



The Characterisation of Human X-linked  
Polymorphic Markers and Their Use in  
Disease Gene Localisation and  
Identification

Andrew James Donnelly B.Sc.

Thesis submitted for the Degree of

Doctor of Philosophy

May, 1997

The University of Adelaide

(Department of Genetics)

## TABLE OF CONTENTS

	Page
Summary	I
Declaration	IV
List of Publications	V
List of Tables	VI
List of Figures	VIII
Abbreviations and Glossary	XI
Acknowledgements	XV
Chapter 1: Introduction	1
Chapter 2: Literature Review	5
Chapter 3: Materials and Methods	24
Chapter 4: Isolation and Characterisation of Dinucleotide Repeat Markers	68
Chapter 5: A Genetic Map of the X Chromosome	110
Chapter 6: Linkage Mapping of Genes for X-linked Mental Retardation	149
Chapter 7: Investigation of Candidate Genes for X-linked Mental Retardation	199
Chapter 8: X-linked Diseases and the Identification of Mutated Candidate Genes	246
Chapter 9: Other Approaches to Isolate Genes for X-linked Disorders	290
Chapter 10: Conclusion	318
References	321
Appendix: Publications	

Addendum in response to examiner CES:

Figure 6.4. The line connecting II-7 should originate from the line connecting I-2 and I-3 and then bend around the line connecting II-7 to her sibs. This figure differs from that presented in Appendix A4 to maintain confidentiality regarding non-paternity in the published manuscript.

Figure 6.6. The line connecting II-7 to her parents should bend around the line connecting II-7 to her sibs.

Page 176 (6.7.2) & Fig. 6.10. A common haplotype in two disparate regions is possessed by all the affected males and not any of the unaffected males, hence the high lod score in both regions. This observation has been made previously (Robledo *et al.*, 1996b).

Figure 6.10. The haplotype of I-1 could be 1,2,2,1,1,4,1,1,1,1,4,4,1,2,1 and the haplotype of I-2 could be 2,6,3,2,2,3,3,2,3,2,2,1,2,1,2/3,2,1,1,1,1,2,3,1,3,4,2,3,1,3 so that II-2, II-3 and II-12 are no longer double recombinants. Instead II-2, II-3, II-5 and II-12 are now single recombinants.

Page 181 (6.8.2) & Table 6.7. Haplotype analysis failed to resolve which of the two regions is the most likely to harbour the disease gene. Multipoint analysis would not solve the dilemma because the result would be sensitive to accuracy of the genetic background map in each region and sensitive to interference in each region (which is not known). There is too much missing data as key family members refused to take part in this study. Neither the computer nor the investigator can distinguish between shared markers identical by state and identical by descent.

Page 190 (6.9.2) Person II-8 should be II-9 throughout the first paragraph.

Pages 192-194 (6.9.4).

Given that alleles of DXS153 and DXS424 have apparently been co-inherited by the affected males, it is possible but unlikely that each region harbours an XLMR gene. The discovery of an affected male with a different allele (ie a recombinant) will eliminate linkage to this marker/XLMR gene. Investigation of enough males ("recombination events") would presumably lead to the exclusion of any linkage to any part of the X chromosome, if the linkage analysis is done under the assumption that there is only one XLMR gene segregating in the family, when in fact there are 2 genes. Sample mix-up involving IV-7 was excluded by examination of all markers, not just those shown in Fig. 6.12. All markers known to map to these two regions of possible linkage were tested. The argument applied to IV-6 also can be applied to IV-7, which would rule out both regions.

Figure 9.6 (9.5.2). The FISH results of YAC WXD2446 could be explained by a deletion at the breakpoint.

## Summary

At the start of this project a number of groups were beginning to develop chromosome specific microsatellite based genetic linkage maps. In this project AC dinucleotide repeat markers were isolated from a bacteriophage library for application to the genetic localisations of X-linked disease genes, particularly those responsible for non-specific mental retardation (*MRX*). Of eleven novel microsatellite markers isolated and characterised (Chapter 4), seven were genotyped through the CEPH families enabling their incorporation into a PCR-based linkage map of the X chromosome. In addition, six other pre-existing microsatellite markers were also genotyped through the CEPH families, and included in the map (Chapter 5). The map consists of seventy-eight markers, sixty-nine of which are PCR-based with an average heterozygosity of 66%. The map spans 236 centimorgans, from the *XG* locus near Xpter to *DXS52* near Xqter, and the average spacing between the seventy-eight markers is 3.1 cM.

This map was then used to refine the localisation of the disease gene segregating in five families affected with X-linked mental retardation (Chapter 6). Three of these families (*MRX19*, *MRX30* and *MRX31*) had males with non-specific mental retardation and the other two families (W and O'H) had males with additional symptoms. The *MRX19* gene has been localised to Xp22, within a 19 cM region between the loci *DXS1043* and *DXS1052*. The *MRX30* gene has been localised to a 28 cM region between the loci *DXS990* (Xq21.33) and *DXS424* (Xq23). The *MRX31* gene has been localised to a 12 cM region between the loci *DXS1126* (Xp11.23) and *DXS1124* (Xq13). These results enabled carrier risk analyses to be offered to the females of these three families. Linkage analysis of family W excluded the gene from most of the long arm of the X chromosome, but revealed significant lod scores to two large regions mainly confined to the short arm, in which the disease gene may lie. The first region is from Xqter to *DXS1068* (Xp11.4) and the second region lies between *MAOA* (Xp11.4) and *DXS1002* (Xq13.3-q21.1). The linkage analysis of family O'H has shown significant lod scores to two relatively small regions: the disease gene lies either between *DXS1275* (Xq12) and *DXS7113* (Xq12-q13.1) or in Xq23 between *DXS1220* and *DXS8067*.

A large proportion of those families affected with non-specific X-linked mental retardation have had the causative gene localised to regions that span the centromere. The isolation and characterisation of genes, with expression in the brain, near this region

would therefore generate candidate genes for these MRX loci. To this end, YACs were identified using primers from one microsatellite marker (*DXS1125*) which mapped to this region. From these YACs, *Alu*-PCR products were generated to identify cosmids from which an attempt was made to isolate exons. Five putative exons were trapped (Chapter 7). This approach was abandoned, however, because the trapped sequences were not exons, and because of the recent generation of thousands of expressed sequence tags *via* single-pass sequencing of cDNA by other investigators. Instead, two candidate genes (*CR73E* and *DXS6673E*) were screened for mutations, by SSCA, using the DNA from members of several families with X-linked mental retardation whose disease gene localisations overlapped the positions of these two genes (Chapter 7). Two polymorphisms were discovered in the *CR73E* gene that enabled it to be eliminated as a candidate for the *MRX12* family.

The genetic map described in Chapter 5 was also used in the linkage analysis of a family with males affected with spastic paraplegia (Chapter 8). The results of this analysis led to an investigation of a candidate gene: *PLP*. A point mutation, T<sup>707</sup>C, was found in exon 6 of this gene that would lead to a Phe<sup>236</sup> to Ser amino acid substitution. A sporadic case of a boy affected with myotubular myopathy was also investigated. The hypothesis was made that the disease in this boy was X-linked and this led to the screening of the *MTMI* gene for any mutations. A C<sup>775</sup>T mutation was discovered in exon b. This would lead to a Arg<sup>259</sup> to Cys amino acid substitution. A third family was also investigated. This family had males affected with Wiskott-Aldrich syndrome. Mutation screening of the *WAS* gene identified a G<sup>407</sup>A transition in exon 4. This would lead to a Gly<sup>125</sup> to Arg amino acid change. In all three cases, evidence is provided that strongly suggests the mutations found can be held responsible for each disease phenotype, and as a consequence, direct tests are now available to the members of these families. In addition, the significance of genotype in relation to phenotype may prove useful in the prognosis of affected males from other families with these diseases.

An alternative to the approach of finding disease genes as given in Chapter 7, is to investigate the molecular basis of cytogenetic abnormalities. Chapter 9 describes the preliminary physical mapping of two translocation breakpoints and one duplication event. The first translocation occurs in Xp22 and may be responsible for the global

developmental delay experienced by the female carrier. The second translocation occurs in Xq24 in a female suffering from manic depression. This region has also been implicated, by linkage analysis, in the manic depression segregating within a Finnish family. A third female is the mother of a boy who had symptoms of Borjeson-Forsman-Lehmann syndrome. She carries the same duplication as her son. The point at which the duplication begins occurs in Xq26 and this region is included in the localisation of this disease gene from linkage analysis of a family with this syndrome. YACs have been selected from published and unpublished physical maps and used for fluorescence *in situ* hybridisation to progressively narrow the regions within which these two translocations and the point of duplication have occurred. This work provides the basis for the cloning of any genes involved in the manifestation of each of these three X-linked disorders.

### **Declaration**

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being available for loan and photocopying.

Signed:

Date: 23/5/97

## List of Publications

Some of the work presented in this thesis can be found in the following publications. Reprints or the final manuscripts as accepted for publication of all these are given in the appendix.

Donnelly A, Kozman H, Gedeon AK, Yu PWL, Webb S, Lynch M, Sutherland G R, Richards RI, and Mulley JC. 1994. A linkage map of microsatellite markers on the human X chromosome. *Genomics* **20**: 363-370.

Mulley JC, Yu S, Loesch DZ, Hay DA, Donnelly A, Gedeon AK, Carbonell P, Lopez I, Glover G, Gabarron I, Yu PWL, Baker E, Haan EA, Hockey A, Knight SJL, Davies KE, Richards RI, and Sutherland GR. 1995. FRAXE and mental retardation. *J Med Genet* **32**: 162-169.

Donnelly AJ, Choo KHA, Kozman HM, Gedeon AK, Danks DM, and Mulley JC. 1994. Regional localisation of a non-specific X-linked mental retardation gene (*MRX19*) to Xp22. *Am J Med Genet* **51**: 581-585.

Donnelly AJ, Partington MW, Ryan AK, and Mulley JC. 1996. Regional localisation of two non-specific X-linked mental retardation genes (*MRX30* and *MRX31*). *Am J Med Genet* **64**: 113-120.

Gedeon AK, Donnelly AJ, Mulley JC, Kerr B, Turner G. 1996. How many X-linked genes for non-specific mental retardation (MRX) are there? *Am J Med Genet* **64**:158-162.

Donnelly A, Colley A, Crimmins D, and Mulley J. 1996. A novel mutation in exon 6 (F236S) of the proteolipid protein gene is associated with spastic paraplegia. *Human Mutation* **8**: 384-385.

Donnelly A, Haan E, Manson J and Mulley J. 1997. A novel mutation in exon b (R259C) of the *MTM1* gene is associated with myotubular myopathy. (*Human Mutation* accepted for publication).

## List of Tables

- 3.1 Microsatellite markers used in the linkage analyses
- 4.1 Stage to which AC markers have been developed
- 4.2 Development of AC repeat markers isolated from a flow-sorted X *lambda* library
- 4.3 Characteristics of seven new polymorphic AC repeat markers and three non-polymorphic STS loci
- 4.4 Previously characterised X-linked microsatellites (Xpter to Xqter) now genotyped in the CEPH families
- 4.5 Mutations
- 5.1 Distance (cM) between loci in the comprehensive genetic map together with odds against inversion with adjacent loci and approximate cytogenetic localisation
- 5.2 Recombination events within the 8 largest CEPH families
- 5.3 Comparison of methods and results between published maps of the X chromosome
- 5.4 Discrepancies of  $\geq 5$  cM between loci given in Table 5.1 and those in other maps
- 6.1 Two-point analysis between *MRX19* and marker loci
- 6.2 Two-point analysis between *MRX30* and marker loci
- 6.3 Two-point analysis between *MRX31* and marker loci
- 6.4 Two-point analysis between the family W XLMR gene and marker loci
- 6.5 Two-point analysis between the family O'H XLMR gene and marker loci
- 6.6 Population allele frequencies
- 6.7 Two-point analysis between the family O'H XLMR gene and additional marker loci, using population allele frequencies
- 6.8 Regional localisations of the *MRX* loci
- 7.1 Primers used to amplify portions of the *CR73E* gene
- 7.2 Primers used to amplify the exons of the *DXS6673E* gene
- 7.3 YAC clones positive for microsatellite *DXS1125*

- 7.4 Cosmid clones selected using YAC inter-*Alu* sequences
- 7.5 Exon trapped products - size classing of colony PCR results
- 7.6 Exon trapped products - sequencing results
- 7.7 Sequences (5' - 3') of putative exons
- 7.8 Flanking sequences of other inserts showing consensus splice site nucleotides
- 7.9 cDNA PCR results from phage libraries
- 7.10 cDNA PCR results using RNA from infant brain tissue
- 7.11 cDNA PCR results using pooled infant brain mRNA samples
- 7.12 cDNA PCR results using DNase treated pooled mRNA samples
- 7.13 *CR73E* sequence variants
- 8.1 Primers used to amplify the exons of the *PLP* gene
- 8.2 Two-point analysis between the gene for spastic paraplegia and marker loci
- 8.3 Two-point analysis between the gene for spastic paraplegia and marker loci (*AR* to *DXS424*)
- 9.1 Probes used for FISH to define the Xp22 breakpoint
- 9.2 YACs used for FISH to define the Xq24 breakpoint
- 9.3 Probes used for FISH to define the Xq26 point and extent of duplication

## List of Figures

- 1.1 Flow diagram summarising the aims of this project.
- 2.1 A typical CEPH pedigree.
- 4.1 Examples of the results of dideoxy sequencing of microsatellite-containing M13 phage.
- 4.2 Autoradiograph showing the polymorphic nature of marker 26 (locus *DXS1126*).
- 4.3 Sequencing results of markers that can be amplified by PCR.
- 4.4 CEPH pedigree showing the phenomenon of cryptic alleles associated with marker 21 (locus *DXS1153*).
- 4.5 Autoradiograph showing the mutations referred to in Table 4.5.
- 4.6 Phase analysis to determine the allelic origins of the mutations referred to in Table 4.5.
- 5.1 Genetic map of the human X chromosome.
- 5.2 Meiotic breakpoint maps of 8 of the CEPH pedigrees, generated using CHROMPIC.
- 5.3 Number of crossovers *per* X chromosome.
- 6.1 The *MRX19* pedigree.
- 6.2 *MRX19* pedigree with haplotype data.
- 6.3 Placement of *MRX19* into the multipoint background map.
- 6.4 The *MRX30* pedigree.
- 6.5 Placement of *MRX30* into the multipoint background map.
- 6.6 The *MRX30* pedigree with haplotype data.
- 6.7 The *MRX31* pedigree.
- 6.8 The *MRX31* pedigree with haplotype data.
- 6.9 The W family.
- 6.10 Family W with haplotype data.
- 6.11 The O'H family.

- 6.12 Family O'H with haplotype data.
- 6.13 Genetic map showing the regional localisation of the *MRX19* gene.
- 6.14 Map of the regional localisations of the *MRX* loci.
- 7.1 The exon trapping procedure.
- 7.2 Fingerprinting of cosmids.
- 7.3 Diagram showing the trapping of vector (HIV) sequences through the recognition of cryptic splice sites.
- 7.4 Northern blot of total RNA showing hybridisation of control probe, ESD.
- 7.5 The *MRX12* family with *CR73E* locus A and B alleles.
- 8.1 The family affected with spastic paraplegia.
- 8.2 Autoradiograph showing single strand conformational band shift of the *PLP* exon 6 PCR product.
- 8.3 Autoradiograph showing single strand conformational band shift of the *PLP* exon 6 PCR products from members (if DNA available) of the pedigree.
- 8.4 Part of the *PLP* exon 6 DNA sequence from the PCR product of affected male III-23.
- 8.5 Agarose gel showing *PLP* exon 6 PCR products cut with *Mnl* I
- 8.6 Results of the *Mnl*I digest of the *PLP* exon 6 PCR products (where DNA available).
- 8.7 The family with an isolated case of myotubular myopathy.
- 8.8 Autoradiograph showing single strand conformational band shift of the *MTM1* exon b PCR product.
- 8.9 Part of the sequence of *MTM1* exon b from the affected male (IV-1) showing the mutated base.
- 8.10 Procedure used to introduce a *Cfo*I restriction site into the wildtype *MTM1* exon b PCR products.
- 8.11 *MTM1* exon b PCR products cut with *Cfo*I.
- 8.12 Results of the SSCA and *Cfo*I digestion of *MTM1* exon b PCR products.
- 8.13 The family affected with Wiskott-Aldrich syndrome.

- 8.14 Autoradiograph showing single strand conformational band shift of the *WAS* exon 4 PCR product.
- 8.15 Part of the sequence of *WAS* exon 4 from the affected male V-1, showing the mutated base.
- 8.16 Expected restriction fragments from the PCR product of a wildtype *WAS* exon 4 allele, following digestion with *AclI*.
- 8.17 *WAS* exon 4 PCR products cut with *AclI*.
- 8.18 Results of the SSCA and, for IV-4, IV-5 and V-1 only, *AclI* restriction enzyme digestion of *WAS* exon 4 PCR products.
- 9.1 Idiogram showing the translocation possessed by female S. Her karyotype is 46,X,t(X;6)(p22.3;q14).
- 9.2 Metaphase, from case S, showing FISH with probe cU91B4.
- 9.3 Physical mapping of the Xp22.2 breakpoint, showing the relative position of CEPH YACs as determined by FISH.
- 9.4 Idiogram showing the translocation possessed by female M. Her karyotype is 46,X,t(X;12)(q24;q15).
- 9.5 Metaphase, from case M, showing FISH with probe y896c5.
- 9.6 Physical mapping of the Xq24 breakpoint, showing the relative position of YACs as determined by FISH.
- 9.7 Idiogram showing the duplication possessed by female B. Her karyotype is 46,XX,dup(q26-q28).
- 9.8 Metaphase, from case B, showing FISH with probe 4D-8.
- 9.9 Physical mapping of the Xq26 point of duplication, showing whether or not YACs have been duplicated, as determined by FISH.
- 9.10 The regional localisations of the genes responsible for Pettigrew, Gustavson, XPH and Borjeson-Forssman-Lehmann syndromes.

## Abbreviations and Glossary

(AC)<sub>n</sub> - dinucleotide repeat microsatellite of n repeat units in length.

AIC - Aicardi Syndrome.

BFLS - Borjeson-Forssman-Lehmann Syndrome.

BP - bipolar disorder.

cM - centimorgan. The genetic distance within which there is expected to be one crossover every 100 meioses.

CEPH - Centre d'Etude du Polymorphisme Humain (Centre for the Study of Human Polymorphism).

distal - indicates that the position of a gene or marker is closer to the telomere relative to some other gene or marker (cf proximal).

EST - expressed sequence tag. A segment of cDNA whose sequence is known and can therefore be amplified by PCR (cf STS).

FISH - fluorescence *in situ* hybridisation (of a DNA probe to metaphase or interphase chromosomes).

functional hemizyosity - refers to the expression of only one allele of a gene because of the silencing of the alternative allele due to mutation or X inactivation.

GDB - Genome Database.

gDNA - genomic DNA.

hemizygous - refers to the genes or markers of the X (or Y) chromosome of males, ie genes or markers that are present only once in the genome (cf functional hemizyosity).

heterozygous - refers to the presence of two different alleles of a marker or gene locus, one on each chromosome.

HGP - Human Genome Project.

HIV - human immunodeficiency virus (part of the exon trapping vector).

ICRF - Imperial Cancer Research Fund.

LCL - lymphoblastoid cell line. Derived from a transformed eukaryotic cell.

linkage - when the alleles of two different loci are inherited together more often than not.

The degree of linkage is measured by the recombination fraction ( $\theta$ ).

lod score - the decimal logarithm of the likelihood ratio of linkage vs no linkage.

microsatellite - segment of DNA, often polymorphic, composed of a repeated unit consisting of one to several base pairs.

MIM - the MIM number refers to an entry in McKusick's *Mendelian Inheritance in Man*.

MLE - maximum likelihood estimate (of the recombination fraction between two loci).

MRX - non-specific (non-syndromal) X-linked mental retardation.

MTM - myotubular myopathy.

*MTM1* - gene in Xq28 responsible for myotubular myopathy when mutated.

NIH - National Institutes of Health.

ORF - open reading frame.

PAGE - polyacrylamide gel electrophoresis.

PCR - polymerase chain reaction.

pfu - plaque forming units (= number of bacteriophage).

pleiotropic - refers to the heterogeneous phenotypic consequences of a particular genetic mutation.

PLP - proteolipid protein gene.

proximal - indicates that the position of a gene or marker is closer to the centromere relative to some other gene or marker (cf distal).

$\theta$  - recombination fraction.

RACE - rapid amplification of cDNA ends.

RFLP - restriction fragment length polymorphism.

RT - reverse transcriptase.

SBMA - spino-bulbar muscular atrophy; Kennedy's disease.

SCH - somatic cell hybrid. A mouse host cell used to propagate human chromosomes or chromosomal fragments.

SPG - spastic paraplegia.

SSCA (SSCP) - single stranded conformational analysis (polymorphism). A technique used to distinguish different alleles of a given locus.

SSR - simple sequence repeat (microsatellite).

STS - sequence tagged site. A segment of DNA whose sequence is known and can be amplified by PCR.

UTR - untranslated region.

VNTR - variable number of tandem repeats, also known as a "minisatellite".

WAS - Wiskott-Aldrich syndrome.

WCH - Women's and Children's Hospital (North Adelaide).

WUSM - Washington University School of Medicine.

**XLMR - X-linked mental retardation, either syndromal or non-syndromal.**

**XLMTM - X-linked myotubular myopathy.**

**XLT - X-linked thrombocytopenia.**

**XPH - X-linked recessive panhypopituitarism.**

**YAC - yeast artificial chromosome. A recombinant yeast chromosome used for propagating relatively large (>100kb) fragments of human DNA (in yeast cells).**

## **Acknowledgements**

This work was performed in the Department of Cytogenetics and Molecular Genetics of the Women's and Children's Hospital, Adelaide, on a half-time basis from the 1st of September 1992 to December 1994 and on a full-time basis from the 1st of January 1995 to 30th May 1997.

I would like to thank Liz Baker, Helen Eyre, Kathy Friend, Jozef Gecz, Agi Gedeon, Helen Kozman, Eric Kremer, Sharon Lane, Michael Lynch, Hilary Phillips, Melanie Pritchard, Rob Richards, Grant Sutherland, Selena Webb, Scott Whitmore, Sui Yu and in particular my supervisors Rory Hope and John Mulley for their help, advice and encouragement during this project. I also thank the clinicians responsible for characterising the families presented in this thesis and Dr Roger Byard for material.

In addition, I thank those members of the Department of Cytogenetics and Molecular Genetics who made my stay enjoyable and memorable, especially: Joanna Crawford, Georgina Hollway, Ingrid Lensink and Marie Mangelsdorf.

I gratefully acknowledge the financial support provided by the University of Adelaide Scholarship and the WCH Department of Cytogenetics and Molecular Genetics.

# Chapter 1

## Introduction

## TABLE OF CONTENTS

	Page
<b>1.1 Historical review of genetic mapping (1866-1987)</b>	1
<b>1.2 The Human Genome Project (1988-present)</b>	2
<b>1.3 Aims of the project</b>	3



## 1.1 Historical review of genetic mapping (1866-1987)

Linkage analysis determines the most likely order and genetic distance between two or more loci given their pattern of inheritance through families. Mendel's experiments with pea plants in the 19th century led to his formulation of two laws referring to the inheritance of different alleles of the same trait and the inheritance of different traits. These are the law of segregation and the law of independent assortment, respectively. Linkage between, or the co-inheritance of, two or more traits is thus a violation of his second law, and has been exploited to construct genetic maps. Bateson, in 1905, was the first to report linkage between two loci: that of pollen shape and flower colour in *Lathyrus odoratus* (sweet peas) (Smith, 1986). The idea that linear genetic maps could be constructed was developed by Sturtevant in 1913 who went on to order 5 sex-linked genes of *Drosophila* by observing the relative frequency of crossing-over between them (Conneally and Rivas, 1980). The success of this approach was for many years limited to laboratory species that could be easily bred and scored for mutations, such as *Escherichia coli*, *Saccharomyces cerevisiae*, *Drosophila melanogaster*, *Zea mays* and *Mus musculus*.

The history of the development of human linkage analysis has been reviewed by Smith (1986) and Dronamraju (1987). The first linkage between any two human loci was made by Bell and Haldane (1937), who estimated the frequency of recombination between the X-linked genes for haemophilia and colour-blindness. This red-green colour-blindness marker remained the most useful X-linked marker until the late 1950s, because of the relatively high frequency of affected males. In 1962, the far more polymorphic Xg blood group system was discovered (Mann *et al.*, 1962). Since then, many more markers such as antigens, isozymes and restriction enzyme fragment length polymorphisms (RFLPs) have been discovered and by 1979 about 360 of these types of markers had been mapped by linkage analysis to human chromosomes (Shows and McAlpine, 1979). It was Botstein and colleagues (1980) who were the first to propose the construction of a linkage map of the entire human genome through a systematic attempt to clone polymorphic DNA sequences, specifically RFLPs, that were not necessarily associated with any gene. The first such map, using 393 RFLPs, was published by Donis-Keller *et al.* in 1987.

## 1.2 The Human Genome Project

The Human Genome Project (HGP) was established in 1988 under the auspices of the U.S. National Institutes of Health and the U.S. Department of Energy, with the aims of reducing human suffering and the cost of health care and to provide a greater understanding of human biology (Watson, 1990; Collins and Galas, 1993; Guyer and Collins, 1993, 1995; Olson, 1993). To achieve these ends the specific objectives of the current five-year plan (1/10/93-30/9/98) were to:

- 1) produce a genetic map, with a resolution of 2 to 5 cM, of the human genome by 1995;
- 2) produce physical maps of the human and mouse genomes with an STS every 100 kb and 300 kb respectively, by 1998;
- 3) increase the rate and reduce the costs of large scale DNA sequencing;
- 4) develop efficient methods for identifying and mapping genes;
- 5) sequence the genomes of the well characterised organisms: *Escherichia coli*, *Saccharomyces cerevisiae*, *Caenorhabditis elegans* and *Drosophila melanogaster*;
- 6) develop and improve databases and the software for analysing the data;
- 7) address the ethical, legal and social implications arising from the above.

The achievement of producing a genetic index map of the human genome, that is, a map consisting of markers with heterozygosities of at least 70% (so that at least half the people studied for any two markers will be informative for both) and separated by 2-5 cM, has been made possible by:

- 1) the discovery of many widely distributed highly polymorphic (microsatellite) markers;
- 2) technological advances to analyse these markers, such as PCR and, more recently, automated genotyping;
- 3) the establishment of reference families so that all markers could be genotyped on the same set of families and linkage analysis could be carried out using a single database;

4) the development of computer algorithms and statistical methodology to determine the order and distances between each by multipoint methods.

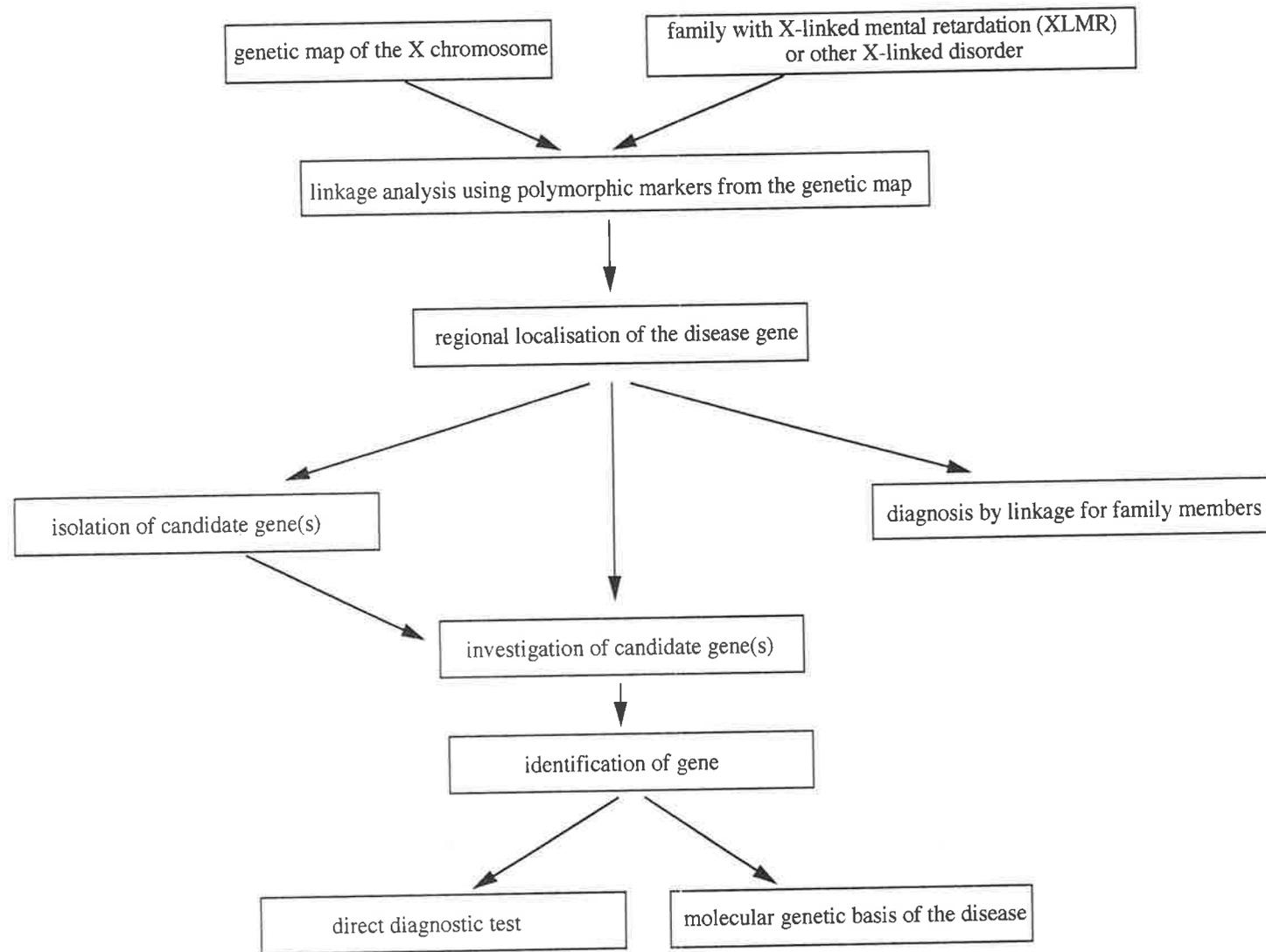
(Each of these four points will be discussed further in the next chapter).

Genetic maps are essential for the positional, as distinct from functional, cloning of disease genes and for genetic counselling. These maps will not only be useful for mapping the genes responsible for monogenic disorders but also those involved in polygenic and multifactorial diseases (Collins, 1992; Sefton and Goodfellow, 1992; Kruglyak and Lander, 1995). They can also be integrated with physical maps (constructed from cloned contiguous DNA segments) since primers used to amplify genomic DNA for genotyping family members can also be used as STS markers on the physical map.

The ultimate aim of the HGP is to determine the DNA sequence of the human genome and reveal all the genes; a task that is expected to be completed early next century (Smith, 1993).

### **1.3 Aims of the project**

In accordance with the first goal of the HGP presented above, the aim of the project presented in this thesis was to isolate microsatellite markers and to construct a high resolution genetic map of the human X chromosome using these and pre-existing microsatellite markers. Such a map would then be put to use in the fine mapping of X-linked diseases. Once a disease gene has been regionally localised, linkage analysis can be used for the determination of carrier status and to provide prenatal diagnosis. In addition, regional localisation is the first step towards the positional cloning of the gene involved *via* the investigation of candidate genes. Identification of the gene responsible may lead to a direct diagnostic test and/or the implementation of preventative or palliative therapy. These steps have been summarised in Figure 1.1.



**Figure 1.1** Flow diagram outlining the aims of this project.

# Chapter 2

## Literature Review

## TABLE OF CONTENTS

	Page
<b>2.1 Hypervariable markers</b>	5
<b>2.2 CEPH reference pedigrees</b>	8
<b>2.3 Linkage mapping</b>	9
<b>2.4 Physical mapping</b>	18
<b>2.5 Morbid anatomy of the human X chromosome</b>	19
<b>2.6 Positional cloning</b>	20
<b>2.7 Identification of disease-causing genes</b>	21
<b>2.8 Transcriptional mapping</b>	22

## 2.1 Hypervariable markers

RFLPs contributed greatly to the construction of the first comprehensive linkage map of the human genome (Donis-Keller *et al.*, 1987). Although RFLPs are common (occurring every 200-300 bp on average; Cooper *et al.*, 1985), they often have low heterozygosities (an average of 40%; cited by Lathrop and Lalouel, 1991) because the molecular basis of their polymorphism is usually just the presence or absence of a single restriction enzyme site. This relatively low level of heterozygosity compared with later markers decreased both the accuracy of inter-marker distance estimates and the support for marker order during linkage map construction, when analysing a similar number of families. Also, the detection of linkage between disease and marker loci is more difficult than would be the case with markers of greater heterozygosity.

Although most RFLPs are biallelic, with a maximum theoretical heterozygosity of 50%, some are multiallelic and have high heterozygosities. The first such hypervariable RFLP was reported by Wyman and White (1980). The cause of the hypervariability of some RFLPs was a difference in the number of copies of a tandemly repetitive sequence. It was Jeffreys *et al.* (1985) who went on to develop probes from these so called “minisatellites” that could detect many unique hypervariable loci. These minisatellites are also called “VNTRs”, the acronym for “variable number of tandem repeats”. The repeat unit can range in length from about 9 to 100 base pairs (bp). Most of these minisatellites map near to the telomeres of chromosomes and thus are not ideal for generating a linkage map consisting of uniformly distributed loci (Donis-Keller *et al.*, 1987; Nakamura *et al.*, 1988; Royle *et al.*, 1988; Rouyer *et al.*, 1990).

There is another group of markers based on a variable number of tandem repeats, designated “microsatellites”, whose repetitive units are smaller than the repeat unit of a minisatellite; about 1 to 6 bp. Microsatellites are also often referred to as “simple sequence repeats” (SSRs) or “short tandem repeats” (STRs). Tautz (1993) has sought to clarify the sometimes confusing nomenclature regarding satellite sequences. Unlike RFLPs, the polymorphic nature of microsatellites is most often assayed *via* the use of the polymerase chain reaction (PCR) technique and not by the use of restriction enzymes.

Although the existence and high abundance of polymorphic microsatellites in eukaryotic genomes had been known since the early 1980s (Hamada *et al.*, 1982; Tautz and Renz,

1984; Gross and Garrard, 1986), it was not until the late 1980s that their variability was recognised, enabling their application to linkage analysis of human populations (Litt and Luty, 1989; Smeets *et al.*, 1989; Tautz, 1989; Weber and May, 1989). By 1991 10% of the approximately 3,000 polymorphic markers characterised had heterozygosity greater than 50% (Weissenbach, 1993). It has been the discovery and application of these highly polymorphic loci that has enabled the rapid development of high resolution linkage maps.

The advantages of using microsatellites over most RFLP-based markers are numerous. Firstly, they are likely to exhibit more polymorphism (Weber, 1990). This aspect is of particular importance when considering the X chromosome because RFLPs, including minisatellites, are relatively less frequent on the X compared to the autosomes (Cooper *et al.*, 1985; Hofker *et al.*, 1986; Dietz-Band *et al.*, 1990; Hendrickx *et al.*, 1993). In addition, as mentioned above, unlike minisatellites, the distribution of microsatellites in the human genome is more uniform (Hamada *et al.*, 1984; Willard *et al.*, 1986; Nakamura *et al.*, 1988; Royle *et al.*, 1988; Barker *et al.*, 1989; Stallings *et al.*, 1991; Hudson *et al.*, 1992; Weissenbach, 1993). There have, however, been reports of regions sparsely populated with microsatellites (for example, subtelomeric regions; Hazan *et al.*, 1992; Weissenbach *et al.*, 1992). The apparent lack of microsatellites in these regions may be artefactual, and could be explained by biases in subcloning and/or inflated recombination rates. Clusters of AC repeats have also been noted and speculated upon (Coleman *et al.*, 1994).

Other advantages in using microsatellites relate to the Nobel prize winning PCR technique, developed by Kary Mullis (Saiki *et al.*, 1985; Mullis and Faloona, 1987). This process produces many copies of any given sequence from genomic DNA to levels that are detectable by a number of methods. PCR is quicker, requires less DNA than the Southern technique (Southern, 1975) required for RFLP analysis and is automatable. Minisatellites up to 10 kilobases (kb) in length can be PCR amplified but such large alleles can only be detected by Southern blotting of the PCR products (Jeffreys and Pena, 1993). PCR can be routinely performed on picogram amounts of DNA and the products are designed to be usually 100-300 bp in length to facilitate resolution of the size differences between alleles and permit successful amplification from degraded samples. Alleles of a given microsatellite locus differ in multiples of the repeat unit, for example 2, 3, or 4 bp for di-, tri-, and tetra-nucleotide repeats respectively. This size difference is readily resolved using

polyacrylamide gels. Methods by which several microsatellite loci can be analysed simultaneously, a technique called “multiplexing”, have also been developed (Church and Kieffer-Higgins, 1988; Hazan *et al.*, 1992).

Another advantage is that the primers flanking microsatellite markers can be made by any laboratory with access to an oligonucleotide synthesiser or can be purchased commercially, rather than having to rely on the investigator who characterised an RFLP to provide the cloned probe. In summary, the utility of PCR-based microsatellite markers far exceeds that of RFLP-based markers.

Most researchers initially concentrated on isolating and characterising poly(AC) dinucleotide repeats as opposed to any other type of microsatellite. In humans the most plentiful microsatellites, in order of decreasing frequency are: poly(A/T), poly(AC/TG), poly(AAAT/TTTA), poly(AG/TC) and poly(AT/TA). Most poly(A) and poly(A<sub>2-3</sub>N) repeats, where N=T, G or C, are found within *Alu* repetitive sequences which makes it difficult to design unique flanking primers (Economou *et al.*, 1989; Weber *et al.*, 1991; Beckmann and Weber, 1992). It has been estimated there are 50,000-100,000 poly(AC)<sub>n</sub> repeat sequences, where “n” represents about 5 to 15 copies (10-30 bp), in the haploid human genome (Hamada and Kakunga, 1982; Tautz and Renz, 1984; Gross and Garrard, 1986). As a consequence, assuming a genome length of  $3 \times 10^9$  bp, there should be a repeat every 30-60 kb, on average. Dinucleotide repeats with less than 13 units are not very polymorphic but the proportion of markers with a repeat length of 40 bp or more (that is,  $n \geq 20$ ) that have heterozygosities  $\geq 0.7$  is 80% (Weber, 1990). It has been calculated that a genetic map with a resolution of 0.3-0.5 cM could be achieved with about 12,000 microsatellite markers, although a great many more will have to be isolated to attain this degree of resolution (Weber, 1990).

Microsatellites not only differ in the length and number of their repeat units but also in their structure (Weber, 1990). The most common type of repeat consists of an uninterrupted stretch of repeat units, for example ACACACACACAC. Such a microsatellite has been classified as a “perfect” repeat. “Imperfect” repeats have interruptions in the run of repetitive units. The least frequent, “compound” repeats are conglomerations of two or more types of repeat. Generally, the heterozygosity of any given microsatellite marker correlates with increasing length of the uninterrupted repeat sequence (Weber, 1990). The

polymorphic nature of microsatellites (the variation in copy number of the repeat unit) may be generated by unequal exchange between sister chromatids and/or strand slippage during DNA replication or repair (reviewed by Levinson and Gutman, 1987).

Recently, interest in trinucleotide repeats has grown because of their involvement in neurological diseases (Richards and Sutherland, 1992; Riggins *et al.*, 1992; Armour *et al.*, 1994; Miwa, 1994; Willems, 1994), despite the fact that microsatellites with a repeat unit greater than two are generally less frequent, occurring on average every 400 kb (Edwards *et al.*, 1991). The characterisation of such trinucleotide repeats has three potential benefits: 1) the isolation of a polymorphic sequence; 2) the identification of a sequence which is potentially unstable, that may account for any associated disease (Richards and Sutherland, 1994); and 3) the isolation of an associated gene. The search for trinucleotide repeats not only contributes to the construction of a linkage map but also to the generation of a transcriptional map of the genome. Similarly, gene-based linkage maps may also be constructed using sequence polymorphisms within introns or the 3' untranslated regions of heteronuclear or mRNAs (Poduslo *et al.*, 1991; Sheffield *et al.*, 1992; Beier, 1993).

Tetrameric repeats have also been isolated and used in the construction of linkage maps (Edwards *et al.*, 1991; Armour *et al.*, 1994; Sheffield *et al.*, 1995; UTAH marker development group, 1995). The advantage of tetra- and trinucleotide repeat markers over dinucleotide markers is that the alleles are easier to score because they are more easily separated upon gel electrophoresis and are less susceptible to polymerase slippage during the PCR. These types of markers may also be isolated from, and thus used to fill, the "gaps" in dinucleotide linkage maps (Sheffield *et al.*, 1995).

## **2.2 CEPH reference pedigrees**

A genetic map is constructed by following the pattern of inheritance from parents to offspring of any given marker and relating it to the pattern of inheritance of other markers. Markers that are linked will be inherited together more often than not, at a statistically significant frequency. Unlinked markers are those that are either on different chromosomes or syntenic but far enough apart such that there is a 50% chance of recombination between them.

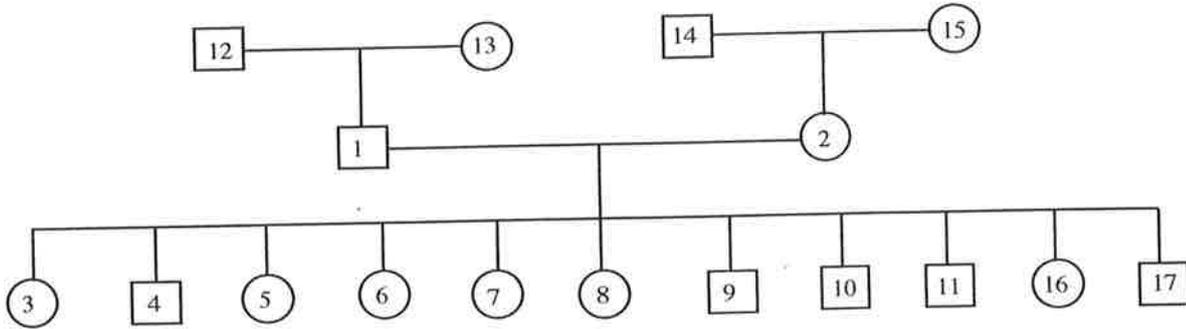
A common set of reference families facilitated the construction of a genetic map of entire chromosomes using markers characterised and genotyped by many different laboratories. To this end, the Centre d'Etude du Polymorphisme Humain (CEPH) research institute was established in 1984. Its function has been reviewed by Dausset *et al.* (1990). CEPH maintains and distributes to over 60 collaborating laboratories around the world, DNA from EBV-transformed lymphoblastoid cell lines (LCLs) of each member of the 40, three generation, reference families. (The number of families has since been increased to 60). These families are from the USA, France and Venezuela and most consist of all four grandparents and an average of 8.5 children (Fig. 2.1). The linkage phase (that is, whether particular alleles of different loci are on the same [*cis*] chromosomal homologue or alternative [*trans*] homologues) of alleles of heterozygous parents can often be inferred from the grandparents and thus the grandparental origin of the alleles in the third generation can be determined, permitting the unequivocal identification of recombinants if the marker is fully informative. There are about 330 potentially informative X-chromosomal meioses provided by the CEPH pedigrees. Collaborating laboratories submit marker genotype data from all of the original (or "standard") 40 families at least, thus contributing to a single and steadily increasing database.

### **2.3 Linkage mapping**

A centimorgan (cM) is, by definition, the distance within which there is expected to be one crossover every 100 meioses (that is, 1 cM is equivalent to a recombination frequency of 1%). The recombination frequency or fraction ( $\theta$ ) is the frequency of recombination between two or more loci, as observed in offspring. The loci can be anonymous polymorphic DNA sequences or one or both could be disease loci.

The relationship between recombination frequency and map distance (cM) is approximately linear over short physical distances but, due to the possible occurrence of multiple crossovers, the recombination frequency underestimates map distance over longer intervals. In other words, the larger the physical distance between two loci the greater the chance of multiple (undetected) crossovers occurring and therefore the true map distance will be underestimated.

CEPH pedigree 1331



**Figure 2.1** A typical CEPH pedigree.

Positive interference is the phenomenon whereby the presence of one chiasma (considered to be the physical manifestation of a recombination event and in humans is most conveniently observed cytologically in male meioses) inhibits the formation of another “nearby”. As a consequence, positive interference also leads to the underestimation of map distances by recombination frequency. These two factors, multiple crossovers and interference, preclude the summing of recombination fractions of adjacent intervals to give the total recombination fraction across the whole region. Algebraically:  $\theta_{13} = \theta_{12} + \theta_{23} - 2r_{123}$ , where  $r$  is the probability of recombination in both intervals (1-2 and 2-3) simultaneously. With positive interference operating,  $r < \theta_{12} \times \theta_{23}$ . The intensity of interference can be investigated by calculating the coincidence ratio which is  $r_{123} / (\theta_{12} \times \theta_{23})$ . For small distances between loci,  $r \approx 0$  and thus  $\theta_{13} = \theta_{12} + \theta_{23}$ .

These non-additive recombination fractions can be converted to additive map distances by the use of an equation called a “map function”. The choice of which map function to use depends on the assumptions made regarding interference. Interference has been studied primarily in *Drosophila* and mice (see for example Weeks *et al.*, 1994), which provide sample sizes sufficient to detect this phenomena, and hence the choice of which mapping function to use for a human gene map is somewhat arbitrary. Nevertheless, choice of the appropriate mapping function (that is, the true model of interference) can be important for relating recombination frequency to map distance and thus for ordering loci correctly. As the evidence for interference comes from the observations of those (rare) chromosomes with multiple crossovers, it has been calculated that to detect this phenomenon in humans more than 490 phase-known meioses, fully informative for 3 markers separated from each other by a recombination fraction of 0.18, would be required (Lathrop and Lalouel, 1991).

A widely used map function devised by Kosambi (1944), using data from *Drosophila* is based on the assumption that the level of interference decreases linearly from complete to none as the recombination fraction between two loci increases. This map function seems to fit empirical observations in *Drosophila*, but does not apply to large recombination fractions in mice (cited by Conneally and Rivas, 1980). The mapping function developed by Rao *et al.*, (1977) was the first to be based on human (male) chiasmata data, nevertheless, citing data from experiments with *Drosophila*, these authors considered that the Kosambi mapping function (which is a special case of their “general” function) may be

appropriate for human female meiosis and thus crossovers involving the X chromosome. The Kosambi function is mathematically simple, but not realistic in the sense that interference, in reality, could be species, sex and position specific. (The first and second factors could be taken into account but the third factor makes any map function unrealistic). The Kosambi function is also unrealistic when used for multipoint mapping of more than 3 loci (and as a consequence is described as being “not multilocus feasible”). This is because when it is necessary to calculate the probability of every possible haplotype that may be passed on by a phase unknown parent, negative probabilities (which are impossible in reality) result (Ott, 1991). To overcome this problem other map functions have been developed which *are* multilocus feasible but they have the disadvantage of assuming interference is not complete at small recombination fractions and therefore they also are unrealistic. Despite the reservations, the Kosambi map function is still predominantly used for the construction of human (multi-locus) linkage maps. The formula is:

$$x \text{ cM} = 25 \ln[(1 + 2\theta) / (1 - 2\theta)]$$

As the density of markers increases the choice of which function to use becomes a moot point because at low recombination fractions (< 0.10) the equations show the recombination fraction to be linearly proportional to the map distance (Conneally and Rivas, 1980) because of the unlikely occurrence of recombination between closely linked loci.

Sex differences in recombination frequency exist and the often quoted sex-averaged (or neuterized) autosomal length is 3,300 cM (Renwick, 1971), however more recent mapping efforts give longer lengths, ranging from about 3,400 to 4,700 cM (NIH/CEPH Collaborative Mapping Group, 1992; Weissenbach *et al.*, 1992). There are very few measures of map distance, as determined cytologically from the observations of female meiotic chiasmata, for the human X chromosome because of the difficulty in obtaining suitable ovarian tissue. An estimate of 225 cM is based on the length of the X chromosome as seen at metaphase relative to the total estimated haploid autosomal length in females of 3,850 cM (Renwick, 1971). More recently, Morton (1991) cites the length of the X chromosome as 220 cM although he does not correctly indicate the source of this estimate (but it was probably calculated relative to the autosomal length of the genome). The physical length of the X chromosome is about 160 Mb (Morton, 1991). (Physical

measurements of distance, as opposed to genetic measures ( $\theta$  or cM), are determined from the sizes of DNA restriction fragments containing the loci of interest or the frequency of the co-inheritance of markers following the shearing of the DNA, either mechanically or by the use of X rays (Cox *et al.*, 1990; Dear and Cook, 1993)).

Statistical support for marker order and determination of the genetic distance between two linked markers is dependent upon the number of informative meioses, that is, the number of gametes (as revealed in the genotypes of the children of the CEPH pedigrees, for example) that can be clearly scored as recombinant or non-recombinant. This number of informative meioses is dependent on the size of the family studied, or the number of families, and the heterozygosity of the markers used. In an ideal situation, where the phase of a doubly heterozygous parent is known from the grandparental genotypes and the other parent is doubly homozygous (that is, a phase-known double backcross mating) or if both parents are double heterozygotes with distinct phase-known genotypes, the recombination fraction between the two loci is simply the number of observed recombinants divided by the total number of offspring. In human pedigrees however, because of the limitation imposed by both the joint heterozygosity of the markers under question and the pedigree structures, this situation is rarely encountered (hence the advantage of the CEPH pedigrees and development of highly informative markers for constructing a genetic map) and so recombinant offspring are not always obvious. Statistical analyses, most commonly maximum likelihood estimation (MLE) methods, are then used to determine the distance between two loci or to order loci on a map. Mapping the X chromosome is simplified relative to the autosomes because the phase of the X chromosome in the mother can always be deduced from the genotype of her father, who is hemizygous, thus facilitating the identification of recombinant children in the third generation. Also, because the father is hemizygous every mating involving a heterozygous mother is informative because this is equivalent to a test cross where one parent is homozygous (in the autosomal situation).

There have been a number of direct and indirect methods applied to the estimation of linkage (for a review see Ott, 1991). The likelihood test of Haldane and Smith (1947) is the pre-cursor of the methods now used. It involves comparing the likelihood (use of the term "probability" is not technically correct: probability refers to events, likelihood refers to parameters; Edwards, 1989) of observing the pattern of inheritance of the alleles of two loci (one of which may be a disease locus) at a given recombination fraction ( $\theta < 0.5$ ) with

the likelihood of observing this same pattern assuming there is no linkage between each locus ( $\theta = 0.5$ ). (Determining the linkage relationship between more than 2 loci will be considered below). The  $\log_{10}$  of this likelihood (L) ratio was used to permit statistical tests of linkage and allow such “lod” (log of the odds ratio) scores between families to be easily summed (Smith, 1953). Mathematically the lod score can be represented as;

$$Z = \log_{10}[L(\theta)/ L(0.5)]$$

For example in a family of 8 siblings where phase is known and 3 children are recombinants ( $\theta$ ) and 5 are non-recombinants ( $1-\theta$ ) the relative likelihood of linkage, when considering two loci, is  $\theta^3(1-\theta)^5 / (1/2)^8$ , which in terms of the lod score becomes:

$$\begin{aligned} Z &= \log [2^8 + \theta^3 + (1-\theta)^5] \\ &= 8 \log(2) + 3 \log(\theta) + 5 \log(1-\theta) \end{aligned}$$

(using equation 3.2 of Ott, 1991, when  $\theta > 0$ ). Note that 3 recombinants out of 8 offspring gives a recombination fraction of 0.375, which when substituted into the above formula gives the maximum lod score of 0.101. Using any other recombination fraction in the formula gives a lower lod score.

When phase is not known every possible haplotype has to be considered:

$$Z = \log\{(1/2 [\theta^3(1-\theta)^5] + 1/2[\theta^5(1-\theta)^3]) / (1/2 [0.5^3 (1 - 0.5)^5] + 1/2 [0.5^5(1-0.5)^3])\}$$

(using equation 4.7 of Ott, 1991). The first term in the denominator and numerator relates to the probability that the two alleles are in *cis* (and therefore 5 of the children are non-recombinants) and the second term relates to the probability that they are in *trans* (and therefore 3 of the children are non-recombinants). The division by 2 is a consequence of each of the two possible phases having a probability of occurrence of 1 in 2.

The maximum likelihood estimate is that recombination fraction ( $\theta$ ) which yields the highest lod score. Morton (1955) proposed a test of significance of these lod scores, resulting from two-point linkage analyses, and implemented a sequential test, whereby data is progressively collected from independent families until, upon summing the lod scores, certain critical values are exceeded. These critical lod score values are 3 and -2. If the lod score is 3 or greater, the frequency of recombination is *significantly* less than 50% (that is, independent assortment is disproved and linkage is demonstrated), whereas if the lod score is -2 or less, the frequency of recombination is *significantly* greater than the recombination fraction for which the lods were calculated. For scores between -2 and 3, conclusions regarding linkage or otherwise are withheld, pending more data (that is, from the analysis of more individuals of the same family, more families or more markers to render additional parts of the family informative).

The selection of 3 and -2, as cut-off points or critical values, are a consequence of Morton's desire to have a low frequency of Type I errors (accepting linkage between loci when in reality there is none) and Type II errors (rejecting linkage when in reality the loci *are* linked), but still have the power of the test equal to 0.99. These two errors occur with a probability of  $\alpha$  and  $\beta$  respectively. The significance level of the test is the probability of a Type I error occurring ( $\alpha$ ). The power of the test is the probability that it correctly rejects the null hypothesis of no linkage ( $\theta = 0.5$ ) when it is false, and is equal to  $1-\beta$ . To maximise the chances of arriving at the correct conclusion  $\alpha$  and  $\beta$  should be minimised, but as the probability of making a Type I error is reduced, that is, making  $\alpha$  more stringent, the probability of making a Type II error is increased and hence the power of the test is decreased. At the critical lod score for accepting linkage of 3 (odds of 1000:1 against no linkage)  $\alpha$  is approximately equal to 0.001. At the critical lod score for rejecting linkage of -2,  $\beta$  is approximately equal to 0.01. (There is some doubt regarding the theoretical justification of using these critical values and Chotai (1984) concludes that they are unnecessarily stringent).

Specifically, the probability of any two autosomal loci selected at random being on the same chromosome is about 5%, which is based on the relative lengths of the autosomes

(Renwick, 1969; Elston and Lange, 1975). However, these two loci, when they are on the same chromosome, also have to be within measurable distance (arbitrarily chosen as < 40% recombination). Taking these prior probabilities into account, the chance of any two randomly chosen autosomal loci being linked is 2% (Renwick, 1969; Elston and Lange, 1975). If the significance level ( $\alpha$ ) of the test is set at the usual 5%, given the relatively low prior probability of linkage, such a test will yield more false positive than true results. To avoid this,  $\alpha$  is made more stringent (0.1%). This is achieved by using a lod score of 3 as the critical point for accepting linkage (equations 3.7 and 4.10 Ott, 1991). For Morton's required power of 99%,  $\beta$  is equal to 1%, hence the use of -2 as the critical lod score for declaring no linkage. The lower lod score of 2 is used as the critical point for declaring linkage between X-linked loci and this is a consequence of the fact that, from their pattern of inheritance, the loci are known to be on the X (but they must still be within measurable distance).

In multipoint linkage analysis (see below), one wishes to construct a map of more than 2 loci or position a locus (the "test" locus, which may be a disease gene) within an existing map. Considering many loci increases the probability of a Type I error because of the individual Type I errors associated with each comparison between the test locus and the other loci. This effect is offset by the increased prior probability of linkage between the test locus and one or more of the other loci (equation 4.16 Ott, 1991). As a consequence, the critical lod score of 3 (or 2 for the X chromosome) applies to both two-point and multipoint analyses.

Morton's (1955) sequential testing method (see above) was only applicable to nuclear families (parents and offspring). This was because the difficulty of the calculations required that larger pedigrees be divided into smaller more manageable units, but the lod scores from each of these units could not be summed as the units were not statistically independent. To facilitate the analysis of extended families, Elston and Stewart (1971) devised an algorithm that "dissected" a pedigree. This algorithm was used in a computer program developed by Ott (1974, 1976) called "LIPED" (**l**ikelihood in **p**edigrees). The Elston-Stewart algorithm is, however, not suited for maps of many loci because the computing time required grows exponentially with the number of markers used. To overcome this, Lander and Green (1987) devised an algorithm that permits computer time

to increase linearly with the number of loci, which has made multipoint analyses feasible.

Calculating linkage between 3 or more loci simultaneously (multipoint analysis), as when constructing a map of marker loci or positioning a disease locus within an existing map, is more efficient in determining distance and gene order, compared to two-point analysis. This is because the more loci used in an analysis the more information is available, in the sense that an incorrect order of 3 or more loci will require an unlikely number of recombination events to account for it and there is a greater chance that individuals uninformative for one marker will be informative for one of the others (this is especially important if small families are used for map construction; Lathrop *et al.*, 1985). This multipoint method evaluates the likelihood of a locus order given the genotypic data and this is then compared with the likelihood of other locus orders, using the likelihood ratio test. When only a few loci are to be ordered, the maximum likelihood locus order can be determined by comparing the likelihood of *all* possible locus orders. The number of orders that are possible equals  $n!/2$ , where  $n$  is the number of loci and the division by 2 reflects the fact that an order of ABCD is the same as DCBA. As the number of loci to be ordered increases, the number of possible orders increases exponentially, which becomes computationally difficult. The CRI-MAP program deals with this problem by using the algorithm of Lander and Green (1987). Loci are added sequentially to a starter map first constructed with a nucleus of two highly informative loci. More loci are added to the map providing they can be positioned with odds of at least 1000:1 against any other location. To assess the final map, support for local order is tested by inverting adjacent groups of loci.

It was not until 1984 that computer programs which permitted the analysis of many more than four loci simultaneously were available. The most commonly used computer programs used to construct linkage maps are: MAP (Morton and Andrews, 1989), CRI-MAP (Lander and Green, 1987), MAPMAKER (Lander *et al.*, 1987) and LINKAGE (Lathrop *et al.*, 1984, 1985). All use the maximum likelihood method for ordering loci and determining the distance between them, however they differ in the way the likelihood is maximised and the assumptions regarding interference. The MAP program uses a multiple two-point analysis and assumes interference. The other three programs use multipoint analysis and calculate adjacent recombination fractions assuming no interference during the calculation of gamete probabilities (although MAPMAKER and CRI-MAP can then convert

these recombination fractions to map distances using the Kosambi mapping function).

As linkage maps increase in density it will become increasingly difficult to order markers with odds of >1000:1 because of the rarity of recombination events between closely spaced loci. Wang and Weber (1992) have calculated that 3 to 4 recombination events between markers should be sufficient to order them with odds of >1000:1. There are two options to increase the number of observed recombinations between closely linked loci: 1) investigate additional markers and/or 2) genotype more individuals within existing families or use new families. As regards the X chromosome, there are about 330 potentially informative X chromosomal meioses provided by the 40 standard CEPH families. This number permits the ordering of X linked loci that are separated by only 2.7 cM, assuming the X chromosome is 220 cM in length and using the formula and assumptions given by Ott (1991, page 137).

As the resolution of genetic mapping using family data is limited to markers separated by more than about 1 cM, to obtain a correctly ordered higher resolution map one needs to resort to physical methods (Chakravarti *et al.*, 1991), although “linkage disequilibrium” mapping may in the future also be used to refine the order of closely spaced markers because this procedure analyses the recombination events within a *whole* population (that has descended from a small founder population and remained relatively isolated) (Peterson *et al.*, 1995).

## 2.4 Physical mapping

Apart from the goal of generating a 2-5 cM genetic map by 1995, the Human Genome Project aims to produce a physical map by 1998 that has a STS (sequence tagged site) every 100 kb (Collins and Galas, 1993). Integration of the genetic and physical maps can be achieved by using polymorphic STSs, of known order, to “seed” the construction of the physical map.

High resolution physical mapping can be achieved by observing the co-segregation of polymorphic markers following irradiation or mechanical shearing, known as “radiation hybrid” and “happy” mapping respectively (Cox *et al.*, 1990; Dear and Cook, 1993). These methods are the physical analogues of meiotic (recombination) mapping and can be analysed in the same manner. The limit to their resolution is dependent on the frequency of

breakage and the statistical methods used to analyse the results. Natural examples of chromosome breakage are revealed in patients with contiguous disease syndromes and the characterisation of the deletions involved can also reveal marker orders (see for example Philippe *et al.*, 1993).

Alternatively, fluorescence *in situ* hybridisation (FISH) of cloned probes to metaphase or interphase chromosomes or to extended DNA strands (“fibre fish”) can be used to locate and order markers. The limit of resolution of each of these *in situ* methods is 1 Mb, 50 kb and 20 kb respectively (Trask, 1991; Brandiff *et al.*, 1991; Buckle and Kearney, 1994; Haaf and Ward, 1994).

Restriction maps of large regions of DNA can also be used to position loci relative to other loci by observing the pattern of restriction fragments that contain these loci.

Finally, large regions of the genome (from about 100 kb up to about 1 Mb) can be cloned and propagated using yeast vectors and subsequently analysed to determine which STS markers they contain. In this manner contiguous sets of clones (for example a yeast artificial chromosome (YAC) “contig”) can be established across the region of interest, from which smaller more manageable segments of DNA (cloned into: bacterial artificial chromosomes, cosmids, bacteriophage or plasmids) can be generated. A first generation physical (YAC contig) map of the human genome has been generated by Cohen *et al.* (1993). Almost complete YAC coverage of the X chromosome has recently been achieved (Roest Crollius *et al.*, 1996; Nagaraja *et al.*, 1997).

The generation of overlapping contigs will ultimately lead to the physical map with the highest resolution; that is, the entire nucleotide sequence of the genome. The goal for completion of this map is 2006 (Smith, 1993), but this depends upon further advances in sequencing technology.

## **2.5 Morbid anatomy of the human X chromosome**

In 1980, Dr. Victor McKusick coined the term “morbid anatomy” to refer to the assignment of disease loci to “specific” chromosomal sites, that is, the construction of a disease map of the human genome (McKusick, 1980). Some diseases are allelic (that is, clinically distinct conditions can arise from mutations in the same gene; Saugier-Weber *et*

*al.*, 1994; Chapter 8) and conversely some diseases are heterogenous (the same clinical phenotype caused by mutations in different genes; Gedeon *et al.*, 1996b; Chapter 6) hence the number of diseases (or clinically defined phenotypes) does not reflect the number disease-gene loci. Historically the X chromosome has been well studied because of the revelation of X-linked recessive diseases in hemizygous males that are carried by normal females. As at 1990 about 100 disorders had been shown to be X-linked; which was the highest disorder density of any chromosome (McKusick, 1990). In comparison, about 470 autosomal disorders had been allocated to their respective autosomes. As at 1992, the genes responsible for 26 of the 100 known X-linked disorders had been cloned (Mandel *et al.*, 1992).

By 1994, the number of disorders localised to the X chromosome had increased by 50%, and the number localised to autosomes had increased by 73% (McKusick, 1994). This rapid rate of localisation was facilitated no doubt by the availability of high resolution linkage maps and highly polymorphic PCR-based genetic markers, enabling more information to be extracted from disease families.

## **2.6 Positional cloning**

Most disease genes have been only regionally localised, because of the lack of any information regarding their function. Identification of genes through a knowledge of their protein sequence, structure or function is referred to as “functional cloning”. In contrast, “positional cloning” is the process whereby the disease gene is isolated and identified solely because it lies within the regional localisation of the disease locus or is disrupted by some chromosomal abnormality. The “positional candidate” approach to cloning relies on the investigation of one or more *previously* identified candidate genes that are known to lie within the disease gene regional localisation.

The first gene to be isolated by positional cloning was that responsible for chronic granulomatous disease in 1986, and by 1992 the genes of 12 other inherited diseases had been identified by this approach (Collins, 1992; Ballabio, 1993). Six of these genes are located on the X chromosome. This bias of X-linked genes cloned by positional cloning is because of the functional hemizyosity of this chromosome, the consequences of which are the clinical detection of (hemizygous) males and those females with cytogenetic abnormalities involving the X chromosome. The presence of obvious cytogenetic lesions

makes identification of the associated genes relatively easy, as was the case for the genes responsible for all 6 X-linked disorders.

In most cases, however, there is no structural abnormality, nor independently isolated candidate gene, known to lie within the same region as the disease locus, and therefore it may be necessary to randomly clone genes from this region and screen them for mutations in the affected members of the family under study. Knowledge of the tissue of expression, the developmental stage of expression and the putative function of any genes isolated can be used to prioritize them for subsequent mutation screening.

Although gene density varies greatly, the regions of greatest density may have one gene every 20 kb (Kendall *et al.*, 1990; Vyas *et al.*, 1992; Coleman *et al.*, 1994). As a consequence it is advantageous to minimise the region to be searched, so as to minimise the effort expended in isolating and characterising those genes present. Using high resolution genetic maps is very often the only way to minimise the regional localisation of a disease gene segregating within a family. These high resolution maps enable the investigator to progressively “close in” on the actual site(s) of recombination. In addition, with increasing resolution of genetic maps it may be possible, as has been done in studies with rodents, to reveal the genes involved in polygenic and multifactorial disorders (Collins, 1992; Guyer and Collins, 1993; Davies *et al.*, 1994; Lander and Schork, 1994; Epplen *et al.*, 1995; Kruglyak and Lander, 1995).

## **2.7 Identification of disease-causing genes**

There are a number of methods to isolate genes from relatively small (<1 Mb) regions of the genome (see for example Hochgeschwender and Brennan, 1991). In the absence of any other information other than position, all such genes are candidates for causing the disease that maps to the same region. These genes then have to be investigated for mutations. Sub-microscopic deletions are relatively easy to detect but occur with a frequency of only about 5% for each well-studied X-linked disease (Mandel *et al.*, 1992). Most diseases are caused by point mutations (within exons) and as a consequence a relatively sensitive method will need to be employed to detect and characterise any particular lesion. Such methods have been reviewed by Landegren (1992). Sequencing is the ultimate means of revealing any differences between the DNA or cDNA of affected and unaffected individuals. This procedure is tedious if there are many exons, unless one is

sequencing a small cDNA. Prior to sequencing, more rapid, but less sensitive, screening methods may help reduce the region to be sequenced. These often involve detecting mismatches between heteroduplex DNA strands (composed of wild-type and mutated sequences) or rely on differences in the secondary structure adopted by wild-type and mutated single-stranded DNA or RNA as revealed by single-stranded conformational analysis (SSCA) (Grompe, 1993).

Once a difference has been found it has to be demonstrated that it is responsible for manifestation of the disease. This can be done by combinations of the following:

- 1) population screening for the same variant, to exclude polymorphism;
- 2) demonstrating segregation of the variant with the disease within the family and in other families with the same disease (however, different mutations in the same gene may cause the same phenotype);
- 3) showing conservation of nucleic acid or amino acid sequence across species, at the site of the variant;
- 4) identification of the same mutation in animal models of the disease;
- 5) generating transgenic animal models with the same mutation;
- 6) determining any effect on transcription;
- 7) assay for protein function;
- 8) molecular modelling of the protein structure.

## **2.8 Transcriptional mapping**

Isolating and identifying genes responsible for diseases is one means by which a transcriptional (or “functional”) map of the genome is constructed. A more efficient and comprehensive process is the isolation and sequencing of (partial) cDNA sequences (see for example Hillier *et al.*, 1996). The rate-limiting step in the construction of a transcriptional map of the genome is not the isolation of these expressed sequence tags (ESTs) but their subsequent genetic localisation and characterisation of the genes to which

they belong.

The physical mapping of the genome will facilitate the mapping of ESTs. Expressed sequence tags have been regionally mapped primarily by FISH and PCR from cell hybrids (see for example Parrish and Nelson, 1993) and, more recently, positioned with higher precision *via* PCR from mapped YACs (Schuler *et al.*, 1996). A transcriptional map of the genome will greatly assist positional cloning projects, because all transcripts that map within the relevant region are candidates for being the disease gene (although some will be better candidates than others).

There is some debate regarding the necessity of sequencing the entire (coding and non-coding) genome when the “important bits” can be found much more readily by cDNA cloning (Chen *et al.*, 1994). It is likely, however, that each of these approaches will serve as catalysts for the other; with the analysis of many cDNAs ultimately requiring the elucidation of their genomic organisation and, conversely, the analysis of “anonymous” raw sequence data revealing genes by the application of exon-recognising computer algorithms.

# Chapter 3

## Materials and Methods

## Table of Contents

Page

### **3.1 Materials**

3.1.1 Chemicals and compounds and their suppliers

3.1.2 Solutions

3.1.3 Kits

3.1.4 Media

3.1.4.1 Liquid media

3.1.4.2 Solid media

3.1.5 Antibiotics

3.1.6 Bacterial strains

3.1.7 Vectors

3.1.8 Enzymes

3.1.9 DNA size markers

3.1.10 Libraries

3.1.11 Miscellaneous

### **3.2 Methods**

3.2.1 Propagation of bacteriophage

3.2.1.1 Propagation of *lambda* (Charon 35)

3.2.1.2 Isolation of *lambda* DNA

3.2.1.3 Propagation of M13 bacteriophage

3.2.1.4 Isolation of M13 DNA

3.2.2 Plaque-lifting

3.2.3 Southern blotting

3.2.4 Radio-labelling DNA

3.2.5 Pre-reassociation of DNA

3.2.6 Hybridisation

3.2.7 DNA extraction and purification

3.2.7.1 Phenol/chloroform extractions and DNA precipitation

3.2.7.2 Phage DNA

3.2.7.3 YAC/Yeast DNA

3.2.7.4 Peripheral lymphocyte DNA

3.2.7.5 PCR products

- 3.2.7.6 Cosmid DNA
- 3.2.8 Restriction digests
- 3.2.9 Dephosphorylation
- 3.2.10 Ligations
- 3.2.11 Transfections/transformations
  - 3.2.11.1 Transformation of recombinant M13 DNA
  - 3.2.11.2 Transfection of recombinant pSPL3 (Exon Trapping)
- 3.2.12 DNA sequencing
  - 3.2.12.1 Single-stranded M13 DNA
  - 3.2.12.2 Double-stranded *lambda* DNA
  - 3.2.12.3 PCR products
  - 3.2.12.4 Cloned PCR products
- 3.2.13 Sequence analysis
- 3.2.14 Oligonucleotide synthesis and purification
- 3.2.15 PCR conditions
  - 3.2.15.1 Microsatellites
  - 3.2.15.2 *Alu*-PCR
  - 3.2.15.3 Colony PCR
- 3.2.16 Electrophoresis
  - 3.2.16.1 Agarose gel
  - 3.2.16.2 Sequencing gels
  - 3.2.16.3 Resolving microsatellite alleles
  - 3.2.16.4 Non-denaturing (SSCA) gels
- 3.2.17 Fluorescence *in situ* hybridisation (FISH)
- 3.2.18 RNA isolation
  - 3.2.18.1 RNA isolation from cultured eukaryotic cells
  - 3.2.18.2 RNA isolation from fresh tissue
- 3.2.19 cDNA synthesis
  - 3.2.19.1 Reverse transcription of exon trapped products
  - 3.2.19.2 Reverse transcription of RNA isolated from fresh tissue
  - 3.2.19.3 *rTth* PCR
- 3.2.20 Northern blotting

### 3.2.21 Linkage analysis

#### 3.2.21.1 Two-point linkage analysis

#### 3.2.21.2 Multipoint linkage analysis using single (disease) family data

#### 3.2.21.3 Multipoint linkage analysis using CEPH data

### 3.1 Materials

#### 3.1.1 Chemicals and compounds and their suppliers

Acrylamide (99.9%)	BioRad
40% (w/v) Acrylamide:Bisacrylamide (19:1) solution	BioRad
30% or 40% (w/v) Acrylamide solution	BioRad
Adenine	Sigma
Agarose: NA grade	Pharmacia
Agarose: Low Melting Point (LMP)	Seaplaque (FMC)
Ammonium acetate	Ajax
Ammonium chloride	Ajax
Ammonium hydroxide	Ajax
Ammonium hydroxide; Conc. (32% w/w)	Merck
Ammonium persulphate	BDH
Ammonium sulphate	Ajax
Ampicillin	Sigma
[ $\gamma$ - <sup>32</sup> P]dATP; 5000 Ci/mmol	Amersham
Bactoagar	Sigma
Bactotryptone	Difco
Benzyl penicillin	CSL
Bis(N,N-methylene- <i>bis</i> -acrylamide)	BioRad
2% (w/v) Bis solution	BioRad
Boric acid	Ajax
Bromophenol blue	BDH
BSA (bovine serum albumin)	Sigma
<i>N</i> -Butanol (butan-1-ol)	Ajax

Calcium chloride	Ajax
Casein hydrolysate acid	Gibco
Chloroform (trichloromethane)	Ajax
Chloroform (for purification of cycle sequencing products)	Applied Biosystems
DAPI (diamidino phenylindole dihydrochloride)	Sigma
[ $\alpha^{32}\text{P}$ ]dCTP; 3000 Ci/mmol	Amersham
[ $\alpha^{35}\text{S}$ ]dCTP; 1177 Ci/mmol	Bresatec
DEPC (diethylpyrocarbonate)	Ajax
Deoxy- and dideoxy-nucleotide triphosphates	Pharmacia
Dextran sulphate	Pharmacia
<i>N,N</i> -dimethyl formamide	Ajax
DMSO (dimethylsulphoxide)	Ajax
DTT (dithiothreitol)	BioRad
EDTA (ethylenediaminetetracetic acid; $\text{Na}_2\text{EDTA}\cdot 2\text{H}_2\text{O}$ )	Ajax
Ethanol (99.5% v/v)	Ajax
Ethidium bromide	Sigma
FCS (Fetal Calf Serum)	CSL
Formaldehyde 37%	Ajax
Formamide (deionised before use)	Fluka Chemika
Gaunidium thiocyanate	ICN
Gelatin	-
Gelslick	AT Biochem
Glacial acetic acid	Ajax
Glucose	Ajax
Glutamine	Trace Biosciences
Glycerol	Ajax

Hams F12 (+ L-glutamine; - NaHCO <sub>3</sub> )	Trace Biosciences
Human placental DNA	Sigma
Hydrogen chloride (36% w/w)	Ajax
IPTG (isopropylthio-β-D-galactosidase)	BRL
Isoamyl alcohol	Ajax
Isopropanol	Ajax
Kanamycin	Boehringer Mannheim
<i>N</i> -lauroylsarcosine (Sarkosyl)	Sigma
LipofectACE	Gibco BRL
Magnesium chloride (MgCl <sub>2</sub> .6H <sub>2</sub> O)	Ajax
Magnesium sulphate	Ajax
Maltose	BDH
MDE (2 x gel solution)	FMC
MEM I (reduced serum medium)	Gibco BRL
βMe (β-mercaptoethanol)	BDH
NP-40 (Nonidet P-40)	Sigma
Paraffin	Ajax
PEG (polyethylene glycol) 3350 and 8000	Sigma
PEG 6000	BDH
Phenol	Wako
Phenol:H <sub>2</sub> O:chloroform (68:18:14)	Applied Biosystems
Phenol red	CSL
Phosphate Buffered Saline (-Mg,-Ca,-NaHCO <sub>3</sub> )	Trace Biosciences
Potassium acetate	Ajax
Potassium chloride	Ajax

di-Potassium hydrogen orthophosphate (K <sub>2</sub> HPO <sub>4</sub> )	Ajax
Potassium dihydrogen orthophosphate (KH <sub>2</sub> PO <sub>4</sub> )	Ajax
RNAsin	Promega
Salmon sperm DNA	Calbiochem
Sigmacoate	Sigma
SDS (sodium dodecyl sulphate)	Sigma
Sodium acetate	Ajax
Sodium bicarbonate	Astra pharmaceuticals
Sodium chloride	Ajax
tri-Sodium citrate	Ajax
di-Sodium hydrogen phosphate	May and Baker Aust.
Sodium hydroxide	Ajax
Sodium phosphate	Ajax
Sorbitol	Sigma
Spermidine-trihydrochloride	Calbiochem Corp.
Sucrose	Ajax
TEMED (N,N,N',N'-tetramethylethylenediamine)	BioRad
Tetracycline	Sigma
Tetrazole	Applied Biosystems
Thiamine HCl	Sigma
Trichloroacetic acid	Applied Biosystems
Tris (hydroxymethyl aminomethane) base	Boehringer Mannheim
Tris-HCl	Boehringer Mannheim
Triton X 100	Ajax
Tween 20 (polyoxyethylene (2) sorbitan mono-laurate)	Ajax
Urea	Ajax

Urea (for DNA sequencing)	BioRad
Yeast extract	Gibco
Yeast nitrogenous base (with no amino acids)	Difco
X-gal (5-Bromo-4-chloro-3-indoyl- $\beta$ -D galactosidase)	BRL
Xylene cyanol	Serva
Zinc chloride	Mallinckrodt Inc.

### 3.1.2 Solutions

All solutions were made with deionised water (dH<sub>2</sub>O). Solutions provided with kits are not necessarily mentioned.

6% (w/v) acrylamide/7 M urea: 0.3% w/v bisacrylamide, 5.7% w/v acrylamide, 42% w/v urea, 1 x TBE. This was filtered through 3mm Whatman chromatography paper.

2 x *Alu*-PCR mix: 100 mM KCl, 20 mM Tris-HCl, 3.0 mM MgCl<sub>2</sub>, 0.02% (w/v) gelatin, 0.1% (v/v) NP-40, 600  $\mu$ M of each dNTP, (pH 8.0).

10 x CAP buffer: 500 mM Tris-HCl, 10 mM MgCl<sub>2</sub>, 1 mM ZnCl<sub>2</sub>, 10 mM spermidine.

Cell lysis buffer: 0.32 M sucrose, 10 mM Tris-HCl, 5 mM MgCl<sub>2</sub>, 1% v/v Triton X-100, (pH7.5).

Denaturing solution: 1.5 M NaCl, 0.5 M NaOH.

Dilution (AmpliTaq) buffer: 0.1 mM EDTA, 0.15% (v/v) Tween 20, 0.15% (v/v) NP-40, 25 mM Tris-HCl, (pH 8.8).

Filter denaturing solution: 0.5 M NaOH.

Filter neutralising solution: 0.2 M Tris-HCl, 2 x SSC, (pH 7.5).

Gel denaturing solution: 2.5 M NaCl, 0.5 M NaOH.

Gel neutralising solution: 1.5 M NaCl, 0.5 M Tris-HCl, (pH 7.5).

Hybridisation solution: 5 x SSPE, 1% w/v SDS, 50% v/v deionised formamide, 0.15 M dextran sulphate.

Labelling mix: 1.5  $\mu$ M each of dGTP, dATP, and dTTP. (To improve resolution, dGTP was replaced with  $dc^7$ GTP; Mizusawa *et al.*, 1986).

5 x Ligation buffer: 250 mM Tris-HCl, 50 mM MgCl<sub>2</sub>, 5 mM DTT, 5 mM ATP, 25% w/v PEG 6000, (pH 7.6).

2 x Loading buffer: 96% v/v deionized formamide, 10 mM EDTA, 0.05% w/v bromophenol blue, 0.05% w/v xylene cyanol, (pH 8.0).

10 x Loading buffer: 1% w/v SDS, 0.05% w/v bromophenol blue, 0.05% w/v xylene cyanol, 50% v/v glycerol, 0.1 M EDTA.

M13 precipitation solution: 25% w/v PEG 6000, 2.5 M NaCl.

MDE gel solution: 37.5% (v/v) 2 x concentrated FMC solution; 0.6 x TBE.

Neutralising solution: 1.5 M NaCl, 0.5 M Tris-HCl, 1 mM EDTA, (pH 7.2).

PBS (Phosphate Buffered Saline): 138 mM NaCl, 2.7 mM KCl, 8.1 mM Na<sub>2</sub>HPO<sub>4</sub>·7H<sub>2</sub>O, 1.2 mM KH<sub>2</sub>PO<sub>4</sub>, (pH 7.4).

2 x PCR himix (as used by Kogan *et al.*, 1987): 33 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 133 mM Tris-HCl, 20 mM  $\beta$ ME, 0.013 mM EDTA, 0.34 mg/ml BSA, 20% v/v DMSO, 3 mM dATP, 3 mM dGTP, 3 mM dCTP, 3 mM dTTP, (pH 8.8).

2 x PCR lomix: as above except all deoxynucleotide triphosphates at 0.4 mM.

Phenol: buffered with Tris-HCl (pH 7.4), unless otherwise stated.

3 M Potassium acetate: 30 ml of 5 M Potassium acetate solution, 5.75 ml of glacial acetic acid, (pH 4.3), *per* 50 ml.

3 x Proteinase K buffer: 10 mM NaCl, 10 mM Tris-HCl, 10 mM Na<sub>2</sub>EDTA, (pH 8.0).

RNA denaturing solution: 10 µl 5 x TBE, 9 µl formaldehyde, 25 µl deionised formamide. (In making the TBE, a 5 x solution of borate/EDTA was treated with DEPC-dH<sub>2</sub>O to which the appropriate amount of Tris-HCl was added, after autoclaving).

RNA loading dye: 1 mM EDTA, 500 µl glycerol, 500 µl DEPC-dH<sub>2</sub>O, 0.05% w/v bromophenol blue, 0.05% w/v xylene cyanol.

RNAse denaturing solution: 4 M guanidinium thiocyanate; 25 mM sodium citrate, 0.1 M βMe, 0.5% (w/v) *N*-lauroylsarcosine, (pH 7.0). This working solution must be used within 1 month.

5 x Sequencing reaction buffer: 35 mM MgCl<sub>2</sub>, 250 mM Tris-HCl (pH 8.0).

3 M Sodium acetate: 24.6 g sodium acetate/100 ml, pH to 5.2 with glacial acetic acid.

SM buffer: 50 mM Tris-HCl, 100 mM NaCl, 8 mM MgSO<sub>4</sub>, 0.01% v/v gelatin (pH 7.5).

20 x SSC: 3 M NaCl, 0.3 M tri-sodium citrate, pH to 7.0 with 36% HCl.

5% SSCA (1:60) gel: 0.08% w/v bisacrylamide, 4.92% w/v acrylamide, 0.5 x TBE, 5% v/v glycerol. The resultant solution was treated as for 6% (w/v) acrylamide/7 M urea.

10% SSCA (1:49) gel: 0.2% w/v bisacrylamide, 9.8% w/v acrylamide, 1 x TBE, 10% v/v glycerol. The resultant solution was treated as for 6% (w/v) acrylamide/7 M urea.

20 x SSPE: 3.6 M NaCl, 0.2 M NaH<sub>2</sub>PO<sub>4</sub>·2H<sub>2</sub>O, 0.02 M EDTA.

50 x TAE: 2 M Tris base, 57.1 ml glacial acetic acid, 0.05 M EDTA (pH 8.0).

5 x TBE: 0.45 M Tris base, 0.45 M Boric acid, 0.01 M EDTA (pH 8.0).

TE: 10mM Tris-HCl, 1mM EDTA, (pH 8.0).

Termination mixes:

G - 20  $\mu$ M each of dGTP, dATP, dTTP, dCTP and 60  $\mu$ M ddGTP

A - 20  $\mu$ M each of dGTP, dATP, dTTP, dCTP and 800  $\mu$ M ddATP

T - 20  $\mu$ M each of dGTP, dATP, dTTP, dCTP and 800  $\mu$ M ddTTP

C - 20  $\mu$ M each of dGTP, dATP, dTTP, dCTP and 400  $\mu$ M ddCTP

(To improve resolution, dGTP was replaced with dc<sup>7</sup>GTP; Mizusawa *et al.*, 1986).

TKM: 10 mM Tris-HCl, 10 mM KCl, 1 mM MgCl<sub>2</sub>, (pH 7.5). Note: this was not treated with DEPC.

TSB: LB broth, 10% w/v PEG 3350, 5% v/v DMSO, 50 mM MgCl<sub>2</sub>.

Washing solution I: 2 x SSC, 1% w/v SDS.

Washing solution II: 0.1 x SSC, 0.1% w/v SDS.

X-gal: 50 mM (2% w/v) solution, using N,N-dimethyl-formamide as the diluent.

Yeast lysis solution: 0.9 M sorbitol, 0.1 M EDTA, 0.001 M  $\beta$ Me.

### 3.1.3 Kits

Exon trapping system (Gibco BRL).

*f*mol DNA Sequencing System (Promega).

Magic Prep PCR purification tubes (Promega).

mRNA Purification Kit (Pharmacia).

Multiprime/Megaprime DNA Labelling kits (Amersham) for radioactively labelling DNA probes.

Prep-a-Gene (BioRad) for purification of PCR products.

PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems) for producing fluorescently labelled products from a PCR product template, prior to sequencing.

PRISM Ready Reaction Dye Primer Cycle Sequencing Kit (Applied Biosystems) for producing fluorescently labelled products from a plasmid or M13 template, prior to sequencing.

Qaigen tips, used to purify plasmid and cosmid DNA from cell lysates.

Qiaquick PCR product purification tubes (Qiagen).

RPM (Rapid Pure Mini-preps) kit, used to purify cosmid DNA (BIO 101).

Taq DyeDeoxy Terminator Cycle Sequencing kit (Applied Biosystems) for producing fluorescently labelled products from a PCR product template, prior to sequencing (NB this is superseded by the PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit).

Thermostable *rTth* reverse transcriptase RNA PCR Kit (Perkin-Elmer).

### **3.1.4 Media**

#### **3.1.4.1 Liquid media:**

LB (Luria-Bertani) broth; containing 5 g yeast extract, 10 g bactotryptone and 5 g NaCl *per* litre, pH to 7.5 with NaOH. This was supplemented with 5.6 mM [0.2% w/v] maltose when culturing *E. coli*. strain DL538.

2 x YT; containing 10 g yeast extract, 16 g bactotryptone and 5 g NaCl *per* litre. This was used when generating single-stranded M13 DNA.

TB (Terrific Broth); containing 12 g bactotryptone, 24 g yeast extract, 4 ml glycerol *per* 900 ml, (pH 7.0). After autoclaving a sterile 100 ml solution of [0.17 M KH<sub>2</sub>PO<sub>4</sub>, 0.72 M K<sub>2</sub>HPO<sub>4</sub>] was added.

AHC broth containing; 6.7 g yeast nitrogenous base (w/o amino acids), 10 g casein hydrolysate acid, 20 mg adenine and 20 g glucose *per litre*, (pH 7.0). This was used for culturing yeast.

COS-7 medium: 13 g DMEM, 100 ml FCS, 44 ml NaHCO<sub>3</sub>, 4 mM glutamine, 5 ml benzyl penicillin, *per litre*, (pH 7.1).

F12: 10.73 g Hams F12 powder (Trace Biosciences), 14 ml NaHCO<sub>3</sub>, 100 ml FCS, 1 ml benzyl penicillin, *per litre*, (pH 7.2).

OMI: 1 sachet of MEM I, 28 ml NaHCO<sub>3</sub>, 50 ml FCS, 1 ml benzyl penicillin, 55 mM βME, 1 ml phenol red, *per litre*, (pH 7.1).

#### 3.1.4.2 Solid media:

LB containing 14 g of bacto-agar *per litre* of LB broth for bottom agar.

H agar containing 10 g bactotryptone, 8 g NaCl and 14 g bacto-agar *per litre*.

PL agarose for plate lysates contained LB agarose supplemented with 16.6 mM (0.3% w/v) glucose, 0.075 mM CaCl<sub>2</sub> and 2 mM MgSO<sub>4</sub>.

M9 minimal medium was used when streaking DH5αF' and XL1-Blue strains and contains 1 x M9 salt solution, 10 mM (0.2% w/v) glucose, 1 mM MgSO<sub>4</sub>, 0.1 mM CaCl<sub>2</sub> and 1 mM thiamine-HCl. (10 x M9 solution: 0.17 M Na<sub>2</sub>PO<sub>4</sub>, 0.22 M KH<sub>2</sub>PO<sub>4</sub>, 0.19 M NH<sub>4</sub>Cl, 86 mM NaCl).

Top agar/agarose contained 7.5 g of agar/agarose *per litre*.

#### 3.1.5 Antibiotics

Ampicillin (amp) was used at a final concentration of 50 or 100 µg/ml in liquid and solid media.

Benzyl penicillin was used at a final concentration of 150 µg/ml when culturing eukaryotic

(mammalian) cells.

Kanamycin (kan) was used at a final concentration of 30 µg/ml when culturing Lawrist 4-cosmid containing bacteria.

Tetracycline (tet) was used at a final concentration of 12.5 µg/ml when culturing (non-recombinant) XL1-Blue bacteria or yeast cells.

### 3.1.6 Bacterial strains

All strains were streaked from stocks kept at -70°C in LB + 15% v/v glycerol.

*Escherichia coli* DH5αF': F'/*endA1 hsdR17(rk-mk<sup>+</sup>) supE44 thi-1 recA1 gyrA(Nal<sup>r</sup>) relA1Δ(lacZYA-argF)U169 (ϕ80dlacΔ[lacZ]M15)*; (Woodcock *et al.*, 1989; Raleigh *et al.*, 1989). This strain was cultured in 2 x YT, and used to propagate M13 phage.

*E. coli* DL538: F-*hsdR(rk-mk<sup>+</sup>) mrcA mrcB supE44 recD1009 sbcC201 pro C600*; (Woodcock *et al.*, 1989; Whittaker *et al.*, 1988). This strain was grown overnight in LB broth supplemented with 5.6 mM (0.2% w/v) maltose. The bacteria were then pelleted by centrifugation at 1000 x g for 15 min., 4°C. The pellet was resuspended in 0.4 x [LB culture volume] of 10 mM MgSO<sub>4</sub> and stored for up to 3 weeks at 4°C. This strain was used to propagate *lambda* phage.

*E. coli* XL1-Blue: F':::Tn10 *proA+B+lacIqΔ(lacZ)M15/recA1 endA1 gyrA96(Nal<sup>r</sup>) thi hsdR17(rk-mk<sup>+</sup>) supE44 relA1 lac*: (New England Biolabs [NEB] 1990-1991 catalogue). This strain was cultured in 2 x YT or LB, and used to propagate M13 phage and for all plasmid cloning. It was purchased from Stratagene.

### 3.1.7 Vectors

M13mp18 RFI DNA (NEB) was used for subcloning (AC)<sub>n</sub>-positive *lambda* clones and for the production of (AC)<sub>n</sub>-positive single-stranded M13 DNA for sequencing.

pSPL3 plasmid vector for trapping exons from cosmids (see Kits).

pBluescript SKII vector (Stratagene) for cloning exon trapped PCR products.

### 3.1.8 Enzymes

All enzymes were used according to the manufacturer's instructions, unless indicated otherwise.

AmpliTaq DNA polymerase (Perkin-Elmer Cetus) was used for sequencing and used at a concentration of 8 units/ $\mu$ l.

CAP (Calf intestinal alkaline phosphatase) was purchased from Boehringer-Mannheim.

DNA polymerase I (Klenow fragment) was supplied in the Multiprime and Megaprime DNA Labelling kits from Amersham.

DNase I (Boehringer Mannheim).

Lysozyme (Boehringer Mannheim).

Lyticase (Sigma).

Polynucleotide kinase (Pharmacia).

Proteinase K (Sigma).

Restriction enzymes (NEB or Progen).

SuperScript II RNase H<sup>-</sup> reverse transcriptase (Gibco BRL).

RNase A (Boehringer Mannheim).

T4 DNA ligase (Promega or Bresatec).

*Taq* polymerase (Boehringer Mannheim) was used for PCR other than for sequencing.

### 3.1.9 DNA size markers

*Eco*R1 restricted bacteriophage SPP-1 (Bresatec).

*Hpa*II restricted pUC19 (Bresatec).

DRigest III (Pharmacia).

### 3.1.10 Libraries

The X chromosome library (LAOXNL01; see also below) had been constructed (by the supplier) by inserting *Sau*3A partially digested human DNA fragments into the *Bam*H1 site of the *lambda* derivative; Charon 35: *lsbhIλ<sup>l</sup>lac5 srI lacZ* [polycloning site] *srIλ3 ΔWL113 KH54 nin5 shndIIIλ6° srIλ5°*. (ATCC product sheet; Sambrook *et al.*, 1989 page 2.29). This library was used to isolate (AC)*n* microsatellite sequences.

Clontech cDNA libraries used were: adult and fetal brain (a 1:1:1 mix of λgt10, λgt11 and λZAPII libraries); adult and fetal kidney (a 1:1 mix of λgt10 libraries); placenta (λgt11).

### 3.1.11 Miscellaneous

GeneScreen plus

DuPont

poly AC/TG probe

Pharmacia

Hybond Nylon membranes

Amersham

## 3.2 Methods

Most of the protocols given below are those routinely used in the Department of Cytogenetics and Molecular Genetics at the WCH and follow, or are based on, those presented in *Molecular Cloning - A Laboratory Manual* (Sambrook *et al.*, 1989), *Current Protocols in Molecular Biology* (Ausubel *et al.*, 1994) and *Current Protocols in Human Genetics* (Dracopoli *et al.*, 1994).

All work mentioned in this and subsequent chapters was performed by the candidate unless specifically indicated otherwise.

### 3.2.1 Propagation of bacteriophage

#### 3.2.1.1 Propagation of *lambda* (Charon 35)

(Sambrook *et al.*, 1989, page 2.118)

Charon 35 *lambda* bacteriophage, from the library LAOXNL01, were screened for (AC)n sequences.

To investigate (by hybridisation - see below) or isolate single phage clones, phage titres were determined by spotting dilutions onto a freshly poured PL plate containing DL538 bacteria, embedded in the agarose, and counting the number of plaques after an overnight incubation. To achieve well separated plaques within the lawn of bacteria, on a 15 cm diameter plate, about 20,000 plaque-forming units (pfu) were used to inoculate 300  $\mu$ l of DL538 (OD<sub>600</sub> at least 0.3). After 20 min. incubation at 37°C this mix was added to 7 ml of top PL agarose at a temperature no greater than 45°C and poured onto pre-dried and pre-warmed 15 cm PL agarose plates. Incubation was for about 16 hours at 37°C. Individual clones were picked from the PL plates using sterile pasteur pipettes. The agarose/phage plug was deposited into 500  $\mu$ l of SM buffer containing chloroform (0.3% v/v) and stored at 4°C.

If necessary, to isolate pure phage clones, more than one round of purification was performed. An aliquot of Charon 35 stored in SM was used to infect 100  $\mu$ l of DL538. This was then poured onto a pre-dried and pre-warmed, 9 cm diameter, PL agarose plate

and incubated overnight at 37°C.

### 3.2.1.2 Isolation of *lambda* DNA.

To obtain sufficient (AC)<sub>n</sub>-positive *lambda* DNA for sub-cloning into M13, the plate lysate method was used (based on Sambrook *et al.*, 1989, page 2.118). To achieve confluent lysis of the lawn of bacteria, 75,000 to 125,000 pfu were used to inoculate 300 µl of DL538. After 20 min. incubation at 37°C this mix was added to 7 ml of top PL agarose at a temperature no greater than 45°C and poured onto pre-dried and pre-warmed 15 cm PL agarose plates. At least two plates were used *per* clone. Incubation was for about 16 hours at 37°C or until confluent lysis had occurred, after which 20 ml of SM buffer was poured onto each plate and left overnight at 4°C. Next day the SM was decanted into a 50 ml tube, to which DNase I (1 µg/ml) and RNase A (10 µg/ml) were added. Incubation was at 37°C for 60 min., after which NaCl (0.0584 g/ml) was added and dissolved. Then PEG 8000 (10% w/v) was added and dissolved, by shaking at 37°C. Tubes were put on ice for several hours or 4°C overnight to precipitate the phage. The phage were pelleted by centrifugation at 17,400 x g, 4°C, for 10 min. The supernatant was gently removed and the pellet resuspended in 500 µl TE and transferred to a 1.5 ml eppendorf tube. A quick spin was used to pellet any remaining debris. 5 µl of 10% (w/v) SDS was added to the supernatant and incubated at 68°C for 5 min. Then 10 µl of 5 M NaCl was added followed by phenol:chloroform:isoamyl-alcohol (25:24:1) extraction (see below) and precipitation of the DNA with ammonium acetate. The DNA was recovered by centrifugation at 12,000 x g for 15 min., 4°C, followed by washing in 70% (v/v) ethanol. After drying, the pellet was dissolved in TE.

### 3.2.1.3 Propagation of M13 bacteriophage

(Sambrook *et al.*, 1989, page 4.22)

The same procedure as described above was used to propagate (AC)<sub>n</sub>-positive M13 phage, except that the host strain was either DH5αF' or XL1-Blue, an incubation time for adsorption was not required and LB plates were used. The bacteria had been grown overnight and could be stored at 4°C for up to several days, prior to infection.

M13 plaques were picked and stored at 4°C in 500 µl of SM buffer (chloroform was not

added). Again, if necessary, a couple of rounds of purification were done to isolate pure M13 clones.

#### 3.2.1.4 Isolation of M13 DNA

To obtain single-stranded M13 DNA for sequencing the following method was used (based on Sambrook *et al.*, 1989 page 4.29). Of an overnight host culture 200  $\mu$ l was inoculated into 2 ml of 2 x YT. To this was added 0.1 x volume of the SM stock of a particular clone. This was incubated for at least 8 hours at 37°C, shaking. The culture was then divided into 2 x 1.5 ml tubes and centrifuged at 15,900 x g for 5 min. The supernatant was decanted and centrifuged again to remove residual bacteria. (A glycerol stock was made using the first bacterial pellet by resuspending it and then adding 80% (v/v) glycerol and storing at -20°C). To the 1 ml (approx.) of cell-free supernatant, 300  $\mu$ l of M13 precipitation solution was added. This was mixed and left at 4°C overnight or 15 min. at room temperature. The phage were pelleted using 15,900 x g for 5 min. All of the supernatant was removed and the pellet resuspended in 100  $\mu$ l of TE. Phenol (50  $\mu$ l) was added and the solution vortexed for 30 sec., then kept at room temperature for 1 min., and vortexed again for 30 sec. This was then centrifuged for 1 min. and the aqueous layer transferred to a new tube. To this was added 50  $\mu$ l of chloroform:isoamyl-alcohol (24:1), followed by vortexing for 30 sec. and centrifugation for 1 min. The aqueous layer was removed to a tube containing a 300  $\mu$ l solution of a 25:1 ratio of absolute ethanol:3 M sodium acetate (pH 5.2). This was mixed and left at room temperature for 15 min., before centrifuging at 12,000 x g for 10 min., 4°C. All supernatant was removed and the pellet washed in 70% (v/v) ethanol. The pellet was then dried at room temperature and resuspended in 30  $\mu$ l of dH<sub>2</sub>O.

#### 3.2.2 Plaque-lifting

Plaque-lifting/blotting, to screen for (AC)n-positive *lambda* or M13 phage, was based on the Benton and Davis (1977) protocol in *Blotting and hybridisation protocols for Hybond membranes* (Amersham). Plaque-lift filters (Hybond N or N<sup>+</sup>, Amersham), were placed onto the plates and permitted to soak through with moisture. The orientation of the filter was recorded by stabbing the filter, and underlying media, with a needle at three points around the circumference. After removing from plates, the filters were layed, for 2 min.,

onto 3mm Whatman paper soaked in denaturing solution, followed by 2 min. on Whatman soaked in neutralising solution, then rinsed in 2 x SSC. After drying, often by using a microwave oven set on 'high' for 40 secs, Hybond N (but not N<sup>+</sup>) filters were exposed to UV radiation using either 1) a light-box and assuming a wavelength of 312 nm, for 3 min. or 2) a BioRad GS Gene Linker UV chamber, to cross-link the DNA.

### 3.2.3 Southern Blotting

Electrophoresis of cut or uncut DNA, or RNA, was performed by adding 10% (v/v) of 10 x loading buffer to the nucleic acid solution, and running (at usually no more than 100 V) through a 0.8% or 2% (w/volume of 1 x TBE) agarose gel. The appropriate DNA size markers were also loaded. After staining in a solution of 0.5  $\mu\text{g/ml}$  ethidium bromide (or incorporating the ethidium bromide into the melted agarose at 0.25  $\mu\text{g/ml}$ ), a photograph of the UV fluorescing gel and scale permitted the determination of the size and quality of the DNA/RNA. An estimate of the quantity of DNA could be obtained by loading several serial dilutions of the sample and comparing the fluorescence with that of the size marker (of known quantity).

Gels were blotted using either one of two methods (based on the methods of Sambrook *et al.*, (1989, page 9.38) or Reed and Mann (1985) respectively):

- 1) Soaking the gel in gel denaturing solution for 30 min., and replacing with gel neutralising solution for 30 min. Avoiding air bubbles, the gel was then placed face down on a transfer tray containing 10 x SSC and overlaid with GeneScreen Plus (DuPont) or Hybond N<sup>+</sup>, cut to size. On top of this was placed 4 sheets of 3mm Whatman paper and then a wad of paper towels. Transfer of the DNA to the filter occurred overnight. In the morning the positions of the wells were marked on the filter, which was then placed in filter denaturing solution for 1 min. and then filter neutralising solution for 1 min. The filter was then air or oven (65°C) dried.

- 2) Soaking the gel in 0.4 M NaOH for 20 min. The rest of the procedure was as described above except that the transfer tray contained 0.4 M NaOH.

### 3.2.4 Radio-labelling DNA

DNA was labelled using the reagents and protocols supplied with the Multi- or Mega-prime kits (Amersham), except for the following: 1) when labelling the poly(AC) probe (Pharmacia), the random hexanucleotide mix could be omitted and 2) when (Megaprime) labelling DNA embedded in LMP agarose slices, the agarose was first melted, an appropriate aliquot removed and added to the labelling reaction, and incubation was for 30 min. rather than the usual 10 min. All reactions were stopped by boiling for 5 min.

### 3.2.5 Pre-reassociation of DNA

To block high copy number repeat sequences, when probing cosmid grids, radio-labelled probes were pre-reassociated. This procedure was based on that given by Sealey *et al.* (1985). In a final volume of 125  $\mu$ l, the following components were mixed: 25 ng of Megaprime-labelled probe (ie entire labelling reaction), 25  $\mu$ l of 20 x SSC, and 337.5  $\mu$ g of sheared human placental DNA. This was then incubated at 100°C for 10 min., put on ice for 1 min., and then incubated at 65°C for 120 min.

### 3.2.6 Hybridisation

Plaque-lifts, Southern blot filters or cosmid grids to be radioactively probed were placed in plastic bags or Hybaid™ bottles and soaked in 5 x SSPE for about 15 min. at 42°C. This solution was replaced with hybridisation solution, air bubbles removed, and pre-hybridised for about 60 min., while shaking/rotating at 42°C. When probing plaque-lifts of poly(AC)-positive phage, there were no more than 10 filters *per* bag and, to avoid blocking of AC repeats, salmon sperm DNA was not a component of the hybridisation solution.

The labelled probe (3.2.4) was denatured by boiling for 10 min. and then added to the bag/bottle of filters. Hybridisation occurred overnight at 42°C, with shaking/rotating.

Washing of the filters (Sambrook *et al.*, 1989, page 9.54) was performed at 65°C, usually with three changes of wash solution I, or wash solution II if more stringent conditions were required. Each wash duration was about 15 min. After washing, the moist filters, sandwiched between plastic, were exposed for the appropriate time to X-OmatK XK-1 Kodak diagnostic film. Exposures were usually done at -70°C with intensifying screens,

although room temperature exposures resulted in higher resolution, which was preferred when exposing sequencing gels (see 3.2.12.1).

Filters were stripped of radioactive probe by shaking at 42°C for 30 min. in 0.4 M NaOH, then replacing this denaturing solution with [0.2 M Tris-HCl (pH 7.5), 0.1 x SSC, 0.1 x SDS]. Alternatively, Hybond (but not GeneScreen) filters were added to a boiling solution of 0.1% (w/v) SDS, simmered for 20 min., and left to cool to room temperature before drying.

### 3.2.7 DNA extraction and purification

#### 3.2.7.1 Phenol/chloroform extractions and DNA precipitation

(Based on Sambrook *et al.*, 1989 page E.3)

Unless otherwise stated, the phenol used was buffered with Tris-HCl (pH 7.5) and “chloroform” refers to a 24:1 solution of chloroform:isoamyl-alcohol.

When purifying DNA the solutions were usually treated by adding an equal volume of phenol, vortexing, then centrifuging for about 1 min., 16,500 x g, room temperature. The supernatant was removed and treated with an equal volume of a 1:1 phenol:chloroform mix as for the phenol alone. The supernatant of this step was similarly treated with an equal volume of chloroform.

DNA was precipitated from the supernatant by adding 0.1 x volume of 3 M Na-acetate (pH 5.2) or 0.33 x volume of 7.5 M ammonium acetate, followed by 2 volumes (DNA solution + acetate solution) of cold ethanol and incubation at -70°C for 30 min., or -20°C overnight. The solution was then centrifuged for 15 min., 16,500 x g. The pellets were washed with 70% ethanol, briefly re-centrifuged and dried (in air or under vacuum). The DNA was resuspended in TE or dH<sub>2</sub>O.

#### 3.2.7.2 Phage DNA (see 3.2.1.2 and 3.2.1.4).

### 3.2.7.3 YAC/Yeast DNA

For glycerol stocks, yeast colonies were grown in 10 ml AHC/tet at 30°C for 48 hrs, centrifuged at 1,000 x g, 4°C for 15 min. The pellets were resuspended and 2 ml of [20% (v/v) glycerol, 80% (v/v) AHC/tet] was added. These stocks were stored at -70°C.

Total yeast DNA was prepared using the method of Sherman *et al.*, (1986), with some modifications:

Yeast cells were grown for 48-72 hrs, shaking, 30°C in 100 ml of AHC/tet. The culture was then centrifuged at 4,500 x g, 4°C, for 5 min. The pellet was resuspended in 3 ml of yeast lysis solution. To this was added 100 µl of lyticase (at 5 mg/ml), followed by a 37°C, 60 min. incubation. The centrifugation was repeated. The pellet was resuspended in a 5 ml solution of [50 mM Tris-HCl (pH 7.4), 20 mM EDTA]. To this was added 500 µl of 10% (w/v) SDS followed by a 65°C, 30 min. incubation. The next addition was 1.5 ml of 5 M potassium acetate. The solution was then incubated on ice for 60 min., followed by centrifuging at 12,000 x g, 4°C, for 10 min. The supernatant was transferred to a new tube, then 2 volumes of room temperature ethanol were added and the solution mixed, then centrifuged at 4,500 x g, room temperature, for 15 min. The pellet was dried and resuspended in 3 ml of TE. This solution was re-centrifuged at 12,000 x g for 15 min. The supernatant was transferred to a new tube and 150 µl RNase (at 1 mg/ml) was added, followed by a 37°C, 30 min. incubation. One volume of isopropanol was added and the solution mixed gently. This was centrifuged at 12,000 x g, room temperature, for 15 min. The pellet was dried and resuspended in 200 µl TE.

To this DNA solution, 100 µl 3 x proteinase K buffer, 4 µl proteinase K (at 10 mg/ml) and 2 µl 10% (w/v) SDS were added followed by an incubation at 37°C for 30 min. After this, a phenol/chloroform extraction was done (3.2.7.1) followed by sodium acetate precipitation. The pellet was resuspended in 100 µl of TE.

A simpler method was employed to prepare YAC DNA for FISH. Yeast cells were grown overnight in 10 ml AHC/tet at 30°C. Next day, 1 ml of this was used to inoculate 150 ml

AHC/tet. This was grown overnight 30°C, then spun at 4,500 x g for 5 min. The pellet was resuspended in 3 ml [0.9 M sorbitol, 0.1 M EDTA]. A few grains of lyticase were added and the solution incubated for 60 min. at 37°C. This was again centrifuged as above. The pellet was resuspended in 5 ml [50 mM Tris-HCl, 20 mM EDTA], to which 0.5 ml of 10% (w/v) SDS was added. Following incubation at 65°C for 30 min., 1.5 ml 5 M KAcetate was added and incubation occurred for 60 min., on ice. This was then centrifuged at 12,000 x g for 10 min. To the supernatant 2 volumes of room temperature ethanol was added and mixed, followed by centrifugation at 3,000 x g for 15 min. The pellet was dried and resuspended in 3 ml TE, by vortexing and incubation at 37°C, followed by centrifugation at 12,000 x g for 15 min. To the supernatant, 150 µl of RNase (at 1 mg/ml) was added and incubation occurred at 37°C for 60 min. One volume of isopropanol was added and mixed, then the solution was spun at 12,000 x g for 30 min. The pellet was dried and resuspended in 0.5 ml TE.

#### 3.2.7.4 Peripheral lymphocyte DNA

For linkage analyses of X-linked diseases, lymphocyte DNA was extracted from whole blood samples from family members.

This method was modified from that given by Wyman and White (1980) and DNA extraction was carried out by Jean Spence. Blood samples were collected in 10 ml EDTA tubes and stored at -20°C. Upon thawing, 20 ml of cell lysis buffer was mixed with the 10 ml of blood in a 50 ml tube, then left on ice for 30 min. The tubes were centrifuged at 2,600 x g, 4°C for 15 min. 20 ml of the supernatant was removed by aspiration. Cell lysis buffer (20 ml) was again added and the centrifugation repeated. The supernatant was removed and 3.25 ml of 3 x proteinase K buffer, 500 µl of 10% (w/v) SDS and 200 µl of proteinase K (at 10 mg/ml) were added and mixed with the pellet. This solution was then mixed overnight at 37°C. Phenol (5 ml) was then mixed into the solution for 15 min., at room temperature. This was followed by centrifugation at 2,600 x g, 15°C for 10 min. The supernatant was transferred to a new tube and the volume made up to 10 ml with phenol and mixed for 10 min. at room temperature. This was then centrifuged at 2,600 x g for 10 min., the supernatant transferred to a new tube and made up to 10 ml with chloroform. The solution was mixed and re-centrifuged. DNA was precipitated by mixing with 300 µl of 3

M sodium acetate (pH 5.2) and adding cold ethanol to 10 ml. After mixing, the tubes were centrifuged and the DNA pellets washed with 70% ethanol, transferred to a 1.5 ml tube, centrifuged briefly at 16,500 x g, then air dried. The pellet was dissolved in 100 µl of TE and quantitated spectrophotometrically.

### 3.2.7.5 PCR products

PCR products were purified using the reagents and protocols of the Prep-a-Gene (BioRad) or Qiaquick (Qiagen) PCR purification kits. If necessary the products from several separate reactions were precipitated with 0.1 x volume of 3M NaAcetate (pH 5.2) and 2 volumes of ethanol and then the resuspended precipitate was cleaned up using Qiaquick tubes.

### 3.2.7.6 Cosmid DNA

Qiagen tip 100s were used to purify cosmid DNA (to be sub-cloned into the exon trapping vector pSPL3) from 150 ml TB/kan overnight cultures, according to the protocol supplied.

A second method involved centrifuging the overnight culture at 4,420 x g for 10 min., 4°C. The pellet was resuspended in 1.2 ml of [50 mM glucose, 50 mM Tris-HCl, 20 mM EDTA, pH 8.0]. To this was added 120 µl of lysozyme (at 40 mg/ml), mixed and incubated at room temperature for 4 min., then 1 min. on ice. Then 2.4 ml of freshly made [0.2M NaOH; 1% (w/v) SDS] was mixed in and incubated on ice for 5 min. To this was added 1.8 ml of ice cold 3M potassium acetate (pH 4.3; see Solutions), which was mixed in by inverting and then incubated on ice for 10 min. Following centrifugation at 27,200 x g for 10 min. the supernatant was removed and 2 x volumes of ethanol were added. After a 5 min. incubation at room temperature, the tubes were spun at 11,500 x g. The pellet was washed in 70% ethanol, vacuum dried, then resuspended in 200 µl of TE10. 10 µl of RNase (at 1 mg/ml) was added and incubation occurred for 30 min., at 37°C. This was followed by a 1 hour incubation at 37°C with [100 µl of 3 x proteinase K buffer; 10 µl of 10% (w/v) SDS; 2 µl of proteinase K (at 10 mg/ml)]. An equal volume of phenol was then added and mixed. Following a 5 min. centrifugation, the top layer was removed and mixed with an equal volume of chloroform. Again, following centrifugation the top layer was removed, to which 0.1 x the volume of 3 M sodium acetate (pH 5.2) was added and 2

volumes of ethanol. The nucleic acid was precipitated by incubating at  $-70^{\circ}\text{C}$  for 30 min. Following a 15 min. centrifugation at  $17,560 \times g$ , the pellet was washed with 70% ethanol, vacuum dried and resuspended in  $100 \mu\text{l}$  of TE.

To remove any contaminating RNA, which could interfere with FISH, the resuspended cosmid DNA was further processed using the reagents and protocol of a RPM kit.

### 3.2.8 Restriction digests

Most restriction digests were performed overnight using 4 x the recommended amount of enzyme for a one hour digest. To avoid inhibition by glycerol, the amount of enzyme added did not exceed 0.1 x the volume of the reaction. Reactions were stopped by freezing or addition of 10 x loading buffer.

### 3.2.9 Dephosphorylation

To prevent re-circularisation of M13 or plasmid vector, cut with a single enzyme, during ligation, the 5' ends were dephosphorylated using calf intestinal alkaline phosphatase (CAP) according to the protocol from Boehringer Mannheim. One unit of CAP was used *per* 50 pmol of vector termini, in a reaction that was incubated for 60 min. at  $37^{\circ}\text{C}$ . After the incubation the reaction was incubated at  $65^{\circ}\text{C}$  for 30 min., to inactivate the enzyme, followed by a phenol/chloroform extraction and precipitation with sodium acetate.

A dephosphorylated vector control was used in a ligation reaction to assess the efficiency of dephosphorylation.

### 3.2.10 Ligations

Usually three ligation reactions of three different (ng) ratios (0.1:1, 1:1, 10:1) of vector:insert DNA were performed *per* sample to be transformed, to increase the chance of success. The reactions were incubated  $16^{\circ}\text{C}$  overnight. The efficiency of the ligation was assessed by ligating SPP-I marker DNA, then running the reaction through a 0.8% (w/v) agarose gel to view the results. Attempted ligation of dephosphorylated M13 vector alone was used as a dephosphorylation control.

### 3.2.11 Transfections/Transformations

#### 3.2.11.1 Transformation of recombinant M13 DNA

Microsatellite-positive *lambda* clones were “shotgun” sub-cloned into M13. Two unequal aliquots from each M13 ligation reaction were added to Tris-HCl (pH 7.5), to a final volume of 50  $\mu$ l. To this ice-cold ligation mix was added 300  $\mu$ l of ice-cold competent bacteria. Incubation on ice was for 10 to 30 min. before plating-out onto 15 cm plates, as described previously except that the top agar contained 0.05 mM glucose, 4 mM X-gal and 1 mM IPTG.

Transformation of 1 ng of uncut M13 and dephosphorylated/ligated M13 were used as competent cell and phosphatase controls respectively.

Cells were made competent using the method of Chung *et al.*, (1989). Bacteria were grown to an OD<sub>600</sub> of 0.3, pelleted (1,000 x g for 15 min. at 4°C) and resuspended in 0.1 x culture volume of ice-cold TSB. The cells were ready to use after a 10 min. incubation on ice. The competent cells could be frozen in liquid nitrogen and stored at -70°C for future use.

#### 3.2.11.2 Transfection of recombinant pSPL3 (Exon Trapping)

Cosmids containing DNA from Xq12-q13 were shot-gun sub-cloned into the exon trapping vector pSPL3, then transfected as follows.

COS-7 eukaryotic cells, purchased from ATCC and passaged by Sharon Lane, were grown to 60%-70% confluence by incubating at 37°C, in COS-7 medium and flushing the (75 cm<sup>2</sup>) flasks with CO<sub>2</sub>. They were then transfected, using LipofectACE and 1  $\mu$ g of the cosmid/pSPL3 DNA.

Transfection was performed according to the protocol from Gibco-BRL: LipofectACE (30  $\mu$ l) was added to 500  $\mu$ l of OMI, gently mixed and left at room temperature for 5 min. To 1  $\mu$ g of the DNA to be transfected, 500  $\mu$ l of OMI was also added. The lipofectACE/OMI mix was added to the DNA/OMI mix, mixed and left at room temperature for 10 min., after

which 4 ml of OMI was added. The COS-7 medium was decanted from the 70% confluent COS cells and replaced with 10 ml of OMI. The cells were gassed with CO<sub>2</sub> and incubated at 37°C for 5 min. The OMI was then poured off and the 5 ml DNA/lipofectACE solution was added and the cells incubated overnight, as described above. The OMI was then replaced with 10 ml of COS-7 medium.

### 3.2.12 DNA Sequencing

#### 3.2.12.1 Single-stranded M13 DNA

Sequencing of (AC)<sub>n</sub>-positive M13 clones was by the dideoxy sequencing reaction, according to the protocol supplied by Perkin-Elmer Cetus, using reagents purchased from Pharmacia (except AmpliTaq). Up to 10  $\mu$ l of M13 single-stranded template solution (3.2.14) was added to 1 x reaction buffer and 5 ng of primer (M13 -20, M13 -40, or a human repeat-specific primer). This mixture was heated for 2 min. at 70°C and cooled to room temperature. Then 2  $\mu$ l of labelling mix, 0.5  $\mu$ l (5  $\mu$ Ci) [ $\alpha$ <sup>32</sup>P]dCTP and 2 units of AmpliTaq were added and incubated at 45°C, for 5 min. Of this labelling reaction, 4  $\mu$ l was transferred to each of four termination mixes: G, A, T and C. These were incubated at 70°C for 5 min. and then cooled to room temperature, after which 4  $\mu$ l of 2 x loading buffer was added. The termination mixes were heated for 2 min. at 90°C before loading 3.5  $\mu$ l onto a 6% (w/v) polyacrylamide/7 M urea, 52 cm x 21 cm gel with wedge spacers and 0.25 mm shark's teeth.

The gels were made using BioRad Sequencing Cell equipment. Before pouring the gel, plates were washed with detergent and rinsed with 100% ethanol. After drying, sigmacoate or gelslick was applied to the back plate to prevent the gel sticking. An acrylamide plug in the base was formed by mixing 30 ml of the acrylamide/urea solution with TEMED and 1.1 M ammonium persulphate in the ratio 142:1:1. This was allowed to set (15 min.) before pouring the bulk of the gel using the same reagents but in the ratio 1000:1:1. Shark's teeth (inverted to form a level edge) were inserted and the gel left to set for at least 2 hours. Gels were pre-electrophoresed (and as a consequence heated) for about 30 min. before loading any samples.

The termination mixes were loaded in the order: GATC and run into the gel at 800 V. The voltage was then increased to a level that kept the gel at a temperature of 50°C, usually

2,000 V. After electrophoresis the gels were blotted onto 3mm Whatman paper and dried under vacuum at 80°C using a BioRad gel-dryer (model 583). For the best resolution autoradiographic exposures were done at room temperature.

To resolve sequencing products larger than about 400 bp an 80 cm x 21 cm, 4% (w/v) polyacrylamide/7 M urea gel was used.

### 3.2.12.2 Double-stranded *lambda* DNA

*Lambda* DNA was sequenced using the *fmol* DNA Sequencing System protocol and reagents except for the polynucleotide kinase (from Pharmacia) and the sequencing enzyme (from Perkin-Elmer Cetus).

Firstly, one primer was end-labelled by mixing 80 ng of this primer with 30  $\mu$ Ci [ $\gamma^{32}$ P]dATP and 5 units of polynucleotide kinase plus buffer in a final reaction volume of 10  $\mu$ l. This was incubated at 37°C for 30 min., then the reaction stopped by incubating at 95°C for 2 min.

Of this labelling reaction, 1.5  $\mu$ l was added to (approx.) 500 ng of *lambda* DNA, buffer and 8 units of AmpliTaq, in a final volume of 17  $\mu$ l. After mixing, 4  $\mu$ l of this mix was added to each of 4 PCR tubes. Each tube contained one of the 4 (G, A, T or C) dideoxy nucleotides in a 2  $\mu$ l solution.

Each tube was then placed in a pre-heated PCR machine and subjected to the following: 95°C 2 min., then [95°C 30 secs, 70°C 30 secs] x 30. When completed 3  $\mu$ l of *fmol* loading buffer was added, heated at 70°C for 2 min. and 3  $\mu$ l loaded onto a 6% sequencing gel as described above (3.2.12.1).

### 3.2.12.3 PCR products

Cycle sequencing was used to sequence double-stranded PCR products using the reagents and conditions specified in the: *Taq DyeDeoxy Terminator Cycle Sequencing Kit* protocol

(Applied Biosystems). The labelled products were then separated and analysed by an Applied Biosystems Model 373A DNA sequencer, according to the methods in the *373A DNA Sequencing System User's Manual 1/92*.

Briefly, 100 ng of PCR product was added to 9.5 µl of the reaction premix and 20 ng of primer. The final volume of 20 µl was overlaid with paraffin. This mix was then subjected to the following thermal cycles: [96°C 30 sec.; 50°C 15 sec.; 60°C 4 min.] x 25. The reaction volume was then increased to 100 µl and the labelled products were extracted twice using 100 µl phenol:H<sub>2</sub>O:chloroform (68:18:14; ABI). The products were then precipitated using 15 µl of 2 M sodium acetate (pH 4.5) and 300 µl of 100% ethanol. Following centrifugation for 15 min. at room temperature the pellets were washed in 70% ethanol, vacuum dried and then stored (in the dark) at -20°C.

The purifying of the PCR products was subsequently simplified, based on the *ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit* (Revision A; August 1995) protocol. The 20 µl reaction volume was added to 2 µl of 3 M sodium acetate (pH 5.2) and 50 µl of 95% ethanol, mixed, and incubated on ice for 10 min, followed by centrifugation at 17,560 x g, for 30 min. The pellet was washed with 70% ethanol, re-centrifuged for 15 min, then vacuum dried.

Denaturing acrylamide gels, used for resolving fluorescently labelled dideoxy-terminated PCR fragments in conjunction with the software of the ABI 373A DNA sequencer, were made, according to the protocol supplied, by mixing 40 g of urea, 12 ml of 40% (w/v) acrylamide:bisacrylamide (19:1), 20 ml dH<sub>2</sub>O and 8 ml of 10 x TBE (pH 8.3). After dissolving with the aid of hot running water, the volume was made up to 80 ml with dH<sub>2</sub>O and the solution was filtered through a Nalgene 0.2 µm filter unit. After cooling to room temperature, 350 µl of freshly made 10% (w/v) potassium persulphate was added and mixed, followed by 45 µl of TEMED. This solution was then poured between two plates, using a syringe. These plates had been prepared by cleaning with a non-fluorescing detergent (Alconox), rinsing with water, separated by 0.4 mm spacers, taped around the

edges and clamped. After pouring in the acrylamide solution, a plastic strip was applied to the top edge (where the shark's teeth were to be positioned) and the gel left to set for at least 2 hours. After setting, the clamps and tape were removed and the plates again cleaned, and the teeth inserted.

The PCR product to be sequenced (3.2.7.5) was resuspended in 3  $\mu$ l of a 5:1 solution of formamide:50 mM EDTA (pH 8.0), heated at 90°C for 2 min. and then loaded onto the gel. The gel was run for 12 hours using 1 x TBE, diluted from a 3 mm Whatman-filtered 10 x TBE stock.

#### 3.2.12.4 Cloned PCR products

Exon-trapped PCR products cloned into Bluescript were subjected to forward and reverse double-strand sequencing using the reagents, including the M13 forward (-21) and reverse primers, and slightly modified protocol of the PRISM dye-primer cycle sequencing kit:

To two PCR tubes, one containing dideoxy-ATP and the other containing dideoxy-CTP, was added 200 ng of template. To two other tubes, one containing dideoxy-GTP and the other containing dideoxy-TTP, 400 ng of template was added. In addition to the dideoxy nucleotide triphosphates all tubes contained dNTPs, buffer, primer and enzyme (pre-mixed by the supplier). The thermal cycling program was: 95°C 30 secs then [55°C 30 secs, 70°C 1 min.] x 15, then [95°C 30 secs, 70°C 1 min.] x 15.

The reactions from each of the four tubes were pooled and added to a tube containing 3  $\mu$ l of 2 M sodium acetate (pH 4.5) and 100  $\mu$ l of 100% ethanol. The tube was left at room temperature for 15 min. before centrifuging at 16,500 x g, room temperature, for 15 min. The pellet was washed with 70% ethanol, re-centrifuged for 5 min., and vacuum dried for 3 min. Dried pellets were stored (in the dark) at -20°C.

The gel and running conditions used were the same as those given in 3.2.12.3.

#### 3.2.13 Sequence analysis

To avoid characterising poly(AC) microsatellite markers previously isolated by other laboratories the sequences flanking each repeat were subjected to a FASTA search of

GenBank (Pearson and Lipman, 1988; Pearson, 1990).

The BLASTN program (email address: blast@ncbi.nlm.nih.gov) was used to search for sequence homology between exon trapped products and anything on the database (Altschul *et al.*, 1990).

The BLASTX program (Gish and States, 1993) was used to search for any sequence homology between predicted amino acid sequence and any protein sequence on the database.

### 3.2.14 Oligonucleotide Synthesis and Purification

Where possible, the design of primers used the suggestions of Lowe *et al.*, (1990). Briefly, the primers should: 1) be 40-60% in GC content; 2) not span a segment greater than 300 bp; 3) be at least 18 bp in length (24 bp preferred); 4) not contain runs of identical bases; 5) possess a GC, CC or CG pair at their 3' end; 6) not contain four contiguous base-pairs of inter-strand nor intra-strand complementarity.

Primers were synthesized using either: 1) an Applied Biosystems Model 391 PCR-MATE EP DNA synthesizer according to the protocol in the *Applied Biosystems User Manual* of May 1989 or 2) a Beckman 1000 DNA synthesizer.

After synthesis, primers were cleaved from their columns and deprotected according to the procedure recommended in the *Applied Biosystems User Manual* (May 1989). The column was connected to a 1 ml syringe which was used to bathe the bound oligonucleotide in 32% w/w NH<sub>4</sub>OH for 20 min. at room temperature, after which the solution was expelled and replaced. This was done three times. The expelled NH<sub>4</sub>OH was made up to 3 ml by the addition of more NH<sub>4</sub>OH. This solution was incubated at 55°C overnight or 70°C for 2 hours. This procedure was slightly modified during the course of the project in that only 1 ml of NH<sub>4</sub>OH was used, with periodic flushing of the same solution through the column using the attached syringe. In addition, 55°C was the preferred incubation temperature because of the tendency of the vapour pressure to explode tubes incubated at 70°C, unless the volume *per* tube was reduced.

Purification was by the n-butanol method of Sawadogo and van Dyke (1991). To the NH<sub>4</sub>OH solution, cooled to room temperature, a 10 x volume of n-butanol was added. This was vortexed for 15 sec., then centrifuged for 45 min. at 2,600 x g, room temperature. The supernatant was discarded and the pellet resuspended in a volume of dH<sub>2</sub>O equal to the original NH<sub>4</sub>OH volume (1 ml). A 10 x volume of n-butanol was again added and processed as before. The pellet was air dried and resuspended in dH<sub>2</sub>O. The concentration of the primers was determined spectrophotometrically (assuming an OD<sub>260</sub> of 1 = 30 µg/ml) and adjusted to 1 µg/µl. The primers were stored at -20°C.

### 3.2.15 PCR conditions

All PCRs were done using a Perkin-Elmer Cetus DNA Thermal Cycler (model 480).

#### 3.2.15.1 Microsatellites

For the amplification of most microsatellite repeats, from anonymous blood bank, CEPH or family DNA samples, the following reagents were used *per* 10 µl reaction volume: 100 ng of genomic DNA sample, 5 µl 2 x PCR himix, 150 ng (≈25 pmol) of each primer and 0.5 units of *Taq* polymerase. For radioactively labelled products 0.5 µl (5 µCi) of [ $\alpha^{32}\text{P}$ ]dCTP was included. MgCl<sub>2</sub> was also added, but its concentration for each pair of primers had to be optimised, although most worked at a final concentration of 6.0 mM. Liquid paraffin was added to each tube and the PCR performed using the thermal cycling program given below. Where possible, taking into consideration PCR product sizes, thermal cycling conditions and MgCl<sub>2</sub> requirements, multiplexing was performed whereby up to four pairs of primers were added to each tube to simultaneously amplify four different loci.

A subsequent discovery made by Kathy Friend was that the PCR also worked using a 7.5 fold lower concentration of deoxynucleotide triphosphates (as routinely used for SSCA). The advantages of this "lomix" relative to the "himix" mentioned above are that: it is less expensive; most reactions work using 1.5 mM MgCl<sub>2</sub> and the thermal cycling program given below; and autoradiographic results from radioactively labelled products are available sooner because of the shorter exposure times required (whether this is because of a higher

specific activity of the PCR product and/or a greater product yield has not been determined). The shorter exposure times enable half the amount of [ $\alpha^{32}\text{P}$ ]dCTP (that is, 2.5  $\mu\text{Ci}$  per tube) to be used. Although most primers work under these conditions, some do not in the sense that no product is seen or the presence of many non-specific products make it difficult to identify the desired polymorphism and reduce the opportunities for multiplexing.

The thermal cycling program most often used was: [94°C 1 min., 60°C 1.5 min., 72°C 1.5 min.] x 10 cycles, then [94°C 1 min., 55°C 1.5 min., 72°C 1.5 min.] x 25 cycles, then 72°C 10 min. This program is referred to as cycling option A in 4.3.4. Other regimes were used and are given in 4.3.4.

Optimisation of  $\text{MgCl}_2$  concentration, when using the himix, was made by performing the PCR with final  $\text{MgCl}_2$  concentrations that varied in 1.0 mM increments usually from 3.0 to 9.0 mM. The reaction products were then electrophoresed through a 1.5% w/v agarose gel along with the pUC19 size marker and stained with ethidium bromide. The  $\text{MgCl}_2$  concentration that resulted in the most intensely fluorescing product was used in subsequent reactions. Other thermal cycling options were investigated if option A failed to yield any product. Improvements in the quality of radio-labelled PCR product (that is, increases in the intensity of the radioactive signal and reduction of non-specific background) could often be achieved by using  $\text{MgCl}_2$  concentrations 0.5 mM more or less than that determined initially.

### 3.2.15.2 *Alu*-PCR

For probing cosmid grids, human-specific products were generated from YACs using inter-*Alu* (“*Alu-Alu*”) or *Alu*-bubble PCRs. The latter technique was developed by Munroe *et al.* (1994) and involves cutting the YAC/yeast DNA and ligating the bubble-adaptor to the restriction fragments. The purpose of the “bubble” in the adaptor is to ensure that first strand synthesis begins from the (human-specific) *Alu* primer, and not from the oligonucleotides ligated to the ends of the restriction fragments. This technique was designed to overcome the problem of large inter-*Alu* distances not amenable to PCR amplification, and thus increase the number of human specific products that could be

obtained from relatively *Alu*-poor regions of the genome.

The *Alu* primers used, in conjunction with the bubble primer, were *Alu* 1, 2, 3, 4, 5 and 6. Various combinations of pairs of these primers were used in conjunction with the bubble primer to determine which “set” gave the optimum results in terms of the number of PCR products and their yield *per* reaction.

*Alu* primers 1 to 6 constitute part of the *Alu* consensus sequence and their sequences (5'-3') are:

*Alu* 1: TGA GCC GAG ATC GCG CCA CTG CAC TCC AGC CTG GG

(includes 3' portion of primer TC-65 from Nelson *et al.*, 1989).

*Alu* 2: AAG TCG CGG CCG CTT GCA GTG AGC CGA GAT

(5' portion of primer TC-65 from Nelson *et al.*, 1989).

*Alu* 3: CGA CCT CGA GAT CTY RGC TCA CTG CAA [Y=T or C; R=A or G]

(primer 517 from Nelson *et al.*, 1989).

*Alu* 4: GGA TTA CAG GCG TGA GCC AC

(includes 3' portion of primer 278 from Nelson *et al.*, 1989).

*Alu* 5: GAT CGC GCC ACT GCA CTC

(*Alu* -3' primer of Tagle and Collins, 1992).

*Alu* 6: CAC TGC ACT CCA GCC TGG GCG AC

(primer A33 from Chumakov *et al.*, 1992).

The following procedure is based on the method by Munroe *et al.* (1994). Firstly, the bubble oligonucleotides were annealed by mixing: 16.5 µg (1 nmol) of bubble-top primer, 17.5 µg (1 nmol) bubble-bottom primer, 222 µl of [3 x SSC, 20 mM Tris-HCl, pH 8.0], dH<sub>2</sub>O to 333 µl. This was boiled for 10 min., then cooled slowly to room temperature. Of this annealing mixture, 16 µl (= 48 pmol of annealed primer) was ligated to 1 µg of YAC restriction digest using 4.5 units of ligase in a 100 µl overnight reaction at 16°C.

The 100 µl *Alu*-bubble PCR consisted of: 10 µl of the ligation reaction, 50 µl of 2 x *Alu*-PCR mix, 1 µM of Not 1-A primer, 1 µM *Alu* primer x, 1 µM *Alu* primer y and 5 units of

*Taq* polymerase. Note: x and y represent 2 of the 6 *Alu* primers. The reaction for inter-*Alu* PCR was the same except that, instead of the ligation reaction, 100 ng of total yeast DNA (uncut) was used and the Not 1-A primer was omitted.

The thermal cycling program was: 94°C, 3 min. then [94°C 15 sec., 72°C 90 sec.] x 30 then 72°C, 5 min.

### 3.2.15.3 Colony PCR

For PCR from the pSPL3 exon trapping vector present within XL1Blue colonies, bacteria were picked using a sterile pipetman tip and inoculated into a PCR tube containing paraffin. Prior to adding the PCR reagents, the tubes were incubated at 100°C for 10 min. The reaction was then performed using the following, in a 50 µl volume: 5 µl of 10 x Boehringer buffer, 1 µl of 10 mM dNTPs, 20 pmol (150 ng) of primer SA5, 20 pmol (125 ng) of primer SD5 and 1.25 units of *Taq* polymerase. The thermal cycling program was: [94°C 30 secs, 60°C 30 secs, 72°C 2 min.] x 30.

### 3.2.16 Electrophoresis

#### 3.2.16.1 Agarose gel

Most agarose gels were made with and run in 1 x TBE buffer, usually at no more than 100 V. Low Melting Point (LMP) agarose gels, used for purifying DNA fragments, were made with and run in 1 x TAE buffer, at 50 V. The electrophoresis apparatus was placed in a 4°C fridge while in operation to keep the buffer cool, thus reducing the chance of the gel melting.

#### 3.2.16.2 Sequencing gels

The gels and running conditions used for DNA sequencing of M13 or PCR products (cloned or uncloned) are given in the appropriate sections above.

### 3.2.16.3 Resolving microsatellite alleles

The gels used to resolve microsatellite alleles, for genotyping, were 42 cm x 50 cm, 0.4 mm thick 5% (w/v) polyacrylamide/7 M urea. In other respects these gels were similar to that used for M13 sequencing (3.2.12.1) except, the urea used was of poorer (cheaper) quality and wells (rather than shark's teeth) were preferred.

To each 10  $\mu$ l PCR, 30  $\mu$ l of 2 x loading buffer was added. These were then heated for 5 min. at 95°C and 5  $\mu$ l was electrophoresed through the 5% gel at 900 V, until the sample had run in, and then the voltage was increased to 2,000. When the run was finished, after about 3 to 4 hours, the gels were blotted and dried as described in 3.2.12.1. Autoradiographs were exposed at -70°C with intensifying screens.

### 3.2.16.4 Non-denaturing (SSCA) gels

To reveal any conformational differences between PCR products from affected and unaffected individuals, non-denaturing gels were used. The gel running equipment used was the same as that used for microsatellite analysis (3.2.16.3).

The conditions for polyacrylamide gel electrophoresis were based on those used by Kauppinen *et al.* (1995). Briefly, 3  $\mu$ l of the PCR was diluted 1:5 in 1% (w/v) SDS/10 mM EDTA followed by a 1:1 dilution (3  $\mu$ l + 3  $\mu$ l) in 95% (v/v) formamide/20 mM EDTA containing 0.05% (w/v) bromophenol blue and 0.05% (w/v) xylene cyanol and mixed. (This dilution process was subsequently altered by simply adding 10  $\mu$ l of 2 x loading buffer to the 10  $\mu$ l PCR). After heating the samples at 100°C for 3 min. they were placed on ice and then 3  $\mu$ l were loaded onto a 5% (w/v) bis/acrylamide (1:60), 5% (v/v) glycerol gel (38 cm x 50 cm x 0.4 mm) and electrophoresis was done using 0.5 x TBE and 600 V at room temperature for 20 hours. An alternative gel used for the analysis of each PCR sample (based on the experience of Savov *et al.*, 1992), consisted of 10% (w/v) bis/acrylamide (1:49) and 10% (v/v) glycerol and was run at 600 V for about 30 hours in 1 x TBE. After electrophoresis the gels were blotted onto filter paper and dried under vacuum. Autoradiography was most often overnight at -70°C, with an intensifying screen.

On occasion, to increase the chance of detecting any mutation, a MDE (“mutation detection enhancement”) gel was used. The gel solution consisted of [37.5% of 2 x concentrated MDE solution (purchased from FMC), 0.6 x TBE]. The running conditions were as for the 10% acrylamide gels, except that the buffer was 0.6 x TBE.

### 3.2.17 Fluorescence *in situ* hybridisation (FISH)

FISH was performed by Elizabeth Baker, Helen Eyre and Paul Yu using *lambda*, YAC and cosmid DNA (Callen *et al.*, 1990). These probes were hybridised to metaphase chromosomes from females. In order to facilitate identification of the X chromosome the metaphase spreads were also hybridised with an X peri-centromeric alphoid repeat (TRX; Choo *et al.*, 1987).

### 3.2.18 RNA Isolation

Where appropriate, to destroy any RNases, solutions (and bottles) were made (or treated) with 0.2% v/v DEPC, incubated at 37°C for several hours and then autoclaved. Where possible, all solutions and manipulations were kept/performed at 4°C, to reduce any RNase activity.

#### 3.2.18.1 RNA isolation from cultured eukaryotic cells

After transfection, with recombinant pSPL3, and incubation for 48 hours (3.2.11.2), the COS-7 cells were rinsed twice with 10 ml of PBS and then scraped off the bottom of the flask using a scraper, in the presence of 10 ml of ice cold PBS. The resuspended cell/PBS solution was transferred to a 10 ml tube and centrifuged at 300 x g, 4°C, for 8 min. All the supernatant was removed and the cell pellet resuspended in 300 µl of ice cold TKM, then left on ice for 5 min. Then 15 µl of NP-40 was added and mixed, then left on ice for 5 min. Centrifugation was at 450 x g for 5 min. After transferring the supernatant to a 1.5 ml tube, 20 µl of 5% (w/v) SDS and 300 µl of phenol were added and mixed. Following centrifugation at 16,500 x g for 5 min., the supernatant was added to 150 µl of phenol and 150 µl chloroform, mixed then re-centrifuged for 3 min. To the supernatant 12 µl of 5 M NaCl and 750 µl of 100% ethanol were added, followed by vortexing and incubation at

-70°C for 30 min. This was followed by centrifuging at 14,000 x g, 4°C, for 30 min. The pellet was washed using 500 µl of ice cold 70% ethanol, centrifuged 14,000 x g, 4°C, for 5 min., and then air dried for 15 min. The RNA was resuspended in 50 to 100 µl of DEPC-treated H<sub>2</sub>O.

Usually 100 to 150 µg of total RNA was obtained from this procedure, as determined spectrophotometrically (assuming an OD<sub>260</sub> of 1 = 40 µg/ml). The RNA was stored at -70°C. To assess the integrity of the RNA 2 µg was electrophoresed through a 1% (w/v) agarose gel.

### 3.2.18.2 RNA isolation from fresh tissue

With the aim of proving that exon trapped products were expressed, RNA was isolated from the following regions of the brain of a normal infant: basal ganglia, brain stem, cerebellum, frontal lobe and occipital lobe. RNA was isolated from 1 g of each region according to the “single-step RNA isolation” protocol of Chomczynski and Sacchi (1987) as given in *Current Protocols in Molecular Biology* (section 4.2.4), with some minor differences, including: the frozen tissue was first crushed to a powder, using a pre-cooled mortar and pestle and frequent application of liquid nitrogen, before adding the RNase denaturing solution. The mortar and pestle had been previously washed with RNase denaturing solution, rinsed with DEPC-treated water, then heat treated for 4 hours at 160-180°C to inactivate any RNases.

After crushing 1 g of tissue, the powder was mixed with 10 ml of RNase denaturing solution and transferred to a 50 ml tube and mixed thoroughly with 1 ml of 2 M sodium acetate (pH 4.5). After this, 10 ml of water-saturated phenol was added and shaken followed by 2 ml of chloroform:isoamyl (49:1) then shaken again. This was incubated on ice for 15 min., then poured into 30 ml autoclaved glass test tubes and centrifuged at 10,900 x g, 4°C, for 20 min. To the supernatant was added 1 volume of isopropanol. The sample was incubated at -20°C for 30 min., then spun again at 10,900 x g, 4°C, for 10 min. The pellet was dissolved in 300 µl of RNase denaturing solution, transferred to a 1.5

ml tube, precipitated with 1 volume of isopropanol for 30 min. at  $-20^{\circ}\text{C}$ , then centrifuged at  $16,500 \times g$ , room temperature, 10 min. The pellet was washed in 70% ethanol, incubated at room temperature for 15 min., re-centrifuged at  $16,500 \times g$ , room temperature, for 15 min. and dried under vacuum. The RNA was dissolved in  $200 \mu\text{l}$  of DEPC- $\text{H}_2\text{O}$  and stored at  $-70^{\circ}\text{C}$ .

The yield of total RNA from 1 g of tissue using this method was usually greater than  $500 \mu\text{g}$ .

### 3.2.19 cDNA synthesis

#### 3.2.19.1 Reverse transcription of exon trapped products

To generate cDNA from pSPL3-transcribed message isolated from transfected COS-7 cells,  $2 \mu\text{g}$  of RNA and  $20 \text{ pmol}$  ( $138 \text{ ng}$ ) of SA2 primer were combined. This was incubated at  $70^{\circ}\text{C}$  for 5 min., followed by 1 min. on ice. Then  $4 \mu\text{l}$  of Superscript II buffer,  $2 \mu\text{l}$  of  $0.1 \text{ M}$  DTT and  $1 \mu\text{l}$  of  $10 \text{ mM}$  dNTP stock solution were added. This was incubated at  $42^{\circ}\text{C}$  for 2 min., prior to adding  $1 \mu\text{l}$  of a 0.1 dilution of RNasin and 200 units of reverse transcriptase. The final volume was  $20 \mu\text{l}$ . Incubation was continued at  $42^{\circ}\text{C}$  for 60 min. The cDNA was stored at  $-20^{\circ}\text{C}$ .

First round PCR ( $100 \mu\text{l}$  reaction):

Reverse transcription reaction ( $10 \mu\text{l}$ ),  $10 \mu\text{l}$  of  $10 \times$  Boehringer buffer (supplied with the polymerase),  $2 \mu\text{l}$  of  $10 \text{ mM}$  dNTP stock solution,  $90 \text{ pmol}$  ( $621 \text{ ng}$ ) primer SA2,  $100 \text{ pmol}$  ( $650 \text{ ng}$ ) primer SD6 were combined and, after heating the tubes at  $94^{\circ}\text{C}$  for 5 min., 2.5 units of *Taq* polymerase were added. The thermal cycling program was:  $94^{\circ}\text{C}$  5 min. (hot start) then [ $94^{\circ}\text{C}$  1 min.,  $60^{\circ}\text{C}$  1 min.,  $72^{\circ}\text{C}$  5 min.]  $\times 6$ .

To prevent the amplification of false-positive (vector-only) products 40 units of *Bst*XI restriction enzyme were added to the PCR and incubated overnight at  $55^{\circ}\text{C}$ . In the morning

10 more units were added and the incubation continued for 60 min.

Second round PCR (100  $\mu$ l reaction):

*Bst*XI-treated 1st PCR (10  $\mu$ l), 10  $\mu$ l of 10 x Boehringer buffer, 2  $\mu$ l of 10 mM dNTP stock, 100 pmol (725 ng) of primer SA4, 100 pmol (695 ng) of primer SD2 were combined and, after heating the tubes at 94°C for 5 min., 2.5 units of *Taq* polymerase were added. The thermal cycling program was: 94°C 5 min. (hot start) then [94°C 1 min., 60°C 1 min., 72°C 3 min.] x 30.

Of this PCR, 10  $\mu$ l was electrophoresed through a 2.0% (w/v) agarose gel. The PCR products were purified from the reactions using phenol/chloroform extraction and precipitated using sodium acetate and ethanol (3.2.7.1).

3.2.19.2 Reverse transcription of RNA isolated from fresh tissue (performed by Dr Jozef Gecz).

To remove DNA contamination of mRNA the following procedure was carried out: 17  $\mu$ l mRNA (6  $\mu$ g), 2  $\mu$ l BRL 10 x buffer and 1  $\mu$ l DNase I (= 1 unit) were mixed and incubated at 37°C for 30 min. The DNase was inactivated by heating at 70°C for 10 min.

The RNA was reverse transcribed using the protocol supplied with the SuperScript II RNase H<sup>-</sup> reverse transcriptase (BRL):

1  $\mu$ g mRNA (or 10-15  $\mu$ g of total RNA)

2  $\mu$ l primer (random and/or dT)

dH<sub>2</sub>O to 24  $\mu$ l

This was incubated at 70°C for 10 min., then mixed with:

8  $\mu$ l 5 x 1st strand buffer

4  $\mu$ l 0.1 M DTT

2  $\mu$ l of 10 mM dNTPs

This was incubated at 37°C for 2 min., before adding 2 µl of reverse transcriptase then continuing the incubation for 60 min. The reaction was terminated by incubating at 95°C for 5 min.

### 3.2.19.3 *rTth* PCR

This technique enables the reverse transcription reaction to occur at 70°C, due to the reverse transcriptase activity of *Thermus thermophilus* DNA polymerase in the presence of Mn<sup>++</sup> ions, thus facilitating the denaturing of any secondary structure of the RNA. The reagents of the *rTth* kit, supplied by Perkin-Elmer, were used which included a positive cDNA control (Interleukin exon primer #152). The following protocol was used:

10.4 µl dH<sub>2</sub>O, 2 µl of 10 x RT buffer, 1 µl (130 ng; 15 pmol) of gene-specific primer, 2 µl RNA, were mixed and incubated for 5 min. at 70°C, then:

2 µl MnCl<sub>2</sub>, 1.6 µl 10 mM dNTPs, 1 µl (2.5 units) *rTth* polymerase, were added and incubated for 30 min. at 70°C.

Then the PCR was performed by adding:

65 µl dH<sub>2</sub>O, 8 µl 10 x chelating buffer, 6 µl MgCl<sub>2</sub>, 1 µl (130 ng; 15 pmol) of second gene-specific primer, and using the following thermal cycles: [94°C 30 sec., 60°C 30 sec., 72°C 2 min.] x 37.

Following the PCR, the sample was electrophoresed through a 2% (w/v) agarose gel.

### 3.2.20 Northern blotting

The procedure used was based on that of Goda and Minton (1995).

The gel teeth, tank and pouring tray were soaked in 1% (w/v) SDS overnight to minimise RNase contamination, and rinsed twice with DEPC-treated dH<sub>2</sub>O. The 1% agarose gel was made with 1 x TBE (from a 5 x borate/EDTA DEPC-treated solution, to which Tris-HCl was added), 5 mM guanidium isothiocyanate and 0.2 µg/ml ethidium bromide. The gel tank was also soaked in 1 M NaOH for 5 min. then rinsed with DEPC-dH<sub>2</sub>O.

To each 15 µg sample of total RNA was added 3.5 x volume of RNA denaturing solution. To this was added 10 µl of RNA loading dye. The samples were run at 60 V for 2.5 hours. The gel was blotted to Hybond N<sup>+</sup> overnight, using a 10 x SSPE transfer solution. The Hybond filter was treated, prehybridised, probed and washed as described previously for Southern blots (3.2.3, 3.2.4, 3.2.6).

### 3.2.21 Linkage Analysis

Two-point linkage analyses were done to regionally localise disease genes segregating within families. Multipoint analyses were done (by Dr Helen Kozman) to generate a genetic map of the X chromosome or to position a disease locus relative to a subset of mapped loci.

The markers used in one or more of the linkage analyses, and their references, are given in Table 3.1.

#### 3.2.21.1 Two-point linkage analysis

To rapidly localise a disease gene, two-point linkage analysis between the disease locus and each marker was performed using the MLINK program of the LINKAGE package, version 5.1 (Lathrop and Lalouel, 1984). The alleles were “down-coded” to facilitate multipoint LINKMAP analysis so that a maximum of 4 alleles, each allocated a frequency of 0.25, were used in the analysis. In addition, unless indicated, the disease was considered to be inherited as an X-linked recessive trait. The disease gene frequency was arbitrarily chosen as 0.0001. Two-point lod scores are presented in the table format recommended by the Committee on Methods of Linkage Analysis and Reporting (Conneally *et al.*, 1985). The regional localisation of the disease gene was delimited by those markers showing recombination with the disease locus that flanked those markers demonstrating close linkage.

#### 3.2.21.2 Multipoint linkage analysis using single (disease) family data

Upon close linkage to one or more markers, multipoint linkage analysis may then be done to reduce the localisation of the disease gene (using the lod-1 principle) and/or provide

greater statistical support for its localisation to the region if no single marker within the regional localisation is fully informative. Multipoint analysis was performed using the LINKMAP program of the LINKAGE program package, version 5.1 (Lathrop *et al.*, 1984). The order and distances between the markers used were those given in the comprehensive map of the X chromosome (Chapter 5; 5.4.1), unless stated otherwise. Due to computer memory limitations, usually no more than 5 loci (4 markers plus the disease locus) could be analysed simultaneously. To incorporate more loci into the multipoint map additional analyses were performed by sliding this 4-marker-window along the map, one marker at a time.

### 3.2.21.3 Multipoint linkage analysis using CEPH data

Multipoint linkage analyses, using genotypic data from the CEPH families, were performed by Dr. Helen Kozman using CRIMAP version 2.4 on a Sun SPARC station IPC (Lander and Green, 1987; Green *et al.*, 1990; Kozman *et al.*, 1993). Briefly, the CRIMAP-BUILD option was used to construct a framework map, beginning with a nucleus of two highly informative markers. Additional markers were sequentially added if the interval support was at least 3 (that is, odds of at least 1000:1 against any alternative position). A comprehensive map was generated by inserting the remaining markers into the framework map using the pairwise linkage analysis algorithm of CRIMAP-ALL.

Support for local order was investigated using the FLIPS $n$  option. This algorithm was used to determine the odds against inversion of adjacent pairs of markers, sliding one locus at a time along the chromosome.

The CHROMPIC option reveals the grandparental origin of alleles and thus any recombinants and was used to flag potential genotyping errors and to construct meiotic breakpoint maps (5.4.4). Genotyping errors were suspected if a double recombination event occurred within 15 cM of any of the markers used in the map. On encountering such events, autoradiographs were re-read and if necessary the genotyping was repeated to eliminate the possibility of sample mix-up. Errors suspected in the genotyping of markers developed by other laboratories were dealt with by removing these genotypes from the analysis. Similarly, the marker genotypes of individuals exhibiting mutations (4.4.2.3) were not included in the analysis.

**Table 3.1 Microsatellite Markers Used in the Linkage Analyses  
(Xpter to Xcen)**

Locus	Marker	Cytogenetic localisation	Reference
DXS1060	AFM205tf2	Xp22.32	Weissenbach <i>et al.</i> , 1992
DXS996	AFM212xe5	Xp22.32	Weissenbach <i>et al.</i> , 1992
DXS237	GMGX9AC	Xp22.32	Gedeon <i>et al.</i> , 1992
KAL	KAL(AC)	Xp22.32	Bouloux <i>et al.</i> , 1991
DXS85	L782	Xp22.31	Chang <i>et al.</i> , 1994
DXS1043	AFM126zd2	Xp22.31	Gyapay <i>et al.</i> , 1994
DXS1224	AFM311vf5	Xp22.31	Gyapay <i>et al.</i> , 1994
DXS16	pSE3.2L	Xp22.31	Chang <i>et al.</i> , 1994
DXS987	AFM120xa9	Xp22.31-22.2	Weissenbach <i>et al.</i> , 1992
DXS207	pPA4B(AC)	Xp22.2	Oudet <i>et al.</i> , 1992b
DXS999	AFM234yf12	Xp22.2	Weissenbach <i>et al.</i> , 1992
DXS443	RX324	Xp22.13	Browne <i>et al.</i> , 1992a
DXS365	RX314	Xp22.13	Browne <i>et al.</i> , 1992a
DXS1052	AFM163yh2	Xp22.13	Rowe <i>et al.</i> , 1994
DXS274	mpA274	Xp22.13	Rowe <i>et al.</i> , 1994
DXS989	AFM135xe7	Xp22.13	Weissenbach <i>et al.</i> , 1992
DXS451	kQST-80	Xp22.13	Browne <i>et al.</i> , 1992a
DYS	3'Dys MS	Xp21.2-21.1	Oudet <i>et al.</i> , 1990
DXS992	AFM184xg5	Xp21.3-21.1	Weissenbach <i>et al.</i> , 1992
DXS997	AFM217xa5	Xp21.3-21.1	Weissenbach <i>et al.</i> , 1992
DYS	5'Dys II	Xp21.1	Feener <i>et al.</i> , 1991
DXS538	pXL27B	Xp21.1-11.21	Browne <i>et al.</i> , 1991
DXS1068	AFM238yc11	Xp11.4	Weissenbach <i>et al.</i> , 1992
MAOA	MAOA	Xp11.4-11.3	Black <i>et al.</i> , 1991
DXS1003	AFM27xf5	Xp11.3-11.23	Weissenbach <i>et al.</i> , 1992
PFC	PFC(AC)	Xp11.3-11.23	Coleman <i>et al.</i> , 1991
DXS1126	EAD	Xp11.23	Donnelly <i>et al.</i> , 1994b
ALAS2	I7CA	Xp11.22-11.21	Cox <i>et al.</i> , 1992
DXS991	AFM15xf6	Xp11.21	Weissenbach <i>et al.</i> , 1992

**Table 3.1 Microsatellite Markers Used in the Linkage Analyses (cont'd)**  
(Xcen to Xqter)

Locus	Marker	Cytogenetic localisation	Reference
AR	AR(CAG)	Xq11.2-12	Donnelly <i>et al.</i> , 1994b
PGK1P1	PGK/5	Xq12	Browne <i>et al.</i> , 1992b
DXS1125	46AD	Xq12	Donnelly <i>et al.</i> , 1994b
DXS1275	AFM261zh5	Xq12	Gyapay <i>et al.</i> , 1994
DXS8111	AFMa099wb9	Xq12	Dib <i>et al.</i> , 1996
DXS453	Mfd66	Xq12-13.1	Weber <i>et al.</i> , 1990
DXS983	AFM078za1	Xq12-13.1	Weissenbach <i>et al.</i> , 1992
DXS7113	4704	Xq13.1	GDB
DXS6673E	(GA) <sub>n</sub>	Xq13.1	van der Maarel <i>et al.</i> , 1996
DXS559	HX28	Xq13.1	Roustan <i>et al.</i> , 1992
DXS1124	45AD	Xq13	Donnelly <i>et al.</i> , 1994b
DXS566	HX60	Xq13.3	Porteous <i>et al.</i> , 1992
DXS986	AFM116xg1	Xq21.1	Weissenbach <i>et al.</i> , 1992
DXS995	AFM207zg5	Xq21.1	Weissenbach <i>et al.</i> , 1992
DXS1002	AFM249vh5	Xq21.2-21.31	Weissenbach <i>et al.</i> , 1992
DXS3	p19.2	Xq21.33	Stanier <i>et al.</i> , 1991
DXS990	AFM136yc7	Xq21.33	Weissenbach <i>et al.</i> , 1992
DXS458	Mfd79	Xq21.33	Weber <i>et al.</i> , 1990
DXS454	Mfd72	Xq22.33	Weber <i>et al.</i> , 1990
DXS1153	21AD	Xq22.1	Donnelly <i>et al.</i> , 1994b
DXS1120	GAD	Xq22.3	Donnelly <i>et al.</i> , 1994b
DXS456	XG30B	Xq22.3-23	Luty <i>et al.</i> , 1990
DXS1072	AFM276wa5	Xq23	Gyapay <i>et al.</i> , 1994
DXS1059	AFM203yd8	Xq23	Gyapay <i>et al.</i> , 1994
DXS8081	AFMc013xd5	Xq23	Dib <i>et al.</i> , 1996
DXS1220	AFM302xc9	Xq23	Dib <i>et al.</i> , 1996
DXS424	XL5A	Xq23	Luty <i>et al.</i> , 1990
DXS8064	AFMb319ya5	Xq23	Dib <i>et al.</i> , 1996
DXS8067	AFMb331yc9	Xq23	Dib <i>et al.</i> , 1996
DXS1001	AFM248we5	Xq24	Weissenbach <i>et al.</i> , 1992

**Table 3.1 Microsatellite Markers Used in the Linkage Analyses (cont'd)**  
(Xcen to Xqter)

Locus	Marker	Cytogenetic localisation	Reference
DXS8059	AFMb303xd1	Xq24	Dib <i>et al.</i> , 1996
DXS1212	AFM280vf5	Xq24	Gyapay <i>et al.</i> , 1994
DXS425	XL90A3	Xq25	Luty <i>et al.</i> , 1990
DXS1114	-	Xq26.1	Weber C <i>et al.</i> , 1993
HPRT	HUMHPRT	Xq26.1	Edwards <i>et al.</i> , 1991
DXS102	cX38.1	Xq26.3	Gedeon <i>et al.</i> , 1992
DXS984	AFM105xc5	Xq27.1	Weissenbach <i>et al.</i> , 1992
DXS292	VK14AC	Xq27.2	Richards <i>et al.</i> , 1991b
DXS297	VK23AC	Xq27.3	Richards <i>et al.</i> , 1991b
DXS548	-	Xq27.3	Verkerk <i>et al.</i> , 1991
DXS1123	41AD	Xq28	Donnelly <i>et al.</i> , 1994b
GABRA3	MGD34-1/2	Xq28	Hicks <i>et al.</i> , 1991
DXS15	p39	Xq28	Wehnert <i>et al.</i> , 1993
DXS1108	sDF-2	Xq28	Freije <i>et al.</i> , 1992

Marker order and cytogenetic localisations were taken from Fig. 5.1 or obtained from the consensus map of Nelson *et al.* (1995), or from the genetic map of Dib *et al.*, (1996).

**Chapter 4**

**Isolation and Characterisation**

**of**

**Dinucleotide Repeat Markers**

## Table of Contents

	Page
<b>4.1 Summary</b>	68
<b>4.2 Introduction</b>	69
<b>4.3 Materials and Methods</b>	71
4.3.1 Identification of poly(AC)-containing <i>lambda</i> phage	71
4.3.2 Shot-gun sub-cloning into M13 phage	71
4.3.3 Sequencing	72
4.3.3.1 M13 Sequencing	72
4.3.3.2 <i>Lambda</i> DNA sequencing	72
4.3.4 PCR conditions	72
4.3.5 CEPH genotyping	73
4.3.6 Characterisation of mutations	74
<b>4.4 Results</b>	76
4.4.1 Isolation and characterisation of microsatellite markers	76
4.4.2 CEPH genotyping	87
4.4.2.1 Cryptic alleles	90
4.4.2.2 Androgen receptor repeats	92
4.4.2.3 Mutations	93
<b>4.5 Discussion</b>	102
4.5.1 Isolation of microsatellite markers	102
4.5.2 Cryptic alleles	104
4.5.3 Androgen receptor repeats	105
4.5.4 Mutations	107
<b>4.6 Conclusion</b>	109

## 4.1 Summary

Eleven microsatellite markers were isolated from a human X chromosome-specific *lambda* library and characterised. Seven new polymorphic markers were genotyped through the forty CEPH families, enabling their incorporation into an entirely PCR-based genetic map of the X chromosome (Chapter 5), as the basis for mapping X-linked disorders. In addition, six polymorphic microsatellites characterised by others were also genotyped through the CEPH families for the purpose of incorporation into the map. During the genotyping, a number of mutations were found. The published version of the work presented in this chapter (Donnelly *et al.*, 1994b) is given in appendix A.1.

## 4.2 Introduction

A polymorphism is described as a Mendelian trait that has two or more alleles, with the most common allele having a population frequency of less than 95% (Ott, 1991). Construction of a genetic map requires polymorphic markers which are usually anonymous single-copy neutral segments of DNA. Such segments upon characterisation are allocated a "D" number by the Genome Data Base (GDB) (Williamson *et al.*, 1991). For example, X-linked DNA segments are designated "DXS#" where D refers to a segment of DNA, X refers to the segment being X-linked and S referring to it being a single copy. Variations of this prefix occur to designate DNA segments that belong to a family of loci (DXF#) or are aliphoid repeat sequences (DXZ#). Coding sequences are designated by the suffix E (DXS#E) or an acronym of the name of the gene to which they belong (for example, AR for the androgen receptor gene).

As mentioned in Chapter 2, microsatellites are now the preferred markers for genetic mapping because of their high heterozygosity, ease of analysis and ubiquity. These characteristics have also made them useful in other fields such as population genetics, forensic science and paternity testing. In addition, the specific detection and isolation of trinucleotide microsatellites may also have direct diagnostic benefits where they are the basis for unstable triplet repeat disorders (Richards and Sutherland, 1994).

PolyAC repeat markers with more than 13 units are usually polymorphic and the greater the length of a repeat the greater the instability, although most are composed of less than 26 units (Weber, 1990). The length variation between alleles of a given microsatellite is thought to be generated by slipped strand mispairing during replication or unequal crossover between sister chromatids (Levinson and Gutman, 1987; Schlotterer and Tautz, 1992).

The ability of alternating purine/pyrimidine sequences, such as polyAC repeats, to adopt a Z-DNA conformation *in vitro* (Nordheim and Rich, 1983), has led to suggestions that these sequences may serve some function *in vivo*, by influencing transcription rates of adjacent genes or promoting recombination (Tautz and Renz, 1984; Miller *et al.*, 1985). Other evidence cited in support of a function is: polyCT repeats are involved in the regulation of heat shock genes in *Drosophila melanogaster* (Lu *et al.*, 1993); the degree of conservation between primates, in terms of repeat structure and position, is higher than

expected given accepted times of divergence (Deka *et al.*, 1994); and proteins that bind to microsatellite sequences have been discovered (Yee *et al.*, 1991; Aharoni *et al.*, 1993; Richards *et al.*, 1993). An alternative point of view, however, is that microsatellites exist, and are so widespread, because they are under no selective pressure. In summary, in some cases microsatellites may serve a specific function but in the absence of such evidence they may merely represent DNA of no known function.

### 4.3 Materials and Methods

Only brief accounts of the methods used are given below. For more detail refer to the relevant Chapter 3 sections cited.

#### 4.3.1 Identification of poly(AC)-containing *lambda* phage

A *lambda* bacteriophage library (LAOXNLO1) was purchased from American Type Culture Collection (ATCC). It had been constructed from flow-sorted chromosomes from cell line 1635, the cells of which possess five X chromosomes. Contaminating autosomal fragments are present at a level of 10% (Luty *et al.*, 1990).

The library was plated-out (3.2.1.1) and plaque-lift filters (3.2.2) probed with a radioactively labelled polyAC sequence (3.2.4; 3.2.6). PolyAC-positive *lambda* phage were picked, purified, propagated and their DNA isolated as described in 3.2.1.1 and 3.2.1.2.

#### 4.3.2 Shot-gun sub-cloning into M13 phage

An aliquot of DNA from individual *lambda* clones was cut with the restriction enzymes: *AluI*, *HaeIII*, *RsaI* and *Sau3A*, in separate reactions. These digests were electrophoresed and Southern blotted. The filters were probed with the polyAC sequence to determine which restriction fragment(s) contained an AC repeat. The restriction enzyme that resulted in polyAC-positive fragments of about 500 bp or less were used to cut the remaining phage DNA.

M13mp18 DNA was digested with an enzyme that generated the appropriate site for insertion of the cut *lambda* DNA. In the case of *Sau3A*, *BamHI* was used to produce the complementary (staggered) cloning site. For the other three enzymes *SmaI* was used to generate blunt ends. Three ligation reactions of three different ratios of M13:insert DNA were performed, to increase the chance of successful ligations. The recombinant M13s were transformed as described in 3.2.11.1. PolyAC-positive M13 clones were selected as was done for the *lambda* clones. M13 phage were propagated (3.2.1.3), screened for AC repeats (as for *lambda*; 4.3.1) and the DNA isolated (3.2.1.4).

### 4.3.3 Sequencing

#### 4.3.3.1 M13 Sequencing

Single-stranded DNA, from polyAC-positive M13 clones, was sequenced (3.2.12.1) and where possible, primers flanking the repeat sequence were designed (3.2.14).

A GenBank database (FASTA; 3.2.13) search was performed to ensure that none of the polymorphic microsatellites had been previously isolated.

The AC (or complementary TG) repeats were categorised as: perfect, imperfect or compound (Weber, 1990).

#### 4.3.3.2 *Lambda* DNA sequencing

The *lambda* DNA of clone 21 was sequenced using the *fmol* DNA Sequencing System protocol (3.2.12.2).

The radiolabelled products were loaded onto a 6% sequencing gel as described in 3.2.12.1.

### 4.3.4 PCR conditions

Most PCRs used in this chapter used the “himix” reagents (3.1.2) and the appropriate MgCl<sub>2</sub> concentration. For the amplification of most repeats thermal cycling option A was sufficient, however other thermal cycling programs needed to be implemented in a few cases. These thermal cycling options are:

A. {94°C 1 min., 60°C 1.5 min., 72°C 1.5 min} x 10 cycles, then {94°C 1 min., 55°C 1.5 min., 72°C 1.5 min} x 25 cycles, then 72°C 10 min.

B. {55°C 2 min., 72°C 3 min., 94°C 2 min} x 25 cycles, then 55°C 2 min., then 72°C 10 min.

C. {50°C 2 min., 72°C 3 min., 94°C 2 min} x 25 cycles, then 55°C 2 min., then 72°C 10 min.

D. Amplification of the androgen receptor (GGN)<sub>n</sub> repeat (AR<sub>2</sub>), where N=T or C, required the reaction components and thermal cycling (annealing temperature of 60°C) of Riggins *et al.* (1992), except that 150 ng of each primer, 5 μCi [ $\alpha^{32}$ P]dCTP, and 0.5 units *Taq* polymerase were used in a reaction volume of 10 μl. The forward and reverse primers used for the amplification of the AR(CAG) and AR(GGN) repeats were designed by Dr Rob Richards (WCH) from the sequence given in Tilley *et al.* (1989), and are:

AR(CAG) 5'-TCCAGAATCTGTTCCAGAGCGTGC-3' and

5'-GCTGTGAAGGTTGCTGTTTCCTCAT-3'

AR(GGN) 5'-ACAGCCGAAGAAGGCCAGTTGTAT-3' and

5'-CAGGTGCGGTGAAGTCGCTTTCCT-3'.

These two repeats are within exon 1 of the AR gene and are separated by 1.1 kb (Lubahn *et al.* 1989; Tilley *et al.* 1989).

#### 4.3.5 CEPH genotyping

Microsatellite alleles were amplified from the DNA (3.2.15.1) of each of the 40 CEPH mothers (see note below) and 40 fathers. The alleles were detected as described in 3.2.16.3.

The observed heterozygosity of the newly characterised microsatellite markers and markers previously characterised in the Department of Cytogenetics and Molecular Genetics was calculated from the mothers' genotypes, according to the formula:

$$\frac{\# \text{ heterozygotes}}{\#(\text{homozygotes} + \text{heterozygotes})}$$

It is often claimed that the standard CEPH collection of families consists of 40 unrelated mothers, however, mother 104/02 is a daughter of 102/02 and so her genotype depends on the genotypes of her parents (102/02 and 102/01), which are in turn dependent on the allele frequencies. So, because the chance of 104/02 being a homozygote or heterozygote is not independent of the genotype of 102/02 (who has been incorporated into the calculation) she was omitted from the heterozygosity calculation.

The expected heterozygosity of any marker was calculated using the formula :

$$1 - \sum(p_n^2)$$

where  $p_n$  is the frequency of the  $n$ th (1, 2,... $n$ ) allele. Allele frequencies were calculated from the 113 X chromosomes of the 74 unrelated CEPH parents, assuming Hardy-Weinberg equilibrium. Again the alleles of the CEPH mother 104/02 and those of the CEPH fathers; 1375/01, 13291/01, 13292/01, 13293/01 and 13294/01 were not included because they are related to people (parents or sisters) whose genotypes were already included in the calculation.

Statistical comparisons between observed and expected heterozygosities were performed using Chi-square analysis with Yates' correction factor:

$$\frac{(\text{observed heterozygosity} - \text{expected} - 0.5)^2}{\text{expected heterozygosity}} \quad (1 \text{ d.f.})$$

Allele sizes were determined relative to the product amplified from the M13 clone from which the sequence was derived. The allele sizes given in Table 4.4 were derived from the references cited.

Whole CEPH families were genotyped for the markers in Tables 4.3 and 4.4 if the mother was heterozygous. This data was then used in construction of the genetic map (by Dr Helen Kozman; Chapter 5).

#### 4.3.6 Characterisation of mutations

Sexing of DNA samples was done by using primers that amplified an X-specific amelogenin sequence as well as an amelogenin-like Y-specific sequence (Nakahori *et al.*, 1991).

Mutation rates were calculated by dividing the number of mutated alleles by the total number of X chromosomes in which it would be possible to detect such events, that is, chromosomes of individuals whose parents had also been genotyped.

Phase analysis was used to determine the parental and/or allelic origin of the mutations in those cases where this data was not obvious. This involved determining the alleles of the locus in question and that of the closest known informative flanking loci in each individual possessing the mutation and their siblings or children. The most common haplotype in the family concerned was considered as the non-recombinant phase, thus establishing the most probable chromosomal origin of the mutated allele. In most cases the genetic distance from the mutated locus to the nearest informative flanking loci was < 10 cM, hence the chance of an undetected double crossover occurring was negligible.

## 4.4 Results

### 4.4.1 Isolation and characterisation of microsatellite markers

Initially, 264 poly(AC)-containing Charon 35 (*lambda*) phage plaques were picked from 10 x 15 cm plates onto which the X library, LAOXNL01, had been plated. DNA preparations were done for 72 of these phage plaques. Forty eight of these 72 clones, when cut with restriction enzymes, yielded an AC-positive fragment of about 500 bp or less. The stage in development that each of these 48 clones reached is given in Table 4.1. The overall success rate at each step relative to the original number of 72 clones is given in Table 4.2. Of the 48 *lambda* clones, 34 were successfully subcloned into M13mp18, in the sense that AC-positive M13 phage resulted. Sequence ladders were produced from 28 of these 34 M13 clones. The quality of the sequencing results enabled both primers flanking the repeat sequence to be designed in 15 cases (see for example Fig. 4.1). Eleven of these 15 repeats could be amplified from genomic DNA by PCR, however only 7 were polymorphic (see for example Fig. 4.2).

Most of those clones given in Table 4.1 that did not progress to the sequencing stage fall into three categories. Firstly there are the O-like clones (O, P, Q and S) for which relatively low yields of *lambda* DNA were obtained, as judged from ethidium bromide stained agarose gels and hybridisation of capillary blots with the poly(AC) probe. No attempt was made to sub-clone this DNA into M13.

Secondly, the B-like clones (B, C, L, 12, 13, 15, 17, 39, 43 and 50) were *lambda* clones whose AC repeat was not successfully sub-cloned into M13.

The third category is that of the H-like clones (H, X, 11, 25, 27 and 69). These were subcloned into M13 but plaque purification or amplification of the M13 DNA was difficult. Specifically, plaques that, because of the number of rounds of purification, were expected to be AC-positive were not, or sometimes no plaques were produced at all. Attempts at generating M13 DNA for sequencing yielded no DNA as determined by the lack of ethidium bromide fluorescence or no AC-positive DNA as revealed by the failure to hybridise with the polyAC probe to capillary blot filters. Single-stranded M13 DNA was prepared from clone X but upon sequencing only a compacted band (no readable ladder) at the top of the sequencing gel was observed.

**Table 4.1 Stage to which AC markers have been developed**

λ clone	Enzyme used	Subcloned into M13	Transformed into <i>E.coli</i>	M13 ss template	Sequenced	Primers designed		Comments
						F	R	
A	S	+	+	+	+	+	+	Cryptic repeat; not polymorphic
B	R	-						No AC+ve M13 clones
C	R	-	-	-	-	-	-	No AC+ve M13 clones
D	S	+	+	+	+	+	-	No consensus sequence obtained
E	S	+	+	+	+	+	+	Polymorphic(TC) <sup>17</sup> (TG) <sup>18</sup>
F	R	+	+	+	+	+	-	Same sequence as 41
G	H	+	+	+	+	+	+	Polymorphic (AC) <sup>23</sup>
H	R	+	+	-				M13 refractory to amplification
I	A	+	+	+	+	+	+	(GT) <sup>25</sup> ; no PCR product
K	H	+	+	+	+	+	+	(TG) <sup>19</sup> ; no PCR product
L	A	-						No AC+ve M13 clones
M	H	+	+	+	+	-	-	(AC) <sup>22</sup> too close to cloning site
O	S	-						Low <i>lambda</i> DNA yield
P	S	-						Low <i>lambda</i> DNA yield
Q	A	-						Low <i>lambda</i> DNA yield
R	H	+	+	+	+	-	-	(AC) <sup>15</sup> but sequence mostly unreadable
S	A	-						Low <i>lambda</i> DNA yield
T	A	+	+	+	+	-	-	No repeat seen
W	R	+	+	+	+	+	+	(TG) <sup>21</sup> ; not polymorphic
X	R	+	+	+	+	-	-	No sequence obtained
11	H	+	+	-				M13 refractory to amplification
12	H	-						No AC+ve M13 clones
13	R	-						No AC+ve M13 clones
15	H	-						No AC+ve M13 clones

+ stage completed; - stage not completed; A *AluI*; H *HaeIII*; R *RsaI*; S *Sau3A*; ss single stranded; F forward primer; R reverse primer.

**Table 4.1 Stage to which AC markers have been developed (cont'd)**

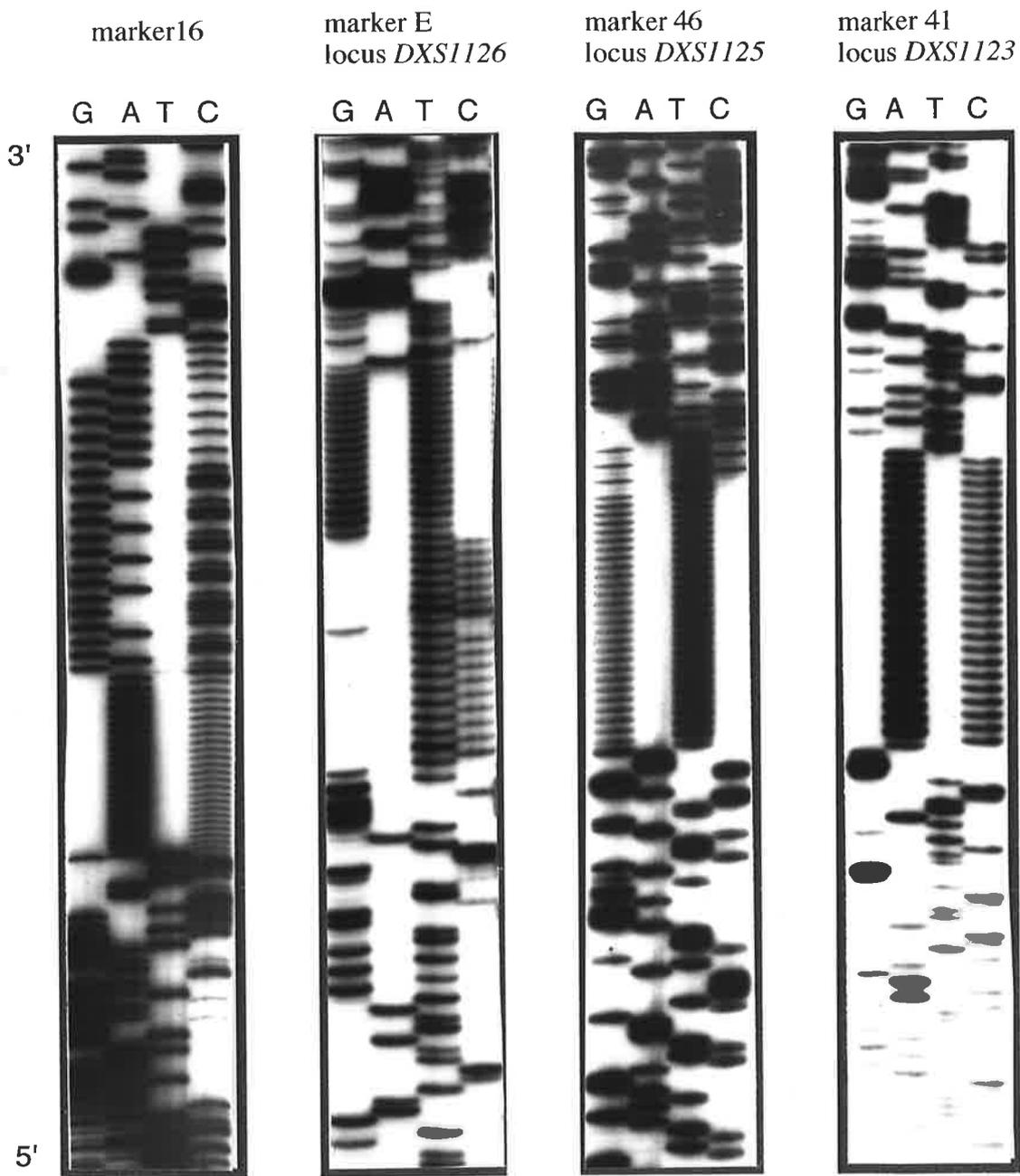
λ clone	Enzyme used	Subcloned into M13	Transformed into <i>E.coli</i>	M13 ss template	Sequenced	Primers designed		Comments
						F	R	
16	H	+	+	+	+	+	+	cryptic repeat; not polymorphic
17	H	-						No AC+ve M13 clones
20	H	+	+	+	+	-	-	(AC)>21 but sequence unreadable
21	H	+	+	+	+	+	+	(TG)22; polymorphic but cryptic alleles; cf DXS178
25	H	+	+	-				Poor template preparation
26	H	+	+	+	+	+	+	(TG)24; polymorphic
27	H	+	+	-				Poor template preparation
28	H	+	+	+	+	-	-	(TG)15 but sequence unreadable
29	S	+	+	+	+	+	+	Not polymorphic
31	H	+	+	+	+	-	-	No repeat seen
32	S	+	+	+	+	+	+	(TG)17(TGG)5 but no PCR product
39	H	-						No AC+ve M13 clones
41	S	+	+	+	+	+	+	(CA)24; polymorphic
42	S	+	+	+	+	+	+	(GT)20; no product
43	S	-						No AC+ve M13 clones
45	S	+	+	+	+	+	+	(TG)6TT(TG)17; polymorphic
46	S	+	+	+	+	+	+	(GT)28; polymorphic
47	S	+	+	+	+	-	-	Sequence faint/unreadable
48	S	+	+	+	+	-	-	Sequence unreadable due to cross-banding
49	S	+	+	+	+	+	-	(GT)21; sequence 3' of repeat is AT-rich
50	H	-						No AC+ve M13 clones
52	S	+	+	+	+	-	-	No repeat seen
67	S	+	+	+	+	+	-	(GT)12T(GT)4; sequence 3' of repeat is AT-rich
69	S	+	+	-				M13 refractory to amplification

+ stage completed; - stage not completed; A *AluI*; H *HaeIII*; R *RsaI*; S *Sau3A*; ss single stranded; F forward primer; R reverse primer.

**Table 4.2 Development of AC repeat markers isolated from a flow-sorted X *lambda* library**

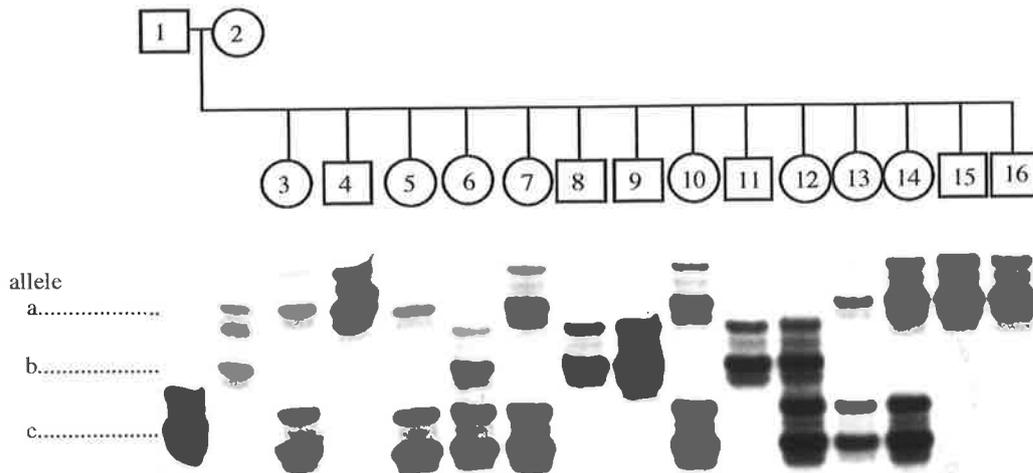
Stage of development	Numbers obtained	% recovery <sup>a</sup>
1. AC-positive <i>lambda</i> clones cut with restriction enzymes	72	----
2. Clones yielding AC-positive restriction fragments <500 bp	48	66.7
3. AC-positive M13 subclones	34	47.2
4. AC-positive M13 clones successfully sequenced	28	38.8
5. Sequence suitable to design primers flanking the repeat	15	20.8
6. PCR product obtained	11	15.3
7. Polymorphic markers	7	9.7

a. relative to step 1.



**Figure 4.1** Examples of the results of dideoxy sequencing of microsatellite-containing M13 phage. The order of loading is given immediately above each sequence.

CEPH pedigree 102



**Figure 4.2** Autoradiograph showing the polymorphic nature of marker 26 (locus *DXS1126*). The most intense band(s) in each track is considered to be the allele. For example, the mother 102/02 has the genotype ab and her daughter, 102/03, is ac. The fainter "stutter" bands associated with each allele are typical of dinucleotide repeats and are thought to result from 2 base-pair strand slippage during the PCR (Hauge and Litt, 1993; Murray *et al.*, 1993).

Of those clones that underwent sequencing, the clarity of the sequencing results from clones R, X (see above), 20, 28, 47, 48 and 67 was not adequate to design one or both primers flanking the repeat. Twenty eight clones were successfully sequenced in that a readable ladder was produced. No repeat sequence was seen in the sequence from clones T, 31 and 52. In a number of cases the repeat was too far from the M13 sequencing primer to enable a reverse primer to be designed so a new forward primer was made to use as a sequencing primer and thus “walk” the sequence. Using this method the sequence flanking the repeat of clone F was revealed to be the same as that previously shown for clone 41. Also, using new forward primers on clones 49 and 67 revealed that the sequence beyond the repeat was not suitable for designing a reverse primer because it was AT-rich.

Instability of a few of the M13 clones seemed to occur. Apart from the H-like clones mentioned above, there were problems of instability with clones D, E and M. The forward primer of clone D failed to work on a new M13 template preparation of this clone. Using the forward M13 sequencing primer on this new template revealed that the sequences flanking the repeat bore no resemblance to the sequence initially revealed (and used to design the forward primer). The repeat from the new template preparation is also slightly different from the original, but still distinctive. This phenomenon of sequence differences between clones that were expected to be identical was also encountered for clones E and M. The repeat of clone M was too close to the cloning site to design a forward primer.

Primers flanking the repeat sequences were designed for clones A, E, G, I, K, W, 16, 21, 26, 29, 32, 41, 42, 45 and 46. No PCR product from genomic DNA was produced using the primers designed from clones I, K, 32 and 42. In the case of the latter 2 markers this was probably due to the fact that there was an (overlooked) *Sau3A* restriction site between the repeat and one of the flanking primers, hence these two clones are most likely chimaeric.

The sequences surrounding those 11 repeats that can be amplified from genomic DNA are given in Figure 4.3. However 4 of these (A, W, 16 and 29) are not polymorphic (based on CEPH genotyping - see below).

**Figure 4.3** Sequencing results of markers that can be amplified by PCR.

The following are sequences (5'-3') derived from the M13 clones containing markers suitable for STSs, that is those markers that could be amplified by PCR. Repeat sequences with  $n > 4$  (including runs of mononucleotides) are bracketed and in bold (except in the case of marker A, for the sake of simplicity). Where possible, the repeat sequences are classified according to Weber (1990). The sequences used for PCR primers are underlined. Bases that could not be identified with confidence are indicated in lower case or by a "N".

**Figure 4.3.1** Marker W

CTGAACAGGAAACAGAGTGAGCTGGGCTTGACCAGGACACATGGCAATGGAG  
AAAAGCAGATGTGAGATGCTTATAGTGGTACTTG[GT]<sub>21</sub>GGTgTgGcATGTAAAT  
GCAGAGG[A]<sub>5</sub>GcTCTGAAATTAACACTCAGAACTGCCCTCAGTAGCTCACATC  
TGGGGAGAGAGGAGGTAGCTACTGTTCTATCGAGAGAATACCTGATAATACC

The repeat is perfect.

**Figure 4.3.2** Marker E (*DXS1126*)

AGTCGGGGCGCGCGTAACCGTGCCCTGCCCCACCTCTCCGTGCCCCGGTGCC  
TGTGACGCTGTCCACGGCACTAAGGCGCTTTAAGAAAGGGCGCGCGCGATCA  
GTGTGTGTGCGTGTGATGTGAGGAGAGCGGTGTGGAGGTTGTGTCCTTGAAA  
GTGTGTGTTCTTTGGTTTAGCGTTCAAGTTTCTAGAAAGGTGCGTGTCTGCTGG  
TGTGTAGTGTGTGTGATACGATATTTTGGGATTTTGTGGCTGCATGTTTTAGGT  
ACACTTGAATTGATCTATATTTTGTGCTATGGC[CT]<sub>x</sub>[TG]<sub>y</sub>

The sequence underlined above was then used to make a sequencing primer to reveal the repeat sequence and additional 3' sequence to enable the design of the reverse primer:

GGGATACTCTGTGCGCGATTCTTTGTGCGTCCGATTGTTTTCTCAGTGACACCG  
TGTGGATATTTGTGAATCCTTATTATGTGTGTTGGCTTGGCCTATGGGCTGTG[T  
C]<sub>17</sub>[TG]<sub>18</sub>ATTTCGTTTGTGTTTGAGAGGGAATGGTCTtCgCATCTCCGTG

Curiously, the sequence between the forward sequencing primer and the repeat of this second sequence bears little resemblance to that in the first sequence (see Discussion); nevertheless a PCR product is obtained using the underlined primers. The size of the PCR product from the M13 clone used for sequencing is based on the sum of the bases comprising the PCR primers and the intervening sequence given in the second sequence above. Comparison with other PCR products of known size confirmed this estimate of product size. The repeat is classified as “compound”.

Figure 4.3.3 Marker 16

CTAGGCAGGCCAGGCCAGGAGAAACACAGGACTGAATTCCATTTAGTTGGTTC  
TGGgCTGCAGATAACCaGAGGGAAGTGTG[TGC]<sub>7</sub>GGCTGCGGCTGCGGCTGCG  
GCTGCGGCGGCTGCGGCTGC[GT]<sub>24</sub>CAAGGGGAGATTGTTGGGGAGCGGGCG  
AGGCCCATCCCTGCTTGGGTCTCCCTACCGCTCCTCCTG[C]<sub>7</sub>ACCCCTTTATTT  
TCTCTCCCTTCCACTCCITTCCTCTCTCGGCCAGTATCCGGATCCGGAGTC

Note that the sequence between the [TGC] and [GT] repeats also has a repetitive structure [GGCTGC] hence this is a cryptic repeat, as is A (Fig. 4.3.10). It does not qualify under Weber’s rules to be designated as a compound repeat because the uninterrupted repeat unit copy number (n) of [GGCTGC] is less than 5.

Figure 4.3.4 Marker 46 (*DXS1125*)

CCTCTGGGCCCCACAAGTCAGGGGCAGTGCAGCTCAGCCACCAGCCAGCAGC  
CCTGCCCTAGGCTCTAGATAGGACTCTAAAGCTCCCcATGCITTTAGGGAGGTGA  
GCTTTCAGGTTcCAGGcCAA[GT]<sub>28</sub>TTGTCTGTCTGTCTTACATCCTAGCAGTA  
GAGCACTCCCAGAGCCATGATCTCTTTCATCGGGCGGTTAAGGTCAACTTTCAC  
CGCTTCGCCTTACGGCTgGCGCAGTTCTTTGGATGTACTGCTgGTGTGTcCAGT  
TTCG

The repeat is classified as perfect.

Figure 4.3.5 Marker 45 (*DXS1124*)

CCTGGTAAGGCCAGCTCCAAAGCAGACTGCCTGTGAGAAGCCTCTGTTTACC  
TCAGATGGAGTCCTGGgCACCTTCCTCCTTTGGCCCACTATGCCTTTCTCCTAAC  
CACTTACTTGGTAAGAGAA[TG]<sub>6</sub>TT[TG]<sub>1</sub>7TAATGAAATATTTcCCCTCAATATAT  
GCTTCAGGTAAAAaCTTTCAGATAATACAGATAATGGTTCCCcTCATATCCcTATA  
CTCAGTGAC

The repeat is classified as imperfect.

Figure 4.3.6 Marker 21 (*DXS1153*)

CCTGAGGATGGCTTGAAGGTG[TG]<sub>21</sub>ATGTGAGTGTTCCCTGGTgGGCAAGCTA  
ATTCCGAAGgTTGCAAAACACCCTGAGACCCTGGGTGTATTGTGTGATGTGAGT  
{GGTACCcGAGCTCGAATTCgTAATCATGG}

The sequence bounded by the curly brackets is homologous to that within the *DXS178* locus (GenBank accession no. X65235). Note also that the forward primer includes part of the perfect [TG]<sub>22</sub> repeat in its sequence. In an attempt to overcome the cryptic allele problem by using another primer, additional sequence 5' to the forward primer was obtained by sequencing the *lambda* template from which the M13 subclone was generated, using the reverse primer:

GACGGGGAGACTGAGAAGCAGCTGTGCTAGGGaCCATGAGGGATGGCTTGAG  
[TG]<sub>22</sub>

The new forward primer is underlined and the old primer (at least that part of the sequence which most closely resembles it) is crossed-out.

Figure 4.3.7 Marker G (*DXS1120*)

GGATTCTAAGACCCTCTCACACCAGAACTCAATCTAAATTATTTCCCCCA[AC]<sub>23</sub>  
GATGGTTATCTTTCCCTGAATAAAACTGAATGGTTTGA[T]<sub>5</sub>GACCGTGGCAGT  
CCGATCCCAAACGGATGTTTCAGGTCATAAGCCTAGATTCAGTGTGCCGTCTGG

The repeat is perfect.

Figure 4.3.8 Marker 26 (*DXS1122*)

CTGGGTTGCTGTTGAAGTTTTCCCATCTGTTACACTGAAAGGACTGCACCAGGTT  
TT[**TG**]<sub>24</sub>TTtGGTCATGGAGCTCCTATTCAGGATTGGCTCTTGAAtTTTTGCGGGC  
CGAGCTCGAATTCGCaATCATGTCATAGCTGTTTC

The repeat is perfect.

Figure 4.3.9 Marker 41 (*DXS1123*)

CCcTGCCTAAATGTTTCGCAAGCCCATTCCCTATTTCCAGAGAGTTCACAGACTCC  
ATTCCTGGTTCTGTATTCCTGGGgA[**CA**]<sub>24</sub>TTTGTATGATTCCTGTATGCTTAAG  
GGGTTTCTAGGAGGCAGCTGTTTGTtATTTGGGgAGAGTGTGGATGAACTTA  
GAAGCATGGGCTATAGGGG

The repeat is perfect.

Figure 4.3.10 Marker A

GCAGACAGCCTATTGGTGGGACCTTGTGATTATGTAAGTTAA[**TG**]<sub>17</sub>TGTCTAT  
CTCCCATATGTATCTCCC[TA]<sub>6</sub>[CA]<sub>2</sub>[TA]<sub>5</sub>[CA]<sub>3</sub>[TA]<sub>4</sub>[CA]<sub>3</sub>[TA]<sub>3</sub>[CA]<sub>4</sub>[TA]<sub>4</sub>[C  
A]<sub>2</sub>[TA]<sub>4</sub>[CA]<sub>3</sub>[TA]<sub>3</sub>[CA]<sub>2</sub>[TA]<sub>4</sub>[CA]<sub>2</sub>[TA]<sub>6</sub>AAAATTCATTCTGTCCCCCTAGAGA  
ACCTGACTAATACATACAT[**AG**]<sub>12</sub>ATTAAATATAGTTAGC

A PCR product is obtained when using the middle and last primers only.

This is a cryptic repeat flanked by two perfect repeats.

Figure 4.3.11 Marker 29

CAGTTCCTTGTACA<sub>c</sub>CCAGGTCCTGTAGTTGAAGCTGTGTGTTGGAAGTCTGCA  
ATCTGACATTTTAGATGCAGTTTTACCTCTAGCTTTATCCCTGCACATC[**GT**]<sub>8</sub>[**T**  
**G**]<sub>12</sub>ATCTTACTATTCACAGGCACATTTT<sub>c</sub>CTGAGTCTGGAGCATTCACTTACCAA  
AA

This is a compound repeat.

The repeats of clones A and 16 are complex (Figs 4.3.10 and 4.3.3). It was hypothesised that by amplifying the (TG)<sub>17</sub> component of the repeat of clone A that a suitable marker for genetic mapping would be defined. However, despite varying the annealing temperature and PCR buffer components, no PCR product could be produced. Primers flanking the more complex 120 bp portion of the repeat were made to investigate if this portion was polymorphic but only 1 female of 26 genotyped possessed two alleles (a heterozygosity of 3.8%), that differed in size by about 2 bp. As a consequence, the (complex) repeat sequence of A is not polymorphic according to the definition that a marker is polymorphic only when its most common allele has a population frequency of less than 95% (Ott, 1991).

#### 4.4.2 CEPH genotyping

Generally, based on the structure and length for markers identified in this study (Table 4.3) and markers identified by others, but now genotyped through the CEPH families (Table 4.4), the observed heterozygosities of the repeats are in accordance with Weber's (1990) observations. These observations were that the longer the run of uninterrupted dinucleotides the greater the heterozygosity of the repeat. The notable exceptions being those repeats with a lower-than-anticipated heterozygosity (*DXS1153*, *DXS1122* and *DXS237*) and one repeat (*3'Dys MS*) with a higher-than-anticipated heterozygosity. The length of the repeat sequences of these loci, given in Tables 4.3 and 4.4, generally corresponds to that of the most common allele. The "anticipated heterozygosity" was not based on the repeat length of one of the less common alleles, which could have accounted for the unanticipated results. Chi square analysis shows however that there is no significant difference between the observed heterozygosities (in females) and the expected heterozygosities (determined from sex pooled allele frequencies) of *DXS1122*, *DXS237* and *3'Dys MS* ( $\chi^2_1$   $p > 0.5$ ,  $> 0.9$ ,  $> 0.5$  respectively).

**Table 4.3 Characteristics of seven new polymorphic AC repeat markers (Xpter to Xqter) and three non-polymorphic STS loci**

D No. (Accession No.)	M13 repeat 5'-3'	Primer 5'-3' (forward/reverse)	[MgCl <sub>2</sub> ] (mM) (a)	Thermal cycle (b)	% Heterozygosity		No. of alleles	Product size (bp) (predominant allele)	CEPH
					Obs	Exp			Genotype (c)
<i>DXS1126</i> (L21179)	(TC) <sub>17</sub> (TG) <sub>18</sub>	TTCTAGAAAGGTGCGTGTCTGCTGG GACCATTCCCTCTCAAACACAAACG	6.5	A	51	67	8	230-252 (244)	248 : 246, 248
<i>DXS1125</i> (L21178)	(GT) <sub>28</sub>	CTTTAGGGAGGTGAGCTTTCAGGTT GGAGTGCTCTACTGCTAGGATGTAA	4.5	C	77	79	9	122-138 (128)	130 : 126, 128
<i>DXS1124</i> (L21177)	(TG) <sub>6</sub> TT(TG) <sub>17</sub>	GCCTTTCTCCTAACCACCTTACTTGG CCTGAAGCATATATTGAGGG	8.0	A	54	54	8	112-126 (118)	118 : 118, 120
<i>DXS1153</i> (L21180)	(TG) <sub>22</sub>	CTGAGGATGGCTTGAAGGT CATCACACAATACCCCAGGGTCT	6.0	A	46	86	13	134-160 (136)	150 : 136, (140)
<i>DXS1120</i> (L21174)	(AC) <sub>23</sub>	GGATTCTAAGACCCCTCTCACACCA CCATTCAGTTTTATTTCAGGGGAAAG	5.5	B	74	72	7	122-134 (130,132)	132 : 130, 130
<i>DXS1122</i> (L21175)	(TG) <sub>24</sub>	TACACTGAAAGGACTGCACCAGGT CAAGAGCCAATCCTGAATAGGAGC	5.5	A	33	38	6	106-116 (110)	110 : 110, 112
<i>DXS1123</i> (L21176)	(CA) <sub>24</sub>	TGCCTAAATGTTTCGCAAGCCCATTC ACAAACAGCTGCCTCCTAGAAACCC	8.0	A	67	68	7	168-180 (172)	172 : 172, 174
STS loci									
W	(GT) <sub>21</sub>	TTGACCAGGACACATGGCAATGGAGA TGACTACTGAGGGCAGTTCTGAGTGT	7.5	A			1	168	
16 (d)		CTGAATTCATTTAGTTGGTTCTGG TCGCCCCGCTCCCCAACAATCT	4.0	A			1	184	
A (d)		GTCTATCTCCCATATGTATCTCCC CTCTAGGGGGACAGAATGAATTTT	8.5	A	4		2	120	
29 (e)	(GT) <sub>8</sub> (TG) <sub>12</sub>	ACCTCTAGCTTTATCCCTGCACATC TGGTAAGTGAATGCTCCAGACTCAG	6.0	A			1	116	

a Concentration of the MgCl<sub>2</sub> used in the PCR when using "himix". Alternatively, the 7 polymorphic markers could be amplified using 1.5 mM MgCl<sub>2</sub> and thermal cycle A, when using "lomix" (see 3.2.15.1).

b Thermal cycling conditions are given in 4.3.4.

c Alleles given are those possessed by the parents of the CEPH pedigree 1331 (01 father, 02 mother). For marker *DXS1153*, 1331/02 possesses a cryptic allele (140). See 4.4.2.1 for further details.

d Complex repeats. See Figure 4.3 for their structure.

e Not physically nor genetically mapped.

**Table 4.4 Previously characterised X-Linked microsatellites (Xpter to Xqter) now genotyped in the CEPH families**

Locus	Marker	Repeat structure (a)	References	[MgCl <sub>2</sub> ] (mM) (b)	Thermal cycle (c)	% Heterozygosity		No. of alleles	Product size (bp) (d) (predominant allele)	CEPH
						Obs	Exp			Genotype (e)
										1331/01:1331/02
<i>DXS237</i>	GMGX9AC	(GT) <sub>23</sub>	Gedeon <i>et al.</i> , 1992	6.0	A	51	52	8	153-167 (161)	161 : 161, 161
<i>DMD</i>	5' Dys-II	(CA) <sub>23</sub>	Feener <i>et al.</i> , 1991	7.5	A	77	72	10	214-236 (214)	214 : 214, 214
	3' Dys MS	(CA) <sub>8</sub> TA(CA) <sub>7</sub>	Oudet <i>et al.</i> , 1990, 1991 Beggs and Kunkel, 1990	7.5	A	38	36	5	121-139 (135)	137 : 127, 135
<i>AR</i> (f)	AR(CAG)	(CAG) <sub>19</sub>	Tilley <i>et al.</i> , 1989	7.5 or 2.0	A or D	77	87	17	240-309 (279)	288 : 279, 291
	AR(GGN)	(GGN) <sub>23</sub>	Tilley <i>et al.</i> , 1989	2.0	D	64	62	10	163-211 (184)	187 : 184, 184
<i>PGK1P1</i>	PGK/5	(CA) <sub>14</sub>	Browne <i>et al.</i> , 1992b	7.5	A	28	40	4	198-204 (202)	198 : 202, 202

a As given in the references cited.

b Alternatively, *DXS237*, and *PGK1P1* can be amplified using 1.5 mM MgCl<sub>2</sub> thermal cycle A and "lomix". *AR*(CAG) cannot be amplified using "lomix".

The other markers have not been tried using the "lomix" conditions.

c Thermal cycling conditions are given in 4.3.4.

d Sizes of the most frequent alleles of the markers *GMGX9AC*, *PGK/5*, *5'Dys-II*, and *3'Dys MS* were assumed to be the same as those given in the cited references.

*AR*(CAG) and *AR*(GGN) allele sizes were determined from a size standard (a gift from Dr W. Tilley).

e Alleles given are those possessed by the parents of the CEPH pedigree 1331 (01 father, 02 mother).

f Primers used were those cited in the references given for each marker, except those for *AR*(CAG) and *AR*(GGN), which were designed by Dr R. Richards from the sequence given in Tilley *et al.* (1989), see 4.3.4. The combined heterozygosity of *AR*(CAG) and *AR*(GGN) is 95%.

#### 4.4.2.1 Cryptic alleles

There *is* a significant difference between the observed and expected heterozygosities of *DXS1153* ( $\chi^2_1$ ,  $p < 0.01$ ). Although 13 alleles were seen for this marker it was noted during the CEPH typing that some daughters who were expected to be heterozygotes, on the basis of parental genotypes, only exhibited one allele or the second allele was much fainter. These alleles always amplified from the DNA of (hemizygous) males and were designated as “cryptic” because of their failure to amplify efficiently (from the DNA of females) in the presence of another allele.

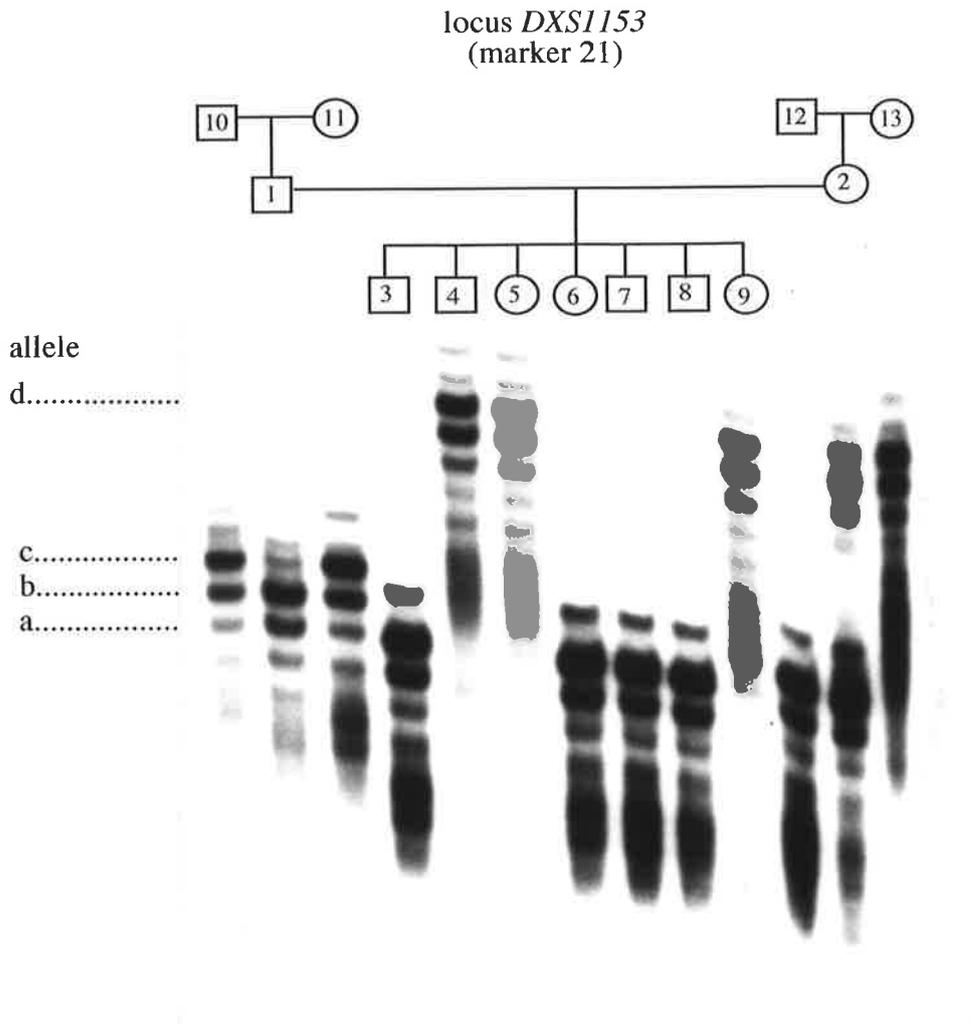
These cryptic alleles were noted in 5 of the 18 families that had heterozygous mothers. An example of this phenomenon is given in Figure 4.4. In this pedigree (13291) the paternal allele is not seen in any of the 3 daughters. The same phenomenon occurs in 4 other families (35, 45, 1418 and 1421) but involves alleles of different sizes.

As a consequence it was hypothesised that some of the mothers originally scored as homozygotes were in fact heterozygotes with a cryptic allele. This was confirmed by amplifying the DNA from the sons of apparently homozygous mothers to reveal any alleles not amplified from the mother’s DNA. In this manner, 15 mothers were shown to be cryptic heterozygotes. This is a lower estimate because in some families the mothers will not be detectable as heterozygotes if they have not passed on their cryptic allele to any of their sons. Taking into account these additional 15 mothers there is no longer any significant difference between observed and expected heterozygosities for this marker.

The non-amplification of alleles is consistent with repeated PCRs from DNA of the relevant individuals. Different PCR conditions have been used in an attempt to enhance the amplification of the cryptic alleles from selected individuals but to no avail.

The molecular basis for this phenomenon of preferential amplification of one allele compared with another is not known. It may be due to a polymorphism in one of the primer target sites causing differential binding of the primer. An attempt was made to investigate this problem further. As the repeat sequence was very close to the cloning site there was no choice in regard to selection of the forward primer sequence, so to obtain

## CEPH pedigree 13291



**Figure 4.4** CEPH pedigree showing the phenomenon of cryptic alleles associated with marker 21 (locus *DXS1153*). By looking at the most intense band(s) in each track (indicated by the letters representing the different alleles), it can be seen that the daughters (5, 6 and 9) have apparently not inherited their father's allele (allowing for the curvature of the image). In addition, the paternal grandmother (11) does not show the allele possessed by her son (the allele b band is not as dense as the allele c band). The many faint "stutter bands" in each track are typical of dinucleotide repeats and are thought to be the result of 2 base-pair strand slippage during the PCR (Hauge and Litt, 1993; Murray *et al.*, 1993).

more sequence preceding the M13 cloning site the original *lambda* clone was sequenced using the reverse PCR primer. The sequence revealed using this approach did not coincide exactly with that derived from the M13 clone (Fig. 4.3.6); there appeared to be an insertion of two bases and a deletion of another two relative to the sequence from M13. Using a new forward primer, designed from this new sequence information (Fig. 4.3.6), in conjunction with the reverse primer resulted in a very poor yield of PCR product and therefore this approach did not resolve the cryptic allele problem. Shifting the reverse primer target site was not attempted.

Before encountering this cryptic allele phenomenon, a sequence database search using the FASTA program revealed that 29 of the 132 bp of sequence flanking the repeat of clone 21 had almost complete homology to part of the *DXS178* locus, which also contains a microsatellite (GenBank accession no. X65235). This was the minimum region of homology because, as can be seen in Figure 4.3.6, it encompasses the last 29 bases of the readable sequence derived from the M13 clone. As the other 103 bases flanking the (TG) repeat were not within this region of homology and that the orientation of the microsatellite sequence associated with *DXS178* was inverted (that is, AC rather than TG) it was concluded that they were not the same marker and so a DXS number (*DXS1153*) was obtained from GDB and the locus genetically mapped (Chapter 5).

#### 4.4.2.2 Androgen receptor repeats

The androgen receptor gene (*AR*) contains two microsatellites within exon 1: a CAG repeat, and a GGN (N=T or C) repeat. The heterozygosity of the CAG repeat is 77%. To increase the informativeness of the CEPH families for the *AR* locus, those families whose mothers were heterozygous for *AR*(GGN) but homozygous for *AR*(CAG) were also genotyped (for the GGN repeat). Alone, the GGN repeat has a heterozygosity of 67% but as the two repeats (CAG and GGN) are separated by only 1.1 kb they can be considered as one locus that has a (combined) heterozygosity of 95%.

Difficulty was encountered in amplifying the larger alleles of CEPH individuals heterozygous for the *AR*(GGN) repeat using those PCR conditions used successfully to amplify the *AR*(CAG) alleles. Using thermal cycling option D (4.3.4) permitted both alleles of a heterozygote to be amplified (although not in the same reaction).

#### 4.4.2.3 Mutations

Nine mutations were observed and these are given in Table 4.5 and shown in Figure 4.5. Where it was not obvious, the parental origins of these mutant alleles were determined by phase analysis (Table 4.5; Fig. 4.6). This involved determining the most likely haplotype of the marker under question and two of the closest informative flanking markers available. From this information the most likely chromosomal origin of the mutation was determined and as a consequence the magnitude and sign of the change. There is no suggestion of any bias as to: 1) the parental origin of the mutation events nor 2) the direction of any changes in size. However the number of mutations observed (9) was too small for more detailed analysis.

DNA samples from the males 23/01, 1375/08 and 1418/08 possessed two alleles for the markers *DXS237*, *AR* and *DXS1123* respectively (Table 4.5). Using primers that amplify an X-specific amelogenin sequence as well as an amelogenin-like Y-specific sequence enabled these samples to be sexed and showed that they were in fact from males, and not mis-labelled samples from females. Supporting this conclusion is that if the sample from 1418/08 and 1375/08 was a mis-labelled sample from a female it would have had to be heterozygous for markers other than *DXS1123*, but was not. The proposition that these samples had been contaminated with the DNA from other males has been discounted because of the unlikely possibility the contaminating male had the same genotype for all bar one of the markers for which these males have been typed. Similarly the possibility that these markers reside in an uncharacterised pseudoautosomal region was rejected because the majority of males are not heterozygotes.

The heterozygous father 23/01 has the same genotype for marker *DXS237* as his wife. One of his daughters (23/05) is a homozygote so it is known that he has passed on this particular allele (#4). Phase analysis, using the principal of crossover minimisation, has shown that he has not passed on the second allele (#6) to any of his other three heterozygous daughters (Fig. 4.6.5).

Considering these 9 mutations and the total number of alleles scored by the candidate upon the genotyping of 13 polymorphic markers through the CEPH families, the overall mutation rate *per gamete per locus* is 0.015% ( $[7/3571]/13$ ). (The two males with two

alleles are not included in this calculation because their mutations are post-zygotic (or artifactual - see Discussion) and thus don't contribute to the "*per gamete*" rates which is the conventional way of expressing these results). This figure of 3571 alleles does not include those alleles that could not be unequivocally excluded as new mutations, that is, those alleles of individuals whose parents had not been genotyped.

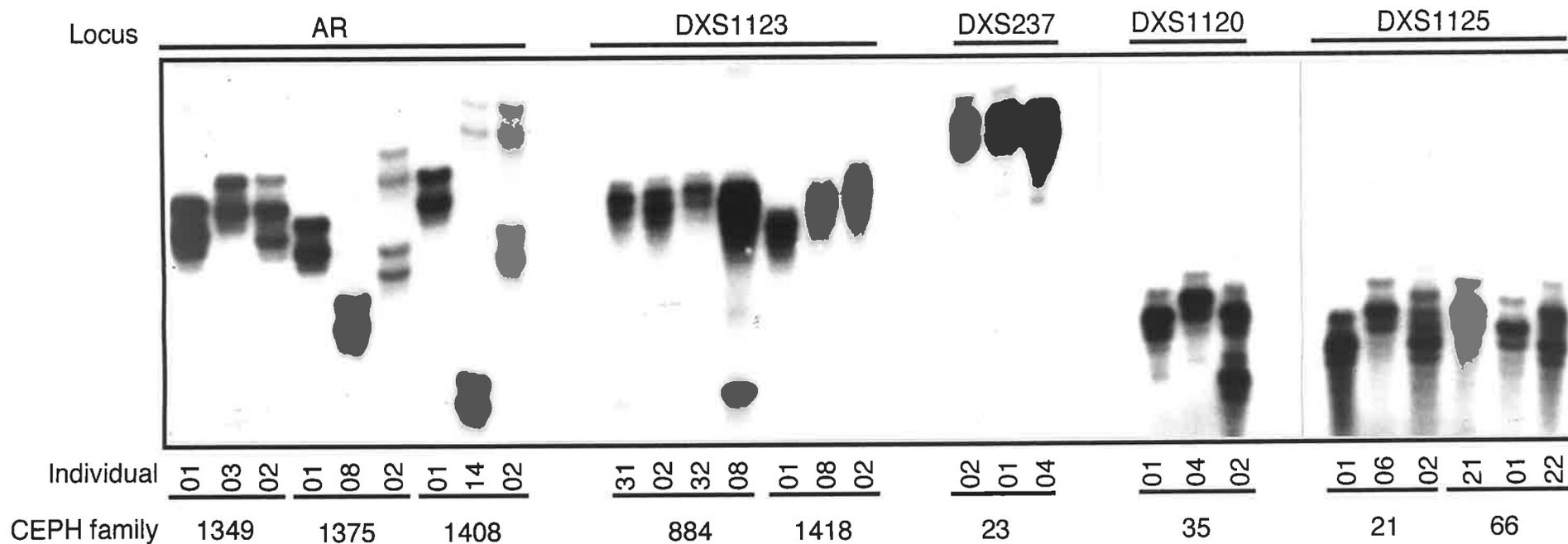
**Table 4.5 Mutations**

CEPH family	individual (sex)	Locus	Mutation	Parental origin	Mutation rate % <sup>a</sup>
1349	03 (F)	<i>AR (CAG)</i>	increase of 1 repeat unit	paternal	0.67
1375	08 (M)	<i>AR (CAG)</i>	decrease of 3 repeat units <sup>b</sup>	maternal	
1408	14 (F)	<i>AR (CAG)</i>	decrease of 13 repeat units	paternal	
884	02 (F)	<i>DXS1123</i>	decrease of 1 repeat unit	paternal <sup>b</sup>	0.55
1418	08 (M)	<i>DXS1123</i>	two alleles	NA	
23	01 (M)	<i>DXS237</i>	two alleles	NA	0.38
35	04 (M)	<i>DXS1120</i>	increase of 1 repeat unit <sup>b</sup>	maternal	0.24
21	06 (M)	<i>DXS1125</i>	increase of 1 repeat unit <sup>b</sup>	maternal	0.45
66	01 (M)	<i>DXS1125</i>	increase or decrease of 1 unit	maternal	

<sup>a</sup> (Number of mutations/number of X chromosomes) x 100.

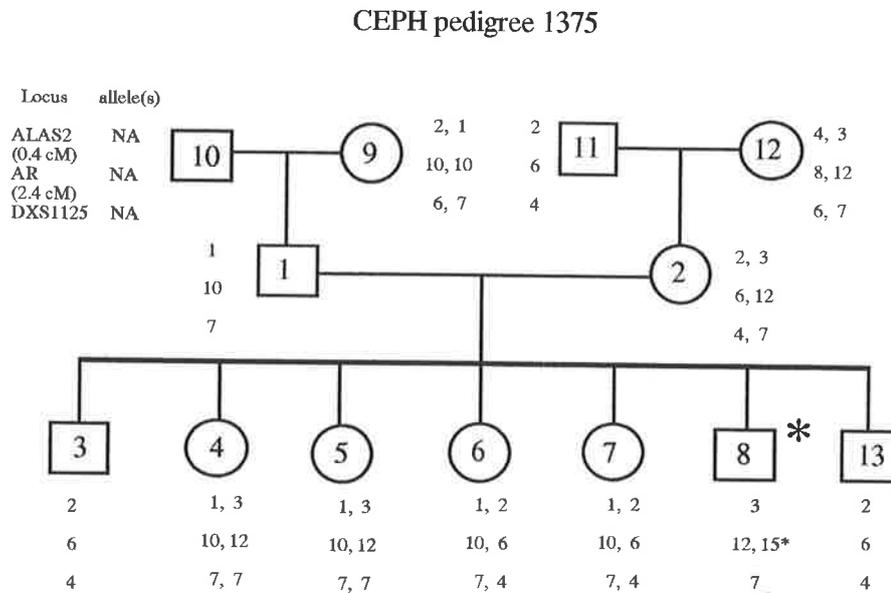
<sup>b</sup> As determined from phase analysis, see Figure 4.6 (The most common haplotype was considered as the non-recombinant phase from which the type and origin of the mutation were determined).

NA=Not applicable.

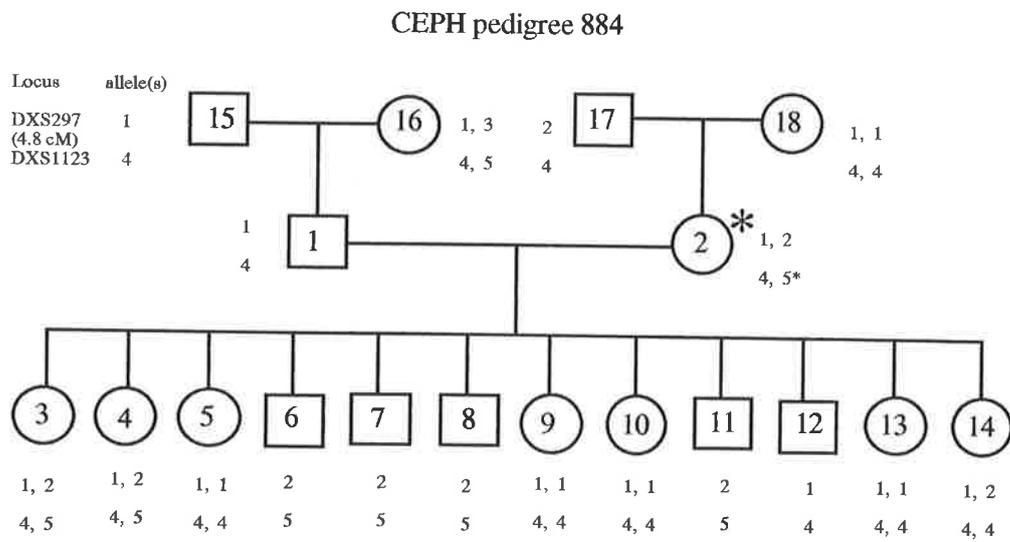


**Figure 4.5** Autoradiograph showing the mutations referred to in Table 4.5. The locus is shown above each set of PCR products. The lanes with the mutated alleles are flanked by the alleles of the father (01) and mother (02), except in the case of 23/01, which is flanked by the spouse's and a son's alleles (not clearly reproduced in the photo). In addition, 884/08 is included to show that 884/02 does possess an allele not carried by either parent (not clearly reproduced in the photograph).

**Figure 4.6** Phase analysis to determine the allelic origins of the mutations referred to in Table 4.5. The most likely parental phases are shown. The distances (cM) between the loci were taken from Table 5.1. Alleles differ in size by multiples of the repeat unit; the lower the allele number the greater the length of the repeat. NA=not available.

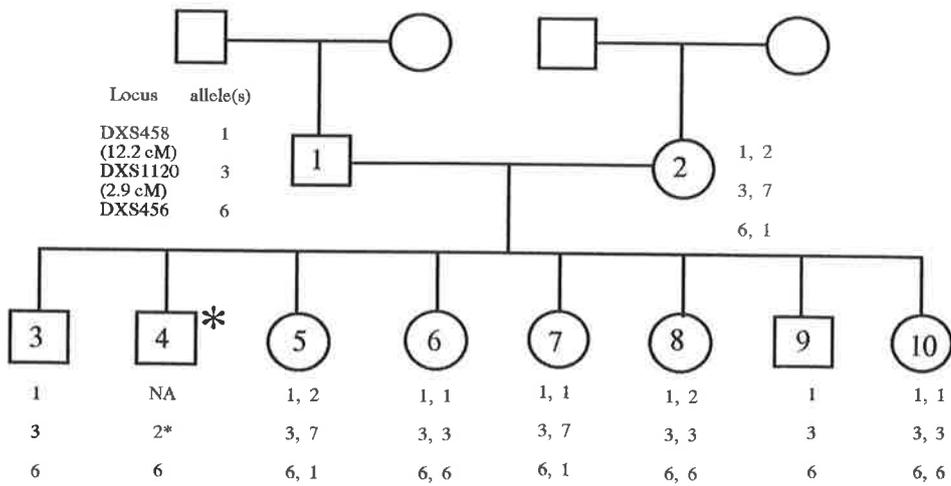


**Figure 4.6.1** Phase analysis to determine the allelic origin of the mutated AR allele (\*) possessed by 1375/08.



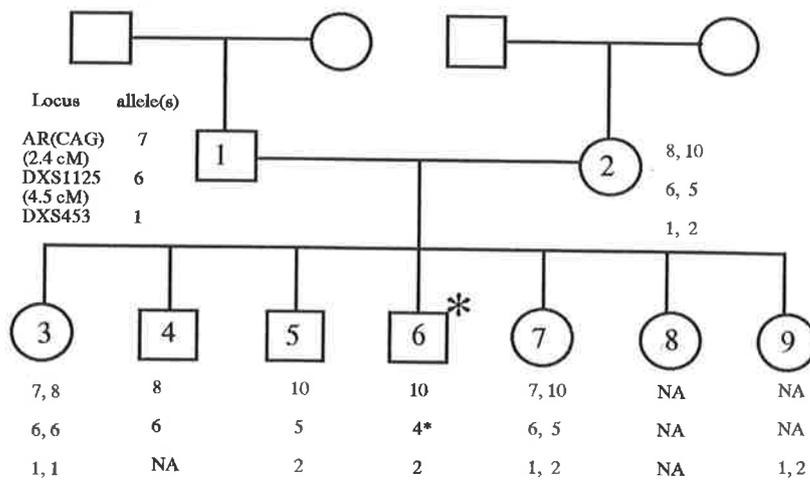
**Figure 4.6.2** Phase analysis to determine parental origin of the mutated *DXS1123* allele (\*) possessed by 884/02. Markers distal to *DXS1123* were either uninformative or had not been genotyped.

CEPH pedigree 35



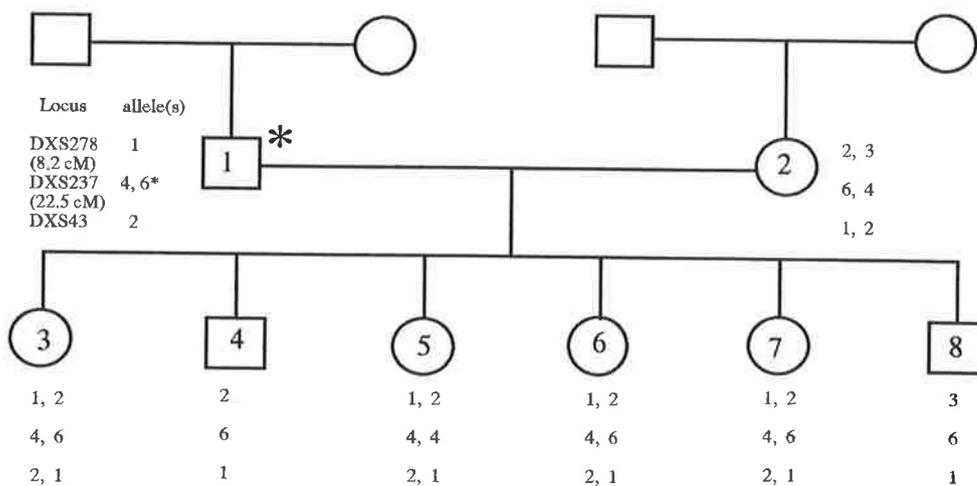
**Figure 4.6.3** Phase analysis to determine the maternal allele origin of the *DXS1120* mutation possessed by 35/04. Alleles for markers *DXS458* and *DXS456* were taken from the CEPH database.

CEPH pedigree 21



**Figure 4.6.4** Phase analysis to determine the allelic origin of the mutated maternal *DXS1125* allele (\*) possessed by 21/06. The alleles for the marker *DXS453* were taken from the CEPH database.

CEPH pedigree 23



**Figure 4.6.5** Phase analysis to determine whether or not the heterozygous father 23/01 has passed on *DXS237* allele #6 (\*) as well as allele #4. The alleles for the markers *DXS278* and *DXS43* were taken from the CEPH database.

## 4.5 Discussion

### 4.5.1 Isolation of microsatellite markers

From 72 AC positive lambda clones, 7 new polymorphic PCR based markers were characterised (Tables 4.2 and 4.3). This success rate (10% from 72 AC positive clones) in obtaining new markers for mapping the X chromosome was lower than anticipated. In hindsight, the propensity for re-arrangement of at least some of the M13 clones was underestimated. Different template preparations of the same clone could have been sequenced to reveal a “consensus” sequence before primers were made, to eliminate those clones that appeared to be unstable. To investigate the phenomenon of clone instability further, PCR products (using M13 forward and reverse primers) from different template preparations of the clones D (suspected of being unstable, see 4.4.1), F, K, 32, 42 and 49 were compared (data not shown). Variable results were obtained between the different template preparations of the clones D, F, K, 42 and 49. Only the preparations of clone 32 gave consistent PCR results in terms of product size and yield.

Four of the eleven PCR products were not polymorphic (Table 4.2). The structure of the repeats of 3 of these 4 was consistent with this observation. The unusual structure of the cryptic repeats of clones A and 16 made them intrinsically interesting despite the fact that it was realised that they were unlikely to be useful for genetic mapping. The non-polymorphic nature of the repeat of clone 29 is probably due to its compound structure, although the possibility existed that this repeat was merely one of the smaller alleles of a polymorphic marker. The repeat of clone W was expected to be highly polymorphic given that it has a perfect structure and is 21 units in length. There does not seem to be anything unusual in the sequence surrounding this repeat (Fig. 4.3.1) that could account for this lack of variability within the CEPH parents.

A relatively low microsatellite recovery rate of 20% was also experienced by Hazan *et al.* (1992) who used methods similar to those described herein. One third of their AC-positive lambda clones did not yield AC-positive M13 subclones. They thought the cause may be Z-DNA related and recommended the use of a *sbcC*, *recB*, *recJ* host strain to reduce the frequency of M13 rearrangement.

The failure to generate PCR products from clones I, K, 32 and 42 could be explained by a chimaeric nature of the M13. Restriction enzyme sites between one of the primers and the repeat sequence were overlooked in clones 32 and 42, therefore the M13 sequenced could be the result of the co-ligation of non-contiguous restriction fragments. There was insufficient sequence between the repeat and the reconstituted restriction sites to design a second primer. For clone K the same explanation could apply although, if so, the co-ligating fragments must have been blunt-ended.

Although the repeats of clones A, W, 16 and 29 were not polymorphic and thus cannot be used to construct a genetic map they can be used as sequence tagged sites (STSs) in physical mapping. The size and structure of the repeat of clone A is similar to an evolutionary conserved repeat within intron 1 of the *factor IX* gene (Sarkar *et al.*, 1991; Jacobson *et al.*, 1993). These types of repeats are part of a class referred to as “cryptic repeats” because no single tandem repeat predominates. The *factor IX* repeat of Sarkar *et al.* (1991) was polymorphic (in Caucasians but not South Koreans) due to the insertion/deletion of additional repetitive units 24-26 bp in length. Linkage equilibrium between these various alleles and polymorphisms on opposite sides of the repeat led these authors to conclude that these types of repeats are hot-spots for recombination, possibly *via* their potential to form Z-DNA. The mechanism of generation of these types of repeats may therefore be due predominantly to recombination or sister chromatid exchange rather than polymerase slippage.

The sequence surrounding the repeat of *DXS1122* (Fig. 4.3.8), which, based on its length and perfect structure was anticipated to be highly polymorphic, is not remarkable in terms of the presence of sequence structures which may inhibit the mechanism(s) generating length variation (reviewed by Levinson and Gutman, 1987). Of course sequences further away than those given in Figure 4.3.8 may be involved. Alternatively, random genetic drift could have resulted in the loss of some alleles and the consequent low heterozygosity or, as with coding trinucleotide repeats, there may be some selective pressure maintaining a low mutation rate at this locus. Curiously, following genetic mapping (Chapter 5) a marker, *DXS994*, isolated by Weissenbach *et al.* (1992) was shown to have a regional localisation that coincides with that of *DXS1122* and also has a heterozygosity of only 33%, and a repeat copy number, as obtained from GenBank, of 19 units. The repeat copy number range for *DXS994* was not available, so it is not known whether 19 copies

represents a relatively large rare allele, of a typically much smaller (and consequently less polymorphic) microsatellite, or whether 19 copies is the most frequent allele (which would suggest that, as for *DXS1122*, some other mechanism other than length, is inhibiting the potential variability at this locus). The sequence flanking the repeat of *DXS994* is not the same as that of *DXS1122*, so they are not the same marker.

#### 4.5.2 Cryptic alleles

The term “cryptic alleles” (4.4.2.1) was coined by the candidate to distinguish them from previously reported “null alleles”, which are characterised by a consistent absence of product (Phillips *et al.*, 1991; Weber *et al.*, 1991; Callen *et al.*, 1993; Gyapay *et al.*, 1994).

The reproducible nature of the results with the marker *DXS1153* is evidence against the interpretation that this marker is intrinsically of low informativeness and that the profusion of alleles seen during genotyping are the consequence of *in vitro* mutation events during the PCR. The fact that usually no more than 1 or 2 alleles are seen in samples from males and females respectively can also be used as evidence against this hypothesis.

This marker, *DXS1153*, resembles *IL9* (Weber *et al.*, 1991) and *5'Dys MS* (Oudet *et al.*, 1991) in the sense that they also have alleles that amplify poorly or not at all. It also resembles the marker *L5.62* of Koorey *et al.* (1993) in that in both cases one of the primers encompasses part of the repeat sequence and some alleles are amplified poorly. Koorey *et al.* (1993) solved their problem, which was due to a polymorphism in the primer site, by selecting a different site, but this failed to correct the problem for *DXS1153* (see Results). Koorey *et al.* (1993) report that sites adjacent to repeats may themselves be of a (less obvious or “cryptic”) repetitive nature and as a consequence hypermutable and therefore best avoided as sites for primers.

The reverse primer of *DXS1153* has the potential to form a hairpin structure *via* the sequence ...ACCCTGGGT..., but the consistent failure to amplify cryptic alleles from the same DNA sample and the fact that other alleles are routinely amplified argues against this being the cause of the difficulties.

As a consequence of the amplification problems, *DXS1153* can only be confidently used in linkage analyses when the parents have different alleles so that the occurrence of (apparently) homozygous daughters can be attributed to the presence of a cryptic allele.

A map from the 5th X chromosome workshop (Willard *et al.*, 1994) has positioned *DXS1153* near to the locus *DXS178*, with which it has some homology (Fig. 4.3.6). Thus the TG repeat of *DXS1153* may be a previously uncharacterised microsatellite at *DXS178*. Given that the region of homology is *within* the sequence designated as *DXS178*, a reason as to why the unique sequence immediately flanking the repeat of *DXS1153* did not show up in the FASTA homology search could be that the M13 clone sequenced may have multiple inserts. In other words, if the *lambda* clone 21 does contain the locus *DXS178*, in addition to *DXS1153* (there were three AC-positive *HaeIII* restriction fragments derived from this *lambda* clone), upon shotgun sub-cloning, 2 non-contiguous restriction fragments could have co-ligated to form the M13 sub-clone, with one of those fragments being part of the previously characterised region of *DXS178* adjacent to an AC repeat designated A01180CA by De Weers *et al.* (1992). If this is so, one would expect the *HaeIII* restriction enzyme site (5'GG'CC3') to be reconstituted at the junction between the non-homologous and homologous regions, which did not occur (Fig. 4.3.6), although the co-ligation could have been between two sheared DNA fragments (that is, fragments not necessarily terminating in GG or CC). An alternative explanation may be that the homologous sequence could be from a duplicated region and that the *DXS1153 lambda* and M13 clones contain some or all of this duplication. The former co-ligation hypothesis would be supported if a PCR product could be generated from *lambda* clone 21 using primers that flank the AC repeat A01180CA. This has not been attempted.

#### 4.5.3 Androgen receptor repeats

The difficulty in amplifying larger alleles, as for individuals heterozygous for *AR*(GGN), has also been reported as occurring for the CCG repeat of *FRAXA* (Yu *et al.*, 1992) and for the 5'*Dys* MSA marker of the dystrophin gene (Oudet *et al.*, 1991). Oudet and colleagues suggest this may be caused by the variation in the stability of a stem-loop structure between different alleles of the MSA repeat. The *AR*(GGN) and *FRAXA* repeats are GC-rich and the difficulty in amplifying long stretches may be caused by rapid renaturation and/or the formation of secondary structure following base-pairing between intra-strand cytosine and

guanosine bases inhibiting: 1) primer access and 2) the progress of the polymerase.

Kennedy's disease, otherwise known as spino-bulbar muscular atrophy (SBMA; MIM 313200), is one of several diseases known to be caused by or associated with an expanding trinucleotide repeat (for brief reviews see Miwa, 1994; Richards and Sutherland, 1994). SBMA is primarily a motor neuron disease resulting in muscle weakness, although other symptoms may be present. It is caused by an expansion of the (CAG)<sub>n</sub> repeat, in exon 1 of the androgen receptor gene, that codes for an NH<sub>2</sub>-terminal poly-glutamine tract (Biancalana *et al.*, 1992). The androgen receptor is a DNA-binding protein that, upon activation by an androgen, regulates transcription from a number of target genes. Exon 1 specifically encodes that part of the protein that is involved in this final *trans*-activation step (Simental *et al.*, 1991). It has been hypothesised that the (CAG)<sub>n</sub> expansion may result in a loss of androgen sensitivity and in a gain of motor neuron toxicity (Mhatre *et al.*, 1993).

As an aside, by convention, the nucleotide sequence of a microsatellite should begin with the letter that comes first in alphabetical order, for example AC repeat rather than CA repeat (Jin *et al.*, 1994); however, as the androgen receptor repeats are protein coding sequences, for clarity they are written in their sense form, that is CAG rather than AGC and GGN rather than CCN.

The number of CAG repeats in those suffering from the SBMA ranges from 40 to 62, which does not overlap with the range of alleles within the normal population (La Spada *et al.*, 1992). With more observations however, the range of the normal copy number and/or that of the abnormal copy number may change. As presented herein (Table 4.4), 17 alleles ranging from 6 to 29 repeats were observed in the 40 CEPH families, which are assumed to be composed of normal individuals with respect to the androgen receptor gene. La Spada *et al.* (1992) recorded the normal range in the population they studied as from 13 to 30 repeat units. The smallest CAG repeat (n=6) in the CEPH population was possessed by a single individual who was a new mutant (Table 4.5); if this result is due to an *in vitro* event (see below) and thus disregarded, the normal range now extends from 9 to 30 repeat units.

One question that arises is; how small can the CAG repeat be before disease symptoms arise (if at all)? Precise deletion of the CAG tract has been observed to have no effect on the *trans*-activating function of the receptor (Mhatre *et al.*, 1993). Surprisingly however, a deletion leaving 12 copies of the CAG repeat has been shown to enhance the detrimental effects of an amino acid substitution in exon 5, to result in partial androgen insensitivity (McPhaul *et al.*, 1991).

No disease is known to be associated with the other, imperfect, trinucleotide repeat (GGN), where N is C, G or T, also in exon 1 of the androgen receptor gene. This repeat codes for poly-glycine. The repeat copy number in the CEPH population ranges from 16 to 32, with the most frequent being 23 (Table 4.4). Patterson *et al.* (1994) report the range as from 12 to 31 units, from an unspecified population.

#### 4.5.4 Mutations

PCR of DNA samples from some individuals revealed alleles not present in either parent. These are given in Table 4.5. Pipetting error was eliminated as a source of these apparent mutations by repeating the PCR from any available DNA stock tubes. Also, sample mix-up nor PCR artifacts cannot be invoked to explain the 9 mutations because all the samples have been genotyped for many other markers and re-typed for those under question. The three *AR* (CAG) mutations have been confirmed by Weeks *et al.* (1995).

A bias in the parental origin and direction of the change in size of mutated microsatellites has been noted previously (Weber and Wong, 1993). Mutations occurring predominantly in the male germline were reported, but may simply reflect the greater number of cell divisions between the zygote and sperm (Weber and Wong, 1993). The mutation rates given in Table 4.5 are also minimum estimates, in that mutation events would not be detected if they produced alleles that: 1) did not violate Mendelian inheritance and 2) escaped the error checking procedures during map construction (Chapter 5).

The mutation rate of 0.20% (mutations *per* gamete) found in this study is twice as great as that reported by Weissenbach *et al.* (1992) and Gyapay *et al.* (1994) who together, at the Genethon institute in Paris, have typed 2066 microsatellite markers throughout the human

genome. Other observations of the mutation rates of microsatellite markers range from 0.002% to 1.5%, although these figures do not distinguish between *in vivo* and *in vitro* events (Kwiatowski *et al.*, 1992; Beckmann *et al.*, 1993; Mahtani and Willard, 1993; McInnis *et al.*, 1993; Petrukhin *et al.*, 1993; Weber and Wong, 1993; Banchs *et al.*, 1994). The *in vivo* (germline) microsatellite mutation rate has been estimated as less than 0.01%, whereas the *in vitro* cell line mutation rate is 10 to 100 times greater (Banchs *et al.*, 1994).

The calculated mutation rates can be influenced by: 1) the type of markers used (for example tetranucleotide repeats have a relatively high mutation rate compared to dinucleotide repeats; Weber and Wong, 1993); 2) their heterozygosities (that is, their relative instability); 3) the number of markers genotyped; 4) care in genotyping; and 5) whether any distinction is made between *in vivo* and *in vitro* events. Determination of the mutation rate is important because (undetected) mutation events (that is, mutation to another allele already segregating within the family being genotyped) will limit the resolution of any linkage map.

Clonal evolution of an initially mosaic cell population has been reported (Migeon *et al.*, 1988). As a consequence, the mutation rates cited in Table 4.5 may reflect the incidence of either *in vivo* somatic mutations or those that occur during cell culturing. Assuming the latter, the cells harbouring these mutations either co-exist with those cells possessing the expected (Mendelian) alleles or come to dominate the cell culture. The possibility of culturing artifacts is supported by the observations of Weber and Wong (1993) and Banchs *et al.* (1994). For example, of 31 cases of mutation observed in DNA from cell cultures (11 of these involved DNA samples exhibiting 3 alleles) only in 12 cases was the mutation shown to be present also in the corresponding untransformed lymphocyte samples (none of these confirmations involved 3 alleles (Weber and Wong, 1993)). B-lymphocytes are used to generate the LCLs distributed by CEPH because it is thought that they maintain a relatively stable karyotype over many generations (Neitzel, 1986). Perhaps another cell type may be more stable at the molecular level. Given that mutations may lead to suspicions of sample mix-up, non-paternity and errors in linkage analyses it is appropriate that DNA from untransformed cells be used whenever possible.

## 4.6 Conclusion

To conclude, 11 new microsatellites were isolated and characterised. Seven of these were polymorphic. Together with 6 others they were genotyped through the CEPH families for subsequent incorporation into a PCR-based genetic map of the X chromosome (Chapter 5), for application to mapping X-linked disorders (Chapters 6 and 8). In addition, information was gained relating to the frequency of mutation events that can be expected when typing such markers, which has a bearing on genetic map construction. These markers were then made available, through primer sequences in GenBank and genotypes in the CEPH database, to the rest of the scientific community, for the construction of genetic and physical maps of the X chromosome and for localisation of X-linked disease genes. The published version (Donnelly *et al.*, 1994b) of the work presented in this chapter is given in appendix A.1.

## Chapter 5

# A Genetic Map of the X Chromosome

## Table of Contents

	Page
<b>5.1 Summary</b>	110
<b>5.2 Introduction</b>	111
<b>5.3 Materials and Methods</b>	113
5.3.1 Construction of the genetic map	113
5.3.2 Physical mapping	114
5.3.3 Phase analysis	114
<b>5.4 Results</b>	115
5.4.1 Genetic map of the X chromosome	115
5.4.2 Physical mapping data	115
5.4.3 Marker heterozygosity	115
5.4.4 Meiotic breakpoint map	122
<b>5.5 Discussion</b>	134
5.5.1 Chromosome arm lengths	134
5.5.2 Distribution of new markers	134
5.5.3 Physical mapping data	134
5.5.4 Recombination across the dystrophin locus	135
5.5.5 Recombination across the centromere	136
5.5.6 Comparisons with other maps	137
5.5.6.1 The order of loci	137
5.5.6.2 The distances between loci	139
5.5.6.3 Factors contributing to differences between maps	140
5.5.7 Meiotic breakpoint map	146
<b>5.6 Conclusion</b>	148

## 5.1 Summary

The polymorphic markers described in the previous chapter were incorporated into a genetic map of the X chromosome along with a number of pre-existing PCR and RFLP-based loci. At completion, the map consisted of 78 markers, including 69 that are represented by PCR-based microsatellites. It spans a distance of 236 centimorgans, from the *XG* locus near *Xpter* to *DXS52* near *Xqter*. These two extreme loci encompass almost 100% of the sex-limited portion of the X chromosome (that is, the pseudoautosomal regions at the p and q termini are not included). The framework map consists of 30 loci whose order is supported by odds of at least 1000:1 against any alternative order. The average distance between these uniquely placed loci is 8.1 cM. The remaining 48 other markers were positioned with odds of less than 1000:1. Considering all 78 loci, the average spacing is 3.1 cM. The average heterozygosity of the PCR-based markers is 66%.

This map of highly polymorphic markers was constructed to expedite the mapping by linkage of X-linked disease loci (Chapters 6 and 8). It also can be used to refine the existing localisations of disease loci, and thus potentially increase the accuracy of genetic diagnoses based on linked markers. Ultimately, by facilitating the regional localisation of disease genes the positional cloning of these genes may be achieved. The published version (Donnelly *et al.*, 1994b) of the map presented in this chapter is given in appendix A.1.

## 5.2 Introduction

Prior to starting this project a whole human genome genetic map, based on RFLP markers, had been published by Donis-Keller *et al.* (1987). Soon after starting this project two other whole genome maps were published by the NIH/CEPH Collaborative Mapping Group (1992) and Weissenbach *et al.* (1992). The NIH/CEPH map was constructed primarily from RFLP and VNTR markers but the Weissenbach map was composed of only microsatellite, PCR-based, markers. The advantage of the X chromosome map presented in this chapter is that it integrated most of the X-linked loci of the Weissenbach map with some pre-existing, pre-dominantly RFLP, loci (at which, in some cases, microsatellites have subsequently been found) which thus served to relate any new markers to well known ones. After publication of the genetic map of the X chromosome, other whole genome maps were published by Matise *et al.* (1994), Buetow *et al.* (1994), Gyapay *et al.* (1994), Spurr *et al.* (1994) and Dib *et al.* (1996). Those genetic linkage maps specifically of the entire X chromosome (except the pseudoautosomal regions) have been constructed by Drayna *et al.* (1984), Drayna and White (1985) (both of which used RFLP-based markers), Donnelly *et al.* (1994b; this Chapter) and, most recently and comprehensively, Fain *et al.* (1995).

There are two levels of statistical support for positioning a locus on a map: global and interval (Keats *et al.*, 1991). Global support is a measure (expressed in lod scores) of the evidence that a locus belongs to a map. Interval support is a measure of the evidence that a locus belongs to a particular interval on the map. Loci that have an interval support lod score of  $\geq 3.0$  are "framework" loci. A map of framework loci is a framework map. Such a map is considered to be highly reliable in terms of locus order. In comparison, a "comprehensive" map comprises the loci of a framework map in addition to any other syntenic loci. These additional loci are placed in their most likely position with less confidence ( $\text{lod} \leq 3.0$ ), than is the case with framework loci. This is because fewer recombination events are known to have occurred between these and the framework loci because the markers: 1) are less informative; 2) have only been genotyped on a subset of the CEPH families; or 3) map very close to one or more existing framework loci.

Apart from providing the order of markers that can be used in the localisation of disease genes, the construction of the maps themselves can reveal interesting biological phenomena, such as: the variation in the frequency of recombination along chromosomes

and differences in recombination between the sexes; the relationship between genetic distance and physical distance; variations in the distribution of markers; and the frequency of mutational events. In addition, PCR-based genetic maps can be easily integrated with physical maps since the primers used for genotyping are also applicable as STSs in contig assembly.

### 5.3 Materials and Methods

Only brief accounts of the methods used are given below, further details are given in the Chapter 3 sections indicated.

#### 5.3.1 Construction of the genetic map

The genetic map, incorporating markers characterised and genotyped by the candidate (Chapter 4), was computed by Dr Helen Kozman as described in 3.2.23.3. A total of 78 loci were used to build the map. Analysis was restricted to those markers which had their microsatellite genotypes determined on the CEPH families or those markers with characterised microsatellites useful for disease mapping but only present on the CEPH database as RFLP genotypes. Twenty of the microsatellite markers used in the analysis were genotyped in the Department of Cytogenetics and Molecular Genetics (WCH). The remainder were extracted from contributions made by other CEPH collaborators to the CEPH database (version 5.0), except for the markers isolated and characterised by Weissenbach *et al.* (1992) which appear in version 6.0.

Suspected genotyping errors in markers obtained from the CEPH database were detected using the CHROMPICS option of CRI-MAP. These were displayed as two or more apparent crossovers within a small chromosomal region (less than 15 cM). Since they were very likely due to genotyping error they were dealt with by removing the suspect genotype from the analysis. The marker genotypes of individuals with new mutations (Chapter 4; 4.4.2.3) were also omitted from the analysis. Suspected genotyping errors in markers submitted to the CEPH database from the Department of Cytogenetics and Molecular Genetics were genotyped again and corrected or omitted from the analysis.

Genotypes for several markers had previously been incorporated into multipoint linkage maps of specific regions. These markers were: *DXS292* and *DXS297* (Richards *et al.*, 1991b), *FRAXAC1* and *FRAXAC2* (Richards *et al.*, 1991a) and *ALAS2* (Cox *et al.*, 1992).

Following construction of the map, the data generated were analysed by the candidate.

### 5.3.2 Physical mapping

Fluorescence *in situ* hybridisation (FISH) of *lambda* clones, isolated by the candidate, was performed as described in 3.2.17 by Liz Baker, Helen Eyre or Paul Yu, Department of Cytogenetics and Molecular Genetics (WCH).

FISH to a metaphase spread from an individual expressing the fragile X E site, was performed by Paul Yu using the *lambda* clone DNA containing microsatellite *DXS1123* (3.2.17), in order to determine its orientation relative to this site (Mulley *et al.*, 1995; appendix A.2). The *FRAXE* site was induced as given in Sutherland and Baker (1992).

### 5.3.3 Phase analysis

Meiotic breakpoint maps of 8 CEPH families were constructed, by the candidate, from the CHROMPICS output (generated by Dr Helen Kozman) of the CRI-MAP program (3.2.23.3). These are the largest CEPH families and were used in the construction of the Genethon genetic maps (Weissenbach *et al.*, 1992; Gyapay *et al.*, 1994; Dib *et al.*, 1996).

## 5.4 Results

### 5.4.1 Genetic map of the X chromosome

The genetic map locations and distances of all 30 framework markers are given in Figure 5.1. The idiogram shown in this figure is a composite of the Giemsa staining pattern at the 550 and 850 band resolution for metaphase chromosomes (ISCN, 1995). The order of these markers is supported by odds of at least 1000:1 against any alternative order. This framework map extends from the *XG* locus near Xpter to *DXS52* near Xqter. The average distance between the uniquely placed framework loci is 8.1 cM. An additional 48 other loci were positioned against this background map, but with odds of less than 1000:1. Considering all 78 loci comprising the comprehensive map (Fig. 5.1 and Table 5.1), the total distance spanned is 236 cM and the average inter-locus spacing is 3.1 cM, however the markers are *not* evenly distributed. There is one interval greater than 15 cM (19.5 cM between *XG* and *DXS278*) and there are 18 groupings of 2 or more markers with no recombination between them. The local support for order of each locus in the comprehensive map is given as odds against inversion with adjacent loci in Table 5.1. Genetically, the short (p) arm is slightly longer than the long (q) arm (121 cM *versus* 114 cM).

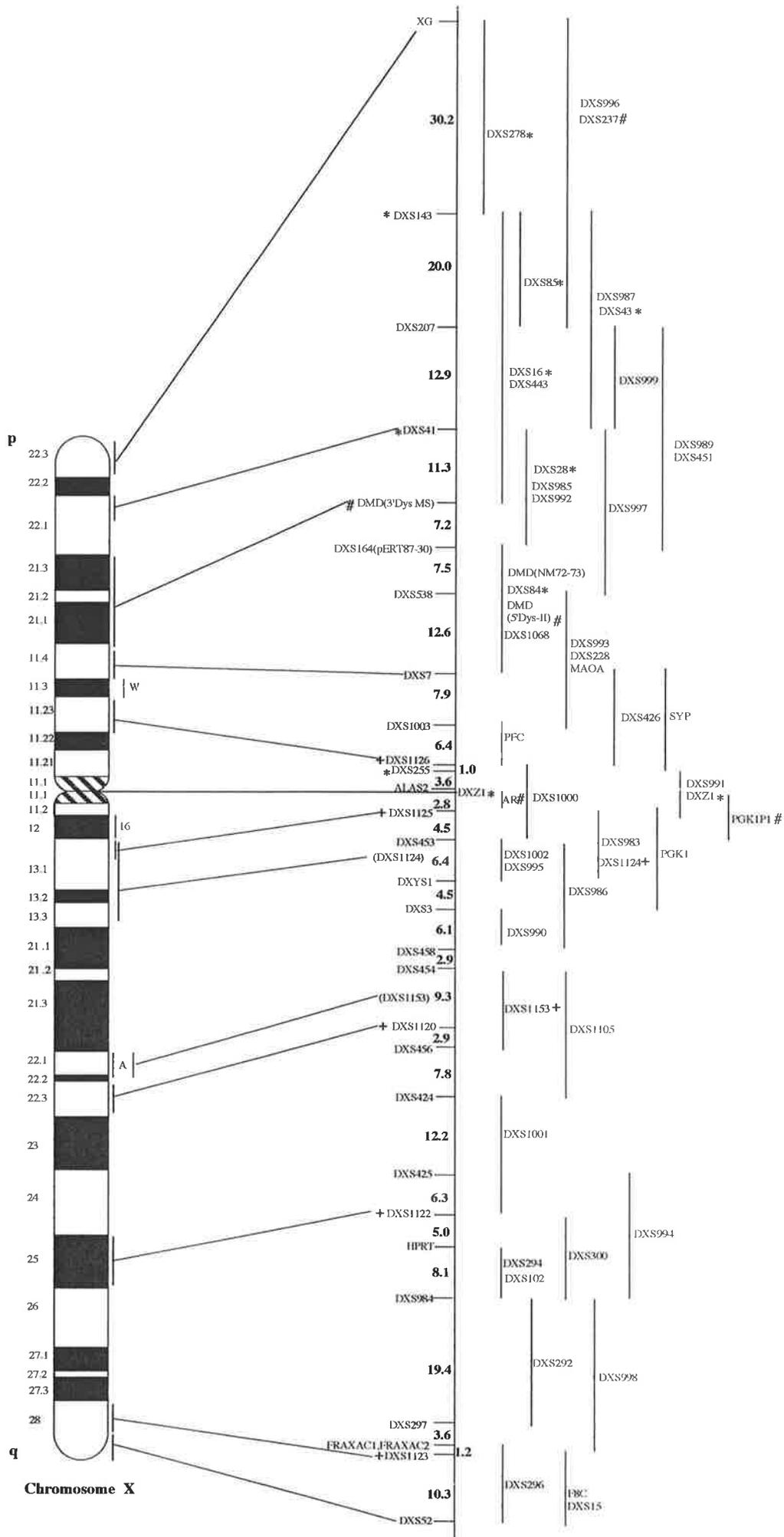
### 5.4.2 Physical mapping data

Fifty of the 78 loci used in the map have been physically mapped. The cytogenetic localisations of markers were taken from Nelson *et al.* (1995). The cytogenetic localisations of 7 of those markers isolated and characterised by the candidate (*DXS1123*, *DXS1124*, *DXS1125*, *DXS1126*, A, W and 16) had, prior to the Nelson *et al.* (1995) publication, been determined by FISH and are shown in Figure 5.1. Microsatellite *DXS1123* is distal to the fragile X E site in Xq28.

### 5.4.3 Marker heterozygosity

The average heterozygosity of the PCR-based markers in the map is 66% ranging from 33% for *DXS1122* and *DXS994* to 95% for the androgen receptor (*AR*) locus. Of the 30 RFLP and PCR-based framework markers 13 are index markers, that is, they have heterozygosities of at least 70%.





**Table 5.1**  
**Distance (cM) between loci in the comprehensive genetic map together**  
**with odds against inversion with adjacent loci and approximate cytogenetic**  
**localisation (Xpter to Xqter)**

Locus	cM	Odds (:1)	Cytogenetic localisation
XGJ			p22.33
<i>DXS278*</i>	19.5	10 <sup>11</sup>	
<i>DXS996</i>	7.7	10 <sup>2</sup>	p22.32
<i>DXS237</i>	0.5	1.0	p22.32
<i>DXS143*</i>	2.5	10 <sup>3</sup>	p22.32-q31
<i>DXS85</i>	9.9	10 <sup>6</sup>	p22.31
<i>DXS16</i>	8.4	10 <sup>7</sup>	p22.31
<i>DXS987</i>	1.7	4.0	p22.31-q22.2
<i>DXS207</i>	0.0	1.0	p22.2
<i>DXS43</i>	0.0	1.0	p22.2
<i>DXS999</i>	2.7	10 <sup>5</sup>	p22.2
<i>DXS443</i>	0.0	1.0	p22.13
<i>DXS41*</i>	10.2	5.0	p22.13
<i>DXS989</i>	6.3	20.0	p22.13
<i>DXS451</i>	0.0	1.0	p22.13
<i>DXS28</i>	3.9	2.0	p21.3
DMD (3'DYS MS)	1.1	1.0	p21.3-q21.1
<i>DXS985</i>	0.9	1.0	
<i>DXS992</i>	0.0	1.0	p21.3-q21.1
<i>DXS997</i>	4.7	10 <sup>3</sup>	p21.3-q21.1
<i>DXS164*</i>	1.6	10 <sup>3</sup>	
DMD (NM72-73)	6.8	10 <sup>7</sup>	
<i>DXS84</i>	0.7	10 <sup>4</sup>	
	0.0	1.0	

**Table 5.1 (cont'd)**

Locus	cM	Odds (:1)	Cytogenetic localisation
DMD (5'DYS-II)			p21.1
<i>DXS538</i>	0.0	1.0	
<i>DXS1068</i>	1.6	10 <sup>4</sup>	p11.4
<i>DXS993</i>	9.8	10 <sup>14</sup>	
<i>DXS228</i>	1.2	10 <sup>3</sup>	p11.4
<i>DXS7</i>	0.0	1.0	p11.4
MAOA	0.0	1.0	p11.4-q3
<i>DXS1003</i>	7.9	10 <sup>14</sup>	p11.3
PFC	5.2	4.0	p11.3-q23
<i>DXS426</i>	0.0	1.0	p11.3-q23
<i>DXS1126</i>	1.2	10 <sup>4</sup>	p11.23
SYP*	0.0	1.0	p11.23
<i>DXS255*</i>	1.0	10 <sup>4</sup>	p11.23
ALAS2	3.6	10 <sup>5</sup>	p11.21
<i>DXS991</i>	0.0	1.0	p11.21
<i>DXZ1*</i>	0.0	1.0	CEN
AR	0.4	10 <sup>5</sup>	q11.2-q12
PGK1P1	0.0	1.0	q12
<i>DXS1125</i>	2.4	10 <sup>3</sup>	q12
<i>DXS1000</i>	0.0	1.0	
<i>DXS983(DXS453)</i>	4.5	10 <sup>4</sup>	q12-q13.1
<i>DXS1124</i>	0.3	1.0	
PGK1	0.0	1.0	q13.3
	2.6	10 <sup>8</sup>	

**Table 5.1 (cont'd)**

Locus	cM	Odds (:1)	Cytogenetic localisation
<i>DXS986</i>	0.0	1.0	q21.1
<i>DXS995</i>	0.0	1.0	q21.1
<i>DXS1002</i>	3.5	10 <sup>3</sup>	q21.2-q21.31
<i>DXYS1</i>	4.5	10 <sup>5</sup>	q21.31
<i>DXS3</i>	0.7	1.0	q21.33
<i>DXS990</i>	5.4	10 <sup>8</sup>	q21.33
<i>DXS458</i>	2.9	10 <sup>4</sup>	q21.33
<i>DXS454</i>	7.2	10 <sup>14</sup>	q21.33
<i>DXS1153</i>	2.1	1.0	q22.1
<i>DXS1105(DXS571)</i>	0.0	1.0	q22.3
<i>DXS1120</i>	2.9	10 <sup>4</sup>	q22.3
<i>DXS456</i>	7.8	10 <sup>12</sup>	q22.3
<i>DXS424</i>	10.2	10 <sup>12</sup>	q23
<i>DXS1001</i>	2.0	10 <sup>2</sup>	q24
<i>DXS425</i>	6.3	10 <sup>8</sup>	q25
<i>DXS1122</i>	5.0	10 <sup>1</sup>	
<i>DXS994</i>	0.0	1.0	q25-q.26.1
HPRT	0.0	1.0	q26.1
<i>DXS300</i>	4.1	10 <sup>8</sup>	q26.2
<i>DXS294</i>	1.1	10 <sup>2</sup>	q26.3
<i>DXS102</i>	2.9	10 <sup>5</sup>	q26.3
<i>DXS984</i>	14.9	10 <sup>11</sup>	q27.1
<i>DXS292</i>	4.5	10 <sup>9</sup>	q27.2

**Table 5.1 (cont'd)**

Locus	cM	Odds (:1)	Cytogenetic localisation
<i>DXS297</i>			q27.3
	3.6	3.0	
<i>DXS998</i>			q27.3
	0.0	1.0	
<i>FRAXAC1</i>			q27.3
	0.0	1.0	
<i>FRAXAC2</i>			q27.3
	1.2	10 <sup>8</sup>	
<i>DXS296</i>			q28
	0.0	1.0	
<i>DXS1123</i>			q28
	10.3	10 <sup>24</sup>	
<i>DXS52</i>			q28
	0.8	10 <sup>1</sup>	
<i>DXS15</i>			q28
	0.0	1.0	
F8C			q28

¶ Locus associated with serological polymorphism.

\* Locus associated with RFLP.

All other loci have microsatellite PCR-based markers.

Markers associated with genes or pseudogenes (PGK1P1) are shown in plain text.

In relation to marker order, this Table is an updated version of that appearing in Donnelly *et al.*, (1994b). Where marker order was not resolved by linkage (ie 1:1 odds), the order now given is consistent with the physical order given by the consensus map of Nelson *et al.* (1995). Note that the position of *DXS1000* is not in accordance with that of Nelson *et al.* (1995), refer to 5.5.6.1.

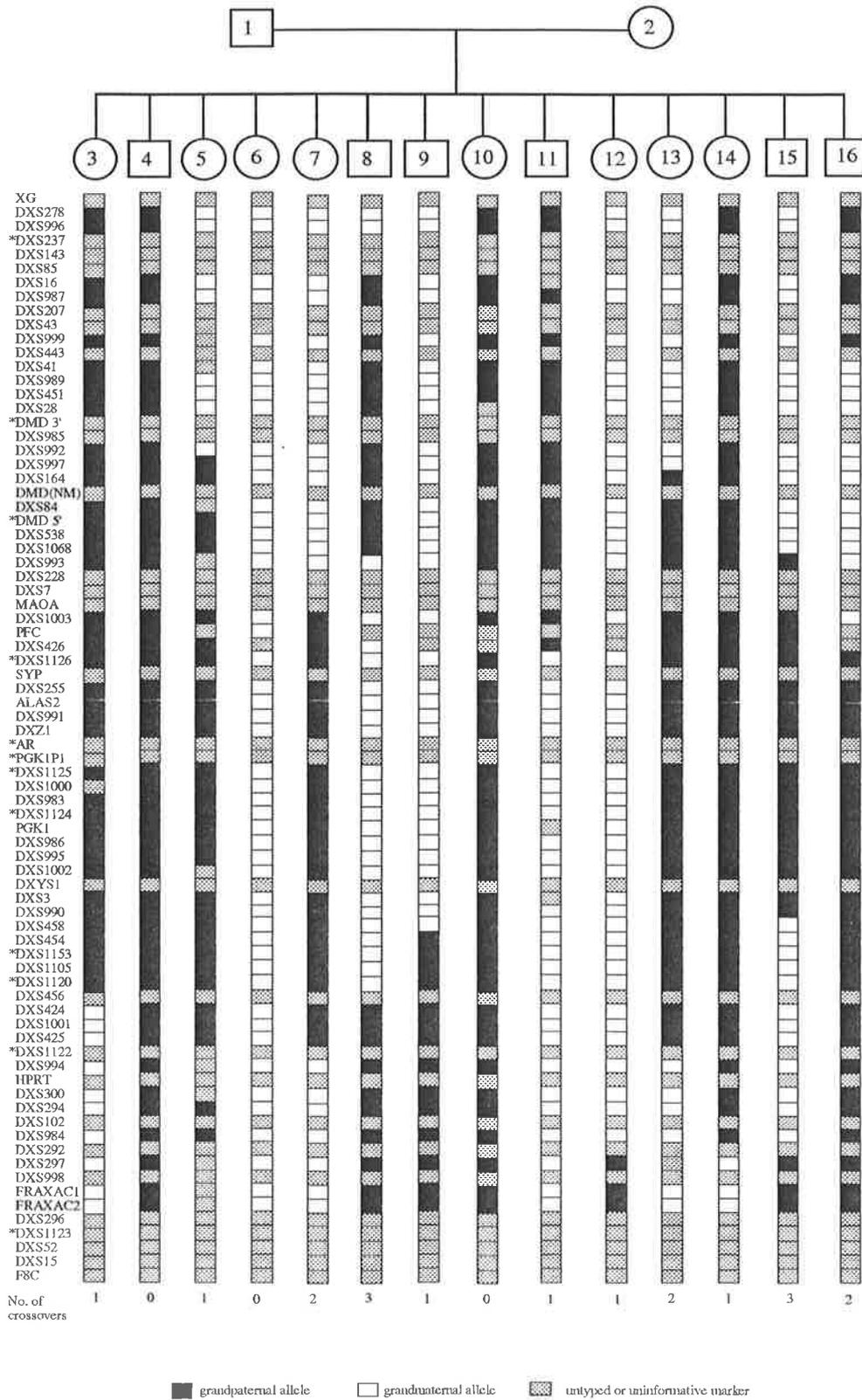
#### 5.4.4 Meiotic breakpoint map

To expedite the genetic mapping of newly isolated markers that have been assigned to a position on the X chromosome by physical or preliminary genetic mapping, and to investigate the frequency and distribution of recombination along the X chromosome, meiotic recombination maps can be constructed. Phase analysis, using the CHROMPIC option of CRI-MAP, permitted the determination of the most likely grandparental origin of alleles and thus the crossover points along the X chromosomes of the children of the 8 largest standard CEPH families (Fig. 5.2).

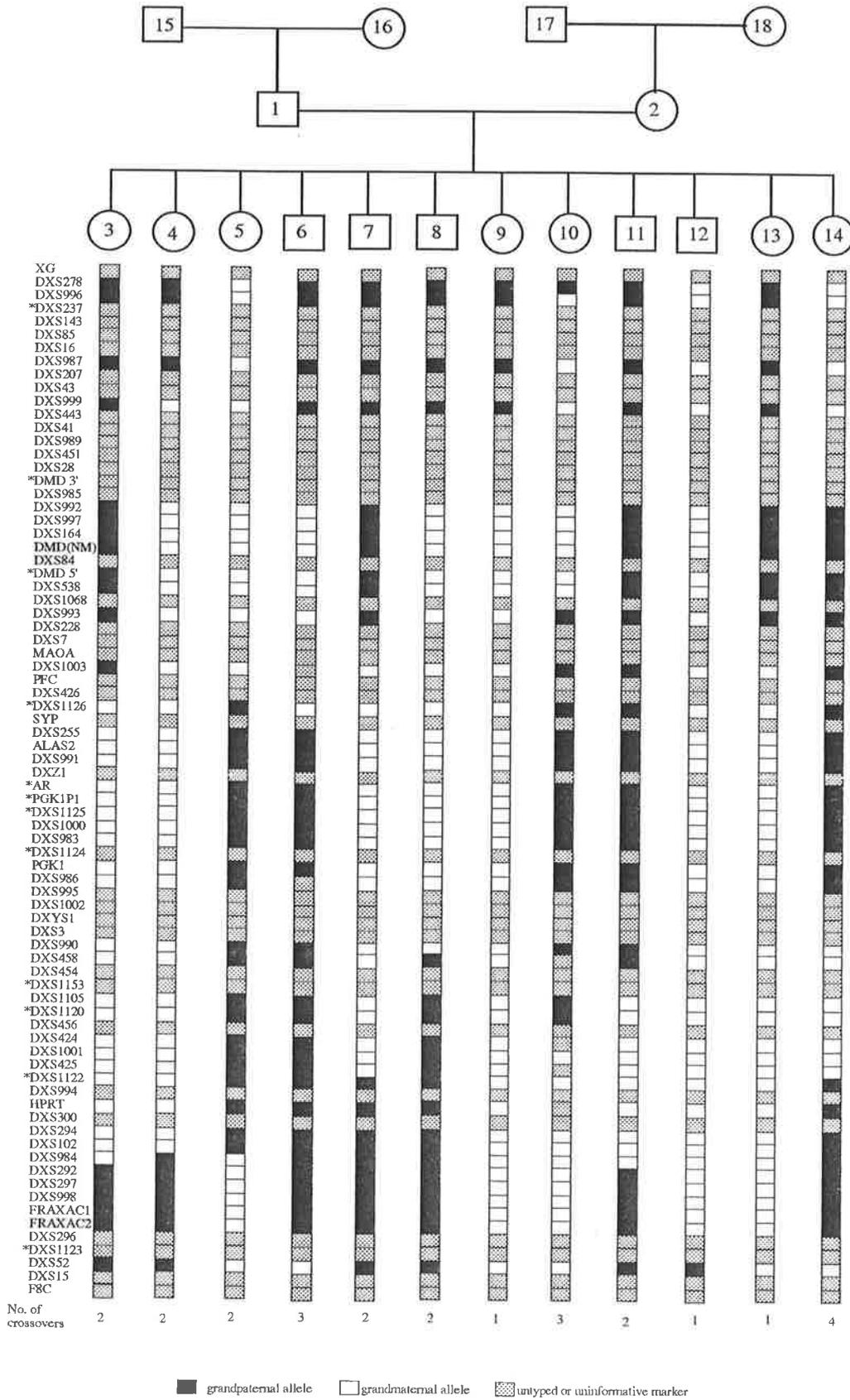
At the level of resolution provided by the map, there were 168 recombination events in the maternally-inherited X chromosomes of the 94 CEPH children shown in Fig. 5.2. This gives an average of 1.8 recombination events *per* chromosome. The number of crossovers *per* chromosome ranges from 0 to 5, with the most frequent being 2 (36%) (Table 5.2; Fig. 5.3). The other frequencies of the recombination events are given in Table 5.2. The maximum number of crossovers (5) occurred in individual 1413/04. The smallest interval in which a double crossover has occurred is at most 7.4 cM between *DXS1003* and *DXS255* in person 1332/07. This is assuming that the genotyping of locus *PFC* (the only informative or typed marker of the four within this region) is correct and that the observations are not the result of a mutation event.

**Figure 5.2** Meiotic breakpoint maps of 8 of the CEPH pedigrees, generated using CHROMPIC. The idiograms beneath each child represent the X chromosome inherited from the mother (02). The order of the loci is the same as that given in Table 5.1. Those markers genotyped by the candidate are indicated (\*). The genotypes of the other markers were obtained from the CEPH database (version 6). Error checking of telomeric recombination events and double recombination events occurring within relatively short distances has not been performed.

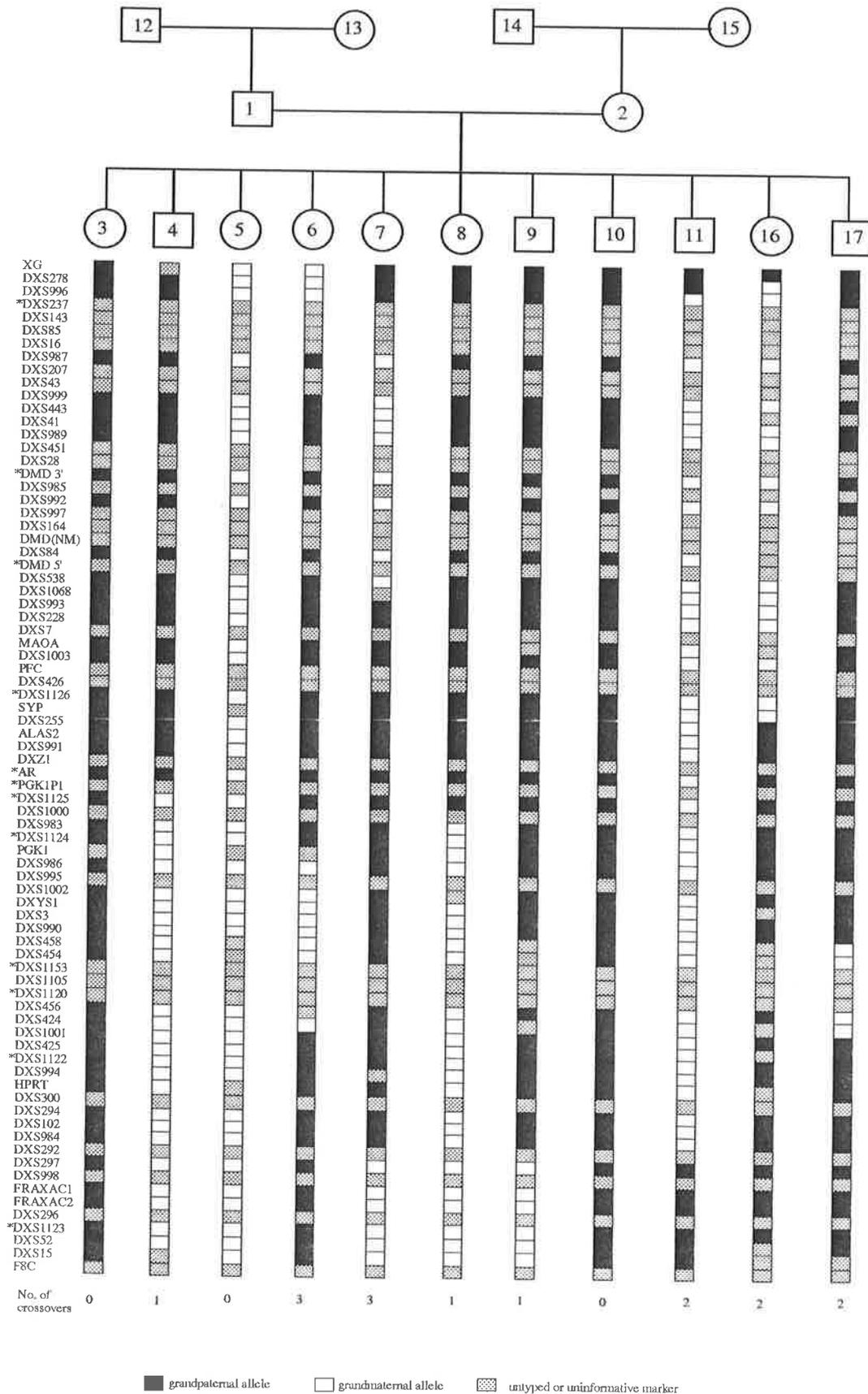
CEPH pedigree 102



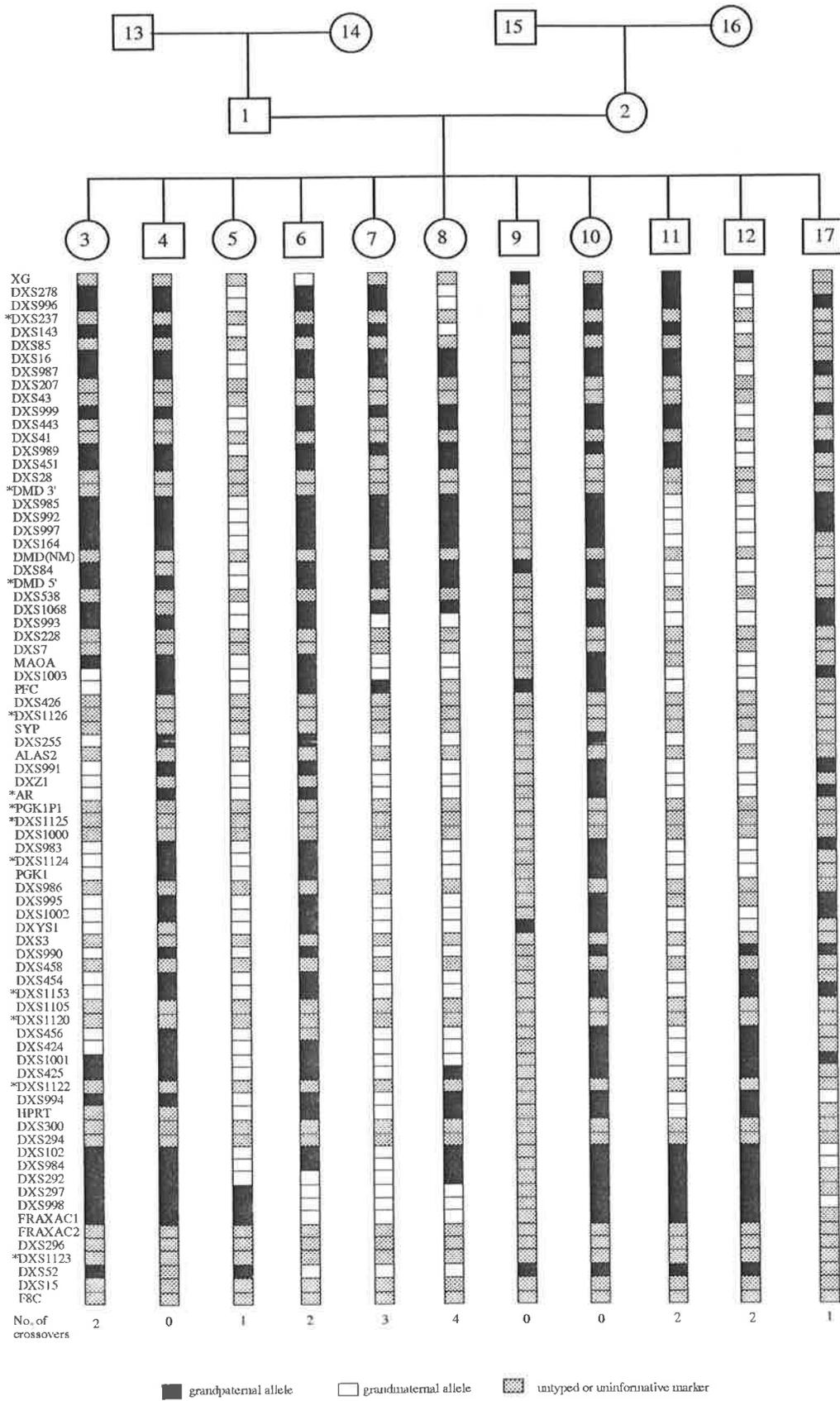
CEPH pedigree 884



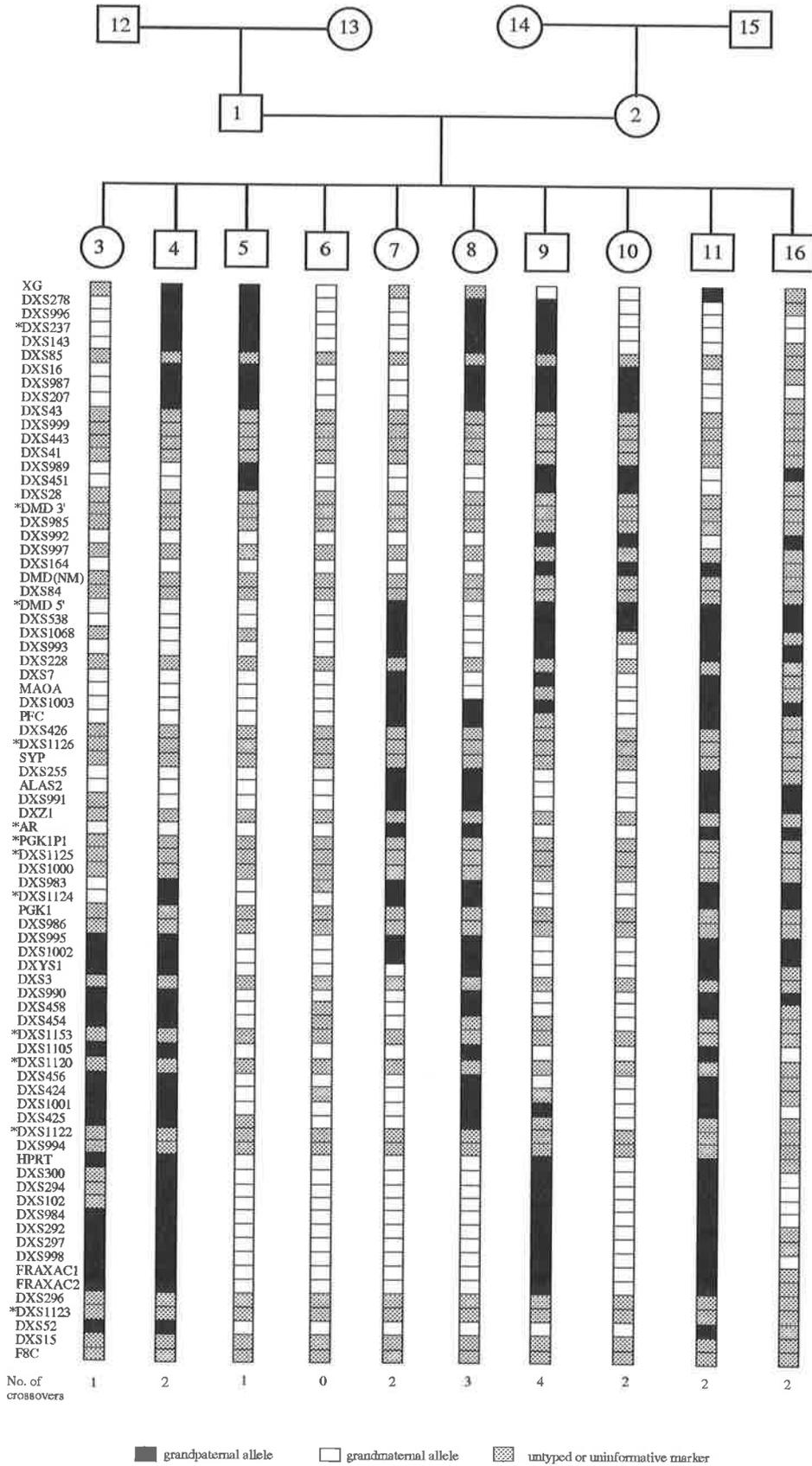
CEPH pedigree 1331



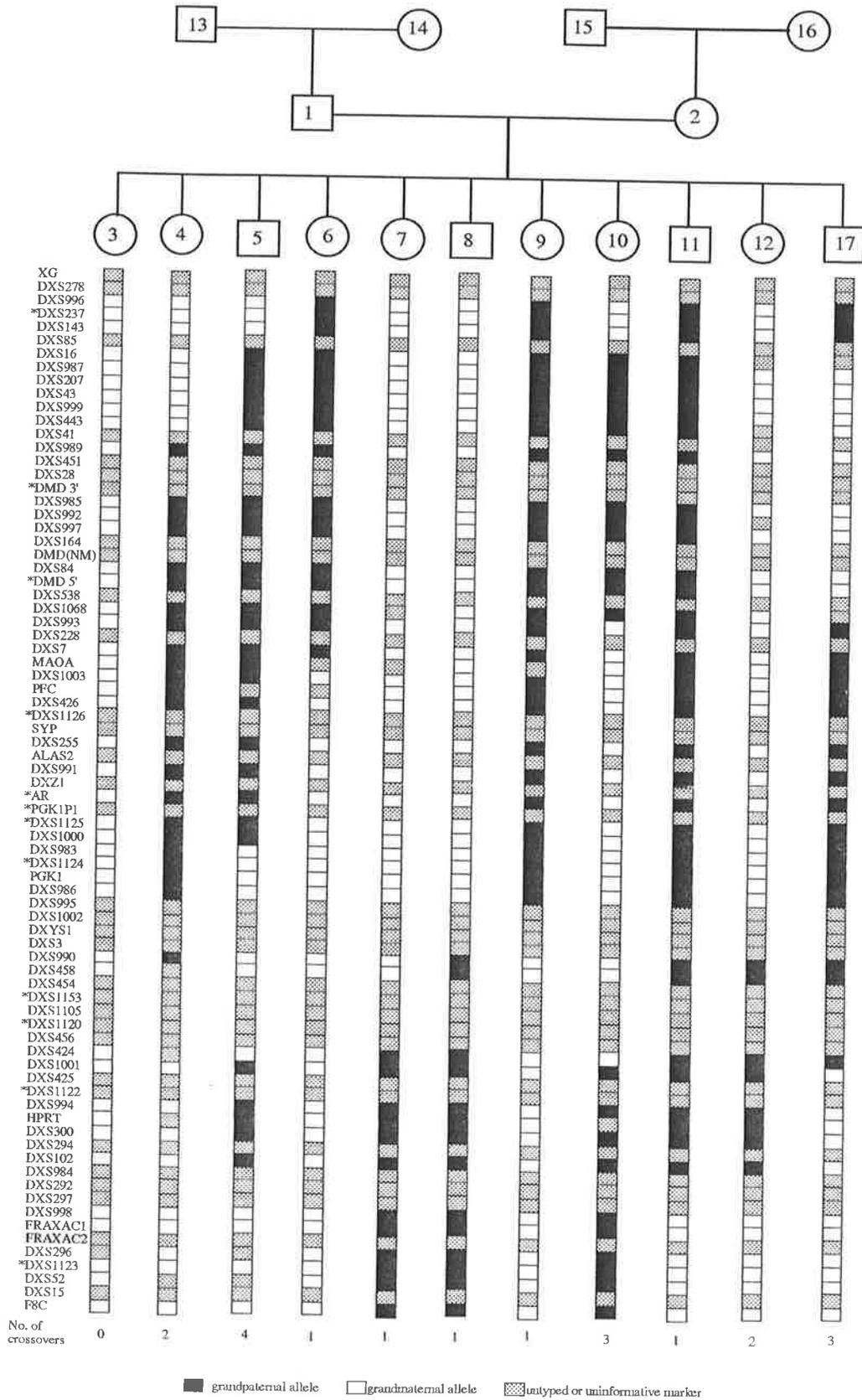
CEPH pedigree 1332



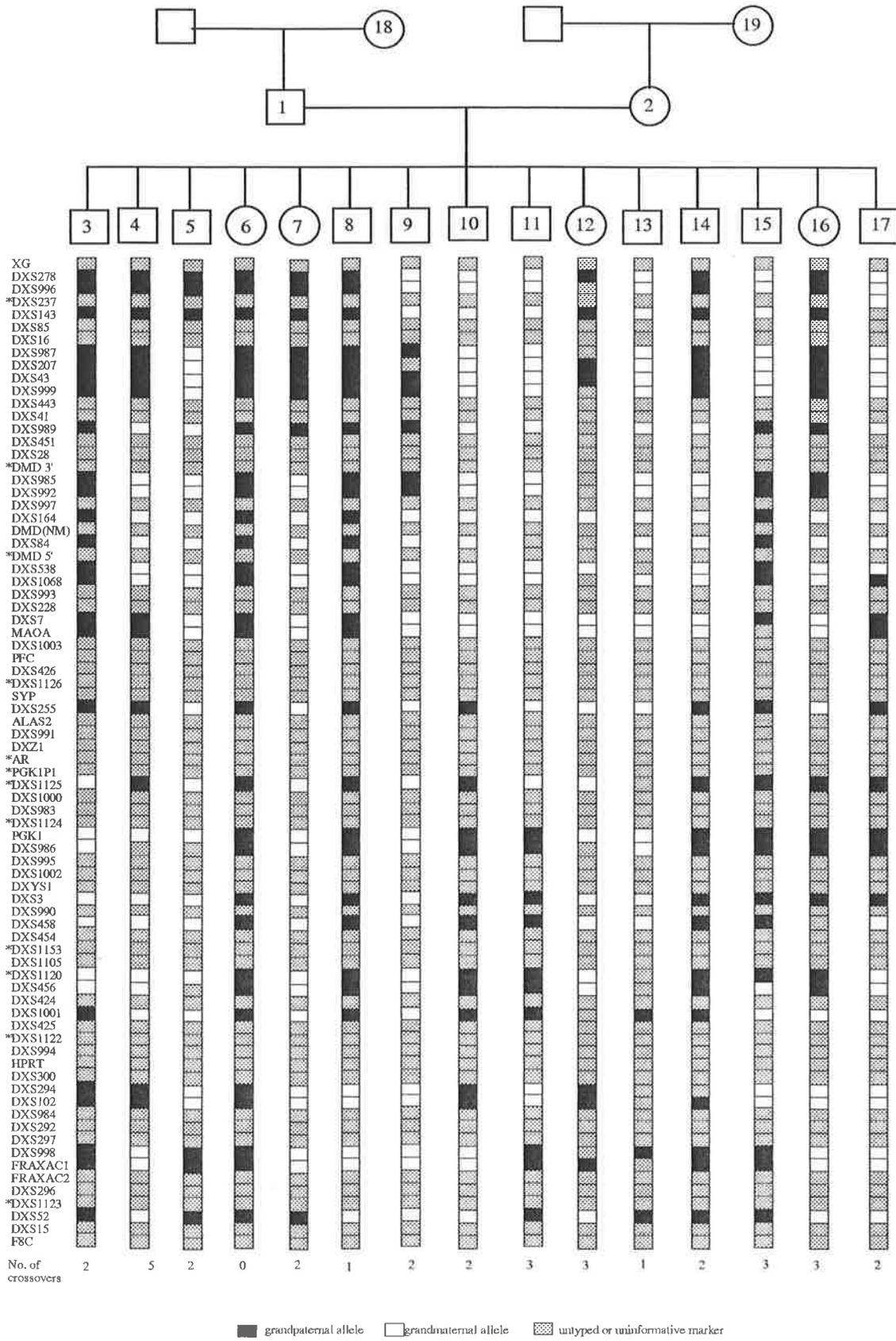
CEPH pedigree 1347



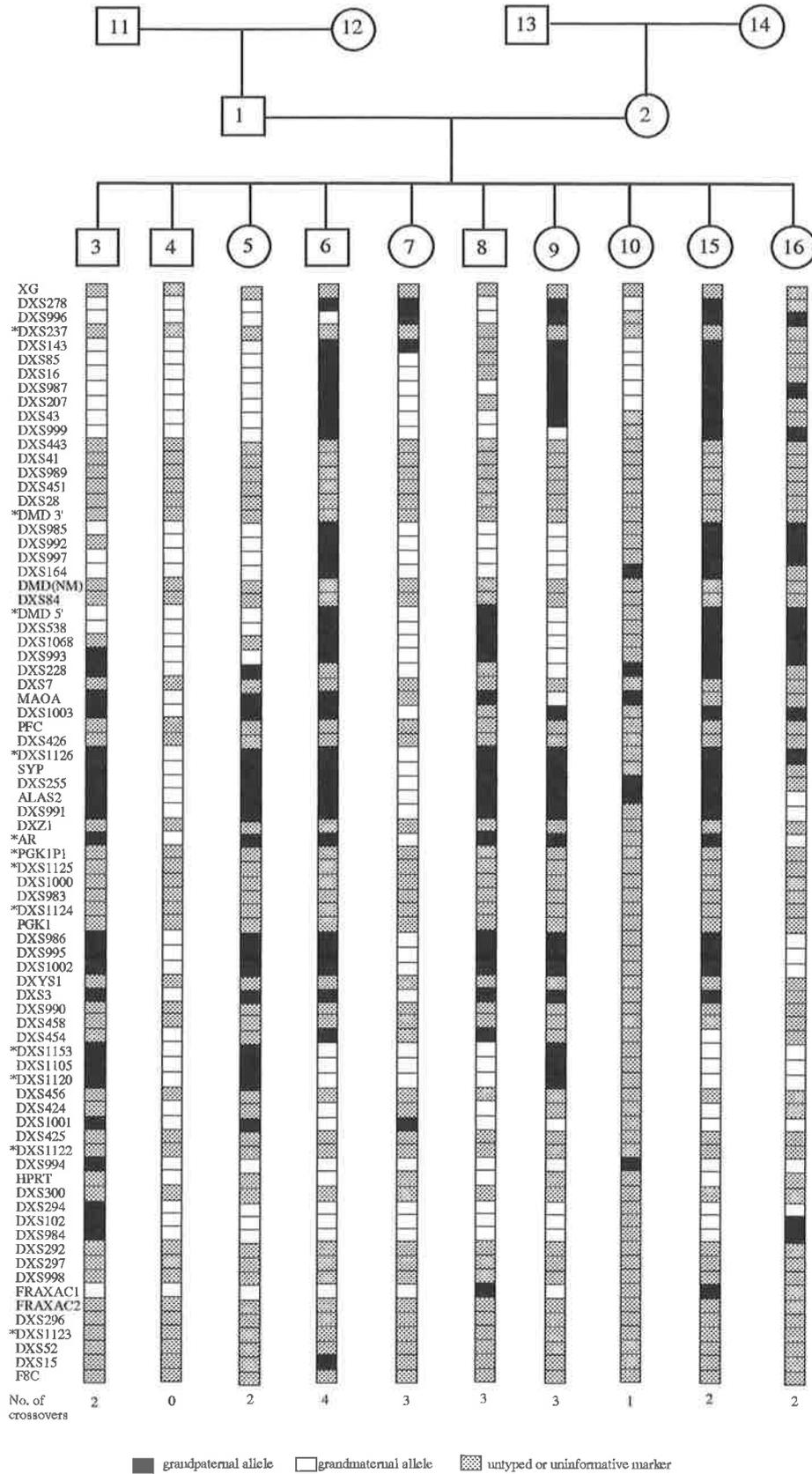
CEPH pedigree 1362



CEPH pedigree 1413



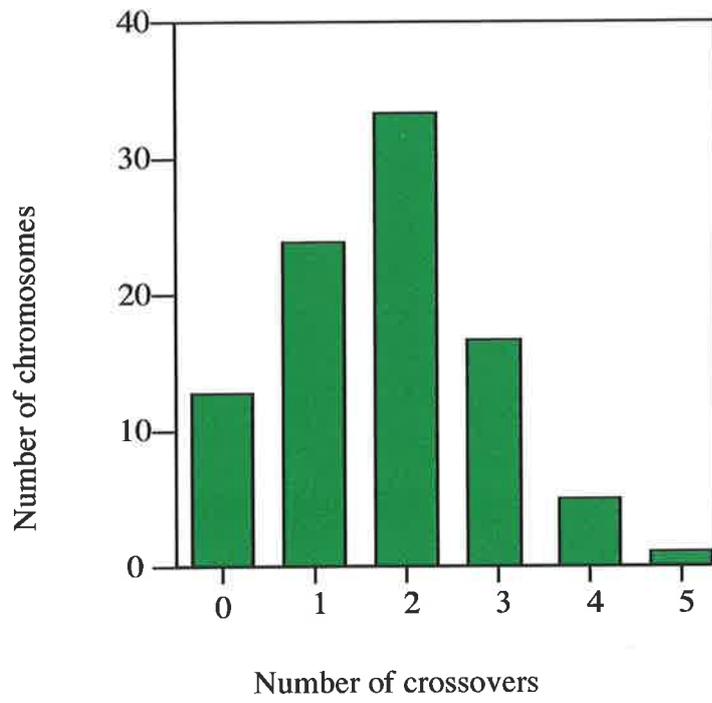
CEPH pedigree 1416



**Table 5.2**  
**Recombination events within the 8 largest CEPH families**

Number of recombination events per chromosome	Number of chromosomes	Frequency (%)
0	13	13.8
1	24	25.5
2	34	36.2
3	17	18.1
4	5	5.3
5	1	1.1
Total	94	100

Note: these figures were derived from the pedigrees in Fig. 5.2.



**Figure 5.3** Number of crossovers *per* X chromosome.

## 5.5 Discussion

### 5.5.1 Chromosome arm lengths

Despite the disparate cytogenetic lengths of the arms of the sub-metacentric X chromosome as seen at metaphase, and as shown in Figure 5.1, genetically, the short arm is slightly longer than the long arm (121 cM *versus* 114 cM). This observation of roughly equivalent lengths was also made by Drayna and White (1985) but is in disagreement with the oft quoted lengths of Morton (1991), which are: 87 cM for the short arm and 133 cM for the long arm. The most recent genetic map of the X chromosome has the short arm at 87 cM and the long arm 103 cM in length (Fain *et al.*, 1995). If the short arm is truly genetically shorter than the long arm, the discrepancy between maps may be explained by: 1) genotyping errors; 2) a difference in the distribution of markers within each arm; or 3) as about equal numbers of loci were investigated on each arm (38 *versus* 39) and given that the short arm is physically smaller, the higher density of short arm markers investigated may have revealed a greater proportion of recombination events.

### 5.5.2 Distribution of new markers

Ten of the 11 new microsatellite markers described in the last Chapter (Table 4.3), were either physically and/or genetically mapped. Eight of these (Table 5.1; Fig. 5.1) mapped to Xq. The physical length of each of the arms of the X chromosome have been calculated as 102 Mb for Xq and 62 Mb for Xp, a ratio of 1:1.6 (Morton, 1991). Therefore one would expect that, upon the random isolation of 10 markers from a library of X loci, on average, 4 of these would lie on the p arm and 6 on the q arm. The actual results are not significantly different from the expected ( $\chi^2$   $p < 0.50$ ). Skewed localisations of randomly isolated markers have been reported and could be due either to actual differences in the distribution of markers along chromosomes or to some bias in the methods used to isolate them (Dracopoli *et al.*, 1988; Hazan *et al.*, 1992; Kwiatkowski *et al.*, 1992; Weissenbach *et al.*, 1992; Buetow *et al.*, 1993).

### 5.5.3 Physical mapping data

Most of the 78 loci used in the map have been physically mapped. The cytogenetic

localisations of the markers were taken from the consensus map of Nelson *et al.* (1995). This consensus map also includes the cytogenetic data of those markers characterised by the candidate.

The marker *DXS1123* was of particular interest as it genetically mapped near to the fragile X E site in Xq28, and as mentioned above, was subsequently shown to be distal, by FISH analysis (Mulley *et al.*, 1995; appendix A.2). At the time the *FRAXE* site was thought to be involved in the manifestation of a mild form of inherited non-syndromal mental retardation. Since then the *FRAXE* site has been cloned (Knight *et al.*, 1993) and the associated gene identified (Gecz *et al.*, 1996; Gu Y *et al.*, 1996).

#### 5.5.4 Recombination across the dystrophin locus

Duchenne muscular dystrophy is the most common lethal X-linked disorder, having an incidence of 1/3500 male births. A multitude of often uncharacterised lesions within the dystrophin gene can result in either the Duchenne or Becker forms of muscular dystrophy (MIM 310200; Koenig *et al.*, 1989) thus linkage analysis, as distinct from a mutation screening approach, is required for genetic counselling in about 40% of cases. The gene spans about 2.3 Mb in Xp21 (Koenig *et al.*, 1987; Den Dunnen *et al.*, 1989). Previous calculations of the genetic risk to individuals with relatives suffering from Duchenne or Becker muscular dystrophies were often based on recombination frequencies observed in disease manifesting families themselves (for example Chen *et al.*, 1989). By genotyping the 5'Dys-II and 3'Dys MS microsatellites, which lie at the ends dystrophin gene, in the CEPH population and incorporating them by multipoint analysis into a high density map it was expected that a more precise and statistically robust estimate of the recombination frequency/genetic distance across this region may be obtained. Assuming the genome average recombination rate applies to this region, one would expect a genetic distance of 2.4 cM (1 cM/Mb) however the genetic distance observed in the CEPH families was 14.7 cM (Table 5.1). This gives an average of 6.1 cM/Mb, six times greater than the genome average. This relatively high ratio is the same as a previous estimate of 6 cM/Mb (confidence interval: 4.1-24.4 cM) by Abbs *et al.* (1990). These authors also used the CEPH families but relied on relatively low informative RFLP loci that were distal to 3'Dys MS (ie *DXS41*) and lay between this locus and 5'Dys-II (*DXS503-DXS164-DXS142*).

Oudet *et al.* (1992a) genotyped 5'Dys-II and 3'Dys MS and several intragenic markers, using the CEPH population, also to determine the recombination frequency across the gene. They only typed 5'Dys-II and 3'Dys MS in families not informative for adjacent markers. They, however, arrived at the smaller ratio of 4 cM/Mb across the dystrophin gene.

Whatever the actual figure is, it is clear that recombination across the dystrophin gene occurs with a relatively high frequency, which makes the use of flanking *and* intragenic markers essential for carrier determination in those families where the specific mutation is unknown. In contrast, the possibility of intragenic recombination is considered negligible for most genetic disorders which can be diagnosed by linkage using *flanking* markers only.

#### 5.5.5 Recombination across the centromere

A reduced frequency of chiasmata near the centromeres of human chromosomes has been reported. These observations have been confirmed by comparisons between genetic and physical maps (Murray *et al.*, 1994). The locus that represents the X centromere is *DXZI*. This is composed of about 5,000 copies of a 2 kb repeat that is itself made of 12 monomers, each about 171 bp in length. As expected, this locus is polymorphic and, from observations of 29 X chromosomes, it ranges from about 1,400 to 3,700 kb in size (Mahtani and Willard, 1990). According to Mahtani and Willard (1988) the physical distance from *DXS7* (Xp11.4) to *PGK1* (Xq13.3) has been estimated at 40 Mb, although no evidence for this figure is given. The genetic distance across this region is only 26.5 cM (Table 5.1). If the assumption is made that on average 1 Mb = 1 cM, one would expect a genetic distance of 40 cM, hence it appears that there is a lower than expected frequency of recombination occurring near the X centromere, although there may be variations at a finer level of resolution. Evidence for such variation has been provided by Weeks *et al.*, (1995) who defined two regions of reduced recombination: around *DXZI* and around *PGK1* (from *DXS441* to *DXS995*).

Until the distance between the centromere and unique flanking sequences has been spanned by YAC contigs, no firm conclusion regarding the relationship between physical and genetic distances can be made.

### 5.5.6 Comparisons with other maps

Many genetic maps of sub-chromosomal regions of the X chromosome exist. These are most often constructed using a small number of markers genotyped in the families suffering from diseases that map to these regions. The features of: 1) the often low heterozygosity of the markers (until recently mainly RFLP-based), 2) the small sizes of the families, and 3) the small number families associated with rare disorders, combine to result in relatively few informative meioses and hence relatively low statistical support for marker distance and order. Groups that have published maps of the whole X chromosome, either exclusively or as part of a whole-genome effort are given in Table 5.3. The X chromosome Workshop consensus maps by Schlessinger *et al.* (1993), Willard *et al.* (1994) and Nelson *et al.* (1995) are not included because they are compilations of many of those maps given in Table 5.3, including the map presented in this thesis, which was incorporated into the report of Willard *et al.* (1994). These maps are based on the CEPH reference panel (or its precursor) although the number of families genotyped for any given marker does vary.

#### 5.5.6.1 The order of loci

There are two significant discrepancies between the order of loci common to the map presented in this thesis (Table 5.1) and the maps cited in Table 5.3. Both discrepancies are with the Fain *et al.* (1995) map. Firstly, Fain *et al.* (1995) have the *DMD* 3' locus positioned proximal and not distal to *DXS997*. The order presented in Table 5.1 is however consistent with the consensus physical map of Nelson *et al.* (1995). Secondly, Fain *et al.* (1995) have *DXS1000* in Xp (and not Xq) between *DXS255* and *DXS991*. This position is in accordance with that given by Nelson *et al.* (1995) and therefore the position of *DXS1000* as given in Table 5.1 can be considered as incorrect, but the regional localisation as shown in the comprehensive map (Fig. 5.1) does overlap that of *DXS255* and *DXS991*. Weissenbach *et al.* (1992), Matise *et al.* (1994) and Gyapay *et al.* (1994) also included *DXS1000* in their maps and similarly the position of this locus in these maps ranged from that given by Fain *et al.* (1995) to that presented in Table 5.1. The other differences that have occurred between the order of loci given in Table 5.1 and those maps cited in Table 5.3 are a consequence of the relatively low confidence with which the markers concerned have been positioned, that is with odds of less than 1000:1.

**Table 5.3 Comparison of methods and results between published genetic maps of the X chromosome (in order of publication)**

Reference	Minimum map length (cM)	Terminal loci (pter-qter)	No. of loci used	No. of loci in common with Donnelly <i>et al.</i> (a)	No. of CEPH families genotyped	Predominant marker type	Map function	Computer program
Drayna <i>et al.</i> , 1984	215	<i>XG-DXS15</i>	9	5	23	RFLP	unknown	ILINK
Drayna and White, 1985	185	<i>XG-DXS52</i>	21	8	38	RFLP	unknown	(direct counting)
Donis-Keller <i>et al.</i> , 1987	193	<i>DXS278-DXS52</i>	16	6	23	RFLP	Kosambi	CRI-MAP & MAPMAKER
NIH/CEPH CMG, 1992	208	<i>DXS207-DXS52</i>	71	25	V	RFLP	Kosambi	CRI-MAP
Weissenbach <i>et al.</i> , 1992	170	<i>205TF2-DXS998</i>	25	23	8	SSR	Kosambi	LINKAGE
Donnelly <i>et al.</i> , 1994 (this thesis)	236	<i>XG-DXS52</i>	78	-	V	SSR	Kosambi	CRI-MAP
Matise <i>et al.</i> , 1994	177	<i>205TF2-DXS52</i>	57	42	V	SSR	Kosambi	MultiMap
Buetow <i>et al.</i> , 1994	210	<i>DXS143-DXS15</i>	32	25	V	SSR	Kosambi	CRI-MAP
Gyapay <i>et al.</i> , 1994	168	<i>DXS1233-DXS1193</i>	80	22	8	SSR	Kosambi	LINKAGE
Aldred <i>et al.</i> , 1994	214	<i>DXS143-DXS15</i>	38	26	V	SSR	unknown	CRI-MAP
Murray <i>et al.</i> , 1994	224	<i>DXYS218-DXS15</i>	34	18	15	SSR	unknown	CRI-MAP
Fain <i>et al.</i> , 1995	190	<i>DXYS14-DXS1108</i>	243	67	V	SSR	Kosambi	CRI-MAP

(a) the loci in common may not be represented by the same type of marker, that is, RFLPs used in earlier maps may have been made obsolete by SSRs found at the same locus.

(b) V= variable. In these instances the data used for most or all of the markers used in the map construction were obtained from the CEPH database and the number of families genotyped can vary between markers, from 8 to about 60.

RFLP = restriction fragment length polymorphism. SSR = simple sequence repeat (microsatellite).

The maps produced by NIH/CEPH Collaborative Mapping Group (1992), Matisse *et al.* (1994) and Aldred *et al.* (1994) are presented in a framework and comprehensive format as is done in Figure 5.1. The regional localisations of loci given in Figure 5.1 are within, span or overlap the regional localisations given in these other three maps. In other words, in no case does a marker given in Table 5.1 reside wholly outside the regional localisations of the same markers in the other three maps and *vice versa*.

Where possible, physical mapping data were used to order adjacent loci with high odds of inversion (that is 1:1), as indicated in Table 5.1. The only discrepancy between the genetic map and known physical mapping data are the relative positions of *DXS237* and *DXS278*. The consensus map of Schlessinger *et al.* (1993) and Willard *et al.* (1994) have *DXS278* positioned proximal, not distal, to *DXS237*. (*DXS278* has been omitted from the more recent consensus map of Nelson *et al.*, 1995). The odds against inversion of these loci are not 1000:1 but are still relatively high (100:1), nevertheless it is possible that the genetic map, being statistically based, is inaccurate. An alternative explanation exists for the discrepancy: *DXS278* is part of a repeated DNA family, now designated as *DXF30* (Knowlton *et al.*, 1989; Ballabio *et al.*, 1990; Li *et al.*, 1990; Meitinger *et al.*, 1990; Johnson *et al.*, 1991). Copies of this family do flank *DXS237*: *DXF30S1* and *S4* (CRI-S232a and d respectively) are distal and *DXF30S2*, *S3*, *S5* and *S6* (CRI-S232b, c, e, f) are proximal (Petit *et al.*, 1990; Wapenaar *et al.*, 1992). Thus the genotyping data used to position *DXS278* in this thesis may in fact belong to one of the distal *DXF30* loci.

#### 5.5.6.2 The distances between loci

For genetic risk analyses it is necessary to be certain that the loci chosen to provide the risk assessment do in fact flank the disease locus. In this context, an accurate figure for the genetic distance between loci is not as important as the locus order. The order of loci can be confirmed by physical mapping but genetic distances cannot. Inaccurate genetic distances between correctly ordered flanking loci may result in an over- or under-estimate of the risk involved in inheriting an intervening disease gene. However, relatively large errors in map genetic distance have little effect on estimates of double crossover rates (the diagnostic error rate using flanking markers) compared with the effect on estimates of single crossover rates (the diagnostic error rate determined when using a single non-intragenic linked marker).

As for the dystrophin region given above, one can compare the genetic distances between markers elsewhere on the map given in Figure 5.1 with those of the other published maps given in Table 5.3. Differences in distances do exist but only those discrepancies chosen arbitrarily as greater than 5 cM will be highlighted and are given in Table 5.4. It is not known what map function was used by Drayna and White (1985) to convert recombination fractions to centimorgans, but the Kosambi function is assumed. The maps of Drayna *et al.* (1984) and Weissenbach *et al.* (1992) only give recombination fractions between loci. These values were converted to centimorgans using the Kosambi map function.

Some of the distance disparities given in Table 5.4 can be explained by the failure to position one or both loci of a pair with high odds within their respective maps, that is, relatively large confidence intervals apply. Reasons as to why inter-loci distances and loci orders can vary between maps will be discussed below.

The 30 cM distance between the *XG* locus (whose alleles are distinguished using a serological assay) and the next framework marker, *DXS143*, is too great compared to previous reports. For example Johnson *et al.* (1991) have it positioned, by integration of its physical location with a genetic map, about 9.5 cM distal to *DXS143*, whereas Yates *et al.* (1987) have the distance as 18 cM (although their map was constructed using the CEPH data of Drayna and White (1985) as well as that from disease families).

#### 5.5.6.3 Factors contributing to differences between maps

Despite the differences between the map given in this chapter and others, at the time this map was constructed it was the most comprehensive map of the X chromosome available, and was used as the basis for the linkage analyses presented in Chapters 6 and 8.

Differences in lengths between maps given in Table 5.3 could be attributed to differences in the: 1) terminal markers used; 2) choice of internal markers used; 3) methods of map construction; 4) numbers of informative meioses; 5) mapping functions used; 6) error rates and 7) numbers of internal markers used. Each of these points will be dealt with in more detail.

**Table 5.4**

**Discrepancies of  $\geq 5^*$  cM between loci given in Table 5.1 and those in other maps**

Loci	Distance (to nearest 0.5 cM)	
	Table 5.1	Other (reference)
DXS7 - DXYS1	32.5	>38 (Drayna <i>et al.</i> , 1984) <sup>a</sup>
DXS143 <sup>^</sup> - DXS43*	20	15 (Donis-Keller <i>et al.</i> , 1987) <sup>b</sup>
DXS43* - DXS41	13	7
DXYS1 - HPRT	57	43
HPRT - DXS15*	43.5	53
DXS15* - F8C*	0	7
XG - DXS143	30	14 (Drayna and White 1985) <sup>c</sup>
DXYS1 <sup>^</sup> - HPRT	57	31
HPRT - DXS15*	43.5	34
DXS207 - DXS7	51.5	64 (NIH/CEPH CMG, 1992)
DXS990* - DXS1105* <sup>^</sup>	17.5	8 (Weissenbach <i>et al.</i> , 1992)
DXS990* - HPRT	52	44 (Matisse <i>et al.</i> , 1994)
DXS456 - DXS425	20	29 (Buetow <i>et al.</i> , 1994)
DXS996-DXS207	23	17.5 (Murray <i>et al.</i> , 1994)
DXS3-DXS456	21	10.5
DXS456-DXS425	20	41
DXS443-DXS41	10	5 (Fain <i>et al.</i> , 1995)

\* The figure of 5 cM is arbitrary. It was chosen so show that genetic distances can vary considerably between maps.

a: Drayna *et al.* (1984) gave recombination fractions with 90% confidence limits.

Distances (Kosambi cM) separating markers in common between the two maps fell within these confidence limits except for the pair given.

b: Donis-Keller *et al.* (1987) included the loci of Drayna *et al.* (1984).

c: Drayna and White (1985) did not specify the mapping function used. Kosambi was assumed.

Loci indicated with \* were not uniquely positioned, as shown in Fig. 5.1.

Loci indicated with <sup>^</sup> were not uniquely positioned in the other map under consideration.

1) All of the maps span about the same physical length of the X chromosome, from or near the *XG* locus to *DXS52/DXS15*, hence the differences in lengths cannot be attributed to variation in the choice of telomeric loci.

2) The markers used are of course limited by the dataset available. Those maps given in Table 5.3 published prior to 1992 relied heavily on markers of low heterozygosity, predominantly RFLPs. The ordering and distances between such markers is thus determined by relatively few recombination events, and the statistical support can be relatively low. In addition, double recombination events could go undetected because of the scarcity of informative meioses. Maps using highly informative PCR-based markers are less susceptible to these problems. However, these high resolution maps can vary in their final length, even if they differ in only a few of the (internal) markers used in their construction, because of the varying amounts of information associated with each marker and the inter-relationship with the data provided by surrounding markers. As map construction in the future draws on the same database one would expect differences due to marker heterogeneity to disappear because the numbers of markers to choose from will be great and only those with the greatest heterozygosity will be used. The physical locations of the markers used in the map may also affect map length in that, given most recombination events occur within the sub-telomeric regions, the greater the marker density in these regions the less likely double recombination events will be overlooked. So maps that use markers that happen to be located in these regions will be longer than those that do not. Conversely recombination in pericentric regions is relatively rare hence increasing marker density in these regions will have no effect on detecting non-existent crossovers.

3) The choice of algorithm used to construct a map may affect the outcome given the same dataset.

4) The numbers of informative meioses depends on the heterozygosities of the markers used and the numbers of families genotyped. The numbers of CEPH families used, when known, in the construction of each map is given in Table 5.3. The fact that many of the markers generated by Weissenbach *et al.* (1992) could not be positioned into the framework map (that is with odds of 1000:1 against any other position), despite the fact that they are highly polymorphic (average heterozygosity of 68%), is solely because they were typed in only 8 of the 40 to 60 CEPH families. Many of the other markers, including

all those genotyped in the Department of Cytogenetics and Molecular Genetics were genotyped on 40 CEPH families.

5) Mapping functions are models of the phenomenon of interference and they are used to convert recombination frequencies ( $\theta$ ) into genetic distances (cM). They have been discussed in Chapter 2 (2.3). Using an inappropriate function may result in incorrect loci orders and distances. For example, a mapping function that assumes no interference may permit the juxtaposition of multiple crossovers that in reality are separated by a substantial physical distance. Most of the maps cited in Table 5.3 are known to have been constructed using the Kosambi function, so any differences between these maps cannot be attributed to the choice of map function. Also, as stated in Chapter 2 (2.3), as marker density increases the choice of mapping function used becomes irrelevant because the recombination frequency is equivalent to the genetic distance. Of concern is the claim that the Kosambi function and some others are not “multi-locus feasible”, that is, they cannot be used to generate the probabilities of the gamete (haplotype) frequencies for more than 3 loci simultaneously (Weeks *et al.*, 1994). Calculation of these gamete probabilities is in turn required for the calculation of the expected recombination frequencies (Ott, 1991). Conversely, Lathrop *et al.* (1985) claim it is reasonable to assume, when multipoint mapping, that no interference is operating.

Under independence (no interference), the probability of a double recombination event occurring is equal to the product of the recombination fractions of the two intervals under consideration that is,  $\theta_{ab}\theta_{bc}$ . The coefficient of coincidence ( $c$ ) is a measure of the deviation from independence due to interference ( $I$ ). Mathematically;

$$c = \frac{(\theta_{ab} + \theta_{bc} - \theta_{ac})}{2\theta_{ab}\theta_{bc}}$$

$$\text{and } I = 1 - c$$

Estimates of  $c$  have a large standard error so distinguishing  $c < 1$  from  $c = 1$  (no interference) can be difficult. As a consequence, it has been determined that to test for interference would require 800 offspring from phase-known triple backcross parents (Ott 1991, page 123). Evidence for positive interference along human chromosomes has been presented by Kwaitkowski *et al.* (1993) and Weber JL *et al.* (1993).

6) As maps of higher resolution are sought the less accurate they may become in terms of local marker order and especially inter-marker distances. This is because as the marker density increases the genotyping error rate approaches the recombination frequency between markers, leading to support for incorrect loci order and a “blow-out” in map length (Buetow, 1991; Lasher *et al.*, 1991; Lincoln and Lander, 1992). This map inflation *per interval* has been estimated as twice the error rate for each marker/locus (Buetow, 1991; Lasher *et al.*, 1991). Genotyping errors have been estimated to occur with a frequency of about 0.04 to 1.0%, (and specifically for the CEPH data: 1%) (Buetow, 1991; Shields *et al.*, 1991; Buetow *et al.*, 1993; McInnis *et al.*, 1993). So, for example, assuming an error rate of 0.1% and a map of 78 markers (= 78-1 intervals), as used in this thesis, the expected inflation is  $(2 \times 0.1 \times 77)$  15 cM.

Apparent double recombination events occurring within relatively short distances are generally suspected to be the result of genotyping errors, but this assumption may have to be reconsidered given that such an event has been observed occurring within 700 kb of the Xq pseudoautosomal region (Rappold *et al.*, 1994).

Errors can arise at a number of steps: mislabelling of DNA samples; pipetting errors; autoradiograph reading errors; transcription errors. Another irony is that markers of the highest heterozygosity are preferred in map construction, yet this great degree of polymorphism is a consequence of the relatively high instability of these markers. Alleles mutating to lengths of other alleles segregating within a pedigree will produce a pseudo-recombination event.

Such human and biological errors are not a problem if they are revealed by the apparent non-Mendelian inheritance of alleles. However, if they are not detected they can result in apparent increases in recombination frequencies between the marker involved and flanking markers. This effect may affect the model of interference (the mapping function) one wishes to use for subsequent mapping programs. Errors resulting in apparent decreases in map length, by masking a true recombination, are also possible but would be very much rarer events because (true) recombination events are themselves relatively rare events.

Methods can be used to combat these errors. One involves flagging either double recombination events that occur within relatively small regions (or double cross-overs

involving single markers) or parents that appear to transmit recombinant chromosomes to most of their children, using for example the CHROMPIC option of CRI-MAP. Such highlighted occurrences can be investigated in more detail by re-reading autoradiographs or re-genotyping the suspect individual.

An alternative procedure is to successively withdraw each marker from the map and observe how much, if at all, the map length decreases (Buetow *et al.*, 1993; McInnis *et al.*, 1993). For a true map interval of  $x$  cM the calculated map length is  $[x + 2r]$  cM, where  $r$  is the percentage error rate of a marker in the interval (see above). In this way, by removing each locus and comparing the new map length with the original, those loci that contribute to the map inflation most can be identified. The marker genotypes of the loci that are responsible for the inflation are investigated further or omitted from the final map construction. Of course there exists the possibility that those genotypings removed may reflect true recombination events (if so, the modified data may be later misinterpreted as evidence for positive interference, or a lack of evidence for negative interference). Gene conversion, the replacement of a segment of DNA with that copied from another locus or allele (see for example Collier *et al.*, 1993), is another process that may be misconstrued as a genotyping error because it mimics a double recombination event over a very small region. The conversion of one microsatellite allele to another has been estimated to occur with a frequency of 0.0003/meiosis (Weber JL *et al.*, 1993).

Similarly, another approach is to apply an “error filter” which is an equation or computer algorithm that reduces the apparent recombination values by an amount dependent on the estimated error rate (Shields *et al.*, 1991; NIH/CEPH Collaborative Mapping Group, 1992).

The most common expected locus order error is the permutation of groups of 2 or more loci (Buetow and Chakravati, 1987). This can be investigated using the FLIPS option of CRI-MAP to invert groups of loci and assess the support for the various orders. Multiple pairwise analysis is less sensitive to errors and can also be used to check the local order of a multipoint map (Buetow, 1991).

7) As the numbers of markers used in map construction increases one would expect map lengths to increase as map resolution improves, until all recombination events, within the family groups studied, are identified. Conversely to point (6), as marker density increases genotyping errors, causing false recombinants, are more readily detected, enabling the data to be corrected or discarded.

#### 5.5.7 Meiotic breakpoint map

The 8 families used to construct the map were the ones also used by Weissenbach *et al.* (1992), Gyapay *et al.* (1994) and Dib *et al.* (1996; who also used 12 other CEPH families) for the construction of their maps (although the new X chromosome markers characterised by Gyapay *et al.* (1994) and Dib *et al.* (1996) were reported subsequent to the analysis shown in this thesis and their locations are not presented herein).

Once a broad localisation of a new marker has been made, more efficient fine scale positioning can be achieved by genotyping only those CEPH children that are recombinants within that interval (Fain *et al.*, 1989; Elsner *et al.*, 1995; Cox *et al.*, 1996). For example, a hypothetical marker, *x*, is known (either from preliminary physical or linkage mapping) to lie within a 20 cM interval. Recombination within this interval will, on average, only be exhibited in about 20% of the CEPH grandchildren. Therefore only these 20% need to be genotyped to position this hypothetical marker more accurately, providing their mothers are heterozygous for the marker alleles. Specifically, say marker *x* is known to be between *DXS1125* and *DXS1153* in Xq. From the CEPH pedigrees in Figure 5.2 the maternally inherited X chromosomes of only 20 of the 94 children can provide information to refine the position. A non-recombinant chromosome from each sibship is also typed for marker *x*, as is the mother. Fain *et al.* (1995) have recently put this principle into practice by constructing a comprehensive map of the human X chromosome.

The frequency of recombination can also be investigated. Drayna and White (1985) determined the frequency and distribution of crossovers in 105 X chromosomes. They observed an average of 1.5 recombination events *per* chromosome, with 14% of chromosomes showing no recombination, 35% (the largest category) experiencing 2 crossovers and 1% experiencing 4 crossovers. Drayna and White (1985) also noted a significant deviation between the observed and expected distribution of events and

concluded that recombination points may be non-random. The average number of human X chromosome chiasmata observed in 27 oocytes during diakinesis I, was 2.76 *per* bivalent (Jagiello *et al.*, 1976). A comparison, between families, of crossover frequencies near the centromere of the X chromosome has also been performed (Mahtani and Willard, 1988).

An average of 1.8 crossovers *per* X chromosome was observed in the 8 CEPH families (Table 5.2; Fig. 5.3). This number was also observed by Fain *et al.* (1995). Given that, by definition, one crossover is expected within a distance of 1 Morgan (100 cM), then the X chromosome by this measure is equal to 180 cM. The autosomal length of the female genome has been calculated as 3,900 cM (Renwick, 1971) and the X is considered to be 5.1% of the total genome (Heslop-Harrison *et al.*, 1989), therefore this figure of 3,900 cM represents 94.9% of the total female genome. It follows that the X should be equal to 210 cM in length, which is similar to the 180 cM predicted from the average number of crossovers seen in the 8 largest CEPH families and to the genetic length of 236 cM (Table 5.1).

Disregarding the pseudoautosomal regions, which were not analysed, the 168 recombination events revealed in Figure 5.2 may not be the true number because of: 1) the distances between the markers of the map; 2) many of these markers have not been typed or are uninformative; 3) not all the children have been genotyped to the same extent (see for example 1332/09); and 4) typing errors and mutations may generate spurious recombination events. Typing errors may have occurred for example in the scoring of the alleles of: *DXS278* in person 884/10; *DXS996* in 1416/06; and *XG* in 1332/06, 1332/12 and 1347/09 (see Fig. 5.2). All these loci occur near the Xp pseudoautosomal boundary and as telomeres experience relatively frequent recombination events these observations do not necessarily reflect genotyping errors.

## 5.6 Conclusion

One of the advantages of the high resolution genetic map presented above is that it integrated most of the microsatellite markers of Weissenbach *et al.* (1992) and some pre-existing loci (which had been converted to a PCR format). As a consequence, the markers of the Weissenbach map can be related to those markers used in previous disease gene mapping efforts (which relied on RFLPs and VNTRs). When originally constructed, this map (Donnelly *et al.*, 1994b; appendix A.1) was put to use in the localisation of a number of X-linked disease genes, to be described in Chapters 6 and 8. Subsequently, a more detailed map of the X chromosome was published by Fain *et al.* (1995).

**Chapter 6**

**Linkage Mapping of Genes**

**for**

**X-linked Mental Retardation**

## Table of Contents

	Page
<b>6.1 Summary</b>	149
<b>6.2 Introduction</b>	150
<b>6.3 Materials and Methods</b>	153
6.3.1 DNA extraction	153
6.3.2 PCR	153
6.3.3 Linkage analyses	153
6.3.4 Haplotype analysis	154
<b>6.4 Mapping of <i>MRX19</i></b>	155
6.4.1 Materials and Methods	155
6.4.1.1 The family	155
6.4.1.2 Linkage analysis	155
6.4.2 Results	158
<b>6.5 Mapping of <i>MRX30</i></b>	163
6.5.1 Materials and Methods	163
6.5.1.1 The family	163
6.5.2 Results	165
<b>6.6 Mapping of <i>MRX31</i></b>	169
6.6.1 Materials and Methods	169
6.6.1.1 The family	169
6.6.2 Results	171
<b>6.7 Mapping of the <i>XLMR</i> gene affecting the W family</b>	174
6.7.1 Materials and Methods	174
6.7.1.1 The family	174
6.7.2 Results	176
<b>6.8 Mapping of the <i>XLMR</i> gene affecting the O'H family</b>	179
6.8.1 Materials and Methods	179
6.8.1.1 The family	179
6.8.1.2 Linkage analysis	179

6.8.2 Results	181
<b>6.9 Discussion</b>	186
6.9.1 <i>MRX19</i>	186
6.9.2 <i>MRX30</i> and <i>MRX31</i>	190
6.9.3 Family W	192
6.9.4 Family O'H	192
6.9.5 Overview	195
<b>6.10 Conclusion</b>	198

## 6.1 Summary

The genetic map of the X chromosome presented in Chapter 5 was applied to the regional localisation of the X-linked mental retardation disease genes segregating in five families. Three of these families (*MRX19*, *MRX30* and *MRX31*) had males with non-specific mental retardation and the other two families (W and O'H) had males with additional symptoms. The *MRX19* gene has been localised to Xp22, within a 19 cM region between *DXS1043* and *DXS1052*. The *MRX30* gene has been localised to a 28 cM region between *DXS990* (Xq21.33) and *DXS424* (Xq23). The *MRX31* gene has been localised to a 12 cM region between *DXS1126* (Xp11.23) and *DXS1124* (Xq13). These results have enabled carrier risk analyses to be offered to the females of these three families. Linkage analysis of family W excluded most of the long arm of the X chromosome, but revealed significant lod scores to two large regions mainly confined to the short arm, in which the disease gene may lie. The first region is from Xqter to *DXS1068* (Xp11.4) and the second region lies between *MAOA* (Xp11.4) and *DXS1002* (Xq13.3-q21.1). The linkage analysis of family O'H has shown significant lod scores to two relatively small regions: the disease gene lies either between *DXS1275* (Xq12) and *DXS7113* (Xq12-q13.1) or in Xq23 between *DXS1220* and *DXS8067*.

The analyses of the three families affected with non-specific mental retardation have been published and are given in the appendix; *MRX19* (Donnelly *et al.*, 1994a; appendix A.3), *MRX30* and *MRX31* (Donnelly *et al.*, 1996b; appendix A.4).

## 6.2 Introduction

The identification of X-linked disorders has usually been relatively easy because of their characteristic pattern of inheritance. It is the hemizyosity of males that reveals rare recessive disorders that if autosomal would be masked by expression from the wildtype allele. As at 1992, approximately 200 X-linked gene loci (not all of which are responsible for diseases) and 2,300 autosomal gene loci had been described (McKusick, 1992). It has been estimated that there are about 70,000 genes in the human genome (Fields *et al.*, 1994) and assuming an even distribution there should be about 3,500 genes on the X chromosome. Therefore only a relatively small number of X-linked gene loci have a known function or are associated with a recognisable phenotype.

Genes for X-linked disorders can be rapidly localised within families of sufficient size given the rapid improvement of the genetic linkage map (Chapter 5). Mapping has in the past been restricted by the scarcity of markers and their low heterozygosity. Generally, given that the average spacing of markers along the comprehensive map of the X chromosome presented in Chapter 5 is 3 cM, it is the number of potentially informative meioses and the chance occurrence of recombination events that determine the extent of the regional localisation of the disease locus under investigation. Generally, the rarer the disease, the fewer the families and the larger the interval to which it is likely to be mapped, simply because of the lack of informative meioses. By chance, however, a disease gene could be localised to a relatively small region (for example 1 cM) from the observation of very few informative offspring. A lod score of 2 or more (indicating statistically significant linkage between an X-linked gene and a marker) requires the equivalent of at least 7 fully informative (phase known) meioses, that is 7 children, of known phenotype, who can be unambiguously scored as being non-recombinant with respect to the marker locus concerned (Ott, 1991 equation 3.2).

Genetic heterogeneity of clinically similar disorders and the presence of phenocopies can complicate genetic mapping. A phenocopy is a person in a family whose phenotype is the same as that of other affected members of that family, but this phenotype is not caused by the mutation segregating within that family (instead, for example, environmental factors may be responsible). Genetic heterogeneity refers to mutations in different genes that cause the same phenotype in affected individuals from different families. For example, past

experience has shown that non-specific X-linked mental retardation is a genetically heterogeneous disease because the regional localisations of the disease loci in some families do not overlap the regional localisations of the disease loci in others (Mulley *et al.*, 1992). This precludes the combined linkage analysis of families, which, when one is dealing with a genetically homogeneous disease, can be used to boost statistical evidence for linkage to a particular marker(s) and reduce the regional localisation of the disease gene, even to the point where positional cloning may be possible.

As a consequence of the hemizyosity of males, the selection pressure acting to eliminate severe X-linked recessive diseases is greater than for autosomal recessive diseases. Different families with the same X-linked disorder are therefore likely to have different mutations (in the same gene). This may lead to variable phenotypes (allelic heterogeneity), and the expected variety of mutations will complicate procedures for routine mutation screening. There are other interesting aspects to the study of the X chromosome: the phenomenon of X inactivation; the presence of the pseudoautosomal regions; and its evolutionary conservation which has facilitated gene mapping because genes X-linked in one species are most likely to be X-linked in another (Ohno's Law; see for example De Kok *et al.*, 1995).

X-linked mental retardation (XLMR) is the most common cause of retardation in males. The frequency of XLMR has been estimated at 1/600 males, with the fragile X A (*FRAXA*) syndrome accounting for about 1/5700 (Herbst and Miller, 1980; Sutherland and Hecht, 1985; reviewed by Glass, 1991, Kerr *et al.*, 1991 and Morton *et al.*, 1997). Mental retardation is the primary characteristic of 98 of the 147 known X-linked diseases that feature this affliction (this figure includes those families with non-specific mental retardation - see below) (for reviews see Schwartz, 1993 and Lubs *et al.*, 1996).

Non-syndromal (or non-specific) forms of X-linked mental retardation (designated "MRX") have no other known abnormalities in common between the affected individuals, whereas syndromal forms of mental retardation ("MRXS") do involve characteristic secondary neurological, biochemical and/or morphological defects. If a statistically significant lod score (2 or greater) is obtained with a particular *MRX* locus, a progressive number is assigned to that disease gene by the Human Gene Mapping Nomenclature Committee, for example *MRX1*, *MRX2* etc, where each *MRX* number represents the

analysis of one family (Mulley *et al.*, 1992), but does not necessarily imply the involvement of a different gene. As at 1995, there were about 40 families afflicted with *MRX* that were large enough to enable sub-localisation of the disease gene by linkage analysis to one or more X-linked markers (Gedeon *et al.*, 1996b; appendix A.5). The known regional localisations of the disease gene(s) segregating in a large proportion of these *MRX* families spans the centromere with the remainder spread along the rest of the X chromosome (Gedeon *et al.*, 1996b). The genetic heterogeneity of this homogeneous clinical entity precludes the combined linkage analysis of families whose *MRX* genes have overlapping localisations because it cannot be assumed the same gene is involved.

The sections to follow describe the genetic mapping of five XLMR mental retardation loci, including *MRX19*, *MRX30* and *MRX31*.

## 6.3 Materials and Methods

Only brief accounts of the methods used are given below, further details are given in the Chapter 3 sections indicated.

### 6.3.1 DNA extraction

DNA was usually extracted from lymphocytes of the family members by Jean Spence (3.2.7.4). Alternatively, extracted DNA was sent from laboratories affiliated with the clinicians who had clinically characterised the families concerned and were responsible for genetic counselling.

### 6.3.2 PCR

Amplification of microsatellites and the subsequent genotypic analysis of the PCR products were performed as described in 3.2.15.1, most often using the “Iomix” conditions. Multiplexing was performed when permitted by compatible MgCl<sub>2</sub> concentrations, low non-specific background and product sizes.

Primer sequences were obtained either from the references given in Table 3.1 or, in the case of those characterised by Weissenbach *et al.* (1992), from the Genome Data Base (GDB).

### 6.3.3 Linkage analyses

To rapidly localise a disease gene, two-point linkage analysis between the disease locus and each marker was performed using MLINK as described in 3.2.21.1.

To refine the localisation of a disease gene additional two-point analyses were performed using informative markers from the region of interest and where necessary, to increase the number of informative meioses, multipoint linkage analysis was done using LINKMAP as described in 3.2.21.2.

Risk analysis was performed using MLINK of the LINKAGE program (version 5.1). The distances between the flanking informative markers used in the analyses were taken from Table 5.1.

#### 6.3.4 Haplotype analysis

Haplotypes were constructed to reveal in whom recombination between marker loci and the disease gene had occurred. The haplotype of a carrier female could be readily determined if her father had also been genotyped. As a consequence recombination events, as revealed in her male children could easily be determined if she was heterozygous for the loci concerned. If the father had not been genotyped, her haplotype was determined on the basis of minimising the number of recombination events, using the data from other members of the pedigree.

More specific materials and methods are presented in the sub-sections dealing with each of the five XLMR families as appropriate.

## 6.4 Mapping of *MRX19*

### 6.4.1 Materials and Methods

#### 6.4.1.1 The family

The *MRX19* pedigree is shown in Figure 6.1. Clinical assessment and previous mapping studies were performed by Prof. David Danks and Dr K. Choo and colleagues at the Murdoch Institute in Melbourne. The six affected males are moderately retarded (IQ 35-50) and have no physical, karyotypic nor other neurological abnormalities. In addition, there are three mildly retarded (IQ 50-70) females, including I-2 in Figure 6.1 who had not been assessed, but was described as “dull” by other family members.

#### 6.4.1.2 Linkage analysis

The DNA samples, from the people indicated in Figure 6.1, were initially screened using 17 polymorphic microsatellite markers spanning the X chromosome, followed by another 7 markers from the region of interest.

Two-point linkage analyses between the disease locus and each marker were first performed assuming X-linked recessive inheritance (that is, no penetrance in female carriers) and then, because there are mildly affected women in the family, repeated using a penetrance of 0.8 in females, as used by Arvelier *et al.* (1988). All females were scored as unaffected for the first linkage run but II-2 and III-2 were scored as affected for the second run.

Multipoint (six-point) analysis using the LINKMAP program was used to position *MRX19* into the background map. No penetrance in females was assumed (as for the first two-point run described above). The background map was constructed from the CEPH database by Dr Helen Kozman as previously described (3.2.21.3), and is as follows (with distances in cM): *DXS996*-1.8-*KAL*-7.9-*DXS1043*-11.7-(*DXS987*, *DXS207*)-3.5-*DXS999*-0.1-*DXS443*-2.8-*DXS365*-0.8-*DXS1052*-0.8-*DXS274*-9.6-*DXS989*. The *KAL*, *DXS1043*, *DXS365*, *DXS1052* and *DXS274* loci are not present in the map given in Chapter 5. The genotyping data required for the inclusion of *DXS1043*, *DXS365*, *DXS1052* and *DXS274* came from version 7 of the CEPH database, whereas the positions of the *KAL* and *DXS1052* loci were estimated, by the candidate, from the data of Zhang *et al.* (1993) and Rowe *et al.* (1994) respectively.

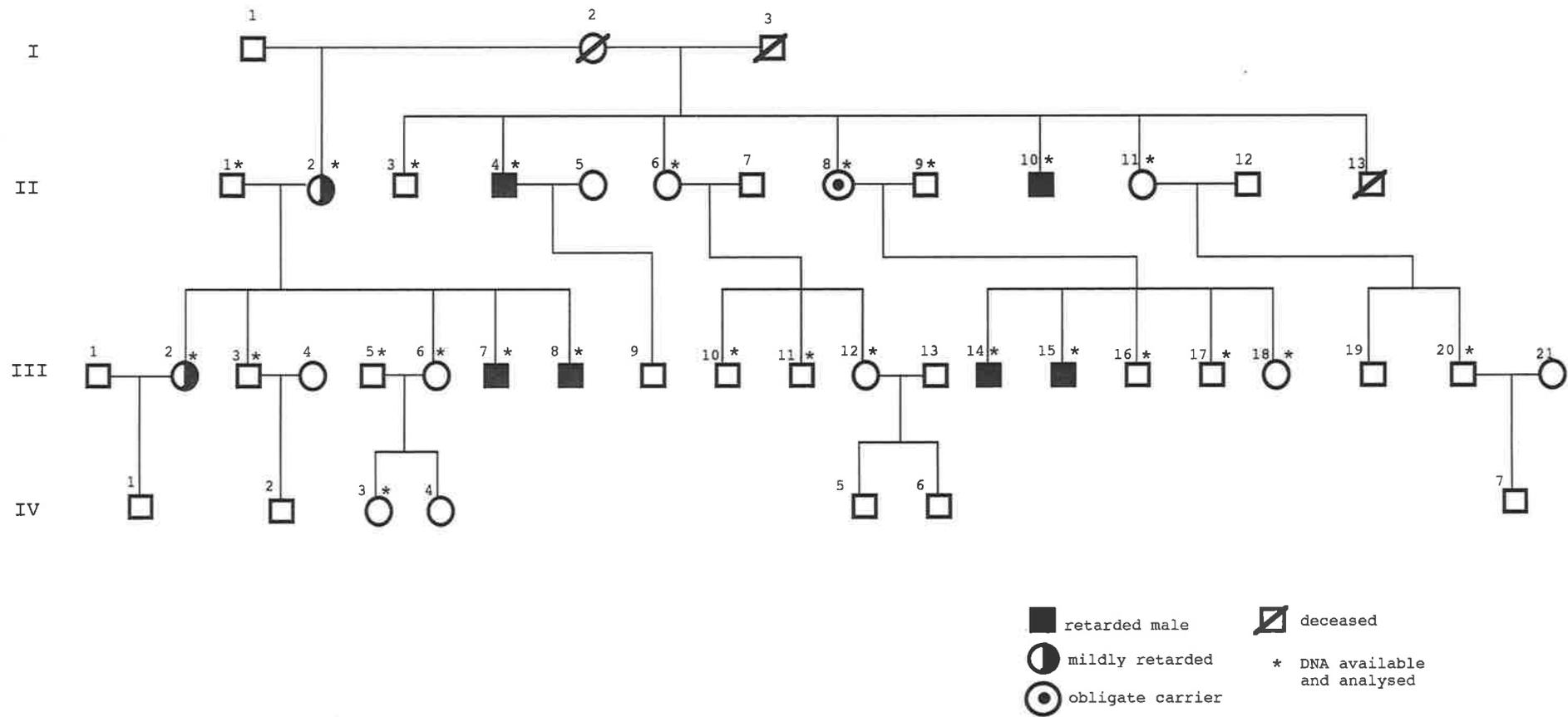


Figure 6.1 The *MRX19* pedigree.

The lod-1 method was applied to the multipoint lod score to determine the (approximate) 90% support interval for the location of the gene (Conneally *et al.*, 1985). The lod-1 method is routinely used for the estimation of support intervals from two-point lod scores, but the significance level when applied to multipoint lod scores is unclear (Keats *et al.*, 1989; Terwilliger and Ott, 1994 p121). The 90% support interval for *MRX19* can therefore only be regarded as a crude approximation designed to indicate the most likely interval for the selection of candidate loci.

An earlier version of this analysis appears in Donnelly *et al.* (1994a; appendix A.3). The analysis presented below differs in that additional microsatellite markers are used that were not available at the time of preparation of the paper and, as a consequence, minor inter-marker genetic distances differences have arisen between the maps derived from the multipoint analyses.

Risk analysis for potential female carriers in this family was performed using a heterozygote penetrance of 0.8 and scoring II-2 and III-2 as affected. This illustrates the clinical value of the disease gene localisation.

## 6.4.2 Results

Of the 24 marker genotypings for individual II-2, 14 were incompatible with I-3 being her biological father. This required the inclusion of the inferred father (I-1) shown in Figure 6.1.

Two-point lod scores between the disease locus and the 24 markers used to screen the *MRX* family are given in Table 6.1. No recombination occurred between the *MRX19* locus and the loci *DXS987*, *DXS207* and *DXS999* to result in the significant lod scores of 3.58, 3.58 and 3.28 respectively. Higher lod scores were obtained when the symptomatic females, II-2 and III-2, were scored as affected and penetrance was considered to be 0.8 (Table 6.1).

Pedigree inspection (Fig. 6.2) of the affected and carrier members of the family allowed identification of the flanking recombinants. III-7 and II-8 are recombinant for *DXS1043* and II-2 exhibits recombination with the proximal locus represented by marker *DXS1052*. Therefore, the *MRX19* gene is located within the 19 cM region flanked by the loci *DXS1043* (Xp22.31) and *DXS1052* (Xp22.13).

Only genotyping data for the most informative markers of the background map were used in the multipoint analysis, because of computer memory limitations. The markers used and the distances (cM) between them were: *DXS1043*-11.7-*DXS207*-3.5-*DXS999*-2.9-*DXS365*-0.8-*DXS1052*. The multipoint lod score peaks at *DXS207* and *DXS999* with a value of 3.58 (Fig. 6.3). The lod-1 (90%) support interval gives a 15 cM region (Fig. 6.3) in which there is a high probability of finding the gene. This distance is only slightly less than the 19 cM between the 2 flanking loci, *DXS1043* and *DXS1052*, determined by examination of the closest crossover points.

Upon mapping the disease gene to Xp22, the female (III-6) requested a risk analysis be performed to determine her likely carrier status. Haplotype analysis, constructed on the basis of crossover minimisation, (Fig. 6.2) indicated that she had not received the same chromosomal segment between (and including) the loci *DXS207* and *DXS1052* as her mildly affected sister (III-2) and her affected brothers (III-7 and III-8). However there is the possibility of an undetected double-crossover occurring between any two markers within the interval *DXS1043*-*DXS1052*. Such an event could have transferred the disease

gene to the low risk haplotype. The largest distance separating any two loci within this interval is 12 cM between *DXS1043* and *DXS987/207*. Hence the maximum chance of a double crossover occurring is  $(0.12/2)^2$  or 0.36% (assuming no interference and 1 cM=1% recombination). As a consequence there is a 99.64% chance that III-6 is a non-carrier. A slightly higher risk (0.7%) of being a carrier was obtained using the LINKAGE program, (assuming 0.8% penetrance and scoring II-2 and III-2 as affected). This was because only the loci *DXS1043* and *DXS1052* were used in the analysis and as 19 cM separates these two loci there is a greater chance of a double crossover going undetected.

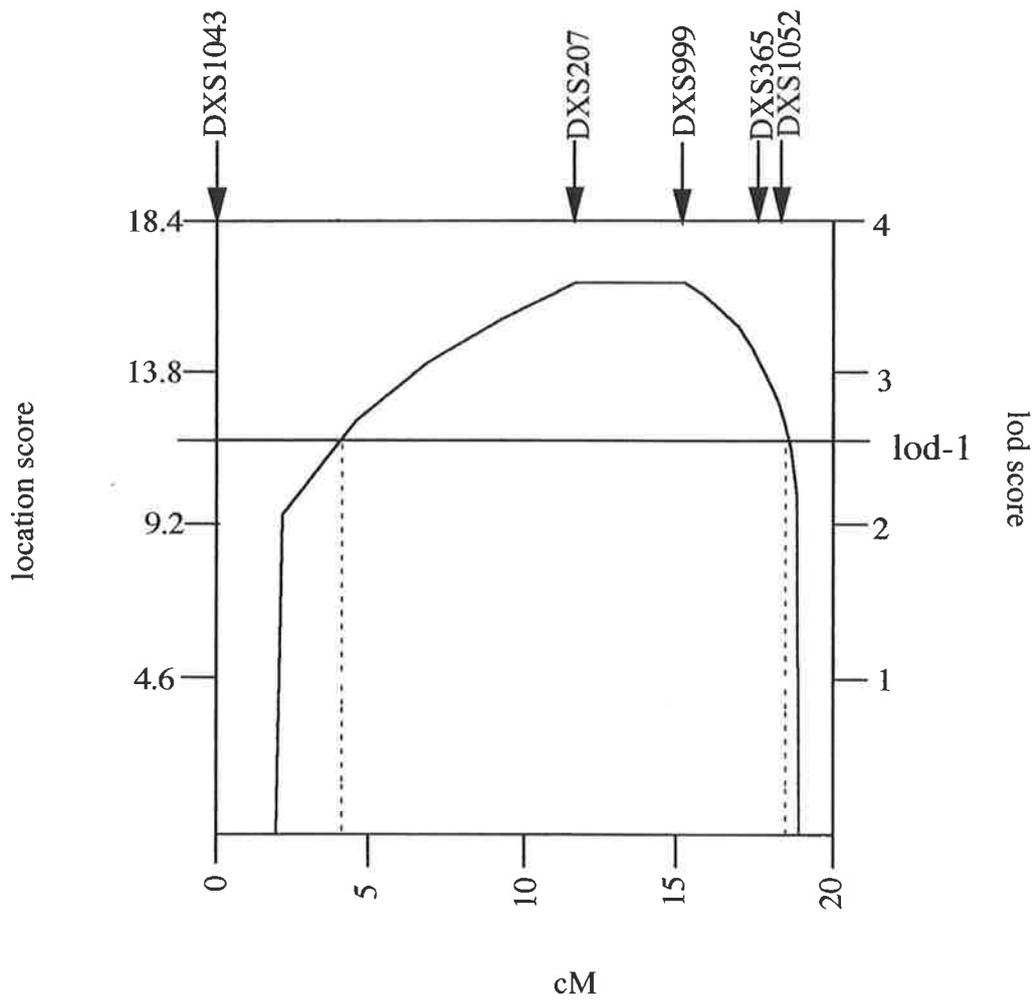
**Table 6.1 Two-point analysis between *MRX19* and marker Loci (Xpter to Xqter)**

Locus	$\theta$							Z max	$\theta$ max
	0.00	0.01	0.05	0.1	0.2	0.3	0.4		
DXS996	$-\infty$	-4.45	-1.79	-0.76	0.03	0.26	0.21		
KAL	$-\infty$	-1.81	-0.51	-0.03	0.29	0.32	0.22		
DXS1043	$-\infty$	-0.47	0.74	1.09	1.15	0.88	0.46		
	$-\infty$	0.61	1.75	2.01	1.87	1.38	0.71		
DXS987	3.58	3.52	3.30	3.00	2.36	1.61	0.78	3.58	0.0
	4.67	4.60	4.31	3.92	3.07	2.11	1.03	4.67	0.0
DXS207	3.58	3.52	3.30	3.00	2.36	1.61	0.78	3.58	0.0
	4.67	4.60	4.31	3.92	3.07	2.11	1.03	4.67	0.0
DXS999	3.28	3.22	3.00	2.70	2.05	1.33	0.57	3.28	0.0
	4.37	4.30	4.00	3.62	2.77	1.81	0.77	4.37	0.0
DXS443	1.94	1.91	1.80	1.66	1.34	0.96	0.52	1.94	0.0
	2.66	2.62	2.46	2.25	1.80	1.29	0.69	2.66	0.0
DXS365	1.91	1.89	1.79	1.65	1.34	0.96	0.52	1.91	0.0
	2.88	2.83	2.66	2.43	1.94	1.39	0.75	2.88	0.0
DXS1052	$-\infty$	1.23	1.72	1.75	1.49	1.04	0.49		
	$-\infty$	2.31	2.73	2.66	2.17	1.46	0.66		
DXS274	$-\infty$	1.53	2.01	2.04	1.74	1.23	0.62		
DXS989	$-\infty$	0.34	0.90	1.02	0.92	0.65	0.30		
DYS (3'Dys MS)	$-\infty$	0.04	0.60	0.73	0.66	0.44	0.17		
DYS (5'Dys-II)	$-\infty$	0.47	0.71	1.04	1.06	0.79	0.41		
DXS1068	$-\infty$	-3.84	-1.23	-0.27	0.41	0.53	0.37		
PFC	$-\infty$	-0.47	0.72	1.06	1.10	0.85	0.46		
AR	$-\infty$	-0.47	0.73	1.07	1.09	0.81	0.41		
DXS1125	$-\infty$	0.48	1.04	1.16	1.07	0.81	0.45		
DXS454	$-\infty$	-2.74	-0.74	0.01	0.53	0.59	0.39		
DXS1153	$-\infty$	-2.37	-0.46	0.21	0.62	0.62	0.41		
DXS425	$-\infty$	-3.61	-1.62	-0.85	-0.22	0.02	0.07		
HPRT	$-\infty$	-2.59	-0.67	0.01	0.44	0.47	0.30		
DXS102	$-\infty$	-3.89	-1.86	-1.05	-0.34	-0.05	0.05		
DXS984	$-\infty$	-3.80	-1.78	-0.97	-0.29	0.04	0.08		
DXS297	$-\infty$	-5.89	-2.57	-1.29	-0.29	0.04	0.08		

For markers in the interval *DXS1043-DXS1052*, lod scores were determined assuming no penetrance (ie X-linked recessive inheritance) in heterozygotes (upper line) and 0.8 penetrance in heterozygotes (lower line).

For the remainder, lod scores represent exclusions and were determined assuming X-linked recessive inheritance.





**Figure 6.3** Placement of *MRX19* into the multipoint background map.

The maximum location score was 16.5 at *DXS207* and *DXS999*, equivalent to a multipoint lod score of 3.6.

## 6.5 Mapping of *MRX30*

### 6.5.1 Materials and Methods

#### 6.5.1.1 The family

The 4 generation *MRX30* pedigree, is shown in Figure 6.4. This family was clinically investigated by Dr Michael Partington and Dr Anna Ryan, of the Royal Children's Hospital, Brisbane, when III-2 asked for genetic counselling because of the mental retardation present in several males in her family. The following clinical descriptions and family history were provided by Dr Partington:

The brother (III-1) of III-2 had been born at term after a normal pregnancy with a spontaneous vertex (head first) delivery. The birth weight was 3.26 kg. Feeding was difficult and he thrived poorly. He was slow to pass his developmental milestones and significant delay was recognised by age 2 years. He repeated kindergarten and had special schooling: his IQ was measured at 63 (mildly retarded). He was described as restless and hyperactive. After school he spent time in a sheltered workshop and then got episodic, casual employment in the open work force; he lived at home with his mother. On examination at age 31 years his height (167 cm) was at the lower limit of normal and his head circumference (52 cm) some 2.5 standard deviations below the mean. No physical abnormalities were found; his ear length was 60 mm and mean testicular volume 18 ml.

The clinical histories of III-4 and III-9 were very similar though with less feeding difficulties. The certificates admitting II-3 to institutional care as a teenager in 1944 told the same story, but in addition, there had been physical violence towards his mother. Physical examination of II-3 and III-9 showed no abnormalities.

The overall clinical picture was of mild intellectual handicap with heights and head circumferences below average but mostly within normal limits (refer to Donnelly *et al.*, 1996b; appendix A.4).

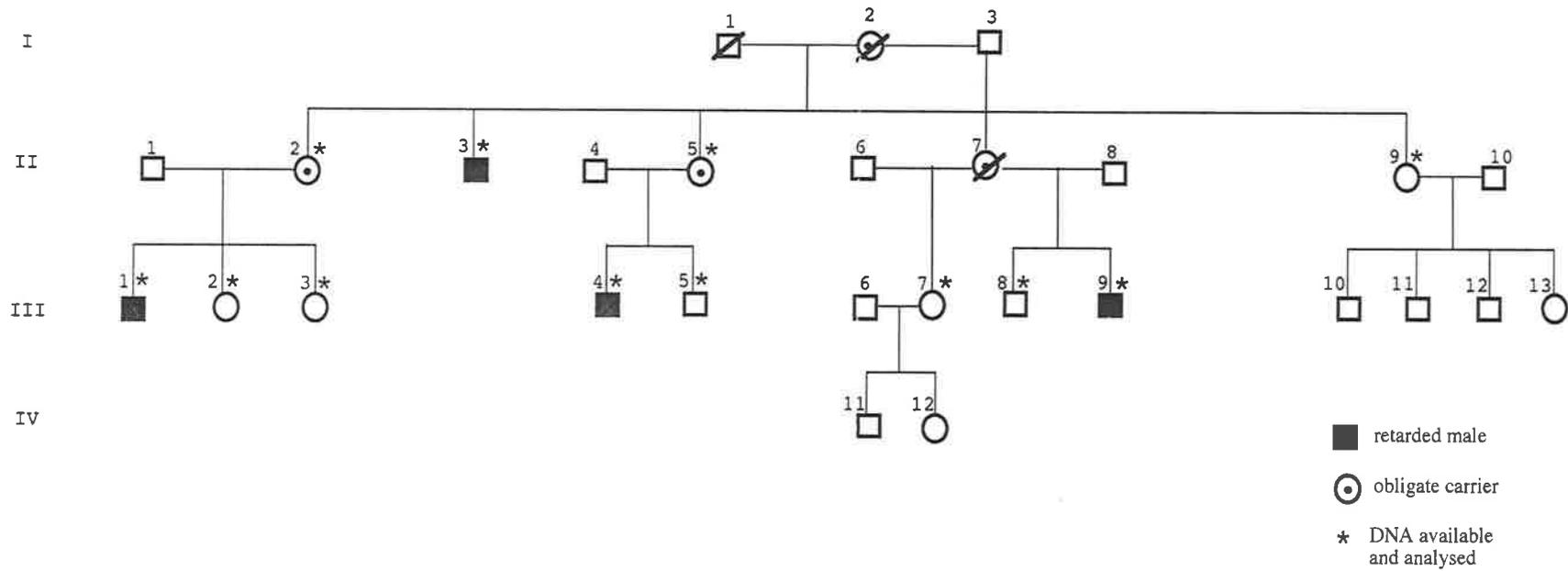


Figure 6.4 The *MRX30* pedigree.

### 6.5.2 Results

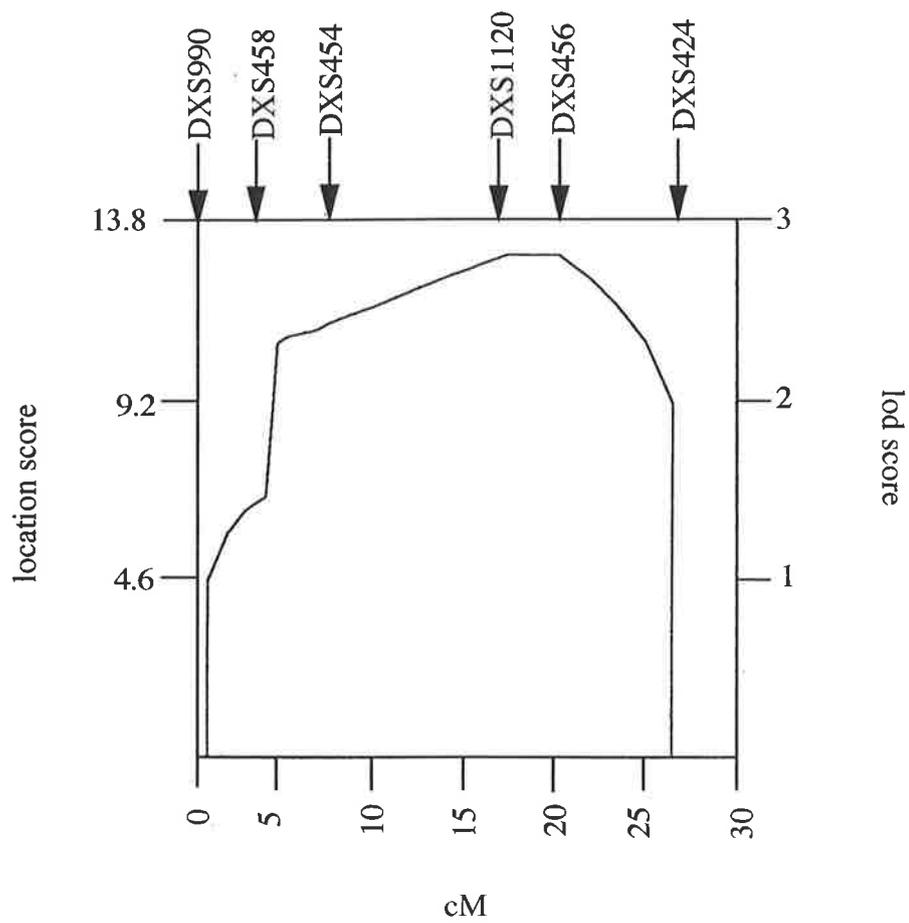
Of 24 marker genotypings inferred for individual II-7, 6 were incompatible with I-1 being her biological father. This required the inclusion of the inferred father (I-3) shown in Figure 6.4.

Two-point linkage analysis (Table 6.2) of the data from the *MRX30* family indicated that the locus most likely resides within a 28 cM region between the loci *DXS990* (Xq21.33) and *DXS424* (Xq23). None of the markers within this region were fully informative, which prevented the achievement of a significant lod score by two-point analysis. Multipoint analysis, however, yielded a significant multipoint lod score of 2.78 (Fig. 6.5), by incorporating informative markers from different parts of the pedigree into a single lod score. This is also the theoretical maximum from a two-point analysis and is consistent with the observation that every meiosis was informative for at least one of the markers within the linkage map used. Haplotyping revealed that recombination between the disease locus and the proximal (*DXS990*) and distal (*DXS424*) loci was exhibited in the female carrier II-5 and her son III-4, respectively (Fig. 6.6).

MLINK risk analysis using the informative flanking markers *DXS990* and *DXS424* resulted in a <1% chance of II-9 being a carrier. The chance of III-2 and III-3 being carriers is 6%, using the informative flanking markers *DXS990* and *DXS425*. The risk of III-7 being a carrier was 1%, using the flanking markers *DXS990* and *DXS424*. This information was conveyed to these women *via* the clinicians.

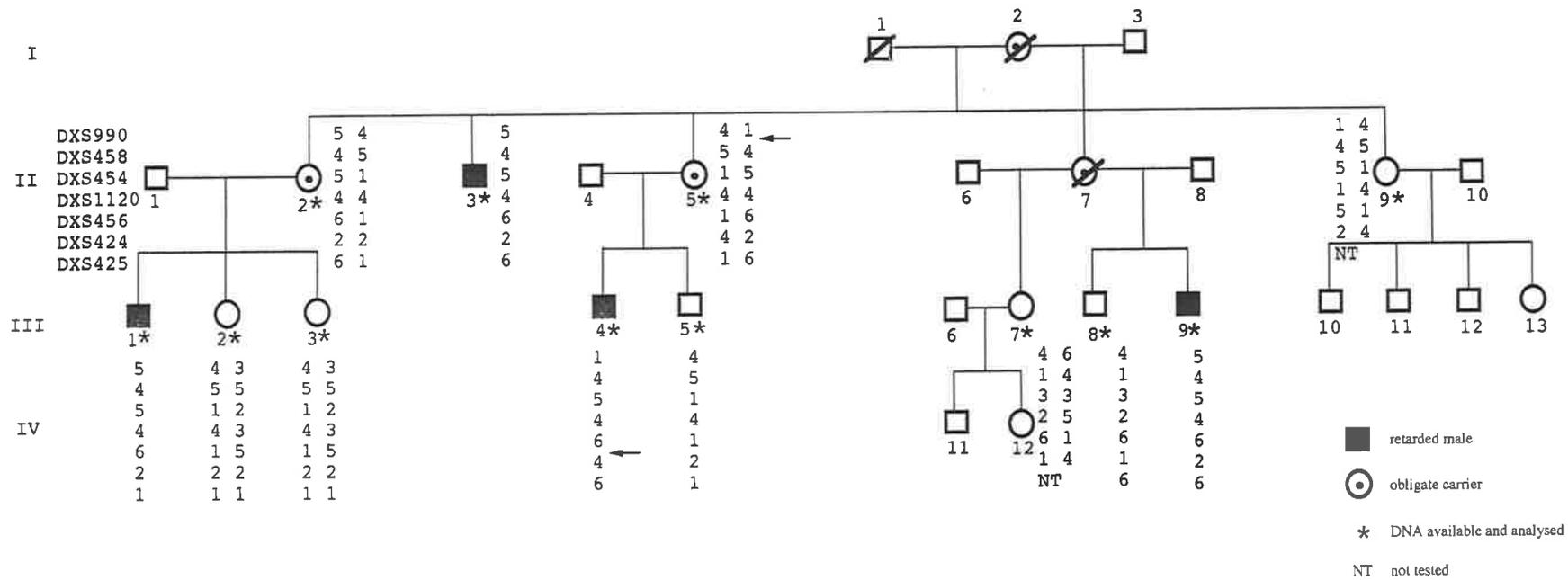
**Table 6.2 Two-point analysis between *MRX30* and marker loci  
(Xpter to Xqter)**

Locus	$\theta$							Z max	$\theta$ max
	0.00	0.01	0.05	0.1	0.2	0.3	0.4		
DXS996	$-\infty$	-7.53	-4.11	-2.70	-1.40	-0.73	-0.31		
DXS85	$-\infty$	-5.84	-3.13	-2.03	-1.03	-0.53	-0.21		
DXS999	$-\infty$	-3.67	-1.69	-0.94	-0.35	-0.15	-0.07		
DXS365	$-\infty$	-3.23	-1.77	-1.16	-0.64	-0.39	-0.20		
DXS989	$-\infty$	-3.49	-2.03	-1.38	-0.75	-0.40	-0.17		
DXS451	$-\infty$	-3.45	-1.94	-1.27	-0.64	-0.33	-0.13		
DYS(5')	$-\infty$	-3.30	-1.82	-1.15	-0.52	-0.21	-0.05		
DXS1068	$-\infty$	-2.73	-1.35	-0.78	-0.29	-0.07	0.01		
MAOA	$-\infty$	-3.40	-1.87	-1.19	-0.53	-0.21	-0.05		
DXS1003	$-\infty$	-4.40	-2.32	-1.45	-0.65	-0.26	-0.06	-0.06	0.4
DXS986	0.76	0.75	0.71	0.65	0.52	0.35	0.18	0.76	0.0
DXS990	$-\infty$	0.74	1.27	1.36	1.19	0.86	0.45	1.36	0.1
DXS458	1.51	1.49	1.41	1.30	1.03	0.72	0.38	1.51	0.0
DXS454	1.64	1.62	1.53	1.40	1.11	0.77	0.40	1.64	0.0
DXS1120	1.88	1.85	1.72	1.54	1.17	0.76	0.32	1.88	0.0
DXS456	1.85	1.82	1.70	1.55	1.20	0.80	0.36	1.85	0.0
DXS424	$-\infty$	0.45	0.99	1.10	0.99	0.71	0.37	1.10	0.1
DXS425	$-\infty$	-1.25	-0.55	-0.27	-0.03	0.05	0.05		
HPRT	$-\infty$	-3.16	-1.71	-1.06	-0.44	-0.14	-0.01		
DXS984	$-\infty$	-3.10	-1.66	-1.02	-0.40	-0.11	0.01		
DXS1123	$-\infty$	-7.48	-4.04	-2.61	-1.27	-0.59	-0.20		
DXS1108	$-\infty$	-3.91	-1.89	-1.08	-0.40	-0.12	-0.02		



**Figure 6.5** Placement of *MRX30* into the multipoint background map.

The maximum location score was 12.8 between *DXS1120* and *DXS456*, equivalent to a multipoint lod score of 2.8.



**Figure 6.6** The *MRX30* pedigree with haplotype data. The points of recombination with the disease gene are arrowed.

## 6.6 Mapping of *MRX31*

### 6.6.1 Materials and Methods

#### 6.6.1.1 The family

The 4 generation *MRX31* pedigree is shown in Figure 6.7. This family was also clinically assessed by Dr Michael Partington and Dr Anna Ryan in response to the parents' concern about the slow development of 3 of their sons (IV-1, 2 and 4). The family history showed that there was a similarly affected maternal cousin (IV-10) and great maternal uncle (II-7) suggesting X linkage. The following clinical descriptions and family history were provided by Dr Partington:

All 3 brothers and their cousin (IV-10) were born without assistance at or near term after normal pregnancies: birth weight ranged from 3.1 to 3.94 kg. There were no feeding difficulties. IV-1 and 2 both passed their gross motor milestones at the right times but were slow to start speaking: global developmental delay was recognised at 4 years in IV-1 and 2 years in IV-2. By clinical observation and the mother's reports, both seemed to have moderate (IQ 35-50) intellectual handicap. IV-4 and IV-10 were said to be hypotonic as young children. IV-10 had been assessed at 6 years of age and was said to have an IQ "in the range 64 to 74"; IV-4 seemed at the same level or a bit below. On examination no major abnormalities were found. IV-1 had large ears (length 65 mm) and a facial appearance suggesting the fragile X syndrome. IV-2 had a strabismus (squint) and wore glasses. Both these boys had a diastema (gap) between the upper central incisors, but so did their unaffected brother (IV-3). IV-10 had a single palmar crease on one hand and mild macrostomia (large mouth). The physical measurements of the 5 affected males are given in Donnelly *et al.* (1996b; appendix A.4).

II-7 had worked on a dairy farm for most of his adult life. He was illiterate and had lived with his mother and then in a nursing home. On examination he could carry on a simple conversation; no physical abnormalities were found.

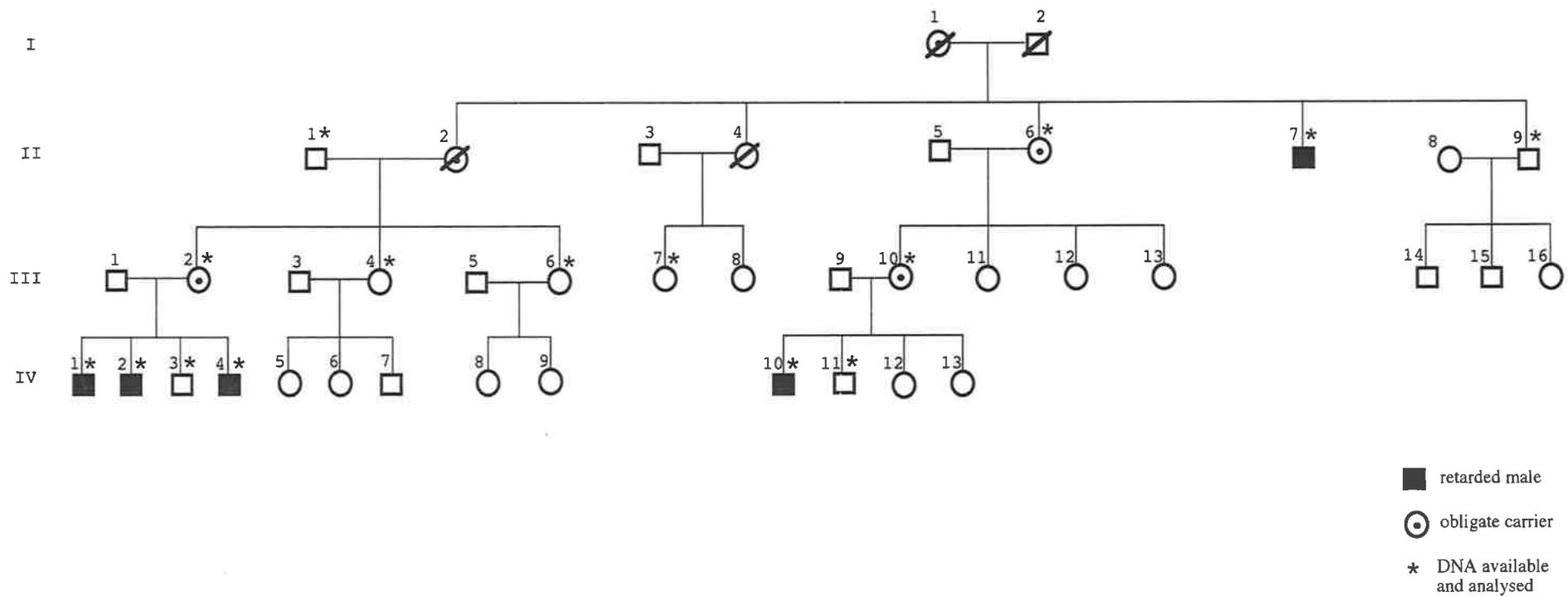


Figure 6.7 The *MRX31* pedigree.

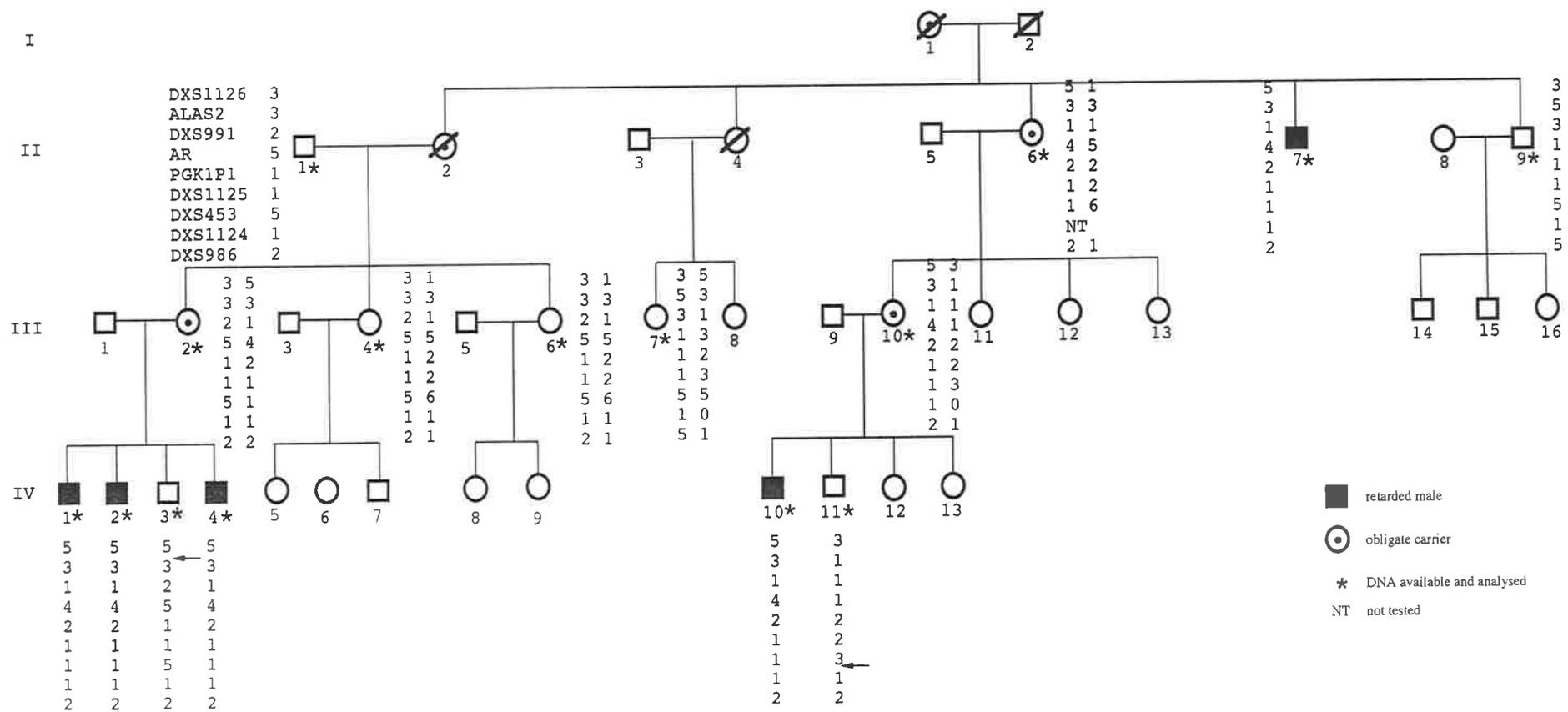
### 6.6.2 Results

Two-point linkage analysis of the data from the *MRX31* family (Fig. 6.7) showed that the localisation encompasses 12 cM from *DXS1126* (Xp11.23) to *DXS1124* (Xq13). Significant lod scores were obtained with the loci *DXS991* ( $Z_{\max}=2.06$ ), *AR* ( $Z_{\max}=3.44$ ), *PGK1P1* ( $Z_{\max}=2.06$ ) and *DXS453* ( $Z_{\max}=3.31$ ), all at recombination fractions of zero (Table 6.3). Recombination of the disease locus with the proximal locus (*DXS1126*) and the distal locus (*DXS1124*) was exhibited in the unaffected males IV-3 and IV-11 respectively (Fig. 6.8).

MLINK risk analysis using the informative flanking markers *DXS1126* and *DXS986* resulted in a <1% chance of III-4 and III-6 being carriers and a 12% chance of III-7 being a carrier. This information was conveyed to these women *via* the clinician.

**Table 6.3 Two-point analysis between *MRX31* and marker loci (Xpter to Xqter)**

Locus	$\theta$							Z max	$\theta$ max
	0.00	0.01	0.05	0.1	0.2	0.3	0.4		
DXS237	−∞	-2.24	-0.88	-0.36	0.05	0.16	0.13		
DXS16	−∞	-6.32	-3.53	-2.34	-1.21	-0.60	-0.23		
DXS999	−∞	-5.59	-2.87	-1.77	-0.77	-0.31	-0.07		
DXS365	−∞	-4.09	-2.09	-1.30	-0.57	-0.21	-0.04		
DXS989	−∞	-6.89	-3.49	-2.12	-0.89	-0.32	-0.06		
DXS451	−∞	-3.79	-1.80	-1.03	-0.39	-0.13	-0.02		
DXS992	−∞	-6.42	-3.05	-1.71	-0.56	-0.09	0.07		
DYS(5')	−∞	-6.59	-3.21	-1.86	-0.69	-0.18	0.02		
DXS1068	−∞	-6.30	-2.93	-1.61	-0.48	-0.03	0.10		
MAOA	−∞	-0.34	0.31	0.52	0.60	0.50	0.30		
DXS1003	−∞	-0.61	0.60	0.96	1.05	0.84	0.48		
PFC	−∞	0.62	1.16	1.25	1.10	0.77	0.38		
DXS1126	−∞	1.39	1.88	1.91	1.63	1.15	0.58	1.91	0.1
ALAS2	1.45	1.43	1.32	1.18	0.88	0.56	0.24	1.45	0.0
DXS991	2.06	2.02	1.88	1.69	1.29	0.85	0.40	2.06	0.0
AR	3.44	3.38	3.16	2.86	2.23	1.52	0.76	3.44	0.0
PGK1P1	2.06	2.02	1.88	1.69	1.29	0.85	0.40	2.06	0.0
DXS1125	0.83	0.83	0.80	0.74	0.57	0.35	0.15	0.83	0.0
DXS453	3.31	3.26	3.04	2.76	2.15	1.47	0.73	3.31	0.0
DXS1124	−∞	-1.54	-0.83	-0.52	-0.23	-0.09	-0.02	-0.02	0.4
DXS986	−∞	-0.10	0.49	0.63	0.61	0.42	0.18		
DXS1002	−∞	0.62	1.16	1.25	1.10	0.77	0.38		
DXS454	−∞	-1.38	-0.12	0.30	0.52	0.47	0.29		
DXS1120	1.47	1.47	1.42	1.34	1.12	0.81	0.44		
DXS456	−∞	-0.84	-0.17	0.07	0.22	0.22	0.13		
DXS424	−∞	-2.18	-0.86	-0.36	0.02	0.12	0.10		
DXS425	−∞	-4.60	-1.96	-0.95	-0.15	0.11	0.13		
HPRT	−∞	-5.87	-3.10	-1.95	-0.90	-0.39	-0.12		
DXS984	−∞	-4.09	-2.07	-1.27	-0.57	-0.25	-0.09		
DXS297	−∞	-4.94	-2.82	-1.91	-1.02	-0.53	-0.21		
DXS1123	0.63	0.60	0.48	0.34	0.13	0.02	-0.00		
DXS1108	−∞	-0.39	0.21	0.38	0.43	0.34	0.19		



**Figure 6.8** The *MRX31* pedigree with haplotype data. The points of recombination with the disease gene are arrowed.

## **6.7 Mapping of the XLMR gene affecting the W family**

### **6.7.1 Materials and Methods**

#### 6.7.1.1 The family

The 3 generation W family, is shown in Figure 6.9. This family was clinically investigated by Dr Martin Delatycki of the Victorian Clinical Genetics Services at the Royal Children's Hospital.

Clinical examination of the 6 affected males revealed mild intellectual disability, mildly coarse features, normal stature, normal head circumference, clinodactyly (curved fingers) and symmetrical brachydactyly (short digits). The unaffected male III-7 did not have the same appearance and his development was normal. The mothers of the affected males were normal in appearance and intelligence.

Amplification of the trinucleotide repeat at the *FRAXA* locus (Xq27.3) of II-12 was investigated by karyotyping and PCR, by the staff at the Royal Children's Hospital: no abnormality was detected.

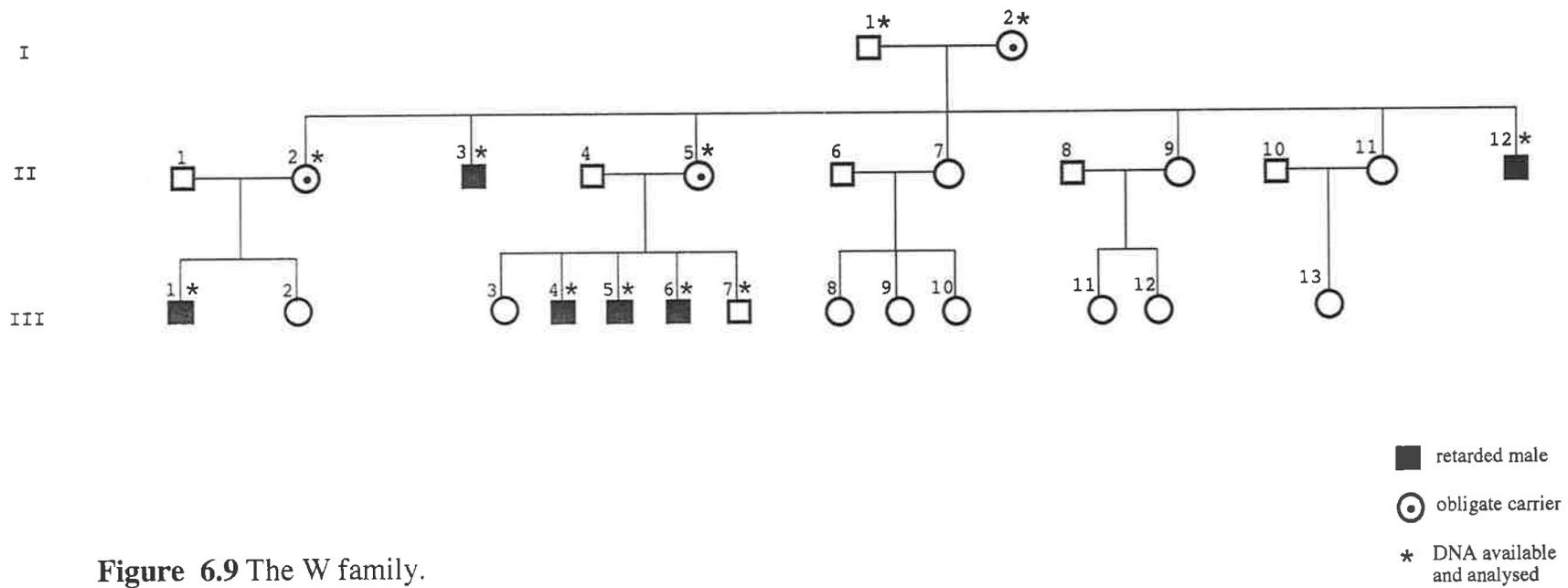


Figure 6.9 The W family.

### 6.7.2 Results

The results of the two-point linkage analysis between the disease locus and loci spanning the X chromosome are given in Table 6.4. A significant lod score of 2.41 was obtained with loci from two regions of the chromosome. The disease locus is therefore located somewhere between Xpter to *DXS1068* (Xp11.4) or between *MAOA* (Xp11.4-11.3) and *DXS1002* (Xq13.3-21.1), two very large regions that span  $\geq 91$  cM and 29 cM respectively. The locus has been excluded from most of the long arm of the X, distal to (and including) *DXS1002* (*DXS456*, in Xq22.3-q23, was not fully informative).

The haplotypes of the two regions are given in Figure 6.10.

**Table 6.4 Two-point analysis between the family W XLMR gene and marker loci (Xpter to Xqter)**

Locus	$\theta$							Z max	$\theta$ max
	0.00	0.01	0.05	0.1	0.2	0.3	0.4		
DXS1060	2.41	2.37	2.21	2.00	1.54	1.03	0.49	2.41	0.0
DXS996	2.41	2.37	2.21	2.00	1.54	1.03	0.49	2.41	0.0
KAL	2.41	2.37	2.21	2.00	1.54	1.03	0.49	2.41	0.0
DXS1043	2.41	2.37	2.21	2.00	1.54	1.03	0.49	2.41	0.0
DXS1224	2.41	2.37	2.21	2.00	1.54	1.03	0.49	2.41	0.0
DXS365	2.41	2.37	2.21	2.00	1.54	1.03	0.49	2.41	0.0
DXS451	2.41	2.37	2.21	2.00	1.54	1.03	0.49	2.41	0.0
DXS997	2.41	2.37	2.21	2.00	1.54	1.03	0.49	2.41	0.0
DYS (5'Dys-II)	2.41	2.37	2.21	2.00	1.54	1.03	0.49	2.41	0.0
DXS1068	$-\infty$	0.37	0.93	1.05	0.96	0.72	0.40		
MAOA	$-\infty$	-2.80	-1.44	-0.89	-0.39	-0.15	-0.04		
DXS1003	2.41	2.37	2.21	2.00	1.54	1.03	0.49	2.41	0.0
DXS991	2.41	2.37	2.21	2.00	1.54	1.03	0.49	2.41	0.0
DXS1125	1.51	1.48	1.39	1.28	1.02	0.73	0.40	1.51	0.0
DXS1002	$-\infty$	0.37	0.93	1.04	0.94	0.66	0.31		
DXS995	$-\infty$	-0.51	0.12	0.32	0.42	0.36	0.22		
DXS990	$-\infty$	-0.51	0.12	0.32	0.42	0.36	0.22		
DXS1120	$-\infty$	-2.51	-1.16	-0.63	-0.18	-0.01	0.04		
DXS456	0.90	0.89	0.81	0.72	0.52	0.30	0.09		
DXS424	$-\infty$	-3.62	-1.63	-0.87	-0.27	0.08	-0.04		
DXS425	$-\infty$	-5.61	-2.91	-1.82	-0.85	-0.38	-0.13		
HPRT	$-\infty$	-6.50	-3.72	-2.54	-1.39	-0.74	-0.31		
DXS984	$-\infty$	-2.80	-1.44	-0.89	-0.39	-0.15	-0.04		
DXS1123	$-\infty$	-2.80	-1.44	-0.89	-0.39	-0.15	-0.04		
DXS15	$-\infty$	-5.61	-2.91	-1.82	-0.87	-0.44	-0.21		

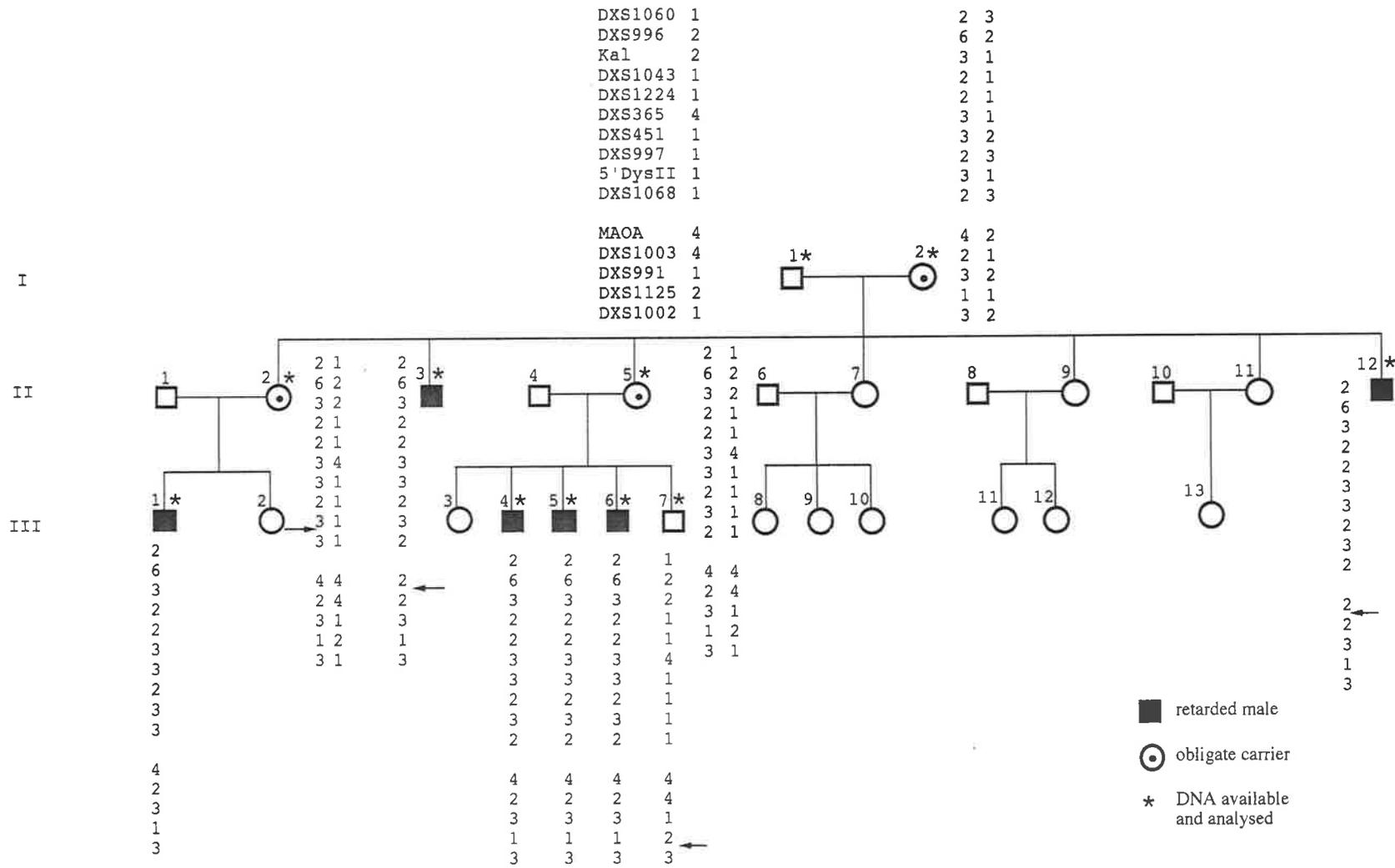


Figure 6.10 Family W with haplotype data. The points of recombination with the disease gene are arrowed.

## 6.8 Mapping of the XLMR gene affecting the O'H family

### 6.8.1 Materials and Methods

#### 6.8.1.1 The family

A very small portion of the 5 generation O'H family, is shown in Figure 6.11. The family was first investigated by Dr John Ferguson in 1987 and for this study follow-up clinical investigations were performed by Dr Michael Partington and Dr Gillian Turner of the Royal Children's Hospital, Queensland.

The affected males shown in Figure 6.11 were microcephalic and mild to moderately retarded, except for V-6 who, although mildly retarded, is physically different from the other affected males. IV-7 is also retarded but this is, according to his mother (III-9), believed to be due to a head injury acquired at 7 years of age, after developing normally.

#### 6.8.1.2 Linkage analysis

Lod scores are sensitive to the population frequencies of the alleles used in linkage analyses (Freimer *et al.*, 1993). To determine population allele frequencies for the loci *DXS453*, *DXS7113* and *DXS424* they were amplified and genotyped from the DNA of unrelated females. Forty alleles were scored for the locus *DXS453*, 40 alleles for the locus *DXS7113* and 76 alleles for the locus *DXS424*.

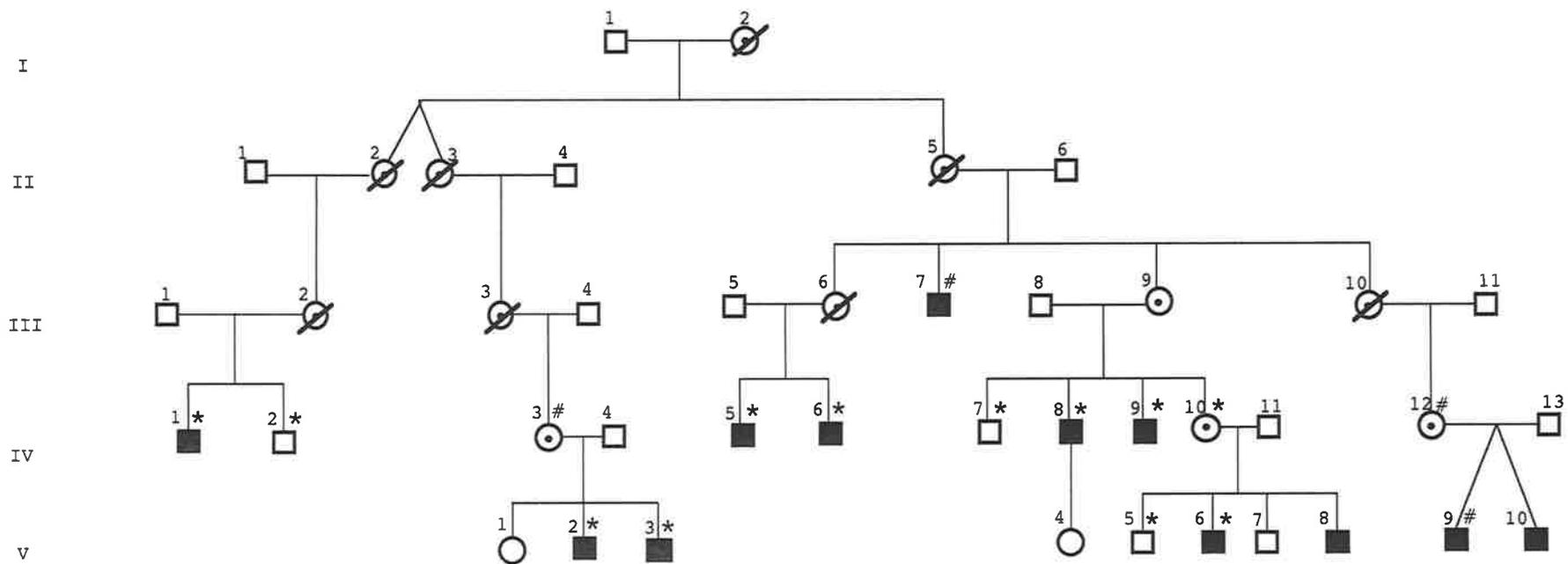


Figure 6.11 The O'H family.

- retarded male
- ⊙ obligate carrier
- /— deceased
- \* DNA available and analysed
- # DNA subsequently obtained and analysed

## 6.8.2 Results

Initial two-point lod scores excluded most of the X chromosome except for two relatively small regions on the long arm. If the affection status of the male V-6 is considered to be unknown (he is physically dis-similar to his affected relatives, see 6.8.1.1) the only positive lod scores, at zero recombination fraction, are 1.79, 1.0, and 1.79 with *DXS453* (Xq12), *DXS7113* (Xq12-q13.1) and *DXS424* (Xq23) respectively (Table 6.5). If V-6 is considered not to carry the XLMR gene there is no tight linkage with any of the markers used, at zero recombination fraction (data not shown), that is, the gene is excluded from the X chromosome, at least at the level of resolution investigated. If V-6 is considered to carry the XLMR gene the lod scores for *DXS453*, *DXS7113* and *DXS424* increase to 2.10, 1.30 and 2.10 respectively (second line in Table 6.5).

In an attempt to resolve which region is truly linked to the XLMR gene, further investigations were performed using: 1) additional family members not available at the time of the initial screen of the X chromosome (refer to Fig. 6.11); 2) population allele frequencies for *DXS453*, *DXS7113* and *DXS424* (Table 6.6), and 3) genotypic data from additional loci near the two regions of interest. The linkage results using this data are shown in Table 6.7. A recombination event revealed in V-9 excluded *DXS7113* from being tightly linked to the disease gene, but the lod scores for both *DXS453* and *DXS424* had increased to significant levels (2.55 and 2.14 respectively). As this additional information still did not indicate which region was most likely to contain the XLMR gene, haplotypes of the family members were constructed (Fig. 6.12), but these also failed to resolve the dilemma.

**Table 6.5 Two-point analysis between the family O'H XLMR gene and marker loci (Xpter to Xqter)†**

Locus	$\theta$							Z max	$\theta$ max
	0.00	0.01	0.05	0.1	0.2	0.3	0.4		
DXS996	∞	-2.33	-0.42	0.23	0.61	0.58	0.35		
DXS16	∞	-5.20	-2.50	-1.43	-0.53	-0.17	-0.02		
DXS999	∞	-2.86	-1.46	-0.88	-0.38	-0.17	-0.06		
DXS365	∞	-6.25	-3.28	-2.05	-1.00	-0.53	-0.25		
DXS989	∞	-4.48	-1.82	-0.80	0.01	0.26	0.23		
DXS451	∞	-4.34	-2.27	-1.42	-0.69	-0.34	-0.14		
DXS992	∞	-3.90	-1.87	-1.06	-0.38	-0.11	-0.01		
DYS (5'Dys-II)	∞	-2.04	-0.70	-0.20	0.16	0.22	0.14		
DXS1068	∞	-5.12	-2.90	-1.93	-1.00	-0.53	-0.23		
MAOA	∞	-0.69	-0.03	0.21	0.33	0.28	0.16		
DXS1003	∞	-1.31	-0.61	-0.33	-0.12	-0.06	-0.03		
DXS991	∞	-0.58	0.04	0.22	0.25	0.16	0.06		
ALAS2	∞	-0.54	0.08	0.27	0.30	0.19	0.08		
DXS1125	∞	-0.54	0.08	0.27	0.30	0.19	0.08		
DXS1275	∞	-1.03	-0.34	-0.08	0.07	0.08	0.04		
DXS453	1.79	1.77	1.65	1.48	1.09	0.69	0.31	1.79	0.00
	2.10	2.07	1.93	1.74	1.30	0.83	0.39	2.10	0.00
DXS7113	1.00	0.99	0.96	0.88	0.67	0.42	0.19	1.00	0.00
	1.30	1.29	1.24	1.14	0.87	0.57	0.27	1.30	0.00
DXS559	∞	-1.97	0.22	0.55	0.60	0.42	0.18		
DXS1124	∞	-1.08	0.10	0.44	0.51	0.35	0.15		
DXS986	∞	0.10	0.66	0.77	0.67	0.45	0.21		
DXS995	∞	-10.81	-6.03	-4.06	-2.17	-1.13	-0.45		
DXS990	∞	-3.06	-1.09	-0.37	0.14	0.25	0.18		
DXS454	∞	-0.79	0.40	0.73	0.78	0.58	0.29		
DXS1120	∞	-3.18	-1.20	-0.46	0.09	0.22	0.17		
DXS456	∞	-1.60	-0.33	0.10	0.33	0.29	0.16		
DXS1072	∞	0.10	0.66	0.77	0.67	0.45	0.21		
DXS1059	∞	-0.79	0.40	0.73	0.78	0.58	0.29		
DXS424	1.79	1.77	1.65	1.48	1.09	0.69	0.31	1.79	0.00
	2.10	2.07	1.93	1.74	1.30	0.83	0.39	2.10	0.00
DXS1212	∞	-3.14	-1.72	-1.13	-0.59	-0.32	-0.14		
DXS1001	∞	-0.63	0.55	0.87	0.89	0.64	0.32		
DXS425	∞	-3.86	-1.83	-1.03	-0.37	-0.10	-0.01		
HPRT	∞	-7.91	-4.30	-2.77	-1.35	-0.65	-0.24		
DXS102	∞	-0.94	-0.28	-0.04	0.12	0.14	0.09		
DXS292	∞	-5.73	-2.98	-1.86	-0.87	-0.41	-0.16		
DXS297	∞	-3.78	-1.74	-0.93	-0.27	-0.03	0.04		
DXS1123	∞	-0.82	-0.24	-0.07	-0.03	-0.04	-0.03		
DXS1108	∞	-4.40	-2.20	-1.28	-0.50	-0.18	-0.05		

†Using genotypes of only those people indicated with \* in Fig. 6.11.

The lod scores were calculated with the affection status of V-6 classed as unknown, except those given in the second line for the loci *DXS453*, *DXS7113* and *DXS424*, where V-6 was classed as affected.

**Table 6.6 Population allele frequencies**

locus	allele	number observed	frequency
<i>DXS453/983</i>	1*	5	0.125
	2	1	0.025
	3~	22	0.55
	4	4	0.1
	5	1	0.025
	6	7	0.175
<i>DXS7113</i>	1*~	16	0.4
	2	10	0.25
	3	2	0.05
	4	1	0.025
	5	11	0.275
<i>DXS424</i>	1~	38	0.5
	2*	6	0.08
	3	16	0.21
	4	5	0.06
	5	4	0.05
	6	3	0.04
	7	2	0.03
	8	2	0.03

\* allele possessed by unaffected male IV-2.

~ allele possessed by affected male IV-9.

**Table 6.7 Two-point analysis between the family O'H XLMR gene and additional marker loci, using population allele frequencies\* (Xpter to Xqter)**

Locus	$\theta$							Z max	$\theta$ max
	0.00	0.01	0.05	0.1	0.2	0.3	0.4		
DXS1275	$-\infty$	-2.20	-0.81	-0.29	0.06	0.11	0.07		
DXS8111	$-\infty$	-2.59	-0.78	-0.06	0.36	0.34	0.18		
DXS453	2.55	2.58	2.54	2.35	1.82	1.19	0.55	2.55	0.00
	2.85	2.87	2.81	2.61	2.02	1.33	0.63	2.85	0.00
DXS7113	$-\infty$	0.00	0.65	0.83	0.78	0.53	0.25		
DXS559	$-\infty$	-2.50	-0.06	0.05	0.40	0.37	0.20		
DXS1059	$-\infty$	-0.72	0.47	0.81	0.87	0.66	0.35		
DXS8081	$-\infty$	-1.31	-0.08	0.30	0.45	0.34	0.17		
DXS1220	$-\infty$	-0.67	0.50	0.82	0.85	0.63	0.33		
DXS424	2.14	2.11	1.96	1.74	1.27	0.80	0.36	2.14	0.00
	2.44	2.41	2.23	2.00	1.48	0.95	0.44	2.44	0.00
DXS8064	$-\infty$	-0.19	0.93	1.18	1.07	0.73	0.34		
DXS8067	$-\infty$	-1.26	-0.06	0.30	0.40	0.28	0.11		
DXS8059	$-\infty$	-4.46	-1.86	-0.91	-0.20	0.02	0.06		
DXS1212	$-\infty$	-2.47	-1.15	-0.65	-0.29	-0.15	-0.07		

Lod scores were calculated using additional members from the family (ie using \* and # in Fig. 6.11).

The affection status of V-6 was classed as unknown, except for the second line for the loci *DXS453* and *DXS424*, where he was classed as affected.

\*Lod scores for loci *DXS453*, *DXS7113* and *DXS424* were obtained using population allele frequencies (Table 6.6).

The order of loci was deduced from Nelson *et al.* (1995) and Dib *et al.* (1996).

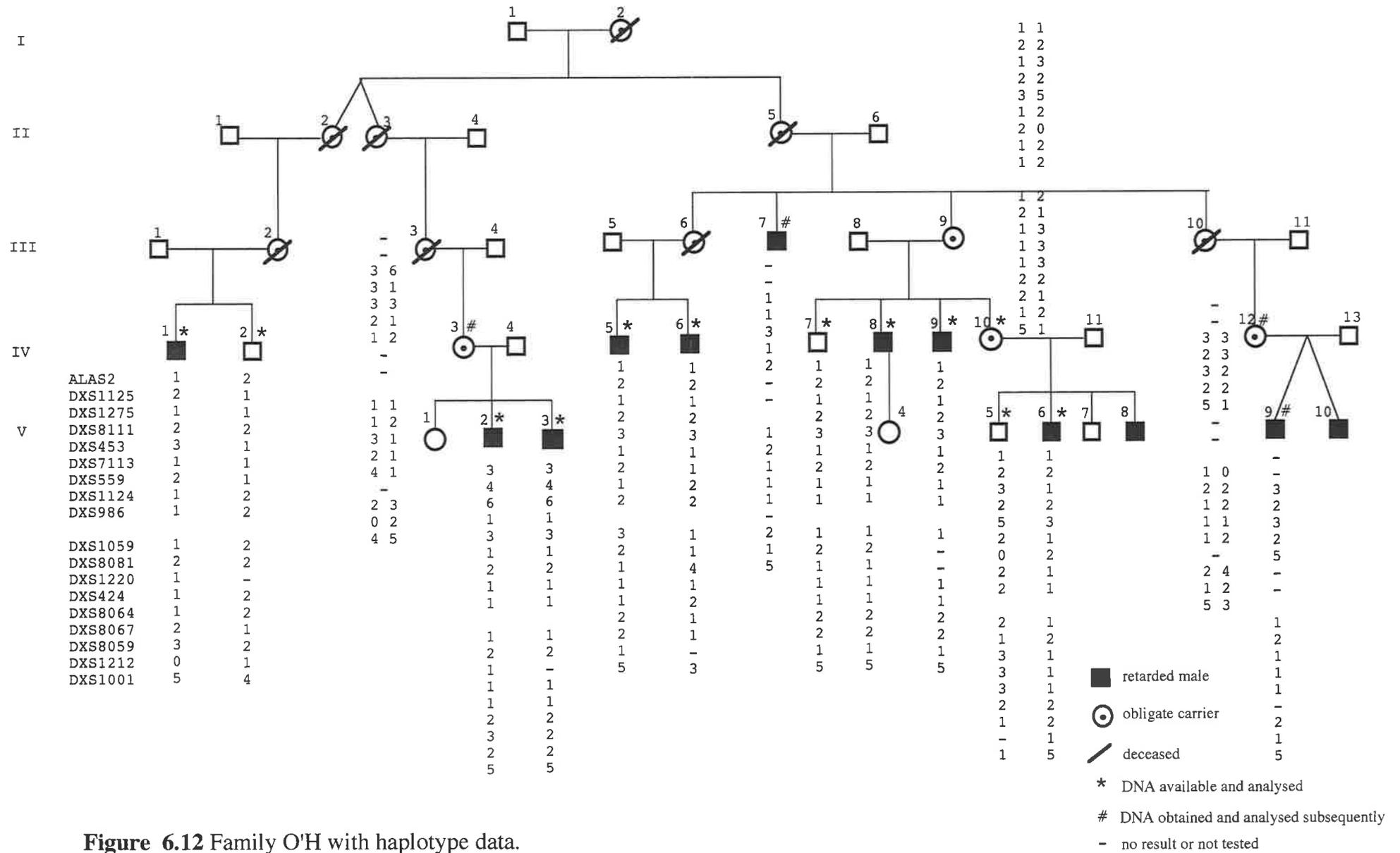


Figure 6.12 Family O'H with haplotype data.

## 6.9 Discussion

### 6.9.1 *MRX19*

*MRX19* has now been mapped to Xp22, flanked by the *DXS1043* and *DXS1052* loci. Eight other *MRX* localisations overlap that of *MRX19* (Table 6.8; Fig. 6.13). The non-overlapping *MRX2* and *MRX37* localisations indicate that there are at least 2 genes responsible for non-specific mental retardation in Xp22. Given that there are two individuals (III-7 and II-8) in the *MRX19* family that are recombinant with the locus *DXS1043*, there is the potential to reduce the distal localisation of *MRX19* further, using informative microsatellite markers between *DXS1043* and *DXS987/207*, so that one of these individuals is no longer a recombinant. This is reflected in the multipoint lod-1 (90%) support interval (Fig. 6.3) which shows that the region within which the disease gene most likely resides begins about 2.5 cM proximal to *DXS1043*.

Three of the 4 carrier females from the *MRX19* family are mildly affected. The *MRX2*, *MRX21*, *MRX24*, *MRX36* and *MRX37* families, which span or overlap *MRX19*, also have affected female carriers. Heterozygote manifestation of retardation is not confined to these *MRX* families, but has also been observed in families whose disease genes map to other regions of the X chromosome (Hane *et al.*, 1996). Observations of the number of those families with affected females may be biased however, because such families may be more easily ascertained due to the relatively increased proportion of affected members.

The manifestation of symptoms in some of the carrier females in families in which a rare disease allele is inherited can be most readily explained by skewed lyonisation (X inactivation; reviewed by Belmont, 1996). Lyonisation is the process whereby one of the two X chromosomes is inactivated early in development (reviewed by McBurney, 1993; Tan *et al.*, 1993). Detrimental effects may occur if there is a high proportion of cells, in any given tissue, in which the X chromosome with the defective gene remains active. The skewing of X inactivation, so that the structurally or genetically defective X remains active, has been reported and is typical of, but not unique to, cases involving balanced X-autosome translocations (Mattei *et al.*, 1982; Verellen-Dumoulin *et al.*, 1984; Nisen *et al.*, 1986; Tihy *et al.*, 1994).

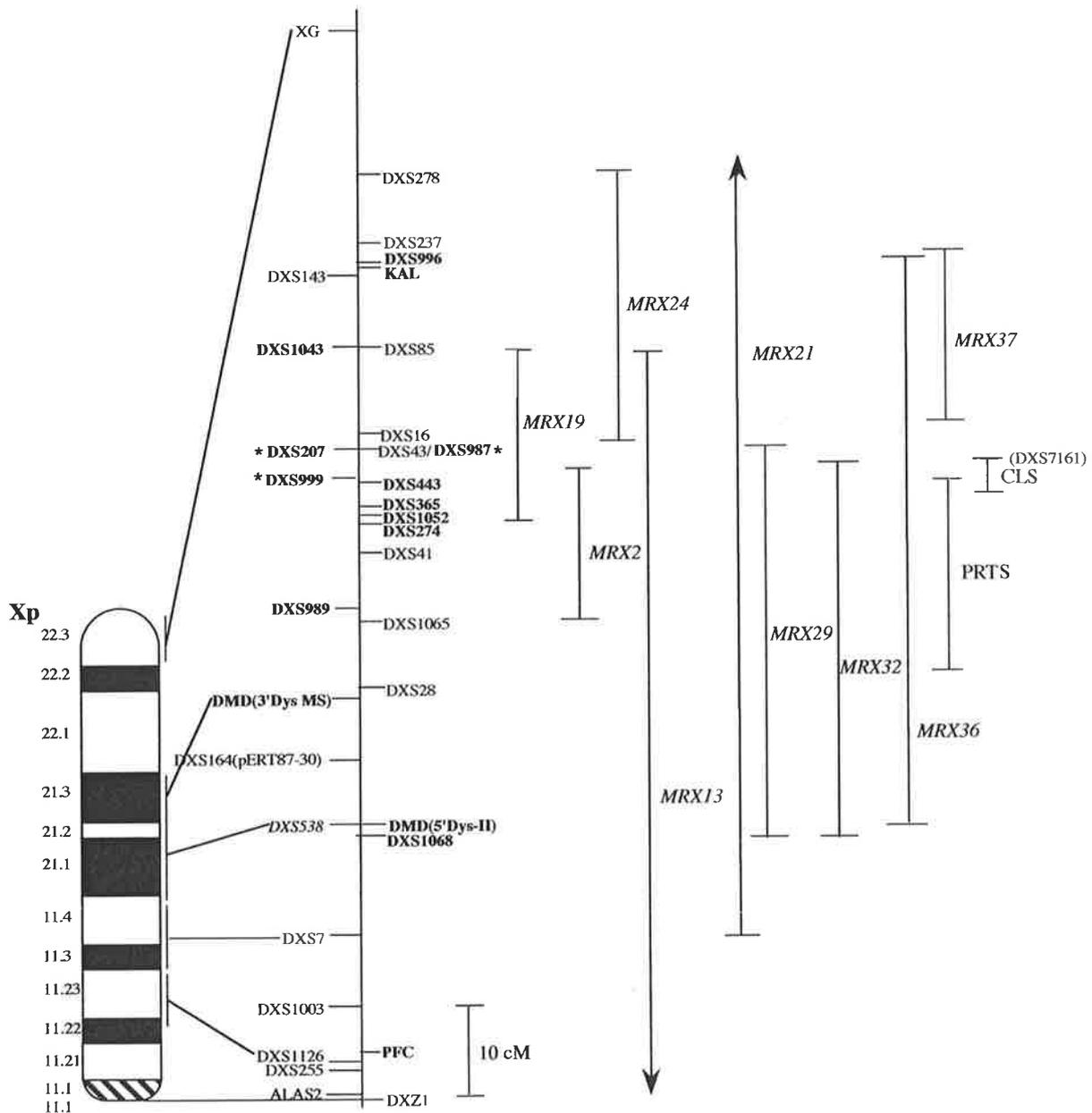
**Table 6.8 Regional localisations of the *MRX* loci (where known)**

<i>MRX</i>	localisation (Xpter to Xqter)	Reference
1	DXS426-DXS1125	Gedeon <i>et al.</i> , 1996b
2	DXS999-DXS1065	Hu <i>et al.</i> , 1994
3	DXS304-qter	Gedeon <i>et al.</i> , 1991
4	DXS255-DXYS1	Hu <i>et al.</i> , 1994
5	OTC-DXS95	Samanns <i>et al.</i> , 1991
6	DXS369 <sup>a</sup>	Kondo <i>et al.</i> , 1991
7	DXS7-DXS3	Jedele <i>et al.</i> , 1990
8	MAOA-DXS454	Schwartz <i>et al.</i> , 1992
9	DXS164-DXS159	Willems <i>et al.</i> , 1993
10	DXS28-DXS228	Gedeon <i>et al.</i> , 1996b
11	DXS164-DXS7	Gedeon <i>et al.</i> , 1996b
12	DXS1238-ALAS2	Gedeon <i>et al.</i> , 1996b
13	DXS85-DXS456	Kerr <i>et al.</i> , 1992
14	PFC-DXS135	Gendrot <i>et al.</i> , 1994; Gedeon <i>et al.</i> , 1996b
15	DXS164-DXS988	Raynaud <i>et al.</i> , 1996
16	DXS52-qter	Moraine <i>et al.</i> , 1994
17	DXS255-DXS106	Gedeon <i>et al.</i> , 1996b
18	DXS538-DXS1126	Gedeon <i>et al.</i> , 1994
19	DXS1043-DXS1052	Donnelly <i>et al.</i> , 1994a (this thesis)
20	DXS1068-DXS454	Lazzarini <i>et al.</i> , 1995
21	pter-DXS7	Kozak <i>et al.</i> , 1993
22	DXS84-DXYS1	Passos-Beuno <i>et al.</i> , 1993
23	DXS1220-DXS424	Gregg <i>et al.</i> , 1996
24	DXS278-DXS207	Martinez <i>et al.</i> , 1995
25	DXS297-qter	Nordstrom <i>et al.</i> , 1992
26	MAOB-DXS454	Robledo <i>et al.</i> , 1996a
27	DXS424-DXS102	Gedeon <i>et al.</i> , 1996a
28	DXS1113-qter	Holinski-Feder <i>et al.</i> , 1996
29	DXS987-DXS1068	Hane <i>et al.</i> , 1996
30	DXS990-DXS424	Donnelly <i>et al.</i> , 1996b (this thesis)
31	DXS1126-DXS1124	Donnelly <i>et al.</i> , 1996b (this thesis)
32	DXS1053-DXS1110	Hane <i>et al.</i> , 1995
33	DXS365-MAOB	Holinski-Feder <i>et al.</i> , 1996
34	DXS989-DMD	Raeymaekers <i>et al.</i> , 1996
35	DXS178-HPRT	Gu XX <i>et al.</i> , 1996
36 <sup>b</sup>	DXS996-DysII	Claes <i>et al.</i> , 1996
37	DXS1223-DXS1224	Bar-David <i>et al.</i> , 1996
38	DXS1226-DXS1238	Schutz <i>et al.</i> , 1996
39 <sup>c</sup>	DXS6673E	van der Maarel <i>et al.</i> , 1996
40	(not available)	
41	DXS52-qter	Hamel <i>et al.</i> , 1996
FRAXE	Xq28	Knight <i>et al.</i> , 1993

<sup>a</sup>Flanking recombinant markers are unpublished or unknown. The loci with which maximum lod scores were obtained are given.

<sup>b</sup>Localisation is dependent on the model used. The maximum region is shown here.

<sup>c</sup>The *MRX* designation should only be applied to families where linkage between the disease gene and X-linked markers has been demonstrated with a lod score of  $\geq 2.0$  (Mulley *et al.*, 1992), and not to sporadic cases with translocations, as has happened here.



**Figure 6.13** Genetic map showing the regional localisation of the *MRX19* gene. The Xp loci used in the linkage analysis are shown in bold. A new comprehensive map was constructed using all the markers between and including *DXS996* and *DXS989*, except for the loci: *KAL* and *DXS1052* (see 6.4.1.2). The position of *DXS1065* was taken from Hu *et al.* (1994). This new map was positioned between the flanking Xp loci (*DXS237* and *DXS28*) of the map given in Chapter 5. Framework markers are shown on the left hand side. Fully informative loci with no recombinants are indicated (\*). The regional localisations of the *MRX* genes from 8 other families are shown, as are the localisations of the genes responsible for Coffin-Lowry Syndrome (CLS) and Partington Syndrome (PRTS).

Skewing of X inactivation could be due to 3 factors:

1) Chance. Given that X inactivation occurs early in embryonic development, the X inactivation pattern of those cells that ultimately populate a particular tissue is a consequence of the number of precursor cells and *their* pattern of X inactivation. Disease symptoms can arise if all or most cells of a tissue are derived, by chance, from precursors in which the X chromosome that remained active contained the defective gene. Chance may account for the skewed inactivation of the X in about 10% of normal females, as observed by Harris *et al.* (1992) and Gibbons *et al.* (1992). Variations in the tissue patterns of X inactivation may also explain intrafamilial variation in the phenotypes of carriers.

2) Selective (non-random) inactivation of the X carrying the normal allele. Which parental X chromosome is inactivated may be determined by genetic factors that control the process, that is, there is "X inactivation" allelic variation in the population (Naumova *et al.*, 1996).

3) Cell selection. This process would normally protect carrier females from symptoms. However, selection against cells with an X chromosome with a major defect (in terms of cell survival) can still result in symptoms of an X-linked recessive disease if the X that remains active has a minor gene defect (see for example Nisen *et al.*, 1986). Alternatively, a gene defect may imbue a cell with a proliferative advantage (ie this mutation would be described as "dominant"). Those cells in which the mutated X remains active would outgrow those that do not. This process may apply to those carrier females who exhibit symptoms of adrenoleukodystrophy (Migeon *et al.*, 1981).

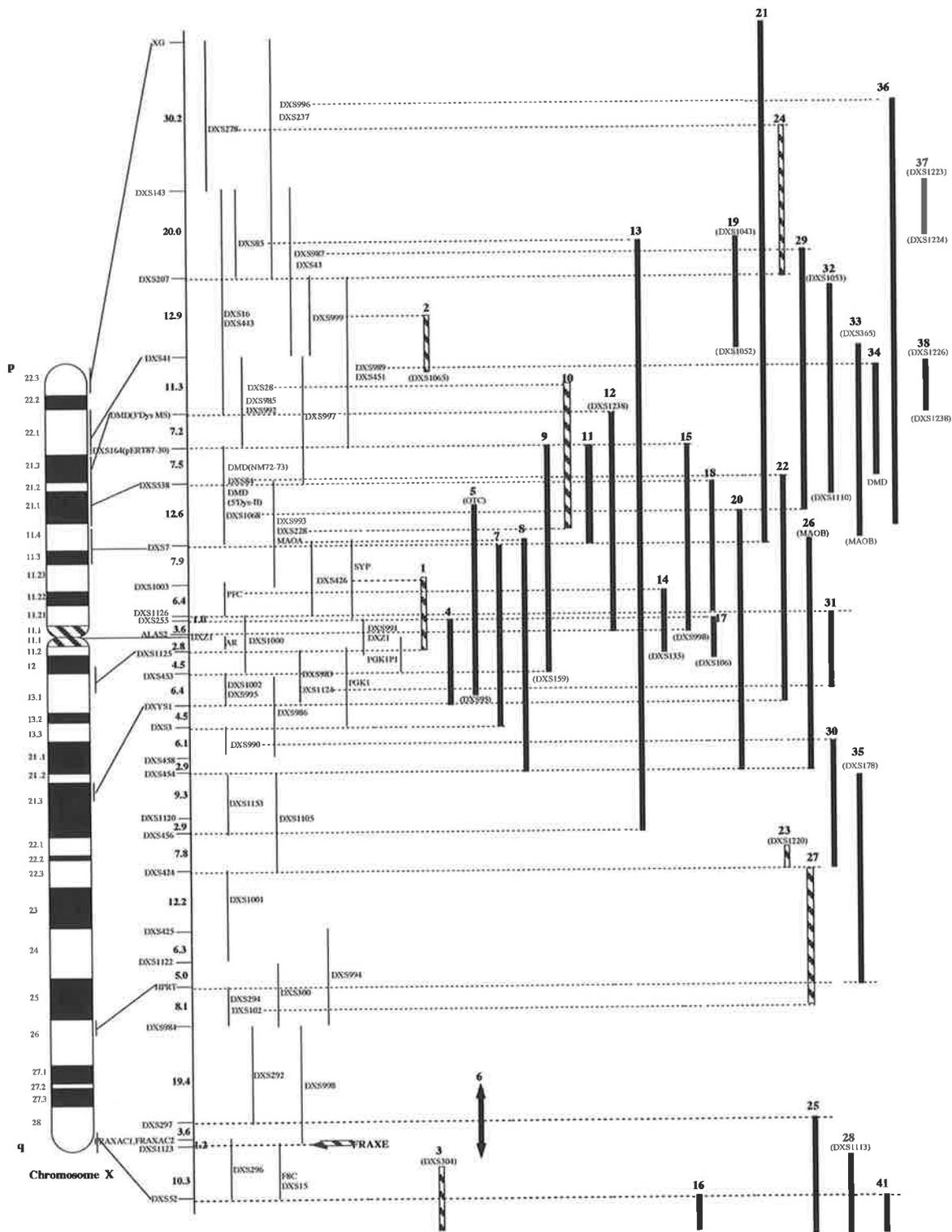
Despite the fact that the *MRX2* family was considered to have non-specific retardation, and hence assigned the *MRX* notation (Arvelier *et al.*, 1988), most of the affected males have additional features in common: short stature, highly arched palates and macro-orchidism or testicular volumes in the upper normal range (Proops *et al.*, 1983). (It was not specifically mentioned, however, that these features were not present in the mentally normal males of the pedigree, so they may be familial traits (albeit unusual in terms of population statistics unrelated to the *MRX* gene). If the gene involved is the same as that in *MRX19* these additional symptoms may be explained by allelic heterogeneity or the involvement of other genes possibly *via* a microdeletion or some form of position effect.

There are two syndromal forms of mental retardation that have also been confined to Xp22 (Fig. 6.13). Partington syndrome (PRTS or *MRXS1*; MIM 309510) has the additional neurological features of dysarthria, ataxia and seizures (Partington *et al.*, 1988). The locus has been mapped to Xp22.1-21.3 between *DXS365* and *DXS28* (Gedeon *et al.*, 1994b). The second form is Coffin-Lowry syndrome (CLS; MIM 303600), characterised by facial and digital abnormalities and progressive skeletal malformation (Hunter *et al.*, 1982; Young, 1988). CLS has been mapped to a 3.4 cM region between the loci *DXS7161* and *DXS365*, which fits within the overlap of *MRX2* and *MRX19* (Bird *et al.*, 1995). The CLS cDNA (*Rsk-2*) has recently been identified (Trivier *et al.*, 1996) and is thus a candidate for also being responsible for *MRX19*. This possibility has not yet been investigated because the genomic structure of *Rsk-2* has not yet been elucidated.

#### 6.9.2 *MRX30* and *MRX31*

Inspection of the *MRX30* pedigree (Fig. 6.4) prior to collection of blood samples revealed that the *MRX* gene might be unmappable, in the sense that a significant lod score would be unattainable, because of the small size of the family. To obtain a significant lod score from this family the meioses leading to II-2, II-3, II-5 and II-7 needed to be informative. II-2, II-3, II-5 and II-7 all received the *MRX* gene, hence their chromosomal segment within the immediate vicinity of the *MRX* gene would be derived from the same homologue, which would preclude inference of heterozygosity for I-1 for tightly linked markers. This would remain the case if II-8 was a carrier, however, given that she had three normal sons it was considered likely that she received the wildtype *MRX*-containing segment from the alternative homologue and therefore it was hoped that the marker genotypes of II-8 could be used to reveal any marker heterozygosity in I-1. This was in fact the case and II-8 was crucial for the successful mapping of *MRX30* even though her carrier status was unknown.

The Xq *MRX30* localisation overlaps those of *MRX8*, *13*, *20*, *23*, *26* and *35* (Table 6.8; Fig. 6.14). The *MRX31* localisation spans the centromere and overlaps most of the remaining *MRX* loci.



**Figure 6.14** Map of the regional localisations of the *MRX* loci.

The striped localisations highlight the existence of non-overlapping *MRX* genes.

### 6.9.3 Family W

Unfortunately, the very large regions in which the disease locus may be found makes it pointless to offer pre-natal or carrier risk analysis to any females from this family whose carrier status is unknown. This is because the chance of recombination within these two regions is considerable. In addition, there is no p arm terminal locus known to have undergone recombination with the disease locus. The only way that the regional localisations could be reduced would be to genotype other distantly related males from the pedigree (if any exist), in the hope of encountering an informative recombination event. The carrier status of the females (not yet genotyped) could be predicted by assaying which of the grandmaternal (I-2) haplotypes each has inherited for the two potential *MRX*-containing regions of the chromosome, providing these regions have remained intact. For example, if female III-3 had the same grandmaternal haplotype as the unaffected male III-7 for the two high risk regions of the X chromosome, then it can be stated that she is unlikely to be a carrier. On the other hand, if she has the same haplotype as the affected males, for both high risk regions, it can be concluded she is a carrier. If however, she has inherited a chromosome that has undergone recombination within the two high risk regions then it would not be possible to determine her carrier status because the *MRX* gene may be within the region of recombination. Haplotypes of the two regions are given in Figure 6.10.

### 6.9.4 Family O'H

Given the size of this family, and therefore the potential for a large number of recombination events, it was hoped that the disease gene would be localised to a relatively small region. This is in fact the case, but unfortunately to *two* relatively small regions. The significant lod score obtained with one of these two regions is an example of false close linkage, and demonstrates that a lod score of 2 is not sufficient evidence to conclude tight linkage to a particular locus unless the remainder of the X chromosome has been excluded (this has also been demonstrated by Robledo *et al.*, 1996b).

Much genotypic information is missing from the family (Fig. 6.11) and thus it is not always possible to track the inheritance, between and within branches, of the alleles inherited by the affected males (that is, the alleles may be identical by state and not by

descent). If the disease gene was being inherited with a rare allele, this would boost the lod score and conversely if the disease gene was inherited with a common allele the lod scores would decrease (recall that routinely for two-point analyses, alleles are arbitrarily assigned equal frequencies; 3.2.21.1). So, to get a more accurate lod score, given the absence of DNA from key family members, population allele frequencies were determined for the loci *DXS453*, *DXS7113* and *DXS424* (Table 6.6) and used in a second two-point linkage run (Table 6.7). Inheriting a rare allele for any of these 3 loci would provide evidence that this allele was derived from the female at the top of the pedigree (I-2), that is, the allele is likely to be identical-by-descent and unlikely to have entered the pedigree from a married-in individual. Unfortunately, for all three loci, the affected males had the most common allele segregating in the general population. Genotyping additional family members and additional markers also failed to provide evidence supporting one region over the other.

Haplotyping of the two candidate regions was done in the hope of revealing: 1) an unaffected male with a high risk haplotype flanking either *DXS453* or *DXS424* or 2) an affected individual with a low risk haplotype flanking either *DXS453* or *DXS424*. However, it is difficult to identify putative recombination events involving the regions around *DXS453* and *DXS424* because the DNA from most of the carrier females of the family was unavailable for genotyping.

To explain the presence of an unaffected male with a flanking high risk haplotype, either a double recombination event has removed the disease gene from the high risk haplotype or, as is more likely, no double recombination event has occurred and the disease gene lies elsewhere. To explain the presence of an affected male with a flanking low risk haplotype, either a double recombination event has added the disease gene to the low risk haplotype or, as is more likely, the disease gene is located elsewhere. Applying this argument to the O'H family; the affected male IV-6 has a low risk haplotype flanking *DXS424*. The probability of a double recombination event (that transfers the putative XLMR gene to this haplotype) within this relatively small region is at least 1 in 870 ( $0.025 \times 0.046$ ; see map below). Therefore it is more likely that his mother (III-6) is homozygous for *DXS424* and that IV-6 does not have a recombinant haplotype. On this basis it is the Xq12 region that is truly linked to the XLMR gene.

The male IV-7 is mentally retarded but this is thought to be due to a head injury and so he is not considered to carry the disease gene. He, however, apparently has the same haplotypes for both regions as his affected brothers and carrier sister. For him not to carry the disease gene a double recombination event *is* likely to have occurred over a relatively small region, either in Xq12 or Xq23 (his mother (III-9) is unlikely to be homozygous for most of the loci haplotyped so her being homozygous cannot be used to explain away his apparently high risk haplotypes). If IV-7 does not carry the disease gene, the double recombination event (and hence the XLMR gene) is perhaps more likely to occur further away from the centromere (where the recombination frequency is relatively low, on average), that is, in Xq23 rather than Xq12. The Xq23 localisation overlaps the coincident localisations of *MRX30* and *MRX23*. Affected males of all three families have microcephaly. The phenotype of affected males of the *MRX30* family is given in 7.5.1.1 and Donnelly *et al.* (1996b; appendix A.4). The *MRX23* family has been clinically assessed by Lehrke (1974; Family 1) and Howard-Peebles and Roberts (1984). A defective gene common to all three families may explain the similar phenotypes. Alternatively IV-7 *does* carry the disease gene and his mother's claim that he developed normally until his accident (6.8.1.1) is incorrect.

The genetic distances (cM) for the two regions have been obtained or estimated from the genetic maps of Fain *et al.* (1995) and Dib *et al.* (1996) are as follows:

*DXS1275* - 1.7 - *DXS453* - 0.9 - *DXS1124*

and

*DXS1220* - 2.5 - *DXS424* - 4.6 - *DXS8067* - 2.9 - *DXS1001*

From Nelson *et al.*, (1995) the physical distances (Mb) are estimated as:

*DXS1275* - 2.0 - *DXS453* - 1.0 - *DXS7113* - 2.0 - *DXS559*

and

*DXS1059* - 3.0 - *DXS424* - 3.0 - *DXS1212* - 2.0 - *DXS1001*

Another scenario is that neither region is the true site of the XLMR gene because the male V-6 (who is physically different from the other affected males) may *not* carry the XLMR gene segregating within his family and therefore this gene lies elsewhere on the X chromosome, although this is unlikely given the number of loci (and their relative genetic distances) investigated (Table 6.5).

### 6.9.5 Overview

From a consideration of the *MRX* localisations given in Figure 6.13, there are at least 8 distinct loci (striped in Fig. 6.14) responsible for non-specific retardation on the X chromosome if *FRAXE* mental retardation is included. The other 7 regions from pter to qter can be represented by *MRX24*, *MRX2*, *MRX10*, *MRX1*, *MRX23*, *MRX27* and *MRX3*. This degree of genetic heterogeneity strongly argues against the combination of data from *MRX* families with overlapping regional localisations for the purpose of refining regional localisations.

Whether the overlapping regional localisations of 2 or more *MRX* loci truly reflects the involvement of the same gene will only be resolved upon cloning of candidate genes and the subsequent screening for mutations in the affected individuals. It has been estimated that 20,000 genes may be involved in brain development and function (Adams *et al.*, 1991). Given that the X chromosome is 5% of the genome (Morton, 1991) and assuming an equal distribution of these loci one can expect 1,000 X-linked genes to be involved in brain function.

What constitutes a candidate gene for *MRX* is a problem raised by the lack of any known biochemical abnormalities in those affected. More extensive clinical reassessment of the affected members of the families may reveal abnormalities other than mental retardation, when compared with their normal relatives. The identification of interfamilial similarity in such secondary symptoms may be used to justify the pooling of linkage data and perhaps lead to a reduction in the gene localisation. The gene (*FMR2*) associated with *FRAXE* mental retardation (which can also be considered a non-specific form of mental retardation) has been isolated (Gecz *et al.*, 1996; Gu Y *et al.*, 1996), and the characterisation of this gene may help to predict the type of genes responsible for the *MRX* disorders.

Currently, efforts are underway to position X-linked ESTs and those that map within regions of interest then become candidates (Parrish and Nelson, 1993; Durkin *et al.*, 1994; Mazzarella and Srivastava, 1994). Expression in the brain may be a more stringent criterion to select candidate genes that map within the regional *MRX* localisations (see for example Gecz *et al.*, 1994; Colleaux *et al.*, 1996; van der Maarel *et al.*, 1996). Other candidate genes may be identified from the use of breakpoint maps constructed from the

cytogenetic and molecular genetic analyses of mentally retarded male or female patients with deletions, duplications, X/Y or X/autosome translocations (although the retarded phenotype may be a consequence of the pleiotropic effects due to the loss (or gain) of a number of genes rather than to the specific loss of a single gene).

For example, people with deletions of or within Xp22 often have contiguous gene syndromes that sometimes include mental retardation (Ballabio *et al.*, 1989; Ballabio and Andria, 1992; Schaefer *et al.*, 1993). PCR analysis of the DNA of patients with deletions in this region and suffering from mental retardation have shown that the putative locus lies between *DXS1145* (near the pseudoautosomal boundary) and the more proximal *DXS1130* (distal to *DXS237*) (Herrell *et al.*, 1995). This region is encompassed by the localisations of *MRX21*, *MRX24* and *MRX36*. There are however males with apparent deletions of this region that are not retarded (Schaefer *et al.*, 1993). More recently, STS PCR screening of MRX males has revealed a microdeletion in Xp21.3-p22.1 between the loci *DXS6764* and *DXS7188*, in one individual (Billuart *et al.*, 1996); this expanse is encompassed by *MRX2*, *10*, *13*, *21*, *29*, *32*, *33*, *36* and *38*. Characterisation of cytogenetically visible deletions in the Xq21 region using markers, and then correlating the extent of these deletions with the phenotypes of the patients from which the cells were obtained, has narrowed a locus responsible for mental retardation to a region bounded by *DXS26* and *DXS121* (Bach *et al.*, 1992; Philippe *et al.*, 1995). May and colleagues claim, however, that the locus is more distal, between *DXS233* and *CHM* (May *et al.*, 1995). The occurrence of cytogenetically visible deletions in viable males implies that relatively few genes are within these regions (such as Xp22.3 and Xq21.3), which may simplify the isolation and identification of the genes responsible for particular phenotypes.

Another example is that of a mentally retarded female possessing an X/autosome translocation [t(X;13)(q13;q34)], with the X breakpoint occurring between the loci *CCG1* and *DXS348* (van der Maarel *et al.*, 1994), a region spanned by many of the pericentromeric *MRX* loci. Her retardation may be a consequence of the disruption of a gene involved in brain function, either on the X or chromosome 13. The X-linked gene disrupted by the break is currently under investigation (van der Maarel *et al.*, 1994; 1996), and has been screened for mutations in a number of families with XLMR (Chapter 7).

It has been observed that the presence of small supernumerary ring X chromosomes, that have lost the *XIST* gene and thus escape inactivation, may increase the risk of mental retardation in those (usually females) possessing them (van Dyke *et al.*, 1991; Dennis *et al.*, 1993; Migeon *et al.*, 1994; Wolff *et al.*, 1994; Jani *et al.*, 1995). Characterisation of the genes present on these rings, which consist of centromeric heterochromatin and presumably functional pericentromeric euchromatin may lead to the identification of candidate genes for those *MRX* localisations that span the centromere. If a gene carried by ring chromosomes *is* responsible for the retardation of the carriers of these chromosomes, the phenotype is likely to be due to increased dosage (that is, two functional copies of the gene).

Ultimately, the characterisation of the loci involved in mental function will require the screening of candidate genes for mutations. With this in mind, Mandel (1994) has proposed the establishment (which has yet to be implemented) of a repository for lymphoblastoid cell lines (LCLs) from at least one affected member of each *MRX* family, so that DNA and RNA will be available to any researcher who wishes to screen a candidate sequence for mutations. Although finding a mutation in one *MRX* family need not mean that the same lesion will be found in other co-localising *MRX* families, because, as mentioned above, of the consequences of the relatively brief propagation of any particular X-linked mutation that reduces the reproductive fitness of males and because of the large number of genes on the X chromosome which are likely to be expressed in neural tissue and therefore represent candidates for *MRX*.

## 6.10 Conclusion

Five loci responsible for mental retardation have been regionally localised to the X chromosome. These investigations have enabled carrier status to be predicted by linkage using flanking markers and may prove useful in the future for prenatal diagnosis. To increase the confidence of any risk assessment, and to narrow the regions within which any candidate genes must lie, would require the genotyping of additional family members and/or (informative) markers (between those markers that define the closest recombination points known and those markers that show no recombination). The investigations of the 3 families affected with non-specific mental retardation have been published (Donnelly et al., 1994a, 1996b; appendix A.3, A.4).

The linkage analyses have also contributed to the “morbid anatomy” of the X chromosome. It is now known that there are at least 8 loci responsible for non-specific mental retardation that are distributed along most of the X chromosome, but the only gene to be discovered to date has been the *FMR2* gene, found at the *FRAXE* locus in Xq28 (Gecz et al., 1996; Gu Y et al., 1996).

## Chapter 7

Investigation of Candidate Genes

for

X-linked Mental Retardation

## TABLE OF CONTENTS

	Page
<b>7.1 Summary</b>	199
<b>7.2 Introduction</b>	200
<b>7.3 Materials and Methods</b>	204
7.3.1 Selection of clones from Xq13-q12	204
7.3.1.1 YAC/Yeast DNA preparation	204
7.3.1.2 <i>Alu</i> -PCR	204
7.3.1.3 Cosmid grids	205
7.3.1.4 Cosmid DNA preparation	205
7.3.2 Exon trapping	206
7.3.2.1 Cosmid digests	206
7.3.2.2 Cosmid-pSPL3 ligations	206
7.3.2.3 Transfection of COS-7 cells and RNA isolation	207
7.3.2.4 cDNA synthesis and PCR	207
7.3.2.5 Cloning of the trapped products	207
7.3.2.6 Colony PCR	208
7.3.3 Characterisation of trapped products	208
7.3.3.1 Plasmid sequencing	208
7.3.3.2 Sequence analysis	209
7.3.3.3 PCR product sequencing	209
7.3.3.4 cDNA libraries	209
7.3.3.5 Total RNA isolation	210
7.3.3.6 PolyA <sup>+</sup> RNA isolation	210

7.3.3.7 DNase treatment	210
7.3.3.8 cDNA synthesis	210
7.3.3.9 $\lambda$ gt11 cDNA library screening	211
7.3.3.10 <i>rTth</i> PCR	211
7.3.3.11 Northern blotting	211
7.3.4 MRX candidate genes: <i>CR73E</i> and <i>DXS6673E</i>	211
7.3.4.1 PCR	212
7.3.4.2 SSCP	214
7.3.4.3 Sequencing	214
<b>7.4 Results</b>	217
7.4.1 Selection of YAC clones that map to Xq12-q13	217
7.4.2 Selection of cosmid clones that map to Xq12-q13	217
7.4.3 Exon trapping	217
7.4.4 Sequence analysis	221
7.4.5 Characterisation of expression	227
7.4.5.1 cDNA library PCR	227
7.4.5.2 RT-PCR	230
7.4.5.3 Northern blotting	233
7.4.6 <i>CR73E</i>	236
7.4.7 <i>DXS6673E</i>	239
<b>7.5 Discussion</b>	240
7.5.1 Exon trapping	240
7.5.2 <i>CR73E</i> and <i>DXS6673E</i>	242
<b>7.6 Conclusion</b>	244

## 7.1 Summary

To begin isolating and characterising genes that may be responsible for one or more of the *MRX* loci that span the centromere of the X chromosome, YACs positive for the microsatellite marker *DXS1125* (mapping to Xq12-q13) were obtained from the Washington University School of Medicine YAC library. *Alu*-PCR was performed to generate human specific sequences from total yeast/YAC DNA. These PCR products were then used to probe an X chromosome-specific cosmid grid obtained from the Imperial Cancer Research Fund. Cosmids that mapped, by FISH, to the same region as the YACs were then sub-cloned into the exon trapping vector; pSPL3. The trapped putative exons were sub-cloned into bluescript, sequenced and used to probe Southern blots of YAC, cosmid and human genomic DNA. Those probes that hybridised to the appropriate YACs and cosmids and to human genomic DNA were investigated further. PCR amplification was attempted from brain cDNA to demonstrate that the trapped products were expressed in neural tissue. Failure to generate consistent evidence that these exon trapped products were expressed led instead to the investigation, by SSCA, of two strong candidate genes for XLMR (*CR73E* and *DXS6673E*). Neither of these genes exhibited XLMR-family specific genetic variation, a finding which would have been suggestive of the presence of a deleterious mutation.

## 7.2 Introduction

The purpose of the work presented in this chapter was to extend the gene localisation studies for XLMR to the identification and characterisation of potential candidate genes for X-linked mental retardation (XLMR).

To identify genes responsible for genetic diseases, in the absence of any cytogenetically detectable abnormality, it will be necessary to screen genes, that lie within the regions defined by linkage analyses, for mutations in the affected members of the pedigrees concerned (Mandel, 1994). Candidate genes will be prioritised on the basis of tissue expression. Genes expressed in neural tissue represent strong candidates for causing XLMR, but for this positional cloning approach to succeed a transcriptional (or functional) map of the relevant region needs to be generated. Such a map would show the relative positions of any genes, or at least ESTs (expressed sequence tags). Construction of such maps initially proceeded at a relatively slow pace primarily because many of the thousands of ESTs isolated were slow to be physically or genetically mapped (Davies, 1993). For example, of about 20,000 ESTs in GDB in early 1994, only 18 were known to be X-linked (Parrish and Nelson, 1993; Mazzarella and Srivastava, 1994). Attempts are being made to rapidly improve this situation as shown by the recent mapping of over 16,000 genes by using the PCR technique to amplify known ESTs from radiation hybrid panels and physically mapped YACs (Schuler *et al.*, 1996).

There are currently many methods to identify and isolate exons and genes from genomic DNA. These methods have been reviewed by Hochgeschwender (1992), Lovett (1994) and Brennan and Hochgeschwender (1995). Briefly, the procedures can be grouped into three categories: hybridisation-based, structure/function-based, and sequence-based techniques (Collins, 1992).

Hybridisation-based techniques involve:

- 1) the identification of evolutionary conserved sequences using zoo blots (Monaco *et al.*, 1986);
- 2) the identification of expressed sequences using Northern blots;

- 3) the detection of splice sites (Melmer and Buchwald, 1992);
- 4) the isolation of human-specific pre-processed (hn)RNA from somatic cell hybrids using *Alu*-PCR (Corbo *et al.*, 1990);
- 5) the use of genomic DNA to select cDNA clones (“direct selection”) (Lovett *et al.*, 1991; Parimoo *et al.*, 1991);
- 6) subtractive hybridisation of cDNAs from somatic cell hybrids (Jones *et al.*, 1992).

Structure-based techniques involve the isolation of genes on the basis of the presence of gene associated features such as:

- 1) unmethylated CpG islands (“island rescue”) (Lindsay and Bird, 1987; Valdes *et al.*, 1994);
- 2) splice sites (“exon trapping” or “exon amplification”) (Duyk *et al.*, 1990; Buckler *et al.*, 1991; Hamaguchi *et al.*, 1992).

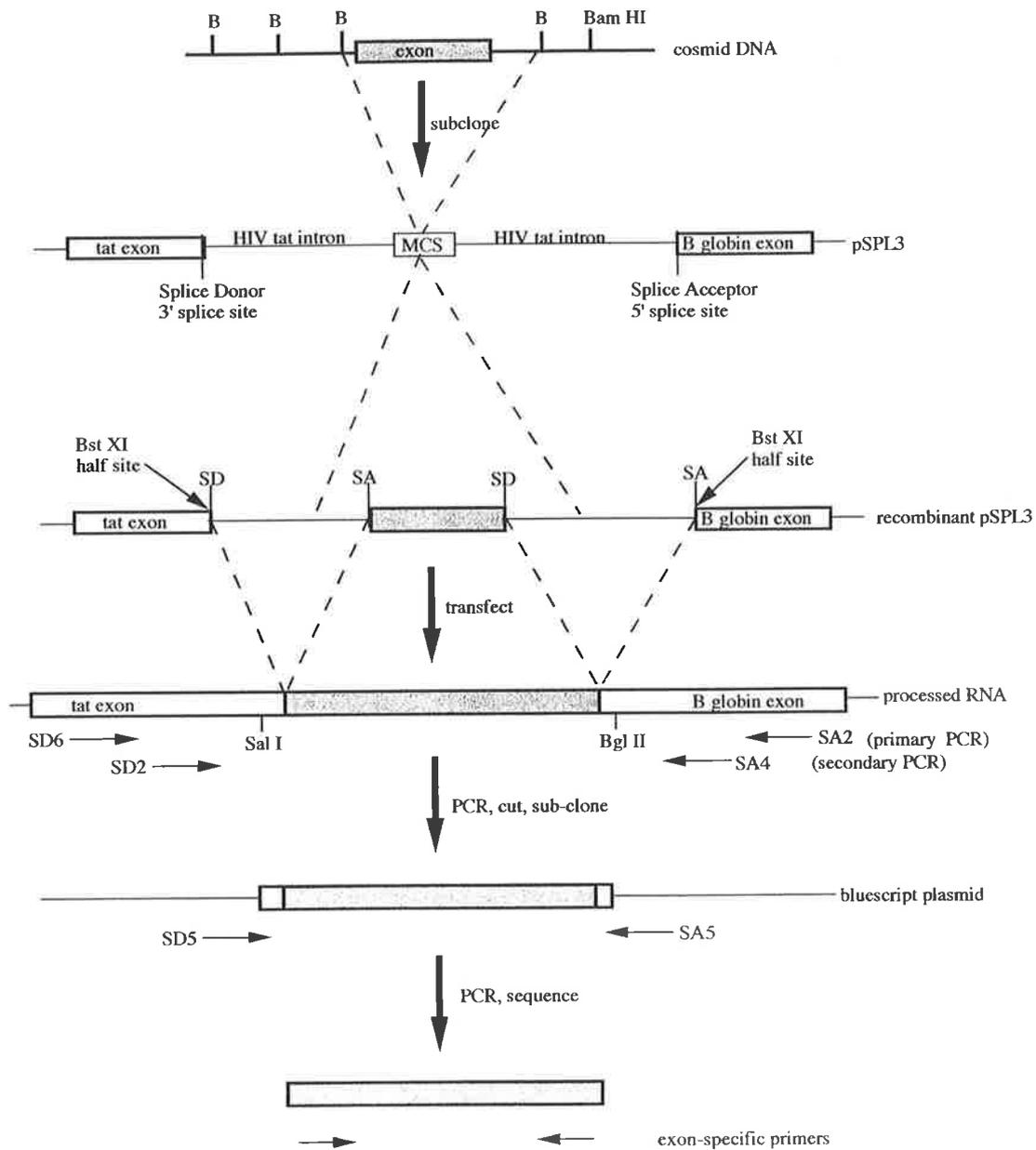
Sequence-based techniques are similar in that they rely on neural-net computer programs to identify gene-associated features, such as open reading frames, codon usage or splice site signals, from raw sequence data (Uberbacher and Mural, 1991; Hutchinson and Hayden, 1992; Snyder and Stormo, 1993). Alternatively, a different sequencing approach is the single-pass sequencing of cDNAs to generate random ESTs, followed by their physical mapping using FISH or PCR from somatic cell/radiation hybrids or YACs (Adams *et al.*, 1991; Khan *et al.*, 1991; Wilcox *et al.*, 1991; Berry *et al.*, 1995; Hillier *et al.*, 1996). This latter technique and those involving direct selection and exon trapping are the methods predominantly used to initially isolate putative expressed sequences. Each technique cited above has its advantages and disadvantages and a combined approach is most often necessary to facilitate the identification of any particular gene.

Exon trapping was the procedure used herein. This method does not rely on the expression of the genes trapped and thus is not affected by the developmental stage of expression, the tissue of expression nor the level of expression. It does rely on the presence of splice sites flanking the exon, and as a consequence terminal (5' and 3') exons are not selected. Exons from genes that have less than three exons are therefore not detected. Once a putative exon is trapped its

expression needs to be confirmed and the remaining exons of the gene need to be found. Proving expression is achieved by showing that the sequence is present in: 1) cDNA from tissue or libraries and 2) RNA (Northern blots). Isolating the whole gene may be achieved by probing the appropriate cDNA libraries with the trapped exon and/or using the RACE technique and cDNA to obtain additional sequence flanking the exon.

The exon trapping technique (Fig. 7.1) involves ligating cloned DNA into the exon trapping vector (Buckler *et al.*, 1991; Church *et al.*, 1994). This vector contains a splice donor and acceptor site, from the Human Immunodeficiency Virus (HIV), that flank the cloning site. Following transfection into eukaryotic cells, transcription proceeds from an SV40 promoter. After transcription, if an exon is within the DNA inserted into the vector, splicing occurs between the vector splice donor site and the exon splice acceptor site and *vice versa*. The result is a hybrid (exon with flanking vector sequences) mRNA. These hybrids can be selected using vector primers and the PCR products cloned into plasmids for amplification and subsequent analysis.

The two XLMR candidate genes investigated were *CR73E* and *DXS6673E*. Both were isolated and characterised by other researchers and shown to map to regions of the X chromosome that are spanned by the regional localisations of a number of XLMR genes. *CR73E* is strongly expressed in the brain and is therefore a good XLMR candidate gene. *DXS6673E* is an excellent candidate gene for XLMR because it is also strongly expressed in the brain and was found to be disrupted by a balanced X;13 translocation in a mentally retarded female (van der Maarel *et al.*, 1996). Females with cytogenetic abnormalities involving the X chromosome often survive the phenotypic consequences, whereas their male counterparts do not. Hence they are a valuable resource for the cloning of X-linked genes that are affected by the defects in chromosomal structure (Collins, 1995).



**Figure 7.1** The exon trapping procedure.

## 7.3 Materials and Methods

Only brief accounts of the methods used are given below, further details are given in the Chapter 3 sections indicated.

### 7.3.1 Selection of clones from Xq13-q12

#### 7.3.1.1 YAC/Yeast DNA preparation

YAC clones were screened with the primers of the microsatellite *DXS1125* (Chapter 4) by the staff of the Washington University School of Medicine (WUSM; Burke *et al.*, 1987). Positive clones were sent as stabs. Total yeast DNA was prepared using the method described in 3.2.7.3. The presence of microsatellite *DXS1125* in all YACs was confirmed by PCR analysis.

FISH was performed, by Liz Baker or Helen Eyre (WCH), to determine the integrity and cytogenetic localisation of each YAC (3.2.17).

#### 7.3.1.2 *Alu*-PCR

To generate human-specific products from YACs, inter-*Alu* or *Alu*-bubble PCRs were used (3.2.15.2). The latter modification was employed because of the failure to generate inter-*Alu* PCR products from the YACs: i43F5, i7F10 and i76H10. *Alu* repeats are present, on average, every 4 kb in the human genome (Moyzis *et al.*, 1989).

The inter-*Alu* primer pair used to PCR from YAC i318A12 were *Alu* 5 and 6. The *Alu* primers used on the other 3 YACs, in conjunction with the bubble primer, were *Alu* 1, 2, 3, 4, 5 and 6. Various combinations of pairs of these primers were used in conjunction with the bubble primer to determine which “set” gave the optimum results in terms of the number of PCR products and their yield *per* reaction. For example, PCR from *RsaI* restricted YAC i76H10 using *Alu* primers 4 and 6, in the same reaction with the bubble primers, resulted in more PCR products than that when primers 1 and 2 were used.

### 7.3.1.3 Cosmid grids

An X chromosome-specific cosmid grid (in duplicate) was obtained from the ICRF (Lehrach *et al.*, 1990; Nizetic *et al.*, 1991). The grid (ICRF library no. 104) contained 20,736 spots of cosmid DNA (equivalent to a 4 to 5 fold coverage of the X chromosome).

Radioactive probing was based on the protocol accompanying the grid. The grids were pre-hybridised at 42°C using hybridisation solution with denatured sheared salmon sperm DNA (at a final concentration of 100 µg/ml). 50 ng of each precipitated *Alu*-PCR was labelled using the Megaprime kit reagents in separate 50 µl reactions. After incubation for 10 min. at 37°C, the reaction was stopped by heating at 100°C for 5 min. Each probe was then pre-associated to block high copy number repeat sequences (3.2.5).

Cosmid vector (50 ng of lawrist 4) was also labelled but with [ $\alpha^{35}\text{S}$ ]dCTP rather than with [ $\alpha^{32}\text{P}$ ]dCTP (without blocking). The purpose of this was to generate a relatively faint uniform radioactive background so that the x/y co-ordinates of any  $^{32}\text{P}$ -positive spots could be readily determined.

The *Alu* and lawrist probes were pooled, added to 15 ml of fresh hybridisation solution (with salmon sperm DNA), and then added to the grids. Hybridisation occurred overnight. Washing of the filters was as previously described (3.2.6). Prior to probing the grids with the *Alu*-PCR products, they were probed with the cosmid vector alone to ensure that any positives subsequently revealed by the labelled *Alu*-PCR products were not due to inconsistencies in the amount of DNA spotted onto the grid. This lawrist 4 vector was supplied by ICRF. The autoradiographic results and the co-ordinates of 10 separate strong positive signals were sent to ICRF who sent back 10 cosmid stabs.

### 7.3.1.4 Cosmid DNA preparation

Qiagen tip 100s were used to purify the cosmid DNA from 150 ml LB/kan overnight cultures (3.2.7.6). FISH was performed by Liz Baker or Helen Eyre to confirm the integrity and

cytogenetic localisation of each cosmid (3.2.17).

Restriction fragment and repeat-sequence fingerprinting were done to determine if any of the cosmids were identical. Fingerprinting was performed by hybridising 50 ng of radiolabelled total human DNA to a Southern blot of the restriction digests.

### **7.3.2 Exon trapping**

Exon trapping was carried out using the protocol supplied by Gibco-BRL. This protocol specifically refers to the use of the vector: pSPL1 (Buckler *et al.*, 1991), however the vector used herein was pSPL3 (a gift from Dr A. Buckler). This pSPL3 vector is an improved derivative of pSPL1 (Church *et al.*, 1994). It was transformed into XL1Blue bacteria to generate sufficient DNA for cloning the cosmids.

#### **7.3.2.1 Cosmid digests**

The cosmids were cut with *Bgl*III and then *Bam*HI. Phenol/chloroform/isoamyl extraction was used to purify the restriction fragments. DNA was precipitated with 0.1 x volume of 3 M potassium acetate (pH 5.2) and 2 x volume of ethanol, followed by incubation at -70°C for 30 min.

#### **7.3.2.2 Cosmid-pSPL3 ligations**

Individual cosmid digests were then ligated into the dephosphorylated *Bam*HI site of pSPL3, and incubated overnight at 16°C. A 1:1 ratio (50 ng each) of both vector and insert was usually used in a 10 µl reaction with 1 unit of T4 ligase. Half of the ligated DNA was transformed into XL1Blue cells. An aliquot (400 µl) of the transformation mixture was plated out to assess the efficiency of ligation and transformation. Another 400 µl was inoculated into 20 ml LB/amp, grown overnight and the plasmid DNA purified using Qiagen tip 5.

*Bam*HI digests were done to linearise 500 ng of this plasmid DNA which was then run through a 0.8% (w/v), ethidium bromide containing, agarose gel to confirm that inserts had been ligated into the vector.

#### 7.3.2.3 Transfection of COS-7 cells and RNA isolation

Transfection and RNA isolation from COS-7 cells was performed according to the protocol from Gibco-BRL (3.2.11.2). Each cosmid preparation was used to transfect a separate flask of cells. In addition two control flasks were used: 1) one flask of cells was transfected with 100 ng of a positive control supplied by Gibco-BRL (this control consisted of pSPL1 with a insert containing a 66 bp exon from the *tau* gene); 2) the other flask of cells was transfected with 1 µg of pSPL3 vector only (no insert).

#### 7.3.2.4 cDNA synthesis and PCR

Following isolation of RNA from the COS-7 cells the RNA was reverse transcribed, before being assayed for the presence of any trapped products. The procedures followed for reverse transcription (3.2.19.1) and the subsequent PCR, were a combination of the protocols supplied by Gibco-BRL and Dr Buckler.

#### 7.3.2.5 Cloning of the trapped products

The cDNA PCR products were then cut simultaneously with *Sal*I and *Bgl*II, dephosphorylated (to avoid ligation of multiple inserts), extracted with phenol/chloroform, precipitated with sodium acetate/ethanol and resuspended in 20 µl of dH<sub>2</sub>O.

These PCR products (100-250 ng) were then ligated into 100 ng of *Sal*I/*Bgl*II cut bluescript plasmid vector and transformed into XL1Blue bacteria. Half of the transformation was plated onto 19 mm diameter LB/amp plates, previously covered with a solution of: 40 µl 2% (w/v) Xgal, 30 µl 100 mM IPTG and 400 µl LB, for blue/white selection of transformants.

About 15 white colonies were picked from each plate. Each plate contained the exon trapped products from one cosmid. So in total, about 75 colonies (5 cosmids x 15 colonies) were picked to be analysed. A colony was picked with a sterile pipetman tip, streaked “into the square” of a second (gridded) plate and inoculated into a PCR tube containing a drop of paraffin, for subsequent colony PCR.

#### 7.3.2.6 Colony PCR

Colony PCR (3.2.15.3), using primers SA5 and SD5, was performed and the products from each cosmid experiment were size classed, with the assumption being that products of the same size were of the same sequence. One colony from most of the different size classes was chosen for sequencing. These colonies were picked from the gridded plates, and grown overnight in 20 ml LB/amp. Plasmid DNA was prepared using Qiagen tip 5.

### 7.3.3 Characterisation of trapped products

Southern blots of *BglIII/BamHI* digests of cosmids, YACs and total human DNA were probed with putative exons to confirm that they map back to original DNA source and hybridised to restriction fragments of the expected size. The PCR products used as probes were generated from the bluescript plasmid templates using primers (SA5 and SD5) that immediately flank the insert. The products were purified from the reaction using Qiaquick spin columns and labelled using the Megaprime kit.

#### 7.3.3.1 Plasmid sequencing

Forward and reverse double strand sequencing was done using the reagents, including the M13 forward (-21) and reverse primers, and protocol of the PRISM dye-primer cycle sequencing kit (3.2.12.4). The sequence generated from the M13 -21 primer was the “sense” sequence, that is, the 5’-3’ orientation of the putative exon as it would have been in the pSPL3 vector.

### 7.3.3.2 Sequence analysis

The BLASTN program (3.2.13) was used to search for nucleotide sequence homology between the trapped products and any DNA sequence on the GenBank database. The critical p value indicating significant homology was taken to be  $\leq 10^{-5}$  (Altschul *et al.*, 1990; Church *et al.*, 1994).

Similarly, the BLASTX program (3.2.13) was used to search for any sequence homology between the predicted amino acid sequence of the trapped products and any protein sequence on the database. The critical p value indicating significant homology was again taken to be  $\leq 10^{-5}$  (Altschul *et al.*, 1990; Church *et al.*, 1994).

To reveal any open (uninterrupted) reading frames (ORFs) within the sequences the Applied Biosystems SeqEd (version 1.0.3) software was used.

The sequence of splice sites of exon trapped products were compared to the consensus splice donor and acceptor sequences reported by Fichant (1992).

### 7.3.3.3 PCR product sequencing

PCR products from genomic DNA were generated, purified and sequenced as described previously (3.2.7.5, 3.2.12.3).

### 7.3.3.4 cDNA libraries

The primers of the putative exons, and a positive control (*TR*; 7.3.3.8), were used to screen the following commercially produced human cDNA phage libraries:

Adult and fetal brain (a 1:1:1 mix of Clontech  $\lambda$ gt10, lgt11 and Stratagene  $\lambda$ ZAPII libraries);

Adult and fetal kidney (a 1:1 mix of Clontech  $\lambda$ gt11);

Lung (source unknown);

Placenta (Clontech;  $\lambda$ gt11)

### 7.3.3.5 Total RNA isolation

RNA isolation from fresh tissue was performed as described in 3.2.18.2.

### 7.3.3.6 PolyA<sup>+</sup> RNA isolation

The mRNA fraction of total RNA was purified using oligo(dT)-cellulose columns according to the protocol supplied (mRNA purification kit; Pharmacia).

The following total RNA samples were pooled to isolate the mRNA:

basal ganglia 120 µg;

cerebellum 471.5 µg;

frontal lobe 471.5 µg;

occipital lobe 187 µg

equivalent amounts of total RNA from each region were not used because the amount of RNA from basal ganglia and occipital lobe were limiting.

The yield of mRNA obtained using this method was 11.7 µg (1% recovery from total RNA).

### 7.3.3.7 DNase treatment

DNA contamination of mRNA was removed as described in 3.2.19.2.

### 7.3.3.8 cDNA synthesis

RNA was reverse transcribed as described in 3.2.19.2.

To assess the integrity of the cDNA, inter-exon primers were used from the:

Transferrin receptor (*TR*) gene (McClelland, 1984):

5' GGATAAAGCGGTTCTTGGTACCAG 3'

5' TGGAAGTAGCACGGAAGAAGTCTC 3'

The PCR product size is 512 bp (from cDNA only)

or the

Esterase D (*ESD*) gene (Lee and Lee, 1986; GenBank accession no. M13450):

5' GGAGCTTCCCCAACTCATAAATGCC 3'

5' GCATGATGTCTGATGTGGTCAGTAA 3'

The PCR product size is 453 bp (from cDNA only).

#### 7.3.3.9 $\lambda$ gt11 cDNA library screening

A *lambda* gt11 phage fetal brain cDNA (Clontech) library was amplified (Scott Whitmore, WCH) and divided into 100 sub-pools. These sub-pools were then pooled into 40 aliquots, which were stored at 4°C. These aliquots were screened by PCR using the putative exon primers, in a 25  $\mu$ l reaction.

#### 7.3.3.10 *rTth* PCR

This technique enables the reverse transcription reaction to occur at 70°C, facilitating the denaturing of any secondary structure of the RNA. The forward and reverse primers of the putative exons D2425.39 and D0814.43 were used, as described in 3.2.19.3, in an attempt to reverse transcribe a mix of total RNA from: frontal lobe, occipital lobe, brain tumour, bladder, Hela cells, and rhabdomyosarcoma.

#### 7.3.3.11 Northern blotting

Northern blotting was performed in an attempt to reveal any transcripts associated with two of the putative exons, as this would be considered proof of their expression. The procedure is described in 3.2.20, using 15  $\mu$ g of total RNA from cerebellum, frontal lobe, lymphoblastoid and rhabdomyosarcoma tissue.

#### 7.3.4 MRX candidate genes: *CR73E* and *DXS6673E*

The *CR73E* cDNA full-length 1669 bp sequence was kindly provided by Dr Philippe Berta and Philippe Jay of the Centre National De La Recherche Scientifique (CNRS), in France. They discovered that the monoexonic gene maps to Xp11.2-p11.1 (radioactive *in situ*

hybridisation) and that it is strongly expressed in fetal and adult brain and hybridises to a 4 kb transcript.

The WCH has access to the DNA from members of nine XLMR families where the disease gene spans the cytogenic localisation of *CR73E*. The families with non-syndromal mental retardation are *MRX1*, 12, 13, 17, 18 and 31 (see Fig. 6.13 and Table 6.8). The 3 syndromal mental retardation families investigated were ones affected with: Sutherland-Haan syndrome (SHS; MIM 309470; Gedeon *et al.*, 1996d); XLMR with macrocephaly (Turner *et al.*, 1994); and Wilson-Turner syndrome (WTS; MIM 309585; Gedeon *et al.*, 1996e).

The XLMR candidate gene *DXS6673E* (EMBL accession no. X95808) was isolated and characterised by van der Maarel *et al.* (1996). The gene maps to Xq13.1 and was disrupted in the 5' UTR of a mentally retarded female and thus may be a candidate for *MRX13*, *MRX31* or WTS (whose disease gene regional localisations were determined at the WCH and they encompass *DXS6673E*). The mRNA ORF is 4074 bp in length and the gene contains 25 exons. The ORF extends from within exon 2 to within exon 25.

#### 7.3.4.1 PCR

PCR was performed using the "lomix" reagents and the following thermal cycles:

[ 95°C, 30 secs; AT, 30 secs; 72°C 1 min.] x 35

where AT is the appropriate annealing temperature for the primer pair used. The annealing temperature that usually worked was calculated according to the formula:

$$4^{\circ}\text{C}/[\text{G or C}] + 2^{\circ}\text{C}/[\text{A or T}]$$

The 5 primer pairs used for the amplification of *CR73E* (Table 7.1) were designed from the cDNA sequence provided by Dr Berta and Philippe Jay (unpublished data). To avoid the possibility of any cross-hybridisation, a BLASTN search (3.2.13) was performed to ensure that the primers were not likely to amplify any other sequence.

**Table 7.1 Primers used to amplify the *CR73E* ORF for SSCA**

Primer set	Forward	Primers Reverse	Product size (bp)
1	ATCGAGTACCGGACTGGCTG	GCAGACGGCAGAGCGTATGC	281
2	TCCTTACTACTTCCTGCTGG	CTCCCGAATAAACTTGTAGG	343
3	TCATCTGCATGGCCTGGACC	CTGGCTGCATGCCATTCTG	379
4	AGGCTGCTGCCAACTGGATC	TCAGGCACTTCTTGAGGTCC	336
5	CAACCCAATTGTCTGCTTCC	ATATGCCTCTTTCCCAGTCC	319

Sequence of the gene provided by Phillippe Jay (CNRS). The primers sets were designed so that they overlapped.

Twenty-five pairs of primers were used to amplify the ORF of *DXS6673E* and were kindly provided by Dr van der Maarel of the Max-Planck Institut für Molekulare Genetik (MPIMG) in Berlin. Their sequences are given in Table 7.2. They were also subjected to a BLASTN search.

A polymorphic microsatellite within exon 1 of the *DXS6673E* gene (primers 736 and 737; van der Maarel *et al.*, 1996) was genotyped through the *MRX13*, *MRX31* and WTS families to determine if recombinants were present. Note that the *MRX13* and WTS families have been genotyped for *non-recombinant* markers flanking the *DXS6673E* gene (Kerr *et al.*, 1992 and Gedeon *et al.*, 1996e respectively)

#### 7.3.4.2 SSCA

To detect any sequence variants both MDE and 10% (w/v) acrylamide gels were used, as described in 3.2.16.4.

The DNA from one affected male from each of the families mentioned above was subjected to SSCA. The DNA from two unrelated unaffected males was used as normal controls.

#### 7.3.4.3 Sequencing

PCR products were generated and sequenced as described in 3.2.7.5 and 3.2.12.3.

**Table 7.2 Primers used to amplify the *DXS6673E* ORF for SSCA**

Exon	Primers	
	No.	Forward/Reverse
2	1263 968	GGAATTCAGGCTAAGACCACC CCAGCAACTCAGTGGCTCCA
	1264 970	TTGATACCCCTGCTGGCCTG CTCTAGTAGCCCCTCAGGTG
	1265 1266	GGCTGGGGCAAATTCCTGTT CACCTCTCCCAGCCCTG
3	1199 1232	TTCTTCCCCAGGAATGGAGG GTCCTTCCCAGAGCTAAGGG
4	1277 1233	CTGGCCAGGGTGTCAGTTTTT GGACCTTTTCCACACCCAGAAC
5	1200 920	CTGTTGGTCAGGGGATGGTTG CACAAGTCATCTTGGTGCC
	1234 1235	GGCCGGGCGCTCTGCAGT CTCATACTCCTGCCAGATC
6	1201 1236	CATTGATCCTGGAGGAGGGG GAGGGGCAGGTCAGGGCTG
7	1238 1237	CCGGGCCAACAAGGGACTG AACCAAGGAACCACACCCC
8	1203 1239	AATAAAAGTGAAGGGACAAATCC GGGAGTATGGGAGGTAGAGAG
9	1240 1241	CCCTCTCTCTTTCTCTTTTCC TTCTCCAACCCTCTTTATGTT
10	1242 1243	GCTGCACACATCTTGCCCCT GACAGTCTGGTCGTGTGGC
11	1204 1244	GAAGGGGTCCTTCCACTTGG CTGGCTGGGAGTTATAGTGTC
12	1245 1246	CTCTTACCCTTGGGTATCTGTGT GCTCCCCACTGCCCTCTTTC
13	1223 1205	GGGAAAGAAAGGGCCCAAC AGGATTAGGAGAGTCCAGGC
14	1247 1248	AAATGGAGAGCTAGTCCCTTG TGCCCCGACCTTGCCCAAAT

Primers provided by Dr van der Maarel (MPIMG).

**Table 7.2 Primers used to amplify the *DXS6673E* ORF for SSCA (cont'd).**

Exon	Primers	
	No.	Forward/Reverse
15	1249	AGGAGGGTACGGCGGTTGG
	1206	GGCAGGCCTAGGACTCTGG
16	1207	AGATGTGACCAGTCCAGCAG
	1250	TGATGAGAGTAAGGGGAAAGG
17	1251	CAACTGCTGGTGTGGTTCTTG
	1208	GGCAGAGACATGCTGGTAG
18	1209	CTGAAAGCTTGTGAGGGTGC
	1252	CCTGATGGCACAAATGCTACC
19	1253	GGAGGGCTGCAGGGTTGAG
	1278	CCTTTGGGCCTCCATTCTTC
20	1279	CTCTGGCCTTACCTCTGCCA
	1254	CTCCTCCTCCCTGTCCCTG
21	1255	ATGGAACTAAGTGAAATTCCCCAT
	1256	CCTGCTGTGGTAAGAGTGACAG
22	1257	CTTTCITCTAGGACCTACAGTC
	1258	GATGGTGGATTGGAAGAATGG
23	1214	GGTTGGGTTTGGGTAACCAC
	904	GAGAGTTGCATGTGTTCCCTC
	1259	TTTTGGGCTGCAGACAGCTG
	1215	TGGGTTATGGCAGGGCATGG
24	1260	TGTGGGCTCATGGAGCCTGTG
	1261	ATCCCTCGGGTCCCCATGC
25	1262	TGGTACTGATCTTCTGTCCTG
	918	TGGCCACAGGACAGACATTG

Primers provided by Dr van der Maarel (MPIMG).

## 7.4 Results

### 7.4.1 Selection of YAC clones that map to Xq12-q13

Primers of the microsatellite marker *DXS1125* were used to select YACs from the Washington University School of Medicine YAC library. This microsatellite was isolated by the candidate (Table 4.3) and maps near the centromere (Fig. 5.1), which is spanned by most of the *MRX* loci (Fig. 6.13). Five YACs were positive for this marker (Table 7.3). Upon receipt of these YACs, and following preparation of DNA, the presence of microsatellite *DXS1125* in each was confirmed by PCR analysis. FISH was performed to determine the integrity and cytogenetic localisation of each YAC. YAC clone i84A5 was shown to be chimaeric and was not used in any subsequent steps (Table 7.3).

### 7.4.2 Selection of cosmid clones that map to Xq12-q13

*Alu*-PCR products from these YACs were used to select cosmid clones from the ICRF gridded library. Ten of the clones that hybridised strongly to the pre-reassociated *Alu*-PCR products were chosen, however, 4 were shown by FISH to map elsewhere in the genome and one other could not be localised (Table 7.4). None of the clones were positive for the marker *DXS1125*. Restriction analysis and repeat-sequence fingerprinting indicated that, of the X-specific cosmids, J0434 and P1110 had some overlapping sequence (Fig. 7.2). The PCR product of *DXS1125* itself was not used to probe the cosmid grids because the presence of the AC repeat would be expected to severely reduce the specificity of the hybridisation.

### 7.4.3 Exon trapping

Each of the remaining five cosmid clones were cut with *Bgl*III and *Bam*HI, then separately shotgun subcloned into the *Bam*HI site of the exon trapping vector pSPL3. The plasmid DNA, containing a range of insert sizes, from the transformed bacteria was isolated and used to transfect eukaryotic COS-7 cells (1 flask of cells/ "cosmid"). The RNA from these cells was isolated and reverse transcribed.

**Table 7.3 YAC clones positive for microsatellite *DXS1125***

WUSM <sup>a</sup> clone	size (kb) <sup>b</sup>	FISH results
i318A12	260	Xq13
i43F5	230	Xq13
i7F10	175	Xq13
i76H10	285	Xq13
i84A5	400	Xq13+15pter

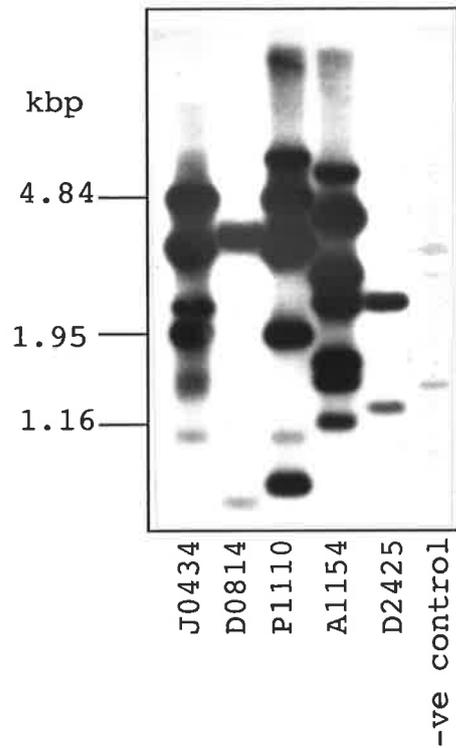
<sup>a</sup> Washington University School of Medicine.

<sup>b</sup> Information provided by WUSM.

**Table 7.4 Cosmid clones selected using YAC inter-*Alu* sequences**

ICRFc104 <sup>a</sup> clone	FISH results
A1154	Xq13
B0144	10q24
D0814	Xq12-q13
D2425	Xq12
J0434	Xq13
L034	8q24.3
M2219	22q13
N0110	7pter
P1110	Xq13
P1128	no specific signal

<sup>a</sup> Imperial Cancer Research Fund cosmid reference library #104.



**Figure 7.2** Fingerprinting of cosmids.

The autoradiograph shows *Bam*HI/*Bgl*II cosmid digests probed with total human DNA. The negative control is a cosmid that maps to Xq27. Its purpose was to reveal any cross-hybridising vector fragments.

Exon trapped products were amplified by PCR, using primers that flank the vector splice donor and vector splice acceptor sites (Fig. 7.1). The trapped products from each reaction were precipitated and cut simultaneously with *SalI* and *BglIII* to enable cloning into the *SalI* and *BglIII* sites of bluescript. Fourteen to 20 white colonies from each transformation were picked onto fresh plates and also subjected to PCR. Each set of colonies picked represent the products trapped from one of the cosmids. The inserts within the plasmids were amplified by PCR and run on an agarose gel. Based on a comparison of the sizes, the PCR products could be grouped into size classes (Table 7.5). It was assumed that PCR products of the same size were of the same sequence so only one representative from most of the size classes was chosen to be investigated further.

#### 7.4.4 Sequence analysis

Plasmid DNA from each of the representative colonies was amplified and sequenced. Results are given in Tables 7.6 and 7.7. Assuming PCR products of the same size were the same sequence, a high proportion of clones derived from cosmids J0434 and P1110 were of the same sequence. This is not surprising given the similar fingerprints of these two cosmids (Fig. 7.2).

In addition, to test the assumption that PCR products of the same size *were* of the same sequence, additional PCR products from Table 7.5 were sequenced: A1154.17, D0814.44, D0814.45, D2425.44, D2425.46, J0434.5, J0434.13 and P1110.6 all possessed the same sequence as the exon trapped product representative of their size class (marked with an asterisk).

Sequences trapped from the HIV *tat* intron within the pSPL3 vector also comprised a relatively high proportion of clones. The sequence of A1154.8 contains part of the pSPL3 sequence from nt699 to nt848 (designated herein as “HIV 1”) and joined to this is another part of the pSPL3 sequence from nt3046 to nt3094 (designated as “HIV 2a”). It appears as if the “correct” splice sites of the vector have been ignored and cryptic splice sites recognised instead as shown in Figure 7.3. As a consequence, the *BstXI* half sites were never joined and thus this artifact was not eliminated after the first round PCR (3.2.19.1).

**Table 7.5 Exon trapped products - size classing<sup>†</sup> of colony PCR results**

---

cosmid .colony (grid position)

---

A1154 .1\*,.4,.5  
           .2\*,.11,.12,.15,.16  
           .3\*,.17,.19  
           .6,.9,.10\*,.13,.18,.20  
           .8\*

D0814 .39\*, .40,.45,.48,.49,.51  
           .43\*,.44,.46,.47  
           .41,.42  
           .50\*

D2425 .3\*  
           .39\*,.42  
           .41\*,.44,.45,.48,.50,.51  
           .43\*,.46,.55,.56,.47  
           .49  
           .52,.53

J0434 .1\*,.2,.3,.5,.6,.7,.8,.9,.10,.11,.12,.13,.14,.15,.16,.17,.18,.19,.20  
           .4\*

P1110 .1\*,.3,.4,.5,.7,.9,.14,.15  
           .2\*,.6,.8,.10,.12.,13  
           .11

---

<sup>†</sup>For example, A1154.1 was the colony grown at agar grid position #1. It contains a trapped product from the experiment using restriction fragments from cosmid A1154. The trapped product was the same size as the products present in colonies grown at grid positons #4 and #5, as determined by comparing the PCR products on an agarose gel.

\* sequencing of clone attempted, see Table 7.6.

**Table 7.6 Exon trapped products - sequencing results**

---

cosmid.colony	comments
A1154.1	13 bp trapped product; not investigated further
A1154.2	does not detect band in genomic DNA (ie gDNA -ve)
A1154.3*	no homology to database sequences; ORF -ve; gDNA +ve
A1154.8	HIV 1 sequence (pSPL3 nt699-848) + HIV 2a (pSPL3 nt3046-3094)
A1154.10	HIV 3 sequence (pSPL3 nt1324-1439)
D0814.39	suspected repeat sequence; gDNA -ve
D0814.43*	no homology to database sequences; ORF +3; gDNA +ve
D0814.50	no sequence data
D2425.3*	no homology to database sequences; ORF -ve; gDNA +ve
D2425.41	HIV 3
D2425.39*	no homology to database sequences; ORF -ve; gDNA +ve
D2425.43*	no homology to database sequences; ORF +1; gDNA +ve
J0434.1	ORF +2 and +3; gDNA -ve; same sequence as P1110.2; suspected repeat
J0434.4	HIV 2b sequence (pSPL3 nt3072-3094)
P1110.1	no sequence data
P1110.2	same sequence as J0434.1; suspected repeat

---

\* Putative exon based on the presence of specific hybridisation to Southern blotted total human DNA (gDNA). Primers made.

ORF +ve = open (uninterrupted) reading frame.

HIV = human immunodeficiency virus/pSPL3 vector sequence.

**Table 7.7 Sequences (5'-3') of putative exons**

---

**A1154.3**

CAGGTCTCCAAGAGAGAAGAACTGCAGTTGAGCTGGCCTCATGAAGATATGATG  
TGGGGGCTGGAAGGTGGCCCAGCACTCAAAGCAGGTCCACCATGANCTCAG

**D0814.43**

AA CGT TCC TGC CAG CTG GGG CAG CTG CAT GTN ACC TCT CTG GGN CTG  
ACG CCA CCC ACA CCC CAG CAG ACT CAG CTN TCA GCT TTA TTC TTT GTG  
ACA GTG AGA CCT GGG ACG ATG ATA TGT GTG TGT CGA G

**D2425.3**

TGGCACAACTACAAGACAGGAACTAAGAAGAAGTAACTGAATGCCTACTCTAGGC  
CAGGTCCTTGCTAGGACTGAACATGTGTTTGTTCGAGGGTGGCTAAGTCTCAGCTG  
AGGCCGACCAGCACCACCAGACGTCATCCCGTTGTGGACTTCTACCTCCACACCA  
GATGGCTTCTTCCACCAGTcAAACCAGGGTCTCCCTGATCTNTNGNAAT

**D2425.39**

GAGTTTCAGCATAAGCTAGCAGCAGCCTGgCTTGGAGGACTGCTTGGCCTTTAGA  
CACCTACATCCGCTCTGCCCGGTCTGTGCAAGGAGCTCGGAGGCAGCAGCCATT  
TTCCCTGCATACCAGGAGAGGCAGGTGATGGCAGGAGAGGAGGGAGGCCCCAGA  
GAGG

**D2425.43**

CTA TAC GGC AAA ATT AAG GAC TGC AGC CTG GAG AAA AAA GTT CCT TCC  
CAC TGT TGC TTG ACT GAG CTG CCA TGG GCA ATG CCA GGA GGT GCA TGG  
GTG TCG GTG AAT AGG ACA GAG AGA CCT GCA GAG TGC AAT TCC AAC CCT  
GCC ACT ATT CGC TGT GTA ACC TTG G

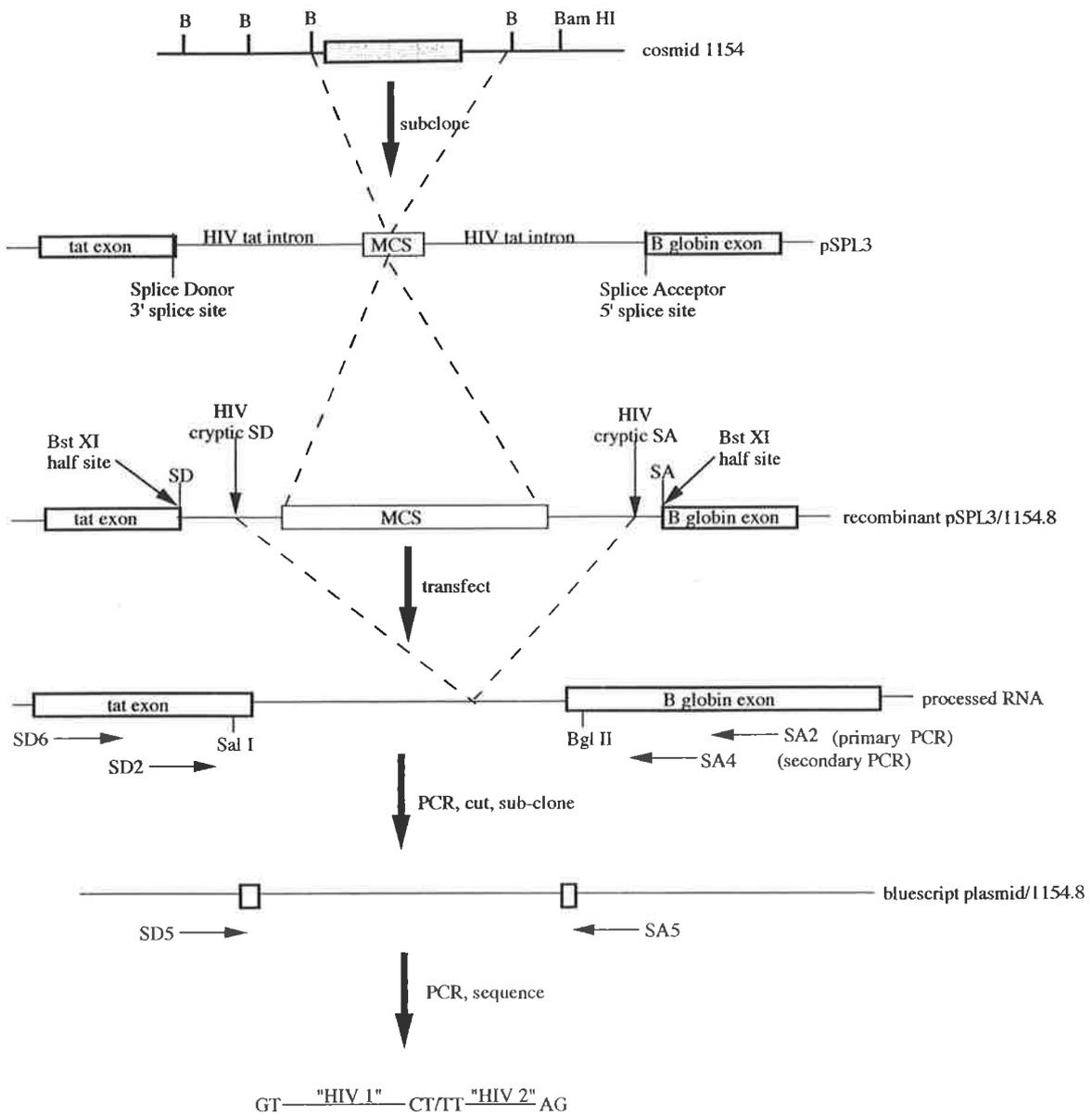
---

Primer sites are underlined. Terminal nucleotides that are conserved at splice sites are outlined.

The open reading frames of D0814.43 and D2425.43 are indicated by grouping bases into codons.

Stop codons are present in all 3 reading frames of the other 3 sequences.

Identity of bases in lower case are uncertain. N = unknown.



**Figure 7.3** Diagram showing the trapping of vector (HIV) sequences through the recognition of cryptic splice sites.

Similarly 32% of trapped products from the colonies represented by A1154.10 and D2425.41 were composed entirely of another portion of the vector sequence (pSPL3 nt1324-1439; designated as "HIV 3"). In this instance the correct vector splice donor and acceptor sites were recognised but joined to cryptic acceptor and donor sites.

The sequence of D0814.39 was 81% identical over 82 bases to a sequence within the 3'UTR of the insulin-like growth factor I (*IGF-I*) gene (GenBank accession no. X57025; nt1854-1936). Highly significant scores ( $\leq 10^{-5}$ ) were also obtained with 37 other loci, which raised the suspicion that this was a repetitive element. However, no human specific band, nor DNA smear, could be detected upon hybridisation of the PCR product from this clone to restricted genomic DNA. Also no bands unique to the YACs used to isolate the cosmids were present, (although cosmid D0814 itself did give a positive signal). The type of repeat is not known.

Similarly the identical sequences of J0434.1 and P1110.2 had very significant scores with 5 loci on the GenBank database. The greatest homology was 83% over 67 bases with the human retinoblastoma susceptibility gene (GenBank accession no. L11910), but no band in restricted human DNA was observed.

Only the sequences from 5 clones (D0814.43, A1154.3, D2425.3, D2425.39 and D2425.43) were considered to be that of putative exons. The sequences are given in Table 7.7. This assessment was based on the detection of single bands upon probing a Southern blot of restricted human, cosmid and YAC DNA with the PCR amplified insert. Open reading frames (ORFs), +3 or +1, were present in the sequences from clones D0814.43 and D2425.43 respectively. There was no ORF in the sequence from clones A1154.3, D2425.3 and D2425.39. Sequencing of the PCR products, from genomic DNA, of A1154.3, D2425.3 and D2425.39 confirmed the presence of the stop codons. These clones were, nevertheless, investigated further because they could be penultimate exons, that is, the 3'UTR may be interrupted by an intron so that the terminal 3' exon, as well as part of the penultimate exon, may not code for any amino acids.

A comparison can be made between the exon splice site sequences of the sequences given in Table 7.7 and the consensus splice site sequences. Proceeding from the last 3 nucleotides of the exon into the intron, the 5' (acceptor) splice site consensus is (A/C)AG:GT(G/A).

Proceeding from the last 14 nucleotides of the intron into the exon, the 3' (donor) splice site consensus is (T/C)10NCAG:G. The most highly conserved bases are outlined. Using these consensus sequences, the 5' splice site consensus is maintained by the putative exon D2425.39 and the 3' splice site consensus is maintained by D0814.43, A1154.3 and D2425.43 (Table 7.7). Neither the 5' nor 3' splice site consensus sequences are present in D2425.3.

Similarly a comparison with those sequences given in Table 7.8 shows that the consensus is most strongly maintained in the trapped product from colony J0434.4 (which was designated "HIV 2b" because it is part of the sequence HIV 2a mentioned in relation to A1154.8).

The majority of trapped products from cosmids J0434 and D0814 probably contained the sequence present in the clones J0434.1 and P1110.2. Despite the apparent frequency with which this sequence was trapped, the 5' and 3' "exon" splice site sequences do not conform to the consensus sequence (Table 7.8).

The 3' and especially 5' splice sites of the HIV 3 trapped product do conform well to the consensus sequence, which may explain the relatively high frequency with which this sequence was trapped (about 18% of all colonies picked and 32% of colonies from cosmids A1154.10 and D2425.41).

#### 7.4.5 Characterisation of expression

##### 7.4.5.1 cDNA library PCR

In an effort to confirm that these 5 putative exons (Table 7.7) were in fact expressed, primers were designed and attempts made to amplify these sequences from cDNA phage libraries, and from several tissues at different stages of development (Table 7.9). PCR products, of the expected size, were obtained using primers from the sequences of clones A1154.3 and D2425.39. Clone A1154.3 was of particular interest because its apparent expression in the brain made it a stronger candidate than D2425.39 for being involved in X-linked mental retardation.

**Table 7.8 Flanking sequences of other inserts showing consensus splice site nucleotides (outlined)**

cosmid.colony	flanking bases 5'-3'	sequence length (bp)
<b>A1154.1</b>	GTA.....CAT	13
<b>A1154.2</b>	AAC.....AAG	78
<b>A1154.8*</b>	GGT.....ACT/TTC.....CAG	150 + 49
<b>A1154.10/D2425.41</b>	TTT.....GAG	116
<b>D0814.39</b>	GAG.....TTG	85
<b>J0434.1/P1110.2</b>	AGA.....GTG	67
<b>J0434.4</b>	GGA.....CAG	24

\* A1154.8 has resulted from the splicing together of disparate vector (HIV) sequences.

**Table 7.9 cDNA PCR results from phage libraries**

cosmid.colony (PCR product size)	cDNA library				
	gDNA	adult+fetal brain	adult+fetal kidney	lung	placenta
A1154.3 (98 bp)	+	+	+	-	-
D0814.43 (101 bp)	+	-	-	-	-
D2425.3 (198 bp)	+	-	-	-	-
D2425.39 (151 bp)	+	-	+	-	-
D2425.43 (163 bp)	+	-	-	-	-
<i>TR</i> control (512 bp)	-	+	+	+	+

Results are only from viewing agarose gels (ie not blotted and probed).

gDNA = genomic DNA (positive control). *TR* = transferrin receptor cDNA.

+ PCR product (of the expected size) obtained.

This presence in cDNA libraries was, however, not considered sufficient evidence for expression because of the possibility of genomic DNA contamination, especially since the sequences of both A1154.3 and D2425.39 had stop codons (ie no ORFs). The absence of the other three sequences within these commercially purchased libraries does not necessarily mean that they are not exons; they could be expressed at a low level or in other tissues or at other stages of development.

PCR from a commercial phage library (7.3.3.9) failed to demonstrate the presence of the putative exons.

#### 7.4.5.2 RT-PCR

To gather additional evidence for expression, PCR amplification of RNA isolated from infant brain tissue was also performed. The results are given in Table 7.10 and show that, apart from brain stem, the RNA appeared to be contaminated with genomic DNA because of the generation of PCR products in the control reactions from which the reverse transcriptase had been omitted.

In an attempt to eliminate this genomic DNA contamination, of the total RNA samples, mRNA was isolated. The results shown in Table 7.11 show that apparently, given the absence of any visible product when reverse transcriptase was omitted from the reaction, at least three of the five exons are expressed in brain (the primers of D0814.43 seemed to be contaminated). However, a control for genomic contamination (microsatellite *DXS1125*), did show that genomic DNA was still present (unless this microsatellite was located in the UTR of a gene). In addition, a number (4) of RT-PCR products were generated using primers for D2425.3. Three of these products were also generated from genomic DNA (data not shown), which raised the suspicion that the PCR products were derived from the amplification of (contaminating) genomic DNA and not cDNA. Although, assuming DNA contamination is the explanation, why the exon-trapped sequences were not also amplified in the reactions without reverse transcriptase is not known. Similarly, given the contamination of the primers of D0814.43, why were there not products generated in all the reactions (with and without reverse transcriptase)?

Table 7.10 cDNA PCR results using RNA from infant brain tissue

	brain region										controls	
	basal ganglia		brain stem		cerebellum		frontal lobe		occipital lobe		no DNA	gDNA
	+	-	+	-	+	-	+	-	+	-		
reverse transcriptase	+	-	+	-	+	-	+	-	+	-		
A1154.3	+	+	+	-	+	(+)	+	+	+	(+)	-	+
D0814.43	+	+	+	-	+	+	+	+	+	+	(+)	+
D2425.3	+	+	+	-	-	-	+	-	+	+	-	+
D2425.39	+	+	-	-	-	+	+	+	-	(+)	-	+
D2425.43	+	+	-	-	-	-	+	(+)	+	-	-	+
<i>ESD</i> control	+	-	+	-	+	-	+	-	+	-	-	-

The brain stem RNA was mRNA prepared by Dr Jozef Gecz. The remaining samples were total RNA.

Results are from viewing agarose gels.

+ correct product size; (+) relatively low yield.

*ESD* = esterase D.

**Table 7.11 cDNA PCR results using pooled infant brain mRNA samples**

	oligo(dT) primed		random oligo primed		controls	
	reverse transcriptase +	-	+	-	no DNA	gDNA
A1154.3	-	-	+	-	-	+
D0814.43	-	-	+	-	(+)	+
D2425.3	-	-	+	-	-	+
D2425.39	-	-	-	-	-	(+)
D2425.43	(+)	-	+	-	-	+
<i>ESD</i> control	(+)	-	+	-	-	-
microsatellite ( <i>DXS1125</i> )	+	+	+	+	-	+

Results from viewing agarose gels.

+ correct product size; (+) relatively low yield.

Given that the contaminating genomic DNA had not been wholly eliminated by isolating mRNA by oligo(dT) column purification, it was decided to treat the pooled mRNA sample with DNase I. The results of the RT-PCR of this sample (and that of a cerebellum/placental RNA mix) are given in Table 7.12. Absence of the microsatellite control PCR product indicated that the DNase treatment had been successful and that the microsatellite was not transcribed). RT-PCR products were generated using the primers for D2425.39, A1154.3 and D0814.43, although there is some doubt regarding the reliability of the results of these last two because of the relatively low yield and the possibility of primer contamination respectively.

RT-PCR using the reagents of the *rTth* kit and “exon”-specific primers, rather than oligo(dT) or random priming, to generate the cDNA was also attempted. Apart from the control reaction, no RT-PCR products were generated using the primers for D2425.39 and D0814.43.

#### 7.4.5.3 Northern blotting

Conclusive evidence proving that D2425.39 and/or D0814.43 were indeed expressed could be obtained if larger products were detected upon hybridisation of these putative exons to a Northern blot. A blot was made that contained total RNA from the following tissues: cerebellum, frontal lobe, lymphoblast and rhabdomyosarcoma (muscle and connective tissue tumour). The control probe (ESD), which detects a 1.3-1.4 kb message, did hybridise to the blot (Fig. 7.4) but no signal was detected when the PCR products of D2425.39 and D0814.43 were hybridised.

**Table 7.12 cDNA PCR results using DNase treated pooled mRNA samples**

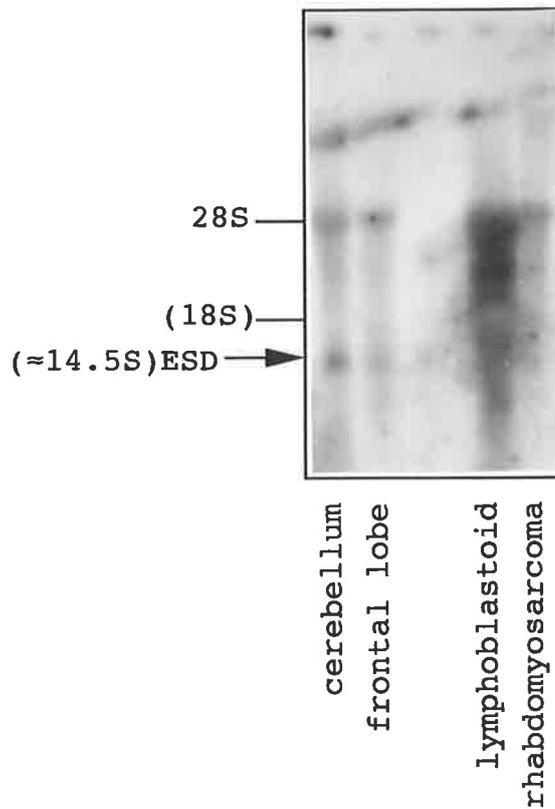
	brain mix		cerebellum & placenta		___controls___	
	reverse transcriptase +	-	+	-	no DNA	gDNA
A1154.3	(+)	-	-	-	-	+
D0814.43	+	-	-	-	(+)	+
D2425.3	-	-	-	-	-	+
D2425.39	+	-	-	-	-	+
D2425.43	-	-	-	-	-	+
<i>ESD</i> control	+	-	+	-	-	-
microsatellite ( <i>DXS1125</i> )	-	-	-	-	-	+

The brain mix consists of pooled DNase I treated infant brain mRNA from the following regions: basal ganglia, cerebellum, frontal lobe and occipital lobe. This mRNA was reverse transcribed using random oligo priming.

The cerebellum and placental mRNA mix was prepared by Dr Jozef Gecz.

Results from viewing agarose gels.

+ correct product size; (+) relatively low yield.



**Figure 7.4** Northern blot of total RNA showing hybridisation of control probe, ESD. From the photograph of the gel, the 18S rRNA band is 0.5 cm above the position to which the ESD probe has hybridised . (The hybridisation to the 28S rRNA band is presumably non-specific).

#### 7.4.6 *CR73E*

Database searching, to determine whether or not the primers designed from the *CR73E* sequence would amplify any other sequence, led to the discovery that the forward primer of primer set 3 was 95% homologous to a small (complementary) sequence of “exon x1” on chromosome 9 (GenBank accession no. M19696). This region of homology did not extend beyond the primer sequence. The reverse primer of primer set 3 was not homologous to any published sequence. The size of the PCR product obtained from these two primers was as expected from the *CR73E* sequence.

SSCA using *CR73E* primer sets 1 and 5 revealed two variants (Table 7.13). The locus amplified by primer set 1 has been designated A and the locus amplified by primer set 5 has been designated B. Each of these loci has two SSC variants: A1 and A2 and B1 and B2. Both A and B SSC variants were seen to segregate in the *MRX12* family. The B variant was also seen in the *MRX31* family. Genotyping the members of these families for these polymorphisms revealed three recombinants, II-7, III-2 and III-12, in the *MRX12* family (Fig. 7.5). The carrier III-2 is a recombinant because she must have inherited the A2 and B2 alleles from her mother (assuming there was no intragenic recombination). The male III-12 died sometime after twenty years of age and had he been mentally retarded it would have been observed earlier in development. The presence of this recombinant eliminates *CR73E* from being a candidate gene for the *MRX12* family.

Sequencing of the PCR products from members of the *MRX12* family (II-3, II-5 and III-5) and *MRX31* family (IV-3 and IV-4) showed that the A2 variant is a silent point mutation (G347A) and that the B1 variant is also a silent point mutation (C1322T) (Table 7.13). The B1 base change is the same in both families.

From the SSCA banding pattern there were no obvious recombinants between the B allele and the *MRX31* gene (data not shown), however, it was difficult to distinguish homozygotes from heterozygotes. III-10 has two sons, one affected (IV-10) and one unaffected (IV-11) (Fig. 6.7). Both of these boys, and their mother, have the B2 allele. If their mother was a heterozygote, then IV-11 would be a recombinant. Sequencing of the PCR product generated by primer set 5 of the carrier mother showed that she was a homozygote (there was no C/T double peak at nucleotide 1322). As a consequence *CR73E* has not been excluded as a candidate gene for the *MRX31* family.

**Table 7.13 CR73E sequence variants**

		SSCA band	Population frequency
locus A variants:			
minor		A2	18%
major	333 AAG CTG GTA CTG CTG GGA CTG ATT ATG	A1	82%
amino acid	25 K L V L L G L I M		
locus B variants:			
minor		B2	20%
major	1311 CTG AGG ACT CAC GCC CCC TGC GGG	B1	80%
amino acid	351 L R T H A P C W		

Sequences were determined from *MRX12* family members II-3, II-5 and III-5 (both loci) and *MRX31* family members IV-3 and IV-4 (locus B only).

The population frequencies were determined from the observation of 45 (locus A) and 40 (locus B) X chromosomes.

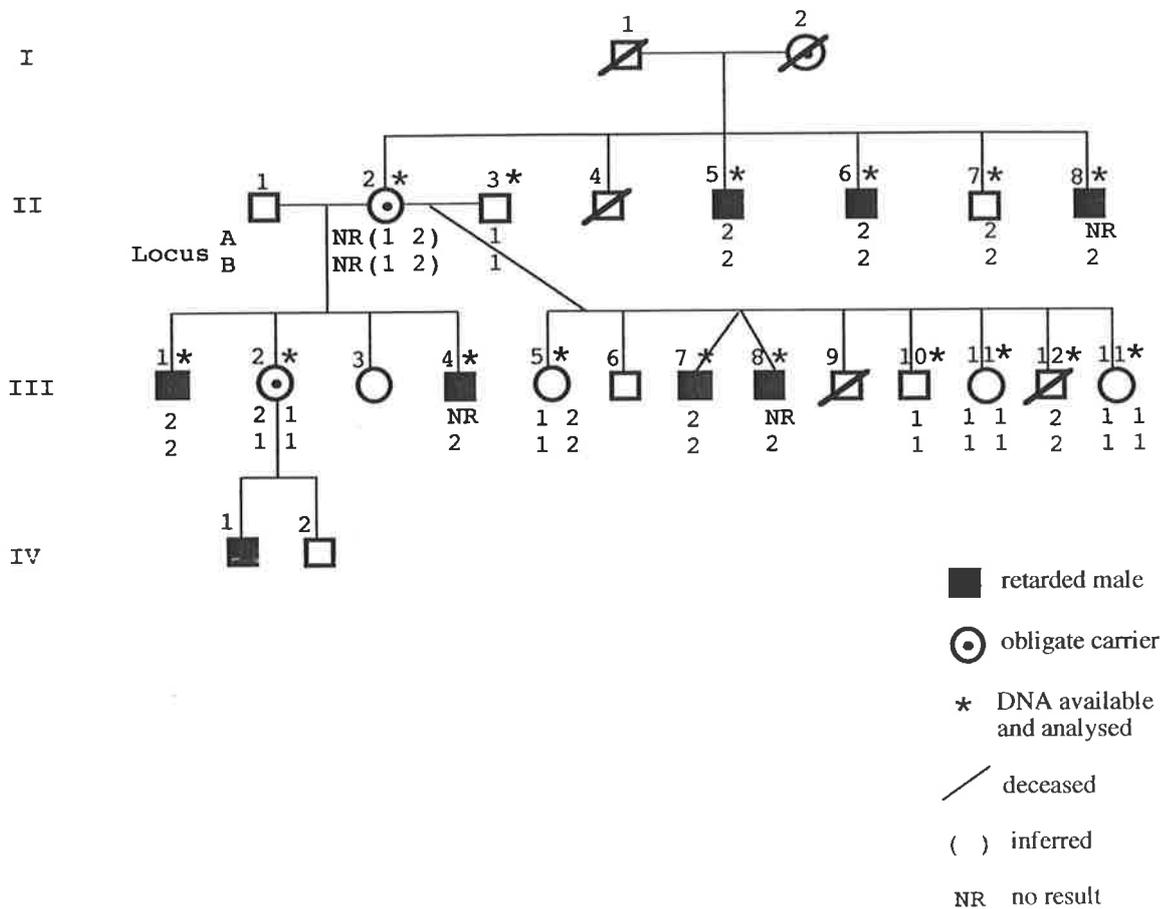


Figure 7.5 The MRX12 family with CR73E locus A and B alleles.

The population frequencies of the alleles of the A and B loci were determined by SSCA of the PCR products generated from the DNA of 40 to 45 unrelated males (Table 7.13). Both A2 and B2 occur at a frequency of about 20% in the normal population. In addition, of 10 unrelated males from the *MRX12* and *MRX31* families and the general population that possessed the rarer B2 allele, 8 also possessed the rarer A2 allele indicating linkage disequilibrium.

#### 7.4.7 *DXS6673E*

Results of using the *DXS6673E* primer sequences in a BLASTN search revealed that exon 5 reverse primer 920 (located within exon 5, that is, it is not intronic) has 95% homology to a sequence within the mouse hexokinase gene (GenBank accession no. L16949). This homology does not extend beyond the primer sequence. Exon 12 forward primer 1245 (located within an intron) has 87% homology to a locus on chromosome 6 (GenBank accession no. U53588).

A polymorphic microsatellite within exon 1 of the *DXS6673E* gene was genotyped through the *MRX31* family to determine if recombinants were present, and if so, exclude *DXS6673E* as a candidate. *DXS6673E* lies between *DXS453* (which shows no recombination with the *MRX31* gene) and *DXS1124* (which is recombinant with the *MRX31* gene) (see Table 6.3; Fig. 6.8). No recombination was observed between *DXS6673E* and the *MRX31* gene, and thus *DXS6673E* was not excluded as a candidate gene nor was the regional localisation of the *MRX31* gene narrowed. The *MRX13* and WTS families had previously been genotyped for (non-recombinant) markers flanking the *DXS6673E* gene and thus linkage analysis was unlikely to exclude *DXS6673E* as a candidate in these families. Nevertheless the microsatellite was genotyped through these families and, as expected, no recombinants were observed.

SSCA of the 24 coding exons of the *DXS6673E* gene failed to reveal any sequence variants and therefore it has not been excluded as the disease gene for the three families tested, although the probability that it is the causative gene in each of these families has been reduced, given the sensitivity of this screening method.

## 7.5 Discussion

### 7.5.1 Exon trapping

Of the 10 cosmids selected using *Alu* PCR products from YACs, 4 mapped to regions other than the expected Xq12-q13 (Table 7.4). These 4 cosmids appeared to have been true positives perhaps because they contain repeat sequences that hybridised to repeats within the inter-*Alu* sequences used to probe the cosmid grids (ie the repeat sequences within the probes were insufficiently blocked). This however does not explain why these particular cosmids, and not a great many more, were false positives. An alternative explanation is that the blocking was efficient but only for high copy number repeats, thus enabling any lower copy number repeats, with a more restricted distribution, to hybridise.

The sizes of the putative exons (Table 7.9) are consistent with sizes of most other internal exons (20-300 bp; Hawkins, 1988). Given that exons are restricted in their size range, it may have been a mistake to assume that *all* clones in a particular size class (Table 7.5) contained the same sequence, and as a consequence true exons may have gone unrecognised. However, sequencing of other PCR products from the same size class (7.4.4) indicated that they were identical.

With about 70,000 genes in a  $3 \times 10^9$  bp genome, there is expected to be 1 gene every 40 kb, on average. Previous experiments have given trapping rates of 2-3 exons *per* 40 kb of cosmid DNA (Church *et al.*, 1994; Nisson and Watkins, 1994). From fingerprinting, there were 4 distinct cosmids used in this experiment (J0434 and P1110 overlapped; Fig. 7.2) therefore 8 exons were expected. The failure to trap the expected number of exons could be due to one or more of the following reasons.

- 1) The cosmids are relatively gene or exon poor. (Genes are not evenly distributed over the X chromosome, see for example Nagaraja *et al.*, 1997).
- 2) The exons were in the incorrect (antisense) orientation within the pSPL3 vector.
- 3) The exons had *Bgl*III and/or *Bam*HI sites, and therefore they would have been disrupted during the cloning process prior to insertion into the vector. Based on the expected frequency with which these restriction sites occur in the genome (each occurs every 4096 bp on average), about one in every 20 exons of 100 bp would be disrupted. One splice site would

have remained however, so chimaeric vector/insert trapped products may have been expected. Partial *BglIII/BamHI* digests, or using other enzyme digests to cut the pSPL3 vector and cosmids, may have reduced this potential problem.

4) The exons had *BstXI* sites (same frequency as above), in which case they would have been eliminated after the first round PCR, following reverse transcription.

5) The bacterial host used can influence the success of the trapping experiment (Dr A. Buckler, personal communication), although Yaspo *et al.* (1995) have successfully used XL1-blue cells.

6) The predominance of the trapped HIV sequences, when using cosmids A1154 and D2425, may have saturated the system, reducing the PCR amplification and detection of any other trapped products.

The trapping of HIV artifacts has been reported to occur with a frequency ranging from 5 to 25% (Nisson and Watkins, 1994; Hutton *et al.*, 1995). The trapping of the HIV 1 and 2 sequences are examples of cryptic splice site recognition.

The 3' and especially 5' splice sites of the HIV 3 trapped product do conform well to the consensus sequence. Why HIV 3 was predominantly trapped (about 32% of colonies) during the experiments with the cosmids A1154 and D2425 is unknown. Perhaps any true splice sites within the restriction fragments cloned into the pSPL3 vector were recognised with far less efficiency. Alternatively, perhaps there were many non-recombinant pSPL3 vectors in which the HIV 3 splice sites did not have to compete with the sites present within any inserts. The trapping of HIV 3 in the future will not be a problem because a new generation vector (pSPL3b) has had this sequence removed (Burn *et al.*, 1995).

In general, many of the trapped products, apart from the HIV 3 sequence, appear to have been trapped through the recognition of cryptic splice sites, either the 3', 5' or both. This may reflect a lack of consensual splice sites within the cosmids used, perhaps due to a paucity of genes. Screening transformed colonies with the products trapped from the pSPL3 vector alone, prior to colony PCR, may be used to eliminate many clones from consideration.

The attempts to prove that the trapped products were expressed resulted in inconsistent results. Firstly, the results from screening the phage libraries (Table 7.9) did not concur with

those from RT-PCR of the brain mix mRNA (Table 7.11) in that PCR products appeared from the latter that were not generated from the former. This could be explained by the phage libraries being unrepresentative of the true RNA population.

Also, the amplification of microsatellite *DXS1125* from reactions without reverse transcriptase (Table 7.11) indicated that contaminating genomic DNA was present. Given this fact, it would have been expected that the putative exons would also have been amplified in the absence of any cDNA, which was not the case.

Treating the mRNA with DNase resulted in the failure to (re-)generate the PCR products of D2425.3 and D2425.43 (cf Tables 7.11 and 7.12). Does this imply that they were originally generated from contaminating DNA and if so, as stated above, why were there no products generated in the reactions without reverse transcriptase? Conversely the 2425.39 PCR product was generated after DNase treatment and not before.

The failure to detect any signal from D2425.39 or D0814.43 upon hybridisation to the Northern blot is not surprising, even if they were true exons. The labelled probes were only about 100 bp in length and in addition Northern blotting is relatively insensitive compared to RT-PCR (Ma *et al.*, 1994). Alternatively the genes may not have been expressed in the tissues used for the blot.

Given the inconsistent results and the failure to prove conclusively that any of the trapped products belonged to genes, this approach was abandoned. A second attempt to trap exons was not performed because vast number of ESTs are now being generated by single-pass sequencing and mapped. This whole genome approach represents a far more efficient method for the identification of candidate genes.

### 7.5.2 *CR73E* and *DXS6673E*

Dr P. Berta (personal communication) discovered *CR73E* to be monoexonic, expressed in the brain, and to have homology with serotonin and dopamine receptors. The predicted polypeptide has 373 amino acids and hydrophobicity analysis revealed the presence of 7 transmembrane domains. The fact that the gene is monoexonic greatly simplified mutation

screening of the XLMR families because it meant that it wasn't necessary to elucidate any intron/exon boundaries. The identification of polymorphisms within *CR73E* may allow linkage analysis to determine whether this gene is a candidate for other families and other diseases that map to Xp11.

The function of the *DXS6673E* gene remains unknown. It is highly conserved, expressed strongly in brain and encodes a 1358 amino acid polypeptide that may span membranes (van der Maarel *et al.*, 1996). It is possible that *DXS6673E* is not responsible for the translocation patient's phenotype and that instead the locus on chromosome 13 (13q31) may be involved, although the translocation does occur in the 5'UTR in exon 1 of the *DXS6673E* gene.

The SSCA technique is not 100% efficient at detecting sequence changes under one set of gel and electrophoresis conditions. The efficiency is also dependent on the DNA fragment size and possibly the DNA sequence of the fragment and the type and position of the mutation in the fragment (Hayashi and Yandell, 1993; Sheffield *et al.*, 1993). For this reason, and that neither the untranslated regions nor the regulatory regions of either of these genes were investigated, both remain candidates for those families not excluded by linkage.

Although the work described in this chapter was unsuccessful, the positional candidate approach to cloning a gene for XLMR has proved successful in the past. For example, Gibbons *et al.* (1995) describe the cloning of the gene responsible for X-linked mental retardation with  $\alpha$ -thalassaemia (ATR-X) syndrome. Linkage analysis using several families defined critical region for the disease gene. YACs from this region were used to select candidate genes in the form of cDNA clones. These clones were then hybridised to patient genomic DNA. One clone was shown to be deleted in one individual. This clone was used to isolate other cDNA clones, enabling a contig to be constructed. SSCA using these clones revealed additional mutations in other ATR-X patients.

## 7.6 Conclusion

The experiments described above revealed that splicing artifacts were a major problem, although most of these should be eliminated by using the new pSPL3 vector (pSPL3b). Not only was the trapping of putative exons difficult, but so were the attempts to prove that these putative exons were expressed. In relation to this it was essential that DNase I treatment of RNA was performed to eliminate DNA contamination.

Additional techniques that could have been employed to help prove that the trapped products were expressed include: 1) the hybridisation to zoo blots and 2) use of the RACE technique in an attempt to generate sequence extending 3' and 5' from that of the known sequence. This latter procedure would have the advantage of revealing sequences that flank exons and thus any genomic DNA contamination of RNA would no longer be of concern because the PCR product sizes could be distinguished.

Despite the lack of success of the candidate in terms of trapping true exons, the exon trapping technique *has* proved valuable in the isolation of exons from genomic clones. The cDNA selection technique may prove to be more productive for the isolation of cDNAs; once having selected a cDNA clone using a genomic target, it should be relatively easy to prove that the clone is truly an expressed sequence because of the likelihood that two or more exons will be trapped that span an intron. One of the disadvantages of this technique is, however, that the gene encoded by the genomic target may not be expressed in the tissue(s) used to construct the cDNA library. The single-pass sequencing and mapping of ESTs is, however, a more efficient approach for selecting large numbers of candidate sequences that can be investigated apropos those diseases that have been mapped to relatively large regions, and it will be the investigation of these ESTs that will form the basis for the identification of disease genes by the positional candidate approach.

With this in mind, the investigation of two XLMR candidate genes was undertaken. *CR73E* was considered a good candidate because of its expression in fetal brain. *DXS6673E* was considered an excellent candidate because it was discovered to be disrupted in a mentally retarded female by van der Maarel *et al.*, (1996). Unfortunately neither *CR73E* nor *DXS6673E* could be shown to be responsible for the disease in any of the XLMR families tested.

At this time additional patients possessing chromosomal abnormalities involving the X chromosome were examined as potential resources for positional cloning projects. The mapping of the breakpoints using reagents now available because of the progress of the Human Genome Project will ultimately lead to the cloning of additional candidate genes for XLMR (Chapter 9).

Chapter 8

X-linked Diseases

and the

Identification

of

Mutated Candidate Genes

## TABLE OF CONTENTS

	Page
<b>8.1 Summary</b>	246
<b>8.2 X-linked spastic paraplegia</b>	247
8.2.1 Introduction	247
8.2.2 Materials and Methods	249
8.2.2.1 The family	249
8.2.2.2 Linkage analysis	249
8.2.2.3 Mutation screening using SSCA	249
8.2.2.4 Sequencing of PCR products	252
8.2.2.5 Restriction enzyme digestion	252
8.2.3 Results	253
8.2.3.1 Linkage analyses	253
8.2.3.2 SSCA	253
8.2.3.3 Sequencing	258
8.2.3.4 Restriction enzyme digestion	258
<b>8.3 Myotubular myopathy</b>	262
8.3.1 Introduction	262
8.3.2 Materials and Methods	263
8.3.2.1 The family	263
8.3.2.2 Mutation screening using SSCA	265
8.3.2.3 Sequencing of PCR products	265
8.3.2.4 Introduction of restriction enzyme site and digestion	265
8.3.3 Results	266
8.3.3.1 SSCA	266
8.3.3.2 Sequencing	266
8.3.3.3 Restriction enzyme digestion	266
<b>8.4 Wiskott-Aldrich syndrome</b>	274
8.4.1 Introduction	274
8.4.2 Materials and Methods	275
8.4.2.1 The family	275
8.4.2.2 Mutation screening using SSCA	275

8.4.2.3 Sequencing of PCR products	275
8.4.2.4 Restriction enzyme digestion	275
8.4.3 Results	277
8.4.3.1 SSCA	277
8.4.3.2 Sequencing	277
8.4.3.3 Restriction enzyme digestion	277
<b>8.5 Discussion</b>	<b>284</b>
8.5.1 X-linked spastic paraplegia	284
8.5.2 Myotubular myopathy	286
8.5.3 Wiskott-Aldrich syndrome	286
<b>8.6 Conclusion</b>	<b>289</b>

## 8.1 Summary

Comprehensive genetic maps have enabled the genetic localisation of many genes responsible for X-linked disorders. Many of the genes now being “tagged”, isolated and characterised are candidates for these various X-linked diseases. Regularly, connections are found between a candidate gene and a disease which opens the way for the determination of the molecular defect in all families affected with that particular disorder.

Linkage analysis of a family having males affected with spasticity of the legs localised the defective gene to Xq21, which excluded the candidate gene *LI* in Xq28, and led to the investigation of a candidate gene (*PLP*). A point mutation, T<sup>707</sup>C, was identified within the *PLP* gene (Donnelly *et al.*, 1996a; appendix A.6). Another family had a boy affected with myopathic features. This family was too small for linkage analysis to be of any benefit, nevertheless investigation of a prime candidate gene (*MTMI*) led to the identification of a point mutation, C<sup>775</sup>T (Donnelly *et al.*, 1997; appendix A.7). This had important implications for genetic counselling. A third family had previously been diagnosed as having males affected with Wiskott-Aldrich syndrome. The recent identification of the gene responsible for this disease led to the identification of the mutation (G<sup>407</sup>A) carried by this family. This finding has not been published because the same mutation in another individual has been reported elsewhere.

The identification of molecular defects, to provide direct diagnosis for other family members, and to examine the range of mutations with a view to assessing the implications for prognosis, is now realistic for an ever increasing number of genetic disorders.

## 8.2 X-linked spastic paraplegia

### 8.2.1 Introduction

Many X-linked disorders feature some degree of spasticity but there are families in which it appears as the dominant characteristic. These X-linked forms of hereditary spastic paraplegia (SPG) have been classified as complicated (MIM 312900) or uncomplicated (or pure) (MIM 312920) (McKusick, 1992) and feature a slowly progressive spasticity of the lower limbs.

Apart from spasticity, complicated forms can exhibit both inter- and intra-familial variability in other neurological disturbances, such as; cerebellar ataxia (uncoordinated movement), sensory loss, nystagmus (jerking movements of the eyes), optic atrophy and mental retardation (Bonneau *et al.*, 1993). One complicated form has been mapped to Xq28, and it was subsequently postulated that it could be allelic to the gene responsible for MASA (mental retardation, aphasia, shuffling gait and adducted thumbs) syndrome (Kenwrick *et al.*, 1986; Winter *et al.*, 1989). This was confirmed by the identification of mutations in the *L1* gene for both SPG and MASA (Jouet *et al.*, 1994). Two other families suffering from the complicated form have however revealed linkage of the gene(s) involved to Xq13-q21.1 (Goldblatt *et al.*, 1989; Bonneau *et al.*, 1993). This localisation overlaps that determined from two other pedigrees with members suffering from uncomplicated paraplegia, which gave a localisation to Xq21-q22 (Keppen *et al.*, 1987; Le Merrer *et al.*, 1989). Yet a third family with uncomplicated paraplegia does not show linkage of the gene to this region (Kruse *et al.*, 1991). In short, many X-linked diseases have spasticity as a symptom although there are clearly two broad regions of interest: Xq28 and Xq13-22 (designated as the *SPG1* and *SPG2* loci respectively).

The clinical differences between the complicated and uncomplicated forms linked to Xq21, and within the complicated classification itself, could be explained by genetic heterogeneity, but given that they are all closely linked to *DXYS1*, allelic heterogeneity within a single gene and/or the involvement of other nearby genes (a contiguous gene syndrome) could account for the variations in phenotype.

Recently, the genetic lesion responsible for causing the complicated form of spastic paraplegia reported by Bonneau *et al.* (1993) was identified. Linkage analysis of this

family showed close linkage of the disease gene to a polymorphism in the proteolipid protein (*PLP*) gene and the researchers concluded that this was a candidate. Sequencing of the seven exons of this gene from an affected male revealed a base change in a highly conserved region, that co-segregated with all the affected individuals in the pedigree (Saugier-Veber *et al.*, 1994). Screening for this base change in 100 normal females indicated that it is not a polymorphism. The *PLP* gene had previously been identified as responsible for Pelizaeus-Merzbacher Disease (PMD; MIM 312080). PMD and complicated spastic paraplegia do share some symptoms but are otherwise clinically distinct.

## 8.2.2 Materials and Methods

### 8.2.2.1 The family

Dr Alison Colley of the Newcastle Western Suburbs Hospital in New South Wales was counselling a family, characterised by Dr Denis Crimmins, in which affected individuals were suffering from spastic paraplegia, that is, paralysis of the lower extremities. In addition, one of these affected males had nystagmus (jerking movements of the eyes). The pattern of inheritance was indicative of an X-linked gene.

### 8.2.2.2 Linkage analysis

The pedigree is shown in Figure 8.1. DNA, extracted from peripheral lymphocytes of the members indicated, was sent from the clinician (Dr Colley) in Newcastle. The PCR conditions have been described in 3.2.15.1. The references for the markers used can be found in Table 3.1.

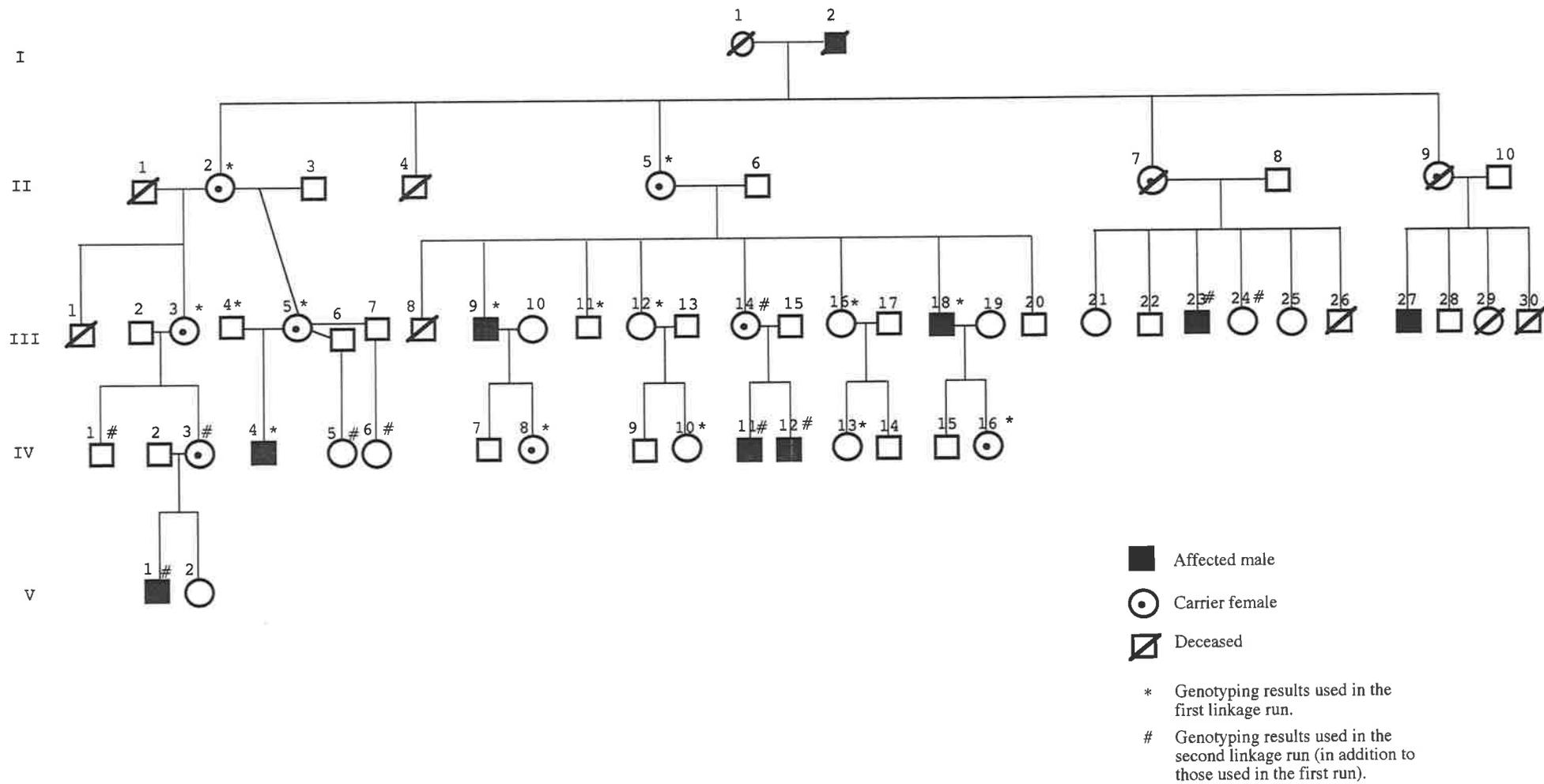
Two-point analyses were performed as described in 3.3.21.1.

### 8.2.2.3 Mutation screening using SSCA

The 7 exons (and parts of the flanking introns) of the *PLP* gene were amplified from the DNA of a carrier (II-5), an affected (III-23) and a normal (III-4) member of the family, and subjected to single-strand conformational analysis (SSCA). The PCR primers used are given in Table 8.1, and were either designed from the genomic sequence of the *PLP* gene (GenBank accession nos M15026 to M15032) or were the same as those used by Dr Pascale Saugier-Veber (personal communication) of the Service de Genetique, Hopital des Enfants Malades, France.

The thermal cycling program was option A and the "Iomix" PCR reagents were used (3.2.15.1).

The conditions for the polyacrylamide gel electrophoresis are described in 3.2.16.4. Both 5% and 10% (w/v) non-denaturing gels were used to screen all 7 exons for mutations.



**Figure 8.1** The family affected with spastic paraplegia.

**Table 8.1 Primers used to amplify the exons of the *PLP* gene**

Exon	Primers		Product size (bp)
	Forward	Reverse	
1	AGCAGGCCTGTCCCTTTAAG	CCCGAACCCCAAAGTTGGAA	284
	AGGTTTCCAGGACATCTTCA	GCTTTCTGATTGACAGCCAA	271
2	TGAGCACAGGGCCTGGCAGA	GCCAGTGAGGGCTTCATGTC	206
	CACTGGATTGTGTTTCTTTG	TGCTAGTTACTAAGGTACTG	220
3	CCGGCGCAGTCAGGCAGATC	ACGAGGCCACAGACTCGCGC	244
	ACCTCACTTATGTCGGGAAA	TTGCCGCAGATGGTGGTCTT	220
4	ACTCCAGGATCTCCCAGTTT	GCAATAGACTGGCAGGTGGT	189
	TGGTGTTTGCCTGCTCTGCT	ACCACCCTCCTTACACTAAG	216
5	AGAGATGGAAGAAGGGCTCT	ATTGAAGGCCATGGGTGTAA	236
6	CACTGGAGCATATTACTGC	AAGGATGGAAGCAGTCTACC	227
7	CCTCATTCCCAAAGGGATTGAGG	AGATGGGAGACGCAGCATTGTAGG	226

NB exons 1 to 4 have two sets of primers which generate (overlapping) PCR products. When selecting primer sites, the size of the PCR product was taken into consideration because SSCA sensitivity is dependent on fragment size (Sheffield *et al.*, 1993).

SSCA of *PLP* exon 6 PCR products from the DNA of unrelated males and the entire family was performed using a 10% (w/v) acrylamide gel.

#### 8.2.2.4 Sequencing of PCR products

Two 100  $\mu$ l PCRs *per* sample were used to generate exon 6 products from the DNA of the carrier (II-5), the affected (III-23) and the normal (III-4) members of the family. Only 100 ng of template DNA was used *per* 100  $\mu$ l reaction to minimise the amount of genomic DNA co-purified with the PCR products when using Qiaquick PCR purification tubes. Forward and reverse sequencing of about 50-100 ng of the exon 6 PCR product was done using the gene-specific primers (Table 8.1) and the reagents and protocol of the Taq Dye-Deoxy Terminator Cycle Sequencing kit (3.2.12.3).

The sequences obtained were compared to the published sequence given by Diehl *et al.* (1986).

#### 8.2.2.5 Restriction enzyme digestion

To generate a sufficient quantity of exon 6 PCR product for agarose gel analysis following *MnII* restriction enzyme digestion, 100  $\mu$ l reactions were used. These had 10 fold the components of the 10  $\mu$ l reactions used for generating radiolabelled products for SSCA, with the exception that the  $^{32}\text{P}$  was omitted and 300 ng (not 1  $\mu$ g) of genomic DNA was used. Prior to digestion, the PCR products were purified using Qiaquick PCR product purification tubes. Of the resultant DNA solution, 16.8  $\mu$ l was used in a 20  $\mu$ l reaction, along with 0.5  $\mu$ l (2.5 units) of *MnII*. After an overnight incubation, the reaction was electrophoresed through a 2% (w/v) agarose gel containing ethidium bromide.

## 8.2.3 Results

### 8.2.3.1 Linkage analyses

Given the genetic heterogeneity of diseases that feature spastic paraplegia the family (Fig. 8.1) was screened with markers along the whole of the X chromosome. Initially, two-point lod score analysis was performed using data obtained from a subset of the individuals from the pedigree (Fig. 8.1; Table 8.2). Nevertheless, linkage analysis revealed a cluster of markers in Xq12-q21 that resulted in low, but positive, lod scores (0.84 at  $\theta=0.0$  with markers *DXS1125*, *DXS1124* and *DXS986*; Table 8.2). In the absence of a suggestion of linkage to the short arm and the distal long arm of the X chromosome (Table 8.2) additional samples (# in Fig. 8.1) were collected and genotyped only for the loci between and including *AR* and *DXS424* (Table 8.3).

Statistically significant two-point lod scores were obtained with the loci *DXS1125*, *DXS983*, *DXS1124*, *DXS566*, *DXS1002* and *DXS3* in the proximal region of the long arm (Table 8.3). In addition, flanking recombinants indicate that the disease locus is located somewhere within the 47 cM separating the *AR* locus in Xq12 and *DXS424* in Xq23.

### 8.2.3.2 SSSA

SSSA was performed using the PCR products of the 7 *PLP* exons (Table 8.1) from the DNA of a carrier (II-5), an affected (III-23) and a normal (III-4) member of the family. A difference in the banding pattern of the PCR products containing *PLP* exon 6 was observed between these three individuals (Fig. 8.2). This difference was only seen using the 10% polyacrylamide gel, not the 5% gel. Figure 8.2 shows that the carrier (II-5) sample consists of PCR products that adopt the same conformation (that is, run at the same level) as those seen in samples from both the normal and affected males.

SSSA using DNA from the other members of the family showed that the exon 6 PCR products from affected males migrated differently from those of unaffected males (Fig. 8.3). However, unlike the carrier sample shown in Figure 8.2, it was not possible to confidently distinguish two bands in the obligate carrier females, because of the small degree of separation.

**Table 8.2 Two-point analysis between the gene for spastic paraplegia and marker loci\* (Xpter to Xqter)**

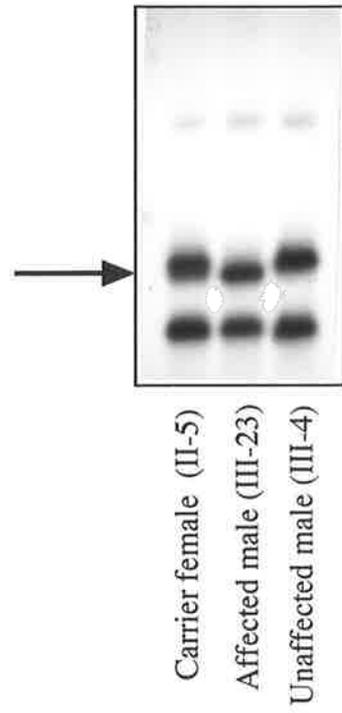
Locus	$\theta$							Z	max $\theta$	max
	0.00	0.01	0.05	0.1	0.2	0.3	0.4			
DXS237	$-\infty$	-1.18	-0.52	-0.28	-0.09	-0.02	0.01			
DXS987	$-\infty$	-0.96	-0.30	-0.06	0.10	0.12	0.08			
DXS989	$-\infty$	-2.60	-1.25	-0.71	-0.25	-0.06	0.01			
DYS (5'Dys-II)	$-\infty$	-2.64	-1.29	-0.76	-0.30	-0.10	-0.02			
DXS538	$-\infty$	-2.64	-1.29	-0.75	-0.28	-0.07	0.00			
DXS1068	$-\infty$	-2.89	-1.52	-0.96	-0.44	-0.19	-0.06			
DXS1003	$-\infty$	-1.18	-0.52	-0.28	-0.09	-0.02	0.01			
AR	$-\infty$	-1.21	-0.54	-0.28	-0.06	-0.01	0.02	0.02	0.40	
DXS1125	0.84	0.82	0.76	0.67	0.48	0.27	0.09	0.84	0.00	
DXS1124	0.84	0.82	0.76	0.67	0.48	0.27	0.09	0.84	0.00	
DXS986	0.84	0.82	0.76	0.67	0.48	0.27	0.09	0.84	0.00	
DXS424	$-\infty$	-1.21	-0.54	-0.28	-0.06	0.01	0.02	0.02	0.40	
DXS425	$-\infty$	-0.99	-0.37	-0.15	-0.02	0.01	0.01			
DXS984	$-\infty$	-1.14	-0.48	-0.23	-0.04	0.01	0.02			
DXS292	$-\infty$	-0.94	-0.32	-0.11	-0.00	0.00	-0.01			
DXS297	$-\infty$	-2.89	-1.52	-0.96	-0.44	-0.19	-0.06			
DXS1123	-0.02	-0.02	-0.01	-0.00	0.00	0.01	0.00			
GABRA3	$-\infty$	-0.96	-0.30	-0.06	0.10	0.12	0.08			

\*using the genotyping data from individuals specified (\*) in Fig. 8.1. The order of loci was taken from Fig. 5.1, except that of GABRA3 (Nelson *et al.*, 1995).

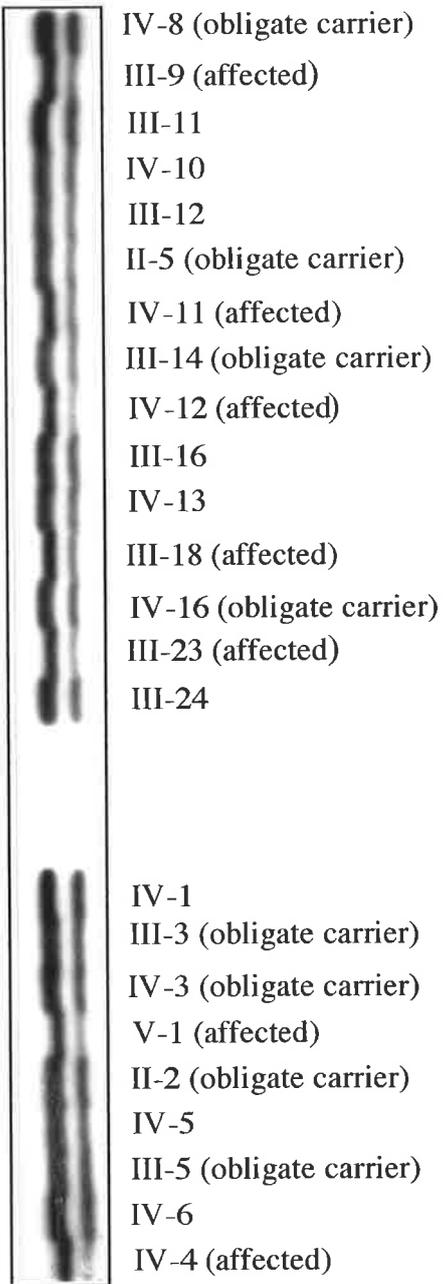
**Table 8.3 Two-point analysis between the gene for spastic paraplegia and marker loci (AR to DXS424)\***

Locus	$\theta$							Z max	$\theta_{\max}$
	0.00	0.01	0.05	0.1	0.2	0.3	0.4		
AR	$-\infty$	0.45	1.00	1.11	1.01	0.75	0.41	1.11	0.10
DXS1125	3.61	3.56	3.32	3.02	2.36	1.63	0.84	3.61	0.00
DXS983	2.83	2.79	2.63	2.42	1.94	1.40	0.76	2.83	0.00
DXS1124	2.73	2.68	2.45	2.16	1.55	0.94	0.38	2.73	0.00
DXS566	3.37	3.32	3.11	2.83	2.22	1.54	0.77	3.37	0.00
DXS986	1.67	1.64	1.52	1.36	1.03	0.67	0.29	1.67	0.00
DXS1002	3.74	3.68	3.47	3.18	2.56	1.84	1.00	3.74	0.00
DXS3	2.47	2.43	2.27	2.06	1.61	1.10	0.53	2.47	0.00
DXS424	$-\infty$	0.16	0.74	0.90	0.88	0.69	0.40	0.90	0.10

\*using the genotyping data from all the individuals specified (\*#) in Fig. 8.1. The order of loci was taken from Fig. 5.1, except that of DXS566 (Nelson *et al.*, 1995).



**Figure 8.2** Autoradiograph showing single strand conformational band shift (arrowed) of the *PLP* exon 6 PCR product.



**Figure 8.3** Autoradiograph showing the single strand conformational band shifts of the *PLP* exon 6 PCR products from members (if DNA available) of the pedigree.

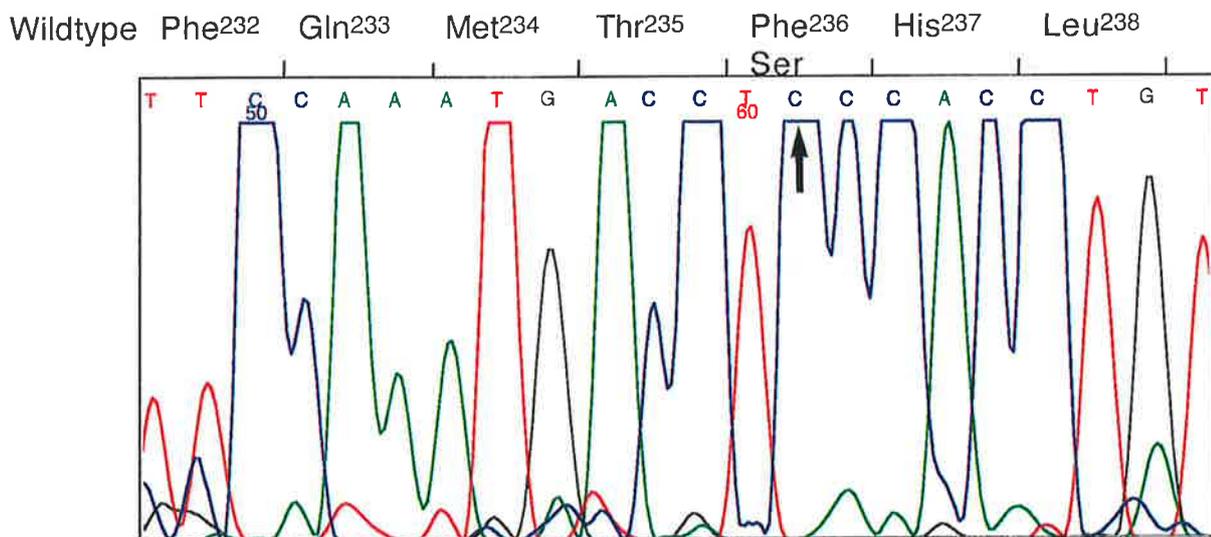
To help determine whether or not this conformational difference reflected a polymorphism (either exonic or intronic) or a disease-causing mutation, SSCA was performed using the DNA of 100 unrelated males. DNA from females was not used because of the difficulty in distinguishing the banding patterns of heterozygotes from homozygotes. None of these 100 samples adopted the same conformation as that given by the PCR products from affected males (data not shown).

#### 8.2.3.3 Sequencing

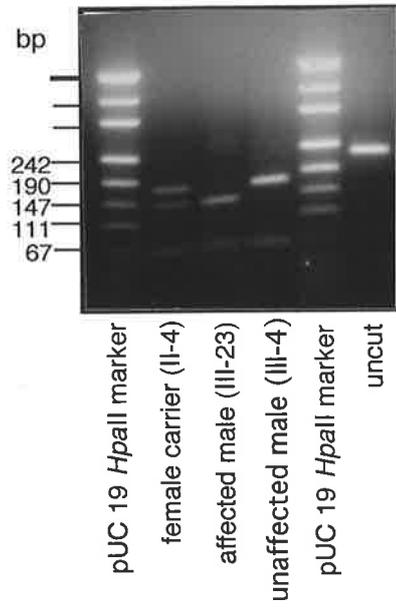
To determine the DNA sequence difference causing the conformational change, the exon 6 PCR products from the affected male (III-23) and the unaffected male (III-4) were sequenced using the intronic primers given in Table 8.1. Comparison of the sequences revealed a base transition, of T<sup>707</sup> to C (Fig. 8.4), as seen in the affected male sample relative to the unaffected male sample and to the published sequence.

#### 8.2.3.4 Restriction enzyme digestion

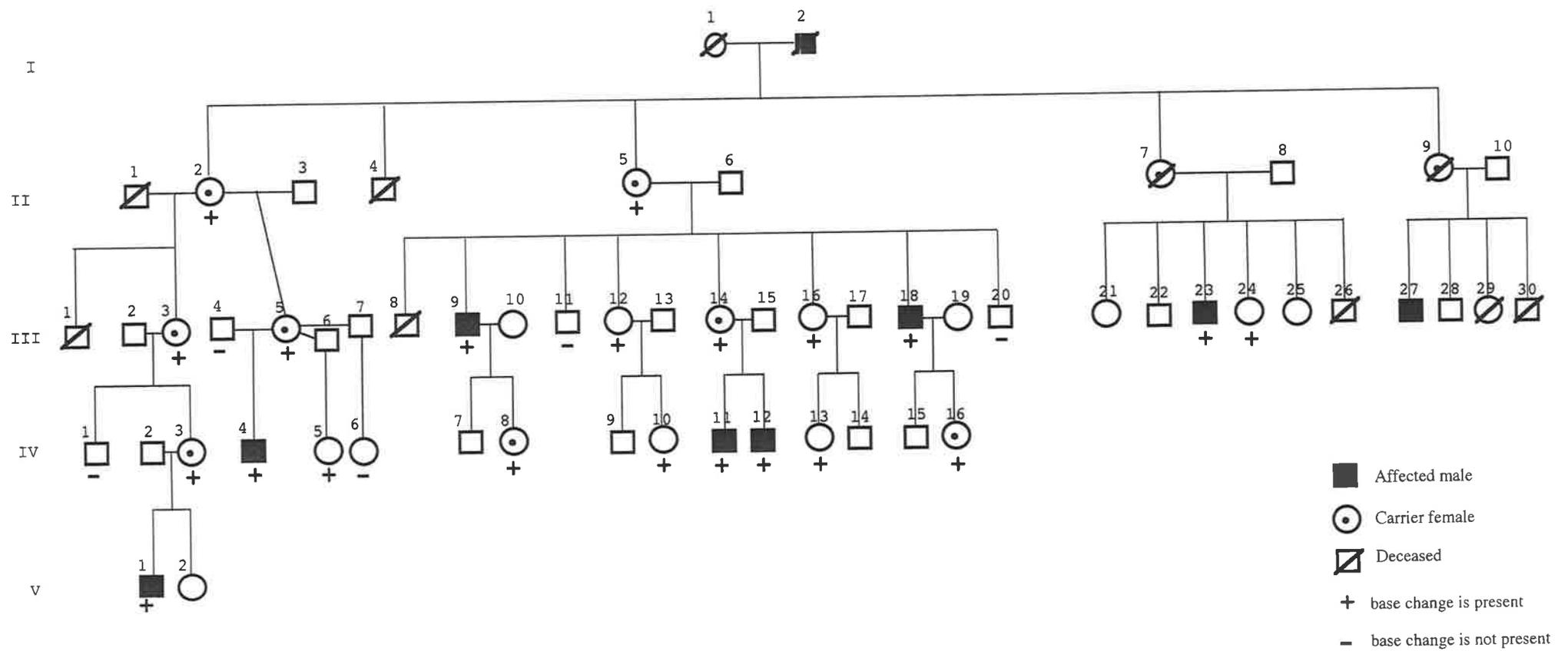
This base change creates a *MnII* restriction enzyme site (CCTC<sup>(N)7</sup>) (Fig. 8.5). From a consideration of the sequence amplified by the primers flanking exon 6, digestion with this enzyme should result in two fragments (47 and 180 bp) from the PCR products from non-carriers of the base change and three fragments (47, 30 and 150 bp) if the change is present. This restriction site change cosegregates with the obligate carrier and affected members of the family (Fig. 8.6) and was not present in the 120 X chromosomes of 60 unrelated females (data not shown). In addition, 6 other females of the family (III-12, III-16, III-24, IV-5, IV-10 and IV-13) who were not obligate carriers of the disease locus (because they have no affected sons or aren't daughters of affected males) have been shown to possess this base change (Fig. 8.6) and are therefore at risk of having affected sons.



**Figure 8.4** Part of the *PLP* exon 6 DNA sequence from the PCR product from the affected male III-23. The base change is arrowed.



**Figure 8.5** Agarose gel showing *PLP* exon 6 PCR products cut with *MnlI*.



**Figure 8.6** Results of the *MnII* digest of the *PLP* exon 6 PCR products (where DNA was available).

NB. DNA from III-20 was received after the linkage analyses reported in Tables 8.2 and 8.3.

## 8.3 Myotubular myopathy

### 8.3.1 Introduction

There are three modes of inheritance of the rare disorder of myotubular myopathy (MTM): X-linked recessive (MIM 310400); autosomal dominant (MIM 160150) and possibly autosomal recessive (MIM 255200) (Wallgren-Pettersson *et al.*, 1995). The X-linked form of the disease (XLMTM) exhibits the most severe phenotype, which includes polyhydramnios (excessive amniotic fluid), severe neonatal hypotonia and general muscle weakness often resulting in respiratory failure and perinatal death of those males affected. Carrier females have no overt clinical symptoms but may exhibit abnormal muscle histology (Wallgren-Pettersson *et al.*, 1995). Clinical suspicion of XLMTM can be confirmed by muscle biopsy which would reveal hypotrophic immature muscle fibres, with central nuclei (hence the often used alternative name of “centronuclear myopathy”), resembling fetal myotubes (Heckmatt *et al.*, 1985). The autosomal forms of myopathy have not yet been genetically mapped and until recently it was thought there may be two X-linked loci: one in Xq28 and another, reported by Samson *et al.* (1995), excluded from Xq26-Xqter. The XLMTM gene in Xq28 has recently been identified (*MTM1*; Laporte *et al.*, 1996). It has now been shown that the family described by Samson *et al.* (1995) does carry a *MTM1* mutation and the original exclusion from the Xq28 locus was due to misinterpretation of biopsy data (Guiraud-Chaumeil *et al.*, submitted, J-L Mandel, personal communication).

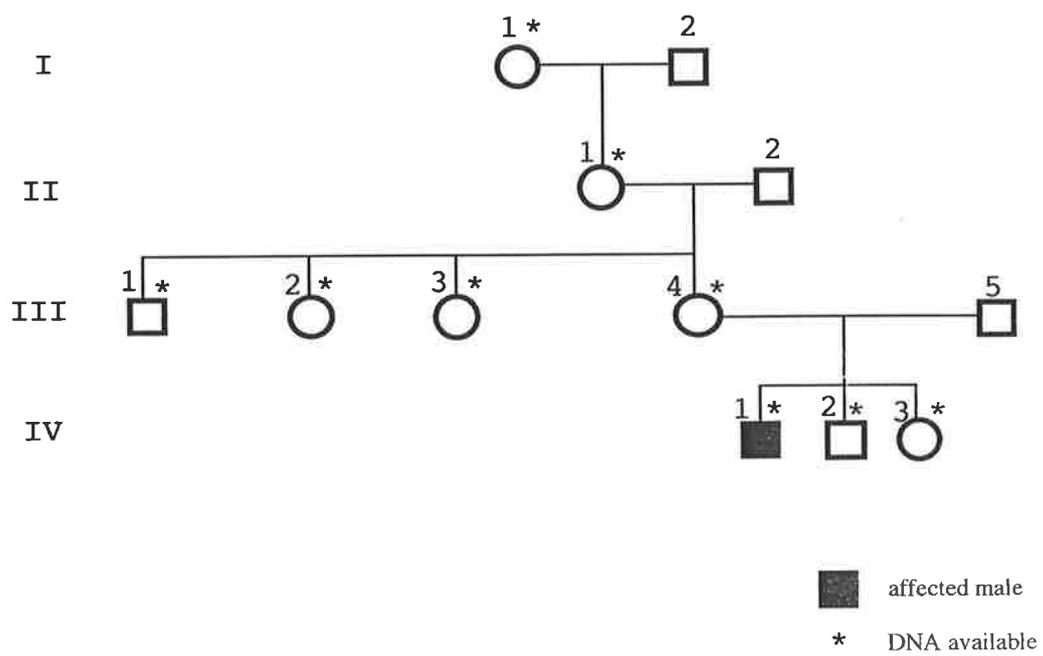
Differential diagnosis of MTM can be problematic when dealing with sporadic cases, although the autosomal cases can be distinguished from the X-linked cases by a later age of onset and milder phenotype. There is, however, the possibility of misdiagnosis of congenital myotonic dystrophy or other congenital myopathies (Wallgren-Pettersson *et al.*, 1994; 1995).

### 8.3.2 Materials and Methods

#### 8.3.2.1 The family

The family is shown in Figure 8.7. The neurological examination of the proband was performed by Dr James Manson (Department of Neurology, Women's and Children's Hospital, Adelaide). The following clinical details are from Dr Manson's report:

The affected male was born at term after a normal pregnancy, labour and delivery, with Apgar scores (assessment of the condition of the infant in relation to heart rate, muscle tone etc) of 4 at 1 minute and 6 at 5 minutes, birth weight 3665 g, length 53 cm and head circumference 35 cm. The mother had noted reduced fetal movement compared to her previous pregnancies. Poverty of movement, hypotonia and poor cry were present from birth. However, the child fed adequately and thrived. Examination at 3 months of age demonstrated persistence of generalised muscle weakness, including facial weakness with bilateral ptosis (drooping eyelids), ophthalmoplegia (paralysis of the eye muscles) involving vertical and horizontal gaze, and truncal and limb hypotonia. Muscle biopsy revealed the features of centronuclear myopathy with marked variability in fibre size, type I fibre atrophy (or hypotrophy), and increased numbers of central nuclei. When last reviewed at 3 years of age, it was noted that he had crawled between 9 and 10 months of age and had walked independently from 1 year. Generalised weakness persisted however, with ptosis, ophthalmoplegia involving vertical and horizontal gaze and a strongly positive Gowers' sign (oscillation of the iris when exposed to light). He could walk 900 m and though he could walk rapidly, he could not run. He was able to feed himself with a spoon and fork. Hypotonia persisted with significant head lag on being pulled to sit. The calves were tight. There was no evidence of progression of muscle weakness. Growth was normal and intelligence appeared to be within normal limits. In summary, the boy had a relatively mild myopathic phenotype.



**Figure 8.7** The family with an isolated case of myotubular myopathy.

### 8.3.2.2 Mutation screening using SSCA

The five published exons (and parts of the flanking introns) of the *MTMI* gene (GenBank accession no. U46024) were amplified from the DNA of the mother (III-4), the proband (IV-1) and his unaffected brother (IV-2), and subjected to single-strand conformational analysis (SSCA), using MDE and 10% gels, as described in 3.2.16.4. The PCR primers used are given by Laporte *et al.*, (1996). The PCR conditions used are given in 3.2.15.1.

### 8.3.2.3 Sequencing of PCR products

This was performed as described in 8.2.2.4.

The sequences obtained were compared to the published sequence given by Laporte *et al.* (1996).

### 8.3.2.4 Introduction of restriction enzyme site and digestion

Using the experience of Haliassos *et al.* (1989), a *CfoI* restriction site was introduced into the PCR product of the wildtype exon b allele by using the following forward primer (in conjunction with the reverse primer used by Laporte *et al.*, 1996):

5'-AGAAAATAAGACGGTCATTGcG-3' (the mismatch base is in small font size). The thermal cycling conditions were: [95°C 60 secs; 60°C 60 secs; 72°C 60 secs] x 35. The PCR products were generated and purified as described in 8.2.2.5. One µg of PCR product was incubated overnight with 10 units of *CfoI* according to the manufacturer's (Boehringer Mannheim) instructions.

### 8.3.3 Results

The family (Fig. 8.7) was too small to demonstrate linkage to any particular autosomal or X-linked locus, thus mutation screening represented the best chance of establishing the molecular defect and hence the mode of inheritance. Given the recent identification of the *MTMI* gene in Xq28 with mutations associated with myotubular myopathy (Laporte *et al.*, 1996) it was decided to screen this gene for a possible mutation. The 5 published *MTMI* exons (there are 15; J-L Mandel, personal communication) were amplified by PCR from the genomic DNA of the mother (III-4), the affected male (IV-1) and his unaffected brother (IV-2). These PCR products were then subjected to SSCA.

#### 8.3.3.1 SSCA

Variation of the exon b PCR products from these three samples was seen (Fig. 8.8), using a 10% acrylamide gel. The band shift, seen with the sample from the affected male and his mother, was not observed upon analysis of the PCR products from from other members of the family (where DNA was available) nor from the DNA of 47 unrelated males and 56 unrelated females (data not shown), a total of 159 X chromosomes.

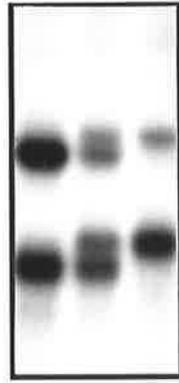
#### 8.3.3.2 Sequencing

Sequencing of the three family samples revealed a base transition, of C<sup>775</sup> to T, as seen in the DNA from the mother (III-4) and her affected son (IV-1) relative to the unaffected male (IV-2) and to the published sequence (Fig. 8.9).

#### 8.3.3.3 Restriction enzyme digestion

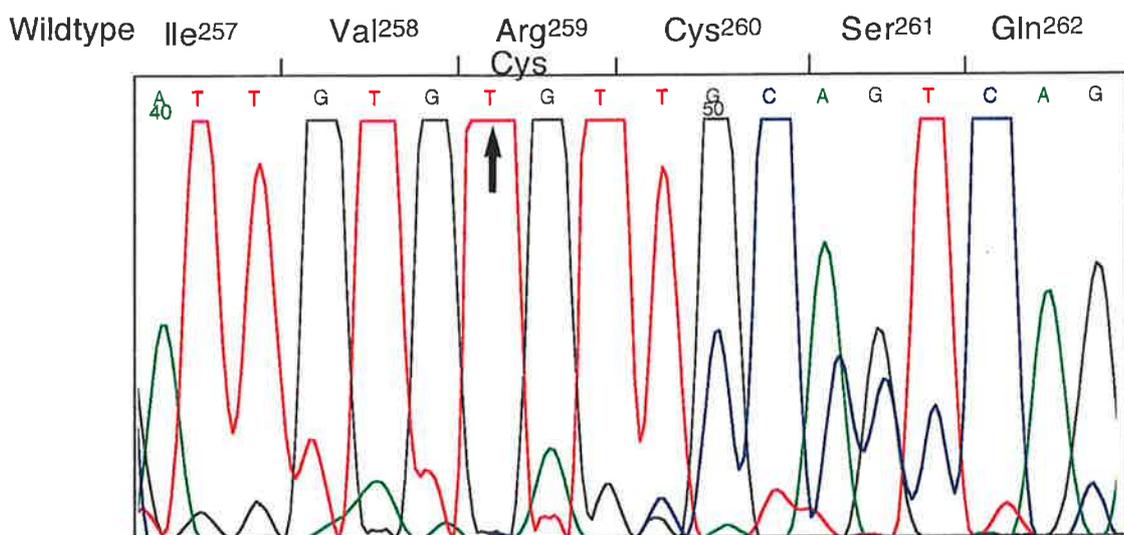
To confirm the presence of this base change, a restriction enzyme (*CfoI*) site was introduced into the PCR product of the wildtype allele (Fig. 8.10). The results of incubating the exon b PCR products from the proband, his mother and unaffected brother with this enzyme were as expected, given the sequencing results (Fig. 8.11). Given these results, DNA samples from additional members of the pedigree were analysed by this method for any base change. No other family members, from whom DNA was obtained, were shown to carry the base change (Fig. 8.12), (which was consistent with the SSCA

results). As a consequence, either the proband's mother (III-4) carries a new mutation (acquired early after fertilisation or from a lone mutated gamete) or one of her parents is a gonadal mosaic (the parental allele mutated could be determined by haplotype analysis). The father (II-2) of the carrier (III-4) was not available for study but the possibility that he carries the C<sup>775</sup> to T base transition as a neutral variant can be excluded by its absence in his other two daughters (III-2 and III-3), assuming he is the biological father of all three women.

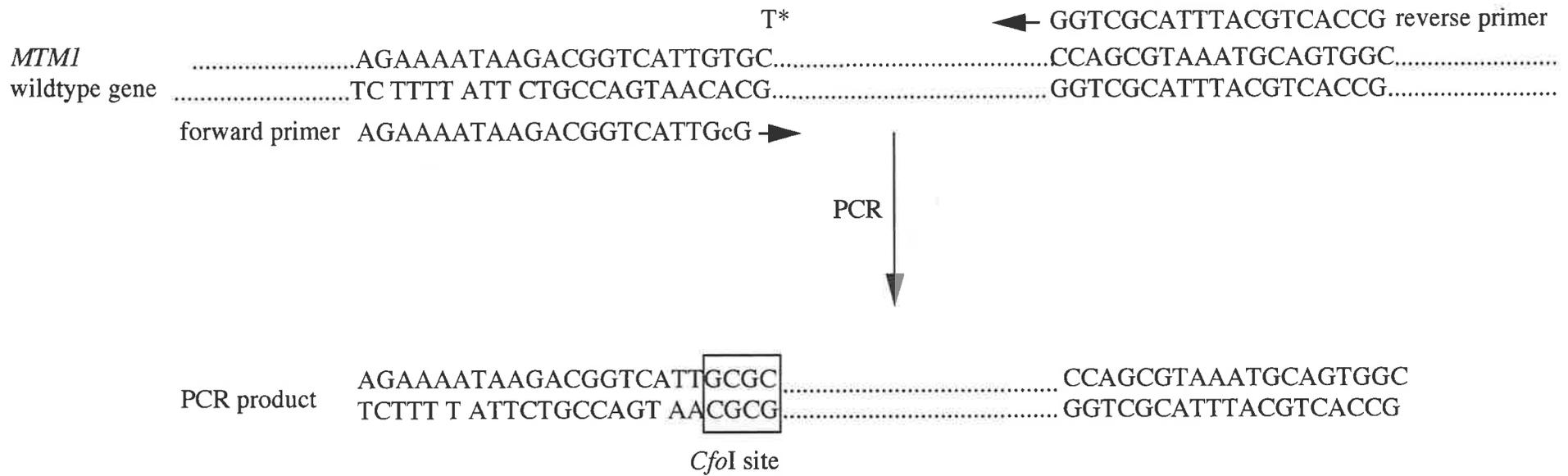


unaffected brother (IV-2)  
carrier mother (III-4)  
proband (IV-1)

**Figure 8.8** Autoradiograph showing single strand conformational band shift of the *MTM1* exon b PCR product.



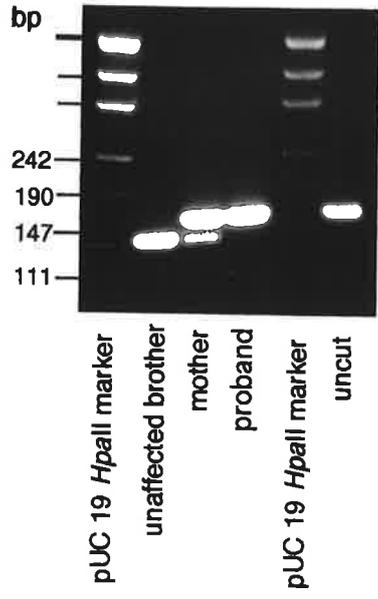
**Figure 8.9** Part of the sequence of *MTM1* exon b from the affected male (IV-1) showing the mutated base.

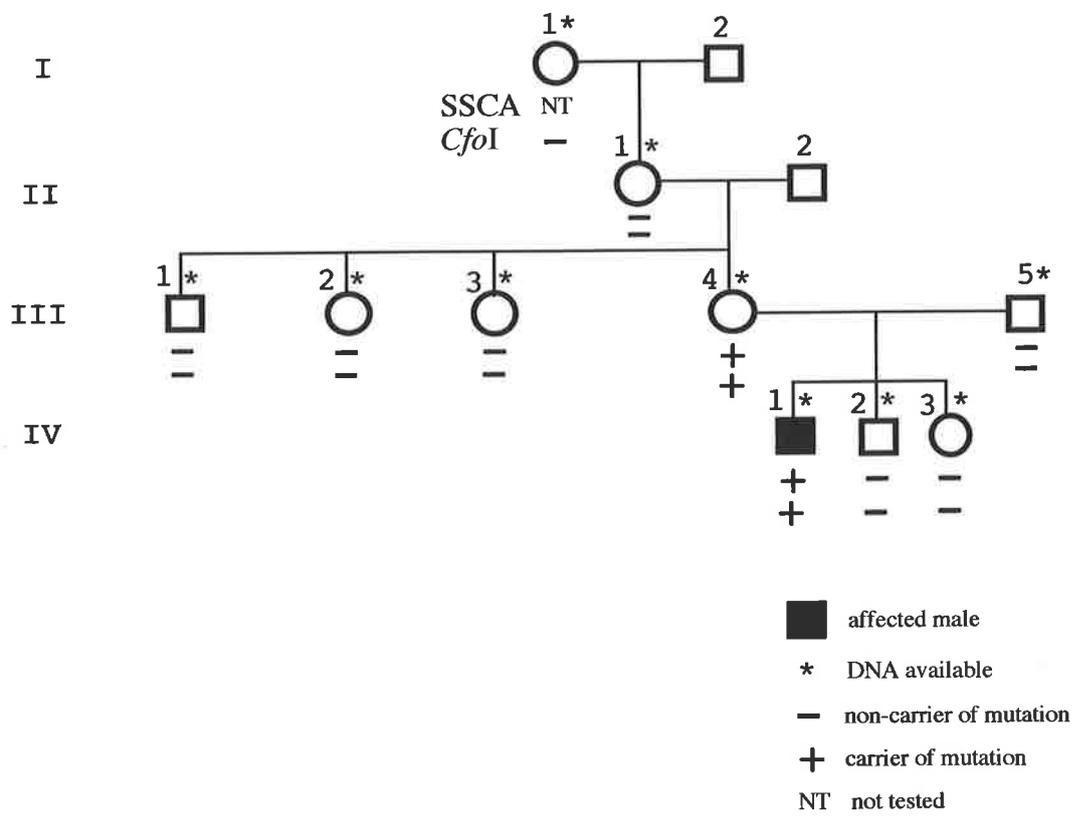


**Figure 8.10** Procedure used to introduce a *CfoI* restriction site into the wildtype *MTM1* exon b PCR products.

\* base substitution

**Figure 8.11** *MTM1* exon b PCR products cut with *Cfo*I.





**Figure 8.12** Results of the SSCA and *CfoI* digestion of *MTM1* exon b PCR products.

## 8.4 Wiskott-Aldrich syndrome

### 8.4.1 Introduction

Wiskott-Aldrich syndrome (MIM 301000) is an X-linked recessive disease characterised by eczema (skin inflammation), thrombocytopenia (abnormal platelets and lymphocytes) and immunodeficiency (Aldrich *et al.*, 1954; Rosen *et al.*, 1995). It occurs with a frequency of 4 *per* million male live births (Perry *et al.*, 1980). Female carriers rarely exhibit symptoms due to non-random X inactivation in haematopoietic cells (Wengler *et al.*, 1995). There have, however, been reports of affected females which has led various authors to suggest that a defective autosomal locus may also lead to the WAS phenotype (Conley *et al.*; 1992; Kondoh *et al.*, 1995; Rocca *et al.*, 1996). The X-linked form exhibits wide clinical variability, with symptoms ranging from mild to severe. Severely affected males die within the first decade of life due to infection, haemorrhaging or lymphoreticular malignancy, unless treated successfully with bone marrow transplantation (Perry *et al.*, 1980; Sullivan *et al.*, 1994). The relatively mild disease: X-linked thrombocytopenia (XLT; MIM 313900) has recently been shown to be allelic to WAS (Derry *et al.*, 1995a; Kolluri *et al.*, 1995). The *WAS* gene was mapped to Xp11 (Kwan *et al.*, 1991) and positionally cloned by Derry *et al.* (1994).

## 8.4.2 Materials and Methods

### 8.4.2.1 The family

The affected male (IV-1) shown in Figure 8.13 was clinically diagnosed as suffering from Wiskott-Aldrich syndrome by Dr Chris Pearson (Flinders Medical Centre, Adelaide). At 8 weeks of age he had a runny nose and oral thrush, with discharge from the ears due to otitis externa (inflammation of the outer ear passage). By 3 to 4 months eczema had appeared with some areas infected. He also presented with thrombocytopenia (low numbers of platelets), small platelets and minor (but not diagnostic) abnormalities of immune function tests.

The boy's uncle (IV-2) died at 3 to 4 months of age. He had eczema, ear infections, osteomyelitis (inflammation of the interior of the bone) and failed to thrive. The other uncle (IV-3) died at 12 months. He had similar symptoms as his brother but also: anaemia (requiring transfusion), pneumonia and asthma.

### 8.4.2.2 Mutation screening using SSCA

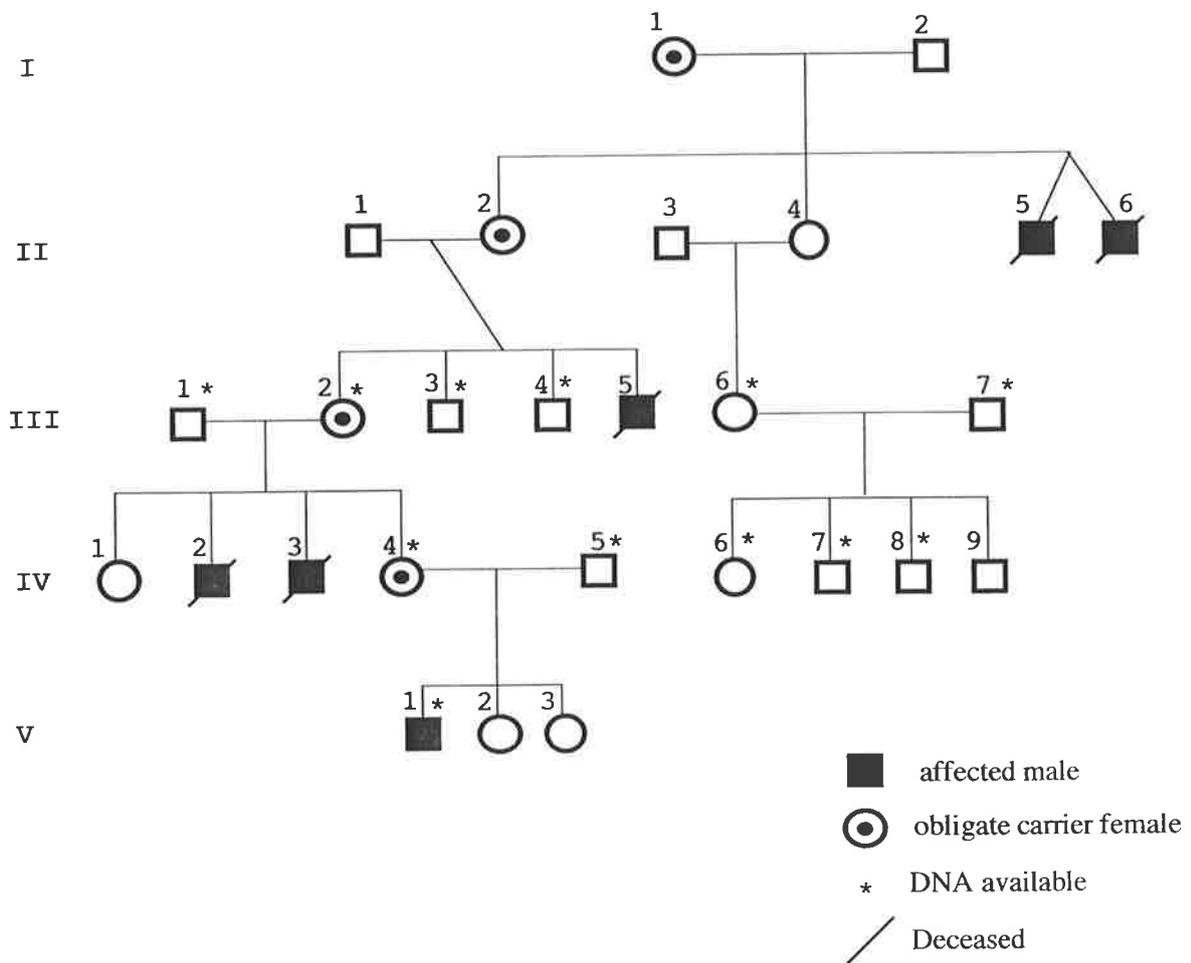
The primers used for amplification of the 12 exons of the WAS gene (GenBank accession no. U19927) are given by Kwan *et al.* (1995). The thermal cycling conditions were: [95°C 30 secs; 58°C 30 secs; 72°C 60 secs] x 35. SSCA was performed as described previously (3.2.16.4) using both MDE and 10% (w/v) acrylamide non-denaturing sequencing gels.

### 8.4.2.3 Sequencing of PCR products

PCR products were generated, purified and sequenced as described in 8.2.2.4. Sequences were compared to the WAS sequence published by Kwan *et al.* (1995).

### 8.4.2.4 Restriction enzyme digestion

Restriction enzyme digestion of 2 µg of PCR product (generated and purified as described in 8.2.2.5) was performed in a volume of 30 µl, according to the manufacturer's instructions. The restriction fragments were separated on a 2% (w/v) agarose gel.



**Figure 8.13** The family affected with Wiskott-Aldrich syndrome.

### 8.4.3 Results

#### 8.4.3.1 SSCA

The *WAS* gene was screened for mutations by amplifying its 12 exons by PCR from the genomic DNA of the affected male (IV-1), his mother (III-2) and his unaffected father (III-1). These PCR products were then subjected to SSCA.

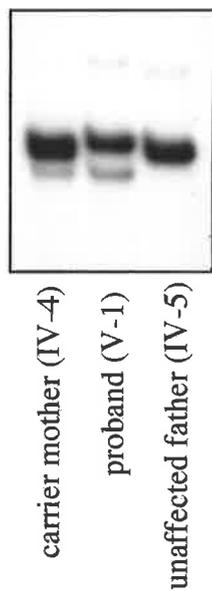
A conformational difference was detected with the exon 4 PCR products from the affected male and his mother, compared to that of the unaffected father (Fig. 8.14), but only using an MDE gel and not a 10% gel. Further analysis revealed that the allele carried by the proband was also carried by his grandmother (as expected from her obligate carrier status), but was not seen upon analysis of the DNA from other members of the family (where DNA was available nor from 99 X chromosomes from 35 unrelated males and 32 unrelated females (data not shown).

#### 8.4.3.2 Sequencing

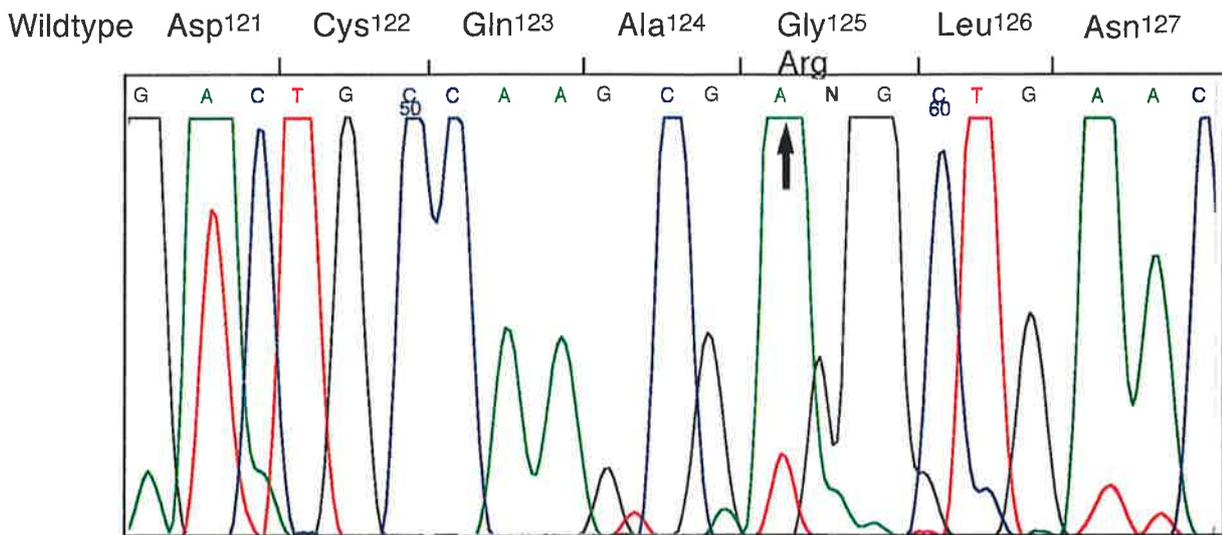
Sequencing of the exon 4 PCR products from the proband and his parents revealed a base transition, G<sup>407</sup> to A, in the affected male relative to the sequence from his unaffected father and to the published sequence (Fig. 8.15). The raw sequence data of the mother's DNA showed a double (G/A) peak at this position, as expected.

#### 8.4.3.3 Restriction enzyme digestion.

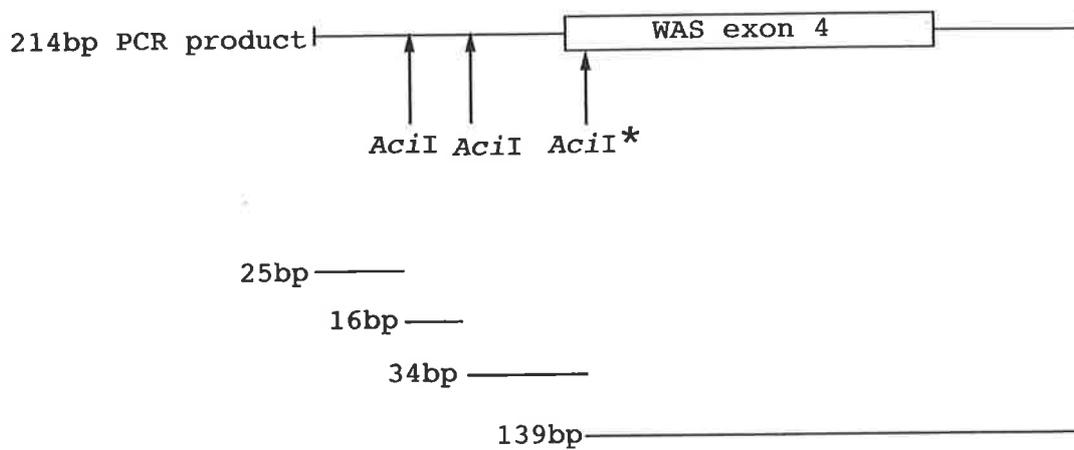
This base change should eliminate an *AciI* restriction enzyme site (5'G<sup>407</sup>CGG3'; Fig. 8.16). This was confirmed by incubating the exon 4 PCR products of the proband (IV-1), his mother (III-2) and his father (III-1) with this enzyme (Fig. 8.17). These results are consistent with the SSCA data (Fig. 8.18). This mutation is predicted to result in a Gly to Arg substitution at codon 125.



**Figure 8.14** Autoradiograph showing single strand conformational band shift of the *WAS* exon 4 PCR product.



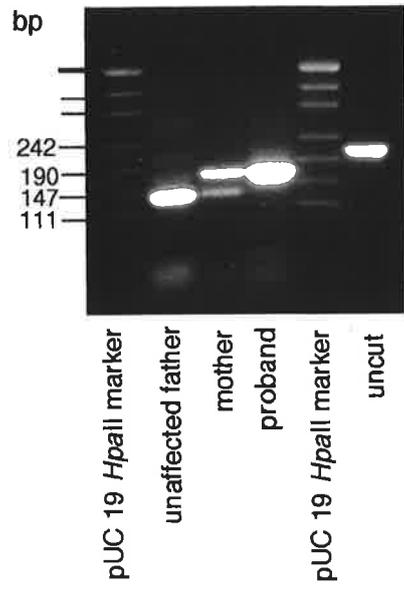
**Figure 8.15** Part of the sequence of *WAS* exon 4 from the affected male V-1, showing the mutated base.

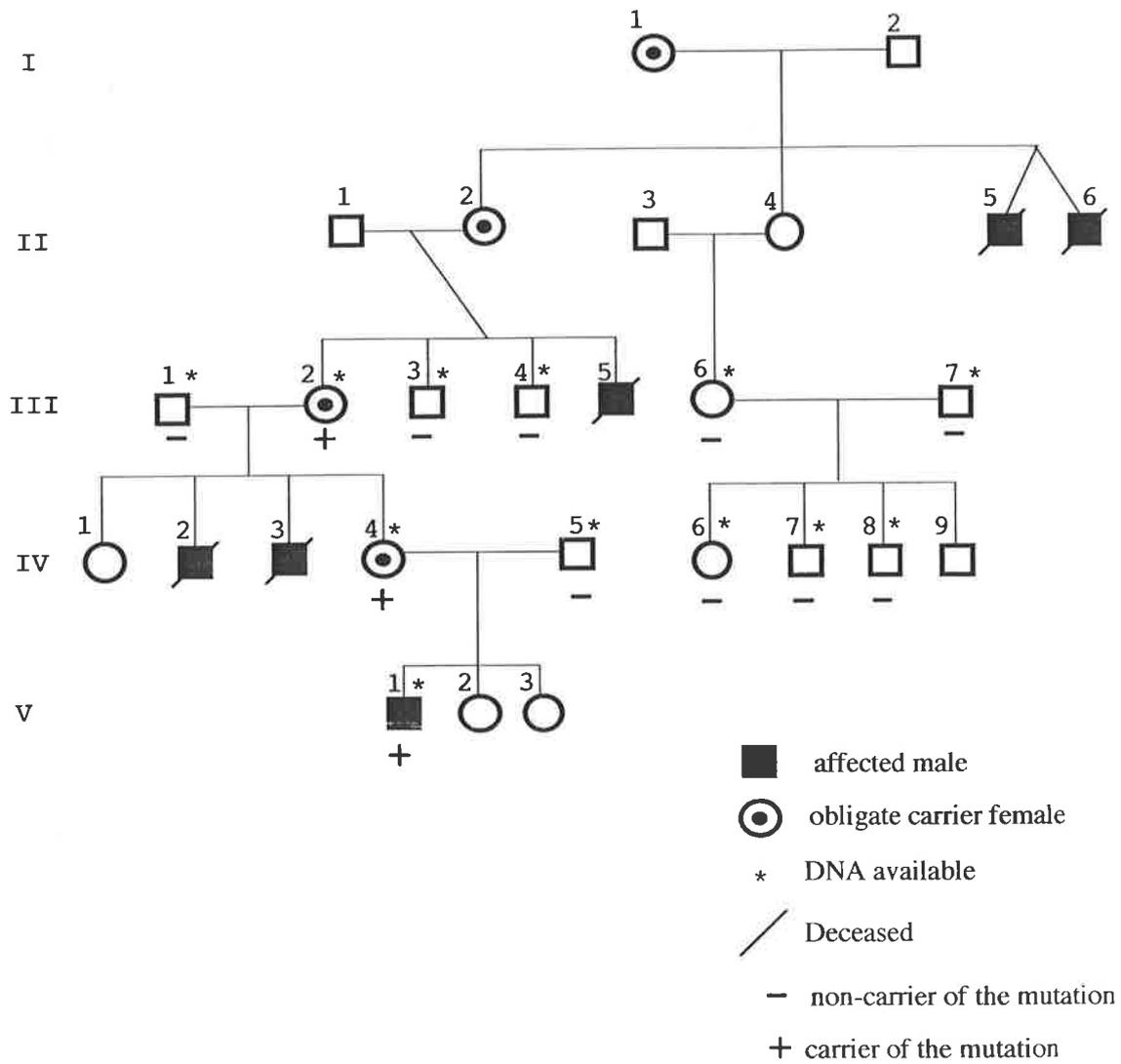


**Figure 8.16** Expected restriction fragments from the PCR product of a wildtype *WAS* exon 4 allele, following digestion with *AcilI*. \* site eliminated in carriers.

(Note that the first intronic *AcilI* site may be polymorphic because it was observed by the candidate and reported by Derry *et al.*, (1995a), but was not present in the sequence reported by Kwan *et al.*, (1995)).

**Figure 8.17** *WAS* exon 4 PCR products cut with *AclI*.





**Figure 8.18** Results of the SSCA and, for IV-4, IV-5 and V-1 only, *AciI* restriction enzyme digestion of the *WAS* exon 4 PCR products.

## 8.5 Discussion

### 8.5.1 X-linked spastic paraplegia

During the course of the linkage analyses presented in 8.2.3.1, a mutation segregating with the affected members of a family suffering from spastic paraplegia (Bonneau *et al.*, 1993) was found within the third exon of the *PLP* gene, which maps to Xq22.2, between *DXS3* and *DXS424* (Saugier-Veber *et al.*, 1994; Nelson *et al.*, 1995). Given that the localisation of the disease gene in the family presented herein was between the *AR* (Xq12) and *DXS424* (Xq23) loci the *PLP* gene became a prime candidate for the disease causing gene.

The T<sup>707</sup> to C mutation discovered within the *PLP* gene would result in the change of amino acid Phe<sup>236</sup> to Ser (Fig. 8.4). According to one PLP model (Popot *et al.*, 1991), this residue lies within the fourth transmembrane alpha-helix (helix D) (and is not one of the 35 amino acids deleted, as a consequence of alternative splicing, in the isoform: DM-20). Given that a hydrophobic amino acid (phenylalanine) is replaced by a hydrophilic amino acid (serine) there is the potential to disrupt this alpha-helix and/or its interaction with other molecules, such as inositol hexakisphosphate, which binds to PLP *in vitro* (Yamaguchi *et al.*, 1996).

This is the first mutation within exon 6 of the human *PLP* gene to be reported in association with the spastic paraplegia phenotype and the second to be found in exon 6 of the human *PLP* gene. The first mutation (a G<sup>742</sup> to C transversion) was discovered in the DNA of a boy suffering from PMD (Pratt *et al.*, 1995a). The *jimpy*<sup>(msd)</sup> mouse also has a mutation within this exon (Gencic *et al.*, 1990). All three mutations would result in the substitutions of amino acids within the fourth alpha-helix of the protein, according to the model cited above. Different mutations affecting different PLP functions may explain why one of the mutations in the human gene causes PMD and the other spastic paraplegia. The remainder of the *PLP* mutations so far reported have been in other exons and most human *PLP* mutations so far detected have been different (Kaye *et al.*, 1994; Pratt *et al.*, 1995b). To date, only two other *PLP* mutations have been associated with SPG2 (Kobayashi *et al.*, 1994; Saugier-Veber *et al.*, 1994), and both, unlike the mutation reported here, are expected to result in amino acid substitutions within hydrophilic domains of the protein

(Popot *et al.*, 1991). The other 38 known mutations, within the protein coding exons 2 to 7, have been associated with the more severe phenotype exhibited in PMD (Hodes and Dlouhy, 1995; Kaye *et al.*, 1994; Pratt *et al.*, 1995b).

Additional evidence in support of the exon 6 base change being responsible for the disease in the family described herein could be obtained from histopathological observations of an abnormal amount and/or appearance of CNS myelin of affected males or perhaps from the creation of a similarly affected transgenic mouse with an identical mutation.

Proteolipid protein constitutes 50% of the protein of adult central nervous system (CNS) myelin (Eng *et al.*, 1968). Myelin is an extension of the plasma membrane of oligodendrocytes that enwraps the axons of neurons (Peters *et al.*, 1970). PLP is thought to be required for oligodendrocyte maturation and survival and for maintaining myelin structure (Duncan, 1990; Boison *et al.*, 1995). The amino acid sequence of this protein is highly conserved: for example, the human, rat and mouse sequences are identical (Macklin *et al.*, 1987; Schliess and Stoffel, 1991; Kitagawa *et al.*, 1993). Heterologous *in vitro* experiments investigating the consequences of the *jimpy*<sup>(msd)</sup> mutation have shown that the defective PLP accumulates within the rough endoplasmic reticulum which, extrapolating to oligodendrocytes, is thought to result in hypomyelination and early cell death (Gow *et al.*, 1994). The relatively mild spastic paraplegia phenotype compared to that exhibited by those affected with PMD may be explained by the former exhibiting dysmyelination (disruption of the myelin structure and as a consequence reduced conduction velocities) and the latter hypomyelination. The loss of oligodendrocytes, perhaps due to the progressive accumulation of toxic mis-folded PLP, may explain the *progressive* nature of the paraparesis seen in the affected members of the family shown in Figure 8.1, but does not account for the specificity of these symptoms. A role for PLP in the peripheral nervous system has been proposed to account for the symptoms of some PMD patients (Kaye *et al.*, 1994).

Phenotypic diversity associated with allelic heterogeneity has been noted previously (Suthers and Davies, 1992; Romeo and McKusick, 1994) and given that spastic paraplegia (both complicated and uncomplicated) and PMD have now been shown to be allelic disorders, other families with symptoms of spasticity that map to Xq21-q22 (Keppen *et al.*, 1987; Goldblatt *et al.*, 1989; Le Merrer *et al.*, 1989) should be investigated for

mutations in the *PLP* gene.

### 8.5.2 Myotubular myopathy

The *MTMI* C<sup>775</sup> to T base change would result in an Arg to Cys substitution at codon 259. This residue has been conserved in the predicted homologous proteins of both *Saccharomyces cerevisiae* and *Caenorhabditis elegans* (Laporte *et al.*, 1996). Given that a non-conservative (in terms of structure and evolution) amino acid substitution has occurred this mutation potentially has functional significance for the putative protein (myotubularin).

At the time the manuscript (appendix A.7) describing the C<sup>775</sup> to T change was submitted for publication, this was the second mutation within exon b of the *MTMI* gene to be reported in association with myotubular myopathy. The first mutation, a deletion (AC763-764), is expected to cause a frameshift and a truncated protein (Laporte *et al.*, 1996). Subsequently, the same C<sup>775</sup> to T base change has been observed in another mildly affected individual (J-L Mandel, personal communication).

Phenotypic confirmation of the carrier status of the boy's mother may be obtained if she has pathological muscle histology (although her carrier status cannot be excluded if she does not have any abnormal muscle phenotype). Given that some phenotypic heterogeneity has been reported in association with XLMTM (Wallgren-Pettersson *et al.*, 1995) (whether this is due to allelic heterogeneity and/or genetic heterogeneity is, as yet, undetermined) the identification of the particular mutation reported herein will contribute to a database showing the correlation between specific *MTMI* gene mutations and the resultant phenotypes. (The mutation reported herein is associated with a relatively mild myopathic phenotype). Such knowledge will assist in the prognosis of any other sporadic males from other families.

### 8.5.3 Wiskott-Aldrich syndrome

The G<sup>407</sup>A change is the 35th *WAS* gene missense mutation to be reported in association with either Wiskott-Aldrich syndrome or X-linked thrombocytopenia (Human Gene Mutation Database, January 1997). Most mutations are found within the first 4 exons

suggesting that these encode for functionally important domains.

Evidence that the *WAS* mutation reported above is deleterious, and not just a neutral variant, includes:

- a) The conformational change was not observed in unaffected males of the family, nor in 99 unrelated X chromosomes;
- b) The base change would result in an amino acid change. Although charged and basic arginine is replacing non-polar glycine, this change is thought to be relatively conservative because both are hydrophilic and the residue lies in a hydrophilic portion of a generally hydrophilic protein (Derry *et al.*, 1994). Therefore the change is unlikely to be involved in disrupting the structure of the molecule, but may disrupt its interaction with other molecules.
- c) This amino acid residue has, however, been conserved in the predicted homologous protein of mice (Derry *et al.*, 1995b), which is evidence that it does serve a critical function.
- d) This amino acid is also considered part of the putative “WASP homology 1 domain” (WH1). This domain has been evolutionarily conserved among proteins involved in cytoskeletal organisation (Symons *et al.*, 1996).
- e) This same mutation has also been observed in the DNA of a boy with a severe form of *WAS* (Zhu *et al.*, 1995).

The function of the 502 amino acid *WAS* protein (WASP) is not known, but is generally considered to have multiple functions (Featherston, 1997). The protein does bind, *in vitro*, to domains of cytoplasmic tyrosine kinases involved in lymphoid cell signaling pathways (Cory *et al.*, 1996), and thus it may be involved in signal transduction. As mentioned above, WASP may instead, or also, be involved in maintaining the cytoskeletal architecture of T cells (Molina *et al.*, 1992; Symons *et al.*, 1996). Another suggestion has been made that, because of the presence of a leucine zipper-like motif, the protein has a role as a transcription factor (Schindelhauer *et al.*, 1996). Its function, or the importance of its function, does seem to vary in different cell types, given that most mutations affect megakaryocytes (from which platelets are derived), thus producing the relatively mild phenotype associated with XLT, whereas other mutations also affect the function of immune response cells, thus giving rise to more severe symptoms.

In general however, it has proven difficult to elucidate any positive genotype/phenotype correlations, that is, one cannot predict with confidence the disease symptoms that will arise from a particular *WAS* mutation because some mutations can cause either XLT or WAS in different families (Kolluri *et al.*, 1995; De Saint Basile *et al.*, 1996; Greer *et al.*, 1996; Schindelhauer *et al.*, 1996). Most patients with XLT have conservative missense mutations, whereas both attenuated *and* severe forms of WAS are associated with nonsense, frameshift or missense mutations (Derry *et al.*, 1995a; Villa *et al.*, 1995; De Saint Basile *et al.*, 1996; Greer *et al.*, 1996; Schindelhauer *et al.*, 1996). Apart from interfamilial variability, there is also intrafamilial variability (Sullivan *et al.*, 1994; Schindelhauer *et al.*, 1996), which also suggests that other genes, environmental factors and/or additional (undetected) *WAS* mutations affect the phenotype.

## 8.6 Conclusion

A point mutation within the *PLP* gene has been found to cosegregate with the affected and carrier members of a family suffering from progressive spasticity of the legs. Given the non-conservative nature of the predicted amino acid substitution, the highly conserved sequence of the protein, the detrimental effects of two previously reported mutations within the same exon and the absence of this base change in 220 unrelated X chromosomes, this mutation is unlikely to be a rare, selectively neutral, variant or polymorphism. Identification of the mutation has allowed the identification of female carriers, and will permit the future determination of the carrier status of other males and females of the family if they wish to avail themselves of this direct test. The published manuscript of this work appears in the appendix (A.6; Donnelly *et al.*, 1996a).

The *MTMI* exon b mutation is also unlikely to be a rare neutral variant because of the non-conservative nature of the predicted amino acid substitution, the phenotypic consequences of the same and other mutations in the *MTMI* gene, the absence of this base change in 159 unrelated X chromosomes, and the fact that the mother is a new mutant (or has inherited the base change from a gonadal mosaic parent). The identification of this mutation has shown that the disease in the family studied is the form associated with *MTMI* in Xq28, as distinct from any other X-linked or autosomal myotubular myopathy. Future prenatal diagnosis can now be offered to the carrier mother, should she request it. Other potential female carriers from the family have been assessed for carrier status and found to be non-carriers. The manuscript (in press) of this work appears in the appendix (A.7; Donnelly *et al.*, 1997).

The *WAS* mutation reported herein is unlikely to be a rare neutral variant given the absence of this base change in 99 unrelated X chromosomes, its predicted (evolutionary conserved) amino acid substitution, and the existence of an affected boy with the same mutation. Prenatal diagnosis of future pregnancies is now available to the mother of the proband, and the carrier status of the proband's sisters and any other, as yet untested, females of the extended family can be easily determined.

Chapter 9

Other Approaches to

Isolate Genes for

X-linked Disorders

## TABLE OF CONTENTS

	Page
<b>9.1 Summary</b>	290
<b>9.2 Introduction</b>	291
<b>9.3 Materials and Methods</b>	293
9.3.1 Clinical material and cytogenetic analyses	293
9.3.1.1 Case S	293
9.3.1.2 Case M	293
9.3.1.3 Case B	293
9.3.2 Selection of YAC and cosmid clones	294
9.3.3 YAC and cosmid DNA preparation	294
9.3.4 FISH	295
<b>9.4 Results</b>	295
9.4.1 Xp22 breakpoint	295
9.4.2 Xq24 breakpoint	301
9.4.3 Xq26-q28 duplication	307
<b>9.5 Discussion</b>	313
9.5.1 Xp22 breakpoint	313
9.5.2 Xq24 breakpoint	313
9.5.3 Xq26-q28 duplication	314
<b>9.6 Conclusion</b>	317

## 9.1 Summary

A female with global developmental delay, another with mental retardation and manic-depression and a third who was the mother of a boy with features of Borjeson-Forssman-Lehmann Syndrome were characterised cytogenetically. The first two females had X;autosome translocations and the third had a duplication of the distal portion of the q arm of the X chromosome. Genes identified at or near these X chromosome alterations are candidates for disorders which affect the central nervous system. The X chromosome breakpoints were determined by FISH, primarily using YACs. The first female has an X;6 (p22.2;q14) breakpoint that most likely lies between *DXS1317* and *DXS999*. The second female has an X;12 (q24;q14) breakpoint that most likely lies between *DXS8059* and *DXS1212*. The third female possesses a duplication of the X chromosome that begins between *DXS1062* (Xq26.2) and *DXS294* (Xq26.3) and extends into Xq28. This targetted approach to gene identification represents an alternative to the positional candidate approach used in Chapter 7.

## 9.2 Introduction

Visible cytogenetic abnormalities such as translocations and deletions have greatly facilitated the localisation and ultimately the identification (positional cloning) of many genes. Of 42 genes identified by positional cloning up until 1995, the cloning process in 26 cases had been greatly assisted by the presence of cytogenetic abnormalities (Collins, 1995). With the currently available STS- and YAC-based physical maps it is relatively easy to accurately position the breakpoints of any chromosomal rearrangement.

X chromosome rearrangements have been particularly useful for mapping and identifying X-linked genes because carrier females either bear affected males or themselves suffer phenotypic consequences and thus undergo clinical and cytogenetic investigation. This chapter presents the preliminary molecular investigations of the cytogenetic abnormalities possessed by three such females. It is hypothesised that these abnormalities are associated with the disease phenotype in the affected females or their male offspring. Characterisation of these molecular defects will lead to the identification of the disease genes and may help elucidate the mechanisms of the translocations and duplication events by examination of the DNA sequence near the breakpoints.

To explain the existence of females with symptoms attributable to apparently balanced translocations involving the X chromosome one or more of the following scenarios can be envisaged:

- 1) Mosaicism of the abnormality. The clinical consequences will depend on the proportion of cells that contain the abnormality.
- 2) An X chromosome gene is disrupted by the translocation. Those cells in which the normal X is inactivated will be functionally nullisomic for that gene, or monosomic if that gene is one that escapes inactivation.
- 3) An autosomal gene is disrupted by the translocation. All cells will then be functionally monosomic for that gene.
- 4) Spreading of inactivation from the derived X into the autosome (that is, the normal X remains active). This will result in functional monosomy of those autosomal genes that become inactivated, if the translocation is balanced. The extent of spreading may vary

between cells and hence the phenotype may also vary.

5) The normal X remains active. This means the cell will be functionally disomic for those X genes translocated to the autosome (Schmidt and Du Sart, 1992; Hatchwell *et al.*, 1996).

6) Position effects. This refers to the up- or down-regulation of expression of genes because of the juxtaposition of regulatory sequences of other genes (Milot *et al.*, 1996).

7) The consequences of (2) to (6) may depend on the cell type and the genes involved. Some cells may not be able to mature or survive, whereas other cell types will, that is, the phenotypic consequences will depend on which part of the X and how much of the X is translocated, and similarly for the autosome. The pattern of cell survival will be reflected in the pattern of X inactivation. Inactivation will be skewed in those cell types which cannot tolerate the consequences of the translocation.

A review of 122 females with balanced X; autosome translocations showed that in 77% of cases it is the translocated X that remains active (Schmidt and Du Sart, 1992). The other 23% of females had some proportion of lymphocytes and/or fibroblasts that were functionally disomic for that part of the X chromosome on the derived autosome because the normal X was early replicating (that is, active) and those cases with the greatest proportion of cells with an active normal X were translocations involving Xp22 or Xq28. In other words, functional disomy for regions of Xp22 or Xq28 is often compatible with cell survival and persistence.

Reports of females with partial duplications of the X chromosome reveal that in almost all cases it is the X with the duplication that is inactivated (Aughton *et al.*, 1993), which would account for their lack of disease symptoms. However, these same duplications result in affected male offspring when compatible with life (Schmidt *et al.*, 1991). The phenotype of the males can be explained by disomy for those genes duplicated and/or disruption of the gene by one of the breakpoints, especially if the phenotype is similar to that of a gene independently found to map to the same location.

### 9.3 Materials and Methods

Only brief accounts of the methods used are given below, further details are given in the Chapter 3 sections indicated.

#### 9.3.1 Clinical material and cytogenetic analyses

Cytogenetic analyses were performed by staff of the Cytogenetics Unit of the Department of Cytogenetics and Molecular Genetics, Women's and Children's Hospital (WCH), unless indicated otherwise.

##### 9.3.1.1 Case S

Patient S is a female who was born on 30/5/94. A tentative diagnosis of partial Aicardi syndrome (Aicardi *et al.*, 1965) was made based on her infantile spasms and her global developmental delay. Cytogenetic analysis revealed that she had the karyotype: 46,X,t(X;6)(p22.31;q14). The translocation was balanced at the level of light microscopy. No mosaicism was evident in cultured leucocytes. X inactivation was random.

##### 9.3.1.2 Case M.

Patient M is a female who was born on 6/9/78. She was diagnosed as being mentally retarded and having bipolar disorder (BP; Manic-depression; MIM 309200) by Dr Eskavari (WCH). Cytogenetic analysis revealed that she had the karyotype: 46,X,t(X;12)(q24;q15). The translocation was balanced at the level of light microscopy. No mosaicism was evident in the cultured leucocytes. X inactivation was skewed, with 100% of cells having a late replicating normal X (ie inactivated).

##### 9.3.1.3 Case B.

Case B is a female who is the mother of a male born in 1974. This boy was brought to the attention of the Department of Cytogenetics and Molecular Genetics by Drs Ming and Zackai of the Children's Hospital of Philadelphia (CHP). They suspected he was suffering from Borjeson-Forssman-Lehmann Syndrome (BFLS; MIM 301900; Borjeson *et al.*, 1962). He had a history of early hypotonia and poor feeding. Chromosome analysis (CHP) showed a duplication of Xq27. Features consistent with BFLS were: profound

developmental delay, neonatal hypotonia, short stature, bilateral cryptorchidism with a small penis, deep set eyes and obesity. In addition he had one kidney, conductive hearing loss and hypertelorism. The patient died from surgical complications and no tissue was kept for cytogenetic or molecular genetic analysis. Chromosome analysis of the phenotypically normal mother showed that she also carried the duplication. It is unknown if she is mosaic for the duplication. X inactivation was skewed; the duplicated X was late-replicating (ie inactivated) in 25/25 metaphases.

### 9.3.2 Selection of YAC and cosmid clones

Cosmid clones mapping to Xp22.3 were selected, by the candidate, from the physical map given by Wapenaar *et al.* (1994). YAC clones mapping to Xp22 were selected from the physical map given by Ferrero *et al.* (1995). Both cosmid and YAC clones were kindly provided by Dr Brunella Franco of the Telethon Institute of Genetics and Medicine in Milan, Italy.

CEPH YAC clones mapping to Xq24 were selected from a physical map available from the Whitehead Institute web site ([http://www-genome.wi.mit.edu/cgi-bin/contig/phys\\_map](http://www-genome.wi.mit.edu/cgi-bin/contig/phys_map); Hudson *et al.*, 1995). These clones were purchased from Genome Systems Inc. The Washington University School of Medicine (WUSM) YAC clones (designated WXD...) mapping to Xq24 were selected from a physical map kindly provided by Dr Ramaiah Nagaraja (WUSM Center for Genetics in Medicine). Details of the YACs and STSs on this map are available from web site <http://genome.wustl.edu/cgm/cgm.html>. They were purchased from ATCC.

CEPH YAC clones mapping to Xq26 were also selected from the physical map available at the Whitehead Institute web site ([http://www-genome.wi.mit.edu/cgi-bin/contig/phys\\_map](http://www-genome.wi.mit.edu/cgi-bin/contig/phys_map); Hudson *et al.*, 1995). These clones were purchased from Genome Systems Inc. WUSM YAC clones mapping to Xq26 were selected from the physical map of Pilia *et al.* (1996). They were purchased from ATCC. Clones: D2425, 4D-8 (*DXS98*), VK16 (*DXS293*), VK21 (*DXS296*), 910H1 and 49.8 (*DXS295* and *DXS296*) had been previously used for investigations unrelated to those reported herein.

### 9.3.3 YAC and cosmid DNA preparation

YAC DNA was prepared, by the candidate, as described in 3.2.7.3. Cosmid DNA was

prepared as described in 3.2.7.6.

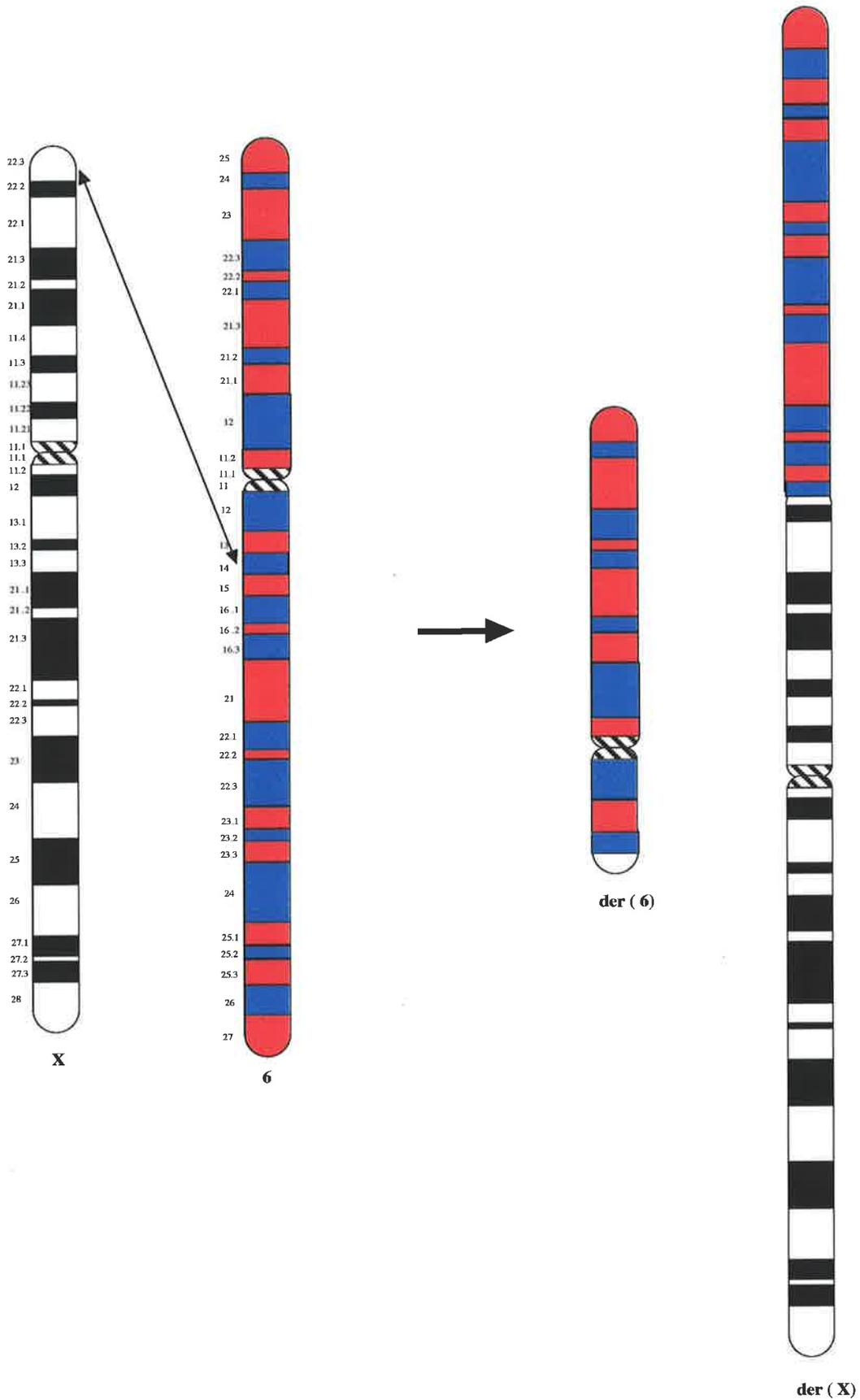
### 9.3.4 FISH

Fluorescence *in situ* hybridisation to metaphase spreads, from short term leucocyte cultures (for cases S and M) and lymphoblastoid cells (for case B), was performed by Liz Baker and Helen Eyre (WCH) using YAC, cosmid, lambda or plasmid DNA as described in 3.2.17.

## 9.4 Results

### 9.4.1 Xp22 breakpoint

The tentative diagnosis of Aicardi syndrome (AIC) for this female led the candidate to select cosmid clones from the putative AIC critical region in Xp22.31. Figure 9.1 shows the break diagrammatically and, as an example, the FISH of cosmid U91B4 is shown in Figure 9.2. The cosmid clones hybridised distal to the breakpoint. This result led to the selection of YAC clones that were expected to flank the breakpoint. Flanking YAC clones were obtained and the breakpoint was progressively approached by several rounds of YAC selection and hybridisation. Table 9.1 shows the cosmids and YACs used for FISH and their position relative to the breakpoint. Figure 9.3 shows the physical map of the region and the position of each YAC relative to the breakpoint, which is in Xp22.2 (between *DXS1195* and *DXS999*) and not Xp22.31. YAC 434e8 has been tentatively assigned a position distal to the breakpoint but the hybridisation signal:noise ratio was poor.



**Figure 9.1** Idiogram showing the translocation possessed by female S.

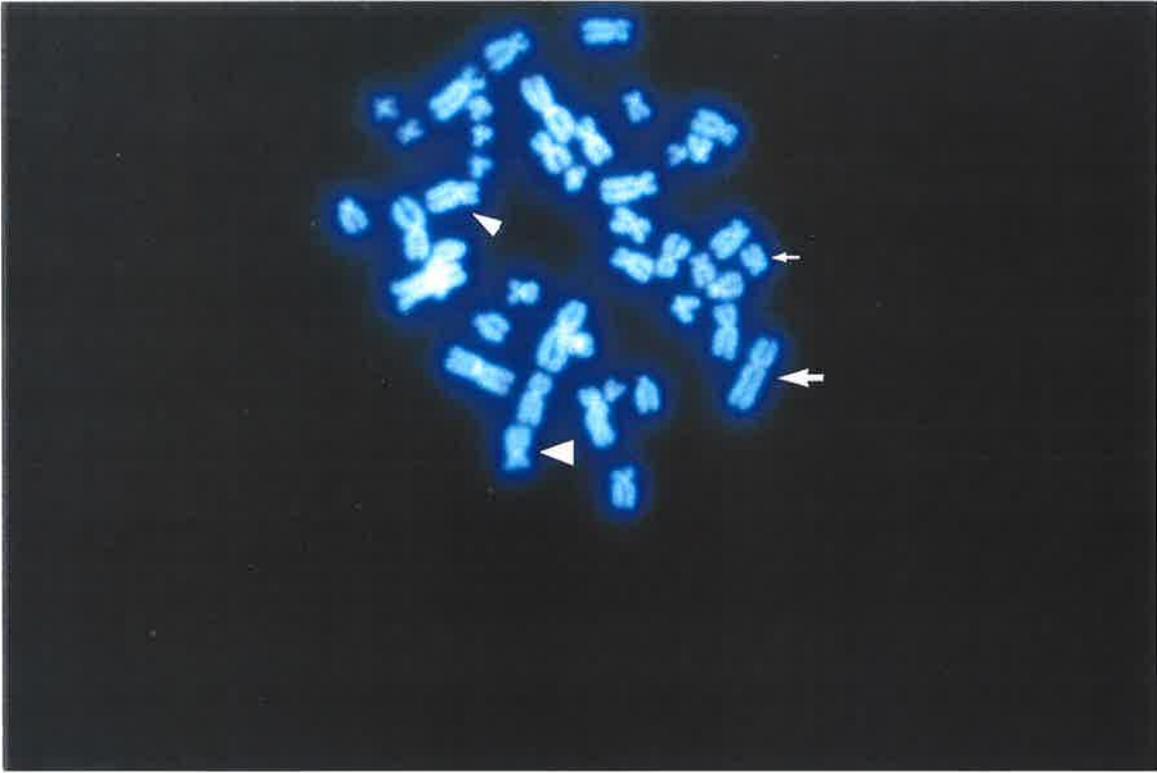
Her karyotype is 46,X,t(X;6)(p22.3;q14).

**Figure 9.2** Metaphase, from case S, showing FISH with probe cU91B4.

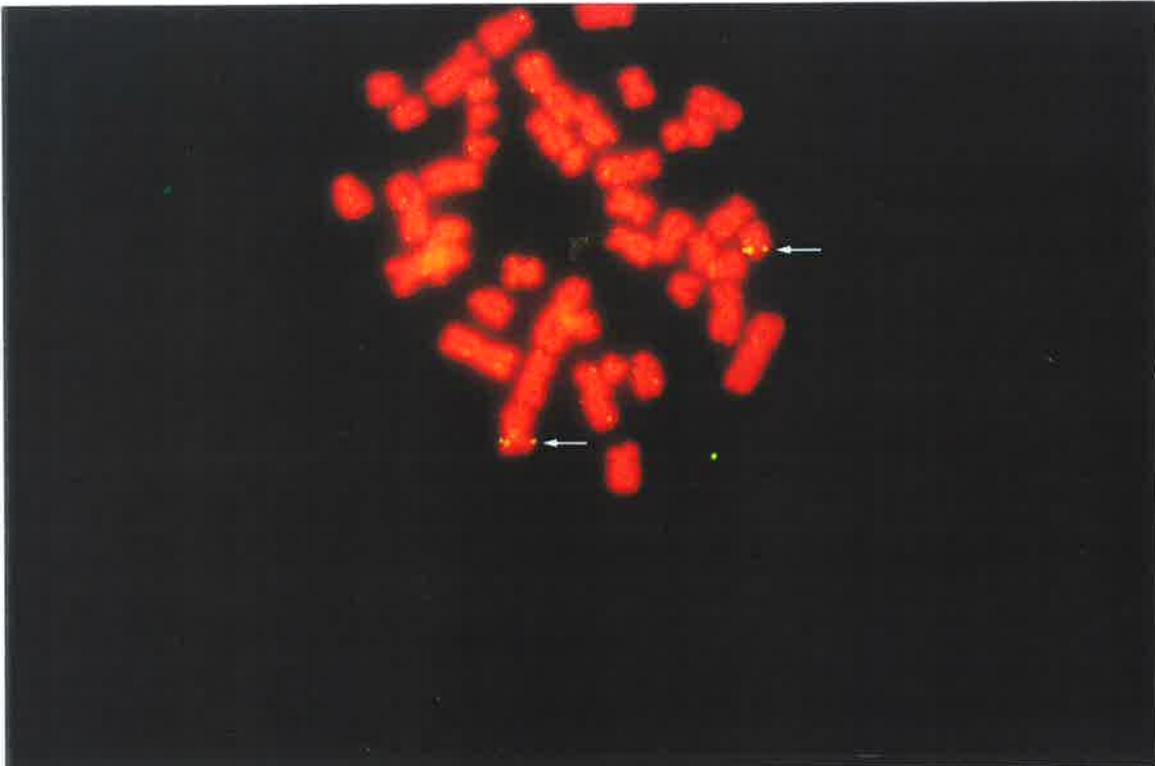
A: Chromosomes stained with DAPI for chromosome identification. The normal X chromosome is indicated by the large arrowhead. The derivative X chromosome is indicated by the large arrow. The normal 6 chromosome is indicated by the small arrowhead. The derivative 6 chromosome is indicated by the small arrow.

B: The same metaphase as above counter-stained with propidium iodide. Fluorescent signal from hybridisation of the probe cU91B4, which hybridises distal to the Xp22 breakpoint, is indicated by the arrows.

A



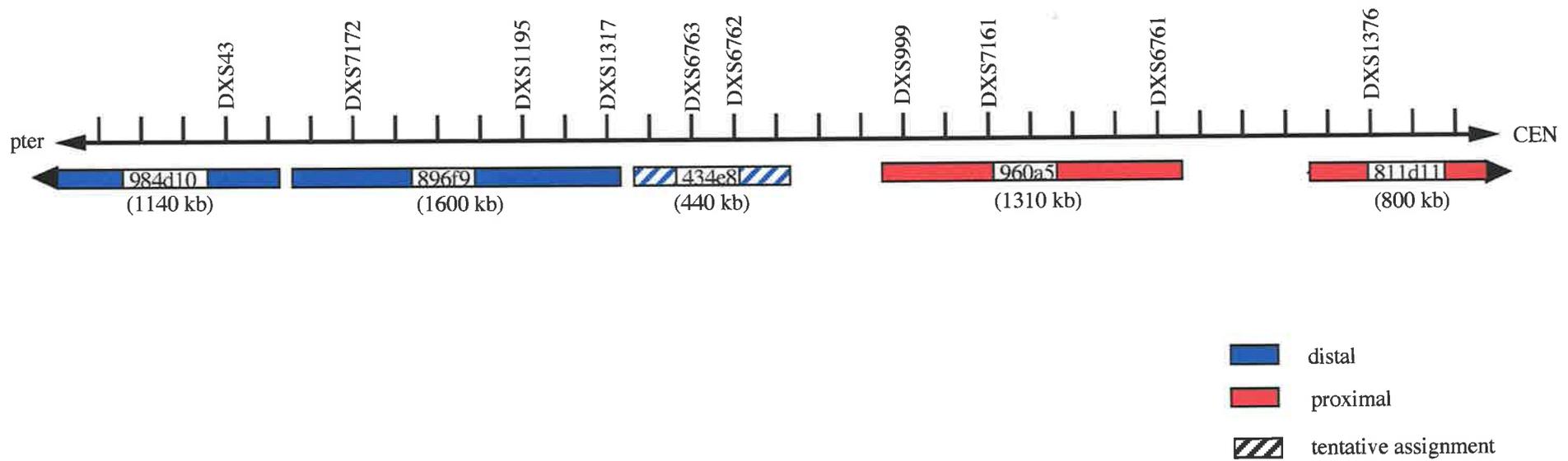
B



**Table 9.1 Probes used for FISH to define the Xp22 breakpoint**

probe (Xpter-cen)	FISH result	No. of cells scored
cU91B4	distal	10
c62H3	distal	10
y926d3	distal	10
y618f6	signal on 6q & 10q	-
y984d10	distal	10
y896f9	distal	10
y434e8	distal (but poor signal)	4
y47f10	poor signal	-
y718b8	no signal	-
y960a5	proximal	5
y811d11	proximal	10
y693d1	proximal	10
y954g6	proximal	10

c = cosmid; y = YAC.

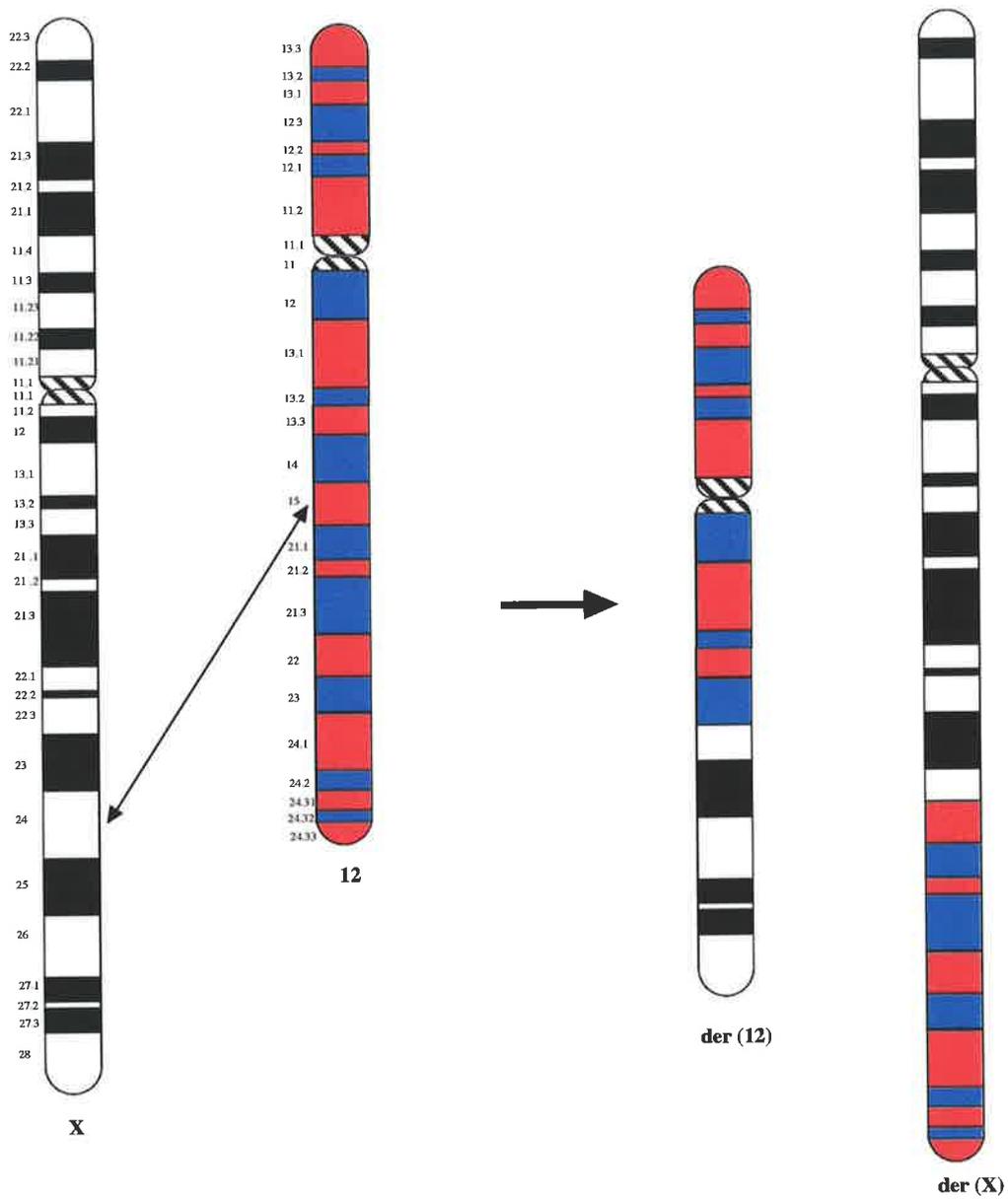


**Figure 9.3** Physical mapping of the Xp22.2 breakpoint, showing the relative position of CEPH YACs as determined by FISH.

The map is derived from Ferrero *et al.* (1995). YACs are not to scale. STSs (vertical lines) are shown to be equi-distant for simplicity.

#### 9.4.2 Xq24 breakpoint

This breakpoint is shown diagrammatically in Figure 9.4. CEPH YACs from Xq24 that were expected to flank the break were selected (Table 9.2). An example of the FISH is given in Figure 9.5 for YAC 896c5. Again, rounds of selection and hybridisation were performed to close in on the breakpoint. This proceeded until it was shown that the breakpoint occurred within a gap in the CEPH YAC map, between markers *DXS1001* and *DXS1212* (Fig 9.6) a region of at least 3 Mb (personal communication with Dr David Schlessinger). This gap has been mostly filled by Dr Schlessinger's group at the WUSM (unpublished data). The results from the YACs selected from this map are also given in Figure 9.6. From PCR analysis, the distal YAC 949f11 contains the locus *DXS8059*, as do YACs WXD6788 and WXD2446, both of which map proximal to the break. PCR analysis also confirmed that *DXS1212* is present in 949f11 and WXD2446.



**Figure 9.4** Idiogram showing the translocation possessed by female M.

Her karyotype is 46,X,t(X;12)(q24;q15).

**Table 9.2 YACs used for FISH to define the Xq24 breakpoint**

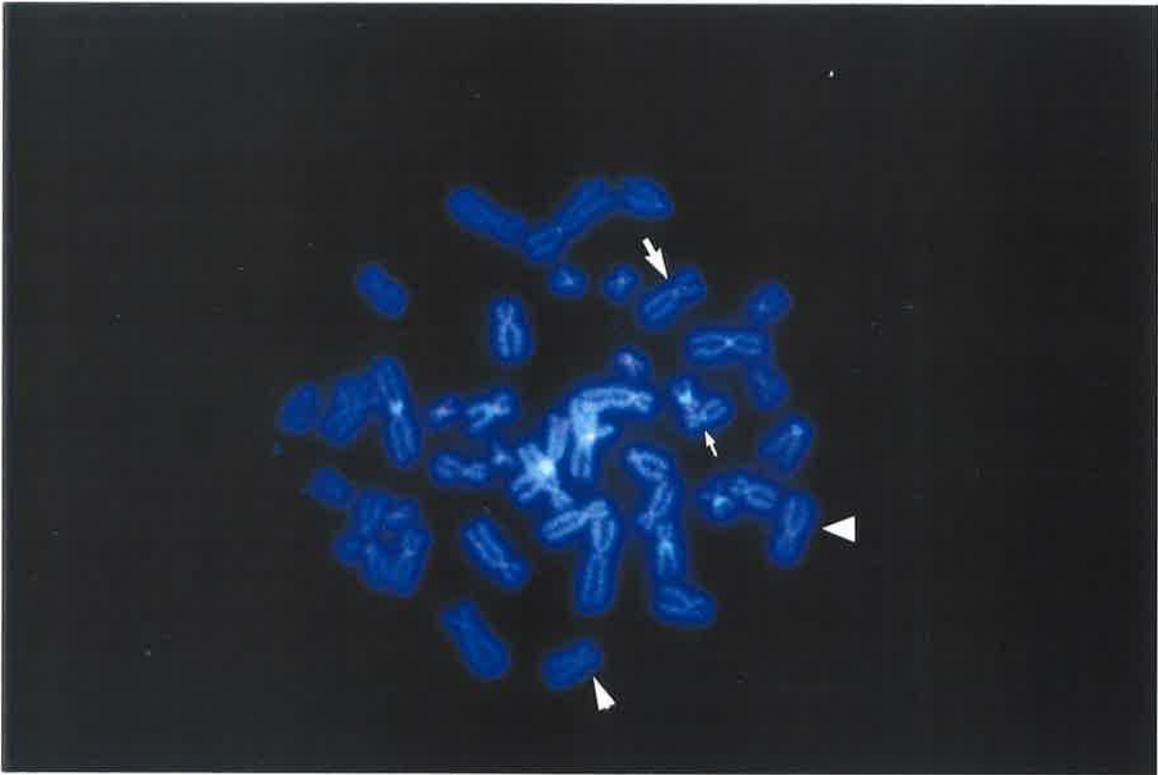
probe (cen-Xqter)	FISH result	No. cells scored
y923d3	proximal	5
y874f9	proximal	10
y705g4	proximal	10
yWXD8	proximal	10
yWXD284	proximal	10
yWXD444	proximal	6
yWXD6788	proximal	4
yWXD6792	proximal	6
yWXD2446	proximal	5
y949f11	distal	10
y779c2	distal	10
y803g7	distal	10
y896c5	distal	20

**Figure 9.5** Metaphase, from case M, showing FISH with probe y896c5.

A: Chromosomes stained with DAPI for chromosome identification. The normal X chromosome is indicated by the large arrowhead. The derivative X chromosome is indicated by the large arrow. The normal 12 chromosome is indicated by the small arrowhead. The derivative 12 chromosome is indicated by the small arrow.

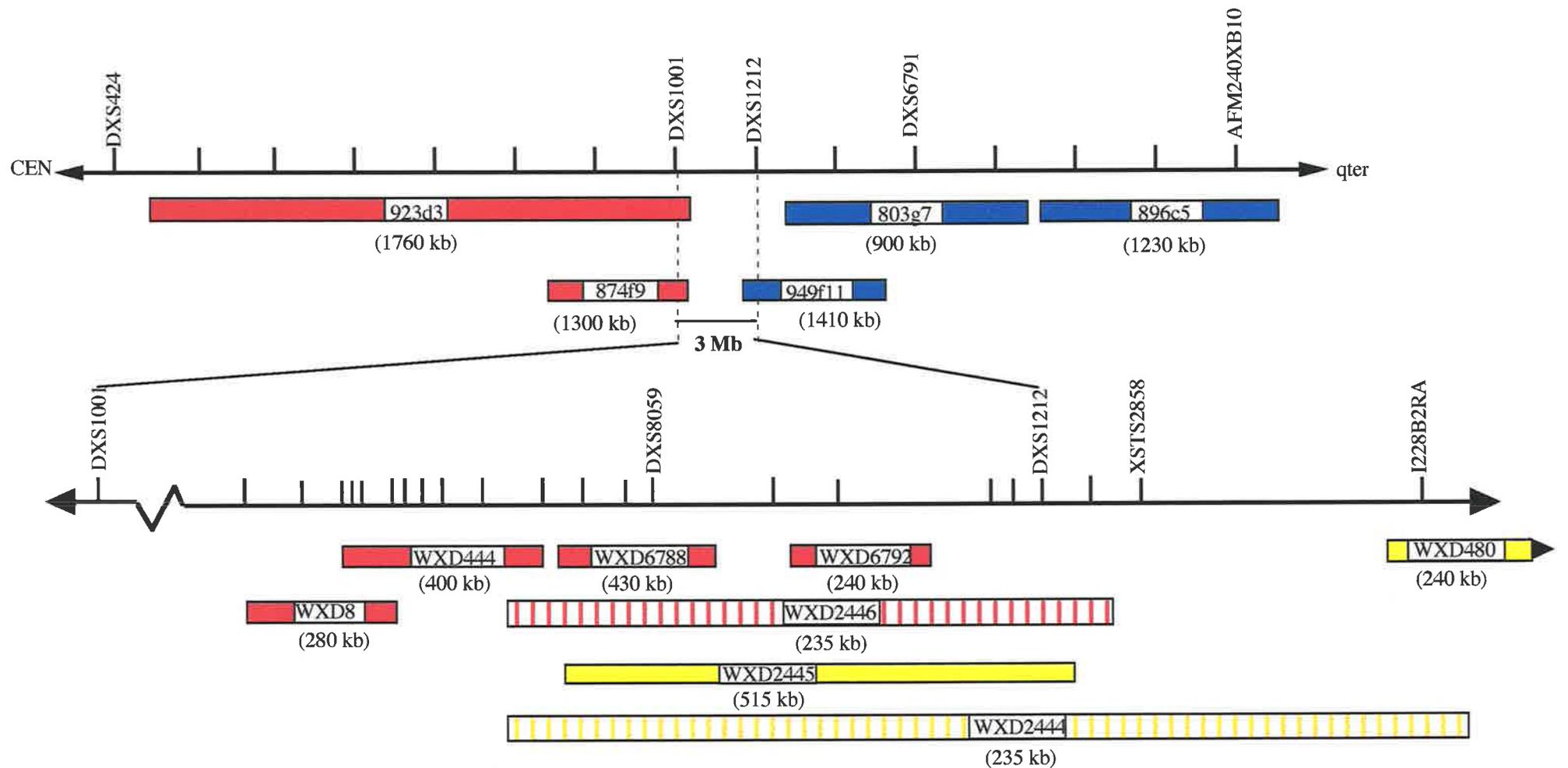
B: The same metaphase as above counter-stained with propidium iodide. Fluorescent signal from hybridisation of the probe y896c5, which hybridises distal to the Xq24 breakpoint, is indicated by the arrows.

A



B





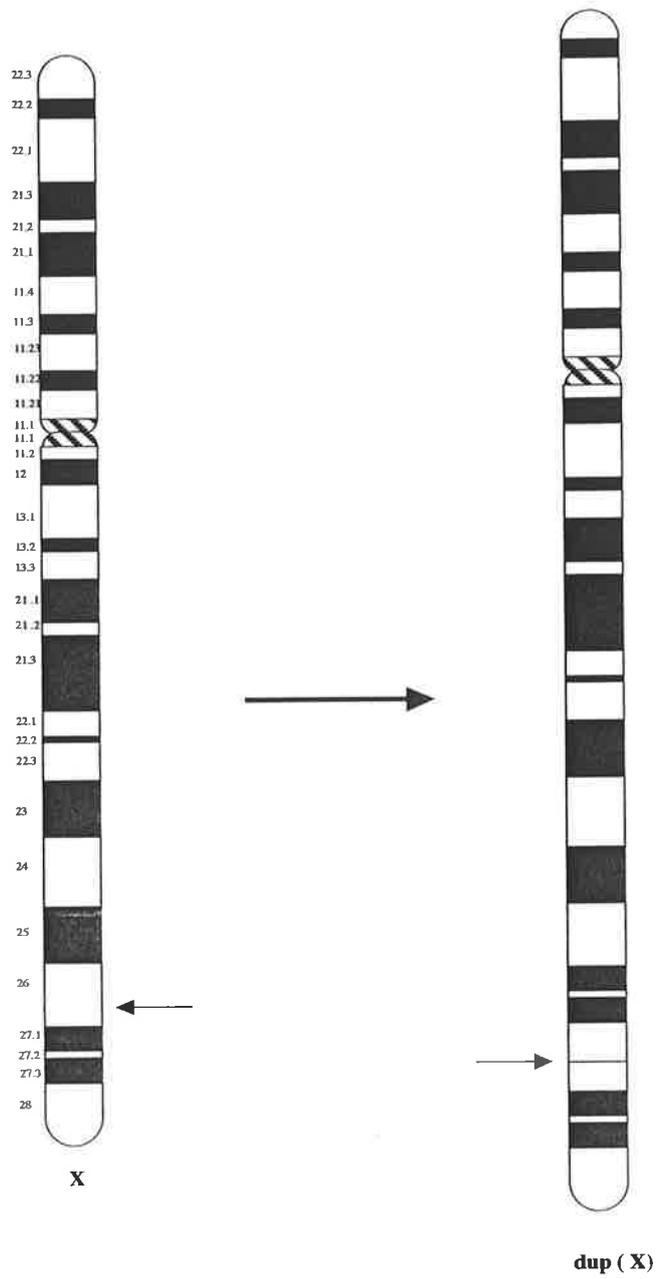
- distal
- proximal
- internally deleted
- awaiting FISH

**Figure 9.6** Physical mapping of the Xq24 breakpoint, showing the relative position of YACs as determined by FISH.

The top map is derived from Hudson *et al.* (1995). YACs are not to scale. STSs (vertical lines) are shown to be equi-distant for simplicity. The expanded map is from Dr. Ramaiah Nagaraja (unpublished data). STSs are positioned to scale.

### 9.4.3 Xq26-q28 duplication

The duplication is shown diagrammatically in Figure 9.7. (The extra material present in the long arm of the X chromosome was shown to be derived from the X by hybridisation with a total X paint). As an example, Figure 9.8 shows the duplication of probe 4D-8. The plasmid and YAC probes used for FISH are given in Table 9.3, with results of whether they have been duplicated or not. A physical map of the YAC clones used for FISH is given in Figure 9.9. At this stage of the investigation, YAC WXD404 (which contains the *DXS294* locus; Xq26.3) is the most centromeric probe to be shown to have been duplicated. The most telomeric YAC which is not duplicated is WXD491 (which contains the locus *DXS1062*; Xq26.2).



**Figure 9.7** Idiogram showing the duplication possessed by female B.

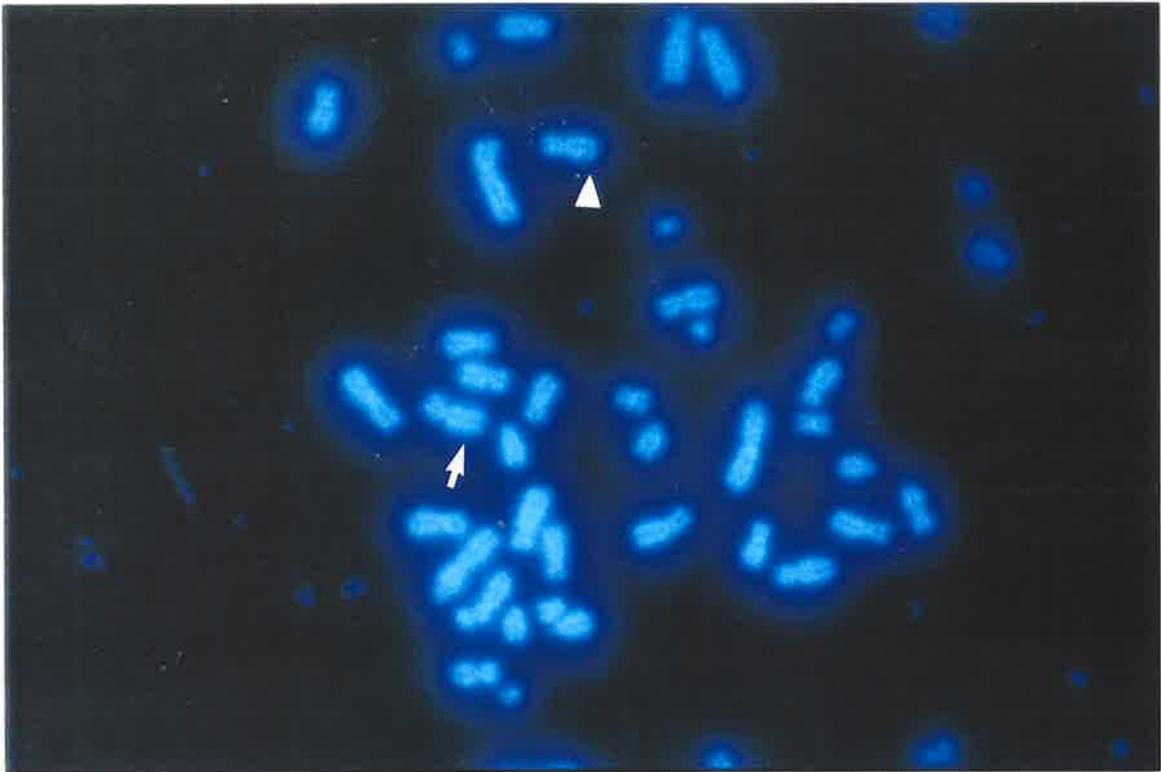
Her karyotype is 46,XX,dup(q26-q28).

**Figure 9.8** Metaphase, from case B, showing FISH with probe 4D-8.

A: Chromosomes stained with DAPI for chromosome identification. The normal X chromosome is indicated by an arrowhead. The abnormal X is indicated by an arrow.

B: The same metaphase as above counter-stained with propidium iodide. Fluorescent signal from hybridisation of the probe 4D-8 is indicated by the arrows. Duplication is evident on the abnormal X chromosome.

A



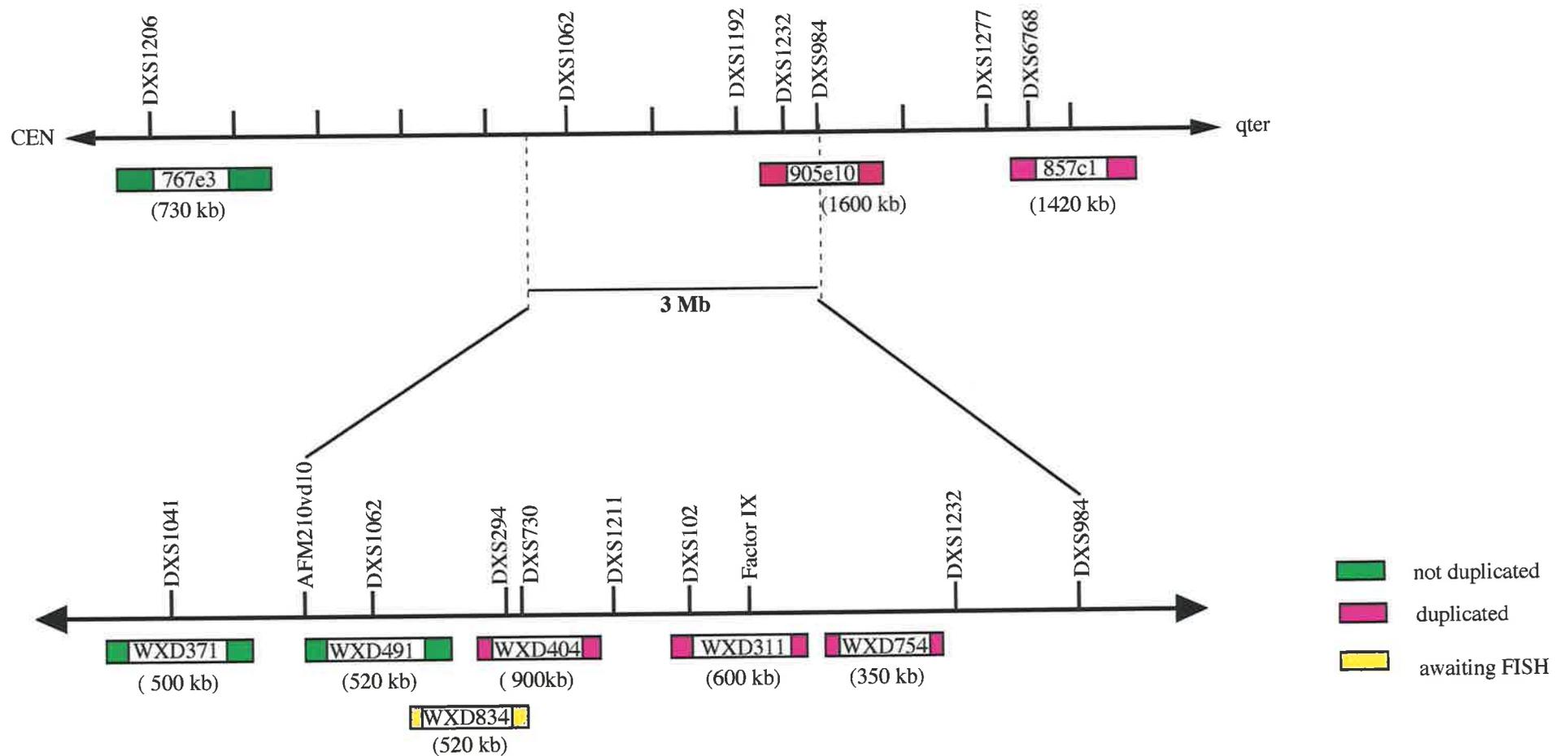
B



**Table 9.3 Probes used for FISH to define the Xq26 point and extent of duplication**

	probe (Xpter-Xqter)	FISH result	No. of cells scored
p22.3	c62H3	not duplicated	10
q12	cD2425	not duplicated	2
q13	y910H1	not duplicated	10
q24	y874f9	not duplicated	5
q24	y896c5	not duplicated	5
q25	y767e3	not duplicated	10
q26.2	y802f8	signal on 6q & 8q	-
q26.1-q26.2	yWXD371	not duplicated	11
q26.2	yWXD491	not duplicated	6
q26.3	yWXD404	duplicated	17
q26.3	yWXD311	duplicated	15
q26.3	yWXD754	duplicated	10
q27.1	y905e10	duplicated	10
q27.1	y857c1	duplicated	10
q27	p4D-8	duplicated	10
q27.3	λVK47	duplicated	2
q28	λVK21	duplicated	5
q28	y49.8	duplicated	7

c = cosmid; λ = lambda; p = plasmid; y = YAC.



**Figure 9.9** Physical mapping of the Xq26 point of duplication, showing whether or not YACs have been duplicated, as determined by FISH.

The top map is derived from Hudson *et al.* (1995). YACs are not to scale. STSs (vertical lines) are shown to be equi-distant for simplicity.

The expanded map is derived from Pilia *et al.* (1996). STSs are positioned to scale.

## 9.5 Discussion

### 9.5.1 Xp22 breakpoint

Aicardi syndrome (MIM 304050) is a male-lethal disorder that features mental retardation, infantile spasms, agenesis of the corpus callosum and chororetinal lacunae (Aicardi *et al.*, 1965). A few AIC patients have gross cytogenetic abnormalities involving Xp22 (reviewed by Ballabio and Andria, 1992). The AIC phenotype overlaps that of two other disorders, Goltz syndrome (also known as focal dermal hypoplasia (FDH; MIM 305600)) and microphthalmia with linear skin defects (MLS; MIM 309801), which has led to the suggestion that the same gene(s) may be involved (Lindsay *et al.*, 1994; Ballabio, 1995). Cytogenetic observations of deletions and translocations involving Xp22 has enabled the MLS critical region to be positioned to a 900 kb region between *DXS85* and *DXS1135* in Xp22.31 (Schaefer *et al.*, 1993; Wapenaar *et al.*, 1993; 1994). Given that the MLS region has been defined and that the MLS gene(s) may also be involved in AIC, cosmids flanking the MLS critical region (Wapenaar *et al.*, 1994) were used for FISH against the chromosome spreads of female S. However these cosmids (U91B4 and U62H3) mapped distal to the breakpoint. This suggests that although her phenotype shares some characteristics with AIC the AIC gene(s) is not involved. Nevertheless, any gene found at the breakpoint still represents a candidate for one or more of the *MRX* families whose disease gene localisations span this point (*MRX13*, *19*, *21*, *32* and *36*; Table 6.10; Fig. 6.15).

The next step was to use YACs that were expected to flank the breakpoint. These YACs were chosen from the map of Ferrero *et al.* (1995) and the region of the breakpoint was progressively narrowed. These YAC FISH results have shown that the break is in Xp22.2 and not p22.31 (as originally suggested by karyotype analysis) and that the breakpoint lies between YACs 896f9 (containing locus *DXS1317*) and 960a5 (containing locus *DXS999*).

### 9.5.2 Xq24 breakpoint

YAC WXD2446 (which contains internal deletions) was expected to span the break because it overlaps (data from the WUSM map and confirmed by PCR) a YAC that maps proximal (WXD6788) and a YAC that maps distal (949f11). This inconsistency could be explained by the possibility that both YACs 949f11 and WXD2446 do in fact span the

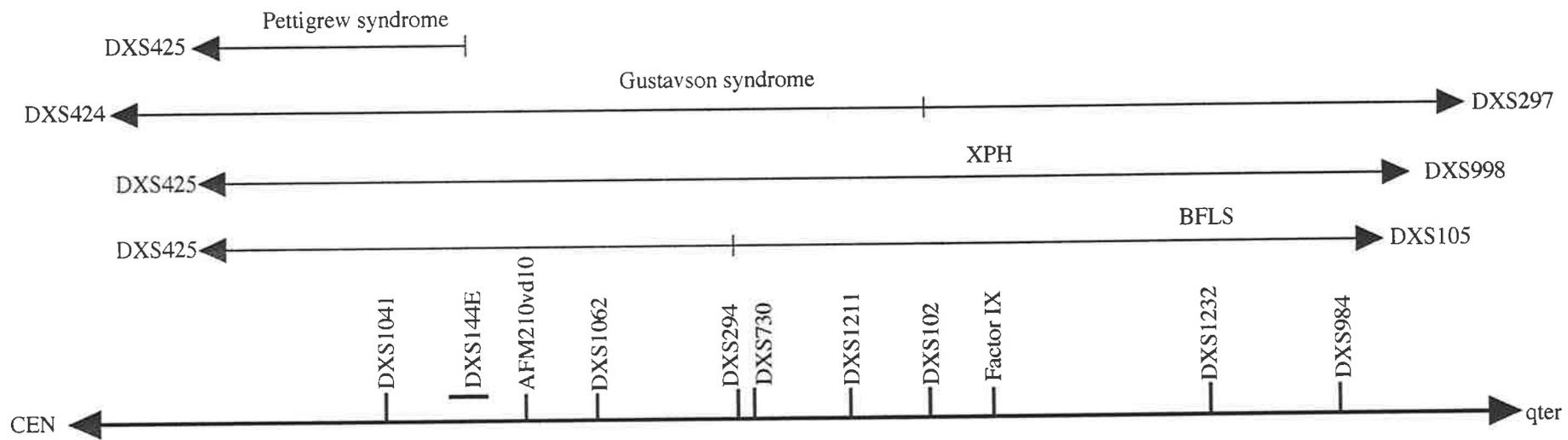
break but only by a small amount of sequence. Additional YACs are to undergo FISH to clarify this situation.

The female with the translocation has been diagnosed with bipolar disorder (BP; MIM 309200) and mental retardation. A number of linkage analyses have implicated distal Xq as harboring genes that may contribute to BP (reviewed by Risch and Botstein, 1996). One of these studies reported linkage to Xq24-q27.1 (Pekkarinen *et al.*, 1995). These authors performed linkage analyses using different modes of inheritance and different diagnostic categories. Using a particular set of diagnostic criteria and assuming one major locus was involved this study resulted in significant and suggestive lod scores for loci between *DXS1001* and *DXS292*. As shown in Figure 9.6, *DXS1001* lies proximal to the breakpoint in female M and thus it is possible that any gene disrupted by the break may also be responsible for the disease in the family reported by Pekkarinen *et al.* (1995).

### 9.5.3 Xq26-q28 duplication

As shown in Figure 9.10, BFLS has been determined by linkage analysis to lie between *DXS425* (Xq25) and *DXS105* (Xq27.1) (or possibly *DXS294* (Xq26.3) if one of the females of the family is considered a carrier) (Gedeon *et al.*, 1996c).

The symptoms affecting the boy could be due to a duplication of the BFLS gene(s) and/or disruption or loss of the gene(s) as a consequence of the event that also led to the duplication. The absence of any symptoms in the mother can be explained by inactivation of the duplicated chromosome. The fact that her cultured lymphoblasts exhibit skewed X inactivation is evidence that activity of the duplicated chromosome is incompatible with cell survival or persistence (*in vitro* at least) and thus it is puzzling that her son was born at all. (There has been a previous report of a duplication from Xq26.3 to qter in a dysmorphic and developmentally delayed female where the duplicated X remained active in all cells studied (Magenis *et al.*, 1984)). Given that the duplication encompasses up to 25 Mb, it would be expected that many genes would be duplicated and, as a consequence, that the transcription level of those expressed would be doubled. Conversely a male, only exhibiting haemophilia, possessing a 7 Mb deletion between *DXS730* (Xq26.3) and *DXS1200* (Xq27.3) has been reported (Gebbia *et al.*, 1996). The position of *DXS730* is shown in Figures 9.9 and 9.10. In addition, another report of a male-viable deletion extends this region telomerically another 5 Mb to the *FRAXA* locus (Quan *et al.*, 1995).



**Figure 9.10** The regional localisations of the genes responsible for Pettigrew, Gustavson, XPH and Borjeson-Forssman-Lehmann syndromes. The points at *DXS294* (BFLS) and *DXS102* (Gustavson syndrome) represent alternative telomeric limits to the disease gene localisation (refer to Gedeon *et al.*, 1996c and Malmgren *et al.*, 1993 respectively for details).

This is evidence that there are no genes in this 11 Mb region that are essential for viability and thus increased dosage of these genes may also be of little consequence, although the duplication possessed by female B extends about 2 Mb beyond *FRAXA* to *DXS295* at least (data from YAC 49.8), which is in Xq28; a band that has regions of high gene content (see for example De Sario *et al.*, 1996).

Given the FISH results the point of duplication lies somewhere between *DXS1062* and *DXS294*. If the distal limit of BFLS is taken to be *DXS294* (Gedeon *et al.*, 1996c) and if the boy does have BFLS, (as distinct from being a phenocopy) and that the duplication is responsible for generating the symptoms, then the most optimistic estimate of the critical region for BFLS is, at this stage, between *DXS1062* and *DXS294*, a distance of 540 kb (Pilia *et al.*, 1996).

In addition, the deletion patient reported by Gebbia *et al.* (1996) does not have BFLS, therefore if BFLS is the result of gene knockout the gene is likely to lie proximal to *DXS730* (Figs 9.9, 9.10).

There are three other syndromal disorders that map to this region: Pettigrew syndrome (Huang *et al.*, 1991; Pettigrew *et al.*, 1991), Gustavson syndrome (Gustavson *et al.*, 1993; Malmgren *et al.*, 1993) and X-linked recessive pan hypopituitarism (XHP) (Phelan *et al.*, 1971; Lagerstrom-Fermer *et al.*, 1997). Gustavson syndrome has some features in common with Pettigrew syndrome (mental retardation, spasticity and seizures) and some with BFLS (mental retardation, large ears and optic abnormalities) and it has been proposed that Gustavson syndrome may be a contiguous gene syndrome that has one or more genes in common with these other two syndromes (Gustavson *et al.*, 1993). The regional localisations of the genes of each of these syndromes are shown in Figure 9.10, and their relative positions does lend support to the "contiguous gene syndrome" hypothesis.

Recently, linkage and haplotype analysis of a family with males affected with X-linked recessive panhypopituitarism (XPH) showed the gene to be most likely between *DXS425* (Xq25) and *DXS998* (Xq27); a localisation which overlaps that of BFLS, as shown in Fig. 9.10 (Lagerstrom-Fermer *et al.*, 1997). The XPH clinical phenotype also overlaps that of BFLS; both feature short stature, genital abnormalities and mental retardation. Two of the affected males of the family were heterozygous for alleles of the *DXS102* locus.

Quantitative analyses confirmed an extra dosage of one of the *DXS102* alleles in carrier females. The same results have been observed for two other unspecified loci. From this evidence Lagerstrom-Fermer *et al.* (1997) have suggested that this family has a duplicated segment of the X chromosome that spans at least 4 cM. Therefore the phenotypic features shared by BFLS and XPH may be explained by duplication of the same gene(s). The only other disease which has symptoms similar to BFLS is Wilson-Turner syndrome, which maps to Xq11.3-Xq21.3 (Wilson *et al.*, 1991; Gedeon *et al.*, 1996e).

## 9.6 Conclusion

Once YACs spanning the two breakpoints and point of duplication have been found they can be subcloned into cosmid or lambda phage. These smaller clones can then be assembled into contigs and used for further FISH, to home in on the molecular lesions. Alternatively, STSs from the YACs could be used to screen bacterial artificial chromosome (BAC) or P1 artificial chromosome (PAC) gridded libraries. Clones that span these points of disruption can be used in exon trapping and/or cDNA selection experiments to find candidate genes. Those genes associated with the Xq24 breakpoint or the Xq26 duplication then become candidates for the disease genes in those families affected with BFLS or BP disorder respectively. Any gene associated with the Xp22.2 breakpoint can be investigated in those MRX families whose disease genes overlap this region.

## Chapter 10

## Conclusion

The Human Genome Project has advanced greatly since the implementation of the first 5 year plan in 1990 (Jordan, 1992; Collins and Galas, 1993). Its progress can be summarised as proceeding from genetic maps to physical maps to transcriptional maps. The ultimate aim is to sequence the entire genome and in the process identify all genes. This knowledge will be used to improve the health of individuals and society as a whole by providing therapies and preventative strategies.

With the generation of comprehensive linkage maps of the whole genome, linkage analysis of single gene disorders is now a routine procedure. The polymorphic markers used for linkage analyses have been used to select and characterise clones which are assembled to form physical maps. Upon these physical maps are being positioned ESTs, thus creating transcriptional maps. Most of the few hundred thousand ESTs representing perhaps half of the 70,000 human genes are expected to be assigned to their cognate genes and mapped using human-rodent radiation hybrids within the next few years (Berry *et al.*, 1995; Gerhold and Caskey, 1996). These ESTs then become candidates for disease genes regionally localised by linkage analysis. At the moment the isolation of genes from seed ESTs is being pursued by individual investigators interested in a particular phenotype, but in the future it is likely that large scale centralised concerted efforts will be made to isolate whole genes from these ESTs. Soon one will be able to identify all ESTs present within a defined genomic region by searching a database (Schuler *et al.*, 1996) or using that region to probe gridded cDNAs (Lee *et al.*, 1995). Already transcriptional maps are being published and described as being "body" maps or "human gene anatomy" maps as the tissue and developmental patterns of expression of ESTs are determined (see for example Adams *et al.*, 1995). The specific function(s) of the vast majority of the genes being discovered remains unknown and the pre-eminence of "classical" (or "forward") genetics, whereby one proceeds from phenotype to identification of the gene and its mutation, is likely to progressively give way to the pursuit of "reverse" genetics, whereby the function of an anonymous gene is determined by the directed mutation of this gene, its transference to a cell or organism, and the observation of any phenotypic consequences.

The work presented in this thesis has paralleled the progress of the Human Genome Project in miniature. It began with the generation of a microsatellite linkage map of the X chromosome (Chapters 4 and 5). This map was then used for linkage analyses which led to the regional localisation of five disease genes responsible for mental retardation (Chapter 6) and one for spastic paraplegia (in which the molecular defect was

subsequently identified; Chapter 8). With the aim of identifying genes involved in mental retardation, an attempt was made to isolate candidate genes from a region spanned by many of the *MRX* localisations, followed by the investigation of two genes which were good candidates for XLMR (Chapter 7). The mapping of disease genes presented in this thesis has been of direct benefit to potential carriers in these families to whom genetic risks could be communicated for the first time. The detection of mutations in other X-linked disease genes segregating in three other families (Chapter 8) allowed the application of a direct test to these families for the first time.

Despite the rapid progress of the HGP, it is still a daunting task to identify the genes responsible for simple Mendelian disorders, let alone those involved in complex traits. The frequency of X-linked non-specific mental retardation has been estimated at about 1/1000 males (Kerr *et al.*, 1991) and thus represents a significant burden on society and the families concerned. Finding and prioritising candidate MRX genes, is going to be particularly difficult because the regions defined by linkage analyses are often large, and within which a great many suitable candidates are likely to lie. The characterisation of cytogenetic abnormalities such as translocations, duplications and deletions is a means of targetted isolation of candidate genes, as exemplified by the positional cloning of *FMR2* (Gecz *et al.*, 1996; Gu Y *et al.*, 1996) and another MRX candidate gene: *DXS6673E* (van der Maarel *et al.*, 1996). It is with this in mind that the characterisation of three cytogenetic abnormalities, associated with the development and/or functioning of the brain, has begun (Chapter 9) in our laboratory as the basis for the future characterisation of candidate genes for X-linked disorders.

## References

- Abbs S, Roberts RG, Mathew CG, Bentley DR, and Bobrow M. 1990. Accurate assessment of intragenic recombination frequency within the Duchenne Muscular Dystrophy Gene. *Genomics* **7**: 602-606.
- Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merrill CR, Wu A, Olde B, and Moreno RF. 1991. Complementary DNA sequencing: Expressed sequence tags and human genome project. *Science* **252**: 1651-1656.
- Adams MD, Kerlavage AR, Fleischmann RD *et al.* 1995. Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. *Nature* **377**(Supp): 3-174.
- Aharoni A, Baran N, and Manor H. 1993. Characterisation of a multisubunit human protein which selectively binds single stranded d(GA)<sub>n</sub> and d(GT)<sub>n</sub> sequence repeats in DNA. *Nuc Acids Res* **21**: 522-5228.
- Aicardi J, Lefebvre J, and Lerique-Koechlin A. 1965. A new syndrome: spasm in flexion, callosal agenesis, ocular abnormalities. *Electroenceph Clin Neurophysiol* **19**: 606-612 (Abstract 18).
- Aldred MA, Flint T, Vossen R, *et al.* 1994. The EUROGEM map of human chromosome X, in Spurr NK, *et al.* European gene mapping project (EUROGEM): genetic maps based on the CEPH reference families. *Eur J Hum Genet* **2**: 193-252.
- Aldrich RA, Steinberg AG, and Campbell DC. 1954. Pedigree demonstrating a sex-linked recessive condition characterized by draining ears, eczematoid dermatitis and bloody diarrhea. *Pediatrics* **13**: 133-139.
- Altschul SF, Gish W, Miller W, Myers MW and Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* **215**: 403-410.
- Armour JAL, Neumann R, Gobert S, and Jeffreys AJ. 1994. Isolation of human simple repeat loci by hybridization selection. *Hum Mol Genet* **3**: 599-605.

Arveiler B, Alembik Y, Hanauer A, Jacobs P, Tranebjaerg L, Mikkelsen M, Puissant H, Larget Piet L, and Mandel J-L. 1988. Linkage analysis suggests at least two loci for X-linked non specific mental retardation. *Am J Med Genet* **30**: 473-483.

Aughton DJ, AlSaadi AA, Johnson JA, Transue DJ, and Trock GL. 1993. Dir Dup(X) (q13-qter) in a girl with growth retardation, microcephaly, developmental delay, seizures, and minor anomalies. *Am J Med Genet* **46**: 159-164.

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (eds). 1994. *Current Protocols in Molecular Biology*. John Wiley and Sons.

Bach I, Robinson D, Thomas N, Ropers H-H, and Cremers FPM. 1992. Physical fine mapping of genes underlying X-linked deafness and non fra(X)-X-linked mental retardation at Xq21. *Hum Genet* **89**: 620-624.

Ballabio A. 1993. The rise and fall of positional cloning? *Nature Genet* **3**: 277-279.

Ballabio A. 1995. MLS, Aicardi and Goltz Syndromes: How many genes involved? *Am J Med Genet* **59**: 100.

Ballabio A, Bardoni B, Carrozzo R, Andria G, Bick D, Campbell L, Hamel B, Ferguson-Smith MA, Gimelli G, Fraccaro M, Maraschio P, Zuffardi O, Guioli S, and Camerino G. 1989. Contiguous gene syndromes due to deletions in the distal short arm of the human X chromosome. *Proc Natl Acad Sci USA* **86**: 10001-10005.

Ballabio A, Bardoni B, Guioli S, Basler E, and Camerino G. 1990. Two families of low-copy-number repeats are interspersed on Xp22.3: Implications for the high frequency of deletions in this region. *Genomics* **8**: 263-270.

Ballabio A and Andria G. 1992. Deletions and translocations involving the distal short arm of the human X chromosome: review and hypotheses. *Hum Mol Genet* **1**: 221-227.

Banchs I, Bosch A, Guimera J, Lazaro C, Puig A, and Estivill X. 1994. New alleles at microsatellite loci in CEPH families mainly arise from somatic mutations in the lymphoblastoid cell lines. *Human Mutation* **3**: 365-372.

Bar-David S, Lerer I, Sarfaty CK, Kohan ZG, Meiner V, Zlotogora J and Abeliovich D. 1996. Localisation of two X-linked mental retardation (XLMR) genes to Xp: MRX37 gene at Xp22.31-p22.32 and a putative MRX gene on Xp22.11-p22.2. *Am J Med Genet* **64**: 83-86.

Barker DF, Dietz-Band JN, Donaldson CW, Andersen WL, Turco AE, Pole AR, Willard HF, and Vincent A. 1989. Further isolation, characterisation and physical localisation of X chromosome RFLP markers, comparing VNTR-directed and random isolation strategies. *Cytogenet Cell Genet* **51**: 958.

Beckmann JS and Weber JL. 1992. Survey of human and rat microsatellites. *Genomics* **12**: 627-631.

Beckmann JS, Tomfohrde J, Barnes RI, Williams M, Broux O, Richard I, Weissenbach J and Bowcock AM. 1993. A linkage map of human chromosome 15 with an average resolution of 2 cM and containing 55 polymorphic microsatellites. *Hum Mol Genet* **2**: 2019-2030.

Beggs AH and Kunkel LM. 1990. A polymorphic CACA repeat in the 3' untranslated region of dystrophin. *Nuc Acids Res* **18**: 1931.

Beier DR. 1993. Single-strand conformation polymorphism (SSCP) analysis as a tool for genetic mapping. *Mammalian Genome* **4**: 627-631.

Bell J and Haldane JBS. 1937. The linkage between the genes for colour-blindness and haemophilia in man. *Proc Roy Soc B* **123**: 119-150.

Belmont JW. 1996. Genetic control of X inactivation and processes leading to X-inactivation skewing. *Am J Hum Genet* **58**: 1101-1108.

Benton WD and Davis RW. 1977. Screening  $\lambda$ gt recombinant clones by hybridisation to single plaques in situ. *Science* **196**: 180-182.

Berry R, Stevens TJ, Walter NAR, Wilcox AS, Rubano T, Hopkins JA, Weber J, Goold R, Bento Soares M and Sikela JM. 1995. Gene-based sequence-tagged-sites (STSs) as the basis for a human gene map. *Nature Genet* **10**: 415-423.

Biancalana V, Serville F, Pommier J, Julien J, Hanauer A, and Mandel JL. 1992. Moderate instability of the trinucleotide repeat in spino bulbar muscular atrophy. *Hum Mol Genet* **1**: 255-258.

Billuart P, Vinet MC, des Portes V, Llense S, Richard L, Moutard ML, Recan D, Bruls T, Bienvenu T, Kahn A, Beldjord C and Chelly J. 1996. Identification by STS PCR screening of a microdeletion in Xp21.3-22.1 associated with non-specific mental retardation. *Hum Mol Genet* **5**: 977-979.

Bird H, Collins AL, Oley C and Lindsay S. 1995. Crossover analysis in a British family suggests that Coffin-Lowry Syndrome maps to a 3.4-cM interval in Xp22. *Am J Med Genet* **59**: 512-516.

Black GCM, Chen Z-Y, Craig IW and Powell JF. 1991. Dinucleotide repeat polymorphism at MAOA locus. *Nuc Acids Res* **19**: 689.

Boison D, Bussow H, D'Urso D, Muller H-W, and Stoffel W. 1995. Adhesive properties of proteolipid protein are responsible for the compaction of CNS myelin sheaths. *J Neurosci* **15**: 5502-5513.

Bonneau D, Rozet J-M, Bulteau C, Berthier M, Mettey R, Gil R, Munnich A, and Le Merrer M. 1993. X linked spastic paraplegia (SPG2): clinical heterogeneity at a single locus. *J Med Genet* **30**: 381-384.

Borjeson M, Forssman H, Lehmann O. 1962. An X-linked, recessively inherited syndrome characterized by grave mental deficiency, epilepsy and endocrine disorder. *Acta Med Scand* **171**: 13-21.

Botstein D, White RL, Skolnick M, and Davis RW. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* **32**: 314-331.

Bouloux P-M G, Hardelin J-P, Munroe P, Kirk JMW, Legouis J, Levilliers J, Hazan J, Weissenbach J, and Petit C. 1991. A dinucleotide repeat polymorphism at the Kallmann locus (Xp22.3). *Nuc Acids Res* **19**: 5453.

- Brandiff B, Gordon L, and Trask B. 1991. A new system for high-resolution DNA sequence mapping in interphase pronuclei. *Genomics* **10**: 75-82.
- Brennan MB and Hochgeschwender U. 1995. So many needles, so much hay. *Hum Mol Genet* **4**: 153-156.
- Browne DL, Luty JA, and Litt M. 1991. Dinucleotide repeat polymorphism at the DXS538 locus. *Nuc Acids Res* **19**: 1161.
- Browne D, Barker D, and Litt M. 1992a. Dinucleotide repeat polymorphisms at the DXS365, DXS443 and DXS451 loci. *Hum Mol Genet* **1**: 213.
- Browne DL, Zoanna J, and Litt M. 1992b. Dinucleotide repeat polymorphism at the PGK1P1 locus. *Nuc Acids Res* **20**: 1169.
- Buckle VJ and Kearney L. 1994. New methods in cytogenetics. *Current Opinion in Genetics and Development* **4**: 374-382.
- Buckler AJ, Chang DD, Graw SL, Brook JD, Haber DA, Sharp PA and Housman DE. 1991. Exon amplification: a strategy to isolate mammalian genes based on RNA splicing. *Proc Natl Acad Sci USA* **88**, 4005-4009.
- Buetow KH. 1991. Influence of aberrant observations on high-resolution linkage analysis outcomes. *Am J Hum Genet* **49**: 985-994.
- Buetow KH and Chakravarti A. 1987. Multipoint gene mapping using seriation. II. Analysis of simulated and empirical data. *Am J Hum Genet* **41**: 189-201.
- Buetow KH, Duggan D, Yang B, Ludwigsen S, Puck J, Porter J, Budarf M, Spielman R, and Emanuel BS. 1993. A microsatellite-based multipoint index map of human chromosome 22. *Genomics* **18**: 329-339.
- Buetow KH, Weber JL, Ludwigsen S, Scherpbier-Heddema T, Duyk G, Sheffield VC, Wang Z, and Murray JC. 1994. Integrated human genome-wide maps constructed using the CEPH reference panel. *Nature Genet* **6**: 391-393.

- Burke DT, Carle G, and Olson MV. 1987. Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* **236**: 806-812.
- Burn TC, Connors TD, Klinger KW, and Landes GM. 1995. Increased exon-trapping efficiency through modifications to the pSPL3 splicing vector. *Gene* **161**: 183-187.
- Callen DF, Baker E, Eyre HJ, Chernos JE, Bell JA, and Sutherland GR. 1990. Reassessment of two apparent deletions of chromosome 16p to an ins(11;16) and a t(1;16) by chromosome painting. *Ann Genet* **33**: 219-221.
- Callen DF, Thompson AD, Shen Y, Phillips HA, Richards RI, Mulley, JC, and Sutherland GR. 1993. Incidence and origin of "null" alleles in the (AC)<sub>n</sub> microsatellite markers. *Am J Hum Genet* **52**: 922-927.
- Chang Y-PC, Smith KD, and Dover GJ. 1994. Dinucleotide repeat polymorphisms at the DXS85, DXS16 and DXS43 loci. *Hum Mol Genet* **3**: 1029.
- Chen E, d'Urso M, and Schlessinger D. 1994. Functional mapping of the human genome by cDNA localization *versus* sequencing. *Bioessays* **16**: 693-698.
- Chen JD, Hejtmancik JF, Romeo G, Lindlof M, Boehm C, Caskey CT, Kress W, Fischbeck KH, Dreier M, Serravalle S, Grimm T, Kaariainen H, Ferrari M, Pfender E, Meng G, De La Chapelle A, Melis MA, Muller B, MacKinlay AG, Muller CR, and Denton MJ. 1989. A genetic linkage map of five marker loci in and around the Duchenne muscular dystrophy locus. *Genomics* **4**: 105-109.
- Chomczynski P and Sacchi N. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* **162**: 156-159.
- Choo KH, Brown R, Webb G, Craig IW, and Filby RG. 1987. Genomic organisation of human centromeric alpha satellite DNA: characterization of a chromosome 17 alpha satellite sequence. *DNA* **6**: 297-305.
- Chotai J. 1984. On the lod score method in linkage analysis. *Ann Hum Genet* **48**: 359-378.

Chumakov IM, Le Gall I, Billault A, Ougen P, Soularue P, Guillou S, Rigault P, Bui H, De Tand M-F, Barillot E, Abderrahim H, Cherif D, Berger R, Le Paslier D and Cohen D. 1992. Isolation of chromosome 21-specific yeast artificial chromosomes from a total human genome library. *Nature Genet* **1**: 222-225.

Chung CT, Niemela SL, and Miller RH. 1989. One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. *Proc Natl Acad Sci USA* **86**: 2172-2175.

Church DM, Stotler CJ, Rutter JL, Murrel JR, Trofatter JA, and Buckler AJ. 1994. Isolation of genes from complex sources of mammalian genomic DNA using exon amplification. *Nature Genet* **6**: 98-104.

Church GM and Kieffer-Higgins S. 1988. Multiplex DNA sequencing. *Science* **240**: 185-188.

Claes S, Gu XX, Legius E, Lorenzetti E, Marynen P, Fryns JP, Cassiman JJ and Raeymaekers P. 1996. Linkage analysis in three families with nonspecific X-linked mental retardation. *Am J Med Genet* **64**: 137-146.

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

Coleman MP, Murray JC, Willard HF, Nolan KF, Reid KBM, Blake DJ, Lindsay S, Bhattacharya SS, Wright A, and Davies KE. 1991. Genetic and physical mapping around the properdin P gene. *Genomics* **11**: 991-996.

Coleman MP, Nemeth AH, Campbell L, Raut CP, Weissenbach J, and Davies KE. 1994. A 1.8-Mb YAC contig in Xp11.23: Identification of CpG islands and physical mapping of CA repeats in a region of high gene density. *Genomics* **21**: 337-343.

Colleaux L, May M, Belougne J, Lepaslier D, Schwartz C and Fontes M. 1996. Localisation of two candidate genes for mental retardation using a YAC physical map of the Xq21.1-21.2 subbands. *J Med Genet* **33**: 353-357.

Collier S, Tassabehji M, and Strachan TA. 1993. *De novo* pathological point mutation at the 21-hydroxylase locus: implications for gene conversion in the human genome. *Nature Genet* **3**: 260-265.

Collins FS. 1992. Positional cloning: let's not call it reverse anymore. *Nature Genet* **1**: 3-6.

Collins FS. 1995. Positional cloning moves from perditional to traditional. *Nature Genet* **9**: 347-350.

Collins F and Galas D. 1993. A new five-year plan for the U.S. Human Genome Project. *Science* **262**: 43-46.

Conley ME, Wang WC, Parolini O, Shapiro DN, Campana D, and Siminovitch KA. 1992. Atypical Wiskott-Aldrich syndrome in a girl. *Blood* **80**: 1264-1269.

Conneally PM and Rivas ML. 1980. Linkage analysis in man, in *Advances in Human Genetics*, vol **10**, edited by Harris H and Hirschhorn K, New York, Plenum Press, pp 209-266.

Conneally PM, Edwards JH, Kidd KK, Lalouel J-M, Morton NE, Ott J, and White R. 1985. Report of the committee on methods of linkage analysis and reporting. *Cytogenet Cell Genet* **40**: 356-359.

Cooper DN, Smith BA, Cooke HJ, Niemann S, and Schmidtke J. 1985. An estimate of unique DNA sequence heterozygosity in the human genome. *Hum Genet* **69**: 201-205.

Corbo L, Maley JA, Nelson DL and Caskey CT. 1990. Direct cloning of human transcripts with hnRNA from hybrid cell lines. *Science* **249**: 652-655.

Cory GOC, MacCarthy-Morrogh L, Banin S, Gout I, Brickell PM, Levinsky RJ, Kinnon C, and Lovering RC. 1996. Evidence that the Wiskott-Aldrich syndrome protein may be involved in lymphoid cell signaling pathways. *J Immunol* **157**: 3791-3795.

Cox DR, Burmeister M, Price ER, Kim S, and Meyers RM. 1990. Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. *Science* **250**: 245-250.

Cox SA, Attwood J, Bryant SP, *et al.* 1996. European Gene Mapping Project (EUROGEM): Breakpoint panels for human chromosomes based on the CEPH reference families. *Ann Hum Genet* **60**: 447-486.

Cox TC, Kozman HM, Raskind WH, May BA, and Mulley, JC. 1992. Identification of a highly polymorphic marker within intron 7 of the ALAS2 gene and suggestion of at least two loci for X-linked sideroblastic anemia. *Hum Mol Genet* **1**: 639-641.

Dausset J, Cann H, Cohen D, Lathrop M, Lalouel JM, and White R. 1990. Centre d'Etude du Polymorphisme Humain (CEPH): Collaborative genetic mapping of the human genome. *Genomics* **6**: 575-577.

Davies JL, Kawaguchi Y, Bennett ST, Copeman JB, Cordell HJ, Pritchard LE, Reed PW, Gough SCL, Jenkins SC, Palmer SM, Balfour KM, Rowe BR, Farrall M, Barnett AH, Bain SC and Todd JA. 1994. A genome-wide search for human type1 diabetes susceptibility genes. *Nature* **371**: 130-136.

Davies K. 1993. The EST express gathers speed. *Nature* **364**: 554.

Dear PH and Cook PR. 1993. Happy mapping: linkage mapping using a physical analogue of meiosis. *Nuc Acids Res* **21**: 13-20.

Deka R, Shriver MD, Yu LM, Jin L, Aston CE, Chakraborty R, and Ferrell RE. 1994. Conservation of human chromosome 13 polymorphic microsatellite (CA)<sub>n</sub> repeats in chimpanzees. *Genomics* **22**: 226-230. 1994

De Kok YJM, van der Maarel SM, Bitner-Glindzicz M, Huber I, Monaco AP, Malcolm S, Pembrey ME, Ropers H-H, Cremers FPM. 1995. Association between X-linked mixed deafness and mutations in the POU domain gene *POU3F4*. *Science* **267**: 685-688.

Den Dunnen JT, Grootsholten PM, Bakker E, Blonden LAJ, Ginjaar HB, Wapenaar MC, van Paassen HMB *et al.*, 1989. Topography of the Duchenne muscular dystrophy (DMD) gene: FIGE and cDNA analysis of 194 cases reveals 115 deletions and 13 duplications. *Am J Hum Genet* **45**: 835-847.

- Dennis NR, Collins AL, Crolla JA, Cockwell AE, Fisher AM, and Jacobs PA. 1993. Three patients with ring (X) chromosomes and a severe phenotype. *J Med Genet* **30**: 482-486.
- Derry JMJ, Ochs HD, and Franke U. 1994. Isolation of a novel gene mutated in Wiskott-Aldrich syndrome. *Cell* **78**: 635-644 and erratum *Cell* **79**: 923.
- Derry JMJ, Kerns JA, Weinberg KI, Ochs HD, Volpini V, Estivill X, Walker AP, and Franke U. 1995a. WASP gene mutations in Wiskott-Aldrich syndrome and X-linked thrombocytopenia. *Hum Mol Genet* **4**: 1127-1135.
- Derry JMJ, Wiedemann P, Blair P, Wang Y, Kerns JA, Lemahieu V, Godfrey VL, Wilkinson JE, and Franke U. 1995b. The mouse homolog of the Wiskott-Aldrich syndrome protein (WASP) gene is highly conserved and maps near the scurfy (sf) mutation on the X chromosome. *Genomics* **29**: 471-477.
- De Saint Basile G, Lagelouse RD, Lambert N, Schwarz K, Le Mareck B, Odent S, Schlegel N, Fischer A. 1996. Isolated X-linked thrombocytopenia in two unrelated families is associated with point mutations in the Wiskott-Aldrich syndrome protein gene. *J Pediatrics* **129**: 56-62.
- De Sario A, Geigl E-M, Palmieri G, D'Urso M, and Bernadi G. 1996. A compositional map of human chromosome band Xq28. *Proc Natl Acad Sci USA* **93**: 1298-1302.
- De Weers M, Mensink RGJ, Kenter M, and Schuurman RKB. 1992. Three dinucleotide repeat polymorphisms at the DXS178 locus. *Hum Mol Genet* **1**: 653.
- Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, Marc S, Hazan J, Seboun E, Lathrop M, Gyapay G, Morissette J and Weissenbach J. 1996. A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* **380**: 152-154.
- Diehl H-J, Schaich M, Budzinski R-M and Stoffel W. 1986. Individual exons encode the integral membrane domains of human myelin proteolipid protein. *Proc Natl Acad Sci USA* **83**, 9807-9811.

Dietz-Band JN, Turco AE, Willard HF, Vincent A, Skolnick MH, and Barker DF. 1990. Isolation, characterisation, and physical localisation of 33 human X-chromosome RFLP markers. *Cytogenet Cell Genet* **54**: 137-141.

Donis-Keller H, Green P, Helms C, Cartinhour S, Weiffenbach B, Stephens K, Keith TP, Bowden DW, Smith DR, Lander ES, Botstein D, Akots G, Rediker KS, Gravius T, Brown VA, Rising MB, Parker C, Powers JA, Watt DE, Kauffmann ER, Bricker A, Phipps P, Muller-Kahle H, Fulton TR, Ng S, Schumm JW, Braman JC, Knowlton RG, Barker DF, Crooks SM, Lincoln SE, Daly MJ, and Abrahamson J. 1987. A genetic linkage map of the human genome. *Cell* **51**: 319-337.

Donnelly AJ, Choo KHA, Kozman HM, Gedeon AK, Danks DM, and Mulley JC. 1994a. Regional localisation of a non-specific X-linked mental retardation gene (*MRX19*) to Xp22. *Am J Med Genet* **51**: 581-585.

Donnelly A, Kozman H, Gedeon AK, Webb S, Lynch M, Sutherland G R, Richards RI, and Mulley JC. 1994b. A linkage map of microsatellite markers on the human X chromosome. *Genomics* **20**: 363-370.

Donnelly A, Colley A, Crimmins D, and Mulley J. 1996a. A novel mutation in exon 6 (F236S) of the proteolipid protein gene is associated with spastic paraplegia. *Human Mutation* **8**: 384-385.

Donnelly AJ, Partington MW, Ryan AK and Mulley JC. 1996b. Regional localisation of two non-specific X-linked mental retardation genes (*MRX30* and *MRX31*). *Am J Med Genet* **64**: 113-120.

Donnelly A, Haan E, Manson J, and Mulley J. 1997. A novel mutation in exon b (R259C) of the *MTM1* gene is associated with myotubular myopathy. (*Human Mutation* accepted for publication).

Dracopoli NC, Stanger BZ, Ito CY, Call KM, Lincoln SE, Lander ES, and Housman DE. 1988. A genetic linkage map of 27 loci from PND to FY on the short arm of human chromosome 1. *Am J Hum Genet* **43**: 462-470.

- Dracopoli NC, Haines JL, Korf BR, Moir DT, Morton CC, Seidman CE, Seidman JG, Smith DR. (eds). 1994. *Current Protocols in Human Genetics*. John Wiley and Sons.
- Drayna D, Davies K, Hartley D, Mandel J-L, Camerino G, Williamson R, and White R. 1984. Genetic mapping of the human X chromosome by using restriction fragment length polymorphisms. *Proc Natl Acad Sci USA* **81**: 2836-2839.
- Drayna D and White R. 1985. The genetic linkage map of the human X chromosome. *Science* **230**: 753-758.
- Dronamraju KR. 1987. The origins of human gene mapping. *Genomics* **1**: 270-276.
- Duncan ID. 1990. Dissection of the phenotype and genotype of the X-linked myelin mutants. *Ann N.Y. Acad Sci* **605**: 110-12.
- Durkin AS, Nierman WC, Zoghbi H, Jones C, Kozak CA, and Maglott DR. 1994. Chromosome assignment of human brain expressed sequence tags (ESTs) by analysing fluorescently labeled PCR products from hybrid cell panels. *Cytogenet Cell Genet* **65**: 86-91.
- Duyk GM, Kim S, Myers RM and Cox DR. 1990. Exon trapping: a genetic screen to identify candidate transcribed sequences in cloned mammalian genomic DNA. *Proc Natl Acad Sci USA* **87**: 8895-8999.
- Economou EP, Bergen AW, and Antonarakis SE. 1989. Novel DNA polymorphic system: variable poly A tract 3' to Alu1 repetitive elements. *Am J Hum Genet* **45**: A138.
- Edwards AWF. 1989. Probability and likelihood in genetic counselling. *Clinical Genetics* **36**: 209-216.
- Edwards A, Civitello A, Hammond HA, and Caskey CT. 1991. DNA typing and genetic mapping with trimeric and tetrameric repeats. *Am J Hum Genet* **49**: 746-756.
- Elsner TI, Albertsen H, Gerken SC, Cartwright P, and White R. 1995. Breakpoint analysis: precise localization of genetic markers by means of nonstatistical computation using relatively few genotypes. *Am J Hum Genet* **56**: 500-507.

Elston RC and Stewart J. 1971. A general method for the genetic analysis of pedigree data. *Hum Hered* **21**: 523-542.

Elston RC and Lange K. 1975. The prior probability of autosomal linkage. *Ann Hum Genet* **38**: 341-350.

Eng LF, Chao FC, Gerstl B, Pratt D, Tavaststjerna MJ. 1968. The maturation of human white matter myelin: fractionation of the myelin membrane proteins. *Biochemistry* **7**: 4455-4465.

Epplen JT, Buitkamp J, Bocker T and Epplen C. 1995. Indirect gene diagnoses for complex (multifactorial) diseases—a review. *Gene* **159**: 49-55.

Fain PR, Wright E, Willard HF, Stephens K, and Barker DF. 1989. The order of loci in the pericentric region of chromosome 17, based on evidence from physical and genetic breakpoints. *Am J Hum Genet* **44**: 68-72.

Fain PR, Kort EN, Chance PF, Nguyen K, Redd DF, Econs MJ and Barker DF. 1995. A 2D crossover-based map of the human X chromosome as a model for map integration. *Nature Genet* **9**: 261-266.

Featherston C. 1997. The many faces of WAS protein. *Science* **275**: 27-28.

Feener CA, Boyce FM, and Kunkel LM. 1991. Rapid detection of CA polymorphisms in cloned DNA: application to the 5' region of the Dystrophin gene. *Am J Hum Genet* **48**: 621-627.

Ferrero GB, Franco B, Roth EJ, Firulli BA, Borsani G, Delmas-Mata J, Weissenbach J, Halley G, Schlessinger D, Chinault AC, Zoghbi HY, Nelson DL, and Ballabio A. 1995. An integrated physical and genetic map of a 35 Mb region on chromosome Xp22.3-Xp21.3. *Hum Mol Genet* **4**: 1821-1827.

Fichant GA. 1992. Constraints acting on the exon positions of the splice site sequences and local amino acid composition of the protein. *Hum Mol Genet* **1**: 259-267.

Fields C, Adams MD, White O, and Venter JC. 1994. How many genes in the human genome? *Nature Genet* **7**: 345-346.

- Freije D, Helms C, Watson MS, and Donis-Keller H. 1992. Identification of a second pseudoautosomal region near the Xq and Yq telomeres. *Science* **258**: 1784-1787.
- Freimer NB, Sandkuijl LA, and Blower SM. 1993. Incorrect specification of marker allele frequencies: effects on linkage analysis. *Am J Hum Genet* **52**: 1102-1110.
- Gebbia M, Lind TJ, Ketterling RP, Miller C, Nelson DL, Sommer SS, and Casey B. 1996. A male with isolated factor IX deficiency harbors a deletion spanning more than 7 Mb in Xq26-27. *Am J Hum Genet* **49**: A259 (Abstract 1497).
- Gecz J, Pollard H, Consalez G, Villard L, Stayton C, Millasseau P, Khrestchatisky M, and Fontes M. 1994. Cloning and expression of the murine homologue of a putative human X-linked nuclear protein gene closely linked to PGK1 in Xq13.3.
- Gecz J, Gedeon AK, Sutherland GR, and Mulley JC. 1996. Identification of FMR2: a gene associated with FRAXE mental retardation. *Nature Genet* **13**:105-108.
- Gedeon A, Kerr B, Mulley J, and Turner G. 1991. Localization of the MRX3 gene for nonspecific X-linked mental retardation. *J Med Genet* **28**: 372-377.
- Gedeon AK, Holman K, Richards RI and Mulley J. 1992. Characterisation of new PCR based markers for mapping and diagnosis: AC dinucleotide repeat markers at the DXS237 (GMGX9) and DXS102 (cX38.1) loci. *Am J Med Genet* **43**: 255-260.
- Gedeon AK, Kerr B, Mulley J, and Turner G. 1994a. Pericentromeric genes for non-specific X-linked mental retardation (MRX). *Am J Med Genet* **51**: 553-564.
- Gedeon AK, Partington M, and Mulley J. 1994b. X-linked mental retardation with dystonic movement of the hands (PRTS): revisited. *Am J Med Genet* **51**: 565-568.
- Gedeon AK, Glass IA, Connar JM, Mulley JC. 1996a. Genetic localisation of MRX27 to Xq24-26 defines another discrete gene for non-specific mental retardation. *Am J Med Genet* **64**: 121-124.
- Gedeon AK, Donnelly A, Mulley JC, Kerr B, Turner G. 1996b. How many X-linked genes for non-specific mental retardation (MRX) are there? *Am J Med Genet* **64**: 158-162.

Gedeon AK, Kozman HM, Robinson H, Pilia G, Schlessinger D, Turner G and Mulley JC. 1996c. Refinement of the background genetic map of Xq26-q27 and gene localisation for Borjeson-Forssman-Lehmann Syndrome. *Am J Med Genet* **64**: 63-68.

Gedeon A, Mulley J and Haan E. 1996d. Gene localisation for Sutherland-Haan syndrome (SHS:MIM 309470). *Am J Med Genet* **64**: 78-79.

Gedeon A, Mulley J and Turner G. 1996e. Gene localisation for Wilson-Turner syndrome (WTS:MIM 309585). *Am J Med Genet* **64**: 80-81.

Gencic S and Hudson LD. 1990. Conservative amino acid substitution in the myelin proteolipid protein of jimpy<sup>msd</sup> mice. *J Neuroscience* **10**, 117-124.

Gendrot C, Ronce R, Toutain A, Moizard M-P, Muh J-P, Raynaud M, Dourlens J, Briault S, and Moraine C. 1994. X-linked mental retardation exhibiting linkage to DXS255 and PGK1P1: a new MRX family (MRX14) with localisation in the pericentromeric region. *Clin Genet* **45**: 145-153.

Gerhold D and Caskey CT. 1996. It's the genes! EST access to human genome content. *BioEssays* **18**: 973-981.

Gibbons RJ, Suthers GK, Wilkie AOM, Buckle VJ, and Higgs DR. 1992. X-linked  $\alpha$ -thalassemia/mental retardation (ATR-X) syndrome: localization to Xq12-q21.31 by X inactivation and linkage analysis. *Am J Hum Genet* **51**: 1136-1149.

Gibbons RJ, Picketts DJ, and Higgs DR. 1995. Syndromal mental retardation due to mutations in a regulator of gene expression. *Hum Mol Genet* **4**: 1705-1709.

Gish W and States DJ. 1993. Identification of protein coding regions by database similarity search. *Nature Genet* **3**: 266-272.

Glass IA. 1991. X linked mental retardation. *J Med Genet* **28**: 361-371.

Goda SK and Minton NP. 1995. A simple procedure for gel electrophoresis and Northern blotting of RNA. *Nuc Acids Res* **23**: 3357-3358.

- Goldblatt J, Ballo R, Sachs B, and Moosa A. 1989. X-linked spastic paraplegia: evidence for homogeneity with a variable phenotype. *Clin Genet* **35**: 116-120.
- Gow A, Friedrich VL and Lazzarini RA. 1994. Many naturally occurring mutations of myelin proteolipid protein impair its intracellular transport. *J Neurosci Res* **37**: 574-583.
- Green P, Falls K, and Crooks S. 1990. Documentation for CRI-MAP, version 2.4 (3/26/90).
- Greer WL, Shehabeldin A, Schulman J, Junker A, and Siminovitch KA. 1996. Identification of WASP mutations, mutation hotspots and genotype-phenotype disparities in 24 patients with the Wiskott-Aldrich syndrome. *Hum Genet* **98**: 685-690.
- Gregg RG, Palmer C, Kirkpatrick S and Simantel A. 1996. Localisation of a non-specific X-linked mental retardation gene, MRX23, to Xq23-q24. *Hum Mol Genet* **5**: 411-414.
- Grompe M. 1993. The rapid detection of unknown mutations in nucleic acids. *Nature Genet* **5**: 111-117.
- Gross DS and Garrard WT. 1986. The ubiquitous potential Z-forming sequence of eukaryotes, (dT-dG)<sub>n</sub>-(dA-dC)<sub>n</sub>, is not detectable in the genomes of eubacteria, archaebacteria, or mitochondria. *Mol Cell Biol* **6**: 3010-3013.
- Gu XX, Decorte R, Marynen P, Fryns J-P, Cassiman J-J, Raeymaekers P. 1996. Localisation of a new gene for non-specific mental retardation to Xq22-q26 (MRX35). *J Med Genet* **33**: 52-55.
- Gu Y, Shen Y, Gibbs RA, Nelson DL. 1996. Identification of *FMR2*, a novel gene associated with *FRA3E* CCG repeat and CpG island. *Nature Genet* **13**:109-113.
- Gustavson K-H, Anneren G, Malmgren H, Dahl N, Ljunggren C-G, Backman H. 1993. New X-linked syndrome with severe mental retardation, severely impaired vision, severe hearing defect, epileptic seizures, spasticity, restricted joint mobility, and early death. *Am J Med Genet* **45**: 654-658.
- Guyer MS and Collins FS. 1993. The Human Genome Project and the future of medicine. *Am J Dis Child* **147**: 1145-1152.

- Guyer MS and Collins FS. 1995. How is the Human Genome Project doing, and what have we learned so far? *Proc Natl Acad Sci USA* **92**: 10841-10848.
- Gyapay G, Morissette J, Vignal A, Dob C, Fizames C, Millasseau P, Marc S, Bernadi G, Lathrop M, and Weissenbach J. 1994. The 1993-1994 Genethon human genetic linkage map. *Nature Genet* **7**: 246-339.
- Haaf T and Ward DC. 1994. Structural analysis of a-satellite DNA and centromere proteins using extended chromatin and chromosomes. *Hum Mol Genet* **3**: 697-709.
- Haldane JBS and Smith CAB. 1947. A new estimate of the linkage between the genes for colour-blindness and haemophilia in man. *Ann Eugen* **14**: 10-31.
- Haliassos A, Chomel JC, Tesson L, Baudis M, Kruh J, Kaplan JC and Kitzis A. 1989. Modification of enzymically amplified DNA for the detection of point mutations. *Nuc Acids Res* **17**: 3606.
- Hamada H and Kakunga T. 1982. Potential Z-DNA forming sequences are highly dispersed in the human genome. *Nature* **298**: 396-398.
- Hamada H, Petrino MG, and Kakunga T. 1982. A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. *Proc Natl Acad Sci USA* **79**: 6465-6469.
- Hamada H, Petrino MG, Kakunga T, Seidman M, and Stollar BD. 1984. Characterization of genomic poly(dT-dG).poly(dC-dA) sequences: structure, organization and conformation. *Mol Cell Biol* **4**: 2610-2621.
- Hamaguchi M, Sakamoto H, Tsuruta H, Sasaki H, Muto T, Sugimura T and Terada M. 1992. Establishment of a highly sensitive and specific exon-trapping system. *Proc Natl Acad Sci USA* **89**: 9779-9783.
- Hamel BCJ, Kremer H, Wesby-van Swaay E, van den Helm B, Smits APT, Oostra BA, Ropers H-H and Mariman ECM. 1996. A gene for nonspecific X-linked mental retardation (MRX41) is located in the distal segment of Xq28. *Am J Med Genet* **64**: 131-133.

Hane B, Schroer RJ, Howard-Peebles PK, Arena JF, Lubs HA, Stevenson RE, Schwartz CE. 1995. Non-specific X-linked mental retardation: linkage analysis of two families (MRX29 and MRX32) and review of MRX entities. Seventh International Workshop on the Fragile X and X-linked Mental Retardation, Tromso.

Hane B, Schroer RJ, Arena JF, Lubs HA, Schwartz CE and Stevenson RE. 1996. Nonsyndromic X-linked mental retardation: review and mapping of MRX29 to Xp21. *Clin Genet* **50**: 176-183.

Harris A, Collins J, Vetrie D, Cole C, and Bobrow M. 1992. X inactivation as a mechanism of selection against lethal alleles: further investigation of incontinentia pigmenti and X linked lymphoproliferative disease. *J Med Genet* **29**: 608-614.

Hatchwell E, Robinson D, Crolla JA, and Cockwell AE. 1996. X inactivation analysis in a female with hypomelanosis of Ito associated with a balanced X;17 translocation: evidence for functional disomy of Xp. *J Med Genet* **33**: 216-220.

Hauge XY and Litt M. 1993. A study of the origin of 'shadow bands' seen when typing dinucleotide repeat polymorphisms by the PCR. *Hum Mol Genet* **2**: 411-415.

Hawkins JD. 1988. A survey on intron and exon lengths. *Nuc Acids Res* **16**: 9893-9908.

Hayashi K and Yandell DW. 1993. How sensitive is PCR-SSCP? *Human Mutation* **2**: 338-346.

Hazan J, Dubay C, Pankowiak MP, Becuwe N, and Weissenbach J. 1992. A genetic linkage map of human chromosome 20 composed entirely of microsatellite markers. *Genomics* **12**: 183-189.

Heckmatt JZ, Sewry CA, Hodes D, and Dubowitz V. 1985. Congenital centronuclear (myotubular) myopathy: a clinical, pathological and genetic study in eight children. *Brain* **108**: 941-964.

Hendrickx J, Coucke P, Bossuyt P, Wauters J, Raeymaekers, Marchau F, Smit GPA, Stolte I, Sardharwalla IB, Berthelot J, van den Bergh I, Berger R, van Broeckhoven C,

- Baussan C, Wapenaar M, Fernandes J, and Willems PJ. 1993. X-linked liver glycogenosis: localization and isolation of a candidate gene. *Hum Mol Genet* **2**: 583-589.
- Herbst DS and Miller JR. 1980. Non-specific X-linked mental retardation II : the frequency in British Columbia. *Am J Med Genet* **7**: 461-469.
- Herrell S, Novo FJ, Charlton R, and Affara NA. 1995. Development and physical analysis of YAC contigs covering 7 Mb of Xp22.3-p22.2 *Genomics* **25**: 526-537.
- Heslop-Harrison JS, Leitch AR, Schwarzacher T, Smith JB, Atkinson MD, and Bennett MD. 1989. The volumes and morphology of human chromosomes in mitotic reconstructions. *Hum Genet* **84**: 27-34.
- Hicks AA, Johnson KJ, Barnard EA, and Darlison MG. 1991. Dinucleotide repeat polymorphism in the human X-linked GABAA receptor  $\alpha 3$ -subunit gene. *Nuc Acids Res* **19**: 4016.
- Hillier L, Lennon G, Becker M *et al.* 1996. Generation and analysis of 280,000 human expressed sequence tags. *Genome Research* **6**: 807-828.
- Hochgeschwender U. 1992. Toward a transcriptional map of the human genome. *Trends Genet* **8**: 41-44.
- Hochgeschwender U and Brennan MB. 1991. Identifying genes within the genome: new ways for finding the needle in a haystack. *BioEssays* **13**: 139-144.
- Hodes ME and Dlouhy SR. 1995. The changing phenotype of Pelizaeus-Merzbacher disease. *Am J Hum Genet* **57**: A92.
- Hofker MH, Skraastad MI, Bergen AA, Wapenaar MC, Bakker E, Millington-Ward A, van Ommen GJ, and Pearson PL. 1986. The X chromosome shows less genetic variation at restriction sites than the autosomes. *Am J Hum Genet* **39**: 438-451.
- Holinski-Feder E, Golla A, Rost I, Seidel H, Rittinger O, Meindl A. 1996. Regional localisation of two MRX genes to Xq28 (MRX28) and to Xp11.4-Xp22.12 (MRX33). *Am J Med Genet* **64**: 125-130.

- Howard-Peebles PN and Roberts SF. 1984. X-linked mental retardation revisited. *Am J Med Genet* **17**: 95-99.
- Hu L-J, Blumenfeld-Heyberger S, Hanauer A, Weissenbach J, and Mandel J-L. 1994. Non-specific X-linked mental retardation: linkage analysis in MRX2 and MRX4 families revisited. *Am J Med Genet* **51**: 569-574.
- Huang TH-M, Hejtmancik JF, Edwards A, Pettigrew AL, Herrera CA, Hammond HA, Caskey CT, Zoghbi HY, and Ledbetter DH. 1991. Linkage of the gene for an X-linked mental retardation disorder to a hypervariable (AGAT)<sub>n</sub> repeat motif within the human hypoxanthine phosphoribosyltransferase (HPRT) locus (Xq26). *Am J Hum Genet* **49**: 1312-1319.
- Hudson TJ, Engelstein M, Lee MK, Ho EC, Rubenfield MJ, Adams CP, Housman DE, and Dracopoli NC. 1992. Isolation and chromosomal assignment of 100 highly informative human simple sequence repeat polymorphisms. *Genomics* **13**: 622-629.
- Hudson TJ, Stein LD, Gerety SS, *et al.* 1995. An STS-based map of the human genome. *Science* **270**: 1945-1954.
- Hunter AGW, Partington MW, and Evans JA. 1982. The Coffin-Lowry syndrome. Experience from four centres. *Clin Genet* **21**: 321-335.
- Hutchinson GB and Hayden MR. 1992. The prediction of exons through an analysis of spliceable open reading frames. *Nuc Acids Res* **20**: 3453-3462.
- Hutton M, Crook R, Shaw D and Hardy J. 1995. Exon trapping from a yeast artificial chromosome (YAC) that contains the *c-fos* gene. *Focus* **17**: 24-27.
- ISCN. 1995. An International System for Human Cytogenetic Nomenclature, Mitelman F (ed); S. Karger, Basel, 1995.
- Jacobson DP, Schmeling P, and Sommer SS. 1993. Characterization of the patterns of polymorphism in a "cryptic repeat" reveals a novel type of hypervariable sequence. *Am J Hum Genet* **53**: 443-450.

Jagiello G, Ducayen M, Fang J-S, and Graffeo J. 1976. Cytogenetic observations in mammalian oocytes. *Chromosomes Today* **5**: 43-63.

Jani MM, Torchia BS, Pai GS and Migeon BR. 1995. Molecular characterisation of tiny ring X chromosomes from females with functional X chromosome disomy and lack of *cis* X inactivation. *Genomics* **27**: 182-188.

Jedele KB, Michels VV, Wagner KV, Schaid DJ, and Thibodeau SN. 1990. Linkage analysis demonstrating association of DXYS1 and non-specific X-linked mental retardation. *Am J Hum Genet* **47**: A62.

Jeffreys AJ, Wilson V, and Thein SL. 1985. Hypervariable 'minisatellite' regions in human DNA. *Nature* **314**: 67-73.

Jeffreys AJ and Pena SDJ. 1993. Brief introduction to human DNA fingerprinting in *DNA Fingerprinting: State of the Science*, pp1-20. Pena SDJ, Chakraborty R, Epplen JT and Jeffreys AJ (eds). Birkhauser Verlag Basel, Switzerland.

Jin L, Zhong Y, and Chakraborty R. 1994. The exact numbers of possible microsatellite motifs. *Am J Hum Genet* **55**: 582-583.

Johnson CL, Charmely P, Yen PH, and Shapiro LJ. 1991. A multipoint linkage map of the distal short arm of the human X chromosome. *Am J Hum Genet* **49**: 261-266.

Jones KW, Chevrette M, Shapero MH and Fournier REK. 1992. Generation of region- and species-specific expressed gene probes from somatic cell hybrids. *Nature Genet* **1**: 278-283.

Jordan E. 1992. The human genome project: where did it come from, where is it going? *Am J Hum Genet* **51**: 1-6.

Jouet M, Rosenthal A, Armstrong G, MacFarlane J, Stevenson R, Paterson J, Metzberg A, Ionasescu V, Temple K, and Kenwrick S. 1994. X-linked spastic paraplegia (SPG1), MASA syndrome and X-linked hydrocephalus result from mutations in the *L1* gene. *Nature Genet* **7**: 402-407.

Kauppinen R, Mustajoki S, Pihlaja H, Peltonen, and Mustajoki P. 1995. Acute intermittent porphyria in Finland: 19 mutations in the porphobilinogen deaminase gene. *Hum Mol Genet* **4**: 215-222.

Kaye EM, Doll RF, Natowicz MR and Smith FI. 1994. Pelizaeus-Merzbacher disease presenting as spinal muscular atrophy: clinical and molecular studies. *Ann Neurol* **36**: 916-919.

Keats BJB, Sherman SL, Morton NE, Robson EB, Buetow KH, Cann HM, Cartwright PE, Chakravarti A, Francke U, and Green PP. 1991. Guidelines for human linkage maps: an international system for human linkage maps (ISLM, 1990). *Genomics* **9**: 557-560.

Keats B, Ott J, and Conneally M. 1989. Report of the committee on linkage and gene order. *Cytogenet Cell Genet* **51**: 459-502.

Kendall E, Sargent CA and Campbell RD. 1990. Human major histocompatibility complex contains a new cluster of genes between the HLA-D and complement C4 loci. *Nuc Acids Res* **18**: 7251-7257.

Kenwrick S, Ionasescu V, Ionasescu G, Searby Ch, King A, Dubowitz M, and Davies KE. 1986. Linkage studies of X-linked recessive spastic paraplegia using DNA probes. *Hum Genet* **73**: 264-266.

Keppen LD, Leppert MF, O'Connell P, Nakamura Y, Stauffer D, Lathrop M, Lalouel J-M, and White R. 1987. Etiological heterogeneity in X-linked spastic paraplegia. *Am J Hum Genet* **41**: 933-943.

Kerr B, Turner G, Mulley J, Gedeon A, and Partington M. 1991. Non-specific X linked mental retardation. *J Med Genet* **28**: 378-382.

Kerr B, Gedeon A, Mulley J, and Turner G. 1992. Localisation of non-specific X-linked mental retardation genes. *Am J Med Genet* **43**: 392-401.

Khan AS, Wilcox AS, Polymeropoulos MH, Hopkins JA, Stevens TJ, Robinson M, Orpana AK and Sikela JM. 1991. Single pass sequencing and physical mapping of

human brain cDNAs. *Nature Genet* **2**: 180-185.

Kitagawa K, Sinoway MP, Yang C, Gould RM, and Colman DR. 1993. A proteolipid gene family: expression in sharks and rays and possible evolution from an ancestral gene encoding a pore-forming polypeptide. *Neuron* **11**: 433-448.

Knight SJL, Flannery AV, Hirst MC, Campbell L, Christodoulou Z, Phelps SR, Pointon J, Middleton-Price HR, Barnicoat A, Pembrey ME, Holland J, Oostra BA, Bobrow M, Davies KE. 1993. Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE mental retardation. *Cell* **74**: 127-134.

Knowlton RG, Nelson CA, Brown VA, Page DC, and Donis-Keller H. 1989. An extremely polymorphic locus on the short arm of the human X chromosome with homology to the long arm of the Y chromosome. *Nuc Acids Res* **17**: 423-437.

Kobayashi H, Hoffman EP and Marks HG. 1994. The *rumpshaker* mutation in spastic paraplegia. *Nature Genet* **7**: 351-352.

Koenig M, Hoffman EP, Bertelson CJ, Monaco AP, Feener C, and Kunkel LM. 1987. Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. *Cell* **50**: 509-517.

Koenig M, Beggs AH, Moyer M *et al.* 1989. The molecular basis for Duchenne versus Becker muscular dystrophy: Correlation of severity with type of deletion. *Am J Hum Genet* **45**: 498-506.

Kogan SC, Doherty M, and Gitschier J. 1987. An improved method for prenatal diagnosis of genetic disease by analysis of amplified DNA sequences. *N Engl J Med* **317**: 985-990.

Kolluri R, Shehabeldin A, Peacocke M, Lamhonwah A-M, Teichert-Kuliszewska K, Weissman SM, and Siminovitch KA. 1995. Identification of WASP mutations in patients with Wiskott-Aldrich syndrome and isolated thrombocytopenia reveals allelic heterogeneity at the WAS locus. *Hum Mol Genet* **4**: 1119-1126.

Kondo I, Tsukamoto K, Niikawa N, Okano K, Kanazawa I, and Hupkes PE. 1991. A new form of X-linked mental retardation (XLMR) linked to DXS369 (RN1). *Cytogenet Cell Genet* **58**: 2071.

Kondoh T, Hayashi K, Matsumoto T, Yoshimoto M, Morio T, Yata J, and Tsuji Y. 1995. Two sisters with clinical diagnosis of Wiskott-Aldrich syndrome: Is the condition in the family autosomal recessive? *Am J Med Genet (Neuropsychiatric Genet)* **60**: 364-369, 1995.

Koorey DJ, Bishop GA, and McCaughan GW. 1993. Allele non-amplification: A source of confusion in linkage studies employing microsatellite polymorphisms. *Hum Mol Genet* **2**: 289-291.

Kosambi DD. 1944. The estimation of map distances from recombination values. *Annals of Eugenics* **12**: 172-175.

Kozak L, Chiurazzi P, Genuardi M, Pomponi MG, Zollino M, and Neri G. 1993. Mapping of a gene for non-specific X linked mental retardation: evidence for linkage to chromosomal region Xp21.1-Xp22.3. *J Med Genet* **30**: 866-869.

Kozman HM, Phillips HA, Callen DF, Sutherland GR, and Mulley J. 1993. Integration of the cytogenetic and genetic linkage maps of human chromosome 16 using 50 physical intervals and 50 polymorphic loci. *Cytogenet Cell Genet* **62**: 194-198.

Kruglyak L and Lander ES. 1995. High-resolution genetic mapping of complex traits. *Am J Hum Genet* **56**: 1212-1223.

Kruse TA, Byskov A, Oftedal G, Wirsching J, Lovseth TW, and Gedde-Dahl Jr. T. 1991. Linkage studies in a Norwegian paraplegia 2 family (Abstract). *Cytogenet Cell Genet* **58**: 2071.

Kwan S-P, Lehner T, Hagemann T, Lu B, Blaese M, Ochs H, Wedgwood R, Ott J, Craig IW, and Rosen FS. 1991. Localisation of the gene for Wiskott-Aldrich syndrome between two flanking markers, TIMP and DXS255, on Xp11.22-Xp11.3. *Genomics* **10**: 29-33.

Kwan S-P, Hageman TL, Radtke BE, Blaese RM, and Rosen FS. 1995. Identification of mutations in the Wiskott-Aldrich syndrome gene and characterisation of a polymorphic dinucleotide repeat at *DXS6940*, adjacent to the disease gene. *Proc Natl Acad Sci* **92**: 4706-4710.

Kwiatkowski DJ, Henske EP, Weimer K, Ozelius L, Gusella JF, and Haines J. 1992. Construction of a GT polymorphism map of human 9q. *Genomics* **12**: 229-240.

Kwiatkowski DJ, Dib C, Slaugenhaupt SA, Povey S, Gusella JF, and Haines JL. 1993. An index marker map of chromosome 9 provides strong evidence for positive interference. *Am J Hum Genet* **53**: 1279-1288.

Lagerstrom-Fermer M, Sundvall M, Johnsen E, Warne GL, Forrest SM, Zajac JD, Rickards A, Ravine D, Landegren U, and Pettersson U. 1997. X-linked recessive panhypopituitarism associated with a regional duplication in Xq25-q26. *Am J Hum Genet* **60**: 910-916.

Landegren U. 1992. Detection of mutations in human DNA. *GATA* **9**: 3-8.

Lander ES and Green P. 1987. Construction of multilocus genetic linkage maps in humans. *Proc Natl Acad Sci USA* **84**: 2363-2367.

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln S and Newburg L. 1987. MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* **1**: 174-181.

Lander ES and Schork NJ. 1994. Genetic dissection of complex traits. *Science* **265**: 2037-2048.

Laporte J, Hu LJ, Kretz C, Mandel J-L, Kioschis P, Coy JF, Klauck SM, Poustka A and Dahl N. 1996. A gene mutated in X-linked myotubular myopathy defines a new putative tyrosine phosphatase family conserved in yeast. *Nature Genet* **13**: 175-182.

Lasher L, Reefer J, and Chakravarti A. 1991. Effects of genotyping errors on the estimation of chromosome map length. *Am J Hum Genet* **49**: 985-994.

La Spada AR, Roling DB, Harding AE, Warner CL, Spiegel R, Hausmanowa-Petrusewicz I, Yee W-C, and Fischbeck KH. 1992. Meiotic stability and genotype-phenotype correlation of the trinucleotide repeat in X-linked spinal and bulbar muscular atrophy. *Nature Genet* **2**: 301-304.

Lathrop GM and Lalouel J-M. 1984. Easy calculations of lod scores and genetic risks on small computers. *Am J Hum Genet* **36**: 460-465.

Lathrop GM, Lalouel J-M, Julier C, and Ott J. 1984. Strategies for multilocus linkage analysis in humans. *Proc Natl Acad Sci USA* **81**: 3443-3446.

Lathrop GM, Lalouel J-M, Julier C, and Ott J. 1985. Multilocus linkage analysis in humans: detection of linkage and estimation of recombination. *Am J Hum Genet* **37**: 482-498.

Lathrop GM and Lalouel J-M. 1991. Statistical methods for linkage analysis. In: Rao CR and Chakraborty R, eds; *Handbook of Statistics* vol **8**: 81-123, North-Holland, Amsterdam.

Lazzarini A, Stenroos ES, Lehner T, McKoy V, Gold B, McCormack MK, Reid CS, Ott J and Johnson WG. 1995. Short tandem repeat polymorphism linkage studies in a new family with X-linked mental retardation (MRX20). *Am J Med Genet* **57**: 552-557.

Lee EY and Lee WH. 1986. Molecular cloning of the human esterase D gene, a genetic marker of retinoblastoma. *Proc Natl Acad Sci USA* **83**: 6337-6341.

Lee CC, Yazdani A, Wehnert M, Zhao ZY, Lindsay EA, Bailey J, Coolbaugh MI, Couch L, Xiong M, Chinault AC, Baldini A, and Caskey CT. 1995. Isolation of chromosome-specific genes by reciprocal probing of arrayed cDNA and cosmid libraries. *Hum Mol Genet* **4**: 1373-1380.

Lehrach H, Drmanac R, Hoheisel J, Larin Z, Lennon G, Monaco AP, Nizetic D, Zehetner G, and Poustka A. 1990. Hybridization fingerprinting in genome mapping and sequencing. In *Genome Analysis Volume 1: Genetic and Physical Mapping*. Davies KE and Tilgham SM (eds.). Cold Spring Harbour Laboratory Press, Cold Spring Harbour, pp.39-81.

- Lehrke RG. 1974. X-linked mental retardation and verbal disability. *Birth Defects Orig Art Ser* **12**: 1-100.
- Le Merrer M, Rosey JM, Bonneau D, Kaplan J, Munnich A, and Frezal J. 1989. A family study of X-linked spastic paraplegia, SPG2. *Cytogenet Cell Genet* **51**: 1030.
- Levinson G and Gutman GA. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol* **4**: 203-221.
- Li X-M, Yen P, Mohandas T, and Shapiro LJ. 1990. A long range restriction map of the distal human X chromosome short arm around the steroid sulfatase locus. *Nuc Acids Res* **18**: 2783-2788.
- Lincoln SE and Lander ES. 1992. Systematic detection of errors in genetic linkage data. *Genomics* **14**: 604-610.
- Lindsay EA, Grillo A, Ferrero GB, Roth EJ, Magenis E, Grompe M, Hulten M, Gould C, Baldini A, Zoghbi HY, and Ballabio A. 1994. Microphthalmia with linear skin defects (MLS) syndrome: clinical, cytogenetic, and molecular characterization. *Am J Med Genet* **49**: 229-234.
- Lindsay S and Bird AP. 1987. Use of restriction enzymes to detect potential gene sequences in mammalian DNA. *Nature* **327**: 336-338.
- Litt M and Luty JA. 1989. A hypervariable microsatellite repeat revealed by in vitro amplification of a dinucleotide repeat within the cardiac actin muscle gene. *Am J Hum Genet* **44**: 397-401.
- Lovett M. 1994. Fishing for complements: finding genes by direct selection. *Trends Genet* **10**: 352-357.
- Lovett M, Kere J, and Hinton LM. 1991. Direct selection: A method for the isolation of cDNAs encoded by large genomic regions. *Proc Natl Acad Sci USA* **88**: 9628-9632.
- Lowe T, Sharefkin J, Yang SQ, and Dieffenbach CW. 1990. A computer program for selection of oligonucleotide primers for polymerase chain reactions. *Nuc Acids Res* **18**: 1757-1761.

Lu Q, Wallrath LL, Granok H and Elgin SCR. 1993. (CT)<sub>n</sub>.(GA)<sub>n</sub> repeats and heat shock elements have distinct roles in chromatin structure and transcriptional activation of the *Drosophila hsp26* gene. *Mol Cell Biol* **13**: 2802-2814.

Lubahn DB, Brown TR, Simental JA, Higgs HN, Migeon CJ, Wilson EM, and French FS. 1989. Sequence of the intron/exon junction of the coding region of the human androgen receptor gene and identification of a point mutation in a family with complete androgen insensitivity. *Proc Natl Acad Sci USA* **86**: 9534-9538.

Lubs HA, Chiurazzi P, Arena JF, Schwartz C, Tranebjaerg L, and Neri G. 1996. XLMR genes: update 1996. *Am J Med Genet* **64**: 147-157.

Luty JA, Guo Z, Willard HF, Ledbetter DH, Ledbetter S, and Litt M. 1990. Five polymorphic microsatellite VNTRs on the human X chromosome. *Am J Hum Genet* **46**: 776-783.

Ma W, Hu Z-B and Drexler HG. 1994. Sensitivity of different methods for the detection of myeloperoxidase in leukemia cells. *Leukemia* **8**: 336-342.

Macklin WB, Campagnoni CW, Deininger PL, and Gardinier MV. 1987. Structure and expression of the mouse myelin proteolipid protein gene. *J Neurosci Res* **18**, 383-394.

Magenis RE, Donlon TA, Canfield WM, Koler R, and Bigley R. 1984. Partial duplication of X long arm, q26.3-qter: gene dosage effects of G6PD, HPRT, factor VIII:C, and factor IX. *Cytogenet Cell Genet* **37**: 528 (Abstract).

Mahtani MM and Willard HF. 1988. A primary genetic map of the pericentromeric region of the human X chromosome. *Genomics* **2**: 294-301.

Mahtani MM and Willard HF. 1990. Pulsed-field gel analysis of  $\alpha$ -satellite DNA at the human X chromosome centromere: high frequency polymorphisms and array size estimate. *Genomics* **7**: 607-613.

Mahtani MM and Willard HF. 1993. A polymorphic X-linked tetranucleotide repeat locus displaying a high rate of new mutation: implications for mechanisms of mutation at short tandem repeat loci. *Hum Mol Genet* **2**: 431-437.

Malmgren H, Sundvall M, Dahl N, Gustavson K-H, Anneren G, Wadelius C, Steen-Bondeson M-L, and Pettersson U. 1993. Linkage mapping of a severe X-linked mental retardation syndrome. *Am J Hum Genet* **52**: 1046-1052.

Mandel J-L. 1994. Towards identification of X-linked mental retardation genes: a proposal. *Am J Med Genet* **51**: 550-552.

Mandel J-L, Monaco AP, Nelson DL, Schlessinger D, and Willard H. 1992. Genome analysis and the human X chromosome. *Science* **258**: 103-109.

Mann JD, Cahan A, Gelb AG, Fisher N, Hamper J, Tippett P, Sanger R, and Race RR. 1962. A sex-linked blood group. *Lancet* **i**: 8-10.

Martinez F, Gal Andreas, Palau F, and Prieto F. 1995. Localization of a gene for X-linked non-specific mental retardation (MRX24) in Xp22.2-p22.3. *Am J Med Genet* **55**: 387-390.

Matise TC, Perlin M, and Chakravarti A. 1994. Automated construction of genetic linkage maps using an expert system (MultiMap): a human genome linkage map. *Nature Genet* **6**: 384-390.

Mattei MG, Mattei JF, Ayme S, and Giraud F. 1982. X-autosome translocations: cytogenetic characteristics and their consequences. *Hum Genet* **61**: 295-309.

May M, Colleaux L, Murgia A, Aylsworth A, Nussbaum R, Fontes M and Schwartz C. 1995. Molecular analysis of four males with mental retardation and deletions of Xq21 places the putative MR region in Xq21.1 between DXS233 and CHM. *Hum Mol Genet* **4**: 1465-1466.

Mazzarella R and Srivastava AK. 1994. Physical linkage of expressed tags (ESTs) to polymorphic markers on the X chromosome. *Hum Mol Genet* **3**: 1095-1101.

McBurney MW. 1993. X chromosome inactivation: the feminine mystique continues. *BioEssays* **15**: 825-826.

McClelland A, Kuhn LC and Ruddle FH. 1984. The human transferrin receptor gene: genomic organization, and the complete primary structure of the receptor deduced from a

cDNA sequence. *Cell* **39**: 267-274.

McInnes MG, Chakravarti A, Blaschak J, Petersen MB, Sharma V, Avramopoulos D, Blouin J-L, Konig U, Brahe C, Matisse TC, Warren AC, Talbot Jr. CC, van Broeckhoven C, Litt M, and Antonarakis SE. 1993. A linkage map of human chromosome 21: 43 PCR markers at average intervals of 2.5 cM. *Genomics* **16**: 562-571.

McKusick VA. 1980. The anatomy of the human genome. *J Hered* **71**: 370-391.

McKusick VA. 1990. "Mendelian Inheritance in Man. Catalogs of Autosomal Dominant, Autosomal Recessive, and X-linked Phenotypes", 9th edition: Baltimore, John Hopkins University Press.

McKusick VA. 1992. *ibid*, 10th edition.

McKusick VA. 1994. *ibid*, 11th edition.

McPhaul MJ, Marcelli M, Tilley WD, Griffin JE, Isidro-Gutierrez, and Wilson JD. 1991. Molecular basis of androgen resistance in a family with a qualitative abnormality of the androgen receptor and responsive to high-dose androgen therapy. *J Clin Invest* **87**: 1413-1421.

Meitinger T, Heye B, Petit C, Levilliers J, Golla A, Moraine C, Dalla Piccola B, Sippell WG, Murken J, and Ballabio A. 1990. Definitive localization of X-linked Kallmann syndrome (hypogonadotropic hypogonadism and anosmia) to Xp22.3: close linkage to the hypervariable repeat sequence CRI-S232. *Am J Hum Genet* **47**: 664-669.

Melmer G and Buchwald M. 1992. Identification of genes using oligonucleotides corresponding to splice site consensus sequences. *Hum Mol Genet* **1**: 433-438.

Mhatre AN, Trifiro MA, Kaufman M, Kazemi-Esfarjani P, Figlewicz D, Rouleau G, and Pinsky L. 1993. Reduced transcriptional regulatory competence of the androgen receptor in X-linked spinal and bulbar muscular atrophy. *Nature Genet* **5**: 184-188.

Migeon BR, Hugo WM, Moser AB, Axelman J, Sillence D, and Norum RA. 1981. Adrenoleukodystrophy: Evidence for X linkage, inactivation, and selection favoring the mutant allele in heterozygous cells. *Proc Natl Acad Sci USA* **78**: 5066-5070.

- Migeon BR, Axelman J, and Stetten G. 1988. Clonal evolution in human lymphoblast cultures. *Am J Hum Genet* **42**: 742-747.
- Migeon BR, Luo S, Jani M, and Jeppesen P. 1994. The severe phenotype of females with tiny ring X chromosomes is associated with inability of these chromosomes to undergo inactivation. *Am J Hum Genet* **55**: 497-504.
- Miller FD, Dixon GH, Rattner JB, and van de Sande JH. 1985. Assembly and characterization of nucleosomal cores on B-vs. Z-form DNA. *Biochemistry* **24**: 102-109.
- Milot E, Fraser P, and Grosveld F. 1996. Position effects and genetic disease. *Trends Genet* **12**: 123-126.
- Miwa S. 1994. Triplet repeats strike again. *Nature Genet* **6**: 3-4.
- Mizusawa S, Nishimura S, and Seela F. 1986. Improvement of the dideoxy chain termination method of DNA sequencing by use of deoxy-7-deazaguanosine triphosphate in place of dGTP. *Nuc Acids Res* **14**: 1319-1324.
- Molina IJ, Kenney DM, Rosen FS, and Remold-O'Donnell E. 1992. T cell lines characterize events in the pathogenesis of the Wiskott-Aldrich syndrome. *J Exp Med* **176**: 867-874.
- Monaco AP, Neve RL, Colletti-Feener C, Bertelson CJ, Kurnit DM and Kunkel LM. 1986. Isolation of candidate cDNAs for portions of the Duchenne muscular dystrophy gene. *Nature* **323**: 646-650.
- Moraine CI, Dessay B, Toutain A, Briault S, Gendrot Ch and Ronce N. 1994. X-linked mental retardation: linkage results in five unrelated families. *Am J Hum Genet* **55**: A196.
- Morton JE, Bunday S, Webb TP, MacDonald F, Rindl PM, and Bullock S. 1997. Fragile X syndrome is less common than previously estimated. *J Med Genet* **34**: 1-5.
- Morton NE. 1955. Sequential tests for the detection of linkage. *Am J Hum Genet* **7**: 277-318.
- Morton NE. 1991. Parameters of the human genome. *Proc Natl Acad Sci USA* **88**: 7474-

Morton NE and Andrews V. 1989. MAP, an expert system for multiple pairwise linkage analysis. *Ann Hum Genet* **53**: 263-269.

Moyzis RK, Torney DC, Meyne J, Buckingham JM, Wu J-R, Burks C, Sirotkin KM and Goad WB. 1989. The distribution of interspersed repetitive DNA sequences in the human genome. *Genomics* **4**: 273-289.

Mulley JC, Kerr B, Stevenson R, and Lubs H. 1992. Nomenclature guidelines for X-linked mental retardation. *Am J Med Genet* **43**: 383-391.

Mulley JC, Yu S, Loesch DZ, Hay DA, Donnelly A, Gedeon AK, Carbonell P, Lopez I, Glover G, Gabarron I, Yu PWL, Baker E, Haan EA, Hockey A, Knight SJL, Davies KE, Richards RI, and Sutherland GR. 1995. FRAXE and mental retardation. *J Med Genet* **32**: 162-169.

Mullis KB and Faloona FA. 1987. Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. *Methods Enzymol* **155**: 335-350.

Munroe DJ, Haas M, Bric E, Whitton T, Aburatani H, Hunter K, Ward D, and Housman DE. 1994. IRE-bubble PCR: a rapid method for efficient and representative amplification of human genomic DNA sequences from complex sources. *Genomics* **19**: 506-514.

Murray JC, Buetow KH, Weber JL *et al.* 1994. A comprehensive human linkage map with centimorgan density. *Science* **265**: 2049-2054.

Murray V, Monchawin C and England PR. 1993. The determination of the sequences present in the shadow bands of a dinucleotide repeat PCR. *Nuc Acids Res* **21**: 2395-2398.

Nagaraja R, MacMillan S, Kere J *et al.* 1997. X chromosome map at 75-kb STS resolution, revealing extremes of recombination and GC content. *Genome Research* **7**: 210-222.

Nakahori Y, Hamano K, Iwaya M, and Nakagome Y. 1991. Sex identification by the polymerase chain reaction using X-Y homologous primer. *Am J Med Genet* **39**: 472-473.

Nakamura Y, Carlson M, Krapcho K, Kanamori M, and White R. 1988 New approach for isolation of VNTR markers. *Am J Hum Genet* **43**: 854-859.

Naumova AK, Plenge RM, Bird LM, Leppert M, Morgan K, Willard HF, and Sapienza C. 1996. Heritability of X chromosome-inactivation phenotype in a large family. *Am J Hum Genet* **58**: 1111-1119.

Neitzel H. 1986. A routine method for the establishment of permanent growing lymphoblastoid cell lines. *Hum Genet* **73**: 320-326.

Nelson DL, Ledbetter SA, Corbo L, Victoria MF, Ramirez-Solis R, Webster TD, Ledbetter DH and 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**: 6686-6690.

Nelson DL, Ballabio A, Cremers F, Monaco AP and Schlessinger D. 1995. Report of the sixth international workshop on X chromosome mapping 1995. *Cytogenet Cell Genet* **71**: 308-342.

NIH/CEPH Collaborative Mapping Group. 1992. A comprehensive genetic linkage map of the human genome. *Science* **258**: 67-86.

Nisen P, Stamberg J, Ehrenpreis R, Velasco S, Shende A, Engelberg J, Karayalcin G, and Waber L. 1986. The molecular basis of severe hemophilia B in a girl. *N Engl J Med* **315**: 1139-1142.

Nisson PE and Watkins PC. 1994. Isolation of exons from cloned DNA by exon trapping. In *Current Protocols in Human Genetics*. Dracopoli NC, Haines JL, Korf BR, Moir DT, Morton CC Seidman CE, Seidman JG, Smith DR. (eds). John Wiley and Sons.

Nizetic D, Zehetner G, Monaco AP, Gellen L, Young BD, and Lehrach H. 1991. Construction, arraying, and high-density screening of large insert libraries of human chromosomes X and 21: their potential use as reference libraries. *Proc Natl Acad Sci USA* **88**: 3233-3237.

- Nordheim A and Rich A. 1983. The sequence (dC-dA)<sub>n</sub>.(dG-dT)<sub>n</sub> forms left-handed Z-DNA in negatively supercoiled plasmids. *Proc Natl Acad Sci USA* **80**: 1821-1825.
- Nordstrom AM, Penttinen M, and von Koskull H. 1992. Linkage to Xq28 in a family with non-specific X-linked mental retardation. *Hum Genet* **90**: 263-266.
- Olson MV. 1993. The human genome project. *Proc Natl Acad Sci USA* **90**: 4338-4344.
- Ott J. 1974. Estimation of the recombination fraction in human pedigrees: efficient computation of the likelihood for human linkage studies. *Am J Hum Genet* **26**: 588-597.
- Ott J. 1976. A computer program for linkage analysis of general human pedigrees. *Am J Hum Genet* **28**: 528-529.
- Ott J. 1991. "Analysis of Human Genetic Linkage" revised ed., Johns Hopkins Univ. Press, Baltimore, MD.
- Oudet C, Heilig R, and Mandel J-L. 1990. An informative polymorphism detectable by polymerase chain reaction at the 3' end of the dystrophin gene. *Hum Genet* **84**: 283-285.
- Oudet C, Heilig R, Hanauer A, and Mandel J-L. 1991. Non-radioactive assay for new microsatellite polymorphisms at the 5' end of the dystrophin gene, and estimation of intragenic recombination. *Am J Hum Genet* **49**: 311-319.
- Oudet C, Hanauer A, Clemens P, Caskey T, and Mandel J-L. 1992a. Two hot spots of recombination in the DMD gene correlate with the deletion prone regions. *Hum Mol Genet* **1**: 599-603.
- Oudet C, Weber C, Kaplan J, Segues B, Croquette M-F, Roman EO, and Hanauer A. 1992b. Characterisation of a highly polymorphic microsatellite at the DXS207 locus: confirmation of a very close linkage to the retinoschisis gene. *J Med Genet* **30**: 300-303.
- Parimoo S, Patanjali SR, Shulka H, Chaplin DD, and 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**: 9623-9627.

Parrish JE and Nelson DL. 1993. Regional assignment of 19 X-linked ESTs. *Hum Mol Genet* **2**: 1901-1905.

Partington MW, Mulley JC, Sutherland GR, Hockey A, Thode A, and Turner G. 1988. X-linked mental retardation with dystonic movements of the hands. *Am J Med Genet* **30**: 251-262.

Passos-Bueno MR, Byht BC, Rosenberg S, Takata RI, Bakker E, Beggs AH, Pavanello RC, Vainzof M, Davies KE, and Zatz M. 1993. Severe nonspecific X-linked mental retardation caused by a proximally Xp located gene: intragenic heterogeneity or a new form of X-linked mental retardation? *Am J Med Genet* **46**: 172-175.

Patterson MN, Hughes IA, Gottlieb B, and Pinsky L. 1994. The androgen receptor gene mutations database. *Nuc Acids Res* **22**: 3560-3562.

Pearson WR. 1990. Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods in Enzymology* **183**: 63-98.

Pearson WR and Lipman DJ. 1988. Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA* **85**: 2444-2448.

Pekkarinen P, Terwilliger J, Bredbacka P-E, J Lonnqvist, and Peltonen L. 1995. Evidence of a predisposing locus to bipolar disorder on Xq24-q27.1 in an extended Finnish pedigree. *Genome Research* **5**: 105-115.

Perry GS, Spector BD, Schuman LM, Mandel JS, Anderson VE, McHugh RB, Hanson MR, Fahlstrom SM, Krivit W, and Kersey JH. 1980. The Wiskott-Aldrich syndrome in the United States and Canada (1892-1979). *J Pediatr* **97**: 72-78.

Peters A, Palay SL, de F Webster H. 1970. *The Fine Structure of the Nervous System: The Cells and Their Processes*. New York, Harper and Row, pp70-104.

Peterson AC, Di Renzo A, Lehesjoki A-E, de la Chapelle A, Slatkin M, and Freimer NB. 1995. The distribution of linkage disequilibrium over anonymous genome regions. *Hum Mol Genet* **4**: 887-894.

- Petit C, Levilliers J, and Weissenbach J. 1990. Long-range restriction map of the terminal part of the short arm of the human X chromosome. *Proc Natl Acad Sci USA* **87**: 3680-3684.
- Petrukhin KE, Speer MC, Cayanis E, de Fatima Bonaldo M, Tantravahi U, Bento Soares M, Fischer SG, Warburton D, Gilliam TC, and Ott J. 1993. A microsatellite genetic linkage map of human chromosome 13. *Genomics* **15**: 76-85.
- Pettigrew AL, Jackson LG, Ledbetter DH. 1991. A new X-linked mental retardation disorder with Dandy-Walker malformation, basal ganglia disease and seizures. *Am J Med Genet* **38**: 200-207.
- Phelan PD, Connelly J, Martin FIR, Wettenhall HNB. 1971. X-linked recessive hypopituitarism. *Birth Defects: Original Article Series* **7**: 24-27.
- Philippe C, Cremers FPM, Chery M, Bach I, Abbadi N, Ropers HH, and Gilgenkrantz S. 1993. Physical mapping of DNA markers in the q13-q22 region of the human X chromosome. *Genomics* **17**: 147-152.
- Philippe C, Arnould C, Sloan F, van Bokhoven H, van der Velde-Visser SD, Chery M, Ropers HH, Gilgenkrantz S, Monaco AP and Cremers FPM. 1995. A high-resolution interval map of the q21 region of the human X chromosome. *Genomics* **27**: 539-543.
- Phillips HA, Hyland VJ, Holman K, Callen DF, Richards RI, and Mulley JC. 1991. Dinucleotide repeat polymorphism at D16S287. *Nuc Acids Res* **19**: 6664.
- Pilia G, MacMillan S, Nagaraja R, Mumm S, Weissenbach J, and Schlessinger D. 1996. YAC/STS map of 9 Mb of Xq26 at 100-kb resolution, localizing 6 ESTs, 6 genes and 32 genetic markers. *Genomics* **34**: 55-62.
- Poduslo SE, Dean M, Kolch U, and O'Brien SJ. 1991. Detecting high-resolution polymorphisms in human coding loci by combining PCR and single-stranded conformation polymorphism (SSCP) analysis. *Am J Hum Genet* **49**: 106-111.
- Popot J-L, Pham Dinh D, and Dautigny A. 1991. Major myelin proteolipid: the 4- $\alpha$ -helix topology. *J Membrane Biology* **120**: 233-246.

Porteus MEM, Curtis A, Lindsay S, Williams O, Goudie D, Kamakari S, and Bhattacharya SS. 1992. The gene for Aarskog syndrome is located between DXS255 and DXS566 (Xp11.2-Xq13). *Genomics* **14**: 298-301.

Pratt VM, Dlouhy SR, and Hodes ME. 1995a. Pelizaeus-Merzbacher disease: a point mutation in exon 6 of the proteolipid protein (*PLP*) gene. *Clin Genet* **47**, 99-100.

Pratt VM, Boyadjiev S, Dlouhy SR, Silver K, Der Kaloustian VM, and Hodes ME. 1995b. Pelizaeus-Merzbacher disease in a family of Portuguese origin caused by a point mutation in exon 5 of the proteolipid protein gene. *Am J Med Genet* **55**: 402-404.

Proops R, Mayer M, and Jacobs PA. 1983. A study of mental retardation in children in the island of Hawaii. *Clinical Genetics* **23**: 81-96.

Quan F, Zonana J, Gunter K, Peterson KL, Magenis RE, and Popovich BW. 1995. An atypical case of fragile X syndrome caused by a deletion that includes the *FMR1* gene. *Am J Hum Genet* **56**: 1042-1051.

Raeymaekers P, Lin J, Gu XX, Soekarman D, Cassiman J-J, Fryns J-P, and Marynen P. 1996. A form of non-specific mental retardation is probably caused by a microdeletion in a Belgian family. (Abstract 7) *Am J Med Genet* **64**: 15-20.

Raleigh EA, Lech K, and Brent R. 1989. In *Current Protocols in Molecular Biology*. Ausubel FM *et al.*, eds. Publishing Associates and Wiley Interscience, New York.

Rao DC, Morton NE, Lindsten J, Hulten M, and Yee S. 1977. A mapping function for man. *Hum Hered* **27**: 99-104.

Rappold GA, Klink A, Weiss B, and Fischer C. 1994. Double crossover in the human Xp/Yp pseudoautosomal region and its bearing on interference. *Hum Mol Genet* **3**: 1337-1340.

Raynaud M, Gendrot C, Dessay B, Moncla A, Ayrault A-D, Moizard M-P, Toutain A, Briault S, Villard L, Ronce N and Moraine C. 1996. X-linked mental retardation with neonatal hypotonia in a French family (MRX15): Gene assignment to Xp11.22-Xp21.1. *Am J Med Genet* **64**: 97-106.

Reed KC and Mann DA. 1985. Rapid transfer of DNA from agarose gels to nylon membranes. *Nuc Acids Res* **13**: 7207-7221.

Renwick JH. 1969. Progress in mapping human autosomes. *Brit Med Bull* **25**: 65-73.

Renwick JH. 1971. The mapping of human chromosomes. *Ann Rev Genet* **5**:81-120.

Richards RI, Holman K, Kozman H, Kremer E, Lynch M, Pritchard M, Yu S, Mulley JC, and Sutherland GR. 1991a. Fragile X syndrome: Genetic localisation by linkage mapping of two microsatellite repeats FRAXAC1 and FRAXAC2 which immediately flank the fragile site. *J Med Genet* **28**: 818-823.

Richards RI, Shen Y, Holman K, Kozman H, Hyland VJ, Mulley JC, and Sutherland GR. 1991b. Fragile X syndrome: diagnosis using highly polymorphic microsatellite markers. *Am J Hum Genet* **48**: 1051-1057.

Richards RI and Sutherland GR. 1992. Heritable unstable DNA sequences. *Nature Genet* **1**: 7-9.

Richards RI, Holman K, Yu S, and Sutherland GR. 1993. Fragile X syndrome unstable element, p(CCG)n, and other simple tandem repeat sequences are binding sites for specific nuclear proteins. *Hum Mol Genet* **2**: 1429-1435.

Richards RI and Sutherland GR. 1994. Simple repeat DNA is not replicated simply. *Nature Genet* **6**: 114-116.

Riggins GJ, Lokey LK, Chastain JL, Leiner HA, Sherman SL, Wilkinson KD, and Warren ST. 1992. Human genes containing polymorphic trinucleotide repeats. *Nature Genet* **2**: 186-191.

Risch N and Botstein D. 1996. A manic depressive history. *Nature Genet* **12**: 351-353.

Robledo R, Melis P, Laficara F, Marchi J, Rinaldi A, Siniscalco M and Filippi G. 1996a. Further linkage evidence for localisation of mutational sites for nonsyndromic types of X-linked mental retardation at the pericentromeric region. *Am J Med Genet* **64**: 107-112.

- Robledo R, Melis P, Siniscalco M, Marchi J, Laficara F, Rinaldi A, Rocchi M, and Filippi G. 1996b. +2.71 lod score at zero recombination is not sufficient for establishing linkage between X-linked mental retardation and X-chromosome markers. *Am J Med Genet* **64**: 134-136.
- Rocca B, Bellacosa A, De Cristofaro R, Neri G, Ventura MD, Maggiano N, Rumi C, and Landolfi R. 1996. Wiskott-Aldrich syndrome: Report of an autosomal dominant variant. *Blood* **87**: 4538-4543.
- Roest Crollius H, Ross MT, Grigoriev A *et al.* 1996. An integrated YAC map of the human X chromosome. *Genome Research* **6**: 943-955.
- Romeo G and McKusick VA. 1994. Phenotypic diversity, allelic series and modifier genes. *Nature Genet* **7**: 451-453.
- Rosen FS, Cooper MD, and Wedgewood RJP. 1995. The primary immunodeficiencies. *N Eng J Med* **333**: 431-440.
- Roustan P, Curtis ARJ, Kamakari S, Thiselton D, Lindsay S, and Bhattacharya SS. 1992. Dinucleotide repeat polymorphism at the DXS559 locus. *Hum Mol Genet* **1**: 778.
- Rouyer F, de la Chapelle A, Andersson M, and Weissenbach J. 1990. An interspersed repeated sequence specific for human subtelomeric regions. *EMBO* **9**: 505-514.
- Rowe PSN, Goulding J, Read A, Lehrach H, Francis F, Hanauer A, Oudet C, Biancalana, Kooh SW, Davies KE, and O'Riordan JLH. 1994. Refining the genetic map for the region flanking the X-linked hypophosphataemic rickets locus (Xp22.1-22.2). *Hum Genet* **93**: 291-294.
- Royle NJ, Clarkson RE, Wong Z, and Jeffreys AJ. 1988. Clustering of hypervariable minisatellites in the proterminal regions of human autosomes. *Genomics* **3**: 352-360.
- Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, and Arnheim N. 1985. Enzymatic amplification of  $\beta$ -globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**: 1350-1354.

- Samanns C, Albrecht R, Neugebauer M, and Neri G, Gal A. 1991: Gene for non-specific X-linked mental retardation maps in the pericentromeric region. *Am J Med Genet* **38**: 224-227.
- Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular Cloning-A Laboratory Manual*. 2nd edition. CSH press, N.Y., U.S.A.
- Samson F, Mesnard L, Heimbürger M, Hanauer A, Chevally M, Mercadier JJ, Pelissier JF, Feingold N, Junien C, Mandel J-L and Fardeau M. 1995. Genetic linkage heterogeneity in myotubular myopathy. *Am J Hum Genet* **57**: 120-126.
- Sarkar G, Paynton C, and Sommer SS. 1991. Segments containing alternating purine and pyrimidine dinucleotides: patterns of polymorphism in humans and prevalence throughout phylogeny. *Nuc Acids Res* **19**: 631-636.
- Saugier-Veber P, Munnich A, Bonneau D, Rozet J-M, Le Merrer M, Gil R, and Boespflug-Tanguy O. 1994. X-linked spastic paraplegia and Pelizaeus-Merzbacher disease are allelic disorders at the proteolipid protein locus. *Nature Genet* **6**: 257-262.
- Savov A, Angelicheva D, Jordanova A, Eigel A, and Kalaydjieva L. 1992. High percentage acrylamide gels improve resolution in SSCP analysis. *Nuc Acids Res* **20**: 6741-6742.
- Sawadogo M and van Dyke MW. 1991. A rapid method for the purification of deprotected oligonucleotides. *Nuc Acids Res* **19**: 674.
- Schaefer L, Ferrero GB, Grillo A, Bassi MT, Roth EJ, Wapenaar MC, van Ommen G-JB, Thuluvancheri KM, Rocchi M, Zoghbi HY, and Ballabio A. 1993. A high resolution deletion map of human chromosome Xp22. *Nature Genet* **4**: 272-277.
- Schindelhauer D, Weiss M, Hellebrand H, Golla A, Hergersberg M, Seger R, Belohradsky BH, and Meindl A. 1996. Wiskott-Aldrich syndrome: no strict genotype-phenotype correlations but clustering of missense mutations in the amino-terminal part of the WASP gene product. *Hum Genet* **98**: 68-76.

- Schlessinger D, Mandel J-L, Monaco AP, Nelson DL, and Willard HF. 1993. Report of the Fourth International Workshop on Human X Chromosome Mapping 1993. *Cytogenet Cell Genet* **64**: 148-194.
- Schliess F and Stoffel W. 1991. Evolution of the myelin integral membrane proteins of the central nervous system. *Biol Chem Hoppe-Seyler* **372**: 865-874.
- Schlotterer C and Tautz D. 1992. Slippage synthesis of simple sequence DNA. *Nuc Acids Res* **20**: 211-215.
- Schmidt M, Du Sart D, Kalitsis P, Leversha M, Dale S, Sheffield L, and Toniolo D. 1991. Duplications of the X chromosome in males: evidence that most parts of the chromosome can be active in two copies. *Hum Genet* **38**: 411-415.
- Schmidt M and Du Sart D. 1992. Functional disomies of the X chromosome influence the cell selection and hence the X inactivation pattern in females with balanced X-autosome translocations. *Am J Med Genet* **42**: 161-169.
- Schuler GD, Boguski MS, Stewart EA, et al. 1996. A gene map of the human genome. *Science* **274**: 540-546.
- Schutz CK, Ives EJ, Chalifoux M, MacLaren L, Farrell S, Robinson PD, White BN and Holden JJA. 1996. Regional localisation of an X-linked mental retardation gene to Xp21.1-Xp22.3 (MRX38). *Am J Med Genet* **64**: 89-96.
- Schwartz CE. 1993. X-linked mental retardation: In pursuit of a gene map. *Am J Hum Genet* **52**: 1025-1031.
- Schwartz CE, May M, Huang T, Ledbetter D, Anderson G, Barker DF, Lubs HA, Arena F, and Stevenson RE. 1992. MRX8: an X-linked mental retardation condition with linkage to Xq21. *Am J Med Genet* **43**: 467-474.
- Sealey PG, Whittaker PA and Southern EM. 1985. Removal of repeated sequences from hybridisation probes. *Nuc Acids Res* **13**: 1905-1922.
- Sefton L and Goodfellow PN. 1992. The human genetic map. *Current Opinion in Biotechnology* **3**: 607-611.

- Sheffield VC, Beck JS, Nichols B, Cousineau A, Lidral AC, and Stone EM. 1992. Detection of multiallele polymorphisms within gene sequences by GC-clamped denaturing gradient gel electrophoresis. *Am J Hum Genet* **50**: 567-575.
- Sheffield VC, Beck JS, Kwitek AE, Sandstrom DW, and Stone EM. 1993. The sensitivity of single-strand conformation polymorphism analysis for the detection of single base substitutions. *Genomics* **16**: 325-332.
- Sheffield VC, Weber JL, Beutow KH *et al.* 1995. A collection of tri- and tetranucleotide repeat markers used to generate high quality, high resolution human genome-wide linkage maps. *Hum Mol Genet* **4**: 1837-1844.
- Sherman F, Fink GR, and Hicks JB. 1986. *Laboratory Course Manual for Methods in Yeast Genetics*, Cold Spring Harbour.
- Shields DC, Collins A, Buetow KH, and Morton NE. 1991. Error filtration, interference, and the human linkage map. *Proc Natl Acad Sci USA* **88**: 6501-6505.
- Shows TB and McAlpine PJ. 1979. The 1979 catalog of human genes and chromosome assignments. *Cytogenet Cell Genet* **25**: 117-127.
- Simental JA, Sar M, Lane MV, French FS, and Wilson E. 1991. Transcriptional activation and nuclear targeting signals of the human androgen receptor. *J Biol Chem* **266**: 510-518.
- Smeets HJM, Brunner HG, Ropers H, and Wieringa B. 1989. Use of variable simple sequence motifs as genetic markers: Application to study of myotonic dystrophy. *Hum Genet* **83**: 245-251.
- Smith CAB. 1953. The detection of linkage in human genetics (with discussion). *J Roy Stat Soc B* **15**: 153-192.
- Smith CAB. 1986. The development of human linkage analysis. *Ann Hum Genet* **50**: 293-311.
- Smith LM. 1993. The future of DNA sequencing. *Science* **262**: 530-532.

- Snyder EE and Stormo GD. 1993. Identification of coding regions in genomic DNA sequences: an application of dynamic programming and neural networks. *Nuc Acids Res* **21**: 607-613.
- Southern EM. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* **98**: 503-517.
- Spurr NK, Bryant SP, Attwood J, Nyberg K, Cox SA, Mills A, Bains R, Warne D, Cullin L, Povey S, Sebaoun J-M, Weissenbach J, Cann HM, Lathrop M, Dausset J, Marcadet-Troton A, and Cohen D. 1994. European gene mapping project (EUROGEM): genetic maps based on the CEPH reference families. *Eur J Hum Genet* **2**: 193-252.
- Stallings RL, Ford AL, Nelson D, Torney DC, Hildebrand CE, and Moyzis RK. 1991. Evolution and distribution of (GT)<sub>n</sub> repetitive sequences in mammalian genomes. *Genomics* **10**: 807-815.
- Stanier P, Newton R, Forbes SA, Ivens A, and Moore GE. 1991. Polymorphic dinucleotide repeat at the DXS3 locus. *Nuc Acids Res* **19**: 4793.
- Sullivan KE, Mullen CA, Blaese RM, and Winkelstein JA. 1994. A multi-institutional survey of the Wiskott-Aldrich syndrome. *J Pediatr* **125**: 876-885.
- Sutherland GR and Baker E. 1992. Characterisation of a new rare fragile site easily confused with the fragile X. *Hum Mol Genet* **1**: 111-113.
- Sutherland GR and Hecht F. 1985. Fragile sites on human chromosomes. New York: Oxford University Press.
- Suthers GK and Davies KE. 1992. Phenotypic heterogeneity and the single gene. *Am J Hum Genet* **50**: 887-891.
- Symons M, Derry JMJ, Karlak B, Jiang S, Lemahieu V, McCormick F, Franke U, and Abo A. 1996. Wiskott-Aldrich syndrome protein, a novel effector for the GTPase CDC42Hs, is implicated in actin polymerization. *Cell* **84**: 723-734.
- Tagle DA and Collins FS. 1992. An optimized *Alu* -PCR primer pair for human-specific amplification of YACs and somatic cell hybrids. *Hum Mol Genet* **1**: 121-122.

- Tan S-S, Williams EA, and Tam PPL. 1993. X-chromosome inactivation occurs at different times in different tissues of post-implantation mouse embryo. *Nature Genet* **3**: 170-174.
- Tautz D. 1989. Hypervariability of simple sequences as a general source for polymorphic markers. *Nuc Acids Res* **17**: 6463-6471.
- Tautz D. 1993. Notes on the definition and nomenclature of tandemly repeated DNA sequences. In *DNA fingerprinting: State of the Science*, pp 21-28. Pena SDJ, Chakraborty J, Epplen JT and Jeffreys AJ eds. Birkhauser Verlag Basel, Switzerland.
- Tautz D and Renz M. 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nuc Acids Res* **12**: 4127-4138.
- Terwilliger JD and Ott J. 1994. *Handbook of Human Genetic Linkage*. The Johns Hopkins University Press, London.
- Tihy F, Vogt N, Recan D, Malfoy B, Leturcq F, Coquet M, Serville F, Fontan D, Guillard J-M, Kaplan J-C, Dutrillaux B, and Lemieux N. 1994. Skewed inactivation of an X chromosome deleted at the dystrophin gene in an asymptomatic mother and her affected daughter. *Hum Genet* **93**: 563-567.
- Tilley WD, Marcelli M, Wilson JD, and McPhaul MJ. 1989. Characterisation and expression of a cDNA encoding the human androgen receptor. *Proc Natl Acad Sci USA* **86**: 327-331.
- Trask BJ. 1991. Fluorescence *in situ* hybridization: applications in cytogenetics and gene mapping. *Trends Genet* **7**: 149-154.
- Trivier E, De Cesare D, Jacquot S, Pannetier S, Zackai E, Young I, Mandel J-L, Sassone-Corsi, and Hanauer A. 1996. Mutations in the kinase Rsk-2 associated with Coffin-Lowry syndrome. *Nature* **384**: 567-570.
- Turner G, Gedeon A, Mulley J. 1994. X-linked mental retardation with heterozygous expression and macrocephaly: pericentromeric gene localization. *Am J Med Genet* **51**: 575-580.

Uberbacher EC and Mural RJ. 1991. Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach. *Proc Natl Acad Sci USA* **88**: 11261-11265.

UTAH Marker Development Group. 1995. A collection of ordered tetranucleotide-repeat markers from the human genome. *Am J Hum Genet* **57**: 619-628.

Valdes JM, Tagle DA and 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**: 5377-5381.

van der Maarel SM, Rohme D, Bach I, Gilgenkranz S, Olde Weghuis D, Kere J, Cremers FPM and Ropers H-H. 1994. Cloning of a candidate gene for X-linked mental retardation in Xq13.1 *Cytogenet Cell Genet* **67**: 349.

van der Maarel SM, Scholten IHJM, Huber I, Philippe C, Suijkerbuijk RF, Gilgenkranz S, Kere J, Cremers FPM and Ropers H-H. 1996. Cloning and characterisation of, *DXS6673E*, a candidate gene for X-linked mental retardation in Xq13.1. *Hum Mol Genet* **5**: 887-897.

van Dyke DL, Wiktor A, Roberson JR, and Weiss L. 1991. Mental retardation in Turner syndrome. *J Pediatr* **118**: 415-417.

Verellen-Dumoulin Ch, Freund M, Meyer RDe, Laterre Ch, Frederic J, Thompson MW, and Markovic VD. 1984. Expression of an X-linked muscular dystrophy in a female due to translocation involving Xp21 and non-random inactivation of the normal X chromosome. *Hum Genet* **67**: 115-119.

Verkerk AJMH, Pieretti M, Sutcliffe JS *et al.* 1991. Identification of a gene (FMR-1) containing a CCG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* **65**: 905-914.

Villa A, Notarangelo L, Macchi P, Mantuano E, Cavagni G, Brugnoli D, Strina D, Patrosso MC, Ramenghi U, Grazia Sacco M, Ugazio A, and Vezzojini P. 1995. X-linked thrombocytopenia and Wiskott-Aldrich syndrome are allelic diseases with mutations in the WASP gene. *Nature Genet* **9**: 414-417.

Vyas P, Vickers MA, Simmons DL, Ayyub H, Craddock CF, and Higgs DR. 1992. Cis-acting sequences regulating expression of the human  $\alpha$ -globin cluster lie within constitutively open chromatin. *Cell* **69**: 781-793.

Wallgren-Pettersson C and Thomas NST. 1994. Report on the 20th ENMC sponsored international workshop: myotubular/centronuclear myopathy. *Neuromusc Disord* **4**: 71-74.

Wallgren-Pettersson C, Clarke A, Samson F, Fardeau M, Dubowitz V, Moser H, Grimm T, Barohn RJ, Barth PG. 1995. The myotubular myopathies: differential diagnosis of the X linked recessive, autosomal dominant, and autosomal recessive forms and present state of DNA studies. *J Med Genet* **32**: 673-679.

Wang Z and Weber JL. 1992. Continuous linkage map of human chromosome 14 short tandem repeat polymorphisms. *Genomics* **13**: 532-536.

Wapenaar MC, Petit C, Basler E, Ballabio A, Henke A, Rappold GA, van Paassen HMB, Blonden LAJ, and van Ommen GJB. 1992. Physical mapping of 14 new DNA markers isolated from the human distal Xp region. *Genomics* **13**: 167-175.

Wapenaar MC, Bassi MT, Schaefer L, Grillo A, Ferrero GB, Chinault AC, Ballabio A, and Zoghbi HY. 1993. The genes for X-linked ocular albinism (OA1) and microphthalmia with linear skin defects (MLS): cloning and characterization of the critical regions. *Hum Mol Genet* **2**: 947-952.

Wapenaar MC, Schiaffino MV, Bassi MT, Schaefer L, Chinault AC, Zoghbi HY and Ballabio A. 1994. A YAC-based binning strategy facilitating the rapid assembly of cosmid contigs: 1.6 Mb of overlapping cosmids in Xp22. *Hum Mol Genet* **3**: 1155-1161.

Watson JD. 1990. The Human Genome Project: Past, Present, and Future. *Science* **248**: 44-49.

Weber C, Oudet C, Johnson S, Pilia G, Schlessinger D, and Hanauer A. 1993. Dinucleotide repeat polymorphism at Xq26.1 (DXS1114). *Hum Mol Genet* **2**: 612.

Weber JL. 1990. Informativeness of human  $(dC-dA)_n.(dG-dT)_n$  polymorphisms.

*Genomics* 7: 524-530.

Weber JL and May PE. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am J Hum Genet* 44: 388-396.

Weber JL, Kwitek AE, May PE, Polymeropoulos MH, and Ledbetter S. 1990. Dinucleotide repeat polymorphisms at the DXS453, DXS454, DXS458 loci. *Nuc Acids Res* 18: 4037.

Weber JL, Polymeropoulos MH, May PE, Kwitek AE, Xiao H, McPherson JD, and Wasmuth JJ. 1991. Mapping of human chromosome 5 microsatellite DNA polymorphisms. *Genomics* 11: 695-700.

Weber JL, Wang Z, Hansen K, Stephenson M, Kappel C, Salzman S, Wilkie PJ, Keats B, Dracopoli NC, Brandriff BF, and Olsen AS. 1993. Evidence for human meiotic recombination interference obtained through construction of a short tandem repeat-polymorphism linkage map of chromosome 19. *Am J Hum Genet* 53: 1079-1095.

Weber JL and Wong C. 1993. Mutation of human short tandem repeats. *Hum Mol Genet* 2: 1123-1128.

Weeks DE, Ott J, and Lathrop GM. 1994. Detection of genetic interference: simulation studies and mouse data. *Genetics* 136: 1217-1226.

Weeks DE, Nygaard TG, Neystat M, Harby LD, and Wilhelmsen KC. 1995. A high-resolution genetic linkage map of the pericentromeric region of the human X chromosome. *Genomics* 26: 39-46.

Wehnert M, Reiner O and Caskey CT. 1993. Four STR polymorphisms map to a 500 kb region between DXS15 and DXS134. *Hum Mol Genet* 2: 1503.

Weissenbach J. 1993. A second generation linkage map of the human genome based on highly informative microsatellite loci. *Gene* 135: 275-278.

Weissenbach J, Gyapay G, Dib C, Vignal A, Morissette J, Millasseau P, Vaysseix G, and Lathrop M. 1992. A second generation linkage map of the human genome. *Nature* 359: 794-801.

Wengler G, Gorlin JB, Williamson JM, Rosen FS, and Bing DH. 1995. Nonrandom inactivation of the X chromosome in early lineage hematopoietic cells in carriers of Wiskott-Aldrich syndrome. *Blood* **85**: 3471-2477.

Whittaker PA, Campbell AJB, Southern EM, and Murray NE. 1988. Enhanced recovery and restriction mapping of DNA fragments cloned in a new  $\lambda$  vector. *Nuc Acids Res* **16**: 6725-6735.

Wilcox AS, Kahn AS, Hopkins JA and Sikela JM. 1991. Use of 3' untranslated sequences of human cDNAs for rapid chromosome assignment and conversion to STSs: implications for an expression map of the genome. *Nuc Acids Res* **19**: 1837-1843.

Willard HF, Waye JS, Skolnick MH, Schwartz CE, Powers VE, and England SB. 1986. Detection of restriction length polymorphisms at the centromeres of human chromosomes by using chromosome-specific satellite DNA probes: implications for development of centromere-based genetic linkage maps. *Proc Natl Acad Sci USA* **83**: 5611-5615.

Willard HF, Cremers F, Mandel JL, Monaco AP, Nelson DL and Schlessinger D. 1994. Report of the fifth international workshop on human X chromosome mapping 1994. *Cytogenet Cell Genet* **67**: 295-358.

Willems PJ. 1994. Dynamic mutations hit double figures. *Nature Genet* **8**: 213-215.

Willems P, Vits L, Buntinx I, Raeymaekers P, van Broeckhoven C, Ceulemans B. 1993. Localisation of a gene responsible for nonspecific mental retardation (MRX9) to the pericentromeric region of the X chromosome. *Genomics* **18**: 290-294.

Williamson R, Bowcock A, Kidd K, Pearson P, Schmidtke J, Ceverha M, Chipperfield M, Cooper DN, Coutelle, C, Hewitt J, Klinger K, Langley K, Beckmann J, Tolley M, and Maidak B. 1991. Report of the DNA committee and catalogues of cloned and mapped genes, markers formatted for PCR and DNA polymorphisms. *Cytogenet Cell Genet* **58**: 1190-1832.

Wilson M, Mulley J, Gedeon A, Robinson H, and Turner G. 1991. New X-linked syndrome of mental retardation, gynaecomastia and obesity is linked to DXS255. *Am J Med Genet* **40**: 406-413.

Winter RM, Davies KE, Bell MV, Huson SM, and Patterson MN. 1989. MASA syndrome: further clinical delineation and chromosomal localisation. *Hum Genet* **82**: 367-360.

Wolff DJ, Brown CJ, Schwartz S, Duncan AMV, Syrti U, and Willard HF. 1994. Small marker X chromosomes lack the X inactivation center: implications for karyotype/phenotype correlations. *Am J Hum Genet* **55**: 87-95.

Woodcock DM, Crowther PJ, Doherty J, Jefferson S, DeCruz E, Noyer-Weidner M, Smith SS, Michael MZ, and Graham MW. 1989. Quantitative evaluation of *Escherichia coli* host strains for tolerance to cytosine methylation in plasmid and phage recombinants. *Nuc Acids Res* **17**: 3469-3478.

Wyman AR and White R. 1980. A highly polymorphic locus in human DNA. *Proc Natl Acad Sci USA* **77**: 6754-58.

Yamaguchi Y, Ikenaka K, Niinobe M, Yamada H, and Mikoshiba K. 1996. Myelin proteolipid protein (PLP), but not DM-20, is an inositol hexakisphosphate-binding protein. *J Biol Chem* **271**: 27838-27846.

Yaspo M-L, Gellen L, Mott R, Korn B, Nizetic D, Poustka A, and Lehrach H. 1995. Model for a transcript map of human chromosome 21: isolation of new coding sequences from exon and enriched cDNA libraries. *Hum Mol Genet* **4**: 1291-1304.

Yates JRW, Goudie DR, Gillard EF, Aitken DA, Affara NA, Clayton JF, Tippett PA, and Ferguson-Smith MA. 1987. Multipoint linkage analysis of steroid sulfatase (X-linked ichthyosis) and distal Xp markers. *Genomics* **1**: 52-59.

Yee HA, Wong AKC, van de Sande JH and Rattner JB. 1991. Identification of novel single-stranded d(TC)<sub>n</sub> binding proteins in several mammalian species. *Nuc Acids Res* **19**: 949-953.

Young ID. 1988. The Coffin-Lowry syndrome. *J Med Genet* **25**: 344-348.

Yu S, Mulley J, Loesch D, Turner G, Donnelly A, Gedeon A, Hillen D, Kremer E, Lynch M, Pritchard M, Sutherland GR, and Richards RI. 1992. Fragile-X syndrome:

Unique genetics of the heritable unstable element. *Am J Hum Genet* **50**: 968-980.

Zhang Y, McMahon R, Charles SJ, Green JS, Moore AT, Barton DE, and Yates JRW. 1993. Genetic mapping of the Kallmann syndrome and X-linked ocular albinism gene loci. In Schlessinger D, Mandel J-L, Monaco A P, Nelson D L, and Willard H F (eds): Report of the 4th International Workshop on the Human X Chromosome Mapping 1993. *Cytogenet Cell Genet* **64**: 147-194 and *J Med Genet* **30**: 923-925.

Zhu Q, Zhang M, Blaese M, Derry JMJ, Junker A, Franke U, Chen S-H and Ochs HD. 1995. The Wiskott-Aldrich syndrome and X-linked congenital thrombocytopenia are caused by mutations of the same gene. *Blood* **86**: 3797-3804.

# Appendix



Donnelly, A., Kozman, H., Gedeon, A.K., Webb, S., Lynch, M., Sutherland, G.R., Richards, R.I., Mulley, J.C. (1994) A Linkage Map of Microsatellite Markers on the Human X-Chromosome.  
*Genomics*, v. 20 (3), pp. 363-370, April 1994

NOTE: This publication is included in the print copy of the thesis held in the University of Adelaide Library.

It is also available online to authorised users at:

<http://dx.doi.org/10.1006/geno.1994.1189>



Mulley, J.C. et al. (1995) *FRAXE and mental retardation*.  
*Journal of Medical Genetics*, v. 32 (3), pp. 162-169, March 1995

NOTE: This publication is included in the print copy of the thesis  
held in the University of Adelaide Library.

It is also available online to authorised users at:

<http://dx.doi.org/10.1136/jmg.32.3.162>



Andrew J. Donnelly, K.H. Andy Choo, Helen M. Kozman, Agi K. Gedeon, David M. Danks, and John C. Mulley (1994) Regional localisation of a non-specific X-linked mental retardation gene (MRX19) to Xp22.  
*American Journal of Medical Genetics*, v. 51 (4), pages 581–585, 15 July 1994

NOTE: This publication is included in the print copy of the thesis held in the University of Adelaide Library.

It is also available online to authorised users at:

<http://dx.doi.org/10.1002/ajmg.1320510457>



Donnelly, A.J., Partington, M.W., Ryan, A.K. and Mulley, J.C. (1996) Regional localisation of two non-specific X-linked mental retardation genes (MRX30 and MRX31).  
*American Journal of Medical Genetics*, v. 64 (1), pp. 113-120, July 1996

NOTE: This publication is included in the print copy of the thesis held in the University of Adelaide Library.

It is also available online to authorised users at:

[http://dx.doi.org/10.1002/\(SICI\)1096-8628\(19960712\)64:1<113::AID-AJMG19>3.3.CO;2-5](http://dx.doi.org/10.1002/(SICI)1096-8628(19960712)64:1<113::AID-AJMG19>3.3.CO;2-5)



Gedeon, A.K., Donnelly, A.J., Mulley, J.C., Kerr, B. and Turner, G. (1996) Letter to the editor: How many X-linked genes for non-specific mental retardation (MRX) are there?  
*American Journal of Medical Genetics*, v. 64 (1), pp. 158-162, July 1996

NOTE: This publication is included in the print copy of the thesis held in the University of Adelaide Library.

It is also available online to authorised users at:

[http://dx.doi.org/10.1002/\(SICI\)1096-8628\(19960712\)64:1<158::AID-AJMG26>3.0.CO;2-L](http://dx.doi.org/10.1002/(SICI)1096-8628(19960712)64:1<158::AID-AJMG26>3.0.CO;2-L)



Andrew Donnelly, Alison Colley, Denis Crimmins and John Mulley (1996) A novel mutation in exon 6 (F236S) of the proteolipid protein gene is associated with spastic paraplegia.

*Human Mutation*, v. 8 (4), pp. 384–385, 1996

NOTE: This publication is included in the print copy of the thesis held in the University of Adelaide Library.

It is also available online to authorised users at:

[http://dx.doi.org/10.1002/\(SICI\)1098-1004\(1996\)8:4<384::AID-HUMU17>3.0.CO;2-Z](http://dx.doi.org/10.1002/(SICI)1098-1004(1996)8:4<384::AID-HUMU17>3.0.CO;2-Z)



## Summary

We report a novel mutation, R259C, in exon b of the *MTMI* gene which is unusual in that it is associated with a relatively mild phenotype.

## ABSTRACT

The genetic basis of the relatively mild myopathic symptoms exhibited in a male was investigated. Mutation screening of a candidate gene, *MTMI*, represented a chance of establishing the molecular defect and the mode of inheritance. SSCA detected variation of the exon b PCR products from the proband and his mother, compared to that observed upon analysis of the PCR products from other members of the family and 159 unrelated X chromosomes. Sequencing revealed a C<sup>775</sup> to T transition, in the proband and his mother, but not in his unaffected brother. To confirm the presence of this base change, a *Cfo*1 site was introduced into the PCR product of the wildtype allele by using the forward primer 5'-AGAAAATAAGACGGTCATTGcG-3' (mismatch base in small font) with the exon b reverse primer as used by Laporte *et al* (1996). Analysis of DNA from other members of the family using this method revealed that either the proband's mother carries a new mutation or one of the proband's grandparents is a gonadal mosaic. This mutation would result in a Arg259Cys substitution. This is the first missense mutation within exon b of the *MTMI* gene to be reported in association with myotubular myopathy.

**MUTATION IN BRIEF**

**A novel mutation in exon b (R259C) of the *MTM1* gene is associated with myotubular myopathy.**

**Andrew Donnelly<sup>1,2</sup>, Eric Haan<sup>3</sup>, James Manson<sup>4</sup> and John Mulley<sup>1,2\*</sup>**

<sup>1</sup>Centre for Medical Genetics, Department of Cytogenetics and Molecular Genetics, Women's and Children's Hospital, <sup>2</sup>Department of Genetics, University of Adelaide, <sup>3</sup>Centre for Medical Genetics, Department of Medical Genetics, Women's and Children's Hospital and <sup>4</sup>Department of Neurology, Women's and Children's Hospital, Adelaide, South Australia, 5006.

\* To whom correspondence should be addressed:

Phone: (618) 08 8204 7111

Fax: (618) 08 8204 7342

email: [gsutherl@mad.adelaide.edu.au](mailto:gsutherl@mad.adelaide.edu.au)

Running title: MTM1 exon b mutation

Key words: MTM1, myotubularin, myotubular myopathy.

Address for correspondence:

72 King William Road

North Adelaide, SA 5006

Australia

Myotubular myopathy is a rare congenital myopathy which can be inherited in a X-linked recessive (MIM 310400), autosomal dominant (MIM 160150) and possibly autosomal recessive (MIM 255200) fashion (Wallgren-Pettersson et al., 1995). The X-linked form of the disorder (XLMTM) can be especially severe with polyhydramnios and reduced fetal movements noted antenatally, severe neonatal hypotonia, and generalised muscle weakness (including ptosis and weakness of the external ocular muscles) often resulting in respiratory failure and neonatal death in affected males. Female carriers of XLMTM have no overt clinical symptoms but may exhibit abnormal muscle histology (Wallgren-Pettersson et al., 1995). The X-linked form appears heterogenous with one locus in Xq28 (*MTMI*) (Laporte et al., 1996) and another excluded from Xq26-Xqter (Samson et al., 1995). Neither the two X-linked forms nor isolated X-linked and autosomal cases can be distinguished clinically or by muscle histology (Wallgren-Pettersson et al., 1995). Recently, Laporte et al. (1996) demonstrated that the *MTMI* gene in Xq28 is a putative tyrosine phosphatase and identified mutations in this gene in 7 out of 60 families with XLMTM. We report a novel mutation in *MTMI* which is unusual in that it is associated with a relatively mild phenotype.

The male proband was born at term after a normal pregnancy, labour and delivery with Apgar scores 4 at 1 minute and 6 at 5 minutes, birth weight 3665g, length 53cm and head circumference 35cm. The mother had noted reduced fetal movement compared to her previous two pregnancies. Poverty of movement, hypotonia and poor cry were present from birth. However, he fed adequately and thrived. Examination at 3 months of age demonstrated persistence of generalised muscle weakness, including facial weakness with bilateral ptosis, ophthalmoplegia involving vertical and horizontal gaze, and truncal and limb hypotonia. Muscle biopsy revealed the features of centronuclear myopathy with marked variability in fibre size, type I fibre atrophy (or hypotrophy), and increased numbers of central nuclei. When last reviewed at 3 years of age, it was noted that he had crawled between 9 and 10 months of age and had walked independently from 1 year. Generalised weakness persisted however, with ptosis, ophthalmoplegia involving vertical and horizontal gaze and a strongly positive Gowers' sign. He could walk 900m and though he could walk rapidly, he could not run. He was able to feed himself with a spoon and fork. Hypotonia persisted with significant head lag on being pulled to sit. The calves were tight. There was no evidence of progression of muscle weakness. Growth was normal and intelligence appeared to be within normal limits.

The family was too small to demonstrate linkage to any particular autosomal or X-linked locus, thus mutation screening of the *MTMI* gene in Xq28 (Laporte et al., 1996; GenBank Acc. No. U46024) represented a chance of establishing the molecular defect and hence the mode of inheritance. Five exons of the *MTMI* gene exons were amplified by PCR (using primers specified by Laporte et al., 1996) GenBank accession No. U46024 from the genomic DNA of the affected male, his mother and his unaffected brother. Single-strand conformation analysis (SSCA) detected variation of the exon b PCR products from the affected male and his mother, compared to that observed upon analysis of the PCR products from other members of the family and 159 X chromosomes from 47 unrelated males and 56 unrelated females.

Sequencing revealed a base transition, C<sup>775</sup> to T, in the affected male and his mother, but not in his unaffected brother as compared with published sequence (Laporte et al., 1996). This missense mutation did not create or destroy a restriction site. To confirm the presence of this base change, a *CfoI* site was introduced into the PCR product of the wildtype allele (Fig.1A) by using the forward primer 5'-AGAAAATAAGACGGTCATTGcG-3' (mismatch base in small font) with the exon b reverse primer of Laporte *et al.* (1996). Analysis of lymphocyte DNA from other members of the family using this method revealed that no others, including the proband's grandmother, carried the base change (Fig.1B). Either the proband's mother (III-4) carries a new mutation or one of the proband's maternal grandparents is a gonadal mosaic. The proband's maternal grandfather (II-2) was unavailable for study; however the possibility that he carries the C<sup>775</sup> to T base transition as a neutral variant can be excluded by its absence in two daughters (aunts of the proband) who are III-2 and III-3 in the pedigree (Fig.1B)

This mutation would result in an Arg to Cys substitution at codon 259. This amino acid residue has been conserved in the predicted homologous proteins of both *Saccharomyces cerevisiae* and *Caenorhabditis elegans* (Laporte et al., 1996). Given that a non-conservative (in terms of structure and evolution) amino acid substitution has occurred this mutation potentially has functional significance for the putative protein (myotubularin). This is the first missense mutation within exon b of the *MTM1* gene to be reported in association with myotubular myopathy. Another mutation in exon b, a deletion (AC763-764), has been reported and is expected to cause a frameshift and a truncated protein (Laporte et al., 1996).

The identification of this mutation has shown that the disease in the family studied is the form associated with *MTM1* in Xq28, as distinct from any other X-linked or autosomal myotubular myopathy. Precise genetic counselling can now be offered to the family on this basis. The females in the family (including the sister of the proband) who were at risk of carrying this mutation (III-2, III-3 and IV-3) are now excluded as carriers, (Fig.1B) and the option of prenatal diagnosis of future pregnancies is now available to the mother of the proband.

#### FIGURE CAPTION

Fig. 1 A: *MTM1* exon b PCR products digested with CfoI. The affected male has a 161bp uncut PCR product, non-carrier individuals have a 139 bp restriction fragment and a 22 bp fragment (not visible) and the female carrier has fragments of 139 bp and 22 bp (not visible) and 161 bp (uncut mutated allele).

B: The family studied, with results of mutation analysis + carrier, - noncarrier of the mutation.

## ACKNOWLEDGMENTS

This work was supported by the National Health and Medical Research Council of Australia.

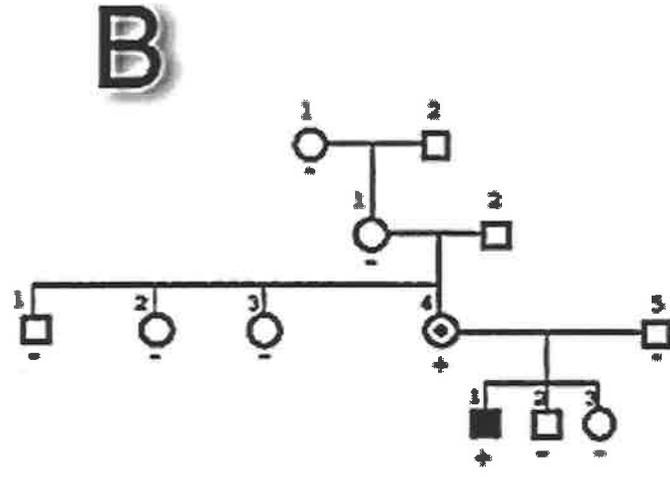
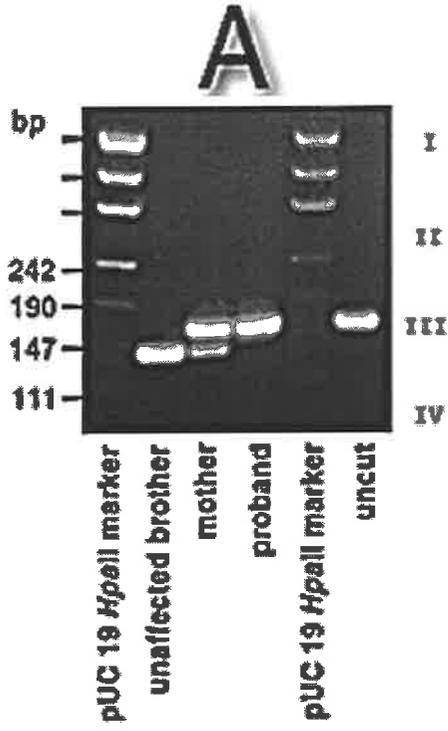
Thanks to Georgina Holloway for technical assistance.

## REFERENCES

Laporte J, Hu LJ, Kretz C, Mandel J-L, Kioschis P, Coy JF, Klauck SM, Poustka A and Dahl N. (1996). A gene mutated in X-linked myotubular myopathy defines a new putative tyrosine phosphatase family conserved in yeast. *Nature Genet* 13: 175-182. Medline I.D. 96225444.

Samson F, Mesnard L, Heimburger M, Hanauer A, Chevally M, Mercadier JJ, Pelissier JF, Feingold N, Junien C, Mandel J-L, and Fardeau M. (1995). Genetic linkage heterogeneity in myotubular myopathy. *Am J Hum Genet* 57: 120-126. Medline I.D. 95335665.

Wallgren-Pettersson C, Clarke A, Samson F, Fardeau M, Dubowitz V, Moser H, Grimm T, Barohn RJ and Barth PG. (1995). The myotubular myopathies: differential diagnosis of the X-linked recessive, autosomal dominant, and autosomal recessive forms and present state of DNA studies. *Med Genet* 32: 673-679. Medline I.D. 96091867.





Addendum in response to examiner NL:

Page X. "Idiogram" is preferable to "ideogram". Medical dictionaries and the International System for Cytogenetic Nomenclature [*Cyto Cell Genet* 21p309, (1978)] mention idiogram as a "graphic representation of a karyotype".

Figures 8.11 and 8.17. A possible explanation is that the faint intensity of the band of lower molecular weight seen in the "mother" track in *both* photos could be explained by exhaustion of the primers before the last cycle of the PCR. If so, during the last cycle, renaturation of the denatured strands will result in 3 types of PCR products: wildtype/wildtype; wildtype/mutant; mutant/mutant. Only the wildtype/wildtype product, representing about 1/3 of all the PCR products, will be a substrate for the restriction enzyme. This could be investigated by reducing the number of cycles, or increasing the concentration of primers, and digesting the PCR products.

Figure 8.13. Only those members of the family known by the candidate to be recorded as deceased were shown as such. Those members not shown as deceased are not necessarily still alive.

#### Amendments

Page 1, para 1: *Drosophila melanogaster*

Page 2, para 6: *Drosophila melanogaster*

Page 7, para 3: "to be isolated"

Page 9, para 3: "occurrence"

Page 32, para 4: "Qiagen"

Page 33, para 1: w/o = without

Page 45, para 3: "occurred"

Page 46, para 6: "incubated at 16"

Page 49, 1st line: "3MM Whatman"

Figure 4.3.2: "Discussion" should be replaced by "possibly due to M13 rearrangement - see Results page 82"

Page 106, para 4: "individual who was a new mutant" should be replaced by "individual who had a new mutation".

Page 137, para 1: "number of families"

Figures 6.4 and 6.6: The line connecting II-7 should originate from the line connecting I-2 and I-3 and then bend around the line connecting II-7 to her sibs.

Figure 6.10. The haplotype of I-2 can be re-drawn so that II-2, II-3 and II-12 are no longer double recombinants.

Instead I-2, II-3, II-5 and II-12 are now single recombinants.

Page 190, para 1: Person II-8 should be II-9.

Figure 8.13: Only those members of the family known by the candidate to be recorded as deceased were shown as such. Hence those members not shown as deceased are not necessarily still alive.

Page 289, para 2: "the mother is a new mutant" should be replaced by "the mother has a new mutation"

Appendix 7, page 5, para 2: "is a gonadal mosaic" should be replaced by "has gonadal mosaicism".