

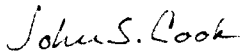
To the Graduate Council:

I am submitting herewith a dissertation written by Mitchell J Walkowicz entitled "Mutations within the *Ery1* transcriptional unit are associated with juvenile lethality, neuromuscular tremors and germ cell defects in *jdf2* mutant mice and pigmentation abnormalities in p^x and p^m alleles." I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Biomedical Sciences.

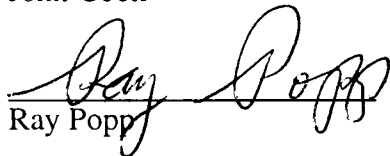


Lisa Stubbs, Major Professor

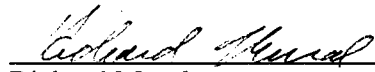
We have read this dissertation
and recommend it acceptance:



John Cook



Ray Popp



Richard Mural

Accepted for the Council:



Associate Vice Chancellor and
Dean of The Graduate School

Mutations within the *Ery1* transcriptional unit are associated with juvenile lethality, neuromuscular tremors and germ cell defects in *jdf2* mutant mice and pigmentation abnormalities in p^x and p^m alleles.

A Dissertation Presented for the Doctor of Philosophy Degree

The University of Tennessee, Knoxville

Mitchell John Walkowicz

August 1998

This dissertation is dedicated to the unparalleled support
of my family and friends.

“eternity comes down to hours, and a minute can make a
tough guy a coward”

Leo Kottke

ACKNOWLEDGMENTS

I would like to thank the members of my dissertation committee; John Cook, Richard Mural, Ray Pop and my major professor Lisa Stubbs particularly for her help in editing this manuscript. I am also indebted to Lee Russell and Gene Rinchik for providing such a wonderful problem as well as providing the reagents which allowed its resolution. Finally I would like to personally thank all the folks in the 'mouse house' and 'Biology' as they once were, for making my stay in Oak Ridge all the better. See you in the funny papers.....!

Bascnosc, Bascnosc, Professor yea twoje mlode.

Abstract

This dissertation describes the cloning of a large gene, called *Eryl*, and presents evidence to link mutations in this gene to pleiotropic affects associated with certain deletions in the mouse pink-eyed dilution (*p*) region of mouse chromosome 7. This complex phenotype, called the 'pink-eye effect' includes reduced-life span, severe runting, neurologic symptoms, and a characteristic sperm cell defect. The identification of several EtNU-induced alleles that fail to complement defects associated with *p*^{6H}, a *p* deletion allele historically associated with the 'pink eye effect', has indicated that this complex phenotype is due to mutations in a single locus, termed juvenile development and fertility locus 2 (*jdf2*).

Here we report the construction of a physical map of the *jdf2* region and cloning of a highly conserved gene, called *Eryl* (End Repeat family 1). Alterations of genomic structure and/or expression of *Eryl* were detected in seventeen of eighteen *jdf2* alleles examined, providing strong evidence that defective *Eryl* function is the cause of the pink-eye/*jdf2* effect in mutant mice. In the course of these studies we have also identified rearrangements that permit us to predict the location of sequences associated with intermediate(*p*^x) and mottled (*p*^m) *p* pigmentation defects in several mutant alleles.

In a collaborative effort, we have cloned and sequenced the complete >15 kb *Eryl* cDNA sequence. The predicted protein contains five recognized protein motifs including three complete copies of a motif related to RCC1, a protein associated with defects in nuclear transport, structure and cell cycle control. *Eryl* also contains a "ZZ"-type zinc finger structure and a HECT domain, the latter of which has been proposed for a role in the ubiquitin intracellular signaling pathway. *Eryl* may act as a guanine nucleotide exchange factor (GEF), as documented for p532, a related HECT domain-containing protein. Alternatively, I propose that *Eryl*'s essential function is in the ubiquitin tagging or degradation pathway, and that the pleiotropic phenotype expressed by *jdf2* mutant mice reflects the involvement of this essential gene at a crosspoint between several fundamental cellular processes.

TABLE OF CONTENTS

<u>SECTION</u>	<u>PAGE #</u>
CHAPTER 1. INTRODUCTION	1
SECTION 1. THE MOUSE PINK-EYE DILUTION LOCUS	4
ADDITIONAL PHENOTYPES ASSOCIATED WITH CERTAIN <i>p</i> ALLELES: THE 'PINK-EYE SYNDROME'	7
PITUITARY AXIS DEFECTS AS A COMMON CAUSE FOR PLEIOTROPIC PINK-EYE SYNDROME.	10
COMPLEMENTATION ANALYSIS OF OAK RIDGE <i>p</i> LOCUS MUTATIONS.	15
PINK-EYE SYNDROME IS DUE TO MUTATION OF A SINGLE GENE DESIGNATED <i>jdf2</i> .	18
DESCRIPTION OF RADIATION -INDUCED AND 'OTHER' ALLELES WHICH DEFINE THE PINK-EYE/ <i>jdf2</i> SYNDROME.	21
DESCRIPTION OF KNOWN LOCI PROXIMAL TO <i>p</i> .	22
SPECIFIC AIMS AND SUMMARY OF RESULTS	26
OBSERVATIONS ON THE PATHOLOGY OF THE <i>jdf2</i> PHENOTYPE.	33
CHAPTER 2. MATERIALS AND METHODS	37
MOUSE BREEDING	37

CDNA SUBTRACTION	3 8
MOLECULAR PROBES SEQUENCING AND HYBRIDIZATIONS.	3 9
GENOMIC CLONE ISOLATION AND CHARACTERIZATION	4 0
RNA ISOLATION, NORTHERN BLOTTING, AND RNASE PROTECTION.	4 2
SUBCLONING AND SEQUENCING <i>mupc7e8.5</i> AND <i>jsub5.1</i> cDNAS.	4 3
CHAPTER 3. RESULTS	4 5
SECTION 1. DERIVATION OF MOUSE PROBES FOR THE <i>D7H15F37S1</i> AND <i>p</i> LOCI AND ANALYSIS OF SELECTED <i>p</i> ALLELES.	4 7
SECTION 2. CONSTRUCTION OF A BACTERIOPHAGE P1 CLONE CONTIG SPANNING THE <i>jdf2</i> INTERVAL.	5 6
SECTION 3. ISOLATION OF cDNAS FROM THE <i>D7H15F37S1-p</i> INTERVAL.	5 9
SECTION 4. CONCORDANCE OF <i>Ery1</i> MISSEXPRESSION WITH THE <i>jdf2</i> PHENOTYPE AND INITIAL CHARACTERIZATION OF DEFECTS LEADING TO <i>p^s</i> AND <i>p^m</i> PHENOTYPES.	6 6
SECTION 5. <i>Ery1</i> EXPRESSION IN WILD TYPE AND <i>jdf2</i> ALLELES.	7 5
CHAPTER 4. DISCUSSION	8 0
SECTION 1 <i>Ery1</i> DEFECTS ARE RESPONSIBLE FOR THE	8 2

jdf2 PHENOTYPE

SECTION 2. OUTLIERS AND REMAINING PUZZLES	8 6
SECTION 3. <i>Eryl</i> PROTEIN.HOMOLOGIES.	9 1
SECTION 4. CONSEQUENCE OF MUTATION	10 7
SECTION 5. <i>Eryl ljd2</i> LINK.	11 1
SECTION 6. FUTURE DIRECTIONS.	11 5
SECTION 7. PLEIOTROPY AND SEQUELA	11 9
REFERENCES	12 3
APPENDICES	13 3
VITA	17 3

List of Figures

Fig. 1. Summarized Complementation maps for the Oak Ridge and Harwell <i>p</i> locus mutations on Mmu7.	135
Fig. 2. EtNU-induced point mutations within the <i>p</i> ^{6H} deletion define the <i>jdf2</i> locus proximal to <i>p</i> on Mmu7.	137
Fig. 3. Early genomic and deletion breakpoint mapping at the <i>p</i> and <i>D7H15F37S1</i> loci on Mmu7.	139
Fig. 4. An informal allelic series reflecting the severity of the <i>jdf2</i> phenotype.	141
Fig. 5. <i>Ery1</i> cDNA and P1 endclone mapping.	143
Fig. 5c. Deletion mapping of selected probes.	144
Fig. 6. Isolation of <i>Ery1</i> cDNA jSub5.1 by Subtractive hybridization.	146
Fig. 7. cDNA clones comprising mouse and human <i>ERY1</i> .	148
Fig. 8. A composite map of the <i>Ery1-p</i> genomic interval.	150
Fig. 9. Summary of genomic clones isolated with <i>D7H15F37S1</i> and <i>p</i> probes.	152
Fig. 10. <i>Ery1</i> expression analysis in wild type and <i>jdf2</i> and <i>p</i> alleles.	154
Fig. 11. <i>Ery1</i> expression, RPA controls and dosimetry.	157
Fig. 12. Identification of RFLVs specific for <i>p</i> ^x and <i>p</i> ^m alleles.	159
Fig. 13. Identification of point mutations in several EtNU-induced <i>jdf2</i> ^{SJ} alleles.	161
Fig. 14. Proteins with homology to <i>ERY1</i> .	163

List of Tables

Table I. <i>jdf2</i> alleles analyzed in this study.	165
Table II. Summary of <i>EryI</i> expression data.	167
Table III. A summary of deletion mapping data.	169

1. INTRODUCTION

This dissertation describes work that provides a molecular explanation for a group of pleiotropic mutations long known in the coat-color genetics literature. Genetic data has long predicted that the region proximal to *p* was bound to be interesting in structure and function, since several different effects including and in addition to pigmentation effects have been localized within the interval proximal to *p* (Russell *et al.*, 1995; Rinchik *et al.*, 1995; Lyon *et al.*, 1992). We have now cloned the gene responsible for the complex neurologic phenotype originally associated with certain alleles of the pink-eyed dilution (*p*) gene and in the process, have also determined that sequences controlling *p* gene expression are located within the boundaries of this large gene, positioned just proximal of *p* on mouse chromosome seven (Mmu7).

We report here a molecular analysis of the 'pink-eye syndrome' historically associated with certain *p* alleles (Hollander *et al.*, 1959; Hunt *et al.*, 1971). The work described is therefore based on over forty years of research on this classical coat color locus, and owes its success primarily to the spectrum and number of *p* mutations generated primarily at the Biology Division at Oak Ridge National Laboratory and the Radiobiologic Unit at Harwell, England. Independent spontaneous mutations have also arisen in colonies around the world. The term 'pink-eye syndrome' refers to defects in addition to pigmentation that are associated with a subset of *p*

alleles. It describes a complex sub-lethal phenotype which includes sterility, runting, and neurologic/behavioral abnormalities. Mutations associated with this phenotype map proximal to *p*, and have been shown to be due mutation of a single locus, termed *jdf2* (juvenile development and fertility locus 2; Rinchik *et al.*, 1995). The goal of the present study has been identification of the gene responsible for the *jdf2* phenotype, as predicted by an extensive body of previous work. This study integrates all known types of *p* mutations in the context of an in depth complementation, molecular and physical mapping analysis.

Over sixty *p* alleles, including mutations associated with the *jdf2* phenotype and alleles of the *p* deletion complex were analyzed in this study with respect to physical maps that were generated in this region. This effort, in collaboration with parallel studies conducted in the homologous human region, has led to the identification of a novel gene, mutations of which are responsible for a complex phenotype of neuromuscular tremor, severe runting, male sterility, female irregular fertility, and reduced life span. The gene associated with the *jdf2* phenotype has been previously mapped proximal to *p* on Mmu7 and seems to be due to mutation of a single pleiotropic gene. Therefore this dissertation has as its primary goal the identification of the gene responsible for the *jdf2* phenotype predicted by Rinchik and colleagues (1995).

This work includes the cloning of complete cDNAs for the mouse and human orthologs of a large protein which is a member of an evolutionary-conserved gene family (*ERY1/Ery1*, human and mouse, respectively). The gene has been named to reflect the derivation of the human homologue (see below). While humans contain multiple *ERY1*-related loci, *Ery1* is the single transcribed copy of the gene corresponding to the previously described *D7H15F37S1* locus in the mouse. This work also provides a genomic clone framework that permits direct molecular access to *p* gene regulatory elements, including those responsible for the somatically mosaic ('mottling'; p^m) and intermediate (p^x) pigmentation phenotypes. This detailed physical map includes the 5' portion of the *p* gene and extends proximally to include *D7H15F37S1/Ery1*.

Cloning and sequencing of mouse and human *ERY1* was accomplished in close collaboration between myself and Yonggang Ji, a graduate student in the laboratory of Dr. R. Nicholls, and the core sequencing facility at Case Western Reserve University (CWRU) in Cleveland. This collaboration was essential for the isolation of the 15.2 kb mouse and human, full-length, *ERY1* cDNA clones. An accompanying paper, to be submitted by our collaborators, will describe the cloning and complete sequence of the mouse and human genes, as well as characterisation of a low-copy sequence repeat element corresponding to part of the ancestral *ERY1* gene (Ji *et al.*, in preparation). Amplification of this repeat element termed the 'END' repeat, may be responsible for several *ERY1* pseudogenes scattered

throughout HU15q11-13 and elsewhere, and has been proposed to be responsible for the clustering of deletion breakpoints associated with Prader-Willi (PWS) and Angelman (AS) syndromes in humans (Amos-Landgraf *et al.*, submitted). The association with 'END' repeat sequences has lead to the name *ERY1* (End Repeats Family 1) for this large conserved gene.

This study establishes an almost complete concordance between disruption of the *Ery1* gene and the *jdf2* phenotype. This work also contributes to the further molecular dissection of the *p* deletion complex by providing extensive genomic clone coverage of the *ERY1/jdf2-p* interval, in the form of YAC, BAC, P1 and cosmid clones. BAC and YAC clones identified in this study provide a link between the region associated with a prenatal lethal locus, *pl1*, located proximal of *Ery1*, and *p*. Analysis of 5'-*Ery1* positive BAC clones, with respect to proximally-located *p* lethal mutation (*pl1*) will also contribute to defining the extent of conserved synteny between HU15q11-13 and Mmu7.

The mouse pink-eyed dilution locus.

As outlined above, the work presented in this dissertation has identified the gene responsible for the juvenile lethality phenotype, located proximal to the *p* gene, and has idicated the likely location of sequences necessary for the proper expression of the *p* gene itself. Therefore this introduction will start with background information

on the *p* gene and the associated deletion complex that has been generated and characterized in the mouse.

The pink-eyed dilution locus (*p*) is a very old mutation carried in varieties of 'fancy' mice that were propagated by amateur collectors at the turn of the century. Together with the albino (*c*) locus, *p* formed the first linkage group identified in the mouse (Haldane, 1915). These two readily distinguishable phenotypic traits have a long history of study, and are represented by many mutant alleles. Therefore, the *p* and *c* loci were logical choices as two of the seven visible marker loci that were used to establish the danger of radiation as a mutagen, as part of "specific locus test" experiments conducted at the Biology Division at Oak Ridge National Laboratory (Russell *et al.*, 1951, 1959). Spontaneous and mutagen-induced alleles of *p* have also been generated at Harwell, England (Lyon *et al.*, 1992; Phillips *et al.*, 1977). The majority of mutations from both collections are large, radiation-induced, multi-locus deletions (Russell *et al.*, 1995; Lyon *et al.*, 1992). Complementation analysis of a subset of the Harwell *p* alleles has been reported and are described in detail below. Forty-five *p* alleles generated at Oak Ridge have been analyzed in a recently completed complementation analysis, and these studies have generated an ordered map of five phenotypically distinguishable genetic functions in the region surrounding the *p* gene (Russell *et al.*, 1995).

Mutations of the mouse *p* gene affect the melanin biosynthetic pathway and lead to a characteristic reduction in pigmentation of the coat, skin, and choroid and retina of the eye in affected animals. The mutation leads to reduction in the level of black-brown (eumelanin) pigment, whereas yellow-red (phaeomelanin) levels are almost normal (Silvers, 1979). The mouse and human *p* genes have been cloned, and mutations in these genes have been correlated with hypopigmentation defects in both species (Rinchik *et al.*, 1993; Gardner *et al.*, 1993; Witkop *et al.*, 1989). Deletion and point mutations in the human *P* gene have been identified and are associated with the human pigmentation defect, tyrosinase-positive (type II) oculocutaneous albinism (MIM 203200; Lee *et al.*, 1995).

The human *P* protein contains twelve transmembrane domains that may function as an integral membrane transport protein for tyrosine, the precursor to melanin biosynthesis (Lee *et al.*, 1995). Consistent with this postulate is the observation that hair bulbs from patients with OCA2 can be induced to produce pigment when incubated with tyrosine (Witkop *et al.*, 1989). Similar experiments have also been done in the mouse (Sidman *et al.*, 1965), suggesting that the gene serves a conserved function in the establishment and maintenance of skin and eye pigmentation.

Over the years, over one hundred independent *p* gene mutations have been generated by a variety of groups. Two reports which summarize the bulk of the collection at Oak Ridge and Harwell

have been published (Russell, *et al.*, 1995; Lyons *et al.*, 1992). Both groups report a variety of mutations, and many of these have recently been characterized on the molecular level (Johnson *et al.*, 1995; Lyon *et al.*, 1992). Both groups also describe alleles with an intermediate *p*-effect on coat color, with normal to slightly less pigmented eyes. These have been termed p^d (dark-eyed, *p*) and p^x (for intermediate) by the two groups, respectively. Russell (1995) also reported two new alleles at *p*, which express the affects of the mutation in a mosaic pattern in somatic cells; this mosaic expression leads to the appearance of patches of hair with diluted pigment on an otherwise normally pigmented background, and give the animal's fur a 'mottled' appearance. It has generally been assumed that both the mottling and intermediate *p* alleles are due to mutation of *p* gene regulatory elements. The first indications of molecular defects associated with intermediate and mottling *p* alleles are reported in this study.

**Additional phenotypes associated with certain *p* alleles:
the 'pink-eye syndrome'.**

Reports of other phenotypes, in addition to pigmentation, associated with certain *p* alleles are present in the early literature. A negative effect on size and body weight, originally attributed to *p*, was observed in backcross animals carrying certain *p* alleles (Castle, 1936). The first report of an allele of *p* which, in addition to pigmentation defects, showed "poor growth, slightly uncoordinated

behavior, difficulty in chewing, inadequate maternal care, sterility in males, premature senility, and abnormal wearing of the incisors" was reported in 1959. This allele was designated p^s for p associated sterility (Hollander *et al.*, 1959). Ironically, despite citing early results from the specific locus test to suggest radiation induces primarily deficiencies (deletions), the authors believed *their* radiation-mutation behaved as a defect in a single pleiotropic gene. However, no satisfactory explanation could be found for the multiplicity of effects seen in homozygous p^s mutants in this early study.

Over the years several groups have studied additional effects attributed to certain p alleles specifically with respect to spermatogenesis. Sterility in p^s homozygotes was attributed to both poor mating response and a high proportion of abnormally shaped sperm (Hollander *et al.*, 1959). The p^s phenotype was later analyzed in two new neutron-induced p -sterile alleles, p^{6H} and p^{25H} , which had been generated at Harwell, England. The term "pink-eye syndrome" was used to describe effects that were inherited by p locus mutants in addition to dilution of pigmentation (Hunt *et al.*, 1971). Homozygotes of all three alleles exhibited pink-eyes and dilute coat color, were smaller than normal litter mates at birth, and remained runted throughout their life span. The animals were described as being active, "nervous", and uncoordinated. Malocclusion of the incisors, which was originally reported for p^s homozygotes, was not seen in homozygotes of either of these three

alleles. Male sterility associated with these mutations was attributed to testicular hypertrophy and abnormal sperm. Sperm cell defects, namely, giant sperm with abnormal heads and multiple tails, were characterized as defects in proacrosome formation and cytokinesis respectively. Looking for a way to link defects in pigmentation and spermatogenesis, the author postulated a pituitary defect to account for the genital hypoplasia and melanogenesis defects (Hunt *et al.*, 1971).

Two independent p^s alleles were also analyzed in all possible combinations with the Harwell mutations p^d (dark pink-eyed, normal viability) and p^{bs} (p -black-eyed sterile) expressing "reduced viability" (Wolfe *et al.*, 1976). An attempt to systematically analyze additional phenotypes that had been reported for these different p alleles was made by Lyon (1992). Harwell alleles p^d and p^{bs} these alleles both express intermediate p -effects on pigmentation 'reduced viability' is reported for the allele only. Harwell alleles p^d and p^{bs} , correspond to the Oak Ridge designations p^x , viable and p^x , jlm respectively (see Fig.1b). A similar range of sperm cell defects as reported previously for a subset of these alleles was also observed for other mutants of this type (Hunt *et al.*, 1971). Based on the distributions of different sperm cell morphologies for each allelic combination it was concluded that this data did not support an obvious deletion, complementation relationship. The author was also careful to point out that the data also did not preclude the presence of deletions.

The p^s mutation was also analyzed in combination with the unlinked mutation, *hpy* (hydrocephalic-polydactyl), located on mouse chromosome six (Mmu6), which affects flagellar axoneme assembly and leads to a total absence of intact flagella (Bryan, 1977; 1981). In this study, p^s/hpy compound heterozygotes were observed to have no new phenotypes not found in either mutant homozygote despite the fact that both mutations affect concurrent processes in sperm development. These observations led to the conclusion that each mutation acts on separate pathway leading to a mature sperm cell (Bryan, 1981).

Aggregation chimera experiments have also been used to determine additional effects of mutations at p by constructing genetically mixed embryos containing p^{6H} and wild type components. It was found that the presence of wild type donor cells in the testis of recipient p^{6H} homozygous embryos was unable to rescue sperm cell defects in the resulting chimeric animals. It was concluded that, at least with respect to spermatogenesis, the p^{6H} mutation behaves in a cell-autonomous fashion (Handel *et al.*, 1977).

Pituitary axis defects as a common cause for pleiotropic pink-eye syndrome

In an effort to reconcile the wide range of effects seen in certain p alleles, several groups have tried to find a common defect

or pathway responsible for all manifestations of the pink-eye syndrome. These efforts were hampered by the fact that, at the time, the molecular basis of the mutant alleles was unknown. Subsequent molecular analysis has revealed that these earlier studies were conducted on a variety of different types of mutations, including a combination of small and large multilocus deletions. This added extra complications to early attempts to provide unifying explanation for this complex phenotype.

The first study to specifically address the effects of *p* in addition to pigmentation, was aimed at determining the effects of *p* on reproduction. The study compared the number of ova released by female mice in response to injections of pregnant mare's serum (PMS) and human chorionic gonadotropin (HCG; Wolfe *et al.*, 1970). This study examined the influence of the original *p* allele and pink-eyed unstable (p^{un} ; Melvold *et al.*, 1971) on ova production by comparing the inbred strains C57BL/6 (+/+) and 129/Re (*p/p*) as well as lines congenic to C57BL/6 and 129/Re, differing only in the fact that they carried specific *p* locus alleles. Their findings suggested that endogenous levels of gonadotropic activity were elevated in pituitary homogenates and supernatants in *p* homozygous mice as compared to wild type controls. These elevated levels of gonadotropic activity were assumed to be the cause for the increased ovulatory response in pink-eyed females mice in response to exogenous gonadotropin injections. No satisfactory link between pituitary function and the pleiotropic effects on behavior,

pigmentation, and fertility was found. It was speculated that defective function or production of the gonadotropin, luteinizing hormone (LH), could be involved in *p*-associated sterility, due the hormone's involvement in growth and maintenance of interstitial cells of the testis (Wolfe *et al.*, 1970).

The recessive alleles p^{6H} , p^{25H} and p^{bs} , each cause sterility in males and semi-fertility in females (Melvold *et al.*, 1974). To study the effects of these *p* alleles on fertility they were each analyzed as compound heterozygotes in combination with p^d , p^{un} and p , three alleles which by themselves do not reduced fertility. Examination of pituitaries from sterile male combinations showed significantly lower proportions of gonadotropic cells as compared to the pituitaries from sibling controls. In addition, the ovaries of homozygous semi-fertile females contained large numbers of developing follicles, but no corpora lutea or corpora hemorrhagica. The ovaries of semi-fertile mutant females also contained large numbers of polyovular follicles. The author noted that other groups had postulated that the reproductive effects of certain *p* alleles might potentially be due to pituitary defects. In line with this hypothesis, he notes that production of interstitial-cell-stimulating hormone (ICSH; the male equivalent of LH), and LH is required for the maintenance of the seminiferous tubules and uterus, respectively, and that secretion of both hormones is controlled by the pituitary.

The effect of *p* on reproduction was also studied in the radiation-induced allele *p*^{25H} (Johnson *et al.*, 1974). Electron microscopy of the pituitary of homozygous animals revealed degenerating axons in the pars nervosa of both sexes, which was readily visible in mutant mice as early as fourteen days after birth. The authors also noted a reduced ability of mutant tissues to bind radioactive estradiol. In addition thyroid iodine binding was reduced in the homozygous mutant mice. Coupled with the observation of a "threadlike uterus" in mutant females and runting in both sexes, They concluded that the *p* mutations lead to "multiple endocrinological abnormalities". Subsequent analysis has determined that the *p*^{25H} allele displays a sub lethal phenotype with runting, jerky gait and sterility (Lyon's 'R, JG, S'). We regards to semi-fertility in mutant females, the authors confirm that some females were sterile whereas others produced litters but failed to rear them. This suggests that the irregular fertility in these cases is influenced by additional factors possibly including maternal care or lactation (Lyon *et al.*, 1992).

Their conclusion, based upon the hypothesis of pigmentation, reproductive, and other disorders as manifestations of a single-gene defect, was that the *p* gene might code for a common subunit for the gonadotropins follicle stimulating hormone (FSH), thyroid stimulating hormone (TSH) and lutenizing hormone (LH). While industrious, this theory gave plausible explanations for growth reduction and sterility due to hypertrophy of the ova and uterus, it completely

failed to account for abnormal sperm (which the authors acknowledge). Their theory also failed to explain how the same hormone subunit could also play a role in pigmentation.

In hindsight we now know that neither model needed to account for *p* effects since *p*^{25H} and *p*^{6H} both delete *p*. Later experiments have since established that any neuroendocrine defects in these experiments are likely due to deletion of the interval proximal to *p*, that is due to *Ery1/jdf2* mutation. In summary, early efforts, to explain the *p*-associated effect in terms of mutations in the *p* gene itself, were unsuccessful. However, now we know that the additional effects are due to the adjacent *Ery1* locus. Therefore early efforts gave indications that there may be a underlying neuronal defect as the root cause of the pink-eye effect. Since it is still possible that such effects could be downstream results, future efforts will need to address defects in pituitary and gonad function in these mutant alleles.

Another theme from this early work is the observation that fertility defects in both sexes seem to be reminiscent of both structural and proliferative problems. For instance the pink-eye mutant testis contains abnormal sperm and signs of hyperplasia in cells of the seminiferous tubules. Though it is particularly difficult to interpret these results they are consistent with pituitary dysfunction and should be addressed in future studies.

Complementation analysis of Oak Ridge *p* locus mutations

Pairwise complementation analysis of forty-five Oak Ridge alleles, representing all known types of *p* mutation and the result of several decades of study, has recently been reported. This analysis allows a linear complementation map containing five genetically determined functions to be drawn with respect to the centromere, *p* and the Mmu7 telomere (Russell *et al.*, 1995; summarized in Fig 1a).

In the region proximal of *p*, this analysis identified a juvenile-lethal phenotype characterized by greatly reduced life span, neuromuscular abnormalities, severe runting, male sterility and irregular female fertility. The genomic interval containing sequences controlling intermediate and mottled expression of *p* was also mapped proximal to *p* based on the ability of alleles such as *p*^{12DTR} and *p*^{39DSD} to fully complement *p*^{4THO-II} a deletion normally perinatally lethal known to remove the 3' half of the large *p* transcription unit leaving 5' and proximal regions intact (Nicholls *et al.*, 1993).

The results of complementation analysis of eleven *p* mutations generated at Harwell, England have also been reported (Lyon *et al.*, 1992). The resulting map identifies four genetic functions and corroborates the basic features of the Oak Ridge map. Specifically, Lyon and colleagues identified a function which they termed "R, JG, S" for runting, jerky gait and sterility mapping in the vicinity of *p*.

Mutations effecting the R, JG, S functional unit include those which have historically defined the pink-eye syndrome (p^{6H} and p^{25H}) and are assumed to effect the same locus as the juvenile lethal function mapping proximal to p described in Russell (1995).

The Oak Ridge and Harwell complementation maps are summarized in figure 1a and 1b. Mutant alleles depicted in the Oak Ridge map include those which define juvenile lethality and intermediate and mottled p expression. Additional alleles included are those that have been used as indispensable mapping reagents. The figure legend contains descriptions of representative crosses which were used to define each functional unit. The legend also contains the rationale and information derived from the mapping alleles mentioned above. Combined with the evolving molecular mapping data (Nicholls *et al.*, 1993; Culiati *et al.*, 1992; Johnson *et al.*, 1995) the complementation data allows the Oak Ridge map to be oriented with respect to the centromere as well as giving the transcriptional orientation of the p gene.

The Harwell complementation data is oriented with respect to the more extensive Oak Ridge map as described above and in the legend of Figure 1b. Again, only salient features and representative mutations are included. To summarize, the Harwell map confirms the basic features of the OR map in that it identifies a phenotype including runting, jerky gait (neuromuscular abnormalities) and sterility as well as a region controlling intermediate and mottled p

expression mapping at *p*. The published orientation is preserved in this figure and has since been shown to be accurate though the authors acknowledge that at the time they were unable to orient their mutations with respect to the centromere.

Complementation studies conducted at Oak Ridge identified two distinct classes of juvenile lethality, termed juvenile lethal mild (jlm) and severe (jls). These two classes are readily distinguishable based on their relative survival to weaning. For instance, of progeny from crosses involving jlm 61% survive through weaning or beyond and 32% survive to 237 days. By contrast, only 28% of jls combinations survive through weaning or beyond and only 4% survive to 237 days. Combinations involving either of these functional units exhibit the neuromuscular phenotype seen in the homozygous state (described as twitching of the hindlimbs and severe tremor of the forelimbs when the animal is suspended by the tail) and the authors note that they could find no information to separate the neurologic defect and juvenile lethality (Russell *et al.*, 1995). Furthermore the authors cite evidence of ethyl nitroso urea-induced (EtNU) mutations which map to the jls, jlm interval suggesting that both phenotypic classes may be due to single gene with highly pleiotropic effects (Rinchik *et al.*, 1995). This issue is revisited below and in the discussion of this thesis in terms of the consequences of specific types of mutations.

Pink-eye syndrome is due to mutation of a single gene designated *jdf2*.

A large-scale mutagenesis screen to identify additional phenotypes within deletions including those of the *p* complex has been undertaken by Rinchik and colleagues (Rinchik *et al.*, 1991; 1995). Briefly, BJR (*p/p*) males were mutagenized by fractionated, intra peritoneal (ip) injection of EtNU in phosphate buffer. Newly induced mutations in the vicinity of *p* are identified as hemizygotes when crossed to the largest *p* deletion *p^{46DFioD}*. These new mutations can then be rapidly mapped with respect to subregions of the deletion complex by further matings to selected deletions. So called repeat mutations, independent mutations displaying the same phenotype and mapping to the same genomic interval, have been identified and verified by allelism testing (EMR, DAC, unpublished).

Originally, three non-complementing, ethyl-nitrosourea (EtNU)-induced mutations that fail to complement the Harwell deletion *p^{6H}* were identified (Rinchik *et al.*, 1995 and EMR, DAC unpublished). The *p^{6H}* mutation has been shown to be a deletion by molecular probe analysis (Lyons *et al.*, 1992; Nicholls *et al.*, 1993; Johnson *et al.*, 1995). Homozygous *p^{6H}* are dilute in pigmentation, runted compared to littermates, and exhibit the neuromuscular/jerky gait that is typical of animals inheriting the pink-eye syndrome (equivalent to the "*p^{6H}* deletion syndrome" described by Rinchik). Each of these EtNU-induced alleles produced the trio of effects on size,

neuromuscular/jerky gait, and spermiogenesis described for the pink-eye syndrome and the p^{6H} deletion phenotype.

Since these original reports, four additional EtNU-induced mutations that produce the full range of " p^{6H} deletion phenotype" effects have been described (Rinchik and colleagues, unpublished; Walkowicz *et al.*, manuscript in preparation). EtNU is thought to produce primarily intragenic base-pair-changes (Russell *et al.*, 1987) and has been found to induce single-base-pair changes in most mutations induced in spermatogonial stem cell stages that have been examined to date (Peters *et al.*, 1990; Marker *et al.*, 1997). The fact that these mutations recapitulate the runting/neurologic/sterility aspects of the pink-eye syndrome phenotype and fail to complement p^{6H} , suggests that the entire phenotype is due to mutations of a single gene with highly pleiotropic effects. This locus, which maps proximal to p but within the p^{6H} deletion, has been designated *jdf2* (juvenile development and fertility locus, 2). *jdf2* can be considered equivalent to the "RGS" function described in Lyon (1992) and the juvenile lethality phenotype (*jlm* and *jls*) proximal to p in the Oak Ridge complementation map (Russell *et al.*, 1995).

More detailed confirmation of the similarities between the phenotype seen in EtNU-induced mutations and p deletion alleles which exhibit the *jdf2* phenotype was found in an examination of various morphologic aspects of spermiogenesis by electron microscopy. Spermiogenesis in each of the EtNU alleles was

examined in direct comparison to the deletion historically associated with the pink-eye syndrome in the literature, p^{6H} . EtNU alleles (generically of the "SJ-series") were analyzed as hemizygotes opposite various p deletions (for instance $jdf2^{932SJ}$, $p/p^{46DFiOD}$; see Figure 2 and legend) and compared with p^{6H} as a homozygote. Consistent with the sperm-cell defects previously reported for certain pink-eye syndromes mutations (see above), affected males (both p^{6H} homozygotes and EtNU-induced point mutants) have a large proportion of spermatids with two acrosomes per spermatid nucleus. This particular type of defect accounts for very few of all male sterile defects that have been characterized so far (Rinchik *et al.*, 1995).

A second defect that was frequently observed in both mutant types was the presence of two spermatid nuclei contained within a single proacrosome. Animals heterozygotes for each of the three originally reported EtNU-induced alleles exhibited apparently normal spermiogenesis. Males hemizygous for any of the three SJ alleles (EtNU-induced mutations inherited with a deletion) displayed the same characteristic defect of acrosome assembly as is seen in the male sterile mutation p^{6H} . By crossing the EtNU-induced alleles with various p deletion mutations (p^{6H} , $p^{4THO-II}$, p^{25DVT} see Figure 2) the SJ alleles were mapped proximal to p in the vicinity of locus $D7H15F37S1$, the interval previously associated with runting/nervous behavior/male sterility (Nicholls *et al.*, 1993; Lyon *et al.*, 1992).

Description of radiation-induced and 'other' alleles which define the pink-eye/*jdf2* syndrome.

This dissertation focuses on the molecular analysis of seven *p* mutations that were included in the complementation analysis of Russell and colleagues (1995), but have not yet been analyzed on the molecular level. Two of the seven are new mutations associated with a mottled *p* phenotype (p^m). Five of the seven mutations that have served as a primary focus of this work are intermediate (p^x) *p* alleles (designated dark-pink-eyed, p^d , by Lyon, 1992). The previously described intermediate alleles, p^d and p^{bs} , are viable or of reduced viability respectively (Lyon *et al.*, 1992).

Animals homozygous for all seven mutations studied here exhibit decreased viability and neurologic defects in addition to their defects in pigmentation. However it should be noted that the p^m and or p^x phenotypes are not a prerequisite for these neurologic effects. This is evidenced by the existence of the p^x -viable mutations at Oak Ridge and the p^d Harwell mutation (Russell *et al.*, 1995). The mutation p^{9DTW} , a small deletion is a *p* null mutation owing to deletion of 5'*p* sequences (Johnson *et al.*, 1995; Walkowicz, this work) also proves p^x or p^m phenotypes are not necessarily linked to the *jdf2* phenotype since the deletion in p^{9DTW} is centered on a small region surrounding 5'*p*. Both deletion breakpoints in p^{9DTW} have been elucidated in the course of this study and in collaboration with Y. Ji.

Although a linear map [oriented cent-*pl-1-jls-jlm-p*-tel; Fig.1a] can be used to order all available *jdf2* mutations on the complementation map, the possibility that some of these mutations may correspond to complex rearrangements still exists. Of thirty six *p* locus mutations that have been analyzed with molecular probes, there is evidence for several non-linear or 'skipping' deletions (Johnson *et al.*, 1995). Rare wild type revertants have also been identified in intercrossees involving the *p^m* mutation *p^{39DSD}* and *p^{48PB}* (Russell *et al.*, 1995). A plausible explanation involving illegitimate recombination events affecting a closely linked regulatory element that controls *p* expression has been presented by Russell and colleagues (1995). The spontaneous allele *p^{48PB}* is shown in this study to be a small deletion involving the 3' end of *Ery1* and is discussed below.

Description of known loci proximal to *p*.

The ultimate purpose of this dissertation has been the mapping and isolation of the gene associated with the *jdf2* phenotype. With the considerable body of reagents and mapping information that were available at the onset of this effort, we were poised to launch a "traditional" positional cloning approach to analyze the *jdf2* mutant locus. Logically, any positional cloning strategy should start with the identification of a minimal physical interval, commonly called a "critical region", containing the mutant interval. Once the interval is

defined, positional cloning strategies generally proceed with identification of any candidate genes mapping into the genetically defined interval (Stubbs, 1992). Probes defining the locus *D7H15F37S1* (originally *D7H15F37S1h*) had previously been shown to map to a region located just proximal of the *p* gene (Nicholls *et al.*, 1993). These probes, which had been isolated by microdissection, were known to represent portions of an expressed gene. (Buiting *et al.*, 1992).

Identification of probes for the *D7H15F37S1* locus came about in a manner analogous to the isolation of probes for the mouse *p* locus (Rinchik, *et al.*, 1993). The common theme in these two efforts was the ability to map with certainty, human probes with respect to well characterized regions of the genome, in this case deletion complexes (Stubbs, 1992; Nicholls *et al.*, 1993).

A clone derived by chromosome microdissection which detected a family of sequences on human chromosomes fifteen and sixteen, designated HSMN7, was isolated by physical dissection of human chromosome 15q11-13 (Buiting *et al.*, 1990; Nicholls *et al.*, 1990; see also Fig. 2). That Mmu7 markers were arranged in the order, centromere-*D7H15F37S1*-*p*-telomere, was first shown by Nicholls and colleagues (Nicholls *et al.*, 1990). Deletion mapping also revealed that the conserved HSMN7 probe detected murine sequences that were deleted in the *p*^{6H} mutation. The order of markers in the mouse *p* region has more recently been confirmed

and extended in the process of mapping of two mutations *p^{12R30Lb}* and *p^{9DTW}* both of which break in the body of the *p* gene and extend proximally (this study, see below). *D7H15F37S1* marks the proximal extent of the documented region of synteny between Mmu7 and HU15q11-13 (Nicholls *et al.*, 1993). Since probes for this locus have played a central role in the analysis of the *jdf2* mutant region, details of its derivation and subsequent analysis are worthy of additional comment.

The majority of patients with Prader-Willi and Angelman syndromes (PWS and AS) have cytologically visible deletions of chromosome HU15q11-q13 although the syndromes are clinically very different. PWS is characterized by hypotonia and failure to thrive as infants. Early in childhood patients develop hyperphagia often leading to cardio-respiratory failure without intervention. PWS patients also display mild mental retardation, hypogonadism, continuing hypotonia and short stature. By contrast mental retardation in AS patients is severe with symptoms including the absence of speech. Patients also exhibit inappropriate laughter and 'puppet-like' ataxic movements (Nicholls *et al.*, 1992). These two distinct disorders map to a common chromosomal region, with inheritance of PWS or AS symptoms depending upon "parent-of-origin" effects. Normal development requires the unique contributions of both parental chromosomes, due to the differential maternal or paternal expression (imprinting) of specific genes in the region. Both syndromes are associated predominately with deletions

(60-75%) but can also occur by uniparental disomy (UDP) of the 15q11-q13 region. UDP of HU15, a condition which results when both chromosomes are inherited from the same parent (in the case, maternal), has been demonstrated as being responsible for the remaining 20-25% of PWS patients (Nicholls *et al.*, 1991).

Probes for *D7H15F37S1* were originally derived by microdissection of geimsa-trypsin banded human metaphase chromosomes from normal amniotic fluid cell cultures. The purpose of these experiments was to derive probes mapping within the interval HU15q11-q13 that had been found to be commonly deleted in patients afflicted with PWS and AS. One deleted clone, designated locus *D15S62*, was thought to contain gene sequences based on the fact that a probe derived from this sequence (a 300 bp fragment called MN7) detected evolutionary-conserved sequences in other species (Buiting *et al.*, 1990). Microclone MN7 was found to detect multiple loci in the human genome, which were mapped by somatic cell hybrids and chromosome *in situ* hybridization to human chromosomal regions 15q11-13 and 16p11.2, respectively. Based on analysis of several YAC and genomic phage clones with human somatic cell mapping reagents, it was determined that several copies of the MN7 sequence were present on 15q11-13, arranged in a tandem array and spread over a large distance (Buiting *et al.*, 1992). By contrast, the HSMN7 detected a unique locus in the mouse which was mapped to Mmu7 roughly in the vicinity of *p* using recombinant inbred strains (Chaillet *et al.*, 1991).

Because of its interspecies conservation, HSMN7 was used to screen a human placental cDNA library, and two types of clones were identified. Clone HSC7.681 (~1.6kb) was subsequently mapped by *in situ* hybridization to H16p11.2. Clone HSC7.541 (~2.0 kb), was determined to be located in H15q11-q13, and thus represented one of the copies of the locus in the PWS/AS region. Expression analysis showed that the gene represented by clone HSC7.541 produced two transcripts, 8 and >14 kb length, in most human tissues. Sequence analysis indicated that the two HSMN7-related transcripts shared no significant sequence homologies to any known genes from human or other species (Buiting *et al.*, 1992). Subsequent analysis has determined that HSMN7 and HSC7.541 are part of a truncated, presumably transcribed pseudogene. Isolation of independent cDNAs in mouse and human has determined that HSC7.541 is part of the *ERY1* gene which is present as complete and pseudogene copies on Hu15q11-q13 and is coded for from a single locus proximal to *p* near the locus *jdf2* (Walkowicz, this study; Ji *et al.*, in preparation).

Specific aims and summary of results

The goal of this work was to identify a candidate gene for the genetically defined locus *jdf2*. Based on all the data available from complementation and molecular maps, and inspired by the conserved nature and expression patterns of MN7-related sequences, it was decided that *D7H15F37S1* should be considered not only as an

important flanking marker but also as a candidate gene that might be responsible for the *jdf2* phenotype. As an integral part of this work, we therefore set out to complete the physical map from this locus extending distally to include the 5'-end of *p*.

The locus *D7H15F37S1* was considered a good candidate gene for the spectrum of phenotypes responsible for the 'pink-eye/*jdf2* syndrome' based on the following criteria:

Map position Locus *D7H15F37S1* maps just proximal to *p* within the proximal portion of the *p*^{6H} deletion, a region which has been associated with the pink-eye/*jdf2* syndrome (Nicholls *et al.*, 1993; Rinchik *et al.*, 1995). Preliminary genomic physical mapping data (obtained in long range restriction studies) suggested the distance between *D7H15F37S1* and *p* to be ≤ 600 Kb (L. Stubbs, unpublished). This genomic data was later supported by the isolation YAC clones containing sequences from both loci, indicative of markers separated by 1 Mb or less (see below).

Expression profile As a naive first approximation one would expect a candidate gene for a phenotype which included neurologic symptoms and male sterility to be expressed at least in the brain and testis. HSC7.541 and related mouse cDNAs meet this basic requirement in that they are expressed in most human tissues tested (Buiting *et al.*, 1992; Walkowicz, unpublished).

"Mutability The existence of three (at the time and now seven in total) allelic, EtNU-induced mutations suggested that this locus was a good target for EtNU mutagenesis. The generation of seven so-called repeat mutations would seem to suggest either a highly mutable locus or one that presented a large target for mutagenesis. That probes for *D7H15F37S1* detect a polyA(+) RNA of >14 kb was therefore viewed as consistent with it being a candidate gene.

Proximity The existence of animals that exhibit the intermediate *p* phenotype in addition to the pink-eye/*jdf2* syndrome (designated by Russell as *p^{xjl}*) suggested, particularly in view of the evolving molecular map, that the defect responsible for the juvenile lethality should lie near the sequences associated with the intermediate *p* phenotype. Since it is thought that the *p^x* phenotype is due to mutation of a regulatory element controlling *p* expression which are traditionally linked in *cis* (by some definition) with the coding region of the gene and complementation data localizing *p^x* and *p^m* proximal of *p*, argues by proximity for *D7H15F37S1* as a candidate gene locus.

Proof: Concordance/Discordance Although the *D7H15F37S1* locus represented an interesting *jdf2* candidate, the size of the *jdf2* interval and the number of other potential gene sequences that might be located within that interval were unknown. In parallel with studies designed to test *D7H15F37S1* as a candidate gene we

therefore also set out to complete a physical map of the region and to use deletions to determine a minimal interval of overlap between mutations associated with the *jdf2* phenotype. Any potential candidate gene would need to show molecular defects, either on the level of genomic DNA rearrangements or RNA expression levels, in animals carrying mutations associated with the *jdf2*/juvenile lethal phenotype.

SUMMARY OF RESULTS

This work has as its ultimate result identification of a novel gene *Ery1* (for End Repeats Family member 1) as the gene responsible for the defects observed in *jdf2* mutant mice. These results establish an almost complete concordance between mutations associated with the *jdf2* phenotype and rearrangement and/or misexpression of *Ery1*. In addition to leading to the complete cloning and sequencing of the mouse and human orthologous of this large transcript, we have completed a physical map between the *D7H15F37S1-p* interval and have identified the first molecular defects in both mottling and 'intermediate' *p* alleles. Finally, we have identified a set of non-chimeric genomic clones centered on the *Ery1-p* interval on mouse Mmu7 (see Fig.1)*.

(* All tables and figures maybe found in the appendix.)

Our experimental goal was to complete the physical map of the genomic interval extending proximally from *p* and including the known locus *D7H15F37S1*. Complementation analysis had identified several additional phenotypes mapping proximal to *p*. These include juvenile lethality, mottling and intermediate expression of the *p* phenotype. As described above it was known both from linkage and preliminary PFGE analysis that *D7H15F37S1* was tightly linked to *p*. Since the genomic data suggested the *D7H15F37S1-p* distance was on the order of 600 kb our first approach was to isolate genomic clones, in the form of YAC clones that carried markers available at that time for the *D7H15F37S1h* and *p* loci.

In the course of this work we have completed the physical map from *D7H15F37S1* distal to *p* using a variety of different types of genomic clones (including yeast artificial chromosome, or YACs, bacterial artificial chromosomes, or BACs, bacteriophage P1, and cosmid clones), isolated with *p* and *D7H15F37S1* mouse cDNAs. Mapping of probes derived from these clones with respect to *jdf2*/juvenile lethal mutations allowed us to identify the first molecular defects in several alleles of the *p^x* class of *jdf2* mutations (Russell's *p^{xjl}*). Furthermore, integration of endclone and cDNA probe mapping data allowed us to identify the minimal intervals necessary for the *jdf2* phenotype as well as the minimal intervals responsible for intermediate *p^x* and *p^m* pigmentation. Figure 8 depicts the integration of the *p* locus complementation map (Russell *et al.*, 1995) with our physical mapping results.

Our map shows that the minimal interval necessary for the *jdf2* phenotype, and intermediate and mottled *p* phenotypes are coincident and located just proximal of 5'*p*. These data are consistent with the Oak Ridge complementation map (Russell *et al.*, 1995). Based on the physical mapping data, 3'*Ery1* and 5'*p* are located on the order of 10 kb apart, making it unlikely that another gene will be found in this intergenic region. Based on physical mapping, deletion mapping, expression studies and mutant cDNA sequencing studies (completed in collaboration with Y. Ji and R. Nicholls), we propose that mutations in the gene represented by the *D7H15F37S1* locus, *Ery1* are responsible for the *jdf2* phenotype.

This work includes the cloning of complete cDNAs for the mouse and human orthologs of a member of an evolutionary-conserved gene family that is predicted to encode one of the largest mammalian proteins identified to date. In the mouse, this gene is a unique transcribed gene, portions of which are represented by cDNA sequences originally used to define the *D7H15F37S1* locus. In addition to providing a molecular explanation for *jdf2* mutations, this study has also provided mapping information for and direct molecular access to the *p* gene regulatory elements including those responsible for the somatically mosaic ('mottling') and intermediate *p* mutations. Analysis of these elements will provide the basis for additional studies at ORNL and elsewhere in future years.

This study establishes an almost complete concordance between disruption of the *Eryl* gene and the *jdf2* phenotype. *Eryl* may act as a guanine nucleotide exchange factor (GEF) for several different classes of GTPases as has been documented for p532 a HECT protein, which is related to *Eryl* in size and homology to RCC1(Rosa *et al.*, 1996). One function of *Eryl* may be in regulating vesicle sorting/partitioning which could be reflected in the acrosomal defects of *jdf2* mice. It is probable that *Eryl*'s essential function is as a E3 protein ubiquitin ligase. Recent insights into the multifunctionality of E3 proteins as well as *Eryl*'s homology to RCC1, itself a hallmark of pleiotropy, make it likely *Eryl* is an essential gene involved at a checkpoint between several fundamental cellular processes.

As will rapidly become clear many questions concerning both details of the *jdf2* phenotype and function of *Eryl* remain. A detailed pathologic examination of the brain and nervous system in *jdf2* animals will determine whether *jdf2* includes neurodegeneration in the CNS or in neuromuscular tissues. Recent data reveals that in addition to traditional apoptotic/necrotic cell death phenotypes more subtle vacuolar and autophagic effects have been seen suggesting there are several aspects or pathways of cell death (Migheli *et al.*, 1997). Therefore more detailed analysis will no doubt be required to determine whether preliminary indications of a brain defect in *jdf2* animals is indeed part of the pathology in *jdf2* animals.

Observations on the pathology of the *jdf2* phenotype

Constraints associated with the low fecundity of *jdf2* mutant stocks as well as an apparent absence of gross brain defects in our limited analysis of *jdf2* animals hampered efforts to make a significant contribution to the pathology of the *jdf2* phenotype. Results of preliminary pathologic analysis are reported here for posterity and completeness.

In order to gain some understanding of the *jdf2* phenotype gross pathology analysis of homozygotes of several *jdf2* alleles was done by Dr. Virginia Godfrey. Examination of hemotoxylin and eosin stained tissue-sections from a limited number of affected animals and control sibs indicated no obvious gross pathology after examination of a representative panel of tissues. Dr. Godfrey did make the following observations. Mice homozygous for the *p^{9DTW}* and *p^{103G}* mutations are born small, but catch up to their control sibs in size and weight with age. 'Tremors' that are exhibited by young mutant animals also appeared in these studies to become less severe as the animals aged. 'Tremors' in *jdf2^{932SJ}* homozygotes appear more severe than in *p^{9DTW}* homozygotes. Also one affected *jdf2^{932SJ}* animal shows microphthalmia in one eye. This is the extent of the formal pathological analysis by Dr. Godfrey on the *jdf2* phenotype.

What follows are observations made by colleagues and collaborators with experience in the field of neurobiology. These observations are listed here in order to provide all available information on the *jdf2* phenotype and are not meant to represent a systematic analysis. In all cases we used animals that were available at ORNL at the time collaborators were available to perform the analysis. Providing mutant animals of specific genotypes of *jdf2* alleles to collaborators was also complicated by difficulties in breeding these animals and the fact that they die as juveniles. Dr. E. Deihl (Meharry Medical College, Nashville TN) and Dr. U. Schambra at (ETSU, Johnson City, TN) both examined hemotoxylin and eosin-stained brain and spinal column material and other tissues from a male *jdf2*^{932SJ} homozygote approximately 20 days old. Comparison with heterozygous sibs of the genotype (*jdf2*^{932SJ/ru-2} + +) or (*jdf2*^{932SJ/+} + *p*) revealed no gross morphologic differences despite that fact that *jdf2*^{932SJ} is a rather severe *jdf2* allele.

Dr. A. Messer from the Jackson labs (Bar Harbor Me.) observed a video tape taken by Dr. W. Russell of a litter of 1971SJ/850SJ compound heterozygotes. Dr. Messing said in a letter to us that the "affected animals seemed to display excessive exploratory behavior and may be a good mouse-model for hyperactivity". Dr. J. Harry from (NIEHS Research Triangle Park, NC) observed a *p*^{6H}/*p*^{6H} homozygous animal that was over 4 months old and said that the animal displayed behavior when suspended by the tail consistent

with a neuro muscular defect. This is consistent with the original observation made by Dr. Russell(Russell *et al*, 1995).

Based on these observations and discussions with D. Carpenter of Dr. Rinchik's group at ORNL I have constructed an informal allelic series reflecting the severity of the *jdf2* phenotype. This allelic series is depicted in Fig.4. This series is compared with the proven molecular defects associated with each mutation in the section "Consequence of Mutation" in the discussion. Very generally the most severe phenotype is seen in compound heterozygotes (m1/m2) between any of the EtNU-induced SJ *jdf2* alleles. These animals are the most runted, exhibit constant tremor, and appear weak and sickly with matted fur. Next in severity are homozygotes of a given SJ allele (m1/m1), followed closely by hemizygotes of a SJ allele (m1/ del)

Compound deletion heterozygotes such as p^{48PB}/p^{6H} and p^{17FATw_b}/p^{6H} are less severely runted than (m1/m2) or (m1/m1) combinations and do not exhibit obvious tremors. These animals are active and clearly exhibit the ataxic or jerky gait described by Lyon and colleagues (Lyon *et al*, 1992). It should be stressed that all of the combinations mentioned thus far all exhibit greatly reduced lifespan characteristic of a severe juvenile lethal mutation. Genotypes associated with mild neurological defects and longer lifespans include p^{6H} and p^{103G} homozygotes mentioned above. Finally the p^{mjlm} alleles, appear closer to (m1/m2) combinations

than p^{103G} homozygotes but more severe than p^{6H} . Figure 4 is an attempt to represent the above as a linear series with respect to severity.

2. MATERIALS AND METHODS

Mouse breeding

This study focuses on a series of mutations generated at the *p* locus during specific locus experiments (summarized in Russell *et al.*, 1995). The *p*^{6H} mutation, known to be a large deletion and producing the *jdf2* phenotype when carried in homozygous form, was generated at the Radiobiological Unit at Harwell England. Four previously undescribed *jdf2* EtNU-induced alleles, *jdf2*^{1971SJ}, *jdf2*^{1881SJ}, *jdf2*^{374SJ}, and *jdf2*^{982SJ}, were generated as described in a previous report (Rinchik *et al.*, 1993; Rinchik *et al.*, 1995). *Mus Musculus* x *Mus Spretus* F1 hybrids and compound heterozygotes involving overlapping *p* deletions used for genomic mapping, as well as several *p*^{xjl}/*p*^{6H} compound heterozygote combinations used for expression analysis, were also produced as previously described (Rinchik *et al.*, 1993; Rinchik *et al.*, 1995). Compound heterozygotes among different deletions exhibiting the *jdf2* phenotype were made by mating *p*^{xjl-1}/*p*^{7R75M} with *p*^{xjl-2}/*p*^{7R75M}, where *p*^{xjl-1} and *p*^{xjl-2} are independent *p*^{xjl} alleles (e.g. *p*^{12DTR}/*p*^{9DTW}) and *p*^{7R75M} allele is a viable intermediate *p* allele.

cDNA subtraction

Isolation of cDNAs by subtractive hybridization was performed using the "Subtractor" kit (Invitrogen San Diego, CA.). The

hybridization reaction contained 10 mg total brain RNA isolated from an adult p^{6H} homozygous animal, and 1.0 mg of biotinylated first-strand cDNA derived from total RNA from the brain of a normal C3Hf adult. One quarter of the selected cDNA population was labeled, preassociated with a cocktail including sheared mouse genomic DNA, and hybridized to a library constructed with randomly-primed adult mouse brain RNA cDNA library (Gin-Wen Chen and D. Johnson, unpublished). Inserts from positively-hybridizing phage clones were isolated from vector by *Eco* RI-digestion, ^{32}P -labeled, and mapped by hybridization to Southern blots containing p^{6H} homozygous DNA. Two independent adult brain cDNA clones that were deleted in p^{6H} DNA were identified in the first four replicating positives tested. A 2.1 kb insert from one of these clones (jsub5.1) was subcloned into pBSKII+ plasmid vector for sequencing. Clones were mapped with respect to p locus deletions and the P1 clone contig using standard Southern blotting and hybridization protocols (Stubbs *et al.*, 1990).

Molecular Probes, sequencing and hybridizations

A 200 bp probe representing the extreme 5'-end of mouse p gene coding sequences (5' p TADJ) was generated by reverse transcript-polymerase chain reaction (RT-PCR) using primers derived from the published sequence, as follows: DJDN10.1 5' cactgctggagagagatcagcgag 3', which starts at nt 82 of the mouse p cDNA sequence, and DJDN10.2 5' cgttttctctgaggaggcaactgcaga 3',

reverse complement of the *p* sequence from nt 469 (Gardener *et al.*, 1992). First strand cDNA was generated from 10 micrograms of mouse total brain RNA, and was primed with random hexamers using Superscript reverse transcriptase (Bethesda Research Labs, Bethesda, Md) according to manufacturers instruction. The mupc78.5 cDNA was a 2.0 kb clone isolated from a d8.5pc mouse embryo library (a gift of Dr. B. Hogan) using the human cDNA HSC7.541 as probe (Buiting *et al.*, 1992). RNase protection experiments were carried out using probe E550, a 703 bp EcoRI-PstI fragment subcloned from the subtracted cDNA clone jSub5.1, and corresponding to *Eryl* nt 4085-4788(Fig.8c). Northern blots were hybridized with 5504K, a 4 kb *Eryl* cDNA clone isolated from a random-primed adult mouse brain cDNA library, and corresponding to *Eryl* nts 1758-6472.

Probes YJ1.7 and YJ1.1 were human probes derived from a YAC which spans a deletion breakpoint cluster in Hu15q11-13(see below)(Ji *et al.*, accompanying paper). Probes RN5561.9 and RN5562.3 were EcoRI fragments of cDNA (RN556A-2) isolated from a from a size-selected, oligo dT-primed mouse day 14 postnatal cerebellum cDNA library (Doyle *et al.*, 1997). RN5561.9 represented nt 6204-8100 of the *Eryl* sequence, while RN5562.3 corresponded to nt 8100-10200 of the *Eryl* cDNA. The probe '7KCD' was a 7 kb cDNA fragment generated by marathon RACE and representing the 3'-end of the mouse *Eryl* gene from *Eryl* nt 8300-15,286. Probe 1.5PA(+) was a 1.55 kb fragment generated from the RACE clone 7KCD and

corresponds to the 3'-terminal end of the *Eryl* transcript from approximately *Eryl* nt 13,700 (Ji *et al.*, accompanying paper).

Restriction enzyme-digested genomic and P1 clone DNA samples were electrophoresed through 0.8% agarose gels, blotted to nylon filters, and hybridized with radiolabeled probes as previously described (Stubbs *et al.*, 1990). All probes were labeled using random hexamers as primers (Feinberg and Vogelstein, 1983). Repetitive probes were preassociated with both sheared mouse genomic DNA and/or Cot-1 DNA (Gibco BRL) as described (Stubbs *et al.*, 1990). cDNA clones mupc78.5 and jsub5.1 were subcloned into pBSKII+ vector (Stratagene, Palo Alto, CA) and partially sequenced using a combination of manual (Sequenase 2.0 USB) and automated fluorescence-base methods (ABI Prism Dye Terminator Cycle Sequencing Ready Reaction kit). Fluorescently-labeled reactions were analyzed on an ABI 373 automated sequencer, according to manufacturers instructions. Sequences were analyzed with the Blast and FASTA algorithms of GCG software, version 7 (Genetics Computer Group Inc. Madison, WI).

Genomic clone isolation and characterization

Bacteriophage P1 clones used in this study were isolated by hybridization of high density gridded arrays of a library constructed from C57BL/6 mouse DNA (F. Francis, unpublished) and distributed by the Reference Library Database (RLDB; Max Planck Institute for

Molecular Genetics, Berlin, Germany). The P1 contig was completed by rescreening the library with a pool of endclones generated by PCR from the originally-identified N1 and N2 clones, which were mapped to the *jdf2* interval using existing P1 clones and *p*-locus deletions.

P1 endclones were generated by a modified version of the ligation mediated PCR technique originally described by Mueller and Wold (1992); BAC ends were generated using a single-primer PCR method (Kim *et al*, in press). P1 endclones were generated using a modification of the ligation-mediated PCR method of Muller and Wold (1992). DNA prepared from P1-containing colonies using standard alkaline lysis methods (approximately 0.2 mg) was digested with EcoRV, RsaI and either TaqI or Pst I for left or right vector ends, respectively. After digestion the DNA was phenol extracted, precipitated with 70% ethanol, and the pellet resuspended in 10 ml TE. Two microliters (mls) of each ligation was ligated to 5 pmol of annealed linker (oligonucleotides L and LP; Mueller and Wold, 1992) in a 10 ml reaction at 14° C overnight. Two ml of the ligated material was used as template in a 100 ml PCR reaction, with 100 ng each of either LP and standard T7 oligonucleotides, or LP and SP6, as primers (to represent sequences located to the "left" and "right" of the P1 cloning site, respectively). Samples were amplified in 35 cycles at an annealing temperature of 65°C. Two ml of the primary PCR reaction was reamplified using the same primers and conditions as above. The resulting end clone products were purified by running reactions on LMP agarose.

The 5'-*p* containing cosmid was isolated by hybridization of the 5'-*p* TADJ probe to a cosmid library generated by MboI partial digestion of C3Hf mouse DNA and cloned into the pCos2 vector (Stubbs *et al*, unpublished). Bacterial artificial chromosome clones containing 5'-*p*TADJ and 1.5A(+) probes were isolated by hybridization of commercially purchased BAC screening filters (Research Genetics, Inc. Huntsville, AL.).

RNA isolation, Northern blotting, and RNase protection

Total RNA was isolated and polyA-selected using commercially available kits (5 Prime 3 Prime, Inc. Boulder, CO.). PolyA(+)RNA (~2.5 mg per lane) was electrophoresed through 0.8% agarose gels containing 19% formaldehyde, which were soaked 30 min. in 50 mM NaOH, neutralized briefly in 0.1 M Tris-HCl, and blotted to Duralon nylon membranes after electrophoresis (Stratagene, Palo Alto CA) in 20X SSC. Blots were rinsed in 2X SSC and baked 2 hr. under vacuum, and UV crosslinked at 250 mJoules/mm² in a Stratalinker (Stratagene) before hybridization.

RNase A protection assays were performed using the RPA II kit (Ambion, Inc. Austin, TX) essentially as per manufacturer's instructions. Briefly, 10 mg total RNA was hybridized overnight with ~5.0 x 10⁴ cpm of ³²P-UTP labeled E550 subclone and B-actin control (provided by the manufacturer). Hybrids were digested with RNase

A, and digests were precipitated and resolved on a 5% denaturing polyacrylamide gel. Gels were exposed using the Fujitsu Phosphorimager II system and dosimetric analysis was performed using MacBas (version2) software. Sense and antisense strands were determined experimentally. Probe titrations and negative controls were performed for each probe reaction.

Subcloning and sequencing mupc7e8.5 and jSub5.1 cDNAs.

Fig.7c depicts the subclones and sequencing of the two original mouse *Eryl* cDNAs mupc78.5 and jSub5.1. The first sequence match identified was between a manual sequencing run from the SP6 side of mupc78.5 to HSC7.541. This sequence includes the mouse version of HSMN7. This sequence was cloned and sequenced independently from mupc78.5 as a 0.4 kb EcoRI-PstI fragment (muMN7-5) confirming the proximal extent of the mupc78.58.5 cDNA. muMN7-5 is the mouse equivalent of HSMN7. Subclone muMN7-2 was also subcloned at this time and it was later found to be contained within a 1.2 Kb sequence block of mupc78.5 by the same logic as mentioned above this provides confirmation by an independent experiment, of the location of the remainder of the embryonic mupc78.5 cDNA.

This point is stressed in order to highlight the fact that mupc78.5 appears to be discontinuous with respect to the corresponding regions in mouse and human *ERY1*. There are several plausible explanations for these results and ultimately this question

will only be resolved by isolation and sequencing of additional *ERY1* positive mouse embryonic cDNAs. The two portions of the cDNA could appear discontinuous simply because this particular cDNA insert was rearranged during the preparation of that particular cDNA library. A more speculative explanation is that the presence of a repeated sequence (MN7 repeat) in this region of the *Ery1* cDNA influences the stability of cDNA inserts. mupc78.5 was isolated from an oligo dt-primed cDNA library with a ~700bp probe that is located at nucleotide ~6000 of a 15 kb RNA. One possible explanation for this result is false-priming by oligo dt caused by its binding to internal poly A/T-rich regions of the cDNA, as has recently be shown for E6AP/UbE3A (Kishino *et al.*, 1998).

3. RESULTS

These results will be presented in four sections.

Section one details early work to define the *jdf2* interval and the construction of a physical map of the *D7H15F37S1-p* genomic interval. This section includes the derivation of mouse cDNA clones for the *D7H15F37S1* and *p* loci and genomic YAC and cosmid clones containing these probes. This section includes deletion mapping data which localizes the genomic deletion breakpoints of several *p* alleles which were identified in the course of this early work. These early results are presented and summarized in Fig.3.

Section two details the generation of a P1 clone contig spanning the *D7H15F37S1-p* interval and genomic mapping data, obtained with P1 end clones, which led to the identification of the first molecular defect in several key *jdf2* alleles. This section also describes the search for gene sequences in the *jdf2* "critical region" defined by physical mapping results. As part of this section I describe collaborative efforts to isolate full-length cDNA sequences representing the *D7H15F37S1* locus gene, and a search for potentially new transcribed sequences in this region. These studies led to identification and full description of a highly conserved gene, *Eryl*, isolated in parallel by our group at ORNL (mouse) and by our collaborators at Case Western Reserve University (in human). The final portion of this section describes the structure and sequence of

full-length human and mouse *ERY1* cDNA clones, and summarizes functional motifs contained within the highly conserved gene.

Section three contains a complete examination of a large selection of *jdf2*, p^x , and p^m alleles relative to genomic clone and cDNA probes. In addition to providing evidence that *Ery1* mutations are the cause of the *jdf2* mutant phenotype, these studies have permitted us to map sequences located within the body of the *Ery1* gene that, when deleted or otherwise disturbed, are responsible for intermediate or mottled expression of *p*. This section also presents additional mapping data that has led to the current composite map of the *D7H15F37S1-p* interval (Figs 8 and 9). It includes Figure 7 which summarizes the sequence alignment of the mouse and human cDNA clones used to determine the full-length mouse and human *ERY1* cDNAs (Fig.7a and 7b) sequences. This section also summarizes all *p* region genomic clones isolated in the course of this work including *Ery1-p* positive BAC clones and distal *p* locus YAC clones (Fig.10) and the subcloning and sequencing of mouse *D7H15F37S1* cDNA clones mupc78.5 and jSub5.1 as well as a map of their orientation with respect to the complete *ERY1* cDNA sequence (Fig. 7c).

Section four focuses upon the identification of the gene representing the *jdf2* locus, called *Ery1*, and experiments detailing the expression profile of the *Ery1* gene in normal mouse tissues radiation-induced *jdf2* alleles and six of the seven known EtNU-

induced *jdf2* alleles. Together with physical mapping data these studies established concordance between *Ery1* genomic rearrangements and/or gene misexpression and the *jdf2* phenotype.

Section 1. Derivation of mouse probes for the *D7H15F37S1* and *p* loci and analysis of selected *p* alleles.

The original experiments aimed at identification of genomic deletion breakpoints of the collection of *p^{jl}* mutations were done by hybridization of genomic Southern blots with human probes for the *D7H15F37S1* and *p* loci. Specifically, we used the original 1.8 kb human cDNA, HSC7.541 originally described by Buiting and colleagues (1992) to represent the *D7H15F37S1* locus and various *TaqI* fragments isolated from the human *P* cDNA (HSCDN10; Rinchik *et al.*, 1993). Although these human probes provided some useful initial data (Rinchik *et al.*, 1993) the sequences represented in both cDNA clones were not well enough conserved to be used effectively in mouse hybridization experiments, such as those involved in screening for longer *D7H15F37S1* cDNAs.

1.1 Derivation of *D7H15F37S1* locus probes.

In an attempt to obtain a better probe for the mouse *D7H15F37S1* locus, a 780 bp *EcoRI*-*RsaI* fragment of HSC7.541 cDNA (isolated from a region flanking the conserved HSMN7 micro dissection clone sequence), was directly compared to the intact

HSC7.541 sequence as a probe on duplicate brain polyA(+) RNA northern blots. The 780 bp *EcoRI-RsaI* (hereafter referred to as pc7.7) yielded an improved signal when compared with the entire HS7.541 cDNA probe. The pc7.7 fragment detected transcripts in polyA(+) RNA isolated from adult mouse brain, liver, lower and upper jaw and palate as well as embryonic 14.5 dpc (days post coitum) brain and jaw (M.W. and D. Johnson unpublished). Based on these trial experiments pc7.7 was chosen as a probe for the initial cDNA library screening attempts to obtain a mouse cDNA clone from the *D7H15F37S1* locus.

Repeated attempts to isolate mouse cDNAs for the *D7H15F37S1* locus were made, involving the screening of two size fractions of a size-selected d14.5pc (days post coitum) embryo library (L. Stubbs, unpublished). From this library a single 1.8 Kb clone named LS-1 was rescreened and isolated as a purified lambda phage clone. The pc7.7 probe was also used to screen a randomly-primed, adult mouse brain cDNA library (GW Chen unpublished). Three primary positives were isolated from this library after screening ~ 300,000 plaque forming units (pfu). A single clone of ~ 3.0 kb, called M23.0 was selected for further study. Concurrently, ~ 300,000 pfu of a mouse testis cDNA library (a gift from Dr. Lee Silver via S. Bultman) was screened with the pc7.7 probe. Two weak positive clones identified in the testis library screens did not replicate in secondary screens.

Finally, a mouse 8.5 dpc embryo cDNA library (a gift from Dr. B. Hogan) was also screened with the pc7.7 probe. A single, replicating, positive clone was isolated in secondary screens (hereafter referred to as mupc78.5). This phage clone contains a *EcoRI* insert of ~1.9 kb that hybridized to YAC clones covering the *D7H15F37S1-p* interval. Hybridization of the phage insert to Southern blots containing genomic DNA from a panel of proven *p* deletion alleles confirmed that this clone mapped proximal of *p* near the *D7H15F37S1* locus. These mapping results are summarized in Table 2 and Figures 5 and 8. Partial sequencing of the mupc78.5 insert from SP6 vector primers verified that the mouse clone contained a region of sequence homology with the human HSC7.541 clone, centered around the HSMN7 conserved microclone sequence. With the isolation of mupc78.5, we felt we had a good probe for further attempts to isolate the complete *D7H15F37S1* cDNA sequence.

Here it should be noted that when the complete sequence of the embryonic cDNA clone, mupc78.5 was aligned with the complete *Ery1* gene sequence (see below), the embryonic subclone was shown to contain two segments of cDNA that are not immediately contiguous in the adult brain cDNA sequence. One 1.2 kb portion of mupc78.5 (total size ~ 1.8 kb) is homologous to *Ery1* nts 2834-4097. The remaining portion of mupc78.5, which overlaps with *Ery1* nts ~5335-5800 represents the mouse version of the human MN7 repeat. Therefore, the mupc78.5 cDNA clone contains sequences that align with the *Ery1* cDNA, but that in the brain cDNA are separated by an

additional cDNA segment of 1 kb. These data are depicted in Fig.7c. These results may reflect an artifact in the mupc78.5 clone, but could also reflect the absence of exon(s) in the embryonic clone due to tissue-specific alternative splicing. For further details on cloning and sequencing of this embryonic cDNA see Materials and Methods. Determining the significance of this observation will require further study.

1.2 Derivation of mouse *p* probes

To identify the 5'-end of the mouse *p* gene, a probe corresponding to the first ~200 bp including the initiation methionine of the published mouse *p* sequence (Gardner *et al.*, 1992) was generated by reverse transcription PCR, (RTPCR) from total brain RNA (D Johnson unpublished). Later, a mouse *p* cDNA, pmc2071 was obtained from Dr. M. Brilliant (Gardner *et al.*, 1992). Since pmc2071 does not include the 5' end of the mouse *p* gene, pmc2071 and 5'*p*TADJ constitute the complete mouse *p* cDNA sequence. Finally the probe 5'*p*TADJ was used to screen a genomic cosmid library (created from DNA of a mutant of the mouse steel locus (*Sl*), on a C3Hf inbred background; L. Stubbs unpublished). Two positively-hybridizing ~35 Kb clones, 5'*p*cos1.2 and 5'*p*cos2.1, were identified. Preliminary mapping of endclones derived from these cosmids confirmed that these cosmids included 5'*p*TADJ and surrounding genomic sequences. One end of cosmid 5'*p*cos1.2, called 1.2E, was found to be deleted in *p*^{9DTW} (Fig.3).

1.3 Preliminary mapping of p^{9DTW} , $p^{12R30lb}$, p^{2DFioD} , alleles relative to p and $D7H15F37S1$ probes.

Molecular analysis was undertaken first for the p^{9DTW} , $p^{12R30lb}$, and p^{2DFioD} alleles. It was predicted that knowledge of the molecular defect in these alleles would be especially informative relative to complementation studies aimed at localization of the genes associated with juvenile lethality, mottling and intermediate expression of p . Dr. Russell and Dr. Rinchik had deduced that these effects would be located proximal to p in part because $p^{4THO-II}$ was able to complement these phenotypic defects (Russell, *et al.*, 1995). Analysis of $p^{4THO-II}$ also permitted the transcriptional orientation of p to be determined with respect to the centromere, since $p^{4THO-II}$ deletes 3' but not 5' portions of the p gene (Nicholls *et al.*, 1993).

The p alleles p^{9DTW} and $p^{12R30lb}$ were thought to be especially good starting points for analysis of the $jdf2$ phenotype for several reasons. First, p^{9DTW} is associated with a severe juvenile lethal (jls) phenotype (Russell *et al.*, 1995) but was thought to be a 'small' deletion since it was shown to be deleted for a small portion of the of p but not for any portion of the available $DH15F37S1$ locus probes (E. Rinchik unpublished). The $p^{12R30lb}$ allele is a prenatal lethal, p -null mutation, and it was expected that cloning of the deletion-breakpoint fusion-fragment in this mutation would allow the physical distance between p and the gene associated with the proximally-located lethal factor to be estimated (Johnson *et al.*, 1995). Knowledge of this

distance was expected to provide information regarding the length of the interval containing *jdf2* and factors responsible for p^x and p^m phenotypes.

Initial indication that p^{9DTW} might represent a deletion came from experiments using the HSCDN10 cDNA as probe (E. Rinchik unpublished). In order to localize the p^{9DTW} deletion breakpoint more precisely we hybridized various *TaqI* fragments of the HSCDN10 cDNA to Southern blots containing p^{9DTW} homozygote genomic DNA samples and appropriate controls. Restriction fragment length variants (RFLVs) specific for p^{9DTW} were detected with four of six restriction enzymes tested, using a 0.9 kb portion of the mouse *p* cDNA pmc2071 (Gardner *et al.*, 1995). Hybridizations with the 0.9 kb probe revealed partial deletion of some portions of the mouse *p* cDNA but, failed to detect size-altered fragments in genomic digests of p^{9DTW} DNA. These data suggest that the deletion breaks within an intron of *p* (Fig.3). Using a similar approach the distal deletion breakpoint of the prenatally-lethal mutation, $p^{12R30lb}$, was localized to a 1.4 kb *TaqI* portion of HSCDN10 (2 of 6 enzymes detected RFLVs in $p^{12R30lb}$ DNA; Fig. 3).

Additional experiments were attempted to localize the $p^{12R30lb}$ deletion breakpoint more precisely. Preliminary experiments localized the distal (*p*-side) breakpoint in $p^{12R30lb}$ within the 1.4 kb HSCDN10 cDNA as mentioned above. The proximal extent of this deletion included all portions of *D7H15F37S1* cDNA that were

available at that time (Johnson *et al.*, 1995). These results are depicted in Fig. 3b and graphically in Fig. 3a. Hybridizations of *p*^{12R30lb} homozygote genomic DNA with the remaining fragments of the mouse *p* cDNA clone (pmc2071) failed to detect additional rearrangements (MW, M.Wu unpublished). These data, suggested that the *p*^{12R30lb} deletion also breaks within an intron of *p*. Since the human *P* gene contains twenty six exons spread over a genomic distance of 200-600 kb, this interpretation is not unlikely (Lee *et al.* 1993).

Since the *p*^{2DFioD} allele complements all other *p* alleles for all genetic functions except pigmentation, it was theorized that this allele may represent an inversion (L. Russell, E. Rinchik unpublished). Consistent with this hypothesis, hybridizations with *p*, *D7H15F37S1*, and *Gabrg3* probes (the *Gabrg3* gene is the closest distal gene marker to *p*, see Figure 3) failed to detect DNA rearrangements in *p*^{2DFioD} DNA by Southern blots prepared from genomic pulsed-field gel electrophoresis (PFGE). This approach was later supplanted by the use of large-insert genomic clones (in this case YACs). Preliminary PFGE data indicated a *p*^{2DFioD}-specific RFLV located ~100 kb from 5'*p* (MW unpublished). The *p*^{2DFioD} mutation has since been shown by cytological analysis to be an inversion with one end within *p* and the other end closely linked to the *c* locus (Russell *et al.*, 1995).

1.4 Isolation and analysis of *p* and *D7H15F37S1* YAC clones.

The original screens for genomic clones corresponding to the *D7H15F37S1-p* interval were done using the Imperial Cancer Research Fund (ICRF) YAC library constructed in the laboratory of Dr. H Lehrach (Larin *et al.*, 1991). Hybridizations were done by Dr. L Stubbs using HSCpc7.541 and a 0.9 *TaqI* fragment of HSCDN10 (corresponding to the most 5' portion of that clone then available). Analysis of two doubly positive clones revealed they were virtually identical in size. Generation of end clones from these YACs and mapping of the end sequences with respect to the *p* deletion complex determined that both clones contained chimeric inserts. YAC clone ICRFY12121 was found to positive for some portions of the HSCDN10 cDNA but was also proven to be chimeric (see Fig. 3). Further rescreening from the original YAC clone stocks identified a non-chimaeric, *D7H15F37S1+/p+* positive clone, ICRFYDO352 (called here YD0352) which was determined by PFGE analysis to be 1 Mb in size (data not shown).

Genomic end clone sequences for YDO352 were generated by the plasmid-rescue technique for the 'left' end (YE352X3) and by inverse polymerase chain reaction (IPCR) for the right (YDO352RT; see M&M). YAC end clone sequences were mapped with respect to *p* by hybridization to several *p* deletions which allowed the proximal-distal orientation of the YAC to be determined (Fig. 3). A long-range PFGE partial-digest restriction map of YD0352 was generated.

Mapping results suggested that, although it was not chimaeric, this clone contained an inversion that reversed the order of *p* and *D7H15F37S1* markers (data not shown). These data were born out by later experiments (see below).

Nevertheless, complete restriction digests of the clone indicated that mupc78.5 and the 5'*p*TADJ probes were located quite close together within YD0352. The two probes detected several of the same small restriction fragments, including a ~280 kb *Bss*HII fragment from YD0352. These results suggested that *D7H15F37S1* probes and the 5' end of *p* were a maximum of 100-200 kb apart. Since we needed a way to confirm and refine these results, it was decided to screen a newly available P1 clone library with *D7H15F37S1* and *p* gene probes (F Francis, unpublished).

The ICRF YAC libraries were screened with the proximal (centromeric) YAC end clone YE352X3, in order to attempt a chromosome walk further in the proximal direction. Unfortunately, clones extending proximally from the *D7H15F37S1* region were absent from the ICRF YAC library, despite the library's depth (approximately 6X genomic coverage) and the Stubbs laboratory's excellent success in obtaining clones in the *p* deletion interval and many other regions (L. Stubbs and colleagues, unpublished). These and other data, including BAC, P1, and cosmid genomic library screening results, have suggested that this proximal region may be

unusually unstable when carried in bacterial or yeast cloning systems.

Section 2. Construction of a bacteriophage P1 clone contig spanning the *jdf2* interval.

Results from the long-range map of YDO352 suggested that *D7H15F37S1* and *p* sequences were located a maximum distance of 280 kb apart. As a means of verifying these preliminary results and obtaining better physical map resolution we screened a bacteriophage P1 library (insert sizes ~80-90 kb). High-density gridded colony filters representing the ICRF P1 library (Francis *et al*, unpublished) were screened with three pooled probes: 5'*p*TADJ, YE352X3 (proximal end clone from YDO352) and the *mupc78.5* cDNA clone. Originally, ten P1 clones were identified by the pooled probes. Upon rescreening five clones were found to contain *mupc7.541* sequences, while only one clone, ICRFP1N15109 (abbreviated here N1) carried both YE352X3 and 5'*p*TADJ. P1 clone N1 was also shown to extend in the distal direction by hybridization with HSCDN10 fragments, 0.9 *Taq*I and 1.4 *Taq*I.

The genomic end clone sequences of clone N1 and one *D7H15F37S1* positive clone, ICRFP1N21306 (abbreviated here N2) were isolated and mapped with respect to the remaining P1 clones and the *p* deletion complex (Fig. 5a and Table 2). These studies revealed that there was no overlap between *D7H15F37S1* and 5'-*p*

positive P1 clones. To fill the gap between the two clones, we used end clones N2PV7 and N1RS7 to rescreen the P1 library. (Note: Generation of terminal or end clone sequences are described in Materials and Methods; the methods employ a PCR step utilizing either T7 or SP6 primers. Therefore for each P1 clone there are two end clones whose designation ends in the either the number 6 or 7). Endclones were chosen for screening based on mapping data, which revealed that the two end clones were located in the interior portion of the *jdf2* region (Fig.5). Library screening identified a single P1 clone, ICRFP1B11152 (abbreviated here B1) that contained sequences corresponding to both N2RV7 and N1RS7 P1 endclones. Mapping the ends of this "bridging clone" with respect to *p* deletion alleles confirmed that it did link *D7H15F37S1* and 5'-*p* positive P1 clones (Fig.5).

Mapping the ends of P1 clones N1, N2, and B1 revealed the first indications of the molecular defects in any *p^{xjl}/jdf2* allele, and also allowed us to orient the P1 contig relative to *p* and the centromere. Endclone and cDNA mapping data are depicted graphically in figures 5a and 8 and are summarized in Table 2. Specifically, hybridization with the endclone N1RS7 indicated for the first time that *p^{48PB}* and *p^{12DTR}* may be small deletions (Fig.5a). These data provided the first indication that the *jdf2* phenotype associated with these alleles might be due to deletion of 3' portions of the transcript associated with the *D7H15F37S1* locus. Furthermore, the interval deleted in these alleles must also include sequences required for the proper expression of

the *p* gene, since these alleles exhibit both *jdf2* and pigmentation defects. Therefore the deleted interval must contain sequences responsible for both viability and pigmentation defects.

In addition to providing the first indications of the molecular defects in several *p^{xjl}* alleles, the P1 contig was shown to extend proximally from *p* to include the *D7H15F37S1* locus. This follows from the fact that clone N1 contains 5'-*p* sequences and endclones sequences from the overlapping clone B1 are both deleted in *p^{6H}*. That the contig extends in the proximal direction, beyond the proximal *p^{6H}* deletion breakpoint, is confirmed by the fact that the most proximal endclone of clone N2, (N2RS6) is not deleted in this mutant (Fig.5 and Table 2). Since *p^{6H}* has historically defined the phenotype and map position of the juvenile lethal/*jdf2* phenotype it was clear that the P1 contig contained at least part of the gene responsible for the *jdf2* phenotype. The P1 contig also completed a physical map of the region between *D7H15F37S1* and *p* and was also therefore likely to contain the regulatory elements responsible for the *p^x* and *p^m* phenotypes. These inferences were strengthened by the detection of deletions within *p^x* mutations *p^{12DTR}* and *p^{48PB}*.

Section 3. Isolation and analysis of cDNAs from the *D7H15F37S1-p* interval.

3.1 Isolation of cDNAs from the *D7H15F37S1-p* interval by subtractive hybridization

Because of its map position and expression patterns, the *D7H15F37S1* locus gene represented a strong candidate gene for the *jd2* locus. However, only 3 kb of the 15 kb gene had been isolated, and repeated attempts to obtain additional segments of the gene through screening size-selected and randomly-primed cDNA libraries with existing probes were unsuccessful. Since the mupc78.5 cDNA and other *D7H15F37S1* locus probes were located on the most proximal P1 clone N2, it seemed likely that other portions of that gene, or perhaps additional unknown transcriptional units, would be found within genomic P1 clones B1 and N1 (Fig. 5).

In an attempt to isolate more of the *D7H15F37S1* locus cDNA and at the same time, address whether there are other genes located between *D7H15F37S1* and *p*, we used a subtractive hybridization scheme. We isolated cDNA sequences corresponding to transcripts that were present in normal animals but absent from the brains of *p^{6H}* mice. This scheme is outlined in Figure 6 (details can be found in the Materials and Methods). This technique relies on the favorable kinetics of solution hybridization to allow enrichment of sequences present in wild type tissues but absent from tissues isolated from

animals carrying the deletion. First strand cDNA fragments and corresponding RNAs form heteroduplexes in all regions in common between the two populations (cDNA:RNA duplexes can form for every transcript/gene except those that are not expressed in p^{6H} animals).

Heteroduplexes are removed from solution after the addition of streptavidin which is bound by biotin that has been photolinked to the p^{6H} 'driver' mRNA prior to hybridization. During phenol/chloroform extraction duplexes are partitioned to the organic phase based on increased molecular mass and decreased organic phase solubility due to binding of streptavidin and multiple biotin molecules. In the aqueous phase remains a single-stranded population of cDNAs that are present in wild type but not the deletion populations. Since we had no reliable method of quantifying this material, one fourth of it by volume was radioactively-labeled with ^{32}P dCTP using random hexamers. The subtracted material was used as probe to screen a random-primed mouse brain cDNA library. Ten of the most strongly hybridizing of over fifty replicating positives obtained in this screen were picked from ~300,000 pfu of the adult brain library. Ten replicating positives were also picked from ~100,000 pfu of the testis library.

Of these twenty original positives, eight were purified in two rounds of screening with the subtractive material as probe until plaque pure. Inserts of these phage clones were tested to determine whether they were deleted in p^{6H} . Of two clones isolated from the

testis library, one was an empty vector, and the other contained an insert which did not map to p^{6H} . Of the remaining six clones, two did not contain inserts, two were not deleted in p^{6H} and two (JSAB4.1 and jSub5.1.2) were found to be deleted in p^{6H} . jSub5.1.2 was mapped with respect to p mutations and P1 clones, and was manually sequenced from vector primers. At the same time this clone was subcloned for complete sequencing and to prepare a short template for riboprobes for *in situ* hybridization experiments (see below). A ~700 bp fragment of jSub5.1 called E550, which represented one end of the original cDNA was subcloned and manually sequenced. This analysis detected overlap between jSub5.1 (via E550) with sequences from the embryonic cDNA mupc78.5. Sequencing of mupc78.5 from SP6 vector primers identified sequence overlap with HSC7.541 that was centered on the HSMN7 conserved repeat sequence. Both mouse probes and the HSC7.541-related sequences mapped to P1 clone N21306, were deleted in the expected p alleles including p^{6H} , and were shown to be overlapping by sequence analysis. We were therefore confident that we had isolated additional cDNA fragments derived from the gene representing the *D7H15F37S1* locus.

3.2 Identification of homology between human and mouse *D7H15F37S1* locus cDNA clones and *ERY1*, a novel human gene isolated from the 15q11-13 region.

As sequencing of the subtracted clone jSub5.1 neared completion, we received human cDNA clones that had been isolated from the syntenically homologous interval of human chromosome 15q11-13 by our collaborators Yonggang Ji and Dr. Rob Nicholls. These human cDNAs were isolated as part of the analysis of a YAC clone spanning a region that is commonly deleted in PW/AS patients (see discussion final section). Hybridization of these human cDNAs to *p* deletions and P1 clones in the region demonstrated that related mouse sequences were located in the *jdf2* interval. Specifically human cDNA YJ1.1 was mapped to P1 clone N21306, and was deleted in several informative *p* locus mutations (see Section 4). Further studies verified sequence homology between human probe YJ1.1 and the proximal end of the jSub5.1 cDNA (Fig. 7c). Sequence homology was also identified between the distal end of YJ1.1 and the mupc78.5 embryonic cDNA. Both sequence matches were of several hundred basepairs at >95% identity (further details can be found in the legend of Fig. 7c). Human cDNA clone YJ1.7, the next adjacent cDNA sequence to YJ1.1, identified restriction fragments of altered size in *p*^{6H} homozygous DNA and mapped to P1 clone N21306. These results were subsequently confirmed by mapping a larger mouse cDNA clone (5504K; see Fig. 7c), which contains part of these human sequences, with respect to *p*^{6H} and P1 clones.

We now know that the YJ1.7 human cDNA corresponds to *ERY1* nts ~900-26000 (Y. Ji *et al.*, in preparation). These data identify *p^{6H}* as a deletion that removes almost all of the *ERY1* cDNA, and that breaks at the proximal side of the P1 N21306. These results were consistent with our original observation that the proximal end clone from N21306 was not deleted in *p^{6H}* homozygous DNA. cDNA mapping experiments therefore verified the fidelity of our P1 genomic clone contig and anchored the contig with respect to the deletion complex (Fig.5). Additional sequencing and mapping data have confirmed that human (YJ1.1, YJ1.7 and HSC7.541 and mouse cDNA clones (mupc78.5 and jSub5.1) correspond to homologous sets of cDNA sequences representing the same conserved gene, called *ERY1* in humans and *Ery1* in mouse respectively (Ji *et al.*, in preparation; Walkowicz *et al.*, in preparation). We will therefore refer to the human gene as *ERY1* and the gene corresponding to the *D7H15F37S1* locus in mouse as *Ery1*.

3.3 Isolation of cDNA clones corresponding to the full-length mouse *Ery1* gene.

To obtain additional cDNA fragments for the >14 kb mouse *Ery1* message the subtracted clone jSub5.1 was used to screen an adult brain cDNA library. Two of three primary positive clones were completely analyzed. Clone 5504K was found to contain an insert of ~4.5 kb, subsequently sequencing has demonstrated this clone

extends from *Ery1* nt 1758-6472 (Fig. 7). This cDNA was subcloned and sequenced by Y Ji and the core sequencing facility at Case Western Reserve University (CWRU). A 0.6 kb subclone termed pH3 representing the extreme 3' end of the 5504K was subsequently used to screen a size-selected, 12 day-postnatal cerebellar cDNA library (Doyle *et al.*, 1997). Despite the proven high complexity and quality of the library and data suggesting that mouse *Ery1* is expressed at high levels in cerebellum (see Section 4, below), no positive mouse clones were isolated. These results were consistent with earlier attempts to isolate the 3'-end of the *D7H15F37S1* locus gene in several previous mouse and human library screening attempts (M. Walkowicz and Y. Ji, unpublished), suggesting that the 3'-end of the *Ery1*/*ERY1* genes may not be carried stably in cloned form. However, we were able to obtain an additional large mouse *Ery1* cDNA from this library using 3'-RACE (rapid amplification of cDNA ends) generated by Y. Ji corresponding to *ERY1* nts 7600-8300 (termed YJR2). The mouse cDNA segment YJR2 was used to screen the cerebellar cDNA library mentioned above. Two primary positives were analyzed and we were able to isolate a 4 kb cDNA clone, termed 556A-2 (*ERY1* nt 6204-10,200; see Fig.7).

The complete *Ery1* cDNA was obtained with additional 3' RACE experiments by Y. Ji. Essentially, the entire 3' half of *Ery1*, beginning at nucleotide 8300 was generated as a single RACE product, termed 7KCD. The 5'-end of *Ery1* was obtained by screening a mouse testis cDNA library with human 5' *ERY1* as probe (Ji *et al.*, in preparation).

Positions of these and other clones which were used to determine the sequence of the full-length mouse *Eryl* cDNA, relative to the sequence of human *ERYL*, can be found below and in Fig.7 and its legend.

Isolation of full length versions of mouse and human *Eryl* cDNAs were done in tandem with our collaborator Y. Ji. Sequencing of mouse and human *Eryl* cDNAs was done in parallel by Y. Ji and the CWRU core sequencing facility. Complete sequence and protein analysis appears in an accompanying paper (Ji *et al.*, in preparation). Alignments between the predicted translation of *Eryl* and those in the data base, which form the basis of predictions on *Eryl* protein motifs and function are presented in Appendix 2 (Ji, Nicholls and colleagues unpublished). The salient features of this analysis are discussed below and are depicted in Figures 7 and 14.

The full-length human *ERYL* has been determined to be 15,286 nts. The human *ERYL* open reading frame (ORF) ends with an ochre stop codon at nt 14,564 and has 720 nt 3' untranslated region (UTR) followed by two polyadenylation sites. The human *ERYL* ORF is 14,505 nts coding for a predicted polypeptide of 4,834 amino acids (Fig. X). Cloning of mouse *Eryl* is described here. Sequence analysis revealed the full-length transcript to be 15,247 nts in length. The overall structure of the mouse *Eryl* gene is similar to that of the human with short 5' and 3' UTRs and two polyadenylation sites. The

mouse transcript contains a single ORF of 14,511, encoding a predicted polypeptide of 4836 amino acids (Ji *et al.*, in preparation).

Comparison of mouse and human *Eryl* cDNA sequences revealed the two genes are highly conserved; within the ORF there is 87% nucleotide identity between the mouse and human genes. Amino acid sequence comparison shows the two proteins are highly conserved, with 95% identity and 99% similarity. Analysis of both mouse and human proteins revealed the presence of several protein motifs. Each predicted protein contains three copies of a RCC1-like domain, a HECT domain and a putative zinc finger. These motifs and what is known of their function is discussed in detail below.

Section 4. Concordance of *Eryl* misexpression with the *jdf2* phenotype and initial characterization of defects leading to p^x and p^m pigmentation phenotypes.

4.1 Description of *jdf2* mutants analyzed.

Table 1 lists nineteen mutations which fail to complement the *jdf2* phenotype or themselves lead to *jdf2* when homozygous. The first four mutations p^{6H} , $p^{7FR601b}$, p^{132G} and p^{116G} are proven deletion alleles, which delete the entire *p* locus and lead to *jdf2* when homozygous. This group includes the canonical *jdf2* mutant allele p^{6H} . Since these deletions included the earliest *D7H15F37S1* cDNA probes available, these alleles were not included in the subsequent

expression analysis. The next eight alleles in Table 1 (*p*^{12R250M}, *p*^{17FATw_b}, *p*^{48PB}, *p*^{12DTR}, *p*^{9DTW}, *p*^{103G}, *p*^{18CoS}, and *p*^{39DSD}) are complex mutations which exhibit both pigmentation effects and decreased lifespan. Finally, seven EtNU-induced 'repeat' alleles which exhibit the same aspects of the *jdf2* phenotype seen in deletion alleles, though with greater severity are listed (generically of the SJ series; eg *jdf2*^{322SJ}.)

4.2 Analysis of DNA structure in radiation-induced mutations.

Data obtained by hybridizing *jdf2* region probes (including *Ery1* cDNA fragments and various genomic endclones) to DNA representing *jdf2* alleles and deletion alleles of the *p* deletion complex are summarized in Table 2. These data contributed to our composite physical and mutational map of the *jdf2-p* interval depicted in Figure 8. Eighteen of nineteen mutations analyzed were found to contain molecular defects that affect the structure and/or expression of *Ery1* coding sequences. Together with expression studies (described in Section 5) these data provide strong evidence that defects in the *Ery1* gene are the primary cause of the disorder expressed by *jdf2* mutant mice. The alleles analyzed in this work fall into several classes. First, there are radiation-induced *p* alleles that delete 3' portions of *Ery1* at their proximal ends but whose phenotype is determined by the distal extent of the deletion. For example, *p*^{116G} fails to complement other *jdf2* mutations but also

studies (described in Section 5) these data provide strong evidence that defects in the *Ery1* gene are the primary cause of the disorder expressed by *jdf2* mutant mice. The alleles analyzed in this work fall into several classes. First, there are radiation-induced *p* alleles that delete 3' portions of *Ery1* at their proximal ends but whose phenotype is determined by the distal extent of the deletion. For example, *p*^{116G} fails to complement other *jdf2* mutations but also fails to complement mutations of the *cpl* complementation group (Culiat *et al*, 1994; see Fig. 1). These mutations also delete the *p* gene sequences and express a complete *p* phenotype. Of four recently described *pxjl* alleles, three (*p*^{12R250M}, *p*^{17FATw_b}, *p*^{48PB}) are small deletions that remove 3' portions of *Ery1* but not the 5'-end of *p* (Fig. 8).

Mapping data obtained with probes YJ1.1, YJ1.7 and the composite clone YJ2.9 allowed us to localize the proximal extent of the deletions *p*^{12R250M} and *p*^{17FATw_b}. Both mutations completely delete the probeYJ1.1, while the longer and contiguous probes YJ1.7 and YJ2.9 detect size-altered restriction fragments in genomic DNAs from these mutants (Fig.5, Table 2). As mentioned above a ~7.0 kb cDNA clone representing the 3' portion of the mouse *Ery1* sequence, 7KCD (ERY1 nt 8300-15247) was generated by 'marathon' RACE cloning from mouse brain RNA by Y. Ji. The 7KCD cDNA clone was mapped with respect to *p* and *jdf2* alleles and to P1 and BAC genomic clones (see below); its position is shown in Figures 7 and 8. Deletion

mapping experiments conducted using the entire 7KCD clone allowed us to derive information about several additional mutations.

For example, we were able to confirm that *p*^{12R250M}, *p*^{48PB}, *p*^{17FAT_{wb}} are all simple deletions affecting 3' portions of the *Eryl* cDNA. Mapping results obtained with a subclone representing the terminal 1.5 kb of this clone '1.5A(+)' were most informative and are described in detail below. Finally, mapping experiments using cDNA 556A-2, allowed us to determine the proximal extent of several deletions. Specifically the proximal extent of deleted material in *p*^{12DTR} is located at *Eryl* nt ~8500 since hybridization of *p*^{12DTR} with a 1.9 kb segment of *Eryl* (nt 6200-8300) detects normal and size-altered restriction fragments (Table 2). It should be noted that in mapping of both large mouse cDNA clones (5504K and RN556) and the 7KCD probe, which together cover *Eryl* nt ~2000-15,286, did not detect positively hybridizing fragments in genomic DNA derived from *p*^{6H} homozygotes. This confirms that in mouse, *Eryl*, is coded from a single locus located just proximal to *p* on Mmu7. This is not the case with human *ERY1* (see final section).

Altogether 9 of 12 radiation-induced mutations analyzed represented clear deletions within the *Eryl* gene region. Three other alleles, *p*^{18CoS}, *p*^{103G}, and *p*^{39DSD}, did not contain clear deletions of *Eryl* sequences. However, specific segments of the *Eryl* cDNA did detect size-altered fragments in DNA isolated from *p*^{39DSD} and *p*^{103G} mutant animals, indicating the presence of DNA rearrangements in

3'-portions of the gene (Fig. 12). No rearrangements have yet been detected with *Eryl* region probes in DNA from the *p^{mjl}* mutation, *p^{18CoS}*.

The three radiation-induced alleles which showed no clear sign of deletion with any probes in the *Eryl*-*p* interval, *p^{18CoS}*, *p^{39DSD}*, and *p^{103G}* were analyzed in greater detail. 1.5A(+) and probes derived from sequentially more proximal locations within the *Eryl* gene (using *BstEII* subfragments of the 7KCD PCR fragment; see Fig.16). Southern blots containing *p^{103G}*, *p^{39DSD}*, *p^{18CoS}* and several *p^x*-viable DNAs were hybridized with probe 1.5A(+) detecting size-altered fragments in *p^{103G}* mutant samples. Finally, a probe corresponding to *Eryl* nt 5700-7500 (called c5561.9, representing the proximal portion of mouse cDNA556A-2), detected size-altered restriction fragments in the *p^{39DSD}* mutation (Fig. 12). These results indicated that genomic rearrangements are present within the region occupied by *Eryl* coding sequences extending from nt 5700-7500 in *p^{39DSD}* DNA. Although the nature of this rearrangement is not yet known, these data provide tools that will simplify further studies of the *p^{39DSD}* allele.

As mentioned above, probes from the *Eryl* region did not detect genomic rearrangement in *p^{18CoS}* DNA. Mutant *p^{18CoS}* mice also appear to express full-length, wild type levels of *Eryl* (see Section 5, below). Therefore, *p^{18CoS}* represents the only one of nineteen radiation-induced *jdf2* alleles that does not carry an

obvious structural defect in the *Ery1* gene. However, since the mutant gene in *p*^{18CoS} has not been sequenced and probes corresponding to intron and intergenic (e.g. *Ery1-p*) regions have not been examined, the structure and activity of *Ery1* in *p*^{18CoS} mice cannot be judged at present.

4.2.2 Mapping sequences mutated in *p*^x and *p*^m alleles.

The molecular defect in the *p*^{103G} allele (which expresses a *p*^{xjlm} phenotype when homozygous) implicates a region located just proximal to 5'*p* and within the 3' end of the *Ery1* cDNA to be necessary for the proper expression of *p* (above and Fig.16). Preliminary RTPCR analysis of the *Ery1* transcript in *p*^{103G}, using primer pairs spanning the entire cDNA detects a 15 kb RNA. Further experiments, such as sequencing the respective genomic and cDNA intervals in this mutation are in progress (Y. Ji pers. comm.).

The *p*^{12DTR} represents an interstitial deletion spanning *Ery1* nt ~9500 to ~13,700, and retains the terminal 1.5 Kb of cDNA including poly adenylation sequences [probe 1.5A(+)]. Since *p*^{12DTR} is also a *p*^x allele, data derived from this mutation also suggests that sequences required for *p* gene expression are located within the genomic interval containing 3'-exons of the *Ery1* gene. Several other *p*^x alleles for example, *p*^{17FAT_{wb}} and *p*^{12R250M}, delete the same portions of *Ery1* that are absent in *p*^{12DTR} mice in addition to other 3'-*Ery1* sequences (Table 2, Fig. 12).

Together these results indicate that defects within the genomic region containing coding sequences for the terminal ~2000 bp of *Eryl* are associated in p^x pigmentation effects as well as the effects seen concomitantly in milder *jdf2* alleles like p^{103G} . Although these studies indicate that an enhancer or other type of regulatory sequence for *p* may be located within a 3'-intron of *Eryl*, the absence of genomic rearrangements in this interval in several p^x -viable alleles suggests that mutations in other sequences in the *Eryl*-*p* interval may also be associated with p^x or p^m phenotypes.

4.3 Mapping *Eryl* and *p* gene sequences relative to the physical map of the *jdf2* region

4.3.1 Assignment of segments of the *Eryl* and *p* genes to specific genomic clone regions.

Hybridization data localizing the *Eryl* and *p* gene cDNAs to various types of genomic clones, and orientation of these clones with respect to the *p* deletion complex, are depicted in Figure 10. Briefly, the 5'-end of *Eryl* is localized to a several BAC clones (BAC.239I3, BAC.241J2, BAC.92D11). *Eryl* probes YJ2.9, the subtracted clone jSub5.1, and the embryonic cDNA mupc78.5 (which are contained within the large mouse *Eryl* cDNA, 5504K) are located on P1 clone N2. The second large mouse cDNA clone 556A-2, (corresponding to nts 6200-10,200) extends across P1 clones N2 and the bridging clone

B1. Finally, the *Ery1* probe 1.5A(+) in addition to hybridizing to 5'-*p* cosmids as mentioned above, detects sequences located on P1 clone N1 which also contains the 5'-half of the *p* gene. BAC clones, covering the *Ery1-p* interval as well as additional genomic clones not explicitly mentioned in this text are presented in Figure 9.

4.3.2 The 3'-end of *Ery1* and 5'-end of *p* are separated by a distance of no more than 10 kb.

In addition to P1 and cosmid clones, we also isolated Bacterial artificial chromosome (BAC) clones with *jdf2* region probes. Hybridization of the *Ery1* 1.5 kb-terminal, PolyA(+)-containing fragment ('1.5A(+)') and 5'*p*TADJ identified commonly hybridizing restriction fragments ~10 kb in length in two independent BAC clones (clones BAC224J13 and BAC258A22). These results suggested that the distance between the two genes is on the order of 10 kb (Fig.8). Both 1.5A(+) and 5'*p*TADJ were also contained within a ~35 kb cosmid clone, 5'pcos1.2, verifying their close juxtaposition in genomic DNA. The close spacing of these two genes has also been confirmed in humans (Ji *et al.*, unpublished)

4.3.3 Mutations that potentially represent complex rearrangements

One further puzzle involves the structure of p^{jl} mutant, p^{17FATw_b} . DNA isolated from p^{17FATw_b} is deleted for a large segment of the 3'-portion of *Eryl* as judged by the fact that sequences detected by probe 7KCD and the 1.5A(+) probe are not present in these animals (Table 2, Fig. 12). However, a probe containing sequences within this 3'-region, B1SP6 (representing the distal endclone of P1 clone B1) are present in p^{17FATw_b} mice, detecting a fragment of altered size in the mutant animals. Since sequence analysis of B1SP6 has shown that the endclone contains *Eryl* coding sequences (nt 13,300 to ~13,550), this enclone should be (1) contained within the larger 7KCD segment and (2) located proximal of the 3'-terminal 1.5.PA+ probe (Fig. 12). These results may suggest that p^{17FATw_b} is a skipping (non linear) deletion. In fact complex, non-linear p alleles affecting the *D7H15F37S1* interval have been identified before (Johnson *et al.*, 1995). Further investigation of this genomic region, including analysis of B1SP6-related sequences in p^{17FATw_b} chromosomes, may eventually explain this discrepancy.

Interestingly, p^{17FATw_b} is a prenatal lethal allele despite the fact that deletion of a larger region that includes the p^{17FATw_b} interval produces a mild *jdf2* phenotype in mutants such as p^{6H} (Russell *et al.*, 1995). A similar early lethality is seen in animals carrying another allele that appears to be a short deletion of the 3'-

Eryl region, p^{48PB} (Table 2, Fig. 8). The notion that these mutations are more complex than current data permits us to evaluate is underscored by the rare revertants reported in matings of p^{48PB} heterozygotes (Russell *et al.*, 1995). Regardless of the relationship between the structure of these alleles and severity of the *jdf2* phenotype examination of this interval in wild type and the mutant chromosomes listed above will be the most direct route to identifying sequences necessary for the p^x phenotype. These results are revisited in the discussion.

Section 5. *Eryl* expression in wild type animals and *jdf2* alleles

5.1 *Eryl* expression in normal mouse tissues.

As mentioned above the human cDNA HSC7.541 has been shown to be expressed in all tissues tested so far (Buiting *et al.*, 1992). Our preliminary experiments and those of our collaborators confirm and extend the idea that *Eryl* is expressed in most mammalian tissues including embryonic palate and lower jaw (MW/D.Johnson unpublished). Since the mouse cDNA mupc78.5 was isolated from a d8.5pc embryonic library, *Eryl* must be expressed at least from d8.5pc onward. Embryonic expression of *Eryl* as determined by RPA is presented in Figure 10b. These data suggest that *Eryl* is present during early development from as early as d8.5pc through at least d16.5pc (MW, CC unpublished).

We also made several attempts to analyze embryonic *Eryl* expression by whole-mount *in situ* hybridization experiments. *Eryl* probe, E550 (see Fig. 7) which was used in RPA experiments described below, was used as an *in situ* hybridization probe on whole mount 8.5 and 10.5 dpc embryos by Dr. Brad Yoder (ORNL). No distinctive pattern was detected in several independent experiments. Obviously these are only preliminary results, but it is entirely possible that this is the *bona fide* expression profile of *Eryl*. That is, it is possible that in the embryo *Eryl* is expressed at low levels in a uniform manner (consistent with expression in most tissues tested so far) that cannot easily be distinguished from background staining. However, definitive analysis will obviously require further experiments.

The expression profile of mouse *Eryl* in normal adult tissues was determined by hybridizing the 4 kb mouse *Eryl* cDNA, 5504K (see Fig. 7) to Northern blots containing polyA-selected RNA from a variety of C3H tissues (prepared by Xiaojia Ren (ORNL); Fig. 10a). These blots were rehybridized with a mouse B-tubulin cDNA probe as a loading control. Our results confirmed that mouse *Eryl* is expressed in most or all mouse tissues tested as a single, >15 kb RNA message. These data are consistent with expression studies conducted on human *ERYL* (Ji *et al.*, unpublished). *Eryl* appears to be expressed at different levels in specific tissues with highest levels

of expression in the brain, and lower levels of transcription in muscle and spleen (Fig. 10a).

In view of the sperm cell defects in *jdf2* animals, *Eryl* expression in C3H testis was compared to total RNA derived from the testis of a W^v homozygous male by RNaseA protection assays (RPA). The W^v mutation leads to a testis with somatic cells but few germ cells (Packer *et al.*, 1995). Preliminary analysis revealed that *Eryl* expression remains in W^v testis, although at lower levels than in C3H testis (Fig 10c). Although preliminary, these experiments are consistent with the hypothesis that *Eryl* is a germ-cell-type specific product (with the reduction of *Eryl* expression in W^v due to the lack of germ cells). This is in keeping with expression of *Eryl*, in all tissues tested thus far, despite the distinct germ cell defects seen in *jdf2* animals.

5.2 *Eryl* expression in *jdf2* mutant mice.

Allelic combinations that were used in expression analysis were chosen in part due to availability of animals, as mentioned above. To summarize, several allelic combinations of both deletion and EtNU-induced *jdf2* alleles, were analyzed. Both EtNU-induced and deletion alleles were analyzed in hemizygous form, eg as compound mutations with known deletion alleles. For instance in the allelic combinations p^{48PB}/p^{6H} and $jdf2^{932SJ}/p^{46DFioD}$ ($p^{46DFioD}$ is abbreviated "46D") expression analysis would test the contributions

of the deletion allele *p*^{48PB} and the point mutation *jdf2*^{932SJ} since both RPA and Northern probes used are deleted in *p*^{6H} and *p*^{46DFioD}. Further details of these expression studies are provided in the legend to Figure 10.

Eryl expression was analyzed in *jdf2* animals by Northern blots and by RPA. Following hybridization with the mouse *Eryl* cDNA clone 5504K, Northern blots were rehybridized with a probe for the mouse dilute gene (*d*; a gift to EMR from Dr. N. Jenkins). This cDNA probe detects a large (~6 kb) mRNA in brain and other tissues and therefore provided a measure of high molecular weight RNA integrity while serving as a loading control. These results established the concordance between *Eryl* expression and the *jdf2* phenotype in two major ways. First, brains of animals carrying several EtNU-induced alleles express lower levels of *Eryl* as compared to brains of normal age-matched controls. Reduced *Eryl* expression was seen in these mutant alleles with both methods (Fig.10d and e). Since the alleles *jdf2*^{322SJ}, *jdf2*^{932SJ}, *jdf2*^{1971SJ} and have subsequently been proven to be point mutations in *Eryl* (see below), the reduced expression of *Eryl* seen in 5 of 6 EtNU-induced *jdf2* alleles analyzed represent strong evidence that defects in *Eryl* are responsible for the *jdf2* phenotype. It should be noted that one EtNU allele, *jdf2*^{932SJ}, seems to produce greater than wild type levels of *Eryl* RNA. These data suggest that the *jdf2* phenotype may arise by other mechanisms than 'apparent' under-expression of *Eryl* message (see Fig.11 lower panel, also Fig.10c).

Evidence that *Eryl=jdf2* is also seen in the fact that, *Eryl* probes detect less abundant mRNAs that are <15 kb in length in animals inheriting *jdf2* alleles *p^{48PB}* and *p^{12DTR}*. Truncated RNAs in these alleles are consistent with the results of physical and deletion mapping which showed that both of these alleles are deletions involving 3' portions of the *Eryl* cDNA. Interestingly, as shown above the *p^{12DTR}* allele is an interstitial deletion; however the presence of the polyA containing 3' terminus of *Eryl* does not seem to significantly protect this transcript from degradation (Fig.10d). Information on various transcripts (arrows) produced by additional alleles can be found in the legend of Figure 10. Expression analysis of these alleles is summarized in Table II. Dosimetry and RPA control runs are depicted in Figure 11. Experimental details can be found in its legend and Materials and Methods. These results are described again in terms of consequence of mutation below.

4. Discussion

Originally we believed *D7H15F37S1h* to be a strong candidate gene for the *jdf2* phenotype based on its location with respect to *p*, expression pattern and transcript size. We have now cloned the entire cDNA for the gene coded by the locus *D7H15F37S1h*. The gene, *Ery1*, fills the *D7H15F37S1h-p* interval. We have determined that rearrangement or mis-expression of *Ery1* in the juvenile brain is almost completely concordant with the *jdf2* phenotype; that is we have been able to detect structural defects in *Ery1* coding sequence and/or abnormal levels or types of mRNA in seventeen of eighteen *jdf2* alleles analyzed. The *Ery1* gene contains a contiguous open reading frame (ORF) encoding for 4836 amino acids, encoded by a genomic locus with exons covering a distance of ~120 kb consistent with the large 'target-size' of this locus as suggested by the isolation of seven allelic EtNU-induced point mutations. Finally, in the process of cloning the full-length *Ery1* gene we have identified restriction fragment length variants (RFLVs) specific for two types of pigmentation affects of *p* which map proximal to *p* as predicted by complementation analysis.

Discovery that mutations in this single gene are responsible for the complex sub-lethal neurologic phenotype termed the 'pink-eye syndrome' now provides a molecular explanation for one of the longest known cases of pleiotropy in the mouse genetics literature.

Eryl contains several highly conserved protein motifs, which are defined by homology to genes that also produce highly-pleiotropic phenotypes when mutated. It is therefore safe to conclude that *Eryl* will prove to be a protein with complex biology. Unfortunately, this highlights the greatest shortcoming of this study, namely the fact that so little is known about the pathology of the *jdf2* phenotype, including the fundamental cause of juvenile lethality. Production of effective antibodies, cellular localization experiments and embryonic expression of *Eryl* continue to be a top priority and are being pursued by our collaborators

This discussion will be divided into seven sections. Then first section summarizes data that implicates *Eryl* as the gene responsible for the *jdf2* phenotype. This section also summarizes results on 'regulatory' mutations of *p*, both mottling and intermediate types, and predictions based on these results. Section 2 discusses one potentially discordant *jdf2* allele and other 'problems' encountered in correlating phenotype with molecular structure for certain alleles. Section 3 details features of the predicted *Eryl* protein based amino acid homologies identified in *Eryl*. Section 4 'Consequence of mutation' includes observations and predictions on the phenotypic consequence of various *jdf2* alleles with regard to molecular defect. This section also details the results of sequence analysis of the EtNU-induced *jdf2* alleles and their predicted outcomes from our collaborator Yonggang Ji (summarized in Fig.13). Section 5 '*Eryl/jdf2* link' presents a best guess as to the function of the *Eryl* protein.

Section 6 'Future directions' details a series of proposed experiments based in part on the analysis of *jdf2* mutant protein and on predictions from the proposed function outlined in section five.

The final section, Sequela describes insights into chromosome structure and the etiology of deletions leading to PW/AS syndromes which involve degenerate copies of the ancestral *ERY1* gene located along HU15q11-13. This work is described in an accompanying study of our collaborators (Amos-Landgraph *et al.*, in preparation).

Section 1. *Ery1* defects are responsible for the *jdf2* phenotype.

Physical mapping results are depicted in Figure 8 and 9. These maps represent the integration of four different types of overlapping genomic clones containing portions of the 14.6 kb mouse *Ery1* cDNA, with respect to all known classes of *p* mutation. As such, they represent our comprehensive maps of the *Ery1-p* interval. Table III summarizes the strength of the concordance between misexpression of *Ery1* with mutations that exhibit the *jdf2* phenotype.

Specifically, eighteen of twenty mutations that are associated with and/or fail to complement *jdf2* were analyzed in these studies (see Table III legend). Seventeen of these eighteen mutations have been shown to contain defects in either the structure or expression of *Ery1* and are therefore concordant with the hypothesis that *Ery1*

mutations are responsible for the *jdf2* phenotype. *Eryl* defects were not found in one *p^{mjlm}* allele, *p^{18CoS}*. Since our analysis was by no means exhaustive and the genomic structure of this 15 kb gene is largely undetermined further examination of the *Eryl* locus in *p^{18CoS}* mice will be required before any definitive conclusions can be made. In fact potential RFLVs have been observed in this allele, and a more intensive analysis based on the 3' portion of the *Eryl* cDNA facilitated by clone reagents isolated in this study should allow resolution of the molecular defect in this allele (MW, unpublished).

Of the seventeen mutations with verified defects in *Eryl* structure or expression, ten alleles were found to be deletions including varying extents of the *Eryl* transcription unit. Several of these alleles express size-altered *Eryl* transcripts including *p^{12DTR}*, which is an interstitial deletion of the *Eryl* gene. Four of the seven existing EtNU-induced *jdf2* alleles were also shown to express *Eryl* transcripts at lower-than-wild type levels. One allele *jdf2^{932SJ}*, was consistently found to express higher-than-normal levels of *Eryl* transcripts in the brains of adolescent mice.

Identification of allele specific polymorphisms for *p^x* and *p^m* alleles.

Structural alterations of genomic DNA sequences were detected in the remaining two mutations (*p^{103G}* and *p^{39DSD}*) examined in this study. Both mutations are polymorphic on the genomic level with

respect to different portions of the *Ery1* cDNA. The contributions of this work with respect to the *jdf2* phenotype as well as insights into various *p*-pigmentation effects are highlighted by examination of the results obtained with three mutations *p*^{103G}, (a *p*^x *jdf2* allele), *p*^{9DTW} (a *p*-null *jdf2* allele, designated jls) and *p*^{18CoS} (a *p*^m *jdf2* allele, designated jlm) (Russell *et al.*, 1995).

Animals homozygous for the *p*^{103G} allele exhibit both *jdf2* and intermediate pigmentation defects. Analysis of *p*^{103G} genomic DNA identified a genomic rearrangement located at the the 3' end of *Ery1* and just proximal to 5'*p*. Probe B1RS6 (the endclone of bridging P1 clone B11152) and an *Ery1* 3' probe, YJB1 (a 1.5 Kb fragment containing both *Ery1* polyA(+) addition sites), both detect *p*^{103G}-specific RFLVs with multiple enzymes. These data indicate the presence of a genomic rearrangement at this site. The probe YJB1, detected *p*^{103G}-specific polymorphic fragments with six of six enzymes tested, whereas B1RS6 detects polymorphisms in only three of six enzymes. These results suggest the rearrangement in *p*^{103G} is closer to the 3' end of *Ery1*.

These results also suggest that sequences required for normal expression of *p* are localized in the small interval including the 3' end of *Ery1* and 5' *p*, an interval identified by these studies to be on the order of 10 kb (see Fig. 8). Preliminary RTPCR experiments indicate that the *p*^x *jdf2* allele, *p*^{103G} produces a full-length *Ery1* message in mouse brain, suggesting the genomic rearrangement in this allele

maybe contained within an intron of *Eryl* (Y. Ji pers. comm.). This result will require confirmation in order to rule out the possibility that the *Eryl* message in *p*^{103G} is not a small deletion which removes some exons, and DNA sequencing maybe needed to rule out a translational frameshift. However, the presence of a full-length *Eryl* message in *p*^{103G} is not incompatible with the genomic rearrangement reported above for this allele. For instance the rearrangement could affect the context or structure of polyA addition sites, resulting in a full-length but unstable message.

This same interval is deleted *p*^{9DTW}, a null allele of *p* which was designated a severe form of juvenile lethality in complementation analysis (Russell *et al.*, 1995). Therefore it is clear that deletions that remove as little as 1.5 kb of *Eryl* cDNA, as in *p*^{9DTW}, are sufficient to produce a severe form of the *jdf2* phenotype. Molecular mapping of other *p*^x alleles also suggests that sequence alterations in the interval containing 3' *Eryl* and 5' *p* as the site of regulatory elements controlling *p*^x expression (see Fig.12). These results have identified a minimal critical region, that contains at least regulatory sequences responsible for the *p*^x phenotype. Genomic sequencing and expression analysis of *p*^{103G}, *p*^{9DTW}, *p*^{12DTR} is ongoing. The deletion breakpoint fusion fragments have been identified in *p*^{12DTR} and *p*^{9DTW} and are being sequenced (Y. Ji, pers. comm.). These results should add to our understanding of this phenomena directly.

Section 2. Outliers and remaining puzzles.

One allele for which we did not detect verified genomic rearrangements, reduced mRNA expression, or aberrant sized transcripts for *Eryl* is the p^m , *jdf2* allele, p^{18CoS} . DNA isolated from animals carrying both p^m alleles, p^{103G} and several p^x -viable alleles were probed in a step wise manner with subclones of the *Eryl* probe 7KCD (which represents the 3' half of *Eryl*; depicted in Fig.12). No size altered fragments were seen in any of these alleles with subclones of the 7KCD cDNA probe located proximal of the 1.5A(+) probe. RFLVs detected in p^{103G} are reported above. A p^{39DSD} -specific RFLV was subsequently detected using a portion of the mouse cDNA clone 556A-2, termed 5562.3, which corresponds to *Eryl* nts ~6200-8100. Polymorphic restriction fragments were seen in two of four enzyme digests tested (Fig.12). The 5562.3 cDNA fragment detects additional RFLVs specific for p^{39DSD} (currently 4 of 7 restriction digests tested; R. Uphoff unpublished). As mentioned above no RFLVs have been detected in the two p^x -viable alleles tested. The study of these alleles included sequences detected not only the 7KCD cDNA probe but also a large portion of cDNAs and genomic endclones reported in this study (MW unpublished). These results suggest that p^{18CoS} and p^x -viable mutations may represent quite subtle DNA rearrangements, consistent with their mild effects on both the *p* and *Eryl* genes. Alternatively, further analysis with *Eryl* probes may reveal a substantial genomic rearrangement as the defect in these mutations. Our analysis nevertheless implicates mutation of the

genomic interval immediately proximal of p to be associated with the p^x phenotype, although at this point we cannot identify a specific sequence.

Remaining puzzles...

Several anomalies which were discovered during the p locus complementation analysis are relevant to this work including the report of rare wild type revertants in matings involving the alleles p^{48PB} and p^{39DSD} (Russell *et al.*, 1995). The p^{48PB} allele is a spontaneous mutation that arose in a control animal during specific locus experiments. Two revertants have been identified in six thousand progeny from crosses between p^{48PB}/p animals. Analysis of these animals revealed one was wildtype and the other was mosaic for wild type and p^{48PB} chromosomes (Russell *et al.*, 1995). With regards to p^{48PB} , Russell speculated that the presence of mosaicism and revertants themselves may suggest that this allele is a complex mutation such as a duplication and deletion (see below). Non-linear deletions which include the region proximal of p have been reported in a recent report on deletion mapping of p locus deletions (Johnson *et al.*, 1994). Regarding reversion of the p^m allele p^{39DSD} , the author notes that the mechanism responsible for mottled (somatically mosaic expression) would presumably not be analogous to reversion seen in an analogous c locus mottling allele c^m . This allele is also known to revert and has been shown to be due to the presence of a retroviral integration (IAP; Russell *et al.*, 1995; Wu *et*

al., 1994). This mechanism is thought to be unlikely since both p^m alleles p^{39DSD} and p^{18CoS} are both radiation induced (see Table I). However, the experiments described in this work would suggest that these alleles are not typical of radiation induced alleles and particularly those at p since we find no evidence that the p^{39DSD} and p^{18CoS} alleles are 'large' deletions.

The final puzzle is reminiscent of issues raised in the consequence of mutation section above. Specifically it concerns the report that *jdf2* alleles p^{48PB} and p^{17FATw_b} have been classified as prenatal lethal in complementation analysis. Examination of the molecular defects in these alleles in comparison to p^{6H} or $p^{12R250M}$, both of which exhibit less severe forms of the *jdf2* phenotype (see Fig. 8), reveals no obvious explanation for this difference. It remains possible that this difference in severity was due to background effects including combined effects of deletions of different lengths (Russell *et al.*, 1995). In fact similar effects, namely decreased viability, have been observed during the generation and propagation of EtNU-induced *jdf2* alleles (Rinchik *et al.*, 1995; D. Carpenter pers. comm.).

This work shows p^{48PB} to be a deletion which extends from *ERY1* nt ~10,500 and deletes the last ~4.5 kb of the *Ery1* cDNA. The deletion stops proximal of 5' p (Fig.8). Though not common there are precedents for spontaneous deletion mutations in the mouse. Particularly analogous is A^Y a pleiotropic allele of the agouti (*a*)

locus which has been shown to be a large (170 kb) spontaneous deletion which produces a pleiotropic phenotype which includes effects on pigmentation and lethality (Michaud *et al.*, 1994). In fact deletion mechanisms which involve repeated sequences are well known in prokaryotes and are becoming recognized as an established mechanism of mutagenesis in humans as well (Section 7). Admittedly more difficult to explain is the reversion of a deletion mutation.

Regardless of whether p^{48PB} is subsequently shown to be more than a simple deletion, reversion remains the only logical explanation for observed wild type progeny from crosses of deletions. Revertant mosaicism is a term coined to describe the presence of normal cells in a field of phenotypically mutant cells. Mitotic gene conversion, that is the repair of a mutation by nonreciprocal transfer of a normal portion of a parental allele for the DNA affected in the mutated chromosome, has recently been reported in humans. The case involves collagen defects which lead to a syndrome which includes somatically expressed blistering of the skin (Jonkman *et al.*, 1997). A compound heterozygote patient was found to have wild type keratinocytes in a field of mutant cells. Haplotype analysis using markers flanking the identified mutations in these revertant (normal) cells was used to show that reversion of one mutation in these cells was due to gene conversion (Jonkman *et al.*, 1997). The authors also report that *meiotic* gene conversion has been postulated to be associated with several human gene families. Meiotic

conversion would be required to explain the reversion of a lethal deletion mutation in a cross necessarily involving deletion over wildtype compound heterozygotes. It should be stressed that whether gene conversion is actually a prevalent mechanism in mammals and whether this phenomena is characteristic of large, complex genes will require further analysis.

Finally, in the course of compiling the tables and maps presented here I have uncovered a mapping discrepancy with respect to the p^{17FATw_b} allele. Specifically, B111RS6, the proximal endclone of the bridging P1, B11152 is polymorphic not deleted in p^{17FATw_b} . This result has been confirmed by independent hybridizations and Southern blots. The p^{17FATw_b} allele has been shown in this study to be a deletion extending distally from *ERY1* nt ~8300 has also been verified in independent experiments.

The P1 end clone B111RS6 has been sequenced and shown to contain sequences corresponding to *ERY1* nt ~13,300 to at least 13,500 (R. Uphoff unpublished). Since this clone is part of the 3' end of *Ery1* probe 7KCD, which is completely deleted in the p^{17FATw_b} allele some contradiction exists (Fig. 8). Since B111RS6 is deleted in all other diagnostic alleles in this region it is clear that this probe does map proximal to p . Three possible explanations for this discrepancy are apparent. First, this result could simply be due to mechanical error, mislabeling or other trivial errors. As mentioned above this seems less likely since identical results have been

obtained in independent experiments. The remaining explanation would implicate the presence of a complex deletion, which are known to exist amongst the Oak Ridge *p* deletion stocks (Johnson *et al.*, 1995). This issue will require at the least sequencing of *Ery1* cDNA and genomic DNA in *p^{17FATw_b}* before a definite explanation for these results can be determined.

Section 3. *Ery1* protein homologies.

The derivation of various cDNA clones which were used to assemble complete cDNAs of mouse and human *Ery1* are described in Figure 7 and its legend. As various cDNA clones were isolated 'end sequences' were generated (both manual and automated methods) and compared to human *ERY1* sequencing data and to the databases using BLAST and FASTA searches using GCG software. Comprehensive analysis of various protein motifs in Fig.18 and described below are detailed in an accompanying paper (Ji *et al.*, 1997).

RCC1

Sequence comparison of initial partial mouse and human *Ery1* cDNAs representing the first 5.0 kb of *Ery1*, revealed significant amino acid homology (initially, 40% similarity, 18-20% identity) to an imperfect tandem repeat consisting of seven copies of a 50-60 amino acid motif repeat found in the cell cycle regulator RCC1,

(Required for Chromatin Condensation). RCC1 was first identified based on its ability to rescue a temperature-sensitive mutation in a hamster cell line which leads to premature chromosome condensation (PCC; Ohtsubo *et al.*, 1987). This pleiotropic mutant cell line also shows defects in transcription, translation and nuclear structure (reviewed in Dasso *et al.*, 1993). RCC1 is an abundant, nuclear protein and is associated with the nuclear matrix as part of a high molecular weight complex, where it is thought to mediate structural and temporal changes in nuclear structure. It has also been shown to act as a guanine nucleotide exchange factor (GEF) for the Ras-like guanine triphosphatase (GTPase) Ran, which promotes the exchange of GDP for GTP and activating the GTPase (Bischoff *et al.*, 1991).

Since its initial discovery, homologs of RCC1 have been isolated in mutant screens from several different experimental systems. RCC1 mutations have been implicated in a host of processes including cell cycle control, specifically those involved in resetting the interphase state after mitosis, protein and RNA transport out of and within the nucleus (to the nucleolus and NPC), protein transport into the nucleus, and RNA processing. RCC1 function has also been implicated in processes controlling chromatin structure and chromosome stability (Dasso *et al.*, 1993; Seki *et al.*, 1996).

Current theories on the mechanism of RCC1 function strive to explain two major functions/features of the RCC1 protein. Each

round of DNA synthesis and replication requires the cell to wait for a signal that synthesis is complete before initiating mitosis. The signal marking the transition from synthesis to mitosis is believed to be mediated through regions of unreplicated DNA. RCC1 is a nuclear protein which associates with DNA as part of a complex of proteins which dissociates from DNA during mitosis (Dasso *et al.*, 1993). Considering the phenotype of RCC1 mutant cells and the protein's documented role as a GEF, one possibility is that RCC1 mediates this signal while acting as a DNA/chromatin-associated GEF. In addition to PCC, which is likely to be a consequence of the cell cycle defect, RCC1 mutant cells exhibit characteristic degeneration of nuclear structure, leading to a series of fused interchromatin granules containing splicing factors and polyA(+) RNA (Huang *et al.*, 1997). It is postulated that RCC1 may mediate structural changes in the nucleus in a manner analogous to Rho and Rac, small GTP binding proteins that regulate the organization of the actin cytoskeleton (Hall *et al.*, 1992).

The micronucleation phenotype exemplified in the RCC1 tsBN2 cell line can be partially complemented by over-expression of nuclear/mitotic apparatus protein (NuMA; Compton *et al.*, 1993). NuMA participates in the organization of the mitotic spindle despite being localized in the nucleus at interphase (Saredi *et al.*, 1996). Due to its observed distribution to the pericentrosomal region of each daughter cell during mitosis and a micronucleation phenotype seen in cells expressing NuMA deletions which remove the globular

domain it has been suggested that NuMA is involved in mitosis in the separation of chromosomes and/or nuclear reassembly (Compton *et al.*, 1993). When NuMA expression is forced in the cytoplasm, by deletion of the nuclear localization signal (NLS), NuMA accumulates as a large mass approximately the size of the nucleus. Transmission electron examination revealed an extensive array of ~ 5 nanometer (nm) filaments incorporating dynamic microtubules into a dense web of 23 nm cables. This result may reflect a structural role for NuMA during interphase role as a component of the core filaments of the nuclear matrix.

Partial rescue of the micronucleation phenotype in RCC1-minus cells by NuMA indicates these proteins are part of the machinery that causes nuclear envelope disassembly/reassembly and formation of an ordered spindle array. I believe that this function will be true of all small RCC1-repeat-containing proteins, as distinct from large proteins containing RCC1 like domains (RLDs) (see below). This view does not necessarily address or preclude evidence that RCC1 can act as a GEF for specific GTPases. Clearly RCC1/Ran interactions are required for various aspects of nuclear transport and structure (Richards *et al.*, 1997). Also site-directed mutagenesis of the yeast RCC1 homolog Prp20, indicates that the amino and carboxy portions of RCC1 are responsible for different cellular functions (Lee *et al.*, 1994).

Eryl has three copies of the RCC1 repeat (RCC1-like domains see Fig.17). Our working assumption is that RLDs in *Eryl* serve to target other proteins that interact with *Eryl*. RLDs may act as GEFs and/or protein targeting sites. However, I do not believe *Eryl* to be an abundant, structural component of the nuclear matrix analogous to RCC1.

Until recently no genes have been identified with significant homology to RCC1, other than direct orthologs from other species. However, recently the gene for the X-linked form of the retinitis pigmentosa type three has been cloned. The gene XLRP3/RP3 shows homology to RCC1 via RCC1-like repeats and has been postulated to be the GEF for a retinal specific GTPase (Roepman *et al.*, 1996; Meindl *et al.*, 1996). The RP3 gene encodes a protein of approximately 500 amino acid residues which is roughly the size of the RCC1 gene. Also the RP3 message is relatively abundant and expressed in most tissues tested as is *Eryl*. However, RP3 is thought to act as a retinal specific GEF with mutations leading the reduction in the visual field caused by degeneration of photoreceptors (Roepman *et al.*, 1996).

'ZZ' type zinc fingers

Eryl also contains a cysteine-rich region which conforms to newly identified zinc finger motif termed the 'ZZ' type finger (Ponting *et al.*, 1996). This cluster of cysteines is thought to form a finger structure incorporating tetrahedral coordinate binding of a zinc ion.

role of the Zn(++) ion in this structure but also refers to the WW amino acid motif thought to mediate protein:protein interactions and includes proteins thought to act as transcriptional adaptor /activator proteins.

The ZZ type finger is found in several interesting proteins all of which seem to reflect different aspects of the *jdf2* phenotype. First among these proteins is Dystrophin, the gene responsible for Duchenne muscular dystrophy (*DMD*). Dystrophin is a large (427 kDa) protein composed of an actin-binding region and a rod-like coiled coil region that contains 24 spectrin repeats, a WW domain and a cysteine-rich region containing a ZZ finger domain.(reviewed in Blake *et al.*, 1995).

ZZ type zinc fingers are also found in the *ref(2)P* gene from *Drosophila melanogaster* (Dezelee *et al.*, 1989). The name denotes alleles that were determined to suppress or were *refractory* to the replication of sigma rhabdovirus multiplication. The homozygous males of *ref(2)* loss-of-function alleles are viable but sterile due to failure of spermiogenesis. Homozygous females are completely normal. The spermatogenic defect seems to involve mitochondrial degeneration specific to spermatids. In addition to the ZZ type finger, *ref(2)* also contains amphiphilic sequences which are thought to be highly hydrophobic regions associated with mitochondrial proteins (Dezelee *et al.*, 1989).

p62 is a protein that was identified as a target for phosphorylation by the lymphoid-specific Src-related kinase p56^{lck}, which is mediated by binding of an SH2 adaptor domain to a phosphotyrosine moiety in the target (Cohen, *et al.*, 1995). p62 has recently been shown to bind ubiquitin in a non covalent, ATP-independent manner. Since p62 shows no homology to either E2 and E3 proteins of the ubiquitin activating pathway or the 23S proteosomal related proteins, p62 represents a novel class of ubiquitin binding protein. p62 has been cloned and found to be expressed in all tissues tested (Joung *et al.*, 1996). This suggests that p62 may act as a common signal transducer in many cell types. p62 represents a link between signal transduction and protein degradation pathways. It is to be expected that the biology of such checkpoint genes will be complex. However several possibilities are evident. p62 could target a substrate (a kinase or itself) for degradation through its interaction with ubiquitin. Tight regulation of phosphorylation/activation and ubiquitination/degradation cycles are known to be involved in controlling essential cellular processes (see below).

Ubiquitination and HECT DOMAINS

Eryl also contains a carboxy terminal HECT domain (homologous to E6AP Carboxyl terminus), a ~300 aa motif with an absolutely conserved cysteine residue (Huibregtse *et al.*, 1995). E6AP is a 100 kDa cellular protein which interacts with the human papillomavirus protein E6, leading to the selective targeting of the tumor suppressor protein p53 for degradation by the ubiquitin/proteasome-mediated protein degradation pathway. The E6AP/E6 complex acts as a E3 ubiquitin ligase initiating the selective degradation of the tumor suppressor p53 and therefore carcinogenesis (Scheffner *et al.*, 1993).

Ubiquitin (Ub) is a highly stable 76 aa polypeptide that is found universally in eukaryotes. Experiments in yeast have established that aberrant and short-lived proteins are targeted for degradation in the proteasome by the addition of polyubiquitin (Bartel *et al.*, 1990). Ubiquitin is added to target via a sequential transfer of Ub in the form of a high-energy thiol-ester bond. Ub-activating (E1) enzymes activate Ub by the formation of thiol-ester linkage in an ATP dependent process. Ub is then transferred to an E2 (Ub-conjugating enzymes) which associate alone or with other different E2's and usually into a complex with an E3 type Ub protein ligase complex. This complex transfers Ub to the target protein (Huibregtse *et al.*, 1995). The specificity needed to ensure that only certain proteins are targeted for destruction is mediated at the level of the

E3 enzymes, possibly in part by their ability to interact with more than one E2 (Nuber *et al.*, 1996). A large number of E2 s have been identified and the existence of many E3 s is assumed, in part since mutations in cells defective in the yeast N rule degradation pathway are capable of bulk protein transport (Bartel *et al.*, 1990).

Ubiquitination of certain proteins, as elucidated in yeast and later for the human papillomavirus E6 protein, ultimately ends with the degradation of target proteins via the proteosome. Ubiquitination has now been implicated in a diverse set of essential cellular. It is logical that mutations affecting the many constituents of this pathway would have pleiotropic effects. This is exemplified by the example of Ubc9p an E2-type ubiquitin conjugating enzyme. Ubc9p's interaction with RanGAP1, a GTP-binding protein which is part of the Ran/RCC1 complex necessary for nuclear transport and cell cycle progression discussed below, is particularly useful in highlighting the pleiotrophy expected when essential/checkpoint genes are mutated.

RanGAP1 is a GTPase-activating protein which works together with RCC1 as GEF to control the hydrolysis GTP by Ran (Saitoh *et al.*, 1997). Disruption of any of these genes lead to characteristic defects in mRNA processing, nuclear transport and nuclear structure. RanGap1, which was originally thought to be a cytosolic part of the nuclear import machinery, has recently be found associated with the cytoplasmic filaments of the nuclear pore complex (NPC). This

localization correlates with covalent modification RanGAP1 by the addition of a ubiquitin-like molecule termed GMP (GAP modifying protein1). GMP-modified RanGAP1 relocates from the NPC to the mitotic spindle apparatus (Matunis *et al.*, 1996). The yeast homologue of GMP1, smt3p was identified as a supressor of MIF2 a cenromeric protein required from spindle integrity and is also known to be ligated to certain proteins (Meluh *et al.*, 1995). Two different approaches have shown that Ubc9p, itself an E2 in the Ub pathway, also acts as an E2 in the Smt3p pathway (Johnson *et al.*, 1997; Saitoh *et al.*, 1997). Since both of these tagging/modification pathways are responsible for the normal progression of the cell cycle and Ubc9p intersects those pathways it is easy to see that defects in Ubc9p function would have pleiotropic effects. Recent attempts to identify Ubc9p catalyzed ubiquitination have been unsuccessful. Though not definitive these data do suggest more details are needed to complete our understanding of these systems.

Overexpression of RCC1 has been shown to rescue certain yeast microtubule mutations, reinforcing RCC1's role as structural part of the nucleus which also acts as a part of the mitotic spindle machinery. Cellular divison involves the partioning of all cellular constituents (including vesicle and cytoskeleton) and reestablishment of functioning systems in daughter cells (including appropriate nuclear architecture and a functioning nuclear import/export system). Defects in RCC1 are therefore inherently pleiotropic, due to this protein's role in this essential transition. The relationship

between RCC1 function and the RLD elements in *Eryl* remains to be determined experimentally. For instance it seems logical that RCC1's normal function maybe disrupted in the sperm of *jdf2* males. What remains to be determined is whether RCC1 is merely affected or defects due to the lack of a RCC1 function, provided in the RLDs of *Eryl* are responsible for aspects of the *jdf2* phenotype like the sperm cell defects as well as the ultimate pathology in these animals.

Finally several other ubiquitin pathway proteins have been identified that are expressed conspicuously in the testis. Ubely, a candidate spermatogonal proliferation gene, and its X chromosome copy are homologous to the E1-Ub activating enzyme, UBE1. UBE1, is itself essential at two stages of the mitotic cell cycle (Odorisio *et al.*, 1996). HHR6A and B are homologues of the yeast N-rule E2, RAD6. Mutations in RAD6 are pleiotropic, implicating this protein in post replication repair, repression of retrotransposition and sporulation. Rad6 has also been shown in vitro to attach Ub to histones leading to E3-dependent degradation. HHR6A and B are expressed in all tissues and cell lines examined so far and increased expression is seen in the round spermatids of the testis. These data are consistent with a role for these E2 ubiquitin enzymes in the removal of histones and their replacement with transition proteins and protamines (Koken *et al.*, 1996).

As with RCC1 homologies, attributing specific, definitive functions to proteins associated with the Ub pathway is difficult due

to the wide range of interconnected processes which involve Ub. Also as the data on Stm3p and recent results which fail to detect Ubc9p-Ub thioesters is meant to highlight what is an evolving field with broad ramifications for many aspects of cell biology. Disruption of an essential complex involving members of the Ub and RCC1 pathways would be consistent with the pleiotropic phenotype associated with *Ery1* mutations in *jdf2* mice. Any explanation of *Ery1*'s function must include both the ultimate pathology in *jdf2* animals as well as explaining specific aspects like those seen in male and female germ cells. Because of its size and several protein-protein interaction domains (ZZ-type fingers and RLDS) it is clear *Ery1* has the potential to be multifunctional and almost certainly has several cellular partners. Due to the presence of the conserved HECT domain I think the safest assumption is that *Ery1* functions as part of a Ub pathway, specifically as a E3 type ubiquitinating factor.

p619/p532

Recently another large human protein has been identified that, like *Ery1*, has both amino and carboxy terminal RCC1-like repeats (RLDs) and a carboxy terminal HECT domain. Originally isolated as part of a transforming/oncogenic DNA fragment, p619 is encoded by a 14 kb message that is expressed in most tissues including brain and testis. Preliminary *in vivo* analysis suggests p619, which is localized to the Golgi complex, has a specific GEF activity associated with both RLD motifs. *In vitro* experiments show that both amino

and carboxy terminal RLD domains stimulate guanine exchange on ARF1 (ADP-ribosylation factor). The amino terminal RLD1 can also act as a GEF for certain Rab proteins though neither domain will interact with or act as a GEF for another small GTPase, Ran (Rosa *et al.*, 1996). ARF and Rab are members of a family of small GTP-binding proteins that are thought to control the specificity and directionality of membrane transport. ARF proteins have been found to be involved in intracellular vesicular transport and phospholipid metabolism (Kahn *et al.*, 1995). Rab (Ran-like GTPase) proteins are associated with docking of transport vesicles to their appropriate membranes. The activity of ARF and Rab is associated with GTP-bound, membrane-associated forms whereas inactive GDP associated forms are mostly cytosolic (Boguski *et al.*, 1993). Therefore p619 is thought to be a part of the cellular machinery that directs the proper polarity of vesicle transport.

Subsequent analysis has resized the p619 protein and renamed it p532. This work also determined that p532 can form stable complexes with heat shock protein 70 (HSP70) and clathrin: this interaction is specific to the cytosol (Rosa *et al.*, 1997). Clathrin is a structural protein of coated membranes involved in receptor induced endocytosis and Golgi sorting (Pleasure *et al.*, 1993). Clathrin has been extensively studied in terms of its role in vesicle formation and transport amongst different Golgi compartments. A role for clathrin in cytokinesis has recently been demonstrated, where it is postulated either is a plasma membrane protein associated with the actomyosin

cytoskeleton or is cytosolic and required for the proper processing/trafficking of membrane bound components of the contractile ring (Niswonger *et al.*, 1997).

Heat shock proteins (HSPs) were first identified because they are preferentially expressed in response to cellular insult. HSPs can be divided into five major families and are amongst the most well conserved proteins known (Schlesinger *et al.*, 1997). Heat shock proteins act as molecular chaperones and are involved in a host of cellular processes where they facilitate interactions between other proteins/complexes/structures. Molecular chaperones assist in the proper folding of proteins including those destined for the mitochondria or chloroplasts, influence clathrin dynamics, and under conditions of stress may refold damaged proteins, remove and refold aggregates or target proteins for degradation or protection (Liang *et al.*, 1997).

Several examples of chaperones that perform functions related to the cytoskeleton and mitotic apparatus have recently been indentified. TCP1, a stress-induced complex of several proteins, is required for proper processing of tubulin and actin and has been shown to be a centrosomal component. HSP70 protects the centrosome during stress and has also been found to bind microtubules (Liang *et al.*, 1997). In the mouse, absence of the male-specific protein HSP70-2, leads to increased apoptosis in pachytene spermatocytes, failure of the synatonemal complex (SC) to desynapse,

and a block in meiotic prophase (Dix *et al.*, 1997). HSP70-2 is also necessary to chaperone cdc2's (cell division checkpoint) binding to its partner cyclin B1, which activates the associated kinase activity required for progression thru the G2/M transition in both mitosis and meiosis (Zhu *et al.*, 1997). HSP70-2 has also been shown by this same group to be a structural component of the lateral element of the (SC; Allen *et al.*, 1996).

Association between p532, clathrin and HSC70 and the previously demonstrated interaction between p532 and ARF1 are consistent with the idea that p532 participates in vesicle transport processes. Proteins with homology to RCC1 are depicted in Figure 19. Initially, we believed that p532 might serve as a good model for *Ery1* since the two were the only large proteins with multiple copies of the RCC1 repeat identified at that time. *Ery1* and p532 are also similar in structure with short 5' and 3' untranslated regions (UTRs) and ORFs of over 4500 aa. Also both genes seem to be expressed ubiquitously at variable levels. That p532 has been implicated in transport processes including Golgi sorting events, and that these processes are involved in specialized transport events such as generation of preacrosomal vesicles, suggests that some parallels between *Ery1* and p532 may be instructive. It is clear that a better understanding of the biology of each of these proteins will require time and careful interpretation. The potential for misleading results and pleiotropic effects that may be expected from mutation of complex proteins like p532 are reinforced by observations that

lethality in clathrin-minus yeast can be rescued by the overexpression of ubiquitin as well as the observation that certain substrates require HSP70 for Ub-mediated degradation (Nelson *et al.*, 1993; Bercovich *et al.*, 1997).

Both *Eryl* and p532 are expected to have multiple cellular partners, some of which may have already been identified for p532. However it is clear that much more work will be needed to determine the significance of these interactions. Proteins that interact with the HECT domain of p532 have been isolated and results on the identity of these proteins is expected (Rosa *et al.*, 1996). I believe that *Eryl* and p532 are similar in that their essential function is as a E3 type Ub protein ligase. Analysis of this emerging family suggests that functionality is conferred to these proteins in part through different protein/protein interaction domains attached to the carboxy HECT domain.

Eryl and p532 clearly differ in this respect (see Fig.18). For example, p532 contains beta repeats domains characteristic of heterotrimeric G proteins and consistent with its proven interaction with ARF1, while this domain is not found in *Eryl* (Rosa *et al.*, 1996; Ji *et al.*, in prep). These structural differences are likely to be reflected in differences in cellular partners for *Eryl* and p532. Finally analysis of recently identified *Drosophila* and chicken expressed sequence tagged sites (ESTs) suggest that *Eryl* is a unique,

conserved E3 type Ub ligase. These data are detailed in terms of a *Ery1/jdf2* functional link in Section 5 below.

Section 4. Consequence of mutation.

Despite the rapid increase in the number of genes that have been cloned and human disease loci that have been characterized, few general rules which predict the consequence of a particular type of mutation have been identified. For instance in-frame deletions in either the actin-binding or rodlike domains of dystrophin lead to a range of mild to relatively severe forms of Becker-type muscular dystrophy whereas mutations that involve the cystine-rich region or carboxy terminal domain are invariably severe (Ponting *et al.*, 1996).

The *jdf2* alleles characterized in this study vary widely with respect to severity of the *jdf2* phenotype, but are not readily distinguishable on the molecular level. For instance *jdf2* alleles *p12R250M*, *p17FATw_b* and *p6H* are all deletions that remove varying amounts of the *Ery1* cDNA starting from the 3' end (see Fig.8). However, these alleles exhibit mild, prenatal lethal and mild forms of *jdf2*, respectively (Russell *et al.*, 1995). At present there is no good explanation for the difference in phenotypic severity in light of these molecular data. An extreme counter-example of this phenomenon is seen in the observation that deletions of any length affecting the ATM gene (ataxia telangiectasia mutated) are all loss of function alleles and produce virtually the same phenotype (Byrd *et al.*, 1996).

It seems that at this point the consequence of a particular kind of mutation is specific to each affected locus.

Figure 4 represents an attempt to group different *jdf2* alleles into an allelic series. It should be stressed that some of this information is based on personal observation and anecdotal information in addition to published data (Russell *et al.*, 1995) and in no way represents a systematic analysis. This conceptual grouping is certainly only one way to depict these results at present and will ultimately be changed when complete information including protein expression data are available. However the observation that point mutations inherited in compound heterozygotes, produce the most severe phenotypic consequence is consistent with the idea that it is more detrimental to produce mutant protein (in this case two types) than none at all (deletions). These data are best explained by the idea that *Ery1* participates in a multi-protein complex, where mutations in an active site allow subunits to bind, and deplete the other members of the complex by incorporating them into non-functional complexes (Maquat, 1996).

Splicing defects usually result from mutations that either destroy or create cryptic splice sites. Defects in splice sites can lead to intron retention, exon skipping or use of a cryptic splice site (Maquat, 1996). Most splicing defects lead to nonsense codons due to shift in translational reading frame. Nonsense coding RNAs result in a reduction in mRNA abundance due to nonsense-mediated

degradation, a process that requires translation. Degradation of missense RNAs is believed to take place at the polysome, preventing the translation of improperly spliced message into defective or truncated protein (He *et al.*, 1995). In general mutations of a given gene produce less severe consequences the closer they are located to the normal termination codon (Maquat, 1996). It will therefore be interesting to look at protein expression in alleles like *p^{12DTR}* and *p^{48PB}*. Both of these alleles, contain intact 5' *Eryl* sequences which would mean the *Eryl* message could be translated for over ~2000 amino acids before the defect in these alleles would be reached. The *p^{12DTR}* allele is an interstitial deletion which retains the 3' end of the *Eryl* mRNA. This may be expected to confer some protective effects on the stability of *Eryl* message regardless of whether this deletion preserves the *Eryl* reading frame (this will be known shortly). However any conclusions concerning the ultimate consequence of these mutations will require protein expression analysis.

Figure 17 depicts a summary of the RTPCR/sequencing analyses which identified point mutations in certain EtNU-induced *jdf2* alleles (Ji *et al.*, in prep). Defects have been identified in three of seven SJ alleles by RTPCR scanning of the *Eryl* transcript followed by sequencing of allele-specific PCR products in parallel with the appropriate genomic controls (BJR). In each case the basic defect has been shown to be due to single base pair change (transitions and transversions) in consensus splice sequences, consistent with the spectrum of EtNU-induced mutations observed at other loci

(compiled in Marker *et al.*, 1997). A strikingly analogous situation to *p^{xjl}* alleles in this study, is the short-ear allele *se^x* (5PDENUS^h), which due to its mild affect on ear size and shape was characterized as an intermediate *se* allele. Molecular analysis of this mutation reveals a splicing defect that would lead to a truncated protein, but still allows production of some wild type message.(Marker *et al.*, 1997).

Briefly, molecular analysis of several EtNU-induced mutations has shown that a splice donor mutation gives rise to exon skipping in the *jdf2^{1971SJ}* allele. The defect results in a translational frame shift and a truncated protein that stops just prior to the ZNF domain. The *jdf2^{932SJ}* point mutation results in an in-frame deletion of a single exon, leading to a deletion of 53 aa, as discussed further below. RTPCR analysis of *jdf2^{322SJ}* revealed the presence of two allele-specific PCR products when amplified with *Ery1* 3' centered primers. Subsequent analysis revealed that one product is produced by exon-skipping, which results in an in-frame deletion of 28 aa. The second product contains a deletion of 7 bp generated through the use of a cryptic splice site; this deletion leads to a translational frame shift and a premature stop (Ji *et al.*, in preparation). The relative amounts of protein produced by each transcript will need to be determined experimentally. It is equally likely that these two transcripts have different stabilities and/or translation efficiencies and these questions will need to be addressed in order to determine the phenotypic consequence of these mutations. However since the splice site mutation in this allele is within the HECT domain homology

it is reasonable to assume that the phenotypic consequences seen in are due to disruption of the *Ery1* HECT domain.

Deleted versions of *Ery1* message seem to be unstable based on limited Northern analysis, whereas exon skipping alleles express less abundant but full-length *Ery1* message (Fig. 13). The absence of any gross pathology in *jdf2* animals could suggest that the function of the *Ery1* protein may be compensated for in some cellular contexts. Further analysis will determine whether *Ery1* defects effect every cell, or that pathology is due to *Ery1* defects in a specific tissue. If the former is the case the assumption would be that *Ery1*'s function can be replaced (compensated for) in unaffected tissues. *jdf2*^{932SJ} would seem a perfect reagent to address some of these questions with respect to *Ery1* as it has been shown to be an in-frame deletion of 53 aa, in a region between obvious protein motifs, in a large multi-domain protein.

Section 5. *Ery1/jdf2* link.

Ery1, a HECT domain-containing protein, is predicted to be a member of the E3 Ub protein ligase family. *Ery1* and p532 are unique among HECT containing proteins in that they share homology to RCC1 (Fig.19). Furthermore, neither p532 nor *Ery1* appear to contain WW and C2 domains, implicated in protein/protein interactions and calcium mediated binding of proteins to the plasma membrane respectively, present in many other E3/HECT proteins.

interactions and calcium mediated binding of proteins to the plasma membrane respectively, present in many other E3/HECT proteins.

Recent database searches have identified *Drosophila* and Chicken expressed sequence tagged sites (ESTs) with homology centered on the carboxy HECT domain of *Eryl*. Comparison of *Eryl* with the fly EST sequence reveals 78% identity over 191 aa, while the same fly sequence and p532 are only 50% identical. The chicken sequence is 100% identical to *Eryl* over 100 aa, in a region located just C terminal to RLD2 of *Eryl* that roughly corresponds to the B repeats in p532 (MW, YJ unpublished). Together with the fact that yeast seem to lack a clear homologue of *Eryl* (R. Nicholls personnel communication), analysis of partial *Eryl* orthologues suggest *Eryl* is derived from an old novel E3/HECT protein distinct from its closest cousin, p532.

Eryl, in addition to the carboxy HECT domain, also encodes several recognized protein:protein interaction domains. It is likely that RLD and ZZ domains in *Eryl* serve to localize or anchor *Eryl*-E3 activity either in time or at a specific location, and in that way regulate degradation or tagging of specific cellular components. It is also possible that *Eryl* and E6AP are similar in the way their E3 activities are regulated. A complex of host E6AP and HPV structural protein E6 is required for ubiquitination p53, and is one of two HPV mechanisms leading to carcinogenesis. E6AP is also thought to bind a host protein equivalent of viral E6. By analogy, *Eryl* may be

proteins that would also restrict/regulate expression of the E3 activity.

E3 proteins share a similar history with RCC1. Initially isolated from a variety of independent screens, subsequent analysis has determined that the same protein is involved in several essential processes, which previously had not seemed related. Such unanticipated, essential function is also likely to be the case with *Eryl*. We have described a highly conserved, novel protein whose expression is seemingly dispensible (as evidenced by long-lived *p^{6H}* animals) at least in terms of survival past one week of age, but whose absence is clearly deleterious to the animal. This would seem to indicate at least two major functional requirements for the *Eryl* protein. One would be systemic: although complete absence of the protein can be compatible with survival, all *Eryl* mutant alleles are runted and sickly, and certain *Eryl* defects lead to a progressive lethal condition and juvenile death. The second function can be seen as the requirement for this protein for male and possibly female fertility. It is also possible that fertility effects are the consequence of the systemic effect. Also it is unclear at this point whether systemic means just that, a whole body requirement, or expression is required at a specific point in embryonic development.

Eryl is predicted to be a novel protein linking the Ub pathway with RCC1. If this association actually occurs *in vivo* the variety of consequences expected from mutation of this gene would be

enormous. As mentioned in section 3, RCC1 already has its own links to the Ub pathway. Furthermore RCC1, by virtue of its association with RanGAP1 and Ubc9p is connected to pathways which lead to the covalent modification of target proteins with Ub and Stmp3. Stmp3/GMP modification has been identified as part of the glucocorticoid receptor transcription initiation complex, and has also been postulated as a tag for mitotic/spindle proteins. Potential links to the *jdf2* phenotype include; breakdown and reassembly of proper nuclear structure and functional nuclear transport machinery concomitant with the assembly/disassembly of the mitotic spindle and cytokinesis, a process that is obviously perturbed in *jdf2* male meiosis. It should also be noted that E3-type equivalents of the Stmp3 pathway have yet to be identified.

Any unifying explanation of *Eryl* function must contend with a host of potential effects. For example, binucleate or biflagellar sperm and polyova defects could be seen as nuclear division defects reminiscent of RCC1. At the same time these defects necessarily involve the actinomyosin cytoskeleton and the nuclear matrix (skeleton). Acrosomal defects and pars nervosa degeneration seen in *p^{6H}* animals both suggest defects in some aspect of vesicular transport. Finally, spermiogenesis, oogenesis and formation of an ordered neurotransmitter vesicle transport system along an axon can also be viewed as examples of specialized cytokinesis, while at the same time being examples of regulated/specialized secretory events.

I believe that *Eryl* will turn out to have an essential role in several distinct cellular processes. This in some sense seems contradictory since *jdf2* animals at least by gross appearance develop almost normally. As mentioned above one possibility is that certain functions of *Eryl* can be, compensated for by other proteins with the most stringent requirement being during male and female meiosis. However, the pars nervosa defect seen in *p^{6H}* was quite subtle when viewed by standard histology and light microscopy. Recent evidence suggests that there are additional death pathways in addition to those that lead to apoptosis and necrosis(REF). This highlights the need for a more careful analysis to determine the pathology of *jdf2* mice.

Section 6. Future directions

The next crucial experiments needed to understand *Eryl*'s function will be the production of high specificity antibodies to the *Eryl* protein. When this reagent can be obtained, large advances into the role of this protein can be expected. This section outlines a series of experiments to understand the role of the *Eryl* protein. These experiments use different *jdf2* alleles identified in this study to produce mutant forms of the *Eryl* protein. These mutant proteins could then be analyzed alone with respect to wild type (cellular distribution for instance) and in combination by crossing different *jdf2* 'mutant protein' alleles. It should be stressed however, that

protein production in these alleles has yet to be determined experimentally.

Another line of experiments that may contribute to our understanding of *Eryl*'s function is the analysis of embryonic expression profile by *in situ* hybridization to whole mount or sectioned mouse embryos over an early embryonic time course. These experiments will necessarily contribute to the overall understanding of *Eryl*'s biology. The success of *in situ* analysis is in ways that are still not fully understood, influenced by the portion of the cDNA used as probe. Therefore failure to obtain specific information from *in situ* analysis using early *D7H15F37S1* cDNA probes described here, may reflect an experimental problem and not a real biological result. *In situ* experiments using human *ERYL* probes are currently being pursued by our collaborators (YJ personnel communication). Embryonic expression analysis would also be complemented by immunohistochemical localization of *Eryl* protein when antibodies become available. However, even distribution of the transcript and protein throughout many tissues may also represent the real *in vivo* state of *Eryl*.

This work has established that *Eryl* is very likely the gene responsible for the *jdf2* phenotype. However formal proof of this relationship requires either ablation of *Eryl* in wild type animals (which would produce the *jdf2* phenotype) or rescue of the *jdf2* phenotype by providing *Eryl* function to mutant animals. Due to

Eryl's large size, complementing *jdf2* by producing *Eryl* transgenic animals will probably be difficult.

Therefore the most feasible approach to prove defects in *Eryl* are responsible for the *jdf2* phenotype would be traditional knock out/homologous recombination experiments. In addition results of *Eryl* protein expression studies could indicate certain defects would be better tolerated. Deletion or disruption of specific domains of *Eryl* may then allow dissection of different functions of this protein. Alternatively, *Eryl* null alleles could be generated by recombination-mediated deletion of a portion of the locus, though it is possible that the true *Eryl*-null phenotype is prenatal lethal. If this is not the case it will be interesting to see if a true null phenotype is equivalent to mild *jdf2* alleles such as *p^{6H}* and *p^{103G}*.

Due to the uncertainty as to whether *jdf2* animals produce any *Eryl* protein some *in vitro* experiments will no doubt be needed to determine *Eryl*'s function. Since it is likely *Eryl* interacts with several other proteins, experiments using portions of *Eryl* as 'bait' in yeast two-hybrid experiments are an obvious choice. These experiments could be expected to address whether the RLD domains of *Eryl* interact with various GTP-proteins, analogous to p53/ARF1 interactions. These studies might also identify an *Eryl*/E3 modifying protein analogous to the cellular form of the HPV encoded E6 protein.

Antibodies to portions of *Eryl* will react with full-length *Eryl* in wild type cells and possibly, in cells carrying certain *jdf2* alleles. As mentioned above, first priority would be to use these antibodies to determine the normal cellular distribution of this protein. Also, antibody staining of *jdf2* mutant testes and ovaries may contribute to our understanding of *Eryl*'s role in these tissues when compared to wild type tissues.

Indications from two-hybrid experiments could then be followed with experiments using anti *Eryl* antibodies in immunoprecipitation experiments to confirm these interactions *in vivo*. The presence of RLD domains would seem to suggest that *Eryl* can act as a GEF for certain GTPases, like Ran, Rab or ARF1. This could be tested by expressing individual RLD domains of *Eryl* in an *in vitro* expression system. Following purification these peptides could be incubated with [H^3] GTP-loaded GTPases (Ran, Rab, ARF1, Rac, Rho) with GEF activity indicated by the release of free [H^3] GTP, analogous to experiments performed with p532 (Rosa *et al.*, 1996).

Another potential avenue to understand *Eryl*'s function could be to test whether *Eryl* retains any functions attributed to RCC1. tsBN2, the original RCC1-mutant cell line, is a mammalian (hamster) leukocyte cell line that exhibits a temperature sensitive phenotype of premature chromatin condensation (PCC; Ohtsubo *et al.*, 1989). The possibility exists that a similar cell line could be derived from *jdf2* animals. Mouse cell lines with temperature-sensitive mutations in

Ub/proteosome pathway proteins have been described (Ikehata *et al.*, 1997; Chowdary *et al.*, 1994). Whether or not such cells are fully viable would be informative in itself, particularly if it turns out to be possible with only certain *jdf2* alleles. A *jdf2* cell line that displays PCC or other obvious defects would be a valuable reagent in determining *Ery1* function. This system could potentially be used to isolate suppressors of the *Ery1* defect by expression from multicopy plasmids analogous to experiments routinely done in yeast.

Obviously no outline of proposed experiments can be expected to be comprehensive since approaches change as new information is acquired. For instance, the *p^{12DTR}* deletion junction fragment and the corresponding genomic interval will be known shortly. This analysis may prove this allele is in fact an in-frame, interstitial deletion. This could indicate a more stable mRNA, which may be translated however the presence of a truncated protein still needs to be proven.

Section 7. Pleiotropy and Sequela.

Pleiotropy

Pleiotropy is defined as the control or determination of more than one characteristic or function by a single gene. In the terms of *jdf2* clarification is necessary. This work shows that *jdf2* which has been associated with certain alleles of *p* is in fact due to mutations in a tightly linked gene, *Ery1*. The *jdf2* phenotype is not due to

pleiotropic effects of the *p* gene as was once proposed. However mutations like *p*^{103G} and *p*^{12DTR} are pleiotropic mutations in that they lead to *jdf2* and also affect (intermediate, mottling) *p* expression.

Eryl mutations now appear pleiotropic and it is expected that *Eryl* has multiple functions. However it is also possible that when *Eryl* is understood one unifying explanation for all *jdf2* effects will emerge. I believe that in addition to specific effects seen in sperm and ova there is a systemic requirement for *Eryl* and this defect is responsible for some undiscovered pathology. It is also possible that both fertility and runting affects have a similar root cause, eg. improper innervation caused by inappropriate *Eryl* expression during embryonic and or juvenile development. In this case better data would explain what at first seemed to be pleiotropy.

Sequela

As was reported above the mouse and human *ERYL* genes are very homologous (~95% aa identity); and this work argues that mutations in *Eryl* in the mouse are responsible for the pathology in animals exhibiting the juvenile lethal *jdf2* phenotype. Therefore it seems a safe assumption that *Eryl* participates in an essential function in both mouse and man. In fact our collaborators are currently searching for mutations in human *ERYL*. Human *Eryl* was isolated by virtue of its proximity to a common deletion breakpoint

cluster seen in a majority of Prader/Willi and Angelmans patients (PW/AS; Amos-Landgraph in press). This work also reports evidence to suggest the common site of deletion in these patients is mediated by low-copy repeat elements ("END" repeats) which are part of *Ery1* coding sequence (Ji *et al.*, 1998 in preparation; Amos-Landgraph *et al.*, 1998).

Repeat mediated deletion/inversion mechanisms have also been proposed in other human disorders including Charcot-Marie-Tooth type-1A neuropathy, Williams disease and others (in Chen *et al.*, 1997). In fact preliminary data on this mechanism in PW/AS (Ji *et al.*, 1997, ASHG 1996 abstract.) has been cited as part of a growing list of examples which suggest that repeat mediated instability may be a common mechanism leading to genomic instability and human disease (Chen *et al.*, 1997; Carrozzo *et al.*, 1997).

Despite the fact that mouse and human *ERY1* genes are highly related, the human locus has diverged significantly. In addition to *ERY1* in the human region of HU15q11-13, there is a closely related locus, *ERY2* and several pseudogene copies of *Ery*-like sequences. The situation therefore seems to be that despite the conservation of *Ery1* coding sequence in mouse and human the chromosomal regions from which the two homologues are coded are significantly diverged. The conservation of the *ERY1* protein argues for an essential role for this protein in both mice and humans. The *ERY1* locus in humans, is in a rapidly evolving region of HU15. One of the more interesting

questions to be addressed in the future will be to determine whether conservation of *ERY1* is assured through an active mechanism, or whether *ERY1* is preserved solely due to its location with respect to a region of chromosomal instability, in this case the PWS/AS deletion breakpoint cluster.

REFERENCES

Allen, J.W.; Dix, D.J.; Collins, B.W.; Merrick, B.A.; He, C.; Selkirk, J.K.; Poorman-Allen, P.; Dresser, M.E.; Eddy, E.M.(1996) HSP70-2 is part of the synaptonemal complex in mouse and hamster spermatocytes. *Chromosoma* **104**:414-421.

Bartel, B.; Wunning, I.; Varshavsky, A. (1990) The recognition component of the N-end rule pathway. *The EMBO Journal* **9**:3179-3189.

Bercovich, B.; Stancovski, I.; Mayer, A; Blumenfeld, N.; Laszlo, A; Schwartz, A. L. Ciechanover, A. (1997) Ubiquitin-dependent degradation of certain protein substrates in vitro requires the molecular chaperone Hsc70. *J Bio Chem* **272**:9002-10.

Bischoff, F.R.; Ponstlingl, H. (1991) Mitotic regulator protein RCC1 is complexed with a nuclear ras-related polypeptide. *Proc. Natl. Acad. Sci.* **88**:10830-10834.

Blake D.J.; Tinsley J.M.; Davies K.E.; Knight A.E.; Winder S.J.; Kendrick-Jones J. (1995) Coiled-coil regions in the carboxy-terminal domains of dystrophin and related proteins: potentials for protein-protein interactions. *TIBS* **20**(4):133-5.

Boman AL; Kahn RA. (1995) Arf proteins: the membrane traffic police? *TIBS* **20**:147-50.

Bryan, J.H.D. (1981) Spermatogenesis revisited. V. Spermiogenesis in mice homozygous for two different male-sterile mutations (p^s and hpy). *Cell Tissue Res* **221**:169-180.

Bryan, J.H.D. (1977) Spermatogenesis Revisited III. The course of spermatogenesis in a male-sterile pink-eyed mutant type in the mouse. *Cell Tiss. Res.* **180**:173-186

Brillant, M.H.(1992) Themouse pink-eyed dilution gene: association with human Prader-Willi and Angelman Syndromes. *Science* **257**:1121-1124.

Buiting, K.; Greger, V.; Brownstein, B.H.; Mohr, R.M.; Voiculescu, I.; Horsthemke, B.(1992) A putative gene family in 15q11-13 and

16p11.2:possible implications for Prader-Willi and Angelman syndromes. *Proc. Natl. Acad. Sci.* **89** 5457-5461.

Buiting, K.; Neumann, M.; Ludecke, H.; Senger, G.; Claussen, U.; Horsthemke, B.(1990) Microdissection of the Prader-Willi syndrome chromosome region and identification of potential gene sequences. *Genomics* **6** 521-527.

Buckwaiter,M.S.; Cook,A.S.; Davisson,M.T.; White,W.F.; Camper, S.A.(1994) A frameshift mutation in the mouse $\alpha 1$ glycine receptor gene (*Glr1*) results in progressive neurological symptoms and juvenile death. *Hum. Mol. Gen.* **3**:2025-2030.

Bercovich, B; Stancovski I.; Mayer, A.; Blumenfeld, N.; Laszlo, A.; Schwartz, A.L.; Ciechanover, A.(1997) Ubiquitin-dependent degradation of certain protein substrates in vitro requires the molecular chaperone Hsc70. *J. Biol. Chem.* **272**:9002-10.

Bondeson, M.L.; Dahl, N.; Malmgren, H.; Kleijer, W.J.; Tonnesen, T.; Carlberg, B.M.; Petterson, U. (1995) Inversion of the IDS gene resulting from recombination with IDS-related sequences is a common cause of the Hunter syndrome. *Hum Mol Gen.* **4**:615-621.

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

Byrd, P.J.; McConville, C.M.; Cooper, P.; Parkhill, J.; Stankovic, T.; McGuire, G.M.; Thick, J.A.; Taylor, A.M.R. (1996) Mutations revealed by sequencing the 5'-half of the gene for ataxia-telangiectasia. *Hum Mol Genet.* **5**:145-149.

Carrozzo, R.; Rossi, E.; Christian, S.L.; Kittikamron, K.; Livieri, C.; Corrias, A.; Pucci, L.; Fois, A.; Simi, P.; Bosio, L.; Beccaria, L.; Zuffardi, O.; Ledbetter, D.H. (1997) Inter- and intrachromosomal rearrangements are both involved in the origin of 15q11-13 deletions in Prader-Willi syndrome. *Am. J. Hum. Genet* **61**:228-231.

Castle W.E., Gates W.H., Reed S.C., Law L.W.(1936) A study of a size cross in mice II. *Genetics* **2** 161-163.

Cheng, K.S.; Manian, P.; Koeuth, T.; Potocki, L.; Zhao, Q.; Chinault, A.C.; Lee, C.C.; Lupski, J. R. (1997) Homologous recombination of a flanking repeat gene cluster is a mechanism for a common contiguous gene deletion syndrome. *Nat Gen* **17**:154-162.

Cheng, Y.; Dahlberg, J.E.; Lund, E. (1995) Diverse effects of the guanine nucleotide exchange factor RCC1 on RNA transport. *Science* **267**:1807-1810.

Cohen G.B.; Ren R; Baltimore D.(1995) Modular binding domains in signal transduction proteins. *Cell* **80**: 237-48.

Compton, D.A.; Cleveland, D.W.(1993) NUMA is required for the proper completion of mitosis. *J. Cell Bio.* **120**:947-957.

Culiat C.T.; Stubbs L.; Nicholls R.D.; Montgomery C.S.; Russell L.B.; Johnson D.K.; Rinchik, E.M.(1993) Concordance between isolated cleft palate in mice and alterations within a region including the gene encoding the beta 3 subunit of the type A gamma-aminobutyric acid receptor. *Proc. Natl. Acad. Sci.* **90**:5105-9

Dasso, M.(1993) RCC1 in the cell cycle: the regulator of chromosome condensation takes on new roles. *Trends in Genetics*. ????

Dezelee, S.; Bras, F.; Contamine, D.; Lopez-Ferber, M.; Segretain, D.; Teninges, D. (1989) Molecular analysis of ref (2)P, a *Drosophila* gene implicated in sigma rhabdovirus multiplication and necessary for male fertility. *The EMBO Journal* **8**:3437-3446.

Dix, D.J.; Allen.J.W.; Collins, B.W. Poorman-Allen, P.; Mori, C.; Blizard, D.R.; Brown, P.R.; Goulding, E.H.; Strong, B.D.; Eddy, E.M. (1997) HSP70-2 is required for desynapsis of synaptonemal complexes during meiotic prophase in juvenile and adult mouse spermatocytes. *Development* **124**: 4595-4603.

Gardner, J.M.; Nakatsu, Y.; Gondo, Y.; Lee, S.; Lyon, M.; King, R.A.; Brilliant, M. H. (1992) The mouse pink-eyed dilution gene: association with human Prader-Willi and Angelman Syndromes. *Science* **257** 1121-1124.

Haldane JBS; (1915) *Journal of Genetics* **5** 133-137.

Hall, A.(1992) Ras-related GTPases and the cytoskeleton. *Mol. Biol. Cell* **3**:475-479.

Handel, M.A.; Washburn, L.L.; Rosenberg, M.P.; Eicher, E.(1987) Male sterility caused by p^{6H} and qk mutations is not corrected in chimeric mice. *J. Exper. Zoo.* **243**:81-92

He, F.; Jacobson, A. (1995) Identification of a novel component of the nonsense-mediated mRNA decay pathwayby use of an interacting protein screen. *Genes and Dev.* **9**:437-454.

Hollander, W.F.; Bryan, J.H.D.; Gowen, J.W. (1959) Pleiotropic effects of a mutant at the p locus from X-irradiated mice. *Genetics* **45** 413-418.

Huang, S.; Mayeda, A.; Krainer, A.R.; Spector, D.L.(1997) RCC1 and Nuclear Organization. *Mol. Bio Cell.* **8**:1143-1157.

Hunt, D.M.; Johnson, D.R. (1971) Abnormal spermiogenesis in two pink-eyed sterile mutants in the mouse. *J. Embryol. exp. Morph.* **26**: 111-121.

Huibregtse J.M.; Scheffner M.; Beaudenon S.; Howley P.M.(1995) A family of proteins structurally and functionally related to the E6-AP ubiquitin-protein ligase *Proc. Natl. Acad. Sci.* **92** 2563-7.

Johnson, D.K.; Stubbs, L.J.; Culiati,C.T.; Montgomery,C.S.; Russell, L.B.; Rinchik, E.M. (1995) Molecular analysis of 36 mutations at the mouse pink-eyed dilution (p) locus. *Genetics* **141** 1563-1571.

Johnson,D.R.; Hunt, D.M.(1975) Endocrinological findings in sterile pink-eyed mice. *J. Reprod. Fert.* **42**:51-58.

Johnson, E.S.; Blobel, G.(1997) Ubc9p is the conjugating enzyme for the ubiquitin-like protein Smt3p. *J. Biol. Chem.* **272**:26799-802.

Jonkman, M.F.; Scheffer, H.; Stulp, R.; Pas, H.H.; Nijenhuis, M.; Heeres, K.; Owaribe, K.; Pulkkinen, L.; Uitto, J. (1997) Rvertant mosaicism in Epidermolysis Bullbosa caused by mitotic gene conversion. *Cell* **88**:543-551.

Joung I; Strominger JL; Shin J. (1996) Molecular cloning of a phosphotyrosine-independent ligand of the p56lck SH2 domain. *Proc Natl Acad Sci* **93**:5991-5.

Kishino, T.; Wagstaff, J. (1998) Genomic organization of the *UBE3A/E6-AP* gene and related pseudogenes. *Genomics* **47**:101-107.

Koken, M.H.; Hoogerbrugge, J.W.; Jasper-Dekker, I.; de wit, J.; Willemsen, R.; Roset, H.P.; grootegoed, J.A.; Hoeijmakers, J.H.(1996) Expression of the ubiquitin-conjugating DNA repair enzymes *HHR6A* and B suggests a role in spermatogenesis and chromatin modification. *Dev. Bio.* **173**:119-32.

Kreig, P.A.(1990) Improved synthesis of full-lengthRNA probes at reduced incubation temperatures. *Nuc. Acids Res.* **18**:6463.

Kim U.J.; Birren B.W.; Slepak T.; Manico V.; Boysen C.; Kang H.L.; Simon M.I.; Shizuya H. (1996) Construction and characterization of a human bacterial artificial chromosome library. *Genomics* **34**:213-8

Lee, A.; Clark, K.L.; Fleischmann, M.; Aebi, M.; Clark, M.W. (1994) Site-directed mutagenesis of the yeast *PRP20/SRM1* gene reveals distinct activity domains in the protein product. *Mol Gen Genet* **245**:32-44.

Liang, P.; MacRae, T.H. Molecular chaperones and the cytoskeleton. (1997) *J. Cell Sci.* **110** 1431-40.

Larin Z; Monaco AP; Lehrach H. (1991) Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. *Proc. Natl. Acad. Sci.* **88**:4123-7.

Lyon, M.F.; King, T.R.;Gondo,Y;Gardener, J.M.;Nakatsu,Y;Eicher,E.M.; Brillant E.M.(1992) Genetic and molecular analysis of recessive alleles at the pink-eyed dilution (*p*) locus of the mouse.*Proc. Natl. Acad.Sci.* **89**:6968.

Malandrini, A.; Villanova, P.; Piomboni,G.; Collodel, M; Spadaro, P.; Guazzi, G.C.(1993)Ultrastructural sperm abnormalities and cerebellar atrophy: does a correlation exist? Report of two cases without endocrine hypogonadism. *J. Submicrosc. Cytol. Pathl.* **25** 371-375.

Maquat, L.E. defects in RNA splicing and the consequence of shortened translational reading frames.(1996) *Am. J. Hum. Genet.* **59**:279-286.

Marker, P.C.; Seung, K.; Bland,A.E.; Russell, L.B.; Kingsley, D.M.(1996) Spectrum of Bmp5 mutations from germline mutagenesis experiments in the mouse. *Genetics* **145**:435-443.

Matunis, M.J.; Coutavas, E.; Blobel, G.(1996) A novel ubiquitin-like modification modulates the partitioning of the Ran-GTPase-activating protein RanGAP1 between the cytosol and the nuclear pore complex. *J. Cell Bio.* **135**:1457-1470.

Meluh PB; Koshland D. (1995) Evidence that the MIF2 gene of *Saccharomyces cerevisiae* encodes a centromere protein with homology to the mammalian centromere protein CENP-C. *Mol. Bio. Cell*, **6**:793-807.

Melvold, R.W. (1974) The effects of mutant *p*-alleles on the reproductive system in mice. *Genet. Res.* **23**:319-325.

Meindl, A.; Dry, K.; Herrmann, F.; Manson,A.; Ciccodicola, A.; Wright, A. (1996) A gene (*RPGR*) with homology to the RCC1 guanine nucleotide exchange factor is mutated in X-linked retinitis pigmentosa (RP3). *Nat. Gen.* **13**:35-42.

Michaud E.J.; Bultman S.J.; Klebig M.L.; van Vugt M.J.; Stubbs L.J.; Russell L.B.; Woychik R.P. (1994) A molecular model for the genetic and phenotypic characteristics of the mouse lethal yellow (Ay) mutation. *Proc Natl Acad Sci* **91**(7):2562-6.

Migheli, A.; Piva R; Wei J.; Attanasio A; Casolino, S; Hodes, M.E.; Dlouhy, S.R.; Bayer, S.A.; Ghetti, B. Diverse cell death pathways result from a single missense mutation in weaver mouse. **151**:1629-38.

Melvold, R.W.(1974) The effects of mutant *p*-alleles on the reproductive system in mice. *Genet. Res., Camb.* **23**:319-325.

Nelson, K.K.; Lemmon S.K.(1993) Supressors of clathrin deficiency: overexpression of ubiquitin rescues lethal strains of clathrin-deficient *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **13**:521-32.

Nicholls, R.D.; Rinchik, E.M.; Driscoll, D.J. (1992) Genomic imprinting in mammalian development: Prader-Willi and Angelman syndromes as disease models. *Sem. Dev. Bio.* **3** 139-152.

Nicholls, R.D.; Gottlieb, W.; Russel, L.B.; Davda, M.; Horsthemke, B.; Rinchik, E.M. (1993) Evaluation of potential models for imprinted and nonimprinted components of human chromosome 15q11-q13 syndromes by fine-structure homology mapping in the mouse. *Proc. Natl. Acad. Sci.* **90** 2050-2054.

Niswonger, M.L.; O'Halloran, T.J. (1997) A novel role for Clathrin in cytokinesis. *Proc. Natl. Acad. Sci.* **94** 8575-8578.

Nomura, N.; Miyajima, N.; Sazuka, T.; Tanka, A.; Kawarabayashi, Y.; Sato, S.; Nagase, T.; Seki, N.; Ishikawa, K.; Tabata, S. (1995) Prediction of the coding sequences of unidentified human genes. I. The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by analysis of randomly sampled cDNA clones from human immature myeloid cell line KG-1. *DNA RES.* **1**:27-35

Nuber, U.; Schwartz, S.; Kaiser, P.; Schneider, R.; Scheffner, M. (1996) Cloning of human ubiquitin-conjugating enzymes UbcH6 and UbcH7 (E2-F1) and characterization of their interaction with E6-AP and RSP5. *J. Bio Chem.* **271** #5:2795-2800.

Odorisio, T.; Mahadevaiah, S.K.; McCarrey, J.R.; Burgoyne, P.S. (1996) Transcriptional analysis of the candidate spermatogenesis gene *Ubely* and of the closely related *Ubex1* shows that they are coexpressed in spermatogonia and spermatids but are repressed in pachytene spermatocytes. *Dev Bio* **180**:336-343.

Ohtsubo, M.; Kai, R.; Furuno, N.; Sekiguchi, T.; Sekiguchi, M.; Nishimoto, T. (1987) Isolation and characterization of the active cDNA of the human cell cycle gene (RCC1) involved in the regulation of onset of chromosome condensation. *Genes & Dev.* **1**:585-593.

Pleasure, I.T.; Black, M.M.; Keen, J.H. (1993) Valosin-containing, VCP, is a ubiquitous clathrin-binding protein. *Nature* **365**:459-461.

Ponting, C.P.; Blake, D.J.; Davies, K.E.; Kendrick-Jones, J.; Winder, S.J. (1996) ZZ and TAZ: new putative zinc fingers in dystrophin and other proteins. *TIBS* **21**:11-13.

Richards, S.A.; Carey, K.L.; Macara, I.G. (1997) Requirement of guanosine triphosphate-bound Ran for signal-mediated nuclear protein export. *Science* **276**:1842-1844.

Rinchik, E.M.; Carpenter, D.A.; Handel, M.A. (1995) Pleiotropy in microdeletion syndromes: Neurologic and spermatogenic abnormalities in mice homozygous for the p^{6H} deletion are likely due to dysfunction of a single gene. *Proc. Natl. Acad. Sci.* **92** 6394-6398.

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

Roepman, R.; van Duijnhoven, G.; Rosenberg, T.; Pinckers, A.J.L.G.; Bleeker-Wagemakers, L.M.; Berger, W. (1996) Positional cloning of the gene for X-linked retinitis pigmentosa 3: homology with the guanine-nucleotide-exchange factor RCC1. *Hum. Mol. Gen.* **7**:1035-1041.

Rosa J.L.; Casaroli-Marano R.P.; Buckler A.J.; Vilaro S.; Barbacid M. (1996) p619, a giant protein related to the chromosome condensation regulator RCC1, stimulates guanine nucleotide exchange on ARF1 and Rab proteins. *Embo J.* **15**:4262-73.

Rosa, J.L.; Barbacid, M. (1997) a giant protein that stimulates guanine nucleotide exchange on ARF1 and Rab proteins forms a cytosolic ternary complex with clathrin and Hsp70. *J. Bio. Chem.* ?????

Russell, L.B.; Montgomery, C.S.; Cacherio, N.L.A.; Johnson, D.K. (1995) Complementation analysis for 45 mutations encompassing the *pink-eyed dilution* locus of the mouse. *Genetics* **141** 1547-1562.

Russell, W.L. et al. Specific-locus test shows ethylnitroso urea to be the most potent mutagen in the mouse. *Proc. Natl. Acad. Sci.* **76**: 5818-5819 (1979)

Saitoh H; Pu R; Cavenagh M; Dasso M. (1997) RanBP2 associates with Ubc9p and a modified form of RanGAP1. *Proc Natl Acad Sci* **94**:3736-41.

Saredi, A.; Howard, L.; Compton, D.A. (1996) NuMA assembles into an extensive filamentous structure when expressed in the cell cytoplasm. *J. Cell Sci.* **109**:619-630.

Scheffner, M.; Huibregtse, J.M.; Vierstra, R.D.; Howley, P.M. (1993) The HPV-16 E6 AND E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53. *Cell* **75**:495-505.

Seki, T.; Hayashi, N.; Nishimoto, T. (1996) RCC1 in the Ran pathway. *J. Biochem.* **120**:207-214.

Shiga, N.; Takeshima, Y.; Sakamoto, H.; Inoue, K.; Yokota, Y.; Yokoyama, M.; Matsuo, M.(1997) disruption of the splicing enhancer sequence within exon 27 of the dystropin gene by a nonsense mutation induces partial skipping of the exon and is responsible for Becker muscular dystrophy. *J. Clin. Invest.* **100**:2204-10.

Stubbs, L.; Huxley, C.; Hogan, B.; Evans, T.; Freid, M.; Duboule, D.; Lehrach, H. (1990) The hox-5 and surfeit gene clusters are linked in the proximal portion of mouse chromosome 2. *Genomics* **6**: 645-650.

Symth, J.R.; Jerszyk, M.M.; Montgomery, N.(1985) Congenital quiver, an inherited neurological defect in the chicken. *J Hered.* **76** 263-266.

Wolfe, H.G.; Erickson, R.P.; Schmidt, L.C.(1976) Effects on sperm morphology by alleles at the pink-eyed dilution locus in mice. *Genetics* **85**: 303-308.

Wolfe, H.G. (1971) Genetic influence on gonadotropic activity in mice. *Bio Repro* **4**:161-173.

Wu M.; Rinchik E.M.; Wilkinson E.; Johnson D.K. (1997) Inherited somatic mosaicism caused by an intracisternal A particle insertion in the mouse tyrosinase gene. *Proc Natl Acad Sci* **94** (3):890-4.

Zhu, D.; Dix, D.J.; Eddy, E.M.(1997) HSP70-2 is required for CDC2 kinase activity in meiosis I of mouse spermatocytes. *Development* **124**:3007-3014.

APPENDICES

Figure 1. Summarized Complementation maps for the Oak Ridge and Harwell *p* locus mutations on Mmu7.

1a. An ideogram of the Oak Ridge complementation map of the pink-eye dilution locus on Mmu7. Various mutations are depicted below with lines representing absent material. Functional units were determined as follows; a cross of $p^{12R301b}$ and p^{25DVT} would lead to prenatal death ($pl-1$). A cross between p^{132G} and p^{116G} would lead to cleft palate and perinatal lethality and these alleles belong to the *cp1* complementation group. Phenotypes above the line are those which have been localized proximal to *p* by Russell *et al.*, 1995. One mutation not included in Oak Ridge testing is p^{6H} and is used as a reference point between the two maps (from Russell *et al.*, 1995).

1b. An ideogram of the Harwell *p* locus complementation map. PL = prenatal lethal, R,JG,S = runted, jerky gait, sterile and CP= cleft palate are proposed functional units. p^{un} is a duplication of >100 kb extending distally from the 3' end of *p*. The star with p^d is meant to represent what was thought to be a small defect (non deletion). Oak ridge mutations such as p^{103G} are thought to be analogous to Harwell mutation p^{bs} in that they both have affects on pigmentation and viability (see text) from (Lyon *et al.*, 1992)

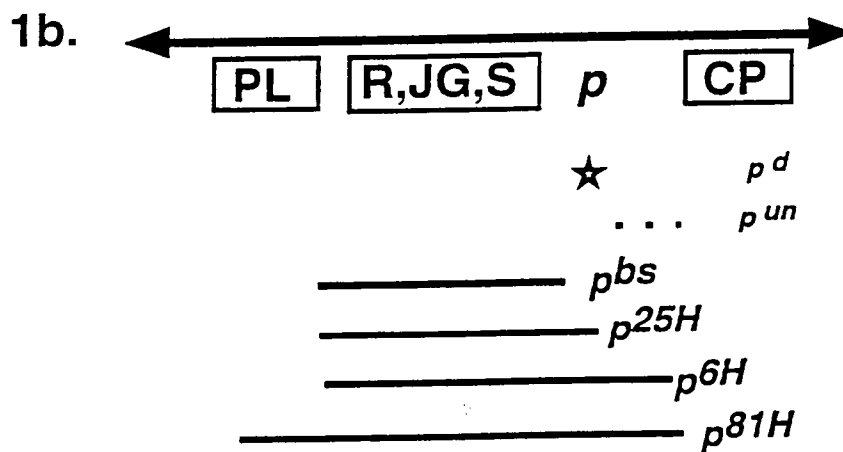
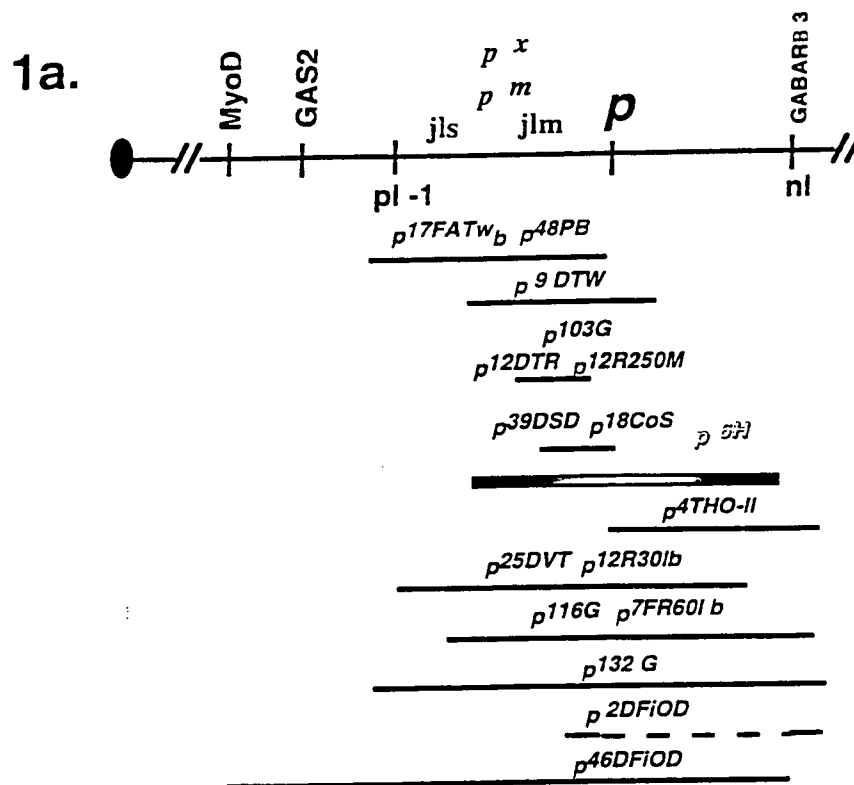


Figure 2. EtNU-induced point mutations within the deletion p^{6H} define the *jdf2* locus proximal to *p* on Mmu7.

Mapping of several EtNU-induced alleles which exhibit the pink eye/*jdf2* syndrome. A complex phenotype seen in certain EtNU alleles and was identified as being linked to *p* since these alleles exhibit the *jdf2* phenotype when opposite a large *p* deletion such as $p^{46DFioD}$. More precise localization of presumed point mutant *jdf2* alleles was achieved by further crosses to $p^{4THO-II}$ and p^{6H} which complement and failed to complement the EtNU *jdf2* alleles respectively (from Rinchik *et al.*, 1995).

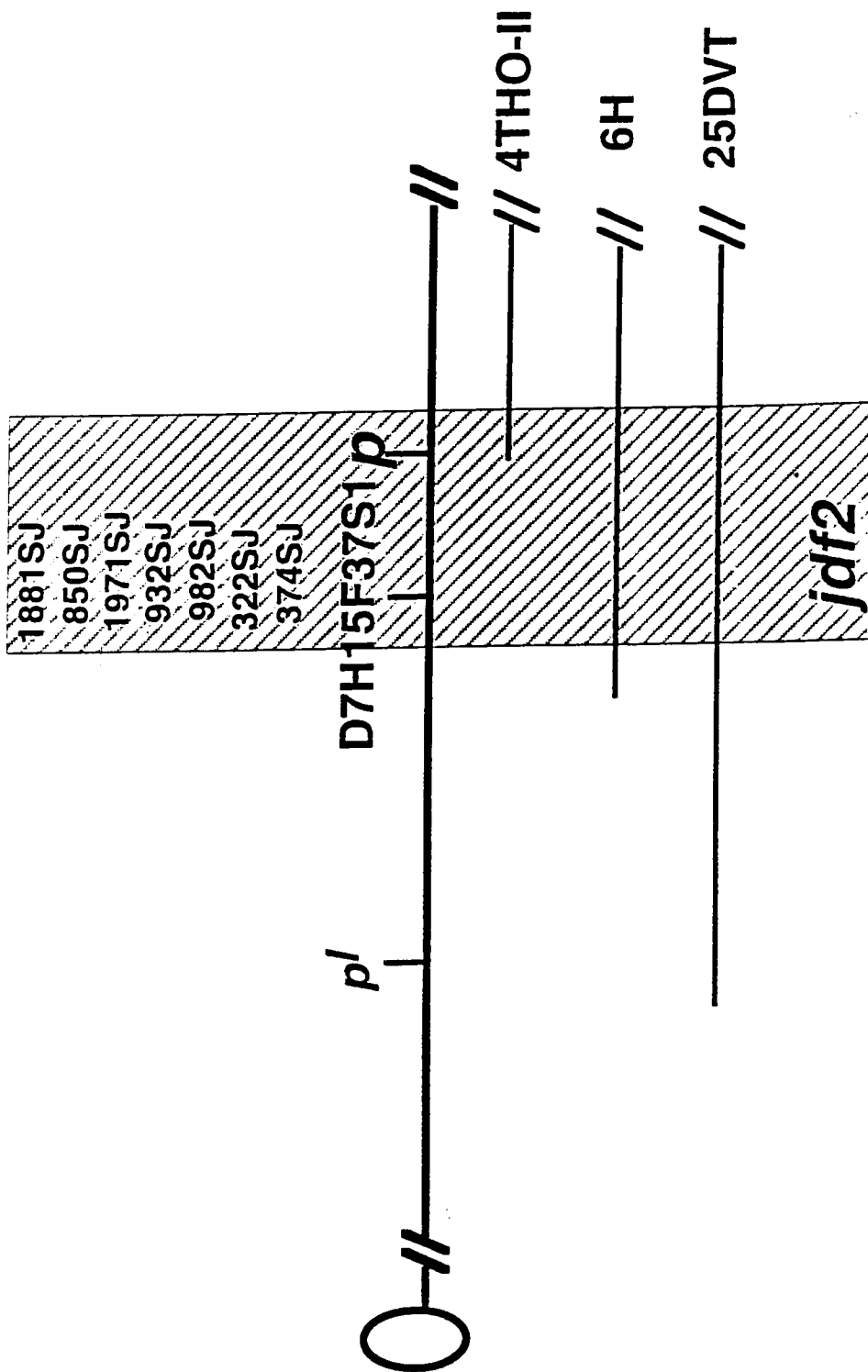


Figure 3. Early results from mapping genomic clones for *D7H15F37S1* and *p*.

This figure depicts cosmid and YAC genomic clones isolated with mouse *D7H15F37S1* and *p* probes (above chromosome) with *p* deletion alleles (below). It includes YAC clone YDO352 which was used to determine these two genes were on the order of 280 kb apart. The location of the proximal deletion breakpoints in *p*^{9DTW} and *p*^{12R301b} alleles are marked with an open box below the probe which detects the breakpoint. Southern blotting data identifying these breakpoints is presented directly below.

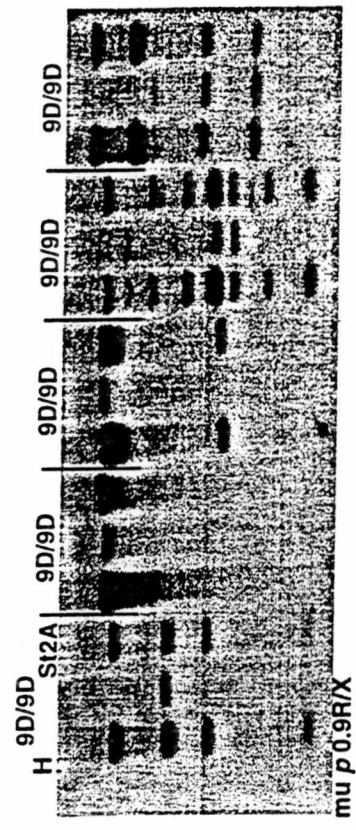
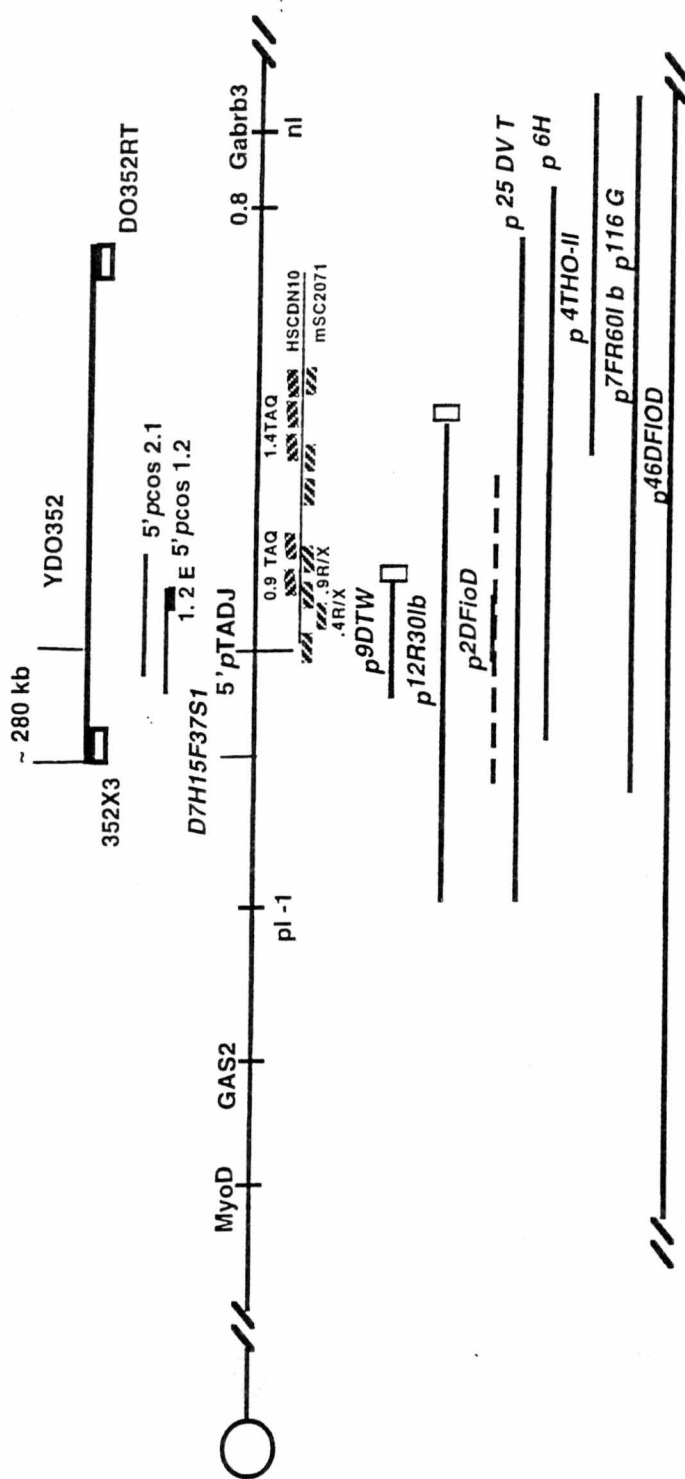


Figure 4. An informal allelic series reflecting the severity of the *jdf2* phenotype.

This figure is an attempt to graphically organize the various allelic combinations which can give rise to the *jdf2* phenotype. Severity of the phenotype decreases from left to right and sterility is a constant, if sometimes immaterial, trait of all these combinations. It should be stressed that this is a generalization, not all combinations are included and relative positions on the continuum are not absolute.

Severe

Mild

$\frac{SJ\ m1}{SJ\ m1}$	$\frac{p18CoS}{p18CoS}$	$\frac{p39DSD}{p39DSD}$	$\frac{p6H}{p6H}$
$\frac{SJ\ m1}{SJ\ m1}$	$\frac{p48PB}{p6H}$	$\frac{p12DTR}{p9DTW}$	$\frac{p6H}{p6H}$
$\frac{SJ\ m1}{SJ\ m2}$	$\frac{SJ\ m2}{SJ\ m2}$	$\frac{p17FATw_b}{p6H}$	$\frac{p103G}{p103G}$
$\frac{SJ\ m1}{SJ\ m2}$	$\frac{SJ\ m1}{del}$	$\frac{p9DTW}{p9DTW}$	$\frac{p9DTW}{p9DTW}$

death by 2.5 wks

sterility

Increased lifespan

Tremors

no obvious tremor

> 50% smaller

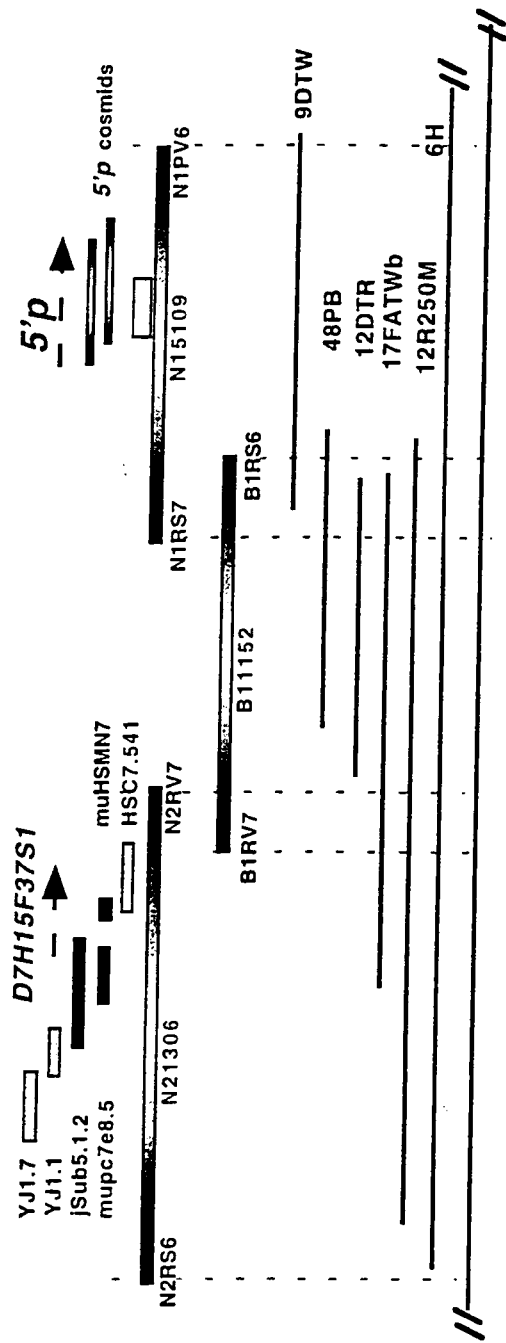
no obvious runting

Figure 5. *Ery1* cDNA and P1 endclone mapping.

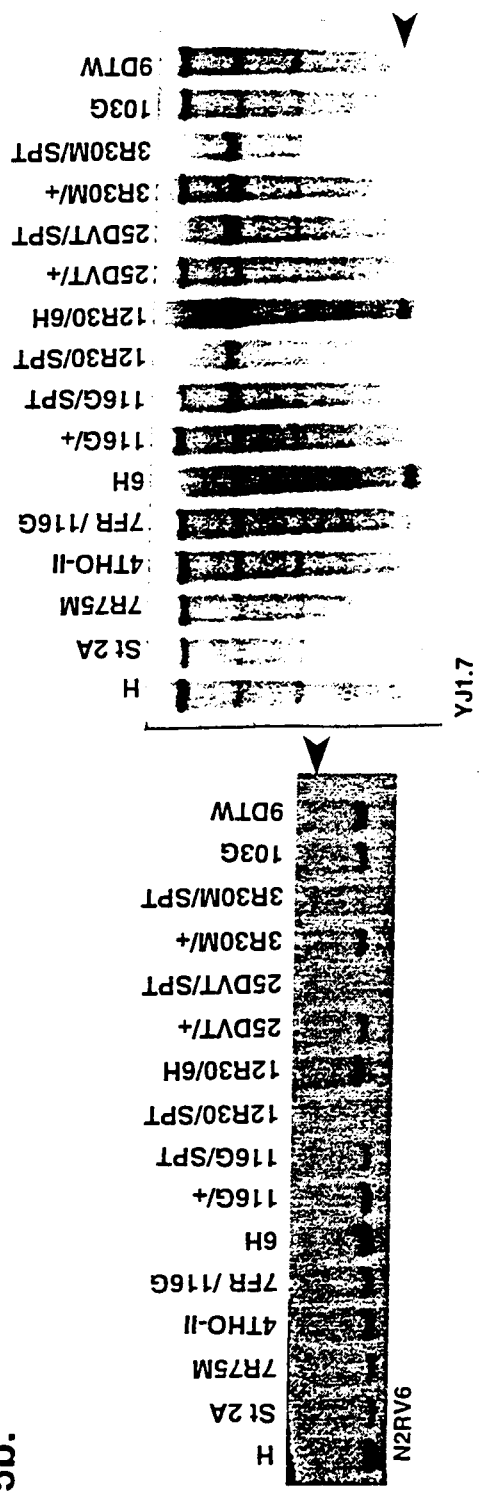
A map of *Ery1* and *p*, P1 clones and cDNAs with respect to various *jdf2* and *p* deletion complex alleles. Solid black lines represent the deleted material in each mutant chromosome. Rectangles above represent genomic P1 and mouse and human *D7H15F37S1* cDNAs. This map depicts a minimal set of P1 clones and initial *Ery1* cDNA clones, a comprehensive genomic clone map is presented in Fig.10. Deletion mapping data for the P1 end clone N2RV6 and human *Ery1* probe YJ1.7 extend beyond the proximal deletion end point of *p*^{6H}, an allele historically associated with the *jdf2* phenotype. The arrow in the N2RV6 panel points to the polymorphic *Mus spretus* band, which seems to be less conserved. The arrow in the YJ1.7 panel indicates a *p*^{6H} deletion breakpoint fragment.

Figure 5c. presents deletion mapping for several informative P1 end clones which allowed the generation of the P1 contig spanning the *jdf2* interval.

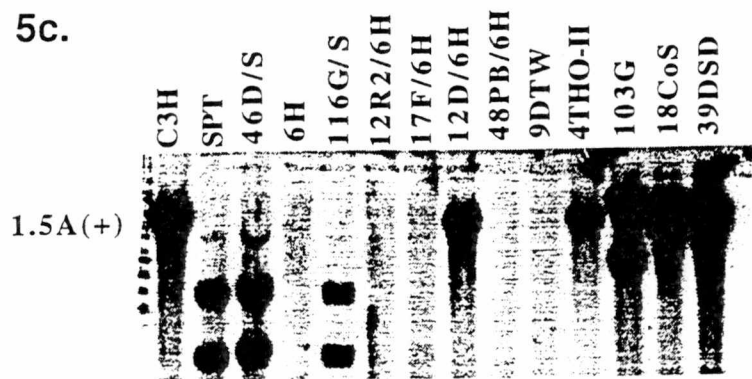
5a.



5b.



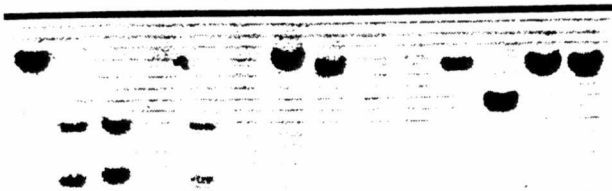
5c.



N1RS7



B1RS6



B1RS6



5'p

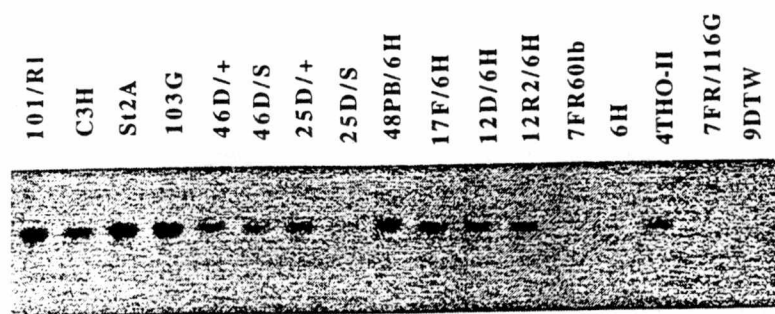


Figure 6. Isolation of *EryI* cDNA jSub5.1 by subtractive hybridization

A scheme to isolate cDNAs for the *EryI* locus from the p^{6H} deletion allele by subtraction. Rationale and experimental details are outlined in Results and Materials and Methods. Formally, this approach could be expected to isolate either cDNAs affected by the genomic deletion in p^{6H} or cDNAs whose expression is affected by the lack of a product absent in p^{6H} . Deletion mapping of one clone isolated by this method jSub5.1 shown in the lower panel, reveals it to be of the former type, that is coded from a locus disrupted in the p^{6H} deletion.

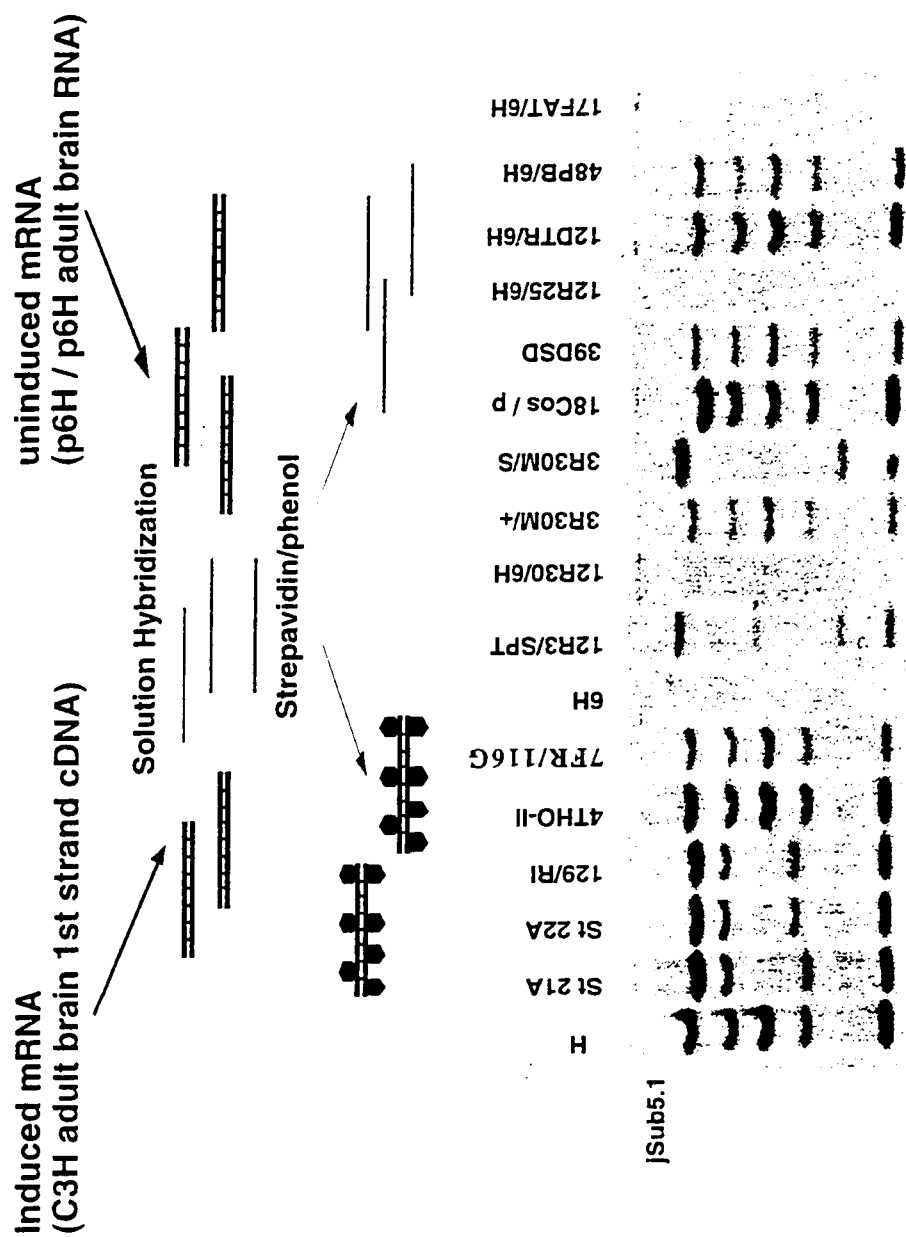
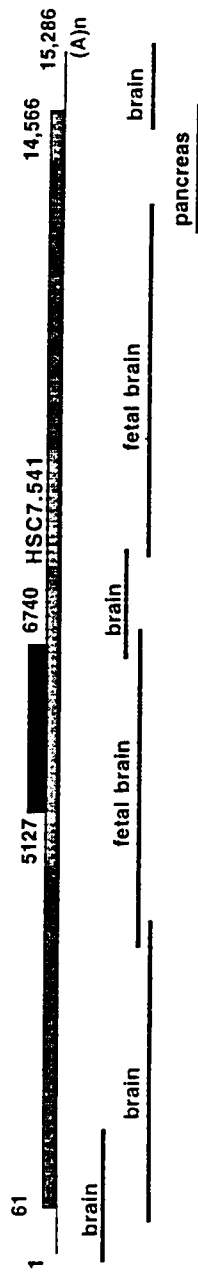


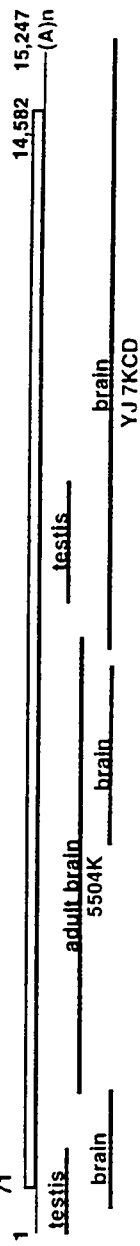
Figure 7. cDNA clones comprising mouse and human *ERY1* genes.

Schematic representation of cDNA clones used to assemble full length human (7a) and mouse (7b) *Ery1* cDNAs. Shaded areas represent ORFs with numbers above line denoting transcriptional start and stop sites. Both genes are transcribed in left 5', right 3' manner. Human cDNAs and sequencing of both full length genes is described in (Ji *et al.*, in preparation). Briefly, the 5' ends of mouse and human *Ery1* were generated by rapid amplification of cDNA ends (RACE) from testis RNA and brain RNA respectively(YJ). The mouse cDNA 5504K, was subcloned and sequenced at CWRU, the corresponding human cDNA was isolated by screening a human fetal brain cDNA library with a 138 bp genomic DNA probe with high homology to a human EST(05046). The mouse cDNA 556A-2 was isolated by screening with a mouse testis 3' RACE clone(YJ3'-2), 556A-2 represents *Ery1* nts ~6000-10,000 and is described further below. While this clone was being isolated the 3' end of mouse *Ery1* was isolated as a single 7.0 kb cDNA by Marathon RACE. This made possible the human *ERY1* RACE experiments which were used to assemble the 3' half of *ERY1* from various cDNAs and EST clones (Ji *et al.*, in prep). Panel 7c represents mouse *Ery1* cDNAs isolated and sequencing done at Oak Ridge, including sequencing of the mouse versions HSMN7 and HSC7.541 and a mouse embryonic cDNA mupc78.5 containing these sequences. Details of this analysis can be found in the final section of the results. Stars represent sequence matches identified between mupc78.5 and jSub5.1 to YJ1.1

7a. Human *ERY1*



7b. mouse *Ery1*



7c. 'OakRidge' mouse *Ery1* cDNAs

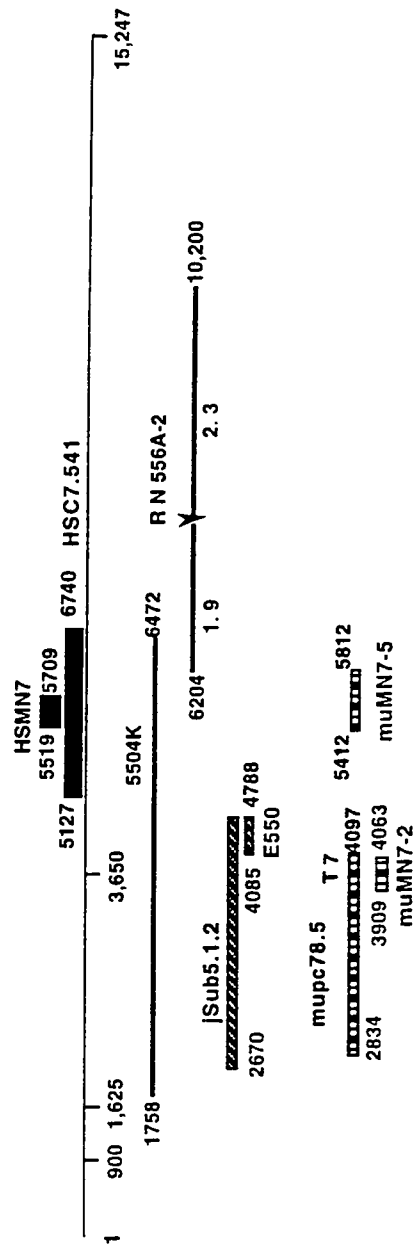


Figure 8. A composite map of *Ery1-p* interval

This figure includes the minimal number of P1, BAC and cosmid genomic clones and cDNAs with respect to various *jdf2* and *p* deletion complex alleles. The major features of this map are 1. *Ery1* in mouse is a large single locus that corresponds to *D7H15F37S1*, 2. *Ery1* is a 15 kb gene coded from a genomic interval of ~200 kb with 3' *Ery1* lying ~10 kb proximal of 5'*p*. 3. Dashed lines in *p*^{103G} and *p*^{39DSD} represent genomic RFLVs in these alleles which suggest the location of sequences required for proper expression of *Ery1* and *p* (resulting in intermediate *p*) in *p*^{103G} and for *Ery1* expression and mottling in *p*^{39DSD}. Open boxes at the end of the lines representing specific deletions signify RFLVs detected in that allele with a given probe. Several of these RFLVs are identified in the deletion mapping grid and its legend (Table II).

**Figure 9. Summary of genomic clones isolated with
D7H15F37S1-p probes.**

This figure serves to catalogue all genomic clones isolated in the course of this work. Note this figure is not drawn to scale. BAC clones were not completely analyzed, though it is known that 5' *Ery1* positive BACs do not overlap with the P1, N21306. Wavy lines at the ends of particular clones represents chimeric clones (e.g.Y12121).

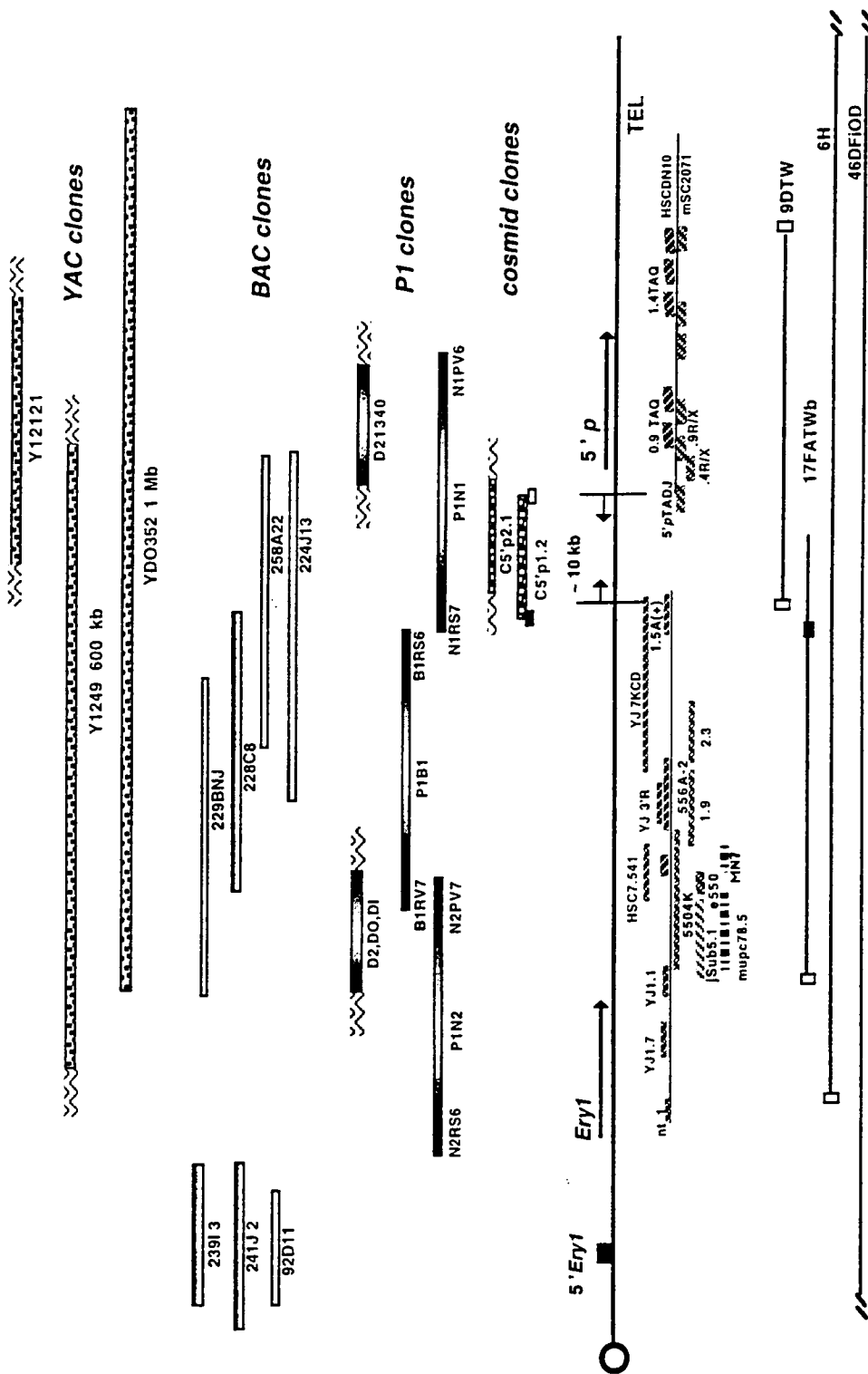
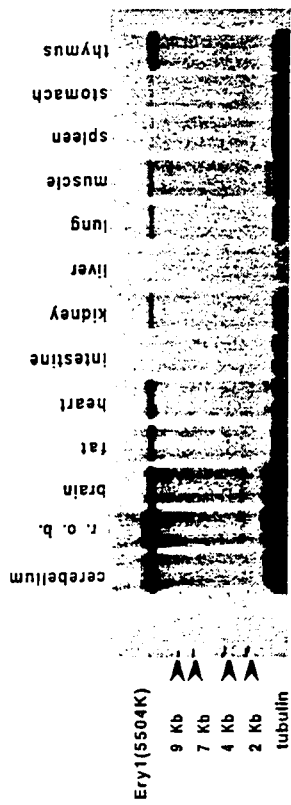


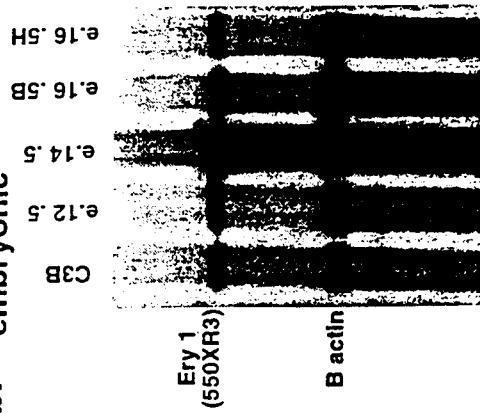
Figure 10. *Eryl* expression analysis in wild type and *jdf2* and *p* alleles.

10a is a Northern blot containing total RNA from a panel of normal (C3H) tissues, probed with the mouse *Eryl* cDNA 5504K followed by rehybridization with a probe for B-tubulin as a loading control. Molecular size standards are depicted by arrows at the left. **10b.** *Eryl* embryonic expression profile. The lower doublet marked with an arrow represents the signal for B actin which serves as an internal loading control. **10c.** Northern blot analysis of selected *jdf2* brain polyA(+) RNAs, probed with *Eryl* cDNA 5504K, followed by rehybridization with tubulin. Arrows in *jdf2*^{932SJ/p116G} lane indicated *p*^{116G}-specific putative fusion transcript and in *p*^{12DTR} mark the approximate position of hybridization with probe 1.5 A(+). **10d.** *Eryl* expression analysis by RPA in selected EtNU-induced *jdf2* alleles. As in 13a the lower doublet represents B-actin expression and serves as an internal loading control. This standard was used to quantify *Eryl* expression levels dosimetry and controls are presented in Fig.11 and its legend.

10 a. Adult



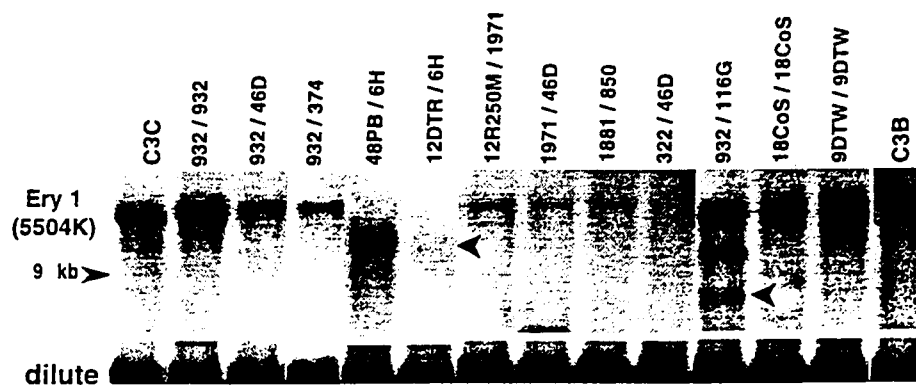
10 b. embryonic



10 c. ERY1 expression in W^v testis



10d. Northern



10e. RNAase Protection

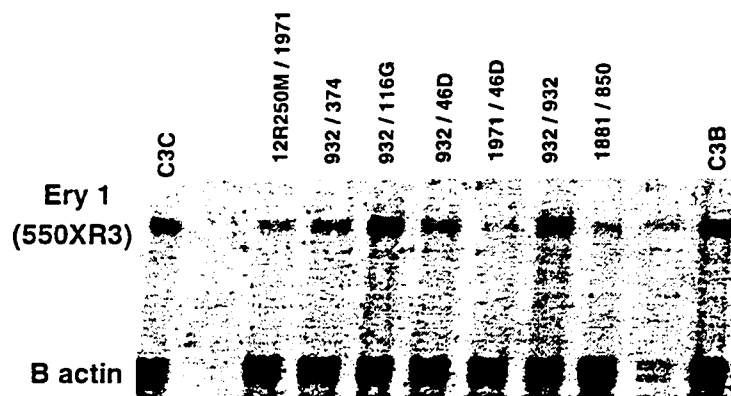


Figure 11. *Ery1* expression, RPA controls and dosimetry.

RPA experiments were performed essentially as per manufacturer's instructions. Controls as depicted at the left of the top panel include sham hybridizations to rule out artifactually protected fragments. In addition these experiments were designed such that the E550 probe was chosen because it is deleted in certain alleles and therefore serves as a negative control. The location of the probe E550 can be seen in Figures 7 and 8.

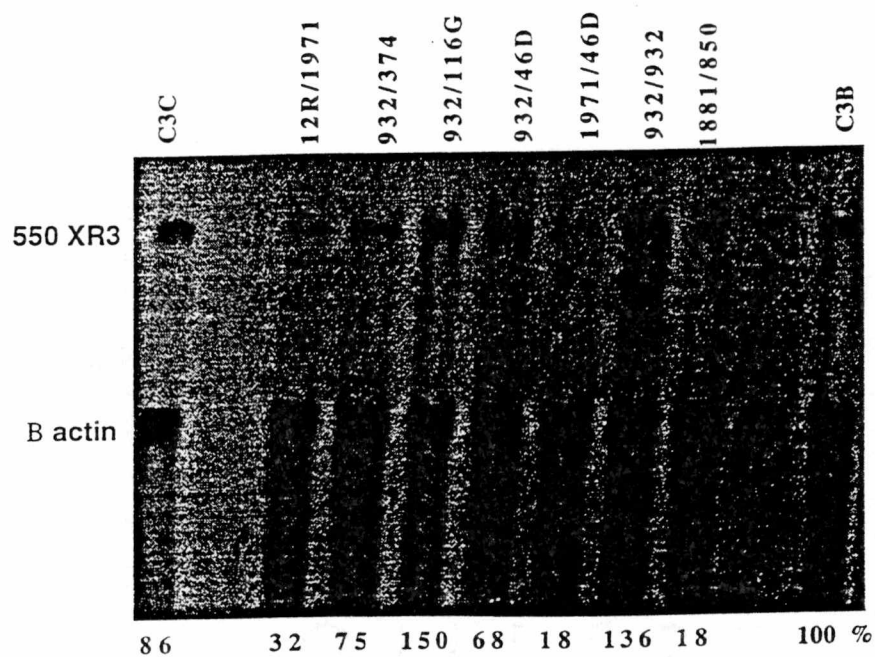
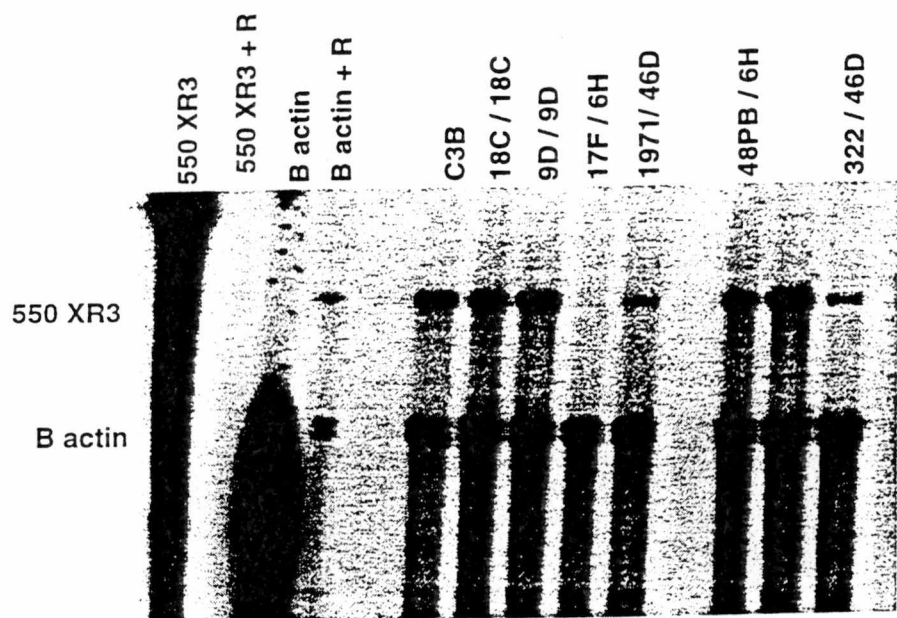


Figure 12. Identification of RFLVs specific for p^x and p^m alleles.

Genomic rearrangements in p^x and p^m alleles detected with *Eryl* cDNA probes. The top panel shows RFLVs detected with the *Eryl* probe 1.5A(+) in p^{103G} homozygotes. The lower panel shows RFLVs (arrows) detected with the probe RN556A-2, 2.3 (see map) in p^{39DSD} homozygous animals. This result has recently been extended in that this probe is polymorphic in p^{39DSD} chromosomes now with 3 of 5 restriction enzymes tested (R. Uphoff pers. comm.)

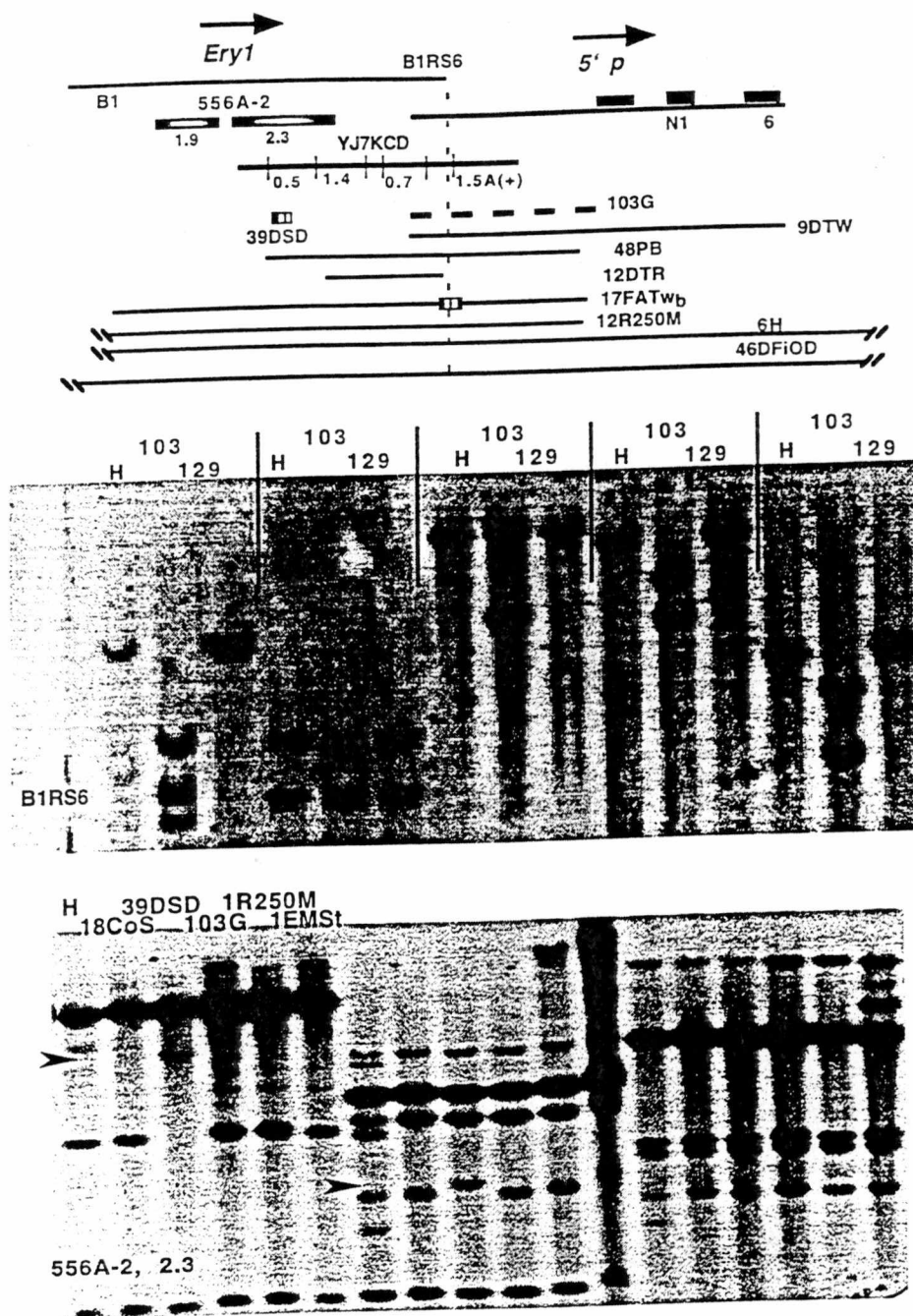
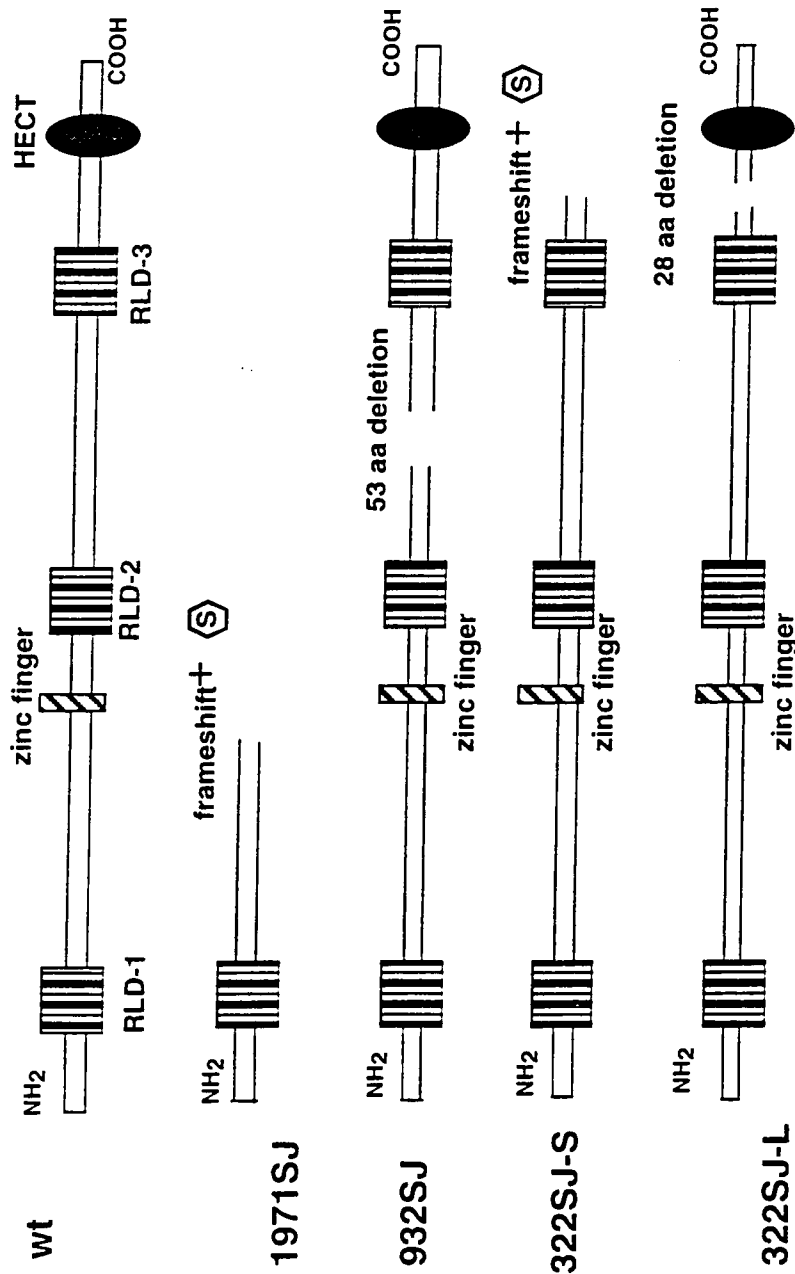


Figure 13. Identification of point mutations in several EtNU-induced *jdf2*^{SJ} alleles.

Sequencing of *Eryl* in several EtNU-induced *jdf2* alleles confirm single base pair changes as the primary defect with various consequences for the *Eryl* transcript. The top panel represents the full-length *Eryl* with various protein domains (RLD, ZNF, HECT). Mutant cDNA fragments were detected as size variants after RTPCR with primers spanning the 15 kb *Eryl* cDNA. Mutant cDNA fragments were sequenced in parallel with genomic controls (BJR/BJR). This analysis revealed *jdf2*^{1971SJ} is due to a small skipped exon which leads to a frameshift and a premature translational stop. An (A:G) transition in the consensus splice-donor sequence of the skipped exon was identified. A similar substitution was found to lead to an in frame skipped exon which removes 53 AA residues in the *jdf2*^{932SJ} allele. Finally RTPCR analysis of *jdf2*^{322SJ} revealed two size-altered fragments due to splice acceptor substitution and the use of a cryptic splice site which lead to an in-frame deletion of 28 aa and a 7 bp deletion respectively (from Ji *et al.*, in prep.)



(from Ji et al, in preparation)

Figure 14. Proteins with homology to *Eryl* .

This figure is a compilation of all database searches done by both of our groups. XIRP/RP3 is the X-linked retinitis pigmentosa gene 3. D25215 is an anonymous cDNA which has yet to be completely characterized (Nomura *et al.*, 1995). Detailed analysis of this homology and other relationships between *Eryl* homologies is presented in (Ji *et al.*, in preparation). The chicken EST shows 100% over a region of >100bp and is being further characterized by Y. Ji. The drosophila EST is also being characterized. The original match was 78% amino acid identity over 191 aa.

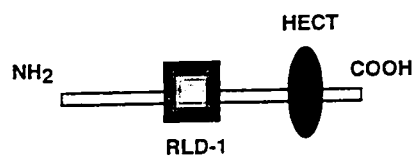
RCC 1



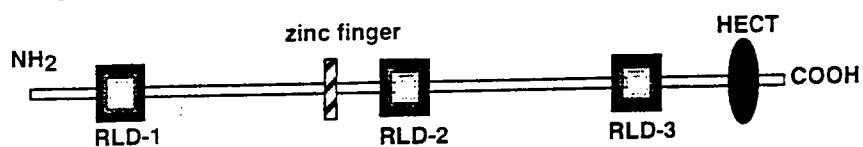
XIRP/RP3



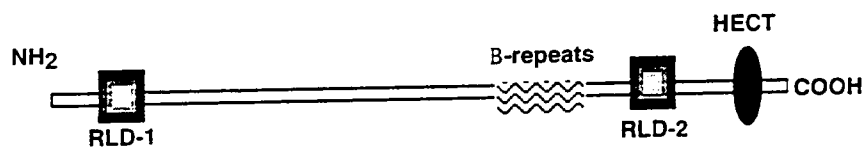
D25215



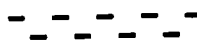
Ery1



p532



Chicken EST



Drosophila EST



Table I. *jdf2* alleles analyzed in this study.

Table I lists alleles analyzed in this study, it includes each allele's phenotype and origin.

TABLE I.

Mutation	Mutagen	<i>p</i> phenotype	<i>jdf2</i> phenotype	Reference
<i>p7FR60lb</i>	O / N	null	j l s	Russell et al '95
<i>p6H</i>	radiation induced	null	N.D.	Phillips, R.J.S. '77
<i>p132G</i>	PG / X	null	j l s	Russell et al '95
<i>p116G</i>	PG / X	null	j l s	Russell et al '95
<i>p12R250M</i>	S / N	<i>p x</i>	j l m	Russell et al '95
<i>p17FATW_b</i>	O / X	<i>p x</i>	<i>p n l</i>	Russell et al '95
<i>p48PB</i>	spontaneous	<i>p x</i>	<i>p n l</i>	Russell et al '95
<i>p12DTR</i>	S / X	<i>p x</i>	j l m	Russell et al '95
<i>p9DTW</i>	S / X	null	j l m	Russell et al '95
<i>p103G</i>	PG / X	<i>p x</i>	j l s	Russell et al '95
<i>p18CoS</i>	S / G	<i>p m</i>	j l m	Russell et al '95
<i>p39DSD</i>	S / X	<i>p m</i>	j l m	Russell et al '95
<i>jdf2932SJ</i>	EtNU	<i>p</i>	N.D.	Rinchik et al '95
<i>jdf2322SJ</i>	EtNU	<i>p</i>	N.D.	Rinchik et al '95
<i>jdf2850SJ</i>	EtNU	<i>p</i>	N.D.	Rinchik et al '95
<i>jdf21971SJ</i>	EtNU	<i>p</i>	N.D.	this study
<i>jdf2347SJ</i>	EtNU	<i>p</i>	N.D.	this study
<i>jdf21881SJ</i>	EtNU	<i>p</i>	N.D.	this study
<i>jdf2982SJ</i>	EtNU	<i>p</i>	N.D.	this study
X = X rays	O = Oocytes			<i>pnl</i> = prenatal lethal
N = neutrons	S = spermatogonia			<i>p^m</i> = mottled <i>p</i> expression
G = Gamma rays	PG = post/spermatogonial stages	EtNU = ethylnitroso urea		<i>p^x</i> intermediate <i>p</i> expression

Table II. Summary of *Eryl* expression data.

Concordance between defects in *Eryl* and the *jdf2* phenotype. Deletion alleles were not included in RPA analysis for obvious reasons (ND, not done). The first ten alleles are arranged in order of severity from most severe at top to more mild defect and phenotype (*p*^{103G}) (FL, full length; WT, wild type). Initial RFLVs reported here for have been confirmed (R. Uphoff pers comm). No confirmed RFLVs have been identified for *p*^{18CoS}. 'Deduced' refers to the assumptions made about the relative contribution of each allele of a compound heterozygote based on a 'known' value based on determining one alleles contibution as a hemizygote (compound heterozygote with a deletion). For example 347SJ levels are assumed to be quite low since the signals in 932/347 lane is roughly equivalent to that in the 932/46D lane. RTPCR-scanning and sequencing of *Eryl* cDNAs in several EtNU-induced *jdf2* alleles, relative to genomic control DNAs are reported elsewhere and are summarized here in Fig.16 with permission (Ji *et al*, in preparation). Cloning and sequencing of several deletion breakpoints and corresponding genomic intervals are discussed below and will be reported elsewhere (Y.Ji pers. comm).

Table II

	Genomic defect	Ery1 expression
<i>p7FR60lb</i>	deletion extending 3' from Ery1 bp ~5600	ND
<i>p132G</i>	deletion extending 3' from Ery1 bp ~5600	ND
<i>p116G</i>	deletion extending 3' from Ery1 bp ~5600	FL + chimeric message N
<i>p6H</i>	deletion extending 3' from Ery1 bp ~1400	ND
<i>p9DTW</i>	deletion extending 3' from Ery1 bp ~13,700	FL + smear
<i>p17FATW_b</i>	deletion extending 3' from Ery1 bp ~4000	null (RPA neg control)
<i>p12R250M</i>	deletion extending 3' from Ery1 bp ~1600	null by RPA
<i>p12DTR</i>	interstitial deletion ~10,200 - 13,700	< FL smear
<i>p48PB</i>	deletion extending 3' from Ery1 bp ~10,200	< FL smear
<i>p103G</i>	multiple rearrangements at Ery1 13,700	ND
<i>p18CoS</i>	no polymorphism detected	FL WT
<i>p39DSD</i>	genomic rearrangement at Ery1 ~7000	ND
* <i>jdf2932SJ</i>	159 bp in frame deletion	136 % WT
* <i>jdf2322SJ</i>	frame shift and deletion	"< 50%"
* <i>jdf2850SJ</i>	in progress	9 % deduced
* <i>jdf21971SJ</i>	exon skip, frame shift translation stop	20 % avg of two dels
* <i>jdf2347SJ</i>	in progress	< 50 % deduced
* <i>jdf21881SJ</i>	in progress	9 % deduced
* <i>jdf2982SJ</i>	in progress	ND

point mutant genomic analysis Y. Ji

*

Table III. A summary of deletion mapping data

Various *Eryl* and genomic endclone probes with respect to selected *jdf2* and *p* alleles. Representative mapping blots identify RFLVs in specific alleles for instance probe 1.5A(+) detects two fragments in *p*^{103G} homozygous genomic DNA. Also probe N1RS7 detects a polymorphic band in *p*^{9DTW}. (see Fig 5c). This grid incorporates all informative probes except those described in Fig. 9 and its legend. Abbreviations ND (not done); Par, partially deleted only; Poly, polymorphic; (+) wild type; triangles, deleted.

TABLE III.

PROBES

Mutations	N2RV6	5504K-1	RE1.7	RE1.1	E550	Sub5.1	mupc78.5	Hupc7.541	N2PV7	B111RV7	B111RS6	N1RS7	YE352X3	5' p	N1PV6	5562.3
p 46DFiod	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	ND	ND	Δ	Δ	Δ	Δ	Δ
p 4THO-II	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
p 25DVT	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	ND
p 12R30lb	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ
p 7FR60lb	Δ	Par	+	+	Δ	+	Δ	+	Δ	Δ	Δ	Δ	Δ	Δ	Δ	ND
p 116G	ND	Par	+	Par	+	Par	Par	+	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ
p 6H	+	Δ	Par	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ
p 17FATWb	+	Δ	+	Δ	Δ	Δ	Δ	Par	Δ	Δ	+	Δ	+	+	+	Δ
p 12DTR	+	+	+	+	+	+	+	+	Δ	+	+	Δ	+	+	+	★ Poly
p 12R250M	+	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	Δ	+	+	+	Δ
p 48PB	+	+	+	+	+	+	+	+	+	+	Δ	Δ	+	+	+	+
p 18Cos	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
p 9DTW	+	+	+	+	+	+	+	+	+	+	Δ	★ Poly	+	+	Δ	+
p 103G	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Appendix 2

Sequence alignment of the RLD, zinc-finger and HECT domains of human *ERY1* and related sequences.

(a). Alignment of the seven internal repeats of the RLD-1, RLD-2 and RLD-3 motifs of ERY1 with the corresponding repeats of human RCC1. (b). Alignment of the *hect* domain of human ERY1 with the *hect* domain of UBE3A (E6AP) and p619 proteins. Identical amino acids are boxed, while similar amino acids are shaded. The conserved cysteine residue involved in ubiquitin transfer is indicated by an arrow. (c). Alignment of the zinc-finger motif of human ERY1 with homologous zinc-finger motifs found in Genbank and SwissProt entries. The six conserved cysteine residues are indicated by asterisks (*).
(taken from Ji *et al.* in prep, and presented here for completeness.)

VITA

Mitchell John Walkowicz is the son of Mary S and Mitchell J Walkowicz. Born May 1, 1963 in Webster, Massachusetts. Educated at St. Joseph School and Shepherd Hill Regional High school, he received a Bachelor of Arts degree at Colby college in Waterville, Maine in 1985. Upon completion of a Doctor of Philosophy degree at the University of Tennessee Graduate school of Biomedical Sciences at Oak Ridge, he will begin a postdoctoral position in developmental biology at the National Institutes for Health, in Bethesda Maryland.