Identification and characterisation of genes from the Angelman and Prader-Willi syndrome region of chromosome 15q11-13.

Megan P. Hitchins

Institute of Child Health,
University College London,
30 Guilford Street,
London WC1N 1EH

Submitted for the degree of Doctor of Philosophy to the University of London, September 1999.

Abstract

Angelman syndrome (AS) and Prader-Willi syndrome (PWS) are clinically distinct neurobehavioural disorders which map to 15q11.2-12. AS is characterised by mental retardation, seizures, ataxia, absent speech, excessive laughter, and distinctive EEG. PWS involves poor suck reflex and hypotonia in infancy, short stature, microgenitalia, and hyperphagia leading to obesity.

Chromosome 15q11.2-12 is subject to genomic imprinting whereby certain genes are functionally active or silent depending on their parental origin of inheritance. AS and PWS are caused by various mechanisms including cytogenetic deletion, uniparental disomy and altered epigenetic modification, with the common factor being lack of a maternal contribution to 15q11.2-12 in AS, and loss of expression of genes on the paternally derived allele in PWS. Approximately 25% of AS patients, including familial cases, are due to mutations of an imprinted gene within 15q11.2-12, which is expressed from the maternal allele. The *UBE3A* gene has been implicated in AS. The lack of a similar category of patients in PWS suggests that this is a contiguous gene disorder.

In order to identify genes which may be involved in the pathogenesis of either AS or PWS, positional cloning studies were performed in the AS and distal portion of the PWS candidate regions. *UBE3A* was located to the AS region, and a novel expressed sequence identified through direct cDNA selection was shown to represent an extended 3' untranslated region of *UBE3A*. No additional genes were identified in the AS region. *UBE3A* was screened for mutations in cases of familial and sporadic AS in which the genetic aetiology of the disease remained unidentified. Pathogenic mutations were identified in 3/5 familial and 4/21 sporadic cases, indicating that *UBE3A* mutations cause AS in a proportion of patients. Two novel transcripts, '395-H22' and '123-E19', were identified from the PWS region. 395-H22 was found to be expressed solely from the paternal allele in an *in vitro* imprinting assay system, indicating that it may represent an additional gene involved in the PWS phenotype.

Acknowledgements

Firstly I'd like to thank my supervisor Prof. Sue Malcolm for her support and guidance, and for sticking with me until the very end! My thanks to each member of the 'Angelman Group', Dr. Jess Buxton, Dr. Jo Chan, Dr. Helen Gilbert, Dr. Una Fairbrother and Sarah Rickard, for all their support and camaraderie in the lab during the course of this project. Special thanks also to Dr. Charlotte Rose for her technical help in the lab and her friendship. I'd like to thank Prof. Marcus Pembrey, Prof. Bryan Winchester and Dr. Dave Muller for their support, good word and helpful advice. Special thanks go to Dr. Gudrun Moore for her collaboration, additional guidance in the writing of this thesis, and for allowing me time to complete it. Thanks also to Prof. Mike Preece for his encouragement.

This project was funded by a Medical Research Council studentship for 3 years, and by the Birth Defects Foundation for a 4 month extension. I'd also like to acknowledge the Angelman Support Group for providing travel expenses.

On a personal note, special thanks go to my mother and grandmother, who have always backed me, and have supported me throughout this endeavour. I'd also like to take this opportunity of thanking my father who paid for my undergraduate university fees – I'd never have got this far otherwise. I'd also like to pay my respects to my late grandfather who always maintained that education was never wasted on women! Finally, I'd like to thank my close friends who have provided inspiration, consolation, practical help and lots of fun; Steve, Michelle, Katie, Vivian, Kerry, Alexia, Kev and Harriet, Mark and Tilly, Des, Greta, Edgar and Sean.

Contents

List of Figures

List of Tables

Abbreviations

A	Deoxyadenosine triphosphate
AS	Angelman syndrome
ASCR	Angelman syndrome candidate region
ATP	Adenosine triphosphate
BAC	Bacterial artificial chromosome
BLAST	Basic local alignment search tool
BSA	Bovine serum albumin
BWS	Beckwith-Wiedemann syndrome
cDNA	Complementary DNA
cM	Centimorgan
C	Deoxycytosine triphosphate
CEPH	Centre d'Etude Polymorphisme Humaine
CF	Cystic fibrosis
CPM	Confined placental mosaicism
CVS	Chorionic villus sampling
dATP	Deoxyadenosine triphosphate
dbEST	EST database
dCTP	Deoxycytosine triphosphate
dGTP	Deoxyguanosine triphosphate
dNTP	Deoxynucleotide triphosphate
dTTP	Deoxythymidine triphoshate
ddATP	Dideoxyadenosine triphosphate
ddCTP	Dideoxycytosine triphosphate
ddGTP	Dideoxyguanosine triphosphate
ddNTP	Dideoxynucleotide triphosphate
ddTTP	Dideoxythymidine triphosphate
DNA	Deoxyribonucleic acid
EDTA	Ethyline diamidetetracetic acid
EEG	Electroencephalogram
EMBL	European Molecular Biology Laboratory
EST	Expressed sequence tag
Fig	Figure
FISH	Fluorescent *in situ* hybridisation
g	Gram
G	Deoxyguanosine triphopshate
HGMP	Human Genome Mapping Project
IC	Imprinting center
I.M.A.G.E.	Integrated molecular analysis of genomes and their expression
IUGR	Intra-uterine growth retardation
kb	Kilobase
kDa	Kilodalton
λ	Lambda phage
LINE	Long interspersed element
LOD	Logarithmic of the odds
mg	Milligram
ml	Millilitre
μg	Microgram
μl	Microlitre
mRNA	Messenger ribonucleic acid

mUPD	Maternal uniparental disomy
M	Molar/moles
MatDp	Maternal duplication/disomy
Mb	Megabase
MCS	Multiple cloning site
Meg	Maternally expressed gene
MI	Meiosis I
MII	Meiosis II
NDUI	Non deletion, uniparental disomy, imprintor mutation
NIX	Nucleotide identify X
OD	Optical density
ORF	Open reading frame
pfu	Plaque forming unit
pmol	picomole
pUPD	Paternal uniparental disomy
PAC	P1 artificial chromosome
PAGE	Polyacrylamide gel electrophoresis
PatDp	Paternal duplication/disomy
PCR	Polymerase chain reaction
Peg	Paternally expressed gene
PEG	Polyethylene glycol
PFGE	Pulsed field gel electrophoresis
PIC	Polymorphic information content
PWS	Prader-Willi syndrome
PWSCR	Prader-Willi syndrome candidate region
rpm	Revolutions per minute
RFLP	Restriction fragment length polymorphism
RFLV	Restriction fragment length variant
RNA	Ribonucleic acid
RT	Reverse transcription/ reverse transcriptase/ reverse transcribed
SDS	Sodium dodecyl sulphate
SINE	Short interspersed element
SRS	Silver-Russell syndrome
SSC	Sodium chloride/ sodium citrate
SSCP	Single-stranded conformational polymorphism
STR	Short tandme repeats
STS	Sequence tagged site
tRNA	Total ribonucleic acid
T	Deoxythymidine triphosphate
TE	Tris EDTA
Tn	Transposon
U	Units of enzyme
UPD	Uniparental disomy
UTR	Untranslated region
VNTR	Variable number tandem repeat
WI	Whitehead Institute
YAC	Yeast artificial chromosome
ZNF	zone of no resolution

Chapter 1: Introduction

1.1 Clinical description of Angelman syndrome

Angelman syndrome (AS), previously referred to as 'Happy Puppet syndrome', was first described by a general paediatrician in three unrelated children. Each displayed mental retardation, jerky movements, ataxia with weakness of the limbs and trunk, frequent epileptic attacks, paroxysms of laughter, absence of speech, protruding tongues, abnormal electroencephalograms (EEGs), brachycephaly associated with microcephaly, and a varying degree of optic atrophy (Angelman, 1965). Table 1.1 lists the clinical features and developmental history of classical AS, and serves as a consensus for diagnostic criteria for AS. Genetic testing is confirmatory in about 80% of cases (Williams *et al.*, 1995). This data was based on the clinical and genetic findings collected for a large number of AS cases of varying age-groups (Zori *et al.*, 1992; Clayton-Smith, 1993; Chan *et al.*, 1993). A girl displaying some of the typical features of AS is shown in Fig 1.1.

Frequency	Clinical Characteristics
Consistent 100% cases	Developmental delay, evident by 6-12 months of age (delayed motor milestones, including sitting, crawling, walking). Absent or minimal speech (maximum of 6 words). Ataxia of gait and / or tremulous limb movements. Neurobehavioural uniqueness including frequent bouts of laughter, happy disposition, hand flapping, hyperactivity.
Frequent > 80% cases	Microcephaly by age 2. Seizures (usually by age 3). Abnormal, characteristic EEG.
Associated 20-80% cases	Protruding tongue, tongue thrusting, excessive drooling. Feeding problems in infancy (poor suck and swallowing). Distinctive craniofacial dysmorphisms: prognathia, wide mouth, widely spaced teeth, strabismus, flat occiput, occipital groove. Hypopimentation of skin, hair and eyes compared to family members. Brisk deep tendon reflexes, uplifting of arms during ambulation. Sleep disturbance. Propensity for water.

Table 1.1. Clinical features of Angelman syndrome, and consensus for diagnostic criteria
The clinical characteristics and the frequency at which these are observed are given. The inclusion criteria for a clinical diagnosis of AS is that the patient has all the clinical features listed in the top two rows. These cases are referred for genetic testing.

Clinical diagnosis of AS in infancy is difficult as prominent characteristics which aid in the differential diagnosis, including outbursts of laughter, absent speech, facial dysmorphisms, onset of seizures, and mental retardation, do not develop or are not recognisable until 1-4 years of age (Yamada and Volpe, 1990; Fryburg *et al.*, 1991; Clayton-

Smith, 1993). The characteristic EEG pattern is invaluable for an early diagnosis of AS (Boyd *et al.*, 1988). AS patients appear to have a normal life span, though only a small number over 30 years of age have been reported as adult patients may have been institutionalised due to mental retardation of unknown cause, prior to the wide recognition of AS (Reish and King, 1995). Clinical diagnosis of older patients may also be difficult since epileptic activity, hyperactivity and frequent laughter may be markedly reduced. Obesity and a worsening of scoliosis were observed with increased age (Clayton-Smith *et al.*, 1993; Buntix *et al.*, 1995). Whilst originally considered rare, the establishment of precise clinical diagnostic criteria and genetic testing has led to an increased rate of detection of AS. The incidence of AS is estimated to be 1 in 20,000. AS may be familial with more than one affected sibling, but most cases are sporadic (Clayton-Smith *et al.*, 1992a; Clayton-Smith and Pembrey, 1992).

Fig 1.1 A girl with classical Angelman syndrome

Photograph of AS patient 3792 at 5 years of age, displaying the typical facial features. This child was also hypopigmented. The behavioural characteristics are not captured in this picture. This photograph reproduced with the kind permission of the probands' mother.

1.2 Clinical description of Prader-Willi syndrome

Prader-Willi syndrome (PWS) was first described in nine patients presenting with severe hypotonia and feeding problems related to poor suck in infancy, with reports of reduced fetal activity during gestation. Hypogonadism, hypogenitalism, short stature, small hands and feet, distinctive facies, delayed psychomoter development, mental retardation, and behavioural problems were also observed. Hyperphagia early in childhood led to obesity, and diabetes mellitus developed in adolescence (Prader *et al.*, 1956). PWS can be variable, although infant hypotonia, poor suck, hyperphagia and obesity are cardinal features. The full clinical spectrum and diagnostic criteria for PWS (Holm *et al.*, 1993), are given in Table 1.2

Category	Diagnostic criteria
Major 1 point	Neonatal and infantile hypotonia and poor suck; improvement with age. Feeding problems in infancy; need for special feeding techniques and failure to thrive. Excessive weight gain after 1 year, but before 6 years of age. Characteristic facial features; dolichocephaly in infancy, narrow bifrontal diameter, almond-shaped eyes, small mouth with thin upper lip, down-turned corners of mouth (3 required). Hypogonadism; a) genital hypoplasia, b) delayed gonadal maturation and puberty in the absence of hormonal treatment. Developmental delay below the age of 6 years (motor, cognitive and language skills) Mental retardation (intelligence quotient 20-90, with an average of 65). Hyperphagia and obsession with food, foraging. Positive genetic test involving disruption of 15q11-13.
Minor 0.5 point	Decreased intrauterine fetal growth or weak cry or lethargy in infancy. Behavioural problems; violent outbursts, tantrums, obsessive/compulsive behaviour, tendency to be stubborn, argumentative, possessive, steal, lie (at least 5 required). Sleep disturbance or sleep apnea. Speech articulation defects. Short stature compared to family members (without growth hormone). Small hands and feet for height. Narrow hands with straight ulna border. Eye abnormalities (myopia, esotropia). Thick viscous saliva with crusting at the corners of the mouth. Hypopigmentation of skin, hair and eyes compared to family members. Skin picking.
Supportive 0 points	High pain threshold, decreased vomiting, temperature instability in infants or altered temperature sensitivity in older patients, scoliosis, osteoporosis, early adrenarche, strong visual-perceptual-motor skills, normal neuromuscular studies.

Table 1.2 Clinical features and diagnostic criteria for Prader-Willi syndrome.

Clinical characteristics are listed in categories of *major, minor* and *supportive*, according to their prevalence in the patient population, and form the basis for diagnosis on a point-scoring system. Patients under three years of age need to score 5 points, 4 derived from the *major* group, for a positive diagnosis. Older children and adults, require 8 points, with at least 5 falling within the *major* category, for a positive diagnosis of PWS.

PWS is difficult to diagnose in infancy as certain distinctive features are not apparent until later in childhood (Butler *et al.*, 1983). PWS is the most common syndromal cause of obesity, and this is probably the most distinctive feature of the disorder (Butler *et al.*, 1986; Butler, 1990). However, excessive weight gain is not apparent until 2-6 years of age. An early diagnosis of PWS is needed if dietary intervention is to occur to prevent gross obesity and resultant morbid sequelae including diabetes and heart conditions (Robinson *et al.*, 1991; Holm *et al.*, 1993). A child with the characteristic features of PWS is shown in Fig 1.2.

The incidence of PWS has been estimated to be 1/15,000 (reviewed by Nicholls, 1993). The vast majority of cases are sporadic, with a limited number of reports describing familial cases, in which more than one sibling is affected (Örstavik *et al.*, 1992; Reis *et al.*, 1994; Buiting *et al.*, 1995).

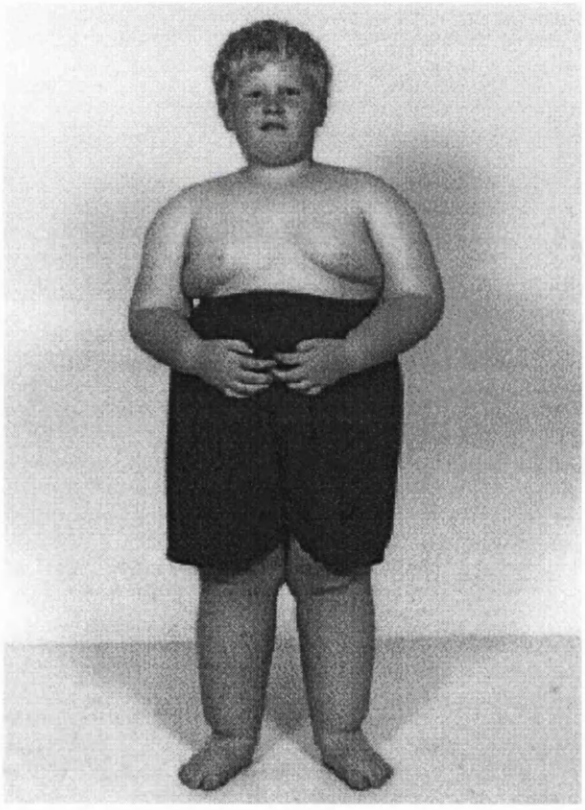


Fig 1.2 A boy with typical features of Prader-Willi syndrome.

Obesity, small hands and feet, and certain characteristic facial features, including almond-shaped eyes and down-turned corners of the mouth, are displayed by this child.

1.3 Genomic Imprinting

Genomic imprinting is defined as the differential expression of a gene or chromosomal region according to the parental origin of inheritance. This is an epigenetic phenomenon which occurs in the normal state, whereby both the maternal and paternal alleles are present, but function unequally in the soma. Genomic imprinting is reversible; inherited maternal and paternal imprints are erased during gametogenesis (Szabo and Mann, 1995), and new imprints established according to the sex of the individual (reviewed by Hall, 1990; Langlois, 1994; Tilghman, 1999). There are several lines of evidence for genomic imprinting based on experiments in mice and observations in human genetic disease.

1.3.1 Imprinting of the maternal and paternal genomes

1.3.1.1 Gynogenetic, parthenogenetic and androgenetic embryos

Diploid mouse embryos have been constructed with the nuclear material derived solely from the maternal (gynogenetic and parthenogenetic) or paternal (androgenetic) genomes. Parthenogenotes derived from a single female parent were produced either by endoreduplication of the haploid nucleus or retention of the second polar body to restore diploidy (reviewed by Solter, 1988). Diparental gynogenotes and androgenotes were created by pronuclear transplantation, whereby one of the pronuclei is removed from a zygote and replaced with a second pronucleus of the opposite parental origin (McGrath and Solter, 1984). Both sets of embryos developed to the blastocyst stage, but failed post-implantation. Parthenogenetic and gynogenetic embryos developed relatively normally to the 25 somite stage, but development of the extraembryonic components was poor (McGrath and Solter, 1984; Mann and Lovell-Badge, 1984; Surani *et al.*, 1984). The converse was observed in the androgenones, with considerably retarded embryonic development, but substantial growth of the trophoblast and yolk sac (McGrath and Solter, 1984; Barton *et al.*, 1984). These experiments demonstrated that both the maternal and paternal genomes are required for normal embryonal development; diploidy is not sufficient. Furthermore they showed that the maternal genome appears to be biased towards fetal development, whereas the paternal genome contributes more towards development of the extraembryonic structures (McGrath and Solter, 1984).

This is also evinced by the reciprocal phenotypes observed in ovarian teratomas and complete hydatidiform moles in humans, which are naturally occurring parthenogenones and androgenones, respectively. Ovarian teratomas are comprised of a disorganised mass of differentiated embryonic tissues, but do not contain any placental tissue (Linder *et al.*, 1975). Complete hydatidiform moles are characterised by extensive growth of the trophoblast in the absence of an embryo (Kajii and Ohama, 1977).

1.3.1.2 Human triploidy

Human diandric triploid fetuses, which have two paternal and one maternal complement, are malformed and have a large cystic placenta with partial molar changes. Digynic triploid fetuses, which have an extra maternal chromosome set, are considerably growth retarded with underdeveloped non-molar placenta. These observations show the differential effect of the maternal and paternal genomes; that the maternal complement contributes more to early fetal development, whereas the paternal genomes is critical for development of the placenta and membranes (McFadden *et al.*, 1993).

1.3.2 Uniparental disomy

Uniparental parental disomy (UPD) is the inheritance of both homologues of a particular chromosome from a single parent. UPD was first suggested as a theoretical mechanism for human genetic disease (Engel, 1980), on the basis that trisomy for several autosomes observed in the products of spontaneous abortions, chiefly stems from gametal disomy (Hassold and Matsuyama, 1979). Engel hypothesised that fusion of a disomic gamete with one nullisomic for the same chromosome (gamete complementation) could form a viable conceptus corrected for chromosomes number, and that this could lead to genetic disease due to reduction to homozygosity of mutant alleles.

1.3.2.1 Mechanisms leading to UPD

Various mechanisms leading to UPD have been proposed, which focus on the reciprocal products of non-disjunction events that occur at meiosis. Complete isodisomy, where the two copies of a chromosome homologue are identical, occurs when a normal monosomic gamete fuses with a nullisomic gamete. The homologue present in the zygote duplicates mitotically to form an exact replica, which demonstrates homozygosity for the

full-length of the chromosome (Spence *et al.*, 1988; Robinson *et al.*, 1993a). This is mechanism is referred to here as 'monosomy rescue'. Heterodisomy is the inheritance of two different homologues of a pair from one parent. This arises when a trisomic conceptus, resulting from fusion of a monosomic gamete with a disomic one, is corrected by post-zygotic loss of one of the extra homologues (Cassidy *et al.*, 1992). This is referred to here as 'trisomy rescue'. Heterodisomy or isodisomy is not always complete. Heterodisomy near the centromere is indicative of a meiosis I (MI) non-disjunction event resulting in co-segregation of the two chromosome homologues into meiosis II (MII). Telomeric markers may be homozygous or heterozygous depending on prior recombinational activity. Reduction to homozygosity (isodisomy) at the centromere accompanied by heterozygosity towards the telomere, is indicative of a MII error, in which the sister chromatids have failed to separate. Heterozygous telomeric loci reflect any recombination events that occurred during MI (Robinson *et al.*, 1993a).

1.3.2.2 UPD in mice

Mice with UPD for whole chromosomes have been bred from mice with Robertsonian translocations, which demonstrate high rates of meiotic non-disjunction. Similarly mice with segmental UPD have been created by crossing mice heterozygous for reciprocal translocations (reviewed by Cattanach, 1986). Mice with whole or partial UPD for chromosomes 2, 6, 7, 11, 12 and 19 have demonstrated imprinting effects, whereby disomy from one parent fails to complement a corresponding nullisomy from the other parent. For instance, early embryonic lethality has been associated with maternal duplication of proximal chromosomes 2 and 6 (reviewed by Cattanach and Beechey, 1990). Opposite phenotypic effects have also been observed for reciprocal disomies of an imprinted region. For example, paternal deficiency/maternal duplication for proximal chromosome 11 resulted in reduced size at birth (70%), whereas maternal deficiency/paternal disomy for the same region was associated with increased birth-weight (130%) (Cattach and Kirk, 1985). These studies provided evidence for genomic imprinting at the chromosomal and sub-chromosomal level. Through the examination of individual segments of the mouse autosomes a map of imprinted regions has been produced (Cattanach, 1986; Cattanach and Beechey, 1990; Cattanach *et al.*, 1997). Refinement of these regions to identify candidate imprinted genes and mouse models for human imprinted disorders is ongoing. The mouse imprinting map is available at the World-Wide Web site: **www.har.mrc.ac.uk/imprinting/maps/**

1.3.2.3 UPD and genetic disease in humans

The first observation of UPD in humans was maternal UPD (mUPD) for chromosome 7 in two individuals with cystic fibrosis (CF) and growth retardation. Both patients were isodisomic with consequent homozygosity of a mutant maternal allele *CFTR*. Genomic imprinting was suggested as the aetiological basis for the observed growth retardation (Spence *et al.*, 1988; Voss *et al.*, 1989). The disease phenotypes associated with UPD in humans may be due to three distinct factors (Ledbetter and Engel, 1995):

(1) ***Reduction to homozygosity of a recessive allele inherited from one parent.***

In addition to CF, this has been observed for several autosomal recessive diseases associated with isodisomy, including congenital chloride loss diarrhoea in paternal UPD (pUPD) for chromosome 7 (Hoglund *et al.*, 1994), and spinal muscular atrophy due to pUPD5 (Brzustowicz *et al.*, 1994).

(2) ***The effects of trisomy on the placenta in cases of 'trisomy rescue'.***

Intrauterine growth retardation (IUGR) in cases of mUPD16 have been shown to be related to the level of trisomy mosaicism in the placenta as opposed to UPD, since cases with biparental inheritance were also growth restricted (Kalousek and Barrett, 1994). However, additional features observed in a proportion of neonates, including hypospadius and anal atresia, may be attributable to disomy (Vaughan *et al.*, 1994).

(3) ***Imprinting effects.***

Genetic disease due to imprinting effects associated with UPD in humans has been confirmed for chromosomes 7, 11, 14 and 15. Most of these regions share homology with regions demonstrating imprinting effects in the mouse (reviewed by Ledbetter and Engel, 1995). Phenotypes due to imprinting effects can be distinguished from uniparental inheritance of a recessive allele by the demonstration of heterodisomy for the full length of the chromosome involved. Maternal heterodisomy for chromosome 15 associated with PWS was the first report of genetic disease due to heterodisomy, and provided strong evidence for imprinting on this chromosome (Nicholls *et al.*, 1989a)(section 1.4.3). mUPD7 accounts for approximately 10% of patients with Silver-Russell syndrome (SRS), characterised primarily by intrauterine and post-natal growth retardation and lateral asymmetry (Kotzot *et al.*, 1995). By analysing a group of mUPD7 patients with mixed regions of isodisomy and heterodisomy, due to prior recombinational activity, it was shown that no particular region of chromosome 7 was consistently isodisomic. This ruled out reduction to homozygosity of a recessive locus as the mechanism causing SRS (Preece *et al.*, 1999). Beckwith-Wiedemann

syndrome (BWS) is a congenital overgrowth disorder often accompanied by Wilm's tumour of the kidney, which is due to mosaic segmental pUPD for 11p15.5 in 20-25% of cases. Although isodisomy has been demonstrated in each case, there is additional evidence for genomic imprinting in this region (section 1.7.6). UPD in this situation arises due to a post-zygotic event (Henry *et al.*, 1993; Slatter *et al.*, 1994).

1.3.3 Mouse transgenes provide evidence for genomic imprinting at the molecular level

Global differences in methylation of endogenous sequences between the male and female gametes in mice had suggested a role for DNA methylation in genomic imprinting (Sanford *et al.*, 1987). Transgenic experiments were conducted to randomly probe the mouse genome for differentially methylated regions, by comparing the methylation patterns of the transgenes after transmission through the male and female germ lines. Transgenic strains were identified in which the methylation state of the transgene was not only dependent on the sex of the transmitting parent, but was switched as the transgene was passed alternately between the sexes in successive generations. The differential methylation of particular transgenes according to parental origin was attributed to the surrounding region of endogenous DNA (although other factors have since been shown to affect the methylation status of transgenes). These experiments showed that the methylation imprint was germ-line specific, persisted through somatic cell division, and was reversible, providing evidence that DNA methylation was the molecular mechanism underlying genomic imprinting (Reik *et al.*, 1987; Sapienza *et al.*, 1987; Swain *et al.*, 1987; Chaillet *et al.*, 1991a).

1.3.4 The paradigm of genomic imprinting in humans

AS and PWS have become the paradigm of genomic imprinting in humans. Not only did these two disorders provide the first definitive example of imprinting in humans, but similar genetic defects involving the same chromosomal location resulted in two distinct disorders. The factor determining the phenotypic outcome was the parental origin of the chromosome that the genetic defect arose on. Through the analysis of the genetic aetiology of AS and PWS and the region involved, much has been learned about genomic imprinting and its processes.

1.4 Genetic mechanisms leading to PWS and AS

1.4.1 Cytogenetic deletions of 15q11-13 in PWS and AS

The first evidence that aberrations of chromosome 15 were associated with PWS came from observations of Robertsonian 15;15 translocations in the karyotype of PWS patients (Hawkey and Smithies, 1976; Fraccaro *et al.*, 1977). Cytogenetically visible deletions of 15q11-13 were first described in PWS patients, with the implementation of high-resolution prophase analysis (Ledbetter *et al.*, 1981). Several cases involving deletions were subsequently reported using G, C and R-banding of prophase and early metaphase chromosomes (Butler *et al.*, 1983; Cassidy *et al.*, 1984).

The causal involvement of chromosome 15 in AS was first reported with the observation of a cytogenetic deletion of 15q11.2 in a patient by Giemsa-trypsin staining of prophase chromosomes (Kaplan *et al.*, 1987). In subsequent reports confirming deletions in AS, attention was drawn to the similarity to the deletions observed in PWS (Magenis *et al.*, 1987; Williams *et al.*, 1989; Pembrey *et al.*, 1989). Direct comparison of the deletions in AS and PWS patients confirmed that both groups were deleted for the same band, 15q11.2 (Magenis *et al.*, 1990).

Chromosome 15 is acrocentric, with heteromorphisms involving variation in short arm length, satellite stalk, and silver-staining activity of the nucleolar organising regions (Wachtler and Musil, 1980). These heteromorphisms were used to determine the parental origin of deletions in PWS and AS patients. In PWS the deletions always arose on the paternal homologue (Butler and Palmer, 1983; Butler *et al.*, 1986), whereas in AS, the deletions were consistently maternal in origin (Williams *et al.*, 1990; Magenis *et al.*, 1990).

1.4.2 Molecular Analysis of deletions

1.4.2.1 Isolation of molecular probes for 15q11-13

Recombinant DNA probes unique to 15q11-13 were identified from two chromosome 15 specific genomic libraries constructed from (i) a supernumerary inverted duplicated 15 (pter→q13::q13→pter), isolated by flow-cytometry (Donlon *et al.*, 1986), and (ii) the microdissected 15q11.2-13 band from a Giemsa-trypsin G-banded metaphase chromosome

(Buiting *et al.*, 1990). These were mapped to the PWS deletion region by quantitative dosage blotting using the DNA from a PWS patient with a cytogenetically defined deletion.

A set of eight non-redundant probes, ML34, TD3-21, IR4-3, IR10-1, 189-1, MN7, PW71 and MN6, identifying loci D15S9-D15S13, D15S62, D15S63 and D15S176 respectively, were originally located within the PWS deletion region (Donlon *et al.*, 1986; Buiting *et al.*, 1990). Some of these were ordered with respect to one another by dosage analysis using DNA from a diverse group of patients with a variety of chromosome 15 aneusomies. However, some of the probes were consistently deleted in PWS patients so it was not possible to order these with respect to one another (Donlon, 1988; Tantravahi *et al.*, 1989) (Fig 1.3).

1.4.2.2 Molecular deletions in PWS and AS

Identification of probes unique to 15q11-13 made it possible to detect or confirm deletions by dosage analysis. Restriction fragment length polymorphisms (RFLPs) were subsequently described for these probes enabling deletions to be detected qualitatively, and allele segregation within families to be determined (Donlon, 1988; Nicholls *et al.*, 1989b). Since cytogenetic deletions in AS and PWS patients appeared not to differ, a molecular comparison was made of the extent of the deletions between the two sets of patients. Both AS and PWS patients were deleted for loci D15S9 to D15S13 (Fig 1.3), demonstrating that the molecular deletion was common to both groups of patients (Donlon, 1988; Knoll *et al.*, 1990). These deletions were estimated to be 3.5-4Mb (Kuwano *et al.*, 1992). RFLP analyses of PWS and AS families also confirmed that the deletions involved in PWS were consistently on the paternal 15, whereas the AS deletions were always maternally inherited (Knoll *et al.*, 1989; Hamabe *et al.*, 1991a; Robinson *et al.*, 1991; Clayton-Smith *et al.*, 1992b; Beuten *et al.*, 1993). Since the extent of the deletions in either disorder was indistinguishable, the parental origin of the deletion was the critical factor in determining the phenotype. Genomic imprinting was implicated in the pathogenesis of PWS and AS (Knoll *et al.*, 1989)

Three molecular categories of AS patients were defined based on presence and extent of the molecular deletion (Fig 1.3). Class 1 patients were deleted for loci D15S9-D15S13 and D15S18 and Class 2 patients for D15S9-D15S13. No molecular deletion was observed in Class 3 patients. This category included familial cases of AS, in which more than one sibling was affected. The possibility of a different molecular aetiology between Class 3 and deletion-positive patients was proposed (Knoll *et al.*, 1990).

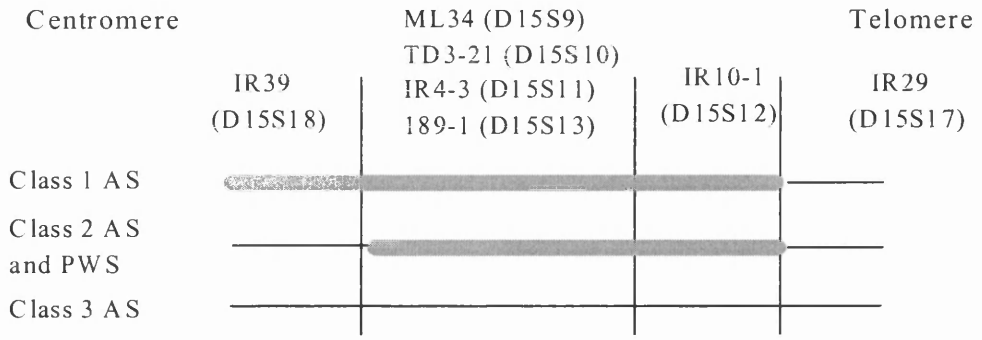


Figure 1.3 Schematic representation of the region deleted in PWS and three molecular classes of AS patients.
The loci identified by probes are in parentheses, and ordered from centromere to telomere. The deleted region is indicated by a thick grey line.

Large deletions of 15q11-13 account for approximately 70% of patients with PWS and AS (Chan *et al.*, 1993; Knoll *et al.*, 1993a; Saitoh *et al.*, 1994; Woodage *et al.*, 1994a; Webb *et al.*, 1995). The recurrence risk of AS or PWS with a cytogenetic deletion in the absence of a predisposing factor has been estimated at <0.1% (Cassidy, 1987; Webb *et al.*, 1993). In a study of 37 AS families with a deletion positive child, the unaffected siblings inherited either copy of the maternal 15 with impunity, and no maternal gonadal mosaicism was detected. This mechanism most likely occurs as a *de novo* event during gametogenesis (Webb *et al.*, 1993). Two cousins with 15q11-13 deletions, one with PWS and one with AS, have been reported. However, the presence of a familial paracentric inversion involving 15q11-13 will have predisposed the region to deletion (Clayton-Smith *et al.*, 1993b).

1.4.3 Uniparental disomy for chromosome 15

During routine RFLP analysis of individuals with PWS, two patients were identified as having maternal UPD for chromosome 15. The probands had inherited two different maternal alleles at six loci within 15q11-13 (D15S9-D15S13 and D15S18), and at a distal variable number tandem repeat (VNTR), indicating heterodisomy. Analysis of the 3' hypervariable region of the α-globin locus on chromosome 16 ruled out non-paternity. A balanced 13q15q Robertsonian translocation was observed in one patient, which would have incurred the non-disjunction event leading to UPD, but was not considered causal to the phenotype as this was also carried by unaffected maternal relatives. The PWS phenotype in

each case was attributed to the lack of inheritance of a paternal 15, analogous to the paternal deletion of 15q11-13, providing further evidence for genomic imprinting in this region (Nicholls *et al.*, 1989a). In subsequent surveys of the molecular pathology of PWS, many more cases of mUPD15 were reported. These were primarily heterodisomic, as detected by RFLP analysis (Robinson *et al.*, 1991; Mascari *et al.*, 1992), or multiplex amplification of dinucleotide repeat markers within 15q11-13 by the polymerase chain reaction (PCR) (Mutirangura *et al.*, 1993a).

AS was first associated with paternal UPD 15 in two patients, one demonstrating heterodisomy, the other isodisomy, at D15S24 and the subtelomeric locus D15S86. Linkage analysis of cytogenetic heteromorphisms confirmed uniparental inheritance of the chromosomes 15 (Malcolm *et al.*, 1991).

The frequency of mUPD observed in PWS is about 28%, accounting for most of the non-deletion cases, and the majority of these involve heterodisomy (Robinson *et al.*, 1991; Mascari *et al.*, 1992; Mutirangura *et al.*, 1993a; Nicholls *et al.*, 1993a; Saitoh *et al.*, 1994; Woodage *et al.*, 1994a; Webb *et al.*, 1995). In AS pUPD is infrequent, occurring in 3-5% of patients, with a ratio of 9:1 cases of isodisomy to heterodisomy (Nicholls *et al.*, 1992; Mutirangura *et al.*, 1993a). The disparity in frequency and type of UPD associated with PWS and AS, may be explained in the light of the mechanisms that lead to UPD, and that the rate of non-disjunction in maternal meiosis exceeds that in paternal meiosis. In the majority of AS cases due to pUPD15, complete isodisomy is observed. The predominant mechansim leading to pUPD in AS is thus thought to be 'monosomy rescue' by the paternal homologue in the zygote, due to an initial non-disjunction error at maternal MI, which left the egg nullisomic for chromosome 15 (Robinson *et al.*, 1993a). In PWS, the observation of maternal heterodisomy in most cases of mUPD15 is indicative of 'trisomy rescue', also resulting from an initial non-disjunction event at maternal MI. This has been definitively observed in PWS: In routine referrals for advanced maternal age in pregnancy, chorionic villus sampling (CVS) detected trisomy 15 of the placenta. Karyotypes of subsequent amniocenteses appeared normal and the pregnancies were continued to term. However the infants were born with PWS (Cassidy *et al.*, 1992; Purvis-Smith *et al.*, 1992; Morichon-Delvallez *et al.*, 1993). Maternal heterodisomy 15 was later demonstrated as the aetiological basis for PWS. This, together with confined placental mosaicism (CPM) for trisomy 15, was consistent with fertilisation of a disomic oocyte, and subsequent loss of the paternal homologue from the embryo to the trophectoderm (Cassidy *et al.*, 1992; Purvis-Smith *et al.*, 1992). Through

analysis of markers along the full length of the chromosome, it was estimated that 71-82% of mUPD15 cases resulting in PWS, and almost all cases of paternal isodisomy in AS, were due to the reciprocal products of maternal MI non-disjunction (Robinson *et al.*, 1993a; 1996). This is consistent with the demonstration that 95% of trisomy 21 cases are due to MI errors during oogenesis (Antonarakis *et al.*, 1991).

1.4.3.1 Advanced maternal age and UPD 15

The recurrence risk for PWS or AS due to UPD has been estimated at <2%. In PWS cases due to mUPD, maternal age was found to be significantly increased when compared to their deletion counterparts (Robinson *et al.*, 1991; Mascari *et al.*, 1992). An exponential increase in frequency of trisomy 15 in spontaneous abortions (Hassold and Chiu, 1985) and mUPD 15 (Robinson *et al.*, 1996) has been observed, when plotted against maternal age.

1.4.3.2 Phenotypic variation between patients with UPD and deletion

A less severe phenotype, including absence of characteristic facies, and less significant epilepsy, ataxia and communication problems, have been reported for a small number of AS patients with UPD (Bottani *et al.*, 1994; Gillessen-Kaesbach *et al.*, 1995a). Hypopigmentation, observed in both PWS and AS, correlates with a 15q11-13 deletion and is the only phenotypic difference consistently seen between the deletion and UPD categories (Butler *et al.*, 1986; Mascari *et al.*, 1992; Bottani *et al.*, 1994). Since this is a manifestation present in both PWS and AS deletion patients, the gene underlying this defect is not subject to imprinting, but due instead to hemizygosity from deletion of one allele (Rinchik *et al.*, 1993). This also indicates that 15q11-13 is the only imprinted region that exists on chromosome 15.

1.4.4 Chromosomal rearrangements

Various chromosome 15 rearrangements have been associated with AS and PWS, including paracentric (Clayton-Smith *et al.*, 1993b) and pericentric inversions (Webb *et al.*, 1992), mosaic supernumerary marker chromosomes 15 (inv dup 15) (Robinson *et al.*, 1993b; Woodage *et al.*, 1994a; Webb, 1994; Spinner *et al.*, 1995), and translocations (Hulten *et al.*, 1991; Smeets *et al.*, 1992; Smith *et al.*, 1994; Toth-Fejel *et al.*, 1996; Horsthemke *et al.*, 1996). However, in these cases the rearrangement *per se* is not usually causal to disease, as

other relatives are often carriers. The rearrangements are usually accompanied by a pathologic deletion or UPD, which was predisposed to.

1.4.4.1 Definitive evidence for imprinting within 15q11-13 is provided by two translocation families

Two translocation families have been described in which both AS and PWS have arisen on either branch, providing definitive evidence for genomic imprinting within 15q11-13. In one report the male ancestor carried a balanced 15;22 translocation which was inherited by several unaffected family members. However, paternal inheritance of an unbalanced form of the translocation in the first generation, led to monosomy for 15pter→q13 and a duplication of 22pter→q11, with consequent PWS. In the second generation, maternal transmission of same unbalanced translocation resulted in AS (Hulten *et al.*, 1991). In the second report, two normal brothers had inherited a Robertsonian 6;15 translocation from the grandmother, which had led to PWS and AS in two cousins. The offspring of one brother had PWS due to a paternal interstitial deletion on the derivative 15, whereas AS in the child of the other brother was due to paternal uniparental heterodisomy (Smeets *et al.*, 1992). These cases demonstrated that AS and PWS were alternative phenotypes resulting from the loss of 15q11-13, dependent on the sex of the transmitting parent.

1.4.5 Imprintor Mutations

1.4.5.1 Parent-of-origin specific methylation in 15q11-13

As evidence of a role for DNA methylation in genomic imprinting emerged from studies of mouse gametes and transgenes, a search for sites of differential methylation between the parental alleles of endogenous sequences was conducted (Driscoll *et al.*, 1992). A method was devised to screen for sites of allele-specific methylation within 15q11-13 by Southern analysis of AS and PWS patients with deletions or UPD, using methylation sensitive restriction enzymes. Patients with deletions are hemizygous for one parental allele at 15q11-13, and those with UPD inherit both their alleles from a single parent (maternal in the case of PWS and paternal in the case of AS). The use of their DNA thus enabled the methylation imprints between the two parental homologues to be examined separately. Lymphocyte DNA was digested with *Hpa*II or *Hha*I, which cleave DNA only if the

recognition site is unmethylated. Where enzyme cut sites are methylated on one parental allele but not the other, differently sized restriction fragments are generated giving distinct banding patterns between the PWS and AS samples. The cDNA probe DN34 detected a striking difference in band intensity between the *Hpa*II pattern for the maternal and paternal alleles at D15S9. This represented the first epigenetic modification associated with an endogenous DNA sequence in mammals (Driscoll *et al.*, 1992).

More distinctive methylation differences were subsequently identified, using the same methodology, at D15S63 by probe PW71 (Dittrich *et al.*, 1992; 1993), and at the 5' CpG island of the *SNRPN* gene, which encompasses the first exon (Sutcliffe *et al.*, 1994). At these two loci, the paternal allele is unmethylated and the maternal allele is hypermethylated. For *SNRPN* exon 1, digestion of DNA with *Not*I and *Xba*I gives a 4.2 kb fragment in PWS DNA, corresponding to the maternal allele, and a 0.9 kb fragment in AS, corresponding to the paternal allele (Fig 1.4). Both fragments were present in normal individuals and AS patients with biparental inheritance of chromosome 15 and no deletion (Kubota *et al.*, 1996a).

Since the allele-specific methylation differences enabled the rapid detection of AS or PWS with deletions or UPD, through the observation of just the maternal or paternal band in the DNA, methylation testing was employed for the genetic diagnosis of these disorders (Clayton-Smith *et al.*, 1993; Glenn *et al.*, 1993a; Reis *et al.*, 1994).

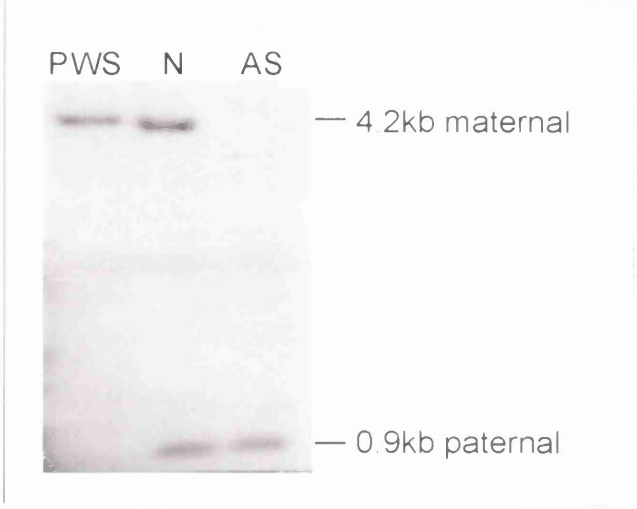


Fig 1.4 Allele-specific methylation pattern at *SNRPN* exon 1.

Lymphocyte DNA from PWS and AS patients with 15q11-13 deletions, and normal DNA (N) digested with *Xba*I and methylation-sensitive enzyme *Not*I, and probed with *SNRPN* exon 1. The 4.2kb maternal allele seen in PWS DNA is methylated and so uncut by *Not*I. The paternal allele detected in AS DNA is digested to 0.9kb by *Not*I, since it is unmethylated.

1.4.5.2 Imprintor Mutations

During the course of investigating AS and PWS patients at differentially methylated loci within 15q11-13, a small number of patients demonstrated the typical AS or PWS methylation patterns in the absence of a deletion or UPD. In the case of AS only the paternal band was detected, despite biparental inheritance of chromosome 15, indicating that the maternal homologue carried the paternal methylation imprint. In PWS cases, only the maternal allele was observed indicating that the paternal homologue bore the maternal epigenotype. These were termed imprintor mutations (Glenn *et al.*, 1993; Reis *et al.*, 1994).

Imprintor mutations occur in both familial and sporadic cases of AS and PWS. Familial cases have been associated with microdeletions of 6-80kb, but the deletions do not cause disease *per se,* as these were also found in normal parents and ancestors. In PWS the deletions include the first exon of *SNRPN* and occur on the paternal allele, but transmission through the maternal germ line is asymptomatic. In AS the deletions map further proximal, sometimes including D15S63 (Fig 1.8). These are always present on the maternal allele, but may be transmitted silently through the paternal germline (Sutcliffe *et al.*, 1994; Buiting *et al.*, 1994, 1995; Saitoh *et al.*, 1996). On the basis of these data, the existence of a bipartite imprinting center (IC) that is involved in resetting the imprint signal to correspond to the sex of the parent, was proposed. A model explaining the occurrence of imprintor mutations was put forward in which *cis*-acting deletions or mutations of the IC would fix the epigenotype of the chromosome on which it occurred (Fig 1.5). Transmission through the same sex would be silent. However, transmission of the chromosome through the germ line of the opposite sex would prevent the reversal of the epigenetic modification. This would be manifested in the offspring as AS or PWS (Reis *et al.*, 1994; Buiting *et al.*, 1995; Saitoh *et al.*, 1996).

Imprintor mutations occur in <2% of PWS patients and 8% of AS patients. Where a microdeletion is present on the transmitting parental allele the recurrence risk is 50%. However, mosaicism for a microdeletion has been observed in the mothers of several AS imprintor cases, so the recurrence risk in these cases is inestimable (Saitoh *et al.*, 1996; Bürger *et al.*, 1997). *De novo* events resulting in sporadic imprintor cases of AS have also been observed in the absence of microdeletions or mutations of the IC. In a few cases the incorrect imprint arose on the grandmaternal allele, indicating that these are not due to a failure of the imprint switch, but may have occurred in the maternal germ line or postzygotically (Buiting *et al.*, 1998).

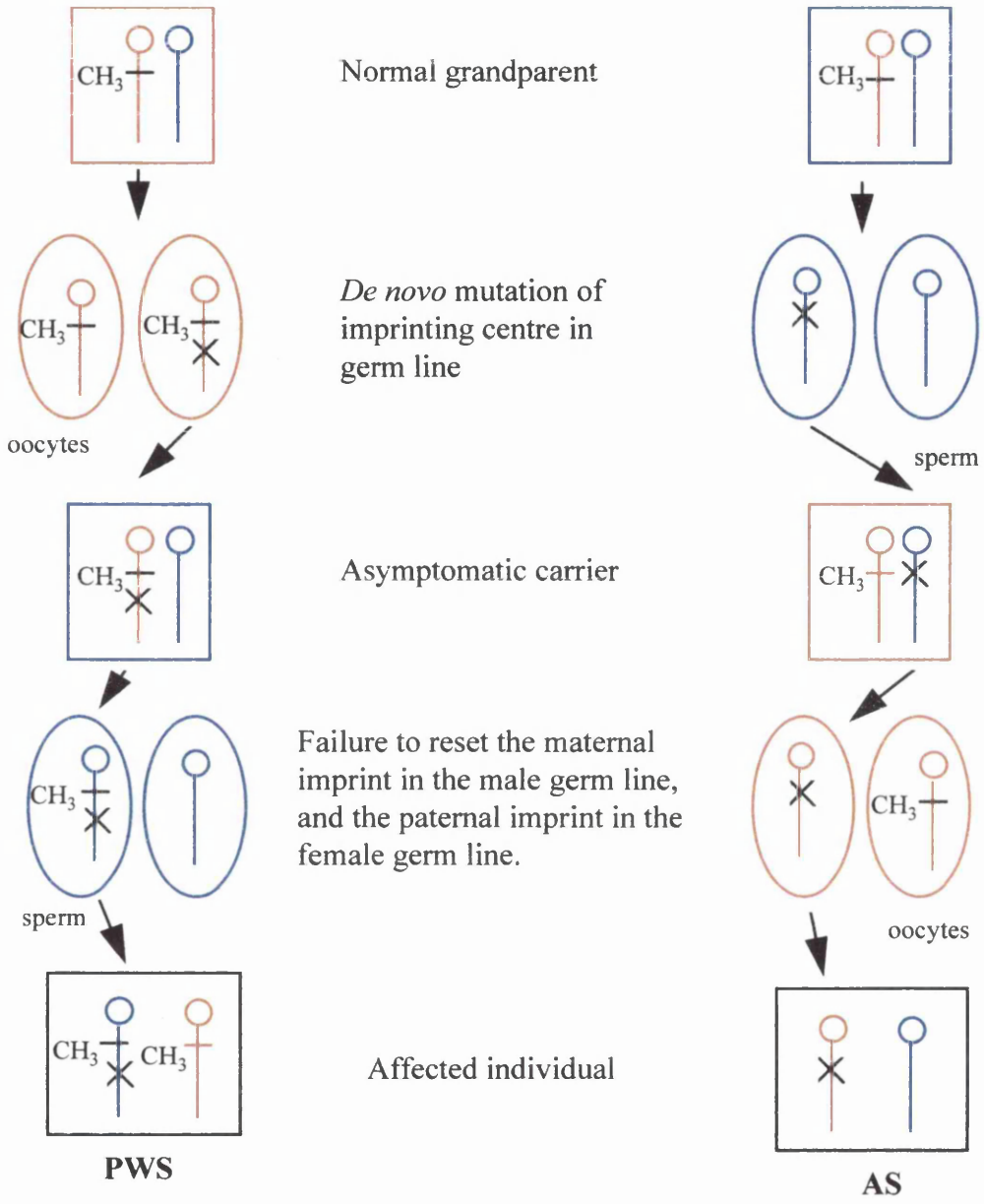


Fig 1.5 Model for the effect of IC mutations on resetting of the epigenotype in 15q11-13

Somatic cells or individuals are represented by boxes, blue for males and red for females. The germ cells are represented by ovals. The paternally inherited chromosome is in blue and the maternal homologue in red. Methylated alleles are labeled CH_3. IC mutations are denoted by a cross (x) on the respective chromosome. Imprintor mutations causing PWS result from a failure to reset the female epigenotype from the grandmother, to the male epigenotype in the paternal germ line. Inheritance of a paternal chromosome 15 with a maternal methylation imprint leads to PWS. Conversely in most AS imprintor mutation cases, the methylation status of the grandpaternal homologue is not reversed to the female epigenotype the maternal germ line. The maternally inherited chromosome 15 thus harbours the paternal methylation imprint.

1.4.6 Dominant imprinted mode of inheritance of AS in familial cases

Over 20% of AS patients remain who have no detectable deletion, disomy, imprintor mutation or other chromosomal rearrangement. These are termed non-deletion, UPD or imprintor (NDUI) cases, and are often familial (Clayton-Smith, 1992; Clayton-Smith and Pembrey, 1992).

To determine whether the underlying genetic aetiology of AS in NDUI patients was the same as in the other mechanisms for disease, the segregation of alleles on chromosome 15 was studied and linkage analyses performed in NDUI AS families. Affected siblings were consistently shown to have inherited the same maternal alleles for chromosome 15, but different paternal alleles. This provided strong evidence against a recessive mode of inheritance of a locus on chromosome 15. Unaffected siblings inherited different maternal 15q11-13 alleles. In several extended pedigrees spanning 3 to 5 generations, transmission of a common 15q11-13 allele occurred asymptomatically through male ancestors, but AS resulted when maternally inherited (Clayton-Smith *et al.*, 1992a; Meijers-Heijboer *et al.*, 1992; Wagstaff *et al.*, 1992; 1993). A dominant mode of inheritance modified by genomic imprinting was proposed whereby AS would result only if the mutant allele was maternally inherited, whereas a latency of the phenotype would result when the mutation was paternally transmitted (Clayton-Smith *et al.*, 1992a). Linkage to GABRβ3 and D15S10 (Fig 1.8) was demonstrated in AS pedigrees, using this model for imprinted inheritance, with maximum lod scores of 5.40 and 3.52 at $\theta = 0$ (Meijers-Heijboer *et al.*, 1992; Wagstaff *et al.*, 1993). The odds ratio in favour of genomic imprinting versus reduced penetrance, was calculated at 9.25×10^5 (Meijers-Heijboer *et al.*, 1992).

These data provided evidence for a common molecular aetiology for AS in NDUI patients and cases due to other chromosome 15 anomalies. A recurrence risk of 50% is indicated in these cases (Clayton-Smith *et al.*, 1992a; Meijers-Heijboer *et al.*, 1992; Wagstaff *et al.*, 1993). This category of patients also evince that AS results from disruption of a single gene on the maternally inherited 15q11-13 allele. Mutation, absence or lack of expression of this gene on the maternal homologue would result in AS (reviewed by Nicholls, 1993b).

1.4.7 The underlying genetic aetiology of PWS and AS

PWS is caused by the loss of a functional paternal chromosome 15q11-13, due to deletion, disomy, imprintor mutations, or through chromosomal rearrangements (Fig 1.6). Since these mechanisms account for all PWS patients, and no mutations have been identified in a single gene in PWS patients, this is considered to be a contiguous gene disorder, involving perhaps several paternally expressed imprinted genes (reviewed by Nicholls, 1993b; 1993c).

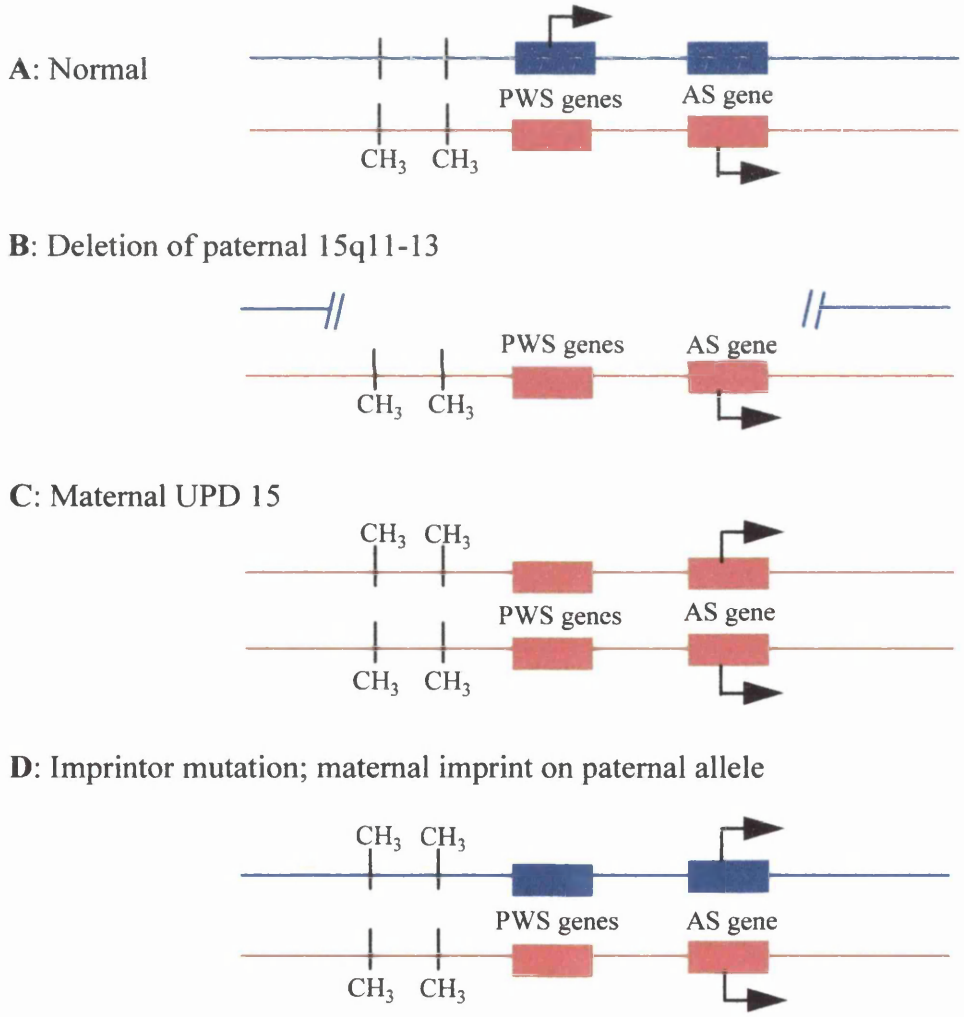


Fig 1.6 Diagrammatic representation of the common mechanisms causing PWS.

The maternal 15q11-13 allele is coloured in red and the paternal allele in blue. Genes are represented by boxes. Arrows denote functional activity. The parental epigenotype is represented by vertical black lines; methylated loci are labeled CH_3. Each mechanism results in absence (B and C) or loss of function (D) of paternally expressed genes within 15q11-13.

AS results from the lack of a maternal contribution to 15q11-13, whether by deletion, disomy, imprintor mutations and chromosomal rearrangements (Fig 1.7). The existence of a category of NDUI patients provides evidence that AS is a single gene defect. Linkage of AS to 15q11.2 in NDUI pedigrees and demonstration of maternal transmission of the disorder, provides strong evidence for the existence of an imprinted gene expressed exclusively from the maternal 15q11-13 (Fig 1.7E) (reviewed by Nicholls, 1993b).

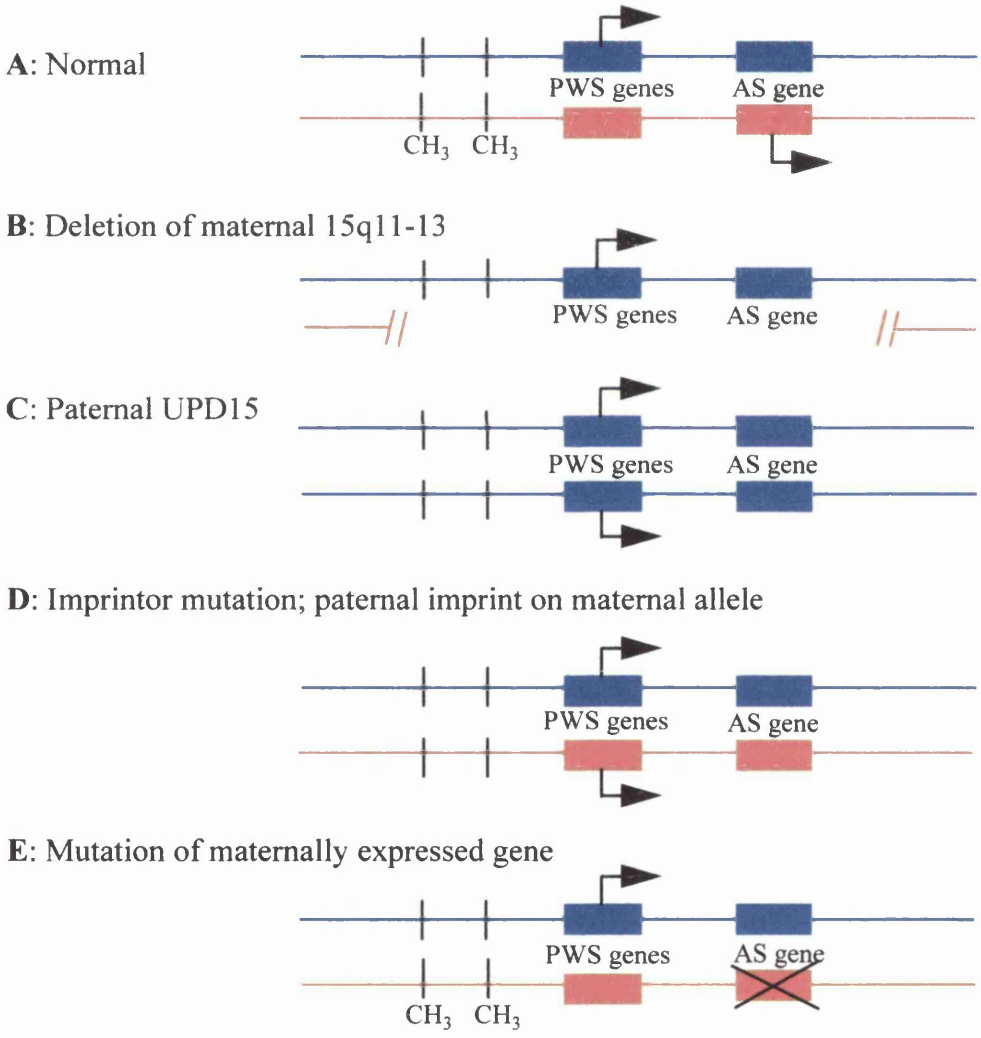


Fig 1.7 Diagrammatic representation of the mechanisms leading to AS.

The paternal 15q11-13 allele is depicted in blue, the maternal allele in red. Genes are represented by boxes, and arrows denote transcription. The vertical black lines represent the epigenetic modification, with methylated alleles labelled. Each mechanism causes absence (B, C) or loss of function (D, E) of a maternally expressed imprinted gene within 15q11-13.

1.5 Genetic diagnosis of PWS and AS

Infants with AS and PWS are often difficult to diagnose clinically since they do not display the full disease spectrum (reviewed by Holm *et al.*, 1993; Williams *et al.*, 1995). Certain aspects of the PWS phenotype including obesity, short stature and delayed puberty, can be treated prior to presentation, if the condition is diagnosed early. Therefore, early confirmation of a clinical diagnosis is important (Malzac *et al.*, 1983; Gillessen Kaesbach *et al.*, 1995). Furthermore, the mechanism of disease in individual patients with PWS and AS needs to be ascertained, in order to assess the sibling recurrence risk. However, no single test is sufficient for a complete classification of the underlying defect. A strategy combining at least two tests is usually employed (Cassidy *et al.*, 1996).

1.5.1 Methylation testing

Methylation testing at D15S63 and *SNRPN* exon 1 (section 1.4.5.1) detects all the categories of PWS patients, and 80% of AS patients in a single test, but does not distinguish between the mechanisms (Gillessen-Kaesbach *et al.*, 1995; Kubota *et al.*, 1996a; Beuten *et al.*, 1996a, b). AS patients with biparental inheritance and no deletion or imprintor mutations, exhibit a normal methylation pattern, so that a diagnosis of AS can not be ruled out by this method. For both disorders, methylation testing is a requirement for the detection of imprintor mutation cases (Reis *et al.*, 1994; Beuten *et al.*, 1996a,b). Parental DNA samples are not required for methylation analysis (van den Ouweland *et al.*, 1995; Cassidy *et al.*, 1996).

Methylation testing is routinely applied in diagnostic laboratories for screening large numbers of samples referred for suspected AS, PWS, or to rule out PWS in infants with hypotonia of unknown cause. A positive diagnosis will be followed by additional tests to determine the molecular defect for genetic counseling purposes (Gillessen-Kaesbach *et al.*, 1995; Cassidy *et al.*, 1996). The efficacy of the PW71 and the *SNRPN* exon 1 probe have been validated in studies of over 500 PWS patients, and patients with infant hypotonia, and are the preferred probes for routine diagnostic purposes (Gillessen-Kaesbach *et al.*, 1995; Kubota *et al.*, 1996a). Prenatal diagnosis by CVS or amniocentesis has been advised against however, due to inconsistent methylation patterns for D15S63 in these cell types (Dittrich *et al.*, 1993; Kubota *et al.*, 1996b).

1.5.2 Detection of 15q11-13 deletions and chromosomal rearrangements

Since the majority of AS and PWS patients have a deletion of 15q11-13, high resolution cytogenetic analysis or fluorescence *in situ* hybridisation (FISH) are often employed as primary or supplementary tests to methylation analysis to identify deletions and chromosomal rearrangements (Erdel *et al.*, 1996; Cassidy *et al.*, 1996). Unique PWS patients with balanced translocations have been identified with normal methylation imprints at D15S63 and *SNRPN* (Schulz *et al.*, 1996; Sun *et al.*, 1996). This also allows the detection of rearrangements of chromosomes other than 15 for exclusion and referral (Crolla *et al.*, 1995). FISH analyses have employed yeast artificial chromosomes (YACs) (Kuwano *et al.*, 1992), the Alu-PCR products of YACs (Erdel *et al.*, 1996), cosmids (Bettio *et al.*, 1995; Butler, 1995) and phage clones (White and Knoll, 1995) derived from 15q11-13, as probes. FISH supercedes cytogenetic analysis in the detection rate of deletions, and may also identify microdeletions (Teshima *et al.*, 1996). Cosmid probes for *SNRPN*, D15S10, *GABRB3* and D15S11 within 15q11-13 and chromosome 15 centromeric probes (such as DZ15z1) are commercially available (Oncor, Gaithersburg, MD). Probe D15Z1, is used to orientate the chromosome 15 homologues in the identification of translocations, duplications, inversions and inv dup (15) (Crolla *et al.*, 1995; Bettio *et al.*, 1995). Parental DNA samples are not required for FISH analysis except where a recurrence risk is needed in the case of chromosomal rearrangement (Cassidy *et al.*, 1996). FISH has also been applied for prenatal testing for chromosome 15 anomalies. Two cases with 15q11-13 deletions and one with a paracentric inversion involving 15q11-13 were detected in amniotic cells after referral for advanced maternal age and abnormal cytogenetic results (Toth-Fejel *et al.*, 1995).

1.5.3 Detection of UPD15

Short tandem repeat markers (STRs) with high polymorphic information content (PIC) are currently used to demonstrate UPD15, where methylation analysis is positive for PWS or AS, and no deletion has been observed. Parental DNA samples are required for segregation analysis and to determine heterodisomy or isodisomy. Paternity may be confirmed by analysis of STRs on other chromosomes (reviewed by Cassidy *et al.*, 1996). This method has been employed for prenatal diagnosis of UPD15, using fetal blood or amniotic fluid, following the observation of CPM for trisomy 15. Sequences for over 100 chromosome 15 STRs with high PIC values are available from the Genome Database (Christian *et al.*, 1996).

1.6 Physical mapping in 15q11-13

1.6.1 Construction of a YAC contig spanning 15q11-13

A YAC contig was constructed which extended from D15S18 to D15S12, across the 3.5-4Mb PWS and AS deletion region. This comprised of 24 overlapping YACs identified from the CEPH and St. Louis YAC libraries by screening with the eight single-copy DNA probes from 15q11-13 previously identified (section 1.4.2.1) (Kuwano *et al.*, 1992; Mutirangura *et al.*, 1993b). Two gaps remained at the centromeric and telomeric ends of the contig, in the region of the common deletion breakpoints (Mutirangura *et al.*, 1993) (Fig 1.8). This YAC contig formed the basis for cloning of specific regions within 15q11-13 into smaller vector systems including cosmids and phage (Malcolm and Donlon, 1994; Robinson *et al.*, 1997), mapping markers and genes within 15q11-13 (Mutirangura *et al.*, 1993b; Knoll *et al.*, 1993; Nakao *et al.*, 1994), and positional cloning (Wevrick *et al.*, 1994). Individual YAC clones from this contig were used in this project for physical mapping and positional cloning.

1.6.2 Ordering of DNA markers and genes within 15q11-13

Deletion mapping of the standard eight markers from 15q11-13 had failed to resolve the mapping order of five of these markers since they were consistently deleted in PWS patients (section 1.4.2.1). These probes, and the *GABRB3*, *GABRA5* and *SNRPN* genes, were ordered by multicoloured FISH analysis of interphase. YACs and phage clones containing the DNA markers and genes were used as probes. These were individually labelled to give differently coloured fluorescent signals, and hybridised simultaneously in groups of two or three probes, to interphase nuclei. Centromeric and telomeric probes were used to anchor the chromosomes and orientate the probes. STS markers were mapped to the YAC contig, giving an average resolution of 1 STS per 200 kb (Kuwano *et al.*, 1992; Knoll *et al.*, 1993b; Mutirangura *et al.*, 1993b). The *UBE3A* gene was mapped by PCR to YACs 132D4 and B230E3 (Nakao *et al.*, 1994), but the order with respect to D15S10 was not established. The order of the genes and markers within 15q11-13 are shown in Fig 1.8.

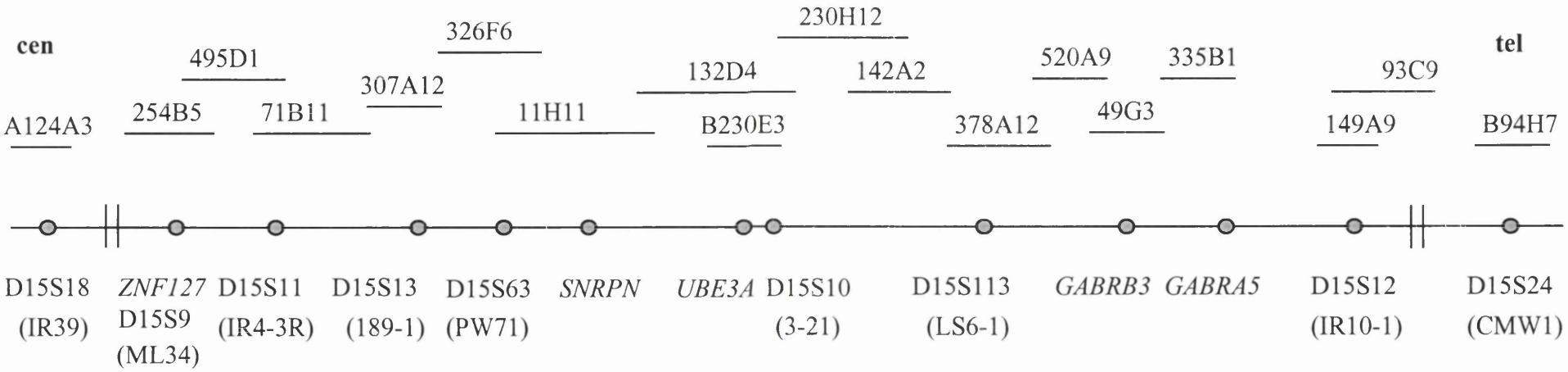


Fig 1.8 YAC map spanning the AS and PWS cytogenetic deletion region and marker order within 15q11-13.

YACs are drawn as horizontal bars with identification number above. The vertical double lines indicate the common deletion breakpoints. The proximal and distal deletion breakpoints were spanned by YACs 254B5 and 93C9 respectively. Two gaps were identified in the YAC contig, between D15S18 and D15S9 at the centromeric end, and *GABRA5* and D15S12 at the telomeric end. Markers and genes are denoted by circles, and the probes identifying them are given in parentheses. The order of the markers and genes was established by multicolour FISH analysis of interphase chromosomes.

1.6.3 Restriction mapping by pulsed-field gel electrophoresis (PFGE)

PFGE analysis using rare-cutting restriction enzymes such as *Not*I identified several CpG islands within 15q11-13, which have subsequently been associated with genes in the region. CpG islands were identified at D15S12 (the *P* gene promoter), at D15S9 (associated with *ZNF127*), and at D15S10 (Kirkilionis *et al.*, 1991). The CpG island at D15S10 was cloned in a plasmid, and the genomic clone is referred to as OP2 (Woodage *et al.*, 1994). Several CpG islands surrounding the *GABRB3* gene were identified (Sinnett *et al.*, 1993), which have been associated with the $GABA_A$ receptor subunit cluster, *GABRA5* and *GABRG3* (Meguro *et al.*, 1997).

1.6.4 The AS and PWS candidate regions

1.6.4.1 The AS and PWS loci are distinct

Evidence that the AS and PWS loci are separate emerged from the study of AS Family S, in which a microdeletion of D15S10, D15S113 and part of the *GABRB3* gene, was present in three generations. The grandfather and mother were phenotypically normal despite carrying the deletion. However, maternal transmission of this microdeletion resulted in AS in three children, in the third generation (Hamabe *et al.*, 1991; Saitoh *et al.*, 1992, Kuwano *et al.*, 1992). PWS was not manifested in the mother, despite transmission of the microdeletion from the grandfather, indicating that the loci responsible for AS and PWS are distinct (Saitoh *et al.*, 1992).

1.6.4.2 The AS candidate region

The AS candidate region was originally defined by the submicroscopic deletion detected in AS Family S. The distal deletion breakpoint lay within intron 3 of *GABRB3* (Sinnett *et al.*, 1993). The junction fragment of the proximal deletion breakpoint, D15S174, was cloned and located just centromeric to D15S10. The size of the deletion was estimated at 1-1.5Mb (Greger *et al.*, 1993) (Fig 1.9).

In AS family WJK, two affected siblings had inherited different maternal alleles at D15S63, but the same maternal allele at D15S122, D15S113 and *GABRB3* (Greger *et al.*, 1994). A recombination event had occurred between D15S63 and D15S122, indicating that a mutation of the maternal allele that was shared by the AS siblings, lay telomeric to D15S122. Although this recombination family does not preclude D15S122 and the region just proximal

from the AS candidate region, the disease-causing mutation in Family WJK must lie distal to D15S122 (Greger *et al.*, 1993; 1994)(Fig 1.9).

A mother (B.S.) and son (D.H.) who had mild mental retardation but no other features of AS, were described with a deletion spanning D15S113, *GABRB3* and D15S12. Both patients were intact for D15S10, so that the proximal deletion breakpoint lay between this marker and D15S113. Since the deletion in D.H. was maternally inherited, the region most likely to be involved in AS lay proximal to D15S113 (Michaelis *et al.*, 1995). Thus the AS candidate region was defined in 1995, by D15S174 at the centromeric extent and D15S113 at the telomeric extent (Fig 1.9).

1.6.4.3 The PWS candidate region

The PWS candidate region (PWS-CR) is defined by the overlap between the molecular deletion of a PWS patient, and the submicroscopic deletion in AS Family S, which separated the PWS from the AS region. The centromeric breakpoint of a PWS deletion, which maps just upstream of D15S63, defines the proximal extent of the PWS-CR. The distal extent of the PWS-CR is defined by D15S174, the centromeric breakpoint of the microdeletion in AS Family S. Since paternal transmission of this deletion did not confer the PWS phenotype, D15S174 represents the boundary between the PWS and AS candidate regions. The PWS-CR was estimated to span 320 kb (Buiting *et al.*, 1993)(Fig 1.9). The PWS-CR is interpreted with caution however, as the paternally expressed gene *ZNF127* lies proximal to this region (Fig 1.9). *ZNF127* is repressed in PWS patients with imprintor mutations due to a deletion of the *SNRPN* CpG island (Saitoh *et al.*, 1996), indicating that the imprinting processes that regulate the genes involved in PWS act beyond the boundary of the PWS-CR.

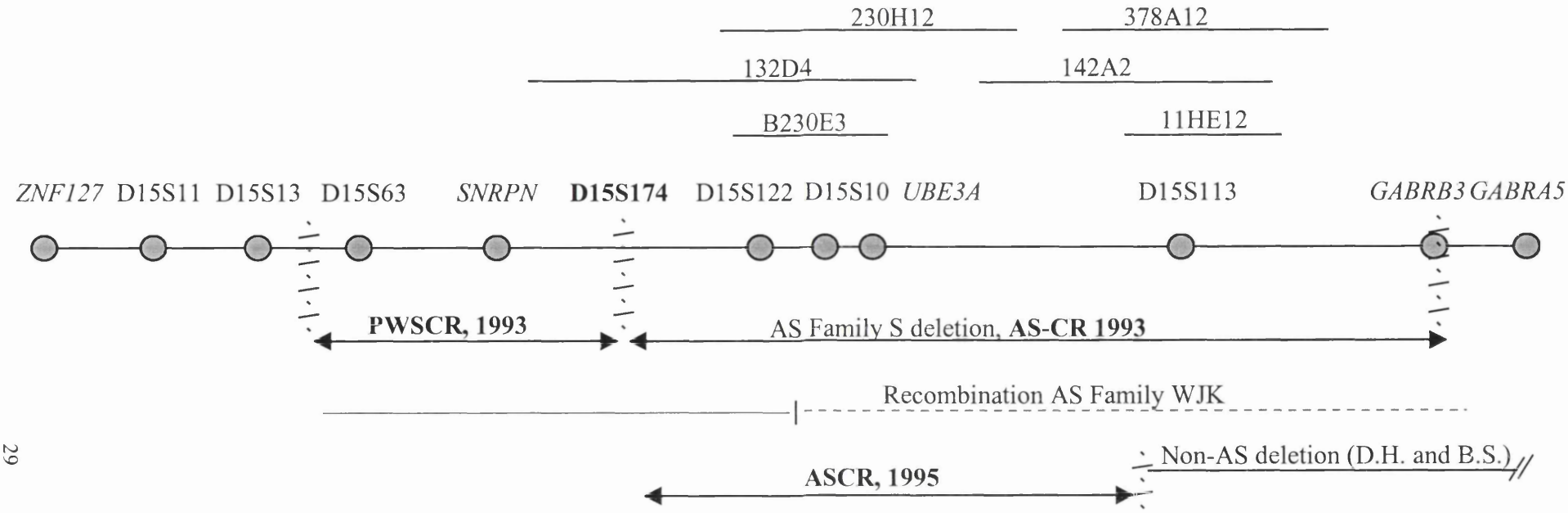


Fig 1.9 Schematic representation of the PWS and the AS candidate regions (PWSCR and ASCR).

A map of PWS and AS candidate regions within 15q11-13 is shown. Markers and genes are represented by grey circles. The jagged lines indicate deletion breakpoints referred to in the text, that define the PWS and AS regions. The double-arrows indicate the PWS and AS candidate regions. The deletion breakpoint between D15S13 and D15S63 defines the proximal boundary of the PWSCR. The proximal deletion breakpoint in AS Family S at D15S174 marks the proximal extent of the ASCR and boundary between the PWSCR and ASCR. For AS recombination family WJK, the full grey line shows the region where the maternal alleles in the two affected siblings differ, and the dashed line indicates the region where both siblings share the same maternal allele. The maternally inherited non-AS deletion seen in patient D.H. is indicated. The redefined ASCR (1995) lies between D15S174 and the centromeric deletion breakpoint in patient D.H. proximal to D15S113. YACs forming the contig across the ASCR (1995) that were used during the course of this project, are drawn as horizontal bars above. The order of *UBE3A,* D15S122 and D15S10 was unknown.

1.7 Imprinting within 15q11-13

Imprinted genes demonstrate four main characteristics: (i) monoallelic expression (ii) clustering in asynchronously replicating domains (iii) conservation with imprinted regions in the mouse (iv) association with parental-allele dependent methylation. In describing these characteristics particular reference is given to 15q11-13 and the genes within this region, but examples of other imprinted genes and regions are also cited.

1.7.1 Monoallelic expression

Monoallelic expression refers to the transcription of a gene exclusively from one parental allele. This is the fundamental characteristic of imprinted genes. A constantly increasing number of imprinted genes have been identified on the human and mouse autosomes. These are listed, with associated phenotypes and references, on the world wide web page: **http://www.mgu.har.mrc.ac.uk/anomaly/anomaly.html**

Within 15q11-13 several maternally imprinted genes and uncharacterised transcripts have been identified, which are expressed exclusively from the paternal allele in all tissues tested, including *ZNF127*, *NDN*, *SNRPN*, *IPW* and certain PAR transcripts (section 1.7.5). Thus imprinting of the paternally expressed genes in this region appears to be ubiquitous.

Imprinted expression of genes may be complex however, demonstrating tissue, developmental-stage, or promoter specificity, as illustrated by the insulin-like growth factor 2 gene (*IGF2*) on chromosome 11. *IGF2* is transcribed exclusively from the paternal allele in most fetal and adult tissues. However, biallelic expression is observed in the choroid plexus, leptomeninges and adult liver, indicating tissue-specific loss of imprinting (Ohlsson *et al.*, 1993; Giannoukakis *et al.*, 1993). *IGF2* expression is monoallelic in fetal liver and until a few months after birth, at which time it becomes biallelic. This biallelic mode of expression is maintained in the mature liver, illustrating developmental-stage specificity of imprinting in this tissue (Kalscheuer *et al.*, 1993). Imprinted expression of *IGF2* is also promoter dependent. *IGF2* has four promoters in humans, P1, P2, P3 and P4. Transcripts derived from promoters P2, P3 and P4 are specific to the paternal allele, whereas expression from P1 occurs from both parental alleles, as witnessed in adult liver (Vu and Hoffman, 1994). Imprinting of the Wilm's tumour suppressor gene *WT1*, is tissue-specific and polymorphic within the human population. Monoallelic or preferential expression from the maternal allele

has been demonstrated in fetal brain and placenta, with biallelic expression in the other major fetal organs. However, biallelic expression has also been detected in a number of placental samples, which was not associated with gestational age, but was attributed to polymorphic imprinting of this gene (Jinno *et al.*, 1994; Nishiwaki *et al.*, 1997).

1.7.2 Asynchronous replication

Asynchronous replication timing was first observed for the X chromosome in female mammalian cells, whereby the active X replicated early and the inactive X late, in S-phase of the cell cycle (Morishima *et al.*, 1962). Asynchronous replication between the two parental homologues has also been observed in regions containing imprinted genes, in both the mouse and human genomes. The early replicating homologue has a euchromatin-like structure, whereas the late replicating copy is heterochromatin-like. Replication timing patterns are currently analysed by FISH using probes specific to the region being analysed. Asynchronous replication timing is visualised as a single hybridisation signal on the unreplicated homologue and a double signal on the early-replicating homologue within the interphase nucleus of a cell, at the time of induced cell-cycle arrest (Izumikawa *et al.*, 1991; Kitsberg *et al.*, 1993).

15q11-13 was shown to replicate asynchronously in normal lymphocytes and lymphoblasts, with the paternal allele replicating earlier during S-phase than the maternal allele at D15S9 (*ZNF127*), D15S11, D15S13, D15S63, *SNRPN*, D15S10 and *GABRB3* (Kitsberg *et al.*, 1993; Knoll *et al.*, 1994; Gunaratne *et al.*, 1995; Watrin *et al.*, 1997). In contrast *GABRA5* replicates earlier on the maternal allele than the paternal allele (Knoll *et al.*, 1994; LaSalle and Lalande, 1995). Replication of the *P* locus was seen to be asynchronous, but was random with respect to the parental homologue (LaSalle and Lalande, 1995). Thus replication asynchrony extends across the entire imprinted region. Aynchronous replication of the PWS region was found to occur earlier during S-phase in neuroblasts where *SNRPN, NDN* and some of the PAR transcripts are predominantly expressed, compared to lymphocytes (Gunaratne *et al.*, 1995).

UPD 15, del(15q11-13) and imprintor mutations within 15q11-13 have been shown to disrupt the replication timing and kinetics of the chromosomes 15 in PWS and AS patients: these replicated at a different stage of S-phase than their normal parental homologues, and

uniparental homologues replicate simultaneously (LaSalle and Lalande, 1995; Gunaratne *et al.*, 1995).

1.7.3 Human-mouse synteny in imprinted regions

15q11-13 is syntenic to central mouse chromosome 7, in the region of the pink-eyed dilution locus, *p* (Chaillet *et al.*, 1991b; Gardner *et al.*, 1992; Nicholls *et al.*, 1993). This was first shown by the localisation of two evolutionarily conserved cDNAs DN34 (*ZNF127*) and MN7 from 15q11-13, to mouse chromosome 7 by somatic cell hybrid mapping and analysis of restriction fragment length variants (RFLVs). MN7 detects a multitude of distinct loci within human 15q11-13 and 16 (Buiting *et al.*, 1992), but identified a single ancestral locus on mouse 7 (Chaillet *et al.*, 1991b). Several *p* mouse mutants characterised by hypopigmentation have been described, which are primarily derived from radiation-induced deletions (Lyon *et al.*, 1992). These served as tools for early mapping of genes homologous to the PWS/AS region. *Dn*34 (*Zpf127*), MN7, *p* and *Gabrb3* were ordered with respect to one another by dosage analysis of *p* mouse mutants with overlapping deletions (Nicholls *et al.*, 1993). Murine homologues have now been identified for all the major genes within human 15q11-13 (section 1.7.5). These are located in the same order but the opposite orientation within the human and mouse syntenic clusters (Fig 1.10). As a rule the imprinting status of human and mouse homologues is conserved. This is the case for all the imprinted genes identified within human 15q11-13 and central mouse 7 (Nicholls *et al.*, 1993; Sutcliffe *et al.*, 1996; Wevrick and Franke, 1997; Macdonald and Wevrick, 1997; Watrin *et al.*, 1997) (Fig 1.10). The region of central mouse 7 that is equivalent to the human PWS and AS region is defined by the centromeric breakpoint of the X-autosome insertional translocation Is(InX;7)Ct (shortened to IsICt) and the T9H translocation breakpoint (Fig 1.10) (Cattanach *et al.*, 1997).

Another example of synteny between the human and mouse imprinted regions is provided by distal mouse chromosome 7 and the imprinted region on human 11p15.5, that is involved in BWS. Growth enhancement has been observed in chimaeric embryos mosaic for paternal disomy of distal 7, analogous to the phenotype of BWS (Ferguson-Smith *et al.*, 1991). This cluster contains at least eight imprinted genes in the mouse, including the paternally expressed genes *Igf2* and *Ins2,* and maternally expressed genes *H19, Mash2,* $p57^{KIP2}$ and *Kvlqt1* (Paulsen *et al.*, 1998; Caspary *et al.*, 1998). The imprinting status of these

individual genes is conserved in humans (Zhang and Tycko, 1992; Giannoukakis *et al.*, 1993; Reik and Maher, 1997; Caspary *et al.*, 1998).

The insulin-like growth factor 2 receptor gene (*Igf2r)* on mouse chromosome 17 is one exception where the imprinting status is not conserved between species. *Igf2r* is expressed monoallelically from the maternal allele in the mouse, but the human homologue is not imprinted (Kalscheuer *et al.*, 1993).

1.7.4 Mouse models for PWS and AS

On the basis of the synteny between human 15q11-13 and mouse central chromosome 7, mouse models for PWS and AS have been described. The PWS mouse model, generated using the IsICt rearrangement, has a maternal duplication (MatDp) and paternal deficiency for central mouse 7 (Fig 1.10). These mice were characterised by post-natal lethality due to failure to thrive, most likely associated with poor suckling activity. This is consistent with the feeding difficulties associated with PWS. These mice also failed to express *Snrpn*, lending further evidence that they represent a PWS mouse model (Cattanach *et al.*, 1992).

The candidate AS mouse model has paternal duplication (PatDp) for proximal and central chromosome 7, in the region centromeric to the T9H translocation breakpoint (PatDp.prox7.T9H) (Fig 1.10). These mice display several features in common with AS including a reduced post-natal growth rate, mild ataxia and an AS-type EEG pattern. The mice are hyperactive, but also display prolonged periods of inactivity. The mice became obese during adulthood, which may be a reflection of the obesity noted in older AS patients (Cattanach *et al.*, 1997). There was some abnormal brain pathology, with a 20% reduction of brain size and marked cortical thinning, which has been noted in AS patients (Jay *et al.*, 1991).

These mouse models have provided a controlled means of analysing the mode of expression of homologous genes from the PWS and AS region in a variety of tissues at different developmental stages (Cattanach *et al.*, 1992; Sutcliffe *et al.*, 1996).

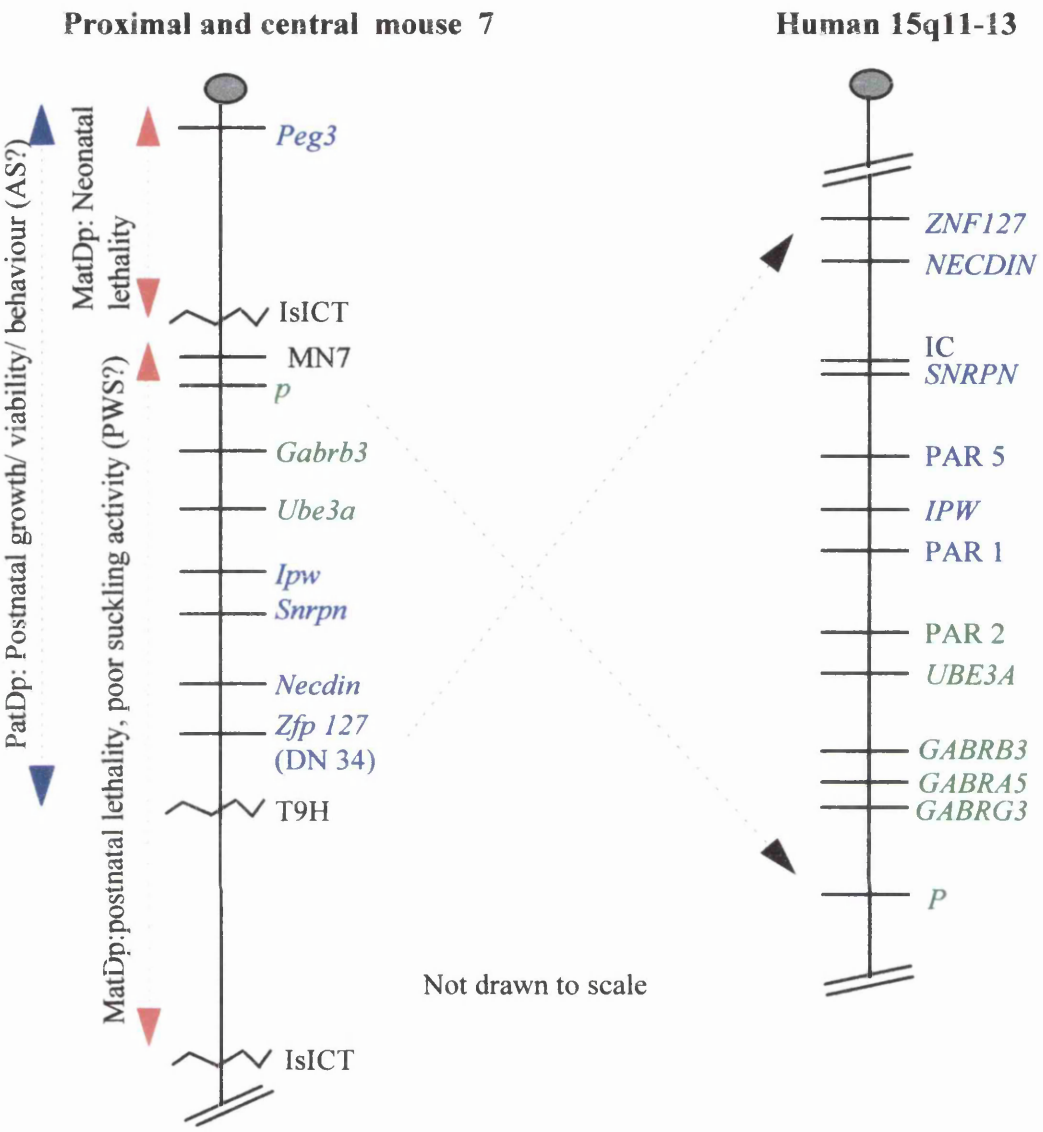


Fig 1.10 Map of the syntenic regions of human 15q11-13 and central mouse 7.

The syntenic regions of central mouse 7 and human 15q11-13 are depicted. The imprinting status and mapping order of the gene homologues is conserved, but they map in the opposite chromosomal orientation. Paternally expressed genes are labeled in blue and biallelically expressed genes in green (accurate to 1997). The regions of proximal and central mouse 7 which show three different imprinted effects, are delineated by dashed arrows. Jagged lines denote the revised T9H and IsICt translocation breakpoints, which define the imprinted region corresponding to the AS and PWS region. All the gene homologues from the PWS and AS region, from *Zpf127* to *p*, lie between the proximal IsICt and T9H breakpoints (Cattanach *et al.*, 1997). The position of MN7 with respect to IsICt has not been reported. A PWS mouse model has been created with MatDp for the region between the Is1Ct breakpoints (Cattanach *et al.*, 1992). This is distinct from the phenotype of mice with MatDp of proximal mouse 7. The AS mouse model has a PatDp of the region proximal to T9H (Cattanach *et al.*, 1997).

1.7.5 Genes mapping to 15q11-13 and their mouse homologues

1.7.5.1 ZNF127 (Zinc finger protein 127)

The differentially methylated locus D15S9, identified by cDNA probe DN34 (Driscoll *et al.*, 1992), encodes two overlapping antisense transcripts, *ZNF127* and *ZNF127AS*. *ZNF127* codes for a zinc-finger protein comprising a RING C_3HC_4 zinc-finger and multiple C_3H finger-like motifs (Jong *et al.*, 1999a). Zinc-finger proteins represent a vast family of proteins which function by interacting with nucleic acids (Reviewed by Pieler and Bellefroid, 1994). *ZNF127* has been proposed to function in gene or RNA regulation. *ZNF127* is intronless and is transcribed from centromere to telomere. The gene is ubiquitously expressed as a 3kb mRNA. The antisense transcript *ZNF127AS* is transcribed from the opposite strand and encompasses the *ZNF127* gene. This is expressed as 7 and 11kb product predominantly in brain and lung (Jong *et al.*, 1999a). The murine homologue of *ZNF127*, *Zfp127* and the antisense transcript, *Zfp127as* have been mapped to central mouse 7 (Jong *et al.*, 1999b).

Functional imprinting of *ZNF127* was demonstrated using PWS and AS fibroblasts, from patients with UPD or deletion (section 1.8.2.1). Absence of expression in PWS cell lines showed that *ZNF127* if expressed solely from the paternal allele (Jong *et al.*, 1999a). *ZNF127* maps 1-1.25 Mb centromeric to *SNRPN*, proximal to the defined PWS candidate region (Fig 1.9). However, silencing of *ZNF127* in the cell lines of PWS imprintor mutation patients indicates that it is under the control of the imprinting centre (section 1.7.5.1) (Saitoh *et al.*, 1996). Thus *ZNF127* may contribute to the PWS phenotype, despite mapping outside the suggested PWS candidate region. Maternal imprinting was also demonstrated for *Zfp127* in the mouse through analysis of a polymorphic in-frame hexanucleotide tandem repeat polymorphism in F1 hybrids. Although paternal expression was observed in several different tissues, a low level of maternal expression was detected in the liver (Jong *et al.*, 1999b).

ZNF127 is completely methylated at the 5' CpG island on the maternal allele in female germ cells (Jong *et al.*, 1999a), and unmethylated in placenta, hydatidiform mole (Mowery-Rushton *et al.*, 1996), and adult sperm (Jong *et al.*, 1999a). Both parental alleles are completely or partially methylated in the majority of fetal tissues. Yet interestingly, studies in AS brain showed that the paternal allele is completely unmethylated, whereas the maternal allele shows the complete and partial methylation pattern (Jong *et al.*, 1999a). Thus the methylation status associated with *ZNF127* varies according to the tissue.

1.7.5.2 SNRPN (Small nuclear ribonucleoprotein-associated polypeptide N)

The *SNRPN* gene encodes the small nuclear ribonucleoprotein-associated polypeptide (snRNP), SmN. snRNPs are immunoprecipitated with anti-Sm antibodies that are present in the sera of patients with the autoimmune disease, systemic lupus erythamatosus. SmN, a 29kDa polypeptide, was identified by immunoblotting various rat tissues with anti-Sm sera (McAllister *et al.*, 1988). The *SNRPN* gene and SmN protein are structurally similar to the *SNRPB* gene and its protein isoforms SmB and SmB' (the latter of which is present in humans but not mouse), that are involved in the nuclear splicing of RNA precursors (Li *et al.*, 1989).

Snrpn is expressed predominantly in the neural cells of the brain and in the heart of rodents (McAllister *et al.*, 1988; Li *et al.*, 1989; Schmauss *et al.*, 1992), but is detectable in most tissues (Barr *et al.*, 1995). This contrasts with *Snrpb*, which is ubiquitously expressed, but with the notable absence of SmB in the brain (Schmauss and Lerner, 1990), where SmN is present at its highest levels. Although the 1.5kb human *SNRPN* transcript is expressed at the highest levels in brain and heart, it is ubiquitously transcribed (Glenn *et al.*, 1996). A role for SmN in neural-specific pre-mRNA processing has been proposed (Schmauss *et al.*, 1992).

SNRPN and its murine homologue were mapped simultaneously to determine whether the gene was associated with any human neurological disorders or mouse mutations. *SNRPN* was initially localised to 15pter→q14 by somatic cell hybrid mapping, then more precisely mapped to 15q11.2-13 by deletion analysis with PWS and AS DNA. An intronless, processed pseudogene *SNRPNP1*, was identified on 6pter→p21 (Özçelik *et al.*, 1992). *SNRPN* falls within the PWS candidate region about 130kb distal to D15S63 (Buiting *et al.*, 1993) (Fig 1.9). Mouse *Snrpn* was mapped by interspecific backcross analysis to the syntenic region on central mouse 7 (Fig 1.10) (Leff *et al.*, 1992).

SNRPN is expressed exclusively from the paternal allele in both the human and mouse. Elevated levels of *Snrpn* were detected in the neural tissues of PatDp.prox7.T9H (AS) mice, but was absent in mice with (MatDp of central chromosome 7 (Cattanach *et al.*, 1992)(section 1.7.4). In an RNAse protection assay using brain RNA derived from reciprocal interspecific mouse crosses, *Snrpn* was expressed monoallelically from the paternal allele, irrespective of the species of the male parent (Leff *et al.*, 1992). Maternal imprinting was shown in human fetal brain, heart and adult cerebral cortex (Reed and Leff, 1994), and in preimplantation embryos (Huntriss *et al.*, 1998), through analysis of expressed polymorphisms (section 1.8.2.2).

SNRPN is transcribed from centromere to telomere and spans approximately 30kb of genomic DNA (Buiting *et al.*, 1993). Initially the gene was thought to consist of 10 exons (Glenn *et al.*, 1996), with the translation initiation codon in exon 4 (Schmauss *et al.*, 1992). However, additional paternally expressed exons have been identified at both the 5' and 3' ends of the gene, which utilise *SNRPN* exons. Several alternatively spliced transcripts termed the BD transcripts, were identified by direct cDNA selection from the region upstream of *SNRPN*, that is deleted in AS imprintor mutation cases. The transcripts differed at the very 5' end, containing either exons BD1A, BD1B' or BD1B*, but all contained exons BD2 and BD3. These transcripts utilised *SNRPN* exons 2-10 at their 3' end. Thus the BD transcripts represent alternative 5' exons of *SNRPN,* which skip the previously identified transcription start site at exon 1, without interrupting the reading frame. Two differentially methylated sites have been identified upstream of BD1A and BD1B, with the maternal allele hypermethylated. BD transcription occurs solely from the paternal allele, at low levels, predominantly in brain, heart, ovary and testis. Exons KB-1, KB2 and KB3/PAR-SN map 2kb distal to *SNRPN.* KB-1 is contiguous with *SNRPN* exon 10, and RT-PCR products and cDNA clones have been identified which contain 3' exons of *SNRPN* and exons KB-1 and 2. However, the KB exons identify a 3.4kb transcript in brain, heart, skeletal muscle, ovary and testis, whereas the *SNRPN* coding exons identify a is 1.8kb mRNA (Ning *et al.*, 1996; Buiting *et al.*, 1997). An RNA precursor consisting of *SNRPN* spliced on to KB/PAR-SN, could form and subsequently undergo endonucleolytic cleavage to release KB/PAR-SN from *SNRPN*, resulting in distinct mRNA transcripts (Buiting *et al.*, 1997). Accordingly, exons KB-1, 2 and KB3/PAR-SN have been renamed SNRPN exons 10a, 11 and 12 respectively (Hosthemke *et al.*, 1997).

SNRPN expression is absent in the lymphoblasts and fibroblasts of PWS patients with deletions, UPD and imprintor mutations (Glenn *et al.*, 1993a; Nakao *et al.*, 1994), indicating that it is a candidate for involvement in the PWS phenotype. However its pathophysiological role has not been delineated, and the precise function of SmN remains elusive. It is possible that alternative splicing of neural-specific transcripts involved in brain development does not occur in PWS individuals, although mice lacking *Snrpn* have apparently normal regulation of neuronal splicing events (Huntriss *et al*, 1994).

1.7.5.3 NDN (Necdin)

Mouse necdin (neurally differentiated EC cell derived factor) is a nuclear protein that is expressed specifically in post-mitotic, differentiated neurons during neurogenesis, from the early stages of development until adulthood. Necdin was found to suppress cell proliferation, hence is thought to be involved in the process of neuronal quiescence (Maruyama *et al.*, 1991; Uetsuki *et al.*, 1996).

The 1.7kb murine Necdin cDNA encoding a 350 amino acid protein, was isolated from a subtraction cDNA library from embroyonal carcinoma cells, treated for the induction of neural differentiation of stem cells (Maruyama *et al.*, 1991). Assignment of an EST representing the 3' end of a clone referred to as 'human Necdin-related mRNA' to the 15pter-D15S156 Whitehead Institute radiation hybrid, led to the isolation of the human homologue *NDN*, and its localisation to the PWS region (MacDonald and Wevrick, 1997). The murine and human homologues display 82% amino acid identity and are contained within one exon (Uetsuki *et al.*, 1996; MacDonald and Wevrick, 1997). *NDN* maps about 100kb distal to *ZNF127* and ~1Mb proximal to *SNRPN* (Jay *et al.*, 1997). *Ndn* was localised to the syntenic in the mouse by interspecific backcross mapping (Fig 1.10) (Macdonald and Wevrick, 1997; Watrin *et al.*, 1997).

In contrast to its murine homologue, *NDN* is ubiquitously expressed in adult and fetal tissues, but is absent in lymphoblasts. Immunohistochemical analyses show that the cell types expressing the highest levels of human NECDIN, were differentiated neurons, such as dorsal root ganglia. *NDN* and *Ndn* may therefore perform a parallel function in human and mouse; that of development of the central nervous system (CNS) (Jay *et al.*, 1997).

Both human and mouse homologues are monoallelically expressed from the paternal allele, and are associated with methylation of the repressed maternal allele. Imprinted expression was demonstrated in the mouse CNS using RNA derived from the progeny of reciprocal interspecific crosses, and in normal human fibroblast and amniocyte RNA through analysis of an expressed polymorphism (MacDonald and Wevrick, 1997; Jay *et al.*, 1997; Watrin *et al.*, 1997). In addition, there was complete loss of *NDN* expression in PWS fibroblasts and brain, although normal expression was observed in corresponding AS samples (Watrin *et al.*, 1997). A role for NECDIN in the pathophysiology of PWS, due to abnormal brain development, has been suggested (Macdonald and Wevrick, 1997; Jay *et al.*, 1997).

1.7.5.4 IPW (Imprinted gene in the PWS region)

IPW was identified by direct cDNA selection of transcripts from lymphoblast, fibroblast and human embryonic kidney cell lines, using YACs from the PWS candidate region (Wevrick *et al.*, 1994). *IPW* is a 2.2kb transcript consisting of 3 exons, and maps about 150kb distal to *SNRPN*. The murine homologue, *Ipw* was identified by screening a mouse genomic library with the highly conserved final exon of human *IPW*. *Ipw* was localised to mouse chromosome 7, using a mouse x rodent somatic cell hybrid panel and *in situ* hybridisation on metaphase spreads (Wevrick and Francke, 1997). *IPW* is widely expressed in adult and fetal tissues and functions in a housekeeping capacity during cell growth and differentiation (Wevrick *et al.*, 1994; Rachmilewitz *et al.*, 1996). In contrast *Ipw* is only expressed substantively in the brain (Wevrick and Francke, 1997).

Monoallelic paternal expression of *IPW* was demonstrated in lymphoblasts, fibroblasts and a variety of fetal tissues, heterozygous for an intrageneic polymorphism (section 1.8.2.2). Expression of *IPW* is not detectable in the cell lines of PWS patients (Wevrick *et al.*, 1994). Maternal imprinting has also been demonstrated for the mouse homologue (Wevrick and Francke, 1997).

Both human and murine homologues have no substantial open reading frame and are thought to play their role as an RNA molecule (Wevrick *et al.*, 1994; Wevrick and Francke, 1997). Thus *IPW* is one of just 3 endogenous mammalian genes that transcribe non-coding RNAs, including *XIST* and *H19* (Erdmann *et al.*, 1999). Both of these genes are monoallelically expressed and are involved in the *in cis* regulation of genes in the region to which they map. *XIST* is involved in the initiation and maintenance of X chromosome inactivation in female mammalian cells (reviewed by Heard *et al.*, 1997). *H19* expression influences the expression of other imprinted genes within the imprinted cluster on 11p15.5 (reviewed by Viville and Surani, 1995; Leighton *et al.*, 1996). However no such role has been identified for *IPW* within 15q11-13, and the precise function of this gene needs to be elucidated before any contribution to the PWS phenotype is discerned (Wevrick *et al.*, 1994).

1.7.5.5 E6-AP (E6-associated protein) / UBE3A (Ubiquitin protein-ligase E3A)

E6-AP was identified as a cellular factor, which mediates the association of the E6 oncoprotein of the high-risk human papilloma viruses (HPVs) 16 and 18 with p53 and its subsequent ubiquitin-dependant degradation *in vitro*. Degradation of p53 in HPV-transformed cells leads to cell proliferation (Huibregtse *et al.*, 1991). The p53 protein is a

growth and tumour suppressor protein that is a key regulator of mammalian cell growth. Mutations of p53 have been associated with various human tumours (Holstein *et al.*, 1991).

The ubiquitin proteolytic pathway that exists in eukaryotes is responsible for the selective turnover of intracellular proteins under basal metabolic conditions. In this system proteins are degraded by the proteasome complex after conjugation with a multitude of ubiquitin molecules. This requires the successive action of three classes of enzymes: The ubiquitin activating enzyme, E1, activates ubiquitin in an ATP-dependent step and transfers it to the ubiquitin-conjugating enzyme, E2. Ubiquitination of the target protein then occurs, sometimes in conjunction with a ubiquitin protein ligase, E3. The E3 enzymes, where required, recognise the substrate and catalyse the transfer of ubiquitin from the E2 to the substrate. Ubiqutin is transferred sequentially from E1 to E2 to E3 and finally, to the target protein by the formation of thioester bonds (Scheffner *et al.*, 1995; Reviewed by Ciechenover, 1994). The E2 enzyme that specifically ubiquitinates E6-AP is UBC4. E6-AP and E6 together function as an E3 protein ligase that mediates the binding and stimulates the ubiquitination of p53 in HPV immortalised cells (Scheffner *et al.*, 1993; 1995). However, E6-AP was found not to recognise p53 in the absence of the HPV 16 or 18 E6 protein (Huibregtse *et al.*, 1993).

A group of 14 *hect* proteins, so named due to their homology to the E6-AP carboxyl terminus, have been identified in *S. cerevisiae*, *Drosophila*, rat, mouse and human. These are predicted to be a class of structurally and functionally related E3 ubiquitin protein ligases on the basis that the last 100 amino acids of the proteins, constituting the *hect* domain, are highly conserved. Two of the proteins share the ability of E6-AP to form a thioester with ubiquitin. A particular cysteine residue at amino acid 833 of E6-AP is conserved among all *hect* proteins, lying 32-36 amino acids away from the protein termini. This was found to be essential for ubiquitin thioester formation, and hence the ability of E6-AP to target p53. The final 6 amino acids of E6-AP are required for transfer of ubiquitin to p53 (Huibregtse *et al.*, 1995).

The gene encoding E6-AP is *UBE3A*. The cDNA was isolated by screening a human keratinocyte cDNA library with an RT-PCR probe, designed using degenerate primers predicted from the amino acid sequence (Huibregtse *et al.*, 1993). ESTs with sequence identity to *UBE3A* were subsequently isolated from a human fetal brain library, whilst using evolutionary conserved genomic fragments from 15q11-13 as probes. PCR amplification of a

short fragment from within exon 3, localised *UBE3A* to YACs 132D4 and B230E3 from the AS region (Fig 1.8) (Nakao *et al.*, 1994).

UBE3A identifies an approximately 5kb transcript that is predominant in skeletal muscle, heart and liver, but is expressed in all the major organs and tissues including brain, lung, placenta and kidney (Nakao *et al.*, 1994). However, the coding region comprises 2.6kb and the 3' untranslated region (3'UTR) prior to the polyadenylation signal is <50bp (Huibregtse *et al.*, 1993). Exon trapping from a PAC telomeric to *UBE3A* led to the identification of additional 5' untranslated exons of the gene, U1, U2, U3, and U4 (Kishino *et al.*, 1997). RT-PCR analysis of the 5' end of the cDNA in various cell lines, identified 5 distinct subtypes of *UBE3A*, formed by differential splicing of the U exons. The 5 subtypes had the potential to encode 3 different protein isoforms, with putative translation initiation codons appearing in exons 1, U3 and U2 in E6-AP isoforms I, II and III respectively (Yamamoto *et al.*, 1997). The genomic structure of *UBE3A* has been defined. The *hect* domain is encoded by exons 3 to 10. The nucleotide sequence coding for the E6-binding site has been identified in exon 3, and the conserved cystein residue is encoded within exon 10. The polyadenylation signal appears in exon 10, downstream of the termination codon (Yamamoto *et al.*, 1997). Mapping near D15S10 (Fig 1.8), was a genomic clone OP2, which contained a CpG island and identified a ~4.5kb transcript in multiple tissues, with expression in skeletal muscle threefold that of other tissues (Woodage *et al.*, 1994b). OP2 was subsequently shown to contain the first exon of *UBE3A* within the CpG island (Kishino *et al.*, 1997).

UBE3A was shown to be expressed from both parental alleles in lymphoblasts and fibroblasts from PWS and AS patients by RT-PCR (section 1.8.2.1) (Nakao *et al.*, 1994). The murine homologue on central mouse 7, *Ube3a* was analysed in a multitude of different tissues in MatDp.prox7.T9H and PatDp.prox7.T9H mice to determine whether *UBE3A* represented a potential candidate for AS. However, *Ube3a* found to be expressed from both parental alleles in mice, making *UBE3A* an unlikely candidate for AS (Sutcliffe *et al.*, 1996; 1997).

1.7.5.6 GABRB3, GABRA5 and GABRG3 ($GABA_A$ receptor subunits β3, α5 and γ3)

The $GABA_A$ receptors are a family of ligand-gated chloride channels activated by GABA, and constitute the principle inhibitory neurotransmitter receptors in the mammalian central nervous system. A role for the $GABA_A$ receptor in the suppression of seizure activity has been indicated (reviewed by DeLorey and Olsen, 1992).

The $GABA_A$ receptor subunit β3 (*GABRB3*) was the first gene to be localised to 15q11-13 through FISH and deletion analysis of PWS and AS patients. A role for the GABRβ3 receptor subunit in epilepsy and movement control disorders featured in AS was suggested (Wagstaff *et al.*, 1991a; Saitoh *et al.*, 1992), and the distal breakpoint of the submicroscopic deletion in AS Family S was located to intron 3 of *GABRB3* (Greger *et al.*, 1993). However, this gene has been excluded from the AS candidate region (Michaelis *et al.*, 1995).

Murine *Gabrb3* was mapped to chromosome 7 by hybridisation to mouse x hamster somatic cell hybrids. Analysis of RFLVs localised the gene to the *p* region (Wagstaff *et al.*, 1991b). *Gabrb3* expression was detected in the brain of mouse mutants with reciprocal UPD for central chromosome 7 (section 1.7.4), indicating that the gene was expressed from both parental alleles (Cattanach *et al.*, 1992). The transcript was similarly detectable in mouse strains with maternal and paternally derived deletions (Nicholls *et al.*, 1993).

The $GABA_A$ receptor α5 (*GABRA5*) has been mapped 100kb distal to *GABRB3* (Knoll *et al.*, 1993; Sinnett *et al.*, 1993) and γ3 (*GABRG3*) further telomeric (Greger *et al.*, 1995). This cluster of genes have been shown to be expressed from both parental alleles using human monochromosomal x rodent hybrids (section 1.8.2.3)(Gabriel *et al.*, 1998). Tissue-restricted expression and lack of expressed polymorphisms in the $GABA_A$ receptor genes has prevented the direct confirmation of expression status in humans (Meguro *et al.*, 1997).

1.7.5.7 The P/p gene

The murine *p* transcript was identified by analysis of the hypopigmented p^{un} mouse mutant. The human homologue *P*, was isolated from a human melanoma cDNA library. *P* was found to correspond to the cDNA DN10, a fragment of probe IR-10, which identified D15S12 (Gardner *et al.*, 1992). PWS patients with hemizygous deletions for D15S12 showed a high incidence of hypopigmentation when compared with patients intact at this locus

(Hamabe *et al.*, 1991a), indicating that this gene may contribute to the hipopigmentation observed in deletion-positive PWS and AS patients. Human and murine P/*p* are expressed in primary melanocytes in the skin and eyes (Gardner *et al.*, 1992; Rinchik *et al.*, 1993). The amino acid sequence of *P* predicted a transmembrane protein with some sequence homology to an *E. coli* tyrosine-specific transporter. The P protein is thought to be a component of the melanosome membranes, and responsible for the transport of tyrosine, the primary precursor in melanin production (Rinchik *et al.*, 1993).

The autosomal recessive disorder Type II (tyrosinase positive) oculocutaneous albinism (OCA II), was linked to D15S10 and D15S13 in a South-African subpopulation, in which the disorder is prevalent (Ramsay *et al.*, 1992). Mutations in *P* were subsequently identified in patients with OCA II and ocular albinism (Rinchik *et al.*, 1993; Lee *et al.*, 1994). OCA II in two PWS patients was found to be due to hemizygosity for a mutant maternally derived *P* allele in one, and a microdeletion of the maternal allele in another (Rinchik *et al.*, 1993; Lee *et al.*, 1994). Haploinsufficiency combined with nonpathologic polymorphisms of *P* are thought to account for the hypopigmentation that manifests in PWS and AS deletion patients (Lee *et al.*, 1994).

1.7.5.8 The PAR transcripts (PWS and AS region transcripts)

Several uncharacterised transcripts, PARs 1, 2, 4, 5, 6 and 7 were identified by screening a human fetal brain cDNA library with genomic fragments that had displayed evolutionary conservation with mouse (Nakao *et al.*, 1994; Sutcliffe *et al.*, 1994). PAR-5 maps 7kb telomeric to *SNRPN* and identifies a 12kb transcript in skeletal muscle (Ning *et al.*, 1996). PAR-7 maps just distal and identifies distinct transcripts of 2.8 and 8.5kb in skeletal muscle and kidney. PAR-1 maps proximal to D15S174, the boundary between the PWS and AS region, and displays a diffuse northern hybridisation signal from 8-14kb in brain. PAR-4 maps distal to D15S174 and identifies 3 and 5kb transcripts in most tissues and additional transcripts of 8kb in muscle and 2kb in pancreas (Sutcliffe *et al.*, 1994). PAR-2 is located centromeric to *UBE3A* and is expressed highly and equally in most tissues tested by RT-PCR (Nakao *et al.*, 1994; Sutcliffe *et al.*, 1994). PARs 1 and 5 are absent in PWS cell lines but detectable in AS cell lines, indicative of paternal expression (section 1.8.2.1). The imprinting status of PARs 4 and 7 could not be determined due to low expression levels (Sutcliffe *et al.*, 1994). PAR-2 is biallelically expressed (Nakao *et al.*, 1994). All the PARs are colinear with genomic DNA and contain no open reading frame (Sutcliffe *et al.*, 1994).

al., 1993). It has been shown that the methylation status of this intronic sequence represents the 'gametic mark' for *Igf2r,* and determines imprinted expression (Wutz *et al.*, 1997).

The importance of differential methylation in maintaining the imprinted status of genes is well documented, but is best illustrated by the effect of erroneous imprinting within 11p15.5 and 15q11-13. The expression and imprinting of *H19* and *IGF2* has been shown to be dependent on the methylation status of the *H19* promoter. Methylation of *H19* on the normally unmethylated maternal allele inactivates *H19*, and results in biallelic expression of *IGF2*. This loss of imprinting has been observed in Wilm's tumour of the kidney (Moulton *et al.*, 1994; Steenman *et al.*, 1994), and in BWS patients who have biallelic inheritance of 11p15.5, but whose maternal allele carries the paternal methylation imprint (Reik *et al.*, 1995). Repression of the paternally expressed genes *ZNF127*, *SNRPN* and *IPW* within 15q11-13 has been demonstrated in PWS imprintor mutation patients, whose paternal alleles bear the maternal methylation imprint. These genes were found to be biallelically expressed in AS patients with imprintor mutations, indicating that lack of methylation resulted in transcription from the normally silent, maternal allele (Saitoh *et al.*, 1996).

1.7.6.1 The 15q11-13 imprinting center

Different models have been proposed to explain the mechanism by which the imprinting center (IC) within 15q11-13 switches the epigenotype of genes within the region in the germ line, to reflect the sex of the parent (Saitoh *et al.*, 1996; Ferguson-Smith, 1996). PWS imprintor mutation lineages have deletions that always include *SNRPN* exon 1 (Sutcliffe *et al.*, 1994; Buiting *et al.*, 1995; Saitoh *et al.*, 1996). In the majority of AS imprintor families exon BD3 is deleted (section 1.7.5.2). BD1A and BD2 were additionally deleted in some families (Dittrich *et al.*, 1996). The BD transcripts have been renamed the IC transcripts, and the exons are now prefixed with 'IC' instead of 'BD' (Horsthemke *et al.*, 1996).

Since the deletions involved in PWS and AS imprintor families are distinct, a model was put forward in which the IC featured a bipartite structure; an imprintor and an imprint switch initiation site. The imprintor, representing the IC transcripts, acts in *cis* on the switch initiation site at *SNRPN* exon 1 or the CpG island, from where the imprint spreads bidirectionally across the region. This might confer the appropriate allele-specific epigenotype, or cause a structural change in the chromatin. In the female germ line, the IC transcript would be expressed from the paternally inherited chromosome. This would enable

a hypothetical oocyte-specific *trans*-acting factor to interact with the switch initiation site, and the maternal imprint would then be conferred. In the male germ line, the trans-acting factor is absent, but erasure of the maternal imprint would occur from the switch initiation site by default (Fig 1.11A) (Dittrich *et al.*, 1996).

An alternative model to explain the action of the IC was proposed, which took into account the global erasure of the methylation imprints of imprinted genes in both male and female germ lines, prior to re-establishment. In this model the switch initiation site is responsible for erasure of the imprint. In the paternal germ line a deletion of *SNRPN* exon 1 would result in retention of the maternal imprint due to a failure of imprint erasure (demethylation). Expression of the IC transcript would occur in the maternal germ line following imprint erasure, and this would incur setting of the maternal imprint. This model would not require the action of a putative *trans*-acting factor in the female germ line (Fig 1.11B) (Ferguson-Smith, 1996). This theory is consistent with the demonstration that *Snrpn* is expressed biallelically in the female germ line in mice (Kono *et al.*, 1996). However, transcription of the IC exons would have to be limited to the maternal germ line if, as this model predicts, they are responsible for invoking the female imprint (Kelsey and Reik, 1997). IC expression has been demonstrated in testis (Dittrich *et al.*, 1996). Therefore, both models are subject to question and require direct testing (Kelsey and Reik, 1997).

A PWS mouse model for imprinting centre mutations has recently been developed with a paternally derived deletion of the IC region, including the promoter and exons 1-7 of *Snrpn*. These mice died neonatally due to inadequate feeding; a similar phenotype to the PWS mouse model with maternal disomy for central mouse 7. Expression of *Zpf127*, *Ndn* and *Ipw* was markedly reduced, as is the case for human PWS imprintor mutation cases. In contrast, mice with intragenic deletions of *Snrpn* exons 5-7 were phenotypically normal, and expression of the paternally expressed genes was unaffected. The IC mouse model will be useful for studying the regulation of imprinted genes in this syntenic region (Yang *et al.*, 1998).

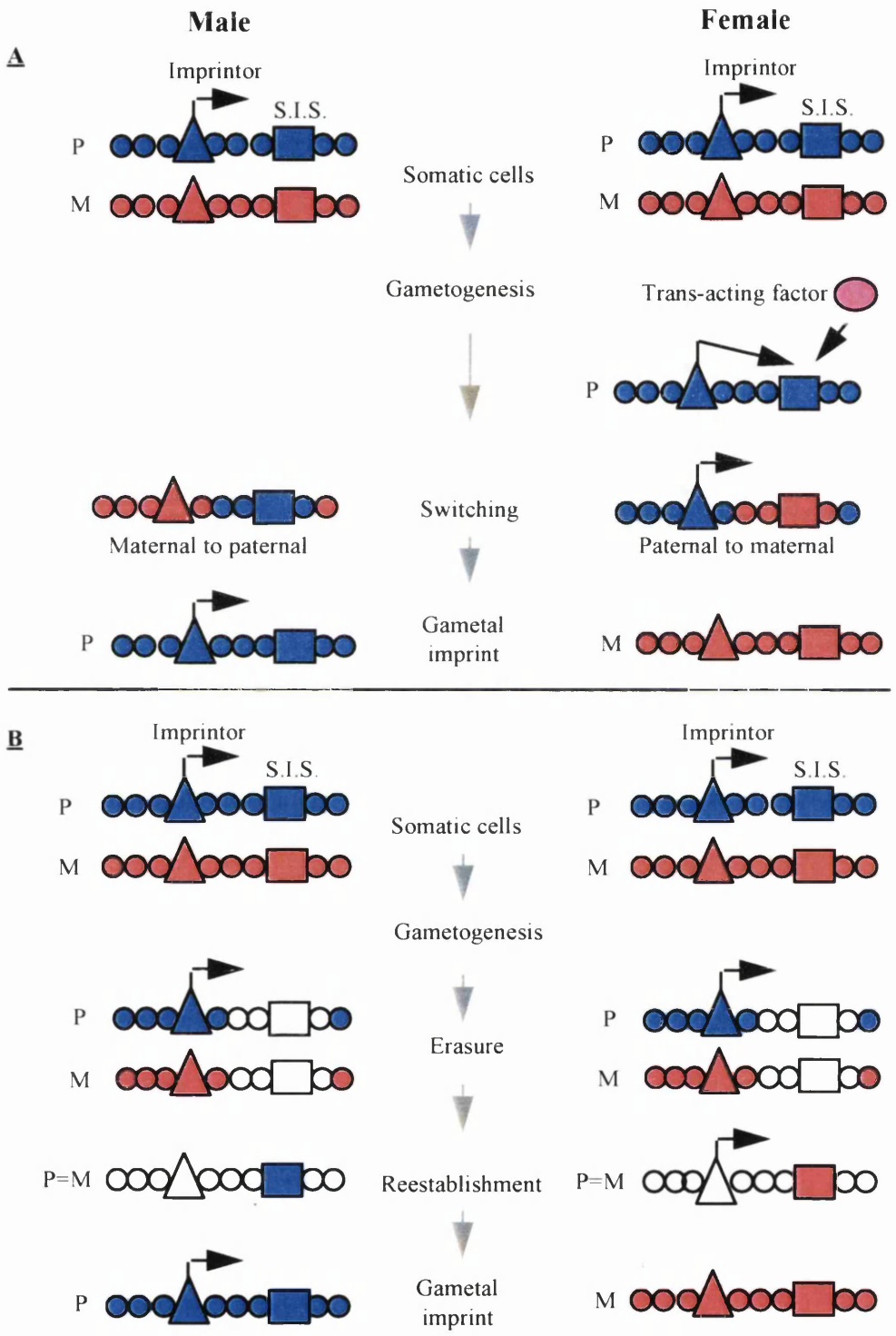


Fig 1.11 Imprint switch models to explain the action of the 15q11-13 imprinting center.

A: Model of Dittrich *et al.*, 1996. B: Model of Ferguson-Smith, 1996.
The paternal allele or methylation imprint is in blue, the maternal imprint in red. Erased methylation imprints are in white. The IC transcripts or imprintor are represented by a triangle. An arrow denotes transcription. The *SNRPN* CpG island or switch initiation site (S.I.S.) is represented by a square. Details of the models are given in the text.

1.8 Identification and analysis of imprinted genes

1.8.1 Identification of imprinted genes

1.8.1.1 Association of phenotype with parental transmission of mutant alleles in mice

Imprinted genes in mice may be identified through the observation of a phenotype that is associated with a given mutation, with a parent-specific mode of inheritance. Imprinting of *Igf2* was revealed by the correlation of a growth deficiency phenotype with paternal transmission of the mutant allele. The phenotype was indistinguishable between heterozygous mice with a paternally inherited mutation, and mice homozygous for the mutation. Heterozygous progeny were normal when the mutation was maternally inherited. *In situ* hybridisation studies and RNase protection assays were used to confirm that *Igf2* expression occurred from the paternal allele (De Chiara *et al.*, 1991).

The *Mash2* gene, which is essential for trophoblast development in the mouse, was identified as paternally imprinted through transmission studies of a null mutation. Embryos heterozygous for the mutation were viable only if the mutation was paternally inherited. If the mutation was of maternal origin, the same embryonic lethal phenotype was observed as in homozygous mutants. *In situ* hybridisation studies confirmed that *Mash2* was maternally expressed in trophoblasts at a crucial stage of development (Guillemot *et al.*, 1995).

1.8.1.2 Identification of differentially methylated loci

Differential methylation has been exploited as a means for identifying new imprinted genes based on the observation of different methylation patterns between the two parental alleles, using methylation sensitive restriction enzymes. The Southern analysis strategy devised by Driscoll *et al.* to screen for sites of allele-specific methylation within 15q11-13 (section 1.4.5.1) has been adopted for other chromosomes. The *Hpa*II methylation patterns detected by expressed sequence tags (ESTs) have been compared between SRS patients with mUPD7 and a case of pUPD7 with normal growth (section 1.3.2.3), to screen for imprinted genes that are involved in IUGR, a major characteristic of SRS. Several ESTs identified differentially methylated sites and are currently being assayed for imprinted expression (Hannula *et al.*, 1998).

The methylation status of novel transcripts identified from the PWS and AS regions in this project was analysed using PWS and AS patients with UPD and hemizygous deletions by Southern analysis with *Hpa*II.

Restriction landmark genomic scanning (RLGS) is a technique where restriction-digested genomic DNA is radiolabelled and electrophoresed in two dimensions to produce an RLGS profile on autoradiography. RLGS with methylation sensitive enzymes has been used to screen the mouse genome for endogenous loci showing imprinted patterns of methylation. The RLGS profiles for F_1 hybrids of reciprocal interspecific crosses were compared, and several landmark variants have been identified with imprinted patterns of transmission (Shibata *et al.*, 1994). Molecular analyses of the imprinted landmarks identified the paternally imprinted gene *U2afbp-rs* within an imprinted region on mouse chromosome 11 (Hayashizaki *et al.*, 1994) and the maternally imprinted gene *Grf1* on chromosome 9 (Plass *et al.*, 1996).

1.8.1.3 Subtraction hybridisation

Paternally expressed imprinted genes (*Pegs*) have been identified by subtraction hybridisation of cDNAs from normal mouse embryos, with cDNAs derived from parthenogenetic embryos of the equivalent developmental stage. In this method biallelic and maternally expressed transcripts are removed from the normal embryo cDNA pool, through competitive hybridisation, leaving a population of paternally expressed genes. This lead to the identification of four *Pegs*, two of which represented the previously identified imprinted genes, *Snrpn* and *Igf2*. *Pegs 1* and *3* represented novel cDNAs. *Peg1*, or the 'mesoderm-specific' transcript *Mest*, maps to mouse chromosome 6 (Kaneko-Ishino *et al.*, 1995). *Peg3* encodes a zinc-finger protein and maps to proximal mouse 7, a region for which maternal duplication/ paternal deficiency is associated with neonatal lethality (Fig 1.10) (Kuroiwa *et al.*, 1996).

Subtraction hybridisation of cDNAs from normal mouse embryos with cDNAs from androgenetic embryos led to identification of several maternally expressed transcripts (*Megs*). *Meg1* was identified as the growth factor receptor-bound protein 10 transcript, *Grb10*, a growth inhibitor that modulates the insulin and insulin-like growth factor pathways (Miyoshi *et al.*, 1998). *Grb10* and the human homologue *GRB10,* map to syntenic regions of proximal mouse 11 and human 7, that are involved in growth-related phenotypes (Jerome *et al.*, 1997; Miyoshi *et al.*, 1998).

1.8.2 Assays for establishing imprinted expression

1.8.2.1 Use of UPD or hemizygous deletions

Analogous to the use of UPD in mouse models to demonstrate parent-specific expression, cultured lymphoblasts and fibroblasts from AS and PWS cases with UPD or hemizygous deletions have been used study the expression of genes mapping to human 15q11-13. Detection of expression in one patient cell line, but not in the other is indicative of imprinting, provided expression of a housekeeping control gene such as *GAPDH* or *HPRT* is demonstrated by RT-PCR. This assay has been used to demonstrate or confirm paternal expression / maternal imprinting of several genes within 15q11-13, and showed that *UBE3A* was biallelically expressed in cultured cell lines (section 1.7.5).

1.8.2.2. Intragenic polymorphisms

Imprinting studies in humans and mice often make use of intragenic polymorphisms which are present in the mRNA, to investigate the mode of expression. Monoallelic expression is demonstrated when a single allele is observed in the RNA of a heterozygous sample. Where parental DNA is available, the allelic origin of expression can also be determined. Where the polymorphisms involves a restriction site, imprinting analysis involves RT-PCR followed by digestion of the product. This strategy has used to demonstrate imprinting of several genes including *SNRPN* (Reed and Leff, 1994) and *H19* (Zhang and Tycko, 1992) in human fetal tissues. This system was also used to demonstrate imprinting of *IGF2* and *SNRPN* in preimplantation embryos. Spare embryos from couples undergoing *in vitro* fertilisation treatment, who were homozygous for opposite alleles of a restriction polymorphism, were analysed. Maternal imprinting was demonstrated for *IGF2* in 8-cell embryos (Lighten *et al.*, 1997), and for *SNRPN* in embryos at the 4-cell stage (Huntriss *et al.*, 1998).

In mice, intragenic polymorphisms are studied in the heterozygous F_1 hybrids of reciprocal interspecific crosses. Where monoallelic expression is observed, the reciprocal cross is used to demonstrate that expression is dependent on parental allele, irrespective of the species of the parent. *Mest* (Kaneko-Ishino *et al.*, 1995), *Snrpn* (Leff *et al.*, 1992) and *Ndn* (MacDonald and Wevrick, 1997) were shown to be imprinted in mouse embryos using this technique.

1.8.2.3 In vitro imprinting assay with human x mouse hybrid cell lines

Microcell-mediated transfer has been used to create rodent A9 hybrid cells containing a single intact human chromosome derived from fibroblasts, for use in mapping and expression analyses (Ning *et al.*, 1992). Somatic cell hybrids containing a single human chromosome 15 (Meguro *et al.*, 1997; Gabriel *et al.*, 1998), and a single human chromosome 11 (Gabriel *et al.*, 1998) of known parental origin, have been produced. Analysis of microsatellite markers of the monochromosomal hybrids, alongside the parental DNA of the fibroblast donors, enabled the parental origin of the human chromosome to be determined for each hybrid. These cell lines were assayed for imprinted expression of known imprinted loci within human chromosomes 11 and 15, and for differential methylation using Southern analysis. *SNRPN, NDN* and *IPW* were expression exclusively from the clones containing the paternal 15; no expression was detected in those containing maternally derived chromosomes 15. The methylation imprint at *SNRPN* exon 1 and intron 9 was maintained in the hybrids containing single human chromosomes 15 (Meguro *et al.*, 1997; Gabriel *et al.*, 1998). *GABRB3, GABRA5, GABRG3* (Gabriel *et al.*, 1998) and *UBE3A* (Meguro *et al.*, 1997) were expressed in both sets of hybrids, consistent with biallelic expression in fibroblast-derived cell lines. Expression of *IGF2* was limited to the hybrids containing a paternal chromosome 11, and *H19* was detected only in those containing a maternally derived 11. Although some of the hybrids were found to contain more than one human chromosome, this did not affect the results obtained for chromosomes 11 or 15, provided the hybrids contained only one homologue for these chromosomes. These somatic cell hybrids therefore represent an efficient model for investigating the imprinting status of candidate imprinted loci on chromosomes 11 and 15 (Gabriel *et al.*, 1998).

This *in vitro* assay system was used in this project to determine whether novel cDNAs mapping to 15q12 were imprinted.

1.9 Positional cloning for the identification of disease genes

Prior to the mid 1980s, genes responsible for disease when mutated, were identified on the basis of fundamental knowledge about the biochemical defect. This process, referred to as functional cloning, began with the purification of the normal protein product followed by partial amino acid sequencing, as a prerequisite to cloning the gene (reviewed by Collins, 1992). For many genetic defects however, the lack of information on the underlying biochemical defect renders this route to gene identification inapplicable, as is the case for AS and PWS. Disease-causing genes may be identified purely on the basis of chromosomal localisation, without any prior knowledge of function, by a strategy referred to as positional cloning. Ultimately, verification that the correct gene has been identified comes with the demonstration of appropriate expression profiles, and mutations in the gene which cause loss or alteration of the gene product (reviewed by Collins, 1995).

1.9.1 Physical mapping

The first step towards positional cloning of a gene is assignment to a chromosomal region. The observation of chromosomal alterations including cytogenetic deletions, duplications, translocations and triplet repeat expansion in affected individuals, provides low-resolution mapping of the disease locus. In the absence of any obvious chromosomal aberration, disease genes may be mapped through linkage analysis using pedigrees in which the disease is segregating. Analysis of multiple polymorphic markers is carried out until linkage to a particular chromosomal locus is established (reviewed by Collins, 1995).

Radiation hybrids containing a human sub-chromosomal fragment in a rodent background have served as useful tools for both physical and genetic mapping. These are generated by irradiation of chromosomes to break them into fragments, and their subsequent recovery as human x rodent somatic cell hybrid clones by cell fusion. The genetic distance between two markers may be determined statistically by establishing whether they remain on the same sub-chromosomal fragment after a given dose of radiation, or are separated by the breakage (Cox *et al.*, 1990).

Pulse field gel electrophoresis (PFGE) allows the separation of DNA fragments of up to 1Mb, enabling long-range restriction mapping to be performed (Schwartz and Cantor,

1984). The physical distances between loci identified by molecular probes in the PWS/AS region were estimated by PFGE of genomic DNA digested with rare-cutting restriction enzymes (Kirkilionis *et al.*, 1991).

For high-resolution physical mapping and gene identification, a candidate region containing disease genes is cloned into a vector system to form an overlapping contiguous series of genomic clones. Yeast artificial chromosomes (YACs), which consist of a vector containing an exogenous segment of DNA of 0.2 to 1Mb long, have provided the means of doing this (Burke *et al.*, 1987; Schlessinger, 1990). The entire human genome has been cloned, with at least ten-fold coverage, into YAC libraries. YAC contigs spanning the entire human genome are near completion (Cohen *et al.*, 1993; Chumakov *et al.*, 1995). Smaller vector systems containing 50-150kb genomic DNA, including phage, cosmids, P1 artificial chromosomes (PACs) and bacterial artificial chromosomes (BACs), are used for more intricate physical mapping.

1.9.2 Identification of coding sequences

Most eukaryotic genes consist of short coding sequences or exons, separated by intervening introns (reviewed by Breathnach and Chambon, 1981). Littered throughout the genome are repetitive sequences belonging to families of short interspersed elements (SINEs) and long interspersed elements (LINEs). These are present in copy numbers of 10^4 to over 10^5 per haploid genome, and represent over 10% of human DNA. The Alu family of SINEs are specific to primates, although Alu-like sequences are present in rodents and other animals (reviewed by Schmid and Jelinek, 1982; Kass and Batzer, 1995). Coding sequences have been estimated to comprise a mere 3% of the human genome (reviewed by Adams *et al.*, 1991), and identifying coding regions within large genomic regions is technically demanding and laborious. Several strategies have been devised to facilitate this process including traditionally, the identification of DNA that is conserved across evolution and more recently, CpG island rescue, exon amplification and direct cDNA selection (reviewed by Brennan and Hochgeschwender, 1995; Parimoo *et al.*, 1995).

1.9.2.1 Evolutionary conservation

DNA encoding biologically significant proteins will have altered little during the course of evolution if the protein is to retain its function, and so coding regions are often conserved between species. In contrast, intervening DNA sequences may have changed considerably. Conventionally, genes were identified by isolating genomic fragments demonstrating evolutionary conservation by zoo analysis (hybridisation to Southern blots of genomic DNA from diverse animal species). These fragments could be employed as probes for northern analysis and screening of cDNA libraries, or were directly sequenced (reviewed by Parimoo *et al.*, 1995).

This strategy was successfully employed in the identification of disease genes including the cystic fibrosis gene (Rommens *et al.*,1989), the adrenoleukodystrophy gene (Mosser *et al.*, 1993) and candidate cDNAs for Duchenne muscular dystrophy (Monaco *et al.*, 1986). However, this method can be extremely laborious; in work leading to the identification of the DCC gene altered in colorectal cancers, 117 *EcoR*I fragments from a 370kb deletion region were isolated for zoo-blotting, of which 24 exhibited cross-species conservation. These fragments then failed to identify any positive clones from several different cDNA libraries, or any transcript on northern blots (Fearon *et al.*, 1990).

1.9.2.2 CpG (HTF) island rescue

CpG islands or *Hpa*II tiny fragments are 500bp-2kb stretches of DNA with a high CpG dinucleotide composition. CpG islands are associated with the promoters of about 60% of genes, including all housekeeping genes analysed and 40% of tissue-specific genes (Bird, 1986; Antequera and Bird, 1993).

The vast majority of CpG islands encompass one or more 5' exons of genes. The gene identification strategy of 'CpG island rescue' thus targets the 5' exon and promoter sequences, which are typically the most difficult to isolate. Crude YAC DNA is digested with rare-cutting restriction enzymes that cleave preferentially in CpG islands. The fragments are ligated to vectorette adapters and PCR amplified with a combination of vectorette and Alu-repeat primers. Amplification products are gel fractionated against *inter*-alu amplification products. Vectorette-alu products potentially containing 5' coding sequences are 'rescued' and used as probes to screen random primed cDNA libraries or for northern analysis (Valdes *et al.*, 1994). An alternative method isolates CpG islands on the basis of their high thermal stability. Cloned genomic DNA is fragmented with restriction enzymes that cut frequently in

bulk DNA such as *Mse*I which recognises TTAA, but which preserve the integrity of the CpG island. The DNA is electrophoresed on a denturing gradient gel and CpG islands with high strand-dissociation kinetics are retained on the gel (Shiraishi *et al.*, 1995). Whole-genome CpG island libraries have been constructed and are available for use in isolating full-length cDNAs and associated promoter constructs (Cross *et al.*, 1994).

CpG island rescue, in conjunction with other methods, has been instrumental in cloning the genes responsible for adrenoleukodystrophy, neurofibromatosis II, Emery-Dreifuss and muscular dystrophy (reviewed by Parimoo *et al.*, 1995).

1.9.2.3 Exon trapping / amplification

Exon amplification is based on the selection of exons flanked by functional donor and acceptor splice sites. This enables the genomic DNA to be screened for coding sequences on a structural basis, independent of their temporal or tissue expression profiles (reviewed by Parimoo *et al.*, 1995). Fragments from genomic DNA clones are introduced into the intron of the HIV I *tat* gene with flanking 5' and 3' splice sites, present within the pSPL expression vector. The plasmid vector construct is transfected into COS cells whereby it replicates and high levels of transcript are produced through the action of the SV40 early promoter within the vector. These are processed according to the splice signals contained by the vector and genomic insert construct. If an exon is present within the genomic fragment, it will be retained in the processed mRNA such that a hybrid consisting of the reporter RNA plus introduced exonic sequence is produced. RT-PCR analysis and sequencing of transcribed products reveals the exon sequences. These may be used to screen cDNA libraries for the isolation of full-length cDNAs, or for northern analysis (Burke *et al.*, 1991). Despite the development of modified vector systems, this method is subject to contamination with artefacts due to the presence of cryptic splice sites within the genomic DNA, and vector RNA due to the introduction of genomic fragments lacking exons (Church *et al.*, 1994).

This system has led to the identification of the genes responsible for Menkes disease (Vulpe *et al.*, 1993) and diastrophic dysplasia (Hästbacka *et al.*, 1994). The main caveat to exon amplification is that small exons, or genes consisting of fewer than three exons, may be missed (reviewed by Parimoo *et al.*, 1995).

1.9.2.4 Direct cDNA selection

Direct cDNA selection is based on the solution hybridisation of a pool of cDNAs to an immobilised target genomic DNA probe, and the subsequent recovery of the specifically bound cDNAs by PCR amplification. After two rounds of selection the cDNA fragments are cloned and a sublibrary enriched for expressed sequences mapping to the genomic region of interest is produced. This technique in based on gene expression and relies on the candidate cDNA being present in the original RNA source used in the selection (Parimoo *et al.*, 1995).

Direct cDNA selection is a highly efficient means of cloning cDNA fragments encoded by large regions of genomic DNA. Originally the target genomic probes were comprised of purified, restriction digested YACs and cosmids (Parimoo *et al.*; Lovett *et al.*, 1991). Genomic probes have also included microdissected genomic DNA (Wei *et al.*, 1996), PACs (Osborne *et al.*, 1996) and crude preparations of YAC DNA. However, the latter was found to be 100-500 fold less sensitive than using purified YACs (Parimoo *et al.*, 1993).

There are two main direct selection strategies (Fig 1.12): DNA may be fixed onto small pieces of nylon membrane and the cDNAs hybridised to the target probe (Parimoo *et al.*; Lovett *et al.*, 1991). Alternatively, the genomic DNA can be labelled with biotin for subsequent immobilisation using streptavidin-coated magnetic beads. In the latter method both the genomic probe and the cDNAs are present in solution during hybridisation (Korn *et al.*, 1992; Morgan *et al.*, 1992). This method is more sensitive than the membrane-based protocol, since the hybridisation conditions follow second order kinetics (Lovett, 1994).

The starting cDNA source usually consists of a pool of PCR-amplified inserts from chosen cDNA libraries. Vector primers, which amplify across the multiple cloning site (MCS), are used to generate cDNA fragments with minimal contaminating vector DNA. The vector arms are used as primer-binding sequences for PCR-amplification of 'selected' cDNAs after hybridisation (Parimoo *et al.*; Lovett *et al.*, 1991). Uncloned cDNAs, derived from tissues in which a specific disease-gene is predicted to be expressed, may also be used as the starting source of cDNAs. Following reverse transcription of RNA and second-strand synthesis, the 'raw' cDNAs are ligated to amplification cassettes via a restriction site, or by blunt-ended ligation. These cassettes contain complementary sites for PCR amplification after rounds of selection, and a restriction site for subsequent cloning of selected cDNAs (Morgan *et al.*, 1992; Simmons *et al.*, 1995; Tassone *et al.*, 1995). In one strategy, uncloned cDNAs from different tissue sources were linked to separate sets of amplification cassettes and pooled for selection process. The different cDNAs were amplified separately according

to tissue source, using PCR primers specific to their cassette (Simmons *et al.*, 1995). The advantage of using uncloned cDNAs extracted directly from the source is that sequence complexity of the tissue is retained. Commercially available cDNA libraries may have a bias in copy number of particular transcripts and rare transcripts may be underrepresented in any one library (reviewed by Lovett, 1994).

High copy number miscellaneous DNA and RNA species, including human repetitive DNA, ribosomal RNA and host vector sequences are preblocked to suppress their selection. This is achieved by prehybridisation of the genomic probe or cDNA pool with a cocktail of quenching agents (Lovett, 1994; Parimoo *et al.*, 1995).

Post-hybridisation the fixed DNA is washed to remove non-specifically bound sequences, prior to elution of the strongly-bound cognate cDNAs. The 'primary selected' cDNAs are recovered by PCR-amplification and processed for enrichment in a second round of selection, using fresh probe (Fig 1.12). During this round some abundance normalisation occurs between transcripts that are present at high frequencies, and those present at low levels, with a net decrease in abundant cDNAs and a net increase in the rare cDNAs. The target genomic DNA is present in limiting amounts in the second round. Due to the increased frequency of cognate cDNAs after the first cycle, the abundant cDNAs will saturate their respective genomic targets in the second cycle, leaving redundant copies in solution. Rare transcripts will not be in excess of their genomic target sites, and a greater fraction of those present will bind to the fixed target (Lovett, 1994; Morgan *et al.*, 1992). The 'secondary selected' cDNA inserts are amplified using nested vector or ligator primers. These are cloned via the restriction site present in the flanking vector sequences to produce rich with cDNAs encoded by the target genomic probe. The sublibraries may be screened with radiolabelled repetitive, ribosomal and yeast DNA, to identify clones for exclusion from further analysis. The remaining cDNA clones are analysed for previously known and novel cDNAs mapping to the original genomic probe (Parimoo *et al.*; Lovett *et al.*, 1991).

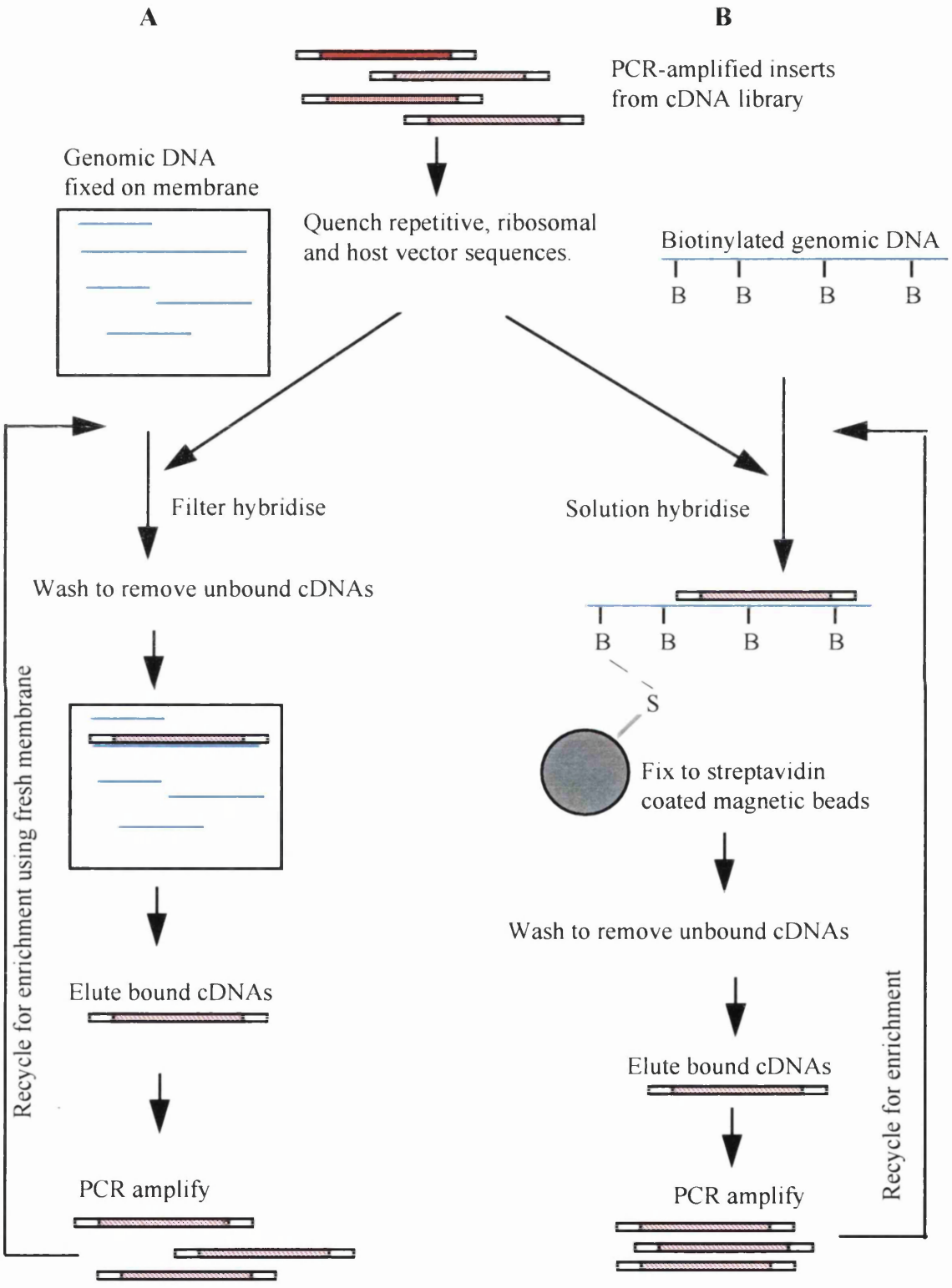


Fig 1.12 Flow diagram depicting the two strategies for direct cDNA selection.

A: Method using a nylon membrane for fixing the genomic DNA for use as probe. The probe is hybridised in a solution of cDNAs. **B**: Strategy employing biotin-labelled genomic DNA as probe. Hybridisation is carried out with both the probe and cDNAs in solution.

The relative abundances of particular cDNA clones in the selected sublibraries were shown to be increased by a factor of 800-4000 compared to the starting cDNA libraries, for membrane-based selection reactions (Parimoo *et al.*, 1991; Lovett *et al.*, 1991). Enrichment factors of control cDNAs may be up to 10^5 in selected sublibraries where the probe was labelled with biotin (Korn *et al.*, 1992; Morgan *et al.*, 1992).

Direct cDNA selection has been instrumental in the identification of the genes for neurofibromatosis I, familial polyposis coli, myotonic dystrophy and others (reviewed by Parimoo *et al.*, 1995). Direct selection has also been used to isolate conserved genomic DNA sequences from the mouse and pig, using cosmids from the human X chromosome (Sedlacek *et al.*, 1993). The membrane-based direct cDNA selection strategy was employed in this project to identify expressed sequences from the PWS and AS region.

1.9.2.5 DNA sequencing and computational analysis

Any positional cloning strategy will ultimately involve determination of the nucleotide sequence, at least for the cDNA. DNA may be sequenced using the dideoxynucleotide chain termination method, followed by polyacrylamide gel electrophoresis to resolve the differently sized products. Radioactive ^{32}P or ^{35}S are incorporated into the DNA during chain elongation and the sequence is determined by reading the radioactive banding pattern produced after electrophoresis and autoradiography. About 200bp of accurate sequence information is acquired with a single primer using this method (Sanger *et al.*, 1977). Fluorescent labelled DNA sequencing also exploits the dideoxy chain termination chemistry, but the dideoxy-nucleotides are labelled with four different fluoresceins to distinguish them. The sequence is read during electrophoresis by detection of the dyes using laser-technology. Between 400 and 600bp of sequence is recorded automatically by computation.

For all sequencing reactions, a certain amount of prior sequence information is required in order to design primers that will bind to and sequence from, complementary sites within the target DNA fragment. Where the sequence is completely anonymous primer-binding sites, which flank the MCS of vectors, may be exploited for sequencing cloned inserts. For fragments of DNA that are too large to be sequenced in one read, there are three main strategies for acquiring the full-length sequence:

(1) The DNA can be restriction-digested and the resultant fragments subcloned. This reduces the target DNA into smaller sections, which can be sequenced with vector

primers alone. The sequences obtained for the small overlapping sections are ordered into a contig according to the restriction map, and a consensus sequence derived.
(2) Transposon-facilitated sequencing exploits the mobile properties of transposons to insert known sequence-priming sites contained within the transposon into the DNA fragment to be sequenced. This also divides the target fragment into sections small enough to be sequenced (Strathmann *et al.*, 1991).
(3) 'Oligomer walking' involves a series of sequencing reactions using different primers that progressively sequence away from the starting point. Sequence tagged sites are generated at each sequencing step enabling further oligomers to be designed with which to sequence the next adjacent section (Strauss *et al.*, 1986). This method was used to sequence the inserts of cDNA clones mapping to 15q11-13 that were larger than 1kb.

Genomic DNA may be sequenced and processed for identification of putative exons by computational analysis. The gene recognition and analysis internet link (GRAIL) program is based on a multiple sensor-neural network which analyses the DNA sequence for protein-coding features such as open reading frames, splice sites and nucleotide distribution. This program advantageously allows some room for sequencing error (Uberbacher and Mural, 1991). During the search for the genes responsible for adrenoleukodystrophy and colorectal cancer linked to 18q, several cross-hybridising fragments were subcloned and sequenced to predict any coding regions by GRAIL. The failure of these fragments to identify positive clones on cDNA library screening was presumably due to their low signal to noise ratio; that of small exons interspersed with large intronic sequences. On the basis of the GRAIL predictions, exon connection PCR was performed between putative exon sequences to generate probes that finally identified positive clones from cDNA libraries (Fearon *et al.*, 1990; Mosser *et al.*, 1993).

Several publicly accessible databases are available in which genomic DNA and cDNA sequences from human, rodents and other organisms are deposited (section 1.9.3). Various tools have been produced which allow these databases to be searched for nucleotide and peptide sequence similarities with a query sequence. The FASTA (Pearson and Lipman, 1988) and basic local alignment search tool, BLAST (Altschul *et al.*, 1990), use scoring parameters and alignment algorithms to display regions of similarity between two sequences.

NIX (Nucleotide Identify X) is a modern, powerful World-Wide Web tool which analyses nucleotide sequences, both genomic and cDNA, for informative regions. NIX

employs an assortment of programs including 'GRAIL', 'Genemark', 'Genefinder', 'BLAST', 'Polyah', and 'Repeatmasker' concurrently, using the default parameters. Thus the query sequence is analysed for particular features, such as coding regions, predicted intron-exon boundaries, repeat elements, polyadenylation signals and regions of homology to sequences from many databases. The results are displayed graphically for simplicity.

1.9.3 The Human Genome Mapping Project (HGMP)

The HGMP, which aims to sequence the entire human genome, has provided the impetus for genetic and physical mapping, and large-scale sequencing. An international collaboration between the Centre d'Etude du Polymorphism Humain (CEPH), the Whitehead Institute and Généthon genome centres, has generated genome-wide YAC contig and radiation hybrid maps, with 99% genome coverage in overlapping clones. These maps are anchored by genetic or physical markers, in the form of STSs and ESTs. This has enabled the integration of information on genetic, physical and transcript maps from multiple genome-wide mapping efforts, providing a framework for sequencing the entire genome (Cohen *et al.*, 1993; Chumakov *et al.*, 1995; Hudson *et al.*, 1995).

Sequence databases including GenBank, Embl and dbEST are rapidly expanding with anonymous EST sequences of 150-500bp in length, derived from various cDNA libraries (Adams *et al.*, 1991; 1992; Khan *et al.*, 1992; Berry *et al.*, 1995; Lennon *et al.*, 1996). A large proportion of these has been assigned to radiation hybrids and the genome-wide YAC contig (Berry *et al.*, 1995). The Genome Sequencing Group at Washington University and Merck and Co set up the I.M.A.G.E. (an integrated molecular analysis of genomes and their expression) Consortium to provide arrayed cDNA libraries for large-scale public domain sequencing. Individual clones are distributed for research purposes through I.M.A.G.E. distribution centres. Sequence tags have been produced to the 5' and 3' ends of these clones, which are publicly available in GenBank and Embl. Since the I.M.A.G.E. Consortium was initiated in 1993, sequence data for tens of thousands of clones is now available. The ultimate aim is to create a 'master array' containing at least one representative cDNA for each transcript in the genome (Lennon *et al.*, 1996).

The dramatic increase in numbers of ESTs in the sequence databases has meant that development of a comprehensive transcription map of the human genome is now feasible. Over 80% of known disease-causing genes demonstrate significant sequence identity to ESTs

in GenBank, making it likely that a high proportion of as yet unidentified genes are represented by ESTs. However, GenBank contains a large number of redundant sequences, due in part to repeated submissions of sequences from different laboratories, but also due to the varying sources of material sequenced, including genomic DNA, alternatively spliced mRNAs and cDNA clones or ESTs. UniGene cluster identifies and groups sequences belonging to the same transcript on the basis of statistically significant sequence homology. The main aim of UniGene cluster is to generate a non-redundant transcription map containing all the expressed sequences in the human genome. To begin with, the ESTs in GenBank were screened by all existing functionally cloned genes to reduce redundancy amongst known genes. Remaining EST entries in GenBank are continuously compared with one-another to identify redundant and overlapping sequences. Where sequence identity is found in the conserved 3' untranslated region, the ESTs are anchored to a single cluster. The clusters are then distributed to mapping centres where they are localised to radiation hybrid and YAC maps. The UniGene Collection is divided by chromosomal location (Schuler *et al.*, 1996; Schuler, 1997).

The production of a human genome transcription map by the HGMP will greatly expedite the identification of disease genes by the 'positional candidate' approach (Boguski and Schuler, 1995). In this strategy a disease locus is mapped or linked to a particular chromosomal region, and likely candidates in the region are screened for mutations. It has been predicted that in future the majority of disease genes will be identified by analysing previously mapped genes or transcripts, rather than through labour-intensive positional cloning methods (Collins, 1995).

All information pertaining to the HGMP, including genetic and physical maps, EST and STS sequences, and Unigene clusters is publicly accessible through the National Centre for Biotechnology Information at the world-wide web site: **http://www.ncbi.nlm.nih.gov/**

1.9.4 Mutation detection

In order to establish whether a candidate gene is causal to a given genetic disease, mutations which disrupt the gene or function of the gene product need to be demonstrated in affected individuals. Mutations range from large deletions, inversions, duplications and triplet repeat expansions to single nucleotide changes. There are a variety of means for detecting unknown molecular mutations (reviewed by Grompe, 1993):

1.9.4.1 Southern blot analysis

Southern blot (Southern, 1975) analysis requires no detailed prior knowledge of the gene under analysis, or the probe used for screening. This method is often used to screen a battery of patients for mutations using a novel probe that is considered a potential candidate. Deletions and insertions in the genomic DNA of patients may be detected as junction fragments; altered bands compared to normal individuals. Deletions and duplications may also be seen as a decrease or increase in signal dosage. Point mutations may be detected if they alter a restriction site that is spanned by the probe. One of the most frequently observed mutations results from oxidative deamination of cytosine to uracil, seen as a C to T transition (reviewed by Lewin, 1994). These frequently occur in CpG dinucleotides where the cytosine is methylated. *Msp*I recognises the tetranucleotide sequence C$^{\downarrow}$CGG and *Taq*I cleaves at the site T$^{\downarrow}$CGA, so these are useful for detecting this type of mutation. Southern blot analysis of AS patients with no deletion, disomy or imprintor mutation was performed using these two enzymes, with candidate cDNAs from the ASCR as probes.

1.9.4.2 PCR and gel electrophoresis based methods

Various methods of mutation detection based on PCR amplification followed by polyacrylamide gel electrophoresis have been developed which rely on differential electrophoretic mobility between the mutant and wild-type molecules. These are visualised as band shifts in the gel, and may be detected by ethidium bromide or silver staining the gel, or through incorporation of ^{32}P or ^{35}S labelled deoxynucleotides in the PCR. These methods indicate the presence of a sequence change within the amplified fragment, but to determine the precise position and nature of the alteration, the fragment then requires sequencing. Through the application of these methods, a gene may be efficiently scanned in a number of patients for the presence and rough location of mutations, minimising the sequencing effort.

Heteroduplex analysis is used for detection of heterozygous mutations. PCR amplification of the DNA forms homoduplex molecules (double-stranded DNA of the same sequence) and heteroduplex molecules (annealing of the mutant and wild-type strands), which exhibit altered migration in non-denaturing gels. The detection rate is 80-90% for DNA fragments of <300bp.

Temperature or denaturing gradient gel electrophoresis involves running PCR fragments through a gel of increasing temperature or concentration of denaturing agent (formamide and urea). The principle underlying these methods is that double-stranded DNA dissociates in regions of low melting temperature. The DNA reaches a point in the gel whereby electrophoretic mobility abruptly decreases. Mutant and wild-type molecules have different dissociation temperatures, detected by altered migration in the gel. The detection rate is about 95% for 600bp fragments.

1.9.4.2.1 Single-stranded conformational polymorphism (SSCP) analysis

SSCP analysis (Orita *et al.*, 1989) is the most widely used scanning technique due to its simplicity. Amplified mutant and wild-type fragments are denatured and electrophoresed alongside one-another through a non-denaturing polyacrylamide gel. The single-stranded DNA molecules form a three-dimensional conformation, which is directly dependent on nucleotide composition. A sequence alteration may cause a different conformation to be assumed, which is detected as an altered migration pattern compared to that of the wild-type. The mutation detection rate for SSCP is 70-95% in fragments of 200bp, but this reduces to 50% in PCR products over 400bp. However, running samples in different temperatures or in the presence of glycerol increases the rate of detection. This method was used for screening the coding region of *UBE3A* for mutations in AS patients with no deletion, UPD, or imprintor mutations.

1.9.4.3 Mutations as opposed to polymorphisms

Mutations may be confirmed by restriction digestion, if they alter an enzyme recognition site, or by repeated sequencing of the alleles. However, for sequence alterations which do not result in a chain termination or frameshift, the biological consequence of the change may be difficult to predict. Missense mutations, which result in single amino acid substitutions, may be functionally insignificant and represent polymorphisms. To confirm if the change is disease-causing *in vitro* expression studies, such as assaying the activity of a mutant protein using patient cell lines or expression vectors, may be performed. Alternatively a given change may be searched for in the DNA of normal individuals. If the change remains unidentified in 100 normal chromosomes, then it is considered pathogenic. Where a genetic disease is familial, the sequence change should segregate with all affected individuals in the pedigree and be absent in unaffected members, in a manner consistent with the mode of inheritance.

1.10 Aims of the project

The main aim of this project was to identify the gene responsible for AS. This was based on two premises. The first was that AS is a single gene defect that is linked to D15S10 and *GABRB3* within 15q11-13 in families in which the disease is segregating. The second was that 15q11-13 is subject to genomic imprinting and AS is transmitted solely on the maternal allele. Therefore, a gene must demonstrate two criteria in order to be considered a candidate for AS: The gene must map to the region demonstrating linkage to AS, and it must demonstrate monoallelic expression from the maternal allele. At the start of this PhD thesis, no gene meeting these criteria had been identified. To prove that a candidate gene is causative of AS, pathogenic mutations must be demonstrated on the maternally derived allele in familial cases, and transmission through the paternal allele must be silent.

The biochemical defect underlying the pathophysiology of AS was unknown, but the position of the gene within the genome, was known. The observation of cytogenetic deletions of the maternal 15q11-13 allele provided the first indication that the gene responsible for AS lay within this region. The analysis of various patients with microdeletions or a recombination event within 15q11-13, led to the delineation of the ASCR, constituting the minimal region in which the gene is most likely to lie. At the start of this project this region had been cloned in YACs. In order to identify gene sequences potentially involved in AS, a positional cloning approach was undertaken within the ASCR. This involved cloning the ASCR in smaller vectors, physical mapping, and direct cDNA selection to identify expressed sequences. These were analysed by a combination of laboratory-based methods and electronic means. Since YACs provided the initial resource for isolating transcripts from the ASCR, the region analysed also overlapped with the distal portion of the PWS SRO. PWS is a multi-gene disorder, due to loss of paternal transcription of genes in 15q11-12. Thus it was considered important to analyse transcripts isolated from the PWSCR, with respect to PWS as well.

During the course of this project, mutations of the *UBE3A* gene were identified in a small number of AS patients in whom the genetic cause for disease had not been established (Matsuura *et al.*, 1997; Kishino *et al.*, 1997). Mutations of *UBE3A* were sought in AS patients referred to The Institute of Child Health for genetic testing, who had no deletion, disomy or imprintor mutation, to determine the molecular aetiology for disease in these cases.

Chapter 2: Materials and Methods

2.1 Materials

2.1.1 Chemicals

All chemicals and reagents were purchased from Sigma, with the following exceptions:

α-^{32}P dCTP and α-^{35}S dATP were supplied by ICN Biochemicals Limited.
Taq Polymerase, 10xNH_4 buffer, 10x KCl buffer and 50mM $MgCl_2$ for PCR were obtained from Bioline.
T4 DNA ligase and ligation buffer, DNA and RNA markers, poly(dT) primers, Rnasin, DTT (0.1M), MMLV first strand buffer and MMLV reverse transcriptase were obtained from Promega.
Streptavidin-coated poly(dT) Dyna beads were obtained from Dynal (UK) Limited.
Restriction enzymes were acquired from Promega, Gibco or New England Biolabs.
PCR primers were synthesised by Genosys. The quantity and melting temperature of the primers was supplied by the manufacturer.
T7 Sequenase version 2.0 DNA polymerase sequencing kit was supplied by United States Biochemicals.
Proteinase K was acquired from BDH.
Agarose gel and the Trizol reagent were obtained from Life Technologies.
The Dye Terminator Cycle Sequencing Ready Reaction kit and the dRhodamine Terminator Ready Reaction kit were bought from Perkin Elmer.
Ultrapure 19:1 acrylamide:bisacrylamide powder, Urea, amberlite, and TEMED were bought from National Diagnostics.
Mutation detection enhancement (MDE) gel was obtained from AT Biochemicals.
S-300 Microspin columns were supplied by Pharmacia.
The Geneclean kit was acquired from BIO.
The LigATor plus ligase DNA cloning kit was supplied by R + D systems.

2.1.2 DNA and Cell lines

Blood samples from AS and PWS individuals were taken from the patient for genetic diagnosis and research purposes, with the informed consent of the patient or guardian. Ethical permission for the use of PWS and AS DNA samples for the study of genes involved in these syndromes was provided by the Great Ormond Street Hospital for Sick Children Ethical Committee, (application 642). Lymphocytes were EBV-transformed to create lymphoblast cell lines by Mr. Paul Rutland, Molecular Genetics Unit, Institute of Child Health.

DNA from anonymous normal individuals for use as controls were provided by the Clinical Genetics Department, Institute of Child Health.

Cultured fibroblasts from normal individuals were kindly provided for RNA extraction by Dr. Jane Hill, Biochemistry Unit, Institute of Child Health.

Mouse genomic DNA used for analysis of evolutionary conservation was a gift from Dr. Karen Woodward, MRC Biochemical Genetics Unit, Galton Laboratory, UCL, London.

2.1.3 Fetal Tissues

Fetal tissues were obtained for RNA extraction from abortuses of 6 to 20 weeks' gestation after termination of pregnancy at Queen Charlotte's and Chelsea Hospital, Imperial College of Medicine, London. Samples were collected in collaboration with Dr. Gudrun Moore and with the kind assistance of Dr. Emma Wakeling and Lisa Sherman from the RPMS Institute of Obstetrics and Gynaecology, Queen Charlotte's and Chelsea Hospital. Ethical permission was granted by the Hammersmith and Chelsea Research Ethics Committee for research into congenital malformations (application 94/4290).

2.1.4 Genomic libraries and clones

The human P1 artificial chromosome (PAC) library and individual PAC clones were provided freely by the HGMP Resource Centre for research purposes. The PAC library was constructed from a normal male donor by Pieter de Jong, Rowell Park Cancer Research

Institute, Buffalo, USA. The vector used was pCYPAC2 and the insert size was on average 110kb. The library was provided as high density gridded filters, on which each clone had been spotted in duplicate, and as PCR pools.

BAC clones 139E9, 119H2, 268A2 and 119G2 were kindly given by Dr. Ung-Jin Kim, Caltech, USA.

2.1.5 cDNA libraries and clones

Two nine week gestation fetal brain libraries, Br32 and Br33, were kindly provided for cDNA library screening and as the cDNA source for direct cDNA selection, by Michel Fontez, INSERM, Marseilles, France. Br32 was produced by poly dT reverse transcription of RNA, and Br33 was produced by random priming. Both libraries were cloned in Lambda Zap 11, via the *EcoR*I and *Xho*I restriction enzyme sites of the multiple cloning site.

I.M.A.G.E. Consortium cDNA clones (whose end-sequences were made publicly available in GenBank), were distributed by the UK HGMP Resource Centre, free of charge for reasearch purposes.

2.1.6 Monochromosomal hybrids containing single chromosomes 15

cDNAs from monochromosomal human-rodent hybrids A15, A9+15, A59.3aZ (containing a paternal human chromosome 15) and 20L.28, ALA-8, t75.2maZ (containing a maternal human chromosome 15) were kindly made available for imprinting assays of novel cDNAs from 15q11-13, by Dr. Robert Nicholls, Case Western Reserve University School of Medicine, Cleveland, USA. The cDNAs and corresponding RT-minus controls were prepared by Mr. James Gabriel.

2.2 Solutions and buffers

Unless otherwise stated, solutions were made up in 1L volumes using distilled deionised water, and autoclaved. Where adjustment of the pH was necessary, the chemicals were dissolved in 800ml water, the pH adjusted, and then the volume made up to 1L.

2.2.1 DNA extraction solutions

Nuclei lysis buffer

10mM Tris

400mM NaCl

2mM EDTA

Suspension buffer

2mM EDTA (pH 8.0)

TE

10mM Tris

1mM EDTA

Adjusted to pH 7.6 with HCl

P1 buffer

25mM Tris-HCl (pH 8.0)

10mM EDTA (pH 8.0)

50mM Glucose

100μg RNase A/ml

P2 buffer

0.2M NaOH

1% SDS

P3 buffer

3.0M Potassium acetate (pH 5.5), stored at 4^0C

CiDIS

100mM NaCl

10mM Tris-HCl (pH 8.0)

1mM EDTA

2% Triton-X

1% SDS

SCEM

1M Sorbitol

0.1M NaCl

10mM EDTA

This was stored at room temperature. 30mM β-mercaptoethanol was added fresh, before use.

2.2.2 RNA extraction solutions

RNA solutions were made up DEPC-treated unless they contained Tris-HCl, in which case they were made up using DEPC-treated water, and autoclaved.

2X binding buffer

20mM Tris-HCl (pH 7.5)

1.0M LiCl

2mM EDTA

Washing buffer

10mM Tris-HCl (pH 7.5)

0.15M LiCl

1mM EDTA

Elution buffer

2mM EDTA (pH 7.5)

2.2.3 Electrophoresis buffers

10X TAE buffer

40mM Tris-acetate

1mM EDTA (pH 8.0)

5X TBE

0.089M Tris

0.089M boric acid

2mM EDTA (pH 8.0)

5X MOPs buffer for electrophoresis of RNA

0.2M MOPs, pH 7.0

0.05M sodium acetate

5mM EDTA, pH 8.0

The pH was adjusted to 7.0 using NaOH pellets.

2.2.4 Loading dyes

Loading dye for agarose gel electrophoresis

0.05% xylene cyanol dye and /or 0.05% bromophenol blue dye

10mM EDTA (pH 8.0)

50% gylcerol

Loading dye for polyacrylamide gel electrophoresis of manual sequencing and SSCP products

0.05% xylene cyanol dye and 0.05% bromophenol blue dye dissolved in 1 part 25mM EDTA (pH8.0). 5 parts deionised formamide was added. The dye was stored at –20°C.

Blue dextran loading dye for fluorescent-labelled automated sequencing

0.05% blue dextran was dissolved in 1 part 25mM EDTA (pH8.0).

5 parts deionised formamide was added. The dye was stored at -20°C.

RNA sample buffer

10ml deionised formamide

3.5ml 37% formaldehyde

2ml 5X MOPs buffer

Stored at -20°C for up to 6 months

RNA loading dye

0.4% bromophenol blue

1mM EDTA pH 8.0

50% glycerol

2.2.5 Southern blot and hybridisation solutions

20X SSC

2.9M NaCl

0.3M Sodium citrate

Adjusted to pH 7.0 with NaOH

Denaturation solution

0.5M NaOH

1.5M NaCl

Neutralisation solution

0.5M Tris-HCl (pH 7.2)

1.5M NaCl

1mM EDTA

Polyethylene glycol (PEG) Prehybridisation solution

5X SSC

10% PEG

1% SDS

1.0 mM EDTA

100µg/ml denatured sonicated salmon or herring sperm DNA

Oligolabelling buffer (OLB)

Solution O: 1.25mM Tris-HCl (pH 8.0)

0.125M $MgCl_2$

Solution A: 1ml solution O

18µl β-mercaptoethanol

5µl 0.1M dATP

5µl 0.1M dGTP

5µl 0.1M dTTP

Solution B: 2.5M HEPES, adjusted to pH 6.6 with 4M NaOH

Solution C: random hexanucleotides in TE (90 OD units/ml)

***OLB*:** Solutions A:B:C combined in a ratio of 10:25:15

2.2.6 Northern hybridisation solutions

50X Denhardts solution

10 mg/ml Ficoll (400)

10mg/ml polyvinylpyrrolidone

10mg/ml BSA

The solution was filtered and stored at -20^{0}C.

Northern hybridisation solution

50% deionised formamide

5X SSPE

2X Denhardt's solution

0.1% SDS

2.2.7 Cell culture media

LB broth (E.coli)

20g/l LB base, autoclaved

LB agar (E.coli)

20g/l LB base

15g/l bacto-agar

Autoclaved

NZY bottom agar (E.coli)

5g/l yeast extract

5g/l NaCl

2g/l $MgSO_4.7H_2O$

10g/l NZ amine (casein hydrolysate)

15g bacto-agar

Autoclaved

NZY top agarose (bacteriophage-infected E.coli)

As for NZY bottom agar, except 0.7% Sea Kem LE agarose replaced the bacto-agar.

AHC= media (S.cerevisiae)

6.7g/l yeast nitrogen base minus amino acids

10g/l tryptone

20mg/l adenine hemisulphate

This was made up to 800ml and autoclaved, prior to the addition of 200ml 10% filter sterilised glucose.

YPD medium (S.cerevisiae)

10g/l yeast extract

20g/l peptone

This was autoclaved prior to adding 10mls 20% filter sterilised dextrose.

SM storage buffer for bacteriophage

5.8g/l NaCl

2g/l $MgSO_4.7H_2O$

0.05M Tris-HCl (pH 7.5)

0.01% gelatin

2.2.8 Media and buffers for human tissues and cell culture

1X RPMI 1640 media (Lymphoblast)

10% heat-activated fetal calf serum

200mM glutamine

50mg/ml penicillin and streptomycin or gentamycin

Phosphate buffered saline (PBS)

1 tablet per 100ml water was dissolved and autoclaved.

2.2.9 Direct cDNA selection solutions

20X SSPE

175.3g NaCl

27.6g NaH2PO4.H20

7.4g EDTA disodium salt

Adjusted to pH 7.4 with NaOH pellets and the volume made up to 1L

20X quench mix

0.5μg/μl total human or Cot 1 DNA

0.5μg/μl yeast AB1380 DNA (for YACs) or *E.coli* DNA (for PACs and BACs)

0.5μg/μl pYAC4 (for YACs) or PAC DNA (for PACs)

0.8μg/μl *Hae*III digested human ribosomal DNA (pR5.8 and pR7.3 in a ratio of 1:0.8)

The ribosomal DNA clones pR5.8 and pR7.3 were in pBR322. These required chloramphenicol amplification, maxiprepping, and excision from the vector by digestion with *EcoR*I and *Xba*I.

2.3 Primers

Primer	Vector	Sequence
M13 universal	pBlueScript/ pT3T7	GTAAAACGACGGCCAGT
M13 reverse	pBlueScript/ pT3T7	GGAAACAGCTATGACCATG
T3	pBlueScript/ pT3T7	AATTAACCCTCACTAAAGGG
T7	pBluescript/ pT3T7	GTAATACGACTCACTATAGGGC
pTAg-5	pTAg	GCTATGACCATGATTACGCCAAG
pTAg-3	pTAg	TGTAAAACGACGGCCAGTGAAT
L-REV*	pMOB	AACAGCTATGACCATGATTACGCCAAG
L-UNI*	pMOB	GTAAAACGACGGCCAGTGAGCGCG
GDIR*	Transposon γδ	TTTCGTTCCATTGGCCCTCAAACCCC
GD1*	Transposon γδ	CAACGAATTATCTCCTT
GD2*	Transposon γδ	TCAATAAGTTATACCAT

Table 2.1 Plasmid vector and transposon primers
An annealing temperature (Tm) of 55°C was used for all vector and transposon primers, for PCR and sequencing. * Primers from Strathmann *et al.*, 1991.

STS marker/ transcript	Primer sequence 5'-3'	Tm (°C)
D15S174	F: AATAAAGATGTGAGGTGC R: TTAGGGTCTCCTTTAAGTTTCA	58
D15S122	L: GATAATCATGCCCCCCA R: CCCAGTATCTGGCACGTAG	56
D15S10 *	GTTCCTTATCTCATTTAGTATATTTGTTGG GGGCTGTTTCAAAACTGAAAGTAACATG	56
D15S210	L: TGAACTAGGCAGCACACACT R: AACCAATGCTTGGACACAG	55
D15S113 + (LS6-1)	CAATTCCTGACATATAGTGGGTGC CTGGGATTACAGGCATAAGCCAC	55
B230E3L +	GACTCACAGATTACTTAGAATTGTG GCAGTCTATCTGTATAAAATGGAACAG	55
B230E3R +	CAGAAGGGAGAAATAGACAACTCAGC CTGGAATTATAGGTGTGTGCCACC	55
PAR-1	F: GCACAAGGCTATGTTCATGCACT R:GCATCAGTAAGGTAGAGGACAGGA	64
PAR-2	F: GGCTTTAATACTCCCAGGGCTTG R: GCACTGTAGTAGGACAGTGTTGGT	63
PAR-4	F: GATCTCCTGCACTGAGCTGTGGTG R: AGGGCTCCAGCCCAAAATACAGTG	70

Table 2.2 PCR primers for amplification of STS markers and transcripts from the PWS and AS region.
Primers from: * Kuwano *et al.*, 1992. + Mutirangura *et al.*, 1993.

EST	Primer sequence 5'-3'	Product size (bp)	Tm (°C)
WI-6654	L: CTAGTTTTGTTCATGAATTGTTTGC R: CAGAGGGGGTGACTGTTTCT	250	62
WI-6780	L: CAATTTAAGACAATAGTTTGGGTGA R: GAATTACTGGGATTAAGTATGTGCA	200	61
WI-16777	L: ATGTTTATGTGCTTTTTAAACATCG R: GATCCCAGGATTGAGATGTCA	125	62
WI-13791	L: GATACAAAAATCCAACAACTTTAGC R: TTTATTGACCGTATCATAACACGC	200	60
WI-6519	L: ACATACACAGAAGACTAGGAAAGGG R: AGAATTACATTGTATAGCCCCACA	154	62

Table 2.3 PCR primers for mapping of Whitehead Institute (WI) ESTs.

Primer	Exon number old (new)	Sequence 5'-3'	Tm (°C)
E6AP-1F E6AP-3R	1 (7) 3 (9)	ATCAGGAGAACCTCAGTCT CTATCACCTATCCTTGAGG	52
E6AP-4F E6AP-8R	4 (10) 8 (14)	TATCTGGAAATGGCTTTGCC GCTCTGTCTGTGCCCGTT	63

Table 2.4 Primers for RT-PCR amplification of the *UBE3A* cDNA.

Primer	Sequence 5'-3'	Tm °C
CL1	TCCCAAAGTGCTGGGATTACAG	50
CL2	CTGCACTCCAGCCTGGG	50

Table 2.5 *Inter*-Alu PCR primers.
(Lengauer *et al.*, 1992).

Primer	Sequence 5'-3'	Tm (°C)
G13-5' G13-3'	ACGAGGGATAAATTTTTGGTG ATATAAAACGCCCTCAGAGT	56
G13-S1	TCATCAGGTTGATCTACAGT	54.3
G13-S2	TGCCTCGATAAAATGGCTGA	65.4
B4-S1	GAAAGGACAGGGATTTTTG	58.5

Table 2.6 PCR and sequencing primers for I.M.A.G.E. clones 1354-G13 and 215-B4.

Primer	Sequence 5'-3'	Tm (°C)
H22-5'	TCCAGCACATTACTCCAACA	58
H22-3'	GTTCAGCTTTCCAAGGAATG	
H22-S1	TATCTGCCTGTCTCTGGCC	63.9

Table 2.7 RT-PCR and sequencing primers for I.M.A.G.E. clone 395-H22.

Primer	Sequence 5'-3'	Tm (°C)
E19-5'	CACAGAGAACTTGGTACAAA	52
E19-3'	CTTTCTTGTCCAGATTAGCT	
E19-S1	TTAGCAGGCATGTGCCCTG	66.9
E19-S2	CTGTTACACATCTTTGCT	64.8

Table 2.8 RT-PCR and sequencing primers for I.M.A.G.E. clone 123-E19.

Selected sequence	Primer sequence 5'-3'	Tm (°C)
16	F: CTAGTAAACTAGGAATAGAGG R: GTCATTTCATCTGCAAACAAAG	48
26	F: TAAGGTTATGTCTCACCTGTT R: AGGGAAAAGCAGCTAAATCAT	55
39	F: GCACAGAAAAGTTCTACAGG R: ACTGACCTGGTCCATTTGGAA	56
86	F: GAGCCCAATCCCATATCTTT R: AGGGCAGTATAAAGAAGGAAG	58
113	F: GTGGACTGTCTCAATTACTC R: GAGCCTAGATGGAAGAGTTA	53
119	F: CACGAGCCAGATGAAAGT R: GGGGGGACAAGCTTATAT	50

Table 2.9 Primers for amplification of selected cDNA sequences by RT-PCR.

cDNA	Primer sequence 5'-3'	Product length (bp)	Tm (°C)
SNRPN	N1: GCTCCATCTACTCTTTGAAGC N4: GGCCATCTTGCAGGATACATC	225	60
HPRT	F: CCACGAAAGTGTTGGATATAAGC R: GGCGATGTCAATAGGACTCCAGATG	220	63

Table 2.10 Primers for RT-PCR controls.
SNRPN primers kindly provided by Dr. John Huntriss, Molecular embryology department, Institute of Child Health.

Primer	Exon number new (old)	Sequence 5'-3'	Tm (°C)
OP2F2 OP2R3	1 (OP2)	CCCGCCCAGCCAGTCCTCCC CTGTCGTCGCCCCCGCGCCT	55
U2F # U2R	2 (U2)	ACCCTGATGTCACCGAATGGC TTTTACAAGCTGTGGCCAT	50
723-A + 723-B +	5 (U3)	TTTCTCTAGAAGTTTTTATAAC ATCGCAGAAAATATGATCAC	52
U4F U4R	6 (U4)	TTGTTTTGCAAGCCAGCTC TTGTGGGTAAGTACCCCAA	53
2F1 2R2 #	8 (2)	GAAGCGCGCAGCTGCAAAGC ACCCTGATGTCACCGAATGGC	54
3F2 3R2	9 (3)	GTGACTTACTTAACAG CTATCACCTATCCTTGAGG	52
3F3 3R3	9 (3)	GCTATGGAAGAAGACTCAG GATATTCAGGACTGTGGAG	53
3F4 3R4	9 (3)	GTTCATTATCGGAATGGAG AAGTTCCTGAAGTGTCAGCTC	52
3F5 3R5	9 (3)	GGAAGTGGACACAAATCAC CGGACAAGTGCATCATCTA	57
E63-C + E63-D +	9 (3)	GATGAAGACAAAGATGAAGAT ATCCACAGACACATCATCAG	52
E63-E + E63-F +	9 (3)	GACAACAATTTGCAAAAATTAG AGTCACATTCCACGTTAGGT	52
E63-G + E63-H +	9 (3)	GCCTTTCTCAATGCACTTGT AGCCATTTCCAGATATTCAG	52
E6-9C + E6-9D +	9 (3)	ATTGTCGAAAACCACTTATC CCCAAATTCTTTGTGACAGCA	55
E610A * E610B *	10 (4)	GCAATCATCTTCTTTTCATGTT CGACACCATAATCACATTAC	56
5F12 5R12	11 (5)	GTTCACATACGATGAATCTAC CAAGTCGACGAAAAGTCCTT	52
6F11 6R11	12 (6)	GTTCTATATCAGAGTTTA CTGTTTTCATTTGTAATTG	50
7F10 7R10	13 (7)	GAATTTGTCAATCTTTATTC CCGGCTTCCACATATAAGCA	51
E6-14F + E6-14R +	14 (8)	AGTAGTATAGCAGATAACTAAGAC CCCTTTGGTGAATCAAATCTTCC	52
9F9 9R9	15 (9)	GGAGTTCTGGGAAATCGTTC CTGTGTCTGGGCCATTTTTG	53
E6ZA * E6ZB *	(10)	ACCATGACTTACAGTTTCCT TGGGACACTATCACCACCAA	55

Table 2.11 Primers used for mutation screening of *UBE3A*.
Primers from: # Yamamoto *et al.*, 1997. + Malzac *et al.*, 1998. * Kishino *et al.*, 1997.

2.4 Methods

2.4.1 Preparation of DNA

2.4.1.1 Extraction of genomic DNA from peripheral blood lymphocytes

To separate the DNA-containing lymphocyte nuclei, 10ml blood was mixed with 90ml ice-cold lysis buffer in a sterile GSA centrifuge bottle, and spun at 10,000rpm for 15 minutes at 4°C in a refridgerated centrifuge (Sorvall). The supernatant containing lysed cells was poured off into disinfectant, and the pellet resuspended in 4.5ml suspension buffer and transferred to a 13ml Sarstedt tube. 50µl 10% SDS and 50µl Proteinase K (20mg/ml stock) were mixed in and incubated at 37°C for 3 hours or at 50°C overnight to allow protein digestion. 5ml phenol equilibrated in 0.1M Tris (pH 8.0) was added to the DNA extraction mix and the tube shaken. The tube was spun at 10,000rpm for 10 minutes at room temperature, and the top aqueous phase containing the DNA removed from the organic phase. Phenol extractions were repeated twice. The aqueous phase was then mixed with an equal volume of chloroform, spun at 10,000rpm for 5 minutes and the aqueous layer removed. This was repeated twice to remove all traces of phenol. The aqueous layer removed from the final chloroform extraction was added to 11ml absolute ethanol in a sterile universal tube. The tube was inverted gently until the DNA precipitated. The DNA was spooled out using a hooked pasteur pipette, washed in 70% ethanol, and dissolved in 1ml TE buffer. The DNA was allowed to dissolve on a rotating rack overnight, and stored at –20°C.

2.4.1.2 Small scale preparation of plasmid DNA using Qiagen columns (miniprep)

Plasmid DNA was extracted using the Qiagen 'mini-prep' kit according to the manufacturers conditions. 3ml overnight bacterial cultures (in LB broth containing 50µg/ml ampicillin or kanamycin) were spun in eppendorfs at 6,500rpm for 5 minutes in a microcentrifuge. The supernatant media was discarded into chloros and the pelleted cells resuspended in 100µl chilled P1 buffer. The cells were lysed by adding 200µl freshly made P2 buffer and mixing by gently inverting the tube (to co-precipitate the chromosomal DNA with the cell wall), until the suspension became clear and viscous. 150µl chilled P3 solution was added and mixed immediately by inversion and left on ice for 15 minutes for neutralisation to occur. The extraction mix was spun at 13,000rpm for 10 minutes to pellet the debris, and the supernatant containing the plasmid DNA was removed and loaded on a Qiagen-tip 20, pre-equilibrated with 1ml QBT buffer. The

bound DNA was washed twice by adding 1ml QC buffer to the column, then eluted into a collection tube, with 800µl QF buffer. The DNA was precipitated by adding 0.7 vols isopropanol and centrifugation at 13,000rpm for 15 minutes in a microcemtrifuge. The pellet was washed with 70% ethanol, air-dried and resuspended in 10-50µl TE or water. This protocol yielded up to 5µg plasmid DNA.

2.4.1.3 Large scale preparation of plasmid DNA using Qiagen columns (maxiprep)

The maxi-prep protocol was used to extract plasmid DNA from 500ml bacterial cultures yielding up to 500µg DNA. This was performed in the same way as the miniprep but with the following exceptions: Cells were harvested in 250ml polypropylene Sorvall bottles by centrifugation at 8000rpm for 5 minutes in a Sorvall centrifuge set at 4°C. The pellet was resuspended in 10ml P1 buffer and transferred to 30ml Sorvall tubes, and 10ml volumes of P2 and P3 buffers were used in the extraction. A Qiagen-tip 500 was equilibrated with 10ml QBT, the bound DNA washed with 30ml volumes of QC, and eluted in 15ml buffer QF. Finally the DNA was dissolved in 25-200µl TE or water.

2.4.1.4 Isolation of PAC and BAC DNA

The mini-prep or maxi-prep methods were followed for the extraction of PAC and BAC DNA, but the DNA solutions were not passed through a Qiagen column. Instead, DNA was precipitated directly by adding an equal volume of ice-cold isopropanol and placing it on ice for 5-20 minutes, or was phenol-chloroform purified and ethanol precipitated.

2.4.1.5 Crude yeast and YAC DNA preparations

Total yeast strain *S.cerevisiae* strain AB1380 DNA and crude YAC DNA preparations were made using the protocol described by Polaina and Adam (1991), with minor adaptations. A single yeast colony or 10µl glycerol stock was inoculated into 5ml YPD broth (50µg/ml ampicillin was included in the broth for culture of YAC-bearing clones) and grown overnight at 30°C in a rotating incubator. The cells were collected by spining at 3000rpm in a Sorvall centrifuge for 10 minutes and the broth removed. The cells were resuspended in 200µl water and transferred to a 1.5ml eppendorf tube. 200µl CiDIS, 0.35g 1000µm acid-washed beads (Sigma) and 200µl phenol/chloroform (1:1) were added and the cells vortexed for 2 minutes to break down the cell walls. The mix was spun at 13000rpm for 5 minutes in a microcentrifuge, and the aqueous phase containing the yeast DNA was removed to a clean tube. RNase to 50µg/ml was added and

the reaction incubated at 37°C for 15minutes. The DNA was phenol/chloroform purified, ethanol precipitated and resuspended in 50-200µl TE.

2.4.1.6 Preparation of YAC DNA in agarose blocks for PFGE

Total yeast DNA was prepared in agarose blocks for PFGE to isolate the artificial chromosome using the method of Sheehan and Weiss (1990). A single yeast colony was used to inoculate 5ml AHC^- media containing 50µg/ml ampicillin and grown in a shaking incubator at 30°C for 24 hours. The 5ml culture was then used to inoculate 200ml AHC^- media in a flask, and grown for a further 36 hours at 30°C. The cells were pelleted in 50ml falcon tubes at 2100rpm for 5 minutes and the media discarded. The yeast pellets were resuspended in 20ml 0.1M EDTA (pH 8.0), and incubated at room temperature for 30 minutes. The cells were pelleted again and resuspended in 3.2ml SCEM (spheroplast medium) containing 1.2mg lyticase. The yeast cells were incubated at 30°C for 1 hour, allowing the lyticase to digest the cell wall. The spheroplasts were pelleted by centrifugation at 2000rpm and resuspended in 2.5ml sorbitol. 1% low melting point (LMP) agarose was made up in 1M sorbitol and held at 50°C. The spheroplasts were mixed with an equal volume of agarose and 80-100µl dispensed into pre-cooled block-formers to set. The blocks were removed from the moulds, placed in fresh Proteinase K solution, and incubated at 50°C for 48 hours. The blocks were washed several times with TE pH 8.0 prior to storage in TE at 4°C.

2.4.2 Purification of DNA

2.4.2.1 Purification of DNA in solution by phenol-chloroform extraction and ethanol precipitation

An equal volume phenol pH7.4/chloroform/isoamyl alcohol in a ratio of 12:12:1 was added to the DNA solution and vortexed. The upper aqueous phase containing the DNA was separated from the organic phase by centrifugation at 13000rpm in a microcentrifuge for 5 minutes, then transfered to a clean tube. This removes protein debris from the DNA. An equal volume of chloroform/isoamyl alcohol (24:1) was added, mixed and centrifuged as before, and the aqueous solution removed. The DNA was precipitated by adding 1 tenth the volume of 3M sodium acetate (pH 4.8-5.2), followed by 2 volumes 100% ethanol, and incubated at -20°C for 15 minutes. The DNA was pelleted by spinning at 13000rpm for 15 minutes. The ethanol was removed and the pellet washed with 70% ethanol and air-dried. The DNA was resuspended in TE or water and stored at -20°C.

2.4.2.2 Purification of DNA using the Geneclean kit II

The Geneclean kit II was used to purify DNA in solution and from agarose gel post-electrophoresis. The Geneclean kit removes salts and PCR primers from DNA solutions, and is specifically designed to recover DNA fragments of over 500bp, with no upper size limit.

To extract DNA from solutions, 3 volumes of NaI were added to the solution. A minimum of 5µl Glassmilk (silica matrix in suspension which binds DNA) for 5µg DNA (with an additional 1µl/µg added if necessary) was added and vortexed. The solution was kept on ice for 5 minutes to allow binding of the DNA to the silica matrix. The matrix was pelleted by centrifugation at 13,000rpm for 10 seconds and the NaI removed. The DNA-bound matrix was washed three times with New wash (a wash provided with the kit that contains ethanol). Each time the pellet was resuspended in New wash then pelleted by spinning for 10 seconds. The pellet was allowed to air-dry for 5 minutes after the final wash. The DNA was eluted by resuspending the pellet in a minimum of 10µl water or TE and incubation at 55^{0}C for 5 minutes. The matrix was pelleted by spinning at 13,000rpm for 30 minutes, and the DNA in solution removed to a clean tube.

To recover DNA from agarose, 3 vols NaI was added if the gel was in TAE buffer, or 0.5 vols TBE modifier followed by 4.5 vols NaI if the gel buffer was TBE. The gel was dissolved in the NaI solution by incubation at 55°C for 5 minutes. Glassmilk was added and the purification continued as above.

2.4.2.3 Purification of PCR products for sequencing and cloning

S-300 microspin columns, pre-packed with sephacryl resin equilibrated in TE buffer, were used to purify PCR products of over 200 bp in length in 50µl volumes, from the primers and salts in the reaction. The columns were vortexed to mix the sephacryl, then the excess TE buffer was removed by snapping off the bottom closure and spinning the column (placed in an eppendorf with the lid removed) for 1 minute at 3000rpm in a microcentrifuge. The PCR reaction was removed from beneath the oil and loaded onto the column. The column was spun for 2 minutes at 3000rpm, and the purified PCR product collected in the eppendorf tube beneath.

2.4.2.4 Gel-purification of DNA

Gel purification was used to excise cloned inserts from plasmid vectors and YAC DNA from host yeast chromosomes after separation of the fragments by electrophoresis. A sterile scalpel was used to cut the appropriate DNA band from the gel, under UV illumination.

2.4.3 Preparation of RNA from fetal tissues and cell lines

2.4.3.1 Collection of fetal tissues for RNA extraction

Fetal tissues were removed from the product of suction-termination of pregnancy and rinsed in sterile phosphate buffered saline (PBS). The tissues were placed in labelled 4ml sterilins and 'flash-frozen' in a flask of liquid nitrogen for transportation, then stored at -70°C prior to use.

2.4.3.2 Creation of a ribonuclease-free environment

RNA is highly sensitive to ribonuclease activity, and so all plastic ware and solutions were treated with diethyl pyrocarbonate (DEPC). DEPC was added to all solutions (except those containing TRIS-HCl) to 0.05% and incubated overnight at room temperature prior to autoclaving. This inactivates the DEPC to carbon dioxide and water. Plastic ware and utensils were soaked overnight in DEPC-water before drying and autoclaving. Aerosol resistance tips (ART) which are sterile and nuclease-free were used for all dealings with RNA.

2.4.3.3 Extraction of total RNA

The Trizol reagent (Life Technologies), which combines phenol and guanidine isothiocyanate, was used to isolate total RNA. Up to 100mg tissue sample was placed frozen, into 1ml Trizol in a large eppendorf, and homogenised using a hand-held motor-driven homogeniser (CAMLAB) supplied with disposable rotor tips. Lymphoblastoid cells were pelleted by centrifugation at 6000rpm for 10 minutes and the media removed. 1ml Trizol per 10^7 cells was added. This lyses the cells immediately. The samples were incubated for 5 minutes at room temperature allowing complete dissociation of nucleoprotein complexes to occur. 200µl chloroform per ml Trizol was added and mixed by vortexing. The samples were incubated for at room temperature for a further 5 minutes then centrifuged at 12000rpm for 15 minutes in a microcentrifuge at 4°C. The aqueous phase containing the RNA was removed into a clean tube and 0.8 volumes isopropanol

added. The RNA was precipitated at room temperature for 10 minutes and pelletted by centrifugation at 12000rpm for 15 minutes at 4°C. The RNA was washed in 70% ethanol, dried and resuspended in DEPC-treated water with 1µl RNasin ribonuclease inhibitor. The RNA was stored at -70°C.

2.4.3.4 DNase treatment of RNA

If performing RT-PCR, RNA samples were DNase treated to eliminate genomic DNA contamination. The RNA was incubated with an equal volume of DNase mix containing 20mM $MgCl_2$, 1mM DTT, 10U RNasin, 10µg DNaseI for 15 minutes at 37°C. The DNase was removed by phenol (pH4.3)/ chloroform purification, and the RNA precipitated with a tenth volume of 3M sodium acetate pH 4.8-5.2 and 0.8 volumes isopropanol, and washed with 70% ethanol.

2.4.3.5 Isolation of mRNA

Polyadenylated messenger RNA (poly A+ mRNA), which comprises about 1-5% of total mammalian RNA, was separated from ribosomal and transfer RNA using oligo $(dT)_{25}$ paramagnetic Dynabeads as instructed by the maufacturer.

The RNA was brought to a volume of 100µl (containing up to 75µg total RNA), with elution buffer. The RNA was heated at 65°C for 2 minutes to disrupt the secondary structure. 200µl of Dynabeads in solution was dispensed into an eppendorf standing in a magnetic particle concentrator rack (MCP-E). The beads were collected at the bottom of the tube and the storage solution removed. The beads were washed in 100µl 2X binding solution, collected by standing the tube in the MCP-E for 30 seconds, and the wash removed. The beads were then resuspended in a further 100µl 2X binding solution and the RNA added to the beads. The polyA+ mRNA was allowed to bind to the beads over 5 minutes. The beads plus bound mRNA were collected by standing the tube in the MCP-E, and the binding solution was removed. The beads were washed twice with 200µl washing buffer, collected at the bottom of the tube, and all traces of buffer removed. The mRNA was eluted by resuspending the beads in a minimum of 10µl elution buffer and heating to 65°C for 2 minutes. The mRNA was removed from the beads by standing the tube in the MCP-E, and the elution buffer containing the mRNA removed to a clean tube. mRNA was stored at -70°C with 0.5µl RNasin.

2.4.4 Quantitation of DNA and RNA in solution

2.4.4.1 Spectrophotometry

The concentration of DNA and RNA in aqueous solutions was measured by measuring the UV absorbance of the DNA/RNA solution at 260nm. Pure DNA solutions have an absorbance peak at 258nm where 1 $O.D._{258}$ = 50μg/ml and for pure RNA, 1 $O.D._{260}$ = 40μg/ml. The purity of the samples was established with additional spectrophotometric assays at 240nm for estimation of polysaccharides and at 280nm for proteins. Pure solutions of DNA had a ratio of 1.8:1 whilst pure RNA solutions had a ratio of 1:2:1 for wavelengths 240:260:280nm.

500μl DNA and RNA solutions, diluted by a factor of 100, were placed in a quartz cuvette and measured against a 'blank' of the diluent using a GeneQuant RNA/DNA spectrophotometer (Pharmacia).

2.4.4.2 Ethidium bromide staining

The concentration of small quantities of DNA and cDNAs was estimated by spotting 1μl of the sample onto agarose gel containing 0.5μg/ml ethidium bromide, set in a small petri-dish, against 1μl spots of a set of DNA concentration standards ranging from 10ng - 1μg/μl. The intensity of staining of the sample was visually compared with that of the standards under UV light.

2.4.5 Restriction digestion of DNA

2.4.5.1. Digestion of DNA in solution

Restriction digestion cleaves DNA at specific sequence recognition sites and was used for excision of cloned inserts from vector DNA and fragmentation of genomic DNA and genomic clones for Southern blot analysis. Reactions were performed in 1X restriction buffer and 1U enzyme per μg DNA up to 1 tenth the total volume. Digestion of plasmids for excision of inserts was performed in 20μl volumes, using 2-5μg DNA. For Southern blots, 200ng YAC, PAC, BAC or cosmid DNA was digested in 20-30μl. Digestion of 5μg genomic DNA was performed in 40μl reactions. Samples were incubated for 3 hours (plasmids, PAC and BAC) or overnight (YAC and genomic DNA) at 37°C unless, otherwise specified by the manufacturer of the restriction enzyme.

2.4.5.2 Digestion of YAC DNA in agarose blocks

Excised agarose blocks containing YAC DNA after PFGE were pre-equilibrated in 1X restriction buffer for 1 hour prior to digestion. Reactions were performed in a total of 300µl, including the 100µl block, for 16 hours at 37°C with 10-20U enzyme.

2.4.6 Polymerase chain reaction (PCR) amplification of DNA

To minimise the risk of DNA contamination, precautions were taken whilst setting up PCR reactions, which included the use of barrier tips, gloves, setting the reactions up in an area separate to where reaction products were processed, and keeping water and reagents separate for PCR.

PCR reactions were carried out in 50 or 100µl volumes containing a final concentration of 1X NH_4 reaction buffer, 25-50pmoles of each oligonucleotide primer, 0.2mM each dNTP, 1.5mM $MgCl_2$, 1-2U DNA *Taq* Polymerase, and 25ng plasmid DNA, 100ng cosmid, PAC or YAC DNA or 200ng human genomic DNA. The volume was made up with sterile distilled water. A reaction 'master mix', containing all reagents except the enzyme and DNA was set up in a volume sufficient to process the necessary number of individual samples simultaneously. The 'master mix' was aliquotted into 0.5ml eppendorf tubes, or 96 well amplification plates, and the DNA sample then added to individual reactions. Water replaced DNA in one tube to act as a negative control, for detection of PCR contamination. The reactions were overlaid with 2-3 drops of mineral oil to prevent evaporation.

The PCR samples were denatured at the start of the PCR cycle by heating the samples to 94°C for 5 minutes during which time the *Taq* Polymerase was added beneath the oil. Generally, amplification reactions were carried out for 30-35 cycles with denaturation at 94°C for 1 minute, followed by annealing for 1 minute and then extension at 72°C for 1 minute. To amplify large fragments of DNA, the extension time was increased; *inter*-Alu PCR reactions were performed with an extension of 4 minutes. A final extension period of 10 minutes was performed prior to cooling the reaction. The annealing temperature, or Tm, varied according to oligonucleotide primer composition. The Tm for each primer was provided by Genosys, but could be calculated using the following equation (Maniatis et al., 1982):

(69.3 + (0.41 x G + C%) - 650 / nucleotide) -12

Where the Tm differed between pairs of primers used in the same reaction, the lower of the two temperatures was used.

PCRs in 96 well plates were performed on a Hybaid Omnigene *Omn-E* PCR machine, with the 96-format block attached. Here the *Taq* Polymerase was added to the master-mix. PCRs in 0.5ml eppendorfs were carried out on a Techne Thermal Cycler PHC-3 or *Gene E.*

2.4.7 Reverse-Transcriptase PCR (RT-PCR)

2.4.7.1 First-strand cDNA synthesis

First strand cDNA synthesis was performed in 20µl volumes with 1-2µg total RNA or 0.25µg mRNA and a final reaction mix of 1X first strand buffer, 500ng oligo $dT_{12\text{-}18}$ or 250ng random primer, 10mM DTT, 0.2mM each dNTP, 4U RNasin ribonuclease inhibitor, and 100U MMLV Reverse Transcriptase enzyme. First, the RNA and DEPC water was heated to 70°C on a Thermal Cycler, to denature the secondary structures. The RNA was then cooled to 37°C and remaining reagents in the reaction mix were added. Reverse transcription was performed for 1 hour at 37°C, then terminated by heating to 75°C for 10 minutes. For each RNA sample, a 'RT minus' control reaction was set up for the subsequent detection of genomic DNA contamination. The reaction was performed with water in place of RT.

2.4.7.2 RT-PCR

2-5µl of the first-strand cDNA (and the RT minus control) was used as template for amplification by PCR. RT-PCR was performed essentially the same as for PCR amplification of DNA. To check the integrity of the RNA/cDNAs, and to detect any genomic contamination of the cDNA sample, RT-PCR amplification of the 'housekeeping' cDNA, *HPRT*, was performed for each cDNA template and its corresponding RT minus control. The *HPRT* cDNA fragment was 200bp, and genomic contamination was detectable as a 1.5kb product in the RT minus control.

2.4.8 Agarose gel electrophoresis

2.4.8.1 Agarose gel electrophoresis of DNA

Agarose gel electrophoresis was routinely used to size fractionate DNA ranging from 100bp to 15kb. The agarose concentration was adjusted according to the size of the fragments to be separated. 0.8% gels were used to separate large DNA fragments (2-20 kb) including for Southern blotting, and 1.5-2% gels used to separate smaller DNA fragments (<2kb). The agarose was melted in 1X TBE or 1X TAE (for Southern blot or gel purification) buffer in a microwave oven. The gel mix was allowed to cool to ~50°C, ethidium bromide added to a final concentration of 0.5μg/ml, and the gel poured into the assembled gel tray. Gels of 8 X 10.5cm were used with 8 to 16-tooth combs for visualisation of PCR products, 20 X 20cm gels with 20-tooth combs for Southern blotting, and 20 X 26 cm gels used with three 36-tooth combs for electrophoresis of PCR products performed in 96-well plates.

Linear DNA carries an overall negative charge and migrates through the gel towards the anode, at a rate directly proportional to $1/\log_{10}$ x molecular weight. The size of the DNA was estimated with reference to DNA size markers, which consist of fragments of known but differing size. These are electrophoresed alongside the sample and display a given banding pattern or ladder. A ruler was placed beside the DNA marker to aid in sizing the fragments if the gel was to be blotted. Samples were loaded with 1X loading buffer and run at 50-100V until the bromophenol blue dye had run two thirds of the way through the gel.

2.4.8.2 Pulsed field gel electrophoresis (PFGE)

Large DNA molecules of >1Mb can be resolved in the presence of an electric field that changes orientation with respect to the gel (Swartz and Cantor, 1984). This is achieved by alternating the current between two separate electrodes, facing north-south and east-west with respect to the gel. The length of time a fragment takes to reorientate and continue to move through the gel depends on its size. The level and duration of the voltage pulses applied across the sets of electrodes determine the separation of the DNA on the gel. Typically a 50-90 second switch time at 200V for 18-30 hours will resilved fragments from 300kb to 2Mb.

A 1% agarose gel in 0.5X TBE buffer was poured into the tray, with the gel-former in place. A trough was formed at the top of the gel. Agarose blocks containing the yeast DNA were loaded into the trough. The gel was submerged into 0.5X TBE buffer in

a Pharmacia-LKB Pulsaphor tank, pre-cooled to 8°C. The hexagonal electrode was slotted into place around the gel. The buffer was circulated continuously and kept at 8°C. An electric field pulse-switcher (Pharmacia) was used to alternate the current at fixed intervals. Host yeast strain *AB1380* was used as a size marker to distinguish the YAC DNA from the native yeast chromosomes, and run alongside the YAC DNA. After electrophoresis the gel was stained in 0.5X TBE containing 5μg/ml ethidium bromide for 20 minutes on a rotating platform for visualisation and photography. The artificial chromosome band was excised, and the DNA digested in agarose blocks, prior to purification from the gel.

2.4.8.3 Electrophoresis of RNA for northern blotting

A 140ml 1% agarose/formaldehyde gel was prepared by boiling 1.4g agarose with 115ml 1X MOPs buffer. The agarose was allowed to cool to below 50°C and 25ml 37% formaldehyde was added. The gel was poured and allowed to set. Due to the toxicity of formaldehyde, these steps were carried out in a fume hood. Ethidium bromide was not added to gels that were to be blotted.

20-60μg total RNA was added to twice the volume of RNA sample buffer. This was heated to 65°C for 5 minutes, cooled to room temperature, and a tenth volume of RNA loading buffer added to the sample. An RNA size marker was treated similarly, except that 0.5μl 10mg/ml ethidium bromide was added prior to loading, in order to visualise the ladder. Where RNA test samples were electrophoresed in duplicate to check the integrity of the RNA, ethidium bromide was added to the individual RNA samples. The gel was run in 1X MOPs buffer at 50V, until the bromophenol blue was about a third of the way down the gel.

2.4.9 Visualisation and photography of DNA and RNA

DNA and RNA were stained with ethidium bromide enabling visualisation under ultraviolet (UV) light at 302nm. After electrophoresis gels were placed on a UV transilluminator and photographed using a UVP videocopy processor (Mitsubishi) or Polaroid camera.

2.4.10 Southern blotting

DNA was transferred onto nylon membranes by the method of Southern (1975). The DNA was denatured by soaking the gel in denaturation solution for 45 minutes on a rotating platform. A tray containing a sponge overlaid with 3MM Whatman paper (to act as a wick) was filled with 20X SSC and the gel placed on the platform. Hybond N^+ nylon membrane (Amersham) was cut to the size of the gel, prewet in 2X SSC and placed on top of the gel. Clingfilm was placed around the 4 sides of the gel to prevent evaporation and the fluid from by-passing the membrane. Three pieces of 3MM paper were prewet in 2X SSC and placed over the membrane, followed by a 5-inch stack of paper towels. Blots were left for at least 6 hours to allow transfer of the DNA to the membrane. The blot was disassembled and the positions of the wells were marked on the filter. The membrane was rinsed in 2X SSC and the DNA fixed by soaking the membrane in 0.4M NaOH for 20 minutes. The membrane was neutralised by rinsing in 2X SSC until it reached pH7.

2.4.10.1 'Sandwich' Southern blotting

'Sandwich' Southern blotting was performed to produce duplicate filters of DNA from a single electrophoresed gel. This method was suitable for transferring PCR products (where a single-sized DNA product is present in large quantities) to filters for more efficient hybridisation screening. The blot was set up in the form of a sandwich: About 5 inches of paper towels formed the base of the blot, onto which was laid 3 squares of 3MM Whatman paper prewet in 2X SSC. A 'lower' sheet of Hybond N was prewet and placed over the squares of 3MM paper. The denatured gel was placed ontop of the 'lower' filter, and a second 'upper' piece of prewet membrane was laid on the gel. This was followed by a further 3 sheets of prewet 3MM paper and finally, another wadge of paper towels. A weight was placed ontop of the 'sandwich' blot and transfer allowed to occur for at least 2 hours. The two blots thus formed a mirror image of the gel. On dismantling, the blots were labelled 'upper' and 'lower', and the positions of the gel lanes were marked.

2.4.11 Northern blotting

RNA for northern analysis was transferred to Hybond N membrane using a similar method to that of Southern blotting. The electrophoresed gel was photographed with a ruler beside the RNA marker to determine the migration distance of the fragments. Where

RNA samples had been stained with ethidium bromide, the migration distance of the 28S and 18S ribosomal RNA bands was also determined. The gel was transferred directly onto the blotting apparatus, which was assembled in exactly the same way as a Southern blot, using Hybond N membrane. The filter was removed from the blot, rinsed in 5X SSC and placed on a piece of 3MM Whatman paper. The RNA was fixed to the membrane by UV cross-linking. The blot was placed 'RNA-side-up' in a UV Crosslinker (Stratalink) and the Auto-Crosslink setting was used to irradiate the blot with 120mJoules.

2.4.12 Hybridisation of Southern and northern blots and library screen filters with ^{32}P labelled probes

2.4.12.1 Oligolabelling of probes with ^{32}Pα-dCTP

Probes were labelled by random priming which uses the *E.coli* large fragment DNA polymerase I (Klenow fragment). This initiates second strand DNA synthesis by random-primer extension with the single-stranded DNA serving as template. Radiolabelled dCTP becomes incorporated into the DNA probe during second strand synthesis.

About 20ng monotypic probe or 100ng genomic probe was denatured in a 35µl volume by boiling for 5 minutes and placed on ice for 2 minutes to prevent renaturation. 10µl 5X oligo-labelling buffer (OLB), 1µl BSA (10mg/ml), 3µl ^{32}Pα-dCTP and finally 1µl (2U) Klenow fragment (Bioline) was added to the probe bringing the volume to 50µl. The labelling reaction was incubated at room temperature in a perspex container for 3 hours. The labelled probe was separated from unincorporated nucleotides by Sephadex filtration. A 1ml syringe was plugged with a glass bead, placed in a disposable test tube, and filled with 1ml Sephadex G50 equilibrated in TE buffer. The column was equilibrated with 300µl TE. The volume of the probe was brought up to 300µl with TE and loaded onto the column. The eluate was discarded, the column transferred to a clean test tube and the probe displaced from the column by adding 300µl TE. The level of radioactive incorporation was measured by counting the emission from a 4µl aliquot of the labelled probe in a calibrated Bioscan QC2000 β-counter. A value exceeding 100,000 counts per minute indicated successful incorporation of ^{32}Pα-dCTP. The radiolabelled probe was denatured by boiling for 5 minutes, then placed on ice.

2.4.12.2 Preannealing of DNA probes to quench repeat sequences

DNA probes such as YACs, PACs and BACs were preannealed with total human DNA prior to filter hybridisation, to suppress repeat sequences. This was performed for mapping of PACs and BACs to the 15q11-13 YAC contig and for cDNA library screening using YAC, PAC and BAC DNA to identify clones potentially mapping to the PWS and AS region. The 300µl radioactively labelled eluted probe was added to 30µl (10µg/µl) Cot 1 or human placental DNA, boiled for 8 minutes and placed on ice. The probe was then added to 45µl 20X SSC, and 1.43ml prehybridisation buffer prewarmed to 65°C. The probe was preannealed at 65°C for 90 minutes and added to the prehybridisation buffer in the hybridisation bottle.

2.4.12.3 Hybridisation of Southern and northern blots

Membranes were wetted in 2X SSC, rolled round a pasteur pipette and unrolled into hybridisation bottles (Hybaid). Where more than one filter was placed in a single bottle, a piece of gauze was placed between them. 15ml prehybridisation solution was added to each bottle, which was then placed horizontally on the rotating rack in an oven for 3 hours. Southern blots were prehybridised at 65°C and northern blots at 42°C. The denatured radiolabelled probe was added to the bottles containing the prehybridised filters. These were returned to the rotating rack in the oven. The hybridisation reaction was left overnight. Hybridised filters were removed from the bottles and washed in sealed containers firstly in 3X SSC/0.1% SDS at room temperature for 30 minutes to remove background radioactivity. Filters were then washed to varying stringency depending on the size of the probe and the hybridisation reaction. Southern blots were washed at 2X, 1X 0.5X, 0.2X or 0.1X SSC/0.1% SDS for 20 minutes in a shaking waterbath at 65°C. Northern blots were washed serially in 3X, 2X, 1X, 0.5X and 0.1X SSC/0.1% SDS at 42°C for 20 minutes until the background signal was reduced. The filters were rinsed in 2X SSC at room temperature to remove the SDS and wrapped in Saran film to prevent them from drying out. Hybridisation signals were detected by autoradiography.

2.4.13 Autoradiography

Autoradiography was performed using X-OMAT film (Kodak) to detect radiolabelled DNA. Filters hybridised with ^{32}Pα-dCTP labelled probes were placed in an autoradiographic cassette containing two intensifying screens. A piece of film was lain directly over the filter, between the two screens. The cassette was placed at -70°C and the

film exposed for between 1 hour and 3 days for Southern blots, and for 2 days to 1 week for northern blots, depending on signal strength. For the detection of DNA sequence where ^{35}S was incorporated, the dried sequencing gel was placed in an autoradiographic cassette without intensifying screens. The film was exposed at room temperature for 1-3 days. Films were developed using a Xray film processor (Fuji).

2.4.14 Determination of size of DNA and RNA fragments identified by Southern and northern analysis

The size of unknown DNA and RNA fragments detected by Southern and Northern analysis was determined by plotting a graph of $1/\log_{10}$ size (kb) versus migration distance (cm) for the DNA or RNA marker included on the gel. This was performed using photographs of the gels with a ruler placed alongside the marker, or by hybridisation of the DNA or RNA markers with λ DNA followed by autoradiography. The migration distance of the unknown fragment was measured from the autoradiograph using a ruler and the 1/log10 value for fragment size read from the vertical axis of the graph.

2.4.15 Cloning of DNA in plasmids

2.4.15.1 Cloning into plasmids by restriction digestion

2-5μg of the plasmid vector was digested in a 30μl volume with the appropriate restriction enzyme at the unique site in the polylinker. The linear vector DNA was then dephosphoylated by additing 20U (1μl) Calf intestinal alkaline phosphatase to the digestion reaction and incubating at 37°Cfor 1 hour. The reaction was stopped with the addition of 1μl 0.5M EDTA. The phosphatased DNA was purified using phenol:chloroform and ethanol precipitated. Dephosphorylation of the vector removes the terminal 5' phosphate, which suppresses self-ligation of the vector and increases the efficiency of recombinant formation.

Ligation reactions were performed in minimal volumes to concentrate the DNA, with 4-10 times the amount of insert DNA to vector DNA. 20μl ligation reactions were set up with 10μg insert DNA, 1μg dephosphorylated vector DNA, 2μl 10X ligation buffer (including 10mM ATP), 1-2μl (1-2U) *E.coli* DNA Ligase enzyme (added last). The ligation reaction was incubated for 2-16 hours at 16°C. Control reactions included: 1) dephosphorylated vector alone to check the level of self-ligation 2) untreated linearised vector DNA to check efficacy of dephosphorylation reaction.

The plasmid plus insert was transformed into competent *E.coli* cells. The cells were removed from storage at -70°C and kept on ice to prevent them from rising above 0°C. 10µl of a tenfold dilution of the ligation mix was added to 200µl competent cells and left on ice for 30 minutes. The cells were heat-shocked by placing them at 42°C for 90 seconds, then replacing them on ice for 2 minutes. This induces the uptake of the plasmid DNA by the cells. LB broth (1ml), containing no antibiotics was added, and the cells incubated at 37°C, on a shaker, for 1 hour. The cells were spun at 7,500rpm for 5 minutes and the top 900µl media removed. The cells were resuspended in the remaining 100µl of media and plated out onto a LB agar plate containing the appropriate antibiotic(s). To assesss the efficiency of transformation, and proportion of self-ligated transformants, the vector alone was used as a control.

2.4.15.2 Cloning of PCR products using the 'LigATor' kit

The 'LigATor plus ligase' kit was used to clone PCR products. The LigATor kit exploited the addition of the single 3' adenine overhangs produced at the ends of PCR products by *Taq* DNA polymerase. The plasmid vector pTAg was provided linearised with single thymine overhangs to complement the ends of the PCR products, facilitating ligation. pTAg contains the α-lactamase gene (LacZα), giving blue coloured colonies in the presence of X-galactosidase and β-D-isoprpyl-thiogalactopyranoside (IPTG). Formation of recombinants disrupts the LacZα gene, producing white colonies. pTAg also contains the selectable markers which confer ampicillin and kanamycin resistance on the bacterium.

Up to 2µl direct PCR product or approximately 50ng purified insert in a volume of up to 6.5µl water was added to 1µl 10X Ligase buffer, 1µl (50ng) pTAg, 0.5µl 10mM ATP and 0.5µl 0.1M DTT. About 2U, 0.5µl T4 DNA Ligase was mixed in to make the volume up to 10µl. The ligation reaction was incubated overnight in a 16°C waterbath. A control insert, with single A overhangs, was included in the kit to test ligation efficiency. Test ligations were set up with 2µl control insert.

The competent cells, provided in the kit, were placed on ice to thaw. 20µl cells were placed into a pre-cooled sterile 1.5ml eppendorf. 1µl ligation reaction was added to the cells, gently mixed in, and incubated on ice for 30 minutes. The cells were placed at 42°C for 40 seconds then directly onto ice for 2 minutes, to heat-shock them. 80µl SOC medium was added to the cells, making the volume up to 100µl, and incubated at 37°C on a rotary shaker (set at 200-250rpm) for at least 1 hour. An Amp^r test plasmid was provided

with the kit to test transformation efficiency of the cells. In place of a ligation reaction, 1μl test plasmid was used in the transformation.

The transformants were spread on large LB agar plates containing 0.5μM IPTG, 80μg/ml X-Gal, 20μg/ml tetracycline and 50μg/ml kanamycin (ampicillin for test plasmid). (Use of tetracycline ensured that the F^1 plasmid bearing the LacZα gene was maintained in the competent cells). Usually 50μl - half the transformation mix- was spread on each plate. The plates were incubated inverted, at 37°C overnight. To enhance colony colour, the plates were cooled to 4°C for 1 hour. At least 50% of the colonies were white when the ligation and transformation reactions were optimal, although up to 10% white colonies were predicted to be non-recombinants.

2.4.16 Transposon-integration into cloned sequences by bacterial mating

In order to sequence cloned fragments of DNA by the transposon-facilitated method, transposons are inserted into the cloned sequence (Strathmann *et al.*, 1991). This is achieved by bacterial mating between designated strains, and was carried out for sequencing of the evolutionarily conserved genomic fragment, P1.1-6. P1.1-6 was cloned into the pMOB miniplasmid vector, which presents the insert as the target for transposition. P1.1-6 was firstly excised from pBlueScript KS2 in which it was cloned, by digestion with *Pst*I followed by gel fractionation, and purified. P1.1-6 was ligated into the pMOB miniplasmid via the *Pst*I site. The pMOB-P1.1-6 construct was transformed into the highly competent DH5α cells, which readily uptake plasmid DNA under the high salt conditions of a ligation mix. Transformants were grown on agar plates containing 50μg/μl ampicillin to select for colonies with ampicillin resistance conferred by pMOB. Six colonies were picked and a subclone containing the pMOB plasmid with P1.1-6 insert was identified by PCR-amplification with T3 and T7 primers. 1μg miniprepped plasmid DNA (pMOB-P1.1-6) was from one DH5α subclone was transformed into the less competent F+ donor DPWC cells. Six DPWC colonies were tested for containment of the pMOB-P1.1-6 construct by PCR, as before.

The pMOB-P1.1-6 construct forms a cointegrate with the transposon bearing F factor in these cells. Transposon mobilisation into the target DNA (P1.1-6) occurs by mating the DPWC cells with recipient F- JGM cells, allowing conjugal transfer of the cointegrate into the recipient strain. The cointegrate is resolved back to its

derivative plasmids in these cells, leaving a copy of the transposon in the F+ factor and the target DNA sequence within pMOB.

The DPWC cells were cultured overnight in LB containing ampicillin, and the JGM cells were grown overnight in LB containing 50μg/ml kanamycin. 100μl each culture was inoculated into 2ml LB without any antibiotic, and incubated for 3 hours on a rotating platform at 37°C. 100μl of 100-fold dilution of the mating mix was spread onto agar plates containing 50μg/ml each of ampicillin and kanamycin. This selects for JGM cells containing the pMOB plasmid. The subclones were picked for transposon mapping by PCR with transposon primer GDIR and pMOB primers LREV and LUNI.

2.4.17 Screening of cDNA libraries

Approximately 1 x 10^6 plaque forming units (p.f.u.) from each of the 9 week fetal brain cDNA libraries Br32 and Br33 were screened by hybridisation with a probe derived from the conserved genomic *Pst*I fragment, P1.1-6.

2.4.17.1 Preparation of plating cells

An inoculating loop (5μl) of XL1-blue cells stored in glycerol was streaked onto LB agar plates with tetracycline (12.5μg/ml final concentration) and incubated at 37°C overnight. A single colony was used to inoculate a 50ml volume of LB, supplemented with 0.2% maltose and 10mM $MgSO_4$ and grown overnight at 30°C. The cells were pelleted by centrifugation at 6000rpm for 5 minutes and resuspended in an equal volume of 10mM $MgSO_4$.

2.4.17.2 Library titering

Phage cDNA library titering was achieved using serial dilutions. The phage were diluted in SM buffer and an aliquot (usually 10μl of 500μl) added to 200μl of plating cells and gently mixed. After incubation of phage and host cells for 15 minutes at 37°C, the suspension was added to 10ml of top agarose (pre-heated to 48°C) and poured into twenty 140mm diameter petri dishes (Sterilin) containing NZY agar. Plates were left to stand for 10 minutes, and incubated overnight at 37°C. The titre, in p.f.u. per ml, was determined by counting the plaque numbers contained on plates of known phage dilution.

2.4.17.3 Library plating

Twenty 140mm diameter petri dishes (Sterilin) per cDNA library were prepared with 100ml NZY bottom agar. Approximately $5x10^4$ recombinant bacteriophage were mixed with 800µl of plating cells in 10mM $MgSO_4$ and incubated at 37°C for 15 minutes. Each aliquot of infected bacteria was added to 7ml of melted (48°C) top agarose and poured onto the NZY bottom agar layer of each petri dish. The plates were incubated overnight at 37°C.

2.4.17.4 Plaque blotting

Duplicate 130mm diameter Hybond-N 'lifts' were taken directly from each plate. The membrane and agar were marked asymmetrically using a sterile needle to allow orientation. The first membrane (marked 'A') was placed on a library plate for 30 seconds and the second, duplicate membrane (marked 'B'), for 2 minutes. The membranes were placed 'DNA side up' sequentially onto 3MM Whatman paper containing denaturing solution for 2 minutes and neutralising solution for 5 minutes. Membranes were rinsed in 3X SSC, dried and baked for 2 hours at 80°C. The filters were hybridised in duplicate with the appropriate $^{32}P\alpha$-dCTP labelled DNA probe. Only hybridisation signals detected in duplicate were considered a positive identification.

2.4.18 Dideoxy-DNA sequencing

2.4.18.1 Radioactive-labelled (manual) sequencing

Dideoxy DNA sequencing was carried out using the USB Sequenase Version 2.0 DNA Sequencing Kit according to the manufacturers instructions, on a Biometra TRIO-Thermoblock, using radioactive ^{35}S-dATP as label.

Between 5 and 10µg plasmid 'miniprep' DNA in 15µl TE was denatured with 5µl freshly made 4M NaOH for 10 minutes at room temperature, to produce the sequencing template. The volume was brought up to 90µl with water, then neutralised with the addition of 10µl 3M Sodium acetate. The single-stranded DNA was then precipitated with 2 volumes 100% ethanol and centrifugation at 13,000rpm for 15 minutes. The template was washed with 70% ethanol, air dried and stored dessicated.

An annealing step was included to preanneal the sequencing primers to the single-stranded DNA template: The dessicated DNA pellet was resuspended in 14µl water and divided into two sequencing reactions (one for each primer). For each reaction 2µl sequenase buffer and 10 moles sequencing primer were added to the template in a 10µl

volume. This was heated to 65°C for 2 minutes then allowed to cool over 30 minutes to room temperature. For the elongation reaction, 1µl 0.1M DDT, 2µl 5X labelling mix, 0.5µl ^{35}S-dATP (1000Ci/mmol) and 3U sequenase version 2.0 T7 polymerase enzyme were added to the preannealed template, and incubated at room temperature for 2-8 minutes. The duration of the extension reaction determined the length of sequence from the primer to be read. Prior to termination, 2.5µl each ddATP, ddTTP, ddCTP, ddGTP were held at 42°C for 2 minutes. The termination reactions began with the addition of 3.5µl extension mix to each ddNTP, incubation at 42°C for 5 minutes each, and finally stopped with the addition of 4µl Stop solution (which also acts as formamide loading dye). Samples were denatured at 94°C for 3 minutes before loading 3.5µl onto a denaturing salt-gradient polyacrylamide gel.

2.4.18.2 Fluorescent-labelled (dye terminator) cycle sequencing

Either the Dye Terminator Cycle Sequencing Ready Reaction kit, or the more recent dRhodamine Terminator Ready Reaction kit were used for fluorescent-labelled sequencing of PCR products or cloned inserts. Both kits included the fluorescently-labelled dideoxynucleotides, dNTPs, and AmpliTaq DNA Polymerase, FS enzyme in a Ready Reaction mix, formulated for optimal sequencing results.

Sequencing reactions were set up in 10µl volumes in small eppendorf tubes. About 500ng of plasmid DNA, or 75-100ng PCR product, was mixed with 4µl Ready Reaction mix, and 3.2 ρmoles primer. MilliQ water was used to make up the volume. A drop of mineral oil was placed on top of each sample. Cycle-sequencing reactions were carried out on a Techne Thermal Cycler *Gene-E* for 25 to 30 cycles as follows: 96°C for 30 seconds denaturation, 50-55°C annealing for 15 seconds, followed by 60°C extension for 4 minutes. Heating between temperatures was carried out at rapid thermal ramp rates of 25. The samples were held at 4°C prior to purification.

The sequencing samples were removed from under the oil and added to 1µl 3M sodium acetate (pH4.6) in a fresh eppendorf. 25µl 95% ethanol was added and mixed, and the samples held on ice for 10 minutes to allow precipitation. The samples were spun at 13000rpm for 15 minutes to collect the sequencing reaction, the ethanol removed. The reactions were washed with 450µl 70% ethanol (diluted with MilliQ water) and spun for 5 minutes at 13000rpm. The ethanol was removed and the samples air-dried. The sequencing reactions were resuspended in 3-4µl of dextran loading dye. 2µl sample was loaded on the polyacrylamide sequencing gel.

2.4.19 Polyacrylamide gel electrophoresis (PAGE)

2.4.19.1 Salt-gradient PAGE of ^{35}S-dATP labeled DNA sequences

The gel plates used to form the sequencing gel were 33.5 x 42 cm (bottom plate) and 33.5 x 39.5 cm top plate (Biorad). The plates were cleaned with distilled water and detergent, followed by ethanol, prior to use. The inside face of the top plate was then liberally covered with Sigmacote (Sigma), to prevent the gel from sticking to that plate. The plates were assembled with 0.4 mm spacers between them and tape or bulldog clips around the edges to hold them together. 60ml of denaturing gel mix consisting of 6% 19:1 acrylamide:bisacrylamide acugel, 7M urea was prepared and 35µl TEMED and 300µl 10% ammonium persulphate (APS) were added to form the gel complex. The gel was poured using a 50ml syringe, and the straight edge of a sharks tooth comb placed at the top of the gel to form a base for well formation. The gel was allowed to set for 45 minutes, and the comb was inverted to form wells into which the samples were loaded. The gel was placed vertically in the electrophoresis tank (Gibco BRL). 0.5X TBE buffer was placed in the top buffer chamber, and 0.5X TBE, 1M NaCl was placed in the bottom chamber. The gel was prewarmed by running at 1500V for 20 minutes before loading the samples. The wells were washed with buffer to remove the urea, ensuring even entry of the samples into the gel. Gels were electrophoresed at 60W with the voltage limited to 1500V for up to 3 hours, until the bromophenol blue dye reached the bottom of the gel. During electrophoresis the salt travels up through the gel creating a salt gradient, which condenses the DNA bands, allowing more sequence to be obtained from a single gel run.

The plates were separated, with the gel adhering to the bottom plate. A piece of 3MM Whatman paper was laid over the gel, and the gel lifted from the glass onto the paper, and covered with clingfilm. The gel was dried on a gel drier set at 80°C attached to to a vacuum pump, for 1 hour. The cling film was removed and the dried gel autoradiographed.

2.4.19.2 PAGE of fluorescent-labeled sequencing products

Fluorescently-labeled DNA sequences were electrophoresed in 5% denaturing polyacrylamide gels prepared with ultrapure reagents to preclude any fluoreceins from the gel that could interfere with reading the sequence.

The glass plates (Genex) were washed with non-fluorescent alconox detergent, rinsed in water, then rinsed in MilliQ water and allowed to dry. The plates were assembled in the cassette with 0.2mm spacers separating them, and clamped in place. The

gel was made with 5.2ml 40% 19:1 ultrapure acrylamide:bisacrylamide for short-range sequencing (up to 400bp), or 5.8ml 40% 29:1 ultrapure polyacrylamide for long-range sequencing (400-600 bp), 18g Ultrapure urea, 25ml MilliQ water and 0.5g Amberlite resin. The gel mix was stirred until the urea had dissolved, then filtered through a 0.2μM cellulose nitrate filter. 5ml 10X TBE (Flowgen) was put in the receptacle at the base of the filter to de-gas it. The gel mix was made to a volume of 50ml with MilliQ. 250μl 10% APS and 35μl TEMED were mixed in, and the gel poured with a 50ml syringe. A trough was formed at the top of the gel by inserting the straight edge of a sharks-tooth comb into the gel. The gel was allowed to set for 1 hour. The comb was removed, the trough rinsed with 0.5X TBE, and re-inserted with teeth jutting into the gel to form the wells. 24 or 36-tooth combs were used accordingly.

The gel was assembled in the ABI 377 PRISM automated DNA sequencer (Perkin Elmer), and 0.5X TBE placed in the top and bottom buffer chambers. The cooling plate was clamped over the front of the gel. The gel was pre-run on the Prerun *2XA* program until the gel reached 51°C with the voltage at 0.97V. The wells were washed out with buffer and 2μl sample denatured in dextran-formamide dye loaded. The samples were electrophoresed for 7 hours on 36 well-to-read module, and the reading distance stipulated at either 24 or 36-sharks-tooth comb. For samples sequenced with the Dye Terminator Cycle Sequencing Ready Reaction kit, the Run program *2XA* was used with the *377 matrix* and primer parameter set at *Any set primer A*. For samples sequenced with the dRhodamine Terminator Ready Reaction kit the Run program *36E-1200*, with the *dRhod matrix* and *dR set any primer* parameters used.

After electrophoresis the tracking of the lanes was checked and rectified if necessary, then the sequences read automatically using the Sample Sheet settings. These were placed into electropherogram and text files for analysis.

2.4.19.3 Single-stranded conformation polymorphism (SSCP) gel electrophoresis

SSCP analysis was performed using mutation detection enhancement (MDE) gel. 40ml 2X MDE gel was mixed with 36ml water and 4ml 10X TBE buffer. Where 15% glycerol was included, 12ml glycerol was added in place of water. 40μl TEMED and 400μl 10% APS was added prior to pouring, to set the gel. The gel apparatus was assembled in the same way as for radioactive sequencing, but the plates were separated by 1mm spacers. A 5μl aliquot of PCR product was mixed with 3μl formamide dye and denatured for 3 minutes at 95°C, prior to loading 4μl. The gel was electrophoresed in

0.5X TBE under two condidtions: at 15W overnight in the cold room at 4°C, or at 60W at room temperature for approximately 3 hours.

2.4.20 Silver-Staining of SSCP gels

The electrophoresed MDE gels were 'silver-stained' to visualise the DNA. The gel was removed, placed in a tray, and soaked in 1L of the various silver-staining solutions. Firstly the DNA was fixed in 10% ethanol for 5 minutes, then oxidised in 1% nitric acid (15.4ml/L) for 3 minutes. The gel was rinsed in distilled water then 0.1% silver nitrate solution (stored at 4°C) was added and allowed to impregnate the gel for 20 minutes, followed by another rinse in water. Freshly prepared developing solution (0.28M Na_2CO_3, 0.19% formaldehyde) was added and the tray rocked until the DNA bands appeared. The reaction was stopped by adding 10% acetic acid for 2 minutes, and the gel rinsed in distilled water. The gel was then shrunk in 50% ethanol for 30 minutes, laid on 3MM Whatman paper and dried for 1 hour on a vacuum-pumped gel drier set at 80°C.

2.4.21 Lymphoblast cell culture

2.4.21.1 Defrosting cell lines from frozen

EBV-tranformed lymphoblastoid cells were removed from storage in liquid nitrogen for culture. The cells were defrosted quickly in a 37°C waterbath and spun for 5 minutes at 1000rpm in a centrifuge. The supernatant was removed and the cells resuspended in a few ml of RPMI medium containing 20% fetal calf serum (FCS), whereby they were transferred to a 25ml flask, and a further 5ml RPMI plus FCS added. The cells were incubated in a humidified 37°C incubator with 5% CO_2.

2.4.21.2 Cell culture and harvesting

At 2-day intervals about half of the yellowing medium was removed from the top of the culture into Chloros disinfectant using a suction pump, and replaced with fresh medium. As the cells approached confluency (approximately 1 x 10^6 cells per ml), the culture was separated into three 50ml flasks and cultured in 20ml volumes of RPMI medium with 20% FCS. For large quantities of cells, these were then transferred to 75ml flasks and cultured in 30ml volumes, with feeding on alternate days. For RNA extraction, as much of the media as possible was removed, and the remainder containing the cells

was transferred to falcon tubes. The cells were spun at 3000rpm for 10 minutes and the supernatant removed prior to the addition of the Trizol reagent.

2.4.21.3 Freezing cell lines for storage

To maximise viability, over 2 x 10^6 cells /ml were frozen. The final 10ml of a 20ml culture was transferred to a falcon tube and spun at 1000rpm for 5 minutes. All but 1.5ml supernatant was removed and the tube shaken gently to resuspend the cells. The cells were cooled on ice and 1.5ml 20% dimethyl sulphoxide in RPMI medium with 20% FCS was added slowly, drop by drop, to the cells. 1ml aliquots were placed in cryotubes pre-cooled on ice. To effect a gradual freezing of the cells, the cryotubes were placed in insulated boxes at –20°C for 2 hours, then at -70°C overnight and finally into liquid nitrogen.

2.4.22 Direct cDNA Selection

2.4.22.1 Preparation of the genomic probes

Genomic probes for direct cDNA selection consisted of YAC DNA, which had been purified by PFGE and restriction digested, the purified inter-*Alu* PCR products of YACs, and PAC and BAC DNA, purified by miniprep or maxiprep.

The genomic probes were immobilised on nylon membrane. Squares of N^+ membrane were cut out to roughly 2mm x 2mm and laid on 3MM Whatman paper. Depending on the size of the genomic probe, 15-50ng (1ng per 10-20 kb) DNA was spotted onto each square membrane. Sufficient DNA to make 5-7 membranes (75-200ng) was mixed with 10X the amount of carrier DNA, *Hae*III digested φX 174 (0.75-2μg), and made to a volume of 10μl with water. The DNA mix was denatured by boiling for 5 minutes and placed on ice for 2 minutes. 5μl 20X SSC was added to the mix, on ice, bringing the volume to 15μl. 1μl DNA mix was spotted onto each membrane square 2-3 times, and air-dried between each application. A dot of ink dot was placed in the corner of the square to denote 'DNA-side-up'.

The DNA was fixed to the membranes in denaturation solution for 5 minutes, then neutralised in neutralisation solution twice, for 5 minutes. This was carried out in separate petri-dishes containing 3MM Whatman paper soaked in the respective solution. The membranes were placed 'DNA-side-up' onto the soaked 3MM. The membranes were rinsed in 2X SSC, air-dried, and stored in labeled eppendorfs at 4°C.

2.4.22.2 Preparation of the 'starting cDNAs'

The cDNA inserts from the 9 week fetal brain cDNA libraries, Br32 and Br33, were amplified to produce the source of cDNAs for selection. About 1 $x10^6$ p.f.u. (1μl) of each of the cDNA library glycerol stocks was PCR-amplified using M13 vector primers, M13U and M13R. These primer binding sites are external to the T3 and T7 primers binding sites, either side of the MCS. Several 50μl volume reactions were set up for each library. Amplification after a hot-start was carried out for 30 cycles, with annealing at 60°C, and extension for 2 minutes to select for longer-length products. An 8μl aliquot of PCR product was electrophoresed on an agarose gel, alongside a λ DNA marker, to size the amplified inserts and ensure that no single length products were predominant. The amplified inserts were purified using the Geneclean kit, which also removes excess primers from the product. The cDNAs were quantified with reference to DNA standards spotted onto agarose containing 0.5 μg/ml ethidium bromide. Equal amounts of the purified cDNAs from the two libraries were combined and diluted to 500ng/μl to form the 'starting cDNA' pool. 500ng starting cDNAs were used in the first round of selection. This was increased to 1μg in repeat-reactions, where yields for certain probes had been low.

2.4.22.3 Pre-blocking of the genomic targets and 'starting cDNAs'

Both the genomic probes and the cDNAs were treated with the quench mix prior to the selection reaction to prevent cDNAs from binding to repeat elements, mitochondrial sequences and human and yeast ribosomal sequences. If the probe was derived from yeast then *S. cerevisiae* AB1380 DNA was included in the quench to block host vector sequences. If the probe was derived from bacteria then total *E. coli* DNA was used. Pre-blocking the genomic probe was especially important where YAC DNA had co-migrated with a native yeast chromosome during PFGE. The cDNAs were pre-blocked to exclude such sequences from the reaction (Lovett, 1994).

2.4.22.4 Prehybridisation of membranes

The fixed genomic probes were prehybridised in 50μl volumes of prehybridisation solution containing 1X quench, in 1.5ml eppendorfs. 5μl 10X quench mix and 27.5μl water was boiled for 5 minutes to denature the DNA, then placed on ice for 2 minutes. The eppendorf was then hand-warmed to prevent precipitation whilst 2μl 10% SDS, 5μl 50X Denhardt's solution and 12.5μl 20X SSPE (to a final concentration of 0.1% SDS, 5X

Denhardt's and 5X SSPE) was added in and mixed to form the prehybridisation solution. The membrane was submerged in this and 50µl mineral oil added to the surface to prevent evaporation. Prehybridisation was carried out in a waterbath at 65°C for 24 hours.

After prehybridisation the layer of mineral oil was carefully pipetted off the surface of the prehybridisation solution. The membrane was rinsed at least 3 times in 600µl 5X SSPE, 0.1% SDS at 65°C. At no time was the membrane allowed to come into contact with the oil.

2.4.22.5 Pre-blocking of cDNAs

500ng of the starting cDNAs for the first round of selection, or 50-100ng of the amplified primary-selected cDNAs for the second round of selection, were preblocked with 1X quench. This was performed in 50µl in 1.5ml eppendorfs, after the membrane had been prehybridising for 22 hours. The cDNAs and 5µl 10X quench solution were made up to a volume of 30.5µl with water. The mix was denatured by boiling for 5 minutes and placed on ice for 2 minutes. SDS, Denhardt's and SSPE were added to the mix as in the previous section. The cDNAs were preblocked at 65°C for 2 hours with no mineral oil overlay.

2.4.22.6 Hybridisation of the cDNAs to the fixed genomic target

The preblocked cDNAs were removed from the waterbath and spun to collect. The rinsed prehybridised membrane was transferred directly into the tube containing the cDNAs. With the membrane submerged in the cDNA solution, 50µl mineral oil was lain over the surface. The hybridisation reaction was incubated at 65°C for 48 hours.

2.4.22.7 Membrane Washing

The mineral oil was pipetted off from the surface of the hybridisation mix and the membrane rinsed several times in 5X SSPE, 0.1% SDS preheated to 65°C. A series of washes of gradually increasing stringency were then performed in 700µl volumes to elute non-specifically bound cDNAs, whilst retaining the cognate cDNA hybrids. Two washes with 2X SSC, 0.1% SDS were performed at room temperature for 5 minutes each, followed by two further washes at 65°C for 20 minutes each. Single washes were then carried out with 1X SSC, 0.1% SDS followed by 0.5%X SSC, 0.1% SDS at 65°C for 20 minutes. A final two washes in 0.2 or 0.1X SSC, 0.1% SDS were performed at 65°C for 20 minutes. The membrane was rinsed in 0.2 or 0.1X SSC at room temperature to remove the SDS.

2.4.22.8 Elution of the cognate cDNAs

To elute the specifically bound cDNAs, the membrane was transferred directly to a small eppendorf containing 40µl water. The tube was heated to 96°C for 8 minutes, then placed on ice for 2 minutes. The selected cDNAs were stored at -20°C.

2.4.22.9 PCR 'rescue' of the selected cDNAs and recycling

The selected cDNAs from the first round of selection were PCR-amplified using the outer M13 vector primers. Half (20µl) of the eluted product was amplified in 50µl volumes for 35 cycles with elongation for 1minute 30 seconds. A positive control PCR of 100ng of the starting cDNA pool was included. A 10µl aliquot of the PCR product was kept aside for agarose gel electrophoresis (to visualise the PCR product), and Southern analysis with the positive and negative control cDNAs. The remainder of the PCR product was purified using the Geneclean kit. 1µl of the purified product was spotted against DNA standards on agarose containing ethidium bromide, to quantify the cDNAs. The purified product was adjusted to 50ng/µl for recycling in a second round of direct selection.

The second round of selection was performed using a fresh membrane, with the corresponding genomic probe. 50-100ng of the M13-amplified primary-selected cDNAs, a tenth the quantity of cDNAs used in the first round, was recycled in the second round. This was because the primary-selected cDNAs were now enriched for cDNAs mapping to the target genomic probes. The secondary-selected cDNAs were rescued by amplification with the nested T3 and T7 vector primers. Amplification and quantitation were performed in the same manner as for the primary-selected cDNAs. 50ng of the starting cDNAs were included as a control for the PCR reaction. 10µl of the amplified secondary-selected cDNAs were retained for electrophoresis and Southern analysis with the reporter cDNAs. The remainder of the PCR product was Genecleaned and resuspended in a minimal volume of 10µl water.

2.4.22.10 Positive and negative control assays

Each 10µl aliquot of the amplified primary and secondary selected cDNAs, that had been retained were electrophoresed on a 0.8% agarose gel alongside 10µl (5ηg) of the starting cDNAs and a size marker. The electrophoresed gels were Southern blotted and the filters hybridised with the appropriate *UBE3A* positive control probe. This consisted of exons U3-U4 for cDNAs selected by the BACs, and exons 4-8 (renumbered

10-14) for cDNAs selected by YACs 132D4 and B230E3, and PACs 50I2, 142A2 and 318N23. The filters were stripped and probed with β-actin as the negative control.

2.4.23 Production of arrayed cDNA sublibraries of selected cDNAs

2.4.23.1 Cloning of secondary selected cDNAs

About 50ng of the amplified secondary selected cDNA inserts were cloned using the 'LigATor' kit. Half (50μl) of the transformation mix (in SOC medium) was plated onto large LB agar plates containing Xgal, IPTG, kanamycin and tetracycline, for picking and arrayal of clones. The remaining transformation mix was inoculated into LB broth containing kanamycin and tetracycline and grown overnight. A 50% glycerol stock of each sublibrary was made by vortexing an equal amount of glycerol with culture. This was stored at -70°C.

2.4.23.2 Gridding of clones from selected cDNA sublibraries

Approximately one clone per kilobase of genomic probe was picked, not allowing for overlaps where contigs of PACs and BACs formed a probe. Clones were arrayed in 96-well microtitre plates with columns labelled 1-12 and rows labelled A-H. The plates were labelled according to the genomic clone used in their selection and with Greek letters denoting the plate number per sublibrary.

White colonies (transformants) were picked at random from the agar plates using autoclaved tooth-picks, and transferred to individual wells of sterile 96-well microtitre plates containing 150μl of LB broth with 50μg/ml kanamycin and 20μg/ml tetracycline. Assuming the cloning procedures were optimal, each large LB agar plate contained 200-400 white colonies. The sublibraries were thus arrayed in co-ordinates within numbered plates for future reference of individual clones. The clones were grown overnight at 37°C then stored shorterm at 4°C. After PCR amplification of the gridded cDNA inserts (section 2.4.24.1), 50μl 40% glycerol in LB was mixed into each well and the plates stored long-term at -70°C.

2.4.24 Screening of the arrayed selected cDNA sublibraries

2.4.24.1 PCR-amplification of the cDNA inserts

The cDNA inserts of the clones were amplified using the pTAG vector primers, pTAG-5' and pTAG-3'. PCRs were performed in bulk, in 96-well PCR plates on an Omn-E thermocycler (Hybaid), specially adapted for microtitre plates. PCRs were set up in the sterile hood. Amplification was performed in a 50µl volume per well. A 5.5ml PCR premix per plate was made up in a trough using the combined KCl and $MgCl_2$ buffer. *Taq* Polymerase was included in the premix. The cDNA template consisted of 5µl of the overnight culture. (A multichannel pipette was used for the dispensation of both premix and template). The cDNA inserts were thus amplified in the same 96-well format as the LB overnight culture plate, so that corresponding co-ordinate positions were retained. A multidispenser pipette was used to overlay the PCRs with 25µl mineral oil, and a lid was placed on the plate. A hot start of 5 minutes at 95°C was performed followed by 30 cycles of amplification with annealing at 55°C and an elongation time of 1 minute 15 seconds.

2.4.24.2 Electrophoresis of amplified inserts

The 96 individual PCR products per plate were electrophoresed on a single 108-well 0.8% agarose gel. Three 36-tooth well formers were positioned in the 26cm long by 20cm wide gel tray, dividing the gel into three sections. The PCR products were electrophoresed in groups corresponding to the columns of the microtitre plate, with columns 1-4 in the top third of the gel, 5-8 in the middle, and 9-12 in the bottom section. Each columnar group was loaded from A to H. 500ng 1kb ladder was included on each third of the gel to size the inserts. Also included on the gel were a lane each of total human and ribosomal as hybridisation positive controls. Electrophoresis was performed for 2-3 hours at 100V. The gels were photographed to aid in the identification of individual clone co-ordinates following screening by hybridisation. The gels were 'sandwich' Southern blotted to produce duplicate filters of the cDNA inserts for screening.

2.4.24.3 'Counter-screen' to identify unwanted cDNA inserts

Despite pre-blocking, a certain percentage of clones in the selected sublibraries will be unwanted contaminants. These can be identified for exclusion from analysis by hybridisation screening the arrayed sublibraries (Parimoo et al., 1995). The cDNA filters were individually hybridised with Cot1 DNA and ribosomal DNA to identify repetitive

and ribosomal cDNAs contaminating the sublibraries. cDNAs selected with YACs were additionally hybridised with labelled *S. cerevisiae* AB1380 DNA and vector pYAC4 DNA. Clones selected with PACs or BACs were additionally hybridised with *E. coli* DNA. The blots were orientated with respect to the hybridisation signals generated for the Cot1 and ribosomal DNA hybridisation controls. Clone co-ordinates were identified for exclusion from further analyses by referring back to the photographs of the electrophoresed cDNA inserts.

2.4.24.4 'Positive' hybridisation screen to identify clones potentially mapping to the target genomic region

The Southern blots were subsequently hybridised with the original genomic probes employed for their selection. For example, cDNAs selected by the PAC-1 contig were hybridised with pooled DNA from PACs 5012, 318N23, and 14L12. cDNAs selected by the YACs or *inter*-Alu PCR products of YACs were hybridised with PFGE-purified YAC DNA or crude preparations of YAC DNA. The hybridisation probes were preannealed with Cot1 DNA to compete repeat sequences. Clones identified by a positive hybridisation signal were firstly checked against the panel of co-ordinates identified as contaminants in the 'counter-screen'. For instance, crude YAC probes hybridised to yeast cDNA inserts, as well as those potentially mapping back to 15q12. Clones giving a positive hybridisation signal with the target genomic probe, and not in the 'counter-screen', were picked from the 96-well LB plates under storage. These were grown up in 5ml LB broth containing kanamycin and tetracycline, and miniprepped for sequencing. 500µl glycerol stocks were made of each individual clone for long-term storage at -70°C.

2.4.25 Sequence Analysis

Sequence analysis was carried out using the Wisconsin Sequence Analysis package (GCG) and Sequence analysis systems, available through the HGMP at the world-wide web site **www.hgmp.mrc.ac.uk**. Several programs in the Wisconsin package enabled sequences to be compared, edited, aligned, assembled into contigs, and consensus sequences to be derived. Each program functioned on the basis of an algorithm for which the parameters could be modified (or the default specifications used). This set a precedent for the speed and stringency of the computations.

2.4.25.1 Sequence comparisons

Sequences obtained by manual radiolabelling were entered into GCG by typing the nucleotide sequence in via 'GelEnter'. Sequences obtained through automated sequencing were entered into GCG files through a file transfer protocol, and reformatted. The 'EXCise' tool allowed any vector sequences specified in 'GelStart' to be ignored in all subsequent comparisons.

The 'GAP' program directly compared pairs of sequences in forward or reverse orientation, as specified. The sequences were aligned at each nucleotide position until a match or overlap of was found which met the specifications set in the parameters. The program displayed the size and positions of overlaps with respect to each sequence, aswell as the degree of similarity at the overlapping positions. Conversely, this enabled the identification of discrepancies between single nucleotides within overlapping sequences, allowing sequencing errors to be rectified through 'Seqed'. 'Pileup' automatically compared each sequence with every other sequence entered, and aligned similar sequences beneath one-another, for visualisation.

'GelMerge' was used to create an assembly of contiguous, overlapping sequences, between a set of individual sequences entered into the directory. 'GelView' assembled this contig for visualisation on a graph. Small overlaps, or those internal to larger sequences, were not displayed on the contig map unless the 'GelMerge' algorithm parameters were reduced. The final consensus sequence was derived by 'GelMerge' and displayed in an output directory, 'Consensus'.

The 'Compare' program gave the degree of similarity between two query sequences. Alternatively where a sequence was compared with itself, the program detected the presence of repetitive elements or motifs that occurred more than once within the sequence. The query sequence was divided into 'words' of a specifed size. Each 'word' was compared with every other 'word' within the sequence in both orientations. 'Dotplot' plotted the sequence against itself. A dot or line was drawn against those 'words' that were also present elsewhere in the sequence. Thus the presence of repeats and their positions within the sequence could be detected.

'MAP' displayed both strands of a DNA sequence and the possible protein translations in three or six frames. Potential open reading frames (ORFs) were noted as long stretches of coding sequences which had not been interrupted by stop codons. This program also indicated the presence and positions of restriction sites within the sequence. 'MAPSORT' additionally gave the sizes of fragments that would be generated on cleavage with a given enzymes.

2.4.25.2 Sequence database searches

Sequence databases were searched to identify sequence similarity between input, or query sequences, and sequences entered into the various databases. These searches ascertained whether a query sequence was known, if it belonged to a gene family, was taxonomically conserved, contained repeats, or identified anonymous ESTs generated through the HGMP.

'BLAST' was used to search various sequence databases for sequence homologies both at the nucleotide sequence and peptide sequence levels. Databases of nucleotide sequences that were searched included GenBank, Embl, New Embl sequences, Human STS and Human EST, Rodent, Mammalian, Fugu, and Alu-like repeats. 'SwissProt' was searched for conserved peptide sequences or motifs.

2.4.25.3 NIX (nucleotide identify X) analysis

NIX is a multifactorial www system which combines several sequence analysis programs including 'BLAST', 'Genemark', 'Repeatmasker', and 'GRAIL' for a rapid, but overall analysis of nucleotide sequences derived from genomic or cDNA. The default parameters are applied for each individual program. Predicted intron-exon boundaries and open reading frames, repetitive elements, polyadenylation signals, regions of similarity to sequences within the databases, and other features, are displayed graphically. The query sequence is entered into the NIX menu from the GCG account.

2.4.25.4 Identification of mutations and polymorphisms within UBE3A exonic sequences

Mutation analysis was performed using the Sequence Navigator software (Applied Biosystems). The patient sequences, sequences from normal individuals and the appropriate published *UBE3A* sequences were entered into files for direct comparison using the 'Factura' program. The sequences were aligned in 'Clustal' and any sequence changes between the wild-type and mutant sequences were highlighted automatically in the sequence text. The electropherograms were then aligned synchronously and compared visually.

Chapter 3: Results

Positional cloning in the Angelman syndrome candidate region

3.1 Physical mapping in the AS candidate region

Positional cloning efforts to identify the gene responsible for AS were concentrated in the region considered most likely to contain the AS gene, which was flanked by markers D15S174 and D15S113 (Fig 1.9). The AS candidate region (ASCR) was encompassed by YACs 132D4, 230H12 and 142A2, although YACs B230E3 (which is internal to 132D4) and 230H12 filled most of this region (Fig 1.9). As a first step towards characterisation of the ASCR, an overlapping series of contiguous PAC and BAC clones was constructed within the region. STS markers, ESTs and the *UBE3A* gene were subsequently mapped to this contig. The work described in this section was performed as a group, with equal contributions from Dr. Jess Buxton and Sarah Rickard.

3.1.1 Construction of an integrated YAC, PAC, BAC and STS contig map of the ASCR

3.1.1.1 Identification of PAC clones from the ASCR by screening of a PAC library

A gridded human PAC library (section 2.1.4) was screened with YACs B230E3 and 230H12 to identify PACs from the ASCR. Fourteen PAC clones gave duplicate positive hybridisation signals, including 9F17, 50I2, 14I12, 26N4, 55O1, 113G10, 156B12, 163O7, 209N9, 216L13, 265H21, 314G1, 318N23 and 117N20. Identification of PACs 9F17 and 14I12 from the PAC library is shown in Fig 3.1.1. The 14 individual PAC clones were provided by the HGMP distribution center, Cambridge, UK.

3.1.1.2. PAC 'fingerprinting' to identify redundancy and overlaps between the clones

The 14 PAC clones were 'fingerprinted' by Southern analysis of *EcoR*I digested PAC DNA with total human DNA, to identify redundant clones and those that were overlapping. Where PACs overlapped significantly, or were internal to others, similar *EcoR*I banding patterns were observed, with several fragments in common. Eight non-redundant PACs were identified (which were not completely internal to others), including 50I2, 113G13, 26N4, 14L12, 9F17, 55O1, 318N23, and 209N9 (Fig 3.1.2). The overlaps between the individual PACs are indicated in Table 3.1.1.

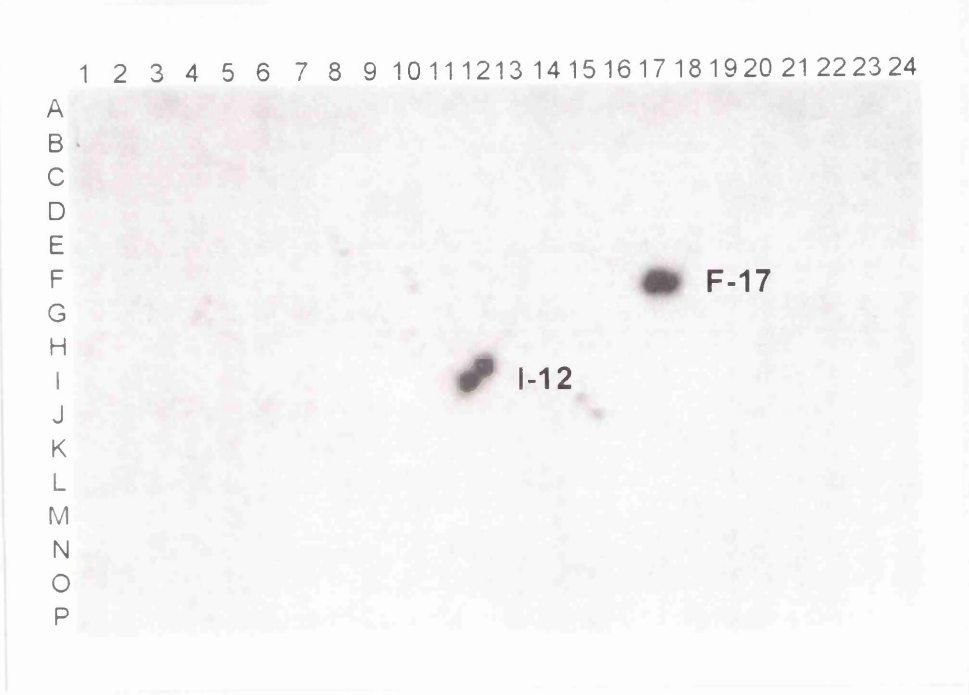


Fig 3.1.1 Screening of a PAC library with YACs B230E3 and 230H12

Identification of PACs 9F17 and 14I12 from the HGMP gridded PAC library, using YACs B230E3 and 230H12 as probes. The YACs were purified by pulsed-field gel electrophoresis, digested with *EcoR*I, and pooled in equal proportions. The radiolabelled probe was preannealed with total human DNA to prevent hybridisation due to the presence of repetitive sequences. The filters were washed to 0.1X SSC / 1% SDS and exposed to X-ray film for 2 days. True positives were identified by duplicate signals in a given pattern. Individual PAC clones were identified by their plate-row-column co-ordinates respectively, for instance 9-F-17.

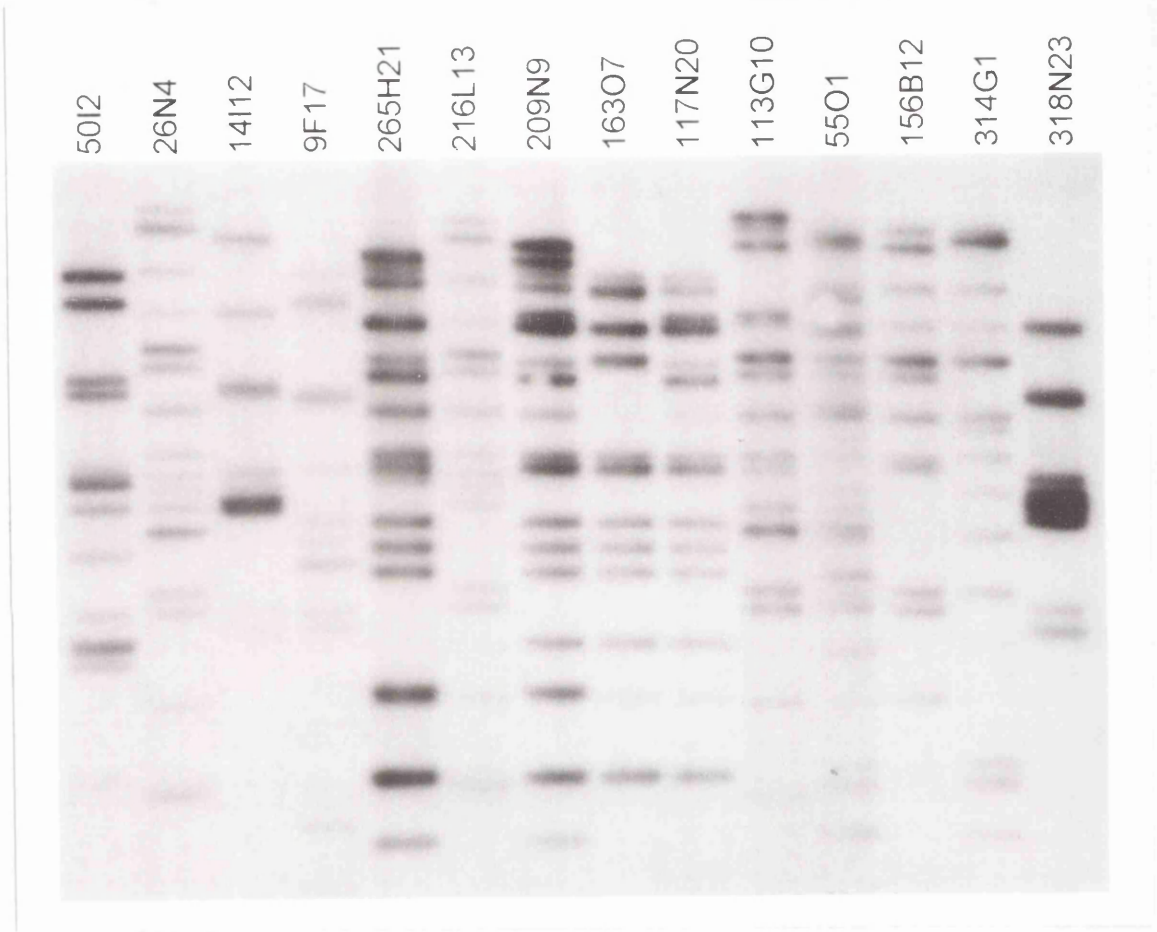


Fig 3.1.2. *EcoR*I Fingerprinting of PACs identified by YACs B230E3 and 230H12

*EcoR*I digested PACs hybridised with total human DNA, and washed to 0.1X SSC/1% SDS, and exposed to X-ray film for 4 hours. PAC 209N9 had a very similar banding pattern to PACs 163O7, 117N20 and 265H21, but had additional fragments. Thus 209N9 encompassed the other three clones. PAC 216L13 appeared internal to 26N4. PACs 314G1 and 156B12 were contained within 55O1. 113G10 had several bands in common with 55O1, indicating a substantial overlap.

3.1.1.3. Mapping of the non-redundant PAC clones to the YAC contig, and identification of contiguous PAC clones

The 8 non-redundant PACs listed in 3.1.1.2 were anchored to the YAC contig from the AS / PWS region, and mapped to one-another to assess the overlaps between clones, by Southern analysis. Miniprepped DNA from each PAC was radiolabelled, preannealed with total human DNA to suppress repeats, and hybridised to Southern blots containing *EcoR*I-digested YAC and PAC DNA. This positioned each PAC on the YAC contig, and confirmed the overlaps between individual PAC clones that had been identified by 'fingerprinting'. The mapping of PACs to the YAC contig by Southern analysis is illustrated by PAC 26N4 in Fig 3.1.3. The mapping data is summarised in Table 3.1.1, and represented diagrammatically in Fig 3.1.7.

PAC 9F17 hybridised to all of the YACs from the ASCR (Table 3.1.1), and most of the PAC clones (data not shown), but did not contain any of the STS markers (section 3.1.1.4; Table 3.1.2) or ESTs from the ASCR (section 3.1.2.2; Table 3.1.4). PAC 9F17 most likely contained a low copy-number repeat element, which cross-hybridised to genomic clones from 15q11-13, and may not have mapped to this region. This PAC was therefore not placed on the 15q12 contig, nor used for any positional cloning purposes.

PAC clone	YAC clones identified by PACs	Overlapping PAC clones
50I2	132D4, B230E3	318N23
9F17	132D4, B230E3, 230H12, 142A2, 11HE12, 268A12	Repetitive
14I12	132D4, B230E3, 230H12	318N23
318N23	132D4, B230E3, 230H12	14I12
209N9	230H12	26N4, 113G10
26N4	230H12, 142A2	209N9, 113G10, 55O1
113G10	230H12, 142A2	209N9, 26N4, 55O1
55O1	230H12, 142A2, 11HE12	26N4, 113G10

Table 3.1.1 Mapping of PACs to YACs spanning the ASCR
Summary of the overlaps between the 8 non-redundant PAC clones and their mapping positions with respect to YAC clones spanning the ASCR.

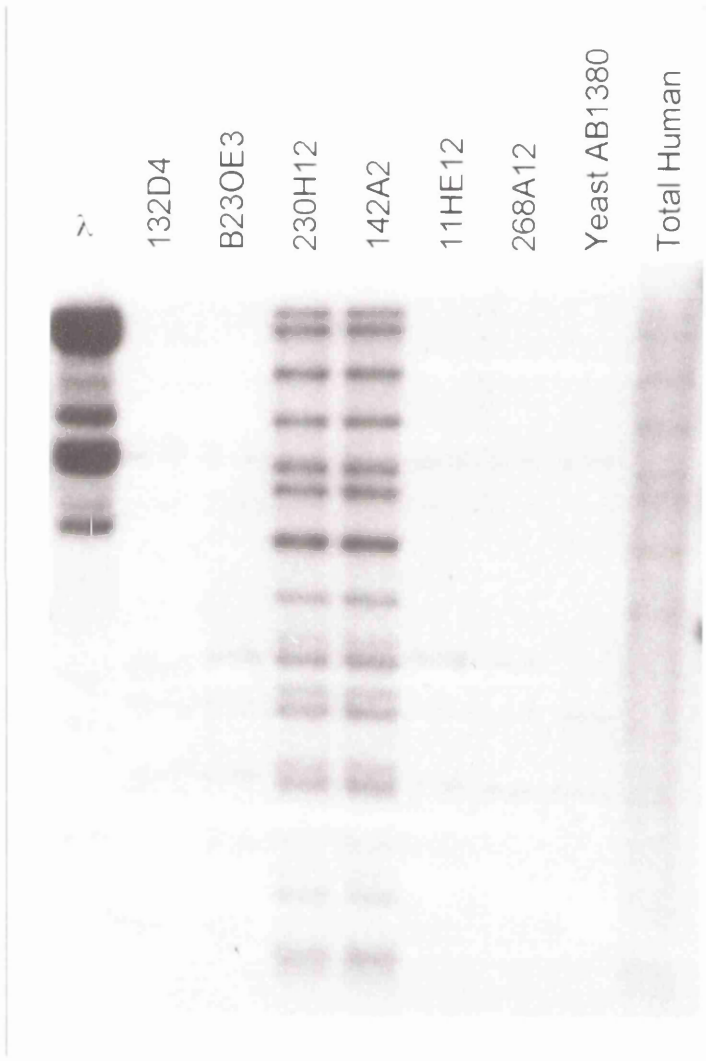


Fig 3.1.3 Mapping of PAC 26N4 to YACs 230H12 and 142A2 from the ASCR

Autradiograph of *EcoR*I digested YAC DNA, hybridised with PAC 26N4, washed to 0.1X SSC/1% SDS and exposed overnight. The radiolabelled probe was preannealed with total human DNA to suppress hybridisation between repetitive sequences.

3.1.1.4 Integration of YACs, PACs and BACs from the ASCR into an STS-based contig map

STS mapping was performed by PCR for markers D15S174, D15S122, D15S10, D15S210, D15S113, and the left and right hand ends of YAC B230E3, to integrate the PACs and YACs into an STS-based contig map. The STS mapping data is summarised in Table 3.1.2. and represented diagrammatically in Fig 3.1.7. None of the initial 14 positive clones identified from the PAC library contained STS D15S210. On the basis of the hybridisation and STS mapping data, PACs 50I2, 318N23 and 14I12 were contiguous and mapped within the region spanned by YAC B230E3 (PAC-1, Fig 3.1.7), and PACs 209N9, 26N4, 113G10 and 55O1 formed a second separate contig, encompassed by YACs 230H12 and 142A2 (PAC-2, Fig 3.1.7). A clone containing marker D15S210 was needed in order to bridge the gap between the two groups of overlapping PACs. PCR-pools of the entire PAC library were screened for D15S210 by PCR, in an attempt to identify a clone positive for this marker. However, no PACs were identified with this marker in this library (data not shown). Four BAC clones 139E9, 268A2, 119H2 and 119G2, which contained D15S210, were kindly provided by Dr. Ung-Jin Kim, Caltech, USA to close the gap. The BAC clones were STS-mapped by PCR (Table 3.1.2) and hybridised to *EcoR*I blots of the YACs and PACs from the ASCR. BAC 139E9 hybridised to PACs 318N23 and 14I12 from the PAC-1 group, and BACs 268A2, 119H2 and 119G2 overlapped the PACs from the PAC-2 group (data not shown). Thus the 4 contiguous BACs joined the two separate groups of PACs, across marker D15S210, to form a single PAC and BAC contig across the ASCR. These data are represented in the form of a map in Fig 3.1.7. Mapping of ESTs to the YAC, PAC and BAC contig also aided in the precise location of the PACs and BACs with respect to one another (section 3.1.2).

Genomic clone	D15S174	D15S122	D15S10	D15S210	B230E3-L	B230E3-R	D15S113
YAC 132D4	+	+	+	+	+	+	-
YAC B230E3	-	+	+	+	+	+	-
YAC 230H12	-	+	+	+	-	+	-
YAC 142A2	-	-	-	-	-	-	+
PAC 50I2	-	-	-	-	+	-	-
PAC 9F17	-	-	-	-	-	-	-
PAC 14L12	-	+	+	-	-	-	-
PAC 318N23	-	+	+	-	-	-	-
PAC 209N9	-	-	-	-	-	-	-
PAC 26N4	-	-	-	-	-	-	-
PAC 113G10	-	-	-	-	-	-	-
PAC 55O1	-	-	-	-	-	-	-
BAC 139E9	-	+	+	+	-	+	-
BAC 119H2	-	-	-	+	-	-	-
BAC 268A2	-	-	+	+	-	+	-
BAC 119G2	-	-	-	+	-	+	-

Table 3.1.2 STS mapping of the YACs, PACs and BACs from the AS / PWS region.

Summary of the results of the PCR-mapping of STS markers from within the AS and PWS region to YAC, PAC and BAC clones. Positive PCR reactions, locating an STS to the genomic clone listed, are denoted by a plus (+) sign. Amplification reactions for which no PCR product was visualised, indicating the absence of that marker in the genomic clone, are shown by a minus (-) sign.

3.1.2 Transcript Mapping in the ASCR

As a first step towards identifying candidate genes for AS, a transcript map of previously identified genes and ESTs was produced. This was to determine which genes and ESTs were located within the ASCR, and to map these more precisely.

3.1.2.1 Mapping of previously known genes to the YAC, PAC and BAC contig

Previously known expressed sequences, *IPW, UBE3A* and PARs 1, 2 and 4 were mapped by PCR or hybridisation to the YAC and PAC/BAC contig. The mapping positions of these sequences within the contig are shown in Fig 3.1.7.

UBE3A had previously been located to YACs 132D4 and B230E3 by PCR amplification of a 377bp product within exon 3 (Nakao *et al.*, 1994). To map this gene more precisely, *UBE3A* exons were amplified and used as hybridisation probes to map the exons to YACs, PACs and BACs from the ASCR contig. *UBE3A* exons 1-3 (later numbered exons 7-9)(Fig 3.1.4), and exons 4-8 (renumbered 10-14) (Fig 3.2.6) were generated by RT-PCR amplification from lymphoblast and fibroblast cDNAs. The final exon 10 was amplified from genomic DNA using intronic primers E6ZA and E6ZB (Kishino *et al.*, 1997). The locations of the *UBE3A* exons are given in Table 3.1.3. and shown diagrammatically in Fig 3.1.7. The mapping of exons 4-8 to PACs 50I2, 14I12 and 318N23 is shown (Fig 3.1.5). Exon 10 was not single copy and so could not be placed on the contig. Mapping of the 5' coding exons to PACs 14I12 and 318N23, and the more 3' exons to PAC 50I2 showed the transcriptional orientation of *UBE3A* to be from telomere to centromere. Thus exons 1-8 (renumbered 7-14) were contained within a 200kb region of the ASCR, but mapped distal to the PWS candidate region (Fig 3.1.7). With the identification of four additional 5' untranslated exons of *UBE3A*, U1-U4 (Kishino *et al.*, 1997), a cDNA probe consisting of exons U3-U4 was similarly generated by RT-PCR of lymphoblasts, and mapped to the PACs and BACs (Table 3.1.3, Fig 3.1.7).

No positive hybridisation signal was obtained with any of the *UBE3A* probes for YAC 230H12. Exon 3 was previously shown to be absent in this YAC, and a deletion within 230H12 was suggested at the time (Nakao *et al.*, 1994). This is consistent with the

STS mapping data presented in Table 3.1.2, which shows that 230H12 did not contain markers D15S10, D15S210 (Table 3.1.2).

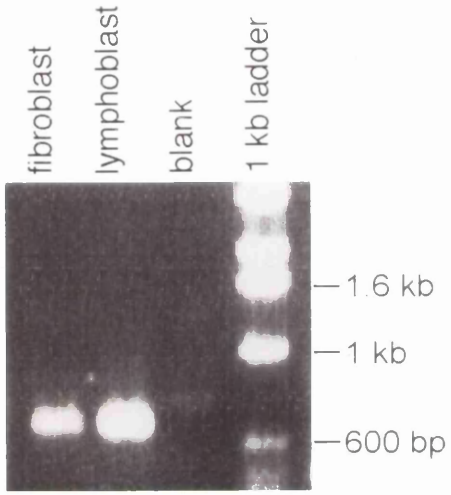


Fig 3.1.4 PCR-amplification of *UBE3A* exons 1-3 for use as a probe

A 600bp RT-PCR product across *UBE3A* exons 1 to 3 was generated with primers E6AP-1F and E6AP-3R from fibroblast and lymphoblast cDNAs. The products were purified for use as a hybridisation probe.

***UBE3A* probe (exons)**	**Mapping position: YAC, PAC, and BAC**	**Size of *EcoR*I fragments identified**
U3-U4	318N23, 14I12, 139E9, 268A2	~ 3.3kb and 4kb
1-3	132D4, B230E3, 318N23, 14I12	~ 4.3kb and 2.9kb
4-8	132D4, B230E3, 318N23, 14I12, 132D4, B230E3, 50I2	~ 6.3kb and 2.3kb ~ 4.5kb

Table 3.1.3 Location of *UBE3A* exons within the YAC, PAC and BAC contig spanning the ASCR.

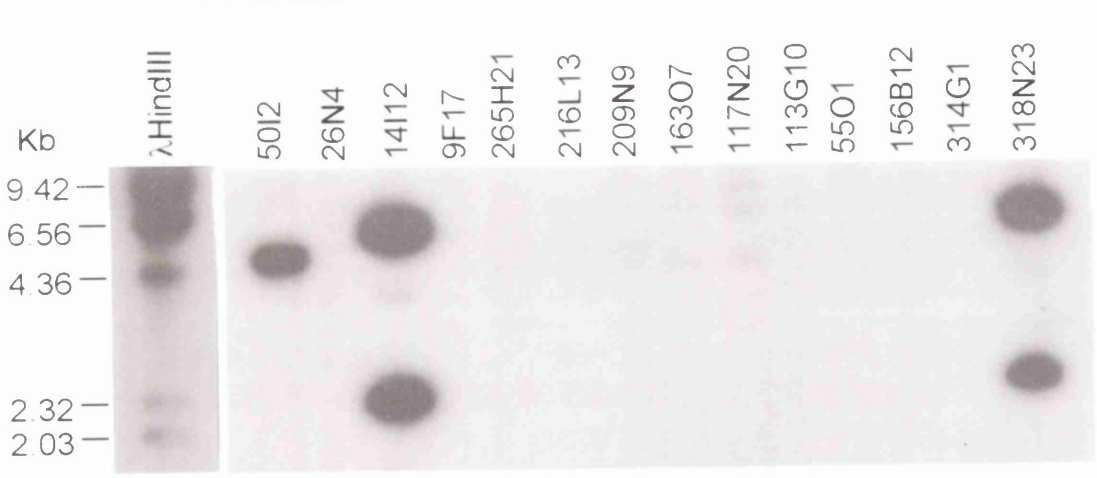


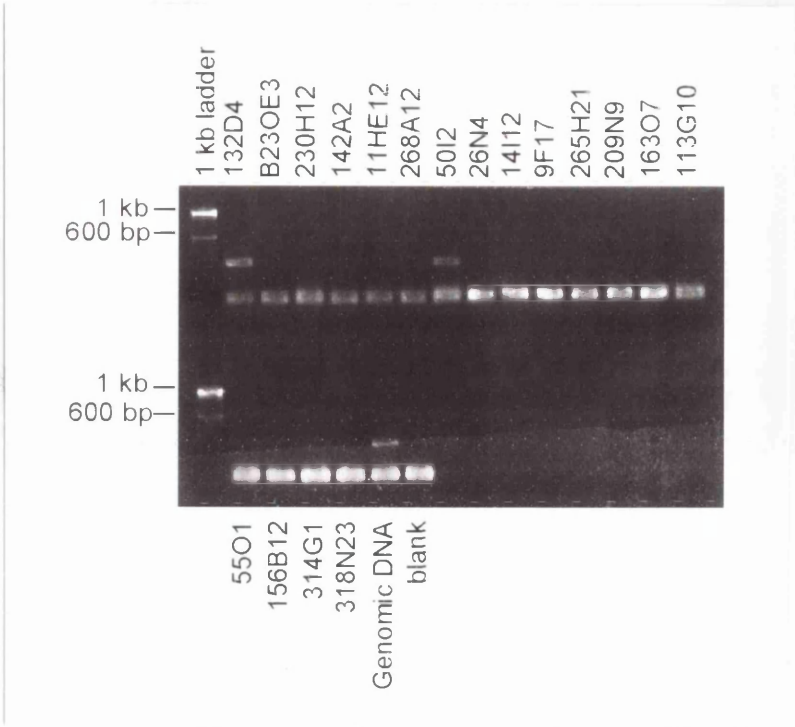
Fig 3.1.5 Hybridisation of *UBE3A* exons 4-8 to PACs 50I2, 14I12 and 318N23.

UBE3A exons 4-8 (renumbered 10-14) identified 3 separate *EcoR*I fragments in separate PACs. These data indicated that the 3' coding exons mapped to PAC 50I2, centromeric to the 5' coding exons 1-3.

3.1.2.2 Mapping of ESTs from the Whitehead Institute (WI) radiation hybrid map to the YAC, PAC and BAC contig in the ASCR

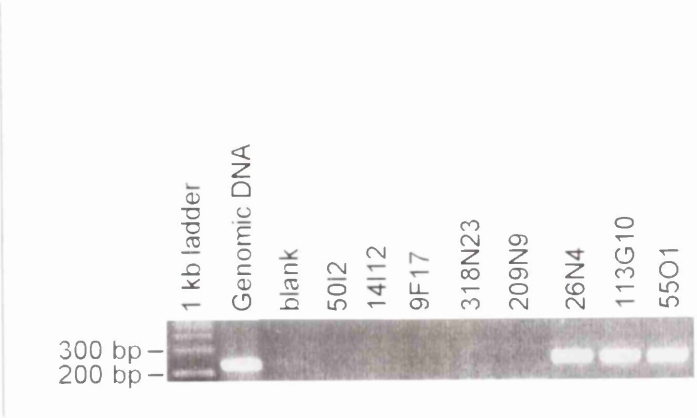
The National Center for Biotechnology Information (NCBI) Human Gene Map lists ESTs that have been located to particular chromosome regions by radiation hybrid mapping at the Whitehead Institute (Hudson *et al.*, 1995) (section 1.9.3). This map is publicly accessible through the World Wide Web at the following address: **http://www.ncbi.nlm.gov/cgi-bin/SCIENCE96/**

A map search was conducted in the 6-14cM interval between D15S128 (which maps proximal to *SNRPN*) and D15S156 (which lies distal to *GABRA5*), to identify ESTs which had been assigned to 15q11-13 by radiation hybrid mapping. 'Unidentified transcripts' WI-6780, WI-13791, WI-6654, WI-6519/WI-11918 and WI-16777 were located to this interval. The GenBank entry information for each sequence, which was linked directly to the Human Gene Map website, listed the primers which could be used to analyse each EST by PCR. These were acquired and used to map the ESTs listed above to the YACs, PACs and BACs from the ASCR, by PCR amplification of the EST sequence using the genomic clones as DNA template. Fig 3.1.6 shows the PCR mapping of ESTs WI-6654 and WI-16777 to YACs and PACs from 15q11-12. The mapping data for each EST is summarised in Table 3.1.4, and represented on the contig map in Fig 3.1.7. ESTs WI-6654, WI-6519, and WI-16777 map within the ASCR representing positional candidates for AS. EST WI-6780, which mapped to the PWS region, was shown by sequence comparison to represent *IPW*. WI-11918 was found to be internal to WI-6519.



A. WI-6654

Amplification of total YAC DNA and 'miniprepped' PAC DNA with WI-6654 primers L and R to give a 250bp product in YAC 132D4 and PAC 50I2.



B. WI-16777

Amplification of EST WI-16777 from 'miniprepped PAC DNA using WI-16777 primers L and R, to give a 125bp product in PACs 26N4, 113G10 and 55O1.

Fig 3.1.6 PCR-mapping of ESTs from the Whitehead Institute radiation hybrid map to YACs and PACs from the contig across the ASCR.

Genomic clone	EST WI-13791	EST WI-6654	EST WI-6519	EST WI-16777	EST WI-6780 (*IPW*)
YAC 132D4	+	+	+	-	+
YAC B230E3	-	-	+	-	-
YAC 230H12	-	-	-	+	-
YAC 142A2	-	-	-	+	-
PAC 50I2	-	+	+	-	-
PAC 9F17	-	-	-	-	-
PAC 14L12	-	-	-	-	-
PAC 318N23	-	-	-	-	-
PAC 209N9	-	-	-	-	-
PAC 26N4	-	-	-	+	-
PAC 113G10	-	-	-	+	-
PAC 55O1	-	-	-	+	-
BAC 139E9	-	-	-	-	N/T
BAC 119H2	-	-	-	-	N/T
BAC 268A2	-	-	-	-	N/T
BAC 119G2	-	-	-	-	N/T

Table 3.1.4 PCR-mapping of ESTs from the Whitehead Institute to the YAC, PAC and BAC contig across the ASCR.

A plus sign (+) denotes amplification of the EST sequence from the genomic clone. A minus (-) indicates that the EST is absent in the respective genomic clone. Mapping reactions not tested are indicated by N/T.

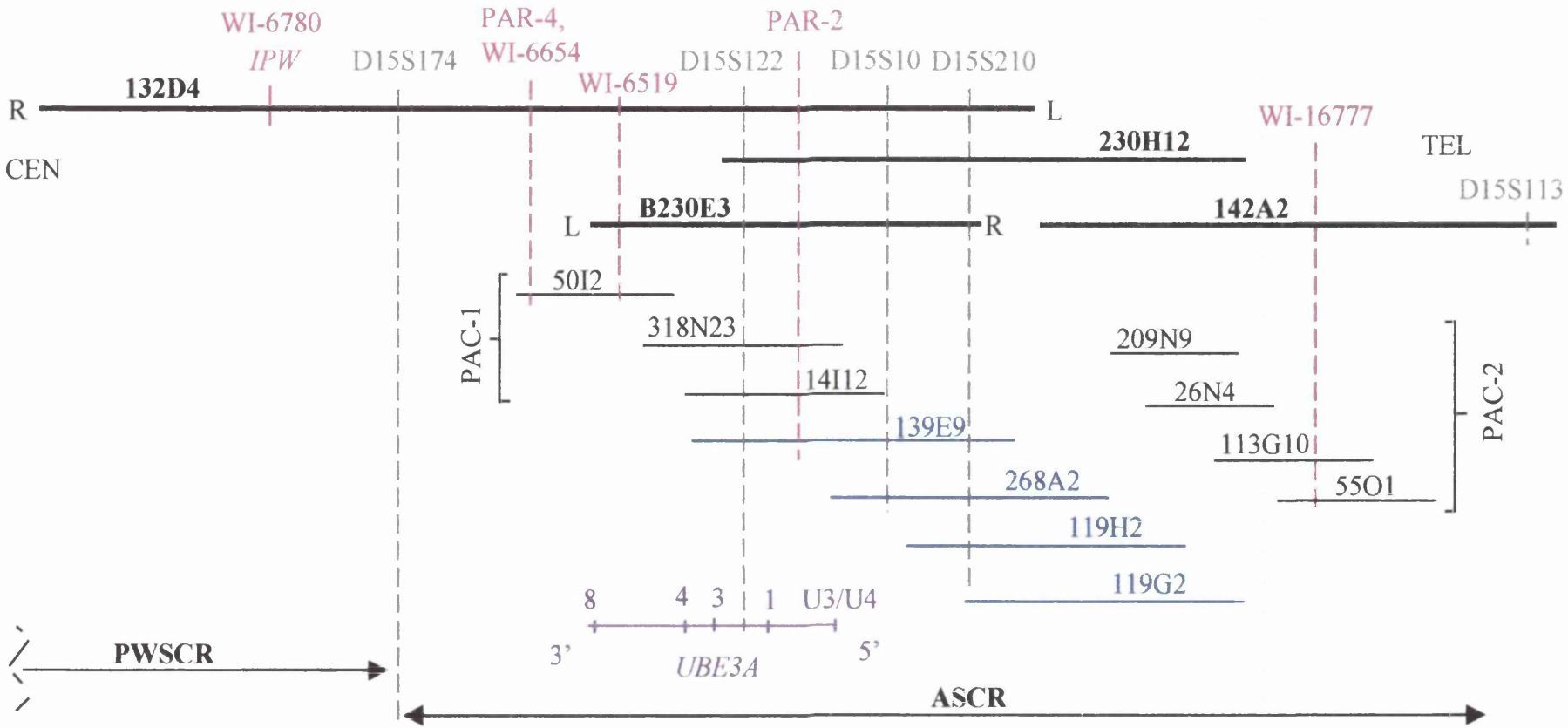


Fig 3.1.7 Integrated YAC, PAC and BAC contig and transcript map of the ASCR.
YACs are denoted by thick black lines, PACs by thin black lines, and BACs by blue lines. STS markers used for anchoring the PACs and BACs are labelled in grey. ESTs from the Whitehead Institute are in pink, and *UBE3A* is in purple. The PWSCR and ASCR are indicated by arrows. PACs were grouped into the PAC-1 and PAC-2 contigs as indicated, for use as genomic probes for the isolation of cDNAs.

3.2 Isolation of expressed sequences from the ASCR and distal portion of the PWSCR by direct cDNA selection.

The use of direct cDNA selection to isolate expressed sequences from the ASCR and distal portion of the PWSCR is described in this section. The principles involved in this technique are detailed in section 1.9.1.4. A revised membrane-based method was employed in this project (Parimoo *et al.*, 1995), with minor alterations for convenience, as outlined in Fig 3.2.1. The inherent advantage of direct cDNA selection is that transcribed sequences may be rapidly identified from large genomic regions using YACs as probes. At the outset of this project the ASCR was cloned in the YAC contig constructed by Mutirangura *et al.*, 1993b. Therefore, prior to the construction of the PAC and BAC contig across the ASCR (section 3.1), YACs were the sole resource of genomic clones available for positional cloning in this region. Direct cDNA selection was therefore considered the most suitable for positional cloning technique in this region at that time.

Direct cDNA selection began with the use of YACs 132D4, B230E3, 230H12 and 142A2 as genomic probes, spanning a region of 1Mb. These four YACs encompassed the ASCR, with D15S174 (the proximal marker for the ASCR) contained in YAC 132D4 and D15S113 (the distal extent of the ASCR) in YAC 142A2 (Fig 1.9). YAC 132D4 overlapped the distal section of the PWSCR, telomeric to *SNRPN*. Therefore any transcripts isolated from this YAC were also analysed with respect to any involvement in the PWS phenotype. The PAC and BAC clones from the ASCR contig (section 3.1) were subsequently used as probes for direct selection as well.

At the time *IPW*, *UBE3A* and the uncharacterised transcripts PARs 1, 2 and 4, were the only expressed sequences that had been mapped to this interval (Wevrick and Franke, 1994; Nakao *et al.*, 1994; Sutcliffe *et al.*, 1994). No AS candidate genes, expressed monoallelically from the maternal allele, had been identified. The aim was to acquire a more comprehensive transcription map of the region, with a view to identifying candidate genes that may contribute to the AS or PWS phenotypes.

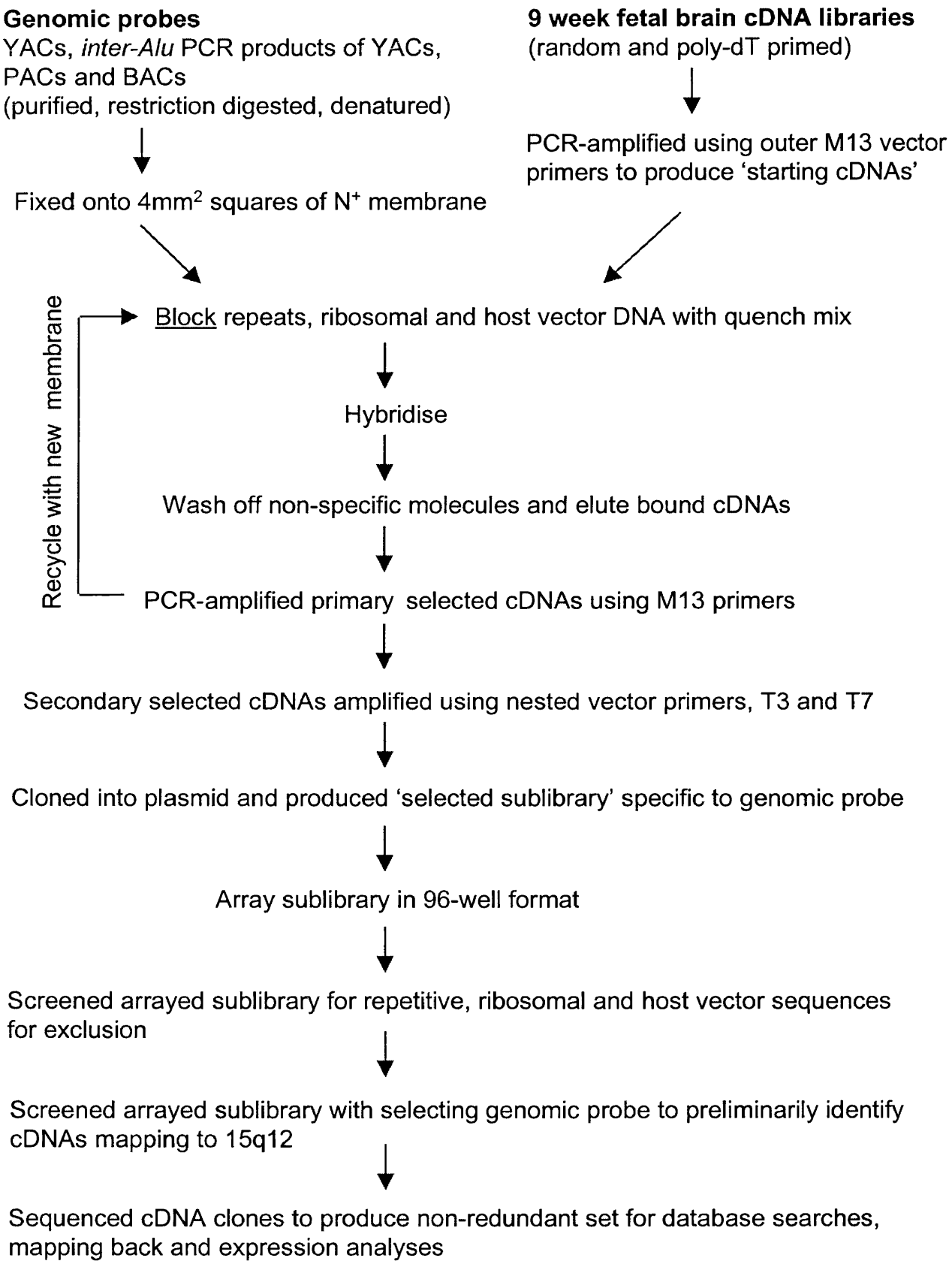


Fig 3.2.1 Flow diagram of the direct cDNA selection protocol applied to 15q12.

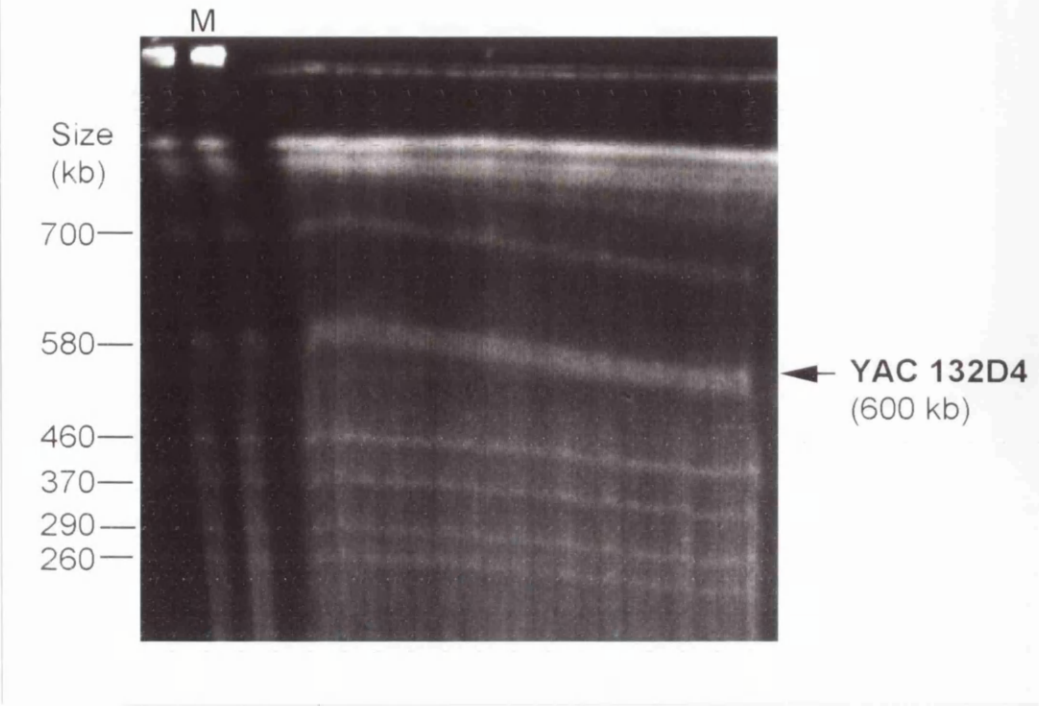
3.2.1 Genomic probes used as targets for direct selection

3.2.1.1 Purified YAC probes

YAC DNA for 132D4, B230E3, 230H12 and 142A2 was purified from the native yeast chromosomes using PFGE (Fig 3.2.2). The sizes of the YACs had previously been estimated at 600kb (132D4), 250kb (B230E3) and 520kb (230H12 and 142A2) (Mutirangura *et al.*, 1993). However, YAC B230E3 was visualised between the 260 and 290kb host yeast chromosomes, bringing the size estimate for this YAC closer to 275kb (Fig 3.2.2B). YACs 132D4 (Fig 3.2.2A), 230H12 and 142A2 (Fig 3.2.2C) co-migrated with the 580kb yeast chromosome so that both the artificial and yeast chromosomes were excised from the gel. This yeast DNA represented a contaminant of the target genomic DNA in the direct selection, requiring competition with total yeast and yeast ribosomal DNA in the blocking agent prior to selection of cDNAs. Restriction-digested YAC DNA was fixed onto pieces of membrane for use as genomic targets for cDNA selection. These probes are refered to simply by their YAC names.

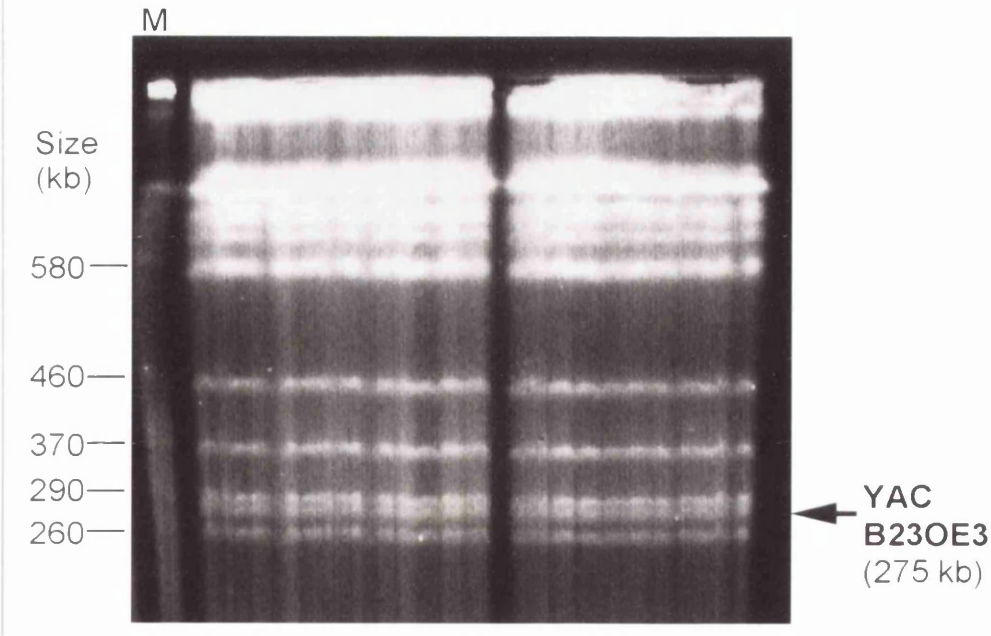
*3.2.1.2 Inter-*Alu *PCR products of YACs*

Inter-Alu PCR amplifies sections of genomic DNA that lie in between adjacent Alu elements. Since Alu repeats are specific to primates, human DNA can be amplified from complex sources such as human x rodent cell hybrids and YACs. *Inter*-Alu PCR may therefore be used to rapidly generate human-specific probes from YAC DNA as this precludes the need for purification from the yeast background. Primers CL1 and CL2 recognise consensus sequences at the 5' and 3' ends respectively of Alu repeats (Lengauer *et al.*, 1992). These were used to generate *inter*-Alu genomic probes for cDNA selection from crude DNA preparations YACs 132D4, B230E3, 230H12 and 142A2. As illustrated in Fig 3.2.3, incorporation of both primers in a PCR reaction amplifies the intervening genomic sequences, irrespective of the orientation of the Alu element. The amplification products retain high sequence complexity but are of reduced size. Amplification with a single primer included in the reaction, generates products of a larger size between Alu elements orientated in opposite directions (Fig 3.2.3), but may result in loss of genomic sequence content (reviewed by Kass and Batzer, 1995).



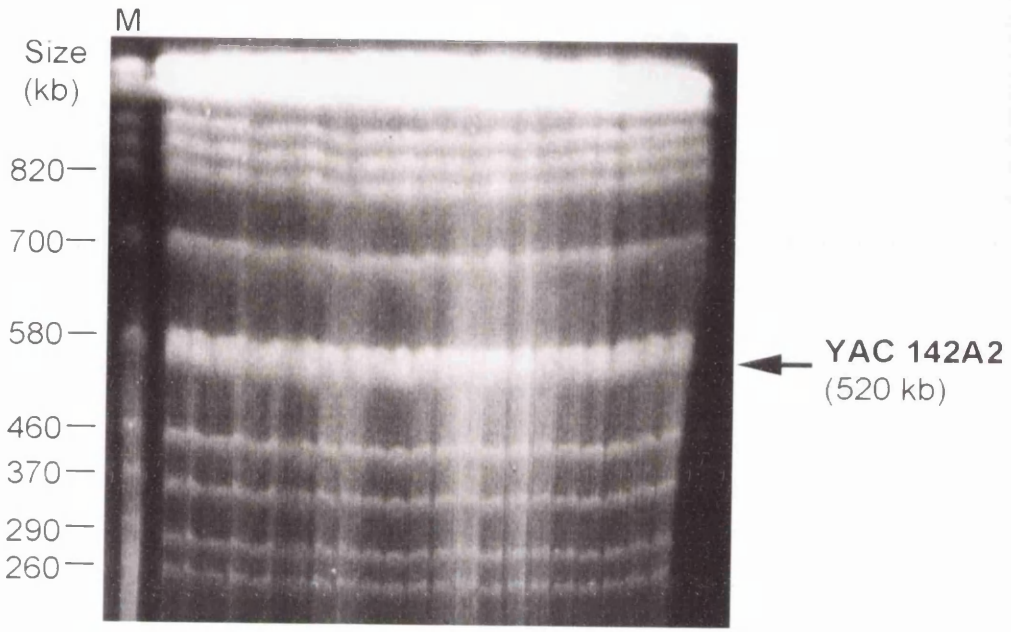
A: PFGE gel of YAC 132D4

Total yeast DNA containing YAC 132D4 was electrophoresed at 185V for 24 hours, with an alternating pulse-time of 60 seconds. The artificial chromosome migrated just above the 580kb yeast chromosome, and could not be fully separated during excision of the YAC band.



B: PFGE gel of YAC B230E3

Electrophoresis conditions for separation of YAC B230E3 were 180V for 22 hours with a 35 second alternating pulse.



C: PFGE gel of YAC 142A2

Yeast DNA containing YAC 142A2 was electrophoresed at 185V for 22 hours with a pulse-time of 50 seconds.The artificial chromosome migrated just ahead of the 580kb yeast chromosome, so the two bands could not be distinguished entirely during excision.

Figs 3.2.2 Purification of YACs using PFGE.

Photographs of the PFGE gels containing YAC DNA separated from the yeast chromosomes are shown. The DNA was stained with ethidium bromide and visualised under UV light. The arrows on the right side of each photograph indicate which bands represent the artificial chromosome. The size of the YAC band is written in bold type on the left. Host strain *S. cerevisiae* AB1380 DNA was used as the marker (M) for the native yeast chromosomes, enabling the YAC band to be identified for excision. The sizes of the host yeast chromosome are labelled on the left of each picture. Several of the higher molecular weight yeast chromosomes are indistiguishable at the top of each gel, in the 'zone of no resolution'(ZNF).

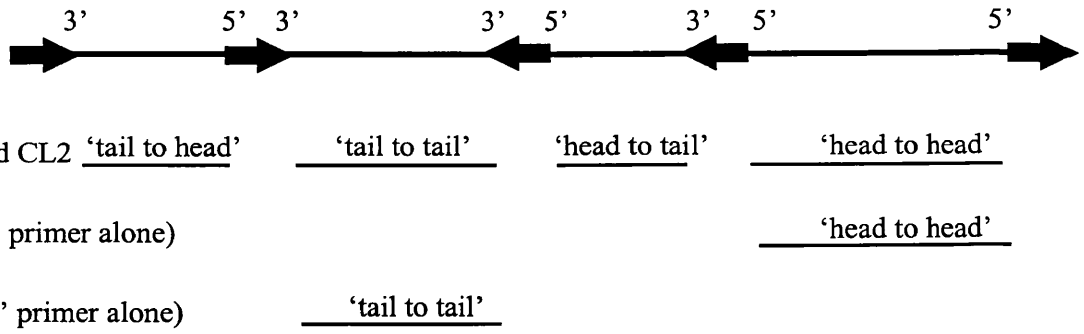


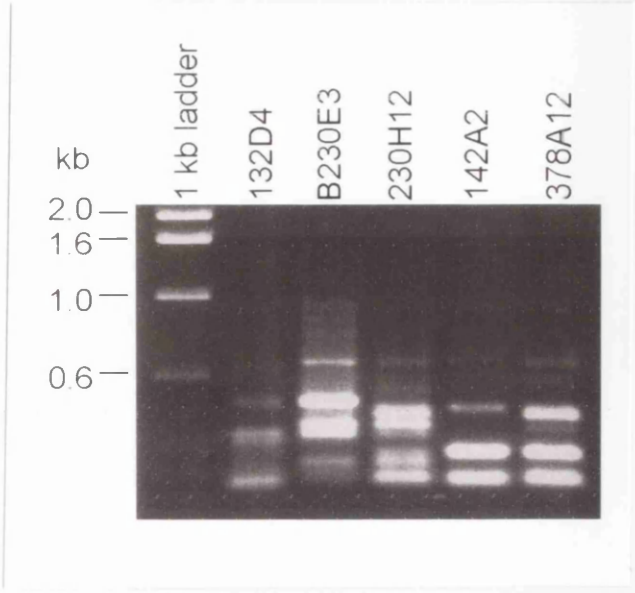
Fig 3.2.3 Possible *inter*-Alu PCR products generated with primers CL1 and CL2.

Alu elements are denoted by arrows, pointing 5' to 3'. Use of Alu primers CL1 and CL2 combined generates PCR products between adjacent Alu elements irrespective of their orientation. When used separately, amplification with CL1 gives a product when adjacent Alu elements are positioned in 'head to head' fashion, whereas amplification with CL2 alone forms products between Alu elements that are orientated 'tail to tail'.

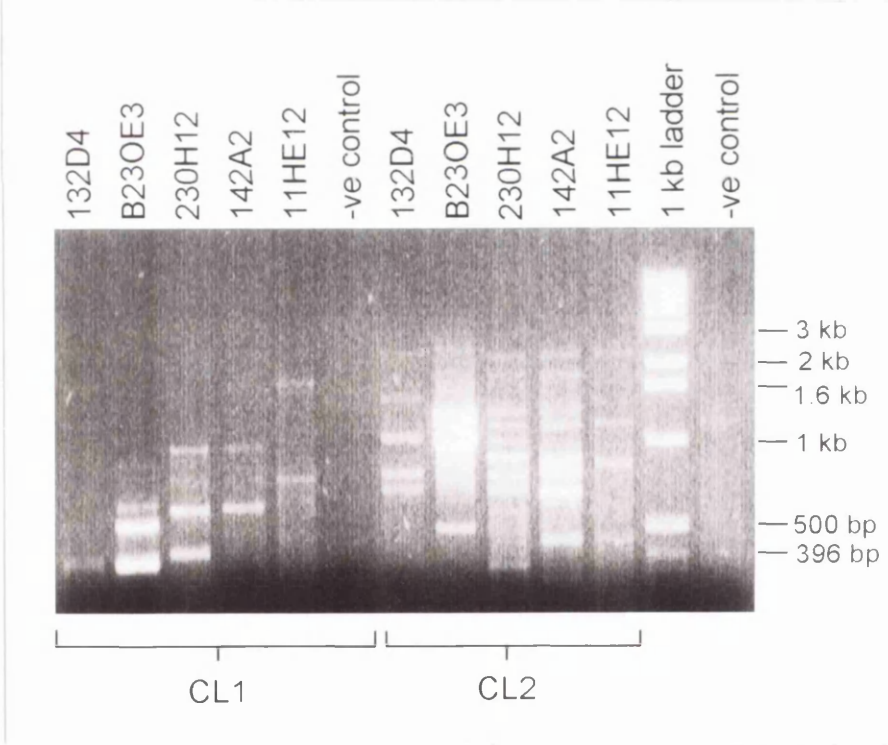
Combination of both Alu primers CL1 and CL2 in the same PCR reaction generated a set of amplification products from of 100-600bp from YACs from 15q11-13 (Fig 3.2.4A). On the basis that there are 500,000 copies of Alu sequences in the haploid genome, the average spacing between Alu elements is 5kb, although actual spacing of 700bp-8kb apart has been observed (Schmid and Jelinek, 1982). This indicates the presence of a large number of closely spaced Alu sequences in the PWS and AS region. To generate larger PCR products without causing a deleterious reduction in sequence complexity, amplification of the YACs was performed using CL1 and CL2 separately. Fragments ranging from 350bp to 3kb were produced (Fig 3.2.4B). The products of the latter reactions were combined in equal proportions and fixed onto nylon membranes for use as probes in the direct selection. The *inter*-Alu PCR probes are referred to as '132D4-Alu', '230H12-Alu', 'B230E3-Alu' and '142A2-Alu' respectively.

3.2.1.3 PAC and BAC clones as genomic probes

Overlapping PAC and BAC clones were combined into three separate groups for use as genomic probes. PACs 50I2, 14I12 and 318N23 were combined to form a single probe, 'PAC-1'. PACs 209N9, 26N4, 113G10 and 5501 were grouped into probe 'PAC-2'. The four BACs formed a single 'BAC' probe (Fig 3.1.7). Equal quantities of DNA from each genomic clone were combined and fixed onto nylon membranes without prior restriction digestion.



A: Amplification of YACs using primers CL1 and CL2 combined in each reaction.



B: Amplification products using primers CL1 and CL2 in separate reactions.

Fig 3.2.4 *Inter*-Alu PCR amplification products of YACs from 15q11-13.

The *inter*-Alu PCR products of YACs from 15q11-13, amplified with primers CL1 and CL2, are shown. The products visible in the B230E3 PCRs, should also be common to 132D4, since 132D4 encompasses B230E3. This is most likely due to poor amplification of 132D4.

3.2.2 cDNAs for selection

3.2.2.1 Source of cDNAs

Direct cDNA selection is based on gene expression, its main limitation is that isolation of cDNAs encoded by a genomic contig, is reliant on the presence of the mRNA in the chosen tissue source. Many genes are expressed in a spatial or temporal specific manner. cDNAs derived from the correct source of tissue at the correct time period are required for such genes to be identified by cDNA selection (Lovett, 1994). Since AS and PWS are congenital disorders characterised predominantly by neurobehavioural anomalies, the genes involved in both phenotypes were predicted to be expressed during early fetal brain development. The majority of neural development and cell differentiation within the central nervous system occurs in the first nine weeks of gestation. Temporally restricted genes that are important for normal brain development are most likely to be expressed within this short window of time. Following this period, the major organs of the body, including the brain, grow and mature (McLauchlan, 1994). Although infant brain and 18 week fetal brain libraries were commercially available, these may not have reflected the full sequence complexity present in a library constructed from first trimester brain.

Two nine week fetal brain cDNA libraries, one constructed using poly dT reverse transcription primers (Br32) and the other random primed (Br33), were used as the cDNA source (refer to Materials, section 2.1).

3.2.2.2 Starting cDNAs

The starting cDNAs for use in the direct selection consisted of pooled amplified inserts from the Br32 and Br33 cDNA libraries. The cDNA inserts from each library were separately amplified using the outer M13 vector primers (Fig 3.2.5). The products were combined in equal proportions to form the 'starting cDNAs'. Size fractionation of the cDNAs for selection was not performed as detailed in the published protocol (Parimoo *et al.*, 1995).

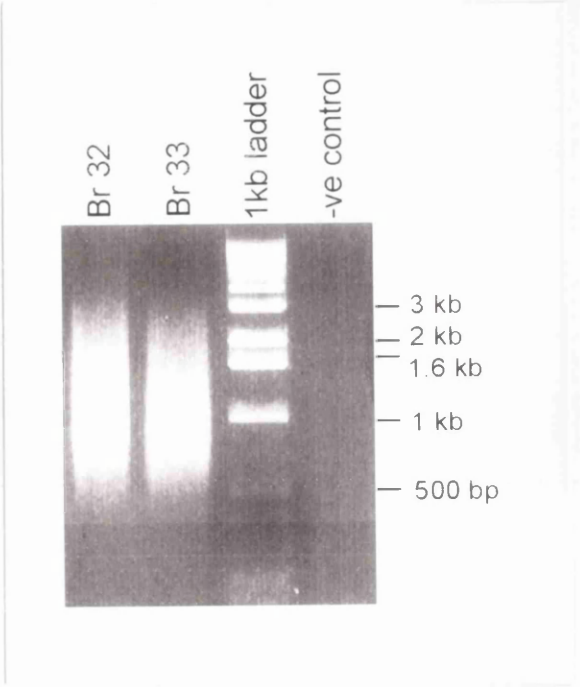


Fig 3.2.5 Amplified inserts from cDNA libraries Br32 and Br33, pooled to form the 'starting cDNAs' for direct selection.

An aliquot of each PCR reaction was electrophoresed on an agarose gel. The cDNAs are visualised as a smear from 500bp to 3kb, with an average length of about 1kb in length.

3.2.3 'Reporter' cDNAs acting as controls for the direct selection

To assess the efficacy of the direct selection, two reporter transcripts acting as 'positive' and 'negative' controls, were traced after each round of selection, and in the 'selected cDNA' sublibraries. The positive control is a known gene encoded by the target genomic contig, for which the cDNA is represented in the starting cDNA pool. A significant increase in the levels of this cDNA should be observed after each round of selection. The positive control cDNA should also be present at a higher frequency in the sublibrary than in the original starting library. The enrichment factor of this cDNA in the selected sublibrary should reflect that of co-selected novel cDNAs. The negative control is a housekeeping gene that maps outside the target genomic contig, but whose cDNA is present in the pool of starting cDNAs. Diminution of this cDNA should be apparent after each round of selection, providing a measure of the specificity of the selection (Lovett, 1994). The negative control used was the β-actin cDNA.

UBE3A had been mapped to 15q12 and demonstrated by RT-PCR, to be expressed in fetal brain of 20 weeks gestation (Nakao *et al.*, 1994). To confirm that *UBE3A* would be a suitable positive control for cDNA selection, PCR was performed to amplify the *UBE3A* transcript from the 'starting cDNAs'. The expected 1.5kb product from exons 4-8 (new exon numbers 10-14) was generated, showing that *UBE3A* was represented in the 'starting cDNA' population, and could be used as a selectable control (Fig 3.2.6). *UBE3A* was also amplified from fibroblast and lymphoblast cDNAs as PCR controls, and to generate hybridisation probes to map exons 4-8 of *UBE3A* to the YAC, PAC and BAC contig from ASCR (section 3.1.2.1; Table 3.1.3).

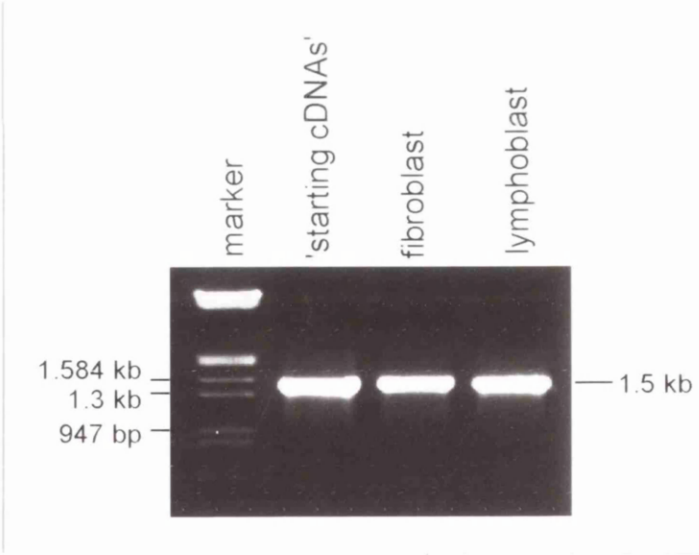


Fig. 3.2.6 Demonstration of *UBE3A* expression in the 'starting cDNA pool' used for the direct selection.

The predicted 1.5kb fragment from exons 4-8 (new numbers 10-14) of *UBE3A* was amplified from the 'starting cDNAs' using primers E6AP-4F and E6AP-8R. Fibroblast and lymphoblast cDNAs were also amplified as positive controls for the PCR, and to generate a hybridisation probe for mapping the exons.

Exons 1-8 (new exon numbers 7-14) of *UBE3A* were initially mapped to YACs 132D4 and B230E3, and subsequently to PAC clones 318N23, 14I12 or 5012 (Table 3.1.3), which were combined to form the 'PAC-1' probe for cDNA selection. The probe containing *UBE3A* exons 4-8 was used as the positive control for direct selection reactions performed using YACs 132D4 and B230E3, their Alu-PCR products and the 'PAC-1' contig as genomic probes (section 3.2.5). Exons U3 and U4, which mapped further distal (Table 3.1.3), were used retrospectively as a positive control for selections performed with the combined BAC clones as the genomic probe (section 3.2.5).

3.2.4 Direct cDNA selection

3.2.4.1 Primary selection

An excess of starting cDNAs was used in the first round of selection as 15q12 encoded transcripts represented a small proportion of the cDNA populatation. The primary selected cDNAs were rescued by PCR amplification using the outer M13 vector primers. An aliquot of each reaction was electrophoresed alongside an equal volume of the starting cDNAs (Figs 3.2.7 and 3.2.8), and Southern blotted for evaluation with the control probes (Figs 3.2.8 and 3.2.9). The remainder of the PCR product was purified, condensed and quantified for recycling in the second round of selection.

The yield of cDNAs selected by probes 132D4, B230E3-Alu, 230H12, and 142A2-Alu in the primary round was high, whereas the yields for 132D4-Alu, B230E3, 230H12-Alu and 142A2 were very low (Fig 3.2.7). In the case of the 132D4-Alu probe, this may have been due to poor *inter*-Alu PCR amplification (Fig 3.2.4). The average sizes of the selected products (500bp) were reduced compared to that of the starting cDNAs (1kb) (Figs 3.2.7, 3.2.8A). This may be attributable to the PCR rescue, which favours amplification of shorter DNA fragments. It may also reflect the hybridisation efficiency between genomic DNA, with its interspersed exons, and cDNAs. A further reduction in size was noted after the second cycle of selection (Fig 3.2.8B).

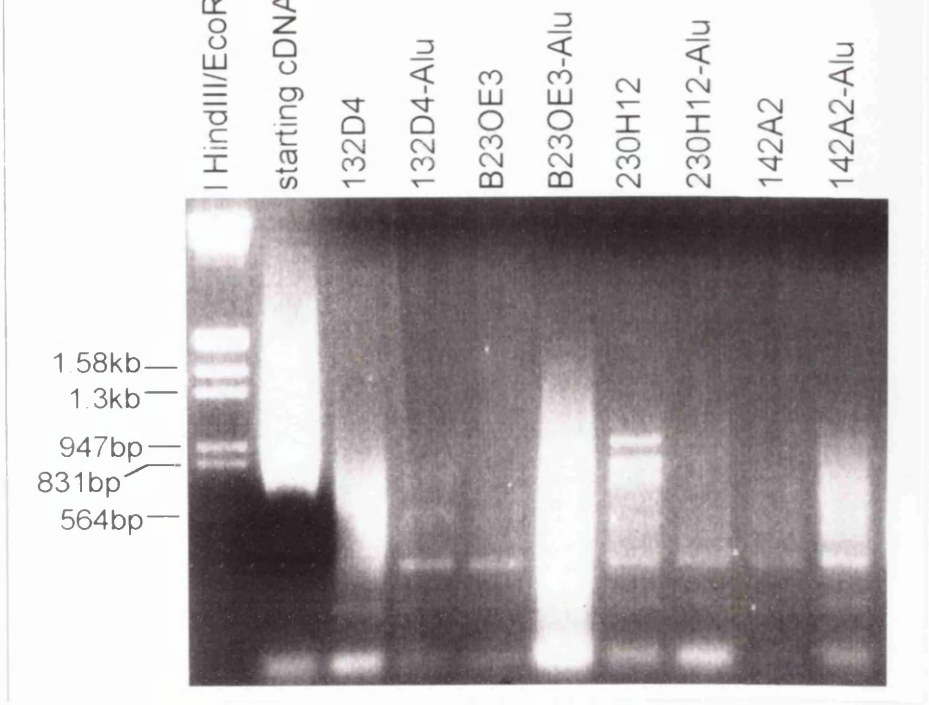


Fig 3.2.7 Primary selected cDNAs for the YAC and YAC-Alu probes

A photograph of the gel containing electrophoresed primary selected cDNAs, amplified with M13 primers, is shown. Strong PCR products are seen for probes 132D4, B230E3-Alu, 230H12 and 142A2-Alu, indicative of a high yield of primary selected cDNAs. Weak PCR products were visualised for probes 132D4-Alu, B230E3, 230H12-Alu and 142A2, indicating low cDNA yields.

3.2.4.2 Secondary selection

A reduced quantity of primary selected cDNAs was recycled in a second round of selection, since these cDNAs are enriched for cognate cDNAs. The secondary selected cDNAs were rescued by amplification with nested vector primers, T3 and T7. As *per* the primary round, an aliquot of the amplified secondary selected products was retained for assay with the reporter cDNAs. The remainder of the PCR product was purified and condensed for cloning. Specific bands appeared in the secondary selected products of B230E3-Alu and 142A2 indicative of selection of a particular cDNA species (Fig 3.2.9). However, abundance normalisation of high copy number cDNAs should have occurred, if the recycled primary selected cDNAs were present in excess of the genomic target (Morgan *et al.*, 1992). Despite repetition, selection with probes 132D4-Alu, B230E3, 230H12-Alu and 142A2 failed, with negligible amounts of cDNAs recovered after the second cycle. These cDNA populations were not cloned or analysed further.

3.2.5 Evaluation of the selected cDNAs with the positive and negative control cDNAs

Enrichment of *UBE3A* transcripts was detected after each round of direct selection specifically in the reactions that employed genomic clones encoding *UBE3A* as targets. This indicated that *UBE3A* transcripts were efficiently isolated with each selection cycle, which should reflect the selection of other cognate cDNAs present in the fetal brain cDNA libraries. Enrichment of *UBE3A* exons 4-8 was demonstrated in the selected products of YAC 132D4: a significant increase in hybridisation signal intensity was seen in the primary selected cDNAs compared to the 'starting cDNAs' (Fig 3.2.8A'), and in the secondary selected cDNAs with respect to the primary selected cDNAs (Figs 3.2.8B'). Similar results were obtained for the PAC-1 probe (data not shown). For the B230E3-Alu probe, increased levels of *UBE3A* were observed after the second round of selection (Fig 3.2.8B'), although the signal intensity in the primary selected product was reduced compared to the 'starting cDNAs (Fig 3.2.8A'). This is most likely due to a bias in quantity of products loaded on the gel, which is evident in the photographs of the electrophoresed cDNAs (Fig 3.2.8A and B). The amount of cDNAs electrophoresed was not equilibrated, as these were quantified after electrophoresis and purification. Increased levels of *UBE3A* exons U3-U4 were also observed after the second round of selection with the BAC contig, relative to the first round (Fig 3.2.9A). As expected, *UBE3A* cDNAs were undetectable in the selected products of YACs 230H12 and 142A2 (Fig 3.2.8A' and B'), and the PAC-2 probe (data not shown), since these probes did not contain the gene (Table 3.1.3). Since the same conditions were employed in all reactions, similar levels of enrichment of cognate cDNAs were expected for each of the probes.

In contrast to the results obtained with the positive control probe, the levels of the negative control β-actin, diminished after respective selections cycles. This is illustrated in the BAC-selected products, where progressively reduced signal intensities for β-actin were observed in the primary and secondary selected products (Fig 3.2.9B). This showed that selection of transcripts outside the genomic target region was inefficient, providing an indication of the specificity of the selection reactions.

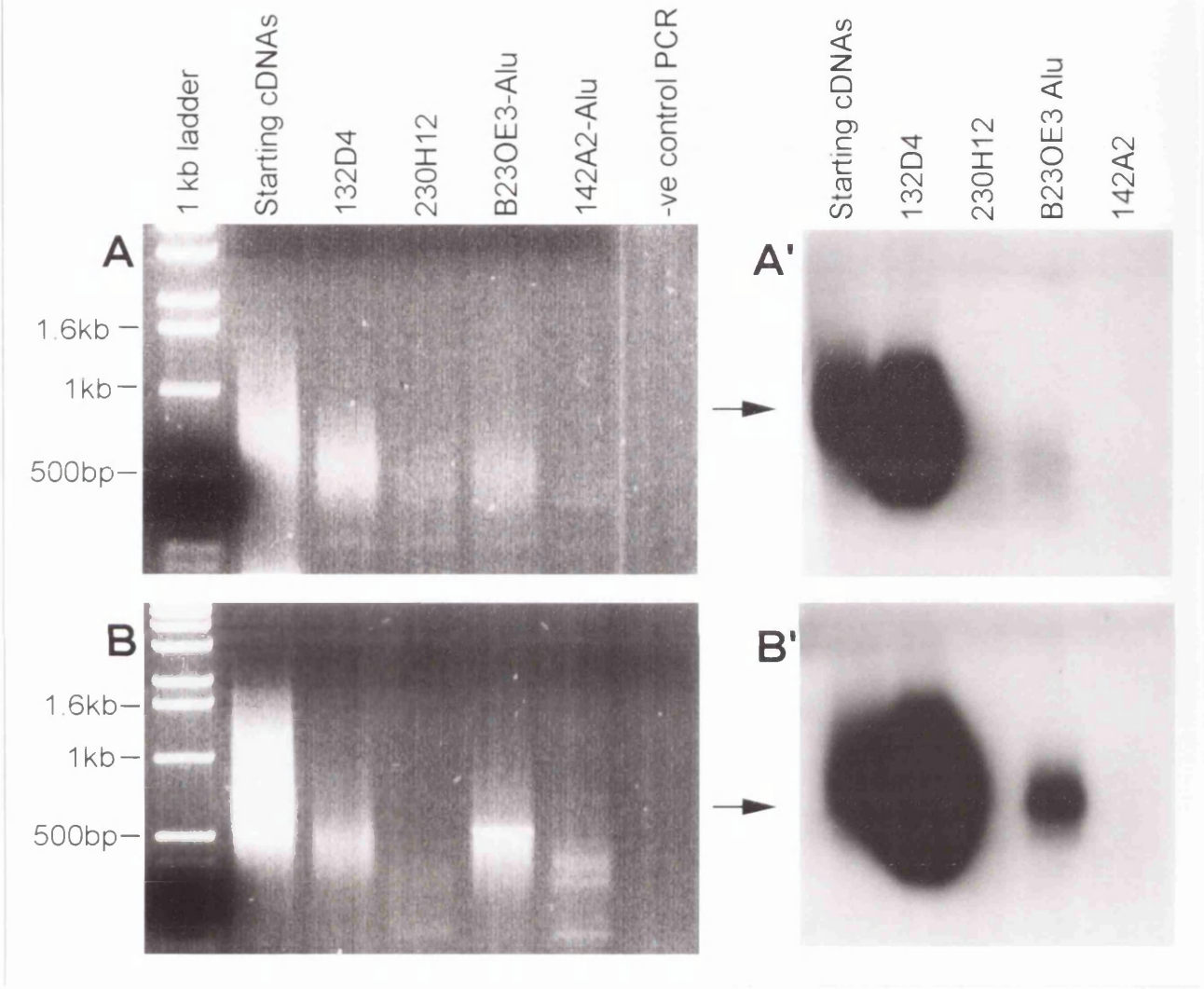


Fig 3.2.8 Evaluation of the primary and secondary selected cDNAs of the YACs, with the positive control reporter cDNA, *UBE3A* exons 4-8.

The gels containing electrophoresed amplified selected products are shown on the left. These were Southern blotted, and hybridisation of the corresponding membranes are shown on the right. Autoradiography was performed overnight in each case.
A Electrophoresed amplified ***primary*** selected cDNAs of YAC probes
A' Southern blot of **A** hybridised with *UBE3A* exons 4-8 (new exon numbers 10-14)
B Electrophoresed amplified ***secondary*** selected cDNAs of YAC probes
B' Southern blot of **B** hybridised with the *UBE3A* probe

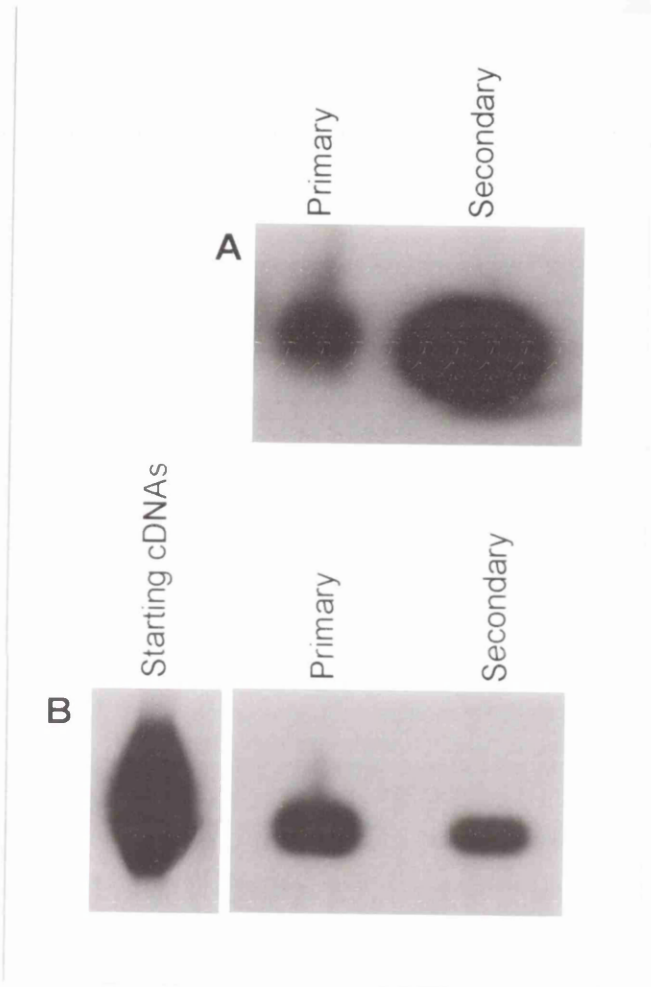


Fig 3.2.9 Assay of the primary and secondary BAC-selected cDNAs with the positive and negative control probes.

A Southern blot of BAC-selected cDNAs hybridised with the *UBE3A* exons U3-U4. An increase in signal intensity was observed with each selection cycle, indicating enrichment of *UBE3A*.
B The same blot subsequently hybridised with β-actin. A decrease in signal intensity with successive rounds of selection showed that β-actin cDNAs were not selected.

3.2.6 Production of arrayed sublibraries of the selected cDNAs

Sublibraries of the cDNAs selected by probes 132D4, B230E3-Alu, 230H12, 142A2-Alu, PAC-1, PAC-2 and BAC were produced for characterisation and storage of cDNAs encoded by 15q12. Thus a selected cDNA sublibrary representing each genomic clone was produced, despite the failure of certain probes to select sufficient yields of cDNAs. Selected cDNA products from the entire target genomic region, from D15S174 to D15S113, were cloned and analysed.

The secondary selected amplification products were cloned in the pTAg plasmid, which is specifically designed for cloning PCR products. This precluded the need for restriction digestion of the cDNAs for cloning, which may have resulted in loss of certain species as a result (Morgan *et al.*, 1992). The sublibraries were arrayed in 96-well format, in LB microtitre plates. The plates were labelled according to the sublibrary (the genomic probe used in the selection), with Greek letters α, β, δ, ε, ϕ, and Ω, denoting the plate number. A column number (1-12) and row letter (A-H), provided each clone with a co-ordinate position within its respective plate. The clones were propagated and stored in the LB plates during analysis of the arrayed sublibraries. Individual clones were traceable via their assigned plate numbers and co-ordinates after characterisation of the inserts by Southern analysis (section 3.2.7).

The protocol suggests that at least one cDNA clone per 2kb of genomic target used in their selection should be characterised (Parimoo *et al.*, 1995). Initially just two 96-well plates were filled for each sublibrary and analysed. Additional clones were arrayed and analysed to a total of 1 cDNA clone per kb of genomic target, from sublibraries yielding results at the 'positive screening' stage (section 3.2.7.2).

The inserts of the arrayed clones from each sublibrary were PCR-amplified in 96-well amplification plates, using vector primers pTAg-5' and pTAg-3'. The amplified inserts were electrophoresed in a 96-well configuration (Fig 3.2.10). These ranged from 200bp to 1kb in size. The gels containing the electrophoresed cDNA inserts were 'sandwich' Southern blotted to produce duplicate filters for rapid characterisation of the sublibaries by Southern hybridisation analyses.

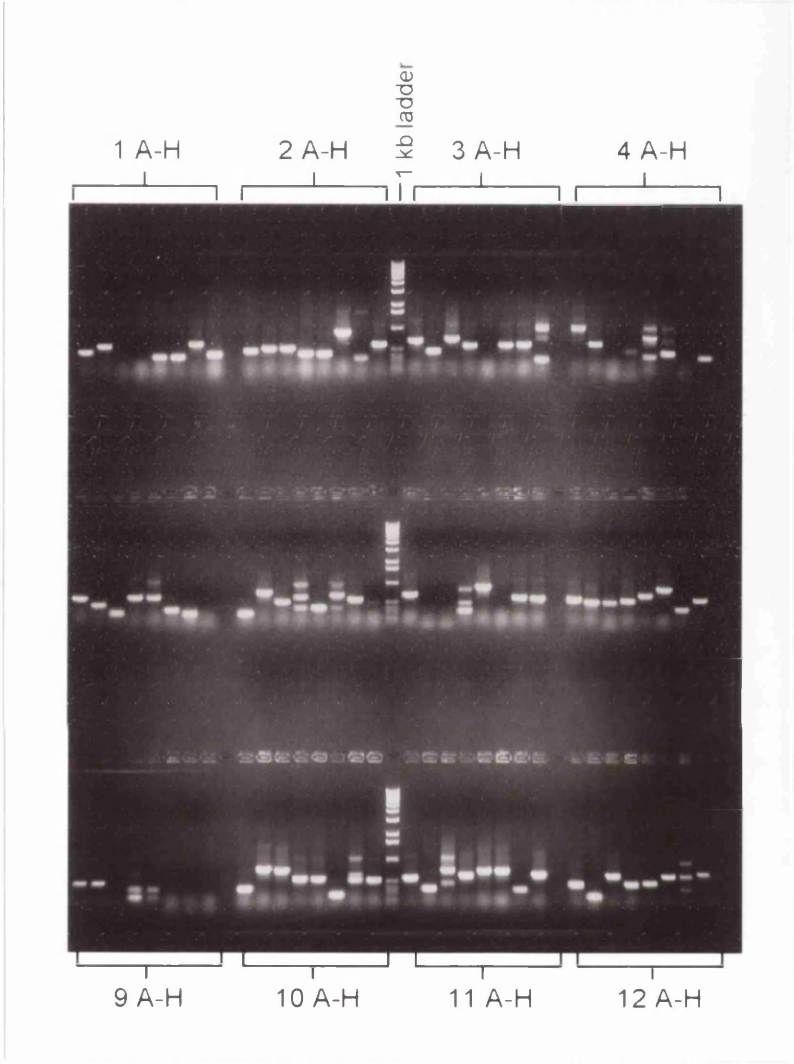


Fig 3.2.10 Electrophoresis of amplified cloned inserts for characterisation of the arrayed selected cDNA sublibraries.

The gel containing electrophoresed amplified inserts of the arrayed 'PAC-1' selected cDNA sublibrary, plate ϕ, is shown. The cDNA inserts were electrophoresed in groups which retained the 96-well configuration of the derivative arrayed sublibrary, for Southern blotting.. In some lanes more than one differently sized PCR product is visible, indicating that the colony picked was not derived from a single clone, or that more than one clone was picked into a single well (human error). Lanes devoid of amplification products most likely represent clones lacking cDNA inserts.

3.2.7 Characterisation of the selected cDNA sublibraries

3.2.7.1 'Counterscreen' of the arrayed sublibraries to identify contaminants

The selected cDNA sublibraries were analysed first for undesirable sequences including repeat elements, ribosomal cDNAs and host vector sequences for exclusion from analysis, in a process referred to as counterscreening (Morgan *et al.*, 1992). This was carried out by hybridisation of the Southern blots of selected cDNA inserts, with Cot1 DNA to identify human repetitive sequences (Fig 3.2.11), either *E. coli* or *S. cerevisiae* AB1380 DNA to identify host vector sequences and ribosomal probes (Fig 3.2.12). The co-ordinates of each insert were identified with reference to photographs taken of the electrophoresed amplification products, and recorded for exclusion from analysis. Table 3.2.1 provides a summary of the population of contaminants and their relative frequencies in each of the arrayed sublibraries.

A high frequency of ribosomal cDNAs are expected in selected sublibraries of PFGE purified YACs, since any co-migrating yeast chromosome included in the probe, will inevitably encode several copies of the yeast ribosomal locus. In selections where these sequences are not preblocked, the selected products may be comprised of up to 70% ribosomal cDNAs (Lovett, 1994). However, the frequency of ribosomal cDNAs in each of the selected sublibraries was less than 5% (Table 3.2.1), indicating that inclusion of ribosomal sequences in the preblocking step had prevented their selection.

In selections using genomic clones propagated in *E.coli*, mitochondrial cDNAs are a potential contaminant as the mitochondrial genome is circular and may have been co-purified with the PACs and BACs (Lovett, 1994). However, due to the low frequencies of ribosomal and host vector sequences identified, the selected sublibraries were not screened by hybridisation for mitochondrial cDNAs. Instead, these were identified after sequencing of cDNA inserts and database searching, and were found to represent less than 5% of the cDNAs in the sublibraries. Human repetitive sequences were the most abundant contaminant of the sublibraries, representing up to 20% of the cDNA clones, despite pre-blocking with Cot1 DNA prior to selection (Table 3.2.1). These latter figures were based on hybridisation analyses alone, but increased after sequence analysis of individual clones. This may be a reflection of the high number of repeat elements present within 15q11-13, and the probes.

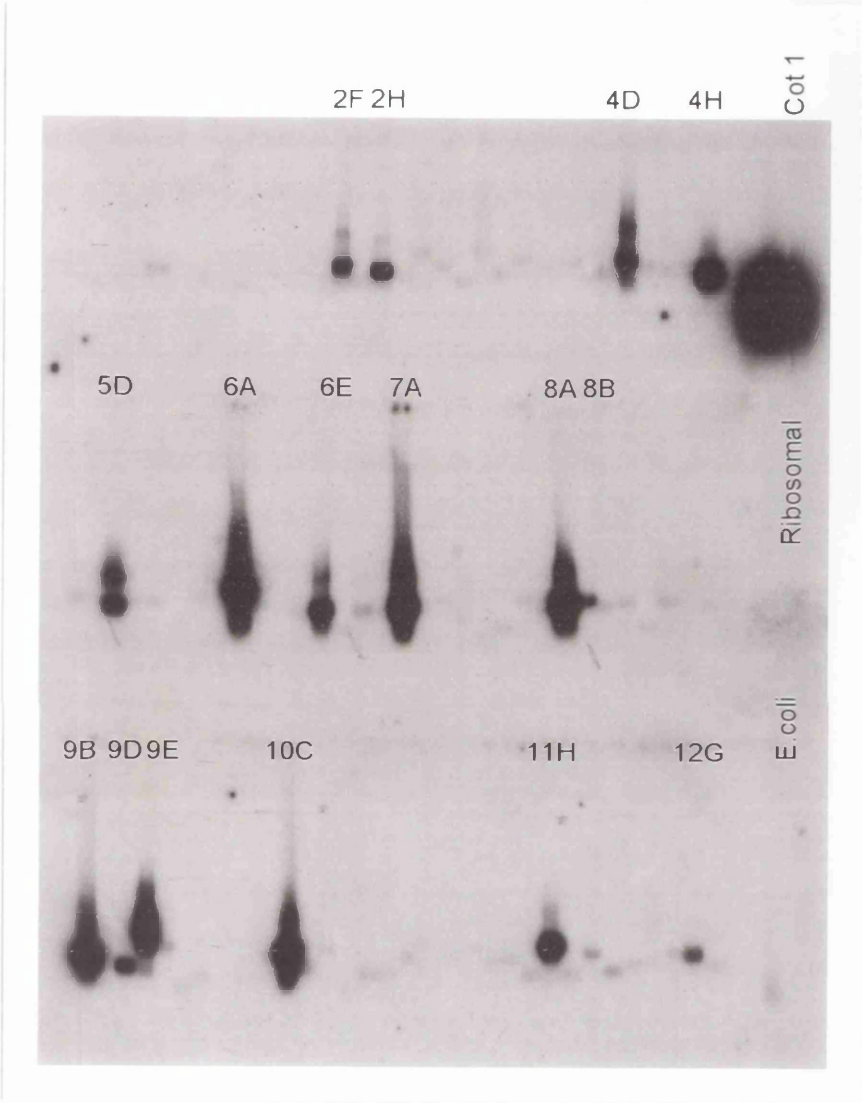
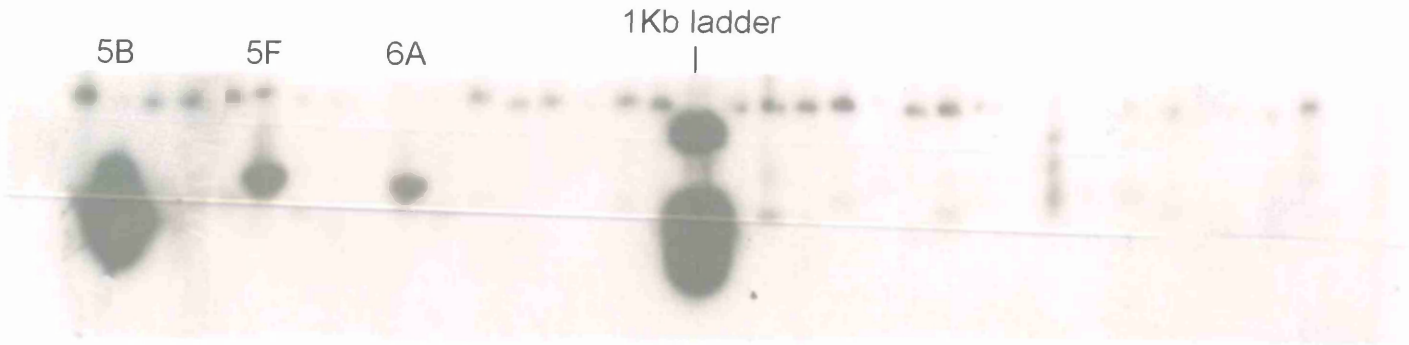
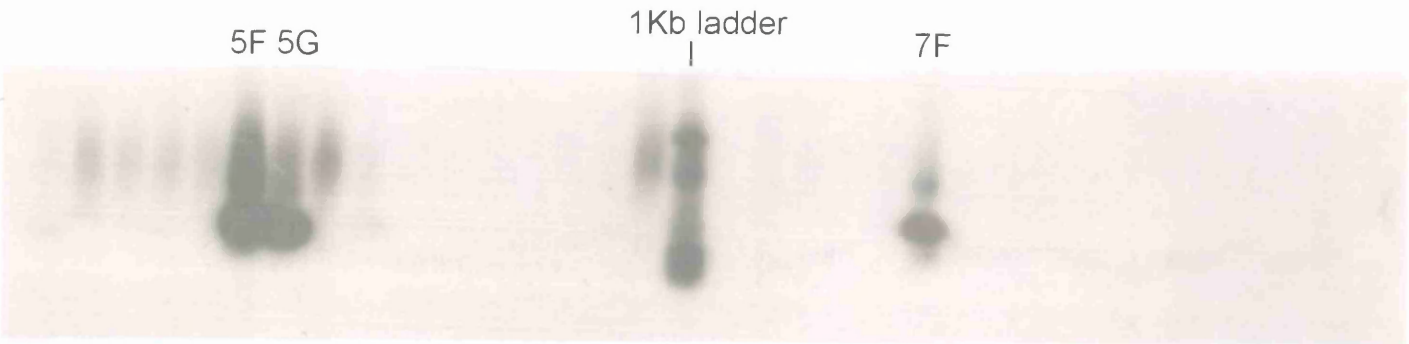


Fig 3.2.11 Identification of unwanted repeat sequences from the PAC-1 selected sublibrary.

The arrayed inserts from the PAC-1 selected sublibrary plate α, hybridised with Cot1 DNA, are shown. The co-ordinates of the repetitive clones giving hybridisation signals are labelled above. The Cot1 DNA loaded on the gel served as a hybridisation control and facilitated orientation of the blot. The blots were washed to 1X SSC/ 0.1% SDS at 65°C and autoradiographed overnight.



Ribosomal: 132D4 plate β, columns 5, 6, 7 and 8, clones A-H



E. coli: PAC-2 plate α, columns 5, 6, 7 and 8, clones A-H

Fig 3.2.12 Identification of unwanted ribosomal and host vector sequences.

A section of the autoradiographs from counterscreens using the human ribosomal probe (top) and total *E. coli* DNA (bottom), are shown for two sublibraries. The plate number and section of the plate from which the inserts are shon, are listed beneath each diagram. The clone co-ordinates identified in each screen are given above. The filters were washed to 1X SSC/0.1% SDS at 65°C for 20 mins, then autoradiographed overnight.

3.2.7.2 'Positive screen' of the arrayed cDNA sublibraries

In order to rapidly identify selected cDNA clones mapping to 15q12, the inserts from each arrayed sublibrary were screened with the target genomic clone originally used for their selection. This process is referred to here as the 'positive hybridisation screen'. In the published protocols, the clones remaining in the sublibraries after the counterscreen were analysed by sequencing or mapping back to the genomic region of interest (Lovett *et al.*, 1991, Parimoo *et al.*, 1995). However, this would be highly labour intensive, requiring sequencing or mapping of numerous individual clones. By prelimarily identifying clones that hybridised to their cognate genomic clones, this effort would be considerably reduced.

Positive hybridisation screens of the 132D4 sublibrary, with YAC 132D4 DNA (Fig 3.2.13A) and the B230E3-Alu sublibrary, with YAC B230E3 DNA (Fig 3.2.15), are shown. The co-ordinates of inserts identified by their respective genomic probes were checked against those identified as contaminants in the counterscreen (section 3.2.7.1), since repetitive sequences present in the sublibraries were often detected in both hybridisation screens. Clones identified solely in the positive screen were retained for individual analysis. Table 3.2.1 gives the number and frequency of clones identified in the positive screen of each sublibrary. As indicated in the table, numerous clones were identified in the positive screen for the genomic region encompassed by YAC 132D4, from the 132D4, B230E3-Alu, PAC-1 and BAC selected sublibraries. However, very few cDNA clones from the sublibraries of genomic probes lying distal to D15S210 (Fig 3.1.7), including 230H12, 142A2 and PAC-2, were identified in this screen.

3.2.7.3 Screen for UBE3A cDNAs in the selected sublibraries

To determine whether *UBE3A* was represented amongst the clones being analysed, the inserts from sublibraries 132D4, B230E3-Alu, PAC-1 and BAC were screened using the designated *UBE3A* probe. The co-ordinates of the inserts identified by *UBE3A* were checked for inclusion in the panel of clones detected in the positive screens, since any selected *UBE3A* transcripts should also have been detected in this screen. *UBE3A* exons 4-8 (new exons 10-14) identified two clones from the 132D4 sublibrary; 9B from plate α and 3H from plate β (Fig 3.2.13B). These inserts were also detected in the positive screen (Fig 3.2.13.A).

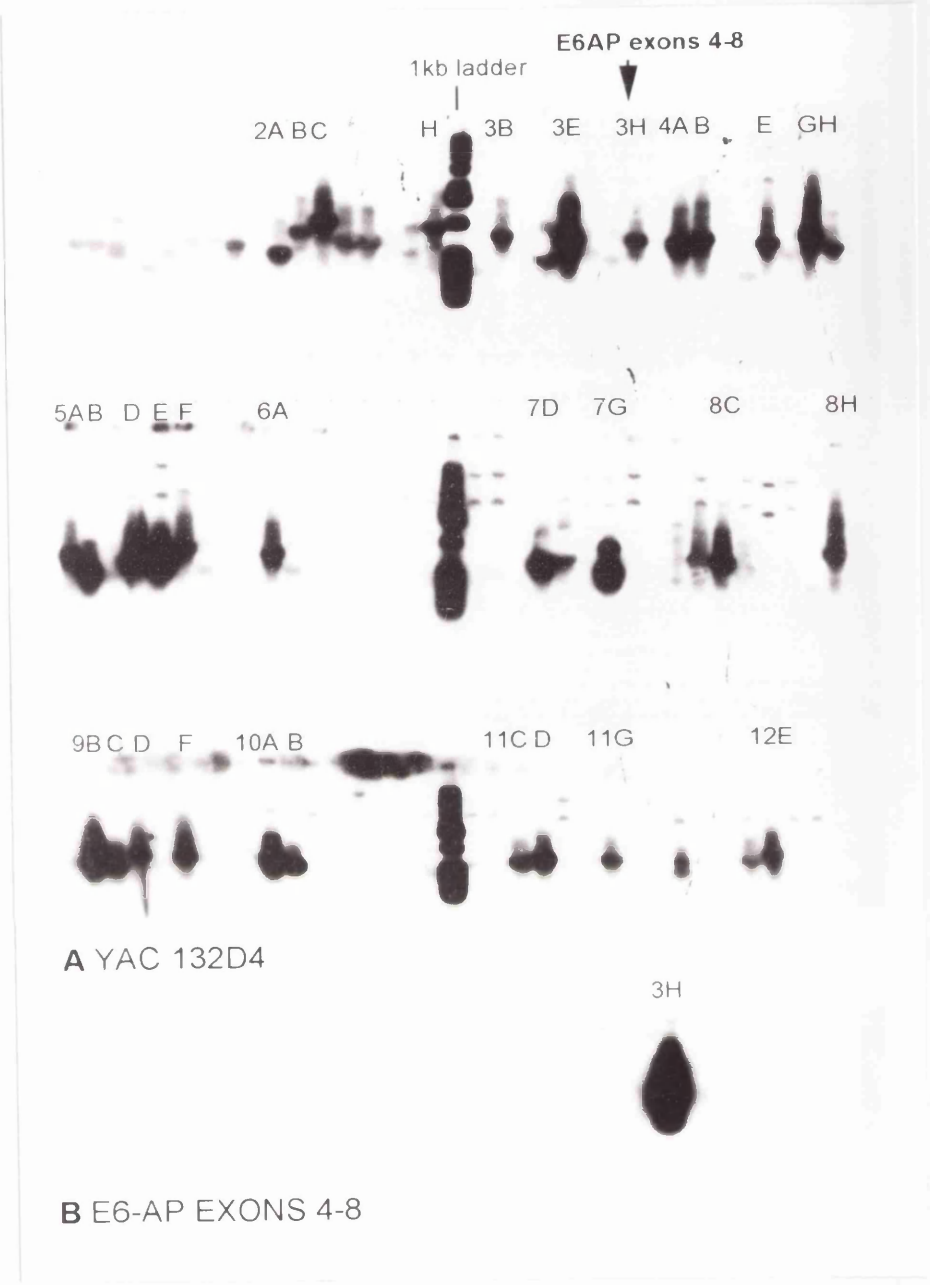


Fig 3.2.13 'Positive' screen of the gridded 132D4-selected cDNA sublibrary for clones potentially mapping to 132D4, and screen for *UBE3A* (E6-AP)

Arrayed inserts from 132D4-selected cDNA sublibrary plate β, hybridised with:
A: Purified YAC 132D4 DNA (positive screen) **B:** *UBE3A* (E6-AP) exons 4-8 (top third of autoradiograph). The co-ordinates of the inserts identified in either screen are labelled above the hybridisation signal. Clone 132D4β-3H was identified by both YAC 132D4 DNA and *UBE3A* exons 4-8.

Clones 3E and 12E from BAC sublibrary plate β were identified by *UBE3A* exons U3 and U4 (Fig 3.2.15). These clones were also identified in the positive hybridisation screen and were sequenced (Table 3.2.2).

The results of the screens for *UBE3A* in the arrayed 132D4 and BAC selected sublibraries showed that the positive control cDNA was amongst the clones being analysed. *UBE3A* had been selected, picked at random, gridded, and identified in the positive hybridisation screen. This indicated that other cognate cDNAs should be present at similar frequencies amongst the clones identified in the positive hybridisation screens.

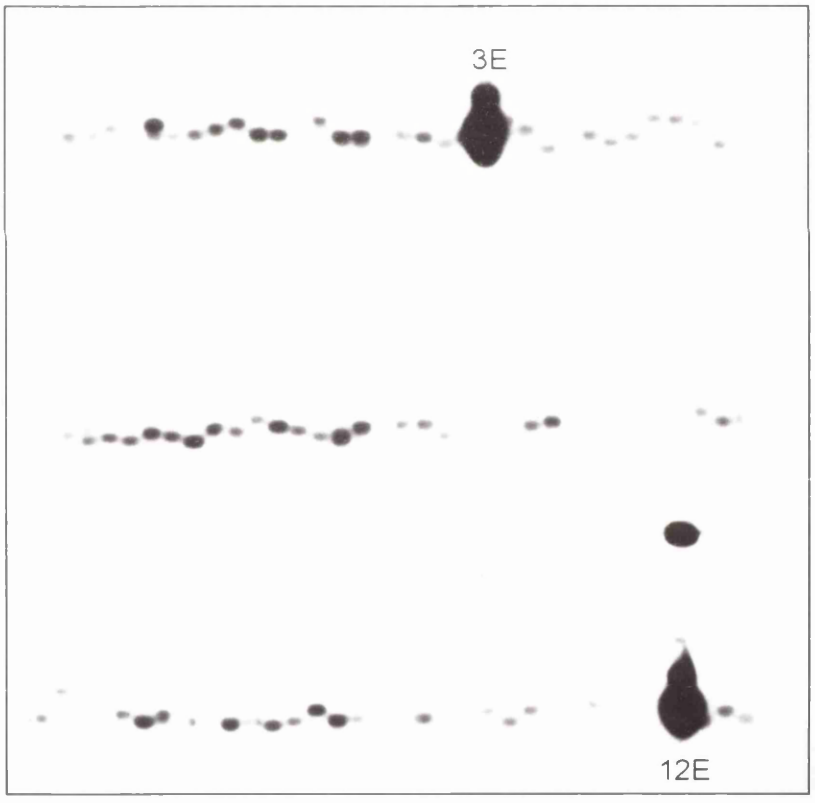


Fig 3.2.14 Screen of the arrayed BAC-selected cDNA sublibrary for *UBE3A* clones.

Autoradiograph showing the inserts from arrayed BAC-selected cDNA sublibray plate β, hybridised with *UBE3A* exons U3-U4. Inserts 3E and 12E identified by *UBE3A*, were also identified in the 'positive screen' with BAC DNA.

*3.2.7.4 Identification of redundant clones in the B230E3-*Alu *selected cDNA sublibrary*

In order to identify redundancy in the selected cDNA sublibraries, pools of clones from rows and columns of the gridded PCR plates may be used as hybridisation probes to screen the remainder of the sublibrary. This reduces the number of clones for sequencing (Parimoo *et al.*, 1995). However, the positive hybridisation screen performed here considerably reduced the number of clones to be analysed by sequencing, so this step of the protocol was not performed. However, the majority of the clones identified in the positive screen of the arrayed B230E3-Alu sublibrary were the notably of the same size; ~750bp. In total, 59 inserts of ~750bp were identified by hybridisation with YAC B230E3 DNA, suggestive of sequence redundancy (Fig 3.2.15). These 59 inserts were electrophoresed, Southern blotted, and screened with one of the 750bp inserts, B230E3-Alu α1E. This identified 47 of the 59 inserts, indicating there was a high level of sequence redundancy in this sublibrary (Fig 3.2.16). These sequences most likely represented the ~750bp band prominent in the electrophoresed secondary selected cDNAs of B230E3-Alu (Fig 3.2.8). The twelve non-redundant B230E3-Alu clones not identified by 1E (labelled in Fig 3.2.16), were retained for sequencing.

Subsequent sequence analysis of clone 1E revealed that it consisted entirely of an Alu repeat element, representing an artifact of the selection process using the *inter*-Alu PCR probe of YAC B230E3 as the target. This sequence is referred to in section 3.3.2 as selected sequence 21 (Table 3.3.1). The use of *inter*-Alu primers CL1 and CL2 amplifies the intervening sequences between adjacent Alu repeats (Fig 3.2.3). However, if used separately, Alu elements may have been amplified across and included within the PCR probe. For instance if CL1 amplifies fragments orientated 'head to head' (Fig 3.2.3), Alu repeats orientated 'head to tail' may have been included within the PCR product. This would have led to the selection of numerous species containing Alu repeats such as the clone 1E.

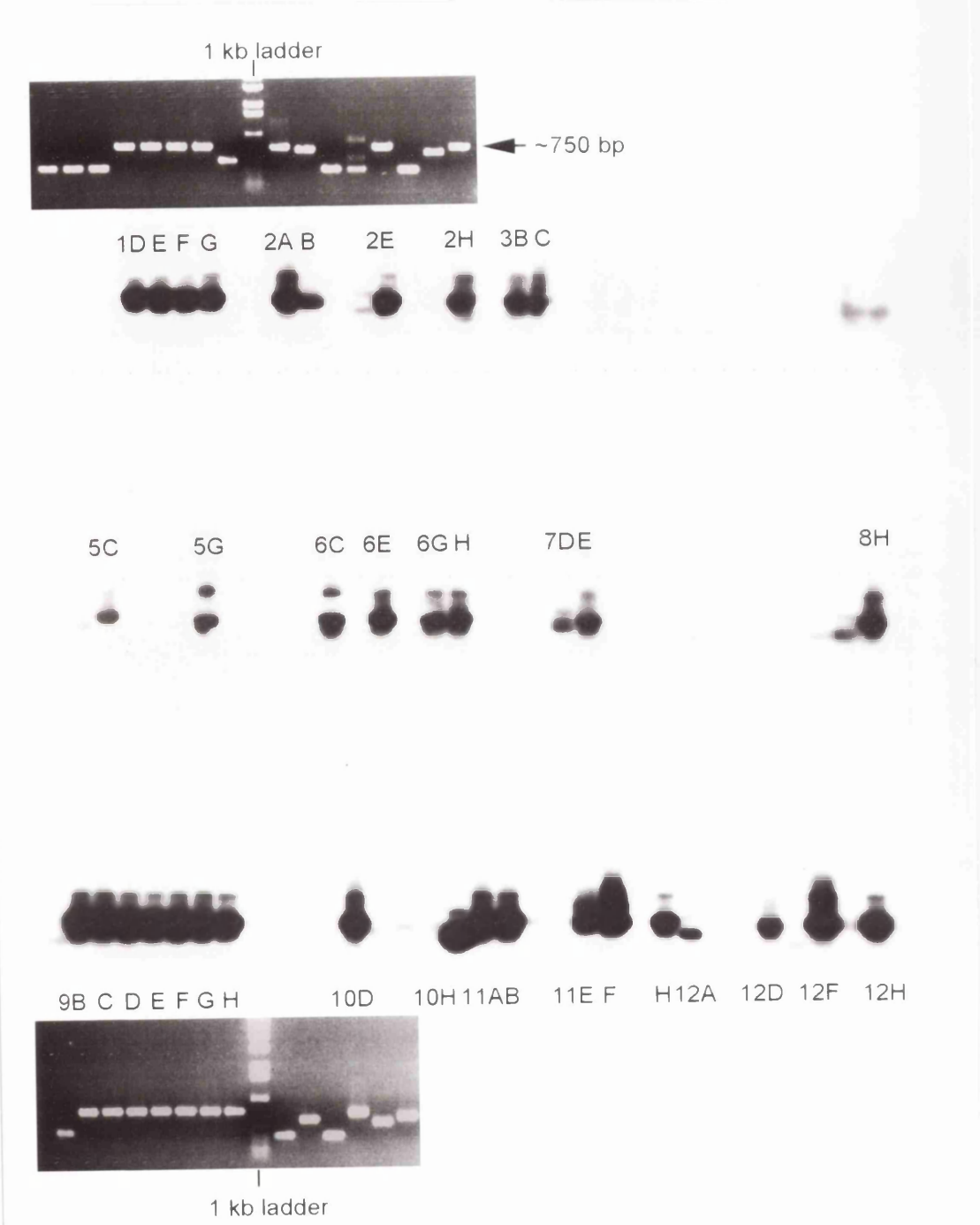


Fig 3.2.15 'Positive' screen of arrayed B230E3-Alu selected sublibrary, plate α.

The centre picture is an autoradiograph of the inserts from B230E3-Alu sublibrary plate α, hybridised with YAC B230E3. Above and below are the corresponding photographs of the electrophoresed inserts. These were sized with reference to the 1kb ladder, included on the gel. The majority of the inserts identified in this screen were the same size, ~750bp.

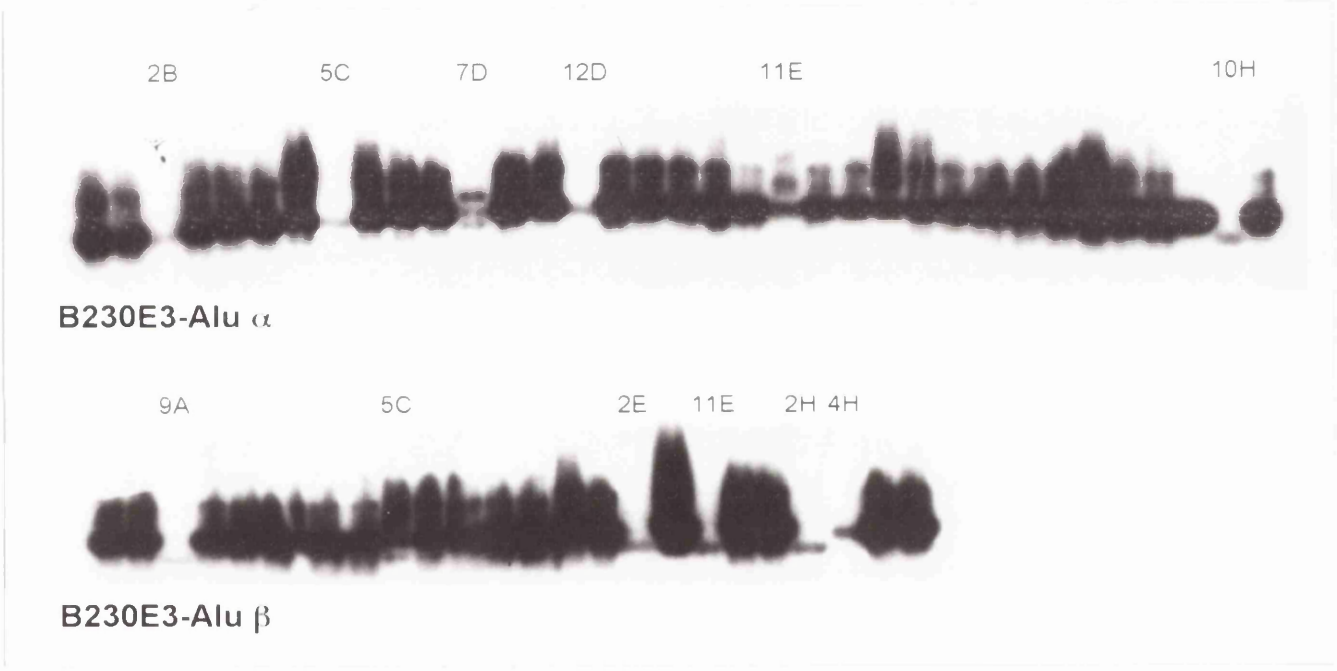


Fig 3.2.16 Assessment of redundancy amongst B230E3-Alu clones identified in the 'positive' screen

Autoradiograph of the 59 inserts of approximately 750bp, identified for plates α (above) and β (below) of the B230E3-Alu sublibrary, screened with clone B230E3-Alu α1E. The non-redundant inserts picked for further analysis are labelled.

Selected cDNA sublibrary	Probe size (kb)	Number of plates and (clones) analysed	Frequency of ribosomal, mitochondrial and host vector sequences (%)	Frequency of repeats (%)	Total frequency of clones detected in the counterscreen	Number and (frequency) of clones detected in 'positive' screen	Number of clones sequenced
132D4	600	6 (576)	5	10	15	45 (7.8%)	45
230H12	520	3 (288)	2.8	4.2	7	3 (1%)	3
B230E3-Alu	275	2 (192)	2.1	19.3	21.4	59 (38.8%)	13
142A2-Alu	520	2 (192)	0	2.6	2.6	6 (3.1%)	6
PAC-C1	~300	4 (384)	2.3	10	12.3	29 (7.6%)	29
BAC	~300	2 (192)	1	6.3	7.3	28 (14.6%)	28
PAC-C2	~300	2 (192)	2	8	10	8 (4.2%)	8

Table 3.2.1 Summary of the characterisation of the arrayed selected cDNA sublibraries by hybridisation analysis.

The number of clones from the selected cDNA sublibraries analysed are given, with the number of clones excluded from the study in the counterscreen and the number identified in the positive hybridisation screen for further analysis. For genomic probes 230H12, 142A2-*Alu* and PAC-2 fewer than 1 clone per kb were analysed as so few clones were positively identified from amongst those analysed. For the B230E3-*Alu* sublibrary just 13 of the clones positively identified were sequenced due to a high level of sequence redundancy in this selected sublibrary.

3.2.8 Sequencing of selected cDNAs identified in the positive screen of the sublibraries

The inserts of the clones identified by the genomic probes in the positive screen of the arrayed sublibraries were individually characterised by sequence analysis. A total of 132 inserts were sequenced on an automated ABI PRISM 377 DNA sequencer. The number of clones from each sublibrary that were sequenced is listed in Table 3.2.1, with the exception of the B230E3-Alu sublibrary, for which 13 clones were sequenced (section 3.2.7.4). For simplicity, each of these clones was assigned a sequence number from 1 to 132, in addition to its sublibrary plate co-ordinate (Tables 3.2.2 and 3.2.3).

The cloned inserts were sequenced from both the 5' and 3' ends using M13 or pTAg vector primers to obtain accurate sequence for the full-length of each insert. The sequence produced with one primer was directly compared to the reverse-complement of the sequence from the other primer, using the GAP program. This showed the overlap between the 5' and 3' end sequences for each cloned insert, enabling a consensus sequence to be derived and the sequence data to be optimised (Fig 3.2.17).

3.2.8.1 Sequence redundancy

Certain sequences were selected repeatedly and represented by several clones in the cDNA sublibraries. To minimise sequence redundancy prior to database searches and sequence analysis, the insert sequences were compared with one-another using the multiple alignment tool, Pileup (Fig 3.2.18). Firstly, the sequences of clones derived from the same selected sublibraries were compared with one another. The non-redundant sequences from each sublibrary were then compared to those of other sublibraries, where the genomic clones used in their selection overlapped (Fig 3.1.7). For instance, the sequences selected by the PAC-1 and B230E3-Alu probes were compared to those selected by YAC 132D4. This reduced the number of individual sequences analysed to 99.

```
GAP of: 11-ptag3.seq (length 378bp)to the reverse-complement of 11-
ptag5.seq (length 466bp).


           Quality:   3187              Length:    480
             Ratio:  8.431               Gaps:      1
 Percent Similarity: 96.429   Percent Identity: 91.369


 11-ptag3.seq x 11-ptag5.rev

                .         .         .         .         .
       1 GGCACGAGTTTTCAAGTTAAGTATTTTTATTAGGTTCCTTTTTATAATTT 50
                                                     |||||
       1 .........................................ATAAATTT 8
                .         .         .         .         .
      51 CTTTTTGTTGATATTCT.CATTTTATTACACATAATTTGTCTGATTTCCA 99
         ||:   |: |||||| | :|||||||| |:|||||||:|||:||||||||
       9 CTNCGGGNGGATATTGTCNATTTTATAANACATAATTNGTCNGATTTCCA 58
                .         .         .         .         .
     100 TTAGTTTTTTTCTTTGTTTTCCTTTAGCTCTTTGAAAATATTTAAGACAT 149
         ||||||||||||:||:||||||||::||||:|| |:||||||||:||||||
      59 TTAGTTTTTTTNTTNGTTTTCCTNNAGCTNTTGGNAAATATTTNAGACAT 108
                .         .         .         .         .
     150 TTTAAAAGTCTTTATCCAAGTTCAATTTCTATGGTTCTGTAGAGATATTT 199
         ||||||||||||:||||||||||||||||||| ||||| ||||||||||||
     109 TTTAAAAGTCTTNATCCAAGTTCAATTTCTAGGGTTCGGTAGAGATATTT 158
                .         .         .         .         .
     200 TCTGCCAGTTTATGTTCTTCTTTTCCATGGGCCATGTTTTCCTGTTTCTT 249
         |||||||||||||||||||||||||||||||||:|||||||||||||||
     159 TCTGCCAGTTTATGTTCTTCTTTTCCATGGGCCANGTTTTCCTGTTTCTT 208
                .         .         .         .         .
     250 TGTATACTTTCTAATTTTTGGTTGAAAACTGAGCATTTGAAAATAGAGCC 299
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     209 TGTATACTTTCTAATTTTTGGTTGAAAACTGAGCATTTGAAAATAGAGCC 258
                .         .         .         .         .
     300 AACTTTCCCAGTTTCTGCAGAGAGTCTTTATGCCACAGTATTCGTTCACT 349
         |||||||||||||||||||||||||||||||||||||||||||||||||
     259 AACTTTCCCAGTTTCTGCAGAGAGTCTTTATGCCACAGTATTCGTTCACT 308
                .         .         .         .         .
     350 GATTAACTGGGTATATCTAAGCTTANGGA..................... 378
         |||||||||||||:||||||||||||:|||
     309 GATTAACTGGGTANATCTAAGCTTAGGGAGCAGCTGAGTCAAAAGTTTAA 358
```

Fig 3.2.17 Obtainment of the full-length insert sequence for the selected cDNAs using 'GAP'

The 'GAP' output for sequence number 11 from the 3' end (11-ptag3) to the reverse complement of the sequence from the 5' end (11-ptag5) is shown. The two sequences demonstrated 96% similarity with a 332bp overlap. The length of sequence obtained with pTAg-3 was 378bp, and with pTAg-5 was 466bp. When corrected for the overlap, the full-length consensus sequence of the insert was 508bp. Indeterminate nucleotides (N) present in the sequence from one end of the clone could be corrected from the overlapping sequence from the other end of the clone, using the Seqed program.

```
125-u  TCCAGAAAAA AGTAACATCT GTTATNGAAA CNTNAACAAG AAAAACAGAA
118-r  NACAGAAAAG CTAAAAATTG ATGATAACCC CATGTNCCAG CTGACTCATT
127-u  GACAGAAGAG CTAAAAATTG ATGATAACCC CATGTTCCAG CTGACTCATT
114-r  CACAAAANAC CTTAATTTTG ATGATNACCC CATGTNCCAN CTGACTCNTT
118-u  CGAGAAAACT TCTAATCTGC CTCTCTTCCC TCAGTGCCAG CAGGTTTATT
127-r  ~~~~AAAACT TCTAATCTGC CTCTCTTCCC TCAGTGCCAG CAGGTTTATT
114-u  ~~~~~~~~CT TCTAATCTGC CTCTCTTCCC TCANTGNCAG CAGGTTTATT
101-u  AGTGGACTGT CTCAATTACT CTAGTTTCAT ANTNAATCCT GACAACCAGC
107-r  AGTGGGCTGT CTCAATTACT CTAGTTTCAT AGTAAATCCT GACAACCAGC
117-u  ANTAAAAGGT GGTAATTTAT ACNAAGTCTT ACCGTGAGAG ACCATNAANC
125-r  TTTGTCTNCC NGAAAGTTTT GCTTCTCATT CGCCANACTC TNCCCANANT
101-r  GCCTAAATG. GAAAATTTAA GGGCTGAAAT ANTTAAAATN TTTCGTCAGG
107-u  GCCTAGATGT GAANANTTNA GGGCTGAAAT AGTTAAANTA TTTCGTCAGG
105-r  NCCCAAAAAA GANCAANTTN ANTCCCANGA ATGCCNAAAN ATNAGAANTC
119-r  AAGCTTATAT GGTAAAATTA CCTGTANTTA ANTGGGCATG TNATATGATG
117-r  ANGTCATAAG TAAAATNTTT CAATTAANAA AAATACNCTG AATACTAAAG
110-u  AAGAAGTGAT TTGGTAAATC ATGTTAAATT TAGAATTTAG GACAGATAAG
110-r  TTAGGGCTTA TCAGTANCAT CATTAATAAT GTNTTCATTT CACTAAAAAT
115-r  TCTATCATTT TAAAAATTAN CATTAGGAAA ATGANTCTAT TTAAAGCATC
105-u  TGTCGATTTT CTCTACTGAT TTTCTTTCCA ATTTTATTGA TTTCTACTGT
115-u  TGTAAAAGAA AACCGGATTG ATGCTTTAAA TAGACTCATT TTCCTAATGC
119-u  GGAGTAAAAC ATGATGCCNA AGTTTTTTGA CCAGAGCAGC TGAGAGAATG
```

Fig 3.2.18 Identification of redundancy amongst the selected sequences using 'Pileup'.

A section of the Pileup output for the 5' and 3' end sequences of the BAC-selected cDNA clones is shown. These sequences were generated with M13 universal (-u) and reverse (-r) primers. Sequences demonstrating similarity were aligned and stacked together. Sequences 114, 118 and 127 (in bold and underlined) were almost identical, as were sequences 101 and 107 (in bold and italics). 101 and 107 were cloned in opposite orientations, since 101-r is similar to 107-u, and vice-versa.

3.2.9 BLAST analysis of the selected sequences

The end product of the direct selection process and sublibrary characterisation is a set of short cloned fragments encoded by the genomic region of interest. These fragments may be used as hybridisation probes to screen cDNA libraries by conventional means to obtain larger cDNA clones for further analysis. However, the huge influx of EST sequences into public databases in recent years has meant that cDNA clones can be identified through BLAST searches of these databases instead, using short sequences as queries or 'cyberprobes'. cDNA clones from the databases are identified on the basis of sequence homology to the cyberprobe. Many of the ESTs in the public databases pertain to the 5' and 3' end sequences of cDNA clones, which are obtainable for research purposes through the I.M.A.G.E. consortium (section 1.9.3).

The 99 non-redundant selected sequences were used as BLAST queries to search the GenBank, Embl, New Embl sequences, EST, STS, Human, Rodent, and 'Alu-like repeat' databases (from the BLAST menu) to identify homologous cDNA clones which would aid in the characterisation of transcripts in the ASCR and PWSCR. These BLAST searches also determined which selected clones represented previously identified genes from 15q12, and which clones would be eliminated from the study on the basis of their sequence content.

3.2.9.1 Rejection of unwanted 'selected' clones

The results of the BLAST searches revealed that 60 selected sequences were mitochondrial, ribosomal, vector, or highly repetitive sequences, and so were eliminated from the study. All 8 of the PAC-2 selected clones identified in the positive screen and sequenced (Table 3.2.1), were found to be vector DNA. The 9 sequences selected by probes 230H12 and 142A2-Alu (Table 3.2.1), consisted of $(CA)_n$, Alu, L1 or Mer repetitive elements, as revealed through searches of the 'Alu-like repeat' database. Thus all the clones identified in the positive screen of selected sublibraries PAC-2, 230H12 and 142A2-Alu, were eliminated from the study. Direct cDNA selection in the region distal to D15S210 therefore failed to identify any expressed sequences(Fig 3.1.7). The number of sequences identified in the positive screen that remained to be studied was reduced to 39. These were derived from the 132D4, B230E3-Alu, PAC-1 and BAC selected sublibraries.

3.2.9.2 Identification of sequences from previously known genes and transcripts

Eight non-redundant selected cDNA sequences demonstrated significant sequence identity to previously known genes or ESTs in the AS and PWS regions. This was in addition to the 2 inserts from 132D4-selected sublibrary, which hybridised to *UBE3A* exons 4-8 (as shown for plate β in Fig 3.2.13B), but were not confirmed by sequencing. Notably, the two major genes that mapped to the target genomic region, *UBE3A* and *IPW*, were represented by several selected cDNA clones (Table 3.3.2). Each of these selected sequences was identified in the sublibrary of the genomic probe to which the known gene or EST mapped (Fig 3.1.7).

BLAST analysis of sequence 114 showed that the first 30bp were identical to the final 30bp of the alternatively spliced 5' UTR exon U3b of *UBE3A* (Vu and Hoffman, 1997). Nucleotides 30-176 of sequence 114 demonstrated 99% identity to the first 146bp of *UBE3A* exon U4 (Table 3.2.2). Sequence 114 therefore overlapped two alternatively spliced exons of *UBE3A* (Fig 3.2.19). Sequence 114 was identical to sequences 118 and 127 (Fig 3.2.18),

which had been identified by hybridisation with *UBE3A* exons U3 and U4 (Fig 3.2.14).

```
AF002225 Homo sapiens Angelman Syndrome Gene, E6-AP ubiquitin protein
ligase 3A (UBE3A) mRNA from promoter P2, 5'UTR.

    "alternatively spliced exon U3" nucleotides 1454. .1649
    "exon U4" nucleotides 1650. .1678

    1401 AAGTTTTGCTGCCTACGTGTTTGAAAACTTCTAATCTGCCTCTCTTCCCT 1450
                                 ||||||||||||||||||||||||||
       1 ......................AAAACTTCTAATCTGCCTCTCTTCCCT 27

    1451 CAGTGCCAGCAGGTTTATTTTTTGTTTTGCAAGCCAGCTCTGCCTCCTTA 1500
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      28 CAGTGCCAGCAGGTTTATTTTTTGTTTTGCAAGCCAGCTCTGCCTCCTTA 77

    1501 CAGTATGACATCTGATGCTGGAGGGTCGCACTTTCAAAAATGAGTCAGCT 1550
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      78 CAGTATGACATCTGATGCTGGAGGGTCGCACTTTCAAAAATGAGTCAGCT 127

    1551 GGTACATGGGGTTATCATCAATTTTTAGCTCTTCTGTCTGGGAGATACAA 1600
         |||||||||||||||||||||||||||||||||||||| |||||
     128 GGTACATGGGGTTATCATCAATTTTTAGCTCTTCTGTCTGGGTGATAC.. 175
```

Fig 3.2.19 BLAST analysis of selected sequence 114.

Sequence 114 (lower) identified the *UBE3A* 5' UTR from the P2 promoter (GenBank accession number AF002225) (upper). Nucleotides from exon U3b appear in bold. Exon U4 appears in normal text.

The frequencies at which the 'known' cDNAs were represented in each selected sublibrary were calculated as a percentage of the total number of clones arrayed (Table 3.2.2). These figures under-represent the true frequencies of the known cDNAs in each sublibrary, since clones that were initially studied as novel cDNAs, were later found to be extended sequences of *IPW* or *UBE3A*. The enrichment factors of *UBE3A* and *IPW* in the selected sublibraries was not assessed as the frequencies of either transcript in the starting fetal brain libraries were not estimated. Up to this point the number of clones identified in the positive hybridisation screen had been minimised through a process of elimination, either by hybridisation in the counterscreen or by sequence analysis. A further 31 remained for analysis.

Known 15q12 cDNA	Query sequence {redundants}	Co-ordinate {redundants}	Size (bp)	Results of BLAST search, hybridisation and direct sequence comparison: Percentage sequence identity (ID) and position of overlap between selected cDNAs and previously known transcripts	Frequency in sublibrary
UBE3A	114 {118} {127}	BACα-5G {BACβ-3E} {BACβ-12E}	176 176 176	99% ID to *UBE3A*, mRNA from promoter P2 5'UTR (AF002225). 99% ID between n30-176 of sequence 114, and *UBE3A* exon U4 (X98037), n1-146	1.56% in BAC sublibrary
	56 78 - -	132D4α-4C 132D4β-11C 132D4α-9B 132D4β-3H	341 326	96% ID to *UBE3A* exon 3, n122-462 (X98023) 97% ID to *UBE3A* exon 3, n120-438 (X98023) Based on hybridisation of *UBE3A* exons 4-8 (10-14) to arrayed 132D4 sublibrary. This was not confirmed by sequencing.	0.7% in 132D4 sublibrary
IPW	55 85 87	132D4α-4A 132D4ϕ-10F 132D4ϕ-12G	403 425 353	96% ID over n1-165 of *IPW*, (U12897) 94% ID to *IPW*, n1083-1449 82% ID to *IPW*, n849-1016	0.52% in 132D4 sulibrary
WI-6654	68	132D4α-4B	467	83% ID to human STS WI-6654 (G06008), n105-268	0.17% in 132D4 sublibray
PAR-2	69	132D4β-4G	411	88% ID over 290bp to DNA sequence from PAC 14I12 (AC00425). Identified as fragment of PAR-2 by 'Pileup' with PAR-2.	0.17% in 132D4 sublibrary

Table 3.2.2 Identity and frequency of 'known' 15q11-13 genes and transcripts represented in the arrayed selected cDNA sublibraries. Details of the 8 non-redundant selected sequences that identified previously known genes or transcripts from 15q11-13 by BLAST analysis are given. The 2 clones identifying *UBE3A* exons 4-8 by hybridisation are included. The clone co-ordinates and query sequence numbers of the selected sequences are provided. The percentage sequence identity and area of nucleotide (n) overlap between the query sequences and the known 15q11-13 transcripts are given. The accession numbers of the known sequences identified in the BLAST searches are in parentheses.

3.2.9.3 Identification of I.M.A.G.E. clones from the sequence databases

Seven non-redundant selected sequences identified the sequence tags of partially sequenced I.M.A.G.E. clones by BLAST searching of the GenBank database. The results of a BLAST search, in which query sequence 79 identified the 3' EST of I.M.A.G.E clone 486309 (GenBank accession AA043709), is illustrated in Fig 3.2.20.

```
BLAST - SEARCH DATABASE GenBank WITH FILE: 79-ptag3.seq
GB:AA043709 AA043709 zk50h11.s1 Soares pregnant uterus NbHPU Homo sapiens
cDNA clone 486309 3'. Length = 475
Identities = 243/251 (96%), Positives = 243/251 (96%), Strand = Minus /
Plus

251 TGAGAAGCCTTTAAGATGANNACAGTTGCACGAAGGTCCCTTTCATCAAGGTAGNGTATG 192
    |||||||||||||||||| |||||||||||||||||||||||||||||| |||||
79  TGAGAAGCCTTTAAGATGACTACAGTTGCACGAAGGTCCCTTTCATCAAGGTAGCGTATG 138

191 TACCCTAACAGTGTTTTAAAGGGTGGCCCAGAAAAACCCCATGTTACCTTATCACAATAT 132
    |||||||||||||| |||||| ||||||||||||||||||||||||||||||||||||
139 TACCCTAACAGTGTTCTAAAGGCTGGCCCAGAAAAACCCCATGTTACCTTATCACAATAT 198

131 GGAAAGCATTGTTTTNTTTTTCCACTAAATTAAATTATGGTGAAAAGTGCCACAGTTTTA 72
    |||||||||||| || ||||||||||||||||||||||||||||||||||||||||||
199 GGAAAGCATTGTCTTCTTTTTCCACTAAATTAAATTATGGTGAAAAGTGCCACAGTTTTA 258

71  TTTAGCATTATGGTACATAACAAACAGTTCTGTTTCAATTATGAAAAAAATTAATTAAAA 12
    |||||||||||||||||||||||||||||| ||||||||||||||||||||||||
259 TTTAGCATTATGGTACATAACAAACAGTTCTGTCTCAATTATGAAAAAAATTAATTAAAA 318

11  TAATCCTGAAA 1
    ||||||||||||
319 TAATCCTGAAA 329
```

Fig 3.2.20 Identification of the 3' EST of I.M.A.G.E. clone 486309 (GenBank accession A043709) by selected sequence 79-ptag3 by BLAST analysis.

The query sequence (upper) and the subject sequence (the sequence identified in the database)(lower) are aligned over the region of overlap. The minus strand of the query showed 96% identity to the plus strand of the subject over 251 nucleotides.

Details of the BLAST results are given in Table 3.2.3. These include the I.M.A.G.E. clone and GenBank accession numbers of sequences identified from the databases by the cyberprobes, and the degree of sequence identity to the query. Where the query sequence was represented by more than one identical clone from the selected sublibraries, the redundant clones and sequence numbers are given. Query sequences 9, 50 and 79 identified several I.M.A.G.E. clones from the databases. Up to 5 of these were included in the table. The nucleotide sequences of the queries are given in the Appendices.

Since the sequence databases are constantly expanding, BLAST searches of 'GenBank', 'Embl' and 'New Embl sequences' were repeated periodically with query sequences that had initially displayed no sequence identity to cDNAs in the databases. For instance, BLAST searches with sequences 50 and 52 failed to identify any sequences from the databases when first used as a query in 1995. Sequence 50 was found to be a single-copy fragment mapping to YAC 132D4 in this project (section 3.2.3). This sequence subsequently identified EST WI-15655 after its submission to GenBank on the 31st May, 1996. Radiation hybrid mapping at the Whitehead Institute localised WI-15655 between D15S128 and D15S156. Sequence 114 also failed to identify any sequences in 1996. Subsequently, sequence 114 identified *UBE3A* exon U4, after its submission to GenBank in January, 1997 (Kishino *et al.*, 1997), and U3b (Vu and Hoffman, 1997), after submission of the 5' UTR sequence on August 13, 1997 (Fig 3.2.19). Repeated BLAST searches with query sequences which had identified I.M.A.G.E. clones were also useful for obtaining additional clones for analysis. For example, clone 139994 (GenBank accession R63959), identified by query sequence 79, failed to grow at the HGMP Resource Centre, which distributes I.M.A.G.E. clones. Clones 562212 and 486309 were identified in subsequent BLAST searches in 1997 (Table 3.2.3), and obtained for analysis (section 3.3). The submission dates of the I.M.A.G.E. and EST sequences to GenBank, are given in Table 3.2.3.

There still remained 24 non-redundant sequences that had not been eliminated from the study, which did not identify any sequences in the databases up until January 31st, 1998. These represented novel sequences and were hybridised to YACs, PACs and BACs from 15q12, to identify those that were single-copy to the ASCR or PWSCR (section 3.3.2).

Query sequence	Clone co-ordinates	Insert size (bp)	BLAST search identities and other sequence information	Date of submission to GenBank
9	PAC-1β 7B B230E3-Aluα 5C	561	95% similarity to 121bp from 5' end of human clone 133446 (R27251) 81% identity over 283bp to human clone 143907 5' (R76744) 75% identity over 215bp to human clone 276665 5' (N46613) 96% identity over 121bp to human clone 683675 3' (AA215706) 76% identity over 185bp to Mus musculus clone 619122 5' (AA170644)	24/4/1995 6/6/1995 14/2/1996 13/8/1997 16/2/1997
50 52	132D4α 1C 132D4α 1G	301 488	98% identity over 223bp to human clone 283995 3' (N5337) 96% similarity over 254bp to human EST WI-15655 (G21562) 96% identity to nucleotides 1-299 human retina clone 360433 5' (AA015722) 92% similarity to nucleotides 241-423 human retina clone 360433 5' (above)	15/2/96 31/5/1996 29/11/1996
71	132D4β5D	338	95% identity over 306bp from 5' end of human clone 241885 (H93815) 81% similarity over 310bp to human retina cDNA (W26979)	4/12/1995 8/5/96
79	132D4-δ 1D 132D4ϕ 8D	352	97% identity over 153bp to human clone 139994 3' (R63959) 96% identity over 159bp to human muscle clone 562212 3' (AA211605) 98% identity over 153bp to pregnant uterus clone 486309 3' (AA043709)	26/5/1995 31/1/1997 10/5/1997
83	132D4δ 1D	376	90% similarity over 275bp to human clone 204357 5' (H59927) 92% identity over 65 bp to human clone 204357 3' (H59928)	6/10/1995 6/10/1995
93	132D4α 7C	356	87% identity over 62bp to human clone 125178 5' (R05682) and human cerebellum EST 26825 (AA324674)	3/4/1995 20/4/1997

Table 3.2.3 Identification of I.M.A.G.E. clones from GenBank by BLAST analysis of selected cDNA sequences.

The BLAST search results of novel selected sequences identified in the positive screen of the selected sublibraries are summarised. The selected sequence numbers used as queries for BLAST searches are given with their clone co-ordinates from the arrayed sublibraries. The clone co-ordinates of redundant query sequences are also listed. The insert sizes of the selected clones are also given. The percentage sequence identity and degree of overlap between the query sequences and I.M.A.G.E. clones or ESTs from the databases are listed. The GenBank accession number of the clones identified are in parentheses. The submission dates of the I.M.A.G.E. and EST sequences to GenBank appear in the final column.

3.2.10 UniGene analysis of I.M.A.G.E. clones identified by selected sequences

The UniGene Collection was searched to determine whether any of the I.M.A.G.E. clones identified by the selected sequences had been anchored to a UniGene cluster and assigned a chromosomal location. Searches of the *Homo sapiens* UniGene Collection at web site www.ncbi.nim.nih.gov/cgi-bin/UniGene/ were performed using the GenBank accession number of each I.M.A.G.E. clone (Table 3.2.3).

I.M.A.G.E. clone 241985 (GenBank: H93815) identified by query sequence 71 (132D4β 5D), belonged to the Chromosome 15 *IPW* UniGene cluster *Hs*.5022. However, sequence 71 had not identified *IPW* in the BLAST searches. Comparison of the sequences within the *IPW* UniGene collection, using the Clustal program, revealed that I.M.A.G.E. clone 241985 (GenBank H93815) overlapped and extended beyond the 3' end of the published *IPW* sequence (GenBank U12897). Selected sequence 71 was homologous to opposite end of clone 241985, and so did not overlap the *IPW* sequence (Fig 3.2.21). These sequences represented 'read-through' transcripts of *IPW*, possibly due to alternative polyadenylation signals within the 3' region (Rachel Wevrick, personal communication). This increased the frequency of *IPW* sequences within selected sublibrary 132D4 from 0.52% (Table 3.2.2) to 0.67%. Since I.M.A.G.E. clone 241985, and its identifying query sequence were found to represent an extended sequence of *IPW*, these clones were eliminated from the study.

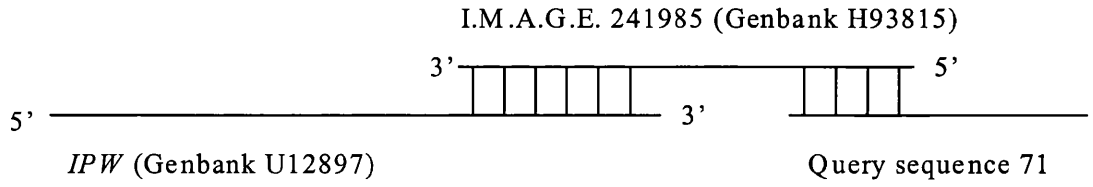


Fig 3.2.21 Overlap between *IPW*, I.M.A.G.E. clone 241985 and selected sequence 71.

Horizontal lines represent the sequences. Vertical lines show the region of overlap between the sequences. The final 530bp at the 3' end of clone 241985 are identical to nucleotides 923-1453 at the 3' end of *IPW* (U12897). The 5' end of clone 241985 extends beyond the 3' end of *IPW*. Sequence 71 identified nucleotides 75-381 at the 5' end of clone 241985, and extend beyond it by 75bp.

I.M.A.G.E. clone 360433 (GenBank accession AA015722) and EST WI-15655, identified by selected sequences 50 and 52 (Table 3.2.3), were anchored to Chromosome 15 UniGene cluster *Hs*.22543. This cluster was identified after clone 360433 (I.M.A.G.E. code number 828-P2) (Table 3.2.4), had been acquired and mapped to YAC 132D4 (section 3.2.12).

None of the I.M.A.G.E. clones identified through sequence homology by selected sequences 9, 79, 83 or 93 were assigned to any UniGene cluster. Therefore these sequences had not previously been previously located to 15q11-13 at the time of identification by BLAST analysis. The UniGene collection is continuously being updated, but repeated searches with these latter I.M.A.G.E. clones failed to identify any cluster prior to January 31st, 1998.

3.2.11 Obtainment of I.M.A.G.E. clones identified by the selected sequences

The I.M.A.G.E clones listed in Table 3.2.4 were obtained from the HGMP distribution centre in Cambridge, UK for analysis, to determine whether any of them represented transcripts involved in AS or PWS. Where several clones had been identified in the BLAST searches, 2-3 clones were requested for analysis. This was firstly because some of the clones could not be supplied due to contamination or failure to grow at the distribution centre, such as clone 13994 (GenBank accession R63959) (Table 3.2.3). Secondly, the sizes of the cloned inserts were variable and not recorded in the GenBank report. Where more than one I.M.A.G.E. clone identified by a single selected sequence was obtained, the clone with the largest insert was analysed.

3.2.11.1 Estimation of insert size and detection of contamination

The inserts from the I.M.A.G.E. clones were amplified using M13 vector primers from lysed cells, prior to colony purification, to estimate the insert size and ensure the clones were not contaminated. For clone 204357 (GenBank accession number H59928, I.M.A.G.E code number 395-H22) identified by selected sequence 83, two PCR products of 1.3 and 0.8kb were obtained, indicative of contamination (Fig 3.2.22A). Clone 395-H22 was colony-purified to separate the correct clone from the contaminating one. Both inserts were partially sequenced with M13 vector primers and compared to the sequence reported in GenBank H59928. The sequence from the 1.3kb insert was correct so this clone was retained for

analysis.

Table 3.2.4 summarises the information available for each I.M.A.G.E. clone obtained, including the GenBank accession and I.M.A.G.E. code numbers, the insert size, the cDNA library it was derived from, and the vector it was cloned in. The clones are referred to in the text by their clone numbers or I.M.A.G.E. code numbers.

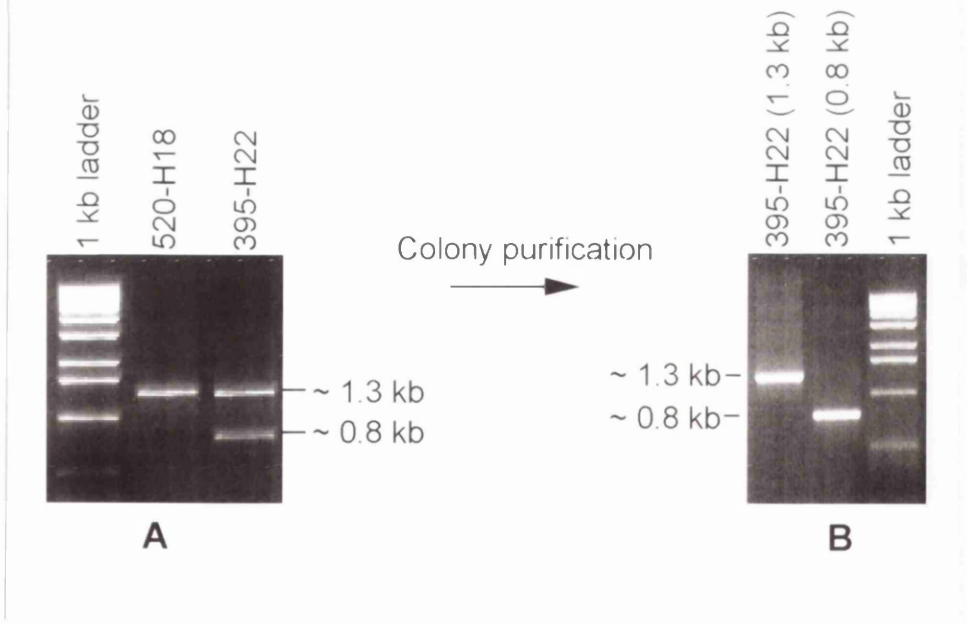


Fig 3.2.22 Colony purification of contaminated I.M.A.G.E. clone 204357 (395-H22)

A. PCR product of 204357 (395-H22), using primers M13U and R from bacterial spread
B. Amplified inserts of the two clones after colony purification.

Query sequence	GenBank accession	Clone number	I.M.A.G.E code	cDNA library	Vector and cDNA library construction	Insert size (kb)
9	R76744	143907	**215-B4**	Soares full-term placenta	Primed with *Not*I oligo-dT primer, ligated to *EcoR*I adaptors and cloned in pT7T3 via *Not*I and *EcoR*I	1.3
50	AA015722	360433	**828-P2**	Soares adult retina	Primed with *Not*I oligo-dT primer, ligated to *EcoR*I adaptors and cloned in pT7T3 via *Not*I and *EcoR*I	2.3
79	AA211605	562212	**1354-G13**	Stratagene adult muscle	Primed with oligo-dT then cloned unidirectionally in pBlueScript, via *EcoR*I	2
	AA043709	486309	**1156-L22**	Soares pregnant uterus	Primed with *Not*I oligo-dT primer, ligated to *EcoR*I adaptors and cloned in pT7T3 via *Not*I and *EcoR*I	0.9
83	H59928	204357	**395-H22**	Soares 20 week fetal liver-spleen	Primed with *Pac*I oligo-dT primer, ligated to *EcoR*I polylinkers and cloned in pT7T3 via *EcoR*I and *Pac*I	1.3
93	R05683	125178	**123-E19**	Soares 20 week fetal liver-spleen	Primed with *Pac*I oligo-dT primer, ligated to *EcoR*I polylinkers and cloned in pT7T3 via *EcoR*I and *Pac*I	1.5

Table 3.2.4 Description of the I.M.A.G.E clones supplied by the HGMP for characterisation.

The selected sequence that identified the I.M.A.G.E. clones are given in the first column. The I.M.A.G.E. clone and GenBank accession numbers are given. The I.M.A.G.E. code number that the sequences are to referred to by in the text are highlighted in bold. The cDNA librariy that each clone was derived from, and the method of construction of these libraries is given. Each clone was derived from a poly-dT primed library. However, where poly-dT primers were linked to a restriction enzyme, the poly-A tails are cleaved during library construction. The poly-A tails are not present in the insert. The insert size, on the basis of PCR amplification across the multiple cloning site, is listed for each clone.

3.2.12 Mapping of the I.M.A.G.E clones to the ASCR and PWSCR

To determine whether the I.M.A.G.E. clones 143907, 1354-G13, 360433, 204357 and 125178 (Table 3.2.4) mapped to the ASCR or PWSCR and were single-copy, the cloned inserts were hybridised to blots containing *EcoR*I digested YAC, PAC and BAC DNA from the 15q12 contig (Fig 3.1.7), and human genomic DNA. The boundary between the PWSCR and the ASCR was taken as marker D15S174, the proximal deletion breakpoint of Family S (Fig 1.9). I.M.A.G.E. clones mapping distal to D15S174, within the region spanned by the PAC and BAC contig, were located to the ASCR. Those mapping to YAC 132D4 alone, and therefore proximal to D15S174, were assigned to the PWSCR (Fig 3.1.7). Following the publication of a PAC contig extending across the PWSCR from D15S63 to D15S10 (Sutcliffe *et al.*, 1997), I.M.A.G.E. clones which had been located to the PWSCR could be more precisely mapped. *SNRPN, IPW, UBE3A* and marker D15S174 had been placed on this PAC contig. PAC 158H23 contained the 3' end of the *UBE3A* gene and PAC 14I12, which was common to the contig constructed in this project, contained the 5' end of *UBE3A*. This enabled the two contigs to be linked (Fig 3.2.28). PACs 5E12, 121D5, 134I14 and 158H23 from the contig published by *Sutcliffe* et al., were encompassed by YAC 132D4. These were obtained from the HGMP Resource Centre for fine-mapping the I.M.A.G.E clones previously located to YAC 132D4. The mapping data for each of the I.M.A.G.E. clones studied is summarised in Table 3.2.5. This includes the genomic clones and the sizes of the *EcoR*I fragments identified, and the region to which the cDNA clones were located. The mapping positions of the I.M.A.G.E. clones are shown diagrammatically in Fig 3.2.28. I.M.A.G.E.

I.M.A.G.E. clone 143907 (code number 215-B4), identified by selected sequence 9 from the PAC-1 sublibrary (Table 3.2.3), identified a single 3.2kb *EcoR*I band in YACs 132D4 and B230E3, PAC 50I2, and human genomic DNA (Fig 3.2.23). I.M.A.G.E. clone 562212 (code number 1354-G13) identified by selected sequence 79 from the 132D4 sublibrary (Table 3.2.3) also hybridised to a 3.2kb *EcoR*I fragment in YACs 132D4 and B230E3 (Fig 3.2.24) and PACs 50I2 and 158H23 (data not shown). These two cDNA clones may thus have identified a common *EcoR*I fragment. These data located clones 215-B4 and 1354-G13 to the ASCR, just downstream or overlapping the 3' coding exons of *UBE3A* (Fig 3.2.28). However, clone 1354-G13 also identified 3 additional distinct *EcoR*I fragments in the human DNA lane alone, indicative of related sequences in the human genome outside 15q12 (Fig 3.2.24).

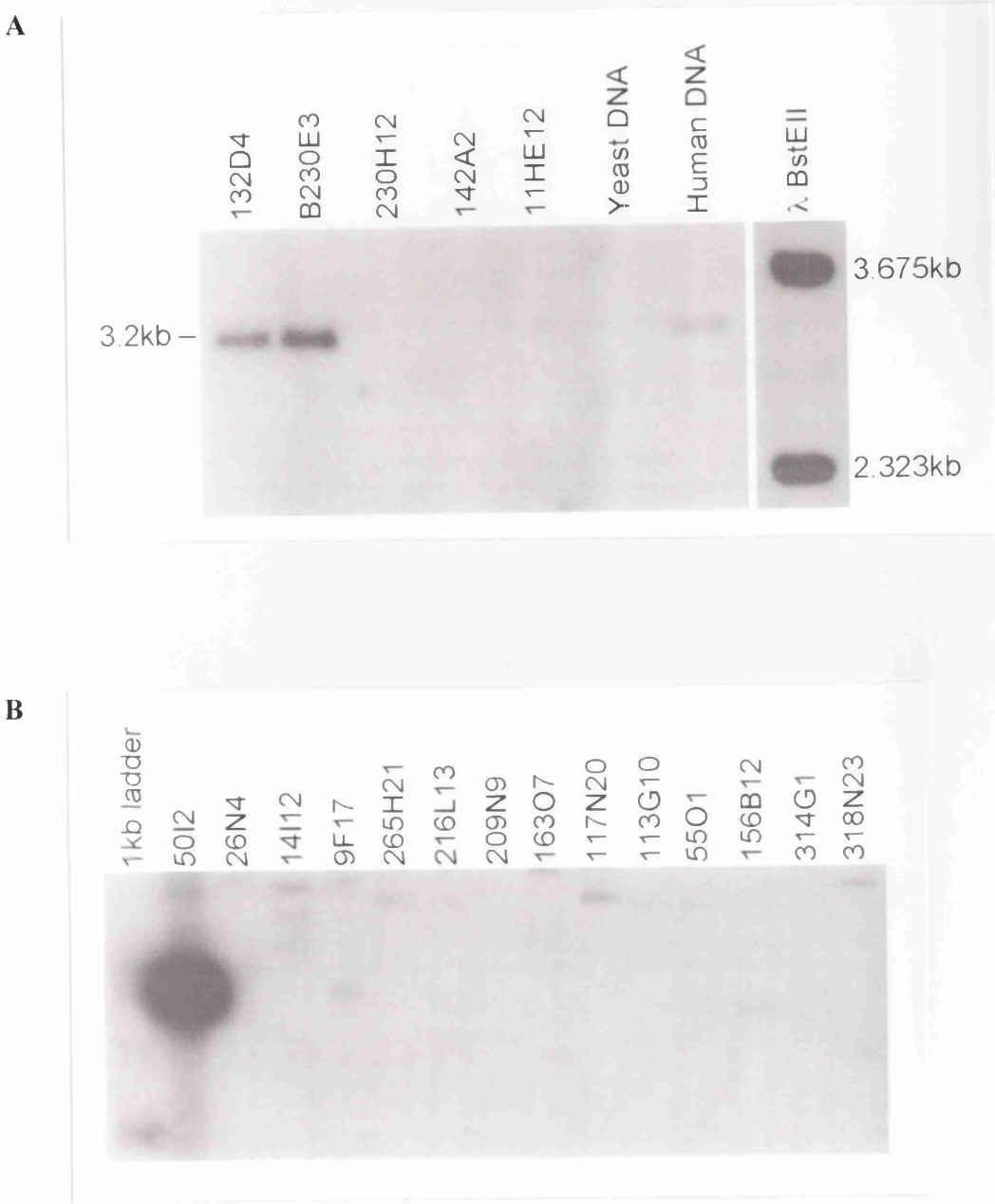


Fig 3.2.23 Mapping of I.M.A.G.E. clone 143507 (215-B4) to the ASCR

Autoradiographs of Southern blots hybridised with the insert from I.M.A.G.E. clone 143507 (215-B4) are shown. **A**: YACs from 15q12, yeast and human genomic DNA digested with *EcoR*I. 215-B4 identified a 3.2kb band in YACs 132D4 and B230E3 and the corresponding fragment in human genomic DNA, indicating it is single-copy to the ASCR. **B**: PAC DNA (as labelled) digested with *EcoR*I. 215-B4 identified a ~3.2kb fragment in PAC 50I2. The blots were washed to 1X SSC/ 0.1% SDS at 65°C for 20 minutes, and autoradiographed overnight.

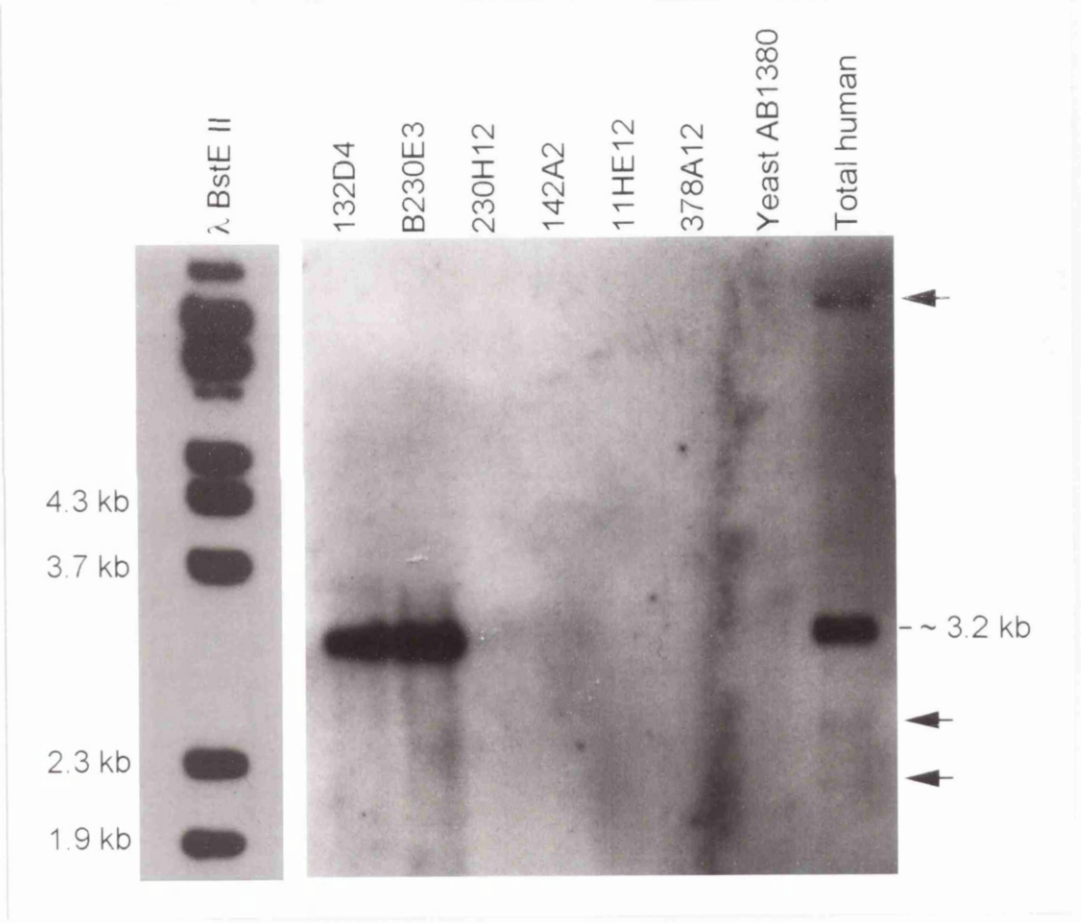


Fig 3.2.24 Mapping of I.M.A.G.E. clone 562212 (1354-G13) to YACs in the ASCR.

1354-G13 identified a 3.2kb *EcoR*I fragment in YACs 132D4 and B230E3 in the ASCR. Additional paler bands in genomic DNA (arrowed) was indicative of related sequences outside the 15q11-13 region. The filter was washed to 0.1X SSC/ 0.1% SDS at 65°C for 20 minutes, and exposed to X-ray film for 2 days.

I.M.A.G.E. clone 360433 (code number 828-P2), identified by query sequence 50 from the 132D4 sublibrary (Table 3.2.3) mapped to YAC 132D4, but was not single-copy to the PWSCR. This clone was highly repetitive, as seen by the smear in the human genomic DNA (Fig 3.2.25), even though selected sequence 50 was single-copy (data not shown). This clone was not analysed further, but was subsequently identified in the UniGene collection (section 3.2.10).

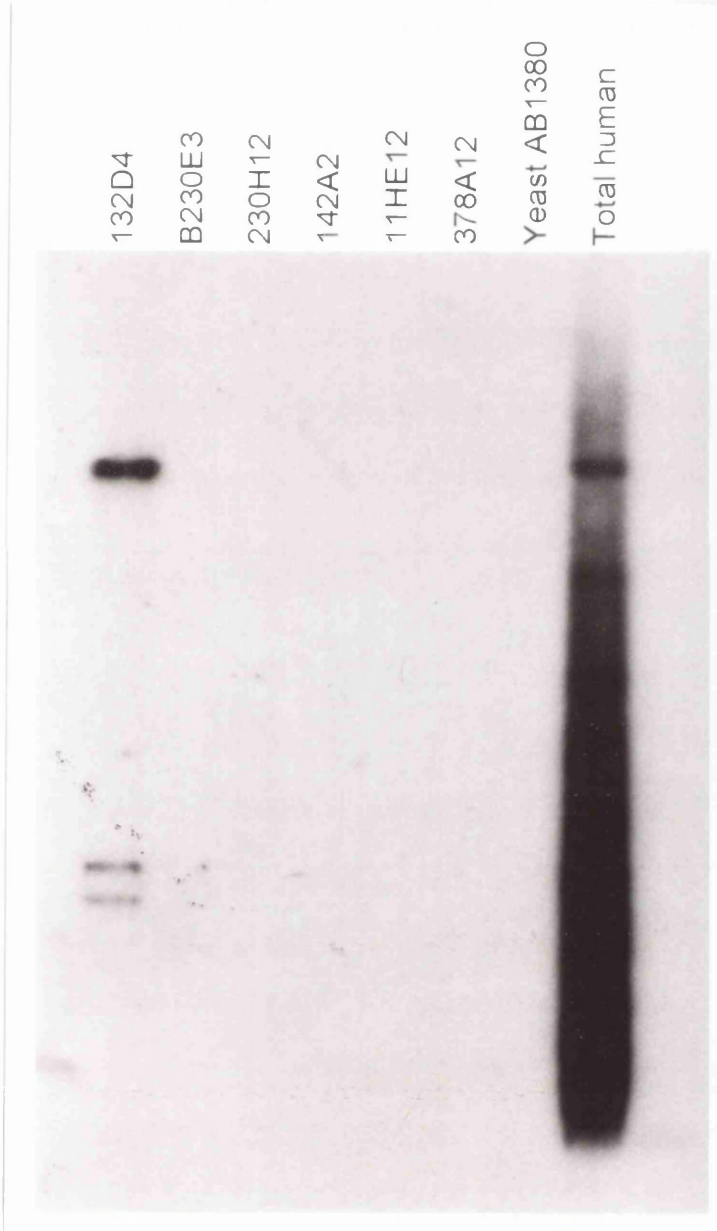


Fig 3.2.25 I.M.A.G.E. clone 360433 (828-P2) maps to YAC 132D4, but is repetitive.

828-P2 identified fragments in YAC 132D4 in the PWSCR, but the black smear in human genomic DNA is indicative of a highly repetitive sequence within the clone. The filter was washed to 0.1X SSC/ 0.1% SDS for 20 mins at 65°C, and exposed to X-ray film overnight.

I.M.A.G.E. clone 204357 (code number 395-H22), identified by selected sequence 83 from the 132D4 sublibrary (Table 3.2.3), hybridised to a single 14-15kb *EcoR*I fragment in 132D4 and human genomic DNA, showing that it is unique-copy to the PWSCR (Fig 3.2.26). 395-H22 was subsequently mapped to PAC 121D5 (Table 3.2.5).

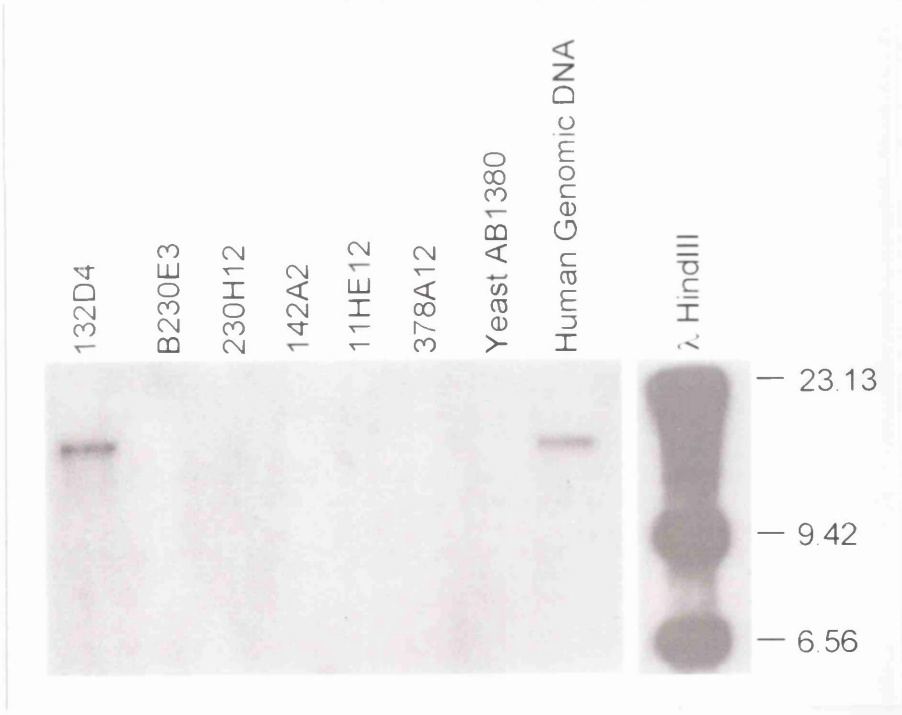


Fig 3.2.26 I.M.A.G.E. clone 204357 (395-H22) maps to YAC 132D4 in the PWSCR.

395-H22 was single-copy to YAC 132D4 in the PWSCR. The filter was washed to 1X SSC/ 0.1% SDS and autoradiographed overnight.

I.M.A.G.E. clone 125178 (code number 123-E19) identified by selected sequence 93 from the 132D4 sublibrary (Table 3.2.3) was also single-copy to YAC 132D4 (Fig 3.2.27) and PAC 121D5 from the PWSCR (Table 3.2.5). This cDNA clone identified 3 *EcoR*I fragments in the YAC and PAC DNA, indicating that it either contained 2 *EcoR*I recognition sites, or had crossed 2 *EcoR*I sites in genomic DNA.

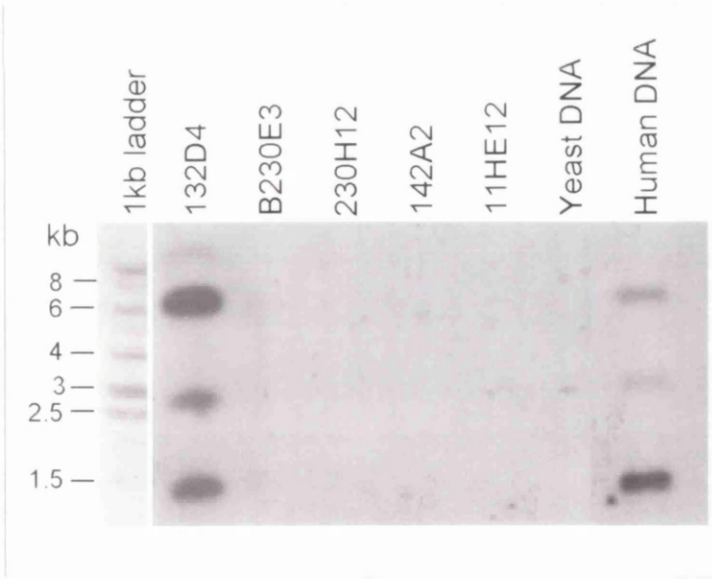


Fig 3.2.27 Mapping of I.M.A.G.E. clone 125178 (123-E19) to the PWSCR.

123-E19 identified was single-copy to the PWSCR, identifying 3 *EcoR*I fragments in YAC 132D4. The filter was washed to 1X SSC / 0.1% SDS at 65°C for 20 minutes and exposed overnight to Xray film.

I.M.A.G.E. clone	I.M.A.G.E. code	*EcoR*I fragment(s) identified (kb)	Genomic clone(s) identified	Region
143907	**215-B4**	~ 3.2	132D4, B230E3, 50I2	ASCR
562212	**1354-G13**	~ 3.2	132D4, B230E3, 50I2, 158H23	ASCR
360433	**828-P2**	Repetitive	132D4	PWSCR
204357	**395-H22**	14-15	132D4, 121D5	PWSCR
125178	**123-E19**	~7.8, ~2.7 and ~1.5	132D4, 121D5	PWSCR

Table 3.2.5 Summary of the mapping positions of the I.M.A.G.E. clones identified by selected sequences.

The *EcoR*I fragments and YACs and PACs identified by the I.M.A.G.E. clones are detailed. Location to the PWSCR or ASCR was assigned according to whether the clones mapped proximal or distal to D15S174 respectively.

With the exception of I.M.A.G.E. clone 828-P2, the clones listed in Table 3.2.5 represented novel single-copy or low-copy number sequences mapping to the AS or PWS regions. These were characterised further to determine if they represented transcripts potentially involved in the PWS or AS phenotypes. Characterisation of I.M.A.G.E. clones 395-H22 and 123-E19, which were single-copy to the PWSCR, is described in Chapter 5. Analysis of I.M.A.G.E. clones 215-B4 and 1354-G13 with respect to AS is described in section 3.3.

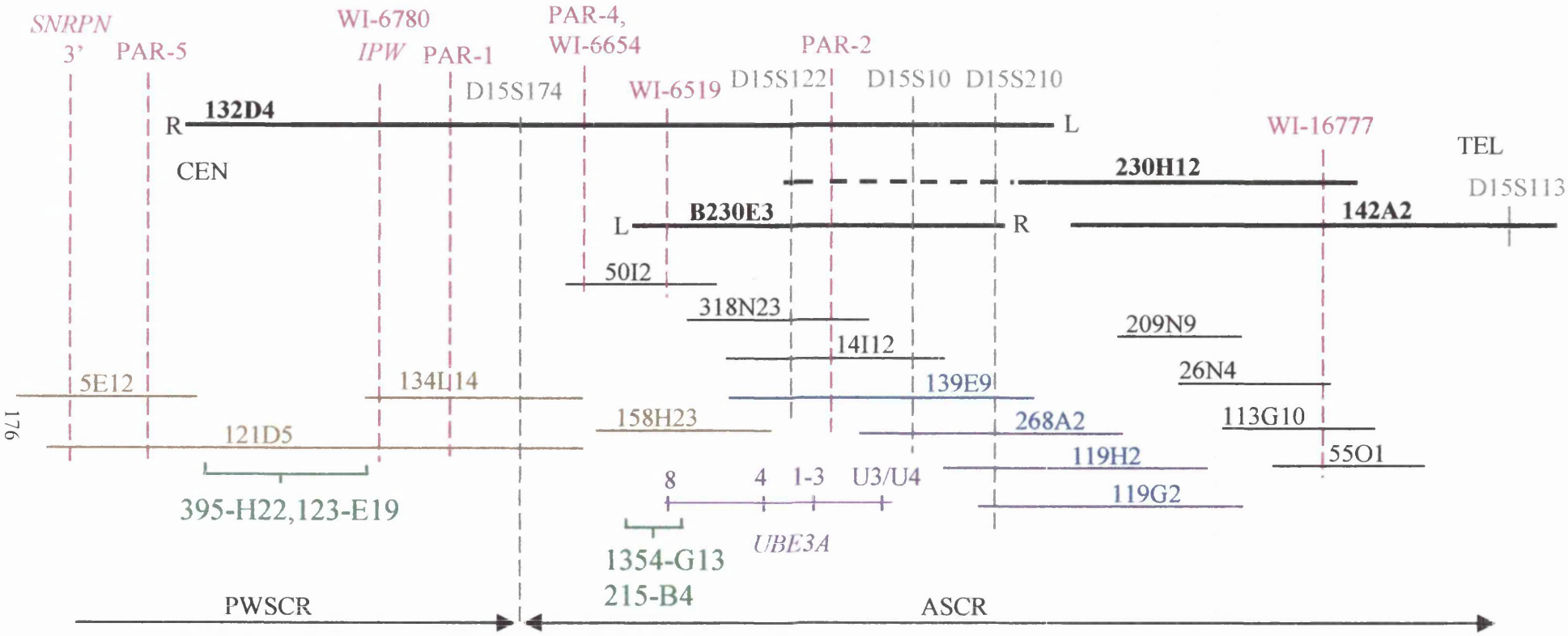


Fig 3.2.28 Map of I.M.A.G.E. clones within the ASCR and PWSCR.

The YAC, PAC and BAC map spanning the ASCR and distal portion of the PWSCR is shown, with the mapping locations of I.M.A.G.E. clones 1354-G13 and 215-B4 in the ASCR downstream or overlapping *UBE3A*, and 395-H22 and 123-19 in the PWSCR. D15S174 marks the boundary between the ASCR and the PWSCR. PACs coloured in brown were published by Sutcliffe *et al.*, (1997).

3.3 Analysis of cDNA clones from the AS candidate region.

3.3.1 I.M.A.G.E. clones 143907 (code 215-B4) and 562212 (code 1354-G13)

3.3.1.1 Sequence analysis

To characterise I.M.A.G.E. clones 143907 (I.M.A.G.E. code 215-B4) and 562212 (I.M.A.G.E. code 1354-G13), the nucleotide sequence for the full-length insert was determined. Firstly, the end sequences for each clone were obtained using M13 vector primers and compared to the GenBank 5' and 3' EST sequences for each clone to check that they corresponded. The sequence for the entire insert was obtained for both clones by 'oligomer-walking' with primers designed from the sequence tags. For instance, primer 'G13-S1' was designed from the sequence generated for 1354-G13 with M13U. 'G13-S2' was designed in turn, from the 'G13-S1' sequence tag, to meet the M13R sequence (Figs 3.3.1 and 3.3.2).

Since 215-B4 and 1354-G13 appeared to map to a common *EcoR*I fragment in genomic DNA (Table 3.2.5), the two sequences were compared by GAP to determine if there was any overlap. The 1.2kb 215-B4 sequence was internal to the 1.85kb 1354-G13 sequence, but cloned in the opposite orientation. Two non-redundant selected sequences, 9 and 79, had identified the ESTs from two different I.M.A.G.E. clones, which represented a common sequence (Fig 3.3.1). The 1.85kb sequence for the 1354-G13 insert is given in Fig 3.3.2.

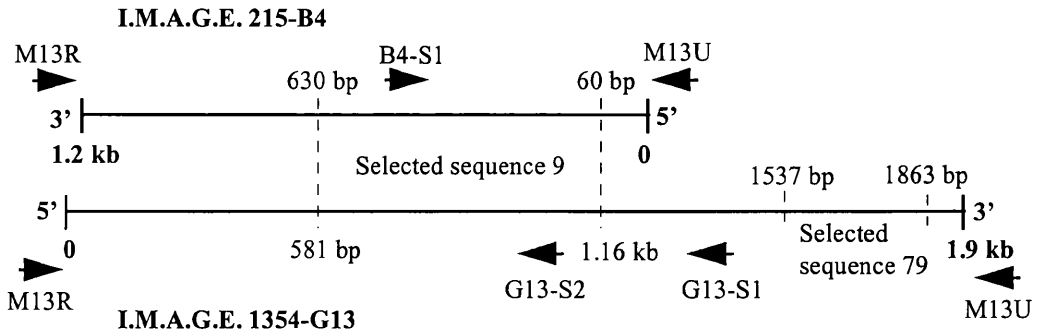


Fig 3.3.1 I.M.A.G.E. clones 125-B4 and 1354-G13 are derived from the same transcript.

Selected sequence 9, which identified the 5' EST of clone 215-B4 by BLAST, corresponded to nucleotides (n) 60-630 of 215-B4, and n581-1160 of clone 1354-G13. Selected sequence 79, which identified the 3' EST for 1354-G13, matched n1537-1863 of 1354-G13, but was not contained in 215-B4. Sequences 9 and 79 identified different segments of a single cDNA.

M13R

```
   1  AAAACAAAAA AAAGGAAAAA AAAAAAGAAA AAATTTAAAA AATTTTAAAA
  51  ATATAACGAG GGATAAATTT TTGGTGGTGA TAGTGTCCCA GTACAAAAAG
                      G13-5'
 101  GCTGTAAGAT AGTCAACCAC AGTAGTCACC TATGTCTGTG CCTCCCTTCT
 151  TTATTGGGGA CATGTGGGCT GGAACAGCAG ATTTCAGCTA CATATATGAA
 201  CAAATCCTTT ATTATTATTA TAATTATTTT TTTGCGTGAA AGTGTTACAT
 251  ATTCTTTCAC TTGTATGTAC AGAGAGGTTT TTCTGAATAT TTATTTTAAG
 301  GGTTAAATCA CTTTTGCTTG TGTTTATTAC TGCTTGAGGT TGAGCCTTTT
 351  GAGTATTTAA AAAATATATA CCAACAGAAC TACTCTCCCA AGGAAAATAT
 401  TGCCACCATT TGTAGACCAC GTAACCTTCA AGTATGTGCT ACTTTTTTGT
 451  CCCTGTATCT AACTCAAATC AGGAACAGTA TTTTTTTTAA TGATTTGCTT
 501  TTGAAACTTG AGTCTTGAAA ACAGTGTGAT GCAATTACTG CTGTTCTAGC
 551  CCCCAAAGAG TTTTCTGTGC AAAATCTTGA GAATCAATCA ATAAAGAAAG
 601  ATGGAAGGAA GGGAGAAATT GGAATGTTTT AACTGCAGCC CTCAGAACTT
 651  TAGTAACAGC ACAACAAATT AAAAACAAAA ACAACTCATG CCACAGTATG
 701  TCGTCTTCAT GTGTCTTGCA ATGAACTGTT TCAGTAGCCA ATCCTCTTTC
 751  TTAGTATATG AAAGGACAGG GATTTTTGTT CTTGTTGTTC TCGTTGTTGT
                      B4-S1
 801  TTTAAGTTTA CTGGGGAAAG TGCATTTGGC CAAATGAAAT GGTAGTCAAG
 851  CCTATTGCAA GAAAGTTAGG AAGTTTGTTG TTTGTTTATT ATAAACAAAA
 901  AGCATGTGAA AGTGCACTTA GGATGGGTTT TTATTAATTA CTTACTTATA
 951  CCTAGATTTT AAATAGACAA TCCAAAGTCT CCCCTTCGTG TTGCCATCAT
1001  CTTGTTGAAT CAGCCATTTT ATCGAGGCAC GTGATCAGTG TTGCAACATA
                      G13-S2
1051  ATGAAAAAGA TGGCTACTGT GCCTTGTGTT ACTTAATCAT ACAGTAAGCT
1101  GACCTGGAAA TGAATGAAAC TATTACTCCT AAGAATTACA TTGTATAGCC
1151  CCACAGATTA AATTTAATTA ATTAATTCAA AACATGTTAA ACGTTACTTT
1201  CATGTACTAT GGAAAAGTAC AAGTAGGTTT ACATTACTGA TTTCCAGAAG
1251  TAAGTAGTTT CCCCTTTCCT AGTCTTCTGT GTATGTGATG TTGTTAATTT
1301  CTTTTATTGC ATTATAAAAT AAAAGGATTA TGTATTTTTA ACTAAGGTGA
1351  GACATTGATA TATCCTTTTG CTACAAGCTA TAGCTAATGT GCTGAGCTTG
1401  TGCCTTGGTG ATTGATTGAT TGATTGACTG ATGTTTTAAC TGATTACTGT
1451  AGATCAACCT GATGATTTGT TTGTTTGAAA TTGGCAGGAA AAATGCAGCT
      G13-S1
```

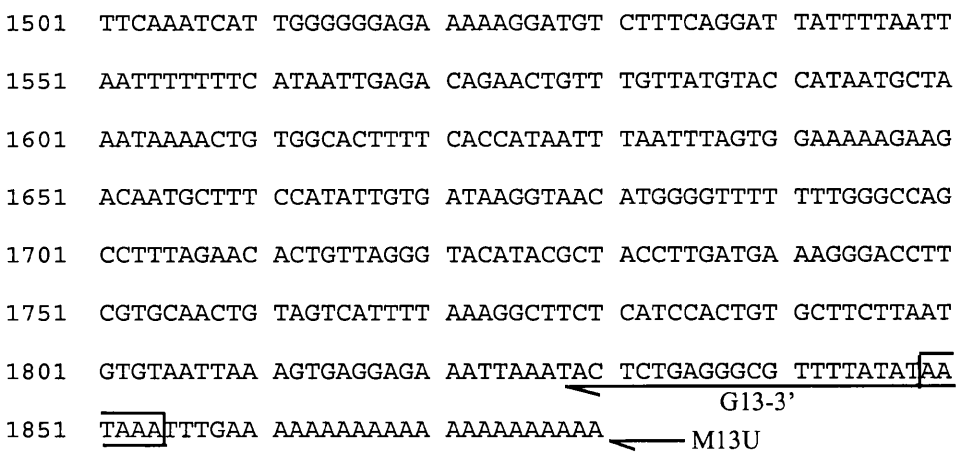


Fig 3.3.2 The nucleotide sequence for I.M.A.G.E. clone 1354-G13

The primer binding sites used for sequencing and PCR are underlined, with the arrow pointing 5' to 3' with respect to the primer sequence. The vector primers used for end-sequencing of the clone, are labelled at either end of the sequence. The consensus polyadenylation signals are boxed. The presence of a poly-A tail and polyadenylation signals indicates that 1354-G13 represents the 3' end of a polyA$^+$ mRNA sequence.

1354-G13 is polyadenylated at the 3' end, and three potential polyadenylation signal sites were identified using NIX analysis (Fig 3.3.2). The nucleotide sequence did not contain an open reading frame in any of the three potential reading frames (Fig 3.3.3). This indicated that 1354-G13 represents the 3' untranslated region (3' UTR) of a derivative transcript. The finding of different polyadenylation signals within 1354-G13 suggests that alternative polyadenylation signalling may occur in this transcript, with variable lengths of the 3'UTR in the mRNA. Since clone 1354-G13 contained the larger of the two inserts, all subsequent analyses were performed using this insert or sequence.

```
      1 AAAACAAAAAAAAGGAAAAAAAAAAAGAAAAAATTTAAAAAATTTTAAAAATATAACGAG
a       K  T  K  K  R  K  K  K  K  K  K  F  K  K  F  *  K  Y  N  E
b        K  Q  K  K  G  K  K  K  R  K  N  L  K  N  F  K  N  I  T  R
c         N  K  K  K  E  K  K  K  E  K  I  *  K  I  L  K  I  *  R  G

     61 GGATAAATTTTTGGTGGTGATAGTGTCCCAGTACAAAAAGGCTGTAAGATAGTCAACCAC
a       G  *  I  F  G  G  D  S  V  P  V  Q  K  G  C  K  I  V  N  H
b        D  K  F  L  V  V  I  V  S  Q  Y  K  K  A  V  R  *  S  T  T
c         I  N  F  W  W  *  *  C  P  S  T  K  R  L  *  D  S  Q  P  Q

    121 AGTAGTCACCTATGTCTGTGCCTCCCTTCTTTATTGGGGACATGTGGGCTGGAACAGCAG
a       S  S  H  L  C  L  C  L  P  S  L  L  G  T  C  G  L  E  Q  Q
b        V  V  T  Y  V  C  A  S  L  L  Y  W  G  H  V  G  W  N  S  R
c         *  S  P  M  S  V  P  P  F  F  I  G  D  M  W  A  G  T  A  D

    181 ATTTCAGCTACATATATGAACAAATCCTTTATTATTATTATAATTATTTTTTTGCGTGAA
a       I  S  A  T  Y  M  N  K  S  F  I  I  I  I  I  I  F  L  R  E
b        F  Q  L  H  I  *  T  N  P  L  L  L  L  *  L  F  F  C  V  K
c         F  S  Y  I  Y  E  Q  I  L  Y  Y  Y  Y  N  Y  F  F  A  *  K

    241 AGTGTTACATATTCTTTCACTTGTATGTACAGAGAGGTTTTTCTGAATATTTATTTTAAG
a       S  V  T  Y  S  F  T  C  M  Y  R  E  V  F  L  N  I  Y  F  K
b        V  L  H  I  L  S  L  V  C  T  E  R  F  F  *  I  F  I  L  R
c         C  Y  I  F  F  H  L  Y  V  Q  R  G  F  S  E  Y  L  F  *  G

    301 GGTTAAATCACTTTTGCTTGTGTTTATTACTGCTTGAGGTTGAGCCTTTTGAGTATTTAA
a       G  *  I  T  F  A  C  V  Y  Y  C  L  R  L  S  L  L  S  I  *
b        V  K  S  L  L  L  V  F  I  T  A  *  G  *  A  F  *  V  F  K
c         L  N  H  F  C  L  C  L  L  L  L  E  V  E  P  F  E  Y  L  K

    361 AAAATATATACCAACAGAACTACTCTCCCAAGGAAAATATTGCCACCATTTGTAGACC
a       K  I  Y  T  N  R  T  T  L  P  R  K  I  L  P  P  F  V  D  H
b        K  Y  I  P  T  E  L  L  S  Q  G  K  Y  C  H  H  L  *  T  T
c         N  I  Y  Q  Q  N  Y  S  P  K  E  N  I  A  T  I  C  R  P  R

    421 GTAACCTTCAAGTATGTGCTACTTTTTTGTCCCTGTATCTAACTCAAATCAGGAACAGTA
a       V  T  F  K  Y  V  L  L  F  C  P  C  I  *  L  K  S  G  T  V
b        *  P  S  S  M  C  Y  F  F  V  P  V  S  N  S  N  Q  E  Q  Y
c         N  L  Q  V  C  A  T  F  L  S  L  Y  L  T  Q  I  R  N  S  I

    481 TTTTTTTTAATGATTTGCTTTTGAAACTTGAGTCTTGAAAACAGTGTGATGCAATTACTG
a       F  F  L  M  I  C  F  *  N  L  S  L  E  N  S  V  M  Q  L  L
b        F  F  *  *  F  A  F  E  T  *  V  L  K  T  V  *  C  N  Y  C
c         F  F  N  D  L  L  L  K  L  E  S  *  K  Q  C  D  A  I  T  A

    541 CTGTTCTAGCCCCCAAAGAGTTTTCTGTGCAAAATCTTGAGAATCAATCAATAAAGAAAG
a       L  F  *  P  P  K  S  F  L  C  K  I  L  R  I  N  Q  *  R  K
b        C  S  S  P  Q  R  V  F  C  A  K  S  *  E  S  I  N  K  E  R
c         V  L  A  P  K  E  F  S  V  Q  N  L  E  N  Q  S  I  K  K  D

    601 ATGGAAGGAAGGGAGAAATTGGAATGTTTTAACTGCAGCCCTCAGAACTTTAGTAACAGC
a       M  E  G  R  E  K  L  E  C  F  N  C  S  P  Q  N  F  S  N  S
b        W  K  E  G  R  N  W  N  V  L  T  A  A  L  R  T  L  V  T  A
c         G  R  K  G  E  I  G  M  F  *  L  Q  P  S  E  L  *  *  Q  H

    661 ACAACAAATTAAAAACAAAAACAACTCATGCCACAGTATGTCGTCTTCATGTGTCTTGCA
a       T  T  N  *  K  Q  K  Q  L  M  P  Q  Y  V  V  F  M  C  L  A
b        Q  Q  I  K  N  K  N  N  S  C  H  S  M  S  S  S  C  V  L  Q
c         N  K  L  K  T  K  T  T  H  A  T  V  C  R  L  H  V  S  C  N
```

```
     721 ATGAACTGTTTCAGTAGCCAATCCTCTTTCTTAGTATATGAAAGGACAGGGATTTTTGTT
a        M  N  C  F  S  S  Q  S  S  F  L  V  Y  E  R  T  G  I  F  V
b         *  T  V  S  V  A  N  P  L  S  *  Y  M  K  G  Q  G  F  L  F
c          E  L  F  Q  *  P  I  L  F  L  S  I  *  K  D  R  D  F  C  S

     781 CTTGTTGTTCTCGTTGTTGTTTTAAGTTTACTGGGGAAAGTGCATTTGGCCAAATGAAAT
a        L  V  V  L  V  V  V  L  S  L  L  G  K  V  H  L  A  K  *  N
b         L  L  F  S  L  L  F  *  V  Y  W  G  K  C  I  W  P  N  E  M
c          C  C  S  R  C  C  F  K  F  T  G  E  S  A  F  G  Q  M  K  W

     841 GGTAGTCAAGCCTATTGCAAGAAAGTTAGGAAGTTTGTTGTTTGTTTATTATAAACAAAA
a        G  S  Q  A  Y  C  K  K  V  R  K  F  V  V  C  L  L  *  T  K
b         V  V  K  P  I  A  R  K  L  G  S  L  L  F  V  Y  Y  K  Q  K
c          *  S  S  L  L  Q  E  S  *  E  V  C  C  L  F  I  I  N  K  K

     901 AGCATGTGAAAGTGCACTTAGGATGGGTTTTTATTAATTACTTACTTATACCTAGATTTT
a        S  M  *  K  C  T  *  D  G  F  L  L  I  T  Y  L  Y  L  D  F
b         A  C  E  S  A  L  R  M  G  F  Y  *  L  L  T  Y  T  *  I  L
c          H  V  K  V  H  L  G  W  V  F  I  N  Y  L  L  I  P  R  F  *

     961 AAATAGACAATCCAAAGTCTCCCCTTCGTGTTGCCATCATCTTGTTGAATCAGCCATTTT
a        K  *  T  I  Q  S  L  P  F  V  L  P  S  S  C  *  I  S  H  F
b         N  R  Q  S  K  V  S  P  S  C  C  H  H  L  V  E  S  A  I  L
c          I  D  N  P  K  S  P  L  R  V  A  I  I  L  L  N  Q  P  F  Y

    1021 ATCGAGGCACGTGATCAGTGTTGCAACATAATGAAAAAGATGGCTACTGTGCCTTGTGTT
a        I  E  A  R  D  Q  C  C  N  I  M  K  K  M  A  T  V  P  C  V
b         S  R  H  V  I  S  V  A  T  *  *  K  R  W  L  L  C  L  V  L
c          R  G  T  *  S  V  L  Q  H  N  E  K  D  G  Y  C  A  L  C  Y

    1081 ACTTAATCATACAGTAAGCTGACCTGGAAATGAATGAAACTATTACTCCTAAGAATTACA
a        T  *  S  Y  S  K  L  T  W  K  *  M  K  L  L  L  L  R  I  T
b         L  N  H  T  V  S  *  P  G  N  E  *  N  Y  Y  S  *  E  L  H
c          L  I  I  Q  *  A  D  L  E  M  N  E  T  I  T  P  K  N  Y  I

    1141 TTGTATAGCCCCACAGATTAAATTTAATTAATTAATTCAAAACATGTTAAACGTTACTTT
a        L  Y  S  P  T  D  *  I  *  L  I  N  S  K  H  V  K  R  Y  F
b         C  I  A  P  Q  I  K  F  N  *  L  I  Q  N  M  L  N  V  T  F
c          V  *  P  H  R  L  N  L  I  N  *  F  K  T  C  *  T  L  L  S

    1201 CATGTACTATGGAAAAGTACAAGTAGGTTTACATTACTGATTTCCAGAAGTAAGTAGTTT
a        H  V  L  W  K  S  T  S  R  F  T  L  L  I  S  R  S  K  *  F
b         M  Y  Y  G  K  V  Q  V  G  L  H  Y  *  F  P  E  V  S  S  F
c          C  T  M  E  K  Y  K  *  V  Y  I  T  D  F  Q  K  *  V  V  S

    1261 CCCCTTTCCTAGTCTTCTGTGTATGTGATGTTGTTAATTTCTTTTATTGCATTATAAAAT
a        P  L  S  *  S  S  V  Y  V  M  L  L  I  S  F  I  A  L  *  N
b         P  F  P  S  L  L  C  M  *  C  C  *  F  L  L  L  H  Y  K  I
c          P  F  L  V  F  C  V  C  D  V  V  N  F  F  Y  C  I  I  K  *

    1321 AAAAGGATTATGTATTTTTAACTAAGGTGAGACATTGATATATCCTTTTGCTACAAGCTA
a        K  R  I  M  Y  F  *  L  R  *  D  I  D  I  S  F  C  Y  K  L
b         K  G  L  C  I  F  N  *  G  E  T  L  I  Y  P  F  A  T  S  Y
c          K  D  Y  V  F  L  T  K  V  R  H  *  Y  I  L  L  L  Q  A  I

    1381 TAGCTAATGTGCTGAGCTTGTGCCTTGGTGATTGATTGATTGATTGACTGATGTTTTAAC
a        *  L  M  C  *  A  C  A  L  V  I  D  *  L  I  D  *  C  F  N
b         S  *  C  A  E  L  V  P  W  *  L  I  D  *  L  T  D  V  L  T
c          A  N  V  L  S  L  C  L  G  D  *  L  I  D  *  L  M  F  *  L
```

```
    1441 TGATTACTGTAGATCAACCTGATGATTTGTTTGTTTGAAATTGGCAGGAAAAATGCAGCT
a         *  L  L  *  I  N  L  M  I  C  L  F  E  I  G  R  K  N  A  A
b          D  Y  C  R  S  T  *  *  F  V  C  L  K  L  A  G  K  M  Q  L
c          I  T  V  D  Q  P  D  D  L  F  V  *  N  W  Q  E  K  C  S  F

    1501 TTCAAATCATTGGGGGGAGAAAAAGGATGTCTTTCAGGATTATTTTAATTAATTTTTTTC
a         F  K  S  L  G  G  E  K  G  C  L  S  G  L  F  *  L  I  F  F
b          S  N  H  W  G  E  K  K  D  V  F  Q  D  Y  F  N  *  F  F  S
c          Q  I  I  G  G  R  K  R  M  S  F  R  I  I  L  I  N  F  F  H

    1561 ATAATTGAGACAGAACTGTTTGTTATGTACCATAATGCTAAATAAAACTGTGGCACTTTT
a         I  I  E  T  E  L  F  V  M  Y  H  N  A  K  *  N  C  G  T  F
b          *  L  R  Q  N  C  L  L  C  T  I  M  L  N  K  T  V  A  L  F
c          N  *  D  R  T  V  C  Y  V  P  *  C  *  I  K  L  W  H  F  S

    1621 CACCATAATTTAATTTAGTGGAAAAAGAAGACAATGCTTTCCATATTGTGATAAGGTAAC
a         H  H  N  L  I  *  W  K  K  K  T  M  L  S  I  L  *  *  G  N
b          T  I  I  *  F  S  G  K  R  R  Q  C  F  P  Y  C  D  K  V  T
c          P  *  F  N  L  V  E  K  E  D  N  A  F  H  I  V  I  R  *  H

    1681 ATGGGGTTTTTTTGGGCCAGCCTTTAGAACACTGTTAGGGTACATACGCTACCTTGATGA
a         M  G  F  F  W  A  S  L  *  N  T  V  R  V  H  T  L  P  *  *
b          W  G  F  F  G  P  A  F  R  T  L  L  G  Y  I  R  Y  L  D  E
c          G  V  F  L  G  Q  P  L  E  H  C  *  G  T  Y  A  T  L  M  K

    1741 AAGGGACCTTCGTGCAACTGTAGTCATTTTAAAGGCTTCTCATCCACTGTGCTTCTTAAT
a         K  G  P  S  C  N  C  S  H  F  K  G  F  S  S  T  V  L  L  N
b          R  D  L  R  A  T  V  V  I  L  K  A  S  H  P  L  C  F  L  M
c          G  T  F  V  Q  L  *  S  F  *  R  L  L  I  H  C  A  S  *  C

    1801 GTGTAATTAAAGTGAGGAGAAATTAAATACTCTGAGGGCGTTTTATATAATAAATTTGAA
a         V  *  L  K  *  G  E  I  K  Y  S  E  G  V  L  Y  N  K  F  E
b          C  N  *  S  E  E  K  L  N  T  L  R  A  F  Y  I  I  N  L  K
c          V  I  K  V  R  R  N  *  I  L  *  G  R  F  I  *  *  I  *  K

    1861 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
a         K  K  K  K  K  K  K  K  K  K  K  K  K  K  K
b          K  K  K  K  K  K  K  K  K  K  K  K  K  K
c          K  K  K  K  K  K  K  K  K  K  K  K  K  K  K
```

Fig 3.3.3 MAP prediction that 1354-G13 is an untranslated sequence

The amino acid sequences for the three possible reading frames a, b, and c, as predicted by the MAP program, are given beneath the DNA sequence. Stop codons, denoted by stars, are present in each frame indicating that this sequence is untranslated.

3.3.1.2 Analysis of expression of the 1354-G13 sequence

3.3.1.2.1 Northern Analysis

Northern analysis was performed for 1354-G13 using fetal brain RNA to determine whether the sequence was expressed in this tissue, and if so, the size of the full-length transcript. The 1354-G13 insert identified a faint 5.2kb transcript in total RNA from 15 week fetal brain (Fig 3.3.4A). However, the probe also hybridised to the ribosomal RNA, so northern analysis was repeated using polyA+ RNA from 19 week gestation brain. This confirmed that 1354-G13 hybridised to a 5.2kb transcript, and that the level of expression of 1354-G13 was low compared to the β-actin control (Fig 3.3.4B). The size of the transcript identified was notably similar to that previously reported for the coding sequence of *UBE3A*. A ~5kb transcript was identified in primary foreskin keratinocytes (Huibregtse *et al.*, 1993) and skeletal muscle (Nakao *et al.*, 1994). The ~5kb hybridising species was shown to be predominant in heart and skeletal muscle, though present at lower levels in other tissues (Sutcliffe *et al.*, 1997).

3.3.1.2.2 RT-PCR analysis of 1354-G13

The distribution of 1354-G13 expression was examined by RT-PCR in a multitude of 15 week gestation human fetal tissues and normal fibroblast cell lines, using primers 'G13-5' and G13-3' (Fig 3.3.2). The 1.8kb RT-PCR product was detected in all fetal tissues examined, including brain, eye, heart, intestine, kidney, limb, liver, lung, placenta, skin and stomach (Fig 3.3.5A) and fibroblasts (Fig 3.3.6). Although northern analyses had indicated low levels of expression in brain and embryo samples, 1354-G13 was shown by RT-PCR to be widely expressed. Reactions in which no reverse transcriptase (RT) had been added were included for each RNA sample as controls for the detection of genomic contamination. No contamination was observed in the RT-minus reactions for 1354-G13 or the housekeeping control gene *HPRT* (3.3.5B).

1354-G13 was also amplified from genomic DNA using primers 'G13-5' and 'G13-3', which gave the same sized (1.8kb) product as the cDNA. This indicated that the cDNA sequence is contiguous with genomic DNA, that 1354-G13 is contained within a single exon (Fig 3.3.6).

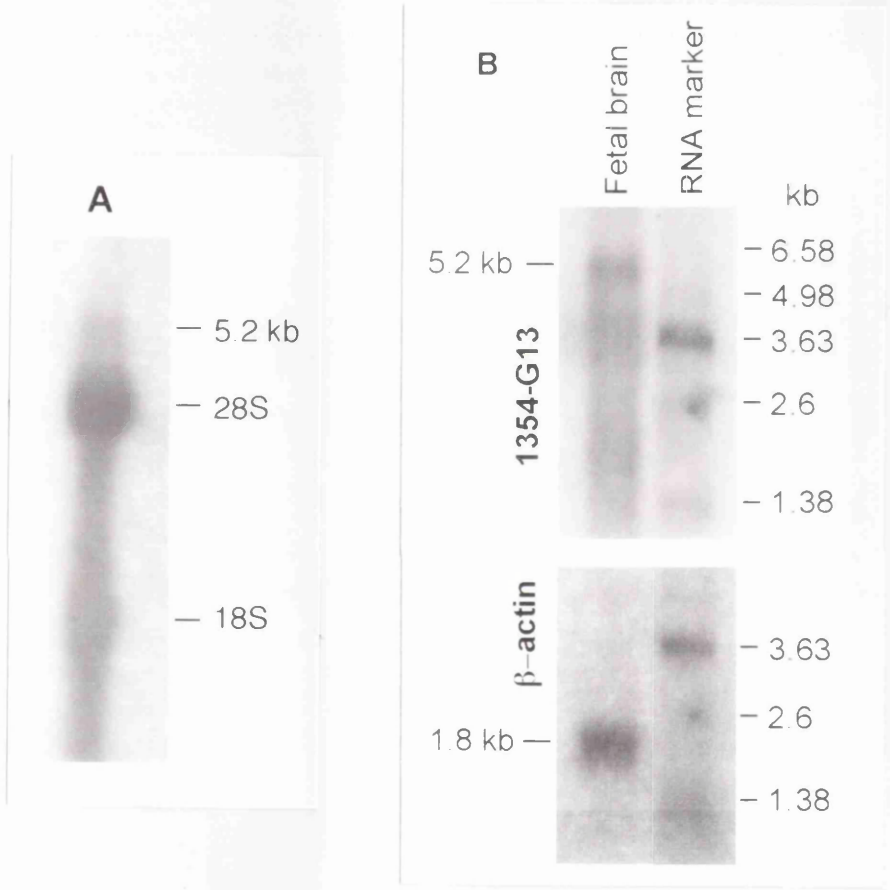


Fig 3.3.4 Northern analysis of 1354-G13

A: Northern blot containing 50μg total RNA from 15 week gestation fetal brain, probed with the 1354-G13 insert. The blot was washed to 1XSSC/0.1%SDS at 42°C and exposed to X-ray film for 3 days. The transcript was sized at 5.2kb by measuring the migration distance of the band identified with reference to the RNA marker and 28S and 18S ribosomal bands.
B: Northern blot containing approximately 2μg polyA+ RNA from 19 week human fetal brain hybridised with 1354-G13 (top) and β-actin control (bottom). The blot was washed to 1X SSC/ 0.1% SDS at 42°C after hybridisation. 1354-G13 was autoradiographed for 1 week, and β-actin 3 days.

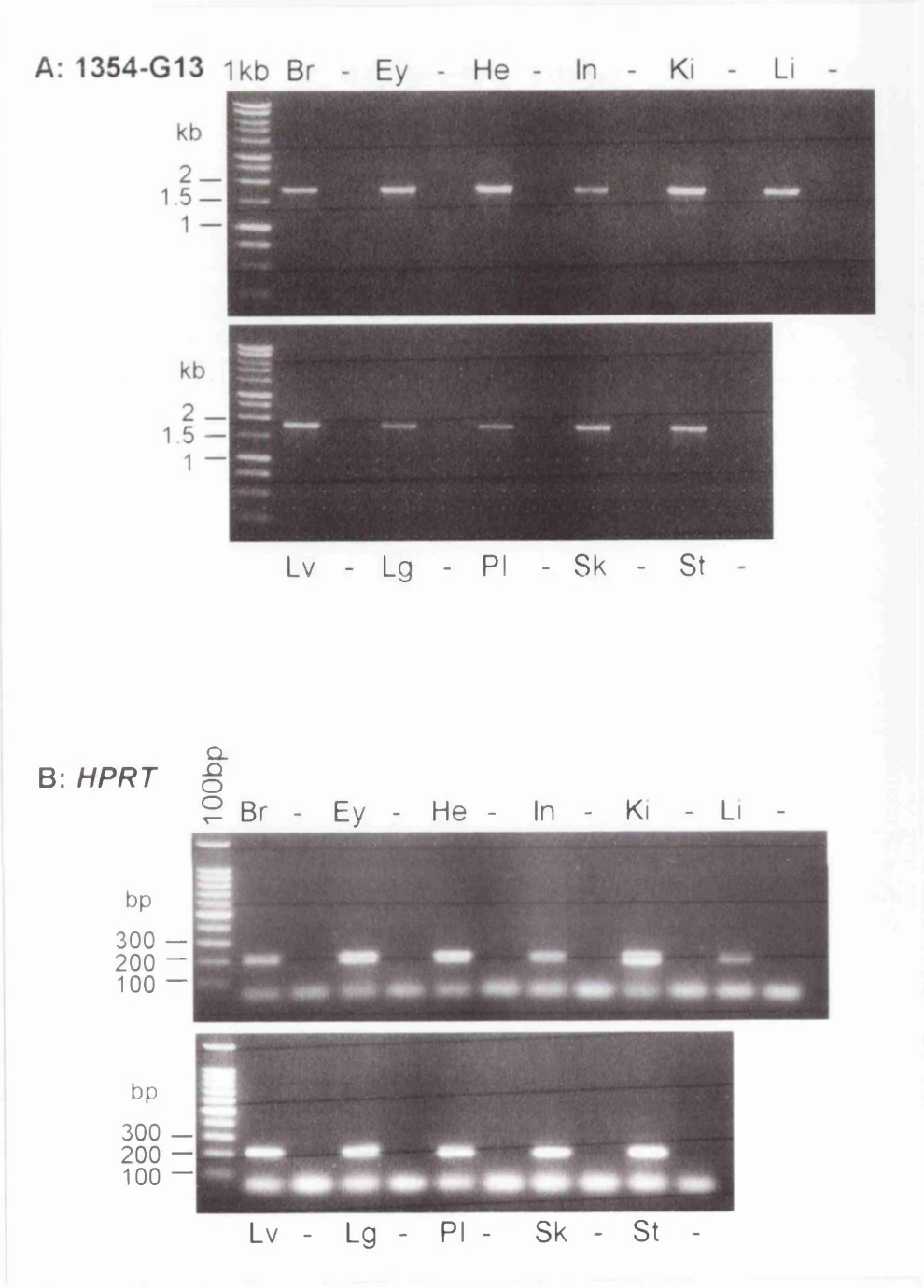


Fig 3.3.5 Analysis of expression of 1354-G13 in fetal tissues by RT-PCR.

RT-PCR analysis was performed on 2µg polyT primed, RNase treated total RNA from 15 week fetal samples; brain (Br), eye (Ey), heart (He), intestine (In), kidney (Ki), limb (Li), liver (Lv), lung (Lg), placenta (Pl), skin (sk) and stomach (St). RT minus (-) controls were included to detect genomic contamination of the cDNA. PCR was performed for 35 cycles. **A**: 1354-G13 amplified with primers G13-5' and G13-3' gave the expected 1.8kb product. **B**: *HPRT*. Amplification of genomic DNA in the *HPRT* reactions is distinguishable from the 220bp cDNA product as a 1.5kb band.

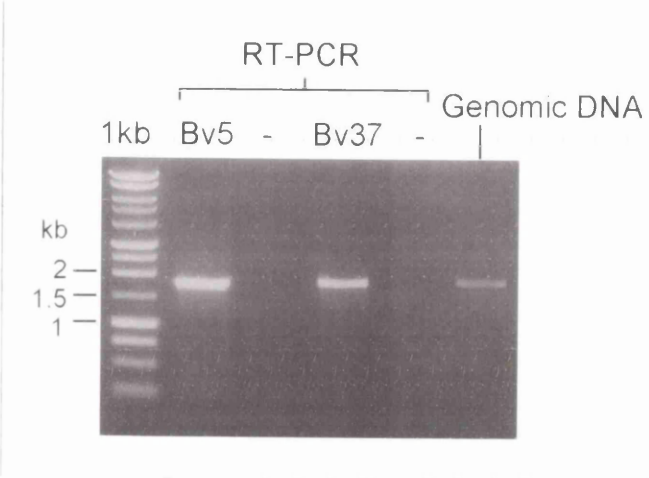


Fig 3.3.6 Detection of expression of 1354-G13 in adult fibroblast cell lines

The 1.8kb cDNA product was observed in two normal adult fibroblast cell lines, Bv5 and Bv37 (supplied by Dr Jane Hill). No genomic contamination was observed in the RT minus controls. RT-PCR was performed for 35 cycles with annealing at 56°C. Amplification from 100ng genomic DNA gave a 1.8kb product, indicating that 1354-G13 does not contain any intron/exon boundaries.

The Bv5 fibroblast and fetal brain RT-PCR products were sequenced with G13-5', G13-S1, G13-S2 and G13-3' to verify the cloned sequence and to confirm the correct cDNA product had been amplified. The Bv5 sequence did not differ from the sequence given in Fig 3.3.2. A homozygous T→C change was identified at position 1769 in the fetal brain sample (data not shown). This most likely represents an expressed sequence polymorphism, which if present at a relatively high frequency in the population, could be used to study the allelic origin of expression of this cDNA. The observation of a homozygous T at this site could be indicative of monoallelic expression, but the corresponding genomic DNA sample was not analysed to determine if it was homozygous or heterozygous for the sequence change.

3.3.1.3 Test for allele-specific methylation associated with 1354-G13

Clone 1354-G13 was tested for association with parental-specific methylation, to determine if it was likely to be imprinted. Southern analysis was performed with the 1354-G13 insert using DNA from PWS and AS patients with UPD 15 or classical deletions of 15q11-13 (who have only a maternal or paternal 15q11-13 allele respectively) with the methylation sensitive enzyme *Hpa*II. This method was first described by Driscoll *et al.* (1992), and is described in detail in section 1.4.5.1.

No differences in banding pattern were detected between the *Hpa*II digested PWS and AS samples, showing that 1354-G13 does not detect any differentially methylated *Hpa*II sites in lymphocytes (Fig 3.3.7). However, allele specific methylation of imprinted genes usually occurs in the promoter region of the gene and sometimes within intronic sequences (reviewed in section 1.7.6). Differential methylation has not been associated with the very 3' end of imprinted genes. Since 1354-G13 represents the 3' UTR of a gene, the finding of similar methylation patterns in both parental alleles does not preclude this gene from being imprinted.

For this experiment, the *Hpa*II enzyme was used in combination with the methylation insensitive enzyme *Hind*III, to reduce the fragment size for electrophoresis. A single-digested *Hind*III lane was included on the blot to estimate the size of the derivative *Hind*III fragment identified by 1354-G13. However, 1354-G13 hybridised to two *Hind*III fragments (Fig 3.3.7), despite containing no *Hind*III reccognition site within the sequence. 1354-G13 also hybridised to additional *EcoR*I fragments in human genomic DNA that were outside the ASCR (Fig 3.2.24). This provided further evidence for the presence of a similar sequence in genomic DNA.

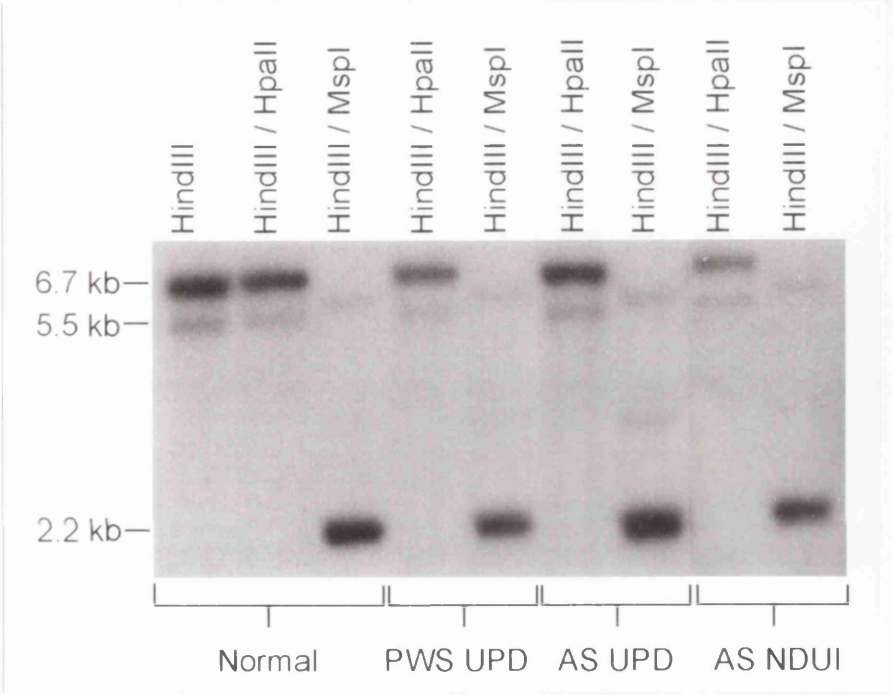


Fig 3.3.7 1354-G13 is not associated with differential methylation of the maternal and paternal 15q11-13 alleles.

Leucocyte DNA from normal, PWS mUPD15, AS pUPD15 and a non-deletion, disomy or imprintor (NDUI) AS patient, digested with *Hpa*II or *Msp*I combined with *Hind*III, was Southern blotted and probed with 1354-G13. The normal DNA sample was additionally digested with *Hind*III alone to provide the size of the derivative *Hind*III fragment(s) detected by the probes. 1354-G13 identified a prominant 6.7kb and a paler 5.5kb, in *Hind*III digested DNA. The same sized bands were detected in *Hind*III/ *Hpa*II digested DNA in all DNA samples. Digestion of the derivative 6.7kb *Hind*III fragment with the methylation insensitive isoschizomer *Msp*I, gave a 5.8kb fragment indicating the presence of methylated *Hpa*II sites within the *Hind*III fragment identified by 1354-G13. Since the *Hind*III / *Hpa*II digested DNA yielded the same sized band as the derivative *Hind*III fragment, all these *Hpa*II sites are methylated on both parental alleles.

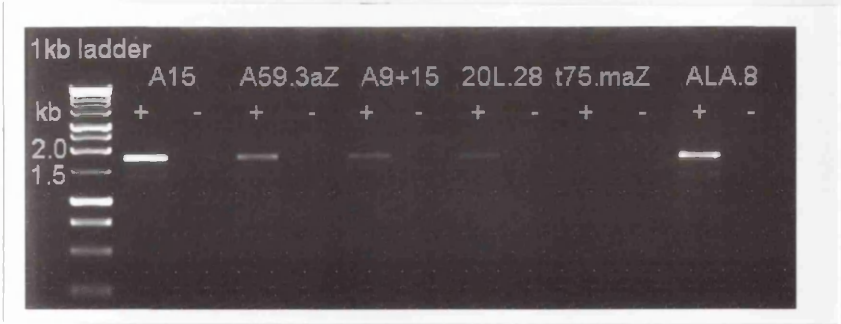
3.3.1.4 In vitro imprinting assay using human monochromosomal 15 x rodent hybrids

To investigate whether 1354-G13 was imprinted, an *in vitro* imprinting assay was performed using human monochromosomal 15 x rodent somatic cell hybrids, containing a chromosome 15 of either maternal or paternal origin (as described in section 1.8.2.3). Hybrid cell lines A15, A9+15 and A59.3aZ contain a paternal 15, and clones 20L.28, ALA-8 and t75.2maZ contain a maternal 15. The human cells used in creating the monochromosomal 15 hybrids were derived from fibroblasts (Gabriel *et al.*, 1998). Since 1354-G13 had been shown to be expressed in fibroblasts (Fig 3.3.6), imprinting assays by RT-PCR of these cell lines were feasible for this transcript. RT-PCR was performed using the cDNAs from each hybrid cell line as template, with the corresponding RT minus control reactions used to detect genomic DNA contamination. Expression of 1354-G13 was detected in all cell lines with the exception of t75.2maZ (Fig 3.3.8A), as was the case for the biallelically expressed control, *UBE3A* (Fig 3.3.8B; see Control reactions). This indicated that 1354-G13 was expressed from both parental alleles in fibroblast-derived cells, so was not imprinted in these cell lines.

Control reactions

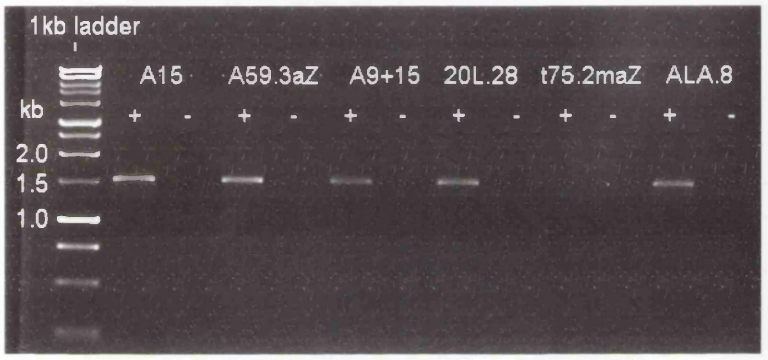
RT-PCR reactions were performed across exons 4-8 (renumbered 10-14) of *UBE3A* as a biallelically expressed control (Fig 3.3.8B), and between exons 2 and 3 of *SNRPN*, as a maternally imprinted control (Fig 3.3.8C), from 15q11-13. *UBE3A* expression was observed in hybrids containing both a maternal and paternal 15 with the exception of t75.2maZ (Fig 3.3.8B). This was most likely due to poor RNA integrity or lack of expression of genes within 15q11-13 in this cell line. Thus absence of an amplification product for this clone in subsequent tests was not considered to be due to any imprinting effect.

SNRPN expression was detected in all three hybrids containing the paternal 15, but not in t75.2maZ or ALA-8 which contained a maternal 15, as expected. However, a weak RT-PCR product was also observed for the 20L.28 hybrid, containing a maternal 15 (Fig 3.3.8C). This may have been due to contamination of the cDNAs with paternally derived chromosome 15 cDNAs, or due to the presence of a supernumerary marker 15 of paternal origin (Gabriel *et al.*, 1998). Alternatively, loss of imprinting may have occurred in this particular hybrid clone due to loss of the methylation imprint, or passage of the hybrid cell line beyond the number of population-doublings for which the imprinting status is maintained. Therefore, only results from hybrid ALA-8 were reliable for the demonstration of expression from the maternal 15.



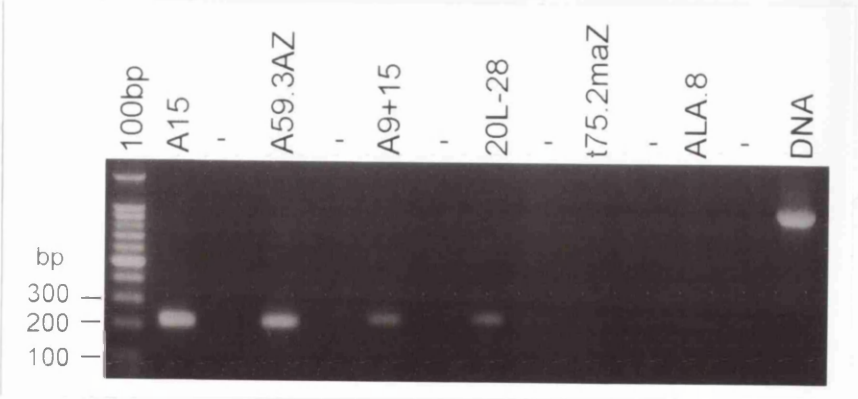
A 1354-G13 is biallelically expressed

1354-G13 was amplified using primers G13-5' and G13-3'. 1354-G13 expression was detected in all cell lines excluding t75.2maZ, as observed for the biallelic control, *UBE3A*.



B *UBE3A*; biallelically expressed control

UBE3A was amplified from previous exons 4-8 (10-14) using primers *E6-AP* 4F and *E6-AP* 8R to give a 1.5kb cDNA fragment. Expression was detected in all cell lines except t75.2maZ.



C *SNRPN*; paternally expressed imprinted control

SNRPN primers N1 and N4 from exons 2 and 3 respectively were used to amplify a ~200bp cDNA fragment. A genomic DNA control reaction was included in the experiment to provide a size reference for any genomic contamination present in the RT minus reactions.

Fig 3.3.8 *In vitro* RT-PCR imprinting assay using cDNAs derived from human monochromosomal 15 x rodent hybrids

Hybrids A15, A59.3aZ, A9+15 contain a paternal chromosome 15. Cell lines 20L.28, t75.2maZ, ALA.8 contain a maternal 15. RT-minus reactions are labelled with a minus (-).

3.3.1.5 1354-G13 represents an extended 3' untranslated sequence of *UBE3A*

There were three main points which suggested that 1354-G13 might represent a previously unidentified, extended 3' UTR of the *UBE3A* transcript. Firstly, 1354-G13 (and 215-B4) mapped to PACs 50I2 and 158H23 in the proximity of the 3' coding exons of *UBE3A* (Table 3.2.5; Fig 3.2.28). However, the 155bp final exon 10 of *UBE3A* was not single-copy and did not identify a single *EcoR*I fragment on the ASCR genomic contig, so the location of 1354-G13 with respect to this exon could not be established. Secondly, sequence and northern analysis showed that 1354-G13 represented the 3'UTR of a 5.2kb transcript. This was remarkably similar to the previously reported transcript size (~5kb) for *UBE3A* coding exons (section 3.3.1.2.1). Thirdly, just 2.6kb of coding sequence, 50bp 3' UTR (Huibregtse *et al.*, 1993), and up to 1kb of alternatively spliced 5' UTR sequence (Kishino *et al.*, 1997; Sutcliffe *et al.*, 1997), totalling <3.6kb had been identified for *UBE3A*. Thus 1.6-2kb of the *UBE3A* transcript remained to be identified, which could be accounted for by a longer 3' UTR.

Close examination of the 5' end of 1354-G13 showed that the first 17bp were the same as the final 16bp of *UBE3A* exon 10, except that 1354-G13 had an additional A inserted at between nucleotides 5 and 12 of 1354-G13 (Fig 3.3.2). To determine if 1354-G13 was an extended 3' sequence of *UBE3A*, exon connection RT-PCR was performed between coding exon 9 of *UBE3A* (renumbered exon 15) and the 3' end of 1354-G13. Forward primers 9F9 from within exon 9 and reverse primer 'G13-3' (Fig 3.3.2) were used to amplify a predicted >2kb RT-PCR product from fetal brain and heart. A ~2.2kb amplification product was observed in the brain, and more strongly in heart, where *UBE3A* is expressed at higher levels (Fig 3.3.9). To confirm that 1354-G13 was linked to exon 9, the RT-PCR product from the fetal heart cDNA was sequenced using primer 9F9. The sequence showed the order to be from the end of exon 9, then the previously reported exon 10 sequence followed by the 1354-G13 sequence, starting at nucleotide 18 with respect to Fig 3.3.2 (Fig 3.3.10). This showed that 1354-G13 was contiguous with the previously reported final exon, and extended beyond it. Thus 1354-G13 represents an additional 3' untranslated sequence of *UBE3A*, making the cDNA sequence up to ~5kb in length. The full 3' UTR of *UBE3A* was also identified and reported by Kishino and Wagstaff, (1998), as discussed in Chapter 6 (section 6.1.2).

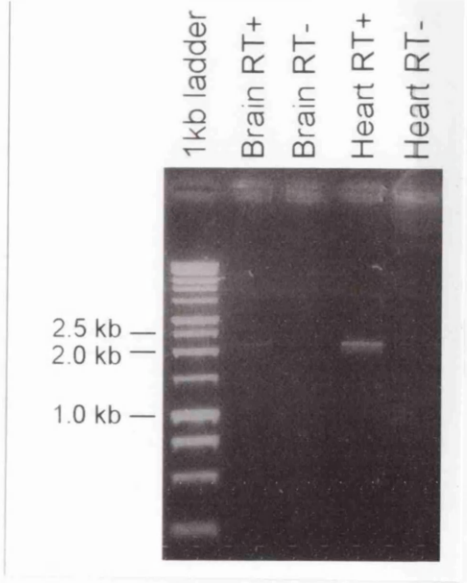


Fig 3.3.9 Exon connection RT-PCR between *UBE3A* exon 9 and 1354-G13.

RT-PCR was performed using *UBE3A* exon 9 forward primer 9F9 and reverse primer 1354-G13. Amplification was carried out for 36 cycles, with extension for 1 minute 15 seconds. A predicted > 2kb RT-PCR product was detected at its higest levels in fetal heart, showing that the 1354-G13 sequence was 3' to *UBE3A* exon 9 in the cDNA product.

TGAGAAACAAGATCTTCTTGCAGTTTACAACGGGCACAGACAGACANNTGTGGGAGGACTAGGAAAATTAAAGA

*TGATTATNGCCNAAAATGGCCCAGACACAGAAAG***GTTACCTACATCTCATACTTGCTTTAATGTGCTTTTACTT**

CCGGAATACTCAAGCAAAGAAAAACTTAAAGAGAGATTGTTGAAGGCCATCACGTATGCCAAAGGATTTGGCAT

*

GCTGTAAAACAAAACAAAACAAAATAAAACAAAAAAAAGGAAGGAAAAAAAAAGAAAAAATTTAAAAAATTTTA

AAAATATAACGAGGGATAAATTTTTGGTGGNGATAGTGTCCCAGTACAAAAAGGCTGTAAGATAGTCAACCACA

GTAGNCACCTATGTCTGGGCCTCCCTTNTTTATTGGGGACATGTGGNCTGGAACAGNANATTTCAGCTACATAT

ATGAACAAATCCTTTATTATTATTATAATTATTTTTTTGNGTGAAAGGNTTACATATTCTTTCACTTGNATNTA

CAGAGAGGGTTTTCTGAATATTTATTTTAAGGNTTAAATCACTTTTTCTTGGGTNTAATTACNTGCTTGGNGGG

TTGANCCCTTTTGNGGGATTTAAAAAATTTTNTTCCCAACAGGAACCTACCTTTCCCAAGGGAAAAATNTTNCC

NCCATTTTNTAAGACCCCCGTAANCCTTTCAAGTATGGGGCCTANTTTTTTTTGNCCCCTGG

Fig 3.3.10 Sequence showing sequence 1354-G13 is an extension of the *UBE3A* 3'UTR.

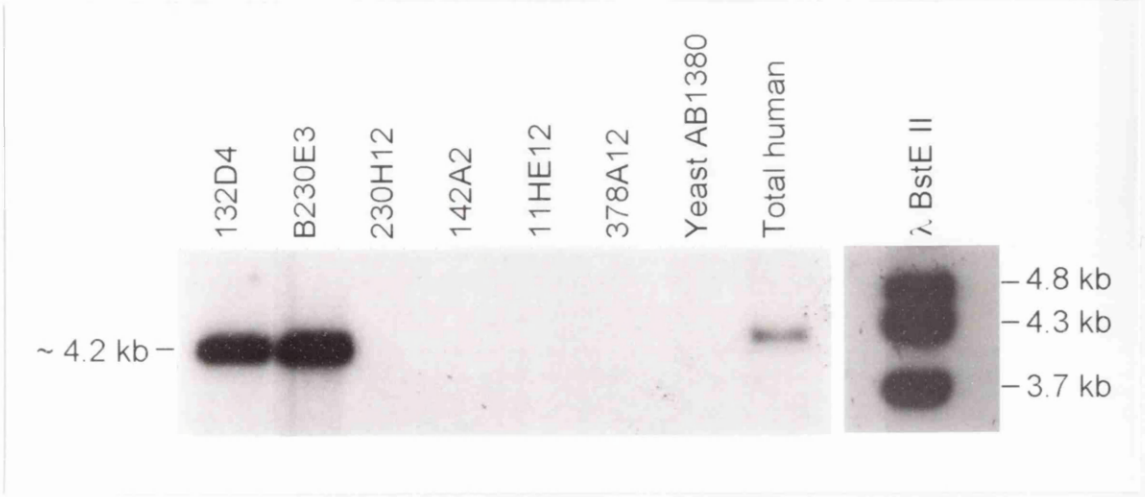
The sequence of the RT-PCR product between *UBE3A* previous exon 9 (exon 15) and 1354-G13, sequenced from the 5' end with primer 9F9, is shown. The sequence from the end of exon 9 (15) is italicised. The previous exon 10 sequence is in bold. The 1354-G13 sequence, underlined, is contiguous with exon 10 (Fig 3.3.2). An additional A nucleotide had been noted at the start of the 1354-G13 sequence, which was not present in the overlapping section of the previous exon 10 sequence. The additional A was also observed in this sequence, indicated directly above by an asterix (*). Nucleotides 'N' are sequencing errors.

3.3.2 Analysis of single-copy short selected fragments mapping to the ASCR.

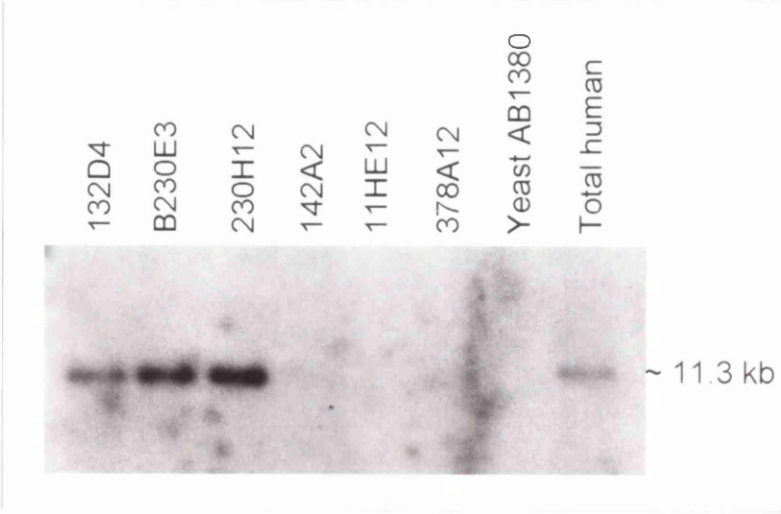
The 24 non-redundant selected sequences that had originally failed to identify any sequences in the databases by BLAST analysis (up to January, 1998) (section 3.2.9.3), were characterised briefly to determine if they represented a novel transcript. The insert sequences from the 24 selected clones were excised from the pTAg vector by digestion with *EcoR*I and *Xho*I, and mapped to the YACs, PACs and BACs from the ASCR contig. The inserts were also hybridised to *EcoR*I blots of total human and mouse genomic DNA to determine which were single-copy, and if any were evolutionarily conserved. Only 8 sequences, 16, 21, 26, 39, 86, 104, 113 and 119 were found to be single-copy and mapped to the ASCR (Table 3.3.1). The nucleotide sequences for these are given in the Appendix. Representative examples of the mapping studies performed are shown for selected sequences 86 and 104 in Fig 3.3.11. None of the 8 sequences was conserved in mouse DNA. The remaining 16 were repetitive, showing a smear in the total human DNA lane, or did not map to 15q11-13.

Sequence analysis of the 8 single-copy sequences was performed using the 'Map' program and NIX analysis to determine whether any were coding or contained any sequence motifs. None of the sequences was coding in any of the six potential reading frames (data not shown). Seven of the sequences (16, 21, 39, 86, 104, 113, and 119) were found to consist largely or entirely of repeat elements, which is surprising since these were single-copy to 15q11-13 (Table 3.3.1). Presumably these were low copy-number repeat sequences, which at stringent post-hybridisation washing conditions of 0.1X SSC/ 0.1% SDS at 65°C, appeared to be unique to the ASCR. It was considered probable that these sequences would hybridise to other repetitive clones if used as probes for conventional cDNA library screening, so this was not performed.

A major sequencing effort within the 15q11-13 was instigated at the Genome Science and Technology Center, University of Texas, USA in 1998, using PAC and BAC DNA as template. PAC 14I12 was sequenced first, and the completed sequence was entered into GenBank in June, 1998. The 8 single-copy selected sequences identified their corresponding sequences within PACs 14I12, 373B1 and 351H23 in GenBank by BLAST searching (Table 3.3.1). Sequence 26 was contained within *UBE3A* intron 3. The query selected sequences were contiguous with genomic DNA, suggesting they were either genomic fragments contaminating the original fetal brain cDNA libraries Br32 and Br33, or were transcribed



A: Selected sequence 86 (clone co-ordinate PAC-1β 10B) was single-copy to YACs 132D4 and B230E3.



B: Selected sequence 104 (clone BACα 3B) was single-copy to YACs 132D4, B230E3 and 230H12.

Fig 3.3.11 Identification of single-copy selected fragments mapping to the ASCR.

Autoradiographs of single-copy selected sequences hybridised to *EcoR*I digested YAC DNA from the ASCR and human genomic DNA are shown.

Sequence number	Clone co-ordinate (redundancy)	Size (bp)	Mapping position	*EcoR*I fragment (kb)	NIX analysis	BLAST analysis (to December, 1998).
16	PAC-1β 10D	298	132D4, B230E3, 318N23, 14I12	8.4	Contains 277 bp L1 (MSR1) repeat	97% identity to 15q11-13 PAC 14I12 sequence (AC004259), n 40,658-40,387
21	B230E3-aluα 1E 46 clones in the B230E3-alu sublibary	401	132D4, B230E3, 318N23, 14I12	4.0	Contains Alu repeat	99% similar to PAC14I12 (AC004259), n 82,331-82,719.
26	B230E3-aluα 11E	283	132D4, B230E3, 318N23, 14I12, 139E9	6.3		98% identity to 15q11-13 PACs 14I12 (AC004259), n13,048-12,830 and PAC 373B1 (AC004600) n81-283 are UBE3A intron 3 (AF016706)
39	PAC-1ε 1C	221	132D4, B230E3, 50I2, 139E9	2.8	Contains MER 37 repeat element	96% identity to 15q11-13 PAC 373B1 sequence (AC004600), n18,057-18,269.
86	132D4φ 11B PAC-1α 11C PAC-1β 4D PAC-1β 5H	419	132D4, B230E3, 50I2	4.2	Contains MER 37 repeat element	93% identity to 15q11-13 PAC 373B1 sequence (AC004600), n103,527-103,939
104	BACα 3B BACα 4C	366	132D4, B230E3, 230H12, 50I2, 139E9, 268A2	11	Contains 217 bp L1 (CA) repeat	97% identity to 15q11-13 PAC 351H23 (AC004738), n112,117-112,469
113	BACα 10H plus 8 BAC / PAC-1 clones	126	132D4, B230E3, 14I12, 139E9, 268A2	3.8	Consists entirely of MER 33 repeat	98% identity to 15q11-13 PACs 14I12 (AC004259), n145,483-145,375, and 351H23 (AC004738)
119	BACβ 3G BACβ 12A	190	132D4, B230E3, 230H12, 139E9	6.3	Consists of L1 (MSR1) repeat	99% identity to 15q11-13 PACs 14I12 (AC004259), n75,655-75,836, and 373B1 (AC004600)

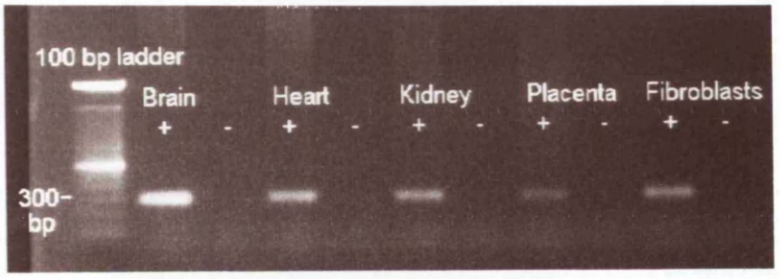
Table 3.3.1 Summary of single-copy sequences isolated from selected cDNA sublibraries, mapping to the ASCR.

The clone numbers, including redundant clones and insert sizes are given for the selected sequences. The mapping positions and sizes of the *EcoR*I fragments identified are listed for the selected clones. The sequence content based on NIX analysis and BLAST search results accurate to December, 1998, are given. The selected sequences were contiguous with the genomic sequences from 15q11-13 PACs in GenBank. The nucleotide (n) positions with respect to the PAC sequences are given.

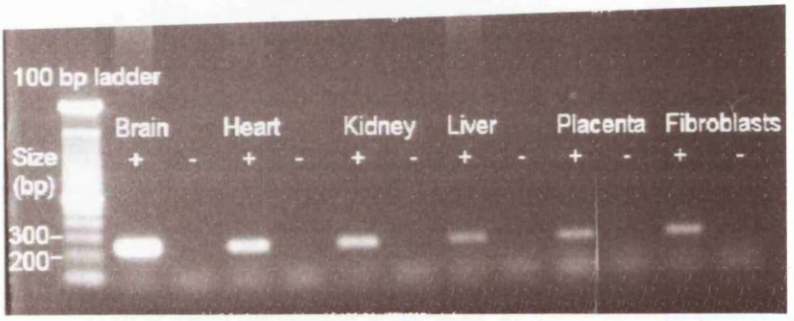
sequences from within single exons that overlapped *UBE3A*.

To determine if the selected sequences were expressed, RT-PCR analysis of various fetal tissues and adult fibroblasts was performed on sequences 16, 21, 26, 39, 86, 113, and 119. Expression analyses were not performed for sequence 104, which was identified as a CA_n dinucleotide repeat (Table 3.3.1). Since the selected sequences were contiguous with genomic DNA, control reactions in which RT had been omitted were included to detect genomic contamination. Each of the 7 selected sequences was found to be expressed in all samples tested, as shown for sequences 21, 26 and 86 in Fig 3.3.12. This suggested the existence of one or more transcripts in the ASCR, which overlap *UBE3A*. However, due to time constraints these sequences were not further analysed.

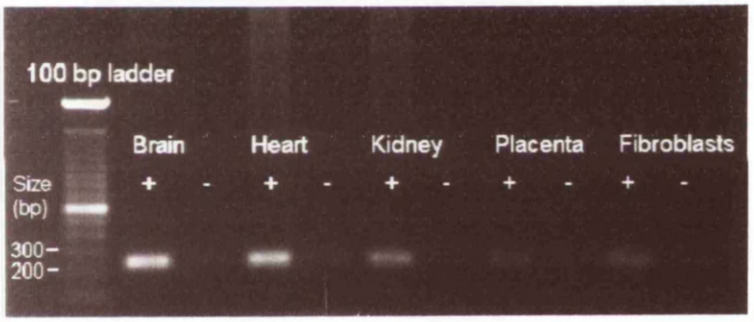
Subsequent to this, an antisense RNA overlapping *UBE3A* was reported. This non-coding RNA was found to be transcribed from *UBE3A* exon 16 to exon 11, and included the intervening intron sequences. The full-length *UBE3A* antisense transcript was not identified, and may include additional exons. This RNA was expressed solely in brain, from the paternal allele (Rougelle *et al.*, 1998). Analysis of the region encompassed by the *UBE3A* antisense RNA within PACs 14I12 and 373B1, indicated that selected sequences 21, 39 and 86 mapped within this region. Thus 3 selected sequences isolated may have represented the *UBE3A* antisense RNA sequence. Although sequence 26 was expressed predominantly in fetal brain (Fig 3.3.12B), expression of sequences 21, 26 and 86 was detected in several fetal tissues and adult fibroblasts (Fig 3.3.12), whereas transcription of the antisense RNA is confined to the brain (Rougelle *et al.*, 1998). The identity of the selected cDNAs was not confirmed.



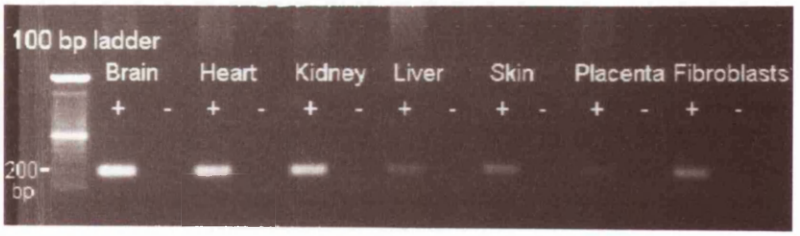
A: Selected sequence 21 (product length 300bp)



B: Selected sequence 26 (product length: 240bp)



C: Selected sequence 86 (product length: 245bp)



D: *HPRT* control

<u>Fig 3.3.12</u> RT-PCR analysis of selected sequences.

RT-PCR was performed in fetal tissues and adult fibroblasts for 35 cycles of amplification to detect expression of selected sequences. PCR Products of the appropriate sizes are shown for selected sequences 21, 26, and 86, indicating they are transcribed. RT minus (-) controls were included to detect genomic contamination. *HPRT* was used as a housekeeping control cDNA.

3.3.3 Analysis of an evolutionarily conserved, unique-copy genomic DNA fragment, P1.1-6

3.3.3.1 Identification of 'P1.1-6'

Prior to the report of a maternal deletion in non-AS patient D.H., which excluded marker D15S113 from the region most likely to contain the AS gene (Fig 1.9) (Michaelis *et al.*, 1995), the region surrounding D15S113 was analysed for coding sequences through the identification of evolutionarily conserved genomic fragments. Cosmids isolated from a 250kb D15S113-positive YAC 11HE12 (Fig 3.3.13), were restriction digested. Single-copy, conserved restriction fragments were sought by hybridisation to human and mouse genomic DNA. A single-copy 1.1kb *Pst*I fragment 'P1.1-6' was identified from cosmid E42 (Fig 3.3.13), which cross-hybridised to mouse genomic DNA (Fig 3.3.14) (Jess Buxton, personal communication). P1.1-6 mapped just distal to D15S113 (Fig 3.3.13)(H. Gilbert, 1996). However, P1.1-6 failed to detect any transcript on Northern analysis, and did not identify any positive clones when used as a probe to screen the Stratagene 18 week fetal brain cDNA library (J. Buxton, personal communication).

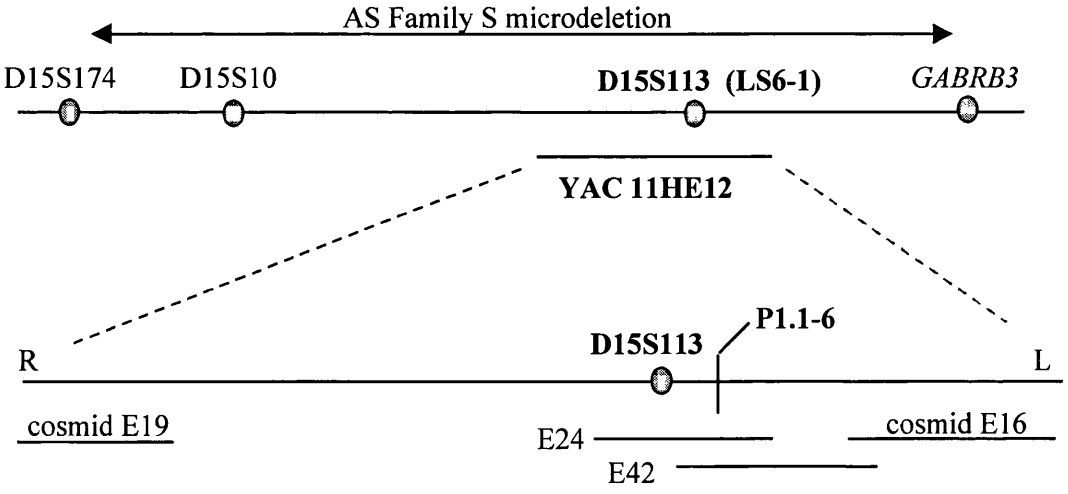


Fig 3.3.13 Map of the region surrounding D15S113.

Arrows indicate the AS Family S deletion, which represented the AS candidate region prior to the report of non-AS patient D.H. Cosmids isolated from YAC 11HE12 and searched for evolutionary conserved fragments are labeled. The position of P1.1-6 with respect to D15S113 is shown.

Fig 3.3.14 P1.1-6 is conserved in mouse genomic DNA.

*EcoR*I blot of total human (H) and mouse (M) genomic DNA hybridised with *Pst*I fragment P1.1-6. The filter was washed to 0.1X SSC/ 1% SDS, and exposed to X-ray film overnight. Hybridisation to two *EcoR*I fragments in human DNA indicates that P1.1-6 contains an *EcoR*I site.

It was considered possible that P1.1-6 contained an open reading frame which was responsible for the cross-hybridisation to mouse DNA. If this represented a small proportion of the 1.1kb fragment, then screening of the cDNA library with P1.1-6 may have failed due to a low 'signal to noise' ratio in the probe. P1.1-6 was sequenced to identify any coding region within the fragment, which could be used as a probe for further library screening.

3.3.3.2 Transposon-facilitated sequencing of P1.1-6

The 1.1kb sequence for P1.1-6 was acquired using a transposon-facilitated sequencing strategy followed by computational analysis. This is depicted in Fig 3.3.15. P1.1-6 was excised from pBluescript KS2 and cloned into the miniplasmid vector pMOB, via the *Pst*I cloning site (Fig 3.3.15A). The pMOB-P1.1-6 construct was transformed into the donor cell strain DPWC, which contains the transposon-bearing F factor. PCR amplification was performed across the MCS using vector primers T3 and T7, giving a product of >1.1kb, to check that the insert was cloned successfully (Fig 3.3.15B). Cointegration of the pMOB-P1.1-6 construct with the F factor occurs within these cells, and the insert is targeted for transposon incorporation. Mating of the donor DPWC cells with recipient F^- JGM cells allows conjugal transfer

of the cointegrate, which is resolved back into its derivative components within the JGM cells.

Transposition into P1.1-6 was random, such that the JGM cell population contained many variants (Fig 3.3.15C). These were plated out and over 20 JGM subclones picked for mapping of the transposon integration sites within P1.1-6. This was achieved by PCR amplification across P1.1-6, between the transposon and either end of the pMOB MCS. Separate amplification reactions were set up for each subclone, with pMOB vector primer LREV or LUNI combined with primer GD1R, which binds to inverted repeat sequences present at both ends of the transposon (Fig 3.3.15D). The sizes of the two PCR products added together, came to 1.1kb, mapping the transposon integration site with respect to either end of the pMOB MCS (Figs 3.3.3.15D and 3.3.16). A map of the various transposon integration sites within P1.1-6 was produced (Fig 3.3.15E). Seven subclones with transposons inserted at 150-250bp intervals within P1.1-6, were selected to provide full sequence coverage of P1.1-6 (Fig 3.3.15F).

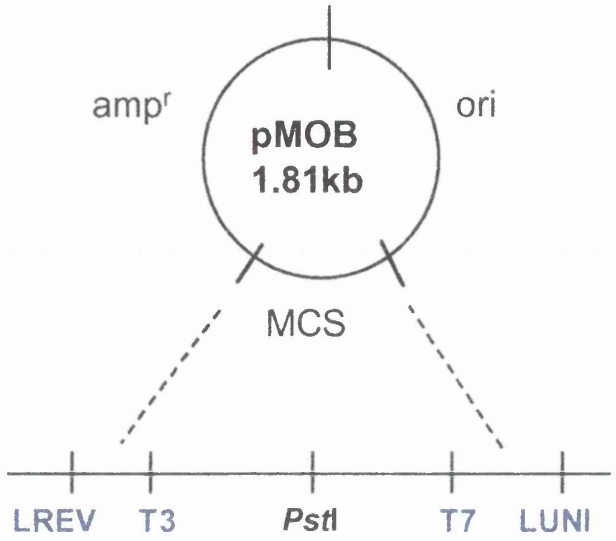
Sequencing primers GD1 and GD2, which are complementary to sequences at the termini of the transposon, were used to sequence P1.1-6 from the point of transposon integration. This occurred bidirectionally, using opposite strands as template (Fig 3.3.15G). Manual sequencing was performed using 'miniprepped' DNA from the seven subclones as template, and detected by incorporation of ^{35}S radiolabel and autoradiography (Fig 3.3.17A). Vector primers T3 and T7 were used to sequence the ends of P1.1-6 using 'miniprepped' DNA from the pBlueScript KS2 clone as template (Fig 3.3.17B). A set of sixteen overlapping sequences was generated for P1.1-6 (Fig 3.3.15H).

3.3.3.3 Derivation of the P1.1-6 consensus sequence by computation.

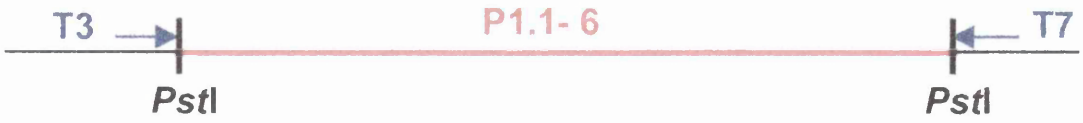
A consensus sequence for P1.1-6 was derived from the pairs of sequences using the GCG Sequence Analysis package. The 'GelMerge' program identified a series of overlaps between sequences from adjacent transposon sites. These were aligned into a contig of short overlapping sequences and represented graphically using the 'GelView' program (Fig 3.3.18). These sequence overlaps were concordant with the PCR mapping data (3.3.15E), and provided dual sequence coverage over much of the P1.1-6 fragment. The 'Consensus' program was used to generate the final consensus sequence for P1.1-6.

Fig 3.3.15 Flow diagram depicting the subcloning, transposition and sequencing of P1.1-6 using a transposon (Tn) facilitated strategy.

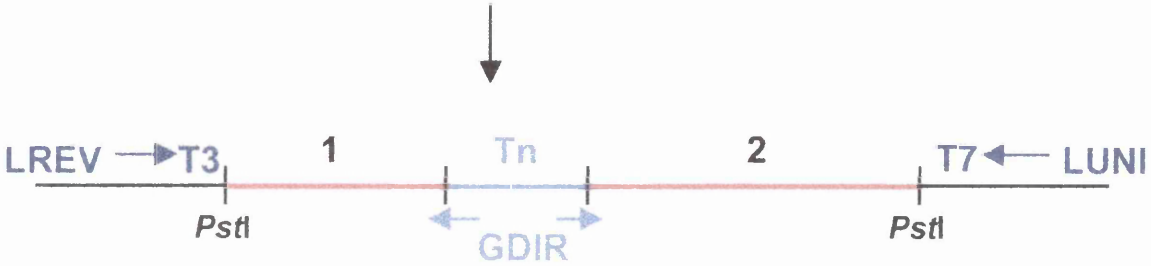
A: pMOB vector presents cloned insert as target for transposition



B: P1.1-6 was cloned in pMOB via the *PstI* site in the MCS. PCR amplification products of 1.1kb with T3 and T7 primers indicated that ligation and transformation had occurred.

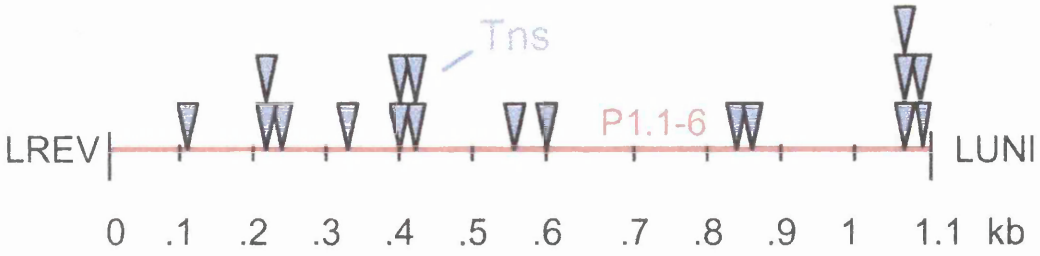


C: Tns randomly, but preferentially integrate into the P1.1-6 target sequence, introducing sequencing primer sites into the fragment. Tns could insert themselves in either orientation.

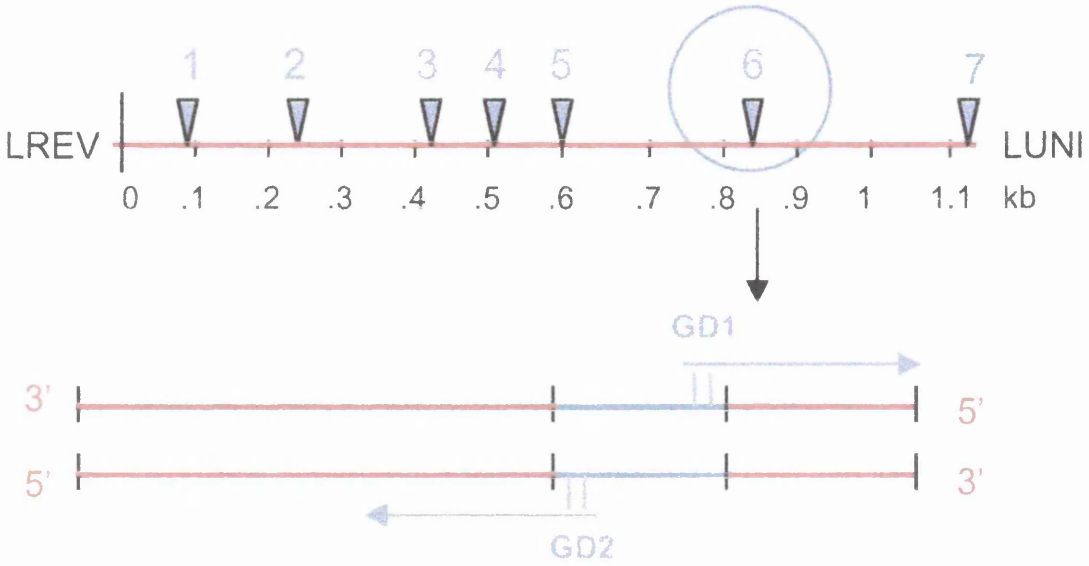


D: PCRs **1** and **2**, with either pMOB primer LREV or LUNI and Tn primer GDIR, mapped each Tn integration site within P1.1-6.

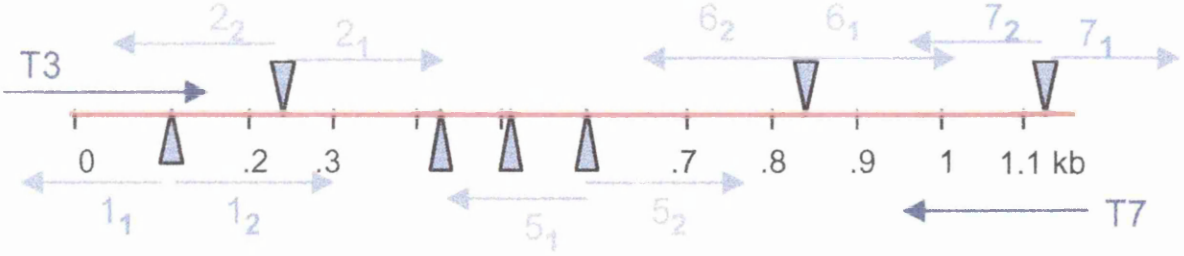
E: Final transposon - integration map within P1.1-6 for various subclones picked at random.



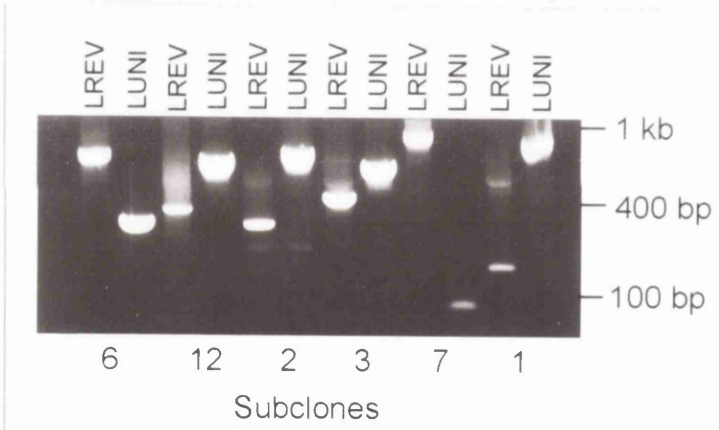
F: Subclones 1- 7, mapping 150-250bp apart, were selected for sequencing



G: Sequencing of both strands of P1.1-6 was bi-directional using primers GD1 and GD2 which are complementary to sequences within the Tn.



H: Sets of P1.1-6 sequences from adjacent Tn sites overlapped, some in reverse order depending on the orientation of the Tn. Vector primers T3 and T7, were used to sequence respective ends of P1.1-6. A consensus sequence was derived by computation



Iig 3.3.16 PCR-mapping of the transposon integration sites within P1.1-6.

Two separate PCR reactions were performed for each subclone; LREV combined vith GD1R gave the distance between the transposon integration site and one end (T3) of the pMOB MCS, and LUNI plus GD1R, gave the distnace bewteen the tansposon and the other end (T7) of the pMOB MCS. The products of the two PCRs vere electrophoresed and sized with respect to a DNA ladder. The two amplification products generated for each subclone added to 1.1kb, enabling the site of transposon iisertion within P1.1-6, to be mapped.

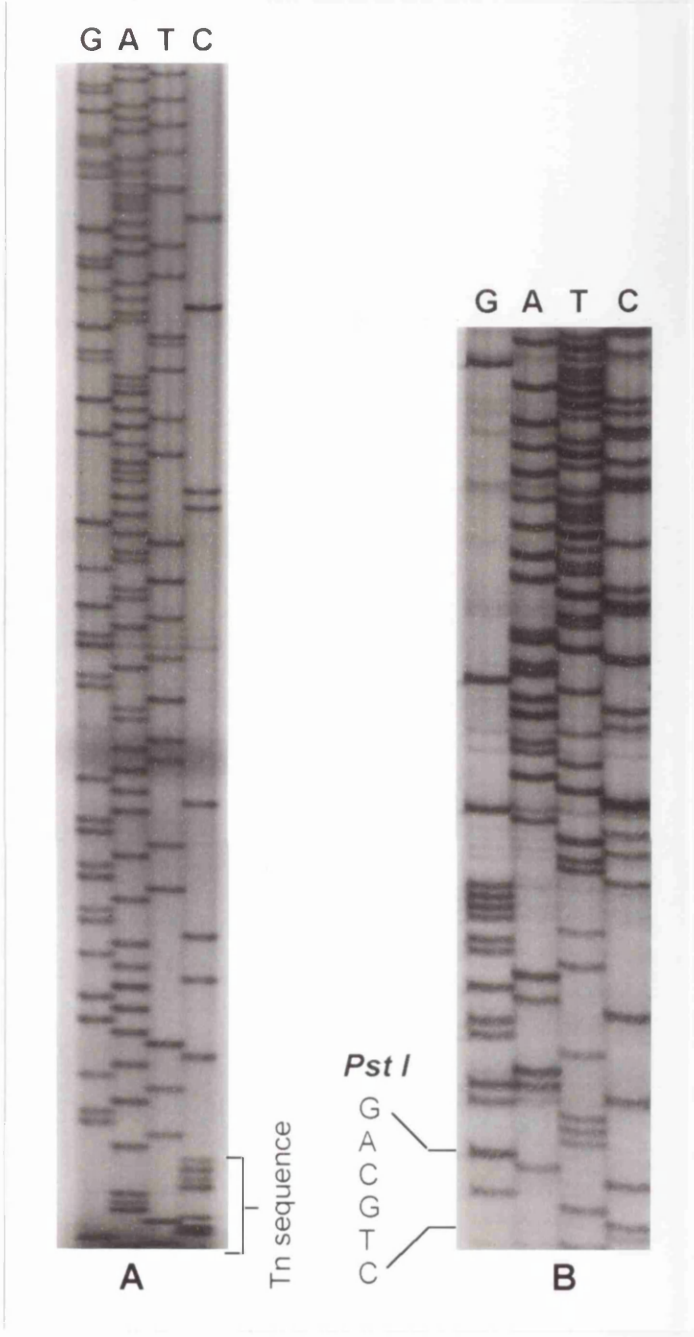


Fig 3.3.17 Manual sequencing of P1.1-6 using ^{35}S radiolabel.

Sections of sequence from P1.1-6 after polyacrylamide gel electrophoresis and autoradiographic detection. **A:** Section of P1.1-6 from subclone 5 using transposon primer GD2, which contains part of a microsatellite repeat region as follows:
ATGGATGACTAGAGACAGACGGATGGATGGATGGACAGATATGAATGGATGGATGGATG
AATGAATAGACACAAATAGATAAGAAATGGTTGAACAGATGGATAGACAAATAGAGATG
GATAGATGGATA
B: End sequence of P1.1-6 using vector primer T3. The sequence following the *Pst*I cloning site is: TTTGC G/A ATGGCAGATGGTGGGGCTTCTCAGCCTATAATCACATGAACT
AATTCCTCATTATCTATTTATCCATCTATCCATCTCTATTTGTCTATC

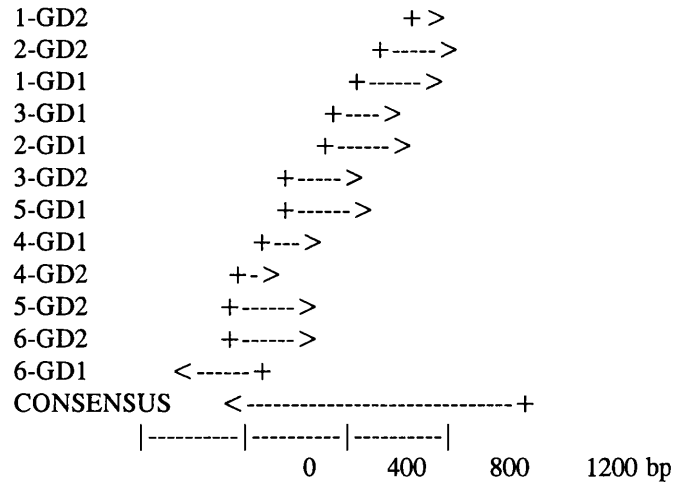


Fig 3.3.18 Alignment of the overlapping sequences to generate a consensus sequence for P1.1-6.

The 'GelView' fragment assembly contig formed by overlapping sequences between adjacent transposons within P1.1-6 is shown. The consensus P1.1-6 sequence was derived from the contig. The sequences are depicted by arrows, and positioned with respect to the final consensus sequence.

3.3.3.4 Sequence analysis of P1.1-6

BLAST searches of the GenBank, Embl, Rodent and 'Alu-like sequence' databases were conducted using the P1.1-6 consensus sequence as a query to determine if there was any homology to sequences in the databases. The final 570bp of P1.1-6 overlapped an 834bp unmapped human STS, UT7715 (GenBank accession: L30335), demonstrating 97% sequence identity within the overlap. UT7715 contained a tetranucleotide repeat sequence and extended beyond P1.1-6 by a further 203bp (Fig 3.3.19). The combined sequences (P1.1-6/UT7715) extended the sequence information available to 1.36kb (Fig 3.3.20).

P1.1-6 also identified a large number of sequences containing $(GATA)_n$ and $(TGGA)_n$ repeats in the BLAST searches, indicating the presence of a tetranucleotide repeat within the fragment. The combined 1.36kb P1.1-6/UT7715 sequence was directly compared to itself, enabling the repeat region to be identified between nucleotides 810 and 1150 (Fig 3.3.21). The presence of a simple $(TGGA)_n$ repeat element between nucleotides 814 and 924 was confirmed by the 'Repeatmasker' facility in the NIX sequence analysis program. A retroviral long tandem repeat (LTR) was identified between nucleotides 1001 and 1359 of P1.1-6/UT7715 (Fig 3.3.20).

```
DEFINITION: Human STS UT7715, 834bp ACCESSION: L30335
KEYWORDS: STS sequence; microsatellite DNA; tetranucleotide repeat
Percent Similarity: 96.678   Percent Identity: 93.706

 501 ATTTTCTGGTGAAACCGACCCTAACCACCACCATTTCGAGTCTCAGCCCC 550
                          |   | | |     ||||||||||||||||
   1 .....................TCGACTCTAGAGGATCGAGTCTCAGCCCC 29

 551 TTACTATGACCTCTTAGCCTCTGTGTGCATTTCCCCTAGAGCCAAGCTTA 600
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  30 TTACTATGACCTCTTAGCCTCTGTGTGCATTTCCCCTAGAGCCAAGCTTA 79

 601 GACATGGCAGCTTGATGTCCCCCCTGAGCCCAGCACAGTGGCTACTAGGT 650
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  80 GACATGGCAGCTTGATGTCCCCCCTGAGCCCAGCACAGTGGCTACTAGGT 129

 651 TAATAGGGGCCAAAAGAAAAACACCATAAAATAATAACATAACAAGCATG 700
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 130 TAATAGGGGCCAAAAGAAAAACACCATAAAATAATAACATAACAAGCATG 179

 701 GTAGGCAGGTAGGCAGGCAGGCAGNACAGCAAGGAAGTTCTCTCTCCAAA 750
     |||||||||||||||||||||| :||||||||||||||||||||||||||
 180 GTAGGCAGGTAGGCAGGCAGGCA.GACAGCAAGGAAGTTCTCTCTCCAAA 228

 751 CCAAAACCTCCTTAAGGGTGGTTGTATTAGTCATGTTCTCCAGAGACACN 800
     |||||||||||||||||||||||||||||||||||||||||||||||||  :
 229 CCAAAACCTCCTTAAGGGTGGTTGTATTAGTCATGTTCTCCAGAGACA.C 277

 801 AACCAATANNNNGGATGGATGGATGGATGGATGGATAGATAGATAAATAG 850
     ||||||||::::|||||||||||||||||||||||||||||:|||||||
 278 AACCAATAGGATGGATGGATGGATGGATGGATAGATAGNTAAATAG 327

 851 ATAGATATGGATATGTGGATNGGATTGATAGATGATTGATAGATGATAGA 900
     |||||||||||||||||| :|||||||||||| ||||||||||||||:
 328 ATAGATATGGATATGTGGA.TGGATTGATAGATG.TTGATAGATGATAGN 375

 901 TAGATAGATGTGGATGGATTACTAGAGCCAGACGGATGGTTGCAGTTTTG 950
     |||||||| |:||||| |||||:|:||||| |||:|||||||||||||:
 376 TAGATAGATTTNGATGG.TTACTNGNGCCAG.CGGNTGGTTGCAGTTTTN 423

 951 ATTGGATGGATGAATGAATAGACCCCAATTGATTGAAATGGTTGAACAGA 1000
     ||||       |     :: | |||||||||:||||||||||||||||||
 424 ATTG.......GTGATGNNTGGCCCCAATTGNTTGAAATGGTTGAACAGA 466

1001 TGGATAGACAAATAGAGATGGATAGATGGATAAATAGATAATGAGGAATT 1050
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 467 TGGATAGACAAATAGAGATGGATAGATGGATAAATAGATAATGAGGAATT 516

1051 AGTTCATGTGATTATAGGGCTGAGAAGCCCCACCATCTGCCATCTTCAAA 1100
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 517 AGTTCATGTGATTATAGGGCTGAGAAGCCCCACCATCTGCCATCTTCAAA 566

1101 CTGCAG..................................................... 1106
     ||| ||
 567 CTGGAGACTCAGTAGAGCTAGTATCTAATTTAGCGTCTGAAGGCCTGAGA 616
```

Fig 3.3.19 Identification of human STS UT7715 by BLAST analysis of P1.1-6
The BLAST output for P1.1-6 is shown. P1.1-6 (top sequence) identified human STS UT7715 (bottom sequence), which contains a tetranucleotide repeat. The two sequences were 97% similar, over a 570bp overlap.

```
   1  CTGCAGGATA CCCTTGCACC GGAAAAGTAA TCAGCTGACT TATTTTTGCG
       PstI                                                HhaI
  51  CATTGAAAGC ACAAAATTCT GAGTAAATGC CACCACGTTT CATCACTAAT

 101  TGGGACAGAC CCATCTGATC TTTCCGTAGA AGGATGTCCT GCCGAGAACT

 151  TCCAAAATTG GATTCACTTT GCCGGGCACA GCAACCCTCA CCTGAACCAC

 201  CTTGCACACA CCCTGGGATG CAGTGGCTGT AGTTAGCATT GTACGGGCAT

 251  CAGAAGAGGG TTGTTTTCAG TGGAAATGAC TTTGAACTCT CCATGTCACA

 301  CACGGAAAGA AATCAGGCTA CAGAAGTCAT CATAATATAT TTTTATGTGG

 351  TCTCAAATGT GTTGTTTCAG GAAGGTCTTG GAAAGCTTTC AGTGCTTCTA

 401  TTTTTAGGGC TGAACGAGAA TTCAGGGACA CGGGAGAGTG AGAAGCAATC
                         EcoRI
 451  ATTTTTGCTC CGAGAGTATT TTCTTCCCCA GGATTTTTTT TTCTCCCCAA

 501  CCTGNTAATT TTCTGGTGAA ACCGACCCTA ACCACCACCA TTTCGAGTCT

 551  CAGCCCCTTA CTATGACCTC TTAGCCTCTG TGTGCATTTC CCCTAGAGCC

 601  AAGCTTAGAC ATGGCAGCTT GATGTCCCCC CTGAGCCCAG CACAGTGGCT

 651  ACTAGGTTAA TAGGGGCCAA AAGAAAAACA CCATAAAATA ATAACATAAC
                    HaeIII
 701  AAGCATGGTA GGCAGGTAGG CAGGCAGGCA GACAGCAAGG AAGTTCTCTC

 751  TCCAAACCAA AACCTCCTTA AGGGTGGTTG TATTAGTCAT GTTCTCCAGA

 801  GACACAACCA ATAGGATGGA TGGATGGATG GATGGATGGA TAGATAGATA

 851  AATAGATAGA TATGGATATG TGGATGGATT GATAGATGAT TGATAGATGA

 901  TAGATAGATA GATGTGGATG GATGACTAGA GCCAGACGGA TGGTTGCAGA

 951  TATGATTGGA TGGATGAATG AATAGACACA AATAGATAGA AATGGTTGAA

1001  CAGATGGATA GACAAATAGA GATGGATAGA TGGATAAATA GATAATGAGG

1051  AATTAGTTCA TGTGATTATA GGGCTGAGAA GCCCCACCAT CTGCCATCTT

1101  CAAACTGCAG ACTCAGTAGA GCTAGTATCT AATTTAGCGT CTGAAGGCCT
          PstI
1151  GAGAACCAGG GAAGCCAATG GTGTAACCCC CAGTTCAAGG ACAGGAAAAG

1201  ATGAGATGAA ATCTTGCAGT TTAAGCAGTG ACGCAGAAAA AAAAGAGGCA

1251  AATGTCTCCT CCCTCCCTTT GTTCTACCCA GGTTGTCACC AGGGTGGATG

1301  CAGTGCCGCA TGTTGGAGAC GGTGATCCAC TTTACTGAGT CCCCCAGTGT

1351  GAATGCTGAC TT
```

Fig 3.3.20 Final sequence of P1.1-6 and UT7715 combined.
The recognition sites of restriction enzymes are in bold, including the *Pst*I cloning sites of P1.1-6. The region containing two predicted open reading frames in the forward and reverse-complement directions is underlined. In italics are a (TGGA)n repeat region (nucleotides 814-924) and an LTR (nucleotides 1001-1359).

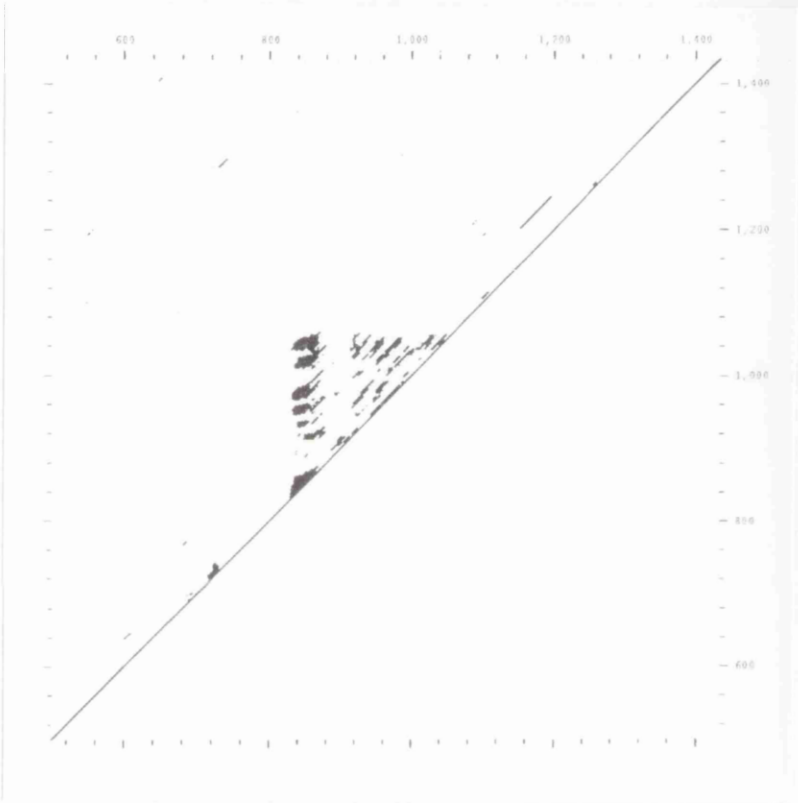


Fig 3.3.21 Identification of the repetitive element within P1.1-6/UT7715 by self-comparison of the sequence.

'Dotplot' output, giving a graphical representation of the self-comparison of the P1.1-6/UT7715 sequence. The diagonal line representing the central axis of the graph indicates that the two sequences compared are exact replicas. Where sequence similarity has been detected between different sites within the length of the sequence, a line is drawn against the respective nucleotide numbers, parallel to the central axis. The symmetrical triangle between nucleotides 810 and 1150, made up of many juxtaposed lines, is indicative of a repeat element at this position.

The 'MAP' program was used to identify any potential open reading frames (ORFs) within P1.16/UT7715, in both the forward and reverse-complement directions. One possible 270 bp ORF, (ORF-1) was predicted between nucleotides 233 and 503 in the forward direction, coding for 90 amino acids (Fig 3.3.22A). An overlapping putative ORF (ORF-2) was identified in the reverse-complement direction, running 5' to 3' from nucleotides 527 to 329, coding for 65 amino acids (Fig 3.3.22B). The 'Genemark' program within the NIX analysis system predicted ORF-2 to be an exon. However, analysis of P1.1-6/T7715 by 'GRAIL' did not predict any intron/exon boundaries.

A: ORF-1

```
233 TTAGCATTGTACGGGCATCAGAAGAGGGTTGTTTTCAGTGGAAATGACTTTGAACTCTCC
    L  A  L  Y  G  H  Q  K  R  V  V  F  S  G  N  D  F  E  L  S

293 ATGTCACACACGGAAAGAAATCAGGCTACAGAAGTCATCATAATATATTTTTATGTGGTC
    M  S  H  T  E  R  N  Q  A  T  E  V  I  I  I  Y  F  Y  V  V

353 TCAAATGTGTTGTTTCAGGAAGGTCTTGGAAAGCTTTCAGTGCTTCTATTTTTAGGGCTG
    S  N  V  L  F  Q  E  G  L  G  K  L  S  V  L  L  F  L  G  L
          EcoRI
413 AACGAGAATTCAGGGACACGGGAGAGTGAGAAGCAATCATTTTTGCTCCGAGAGTATTTT
    N  E  N  S  G  T  R  E  S  E  K  Q  S  F  L  L  R  E  Y  F

473 CTTCCCCAGGATTTTTTTTTCTCCCCAACCTGNTAATTTTCTGGTGAAACCGACCCT
    L  P  Q  D  F  F  F  S  P  T  ?  *
```

B: ORF-2

```
527 GGTCGGTTTCACCAGAAAATTANCAGGTTGGGGAGAAAAAAAAATCCTGGGGAAGAAAAT
    G  R  F  H  Q  K  I  ?  R  L  G  R  K  K  N  P  G  E  E  N
                                                  EcoRI
467 ACTCTCGGAGCAAAAATGATTGCTTCTCACTCTCCCGTGTCCCTGAATTCTCGTTCAGCC
    T  L  G  A  K  M  I  A  S  H  S  P  V  S  L  N  S  R  S  A

407 CTAAAAATAGAAGCACTGAAAGCTTTCCAAGACCTTCCTGAAACAACACATTTGAGACCA
     L  K  I  E  A  L  K  A  F  Q  D  L  P  E  T  T  H  L  R  P

347 CATAAAAATATATTATGA 329
    H  K  N  I  L  *
```

Fig 3.3.22 Predicted open reading frames within P1.1-6/UT7715.

'MAP' output showing the predicted amino acid sequences below the DNA sequence for putative ORFs 1 and 2. Stop codons are denoted by a star. The nucleotides numbered on the left, are given with respect to their position within the P1.1-6/UT7715 sequence. The *EcoR*I site spanned by ORFs 1 and 2 is indicated. In **A** the DNA sequence is in the forward direction. In **B** the DNA sequence is in the reverse complement direction.

3.3.3.5 The region containing ORFs 1 and 2 is evolutionarily conserved, but fails to identify any clones by cDNA library screening.

A probe containing putative ORFs 1 and 2 was generated by restriction digestion of P1.1-6 with *Hha*I and *Hae*III (Fig 3.3.20), to produce a 620bp fragment. This probe identified the same *EcoR*I fragment in mouse DNA identified by P1.1-6 (Fig 3.3.14), indicating that this specific region was evolutionarily conserved and that ORF1 or 2 was likely to be coding (data not shown). Approximately 10^6 plaque forming units from each of two nine week fetal brain cDNA libraries Br 32 and Br 33, were screened using the same probe containing putative ORFs 1 and 2, to identify a cDNA containing either ORF. However, no positive clones were identified by this probe from either cDNA library.

3.3.3.6 Analysis of evolutionary conservation in the D15S113 region failed to identify coding sequences

Sequence analysis of the conserved genomic fragment P1.1-6, followed by cDNA library screening with a fragment from P1.1-6 containing a predicted exon, did not lead to the identification of any cDNA. No other evolutionarily conserved *Pst*I fragments were identified from the cosmids derived from YAC 11HE12. Thus the analysis of the genomic region surrounding D15S113 failed to identify any coding sequences.

Human STS UT7715 (L30335) was subsequently identified as a tetranucleotide microsatellite repeat marker from 15q11-13. UT7715 was localised near D15S113 during an intensive mapping project, in which STSs and ESTs from 15q11-13 were ordered with respect to one-another on a complete YAC map. UT7715 was assigned the locus number D15S540 (Christian *et al.*, 1998).

Chapter 4: RESULTS

Identification of mutations in the *UBE3A* gene in AS patients with no deletion, disomy or imprintor mutation of 15q11-13.

4.1 Screen for mutations of *UBE3A* in NDUI AS patients by Southern analysis

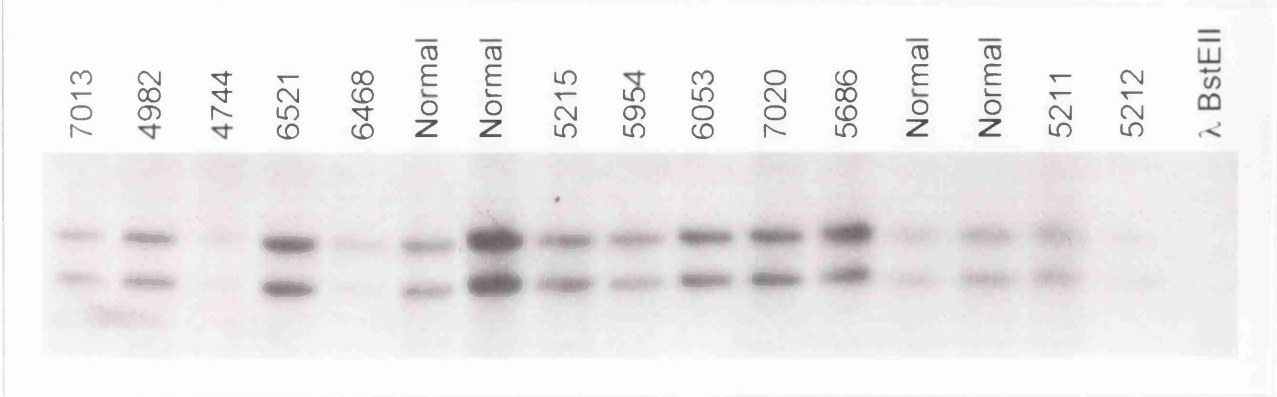
4.1.1 Southern analysis of NDUI AS patients using probe of UBE3A exons 4-8 (10-14).

After the *UBE3A* gene had been located to the ASCR it became a positional candidate for AS (section 3.1.2). NDUI AS patients were screened for *UBE3A* mutations and disruptions involving *UBE3A* by Southern analysis. Southern blots containing genomic DNA from 26 index cases of NDUI AS and normal controls digested with *EcoR*I, *BamH*I, *Hind*III, *Taq*I, *Msp*I, and *Not*I combined with *Xba*I, were hybridised with *UBE3A* exons 4-8 (10-14). No altered fragments were detected in the DNA of any of the NDUI patients with these enzymes, using this probe. The autoradiographs of the blots showing NDUI AS patient DNA digested with *BamH*I, *Taq*I, and *Not*I combined with *Xba*I, hybridised with *UBE3A* exons 10-14 are shown in Fig 4.1.1. Thus no mutations or alterations of *UBE3A* were identified using this method.

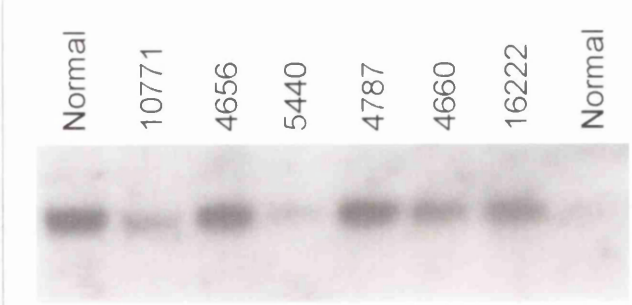
4.1.2 Southern analysis of NDUI AS patients with I.M.A.G.E. clone 143907 (215-B4).

I.M.A.G.E. clone 143907 (215-B4), which was the first clone to be identified through direct cDNA selection and sequence database searching (Table 3.2.3), also mapped to the ASCR and was a potential candidate for AS (Table 3.2.5). This clone was subsequently shown to be internal to clone 1354-G13 and to represent the 3'UTR of *UBE3A* (section 3.3.1.1). To screen for mutations in NDUI AS cases, Southern blots of DNA from 17 NDUI AS patients and controls digested with *Taq*I and *Msp*I (section 1.9.4.1), were hybridised with the insert from clone 215-B4 (Fig 4.1.2).

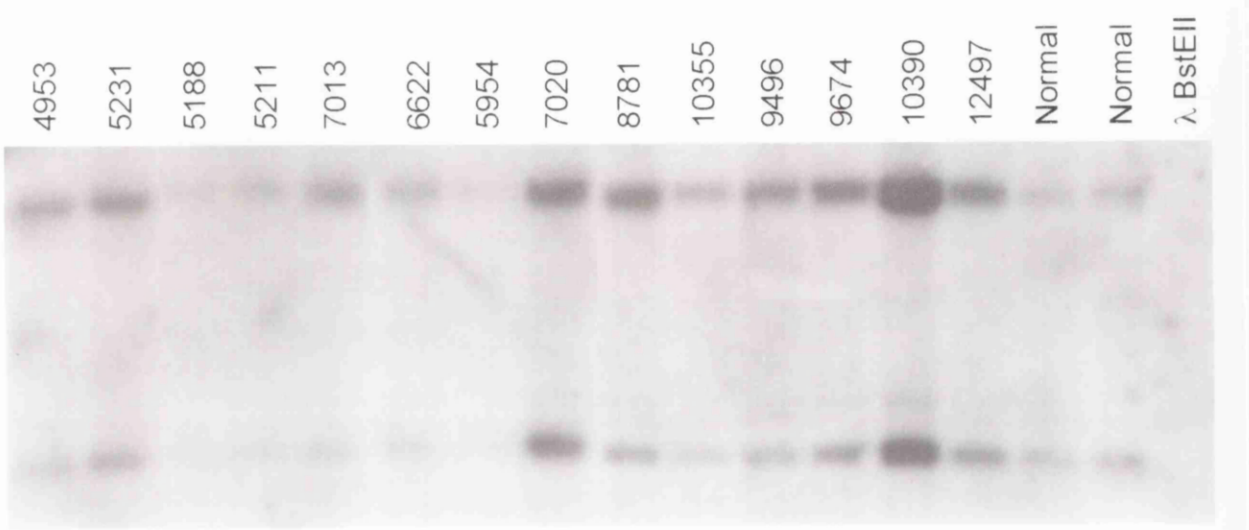
No alterations were identified by clone 215-B4 in patient DNA digested with *Msp*I (Fig 4.1.2A). 215-B4 identified two differently sized upper bands in the *Taq*I digested DNA (Fig 4.1.2B), indicative of a sequence change. The highest molecular weight band was identified in three AS patients, 5211, 9574 and 10390. This band was subsequently identified in normal controls at a frequency of 18% heterozygosity (Sarah Rickard, personal communication). This band therefore represents a *Taq*I RFLP identified by the 215-B4 probe. Thus no mutations were identified in any of the 17 NDUI AS patients by clone 215-B4, using these enzymes.



A: NDUI AS DNA digested with *BamH*I and probed with *UBE3A* exons 4-8.
Patients 5211, 5212 and 5215 are affected siblings from the same family and are recorded as one index case.



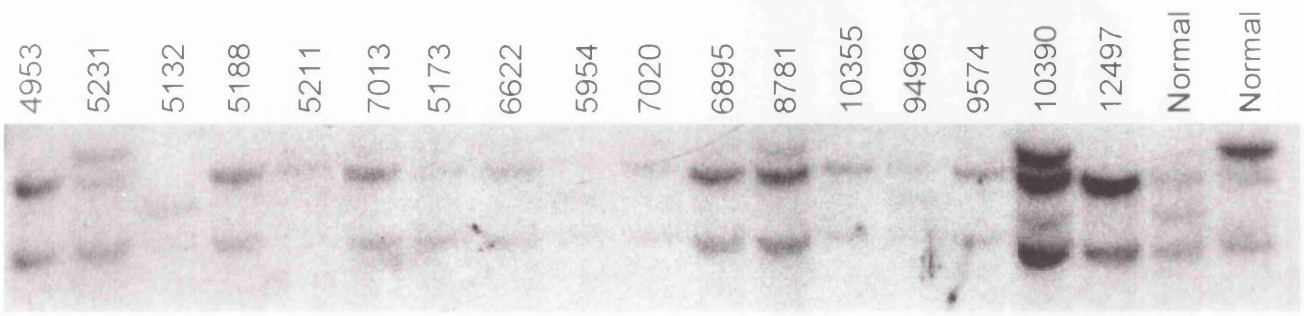
B: NDUI AS patient DNA digested with *Not*I combined with *Xba*I probed with *UBE3A* exons 4-8.
The *Not*I restriction enzyme is methylation sensitive, and so is capable of detecting mutations involving altered methylation, in addition to altered restriction length fragments.



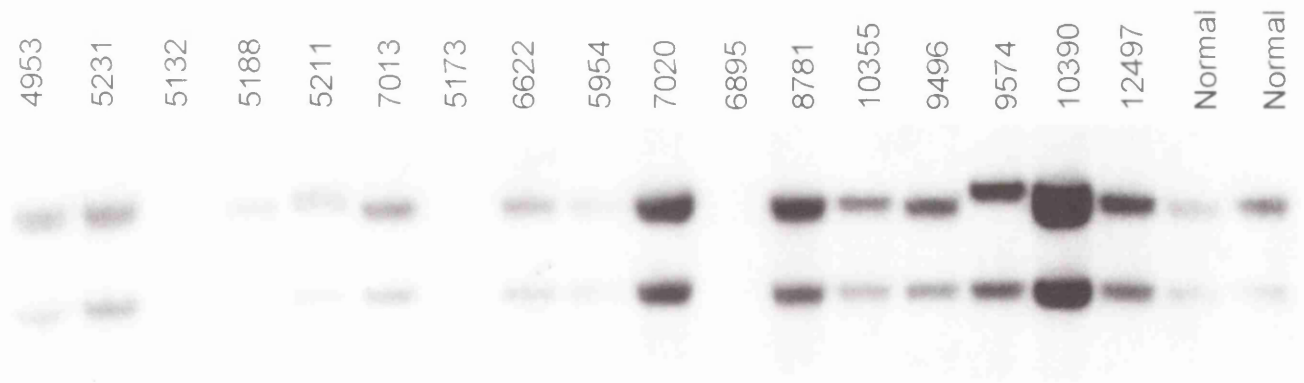
C: NDUI AS DNA digested with *Taq*I

Fig 4.1.2 Screen of NDUI AS patients for mutations of *UBE3A* by Southern analysis.

Autoradiographs of restriction-digested DNA from NDUI AS patients (as labelled), hybridised with *UBE3A* exons 4-8 (renumbered exons 4-8). The blots were washed at 1X SSC/ 0.15 SDS at 65°C for 20 minutes, and autoradiographed for 1-3 days. Southern blots were made by Dr. Jess Buxton. No altered fragments were identified for any of the NDUI AS cases analysed by this method.



A: Genomic DNA digested with *Msp*I.
Some incomplete digestion of the DNA is evident by the identification of higher molecular-weight bands of differing intensities.



B: Genomic DNA digested with *Taq*I.
A *Taq*I RFLP was identified by clone 215-B4.

<u>Fig 4.1.3</u> Mutation screen in NDUI AS patients with I.M.A.G.E. clone 143907 (215-B4)

Autoradiographs of restriction-digested NDUI AS DNA (as labelled), hybridised with the insert from I.M.A.G.E. clone 215-B4. The blots were washed to 0.5X SSC/ 0.1% SDS at 65°C for 20 minutes and autoradiographed for 1 week.

4.2 Screen for mutations of *UBE3A* in NDUI AS patients by SSCP and sequence analysis

Following two reports implicating *UBE3A* in the pathogenesis of AS, NDUI AS patients referred to the Institute of Child Health (I.C.H.) for genetic diagnosis were screened for point mutations of *UBE3A* by SSCP analysis followed by sequencing. In one report, a maternally inherited balanced paracentric inversion within 15q11-13 was identified in an AS child. Locus D15S10 lay between the inversion breakpoints. The centromeric inversion breakpoint was distal to the coding region of *UBE3A*, but proximal to the genomic clone OP2 (section 1.7.5.5). OP2 was found to contain the first exon of *UBE3A*, and exon trapping from a PAC clone spanning the inversion breakpoint led to the isolation of 5' untranslated *UBE3A* exons U1-U4. Thus the inversion disrupted the 5' end of the gene (Kishino *et al.*, 1997) (Fig 4.2.1). This prompted a search for mutations of *UBE3A* in AS patients for whom the genetic basis for disease had not been identified. SSCP analysis and sequencing of the coding exons of *UBE3A* from genomic DNA led to the identification of loss-of-function mutations in a small number of NDUI AS patients (Kishino *et al.*, 1997; Matsuura *et al.*, 1997).

The genomic organisation of *UBE3A* and the position of the paracentric inversion associated with AS is shown in Fig 4.2.1. The original and reassigned exon numbers for *UBE3A* are given. Exons are referred to by the new numbering system in the following sections.

In total 21 sporadic and 5 familial NDUI AS cases referred to I.C.H. were screened for disease-causing mutations of *UBE3A* by SSCP and sequence analysis. The familial cases were previously shown to have dominantly inherited AS, and the affected siblings in each family had the same maternal chromosome 15q11-13 haplotypes (Clayton-Smith *et al.*, 1992). A clinical diagnosis of AS was confirmed by Dr. Jill Clayton-Smith or Prof. Marcus Pembrey at the Great Ormond Street Hospital for Sick Children, according to the diagnostic criteria for AS listed in Table 1.1. This was a joint project, with contributions from Dr Una Fairbrother, Sarah Rickard and Dr Francesca Faravelli.

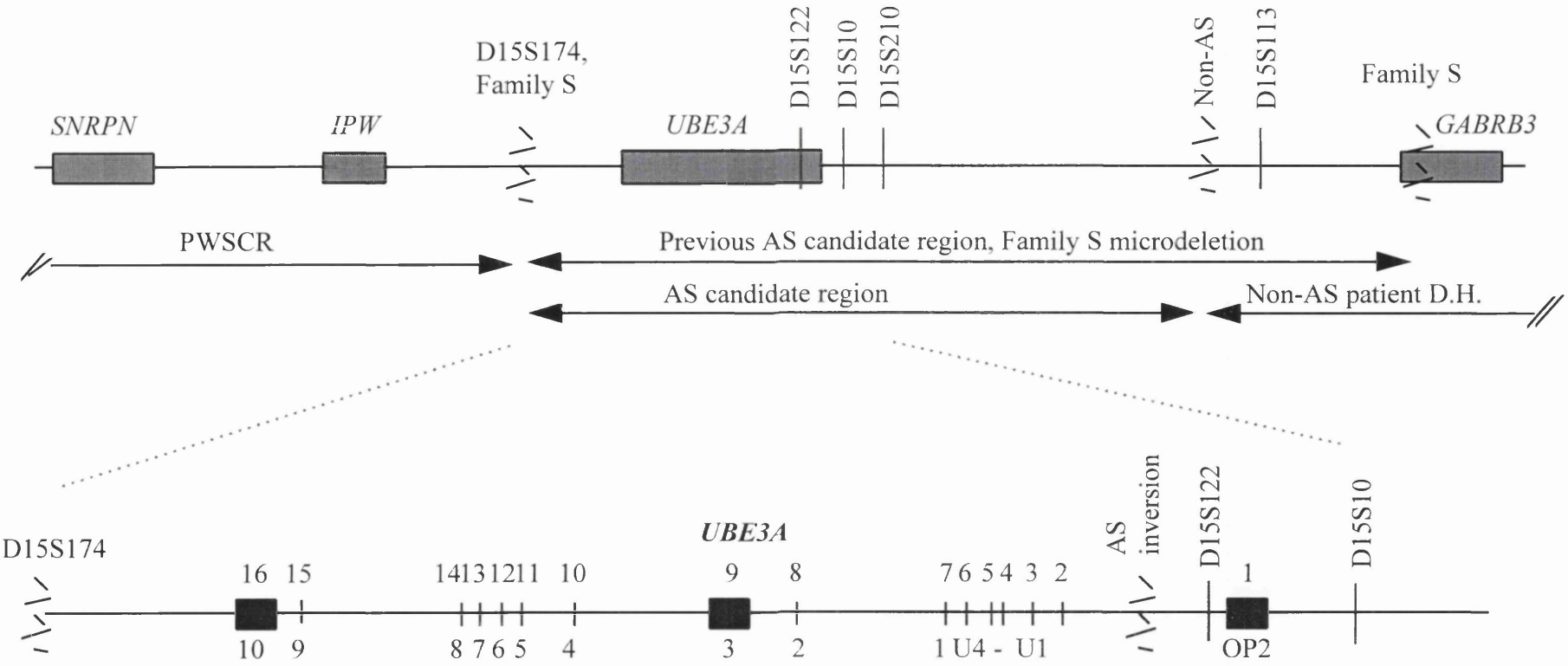


Fig 4.2.1 Genomic structure of the *UBE3A* gene contained within the Angelman syndrome candidate region (ASCR). A physical map of the AS region and distal portion of the PWS region are shown. Genes within the region are boxed in grey. The zig-zag lines denote the breakpoints in the AS Family S microdeletion and the non-AS maternal deletion, which define the PWS and AS regions. The PWSCR and the ASCR are indicated by arrowed lines. An expanded map of the region containing the *UBE3A* gene is given below. The position of the paracentric inversion in an AS child, which disrupted the 5' end of the *UBE3A* gene, is indicated by a zig-zag line. The structure of the *UBE3A* gene is given, with new exons numbers according to Kishino and Wagstaff, (1998) above, and old exon numbers (Huibregtse et al., 1993 and Kishino et al., 1997) below.

4.2.1 Mutation screening strategy

4.2.1.1 Primer design and UBE3A exons screened for mutations by SSCP analysis

The entire coding region of the *UBE3A* gene (exons 8-16), and the 5' UTR exons 1, 4, 5 and 6 were screened for mutations in 21 sporadic and 5 familial cases of NDUI AS by SSCP analysis followed by sequencing of genomic DNA. Exons 7 and 3 (U1) were not screened since they were only 42bp and 30bp in length respectively (Appendix). Exon 2 was identified at a later date, and so was not screened for mutations in this project (Kishino and Wagstaff, 1998). The section of the final exon analysed by SSCP was the former 156bp exon 10 sequence, which contained the final codons (Huibregtse *et al.*, 1993). The exon sequences are given in the Appendix.

The *UBE3A* exons screened and the primer sequences used to amplify the exons are given in Table 4.1. The individual exons within the *UBE3A* cDNA had been delineated (Huibregtse *et al.*, 1993), but the intron sequences were not known at the outset of this screening project. Therefore amplification across the intron-exon boundaries of *UBE3A* was not feasible initially. Primers were designed from within the *UBE3A* exon sequences to PCR amplify fragments of 150-400bp from genomic DNA for SSCP analysis. With the subsequent publication of intronic sequences, or primers used by other groups to screen for *UBE3A* mutations (as indicated in Table 4.1), SSCP and sequence analysis was performed using these.

New exon numbers	Former exon numbers	GenBank accession	Exon size (bp)	Primer pairs
1	OP2	AF009339	539	OP2F2 / OP2R3
4	U2	X98035	120	U2F #/ U2R
5	U3	X98036	113	723-A+/ 723-B+ (both intronic)
6	U4	X98037	196	U4F / U4R
8	2	X98022	299	2F1 / 2R1
9	3	X98023	1247	3F2 /3R2, 3F3 /3R3, 3F4 /3R4, 3F5 /3R5, E63-C+/ E63D+, E63-E +/ E63-F+, E63-G+/ E63-H+, E6-9C+/ E6-9D+
10	4	X98024	145	E610A*/E610B* (both intronic)
11	5	X98025	206	5F12 / 5R12
12	6	X98026	165	6F11 / 6R11
13	7	X98027	156	7F10 / 7R10
14	8	X98028	76	E6-14F^{+} / E6-14R^{+} (both intronic)
15	9	X98029	144	9F9 / 9R9
16	10	X98030	155	E6ZA* (Intron) / E6ZB*

Table 4.1 Primers used for SSCP analysis and sequencing of *UBE3A* in AS patients.

The *UBE3A* exons screened for mutations in NDUI AS patients are listed according to the new and previous exon numbering systems and the GenBank accession numbers are provided. The primers were designed within the exon sequences, except where stated. # Primers from Yamamoto *et al.*, (1997). * Primers from Kishino *et al.* (1997). $^{+}$ Primers from Malzac *et al.* (1998). The nucleotide sequences and annealing temperatures of the primer pairs listed are given in Table 2.11.

4.2.1.2 Strategy of SSCP and sequence analysis employed

SSCP analysis was performed on each exon listed (Table 4.1) for each of the AS probands and 10 normal controls. On identification of an SSCP shift in an AS patient, that was not present in the normal individuals analysed, SSCP analysis of the appropriate *UBE3A* fragment was performed for the entire family, including the AS patient(s), parents and any unaffected siblings. This strategy was employed for both familial and sporadic cases of AS.

In familial cases, where the SSCP shift followed the predicted pattern of disease segregation for AS (the shift was present in all affected siblings and the mother, but not in the father or unaffected siblings), it was considered likely the fragment contained a mutation. The fragment was analysed by sequencing in each family member to determine the nature of the sequence change. If an SSCP shift was present in the father or unaffected siblings, the fragment was thought likely to contain a benign polymorphism and was not sequenced.

In sporadic cases, a *de novo* shift not detectable in either parent was indicative of the presence of a mutation, and the fragment analysed by sequencing. To distinguish a pathogenic mutation from a benign polymorphism, SSCP analysis was performed in at least 50 normal controls (100 chromosomes). The mutation was considered pathogenic if no SSCP shift was detected in the normal controls using the same SSCP conditions. PCR fragments were either cloned in the pTAg vector for sequencing of the individual alleles, or were sequenced directly.

The SSCP and sequencing data are presented here for four AS cases in which SSCP shifts and sequence changes were identified in *UBE3A* exon 9, including sporadic patients 5954 and 7020, and Families H and W. In this chapter data are separated into two sections. The SSCP data are presented first, followed by the identification and anlysis of sequence changes. In the joint project with Dr. Una Fairbrother and S. Rickard, SSCP shifts and loss-of-function mutations of *UBE3*A were identified in 7 NDUI AS index cases including 4 of the 21 sporadic patients and 3 of the 5 families. These included sporadic patients 5954, 5173, 18244 and Patient O, and Families H, W and A. The nature of the mutations identified, and *UBE3A* exons they were identified in, are summarised in Table 4.2.

4.2.2 Identification of SSCP shifts in *UBE3A* in NDUI AS patients

SSCP shifts which fitted the pattern indicative of a mutation causing AS and were not observed in normal controls, were identified in 8 index cases. The SSCP data for AS Families H and W, and sporadic cases 7020 and 5954 are shown in this section.

4.2.2.1 Identification of a de novo SSCP shift in sporadic patient 5954

An SSCP shift was observed in exon 9 of patient 5954, using primers 3F5 and 3R5, but was not seen in either parent (Fig 4.2.2). This was indicative of a *de novo* mutation in this patient. This SSCP shift was not observed in 60 normal individuals using these primers. No other SSCP shifts were detected in patient 5954 in any of the other *UBE3A* exons screened (Table 4.1). This fragment was cloned and sequence analysis performed for the patient and both parents.

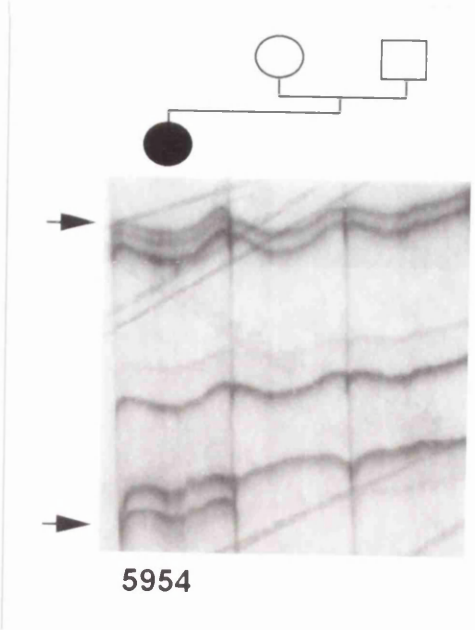


Fig 4.2.2 Detection of a *de novo* SSCP shift in *UBE3A* exon 9 in sporadic patient 5954

The SSCP shift observed in PCR fragment 3F5 and 3R5 within exon 9 of *UBE3A* is shown for patient 5954. This was not observed in either parent. PCR products were electrophoresed on MDE gel overnight at 4°C and 15W, and visualised by silver-staining.

4.2.2.2 *Identification of an SSCP shift in UBE3A exon 9 in sporadic AS patient 7020*

An SSCP shift was detected in exon 9 of *UBE3A* in patient sporadic AS patient 7020 using primers 3F5 and 3R5 (Table 4.1). The shift was not present in either parent, indicating a *de novo* sequence change in the proband (Fig 4.2.3). No further SSCP shifts were detected in this patient in any of the other *UBE3A* exons tested (Table 4.1). No SSCP shift was detected in 60 anonymous normal control individuals (120 normal chromosomes), using the same primer pair, indicating that the change was likely to be pathogenic in patient 7020. The 3F5-3R5 PCR fragment for patient 7020 was cloned and analysed by sequencing.

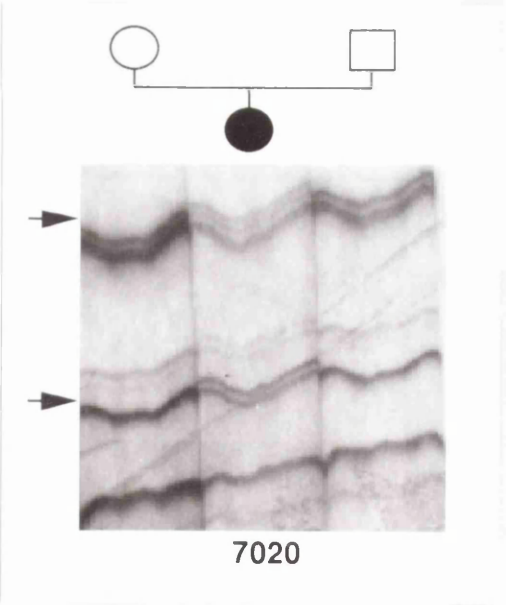


Fig 4.2.3 Detection of a *de novo* SSCP shift in *UBE3A* in sporadic AS patient 7020

The SSCP shift, detected in exon 9 of patient 7020 using primers 3F5-3R5, is shown. This was not present in either parent. The PCR products were electrophoresed overnight at 4°C on an MDE gel, at 15W. The products were detected by silver-staining, and the gel dried.

4.2.2.3 Identification of an SSCP shift in AS Family H.

An SSCP shift was identified in exon 9 with primers 3F5 and 3R5 in the two affected children in Family H (patients 5187 and 5188) and the mother, but not in the father (Fig 4.2.4). No SSCP shifts were detected with these primers in 60 normal individuals, indicating that the sequence change was likely to be disease-causing. This indicated the presence of a maternally transmitted mutation fragment 3F5-3R5 of exon 9 in this family. The PCR products from each family member were cloned for sequence analysis of the individual alleles.

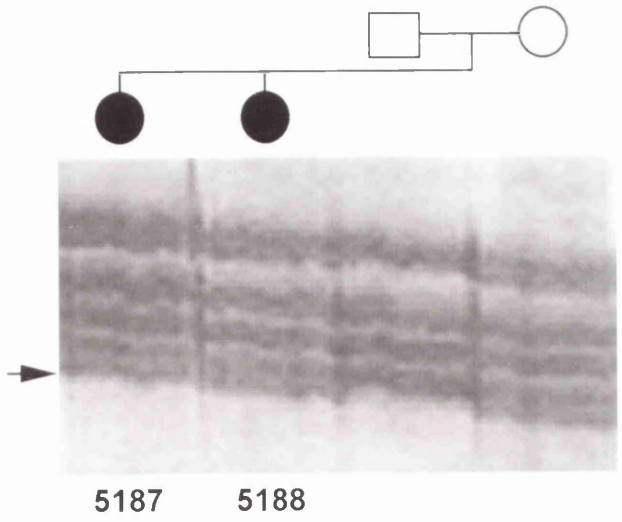


Fig 4.2.4 SSCP shift on the maternal allele of *UBE3A* exon 9 in AS Family H

The 3F5-3R5 PCR products for Family H, run on an MDE gel overnight at 4°C and 15W, is shown. The gel was silver-stained to detect the DNA, and dried.

4.2.2.4 Identification of an SSCP shift in Family W

An SSCP shift was detected in exon 9 of *UBE3A* using primers 3F3 and 3R3 in the three affected siblings and their phenotypically normal mother in Family W. The shift was not observed in the paternal DNA. This indicated the presence of a maternally inherited sequence change in this fragment (Fig 4.2.5). This shift was not identified in the genomic DNA of 150 normal individuals. No other SSCP changes were identified in Family W in any other exons of *UBE3A*. The 3F3-3R3 fragment for each member of Family W was cloned and analysed by sequencing.

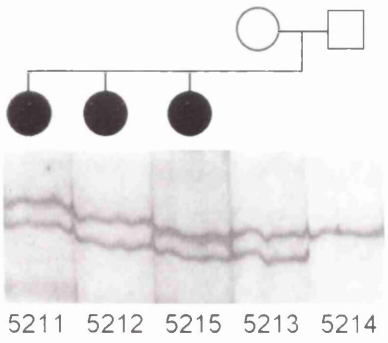


Fig 4.2.5 SSCP analysis of AS Family W

The SSCP shift observed in *UBE3A* exon 9 in three affected siblings (5211, 5212 and 5215), and the unaffected mother (5213) is shown. This was not present in the father (5214). The fragment was amplified from genomic DNA using primers 3F3 and 3R3, from within exon 9. PCR products were electrophoresed overnight on an MDE gel at 4°C and 15W. The DNA was visualised by silver-staining.

4.2.3 Identification of pathogenic mutations of *UBE3A* and contaminating pseudogene sequences by sequence analysis of NDUI AS patients with SSCP shifts

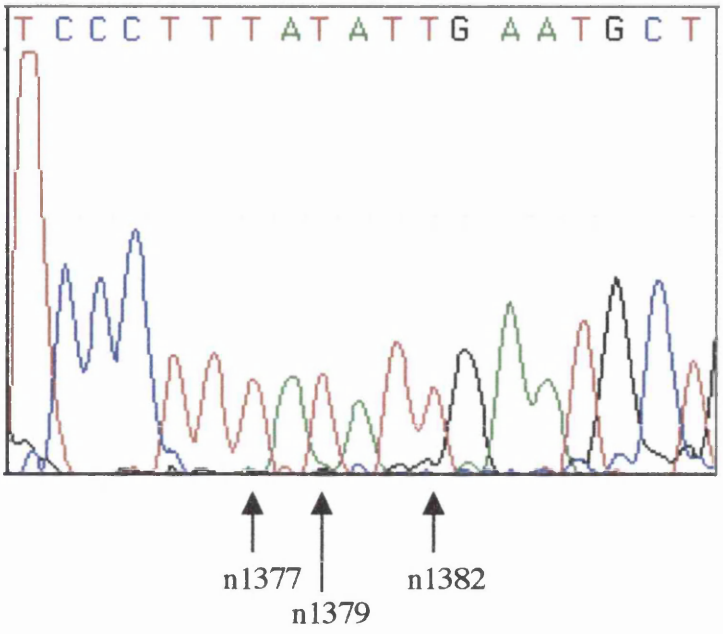
Two processed pseudogenes *UBE3AP1* (GenBank AF10598) and *UBE3AP2* (GenBank AF10599), which share 90-95% sequence similarity to *UBE3A*, were reported subsequent to SSCP analysis and sequencing of the AS patients presented here. *UBE3AP1* demonstrates significant sequence identity to *UBE3A* exons 3, 4, 6 and 7, and *UBE3AP2* shares sequence homology to exons 9-16 (Kishino and Wagstaff, 1998). The presence of the pseudogenes proved, in some cases, to be a confounding factor in the identification of pathogenic mutations of *UBE3A* by sequence analysis.

4.2.3.1 Identification of a frameshift mutation in sporadic patient 5954

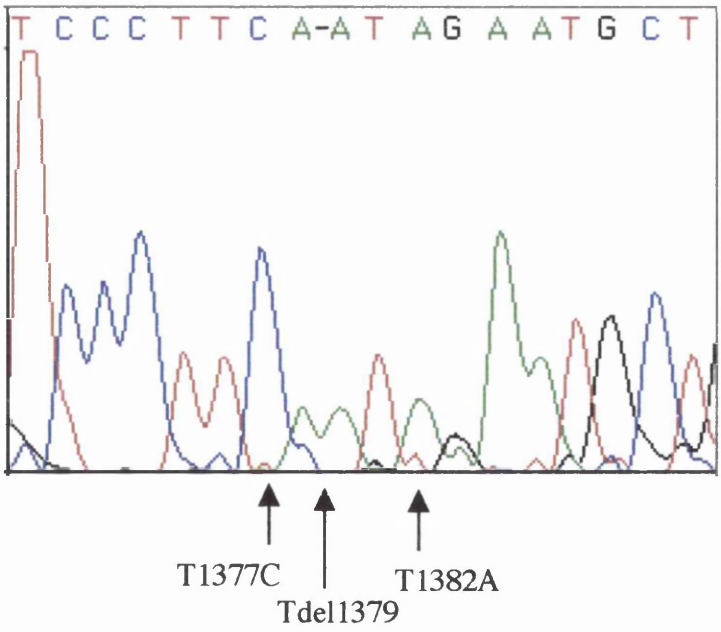
The 3F5-3R5 PCR product which had demonstrated an SSCP shift in patient 5954 (section 2.2.2.1; Fig 4.2.2) was cloned and eight subclones sequenced. A wild-type allele, which did not vary from the published *UBE3A* sequence (GenBank X98032) was sequenced from four separate subclones (Fig 4.2.6). A mutant allele, represented by three separate subclones, was identified which differed from the normal sequence at three nucleotide positions: a T→C transition at nucleotide 1377, a deletion of a T at nucleotide position 1379, and a T→A transversion at nucleotide 1382 (Fig 4.2.6; 4.2.7). Sequence analysis of patient 5954 using primers E6-9C and E6-9D (Malzac *et al.*, 1998), confirmed the three nucleotide changes (Sarah Rickard, personal communication).

The T1377C substitution did not alter an amino acid. The 1379delT mutation was predicted to cause a frameshift leading to the introduction of a stop signal downstream at codon 490 (with respect to *UBE3A* isoform I, X98032). This would result in premature chain termination at amino acid 490 (Fig.4.2.7). This would destroy the *hect* domain at the caroboxy terminus of all three E6-AP protein isoforms, and abolish the ubiquitination activity of E6-AP. The PCR products from the proband's parents were also subcloned and sequenced, but no alterations were identified, consistent with the SSCP data. The inactivating mutation in patient 5954 thus arose *de novo*, providing strong evidence that it is the disease-causing mutation in this patient.

Wild-type allele (subclones 2, 7, 8, 9)

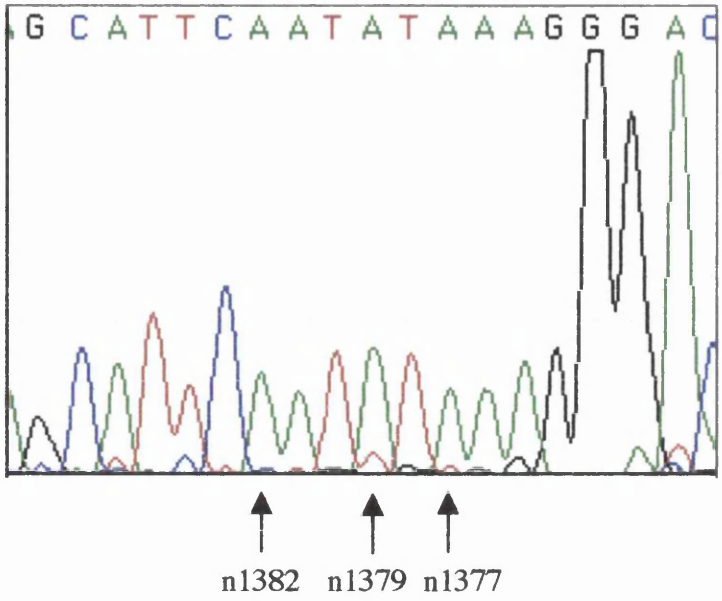


Mutant allele (subclones 1, 10, 12)



A: Sequences for AS patient 5954 in the forward direction

B: **Wild-type allele for 5954** (subclones 2, 7, 8, 9)



Mutant allelefor 5954 (subclones 1, 10, 12)

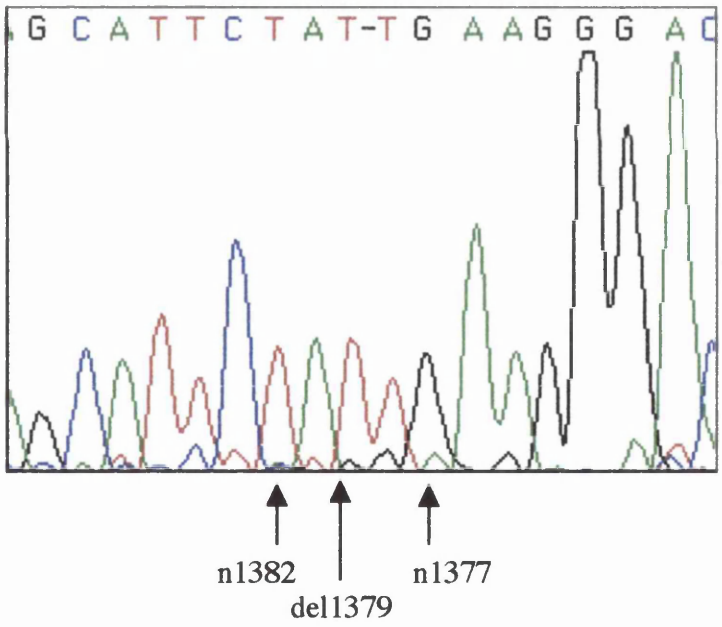


Fig 4.2.6 Sequence changes causing a frameshift mutation of *UBE3A* in sporadic AS patient 5954

Sequence electropherograms from subclones of *UBE3A* fragment F5-3R5 in patient 5954 are shown in the forward (**A**), and reverse-complement (**B**) directions. The wild-type allele was the same as the published *UBE3A* sequence (GenBank X98032). A deletion of a T at nucleotide (n) position 1379, and 2 substitutions, a T to C at n1377, and a T to A at n1382 were identified in the mutant allele, as indicted.

A: Wild-type *UBE3A* allele

```
1321  TATACTTTTTTCAAAGTAGAAACAGAGAACAAATTCTCTTTTATGACATGTCCCTTTATA
 441  Y  T  F  F  K  V  E  T  E  N  K  F  S  F  M  T  C  P  F  I

1381  TTGAATGCTGTCACAAAGAATTTGGGATTATATTATGACAATAGAATTCGCATGTACAGT
 461  L  N  A  V  T  K  N  L  G  L  Y  Y  D  N  R  I  R  M  Y  S

1441  GAACGAAGAATCACTGTTCTCTACAGCTTAGTTCAAGGACAGCAGTTGAATCCATATTTG
 481  E  R  R  I  T  V  L  Y  S  L  V  Q  G  Q  Q  L  N  P  Y  L
```

B: Frameshift mutation in AS Patient 5954

```
                                                    T1377C Tdel1379
                                                          ↓ ↓
1321 TATACTTTTTTCAAAGTAGAAACAGAGAACAAATTCTCTTTTATGACATGTCCCTTCA-AT
 441 Y  T  F  F  K  V  E  T  E  N  K  F  S  F  M  T  C  P  F  N

     T1382A
       ↓
1381 AGAATGCTGTCACAAAGAATTTGGGATTATATTATGACAATAGAATTCGCATGTACAGTG
 461 R  M  L  S  Q  R  I  W  D  Y  I  M  T  I  E  F  A  C  T  V

1441 AACGAAGAATCACTGTTCTCTACAGCTTAGTTCAAGGACAGCAGTTGAATCCATATTTGA
 481 N  E  E  S  L  F  S  T  A  *  F  K  D  S  S  *  I  H  I  *
                                  ↑
```

Chain termination at amino acid 490

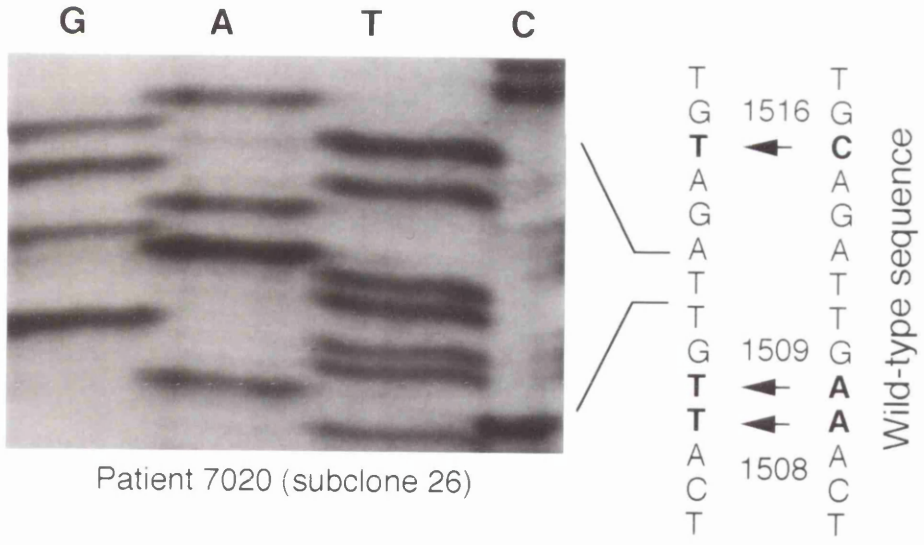
Fig 4.2.7 Predicted effect of the *UBE3A* frameshift mutation in patient 5954 on the amino acid sequence of E6-AP.

The nucleotide and amino acid sequence of *UBE3A* in the region surrounding the mutations identified in exon 9 in AS patient 5954 are shown. **A:** The wild-type allele is shown, with nucleotide and amino acids according to the *UBE3A* isoform I sequence (GenBank X98032). **B:** The mutant allele observed in 5954 is shown, with predicted amino acid changes in the patient in bold. The T1377C substitution does not result in any amino acid change. The T deletion causing the frameshift error is labelled and indicated by a dash (-). The stop codons are indicated in the amino acid sequence by stars (*).

4.2.3.2 Identification of nucleotide substitutions in sporadic AS patient 7020.

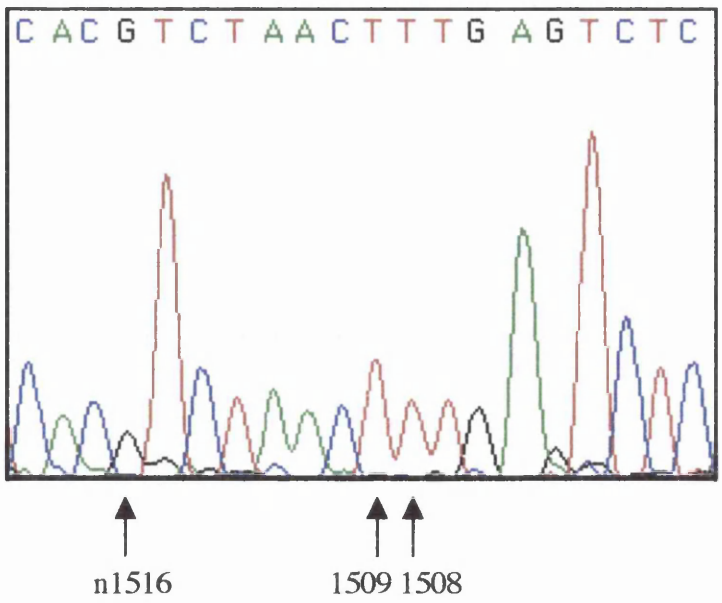
Six subclones containing the 3F5-3R5 fragment from patient 7020 (Fig 4.2.3) were sequenced in both the forward and reverse directions. In three subclones, 22, 23 and 24, the sequence did not differ from the published sequence for *UBE3A* exon 9 (X98023). For three other subclones, 20, 21 and 26, three nucleotide substitutions were identified with respect to *UBE3A*. The sequence changes in the variant allele consisted of two consecutive A→T transversions at nucleotide positions 1508 and 1509, and a C→T transition at nucleotide 1516, with respect to the *UBE3A* isoform I nucleotide sequence (Fig 4.2.8).

Translation of the variant sequence for 7020 predicted that the nucleotide substitutions would result in two missense mutations in *UBE3A,* if these changes were present in the gene itself (Fig 4.2.9A). In the light of future findings, it was postulated that the single base-pair changes were present in a pseudogene sequence, and were not causative of AS in this patient. Direct comparison of the variant sequence for patient 7020 with the recently published *UBE3A* pseudogene sequences showed that the three nucleotide substitutions were present in the sequence of the *UBE3AP2* pseudogene (Fig 4.2.9B). No similarity was observed between the variant allele and *UBE3AP1*. This suggested that the sequence changes in the variant allele were due to amplification of the *UBE3AP2* sequence and were not genuine missense mutations of *UBE3A*.

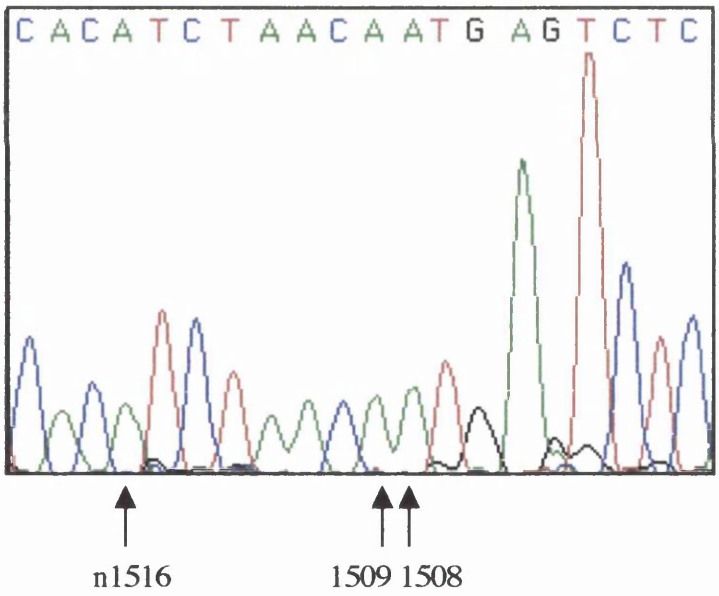


A: Manual sequence of the mutant allele (subclone 26) for patient 7020, within PCR fragment 3F5-3R5, is shown in the forward direction. The normal *UBE3A* sequence (X98023) is given on the left hand side of the patient sequence for comparison. Three nucleotide substitutions were observed: A1508T, A1509T, and C1516T (Sequencing by S. Rickard).

Wild-type allele (subclones 22, 23, 24)



Mutant allele (subclones 20, 21 and 26)



B: The automated sequence electropherograms for wild-type allele and variant allele for patient 7020 are shown in the reverse complement orientation. The positions of the nucleotide (n) changes with respect to *UBE3A* (GenBank X98032), A1058T, A1059T and C1516T, are indicated directly beneath the electropherograms.

Fig 4.2.8 Sequence changes identified in sporadic AS patient 7020

```
Normal   1501 AGACTCAAAGTTAGACGTGACCATATCATAGATGATGCACTTGTCCGGCTAGAGATGATC
         501  R  L  K  V  R  R  D  H  I  I  D  D  A  L  V  R  L  E  M  I

Variant  1501 AGACTCATTGTTAGATGTGACCATATCATAGATGATGCACTTGTCCGGCTAGAGATGATC
         501  R  L  I  V  R  C  D  H  I  I  D  D  A  L  V  R  L  E  M  I
                   ↑        ↑
                 K503I    R506C
```

A: Predicted translation products of the normal and variant alleles of AS patient 7020 are shown. The A1508T and A1509T nucleotide substitutions in 7020 predict a lysine to isoleucine substitution at amino acid 503, and the C1516T change results in a arginine to cysteine substitution at amino acid 506 (indicated by arrows).

```
                                    ↓↓      ↓
  394 GCAGTTGAATCCATATTTGAGACTCATTGTTAGATGTGACCATATCATAGATGATGCACTTGTC 433
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 1249 GCAGTTGAATCCATATTTGAGACTCATTGTTAGATGTGACCATATCATAGATGATGCACTTGTC 1299
```

B: GAP output of the direct sequence comparison between the variant allele in patient 7020 (top) and the *UBE3AP2* pseudogene (GenBank AF010599) (bottom). Arrows indicate the positions of the three nucleotide substitutions identified on the variant allele in patient 7020. These changes are also present in *UBE3AP2*.

Fig 4.2.9 Sequence analysis of the variant allele in patient 7020 suggests that it represents the *UBE3AP2* pseudogene sequence.

4.2.3.3 Sequence analysis of UBE3A exon 9 in Family H identifies a pathogenic frameshift mutation and the UBE3AP2 pseudogene sequence.

Sequence analysis was performed for each member of Family H on the cloned 3F5-3R5 PCR fragment, which had demonstrated an SSCP shift in the mother and two affected siblings (section 2.2.1.3; Fig 4.2.2). Several subclones for each family member were sequenced. On the basis of the sequence data, three distinct 'alleles' were identified in the mother and two affected children, and two alleles were identified in the father. One allele present in each family member was the wild-type *UBE3A* sequence (X98032) (Fig 4.2.10A). A second 'variant allele', which was also identified in all family members, had six nucleotide substitutions with respect to the *UBE3A* sequence (X98032) as follows: A1110G, A1114G, C1146T, A1164G, A1172A and A1179G (Fig 4.2.10B). The third 'mutant' allele was identified in the mother and both affected children, and not in the father (Fig 4.2.10C).

The mutant allele showed a heterozygous 5bp deletion (ATTAT) of nucleotides 1407-1411 with respect to *UBE3A* isoform I (GenBank X98032). The 5bp deletion was initially identified in the mother and two AS children amongst the subclones of fragment 3F5-3R5 amplified from genomic DNA. This was subsequently confirmed in the lymphoblast cDNAs of the mother and both affected children, by RT-PCR amplification of cDNAs with primers 3F5 and 3R5, followed by sequence analysis (Fig 4.2.11). The mutation was also confirmed in the genomic DNA of the mother and affected children using primers E6-9C and E6-9D (Table 4.1), published by Malzac *et al.*, 1998 (Sarah Rickard, personal communication).

The 5bp deletion was predicted to cause a frameshift mutation resulting in premature chain termination at amino acid 473 with respect to E6-AP isoform I (Fig 4.2.12). This would abolish the *hect* domain and the amino acid residues at the carboxy termini that are vital for ubiquitin thioester formation and transfer of ubiquitin to p53, in all three E6-AP isoforms derived from the maternal allele. This mutation causes loss of function of E6-AP and was shown to be maternally transmitted, providing strong evidence that it is causative of AS in Family H.

```
1081 GTGGGAGGGGAAGTGGACACAAATCACAATGAAGAAGATGATGAAGAGCCCATCCCTGAG
1141 TCCAGCGAGCTGACACTTCAGGAACTTTTGGGAGAAGAAAGAAGAAACAAGAAAGGTCCT
1201 CGAGTGGACCCCCTGGAAACTGAACTTGGTGTTAAAACCCTGGATTGTCGAAAACCACTT
1261 ATCCCTTTTGAAGAGTTTATTAATGAACCACTGAATGAGGTTCTAGAAATGGATAAAGAT
1321 TATACTTTTTTCAAAGTAGAAACAGAGAACAAATTCTCTTTTATGACATGTCCCTTTATA
1381 TTGAATGCTGTCACAAAGAATTTGGGATTATATTATGACAATAGAATTCGCATGTACAGT
```

A: Wild-type allele, present in mother, father and two affected siblings

```
                                    A1110G A1114G
                                       ↓   ↓
1081 GTGGGAGGGGAAGTGGACACAAATCACAGTGAGGAAGATGATGAAGAGCCCATCCCTGAG
        C1146T              A1164G    G1172A  A1179G
          ↓                   ↓         ↓       ↓
1141 TCCAGTGAGCTGACACTTCAGGAGCTTTTGGAAGAAGAGAGAAGAAACAAGAAAGGTCCT
1201 CGAGTGGACCCCCTGGAAACTGAACTTGGTGTTAAAACCCTGGATTGTCGAAAACCACTT
1261 ATCCCTTTTGAAGAGTTTATTAATGAACCACTGAATGAGGTTCTAGAAATGGATAAAGAT
1321 TATACTTTTTTCAAAGTAGAAACAGAGAACAAATTCTCTTTTATGACATGTCCCTTTATA
1381 TTGAATGCTGTCACAAAGAATTTGGGATTATATTATGACAATAGAATTCGCATGTACAGT
```

B: 'Variant' allele present in mother, father and two affected siblings

```
1081 GTGGGAGGGGAAGTGGACACAAATCACAATGAAGAAGATGATGAAGAGCCCATCCCTGAG
1141 TCCAGCGAGCTGACACTTCAGGAACTTTTGGGAGAAGAAAGAAGAAACAAGAAAGGTCCT
1201 CGAGTGGACCCCCTGGAAACTGAACTTGGTGTTAAAACCCTGGATTGTCGAAAACCACTT
1261 ATCCCTTTTGAAGAGTTTATTAATGAACCACTGAATGAGGTTCTAGAAATGGATAAAGAT
1321 TATACTTTTTTCAAAGTAGAAACAGAGAACAAATTCTCTTTTATGACATGTCCCTTTATA
                                  1407delATTAT
1381 TTGAATGCTGTCACAAAGAATTTGGG-----ATTATGACAATAGAATTCGCATGTACAGT
```

C: 'Mutant' allele identified in the mother and two affected children

Fig 4.2.10 Nucleotide sequences of the three different alleles identified in members of AS Family H.

The nucleotides are numbered with respect to the *UBE3A* isoformI sequence (GenBank X98032). Nucleotide alterations from the wild-type sequence are given in bold. A dashed line (-) indicates a nucleotide deletion.

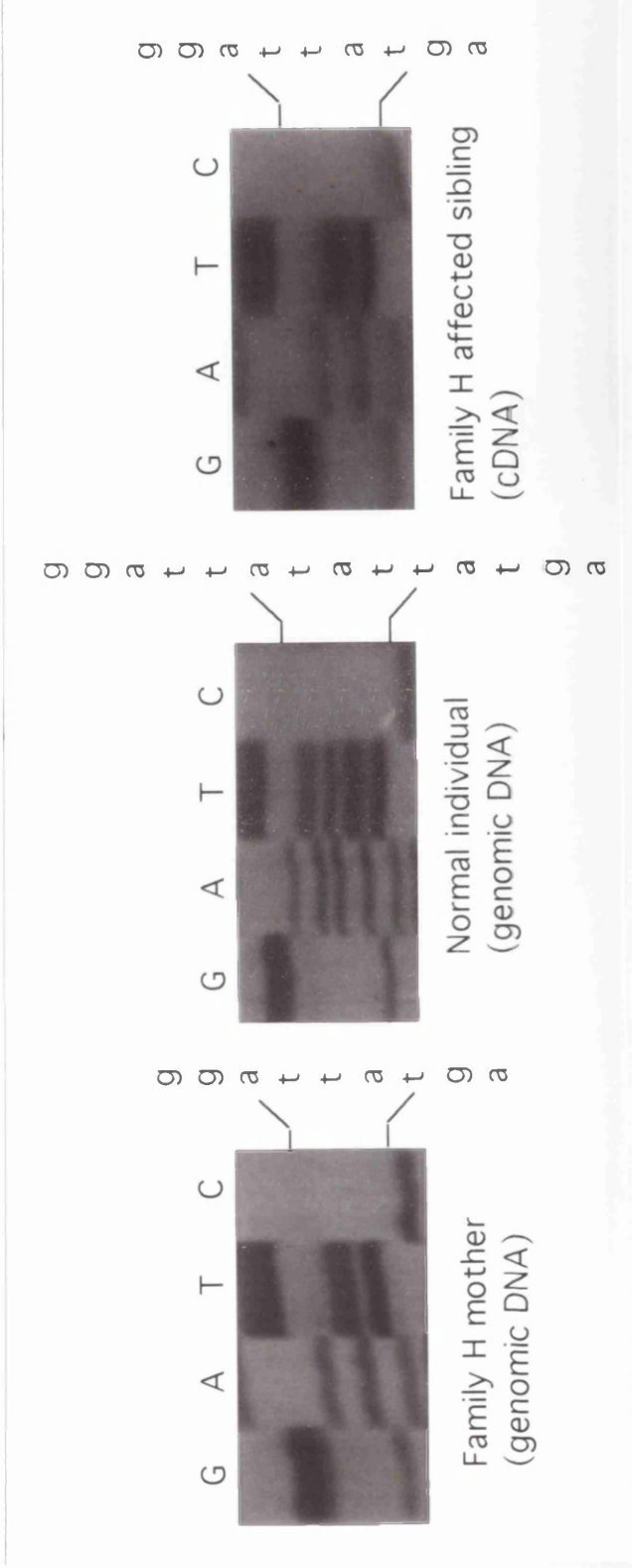


Fig 4.2.11 Frameshift mutation caused by a 5bp deletion on the maternal allele in AS Family H

The manual sequence for the mother and one of the affected children from Family H is given, showing the 5bp deletion in *UBE3A* exon 9, compared to a normal control. The mutation was sequenced from subclones of the amplification product using primers 3F5 and 3R5 that identified an SSCP shift, and from a cDNA template, as indicated in parentheses. The sequencing was performed, and diagram prepared by S. Rickard and U. Fairbrother.

A: Wild-type *UBE3A* and E6-AP sequence

```
1321 TATACTTTTTTCAAAGTAGAAACAGAGAACAAATTCTCTTTTATGACATGTCCCTTTATA
 441 Y  T  F  F  K  V  E  T  E  N  K  F  S  F  M  T  C  P  F  I

1381 TTGAATGCTGTCACAAAGAATTTGGGATTATATTATGACAATAGAATTCGCATGTACAGT
 461 L  N  A  V  T  K  N  L  G  L  Y  Y  D  N  R  I  R  M  Y  S
```

B: Mutant allele

```
1321 TATACTTTTTTCAAAGTAGAAACAGAGAACAAATTCTCTTTTATGACATGTCCCTTTATA
 441 Y  T  F  F  K  V  E  T  E  N  K  F  S  F  M  T  C  P  F  I

                                  5bp deletion
1381 TTGAATGCTGTCACAAAGAATTTGGG-----ATTATGACAATAGAATTCGCATGTACAGT
 461 L  N  A  V  T  K  N  L  G       L  V  E  *
                                                  ↑
                                  Premature chain termination
```

Fig 4.2.12 Predicted effect of the frameshift mutation in AS Family H on E6-AP.

A: The amino acid translation product for the wild-type *UBE3A* sequence (isoform I, GenBank X98032).
B: The predicted amino acid sequence for the maternally inherited mutant allele in Family H. The nucleotide and amino acid numbers are given with respect to *UBE3A* / E6-AP isoform I. The deleted nucleotides in the mutant allele are denoted by dashes (-), and are underlined in the wild-type sequence. The stop codon is indicated by an asterix (*). Where the amino acid sequence differs from the *UBE3A* sequence, the residues are in bold. The 5bp deletion results in truncation of the amino acid sequence at position 473 of E6-AP.

The second 'variant' allele was predicted to represent a pseudogene sequence due to the presence of multiple (six) disparate nucleotide substitutions within the sequence, and because they were observed in each family member showing no pattern of disease segregation. Furthermore, the predicted amino-acid translation product of the 'variant' allele did not demonstrate any obvious detrimental effect on the E6-AP protein, were this sequence part of *UBE3A*. The A1110G and the G1172A changes were predicted to cause missense mutations if present in the *UBE3A* sequence, but the remaining four substitutions were conservative (Fig 4.2.13A). With the identification of the *UBE3A* pseudogenes *UBE3AP1* and *UBE3AP2*, the variant allele in Family H was directly compared to both pseudogene sequences. No homology was with *UBE3AP1*. However, the variant allele showed strong

sequence homology to *UBE3AP2*. Each of the six nucleotide changes detected in the variant allele were also present in the *UBE3AP2* pseudogene sequence (Fig 4.2.13B). This indicated that this 'allele' represented an artefact due to co-amplification of the *UBE3AP2* pseudogene sequence with *UBE3A* exon 9, when primers 3F5 and 3R5 were used.

```
x98032 1081 GTGGGAGGGGAAGTGGACACAAATCACAATGAAGAAGATGATGAAGAGCCCATCCCTGAG
        361 V  G  G  E  V  D  T  N  H  N  E  E  D  D  E  E  P  I  P  E
                                   A1110G↓    ↓A1114G
Fam. H 1081 GTGGGAGGGGAAGTGGACACAAATCACAGTGAGGAAGATGATGAAGAGCCCATCCCTGAG
        361 V  G  G  E  V  D  T  N  H  S  E  E  D  D  E  E  P  I  P  E

x98032 1141 TCCAGCGAGCTGACACTTCAGGAACTTTTGGGAGAAGAAAGAAGAAACAAGAAAGGTCCT
        381 S  S  E  L  T  L  Q  E  L  L  G  E  E  R  R  N  K  K  G  P
          C1146T↓            A1164G↓ G1172A↓      ↓A1179G
Fam. H 1141 TCCAGTGAGCTGACACTTCAGGAGCTTTTGGAAGAAGAGAGAAGAAACAAGAAAGGTCCT
        381 S  S  E  L  T  L  Q  E  L  L  E  E  E  R  R  N  K  K  G  P
```

A: The six nucleotide substitutions observed in the variant allele of Family H and predicted effect on E6-AP are shown with respect to the *UBE3A*/ E6-AP isoform I sequence (X98032).

Family H x *UBE3AP2*

```
                                                  A1110G  A1114G
    1 ..................GGAAGTGGACACAAATCACAGTGAGGAAGATGA 33
                        ||||||||||||||||||||||||||||||||||
  851 CAAATGGTAGTGGGAAGGGAAGTGGACACAAATCACAGTGAGGAAGATGA 900
                                C1146T              A1164G
   34 TGAAGAGCCCATCCCTGAGTCCAGTGAGCTGACACTTCAGGAGCTTTTGG 83
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  901 TGAAGAGCCCATCCCTGAGTCCAGTGAGCTGACACTTCAGGAGCTTTTGG 950
    G1172A A1179G
   84 AAGAAGAGAGAAGAAACAAGAAAGGTCCTCGAGTGGACCCCCTGGAAACT 133
      |||||||||||||||||||||||||| |||||||||||||||||| |
  951 AAGAAGAGAGAAGAAACAAGAAAGGTCCTTGAGTGGACCCCCTGGAAAAT 1000
```

B: GAP output of the direct comparison between the Family H variant allele (top) and the *UBE3AP2* pseudogene sequence (AF010599) (bottom). The six nucleotide substitutions observed in Family H are also present in the pseudogene sequence.

Fig 4.2.13 The 'variant' allele in Family H represents the *UBE3AP2* pseudogene sequence.

The demonstration that sequence alterations identified in AS patient 7020 and Family H were also present in the *UBE3AP2* pseudogene sequence showed that primers 3F5 and 3R5 could not be used to specifically amplify *UBE3A* exon 9 from genomic DNA. This was confirmed by the demonstration of 100% sequence identity between the published *UBE3A* sequence and *UBE3AP2* at the primer-binding sites for 3F5 and 3R5 (Fig 4.2.14). The similarity between *UBE3A* exon 9 and *UBE3AP2* meant that amplification of genomic DNA using these primers would have efficiently amplified both sequences to give similarly sized products. These primers were therefore not suitable for mutation screening of *UBE3A* exon 9 using genomic DNA as template.

***UBE3A* (x98032) x *UBE3AP2* (AF010599)**

A: **3F5 primer-binding site**

```
   1 ..................GGAAGTGGACACAAATCACAATGAAGAAGATGA 33
                       ||||||||||||||||||| ||| ||||||||
 851 CAAATGGTAGTGGGAAGGGAAGTGGACACAAATCACAGTGAGGAAGATGA 900

  34 TGAAGAGCCCATCCCTGAGTCCAGCGAGCTGACACTTCAGGAACTTTTGG 83
     ||||||||||||||||||||||| |||||||||||||||| |||||||
 901 TGAAGAGCCCATCCCTGAGTCCAGTGAGCTGACACTTCAGGAGCTTTTGG 950
```

B:

```
 384 TTCAAGGACAGCAGTTGAATCCATATTTGAGACTCAAAGTTAGACGTGAC 433
     |||||||| ||||||||||||||||||||||||  |||||| |||||
1249 TTCAAGGACGGCAGTTGAATCCATATTTGAGACTCATTGTTAGATGTGAC 1298

 434 CATATCATAGATGATGCACTTGTCCGG..................... 460
     |||||||||||||||||||||||||
1299 CATATCATAGATGATGCACTTGTCCGGCTAGAGATGATCACTATGGAAAA 1348
```

3R5 primer-binding site

<u>Fig 4.2.14</u> Primers 3F5 and 3R5 are complementary to *UBE3A* exon 9 and *UBE3AP2*

The GAP output for the direct comparison between *UBE3A* fragment 3F5 and 3R5 and *UBE3A*P2 is shown in the region of A: primer-binding sequence for 3F5 and B: the primer-binding sequence for 3R5. Complete sequence identity was observed between *UBE3A* and *UBE3AP2* at both primer-binding sites (highlighted in bold).

4.2.3.4.1 Identification of multiple nucleotide substitutions in AS Family W.

The amplified 3F3-3R3 fragment for each family member was cloned to separate the alleles for sequence analysis. In an attempt to reduce the number of subclones for sequencing, the amplified inserts of 8 subclones per family member were electrophoresed on an SSCP gel in order to distinguish the alleles on the basis of SSCP patterns. Two different SSCP bands were expected, representing a 'mutant' or 'wild-type' allele. However, the SSCP pattern observed was uninterpretable. Three differently migrating bands were observed amongst the subclones for the mother and each affected sibling, whilst two bands were observed amongst the subclones from the father (data not shown). It was thought that one of the bands may have been due to lack of an insert in the subclone, or that the PCR amplification reactions had been contaminated. PCR reactions were repeated using primers 3F3 and 3R3 for each family member, in case of contamination of the previous reactions, and the fragment subcloned.

Sequence analysis of the 3F3-3R3 subclones (from the repeated PCR reactions) identified two different alleles present in all family members. One allele was the wild-type *UBE3A* sequence. The second 'variant' allele had ten nucleotide substitutions as follows: 1) G570C, 2) T572C, 3) T638C, 4) C692T, 5) C746T, 6) T756G, 7) G767A, 8) C798A, 9) G800T, 10) T810C (data not shown). Eight of these changes were predicted to cause amino acid substitutions if present in *UBE3A* exon 9, whilst two were conservative (Fig 4.2.15). This was unusual, since complex mutations involving more than one nucleotide substitution are rare. Due to the number and nature of the changes, and the detection of this allele in all family members, the 'variant allele' was predicted to be the result of amplification of a pseudogene sequence. After the publication of the *UBE3A* pseudogene sequences, the variant allele in Family W was directly compared to *UBE3AP1* and *UBE3AP2*. No similarity was detected between the variant allele and *UBE3AP1*, but 100% homology was identified between the variant allele and *UBE3AP2* (Fig 4.2.16). The primer-binding sequences for 3F3 and 3R3 also demonstrated complete sequence identity between the *UBE3A* sequence and *UBE3AP2* (Fig 4.2.17). Comparison of the wild-type *UBE3A* exon 9 3F3-3R3 fragment with the *UBE3AP2* sequence found 97.6% identity, differing only at the positions of the 10 nucleotide changes identified in Family W. This showed conclusively that the changes identified in Family W with primers 3F3 and 3R3 represented the *UBE3AP2* sequence due to co-amplification of *UBE3A* exon 9 and *UBE3AP2.*

```
x98032 541 GAAGAAGACTCAGAAGCATCTTCCTCAAGGATAGGTGATAGCTCACAGGGAGACAACAAT
       181 E  E  D  S  E  A  S  S  S  R  I  G  D  S  S  Q  G  D  N  N
                                  G570C↓ ↓T572C
Fam. W 541 GAAGAAGACTCAGAAGCATCTTCCTCAAGCACAGGTGATAGCTCACAGGGAGACAACAAT
       181 E  E  D  S  E  A  S  S  S  S  T  G  D  S  S  Q  G  D  N  N


x98032 601 TTGCAAAAATTAGGCCCTGATGATGTGTCTGTGGATATTGATGCCATTAGAAGGGTCTAC
       201 L  Q  K  L  G  P  D  D  V  S  V  D  I  D  A  I  R  R  V  Y
                                                  ↓T638C
Fam. W 601 TTGCAAAAATTAGGCCCTGATGATGTGTCTGTGGATACTGATGCCATTAGAAGGGTCTAC
       201 L  Q  K  L  G  P  D  D  V  S  V  D  T  D  A  I  R  R  V  Y


x98032 661 ACCAGATTGCTCTCTAATGAAAAAATTGAAACTGCCTTTCTCAATGCACTTGTATATTTG
       221 T  R  L  L  S  N  E  K  I  E  T  A  F  L  N  A  L  V  Y  L
                                        ↓C692T
Fam. W 661 ACCAGATTGCTCTCTAATGAAAAAATTGAAATTGCCTTTCTCAATGCACTTGTATATTTG
       221 T  R  L  L  S  N  E  K  I  E  I  A  F  L  N  A  L  V  Y  L


x98032 721 TCACCTAACGTGGAATGTGACTTGACGTATCACAATGTATACTCTCGAGATCCTAATTAT
       241 S  P  N  V  E  C  D  L  T  Y  H  N  V  Y  S  R  D  P  N  Y
                                  ↓C746T    ↓T756G     ↓G767A
Fam.W  721 TCACCTAACGTGGAATGTGACTTGATGTATCACAAGGTATACTCTCAAGATCCTAATTAT
       241 S  P  N  V  E  C  D  L  M  Y  H  K  V  Y  S  Q  D  P  N  Y


x98032 781 CTGAATTTGTTCATTATCGGAATGGAGAATAGAAATCTCCACAGTCCTGAATATCTGGAA
       261 L  N  L  F  I  I  G  M  E  N  R  N  L  H  S  P  E  Y  L  E
                     G767A↓ ↓G800T     ↓T810C
Fam. W 781 CTGAATTTGTTCATTATAGTAATGGAGAACAGAAATCTCCACAGTCCTGAATATCTGGAA
       261 L  N  L  F  I  I  V  M  E  N  R  N  L  H  S  P  E  Y  L  E
```

Fig 4.2.15 Sequence changes identified on one 'variant' allele in AS Family W.

The nucleotide and amino acid sequence for fragment 3F3-3R3 of the published *UBE3A* sequence (X98032) is given above the variant allele identified in Family W (Fam. W). The nucleotide sequence changes in Family W are indicated by arrows and numbered 1-10. The amino acid translation product is given below the nucleotide sequence. Amino acids pertaining to the ten substitutions identified in Family W are highlighted in bold.

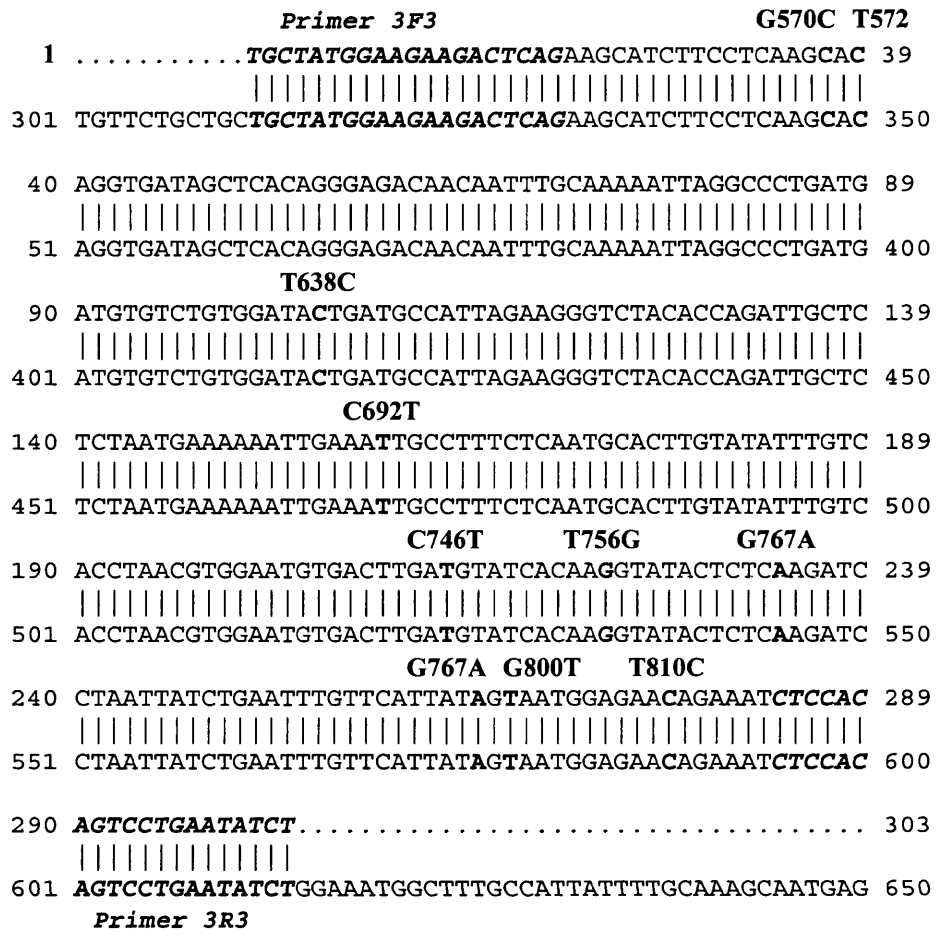


Fig 4.2.16 The variant allele in Family W represents the *UBE3AP2* pseudogene sequence.

The GAP output of the sequence comparison between the variant 3F3-3R3 allele in Family W (top), and *UBE3AP2* (AF010599) (bottom) is shown. The positions of sequence changes observed in the variant allele of Family W are given above the appropriate nucleotide, which is highlighted in bold. The primer-binding sequences for 3F3 and 3R3 are shown in bold-italics. The variant allele and pseudogene sequence are identical, indicating that the sequence changes in Family W are artefacts due to co-amplification of *UBE3AP2* with primers 3F3 and 3R3.

4.2.3.4.2 Identification of a truncating mutation in affected siblings in AS Family W

Although SSCP analysis of Family W with primers 3F3 and 3R3 indicated the presence of a maternally inherited sequence change in *UBE3A* exon 9 in the three affected children (Fig 4.2.5), sequence analysis had revealed nucleotide substitutions which were subsequently shown to be in the *UBE3AP2* pseudogene sequence. Since sequencing of subclones of the 3F3-3R3 PCR fragment had failed to identify any maternally transmitted sequence change in the AS children, sequence analysis of affected siblings 5211 and 5212 was repeated using primers E63-C and E63-D, E63-E and E63-F, and E63G and E63-H (Malzac *et al.*, 1998). These primers were used to amplify sections of exon 9 from genomic DNA, spanning the region previously amplified with primers 3F3 and 3R3. The amplification products of each primer pair were sequenced directly in both the forward and reverse directions. (These analyses were performed by S. Rickard).

A heterozygous 4bp tandem duplication (ATAG) at nucleotide position 575, with respect to the published *UBE3A* isoform I sequence (X98032), was detected in both siblings with primers E63-C and E63-D (Fig 4.2.17) (S. Rickard). This was predicted to cause a frameshift resulting in premature chain termination in all three E6-AP isoforms, at amino acid 194 with respect to isoform I (Fig 4.2.18). This would completely inactivate the ubiquitination function of E6-AP on this allele. This mutation could not be confirmed in affected sibling 5215 or the mother due to lack of DNA. However, on the basis of the SSCP data, this mutation was maternally transmitted (Fig 4.2.5).

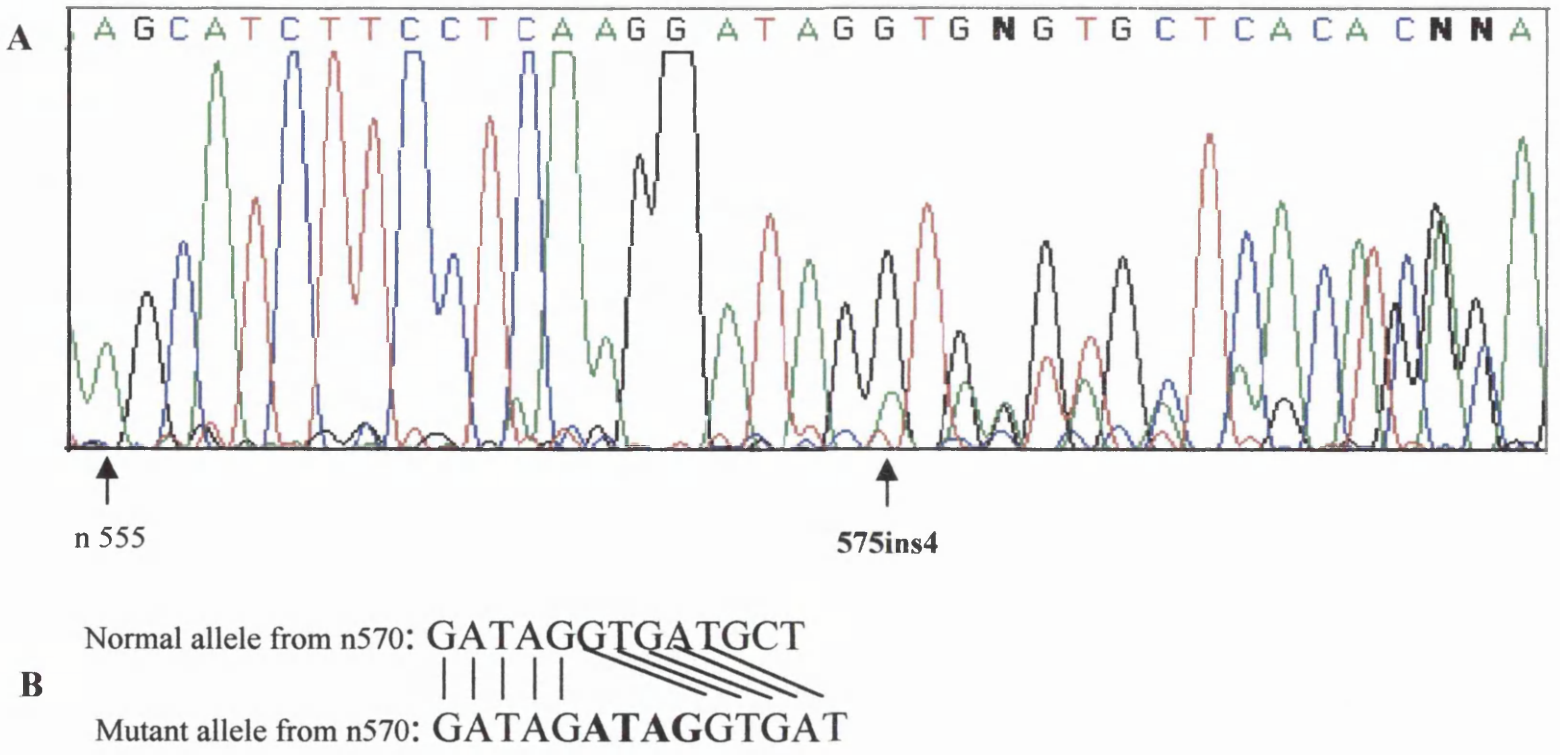


Fig 4.2.17 Heterozygous frameshift mutation detected in affected siblings 5211 and 5212 of Family W.

A: A section of the sequence electropherogram from *UBE3A* exon 9 is shown across the heterozygous frameshift mutation in AS sibling 5211. The region was amplified with primers E63-C and E63-D, then sequenced from both alleles. The sequence shown was acquired using E63-C. The nucleotide sequence is synchronous from nucleotide (n) 555 up to n575, the point of the 4bp insertion. After this the sequences become asynchronous due to the frameshift on one allele. **B**: Depiction of the ATAG insertion by the alignment of the normal sequence (above) with the mutant sequence (below) from n570 to beyond the insertion. The sequences become mis-aligned at n575 and the ATAG insertion is evident.

A: Normal *UBE3A* (X98032)

```
421 GTACAGAGCTTCCGGAAAGTTAAACAACACACCAAGGAAGAACTGAAATCTCTTCAAGCA
141 V  Q  S  F  R  K  V  K  Q  H  T  K  E  E  L  K  S  L  Q  A

481 AAAGATGAAGACAAAGATGAAGATGAAAAGGAAAAAGCTGCATGTTCTGCTGCTGCTATG
161 K  D  E  D  K  D  E  D  E  K  E  K  A  A  C  S  A  A  A  M

541 GAAGAAGACTCAGAAGCATCTTCCTCAAGGATAGGTGATAGCTCACAGGGAGACAACAAT
181 E  E  D  S  E  A  S  S  S  R  I  G  D  S  S  Q  G  D  N  N
```

B: Mutant allele in Family W

```
421 GTACAGAGCTTCCGGAAAGTTAAACAACACACCAAGGAAGAACTGAAATCTCTTCAAGCA
141 V  Q  S  F  R  K  V  K  Q  H  T  K  E  E  L  K  S  L  Q  A

481 AAAGATGAAGACAAAGATGAAGATGAAAAGGAAAAAGCTGCATGTTCTGCTGCTGCTATG
161 K  D  E  D  K  D  E  D  E  K  E  K  A  A  C  S  A  A  A  M

541 GAAGAAGACTCAGAAGCATCTTCCTCAAGGATAGATAGGTGATAGCTCACAGGGAGACAA
181 E  E  D  S  E  A  S  S  S  R  I  D  R  *  *  L  T  G  R  Q
                                      ↑     ↑
                                   575dup4  Chain termination
```

Fig 4.2.18 Heterozygous frameshift mutation in affected siblings of Family W causes truncation of *UBE3A*.

The MAP output for the published *UBE3A* sequence (A) and the mutant allele in Family W (B) is given. The nucleotides are above the predicted amino acid sequence. The ATAG tandem duplication (in bold) causes a frameshift with consequent introduction of a stop codon at amino acid 194. The amino acid sequence from the point of duplication is highlighted in bold.

4.2.5 Summary of mutations identified in the UBE3A gene in NDUI AS patients at the Institute of Child Health

Details of the pathogenic mutations identified in *UBE3A* in NDUI AS cases referred to the I.C.H. are given in Table 4.2. Disease-causing mutations in the *UBE3A* gene were identified in 3 out 5 (60%) familial cases. In two of the familial cases the mutations were confirmed by sequencing in all the affected siblings and the mother, demonstrating maternal transmission in accordance with the inheritance pattern for AS. In the case of Family W, the mutation was identified in affected siblings 5211 and 5212, but no DNA was available to confirm the presence of the mutation in sibling 5215. No DNA was left for either parent either, so maternal transmission could not be confirmed by sequence analysis, although the mother did have the same SSCP shift as the three affected siblings (Fig 4.2.5).

In sporadic Patient O the mutation was present in the proband and mother, but not in the unaffected child. This patient was a sporadic case of AS, since only one child was affected, but the mutation was not *de novo*. Demonstration that the mother was a carrier of the mutation showed that this was potentially a familial case of AS, and a 50% sibling recurrence risk is indicated in this family. The 12bp deletion was not identified in 50 normal individuals, and demonstration that it was maternally inherited provided further support that this was the pathogenic mutation in this patient. If Family O was categorised as a familial case on the basis that the mother was a mutation carrier, the number of *UBE3A* mutations identified in familial cases rose to 4 in 6 (66.7%).

De novo truncating mutations of *UBE3A* were identified in 4/21 (19%) sporadic cases of NDUI AS screened, or 3/20 (15%) if Patient O is recategorised as a familial case, despite using similar screening procedures in familial and isolated AS cases. These results indicate that mutations of *UBE3A* can be a cause for AS, but may not account for all cases.

Family/ proband	Form of AS	Sequence alteration	Site (exon)	Primers	Predicted effect	Inheritance
Family H	Familial	1407delATTAT	9	3F5/3R5 E6-9C/ E6-9D	Frameshift	Present in mother and both affected siblings
Family W	Familial	575dupATAG	9	E63-C/ E63-D	Frameshift	Confirmed in two of three affected siblings
Family A	Familial	2344delTTC	15	9F9/ 9R9	F782del	Present in mother and both affected children
Patient O	Sporadic	2268delTATACCAGGGAC	14	E6-14F/ E6-14R	YTRD757del	Present in mother, but not in unaffected sibling
5954	Sporadic	1379delT	9	3F5/ 3R5	Frameshift	*De novo*
5173	Sporadic	2507delAAGA	16	E6ZA/ E6ZB	Frameshift	*De novo*
18244	Sporadic	2516insAAGTT	16	E6ZA/ E6ZB	Frameshift	*De novo*

Table 4.2 Description of pathogenic mutations identified amongst the non-deletion, disomy and imprintor mutation AS cases at the Institute of Child Health.

Chapter 5: RESULTS

Characterisation of I.M.A.G.E. clones 395-H22 and 123-E19 mapping to the PWS region.

Summary

I.M.A.G.E. clones 204357 (395-H22) and 125178 (123-E19) had been identified by BLAST analysis of selected sequences 83 and 93 respectively, from the 132D4-selected cDNA sublibrary (Table 3.2.3). 395-H22 and 123-E19 were mapped proximal to D15S174 in the PWS region, between *SNRPN* and *IPW* (Table 3.2.5; Fig 3.2.28). PWS is a contiguous gene disorder caused by a deficiency of paternally expressed genes, although the number of genes involved in the phenotype is not known. In addition to *ZNF127*, *SNRPN*, *NDN* and *IPW*, other transcripts might exist that also contribute to PWS. Clones 395-H22 and 123-E19 were characterised further to determine if they represented cDNAs, which could potentially play a role in the PWS phenotype.

5.1. Sequence analysis

To characterise I.M.A.G.E. clones 395-H22 and 123-E19 the insert was sequenced and analysed for coding potential. The internal sections of sequence were obtained by oligomer-walking from the 5' and 3' end-sequences acquired using vector primers. Both sequences had been cloned into the modified pT7T3D vector via the *Pac*I restriction site that was present within the poly dT primer at the 3' end, and the *EcoR*I site at the 5' end (Table 3.2.4). Thus the insert sequences were oriented with respect to the 5' *EcoR*I cloning site.

5.1.1. Clone 204357 (395-H22)

The 1.2kb insert sequence for clone 204357 (395-H22) is given in Fig 5.1. The end sequences generated by M13 vector primers matched the 5' and 3' ESTs of clone 204357 in the databases (GenBank H59927 and H59928 respectively), confirming this to be the clone identified by BLAST analysis of selected sequence 83 (Table 3.2.3). Sequence 83 matched nucleotides 14-416 of the insert sequence. There was no poly A tail in the clone, although digestion with the *Pac*I at the 3' end will have cleaved this from the transcript during cDNA library construction. No consensus polyadenylation signal was identified. The MAP program was used to predict the amino acid translation product for 395-H22 in the three possible reading frames, but no significant open reading frame was identified (Fig 5.2). NIX analysis also failed to identify any protein-coding region. This indicated that 395-H22 is not translated and could potentially comprise the 3' UTR of a derivative transcript, or function as an RNA.

M13-R

```
   1  GAATTCGGCA CGAGACAAAA TGAGTGAAAA CTCTATACTG TCATCCTCGT
      EcoRI
  51  CGAACTGAGG TCCAGCACAT TACTCCAACA GGGGCTAGAC AGAAAAGGCC
                          H22-5'
 101  AACATCCGTT TGTTGACATG GGTTATATCA AGGCGTCTGT TCAGGCTTAG

 151  AATGTGGTCT CTTATGGGTG ATGGGGGTCA CAGGAGAGTG GTGGCTCCCA

 201  CGTATAGGAA ATTTCTTGTT TGAAGGACTG TCAGTAGGGT GGGTAACACA

 251  TGCATTGTCT GCAGGACTAG GTGAATGTCC ATGTGGCCTA GCAAGAGTTA

 301  GCTGGTAGCC CGCCTCTGGT TGCCAATTTG TTCTTGAGTC CTTGTTCTGG

 351  GTTCTCAGGT CCCACGGAGG AAAACAGATC TGTGTGGTTG AGAGGTGGGT

 401  ACAAGGCCGC ATCTTTGTCA TTTGTTGGCT AACTTTGTCC TTGGTTGAGG

 451  ACATTAGAGT TTTGGTCACC AGGCATAGCC TATGTGCCAT TGTGCCCGTG

 501  TTGTATCCCA CGTGTTTTGA GGACATGTAT TTTGCACGTA AAGGTGAGCT

 551  CCTGCTCCAA GCTGGTTCTG ATACCAAAGG AGTCCCTGGC TTATCCTAAA

 601  CTCATGGTAG GTTAAAGCCT TCCTCCTTAG GGGTTCAGGC CCGCAAGGCT

 651  TTTGTGAGTG GCATTGCAGG CGTTGAAGCA GTGATGTTGA GAGGGATGGT

 701  CAATGTCAGT GCTCTTTAGC AGGATGGTGT ACTGCAGGGG CCCCCAGCCC

 751  CGAGACGAGC ATCCCTGCAT CCATGCATTT ATGCCTCCAT GAACAGGGGA

 801  GGCCAGAGAC AGGCAGATAG TAGATAAATT GCAGGGGACT GGATGACATG
                H22-S1
 851  GCCCTCGTGA CCTGTGCACC TGTCTGTCTT TCTGAAGCAC GCCTGTGTTA

 901  ACTCTGCACC TCCCAGGTAG CACTGGCATG GAGGGCAGGC ACATGTTGGT

 951  GAGGGACAAT TGTTACCTTG TGTGAGCTGC GGAGATACCA GGAAGCCCCT

1001  GGACACAAAT GGCAAAGGCT CCTTCGGAAG TTGTTGGATC CCTTCTGAAT

1051  GTAAGCACTT CTTTCCCAGA GCACTCTGAG TTTCCTCATT TGCAGGGACA

1101  AATACTGTGC GTGGATCGAT GATGACTTCC ACATATACAT TCCTTGGAAA
                                                 H22-3'
1151  GCTGAACAAA ATGAGTGAAA ACTCTATACC GTCATCCTCG TCGAACTGAG

1201  GTCC
```

M13-U

Fig 5.1 Nucleotide sequence of I.M.A.G.E. clone 204357 (395-H22)

The full 1.2kb insert sequence for clone 395-H22 is shown. The *EcoR*I site denotes the 5' end of the insert. The primer-binding sites for sequencing and RT-PCR are underlined, with arrows pointing 5' to 3' with respect to the oligonucleotides. M13 vector primers were used to sequence the ends of the insert, as indicated. Primer H22-S1 was designed from the M13U sequence tag and used to sequence the internal section of the insert. Primers H22-5' and H22-3' were used for RT-PCR analysis.

```
      1 GAATTCGGCACGAGACAAAATGAGTGAAAACTCTATACTGTCATCCTCGTCGAACTGAGG
a       E  F  G  T  R  Q  N  E  *  K  L  Y  T  V  I  L  V  E  L  R
b        N  S  A  R  D  K  M  S  E  N  S  I  L  S  S  S  S  N  *  G
c         I  R  H  E  T  K  *  V  K  T  L  Y  C  H  P  R  R  T  E  V

     61 TCCAGCACATTACTCCAACAGGGGCTAGACAGAAAAGGCCAACATCCGTTTGTTGACATG
a       S  S  T  L  L  Q  Q  G  L  D  R  K  G  Q  H  P  F  V  D  M
b        P  A  H  Y  S  N  R  G  *  T  E  K  A  N  I  R  L  L  T  W
c         Q  H  I  T  P  T  G  A  R  Q  K  R  P  T  S  V  C  *  H  G

    121 GGTTATATCAAGGCGTCTGTTCAGGCTTAGAATGTGGTCTCTTATGGGTGATGGGGGTCA
a       G  Y  I  K  A  S  V  Q  A  *  N  V  V  S  Y  G  *  W  G  S
b        V  I  S  R  R  L  F  R  L  R  M  W  S  L  M  G  D  G  G  H
c         L  Y  Q  G  V  C  S  G  L  E  C  G  L  L  W  V  M  G  V  T

    181 CAGGAGAGTGGTGGCTCCCACGTATAGGAAATTTCTTGTTTGAAGGACTGTCAGTAGGGT
a       Q  E  S  G  G  S  H  V  *  E  I  S  C  L  K  D  C  Q  *  G
b        R  R  V  V  A  P  T  Y  R  K  F  L  V  *  R  T  V  S  R  V
c         G  E  W  W  L  P  R  I  G  N  F  L  F  E  G  L  S  V  G  W

    241 GGGTAACACATGCATTGTCTGCAGGACTAGGTGAATGTCCATGTGGCCTAGCAAGAGTTA
a       G  *  H  M  H  C  L  Q  D  *  V  N  V  H  V  A  *  Q  E  L
b        G  N  T  C  I  V  C  R  T  R  *  M  S  M  W  P  S  K  S  *
c         V  T  H  A  L  S  A  G  L  G  E  C  P  C  G  L  A  R  V  S

    301 GCTGGTAGCCCGCCTCTGGTTGCCAATTTGTTCTTGAGTCCTTGTTCTGGGTTCTCAGGT
a       A  G  S  P  P  L  V  A  N  L  F  L  S  P  C  S  G  F  S  G
b        L  V  A  R  L  W  L  P  I  C  S  *  V  L  V  L  G  S  Q  V
c         W  *  P  A  S  G  C  Q  F  V  L  E  S  L  F  W  V  L  R  S

    361 CCCACGGAGGAAAACAGATCTGTGTGGTTGAGAGGTGGGTACAAGGCCGCATCTTTGTCA
a       P  T  E  E  N  R  S  V  W  L  R  G  G  Y  K  A  A  S  L  S
b        P  R  R  K  T  D  L  C  G  *  E  V  G  T  R  P  H  L  C  H
c         H  G  G  K  Q  I  C  V  V  E  R  W  V  Q  G  R  I  F  V  I

    421 TTTGTTGGCTAACTTTGTCCTTGGTTGAGGACATTAGAGTTTTGGTCACCAGGCATAGCC
a       F  V  G  *  L  C  P  W  L  R  T  L  E  F  W  S  P  G  I  A
b        L  L  A  N  F  V  L  G  *  G  H  *  S  F  G  H  Q  A  *  P
c         C  W  L  T  L  S  L  V  E  D  I  R  V  L  V  T  R  H  S  L

    481 TATGTGCCATTGTGCCCGTGTTGTATCCCACGTGTTTTGAGGACATGTATTTTGCACGTA
a       Y  V  P  L  C  P  C  C  I  P  R  V  L  R  T  C  I  L  H  V
b        M  C  H  C  A  R  V  V  S  H  V  F  *  G  H  V  F  C  T  *
c         C  A  I  V  P  V  L  Y  P  T  C  F  E  D  M  Y  F  A  R  K

    541 AAGGTGAGCTCCTGCTCCAAGCTGGTTCTGATACCAAAGGAGTCCCTGGCTTATCCTAAA
a       K  V  S  S  C  S  K  L  V  L  I  P  K  E  S  L  A  Y  P  K
b        R  *  A  P  A  P  S  W  F  *  Y  Q  R  S  P  W  L  I  L  N
c         G  E  L  L  L  Q  A  G  S  D  T  K  G  V  P  G  L  S  *  T

    601 CTCATGGTAGGTTAAAGCCTTCCTCCTTAGGGGTTCAGGCCCGCAAGGCTTTTGTGAGTG
a       L  M  V  G  *  S  L  P  P  *  G  F  R  P  A  R  L  L  *  V
b        S  W  *  V  K  A  F  L  L  R  G  S  G  P  Q  G  F  C  E  W
c         H  G  R  L  K  P  S  S  L  G  V  Q  A  R  K  A  F  V  S  G
```

```
     661 GCATTGCAGGCGTTGAAGCAGTGATGTTGAGAGGGATGGTCAATGTCAGTGCTCTTTAGC
a          A  L  Q  A  L  K  Q  *  C  *  E  G  W  S  M  S  V  L  F  S
b           H  C  R  R  *  S  S  D  V  E  R  D  G  Q  C  Q  C  S  L  A
c            I  A  G  V  E  A  V  M  L  R  G  M  V  N  V  S  A  L  *  Q

     721 AGGATGGTGTACTGCAGGGGCCCCCAGCCCCGAGACGAGCATCCCTGCATCCATGCATTT
a          R  M  V  Y  C  R  G  P  Q  P  R  D  E  H  P  C  I  H  A  F
b           G  W  C  T  A  G  A  P  S  P  E  T  S  I  P  A  S  M  H  L
c            D  G  V  L  Q  G  P  P  A  P  R  R  A  S  L  H  P  C  I  Y

     781 ATGCCTCCATGAACAGGGGAGGCCAGAGACAGGCAGATAGTAGATAAATTGCAGGGGACT
a          M  P  P  *  T  G  E  A  R  D  R  Q  I  V  D  K  L  Q  G  T
b           C  L  H  E  Q  G  R  P  E  T  G  R  *  *  I  N  C  R  G  L
c            A  S  M  N  R  G  G  Q  R  Q  A  D  S  R  *  I  A  G  D  W

     841 GGATGACATGGCCCTCGTGACCTGTGCACCTGTCTGTCTTTCTGAAGCACGCCTGTGTTA
a          G  *  H  G  P  R  D  L  C  T  C  L  S  F  *  S  T  P  V  L
b           D  D  M  A  L  V  T  C  A  P  V  C  L  S  E  A  R  L  C  *
c            M  T  W  P  S  *  P  V  H  L  S  V  F  L  K  H  A  C  V  N

     901 ACTCTGCACCTCCCAGGTAGCACTGGCATGGAGGGCAGGCACATGTTGGTGAGGGACAAT
a          T  L  H  L  P  G  S  T  G  M  E  G  R  H  M  L  V  R  D  N
b           L  C  T  S  Q  V  A  L  A  W  R  A  G  T  C  W  *  G  T  I
c            S  A  P  P  R  *  H  W  H  G  G  Q  A  H  V  G  E  G  Q  L

     961 TGTTACCTTGTGTGAGCTGCGGAGATACCAGGAAGCCCCTGGACACAAATGGCAAAGGCT
a          C  Y  L  V  *  A  A  E  I  P  G  S  P  W  T  Q  M  A  K  A
b           V  T  L  C  E  L  R  R  Y  Q  E  A  P  G  H  K  W  Q  R  L
c            L  P  C  V  S  C  G  D  T  R  K  P  L  D  T  N  G  K  G  S

    1021 CCTTCGGAAGTTGTTGGATCCCTTCTGAATGTAAGCACTTCTTTCCCAGAGCACTCTGAG
a          P  S  E  V  V  G  S  L  L  N  V  S  T  S  F  P  E  H  S  E
b           L  R  K  L  L  D  P  F  *  M  *  A  L  L  S  Q  S  T  L  S
c            F  G  S  C  W  I  P  S  E  C  K  H  F  F  P  R  A  L  *  V

    1081 TTTCCTCATTTGCAGGGACAAATACTGTGCGTGGATCGATGATGACTTCCACATATACAT
a          F  P  H  L  Q  G  Q  I  L  C  V  D  R  *  *  L  P  H  I  H
b           F  L  I  C  R  D  K  Y  C  A  W  I  D  D  D  F  H  I  Y  I
c            S  S  F  A  G  T  N  T  V  R  G  S  M  M  T  S  T  Y  T  F

    1141 TCCTTGGAAAGCTGAACAAAATGAGTGAAAACTCTATACCGTCATCCTCGTCGAACTGAGGTCC
a          S  L  E  S  *  T  K  *  V  K  T  L  Y  R  H  P  R  R  T  E  V
b           P  W  K  A  E  Q  N  E  *  K  L  Y  T  V  I  L  V  E  L  R  S
c            L  G  K  L  N  K  M  S  E  N  S  I  P  S  S  S  S  N  *  G  C
```

Fig 5.2 Predicted amino acid translation product for sequence 204357 (395-H22)

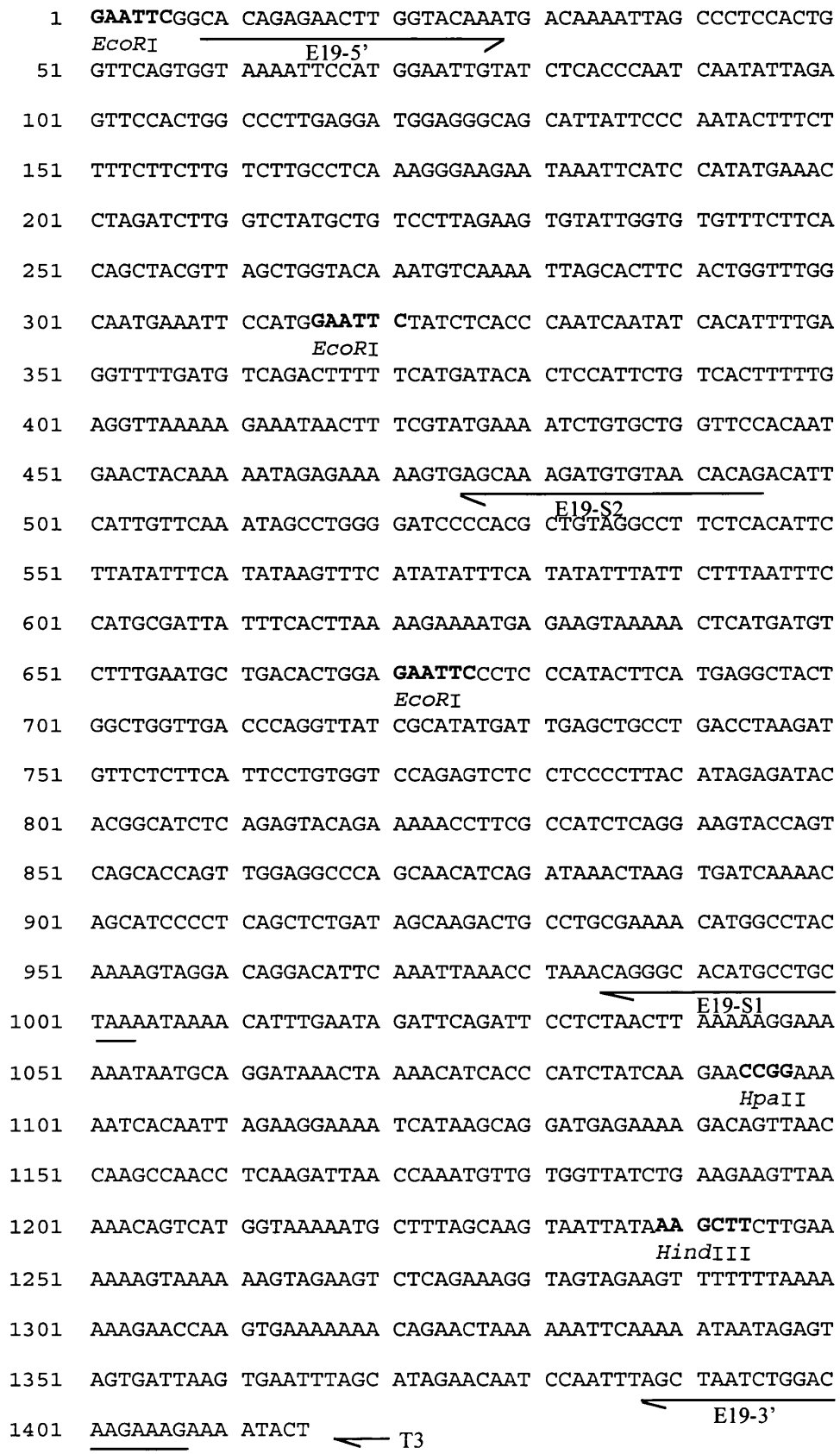
Output from the MAP program, with translation in three reading frames (a, b, c). Stop codons are shown as stars (*). No significant open reading frame was identified, showing that this sequence is untrnslated.

5.1.2 Clone 125178 (123-E19)

The 1.4kb insert sequence for I.M.A.G.E. clone 123-E19 is shown in Fig 5.3. The end sequences of the insert matched the 5' and 3' ESTs for clone 125178 in the sequence databases (GenBank R05682 and R05683 respectively), confirming that 123-E19 was the clone identified by selected sequence 93 (Table 3.2.3). Sequence 93 corresponded to nucleotides 70-114 of 123-E19. 123-E19 contained two *EcoR*I sites, hence the identification of 3 separate *EcoR*I fragments in PAC 121D5, YAC 132D4 and genomic DNA by 123-E19 (Table 3.2.5). The PolyA program in NIX analysis identified a single AATAAA consensus polyadenylation signal at nucleotides 179-184 (Fig 5.3). Due to digestion of the 3' end of the sequence with *Pac*I, the polyA tail was cleaved from the sequence during cDNA library construction. 123-E19 is an untranslated sequence, with no significant open reading frame in any of the three possible reading frames (Fig 5.4). This clone may represent the 3' UTR of a transcript or an untranslated RNA molecule.

Fig 5.3 Nucleotide sequence for I.M.A.G.E. clone 125178 (123-E19)(next page)

The 1.4kb insert sequence of clone 123-E19 is shown. The complementary sites for sequencing and RT-PCR primers are underlined, with the arrow pointing 5' to 3' with respect to the oligonucleotide. *EcoR*I sites are indicated in bold and the polyadenylation signal is boxed. The 3' end sequence was acquired with vector primer T3. The sequence of the full 1.4kb insert was obtained by primer walking from the T3 sequence tag, with oligonucleotides E19-S1, followed in turn by E19-S2.


1 GAATTCGGCA CAGAGAACTT GGTACAAATG ACAAAATTAG CCCTCCACTG
EcoRI
E19-5'
51 GTTCAGTGGT AAAATTCCAT GGAATTGTAT CTCACCCAAT CAATATTAGA
101 GTTCCACTGG CCCTTGAGGA TGGAGGGCAG CATTATTCCC AATACTTTCT
151 TTTCTTCTTG TCTTGCCTCA AAGGGAAGAA TAAATTCATC CATATGAAAC
201 CTAGATCTTG GTCTATGCTG TCCTTAGAAG TGTATTGGTG TGTTTCTTCA
251 CAGCTACGTT AGCTGGTACA AATGTCAAAA TTAGCACTTC ACTGGTTTGG
301 CAATGAAATT CCATGGAATT CTATCTCACC CAATCAATAT CACATTTTGA
EcoRI
351 GGTTTTGATG TCAGACTTTT TCATGATACA CTCCATTCTG TCACTTTTTG
401 AGGTTAAAAA GAAATAACTT TCGTATGAAA ATCTGTGCTG GTTCCACAAT
451 GAACTACAAA AATAGAGAAA AAGTGAGCAA AGATGTGTAA CACAGACATT
E19-S2
501 CATTGTTCAA ATAGCCTGGG GATCCCCACG CTGTAGGCCT TCTCACATTC
551 TTATATTTCA TATAAGTTTC ATATATTTCA TATATTTATT CTTTAATTTC
601 CATGCGATTA TTTCACTTAA AAGAAAATGA GAAGTAAAAA CTCATGATGT
651 CTTTGAATGC TGACACTGGA GAATTCCCTC CCATACTTCA TGAGGCTACT
EcoRI
701 GGCTGGTTGA CCCAGGTTAT CGCATATGAT TGAGCTGCCT GACCTAAGAT
751 GTTCTCTTCA TTCCTGTGGT CCAGAGTCTC CTCCCCTTAC ATAGAGATAC
801 ACGGCATCTC AGAGTACAGA AAAACCTTCG CCATCTCAGG AAGTACCAGT
851 CAGCACCAGT TGGAGGCCCA GCAACATCAG ATAAACTAAG TGATCAAAAC
901 AGCATCCCCT CAGCTCTGAT AGCAAGACTG CCTGCGAAAA CATGGCCTAC
951 AAAAGTAGGA CAGGACATTC AAATTAAACC TAAACAGGGC ACATGCCTGC
E19-S1
1001 TAAAATAAAA CATTTGAATA GATTCAGATT CCTCTAACTT AAAAAGGAAA
1051 AAATAATGCA GGATAAACTA AAACATCACC CATCTATCAA GAACCGGAAA
HpaII
1101 AATCACAATT AGAAGGAAAA TCATAAGCAG GATGAGAAAA GACAGTTAAC
1151 CAAGCCAACC TCAAGATTAA CCAAATGTTG TGGTTATCTG AAGAAGTTAA
1201 AAACAGTCAT GGTAAAAATG CTTTAGCAAG TAATTATAAA GCTTCTTGAA
HindIII
1251 AAAAGTAAAA AAGTAGAAGT CTCAGAAAGG TAGTAGAAGT TTTTTTAAAA
1301 AAAGAACCAA GTGAAAAAAA CAGAACTAAA AAATTCAAAA ATAATAGAGT
1351 AGTGATTAAG TGAATTTAGC ATAGAACAAT CCAATTTAGC TAATCTGGAC
E19-3'
1401 AAGAAAGAAA ATACT
T3

```
      1 GAATTCGGCACAGAGAACTTGGTACAAATGACAAAATTAGCCCTCCACTGGTTCAGTGGT
a       E  F  G  T  E  N  L  V  Q  M  T  K  L  A  L  H  W  F  S  G
b        N  S  A  Q  R  T  W  Y  K  *  Q  N  *  P  S  T  G  S  V  V
c         I  R  H  R  E  L  G  T  N  D  K  I  S  P  P  L  V  Q  W  *

     61 AAAATTCCATGGAATTGTATCTCACCCAATCAATATTAGAGTTCCACTGGCCCTTGAGGA
a       K  I  P  W  N  C  I  S  P  N  Q  Y  *  S  S  T  G  P  *  G
b        K  F  H  G  I  V  S  H  P  I  N  I  R  V  P  L  A  L  E  D
c         N  S  M  E  L  Y  L  T  Q  S  I  L  E  F  H  W  P  L  R  M

    121 TGGAGGGCAGCATTATTCCCAATACTTTCTTTTCTTCTTGTCTTGCCTCAAAGGGAAGAA
a       W  R  A  A  L  F  P  I  L  S  F  L  L  V  L  P  Q  R  E  E
b        G  G  Q  H  Y  S  Q  Y  F  L  F  F  L  S  C  L  K  G  K  N
c         E  G  S  I  I  P  N  T  F  F  S  S  C  L  A  S  K  G  R  I

    181 TAAATTCATCCATATGAAACCTAGATCTTGGTCTATGCTGTCCTTAGAAGTGTATTGGTG
a       *  I  H  P  Y  E  T  *  I  L  V  Y  A  V  L  R  S  V  L  V
b        K  F  I  H  M  K  P  R  S  W  S  M  L  S  L  E  V  Y  W  C
c         N  S  S  I  *  N  L  D  L  G  L  C  C  P  *  K  C  I  G  V

    241 TGTTTCTTCACAGCTACGTTAGCTGGTACAAATGTCAAAATTAGCACTTCACTGGTTTGG
a       C  F  F  T  A  T  L  A  G  T  N  V  K  I  S  T  S  L  V  W
b        V  S  S  Q  L  R  *  L  V  Q  M  S  K  L  A  L  H  W  F  G
c         F  L  H  S  Y  V  S  W  Y  K  C  Q  N  *  H  F  T  G  L  A

    301 CAATGAAATTCCATGGAATTCTATCTCACCCAATCAATATCACATTTTGAGGTTTTGATG
a       Q  *  N  S  M  E  F  Y  L  T  Q  S  I  S  H  F  E  V  L  M
b        N  E  I  P  W  N  S  I  S  P  N  Q  Y  H  I  L  R  F  *  C
c         M  K  F  H  G  I  L  S  H  P  I  N  I  T  F  *  G  F  D  V

    361 TCAGACTTTTTCATGATACACTCCATTCTGTCACTTTTTGAGGTTAAAAAGAAATAACTT
a       S  D  F  F  M  I  H  S  I  L  S  L  F  E  V  K  K  K  *  L
b        Q  T  F  S  *  Y  T  P  F  C  H  F  L  R  L  K  R  N  N  F
c         R  L  F  H  D  T  L  H  S  V  T  F  *  G  *  K  E  I  T  F

    421 TCGTATGAAAATCTGTGCTGGTTCCACAATGAACTACAAAAATAGAGAAAAAGTGAGCAA
a       S  Y  E  N  L  C  W  F  H  N  E  L  Q  K  *  R  K  S  E  Q
b        R  M  K  I  C  A  G  S  T  M  N  Y  K  N  R  E  K  V  S  K
c         V  *  K  S  V  L  V  P  Q  *  T  T  K  I  E  K  K  *  A  K

    481 AGATGTGTAACACAGACATTCATTGTTCAAATAGCCTGGGGATCCCCACGCTGTAGGCCT
a       R  C  V  T  Q  T  F  I  V  Q  I  A  W  G  S  P  R  C  R  P
b        D  V  *  H  R  H  S  L  F  K  *  P  G  D  P  H  A  V  G  L
c         M  C  N  T  D  I  H  C  S  N  S  L  G  I  P  T  L  *  A  F

    541 TCTCACATTCTTATATTTCATATAAGTTTCATATATTTCATATATTTATTCTTTAATTTC
a       S  H  I  L  I  F  H  I  S  F  I  Y  F  I  Y  L  F  F  N  F
b        L  T  F  L  Y  F  I  *  V  S  Y  I  S  Y  I  Y  S  L  I  S
c         S  H  S  Y  I  S  Y  K  F  H  I  F  H  I  F  I  L  *  F  P

    601 CATGCGATTATTTCACTTAAAAGAAAATGAGAAGTAAAAACTCATGATGTCTTTGAATGC
a       H  A  I  I  S  L  K  R  K  *  E  V  K  T  H  D  V  F  E  C
b        M  R  L  F  H  L  K  E  N  E  K  *  K  L  M  M  S  L  N  A
c         C  D  Y  F  T  *  K  K  M  R  S  K  N  S  *  C  L  *  M  L
```

```
     661 TGACACTGGAGAATTCCCTCCCATACTTCATGAGGCTACTGGCTGGTTGACCCAGGTTAT
a          *  H  W  R  I  P  S  H  T  S  *  G  Y  W  L  V  D  P  G  Y
b           D  T  G  E  F  P  P  I  L  H  E  A  T  G  W  L  T  Q  V  I
c           T  L  E  N  S  L  P  Y  F  M  R  L  L  A  G  *  P  R  L  S

     721 CGCATATGATTGAGCTGCCTGACCTAAGATGTTCTCTTCATTCCTGTGGTCCAGAGTCTC
a          R  I  *  L  S  C  L  T  *  D  V  L  F  I  P  V  V  Q  S  L
b           A  Y  D  *  A  A  *  P  K  M  F  S  S  F  L  W  S  R  V  S
c           H  M  I  E  L  P  D  L  R  C  S  L  H  S  C  G  P  E  S  P

     781 CTCCCCTTACATAGAGATACACGGCATCTCAGAGTACAGAAAAACCTTCGCCATCTCAGG
a          L  P  L  H  R  D  T  R  H  L  R  V  Q  K  N  L  R  H  L  R
b           S  P  Y  Y  I  E  I  H  G  I  S  E  Y  R  K  T  F  A  I  S  G
c           P  L  T  *  R  Y  T  A  S  Q  S  T  E  K  P  S  P  S  Q  E

     841 AAGTACCAGTCAGCACCAGTTGGAGGCCCAGCAACATCAGATAAACTAAGTGATCAAAAC
a          K  Y  Q  S  A  P  V  G  G  P  A  T  S  D  K  L  S  D  Q  N
b           S  T  S  Q  H  Q  L  E  A  Q  Q  H  Q  I  N  *  V  I  K  T
c           V  P  V  S  T  S  W  R  P  S  N  I  R  *  T  K  *  S  K  Q

     901 AGCATCCCCTCAGCTCTGATAGCAAGACTGCCTGCGAAAACATGGCCTACAAAAGTAGGA
a          S  I  P  S  A  L  I  A  R  L  P  A  K  T  W  P  T  K  V  G
b           A  S  P  Q  L  *  *  Q  D  C  L  R  K  H  G  L  Q  K  *  D
c           H  P  L  S  S  D  S  K  T  A  C  E  N  M  A  Y  K  S  R  T

     961 CAGGACATTCAAATTAAACCTAAACAGGGCACATGCCTGCTAAAATAAAACATTTGAATA
a          Q  D  I  Q  I  K  P  K  Q  G  T  C  L  L  K  *  N  I  *  I
b           R  T  F  K  L  N  L  N  R  A  H  A  C  *  N  K  T  F  E  *
c           G  H  S  N  *  T  *  T  G  H  M  P  A  K  I  K  H  L  N  R

    1021 GATTCAGATTCCTCTAACTTAAAAAGGAAAAAATAATGCAGGATAAACTAAAACATCACC
a          D  S  D  S  S  N  L  K  R  K  K  *  C  R  I  N  *  N  I  T
b           I  Q  I  P  L  T  *  K  G  K  N  N  A  G  *  T  K  T  S  P
c           F  R  F  L  *  L  K  K  E  K  I  M  Q  D  K  L  K  H  H  P

    1081 CATCTATCAAGAACCGGAAAAATCACAATTAGAAGGAAAATCATAAGCAGGATGAGAAAA
a          H  L  S  R  T  G  K  I  T  I  R  R  K  I  I  S  R  M  R  K
b           I  Y  Q  E  P  E  K  S  Q  L  E  G  K  S  *  A  G  *  E  K
c           S  I  K  N  R  K  N  H  N  *  K  E  N  H  K  Q  D  E  K  R

    1141 GACAGTTAACCAAGCCAACCTCAAGATTAACCAAATGTTGTGGTTATCTGAAGAAGTTAA
a          D  S  *  P  S  Q  P  Q  D  *  P  N  V  V  V  I  *  R  S  *
b           T  V  N  Q  A  N  L  K  I  N  Q  M  L  W  L  S  E  E  V  K
c           Q  L  T  K  P  T  S  R  L  T  K  C  C  G  Y  L  K  K  L  K

    1201 AAACAGTCATGGTAAAAATGCTTTAGCAAGTAATTATAAAGCTTCTTGAAAAAAGTAAAA
a          K  Q  S  W  *  K  C  F  S  K  *  L  *  S  F  L  K  K  V  K
b           N  S  H  G  K  N  A  L  A  S  N  Y  K  A  S  *  K  K  *  K
c           T  V  M  V  K  M  L  *  Q  V  I  I  K  L  L  E  K  S  K  K

    1261 AAGTAGAAGTCTCAGAAAGGTAGTAGAAGTTTTTTTAAAAAAAGAACCAAGTGAAAAAAA
a          K  *  K  S  Q  K  G  S  R  S  F  F  K  K  R  T  K  *  K  K
b           S  R  S  L  R  K  V  V  E  V  F  L  K  K  E  P  S  E  K  N
c           V  E  V  S  E  R  *  *  K  F  F  *  K  K  N  Q  V  K  K  T
```

```
    1321 CAGAACTAAAAAATTCAAAAATAATAGAGTAGTGATTAAGTGAATTTAGCATAGAACAAT
a         Q  N  *  K  I  Q  K  *  *  S  S  D  *  V  N  L  A  *  N  N
b          R  T  K  K  F  K  N  N  R  V  V  I  K  *  I  *  H  R  T  I
c           E  L  K  N  S  K  I  I  E  *  *  L  S  E  F  S  I  E  Q  S

    1381 CCAATTTAGCTAATCTGGACAAGAAAGAAAATACT
a         P  I  *  L  I  W  T  R  K  K  I
b          Q  F  S  *  S  G  Q  E  R  K  Y
c           N  L  A  N  L  D  K  K  E  N  T
```

Fig 5.4 Predicted translation product for 125178 (123-E19)

MAP output for sequence 123-E19. The sequence is translated in three frames in the forward direction (a, b, c). Stop codons are denoted by a star (*). No significant open reading frame was identified, indicating that 123-E19 is untranslated.

5.2 Analysis of expression

5.2.1. 395-H22

The insert of clone 395-H22 identified a 1.4kb transcript of differing intensity in various fetal tissues by northern analysis. The 1.4kb signal was detected in a 7 week gestation whole embryo, 10 week fetal brain, heart and gut (Fig 5.5), 15 week fetal brain, heart, liver, kidney, lung and skin (Fig 5.6), and 17 week brain, heart, kidney, liver and skin (data not shown). Expression was most prominent in the fetal heart of 10, 15 and 17 week gestational ages in comparison to the other tissues tested (Figs 5.5 and 5.6). No expression was detected in 10 week placenta (Fig 5.6). These data show that 395-H22 is expressed predominantly in heart, but in a variety of tissues at different stages of fetal development. The insert for 395-H22 is 1.2kb (Fig 5.1). Detection of a 1.4kb transcript showed that 200bp of sequence remained to be identified to obtain the full-length transcript identified by 395-H22.

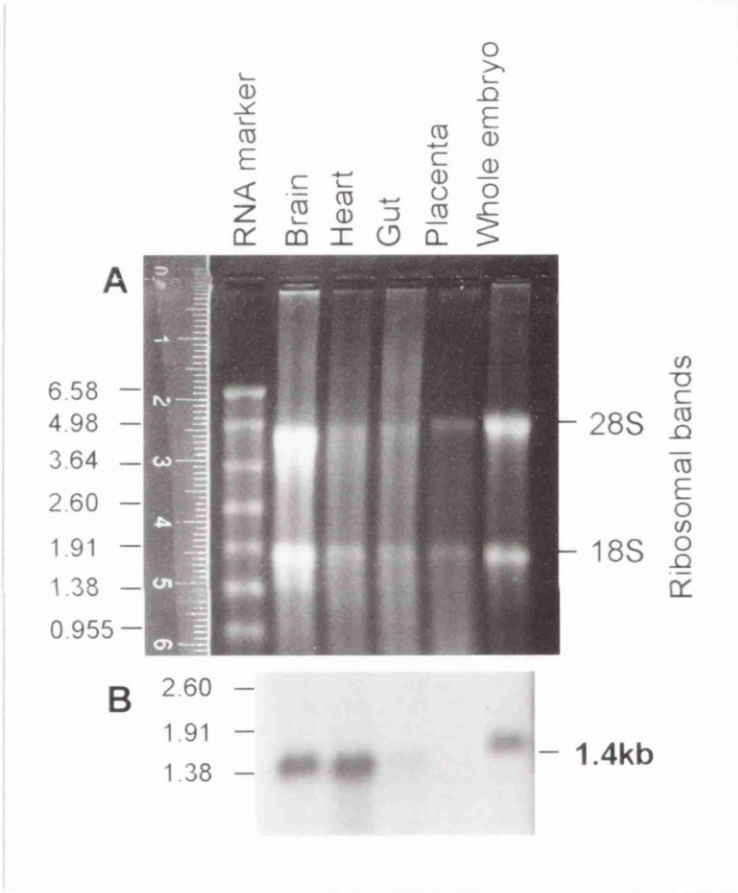


Fig 5.5 Northern analysis of clone 395-H22 in first-trimester fetal samples.

A: Duplicate set of total RNA samples (2μg) electrophoresed alongside those blotted for northern analysis, to check RNA integrity and loading. There was some bias in RNA loading, with an excess of 10 week brain and 7 week embryo samples compared to heart, gut and placenta. An RNA ladder was included as a size reference, with a ruler placed alongside.
B: The 395-H22 insert detected a 1.4kb transcript in 20μg total RNA from a 7 week whole embryo and 10 week fetal brain, heart, and gut, but not in placenta. Expression was strongest in heart, despite unequal loading. The transcript was sized with respect to the RNA marker visualised in **A**. Hybridisation was performed overnight at 65°C. Post-hybridisation washes were performed at 3X SSC/ 0.1% SDS at 65°C for 20 mins. Autoradiography was performed for 72 hours.

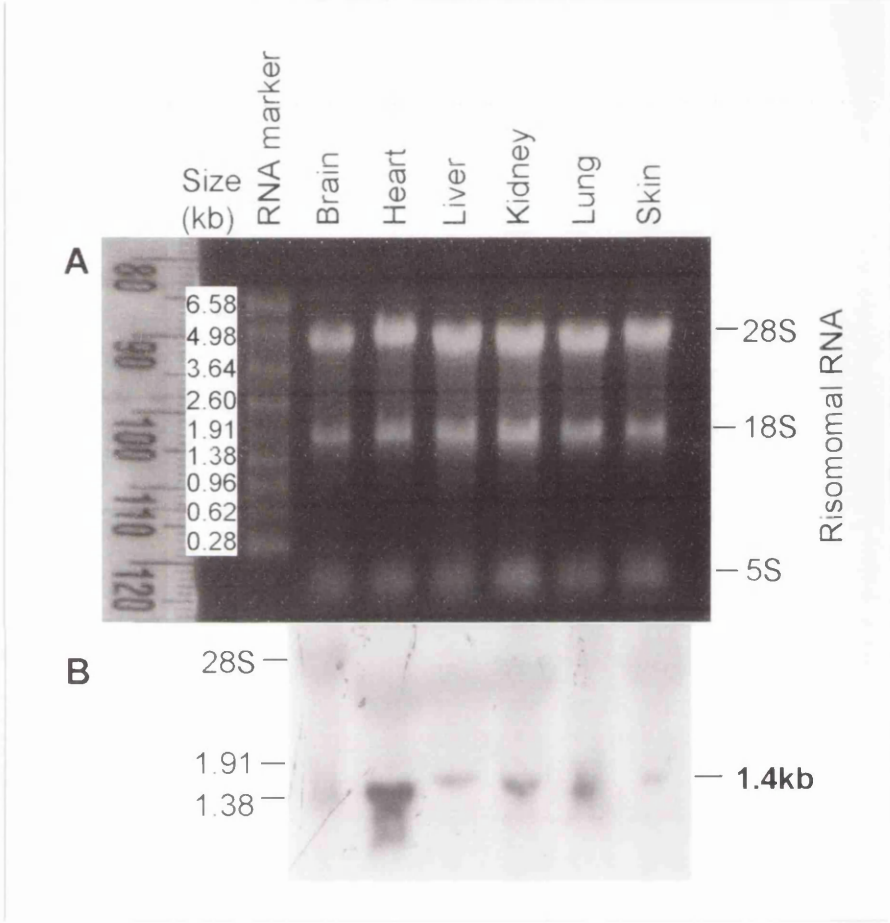


Fig 5.6 Northern analysis of 395-H22 in 15 week gestation fetal tissues.

A: Duplicate set of total RNA samples (5μg) electrophoresed alongside those blotted for northern analysis. RNA integrity was confirmed by visualisation of the 28S and 18S ribosomal bands, and there was even loading of the samples. An RNA marker was included for RNA sizing. The migration distances of the bands of the ladder were measured with a ruler.

B: Autoradiograph of northern blot containing 50μg total RNA samples (as labelled), probed with the 395-H22 insert. A 1.4kb signal was detected in each sample, with the strongest signal seen in heart. The transcript was sized with respect to the RNA marker and ribosomal bands. Hybridisation was performed overnight at 42°C, and post-hybridisation washes carried out to 0.5X SSC/ 0.1% SDS at 42°C. The blot was autoradiographed for 1 week.

RT-PCR analysis confirmed that 395-H22 is expressed widely in fetal tissues, with the highest levels seen in heart, but no expression was detected in placenta (data not shown). RT-PCR analysis showed that 395-H22 was expressed in adult fibroblast cell lines (Fig 5.7). PCR amplification of genomic DNA from a normal individual with H22-5' and H22-3' gave the same sized product as the cDNA, indicating that 395-H22 is contiguous with genomic DNA (Fig 5.7).

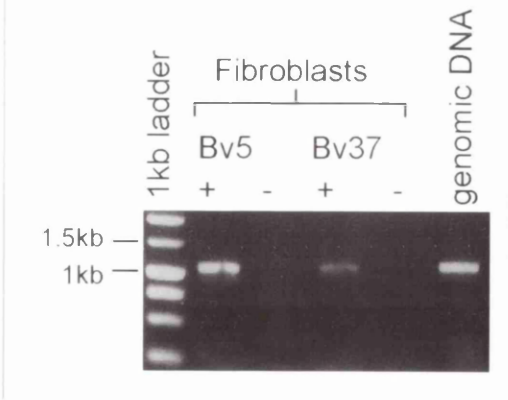


Fig 5.7 Detection of expression of 395-H22 in adult fibroblasts by RT-PCR analysis.

RT-PCR was performed on fibroblast cell lines Bv5 and Bv37, using primers H22-5' and H22-3' to give a 1kb product. These RT-PCR products were sequenced. RT minus reactions are labelled (-). Amplification of genomic DNA also gave a 1kb band.

The RT-PCR amplification products from the fibroblasts (Fig 5.7) and a fetal brain sample were sequenced with primers H22-5', H22-S1 and H22-3' (Fig 5.1), to confirm the cloned sequence, and that the expected cDNA had been amplified by RT-PCR. The Bv5 RT-PCR product was identical to the cloned 395-H22 sequence (Fig 5.1). However, a homozygous **C→T** change was detected at nucleotide position 838 (with respect to the sequence in Fig 5.1) in the fetal brain sample. Three homozygous changes were identified in the Bv37 fibroblasts; an **A** insertion at nucleotide 714, a **G→C** change at nucleotide 503, and a **C→A** at nucleotide position 488. These sequence changes most likely represent expressed polymorphisms. It was noted that in the Bv37 and the fetal brain sample, none of the polymorphisms identified was heterozygous, suggestive of monoallelic expression. Unfortunately the corresponding genomic DNA from these samples was not prepared for analysis to determine whether the DNA was heterozygous for the homozygous alterations that were observed in the cDNA. Thus the possibility that 395-H22 is expressed monoallelically was not investigated through analysis of polymorphisms in the DNA and RNA.

5.2.2 Analysis of expression of sequence 123-E19

The insert from clone 123-E19 failed to detect a transcript on northern analysis using 50μg of total RNA from 15 week gestation fetal brain, heart, liver, kidney, lung and skin, followed by autoradiography. No signal was detected with 123-E19 using 60μg total RNA from 17 week fetal brain either, indicating that this sequence is either not expressed, or expressed at low levels in the tissues tested. Northern analysis with polyA+ RNA was not performed. However, expression of 123-E19 was detected by RT-PCR analysis in adult fibroblasts (Fig 5.8) and 15 week fetal brain, heart, intestine, liver, eye, limb, and skin (Fig 5.9A), using primers E19-5' and E19-S1 (Fig 5.3). No RT-PCR product was observed in fetal lung or placenta cDNAs, although the control cDNA *HPRT* was detected in these samples (Fig 5.9B), indicating lack of 123-E19 expression in these tissues. PCR amplification of genomic DNA using primers E19-5' and E19-S1 amplified the same sized product as the cDNA (1kb), showing that the first 1kb of the 123-E19 sequence is contiguous with genomic DNA (Fig 5.9A). To show that the 123-E19 RT-PCR products were due to amplification of a cDNA template as opposed to amplification of contaminating genomic DNA, RT negative

control samples were included in the analysis. No product was detected in the RT minus samples for 123-E19 (Figs 5.8 and 5.9A), or *HPRT* (5.9B). These data show that 123-E19 is expressed in a multitude of fetal tissues and in adult fibroblast cell lines, though at low levels undetectable by northern analysis.

The RT-PCR amplification products from the fetal brain and fibroblast Bv5 cDNAs were sequenced with primers E19-5', E19-S1 and E19-S2 (Fig 5.3), which confirmed the cloned cDNA sequence and showed that the correct sequence had been amplified by RT-PCR. Bv5 gave the same sequence as the cloned 123-E19 insert sequence. A homozygous deletion of a **T** at nucleotide 927 was identified in the fetal brain sample, presumably representing an intragenic polymorphism. The corresponding genomic DNA sample was not analysed to determine whether it was heterozygous for this change, and so the possibility that the 123-E19 sequence is monoallelically expressed was not investigated by this means.

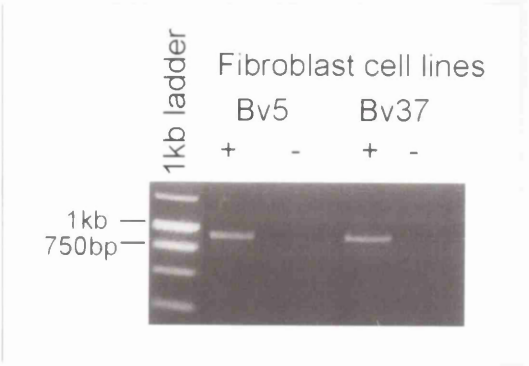


Fig 5.8 Detection of expression of 123-19 in adult fibroblasts.

The 123-E19 sequence was amplified from fibroblast Bv5 cDNAs using primers E19-5' and E19-S1, indicating that 123-E19 is expressed in fibroblasts. The RT negative reaction is blank, showing there was no genomic contamination.

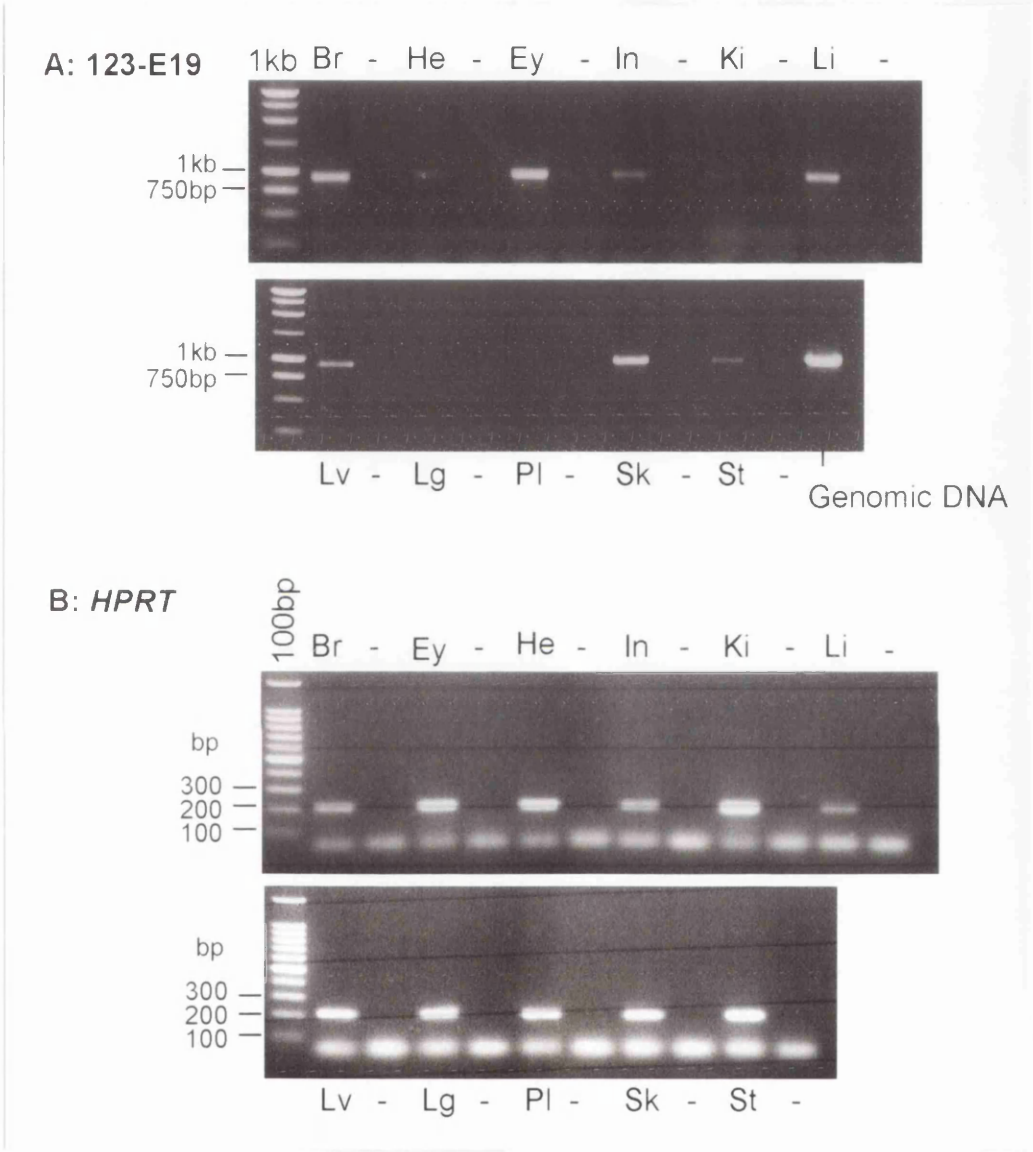


Fig 5.9 Detection of expression of 123-E19 in fetal tissues by RT-PCR

RT-PCR was performed on cDNAs extracted from 15 week gestation fetal tissues (as labelled). **A**: 123-E19 amplified with primers E19-5' and E19-S1. Amplification of genomic DNA gave the same sized product as the cDNA. **B**: Amplification of *HPRT* was performed across an intron using primers listed in Table 2.9. RT minus (-) control reactions were included to detect any genomic contamination. This is detectable as a 1.5kb product in the *HPRT* reactions. These were blank, demonstrating that amplification of 123-E19 occurred from a cDNA template. PCR amplification was performed for 35 cycles.

5.3 Analysis of the cloned inserts for evolutionary conservation

5.3.1. Clone 395-H22

The insert from I.M.A.G.E. clone 395-H22 cross-hybridised strongly to mouse genomic DNA (Fig 5.10), indicating that it is conserved in evolution, even though no protein-coding sequence was identified within the clone (Fig 5.2).

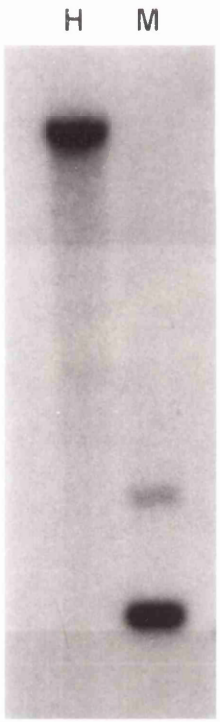


Fig 5.10 Clone 395-H22 is conserved in mouse genomic DNA.

Southern blot containing 5μg human (H) and mouse (M) genomic DNA digested with *EcoR*I and probed with the insert from I.M.A.G.E. clone 395-H22. Post-hybridisation washes were performed to 0.1X SSC/ 0.1% SDS at 65°C, and exposed to X-ray film for 48 hours.

5.3.2. Clone 123-E19

The insert from clone 123-E19 cross-hybridised to mouse genomic DNA, demonstrating that this sequence is phylogenetically conserved (Fig 5.11). This supports the RT-PCR results indicating that the 123-E19 sequence is transcribed, although no signal was detected by northern analysis of 123-E19, and no translated sequence was predicted for the 1.4kb insert (Fig 5.4).

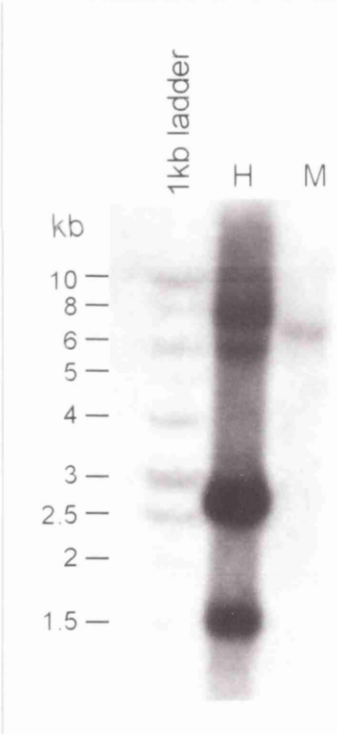


Fig 5.11 Evolutionary conservation of clone 123-E19.

Southern blot of *EcoR*I digested human (H) and mouse (M) genomic DNA hybridised with the insert from I.M.A.G.E. clone 123-E19. The blot was washed to 2X SSC/ 0.1% SDS at 65°C and autoradiographed for 3 days.

5.4 Southern analysis to detect allele-specific methylation

As described in section 3.3.1.3. the inserts of clones 395-H22 and 123-E19 were investigated for parental allele-specific methylation, by Southern analysis using the methylation sensitive enzyme *Hpa*II. PWS and AS DNA, from subjects with UPD15 or 15q11-13 deletions, were used to distinguish between the maternal and paternal 15q11-13 alleles.

5.4.1 Methylation test with clone 395-H22

395-H22 identified a single fragment of >9kb band in *Hind*III and *Hind*III / *Hpa*II digested DNA, which was absent in the *Msp*I digested DNA. A 3.6kb fragment was identified in the *Msp*I DNA, showing that any *Hpa*II/*Msp*I recognition sites within the *Hind*III fragment identified by 395-H22, are methylated. An additional ~ 6.7kb band was detected in the *Hpa*II digested PWS (mUPD15) DNA, but not in AS (pUPD15) DNA (Fig 5.12). However the ~6.7kb band was weak in intensity, and furthermore the derivative >9kb *Hind*III fragment was still present in the PWS DNA, indicating partial digestion of the derivative 9kb band at one or more unmethylated *Hpa*II sites on the maternal allele. Similar results were obtained for PWS and AS DNA with deletions of 15q11-13 (data not shown). This indicates that 395-H22 is associated with a partial differential methylation pattern between the maternal and paternal alleles, but that this is mosaic or heterogeneous, and not ubiquitous in lymphocytes. This may not be functionally relevant in terms of the imprinting status of 395-H22.

5.4.2 123-E19 methylation test

123-E19 identified a fragment of 4.2kb in *Hind*III and *Hind*III/*Hpa*II digested DNA in all samples, which was not observed in the *Msp*I lane. A 3.6kb *Msp*I fragment was identified, indicating that the *Hpa*II/*Msp*I sites within the derivative *Hind*III fragment were methylated in normal DNA. No differences were detected in the *Hpa*II banding patterns between PWS and AS UPD DNA (Fig 5.13) or deletion-positive PWS or AS DNA (data not shown). Clone 123-E19 is therefore not associated with any parent-of-origin methylation of *Hpa*II sites. However, if this sequence lies at the 3' end of a gene, it would be unlikely to be associated with functional methylation differences between the parental alleles.

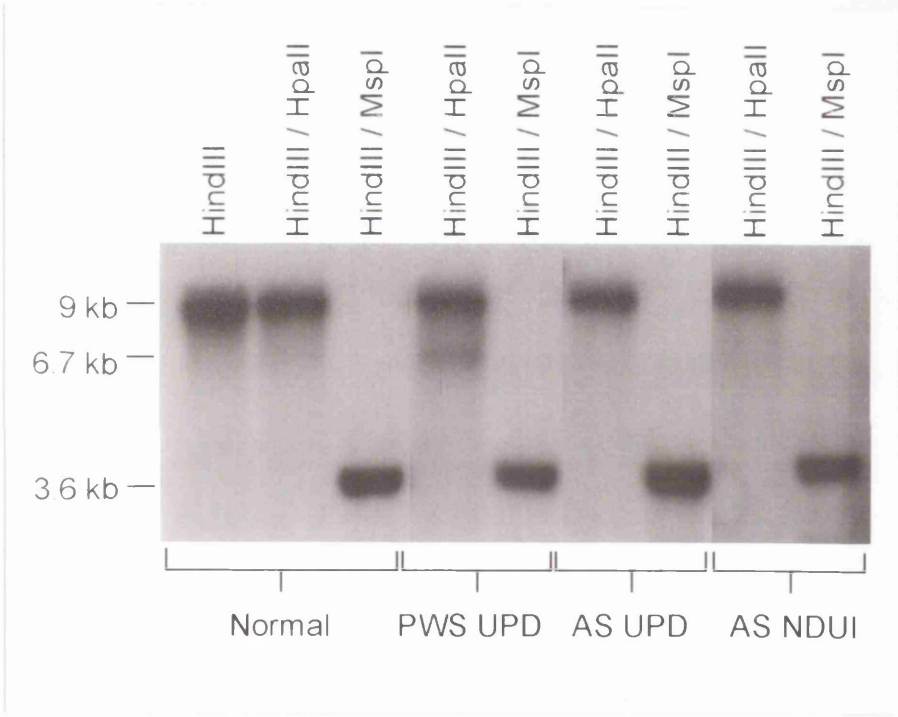


Fig 5.12 Partial differential methylation associated with 395-H22.

Southern blot containing normal, PWS (mUPD15), AS (pUPD15) and AS (non-deletion, disomy or imprintor mutation) DNA digested with *Hpa*II or *Msp*I combined with *Hind*III (as labelled), hybridised with the insert from I.M.A.G.E. clone 395-H22. 395-H22 detected a >9kb band in *Hind*III and *Hind*III/*Hpa*II-digested DNA of all individuals, and a 3.6kb *Msp*I fragment. A 6.7kb *Hpa*II fragment was identified in PWS DNA but not AS DNA. Fragments were sized with respect to a DNA ladder electrophoresed alongside the gel (not shown).

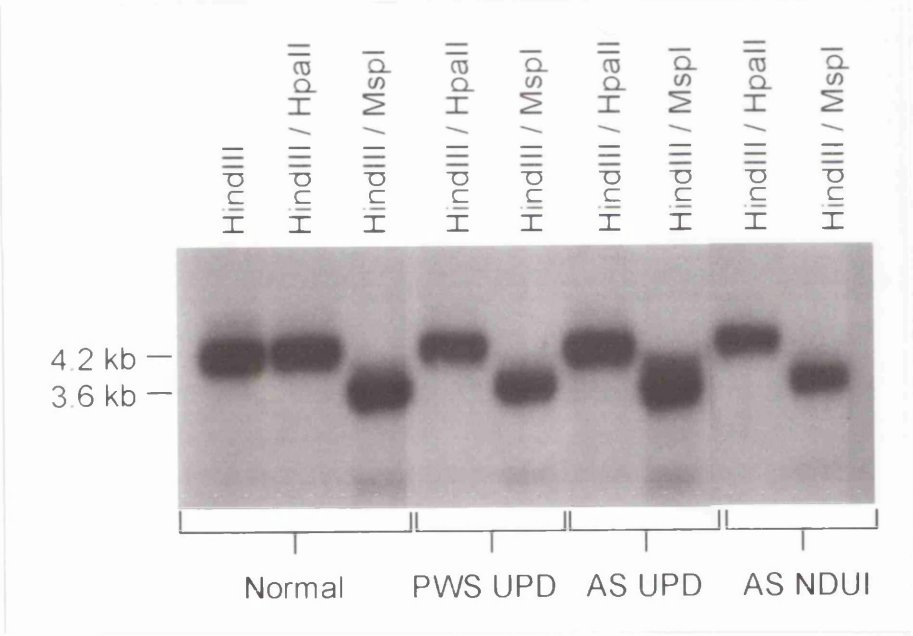


Fig 5.13 Clone 123-E19 has no *Hpa*II methylation imprint

Southern blot containing PWS, AS and normal control DNA (as labelled) digested with a combination of *Hind*III and *Hpa*II or *Msp*I, probed with the insert from I.M.A.G.E. clone 123-E19. A fragment of 4.2kb was identified in *Hind*III and *Hpa*II/*Hind*III digested DNA. A 3.6kb fragment was identified in *Msp*I digested DNA. No differences were observed between the AS and PWS DNA digested with the methylation sensitive enzyme *Hpa*II.

5.5 In vitro imprinting assay using human monochromosomal 15 x rodent hybrids

To investigate whether 395-H22 and 123-E19 were imprinted, an *in vitro* imprinting assay employing human monochromosomal 15 x rodent somatic cell hybrids was used, as described previously in section 3.3.1.4. This assay involved RT-PCR analysis of the hybrid clones, which contained either a maternal or paternal human chromosome 15. The human component was originally derived from fibroblast cells, and both 395-H22 and 123-E19 had been shown to be expressed in fibroblast cell lines (Figs 5.7 and 5.8), and so imprinting analysis of these two sequences using this assay system was feasible.

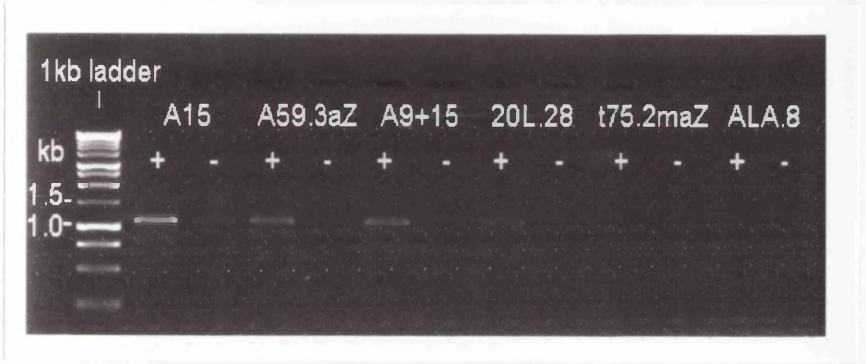
5.5.1 395-H22

RT-PCR analysis of sequence 395-H22 using the set of human monochromosomal 15 hybrids provided an indication that this transcript is imprinted, with expression restricted to the paternal allele. Amplification with primers H22-5' and H22-3' (Fig 5.1), gave the expected 1.1kb product in the three hybrid clones containing a paternally derived chromosome 15, A15+, A59.3aZ and A9+15 (Fig 5.14A). A weak RT-PCR product was observed in hybrid 20L.28 (Fig 5.14A), which contained a maternal 15, but this was considered unreliable since *SNRPN* expression had been detected in this cell line (Fig 5.14B). No RT-PCR product was visualised for t75.2maZ, as expected, since no product was detected for the biallelic control *UBE3A* in this hybrid (Fig 5.14D). However, expression was not detected for 395-H22 in the ALA.8 hybrid (Fig 5.14A), which contains a maternal 15 and did express *UBE3A* (Fig 5.14D).

The overall expression profile seen in the hybrids for sequence 395-H22 was similar to that of the maternally imprinted control *SNRPN* (Fig 5.14B), providing evidence for expression specifically from the paternal allele. This would make the transcript from which cDNA clone 395-H22 is derived a candidate for involvement in the PWS phenotype. However, these are preliminary data based on the absence of expression of 395-H22 in a single hybrid containing a maternal chromosome15, and require confirmation using an alternative system for analysis of imprinting. This could be achieved by analysis of PWS and AS cell lines, or by tracing the allelic origin of expressed polymorphisms (such as those identified in section 5.2.1) in heterozygous samples.

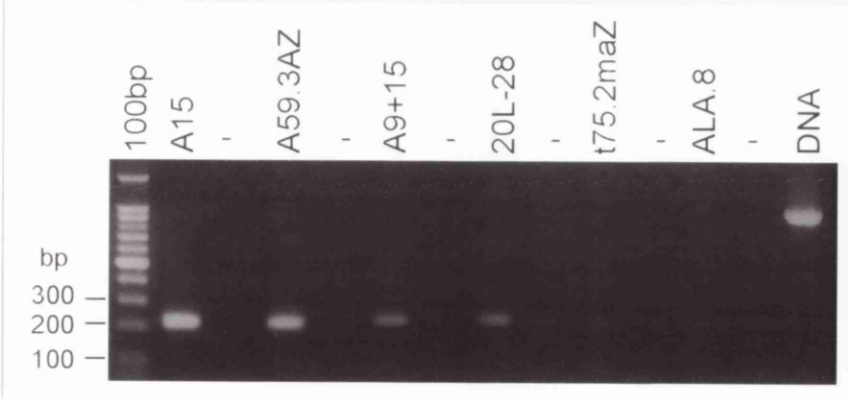
5.5.2 123-E19

A ~1kb RT-PCR amplification product using primers E19-5' and E19-S1 (Fig 5.3) was obtained in the cDNAs from each hybrid cell line except t75.2maZ (Fig 5.14C). Thus RT-PCR analysis using the monochromosomal 15 hybrid assay system indicated that 123-E19 is expressed from both parental alleles in cell lines derived from fibroblasts, as was the case for the *UBE3A* control (Fig 5.14D). These data provide evidence that 123-E19 is not imprinted, but are subject to confirmation using other imprinting assays, such as RT-PCR of PWS and AS patient cell lines, or analysis of expression in different tissues using an expressed polymorphism. If 123-E19 is transcribed from both parental alleles in all tissues in which it is expressed, it may be ruled out as a candidate for involvement in the PWS phenotype.



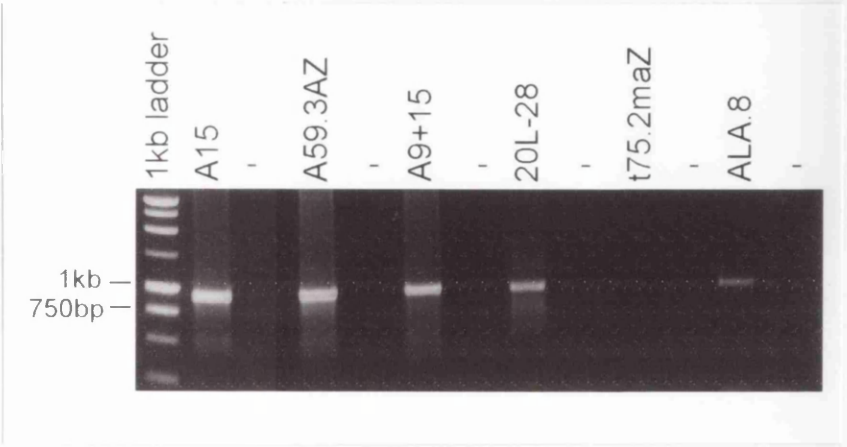
A 395-H22 is expressed in hybrids containing the paternal chromosome 15

Primers H22-5' and H22-3' were used to amplify a 1.1kb product.



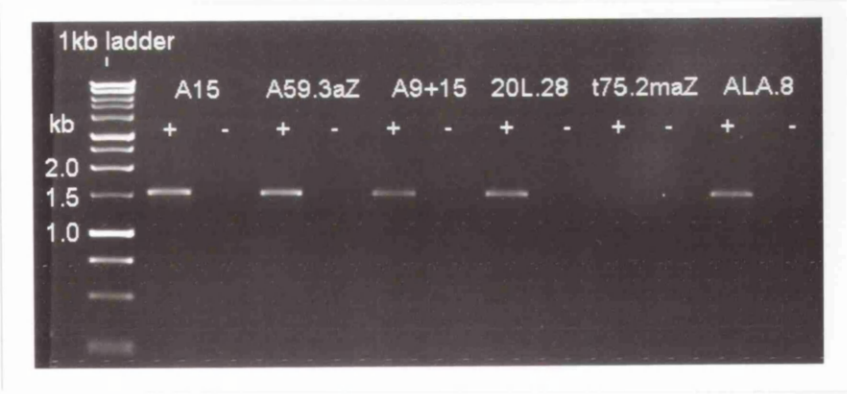
B *SNRPN*; paternally expressed imprinted control

RT-PCR was performed using primers N1 and N4. Amplification due to genomic DNA contamination is distinguishable from the cDNA as a 1.2kb product in the RT- reactions.



C 123-E19 is transcribed from both parental chromosomes 15

Primers E19-5' and E19-S1 were used to amplify a 1kb product.



D *UBE3A*; biallelically expressed control

Amplification was performed using primers E6AP-4F and E6AP-8R.

Fig 5.14 *In vitro* imprinting assay of sequences 395-H22 and 123-E19 using human monochromosomal 15 x rodent hybrids

Hybrids A15, A59.3aZ, A9+15 contain a paternal chromosome 15. Cell lines 20L.28, t75.2maZ, ALA.8 contain a maternal 15. Analysis of imprinting was performed by RT-PCR using cDNAs from the monochromosome 15 hybrids. RT reactions are indicated with a plus (+). Control reactions for the detection of genomic contamination, from which RT was excluded, are labelled minus (-).

Chapter 6: Discussion

6.1 Identification and characterisation of the Angelman syndrome gene

Angelman syndrome is a neurobehavioural disorder, which in 80% of cases is due to maternal deletion of 15q11-13, pUPD15 or epigenetic lesion of the maternally inherited 15q11-13 allele. These genetic mechanisms indicated that AS is an imprinted disorder, due to loss or disruption of one or more maternally expressed imprinted genes within 15q11-13. Approximately 20% of AS patients, including familial cases, have no deletion, UPD or imprintor mutation (NDUI). Linkage to D15S10 and *GABRB3* was established in several AS pedigrees, in which imprintor mutations were subsequently ruled out by methylation testing. Maternal transmission of AS in these families indicated that this gene was expressed solely from the maternal allele. Identification of the AS gene was important in elucidating the pathophysiology of this disorder and would enable accurate genetic counselling and prenatal diagnosis to be performed in AS families.

Through the analysis of an AS family with an atypical microdeletion and a non-AS family with a deletion involving 15q11-13, a 1Mb region between D15S174 and D15S113 was defined as the minimal region in which the AS gene was most likely to be located (Fig 1.9). The *UBE3A* gene was mapped to this region, near to D15S10 (Nakao *et al.*, 1994; section 3.1.2). *UBE3A* was identified as the AS gene with the finding of inactivating mutations in four NDUI AS patients (Kishino *et al.*, 1997; Matsuura *et al.*, 1997). Loss-of-function mutations have now been described in over 40 index cases of AS. The mutations in familial cases were shown to be maternally transmitted, consistent with the matrilineal mode of inheritance of AS (Malzac *et al.*, 1998; Fang *et al.*, 1999).

6.1.1 Positional cloning efforts in the ASCR fail to identify any genes other than *UBE3A*

6.1.1.1 Physical mapping and sequencing within 15q11-13

The PWS and AS candidate regions have now been cloned in PAC or BAC vectors. A PAC and BAC contig, consisting of a minimal set of overlapping clones, was constructed across the ASCR in this project. This contig was anchored to the YAC map previously reported by Mutirangura *et al.*, 1993b, by mapping of PAC and BAC clones to the YACs, and STS mapping. ESTs WI-6519, WI-6654 and WI-16777, which had been imprecisely located to a 6-14cM region between *SNRPN* and *GABRG3* on the Whitehead/MIT radiation

hybrid map, and the *UBE3A* gene were placed on the PAC contig, locating them within the ASCR. The preliminary PAC contig and mapping data was reported at The third International Workshop on Human Chromosome 15 Mapping in 1996 (Robinson and Knoll, 1997) and the updated contig was presented at the European Society for Human Genetics, 1997 (Fairbrother *et al.*, 1997). A PAC contig spanning the PWSCR from D15S63 to *UBE3A* was subsequently reported, which overlapped the contig described in this report (Sutcliffe *et al.*, 1997). PAC clone 14I12 was common to both contig maps. The two contigs thus encompassed the entire PWSCR and the ASCR. Each clone is publicly available from the HGMP as a resource for research purposes.

The entire 15q11-13 interval between the cytogenetic PWS and AS deletion breakpoints at D15S18 (centromeric end) and D15S12 (telomeric end) (Fig 1.8), has now been intensively mapped. A YAC contig consisting of 118 YACs, including the original clones identified by Mutirangura *et al.* (1993b), has been constructed across this region. Two gaps that existed in the original contig have now been filled. This physical map has integrated all the mapping data from the CEPH YAC map, the WI radiation hybrid map and Genethon genetic map, correcting several discrepancies that existed between the different sources of mapping information (Christian *et al.*, 1998). The STSs and ESTs placed on the PAC and BAC contig in the ASCR (sections 3.1.1.4 and 3.1.2) are concordant with this integrated map. Christian *et al.* mapped an additional 5 ESTs to the interval between D15S174 and D15S113. The average physical distance between markers on the new map is 35 kb, considerably improving the previous STS spacing of 200 kb (section 1.6). This will provide the framework for sequencing this entire region, and will facilitate identification of other disease genes from the region (Christian *et al.*, 1998).

In addition to AS, PWS and OCA2, several other diseases have been linked to 15q11-13. These include spastic paraplegia 6 (Fink *et al.*, 1995), obsessive-complusive disorder (Dykens *et al.*, 1996), autism (Cook *et al.*, 1997), and bipolar disease (Edenburg *et al.*, 1997). The complete YAC map will also enable the molecular basis for the predisposition of the regions around D15S18 and D15S12 to deletion to be analysed. In the previous YAC map these regions were not cloned, possibly reflecting their instability (Christian *et al.*, 1998).

Efforts to sequence the entire 4Mb classical PWS and AS deletion region, starting with the ASCR and PWSCR, are currently underway at the Genome Science and Technology Center, University of Texas, USA and the Wellcome Trust Genome Campus,

Cambridgeshire, UK. PAC and BAC clones from the contig reported here, and by Sutcliffe *et al.*, 1997, are being used as sequencing template. Sequencing of PAC 14I12, which contains part of *UBE3A* (Fig 3.1.7), was completed in June 1998 and is available in Genbank (accession number AC004259). Sequencing of the entire deletion region will enable additional disease genes to be identified using computational analysis, which has the power to predict the positions of coding sequences within genomic DNA, combined with laboratory analysis.

6.1.1.2 Search for coding sequences within the ASCR.

UBE3A is the only protein-coding gene that has been identified within the ASCR, despite intensive positional cloning efforts in this region. This was the case in the application of direct cDNA selection in this region, as described in this project. I.M.A.G.E. clones 215-B4 and 1354-G13 were novel sequences at the time of identification, but were later shown to be part of the *UBE3A* transcript (discussed in section 6.1.2). The cDNA selection protocol proved to be sensitive and specific, isolating sequences from both previously known genes from the target genomic region; *IPW* and *UBE3A*. In addition, the PAR2 transcript and ESTs WI-6654 and WI-15655 were identified. *UBE3A* served as the positive control for the direct selection procedure, since it had been mapped to the region of interest and was represented in the starting cDNA library. Coding exons 9 (previous exon 3) and 10-14 (previous exons 4-8), 5' UTR exons U3b and U4, and the extended 3' UTR sequence of *UBE3A* were identified amongst the selected cDNA clones, either by sequencing or by hybridisation. The first estimation of the frequency of *UBE3A* clones in the 132D4-selected cDNA sublibraries was 0.67% (Table 3.2.2). With the demonstration that I.M.A.G.E. clone 1354-G13, identified by BLAST analysis of redundant selected sequences 79 and 89 from this sublibrary was also part of *UBE3A*, this increased the frequency of *UBE3A* clones in the 132D4 sublibrary to 1.33%. Thus the direct selection process identified several exons of the positive control gene in addition to other genes and transcripts previously identified from the region investigated. It is possible that other genes in the region exist which are not expressed in 9 week gestation fetal brain, which was used as the tissue source for isolation of cDNAs. The main limitation of direct cDNA selection is that the isolation of a transcript is dependent on its expression in the source of cDNAs. However, since both AS and PWS are primarily neurobehavioural disorders, any genes involved in the pathophysiology of these disorders were predicted to be

expressed during fetal brain development.

It is possible that the means of analysis of the selected sublibraries and clones identified in the positive screen was not comprehensive and that novel transcripts in the region were isolated by the direct selection procedure but were not recognised as cDNAs.

A significant number of clones were not identified in either the 'counterscreen' as repetitive, vector or ribosomal sequences, or in the 'positive' hybridisation screen. These may have represented low copy-number sequences which did not map to the target genomic contig (Lovett *et al.*, 1994), or clones mapping to 15q11-13 that failed to hybridise in the positive screen. These clones were not characterised due to time constraints and the large sequencing effort this would have entailed.

An antisense RNA to *UBE3A* has been identified, which is transcribed from exons 16 to 11 at the 3' end of *UBE3A,* but may start further downstream. The antisense transcript includes the intervening intron sequences of *UBE3A*. This transcript is expressed specifically in the brain, but was found to be absent in PWS brain, indicating that it is paternally expressed (Rougelle *et al.*, 1998). Single-copy selected sequences 21, 39 and 86 from the ASCR (section 3.2.2) may have represented sequences from this antisense RNA, as BLAST analysis after June 1998 showed that these sequences were located within the genomic region spanned by the antisense transcript, in PACs 14I12 and 373B1. This was only determined retrospectively, after the *UBE3A* antisense RNA had been described. By 1996-1997 an estimated 80% of all genes were represented by ESTs in the sequence databases (Schuler *et al.*, 1996; Schuler 1997). To date nearly 1.5 million *Homo sapiens* ESTs are available in the dbEST database. Despite this, BLAST analysis of query sequences 21, 39 and 86 failed to identify any ESTs from the sequence databases. Thus the antisense RNA may have been represented amongst the single-copy selected sequences isolated, but the limited analysis of these sequences failed to lead to the identification of the *UBE3A* antisense RNA.

No transcripts were identified by cDNA selection from the region encompassed by YACs 230H12 and 142A2, between D15S210 and D15S113. The same reaction conditions were used for all the probes and there was some redundancy in the probes used as targets for selection in this interval, including YAC DNA purified by PFGE, the inter-*Alu* PCR products of the YACs, and purified PAC clones from the PAC-2 contig. This makes it unlikely that the failure to select any transcripts in this region was due to the practical conditions employed in this technique or poor integrity of the target probes. Since the yields of selected products

using probes 230H12 and 142A2 were initially low (Fig 3.2.7), selection reactions were performed under less stringent hybridisation and washing conditions, and with double the amount of starting cDNAs in the first round. Although this improved the quantity of products selected, these were found to be repetitive or vector sequences (Table 3.3.1). The region surrounding D15S113 was also analysed for evolutionary conserved fragments which might lead to the identification of expressed sequences in the region. A single conserved 1.1 kb *Pst*I fragment was identified just distal to D15S113. However, sequence analysis and subsequent cDNA library screening using a segment of this sequence that was potentially coding, failed to identify any cDNAs. P1.1-6 was found to contain a tetranucleotide repeat sequence and identified the microsatellite STS UT7715 (Genbank accession number L30335) by BLAST analysis. This STS was subsequently placed on an integrated YAC, STS and EST map distal to D15S113, and assigned the locus number D15S540 (Christian *et al.*, 1998).
This suggested that the interval between D15S10 and D15S113 may be relatively gene-poor. The entire AS and PWS region within 15q11-13 is reportedly repeat-rich (Donlon *et al.*, 1986; Christian *et al.*, 1998). This was evinced by the close proximity of Alu repeat sequences in YACs 132D4, 230H12, 142A2 and 368A12. *Inter*-Alu PCR demonstrated that Alu repeats in this region were 100-600bp apart (Fig 3.2.4A). A previous estimation of *inter*-Alu spacing in the human genome was 5kb on average, although spacing of 700bp to 8kb was reported (Schmid and Jelinek, 1982). However, the presence of a large number of direct inverted repeats is a feature of imprinted regions (Bartolomei, 1994). This was a confounding factor for positional cloning efforts in the region. P1.1-6 contained a repeat sequence (section 3.3.3.4), and the vast majority of selected sequences analysed from the region were also repetitive despite quenching with human DNA or Cot1 DNA (Table 3.3.1).

The application of a single positional cloning strategy will invariably fail to isolate every coding sequence within a genomic region of interest. However, various positional cloning techniques have been applied in the ASCR by several groups, and yet *UBE3A* remains the only protein coding AS candidate gene identified. Analysis of the ASCR for evolutionarily conserved genomic fragments originally led to the identification of *UBE3A* and the non-coding PAR2 transcript from the region. Pooled cDNAs from mouse cDNA libraries were used as hybridisation probes to identify conserved genomic fragments from cosmids and phage clones from human 15q12 by Southern analysis. A conserved *Eco*RI fragment was identified within a cosmid from the ASCR using a mouse brain cDNA library

as probe. The fragment was used in turn to screen a human cDNA library, which led to the identification of *UBE3A*. Similarly, a conserved genomic fragment identified by mouse liver and testis cDNAs led to the isolation of the PAR2 transcript. No other conserved fragments were identified within the ASCR, although others were identified from the PWSCR (Nakao *et al.*, 1994).

Analysis of 15q11-13 by PFGE restriction analysis with rare-cutting enzymes identified several CpG islands within the 4Mb PWS and AS deletion region, which were subsequently associated with the genes *ZNF127*, *P* and the $GABA_A$ receptor subunit cluster (Kirkilionis *et al.*, 1991; Sinnett *et al.*, 1993). Within the ASCR however, only one CpG island was identified at D15S10 (Kirkilionis *et al.*, 1991). This was subsequently cloned as OP2 (Woodage *et al.*, 1994b), and shown to contain the first exon of *UBE3A* (Kishino *et al.*, 1997). One additional CpG island 600kb distal to D15S174 has been reported, but has not been associated with any gene to date (Bürger *et al.*, 1997b). Although about 40% of genes are not associated with CpG islands, they represent a marker for the presence of genes, suggesting a paucity of genes within the ASCR (Antequera and Bird, 1993). Since the vast majority of imprinted genes have a differentially methylated CpG islands in the promoter region (Neumann and Barlow, 1996), any imprinted gene involved in AS or PWS is likely to be associated with a CpG island.

Exon trapping has been applied to the ASCR, with a multitude of potential exons identified (Bürger *et al.*, 1997b; Kishino *et al.*, 1997). Four of the 5' untranslated exons of *UBE3A*, exons U1-U4 (renumbered exons 3-6), were identified from between OP2 and the coding exons of *UBE3A* using this method (Kishino *et al.*, 1997). However, no additional genes have been described so far as a result of exon-trapped products.

Although each positional cloning strategy has its limitations, for instance exon-trapping is not suitable for the identification of intronless genes, and cDNA selection relies on the expression of genes in the RNA source, the application of several different techniques to the ASCR has failed to identify any gene in addition to *UBE3A*. Direct sequencing of the PWSCR and ASCR is currently underway, and will greatly facilitate the identification of any other genes that may exist in the region.

6.1.2 Characterisation of *UBE3A*

At the start of this project, *UBE3A* had been localised to YACs 132D4 and B230E3 by PCR-mapping of a short sequence within exon 9, and expression from both parental alleles in fibroblasts and lymphoblasts had been demonstrated. Northern analysis had shown that *UBE3A* transcribed a ~5kb mRNA, which was most abundantly expressed in skeletal muscle, heart and liver, but was detectable in most tissues (Nakao *et al.*, 1994). However, just 2.65kb of the cDNA sequence had been identified (Huibregtse *et al.*, 1995). To further characterise *UBE3A* the coding exons were more precisely located to the PAC and BAC contig across the ASCR (Fig 3.1.7). The gene was shown to span 100-200 kb within the ASCR, on the basis that exons 7-14 (previous exons 1-8) mapped to two overlapping PACs of average size 110kb. The direction of transcription was shown to run from telomere to centromere (section 3.1.2) (Fairbrother *et al.*, 1997; Robinson *et al.*, 1997).

The *UBE3A* gene has now been well characterised. The full-length transcript has been identified. Several untranslated exons were identified at the 5' end of the gene. Amplification of the 5' region of *UBE3A* by 5' RACE, and analysis of cDNA clones, has found evidence for multiple splice forms of *UBE3A* (Kishino *et al.*, 1997; Yamamoto *et al.*, 1997; Rougelle *et al.*, 1997; Vu and Hoffman, 1997; Kishino and Wagstaff, 1998). Five *UBE3A* subtypes, generated by differential splicing of the 5' untranslated exons U1-U4 onto the coding exons, were first described. These had the potential to encode 3 different protein isoforms (Yamamoto *et al.*, 1997). *UBE3A* subtypes 6 and 7 were identified specifically in brain (Rougelle *et al.*, 1997). A further 2 major subtypes were identified in fetal tissues, which contain novel untranslated sequences OP2a and b, and U3a and b. The U3 exons are contiguous with genomic DNA, with U3a upstream of U3, and U3b downstream. These subtypes are most likely produced by alternative splicing of exons OP2 and U3, and are thought to encode one of the three protein isoforms (Vu and Hoffman, 1997). The specific function of the different splice forms remains to be elucidated. For each subtype the 5' UTR sequence was less than 1kb in length, with over 1.4kb of *UBE3A* sequence remaining to be identified, to acquire the full-length cDNA.

Direct cDNA selection in the ASCR led to the identification of an additional 1.85kb of 3' untranslated sequence for *UBE3A*, accounting for the full ~5kb *UBE3A* transcript. BLAST analysis of sequences isolated by direct selection identified two I.M.A.G.E. cDNA

clones 215-B4 and 1354-G13 through sequence homology, which were obtained through the I.M.A.G.E. consortium. The insert sequences of both clones mapped just centromeric to, or overlapping the 3' coding exons of *UBE3A*. Sequence analysis of clone 1354-G13 showed that it was non-coding and polyadenylated with three polyadenylation signals, indicating that it represented the 3' UTR of a derivative transcript. Northern analysis of fetal brain showed that 1354-G13 was part of a 5.2kb transcript, which was strikingly similar to the reported size of the *UBE3A* (Nakao *et al.*, 1994; Sutcliffe *et al.*, 1997). Exon connection RT-PCR with the 1354-G13 sequence and the penultimate exon of *UBE3A* showed that the 1.85kb 1354-G13 sequence was part of the *UBE3A* transcript and contiguous with the previously reported final exon sequence (Huibregtse *et al.*, 1993). The final 2kb of *UBE3A* was also identified by Kishino and Wagstaff, 1998, and so the data reported here was not published. Kishino and Wagstaff reasoned that the missing ~1.5kb of *UBE3A* sequence from the ~5kb transcript could be accounted for by a long 3' UTR. A probe was generated from the genomic region immediately 3' to the previously reported final exon sequence and analysed by northern blotting. This hybridised to a ~5 kb transcript, indicating that the genomic sequence 3' to *UBE3A* was part of the mRNA. The same probe was used to screen a cDNA library. Several clones were identified and sequenced. Two of the clones contained the final nucleotides of *UBE3A* exon 15 and the previous terminal exon sequence upstream of the novel untranslated sequence. Furthermore, these clones showed that alternative polyadenylation signalling did occur, with different clones containing polyadenylation tails directly after each of the three polyadenylation signals identified in the 2kb 3'UTR. The previous 155bp final exon, which contained the first consensus polyadenylation signal 50bp upstream of the polyadenylation tail, represented the shortest *UBE3A* transcript. This was originally sequenced from a cDNA clone (Huibregtse *et al.*, 1993). However, an A-rich region is present just 3' to the first polyadenylation signal, and so it was considered possible that this clone may have resulted from artefactual priming of oligo(dT) to this region during cDNA library construction, and may not occur frequently *in vivo* (Kishino and Wagstaff, 1998).

The genomic organisation of *UBE3A* has also been established. *UBE3A* has been shown to span 120kb (Sutcliffe *et al.*, 1997; Kishino and Wagstaff, 1998) and is associated with the CpG island near D15S10, within the genomic clone OP2 (Kishino *et al.*, 1997). Two putative promoters have been identified in the OP2 region, P1 and P2, from which different splice forms of *UBE3A* are transcribed (section 6.2.1.1)(Vu and Hoffman, 1997). The intron-

exon boundaries have been sequenced, which has been useful for mutation screening in NDUI AS patients from genomic DNA (Kishino and Wagstaff, 1998). An antisense RNA, which encompasses exons 16-11 has also been identified (previous section)(Rougelle *et al.*, 1998).

Two *UBE3A* pseudogenes *UBE3AP1* and *UBE3AP2* have been described which map to chromosomes 2 and 21 respectively, and share 90-95% sequence similarity to different *UBE3A* exons. *UBE3AP1* is homologous to exons 6, 7, 3 and 4, and is flanked by direct repeats. It was suggested that this pseudogene was formed by a circular intermediate. *UBE3AP2* was found to be homologous to exons 9-16, with an *Alu* repeat inserted into the exon 11 sequence. Neither pseudogene sequence was detected in fetal brain RNA, indicating that these were not expressed (Kishino and Wagstaff, 1998). The presence of these two sequences in the human genome was a confounding factor in the screen for mutations of *UBE3A* in NDUI AS patients, as discussed in section 6.3.1.

6.1.3 Evidence that *UBE3A* is the AS gene

Although *UBE3A* was mapped to the ASCR, it was not considered a good candidate for the AS gene because it did not demonstrate the expression or imprinting profile predicted for the AS gene. The major pathological defect in AS most likely occurs during development of the brain and so any gene implicated in AS was predicted to be expressed abundantly in this tissue. However, nothern analysis showed that *UBE3A* was expressed at relatively low levels in the brain compared to skeletal muscle, heart and liver. The gene involved in AS must also demonstrate monoallelic expression from the maternal allele, since AS only results if the genetic defect involving 15q11-13 is maternally transmitted or arises on the maternal allele. Imprinting studies using fibroblasts and lymphoblasts from PWS and AS with cytogenetic deletions or UPD showed that *UBE3A* was expressed from both the paternal and maternal alleles in these cell lines (Nakao *et al.*, 1994). In order to investigate tissue-specific imprinting, the allelic origin of expression of *Ube3a* was investigated in a multitude of tissues from mice with reciprocal disomies for central chromosome 7, by northern and RT-PCR analysis. *Ube3a* expression was detected equally in mice with maternal duplication (PWS mouse model) and paternal duplication (AS mouse model) for this region, in all the major organs including the brain. These observations showed that *Ube3a* was not imprinted

in any one tissue, and were interpreted as meaning *UBE3A* was unlikely to be involved in AS (Sutcliffe *et al.*, 1996).

UBE3A was mapped more precisely to the ASCR in this project, and on that basis was considered a positional candidate for AS. To determine whether *UBE3A* was disrupted or mutated in NDUI AS patients, a 1.5kb portion of the *UBE3A* cDNA from coding exons 10-14 (previous exons 4-8) was used as a hybridisation probe for Southern analysis. No altered fragments were identified in 17 patients tested for any of the restriction enzymes used (section 4.1). This screen was brief and only covered a portion of the *UBE3A* coding region. But since *UBE3A* was considered a poor candidate for AS at the time, the identification of other genes in the ASCR was prioritised over a more thorough search for mutations in *UBE3A*.

6.1.3.1 Maternal mutations of UBE3A cause AS in NDUI patients

The first indication that *UBE3A* might be involved in the AS phenotype came with the observation of an AS patient with a maternally inherited paracentric inversion within 15q11-13, which disrupted the 5' end of *UBE3A* (Kishino *et al.*, 1997). The coding region of the gene was then investigated for pathologic mutations in NDUI AS patients by SSCP and sequence analysis. Heterozygous truncating mutations were first identified in 4 NDUI index cases, implicating *UBE3A* in the pathogenesis of the disorder (Kishino *et al.*, 1997; Matsuura *et al.*, 1997). Loss-of-function mutations in *UBE3A* have now been published in over 40 index cases. In all familial cases reported the mutations were maternally transmitted. In some families the authors were able to demonstrate that the mutations were inherited asymptomatically from the maternal grandfather and were present in other maternal relatives (Malzac *et al.*, 1998; Fang *et al.*, 1999). This was consistent with the matrilineal mode of inheritance of AS. The best example of this is given by a large AS pedigree with 8 affected individuals that was previously used to demonstrate linkage of AS to this region (Meijers-Heijboer *et al.*, 1992). A 4bp deletion causing a frameshift was identified in all affected individuals. The mutation was present in the mothers of all the affected individuals, and numerous other relatives. However, the mutation was transmitted with impunity from several male family members, including the male ancestor (Fang *et al.*, 1999). These cases provided strong evidence that *UBE3A* is the gene involved in the pathogenesis of AS due to loss of *UBE3A* product from the maternally derived allele (Kishino *et al.*, 1997; Matsuura *et al.*,

1997; Malzac *et al.*, 1998; Fang *et al.*, 1999).

We have identified heterozygous *UBE3A* mutations in a further 7 index NDUI cases of AS (Chapter 4). The entire coding region and 5' untranslated exons 1, 4, 5 and 6 of *UBE3A* were screened by SSCP and sequence analysis in 21 sporadic and 5 familial NDUI AS patients referred to the Institute of Child Health for genetic testing. Mutations in the coding region of *UBE3A* were identified in 4 sporadic patients (19%) and 3 familial cases (66%) (Table 4.2). The mutations identified consisted of 5 frameshift mutations in 2 families and 3 sporadic patients, caused by deletions or insertions and a tandem duplication in one case. These were predicted to cause premature chain termination. Protein truncation would result in a catalytically inactive 3' carboxy terminus of E6-AP. Lack of the cysteine residue at amino acid 833 would destroy the ability of E6-AP to form a thioester bond with ubiquitin, and loss of the final 6 amino acids would render E6-AP incapable of transferring the ubiquitin to the substrate (Huibregtse *et al.*, 1995). The truncation mutations described in these cases would therefore completely abolish the ubiquitination activity of the E6-AP products derived from the mutant allele. In 2 cases, the mutations were in-frame deletions, causing loss of the phenylalanine at position 782 in Family A and loss of four amino acids at position 757-760 in sporadic patient O. In both these cases the effect of the missing amino acids on the protein was difficult to predict, since no crucial function has been associated with these particular amino acids. However, in each case the amino acids deleted were within the conserved *hect* domain, which comprises the final 100 amino acids of the E6-AP protein. Conservation of this region in genes from species as diverse as humans, rodents, *Drosophila* and yeast, indicates that it serves an important function. The *hect* domain may be significant in substrate recognition (Huibregtse *et al.*, 1995), and disruption of this region may affect the protein's ability to target the protein substrate(s) for ubiquitination. *In vitro* assays for the ubiquitination activity of E6-AP in the cell lines of these AS patients would determine the level of protein activity, and could provide an indication of how these amino acid deletions effect protein function. These mutations were transmitted from the mothers of the affected individuals and were not identified in over 50 normal individuals and so were confirmed to be pathogenic.

In the 3 familial cases the mutations were carried by the normal mothers. These mutations must have been transmitted silently from the maternal grandfather, or arose *de novo* on the grandpaternal allele. However, no grandparental DNA was available to confirm

this. The demonstration of loss-of-function mutations on the maternal allele in the familial cases provides definitive evidence that these represent the molecular mechanism for disease in these patients. In 3 of the 4 sporadic patients for whom mutations were identified, these were not present in either parent, but arose *de novo*. This indicates that the sequence changes were pathogenic in these 3 patients. In sporadic patient O, the mutation was identified in the maternal DNA, indicating a sibling recurrence risk of 50%. An unaffected sibling did not carry the mutation. The clinical implications of this case are discussed further in section 6.3.

There appears to be no hot-spot for *UBE3A* mutations described in the literature, and none of the mutations identified in the 7 AS cases from the Institute of Child Health have been reported by any other group. With one exception, the *UBE3A* mutations described in the literature have occurred in the coding region of the gene. A minority of these are missense mutations for which the effect on protein function has been difficult to predict. However the majority of mutations are blatant frameshift errors or nonsense mutations, which cause premature termination of translation, representing a null allele. In one case a mutation of the stop codon was described that was predicted to cause protein elongation with consequent inactivation of E6-AP. No mutations were identified in the 5' untranslated exons screened in this study, or in exons 4 and 5 screened by Malzac *et al.*, 1998.

The mounting number of loss-of-function mutations identified in *UBE3A* mutations observed in NDUI AS patients provides strong evidence that disruption of *UBE3A*, with consequent loss of the maternally expressed product, causes AS. However, a substantial number of NDUI AS patients remain for whom no *UBE3A* mutations have been identified. These cases are discussed in section 6.3.2.

6.1.3.2 Further reduction of the AS candidate region includes the UBE3A gene

Through the molecular analysis of two key families, the AS candidate region has recently been reduced at the telomeric end. The maternally inherited 15q11-13 deletion in patient D.H., who presented with mental retardation but not AS, was further analysed and narrowed. D.H. and his mother B.S. were shown to be deleted at D15S986, a new microsatellite marker proximal to D15S113, but were intact for D15S210 (Trent *et al.*, 1997). FISH analysis using cosmids from the AS region placed the deletion breakpoint, and hence the telomeric boundary of the ASCR, just distal to PAC 14I12 (Sutcliffe *et al.*, 1997) (Fig 6.1).

An AS family with a reciprocal 14;15 translocation was described, in which the proband had inherited an unbalanced 'cryptic' translocation, with monosomy for 15pter-q11.2 and trisomy for 14pter-q11.2. The proband was found to have a maternally derived deletion of D15S11 and *SNRPN*, but was intact for *GABRB3* (Sutcliffe *et al.*, 1996). The translocation breakpoint was located distal to D15S10 within PAC 14I12, by FISH analysis (Sutcliffe *et al.*, 1997) (Fig 6.1). Methylation studies at *ZNF127*, D15S63 and *SNRPN* gave the typical AS pattern. The mother and unaffected sister carried an apparently balanced 14;15 translocation and had normal methylation status. Two possible mechanisms were presented to explain the occurrence of AS in the proband, whilst the sister was normal. Loss of just the imprinting centre (IC) would mean that the IC and AS structural gene would be separated by the translocation in the unaffected sister, with the AS gene lying on the derivative 14. This would suggest that the IC was capable of acting in *trans*, and that resetting of the maternal imprint occurs after metaphase I during oogenesis. Alternatively, loss of both the IC and AS gene would assume that both lay proximal to the translocation breakpoint, and both were present on the derivative 15 in the unaffected sister. The latter scenario was considered the most likely, which would place the AS gene centromeric to D15S10 (Burke *et al.*, 1996). This was consistent with the region excluded from the ASCR by the non-AS family deletion. The proximal extent of the ASCR remained at D15S174, as defined by the centromeric microdeletion breakpoint in AS Family S.

This narrowed ~250 kb AS critical region contained the *UBE3A* gene, making it the sole positional candidate for AS (Fig 6.1) (Sutcliffe *et al.*, 1997). This was consistent with the previous finding of truncating mutations in 4 NDUI AS patients with no previously identified molecular defect (Matsuura *et al.*, 1997; Kishino *et al.*, 1997).

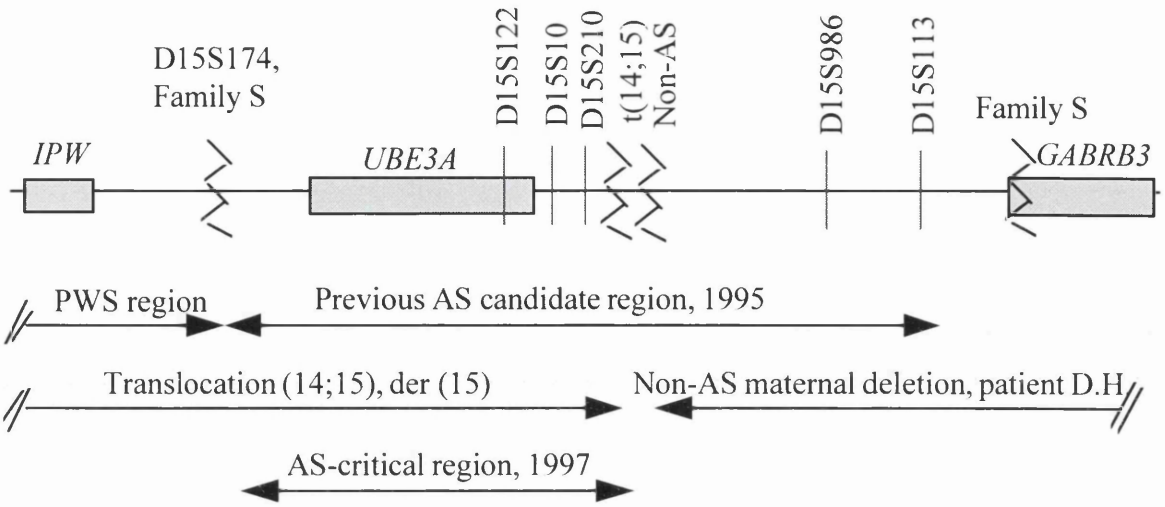


Fig 6.1 The minimal AS critical region contains the *UBE3A* gene.

Map showing the previous AS critical region, defined by D15S174 and D15S113, and the reduced AS critical region, defined by D15S174 and the breakpoints of the t(14;15) AS family and non-AS maternal deletion in patient D.H. The AS critical region spans approximately 250kb and contains the *UBE3A* gene.

6.2 Is *UBE3A* a suitable candidate for AS?

All classes of AS patients dictate that the gene responsible for AS must be expressed exclusively or preferentially from the maternal allele, or an essential transcript must serve its function in the maternal germline. Yet at the time maternally inherited inactivating mutations were identified in AS patients, there was no direct evidence for imprinted expression of *UBE3A*. Biallelic expression had been observed in both humans and mouse (Nakao *et al.*, 1994; Sutcliffe *et al.*, 1996; 1997). Furthermore, the precise function of E6-AP *in vivo*, and the means by which the absence of this protein would lead to AS, remained elusive.

6.2.1. Tissue-specific imprinted expression of *UBE3A*

In recent *in situ* hybridisation studies in PatDup:prox7 mice, the possibility that *Ube3a* is imprinted in a more subtle tissue-specific manner, was investigated, with the emphasis placed on neuronal tissues. Firstly, *Ube3a* expression was studied in wild-type mice at various stages of embryonic development to determine the normal pattern of

expression. Expression was first detected at embryonic day 8.5 in the neuroepithelium and branchial arches. By day 10.5 expression in these areas was substantial and was also visualised in the fronto-nasal mass. By day 15.5, expression became more localised to specific regions of the brain and developing facial tissues, with high levels of mRNA detected in hippocampus, thalamus, hypothalamus, cerebral cortex, basal ganglia, pituitary, olfactory bulb and tongue. The *Ube3a* expression profile in 6-9 month post-natum PatDup.prox7 mice was then compared with that of normal littermates. Several regions of the brain showed distinct differences in the levels of *Ube3a* expression. No expression was detected in regions of the hippocampus, Purkinje neurons, and the mitral cell layer of the olfactory bulb in PatDup:prox7 mice, where expression was abundant in normal mice. A generalised reduction of *Ube3a* transcript was noted in the cerebella and other cell-types. These results indicated that in certain specific regions of the brain *Ube3a* was expressed predominantly, if not exclusively, from the maternal allele. In other cell-types, maternal transcription was more prevalent than paternal transcription, whilst in other areas, expression was equivalent from both alleles (Albrecht *et al.*, 1997).

Recent RT-PCR based studies in humans have also provided evidence for tissue-specific imprinting of *UBE3A*, with a predominance of maternally derived expression seen in the brain (Vu and Hoffman, 1998; Rougelle *et al.*, 1998). In one study, an expressed polymorphism was used to determine whether *UBE3A* was imprinted in various tissues derived from heterozygous fetuses. This consisted of an A to C substitution in exon U4 of the 5' untranslated region (UTR). A single-nucleotide allele-specific RT-PCR assay of the A and C alleles was performed to quantitate the levels of expression of the two parental alleles. In heart, lung, limb, kidney, adrenal, intestine and placenta, an equal molar ratio of the A and C transcripts was observed, indicating that *UBE3A* is biallelically expressed in these tissues. Although both alleles were detected in fetal brain, there was a predominance of expression from a single allele. The parental origin of the alleles could not be traced however, as the parental genotypes could not be determined. The predominant allele was predicted to be the maternal allele (Vu and Hoffman, 1997). In another study, quantitative RT-PCR of the *UBE3A* subtypes was performed in normal and deletion-positive PWS and AS frontal cortex samples. The levels of expression of all subtypes were makedly reduced in AS brain, as compared with the normal and PWS samples. Using the Phosphorimager system, the *UBE3A* amplification products were normalised using a control cDNA in all brain samples. The level

of expression observed in AS brain was found to be 10% of the control signal (Rougelle *et al.*, 1997). This study indicates that *UBE3A* is expressed predominantly from the maternal allele in the human brain, corroborating the work done in mice with *Ube3a.*

The demonstration of paternal repression of *UBE3A* in brain provides the molecular basis by which loss-of-function mutations of *UBE3A* causes AS. The absence of *UBE3A* expression in specific regions and cell-types within the brain must underlie the pathogenesis of AS. These data also indicate which cell-types are involved in the pathophysiology of AS.

One other example of a gene demonstrating an anomalous imprinting pattern is the *KVLQT1* gene on chromosome 11p15.5. *KVLQT1* is globally imprinted during fetal development with expression from the maternal allele, but demonstrates biallelic expression in fetal heart. *KVLQT1* has been implicated in the BWS imprinted disorder, with the finding that in some patients with translocations involving 11p15.5, the breakpoints lay within *KVLQT1* (Lee *et al.*, 1997). However, *KVLQT1* is also associated with non-imprinted disorders involving cardiac defects with variable penetrance, which may be attributed to its imprinting profile. Homozygous missense mutations have been identified in the recessive disorder long-QT syndrome (LQTS), characterised primarily by prolongation of the QT intervals in electrocardiograms (Priori *et al.*, 1998; Saarinen *et al.*, 1998). Heterozygous mutations of *KVLQT1* have been identified in Romano-Ward syndrome, the autosomal dominant form of LQTS (Wang *et al.*, 1996). In addition, homozygous loss-of-function mutations of this gene have been detected in patients with Jervell and Lange-Nielsen syndrome, with congenital bilateral deafness combined with LQTS (Neyroud *et al.*, 1997).

The molecular mechanism underlying the tissue-specific imprinted expression of *UBE3A* is unknown and likely to be complex. This may involve differential regulation of a specific promoter or isoform, or a mechansim that directly represses expression from the paternal allele.

6.2.1.1 Is imprinting of UBE3A specific to a particular promoter, subtype or isoform?

The specific function of the various *UBE3A* subtypes and resultant protein isoforms is not known. However, RT-PCR analysis across the 5' UTR in 2 year and adult brain samples, and lymphocytes, revealed different expression profiles. Subtypes 6 and 7 were present in brain, but were not detected in lymphoblasts, and there was a reduction of subtype 5 in brain. It was postulated that one specific subtype or isoform of *UBE3A* might be imprinted in the

brain, and that AS may result through loss of this essential maternally expressed transcript. To investigate this, RT-PCR was performed across the 5' UTR, using primers that concurrently amplified all major *UBE3A* subtypes, in normal, PWS and AS brain. A generalised reduction of all subtypes expressed in brain to a level of 10% was observed in AS brain, indicating preferential expression of all subtypes from the maternal allele. Therefore, of the 7 subtypes observed in brain, none was expressed exclusively from the maternal allele (Rougelle *et al.*, 1997). However, these results do not preclude the possibility that specific subtypes, perhaps subtypes 6 and 7, serve an essential function in the brain, the absence of which may contribute to the AS phenotype. Furthermore, with the observation of a multitude of subtypes, it is still possible that an as yet unspecified splice-form specific to the brain, is expressed exclusively from the maternal allele. Although these analyses were performed with primers designed to amplify all previously reported *UBE3A* subtypes, others may exist which were not amplifiable with the primer set used. It is still possible that a single subtype of *UBE3A* is primarily involved in the pathogenesis of AS. However, the mutations identified in the coding region affect all three of the predicted protein isoforms, making such possibilities difficult to delineate.

In *Igf2* the imprinted transcripts are only derived from promoter P1, whereas promoters P2, P3 and P4 are active on both alleles. To test whether imprinted expression of *UBE3A* is promoter-specific, allelic expression from the two putative promoters P1 (within OP2) and P2, were assayed. Expression from both promoters was found to occur preferentially from just one allele, indicating that imprinted expression of *UBE3A* is not derived from just one promoter (Vu and Hoffman, 1997).

6.2.1.2 Mechanisms directly repressing the paternal allele

No allele-specific methylation has been reported at the *UBE3A* locus using lymphocytes of PWS and AS patients. The parental alleles at D15S10 showed similar methylation patterns (Driscoll *et al.*, 1992). Although 1354-G13 was used as a probe to test for differential methylation, this was situated at the 3' end of the gene, where methylation differences between the parental alleles are unlikely to be found (section 3.3.1.3). However, tissue-specific differential methylated in the regions where *UBE3A* is imprinted may occur, with a classical pattern of methylation on the inactive paternal allele and hypomethylation of the active maternal allele (Fig 6.2). There is some precedence for this. Tissue-specific

methylation has been observed in non-imprinted genes, which are expressed in a highly tissue-restricted manner. For instance the promoter region of the human beta-amyloid precursor protein gene, which shows varying levels of expression in different regions of the brain, was found to be methylation in some regions of the brain, and unmethylated in others. The methylation patterns correlated with the levels of expression in the different regions, with the promoter unmethylated where expression levels were high (Rogaev *et al.*, 1994). Tissue-specific differential methylation has, more pertinently, been observed for imprinted genes. Expression of the mouse *Igf2r* gene exclusively from the unmethylated maternal allele occurs in most tissues, including liver, kidney, intestine, skin, and skeletal muscle. The repressed paternal allele is methylated in the 5' region of the gene in these tissues. However, in the central nervous system, where *Igf2r* is biallelically expressed, both alleles are unmethylated (Hu *et al.*, 1998). In the imprinted region of chromosome 11p15.5, monoallelic expression of *H19* from the maternal allele occurs when the promoter is unmethylated. This causes inactivation of *IGF2* on the maternal allele. Consequently *IGF2* is expressed solely from the paternal allele. The *H19* promoter is unmethylated on the maternal allele in all tissues in which *H19* and *IGF2* are oppositely imprinted. However, in the choroid plexus and leptomeninges, where *H19* is inactive and *IGF2* is biallelically expressed, the *H19* promoter is methylated on both alleles. Thus the methylation status of the *H19* gene is different in the specific regions of the brain where *IGF2* is exempt from imprinting, than in the tissues where these genes are normally imprinted. This distinct methylation pattern is observed in BWS patients with imprintor mutations and Wilm's tumours of the kidney. Here the methylation pattern is aberrant and consequent loss of imprinting has phenotypic consequences (Moulton *et al.*, 1994; Steenman *et al.*, 1994). It is plausible that *UBE3A* is specifically methylated on one allele in the hippocampus, olfactory bulb, and other regions of the brain, reflecting the tissue-specific imprinted mode of expression that exists in these regions. Aberrant methylation of *UBE3A* in these regions of the brain could represent another molecular mechanism for AS, causing loss of maternal expression.

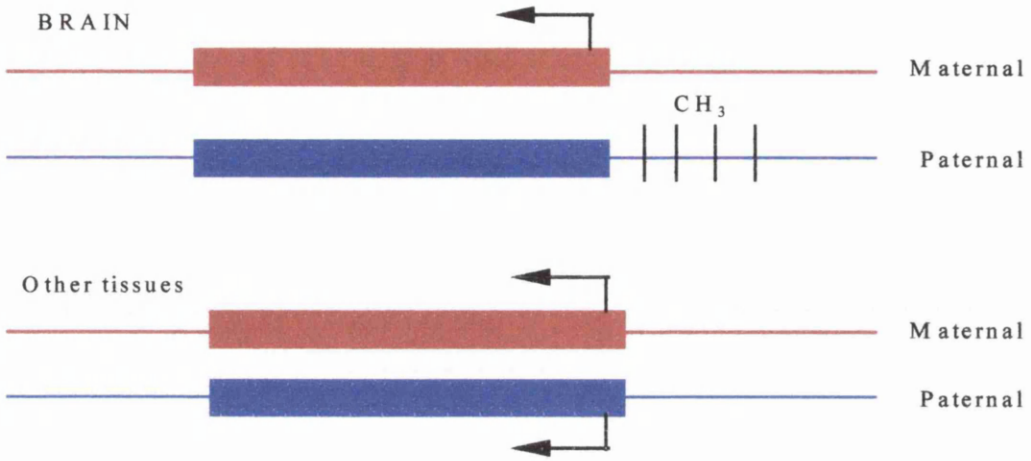


Fig 6.2 Brain-specific differential methylation may regulate paternal imprinting of *UBE3A* in the brain.
UBE3A is represented by boxes. Arrows denote gene expression. The maternal allele is in red, the paternal allele in blue. Vertical lines indicate methylation of the paternal allele in specific regions of the brain where *UBE3A* is imprinted.

Regulation of imprinted expression through the transcription or action of an antisense RNA has been proposed as a mechanism for allele-specific gene silencing (reviewed by Tilghman, 1999). Antisense RNAs have been detected for other imprinted genes including *ZNF127* (*ZNF127AS*) on 15q11.2 and its murine homologue (Jong *et al.*, 1999a,b), *Igf2r* (Wutz *et al.*, 1997), and *Xist* (Lee *et al.*, 1999). In the case of *Igf2r*, which is maternally transcribed, the antisense transcript is expressed exclusively from the paternal allele (Wutz *et al.*, 1997). Prior to X chromosome inactivation in female cells, *Xist* is transcribed from both X chromsomes. The RNA transcript from the future inactive X then becomes stabilised and mediates X inactivation in *cis*. The *Xist* RNA remains closely associated with the inactivated X (reviewed by Heard *et al.*, 1997). The antisense transcript *Tsix* is similarly expressed from both X chromosomes in undifferentiated ES cells, but at the onset of X inactivation, *Tsix* expression occurs monoallelically from the X chromosome that is destined to remain active. *Tsix* expression continues until *Xist* is switched off (Lee *et al.*, 1999). Transcription of the *UBE3A* antisense RNA may indirectly affect expression of the *UBE3A* gene through competition for regulatory elements. In this hypothesis, *UBE3A* would only be expressed from the allele on which the antisense transcript is silent (Fig 6.3). This seems plausible since expression of the antisense RNA is confined to the brain, which correlates with imprinting of

UBE3A specifically in this tissue. This mechanism would also account for paternal imprinting of *UBE3A* in the absence of a methylation imprint in the promoter region. Instead the promoter for the antisense transcript downstream of *UBE3A* may be differentially methylated, with expression occuring from the unmethylated paternal allele. Thus the methylation status at the start site of the antisense RNA may indirectly affect *UBE3A* expression. It is possible that loss of imprinting of the antisense transcript, perhaps through epigenetic alteration, could give rise to AS through inhibition of *UBE3A* transcription.

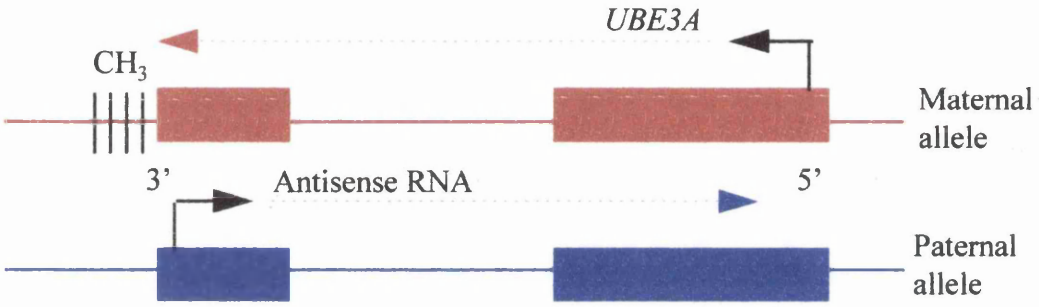


Fig 6.3 Tissue-specific imprinting through antisense RNA transcription
The boxes represent the 5' and 3' ends of *UBE3A*. The maternal allele is in red, the paternal allele in blue. Brain-specific antisense transcription occurs 3' to 5' from the paternal allele. *UBE3A* transcription occurs on the maternal allele in the brain.

6.2.1.3 Possible effects of imprintor mutations on UBE3A expression in AS

Maternal deletions and mutations of a region 5' to *SNRPN*, which cause disruption of expression of the BD or IC exons, result in aberrant methylation of the maternal allele around the *SNRPN* promoter and AS. The maternal chromosome in AS imprintor mutation cases resembles the wild-type paternal allele both epigenetically and functionally. Conversely, deletion of the *SNRPN* promoter on the paternal allele results in PWS, with the paternal allele showing the same methylated imprint as the normal maternal allele. Thus a bipartite imprinting control element was proposed which resets the methylation imprint during gametogenesis to correspond to the sex of the parent. This consequently regulates imprinted gene expression within 15q11-13. However, in the models proposed to explain how the IC effects imprint resetting, the in *cis* regulation of *UBE3A* expression is not addressed. Since IC mutations and loss-of-function mutations in *UBE3A* have the same phenotypic outcome,

UBE3A expression must be inhibited either indirectly through aberrant methylation, or through lack of an IC transcript. In both models for the regulation of tissue-specific imprinted expression of *UBE3A*, the possibility of epigenetic mutations in the *UBE3A* region is invoked. A paternal imprint on the maternal allele at the putative site of tissue-specific methylation of *UBE3A* may cause silencing of the maternal allele. Alternatively, activation of the *UBE3A* antisense transcript on the maternal allele, as occurs for the paternally expressed genes in the PWS region (Saitoh *et al.*, 1996), may lead to maternal repression of *UBE3A* in the brain. This may occur through lack of methylation on the maternal allele in the AS region, at the site of the antisense promoter.

6.2.2 Pathogenesis of AS due to lack of E6-AP

The pathogenesis of AS remains to be defined, but recent work indicates that the phenotypic effects are due to deficiency of functional E6-AP in particular cell types where the paternal *UBE3A* allele is inactive.

In vitro studies showed that E6-AP is involved in the degradation of p53 in the presence of the E6 protein of the HPV types 16 and 18 protein (Huibregtse *et al.*, 1991). The ability of E6-AP to target p53 for ubiquitination *in vivo* has now been demonstrated, but only in an E6-dependent manner in HPV infected cells (Beer-Romero *et al.*, 1997; Talis *et al.*, 1998). E6-AP was also found to target HH23RA, which has sequence homology to the yeast RAD23 protein, in an *in vitro* ubiquitin assay (Kumar *et al.*, 1997). E6-AP was recently shown to serve a second function as a coactivator of hormone-dependent transcriptional activity of the nuclear hormone receptors, including the progesterone, A, E, glucocorticoid, and thyroid hormone receptors. This function was found to be independent of the ubiquitin-protein ligase activity of E6-AP (Nawaz *et al.*, 1999). Functional studies performed on AS disease mutant forms of the E6-AP protein have shown that it is the ubiquitination activity that is impaired in AS. Mutant proteins were incapable of p53 ubiquitination in the presence of E6 *in vitro* (Talis *et al.*, 1998). *In vivo* studies using transfected HeLa cells showed that mutant proteins, including AS truncated and elongated forms, were unable to ubiquitinate HHR23A. However, one AS missense mutant was able to ubiquitinate the HHR23A at a similar level to the wild-type E6-AP. The ability of the same mutant E6-AP proteins to coactivate the progesterone receptor in the presence of progesterone was not affected. This

study indicates that the pathogenesis of AS is likely to be attributable to a lack of ubiquitination activity in the cell types where E6-AP is absent, rather than through loss of coactivation function (Nawaz *et al.*, 1999).

Mice with a null mutation for *Ube3a* have been created by deleting exon 2 and causing a frameshift at the N terminus. Mice with a paternal deficiency of *Ube3a* (m+/p-) show no phenotypic features. However, mice with a maternal deficiency for *Ube3a* (m-/p+), have several features in common with AS, demonstrating motor disfunction, inducible seizures, context-dependent learning deficit, and impaired hippocampal long-term potentiation. As previously found in mice with paternal disomy for proximal chromosome 7, there was an absence of *Ube3a* expression in cerebellar Purkinje cells, hippocampal neurons and mitral cells of the olfactory bulb, compared to the wild-type littermates and m+/p- mice. These data indicate that mice with the maternal null-mutation for *Ube3a* represent an AS mouse model for *UBE3A* mutations seen in AS patients. These mice do not show the same extensive phenotypic features observed in the pUPDprox7 mice, but this may be due to the effects of other genes within the ~30cM region. Immunohistochemical analysis of p53 in brain sections showed a distinct increase in the levels of cytoplasmic p53 in the Purkinje cells and hippocampus in the m-/p+ mice. Analysis of sections of AS brain showed a marked loss of Purkinje cells and internal granule neurons in the cerebellum, with increased levels of p53 in the few that remained. These data indicate that E6-AP regulates the levels of p53 in post-mitotic neurons in the absence of viral E6 in both mouse and humans (Jiang *et al.*, 1998).

The mystery is whether p53 is the effector protein targeted by E6-AP that is primarily involved in the pathogenesis of AS. E6-AP may target several substrates for ubiquitination, and p53 may just be a marker for lack of ubiquitination activity in post-mitotic neurons in AS. Excessive levels of p53 in the nucleus lead to cell cycle arrest and may lead ultimately to apoptosis. Death of cells lacking E6-AP expression during fetal brain development, may result in a reduction of post-mitotic cells in crucial regions of the brain. However, p53 is also ubiquitinated by other E3 ligases, including by MDM2 in the nucleus (Haupt *et al.*, 1997), and so is not regulated by E6-AP alone. There is also the possibility that ubiquitination by E6-AP does not lead solely to proteolysis and degradation of the substrate. Ubiquitination has also been shown to cause functional activation of the NF-κB transcription factor through IκBα protein kinase activity (Chen *et al.*, 1996), and is involved in endocytosis of cell

surface receptors (reviewed by Hochstrasser, 1996). Although a lack of ubiquitination activity has been observed *in vivo* for AS mutant proteins, the effector protein that is involved in AS may not be subject to degradation by E6-AP.

The involvement of the ubiquitin proteolytic pathway in the pathogenesis of several human genetic diseases has recently come to light. For instance, despite normal ion channel function, the cystic fibrosis mutant delta F508 CFTR protein does not reach the cell surface, but is retained in the endoplasmic reticulum and is degraded by the ubiquitin proteasome pathway (Ward *et al.*, 1995). In neurodegenerative disorders, including Parkinson's disease, Alzheimer's disease and Huntington's disease, there is incomplete proteolysis of the mutant proteins. This results in accumulation of ubiquitinated abnormal proteins in endosomal or lysosomal inclusions in the neural cells. This may be due to malfunction of the ubiquitin/ATP dependent pathway, or structural changes in the protein substrates that prevent complete proteolysis (DiFiglia *et al.*, 1997; Alves-Rodrigues *et al.*, 1998). The action of ubiquitin mediated proteolysis has also been implicated in malignancies, immune and inflammatory responses and muscle-wasting (reviewed by Schwartz and Ciechanover, 1999).

6.3 Clinical applications and implications

6.3.1 Difficulties in identification of mutations in *UBE3A*

The presence of two processed pseudogenes of *UBE3A* was a confounding factor in the identification of disease-causing mutations in NDUI AS patients. The identification of numerous nucleotide substitutions within short exonic sequences of *UBE3A* during the mutation screening was indicative of the existence of *UBE3A* pseudogene sequences. Furthermore, in two families the affected siblings and mother had 3 *UBE3A* exon 9 'alleles'. In Family H, SSCP analysis identified a shift in the mother and affected children, consistent with a maternally transmitted mutation. Sequencing of subclones from exon 9 amplified with primers 3F5 and 3R5, identified 3 'alleles' in the mother and affected siblings. These consisted of the normal *UBE3A* sequence, the mutant *UBE3A* sequence, and a sequence containing several nucleotide substitutions. The latter was indicative of the presence of a pseudogene sequence, which had been co-amplified with the *UBE3A* exon 9 sequence. This

was subsequently shown to be part of the *UBE3AP2* pseudogene, which was highly homologous to *UBE3A* exon 9. Amplification of these three different species most likely caused the complicated SSCP patterns observed in this family, consisting of 5 separate bands. In Family W, SSCP analysis of exon 9 with primers 3F3 and 3R3 showed a shift in the DNA of the mother and three affected siblings, indicative of a maternally inherited mutation in this fragment. The PCR products were cloned and SSCP analysis performed for 6 subclones from each family member. Three different SSCP bands were detected in the DNA of the mother and affected siblings, but just two in the father. Initially it was assumed that the DNA samples or PCR products had been contaminated, or one of the bands represented a clone with no insert. Sequence analysis of the mother and affected children revealed ten nucleotide substitutions on one 'allele'. Analysis of this family was repeated using primers published by Malzac *et al.*, and a 4bp insertion causing a pathological frameshift was subsequently identified in two of the affected children analysed. Complex mutations involving more than one nucleotide are a rare occurrence. Mutations involving two or three nucleotide alterations have been reported. For instance, substitution of two consecutive nucleotides have been identified in the FGFR2 gene, causing Apert syndrome (Oldridge *et al.*, 1997). A mutation involving three nucleotide substitutions in the *IsK* gene has been reported in a family with Jervell and Lange-Nielson syndrome. Two substitutions were identified in two adjacent nucleotides, and a third substitution was identified 5bp upstream (Tyson *et al.*, 1997). However, such cases are extremely rare, and there is no precedent for mutations involving up to 10 nucleotide substitutions within one exon, as observed in Family W. Furthermore, I.M.A.G.E. clone 1354-G13, representing the extended exon 16 of *UBE3A,* identified genomic fragments in genomic DNA digested with *EcoR*I and *Hind*III, which mapped outside the ASCR. This also indicated the presence of related low copy-number sequences within the genome, that was possibly a pseudogene sequence. The existence of two pseudogenes, *UBE3AP1* and *UBE3AP2*, and their sequences were subsequently reported. The nucleotide substitutions identified in AS families H and W, and patient 7020 were later shown to represent the *UBE3AP2* pseudogene sequence (section 4.2.3), illustrating the need for primers designed to intronic sequences where possible, for mutation screening of *UBE3A* from genomic DNA in NDUI AS patients. However, certain exons such as exon 9, are large. Where mutation screening is performed by SSCP, these exons need to be divided into <300bp fragments, making the use of exonic primers unavoidable. In this situation, primers

need to be designed to sequences where the *UBE3A* exon and pseudogene sequences differ, in order to achieve specific amplification of the exon sequence. Alternatively, mutation screening could be performed using cDNA from patient cell lines, since the pseudogene sequences are not expressed. In this report, the pseudogene sequences were identified and distinguished through direct sequence comparison of the 'alleles' displaying sequence changes with the published *UBE3AP2* pseudogene sequence.

Most of the mutations observed in AS patients have led to frameshift errors, but some missense mutations have been identified for which the effect on the protein is difficult to predict. Mutations in the phosphomannomutase 2 gene *PMM2* cause carbohydrate-deficient glycoprotein syndrome type 1 (CDG1). This has a pseudogene, PMM2ψ, with high sequence homology, and primers which avoid amplification of the pseudogene are used for mutation detection in CDG1 patients. However, it has been found that some of the mutations in *PMM2,* which have occurred through oxidative deamination of a C to a T, are the same as the sequence changes observed in the pseudogene. Presumably these represent mutation hotspots which have occurred during the divergence process of the functionally redundant pseudogene, which are also liable to mutate in the *PMM2* gene itself (Schollen *et al.*, 1998). Where missense mutations occur in *UBE3A* that are also represented in the pseudogene sequence, it is important that the mutant *UBE3A* allele can be distinguished from the pseudogene.

6.3.2 Mutations of *UBE3A* do not account for all cases of AS with negative deletion, UPD and methylation tests

No mutations of *UBE3A* were identified for 2 of the 5 families and 17 sporadic NDUI cases of AS studied at the Institute of Child Health. However, the proportion of cases for which mutations were identified in this study, 60% in familial cases and 19% in sporadic patients, is comparable with those reported by other groups. This is similarly the case if sporadic patient O is re-categorised as a familial case, with the demonstration that the pathogenic mutation was maternally inherited in this patient. This would increase the number of mutations identified in familial cases to 4/6 (66.7%), and reduce the number in sporadic patients to 3/20 (15%). Malzac *et al.*, identified mutations in 8 of 10 AS families (80%) and

just 5 of 35 (14%) isolated cases. In this study mutation screening was performed for coding exons 7 to 16, plus 5' untranslated exons 4 and 5, by SSCP analysis followed by sequencing (Malzac *et al.*, 1998). Fang *et al.*, (1999) found mutations in 6 out of 8 (75%) families and 11 of 47 (23%) sporadic patients by direct sequencing of exons 7 to 16. Therefore a substantial number of NDUI AS patients remain for whom no pathogenic mutation of the *UBE3A* gene has been identified, and the number of mutations detected in familial cases is considerably higher than in isolated cases.

It is likely that a proportion of the remaining NDUI AS cases have unidentified mutations of *UBE3A* on the maternal allele. It is possible that the use of SSCP for mutation screening failed to detect mutations in some patients, as this method is reported to be 70-95% efficient, depending on the size of the DNA product and the nature of the sequence change (section 1.9.4.2.1). Direct sequencing of all the coding exons of *UBE3A* may have improved the mutation detection rate in the study by Fang *et al.* although they failed to identify mutations in all the patients studied. There may be mutations present in regions of the gene that have not yet been analysed. For instance, the recombination in AS Family WJK places the pathogenic mutation telomeric to D15S122, in exon 1 or the regulatory region of *UBE3A* (Fig 1.9)(Kishino *et al.*, 1997). Where PCR amplification of exons was performed using exonic primers, any splice mutations at intron/exon boundaries will not have been detected. Disease-causing mutations have occasionally been detected in the introns of genes (Dlott *et al.*, 1990; Munroe *et al.*, 1997; Silbergeld *et al.*, 1997; Fujimaru *et al.*, 1998; Li *et al.*, 1998). AS yet, there have been no reports of the investigation of *UBE3A* introns for mutations.

These data suggest that primary sequence changes of *UBE3A* may not account for all patients with a clinical diagnosis of AS, but a negative methylation test. This raises many questions. The AS phenotype has some overlap in clinical phenotype with other disorders, including ataxia, lack of speech and mental retardation. Therefore it is possible that some of these patients have been misdiagnosed. However, stringent criteria have been published for a positive diagnosis of AS (Table 1.1)(Williams *et al.*, 1995), and in recent publications the authors have stipulated that the patients were examined by experienced clinicians (Malzac *et al.*, 1998; Fang *et al.*, 1999). The diagnoses of patients referred to the Institute of Child Health with probable AS were either confirmed or excluded by two paediatricians who are specialists of this particular syndrome. Therefore misdiagnoses are highly unlikely to account for the number of NDUI AS patients without mutations of *UBE3A*. A sporadic condition may

exist with similar clinical findings, which has a distinct genetic aetiology. Locus heterogeneity for AS may exist. However, the AS structural gene region has now been narrowed to approximately 250 kb, of which *UBE3A* spans 120 kb (Sutcliffe *et al.*, 1997; Kishino and Wagstaff, 1998). No additional genes have been identified in this region, despite intensive positional cloning efforts by several groups. Mutations of *UBE3A* have been identified in a greater fraction of familial AS cases, which demonstrate linkage to D15S10, than sporadic cases. Another locus not linked to 15q11-13 may be involved in cases of sporadic AS, which is associated with a low recurrence risk. This may be a gene that regulates *UBE3A* expression, belongs to the same biochemical pathway, or a gene whose protein product interacts with E6-AP in crucial tissues, giving rise to a similar phenotype. Alternatively a distinct, as yet unidentified molecular mechanism may exist, such a localised epigenetic change that leads to maternal silencing of *UBE3A*.

6.3.3 Implications for Genetic Counseling

Genetic counseling in AS is complex and the risk of recurrence of the disease in siblings is dependent on the aetiological mechanism. Recurrence risk for AS or PWS due to cytogenetic deletion or UPD is below 2% in the absence of a predisposing chromosomal rearrangement, but is 50% in cases with imprintor mutations due to deletion or mutation in the imprinting center. In all cases that were negative for the methylation test and demonstrated biallelic inheritance for 15q11-13, a recurrence risk of 50% was given irrespective of whether there was a single or more than one affected child in the family. Sporadic cases due to *de novo* alterations with a low recurrence risk could not be distinguished from familial cases where a recurrence of the disorder could occur. Assessment of carrier status or prenatal diagnosis was previously based on haplotype analysis (reviewed by Stalker and Williams, 1998).

The identification of mutations of *UBE3A* is invaluable for accurate risk assessment and genetic counseling in individual families giving normal methylation results. Knowledge of the pathogenic mutation will enable definitive prenatal diagnosis of repeat pregnancies in the mothers of affected children and determination of carrier status in AS family members. In one family with two affected children, prenatal diagnosis was requested for a subsequent

pregnancy. A G→A substitution causing a nonsense mutation at codon 768 was identified in the two affected children and the mother. Mutation analysis on DNA from cultured amniocytes indicated that the fetus had not inherited the mutation from the mother, and the pregnancy was continued to term. A normal infant was born. Mutation analysis of maternal relatives indicated that none were carriers for the disease (Tsai *et al.*, 1998). In another familial case, the maternal aunt of two AS siblings was pregnant and referred for genetic testing. The aunt was previously shown to have inherited the same grandpaternal 15q11-13 haplotype as the affected children and their mother, and so was given a high risk of carrier status. A single base deletion was subsequently identified in the AS siblings and their mother, but was not present in the maternal grandparents or the pregnant aunt. Thus the mutation had arisen *de novo* on the grandpaternal allele in the mother, or the grandfather was germ-line mosaic for the mutation. Identification of the mutation in this family enabled the carrier status in the aunt to be accurately reassessed (Fang *et al.*, 1999).

In one isolated case of AS (patient O) reported here, the mutation was also identified in the mother. This confirms a recurrence risk of 50% in this family. Detection of carrier status can now be performed accurately in matrilineal relatives in this family and in the 3 other familial cases for which the pathogenic mutation has been identified.

In 4 isolated AS cases studied here the mutations represented *de novo* events as these were not present in the parents, indicating a low recurrence risk. However, these results must be interpreted with caution as maternal somatic and gonadal mosaicism has been described in 3 AS families with *UBE3A* mutations (Malzac *et al.*, 1998). In one family, two affected children shared the same maternal haplotype as an unaffected sibling. An estimated 5% somatic mosaicism was detected in the mother. In another family, the mothers of AS cousins were heterozygous for a mutation, for which the grandfather was found to be mosaic. Somatic mosaicism was also detected in the mother of an isolated AS case, indicating a sibling recurrence risk of up to 50% depending on the level of gonadal mosaicism. A low recurrence risk is indicated in families where a *de novo* mutation has occurred and there is no evidence for somatic mosaicism. However, this does not rule out the possibility of mosaicism confined to the ovaries. Prenatal diagnosis may be performed where there is concern for gonadal mosaicism in mothers who apparently do not carry the mutation. In cases where the molecular aetiology for AS is still unknown, the recurrence remains at 50%.

6.4 Evidence that PWS is a multigene disorder

PWS is a neurogenetic disorder caused by disruption of the paternal 15q11-13 allele through paternal deletion, mUPD15 or imprintor mutations. These mechanisms account for almost all PWS patients indicating that this is a contiguous gene disorder, due to the loss of expression of more than one, perhaps several, paternally expressed imprinted genes from 15q11-12. Futhermore, no mutations have been described in a single gene within 15q11-13 in PWS or PWS-like patients. Absence of expression of paternally expressed genes *ZNF127*, *NDN*, *SNRPN*, *IPW* and transcripts PARs 1 and 5 from 15q11-13 has been demonstrated in each of these categories of PWS patients. Imprinted expression of genes within 15q11-13 is under the control of a bipartite imprinting center, which in the case of the paternally expressed genes, involves the promoter and first exon of *SNRPN*. The total number of genes involved in the PWS phenotype is also not known (reviewed by Nicholls *et al.*, 1998).

SNRPN was the first paternally expressed gene to be mapped to the PWSCR and became the first major candidate for PWS. There are several lines of evidence that suggest *SNRPN* play a role in the PWS phenotype. PWS may be due to loss of a neuronal factor. *SNRPN* is expressed most abundantly in neural tissues, but is absent in PWS patients with deletions, disomy and imprintor mutations. Paternal deletions of the *SNRPN* promoter and first exon result in PWS due to imprintor mutations (Sutcliffe *et al.*, 1994; Buiting *et al.*, 1995; Saitoh *et al.*, 1996). One typical PWS case has been described in whom a *de novo* balanced (15;19) translocation disrupts the *SNRPN* gene, implicating *SNRPN* alone in the PWS phenotype. The translocation breakpoint was identified between *SNRPN* exons 2 and 3, with the 5' end of the gene on the derivative paternal 15, and the coding exons on the derivative chromosome 19. The methylation pattern was normal at *SNRPN* and D15S63, and expression of the paternally expressed genes *ZNF127*, *IPW* and the PAR transcripts was detected in this patient, suggesting these genes do not play a significant role in PWS. However, the level of expression of these transcripts was not determined, but may have been decreased. Although transcription of *SNRPN* exons 1-2 and 3-8 was detected by RT-PCR, the translocation prevented transcription of the full-length RNA (Sun *et al.*, 1996). The recently detected *Snurf* (*SNRPN* upstream reading frame) transcript encoded by *SNRPN* exons 1-3 (R. Nicholls, unpublished), would also have been disrupted.

There is also conflicting evidence to suggest that *SNRPN* may not contribute

significantly to PWS. A patient meeting the diagnostic criteria for PWS but with no history of hyperphagia has been described with a balanced (9;15) translocation, in whom *SNRPN* expression is unaffected. The translocation breakpoint was identified on the paternal allele between *SNRPN* and *IPW. ZNF127*, *SNRPN* and PAR5 were present on the derivative 15 and demonstrated normal expression. *IPW* and PAR1 were on the derivative chromosome 9, and no expression was detected for either of these transcripts. The methylation patterns at *SNRPN* and D15S63 were normal. This suggests that *SNRPN* and the paternally expresed genes proximal to *SNRPN* do not contribute to the PWS phenotype, and implicates *IPW*, PAR1 and any other transcripts distal to *SNRPN* in PWS (Schulze *et al.*, 1996). SmN is thought to function as an RNA splicing protein in neural tissues, where it is most abundant in both humans and mice, and where its functional homologue SmB is absent. However, in mice with maternal disomy for central chromosome 7 (PWS mouse model) normal splicing of RNAs, such as calcitonin, has been observed in neural tissues despite the absence of SmN (Huntriss *et al.*, 1994). Elevated levels of SmB were observed in neural tissues of these mice, suggesting that lack of SmN is compensated for by the action of SmB. However, it is possible that SmN has specific splice targets in the brain which SmB does not process (Huntriss, 1998). In mice created with a paternal deletion of coding exons 5-7 of *Snrpn,* with consequent absence of SmN, appeared phenotypically normal. The differential methylation pattern in the region and expression of *Zpf127*, *Ndn* and *Ipw* was unaffected in these mice. The mice were reproduced in the homozygous state indicating that the *Snrpn* deletion did not affect viability or fertility, although no cognitive tests were performed. In contrast mice with a paternal deletion of the exons 1-6 of *Snrpn*, including the IC region, died neonatally. These mice were hypotonic and died within a few days after birth due to poor feeding abilities. These mice failed to express *Zpf127*, *Ndn* and *Ipw Zpf127*. Lack of an IC in mice therefore causes a phenotype analogous to PWS in humans (Yang *et al.*, 1998). These analyses show that while the protein product of *Snrpn* is not essential for an apparently normal phenotype in mice, transcription of the 5' exons of the *Snrpn* gene is. The IC in mouse and man includes the first exon of *Snrpn*, which is necessary for the normal regulation of imprinted genes in the region.

It is possible that mutations in individual genes in the PWS region may cause phenotypic effects that do not meet the criteria for a positive diagnosis of PWS. Patients with mild phenotypic features that overlap the PWS phenotype may bear mutations in individual

genes in the PWS region. For instance, Sun *et al.* (1996) suggested a role for *SNRPN* in hyperphagia, since the patient in whom *SNRPN* activity was found to be normal was not hyperphagic. The role of individual genes may be established through the analysis of mice in which single genes are targeted for disruption. Not only is the role of individual genes in the PWS region unknown, but the number of genes involved in this syndrome is also unknown. It is possible that additional genes in the PWS region, which may also play a role in PWS, remain to be found.

6.4.1 Identification of a novel, imprinted sequence in the PWS region

In this report, the identification of two additional partial transcripts from the PWS region, represented by cDNA clones 395-H22 and 123-E19, was described. I.M.A.G.E. clones 395-H22 and 123-E19 were identified from the PWS region proximal to D15S174, through direct cDNA selection followed by database searching. These transcripts map between *SNRPN* and *IPW*. Preliminary analyses with a human x mouse hybrid imprinting assay system indicated that 395-H22 expression is repressed on the maternal allele, adhering to the paternal-only mode of expression of other genes and transcripts identified from the region. However, absence of expression of 395-H22 needs to be confirmed in PWS cell lines in order to ascertain whether it represents another candidate gene for PWS. Due to time constraints, little is known about 395-H22, and much further work needs to be performed to determine its significance with respect to PWS. 395-H22 identified a 1.4kb transcript predominantly in fetal heart, but RT-PCR analysis showed that it was widely transcribed. The 1.1kb 395-H22 sequence was non-coding, but analysis of the full-length 1.4kb sequence will determine more conclusively if this transcript codes for any amino acid sequence or functions as an RNA molecule. Several non-coding sequences have been identified in the PWS region, including *IPW*, and PARs 1, 5 and 7. It has been suggested that the PAR transcripts may represent 'read-through' transcripts of *SNRPN* or *IPW*. A novel 'read-through' sequence at the 3' end of *IPW* was identified by cDNA selection in this project, which was subsequently linked to the *IPW* Unigene cluster (section 3.2.10). *SNRPN* is a 1.5kb transcript and 395-H22 identified a transcript of similar size. However, no sequence overlap was observed between 395-H22 and *SNRPN*, making it unlikely that 395-H22 is such a 'read-through' transcript. In order to characterise the 395-H22 transcript further, the full-length sequence needs to be isolated, either by screening a cDNA library, or by 5' and 3' RACE. This cDNA clone has

not been linked to any Unigene cluster, and remains a novel sequence identified in the PWS region.

A second clone, 123-E19 was identified through direct cDNA selection and database searching. Although this clone mapped to the PWSCR, maternal expression was detected in monochromosomal hybrids containing a maternal chromosome 15, indicative of biallelic expression in fibroblast cells. Imprinting analysis needs to be repeated using PWS and AS cell lines before imprinting of this sequence can be ruled out in these cell types. 123-E19 has been poorly characterised so far. The clone failed to identify a signal on northern blots containing fetal brain, heart, liver, kidney, lung and skin, although expression was detected in several tissues by RT-PCR. This indicates that 123-E19 is expressed at low levels in the 15 week fetal tissues analysed by northern hybridisation. However, analysis of expression of sequences that are contiguous with genomic DNA by RT-PCR has to be performed with care to avoid amplification from a DNA template. No genomic contamination was observed in the RT-minus controls for 123-E19 or the *HPRT* gene, indicating amplification of 123-E19 was from a cDNA template. Northern analysis using up to 3µg polyA+ mRNA should be repeated for this clone to confirm expression and establish the size of the transcript 123-E19 is derived from. Following this, the full-length transcript needs to be isolated and sequenced, in order to characterise it further

References

Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF, Kerlavage AR, McCombie WR, Venter JC (1991). Complementary DNA sequencing: Expressed sequence tags and human genome project. Science 252: 1651-6

Adams MD, Dubnick M, Kerlavage AR, Moreno R, Kelley JM, Utterback TR, Nagle JW, Fields C, Venter JC (1992). Sequence identification of 2,375 human brain genes. Nature 355: 632-4

Albrecht U, Sutcliffe JS, Cattanach BM, Beechey CV, Armstrong D, Eichele G, Beaudet AL (1997). Imprinted expression of the murine Angelman syndrome gene, *Ube3a*, in hippocampal and Purkinje neurons. Nat. Genet. 17: 75-8

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic Local Alignment Search Tool. J. Mol. Biol. 215: 403-10

Alves-Rodrigues A, Gregori L, Figueiredo-Pereira ME, (1998). Ubiquitin, cellular inclusions and their role in neurodegeneration. TINS 21: 516-20

Angelman H (1965). 'Puppet' Children. A report on three cases. Dev. Med. Child Neurol 7: 681-688

Antequera F, Bird A (1993). Number of CpG islands and genes in human and mouse. Proc. Natl. Acad. Sci. USA 90: 11995-9

Antonarakis SE (1991). Parental origin of the extra chromosome in trisomy 21 as indicated by analysis of DNA polymorphisms. Down Syndrome Collaborative Group. N. Engl. J. Med. 324: 872-6

Barr JA, Jones J, Glenister PH, Cattanach BM (1995). Ubiquitous expression and imprinting of Snrpn in the mouse. Mamm Genome 6: 405-7

Bartolomei MS, Webber AL, Brunkow ME, Tilghman SM (1993). Epigenetic mechanisms underlying the imprinting of the mouse H19 gene. Genes Dev 7: 1663-73

Bartolomei MS (1994). The search for imprinted genes. Nat. Genet. 6: 220-1

Barton SC, Surani MAH, Norris ML (1984). Role of paternal and maternal genomes in mouse development. Nature 311: 374-6

Beer-Romero P, Glass S, Rolfe M (1997). Antisense targeting of E6AP elevates p53 in HPV-infected cells but not in normal cells. Oncogene 14: 595-602

Berry R, Stevens TJ, Walter AS, Wilcox AS, Rubano T, Hopkins JA, Weber J, Goold R, Bento Soares M, Sikela JM (1995). Gene-based sequence-tagged-sites (STSs) as the basis for a human gene map. Nat. Genet. 10: 415-23

Bettio D, Rizzi N, Giardino D, Grugni G, Briscioli V, Selicorni A, Carnevale A, Larizza L (1995). FISH analysis in Prader-Willi and Angelman syndrome patients. Am. J. Med. Genet. 56: 224-8

Beuten J, Mangelschots K, Buntinx I, Coucke P, Brouwer OF, Hennekam RC, vanBroechhoven C, Willems PJ (1993). Molecular study of chromosome 15 in 22 patients with Angelman syndrome. Hum. Genet. 90: 489-95

Beuten J, Hennekam RCM, Van Roy B, Mangelschots K, Sutcliffe JS, Halley DJJ, Hennekam FAM, Beaudet AL, Willems PJ (1996a). Angelman syndrome in an inbred family. Hum. Genet. 97: 294-8

Beuten J, Sutcliffe JS, Casey BM, Beaudet AL, Hennekam RCM, Willems PJ (1996b). Detection of impriting mutations in Angelman syndrome using a probe for exon α of SNRPN. Am. J. Med. Genet. 63: 414-5

Bird A (1986). CpG-rich islands and the function of DNA methylation. Nature 321: 209-13

Boguski MS, Schuler GD (1995). ESTablishing a human transcript map. Nat. Genet. 10: 369-71

Bottani A, Robinson WP, DeLozier-Blanchet CD, Engel E, Morris MA, Schmitt B, Thun-Hohenstein L, Schinzel A (1994). Angelman syndrome due to paternal uniparental disomy of chromosome 15: A milder phenotype? Am. J. Med. Genet. 51: 35-40

Boyd SG, Harden A, Patton MA (1988). The EEG in early diagnosis of the Angelman (Happy Puppet) syndrome. Eur. J. Paediatr. 147: 508-13

Brennan MB, Hochgeschwender U (1995). Commentary: So many needles, so much hay. Hum. Mol. Genet. 4: 153-6

Breathnach R, Chambon P (1981). Organization and expression of eucaryotic split genes coding for proteins. Annu. Rev. Biochem. 50: 349-83

Brzustowicz LM, Allitto BA, Matseoane D, Theve R, Michaud L, Chatkupt S, Sugarman E, Penchaszadeh GK, Suslak L, Koenigsberger MR, et al. (1994). Paternal isodisomy for chromosome 5 in a child with spinal muscular atrophy. Am. J. Hum. Genet. 54: 482-8

Buckler AJ, Chang DD, Graw SL, Brook JD, Haber DA, Sharp PA, Housman DE (1991). Exon amplification: a strategy to isolate mammalian genes based on RNA splicing. Proc. Natl. Acad. Sci. USA. 88: 4005-9

Buiting K, Neumann M, Ludecke H-J, Senger G, Claussen U, Antich J, Passarge E, Horsthemke B (1990). Microdissection of the Prader-Willi syndrome chromosome region and identification of potential gene sequences. Genomics 6: 521-7

Buiting K, Greger V, Brownstein BH, Mohr RM, Voiculescu I, Winterpacht A, Zabel B, 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. USA 89: 5457-61

Buiting K, Dittrich B, Grob S, Greger V, Lalande M, Robinson W, Mutirangura A, Ledbetter D, Horsthemke B (1993). Molecular definition of the Prader-Willi syndrome chromosome region and orientation of the SNRPN gene. Hum. Mol. Genet. 2: 1991-4

Buiting K, Dittrich B, Robinson WP, Guitart M, Abeliovich D, Lerer I, Horsthemke B (1994). Detection of aberrant DNA methylation in unique Prader-Willi syndrome patients and its diagnostic implications. Hum. Mol. Genet. 3: 893-5

Buiting K, Saitoh S, Gross S, Dittrich B, Scwartz S, Nicholls RD, Horsthemke B (1995). Inherited microdeletions in the Angelman and Prader-Willi syndromes define an imprinting centre on human chromosome 15. Nat. Genet. 9: 395-400

Buiting K, Dittrich B, Endele S, Horsthemke B (1997). Identification of novel exons 3' to the human SNRPN gene. Genomics 40: 132-7

Buiting K, Dittrich B, Gorss S, Lich C, Farber C, Buchholz T, Smith E, Reis A, Burger J, Nothen MM, Barth-Witte U, Janssen B, Abeliovich D, Lerer I, van den Ouweland AM, Halley DJ, Schrander-Stumpel C, Smeets H, Meinacke P, Malcolm S, Gardner A, Lalande M, Nicholls RD, Friend K, Horsthemke B, et al. (1998). Sporadic imprinting defects in Prader-Willi syndrome and Angelman syndrome: implications for imprint-switch models, genetic counselling, and prenatal diagnosis. Am. J. Hum. Genet. 63: 170-80

Buntinx IM, Hennekam CM, Brouer OF, Stroink H, Beuten J, Mangelschots K, Fryns JP (1995). Clinical profile of Angelman syndrome at different ages.
Am. J. Med. Genet. 56: 176-183

B rger J, Buiting K, Dittrich B, Grob S, Lich C, Sperling K, Horsthemke B, Reis A (1997a). Different mechanisms and recurrence risks of imprinting defects in Angelman syndrome. Am. J. Hum. Genet. 61: 88-93

B rger J, Hennies HC, Wessel T, Burkhard E, Sperling K, Reis A (1997b). An integrated restriction-site, STS and transcript based BAC/PAC contig map of human chromosome 15q12. Nat. Genet. 9: Abstracts P4.187

Burke DT, Carle GF, Olson MV (1987). Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. Science 236: 806-12

Burke LW, Wiley JE, Glenn CC, Driscoll DJ, Loud KM, Smith AJW, Kushnick T (1996). Familial cryptic translocation resulting in Angelman syndrome: Implications for imprinting or location of the Angelman Gene? Am. J. Hum. Genet. 58: 777-84

Butler MG, Palmer CG (1983). Parental origin of chromosome 15 deletion in Prader-Willi syndrome. Lancet 1: 1285-6

Butler MG, Meaney FJ, Palmer CG (1986). Clinical and cytogenetic survey of 39 individuals with Prader-Labhart-Willi syndrome. Am. J. Med. Genet. 23: 793-809

Butler MG (1990). Prader-Willi syndrome: Current understanding of cause and diagnosis. Am. J.Med. Genet. 35: 319-32

Butler MG (1995). High resolution chromosome analysis and fluorescence in situ hybridization in patients referred for Prader-Willi or Angelman syndrome. Am. J. Med. Genet. 56: 420-2

Caspary T, Cleary MA, Baker CC, Guan XJ, Tilghman SM (1998). Multiple mechanisms regulate imprinting of the mouse distal chromosome 7 gene cluster. Mol. Cell. Biol. 18: 3466-74

Cassidy SB (1984). Prader-Willi syndrome. Curr. Prob. Pediatr. 14: 1-55

Cassidy SB, Thuline HC, Holm VA (1984). Deletion of chromosome 15 (q11-13) in a Prader-Labhart-Willi syndrome clinic population. Am. J. Med. Genet. 17: 485-95

Cassidy SB (1987). Recurrence risk in Prader-Willi syndrome. Am. J. Med. Genet. 28: 59-60

Cassidy SB, Lai L-W, Erickson RP, Magnuson L, Thomas E, Gendron R, Herrman J (1992). Trisomy 15 with loss of the paternal 15 as a cause of Prader-Willi syndrome due to maternal disomy. Am. J. Hum. Genet. 51: 701-8

Cassidy SB, Beaudet AL, Knoll JHM, Ledbetter DH, Nicholls RD, Schwartz S, Butler MG, Watson M (1996). Diagnostic testing for Prader-Willi and Angelman syndromes: Report of the ASHG/ACMG Test and Technology Transfer Committee. Am. J. Hum. Genet. 58: 1085-8

Cattanach BM, Kirk M (1985). Differential activity of maternally and paternally derived chromosome regions in mice. Nature 315: 496-8

Cattanach BM (1986). Parental origin effects in mice. J. Embryol. exp. Morph. 97 Supplement; 137-150

Cattanach BM, Beechey CV (1990). Autosomal and X-chromosome imprinting. Development: Supplement; 63-72

Cattanach BM, Barr JA, Evans EP, Burtenshaw M, Beechey CV, Leff SE, Brannan CI, Copeland NG, Jenkins NA, Jones J (1992). A candidate mouse model for Prader-Willi syndrome which shows an absence of *Snrpn* expression. Nat. Genet. 2: 270-4

Cattanach BM, Barr J, Beechey CV, Bressler J, Sutcliffe JS, Beaudet AL, Martin JL, Noebels JL, Jones J (1996). A mouse model for Angelman syndrome. Am. J. Hum. Genet. 59: Supplement; A59

Cattanach BM, Barr JA, Beechey CV, Martin J, Noebels J, Jones J (1997). A candidate model for Angelman syndrome in the mouse. Mammalian Genome 8: 472-8

Chaillet JR, Vogt TF, Beier DR, Leder P (1991a). Parental-specific methylation of an imprinted transgene is established during gametogenesis and progressively changes during embryogenesis. Cell 66: 77-83

Chaillet JR, Knoll JHM, Horsthemke B, Lalande M (1991b). The syntenic relationship between the critical deletion region for the Prader-Willi/ Angelman syndromes and proximal mouse chromosome 7. Genomics 11: 773-6

Chan CT, Clayton-Smith J, Cheng XJ, Buxton J, Webb T, Pembrey ME, Malcolm S (1993). Molecular mechanisms in Angelman syndrome: a survey of 93 patients.
J. Med. Genet. 30: 895-902

Chen Z, Parent L, Maniatis T (1996). Site-specific phosporylation of IkBa by a novel ubiquitination-dependent protein kinase activity. Cell 84: 853-62

Christian SL, Smith ACM, Macha M, Black SH, Elder FFB, Johnson JM-P, Resta RG, Surti U, Suslak L, Verp MS, Ledbetter DH (1996). Prenatal diagnosis of uniparental disomy 15 following trisomy 15 mosaicism. Prenatal Diagnosis 16: 323-32

Christian SL, Bhatt NK, Martin SA, J.S. S, Kubota T, Huang B, Mutirangura A, Chinault AC, Beaudet AL, Ledbetter DH (1998). Integrated YAC contig map of the Prader-Willi / Angelman region on chromosome 15q11-q13 with average STS spacing of 35kb. Genome Res 8: 146-57

Chumakov IM, Rigault P, Le Gall I, Bellanne-Chantelot C, Billaut A, Guillou S, Soularue P, Guasconi G, Poullier E, Gros I, et al. (1995). A YAC contig map of the human genome. Nature 377: Supplement; 175-83

Church DM, Stotler CJ, Rutter JL, Murrell JR, Trofatter JA, Buckler AJ (1994). Isolation of genes from complex sources of mammalian genomic DNA using exon amplification.
Nat. Genet. 6: 98-105

Ciechanover A (1994). The ubiquitin-proteasome proteolytic pathway. Cell 79: 13-21

Clayton-Smith J, Webb T, Robb SA, Dijkstra I, Willems P, Lam S, Cheng X-J, Pembrey ME, Malcolm S (1992a). Further evidence for dominant inheritance at the chromosome 15q11-13 locus in familial Angelman syndrome. Am. J. Med. Genet. 44: 256-60

Clayton-Smith J, Webb T, Pembrey ME, Nichols M, Malcolm S (1992b). Maternal origin of deletion 15q11-13 in 25/25 cases of Angelman syndrome. Hum. Genet. 88: 376-8

Clayton-Smith J, Pembrey ME (1992). Angelman syndrome. J. Med. Genet. 29: 412-5

Clayton-Smith J (1992). Angelman's syndrome. Arch. Dis. Child. 67: 889-91

Clayton-Smith J (1993). Clinical research on Angelman syndrome in the United Kingdom: Observations on 82 affected individuals. Am. J. Med. Genet. 46: 12-5

Clayton-Smith J, Driscoll DJ, Waters MF, Webb T, Andrews T, Malcolm S, Pembrey ME, Nicholls RD (1993). Difference in methylation patterns within the D15S9 region of chromosome 15q11-13 in first cousins with Angelman syndrome and Prader-Willi syndrome. Am.J. Med. Genet. 47: 683-6

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

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

Collins FS (1995). Positional cloning moves from perditional to traditional. Nat. Genet. 9: 347-50

Cook EH Jr, Lindgren V, Leventhal BL, Courchesne R, Lincoln A, Shulman C, Lord C, Courchesne E (1997). Autism or atypical autism in maternally but not paternally derived proximal 15q duplication. Am. J. Hum. Genet. 60: 928-34

Cox DR, Burmeister M, Price ER, Kim S, Myers RM (1990). Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. Science 250: 245-50

Crolla JA, Harvey JF, Sitch FL, Dennis NR (1995). Supernumerary marker 15 chromosomes: a clinical, molecular and FISH approach to diagnosis and prognosis. Hum. Genet. 95: 161-70

Cross SH, Charlton JA, Nan X, Bird AP (1994). Purification of CpG islands using a methylated DNA binding column. Nat. Genet. 6: 236-44

DeChiara TM, Robertson EJ, Efstratiadis A (1991). Parental imprinting of the mouse insulin-like growth factor II gene. Cell 64: 849-59

DeLorey TM, Olsen RW (1992). Gamma aminobutyric acid$_A$ receptor structure and function. J. Biol. Chem. 267: 16747-50

DiFiglia M, Sapp E, Chase KO, Davies SW, Bates GP, Vonsattel JP, Aronin N, (1997). Aggregation of Huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain. Science 277: 1990-4

Dittrich B, Robinson WP, Knoblauch H, Buiting K, Schmidt K, Gillessen-Kaesbach G, Horsthemke B (1992). Molecular diagnosis of the Prader-Willi and Angelman syndromes by detection of parent-of-origin specific DNA methylation in 15q11-13. Hum. Genet. 90: 313-5

Dittrich B, Buiting K, Grob S, Horsthemke B (1993). Characterisation of a methylation imprint in the Prader-Willi syndrome chromosome region. Hum. Mol. Genet. 2: 1995-9

Dittrich B, Buiting K, Korn B, Rickard S, Buxton J, Saitoh S, Nicholls RD, Poustka A, Winterpacht A, Zabel B, Horsthemke B (1996). Imprint switching on human chromosome 15 may involve alternative transcripts of the SNRPN gene. Nat. Genet. 14: 163-70

Dlott B, d'Azzo A, Quon DV, Neufeld EF, (1990). Two mutations produce intron insertion in mRNA and elongated beta-subunit of human beta-hexosaminidase.
J. Biol. Chem. 265: 17921-7

Donlon TA, Lalande M, Wyman A, Bruns G, Latt SA (1986). Isolation of molecular probes associated with the chromosome 15 instability in the Prader-Willi syndrome.
Proc. Natl. Acad. Sci. USA 83: 4408-12

Donlon TA (1988). Similar molecular deletions on chromosome 15q11.2 are encountered in both the Prader-Willi and Angelman syndromes. Hum. Genet. 80: 322-8

Driscoll DJ, Waters MF, Williams CA, Zori RT, Glenn CC, Avidano KM, Nicholls RD (1992). A DNA methylation imprint, determined by the sex of the parent, distiguishes the Angelman and Prader-Willi syndromes. Genomics 13: 917-24

Dykens EM, Leckman JF, Cassidy SB (1996). Obsessions and compulsions in Prader-Willi syndrome. J. Child Psychol. Psychiatry 37: 995-1002

Engel E (1980). A new genetic concept: Uniparental disomy and its potential effect, isodisomy. Am. J. Med. Genet. 6: 137-43

Erdel M, Schuffenhauer S, Buchholz B, Barthe-Witte U, Kochl S, Utermann B, Duba H-C, Utermann G (1996). Routine screening for microdeletions by FISH in 77 patients suspected of having Prader-Willi or Angelman syndromes using YAC clone 273A2 (D15S10).
Hum. Genet. 97: 784-93

Erdmann VA, Szymanski M, Hochberg A, de Groot N, Barciszewski J (1999). Collection of MRNA-like non-coding RNAs. Nucl. Acids Res. 27: 192-5

Fairbrother UL, Buxton JL, Hitchins MP, Rickard SJ, Pembrey ME, Malcol S (1997) Mutations in *UBE3A* (E6AP) associated with Angelman syndrome.
Nat. Genet. 9: Abstracts P4.222

Fang P, Lev-Lehman E, Tsai TF, Matsuura T, Benton CS, Sutcliffe JS, Christian SL, Kubota T, Halley DJ, Meijers-Heijboer H, Langlois S, Graham JM Jr, Beuten J, Willems PJ, Ledbetter DH, Beaudet AL (1999). The spectrum of mutations in UBE3A causing Angelman syndrome. Hum. Mol. Genet. 8: 129-135

Fearon ER, Cho KR, Nigro JM, Kern SE, Simons JW, Ruppert JM, Hamilton SR, Preisinger AC, Thomas G, Kinzler KW, Vogelstein B (1990). Identification of a chromosome 18q gene that is altered in colorectal cancers. Science 247: 49-56

Feil R, Walter J, Allen ND, Reik W (1994). Developmental control of allelic methylation in the imprinted mouse Igf2 and H19 genes. Development 120: 2933-43

Ferguson-Smith AC, Cattanach BM, Barton SC, Beechey CV, Surani MA (1991). Embryological and molecular investigations of parental imprinting on mouse chromosome 7.
Nature 351: 667-70

Ferguson-Smith AC (1996). Imprinting moves to the centre. Nat. Genet. 14: 119-21

Fink JK, Wu CT, Jones SM, Sharp GB, Lange BM, Lesicki A, Reinglass T, Varvil T, Otterud B, Leppert M (1995). Autosomal dominant familial spastic paraplegia: tight linkage to chromosome 15q. Am. J. Hum. Genet. 56: 188-92

Fraccaro M, Zuffardi O, Buhler EM, Jurik LP (1977). 15/15 translocation in Prader-Willi syndrome. J. Med. Genet. 14: 275-6

Fryburg JS, Breg WR, Lindgren V (1991). Diagnosis of Angelman syndrome in infants. Am. J. Med. Genet. 38: 58-64

Fujimaru M, Tanaka A, Choeh K, Wakamatsu N, Sakuraba H, Isshiki G, (1998). Two mutations remote from an exon/intron junction in the beta-hexosaminidase beta-subunit gene affect 3'-splice site selection and cause Sandhoff disease. Hum. Genet. 103: 462-9

Gabriel JM, Higgins MJ, Gebuhr TC, Shows TB, Saitoh S, Nicholls RD (1998). A model system to study genomic imprinting of human genes.
Proc. Natl. Acad. Sci. USA 95: 14857-62

Gardner JM, Nakatsu Y, Gondo Y, Lee S, Lyon MF, King RA, Brilliant MH (1992). The mouse pink-eyed dilution gene: Association with human Prader-Willi and Angelman syndromes. Science 257: 1121-4

Giannoukakis N, Deal C, Paquette J, Goodyer CG, Polychronakos C (1993). Parental genomic imprinting of the human *IGF2* gene. Nat. Genet. 4: 98-101

Gillessen-Kaesbach G, Albrecht B, Passarge E, Horsthemke B (1995a). Further patient with Angelman syndrome due to paternal disomy of chromosome 15 and a milder phenotype. Am. J. Med. Genet. 56: 328-9

Gillessen-Kaesbach G, Gross S, Kaya-Westerloh S, Passarge E, Horsthemke B (1995b). DNA methylation based testing of 450 patients suspected of having Prader-Willi syndrome. J. Med. Genet. 32: 88-92

Glenn CC, Porter KA, Jong MTC, Nicholls RD, Driscoll DJ (1993a.) Functional imprinting and epigenetic modification of the human SNRPN gene. Hum. Mol. Genet. 2: 2001-5

Glenn CC, Nicholls RD, Robinson WP, Saitoh S, Niikawa N, Schinzel A, Horsthemke B, Driscoll DJ (1993b). Modification of 15q11-q13 DNA methylation imprints in unique Angelman and Prader-Willi patients. Hum. Mol. Genet. 2: 1377-82

Glenn CC, Saitoh S, Jong MTC, Filbrandt MM, Surti U, Driscoll DJ, Nicholls RD (1996). Gene structure, DNA methylation, and imprinted expression of the human SNRPN gene. Am. J. Hum. Genet. 58: 335-46

Greger V, Woolf E, Lalande M (1993). Cloning of the breakpoints of a submicroscopic deletion in an Angelman syndrome patient. Hum. Mol. Genet. 2: 921-4

Greger V, Reis A, Lalande M (1994). The critical region for Angelman syndrome lies between D15S122 and D15S113. Am. J. Med. Genet. 53: 396-8

Greger V, Knoll JH, Woolf E, Glatt K, Tyndale RF, DeLorey TM, Olsen RW, Tobin AJ, Sikela JM, Nakatsu Y, et al. (1995). The gamma-aminobutyric acid receptor gamma 3 subunit gene (GABRG3) is tightly linked to the alpha 5 subunit gene (GABRA5) on human chromosome 15q11-q13 and is transcribed in the same orientation. Genomics 26: 258-64

Grompe M (1993). The rapid detection of unknown mutations in nucleid acids. Nat. Genet. 5: 111-7

Guillemot F, Caspary T, Tilghman SM, Copeland NG, Gilbert DJ, Jenkins NA, Anderson DJ, Joyner AL, Rossant J, Nagy A (1995). Genomic imprinting of *Mash2*, a mouse gene required for trophoblast development. Nat. Genet. 9: 235-42

Gunaratne PH, Nakao M, Ledbetter DH, Sutcliffe JS, Chinault AC (1995). Tissue-specific and allele-specific replication timing control in the imprinted human Prader-Willi syndrome region. Genes Dev. 9: 808-20

Hall JG (1990). Genomic imprinting: review and relevance to human diseases. Am. J. Hum. Genet. 46: 857-73

Hamabe J, Fukushima Y, Harada N, Abe K, Matsuo N, Nagai T, Yoshioka A, Tonoki H, Tsukino R, Niikawa N (1991a). Molecular study of the Prader-Willi syndrome: deletion, RFLP and phenotype analyses of 50 patients. Am. J. Med. Genet. 41: 54-63

Hamabe J, Kuroki Y, Imaizumi K, Sugimoto T, Fukushima Y, Yamaguchi A, Izumikawa Y, Niikawa N (1991b). DNA deletion and its parental origin in Angelman syndrome patients. Am. J. Med. Genet. 41: 64-8

Hannula K, Lipsanen-Nyman M, Hoglund P, Holmberg C, Scherer SW, Kere J (1998). Systematic analysis of methylation differences in paternal and maternal uniparental disomy for chromosome 7. Am. J. Hum. Genet. 63: Supplement; 1896

Hassold T, Matsuyama A (1979). Origin of trisomies in human spontaneous abortions. Hum. Genet. 46: 285-94

Hassold T, Chiu D (1985). Maternal age-specific rates of numerical chromosome abnormalities with special reference to trisomy. Hum. Genet. 70: 11-7

Hastbacka J, de la Chapelle A, Mahtani MM, Clines G, Reeve-Daly MP, Daly M, Hamilton BA, Kusumi K, Trivedi B, Weaver A, Coloma A, Lovett M, Buckler A, Kaitila I, Lander ES (1994). The diastrophic dysplasia gene encodes a novel sulfate transporter: Positional cloning by fine-structure linkage disequilibrium mapping. Cell 78: 1073-87

Haupt Y, Maya R, Kazaz A, Oren M (1997). Mdm2 promotes the rapid degradation of p53. Nature 387: 296-9

Hawkey CJ, Smithies A (1976). The Prader-Willi syndrome with a 15/15 translocation. Case report and review of the literature. J. Med. Genet. 13: 152-63

Hayashizaki Y, Shibata H, Hirotsune S, Sugino H, Okazaki Y, Sasaki N, Hirose K, Imoto H, Okuizumi H, Muramatsu M, et al. (1994). Identification of an imprinted U2af binding protein related sequence on mouse chromosome 11 using the RLGS method. Nat. Genet. 6: 33-40

Heard E, Clerc P, Avner P (1997). X-chromosome inactivation in mammals. Annu. Rev. Genet. 31: 571-610

Henry I, Puech A, Riesewijk A, Ahnine L, Mannens M, Beldjord C, Bitoun P, Tournade MF, Landrieu P, Junien C (1993). Somatic mosaicism for partial isodisomy in Wiedemann-Beckwith syndrome: a post-zygotic event. Eur. J. Hum. Genet. 1: 19-29

Hochstrasser M (1996). Protein degradation or regulation: Ub the judge. Cell 84: 813-5

Hollstein M, Sidransky D, Vogelstein B, Harris CC (1991). p53 mutations in human cancers. Science 253: 49-53

Hoglund P, Holmberg C, de la Chapelle A, Kere J (1994). Paternal isodisomy for chromosome 7 is compatible with normal growth and development in a patient with congenital chloride diarrhea. Am. J. Hum. Genet. 55: 747-52

Holm VA, Cassidy SB, Butler MG, Hanchett JM, Greenswag LR, Whitman BY, Greenberg F (1993). Prader-Willi syndrome: Consensus diagnostic crieteria. Pediatrics 91: 398-402

Holmquist GP (1987). Role of replication time in the control of tissue-specific gene expression. Am. J. Hum. Genet. 40: 151-73

Horsthemke B, Maat-Kievit A, Sleegers E, van den Ouweland A, Buiting K, Lich C, Mollevanger P, Beverstock G, Gillessen-Kaesbach G, Schwanitz G (1996). Familial translocations involving 15q11-q13 can give rise to interstitial deletions causing Prader-Willi or Angelman syndrome. J. Med. Genet. 33: 848-51

Horsthemke B (1997). Structure and function of the human chromosome 15 imprinting centre. J. Cell. Physiol. 173: 237-41

Hu JF, Oruganti H, Vu TH, Hoffman AR (1998). Tissue-specific imprinting of the mouse insulin-like growth factor II receptor gene correlates with differential allele-specific DNA methylation. Mol. Endocrinol. 12: 220-32

Hudson TJ, L.D. S, Gerety SS, Ma J, Castle AB, Silva J, Slonim DK, Baptista R, Kruglyak L, Xu S-H, al. e (1995). An STS-based map of the human genome. Science 270: 1945-54

Huibregtse JM, Scheffner M, Howley PM (1991). A cellular protein mediates association of p53 with the E6 oncoprotein of human papillomavirus types 16 or 18. EMBO J. 10: 4129-35

Huibregtse JM, Scheffner M, Howley PM (1993). Cloning and expression of the cDNA for E6-AP, a protein that mediates the interaction of the human papillomavirus E6 oncoprotein with p53. Mol. Cell. Biol. 13: 775-84

Huibregtse JM, Scheffner M, Beaudenon S, Howley PM (1995). A family of proteins structurally and functionally related to the E6-AP ubiquitin-protein ligase. Proc. Natl. Acad. Sci. USA 92: 2563-7

Hulten M, Armstrong S, Challinor P, Gould C, Hardy G, Leedham P, Lee T, McKeown C (1991). Genomic imprinting in an Angelman and Prader-Willi translocation family. Lancet 338: 638-9

Huntriss JD, Barr JA, Horn DA, Williams DG, Latchman DS (1994). Mice lacking Snrpn expression show normal regulation of neuronal alternative splicing events. Mol. Biol. Rep. 20: 19-25

Huntriss (1998). 8th Mammalian Genetics and Development Workshop. London, United Kingdom, 19-21 November 1997. Abstracts. Genet. Res. 72: 59-72

Huntriss J, Daniels R, Bolton V, Monk M (1998). Imprinted expression of *SNRPN* in human preimplantation embryos. Am. J. Hum. Genet. 63: 1009-14

Huq AHMM, Sutcliffe JS, Nalao M, Shen Y, Gibbs RA, Beaudet AL (1997). Sequencing and functional analysis of the SNRPN promoter: In vitro methylation abolishes promoter activity. Genome Res. 7: 642-8

Imaizumi K, Takada F, Kuroki Y, Naritomi K, Hamabe J, Niikawa N (1990). Cytogenetic and molecular study of the Angelman syndrome. Am. J. Med. Genet. 35: 314-8

Izumikawa Y, Naritomi K, Hirayama K (1991). Replication asynchrony between homologs 15q11.2: cytogenetic evidence for genomic imprinting. Hum. Genet. 87: 1-5

Jay V, Becker LE, Chan FW, Perry TL Sr (1991). Puppet-like syndrome of Angelman: a pathologic and neurochemical study. Neurology 41: 416-22

Jay P, Rougelle C, Massacrier A, Moncla A, Mattei M-G, Malzac P, Roeckel N, Taviaux S, Berge Lefranc J-L, Cau P, Berta P, Lalande M, Muscatelli F (1997). The human necdin gene, NDN, is maternally imprinted and located in the Prader-Willi syndrome chromosomal region. Nat. Genet. 17: 357-61

Jerome CA, Scherer SW, Tsui LC, Gietz RD, Triggs-Raine B (1997). Assignment of growth factor receptor-bound protein 10 (GRB10) to human chromosome 7p11.2-p12. Genomics 40: 215-6

Jiang Y-H, Armstrong D, Albrecht U, Atkins CM, Noebels JL, Eichele G, Sweatt JD, Beaudet AL (1998). Mutation of the Angelman ubiquitin ligase in mice caused increased cytoplasmic p53 and deficits of contextual learning and long-term potentiation. Neuron 21: 799-811

Jinno Y, Yun K, Nishiwaki K, Kubota T, Ogawa O, Reeve AE, Niikawa N (1994). Mosaic and polymorphic imprinting of the WT1 gene in humans. Nat. Genet. 6: 305-9

Jong MT, Gray TA, Ji Y, Glenn CC, Saitoh S, Driscoll DJ, Nicholls RD (1999a). A novel imprinted gene, encoding a RING zinc-finger protein, and overlapping antisense transcript in the prader-willi syndrome critical region. Hum. Mol. Genet. 8: 783-93

Jong MT, Carey AH, Caldwell KA, Lau MH, Handel MA, Driscoll DJ, Stewart CL, Rinchik EM, Nicholls RD (1999b). Imprinting of a RING zinc-finger encoding gene in the mouse chromosome region homologous to the prader-willi syndrome genetic region. Hum. Mol. Genet. 8: 795-803

Kajii T, Ohama K (1977). Androgenetic origin of hydatidiform mole. Nature 268: 633-4

Kalscheuer VM, Mariman EC, Schepens MT, Rehder H, Ropers H-H (1993). The insulin-like growth factor type-2 receptor gene is imprinted in the mouse but not in humans. Nat. Genet. 5: 74-8

Kalousek DK, Barrett I (1994). Confined placental mosaicism and stillbirth. Pediatr. Pathol. 14: 151-9

Kaneko-Ishino T, Koroiwa Y, Kohda T, Suzuki R, Yokoyama M, Viville S, Barton SC, Ishino F, Surani MA (1995). *Peg1/Mest* imprinted gene on chromosome 6 identified by cDNA subtraction hybridisation. Nat. Genet. 11: 52-8

Kaplan LC, Wharton R, Elias E, Mandell F, Donlon T, Latt SA (1987). Clinical heterogeneity associated with deletions of the long arm of chromosome 15: Report of 3 new cases and their possible genetic significance. Am. J. Med. Genet. 28: 45-53

Kass DH, Batzer MA (1995). Inter-Alu polymerase chain reaction: advancements and applications. Anal. Biochem. 228: 185-93

Kelsey G, Reik W (1997). Imprint switch mechanism indicated by mutations in Prader-Willi and Angelman syndromes. BioEssays 19: 361-5

Khan AS, Wilcox AS, Polymeropoulos MH, Hopkins JA, Stevens TJ, Robinson M, Orpana AK, Sikela JM (1992). Single pass sequencing and physical and genetic mapping of human brain cDNAs. Nat. Genet. 2: 180-5

Kirkilionis AJ, Gregory CA, Hamerton JL (1991). Long-range restriction mapping and linkage analysis of the Prader-Willi chromosome region (PWCR). Genomics 9: 524-35

Kishino T, Lalande M, Wagstaff J (1997). UBE3A/E6-AP mutations cause Angelman syndrome. Nat. Genet. 15: 70-3

Kishino T, Wagstaff J (1998). Genomic organisation of the *UBE3A* / E6-AP gene and related pseudogenes. Genomics 47: 101-7

Kitsberg D, Selig S, Brandeis M, Simon I, Keshet I, Driscoll DJ, Nicholls RD, Cedar H (1993). Allele-specific replication timing of imprinted gene regions. Nature 364: 459-63

Knoll JHM, Nicholls RD, Magenis RE, Graham J, J.M., Lalande M, Latt SA (1989). Angelman and Prader-Willi syndromes share a common chromosome 15 deletion but differ in parental origin of the deletion. Am. J. Med. Genet. 32: 285-90

Knoll JHM, Nicholls RD, Magenis RE, Glatt K, Graham Jr. JM, Kaplan L, Lalande M (1990). Angelman syndrome: Three molecular classes identified with chromosome 15q11q13-specific DNA markers. Am. J. Hum. Genet. 47: 149-55

Knoll JHM, Wagstaff J, Lalande M (1993a). Cytogenetic and molecular studies in the Prader-Willi and Angelman syndromes: An overview. Am. J. Med. Genet. 46: 2-6

Knoll JHM, Sinnett D, Wagstaff J, Glatt K, Schantz Wilcox A, Whiting PM, Wingrove P, Sikela JM, Lalande M (1993b). FISH ordering of reference markers and of the gene for the a5 subunit of the g-aminobutyric acid receptor (GABRA5) within the Angelman and Prader-Willi syndrome chromosomal regions. Hum. Mol. Genet. 2: 183-9

Knoll JHM, Cheng S-D, Lalande M (1994). Allele specificity of DNA replication timing in the Angelman/Prader-Willi syndrome imprinted chromosomal region. Nat. Genet. 6: 41-6

Kono T, Obata Y, Yoshimzu T, Nakahara T, Carroll J (1996). Epigenetic modifications during oocyte growth correlates with extended parthenogenetic development in the mouse. Nat. Genet. 13: 91-4

Korn B, Sedlacek Z, Manca A, Kioschis P, Konecki D, Lehrach H, Poustka A (1992). A strategy for the selection of transcribed sequences in the Xq28 region. Hum. Mol. Genet. 1: 235-42

Kotzot D, Schmitt S, Bernasconi F, Robinson WP, Lurie IW, Ilyina H, Mehes K, Hamel BC, Otten BJ, Hergersberg M, et al. (1995). Uniparental disomy 7 in Silver-Russell syndrome and primordial growth retardation. Hum. Mol. Genet. 4: 583-7

Kubota T, Sutcliffe JS, Aradhya S, Gillessen-Kaesbach G, Christian SL, Horsthemke B, Beaudet AL, Ledbetter DH (1996a). Validation studies of SNRPN methylation as a diagnostic test for Prader-Willi syndrome. Am.J. Med. Genet. 66: 77-80

Kubota T, Aradhya S, Macha M, Smith ACM, Surh LC, Satish J, Verp MS, Nee HL, Johnson A, Christian SL, Ledbetter DH (1996b). Analysis of parent of origin specific DNA methylation at SNRPN and PW71 in tissues: implication for prenatal diagnosis. J. Med. Genet. 33: 1011-4

Kumar S, Kao WH, Howley PM (1997). Physical interaction between specific E2 and Hect E3 enzymes determines functional cooperativity. J. Biol. Chem. 272: 13548-54

Kuroiwa Y, Kaneko-Ishino T, Kagitani F, Kohda T, Li L-L, Tada M, Suzuki R, Yokoyama M, Shioishi T, Wakana S, Barton SC, Ishino F, Surani MA (1996). *Peg3* imprinted gene on proximal chromosome 7 encodes for a zinc finger protein. Nat. Genet. 12: 186-90

Kuwano A, Mutirangura A, Dittrich B, Buiting K, Horsthemke B, Saitoh S, Niikawa N, Ledbetter SA, Greenberg F, Chinault AC, Ledbetter DH (1992). Molecular dissection of the Prader-Willi/Angelman syndrome region (15q11-13) by YAC cloning and FISH analysis. Hum. Mol. Genet. 1: 417-25

Langlois S (1994). Genomic imprinting: a new mechanism for disease. Paediatr. Pathol. 14: 161-5

LaSalle JM, Lalande M (1995). Domain organisation of allele-specific replication within the GABRB3 gene cluster requires a biparental 15q11-13 contribution. Nat. Genet. 9: 386-94

Ledbetter DH, Riccardi VM, Airhart DD, Strobel RJ, Keenan BS, Crawford JD (1981). Deletions of chromosome 15 as a cause of the Prader-Willi syndrome. N. Engl. J. Med. 304: 325-9

Ledbetter DH, Engel E (1995). Uniparental disomy in humans: development of an imprinting map and its implications for prenatal diagnosis. Hum. Mol. Genet. 4: 1757-64

Lee ST, Nicholls RD, Bundey S, Laxova R, Musarella M, Spritz RA (1994). Mutations of the P gene in oculocutaneous albinism, ocular albinism, and Prader-Willi syndrome plus albinism. N. Engl. J. Med. 330: 529-34

Lee MP, Hu R-J, Johnson LA, Feinberg AP (1997). Human KVLQT1 gene shows tissue-specific imprinting and encompasses Beckwith-Wiedemann syndrome chromosomal rearrangements. Nat. Genet. 15: 181-5

Lee JT, Davidow LS, Warshawsky D (1999). *Tsix*, a gene antisense to Xist at the X-inactivation centre. Nat. Genet. 21: 400-4

Leff SE, Brannan CI, Reed ML, Ozcelik T, Francke U, Copeland N, Jenkins NA (1992). Maternal imprinting of the mouse *Snrpn* gene and conserved linkage homology with the human Prader-Willi syndrome region. Nat. Genet. 2: 259-64

Leighton PA, Saam JR, Ingram RS, Tilghman SM (1996). Genomic imprinting in mice: its function and mechanism. Biol. Reprod. 54: 273-8

Lengauer C, Green ED, Cremer T (1992). Fluorescence *in situ* hybridisation of YAC clones after Alu-PCR amplification. Genomics 13: 826-8

Lennon G, Auffray C, Polymeropoulos M, Soares MB (1996). The I.M.A.G.E. Consortium: an integrated molecular analysis of genomes and their expression. Genomics 33: 151-2

Lewin B (1994). Genes V. Oxford University Press, New York

Li S, Klein ES, Russo AF, Simmons DM, Rosenfeld MG (1989). Isolation of cDNA clones encoding small nuclear ribonuclear-associated proteins with different tissue specificities. Proc. Natl. Acad. Sci. USA 86: 9778-82

Li E, Bestor TH, Jaenisch R (1992). Targeted mutation of the DNA methyltransferase gene reults in embryonic lethality. Cell 69: 915-26

Li M, Kuivenhoven JA, Ayyobi AF, Pritchard PH, (1998). T→G or T→A mutation introduced in the branchpoint of intron 4 of lecithin:cholesterol acyltransferase (LCAT) gene: intron retention causing LCAT deficiency. Biochim. Biophys. Acta. 1391: 256-64

Lighten AD, Hardy K, Winston RML, Moore GE (1997). *IGF2* is parentally imprinted in human preimplantation embryos. Nat Genet 15: 122-3

Linder D, Hecht F, B.K. M, Campbell JR (1975). Origin of extragonadal teratomas and endodermal sinus tumours. Nature 254: 597-8

Lovett M, Kere J, Hinton LM (1991). Direct selection: a method for the isolation of cDNAs encoded by large genomic regions. Proc. Natl. Acad. Sci. USA 88: 9628-32

Lovett M (1994). Fishing for complements: finding genes by direct cDNA selection. Trends Genet 10: 352-7

Lyon MF, King TR, Gondo Y, Gardner JM, Nakatsu Y, Eicher EM, Brilliant MH (1992). Genetic and molecular analysis of recessive alleles at the pink-eyed dilution (*p*) locus of the mouse. Proc. Natl. Acad. Sci. USA 89: 6968-72

MacDonald HR, Wevrick R (1997). The necdin gene is deleted in Prader-Willi syndrome and is imprinted in human and mouse. Hum.Mol. Genet. 6: 1873-8

Magenis RE, Brown MG, Lacy DA, Budden S, LaFranchi S (1987). Is Angelman syndrome an alternate result of del(15)(q11q13)? Am. J. Med. Genet. 28: 829-38

Magenis RE, Toth-Fejel S, Allen LJ, Black M, Brown MG, Budden S, Cohen R, Friedman JM, Kalousek D, Zonana J, Lacy D, LaFranchi S, Lahr M, Macfarlane J, Williams CPS (1990). Comparison of the 15q deletions in Prader-Willi and Angelman syndromes: Specific regions, extent of deletions, parental origin, and clinical consequences. Am. J. Med. Genet. 35: 333-49

Malcolm S, Clayton-Smith J, Nichols M, Robb S, Webb T, Armour JAL, Jeffreys AJ, Pembrey ME (1991). Uniparental disomy in Angelman's syndrome. Lancet 337: 694-7

Malcolm S, Donlon TA (1994). Report and abstracts of the Second International Workshop on Human Chromosome 15 Mapping. England February 18-20, 1994. Cytogenet. Cell Genet. 67: 1-22

Malzac P, Moncla A, Voelckel MA, Livet O, Girardot L, Mattei MG, Mattei JF (1993). Prader-Willi syndrome: diagnostic strategy with a cytogenetic and molecular approach. Neuromuscul. Disord. 3: 493-6

Malzac P, Webber H, Moncla A, Graham JM, Kukolich M, Williams C, Pagon RA, Ramsdell LA, Kishino T, Wagstaff J (1998). Mutation analysis of UBE3A in Angelman syndrome patients. Am. J. Hum. Genet. 62: 1353-60

Maniatis T, Fritsch EF, Sambrook J (1982). Molecular cloning, a laboratory manual. Cold spring Harbour Press, Cold Spring Harbour, New York

Mann JR, Lovell-Badge RH (1984) Inviability of parthenogenones is determined by pronuclei, not egg cytoplasm. Nature 310: 66-7

Maruyama K, Usami M, Aizawa T, Yoshikawa K (1991). A novel brain-specific mRNA encoding nuclear protein (necdin) expressed in neurally differentiated embryonal carcinoma cells. Biochem. Biophys. Res. Commun. 178: 291-6

Mascari MJ, Gottlieb W, Rogan PK, Butler MG, Waller DA, Armour JAL, Jeffreys AJ, Ladda RL, Nicholls RD (1992). The frequency of uniparental disomy in Prader-Willi syndrome. Implications for molecular diagnosis. N. Engl. J. Med. 326: 1599-607

Matsuura T, Sutcliffe JS, Fang P, Galjaard R-J, Jiang Y, Benton CS, Rommens JM, Beaudet AL (1997). *De novo* truncating mutations in E6-AP ubiquitin-protein ligase gene (UBE3A) in Angelman syndrome. Nat. Genet. 15: 74-7

McAllister G, Amara SG, Lerner MR (1988). Tissue-specific expression and cDNA cloning of small nuclear ribonucleoprotein-associated polypeptide.
N. Proc. Natl. Acad. Sci. USA 85: 5296-300

McFadden DE, Kwong LC, Yam IY, Langlois S (1993). Parental origin of triploidy in human fetuses: evidence for genomic imprinting. Hum. Genet. 92: 465-9

McGrath J, Solter D (1984). Completion of mouse embryogenesis requires both the maternal and paternal genomes. Cell 37: 179-83

McLachlan J (1994). Medical Embryology. Addison-Wesley

Meguro M, Mitsuya K, Sui H, Shigenami K, Kugoh H, Nakao M, Oshimura M (1997). Evidence for uniparental, paternal expression of the human $GABA_A$ receptor subunit genes, using microcell-mediated chromosome transfer. Hum. Mol. Genet. 6: 2127-33

Meijers-Heijboer EJ, Sandkuijl LA, Brunner HG, Smeets HJM, Hoogeboom AJM, Deelen WH, van Hemel JO, Nelen MR, Smeets DFCM, Niermeijer MF, Halley DJJ (1992). Linkage analysis with chromosome 15q11-13 markers shows genomic imprinting in familial Angelman syndrome. J. Med. Genet. 29: 853-7

Michaelis RC, Skinner SA, Lethco BA, Simensen RJ, Donlon TA, Tarleton J, Phelan MC (1995). Deletion involving D15S113 in a mother and son without Angelman syndrome: refinement of the Angelman syndrome critical deletion region. Am. J. Med. Genet. 55: 120-6

Miyoshi N, Kuroiwa Y, Kohda T, H. S, Yonekawa H, Kawabe T, Hasegawa H, Barton SC, Surani MA, Kaneko-Ishino T, Ishino F (1998). Identification of the Meg1/Grb10 imprinted gene on mouse proximal chromosome 11, a candidate for the Silver-Russell syndrome gene. Proc. Natl. Acad. Sci. USA 95: 1102-7

Monaco AP, Neve RL, Colletti-Feener C, Bertelson CJ, Kurnit DM, Kunkel LM (1986). Isolation of candidate cDNAs for portions of the Duchenne muscular dystrophy gene. Nature 323: 646-50

Morgan JG, Doganov GM, Robbins SE, Hinton LM, Lovett M (1992). The selective isolation of novel cDNAs encoded by the regions surrounding the human interleukin 4 and 5 genes. Nucleic Acids Res. 20: 5173-9

Morichon-Delvallez N, Mussat P, Dumez Y, Vekemans M (1993). Trisomy 15 in chorionic villi and Prader-Willi syndrome at birth. Prenat. Diagn. 13: 307-8

Morishima A, Grumbach MM, Taylor JH (1962). Asynchronous duplication of human chromosomes and the origin of sex chromatin. Proc. Natl. Acad. Sci. USA 48: 756-63

Mosser J, Douar A-M, Sarde C-O, Kioschis P, Feil R, Moser H, Poustka A-M, Mandel J-L, Aubourg P (1993). Putative X-linked adrenoleukodystrophy gene shares unexpected homology with ABC transporters. Nature 361: 726-30

Moulton T, Crenshaw T, Hao Y, Moosikasuwan J, Lin N, Dembitzer F, Hensle T, Weiss L, McMorrow L, Loew T, Kraus W, Gerald W, Tycko B (1994). Epigenetic lesions at the *H19* locus in Wilm's tumour patients. Nat. Genet. 7: 440-7

Mowery-Rushton PA, Driscoll DJ, Nicholls RD, Locker J, Surti U (1996). DNA methylation patterns in human tissues of uniparental origin using a zinc-finger gene (*ZNF127*) from the Angelman/Prader-Willi region. Am. J. Med. Genet. 61: 140-6

Munroe PB, Mitchison HM, O'Rawe AM, Anderson JW, Boustany RM, Lerner TJ, Taschner PE, de Vos N, Breuning MH, Gardiner RM, Mole SE, (1997). Spectrum of mutations in the Batten disease gene, CLN3. Am. J. Hum. Genet. 61: 310-6

Mutirangura A, Greenberg F, Butler MG, Malcolm S, Nicholls RD, Chakravarti A, Ledbetter DH (1993a). Multiplex PCR of three dinucleotide repeats in the Prader-Willi/Angelman critical region (15q11-q13): molecular diagnosis and mechanism of uniparental disomy. Hum. Mol. Genet. 2: 143-51

Mutirangura A, Jayakumar A, Sutcliffe JS, Nakao M, McKinney J, Buiting K, Horsthemke B, Beaudet AL, Chinault AC, Ledbetter DH (1993b). A complete YAC contig of the Prader-Willi / Angelman chromosome region (15q11-13) and refined localisation of the SNRPN gene. Genomics 18: 546-52

Nakao M, Sutcliffe JS, Durtschi B, Mutirangura A, Ledbetter DH, Beaudet AL (1994). Imprinting analysis of three genes in the Prader-Willi/Angelman region: SNRPN, E6-associated protein, and PAR-2 (D15S225E). Hum. Mol. Genet. 3: 309-15

Nawaz Z, Lonard DM, Smith CL, Lev-Lehman E, Tsai SY, Tsai M-J, O'Malley BW (1999). The Angelman syndrome-associated protein, E6-AP, is a coactivator for the nuclear hormone receptor superfamily. 19: 1182-9

Nelson DL, Ledbetter SA, Corbo L, Victoria MF, Ramirez-Solis R, Webster T, Ledbetter DH, Caskey CT (1989). Alu polymerase chain reaction: A method for rapid isolation of human-specific sequences from complex DNA sources.
Proc. Natl. Acad. Sci. USA 86: 6686-90

Neumann B, Barlow DP (1996). Multiple roles for DNA methylation in gametic imprinting.
Curr. Opin. Genet. Dev. 6: 159-63

Neyroud N, Tesson F, Denjoy I, Leibovici M, Donger C, Barhanin J, Faure S, Gary F, Coumel P, Petit C, Schwartz K, Guicheney P, (1997). A novel mutation in the potassium channel gene KVLQT1 causes the Jervell and Lange-Nielsen cardioauditory syndrome.
Nat. Genet. 15: 186-9

Nicholls RD, Knoll JHM, Butler MG, Karam S, Lalande M (1989a). Genetic imprinting suggested by maternal heterodisomy in non-deletion Prader-Willi syndrome.
Nature 342: 281-5

Nicholls RD, Knoll JH, Glatt K, Hersh JH, Brewster TD, Graham J, J.M., Wurster-Hill D, Wharton R, Latt SA (1989b). Restriction fragment length polymorphisms within proximal 15q and their use in molecular cytogenetics and the Prader-Willi syndrome.
Am. J. Med. Genet. 33: 66-77

Nicholls RD, Pai GS, Gottlieb W, Cantu ES (1992). Paternal uniparental disomy of chromosome 15 in a child with Angelman syndrome. Ann. Neurol. 32: 512-8

Nicholls RD (1993a). Genomic imprinting and uniparental disomy in Angelman and Prader-Willi syndromes: A review. Am. J. Med. Genet. 46: 16-25

Nicholls RD (1993b). Genomic imprinting and candidate genes in the Prader-Willi and Angelman syndromes. Curr. Opin. Genet. Dev. 3: 445-56

Nicholls RD (1993c). New insights reveal complex mechanisms involved in genomic imprinting. Am. J. Hum. Genet. 54: 733-40

Nicholls RD, Gottlieb W, Russell LB, Davda M, Horsthemke B, Rinchik EM (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. USA 90: 2050-4

Nicholls RD, Saitoh S, Horsthemke B (1998). Imprinting in Prader-Willi and Angelman syndromes. TIG 14: 194-200

Ning Y, Lovell M, Taylor L, Pereira-Smith OM (1992). Isolation of monochromosomal hybrids following fusion of human diploid fibroblast-derived microcells with mouse A9 cells. Cytogenet Cell Genet 60: 79-80

Ning Y, Roschke A, Christian SL, Lesser J, Sutcliffe JS, Ledbetter DH (1996). Identification of a novel paternally expressed transcript adjacent to snRPN in the Prader-Willi syndrome critical region. Genome Res. 6: 742-6

Nishiwaki K, Niikawa N, Ishikawa M (1997). Polymorphic and tissue-specific imprinting of the human Wilm's tumour gene, *WT1*. Jpn. J. Hum. Genet. 42: 205-11

Ohlsson R, Nysrom A, Pfeifer-Ohlsson S, Tohonen V, Hedborg F, Schofield P, Flam F, Ekstrom TJ (1993). IGF2 is parentally imprinted during human embryogenesis and in Beckwith-Wiedemann syndrome. Nat. Genet. 4: 94-7

Oldridge M, Lunt PW, Zackai EH, McDonald-McGinn DM, Muenke M, Moloney DM, Twigg SR, Heath JK, Howard TD, Hoganson G, Gagnon DM, Jabs EW, Wilkie AO (1997). Genotype-phenotype correlation for nucleotide substitutions in the IgII-IgIII linker of FGFR2. Hum. Mol. Genet. 6: 137-43

Orita M, Iwahana H, Kanazawa H, Hayashi K, Sekiya T (1989). Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. Proc. Natl. Acad. Sci. USA. 86: 2766-70

Örstavik KH, Tangsrud SE, Kiil R, Hansteen I-L, Steen-Johnsen J, Cassidy SB, Martony A, Anvret M, Tommerup N, Brondum-Nielsen K (1992). Prader-Willi syndrome in a brother and sister without cytogenetic or detectable molecular genetic abnormality at chromosome 15q11q13. Am. J. Med. Genet. 44: 534-8

Osborne LR, Martindale D, Scherer SW, Shi XM, Huizenga J, Heng HHQ, Costa T, Pober B, Lew L, Brinkman J, Rommens J, Koop B, Tsui LC (1996). Identification of genes from a 500-kb region at 7q11.23 that is commonly deleted in Williams syndrome patients. Genomics 36: 328-36

Özçelik T, Leff S, Robinson W, Donlon T, Lalande M, Sanjines E, Schinzel A, Francke U (1992). Small nuclear ribonucleoprotein polypeptide N (SNRPN), an expressed gene in the Prader-Willi syndrome critical region. Nat. Genet. 2: 265-9

Parimoo S, Patanjali SR, Shukla H, Chaplin DD, Weissman SM (1991). cDNA selection: efficient PCR approach for the selection of cDNAs encoded in large chromosomal DNA fragments. Proc. Natl. Acad. Sci. USA 88: 9623-7

Parimoo S, Kolluri R, Weissman SM (1993). cDNA selection from total yeast DNA containing YACs. Nucleic Acids Res. 21: 4422-3

Parimoo S, Patanjali SR, Kolluri R, Xu H, Wei H, Weissman SM (1995). cDNA selection and other approaches in positional cloning. Anal. Biochem. 228: 1-17

Paulsen M, Davies KR, Bowden LM, Villar AJ, Franck O, Fuermann M, Dean WL, Moore TF, Rodrigues N, Davies KE, Hu R-J, Feinberg AP, Maher ER, Reik W, Walter J (1998). Syntenic organisation of the mouse distal chromosome 7 imprinting cluster and the Beckwith-Wiedemann syndrome region in chromosome 11p15.5.
Hum. Mol. Genet. 7: 1149-59

Pearson WR, Lipman DJ (1988). Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA 85: 2444-8

Pembrey M, Fennell SJ, Van Den Berghe J, Fitchett M, Summers D, Butler L, Clarke C, Griffiths M, Thompson E, Super M, Baraitser M (1989). The association of Angelman's syndrome with deletions within 15q11-13. J. Med. Genet. 26: 73-7

Pieler T, Bellefroid E (1994). Perspectives on zinc finger protein function and evolution - an update. Mol. Biol. Rep. 20: 1-8

Plass C, Shibata H, Kalcheva I, Mullins L, Kotelevtseva N, Mullins J, Kato R, Sasaki H, Hirotsune S, Okazaki Y, Held WA, Hayashizaki Y, Chapman VM (1996). Identification of Grf1 on mouse chromosome 9 as an imprinted gene by RLGS-M. Nat. Genet. 14: 106-9

Polaina J, Adam AC (1991). A fast procedure for yeast DNA purification. Nucl. Acids Res. 19: 5443

Prader A, Labhart A, Willi H (1956). Ein syndrom von Adipositas, Kleinwuchs, Kryptorchismus und Oligophrenie nach myatonieartigem Zustand im Neugeborenenalter. Schweiz Med Wochenschr 86: 1260-1

Preece MA, Abu-Amero SN, Ali Z, Abu-Amero KK, Wakeling EL, Stanier P, Moore GE (1999). An analysis of the distribution of hetero- and isodisomic regions of chromosome 7 in five mUPD7 Silver-Russell syndrome probands. J. Med. Genet. 33: 457-60

Priori SG, Schwartz PJ, Napolitano C, Bianchi L, Dennis A, De Fusco M, Brown AM, Casari G, (1998). A recessive variant of the Romano-Ward long-QT syndrome? Circulation 97:2420-5

Purvis-Smith SG, Saville T, Manass S, Yip M-Y, Lam-Po-Tang PRL, Duffy B, Johnston H, Leigh D, McDonald B (1992). Uniparental disomy 15 resulting from "correction" of an intial trisomy 15. Am. J. Hum. Genet. 50: 1348-50

Rachmilewitz J, Elkin M, Looijenga LHJ, Verkerk AJMH, Gonik B, Lustig O, Werner D, de Groot N, Hochberg A (1996). Characterization of the imprinted IPW gene: allelic expression in normal and tumourigenic human tissues. Oncogene 13: 1687-92

Ramsay M, Colman M-A, Stevens G, Zwane E, Kromberg J, Farrall M, Jenkins T (1992). The tyrosine-positive oculocutaneous albinism locus maps to chromosome 15q11.2-q12. Am. J. Hum. Genet. 51: 879-84

Razin A, Cedar H (1994). DNA methylation and genomic imprinting. Cell 77: 473-6

Reed ML, Leff SE (1994). Maternal imprinting of human SNRPN, a gene deleted in Prader-Willi syndrome. Nat. Genet. 6: 163-7

Reik W, Collick A, Norris ML, Barton SC, Surani MA (1987). Genomic imprinting determines methylation of parental alleles in transgenic mice. Nature 328: 248-51

Reik W, Brown KW, Slatter RE, Sartori P, Elliott M, Maher ER (1994). Allelic methylation of H19 and IGF2 in the Beckwith-Wiedemann syndrome. Hum. Mol. Genet. 3: 1297-301

Reik W, Brown KW, Schneid H, Le Bouc Y, Bickmore W, Maher ER (1995). Imprinting mutations in the Beckwith-Wiedemann syndrome suggested by an altered imprinting pattern in the *IGF2-H19* domain. Hum. Mol. Genet. 4: 2379-85

Reik W, Maher ER (1997). Imprinting in clusters: lessons from Beckwith-Wiedemann syndrome. TIG 13: 330-4

Reis A, Dittrich B, Greger V, Buiting K, Lalande M, Gillessen Kaesbach G, Anvret M, Horsthemke B (1994). Imprinting mutations suggested by abnormal DNA methylation patterns in familial Angelman and Prader-Willi syndromes. Am. J. Hum. Genet. 54: 741-7

Reish O, King RA (1995). Angelman syndrome at an older age.
Am. J. Med. Genet. 57: 510-1

Riesewijk AM, Hu L, Schulz U, Tariverdian G, Hoglund P, Kere J, Ropers H-H, Kalscheuer VM (1997). Monoallelic expression of human *PEG1/MEST* is paralleled by parent-specific methylation in fetuses. Genomics 42: 236-44

Rinchik EM, Bultman SJ, Horsthemke B, Lee S-T, Strunk KM, Spritz RA, Avidano KM, Jong MTC, Nicholls RD (1993). A gene for the mouse pink-eyed dilution locus and for human type II oculocutaneous albinism. Nature 361: 72-6

Risch N, Stein Z, Kline J, Warburton D (1986). The relationship between maternal age and chromosome size in autosomal trisomy. Am. J. Hum. Genet. 39: 68-78

Robinson WP, Bottani A, Yagang X, Balakrishman J, Binkert F, Machler M, Prader A, Schinzel A (1991). Molecular, cytogenetic, and clinical investigations of Prader-Willi syndrome patients. Am. J. Hum. Genet. 49: 1219-34

Robinson WP, Bernasconi F, Mutirangura A, Ledbetter DH, Langlois S, Malcolm S, Morris MA, Schinzel AA (1993a). Nondisjunction of chromosome 15: Origin and recombination. Am. J. Hum. Genet. 53: 740-51

Robinson WP, Lorda-Sanchez I, Malcom S, Langlois S, Schuffenhauer S, Knoblauch H, Horsthemke B, Schinzel AS (1993b). Increased parental ages and uniparental disomy 15: A paternal age effect? Eur. J. Hum. Genet. 1: 280-6

Robinson WP, Langlois S, Schuffenhauer S, Horsthemke B, Michaelis RC, Christian S, Ledbetter DH, Schinzel A (1996). Cytogenetic and age-dependent risk factors associated with uniparental disomy 15. Prenat. Diagn. 16: 837-44

Robinson WP, Horsthemke B, Leonard S, Malcolm S, Morton C, Nicholls RD, Ritchie RJ, Rogan P, Schulze R, Schwartz S, Sharpe R, Trent R, Wevrick R, Williamson M, Knoll JHM (1997). Report of the Third International Workshop on Human Chromosome 15 Mapping 1996. October 25-27, in Vancouver, B.C., Canada. Cytogenet. Cell. Genet. 76: 1-13

Rogaev EI, Lukiw WJ, Lavrushina O, Rogaeva EA, St George-Hyslop PH (1994). The upstream promoter of the beta-amyloid precursor protein gene (APP) shows differential patterns of methylation in human brain. Genomics 22: 340-7

Rommens JM, Iannuzzi C, Kerem B, Drumm ML, Melmer G, Dean M, Rozmahel R, Cole JL, Kennedy D, Hidaka N, Zsiga M, Buchwald M, Riordan JR, Tsui L-C, Collins FS (1989). Identification of the cystic fibrosis gene: chromosome walking and jumping. Science 245: 1059-65

Rougelle C, Glatt H, Lalande M (1997). The Angelman syndrome candidate gene, *UBE3A/E6-AP*, is imprinted in brain. Nat. Genet. 17: 14-5

Rougelle C, Cardoso C, Fontes M, Colleaux L, Lalande M (1998). An imprinted antisense RNA overlaps *UBE3A* and a second maternally expressed transcript. Nat. Genet. 19: 15-6

Saarinen K, Swan H, Kainulainen K, Toivonen L, Viitasalo M, Kontula K (1998). Molecular genetics of the long QT syndrome: two novel mutations of the KVLQT1 gene and phenotypic expression of the mutant gene in a large kindred. Hum. Mutat. 11:158-65

Saitoh S, Kubota T, Ohta T, Jinno Y, Niikawa N, Sugimoto T, Wagstaff J, Lalande M (1992). Familial Angelman syndrome caused by imprinted submicroscopic deletion encompassing $GABA_A$ receptor β3-subunit gene. Lancet 339: 366-7

Saitoh S, Harada N, Jinno Y, Hashimoto K, Imaizumi K, Kuroki Y, Fukushima Y, Sugimoto T, Renedo M, Wagstaff J, Lalande M, Mutirangura A, A. K, Ledbetter DH, Niikawa N (1994). Molecular and clinical study of 61 Angelman syndrome patients. Am. J. Med. Genet. 52: 158-63

Saitoh S, Buiting K, Rogan PK, Buxton JL, Driscoll DJ, Arnemann J, Konig R, Malcolm S, Horsthemke B (1996). Minimal definition of the imprinting centre and fixation of a chromosome 15q11-q13 epigenotype by imprinting mutations. Proc. Natl. Acad. Sci. USA 93: 7811-5

Sanford JP, Clark HJ, Chapman VM, Rossant J (1987). Differences in DNA methylation during oogenesis and spermatogenesis and their persistence during early embryogenesis in the mouse. Genes Dev. 1: 1039-46

Sanger F, Nicklen S, Coulson AR (1977). DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5463-7

Sapienza C, Peterson AC, Rossant J, Balling R (1987). Degree of methylation of transgenes is dependent on gamete of origin. Nature 328: 251-4

Scheffner M, Huibregtse JM, Vierstra RD, Howley PM (1993). The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53. Cell 75: 495-505

Scheffner M, Nuber U, Huibregtse JM (1995). Protein ubiquitination involving an E1-E2-E3 enzyme ubiquitin thioester cascade. Nature 373: 81-3

Schlessinger D (1990). Yeast artificial chromosomes: tools for mapping and analysis of complex genomes. TIG. 6: 248-58

Schmauss C, Lerner MR (1990). The closely related small nuclear ribonucleoprotein polypeptides N and B/B' are distinguishable by antibodies as well as by differences in their mRNAs and gene structures. J. Biol. Chem. 265: 10733-9

Schmauss C, Brines ML, Lerner MR (1992). The gene encoding the small nuclear ribonucleoprotein-associated protein N is expressed at high levels in neurons. J. Biol. Chem. 267: 8521-9

Schmid CW, Jelinek WR (1982). The Alu family of dispersed repetitive sequences. Science 216: 1065-70

Schollen E, Pardon E, Heykants L, Renard J, Doggett NA, Callen DF, Cassiman J-J, Mtthijs G (1998). Comparative analysis of the phosphomannomutase genes PMM1, PMM2 and PMM2y: the sequene variation in the processed pseodogene is a reflection of the mutations found in the functional gene. Hum. Mol. Genet. 7: 157-64

Schuler GD, Boguski MS, Stewart EA, Stein LD, Gyapay G, Rice K, White RE, Rodriguez-Tome P, Aggarwal A, Bajorek E, Bentolila S, al. e (1996). A gene map of the human genome. Science 274: 547-62

Schuler GD (1997). Pieces of the puzzle: expressed sequence tags and the catalog of human genes. J. Mol. Med. 75: 694-8

Schulze A, Hansen C, Skakkebaek NE, Brondum-Nielsen K, Ledbetter DH, Tommerup N (1996). Exclusion of SNRPN as a major determinant of Prader-Willi syndrome by a translocation breakpoint. Nat. Genet. 12: 452-4

Schwartz DC, Cantor CR (1984). Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. Cell 37: 67-75

Schwartz AL, Ciechanover A, (1999). The ubiquitin-proteasome pathway and pathogenesis of human diseases. Annu. Rev. Med. 50: 57-74

Sedlacek Z, Konecki DS, Siebenhaar R, Kioschis P, Poustka A (1993). Direct selection of DNA sequences conserved between species. Nucleic Acids Res. 21: 3419-25

Sheehan C, Weiss AS (1990). Yeast artificial chromosomes: rapid extraction for high resolution analysis. Nucl. Acids Res. 18: 2193

Shemer R, Birger Y, Riggs AD, Razin A (1997). Structure of the imprinted mouse Snrpn gene and establishment of its parental-specific methylation pattern. Proc. Natl. Acad. Sci. USA 94: 10267-72

Shibata H, Hirotsune S, Okazaki T, Komatsubara H, Muramatsu M, Takagi N, Ueda T, Shoroishi T, Moriwaki K, Katsuki M, al. e (1994). Genetic mapping and systematic screening of mouse endogenously imprinted loci detected with restriction landmark genome scanning method (RLGS). Mamm. Genome. 5: 797-800

Shibata H, Yoshino K, Muramatsu M, Plass C, Chapman VM, Hayashizaki Y (1995). The use of restriction landmark genomic scanning to scan the mouse genome for endogenous loci with imprinted patterns of methylation. Electrophoresis 16: 210-7

Shiraishi M, Lerman LS, Sekiya T (1995). Preferential isolation of DNA fragments associated with CpG islands. Proc. Natl. Acad. Sci. USA 92: 4229-33

Silbergeld A, Dastot F, Klinger B, Kanety H, Eshet R, Amselem S, Laron Z, (1997). Intronic mutation in the growth hormone (GH) receptor gene from a girl with Laron syndrome and extremely high serum GH binding protein: extended phenotypic study in a very large pedigree. J. Paediatr. Endocrinol. Metab. 10: 265-74

Simmons AD, Goodart SA, Gallardo TD, Overhauser J, Lovett M (1995). Five novel genes from the cri-du-chat critical region isolated by direct selection. Hum. Mol. Genet. 4: 295-302

Sinnett D, Wagstaff J, Glatt K, Woolf E, Kirkness J, Lalande M (1993). High-resolution mapping of the γ-aminobutyric acid receptor subunit β3 and α5 gene cluster on chromosome 15q11-q13, and localisation of breakpoints in two Angelman syndrome patients. Am. J. Hum. Genet. 52: 1216-29

Slatter RE, Elliot M, Welham K, Carrera M, Schofield PN, Barton DE, Maher ER (1994). Mosaic uniparental disomy in Beckwith-Wiedemann syndrome. J. Med. Genet. 31: 749-53

Smeets DFCM, Hamel BCJ, Nelen MR, Smeets JHM, Bollen JHM, Smits APT, Ropers H-H, van Oost BA (1992). Prader-Willi syndrome and Angelman syndrome in cousins from a family with a translocation between chromosomes 6 and 15. N. Engl. J. Med. 326: 807-11

Smith A, Deng Z-M, Beran R, Woodage T, Trent RJ (1994). Familial unbalanced translocation t(8;15)(p23.3;q11) with uniparental disomy in Angelman syndrome. Hum. Genet. 93: 471-3

Solter D (1988). Differential imprinting and expression of maternal and paternal genomes. Annu. Rev. Genet. 22: 127-46

Southern EM (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98: 503-17

Spence JE, Perciaccante RG, Greig GM, Willard HF, Ledbetter DH, Hejtmancik JF, Pollack MS, W.E. OB, Beaudet AL (1988). Uniparental disomy as a mechanism for human genetic disease. Am. J. Hum. Genet. 42: 217-26

Spinner NB, Zackai E, Cheng S-D, Knoll JHM (1995). Supernumerary inv dup (15) in a patient with Angelman syndrome and a deletion of 15q11-q13. Am. J. Med. Genet. 57: 61-5

Stalker HJ, Williams CA (1998). Genetic counseling in Angelman syndrome: The challenges of multipple causes. Am. J. Med. Genet. 77: 54-9

Steenman MJC, Rainier S, Dobry CJ, Grundy P, Horon IL, Feinberg AP (1994). Loss of imprinting of IGF2 is linked to reduced expression and abnormal methylation of *H19* in Wilm's tumour. Nat. Genet. 7: 433-9

Stöger R, Kubicka P, Liu C-G, Kafri T, Razin A, Cedar H, Barlow DP (1993). Maternal-specific methylation of the imprinted mouse Igf2r locus identifies the expressed locus as carrying the imprinting signal. Cell 73: 61-71

Strathmann M, Hamilton BA, Mayeda CA, Simon MI, Meyerowitz EM, Palazzolo MJ (1991). Transposon-facilitated sequencing. Proc. Natl. Acad. Sci. USA 88: 1247-50

Strauss EC, Kobori JA, Siu G, Hood LE (1986). Specific-primer-directed DNA sequencing. Anal. Biochem. 154: 353-60

Sun Y, Nicholls RD, Butler MG, Saitoh S, Hainline BE, Palmer CG (1996). Breakage in the SNRPN locus in a balanced 46, XY,t(15;19) Prader-Willi syndrome patient. Hum. Mol. Genet. 5: 517-24

Surani MAH, Barton SC, Norris ML (1984). Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. Nature 308: 548-50

Sutcliffe JS, Nakao M, Christian S, Orstavik KH, Tommerup N, Ledbetter DH, Beaudet AL (1994). Deletions of a differentially methylated CpG island at the SNRPN gene define a putative imprinting control region. Nat. Genet. 8: 52-8

Sutcliffe JS, Galjaard R-J, Jiang Y-H, Kubota T, Fang P, Matsuura T, Ledbetter DH, Cattanach B, Beaudet AL (1996). The Angelman syndrome critical region is narrowed to ~200 kb and includes the *UBE3A* gene. Am. J. Hum. Genet. 59: Supplement; 749

Sutcliffe JS, Jiang Y-H, Galjaard R-J, Matsuura T, Fang P, Kubota T, Christian SL, Bressler J, Cattanach B, Ledbetter DH, Beaudet AL (1997). The E6-AP ubiquitin-protein ligase (UBE3A) gene is localised within a narrowed Angelman syndrome critical region. Genome Res 7: 368-77

Swain JL, Stewart TA, Leder P (1987). Parental legacy determines methylation and expression of an autosomal transgene: a molecular mechanism for parental imprinting. Cell 50: 719-27

Szabo PE, Mann JR (1995). Biallelic expression of imprinted genes in the mouse germ line: implications for erasure, establishment, and mechanisms of genomic imprinting. Genes and Development 9: 1857-68

Talis AL, Huibregtse JM, Howley PM (1998). The role of E6AP in the regulation of p53 protein levels in human papillomavirus (HPV)-positive and HPV-negative cells. J. Biol. Chem. 273: 6439-45

Tantravahi U, Nicholls RD, Stroh H, Ringer S, Neve RL, Kaplan L, Wharton R, Wurster-Hill D, Graham J, J.M., Cantu ES, J.L. F, Kousseff BG, Latt SA (1989). Quantitative calibration and use of DNA probes for investigating chromosome abnormalities in the Prader-Willi syndrome. Am. J. Med. Genet. 33: 78-87

Tassone F, Xu H, Burkin H, Weissman S, Gardiner K (1995). cDNA selection from 10 Mb of chromosome 21 DNA: efficiencey in transcriptional mapping and reflections of genome organization. Hum. Mol. Genet. 4: 1509-18

Teshima I, Chadwick D, Chitayat D, Kobayashi P, Ray P, Shuman C, Siegel-Bartelt J, Strasberg P, Weksberg R (1996). FISH detection of chromosome 15 deletions in Prader-Willi and Angelman syndromes. Am. J. Med. Genet. 62: 217-23

Tilghman SM (1999). The sins of the fathers and mothers: genomic imprinting in mammalian development. Cell 96: 185-93

Toth-Fejel S, Magenis RE, Leff S, Brown MG, Comegys B, Lawce H, Berry T, Kesner D, Webb MJ, Olson S (1995). Prenatal diagnosis of chromosome 15 abnormalities in the Prader-Willi/Angelman syndrome region by traditional and molecular cytogenetics.
Am. J. Med. Genet. 55: 444-52

Toth-Fejel S, Olson S, Gunter K, Quan F, Wolford J, Popovich BW, Magenis RE (1996). The impact of imprinting: Prader-Willi syndrome resulting from chromosome translocation, recombination, and nondisjunction. Am. J. Hum. Genet. 58: 1008-16

Trent RJ, Sheffield LJ, Deng Z-M, Kim WS, Nassif NT, Ryce C, Woods CG, Michaelis RC, Tarleton J, Smith A (1997). The elusive Angelman syndrome critical region.
J. Med. Genet. 34: 714-8

Tsai T-F, Raas-Rothschild A, Ben-Neriah Z, Beaudet AL (1998). Prenatal diagnosis and carrier detection for a point mutation in *UBE3A* causing Angelman syndrome.
Am. J. Hum. Genet. 63: 1561-3

Tycko B (1997). DNA methylation in genomic imprinting. Mutat. Res. 386: 131-40

Tyson J, Tranebjaerg L, Bellman S, Wren C, Taylor JF, Bathen J, Aslaksen B, Sorland SJ, Lund O, Malcolm S, Pembrey M, Bhattacharya S, Bitner-Glindzicz M (1997). IsK and KvLQT1: mutation in either of the two subunits of the slow component of the delayed rectifier potassium channel can cause Jervell and Lange-Nielsen syndrome.
Hum. Mol. Genet. 6: 2179-85

Uberbacher EC, Mural RJ (1991). Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach. Proc. Natl. Acad. Sci. USA 88: 11261-5

Uetsuki T, Takagi K, Sugiura H, Yoshikawa K (1996). Structure and expression of the mouse necdin gene. Idenification of a postmitotic neuron-restrictive core promoter.
J. Biol. Chem. 271: 918-24

Valdes JM, Tagle DA, Collins FS (1994). Island rescue PCR: A rapid and efficient method for isolating transcribed sequences from yeast artificial chromosomes and cosmids. Proc. Natl. Acad. Sci. USA 91: 5377-81

van den Ouweland AMW, van der Est M, Wesby-van Swaay E, Tijmensen TSLN, Los FJ, Van Hemel JO, Hennekam RCM, Meijers-Hoejboer HJ, Niermeijer MF, Halley DJJ (1995). DNA diagnosis of Prader-Willi and Angelman syndromes with the probe PW71 (D15S63). Hum. Genet. 95: 562-7

Vaughan J, Ali Z, Bower S, Bennett P, Chard T, Moore G (1994). Human maternal uniparental disomy for chromosome 16 and fetal development. Prenat. Diagn. 14: 751-6

Viville S, Surani MA (1995). Towards unravelling the Igf2/H19 imprinted domain. Bioessays 17: 835-8

Voss R, Ben-Simon E, Avital A, Godfrey S, Zlotogora J, Dagan J, Tikochinski Y, Hillel J (1989). Isodisomy of chromosome 7 in a patient with cystic fibrosis: could uniparental disomy be common in humans? Am. J. Hum. Genet. 45: 373-80

Vu TH, Hoffman AR (1994). Promoter-specific imprinting of the human insulin-like growth factor-II gene. Nature 371: 714-7

Vu TH, Hoffmann AR (1997). Imprinting of the Angelman syndrome gene, *UBE3A*, is restricted to brain. Nat. Genet. 17: 12-3

Vulpe C, Levinson B, Whitney S, Packman S, Gitschier J (1993). Isolation of a candidate gene for Menkes disease and evidence that it encodes a copper-transporting ATPase. Nat. Genet. 3: 7-13

Wachtler F, Musil R (1980). On the structure and polymorphism of the human chromosome number 15. Hum. Genet. 56: 115-8

Wagstaff J, Knoll JHM, Fleming J, Kirkness EF, Martin-Gallardo A, Greenberg F, Graham Jr. JM, Menninger J, Ward D, Venter JC, Lalande M (1991a). Localization of the Gene encoding the $GABA_A$ receptor β3 subunit to the Angelman/Prader-Willi region of human chromosome 15. Am. J. Hum. Genet. 49: 330-7

Wagstaff J, Chaillet JR, Lalande M (1991b). The $GABA_A$ receptor β3 subunit gene: characterization of a human cDNA from chromosome 15q11q13 and mapping to a region of conserved synteny on mouse chromosome 7. Genomics 11: 1071-8

Wagstaff J, Knoll JHM, Glatt KA, Shugart YY, Sommer A, Lalande M (1992). Maternal but not paternal transmission of 15q11-13-linked nondeletion Angelman syndrome leads to phenotypic expression. Nat. Genet. 1: 291-4

Wagstaff J, Shugart YY, Lalande M (1993). Linkage analysis in familial Angelman syndrome. Am. J. Hum. Genet. 53: 105-12

Wang Q, Curran ME, Splawski I, Burn TC, Millholland JM, VanRaay TJ, Shen J, Timothy KW, Vincent GM, de Jager T, Scwartz PJ, Towbin JA, Moss AJ, Atkinson DL, Landes GM, Connors TD, Keating MT, (1996). Positional cloning of a novel potassium channel gene: *KVLQT1* mutations cause cardiac arrhythmias. Nat. Genet. 12: 17-23

Ward CL, Omura S, Kopito RR, (1995). Degradation of CFTR by the ubiquitin-proteasome pathway. Cell 83: 121-7

Watrin F, Roeckel N, Lacroix L, Mignon C, Mattei M-G, Disteche C, Muscatelli F (1997). The mouse Necdin gene is expressed from the paternal allele only and lies in the 7C region of the mouse chromosome 7, a region of conserved synteny to the human Prader-Willi syndrome region. Eur. J. Hum. Genet. 5: 324-32

Webb T, Clayton-Smith J, Cheng X-J, Knoll JHM, Lalande M, Pembrey ME, Malcolm S (1992). Angelman syndrome with a chromosomal inversion 15 inv(p11q13) accompanied by a deletion in 15q11q13. J. Med. Genet. 29: 921-4

Webb T, Malcolm S, Pembrey ME, Clayton-Smith J (1993). Inheritance of parental chromosomes 15 in Angelman syndrome - implications for the family. Genet. Couns. 4: 1-6

Webb T (1994). Inv dup(15) supernumerary marker chromosomes.
J. Med. Genet. 31: 585-94

Webb T, Clarke D, Hardy CA, Kilpatrick MW, Corbett J, Dahlitz M (1995). A clinical, cytogenetic, and molecular study of 40 adults with the Prader-Willi syndrome.
J. Med. Genet. 32: 181-5

Wei J, Hodes ME, Wang Y, Feng Y, Ghetti B, Dlouhy SR (1996). Direct cDNA selection with DNA microdissected from mouse chromosome 16: isolation of novel clones and construction of a partial transcription map of the C3-C4 region. Genome Res. 6: 678-87

Wevrick R, Kerns JA, Francke U (1994). Identification of a novel paternally expressed gene in the Prader-Willi syndrome region. Hum. Mol. Gen. 3: 1877-82

Wevrick R, Francke U (1997). An imprinted mouse transcript homologous to the human imprinted in Prader-Willi syndrome (*IPW*) gene. Hum. Mol. Genet. 6: 325-32

White L, Knoll JHM (1995). Angelman syndrome: Validation of molecular cytogenetic analysis of chromosome 15q11-q13 for deletion detection. Am. J. Med. Genet. 56: 101-5

Williams CA, Gray BA, Hendrickson JE, Stone JW, Cantu ES (1989). Incidence of 15q deletions in the Angelman syndrome: A survey of twelve affected persons.
Am. J. Med. Genet. 32: 339-45

Williams CA, Zori RT, Gray BA, Cantu ES, Ostrer H (1990). Maternal origin of 15q11-13 deletions in Angelman syndrome suggests a role for genomic imprinting.
Am. J. Med. Genet. 35: 350-3

Williams CA, Angelman H, Clayton-Smith J, Driscoll DJ, Hendrickson JE, Knoll JHM, Magenis EE, Schinzel A, Wagstaff J, Whidden EM, Zori RT (1995). Angelman syndrome: Consensus for diagnostic criteria. Angelman Syndrome Foundation.
Am. J. Med. Genet. 56: 237-8

Woodage T, Deng Z-M, Prasad M, Smart R, Lindeman R, Christian SL, Ledbetter DH, Robson L, Smith A, Trent RJ (1994a). A variety of genetic mecanisms are associated with the Prader-Willi syndrome. Am. J. Med. Genet. 54: 219-26

Woodage T, Lindeman R, Deng ZM, Fimmel A, Smith A, Trent RJ (1994b). Physical mapping studies at D15S10: Implications for candidate gene identification in the Angelman syndrome / Prader-Willi syndrome chromosome region of 15q11-q13. Genomics 19: 170-2

Wutz A, Smrzka OW, Schweifer N, Schellanders K, Wagner EF, Barlow DP (1997). Imprinted expression of the *Igf2r* gene depends on an intronic CpG island.
Nature 389: 745-9

Yamada KA, Volpe JJ (1990). Angelman's syndrome in infancy.
Dev. Med. Child Neurol. 32: 1005-11

Yamamoto Y, Huibregtse JM, Howley PM (1997). The human *E6-AP* gene (*UBE3A*) encodes three potential protein isoforms generated by differential splicing. Genomics 41: 263-6

Yang T, Adamson TE, Resnick JL, Leff S, Wevrick R, Francke U, Jenkins NA, Copeland NG, Brannan CI (1998). A mouse model for Prader-Willi syndrome imprinting-centre mutations. Nat. Genet. 19: 25-31

Zhang Y, Tycko B (1992). Monoallelic expression of the human *H19* gene.
Nat. Genet. 1: 40-4

Zori RT, Hendrickson J, S. W, Whidden EM, Gray B, Williams CA (1992). Angelman syndrome: Clinical profile. J. Child Neurol. 7: 270-80

Appendix

Selected sequences isolated by ' direct cDNA selection' used as queries for BLAST searches.

Sequence 9 (577bp)

```
  1  TGGGGCTATA CAATGTAATC CTTAGGAGTA ATAGTTTCAT TCATTTCCAG
 51  GTCAGCTTAC TGTATGATTA AGTAACACAA GGCACAGTAG CCATCTTTTT
101  CATTATGTTG CAACACTGAT CACGTGCCTC GATAAAATGG CTGATTCAAC
151  AAGATGATGG CAACACGAAG GGGAGACTTT GGATTGTCTA TTTAAAATCT
201  AGGTAATAAG TAAGTAATTA ATAAAAACTC TATCTTAAGT GCACTTTCAC
251  ATGCTTTTTG TTTATAATAA ACAAACAACA AACTTCCTAA CTTTGTTGCA
301  ATAGGCTTGA CTACCATTTC ATTTGGCCAA ATGCACTTTC CCCAGTAAAC
351  TTAAAACAAC NACGAGAACA ACAAGAACAA AAATCCCTGT CCTTTCATAT
401  ACTAANAAAA AAGATTGGCT ACTGAAACAG TTCATTGCCA CACACATGAA
451  AACGACATCT GTGGCTGAAT TGTTTCTATT TTAAACTTGT TTGTGCTGTT
501  ACCAAATTCC GAAGGCGGCA TTTAAAACAT TCATTTCTCC CTCCTTCCCC
551  CTTCCCCTNT GATTAATCCC CAAAATT
```

Sequence 10 (274bp)

```
  1  TGAGTCAGAG GATCAGGACC TGTAAGGGCA GTATAAAGAA GGAAGGAATG
 51  ATGGGAGAGG TTTAGGGGGA AAGAAGTGAC AACCCGTGAC AATTGAGGAA
101  TGAGAGATTT TGATGATTGG GGTGAAGGTT ACATTTCTTA AAAAGAAACA
151  AAGAATGGTC GGTCGATAGA GATGCAAGAT GAAGAATTTT GTTTTTAGGA
201  CTACTGACTA TGAGGTAACA AAGAAATCCC AGAGACAGAG GTCTGAGCAA
251  AAGATATGGG ATTGGGCTCG TGCC
```

Sequence 16 (298bp)

```
  1  TTTACACATA TAGTCATTTC ATCTGCAAAC AAAGGCAGTT GTATTTTTTT
 51  GTTCCCCGTT AGTATACCCT TTATTTCCTT TTCTTGTCTT AGCTAGGACT
101  TCCAGTATGG AGTGGTGAAG TGGGGATATT CTTTTCTTGT TCCTGATCTT
151  AGTAAGAAAG TTTCTAGTTT CTCATCATTA AGGATGGTGT TAGCTGTATG
201  TTTTTTGTAA ACGTTCTTTA TCAAGTTGAA GGTGTTCCCT CTATTCCTAG
251  TTTACTAGTT TCTATCATGA GTGGATGGTC GAGGGGGGCT CGGTACCC
```

Sequence 21 (401bp)

```
  1  GGCACGAAAG GCAGTAAAAC TGACCCCTAG TTCTGCCTAT CACTATTTAA
 51  AATTACTTTA AGTGTTGAGG ACTGTCCCCA CAGCAGATAT CTAAAAGTAG
101  GCCTGAAAGA ACTAGAAAGT ATTTCATGCA TCCAGCAGTT ACAAATGCCT
151  AAAACTGTAC ATGAAAAAAC AAAACAAAGA AATGAACTGA GAGCAAACGT
201  TTACTGTTTT CTTATAATCT TCAATCCAGG ATTTAAGAAA TTATTTAATA
251  TCAGTTGAGA TACGTGACCA ACCCATTCTC CAGCTGAGAT CCATCTATTT
301  TTAAATTCAC TCCCCCAGCC TTTCATTCAG CCTAAAAGAA GGAATAAGAA
351  TGGGAAGGTG GGAAGGACTT ATCCCATACT GAGTTAAATA AGACCACCAG
401  A
```

Sequence 26 (283bp)

```
  1  GAGTAAGGAC ATGAGTAAGG TTATGTCTCA CCTGTTTTTA AGGAAATGTG
 51  GATATAAGAT GGGTTCTAGT CCATTAAAAG GTGGTAATTT ATACTAAGTC
101  TTACTGTGAG AGACCATAAA CTGCTTTAGT ATTCAGTGTA TTTTTCTTAA
151  TTGAAATATT TTACTTATGA CTTAGTAGAT ACTAAGACTT AACCCTTGAG
201  TTTCTATTCT AATAAAGGAC TACTAATGAA CAATTTTGAG GTTAGAACTC
251  TACTCCATTG TTTTTTGCCG AAATGATTTA CCT
```

Sequence 39 (221bp)

```
   1  CCTTCTTTAC TCCTCTATCC CACACATATA TTTGAACTTT TTTTTGGTTT
  51  TTATTACTGA CCTGGTCCAT TTGGAAGTTA TTCCCAAAAG TTTTTCCTCG
 101  GTACAAGGTT TGTTTCTGGT GGGGGTTTGG TATTCGGGTG TTTTATGGGC
 151  CCAAAGCTTA CTACATCATG ATCCCTTTGT AATCATCCGC AAATTAAATC
 201  CTGTAGAACT TTTCTCGTGC C
```

Sequence 50 (624bp)

```
   1  CGAGGAAAAT GATGTGAATC TTCCAATTTT ATCTATTTGA TTCAAATGGT
  51  ATGTTTTTTC ACATTTCTCA TAATTTCCAC GTCAACTCCC ATGATGATCA
 101  TGCAGGAAAG CAGCATGATT GCACTTCTTG TAGTGCATAC CATCACTAAA
 151  ACATAGTTAC TGTCTTTCTG TGGTCTTAAG AAAAATTCAG AAACAGCATA
 201  CCAATGTTCA TATAAGATAA ATAAGCTTAT GTGGTACAGT AAGTGTTCTA
 251  TTCTGAAAAA TACCTTGATT CGTATCCTAA TGTCTGAAGT TATTGGTAAT
 301  TCTCCAAACT TCCCAATTCT GCTTTCATCT AAATTCATCT TATCTATCTA
 351  TTGTTAAATA ATCTATCCAG TCCCATTTTT ATCTATCATC CTATATTCTG
 401  TCTTTATCAA TCATTCGTCG GTAATCAATT TGTGCCGTTT GGCATAAGAG
 451  ACAGAAGGGC ATAGCAGTGG AAACTCAAAG GAGTTACCTC CATGTGAATG
 501  GACAGTGGTG ATGGTGATCT GTGAGGTGGT GATGTATAGA AAGAAAAACT
 551  AGAAGAGAGC ATGCCCTAAC TCAGGGACCT TGTCCAGCAG GCACTCTGGC
 601  CTTGTTATAC TGTGCTTGAA TATG
```

Sequence 71 (338bp)

```
   1  ATGTCAAGGG TACTATTCAA ACAGGTTTTC TGTGGAGACT TTTAAGTGGT
  51  GGGTTTAAAG CAAGCAGGCA AGAGTTCCCT GGGGTCACAC TGTGACTGGG
 101  AGGTGGACAC TGGTTCCATG AGGGGTGTGG GGTTCCAGGG TCTTGTCAAG
 151  TAAATTGCAT CAAGCTGTCC CATGATGAAG TGGTCCAAAG GAGGCAGTTG
 201  GATAAGGCAC CTATCTGGAC CAATCACACT TGCAGTGCAT CGTTTTGATC
 251  TTTCCTGGAA AAAAGAAGCT GGTGCCAATG ACAAAAACAC TACTTCTTTG
 301  GTGTGGAGGT GAAAATTCCA AGCGGCAAAA TGTTGGCC
```

Sequence 79 (325bp)

```
   1  AATTTATTAT ATAAAACGCC CTCAGAGTAT TTAATTTNTC CTCACTTTAA
  51  TTACACATTA AGAAGCACAG TGGATGAGAA GCCTTTAAGA TGACAACAGT
 101  TGCACGAAGG TCCCTTTCAT CAAGGTAGCG TATGTACCCT AACAGTGTTT
 151  TAAAGGGTGG CCCAGAAAAA CCCCATGTTA CCTTATCACA ATATGGAAAG
 201  CATTGTTTTN TTTTTCCACT AAATTAAATT ATGGTGAAAA GTGCCACAGT
 251  TTTATTTAGC ATTATGGTAC ATAACAAACA GTTCTGTTTC AATTATGAAA
 301  AAAATTAATT AAAATAATCC TGAAA
```

Sequence 83 (421bp)

```
   1  AGAGTTGCGG CCTTGTACCC ACCTCTCAAC CACACAGATC TGTTTTCCTC
  51  CGTGGGACCT GAGAACCAGA ACAAGGACTC GAGAACAAAT TGGGCAACCA
 101  GGAGGGGGGG AACCAGCTAA ATTTTGTTAG GCCACATGGA AATTTCACTT
 151  AGTCCTGCAG ACAATGCATG TGTAACCAAC CTTCACTGAC AGTCTTTCAA
 201  ACAAGAAATT TCCTATACAT GGGAGCCACC ATTTTCTTGT GACCCCCATC
 251  ACCCATAAGA GACCACATTC TAAGCCTGAA CAGACGCCTT GATATAACCC
 301  ATGTCAACAA ACGGACGTTG GCCTTTTCTG TNTAGCCCCT GTTGGGGTAA
 351  TGTGCTGGAC CTCAGTTCGA CGAGGANGAC AGTATAGAGT TTTCANTCAT
 401  TTTGTTCAGN TTTCCAAGGA A
```

Sequence 86 (419bp)

```
   1  GGTGGGGGTG AAGGTTATAT TTCTTAAAAA GAANCAAAGA ATGGTCGGTC
  51  GATAGAGATG CAAGATGAAG AATTTTGTTT TTAGGACTAC TGACTATGAG
 101  GTAACAAAAA AATCCCAGAG ACAGAGGTCT GAGCAAAAGA TATGGGATTG
 151  GGGAAAAATC TTTGGGAGTG ACTGAGGTTG ACTACAGGGA ACTGGGCAAG
 201  AACAGGTAAG GACTTTAAGC ACAATTGGAG GAGTAACCAT CTAAAATCTG
 251  GGTAAACTGG GATGGAAAAC CCACTAGCCA GAAGAAGCAT CACCCCAAGC
 301  TAGGGTATTG TCAAGGTTGT AAATCTTACT GATTTCGGCA ATGAGGAGAT
 351  TATCAGGATC CCTTGGAGAC TGTCCTTCCG GTGAAACTAC CATGCATTAG
 401  TAGCCTTTCA CAGTGATAT
```

Sequence 93 (276bp)

```
   1  GGCACGAGAC CAAAGACTAG GTTTCATATG GAAGAGTTTA TTCTTCTACT
  51  TGAGGTGTGA CATGAAGGAA AGTAATGGAA TAATGCTGCC CTCCCATTCT
 101  CAAGGGCCAG GGGGAACTCA AATATTGAAC GGAAGTGATT TTGGAATTGC
 151  ATGGAATTTT ACCTGCCATC CCGAGGGAGG GCTGATCTTG TCATTCGTAC
 201  CTGTTTCATT CCTTGGGGAG CATAACAGAA GGACGCTCTC CTGAGGATCG
 251  GCATTAACCT TGAATTGGGT TTCATA
```

Sequence 104 (366bp)

```
   1  TTCGAGTGTT CTCTCTTTCA TTTCTGATAC TGGTAATTTT TTTTCTTAGC
  51  CTAGCTACAG AGGCTTATCA ATTTTATCGA TCTTTTCAAA GAACCAGCTT
 101  ATAATTTTGT CGATTTTCTC TACTGATTTT CTTTCCAATT TTATTGATTT
 151  CTACTGTAAT TTTGTTATTT CTTATCTTCT GCTTTCTTTG GATTTAATTT
 201  GCTCCTCTTT TGCTAGTTTC CTAAAGTGGA CTGTGAACTT CATAGGAGCT
 251  TCTCAGTTGC CCCTCATCCC TTAGATAGGA CAGGATGGTT ACAGGTAACT
 301  GGTGTTAGGT ATTTCCCTTC CCCTGGGTAG GTTAGGTTCT GATAAAACCC
 351  CAGCATTTCT CGTGCC
```

Sequence 113 (126bp)

```
   1  GGCACGAGCC TAGATGGAAG AGTTAAGGGC TGAAATAGTT AAAATATTTT
  51  GTCAGGAAAG GACACTTCCC TTGCTGGTTG TCTGGATTTA CTATGAAACT
 101  AGAGTAATTG AGACAGTCCA CTTCAA
```

Sequence 119 (190bp)

```
   1  CCTCTTCTCC CCCCACAAGC TTATATGGTA AAATTACCTG TAGTTAAGTG
  51  GGCATGTGAT ATGATGCACA GATTTAATGC AATAATTTAT AAGAATTATC
 101  CAACTCCATT CTCTCAGCTG CTCTGGTCAA AAAACTTTGG CATCATGTTT
 151  TACTCCTCCC TTTATTTCCC ACTTTCATCT GGCTCGTGCC
```

UBE3A nucleotide sequences

UBE3A exon 1 (contained within OP2) GenBank AF009339 Length: 539bp

```
  1  GCCAGTCCTC CCGTCTTGCG CCGCGGCCGC GAGATCCGTG TGTCTCCCAA
 51  GATGGTGGCG CTGGGCTCGG GGTGACTACA GGAGACGACG GGGCCTTTTC
101  CCTTCGCCAG GACCCGACAC ACCAGGCTTC GCTCGCTCGC GCACCCCTCC
151  GCCGCGTAGC CATCCGCCAG CGCGGGCGCC CGCCATCCGC CGCCTACTTA
201  CGCTTCACCT CTGCCGACCC GGCGCGCTCG GCTGCGGGCG GCGGCGCCTC
251  CTTCGGCTCC TCCTCGGAAT AGCTCGCGGC CTGTAGCCCC TGGCAGGAGG
301  GCCCCTCAGC CCCCCGGTGT GGACAGGCAG CGGCGGCTGG CGACGAACGC
351  CGGGATTTCG GCGGCCCCGG CGCTCCCTTT CCCGGCCTCG TTTTCCGGAT
401  AAGGAAGCGC GGGTCCCGCA TGAGCCCCGG CGGTGGCGGC AGCGAAAGAG
451  AACGAGGCGG TGGCGGGCGG AGGCGGCGGG CGAGGGCGAC TACGACCAGT
501  GAGGCGGCCG CCGCAGCCCA GGCGCGGGGG CGACGACAG
```

UBE3A exon 2 GenBank AF009340 length: 133bp

```
  1  AGGGGAGGTA TGGTGTCAAG AAAACATAGT CACCATTATT ACGAAAAGTA
 51  AAATATGGAA GAGATGATCC CTACCATCAA TCAGCTTACA ACTAGAGGCA
101  CTGACAAATG TATACAGATA GCTGTAATGT AAG
```

UBE3A exon 3 (previous exon U1) GenBank X98034 Length: 30bp

```
  1  ACCAGCTCCT CAGAAGTTTG GCGAAATATG
```

UBE3A exon 4 (previous exon U2) GenBank X98035 Length: 120bp

```
  1  AGTTATTAAG CCTACGCTCA GATCAAGGTA GCAGCTAGAC TGGTGTGACA
 51  ACCTGTTTTT AATCAGTGAC TCAAAGCTGT GATCACCCTG ATGTCACCGA
101  ATGGCCACAG CTTGTAAAAG
```

UBE3A exon 5 (previous exon U3) GenBank X98036 Length: 113bp

```
  1  AGAGTTACAG TGGAGGTAAA AGGAGTGGCT TGCAGGATGG AGAAGCTGCA
 51  CCAGTGTTAT TGGAAGTGAG CCACCATTTG AATTTGCTAG CTCATGCTGC
101  AGTATTCAGA TTA
```

***UBE3A* exon 6 (previous exon U4) GenBank X98037 Length: 196bp**

```
  1  TGCCAGCAGG TTTATTTTTT GTTTTGCAAG CCAGCTCTGC CTCCTTACAG
 51  TATGACATCT GATGCTGGAG GGTCGCACTT TCAAAAATGA GTCAGCTGGT
101  ACATGGGGTT ATCATCAATT TTTAGCTCTT CTGTCTGGGA GATACAAGTT
151  TGAAAGCAAT CTTGGGGTAC TTACCCACAA GGCTGGTGGA GACCAG
```

***UBE3A* exon 7 (previous exon 1) GenBank X98021 Length: 42bp**

```
  1  ATCAGGAGAA CCTCAGTCTG ACGACATTGA AGCTAGCCGA AT
```

***UBE3A* exon 8 (previous exon 2) GenBank X98022 Length: 299bp**

```
  1  GAAGCGAGCA GCTGCAAAGC ATCTAATAGA ACGCTACTAC CACCAGTTAA
 51  CTGAGGGCTG TGGAAATGAA GCCTGCACGA ATGAGTTTTG TGCTTCCTGT
101  CCAACTTTTC TTCGTATGGA TAATAATGCA GCAGCTATTA AAGCCCTCGA
151  GCTTTATAAG ATTAATGCAA AACTCTGTGA TCCTCATCCC TCCAAGAAAG
201  GAGCAAGCTC AGCTTACCTT GAGAACTCGA AAGGTGCCCC CAACAACTCC
251  TGCTCTGAGA TAAAAATGAA CAAGAAAGGC GCTAGAATTG ATTTTAAAG
```

***UBE3A* exon 9 (previous exon 3) GenBank X98023 Length: 1.247kb**

```
  1  ATGTGACTTA CTTAACAGAA GAGAAGGTAT ATGAAATTCT TGAATTATGT
 51  AGAGAAAGAG AGGATTATTC CCCTTTAATC CGTGTTATTG GAAGAGTTTT
101  TTCTAGTGCT GAGGCATTGG TACAGAGCTT CCGGAAAGTT AAACAACACA
151  CCAAGGAAGA ACTGAAATCT CTTCAAGCAA AAGATGAAGA CAAAGATGAA
201  GATGAAAAGG AAAAAGCTGC ATGTTCTGCT GCTGCTATGG AAGAAGACTC
251  AGAAGCATCT TCCTCAAGGA TAGGTGATAG CTCACAGGGA GACAACAATT
301  TGCAAAAATT AGGCCCTGAT GATGTGTCTG TGGATATTGA TGCCATTAGA
351  AGGGTCTACA CCAGATTGCT CTCTAATGAA AAAATTGAAA CTGCCTTTCT
401  CAATGCACTT GTATATTTGT CACCTAACGT GGAATGTGAC TTGACGTATC
451  ACAATGTATA CTCTCGAGAT CCTAATTATC TGAATTTGTT CATTATCGGA
501  ATGGAGAATA GAAATCTCCA CAGTCCTGAA TATCTGGAAA TGGCTTTGCC
551  ATTATTTTGC AAAGCGATGA GCAAGCTACC CCTTGCAGCC CAAGGAAAAC
601  TGATCAGACT GTGGTCTAAA TACAATGCAG ACCAGATTCG GAGAATGATG
```

```
 651  GAGACATTTC AGCAACTTAT TACTTATAAA GTCATAAGCA ATGAATTTAA
 701  CAGTCGAAAT CTAGTGAATG ATGATGATGC CATTGTTGCT GCTTCGAAGT
 751  GCTTGAAAAT GGTTTACTAT GCAAATGTAG TGGGAGGGGA AGTGGACACA
 801  AATCACAATG AAGAAGATGA TGAAGAGCCC ATCCCTGAGT CCAGCGAGCT
 851  GACACTTCAG GAACTTTTGG GAGAAGAAAG AAGAAACAAG AAAGGTCCTC
 901  GAGTGGACCC CCTGGAAACT GAACTTGGTG TTAAAACCCT GGATTGTCGA
 951  AAACCACTTA TCCCTTTTGA AGAGTTTATT AATGAACCAC TGAATGAGGT
1001  TCTAGAAATG GATAAAGATT ATACTTTTTT CAAAGTAGAA ACAGAGAACA
1051  AATTCTCTTT TATGACATGT CCCTTTATAT TGAATGCTGT CACAAAGAAT
1101  TTGGGATTAT ATTATGACAA TAGAATTCGC ATGTACAGTG AACGAAGAAT
1151  CACTGTTCTC TACAGCTTAG TTCAAGGACA GCAGTTGAAT CCATATTTGA
1201  GACTCAAAGT TAGACGTGAC CATATCATAG ATGATGCACT TGTCCGG
```

***UBE3A* exon 10 (previous exon 4) GenBank X98024 Length: 145bp**

```
   1  CTAGAGATGA TCGCTATGGA AAATCCTGCA GACTTGAAGA AGCAGTTGTA
  51  TGTGGAATTT GAAGGAGAAC AAGGAGTTGA TGAGGGAGGT GTTTCCAAAG
 101  AATTTTTTCA GCTGGTTGTG GAGGAAATCT TCAATCCAGA TATTG
```

***UBE3A* exon 11 (previous exon 5) GenBank X98025 Length: 206bp**

```
   1  GTATGTTCAC ATACGATGAA TCTACAAAAT TGTTTTGGTT TAATCCATCT
  51  TCTTTTGAAA CTGAGGGTCA GTTTACTCTG ATTGGCATAG TACTGGGTCT
 101  GGCTATTTAC AATAACTGTA TACTGGATGT ACATTTTCCC ATGGTTGTCT
 151  ACAGGAAGCT AATGGGGAAA AAAGGAACTT TTCGTGACTT GGGAGACTCT
 201  CACCCA
```

***UBE3A* exon 12 (previous exon 6) GenBank X98026 Length: 165bp**

```
   1  GTTCTATATC AGAGTTTAAA AGATTTATTG GAGTATGAAG GGAATGTGGA
  51  AGATGACATG ATGATCACTT TCCAGATATC ACAGACAGAT CTTTTTGGTA
 101  ACCCAATGAT GTATGATCTA AAGGAAAATG GTGATAAAAT TCCAATTACA
 151  AATGAAAACA GGAAG
```

***UBE3A* exon 13 (previous exon 7) GenBank X98027 Length: 156bp**

```
  1  GAATTTGTCA ATCTTTATTC TGACTACATT CTCAATAAAT CAGTAGAAAA
 51  ACAGTTCAAG GCTTTTCGGA GAGGTTTTCA TATGGTGACC AATGAATCTC
101  CCTTAAAGTA CTTATTCAGA CCAGAAGAAA TTGAATTGCT TATATGTGGA
151  AGCCGG
```

***UBE3A* exon 14 (previous exon 8) GenBank X98028 Length: 74bp**

```
  1  AATCTAGATT TCCAAGCACT AGAAGAAACT ACAGAATATG ACGGTGGCTA
 51  TACCAGGGAC TCTGTTCTGA TTAG
```

***UBE3A* exon 15 (previous exon 9) GenBank X98029 Length: 144bp**

```
  1  GGAGTTCTGG GAAATCGTTC ATTCATTTAC AGATGAACAG AAAAGACTCT
 51  TCTTGCAGTT TACAACGGGC ACAGACAGAG CACCTGTGGG AGGACTAGGA
101  AAATTAAAGA TGATTATAGC CAAAAATGGC CCAGACACAG AAAG
```

***UBE3A* previous exon 10 GenBank X98030 Length: 155bp**

```
  1  GTTACCTACA TCTCATACTT GCTTTAATGT GCTTTTACTT CCGGAATACT
 51  CAAGCAAAGA AAAACTTAAA GAGAGATTGT TGAAGGCCAT CACGTATGCC
101  AAAGGATTTG GCATGCTGTA AAACAAAACA AAACAAAATA AAACAAAAAA
151  AGGAA
```

***UBE3A* exon 16, 3'UTR GenBank AF009341 Length: 2.009kb**

```
  1  GTTACCTACA TCTCATACTT GCTTTAATGT GCTTTTACTT CCGGAATACT
 51  CAAGCAAAGA AAAACTTAAA GAGAGATTGT TGAAGGCCAT CACGTATGCC
101  AAAGGATTTG GCATGCTGTA AAACAAAACA AAACAAAATA AAACAAAAAA
151  AAGGAAGGAA AAAAAAAGAA AAAATTTAAA AAATTTTAAA AATATAACGA
201  GGGATAAATT TTTGGTGGTG ATAGTGTCCC AGTACAAAAA GGCTGTAAGA
251  TAGTCAACCA CAGTAGTCAC CTATGTCTGT GCCTCCCTTC TTTATTGGGG
301  ACATGTGGGC TGGAACAGCA GATTTCAGCT ACATATATGA ACAAATCCTT
351  TATTATTATT ATAATTATTT TTTTGCGTGA AAGTGTTACA TATTCTTTCA
401  CTTGTATGTA CAGAGAGGTT TTTCTGAATA TTTATTTTAA GGGTTAAATC
451  ACTTTTGCTT GTGTTTATTA CTGCTTGAGG TTGAGCCTTT TGAGTATTTA
```

501 AAAAATATAT ACCAACAGAA CTACTCTCCC AAGGAAAATA TTGCCACCAT

551 TTGTAGACCA CGTAACCTTC AAGTATGTGC TACTTTTTTG TCCCTGTATC

601 TAACTCAAAT CAGGAACTGT ATTTTTTTTA ATGATTTGCT TTTGAAACTT

651 GAAGTCTTGA AAACAGTGTG ATGCAATTAC TGCTGTTCTA GCCCCCAAAG

701 AGTTTTCTGT GCAAAATCTT GAGAATCAAT CAATAAAGAA AGATGGAAGG

751 AAGGGAGAAA TTGGAATGTT TTAACTGCAG CCCTCAGAAC TTTAGTAACA

801 GCACAACAAA TTAAAAACAA AAACAACTCA TGCCACAGTA TGTCGTCTTC

851 ATGTGTCTTG CAATGAACTG TTTCAGTAGC CAATCCTCTT TCTTAGTATA

901 TGAAAGGACA GGGATTTTTG TTCTTGTTGT TCTCGTTGTT GTTTTAAGTT

951 TACTGGGGAA AGTGCATTTG GCCAAATGAA ATGGTAGTCA AGCCTATTGC

1001 AACAAAGTTA GGAAGTTTGT TGTTTGTTTA TTATAAACAA AAAGCATGTG

1051 AAAGTGCACT TAAGATAGAG TTTTTATTAA TTACTTACTT ATTACCTAGA

1101 TTTTAAATAG ACAATCCAAA GTCTCCCCTT CGTGTTGCCA TCATCTTGTT

1151 GAATCAGCCA TTTTATCGAG GCACGTGATC AGTGTTGCAA CATAATGAAA

1201 AAGATGGCTA CTGTGCCTTG TGTTACTTAA TCATACAGTA AGCTGACCTG

1251 GAAATGAATG AAACTATTAC TCCTAAGAAT TACATTGTAT AGCCCCACAG

1301 ATTAAATTTA ATTAATTAAT TCAAAACATG TTAAACGTTA CTTTCATGTA

1351 CTATGGAAAA GTACAAGTAG GTTTACATTA CTGATTTCCA GAAGTAAGTA

1401 GTTTCCCCTT TCCTAGTCTT CTGTGTATGT GATGTTGTTA ATTTCTTTTA

1451 TTGCATTATA AAATAAAAGG ATTATGTATT TTTAACTAAG GTGAGACATT

1501 GATATATCCT TTTGCTACAA GCTATAGCTA ATGTGCTGAG CTTGTGCCTT

1551 GGTGATTGAT TGATTGATTG ACTGATTGTT TTAACTGATT ACTGTAGATC

1601 AACCTGATGA TTTGTTTGTT TGAAATTGGC AGGAAAAATG CAGCTTTCAA

1651 ATCATTGGGG GGAGAAAAAG GATGTCTTTC AGGATTATTT TAATTAATTT

1701 TTTTCATAAT TGAGACAGAA CTGTTTGTTA TGTACCATAA TGCTAAATAA

1751 AACTGTGGCA CTTTTCACCA TAATTTAATT TAGTGGAAAA AGAAGACAAT

1801 GCTTTCCATA TTGTGATAAG GTAACATGGG GTTTTTCTGG GCCAGCCTTT

1851 AGAACACTGT TAGGGTACAT ACGCTACCTT GATGAAAGGG ACCTTCGTGC

1901 AACTGTAGTC ATCTTAAAGG CTTCTCATCC ACTGTGCTTC TTAATGTGTA

1951 ATTAAAGTGA GGAGAAATTA AATACTCTGA GGGCGTTTTA TATAATAAAT

2001 TCGTGAAGA

***UBE3A* sequence for E6-AP protein isoform I (GenBank X98032) Length: 2.559kb**

```
   1  ATGAAGCGAG CAGCTGCAAA GCATCTAATA GAACGCTACT ACCACCAGTT
  51  AACTGAGGGC TGTGGAAATG AAGCCTGCAC GAATGAGTTT TGTGCTTCCT
 101  GTCCAACTTT TCTTCGTATG GATAATAATG CAGCAGCTAT TAAAGCCCTC
 151  GAGCTTTATA AGATTAATGC AAAACTCTGT GATCCTCATC CCTCCAAGAA
 201  AGGAGCAAGC TCAGCTTACC TTGAGAACTC GAAAGGTGCC CCCAACAACT
 251  CCTGCTCTGA GATAAAAATG AACAAGAAAG GCGCTAGAAT TGATTTTAAA
 301  GATGTGACTT ACTTAACAGA AGAGAAGGTA TATGAAATTC TTGAATTATG
 351  TAGAGAAAGA GAGGATTATT CCCCTTTAAT CCGTGTTATT GGAAGAGTTT
 401  TTTCTAGTGC TGAGGCATTG GTACAGAGCT TCCGGAAAGT TAAACAACAC
 451  ACCAAGGAAG AACTGAAATC TCTTCAAGCA AAAGATGAAG ACAAAGATGA
 501  AGATGAAAAG GAAAAAGCTG CATGTTCTGC TGCTGCTATG GAAGAAGACT
 551  CAGAAGCATC TTCCTCAAGG ATAGGTGATA GCTCACAGGG AGACAACAAT
 601  TTGCAAAAAT TAGGCCCTGA TGATGTGTCT GTGGATATTG ATGCCATTAG
 651  AAGGGTCTAC ACCAGATTGC TCTCTAATGA AAAAATTGAA ACTGCCTTTC
 701  TCAATGCACT TGTATATTTG TCACCTAACG TGGAATGTGA CTTGACGTAT
 751  CACAATGTAT ACTCTCGAGA TCCTAATTAT CTGAATTTGT TCATTATCGG
 801  AATGGAGAAT AGAAATCTCC ACAGTCCTGA ATATCTGGAA ATGGCTTTGC
 851  CATTATTTTG CAAAGCGATG AGCAAGCTAC CCCTTGCAGC CCAAGGAAAA
 901  CTGATCAGAC TGTGGTCTAA ATACAATGCA GACCAGATTC GGAGAATGAT
 951  GGAGACATTT CAGCAACTTA TTACTTATAA AGTCATAAGC AATGAATTTA
1001  ACAGTCGAAA TCTAGTGAAT GATGATGATG CCATTGTTGC TGCTTCGAAG
1051  TGCTTGAAAA TGGTTTACTA TGCAAATGTA GTGGGAGGGG AAGTGGACAC
1101  AAATCACAAT GAAGAAGATG ATGAAGAGCC CATCCCTGAG TCCAGCGAGC
1151  TGACACTTCA GGAACTTTTG GGAGAAGAAA GAAGAAACAA GAAAGGTCCT
1201  CGAGTGGACC CCCTGGAAAC TGAACTTGGT GTTAAAACCC TGGATTGTCG
1251  AAAACCACTT ATCCCTTTTG AAGAGTTTAT TAATGAACCA CTGAATGAGG
1301  TTCTAGAAAT GGATAAAGAT TATACTTTTT TCAAAGTAGA AACAGAGAAC
1351  AAATTCTCTT TTATGACATG TCCCTTTATA TTGAATGCTG TCACAAAGAA
1401  TTTGGGATTA TATTATGACA ATAGAATTCG CATGTACAGT GAACGAAGAA
```

1451 TCACTGTTCT CTACAGCTTA GTTCAAGGAC AGCAGTTGAA TCCATATTTG

1501 AGACTCAAAG TTAGACGTGA CCATATCATA GATGATGCAC TTGTCCGGCT

1551 AGAGATGATC GCTATGGAAA ATCCTGCAGA CTTGAAGAAG CAGTTGTATG

1601 TGGAATTTGA AGGAGAACAA GGAGTTGATG AGGGAGGTGT TTCCAAAGAA

1651 TTTTTTCAGC TGGTTGTGGA GGAAATCTTC AATCCAGATA TTGGTATGTT

1701 CACATACGAT GAATCTACAA AATTGTTTTG GTTTAATCCA TCTTCTTTTG

1751 AAACTGAGGG TCAGTTTACT CTGATTGGCA TAGTACTGGG TCTGGCTATT

1801 TACAATAACT GTATACTGGA TGTACATTTT CCCATGGTTG TCTACAGGAA

1851 GCTAATGGGG AAAAAAGGAA CTTTTCGTGA CTTGGGAGAC TCTCACCCAG

1901 TTCTATATCA GAGTTTAAAA GATTTATTGG AGTATGAAGG GAATGTGGAA

1951 GATGACATGA TGATCACTTT CCAGATATCA CAGACAGATC TTTTTGGTAA

2001 CCCAATGATG TATGATCTAA AGGAAAATGG TGATAAAATT CCAATTACAA

2051 ATGAAAACAG GAAGGAATTT GTCAATCTTT ATTCTGACTA CATTCTCAAT

2101 AAATCAGTAG AAAAACAGTT CAAGGCTTTT CGGAGAGGTT TTCATATGGT

2151 GACCAATGAA TCTCCCTTAA AGTACTTATT CAGACCAGAA GAAATTGAAT

2201 TGCTTATATG TGGAAGCCGG AATCTAGATT TCCAAGCACT AGAAGAAACT

2251 ACAGAATATG ACGGTGGCTA TACCAGGGAC TCTGTTCTGA TTAGGGAGTT

2301 CTGGGAAATC GTTCATTCAT TTACAGATGA ACAGAAAAGA CTCTTCTTGC

2351 AGTTTACAAC GGGCACAGAC AGAGCACCTG TGGGAGGACT AGGAAAATTA

2401 AAGATGATTA TAGCCAAAAA TGGCCCAGAC ACAGAAAGGT TACCTACATC

2451 TCATACTTGC TTTAATGTGC TTTTACTTCC GGAATACTCA AGCAAAGAAA

2501 AACTTAAAGA GAGATTGTTG AAGGCCATCA CGTATGCCAA AGGATTTGGC

2551 ATGCTGTAA

E6-AP amino acid sequences

E6-AP isoform I (GenBank X98032)

```
  1  MKRAAAKHLI ERYYHQLTEG CGNEACTNEF CASCPTFLRM DNNAAAIKAL
 51  ELYKINAKLC DPHPSKKGAS SAYLENSKGA PNNSCSEIKM NKKGARIDFK
101  DVTYLTEEKV YEILELCRER EDYSPLIRVI GRVFSSAEAL VQSFRKVKQH
151  TKEELKSLQA KDEDKDEDEK EKAACSAAAM EEDSEASSSR IGDSSQGDNN
201  LQKLGPDDVS VDIDAIRRVY TRLLSNEKIE TAFLNALVYL SPNVECDLTY
251  HNVYSRDPNY LNLFIIGMEN RNLHSPEYLE MALPLFCKAM SKLPLAAQGK
301  LIRLWSKYNA DQIRRMMETF QQLITYKVIS NEFNSRNLVN DDDAIVAASK
351  CLKMVYYANV VGGEVDTNHN EEDDEEPIPE SSELTLQELL GEERRNKKGP
401  RVDPLETELG VKTLDCRKPL IPFEEFINEP LNEVLEMDKD YTFFKVETEN
451  KFSFMTCPFI LNAVTKNLGL YYDNRIRMYS ERRITVLYSL VQGQQLNPYL
501  RLKVRRDHII DDALVRLEMI AMENPADLKK QLYVEFEGEQ GVDEGGVSKE
551  FFQLVVEEIF NPDIGMFTYD ESTKLFWFNP SSFETEGQFT LIGIVLGLAI
601  YNNCILDVHF PMVVYRKLMG KKGTFRDLGD SHPVLYQSLK DLLEYEGNVE
651  DDMMITFQIS QTDLFGNPMM YDLKENGDKI PITNENRKEF VNLYSDYILN
701  KSVEKQFKAF RRGFHMVTNE SPLKYLFRPE EIELLICGSR NLDFQALEET
751  TEYDGGYTRD SVLIREFWEI VHSFTDEQKR LFLQFTTGTD RAPVGGLGKL
801  KMIIAKNGPD TERLPTSHTC FNVLLLPEYS SKEKLKERLL KAITYAKGFG
851  ML*
```