blue: *recA*1, *endA*1, *gyrA*96, *thi*-1, *hsdR*17, *supE*44, *relA*1, *lac*, [F' *proAB*, *lacI*^{q}ZDM15], Tn*10*(tet^R)].

Y1090: Δ(*lac*)U169, Δ(*lon*), *araD*139, *strA*, *supF*, *mcrA*, *trpC*22::Tn*10*(tet^R), [pMC9amp^{R}tet^R].

Long term storage of bacterial strains was at -20°C or -70°C in 50% (v/v) glycerol. For short term storage, strains were maintained as isolated colonies on L-agar plates containing appropriate antibiotic at 4°C, except for *E.coli* strain JM101 which was maintained on M9 minimal media agarose plates.

2.1.12 Tissue culture reagents

Dulbecco's Minimal Essential Media (DMEM), RPMI 1640, Dulbecco's phosphate buffered saline, Trypsin-EDTA, penicillin and streptomycin were supplied by BRL. L-

Glutamine was obtained from Imperial Laboratories. Foetal bovine serum(FBS) was purchased from ICN Flow.

2.1.13 Cell lines

A9 mouse fibroblast cell line, P19 mouse embryonal carcinoma cell line and the XY normal lymphoblastoid cell line were bought from NIGMS Human Genetic Mutant Cell Repository. WA17 was a gift from Dr. D. Patterson. The hamster Ade-C genomic cell line DNA was a gift from Ms.H.O'Donnell. 3T6 was rodent cell line DNA of unknown origin.

2.1.14 Genebridge 4 Radiation Hybrid DNA panel

The Genebridge 4 Radiation Hybrid DNA panel (Gyapay *et al.*, 1996) was obtained from the UK-HGMP Resource Centre: this is a set of whole genome radiation hybrids (human on hamster background) that can be used for mapping by PCR, using human-specific oligonucleotide primers. DNA from 84 cell lines is supplied as template for PCR (as per standard protocol, section 2.2.15), and the results (positive, negative and unknown) are submitted electronically (http://www.hgmp.mrc.ac.uk/cgi-bin/contig/rhmapper.pl). The result is returned by email.

2.2 METHODS

All solutions and media were made with deionised water and, unless otherwise stated, autoclaved at 15 pounds per square inch (psi) for 20 minutes. Electrophoresis and blotting solutions were not autoclaved. RNA solutions were made with chemicals kept separate from general use and with milliQ grade deionised water: they were autoclaved if possible as above or made up in autoclaved water and/or filter sterilised.

2.2.1 Buffers, solutions and media

2.2.1.1 Buffers

E-buffer (TAE)	40mM Tris-acetate, 1mM EDTA, pH7.0
SM buffer	100mM NaCl, 10mM $MgSO_4.7H_2O$, 50mM Tris-HCl, pH7.5, 0.01% gelatin
TE	10mM Tris-HCl, pH7.4-8.0, 1mM EDTA, pH8.0
TBE	89mM Tris, 89mM boric acid, 2mM EDTA, pH8.0
Oligolabelling buffer(OLB)	Solutions A, B and C are mixed in the ratio of 100:250:150
Solution A	1ml solution O, 18µl ß-mercaptoethanol, 5µl 0.1M dATP, 5µl 0.1M dTTP, 5µl 0.1M dGTP
Solution B	2M HEPES, 4M NaOH
Solution C	Random hexanucleotides suspended in TE at 90 OD units/ml.
Solution O	1.25M Tris-HCl, pH8.0, 1.25M $MgCl_2$
Ligation buffer(10x)	200mM Tris-HCl, pH7.6, 100mM $MgCl_2$, 100mM DTT (supplied by BRL)
PCR buffer(10x)	100mMTris-HCl, pH8.4, 500mM KCl, 0.1% gelatin, 15mM $MgCl_2$
RNA Polymerase buffer(1x)	40mM Tris-HCl, pH8.0, 8 mM MgCl2, 50mM NaCl, 2mM spermidine (supplied by Stratagene)
Gel-loading buffer(10x)	25% ficoll (w/v), 0.25% orange G
RNA Gel-loading buffer(10x)	75% (v/v) deionised formamide, 9% (v/v) formaldehyde, 1.5x MOPS buffer
MOPS(10x)	0.4M Morpholinopropanesulfonic acid, 0.1M Na acetate, 10mM EDTA. Final adjusted pH7.2.

2.2.1.2 Solutions

I General solutions

Solution I	50mM glucose, 25mM Tris-HCl, pH8.0, 10mM EDTA
Solution II/Buffer P2	0.2M NaOH, 1% (w/v) SDS
STET buffer	8% (w/v) sucrose, 5% (v/v) Triton X-100, 50mM Tris-HCl, pH8.0, 50mM EDTA, pH8.0
Solution III	5M KOAc, pH4.8
Buffer P1	100µg/ml RNase A, 50mM Tris-HCl, 10mM EDTA, pH8.0
Buffer P3	3M KAc, pH5.5
Buffer QBT	0.75M NaCl, 50mM MOPS, 15% ethanol, pH7.0, 0.15% (v/v) Triton X-100
Buffer QC	1M NaCl, 50mM MOPS, 15% ethanol, pH7.0
Buffer QF	1.25M NaCl, 50mM Tris-HCl, 15% ethanol, pH8.5
Nuclease solution	50 mg DNase I, 50 mg RNase A in 10ml 50% (v/v) glycerol, 30mM Na Acetate, pH6.8
20x SSC	3M NaCl, 0.3M Tri-sodium citrate
Denaturing solution	1.5M NaCl, 0.5M NaOH
Neutralising solution	1M Tris-HCl, pH8.0, 1.5M NaCl
100x Denhardt's	2% (w/v) BSA, 2% (w/v) ficoll400, 2% (w/v) polyvinylpyrrolidine
$CaCl_2$ solution	60mM $CaCl_2$, 15% (v/v) glycerol, 10mM PIPES, pH7.0
Prehybridisation/hybridisation solution	50mM phosphate buffer, pH6.8, 4 x SSC, 5x Denhardt's solution, 0.15 mg/ml salmon sperm DNA,0.3% (w/v) SDS, 0.15% (w/v) sodium pyrophosphate

II Solutions for FISH

10x dNTP mix	0.2mM dCTP, 0.2mM dGTP, 0.2mM dTTP, 0.1m dATP, 0.1mM biotin-14-dATP, 50mM Tris-HCl, pH7.8,50mM $MgCl_2$, 100µg/ml BSA, 100mM β-mercaptoethanol
10x Enzyme mix	0.5U/µl DNA polymerase I, 0.0075U/µl DNase I, 50mM Tris-HCl, pH7.5, 5mM MgAcetate, 1mM β-mercaptoethanol, 0.1mM phenylmethylsulfonyl fluoride, 50% (v/v) glycerol, 100µg/ml BSA

Hybridisation mix	50% (v/v) deionised formamide, 10% (w/v) dextran sulphate, 2x SSC
Blocking buffer	3% (w/v) BSA, 4 x SSC, 0.1% (v/v) Tween 20
FITC buffer	1% (w/v) BSA, 4 x SSC, 0.1% (v/v) Tween 20

III Solutions for RNA work

Denaturing solution:	4M guanidinium thiocyanate, 25mM Na citrate, pH7.0, 0.5% (w/v) sarkosyl, 0.1M β-mercaptoethanol
Column loading buffer(CLB)	20mM Tris-HCl, pH7.6, 0.5M NaCl, 1mM EDTA, 0.1% (w/v) SDS
Elution buffer	10mM Tris-HCl, pH7.6, 1mM EDTA, 0.05% (w/v) SDS
NTE buffer	0.5M NaCl, 10mM Tris-HCl, pH8.0, 5mM EDTA, pH8.0

Northerns

Prehybridisation	5x SSC, 60% (v/v) deionised formamide, 20mM phosphate, pH6.0, 5x Denhardt's solution, 1% (w/v) SDS, heat-denatured 100µg/ml salmon sperm DNA,100µg/ml yeast tRNA, 10µg/ml poly A RNA
Hybridisation	As for prehybridisation, with the addition of 7% (w/v) dextran sulphate

In situ hybridisation solutions

Tissue section prehybridisation/hybridisation solution	50% (v/v) deionised formamide, 0.3M NaCl, 20mM Tris-HCl, pH8.0, 5mM EDTA, pH8.0, 10% (w/v) dextran sulphate, 1x Denhardt's solution, 0.5mg/ml yeast RNA
Whole mount prehybridisation/hybridisation solution:	50% formamide, 5x SSC, pH4.5, 50µg/ml yeast RNA, 1% SDS, 50µg/ml heparin
Solution I	50% (v/v) formamide, 5x SSC, pH4.5, 1% (w/v) SDS
Solution II	0.5M NaCl, 10mM Tris-HCl, pH7.5, 0.1% Tween-20
Solution III	50% (v/v) formamide, 2x SSC, pH4.5, 1% (w/v) SDS
10x TBST	1.M NaCl, 27mM KCl, 0.25M Tris-HCl, pH 7.5, 1% Tween-20
NTMT	100mM NaCl, 100mM Tris-HCl pH9.5, 50mM $MgCl_2$, 0.1% Tween-20
NY Prehybridisation/ Hybridisation solution	50% formamide,1.3x SSC, pH 5.3, 5mM EDTA,

50µg/ml yeast RNA, 0.2% Tween-20, 0.5% CHAPS, 100µg/ml heparin

2.2.1.3 Media

I Microbiological media (per litre)

L-broth (LB)	10g Bacto-tryptone, 5g yeast extract, 10g NaCl
2xYT	16g Bacto-tryptone, 10g yeast extract, 5g NaCl
NZY broth	21g NZY, 2 pellets NaOH
H-broth	10g Bacto-tryptone, 8g NaCl
M9 minimal media	6g Na_2HPO_4, 3g KH_2PO_4, 0.5g NaCl, 1g NH_4Cl and after autoclaving add, 2mM $MgSO_4$, 0.2% glucose, 0.1mM $CaCl_2$
SOB	20g bacto-tryptone, 5g yeast extract, 0.5g NaCl, 10ml 0.25M KCl
SOC	SOB as above, with the addition 20ml 1M glucose and 10ml 2M $MgCl_2$

Plates were made with the addition of 15g of agar per litre.

Soft top was made with the addition of 7g of agarose per litre.

Selective media were made by the addition of antibiotic solution immediately prior to use.

Antibiotic solutions (1000x)

Ampicillin	50mg/ml
Tetracycline	12mg/ml
Kanamycin	25mg/ml

Colour selection (1000x)

X-gal	25mg/ml in dimethyl formamide
IPTG	25mg/ml, filter sterilised

II Tissue culture media

Attached cell lines were grown in DMEM, and suspension lines in RPMI 1640. Media was supplemented with 10% (v/v)FBS, 4mM glutamine, 100IU/ml penicillin, 100UG/ml streptomycin.

2.2.2 Extraction and precipitation of nucleic acids

Extraction of nucleic acids was performed by the addition of an equal volume of equilibrated phenol/chloroform/isoamyl alcohol (IAA, 50:49:1 v/v), and subsequent

vortexing until a homogeneous solution was formed. This was followed by centrifugation to separate the phases. The upper aqueous layer was carefully removed, avoiding any contamination with debris from the interface, and re-extracted with an equal volume of chloroform/IAA. After centrifugation, the upper aqueous layer was removed and sodium acetate (pH5.2) was added, with mixing, to a final concentration of 0.3M. The nucleic acids were precipitated by the addition of 2.5 volumes of ethanol, and incubation at -20°C, -70°C or on dry-ice for at least 20 minutes. The nucleic acids were pelleted by centrifugation in a microcentrifuge or at 10,000g in a Sigma 6K10 for 10 minutes. The supernatant was removed and the pellet was washed with 70% (v/v) ethanol, and centrifuged as before. The supernatant was removed and the pellet air-dried at room temperature for 5 minutes, before being resuspended in water, TE buffer, or another appropriate buffer.

2.2.3 Miniprep preparation of plasmid and cosmid DNA

Plasmid DNA was prepared by one of the three methods described below. Cosmid DNA was solely prepared by the alkaline lysis method (Birnboim and Doly, 1979). 5ml of a selective LB culture was inoculated with a single bacterial colony and incubated overnight with shaking at 37°C. The DNA was then prepared by one of the methods outlined below.

2.2.3.1 One-step method (Chowdhury, 1991)

This method was used for rapid isolation of plasmid DNA to allow for quick verification of subcloned DNA fragments. To 500µl of an overnight culture, an equal volume of equilibrated phenol/chloroform/IAA (50:49:1) was added, and the solution was thoroughly vortexed for 1 minute. After centrifugation in a microcentrifuge at 14 000g for 5 minutes, 450µl of the upper aqueous layer was removed and to this an equal volume of propan-2-ol was added. After mixing, the solution was immediately centrifuged at 14 000g for 10 minutes. The pellet was washed with 70% (v/v) ethanol, air-dried for 5 minutes at room temperature and resuspended by pipetting in 50µl of TER [TE, pH7.4, ribonuclease A(RNase A) 10µg/ml]. 10µl was used for restriction enzyme digestion.

2.2.3.2 Boiling mini-prep method

This was used for preparing plasmid DNA for double-stranded sequencing. 1.5ml of an overnight culture was spun in a microcentrifuge for 30 seconds. The supernatant was

removed and the cell pellet resuspended in 350µl of STET buffer and 25µl of lysozyme (10mg/ml). The tube was then boiled for 40 seconds, followed by centrifugation in a microcentrifuge for 10 minutes. The pelleted cell debris was removed with a toothpick. 33µl of 3M sodium acetate, pH5.2 was added to the tube, followed by 450µl of propan-2-ol. After 5 minutes at room temperature, the tube was spun in microcentrifuge for 10 minutes . The pelleted DNA was washed with 1 ml 70% (v/v) ethanol and spun for 2 minutes in a microcentrifuge. The wash was removed, and the pellet allowed to air-dry for 5 minutes, before being resuspended in 50µl of TE, pH7.6.

2.2.3.3 Alkaline lysis method

1.5ml of the culture was spun in a microcentrifuge at 14 000g for 3 minutes. The supernatant was carefully removed to leave the cell pellet. This was usually sufficient for plasmids but was repeated for cosmid preparations. The cell pellet was resuspended by vortexing in 100µl of solution I, and left at room temperature for 5 minutes. 200µl of freshly prepared solution II was added, mixed gently and incubated on ice for 5 minutes. 150µl of solution III was added and mixed vigorously by vortexing and again placed on ice for 5 minutes. After centrifugation for 5 minutes at 14 000g the supernatant was transferred into a clean tube and extracted once with phenol/chloroform and once with chloroform. The DNA was precipitated with 2 volumes of ethanol, washed, dried and resuspended in 50µl of TE.

2.2.4 Maxiprep preparation of plasmid DNA

Plasmid DNA was prepared by the two methods described below.

2.2.4.1 Alkaline lysis and caesium chloride purification

500ml of selective media was inoculated with a 10ml overnight culture and grown with shaking at 37°C overnight. The cells were pelleted by centrifugation at 6000g for 10 minutes at 4°C in a Sigma 6K10. The media was poured away and the cell pellet was resuspended in 10ml of solution I. After 10 minutes at room temperature, 20ml of freshly prepared solution II was added and mixed gently. The mixture was incubated on ice for 10 minutes and then 15ml of solution III added and mixed vigorously. Following a further 30 minute incubation on ice, the cellular debris which had precipitated was pelleted by centrifugation at 10 000g for 10 minutes at 4°C. The supernatant was decanted into clean tubes through polymer wool, and the DNA was precipitated by the addition of 0.6 volumes of propan-2-ol. After 20 minutes at room temperature the

DNA was pelleted by centrifugation at 10 000g for 20 minutes at room temperature. The pellet was washed with 70% (v/v) ethanol , air-dried then resuspended in 3.9ml TE. 3.9g of caesium chloride was added and allowed to dissolve prior to the addition of 312µl of ethidium bromide (10mg/ml). A clearing spin was then performed to remove proteins and RNA at 14 000rpm for 30 minutes at room temperature. The supernatant was transferred to a mini-ultracentrifuge tube containing 39µl of 1% (v/v) Triton X-100, and topped up with paraffin oil then sealed. Density gradients were obtained by centrifugation at 100 000rpm for 4 hours at 20°C in a Beckman mini-ultracentrifuge. The nucleic acid was then visualised under UV light, and the lower band representing the covalently closed circular plasmid DNA was removed by puncturing the side of the tube with a needle and collecting the DNA into the syringe (Sambrook *et al*, 1989). The ethidium bromide was removed from the solution by extraction with water-saturated butan-1-ol. The sample was then diluted with 3 volumes of water and the DNA ethanol precipitated, then resuspended in 300µl of TE.

2.2.4.2 Qiagen kit method

The method followed was as outlined in the manufacturer's instructions. 150ml of selective media was inoculated with a 5ml overnight culture and grown with shaking at 37°C overnight. The cells were pelleted by centrifugation at 6000g for 15 minutes at 4°C. The media was poured away and the cell pellet was resuspended in 10ml of buffer P1. 10ml of buffer P2 was added with gentle mixing, followed by incubation at room temperature for 5 min. Then, 10ml of chilled buffer P3 was added with mixing, and the solution left to incubate on ice for 20 minutes. Following centrifugation at 13 000g for 30 minutes at 4°C, the supernatant was passed through polymer wool and collected. It was applied to a QIAGEN-tip 500, pre-equilibrated with buffer QBT. The QIAGEN-tip was washed twice with buffer QC, and the DNA eluted with 15ml of buffer QF. The DNA was precipitated with 0.7 volumes of isopropanol, then pelleted at 13 000g for 20 minutes at 4°C. The DNA was washed in 70% (v/v) ethanol, air-dried and resuspended in an appropriate volume of TE.

2.2.5 Preparation of DNA from cultured cells

Attached cell lines were harvested by trypsinization, pelleted by centrifugation at 1500rpm for 5 minutes, resuspended and washed in PBS then repelleted. Suspension cell lines were pelleted and washed twice with PBS. The cell pellets were resuspended in 1ml of PBS thoroughly. This was added in a dropwise manner to the lysis mix (final concentrations: 50mM EDTA(pH8.0), 1% (w/v) Sarkosyl, 200µg/ml proteinase K

0.25µg/ml. The lysis mix was incubated overnight at 55°C then extracted twice with phenol, once with phenol/chloroform and once with chloroform. The DNA was ethanol precipitated and hooked out with a 1.0ml pipette tip. This was washed briefly in 70% (v/v) ethanol and allowed to air dry, before being resuspended in an appropriate volume of TE.

2.2.6 Preparation of bacteriophage λ DNA

Bacteriophage λ was prepared according to the method of Chisholm (1989). Isolated plaques were plugged with a 3ml Pasteur pipette into 1ml of SM and allowed to elute at room temperature for at least 2 hours. 50µl of the phage were added to 500µl of host cells and incubated for 30 minutes at 37°C. The preadsorbed phage were then added to 37ml of NZY broth in a 250ml flask and incubated overnight at 37°C with shaking. The culture was transferred to a centrifuge tube, and 100µl of chloroform and 370µl of nuclease solution were added. This mixture was incubated at 37°C for 30 minutes, then 2.1g of NaCl was added and dissolved gently. After spinning for 20 minutes at 8000g at 4°C, the supernatant was transferred to clean tubes containing 3.7g of PEG 6000-8000. The PEG was dissolved and the tubes placed on ice for 60 minutes. The phage were pelleted by centrifugation at 8000g for 20 minutes at 4°C. The supernatant was removed, and the phage pellet resuspended in 500µl of SM over a period of 30 minutes. They were extracted once with chloroform, and the aqueous phase was transferred to clean Eppendorf tubes to which was added 20µl of 0.5M EDTA, 10µl of 10% (w/v) SDS and 10µl of proteinase K (2.5mg/ml). The tubes were incubated at 65°C for 30 minutes. One phenol and one chloroform extraction were carried out. 170µl of 6M ammonium acetate was added to the aqueous phase followed by 700µl of propan-2-ol to precipitate the DNA. The phage DNA was hooked out with a 1ml pipette tip, washed in 70% (v/v) ethanol and allowed to air dry before being resuspended in an appropriate volume of TE.

2.2.7 Quantification of nucleic acids

Nucleic acid concentration was determined by measuring the absorbance of the solution at 260nm. A value of A_{260} = 1 correlates to a concentration of approximately 50 µg/ml for a solution of double stranded DNA and 40 µg/ml for RNA. Absorbance at 280nm was also measured, and the ratio of $O.D._{260/280}$ determined: the expected ratio is 2.0 for RNA and 1.8 for DNA. For oligonucleotide concentration, the optical density at 260nm was determined and used in the following formula to give an approximate value:

Concentration (pmol/µl)=$O.D._{260}$/(0.01 x N)

where N equals the number of bases in the primer.

2.2.8 Restriction endonuclease digestion of DNA

Restriction endonuclease digests were performed using the reaction buffer provided by the enzyme supplier. Spermidine trihydrochloride was added to give a final concentration of 2 mM (in buffers with a salt concentration greater than or equal to 50mM). Typically 1-2U of enzyme was used to digest 1µg of DNA, in a total volume greater than 10 times the volume of enzyme used. The digests were incubated at the optimum temperature recommended by the supplier for 1-3 hours, except for genomic digests which were incubated for 6 hours to overnight.

2.2.9 Agarose gel electrophoresis of DNA

Agarose gel electrophoresis was used to separate DNA fragments, the size of the fragments to be separated determining the percentage of the agarose [usually 0.8-1.2% (w/v)]. The agarose was dissolved in E buffer by heating in a microwave oven, and when the solution had cooled to 50°C, ethidium bromide was added to a final concentration of 0.5 µg/ml. The solution was poured into a horizontal gel former with a comb in position. After the gel was set, it was placed in a gel tank containing E buffer, and the comb was removed. The samples to be run were mixed with the loading buffer, and loaded into the wells. Commercial size markers, 1kb ladder and/or λ*Hind*III, were also loaded. The DNA was electrophoresed horizontally along a voltage gradient of 1-10V/cm until the required separation was achieved.

2.2.10 Visualisation and photography of nucleic acids

During agarose gel electrophoresis, the nucleic acid becomes intercalated with the ethidium bromide which allows visualisation of the nucleic acid on a UV transilluminator (302nm). It was photographed using a Polaroid DS 34 camera on Polaroid type 667 film.

2.2.11 Southern blotting

DNA fragments separated using agarose gel electrophoresis were transferred to a charged nylon membrane (Hybond-N^+ or PALL Biodyne B) using the sandwich technique (Southern,1975). After electrophoresis and photography, the gel blot was set up with denaturing solution, and left overnight. The filter was washed briefly in 3x SSC and air-dried.

2.2.12 Radiolabelling of DNA

DNA probes were prepared by random priming of DNA template using a modification of the method described by Feinberg and Vogelstein (1984). Template DNA was prepared by excising the appropriate band with a scalpel blade from a 1% (w/v) low melting point agarose gel on a UV transilluminator. The band was weighed and 3ml water per g agarose added. The agarose was melted by heating to 65°C for 10 minutes, and the solution stored at -20°C until used. A heat-denatured aliquot of probe (25-50ng) was radiolabelled in a total volume of 50 µl, in OLB, 0.4 mg/ml BSA, 20µCi [α-^{32}P] dCTP with 1U of the Klenow fragment of *E.coli* DNA polymerase for 1-3 hours at 37°C or overnight at room temperature. Sephadex G-50 columns were prepared in 1ml syringe barrels which had been plugged with polymer wool. The labelling reactions were diluted with 100µl of 3x SSC, and then spun through the prepared column for 5 minutes at 1600rpm, separating the unincorporated nucleotides from the radiolabelled probe. The efficiency of incorporation was then estimated by Cerenkov counting.

2.2.13 Hybridisation of DNA filters

Hybridisation of Southern blots, gridded libraries and plaque lifts with oligolabelled DNA probes was performed according to the standard methods (Sambrook *et al.*, 1989). Filters were prehybridised in hybridisation solution in heat-sealed plastic bags, in hybridisation bottles with interleaved nylon mesh or in petri dishes. They were incubated with shaking or rotation at 65°C for at least 1 hour. Prior to use, double-stranded DNA probes were denatured by boiling for 5-10 minutes and quenched on ice to prevent reannealing. With some genomic probes it was necessary to compete them with total human DNA to prevent hybridisation to repetitive DNA sequences. Low molecular weight placental DNA was added to the probe at a final concentration of 1.25mg/ml in 3x SSC, the sample boiled for 10 minutes, then incubated at 65°C for 1-2 hours. It was then quenched on ice and added to the hybridisation solution. The

prehybridisation solution was replaced with hybridisation solution containing not more than 1x 10^6 dpm/ml of probe. Hybridisation was performed at 65°C overnight.

Following hybridisation, filters were washed at varying stringencies depending on the probe used. Generally gridded libraries and plaque lifts were first washed in a solution of 6x SSC, 0.1% (w/v) SDS for 30 minutes at 65°C, whilst Southern blots were usually washed first in 3x SSC, 0.1% (w/v) SDS under the same conditions. Subsequent washes were in decreasing salt concentrations, but not less than 0.1x SSC, until monitoring with a hand-held Geiger counter indicated an appropriate level of radioactivity (usually 5-10cps).

Washed filters were mounted onto intensifying screens under cling film and exposed to X-ray film in cassettes, with a second screen, usually overnight at -70°C. The exposed film was developed in an automatic developing machine.

2.2.14 Stripping of filters

Radiolabelled probes were removed from nylon filters by immersion in a boiling 0.1% (w/v) SDS solution which was allowed to cool to room temperature with agitation. The filters were allowed to air-dry and re-used. The stripped filters were sometimes exposed to X-ray film overnight to check that the probe had been completely removed.

2.2.15 Amplification of DNA by PCR

PCR reactions were performed in a final volume of 50µl, using 1xPCR buffer, 50-100pmoles of each oligonucleotide, 500ng of genomic DNA or 1µl of an overnight bacterial culture as template and 1U Biopro polymerase. The Mg^+ concentration was varied between 1.5-3.0 mM (final concentration) to allow for different optimal primer conditions. The mix was placed in an 0.5ml Eppendorf tube and overlaid with mineral oil. The DNA was denatured by heating to 94°C for 1 minute and then the reaction was allowed to proceed as programmed in a Hybaid thermal cycler. Standard reactions involved 30-35 cycles of denaturation at 94°C for 15 seconds, annealing at X°C for 30 seconds and extension at 72°C for 1 minute. There was a final extension at 72°C for 10 minutes. The appropriate annealing temperature (X°C) was estimated from the T_m of the oligonucleotide primer pair, where T_m=(2°C x no. of A + Ts) + (4°C x no. of G + Cs), and is 6°C below the T_m. Specific PCR conditions are quoted in the relevant place in the text.

2.2.15.1 Rapid amplification of cDNA ends(RACE)

RACE is a PCR-based method to specifically amplify the 5'ends of cDNAs (Frohman *et al.*, 1988; Belyavsky *et al.*, 1989). The Clontech 5'-RACE-Ready™ cDNA kit (Apte and Siebert, 1993) was used as per manufacturer's instructions to obtain the 5' ends of cDNAs.

2.2.15.2 Subcloning of PCR products

The PCR-Script™ Amp cloning kit (Stratagene®) was used to subclone PCR products, according to the manufacturer's instructions.

2.2.16 Large scale preparation of competent cells

From an overnight culture, 4ml was taken to inoculate 400ml of L-broth. This was grown until an $O.D._{590}$ of 0.375 was reached, then 50ml aliquots were dispensed into prechilled centrifuge tubes, and kept on ice for 5-10 minutes. The cells were pelleted by centrifugation at 1500g for 5 minutes at 4°C. The pellets were gently resuspended in 10ml of prechilled $CaCl_2$ solution on ice. The cells were then centrifuged at 1000g for 5 minutes at 4°C, resuspended as before in $CaCl_2$ solution, and kept on ice for 30 minutes. Next, they were spun at 1000g for 5 minutes at 4°C, and each pellet was well resuspended in 2 ml of ice cold $CaCl_2$ solution. The cells were dispensed in 200µl aliquots into prechilled Eppendorf tubes and snap frozen on a dry ice/ethanol mix. The competent cells were stored at -70°C.

2.2.17 Subcloning

Plasmid vector pBluescript SK^+DNA (1-2µg) was digested with the appropriate restriction endonuclease as described, 1U of calf intestinal alkaline phosphatase (CIAP) was added and the mixture incubated for 30min at 37°C. The DNA was extracted with phenol and chloroform, precipitated with ethanol and resuspended at a concentration of ≈ 25ng/µl. The plasmid or cosmid DNA to be subcloned was usually excised from a LMP agarose, phenol and chloroform extracted and ethanol precipitated. Ligation reactions contained ≈ 25ng of vector and 25-100ng of plasmid/cosmid DNA (depending on insert size), 2µl of 10x ligation buffer, 1µl of 20mM ATP and 1U of T4

DNA ligase in a total volume of 20µl. Appropriate controls were also set up. Reactions were incubated overnight at 12-15°C. Competent cells were either DH5α (subcloning efficiency) cells from BRL or XL1-B cells that were made competent and stored as described (section 2.2.16). After thawing the cells on ice, 2-5µl of the ligation reaction was added with gentle mixing to 50µl of cells, and the mixture kept on ice for 30 minutes. The transformation mix was heat-shocked at 42°C for 90 seconds, returned to ice for 2 minutes, then 500µl of NZY broth was added. The mixture was incubated at 37°C for 45-60 minutes, and 200µl was then plated onto L-agar plates containing the appropriate selection, X-gal (25µg/ml) and IPTG (25µg/ml). The plates were incubated overnight at 37°C. White colonies were miniprepped as in section 2.2.2.2.

2.2.18 M13 subcloning

2.2.18.1 Ligation and transformation

DNA for single stranded sequencing was ligated into M13mp18 and 19. Preparation of insert and vector DNA was as described (section 2.2.17). *E.coli* strain JM101, maintained on a minimal media agar plate, was used to prepare competent cells. A single colony was used to inoculate 10ml of 2 x YT broth and grown overnight at 37°C. 0.5ml of this culture was used to inoculate 50ml of 2x YT broth, which was then grown for 105 minutes at 37°C with shaking. The cells were harvested by centrifugation at 3000 g for 5 minutes at 4°C, and gently resuspended in 10ml of ice cold 50mM $CaCl_2$. After a 40 minute incubation on ice, the cells were harvested as before. The pellet was resuspended in 2ml of 50mM $CaCl_2$ and either used immediately or left overnight at 4°C.

From the ligation reaction, 5µl was removed and gently mixed with 200µl of competent cells, then incubated on ice for 40 minutes. A plating mixture consisting of 200µl of a 1/50 dilution of a fresh overnight JM101 culture in 2x YT broth, 20µl of X-gal, 20µl of IPTG and 3ml of molten H-top was prepared for each transformation and kept at 50°C. The transformation mixture was heat-shocked at 42°C for 90 seconds, placed on ice, then added to the plating mixture. After mixing, it was poured onto a 90mm H-plate, and incubated overnight at 37°C.

2.2.18.2 Preparation of single-stranded DNA

White recombinant plaques were picked with a toothpick and used to inoculate 2ml aliquots of a 1/100 dilution of an overnight JM101 culture in 2x YT broth. The culture was incubated for 5-6 hours with shaking at 37°C, and 1.5ml of culture removed to an Eppendorf tube. The cells were pelleted by centrifugation at 14 000g for 5 minutes. The supernatant was then poured into a clean Eppendorf tube and centrifuged at 14 000g for 30 seconds. A 1.2ml aliquot of supernatant was transferred into a clean Eppendorf tube, to which was added 300µl of 2.5M NaCl, 20% (w/v) PEG 6000. After incubation for 20 minutes at room temperature, the precipitated phage were pelleted by centrifugation for 10 minutes at 14 000g. The supernatant was carefully removed, the tube respun for 30 seconds, and the remainder of the supernatant removed. The pellet was dissolved in 200µl of water, then extracted once with phenol and once with chloroform, and precipitated as described in 2.2.2. Usually a tenth of the final volume was used in a sequencing reaction.

2.2.19 DNA Sequencing

Manual DNA sequencing was performed by the dideoxy chain termination method of Sanger *et al.* (1977) using a commercially available kit (Sequenase® Version 2.0, USB). Single-stranded template was prepared as described in 2.2.18, and ≈1µg of DNA was mixed with 2µl of 5x Sequenase® reaction buffer, 0.5pmoles primer, in a final volume of 10µl.

2.2.19.1 Preparation of double-stranded DNA template

3-5µg of plasmid DNA that was either CsCl-purified or prepared using the boiling miniprep method was alkaline denatured: it was mixed with 2µl of 5M NaOH, 10µl of 10mM EDTA, in a total volume of 50µl, and incubated for 15 minutes at 37°C. The reaction was then neutralised by the addition of 5µl of 3M NaAc, pH5.2, and 125µl of ethanol added to precipitate the DNA. The DNA was pelleted by centrifugation at 14 000g for 10 minutes, washed once with 70% (v/v) ethanol, air dried and resuspended in 10µl of the annealing mix (1x Sequenase® reaction buffer, 0.5pmoles primer).

2.2.19.2 Sequencing reaction

The annealing step was carried out by heating the sample at 65°C for 2 minutes, and allowing it to cool slowly to 35°C, by flotation in a small beaker with water from the

waterbath at 65°C. Once completed, the mix was kept on ice for a maximum of 4 hours before being used.

To the annealed template and primer mix was added 5.5µl of the sequencing reaction mix, consisting of 1µl of 0.1M DTT, 2µl of "labelling mix" (diluted 1:5 with water), 0.5µl of [α- ^{35}S] dATP and 2µl of Sequenase® enzyme (diluted 1:8 with ice-cold Sequenase® dilution buffer). This was mixed and incubated at room temperature for 5 minutes.

During the labelling step, 2.5µl of each ddNTP termination mix was aliquoted into a 60-well microtitre plate (Pharmacia LKB), and heated at 37-42°C for 1 minute prior to use. Once the labelling step was finished, 3.5µl of each reaction was added to each of the four termination mixes and incubated at 37-42°C for 5 minutes. The reaction was then stopped with the addition of 4µl of Stop solution, and stored at -20°C until used.

2.2.19.3 Polyacrylamide gel electrophoresis

The glass plates were washed with detergent, well-rinsed and wiped with ethanol. The eared plate was silanised by wiping a small amount of dimethylchlorosilane solution onto one side. The plates were assembled into a gel former using 0.4mm thick spacers and bulldog clips. The polyacrylamide gels were prepared using Sequagel (National Diagnostics). Typically, 40ml of diluent, 14.4 ml of concentrate, 6ml of 10x TBE were mixed together. The polymerisation was initiated by the addition of 510µl of 10% (w/v) ammonium persulphate and 30µl of TEMED (N,N,N',N' tetramethyl-ethylene-diamine). Using a syringe, the gel was quickly poured between the plates, with an inverted "sharks tooth comb" placed at the top. After it was set, the gel was placed in a Hybaid sequencing tank with 1x TBE and pre-run for 15-30 minutes at 70W. The samples were denatured at 100°C for 2 minutes, placed on ice, and 2-3µl loaded onto the gel and run at 65-75W for the appropriate time.

2.2.19.4 Autoradiography

The two glass plates were prised apart and the gel lifted off the plain glass plate with Whatman 3MM paper. After covering the gel surface with cling film, it was dried at 80°C in a vacuum gel drier. The gel was exposed directly to Kodak BiomaxMR film overnight at room temperature.

2.2.19.5 Sequence analysis

Computer programmes for sequence analysis and database searching were accessed at the UK-Human Genome Mapping Project (HGMP) Resource Centre (http://www.hgmp.mrc.ac.uk). The Basic Local Alignment Search tool (BLAST; Altschul *et al.*, 1990) was used to search databases such as GenBank and NBRF. The GCG programmes, available at the UK-HGMP Resource Centre, were used to analyse and edit sequences.

2.2.20 Library methods

2.2.20.1 Plating and screening cDNA bacteriophage libraries

Bacteriophage libraries were screened according to the standard methods (Sambrook *et al.*, 1989). Plating cells were prepared by inoculating 10ml of NZY broth, supplemented with 0.2% (w/v) maltose and 10mM $MgSO_4$, with a single colony of the appropriate host strain, and incubating overnight at 37°C with shaking. The cells were pelleted by centrifugation at 3000rpm for 5 minutes, resuspended in 0.5 volumes of 10mM $MgSO_4$ and stored at 4°C for up to one week.

To titre the library, serial dilutions of the bacteriophage stock were made in SM, then each mixed with 500µl of the prepared plating cells and allowed to adsorb for 10 minutes at room temperature, followed by 20 minutes at 37°C. 10ml aliquots of NZY top agarose were prepared and kept at 50°C. The top agarose and cell mixture were quickly combined, then poured onto a NZY agar plate. After the top agarose had hardened, the plates were incubated overnight at 37°C. The number of plaque forming units (p.f.u.) was counted, and the titre of the library was estimated.

The library plating was performed using typically 20x 150mm petri dishes, each with 25 000-30 000p.f.u. The plates were made as described above, and incubated overnight at 37°C. They were placed at 4°C for 30-60 minutes to harden the top agarose before lifts were taken. Duplicate replica filters were made from each of the masters: a Hybond-N^+ filter was placed on the plate, and orientation marks were made using a sterile 21G syringe needle. After 3 minutes, the filter was carefully peeled off, and replaced with a second circle. The same orientation marks were made and the duplicate filter was left for 5 minutes before removal. To release the DNA, each filter was placed DNA side up into a tray containing 3MM filter paper soaked in denaturing solution for 5 minutes. The filter was then placed on 3MM soaked in neutralising

solution for 5 minutes, followed by transferral to a tray containing neutralising solution for 3 minutes. The filter was rinsed in 3x SSC and allowed to air-dry (Benton and Davis, 1977). The filters were baked at 80°C for 2 hours to bind the DNA. The filters were prehybridised and hybridised as described in section 2.2.13.

2.2.20.2 Identification and purification of positive bacteriophage

Signals which were duplicated on both filters were taken as identifying positive clones. The master plate corresponding to any positive signal was placed on top of the autoradiograph on a light box. The area covering the positive signal was removed as a plug of agar with a plastic Pasteur pipette. The plug was expelled into an Eppendorf tube containing 1ml of SM and a drop of chloroform, and the phage allowed to elute for 2-3 hours at room temperature or overnight at 4°C. Serial dilutions of the phage stock were made in SM, and typically 100µl of 10^{-4} dilution was plated for secondary screening. The dilution was plated as described in section 2.2.20.1 and further rounds of screening performed until a single plaque could be picked. DNA was prepared as described in section 2.2.6.

2.2.20.3 Screening of genomic libraries

The ICRF chromosome 17 cosmid library and the human PAC library were supplied as duplicate filters, containing either a 96 x 96 gridded array of cosmids (ICRF) or on 7 filters, each containing 36,864 duplicated clones (PAC). These filters were hybridised as described in section 2.2.13.

2.2.20.4 Identification and purification of cosmids

Signals which were duplicated on both filters were taken as identifying positive clones. The co-ordinates for each signal was determined, and the clones requested from the RLDB, ICRF or the UK-HGMP Resource Centre. The clones were sent as stabs, which were immediately streaked onto LB-agar plates containing 20µg/ml of kanamycin. Several bacterial colonies were then prepared from each plate as described in section 2.2.3.3.

2.2.21 Tissue culture methods

2.2.21.1 Fibroblast cell lines

Fibroblast cell lines were grown in DMEM, supplemented as described in section 2.2.1.3, and maintained as monolayer cultures in CO_2-buffered incubators. They were passaged by trypsinization: the media was removed and the cells washed with PBS. Following the removal of the PBS, Trypsin-EDTA solution was added and allowed to cover the cells for a few minutes. The flask was tapped gently to help lift off the cells. An equal amount of media was added to the flask to stop the trypsinization, and the cells were gently resuspended by pipetting. The cell suspension was transferred into a larger flask, or split amongst more flasks, used to prepare DNA (as described in 2.2.5), or frozen down.

To make frozen stocks, the cells were pelleted by centrifugation at 345g for 5 minutes and the media was removed. The cell pellet was gently resuspended in 10ml of PBS and spun again at 345g for 5 minutes. The pellet was resuspended at 5x 10^6 cells/ml in medium containing 20% FBS and 10% DMSO and aliquoted into Nunc cryotubes. The cells were frozen slowly to -70°C overnight and transferred to liquid nitrogen for long term storage. The cells were thawed rapidly by incubation at 37°C, centrifuged out of the freezing solution , resuspended in 5 ml of media and transferred to a 25cm^2 flask.

2.2.21.2 Lymphoblastoid cell lines

Lymphoblastoid cell lines were grown in RPMI 1640 medium, supplemented as described in section 2.2.1.3. The cells were passaged by dilution. Frozen stocks were made as described in section 2.2.21.1. Cells for metaphase spreads were prepared as described in 2.2.22.1.

2.2.22 Fluorescence *in situ* hybridisation (FISH)

Fluorescence *in situ* hybridisation was performed using standard methods (Baldini and Lindsay, 1994).

2.2.22.1 Preparation of metaphase spreads

A 50 ml culture of lymphoblastoid cells was prepared and the media changed 24 hours prior to preparation of metaphase spreads. After 24 hours the cells were mitotically arrested by the addition of 250µl of 10µg/ml colcemid and incubation for 1 hour. The cells were pelleted by centrifugation at 160g for 5 minutes, and the media

aspirated. Cells were resuspended in the remaining media and 50ml of freshly prepared 75mM KCl hypotonic solution was added dropwise. Cells were spun down and resuspended in the remaining supernatant. Ice cold fix (3:1 methanol: glacial acetic acid, freshly prepared) was added dropwise, and the volume made up to 10 ml. Cells were mixed by inversion and incubated on ice for 20 min. Cells were pelleted by centrifugation at 1500 rpm for 5 minutes and resuspended dropwise in a further 10 ml of fresh fix. The cells were fixed twice more and dropped onto slides, previously cleaned with ethanol. Slides were stored at 4°C with desiccant until required. Metaphase chromosomes in fix were kept at 4°C for short term storage and at -20°C for longer term storage.

2.2.22.2 Nick translation of DNA probes

Nick translation was carried out using the BioNick kit from BRL. Cosmid mini-prep DNA was prepared as described in section 2.2.2.3. Prior to use as template DNA, 10μg/ml of RNase A was added, and the mix incubated at 37°C for 1 hour. The DNA was then extracted and precipitated as described in section 2.2.2. The nick translation was set up as follows: 1μg of DNA was mixed with 5μl of 10x dNTP mix, 5μl of the 10x enzyme mix, and the volume increased to 50μl. The reaction was incubated at 12-16°C for 2 hours. The reaction was stopped by the addition of 4μl of 0.5M EDTA and 1μl of 10% (w/v) SDS. The probe was separated from unincorporated nucleotides by spinning through a G-50 Sephadex column, equilibrated in 50 mM Tris-Cl, 1 mM EDTA, 0.1%(w/v) SDS at 355g for 5 minutes. A 10μl aliquot of the probe was electrophoresed in a 2% (w/v) agarose gel to estimate the fragment size and quantity. The probe was kept at -20°C until required.

2.2.22.3 Hybridisation

Prior to use metaphase slides were aged in 2x SSC at 37°C for 30 min and dehydrated through 70% (v/v), 90% (v/v) and absolute ethanol for 3 min each and air dried. Slides were denatured by immersion in 70% (v/v) deionised formamide, 2x SSC equilibrated at 70°C, for 2 minutes exactly. The slides were then dehydrated by dipping successively in 70% (v/v) ethanol (prechilled at -20°C), 90% (v/v) ethanol and absolute ethanol. The slides were left to air-dry at an incline.

The probes were repetitive so it was necessary to compete them with competitor DNA to suppress background prior to the hybridisation. A 10μl aliquot of labelled probe was mixed with 2μl of C_0t 1 human competitor DNA and 3μl of salmon sperm DNA (10mg/ml), and the mix was co-precipitated with the addition of 1.5μl of 3M NaAc,

pH5.2 and 60µl of ethanol. It was incubated on dry ice for 15 minutes or at -70°C for 30 minutes, then pelleted by centrifugation at 14 000 rpm at 4°C for 20 minutes. The pellet was allowed to air-dry, then resuspended in 10µl of the hybridisation solution. The probe was denatured by incubation at 80°C for 5 minutes, followed by annealing at 37°C for 15 minutes, after which it was placed on ice until required.

The slides were prewarmed, then 10µl of the competed probe in the hybridisation solution was pipetted over the selected area , and a coverslip (22 x 22 mm) was gently lowered onto the solution, being careful not to trap any air bubbles. The rim of the coverslip was sealed with Cow Gum rubber cement. The slides were transferred to a moist box, and incubated horizontally overnight at 37°C.

2.2.22.4 Visualisation

Post hybridisation, the slides were washed three times in 50% (v/v) formamide, 2x SSC at 42°C for 5 minutes, followed by three washes in 1x SSC at 60°C for 5 minutes. To decrease non-specific hybridisation, 200µl of blocking buffer was applied to each slide, covered with a 50 x 22mm coverslip, and incubated at 37°C for 30 minutes. The rest of the procedure was carried out in the minimum of light to protect the fluorochromes. The antibodies were then applied in a similar manner diluted in FITC buffer to a final concentration of 7ng/µl, and incubated at 37°C for 30 minutes to 1 hour. This was followed by three successive washes in 4 x SSC, 0.1%(v/v) Tween 20 at 42°C for 5 minutes each. The chromosomes were stained by immersing the slides in DAPI (200ng/ml), 2x SSC for 10 minutes at room temperature. The slides were then mounted using Vectashield mountant from Vector Labs, Inc.

2.2.22.5 Fluorescence microscopy

The slides were examined in diminished lighting on a Ziess Axioscope 20 fluorescence microscope under oil using a 100x objective. A positive result was confirmed after similar results were obtained using more than one probe (if possible) and upon examination of several different metaphase spreads. Images were photographed using a Photometrics Nu200 CCD camera system and manipulated using the SmartCapture software from Digital Scientific. Images were printed from a MacDrawPro format on a Mitsubishi sublimation printer.

2.2.23 RNA Methods

RNA chemicals were kept separate from general stocks. Glassware was washed with detergent, rinsed in milliQ water, and baked overnight at 200°C. Plasticware was either of the single-use disposable type and assumed to be RNase-free, or treated with 2% Absolve (NEN), before rinsing copiously in sterile milli Q water.

2.2.23.1 Animal husbandry

Mating pairs of CD-1 mice were set up in late afternoon. On examination the following morning, the presence of a mucous plug in the vagina of the mouse was taken to indicate pregnancy, with that morning being 0.5 days post coitum (dpc). Animals were sacrificed on the appropriate day by cervical dislocation. Embryos were staged approximately by day of sacrifice, and more accurately by the number of somites present.

The embryos were placed in ice-cold PBS, and dissected out in fresh ice-cold PBS, using if needed an Olympus microscope. They were then either frozen in liquid nitrogen and stored at -70°C until required or further processed for *in situ* hybridisation (sections 2.2.23.8 and 2.2.23.9). Adult tissues were taken after removal of the embryos and processed either for RNA extraction (section 2.2.7.2) or for tissue section *in situ* hybridisation (section 2.2.23.8).

2.2.23.2 Isolation of total RNA

Total RNA from cell lines was isolated using the single-step method by acid guanidinium thiocyanate-phenol/chloroform extraction (Chomczynski and Sacchi, 1987). Total RNA from adult mouse tissues and embryos was extracted using the previous method or an adaptation of the method described by Auffray and Rougeon (1980).

I Single-step method

For the isolation of total RNA from fibroblast cell lines, the media was removed from flasks of rapidly dividing attached cells, and the cells washed with PBS, which was subsequently removed. For each 10^7 cells, 1ml of denaturing solution was added and the resulting viscous fluid transferred into 15ml corex centrifuge tubes. For tissues, 1ml of denaturing solution per 100mg of tissue (usually ≈500mg) was placed in a 15ml corex centrifuge, the frozen tissue added, and immediately homogenised using an electric homogeniser (Sorvall® Omni-Mixer).

The following were then added with mixing after each addition (per ml of denaturing solution): 100µl of 2M sodium acetate, pH 4.0, 1ml of water saturated phenol, and 200µl of chloroform/IAA (49:1). The tubes were vortexed for 10 seconds, incubated on ice for 15 minutes, then centrifuged at 10 000g for 20 minutes at 4°C. The upper aqueous layer was carefully removed to a clean centrifuge tube and mixed with an equal volume of isopropanol to precipitate the RNA. The tube was placed at -20°C for at least 1 hour, prior to pelleting the RNA by centrifugation at 10 000g for 20 minutes at 4°C. The pellet was resuspended in a third of the initial volume of denaturing solution and, if feasible, transferred to an Eppendorf tube. An equal volume of isopropanol was added, and the tube incubated at -20°C for 1 hour to precipitate the RNA. The RNA was pelleted by centrifugation in a microcentrifuge at 14 000g for 10 minutes at 4°C. The pellet was washed with 70%(v/v) ethanol, briefly air-dried and resuspended in an appropriate volume of sterile milliQ water.

II LiCl precipitation method

This protocol is an adaptation of the method described by Auffray and Rougeon (1980). Each gram of frozen tissue was homogenised in 5-10 ml of pre-chilled 3M LiCl, 6M urea for 2 minutes on ice. The homogenate was left overnight at 4°C and then further processed. Occasionally samples were left at this stage for up to 1 week as other samples were collected. The homogenate was centrifuged at 10 000g for 20 minutes at 4°C, and the pellet resuspended by vortexing in half the original volume of 3M LiCl, 6M urea. The mix was centrifuged at 10 000g for 20 minutes at 4°C. The supernatant was discarded and the pellet was dissolved in half its original volume of TE, pH7.6, 0.5% (w/v) SDS. An equal volume of equilibrated phenol/chloroform/IAA was added, and the tube shaken vigorously for 5 minutes. If the pellet was difficult to dissolve, the phenol mix was added, and the tube vortexed for 2-3 minutes. After centrifuging to separate the phases, the upper aqueous layer was removed to a clean tube. The RNA was ethanol precipitated as described in section 2.2.2. The RNA pellet was dissolved in sterile milliQ water.

The RNA was stored at -70°C until required. The RNA was quantified as described in section 2.2.7. The quality of RNA was established by running an aliquot on a formaldehyde denaturing gel (section 2.2.23.4).

2.2.23.3 Isolation of poly(A)$^+$ RNA

Poly(A)$^+$ RNA was isolated from total RNA by affinity chromatography on oligo(dT)-cellulose (Sambrook *et al.*, 1989). An appropriate amount of oligo(dT)-cellulose (≈ 0.2-0.4 g) was resuspended in 0.1M NaOH, and poured into a sterile Dispocolumn

(Bio-Rad). The column was washed with 3 column volumes of sterile milliQ water. The column was then washed with 1x column-loading buffer (CLB) until the pH of the effluent was less than 8.0 (as measured using pHstix, BDH). A quantity of total RNA [up to 10mg of total RNA per 1 ml of packed oligo(dT)-cellulose] was heated at 65°C for 5 minutes, cooled to room temperature quickly, and an equal volume of 2x CLB was added. The solution was applied to the column, and the eluate immediately collected. Once all of the RNA solution had entered the column, one column volume of 1x CLB was added and the eluate collected. The collected eluate was heated at 65°C for 5 minutes, then reapplied to the column. The column was then washed with 10 column volumes of 1x CLB.

The poly(A)$^+$ RNA was eluted from the oligo(dT)-cellulose column by applying 2-3 column volumes of elution buffer. The eluate was collected in fractions equivalent to 1/2 the column volume. The absorbance of the fractions was determined as described in section 2.2.7, and those containing the majority of the poly(A)$^+$ RNA were pooled (typically the first 3 fractions). The RNA was ethanol precipitated as described in section 2.2.2, and the pellet resuspended in sterile milliQ water. The yield of poly(A)$^+$ RNA was determined by UV spectrophotometry (section 2.2.7). The RNA was stored at -70°C until required. The oligo(dT)-cellulose column was regenerated by sequential washing with 0.1M NaOH, water and CLB: it was stored at 4°C until reused.

2.2.23.4 Agarose gel electrophoresis of RNA

Formaldehyde denaturing gel electrophoresis was used to separate RNA samples (Sambrook *et al.*, 1989). Typically, either 10µg of total RNA or 1-2µg of poly(A)$^+$ RNA in a volume of 4-7µl was mixed with RNA loading buffer. The sample was denatured by heating at 65°C for 5 minutes, then quenched on ice. An appropriate amount of RNA size marker was treated similarly. The samples were then mixed with RNA gel loading dye, and if required, 1µl of 5mg/ml ethidium bromide. They were loaded onto a denaturing agarose gel containing 1% (w/v) agarose, 1x MOPS buffer and 6% (v/v) formaldehyde, and run at 5V/cm in 1x MOPS buffer for an appropriate length of time. The RNA was visualised and photographed as for DNA (section 2.2.10).

2.2.23.5 Northern blotting

Northern blotting was performed according to the manufacturer's instructions for Gene Screen *Plus*® membrane (NEN® Research Products). Subsequent to electrophoresis and photography of the RNA gel, it was washed four times in sterile milliQ water to remove excess formaldehyde. During this time, a piece of Gene Screen *Plus* ® membrane was cut

to the appropriate size, briefly wetted in sterile milliQ water, then soaked in 10x SSC for 15 minutes. The blot was then assembled in the usual way (Sambrook *et al.*, 1989) with 10x SSC as transfer solution, and left overnight. Following transfer, the blot was rinsed in 2x SSC to remove any residual agarose, air-dried, then baked at 80°C for 2 hours to reverse the formaldehyde reaction. The blots were stored at -20°C until required.

2.2.23.6 Labelling of RNA probes

Single-stranded, high specific activity RNA probes for northerns were prepared according to the manufacturer's instructions (Stratagene) using [α-^{32}P] UTP (800 Ci/mmol). Non-radioactive single-stranded RNA probes for whole-mount *in situ* hybridisation were made using the same protocol, with DIG-11-UTP as substrate. Single-stranded, high specific activity RNA probes for tissue section *in situ* hybridisation were prepared essentially as described by Wilkinson and Nieto (1993).

I Stratagene protocol

Cs-Cl purified plasmid DNA was used to prepare template DNA. This was linearised with an appropriate restriction endonuclease, followed by treatment with proteinase K (final concentration 50µg/ml) for 30 minutes at 37°C. The DNA was then extracted once with phenol/chloroform, once with chloroform, and ethanol precipitated. The template DNA was resuspended in autoclaved milliQ water at an approximate concentration of 1µg/µl, and stored at -20°C until required. Depending on the orientation of the insert DNA, either T3 or T7 RNA polymerase was used. For non-radioactive labellings, the [α-^{32}P] UTP was substituted with final concentrations of 140µM DIG-11-UTP and 260µM UTP.

The RNA transcription reaction contained 1x transcription buffer, 1µg template DNA, 400µM ATP, 400µM CTP, 400µM GTP, 30mM DTT, 1U RNasin, 50µCi [α-^{32}P] UTP and 10U of the appropriate RNA polymerase in a volume of 25µl. After mixing, the reaction was incubated at 37°C for 30 minutes . Following transcription, the DNA template was removed by adding 1U RNasin, 10µg of yeast tRNA and 10U of RNase-free DNase, followed by incubation at 37°C for 10 minutes. Subsequently the probe was precipitated by the addition of 95µl of autoclaved milliQ water, 10µl of 5M LiCl, 300µl of ethanol and incubation at -70°C for 1 hour or longer or on dry ice for 15 minutes. The RNA probe was pelleted by centrifugation at 14 000g for 20 minutes at 4°C, washed with ice-cold 80% (v/v) ethanol, air-dried for 5 minutes and resuspended in 150µl of autoclaved milliQ water. The efficiency of incorporation was then estimated by Cerenkov counting.

II Labelling of RNA probes for radioactive *in situ* hybridisation

This was as essentially described by Wilkinson and Nieto (1993). Template DNA was prepared as described in the section above. The RNA transcription reaction contained 1x transcription buffer, 10mM DTT, 0.25mM ATP, 0.25mM CTP, 0.25mM GTP, 1μg template DNA, 1U RNasin, 10μCi [α-^{35}S] UTP and 10U of the appropriate RNA polymerase in a final volume of 20μl. After mixing, the probe was incubated at 37°C for 1 hour, then 100μl of autoclaved milliQ water was added. From this, 1μl was removed to determine the percentage incorporation of isotope, and the specific activity of the probe produced.

To determine the percentage incorporation, a 1μl sample was diluted 1000-fold with water, and 10μl aliquots were spotted onto DE81 filter disks, which were air dried. Duplicate filters were divided into two groups: one group was washed 6 times in 0.5 M Na_2HPO_4 for 5 minutes each, to remove unincorporated nucleotides, then both sets were washed twice in water for 1 minute each and twice in 100% ethanol for 1 minute each, then allowed to air dry. The filter disks were counted in liquid scintillation counter, and the percentage incorporation determined. This was usually in the range of 70-85%.

The unincorporated nucleotides were removed by spinning the reaction mix through a G-50 Sephadex column at 1600 rpm for 5 minutes. The RNA probe was precipitated with the addition of 10μl of 5M LiCl, 300μl ethanol and incubation at -70°C overnight. The probe was pelleted by centrifugation at 14 000g for 20 minutes at 4°C, washed with ice-cold 80% (v/v) ethanol, resuspended in an appropriate amount of 100mM DTT, and stored at -70°C.

2.2.23.7 Hybridisation of northern blots

I Using RNA probes

Northern blots were hybridised with RNA probes as described by Rowe *et al.* (1991). The northern blots were hybridised in hermetically sealed plastic bags. They were prehybridised for at least 2 hours in prehybridisation mixture at 65°C with shaking. The hybridisation mixture was the same as for prehybridisation with the addition of 7% (w/v) dextran sulphate, and the labelled RNA probe to a final concentration of 10^6cpm/ml of hybridisation mixture. The probe was heat denatured by heating to 80°C for 5 minutes prior to adding to the mix. Hybridisation was carried out overnight at 65°C with shaking.

After hybridisation, the membranes were transferred to plastic-sandwich boxes, and washed with increasing stringency, starting with two washes at 2x SSC, 0.1% (w/v) SDS for 20 minutes each, one at room temperature and the next at 65°C. The subsequent washes consisted of one wash in 1x SSC, 0.1% (w/v) SDS at 65°C, two washes in 0.1x SSC, 0.1% (w/v) SDS at 70-75°C respectively. Depending on the strength of the signal as indicated by a hand-held Geiger counter, there was an optional further wash in 0.1x SSC, 0.1% (w/v) SDS at 80°C.

II Using DNA probes

The hybridisation method used for DNA probes was essentially the same as described above with the following differences: DNA probes were labelled using random priming as in section 2.2.12; prehybridisation and hybridisation steps were carried out at 42°C; and the washes were less stringent, with the maximum temperature used being 65°C.

III Visualisation and quantification of signal

The membranes were wrapped in cling-film and either exposed to X-ograph blue film at -70°C for an appropriate time or exposed to a phosphorimager screen (Molecular Dynamics).

2.2.23.8 Tissue section in situ hybridisation

I Preparation of tissue

Tissues were collected as described in section 2.2.23.1. All tissues were fixed overnight in freshly prepared 4% (w/v) paraformaldehyde in PBS, at 4°C with agitation. The fixative was removed by washing twice with 0.83% (w/v) NaCl for 30 minutes or longer if appropriate at 4°C with agitation, and the tissues dehydrated by successive washes for 30 minutes in 0.83% (w/v) saline/ ethanol mix (1:1) at 4°C, 70% (v/v) ethanol twice at room temperature, 85% (v/v) ethanol, 95% (v/v) ethanol, and absolute ethanol twice, all at room temperature with gentle agitation. The tissues were then cleared in Histoclear, twice for 30 minutes.

Fixed tissues were gradually saturated in freshly melted, filtered paraffin wax, starting with a Histoclear/wax mix (1:1) for 20 minutes with heating to 60°C, then three changes of wax only at 60°C for 20 minutes. The tissues were then transferred to warmed plastic embedding moulds, orientated in the wax, and cooled to 4°C on a cold plate. The paraffin blocks were stored with desiccant at 4°C for up to several months prior to sectioning.

II Preparation of sections

Glass slides (Superfrost- BDH) were washed by dipping briefly in 10% (v/v) HCl/70% (v/v) ethanol, followed by autoclaved milliQ water and then 95% (v/v) ethanol. The slides were dried at 80°C for 5-10 minutes. The slides were then coated with TESPA (3-aminopropyl-etoxysilane) by dipping for 10 seconds in a freshly prepared solution of 2% (v/v) TESPA in acetone, rinsed briefly in acetone twice, and once in autoclaved milliQ water, before drying overnight at 37°C. The slides were stored at room temperature until required. Sections were cut at 7-10μ thickness on a microtome (American Optical), ribbons collected, and appropriate lengths floated out on a 50°C water bath to remove creases, before being collected on TESPA-coated slides. The sections were dried flat on the slides overnight at 37°C, after which they were stored with desiccant at 4°C until required.

III Pre-treatment of tissue sections

The slides were allowed to warm to room temperature in their sealed box, and subsequently de-waxed by placing in Histoclear for 10 minutes twice, followed by an absolute ethanol rinse for 2 minutes to remove most of the Histoclear. The sections were rehydrated by passing quickly through absolute ethanol twice, then 95%, 85%, 70%, 50% and 30% ethanol.

The slides were then washed in 0.83% (w/v) NaCl, and PBS for 5 minutes each. The sections were fixed by incubation in freshly prepared 4% (w/v) paraformaldehyde in PBS for 20 minutes, followed by two 5 minute rinses in PBS. The slides were incubated in 20μg/ml proteinase K in 50mM Tris-Cl, 5mM EDTA (pH8.0) for 5 minutes, and rinsed in PBS for 5 minutes. The sections were again fixed by incubation in freshly prepared 4% (w/v) paraformaldehyde in PBS for 20 minutes, followed by a brief rinse in autoclaved milliQ water. The sections were acetylated by incubation in acetic anhydride in 0.1M triethanolamine (pH8.0) for 10 minutes, followed by a rinse in PBS, then 0.83% (w/v) NaCl for 5minutes each. The sections were finally dehydrated by passing through 30%, 50%, 70%, 85%, 95% and absolute ethanol for a few seconds each, apart from the 70% (v/v) ethanol rinse which was for 5 minutes to avoid salt deposits. The slides were left to air-dry under cover before subsequent hybridisation procedures were carried out later that day.

IV Hybridisation

The RNA probe (section 2.2.23.6- II) was resuspended in hybridisation solution to a final concentration of 8×10^4 cpm/μl. Prior to use this mix was heated at 80°C for 2 minutes. The probe was pipetted onto the slides at ≈ 2.5μl per cm^2 of coverslip

(usually 30µl was sufficient), and spread out over the sections using a coverslip, being careful to avoid trapping air bubbles. The slides were placed hori ontally in polystyrene boxes lined with 3MM Whatman paper soaked in 50% formamide, 5x SSC, and sealed with Nescofilm. The slides were incubated overnight at 60°C.

After hybridisation, the slides were transferred to 5x SSC, 0.1M DTT at 60°C for 30-60 minutes so that coverslips could fall off. Next, they were placed in 50% (v/v) formamide, 2x SSC, 0.2M DTT at 65°C for 30 minutes. To remove the DTT, the slides were washed three times in NTE buffer (0.5M NaCl, 10 mM Tris-Cl, 5mM EDTA, pH 8.0) at 37°C for 10 minutes each. The slides were then treated with 20µg/ml unboiled RNase A in NTE buffer at 37°C for 30 minutes, and rinsed in NTE buffer at 37°C for 15 minutes. Next, they were placed in 50% (v/v) formamide, 2x SSC, 0.2M DTT at 65°C for 30 minutes. They were rinsed in 2x SSC at 65°C for 15 minutes, followed by 0.1x SSC for 15 minutes, both at room temperature. The sections were dehydrated by passing quickly through 60% (v/v) ethanol, 80% (v/v) ethanol then 95 % (v/v) ethanol all containing 0.3M ammonium acetate, and finally absolute ethanol twice.

V Visualisation

The slides were initially exposed to BioMax-MR film at -70°C overnight, to establish a rough guide to the strength of the signal and background levels.
The slides were exposed to a photographic emulsion to obtain a higher resolution of the signal. Slides were dipped in the dark, in Ilford K5 emulsion diluted 1:1 in water, kept liquid at 42°C. After dipping, the slides were placed in a rack and left to drain vertically for 30 minutes, followed by draining horizontally for 90-120 minutes to allow the emulsion to harden. They were then placed in a light-tight box with desiccant at 4°C, and exposed for the appropriate length of time. Test slides were developed first, by allowing them to warm up to room temperature and developed in Ilford Developer solution for 5 minutes at 17°C, briefly rinsed in water with a few drops of glacial acetic acid, followed by 5 minutes in Ilford fixer solution, and finally copious rinsing in tap water. If the exposure was satisfactory (by briefly inspecting under dark field microscopy), the rest of the slides were developed in a similar fashion.
Sections were stained in 0.02% (w/v) trypan blue, destained in water. Next they were dehydrated by quickly passing through an ethanol series, followed by two 10 minute washes in Histoclear, after which they were immediately mounted in DPX Mountant. The sections were photographed using Olympus BH2.

2.2.23.9 Whole-mount in situ hybridisation

This method is adapted from Wilkinson and Nieto (1993).

I Pre-treatment of embryos

Embryos were obtained as described in 2.22.3.1, and fixed overnight in freshly prepared 4% (w/v) paraformaldehyde in PBS at 4°C with gentle shaking. Excess fixative was removed with two 5 minute rinses in PBT (PBS, 0.1% (v/v) Tween-20) at 4°C. The embryos were dehydrated by passing through a methanol series: 25%, 50%, 75% methanol/PBT, then 100% methanol twice for 5 minutes with gentle shaking. The embryos were then rehydrated by passing them through 75%, 50% and 25% methanol/PBT and washing twice in PBT for 5 minutes each. The embryos were bleached in 6% (v/v) hydrogen peroxide in PBT for 1 hour, then rinsed three times in PBT for 5 minutes each. This was followed by treatment with 10μg/ml of proteinase K in PBT for 15 minutes. washed with freshly prepared 2mg/ml glycine in PBT for 5 minutes, then rinsed twice in PBT for 5 minutes each. The embryos were fixed again by treating with freshly prepared 0.2% (v/v) glutaraldehyde/4% (w/v) paraformaldehyde in PBT for 20 minutes, then rinsed twice in PBT for 5 minutes each. The prehybridisation solution (50% deionised formamide, 5x SSC, pH 4.5, 50μg/ml yeast RNA, 1% (w/v) SDS, 50μg/ml heparin) was added and the embryos incubated overnight at 70°C with gentle shaking. The embryos were stored in the prehybridisation solution at -20°C until required.

II Hybridisation of embryos

The prehybridisation mix was replaced with hybridisation solution containing up to 1μg/ml digoxygenin-labelled RNA probe, and incubated overnight at 70°C.

III Post-hybridisation washes

The washes were as follows: solution 1 twice at 70°C for 30 minutes; solution 1/solution 2(1:1mix) at 70°C for 10 minutes; solution 2 only three times at room temperature for 5 minutes. The embryos were next treated with 100μg/ml RNase A in solution 2 at 37°C for 30 minutes. They were rinsed with solution 2, then solution 3 at room temperature for 5 minutes each. High stringency washes were carried out by twice incubating the embryos in solution 3 at 65°C for 30 minutes. Then they were rinsed three times in freshly prepared 1x TBST at room temperature for 5 minutes each. The embryos were preblocked in 10% (v/v) sheep serum in 1x TBST at room temperature for 60-90 minutes. To preadsorb the antibody, 3 mg of 14 dpc mouse embryo powder in 500μl of 1x TBST was heated to 70°C for 30 minutes, cooled on ice, and 5μl of sheep serum and 1μl of anti-digoxygenin antibody (Boehringer Mannheim)

were added. This was shaken gently at 4°C for 60 minutes, followed by centrifugation in a microcentrifuge for 10 minutes. The supernatant containing the preadsorbed antibody was removed and its volume increased to 2 ml with 1x TBST. The embryos were rocked overnight at 4°C with the preadsorbed antibody.

IV Post-antibody washes and visualisation

The washes were as follows: three times in 1x TBST for 5 minutes each; ten times in 1 x TBST for 30 minutes each; and an overnight wash in 1 x TBST at 4°C.
The embryos were washed three times in NTMT for 10 minutes, then incubated in NTMT containing NBT (75mg/ml in 70% dimethylformamide) and BCIP (50mg/ml in dimethylformamide) for 20 minutes in the Eppendorf tube, before being transferred to glass embryo dishes so that any colour reaction could be more easily observed. The visualisation steps were kept in the dark as much as possible. When the colour had developed to the desired extent, the embryos were washed twice in PBS/0.1% (v/v) Triton X-100, once for 30 minutes and then overnight.

V Photography

The embryos were positioned in agarose wells, viewed using an Nikon SMZ-U microscope and photographed with Nikon AF 801S camera. To section the embryos, they were re-fixed in freshly prepared 4% (w/v) paraformaldehyde in PBS at 4°C for 2 hours, then incubated in tetrahydronapthalene, and finally embedded in wax (section 2.2.23.8) 10μ sections were cut and mounted on TESPA-coated slides with DPX mountant. Slides were examined on a Nikon TMS microscope and photographed using a Nikon AF 801S camera.

VI New York Method for whole-mount *in situ* hybridisation

This method is a simpler version of the Wilkinson and Nieto protocol (Streit *et al.*, 1995). It permits the use of probes up to 3 kb in size, and the increased stringency of hybridisation and post-hybridisation washes is supposed to generate lower background signal. The riboprobe in hybridisation mix is re-used numerous times.
The main differences are in the solutions used for hybridisation and in the post-hybridisation washes, otherwise the protocol is as described above.
Briefly, embryos were pre-treated as described above minus the hydrogen peroxide step, and prehybridised at 70°C overnight. The prehybridisation solution was replaced with hybridisation solution containing DIG-labelled riboprobe at a final concentration of 100-200 ng/ml, and the embryos were incubated overnight at 70°C.
Posthybridisation washes were as follows: two washes with preheated hybridisation solution at 70°C for 5 minutes; three washes with preheated hybridisation solution at

70°C for 30 minutes; one wash with preheated hybridisation solution/ 1x TBST (1:1); three washes with 1x TBST at room temperature for 5 minutes; and three washes with 1x TBST at room temperature for 30 minutes. The embryos were blocked with 10% sheep serum, 1% BSA in 1x TBST for 3 hours, prior to the addition of preadsorbed antibody and incubation overnight at 4°C. Post antibody washes and visualisation were as described above.

CHAPTER 3 RESULTS

3.1 FOREWORD

The results chapter follows a logical order of sequential analysis of the 8dI-positive murine cDNA clones, identifying a novel gene subsequently called *Nfe2l1*. The *a priori* aim was to determine if the murine homologue of the *Drosophila cnc* gene had been identified. The elucidation of the full length sequence is an important first step in answering this question and also assists in the molecular characterisation of this gene's role, if any, in mouse development.

3.2 SEQUENCE ANALYSIS OF MURINE *Nfe2l1*

3.2.1 Preliminary characterisation of murine 8.5dpc cDNA clones

All six 8dI-positive bacteriophage tertiary clones (see section 1.5) were confirmed to be genuine by Southern blot analysis (Figure 3.1). The clone cnc20 produced two fragments upon *Eco* RI digestion (700 bp and 350 bp), but originally only the larger band was identified and subcloned, therefore the smaller fragment was not used in any of the following experiments. The bacteriophage clones were initially subcloned from the bacteriophage vector into the *Eco* RI site of pUC plasmid vector by Dr Johnstone. To make their subsequent analysis easier, the pUC subclones were digested with *Eco* RI and the cDNA fragments subcloned into the *Eco* RI site of pBluescript SK^+ (pBS SK^+) plasmid. The restriction endonuclease (RE) digestion of the pUC subclones showed that a number of isolated phage clones had internal *Eco* RI sites which had not been cut prior to subcloning into the pUC plasmid vector from the phage DNA (data not shown). In total, nine individual *Eco* RI cDNA fragments were subcloned into the pBS SK^+ vector (Table 3.1). These were hybridised to Southern blots containing RE digested mouse and human genomic cell line DNA, and washed to various stringencies (Figure 3.2).

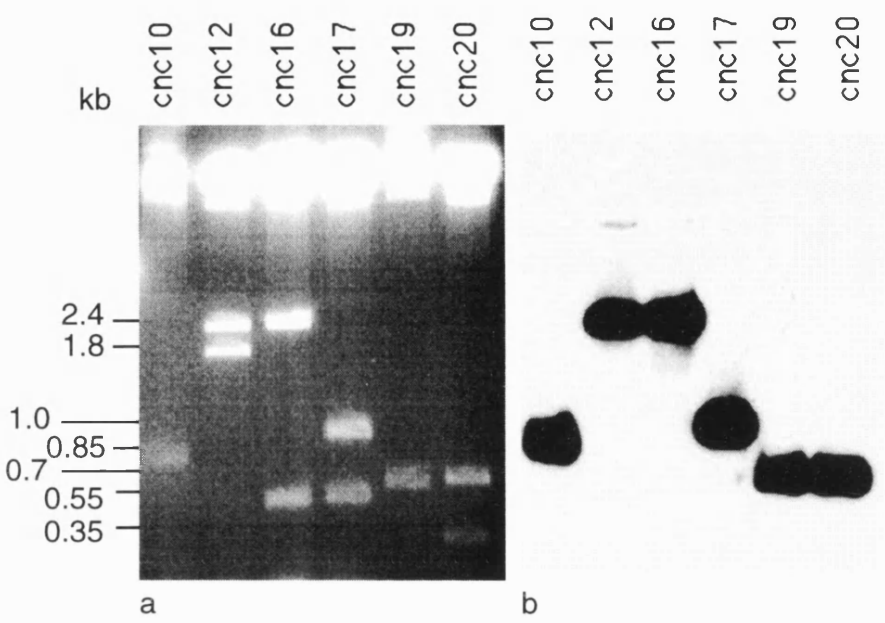


Figure 3.1 Analysis of 8dl-positive bacteriophage clones
(a)Bacteriophage clones were digested with Eco RI and electrophoresed. (b) A Southern blot was prepared, probed with 8dl, washed to a stringency of 0.1x SSC, 0.1% SDS and exposed to film for 1 hour at room temperature.

Initial clone name	Estimated insert size (bp)	Fragment name and size (bp) after RE digestion
8dI	800	800
cncI0	800	800
cncI2	4200	(cncI2SM)I800
		(cncI2LG) 2400
cncI6	2950	(cncI6SM) 550
		(cncI6LG) 2400
cncI7	I550	(cncI7SM) 550
		(cncI7LG) I000
cncI9	700	700
cnc20	700	700 (+ 350*)

Table 3.I. Summary of murine cDNA clones
*The initial six positive tertiary clones were digested with Eco RI, subcloned into pUC vector, and then into pBS SK+, resulting in a total of nine subclones.*350 bp fragment from cnc20 was missed in the subcloning from bacteriophage into plasmid vector.*

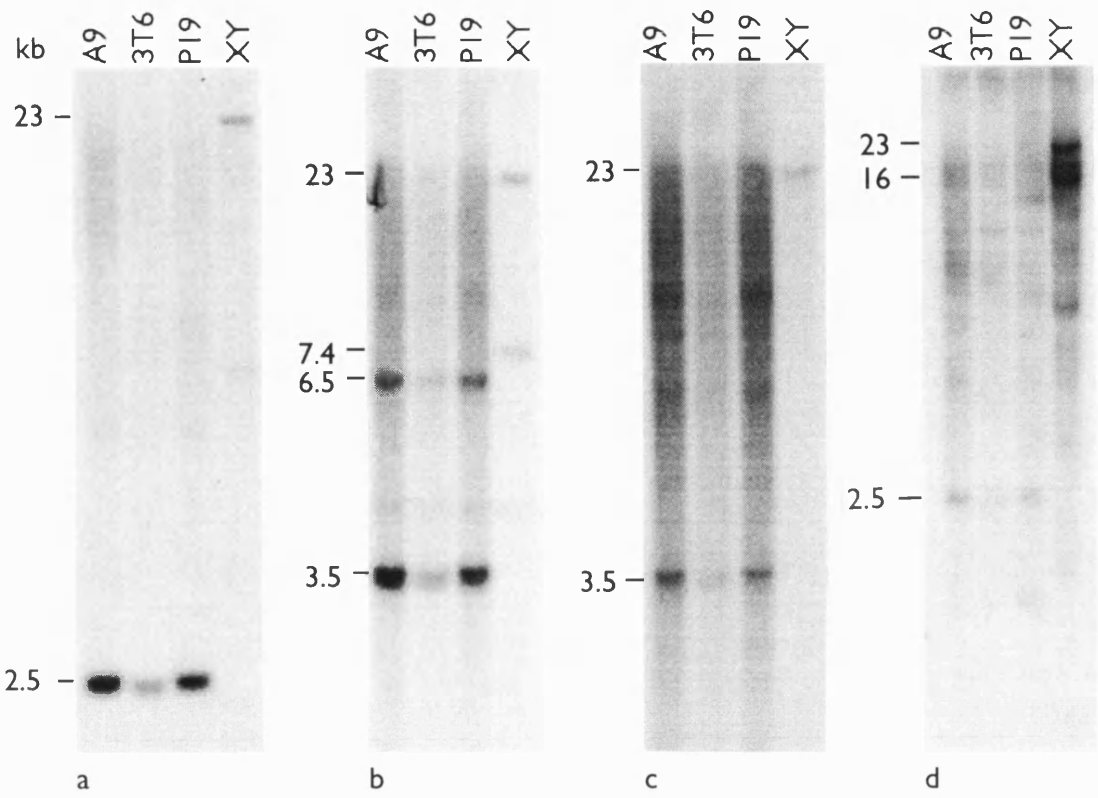


Figure 3.2 Murine cDNA clones hybridised to Eco RI-digested mouse and human genomic DNA

The fragments showed 3 patterns of hybridisation at a washing stringency of 0.1x SSC, 0.1% SDS: (a) probes cnc10,cnc12LG, cnc16LG, cnc17LG, cnc19 and cnc20; (b) cnc12SM and (c) cnc16SM and cnc17SM (the underlined cDNA clone is the actual probe used in the representative hybridisation shown). The result of the same hybridisation as shown in (a) is shown in (d) but at a lower washing stringency (3x SSC, 0.1% SDS). The filters were exposed to film for 1 to 4 nights at -70°C.

The patterns displayed by the *Nfe2l1* cDNAs allowed three distinct subgroups to be formed. The multiple bands in the XY lane detected by cnc16LG (Figure 3.1d) at a low stringency wash (3x SSC, 0.1% SDS) suggested the existence of other related DNA sequences in the human genome.

3.2.2 Sequencing of 8.5 dpc cDNA clones

The cloning site of each end of the murine cDNA fragments was sequenced to determine if it was a linkered *Eco* RI site or an internal *Eco* RI site, as the preliminary sequence analysis of 8dI, one clone, by Dr. K.A. Johnstone had shown that at its 5' end there was an internal *Eco* RI site. Restriction enzyme mapping was also carried out, and along with the Southern analysis (Figure 3.2), the murine cDNA fragments were ordered and

an approximate map of the different cDNA fragments was assembled to facilitate the sequencing strategy (Figure 3.3).

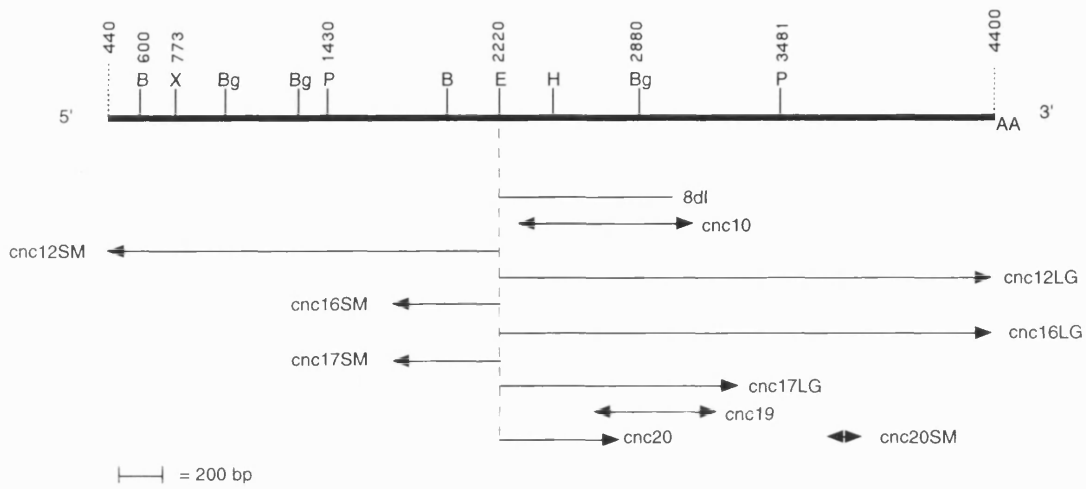


Figure 3.3 Preliminary restriction enzyme map and ordering of murine cnc-like cDNA clones

The numbering refers to the nucleotide position in the full length sequence of Nfe2l1 (Figure 3.10). Good quality sequence was not obtained from cnc19, and it was only accurately placed on the map when cnc16LG was fully sequenced. The presence of cnc20SM was missed in the original bacteriophage subcloning. The arrowheads denote the presence of a linker.
B= Bam HI, Bg= Bgl II, E= Eco RI, H= Hind III, P= Pst I, X= Xho I

The majority of the sequencing was done manually by subcloning various fragments using the above restriction enzyme site map as a guide. Either pBS SK$^+$ for double stranded sequencing or M13 phagemid vector for single stranded sequencing were used and the dideoxy chain termination method employed to sequence bidirectionally. For fragments that were unable to be wholly sequenced by using vector primers, primer walking was employed (Table 2.1). The most 5' cDNA clone, cnc12SM, was found upon sequencing and subsequent translation to have an ORF throughout its length but with the first in-frame methionine at nucleotide position 1268 (Figure 3.10). Also since preliminary expression analysis suggested a mRNA transcript size of approximately 4.4 kb (see section 3.4), concurrent attempts were made to isolate a more 5' cDNA to ensure that the entire coding region was cloned.

3.2.3 Isolation and analysis of full length murine *Nfe2l1* cDNA

In an attempt to clone more 5' cDNA fragments two approaches were taken: screening another mouse cDNA library (random- and oligo-d(T) primed) with a cDNA probe derived from the 5' end of cnc12SM, and 5' RACE (**R**apid **A**mplification of **c**DNA **E**nds).

3.2.3.1 Screening of mouse embryo 10 dpc cDNA library

Approximately 600 000 plaques of a mouse embryo 10 dpc cDNA library were screened with a *Xho* I fragment of cnc12SM (Figure 3.10, nuc $_{440\ to\ 773}$), and one tertiary positive (lab name 10.5) was obtained at a washing stringency of 2x SSC, 0.1% SDS at 65°C. The insert was released from the bacteriophage lambda by a restriction enzyme digest using *Eco* RI and *Hind* III: a 700bp *Eco* RI fragment (10.5RI) and a 600 bp *Eco* RI and *Hind* III fragment (10.5RH) were detected (Figure 3.4). Both fragments were initially subcloned into pBS SK$^+$, and then into M13 for single-stranded sequencing. Primer walking was also employed (Table 2.1).

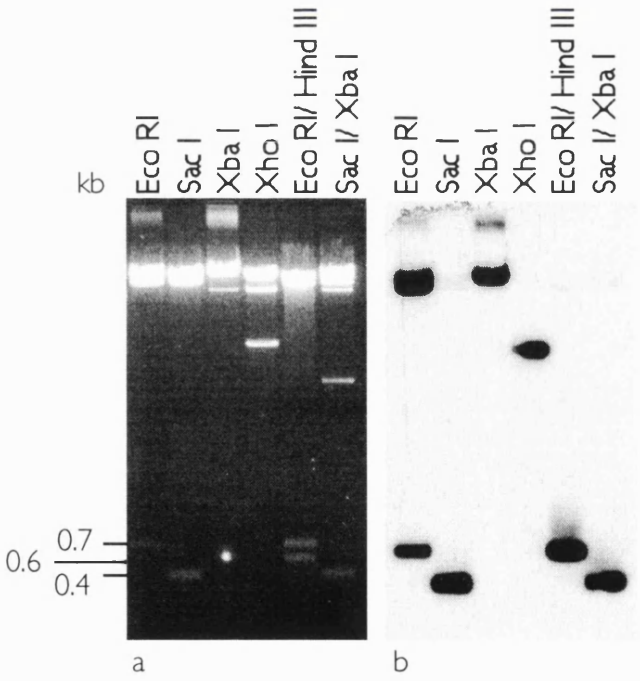


Figure 3.4 RE digests and Southern blot analysis of clone 10.5
The insert was released by digestion with Eco RI and Hind III, as shown. The gel was blotted and probed with the Xho I fragment of cnc12SM used in the library screening. The filter was washed at 0.1x SSC, 0.1% SDS at 65°C, and exposed for 3 hours at room temperature.

Sequence analysis of clone 10.5 revealed that it was not a straightforward transcript; it appeared to be composed of two or three different cDNAs (Figure 3.5).

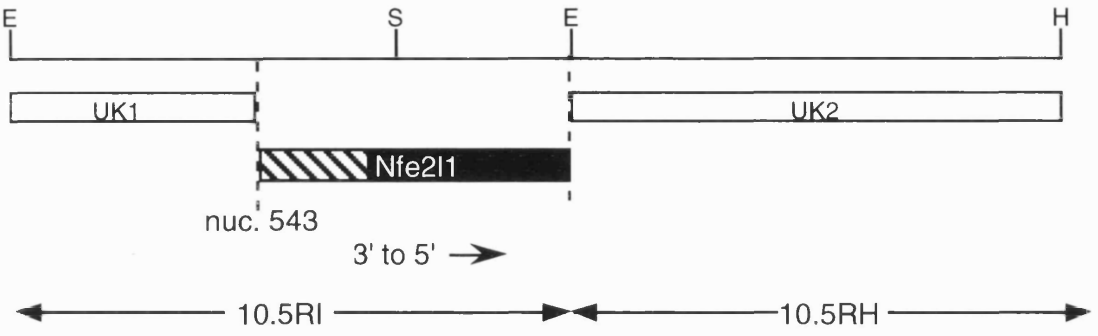


Figure 3.5 Schematic of clone 10.5, incorporating sequence and RE analyses
UK= unknown sequence
E= Eco RI, S= Sac I, H= Hind III
Shaded box= Nfe2l1 coding region
Solid box= Nfe2l1 5'UTR

It is most likely that the *Eco* RI fragment (10.5RI) and the *Eco* RI/*Hind* III fragment (10.5RH) form a co-ligation for a number of reasons; firstly, the incorporation of the size of 10.5RH would be inconsistent with the estimated size of the *Nfe2l1* mRNA; secondly, the library inserts are released using *Eco* RI and *Hind* III, so *Eco* RI/internal *Eco* RI co-ligations could potentially occur during library production; and thirdly, Southern analysis using the 10.5RH fragment showed a band pattern that was very different to that of *Nfe2l1* (Figure 3.6a,b): the multiple bands present suggest that the 10.5RH is either composed of spliced exonic DNA or identifies several highly-related sequences, both inconsistent with its putative role of encoding 5'UTR of *Nfe2l1*. Database comparisons of sequence from the ends of 10.5RH gave no significant matches at either the DNA or protein level (data not shown).

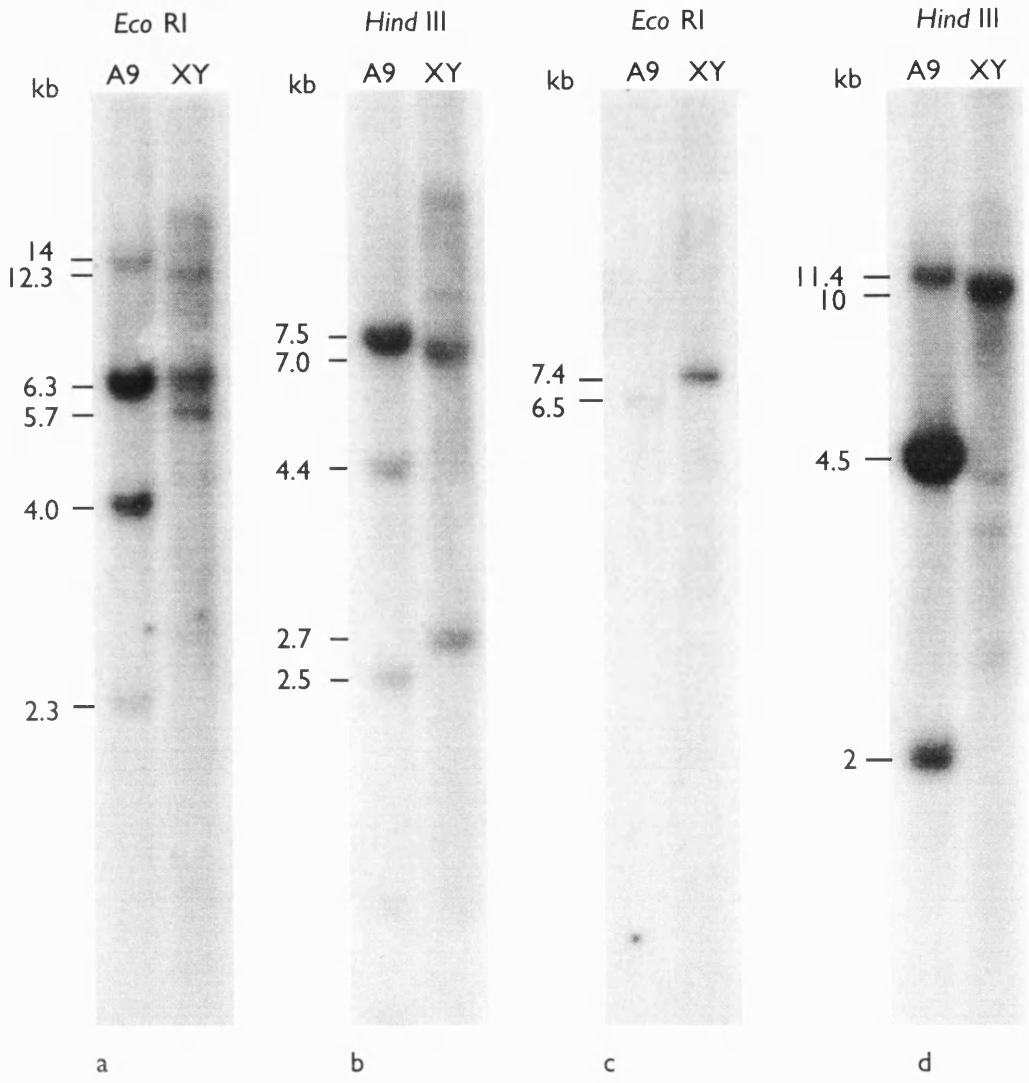


Figure 3.6 Southern genomic DNA analysis of clone 10.5
Southern blots were prepared from Eco RI- and Hind III-digested mouse (A9) and human (XY) genomic DNA. Hybridisations were carried out using 10.5RH (a & b) and 10.5RI (c & d). The filters were washed to a stringency of 2x SSC, 0.1% SDS at 65°C, and exposed to film for 3 nights at -70°C.

The subclone, 10.5RI, was fully sequenced, and the UK1 fragment sequence and translation is shown in Figure 3.7. Southern analysis using the 10.5 RI fragment (Figure 3.6c,d) showed some bands in common with those detected by cnc12SM (Figure 3.2b and 3.16) and only an additional one (11.4 kb) in the *Hind* III-digested mouse genomic lane. The *Eco* RI genomic digest had a single band only in both mouse and human lanes, whose sizes were consistent with those seen with cnc12SM. If the *Eco* RI fragment 10.5RI represented an unprocessed mRNA, UK1 should be linked by a 5' splice site consensus sequence (AG/GT) but it is not (AG/AG). UK1 could represent a splice variant of *Nfe2l* : the consensus 3' exon sequence AG ($Nfe2l1_{nuc542-543}$) is

present but no open reading frame (ORF) in the appropriate direction. However a potential transcriptional start site is present at UK1 ($nuc_{116-118}$), with an adequate Kozak consensus sequence. There is an ORF in the opposite direction, and database searches have shown significant homology at the DNA level with a number of sequences (Figure 3.8), all of which read in the 5' to 3' direction. Sequence from UK1 was amplified in a PCR reaction (Table 2.1), and the resulting 249 bp product was gel purified and used as a probe on a human foetal northern (Clontech) and a mouse embryonic northern. The UK1 PCR-generated probe gave no signal on either northern, after washing to a stringency of 2x SSC, 0.1% SDS at 65°C, and exposure to film for a week at -70°C (data not shown).

Alternatively, UK1 possibly represents another coligation event, with the subsequent loss of restriction enzyme site. The library used was commercially obtained with little detail on its construction and attempts to communicate with the manufacturer have been unfruitful, which may have helped to explain the formation of clone 10.5. The sequence of the *Nfe2l1* portion of clone 10.5 covers nucleotides 19 to 543 (Figure 3.10), with a 6 bp addition at the 5' end (AAGCGA).

```
   AAGCCTTCTGCTCTCCAATGGACGTGATGGTAGCTACCTGCTGCGGGACAGCAATGAGCA
 1 ---------+---------+---------+---------+---------+---------+ 60
   TTCGGAAGACGAGAGGTTACCTGCACTACCATCGATGGACGACGCCCTGTCGTTACTCGT

     A  F  C  S  P  M  D  V  M  V  A  T  C  C  G  T  A  M  S  R

   GACTGGGCTCTACTCTCTCCTCTGTGAGAGCCAAAGACTCTGTCAAACACTTTCATGTTG
61 ---------+---------+---------+---------+---------+---------+ 120
   CTGACCCGAGATGAGAGAGGAGACACTCTCGGTTTCTGAGACAGTTTGTGAAAGTACAAC

     L  G  S  T  L  S  S  V  R  A  K  D  S  V  K  H  F  H  V  E

   AATATACTGGGTACTCGTTTAAATTCGGCTTTAATGAGTATTCATCTCTAAAGGATTTGT
121 ---------+---------+---------+---------+---------+---------+ 180
   TTATATGACCCATGAGCAAATTTAAGCCGAAATTACTCATAAGTAGAGATTTCCTAAACA

     Y  T  G  Y  S  F  K  F  G  F  N  E  Y  S  S  L  K  D  L  S

   CAAGCATTTGCAAATCAGCCTTTGATTGGAAGCGAGACAGGACTCTGATGGTTTTGAAGC
181 ---------+---------+---------+---------+---------+---------+ 240
   GTTCGTAAACGTTTAGTCGGAAACTAACCTTCGCTCTGTCCTGAGACTACCAAAACTTCG

     S  I  C  K  S  A  F  D  W  K  R  D  R  T  L  M  V  L  K  H

   ATCCCTATCCAAGGAAGTGGAAGAACCTTGCATTTACGAATCAGTCCAAATTCACACAGC
241 ---------+---------+---------+---------+---------+---------+ 300
   TAGGGATAGGTTCCTTCACCTTCTTGGAACGTAAATGCTTAGTCAGGTTTAAGTGTGTCG

     P  Y  P  R  K  W  K  N  L  A  F  T  N  Q  S  K  F  T  Q  R

   GATGCAGACAGGAAGGACAGAGAACGACCTCATACCCACTGCGCCTTCT
301 ---------+---------+---------+---------+--------- 349
   CTACGTCTGTCCTTCCTGTCTCTTGCTGGAGTATGGGTGACGCGGAAGA

     C  R  Q  E  G  Q  R  T  T  S  Y  P  L  R  L
```

Figure 3.7 UK1 complete sequence and ORF translation
The sequence was translated using the GCG package at HGMP Resource Centre, Hinxton. The only potential full length ORF (+3) is shown and goes in the opposite direction to that of Nfe2l1 (Figure 3.5). Kozak consensus sequence underlined.

```
Query= 10.5uk, 349 bases, 4BF36591 checksum.
         (349 letters)

Database: est
           1,682,497 sequences; 634,576,516 total letters

Searching.....................................................done
                                                                  High    E
Sequences producing significant alignments:                      Score  Value

EM:MMA64488 AA064488 ml49f05.r1 Stratagene mouse testis (#93730..  85.7  3e-15
EM:HS1166385 AA251658 zs10f03.r1 NCI_CGAP_GCB1 Homo sapiens cDN..  65.9  3e-09
EM:SSC20B10 F14802 S.scrofa mRNA; expressed sequence tag (5'; c..  58.0  7e-07
EM:AA583339 AA583339 nn45f11.s1 NCI_CGAP_GC5 Homo sapiens cDNA ..  38.2  0.65
EM:AA667431 AA667431 vv17g02.r1 Stratagene mouse heart (#937316..  38.2  0.65
EM:MMAA52800 AA152800 mr85f07.r1 Stratagene mouse heart (#93731..  36.2  2.6

> EM:MMA64488 AA064488 ml49f05.r1 Stratagene mouse testis (#937308)
          Mus musculus cDNA clone 515361 5'.
          Length = 461

Score = 85.7 bits (43), Expect = 3e-15
Identities = 50/51 (98%), Positives = 50/51 (98%), Gaps = 1/51 (1%)


Query 220 ggactctgatggttttgaagcatccctatccaa-ggaagtggaagaacctt 269
          ||||||||||||||||||||||||||||||||| |||||||||||||||||
Sbjct 411 ggactctgatggttttgaagcatccctatccaagggaagtggaagaacctt 461

> EM:HS1166385 AA251658 zs10f03.r1 NCI_CGAP_GCB1 Homo sapiens cDNA
          clone IMAGE:684797 5' similar to TR:G1000069 G1000069
          RAC-ALPHA SERINE/THREONINE KINASE HOMOLOG. ;.
          Length = 472

Score = 65.9 bits (33), Expect = 3e-09
Identities = 96/113 (84%), Positives = 96/113 (84%), Gaps = 3/113 (2%)


Query 212 gcgagacagg-actctgatggttttgaagcatccctatccaaggaa-gtggaagaacctt 269
          |||||||||| |||||||||||| | || |||||||| ||||| || |||||||||||| |
Sbjct 1   gcgagacaggcactctgatggttctaaaacatccctacccaagaaaagtggaagaaccct 60


Query 270 gcatttacgaatcagtc-caaattcacacagcgatgcagacaggaaggacaga 321
           |||||| ||||| ||| |   |||||||||| |||||||||||||| |||||
Sbjct 61  ccatttatgaatctgtcgcgggttcacacagcaatgcagacaggaagaacaga 113

> EM:SSC20B10 F14802 S.scrofa mRNA; expressed sequence tag (5';
          clone c20b10)
          Length = 276

Score = 58.0 bits (29), Expect = 7e-07
Identities = 51/59 (86%), Positives = 51/59 (86%)


Query 291 ttcacacagcgatgcagacaggaaggacagagaacgacctcatacccactgcgccttct 349
          |||||||||| |||||||||| | |||||||||||||| | |||||||| ||||||
Sbjct 5   ttcacacagcaatgcagacaggnngaacagagaacgacctggtgcccactgcaccttct 63
```

Figure 3.8 Results of BLAST analysis of UK1 sequence
Both DNA and protein databases were searched with UK1 sequence. Searches with significant matches are shown.

3.2.3.2 5' RACE

The oligonucleotides designed for sequencing the 10.5 cDNA clone (Table 2.1) were also utilised for a 5' RACE experiment using the Clontech 5' RACE kit (section 2.2.15.1). The template was 5'-tagged mouse brain cDNA, and after following the protocol, one band (310bp) against a background smear was detected (Figure 3.9a) and subcloned into *Srf* I-digested PCR-Script SK$^+$ vector . It was initially confirmed to be the 5' portion of *Nfe2l1* by hybridisation (Figure 3.9b), and then sequenced. It matched the sequence obtained from clone 10.5 (section 3.2.3.1) apart from the very 5'end, represented by nucleotides 1 to 18 (Figure 3.10).

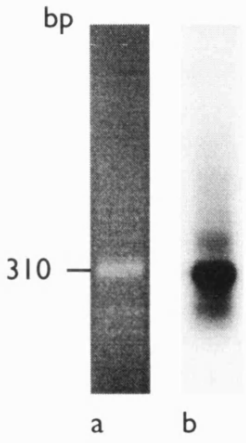
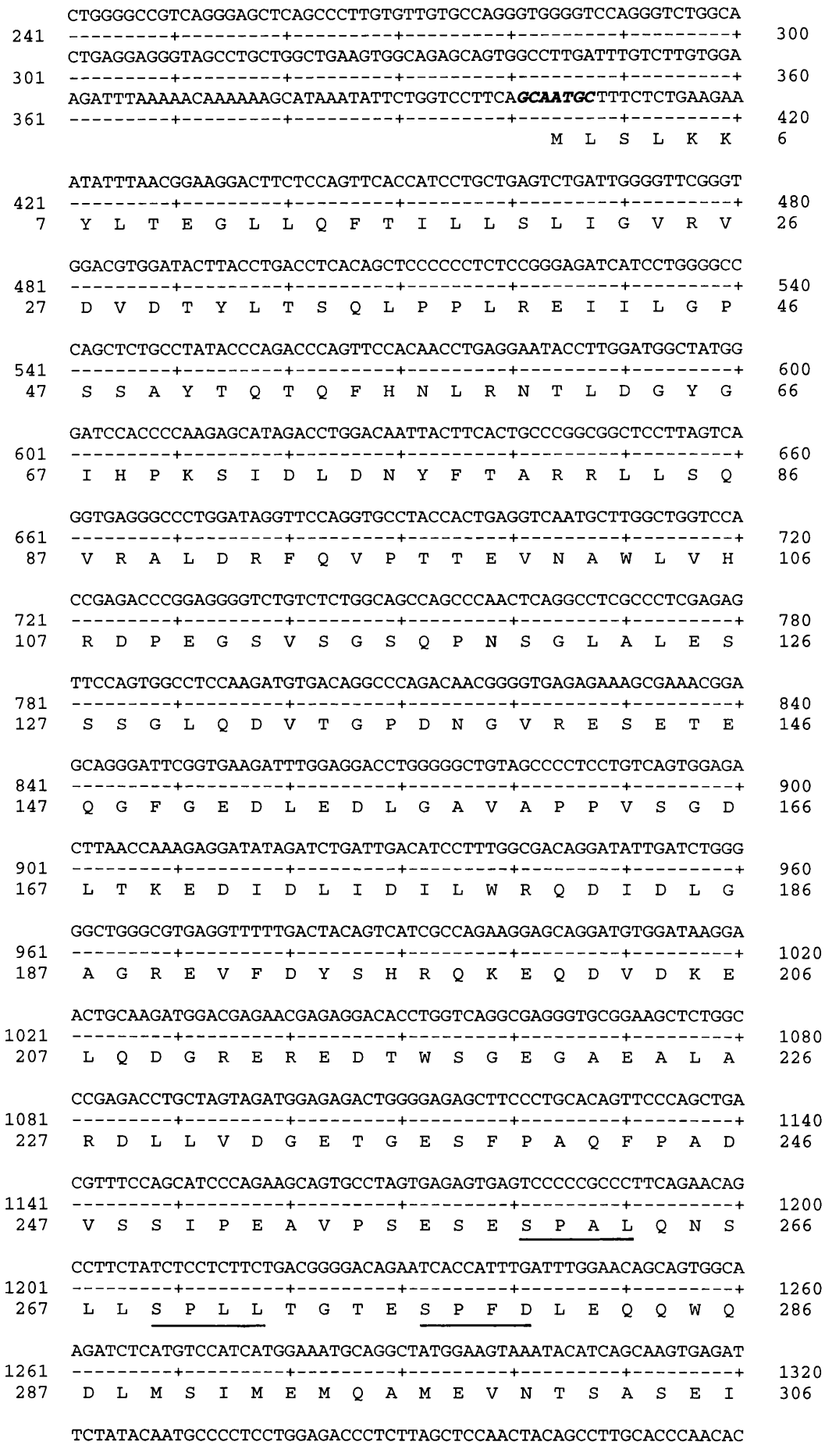


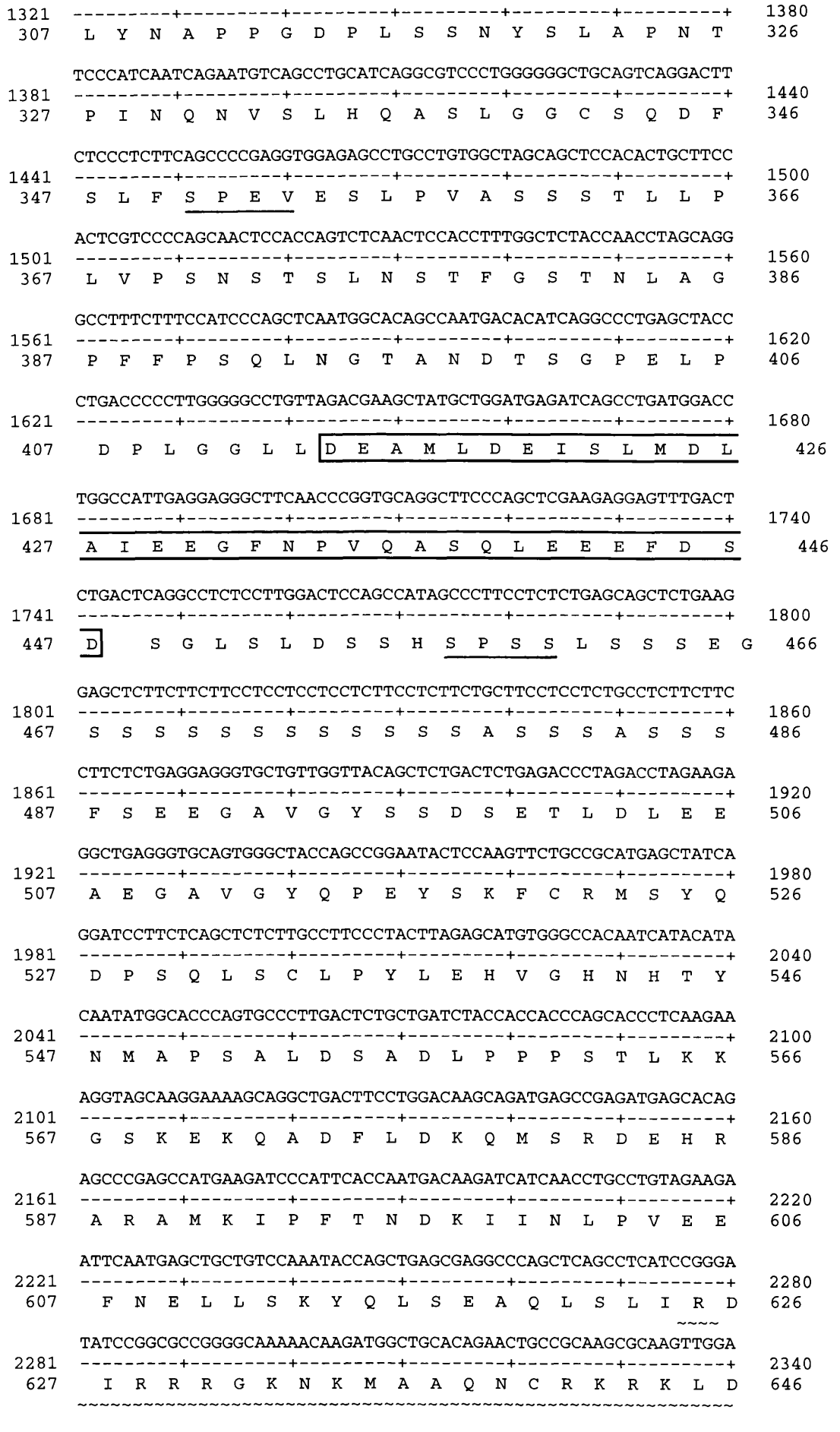
Figure 3.9 5'RACE product for Nfe2l1
(a) After following the manufacturer's instructions (Clontech), the electrophoresed RACE product gave a 310 bp band, detected against a background smear. (b) A Southern blot was prepared from the gel and hybridised with clone 10.5RI fragment, which gave a strong specific signal. The blot was washed to a stringency of 0.1x SSC, 0.1% SDS at 65°C, and exposed to film for 30 minutes at room temperature.

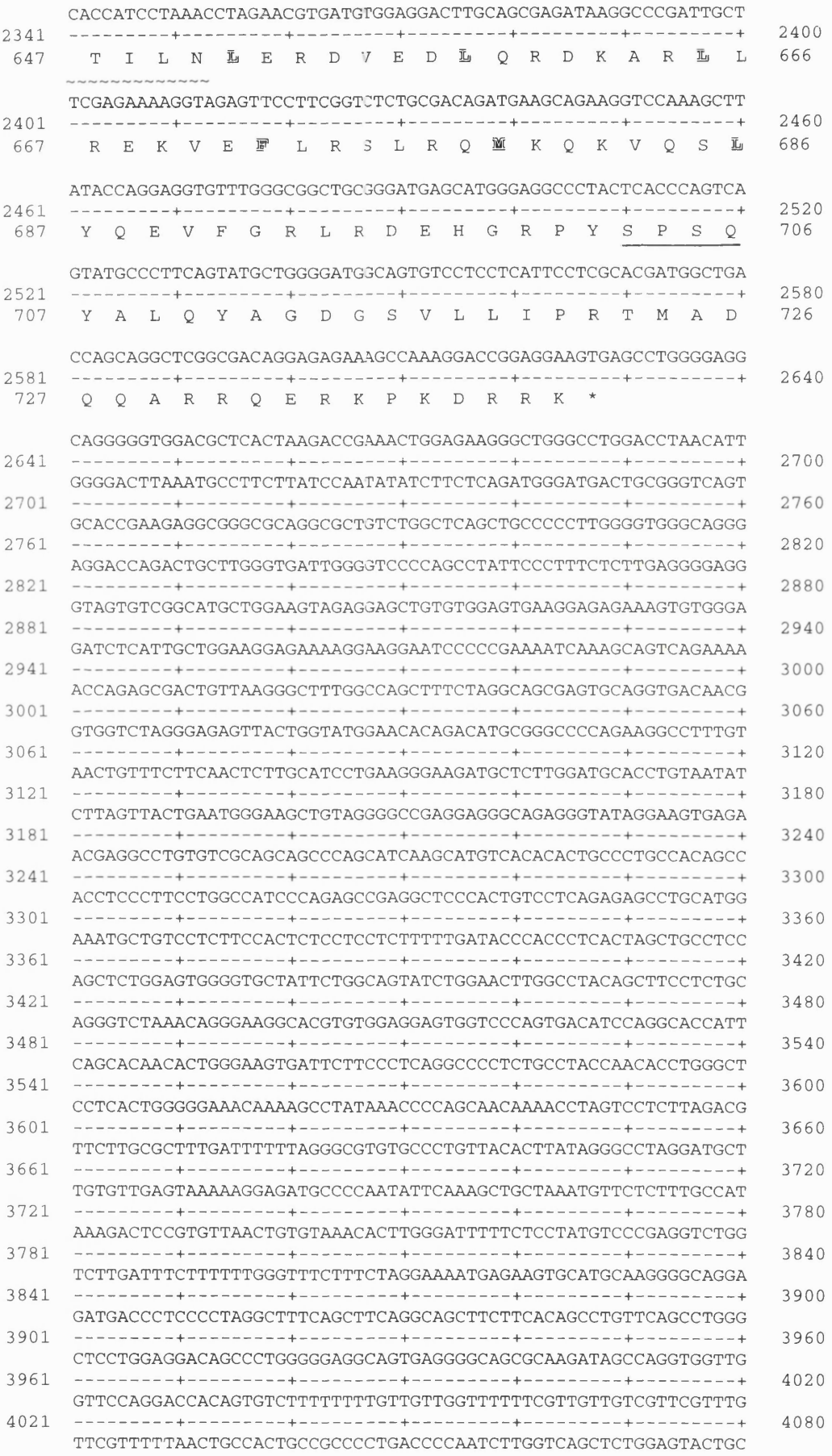
3.2.3.3 Complete sequence and analysis of murine Nfe2l1

The sequence obtained from the 8.5 dpc cDNAs, the 10.5 cnc cDNA and the 5' RACE products was assembled and translated using the GCMG package at the HGMP Resource Centre (Figure 3.10).

```
    GAGAAAAAAGGAGGAAAGAAAGACTCCATCAGCTGTTGGTGTGAGCAGCCTGCAGTGGAT
  1 ---------+---------+---------+---------+---------+---------+   60
    CTCCATAGAGGACAACTAATGTGAATCCGAAAGTGACTGTGTATGGGCTGTGGTGAAGTA
 61 ---------+---------+---------+---------+---------+---------+  120
    AGTTGAGGGCCCTGAGCGCCTGGACTGTGTTAGGAACGATCGAACGGCTCAACTTTGCGA
121 ---------+---------+---------+---------+---------+---------+  180
    GGTGAGGTGTCAAAAAGGGAAAAGTGAATGTGCGTTTCGCTCCACGGGGTGTGCTGTTGT
181 ---------+---------+---------+---------+---------+---------+  240
```


CTGGGGCCGTCAGGGAGCTCAGCCCTTGTGTTGTGCCAGGGTGGGGTCCAGGGTCTGGCA
241 300
CTGAGGAGGGTAGCCTGCTGGCTGAAGTGGCAGAGCAGTGGCCTTGATTTGTCTTGTGGA
301 360
AGATTTAAAAACAAAAAAGCATAAATATTCTGGTCCTTCAGCAATGCTTTCTCTGAAGAA
361 420
M L S L K K 6
ATATTTAACGGAAGGACTTCTCCAGTTCACCATCCTGCTGAGTCTGATTGGGGTTCGGGT
421 480
7 Y L T E G L L Q F T I L L S L I G V R V 26
GGACGTGGATACTTACCTGACCTCACAGCTCCCCCCTCTCCGGGAGATCATCCTGGGGCC
481 540
27 D V D T Y L T S Q L P P L R E I I L G P 46
CAGCTCTGCCTATACCCAGACCCAGTTCCACAACCTGAGGAATACCTTGGATGGCTATGG
541 600
47 S S A Y T Q T Q F H N L R N T L D G Y G 66
GATCCACCCCAAGAGCATAGACCTGGACAATTACTTCACTGCCCGGCGGCTCCTTAGTCA
601 660
67 I H P K S I D L D N Y F T A R R L L S Q 86
GGTGAGGGCCCTGGATAGGTTCCAGGTGCCTACCACTGAGGTCAATGCTTGGCTGGTCCA
661 720
87 V R A L D R F Q V P T T E V N A W L V H 106
CCGAGACCCGGAGGGGTCTGTCTCTGGCAGCCAGCCCAACTCAGGCCTCGCCCTCGAGAG
721 780
107 R D P E G S V S G S Q P N S G L A L E S 126
TTCCAGTGGCCTCCAAGATGTGACAGGCCCAGACAACGGGGTGAGAGAAAGCGAAACGGA
781 840
127 S S G L Q D V T G P D N G V R E S E T E 146
GCAGGGATTCGGTGAAGATTTGGAGGACCTGGGGGCTGTAGCCCCTCCTGTCAGTGGAGA
841 900
147 Q G F G E D L E D L G A V A P P V S G D 166
CTTAACCAAAGAGGATATAGATCTGATTGACATCCTTTGGCGACAGGATATTGATCTGGG
901 960
167 L T K E D I D L I D I L W R Q D I D L G 186
GGCTGGGCGTGAGGTTTTTGACTACAGTCATCGCCAGAAGGAGCAGGATGTGGATAAGGA
961 1020
187 A G R E V F D Y S H R Q K E Q D V D K E 206
ACTGCAAGATGGACGAGAACGAGAGGACACCTGGTCAGGCGAGGGTGCGGAAGCTCTGGC
1021 1080
207 L Q D G R E R E D T W S G E G A E A L A 226
CCGAGACCTGCTAGTAGATGGAGAGACTGGGGAGAGCTTCCCTGCACAGTTCCCAGCTGA
1081 1140
227 R D L L V D G E T G E S F P A Q F P A D 246
CGTTTCCAGCATCCCAGAAGCAGTGCCTAGTGAGAGTGAGTCCCCCGCCCTTCAGAACAG
1141 1200
247 V S S I P E A V P S E S E S P A L Q N S 266
CCTTCTATCTCCTCTTCTGACGGGGACAGAATCACCATTTGATTTGGAACAGCAGTGGCA
1201 1260
267 L L S P L L T G T E S P F D L E Q Q W Q 286
AGATCTCATGTCCATCATGGAAATGCAGGCTATGGAAGTAAATACATCAGCAAGTGAGAT
1261 1320
287 D L M S I M E M Q A M E V N T S A S E I 306
TCTATACAATGCCCCTCCTGGAGACCCTCTTAGCTCCAACTACAGCCTTGCACCCAACAC


1321 307 L Y N A P P G D P L S S N Y S L A P N T 326
TCCCATCAATCAGAATGTCAGCCTGCATCAGGCGTCCCTGGGGGGCTGCAGTCAGGACTT
1381 1440
327 P I N Q N V S L H Q A S L G G C S Q D F 346
CTCCCTCTTCAGCCCCGAGGTGGAGAGCCTGCCTGTGGCTAGCAGCTCCACACTGCTTCC
1441 1500
347 S L F S P E V E S L P V A S S S T L L P 366
ACTCGTCCCCAGCAACTCCACCAGTCTCAACTCCACCTTTGGCTCTACCAACCTAGCAGG
1501 1560
367 L V P S N S T S L N S T F G S T N L A G 386
GCCTTTCTTTCCATCCCAGCTCAATGGCACAGCCAATGACACATCAGGCCCTGAGCTACC
1561 1620
387 P F F P S Q L N G T A N D T S G P E L P 406
CTGACCCCCTTGGGGGCCTGTTAGACGAAGCTATGCTGGATGAGATCAGCCTGATGGACC
1621 1680
407 D P L G G L L D E A M L D E I S L M D L 426
TGGCCATTGAGGAGGGCTTCAACCCGGTGCAGGCTTCCCAGCTCGAAGAGGAGTTTGACT
1681 1740
427 A I E E G F N P V Q A S Q L E E E F D S 446
CTGACTCAGGCCTCTCCTTGGACTCCAGCCATAGCCCTTCCTCTCTGAGCAGCTCTGAAG
1741 1800
447 D S G L S L D S S H S P S S L S S S E G 466
GAGCTCTTCTTCTTCCTCCTCCTCCTCTTCCTCTTCTGCTTCCTCCTCTGCCTCTTCTTC
1801 1860
467 S S S S S S S S S S S S A S S S A S S S 486
CTTCTCTGAGGAGGGTGCTGTTGGTTACAGCTCTGACTCTGAGACCCTAGACCTAGAAGA
1861 1920
487 F S E E G A V G Y S S D S E T L D L E E 506
GGCTGAGGGTGCAGTGGGCTACCAGCCGGAATACTCCAAGTTCTGCCGCATGAGCTATCA
1921 1980
507 A E G A V G Y Q P E Y S K F C R M S Y Q 526
GGATCCTTCTCAGCTCTCTTGCCTTCCCTACTTAGAGCATGTGGGCCACAATCATACATA
1981 2040
527 D P S Q L S C L P Y L E H V G H N H T Y 546
CAATATGGCACCCAGTGCCCTTGACTCTGCTGATCTACCACCACCCAGCACCCTCAAGAA
2041 2100
547 N M A P S A L D S A D L P P P S T L K K 566
AGGTAGCAAGGAAAAGCAGGCTGACTTCCTGGACAAGCAGATGAGCCGAGATGAGCACAG
2101 2160
567 G S K E K Q A D F L D K Q M S R D E H R 586
AGCCCGAGCCATGAAGATCCCATTCACCAATGACAAGATCATCAACCTGCCTGTAGAAGA
2161 2220
587 A R A M K I P F T N D K I I N L P V E E 606
ATTCAATGAGCTGCTGTCCAAATACCAGCTGAGCGAGGCCCAGCTCAGCCTCATCCGGGA
2221 2280
607 F N E L L S K Y Q L S E A Q L S L I R D 626
TATCCGGCGCCGGGGCAAAAACAAGATGGCTGCACAGAACTGCCGCAAGCGCAAGTTGGA
2281 2340
627 I R R R G K N K M A A Q N C R K R K L D 646


CACCATCCTAAACCTAGAACGTGATGTGGAGGACTTGCAGCGAGATAAGGCCCGATTGCT
2341 2400
647 T I L N L E R D V E D L Q R D K A R L L 666
TCGAGAAAAGGTAGAGTTCCTTCGGTCTCTGCGACAGATGAAGCAGAAGGTCCAAAGCTT
2401 2460
667 R E K V E F L R S L R Q M K Q K V Q S L 686
ATACCAGGAGGTGTTTGGGCGGCTGCGGGATGAGCATGGGAGGCCCTACTCACCCAGTCA
2461 2520
687 Y Q E V F G R L R D E H G R P Y S P S Q 706
GTATGCCCTTCAGTATGCTGGGGATGGCAGTGTCCTCCTCATTCCTCGCACGATGGCTGA
2521 2580
707 Y A L Q Y A G D G S V L L I P R T M A D 726
CCAGCAGGCTCGGCGACAGGAGAGAAAGCCAAAGGACCGGAGGAAGTGAGCCTGGGGAGG
2581 2640
727 Q Q A R R Q E R K P K D R R K *
CAGGGGGTGGACGCTCACTAAGACCGAAACTGGAGAAGGGCTGGGCCTGGACCTAACATT
2641 2700
GGGGACTTAAATGCCTTCTTATCCAATATATCTTCTCAGATGGGATGACTGCGGGTCAGT
2701 2760
GCACCGAAGAGGCGGGCGCAGGCGCTGTCTGGCTCAGCTGCCCCCTTGGGGTGGGCAGGG
2761 2820
AGGACCAGACTGCTTGGGTGATTGGGGTCCCCAGCCTATTCCCTTTCTCTTGAGGGGAGG
2821 2880
GTAGTGTCGGCATGCTGGAAGTAGAGGAGCTGTGTGGAGTGAAGGAGAGAAAGTGTGGGA
2881 2940
GATCTCATTGCTGGAAGGAGAAAAGGAAGGAATCCCCCGAAAATCAAAGCAGTCAGAAAA
2941 3000
ACCAGAGCGACTGTTAAGGGCTTTGGCCAGCTTTCTAGGCAGCGAGTGCAGGTGACAACG
3001 3060
GTGGTCTAGGGAGAGTTACTGGTATGGAACACAGACATGCGGGCCCCAGAAGGCCTTTGT
3061 3120
AACTGTTTCTTCAACTCTTGCATCCTGAAGGGAAGATGCTCTTGGATGCACCTGTAATAT
3121 3180
CTTAGTTACTGAATGGGAAGCTGTAGGGGCCGAGGAGGGCAGAGGGTATAGGAAGTGAGA
3181 3240
ACGAGGCCTGTGTCGCAGCAGCCCAGCATCAAGCATGTCACACACTGCCCTGCCACAGCC
3241 3300
ACCTCCCTTCCTGGCCATCCCAGAGCCGAGGCTCCCACTGTCCTCAGAGAGCCTGCATGG
3301 3360
AAATGCTGTCCTCTTCCACTCTCCTCCTCTTTTTGATACCCACCCTCACTAGCTGCCTCC
3361 3420
AGCTCTGGAGTGGGGTGCTATTCTGGCAGTATCTGGAACTTGGCCTACAGCTTCCTCTGC
3421 3480
AGGGTCTAAACAGGGAAGGCACGTGTGGAGGAGTGGTCCCAGTGACATCCAGGCACCATT
3481 3540
CAGCACAACACTGGGAAGTGATTCTTCCCTCAGGCCCCTCTGCCTACCAACACCTGGGCT
3541 3600
CCTCACTGGGGGAAACAAAAGCCTATAAACCCCAGCAACAAAACCTAGTCCTCTTAGACG
3601 3660
TTCTTGCGCTTTGATTTTTTAGGGCGTGTGCCCTGTTACACTTATAGGGCCTAGGATGCT
3661 3720
TGTGTTGAGTAAAAAGGAGATGCCCCAATATTCAAAGCTGCTAAATGTTCTCTTTGCCAT
3721 3780
AAAGACTCCGTGTTAACTGTGTAAACACTTGGGATTTTTCTCCTATGTCCCGAGGTCTGG
3781 3840
TCTTGATTTCTTTTTTGGGTTTCTTTCTAGGAAAATGAGAAGTGCATGCAAGGGGCAGGA
3841 3900
GATGACCCTCCCCTAGGCTTTCAGCTTCAGGCAGCTTCTTCACAGCCTGTTCAGCCTGGG
3901 3960
CTCCTGGAGGACAGCCCTGGGGGAGGCAGTGAGGGGCAGCGCAAGATAGCCAGGTGGTTG
3961 4020
GTTCCAGGACCACAGTGTCTTTTTTTTGTTGTTGGTTTTTTCGTTGTTGTCGTTCGTTTG
4021 4080
TTCGTTTTTAACTGCCACTGCCGCCCCTGACCCCAATCTTGGTCAGCTCTGGAGTACTGC

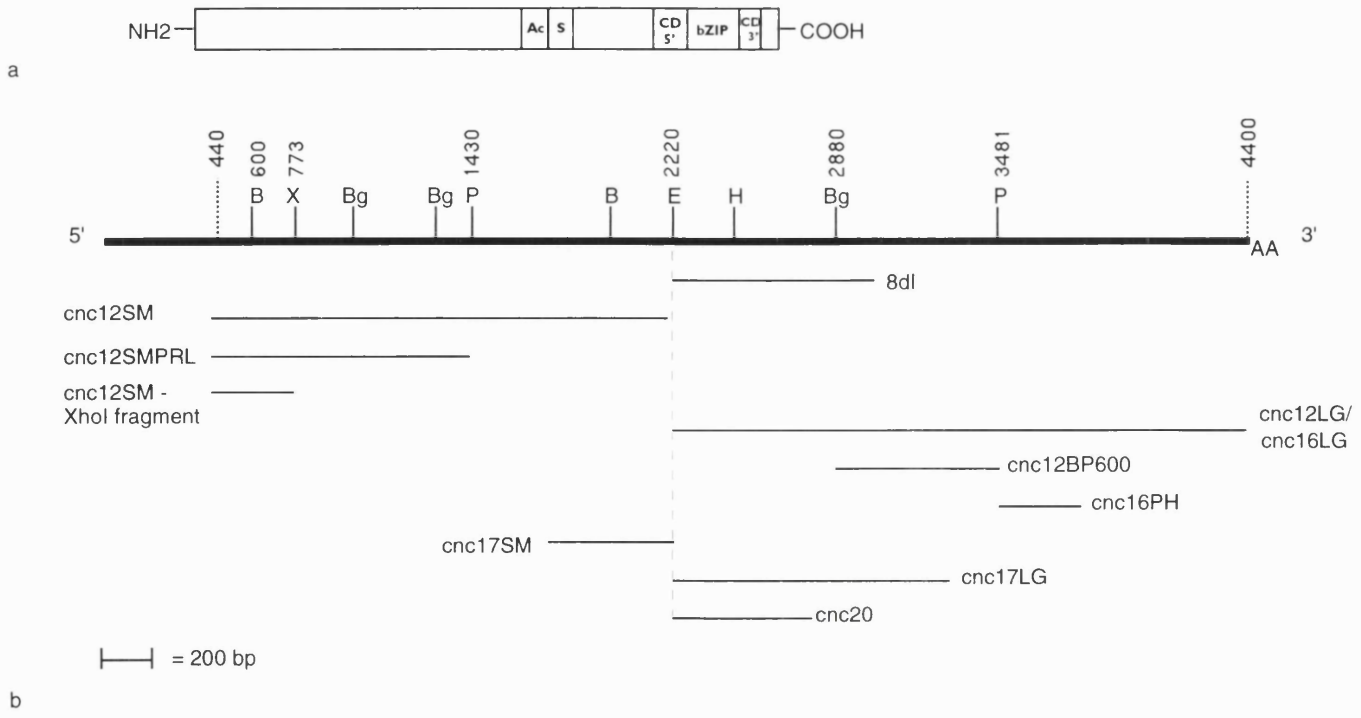


Figure 3.10' ***Schematic of NFE2L1 and the relative position of commonly used DNA probes for Nfe2l1***

(a) Schematic of NFE2L1 and its identified domains (see below). (b) Map of cDNA clones derived from Figure 3.3.The numbering refers to the nucleotide position in the full length sequence of Nfe2l1 (Figure 3.10).
Ac= acidic domain, S= serine repeat region, bZIP= bZIP region, CD5'= CNC5' domain, CD3'= CNC3' domain
B= Bam HI, Bg= Bgl II, E= Eco RI, H= Hind III, P= Pst I, X= Xho I

```
4081  ---------+---------+---------+---------+---------+---------+   4140
      CTGCCCCAGACGAGCAGGGGTTGGGGGGGAGCACTGATCCTCCTCCCTGGGCAGGGCAGA
4141  ---------+---------+---------+---------+---------+---------+   4200
      GGGCTTTCCTAACCGAGCAGTAGGGATAGAAAGCGTGAGCCTGGGAGTGCTTTTTATAAA
4201  ---------+---------+---------+---------+---------+---------+   4260
      TTATTTTCCTTGTAGATTTTATTTTTAATTTATCTCTGTGACCTGCCAGGGAGAGGAGAG
4261  ---------+---------+---------+---------+---------+---------+   4320
      AAAGAAATGCTGTGAGCACATGACAAAATAAAATCAAATAAAATGGATGATTCAGCTTAA
4321  ---------+---------+---------+---------+---------+---------+   4380
      AAAAAAAAAAAAAAAAAAAA
4381  ---------+---------- 4400
```

Figure 3.10 Nucleotide sequence and deduced amino acid sequence of murine Nfe2l1 (GenBank Accession No. x78709)

The basic region is highlighted by a dashed underline, and the leucine/ hydrophobic heptad repeats are outlined. The acidic domain is boxed. Kozak consensus sequences are boldfaced and italicised. S-P-X-X sequences are solid underlined. Potential polyadenylation signals are boldfaced.

There is an initiation codon with an adequate Kozak consensus sequence (Kozak, 1996) at nucleotides 401- 407, with a subsequent open reading frame (ORF) of 741 amino acids. The predicted molecular mass of this ORF is 81.5 kDa: p*I* is predicted to be 7.2. *Nfe2l1* has a bZIP domain towards its C-terminus, consisting of an arginine- and lysine-rich basic domain (amino acids 625 to 650) followed by six heptad repeats of leucine and hydrophobic residues. The predicted protein has several other features compatible with a role as a transcription factor: an acidic domain at amino acids 414-447; six occurrences of the sequence S-P-X-X (relatively over-represented in such proteins, Suzuki *et al.*, 1989); serine/threonine-rich stretches (e.g. amino acids 360-382) and a polyserine tract (immediately C-terminal to the acidic domain). Figure 3.10' depicts the relative position of commonly used *Nfe2l1* cDNA fragments (Figure 3.3) to the derived protein sequence of *Nfe2l1*, and its identified domains.

3.3 MAPPING OF MURINE AND HUMAN *NFE2L1*

3.3.1 Chromosomal localisation of murine *Nfe2l1*

The localisation on the mouse chromosomal map is an important step in a gene's characterisation as it may co-localise with a known mouse mutant or trait. As mentioned previously large numbers of mouse mutants have been mapped and this information is readily available. Also the comparative map between mouse and man is extensive and so a position on a murine chromosome may well have an established homology of synteny grouping on a human chromosome.

The CsCl-purified 2.2 kb cDNA clone, cnc16LG, was sent to a collaborator in Marseille, Dr. M.G. Mattei, who performed radioactive *in situ* hybridisation on metaphase spreads from a WMP male mouse as described in Mattei *et al.* (1985). The results obtained are shown as a schematic in Figure 3.11.

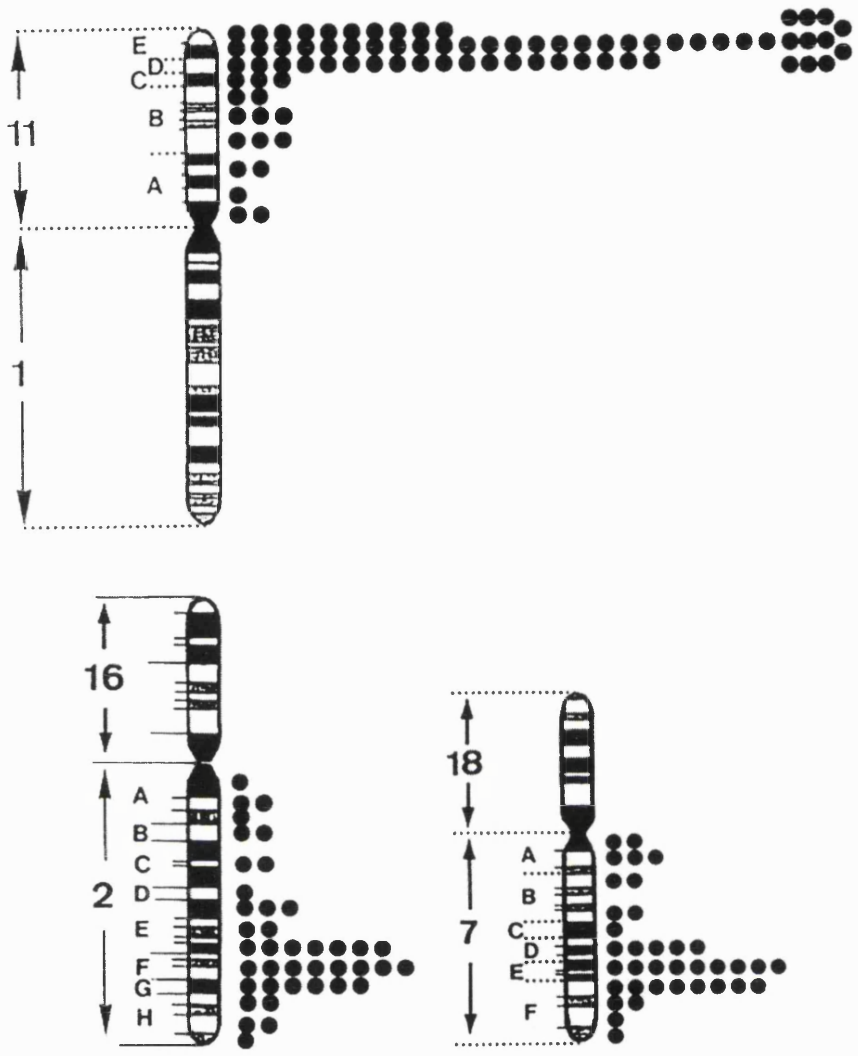


Figure 3.11 Radioactive in situ hybridisation of murine Nfe211
The panel shows the distribution of silver grains detected over different regions of mouse chromosomes 11, 7 and 2. The chromosomes are not normal but are Robertsonian translocations.

In the 200 metaphase cells examined there were 459 silver grains associated with chromosomes, and three hybridisation peaks were detectable. The greatest number of grains was on chromosome 11, with two other peaks of grains on chromosomes 2 and 7. In total 80 grains were located on chromosome 11 (17.4%). The distribution of grains was non-random, with 64/80 (80%) clustered in the 11D-E region. Forty-one grains were located on chromsome 2 (8.9%), with a non random distribution: 21/41 (51%) mapped to the 2E4-G region. The third peak of 36 grains was located on chromosome 7 (7.8%): the cluster of 22 (61%) grains at the 7D1-F1 region was non-random. This data, similar to the Southern results (Figure 3.2d) suggests the

presence within the genome of either pseudogenes or an extended family of '*cnc*-like' mammalian bZIP proteins .

3.3.2 Fine mapping of murine *Nfe2l1* on chromosome 11

To confirm and further refine the map position of *Nfe2l1* on mouse chromosome 11, the Mouse Backcross facility at the UK-HGMP Resource Centre was contacted, which provided murine genomic DNA (*M.spretus* and C57BL/6) for restriction enzyme digestion and Southern blotting (Figure 3.12a): the blot was probed with cnc16LG and cnc12SM. The probe cnc16LG did not show any differences between strains but cnc12SM detected two sets of polymorphisms using *Kpn* I and *Taq* I (Figure 3.12b). The *Taq* I polymorphism generated a C57BL/6-specific allele of 7.0 kb and a *M.spretus*-specific allele of 1.8 kb. The UK-HGMP Resource Centre subsequently provided Southern blots of *Taq* I digested genomic DNA from the 36 progeny of an interspecific backcross (C57BL/6 x *M.spretus*)F_1 x C57BL/6 carried out by the European Collaborative Interspecific Backcross (EUCIB). Hybridisation of the EUCIB blots with cnc12SM suggested that the *M.spretus*-specific allele of 1.8 kb was a doublet. The localisation of *Nfe2l1* was determined by haplotype analysis of the results (Table 3.2). Figure 3.13 is a schematic of chromosome 11 showing *Nfe2l1* relative to other murine markers on the Mbx map (Breen *et al.* 1994; http://www.hgmp.mrc.ac.uk/localdata/mbx/Mbx_Homepage.html).

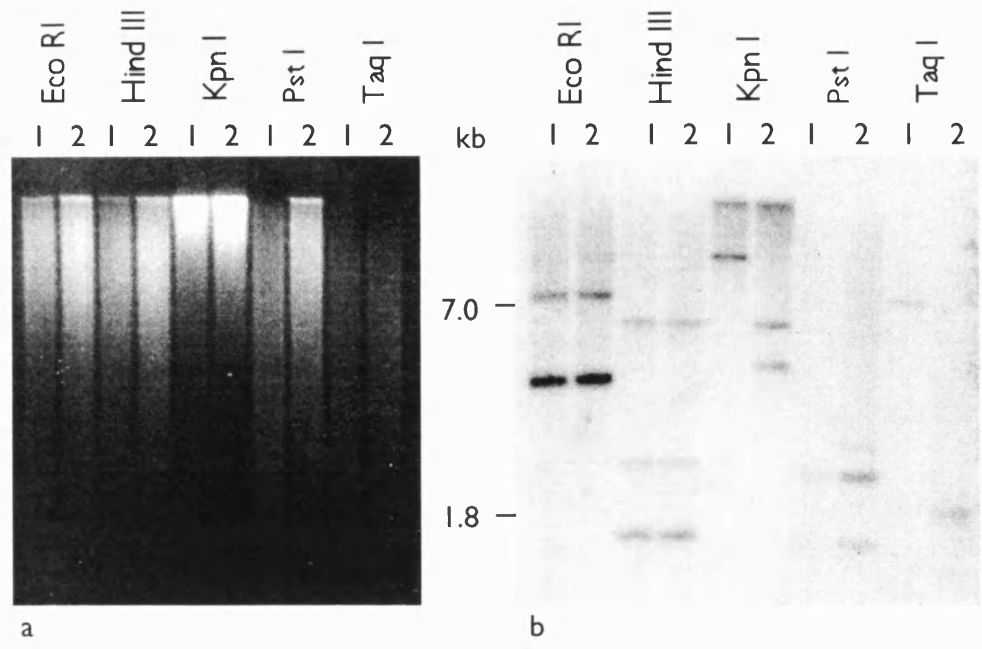


Figure 3.12 Detection of allelic variants of Nfe2l1 in M.spretus and C57BL/6 mouse strains
(a) A Southern blot was prepared, with 5μg of restriction endonuclease-digested genomic DNA in each lane. (b) Polymorphisms were detected for both Kpn I and Taq I digested DNA by hybridisation with cnc12SM. Band sizes for the Taq I polymorphism are labelled. The blot was washed to a stringency of 0.1x SSC, 0.1% SDS at 65°C, and exposed to film for 1 week at -70°C.
Lane 1 = C57BL/6
Lane 2 = M.spretus

Locus 1	Locus 2	Number of recombinants	Genetic distances (cM)
Nfe2l1	*D11Mit10*	4/36	11.11±5.24
Nfe2l1	*D11Mit36*	5/36	13.89±5.76
Nfe2l1	*D11Mit31*	5/36	13.89±5.76
Nfe2l1	*D11Nds19*	8/36	22.22±6.93
Nfe2l1	*D11Mit11*	8/33	24.24±7.46

Table 3.2 Interspecific backcross mapping data for Nfe2l1 provided by UK-HGMP Resource Centre

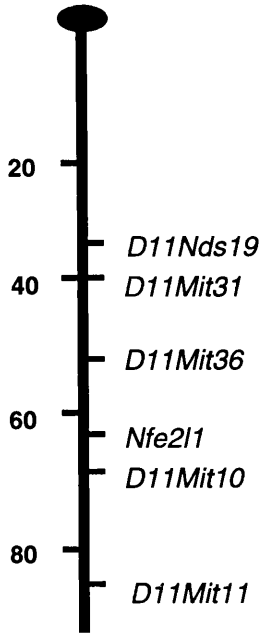


Figure 3.13 The localisation of Nfe2l1 on the Mbx map of Chr 11
Genetic distances are in cM (left)

3.3.3 Chromosomal localisation of human *NFE2L1*

3.3.3.1 Obtaining homologous human genomic probes for murine Nfe2l1

There is a homology of synteny grouping between murine chromosome 11DE region and the long arm of human chromosome 17. An ICRF human chromosome 17-specific cosmid gridded library (Lehrach *et al.*, 1990) was screened with cnc16LG and washed to a stringency of 0.1x SSC, 0.1% SDS (Figure 3.14). Six strong positives (nos. ICRFc105- E028, E09104, H01106, F06143, B11144, C04160) and six weaker positives were requested (nos. ICRFc105- H1039, A04128, F01134, E0311, E0366, F0982), and genomic DNA from these clones was digested with *Hind* III, electrophoresed and blotted (Figure 3.15a). The Southern blot was hybridised with both cnc12SM (Figure 3.15b) and cnc16LG (Figure 3.15c). Two bands were detected by cnc12SM (10 kb and 5 kb) and one by cnc16LG (5 kb) but only in the lanes containing the six strong positives: these overlapped by RE mapping (data not shown) and were extremely likely to contain homologous human genomic DNA for murine *Nfe2l1*. To confirm this suggestion, the 5kb *Hind*III fragment from cosmid no. ICRFc105C04160 that cross-hybridised with cnc16LG and cnc12SM was used as probe on restriction endonuclease digests of mouse and human genomic DNA (Figure 3.16). As it detected a subset of those fragments seen on hybridisation with murine

Nfe2l1 (Figure 3.16a and Figure 3.2a) it was reasonable to assume that the six strong positive clones contained sequences from the human homologue of *Nfe2l1* and consequently to use them as probes in fluorescent *in situ* hybridisation (FISH) mapping of human *NFE2L1*.

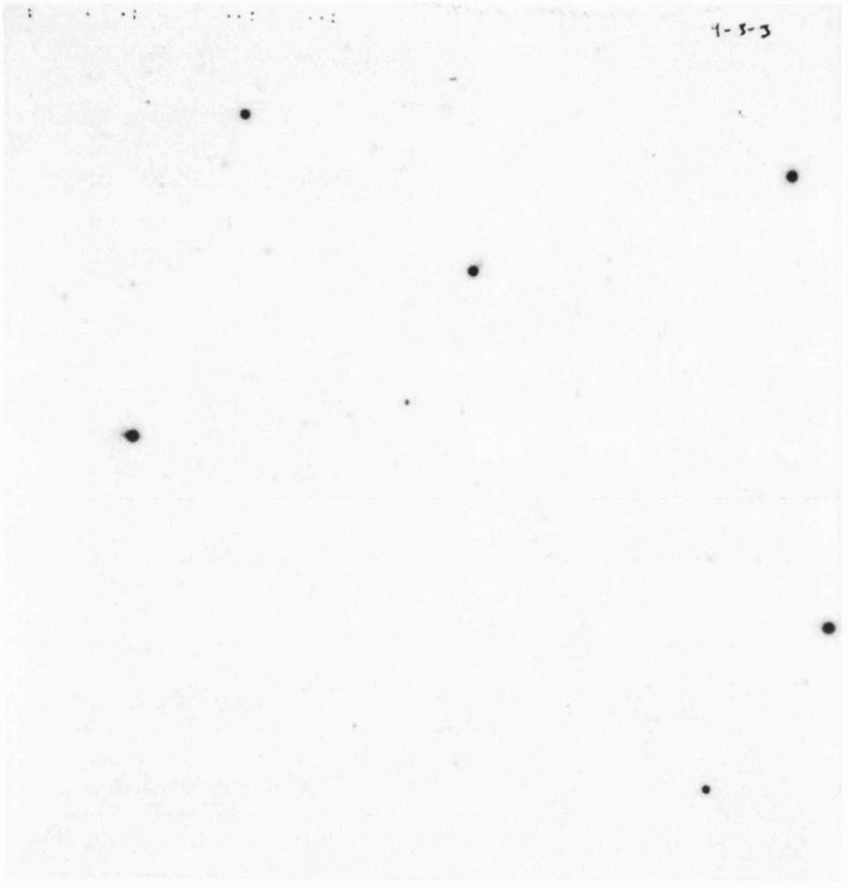


Figure 3.14 Chromosome 17-specific cosmid library no. 105 (L4/FS17) screened with cnc16LG

The gridded cosmid library was represented on a single duplicated filter. Twelve replicated positives were requested from the ICRF Reference Library Database, London. The filters were washed to a stringency of 0.1x SSC, 0.1% SDS at 65°C, and exposed to film for 5 hours at room temperature.

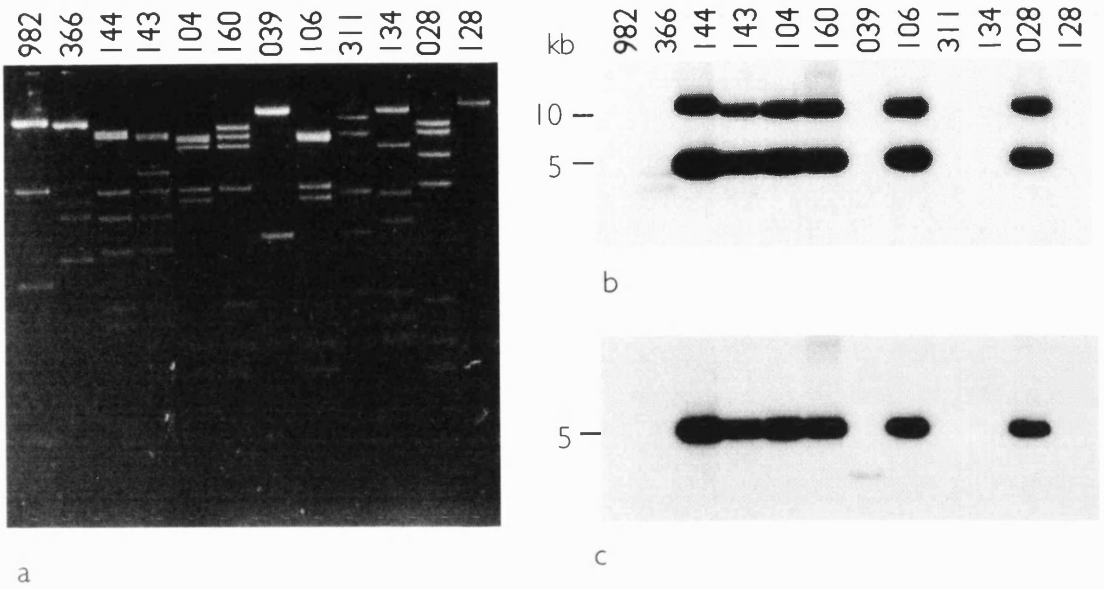


Figure 3.15 Restriction endonuclease digest and Southern analysis of chromosome 17-specific library clones

(a) A Southern blot was prepared with HindIII digested cosmid DNA (individual lanes are marked using the last three digits of the ICRF clone no.). It was hybridised with both cnc12SM (b) and cnc16LG (c), washed to a stringency of 0.1x SSC, 0.1% SDS at 65°C and exposed to film overnight at -70°C. The bands detected by cnc12SM (10kb and 5kb) are the same size as those detected by the same probe on Hind III-digested human genomic DNA (Figure 3.16).

No bands were detected in the murine genomic lanes by the 5kb human cosmid fragment (Figure 3.16b and c), and this could be due to lack of cross-species conservation. Subsequent analysis of the genomic organisation of the *NFE2l1* gene using the deposited genomic sequence from PAC clone no. HRPC890E16 (GenBank accession no. AC004477) revealed that the human cosmid band sizes detected by cnc12SM and cnc16LG were of an appropriate size (data not shown). Sequence analysis also revealed that the 5kb human cosmid fragment used in the Southern analysis contained the majority of the coding sequence (nucleotides 1130 to 4440; data not shown). The coding sequence of *Nfe2l1* is relatively well conserved between man and mouse and the explanation for the absence of murine genomic bands might be a combination of the following reasons: too short an exposure time for the autoradiograph; poor quality of Southern blot; and cross-species differences in sequence. This might be resolved by repeating the Southern analysis.

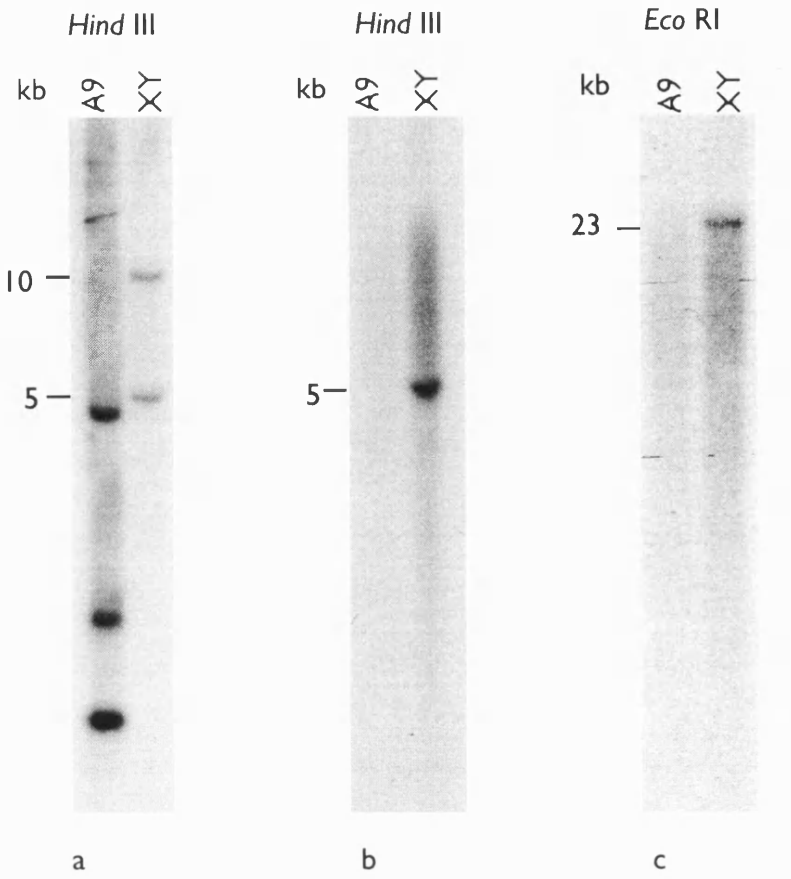


Figure 3.16 Genomic bands detected by 5kb Hind III cosmid fragment
Southern blots were prepared from RE digested mouse (A9) and human (XY) cell line DNA. (a) Cnc12SM detected 2 bands, 10 kb and 5 kb, in the human lane as expected from the cosmid DNA Southern analysis (Figure 3.15). The 5 kb Hind III cosmid DNA fragment was used as a probe: clear signal was only obtained in the XY genomic DNA lane but gave the expected size of bands in both Hind III- (b) and Eco RI- (c) digested Southern blots. All blots were washed to a stringency of 0.1x SSC, 0.1% SDS at 65 °C, and exposed to film for (a) 5 days or overnight (b and c).

3.3.3.2 Fluorescent in situ hybridisation mapping of human NFE2L1

Two of the human cosmids (nos. ICRFc105- C04160 and F06143) were used as probes for FISH analysis on normal XY metaphase spreads as described in section 2.2.22. The human chromosome detected by the *NFE2L1*-containing cosmid was identified using a chromosome 17 specific α-satellite probe (Oncor). Several different metaphase spreads were examined to confirm the localisation and a representative result is shown in Figure 3.17.

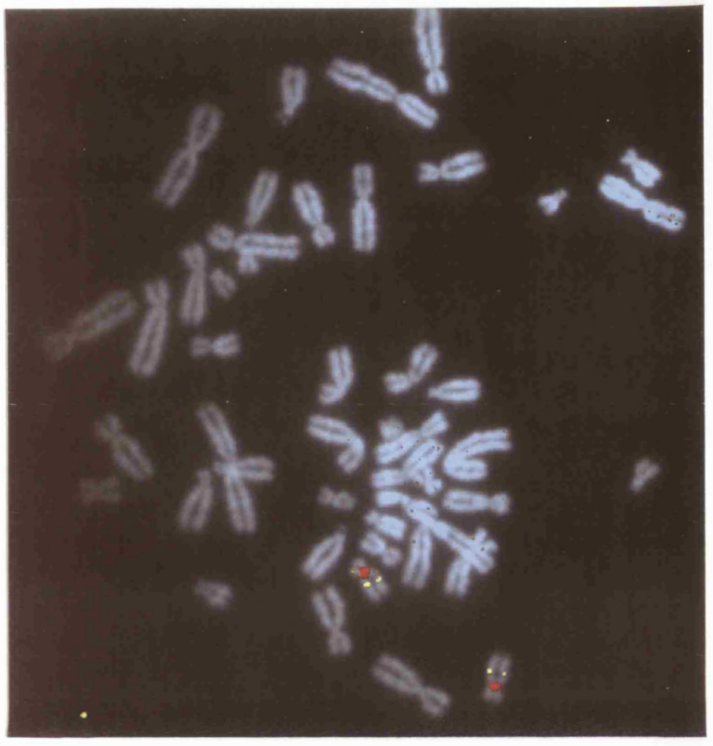


Figure 3.17 FISH mapping of human NFE2L1
The human cosmid DNA (yellow signal) is in the middle of the long arm of chromosome 17, identified by the D17Z1 α-satellite probe (red signal).

3.4 EXPRESSION ANALYSIS OF MURINE *Nfe2l1*

This section is in two parts: firstly, northern analysis is used to determine the transcript size of *Nfe2l1*, presence of alternative transcripts and some gross information on sites of *Nfe2l1* expression. Next, *in situ* hybridisation is used to ascertain finer detail of the spatial and temporal expression pattern of *Nfe2l1* mRNA.

3.4.1 Northern analysis

Initial northern blots were made using total RNA from two cell lines: A9, a mouse fibroblast cell line; and WA17, a somatic cell hybrid cell line containing three copies of human chromosome 21. All of the *Nfe2l1* cDNA subclones were tested and each gave a similar pattern (Figure 3.18). Below the smear of 28S, two distinct bands can be seen in both lanes. To confirm that the two transcripts detected by northern hybridisation were genuine and not non-specific, the same experiment was performed using poly(A)$^+$ RNA (Figure 3.19). The total RNA lane for A9 is under-loaded compared with the other lanes as estimated visually using β-actin (Figure 3.19c), however both bands detected by *Nfe2l1* are present in the total RNA and poly(A)$^+$ RNA lanes.

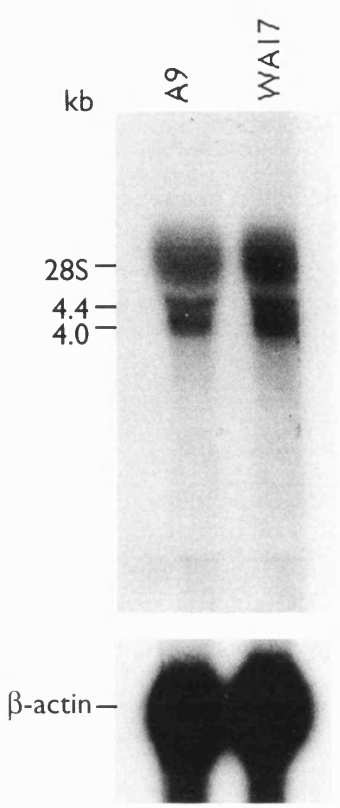


Figure 3.18 Northern analysis of Nfe2l1 against cell line total RNA
A northern blot was prepared from 10μg each of total RNA from the cell lines A9 (mouse) and WA17(human). The blot was hybridised with a DNA probe, cnc16LG, washed to a stringency of 1x SSC, 0.1% SDS at 55°C, and exposed to film for 1 week at room temperature. All Nfe2l1 cDNAs gave a similar pattern result: two distinct bands can be seen (4.4 kb and 4.0 kb) below the smear of 28S. β-actin is shown as a loading control.

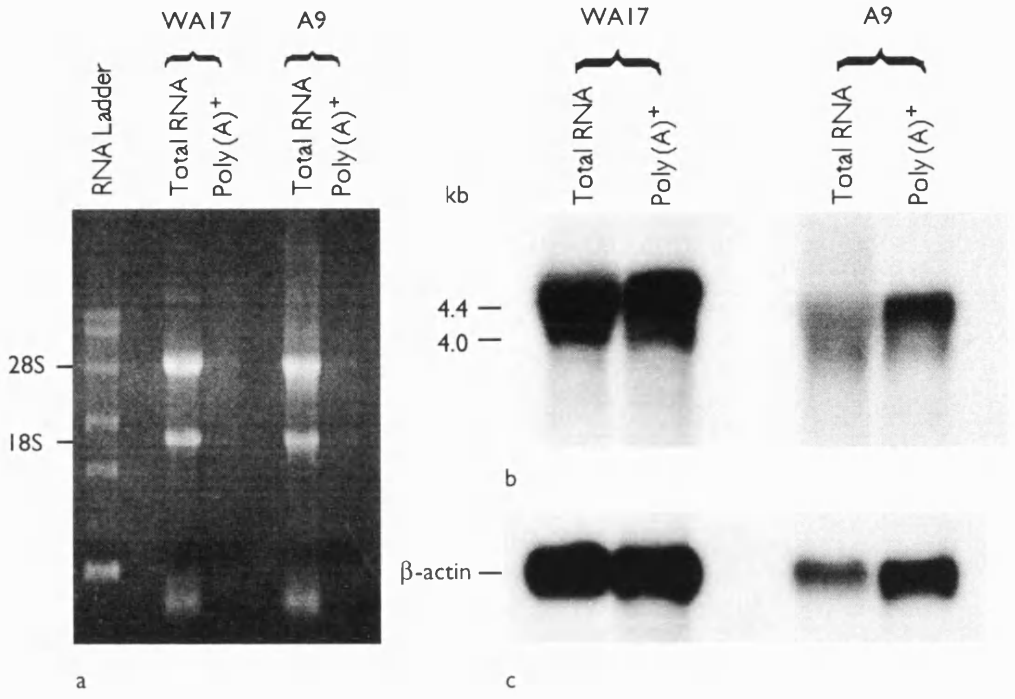


Figure 3.19 Total RNA versus poly(A)⁺ RNA

(a) A northern blot was prepared from 10μg of total RNA and 1μg of poly(A)⁺ RNA. (b) It was hybridised with an antisense cnc17LG riboprobe, washed to a stringency of 0.1x SSC, 0.1% SDS at 80°C and exposed to film overnight at -70°C. (c) β-actin is shown as a loading control.

The expression of *Nfe2l1* was determined in a range of adult mouse tissues and staged mouse embryos (Figure 3.20). The antisense cnc12SM-PRL riboprobe (5' coding; nuc $_{440-1430}$) detected a predominant band at 4.4 kb in all lanes, with a smaller band (4.0 kb) also being seen in a few other lanes. An additional 3.6 kb band was also detected in adult mouse brain: this could be accounted for by alternative splicing or polyadenylation, or by cross-hybridisation with *Nfe2l1*-related genes. For comparison, a Clontech adult human brain northern was hybridised with cnc12SM (nuc $_{440-2221}$) (Figure 3.21). Only two signals were detected (4.8 kb and 4.4 kb), suggesting that the 3.6 kb transcript in adult mouse brain might be murine-specific or that it is expressed in human brain at levels not detectable by northern analysis. The larger band sizes found in humans are accounted for by a longer 5' UTR region (Luna *et al.*, 1994). The antisense cnc16LG-PH (3' UTR; nuc $_{3481-3791}$) riboprobe gave similar results to those obtained with cnc12SM-PRL (data not shown).

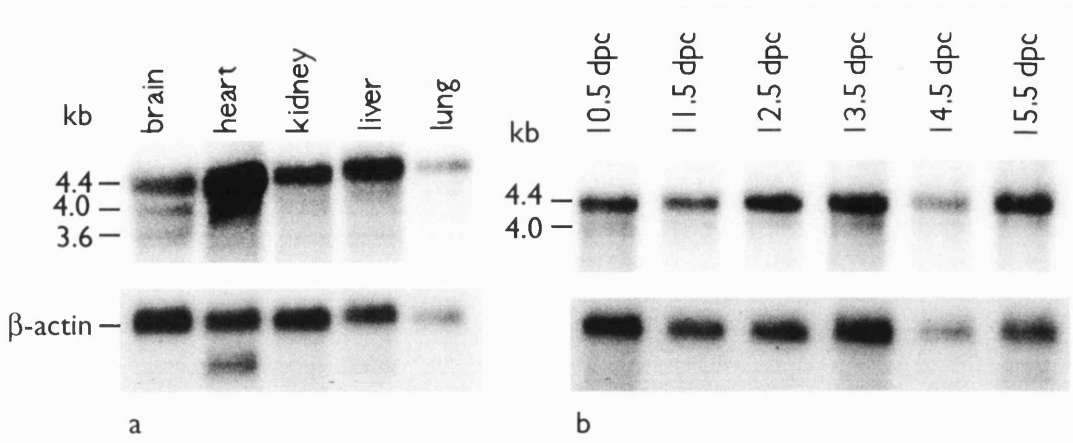


Figure 3.20 Nfe2l1 expression in adult mouse tissues and staged mouse embryos
Northern blots were prepared with ≈1-2μg of poly(A)$^+$ RNA of each tissue or stage shown, and hybridised with an antisense cnc12SM-PRL (nuc $_{440-1430}$) riboprobe: the filters were washed to a stringency of 0.1x SSC, 0.1% SDS at 80°C and exposed to film for 7 hours at -70°C. β-actin. is shown as a loading control.

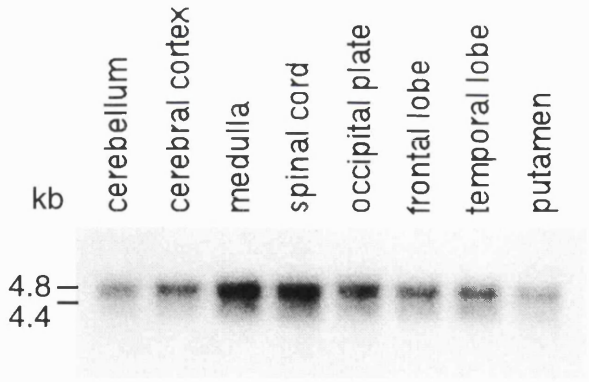


Figure 3.21 Nfe2l1 expression in adult human brain
A Clontech adult human brain northern was hybridised with cnc12SM (nuc $_{440-2221}$), according to the manufacturer's instructions, washed to a final stringency of 0.1x SSC, 0.1% SDS at 68°C, and exposed to film for 6 hours at -70°C.

The radioactive signals (Figure 3.20) were quantified using a phosphorimager (Molecular Dynamics), with β-actin as a loading control (Tables 3.3, 3.4). *Nfe2l1* signals are particularly strong in adult mouse heart: the signal is approximately three and a half times as great as that observed in the brain (Table 3.3). No large changes in expression level are seen in whole mouse embryos (10.5- 15.5 dpc), with only a steady increase with developmental stage (Table 3.4): the level of *Nfe2l1* expression in

the 15.5 dpc mouse embryo is approximately twice that detected in the 10.5 dpc mouse embryo.

	brain	heart	kidney	liver	lung
cnc12SM-PRL	217708	764368	205229	293069	56763
β-actin	145461	154535	164660	117005	42178
RV	1.11	3.69	0.92	1.86	1

Table 3.3 Phosphorimager quantification of hybridisation signal in Figure 3.20
The figures have been adjusted for background levels and relative values (RV) calculated.

	10.5	11.5	12.5	13.5	14.5	15.5
cnc12SM-PRL	131261	97524	169230	187635	71472	173837
β-actin	323585	217570	270654	378043	106899	239286
RV	0.6	0.66	0.94	0.74	1	1.08

Table 3.4 Phosphorimager quantification of hybridisation signal in Figure 3.21
The figures have been adjusted for background levels and relative values (RV) calculated.

3.4.2 *In situ* hybridisation analysis

3.4.2.1 Whole-mount in situ hybridisation

Two different fragments of *Nfe2l1* were used as templates for synthesis of riboprobes (both as sense and antisense probes) to affirm the specificity of the expression pattern: cnc12SMPRL (5' coding, nuc $_{440-1430}$) and cnc16PH (3' UTR, nuc $_{3481-3794}$). The whole-mount *in situ* hybridisation was performed on 7.5- 10 dpc mouse embryos as described (section 2.2.23.9) with on average 4-7 embryos/probe. An antisense riboprobe for *Krox 20* was used as a positive control (Figure 3.22): at this stage it is expressed in rhombomere 5 (Wilkinson *et al.*, 1989a; Sham *et al.*, 1993). The probe cnc12SMPRL was used for whole-mount *in situ* hybridisation on 7.5 dpc and 8.5 dpc mouse embryos. There is widespread low-level expression in 7.5 dpc mouse embryos (Figure 3.23) and some signal seems to be localised, but because of distortion of the embryos it is difficult to name the regions. In the 8.5 dpc mouse embryos a strong ubiquitous signal was seen (data not shown).

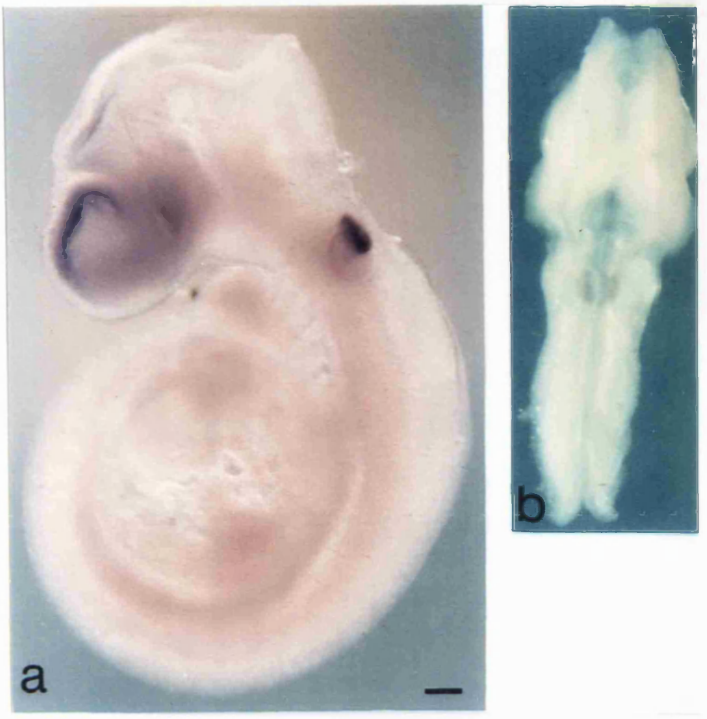
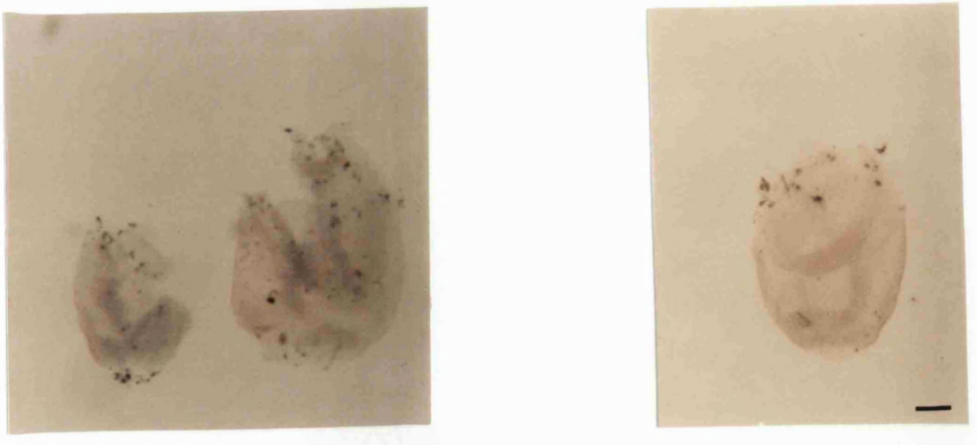


Figure 3.22 Whole-mount in situ hybridisation using Krox 20 riboprobe
(a) Lateral view of 9.5 dpc mouse embryo hybridised with antisense Krox 20 riboprobe: signal is clearly localised to rhombomere 5.(b) Dorsal view of 9.5 dpc mouse embryo hybridised with antisense Krox 20 riboprobe using the NY method: signal is present as expected but is weaker. Bar = 178μm (a), 238μm (b).



a b

Figure 3.23 Whole-mount in situ hybridisation on 7.5 dpc mouse embryos
(a) Embryos hybridised with antisense cnc12SMPRL. (b) Control embryo hybridised with sense cnc 12SMPRL.
Bar = 180μm.

Whole-mount *in situ* hybridisations were undertaken on 9.5- 10 dpc mouse embryos using cnc12SMPRL (Figure 3.24) and cnc16PH (Figure 3.25). The expression pattern seen with each probe was very similar, although signal obtained using cnc16PH was both weaker and slightly less consistent. *Nfe2l1* was expressed in most regions of the embryo, but it was absent or expressed at a relatively low level in the developing brain, neural folds and neural tube (Figure 3.24e). Expression appeared strongest in the first and second branchial arches (Figure 3.24d). To confirm these findings embryos hybridised with antisense cnc12SMPRL were sectioned post-staining, mounted and examined (Figure 3.26). *Nfe2l1* expression is seen throughout the embryo including the developing nervous system and heart, and is strongest in the first branchial arch (Figure 3.26a) and in the dermomyotome component of the somites (Figure 3.26b and c).

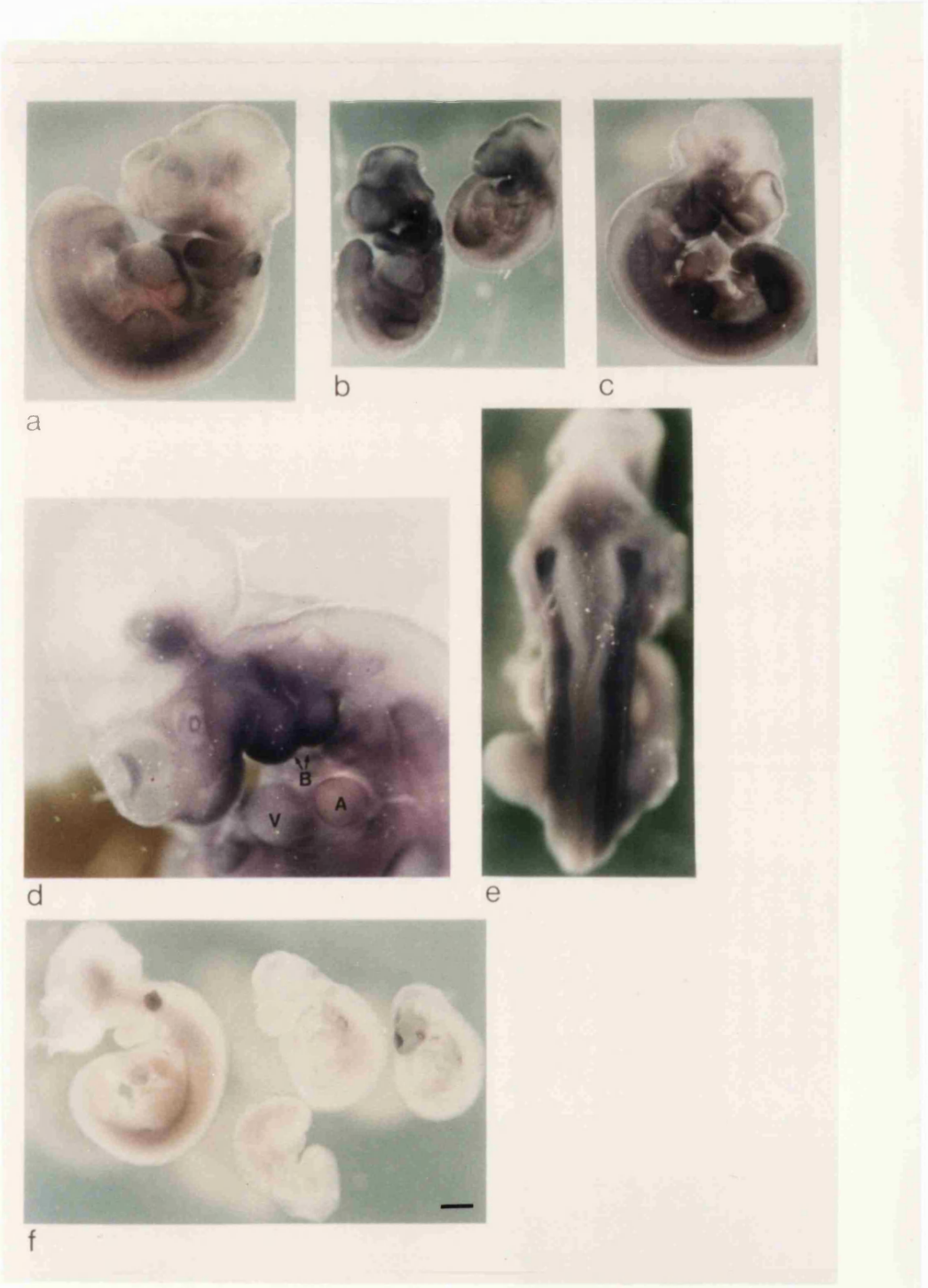


Figure 3.24 Whole-mount in situ hybridisation using cnc12SMPRL riboprobe on 9.5- 10 dpc mouse embryos

(a)-(c) Lateral view of 9.5-10dpc mouse embryos hybridised with antisense cnc12SMPRL. (d) Lateral view of 10 dpc mouse embryo in (c) at higher magnification. (e) Dorsal view. (f) Control embryos hybridised with sense cnc12SMPRL. Bar = 467μm (a), 541μm (b), 656μm (c,f), 327μm (d), 409μm (e).

B= branchial arch

A= developing atrium

V= primitive ventricle

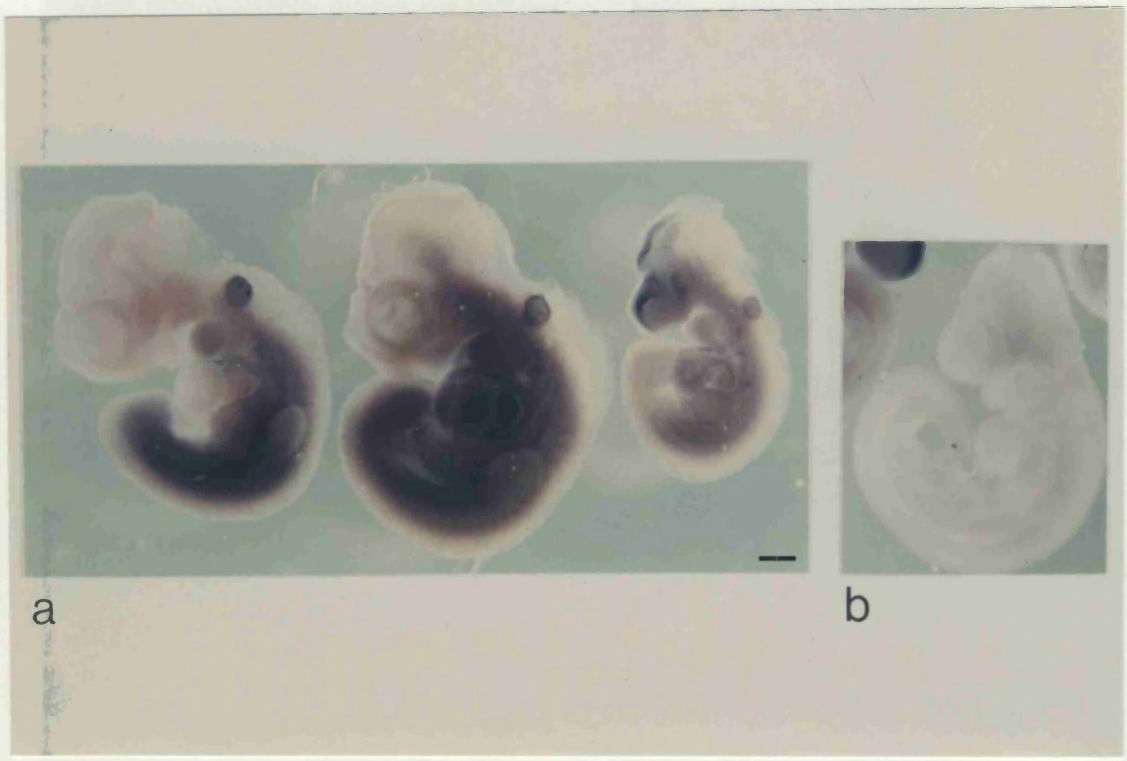


Figure 3.25 Whole-mount in situ hybridisation using cnc16PH riboprobe on 9.5 - 10 dpc mouse embryos

(a) Lateral view of 9.5- 10.5 dpc embryos hybridised with antisense cnc16PH. (b) Control embryo hybridised with sense cnc16PH. Bar = 390μm (a), 500μm (b).

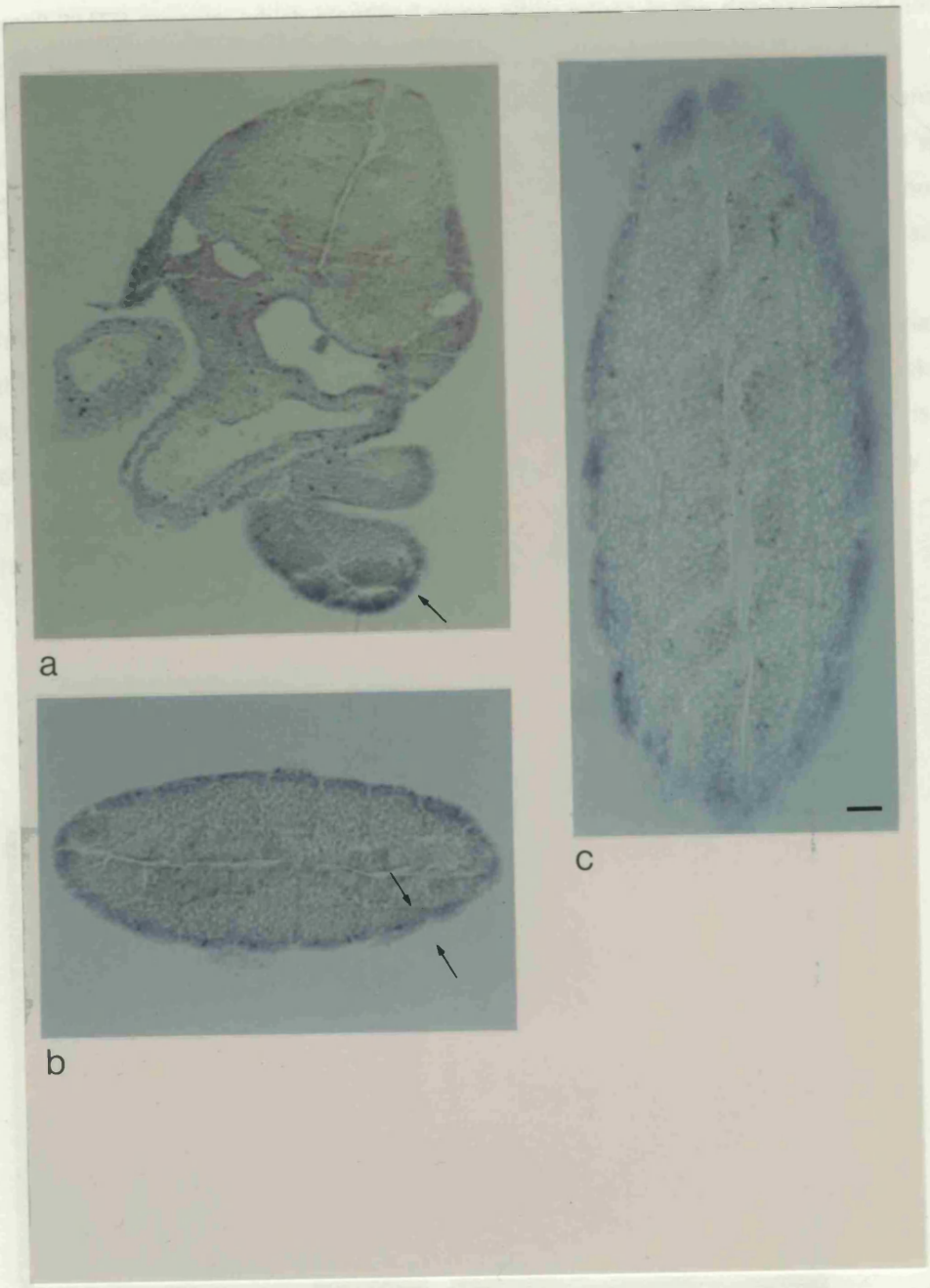


Figure 3.26 Transverse sections of an 9.5 dpc mouse embryo after whole-mount in situ hybridisation using antisense cnc12SMPRL riboprobe

(a) Transverse section at level of branchial arches (arrowed). (b) Caudal transverse section through spinal cord. (c) Similar section to (b) at higher magnification. Dermomyotome is arrowed. Bar = 80μm (a), 62μm (b), 31μm (c).

Database comparisons of the cnc16PH sequence latterly revealed the presence of two small 3'UTR repeats, which identified some other genes in the database and therefore cast doubt on the specificity of the riboprobe. A new *Nfe2l1* 3'UTR fragment cnc12BP600 (nuc $_{2880-3481}$) was used as a template for riboprobe synthesis, and the whole-mount *in situ* hybridisation experiment was repeated. The method used was different to earlier experiments and is referred to as the New York (NY) method (section 2.2.23.9.VI): this protocol is expected to generate lower background and can be used with very large riboprobes (up to 3 kb). The whole-mount *in situ* hybridisation results using the NY method are shown in Figure 3.27: the signal from the positive control, Krox20 (Figure 3.22b), and the antisense riboprobe took much longer to appear and were overall much weaker than that seen previously. It is possible that the conditions used in the NY method were more stringent than previously used. The experiment was only carried out once so further whole-mount *in situ* hybridisations should be undertaken to confirm the results.

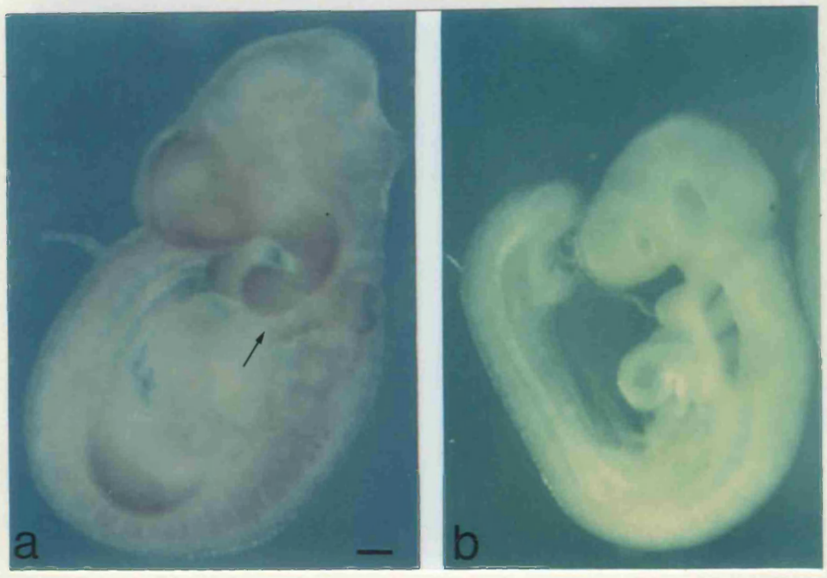


Figure 3.27 Whole-mount in situ hybridisation using NY method
(a) Lateral view of 9.5 dpc mouse embryo hybridised with antisense cnc12BP600. Branchial arches are arrowed. (b) Lateral view of control embryo hybridised with sense 12BP600. Bar = 238μm (a), 312μm (b).

3.4.2.2 *Tissue section in situ hybridisation*

Greater detail of *Nfe2l1* expression, both spatial and temporal, was obtained by tissue section *in situ* hybridisation experiments on a variety of mouse embryonic developmental stages. As for the whole mount *in situ* hybridisations, two fragments of *Nfe2l1* were used as templates for riboprobe synthesis (cnc12SMPRL and cnc16PH), and both gave similar results. Before dipping in emulsion, the slides were exposed to film to give a rough indication of signal strength and background: a selection of the sectioned staged embryos which were typical of the *Nfe2l1* expression pattern are presented (Figures 3.28- 3.29).

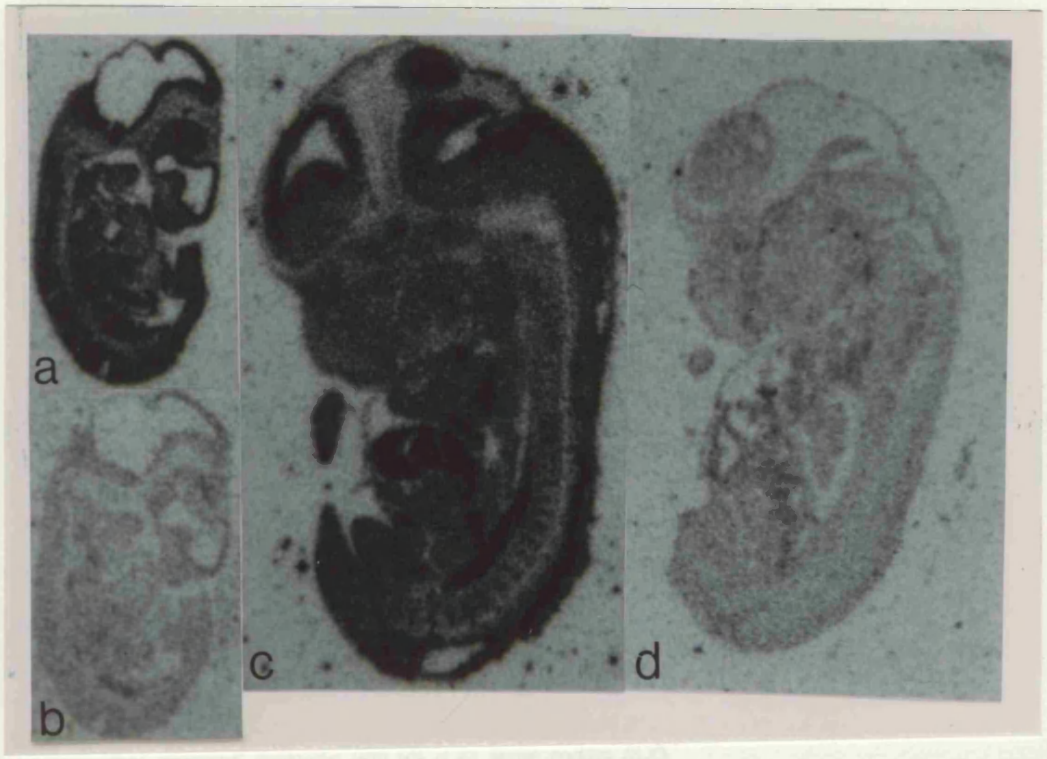


Figure 3.28 Autoradiographs of tissue section in situ hybridisation using 35^S-labelled riboprobes for Nfe2l1

All sections are hybridised with cnc12SMPRL. Sagittal sections of 11.5 dpc mouse embryo hybridised with antisense(a) and sense (b) riboprobe (12.5X). Sagittal sections of 13.5 dpc mouse embryo hybridised with antisense(c) and sense (d) riboprobe (12.5X).

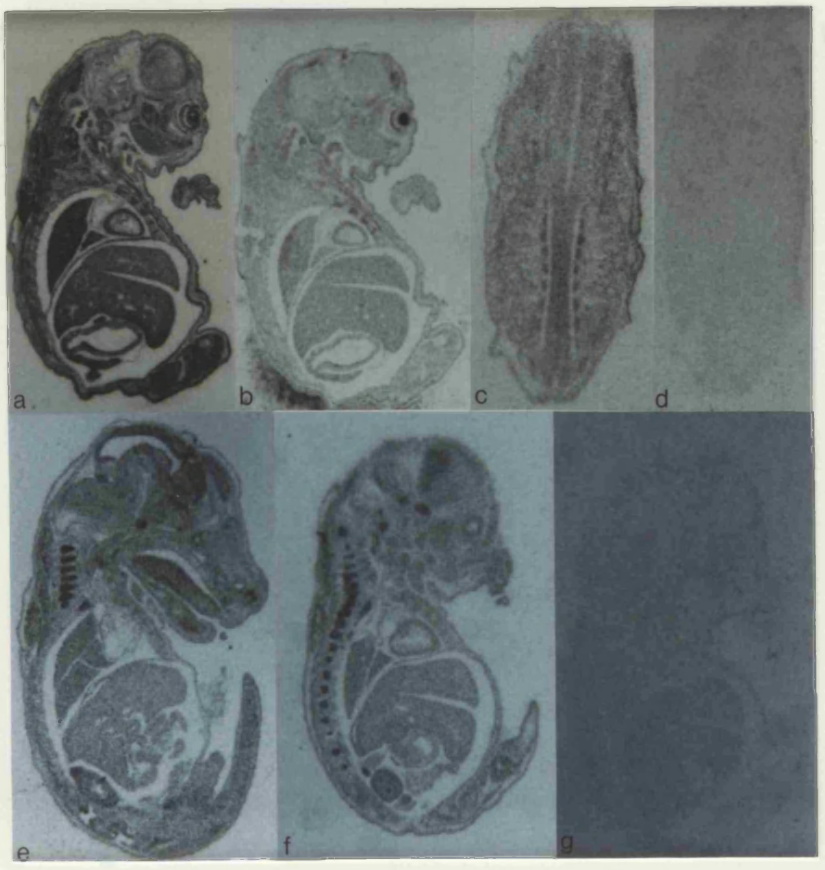


Figure 3.29 Autoradiographs of tissue section in situ hybridisation using 35^S-labelled riboprobes for Nfe2l1

Sections of 16.5 dpc mouse embryo hybridised with cnc12SMPRL. (a), (e) and (f) are sagittal sections hybridised with antisense riboprobe, (b) and (g) are sense controls (all at 4.5X). (c) is a coronal section hybridised with antisense riboprobe and (d) is its sense control (6X).

In the different stages examined (10.5- 16.5 dpc), *Nfe2l1* appeared ubiquitous, although not uniform in its distribution. Tissue sections of 10.5 dpc mouse embryos confirmed the findings of the whole-mount *in situ* hybridisation experiments described above (data not shown). From 11.5 dpc onwards, the expression is stronger in the

developing nervous system than in earlier stages. There appears to be some specific sites of developmental upregulation by 16.5 dpc: examples of sites of the non-uniform distribution of silver grains in the 16.5 dpc mouse embryo are shown (Figures 3.30). The dorsal and ventral grey horns of the spinal cord and the flanking dorsal root ganglia (DRG) are strongly positive for *Nfe2l1* expression (Figure 3.30a, c). A parasagittal section through the DRG and ribs shows clearly the high level of *Nfe2l1* expression in the DRG (Figure 3.30c). The inner nuclear layer of the retina is strongly positive as compared to the outer nuclear layer of the retina, which is also positive but to a lesser degree (Figure 3.30e). There is also strong expression in the heart (Figure 3.30g).

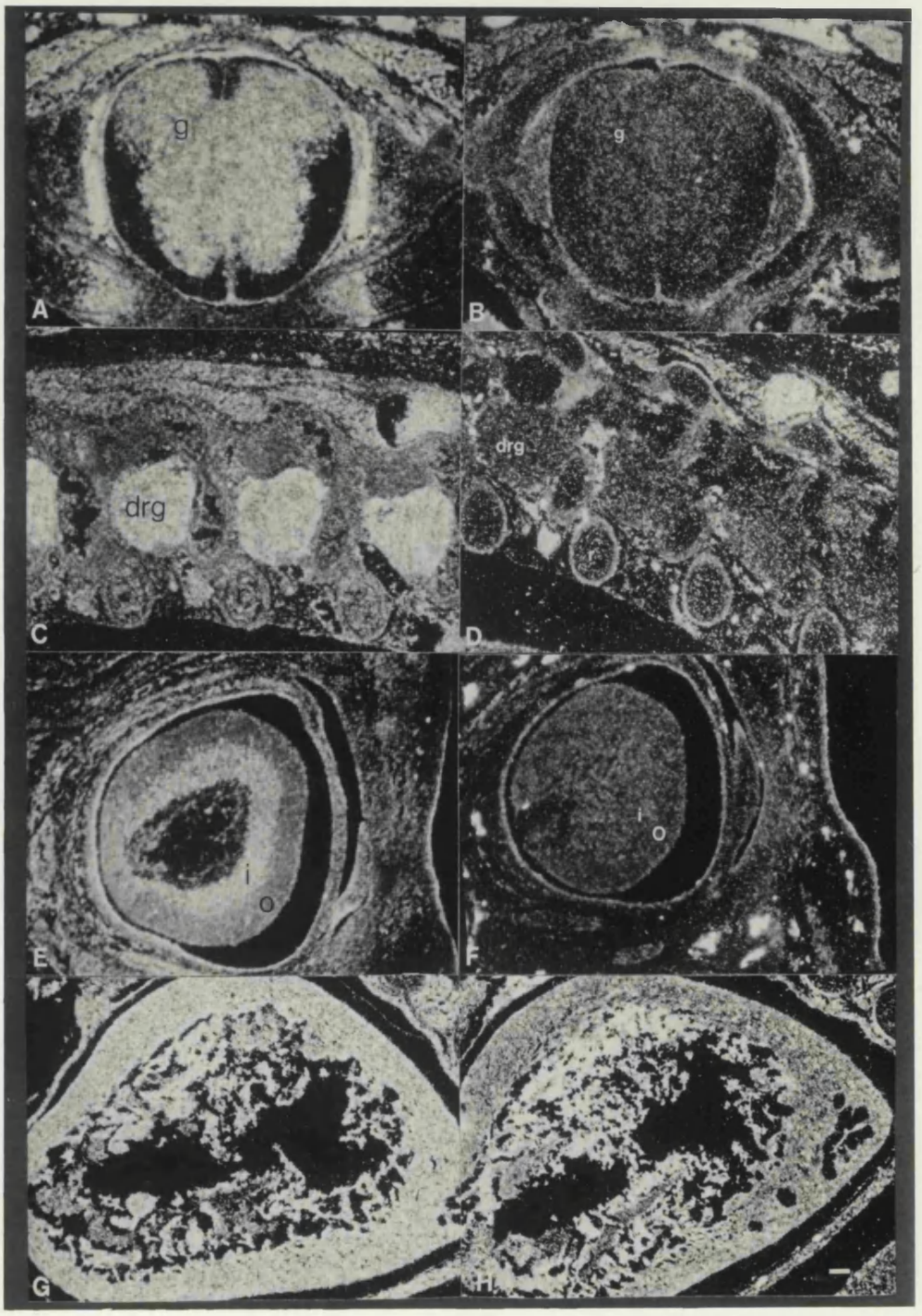


Figure 3.30 Darkfield photography of tissue section in situ hybridisation on 16.5 dpc mouse embryo

Transverse and sagittal sections of a 16.5 dpc mouse embryo were hybridised with antisense (a, c, e, g) and sense (b, d, f, h) cnc12SMPRL riboprobes. The areas selected showed a non-uniform distribution of silver grains as compared to the negative controls: transverse section of spinal cord (a, b); sagittal section through DRG and ribs (c,d); parasagittal section through developing eye (e, f). A sagittal section through heart is also shown (g, h). g= grey matter, drg= dorsal root ganglia, r= rib i=inner nuclear layer of the retina, o=outer nuclear layer of the retina. Bar = 70μm.

3.5 ISOLATION OF OTHER FAMILY MEMBERS

3.5.1 Degenerate PCR

As mentioned earlier in section 1.2.1.6, cross-species related genes can be cloned by using degenerate oligonucleotide primers in a PCR experiment and the same approach can be taken to obtain novel family members within the same species. Two *Nfe2l1*-related genes were cloned whilst this work was ongoing, and will be discussed later (section 4.3): by comparing the homologous regions of the proteins murine NFE2L1, murine NFE2 (Andrews *et al.*, 1993a), human NRF2 (Moi *et al.*, 1994a) and CNC, a degenerate set of primers was designed (Figure 3.31; Table 2.1). The published information on the genomic structure of *NFE2l1* (Luna *et al.*, 1995; Figures 4.3, 4.3') showed that the primers were within a single exon, so human genomic XY cell line DNA was used successfully as template. The primers FAM1A, FAM1B and FAM2 were used as described (section 2.2.15) to generate product at different annealing temperatures (45°C to 51°C) and magnesium concentrations (Figure 3.32). The subcloned products (section 2.2.15.2) were sequenced and analysed by using the BLAST programme (UK-HGMP Resource Centre), and searching the GenBank and SwissProt databases. Out of a total of 60 sequenced clones, 18 clones were *Nfe2l1*, 16 clones were *p45-Nfe2*, 16 clones were *Nrf2* and 10 clones were either unknown or the sequence was of poor quality.

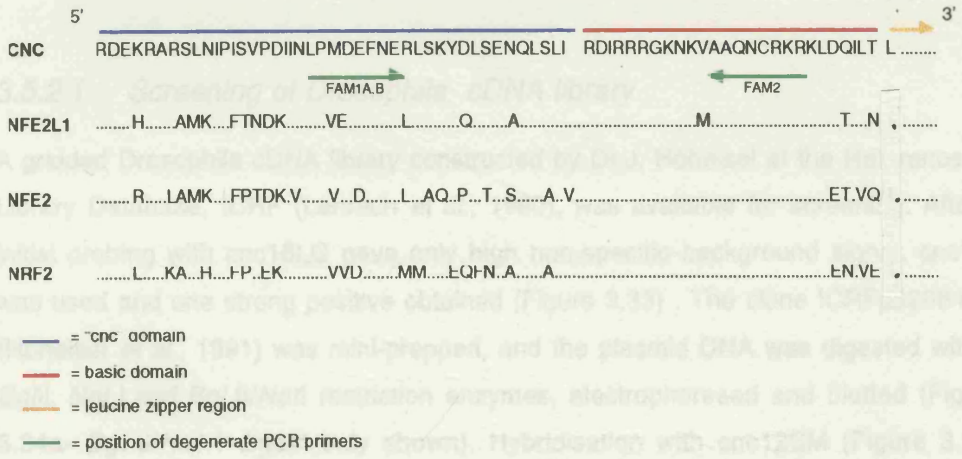


Figure 3.31 Design of degenerate primers by comparison of known Nfe2l1-related family members

The region of homology stretches both 5' and 3' of the bZIP motif: the greatest similarity is seen in the basic DNA-binding domain and the region 5'. Identity is shown as a blank. The primers FAM1A,B and FAM2 should generate a PCR product of 126 bp.

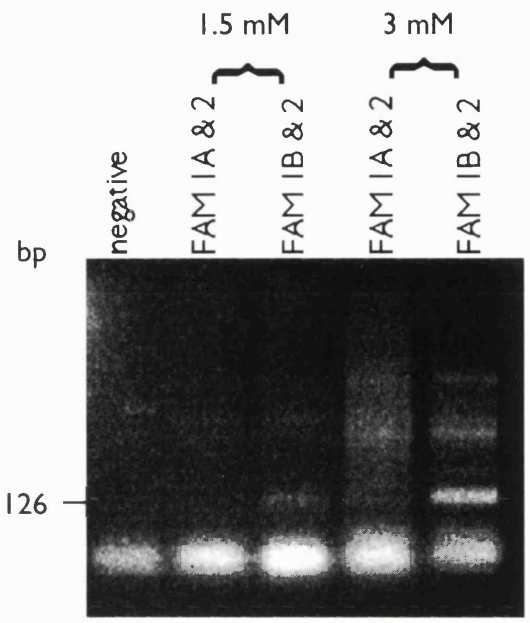


Figure 3.32 Agarose gel of degenerate PCR products
An example of the variation in PCR conditions is shown above: PCR was carried out using either FAM1A/FAM2 primers or FAM1B/FAM2 primers at 51°C annealing temperature, at variable concentrations of magnesium in the buffer (1.5mM or 3mM, final concentration), using mouse genomic DNA as template. The expected product should be 126 bp, and is present in the three right hand lanes. The products were electrophoresed on a 3% agarose gel.

3.5.2 The search for a *Drosophila* homologue of *Nfe2l1*

3.5.2.1 Screening of Drosophila cDNA library

A gridded Drosophila cDNA library constructed by Dr.J. Höheisel at the Reference Library Database, ICRF (Lehrach *et al.*, 1990), was available for screening. After initial probing with cnc16LG gave only high non-specific background signal, cnc12SM was used and one strong positive obtained (Figure 3.33) . The clone ICRFp520B143 (Höheisel *et al.*, 1991) was mini-prepped, and the plasmid DNA was digested with *Bgl*II, *Not* I and *Bgl* II/*Not*I restriction enzymes, electrophoresed and blotted (Figure 3.34a- *Bgl* II/*Not* I digest only shown). Hybridisation with cnc12SM (Figure 3.34b) localised the signal to a 0.6 kb *Bgl*II/*Not*I band (lab name DBN0.6) which was subcloned into *Bam* HI/*Not* I digested pBS SK$^+$vector for sequencing. The entire fragment (470 bp) was manually sequenced, then analysed using the GCG and BLAST programs at the UK-HGMP Resource Centre. The sequence was identical to that of the *Drosophila* gene *Troponin-T* (GenBank accession no. x54504) as illustrated by

BESTFIT analysis (Figure 3.35). Comparison of *Nfe2l1* (EMBL accession no. x78709) and *Drosophila Troponin-T* DNA sequences identified a region of similarity (Figure 3.36), which was presumed to be responsible for the strong hybridisation signal obtained: this region encodes in *Nfe2l1* a serine repeat and in *Troponin-T*, a glutamic acid repeat.

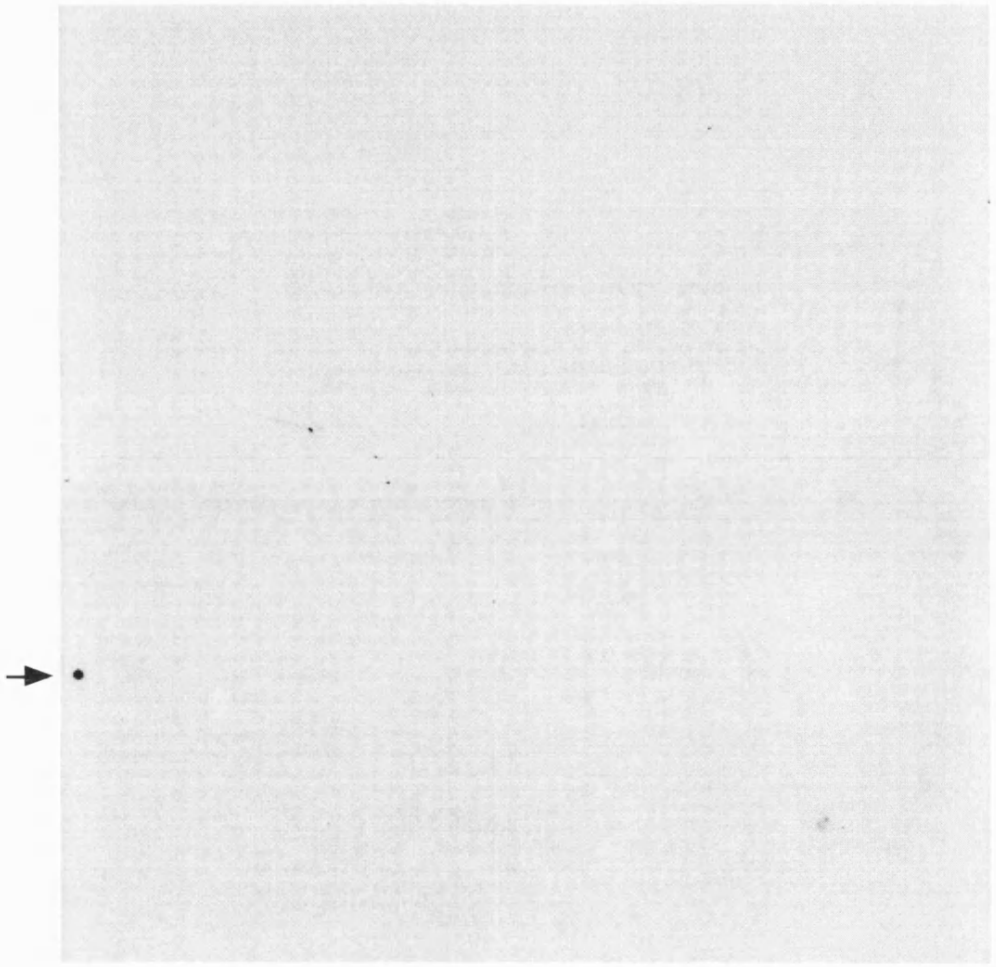


Figure 3.33 Drosophila cDNA gridded library screen with Nfe2l1
The gridded Drosophila cDNA library was hybridised with cnc1 2SM, washed to a stringency of 4x SSC, 0.1% SDS at 65°C, and exposed to film for 5 hours at -70°C. One clear replicated positive was seen (arrowed).

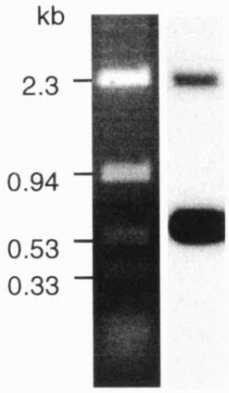


Figure 3.34 Southern analysis of Drosophila cDNA clone ICRFp520B143
(a) Clone ICRFp520B143 was digested with Bgl II and Not I to release the cDNA insert, producing a vector band of 2.3 kb, and three other insert bands (0.94 kb, 0.53 kb and 0.33 kb). (b) A Southern blot was prepared using the digested plasmid DNA, probed with cnc12SM, and washed to a stringency of 4x SSC, 0.1% SDS at 65°C, and exposed to film overnight at -70°C. The signal is localised to the 0.53 kb Bgl II/Not I band (sites determined by other RE digests not shown).

```
SCORES           Init1:   732 Initn:  1245 Opt:  1396
          84.2% identity in 450 bp overlap

        890       900       910       920       930       940
x54504 GATGAGATCTCCAAGGACTCGAACGAGAAGATCTGGAACGAGAAGAAGGAGCAATACACC
                                  ||||||: |||||||||||||||||||:||
DBN0.6                            ATCTGGNCCGAGAAGAAGGAGCAATACNCC
                                         10        20        30

        950       960       970       980       990      1000
x54504 GGCCGTCAAAAATCCAAACTGCCAAAGTGGTTCGGCGAGCGACCAGGCAAGAAGGCCGGT
       |||||||||||||||||||||||:||:||||||||||||||||||||||||||||||||
DBN0.6 GGCCGTCAAAAATCCAAACTGCCNAANTGGTTCGGCGAGCGACCAGGCAAGAAGGCCGGT
                40        50        60        70        80        90

       1010      1020      1030      1040      1050      1060
x54504 GAGCCCGAGACACCCGAGGGCGAGGAGGACGCCAAGGCCGATGAGGACATCGTCGAGGAT
       |||||||||||||||||||||||||||||:|| ||||||||:|||||:|| || |||||
DBN0.6 GAGCCCGAGACACCCGAGGGCGAGGAGGACNCCTAGGCCGATNAGGACNTCTTCTAGGAT
              100       110       120       130       140       150

       1070      1080      1090      1100      1110      1120
x54504 GATGAGGAGGTCGAGGAGGAGGTCGTCGAGGAGGAAGATGAGGAGGACGAGGAGGATGAG
       :|| |||||||| |||||||||| || |||||||:||| ||||:| | ||||:||| ||
DBN0.6 NATTAGGAGGTCTAGGAGGAGGTCTTCTAGGAGGANGATTAGGANGCCTAGGANGATTAG
              160       170       180       190       200       210

       1130      1140      1150       1160      1170      1180
x54504 GAGGAGGAGGAGGAGGAAGAGGAGGAA-GAAGAGGAAGAGGAGGAGGAAGAGGAGGAGGA
       ||||| ||:|||||:||| |||| ||: :|| |:||| |:||||||||: | ||:|||||
DBN0.6 GAGGATGANGAGGANGAATAGGATGANTNAATANGAATANGAGGAGGANTATGANGAGGA
              220       230       240       250       260       270

        1190      1200      1210      1220      1230      1240
x54504 GGAAGAAGAGGAAGAGGAGGAGGAGGAATAGACCAATCCACACTCTGGGCCCGCCCTCGC
       :||| |: |:||| |:|||||||||||||:||||::|||:|:||||||||||||||||||
DBN0.6 NGAATANTANGAATANGAGGAGGAGGAATNGACCNNTCCNCNCTCTGGGCCCGCCCTCGC
              280       290       300       310       320       330

        1250      1260      1270      1280      1290      1300
x54504 GCACCACTTTCCTTCACCACTTTTCGTAAAAACAAACACTACATTTAAATACCAAAGCCA
       |||||||||||||||||||||||||||||||||| | ||||||:||||||:||:||
DBN0.6 GCACCACTTTCCTTCACCACTTTTCGTAAAAACAAACCCCACATTTNAATACCNAANCCC
              340       350       360       370       380       390

        1310      1320          1330      1340      1350      1360
x54504 AACAATTTTAATAAAAGTA----AATGAACTAAGTAGTTAACAACTTGCTATTCAAAATC
       |||||||||||| :||:|:    ||||||||:::|: |||||||||||||||| |||||
DBN0.6 AACAATTTTAATTNAANTNAATGAATGAACTNNNTNTTTAACAACTTGCTATTCCAAATC
              400       410       420       430       440       450


x54504 A

DBN0.6 CAAAAAAAAAAAAAAAAAA
              460       470
```

Figure 3.35 BESTFIT analysis of Drosophila Troponin-T (accession no. X54504) and DBN0.6 subclone

```
SCORES            Init1:   178 Initn:   178 Opt:   192
            87.1% identity in 62 bp overlap

              1219       1209       1199       1189       1179       1169
x54504 TGTGGATTGGTCTATTCCTCCTCCTCCTCTTCCTCTTCTTCCTCCTCCTCCTCTTCCTCC
                                        || ||||||||||||||||||||||||
x78709 CCTTCCTCTCTGAGCAGCTCTGAAGGGAGCTCTTCTTCTTCCTCCTCCTCCTCTTCCTCT
         1420       1430       1440       1450       1460       1470

              1159       1149       1139       1129       1119       1109
x54504 TCCTCTTCCTCTTCTTCCTCCTCTTCCTCCTCCTCCTCCTCATCCTCCTCGTCCTCC
       ||  ||||||| ||| |||| ||||||| |||
x78709 TCTGCTTCCTCCTCTGCCTCTTCTTCCTTCTCTGAGGAGGGTGCTGTTGGTTACAGCTCT
         1480       1490       1500       1510       1520       1530
```

Figure 3.36 BESTFIT analysis of Drosophila Troponin-T (accession no. x54504) and Nfe2l1 (accession no. x78709)

3.5.2.2 Zoo blot

Subsequent to the previous experiment, some species of genomic DNA (including *Drosophila*) became available. A zoo blot containing *Drosophila* genomic DNA was hybridised with a *Xho* I fragment of cnc12SM (Figure 3.37) and gave no signal in the *Drosophila* and rabbit lanes. It is likely that there is no direct *Drosophila* homologue of *Nfe2l1*.

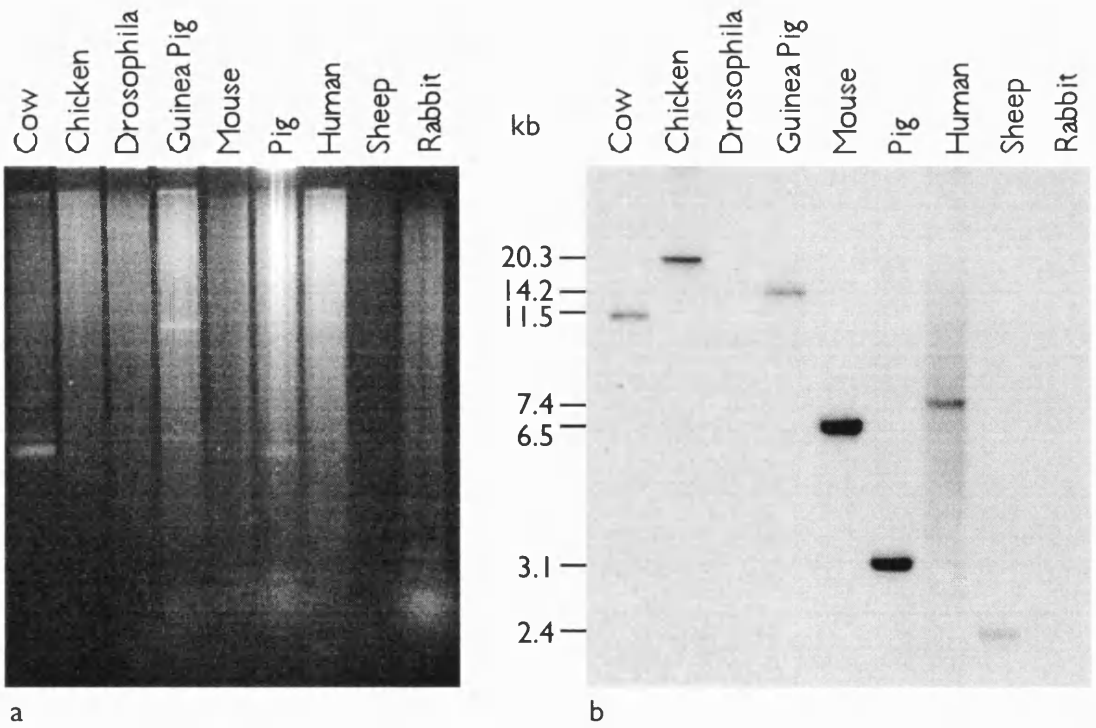


Figure 3.37 Zoo blot hybridised with Nfe2l1

Genomic DNA from a variety of species was digested with Eco RI, electrophoresed and blotted. A Xho I fragment of cnc12SM (nucleotides 440 to 773) was used as a probe, and the blot was washed to a stringency of 6x SSC, 0.1% SDS at 65°C. The blot was exposed to film for one week at -70°C.

3.5.3 The search for a mammalian homologue of *Drosophila cnc*

3.5.3.1 Isolation of Drosophila cnc cDNA

Two oligonucleotide primers, DCNC.FOR and DCNC.REV (Table 2.1) were used in a PCR experiment using as template, random hexamer-primed first strand *Drosophila* embryo cDNA (gift from R.Llevadot). This generated three bands: the expected 1kb band, and two smaller bands, 0.4 and 0.32 kb (Figure 3.38). Various PCR conditions were tested as the two smaller bands were unexpected, however all three bands were consistently amplified. The 1 kb band was subcloned into the PCRscript SK^+ plasmid vector (Stratagene). Upon sequencing and subsequent database analysis the 1 kb fragment was found not to encode *Drosophila cnc* but an unknown DNA fragment (data not shown). Another independent PCRscript SK^+ subclone, containing a 1 kb insert, was sequenced and found to be the same. At least one of the DCNC specific primers was present (as determined by sequencing), as the sequence gel had run too far to read through the other primer. Database analysis of the unknown sequence gave no significant matches. The PCR had therefore been non-specific for *cnc*.

Concurrently, an EST(GenBank accession no. AA392043) encoding the 5' end of *Drosophila cnc* was deposited into the public database from the BDGP/HHMI Drosophila EST Project (Harvey *et al.*, 1997). It was requested (clone name LD12407) and sequenced to confirm its identity: it covered the majority of the *Drosophila cnc* gene (nucleotides 19- 2410). Clone LD12407 was used as a probe to screen a 7.5 dpc and a 10 dpc mouse embryo cDNA libraries, and no positives were identified. A blot of the earlier *Drosophila cnc* PCR (Figure 3.38) and a zoo blot (Figure 3.39) were hybridised with the *cnc* EST, LD12407. The Southern blot of the PCR of *Drosophila cnc* gave no signal after probing with LD12407, confirming the previous conclusion that the PCR was not specific for *cnc*. The zoo blot showed only one band, in the *Drosophila* lane, at a washing stringency of 6x SSC, 0.1% SDS, after an exposure of one week to film at -70°C.

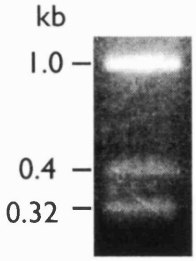


Figure 3.38 PCR for Drosophila cnc
The PCR product was electrophoresed on an 1% agarose gel, and three bands were present (1, 0.4 and 0.32 kb). The gel was blotted and hybridised subsequently with LD12407 (see above), which gave no signal.

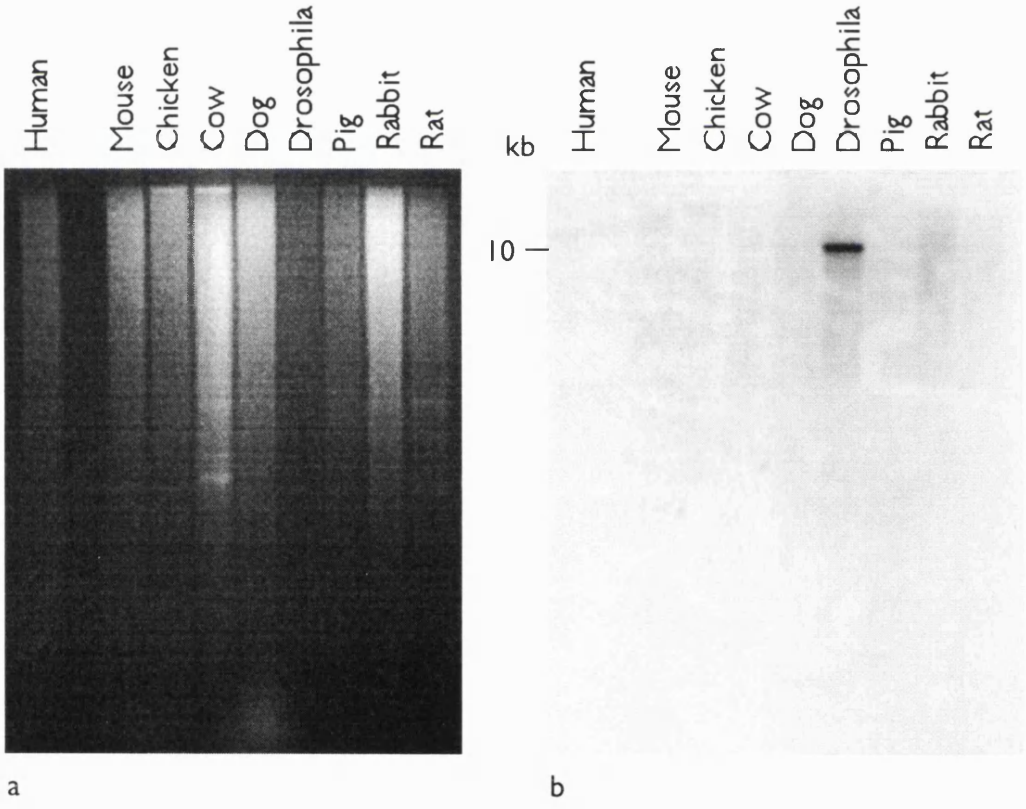


Figure 3.39 Zoo blot hybridised with Drosophila cnc EST
(a) A Southern blot was made using 5μg of Eco RI-digested genomic DNA from the species detailed above. (b) The blot was hybridised with LD12407, washed to a stringency of 6x SSC, 0.1% SDS at 65°C, and exposed to film for an overnight at -70°C. A longer exposure of 1 week at -70°C detected no additional bands (data not shown).

3.5.4 Isolation of a novel family member by dbEST screening

3.5.4.1 Identification of a novel family member

In the last three years there has been an enormous deposition of ESTs into the public domain databases (Adams *et al.*, 1995; Hillier *et al.*, 1995). In a final attempt to isolate a novel family member, a computer-based approach was adopted: the peptide translation of *Nfe2l1* was used to screen the GenBank and EST databases using the BLAST program. Three human ESTs encoding for a novel *cnc*-like gene were identified (Figure 3.40). These ESTs (and their opposite end EST if available from the WashU-Merck EST Project: http://genome.wustl.edu/est) were used in turn to identify other overlapping ESTs in the EST database, and by using a combination of BLAST analysis and the TIGR Human Gene Index (HGI) facility (http://www.tigr.org), a series of linked contigs of ESTs (Figure 3.41) was generated for this novel gene, named NFM (Novel Family Member). Figure 3.42 is a representative example of one of the Tentative Human Consensus (THC) contigs; the others are in Appendix 1 (THC212922 is not included as it was likely to be a false walk).

```
>EM:HSA45573 AA045573 z166a11.r1 Stratagene colon (#937204) Homo sapiens cDNA
             clone 509564 5' similar to PIR:S48097 S48097 LCR-F1 protein -
human
             ;.
             Length = 431

  Plus Strand HSPs:

 Score = 107 (48.2 bits), Expect = 8.4e-05, P = 8.4e-05
 Identities = 22/34 (64%), Positives = 27/34 (79%), Frame = +1

Query:    433 DEISLMDLAIEEGFNPVQASQLEEEFDSDSGLSL 466
              DEI+LM LA E+ F+P+  SQL +E DSDSGLSL
Sbjct:      4 DEINLMSLATEDNFDPIDVSQLFDEPDSDSGLSL 105

>EM:HS1184310 AA279920 zs88h06.r1 Soares NbHTGBC Homo sapiens cDNA clone
704603
             5' similar to TR:G520471 G520471 TRANSCRIPTION FACTOR LCR-F1. [1]
             ;.
             Length = 360

  Plus Strand HSPs:

 Score = 101 (45.5 bits), Expect = 0.00043, P = 0.00043
 Identities = 21/33 (63%), Positives = 26/33 (78%), Frame = +2

Query:    434 EISLMDLAIEEGFNPVQASQLEEEFDSDSGLSL 466
              EI+LM LA E+ F+P+  SQL +E DSDSGLSL
Sbjct:      2 EINLMSLATEDNFDPIDVSQLFDEPDSDSGLSL 100
```

```
>GB:R20139 R20139 yg18g01.r1 Homo sapiens cDNA clone 32759 5'.
          Length = 500

  Plus Strand HSPs:

 Score = 75 (33.8 bits), Expect = 9.0, P = 1.0
 Identities = 15/24 (62%), Positives = 19/24 (79%), Frame = +3

Query:   443 EEGFNPVQASQLEEEFDSDSGLSL 466
             E+ F+P+  SQL +E DSDSGLSL
Sbjct:     6 EDNFDPIDVSQLFDEPDSDSGLSL 77
```

Figure 3.40 Initial ESTs for NFM
Both GenBank and EST databases were searched using the peptide translation of Nfe2l1. Multiple matches were obtained but after the removal of ESTs encoding other cnc-family members, only these three matches were left.

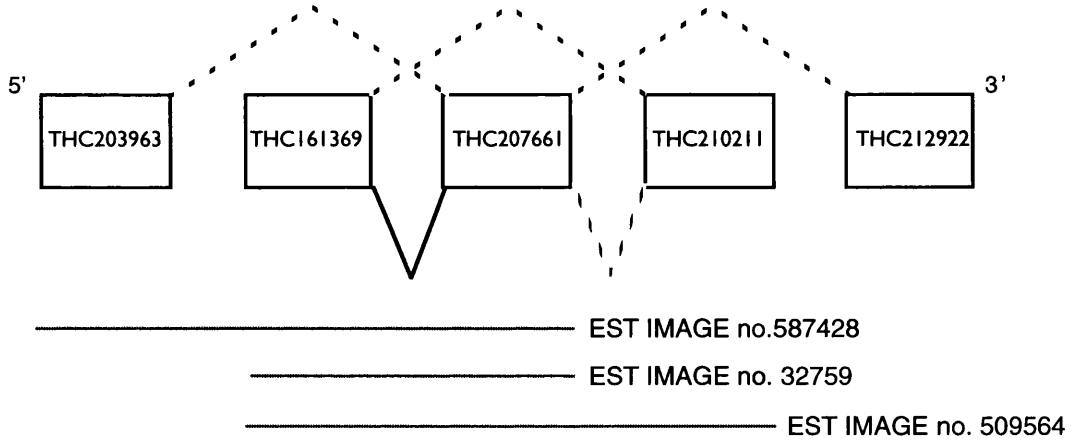


Figure 3.41 NFM EST contigs
A series of five contigs was generated: each contig has a THC number assigned by the TIGR corporation. The links between contigs are shown as a solid line if "likely" (more than one connecting EST), and as a dotted line if "tentative" (only one connecting EST). The orientation of the THC contigs shown is as suggested by their EST content; for example, both THC207661 and THC210211 have ESTs containing poly(A) tails. Three IMAGE ESTs that were obtained for analysis are also shown (section 3.5.4.2).

```
[Image]  The TIGR Human Gene Index (HGI)

HGI THC Report: THC207661

EST IDs are linked to HGI EST reports. HT# s are linked to EGAD HT reports.
GB# s are linked to GenBank accessions. ATCC#s are linked to order forms for
requesting clones.
--------------------------------------------------------------------------------

>THC207661 THC44663 THC100162 THC111612 THC181377
GATATTTTTAGTAGATTAAGAGATGACCAAGGTAGGCCAGTCAATCCCAACCACTATGCTCTCCAGTGTACCCATGATGG
AAGTATCTTGATAGTACCCAAAGAACTGGTGGCCTCAGGCAACAAaAGGAAACCCAAAAGGGAAAGAGAAAGTGAGAAGA
AACTGAAGATGGACTCTATTATGTGAAGTAGTAATGTTCAGAAACTGATTATTTGGATCAGAAACCATTGAAACTGCTTC
AAGAATTGTATCTTTAAGTACTGCTACTTGAATAACTCAGTTAACGCTGTTTTGAAGCTTACATGGACAAATGTTTAGGA
CTTCAAGATCaCACTTGTGGGCAATCTGGGGGAGCCACAACTTTTCATGAAGTGCATTGTATACAAAATTCATAGTTATG
TCCAAAGAATAGGTTAACATGAAAACCCAGTAAGACTTTCCATCTTGGCAGCCATCCTTTTTAAGAGTAAGTTGGTTACT
TCAAAAAGAGCAAACACTGGGGATCAAATTATTTTAAGAGGTATTTCAGTTTTAAATGCAAAATAGCCTTATTTTCATTT
AGTTTGTTAGCACTATAGTGAGCTTTTCAAACACTATTTTAATCTTTATATTTAACTTATAAATTTTGCTTTCTATGGAA
ATAAATTTTGTATTTGTATTAAAAATTA

 1================================THC207661================================668
 ----------------------------1---------------------------->
            -------------2------------->
                 <----------3------------
                     ------------------4------------------>
                            ------------------------5------------------------->
                            <----------------------6-------------------------
                              <-----------------------7-----------------------
                              <---------------------8-----------------------
                                <---------------------9-----------------------
                                        <------------------10------------------

#       EST Id     GB#      ATCC#   left right library
--------------------------------------------------------------------------------
1     F yi82c06.r1 R78318   383651     1   472 placenta Nb2HP, Soares
2     F ze70d09.r1 AA022505 821451    87   322 fetal heart NbHH19W, Soares
3     A EST35507   AA331415 133272   133   333 embryo, 8 week I
4     A EST80449   AA369091 173772   169   488 placenta II
5     A EST176112  AA305111 128600   227   668 Colon carcinoma (Caco-2) cell
                                               line II
6     F yi82c06.s1 R77947   383651   232   638 placenta Nb2HP, Soares
7     F zo20c03.s  AA132523          243   663 colon, Stratagene (#937204)
8     F zr18a05.s1 AA227467 232137   248   660 neuronal precursor NT2,
                                               Stratagene (#937230)
9     F yl32g06.s1 H21401   397907   265   652 breast 3NbHBst, Soares
10    F yh89c02.s1 R36639   374811   321   638 placenta Nb2HP, Soares

Sequence source codes:
F = WashU/Merck
A = TIGR

There are no hits for THC207661.

View Opposite End information.
--------------------------------------------------------------------------------
```

Figure 3.42 THC207661

The THC is generated by the TIGR corporation, as part of their Human Gene Index resource; it is only assembled if there are overlapping TIGR ESTs. The EST content of this particular THC suggests that it represents the 3' end of NFM, and it links up to 4 other THC contigs (Figure 3.41)

3.5.4.2 *Isolation and analysis of full length NFM*

Three of the ESTs (Table 3.5) distributed by the IMAGE consortium were obtained from the UK-HGMP Resource Centre and their approximate position in the NFM EST contig is shown in Figure 3.41.

IMAGE no.	Accession no. 5'	Accession no. 3'	Insert size (kb)	THC location 5'	THC location 3'
5875428	aa132584	aa132523	2.1	203963	207661
32759	r20139	r43198	1.3	161369	207661
509564	aa045573	aa045574	2.2	161369	210211

Table 3.5 NFM ESTs

One million plaques of a human placental cDNA library were screened using IMAGE EST no. 509564 and 42 primary positives were picked. Eight were taken through to tertiary screens, and isolated as individual plaques. Initial PCR analysis using combinations of vector primers and an internal primer (NFMMAP1; section 3.5.4.3) divided the phage clones into two distinct groups (Figure 3.43): there were two different sizes of 3' cDNA ends. One bacteriophage from each group (NFM2 and NFM26) was selected for further analysis. The phage inserts were subcloned into *Eco* RI-digested pBluescript SK$^+$, after Southern blot analysis to confirm that they were true positives (Figure 3.44).

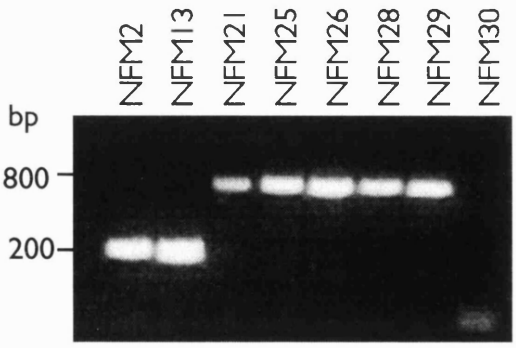


Figure 3.43 PCR analysis of NFM-positive bacteriophage clones
A 3' internal primer (NFMAP1) was used in combination with the vector primers in a PCR using as template the bacteriophage in SM buffer. Two plaques (NFM 2 and NFM13) gave a 200bp product with the 3' primer, and the remaining all gave a 800bp product.

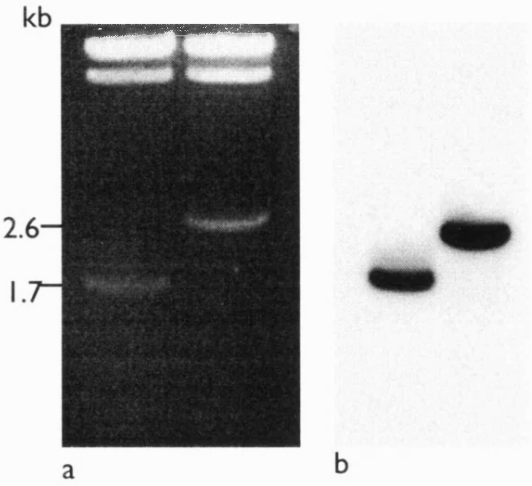


Figure 3.44 Southern blot analysis of NFM-positive bacteriophage clones
Bacteriophage clones were digested with Eco RI and electrophoresed (a). A Southern blot was prepared, probed with EST509564, washed to a stringency of 0.1x SSC, 0.1% SDS at 65°C, and exposed to film for 1 hour at room temperature.

Preliminary sequence analysis of NFM2 and NFM26 (data not shown), and comparison with the NFM EST contigs (Figure 3.41) showed that the cDNAs had a common 5' end, starting at an internal *Eco* RI site in THC and different 3' ends. Both NFM2 and NFM26 had poly(A)$^+$ tails, which corresponded with the 3' ends of the NFM EST contigs THC207661 and THC210211 respectively. Neither cDNA had full length coding sequence.

3.5.4.2 Mapping of NFM

I Localisation by PCR using the Genebridge 4 Radiation Hybrid DNA panel

Oligonucleotide primers, NFMMAP1 and NFMMAP2, were designed using sequence from THC207661 that was likely to be 3' untranslated region. Using human and hamster genomic DNA as template, they gave a human-specific 123 bp product (Figure 3.45). The Genebridge 4 Radiation Hybrid DNA panel (Gyapay *et al.*, 1996) was obtained from the UK-HGMP Resource Centre (section 2.1.14). The result obtained by PCR (Figure 3.46) using the 84 radiation hybrid cell line DNA was submitted, but failed to give a localisation as too many cell lines were positive.

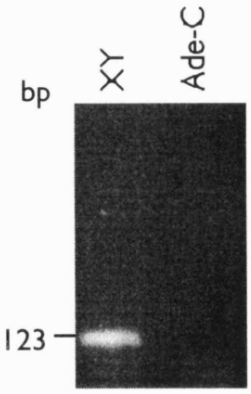


Figure 3.45 Human specificity of NFMMAP1 and NFMMAP2 primers
Standard PCR was carried out using as template human XY and hamster Ade-c genomic cell line DNA. The PCR products were electrophoresed on a 2% agarose gel. The expected 123 bp product was only seen in the human lane.

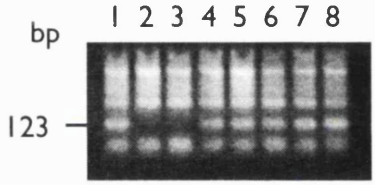


Figure 3.46 Mapping NFM using the Genebridge 4 Radiation Hybrid DNA panel
PCR analysis of 84 radiation hybrid cell line DNAs was carried out using NFMMAP1 and NFMMAP2 primers. The PCR products were electrophoresed on a 2 % agarose gel, and a representative picture is shown above. Although there was a strong background smear, the expected product (arrowed) is clearly either present or absent.

II Localisation by FISH

A gridded human PAC library (UK-HGMP Resource Centre) was screened using EST509564, and 7 positives were identified (Figure 3.47). All seven were obtained from the UK-HGMP Resource Centre, and analysed by Southern blotting (Figure 3.48): two PACS (135D7 and 209O15) were strongly positive by hybridisation with EST509564.

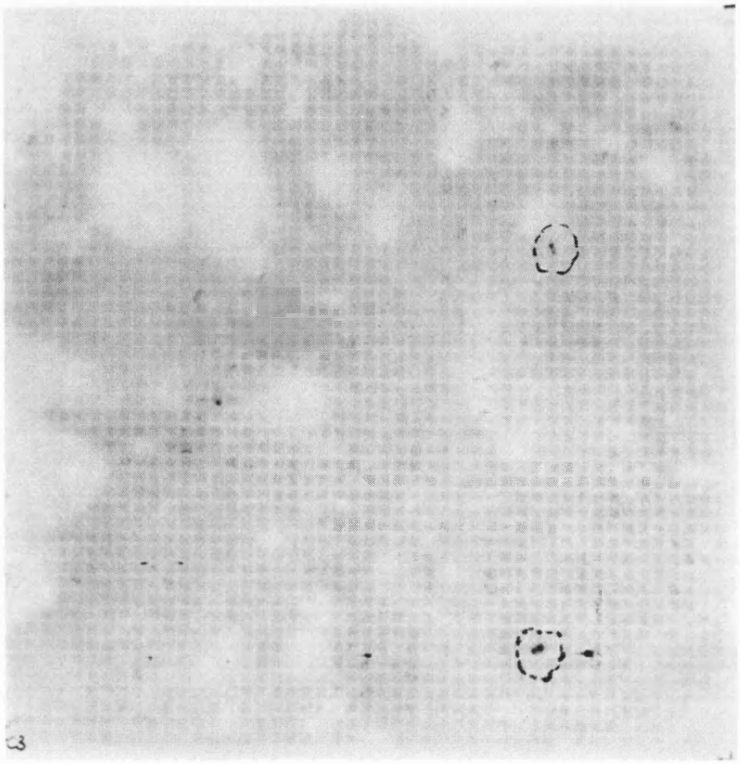


Figure 3.47 Human PAC library screen for NFM
A gridded human PAC library was screened using EST509564. The library is supplied on 7 filters, and a representative result of the hybridisation is shown. The filters were washed to a stringency of 6x SSC, 0.1% SDS at 65°C, and exposed overnight at -70°C.

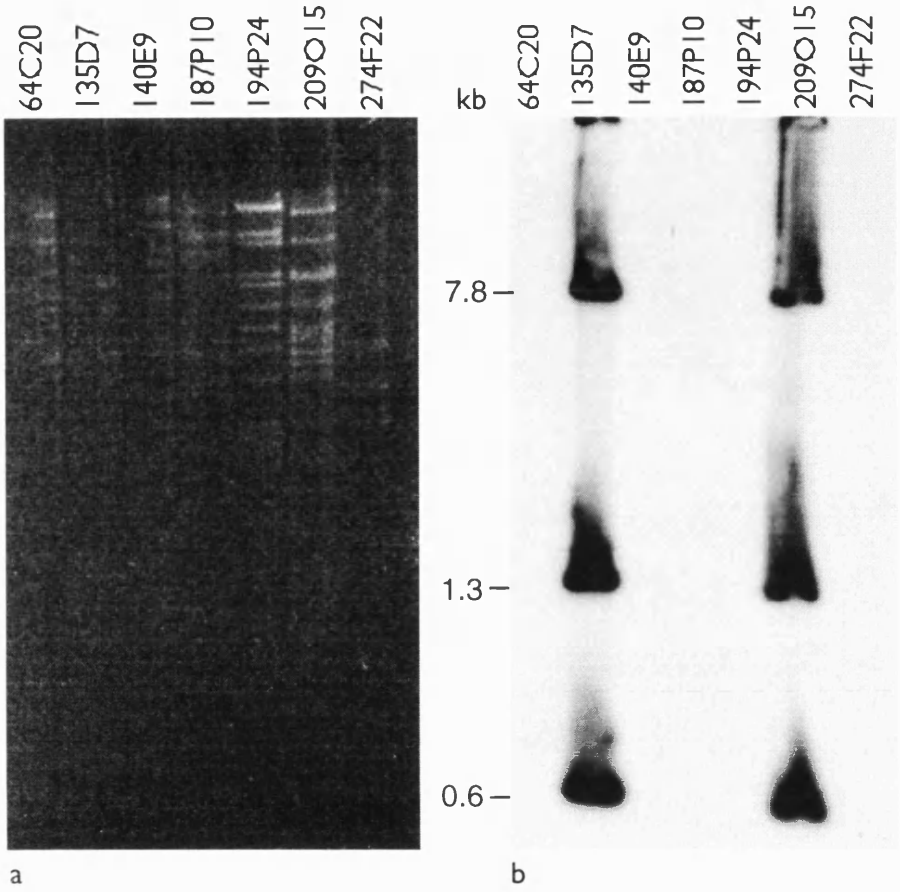


Figure 3.48 Southern analysis of human PACs, positive for NFM
(a) A Southern blot was prepared with Hind III-digested DNA from seven PAC clones. (b) The blot was hybridised with EST509564 and washed to a stringency of 0.1xSSC, 0.1% SDS at 65°C, and exposed to film for 4 hours at -70°C. Three bands in two lanes (PACs 135D7 and 209O15) were clearly identified.

At this time the sequence of human BAC clone RG119C02 (Dubuque *et al.*, 1998: GenBank accession no. AC004520) was deposited into GenBank; it contained a predicted protein and coding region identical to NFM, and mapped to 7p15 as part of a collaborative chromosome 7 genomic sequencing project. With the mapping information available, FISH analysis on normal XY metaphase spreads was carried out (Figure 3.49), using biotin-labelled PAC135D7 or PAC209O15 (green signal) and a DIG-labelled marker probe (red signal) for chromosome 7 centromere and 7q telomere (Oncor).

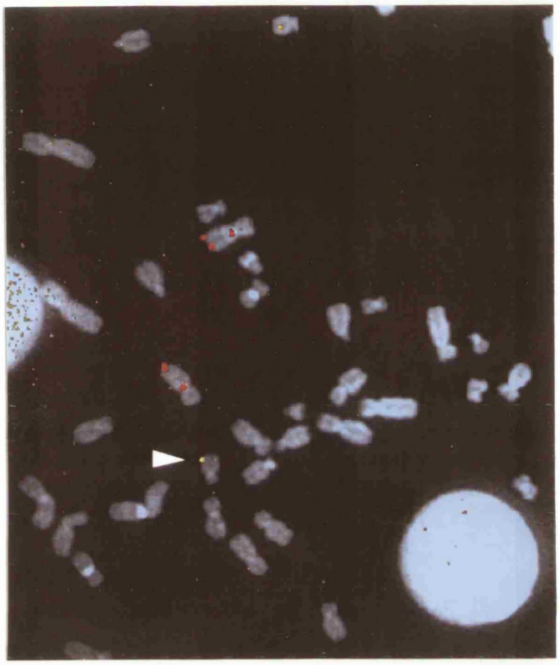


Figure 3.49 FISH localisation of NFM-positive PAC
The result shown is representative of a number of examined metaphases. In this instance, PAC135D7 (yellow signal) is clearly localising to the telomere of the long arms of a chromosome other than 7 (red signal).

Both PAC135D7 and PAC209O15 did map not to chromosome 7, but to the telomere of the long arms of a smaller chromosome.

3.5.4.3 Zoo blot

A zoo blot was hybridised with IMAGE EST no.32759 (Figure 3.50). Four bands were detected in the human lane, and a band was clearly seen in both cow and dog lanes. A faint indistinct band was also seen in the mouse lane. With the release of the sequence for BAC clone RG119C02, restriction endonuclease analysis using MAPSORT (GCG program at UK-HGMP Resource Centre) revealed that IMAGE EST no.32759 should only detect one *Eco* RI-digested human genomic DNA band of 3.374 kb. The washing stringency of the zoo blot was low (6x SSC, 0.1% SDS at 65°C) so possibly other highly related genes were being detected. Alternatively, IMAGE EST no.32759 could be

detecting the presence of NFM pseudogenes. This explanation is further supported by the experimental mapping data.

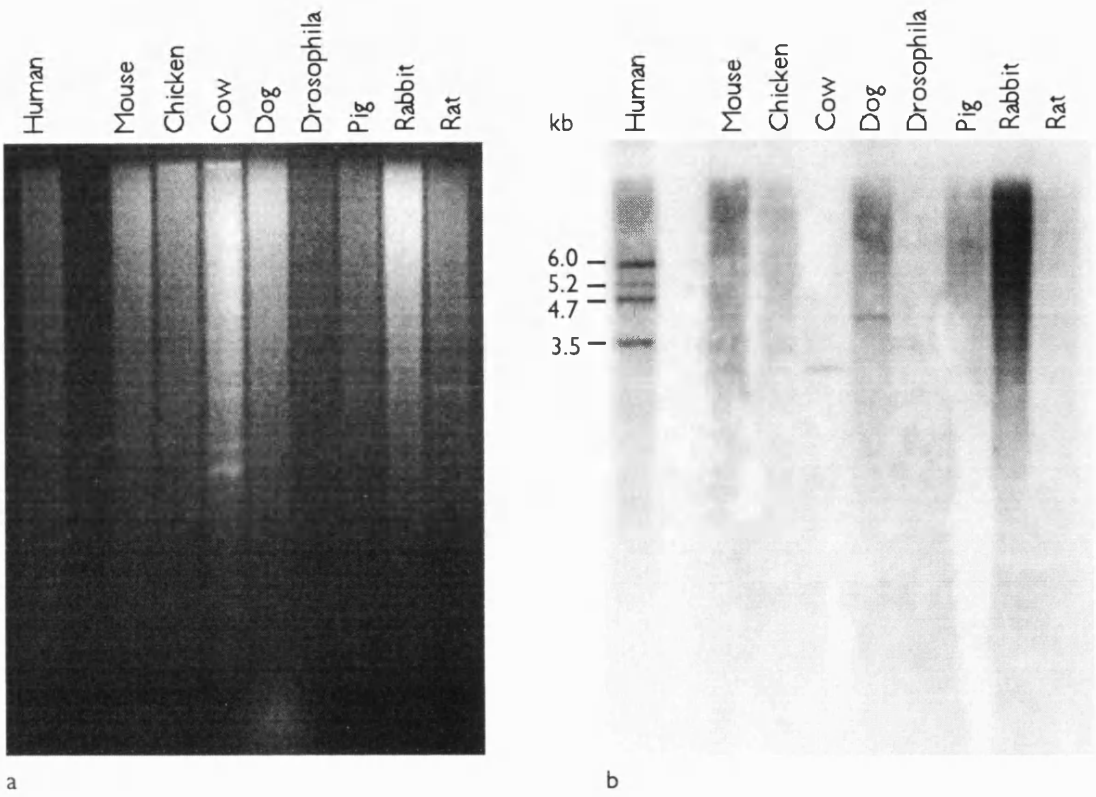


Figure 3.50 Zoo blot with NFM
Genomic DNA from a variety of species was digested with Eco RI, electrophoresed and blotted. IMAGE EST no.32759 was used as a probe, and the blot was washed to a stringency of 6x SSC, 0.1% SDS at 65°C. The blot was exposed to film for one week at -70°C.

3.5.4.5 Preliminary expression analysis of NFM

The sites of NFM expression were determined by northern analysis (Figure 3.51). A predominant band of 2.8 kb is seen, at variable intensities, in all tissues tested. An additional band (3.5 kb) is present in human foetal kidney. On consideration of the poly(A) tails of the NFM2 and NFM26 cDNAs, and comparison with the genomic sequence (GenBank accession no. AC004520), the larger transcript is likely to be due to alternative polyadenylation, reminiscent of the*Nfe2l1* transcripts.

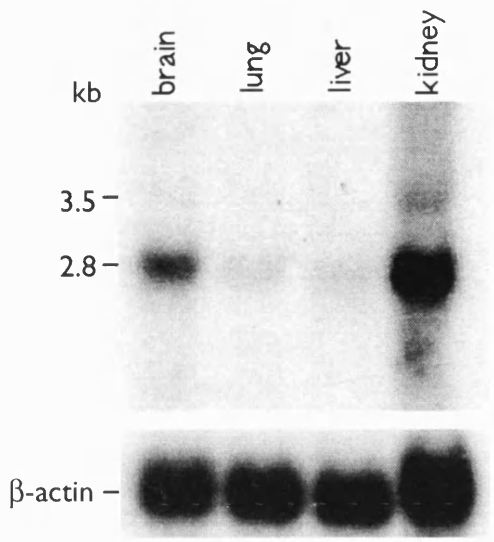


Figure 3.51 Northern analysis of NFM expression

A Clontech human foetal northern was hybridised with EST509564 according to manufacturer's instructions, washed to a final stringency of 0.1x SSC, 0.1% SDS at 68°C, and exposed to film overnight at -70°C. β-actin is shown as a loading control.

CHAPTER 4 DISCUSSION

4.1 FOREWORD

The discussion chapter is divided into four parts. The first section examines the results obtained for murine *Nfe2l1*, including sequence analysis, chromosomal localisation and expression: a description of, and comparison with, its human homologue, *NFE2L1* is also included. The following section summarises the evidence presented in this thesis and the published data, for a novel family of CNC-related bZIP transcription factors. The third section analyses the characterisation of a novel CNC-related gene, *NFM*, and compares the different research approaches taken with *Nfe2l1* and *NFM*. Finally, there is a section considering future work for this project and, on a broader level, for the CNC-family of transcription factors.

4.2 *Nfe2l1*- A NOVEL MURINE bZIP TRANSCRIPTION FACTOR

Nfe2l1 is a novel murine transcription factor, belonging to the bZIP superfamily (Landschulz *et al.*, 1988). The initial premise of the thesis was the similarity of clone 8dl to *Drosophila cnc* (Mohler *et al.*, 1991; sections 1.4.3.1,1.5), and the possibility that a fragment of the murine homologue of *cnc* had been isolated. The original aim of the thesis was to answer these questions: is *Nfe2l1* the murine homologue of *Drosophila cnc*, and what, if any, is its role in development?

4.2.1 Isolation and sequence analysis of *Nfe2l1*

4.2.1.1 Sequence comparison of murine Nfe2l1 and Drosophila cnc

The original *Nfe2l1* cDNA fragment isolated by Dr. Johnstone, 8dl (section 1.5), showed strong homology to a *Drosophila* protein, CNC (section 1.4.3.1). *cnc* is a member of the bZIP family of transcription factors and its segmental expression in the developing *Drosophila* embryo suggested a role in head specification (Mohler *et al.*,

1991), which was later confirmed by the isolation and characterisation of *cnc* mutants (Mohler *et al.*, 1995).

The *cnc* gene encodes a 2.8 kb transcript, with a potential ORF of 533 aa. Towards its C-terminus is a bZIP motif: analysis of the zipper region suggests that it is likely to be an obligate heterodimer, and no dimerisation partners have so far been reported. It contains a potential transactivation domain (Courey *et al.*, 1989), a glutamine-rich region, on the C-terminal side of the bZIP motif and an N-terminal sequence much smaller than that of *Nfe2l1.*

The striking similarity of 8dl and CNC started in a region just 5' to the DNA-binding basic domain of the bZIP motif (Figure 1.2). The similarity amongst bZIP proteins is most conserved in the basic DNA-binding domain, with the heptad repeat of leucines being the most obvious conserved feature in the leucine zipper dimerisation domain (Hurst, 1995). There is limited sequence similarity outwith these regions but families of bZIP proteins have been defined (Hurst, 1995). The degree and position of conservation between 8dl and CNC suggested either a direct homology or a family relationship. The full length sequence of *Nfe2l1* (Figure 3.10) revealed that the similarity to CNC was essentially restricted to domains immediately flanking the bZIP region (Figure 4.1). The CNC5' domain is immediately adjacent to the basic DNA-binding domain, and could possibly be involved in DNA binding itself, or in influencing the specificity and stability of the interaction of the basic domain with DNA. Two members of the Maf bZIP family, *c-maf* and *NRL*, share an N-terminal extended basic DNA-binding domain (Kerrpola and Curran, 1994) similar to *Nfe2l1*, and interestingly, members of this family have been shown to interact *in vitro* with human NFE2L1 (section 4.2.5.1; Moi *et al.*, 1994b; Marini *et al.*, 1997; Johnsen *et al.*, 1996, 1998). The CNC3' domain is immediately C-terminal to the zipper dimerisation region, and could potentially influence dimer stability (Cohen and Curran, 1990; Katagiri *et al.*, 1992). Since the conservation is limited to these domains alone, it can be concluded that *Nfe2l1* is unlikely to be a direct homologue of *Drosophila cnc.*

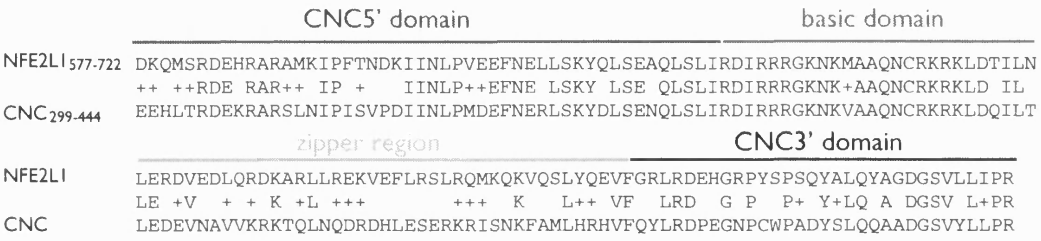


Figure 4.1 'CNC' domains
The similarity between NFE2L1 and CNC is most striking around the bZIP motif.

4.2.1.2 Features of the Nfe2l1 sequence

I General characteristics of *Nfe2l1*

The full length sequence of *Nfe2l1* is 4.4 kb. Initial northern analysis of cell line RNA (Figures 3.18 and 3.19) demonstrated two *Nfe2l1* transcripts of approximately 4.4 kb and 4.0 kb, but only one transcript was derived from analysis of the murine 8.5 dpc cDNAs. Subsequent northern analysis of staged mouse embryos, adult mouse and human tissues showed that there was one predominant transcript of approximately 4.4 kb, and a variably expressed 4.0 kb band (Figures 3.20-22).

An adequate Kozak consensus sequence is found at nucleotide positions 401-407, but there is a cluster of methionines approximately 800 bp downstream (Table 4.1). Translation usually initiates uniquely at the first ATG codon in an adequate context (Kozak, 1995), so it is most likely that the more upstream ATG codon is the initiator codon: consistent with this conclusion, it is preceded by terminator codons in all three reading frames. Using the first ATG codon (nuc $_{401-407}$) as the translational start site, an ORF of 741 amino acids is generated, containing numerous features typical of a transcription factor, the most significant being the bZIP motif towards its C-terminus.

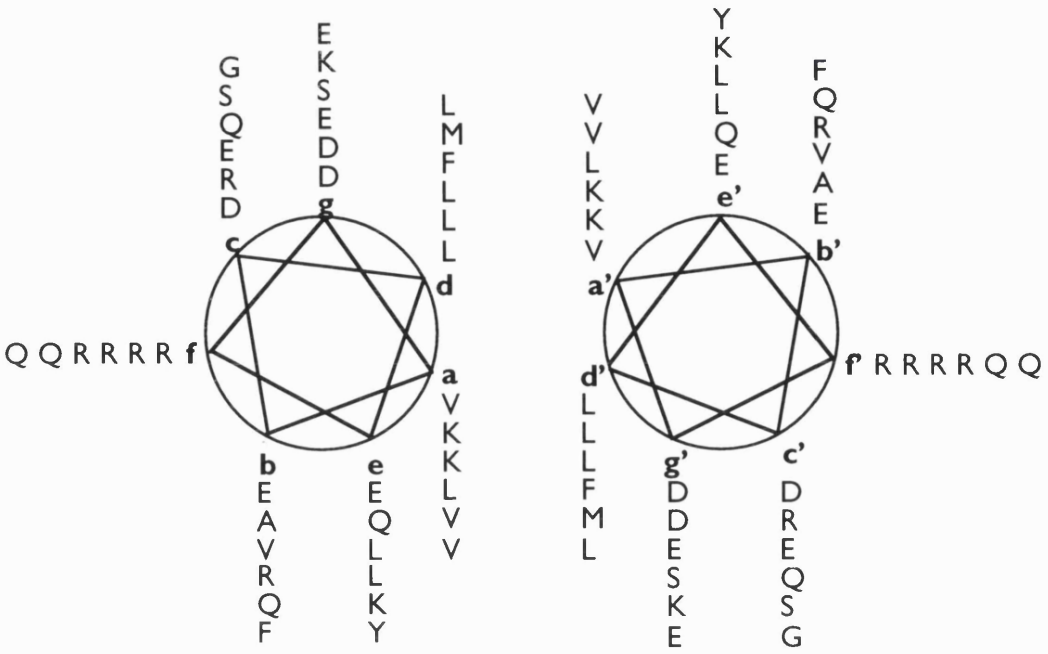
ATG codon no.	Nucleotide position in murine *Nfe2l1*	Mouse	Human	Kozak consensus sequence?
1	401-407	GCAatgC	GCAatgC	adequate
2	1265-1271	CTCatgT	CTCatgT	poor
3	1274-1280	ATCatgG	ATCatgG	near optimal
4	1280-1286	GAAatgC	GAAatgC	adequate
5	1289-1295	GCTatgG	GCCatgG	strong

Table 4.1 Potential initiation sites for translation

The 5'UTR is longer than average (403 bp), which is a common feature of proteins with a strictly regulated expression pattern (Kozak, 1987; Pesole *et al.*, 1994), and may be important for translational control (Jansen *et al.*, 1995). The 3'UTR is 1763 bp long, and contains numerous AT-rich motifs, often found in the 3'UTRs of proto-oncogenes and lymphokines including c-*fos* (Treisman, 1985; Chen and Shyu, 1994): these are thought to promote mRNA degradation (Chen and Shyu, 1995; Zubiaga *et al.*, 1995), and may also be involved in the regulation of translation initiation efficiency (Jackson, 1993). Other bZIP families such as the Maf family (Motohashi *et al.*, 1997) have large 3'UTRs of indeterminate function. Other regions in the 3'UTR (and 5'UTR) may have important functions in the regulation of mRNA stability, translation, RNA processing, nuclear export and intracellular localisation (Theil, 1993; Curtis *et al.*, 1995; St. Johnston, 1995).

II The bZIP motif of NFE2L1

The bZIP motif is characterised by an arginine- and lysine-rich DNA-binding basic domain, followed by six heptad repeats of leucine and hydrophobic residues of the leucine zipper. The basic domain is as described for 8dI, with *Drosophila* CNC being most similar, followed by JUN-related (Ryder *et al.*, 1988) and FOS-related proteins (Franza *et al.*, 1987). Typically, the leucine zipper domain is less well conserved, and the zipper region in NFE2L1 is most similar to *Drosophila* CNC, with similarity essentially limited to the heptad leucine repeat of other bZIP proteins. As NFE2L1 can only function as a dimer, its binding specificity is in part determined by its choice of partner: the leucine zipper region mediates dimerisation, forming either homodimers or heterodimers. The sequence of the zipper region can be informative about the likely choice of partner. The NFE2L1 zipper region is shown diagrammatically (Figure 4.2).



a

a	b	c	d	e	f	g
			L	E	R	D
V	E	D	L	Q	R	D
K	A	R	L	L	R	E
K	V	E	F	L	R	S
L	R	Q	M	K	Q	K
V	Q	S	L	Y	Q	E
V	F	G				

b

Figure 4.2 Helical wheel diagram of the NFE2L1 leucine zipper
(a) The bZIP domain of NFE2L1 is displayed as a helical wheel (section 1.4.1.2) to demonstrate the proximity of the a and d residues, on the inner face of the coiled-coil. Electrostatic interactions between the opposing e and g residues (e' and g; e and g'), are thought to contribute to the stability (or instability) of the dimer. (b) The bZIP domain is shown as a heptad repeat.

Residues at the **a** and **d** positions are typically hydrophobic: the substitution of leucines at position **d** by other hydrophobes or small apolar residues is acceptable (Hu *et al.*, 1990; Kouzarides and Ziff, 1989). Hydrophobic residues are also preferred at position **a**, although polar residues are less destabilising here than at position **d**. The presence of two charged amino acids in position **a** of NFE2L1 is reminiscent of FOS, and may well effect its choice of partner: monomers that stabilise the charge in the appropriate heptad repeat would be favoured, and homodimerisation would be less likely. The presence of several acidic amino acids in positions **e** and **g** is also similar to

FOS. The selective heterodimerisation of FOS and JUN can be accounted for in part by charged residues at positions **e** and **g** (Schuermann *et al.*, 1991; O'Shea *et al.*, 1992), and other studies have supported the importance of the interhelical salt bridges between **e** and **g** residues for dimer formation (Vinson *et al.*, 1993). In NFE2L1, the preponderance of acidic residues at positions **e** and **g** suggest that stable homodimers are unlikely to form. CNC is also likely to be an obligate heterodimer (Mohler *et al.*, 1991). The role of the residues at **b**, **c** and **f** is unknown, and in NFE2L1, there is a striking run of basic amino acids at position **f**.

III Other recognised functional domains and sites of NFE2L1

Sequences within the basic domain of the bZIP are probably sufficient for nuclear localisation (Waeber and Habener, 1991; Morgan *et al.*, 1992), although there is also a potential nuclear localisation domain at amino acids 730-737 (**RR**QE**RK**P**K**, basic amino acids in boldface).

The sequence N-terminal to the bZIP motif is serine/threonine rich and contains a polyserine tract, features characteristic of transcription factors: serine residues may be important in providing a polar interface for protein/protein interactions (Pascal and Tjian, 1991). Other bZIP proteins have been found to have serine/threonine-rich domains, that can undergo post-translational modifications to modify their transcriptional activity (Clauss *et al.*, 1996). An acidic domain (aa $_{441\text{-}447}$) is flanked on the N-terminal side by the serine/threonine rich residues and on the C-terminal the polyserine stretch: it has a net charge of -14, and may be involved in activation of RNA polymerase II transcription factors (Mitchell and Tjian, 1989).

The program MOTIFS (GCG collection at UK-HGMP Resource Centre) identified several known motifs, as defined in the PROSITE database, in the protein NFE2L1 in keeping with its proposed role as a transcription factor (Appendix 3). Numerous potential phosphorylation sites were found, casein kinase II (CK-2) phosphorylation sites (Pinna, 1990) being the commonest: seventeen in total, two of them in the acidic domain and one in the serine repeat domain. CK2 phosphorylation can alter transcriptional activity, both positively and negatively: it has been shown to decrease the DNA-binding activity of c-JUN (Lin *et al.*, 1992), and increase transactivation of CREB-α (Lee *et al.*, 1990). There were also five protein kinase C (PKC) sites (Kishimoto *et al.*, 1985), one tyrosine phosphorylation (Hunter, 1982), one c-AMP- and c-GMP-dependent protein kinase phosphorylation site (Glass *et al.*, 1986). Other potential post-translational modification sites were found including four myristoylation (Towler *et al.*, 1988) sites (one in the serine repeat domain) and eight N-glycosylation sites (Gavel and von Heijne, 1990). Other functional domains are

likely to be present, possibly involving conserved residues, but are unrecognised by current sequence analysis methods.

4.2.2 The human homologue of *Nfe2l1*

The human homologue of *Nfe2l1* was cloned and reported whilst this work was ongoing. Three independent groups have isolated *NFE2L1,* and their findings are briefly summarised in the following sections, ending in a comparison of mouse and human *NFE2L1* genes. *NFE2L1* and its isoforms are variously called *NRF1* (Chan *et al.*, 1993), *TCF11* (Luna *et al.*, 1994) and *LCR-F1* (Caterina *et al.*, 1994). More detailed consideration of all the published functional data and conclusions for NFE2L1 will be discussed later in appropriate sections (4.2.5).

4.2.2.1 Cloning of Nrf1, an NF-E2-related transcription factor, by genetic selection in yeast (Chan et al., 1993)

Using a genetic complementation strategy in yeast, a cDNA from the human K562 erythroleukemia cell line, was isolated which bound to tandem NF-E2/AP1 sites (Collis *et al.*, 1990; Moi and Kan, 1990; Ney *et al.*, 1990), and activated transcription. The NF-E2 binding site (Mignotte *et al.*, 1989) is an extension of the AP1 site (Franza *et al.*, 1988; Halazonetis *et al.*, 1988), and will be discussed in later sections. The 5 kb transcript, named *NRF1*, had a single long ORF encoding potentially 742 amino acids. The predicted translation product from the first ATG codon is 81 kDa; if translation is from the first internal near optimal ATG codon (Table 4.1; ATG codon 3), a protein of 50 kDa is predicted. *In vitro* transcription and translation of the entire coding region of *NRF1* generated two products: a major product of 110 kDa and a minor one of 65 kDa. The smaller product was shown to be likely to be derived from the internal optimal ATG codon. The disparity in sizes between predicted (81 kDa and 50 kDa) and actual (110 kDa and 65 kDa), was postulated to be due to aberrant gel migration (Van Beveren *et al.*, 1983). Immunoblot experiments using K562 whole cell extract, detected only one major endogenous product of 65 kDa. They concluded that the smaller 65 kDa product detected by immunoblotting could be either from usage of the internal ATG codon, or from intrinsic properties of the larger protein product giving an aberrant gel migration pattern. Northern analysis demonstrated two transcripts of approximately 5 kb, at variable levels, in all human adult tissues, erythroid, and non-erythroid cell lines tested. The ability of NRF1 to activate transcription through the tandem NF-E2/AP1 repeat present in HS2 was confirmed in K562 and yeast cells.

4.2.2.2 Molecular cloning of a putative novel human bZIP transcription factor on chromosome 17q22 (Luna et al., 1994)

The original cDNA was cloned fortuitously as a false positive from a placental cDNA library. A total of four cDNA libraries were screened, and 78 cDNA clones were analysed and used to generate a consensus sequence for *TCF11* of 4760 bp. Multiple alternative transcripts were found: variations were found in both the 3' and 5' UTR regions and within the coding region. The alternative coding region exons all occurred in the N-terminal activation domain (Caterina *et al.*, 1994), and could well alter the transcriptional activity of TCF11. Alternative polyadenylation signals were used to generate one 3'UTR 530 bp shorter than the other. Further work by the same group (Luna *et al.*, 1995) on the genomic organisation of the gene *TCF11*, demonstrated the use of alternative splicing to produce the various mRNA isoforms cloned. They proposed that by using the initiator ATG codon (Table 4.1; ATG codon no.1), four different TCF11 protein isoforms would be produced (772 aa, 769 aa, 742 aa, 731 aa respectively). The expression studies essentially repeated the findings of Chan *et al.*, 1993, and confirmed that the larger 5.0 kb transcript contained the longer 3'UTR. FISH localised *TCF11* to 17q22 using a cosmid genomic probe. The analysis of the genomic structure of the *TCF11* gene showed that the alternative transcripts were produced by alternative splicing from a single gene. The precise details of the genomic organisation (Luna *et al.*, 1995) revealed the exon/intron structure of *TCF11,* and is shown schematically below (Figure 4.3a); the composition of *Nfe2l1/NRF1* is shown in Figure 4.3b. The deposited *TCF11* gene (GenBank accession no. X77366) included an extra exon (no.4) relative to *NRF1* and *Nfe2l1.* Their genomic analysis also highlighted that the first 300 bp of *NRF1* was actually a DNA co-ligation event of ribosomal origin.

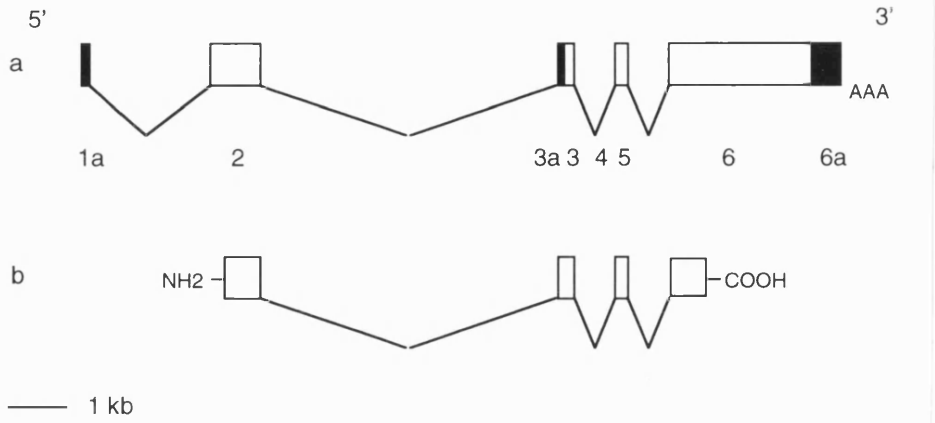


Figure 4.3' ***Schematic representation of the exon/intron structure of Nfe2l1/NRF1 and the relative position of NFE2L1/NRF1 and its domains***

(a) The identified exon/intron structure of Nfe2l1/NRF1 (see Figure 4.3): the exons are represented by boxes and the introns by lines. (b) A schematic of NFE2L1/NRF1: the protein is represented by boxes, and the connecting introns are depicted by lines. Exon 6 encodes from amino acid 296 onwards, and therefore includes the acidic domain, the serine repeat region, the CNC domains and the bZIP region.

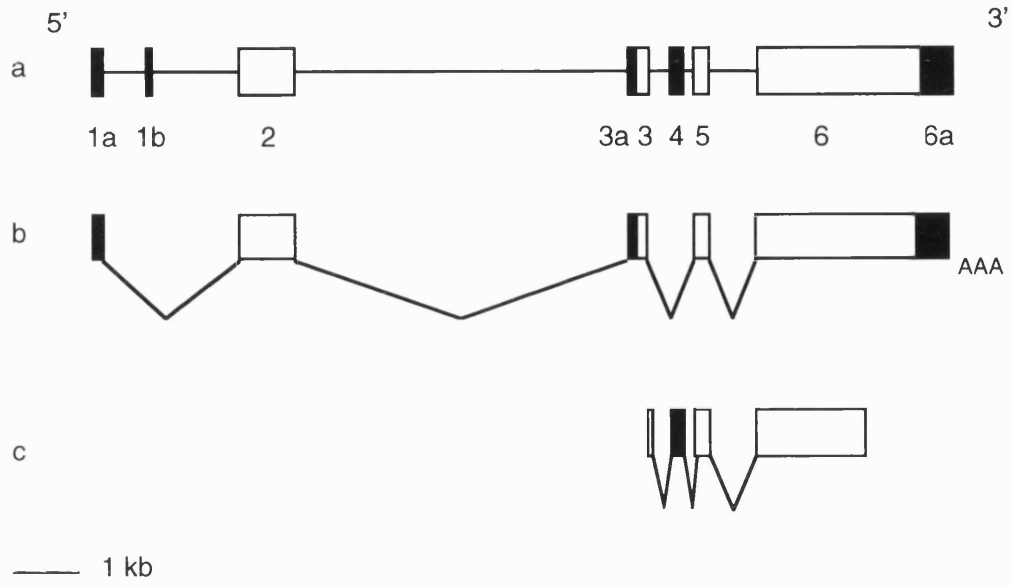


Figure 4.3 Schematic representation of the exon/intron structure of TCF11, Nfe2l1/NRF1 and LCR-F1

(a) All identified exons of TCF11: the exons are represented by boxes (open, common; solid, alternate) and the introns by a line. Luna et al., 1994, detected 4 transcripts equivalent to the following exon combinations: 1a, 2, 3a, 3, 4, 5 ,6; 1a, 2, 3a, 3, 4, 5, 6, 6a (equivalent to TCF11); (b) 1a, 2, 3a, 3, 5, 6, 6a (equivalent to Nfe2l1/NRF1); (c) part of 3, 4, 5 ,6 (the deduced exon structure of LCRF1)

4.2.2.3 Cloning and functional characterization of LCR-F1: a bZIP transcription factor that activates erythroid-specific, human globin gene expression (Caterina et al., 1994)

A K562 cell line cDNA expression library was screened with multimerised AP1-like sites. *LCR-F1* was isolated, and was used to screen another K562 cDNA library, which enabled a 2100 bp consensus sequence to be generated (Figure 4.3c). There is an ORF of 447 amino acids, using the initiator ATG codon no. 5 (Table 4.1). It was noted that *LCR-F1* and *NRF1* were identical except that *LCR-F1* contained a 91 bp insertion (exon 4) located 119 bp upstream of the predicted translational start site . This introduced a stop codon and was thought most likely to be an untranslated exon. By comparison with the *TCF11* genomic sequence (Luna *et al.*, 1995), the exonic structure of *LCR-F1* could be determined (Figure 4.3c). The 5' end of the cDNA started approximately half way through exon 3 and finished two-thirds of the way through exon 6. It contained exon 5, and putatively exon 4 , but the latter exon did not match the genomic sequence exactly (GenBank accession no. X84060), and contained several nucleotide changes that introduced stop codons. These variations may be explained by poor sequence quality or the possibility that a pseudogene had been cloned.

The expression data was limited to cell line total RNA, and was similar to that obtained by Chan *et al.*, 1993, although only the smaller transcript was present in human reticulocytes. Transient transfection assays determined that Gal4/LCR-F1 fusion proteins activated high-level erythroid-specific globin gene expression, possibly through an erythroid-specific co-activator (Luo *et al.*, 1992; Corcoran *et al.*, 1993). The N-terminal end of LCR-F1, containing the acidic domain, functioned as a strong transactivation domain specifically in erythroid cells.

4.2.2.4 Comparison of murine and human NFE2L1

Human and murine NFE2L1 are strikingly similar: 97 % identical and 98% similar at the protein level, with only one extra murine amino acid (E_{214}) and two fewer serines in the murine serine repeat. In the coding region, human *NFE2L1* is 90% identical at the nucleotide level to murine *Nfe2l1*. Conservation is also seen in the 5' UTR (86% identical) and 3' UTR (80% identical).

The conservation seen in the 3' UTR is greatest in two separate blocks with more than 90% identity (Figure 4.4), with the remaining 3'UTR being 70% identical. Both conserved blocks (nuc $_{3645-3887}$ and nuc $_{4208-4373}$) are around the two potential polyadenylation signals (Figure 4.4). It has been estimated that more than 30% of vertebrate mRNAs have conserved regions within their 3'UTRs (Duret *et al.*, 1993): these sequences are essentially unique and therefore match only to corresponding regions of orthologous mRNAs in other species. BLAST searches of the databases (GenBank, EMBL, dbEST) with the two conserved blocks of 3'UTR sequence of *Nfe2l1* brought up only matches with human or mouse *Nfe2l1* cDNAs, and so are potentially unique. It has been proposed that the highly conserved blocks of sequence may form long perfect duplexes with antisense transcripts, which in turn, may be essential for post-transcriptional regulatory systems such as mRNA destabilisation/degradation (Lipman, 1997).

An examination of the available 3' ends of both mouse and human ESTs in dbEST database showed that in the mouse, 100% of ESTs (3/3) had the longer poly(A) tail, and in the human, 22.5% (18/80) of ESTs had the shorter poly(A) tail, and 77.5% (62/80) of ESTs had the longer poly(A) tail. Alternative polyadenylation can reflect tissue specificity, and can influence translational efficiency (De Sauvage *et al.*, 1992).

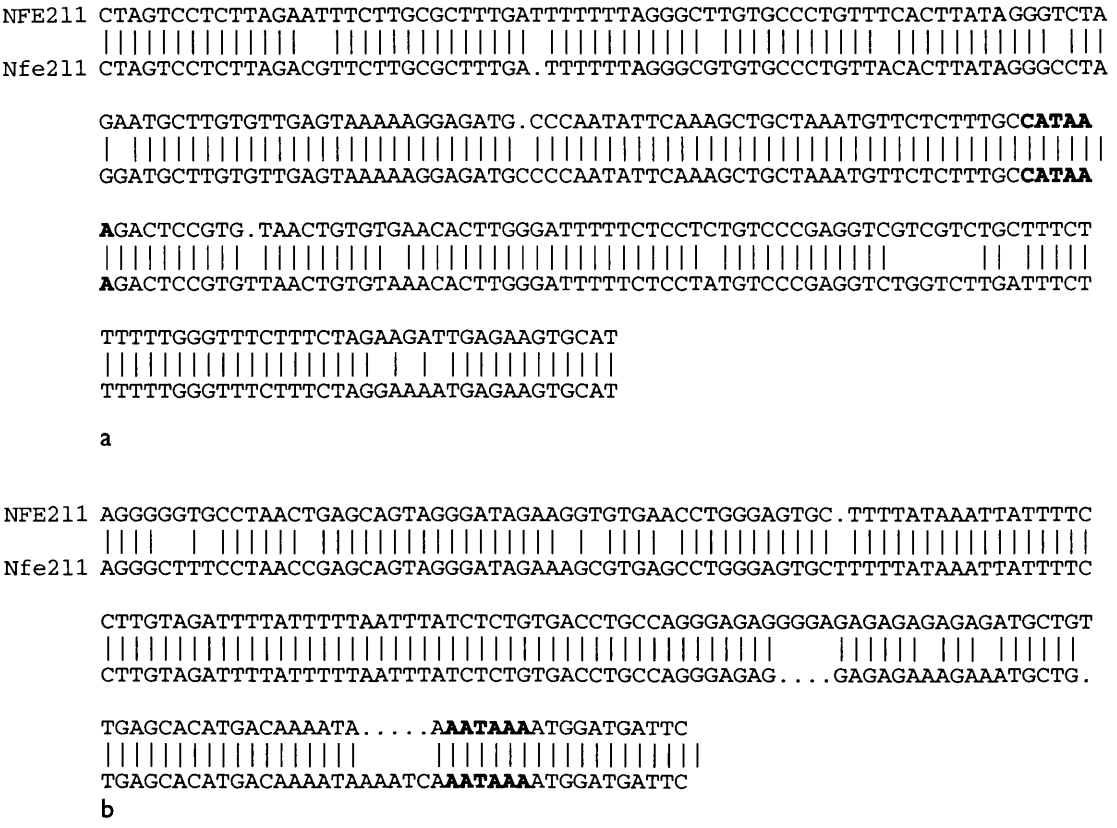


Figure 4.4 BESTFIT analysis of the 3'UTR of human and murine NFE2L1
The program BESTFIT (GCG collection at UK-HGMP Resource Centre) compared the 3' UTR of human and murine NFE2L1. A match (with gaps) of overall 80% identity was found throughout the length of the 3'UTR. The majority of the sequence was actually 70% identical, with two highly conserved blocks of sequence (greater than 90% identity), as shown above. The potential polyadenylation signals are highlighted in bold.

Conservation is 100% in the bZIP domain and N-terminal CNC domain. Conservation is also 100% in the acidic domain. The majority of the potential post-translational modification sites are conserved, with the loss of one CKII site (outwith the acidic domain) and one PKC site.

Examination of the potential initiation translational sites (Table 4.1) in human and mouse shows that ATG codon nos. 1-4 are a 100% match, with one nucleotide change in ATG codon no.5. The regions between the two first potential start sites is strongly conserved at the amino acid and nucleotide level, suggesting that this segment of the protein is functionally important. The absence of any in-frame termination codon in the murine sequence increases the likelihood that the first initiation site is functional. This is backed up by studies mentioned previously which demonstrate that initiation of translation occurs at the first adequate Kozak consensus sequence, and to be able to use the second initiation codon would either require the occurrence of alternative splice

isoforms or leaky scanning, which is uncommon and inefficient in cellular mRNAs (Kozak, 1992, 1996). However there are examples of transcription factors, such as *Pit-1* and c-*myc*, that employ multiple initiation codons to produce N-terminally extended or truncated protein products (Voss *et al.*, 1991; Hann *et al.*, 1992). The choice of initiation site is another point for translational control, such as repression, and often genes involved in growth and development have inefficient or "difficult" translational start sites (e.g. long leader sequences, upstream ATG codons), presumably adding in another level of regulatory control (Jansen *et al.*, 1995). Another possibility is post-translational processing of the protein such as cleavage: the program SIGCLEAVE (GCG package at UK-HGMP Resource Centre) identified a potential signal sequence cleavage site (von Heijne, 1986) at amino acids 358-371, which if functional, would result in a protein of 41.2 kDa.

4.2.3 Chromosomal localisation of *Nfe2l1*

Evidence for a gene's role is compiled from many various sources. The following sections examine the information generated by this thesis and by other researchers.

4.2.3.1 Mouse

The initial localisation of *Nfe2l1* to mouse chromosome 11 (D-E region) by our collaborator Dr. M.G.Mattei, was by radioactive *in situ* hybridisation (section 3.3.1). The cDNA clone, cnc16LG, was used as the probe, and it also detected two other distinct loci on mouse chromosomes 7 (D1-F1 region) and 2 (E4-G region). The significance of these two other localisations is unknown, but it suggests the presence of either *Nfe2l1* pseudogenes or *cnc*/*Nfe2l1*-related genes. The same cDNA probe also detected multiple bands on a genomic Southern of *Eco* RI-digested XY human cell line DNA, although not in the mouse genomic DNA lane (Figure 3.1d). The absence of extra murine genomic bands cannot be explained by differences in washing stringencies (0.5x SSC, 0.1% SDS at 65°C), and perhaps repeating the experiment at a lower washing stringency may resolve the findings. Clone cnc16LG contains DNA sequence from the CNC5' domain onwards to the poly(A) tail, and therefore covers the regions most conserved between human and mouse *Nfe2l1* and *Drosophila cnc.* It also encodes the 3' UTR, which is well conserved between mouse and human (Figure 4.4). Luna *et al.* (1995) determined that the DNA encoding the last two-thirds of the 3'UTR of *TCF11*, and not the *cnc*-related coding sequence, was responsible for the complex hybridisation pattern seen with human genomic DNA (Figure 3.1d). This suggests that the additional human bands are due either to *TCF11* pseudogenes and/or the highly

conserved 3'UTR blocks of sequence (Figure 4.4), present in other unknown genes. If the signals detected by clone cnc16LG on mouse chromosomes 2 and 7 are due to the 3'UTR, as suggested by the findings of Luna *et al.*, 1995, this would suggest the presence of *Nfe2l1* pseudogenes and/or unidentified genes containing the conserved 3'UTR blocks of sequence and not other *cnc*-related genes.

The original localisation of murine *Nfe2l1*, to mouse chromosome 11D-E region, opened up two different avenues of research: by utilising known mouse/human homology of synteny groupings (Buchberg *et al.*, 1989), it allowed for the rapid isolation of a human genomic homologous DNA fragment and its subsequent localisation by FISH in the human genome (section 3.3.3). Secondly, by examining the mouse mapping information available for the D-E region of mouse chromosome 11, it was possible to consider *Nfe2l1* as a candidate gene for any mapped mouse mutants in the region.

The Mouse Chromosome Atlas (Lyon and Kirby, 1992) for chromosome 11 was consulted: a more recent example of this map is shown below (Figure 4.5). In the D-E region of chromosome 11, numerous mutants had been localised, including *Ts* (tail-short), *cod* (cerebellar outflow degeneration), *tn* (teetering), *js* (Jackson shaker) and *bsk* (bare skinned). Given the putative role of *cnc* in *Drosophila* head development, those mutants with neurological and head abnormalities were of particular interest (Table 4.2). During the mapping experiments, a murine *Nfe2l1*-related gene was cloned, *p45-Nfe2* (Andrews *et al.*, 1993a; section 4.3) which was thought to be essential for globin gene expression. Therefore mutants with erythropoietic abnormalities were also of interest: tail-short mapped close to *Nfe2l1* (Table 4.2), although the prenatal anaemia described for this mutant may only be secondary to growth retardation (Brotherton *et al.*, 1979).

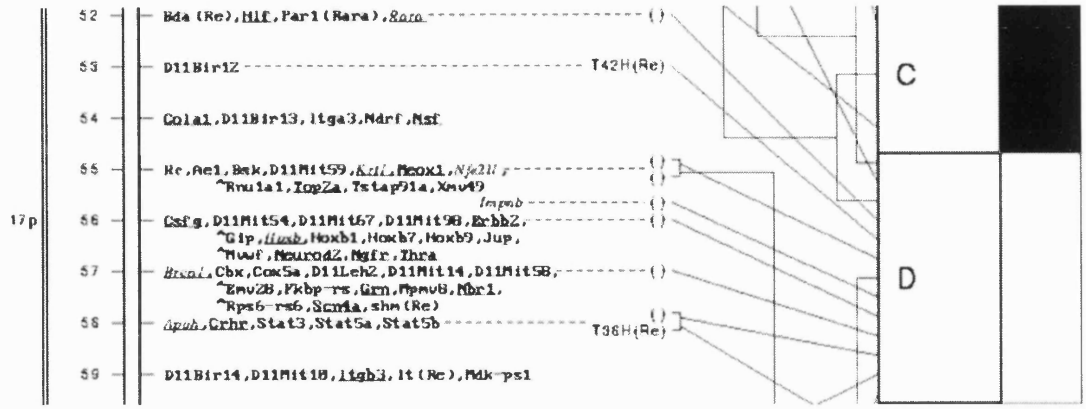


Figure 4.5 Mouse Chromosome Atlas for chromosome 11
This map was obtained from the MRC Mammalian Genetics Unit, Harwell, UK (http://www.mgu.har.mrc.ac.uk). The segment containing the Nfe2l1 localisation is shown as an example. Known human homology of synteny groupings are shown on the right hand side of the diagram, followed by the genetic distance (in cM) of the mapped mouse genes and mutants/phenotypes.

Symbol	Name	Cytogenetic position	Map position (cM)	Features	Reference
co	cocked	11B5-D	46	General abnormalities of the inner ear. Head tossing and abnormal position.	Peterson and Biddle, 1970
oe	open eyelids	11B5-D	46	Absent eyelids; micropthalmia; cloudy cornea	Mackensen, 1960
shm	shambling	11D-E1	57	Ataxia; short; vacuolated lymphocytes	Green, 1967
Ts	tail-short	11D-E1	66	Skeletal abnormalities; prenatal anaemia	Morgan, 1950
cod	cerebellar outflow degeneration	11E1-TER	75	Ataxia; cerebellar degeneration	Sidman, 1967
tn	teetering	11E1-TER	77	Short; cerebral atrophy; abnormal pons/medulla/ganglia	Lane and Green, 1962
js	Jackson shaker	11E1-TER	78	Sensorineural deafness; circling	Dickie and Deol, 1967

Table 4.2 Mouse mutants in the MMU11D-E region
There are many mouse mutants in the 11D-E region, and a few examples with neurological or haematopoietic abnormalities are shown. The localisations are as recorded in the Mouse Chromosome Atlas in 1992 (Lyon and Kirby, 1992).

In order to confirm and further refine the map position of *Nfe2l1*, especially with respect to these mutants, an interspecific backcross provided by the Mouse Backcross facility at the UK-HGMP Resource Centre (Breen *et al.*, 1994), was screened and localised *Nfe2l1* relative to other chromosome genetic markers (section 3.3.2; McKie

and Scambler, 1996). A recent update (March 1998) on the EUCIB mapping data for *Nfe2l1*, shows zero recombination between *Nfe2l1* and two DNA markers, *D11Mit288* and *D11Mit263*, by haplotype analysis of 9 progeny. The map of chromosome 11, in the region surrounding *Nfe2l1*, has been obtained recently from both the MBx facility (Breen *et al.*, 1994) and the Mouse Genome Database (MGD) (Blake *et al.*, 1997). MBx positions *Nfe2l1* at 60.77 cM, close to *D11Mit160* (Figure 4.6a) and MGD positions *Nfe2l1* at 57.0 cM (Figure 4.6b).

The map position of mouse mutant *Ts* has also been further refined (Uchida *et al.*, 1996), which definitely excludes *Nfe2l1* as a candidate gene. The only nearby mapped mouse mutants are bald arthritic (*Bda*; Ferguson and Wallace, 1977), Rex (*Re*; Crew and Auerbach, 1939) and bare skin (*Bsk*; Lyon and Glenister, 1984): all are positioned at 58.0 cM on the MGD map, and all have skin and hair texture abnormalities. The ubiquitous expression of *Nfe2l1* during development would permit its involvement in a wide variety of developmental processes, and in particular its high level expression in the dermomyotome at 9-10 dpc, the precursor of dermis and skeletal muscle, could suggest an involvement in the differentiation of these tissues, making *Nfe2l1* a potential candidate for these hair and skin mutants. The phenotypes of null mutant *Nfe2l1* mice do not clarify the role, if any, of *Nfe2l1* in skin and hair development (section 4.2.5.3; Farmer *et al.*, 1997; Chan *et al.*, 1998). Further mapping studies should establish whether *Nfe2l1* is a candidate gene for these mutants.

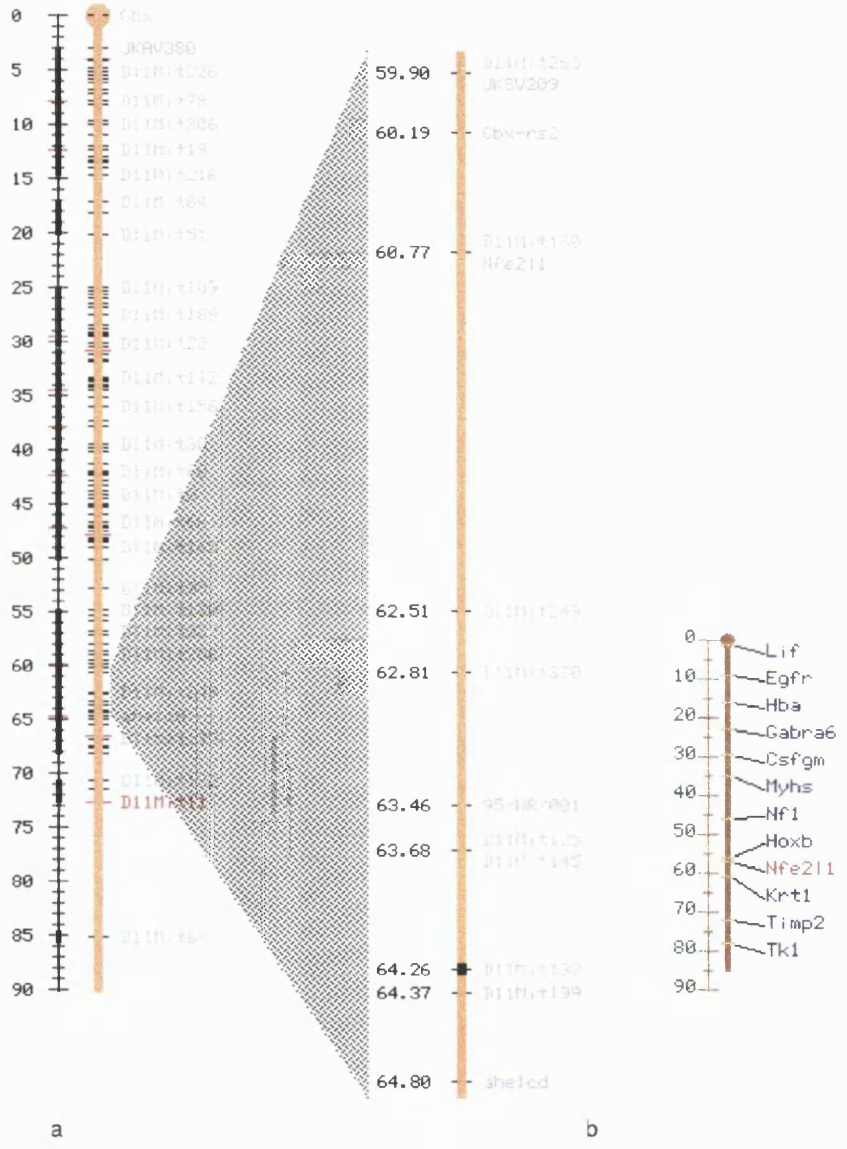


Figure 4.6 Maps of MMU11 in the vicinity of Nfe2l1

(a) MBx map around Nfe2l1. (b) MGD map showing Nfe2l1 in relation to other mapped genes on chromosome 11. Both the MGD and MBx mapping information are available on the World Wide Web (http:www.informatics.jax.org; http:www.hgmp.mrc.ac.uk/MBx/MBxHomepage.html).

4.2.3.2 Human

The human homologue for *Nfe2l1* was localised by FISH analysis to the middle of the long arm of chromosome 17 (section 3.3.3). Luna *et al.* (1994) mapped *TCF11* to 17q22, and further refined its chromosomal position (Luna *et al.*, 1995). Chan *et al.* (1995) also carried out FISH analysis of *NRF1* and localised it to 17q21.3. Various syndromes associated with mutations in the COL1A1 gene map to 17q21-22 (Pope *et al.*, 1985; Byers *et al.*, 1997), but there are no other mapped disease loci in the nearby vicinity.

Transcription factors have been implicated in numerous malignancies through translocation (Cleary, 1991; Cooper, 1996), either giving altered expression and/or through fusion proteins with novel functions (Gauwerky and Croce, 1993; Davis and Barr, 1997). Chromosome 17 is involved with numerous genetic changes associated with cancer, such as translocations (Mitelman, 1993; Van Roy *et al.*, 1997), loss of heterozygosity (Lindblom *et al.*, 1993) and amplifications (Kallioniemi *et al.*, 1994; Borresen *et al.*, 1990). The role, if any, of *NFE2L1* in malignancy is unknown, but its localisation is an essential step in defining any possible involvement. Luna *et al.* (1995) have mapped *TCF11* using a radiation hybrid panel of chromosome 17 (Abel *et al.*, 1993), and characterised a YAC containing the *TCF11* gene. The derived physical map of the YAC has shown that *TCF11* is approximately 120-145 kb distant from BTR, a novel sequence fused to *THRA1* in the BT474 breast cancer lines, as a consequence of a large deletion on chromosome 17q (Futreal *et al.*, 1994).

4.2.4 Expression pattern of *Nfe2l1*

4.2.4.1 Northern analysis

Nfe2l1 has a widespread expression: transcripts of 4.4 kb and 4.0 kb were detected initially by all *Nfe2l1* clones, in mouse and human fibroblast cell lines total and poly(A)$^+$ RNA (Figures 3.18, 3.19). Similarly sized transcripts were found in adult mouse tissues (brain, heart, kidney, liver and lung) and in staged whole mouse embryos (10.5 dpc- 15.5 dpc; Figure 3.20). The larger 4.4 kb transcript was predominant, and adult mouse brain had an extra 3.6 kb transcript, which was not detected in adult human brain (Figure 3.21). From the work by Luna *et al.* (1994, 1995) on human *NFE2L1*, the larger transcript uses the second polyadenylation signal to produce a 540 bp extended 3'UTR (section 4.2.2.2). The smaller, less abundant,

transcript uses the first polyadenylation signal. Multiple *Nfe2l1* mRNA isoforms have been found (Luna *et al.*, 1994) and possibly the 3.6 kb band detected in adult mouse brain is another alternative isoform. It was not detected in adult human brain (Figure 3.21), and so might be murine-specific or only be produced at levels undetectable by northern analysis in humans. Other bZIP proteins have numerous alternative isoforms that have quite individual function. For example, both CREB (Lee and Masson, 1993) and CREM proteins have splice variants that are tissue specific and developmentally regulated; different isoforms can either activate or repress transcription (Hurst, 1995). Also, a truncated isoform of FosB has been isolated that inhibits FOS/JUN transcriptional activity (Nakabeppu and Nathans, 1991).

The relative levels of *Nfe2l1* expression were determined in adult and embryonic mouse poly(A)$^+$ RNA (Tables 3.3, 3.4). In the adult, the most striking observation was the high level of expression in the heart: relative to brain, kidney and lung, the signal was approximately three and a half times stronger (notwithstanding tissue specific differences in β-actin level). These findings in the adult mouse tissues correlate with those found for human *NFE2L1* in adult human tissues (Chan *et al.*, 1993; Luna *et al.*, 1994). In the staged mouse embryos, there was a relatively steady increase in the level of expression through development: at 15.5 dpc, *Nfe2l1* expression was approximately twice as great as that detected in the 10.5 dpc mouse embryos. There is no striking developmental regulation of *Nfe2l1*, as detected by northern analysis. More detailed information on the precise sites of *Nfe2l1* expression in staged mouse embryos was sought, to establish any obvious developmental regulation and/or restricted expression, such as seen for *cnc*.

4.2.4.2 In situ hybridisation

No tightly restricted expression pattern is seen for *Nfe2l1* (section 3.4.2). There is some published information on murine *Nfe2l1* expression during early embryogenesis, undertaken as part of a project to generate a null mutant for *Nfe2l1* (Farmer *et al.*, 1997). *In situ* analyses showed widespread expression of *Nfe2l1* at 6.5 dpc, but RT-PCR analyses demonstrated no expression at 3.5 dpc: *Nfe2l1* is developmentally upregulated between day 3.5 and day 6.5 of mouse development.

At 7.5 dpc, there is a widespread low level of *Nfe2l1* expression, with some localisation of signal (Figure 3.23), and by 8.5 dpc the signal is ubiquitous and stronger (data not shown). Using whole-mount *in situ* hybridisation, expression in 9-10 dpc embryos appears fairly widespread: there is strong expression in the branchial arches (Figures 3.24, 3.25 and 3.26a) and in the dermamyotome component of the

somites (Figure 3.26b, c), whilst the developing brain, neural folds and neural tube have very low expression (Figures 3.24-3.27). The branchial arches are composed of mainly neural crest-derived mesenchyme and contribute to a number of structures including some facial bones and cartilage and the parathyroid and thymus glands. The significance of the high level expression of *Nfe2l1* in this tissue is unknown. Segmentation of the paraxial mesoderm gives rise the somites, which differentiate ventrally into sclerotome and dorsally into dermamyotome. The dermamyotome gives rise to the dermis and skeletal muscle, neither of which show particularly strong expression of *Nfe2l1*. The strong expression in the dermamyotome is distributed evenly, and so does not appear to correlate with neural crest cell migration (Le Dourain, 1982) or with the compartmentalisation of this structure (Marcelle *et al.*, 1997). Many signalling molecules are involved in patterning the somite such as Wnt family members, Bone morphogenetic protein-4 (BMP-4), noggin and Sonic Hedgehog (SHH) (Munsterberg *et al.*, 1995; Marcelle *et al.*, 1997; Tajbakhsh and Spörle, 1998). Transcription factors implicated in the differentiation of the somite include the bHLH MyoD family such as *MyoD* and *Myf5* (Cossu *et al.*, 1996; Ordahl and Williams, 1998), and the homeobox genes Mox1 and Mox2 (Candia *et al.*, 1992, 1996). It is possible that *Nfe2l1* may also play a role in differentiating the somite.

The expression of *Nfel1* was studied at later developmental stages using ^{35}S tissue section *in situ* hybridisation (Figures 3.28- 3.30). There is widespread expression throughout the embryos, and with increasing developmental stage, there appears to be upregulation of *Nfe2l1* at specific sites within the embryo. The developing heart appears to have a low level of *Nfe2l1* expression at 9- 10 dpc (Figures 3.24, 3.25), which is upregulated by 16.5 dpc (Figure 3.30g), consistent with the high levels detected in the adult heart.

The expression of *Nfe2l1* in the developing CNS becomes greater from 10.5 dpc onwards (Figures 3.28- 3.30). A component of the peripheral nervous system (PNS), the dorsal root ganglia (DRG), is strongly positive at 16.5 dpc (Figures 3.29, 3.30a, c). At all earlier stages examined, 9.5- 13.5 dpc, the DRG are positive for *Nfe2l1* but at no greater levels than that of the surrounding tissues (data not shown). This developmental upregulation of *Nfe2l1* would suggest a role for *Nfe2l1* in the DRG. The DRG are mainly neural crest cell derived, arising at 9 dpc in a rostro-caudal direction, and contain the cell bodies of the sensory neurones of both the somatic and autonomic nervous system, as well as satellite glial cells. The cell type positive for *Nfe2l1* is unknown, however the ventral roots appear to be positive also for *Nfe2l1*, suggesting a neuronal site of expression. Many neurotrophic factors are involved in the differentiation of the DRG such as neurotrophin-3 (NT-3), brain-derived

neurotrophic factor (BDNF) and fibroblast growth factor-2 (FGF2) (Murphy *et al.*, 1994; Davies, 1995). Neuronal subpopulations within the DRG differentiate at variable times throughout development (Kitao *et al.*, 1996), and the temporal upregulation of *Nfe2l1* expression may correlate with the differentiation of a specific type of neurone. By 16.5 dpc, some cranial ganglia appear strongly positive for *Nfe2l1* (Figure 3.29e,f). These are also part of the PNS, and so support the suggestion of a role for *Nfe2l1* in particular differentiative pathways in the developing PNS.

The eye in the 16.5 dpc embryo shows an interesting pattern of *Nfe2l1* expression, with relatively high levels seen in the inner nuclear layer of the retina as compared to the outer nuclear layer (Figure 3.30e). The neural retina is initially a simple pseudostratified columnar epithelium, which differentiates into a layered structure. The inner nuclear layer contains bipolar, horizontal and amacrine neuronal cells and Müller glial cells, whilst the outer nuclear layer contains photoreceptor cells. The pattern of *Nfe2l1* expression may reflect an involvement in the differentiation of the neural retina, and interestingly, *nrl*, a member of the large Maf family of bZIP transcription factors shows a restricted pattern of expression in the developing eye (section 1.4.3.2; Farjo *et al.*, 1993; Liu *et al.*, 1996). Many other transcription factors have been implicated in retinal differentiation including homeobox-containing genes such as *Pax6*, *Prox 1* and *Chx10* (section 1.3.2.2; Oliver *et al.*, 1993; Liu *et al.*, 1994).

In summary, there is widespread expression of *Nfe2l1* from early embryogenesis which would permit its involvement in a great number of developmental processes. In addition, there is a spatio-temporal upregulation of Nfe2l1 suggesting a more specific role in specific differentiative pathways. Further speculation on its role *in vivo* is discussed later in the context of published evidence.

4.2.5 Published information on NFE2L1

In recent years, the functional role of human and mouse NFE2L1 has been studied by various research groups, and this section summarises their findings.

4.2.5.1 Dimerisation studies with NFE2L1

There have been numerous *in vitro* interaction and binding site preference studies involving human NFE2L1 and its isoforms, and although the results are not conclusive,

they point to a pivotal role for NFE2L1 in a variety of cellular and developmental processes. Human NFE2L1 (Chan *et al.*, 1993; Luna *et al.*, 1994; Caterina *et al.*, 1994) has been shown to be unable to bind, or at best very weakly, to DNA as a homodimer (Moi *et al.*, 1994b; Toki *et al.*, 1997; Marini *et al.*, 1997), consistent with the predictive examination of the *Nfe2l1* leucine zipper motif (section 4.2.1.2). Instead, it preferentially interacts with members of the small Maf family of proteins (section 1.4.3.2; Johnsen *et al.*, 1996, 1998; Marini *et al.*, 1997; Moi *et al.*, 1994; Toki *et al.*, 1997).

I The Maf family

The Maf family is a subclass of the bZIP family of transcription factors composed of the large Maf proteins (section 1.4.3.2) which contain an activation domain, and the small Maf proteins: MAFK (Fujiwara *et al.*, 1993; Andrews *et al.*, 1993b; Igarashi *et al.*, 1995), MAFF (Fujiwara *et al.*, 1993), and MAFG (Kataoka *et al.*, 1995; Blank *et al.*, 1997; Marini *et al.*, 1997) which all lack an activation domain. They can form homodimers, intra-family heterodimers (Motohashi *et al.*, 1997) and inter-family heterodimers with FOS and JUN (Kataoka *et al.*, 1995; Francastel *et al.*, 1997). Maf homodimers or intra-family heterodimers recognise 13 or 14 bp elements, termed MAREs (Maf-responsive elements), of the TRE(12-*O*-tetradecanoylphorbol-13-acetate-responsive element) or CRE (cyclic AMP-responsive element) type [T-MARE(TGCTGA(C/G)-TCAGCA) and C-MARE(TGCTGACGTCAGCA) respectively].

II Interactions with NFE2L1

The small Maf proteins have all been shown to interact with NFE2L1, and its isoforms, (Johnsen *et al.*, 1996, 1998; Marini *et al.*, 1997; Moi *et al.*, 1994b; Toki *et al.*, 1997) and bind *in vitro* to specific DNA binding sites. The preferred heterodimer binding site contains consensus sequences for both Maf homodimer (T-MARE) and AP1 [(C/G)TCA(T/C)], and is identical to the NF-E2 site [(T/C)GCTGA(C/G)TCA(C/T)], implicated in the regulation of erythroid specific gene expression at the Locus Control Region (LCR) of the β-globin gene cluster (Mignotte *et al.*, 1989; Orkin, 1990). Two independent groups of researchers (Marini *et al.*, 1997; Toki *et al.*, 1997) demonstrated that homomers of NFE2L1 do not bind the T-MARE DNA sequence effectively, but require the presence of human small Maf proteins to bind efficiently and activate transcription at the NF-E2 site. This transactivation is lost as the amount of *MafG* expression vector is increased (Marini *et al.*, 1997).

In contrast, Johnsen and co-workers (Johnsen *et al.*, 1996, 1998) have shown that homomers of NFE2L1 (TCF11) can form multimeric complexes which bind DNA, if

with less specificity and affinity than heterodimers, and transactivate in a dose dependent manner. However, in agreement with Toki *et al.* (1997), and Marini *et al.* (1997), it was found that in the presence of small Maf proteins, there is a preference for heterodimer formation and an increased affinity for the NF-E2 binding site. The TCF11/MAFG heterodimer was shown to repress transactivation, and MAFG interfered in a dose dependent manner with the transactivation seen with TCF11 alone: this could be due to TCF11/MAFG heterodimer formation which lacks transactivation ability or to the formation of MAFG homodimers that compete for binding site access, and act as repressors at the NF-E2 site. However TCF11 preferentially forms heterodimers with MAFG *in vitro* (Marini *et al.*, 1997; Toki *et al.*, 1997; Johnsen *et al.*, 1996, 1998), and is therefore more likely to be responsible for the repression of transactivation. The discrepancy seen in heterodimerisation producing repression versus transactivation may possibly be accounted for by the use of chicken MAFG (Johnsen *et al.*, 1998) as oppose to human MAFG (94% identity at protein level; Marini *et al.*, 1997; Toki *et al.*, 1997). Johnsen *et al.*, 1998, also highlight that the slight increase in reporter activity seen falls well within (Marini *et al.*, 1997), or just outside (Toki *et al.*, 1998), the range of errors for the experiment. Attempts by this group to simulate the conditions used by Toki *et al.* (1997) have failed to repeat the reported transactivation. Of interest though, is a recent report of another CNC-related protein, p45-NFE2 that heterodimerises with MAFK (p18 NF-E2), causing transactivation initially, then switching to repression with only a 4-fold difference in abundance of the small Maf protein (Nagai *et al.*, 1998): their study suggested that MafK may function as an active transcription repressor. The choice of NFE2L1 isoform for these interaction studies might affect its transactivation ability, and this issue has not yet been addressed: TCF11 contains an extra 30 amino acids, N-terminal to the bZIP motif (Luna *et al.*, 1994; Johnsen *et al.*, 1998) relative to NRF1 (Chan *et al.*, 1993; Marini *et al.*, 1997).

4.2.5.2 Potential NF-E2 binding sites

The binding site preferences that have been identified for NFE2L1 and its heterodimers, with the small Maf proteins, have been mentioned briefly above. A detailed examination of binding site preference, utilising a PCR-assisted approach of cloning binding sites following *in vitro* selection of degenerate oligonucleotides (Kataoka *et al.*, 1994b), showed that 86% of examined sites were the consensus sequence 5'-TGCTgaGTCAT-3' (Johnsen *et al.*, 1998). This is identical to the NF-E2 site (Mignotte *et al.*, 1989), the antioxidant response element (ARE; Rushmore *et al.*, 1991) and the heme response element (HRE; Inamdar *et al.*, 1996). These elements have been found in numerous genes, which can be roughly divided into three categories

that are co-regulated in response to a specific signal: antioxidant response [e.g. NAD(P)H:quinone oxidoreductase (NQO_1); Venugopal and Jaiswal, 1996], heme biosynthesis (e.g. porphobilinogen deaminase; Mignotte *et al.*, 1989) and erythroid differentiation (β-globin gene cluster; Ney *et al.*, 1990). However the definitive identification of functional target genes for NFE2L1 is essential for further understanding of its role *in vivo.*

4.2.5.3 Knockouts for Nfe2l1

Two null mutants of *Nfe2l1* have been produced, each with a unique phenotype (Farmer *et al.*, 1997; Chan *et al.*, 1998).

I The bZIP transcription factor LCR-F1 is essential for mesoderm formation in mouse development (Farmer *et al.*, 1997)

Lcrf1 mutant (-/-) embryos developed normally to the late egg cylinder stage with no apparent growth retardation, but development was arrested at approximately 6.5 dpc. *Lcrf1* null embryos failed to form primitive streak and were unable to produce mesoderm. This demonstrated that *Lcrf1* is essential for gastrulation in the mouse, and is involved in the regulation of mesoderm formation. *Lcrf1* expression is developmentally upregulated: RT-PCR failed to detect *Lcrf1* expression either in wild type ES cells, or in normal 3.5 dpc blastocysts, but widespread high level expression was demonstrated by *in situ* hybridisation in tissue sections of 6.5 dpc wild type embryos. This expression pattern is consistent with a role for *Lcrf1* in early mouse development.

II Targeted disruption of the ubiquitous CNC-bZIP transcription factor, Nrf-1, results in anemia and embryonic lethality in mice (Chan *et al.*, 1998)

Nrf1 mutant (-/-) embryos suffer from anemia as a result of abnormal fetal liver erythropoiesis and die at mid- to late-gestation. No abnormalities were seen before 10.5 dpc, and, by comparison with normal littermates, null mutants became progressively growth retarded and anemic, dying from 10.5 dpc onwards. The growth retardation was seen prior to the anemia, possibly reflecting a general role for *Nrf-1* in cellular metabolism or proliferation. Abnormal maturation of haematopoietic progenitor cells in the fetal liver microenviroment caused anemia and, putatively, death from the resultant hypoxia. *Nrf1* mutant (-/-) embryos were otherwise normal, and so *Nrf1* is not essential for early development and organogenesis.

III Comparison of the *Nfe2l1* null mutant experiments

Both papers report the essential nature of NFE2L1 for mouse development, but describe two quite different null mutant phenotypes. Both have also shown that the defect is not cell autonomous as ES (-/-) cells can contribute to all cell types in chimaeras, and that globin gene expression is unaffected. Strain-specific differences in null mutant phenotypes have been reported before (Sibilia and Wagner, 1995; Threadgill *et al.*, 1995), but preliminary data (Farmer *et al.*, 1997; Chan *et al.*, 1998) suggests that this is not the case with the *Nfe2l1* knockouts. Further breeding experiments should answer this possibility. The mutations were generated using two different constructs: Farmer *et al.* (1997) used a replacement vector that removed from exon 3a to close to the end of exon 6 (Figure 4.3') and the intervening genomic sequence, whilst Chan *et al.* ,1997, inserted a neomycin cassette in the opposite transcriptional orientation into exon 6, N-terminal to the bZIP domain. The former approach could interrupt the sequence or regulatory elements of another gene embedded in an intron of *Nfe2l1* (Cawthon *et al.*, 1990), or have some position effect on a gene more distant (Olson *et al.*, 1996). Of note, there are several CpG islands in the genomic region surrounding human *NFE2L1* (Luna *et al.*, 1995). Analysis of the human genomic sequence encoding the *NFE2L1* region, recently deposited in GenBank (accession no. AC004477), may yield useful information on neighbouring/overlapping gene content. The intron/exon structure of *Nfe2l1* is well conserved between man and mouse and the gene is contained in a well conserved homology of synteny grouping, so that flanking human genes might well have murine correlates (Chan *et al.*, 1995). The "Nrf1" construct allowed transcription of the *Nfe2l1* gene, producing a 7 kb transcript, as detected by northern analysis, but containing the inverted neomycin cassette that introduced translational stop codons. There was no evidence for a *Nfe2l1* translation product, normal or truncated, as detected by Western blotting. RT-PCR detected only mutant transcript in poly(A)$^+$ RNA from 12.5-15.5 dpc null mutant embryos. A variable phenotype for the same knockout can be due to a leaky mutation (Dorin *et al.*, 1994), but Chan *et al.* (1998) have concluded that this is improbable. Similar to the "Lcrf1" construct, the introduction of the "Nrf1" construct itself may alter the phenotype of the null mutant embryos. Chan *et al.* (1998) noted that their non-targeted ES cells expressed *Nfe2l1,* but it was not detected by Farmer *et al.* (1997) in their non-targeted ES cells.

4.3 THE CNC FAMILY- A NOVEL SUBCLASS OF THE bZIP FAMILY OF TRANSCRIPTION FACTORS

This section describes the family of *Drosophila cnc*-related genes (Caterina *et al.*, 1994), as reported in the literature in recent years. The role of *Nfe2l1* in this context will be discussed, along with a comparison with other family members.

4.3.1 Isolation of the vertebrate founding member, p45-NFE2

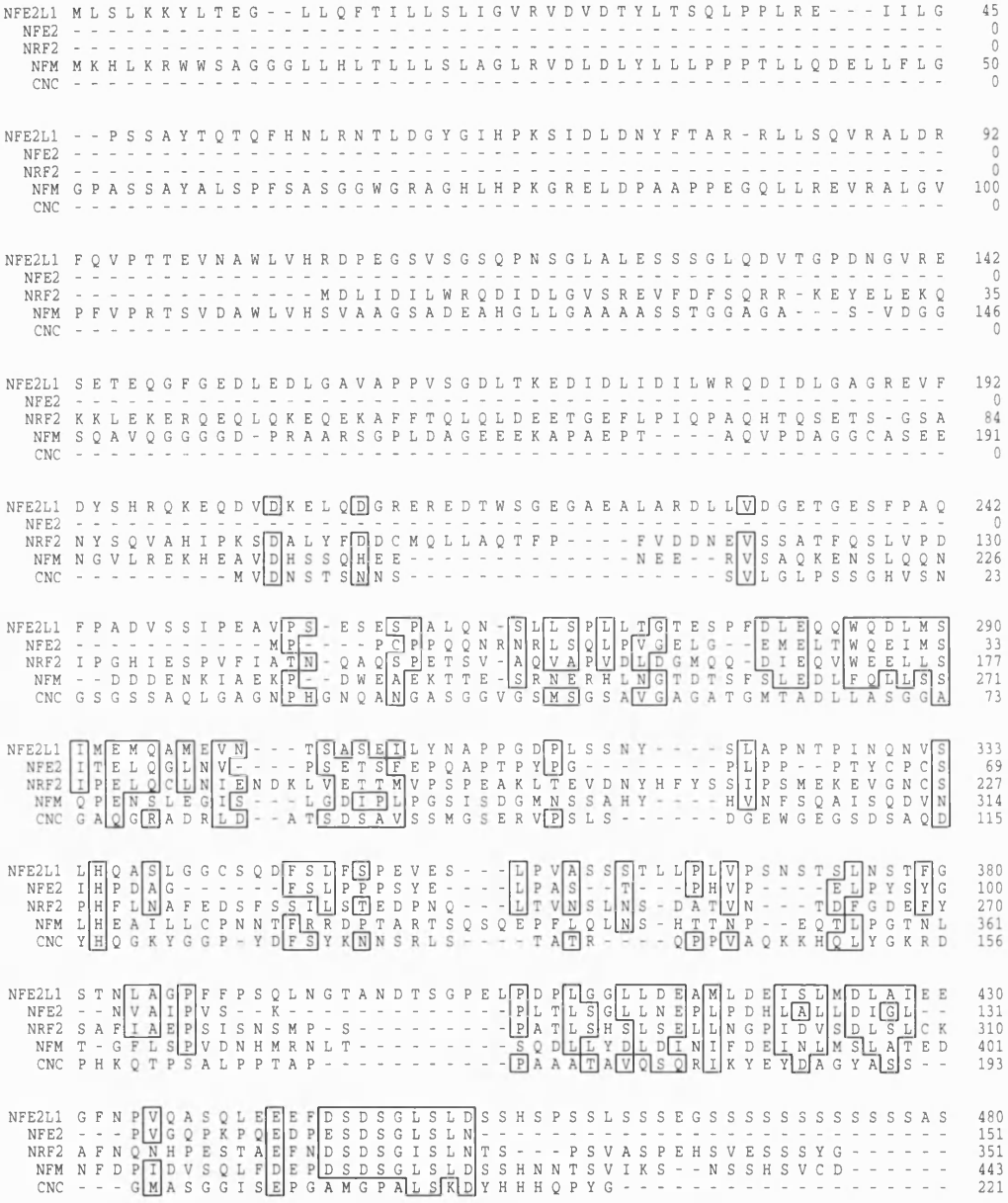
The β-globin gene cluster has upstream regulatory sequences, originally identified for their DNase I hypersensitivity, which are individually called hypersensitive sites (5'-HS1 to HS4), and collectively referred to as the locus control region (LCR). The LCR contains numerous regulatory elements such as enhancers, silencers, putative insulators and origins of replication. The sequences of the specific DNA elements responsible for these activities has been elucidated, and one such sequence, the NFE2 (nuclear factor erythroid 2)/AP1-like motif is found duplicated in the core of the HS2 enhancer. This element is required for high level expression of the β-globin gene cluster 50 kb downstream (Grosveld *et al.*, 1987; Collis *et al.*, 1990), and is recognised by AP1 (Lee *et al.*, 1987), CREB (Gonzalez *et al.*, 1989) and the CNC family. Early functional analyses suggested that an erythroid-specific factor, termed NF-E2, was responsible for activity at the HS2 site (Ney *et al.*, 1990). Subsequently the tissue-specific component of NF-E2 was cloned, *p45-Nfe2*, (Andrews *et al.*, 1993a) and was found to encode a novel bZIP protein with strong similarity to *Drosophila cnc*. The family as a whole and the individual members are discussed in the following sections.

4.3.2 p45-NFE2, NFE2L1 and NRF2

4.3.2.1 General comments

The three proteins most closely related to *Drosophila* CNC are p45-NFE2, NFE2L1 and NRF2. A clustal alignment (Clustal W program at UK-HGMP Resource Centre) of these four proteins and the novel family member, NFM (section 3.5.4), demonstrates the shared regions of similarity in and around the bZIP motif (NFE2L1 $_{577\text{-}722}$; Figure 4.7),

termed the CNC domains (Figure 4.1). It also reveals other conserved regions towards the N-terminal of the vertebrate proteins, such as an 11 amino acid stretch (NFE2L1 $_{442-452}$, ExxDSDSG(I/L)SL): it overlaps with the C-terminal end of the acidic domain of NFE2L1, and is just N-terminal to the serine repeat region of NFE2L1. It is 100% conserved between human and mouse in all three proteins. It has no matches in the PROSITE database, but presumably has a conserved function, possibly as an unknown site for post-translational modification or in protein-protein interactions. A degree of relatedness in the zipper region can also be established for all four proteins, assuming *cnc* as the ancestral gene: cnc> NFE2L1> NRF2> p45-NFE2 (Moi *et al.*, 1994a). The genomic organisation of the three vertebrate genes is also similar, with a large terminal exon containing the bZIP domain and the 3'UTR (Moi *et al.*, 1994a; Luna *et al.*, 1995).



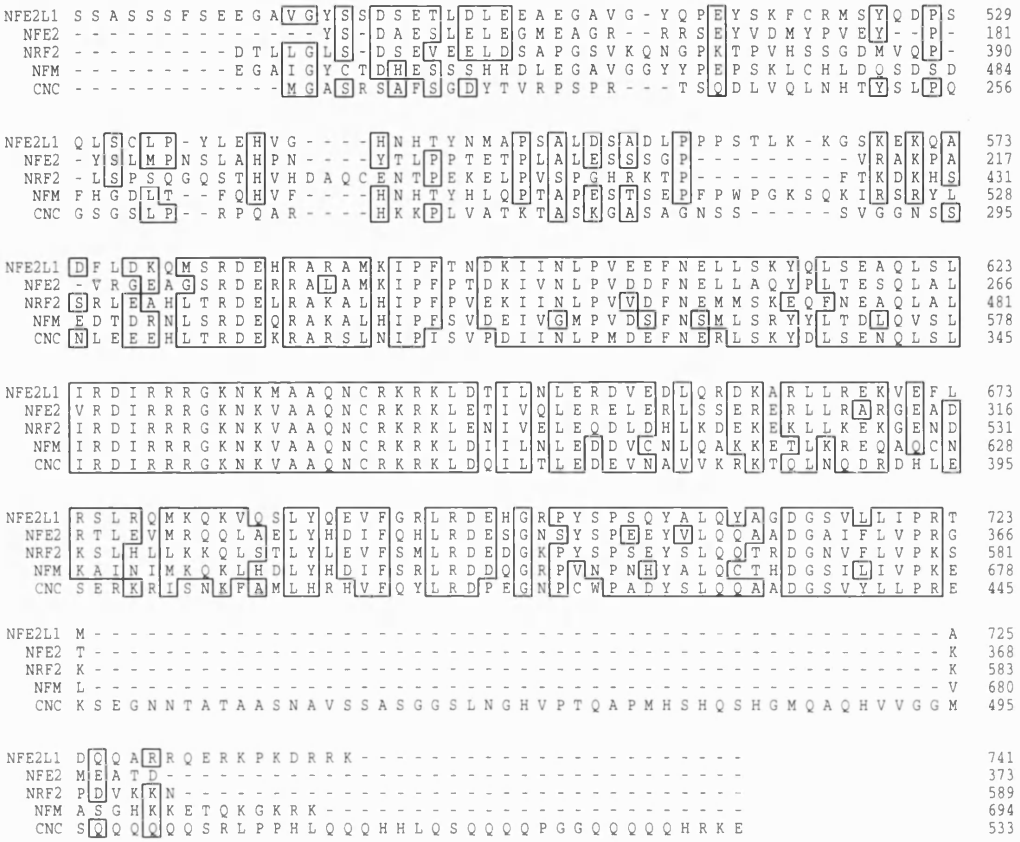


Figure 4.7 Clustal W alignment of selected CNC family members

The program Clustal W (Thompson et al., 1994) was used to generate the multiple alignment of CNC, murine NFE2L1, p45-NFE2, human NRF2 and NFM. The PRETTYPLOT program (UK-HGMP Resource Centre) boxed identical and similar amino acids.

4.3.2.2 *Localisation*

The chromosomal localisations of *p45-NFE2* (Ney *et al.*, 1993; Chan *et al.*, 1995), *NFE2L1* (section 3.3.3.2; Luna *et al.*, 1994; Chan *et al.*, 1995) and *NRF2* (Chan *et al.*, 1995) have been determined (Table 4.3). Comparison of their sequences (Figure 4.7) and localisations suggest that they have arisen through gene duplication and divergence from a single ancestral gene: they all map into a homology of synteny grouping containing, amongst others, the Hox clusters (McGinnis and Krumlauf, 1992), collagen (Pope *et al.*, 1985) and integrin family members. The well-characterised Hox family have four clusters of paralogous genes (section 1.3.1), and so it could be predicted that there would be a fourth member of the CNC family on human chromosome 7 beside the Hox cluster: this has been confirmed by the isolation of the novel family member, NFM (section 3.5.4). All of the human localisations are in areas that have been associated with cancer (Mitelman and Heim, 1988), however there is no evidence for involvement for the CNC family in cancer or any other disease so far. Murine *p45-Nfe2* was mapped to chromosome 15, near the microcytosis (*mk*) locus (Peters *et al.*, 1993): homozygous *mk* mice have severe hypochromic microcytic anaemia and *p45-Nfe2* was therefore an attractive candidate gene for the mutant. The *mk* allele had a single base pair change in *p45-Nfe2*, resulting in a valine to alanine substitution at amino acid 173 (Peters *et al.*, 1993). However the base pair change was subsequently found in normal inbred BALB/c mice, and so was not a causative mutation but a polymorphism (Lu *et al.*, 1994; Peters *et al.*, 1994).

Gene	Human	Mouse
Nfe2l1	17q21-22	11DE
p45-NFE2	12q13.1-13.3	15
NRF2	2q31	ND
NFM	7p15	ND

Table 4.3 Chromosomal localisations of the CNC family
ND = not determined

4.3.2.3 *p45-NFE2*

Mouse and human *p45-NFE2* have been isolated, and the 1.6 kb mRNA encodes a putative protein with an ORF of 373 amino acids, and a predicted molecular weight of

41 kDa. An alternative human mRNA isoform has been isolated with an alternative promoter and 5'UTR: it is developmentally regulated (Pischedda *et al.*, 1995),. Human *p45-NFE2* is 89% identical and 94% similar to murine *p45-Nfe2.* Expression of *p45-NFE2* is limited to haematopoietic cells of the erythroid, megakaryocytic and mast cell lineages (Andrews *et al.*, 1993a) and intestinal epithelia (Peters *et al.*, 1993). *In vitro* translation generates two products of around 45 kDa (Andrews *et al.*, 1993a; Chan *et al.*, 1994). p45-NFE2 is the tissue-specific component of NFE2 (section 4.3.1), and interacts with a ubiquitous component, p-18 NFE2 (also known as MAFK; Andrews *et al.*, 1993b; Igarashi *et al.*, 1994) to form the activity . Surprisingly, *p45-Nfe2* null mutant mice showed very little effect in their erythroid cell lineages, but instead failed to produce platelets, secondary to a maturational arrest in the megakaryocyte lineage, and typically died as a result of haemorrhage (Shivdasani *et al.*, 1995; Shivdasani and Orkin, 1995). p45-NFE2 interacts with several of the small Maf proteins, although it shows a preference for binding MAFK: p45-NFE2-MAFK heterodimers (NFE2 activity) can transactivate at the HS2 site to regulate β-globin gene expression (Kotkow and Orkin, 1995; Amrolia *et al.*, 1997; Bean and Ney, 1997). p45-NFE2-MAFK heterodimers have been reported to bind to chromatin *in vitro* and disrupt nucleosomal structure (Armstrong and Emerson, 1996). p45-NFE2 has also been shown to interact with the thyroid hormone (T3R) and retinoic acid receptors (RAR) to serve as a potent co-activator of these nuclear hormone receptors (Cheng *et al.*, 1997).

4.3.2.4 NRF2

Human NRF2 was isolated from an erythroleukaemia expression library by binding site recognition of tandem NFE2/AP1-like repeats (Moi *et al.*, 1994a), and murine *Nrf2* was isolated by cross-species conservation (Chui *et al.*, 1995; Chan *et al.*, 1996). Chicken *Ech* was isolated from an erythroid cell cDNA library by screening with murine *p45-Nfe2* (Itoh *et al.*, 1995): sequence comparison suggests that is the avian homologue of *NRF2.* The human *NRF2* gene encodes a 2.3 kb transcript, containing an ORF of 589 amino acids with a predicted molecular mass of 66.1 kDa. A schematic of the derived protein domains of NRF2 is shown below (Figure 4.8). Both mouse and human NRF2 are highly homologous (82% identical, 85% similar). Although different from the acidic domain of NFE2L1, NRF2 has a glutamic and aspartic acid-rich N-terminal region that could function as an activation domain (Mitchell and Tjian, 1989), and *in vitro* transactivation analysis has demonstrated that the N-terminal half of the protein contains a powerful activation domain (Moi *et al.*, 1994a). The similarity with the other members of the CNC family is limited essentially to the bZIP motif and flanking CNC domains (Figure 4.7). Analysis of the bZIP region shows

that it is likely to be an obligate heterodimer (Moi *et al.*, 1994): *in vitro* studies have demonstrated interactions between NRF2 and MAFK (Marini *et al.*, 1997), and between ECH and avian small Maf proteins, MAFK, MAFG, and MAFF (Itoh *et al.*, 1995). NRF2 is expressed ubiquitously in human adult tissues and cell lines (Moi *et al.*, 1994a), and mouse adult and foetal tissues (Chan *et al.*, 1996). *Nrf2* null mutant mice showed that NRF2 was not essential for mouse development, as Nrf2(-/-) mice developed normally, had normal blood indices, reached adulthood and could reproduce (Chan *et al.*, 1996; Itoh *et al.*, 1997). However, examination of the regulation of phase II detoxifying enzymes in the *Nrf2* null mutant mice revealed that *Nrf2* (probably as a heterodimer with a small Maf protein) is essential for the transcriptional induction of phase II enzymes (Itoh *et al.*, 1997), confirming previous *in vitro* studies (Venugopal and Jaiswal, 1996).

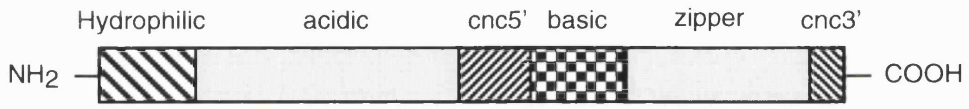


Figure 4.8 Schematic of NRF2
The localisation of known protein domains is shown. At the N-terminus is a hydrophilic stretch of 40 aa, of mainly basic and acidic residues. This is followed by a region that contains an activation domain. The C-terminal half contains the CNC and bZIP motifs.

4.3.3 Other *cnc*-related members

4.3.3.1 The Bach genes

Using a Gal4-MAFK fusion protein in a yeast two-hybrid screen, two novel members of the CNC family were isolated, called *Bach1* and *Bach2* (Oyake *et al.*, 1996). Their CNC5' and bZIP regions are related, although not as strongly, to those of the CNC family; there is no similarity with the CNC family elsewhere, but they are highly related to each other. A splice variant of *Bach2* has also been cloned (Oyake *et al.*, 1996). Both BACH1 and BACH2 contain a protein interaction motif, called variously a BTB (broad-complex, tramtrack, bric a brac) or POZ (pox and zinc finger) domain (Albagli *et al.*, 1995) and BTB-containing proteins can regulate chromatin structure (Dorn *et al.*, 1993; Farkas *et al.*, 1994). Both BACH proteins can bind *in vitro* to NF-E2 sites by forming heterodimers with MAFK. Both act repressively in transfection assays using fibroblast cells, but BACH1 activates transcription in cultured erythroid cells, whilst BACH2 acts as a repressor. Therefore the BACH proteins could act

bifunctionally to regulate transcription and alter chromatin, such as happens at the β-globin LCR in differentiating erythroid cells.

4.3.3.2 skn-1

The isolation of the *C.elegans* gene, *skinhead* (*skn-1*), resulted from a genetic screen for developmental mutants (Bowerman *et al.*, 1992). In *skn-1* mutants, specific cells early in *C.elegans* development, destined to become pharyngeal, intestinal and muscular cells, switch pathways to that of hypodermal differentiation. The maternally expressed gene encodes a 2.9 kb mRNA, with an ORF of 533 amino acids and a predicted molecular weight of 61 kDa. SKN-1 protein is largely hydrophilic, and is rich in serine and threonine residues. At its C-terminus is a basic DNA-binding domain, moderately similar to that of the CNC family. However there is no adjacent leucine zipper or dimerisation motif. Random sequence selection methods have been used to show that SKN-1 binds DNA as a monomer, using a stretch of 85 aa (SKN-1 domain) that includes the basic DNA-binding domain, in a fashion similar to that of homeodomain proteins (Blackwell *et al.*, 1994). Recent work on the three dimensional structure of SKN-1, in solution and upon DNA binding, points to a role for the region just N-terminal to the basic DNA-binding domain in modulating the affinity and specificity of the DNA binding by bZIP proteins (Carroll *et al.*, 1997; Pal *et al.*, 1997). However this N-terminal region shows no similarity at the amino acid level to the CNC5' domain.

4.4 NOVEL FAMILY MEMBER -*NFM*

4.4.1 Isolation of a novel CNC family member

NFM was discovered by screening the sequence databases (GenBank and dbEST) for moderate matches to the NFE2L1 protein sequence (section 3.5.4). Several human *NFM*-related ESTs were identified and used to screen a human placental cDNA library. Sequence analysis of the ESTs and the cDNA clones confirmed *NFM* to be a *Nfe2l1*-related gene, and to be a novel family member of the CNC family of bZIP transcription factors. This thesis contains only the preliminary characterisation of *NFM* as the work is still ongoing: partial sequence for *NFM*, from the ESTs and the human placental cDNAs, has been obtained and analysed. Recently the sequence of a human BAC clone, containing *NFM*, has been deposited into the GenBank database (accession no. AC004520), as part of a collaborative chromosome 7 genomic sequencing project (Dubuque *et al.*, 1998). The human BAC clone RG119C02 contains a putative protein sequence and predicted exon/intron structure for the coding sequence of a gene identical to the partial sequence of *NFM* (Appendix 1). A combination of the data available for *NFM*, generated by this thesis and from the human BAC clone RG119C02, will be discussed in the following sections.

4.4.2 Sequence analysis of *NFM*

4.4.2.1 Characterisation of NFM cDNAs and ESTs

The derived gene structure of *NFM*, and the relative position of the *NFM* cDNAs and an IMAGE EST no. 587428 is shown (Figure 4.9). Its genomic structure is reminiscent of other CNC family members, with the CNC domains, bZIP motif and 3'UTR contained within one large terminal exon (Figure 4.4). The 5'UTR of *NFM* has not yet been determined, as both isolated cDNAs are not full length: northern analysis revealed a predominant *NFM* transcript of 2.8 kb and a fainter band at 3.4 kb, so approximately 300 bp of 5' sequence is missing (Figure 3.51).

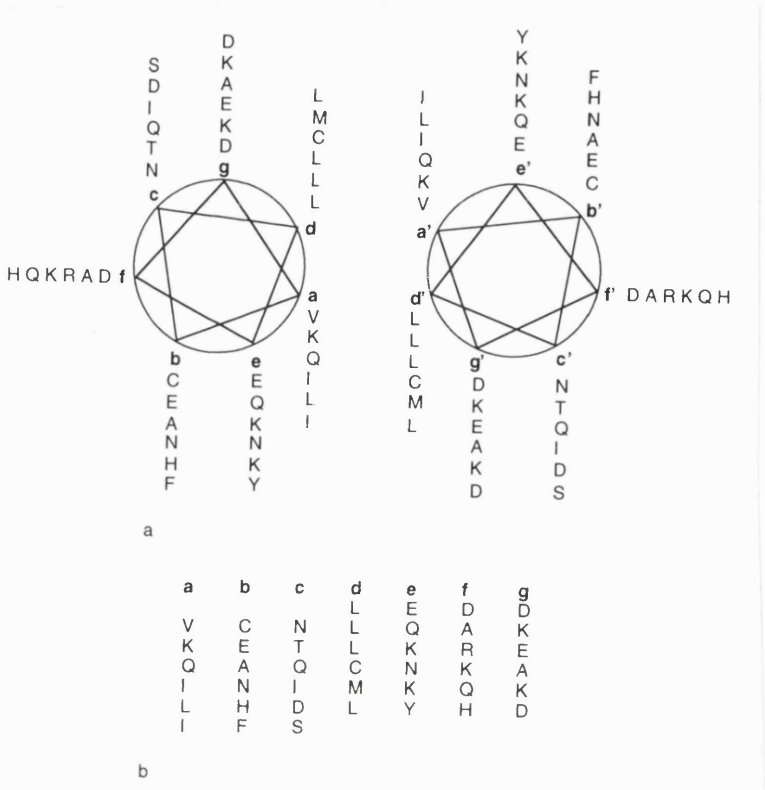


Figure 4.10 Helical wheel diagram of the NFM leucine zipper

(a) The bZIP domain of NFM is displayed as a helical wheel to demonstrate the relative position of the residues in the dimerisation region. (b) The bZIP domain is shown as a heptad repeat.

4.4.2.2 General features of NFM sequence

The potential coding sequence and 3'UTR of *NFM* and its predicted protein sequence were generated by combining the partial cDNA and EST analysis (section 3.5.4.2) and the genomic sequence (Appendix 2). The translation start site selected is in a good context (Kozak, 1996), and is followed by an ORF of 694 amino acids with a predicted molecular weight of 76 kDa. The two NFM mRNA transcripts, 2.8 kb and 3.4 kb, are generated by alternative polyadenylation. Both NFM cDNAs have different 3'UTRs, correlating with the EST contig data (THC207661 and THC210211; Figure 3.42), which by comparison with the genomic sequence, appear to arise from different polyadenylation signals (Figure 4.9). A similar situation is true for *Nfe2l1*, and its significance is discussed earlier (section 4.2.2.4).

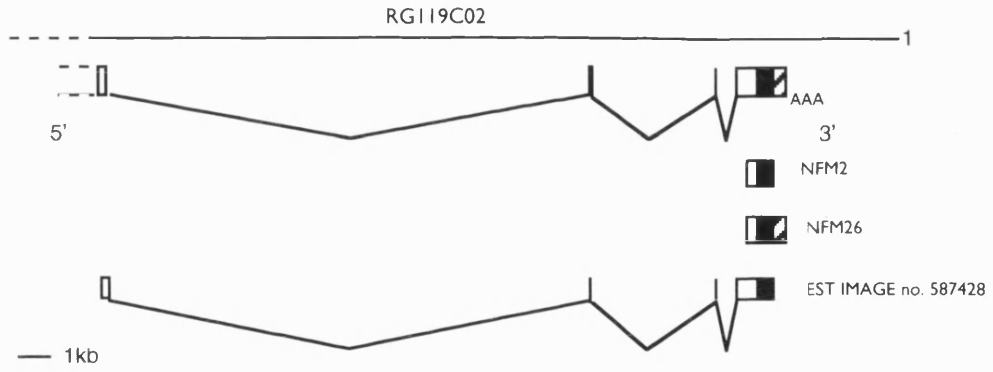


Figure 4.9 Derived genomic structure of NFM
The genomic structure of NFM was derived from the BAC RG119C02, and BESTFIT (HGMP-Resource Centre, UK) analysis of the NFM cDNA sequences. The IMAGE EST no. 587428 goes farthest 5', but is not full length.

4.4.2.3 The bZIP motif

The bZIP motif is located towards the C-terminus of NFM. The basic DNA-binding domain is well conserved (69% identical and 81% similar over CNC5' and basic DNA-binding domain with NFE2L1) with other CNC family members, and is followed by six heptad repeats of leucine and mainly hydrophobic residues of the leucine zipper. The leucine zipper dimerisation domain is most similar to other CNC family members and is represented as a helical wheel (Figure 4.10). Positions **a** and **d** are usually hydrophobic: charged residues will destabilise any dimerisation (Alber, 1992), but can concomitantly influence dimer specificity (O'Shea *et al.*, 1991). The presence of a cysteine amino acid in the **d** position is unusual, but such polar residues have been deemed acceptable, especially if occurring in the third or fourth heptad repeat (Hu *et al.*, 1990; van Heeckeren *et al.*, 1992). Other residues known to influence dimer

stability and specificity include the glutamine and lysine residues in position **a** (NFE2L1 has two basic amino acids in the same positions; Figure 4.2), and a number of charged residues in positions **e** and **g**.

4.4.2.4 Other recognised functional domains and sites

The program MOTIFS (GCG collection at UK-HGMP Resource Centre) identified several known motifs for post-translational modification in NFM (Appendix 4). Numerous potential phosphorylation sites were found, CK-2 sites being the commonest with nineteen in total. There were also six PKC sites and two cAMP- and cGMP-dependent protein kinase sites. Phosphorylation is an essential part of the regulation of transcription factor activity (section 4.2.1.2). Other potential post-translational modifications were found including twelve N-myristoylation sites, eleven N-glycosylation sites and one C-terminal amidation site. The conserved 11 aa stretch (ExxDSDSG(I/L)SL) of unknown function, noted in p45-NFE2, NRF2 and NFE2L1 (section 4.3.2.1), is also present in NFM. However the conservation is extended between NFM and NFE2L1 by 4 aa towards the C-terminus. Unlike NFE2L1, there is no striking acidic domain, but its acidic amino acid content is 15%.

4.4.3 Chromosomal localisation of *NFM* and its putative pseudogene

An initial attempt to localise *NFM* by PCR using the Genebridge 4 Radiation Hybrid DNA panel (Gyapay *et al.*, 1996) failed, as too many hybrid cell lines tested positive for *NFM* (section 3.5.4.2-I). The PCR primers were designed using 3'UTR sequence, which should be unique to NFM, and control PCR experiments showed them to be human-specific. Too many positives could reflect PCR contamination, but the PCR products were of the expected size and the results were repeatable and specific, even under varying PCR conditions (data not shown). Alternatively the radiation hybrid cell lines could contain an over-representation of the genomic region containing NFM, or the PCR primers were not unique for NFM. The presence within the genome of NFM pseudogenes is another explanation, which is supported by Southern analysis of NFM (section 3.5.4.3), and the attempted localisation of NFM by FISH (section 3.5.4.2-II).

The human BAC clone containing *NFM* (GenBank accession no. AC004520) has been localised to human chromosome 7p15. However two genomic clones for *NFM*, isolated from a human PAC library, localised to the telomere of a small chromosome, not

4.4.5 Strategies for isolating CNC family members

In addition to the *in silico* approach described above, alternative strategies were employed in an attempt to isolate other CNC family members (section 3.5). The degenerate PCR approach (section 3.5.1) failed to isolate any genes other than those used in the design of the PCR primers. Of note, the 5' primers chosen would not have been suitable to isolate *NFM*. The primers were contained within an exon (Figure 4.3') so genomic DNA should have been an adequate template, although cDNA generated by RT-PCR of poly(A)$^+$ RNA could also have been used. Low stringency screening of cDNA libraries using a conserved fragment of *Nfe2l1* such as the CNC domains could also have been attempted.

The attempt to isolate other *Drosophila* homologues of *Nfe2l1* resulted only in the isolation of *Troponin-T*, a false positive due to the presence of a trinucleotide repeat present in both genes (section 3.5.2). The homology is not very striking (87% identity over 61bp; Figure 3.36), given the strength of the signal from the library screening (Figure 3.33). It could have been avoided by using a *Nfe2l1* DNA probe that did not contain the repeat, and it is interesting that more false positives were not detected. This may be a reflection either of the frequency of the occurrence of this motif, and/or of the representation of the library screened. A zoo blot hybridised with cnc12SM-Xho I DNA fragment (Figure 3.10') detected bands in all species tested apart from rabbit and *Drosophila*. This probe is from a unique coding region for *Nfe2l1* and it is unexpected that there is no rabbit homologue for *Nfe2l1*. The quality of the rabbit genomic DNA could be tested using other DNA probes known to be conserved in rabbit. The signal in the human lane is smeared, however previous Southern analyses using cnc12SM, from which the Xho I fragment is derived, have not suggested the presence of a repeat. It is therefore likely that the human smear is of a non-specific nature which higher stringency washing would have removed.

The isolation by PCR of *Drosophila cnc* , to be used as a DNA probe, was unsuccessful (section 3.5.3).*cnc* should be present in the *Drosophila* cDNA used as template as it was derived from embryonic stage *Drosophila*, and *cnc* is expressed from early embryogenesis onwards (Mohler *et al.*, 1991). The cDNA template should be of sufficient quality as it has been used successfully to isolate other *Drosophila* genes by PCR (R. Llevadot, personal communication). Different primers and PCR conditions may have overcome the problem, but as an EST for *cnc* became publicly available (Harvey *et al.*,1997), it became unnecessary to resolve this issue. Using the *cnc* EST as a probe, a zoo blot showed no bands in any lanes apart from *Drosophila* (Figure 3.39). This probe covered the majority of the gene, so hopefully it would detect the presence of a mammalian homologue. However failure to do so does not exclude the possibility of the existence of a mammalian *cnc*, as the level of conservation might be greater at the protein level.

chromosome 7 (Figure 3.49). The PACs might be chimaeric for another locus and only contain a minimal portion of *NFM,* and so give the FISH result obtained. However PACs are less likely to be chimaeric than other vectors such as YACs (Cohen *et al.*, 1993), and both PACs were non-identical as determined by their restriction endonuclease pattern. They could contain a closely related sequence of *NFM*, but Southern blot analysis of the RE-digested PACs gave a very strong signal with an NFM-positive EST (EST509564) under high stringency washing conditions (Figure 3.48). Alternatively they could contain a *NFM* pseudogene, with a genuine NFM-positive PAC not being found during the library screen.

By comparing the expected band pattern (as calculated by restriction endonuclease mapping of the BAC clone RG119C02) with the results of the zoo blot and the PAC blot (Figures 3.50 and 3.48 respectively), it was noticed that additional bands were present. These extra bands could represent either cross-hybridisation to closely-related sequences or a pseudogene(s).

A comparison of the neighbouring genes of NFM on chromosome 7p15 includes the HoxA cluster (section 4.3.2.2). This fourth family member and its localisation were predicted after previous mapping of the other CNC family members (Chan *et al.*, 1995).

4.4.4 Comparison with other members of the CNC family

BLAST analysis of the full length protein sequence of NFM suggests that it is most similar to NFE2L1. The conservation in the basic DNA-binding region suggests that it will bind to similar binding sites as the other family members. The preliminary expression analysis suggests that it is not as ubiquitous as *Nfe2l1* or *NRF2*, possibly pointing to a more specialised function. Also the representation of ESTs for *NFM* in the dbEST is less than that for the other family members, and no murine ESTs have matched the *NFM* sequence so far.

4.5 FUTURE WORK

This section is divided into three parts: initially topics with specific relevance for this thesis are considered. The next section assesses the direction for current research for the CNC family, and finally, potential approaches for the identification of downstream targets of the CNC family are examined .

4.5.1 Characterisation of NFM

The conflicting FISH localisation result obtained with the two NFM-positive PACs should be investigated. The presence of a pseudogene in the PACs could be detected by using PCR primers designed either to connect two exons or to be intronic: pseudogenes are most commonly intronless. The exon-connection primers should generate different sized products from NFM and its pseudogene. Southern analysis using a UTR-specific probe may also determine if the PACs contain a closely related gene, and comparison of band sizes generated by different NFM-specific probes should also help, as the genomic structure of NFM is known. The human BAC clone is available from Research Genetics Inc. and so its localisation to human chromosome 7p15 could be confirmed by FISH analysis. Alternatively, the exon-connection PCR primers (designed to characterise the PAC clones) could be used to amplify the Genebridge 4 Radiation Hybrid panel cell line DNA, as a specific band for *NFM* should be detected. The full length sequence for *NFM* must be determined and analysed. The deposition of the genomic region should allow analysis of the promoter region for potential transcription factor binding sites. The precise sites of NFM expression, and any evidence for developmental regulation should be examined. This would be greatly facilitated by the isolation of a murine homologue for *NFM*, possibly by low stringency screening of a murine cDNA library with an *NFM*-specific probe from the coding region. Functional studies concerning binding site preferences, interaction partners and a null mutant mouse for *NFM*, similar to those undertaken for other CNC family members should be carried out.

4.5.2 The CNC family of bZIP transcription factors

4.5.2.1 *The Nfe2l1 knockouts*

The differences in phenotype seen (section 4.1.5.3; Farmer *et al.*, 1997; Chan *et al.*, 1998) might be resolved by breeding onto different genetic backgrounds, although initial results from both groups suggested this would be unlikely. The genomic sequence containing *NFE2L1* has been deposited (GenBank accession no. AC004477), and its analysis may infer the presence of embedded genes or control elements. However more direct evidence would come from analysis of the relevant murine genomic region. Alternatively, a transgene rescue of the *Nfe2l1* mutant mice could be attempted, to demonstrate that the phenotypes seen are due to loss of functional NFE2L1. This has been carried out for other knockouts such as *Src* (Schwartzberg *et al.*, 1997) and utrophin-dystrophin deficient mice (Rafael *et al.*, 1998). Conditional gene targeting is a useful approach (Kühn *et al.*, 1995; Spencer, 1996), particularly for mutations that are embryonic lethal, like the *Nfe2l1* null mutant mice. Mutations can be produced that are lineage- or tissue-specific (Gu *et al.*, 1994; Tsien *et al.*, 1996; Chen *et al.*, 1998) and/or temporally-regulated (St. Onge *et al.*, 1996; Schwenk *et al.*, 1998). For example, if a conditional mutation could be produced that was active only after mesoderm formation, the consequences of the *Nfe2l1* mutation on liver function could be assessed. These protocols commonly rely on two systems: the Cre/*loxP* site-specific recombination system (Sauer and Henderson, 1989; Schwenk *et al.*, 1995) and the tetracycline-controlled system for gene activation (Gossen and Bujard, 1992; Kistner *et al.*, 1996).

4.5.2.2 *Functional redundancy amongst the CNC family*

All four proteins, NFE2L1, p45-NFE2, NRF2 and NFM are well conserved in their shared domains of homology, including the bZIP motif (Figure 4.7). This would suggest both similar DNA-binding site preferences and choice of dimerisation partner, as has been found to be the case for the members so far characterised (Igarashi *et al.*, 1994; Johnsen *et al.*, 1996, 1998; Blank *et al.*, 1997, Marini *et al.*, 1997; Toki *et al.*, 1997). NFE2L1, p45-NFE2 and NRF2 can bind to the NF-E2/AP1-like sequences *in vitro*, and it has been suggested that NFE2L1 is a functional homologue of p45-NFE2 for erythroid-specific β-globin gene expression (Johnsen *et al.*, 1996). In fact the precise factors that do bind *in vivo* at the LCR to regulate β-globin gene expression has become more uncertain as other bZIP proteins, such as BACH1 and 2 and the small Maf family, are shown to bind to the same NFE2/AP1-like sites *in vitro* (Igarashi *et al.*, 1994; Oyake *et al.*, 1996; Igarashi *et al.*, 1998). However distinct differences have been found amongst the CNC family in their choice and affinity for interaction partner, precise DNA-binding site preference and transactivation ability (Igarashi *et al.*,

1994; Chan *et al.*, 1997; Motohashi *et al.*, 1997; Johnsen *et al.*, 1998). Functional experiments such as the knockouts for *Nfe2l1*, *p45Nfe2* and *Nrf2* have failed to pinpoint the key player in globin gene expression, with functional compensation by the other two proteins being held responsible. This explanation can be tested by mating the null mutants to produce double or triple knockouts, as has been done before when functional redundancy is suspected (Lufkin *et al.*, 1993; Li *et al.*, 1993; Mendelson, *et al.*, 1994; Lohnes *et al.*, 1994). However the continued investigation of the null phenotypes generated for all the CNC family members, by examination of potential target gene function (Itoh *et al.*, 1997; Venugopal and Jaiswal, 1996) could still provide further functional clues. With the growing number of contenders for activity at the LCR, future work will be needed to isolate undiscovered interaction partners and splice variants, their precise binding site preferences and transactivational ability. The potential for interacting with other regulatory networks such the nuclear hormone receptors should be another research growth area (Cheng *et al.*, 1997).

4.5.2.3 Characterisation of functional domains for the CNC family

Domains other than the bZIP are important for transcription factor function. The 11aa stretch conserved between all four proteins may well be important and should be characterised, possibly by studying mutant isoforms. Other non-conserved regions are likely to be responsible for the individuality of each family member, and the importance of post-translational modification of the protein, especially phosphorylation, is paramount. This is exemplified by a recent paper describing a splice variant of *Nfe2l1* (Prieschl *et al.* ,1998), which is missing sequences encoding the serine-rich region and some C-terminal sequences (GenBank accession no. AF015881), in a manner reminiscent of a *Bach2* splice variant (Oyake *et al.*, 1996). The *in vitro* phosphorylation of the NFE2L1 isoform (and a N-terminal truncated NFE2L1 protein) by CKII strongly facilitated binding site interactions, which is also likely to be relevant *in vivo* (Prieschl *et al.*, 1998). However the different NFE2L1 isoforms were shown to have differing transcriptional activities. As there are multiple sites for post-translational modification, similar studies to those of Prieschl *et al.*, 1998, should be fruitful.

The significance of the CNC domains that flank the leucine zipper could be investigated by determining the tertiary structure of the proteins upon DNA binding, as has been done for SKN-1 (Carroll *et al.*, 1997; Pal *et al.*, 1997), and mutagenesis and domain swapping studies could help to further characterise these regions.

4.5.3 Identification of functional transcription factor targets

Whatever the findings of *in vitro* studies, their biological relevance must still be proven. The importance of identifying biologically relevant targets is essential as often a protein with certain *in vitro* characteristics, is different *in vivo.* Numerous potential binding sites (binding site preferences) for the CNC family have been determined by a variety of methods. A large number were identified for NFE2L1 using a degenerate PCR-based oligonucleotide strategy, involving repeated rounds of binding, selection and amplification, followed by cloning and analysis of the specifically enriched oligonucleotides (Johnsen *et al.*, 1998). Subsequent searching of the databases [EMBL and EPD (eukaryotic promoter database)] identified potential gene targets. Electrophoretic mobility shift (EMSA) (Johnsen *et al.*, 1996; Igarashi *et al.*, 1998) and transactivation assays (Caterina *et al.*, 1994) can help to confirm these findings. However all three techniques are limited, especially if there are multiple competitors for any potential binding sites, as is the case for the NF-E2/AP1 like site. Other approaches have been developed to augment the identification of possible downstream targets, and by collating and validating information generated in a variety of systems, biologically relevant target molecules will be identified. Targets of Hox and ETS family members have been identified through differential gene expression, promoted by overexpression of the transcription factor (Feinstein *et al.*, 1995; Robinson *et al.*, 1997). However the genes identified in these studies may be direct or indirect targets of the transcription factors. To obtain evidence for a direct control, methods such as whole genome PCR (el-Deiry *et al.*, 1992; Matsugi *et al.*, 1995; Robinson *et al.*, 1997) and chromatin precipitation (Gould *et al.*, 1990) have successfully identified direct downstream targets for the ETS and Pax gene families respectively (Phelps and Dressler, 1996; Robinson *et al.*, 1997). Potential *in vivo* targets can be validated by using other techniques such as RNA differential display (Robinson *et al.*, 1997), and by other *in vitro* methods such as EMSA and gel mobility supershift assays.

4.5.4 Conclusion

The original premise of this thesis was to determine if the murine homologue of *Drosophila cnc* had been cloned, and to investigate its role in development. *Nfe2l1* is not the murine homologue of *cnc*, but it is a member of a well conserved novel bZIP family. The extended CNC family are likely to play crucial roles in numerous pathways including erythroid differentiation, antioxidant response and heme biosynthesis, and to interface with other transcription factor networks including the Fos and Maf bZIP

families and the nuclear hormone receptor network, as well as playing a potential involvement in chromatin regulation. *cnc* appears to have a clearly defined role in *Drosophila* head development as a segment-specific selector gene. In contrast *Nfe2l1* and the CNC family are clearly embroiled in a panoply of transcriptional regulatory networks involving post-translational modifications, multiple protein isoforms, differential affinities and binding site preferences, competition for target sites, and interactions with assisting co-factors. The dissection of the functional specificities for each CNC family member is an exciting challenge for the future.

REFERENCES

Abel, K.J., Boehnke, M., Prahalad, M., Ho, P., Flejter, W.L., Watkins, M., VanderStoep, J., Chandrasekharappa, S.C., Collins, F.S., Glover, T.W., and Weber, B.L. (1993). A radiation hybrid map of the BRCA1 region of the chromosome 17q12-q21. *Genomics* **17**: 632-641.

Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., and White, O. (1995). Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. *Nature* **377**: 3-174.

Albagli, O., Dhordain, P., Deweindt, C., Lecocq, G., and Leprince, D. (1995). The BTB/POZ domain: a new protein-protein interaction motif common to DNA- and actin-binding proteins. *Cell Growth. Differ.* **6**: 1193-1198.

Alber, T. (1992). Structure of the leucine zipper. *Curr. Opin. Genet. Dev.* **2**: 205-210.

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

Amrolia, P.J., Ramamurthy, L., Saluja, D., Tanese, N., Jane, S.M., and Cunningham, J.M. (1997). The activation domain of the enhancer binding protein p45NF-E2 interacts with TAFII130 and mediates long-range activation of the alpha- and beta-globin gene loci in an erythroid cell line. *Proc. Natl. Acad. Sci. U. S. A.* **94**: 10051-10056.

Andrews, N.C., Erdjument-Bromage, H., Davidson, M.B., Tempst, P., and Orkin, S.H. (1993a). Erythroid transcription factor NF-E2 is a haematopoietic-specific basic-leucine zipper protein. *Nature* **362**: 722-728.

Andrews, N.C., Kotkow, K.J., Ney, P.A., Erdjument-Bromage, H., Tempst, P., and Orkin, S.H. (1993b). The ubiquitous subunit of erythroid transcription factor NF-E2 is a small basic-leucine zipper protein related to the v-maf oncogene. *Proc. Natl. Acad. Sci. U. S. A.* **90**: 11488-11492.

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.

Apte, A.N. and Siebert, P.D. (1993). Anchor-ligated cDNA libraries: a technique for generating a cDNA library for the immediate cloning of the 5' ends of mRNAs. *Biotechniques.* **15**: 890-893.

Armstrong, J.A. and Emerson, B.M. (1996). NF-E2 disrupts chromatin structure at human beta-globin locus control region hypersensitive site 2 in vitro. *Mol. Cell Biol.* **16**: 5634-5644.

Asher, J.H.,Jr., Harrison, R.W., Morell, R., Carey, M.L., and Friedman, T.B. (1996). Effects of Pax3 modifier genes on craniofacial morphology, pigmentation, and viability: a murine model of Waardenburg syndrome variation. *Genomics.* **34**: 285-298.

Auffray, C. and Rougeon, F. (1980). Purification of mouse immunoglobulin heavy-chain messenger RNAs from total myeloma tumor RNA. *Eur. J. Biochem.* **107**: 303-314.

Augustine, K. (1997). Antisense approaches for investigating mechanisms of abnormal development. *Mutat. Res.* **396**: 175-193.

Avner, P., Amar, L., Dandolo, L., and Guenet, J.L. (1988). Genetic analysis of the mouse using interspecific crosses. *Trends. Genet.* **4**: 18-23.

Baldini, A. and Lindsay, E.A. (1994). Mapping human YAC clones by fluorescence in situ hybridization using Alu-PCR from single yeast colonies. *Methods Mol. Biol.* **33**: 75-84.

Ballabio, A. (1993). The rise and fall of positional cloning?. *Nat. Genet.* **3**: 277-279.

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

Bannister, A.J., Brown, H.J., Sutherland, J.A., and Kouzarides, T. (1994). Phosphorylation of the c-Fos and c-Jun HOB1 motif stimulates its activation capacity. *Nucleic Acids Res.* **22**: 5173-5176.

Bateson, W. (1894). Materials for the Study of Variation. (New York: Macmillan).

Bean, T.L. and Ney, P.A. (1997). Multiple regions of p45 NF-E2 are required for beta-globin gene expression in erythroid cells. *Nucleic Acids Res.* **25**: 2509-2515.

Belyavsky, A., Vinogradova, T., and Rajewsky, K. (1989). PCR-based cDNA library construction: general cDNA libraries at the level of a few cells. *Nucleic Acids Res.* **17**: 2919-2932.

Benton, W.D. and Davis, R.W. (1977). Screening lambdagt recombinant clones by hybridization to single plaques in situ. *Science* **196**: 180-182.

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

Blackwell, T.K., Bowerman, B., Priess, J.R., and Weintraub, H. (1994). Formation of a monomeric DNA binding domain by Skn-1 bZIP and homeodomain elements. *Science* **266**: 621-628.

Bladt, F., Riethmacher, D., Isenmann, S., Aguzzi, A., and Birchmeier, C. (1995). Essential role for the c-met receptor in the migration of myogenic precursor cells into the limb bud. *Nature* **376**: 768-771.

Blake, J.A., Richardson, J.E., Davisson, M.T., and Eppig, J.T. (1997). The Mouse Genome Database (MGD). A comprehensive public resource of genetic, phenotypic and genomic data. The Mouse Genome Informatics Group. *Nucleic Acids Res.* **25**: 85-91.

Blank, V., Kim, M.J., and Andrews, N.C. (1997). Human MafG is a functional partner for p45 NF-E2 in activating globin gene expression. *Blood.* **89**: 3925-3935.

Bober, E., Franz, T., Arnold, H.H., Gruss, P., and Tremblay, P. (1994). Pax-3 is required for the development of limb muscles: a possible role for the migration of dermomyotomal muscle progenitor cells. *Development.* **120**: 603-612.

Bopp, D., Burri, M., Baumgartner, S., Frigerio, G., and Noll, M. (1986). Conservation of a large protein domain in the segmentation gene paired and in functionally related genes of Drosophila. *Cell* **47**: 1033-1040.

Borresen, A.L., Ottestad, L., Gaustad, A., Andersen, T.I., Heikkila, R., Jahnsen, T., Tveit, K.M., and Nesland, J.M. (1990). Amplification and protein over-expression of the neu/HER-2/c-erbB-2 protooncogene in human breast carcinomas: relationship to loss of gene sequences on chromosome 17, family history and prognosis. *Br. J. Cancer* **62**: 585-590.

Bowerman, B., Eaton, B.A., and Priess, J.R. (1992). skn-1, a maternally expressed gene required to specify the fate of ventral blastomeres in the early C. elegans embryo. *Cell* **68**: 1061-1075.

Brandon, E.P., Idzerda, R.L., and McKnight, G.S. (1995). Knockouts. Targeting the mouse genome: a compendium of knockouts (Part I). *Curr. Biol.* **5**: 625-634.

Breen *et al.* European Backcross Collaborative Group.(1994). Towards high resolution maps of the mouse and human genomes--a facility for ordering markers to 0.1 cM resolution. *Hum. Mol. Genet.* **3**: 621-627.

Brotherton, T.W., Chui, D.H., McFarland, E.C., and Russell, E.S. (1979). Fetal erythropoiesis and hemoglobin ontogeny in tail-short (Ts/+) mutant mice. *Blood.* **54**: 673-683.

Buchberg, A.M., Brownell, E., Nagata, S., Jenkins, N.A., and Copeland, N.G. (1989). A comprehensive genetic map of murine chromosome 11 reveals extensive linkage conservation between mouse and human. *Genetics* **122**: 153-161.

Burri, M., Tromvoukis, Y., Bopp, D., Frigerio, G., and Noll, M. (1989). Conservation of the paired domain in metazoans and its structure in three isolated human genes. *EMBO J.* **8**: 1183-1190.

Byers, P.H., Duvic, M., Atkinson, M., Robinow, M., Smith, L.T., Krane, S.M., Greally, M.T., Ludman, M., Matalon, R., Pauker, S., Quanbeck, D., and Schwarze, U. (1997). Ehlers-Danlos syndrome type VIIA and VIIB result from splice-junction mutations or genomic deletions that involve exon 6 in the COL1A1 and COL1A2 genes of type I collagen. *Am. J. Med. Genet.* **72**: 94-105.

Camper, S.A., Saunders, T.L., Katz, R.W., and Reeves, R.H. (1990). The Pit-1 transcription factor gene is a candidate for the murine Snell dwarf mutation. *Genomics.* **8**: 586-590.

Candia, A.F., Hu, J., Crosby, J., Lalley, P.A., Noden, D., Nadeau, J.H., and Wright, C.V. (1992). Mox-1 and Mox-2 define a novel homeobox gene subfamily and are differentially expressed during early mesodermal patterning in mouse embryos. *Development.* **116**: 1123-1136.

Candia, A.F. and Wright, C.V. (1996). Differential localization of Mox-1 and Mox-2 proteins indicates distinct roles during development. *Int. J. Dev. Biol.* **40**: 1179-1184.

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

Carpenter, E.M., Goddard, J.M., Chisaka, O., Manley, N.R., and Capecchi, M.R. (1993). Loss of Hox-A1 (Hox-1.6) function results in the reorganization of the murine hindbrain. *Development.* **118**: 1063-1075.

Carroll, A.S., Gilbert, D.E., Liu, X., Cheung, J.W., Michnowicz, J.E., Wagner, G., Ellenberger, T.E., and Blackwell, T.K. (1997). SKN-1 domain folding and basic region monomer stabilization upon DNA binding. *Genes. Dev.* **11**: 2227-2238.

Carver, E.A. and Stubbs, L. (1997). Zooming in on the human-mouse comparative map: genome conservation re- examined on a high-resolution scale. *Genome Res.* **7**: 1123-1137.

Caterina, J.J., Donze, D., Sun, C.W., Ciavatta, D.J., and Townes, T.M. (1994). Cloning and functional characterization of LCR-F1: a bZIP transcription factor that activates erythroid-specific, human globin gene expression. *Nucleic Acids Res.* **22**: 2383-2391.

Caubet, J.F. (1989). c-fos proto-oncogene expression in the nervous system during mouse development. *Mol. Cell Biol* **9**: 2269-2272.

Cawthon, R.M., O'Connell, P., Buchberg, A.M., Viskochil, D., Weiss, R.B., Culver, M., Stevens, J., Jenkins, N.A., Copeland, N.G., and White, R. (1990). Identification and characterization of transcripts from the neurofibromatosis 1 region: the sequence and genomic structure of EVI2 and mapping of other transcripts. *Genomics.* **7**: 555-565.

Celli, G., LaRochelle, W.J., Mackem, S., Sharp, R., and Merlino, G. (1998). Soluble dominant-negative receptor uncovers essential roles for fibroblast growth factors in multi-organ induction and patterning. *EMBO J.* **17**: 1642-1655.

Chan, J.Y., Cheung, M.C., Moi, P., Chan, K., and Kan, Y.W. (1995). Chromosomal localization of the human NF-E2 family of bZIP transcription factors by fluorescence in situ hybridization. *Hum. Genet.* **95**: 265-269.

Chan, J.Y., Han, X.L., and Kan, Y.W. (1993). Cloning of Nrf1, an NF-E2-related transcription factor, by genetic selection in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **90**: 11371-11375.

Chan, J.Y., Kwong, M., Lu, R., Chang, J., Wang, B., Yen, T.S., and Kan, Y.W. (1998). Targeted disruption of the ubiquitous CNC-bZIP transcription factor, nrf-1, results in anemia and embryonic lethality in mice. *EMBO J.* **17**: 1779-1787.

Chan, K., Lu, R., Chang, J.C., and Kan, Y.W. (1996). NRF2, a member of the NFE2 family of transcription factors, is not essential for murine erythropoiesis, growth, and development. *Proc. Natl. Acad. Sci. U. S. A.* **93**: 13943-13948.

Chang, D.T., Lopez, A., von Kessler, D.P., Chiang, C., Simandl, B.K., Zhao, R., Seldin, M.F., Fallon, J.F., and Beachy, P.A. (1994). Products, genetic linkage and limb patterning activity of a murine hedgehog gene. *Development.* **120**: 3339-3353.

Chapman, V.M. and Nadeau, J.H. (1992). The mouse genome: an overview. *Curr. Opin. Genet. Dev.* **2**: 406-411.

Chen, C.Y. and Shyu, A.B. (1994). Selective degradation of early-response-gene mRNAs: functional analyses of sequence features of the AU-rich elements. *Mol. Cell Biol.* **14**: 8471-8482.

Chen, C-Y.A. and Shyu, A-B. (1995). AU-rich elements: characterization and importance in mRNA degradation. *Trends. Biochem. Sci.* **20**: 465-470.

Chen, J., Kubalak, S., and Chien, K. (1998). Ventricular muscle-restricted targeting of the RXR(alpha) gene reveals a non-cell-autonomous requirement in cardiac chamber morphogenesis. *Development.* **125**: 1943-1949.

Chen, R.H., Abate, C., and Blenis, J. (1993). Phosphorylation of the c-Fos transrepression domain by mitogen-activated protein kinase and 90-kDa ribosomal S6 kinase. *Proc. Natl. Acad. Sci. U. S. A.* **90**: 10952-10956.

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

Cheng, X., Reginato, M.J., Andrews, N.C., and Lazar, M.A. (1997). The transcriptional integrator CREB-binding protein mediates positive cross talk between nuclear hormone receptors and the hematopoietic bZip protein p45/NF-E2. *Mol. Cell Biol.* **17**: 1407-1416.

Chisaka, O., Musci, T.S., and Capecchi, M.R. (1992). Developmental defects of the ear, cranial nerves and hindbrain resulting from targeted disruption of the mouse homeobox gene Hox-1.6. *Nature* **355**: 516-520.

Chisholm, D. (1989). A convenient moderate-scale procedure for obtaining DNA from bacteriophage lambda. *Biotechniques.* **7**: 21-23.

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

Chowdhury, K. (1991). One step 'miniprep' method for the isolation of plasmid DNA. *Nucleic Acids Res.* **19**: 2792.

Chowdhury, K., Rohdewohld, H., and Gruss, P. (1988). Specific and ubiquitous expression of different Zn finger protein genes in the mouse. *Nucleic Acids Res.* **16**: 9995-10011.

Chui, D.H., Tang, W., and Orkin, S.H. (1995). cDNA cloning of murine Nrf 2 gene, coding for a p45 NF-E2 related transcription factor. *Biochem. Biophys. Res. Commun.* **209**: 40-46.

Clauss, I.M., Chu, M., Zhao, J.L., and Glimcher, L.H. (1996). The basic domain/leucine zipper protein hXBP-1 preferentially binds to and transactivates CRE-like sequences containing an ACGT core. *Nucleic Acids Res.* **24**: 1855-1864.

Cleary, M.L. (1991). Oncogenic conversion of transcription factors by chromosomal translocations. *Cell* **66**: 619-622.

Cohen, D., Chumakov, I., and Weissenbach, J. (1993). A first-generation physical map of the human genome. *Nature* **366**: 698-701.

Cohen, D.R. and Curran, T. (1990). Analysis of dimerization and DNA binding functions in Fos and Jun by domain-swapping: involvement of residues outside the leucine zipper/basic region. *Oncogene.* **5**: 929-939.

Collins, F.S. (1992). Positional cloning: let's not call it reverse anymore. *Nat. Genet.* **1**: 3-6.

Collins, F.S. (1995). Positional cloning moves from perditional to traditional. *Nat. Genet.* **9**: 347-350.

Collis, P., Antoniou, M., and Grosveld, F. (1990). Definition of the minimal requirements within the human beta-globin gene and the dominant control region for high level expression. *EMBO J.* **9**: 233-240.

Cooley, L., Kelley, R., and Spradling, A. (1988). Insertional mutagenesis of the Drosophila genome with single P elements. *Science* **239**: 1121-1128.

Cooper, C.S. (1996). Translocations in solid tumors. *Curr. Opin. Genet. Dev.* **6**: 71-75.

Copeland, N.G., Jenkins, N.A., Gilbert, D.J., Eppig, J.T., Maltais, L.J., Miller, J.C., Dietrich, W.F., Weaver, A., Lincoln, S.E., and Steen, R.G. (1993). A genetic linkage map of the mouse: current applications and future prospects. *Science* **262**: 57-66.

Copp, A.J. (1995). Death before birth: clues from gene knockouts and mutations. *Trends. Genet.* **11**: 87-93.

Copp, A.J., Brook, F.A., Estibeiro, J.P., Shum, A.S.W., and Cockfort, D.L. (1990). The embryonic development of neural tube defects. *Prog. Neurobiol.* **35**: 363-403.

Corcoran, L.M., Karvelas, M., Nossal, G.J., Ye, Z.S., Jacks, T., and Baltimore, D. (1993). Oct-2, although not required for early B-cell development, is critical for later B-cell maturation and for postnatal survival. *Genes. Dev.* **7**: 570-582.

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

Cossu, G., Tajbakhsh, S., and Buckingham, M. (1996). How is myogenesis initiated in the embryo? *Trends. Genet.* **12**: 218-223.

Courey, A.J., Holtzman, D.A., Jackson, S.P., and Tjian, R. (1989). Synergistic activation by the glutamine-rich domains of human transcription factor Sp1. *Cell* **59**: 827-836.

Cowell, I.G., Skinner, A., and Hurst, H.C. (1992). Transcriptional repression by a novel member of the bZIP family of transcription factors. *Mol. Cell. Biol.* **12**: 3070-3077.

Crew, F.A.E. and Auerbach, C. (1939). Rex: a dominant autosomal monogenic coat texture character in the mouse. *J. Genet.* **38**: 341-344.

Curran, T. and Franza, B.R.J. (1988). Fos and Jun: the AP-1 connection. *Cell* **55**: 395-397.

Curran, T., MacConnell, W.P., van Straaten, F., and Verma, I.M. (1983). Structure of the FBJ murine osteosarcoma virus genome: molecular cloning of its associated helper virus and the cellular homolog of the v-fos gene from mouse and human cells. *Mol. Cell Biol* **3**: 914-921.

Curtis, D., Lehmann, R., and Zamore, P.D. (1995). Translational regulation in development. *Cell* **81**: 171-178.

Dahl, E., Koseki, H., and Balling, R. (1997). Pax genes and organogenesis. *Bioessays* **19**: 755-765.

Davies, A.M. and Wright, E.M. (1995). Neurotrophic factors. Neurotrophin autocrine loops. *Curr. Biol.* **5**: 723-726.

Davis, R.J. and Barr, F.G. (1997). Fusion genes resulting from alternative chromosomal translocations are overexpressed by gene-specific mechanisms in alveolar rhabdomyosarcoma *Proc. Natl. Acad. Sci. U. S. A.* **94**: 8047-8051.

Dawid, I.B. and Sargent, T.D. (1988). Xenopus laevis in developmental and molecular biology. *Science* **240**: 1443-1448.

De Sauvage, F., Kruys, V., Marinx, O., Huez, G., and Octave, J.N. (1992). Alternative polyadenylation of the amyloid protein precursor mRNA regulates transcription. *EMBO J.* **11**: 3099-3103.

Deutsch, U., Dressler, G.R., and Gruss, P. (1988). Pax 1, a member of a paired box homologous murine gene family, is expressed in segmented structures during development. *Cell* **53**: 617-625.

Dickie, M.M. and Deol, M.S. (1967). Jackson shaker, js. *Mouse News Lett.* **36**: 39-30.

Dietrich, W.F., Miller, J., Steen, R., Merchant, M.A., Damron-Boles, D., Husain, Z., Dredge, R., Daly, M.J., Ingalls, K.A., and O'Connor, T.J. (1996). A comprehensive genetic map of the mouse genome .*Nature* **380**: 149-152.

Dolle, P., Lufkin, T., Krumlauf, R., Mark, M., Duboule, D., and Chambon, P. (1993). Local alterations of Krox-20 and Hox gene expression in the hindbrain suggest lack of rhombomeres 4 and 5 in homozygote null Hoxa-1 (Hox-1.6) mutant embryos. *Proc. Natl. Acad. Sci. U. S. A.* **90**: 7666-7670.

Dony, C. and Gruss, P. (1987). Proto-oncogene c-fos expression in growth regions of fetal bone and mesodermal web tissue. *Nature* **328**: 711-714.

Dorin, J.R., Stevenson, B.J., Fleming, S., Alton, E.W., Dickinson, P., and Porteous, D.J. (1994). Long-term survival of the exon 10 insertional cystic fibrosis mutant mouse is a consequence of low level residual wild-type Cftr gene. *Mamm. Genome* **5**: 465-472.

Dorn, R., Krauss, V., Reuter, G., and Saumweber, H. (1993). The enhancer of position-effect variegation of Drosophila, E(var)3-93D, codes for a chromatin protein containing a conserved domain common to several transcriptional regulators. *Proc. Natl. Acad. Sci. U. S. A.* **90**: 11376-11380.

Driever, W., Stemple, D., Schier, A., and Solnica-Krezel, L. (1994). Zebrafish: genetic tools for studying vertebrate development. *Trends. Genet.* **10**: 152-159.

Dubuque, T., Smith, A., Elliott, G., and Harmon, G. (1998). The sequence of Homo sapiens BAC clone RG119C02. Unpublished.

Dupin, E., Ziller, C., and Le Douarin, N.M. (1998). The avian embryo as a model in developmental studies: chimeras and in vitro clonal analysis. *Curr. Top. Dev. Biol.* **36**: 1-35.

Duret, L., Dorkeld, F., and Gautier, C. (1993). Strong conservation of non-coding sequences during vertebrates evolution: potential involvement in post-transcriptional regulation of gene expression. *Nucleic Acids Res.* **21**: 2315-2322.

Eichmann, A., Grapin-Botton, A., Kelly, L., Graf, T., Le Douarin, N.M., and Sieweke, M. (1997). The expression pattern of the mafB/kr gene in birds and mice reveals that the kreisler phenotype does not represent a null mutant. *Mech. Dev.* **65**: 111-122.

el-Deiry, W.S., Kern, S.E., Pietenpol, J.A., Kinzler, K.W., and Vogelstein, B. (1992). Definition of a consensus binding site for p53. *Nat. Genet.* **1**: 45-49.

Ellenberger, T. (1994). Getting a grip on DNA recognition: structures of the basic region leucine zipper, and the basic region helix-loop-helix DNA binding domains. *Curr. Opin. Struc. Biol.* **4**: 12-21.

Ellenberger, T.E., Brandl, C.J., Struhl, K., and Harrison, S.C. (1992). The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted alpha helices: crystal structure of the protein-DNA complex. *Cell* **71**: 1223-1237.

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

Epstein, J.A., Shapiro, D.A., Cheng, J., Lam, P.Y.P., and Maas, R.L. (1996). Pax3 modulates expression of the c-Met receptor during limb muscle development. *Proc. Natl. Acad. Sci. U. S. A.* **93**: 4213-4218.

Farjo, Q., Jackson, A.U., Junzhe, X., Gryzenia, M., Skolnick, C., Agarwal, N., and Swaroop, A. (1993). Molecular characterization of the murine neural retina leucine zipper gene, Nrl. *Genomics* **18**: 216-222.

Farkas, G., Gausz, J., Galloni, M., Reuter, G., Gyurkovics, H., and Karch, F. (1994). The Trithorax-like gene encodes the Drosophila GAGA factor. *Nature* **371**: 806-808.

Farmer, S.C., Sun, C.W., Winnier, G.E., Hogan, B.L., and Townes, T.M. (1997). The bZIP transcription factor LCR-F1 is essential for mesoderm formation in mouse development. *Genes. Dev.* **11**: 786-798.

Favor, J., Sandulache, R., Neuhauserklaus, A., Pretsch, W., Chatterjee, B., Senft, E., Wurst, W., Blanquet, V., Grimes, P., Sporle, R., and Schughart, K. (1996). The mouse pax2(1neu) mutation is identical to a human pax2 mutation in a family with renal-coloboma syndrome and results in developmental defects of the brain, ear, eye, and kidney. *Proc. Natl. Acad. Sci. U. S. A.* **93**: 13870-13875.

Feinberg, A.P. and Vogelstein, B. (1984). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **137**: 266-267.

Feinstein, P.G., Kornfeld, K., Hogness, D.S., and Mann, R.S. (1995). Identification of homeotic target genes in *Drosophila melanogaster* including *nervy*, a proto-oncogene homologue. *Genetics* **140**: 573-586.

Ferguson, J.M. and Wallace, M.E. (1977). Bald mice: arthritic? *Mouse News Lett.* **57**: 11-10.

Flaherty, L. (1998). Generation, identification, and recovery of mouse mutations [In Process Citation]. *Methods* **14**: 107-118.

Fortin, A.S., Underhill, D.A., and Gros, P. (1997). Reciprocal effect of waardenburg syndrome mutations on dna binding by the pax-3 paired domain and homeodomain. *Hum. Mol. Gen.* **6**: 1781-1790.

Francastel, C., Augery-Bourget, Y., Prenant, M., Walters, M., Martin, D.I., and Robert-Lezenes, J. (1997). c-Jun inhibits NF-E2 transcriptional activity in association with p18/maf in Friend erythroleukemia cells. *Oncogene.* **14**: 873-877.

Franz, T. (1993). The Splotch (Sp1H) and Splotch-delayed (Spd) alleles: differential phenotypic effects on neural crest and limb musculature. *Anat. Embryol. (Berl).* **187**: 371-377.

Franza, B.R.,Jr., Rauscher, F.J., Josephs, S.F., and Curran, T. (1988). The Fos complex and Fos-related antigens recognize sequence elements that contain AP-1 binding sites. *Science* **239**: 1150-1153.

Franza, B.R.,Jr., Sambucetti, L.C., Cohen, D.R., and Curran, T. (1987). Analysis of Fos protein complexes and Fos-related antigens by high- resolution two-dimensional gel electrophoresis. *Oncogene.* **1**: 213-221.

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

Friedrich, G. and Soriano, P. (1993). Insertional mutagenesis by retroviruses and promoter traps in embryonic stem cells. *Methods Enzymol.* **225**: 681-701.

Frigerio, G., Burri, M., Bopp, D., Baumgartner, S., and Noll, M. (1986). Structure of the segmentation gene paired and the Drosophila PRD gene set as part of a gene network. *Cell* **47**: 735-746.

Frohman, M.A., Dush, M.K., and Martin, G.R. (1988). Rapid production of full-length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. USA* **85**: 5227-5232.

Frohman, M.A., Martin, G.R., Cordes, S.P., Halamek, L.P., and Barsh, G.S. (1993). Altered rhombomere-specific gene expression and hyoid bone differentiation in the mouse segmentation mutant, kreisler (kr). *Development.* **117**: 925-936.

Fuentes, J.J., Pucharcos, C., Pritchard, M., and Estivill, X. (1997). Alu-splice PCR: a simple method to isolate exon-containing fragments from cloned human genomic DNA. *Hum. Genet.* **101**: 346-350.

Fujiwara, K.T., Kataoka, K., and Nishizawa, M. (1993). Two new members of the maf oncogene family, mafK and mafF, encode nuclear b-Zip proteins lacking putative trans-activator domain. *Oncogene.* **8**: 2371-2380.

Futreal, P.A., Cochran, C., Marks, J.R., Iglehart, J.D., Zimmerman, W., Barrett, J.C., and Wiseman, R.W. (1994). Mutation analysis of the THRA1 gene in breast cancer. *Cancer Res.* **54**: 1791-1794.

Gauwerky, C.E. and Croce, C.M. (1993). Chromosomal translocations in leukaemia. *Semin. Cancer Biol.* **4**: 333-340.

Gavel, Y. and von Heijne, G. (1990). Sequence differences between glycosylated and non-glycosylated Asn-X- Thr/Ser acceptor sites: implications for protein engineering. *Protein. Eng.* **3**: 433-442.

Gavin, B.J., McMahon, J.A., and McMahon, A.P. (1990). Expression of multiple novel Wnt-1/int-1-related genes during fetal and adult mouse development. *Genes. Dev.* **4**: 2319-2332.

Gehring, W.J., Affolter, M., and Burglin, T. (1994). Homeodomain proteins. *Annu. Rev. Biochem.* **63**: 487-526.

Gellon, G., Harding, K.W., McGinnis, N., Martin, M.M., and McGinnis, W. (1997). A genetic screen for modifiers of deformed homeotic function identifies novel genes required for head development. *Development* **124**: 3321-3331.

Gellon, G. and McGinnis, W. (1998). Shaping animal body plans in development and evolution by modulation of Hox expression patterns. *Bioessays* **20**: 116-125.

Gendron-Maguire, M., Mallo, M., Zhang, M., and Gridley, T. (1993). Hoxa-2 mutant mice exhibit homeotic transformation of skeletal elements derived from cranial neural crest. *Cell* **75**: 1317-1331.

Gitelman, I. (1997). Twist protein in mouse embryogenesis. *Dev. Biol.* **189**: 205-214.

Gius, D., Cao, X.M., Rauscher, F.J., Cohen, D.R., Curran, T., and Sukhatme, V.P. (1990). Transcriptional activation and repression by Fos are independent functions: the C terminus represses immediate-early gene expression via CArG elements. *Mol. Cell Biol.* **10**: 4243-4255.

Glass, D.B., el-Maghrabi, M.R., and Pilkis, S.J. (1986). Synthetic peptides corresponding to the site phosphorylated in 6- phosphofructo-2-kinase/fructose-2,6-bisphosphatase as substrates of cyclic nucleotide-dependent protein kinases. *J. Biol. Chem.* **261**: 2987-2993.

Goff, D.J. and Tabin, C.J. (1996). Hox mutations au naturel. *Nat. Genet.* **13**: 256-258.

Goldstein, R.S. and Kalcheim, C. (1991). Normal segmentation and size of the primary sympathetic ganglia depend upon the alternation of rostrocaudal properties of the somites. *Development.* **112**: 327-334.

Gonzalez, G.A., Yamamoto, K.K., Fischer, W.H., Karr, D., Menzel, P., Biggs, W., Vale, W.W., and Montminy, M.R. (1989). A cluster of phosphorylation sites on the cyclic AMP-regulated nuclear factor CREB predicted by its sequence. *Nature* **337**: 749-752.

Goodman, F.R., Mundlos, S., Muragaki, Y., Donnai, D., Giovannucci-Uzielli, M.L., Lapi, E., Majewski, F., McGaughran, J., McKeown, C., Reardon, W., Upton, J., Winter, R.M., Olsen, B.R., and Scambler, P.J. (1997). Synpolydactyly phenotypes correlate with size of expansions in HOXD13 polyalanine tract. *Proc. Natl. Acad. Sci. U. S. A.* **94**: 7458-7463.

Gossen, M. and Bujard, H. (1992). Tight control of gene expression in mammalian cells by tetracycline- responsive promoters. *Proc. Natl. Acad. Sci. U. S. A.* **89**: 5547-5551.

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

Gould, A.P., Brookman, J.J., Strutt, D.I., and White, R.A. (1990). Targets of homeotic gene control in Drosophila. *Nature* **348**: 308-312.

Goulding, M.D., Chalepakis, G., Deutsch, U., Erselius, J.R., and Gruss, P. (1991). *Pax-3*, a novel murine DNA-binding protein expressed during early embryogenesis. *Embo J.***10**: 1135-1147.

Graba, Y., Aragnol, D., and Pradel, J. (1997). Drosophila Hox complex downstream targets and the function of homeotic genes. *Bioessays* **19**: 379-388.

Green, E.L. (1967). Shambling, a neurological mutant of the mouse. *J. Hered.* **58**: 65-68.

Grigoriadis, A.E., Heersche, J.N., and Aubin, J.E. (1988). Differentiation of muscle, fat, cartilage, and bone from progenitor cells present in a bone-derived clonal cell population: effect of dexamethasone. *J. Cell Biol* **106**: 2139-2151.

Grigoriadis, A.E., Wang, Z.Q., Cecchini, M.G., Hofstetter, W., Felix, R., Fleisch, H.A., and Wagner, E.F. (1994). c-Fos: a key regulator of osteoclast-macrophage lineage determination and bone remodeling. *Science* **266**: 443-448.

Grigoriadis, A.E., Wang, Z.Q., and Wagner, E.F. (1995). Fos and bone cell development: lessons from a nuclear oncogene. *Trends. Genet.* **11**: 436-441.

Grindley, J.C., Davidson, D.R., and Hill, R.E. (1995). The role of Pax-6 in eye and nasal development. *Development* **121**: 1433-1442.

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

Gruss, P. and Walther, C. (1992). Pax in development. *Cell* **69**: 719-722.

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

Gyapay, G., Schmitt, K., Fizames, C., Jones, H., Vega-Czarny, N., Spillett, D., Muselet, D., Prud'Homme, J.F., Dib, C., Auffray, C., Morissette, J., Weissenbach, J., and Goodfellow, P.N. (1996). A radiation hybrid map of the human genome. *Hum. Mol. Genet.* **5**: 339-346.

Halazonetis, T.D., Georgopoulos, K., Greenberg, M.E., and Leder, P. (1988). c-Jun dimerizes with itself and with c-Fos, forming complexes of different DNA binding affinities. *Cell* **55**: 917-924.

Halder, G., Callaerts, P., and Gehring, W.J. (1995). Induction of ectopic eyes by targeted expression of the eyeless gene in Drosophila. *Science* **267**: 1788-1792.

Hann, S.R., Sloan-Brown, K., and Spotts, G.D. (1992). Translational activation of the non-AUG -initiated c-*myc* 1 protein at high cell densities due to methionine deprivation. *Genes Devel.* **6**: 1229-1240.

Hanson, I.M., Fletcher, J.M., Jordan, T., Brown, A., Taylor, D., Adams, R.J., Punnett, H.H., and Van Heyningen, V. (1994). Mutations at the PAX6 locus are found in heterogeneous anterior segment malformations including Peters' anomaly. *Nat. Genet.* **6**: 168-173.

Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C., Brokstein, P., Lewis, S., and Rubin, G.M. (1997). BDGP/HHMI Drosophila EST Project (Unpublished)

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

Hauptmann, G. and Gerster, T. (1994). Two-color whole-mount in situ hybridization to vertebrate and Drosophila embryos. *Trends. Genet.* **10**: 266.

Hedge, S.P., Kumar, A., Kurschner, C., and Shapiro, L.H. (1998). c-Maf interacts with c-Myb to regulate transcription of an early myeloid gene during differentiation. *Mol. Cell. Biol.* **16**: 2729-2737.

Herr, W., Sturm, R.A., Clerc, R.G., Corcoran, L.M., Baltimore, D., Sharp, P.A., Ingraham, H.A., Rosenfeld, M.G., Finney, M., and Ruvkun, G. (1988). The POU domain: a large conserved region in the mammalian pit-1, oct-1, oct-2, and Caenorhabditis elegans unc-86 gene products. *Genes. Dev.* **2**: 1513-1516.

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

Hicks, G.G., Shi, E.G., Li, X.M., Li, C.H., Pawlak, M., and Ruley, H.E. (1997). Functional genomics in mice by tagged sequence mutagenesis. *Nat. Genet.* **16**: 338-344.

Hilberg, F., Aguzzi, A., Howells, N., and Wagner, E.F. (1993). c-jun is essential for normal mouse development and hepatogenesis. *Nature* **365**: 179-181.

Hill, R.E., Favor, J., Hogan, B.L., Ton, C.C., Saunders, G.F., Hanson, I.M., Prosser, J., Jordan, T., Hastie, N.D., and Van Heyningen, V. (1991). Mouse small eye results from mutations in a paired-like homeobox- containing gene. *Nature* **354**: 522-525.

Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Tan,F. Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R. (1995). WashU-Merck EST Project (Unpublished).

Hodgkin, J., Plasterk, R.H., and Waterston, R.H. (1995). The nematode Caenorhabditis elegans and its genome. *Science* **270**: 410-414.

Hodgkinson, C.A., Moore, K.J., Nakayama, A., Steingrimsson, E., Copeland, N.G., Jenkins, N.A., and Arnheiter, H. (1993). Mutations at the mouse microphthalmia locus are associated with defects in a gene encoding a novel basic-helix-loop-helix-zipper protein. *Cell* **74**: 395-404.

Hoheisel, J.D., Lennon, G.G., Zehetner, G., and Lehrach, H. (1991). Use of high coverage reference libraries of Drosophila melanogaster for relational data analysis. A step towards mapping and sequencing of the genome. *J. Mol. Biol.* **220**: 903-914.

Hol, F.A., Geurds, M.P.A., Chatkupt, S., Shugart, Y.Y., Balling, R., Schranderstumpel, C.T.R., Johnson, W.G., Hamel, B.C.J., and Mariman, E.C.M. (1996). Pax genes and human neural-tube defects - an amino-acid substitution in pax1 in a patient with spina-bifida. *J. Med. Genet.* **33**: 655-660.

Hsieh, F., Kido, M., Swaroop, A., Blank, J.C., and Rich, K.A. (1996). In-situ localization of bzip transcription factor nrl during mouse retinal development demonstrates a temporal correlation with rod photoreceptor differentiation. *Invest. Ophth. Vis. Sci.* **37**: 2884.

Hsu, W., Kerppola, T.K., Chen, P.L., Curran, T., and Chen-Kiang, S. (1994). Fos and Jun repress transcription activation by NF-IL6 through association at the basic zipper region. *Mol. Cell Biol.* **14**: 268-276.

Hu, J.C., O'Shea, E.K., Kim, P.S., and Sauer, R.T. (1990). Sequence requirements for coiled-coils: analysis with lambda repressor- GCN4 leucine zipper fusions. *Science* **250**: 1400-1403.

Hughes, M.J., Lingrel, J.B., Krakowsky, J.M., and Anderson, K.P. (1993). A helix-loop-helix transcription factor-like gene is located at the mi locus. *J. Biol. Chem.* **268**: 20687-20690.

Hunter, K.W., Riba, L., Schalkwyk, L., Clark, M., Resenchuk, S., Beeghly, A., Su, J., Tinkov, F., Lee, P., Ramu, E., Lehrach, H., and Housman, D. (1996). Toward the construction of integrated physical and genetic maps of the mouse genome using interspersed repetitive sequence PCR (IRS-PCR) genomics. *Genome Res.* **6**: 290-299.

Hunter, T. (1982). Synthetic peptide substrates for a tyrosine protein kinase. *J. Biol. Chem.* **257**: 4843-4848.

Hurst, H.C. (1995). Transcription factors 1: bZIP proteins. *Protein. Profile.* **2**: 101-168.

Igarashi, K., Hoshino, H., Muto, A., Suwabe, N., Nishikawa, S., Nakauchi, H., and Yamamoto, M. (1998). Multivalent DNA binding complex generated by small Maf and Bach1 as a possible biochemical basis for b-globin Locus Control Region complex. *J. Biol. Chem.* **19**: 11783-11790.

Igarashi, K., Itoh, K., Motohashi, H., Hayashi, N., Matuzaki, Y., Nakauchi, H., Nishizawa, M., and Yamamoto, M. (1995). Activity and expression of murine small Maf family protein MafK. *J. Biol. Chem.* **270**: 7615-7624.

Igarashi, K., Kataoka, K., Itoh, K., Hayashi, N., Nishizawa, M., and Yamamoto, M. (1994). Regulation of transcription by dimerization of erythroid factor NF-E2 p45 with small Maf proteins. *Nature* **367**: 568-572.

Inamdar, N.M., Ahn, Y.I., and Alam, J. (1996). The heme-responsive element of the mouse heme oxygenase-1 gene is an extended AP-1 binding site that resembles the recognition sequences for MAF and NF-E2 transcription factors. *Biochem. Biophys. Res. Commun.* **221**: 570-576.

Ioannou, P.A., Amemiya, C.T., Garnes, J., Kroisel, P.M., Shizuya, H., Chen, C., Batzer, M.A., and de Jong, P.J. (1994). A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nat. Genet.* **6**: 84-89.

Itoh, K., Chiba, T., Takahashi, S., Ishii, T., Igarashi, K., Katoh, Y., Oyake, T., Hayashi, N., Satoh, K., Hatayama, I., Yamamoto, M., and Nabeshima, Y. (1997). An Nrf2/small Maf heterodimer mediates the induction of phase II detoxifying enzyme genes through antioxidant response elements. *Biochem. Biophys. Res. Commun.* **236**: 313-322.

Itoh, K., Igarashi, K., Hayashi, N., Nishizawa, M., and Yamamoto, M. (1995). Cloning and characterization of a novel erythroid cell-derived CNC family transcription factor heterodimerizing with the small Maf family proteins. *Mol. Cell Biol.* **15**: 4184-4193.

Jackson, R.J (1993). Cytoplasmic regulation of mRNA function: the importance of the 3' untranslated region. *Cell* **74**: 9-14.

Jackson, V.N., Price, N.T., Carpenter, L., and Halestrap, A.P. (1997). Cloning of the monocarboxylate transporter isoform MCT2 from rat testis provides evidence that expression in tissues is species-specific and may involve post-transcriptional regulation. *Biochem. J.* **324**: 447-453.

Jaenisch, R., Harbers, K., Schnieke, A., Lohler, J., Chumakov, I., Jahner, D., Grotkopp, D., and Hoffmann, E. (1983). Germline integration of moloney murine leukemia virus at the Mov13 locus leads to recessive lethal mutation and early embryonic death. *Cell* **32**: 209-216.

Jansen, G., Hazendonk, E., Thijssen, K.L., and Plasterk, R.H. (1997). Reverse genetics by chemical mutagenesis in Caenorhabditis elegans. *Nat. Genet.* **17**: 119-121.

Jansen, M., De Moor, C.H., Sussenbach, J.S., and Van Den Brande, J.L. (1995). Translational control of gene expression. *Pediatric Res.* **37**: 681-686.

Johnsen, O., Murphy, P., Prydz, H., and Kolsto, A.B. (1998). Interaction of the CNC-bZIP factor TCF11/LCR-F1/Nrf1 with MafG: binding-site selection and regulation of transcription. *Nucleic Acids Res.* **26**: 512-520.

Johnsen, O., Skammelsrud, N., Luna, L., Nishizawa, M., Prydz, H., and Kolsto, A.B. (1996). Small Maf proteins interact with the human transcription factor TCF11/Nrf1/LCR-F1. *Nucleic Acids Res.* **24**: 4289-4297.

Johnson, R.S., Spiegelman, B.M., and Papaioannou, V. (1992). Pleiotropic effects of a null mutation in the c-fos proto-oncogene. *Cell* **71**: 577-586.

Johnson, R.S., van Lingen, B., Papaioannou, V.E., and Spiegelman, B.M. (1993). A null mutation at the c-jun locus causes embryonic lethality and retarded cell growth in culture. *Genes Dev.* **7**: 1309-1317.

Johnstone, K.A. (1995). Construction and evaluation of a chromosome 21 specific hncDNA library (London: University of London).

Jordan, T., Hanson, I., Zaletayev, D., Hodgson, S., Prosser, J., Seawright, A., Hastie, N., and Van Heyningen, V. (1992). The human PAX6 gene is mutated in two patients with aniridia. *Nat. Genet.* **1**: 328-332.

Justice, M.J., Zheng, B., Woychik, R.P., and Bradley, A. (1997). Using targeted large deletions and high-efficiency N-ethyl-N- nitrosourea mutagenesis for functional analyses of the mammalian genome. *Methods* **13**: 423-436.

Kallioniemi, A., Kallioniemi, O.-P., Piper, J., Tanner, M., Stokke, T., Chen, L., Smith, H.S., Pinkel, D., Gray, J.W., and Waldman, F.M. (1994). Detection and mapping of amplified DNA sequences in breast cancer by comparative genomic hybridization. *Proc. Natl. Acad. Sci. USA* **91**: 2156-2160.

Karin, M., Liu Zg, , and Zandi, E. (1997). AP-1 function and regulation. *Curr. Opin. Cell Biol.* **9**: 240-246.

Katagiri, F., Seipel, K., and Chua, N.H. (1992). Identification of a novel dimer stabilization region in a plant bZIP transcription factor. *Mol. Cell. Biol.* **12**: 4809-4816.

Kataoka, K., Fujiwara, K.T., Noda, M., and Nishizawa, M. (1994a). MafB, a new Maf family transcription activator that can associate with Maf and Fos but not with Jun. *Mol. Cell. Biol.* **14**: 7581-7591.

Kataoka, K., Igarashi, K., Itoh, K., Fujiwara, K.T., Noda, M., Yamamoto, M., and Nishizawa, M. (1995). Small Maf proteins heterodimerize with Fos and may act as competitive repressors of the NF-E2 transcription factor. *Mol. Cell. Biol.* **15**: 2180-2190.

Kataoka, K., Nishizawa, M., and Kawai, S. (1993). Structure-function analysis of the maf oncogene product, a member of the b-Zip protein family. *J. Virol.* **67**: 2133-2141.

Kataoka, K., Noda, M., and Nishizawa, M. (1994b). Maf nuclear oncoprotein recognizes sequences related to an AP-1 site and forms heterodimers with both Fos and Jun.. *Mol. Cell. Biol.* **14**: 700-712.

Kennison, J.A. (1993). Transcriptional activation of Drosophila homeotic genes from distant regulatory elements. *Trends. Genet.* **9**: 75-79.

Kenyon, C. (1988). The nematode Caenorhabditis elegans. *Science* **240**: 1448-1453.

Kerppola, T.K. and Curran, T. (1991). Transcription factor interactions: basics on zippers. *Curr. Opin. Struc. Biol.* **1**: 71-79.

Kerppola, T.K. and Curran, T. (1994). A conserved region adjacent to the basic domain is required for recognition of an extended DNA binding site by Maf/Nrl family proteins. *Oncogene.* **9**: 3149-3158.

Kim, S.J., Angel, P., Lafyatis, R., Hattori, K., Kim, K.Y., Sporn, M.B., Karin, M., and Roberts, A.B. (1990). Autoinduction of transforming growth factor beta 1 is mediated by the AP-1 complex. *Mol. Cell Biol.* **10**: 1492-1497.

Kingsley, D.M., Bland, A.E., Grubber, J.M., Marker, P.C., Russell, L.B., Copeland, N.G., and Jenkins, N.A. (1992). The mouse short ear skeletal morphogenesis locus is associated with defects in a bone morphogenetic member of the TGF beta superfamily. *Cell* **71**: 399-410.

Kirby, M.L., Hunt, P., Wallis, K., and Thorogood, P. (1997). Abnormal patterning of the aortic arch arteries does not evoke cardiac malformations. *Dev. Dyn.* **208**: 34-47.

Kishimoto, A., Nishiyama, K., Nakanishi, H., Uratsuji, Y., Nomura, H., Takeyama, Y., and Nishizuka, Y. (1985). Studies on the phosphorylation of myelin basic protein by protein kinase C and adenosine 3':5'-monophosphate-dependent protein kinase. *J. Biol. Chem.* **260**: 12492-12499.

Kistner, A., Gossen, M., Zimmermann, F., Jerecic, J., Ullmer, C., Lubbert, H., and Bujard, H. (1996). Doxycycline-mediated quantitative and tissue-specific control of gene expression in transgenic mice. *Proc. Natl. Acad. Sci. U. S. A.* **93**: 10933-10938.

Kitao, Y., Robertson, B., Kudo, M., and Grant, G. (1996). Neurogenesis of subpopulations of rat lumbar dorsal root ganglion neurons including neurons projecting to the dorsal column nuclei. *J. Comp. Neurol.* **371**: 249-257.

Klymkowsky, M.W. and Karnovsky, A. (1994). Morphogenesis and the cytoskeleton: studies of the Xenopus embryo. *Dev. Biol.* **165**: 372-384.

Knittel, T., Kessel, M., Kim, M.H., and Gruss, P. (1995). A conserved enhancer of the human and murine Hoxa-7 gene specifies the anterior boundary of expression during embryonal development. *Development.* **121**: 1077-1088.

Konig, P. and Richmond, T.J. (1993). The X-ray structure of the GCN4-bZIP bound to ATF/CREB site DNA shows the complex depends on DNA flexibility. *J. Mol. Biol.* **233**: 139-154.

Kotkow, K.J. and Orkin, S.H. (1995). Dependence of globin gene expression in mouse erythroleukemia cells on the NF-E2 heterodimer. *Mol. Cell Biol.* **15**: 4640-4647.

Kouzarides, T. and Ziff, E. (1989). Leucine zippers of fos, jun and GCN4 dictate dimerization specificity and thereby control DNA binding. *Nature* **340**: 568-571.

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

Kozak, M. (1992). Regulation of translation in eukaryotic systems. *Annu Rev Cell Biol* **8**: 197-225.

Kozak, M. (1995). Adherence to the first-AUG rule when a second AUG codon follows closely opon the first. *Proc. Natl. Acad. Sci. USA* **92**: 2662-2666.

Kozak, M. (1996). Interpreting cDNA sequences: some insights from studies on translation. *Mamm. Genome* **7**: 563-574.

Kraut, R. and Levine, M. (1991). Mutually repressive interactions between the gap genes giant and Kruppel define middle body regions of the Drosophila embryo. *Development.* **111**: 611-621.

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

Krylov, D., Kasai, K., Echlin, D.R., Taparowsky, E.J., Arnheiter, H., and Vinson, C. (1997). A general method to design dominant negatives to b-hlhzip proteins that abolish dna binding. *Proc. Natl. Acad. Sci. U. S. A.* **94**: 12274-12279.

Kuhn, R., Schwenk, F., Aguet, M., and Rajewsky, K. (1995). Inducible gene targeting in mice. *Science* **269**: 1427-1429.

Kumar, R., Chen, S., Scheurer, D., Wang, Q.L., Duh, E., Sung, C.H., Rehemtulla, A., Swaroop, A., Adler, R., and Zack, D.J. (1996). The bZIP transcription factor Nrl stimulates rhodopsin promoter activity in primary retinal cell cultures. *J. Biol. Chem.* **271**: 29612-29618.

Landschulz, W.H., Johnson, P.F., and McKnight, S.L. (1988). The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* **240**: 1759-1764.

Lane, P.W. and Green, M.C. (1962). Teetering (tn). *Mouse News Lett.* **27**: 38-30.

Le Dourain, N.M. (1982). *The Neural Crest.* (Cambridge: Cambridge University Press,).

Le Mouellic, H., Lallemand, Y., and Brulet, P. (1992). Homeosis in the mouse induced by a null mutation in the Hox-3.1 gene. *Cell* **69**: 251-264.

Lechner, M.S. and Dressler, G.R. (1996). Mapping of Pax-2 transcription activation domains. *J. Biol. Chem.* **271**: 21088-21093.

Lee, C.Q., Yun, Y.D., Hoeffler, J.P., and Habener, J.F. (1990). Cyclic-AMP-responsive transcriptional activation of CREB-327 involves interdependent phosphorylated subdomains. *EMBO J.* **9**: 4455-4465.

Lee, K.A. and Masson, N. (1993). Transcriptional regulation by CREB and its relatives. *Biochim. Biophys. Acta* **1174**: 221-233.

Lee, W., Mitchell, P., and Tjian, R. (1987). Purified transcription factor AP-1 interacts with TPA-inducible enhancer elements. *Cell* **49**: 741-752.

Lehrach, H. (1990). . In Genome Analysis Volume 1: Genetic and Physical Mapping. K.E. Davies and S.M. Tilgham, eds. (Cold Spring Harbor: Cold Spring Harbor Laboratory Press), pp. 39-81.

Lewandoski, M. and Martin, G.R. (1997). Cre-mediated chromosome loss in mice. *Nat. Genet.* **17**: 223-225.

Lewis, E.B. (1978). A gene complex controlling segmentation in Drosophila. *Nature* **276**: 565-570.

Li, S., Crenshaw, E.B., Rawson, E.J., Simmons, D.M., Swanson, L.W., and Rosenfeld, M.G. (1990). Dwarf locus mutants lacking three pituitary cell types result from mutations in the POU-domain gene pit-1. *Nature* **347**: 528-533.

Li, Z.W., Stark, G., Gotz, J., Rulicke, T., Gschwind, M., Huber, G., Muller, U., and Weissmann, C. (1996). Generation of mice with a 200-kb amyloid precursor protein gene deletion by Cre recombinase-mediated site-specific recombination in embryonic stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **93**: 6158-6162.

Li, E., Sucov, H.M., Lee, K.F., Evans, R.M., and Jaenisch, R. (1993). Normal development and growth of mice carrying a targeted disrutpion of the alpha 1 retinoic acid receptor gene. *Proc. Natl. Acad. Sci. USA* **90**: 1590-1594.

Lin, A., Frist, J., Deng, T., Smeal, T., al Alawi, N., Kikkawa, U., Hunter, T., Brenner, D., and Karin, M. (1992). Casein kinase II is a negative regulator of c-Jun DNA binding and AP-1 activity. *Cell* **70**: 777-789.

Lindblom, A., Skoog, L., Andersen, T.I., Rotstein, S., Nordenskjold, M., and Larsson, C. (1993). Four separate regions on chromosome 17 show loss of heterozygosity in familial breast carcinomas. *Human Genetics* **91**: 6-12.

Lipman, D.J. (1997). Making (anti)sense of non-coding sequence conservation. *Nucleic Acids Res.* **25**: 3580-3583.

Liu, I.S., Chen, J.D., Ploder, L., Vidgen, D., van der Kooy, D., Kalnins, V.I., and McInnes, R.R. (1994). Developmental expression of a novel murine homeobox gene (Chx10): evidence for roles in determination of the neuroretina and inner nuclear layer. *Neuron.* **13**: 377-393.

Liu, Q., Ji, X., Breitman, M.L., Hitchcock, P.F., and Swaroop, A. (1996). Expression of the bZIP transcription factor gene Nrl in the developing nervous system. *Oncogene.* **12**: 207-211.

Lohler, J., Timpl, R., and Jaenisch, R. (1984). Embryonic lethal mutation in mouse collagen I gene causes rupture of blood vessels and is associated with erythropoietic and mesenchymal cell death. *Cell* **38**: 597-607.

Lohnes, D., Mark, M., Mendelsohn, C., Dolle, P., Dierich, A., Gorry, P., Gansmuller, A., and Chambon, P. (1994). Function of the retinoic acid receptors (RARs) during development (I). Craniofacial and skeletal abnormalities in RAR double mutants. *Development.* **120**: 2723-2748.

Lu, S.J., Rowan, S., Bani, M.R., and Ben-David, Y. (1994). Retroviral integration within the Fli-2 locus results in inactivation of the erythroid transcription factor NF-E2 in Friend erythroleukemias: evidence that NF-E2 is essential for globin expression. *Proc. Natl. Acad. Sci. U. S. A.* **91**: 8398-8402.

Lufkin, T., Dierich, A., LeMeur, M., Mark, M., and Chambon, P. (1991). Disruption of the Hox-1.6 homeobox gene results in defects in a region corresponding to its rostral domain of expression. *Cell* **66**: 1105-1119.

Lufkin, T., Lohnes, D., Mark, M., Dierich, A., Gorry, P., Gaub, M.P., LeMeur, M., and Chambon, P. (1993). High postnatal lethality and testis degeneration in retinoic acid receptor alpha mutant mice. *Proc. Natl. Acad. Sci. U. S. A.* **90**: 7225-7229.

Luna, L., Johnsen, O., Skartlien, A.H., Pedeutour, F., Turc-Carel, C., Prydz, H., and Kolsto, A.B. (1994). Molecular cloning of a putative novel human bZIP transcription factor on chromosome 17q22. *Genomics.* **22**: 553-562.

Luna, L., Skammelsrud, N., Johnsen, O., Abel, K.J., Weber, B.L., Prydz, H., and Kolsto, A.B. (1995). Structural organization and mapping of the human TCF11 gene. *Genomics.* **27**: 237-244.

Luo, Y., Fujii, H., Gerster, T., and Roeder, R.G. (1992). A novel B cell-derived coactivator potentiates the activation of immunoglobulin promoters by octamer-binding transcription factors. *Cell* **71**: 231-241.

Lyon, M.F. and Glenister, P.H. (1984). Bareskin (Bsk). *Mouse News Lett.* **71**: 26-20.

Lyon, M.F. and Kirby, M.C. (1992). Mouse Chromosome Atlas. *Mouse Genome* **90**: 22-46.

Lyons, L.A., Laughlin, T.F., Copeland, N.G., Jenkins, N.A., Womack, J.E., and O'Brien, S.J. (1997). Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. *Nat. Genet.* **15**: 47-56.

Mackensen, J.A. (1960). "Open eyelids" in newborn mice. *J. Hered.* **51**: 188-190.

Maki, Y., Bos, T.J., Davis, C., Starbuck, M., and Vogt, P.K. (1987). Avian sarcoma virus 17 carries the jun oncogene. *Proc. Natl. Acad. Sci. U. S. A.* **84**: 2848-2852.

Mansouri, A., Stoykova, A., Torres, M., and Gruss, P. (1996). Dysgenesis of cephalic neural crest derivatives in pax7(-/-) mutant mice. *Development* **122**: 831-838.

Manzanares, M., Cordes, S., Kwan, C.T., Sham, M.H., Barsh, G.S., and Krumlauf, R. (1997). Segmental regulation of Hoxb-3 by kreisler. *Nature* **387**: 191-195.

Marcelle, C., Stark, M.R., and Bronner-Fraser, M. (1997). Coordinate actions of BMPs, Wnts, Shh and noggin mediate patterning of the dorsal somite. *Development.* **124**: 3955-3963.

Marini, M.G., Chan, K., Casula, L., Kan, Y.W., Cao, A., and Moi, P. (1997). hMAF, a small human transcription factor that heterodimerizes specifically with Nrf1 and Nrf2. *J. Biol. Chem.* **272**: 16490-16497.

Mashal, R.D. and Sklar, J. (1996). Practical methods of mutation detection. *Curr. Opin. Genet. Dev.* **6**: 275-280.

Mason, I.J. (1994). The ins and outs of fibroblast growth factors. *Cell* **78**: 547-552.

Matsugi, T., Kreider, B.L., Delwel, R., Cleveland, J.L., Askew, D.S., and Ihle, J.N. (1995). The Evi-1 zinc finger myeloid transforming protein binds to genomic fragments containing (GATA)n sequences. *Oncogene.* **11**: 191-198.

Mattei, M.G., Philip, N., Passage, E., Moisan, J.P., Mandel, J.L., and Mattei, J.F. (1985). DNA probe localization at 18p113 band by in situ hybridization and identification of a small supernumerary chromosome. *Hum. Genet.* **69**: 268-271.

McCarthy, L.C., Terrett, J., Davis, M.E., Knights, C.J., Smith, A.L., Critcher, R., Schmitt, K., Hudson, J., Spurr, N.K., and Goodfellow, P.N. (1997). A first-generation whole genome-radiation hybrid map spanning the mouse genome. *Genome Res.* **7**: 1153-1161.

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

McGinnis, W., Hart, C.P., Gehring, W.J., and Ruddle, F.H. (1984b). Molecular cloning and chromosome mapping of a mouse DNA sequence homologous to homeotic genes of Drosophila. *Cell* **38**: 675-680.

McGinnis, W. and Krumlauf, R. (1992). Homeobox genes and axial patterning. *Cell* **68**: 283-302.

McGinnis, W., Levine, M.S., Hafen, E., Kuroiwa, A., and Gehring, W.J. (1984c). A conserved DNA sequence in homoeotic genes of the Drosophila Antennapedia and bithorax complexes. *Nature* **308**: 428-433.

McKay, I.J., Muchamore, I., Krumlauf, R., Maden, M., Lumsden, A., and Lewis, J. (1994). The kreisler mouse: a hindbrain segmentation mutant that lacks two rhombomeres. *Development.* **120**: 2199-2211.

McKie, J. and Scambler, P.J. (1996). The Nfe2l1 gene maps to distal mouse chromosome 11. *Mamm. Genome* **7**: 89.

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

Mendelsohn, C., Lohnes, D., Decimo, D., Lufkin, T., LeMeur, M., Chambon, P., and Mark, M. (1994). Function of the retinoic acid receptors (RARs) during development (II). Multiple abnormalities at various stages of organogenesis in RAR double mutants. *Development.* **120**: 2749-2771.

Meyers, E.N., Lewandoski, M., and Martin, G.R. (1998). An Fgf8 mutant allelic series generated by Cre- and Flp-mediated recombination. *Nat. Genet.* **18**: 136-141.

Mignotte, V., Eleouet, J.F., Raich, N., and Romeo, P.H. (1989). Cis- and trans-acting elements involved in the regulation of the erythroid promoter of the human porphobilinogen deaminase gene. *Proc. Natl. Acad. Sci. U. S. A.* **86**: 6548-6552.

Mitchell, P.J. and Tjian, R. (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* **245**: 371-381.

Mitelman, F. and Heim, S. (1988). Consistent involvement of only 71 of the 329 chromosomal bands of the human genome in primary neoplasia-associated rearrangements. *Cancer Res.* **48**: 7115-7119.

Mitelman, F. (1993). The cytogenetic scenario of chronic myeloid leukemia. *Leuk. Lymphoma.* **11 Suppl 1**: 11-15.

Mohler, J., Mahaffey, J.W., Deutsch, E., and Vani, K. (1995). Control of Drosophila head segment identity by the bZIP homeotic gene cnc. *Development.* **121**: 237-247.

Mohler, J., Vani, K., Leung, S., and Epstein, A. (1991). Segmentally restricted, cephalic expression of a leucine zipper gene during Drosophila embryogenesis. *Mech. Dev.* **34**: 3-9.

Moi, P., Chan, K., Asunis, I., Cao, A., and Kan, Y.W. (1994a). Isolation of NF-E2-related factor 2 (Nrf2), a NF-E2-like basic leucine zipper transcriptional activator that binds to the tandem NF-E2/AP1 repeat of the beta-globin locus control region. *Proc. Natl. Acad. Sci. U. S. A.* **91**: 9926-9930.

Moi, P., Chan, K., Marini, G., Casula, L., Melis, A., and Cao, A. (1994b). Cloning of a small human maf protein that heterodimerizes specifically with nrf1. *Blood* **84**: A 19.

Moi, P. and Kan, Y.W. (1990). Synergistic enhancement of globin gene expression by activator protein- 1-like proteins. *Proc. Natl. Acad. Sci. U. S. A.* **87**: 9000-9004.

Monaco, A.P. (1994). Isolation of genes from cloned DNA. *Curr. Opin. Genet. Dev.* **4**: 360-365.

Monaco, A.P. and Larin, Z. (1994). YACs, BACs, PACs and MACs: artificial chromosomes as research tools. *Trends. Biotechnol.* **12**: 280-286.

Morgan, I.M., Ransone, L.J., Bos, T.J., Verma, I.M., and Vogt, P.K. (1992). Transformation by Jun: requirement for leucine zipper, basic domain and enhancement by Fos. *Oncogene.* **8**: 1135-1140.

Morgan, W.C. (1950). A new tail-short mutation in the mouse - Whose lethal effects are conditioned by the residual genotypes. *J. Hered.* **41**: 208-215.

Mortlock, D.P. and Innis, J.W. (1997). Mutation of HOXA13 in hand-foot-genital syndrome. *Nat. Genet.* **15**: 179-180.

Mortlock, D.P., Post, L.C., and Innis, J.W. (1996). The molecular basis of hypodactyly (Hd): a deletion in Hoxa 13 leads to arrest of digital arch formation. *Nat. Genet.* **13**: 284-289.

Moser, A.R., Luongo, C., Gould, K.A., McNeley, M.K., Shoemaker, A.R., and Dove, W.F. (1995). ApcMin: a mouse model for intestinal and mammary tumorigenesis. *Eur. J. Cancer* **31A**: 1061-1064.

Motohashi, H., Shavit, J.A., Igarashi, K., Yamamoto, M., and Engel, J.D. (1997). The world according to Maf. *Nucleic Acids Res.* **25**: 2953-2959.

Munsterberg, A.E., Kitajewski, J., Bumcrot, D.A., McMahon, A.P., and Lassar, A.B. (1995). Combinatorial signaling by Sonic hedgehog and Wnt family members induces myogenic bHLH gene expression in the somite. *Genes. Dev.* **9**: 2911-2922.

Muragaki, Y., Mundlos, S., Upton, J., and Olsen, B.R. (1996). Altered growth and branching patterns in synpolydactyly caused by mutations in HOXD13. *Science* **272**: 548-551.

Murphy, M., Reid, K., Ford, M., Furness, J.B., and Bartlett, P.F. (1994). FGF2 regulates proliferation of neural crest cells, with subsequent neuronal differentiation regulated by LIF or related factors. *Development.* **120**: 3519-3528.

Nagai, T., Igarashi, K., Akasaka, J., Furuyama, K., Fujita, H., Hayashi, N., Yamamoto, M., and Sassa, S. (1998). Regulation of NF-E2 activity in erythroleukemia cell differentiation. *J. Biol. Chem.* **273**: 5358-5365.

Nakabeppu, Y. and Nathans, D. (1991). A naturally occurring truncated isoform of FosB that inhibits Fos/Jun transcriptional activity. *Cell* **64**: 751-759.

Ney, P.A., Sorrentino, B.P., McDonagh, K.T., and Nienhuis, A.W. (1990). Tandem AP-1-binding sites within the human beta-globin dominant control region function as an inducible enhancer in erythroid cells. *Genes. Dev.* **4**: 993-1006.

Nicklin, M.J. and Casari, G. (1991). A single site mutation in a truncated Fos protein allows it to interact with the TRE in vitro. *Oncogene.* **6**: 173-179.

Nishizawa, M., Kataoka, K., Goto, N., Fujiwara, K.T., and Kawai, S. (1989). v-maf, a viral oncogene that encodes a "leucine zipper" motif. *Proc. Natl. Acad. Sci. U. S. A.* **86**: 7711-7715.

Nonchev, S., Vesque, C., Maconochie, M., Seitanidou, T., Ariza-McNaughton, L., Frain, M., Marshall, H., Sham, M.H., Krumlauf, R., and Charnay, P. (1996). Segmental expression of Hoxa-2 in the hindbrain is directly regulated by Krox-20. *Development.* **122**: 543-554.

O'Brien, S.J., Womack, J.E., Lyons, L.A., Moore, K.J., Jenkins, N.A., and Copeland, N.G. (1993). Anchored reference loci for comparative genome mapping in mammals. *Nat. Genet.* **3**: 103-112.

O'Shea, E.K., Klemm, J.D., Kim, P.S., and Alber, T. (1991). X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel colied coil. *Science* **254**: 539-544.

O'Shea, E.K., Rutkowski, R., and Kim, P.S. (1989). Evidence that the leucine zipper is a coiled coil. *Science* **243**: 538-542.

O'Shea, E.K., Rutkowski, R., and Kim, P.S. (1992). Mechanism of specificity in the Fos-Jun oncoprotein heterodimer. *Cell* **68**: 699-708.

Ogino, H. and Yasuda, K. (1998). Induction of lens differentiation by activation of a bzip transcription factor, l-maf. *Science* **280**: 115-118.

Oliver, G., Sosa-Pineda, B., Geisendorf, S., Spana, E.P., Doe, C.Q., and Gruss, P. (1993). Prox 1, a *prospero*-related homeobox gene expressed during mouse development. *Mech. Dev.* **44**: 3-16.

Olson, E.N., Arnold, H.H., Rigby, P.W., and Wold, B.J. (1996). Know your neighbours: three phenotypes in null mutants of the myogenic bHLH gene MRF4. *Cell* **85**: 1-4.

Opdecamp, K., Nakayama, A., Nguyen, M.T., Hodgkinson, C.A., Pavan, W.J., and Arnheiter, H. (1997). Melanocyte development in vivo and in neural crest cell cultures: crucial dependence on the Mitf basic-helix-loop-helix-zipper transcription factor. *Development.* **124**: 2377-2386.

Ordahl, C.P. and Williams, .B.A. (1998). Knowing cops from chuck: roasting MyoD redundancy. *Bioessays* **20**: 357-362.

Orkin, S.H. (1990). Globin gene regulation and switching: circa 1990. *Cell* **63**: 665-672.

Oyake, T., Itoh, K., Motohashi, H., Hayashi, N., Hoshino, H., Nishizawa, M., Yamamoto, M., and Igarashi, K. (1996). Bach proteins belong to a novel family of BTB-basic leucine zipper transcription factors that interact with MafK and regulate transcription through the NF-E2 site. *Mol. Cell Biol.* **16**: 6083-6095.

Pabo, C.O. and Sauer, R.T. (1992). Transcription factors: structural families and principles of DNA recognition. *Annu. Rev. Biochem.* **61**: 1053-1095.

Pal, S., Lo, M.C., Schmidt, D., Pelczer, I., Thurber, S., and Walker, S. (1997). Skn-1: evidence for a bipartite recognition helix in DNA binding. *Proc. Natl. Acad. Sci. U. S. A.* **94**: 5556-5561.

Paro, R. (1993). Mechanisms of heritable gene repression during development of Drosophila. *Curr. Opin. Cell Biol.* **5**: 999-1005.

Pascal, E. and Tjian, R. (1991). Different activation domains of Sp1 govern formation of multimers and mediate transcriptional synergism. *Genes. Dev.* **5**: 1646-1656.

Perkins, K.K., Admon, A., Patel, N., and Tjian, R. (1990). The Drosophila Fos-related AP-1 protein is a developmentally regulated transcription factor. *Genes. Dev.* **4**: 822-834.

Perkins, K.K., Dailey, G.M., and Tjian, R. (1988). Novel Jun- and Fos-related proteins in Drosophila are functionally homologous to enhancer factor AP-1. *EMBO J.* **7**: 4265-4273.

Pesole, G., Fiormarino, G., and Saccone, C. (1994). Sequence-analysis and compositional properties of untranslated regions of human messenger-rnas. *Gene* **140**: 219-225.

Peters, L.L., Andrews, N.C., Eicher, E.M., Davidson, M.B., Orkin, S.H., and Lux, S.E. (1993). Mouse microcytic anaemia caused by a defect in the gene encoding the globin enhancer-binding protein NF-E2. *Nature* **362**: 768-770.

Peters, L.L., Andrews, N.C., Eicher, E.M., Davidson, M.B., Orkin, S.H., and Lux, S.E. (1994). Mouse microcytic anaemia caused by a defect in the gene encoding the globin enhancer-binding protein NF-E2 [published erratum]. *Nature* **371**: 358-350.

Peterson, A. and Biddle, F. (1970). Cocked,co. *Mouse News Lett.* **43**: 19-10.

Phelps, D.E. and Dressler, G.R. (1996). Identification of novel Pax-2 binding sites by chromatin precipitation. *J. Biol. Chem.* **271**: 7978-7985.

Pinna, L.A. (1990). Casein kinase 2: an 'eminence grise' in cellular regulation? *Biochim. Biophys. Acta* **1054**: 267-284.

Pischedda, C., Cocco, S., Melis, A., Marini, M.G., Kan, Y.W., Cao, A., and Moi, P. (1995). Isolation of a differentially regulated splicing isoform of human NF-E2. *Proc. Natl. Acad. Sci. U. S. A.* **92**: 3511-3515.

Plasterk, R.H. (1992). Reverse genetics of Caenorhabditis elegans. *Bioessays* **14**: 629-633.

Pollock, R.A., Jay, G., and Bieberich, C.J. (1992). Altering the boundaries of Hox3.1 expression: evidence for antipodal gene regulation. *Cell* **71**: 911-923.

Pope, F.M., Nicholls, A.C., McPheat, J., Talmud, P., and Owen, R. (1985). Collagen genes and proteins in osteogenesis imperfecta. *J. Med. Genet.* **22**: 466-478.

Popperl, H., Bienz, M., Studer, M., Chan, S.K., Aparicio, S., Brenner, S., Mann, R.S., and Krumlauf, R. (1995). Segmental expression of Hoxb-1 is controlled by a highly conserved autoregulatory loop dependent upon exd/pbx. *Cell* **81**: 1031-1042.

Prieschl, E.E., Novotny, V., Csonga, R., Jaksche, D., Elbe-Burger, A., Thumb, W., Auer, M., Stingl, G., and Baumruker, T. (1998). A novel splice variant of the transcription factor Nrf1 interacts with the TNFa promoter and stimulates transcription. *Nucleic Acids Res.* **26**: 2291-2297.

Quiring, R., Walldorf, U., Kloter, U., and Gehring, W.J. (1994). Homology of the eyeless gene of Drosophila to the Small eye gene in mice and Aniridia in humans. *Science* **265**: 785-789.

Rafael, J.A., Tinsley, J.M., Potter, A.C., Deconinck, A.E., and Davies, K.E. (1998). Skeletal muscle-specific expression of a utrophin transgene rescues utrophin-dystrophin deficient mice. *Nat. Genet.* **19**: 79-82.

Ramirez-Solis, R., Davis, A.C., and Bradley, A. (1993). Gene targeting in embryonic stem cells. *Methods Enzymol.* **225**: 855-878.

Ramirez-Solis, R., Liu, P., and Bradley, A. (1995). Chromosome engineering in mice. *Nature* **378**: 720-724.

Rancourt, D.E., Tsuzuki, T., and Capecchi, M.R. (1995). Genetic interaction between hoxb-5 and hoxb-6 is revealed by nonallelic noncomplementation. *Genes. Dev.* **9**: 108-122.

Rauscher, F.J., Cohen, D.R., Curran, T., Bos, T.J., Vogt, P.K., Bohmann, D., Tjian, R., and Franza, B.R.J. (1988). Fos-associated protein p39 is the product of the jun proto-oncogene. *Science* **240**: 1010-1016.

Raziuddin, A., Sarkar, F.H., Dutkowski, R., Shulman, L., Ruddle, F.H., and Gupta, S.L. (1984). Receptors for human alpha and beta interferon but not for gamma interferon are specified by human chromosome 21. *Proc. Natl. Acad. Sci. U. S. A.* **81**: 5504-5508.

Read, A.P. and Newton, V.E. (1997). Waardenburg syndrome. *J. Med. Genet.* **34**: 656-665.

Rehemtulla, A., Warwar, R., Kumar, R., Ji, X., Zack, D.J., and Swaroop, A. (1996). The basic motif-leucine zipper transcription factor Nrl can positively regulate rhodopsin gene expression. *Proc. Natl. Acad. Sci. U. S. A.* **93**: 191-195.

Reichert, H. and Boyan, G. (1997). Building a brain: developmental insights in insects. *Trends. Neurosci.* **20**: 258-264.

Riesgoescovar, J.R. and Hafen, E. (1997). Common and distinct roles of dfos and djun during drosophila development. *Science* **278**: 669-672.

Rijli, F.M., Mark, M., Lakkaraju, S., Dierich, A., Dolle, P., and Chambon, P. (1993). A homeotic transformation is generated in the rostral branchial region of the head by disruption of Hoxa-2, which acts as a selector gene. *Cell* **75**: 1333-1349.

Robinson, L., Panayiotakis, A., Papas, T.S., Kola, I., and Seth, A. (1997). ETS target genes: identification of egr1 as a target by RNA differential display and whole genome PCR techniques. *Proc. Natl. Acad. Sci. U. S. A.* **94**: 7170-7175.

Rorth, P., Szabo, K., Bailey, A., Laverty, T., Rehm, J., Rubin, G.M., Weigmann, K., Milan, M., Benes, V., Ansorge, W., and Cohen, S.M. (1998). Systematic gain-of-function genetics in Drosophila. *Development.* **125**: 1049-1057.

Rowe, A., Eager, N.S., and Brickell, P.M. (1991). A member of the RXR nuclear receptor family is expressed in neural- crest-derived cells of the developing chick peripheral nervous system. *Development.* **111**: 771-778.

Rubin, G.M. (1988). Drosophila melanogaster as an experimental organism. *Science* **240**: 1453-1459.

Rushmore, T.H., Morton, M.R., and Pickett, C.B. (1991). The antioxidant responsive element. Activation by oxidative stress and identification of the DNA consensus sequence required for functional activity. *J. Biol. Chem.* **266**: 11632-11639.

Ruther, U., Komitowski, D., Schubert, F.R., and Wagner, E.F. (1989). c-fos expression induces bone tumors in transgenic mice. *Oncogene.* **4**: 861-865.

Ruther, U., Muller, W., Sumida, T., Tokuhisa, T., Rajewsky, K., and Wagner, E.F. (1988). c-fos expression interferes with thymus development in transgenic mice. *Cell* **53**: 847-856.

Ryder, K., Lau, L.F., and Nathans, D. (1988). A gene activated by growth factors is related to the oncogene v-jun. *Proc. Natl. Acad. Sci. U. S. A.* **85**: 1487-1491.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual (New York: Cold Spring Harbor Laboratory Press).

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **74**: 5463-5467.

Sanyanusin, P., Schimmenti, L.A., McNoe, L.A., Ward, T.A., Pierpont, M.E., Sullivan, M.J., Dobyns, W.B., and Eccles, M.R. (1995). Mutation of the PAX2 gene in a family with optic nerve colobomas, renal anomalies and vesicoureteral reflux. *Nat. Genet.* **9**: 358-364.

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

Schedl, A., Ross, A., Lee, M., Engelkamp, D., Rashbass, P., Van Heyningen, V., and Hastie, N.D. (1996). Influence of PAX6 gene dosage on development: overexpression causes severe eye abnormalities. *Cell* **86**: 71-82.

Schuermann, M., Hunter, J.B., Hennig, G., and Muller, R. (1991). Non-leucine residues in the leucine repeats of Fos and Jun contribute to the stability and determine the specificity of dimerization. *Nucleic Acids Res.* **19**: 739-746.

Schwartzberg, P.L., Xing, L., Hoffmann, O., Lowell, C.A., Garrett, L., Boyce, B.F., and Varmus, H.E. (1997). Rescue of osteoclast function by transgenic expression of kinase- deficient Src in src-/- mutant mice. *Genes. Dev.* **11**: 2835-2844.

Schwenk, F., Baron, U., and Rajewsky, K. (1995). A cre-transgenic mouse strain for the ubiquitous deletion of loxP- flanked gene segments including deletion in germ cells. *Nucleic Acids Res.* **23**: 5080-5081.

Schwenk, F., Kuhn, R., Angrand, P.O., Rajewsky, K., and Stewart, A.F. (1998). Temporally and spatially regulated somatic mutagenesis in mice. *Nucleic Acids Res.* **26**: 1427-1432.

Scott, M.P., Tamkun, J.W., and Hartzell, G.W. (1989). The structure and function of the homeodomain. *Biochim. Biophys. Acta* **989**: 25-48.

Scott, M.P. and Weiner, A.J. (1984). Structural relationships among genes that control development: sequence homology between the Antennapedia, Ultrabithorax, and fushi tarazu loci of Drosophila. *Proc. Natl. Acad. Sci. U. S. A.* **81**: 4115-4119.

Serbedzija, G.N. and McMahon, A.P. (1997). Analysis of neural crest cell migration in Splotch mice using a neural crest-specific LacZ reporter. *Dev. Biol* **185**: 139-147.

Sham, M.H., Vesque, C., Nonchev, S., Marshall, H., Frain, M., Gupta, R.D., Whiting, J., Wilkinson, D., Charnay, P., and Krumlauf, R. (1993). The zinc finger gene Krox20 regulates HoxB2 (Hox2.8) during hindbrain segmentation. *Cell* **72**: 183-196.

Sharon-Friling, R., Richardson, J., Sperbeck, S., Lee, D., Rauchman, M., Maas, R., Swaroop, A., and Wistow, G. (1998). Lens-specific gene recruitment of zeta-crystallin through Pax6, Nrl- Maf, and brain suppressor sites. *Mol. Cell. Biol.* **18**: 2067-2076.

Shehee, W.R., Oliver, P., and Smithies, O. (1993). Lethal thalassemia after insertional disruption of the mouse major adult beta-globin gene. *Proc. Natl. Acad. Sci. U. S. A.* **90**: 3177-3181.

Shibata, H., Toyama, K., Shioya, H., Ito, M., Hirota, M., Hasegawa, S., Matsumoto, H., Takano, H., Akiyama, T., Toyoshima, K., Kanamaru, R., Kanegae, Y., Saito, I., Nakamura, Y., Shiba, K., and Noda, T. (1997). Rapid colorectal adenoma formation initiated by conditional targeting of the Apc gene. *Science* **278**: 120-123.

Shivdasani, R.A. and Orkin, S.H. (1995). Erythropoiesis and globin gene expression in mice lacking the transcription factor NF-E2. *Proc. Natl. Acad. Sci. U. S. A.* **92**: 8690-8694.

Shivdasani, R.A., Rosenblatt, M.F., Zucker-Franklin, D., Jackson, C.W., Hunt, P., Saris, C.J., and Orkin, S.H. (1995). Transcription factor NF-E2 is required for

platelet formation independent of the actions of thrombopoietin/MGDF in megakaryocyte development. *Cell* **81**: 695-704.

Sibilia, M. and Wagner, E.F. (1995). Strain-dependent epithelial defects in mice lacking the EGF receptor. *Science* **269**: 234-238.

Sidman, R.L. (1967). Cerebellar outflow degeneration (cod). *Mouse News Lett.* **36**: 33-30.

Sieweke, M.H., Tekotte, H., Frampton, J., and Graf, T. (1996). MafB is an interaction partner and repressor of Ets-1 that inhibits erythroid differentiation. *Cell* **85**: 49-60.

Simon, J. (1995). Locking in stable states of gene expression: transcriptional control during Drosophila development. *Curr. Opin. Cell Biol.* **7**: 376-385.

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

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

Smeyne, R.J., Curran, T., and Morgan, J.I. (1992a). Temporal and spatial expression of a fos-lacZ transgene in the developing nervous system. *Brain Res. Mol. Brain Res.* **16**: 158-162.

Smeyne, R.J., Schilling, K., Robertson, L., Luk, D., Oberdick, J., Curran, T., and Morgan, J.I. (1992b). fos-lacZ transgenic mice: mapping sites of gene induction in the central nervous system. *Neuron.* **8**: 13-23.

Sosa-Pineda, B., Chowdhury, K., Torres, M., Oliver, G., and Gruss, P. (1997). The Pax4 gene is essential for differentiation of insulin-producing beta cells in the mammalian pancreas. *Nature* **386**: 399-402.

Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**: 503-517.

Spencer, D.M. (1996). Creating conditional mutations in mammals. *TIGS* **12**: 181-187.

St.Johnston, D. and Nusslein-Volhard, C. (1992). The origin of pattern and polarity in the Drosophila embryo. *Cell* **68**: 201-219.

St-Onge, L., Furth, P.A., and Gruss, P. (1996). Temporal control of the Cre recombinase in transgenic mice by a tetracycline responsive promoter. *Nucleic Acids Res.* **24**: 3875-3877.

St.Johnston, D. (1995). The intracellular localization of messenger RNAs. *Cell* **81**: 161-170.

Stern, C.D. (1994). The avian embryo: a powerful model system for studying neural induction. *FASEB. J.* **8**: 687-691.

Strachan, T. and Read, A.P. (1994). PAX genes. *Curr. Opin. Genet. Dev.* **4**: 427-438.

Streit, A., Stern, C.D., Thery, C., Ireland, G.W., Aparicio, S., Sharpe, M.J., and Gherardi, E. (1995). A role for HGF/SF in neural induction and its expression in Hensen's node during gastrulation. *Development.* **121**: 813-824.

Suda, T., Takahashi, N., and Martin, T.J. (1992). Modulation of osteoclast differentiation. *Endocr. Rev* **13**: 66-80.

Sutherland, J.A., Cook, A., Bannister, A.J., and Kouzarides, T. (1992). Conserved motifs in Fos and Jun define a new class of activation domain. *Genes. Dev.* **6**: 1810-1819.

Suzuki, M. (1989). SPXX, a frequent sequence motif in gene regulatory proteins. *J. Mol. Biol.* **207**: 61-84.

Swaroop, A., Xu, J., Pawar, H., Jackson, A., Skolnick, C., and Agarwal, N. (1992). A conserved retina specific gene encodes a basic motif/leucine zipper domain. *Proc. Natl. Acad. Sci. USA* **89**: 266-270.

Tachibana, M., Perezjurado, L.A., Nakayama, A., Hodgkinson, C.A., Li, X., Schneider, M., Miki, T., Fex, J., Francke, U., and Arnheiter, H. (1994). Cloning of mitf, the

human homolog of the mouse microphthalmia gene and assignment to chromosome 3p14.1 - p12.3. *Human Molecular Genetics* **3**: 553-557.

Tachibana, M., Takeda, K., Nobukuni, Y., and Urabe, K. (1996). Ectopic expression of mitf, a gene for waardenburg syndrome type-2, converts fibroblasts to cells with melanocyte characteristics. *Nature Genetics* **14**: 50-54.

Tajbakhsh, S. and Sporle, R. (1998). Somite development: constructing the vertebrate body. *Cell* **92**: 9-16.

Takeuchi, T., Yamazaki, Y., Katoh-Fukui, Y., Tsuchiya, R., Kondo, S., Motoyama, J., and Higashinakagawa, T. (1995). Gene trap capture of a novel mouse gene, jumonji, required for neural tube formation. *Genes. Dev.* **9**: 1211-1222.

Tassabehji, M., Newton, V.E., and Read, A.P. (1994). Waardenburg syndrome type-2 caused by mutations in the human microphthalmia (mitf) gene. *Nature Genetics* **8**: 251-255.

Theil, E.C. (1993). The IRE (iron regulatory element) family: structures which regulate mRNA translation or stability. *BioFactors* **4**: 87-93.

Thomas, K.R. and Capecchi, M.R. (1990). Targeted disruption of the murine int-1 proto-oncogene resulting in severe abnormalities in midbrain and cerebellar development. *Nature* **346**: 847-850.

Thomas, K.R., Musci, T.S., Neumann, P.E., and Capecchi, M.R. (1991). Swaying is a mutant allele of the proto-oncogene Wnt-1. *Cell* **67**: 969-976.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673-4680.

Thompson, S., Clarke, A.R., Pow, A.M., Hooper, M.L., and Melton, D.W. (1989). Germ line transmission and expression of a corrected HPRT gene produced by gene targeting in embryonic stem cells. *Cell* **56**: 313-321.

Threadgill, D.W., Dlugosz, A.A., Hansen, L.A., Tennenbaum, T., Lichti, U., Yee, D., LaMantia, C., Mourton, T., Herrup, K., Harris, R.C., Barnard, J.A., Yuspa, S.H., Coffey,

R.J., and Magnuson, T. (1995). Targeted disruption of mouse EGF receptor: effect of genetic background on mutant phenotype. *Science* **269**: 230-234.

Todd, J.A., Aitman, T.J., Cornall, R.J., Ghosh, S., Hall, J.R., Hearne, C.M., Knight, A.M., Love, J.M., McAleer, M.A., and Prins, J.B. (1991). Genetic analysis of autoimmune type 1 diabetes mellitus in mice. *Nature* **351**: 542-547.

Toki, T., Itoh, J., Kitazawa, J., Arai, K., Hatakeyama, K., Akasaka, J., Igarashi, K., Nomura, N., Yokoyama, M., Yamamoto, M., and Ito, E. (1997). Human small Maf proteins form heterodimers with CNC family transcription factors and recognize the NF-E2 motif. *Oncogene.* **14**: 1901-1910.

Torres, M., Gomez-Pardo, E., Dressler, G.R., and Gruss, P. (1995). Pax-2 controls multiple steps of urogenital development. *Development.* **121**: 4057-4065.

Towler, D.A., Gordon, J.I., Adams, S.P., and Glaser, L. (1988). The biology and enzymology of eukaryotic protein acylation. *Annu. Rev. Biochem.* **57**: 69-99.

Townley, D.J., Avery, B.J., Rosen, B., and Skarnes, W.C. (1997). Rapid sequence analysis of gene trap integrations to generate a resource of insertional mutations in mice. *Genome Res.* **7**: 293-298.

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* **83**: 753-760.

Treisman, R. (1985). Transient accumulation of c-fos RNA following serum stimulation requires a conserved 5' element and c-fos 3' sequences. *Cell* **42**: 889-902.

Tsien, J.Z., Chen, D.F., Gerber, D., Tom, C., Mercer, E.H., Anderson, D.J., Mayford, M., Kandel, E.R., and Tonegawa, S. (1996). Subregion- and cell type-restricted gene knockout in mouse brain. *Cell* **87**: 1317-1326.

Uchida, K., Koopman, P., Mita, A., Wakana, S., Wright, E., Kikkawa, Y., Yonekawa, H., Moriwaki, K., and Shiroishi, T. (1996). Exclusion of Sox9 as a candidate for the mouse mutant tail-short. *Mamm. Genome* **7**: 481-485.

Urbanek, P., Wang, Z.Q., Fetka, I., Wagner, E.F., and Busslinger, M. (1994). Complete block of early B cell differentiation and altered patterning of the posterior midbrain in mice lacking Pax5/BSAP. *Cell* **79**: 901-912.

Van Beveren, C., van Straaten, F., Curran, T., Muller, R., and Verma, I.M. (1983). Analysis of FBJ-MuSV provirus and c-fos (mouse) gene reveals that viral and cellular fos gene products have different carboxy termini. *Cell* **32**: 1241-1255.

van der Lugt, N.M., Alkema, M., Berns, A., and Deschamps, J. (1996). The Polycomb-group homolog Bmi-1 is a regulator of murine Hox gene expression. *Mech. Dev.* **58**: 153-164.

van der Lugt, N.M., Domen, J., Linders, K., van Roon, M., Robanus-Maandag, E., te Riele, H., van der Valk, M., Deschamps, J., Sofroniew, M., and van Lohuizen, M. (1994). Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the bmi-1 proto-oncogene. *Genes. Dev.* **8**: 757-769.

van Dijk, M.A. and Murre, C. (1994). extradenticle raises the DNA binding specificity of homeotic selector gene products. *Cell* **78**: 617-624.

van Dijk, M.A., Peltenburg, L.T., and Murre, C. (1995). Hox gene products modulate the DNA binding activity of Pbx1 and Pbx2. *Mech. Dev.* **52**: 99-108.

van Heeckeren, W.J., Sellers, J.W., and Struhl, K. (1992). Role of the conserved leucines in the leucine zipper dimerization motif of yeast GCN4. *Nucleic Acids Res.* **20**: 3721-3724.

Van Roy, N., Laureys, G., Van Gele, M., Opdenakker, G., Miura, R., van der Drift, P., Chan, A., Versteeg, R., and Speleman, F. (1997). Analysis of 1;17 translocation breakpoints in neuroblastoma: implications for mapping of neuroblastoma genes. *Eur. J. Cancer* **33**: 1974-1978.

Venugopal, R. and Jaiswal, A.K. (1996). Nrf1 and Nrf2 positively and c-Fos and Fra1 negatively regulate the human antioxidant response element-mediated expression of NAD(P)H:quinone oxidoreductase1 gene. *Proc. Natl. Acad. Sci. U. S. A.* **93**: 14960-14965.

Vinson, C.R., Hai, T., and Boyd, S.M. (1993). Dimerization specificity of the leucine zipper-containing bZIP motif on DNA binding: prediction and rational design. *Genes. Dev.* **7**: 1047-1058.

Vinson, C.R., Sigler, P.B., and McKnight, S.L. (1989). Scissors-grip model for DNA recognition by a family of leucine zipper proteins. *Science* **246**: 911-916.

von Heijne, G. (1986). A new method for predicting signal sequences cleavage sites. *Nucleic Acids Res.* **14**: 4683-4690.

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

Voss, J.W., Yao, T-P., and Rosenfeld, M.G. (1991). Alternative translation initiation site usage results in two structurally distinct forms of Pit-1. *Journal Of Biological Chemistry* **266**: 12832-12835.

Waeber, G. and Habener, J.F. (1991). Nuclear translocation and DNA recognition signals colocalized within the bZIP domain of cyclic adenosine 3',5'-monophosphate response element-binding protein CREB. *Mol. Endocrinol.* **5**: 1431-1438.

Walther, C. and Gruss, P. (1991). Pax6, a murine paired box gene, is expressed in the developing CNS. *Development* **113**: 1435-1449.

Walther, C., Guenet, J.L., Simon, D., Deutsch, U., Jostes, B., Goulding, M.D., Plachov, D., Balling, R., and Gruss, P. (1991). Pax: a murine multigene family of paired box-containing genes. *Genomics.* **11**: 424-434.

Wang, Z.Q., Grigoriadis, A.E., Mohle-Steinlein, U., and Wagner, E.F. (1991). A novel target cell for c-fos-induced oncogenesis: development of chondrogenic tumours in embryonic stem cell chimeras. *EMBO J.* **10**: 2437-2450.

Wang, Z.Q., Ovitt, C., Grigoriadis, A.E., Mohle-Steinlein, U., Ruther, U., and Wagner, E.F. (1992). Bone and haematopoietic defects in mice lacking c-fos. *Nature* **360**: 741-745.

Ward, J.M. and Young, D.M. (1976). Histogenesis and Morphology of periosteal sarcomas induced by FBJ virus in NIH Swiss mice. *Cancer Res.* **36**: 3985-3992.

Watanabe, A., Takeda, K., Ploplis, B., and Tachibana, M. (1998). Epistatic relationship between waardenburg syndrome genes mitf and pax3. *Nature Genetics* **18**: 283-286.

Wilkinson, D.G., Bhatt, S., Chavrier, P., Bravo, R., and Charnay, P. (1989a). Segment-specific expression of a zinc-finger gene in the developing nervous system of the mouse. *Nature* **337**: 461-464.

Wilkinson, D.G., Bhatt, S., Ryseck, R.P., and Bravo, R. (1989b). Tissue-specific expression of c-jun and junB during organogenesis in the mouse. *Development.* **106**: 465-471.

Wilkinson, D.G. and Nieto, M.A. (1993). Detection of messenger RNA by in situ hybridization to tissue sections and whole mounts. *Methods Enzymol.* **225**: 361-373.

Woessner, J.F.,Jr. (1991). Matrix metalloproteinases and their inhibitors in connective tissue remodeling. *FASEB. J.* **5**: 2145-2154.

Woychik, R.P., Maas, R.L., Zeller, R., Vogt, T.F., and Leder, P. (1990). 'Formins': proteins deduced from the alternative transcripts of the limb deformity gene. *Nature* **346**: 850-853.

Xu, W., Rould, M.A., Jun, S., Desplan, C., and Pabo, C.O. (1995). Crystal structure of a paired domain-DNA complex at 2.5 A resolution reveals structural basis for Pax developmental mutations. *Cell* **80**: 639-650.

You, Y., Bergstrom, R., Klemm, M., Lederman, B., Nelson, H., Ticknor, C., Jaenisch, R., and Schimenti, J. (1997). Chromosomal deletion complexes in mice by radiation of embryonic stem cells. *Nat. Genet.* **15**: 285-288.

Zambrowicz, B.P., Friedrich, G.A., Buxton, E.C., Lilleberg, S.L., Person, C., and Sands, A.T. (1998). Disruption and sequence identification of 2,000 genes in mouse embryonic stem cells. *Nature* **392**: 608-611.

Zeitlinger, J., Kockel, L., Peverali, F.A., Jackson, D.B., Mlodzik, M., and Bohmann, D. (1997). Defective dorsal closure and loss of epidermal decapentaplegic expression in drosophila fos mutants. *EMBO J.* **16**: 7393-7401.

Zhang, K., Chaillet, J.R., Perkins, L.A., Halazonetis, T.D., and Perrimon, N. (1990). Drosophila homolog of the mammalian jun oncogene is expressed during embryonic-development and activates transcription in mammalian-cells. *Proc. Natl. Acad. Sci. U. S. A.* **87**: 6281-6285.

Zubiaga, A.M., Belasco, J.G., and Greenberg, M.E. (1995). The nonamer UUAUUUAUU is the key AU-rich sequence motif that mediates mRNA degradation. *Mol. Cell Biol.* **15**: 2219-2230.

Appendix I

The portion of the GenBank report for human BAC clone RG119C02 containing a putative protein coding sequence which exactly matches NFM is shown below.

```
LOCUS       AC004520   132513 bp    DNA             PRI       01-APR-1998
DEFINITION  Homo sapiens BAC clone RG119C02 from 7p15, complete sequence.
ACCESSION   AC004520
NID         g3004572
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 132513)
  AUTHORS   Dubuque,T., Smith,A., Elliott,G. and Harmon,G.
  TITLE     The sequence of Homo sapiens BAC clone RG119C02
  JOURNAL   Unpublished (1998)
REFERENCE   2  (bases 1 to 132513)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-APR-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108,
USA
COMMENT     SUBMITTED BY:  WUGSC
            Genome Sequencing Center
            Department of Genetics
            Washington University
            St. Louis MO 63108, USA
            http://genome.wustl.edu/gsc
            mailto:sapiens@watson.wustl.edu
SOURCE INFORMATION:
            Clone RG119C02 is from the first release of the human BAC library
            CITB-978SK-B.  The library contains cloned DNA from the male
            fibroblast cell line 978SK.  See:  Shizuya et al., Proc. Natl.
            Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-
8
            (1996).  This clone is available from Research Genetics, Inc.
            (http://www.resgen.com).
            VECTOR:  pBeloBAC11
            Selection:  chloramphenicol

            NEIGHBORING SEQUENCE INFORMATION:
            The actual start of this clone is at base position 1 of RG119C02;
            actual end is at 132513 of RG119C02. The orientation of this clone
            is unknown.

            This clone contains STS sWSS2790 (NID:g1113585).
FEATURES             Location/Qualifiers
     source          1. .132513
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7p15"
                     /clone_lib="CITB-978SK-B"
                     /clone="RG119C02"
gene              complement(6641. .39925)
                     /gene="WUGSC:H_RG119C02.1"
     CDS             complement(join(6641. .7891,8640. .8723,14302. .14481,
                     39356. .39925))
                     /gene="WUGSC:H_RG119C02.1"
                     /note="similar to  NFE2-related transcription factors;
                     similar to I48694 (PID:g2137676); H_RG119C02.1"
```

/codon_start=1
/evidence=not_experimental
/db_xref="PID:g3004573"

/translation="MKHLKRWWSAGGGLLHLTLLLSLAGLRVDLDLYLLLPPPTLLQD
ELLFLGGPASSAYALSPFSASGGWGRAGHLHPKGRELDPAAPPEGQLLREVRALGVPF
VPRTSVDAWLVHSVAAGSADEAHGLLGAAAAASSTGGAGASVDGGSQAVQGGGGDPRAA
RSGPLDAGEEEKAPAEPTAQVPDAGGCASEENGVLREKHEAVDHSSQHEENEERVSAQ
KENSLQQNDDDENKIAEKPDWEAEKTTESRNERHLNGTDTSFSLEDLFQLLSSQPENS
LEGISLGDIPLPGSISDGMNSSAHYHVNFSQAISQDVNLHEAILLCPNNTFRRDPTAR
TSQSQEPFLQLNSHTTNPEQTLPGTNLTGFLSPVDNHMRNLTSQDLLYDLDINIFDEI
NLMSLATEDNFDPIDVSQLFDEPDSDSGLSLDSSHNNTSVIKSNSSHSVCDEGAIGYC
TDHESSSHHDLEGAVGGYYPEPSKLCHLDQSDSDFHGDLTFQHVFHNHTYHLQPTAPE
STSEPFPWPGKSQKIRSRYLEDTDRNLSRDEQRAKALHIPFSVDEIVGMPVDSFNSML
SRYYLTDLQVSLIRDIRRRGKNKVAAQNCRKRKLDIILNLEDDVCNLQAKKETLKREQ
AQCNKAINIMKQKLHDLYHDIFSRLRDDQGRPVNPNHYALQCTHDGSILIVPKELVAS
GHKKETQKGKRK"

Appendix 2

The NFM EST contigs obtained from the TIGR HGI facility (http://www.tigr.org) are shown below (Figure 3.42).

The TIGR Human Gene Index (HGI)

HGI THC Report: THC203963

EST IDs are linked to HGI EST reports. HT# s are linked to EGAD HT reports. GB# s are linked to
GenBank accessions. ATCC#s are linked to order forms for requesting clones.

>THC203963 THC183628
TCGGAGTGGCCCCTTGGACGCCGGGGAAGAGGAGAAGGCACCCGCGGAACCGACGGCTCAGGTGCCGGACGCTGGCGGAn
TGTGCGAGCGAGGAGAATGGGGTACTAAGAGAAAAGCACgAAGCtGTGGATCATAGTTCCCAGCATGAGGAAAATGAAGA
AAGGGTGTCAGCCCAGAAGGAGAACTCACTTCAGCAGAATGATGATGATGAAAACAAAATAGCAGAGAAACCTGACTGGG
AGGCAGAAAAGACCACTGAATCTAGAAATGAGAGACATCTGAATGGGACAGATACTTCTTTCTCTCTGGAAGACTTATTC
CAGTTGCTTTCATCACAGCCTGAAAATTCACTgGAGGGCATCTCaTTgGGGAGATATTCCTCTTCCaGGCAGTATCAGTG
ATGGcATGAATTCTTCAgCACATTaTCATGTAAACTTtCAGCCAGGCTaTAAGTCaGGATGtGAATCTTcATGAAGGnCA
TCnTGGCTTTGTCCCCAAnCAATAGCATTTGGAAGAGnTCCAACAGnCAAGGACTTT

```
 1================================THC203963================================537
 ---------------------------------1---------------------------------->
    ------------------------------------2------------------------------------>
       ------------------------------3------------------------------>
                  -------------------------4------------------------>
```

#	EST Id	GB#	ATCC#	left	right	library
1	F zo20c03.r	AA132584		1	446	Colon, Stratagene (#937204)
2	F zn99f04	AA149705	724743	25	537	Colon, Stratagene (#937204)
3	F zl81a01.r1	AA099770	968098	44	452	Colon, Stratagene (#937204)
4	F zo25h07.r	AA135857		131	473	Colon, Stratagene (#937204)

Sequence source codes:
F = WashU/Merck

There are no hits for THC203963.

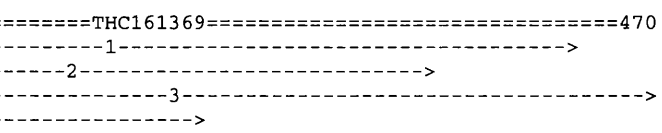
View Opposite End information.

[Image] The TIGR Human Gene Index (HGI)

HGI THC Report: THC161369

EST IDs are linked to HGI EST reports. HT# s are linked to EGAD HT reports.
GB# s are linked to GenBank accessions. ATCC#s are linked to order forms for
requesting clones.

>THC161369 THC66562 THC81180
TTTGATGAGATAAACTTAATGTCATTGGCCACAGAAGACAACTTTGATCCAATCGATGTTTCTCAGCTTTTTGATGAACC
AGATTCTGATTCTGGCCTTTyTTTAGATTCAAGTCACAATAATACCTCTGTCATCAAGTCTAATTCCTCTCACTCTGTGT
GTGATGAAGGTGCTATAGGTTATTGCACTGACCATGAAtCTAGTTCCCATCATGACTTAGAAGGTGCTGTAGGTGGCTAC
TACCCAGAACCCAGTAAGCTTTGTCACTTGGATCAAAGTGATTCTrATTTCCATGGAGATCTTACATTTCAACACGTATT
TCATAACCACACTTACCACTTACAGCCAACTGCACCAGgAATCTACTTCTGgAACCTTTTCCGTGGgCCTGGGaAGtCAC
AGaAGaTaAGGaGTAGaTACCTTGrArGACACAGGATAGGAAACTTTGnGGCCGTGGATGGAACAGCGTG



#		EST Id	GB#	ATCC#	left	right	library
1	F	zl66a11.r1	AA045573	966658	1	434	colon, Stratagene (#937204)
2	F	zs88h06.r	AA279920		27	366	
3	F	yg18g01.r1	R20139	358537	29	470	infant brain 1NIB, Soares
4	A	EST84717	AA372783	177142	42	258	Colon adenocarcinoma IV

Sequence source codes:
F = WashU/Merck
A = TIGR

There are no hits for THC161369.

View Opposite End information.

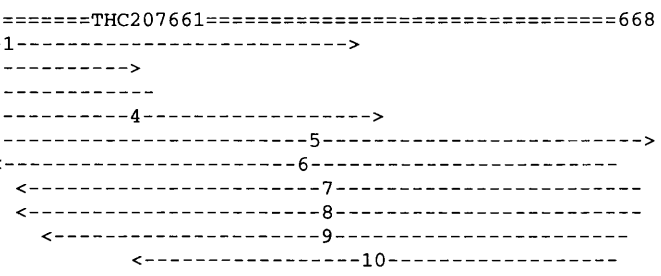
[Image] [Image]

[Image] The TIGR Human Gene Index (HGI)

HGI THC Report: THC207661

EST IDs are linked to HGI EST reports. HT# s are linked to EGAD HT reports.
GB# s are linked to GenBank accessions. ATCC#s are linked to order forms for
requesting clones.

>THC207661 THC44663 THC100162 THC111612 THC181377
GATATTTTTAGTAGATTAAGAGATGACCAAGGTAGGCCAGTCAATCCCAACCACTATGCTCTCCAGTGTACCCATGATGG
AAGTATCTTGATAGTACCCAAAGAACTGGTGGCCTCAGGCAACAAaAGGAAACCCAAAAGGGAAAGAGAAAGTGAGAAGA
AACTGAAGATGGACTCTATTATGTGAAGTAGTAATGTTCAGAAACTGATTATTTGGATCAGAAACCATTGAAACTGCTTC
AAGAATTGTATCTTTAAGTACTGCTACTTGAATAACTCAGTTAACGCTGTTTTGAAGCTTACATGGACAAATGTTTAGGA
CTTCAAGATCaCACTTGTGGGCAATCTGGGGGAGCCACAACTTTTCATGAAGTGCATTGTATACAAAATTCATAGTTATG
TCCAAAGAATAGGTTAACATGAAAACCCAGTAAGACTTTCCATCTTGGCAGCCATCCTTTTTAAGAGTAAGTTGGTTACT
TCAAAAAGAGCAAACACTGGGGATCAAATTATTTTAAGAGGTATTTCAGTTTTAAATGCAAAATAGCCTTATTTTCATTT
AGTTTGTTAGCACTATAGTGAGCTTTTCAAACACTATTTTAATCTTTATATTTAACTTATAAATTTTGCTTTCTATGGAA
ATAAATTTTGTATTTGTATTAAAAATTA



#		EST Id	GB#	ATCC#	left	right	library
1	F	yi82c06.r1	R78318	383651	1	472	placenta Nb2HP, Soares
2	F	ze70d09.r1	AA022505	821451	87	322	fetal heart NbHH19W, Soares
3	A	EST35507	AA331415	133272	133	333	embryo, 8 week I
4	A	EST80449	AA369091	173772	169	488	placenta II
5	A	EST176112	AA305111	128600	227	668	Colon carcinoma (Caco-2) cell line II
6	F	yi82c06.s1	R77947	383651	232	638	placenta Nb2HP, Soares
7	F	zo20c03.s	AA132523		243	663	colon, Stratagene (#937204)
8	F	zr18a05.s1	AA227467	232137	248	660	neuronal precursor NT2, Stratagene (#937230)
9	F	yl32g06.s1	H21401	397907	265	652	breast 3NbHBst, Soares
10	F	yh89c02.s1	R36639	374811	321	638	placenta Nb2HP, Soares

Sequence source codes:
F = WashU/Merck
A = TIGR

There are no hits for THC207661.

View Opposite End information.

[Image] [Image]

[Image] The TIGR Human Gene Index (HGI)

HGI THC Report: THC210211

EST IDs are linked to HGI EST reports. HT# s are linked to EGAD HT reports.
GB# s are linked to GenBank accessions. ATCC#s are linked to order forms for
requesting clones.

>THC210211 THC187372
TnTTAcAGGTCTTGGGAAACTTTATTATAAAAATCCAACGGTATATATAAACTGCATTTCTCTAGCCCAAGATACTTATG
TGATACTCTACACTGTAGCTCCTATGGCAGAAATAATATCTGAGAAACTTCATTTTGAAAATTTGGACATATTAAGCATT
ATAGTACCCGTCAAAATCTTGATGTTCCCCAAACTTGAGTCATCTAGGTGAGGTCATTGCTGTCTTCAAAAAATTCAGCA
ATCTCCTGTTTACTGtTCTAGTGAACTTAAgTGAgGTTTTCATTTCAAATGCTAtTATTTCTAGrATTTaCCTTACCCTG
GGTtTAtTTTCAAATTCACCTGTkyTTTTTCCCTTAATCGCTTTtGCCCATTTTAAACAgATGGTTTCACTGtTGCTgAC
TCTTTCCCCnACTGnTTTATATGGCTTACAGTGGTCTGGGTGA

 1==================================THC210211==================================443
 ------------------------------------1----------------------------------->
 --2----------------------------------->

#		EST Id	GB#	ATCC#	left	right	library
1	F	zl66a11.s1	AA045574	966658	1	408	colon, Stratagene (#937204)
2	F	ze70d09.s1	AA022506	821451	4	443	fetal heart NbHH19W, Soares

Sequence source codes:
F = WashU/Merck

There are no hits for THC210211.

View Opposite End information.

[Image] [Image]

Appendix 3

The mouse protein NFE2l1 was analysed using the programme MOTIFS (part of the GCG collection at UK-HGMP Resource Centre, UK). The results are detailed below.

```
************************
* N-glycosylation site *
************************

Asn_Glycosylation     N~(P)(S,T)~(P)
                         N~P(S)~P
           300: QAMEV      NTSA      SEILY

                         N~P(S)~P
           319: DPLSS      NYSL      APNTP

                         N~P(S)~P
           331: TPINQ      NVSL      HQASL

                         N~P(T)~P
           371: PLVPS      NSTS      LNSTF

                         N~P(T)~P
           376: NSTSL      NSTF      GSTNL

                         N~P(T)~P
           394: FPSQL      NGTA      NDTSG

                         N~P(T)~P
           398: LNGTA      NDTS      GPELP

                         N~P(T)~P
           543: EHVGH      NHTY      NMAPS
```

```
*****************************************************
* bZIP transcription factors basic domain signature *
*****************************************************

Bzip_Basic
(K,R)x{1,3}(R,K,S,A,Q)Nx2(S,A,Q)2x(R,K,T,A,E,N,Q)xRx(R,K)
                                  (R)x{3}(K)Nx{2}(A){2}x(N)xRx(R)
           628: LIRDI                    RRRGKNKMAAQNCRKR
KLDTI

                                  (R)x{2}(K)Nx{2}(A){2}x(N)xRx(R)
           629: IRDIR                     RRGKNKMAAQNCRKR
KLDTI

                                   (R)x(K)Nx{2}(A){2}x(N)xRx(R)
           630: RDIRR                     RGKNKMAAQNCRKR
KLDTI
```

```
**********************************************************************
* cAMP- and cGMP-dependent protein kinase phosphorylation site *
**********************************************************************

Camp_Phospho_Site      (R,K)2x(S,T)
                        (K){2}x(S)
            565: PPSTL     KKGS     KEKQA
```

```
****************************************
* Casein kinase II phosphorylation site *
****************************************

Ck2_Phospho_Site       (S,T)x2(D,E)
                        (T)x{2}(D)
            134: GLQDV     TGPD     NGVRE

                        (S)x{2}(E)
            143: NGVRE     SETE     QGFGE

                        (T)x{2}(D)
            168: VSGDL     TKED     IDLID

                        (S)x{2}(E)
            249: PADVS     SIPE     AVPSE

                        (S)x{2}(E)
            256: PEAVP     SESE     SPALQ

                        (T)x{2}(E)
            273: LSPLL     TGTE     SPFDL

                        (S)x{2}(D)
            277: LTGTE     SPFD     LEQQW

                        (S)x{2}(E)
            290: WQDLM     SIME     MQAME

                        (S)x{2}(E)
            302: MEVNT     SASE     ILYNA

                        (T)x{2}(D)
            396: SQLNG     TAND     TSGPE

                        (S)x{2}(E)
            401: TANDT     SGPE     LPDPL

                        (S)x{2}(D)
            422: MLDEI     SLMD     LAIEE

                        (S)x{2}(E)
            438: NPVQA     SQLE     EEFDS

                        (S)x{2}(E)
            462: SPSSL     SSSE     GSSSS

                        (S)x{2}(E)
            486: SSASS     SFSE     EGAVG

                        (S)x{2}(E)
            497: AVGYS     SDSE     TLDLE

                        (S)x{2}(D)
            524: KFCRM     SYQD     PSQLS
```

```
                (S)x{2}(D)
          551: YNMAP      SALD      SADLP

                (S)x{2}(E)
          581: LDKQM      SRDE      HRARA

                (T)x{2}(D)
          723: LLIPR      TMAD      QQARR
```

```
************************************
* Glycosaminoglycan attachment site *
************************************

Glycosaminoglycan     SGxG
          218: REDTW SGEG AEALA
```

```
*************************
* N-myristoylation site *
*************************

Myristyl              G~(E,D,R,K,H,P,F,Y,W)x2(S,T,A,G,C,N)~(P)
                          G~(E,D,R,K,H,P,F,Y,W)x{2}(G)~P
          111: HRDPE                    GSVSGS                    QPNSG

                          G~(E,D,R,K,H,P,F,Y,W)x{2}(N)~P
          115: EGSVS                    GSQPNS                    GLALE

                          G~(E,D,R,K,H,P,F,Y,W)x{2}(S)~P
          139: TGPDN                    GVRESE                    TEQGF

                          G~(E,D,R,K,H,P,F,Y,W)x{2}(S)~P
          466: LSSSE                    GSSSSS                    SSSSS
```

```
****************************************
* Protein kinase C phosphorylation site *
****************************************

Pkc_Phospho_Site      (S,T)x(R,K)
                        (S)x(K)
            3:    ML      SLK      KYLTE

                        (T)x(R)
           79: LDNYF      TAR      RLLSQ

                        (S)x(R)
          195: EVFDY      SHR      QKEQD

                        (T)x(K)
          563: LPPPS      TLK      KGSKE

                        (S)x(R)
          675: VEFLR      SLR      QMKQK
```

```
****************************************
* Tyrosine kinase phosphorylation site *
****************************************

Tyr_Phospho_Site      (R,K)x{2,3}(D,E)x{2,3}Y
                          (K)x{2}(D)x{3}Y
           70: YGIHP         KSIDLDNY          FTARR
```

Appendix 4

The human protein NFM was analysed using the programme MOTIFS (part of the GCG collection at UK-HGMP Resource Centre, UK). The results are detailed below.

```
******************
* Amidation site *
******************

Amidation              xG(R,K)(R,K)
                         xG(K)(R)
           690: KKETQ      KGKR      K
```

```
************************
* N-glycosylation site *
************************

Asn_Glycosylation      N~(P)(S,T)~(P)
                          N~P(T)~P
           254: NERHL       NGTD       TSFSL

                          N~P(S)~P
           296: ISDGM       NSSA       HYHVN

                          N~P(S)~P
           304: AHYHV       NFSQ       AISQD

                          N~P(T)~P
           324: ILLCP       NNTF       RRDPT

                          N~P(T)~P
           360: TLPGT       NLTG       FLSPV

                          N~P(T)~P
           374: DNHMR       NLTS       QDLLY

                          N~P(T)~P
           428: LDSSH       NNTS       VIKSN

                          N~P(S)~P
           429: DSSHN       NTSV       IKSNS

                          N~P(S)~P
           436: SVIKS       NSSH       SVCDE

                          N~P(T)~P
           497: QHVFH       NHTY       HLQPT

                          N~P(S)~P
           534: EDTDR       NLSR       DEQRA
```

```
*****************************************************************
* cAMP- and cGMP-dependent protein kinase phosphorylation site *
*****************************************************************

Camp_Phospho_Site      (R,K)2x(S,T)
                        (K){2}x(T)
            616: CNLQA     KKET      LKREQ

                        (K){2}x(T)
            685: VASGH     KKET      QKGKR
```

```
****************************************
* Casein kinase II phosphorylation site *
****************************************

Ck2_Phospho_Site       (S,T)x2(D,E)
                        (T)x{2}(D)
            106: PFVPR     TSVD      AWLVH

                        (S)x{2}(E)
            120: SVAAG     SADE      AHGLL

                        (S)x{2}(E)
            206: AVDHS     SQHE      ENEER

                        (S)x{2}(E)
            247: EKTTE     SRNE      RHLNG

                        (S)x{2}(D)
            261: TDTSF     SLED      LFQLL

                        (S)x{2}(E)
            271: FQLLS     SQPE      NSLEG

                        (S)x{2}(D)
            281: SLEGI     SLGD      IPLPG

                        (S)x{2}(D)
            290: IPLPG     SISD      GMNSS

                        (T)x{2}(E)
            350: LNSHT     TNPE      QTLPG

                        (S)x{2}(D)
            366: LTGFL     SPVD      NHMRN

                        (T)x{2}(D)
            376: HMRNL     TSQD      LLYDL

                        (S)x{2}(D)
            440: SNSSH     SVCD      EGAIG

                        (T)x{2}(E)
            451: AIGYC     TDHE      SSSHH

                        (S)x{2}(D)
            457: DHESS     SHHD      LEGAV

                        (S)x{2}(D)
            481: CHLDQ     SDSD      FHGDL

                        (T)x{2}(E)
            505: YHLQP     TAPE      STSEP
```

```
                (S)x{2}(E)
509: PTAPE      STSE      PFPWP

                (S)x{2}(E)
536: TDRNL      SRDE      QRAKA

                (S)x{2}(E)
550: LHIPF      SVDE      IVGMP
```

```
*************************
* N-myristoylation site *
*************************

Myristyl              G~(E,D,R,K,H,P,F,Y,W)x2(S,T,A,G,C,N)~(P)
                          G~(E,D,R,K,H,P,F,Y,W)x{2}(S)~P
 50: ELLFL                        GGPASS                      AYALS

                          G~(E,D,R,K,H,P,F,Y,W)x{2}(A)~P
126: ADEAH                        GLLGAA                      AASST

                          G~(E,D,R,K,H,P,F,Y,W)x{2}(A)~P
129: AHGLL                        GAAAAS                      STGGA


                          G~(E,D,R,K,H,P,F,Y,W)x{2}(A)~P
137: AASST                        GGAGAS                      VDGGS

                          G~(E,D,R,K,H,P,F,Y,W)x{2}(S)~P
138: ASSTG                        GAGASV                      DGGSQ

                          G~(E,D,R,K,H,P,F,Y,W)x{2}(A)~P
145: GASVD                        GGSQAV                      QGGGG

                          G~(E,D,R,K,H,P,F,Y,W)x{2}(S)~P
185: QVPDA                        GGCASE                      ENGVL

                          G~(E,D,R,K,H,P,F,Y,W)x{2}(S)~P
255: ERHLN                        GTDTSF                      SLEDL

                          G~(E,D,R,K,H,P,F,Y,W)x{2}(G)~P
279: ENSLE                        GISLGD                      IPLPG


                          G~(E,D,R,K,H,P,F,Y,W)x{2}(S)~P
294: GSISD                        GMNSSA                      HYHVN

                          G~(E,D,R,K,H,P,F,Y,W)x{2}(T)~P
358: EQTLP                        GTNLTG                      FLSPV

                          G~(E,D,R,K,H,P,F,Y,W)x{2}(G)~P
463: HHDLE                        GAVGGY                      YPEPS
```

```
*************************************
* Protein kinase C phosphorylation site *
*************************************

Pkc_Phospho_Site      (S,T)x(R,K)
                        (T)x(R)
326: LCPNN      TFR     RDPTA
```

```
                (T)x(R)
332: FRRDP       TAR      TSQSQ

                (S)x(K)
520: PWPGK       SQK      IRSRY

                (T)x(R)
531: RYLED       TDR      NLSRD

                (T)x(K)
619: QAKKE       TLK      REQAQ

                (T)x(K)
688: GHKKE       TQK      GKRK
```

PUBLICATIONS

Brief Data Reports

Mammalian Genome

Mammalian Genome 7, 89–90 (1996).

The *Nfe2l1* gene maps to distal mouse Chromosome 11

J. McKie, P.J. Scambler

Molecular Medicine Unit, Institute of Child Health, 30, Guilford Street, London WC1N 1EH, UK

Received: 1 May 1995 / Accepted: 1 September 1995

Species: Mouse
Locus name: Nfe2-like-gene, 1
Locus symbol: *Nfe2l1*
Map position: Centromere–*D11Mit36*–(13.89 ± 5.76)–*Nfe2l1*–(11.11 ± 5.24)–*D11Mit10*
Method of mapping: *Nfe2l1* was localized by haplotype analysis of 36 progeny from an interspecific backcross, (C57BL/6J × *M. spretus*)F_1 × C57BL/6J, carried out by the European Collaborative Interspecific Backcross (EUCIB) and provided by the UK-HGMP Resource Centre [1].
Database deposit information: Detailed map of Chromosome (Chr) 11 localization available via world wide web (http:/www.hgmp.mrc.ac.uk/local-data/mbx/Mbx_Homepage.html).
Molecular reagents: A 1.8-kb cDA fragment [2] corresponding to amino acids 13 through 606 was labeled by random priming and hybridized to Southern blots from the UK-HGMP Resource Centre.
Allele detection: Allele detection was performed by RFLP analysis of a *Taq*I polymorphism. Single bands were detected: the *M. spretus*-specific allele was 1.8 kb and the C57BL/6J-specific allele was 7.0 kb.
Previously identified homologs: The human homolog maps to 17q22 by FISH [3, and our unpublished data]. Two related sequences map to mouse Chr 2E4-2G and 7D1-F1 [2] by radioactive *in situ* hybridization. Murine *Nfe2* maps to mouse Chr 15 [4].
Discussion: NFE2L1 was cloned in a complementation assay designed to identify proteins that activate transcription through the NFE2/AP1 enhancer sequence [5]. The gene encodes a basic leucine zipper (bZIP) protein which shares sequence similarity with p45NF-E2 [6], the CNC protein from *Drosophila melanogaster* [7] and skn-1 from *Caenorhabditis elegans* [8]. Nfe2l1 is widely expressed in adult, fetal and embryonic tissues [5, and our unpublished data]. NF-E2 is believed to be an important regulator of globin gene expression via the Locus Control Region (LCR) [6], and CNC is involved in the specification of head segment identity in the developing fruitfly [9]. One isoform of NFE2l1 may have a role in LCR-mediated human globin gene expression [10], and potential NF-E2 binding sites have been noted in the promoter region of some genes involved in heme biosynthesis and iron metabolism [6].

Acknowledgments: This work was supported by the Dunhill Medical Research Trust and the special trustees at the Hospital for Sick Children, Great Ormand St., London, UK.

Correspondence to: J. McKie

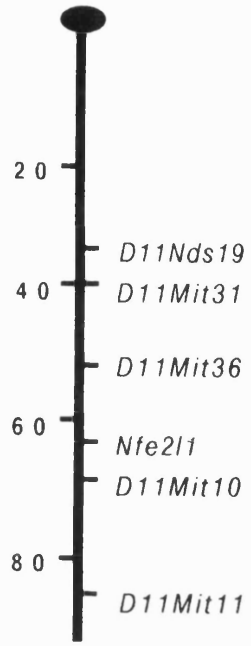


Fig. 1. The localization of *Nfe2l1* on the Mbx map of mouse Chr 11 [1]. Genetic distances are in cM (left).

Table 1. Interspecific backcross mapping data for *Nfe2l1* provided by UK-HGMP Resource Centre.

Locus 1	Locus 2	Number of recombinants	Genetic distances (cM)
Nfe2l1	*D11Mit10*	4/36	11.11 ± 5.24
Nfe2l1	*D11Mit36*	5/36	13.89 ± 5.76
Nfe2l1	*D11Mit31*	5/36	13.89 ± 5.76
Nfe2l1	*D11Nds19*	8/36	22.22 ± 6.93
Nfe2l1	*D11Mit11*	8/36	24.24 ± 7.46

References

1. Breen et al. (1994). Hum. Mol. Genet. 3, 621–627.
2. McKie, J.M., Johnstone, K., Mattei, M-G., Scambler, P.J. (1995). Genomics 25, 716–719.
3. Luna, L., Johnsen, Ø., Skartlien, A., Pedeutour, F., Turc-Carel, C., Prydz, H., Kolstø, A-B. (1994). Genomics 22, 553–562.
4. Peters, L.L., Andrews, N.C., Eicher, E.M., Davidson, M.B., Orkin, S.H., Lux, S.E. (1993). Nature 362, 768–770.
5. Chan, J.Y., Han, X-L., Kan, Y.W. (1993). Proc. Natl. Acad. Sci. USA 90, 11371–11375.
6. Andrews, N.C., Erdjument-Bromage, H., Davidson, M.B., Tempst, P., Orkin, S.H. (1993). Nature 362, 722–728.
7. Mohler, J., Vani, K., Leung, S., Epstein, A. (1991). Mech. Dev. 34, 3–10.
8. Bowerman, B., Eaton, B.A., Priess, J.R. (1992). Cell 68, 1061–1075.
9. Mohler, J., Mahaffey, J.W., Deutsch, E., Vani, K. (1995). Development 121, 237–247.
10. Caterina, J.J., Donze, D., Sun, C.-W., Ciavatta, D.J., Townes, T.M. (1994). Nucleic Acids Res. 22, 2383–2391.

SHORT COMMUNICATION

Cloning and Mapping of Murine *Nfe2l1*

JUDITH MCKIE,*,[1] KAREN JOHNSTONE,* MARIE-GENEVIÈVE MATTEI,† AND PETER SCAMBLER*

*Molecular Medicine Unit, Institute of Child Health, 30 Guilford Street, London WC1N 1EH, United Kingdom; and †INSERM U406, Génétique Médical et Dévelopément, Faculté de Médecine, 27 Boulevard Jean Moulin, Marseille 13385 Cedex 5, France

Received June 30, 1994; revised October 26, 1994

The murine homologue of the human NFE2L1 basic leucine-zipper gene was isolated from an early embryo library. The deduced amino acid sequence shows 97% identity between the two proteins. Significant sequence similarity is also seen with the p45 subunit of NF-E2 and with the *Drosophila* CNC protein. Murine *Nfe2l1* maps to chromosome 11DE with similar sequences at 7D1-7F1 and 2E4-2G. © 1995 Academic Press, Inc.

The NFE2L1 gene encodes a protein with a b-ZIP motif and strong sequence similarity to the p45 NF-E2 transcription factor. p45 NF-E2 is required for activity at the locus control region (LCR) upstream of the globin gene complexes (2). The human NFE2L1 gene was isolated using a complementation assay designed to detect proteins activating transcription at the NF-E2/AP1 binding site (4). In addition to the globin LCRs, NF-E2 binding sites are also found within the promoters of several other genes involved in erythroid gene expression, including enzymes of heme biosynthesis, ferrochetalase, ferritin, and porphobilinogen deaminase (2). p45 NF-E2 is the tissue-specific component of the NF-E2 complex, being expressed only in erythroid and erythroid-progenitor cell lines, liver, spleen, and bone marrow. p45 NF-E2 interacts with a ubiquitously expressed b-ZIP protein, p18 NF-E2 (3), which has sequence similarity to NRL and v-*maf*.

Both p45 NF-E2 and NFE2L1 are more closely related to the *Drosophila* CNC protein than to any other mammalian b-ZIP protein (2, 4, 5). CNC was described as a gene expressed in a segmental pattern during early *Drosophila* embryo development, and it was suggested that the gene may have a role in specification of head structures (9). While pursuing an RT PCR-based gene isolation protocol, we fortuitously isolated a short cDNA clone (8di) with sequence similarity to the b-ZIP region of CNC. An 8.5-day pc and a 10.5-day pc whole mouse embryo library were screened with 8di. Further sequence analysis showed this CNC-like gene to be the murine homologue of the recently described NFE2L1 gene (originally called NRF1) (4).

The sequence data reported in this paper have been deposited with the EMBL database under Accession No. X78709.

[1] To whom correspondence should be addressed. Telephone: 44 71 242 9789 ext. 2275. Fax: 44 71 831 0488.

cDNAs were sequenced using the dideoxy chain termination method (12) and primer walking. The DNA sequence was translated using the GCG package and the derived protein sequence used to search the NBRF and Swissprot databases using BLAST (1), BLOCKS (14), and MPsrch programs (developed by S. Sturrock and J. F. Collins). Phosphorylation sites were predicted using the Prosearch program developed by F. Kolakowski. All programs are available at the HGMP resource center (Northwick Park, Harrow, UK).

The *Nfe2l1* sequence is presented in Fig. 1. There is an initiation codon with a good Kozak consensus sequence (6) at nucleotides 42–48, with a subsequent open reading frame (ORF) of 741 amino acids. The predicted molecular mass of this ORF is 81.5 kDa; p*I* is predicted to be 7.2. *Nfe2l1* has a b-ZIP domain toward the C-terminus of the protein with six heptad repeats of leucine and hydrophobic residues, preceded by an arginine- and lysine-rich basic domain. The predicted protein had several other features compatible with a role as a transcription factor: (i) an acidic domain at amino acids 414–447, (ii) six occurrences of the sequence S–P–X–X (relatively overrepresented in such proteins (13)), (iii) serine/threonine-rich stretches and serine repeats, and (iv) a potential nuclear localization domain at amino acids 730–737 (**RRQERKPK**, basic amino acids in boldface).

Database searches revealed strong sequence similarities with human NFE2L1 (97% identity over the full length of the derived protein sequences) and p45 NF-E2 (79% similarity over the 142 amino acids containing the CNC match). This strongly suggests that the gene that we have isolated is the murine homologue of human NFE2L1.

Several features of the derived mouse and human NFE2L1 sequences are worthy of comment. Conservation is 100% in the basic and leucine zipper regions (human and murine p45 NF-E2 have three amino acid differences in the leucine zipper domain). Conservation is also 100% in the acidic domain, which may be involved in activation of RNA polymerase II transcription factors (8), and in an 11-aa section at amino acids 442–452 (partially overlapping with the acidic domain and immediately N-terminal of the serine repeat). The 11-aa domain and an area immediately N-terminal to the basic region are also conserved in human and murine

0888-7543/95 $6.00

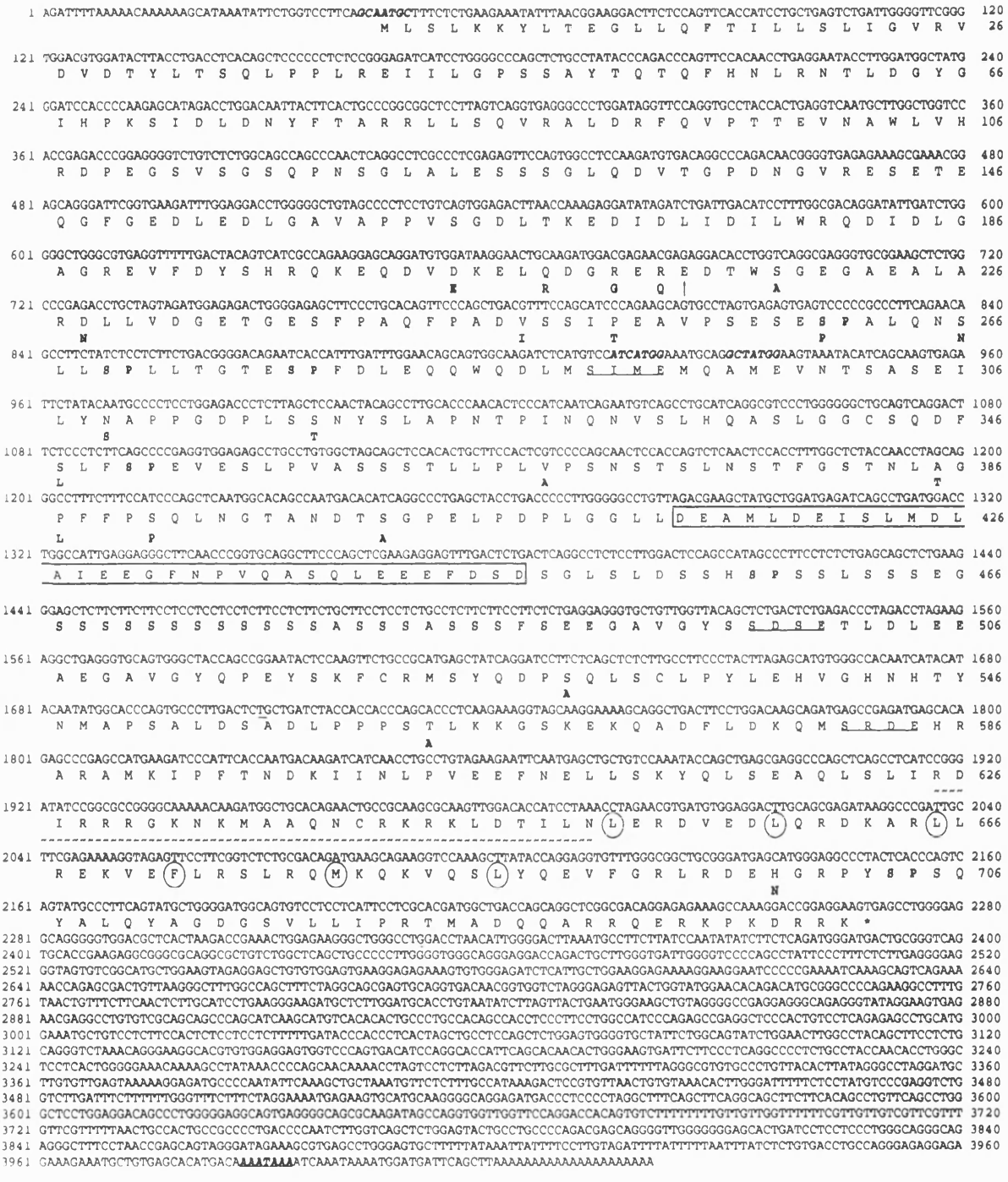


FIG. 1. Nucleotide sequence and deduced amino acid sequence of murine *Nfe2l1*. The basic region is highlighted by a dashed underline, and the leucine/hydrophobic heptad repeats are circled. The acidic domain is boxed. Potential CK2 phosphorylation sites conserved in both mouse and human NFE2L1 and p45 NF-E2 are solid underlined. Kozak consensus sequences are boldfaced and italicized. SP couplets are shown in boldface. Amino acid substitutions in human NFE2L1 appear in boldface beneath their murine counterpart, and an extra murine amino acid is indicated with an arrow.

p45 NF-E2. The suggestion that these regions might have functional significance (5) is thereby strengthened.

Immunoblot experiments with antibodies raised against NFE2L1 detected a 65-kDa product in K562 erythroleukemia cells, implying that the first translation initiation codon is not used or that the protein is cleaved following translation (4). In both human and murine *Nfe2l1* there are two internal Kozak consensus sequences that could serve to initiate synthesis of the 65-kDa protein. Examination of the Kozak consensus sequences in murine *Nfe2l1* shows that the first of the two internal sites is a better match than the second, which has a T at position −1. More importantly, since the region between the first initiation site proposed for human NFE2L1 (but apparently not used in K562 cells)

and the second initiation site is strongly conserved at the amino acid and nucleotide levels, it is very likely that this segment of the protein is functionally important. The absence of any in-frame termination codon in the murine sequence increases the likelihood that the first initiation site will be functional in some tissues.

The serine-rich region and poly-serine stretch are also present in both proteins, with the sole difference being two serines fewer in the murine sequence. Serine residues may be important in providing a polar interface for protein/protein interactions important in transcriptional regulators (10).

Murine *Nfe2l1* has many potential phosphorylation sites as determined by the program Prosearch. Twenty potential creatine kinase 2 (CK2) sites were detected (19 conserved in the human sequence), with 5 protein kinase C sites (4 conserved in the human sequence). The sequence similarity between p45 NF-E2 and NFE2L1 has been presented previously (5). Three potential CK2 sites are present in both human and murine p45 NF-E2 and both human and murine NFE2L1 (highlighted in Fig. 1). Human and murine *Nfe2l1* have long 3′ untranslated regions, but there is little DNA sequence conservation between the two.

p45 NF-E2 interacts with p18 NF-E2 to modulate the expression of erythroid-specific genes such as the globins and heme biosynthesis enzymes (2, 3). In particular, NF-E2 acts at motifs within the globin LCR, and expression of the p45 component is limited to hematopoietic lineages. As shown previously (4) and in our unpublished work, *Nfe2l1* is widely expressed, but the DNA binding specificity of *Nfe2l1* is yet to be determined. However, human NFE2L1 will activate transcription through a tandem NF-E2/AP1 repeat (4), and there is a striking degree of conservation of protein sequence between the basic (DNA-binding) domains of p45 NF-E2 and *Nfe2l1* in both mice and humans. It therefore seems likely that *Nfe2l1* will bind to sequences similar to those bound by NF-E2. p45 NF-E2 does not bind to the NF-E2 motif as a homodimer, the p18 subunit being required for this activity. Since both human and murine *Nfe2l1* have charged residues at positions within the leucine zipper that are likely to cause electrostatic repulsion in homodimeric units (5), it is thought that *Nfe2l1* will also act as an obligate heterodimer.

In situ hybridization experiments were carried out using metaphase spreads from a WMP male mouse, in which all of the autosomes except 19 were in the form of metacentric Robertsonian translocations. The WMP mouse was a gift from J. L. Guenet, Pasteur Institut, Paris. A 2.2-kb fragment from the 3′ end of the *Nfe2l1* cDNA was nick-translated with a tritium label and hybridized to the metaphase spreads as described previously (7). In the 200 metaphase cells examined there were 459 silver grains associated with chromosomes, and 3 hybridization peaks were detectable. The greatest number of grains was on chromosome 11, with two other peaks of grains on chromosomes 2 and 7 (Fig. 2).

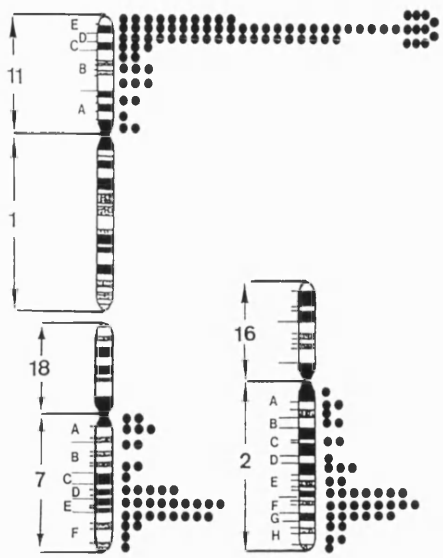


FIG. 2. Radioactive *in situ* hybridization of murine *Nfe2l1* cDNA. The figure shows the number of silver grains detected over different regions of mouse chromosomes 11, 7, and 2.

In total 80 grains were located on chromosome 11 (17.4%). The distribution of grains was nonrandom, with 64/80 (80%) clustered in the 11D–E region. Forty-one grains were located on chromosome 2 (9%), with a nonrandom distribution: 21/41 (51%) mapped to the 2E4–G region. The third peak of 36 grains was located on chromosome 7 (8%). Again, the distribution of grains was nonrandom, with a cluster of 22 grains (61%) mapping to the 7D1–F1 region. None of the murine hybridization peaks corresponds with the locus of NF-E2 on mouse chromosome 15 (11). These data suggest the existence of an extended family of mammalian b-ZIP proteins with sequence similarity to *Nfe2l1*.

ACKNOWLEDGMENTS

This work was supported by the Dunhill Medical Research Trust, and the special trustees at the Hospital for Sick Children, Gt. Ormond St.

REFERENCES

1. Altschul, S. F., Gish, W., Miller, W., Myer, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410.
2. Andrews, N. C., Erdjument-Bromage, H., Davidson, M. B., Tempst, P., and Orkin, S. H. (1993). Erythroid transcription factor NF-E2 is a haematopoietic-specific basic-leucine zipper protein. *Nature* **362:** 722–728.
3. Andrews, N. C., Kotkow, K. J., Ney, P. A., Erdjument-Bromage, H., Tempst, P., and Orkin, S. H. (1993). The ubiquitous subunit of erythroid transcription factor NF-E2 is a small basic-leucine zipper protein related to the v-maf oncogene. *Proc. Natl. Acad. Sci. USA* **90:** 11488–11492.
4. Chan, J. Y., Han, X.-L., and Kan, Y. W. (1993). Cloning of Nfr1, an NF-E2-related transcription factor by genetic selection in yeast. *Proc. Natl. Acad. Sci. USA* **90:** 11371–11375.

5. Chan, J. Y., Han, X.-L., and Kan, Y. W. (1993). Isolation of cDNA encoding the human NF-E2 protein. *Proc. Natl. Acad. Sci. USA* **90:** 11366–11370.
6. Kozak, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44:** 283–292.
7. Mattei, M. G., Philip, N., Passage, E., Moisan, J. P., Mandel, J. L., and Mattei, J. F. (1985). DNA probe localisation at 18p113 band by in situ hybridisation and identification of a small supernumerary chromosome. *Hum. Genet.* **69:** 268–271.
8. Mitchell, P. J., and Tjian, R. (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* **245:** 371–381.
9. Mohler, J., Vani, K., Leung, S., and Epstein, A. (1991). Segmentally restricted, cephalic expression of a leucine zipper gene during Drosophila embryogenesis. *Mech. Dev.* **34:** 3–10.
10. Pascal, E., and Tjian, R. (1991). Different activation domains of Sp1 govern formation of multimers and mediate transcriptional synergism. *Genes Dev.* **5:** 1646–1656.
11. Peters, L. L., Andrews, N. C., Eicher, E. M., Davidson, M. B., Orkin, S. H., and Lux, S. E. (1993). Mouse microcytic anaemia caused by a defect in the gene encoding the globin enhancer-binding protein NF-E2. *Nature* **362:** 768–770.
12. Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74:** 5463–5467.
13. Suzuki, M. (1989). SPXX, a frequent sequence motif in gene regulatory proteins. *J. Mol. Biol.* **207:** 61–84.
14. Wallace, J. C., Agus, S., Henikoff, J. G., and Henikoff, S. (1991). Automated assembly of protein blocks for database searching. *Nucleic Acids Res.* **19:** 6565–6572.