MAPPING GENES FOR X-LINKED DISORDERS

by

Ági Kyra Gedeon B.Sc.(Hons.)

This thesis is submitted as the complete requirement for the degree of Doctor of Philosophy within the Department of Genetics, Faculty of Science, at The University of Adelaide, South Australia.

September 1996

Declaration

This thesis contains no material that has been accepted for the award of any other degree or diploma in any University or other tertiary institution, nor to the best of the candidate's knowledge any material previously published or written by another person, except where due reference is made in the text. Throughout this volume any reference to 'the candidate' signifies the author Agi Kyra Gedeon. The work described in this thesis has been performed by the candidate or under the candidate's supervision, except where duly acknowledged in the text.

NAME: Agi K. GEDEON **COURSE:** Ph.D. (Science)

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

SIGNATURE: ... **DATE:** 28.10.96

Table of Contents

Abbreviations

5'UTR	-	5' untranslated region of a gene
$(AC)_n$	-	a dinucleotide repeat sequence comprised of the bases AC repeated n times
$(CCG)_n$	-	a trinucleotide repeat sequence comprised of the bases CCG repeated n times
A, C, G, T	-	the four nitrogenous organic bases; adenine, cytosine, guanine and thymine
ANGIS	-	Australian National Genomic Information Service
bp	-	basepairs; deoxyribonucleoside pairs
BFLS	-	Börjeson-Forssman-Lehmann syndrome; a syndromal XLMR (Chapter 5)
BTHS	-	Barth syndrome; X-linked cardiomyopathy (Chapter 6)
cDNA	-	complementary DNA or copy DNA; reverse transcribed *in vitro* from mRNA
CEPH	-	Centre d'Étude du Polymorphisme Humain in France
cM	-	centimorgan; the unit of genetic distance
CVS	-	chorionic villus sample
DMD	-	Duchenne muscular dystrophy
DNA	-	deoxyribonucleic acid
DXS...	-	standardised D alphanumeric locus symbol representing a unique segment of the X chromosome identified by a cloned DNA sequence
EST	-	expressed sequence tag; partial sequence of a cDNA clone corresponding to an mRNA.
FISH	-	fluorescent *in-situ* hybridisation
FMR1	-	gene symbol for Fragile X mental retardation 1 gene at FRAXA
FMR2	-	gene symbol for Fragile X mental retardation 2 gene at FRAXE
FraX	-	fragile X syndrome at the FRAXA fragile site
FRAXA	-	folate sensitive fragile site at Xq27.3
FRAXE	-	folate sensitive fragile site in Xq28
FRAXF	-	folate sensitive fragile site in Xq28
GDB	-	Genome Data Base
$(GT)_n$	-	complementary strand to an $(AC)_n$ sequence
HGP	-	Human Genome Project
IQ	-	intelligence quotient; a standardised measure of intellectual functioning
kb	-	kilobase; 1000 basepairs of DNA
M	-	Morgan; 1000 centiMorgans
mA	-	milliampere; unit of electrical current
Mb	-	megabase; 1000 kilobases of DNA
MIM	-	Mendelian Inheritance in Man; a listing and numeric code for human disease genes
mRNA	-	messenger RNA; transcript of a gene and serves as a template for translation
MRX	-	mental retardation, X-linked; used for non-specific forms only (Chapter 4)

MRXS	-	mental retardation, X-linked, syndrome; interim symbol used for syndromal forms only (Chapter 5)
OD_{260}	-	optical density at a wavelength of 260nm
OD_{600}	-	optical density at a wavelength of 600nm
OMIM	-	on-line version of the Mendelian Inheritance in Man (MIM) database
PAGE	-	polyacrylamide gel electrophoresis
PAR	-	pseudoautosomal region of the X chromosome
PC	-	personal computer
PCR	-	polymerase chain reaction; an *in vitro* process for DNA amplification
PDR	-	X-linked reticulate pigmentary disorder (Chapter 6)
PFGE	-	pulsed-field gel electrophoresis
PIC	-	polymorphic information content
POWCH	-	Prince of Wales Children's Hospital, Sydney, New South Wales
PRTS	-	Partington syndrome; XLMR with dystonic movements of the hands (Chapter 5)
RFLP	-	restriction fragment length polymorphism
RNA	-	ribonucleic acid
SHS	-	Sutherland-Haan syndrome; XLMR with spastic diplegia (Chapter 5)
SSCA	-	single stranded conformational analysis
SSCP	-	single stranded conformational polymorphism
STR	-	simple tandem repeat
STS	-	sequence tagged site; a short, unique DNA sequence
VNTR	-	variable number of tandem repeats
v/v	-	volume per volume
w/v	-	weight per volume
WCH	-	Women's and Children's Hospital, Adelaide, South Australia
WTS	-	Wilson-Turner syndrome; XLMR with gynaecomastia and obesity (Chapter 5)
XCW	-	X chromosome workshop
XLMR	-	X-linked mental retardation; general term
XS	-	X-linked Lutheran suppressor gene (Chapter 6)
YAC	-	yeast artificial chromosome; used for cloning large DNA fragments

Acknowledgments

This project was carried out in the Department of Cytogenetics and Molecular Genetics at the Women's and Children's Hospital in Adelaide, South Australia. The work was done on a part-time basis alongside the full-time diagnostic commitment to families with fragile X syndrome, myotonic dystrophy, fascio-scapulo-humeral dystrophy and miscellaneous other rare genetic disorders including those described in this thesis. I gratefully acknowledge the participation of the families and their clinicians in the study. It is particularly gratifying that many of the results in this thesis have already, or are being, put to use in the genetic counselling of the family members.

I wish to express appreciation to Professor Grant R. Sutherland, my supervisor and head of the department, for the opportunity to complete this study under his guidance. Many thanks also to Dr. Rory Hope, my supervisor, Department of Genetics, University of Adelaide and Dr. John Mulley, GM unit section head, who participated in many useful discussions and gave helpful advice on several aspects of this project.

Several clinicians entered into this study by contributing families. I am indebted to them all, but outstandingly to Drs. Gillian Turner and Bronwyn Kerr then of the Prince of Wales Children's Hospital, as well as Drs. Lesley Ades and Meredith Wilson then of Camperdown Children's Hospital. Hazel Robinson, Sam Wake and the Sydney team, have been reliably helpful and pleasant associates. I have great respect for Dr. Michael Partington, Dr. Gillian Turner and Dr. Eric Haan and am honoured to have worked with each of them on some of the families presented here.

The contributions of specialised skills and extra hands to this work by my colleagues Elizabeth Baker, Andrew Donnelly, Helen Kozman and Jozef Gecz are gratefully acknowledged. The opportunity to work together with colleagues toward a common goal have been most valuable and insightful.

My thanks to my supportive family who hold me up from all corners of the globe, especially my husband, Dietmar Blanke, who I lean on the heaviest. My thanks also to my baby daughter, Dominique, who peacefully gestated whilst I was in the throes of writing-up, and continued to be a model baby until thesis submission. I dedicate this effort to the fond memory of my grandfather - Töled jön az erö folytatni s befejezni. Nektek tartozok mindennel, irtoan szeretlek Benneteket.

I am grateful to all for their immense patience with me and their unfailing support and encouragement when the going got rough. Apart from my family, my team of moral supporters included Lesley Ades, Bryone Kuss, Jozef Gecz, Sinoula Apostolou, Kathy Friend and numerous others who periodically boosted my flagging energy towards the end.

A Salient Quote

"If.....IQ is due to sex-linked genes, it is of more importance that a boy should have a clever mother than a clever father."

Hogben, 1932
(quoted in Lehrke, 1974)

Summary

The human genome project aims to sequence the genome, identify genes and thus the causes of inherited disease. The research presented in this thesis reports contributions to the development of the evolving map of the X chromosome, and the localisation and genetic delineation of rare X-linked monogenic neuropsychiatric diseases and other X-linked disorders, culminating in the identification of two new genes.

New polymorphic PCR-based markers were characterised from probes of known location (DXS102, DXS237, DXS294 and DXS300). Three were physically mapped between somatic cell hybrid breakpoints and on a YAC contig of Xq25-27. Development of a regional genetic map contributed to the integration of maps in Xq26. 40 CEPH reference families were genotyped for these markers and another, DXS538, and added to a comprehensive genetic map constructed from the CEPH database (ver 5.0). Additional PCR-based markers gathered from the literature, were ordered by comparisons with other genetic and physical maps, to create a composite background map spanning the entire X chromosome to facilitate localisation of disease genes by linkage.

The disproportionate excess of retarded males in the population may be accounted for by a number of X-linked disorders with mental retardation as a feature. The fragile X syndrome is responsible for less than half of these. To determine the number and distribution of other genes involved in XLMR, fifteen non-FRAXA families segregating mental retardation were collected into two distinct classes: i) MRX, where the mental retardation was the sole and major feature in otherwise clinically apparently normal individuals and ii) MRXS, with associated clinically distinct features. The disease gene in each family was localised by demonstration of significant linkage to markers in an interval flanked by recombinant events. These linkage studies identified 7 distinct regions of the X chromosome involved in MRX and confirmed the genetic heterogeneity of MRX.

The genes for five clinically distinct mental retardation syndromes (PRTS, BFLS, WTS, SHS and XLMR with macrocephaly) were similarly localised or their localisations refined, but overlapping localisations with MRX genes cannot separate these entities from allelic mutations in the same genes. Two candidate loci, identified by position, were analysed by single stranded conformation analysis (SSCA) and sequencing to detect mutation. One at SOX3 in the BFLS region did not yield any differences between normals and BFLS affecteds, the other at the DMD brain promoter detected a C to T base substitution in a normal male from an MRX family and does not appear to have a deleterious effect within this highly conserved promoter.

Other genes for X-linked disorders were genetically localised using the assembled resources and the linkage approach. New localisations include the genes for a pigmentary disorder with systemic manifestations (PDR) previously known as X-linked cutaneous amyloidosis, a Lutheran blood group suppressor (XS) and two forms of cardiomyopathy; Barth syndrome (BTHS) and the possibly allelic form of fatal infantile cardiomyopathy. Localisation of BTHS has lead to collaboration resulting in the identification of the BTHS gene (G4.5) by the positional candidate approach. The disease-causing mutation has been identified in the BTHS family and work is under way to investigate allelism of fatal infantile cardiomyopathy.

Males with submicroscopic deletions of the X chromosome can be valuable for recognising and isolating disease genes, as they identify all the phenotypic associations with the deleted gene(s). Deletion of the FMR1 gene and adjacent sequences was associated with only the characteristic fragile X phenotype in one boy. This was the first case described with fragile X syndrome due to deletion of FMR1 and supported the hypothesis that FMR1 was indeed the gene solely involved in the fragile X syndrome. Deletions detected in Xq28, over 600kb distal to FMR1, were associated with global developmental delay in one boy and with speech delay in another. Characterisation of these deletions and demonstration that the deleted sequence was conserved in evolution, lead to the isolation of the FMR2 gene at FRAXE. This gene is the eighth locus responsible for non-specific X-linked mental retardation (MRX) identified in this study. Others have identified further MRX loci in contiguous gene syndromes at Xp22.3 and Xq21 and a translocation breakpoint in Xq13 to physically discrete intervals. Up to 11 distinct regions/genes causing X-linked mental retardations (including FMR1) have now been established confirming the genetic heterogeneity of XLMR.

These studies have focused on gene localisation in X-linked disorders, particularly those with mental retardation. Genetic risk estimation has been made available to family members when required. A significant contribution has been made to the gene map of the X chromosome and to the delineation of genes responsible for XLMR. At least 10 discrete MRX genes, distinct from FMR1 and each other, account for a proportion of the excess of non-FRAXA retardation in the male population. The regional localisations described upon the background map of the X chromosome, include ten loci for MRX, five syndromal forms of XLMR and four other X-linked conditions. The physical characterisation of the deletion in each of three boys and maintenance of fruitful international collaborations have made possible the identification and eventual cloning of the disease genes involved in two of the clinical conditions studied and have contributed to the mapping of the X chromosome.

Note regarding Publications

Most of the work completed during this project has been published. The published manuscripts have been cited in the text and the papers are included in the Appendices. It is recommended that the publications relevant to each chapter are read first or concurrently with the chapter since some of the data appears only in one or other text to avoid extensive repetition.

CHAPTER 1

Gene Mapping: Review of the literature

1.1. Introduction

The human genome project (HGP), an extensive international collaboration established to map the human genome, has brought about significant advances in the past decade. It has had a fundamental role in the development of technology, particularly in the field of molecular biology, and influenced new approaches to defining genes. Mapping of specific disease genes contributes to completion of the map of the human genome and aids in determination of normal gene function through examination of the phenotypes of disease-causing mutations. Understanding gene functions and interactions will ultimately have a role in ameliorating human disease through gene therapy for example, but will also be fundamental in supporting theories of evolution through comparative studies in other species.

Since around 3% of the population are mentally retarded, and given the human preoccupation with intelligence and its determinants, there is much interest in genes involved in brain function. Genes on the sex chromosomes are of particular interest in that they account for the excess of males in the retarded population and are easily recognised due to their characteristic mode of inheritance. In X-linked recessive diseases, half of the sons of a female heterozygous at the disease locus will be affected, and half of the daughters will be carriers at risk of having affected sons themselves. Mapping of X-linked disease genes can provide the basis for prenatal diagnosis and informed decision making on pregnancy outcomes rather than termination of all male fetuses in such families. Identification of disease genes and their distribution on the X chromosome thus promotes a beneficial service by medical scientists delivering information to families on carrier status and prenatal risk assessments and ultimately contributes to the erudition of the human genome.

This thesis presents a molecular genetic approach to the mapping and delineation of a number of X-linked monogenic neuropsychiatric disorders and others with unknown biochemical defect. Gene mapping proceeds by detection of linkage to genetic markers of known location that can be shown to cosegregate with the disease phenotype. Localisation of disease genes by linkage to chromosomal regions, is the primary step toward the isolation of genes when the gene product is not known and no other source of localisation eg. deletion or translocation, is available. Gene identification by positional cloning is possible once the putative gene interval is minimised and amenable to physical mapping. Candidate genes may be identified if already known to lie in the interval. Close examination of these positional candidates can hasten identification of the mutation responsible for the disease. Application of a strategy for mutation detection in a candidate gene is outlined in Chapter 4, and

demonstrates one direct approach to the characterisation of mutation(s) giving rise to specific disease phenotypes.

Linkage analysis makes use of genetic variation to gather evidence for genomic positioning of the disease gene segregating in a family. The power of the linkage approach is chiefly dependent on the number of potentially informative meioses available and the density of polymorphic loci. The size of the family, for diseases found in single families, is a critical criterion. Genes mapped in this thesis were based on linkage studies in large discrete families (Chapters 4, 5 and 6), except for those described with deletions (Chapter 7). Development of a new generation of polymorphic genetic markers ordered on the X chromosome has had considerable impact on this study (Chapter 3). This introductory chapter provides the background philosophies highlighting the technological revolution which occurred during the period of PhD candidature and concludes by outlining the objectives of the project.

1.2. Mapping the Human Genome

The complete genetic material for any cell is termed the genome and comprises the sequence of basepairs for nuclear and mitochondrial genes interspersed by non-coding sequences. The nuclear DNA sequence defining the human genome is carried on a set of 46 chromosomes and is the main library for heredity and genetic variability of the species. Differences in DNA sequences between individuals are the basis for genetic variation and etiology of genetic disease. By mapping the human genome some of the bases of variability and speciation (section 1.2.1.) may be revealed.

Many scientists through time have had an interest in individuality and heredity of traits, notably they include Darwin, who in 1859 developed a theory of evolution based on natural selection and Mendel, who in 1865 observed a regularity in the transmission ratios of traits in peas. Much of the basis of modern genetics was already hypothesised by 1909, with subsequent works establishing the foundation (Smith, 1986). The laws of segregation and of independent assortment formulated by Mendel have become fundamental principles of genetics. It is the demonstrable adherence of single gene traits to Mendel's law of segregation that is today referred to as Mendelian inheritance. Evidence of Mendelian inheritance is often the only way that phenotype can be connected with genotype. Morgan (in 1910), established evidence for crossing over between genes and calculated recombination fractions by breeding experiments in vinegar flies (*Drosophila melanogaster*). He noted reduced crossovers or co-inheritance of some traits (ie. departure from the law of independent assortment) which is now referred to as linkage. Shortly after (1913), one of his students, Sturtevant, constructed a linear map of *Drosophila* genes based

on recombination frequencies; the first 'genetic-linkage map'. Since in humans it is not possible to do controlled-breeding experiments to analyse thousands of offspring, linkage maps are based on estimating genetic distances from recombinants in subsets of families (section 1.2.2.).

Maps of the chromosomes today include genetic maps based on meiotic recombination rates between polymorphic loci, as well as cytogenetic, physical, and functional maps derived by several different experimental approaches, discussed further in section 1.3. By compiling and integrating these maps with differing levels of resolution (section 1.2.3.), a complete map of the human genome will result.

1.2.1. The Human Genome Project

A consolidated 15 year effort, to construct genetic and physical maps and ultimately to sequence the human genome, was commenced in 1990 and has been called the Human Genome Project (HGP) (Collins and Galas, 1993). The short-term goal of the HGP envisions unambiguously ordered marker resolution at 2-5cM on the genetic map and at 100kb on the physical map, as well as gene identification and sequencing of 50Mb/year by the year 1998 (Collins and Galas, 1993). The concomitant development of new biological tools to rapidly study the genome is aimed at the eventual identification of all genes in these maps and in the complete sequence. An understanding of human diversity in normal and inherited disease populations will unfold as a direct consequence of, and an important reason for, the HGP (Cavalli-Sforza et al, 1991). The HGP is an unprecedented international collaboration with the potential for significant impact on world human health. Knowledge of the mechanisms and etiology of disease genes could have a profound influence on global health by permitting actions to prevent or ameliorate and possibly eliminate common human disorders.

A selected number of non-mammalian model organisms including bacteria *Escherichia coli* (3 million bp), yeast *Saccharomyces cerevisiae* (14 million bp), the nematode *Caenorhabditis elegans* (80 million bp) and the vinegar fly *Drosophila melanogaster* (165 million bp) have been included for close comparative analysis under the umbrella of the HGP (Collins and Galas, 1993). The DNA sequence of model systems can be useful in deducing genes that guide development or define the characteristic phenotype of each organism. The function of a human protein may be inferred from the amino-acid sequence by analogy with genes in experimental organisms. The house mouse *Mus musculus* (3 billion bp) is a model organism with a particularly important contribution to the HGP. Conserved linkage groups between mouse and man allow the identification of mutants, and/or candidate genes that have homology with human disease genes (Brown, 1992). Controlled genetic crosses and new mutations can be produced in mice to provide more

resources for the rapid cloning and functional analysis of a gene. Transgenic mouse models can be created with specific alterations or total disruptions to a gene eg. knockout mice. A mutant mouse provides a model to study the pathophysiology and potential therapies for human diseases. Comparative mapping can also contribute to understanding chromosome evolution as linkage groups conserved since the divergence of the rodent and primate species, imply constraint on changes to the sequence and hence indicate essential biological function. The great evolutionary distance between yeast, *S.cerevisiae,* and man (~1 billion years) is the most compelling argument that regions of high sequence homology must code for proteins involved in the basic structural or functional cellular mechanisms (Tugendreich et al, 1994).

One of the primary HGP objectives was to determine the nucleotide sequence of the human genome (~3 billion base pairs) and thus identify all genes, originally estimated (in 1990) to number ~50,000-100,000 (section 1.2.3.). By 1992, the emphasis had shifted towards the partial sequencing of complementary DNA (cDNA) clones to produce expressed sequence tags (ESTs) in specific tissues, with some research groups concentrating on the brain (Adams et al, 1991; Khan et al, 1992), and others aiming to ultimately produce a 'body map' of expressed genes (Okubo et al, 1992). Libraries of ordered cDNA clones with known patterns of expression in different tissues permit access to sequences responsible for disease. Full sequence information (human and comparative) will provide insight into functional mechanisms, the role of non-coding DNA and possibly speciation. Sequence variation in genomic DNA of different human populations will provide information on human evolutionary history and genetic mechanisms such as selection, mutation and genetic drift (Bowcock et al, 1991).

Significant technological developments arising from the HGP include computer hardware and software improvements such as online databases, for example the Genome Database (GDB), that can handle vast bodies of data (Pearson, 1991), more efficient laboratory methodologies and equipment particularly the automation of many procedures through laser and robotics. Ethical, legal and social issues arising from the HGP, though fascinating, are not dwelt on in this thesis.

1.2.2. CEPH

The Centre d'Etude du Polmorphisme Humain (CEPH) international collaboration to genetically map the human genome was organised in 1984 in order to produce a primary genetic map of the genome. The aim was to place polymorphic markers at approximately 20cM intervals along each chromosome in a reference population (Dausset et al, 1990). The

more recent aim of the CEPH collaboration is to construct high resolution genetic maps with intervals less than 5cM to improve utility for gene localisation (Cann, 1992).

The CEPH reference panel consists of lymphoblastoid cell lines from 40 families with large sibships, spanning three generations. These families, with a total of 656 potentially informative meioses, allow determination of parental phase and therefore recombination frequencies between loci. The relative order and distance between markers can thus be established on a genetic linkage map. Although the panel has been expanded to 61 families to improve map resolution (Dausset et al 1990), many genetic maps are generated from the original set of 40 pedigrees (Cann, 1992) or a subset of these (Weissenbach et al, 1992; Gyapay et al, 1994).

DNA samples are distributed to participating investigators in exchange for genotype data that is accumulated in the CEPH database (currently ver 7), the rationale being that whole chromosome maps produced from data on the same group of families should determine more accurate map distances between markers and be more rapidly created through contributions to a common dataset from a large number of laboratories. Mutations in the lymphoblastoid DNA of the CEPH pedigrees have been reported (Royle et al, 1993, Banchs et al, 1994), however, and are estimated to occur at a rate between 0.001-0.0001 per gene per generation, mainly due to clonal evolution *in vitro* during cell culture (Weber and Wong, 1993). Germline and somatic mutation events *in vivo* can be difficult to distinguish in three generation families, suggesting the germline mutation rate may be as high as $3.9x10^{-4}$ per locus per gamete (Banchs et al, 1994). The true germline mutation rate of *de novo* mutations at microsatellite loci is estimated to be less than $1x10^{-4}$ in DNA directly obtained from cells of an individual, whereas the somatic mutation rate in DNA from cell lines is much higher (10^{-2} -10^{-3}) (Banchs et al, 1994). These mutations may be detected as apparent double crossovers over very small map distances, but may corrupt linkage data for genetic map construction when phase is incorrectly assigned and introduce errors that cause false expansion of the linkage map (Royle et al, 1993). High resolution microsatellite maps may become difficult to generate in CEPH families due to the higher mutation rate in DNA from cell lines (Banchs et al, 1994).

In 1989, there were 63 collaborating laboratories contributing to the worldwide CEPH effort (Dausset et al, 1990), by 1992 there were 88 (Cann, 1992). One of the largest collaborating groups, Généthon, have mapped 2066 markers covering 3690cM of the genome with only one gap greater than 20cM (Gyapay et al, 1994). Several chromosome-specific genetic maps have been completed by CEPH consortia with that of chromosome 10, including the extended panel of families, being the first published (White et al, 1990). Consortium maps represent the consensus genetic marker order determined by linkage and

are therefore reliable background maps for gene mapping and for integration with physical maps of the genome.

1.2.3. Genome Statistics

The unit of genetic distance is the morgan, more usually expressed in centimorgans (cM), and is the probability of crossing over between loci on a chromosome. Markers 1cM apart are separated by a distance such that on average 1% (0.01) recombination occurs in that interval. In humans, under the assumption that crossing over is constant along the length of a chromosome, 1cM roughly corresponds to a physical distance of 10^6 basepairs or one megabase (Mb). In reality, the probability of crossing over varies with chromosomal location (section 1.3.1.) and sex of the individual (females have a higher probability). The physical length of the haploid human genome is estimated to comprise ~3 x 10^9 base pairs (3227Mb - Morton, 1991) and spans an estimated sex-averaged autosomal genetic length of 3300cM based on cytogenetic counts of chiasmata (Renwick, 1969). Genetic lengths estimated from linkage analyses range from 4074cM (Morton, 1991) to 4910cM (NIH/CEPH, 1992). Chiasma counts are known to under-estimate genetic map length, while estimates from linkage analysis may be inflated due to genotyping errors (NIH/CEPH, 1992) or use of an inappropriate mapping function (Morton, 1991). Assuming the errors in the two methods of determining genetic map length compensate for one another, the total genetic map length is likely to be around 4000cM.

The number of genes in the genome can depend on the definition of a gene as containing a transcription unit, or as the translated regions relating to phenotype-causing mutation (Carlson, 1991). Estimates of the number of genes have been based on approximating that 100,000 genes with an average size of 30kb arranged end to end span the $3x10^6$kb of the haploid genome (Fields et al, 1994). Since gene sizes are variable eg. DMD=2.4Mb, while others overlap one another, and their distribution is known to be concentrated in the GC rich R bands (section 1.3.2.), the total number of genes cannot be predicted but 50-100,000 is usually quoted. Up to 30,000 genes may be expressed in the human brain, which reflects the complexity of this organ and the involvement of neurological dysfunction in many genetic diseases (Adams et al, 1991). Experimental approaches to provide direct estimates of gene number can include RNA reassociation kinetics or CpG island detection or sequence comparisons of expressed sequence tags (Fields et al, 1994). Fields (1994) estimate one gene per 40 or 50kb given that 2/3 genes have CpG islands and gene poor regions have 1/10 genes compared to gene rich regions - this results in a total estimate of 60-70,000 genes in the human genome.

Around 1600 different protein coding sequences had been cloned from the human genome by 1990, and a further 800 reports of cloned human DNA were recorded during the following year (Schmidtke and Cooper, 1990 and 1991) many only as partial cDNA or expressed sequence tags (ESTs). In 1994, nearly 60,000 genes and markers are recorded on GDB and represent the logarithmic phase of gene mapping propelled by the new class of polymorphic marker (section 1.4.2.) and the genome project.

The analysis of a gene responsible for a disease phenotype may begin with the search for the gene localisation in 3,000,000kb, but ultimately aims to identify specific changes in only one or a few basepairs in a coding sequence of a few kilobases. For X-linked disorders chromosome localisation is inherently achieved as a consequence of their easily recognisable inheritance pattern (section 1.5.1). Morton (1991) estimated the physical length of the X chromosome to be 164Mb, roughly equivalent to 5% of the total genome. The current genetic estimates of X chromosome length are approximately 210cM (Schlessinger et al, 1993), just over 5% of the ~4000cM human genome. By this premise, if there are an estimated 50,000-100,000 genes in the human genome, between 2500-5000 of these would be expected on the X chromosome assuming a random distribution of genes.

1.3. Gene Mapping

Over ~6600 Mendelian disorders are known to be caused by defects in a single gene (McKusick, 1994). A gene can be cloned by the functional approach (if the function is known) and then mapped, or from the map position, termed positional cloning, then functionally analysed. Locating a gene responsible for an inherited disease to a position on a chromosome is often the first step towards its isolation and can be achieved by genetic or physical means.

In the absence of a clear mode of inheritance where the gene product or the causative mutation is unknown, nonparametric methods of linkage analysis can be applied to determine identity by descent. Linkage mapping under an assumed model (parametric analysis) is usually applied in diseases that can be demonstrated to be due to a single mutant gene traced through generations of affected families and is the subject of this thesis. Using the genetic approach in families segregating a particular disease phenotype, the disease gene can be localised to a specific chromosomal interval by observing linkage or co-inheritance with markers of known position (section 1.3.1.). The markers most often inherited with the disease are physically the closest to the causative gene and can be used to locate and clone DNA from adjacent portions of the physical map (an ordered set of clones). Analysis of candidate genes by position in such a gene interval can circumvent often prolonged physical characterisation of these intervals necessary for the positional cloning of a disease gene.

Physical mapping of a specific disease gene relies either on knowledge of its approximate position in the genome or its biological function. Approaches to the mapping and cloning of genes with known biochemical defect, thus known biological function, take advantage of knowledge of the amino acid sequence of the protein and possibly antibodies against the gene product. Functional cloning often involves the reverse transcription of mRNA to produce cDNA. The isolated cDNA is then physically mapped on somatic cell hybrids or by fluorescence *in situ* hybridisation (FISH) for the cytogenetic placement of the locus. Characterisation of genes of unknown etiology and in the absence of disease association is technically possible. Various approaches to the detection and mapping of transcribed sequences including exon trapping and hybridisation-based direct cDNA selection are being developed to construct a transcriptional map of the human genome (Hochgeschwender, 1992). Ultimately integration of data from genetic, physical and transcriptional maps will provide information on the position, sequence and expression of a gene.

The differences in these approaches are that for 'genetic' or linkage mapping, phenotypic and genotypic family data are required, whereas for the 'functional' approach individual or unrelated patient samples are sufficient.

1.3.1. Genetic Mapping

Gene localisation by linkage is straightforward when a single gene with a defined Mendelian mode of transmission is responsible for the disease phenotype. Classic techniques of linkage analysis together with genotype data at polymorphic markers (section 1.4.), result in the localisation of genes implicated in disease. Computer programs utilising appropriate algorithms hasten such calculations (Chapter 2). The requirements for the parametric approach to linkage analysis include; a) sufficient potentially informative meioses from several clinically and genetically homogenous smaller families or one large family segregating the disease phenotype, b) a genetic background map of sufficient density to locate the disease gene to within a few centiMorgans, c) a recognisable Mendelian mode of inheritance and d) estimates of the disease allele frequency, mutation rate and penetrance.

The analysis tests whether the disease gene and a given marker segregate independently or whether they tend to co-segregate due to close proximity on the same chromosome. The identification of a polymorphic DNA marker linked with statistical significance to the disease phenotype allows association of the gene to a chromosome or region of a chromosome. This is based on whether the offspring (progeny) of an informative meiotic event are classified as recombinant (R) or non-recombinant (NR). Recombinants arise from the physical exchange between paternally and maternally derived homologues

during crossing over at meiosis. Recombination events can only be identified in the offspring of a doubly heterozygous parent. An odd number of crossovers between two markers results in recombinant individuals carrying a recombinant chromosome. When there are none or an even number of events these cannot be distinguished and will be classified as NR. If linkage phase is not known in an informative parent eg. in small families where phase is not known and cannot be inferred, then offspring cannot be identified as R or NR. Recombination fractions (θ) are derived by counting the number of R chromatids or non-parental haplotypes as a proportion of the total number.

The two basic hypotheses of linkage are H_0=free recombination and H_1=linkage. The degree of genetic linkage is expressed as a recombination fraction (θ) where with tight linkage between loci, θ approaches 0 since recombination is rare, and $\theta=1/2$ when loci are segregating independently and have 50% chance of recombination. The **l**og of the **odd**s or **lod** score $Z(\theta)$ defined by Morton (1955) is the logarithm (base10) of the likelihood ratio $L(\theta)/L(1/2)$ where $L(\theta)$ is the likelihood of linkage at any assigned $\theta<1/2$ and $L(1/2)$ is the likelihood under H_0 when $\theta=1/2$ (Ott, 1991). The odds for linkage are determined over a range of recombination fractions $\theta<1/2$, so that the likelihood of the data under linkage, rather than under absence of linkage, is calculated. The maximum likelihood estimated recombination fraction is that value of θ when the lod score (Z) is the highest, expressed as Zmax. Z is the measure of support for linkage versus absence of linkage and represents 'significant evidence for linkage' when it exceeds a critical value. The critical values for Z suggested by Morton (1955), were based on the probability of a Type I (false positive) or Type II (false negative) error of linkage. In an autosomally inherited condition when Zmax=3 the H_0 is rejected and represents 1000:1 odds for linkage. For X-linked loci the prior odds for linkage to the X chromosome, given the recognisable mode of inheritance, makes acceptable the less stringent critical lod score limit Zmax=2 (Ott, 1991) since the two loci in question are known to be on the same chromosome. When $Z\leq-2$ for a given θ, the disease gene can be excluded for the equivalent map distance from that marker.

For markers located on the same chromosome arm, the occurrence of one crossover appears to inhibit another in close proximity and is termed interference. Models of interference assume an inversely proportional relationship with the physical distance, that is the smaller the distance the greater the effect of interference in inhibiting multiple crossovers. One exception to this rule may be within the 2.6Mb PAR at Xpter, a known hot spot of recombination in male meiosis, where double crossover has been shown in a maximum 780kb interval spanning 19.1cM (Rappold et al, 1994). This finding suggests that it is the genetic distance and not the physical distance or steric hindrance that is involved in

interference. Most linkage analyses ignore the effects of interference or incorporate different levels of interference by using map functions (Ott, 1991).

Positioning of the gene in an interval is achieved by finding the closest flanking markers demonstrating recombinations and thereby bracketing the gene. If the regional localisation can be reduced to <1cM (≅1Mb), then the loci closest to the gene serve as the starting points for cloning by chromosome walking and jumping in cosmid or YAC libraries spanning the interval (section 1.3.2.). Single families are particularly subject to the limits of resolution of a linkage study, such that the gene interval may not become sufficiently narrow for approach by physical mapping (Boehnke, 1994). Linkage is a reliable way of confirming genetic heterogeneity among families with the same apparent clinical condition by demonstrating that similar phenotypes segregating in different families map to different and non-overlapping genomic locations. The advantage of studying families segregating a phenotype is that it subsequently allows assessment of individual risks for genetic counselling and development of diagnostic tests for disease determining genes.

The ability to localise disease genes by linkage is strictly dependent on the map of available polymorphic loci. Based on Botstein's (1980) calculations, 165 markers spaced at equal 20cM intervals would span the 33M of the genome so that any gene would be within 10cM of the nearest markers. A decade later over 10,000 loci defined by DNA probes and markers, of which approximately half were polymorphic, suggested that on average, there was at least one known polymorphic marker every Mb, however these markers were not evenly distributed throughout the genome (Williamson et al, 1991). The usefulness of a genetic map increases with marker density and informativeness (section 1.4.). Taking the Genethon microsatellite repeat markers (section 1.4.2.) alone, the current resolution under this premise is less than 2cM, so that any gene is theoretically within 1cM of the nearest marker (Weissenbach et al, 1992; Gyapay et al, 1994).

Construction of genetic maps involve strategies that infer the order of markers from observed recombination values (Lalouel et al, 1986). Although on average, 1cM ≡ 1Mb, recombination is generally suppressed at centromeres, increased at telomeres and occurs in excess at recombinational hotspots. Comparison of physical and genetic maps of chromosome 21 have indicated that 1cM involves only 200-300kb, while on the larger chromosome 4, the 1:1 relationship holds (Sherman, 1991). Background maps established from numerous normal families permit greater precision and statistical confidence in the mapping and localisation of rare genes (Drayna and White, 1985). Present day linkage maps, constructed with microsatellites in the CEPH reference families (Chapter 3), offer a quantum leap in disease gene mapping capabilities.

1.3.2. Physical Mapping

The physical mapping and cloning of a gene can take place following genetic mapping as part of the continuum of positional cloning or during functional cloning. Several technological innovations have enabled the physical dissection of the genome. Vectors for the cloning of DNA fragments of various sizes include plasmids, bacteriophage (λ), cosmids, bacterial artificial chromosomes (BACs) and P1. Yeast artificial chromosome (YAC) and megaYAC vectors can be used to clone >1Mb fragments of genomic DNA. When arranged into a contig (contiguous sequence), ordered physical maps of cloned DNA can be constructed over large regions of the genome (Cohen et al, 1993). Assembly of clones into a contig requires sequence tagged sites (STSs) that act as unique milestones linking overlapping clones along the linear physical map. Contig closure by filling the gaps between existing contigs, involves the directed strategy of walking using unique end probes to add bridging clones. Polymorphic markers used for genetic mapping have an important role as STSs to integrate the genetic and physical maps. Genetic markers flanking a disease gene will then also define the physical interval and designate the cloned fragments that contain the gene. The integration of the physical contigs and genetic markers are the bridge between genetic mapping studies and disease gene identification.

Naturally occurring gross rearrangements such as deletions or translocation breakpoints, when associated with a disease entity, inherently pinpoint the location of the defective gene (section 1.3.3.). Somatic cell hybrid panels produced from such rearrangements, or by *in vitro* exposure to high levels of radiation (Falk, 1991) can be used to determine the physical location and order of DNA markers, genes and expressed sequence tags (ESTs) to assemble a functional map of the genome. Mapped genes serve as STSs on the physical map as well as candidate loci for disease genes. Coding sequences can be identified from cloned DNA fragments by northern blot to test expression in specific tissues or, often highly conserved during evolution, can be recognised in genomic DNA by 'zoo blot' demonstrating cross-species homology (Monaco, 1994).

Pulsed field gel electrophoresis (PFGE) can resolve large DNA fragments and when used in combination with rare cutting restriction enzymes can be used to construct long range physical maps of the genome. Since recognition sites for rare cutters with C and G residues are concentrated in CpG islands (Lindsay and Bird, 1987), the construction of these maps can also indicate genes associated with these islands. All housekeeping and widely expressed genes have a CpG island at the transcript start, while 40% of tissue-specific genes and those with limited expression are associated with CpG islands (Bird, 1987). Overall 57% of genes are CpG island associated (Antequera and Bird, 1994). These islands are

likely to contain the promoters or enhancers of genes and may show bidirectional promoter activity (Larsen et al, 1992). The distribution of CpG islands can be visualised on metaphase chromosomes stained with Giemsa to produce a pattern of dark and light bands. High island density corresponds to cytogenetic Giemsa-light fractions or R bands that are early replicating and GC rich, and also contain a comparatively high frequency of Alu sequences (Craig and Bickmore, 1994). The chromosomal distribution of genes correlates more with GC content than chromosome length (Polymeropoulos et al, 1993). The most GC rich isochore (family H3) with the highest concentration of genes, CpG islands, transcriptional and recombination activity are located in the telomeric bands of metaphase chromosomes (Saccone et al, 1992).

1.3.3. Positional Cloning and the Candidate Gene Approach

Positional cloning, previously called reverse genetics, is a multistep strategy for identifying and isolating disease genes of unknown biological function (Collins, 1992) from knowledge of their positions in the genome. The approach depends upon initially determining the chromosomal localisation of the responsible gene, and narrowing down this region to the smallest possible interval, often by genetic linkage. Detailed physical mapping of the region then precedes DNA isolation, transcript identification, cDNA cloning and mutation searching of isolated candidate genes.

The chronic granulomatous disease gene (CYBB) on the X chromosome was possibly the first to be isolated by positional cloning (Royer-Pokora et al, 1986). Structural rearrangements of genes by deletion or translocation have successfully accelerated the positional cloning of X-linked genes for Duchenne muscular dystrophy (DMD) and Kallmann syndrome (Collins, 1992). A cytogenetically detectable fragile site associated with the fragile X syndrome of mental retardation defined the physical target for cloning of the FMR1 gene (section 1.6.1.). For genes localised by linkage, the isolation of the genomic DNA or coding sequences between flanking markers is most efficiently achieved when the genetic and physical maps have been integrated. Map integration can give a composite location for each locus and subsume partial maps (Wang et al, 1994) to facilitate cloning of disease genes.

Disease gene intervals defined by linkage in large single families depend on flanking recombination events and can therefore span many megabases. Positional cloning will be inefficient in searching for the transcript in such a large physical interval. In these instances candidate genes identified by their mapped position within the interval can be screened for disease causing mutations. Currently referred to as positional candidates, they become candidates for the disease gene by virtue of their location and expedite the discovery

of the mutation causing the disease phenotype. Mutation screening in these genes is more germane if they are transcribed in the appropriate affected tissues.

Candidate genes can also be directly investigated in the absence of linkage data if clear physiological (ie. functional) or physical (eg. translocation) evidence suggest an obvious candidacy. In most cases however, the identification of candidate disease genes requires a targeted search of the region between markers defining the minimum gene localisation by direct selection of cDNA clones from genomic DNA (Lovett et al, 1991) or exon trapping of expressed sequences (Duyk et al, 1990). Genomic DNA sequences stored on the GenBank nucleotide sequence database (Burks et al, 1991) can be analysed by computer algorithms written to assist in finding gene coding sequences. Specific software such as BLAST (Basic Local Alignment Search Tool) can identify exons, the open reading frame, stop codons and search for homology to known genes.

Strategies to identify random candidate genes have generated expressed sequence tagged sites (ESTs) by partial sequencing of cDNA clones directly from tissue-specific cDNA libraries (Adams et al, 1993b). This method of identifying new genes also describes the transcriptional activity of the tissue. Construction of regional transcription maps is a thorough method of detecting all the genes in a defined chromosomal area (Tribioli et al, 1994). The development of integrated transcriptional maps of the genome, and subchromosomal localisations of randomly derived ESTs will make these ESTs available as positional candidates for gene isolation and mutation screening. Examination of transcripts in an area may eventually identify a gene that contains a mutation in individuals who have the disease, but not in normals. Detection of the disease causing mutation can involve several complex techniques (Grompe, 1993) and is likely to become a bottleneck precluding widespread diagnostic application. Once the disease gene is cloned, the encoded protein and its function must be analysed biochemically in an effort to seek understanding of the biological basis of the clinical phenotype.

1.4. Evolution of Polymorphic Markers

The research project described in this thesis was commenced just prior to a vast turning point in the field of molecular genetics. Earlier polymorphic markers (including blood group antigens, allozymes and major histocompatibility types) had been superseded in the 1980's, initially by RFLP and VNTR markers that were the earliest form of DNA based markers (Botstein et al, 1980; Nakamura et al, 1987). RFLPs and VNTRs were labour intensive for gene mapping studies and their exact physical locations were not always known. This was the rate limiting step to the mapping and positional cloning of many disease genes. In 1989, a new class of polymorphic repeat markers (Weber et al, 1989; Litt

and Luty, 1989), revolutionised genetic mapping. This abundant class of highly polymorphic microsatellite markers was more efficient both in construction of the background genetic map and in localisation of disease genes in families. Application of a rapid *in vitro* method for selective amplification of DNA segments, the polymerase chain reaction (PCR) (Saiki et al, 1988), to genotyping microsatellites accelerated gene mapping studies.

The usefulness of a given genetic marker can be correlated with the number and frequencies of the different alleles. The higher the heterozygote frequency of a marker the more meioses will be informative, whether in reference families to enable reliable ordering of markers on the map, or in disease families to efficiently localise the gene. The maximum expected heterozygosity for markers detecting biallelic systems is 50% as defined by the formula; $H = 1 - \Sigma P_i^2$ where P_i is the frequency of allele i. Alternatively, the polymorphic information content (PIC) can be calculated by; $PIC = 1 - (\Sigma_{i=1} P_i^2) - \Sigma_{i=1} \Sigma_{j=i+1} 2P_i^2 P_j^2$ (Botstein et al, 1980). PIC values can theoretically range from 0 to 1 for markers with one allele to those with an infinite number of alleles, however a polymorphic locus is defined as H (or PIC) ≥ 0.01 (Ott, 1991).

1.4.1. RFLPs and VNTRs

In 1980, Botstein proposed the construction of a linkage map of man using restriction fragment length polymorphisms (RFLPs) and their subsequent use in finding disease genes by linkage. Utilising classical linkage analysis they were rapidly developed as a tool for gene mapping. RFLPs have been very useful for primary mapping of disease genes in the past, for example Huntington disease (Gusella et al, 1983), however they have limited utility because of low informativeness. RFLPs are commonly biallelic and detect changes in the genomic sequence within a given fragment that result from the presence or absence of a polymorphic restriction enzyme recognition site. The resultant length polymorphism is assayed by Southern blot hybridisation, one drawback of which is the requirement for a large amount of DNA relative to the PCR method (section 1.4.2.). Often not sufficiently densely placed for refining localised genes to intervals small enough for gene identification, the detection of RFLPs is demanding on labour, time and DNA for generating genotype data. These factors contributed to limitations to the RFLP linkage map itself (Chapter 3.2.1.) and to gene localisations thereupon.

Another class of polymorphic genetic marker with a variable number of tandem repeats (VNTR), are minisatellites that contain >6 nucleotides in every repeat unit (Jeffreys et al, 1985; Nakamura et al, 1987). Genotyping still requires Southerns as the fragment sizes are often too large to analyse by PCR although they represent essentially the same class of

polymorphic marker as the microsatellites (section 1.4.2.). With heterozygosities >70%, minisatellites are more informative than RFLPs but are not as evenly distributed in the genome, tending to cluster at the telomeres of chromosomes (Lathrop et al, 1988). Analysis of VNTRs also requires large amounts of DNA, a limited resource in most family studies, and can have complex patterns of alleles.

1.4.2. Microsatellites and Dinucleotide Repeat Polymorphisms

Microsatellites are hypervariable short tandem repeats (STRs) of 1-6bp motifs that commonly consist of di-, tri- or tetra-nucleotide units (Edwards et al, 1991). Exploitation of PCR technology, (Saiki et al, 1988) for genotyping of polymorphic dinucleotide repeat markers (Weber and May, 1989; Litt and Luty, 1989) revealed a vast new source of highly polymorphic genetic markers spread throughout the genome. Microsatellite markers have since been used for constructing high resolution genetic maps (NIH/CEPH Collaborative Mapping Group, 1992; Gyapay et al, 1994) and localisation of disease genes. Multiple loci can be genotyped simultaneously by multiplex amplification of selected targets. Markers with heterozygosity values greater than or equal to 75% have been predicted to be fully informative in half the families typed when typed two markers at a time (Williamson et al, 1991). Unique primer sequences flanking the repeat motif could also be utilised as STS reagents for establishing contigs and for integration of genetic and physical maps (section 1.3.2.).

The relative abundance of tri- and tetranucleotide repeats is high in mammalian genomes (Beckmann and Weber, 1992), however the most widely developed subclass are the $(CA)_n(GT)_n$ dinucleotide motifs (Weber and May, 1989; Litt and Luty, 1989; Weissenbach et al, 1992; Gyapay et al, 1994) conserved in many eukaryotes (Miesfield et al, 1981). Dinucleotide repeat loci occur at average intervals of 30-60kb (Weber and May, 1989), appear to be distributed ubiquitously and are estimated to number 50-100,000 in the human genome (Tautz and Renz, 1984). Tandem microsatellite repeat sequences can be classified as perfect (64%), imperfect (interrupted) (25%) or compound (together with another repeat type) (11%). The PIC can be predicted from the average length of the uninterrupted 'perfect' repeat, where 16 repeats have an average PIC = 0.5, 20 repeats PIC = 0.7, however beyond 21 repeats PIC values level out below 0.9 (Weber, 1990). Assuming random distribution, and barring genotyping errors and mutations, it has been estimated that 12,000 (AC)n with PIC>0.5 (of which 7000 have PIC>0.7) could theoretically yield high resolution genetic maps of 0.3-0.5cM spanning the genome (Weber, 1990).

A landmark development in the HGP has been the automated production of a genome-wide polymorphic microsatellite marker map at Généthon (Todd, 1992). Initially,

814 genome-wide primer sets were characterised with 25 on the X chromosome (Weissenbach et al, 1992). Segregation anomalies were observed at several loci and correspond to a mutation rate close to 0.1%/locus/generation (Weissenbach et al, 1992). A further Généthon map more than trebling the markers on the X and increasing the number of genome wide markers to 2066 has since been released (Gyapay et al, 1994). These short tandem repeats were mapped with respect to one another by genotyping 150 members of eight large CEPH families to produce a recombination map spanning 3690cM of the genome, including 168cM of the X chromosome (Gyapay et al, 1994). Integration with other existing markers on the X chromosome (Willard et al, 1994) is useful in the mapping and isolation of genes (Chapter 3).

The abundance and apparently random distribution of STRs in eukaryotes from yeast to human suggests some biological significance. Database searches have shown however, that the position of (GT)n repeats is conserved in closely related primates but not in more distant species (Stallings et al, 1991). Potential Z-DNA forming sequences (dA-dC)n.(dT-dG)n may be important in regulating DNA function by mutagenesis or control of gene expression (Hamada and Kakunaga, 1982), modulation of transcription, or a role in recombination (Stallings et al, 1991). Heritable unstable trinucleotide sequences have been found in association with a number of predominantly neuropathologic disorders (section 1.6.1.) some of which have a demonstrable ability to act as binding sites for nuclear proteins (Richards et al, 1993). The complex mechanisms of repeat sequence mutations (Richards and Sutherland, 1994), increasingly associated with single gene disorders and neoplasia (Wooster et al, 1994), substantiate some fundamental function of repeat sequences in the genome.

1.5. The Sex Chromosomes

The 46 human chromosomes are classified as 22 pairs of homologous chromosomes or autosomes, with the morphologically distinct sex chromosomes, X and Y, making up the 23rd set. The X and Y are only homologous over the pseudoautosomal region (PAR), at each telomere of the X. Normal females are XX, while normal males are XY. Exceptions to this rule have provided important clues in discerning mechanisms for sex determination (de la Chapelle, 1988). Individuals without a Y chromosome are female, those with a Y are male irrespective of the number of X chromosomes present.

A male transmits only his X chromosome to his daughters and his Y to his sons. The Y is believed to be silent or at least not affecting the expression of X-linked genes, although recombination can occur within the PARs. The PAR pair during male meiosis and the two chromosomes can exchange sequence information within the 2.6Mb at Xpter (Schmitt et al,

1994) a known hot spot of homologous recombination, and in the 320kb at Xqter (Freije et al, 1992; Kvaløy et al, 1994). A double crossover has been shown in the Xp/Yp region (Rappold et al, 1994).

Sex determination is due to a gene named *SRY* (sex determining region Y gene) located on the Y chromosome (Sinclair et al, 1990; Hawkins, 1994). *SRY* encodes a dominant testis determining factor (TDF) in males and is expressed in several tissues besides the gonads (Clépet et al, 1993). The SOX3 homologue to *SRY* is X-linked (Xq26-27) but not involved in sex determination as evidenced by an XY male deleted for SOX3 (Stevanovic et al, 1993). Ovarian differentiation is considered the default pathway overridden in the presence of *SRY* expression (Nordqvist and Lovell-Badge, 1994). Further studies show that a TDF gene on Xp (TDF-X), that suppresses testis formation, is subject to X inactivation and that interaction with *SRY* in males switches TDF-X off, while in females TDF-X is active on one X chromosome (Ogata and Matsuo, 1994).

One of the X chromosomes in each female somatic cell is randomly rendered inert by lyonisation and forms a Barr body. Inactivation results in transcriptional silencing of the majority of genes on one X in female cells and results in comparable expression of genes on the X in males and females. Expression of a disease gene in females is thus moderated by the presence of the normal X, active in some cells. Failure to inactivate may result in aberrant phenotypes, analogous to disease gene expression in males, from supernormal expression of X-linked genes that would normally be dosage-compensated (Lahn et al, 1994). A unique locus XIC (X inactivation center) within band Xq13, extending from 680-1200kb, is believed to be necessary for inactivation (Heard and Avner, 1994). Within this region the XIST (X_i-specific transcripts) gene is a possible candidate for XIC, but may not act on its own. Methylation of the X has been shown to be involved with maintenance of inactivation (Heard and Avner, 1994). Rearrangements of the X in X:autosome translocations resulting in the expression of a recessive phenotype in a female, indicate the site of the gene since the rearranged X is active in most cells and the breakpoint interrupts the gene.

The X chromosome has been maintained as a single syntenic group in all eutherian mammals tested, so that comparative mapping of conserved single-copy probes between mouse and human X chromosomes can be used to define regions of homology (Laval and Boyd, 1993). That inactivation and sex determination systems were protected from disruption for over 100 million years implies a constraint due to biological function (Marshall Graves and Watson, 1991). Mammalian conservation of genes on the long arm of the human X provide evidence that this is part of an ancestral X chromosome. Genes mapped to the short arm however, were probably autosomal 60-150 million years ago and

reveal rearrangements in sex chromosomal evolutionary history that may account for regions exempt from inactivation (Watson et al, 1991).

1.5.1. X Linkage

Recognising the mode of inheritance is an essential element for mapping the defective gene in human disorders. The specific pattern of X-linked gene transmission arises as a consequence of females producing only X bearing eggs while males produce X or Y sperm. X-linked genes assume haploid status in all males. In X-linkage therefore, each mating represents a test cross[1] as the carrier female is heterozygous, and often asymptomatic for the disease, the male hemizygous. A carrier female will transmit either of her two X chromosomes to her children. All male progeny will be informative as they phenotypically reflect which of her X chromosomes they have inherited. Male offspring therefore have a probability of 1/2 of being affected, while daughters have a probability of 1/2 of being a carrier. Recombinants can be easily detected when the grandpaternal genotype is known, since this information can be used to determine phase in the mother so that the maternal and paternal homologues can be differentiated as haplotypes.

Pedigrees with affected males in sibships connected through females (ie. no male to male transmission) are accepted as evidence of X-linked recessive inheritance, however when the family is small, the mode may be indistinguishable from autosomal inheritance. A sex-limited autosomal dominant gene can have the same pattern, however X-linkage is conclusive if all sons of affected men are normal. X-linked dominant inheritance is accepted in pedigrees with manifesting heterozygous females, where an affected male has only affected female offspring but normal sons. In many conditions affected males do not reproduce making it difficult to distinguish the mode of inheritance in small families. In either X-linked mode, carrier women will pass the gene to 50% of their sons (segregation ratio of 1:1). In X-linked lethal conditions eg. incontinentia pigmenti, all affected individuals appear to be females who bear few but normal live born sons as males inheriting the disorder do not survive to postnatally express the gene. For X-linked diseases involving a single gene these principles permit recognition of the genetic model of inheritance often without a need for formal segregation analysis.

---

[1]The test cross is a genetic method for the demonstration of allelic segregation by crossing the first filial generation (F1) of a true-breeding wildtype x true-breeding recessive to a true-breeding line (often also referred to as a backcross when the F1 is crossed back to the true-breeding parent). The true-breeding parent is homozygous for a recessive gene (or hemizygous in the X-linked case) and contributes 'silent' gametes to the cross and thus tests the genetic makeup of the F1, and confirms the heterozygous state through the phenotypes of the progeny segregating in a 1:1 ratio.

1.5.2. X-linked Disorders

Pedigree inspection can often assign the gene responsible for an X-linked disorder to the X chromosome by the characteristic pattern of inheritance. The haemophilia and colourblindness genes were the first to be recognised and classified as X-linked disorders in the early part of this century. In 1992, 50% of the genes assigned to the X chromosome were disease-related while only 20% of genes assigned to the comparably sized autosome, chromosome 6, were disease-related (Neri et al, 1994). By mid-1994, ~230 X-linked expressed genes of known function were assigned to the X (McKusick, 1994). Comparatively, the almost 3000 validated autosomal loci indicate that in spite of the ease of recognising X-linkage, the number of conditions is not disproportionate to the amount of genetic material.

Many clinically defined disorders are being mapped by linkage to specific chromosomal locations whenever families large enough or numerous enough are ascertained. In families with an X-linked disease of severe, debilitating or lethal manifestation the regional localisation of the gene can provide access to genetic counselling for reproductive choice in subsequent generations. Where the pattern of inheritance is unclear, evidence of linkage of a phenotype to an X-linked marker proves that the disorder is X-linked. Most X-linked conditions are recessive. The disease-causing mutations are expressed with full penetrance of the phenotype in males, and most often not phenotypically apparent in females. Manifestation of the disease in carrier females can be due to dominance of the disease gene, skewed inactivation, homozygosity of a recessive gene or chromosomal rearrangement. In females, random inactivation of one X chromosome reduces the likelihood of extreme manifestations of any deleterious gene. An exception is the fragile X syndrome (section 1.6.1.) that appeared to be inherited as a recessive trait but was found to be expressed in carrier females and not in normal transmitting males (Sutherland and Hecht, 1985).

From 1986 to 1994 at least 11 X-linked disease genes had been identified by positional cloning and a further 5 found by the application of the candidate gene approach for disease-gene identification (McKusick, 1994). Detection of a chromosomal defect such as deletion or translocation has led to mapping and cloning of a number of disease genes. The largest known single gene on the X chromosome, dystrophin (2.4Mb with a 14kb transcript), was the first Mendelian disorder mapped by chromosome rearrangements (Tommerup, 1993). The most common dystrophin mutation is by deletion, however a translocation in an affected female helped characterise the gene. Fourteen balanced translocations of the X lead to a clinical disorder in females, several of these genes have

been cloned by physical mapping of the translocation breakpoint (Frézal and Schinzel, 1991). Fifteen X-linked genes have been located by mapping deletion intervals in patients, particularly those associated with contiguous gene syndromes can be accurately positioned and ordered on the physical map (Frézal and Schinzel, 1991).

1.6. Mental Retardation and the X Chromosome

Mental retardation is defined by three criteria: IQ below 70-75 (less than two standard deviations below the mean); significant limitations in two or more adaptive skills essential for daily functioning; and presence of the condition from childhood (age 18 or less) (AAMR, 1992). Normal intelligence is defined by IQ >85-115. The prevalence of mental retardation in the general population is between 2.5 to 3 percent. Formally assessed IQ can be used to classify the levels of retardation within ranges established by the World Health Organisation (Chapter 4.1.). The majority (80-90%) of retarded individuals are mildly affected, whereas 5% are severely to profoundly impaired.

Historically, Lehrke (1974) demonstrated an excess of males in the retarded population. Although he noted that females appeared to be relatively more frequently institutionalised than males (as a means of preventing them reproducing), estimates of sex ratio in the institutionalised population showed a male excess of 21% in the US and 25% in Colchester, England (Lehrke, 1974; Morton et al, 1977). The proportion of retarded males relative to females in the noninstitutional community was far greater, reflecting the disproportionate institutionalisation of females. This male excess was initially attributed to ascertainment bias given higher expectations of males in schools, the greater number of males in the total population, or simply a greater susceptibility of males to the environment (Lehrke, 1974; Glass, 1991).

Evidence of a genetic component to retardation was provided by Reed and Reed (1965), who showed that 5 of 6 retarded are 1st or 2nd degree relatives of a retarded individual. The preponderance of males in families with two or more mentally retarded sibs was also reported (Wortis, 1966). The hypotheses put forward by Lehrke (1974) asserted that *'there are major genetic loci relating to intellectual functioning on the X chromosome'* that *'if mutated, can lead to ... mental retardation transmitted in an X-linked manner'*. In addition to an excess of retarded males, Lehrke (1974) discussed evidence of males with high IQ. The broader distribution of IQ in males relative to females was used to support the hypothesis of X-linked genes for intelligence by invoking dosage compensation through inactivation of one X chromosome in females. Lehrke concluded that sex-linked genes should account for between 20-36% of all mental retardation and thus the observed excess of males (Lehrke, 1974). This 'conservative' estimate followed a challenge of his more

often quoted original estimate of 25-50% (Opitz, 1986). A study of school age retardates in Hawaii also suggested that 9-26% of all retardation was due to X-linked genes (Proops et al, 1983). The genetic contribution of the X chromosome to mental retardation is clearly not proportional with its relative size (~5%) in the genome.

X-linked mental retardation (XLMR) is a general term including syndromal and non-syndromal forms (sections 1.6.2.-3.) chiefly differentiated by the presence of consistent and characteristic phenotypic manifestations associated with mental retardation in the former, and a lack of definitive difference between normal and retarded individuals within the families of the latter. Turner and Turner (1974) studied retarded sib pairs (IQ 30-55) in mental institutions of NSW, and found that the excess of brother pairs corresponded to a prevalence[1] of XLMR of 1 in 1724 males. Considering the entire non-specific XLMR population of British Columbia, the prevalence was estimated to be 1/546 live male births with a carrier frequency of 1/410 females (Herbst and Miller, 1980). Since a birth cohort was utilised this figure may be regarded as a minimum incidence[2]. Sutherland and Hecht (1985), similarly estimated that about 1 in 600 male births have XLMR. All of these figures related to XLMR with few or no syndromic features, but included the fragile X syndrome which was considered a non-specific entity at that time (section 1.6.1.). An estimated total prevalence of all XLMR of 1/145 accounts for up to one quarter (20-25%) of all mental retardation in humans, given a 3% prevalence of MR in the population (Opitz, 1986).

Until 1990, only 39 X-linked conditions were identified in which mental retardation was the primary characteristic, 17 of these had been regionally localised (Neri et al, 1991). Disorders such as Coffin-Lowry syndrome, Lesch-Nyhan syndrome, Börjeson-Forssman-Lehmann syndrome, Hunter mucopolysaccharidosis and the fragile X syndrome could be differentiated by distinct clinical, biochemical or cytogenetic abnormalities. Gene mapping in rare syndromic forms, catalogued on the basis of dysmorphology, confirmed clinical evidence of the marked genetic heterogeneity of XLMR (Neri et al, 1991). In the 1992 listing of Mendelian Inheritance in Man (MIM), >70 of the 360 X-linked conditions (20%), were associated with mental handicap (McKusick, 1992). The gene for more than half, 40 of the 77 identified XLMR conditions, had been regionally localised (Neri et al, 1992). The most recent update classifies 127 XLMR conditions and includes all known forms as well as those where the MR is a secondary or peripheral feature (Neri et al, 1994).

---

[1]Prevalence is the frequency of cases living in the general population as opposed to the incidence which is the frequency to age 1yr of persons who are retarded. Incidence figures take into account premature mortality that may reduce the prevalence (Morton et al, 1977).

[2]The incidence of XLMR is difficult to estimate accurately especially in the non-syndromal form where survival is not affected but diagnosis is often incomplete prior to school entry.

1.6.1. The Fragile X

The discovery of a chromosomal aberration, fra(X)(q28), associated with mental retardation (Lubs, 1969) and sensitivity to low folic acid tissue culture media (Sutherland, 1977) established a diagnostic test for over one quarter of all X-linked mental retardations (Sutherland and Hecht, 1985). The fragile X syndrome is the most common form of inherited mental retardation (Sutherland, 1977), second only to Down Syndrome as the most common chromosomal cause of mental retardation. The prevalence in males is 1 in 2000-2500 (Sutherland and Hecht, 1985) or as high as 1 in 1350 English schoolboys (Webb et al, 1986). The incidence calculated from the data of Webb et al (1986) is 1 in 363 and includes female carriers and non-manifesting hemizygous males (Opitz, 1986). A possibly increased infant mortality in affected males or variation in ascertainment methods in different studies may contribute to the apparent discrepancy.

Initially, many of the clinical characteristics of males with XLMR were not considered to be more than minor anomalies within the range of normal variation. Macro-orchidism was reported frequently (Herbst, 1980) but was not present in all cases (Jacobs et al, 1979) until correlations with fragile site expression were demonstrated (Turner et al, 1980; Sutherland and Ashforth, 1979). A large pedigree described in 1943 by Martin and Bell had X-linked mental retardation with consistent, but non-syndromal, features and was later shown to express the fragile site (Richards et al, 1981). Once families could be delineated by cytogenetic expression of the fragile site, the clinical phenotype which is subtle in some individuals, was examined as a group and a series of features defining the nosology could be correlated. The specific phenotype including large ears, hyperextensible joints and megalotestes delineated the Martin-Bell syndrome (in deference to the first authors describing a family with the syndrome) and later became known as the fragile X syndrome.

Another family with severe mental retardation but lacking physical signs, described by Renpenning in 1962, was for a time considered to be the same. The eponymic Renpenning syndrome came to represent XLMR in individuals without physical abnormality (Turner et al, 1971), so that parts of the literature discuss subjects with Renpenning syndrome who carry the fragile X (Proops and Webb, 1981). Restudy of the original Renpenning family has since delineated the phenotype from the fragile X (Fox et al, 1980) and demonstrated that the gene for this fragile X negative disorder is localised to the proximal short arm (Schwartz et al, 1994). Little over a decade ago, however, the Martin-Bell and Renpenning syndromes were defined by families with mentally retarded males of normal appearance and stature (Proops and Webb, 1981). These families represented the

first examples of non-specific X-linked mental retardation (section 1.6.2.). With increasing diagnostic criteria the nonspecific XLMR 'entity' was subdivided into at least three phenotypes (Tariverdian and Weck, 1982), though they were mainly differentiated on the basis of expression or lack of expression of the fragile site at Xq27 (Turner et al, 1980; Sutherland and Ashforth, 1979).

Families with the Martin-Bell phenotype but not expressing the fragile site raised the question of a fra(X) negative Martin-Bell syndrome (Neri et al, 1988). Re-evaluation of several 'fragile X negative' retarded males found no evidence to support the existence of such a group, since clinical reassessments could define those positive for the fragile X and others that did not fit the criteria of the Martin-Bell phenotype (Thode et al, 1988). This put some doubt onto the existence of a fraX negative Martin-Bell syndrome and suggested greater genetic heterogeneity to account for XLMR. Deletions involving FMR1, detected by direct molecular analysis, have been recently shown to account for a small proportion of non-FRAXA males with classic Martin-Bell phenotype (Chapter 7.2.).

The folate sensitive, cytogenetically detectable fragile site at Xq27.3, served to delineate the fragile X syndrome as a specific clinical entity. This led to refinement of the gene localisation at or very near to the fragile site by linkage and facilitated the isolation of DNA sequences and cloning of this gene. At the molecular level, the majority of mentally retarded patients are characterised by expansion of an unstable $(CCG)_n$ repeat in the 5' UTR of the FMR1 gene (Fu et al, 1991; Obérle et al, 1991; Verkerk et al, 1991; Yu et al, 1991). The pathogenic mechanism involves silencing of the FMR1 gene by expansion of the CCG repeat beyond a critical level with subsequent hypermethylation of the repeat and the adjacent CpG island and resultant lack of mRNA (Bell et al, 1991; Oberle et al, 1991; Pieretti et al, 1991). Less commonly point mutation (De Boulle et al, 1993) or deletion (Chapter 7) at the FMR1 locus can give rise to the fragile X phenotype.

Direct molecular diagnosis of the fragile X has been made possible by the cloning of the FMR1 gene and genotype-phenotype correlation of the CCG repeat (Yu et al, 1992). The CCG_n copy number is polymorphic over a range of n= 6 to 54, with the most frequent alleles between 27 to 31 repeats (Fu et al, 1992). Expansions beyond ~55 copies are in a class of premutations that are predisposed to further increases in subsequent generations, particularly when transmitted by a female. It is in this class that normal transmitting males and carrier females are found. CCG repeats numbering more than 200 copies (an expansion greater than 600bp) represent the full-mutation and are both meiotically and mitotically unstable (Rousseau et al, 1992). The CCG repeat expansion in the FMR1 gene can be quantitated by PCR or Southern blot to detect carriers of the fragile X and confirm earlier diagnoses based on flanking linked markers (Appendix V, Paper 3). Prenatal diagnosis by direct assay of the

amplification is more reliable than cytogenetic detection of the fragile X chromosome in cultured fetal cells that are prone to 5% misdiagnoses (Appendix V, Paper 1).

Recently the cytogenetic test for fragile X syndrome has been further complicated by characterisation of other folate sensitive fragile sites FRAXE (Sutherland and Baker, 1992) and FRAXF (Hirst et al, 1993) in the Xq27-q28 region. Cloning around the FRAXE (Knight et al, 1993) has revealed CCG amplification in association with non-specific features of mental retardation (Hamel et al, 1994; Knight et al, 1994; Appendix V, Paper 6). The CCG repeat at FRAXF appears not to be associated with any phenotype although ascertainment is often through males with developmental delay and fragile site expression (Parrish et al, 1994; Ritchie et al, 1994), however very few families have been identified. Another, autosomal, folate-sensitive fragile site FRA16A, has been characterised with dynamic CCG repeat mutation but without phenotype in heterozygotes (Nancarrow et al, 1994). Cytogenetic *in vitro* expression of these fragile sites apparently requires CCG repeat expansion, methylation of its CpG residues and thymidine deficient culture medium, although much about their mechanism of expression remains unclear at the molecular level (Sutherland and Richards, 1995).

Unusual inheritance of the fraX syndrome is identified by the presence of nonpenetrant carrier or phenotypically normal transmitting (NT) males, frequent expression in carrier females and variable penetrance between sibships. Incomplete penetrance in females and the occurrence of a special class of mutation (premutation) accounting for normal transmitting males was factored into segregation analyses for genetic counselling (Sherman et al, 1985). These peculiarities of segregation were termed the Sherman paradox (Sherman et al, 1984, 1985). Observations that the unstable sequence caused increased penetrance of the disease in subsequent generations were puzzling to Mendelian geneticists who explained it in terms of poorly understood genetic mechanisms (Sutherland et al, 1991), now known as anticipation. The molecular basis for this phenomenon has been demonstrated at a number of other sites in the human genome in association with trinucleotide repeat disruption of transcription/translation in genes. The class of 'dynamic' mutations of which the fragile X was the first, has now been identified in association with several inherited neuropathological conditions: fragile X, DM, SBMA, HD, FRAXE, SCA1, DRPLA/HRS and Machado-Joseph/SCA3 (Brook et al, 1992; La Spada et al, 1991; Huntington Disease Collaborative research group, 1993; Knight et al, 1993, Orr et al, 1993; Koide et al, 1994; Kawaguchi et al, 1994) but only three different triplet repeats CCG, CAG and CTG. CAG expansions in SBMA, HD, SCA1 and 3, DRPLA are translated into polyglutamine tracts with gain of function from the resultant protein (Monckton and Caskey,

1995; Jennings, 1995). Myotonic dystrophy (DM) is the only disease with an amplified CTG repeat in the 3'UTR.

1.6.2. Non-specific MR (MRX)

Early studies defined non-specific XLMR as a causally heterogeneous category including all cases "without a recognised syndrome" (Proops, 1983) and those with no definite biochemical, chromosomal or environmental cause. Clinical findings could differentiate non-specific XLMR into two groups associated with presence or absence of macroorchidism and the fragile site, however the pattern of inheritance was the only reliable means of identifying XLMR (Herbst et al, 1981). Delineation of the fragile X syndrome as a specific entity diagnosed by clinical and laboratory indicators (section 1.6.1.) redefined the term 'non-specific' in the clinical sense. Individuals with MRX are defined as having no consistent clinically recognisable phenotypic (physical or physiological) features apart from non-progressive idiopathic intellectual handicap segregating in a characteristic X-linked recessive manner (Appendix II, Paper 2). The interchangeable terms non-specific or non-syndromal and the symbol MRX are now largely accepted by the scientific community to refer only to those families in which mental retardation occurs without other distinguishing biochemical, cytogenetic or phenotypic markers (Glass, 1991; Schwartz, 1993). Families with consistent but minor anomalies have still occasionally been reported in this category (Wilson et al, 1992). Families described as bland, simple or non-syndromic forms of XLMR as well as those previously documented as non-fragile X Martin-Bell should be recategorised as MRX or syndromal XLMR (Kerr et al, 1991).

Approximately half of all the 'nonspecific' XLMRs occurring in 1.83 per 1000 male births were associated with the fragile X (Herbst and Miller, 1980). The consensus reached at the Second International Fragile X and XLMR Conference was that mutations in the gene responsible for the highly prevalent fra(X) syndrome (section 1.6.1.), account for about 40% of all 'bland' forms of XLMR (Opitz, 1986). An equal or greater number of non-dysmorphic conditions must exist therefore fitting the definition of MRX. Calculated from the 1 in 363 incidence of fraX (Opitz, 1986), MRX therefore has an expected incidence of 1 in 242.

The gene causing fragile X, now known as FMR1, is a single major gene for mental retardation on the X chromosome. Based on this evidence the remaining proportion of non-specific XLMR might arise from another major single gene or a number of recessive genes for retardation. Indirect estimations assuming a mutation rate of $3\text{-}9\times10^{-5}$ have suggested that there are 7 to 19 X-linked loci causing MRX including the fraX (Herbst and Miller, 1980). Morton (1977) estimated at least 18 loci on the X chromosome causing mental retardation at a mutation rate up to 3.2×10^{-5}/gene/generation. Based on the relative physical size a

proportionate number of X-linked genes to autosomal genes. Localisations of 2 MRX genes to distinct regions of the X chromosome by linkage to polymorphic markers genotyped in families (Arveiler et al, 1988; Suthers et al, 1988) was the first real evidence of the genetic heterogeneity of MRX. Genetic heterogeneity in this clinically homogenous group is discussed in Chapter 4.

The implication that many critical genes for intelligence reside on the X is based on the number of genetically and clinically heterogeneous X-linked mental retardations. MRX genes with no recognisable morphologic, neurologic or metabolic signs have been suggested as the genetic basis of intelligence in humans (Turner and Partington, 1991). Certainly if mutations causing disruption in an MRX gene have a negative effect on intelligence, then these genes could reasonably be assumed to represent the 'pure' determinants for normal brain development or function. The number of MRX genes, by extrapolation, would be directly informative with regard to enumerating the X-linked genes involved in determining the essential substance or processes of intelligence in the brain (Chapter 4). It would be naïve, however, to suggest that intelligence is beholden to the products of any single gene. Rather, one might suppose that MRX genes are heritable components of intelligence that represent a group of independent factors which when defective disrupt any interactions to result in generalised MR or specific defects in speech or verbal functioning.

Since MRX genes affect intelligence without other phenotypic effects, it is tempting to assume that given that the only tissue apparently involved is the brain, the mutated gene involved should be expressed in the brain. Up to 30,000 of the 50-100,000 genes in the human genome may be expressed in the brain (Adams et al, 1991). Human brain cDNA tags (ESTs) may, when chromosomally localised, identify genes associated with neurological diseases and provide an overall view of gene expression in the brain (Adams et al, 1991). Sequence similarity of cDNAs expressed in brain with genes characterised in model organisms and tissues have identified over 700 proteins from cytoskeletal structural proteins to very low abundance neurotransmitter receptors (Adams et al, 1993a). Mapping of 22 ESTs derived from a brain cDNA library against hybrid cell line panels and to YAC contigs has localised potential candidate genes for neurological disorders to specific regions of the X chromosome (Parrish and Nelson, 1993; Mazzarella and Srivastava, 1994). Novel genes containing triplet repeats expressed in brain are also candidates for neuropsychiatric diseases including MRX, pending chromosomal assignment to the X chromosome (Li et al, 1993). Genes that result in mental retardation are not necessarily only expressed in the brain however, and not only involved in development or regulation of function either (section 1.6.3.). A simplistic cause and effect cannot be predicted for any genes affecting intelligence

since the processes involved in normal development and maintenance of brain function are numerous and ill understood.

1.6.3. Syndromal MR (MRXS)

Clinical geneticists define a syndrome as "..a recognisable pattern of anomalies which are known or thought to be causally related" (Opitz, 1994). X-linked mental retardation syndromes are largely classified on the basis of an X-linked clinical phenotype in association with consistent mental retardation. More than 20 syndromes of XLMR had been described by the mid-80s (Opitz and Sutherland, 1984; Turner, 1986). By 1994 over 100 clinically identifiable XLMR syndromes were recognised, including those disorders where the mental retardation was secondary and peripheral to the main pathology, 80 had been assigned a MIM number (Neri et al, 1994).

The XLMR syndromes have been subdivided on the basis of identified specific neuromuscular involvement, patterns of recognisable physical anomalies, dominant disorders lethal in males and inborn errors of metabolism (Neri et al, 1994). The largest subgroup, MRXS, includes only those conditions where the mental retardation is due to an unknown cause rather than secondary to a metabolic disturbance. The greatest proportion of MRXS is accounted for by mutations in FMR1 causing the fragile X syndrome (section 1.6.1.). The phenotype in some syndromes can be subtle, for example the fragile X syndrome was originally described as a non-specific form of mental retardation until the discovery of laboratory methods to induce (and enhance) fragile site expression in affected males, brought about assessment of differential features. Genetic heterogeneity or affinity of XLMR conditions (also termed splitting or lumping respectively) may become apparent with molecular characterisation of MRXS disorders through regional gene localisations and mutation analyses. Clinically heterogeneous conditions may be found to be allelic at the same gene locus and hence lumped, while clinically similar entities may be split by non-overlapping gene localisations. Some of the clinically homogenous MRX group (section 1.6.2.), classified by gene mapping (Chapter 4), may in future be lumped with some syndromal forms of mental retardation (Chapter 5).

There is an enormous potential for learning about brain development through these disorders, however, abnormal brain development may arise from a metabolic lesion that affects the brain only indirectly. Inborn errors of metabolism are often pleiotropic reflecting multiple instances during embryology and fetal development (organogenesis) where a lack of an enzyme or excess of a metabolite can cause damage, for example Niemann-Pick's disease is a fatal childhood inborn error of lipid metabolism where a deficiency of enzymes causes accumulation of a phospholipid (sphingomyelin) in all organs including the brain and

causes mental deficiency. Similarly, HPRT deficiency results in Lesch-Nyhan syndrome where males are mentally retarded and have a bizarre pattern of behaviour, while the normal gene function is in bone marrow. Mutation in a liver specific gene leads to accumulation of neurotoxic metabolites which causes severe retardation in phenylketonuria, while accumulation of complex macromolecules causes mental retardation in lysosomal storage disorders. Many housekeeping genes expressed in diverse tissues can have pleiotropic effects on brain development and function. The X-linked NDP gene causing Norrie disease (a severe neurodevelopmental disorder), is expressed in fetal eye, lung and brain as well as adult brain and muscle (Chen et al, 1993). Pleiotropy is recognised as "the lifeblood of nosography in clinical genetics' and an 'essential quality of genetic syndromes' (Opitz, 1994).

1.7. The Objectives and Purpose of the project

The influence of the HGP has already realised improved molecular technologies for more rapid methods of gene localisation, identification and isolation, as well as computing and linkage methods for construction of genetic maps of the genome. The development of extensive genome maps and polymorphic markers has enabled the steps toward the initial localisation of genes involved with disease to be imminently achievable. Since an X-linked gene can be recognised by virtue of the definitive pattern of inheritance, the search for the genomic position is narrowed to 5% of the genome. These technical innovations were applied to accelerate the exegesis of X-linked disorders.

The most common cause of inherited retardation is X-linked mental retardation (XLMR) as evidenced by an excess of males in the retarded population. Special interest in XLMRs was instilled by the fra(X) syndrome because of its phenotypic expression, high frequency in the population, linkage to the cytogenetically visible fragile site and unusual segregation. The remaining equally large non-syndromal group of X-linked mental retardations (MRX) cannot be classified by clinical evaluation as no distinct features can be used to index these as different entities. The only means of classification lies in genetic delineation based on regional localisation of the gene borne in each family. The clinically heterogeneous syndromal forms of XLMR also await further classification on the basis of molecular evaluation. Regional localisations in all X-linked conditions with mental retardation as a feature, will define genetically heterogeneous disorders and be a step towards recognising potentially allelic disorders. Candidate genes within refined localisations can be analysed for mutations associated with disease, or positional cloning methods can be applied to the cloning of the gene mapped to a small interval.

Regional localisations of the genes causing other X-linked disorders provide the basis for identification and characterisation of defective genes and contribute to the body of information on the X chromosome. Gene identification on this ancient chromosome, conserved through millions of years of evolution, contributes to an understanding of human diversity and to the unravelling of complex mechanisms such as those that cause some genes to escape X inactivation. In the immediate and more altruistic sense, gene mapping in large families segregating disorders that are debilitating or incompatible with life eg. PDR, cardiomyopathy (Chapter 6), can be applied as a diagnostic tool for early ameliorative medical intervention or informed reproductive choice. Such applied research is highly valued by individuals in genetically compromised families.

The criteria for mapping disease genes by linkage include large families segregating the disease phenotype and bearing sufficient potentially informative meiosis to yield a lod score greater than +2 for X-linked conditions. The efficiency and accuracy with which genes can be mapped is dependent on the resolution of the background map and the heterozygosity of the markers within it. Genetic maps constructed from genotype data in the CEPH reference families have improved map resolution for many chromosomes, specifically the X. Regional fine maps with high resolution have been produced for specific intervals of the X and been integrated with physical data in some (Chapter 3).

It was with consideration of these factors that the objectives and strategies for this investigation have been devised. The goals for this project were;

1. To contribute to the map of the X chromosome.
2. To localise the genes for non-specific mental retardation (MRX) in 10 families.
3. To map and refine the localisations of genes for specific syndromal forms of mental retardation in five families.
4. To enumerate the number of X-linked genes for mental impairment and thereby contribute to the understanding of the distribution of XLMR genes.
5. To detect the mutation in a positional candidate gene.
6. To map the genes in large families with rare X-linked disorders, other than mental retardation.
7. To provide the genetic basis for carrier detection and prenatal diagnosis using linked DNA markers in these families.
8. To provide the basis for gene identification towards detection of the specific disease-causing mutation in these families.
9. To characterise the extent of any submicroscopic deletion of the X chromosome associated with mental retardation.

This thesis has been divided into eight chapters dealing with different aspects of these objectives. A total of 25 publications arising from data generated in some chapters have been attached as Appendices I-V.

CHAPTER 2

Materials and Methods

2.1. Introduction

This chapter is presented in two parts representing the 'wet' and 'dry' methods used to carry out this project. The laboratory methods of molecular genetic technology described briefly in Part I were used in generating reagents or data used in more than one of the subsequent chapters. Part II outlines the computer hardware and software required for linkage analysis and computations of risk. Most of these methods are established and in routine use in the Department of Cytogenetics and Molecular Genetics, WCH. Methods used exclusively for particular experiments are outlined in the corresponding chapter.

The main focus was to rapidly generate genotype data from DNA of numerous individuals within families segregating an X-linked disorder and then analyse the findings for statistical significance. The two main methods for genotyping were by Southern blotting and hybridisation and by the polymerase chain reaction (PCR). Analysis of RFLPs was performed by the established technique of Southern, while PCR was used to amplify di-, tri- and tetra-nucleotide repeat polymorphisms characterized by the candidate or extracted from the literature (Chapter 3). Polymorphic loci were also used as physical markers to characterise the extent of deletion mutations (Chapter 7). Genotyping strategies applied to improving efficiency of generating genotype data are described in Chapter 3.

Disease loci were mapped to the X chromosome using cosegregation analysis, which measures the genetic distance between DNA probes/markers and the gene responsible for the disease (Botstein et al, 1980). Linkage analyses that determine the likelihood of cosegregation of polymorphic markers with the disease phenotype were performed using the LIPED and LINKAGE computer programs. Their specific applications and the coding of input files are described in Part II of this chapter. Improvements in both computer hardware and software have permitted faster and larger computations to be performed. Other programs and databases utilised frequently are briefly described.

Part I: Laboratory Protocol

2.2. Patient Samples

The clinical criteria for each disease phenotype studied was determined and defined by the collaborating clinician involved with each family. The gathering of patient samples, clinical examinations and validation of pedigrees were freely contributed and recognised by co-authorship of the clinicians (see publications in Appendices II, III, IV and V). Peripheral blood samples (20mls) were collected with informed consent from normal and affected

individuals. All participating individuals within a family were examined, where possible, by the same clinician so that subjective assessments were founded on the same baseline and diagnosis was standardized. The correct phenotypic classification of each member of the family is critical to the detection of linkage, and was established prior to commencement of genotyping. Assumptions made in the linkage analysis are based on the clinical findings and unreliable or unsubstantiated data may greatly affect the reliability of the resultant map localisations. Linkage analyses in large single families segregating a particular disease phenotype are likely not to be troubled by the possibility of genetic heterogeneity within the family as all affected individuals are assumed to carry the same mutation in a single gene. Recognition of phenocopies or, in XLMR families, mentally impaired family members who are impaired for other reasons depends on the diagnostic skill of the clinician.

The most information for linkage can be obtained from obligate carriers who are doubly heterozygous for a marker locus and the disease locus and are phase known. In this case all her children will provide information for identifying recombinants. A fully genotyped three generation family or one in which the grandparental genotypes can be inferred is more informative than a nuclear family, since phase in the former is directly inferred but in the latter can often only be determined with some probability from the children's genotypes but cannot exclude alternative phases. It is for this reason that the CEPH families are large three generation pedigrees (Dausset et al, 1990). The expected lod score for a family is dependent on its size, structure, and knowledge of phase.

Families were accepted for linkage analysis if an X-linked mode of inheritance was evident as outlined in Chapter 1 and if sufficient potentially informative meioses could be collected to provide significant evidence for linkage. A critical number of eight informative meioses are required to achieve a lod score of +2 (section 2.7.). Members of families with fewer meioses have been collected, the DNA extracted and stored since currently mapping studies cannot localise these genes with statistical significance. These families may be screened in the future when specific candidate genes are available or the mutation known.

Each family in the following chapters was treated as an independent entity for the linkage analysis, therefore contributions of families from multiple investigators had no effect on outcome. In the study of MRX families, however, comparisons of any vague clinical findings based on overlapping map location of the disease gene may be instructive given that the same clinicians or others under their supervision were involved in establishing all of the diagnoses in families MRX1, MRX3, MRX10, MRX11, MRX12, MRX13, MRX17 and MRX18 (Chapter 4). By the same token, similarities in even such soft signs would have been detected by the astute practitioners examining these families.

2.3. Preparation of reagents

Extraction of DNA samples from the various tissue samples, DNA probes and oligonucleotide primers all followed established protocols as detailed in this section. Genomic DNA was extracted from duplicate blood samples of each MRX family member by the candidate. Samples already extracted by a contributing laboratory required confirmation of quantitation (section 2.3.2.) and infrequently repurification with phenol/chloroform (below).

Anonymous DNA probes (Chapter 3) that detect restriction fragment polymorphisms (RFLPs) on Southern analysis (section 2.4.) were prepared from plasmid DNA and frequently required transformation into bacterial hosts, before isolation (section 2.3.3.). Insert probes were excised from most plasmids, isolated and quantitated for labelling. Preparation of reagents for PCR including synthesis and purification of oligonucleotide pairs were as described below (section 2.5.).

2.3.1. Lymphocyte DNA isolation

DNA was extracted from lymphocytes by classical methods (modification of Wyman and White, 1980) including lysis with proteinase K and SDS, extraction with phenol and chloroform to remove proteins and cell debris, and precipitation with ethanol. Peripheral blood samples were collected into 2x10ml EDTA tubes and stored at -20°C until a batch of tubes were accumulated. Extraction was commenced on at least 12 patient samples at a time, duplicates were never included in the same extraction run with the paired tube to reduce the chance of complete loss of a sample or of undetected sample mix-up.

1. Frozen 10ml samples were thawed slowly, fresh samples could also be included into the batch at this time. Each blood was transferred to a correspondingly labelled 50ml polypropylene tube and the EDTA tube rinsed into it once, with cell lysis buffer (0.32M sucrose, 10mM TrisHCl, 5mM $MgCl_2$, 1% v/v Triton X-100) before the volume was made up to 30ml total with cell lysis buffer. Tubes were placed on ice for 30mins - 1hr.
2. The cell suspensions were centrifuged at 2000g, 4°C for 15mins, (3500rpm in Jouan centrifuge) and the supernatant removed by suction to the 10ml mark. A further 20ml cell lysis buffer was added and the cells gently resuspended by inversion before the spin at 2000g was repeated for 15mins.
3. The entire supernatant was removed leaving approximately 1ml of cells in a soft pellet. This pellet was resuspended in 3.25ml proteinase K buffer (10mM TrisHCl, 10mM NaCl, 10mM disodium EDTA), together with 0.5ml 10% w/v SDS and 0.2ml 10mg/ml

Proteinase K (Boehringer Mannheim) and incubated at 37°C overnight, or at least 4hrs, with constant agitation or on a slow wheel.

4. Buffer saturated phenol (equilibrated with 10mM TrisHCl, 1mM EDTA) added in 1:1 ratio, was mixed gently with the cell lysate by inversion on a wheel for 10mins, then centrifuged at 2000g for 10mins. The aqueous, upper phase was transferred to a 10ml polypropylene tube with another 5ml phenol, inverted, then spun at 400g for 10mins.
5. The top phase, again transferred to a fresh 10ml tube (using a wide bore pipette or plastic Pasteur) with care to avoid carryover of proteins (seen as a white layer) at the interface, was then extracted with chloroform:isoamyl alcohol (24:1) added to the 10ml mark and gently mixed by inversion or, since chloroform is volatile, by gentle repeated pipetting through a wide bore tip. The tube was then spun at 400g for 10mins.
6. Again the aqueous top layer was carefully removed to a new tube without carryover of the organic phase, and 1/10 v/v of 3M NaAcetate (pH 5.2) and two volumes of ice-cold pure ethanol were added. DNA can be seen to precipitate by inversion of the tube. This mixture can be placed at -20°C overnight in cases with low yield.
7. The cotton-wool like mass was transferred to an Eppendorf tube and centrifuged at 17000g at 4°C for 10-15mins. The pellet was washed in 70% ethanol, then dessicated under vacuum in a speed vacuum concentrator (model RH40-11, SAVANT Instruments Inc.).
8. The dried down pellet was redissolved in 20-100µl of TE (10mM TrisHCl, 0.1mM EDTA), estimated from the size of the pellet, and was adjusted to a final concentration of 1mg/ml following DNA quantitation.

DNA from other sources such as cultured cells, chorionic villus samples (CVS) or biopsies were extracted as above (from step 3) but scaled up or down based on the approximate mass of cells to be extracted. Repurification of DNA proceeded when necessary from step 4 (above) in a minimum volume of 500µl.

2.3.2. DNA quantitation

A small aliquot of each of the resuspended DNA samples was diluted into H_2O and mixed well. The known dilution was between 1:250 to 1:1000 in a total volume not exceeding 1ml. Samples were read in quartz cuvettes using the deuterium lamp of a spectrophotometer (Cecil CE2020) at a wavelength of 260nm. The first reading of pure H_2O was zeroed as the blank, and the optical density (OD) reading of each test sample recorded. The concentration of DNA was calculated using the formula OD_{260} x dilution factor x 0.05 =

μg/μl, with known OD_{260} and dilution factors. On average 10ml of blood yielded up to 300μg of DNA.

2.3.3. Plasmid DNA preparation

Plasmid DNA was prepared to make hybridisation probes by transformation and/or isolation (modified from Maniatis et al, 1982). The requirement for transformation depended on the way in which probes were supplied, those supplied as lyophilized or in solution plasmid DNA preparations were transformed into bacteria. Bacterial strains used were always *E.coli* that are recA$^-$ such as LE392, to prevent recombination during transformation and growth in media. Plasmids supplied as agar stabs or glycerol stocks could be directly streaked onto agar plates with the appropriate selective antibiotic, and isolation continued from step 3 below;

1. Transformation involves the preparation of fresh *E.coli* cells to make them competent. Following overnight culture in 10mls, fresh culture was inoculated (1:50) and grown to OD_{600} = 0.3-0.5, then cooled on ice 10mins. The cells were pelleted at 3000rpm for 5mins and resuspended in 5mls ice cold 100mM $MgCl_2$. Pelleted again the cells were then resuspended in 10mls ice cold 50mM $CaCl_2$ and held on ice for 20mins. The final pellet of competent cells was resuspended in 2mls of ice cold 50mM $CaCl_2$ and held on ice for transformation. For transformation 200μls of competent cells were mixed with up to 1μg of DNA and held on ice for 60mins. Heat shocked for 2mins at 42°C, the cells were returned to the ice for 10min. One ml of LB was then added and incubated at 37°C.
2. The incubation was left for 1hr and streaked onto plates containing the appropriate selective media. The plates were incubated at 37°C overnight.
3. Ten mls of Luria Bertani (LB) liquid media, most commonly containing ampicillin (50μg/ml), was inoculated with a single colony from a streaked agar plate. The culture was incubated with vigorous shaking at 37°C for up to 6 hours, then transferred to 100ml LB with ampicillin (50μg/ml) and incubation continued overnight.
4. The following morning, the cultures were cooled on ice for 15mins then spun in two 50ml tubes at 3000rpm for 15mins at 4°C in a Jouan CR3000 centrifuge. The cell pellet was resuspended in 300μl TE and glucose (50mM Tris-HCl, 20mM EDTA and 50mM glucose) and 60μl of 80mg/ml lysozyme (Boehringer) and left at room temperature for 5min, then ice for 1min. A further 5min on ice with 1.2ml fresh 0.2MNaOH, 1%SDS added to the cell suspension. Then 900μl of ice cold 3M potassium acetate (pH4.3) was added and mixed by inversion, then spun in the JA20 rotor of a Beckman J2-21M/E centrifuge at 4°C, 15K for 15mins. The supernatant was spun again in a fresh tube and

the pure supernatant transferred and mixed thoroughly by inversion with two volumes ethanol in another clean tube. DNA was pelleted by centrifugation for 15min.

5. The DNA pellet was washed in 70% ethanol, dessicated then resuspended in 200µl TE (10mM Tris-HCl, 1mM EDTA, pH8.0). Purification with 10µl DNase-free RNase (1mg/ml - Boehringer) incubated at 37°C for 15min, then with 2µl proteinase K (10mg/ml - MERCK) in 100µl of 3 x proteinase K buffer (10mM Tris-HCl, 10mM NaCl, 10mM EDTA) and 10µl of 10% SDS incubated at 37°C for 1hour removed RNA and protein contamination.
6. Phenol/chloroform extraction of the DNA was carried out as described (step 4, section 2.3.1.) but in scaled down volumes to fit 1.5ml Eppendorf tubes, and centrifugations were at 10000rpm for 10min in an Eppendorf centrifuge. The final plasmid precipitation was in 1/3 v/v 7.5M ammonium acetate and 2 volumes ethanol, rather than 1/10 v/v 3M sodium acetate, as the ammonia dissipated, whereas residual salts may inhibit labelling and restriction. The DNA was pelleted by centrifugation in an Eppendorf centrifuge (12000rpm, 15mins), vacuum dried and resuspended in 200µl TE.

2.3.4. Insert preparation

Cloned DNA inserts were prepared from the plasmid preparations (above) prior to random primed labelling reactions. Inserts were spliced from their plasmid vectors using restriction enzymes that excised the specific insert and that separated it by size upon electrophoresis from the remaining linear plasmid molecule. The digested plasmid products were run on a preparative agarose gel and the band representing the DNA insert was excised as an agarose block. The insert was isolated by i) freeze squeeze or ii) electroeluted into dialysis tubing.

To freeze-squeeze, the block was placed into screw-capped tubes and snap-frozen by immersion in liquid nitrogen. Tubes were then spun at 15000rpm for 10min in the Hettich centrifuge pre-cooled below 4°C. The supernatant was collected and the freeze, spin cycle repeated. The total supernatant was then clarified by a final 10min spin and transferred to a clean tube. The insert preparation was quantitated against a known molecular weight marker on an agarose gel.

Electroelution of DNA from the agarose gel slice required pretreated dialysis tubing (boiled 10mins in 2% sodium bicarbonate, 1mM EDTA, then rinsed and boiled 10mins in distilled water) containing 0.5xTBE (5x = 0.45M Tris/0.01M EDTA/0.45M boric acid). The gel block was placed into the tubing and electrophoresed at 100V for 1-2 hours in 0.5xTBE to release the DNA into the buffer. The current was reversed for 2mins and DNA recovered

by ethanol precipitation of the buffer within the dialysis tubing. Those probes that were not prepared in one of these ways were labelled by nick translation of the whole plasmid (section 2.4.4.).

2.4. Southern analysis

DNA samples were digested, then size fractionated by electrophoresis through agarose gels (sections 2.4.1. -2.4.2.) and then denatured and transferred to nitrocellulose membranes (Gene Screen Plus, NEN) by the method modified from Southern (1975). The restriction fragments were analysed by hybridisation to labelled specific probes.

2.4.1. Restriction digests

Restriction enzymes were usually supplied together with the appropriate specific buffer eg. *Taq*I (Boehringer Mannheim) and *Eco*RI (New England Biolabs), or universal buffer. Genomic DNA (8μg) was cleaved with 30-40 units of restriction endonuclease in the compatible restriction buffer. To avoid star activity, glycerol levels or restriction enzyme volume did not exceed 1/10 of the final volume. Digests were incubated overnight at the prescribed temperature recommended by the manufacturer.

Digestion efficiency was monitored on an agarose minigel, where complete digestion was visualized, upon ethidium bromide staining, as a smear in the lane and incomplete digestion was indicated by an excess of high molecular weight material. DNA digestion reactions were stopped and prepared for electrophoresis by adding 1/10 v/v of 10x loading buffer: 15% Ficoll 400 w/v, 0.05% bromophenol blue w/v, 0.05% xylene cyanol w/v OR 0.1M Tris-HCl pH8.0, 0.2M EDTA pH8.0, 2% Sarcosyl w/v, 20% Ficoll 400 w/v, 0.1% bromophenol blue w/v, 0.1% xylene cyanol w/v. EDTA in the 10x loading buffer has an enzyme inactivating role, while the two dyes separated at different molecular weights under electrophoresis.

2.4.2. Agarose gel electrophoresis

Agarose gels (0.8%-1.4% w/v in 1xTBE) were cast in 14x11cm gel trays (BRL) sealed with autoclave tape at either end, using 14 tooth combs. Resolution of DNA in the range 0.5-20kb was achieved by electrophoresis in 1xTBE at 12-25mA constant current for 16 hours. Small analytical gels, or mini-gels, were poured to the limit of surface tension on glass slides cleaned with ethanol and electrophoresed at 100-150V for 40min.

Molecular weight standards were run on each gel for determination of band sizes as appropriate; *Eco*RI digested SPPI bacteriophage (DMWS1 - Bresatec) has an effective range

0.5-8.5kb, while lambda digested with *Hind*III and mixed with *Hae*III digest of ϕX-174 DNA (DRIgest III - Pharmacia) has a range 200bp to 23kb. Separation could be estimated from migration of the two dyes in the loading buffer, but DNA was visualized by staining for 30mins in 1xTBE with ethidium bromide (0.0025mg/ml) and UV transillumination. Gels were photographed alongside a ruler, to preserve a record of digestion status and for fragment size determination using the DNASIZE program (section 2.13.). Photographic records were made of each gel with a Polaroid camera (Polaroid film Type 667) through a Wratten red gelatin filter (Kodak).

2.4.3. Southern transfer

To set up DNA transfer, gels were submerged and rocked first in denature solution (2.5M NaCl, 0.5M NaOH) for 30min, then in neutraliser (1.5M NaCl, 0.5M Tris-HCl pH7.5) for 30min. When the expected products of the restriction digest included a population of large fragments the gel was acid nicked in 3.5ml HCl/500ml H_2O for 20mins before denaturation. The denatured gel was inverted and placed onto the filter paper wick bridge of a Southern transfer tray. These trays comprise resevoirs of 10xSSC at either end connected by a paper wick with each end partly submerged in the solution and supported on a glass plate bridge. A piece of Gene Screen Plus membrane (Dupont) cut to size and labelled with the date and gel number, prewet first in water then equilibrated in 10xSSC, was placed face down over the gel ensuring that no bubbles of air were trapped between the gel and filter, and gel and membrane. Several pieces of filter paper cut to size and prewet in 10xSSC were then layered on top, again removing any bubbles. The exposed wick around the gel was masked with Glad Wrap or other plastic. A bundle of absorbent paper placed finally on top was held in place with a glass plate but was not weighed down. The transfer was allowed to proceed overnight.

The following morning the absorbent paper stack and filter paper wad were removed and discarded. The gel and membrane were lifted off together and turned over so that the wells of the gel could be marked on the membrane. The gel was then peeled away and discarded or restained to confirm DNA transfer. The membrane, or Southern blot, was denatured in 0.5M NaOH for about 60secs and neutralised in 200mM Tris-HCl pH7.5, 2xSSC for a further 60secs before being allowed to air dry at room temperature.

2.4.4. ^{32}P Labelling and Hybridization reactions

The deoxynucleoside triphosphate dCTP labelled with phosphorus-32 (^{32}P) in the alpha position was purchased from Amersham. This reagent was supplied in aqueous

solution (1mCi/ml). A probe is single stranded DNA radiolabelled with ^{32}P isotope. All labelling reactions for probes used in Southern blot hybridizations made use of labelling kits supplied by Amersham. The nick translation kit (Amersham Kit N5000) was used to label approximately 250ng of whole plasmid DNA and followed the protocol provided. The purified plasmid (section 2.3.3.) was linearised by digestion with an enzyme with only a single recognition site within the plasmid. Following nick translation unincorporated radioactive nucleotides were removed by separation of the labelled product on a Sephadex G-50 column.

Preparation of excised inserts from plasmids permitted more specific probes to be available for labelling (section 2.3.4.). The Amersham Multiprime DNA labelling kit (RPN 1601) makes use of the random hexanucleotide primed DNA synthesis procedure of Feinberg and Vogelstein (1983). Only 50ng of linearised DNA is required/20ml hyb mix (see below). The Megaprime™ kit (RPN 1604/5) available from Amersham is a modification of this procedure and uses nonamers for faster labelling. The major difference between the three kits is the quantity of source DNA required and the time taken for sufficient incorporation. The nick kits require 1-1.5 hrs incubation at 14°C to reach optimal incorporation, while the Multiprime kit requires only 30mins at 37°C and the Megaprime only 10mins at 37°C.

Southern blots were hybridised in sealed plastic bags or in Hybaid bottles. When two membranes were labelled in the same container, they were placed back to back. Hybridisation mix (5xSSPE, 1%SDS, 50% formamide, 10% w/v dextran sulphate MW500,000) made with freshly deionised formamide[1] was stored at 4°C for up to 1 month.

1. Membranes were prewet in 5xSSPE, and bubbles from between walls of container and the membrane were removed. The 5xSSPE was discarded and replaced with 20mls prewarmed hybridization mix.
2. Salmon sperm DNA[2], freshly denatured by boiling for 5mins, was added 200µl/20mls (final conc. 100µg/ml), mixed well and poured into hybridisation container removing bubbles. Pre-hybridisation by agitation (or rotation of bottles) at 42°C was rarely longer than 30mins while labelling reactions were prepared.

---

[1]Formamide is deionised as follows: 5g mixed bed resin (BioRad) is stirred into 100ml formamide for an hour at room temperature, then filtered to remove resin. Can be stored frozen.

[2]Salmon sperm DNA (Sigma) is prepared by dissolving 250mg in 25ml of water. The DNA (10mg/ml) is sheared through a heavy guage needle and boiled in a microwave. DNA can be fragmented by sonicating however manual shearing has been equally effective.

3. The labelled probe was heat denatured at 100°C for 5mins then snap chilled, or spun and added directly to the warmed solution in each container. Air bubbles were excluded and well sealed containers returned to incubate with agitation at 42°C overnight.

2.4.5. Washing

Following hybridisation, membranes were washed twice (changing the solution) in 3-500ml of wash 1 (2xSSC, 0.5% SDS) at 65°C for 10-15mins with agitation to remove unbound probe. A Geiger counter used to check the signal, gave a good indication of the necessity to proceed to wash 2 (0.1xSSC, 0.1% SDS) twice for 10mins at 65°C. The stringency of wash was increased with time and temperature and by decreasing the concentration of the solutions used. The stringency required depended on the background, though most labelling reactions were washed adequately in wash 1, few needed further washing in wash 2. The membranes were blotted briefly and wrapped in Gladwrap.

2.4.6. Autoradiography

The position of hybridised DNA complementary to the radiolabelled DNA probe was visualised by autoradiography. After washing, membranes were exposed to XO-mat XK-1 (Kodak) film at -70°C in the presence of intensifying screens. Exposure times were estimated from the level of signal detected with the Geiger counter or after developing film following an overnight exposure.

2.4.7. Re-use and Storage

The hybridised probe was stripped from all membranes to permit longterm storage or re-use of the filter membrane with a different probe. Radioactive signal was removed by incubation at 42°C with agitation for 30mins in 0.4M NaOH, then a further 30mins at 42°C in 0.2M TrisHCl pH7.5, 0.1xSSC, 0.1% SDS. Very hot filters were washed in 0.1xSSC, 0.1% SDS at 95°C for up to 1 hour to strip them clean. The membranes were then air dried and stored in envelopes of Whatmann paper for future use.

2.5. Polymerase Chain Reactions (PCR)

The polymerase chain reaction (PCR) was used to amplify target sequences for genotyping of microsatellites, physical mapping, sequencing and mutation detection. All PCR were performed in a Perkin Elmer Cetus Thermal Cycler (model 480). The reaction components and optimisation of conditions are described below. Primer design and source are covered in Chapter 3. Primer pairs were synthesized on an Applied Biosystems 391

DNA Synthesizer or Beckmann Oligosynthesizer by K Holman and J Spence, WCH. Column sizes were 30nM or 200nM. Three thermal cycling conditions with different times and/or temperatures for denaturing, annealing and extension were utilised:

File 21: 94°C (1min), 60°C (1.5min), 72°C (1.5min) for 10 cycles, then 94°C (1min), 55°C (1.5min), 72°C (1.5min) for 25 cycles, ending with extension at 72°C (10min) before switching to a 4°C soak.

File 15: 55°C (2min), 72°C (3min), 94°C (2min) for 25 cycles, then 55°C (2min) ending with extension at 72°C (10min) and a 4°C soak.

File 7: 50°C (2min), 72°C (3min), 94°C (2min) for 25 cycles, then 55°C (2min) ending with extension at 72°C (10min) and a 4°C soak.

2.5.1. Oligonucleotide purification and quantitation

Cleavage of oligonucleotides from columns was achieved by ammonium hydroxide saturation followed by butanol precipitation; oligonucleotides were cleaved and deprotected in 1ml NH_4OH overnight at 55°C or three hours at 70°C. Cooled to room temperature, the oligos were transferred to tubes large enough to accept 20x the volume of n-butanol. Vortexed for 15secs, the tubes were centrifuged in the appropriate centrifuge; Hettich or Jouan for 5min full speed or 40min 3000rpm respectively. The supernatant was carefully poured off and the pellet resuspended in 1ml water. Vortexed with another 20ml n-butanol for 15secs the spin was repeated the supernatant removed and the pellet dried under vacuum. Pellets were resuspended in 50µl for the smaller 30nmole columns or 300µl for the larger 200nmole columns prior to quantitation.

Quantitation was as for genomic DNA samples (section 2.3.2.), except that a 1 in 500 dilution was examined for optical density. 1OD≅40µg/ml. The concentration (µg/ml) was calculated by OD_{260} x 40 x diln factor. Each primer was adjusted to 1mg/ml for storage at -20°C. Only a small aliquot of each pair of primers was mixed and diluted to 150ng/µl ready for use. The molar concentrations of each primer could be calculated by taking into account the base composition in the following formula (Hughes, 1993):

$$\frac{OD_{260} \text{ x dilution}}{A(15,200)+T(8400)+G(12,010)+C(7050)}$$

Application of this formula was necessary only when published primers did not work in any of the standard conditions (section 2.5.3.) and the published method had to be used, or in cases where a primer was supplied with only the molar concentration.

2.5.2. PCR reaction buffers

2 x PCR reaction mix (modified from Kogan et al, 1987):

Standard Mix	Low dNTP Mix
33mM $(NH_4)_2SO_4$	33mM $(NH_4)_2SO_4$
133mM TrisHCl (pH 8.8)	133mM TrisHCl (pH 8.8)
20mM β-mercaptoethanol	20mM β-mercaptoethanol
13μM EDTA (pH 8.0)	13μM EDTA (pH 8.0)
0.34mg/ml bovine serum albumin (BSA)	0.34mg/ml bovine serum albumin (BSA)
20% dimethyl sulphoxide (DMSO)	20% dimethyl sulphoxide (DMSO)
3mM dATP	400μM dATP
3mM dCTP	400μM dCTP
3mM dGTP	400μM dGTP
3mM dTTP	400μM dTTP

2.5.3. Optimisation of priming conditions

All PCR reactions relied on the heat stability of the *Taq* polymerase isolated from the thermophilic bacterium *Thermus aquaticus*. This polymerase requires Mg++ for activity and must be optimised for efficiency over the range 1mM to 10mM. All primer pairs flanking microsatellite markers, including those designed by the candidate and those extracted from the literature, and PCR converted RFLPs (eg. DXS164-Chapter 3) were optimised by varying a range of $MgCl_2$ concentrations and different cycling files in a cold (non-radioactive) reaction.

The 20μl incubation mix included:

10μl 2x PCR reaction mix

6μl $MgCl_2$ at a range of final concentrations from 1.5-9.0mM

1μl oligo primer pair (150ng/μl of each oligo)

1-2μl template DNA (100ng/μl)

1μl *Taq* polymerase (0.5units)

Mineral or paraffin oil overlayed on top of the incubation mix prevented condensation in the tube.

Product size and integrity was checked on 1.4% agarose gels stained with ethidium bromide as for Southerns, against the molecular weight marker DMW-P1 (pUC19 DNA digested with *Hpa*II purchased from Bresatec) with bands ranging from 110 to 501bp. In "standard" 2xPCR mix (section 2.5.2.) most markers amplified in 4.5-6.0mM $MgCl_2$ on the file 21 cycling template. The recent use of low dNTP mix has latterly circumvented the need

to optimise each oligo pair as the majority worked satisfactorily at 1.5mM $MgCl_2$ final concentration. Only those that had not amplified were taken through the optimisation procedure at a lower range of Mg++ concentrations. Primers designed flanking known polymorphic restriction sites take advantage of PCR technology by RFLP-PCR. Following amplification, half of the sample was subjected to restriction digestion with the appropriate enzyme, the other half was run uncut on the same gel to confirm amplification of the correct product.

2.5.4. Hot PCRs

Microsatellite PCR products were radio-labelled by incorporation of ^{32}P-dCTP during amplification. The optimum Mg++ concentration as determined above was used in a 10µl reaction and then run-out on the BioRad PAGE system (see below). Incubation reaction mix:

5µl 2x PCR reaction mix

1µl $MgCl_2$ (at optimal concentration)

1µl oligo primer pair (150ng/µl of each oligo)

1µl template DNA in TE buffer (100ng/µl)

1µl *Taq*I polymerase 1/10 dilution (0.5units) (Boehringer Mannheim)

with 0.25µCi ^{32}P dCTP (Amersham) diluted with H_2O or 10x PCR dilution buffer (16mM ammonium sulphate, 67mM Tris-HCl (pH8.8 at 25°C), 10mM β-mercaptoethanol, 6.7µM EDTA, 170µg BSA/ml soln).The mix is topped with one drop of paraffin oil.

For multiplex PCR involving two or more pairs of primers, 75ng/µl of each oligo were added to the minimal $MgCl_2$. After PCR amplification, three volumes of formamide stop solution (95% deionized formamide, 1mM EDTA, 0.1% xylene cyanol FF, 0.1% bromophenol blue) were added to each reaction, then denatured at 94°C for 5mins before loading on a 5% polyacrylamide gel. The addition of formamide eliminates many additional bands (section 2.5.5.) making scoring of alleles simpler (Todd, 1992).

2.5.5. Polyacrylamide gel electrophoresis (PAGE)

The PAGE system with specific varied gel compositions was used for analysis of microsatellite alleles, sequencing and SSCA (see below). In each case the gel mixture was filtered to remove undissolved crystals of urea and other impurities, and stored up to one month protected from light. In preparation for casting a gel, Biorad IPC sequencing apparatus (21X50 or 38X50cm) was washed with detergent, rinsed with ethanol and dried thoroughly. A thin film of Sigmacote (Sigma) was applied with a tissue to the IPC surface

only, to prevent gel adhering to both plates at end of run. Glass plates were sandwiched using 0.4mm spacers for microsatellite analysis and then clamped together. A sealing gel was first cast in the casting tray which was clamped to the bottom of the IPC apparatus. The gel was then poured between the glass plates following the addition of 0.025% fresh ammonium persulphate (APS) and 0.1% TEMED (N,N,N,N-tetra-methylethylenediamine). Combs were clamped into place taking care to remove bubbles and the gel laid on a slight incline to polymerise for a minimum of two hours or overnight.

Inversion of the sharkstooth combs (2x24 teeth) so that teeth point down into the gel yielded 50 wells, with 4-6µl capacity. Buffer chambers in the base and at the back of the IPC were filled with 1xTBE to cover the gel. Pre-electrophoresis for 10-20mins in 1xTBE buffer prior to sample loading removed impurities and prewarmed the gel to 45-50°C. Wells were flushed out with buffer using a needle with a syringe. 5µl of denatured samples were run in at 900V until the two dyes of the loading buffer separated, then run at 2000-2500V (Biorad model 3000xi power supply) with temperature not exceeding 50°C, for 3-4hrs. Following electrophoresis the gel was transferred to 3MM Whatmann filter paper. The surface covered with plastic wrap was vacuum dried at 80°C for 1-2hrs on a Biorad 583 Gel Drier. The dried gel was exposed to Kodak X-Omat XK-1 film at room temperature overnight or at -70°C for 4hr, overnight or the appropriate length of time to visualise bands.

Polyacrylamide gel of 3.5-5% v/v (acrylamide:bisacrylamide at 19:1) prepared in 1xTBE with 7M urea was used to size fractionate amplified microsatellite products ranging in size from 90 to 400bp. Samples were electrophoresed through approximately 30cm of the substrate (on average) to achieve adequate resolution of alleles separated by only 1 or 2 basepairs. Scoring of dinucleotide microsatellites can be problematic in the presence of 'shadow bands' spaced at 2bp intervals to the major band (Chapter 3.3.3.). These artefactual bands are generated during PCR due to slipped-strand mispairing that lengthen or shorten the product by an integral multiple of the repeat unit length (Hauge and Litt, 1993). Although each allele consists of a series of fragments, the pattern is constant and the major band most intense.

2.5.5.1. Sequencing

Manual sequencing of DNA templates eg. AC positive subclones of anonymous M13 DNA (Chapter 3.3.1.) by the dideoxy chain termination method (Sanger et al, 1977) utilised the United States Biochemical corporation (USB, Cleveland, USA) Sequencing Kit (No.70700) or Perkin Elmer Cetus (PEC, Norwalk, USA) Sequencing Kit (No.N808-0001) according to the USB protocol. The annealing reaction contained 5-7µl (~1µg) template, 1µl

(0.5pmol) primer, 2μl annealing buffer made up to 13μl with H_2O and was heated at 70°C for 2min. When cooled 2μl labelling mix, 0.5μl ^{32}P-dCTP and 2μl (3units) sequenase (1/8 diln, USB) or 2.5units Taq polymerase (PEC) were added. This labelling reaction was mixed and incubated at 45°C for 5min. Aliquots of 4μl were combined with 4μl of each termination solution (A,C,G and T) and incubated at 70°C for 5min. Formamide stop solution (4μl) was added when cooled to room temperature. Termination reactions were heated to 95°C for 5min before PAGE. Gels of 6% polyacrylamide (19:1 acrylamide:bis-acrylamide)/7M urea were poured as above using 0.25-0.75mm wedge spacers. Electrophoresis proceeded for approx. 4-5 hours at 2000-2500V (50°C) then samples re-denatured and loaded again in adjacent wells to run another 2-3 hours. The gel was dried and autradiographed as described above for the analysis of microsatellite repeats.

Sequencing of PCR products was performed for mutation detection in the dystrophin brain promoter (Chapter 4) and in the SOX3 gene (Chapter 5), and to search databases for homology of pS8 to known genes (Chapter 7). Automated cycle sequencing using fluorescent dye labelled terminators utilised the Taq DyeDeoxy™Terminator Cycle Sequencing Kit (No.401113) and protocol (outlined below) and was run on an Applied Biosystems 373A DNA Automated Sequencer (ABI sequencer) (protocol given in Rev.E Applied Biosystems, Inc.). Radio-labelled 'manual' cycle sequencing used the Stratagene Cyclist™ Exo$^-$ *Pfu* kit (supplied with protocol). The dsDNA templates for the sequencing reactions were PCR amplified in 50ul reactions; 25μl standard 2XPCR mix, 0.5μl 1M βME, 5μl primers (150ng/μl each), 1ul Taq polymerase, 5ul 60mM $MgCl_2$. Using the Prep-a-Gene Purification Matrix (Bio-Rad) according to the manufacturers instructions, the PCR product was purified for sequencing and eluted into 10μl sterile deionised water.

For ABI sequencing using the Taq DyeDeoxy™ Terminator Cycle Sequencing Kit; the reaction premix containing each of four dye-labelled dideoxynucleotides (terminators), together with buffer and AmpliTaq DNA polymerase was mixed well and stored at 4°C until required. The sequencing reaction was set up using 9.5μl premix, approximately ~100-150ng of the dsDNA template and 3.2pmol (40ng pUC) primers made up to 20μl. The reaction topped with paraffin oil was put in a Perkin Elmer Cetus thermal cycler preheated to 96°C. Cycling followed the required conditions on file 59; 96°C for 60secs, 50°C for 15secs and 60°C for 4mins, but repeated 28 times (not 25 as specified), or file 67; 96°C for 30secs, 50°C for 15secs, 60°C for 4mins for 25 cycles. After cycling, the unincorporated terminators were extracted using phenol/chloroform and the resuspended pellet was electrophoresed on an ABI automated sequencer by Julie Nancarrow and Jozef Gecz. The resultant sequence was analysed by GeneScan™ (ABI) software.

The Cyclist™ Exo⁻ *Pfu* DNA sequencing kit (Stratagene) was supplied with a protocol for cycle sequencing by linear amplification of the sequencing product. In this method the reaction mixture of template, primer, buffer, labelled dCTP, Exo⁻ *Pfu* polymerase and DMSO were added to ddNTPs in four termination tubes. After initial denaturation at 95°C for 5min, sequencing incorporating radiolabelled nucleotides was done on file 95 for 30 cycles of 95°C for 30secs, 60°C for 30secs and 72°C for 60secs. The products were separated on prewarmed 6% PAGE and visualised following autoradiography.

2.5.5.2. SSCA

Single stranded conformational analysis (SSCA) aims to detect polymorphisms (SSCP) or rare variants associated with disease based on detecting altered electrophoretic mobility of a mutant DNA strand (Orita et al, 1989). Non-denaturing 4.5% gels for SSCA were poured in the BioRad IPC with 0.4mm spacers but the acrylamide composition differed (49:1 acrylamide:bisacrylamide), had no urea and contained 10% glycerol to improve resolution. The target sequence ie. the candidate gene or a segment thereof, for example, the dystrophin brain promoter (DBP) and SOX3 a single exon gene (Chapter 4 and 5) were amplified by PCR in optimised thermal cycling conditions and in the presence of [α-^{32}P]dCTP. A 2µl aliquot of the PCR product was diluted 1:5 in 0.1%SDS, 10mMEDTA (pH8.0) and heat denatured for 5mins in the presence of formamide loading buffer. 5µl of each sample was immediately loaded and run for at least 10hrs at 300-500V. The gel was dried and autoradiographed as described above.

2.6. cDNA screening

In order to isolate transcribed sequences from within the region deleted in a boy (CB) with mental retardation (Chapter 7.3.1.), the human fetal brain 5'-STRETCH PLUS cDNA library (Clontech HL3003b) was screened as follows:

20mls Luria broth with 0.2% maltose (w/v) and 10mM $MgSO_4 7H_2O$ was inoculated with a colony of *E. Coli* strain Y1090r⁻ cells and incubated at 37°C overnight. Host cells were resuspended in 10mls 10mM $MgSO_4 7H_2O$ and incubated for 30mins at 37°C with one library equivalent ie. $2x10^6$ plaque forming units. The library was plated out and plaques lifted as described in Chapter 3 except using Luria agar and making duplicate lifts of each plate. Screening for cDNAs was by hybridization to DXS296 markers, VK21A and pS8, which detected conserved sequences and were deleted in patient CB. Positive plaques found on duplicate plates were picked into 1ml SM buffer and stored at 4°C. Each of these plaques

was plated out, grown, lifted and rescreened on duplicate lifts. Plaques remaining positive after the second round of screening were stored as pS8 positives, VK21A positives and dual positives. Further characterisation of these clones was carried out by Dr. J Gecz.

Part II: Computational Methods

2.7. Linkage Analyses

The LIPED and LINKAGE computer programs were used to calculate lod scores for linkage analysis. In X-linked disorders, the test for linkage was declared significant when the maximum lod score exceeded the value of Zmax = +2 (Chapter 1.3.1.). This corresponds to an odds for linkage of 100:1 and is required to demonstrate linkage with 95% certainty. Once +2 is obtained, additional markers were genotyped in the region to confirm these findings and determine the limits of the regional localisation. Multipoint analysis may increase the lod score (odds in favour of linkage) if the original marker(s) were not fully informative. By convention, linkage was definitively excluded for recombination fraction values (θ) at which $Z\theta < -2$ (Ott, 1991). Initially only LIPED was available for determination of two-point lod scores by pairwise analysis. Later, LINKAGE permitted multipoint analyses and risk calculations to be performed from the same input files. All pedigrees initially coded for LIPED were later translated manually into LINKAGE files to enable multipoint (LINKMAP) and risk analyses (MLINK). Several families were coded directly for LINKAGE only. Both programs required that the mode of inheritance be known or assumed.

The computers used were; an NEC Powermate IV (386) operating under MS-DOS (ver 3.2), with 640K of core memory, a 40Mb hard disc, an 80287 mathematics co-processor and 512K of extended memory OR an IBM compatible 486DX2/66 with 4Mb RAM and 210Mb hard-drive under MS-DOS (ver 6.2). For larger multipoint analyses that required the capability to analyse several loci together, a mainframe computer at Sydney University was accessed through the internet on a MacIntosh IIE. This connection was achieved by logging on to the Adelaide University VAX and then telnet to ANGIS (Australian National Genomic Information Service) in Sydney. Files were transferred as outlined below (section 2.7.3.1.).

2.7.1. LIPED

Named for **li**kelihood in **ped**igrees, the LIPED program calculates two-point lod scores based on the Elston-Stewart algorithm, and has been generally available since 1974

(Ott, 1976). The coding of family files followed the guidelines given with the program. All data on family structure, genotypes and markers as well as penetrances and affection status were coded within one file with a male coded first. The program was chiefly used to screen genotyped markers for positive lod scores against the phenotype - this information being used to assist with the next marker selection, based on the recombination fraction at which the peak positive lod score was achieved, if any.

2.7.2. LINKAGE Programs

In 1985 the more sophisticated LINKAGE programs became available for gene mapping and enabled analyses at multiple loci and of genetic risks (Lathrop and Lalouel, 1984; Lathrop et al, 1984). The LINKAGE analysis package of programs (Version 5.1) comprise four analytical programs MLINK, LINKMAP, LODSCORE and ILINK. Only the first two (detailed below) were utilised in the analysis of affected pedigrees to generate lod scores in two-point analysis, and location scores in multipoint analyses.

MLINK was largely used for gathering evidence of linkage by preparing lod score tables of polymorphic markers against the disease phenotype. The program calculated pairwise lod scores at given recombination fractions for each marker. Adjacent markers with positive lod scores represented linkage groups to be scanned by LINKMAP. Risk analyses (Chapters 4 and 5) were also performed using MLINK as described below (section 2.8.).

Multipoint analyses using LINKMAP considers joint probabilities of genotypes at several adjacent loci simultaneously, subject to limitations outlined (section 2.7.3.) and overcomes missing data at markers that are not fully informative. Sequential analyses were performed to calculate the likelihood for the position of the disease locus with respect to a fixed genetic background map of markers (Chapter 3). One locus at a time was added and another dropped for each analytic window along the map. Genetic distances between markers in the background map were entered as recombination fractions. Location scores (S), twice the natural logarithm of the odds, were calculated at intervals outside and between sets of ordered markers. The peak location score obtained indicated the most likely position for the gene against that set of markers. Often, with the limited data available from single families, the curve tends to be flat over much of the interval however, and may not be useful in indicating the most reliable position to commence positional cloning of the gene. LINKMAP was therefore used only to maximise the linkage data in those families not fully informative at all loci. Multipoint location scores can be transformed back to lod scores by dividing the location score (S) by 4.6. The maximum possible multipoint lod score is equal to the maximum two-point lod score achievable when all meioses are informative.

To run the linkage programs the user must input two different files, the INFILE that contains information about the pedigree structure, affection status, and marker genotypes for each locus on each individual, and the DATAFILE that encodes the disease gene frequency, mode of inheritance and penetrance, as well as allele frequencies and order of the encoded marker loci. All of these parameters are coded as outlined below, then entered into the system using a text editing package and saved with the extension "-.inf" or "-.dat".

2.7.2.1. Construction of Input files

The pedigree INFILE is coded in columns (with each row representing the data on an individual) and follows the format: (1) pedigree number (that is the same for all members of a family), (2) ID number of an individual (in sequential order), (3) ID number of father (0 if unknown), (4) ID number of mother (0 if unknown), (5) the sex of the individual; 1=male, 2=female, (6) affection status or specific phenotype; 1=normal, 2=affected, (7 and on) the genotypes at locus 1, 2 etc coded as binary or allele numbers. Columns 1-4 should be a minimum two digit column width, with each column separated by at least one space from the adjacent one, one space should separate the alleles at each locus ie. each locus column must be a minimum of three digits wide. Any individuals that were not genotyped but necessary to the analysis for establishment of phase, for example the father and/or deceased affected male siblings were coded in columns 1-6 and then 0 0 at each locus. Similarly, in cases where non-paternity was suspected due to inconsistent or non-Mendelian inheritance and laboratory error had been excluded, the father was coded as null for all genotype data. Where the genotypes of more than one child were incompatible with the suggested father, another father was added without genotype data and the original father and his genotypes were retained for his compatible children.

Each INFILE locus corresponded with the description of that locus in the DATAFILE. Genotypes were encoded as binary factors or allele numbers. The latter are the numeric codes for each allele, eg 1 4 indicates that alleles 1 and 4 are carried in a heterozygous individual, while 2 2 is a female homozygous for allele 2, or a male[1]. Binary factors employing strings of 1 and 0 to indicate presence or absence respectively of an allele, were used mostly for two allele systems such as an RFLP. A 1 0 represents the 'slow' allele, 0 1 the 'fast', while 1 1 represents the heterozygote and 0 0 is unknown. The coding of each

[1]The input files for LINKAGE programs are not altered for X-linked conditions except that "sex-linked" is specified in the datafile. Males carrying only one X chromosome are hemizygous at all X-linked loci, but the coding of their genotypes in the infile requires their entry as homozygous for that single allele. This has no effect on the running of the program, which will crash if this is not done.

infile was checked for errors manually, by drawing the pedigree from the coded information to confirm family structure and by ensuring that parental alleles were compatible with mendelian inheritance of the genotypes in the children. Once entered, the INFILE was converted into a PEDFILE with the extension "-.ped" by executing the program MAKEPED.EXE and responding to questions asked.

The corresponding DATAFILE was prepared using a text editor, or in the program PREPLINK, and given the same filename but with the specific extension "-.dat". PREPLINK.EXE ensured construction of a reliable datafile since data were coded from responses to relatively user-friendly layouts. The number and order of loci, mode of inheritance (sex-linked or autosomal), mutation rate and risk locus were all defined here. Penetrance is the probability that a genotype will yield the predicted phenotype and is a reflection of the definition of the phenotype and the ability to recognise the expression of the causative gene (since gene expression may not be fully apparent in the phenotype). The penetrance equals 1 when all those inheriting the predisposing genotype manifest the phenotype. Clinically normal individuals who can be inferred to have the disease genotype are non-penetrant and indicate incomplete penetrance of the disorder. In those MRX families where incomplete penetrance was apparent or where some obligate heterozygotes were manifesting, an arbitrary 0.75 was used (Chapter 4), since reliance on data published in other families is not necessarily valid and estimation from very small families is difficult. Varying penetrance had little impact on gene location in the MRX families (section 2.10.).

The allele frequencies[1] at each locus must be specified as a decimal fraction of 1.0. In most cases where the actual population frequencies of alleles were not determined, the frequencies were coded as 1/n (with n=number of alleles at that locus) such that all alleles had equal frequencies. At some loci, where allele sizes were determined so that frequencies were known, alleles segregating in that family were coded with the published frequencies[2].

---

[1]The allele frequencies become important to the analysis only in rare cases where the genotype of an individual (particularly a founder at the top) is unknown and/or cannot be absolutely inferred and the likelihood analysis relies on the population frequencies of alleles to assign probabilities to possible genotypes. Erroneous designation of frequencies may result in false positive evidence for linkage or failure to correctly exclude linkage (Ott, 1992; Freimer et al, 1993).

[2]Since published allele frequencies represent crossections of each authors' local population and there may be significant variation in different populations, they may not accurately reflect the Australian population. Technically, sizing of a PCR product for comparison with published allele size may not be straightforward depending on method of labelling and type of polymerase (Knowles et al, 1992). In some cases the most common allele was identified from unrelated family members. More recently, as suggested (Knowles et al, 1992; Spurr, 1992), markers have been reported with the genotypes of CEPH individuals to standardize allele sizing across laboratories and methods.

Differences from lods calculated under 1/n were noted in Family M (Chapter 4). Due to program constraints the number of alleles frequently required the recoding of allelic information to reduce computation time (Ott, 1991). Recoding was most easily achieved by coding all married-in genotypes not contributing to the lod score as the same. Reusing allele numbers without affecting segregation of the haplotype could reduce alleles to four when there was only one mate per individual (Ott, 1978). For highly polymorphic microsatellites, a maximum of 6 alleles usually sufficed to cover the alleles encountered in any one family.

2.7.2.2. Linkage support programs (LSP)

Newer versions of LINKAGE have benefited from the addition of linkage support programs (LSP) that form a user-friendly shell for access to the core programs. The linkage control program (LCP), permits the user to set up linkage runs on any of the four programs. Under LCP, it was possible to select the input files, the program, the markers, the type of analysis and recombination fractions quickly on a series of screens. The set up was saved commonly under the default command name PEDIN (ie pedin.bat) or a command selected by the user, which then activated the program when entered at the DOS prompt. The results were viewed using the linkage report program (LRP) that similarly asked for information on the set up and program used, and then converted the output from the LINKAGE programs into tabulated results either to a file or directly to the screen as required.

2.7.3. Limitations on PC vs FASTLINK

The capacity of the computer used can exert limitations on the ability to perform linkage analyses, particularly using LINKMAP. The program constants for LINKMAP on a 386 PC included the delimiting factor of 96 haplotypes ie. no more than 5 biallelic loci or 2 loci of 6 alleles, plus the disease with two alleles. The effects of this constraint on the analyses became more evident with the application of highly polymorphic multiallelic microsatellite markers. The user-defined program constants were altered in the file **genc.pas** on the 486 PC to streamline the program by providing extra memory. For X-linked conditions the parameters were maxlocus=7 (number of loci); maxseg=64 ($2^{\text{maxlocus-1}}$); maxall=9 (alleles at a single locus); maxhap=96 (maximum number of haplotypes, n1 x n2 x ... where ni = number of alleles locus i); maxfem=4656 (joint genotypes for a female = maxhap x (maxhap + 1) /2); maxmal=96 (joint genotypes for a male = maxhap); maxind=50 (maximum number of individuals) and maxped=1 (maximum number of pedigrees). Following changes to constants, the program name identified in the first line of **swgen.pas** was checked and compiled in DOS by typing eg. *compile mlink*.

Faster versions of LINKAGE 5.1 have been dubbed FASTLINK of which version 2.1 became available in mid-1994 (distributed via email). The increased memory available using FASTLINK on ANGIS (in Sydney) allowed larger, swifter computations through analysis of up to 8 loci with up to 13 alleles but to a maximum of 1000 haplotypes. Recompiling of program constants became inaccessible and unnecessary since up to 9 biallelic loci could be analysed simultaneously.

2.7.3.1. File transfer to enable FASTLINK

To use the greater computational ability of the FASTLINK program, files coded on the PC were transferred to ANGIS (the Sydney system). The current set-up required that this be done through Ache (the Adelaide University computer system), accessed by telnet from a Macintosh IIE. Only the coded -.ped and -.dat files were saved to floppy disc (1.44MB) from the PC. Using the *'apple file exchange alias'* files were translated onto the Mac. hard disc. Files were then transferred using the following series of commands; telnet to Ache and logon using personal password, select *'file transfer protocol (ftp) enable'*, *'set transfer directory'* then *'send ftp command'*. At the prompt **ftp>** type *binary* then *mget 'filename'* to transfer all files. Check that files have transferred by typing *'ls'* at the **ache>** prompt. Telnet to ANGIS, logon and *mget* files from Ache, again the *'ls'* command will confirm transfer. Selecting *mapping programs* from the mainmenu then enabled FASTLINK to be set-up by the LSP to run as a job, or immediately on-line.

2.8. Quantitation of Genetic Risk

Risks determined by linkage are based on the pedigree and phenotypes within it. It is the conditional probability that an individual has a disease predisposing genotype given the phenotype and genotype information on all pedigree members. Risk is a derived numerical prediction of phenotype given a specific genotype in an individual and is expressed as a percentage likelihood of developing the disease. Calculations of genetic linkage and risk are interdependent. The accuracy of risk estimations reflects the reliability of the initial gene localisations to the smallest possible interval.

2.8.1. Diagnosis: Prenatal and Carrier Detection

Once the disease gene has been mapped within an interval flanked by informative markers, genetic risks can be conservatively estimated. MLINK can be used to determine genetic risk to an individual and with minor changes uses the established DATAFILE and INFILE for any given pedigree. The INFILE must be converted to a new PEDFILE (through

MAKEPED), selecting the individual to be assessed. For prenatal diagnosis, information on the sex, genotypes, siblings and parents of the fetus must first be added to the INFILE. The DATAFILE must be altered to calculate risks by defining the risk allele at the risk locus. The estimated genetic distances between markers flanking the gene localisation are entered as recombination fractions during the set-up by LCP.

If the genetic distance between the closest markers flanking the disease gene is say 10cM, this means that the gene is not more than 5cM away from the closest one. Flanking markers detect single crossovers between them at a frequency proportional to the distance between loci. Clearly the greater the distance involved the greater the chance of a crossover that renders the diagnosis uninformative even when both markers were heterozygous. When the map distance between two loci is 10cM or 0.1 Morgan, the recombination fraction θ = 0.1 or 10%. Diagnostic error can be expressed as the chance of double crossover that would pass undetected between flanking markers. Over 10cM, this is calculated by 0.05 x 0.05 = 0.0025. Based on this premise, prenatal and carrier diagnoses were offered with at least 99% confidence. In the absence of close flanking markers, diagnosis could be provided by two-point linkage (Chapter 5; Paper 4, Appendix III), but the possibility of error is more difficult to predict. Risks determined in MRX families in this study were calculated from mapping data independently established within each family (Chapter 4) against the background map of markers (Chapter 3). Uncertain parameters, such as penetrance of the disease gene and the recombination fraction between flanking markers, were applied conservatively and not reported within a support interval. Risk intervals may however more accurately reflect uncertainties in the analysis than a point risk estimate.

2.9. Confidence intervals

In two-point linkage, the confidence interval, or support interval, for theta (θ) can be estimated by an established technique (Ott, 1985). Confidence intervals are derived by plotting the lod score curve (unimodal likelihood) and taking values of the recombination fraction (or cM) corresponding to a lod score one unit less than the maximum, once linkage is significant (ie. greater than +2 on the X chromosome). Referred to as 'one-LOD-down' (Conneally et al, 1985), the confidence coefficient is 90-95%. Multipoint analyses generate complex likelihood functions for gene localisations against a background map and may have more than one maximal peak. Varying the genetic distances in the background map preserves the shape of the multipoint curve obtained for a given set of data, however the magnitude of the confidence interval may be affected (Suthers and Wilson, 1990). The significance of confidence intervals obtained by 'one-LOD-down' of the peak obtained by

multipoint linkage analysis, is unclear (Keats et al, 1989) and may be misleading (Lathrop et al, 1984).

The confidence interval for gene location is usually narrow for relatively common disorders where there are many families, and generally rather broad in rare conditions. Determination of a confidence interval is important to the provision of risk estimates in genetic counselling (Suthers and Wilson, 1990). Risk support intervals can be used by a counsellor to evaluate the accuracy of genetic risks calculated in diseases involving linkage heterogeneity (Leal and Ott, 1994) or in private syndromes (Paper 4, Appendix III). Computer simulation or resampling methods have been applied to calculation of the approximate confidence interval of the MRX1 gene location for application to carrier counselling (Suthers and Wilson, 1990; Chapter 4).

2.10. Limitations to linkage analyses

The defective gene(s) in more common disorders can be mapped in large numbers of small families, however genetic heterogeneity may become an issue. Linkage analyses in rare conditions especially, depend on segregation in large families. In X-linked disorders, a suitable pedigree must have a minimum of eight potentially informative meioses and should segregate a clinically defined phenotype which follows a Mendelian mode of inheritance. In this study linkage analysis was performed independently for each family, therefore ascertainment bias would not affect the recombination fraction. Errors in definition of the clinical phenotype or identification of phenocopies may affect gene localisation by linkage, suggesting genetic heterogeneity. Heterogeneity is already evident in MRX families independently of such errors (Chapter 1.6.2.).

Low informativeness and density of markers in given regions can affect gene mapping by linkage. The limitations of RFLPs have been minimised by the introduction of highly polymorphic markers (Chapter 3). These multiallelic markers introduce another limitation to linkage analyses through exceeding computing maxima and requirements for extra memory (section 2.7.3.). Recoding to reduce numbers of alleles can reduce computing time (see 2.7.2.1.), but may also reduce informativeness at a locus.

The correct specification of genetic parameters is of paramount importance to linkage analyses. A number of studies have examined the impact of miss-specifying mandatory factors such as allele frequencies (section 2.7.2.1.), penetrance and age of onset, where incorrect values have been found to inflate the lod score or cause false positive evidence for linkage under certain conditions (Ott, 1992, Freimer et al, 1993). Penetrance can be estimated by counting the proportion of affecteds amongst susceptible individuals. Misspecifying penetrance can lead to a slight underestimate of the lod score, however

Greenberg (1989) shows that penetrance can be estimated by maximizing the lod score. Varying penetrance in the MRX2 family between 0.4-1.0 had little effect on the lod score and did not alter the localisation (Hu et al, 1994). The penetrance in MRX families (Chapter 4) was not associated with age at onset, since in all cases the diseases were ascertained through children and do not appear to be age-dependant. Subclinical expression of the disease cannot be excluded until specific direct tests are available once the gene has been cloned and the mechanism understood.

2.11. Haplotype Analysis

The genotypes of an individual can be determined and arranged to represent the parental chromatid or haplotype associated with the disease. The haplotype is structured to minimize the number of recombinants (in obeisance with Mendelian laws) and in this way can be used to order markers with respect to one another and the disease gene. Haplotype analysis is a simple paper and pen method of examining the transmission of the disease gene through a pedigree in association with a set of linked markers.

In X-linked conditions since the males are hemizygous their genotypes are actually the haplotype inherited from their mother. Her remaining alleles thus correspond to the other haplotype given that the male was not a recombinant. Another son with the other haplotype, or a daughter who's paternal haplotype is known (the father's haplotype is always transmitted without recombination) can be used to confirm that the haplotyping is correct in the family. Recombinant individuals in the family data can be used to define the minimal region for gene localisation. These cross-overs in the closest proximal and distal markers flanking the region of proven linkage are the limits of the disease gene. For fine mapping further markers internal to these limits are genotyped with particular attention to the recombinant individuals. The interval can be refined if the recombination events occur consistently on either side. The appearance of double recombinants within a small interval are highly unlikely and indicate possible genotyping errors in the pedigree. Determination of the haplotype associated with the disease gene becomes useful in determining risk to other individuals in the pedigree.

2.12. Exclusion Mapping

Negative information from both two-point and multipoint analyses can be very useful in gene mapping since every informative meiosis gives data for or against locus position. The criterion of $Z \leq -2$ corresponds to a likelihood ratio of 100:1 against, so that at those values of θ where $Z=-2$, the loci can be excluded from linkage to the disease studied

(section 2.7.). Multipoint analysis using LINKMAP (above) can be used to gather intermarker exclusion data (Ellison et al, 1992) but can be extremely demanding of computer time and memory. Exclusions can be simply demonstrated by examination of two-point lod scores, particularly when large numbers of adjacent markers are used. The EXCLUDE program (Edwards, 1987) can be used to illustrate the increased probability of linkage to other parts of the chromosome as regions are excluded by negative two-point lod scores. This program however, gives no estimate of statistical support and is susceptible to the informativeness and density of the markers in each region.

2.13. Other computer programs/software

The Genome Data Base (GDB) holds information about genetic markers and disease loci, as well as primer sequences and heterozygosities. Online Mendelian Inheritance in Man (OMIM) is a database cataloguing disease phenotypes and clinical and genetic classifications. Both GDB and OMIM can be accessed via Internet connections on the Macintosh IIE for up-to-date information retrieval.

The MAPFUN utility program is part of the linkage package and is useful for conversion of genetic map units (cM) to recombination fractions or vice versa using any one of 7 mapping functions. Mapping functions take into account interference (Chapter 1.3.1.) which is assumed to be complete for closely linked genes. The genetic distance in morgans approximately equals the recombination fraction (θ) when θ is small (less than 0.1), but as the genetic distance between loci increases, θ approaches the 0.5 limit. Since multiple crossovers may occur between loci, the θ values cannot be added to determine map distances. Using MAPFUN, conversion of cM derived by CRIMAP under the Kosambi mapping function (assumes marginal interference) or map distances from publications is possible (Chapter 3.4.1.2.). LINKMAP analysis uses recombination fractions, while the remaining linkage programs use Haldane's mapping function (no interference). For the purposes of most analyses where the distance between markers does not exceed 10cM the difference between Kosambi and Haldane cM are minimal/negligible, therefore it is not necessary to convert over small distances.

DNASIZE was a program written by G. Suthers (1990) and was useful for calculation of hybridization band sizes against molecular weight markers on Southern gels. The distance run (in millimeters) of each known marker band was entered, against which the size of each unknown band could be determined from the relative mobility.

CHAPTER 3

The Map Of The X Chromosome

Note to Publications

Three papers have been published from the work summarised in this chapter. Reprints are bound in Appendix I in the following order:

1. **Gedeon AK**, Richards RI, Mulley JC (1991) Dinucleotide repeat polymorphisms at the DXS294 and DXS300 loci in Xq26. Nucl Acids Res 19/18:5087.

2. **Gedeon AK**, Holman K, Richards RI, Mulley JC (1992) Characterization 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.

3. Donnelly A, Kozman H, **Gedeon AK**, Webb S, Lynch M, Sutherland GR, Richards RI, Mulley JC (1994) A linkage map of microsatellite markers on the human X chromosome. Genomics 20:363-370.

3.1 Introduction

Molecular geneticists are the cartographers of the human genome - the making of maps being a central theme underlying research in genetics. Whether physical, genetic or cytogenetic, the map becomes the critical basis upon which phenotypes and genes can be correlated. One of the first steps to an effective approach to linkage mapping of disease genes is the establishment of the background genetic map of polymorphic markers on a panel of highly informative control families (Chapter 1.2.2.). The linkage analysis (Chapter 2.6.2.) may then assume a fixed map of ordered markers upon which to calculate likelihoods of the disease gene locus in each interval. Accurate mapping relies upon the informativeness and resolution (density) of the background genetic map and its component markers and is most reliable when based on integrated consensus data from several sources.

The unique inheritance of the X chromosome (Chapter 1.5.1.) has allowed intensive characterisation of many genes and markers along its length. Polymorphic markers at frequent intervals spanning the chromosome are a prerequisite to the localisation of a disease gene to a finite chromosomal region. For X-linked diseases, assignment of the corresponding mutated gene, by virtue of recognisable inheritance, requires a search of only 5% (or an estimated 164Mb) of the genome. Fine mapping of the candidate region is a necessary prelude to positional cloning or to screening of positional candidate genes. The gene mapping studies presented in the following chapters (4-6) required the use of polymorphic genetic markers of known order and intermarker distance along the X chromosome.

During the course of this research, genetic markers have evolved from RFLPs to PCR-based microsatellites (Chapter 1.4.) and resolution of the chromosome substantially increased, however the principles for mapping remain. Construction of the genetic map of the X chromosome is complex since there are $n!/2$ possible orders for n markers. Marker order requires validation by various means including integration with physical maps. This chapter discusses the strategies applied to genotyping polymorphic markers, the construction of a background genetic map and integration of different lines of evidence for validation of marker order on the X chromosome. Map integration is on-going and will ultimately improve map resolution to result in a consensus map useful in many applications.

Results presented in the following sections contributed to the development of new highly polymorphic markers (Appendix I, Papers 1 and 2) and construction of a comprehensive genetic map of 62 PCR-based markers spanning the entire X chromosome (Appendix I, Paper 3). Additional information in support of marker order has been extracted from the literature to derive a background map including 90 PCR-based (mainly

microsatellites) and 50 RFLP loci. The composite map has, and continues to, evolve through essential transitory steps over time. The gene mapping undertaken in this project has at various times relied heavily on the best available regional maps that were current at those times. In finalising this research for assessment it has been necessary to call a halt to the continual updating of map information. Gene localisations described in this volume are largely based on the background map below that represents a snapshot in time of the evolving map of the X chromosome.

3.2. Evolution of the maps of the X chromosome

The genome project (Chapter 1.2.1.) has influenced the exponential evolution of the genome map through improvements in technology and the introduction of a new generation of polymorphic marker (section 3.2.2.). Single chromosome workshops (SCW) have been introduced to handle the sheer volumes of data and CEPH consortia have been established to produce genetic maps of some chromosomes. The biannual Human Gene Mapping Workshops (HGM) have been reduced to administrative and overseeing functions to ensure uniform modes of data collection.

In the absence of a CEPH consortium for construction of an X chromosome consensus linkage map, the only available consensus maps are developed at X chromosome workshops (XCW) that focus on the physical relationships of markers and genes and only recently whole chromosome linkage maps. Held annually, the ultimate aim is to integrate the genetic and physical maps and improve marker resolution to achieve 100% continuity in contig coverage. The gradual progress towards contig closure of the physical map will eventually result in integration and a reliable consensus on order and distance between loci. Contributions to the latest XCW4 and XCW5 maps have been built into consensus maps but are rapidly outdated (Schlessinger et al, 1993; Willard et al, 1994). Evolution of the genetic map of the X chromosome is related to the polymorphic markers used in their construction.

3.2.1. RFLPs and VNTRs

An early genetic map of the X chromosome included nine single copy DNA sequences detecting RFLPs from the Xg locus at Xp22 to DXS15 (DX13) at Xqter and spanned 215 recombination units (or 260cM using Haldanes mapping function). It was demonstrated that although physical distances between pairs of loci may be the same, the genetic distances are variable eg. physically equidistant loci in Xq21 are separated by half the genetic distance compared with loci near Xqter (Drayna et al, 1984). One year later more than 20 loci were included in the map based on meiotic recombination frequencies, with a

total length of 185 recombination units, the short arm being roughly equivalent to the long arm genetically (Drayna and White, 1985).

At the commencement of this project the only approach to mapping the genes for X-linked disorders was by linkage analysis to RFLP markers (Chapter 1.4.1.). RFLP markers were collected into a RFLP probe bank (Table 3.1) representing over fifty loci spanning the X chromosome. Plasmids were requested from the originating authors (HGM 9, 10, 10.5, 11) and probes were prepared as inserts (Chapter 2.3.3.). The PIC at these loci varied from 0.07 to 0.49, with some probes detecting RFLPs at more than one enzyme system and some loci detected by more than one probe. Only two VNTR markers were known, one at DXS52 in Xq28, the other at DXS255 in Xp11. A strategy was developed to improve the efficiency of genotyping by Southern analysis (Chapter 2.4.).

3.2.1.1. RFLP Genotyping strategy

The relatively low heterozygosity of RFLP markers, labour intensive filter preparation and greater requirement for a DNA resource that is limited, were factors considered in establishing a strategy to conserve samples. Restriction enzymes were selected that revealed length polymorphism at more than one locus, for example *Taq*I or *Pst*I (Table 3.1). Preparation of initial filter sets included digests of obligate carriers, who by definition are heterozygous for the disease gene. Probes detecting variation at those enzyme systems were screened for informativeness of the marker, those with highest PIC screened first. The descendants of heterozygous females were followed-up, while the first filters were rescreened at other loci. Relatives that connected affected individuals were all genotyped as well as those that would help in inference of missing parental or grandparental genotypes. Unaffected relatives eg. female siblings were included only if they could provide information for linkage by inference of parental genotypes or if they had requested carrier status evaluation.

3.2.2. Emergence of Microsatellites

The introduction of PCR technology (Chapter 1.4.) led to the development of a new class of polymorphic marker that vastly improved the density and informativeness of markers available for mapping. Genotyping by PCR was more rapid than Southern analysis for gathering genotype information in families. Some RFLPs were converted to PCR assay eg. DXS164 (Roberts et al, 1989), whereby primers flanking the polymorphic restriction site selectively amplified the target prior to digestion with restriction endonuclease. The PIC, however, remained the same. Microsatellite markers (Chapter 1.4.2.) had numerous advantages over RFLPs including ease of genotyping by PCR, high PIC, random and

Table 3.1. List of RFLP markers used spanning the X chromosome. Information on source and availability are referenced in HGM.

Locus	Probe	Region	Enzyme	PIC
DXS143	dic56	Xp22.3	BclI	0.49
DXS237	GMGX9	Xp22.3	HindIII	0.44
DXS85	p782	Xp22.3-p22.2	EcoRI	0.48
DXS9	RC8	Xp22.2	TaqI	0.30
DXS16	pXUT23	Xp22.2	BclII,BglII	0.49,0.26
DXS16	pSE3.2-L	Xp22.2	MspI	0.49
DXS43	pD2	Xp22.2	PvuII	0.49
DXS92	pXG-16	Xp22.1	TaqI,HindIII	0.58,0.5
DXS41	p99.6	Xp22.2-22.1	PstI	0.49
DXS67	pB24	Xp21.3	MspI	0.07
DXS28	C7	Xp21.3	EcoRV	0.28
DXS268	J66	Xp21.3-p21.2	PstI	0.54
DXS164	pERT87-15	Xp21.2	TaqI	0.46
DXS269	P20	Xp21.2	MspI,EcoRV	0.48,0.48
DXS270	J-Bir	Xp21.2	BamHI	0.33
DXS148	cx5.7	Xp21.1	MspI	0.42
DXS206	XJ1.1	Xp21.1	TaqI	0.40
DXS206	XJ2.3	Xp21.1	TaqI	0.42
DXS84	p754-11	Xp21.1	EcoRI	0.35
DXS84	754	Xp21.1	PstI	0.49
OTC	OTC	Xp21.1	MspI	0.47
DXS7	L1.28	Xp11.4-p11.3	TaqI	0.35
TIMPI	TIMP-3.9x	Xp11.3-p11.23	BglII	0.44
DXS146	pTAK8A	Xp11.22	XbaI	0.46
DXS255	M27B	Xp11.3-cen	EcoRI	variable
DXS14	p58.1	Xp11.21	MspI	0.45
DXS1	p8	Xq11.2-q12	TaqI,HindIII	0.26,0.34
DXS106	cpX203	Xq12	BglII	0.46
PGKI	pHPGK-7e	Xq13.3	PstI	0.49
DXS72	pX65H7	Xq21.1	HindIII	0.49
DXYS1	pDP34	Xq21.31	TaqI	0.44
DXS178	p212/9	Xq21.33-q22	TaqI, XbaI	0.42,0.49
DXS87	pA13.R1	Xq21.33-q22	BglII	0.48
DXS101	cX52.5	Xq22	MspI	0.45
DXS17	S21	Xq22	MspI,TaqI	0.18,0.45
DXS94	pXG-12	Xq22	PstI	0.48
DXS42	p43-15	Xq25	BglII	0.30
DXS10	36B-2/6A-1	Xq26.1	TaqI	0.44
DXS86	St1	Xq26.1	BglII	0.40
HPRT	HPRT	Xq26.1	BamHI	0.37
DXS51	52A	Xq26.2-q26.3	TaqI	0.49
DXS102	cX38.1	Xq26.3-q27.1	TaqI	0.18
F9	pVIII	Xq26.3-q27.1	TaqI	0.37
DXS105	cX55.7	Xq27.1-q27.2	TaqI	0.11
DXS98	4D-8	Xq27.2	MspI	0.29
DXS369	pRN1	Xq27.2-q27.3	XmnI,TaqI	0.48,0.24
DXS297	VK23B	Xq27.3	XmnI,HindIII	0.49,0.34
DXS296	VK21A	Xq27.3-q28	TaqI	0.22
DXS296	VK21C	Xq27.3-q28	MspI	0.30
IDS	c2S15	Xq27.3-q28	StuI,TaqI	0.49,0.07
DXS304	U6.2	Xq28	TaqI	0.30
DXS374	1A1	Xq28	PstI,TaqI	0.48,0.30
DXS52	St14-1	Xq28	TaqI	0.77
DXS15	DX13	Xq28	BglII	0.5
F8C	F8	Xq28	HindIII	0.29

ubiquitous distribution in the genome. Their abundance represented a rich source of potentially informative markers to expand the genetic linkage map.

The anticipated boost to the resources for linkage mapping in particular was not overlooked by the scientific community and many groups became involved in isolating and mapping polymorphic dinucleotide repeats (section 3.3.). Initially there were only 6 dinucleotide repeat markers available on the X chromosome (Litt and Luty, 1989; Weber, 1989). As more polymorphic X-linked microsatellite markers began appearing in publications however, it became clear that most of these new loci were not mapped against existing RFLP markers therefore their relative positions were not known. Markers selected for use in gene mapping must satisfy certain criteria including heterozygosity and ease of genotyping, but must also be reliably placed with respect to others on the genetic map.

3.3. Development of AC repeat markers on the X chromosome

The construction of a library of PCR-based polymorphic markers was fundamental to increase map resolution and marker availability for gene mapping during the course of this study. Given the physical length of the X chromosome (164Mb) more than 3000 ACn repeats are expected if they occur at ~50kb intervals (Chapter 1.4.2.), while a minimum of 24 markers at 10cM intervals are required to map a disease gene within 5cM of the nearest marker (section 3.4.1.1.). The library of microsatellite markers was compiled from loci published in the literature or characterised by the candidate (section 3.3.1.). PCR primer pairs were selected for oligonucleotide synthesis (Chapter 2.5.) based on two main criteria; i) position - primarily to fill gaps in chromosome coverage, and ii) the PIC or heterozygosity[1] values cited - where markers of greater potential informativeness were selected with priority. Initially all markers published were collected since they fulfilled the latter criteria for informativeness with respect to RFLPs. By the end of 1991, 31 X-linked, PCR-based markers with minimum heterozygosities of 50% had appeared in the literature (Appendix I, Paper 2). Less than 3 years later Généthon alone had published 80 more (Gyapay et al, 1994). Primer sequences for early Généthon markers were extracted from the Généthon Microsatellite Map Catalogue (1992) and later from Gyapay (1994).

The assembled library of microsatellite including 40 loci on the short arm and 50 on the long arm with three PCR-based RFLP markers, are tabulated (Table 3.2) with reference and heterozygosity information. Although referred to as microsatellites, the vast majority of

---

[1]Expected heterozygosity and PIC values are calculated from the allele frequencies (Chapter 1.4.1.), while observed heterozygosities represent the proportion of females with two alleles. For X-linked loci, heterozygosity should equal the PIC value since the male parent is never heterozygous.

Table 3.2 Polymorphic microsatellite repeat markers spanning the X chromosome

LOCUS	Clone Name	Alleles	Het/PIC†	Location	REFERENCES
DXS996	AFM212xe5	10	0.81	p22.3	Weissenbach et al, 1992
DXS237	GMGX9AC	8	0.44p	p22.3	Gedeon et al, 1992
KAL	KAL	5	0.61o/0.72p	p22.3	Bouloux et al, 1991
DXS987	AFM120xa9	10	0.83	p22	Weissenbach et al, 1992
DXS207	pPA4B	9	0.83	p22.1-p22.2	Oudet et al, 1992
DXS418	P122	11	0.83	p22.1	Van de Vosse et al, 1993
DXS999	AFM234yf12	7	0.75	p22	Weissenbach et al, 1992
DXS443	pRX324E1	7	0.60o	p21.2-p22.3	Browne et al, 1992
DXS365	pRX314E1	9	0.57o	p21.2-p22.3	Browne et al, 1992
DXS451	kQST80H1	9	0.80o	p21.2-p22.3	Browne et al, 1992
DXS989	AFM135xe7	9	0.80	p21-p22	Weissenbach et al, 1992
DXS992	AFM184xg5	11	0.87	p21-p22	Weissenbach et al, 1992
DXS1234	3'DMD(MZ18/19)	4	0.35	p21.3	Oudet et al, 1990
DMDC2	DMD1c-2	5	0.50	p21	Powell et al, 1991
DXS1235	STR50	6	0.72	p21	Clemens et al, 1991
DXS1236	STR49	19	0.93	p21	Clemens et al, 1991
DXS997	AFM217xa5	4	0.65	p21-p22	Weissenbach et al, 1992
DXS1237	STR45	13	0.89	p21	Clemens et al, 1991
DXS1238	STR44	12	0.87	p21	Clemens et al, 1991
DYSIII	–	4	0.51o/0.59p	p21.1	Feener et al, 1991
DYSII	–	8	0.82o/0.77p	p21.1	Feener et al, 1991
DYSI	–	5	0.79o/0.61p	p21.1	Feener et al, 1991
5'DYSMSA	5'DMD(NM72/73)	10	0.57	p21	Oudet et al, 1991
DXS538	XL27-B	7	0.72o/0.79p	p11.21-p21.1	Browne et al, 1991
CYBB	–	8	0.76	p21.1	Hardwick et al, 1993
DXS1068	AFM238yc11	11	0.82	p11.4-p21	Weissenbach et al, 1992
DXS228	1aA6	5	0.53o	p11.3-p11.4	Coleman et al, 1991
DXS7	L1.28(AC)	4	0.52p	p11.3-p11.4	Moore et al, 1992
MAOA3	–	15	0.75	p11.23-p11.4	Hinds et al, 1992
MAOA	MAOCA-1	7	0.69p	p11.3	Black et al, 1991
MAOB	MAOBTG-1	6	0.64o/0.73p	p11.23-p22.1	Grimsby et al, 1991
DXS1003	AFM276xf5	10	0.79	p11	Weissenbach et al, 1992
SYNI		7	0.84	p11.2	Kirchgessner et al, 1991
PFC	Properdin	5	0.65	p11.23-p11.3	Coleman et al, 1991
DXS426	XL91B10	12	0.52o	p11.21-p21.1	Luty et al, 1990
DXS1126	EAD	8	0.51o/0.67p	p11	Donnelly et al, 1994
DXS573	HX82	5	0.72o	p11.1-p11.22	Roustan et al, 1993
ALAS2	–	9	0.78	p11.21	Cox et al, 1992
DXS991	AFM151xf6	9	0.81	p11	Weissenbach et al, 1992
AR	AR(CAG)$_n$	-	-	q11-q12	La Spada et al, 1991
PGK1P1	PGK/5	3	0.42p	q12	Browne et al, 1991, 1992
DXS106	–	11	0.81p	q12	Fairweather et al,1993
DXS1125	46AD	9	0.77o/0.79p	q11-q12	Donnelly et al, 1994
DXS453	Mfd66	15	0.63p	p11.23-q21.1	Weber et al, 1990
DXS559	HX28	8	0.63o	q12	Roustan et al, 1992
DXS227	–	11	0.85p	q13.1	Fairweather et al,1993
DXS1124	–	8	0.54o/0.54p	q13	Donnelly et al, 1994
DXS566	HX60	5	0.59o/.064p	q13	Porteous et al, 1992
DXS986	AFM116xg1	10	0.76	q13-q21	Weissenbach et al, 1992
DXS995	AFM207zg5	3	0.60	-	Weissenbach et al, 1992
DXYS1X	DXYS1/4-1	3	0.53p	q21.31	Browne et al, 1991
DXS3	p19-2	3	0.64o	q21.3	Stanier et al, 1991

LOCUS	Clone Name	Alleles	Het/PIC†	Location	REFERENCES
DXS990	AFM136yc7	5	0.75	q21-q22	Weissenbach et al, 1992
DXS101	–	15	0.76o	q21.3-22	Allen et al, 1993
DXS458	Mfd79	7	0.62p	q21.1-q23	Weber et al, 1990
DXS454	Mfd72	4	0.59p	q21.1-q23	Weber et al, 1990
DXS178	p212/9	11	0.65o	q21.3-22	Allen et al, 1992
DXS456	XG30B	10	0.77o	q21-q22	Luty et al, 1990
DXS457	XL1E	4	0.20o	q21.1-q22	Luty et al, 1990
DXS1120	GAD	7	0.74o/0.72p	q21-q25	Donnelly et al, 1994
COL4A5	2B6	7	0.76o	q22	Barker et al, 1992
DXS571	HX79B	6	0.46o	q21.1-q21.3	Curtis et al, 1992
DXS424	XL5A	8	0.79o	q24-q26	Luty et al, 1990
DXS1001	AFM248we5	8	0.80	-	Weissenbach et al, 1992
DXS425	XL90A3	10	0.79o	q26-q27.1	Luty et al, 1990
DXS1122	–	6	0.33o/0.38p	-	Donnelly et al, 1994
DXS692	–	3	0.55	q25-q26.2	Lasser et al, 1993
DXS1114	–	3	0.59	q26	Weber et al, 1993
HPRT	HUMHPRTB	8	0.70	q26.1	Edwards et al, 1991
DXS994	AFM205wd2	5	0.33	-	Weissenbach et al, 1992
DXS300	VK29AC	3	0.40p	q26	Gedeon et al, 1991
DXS294	VK17AC	10	0.75p	q26	Gedeon et al, 1991
DXS102	cX38.1AC	8	0.71p	q26.2-q27.1	Gedeon et al, 1992
DXS984	AFM105xc5	9	0.71	q26-q27	Weissenbach et al, 1992
DXS691	–	5	0.65	q26.2-q27	Lasser et al, 1993
DXS292	VK14AC	-	0.58	q27.3	Richards et al, 1991
DXS297	VK23AC	-	0.67	q27.3	Richards et al, 1991
DXS548	–	9	0.49o/0.72p	q27.3	Riggins et al, 1992
DXS1123	–	7	0.67o/0.68p	q28	Donnelly et al, 1994
DXS1113	–	12	0.75	q27.3-q28	Weber et al, 1993
DXS1684	–	7	0.82	q28	Gong et al, 1994
GABRA3	MGD34	4	0.36o/0.29p	q28	Hicks et al, 1991
–	p26	9	0.81	q28	Wehnert et al, 1993
–	p39	8	0.84	q28	Wehnert et al, 1993
DXS707	2-55	personal communication		q28	Maestrini et al, 1992
DXS605	2-19	personal communication		q28	Maestrini et al, 1992
F8C	FVIII-13	8	0.91o/0.69p	q28	Lalloz et al, 1991
DXS1108	sDF-2	7	0.75	q28	Freije et al, 1992
DXYS154	sDF-1	10	0.70	q28	Freije et al, 1992

† Heterozygosity values with a subscript 'o' were observed in the population examined by the authors, while values with the subscript 'p' were calculated from the allele frequencies. In most cases the derivation of the value was not specified.

the markers included in this resource are simple dinucleotide repeats, although some are compound perfect, tri or tetranucleotide repeats (Chapter 1.4.2.) and three are PCR-RFLPs (DXS164, DXS707 and DXS605). With increasing numbers of markers available for gene mapping, their relative positions and order became an important criterion to assess their usefulness. Characterisation of AC repeats of known location was one aim addressed by the candidate (section 3.3.1; Appendix I, Papers 1 and 2). All primer sequences were synthesised and then optimised for $MgCl_2$ concentration to suit established PCR cycling profiles (Chapter 2). Establishment of this resource, together with the RFLP probe bank (section 3.2.1.), has enabled genotyping of families at numerous polymorphic loci spanning the chromosome (Chapters 4, 5 and 6).

3.3.1. Screening, selecting and subcloning AC repeats

A strategy for isolating polymorphic dinucleotide markers of known order and location was devised. The rationale was that the known chromosomal location of DNA sequences used as RFLP probes was more useful than random subclones and that heterozygosity and speed of genotyping at these RFLP loci could be improved. Forty-five RFLP clones representing 113.4kb of the X chromosome were screened with poly ACn (Appendix I, Paper 2). In addition, a pair of anonymous lambda clones, VK17 (DXS294) and VK29 (DXS300), had been mapped against a somatic cell hybrid panel of the X chromosome (Suthers et al, 1990) and were shown to be ACn positive (Richards et al, 1991). These lambda clones were selected for subcloning because of their location in Xq26 (Appendix I, Paper 1) and therefore potential usefulness for mapping and diagnosis of Borjeson-Forssman-Lehmann syndrome (Chapter 5).

Dot blots of 50 to 100ng of insert from each RFLP probe and the two lambda clones were prepared by spotting up to 10ul onto Genescreen Plus (NEN) membranes. These filters were left to air dry for up to 30 mins and then denatured for 3min in 1M NaOH, neutralised for 3min in 1M TrisHCl, rinsed in 2xSSC (0.3M NaCl/ 0.03M Na citrate) and again left to air dry until hybridisation with ^{32}P-labelled poly(AC.GT) (Pharmacia). Hybridization and washing of filters was as described in Chapter 2, except that no salmon sperm was added to the hybridisation solution. Autoradiography required only a few hours exposure. Positive probes were restricted with the frequent-cutting enzymes *Hae*III, *Sau*3AI, *Hpa*II, *Rsa*I or *Hin*PI and run out on a preparative agarose gel. Fragments were transferred to Genescreen Plus membranes in alkali (0.4N NaOH) overnight. Filters were screened by hybridisation with poly(AC.GT) and required less than 2hrs exposure to X-ray film. Positive bands <700bp were selected for subcloning.

Digests of around 500ng insert DNA with the appropriate enzyme were prepared for subcloning and phenol/chloroform purified. M13mp18 or M13mp19 vector was cut with a complementary enzyme eg. *Sau*3AI and *Bam*HI or *Hae*III and *Sma*I and dephosphorylated[1]. Insert digests were ligated into vector following a method modified from the product handbook of Promega, Madison, USA. Ligation reactions contained 50-100ng insert, 100ng vector, 4µl 5X buffer, 0.25U T4 ligase in a total of 20µl incubated at 4°C overnight or >4hours. Ligations (1-10µl) were transformed into 300µl competent[2] *E.Coli* MV1190 cells on ice for 40mins, heat shocked 42°C for 3mins then returned to ice. Cells were added to 8mls prewarmed H top agar with 80µl 100mM isopropylthio-β-D-galactoside (IPTG), 80µl 200mM 5-bromo-4-chloro-3-insolyl-β-D-galactoside (BCIG), 2µl 20% glucose and 400µl fresh cells and plated on H plates. Plates were inverted and grown overnight at 37°C. Plaques were lifted in duplicate following the method modified from the NEN Colony/Plaque Screen Genescreen Plus handbook; Nitrocellulose filters were placed face-down on cold plates for 3min and marked with ink for future orientation, then carefully removed and placed face-up on Whatmann paper (3MM) soaked with 1.5M NaCl, 0.5M NaOH for 5min followed by 3M NaCl, 0.5M TrisHCl pH7.5 for 2x5min before being rinsed in 2xSSC for 5min. Air dried filters were rescreened with the ACn oligonucleotide. Positive plaques were picked into 2ml 2xTY media and grown at 37°C overnight. Single stranded DNA was extracted and prepared for sequencing by a method modified from the Bio-Rad product handbook; Phage were pelleted from the supernatant after 4°C overnight in 2.5M NaCl with 25% polyethylene glycol (PEG 6000), lysed in 100mM Tris pH7.5, 5% SDS, 1mM EDTA incubated at 80°C for 10mins, then precipitated with 1 in 10 v/v 7.5M NH_4OAc in isopropanol.

The markers DXS102, DXS237, DXS294 and DXS300 yielded purified AC positive clones after two rounds of screening. ACn positive clones were manually sequenced according to the method adapted from the USB protocol (Chapter 2.5.5.1.). Subcloning of DXS102 was repeated using *Hae*III after sequencing of the *Sau*3AI clones showed the AC

---

[1]Vector digested with a single restriction enzyme was dephosphorylated to prevent self-ligation by removal of the 5' terminal phosphate group. To 2µg M13 vector DNA digested to completion in 50µl, was added 40µl H_2O, 10µl 10X buffer and 1µl calf intestine alkaline phosphatase (1U/µl - Boehringer Mannheim Biochemica). The reaction was incubated at 37°C for 1hour to dephosphorylate. Phenol/chloroform extraction removed the phosphatase (alternatively incubation at 65°C for 20mins inactivated the phosphatase). Prepared vector ready for direct use in ligation reactions.

[2]MV1190 cells grown in 20mls 2XTY media to OD_{600}=0.2, were made competent by resuspension of freshly pelleted cells in TSB (10% PEG 3000, 5% DMSO, 10mM $MgSO_4$ and 10mM $MgCl_2$ in 2ml Luria Broth) and chilled on ice for up to 1hour.

repeat was too close to the cloning site to allow for primer design (section 3.3.2.). Conversely, the forward primer of DXS300 had to be made first and used to sequence through the repeat as the near *Sau*3AI site was a distance from the repeat motif.

3.3.2. Primer design

The criteria for oligonucleotide primer design followed the suggestions of Luty (1990) with the addition that each primer be 25 bases long for increased specificity. The GC content was no more than 50% with no stretches of four or more consecutive purines. Primers were designed to recognise single copy DNA sequences flanking the repeat blocks and such that the resultant product size should not exceed 300bp. The size of the products was imposed during primer design to enable multiplex reactions without overlap with PCR products from the then existing primer sets. Other criteria considered were random base distribution, avoiding AT-GC rich regions within each primer. Primers were checked for complementarity to avoid the production of primer-dimers by annealing to one another. PCR conditions were optimised for each primer pair and tested for integrity against the parent clone by amplification of the M13 clone as well as the original plasmid.

3.3.3. Genotyping microsatellites

Microsatellites were genotyped in families and individuals by hot PCR incorporating radiolabelled dCTP (Chapter 2.5.4.) with alleles size fractionated by PAGE (Chapter 2.5.5.). Dinucleotides isolated by the candidate were genotyped in genomic samples alongside PCR of a 1/100 dilution of the sequenced M13 product of known size. This served as a known control for direct size comparison and allowed the sizing of genomic genotypes. Heterozygotes with various allele sizes, established against the M13 clone, were used to standardise allele sizing between gels. Family members were near one another on gels, with controls not further than 10 samples away. Products from published primers, often with no controls, were sized consistently within families but 1/n allele frequencies were used in the linkage analysis where correspondence with the published frequencies could not be established (Chapter 2.6.2.1.).

The microsatellites isolated above (DXS102, DXS237, DXS294 and DXS300), and another named XL27B at the DXS538 locus (Browne et al, 1991a), were genotyped in the 40 CEPH reference families to confirm Mendelian inheritance, to determine heterozygosity values and to contribute data for placement of each locus in the genetic map. These genotypes were submitted to the CEPH database and have been included in the production of an X chromosome linkage map (Appendix I, Paper 3). Population allele frequencies were

established from the genotypes of DNA samples from a cohort of unrelated blood bank donors. Comparative allele frequencies estimated from the CEPH parents, reflected the population frequencies. Heterozygosity values were calculated from the observed allele frequencies (Chapter 1.4.1.). This expected heterozygosity did not vary greatly from the observed numbers of heterozygotes.

An advantage of PCR for genotyping was the relatively negligible use of DNA as compared with Southern analysis of RFLPs. PCR based polymorphisms were preferentially selected for genotyping also because of the relative speed of genotyping. Multiplexing involves the typing of more than one locus at a time or co-amplification of several target DNAs at once. Markers were multiplexed only if at least 20bp separated the largest allele of the smaller product from the smallest allele of the larger product. To aid genotyping by multiplex reactions, primers were optimised to conform to one of two sets of amplification profiles, File 15 and File 21 (Chapter 2.5.). PCR conditions were established so several primers could be combined and still produce adequate products of each.

Shadow bands created during PCR amplification may complicate determination of genotypes. The complementary strands of each allele have been shown to have different electrophoretic mobility (Weber and May, 1989). Other bands may be caused by asymmetric 3' base addition by the DNA polymerase, replication slippage or slipped-strand mispairing *in vitro* during PCR (Hauge and Litt, 1993). The pattern of extra bands was found to be constant at any given locus however and across a given gel, and was reproducible such that allele assignment was not affected.

Genotyping errors in the data used to construct linkage maps reduce the power for determining order, inflate the map length and produce incorrect orders largely due to false double recombinants. The potential for errors in genotyping were minimised and controlled so that inconsistencies could be corrected. To guard against bias, the datasheets for each gel listed the DNA numbers only of individuals, so that when genotypes were assigned it was without prior knowledge of the sex or phenotype of the individual. Discrepancies detected on transfer of genotypes to pedigrees were repeated on duplicate samples and at adjacent loci. Non-paternity was usually evident at several loci spanning the chromosome, while true recombinants were consistent at adjacent loci. Transcription and genotyping errors were notably single events that simulated double recombination between flanking markers in a small interval, did not correspond with the family or with the sex of the individual.

Initial gels were filled with obligate carriers and affecteds from more than one family, under the premise that affected members of a pedigree would be more likely to share genotypes at a linked marker locus than would be expected by chance. Additional family members were only analysed when obligate carriers were informative at that given locus and

when affecteds demonstrated some concordance of the inherited allele. When any marker demonstrated an apparent linkage with the disease, further loci either side were used to define the gene limits by recombination or to establish significant linkage. Markers internal to the nearest flanking markers detecting recombination events were subsequently selected for genotyping to reduce initial localisations for fine mapping of the disease locus. This 'nested' approach resulted in refinement of the gene interval where this was possible. The mapping and fine mapping of disease genes is limited by and particularly reliant on the accuracy of the map of markers relative to one another and on the occurrence of recombination within the family.

3.4. Linkage maps of the X

The evolution of the linkage map of the X chromosome has been rapid since the characterisation of microsatellites. Most research groups, however, construct regional linkage maps of high resolution for refinement of disease gene localisation's (eg. Mahtani et al, 1991; Oudet et al, 1992b; Biancalana et al, 1994; Appendix III, Paper 7). Only recently, whole chromosome maps have been produced that include microsatellites.

In late 1992, the NIH/CEPH Collaboration (Murray et al, 1992) described a map composed of 71 markers, the majority of which were RFLP based. The genetic length of 208cM was estimated to contain 95% of the physical length between the markers DXS207 at Xp and DXS52 at Xq. Many of these loci are now represented by dinucleotide repeats. The first CHLC (Cooperative Human Linkage Center) Report in May 1993 outlined the long range goal of the Center to develop genetic maps enriched for high heterozygosity PCR-based markers (with emphasis on tri- and tetranucleotide repeats) to be revised 6 monthly with online access (Murray et al, 1993a). The 210cM framework map of 32 loci was drawn from the available CEPH database resource and presented at XCW4 and was unaltered in 1994 (Murray et al, 1993b; Buetow et al, 1994). Comparative alignment with other maps showed only minor variation between intervals at 14 loci in common, but differences in genetic distances affected overall lengths ranging from 170 to 236cM (Schlessinger et al, 1993). Given that the same dataset was utilised, namely CEPH reference families, order was in agreement as expected.

The current Généthon map includes 80 microsatellite markers over 166cM of the X, with the greatest gap now 18cM (Gyapay et al, 1994). Of the first 25 Généthon markers on the X, 17 were positioned with odds >1000:1 within the Généthon subset (Weissenbach et al, 1992). As only 150 individuals from 8 large CEPH families were genotyped, however, less meiotic events are available for locus placement into existing linkage maps with similar support. It has been suggested that, rather than production of further markers where already

a 20% repetition is in evidence, more extensive genotyping of the full complement of CEPH reference families and/or physical mapping be pursued to properly integrate existing maps (Schmitt and Goodfellow, 1994). The abundance of microsatellite markers means that many alternative markers now exist to saturate the chromosome with good quality mapping reagents once they are reliably placed in the map. Their ubiquitous abundance and STS capabilities are also advantageous for physical and genetic map integration.

3.4.1. Construction of the Background Genetic Map

The linkage map provides a genetic framework for mapping X-linked disorders. In the absence of a complete map of the X chromosome that included all the polymorphic markers accumulated for this study at the WCH, it was necessary to consider construction of a reliable background map that gave sufficient evidence for marker order of the working set. Two parameters are involved in the construction of a linkage map - genetic distance and genetic order. Linkage maps are based on recombination events used to construct a linear order of markers so that the recombination fractions between adjacent loci, considered in pairs, represent the map distance between these markers. The biologic order of loci must satisfy the condition that recombination between loci flanking an interval be greater than or equal to that between adjacent loci. Computer programs to determine the most likely order and recombination fraction between loci have automated the construction of genetic maps but requires specialised skills that were not encompassed by this thesis.

Establishment of correct order in the background map is an important factor as an incorrect order can have effects both on the interpretation of gene mapping data and on the subsequent provision of risks based on that data. Support for order is determined statistically (Lathrop et al, 1984). Framework markers ordered with interval support of at least 3 are the most reliable in this regard, though they are of low resolution. For fine mapping of disease loci it is necessary to use a comprehensive map that includes all syntenic loci albeit with less emphasis on support for order (Keats et al, 1991). Integration with physical evidence can provide a secondary source for validation of the presented order.

3.4.1.1. The framework map

Framework maps require strong statistical support of 1000:1 odds (lod 3) for order of markers with respect to one another. Construction of the framework map was initiated by selecting loci from the CEPH database (versions 5.0 and 6.0) to be included in the analysis. Genotype data on the CEPH reference families from many of the microsatellite loci in Table 3.2 as well as those generated by the candidate (section 3.3.3.) were included. RFLP marker

genotypes existing on the database but now available as microsatellites at the same loci were also included. Two RFLPs were included as landmarks at Xpter (XG) and at the centromere (DXZ1) while the VNTR at DXS52 defined the region known to be near Xqter. The other VNTR at DXS255 was included since it was highly polymorphic and near the centromere. RFLPs without microsatellites at the same locus were not added since the map was intended to order PCR-based loci to be used in gene mapping. The map was constructed by H Kozman (WCH) on the basis of lod 3 criteria ($\log_{10}1000=3$ or log of the odds LOD=3), using the BUILD option of CRIMAP (Kozman et al, 1993). The resultant order represented a 236cM framework map of nearly 30 polymorphic loci (Appendix I, Paper 3) and provided a structure for the development of a comprehensive map (section 3.4.1.2.) and integration with the cytogenetically banded ideogram of the X chromosome (Figure 3.1). Distances between 7 framework markers spanning much of the long arm from DXS453 to HPRT remained in accord with the original map of the region produced in the CEPH pedigrees (Huang et al, 1992). Comparison of this framework map with other framework genetic maps was compatible (section 3.4.).

The genetic ordering of markers can be difficult where some markers have low informativeness, the odds of alternative orders are often less than 1000:1. Mixing RFLP markers with highly polymorphic PCR markers compounds this effect. The establishment of unambiguous order is also limited by the size of the CEPH panel and the informativeness of the markers included in the analysis. Most markers contributed to the CEPH database have been genotyped in at least the first 40 reference families representing 656 potential meioses. The 'Généthon' markers having been genotyped only on a subset of 8 of these families may affect orders in some areas. The WCH microsatellite library (section 3.3.) contains 16 of the first 25 loci published by Généthon (Weissenbach et al, 1992; Généthon Microsatellite Map Catalogue, 1992) only two of these DXS1003 and DXS984 could be placed with 1000:1 odds on the framework map constructed above.

3.4.1.2. The comprehensive map

The loci in a comprehensive map are delimited into intervals by framework loci. Genetic evidence for order within intervals is dependent on marker informativeness, hence local orders can be unreliable. Further loci were added to the framework map with odds for inversion less than 10^3:1 using the ALL option. A comprehensive map of 78 loci was produced (Appendix I, Paper 3) although some orders or distances between markers will be unstable with the addition of further markers. Interlocus distances were given in cM (Kosambi) and were converted back to recombination fractions using MAPFUN (Chapter 2.13.), when necessary for LINKMAP analyses (Chapter 4).

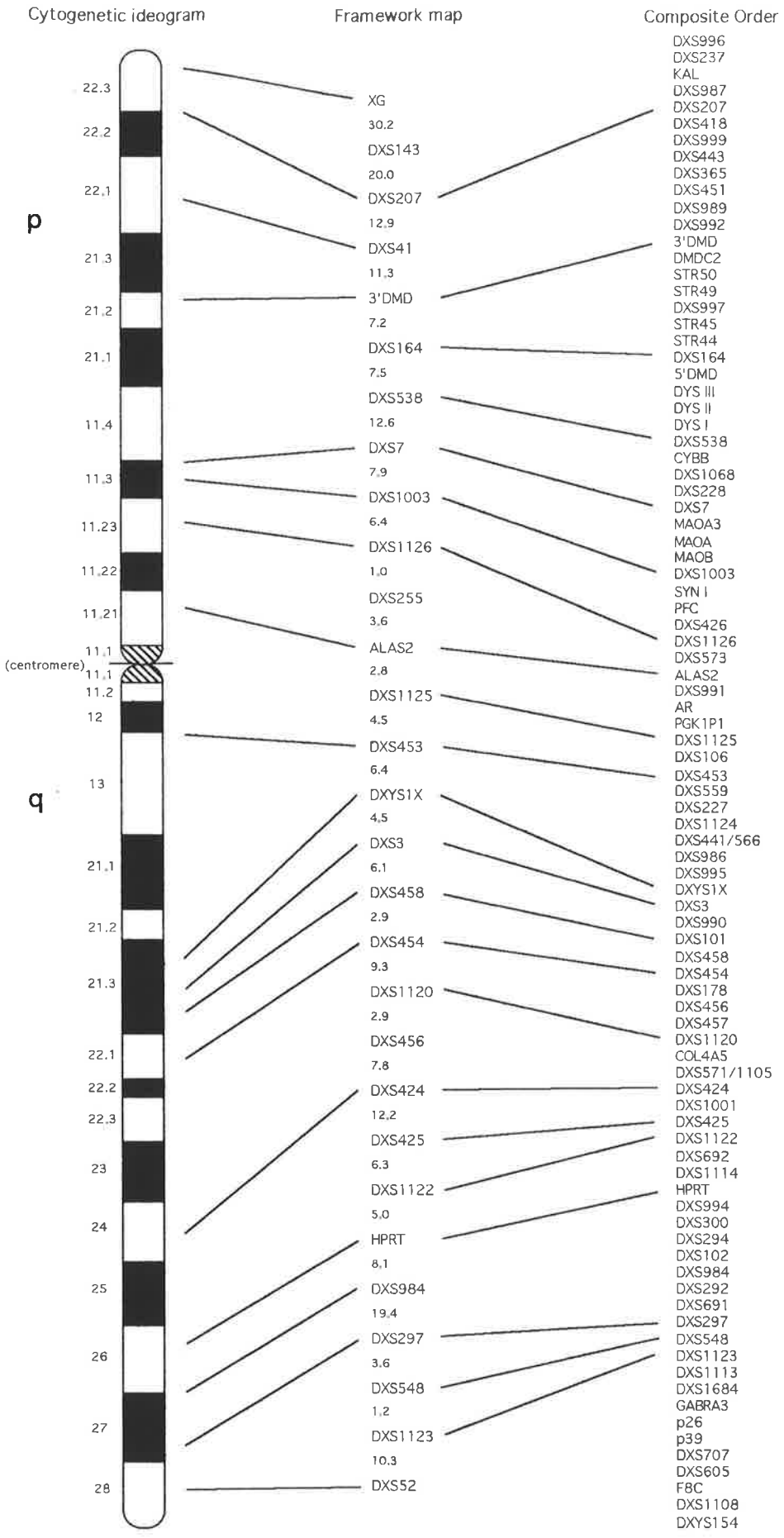


Figure 3.1. Polymorphic markers of the X-chromosome. The framework map of PCR-based markers provides a structure for comprehensive ordering of loci to form a composite map and a means of integration with the cytogenetic ideogram of the X chromosome.

Production of increasingly dense genetic maps will inevitably embrace a nonzero error rate due to genotyping errors (Lincoln and Lander, 1992). These errors introduce spurious crossovers that tend to increase genetic length and decrease support for order. The accuracy of ordering loci in a genetic map depends upon assessing the probability of double recombination events. Error checking for double recombinants was done by CHROMPICS analysis (H Kozman, WCH). Any ambiguous order of markers in the comprehensive genetic map was resolved where possible, by information on order from regional genetic maps and/or when supported by physical mapping data from XCW4 and XCW5 (section 3.2.).

Map distances in distal Xp were later updated by H Kozman as above, by adding KAL and DXS365 from the CEPH ver 6.0 database, to the established comprehensive background map. The KAL locus had been physically mapped in Xp22, between DXS237 and DXS143 (Zhang et al, 1993) and was now genetically at 1.8cM proximal to DXS996 and 19cM distal to DXS207 (Appendix II, Paper 5). Placement of DXS365 proximal to DXS999 and distal to DXS41 has been critical in defining the interval of the PRTS gene (Appendix III, Paper 5).

A separate regional analysis constructed a genetic map across Xq26 for application to diagnosis of BFLS and integration with the physical map of the region (Chapter 5). This high resolution map spanning over 50cM, included 22 PCR based markers from DXS424 to DXS297 (Appendix III, Paper 7). Loci with low odds against inversion as determined from the genetic map were ordered with physical support for order where possible (section 3.4.2.2.). The physical order of markers at the proximal loci DXS424, DXS425 and DXS1001 is; cen-COL4A5-DXS1001-DXS425-DXS424-DXS994-tel, however somatic cell hybrid mapping reverses DXS425 and DXS424 to the order DXS1001-DXS424-DXS425 with DXS424 proximal to DXS425 (Luty et al, 1990; Willard et al, 1994). The odds against inversion determined from the CEPH genotype data are 1:1 between DXS1001-DXS424 and 10^{16}:1 between DXS424-DXS425. A recent two dimensional map of the X chromosome constructed from crossover mapping in CEPH families supports the order DXS424-DXS1001-DXS425 with high confidence (Fain et al, 1995). A recombination event detected in individual IV-9 of the BFLS family (Chapter 5) also supports the placement of DXS424 proximal to DXS1001 and DXS425. Haplotype analysis in the BFLS family further suggests that DXS994 lies distal to the recombination event between HPRT and DXS294 detected in CVS4 (generation IV).

Additional loci used in gene mapping but not on the comprehensive map, were inserted into the background map based on regional publications of locus order. Where possible data derived from CEPH families was preferentially selected as follows:

In the comprehensive map (Appendix I, Paper 3) the dystrophin gene spanned 14.7cM as only single markers were used at either end of the gene (DYSII and 3'DMD) which were informative in less families, thereby inflating the map distance particularly between DXS164-5'DMD. Intragenic recombination frequency of 0.12 has been reported across the dystrophin gene (Abbs et al, 1990). Recombination of 8.7% was determined across the same gene in 39 normal (CEPH) families (Oudet et al, 1992a), nearly 4X the expected frequency (about 1cM/1Mb) on the basis of 2.4Mb gene length. As no recombination was detected between the markers 5'DYS-MSA and DYSII, near the muscle and brain promoters (~125kb), nor STR49 and STR50, nor between DMD1c-2, MP1P and 3'DMD (~500kb) each pair was assumed to constitute single loci (Oudet et al, 1992a). The map distances; 3'DYS-0.6-STR50,STR49-1.6-STR45-2.9-STR44-0.5-DXS164-3.1-5'DMD across the dystrophin locus were used in place of those in the comprehensive map (Appendix I, Paper 3). Hot spots of recombination have been revealed in two regions of the dystrophin gene - one of 4.5% recombination between markers STR50 and STR44 (~440kb) peaking between STR45 and STR44, the second between DXS206 and 5'DMD (Oudet et al, 1992a). To increase information between STR50 to STR44, the DXS997 locus in intron 48 (Saad et al, 1993) was inserted into the background map derived for gene mapping in this thesis with recombination fractions arbitrarily allotted STR50/49-0.6-DXS997-1.0-STR45-2.9-STR44 (NB. STR markers in the dystrophin gene have recently been given DXS numbers at the XCW5 as listed in Table 3.2, but have been used in the earlier format for ease of identification in this thesis).

The CYBB gene for chronic granulomatous disease physically mapped proximal to 5'DMD and distal to DXS1068 (Willard et al, 1994) is an estimated 14cM distal to DXS7 (Thiselton et al, 1995) placing it in the same genetic interval as DXS538 (Appendix I, Paper 3). For the purposes of linkage analysis in the ensuing chapters DXS538 and CYBB were regarded as markers with no recombination between them.

The lack of recombination between PFC and DXS426 in CEPH families suggested their use as a haplotype for linkage analyses (Coleman et al, 1991). A recombinant in MRX4, between PFC and DXS426, confirms the consensus order at XCW5 placing PFC distal to DXS426 (Hu et al, 1994; Willard et al, 1994). Similarly the adjacent group DXS228, DXS7 and MAOA had no recombinants with odds of 1:1 (Appendix I, Paper 3). The consensus map of XCW4 ordered DXS228 two physical intervals away from MAOA and DXS7, and later, YAC contigs at XCW5 centered around DXS7 and DXS228 were linked and showed DXS228 in a distinct interval distal to DXS7, MAOA and MAOB (Black et al, 1994). Data from the CEPH database (ver 3) suggested DXS7-9-SYN/ARAF-4.2-TIMP-2.8-DXS255 (Kirchgessner et al, 1991) so that the map used was DXS228, DXS7,

MAOA/B-7.9-DXS1003-1.0-SYN-4.2-TIMP, PFC, DXS426-1.2-DXS1126, SYP-1.0-DXS255 based on distances from the comprehensive linkage map and physical evidence for ordering the SYN, TIMP and PFC genes (section 3.4.2.).

The dinucleotide repeat markers at DXS106, DXS227, DXS571, DXS559 and DXS566 (the latter detecting the same repeat as DXS441) were not included in the comprehensive map since genotypes were not available on the CEPH database (ver 5). Genetic map distances were therefore not available, however recombinant meiotic events have placed DXS559 proximal to DXS566 (Lindsay et al, 1993). DXS106 and DXS227 were isolated from a YAC library and therefore have known physical locations (Fairweather et al, 1993). Recombination events between the DXS106 and DXS1125 loci in MRX1 and MRX17 (Chapter 4) suggest the order cen-DXS1125-DXS106-tel for this pair of markers. Placement of the rest of these markers in intervals between hybrids, order loci near to the centromere on the long arm with map distances in the composite map (section 3.5.) estimated from distances between markers existing in the comprehensive map.

3.4.2. Physical Maps of the X

Physical mapping of polymorphic markers provides evidence to support marker order and direct access to DNA fragments facilitating positional cloning of genetically linked genes. Physical means of map construction include FISH (Wu et al, 1993), PFGE, somatic and radiation-induced hybrids (Falk, 1991) for ordering markers. Hybrid breakpoints can be visualised cytogenetically and are a means for integration of the cytogenetic and physical maps. YACs are particularly useful for estimation of interval distances as well as ordering of loci on contigs. Two primary properties of gene maps can be compared; the order of loci and the relative distances between them. In linkage maps distance reflects recombination frequency and is location dependent, while physical distance is a direct measurement of base pairs that do not necessarily correlate with recombination frequency (Chapter 1.3.). Map construction by physical and linkage means should be collinear with the most important criterion being unambiguous order.

The XCW consensus map includes alignment of contigs and ordering of breakpoints verified by a number of sources, and places genes and microsatellite markers in order from telomere to telomere on the physical map. The XCW4 listed 118 X-linked multiallelic microsatellites, but placed only 82 on the physical map with 64 ordered on YAC and cosmid contigs and 18 between breakpoints (Schlessinger et al, 1993). Many more markers had been added to XCW5 and ordered on more extensive YACs (Willard et al, 1994) toward a more detailed map integration.

Division of the X chromosome into equal 8-10Mb intervals (~20 hybrids) was proposed at XCW2 as a valuable reference resource for physical mapping (Davies and Craig, 1991). A panel of 18 X-chromosome hybrids have been used to assign 60 markers to the pericentromeric region (p21-q21) of the chromosome including gene order for OTC, MAOA, SYN1, TIMP, PFC, SYP, AR and PGK1P1 (Lafreniere et al, 1991). A further 26 X-chromosome rearrangements overlap in q13-q21 and extend the mapping panel to q22 (Philippe et al, 1993). Deletion panels are also effective for specific high resolution ordering of loci (Ballabio and Andria, 1992).

YAC and cosmid contigs have been constructed to span disease gene candidate regions and to link gaps between genes (Lee et al, 1993; Francis et al, 1994). Long-range region-specific physical maps across 50Mb in Xq24-28 and of 35Mb in Xp22.3-Xp21.3 approach the goal of a complete integrated map of the X chromosome (Schlessinger et al, 1991; Ferrero et al, 1995). At HGM11 in 1991, more than 25% of the X chromosome was contained in YAC contigs. In early 1993, participants at XCW4 from 12 countries contributed data leading to over 50% YAC coverage (Schlessinger et al, 1993) and by 1994 at XCW5, were being directed towards map closure of 85% of the chromosome (Willard et al, 1994). Such contigs may be utilised for ordering or confirming order of polymorphic loci, transcribed sequences and genes. Alternatively, polymorphic sequence tagged sites (STS) ordered by meiotic recombination have been used to construct a physical map of the genome that is integrated (Cohen et al, 1993).

3.4.2.1. YACs

Collaboration was established with Dr David Schlessinger (Washington University School of Medicine, USA) for mapping newly characterised dinucleotide repeats (section 3.3.1.) against a YAC contig around BFLS (Schlessinger et al, 1991). Marker loci were physically ordered within the BFLS interval on an 8Mb YAC contig of Xq26 (Little et al, 1989); DXS1122-DXS79-HPRT-DXS86-DXS144-**DXS300-DXS294**-DXS51-**DXS102**-F9-DXS105. This provided 3 new sequence tagged sites (STS) for linking clones, a means of ordering markers with physical support for order and the estimation of map distance (in Mb) for comparison to the estimated genetic length (which seems to hold to the 1Mb=1cM rule in this region of the chromosome) (Appendix III, Paper 7).

3.4.2.2 Physical Evidence for Marker Order

The physical order of markers within the WCH libraries, both RFLP probes and microsatellite primer pairs, was extracted from the XCW4 and XCW5 consensus maps.

These markers have been linked to the cytogenetic map (Figure 3.1) with many mapped to breakpoints or deletion intervals and can thereby anchor disease gene localisations to specific regions of the chromosome. The following listing gives markers of definite order (support for order by two or more lines of evidence) typed in bold; microsatellites are underlined and RFLP probe names are specified alongside the DXS number. Brackets indicate the interval in which markers of unknown order lie, markers in bold within brackets are of known order in that interval, extra markers (not on the XCW4 and XCW5 consensus maps) are inserted in *italics*;

pter - **DXS996, DXS237**GMGX9, **KAL, DXS143**dic56, **DXS85**p782, **DXS16**pSE3.2L, **DXS9**RC8, **DXS987, DXS207, DXS43**pD2, **DXS418**, (**DXS999**, DXS443, **DXS365, DXS41**p99.6, DXS451, **DXS989**), **DXS67**B24, **DXS28**C7, (**3'DYS, DMD1c-2**, DXS992, **STR50, STR49, DXS997, STR45, STR44, DXS269**P20, **DXS164**pERT87-15, **DXS206**XJ-2.3, **5'DYS, DYSIII, DYSII, DYSI, DXS84**p754, (*DXS538*, **CYBB**), **OTC, DXS1068, DXS228, DXS7**L1.28, **MAOA3, MAOA, MAOB, DXS1003, SYNI, TIMPI, PFC, DXS426**, (**DXS1126**, *DXS573*, **DXS255**M27B), **DXS146**pTAK8A, **ALAS2, DXS991, DXS14**58.1, **cen, DXS1**p8, **AR, PGK1P1**, (**DXS106**cpx203, **DXS453/983**, DXS559, *DXS1125*, **DXS227**, *DXS1124*, **DXS441/566, PGK1, DXS986**), **DXS72**pX65H7, **DXS995, DXYS1X**pDP34, **DXS3, DXS990**, (**DXS458, DXS454**, DXS456, **DXS178**212/9, **DXS101**cX52.5, **DXS94**XG-12, **DXS17**S21, DXS1120, **DXS87**A13, **COL4A5**), **DXS571/1105**, (**DXS1001, DXS425, DXS424, DXS994, DXS42**p43-15, *DXS1122*), **DXS692, DXS1114, HPRT, DXS86**St1, **DXS106**A1, **DXS300, DXS294, DXS51**52A, **DXS102**cx38.1, **F9, DXS984, DXS105**cx55.7, **DXS98**4D-8, **DXS292**VK14, (**DXS369**RN1, DXS691, **DXS297**VK23, **DXS548**), **FRAXA, DXS296**VK21A+C, **DXS1123, IDS, DXS1113, DXS304**U6.2, **DXS1684, DXS374**1A1, **GABRA3, DXS52**ST-14, *p26* and *p39*, **DXS15**DX13, **DXS707**2-55, **DXS605**2-19, **F8C, DXS1108, DXYS154**, - qter.

The following supporting information confirmed placement of microsatellites on the physical map (from original references where available):

In Xp22, KAL maps to Xp22.3 at 8700kb from the Xp telomere (Bouloux et al, 1991) and DXS207 was cloned from a 35kb cosmid selected with the pBA4B probe at DXS207, which shows very tight linkage to DXS43 such that no recombinants have been reported (Oudet et al, 1992b). New microsatellites have recently been described at the DXS85 and DXS16 RFLP loci (Chang et al, 1994). DXS418, derived from a cosmid screened with probe P122 was localised to Xp22.1 between breakpoints in a deletion patient (Van de Vosse et al, 1993). Three clones DXS443, DXS451 and DXS365 were placed at Xp21.2-22.3 on a hybrid cell panel (Browne et al, 1992b).

The loci defining the dystrophin locus are 3'DYS (or 3'DMD) at the extreme 3'end of the gene, within the 3' untranslated region (Oudet et al, 1990), then DMD1c-2 selected from a cosmid containing *Hind*III fragments 46 and 47 (Powell et al, 1991), which correspond to the region between exons 56 and 60 (Oudet et al, 1992a), STR50 and STR49 are located in introns 50 and 49 respectively (Clemens et al, 1991). Amplification of DXS997 (Weissenbach et al, 1992) in known deletions of the dystrophin gene suggest the marker is localised in intron 48 (Saad et al, 1993), while STR45 and STR44 are located in introns 45 and 44 respectively (Clemens et al, 1991), with 440kb between introns 44 and 50, intron 44 is about 170kb (Oudet et al, 1992a). The DXS164 RFLP converted to PCR is about 850kb from the 5' end of dystrophin (Oudet et al, 1992a) with over 70% women expected to be heterozygous for at least one (Roberts et al, 1989). 5'DYS MSA (5'DMD), is localised in the first intron downstream to the muscle-specific first exon within 25kb of the muscle-specific promoter (Oudet et al, 1991) and is about 125kb distal to 5'DYSII. DYSII was one of 4 ACn repeats in 19kb isolated from genomic sequences surrounding the brain promoter, about 100kb upstream of the muscle promoter, and therefore at the extreme 5' terminus of the gene. DYSIV was non-polymorphic, but the remaining three together are informative in 90% of women. DYSI and DYSII are most informative and located just upstream (5') of the transcriptional start site for the brain promoter - 3.5kb and 1.2kb respectively, whereas DYSIII is within the first intron 3.5kb 3' to the transcriptional start site (Feener et al, 1991), therefore the order is tel-III-II-I-cen.

Proximal to the dystrophin gene, DXS538 mapped between DXS164 and DXS7 (Browne et al, 1991a) although the order of DXS538 with respect to DYSII and DXS84 is not resolved on either the physical or comprehensive genetic maps. DYSII is distal to DXS84 on the physical map, but DXS538 is at 0.0cM with 1:1 odds of inversion with either marker in this linkage cluster, so order may be tel-DYSII-DXS84-DXS538 or DYSII-DXS538-DXS84. The dinucleotide repeat sequence CYBB, isolated from a 360kb YAC mapping to Xp21.1, maps to intron 5 of the chronic granulomatous disease gene (Hardwick et al, 1993) and is placed 450kb distal to CYBB between DXS84 and OTC on the consensus XCW5 map (Willard et al, 1994). A compound perfect repeat (CAC_5CA_{18}) at DXS7 was subcloned from an L1.28 positive cosmid (Moore et al, 1992) adjacent to MAOA3, a VNTR contained within a genomic fragment encompassing exon 1 of MAOA (Hinds et al, 1992). The MAOA (monoamine oxidase-A) $(AC)_{18}CG(AC)_3$ repeat in Xp11.3, (Black et al, 1991), is most likely in intron 2 of the gene (Grimsby et al, 1992), while the MAOB $(TG)_4CCTC(TG)_{23}AGAC(AG)_4$ dinucleotide is located 260bp downstream from the exon 2 donor site (Grimsby et al, 1992). The same repeat was reported by Konradi et al, 1992 with different primers, within intron 2 at the 5'end of the MAOB gene. Fine physical mapping of

Norrie disease (NDP) orientated these genes; DXS7-5'MAOA3'-40kb-3'MAOB5'-NDP-cen, each spanning 90kb (Sims et al, 1992). A compound ACn is located about 1000bp from the 3'end polyadenylation signal of the SYN1 gene, and in the last intron of Araf1 gene, these genes lie 3' to 3' and are transcribed from opposite strands (Kirchgessner et al, 1991). Properdin P factor (PFC), is a component of the alternative pathway of complement; the ACn within 16kb from the 3'end of properdin gene lies less than 100kb distal to TIMP. From YAC data DXS426, PFC and TIMP are all on a 440kb YAC, with TIMP and PFC on a 100kb fragment, while PFC and DXS426 are both on a 390kb YAC with suggested order tel-(DXS426, PFC)-TIMP-cen (Coleman et al, 1991). The consensus order at XCW5 was tel-1003-SYN-TIMP-PFC-426-cen (Willard et al, 1994). One of the earliest microsatellite loci DXS426 was derived from subclone XL91B10 in Xp11-p21 (Luty et al, 1990), and is likely to be linked to RP2 in Xp11.4-p11.23, but not RP3 in Xp21 refining localisation between DXS7 and DXS255 (Coleman et al, 1990). DXS573 was assigned to Xp11.1-p11.22 on a cell panel (Roustan et al, 1993), and mapped proximal to DXS426, distal to DXS255 by both genetic and physical means (Thiselton et al, 1995). Recombinants detected in the XS family (Chapter 6) place DXS573 proximal to DXS1126.

Across the centromere additional chromosomal breakpoints localised markers to defined intervals from Xp11-q22 (Barker and Fain, 1993). A compound dinucleotide repeat in intron 7 of the ALAS2 gene showed no recombination with DXS14 or DXZ1 at the centromere (Cox et al, 1992). The physical distance of DXS991 to AR or DXS106 is greater than 10Mb, yet the genetic distance of only a few cM reflects the reduced recombination at the centromere (Willard et al, 1994).

On the long arm of the chromosome, the CAG_n repeat in the coding region of the first exon of the AR gene corresponds to a long tract of glutamine residues. Variable alleles n=17-26 in normals, with n=40-52 result in abnormal instability associated with SBMA phenotype (La Spada et al, 1991; Biancalana et al, 1992). The clone named PGK/5 was initially reported at the more distal PGK1 locus (Browne et al, 1991c), though later physical and genetic evidence placed the microsatellite at Xq12 near the pseudogene PGK1P1 (Browne et al, 1992a). A CA_{19} dinucleotide at the DXS106 locus was identified from cosmid clones at Xq12 (Fairweather et al, 1993). The primers for DXS453-Mfd66 (Weber et al, 1990), and DXS983 (Généthon Microsatellite Map Catalogue), recognise the same locus (Rider and Monaco, 1993). A CA_{24} at DXS559 was assigned to Xq12 on a somatic cell hybrid panel and confirmed by FISH (Roustan et al, 1992). A CA_{18} dinucleotide was isolated from cosmids at the known locus for DXS227 in Xq13.1 (Fairweather et al, 1993). Analysis of the sequences flanking the DXS566 microsatellite (Porteous et al, 1992), identified the primer sequences amplifying DXS441 (Ram et al, 1992), therefore they detect

the same locus (Barker and Fain 1993; Lindsay et al, 1993). Sublocalisation of the derivative cosmids by in situ demonstrated that DXS566 at Xq13, is distal to DXS559, recombinant meiotic events order the loci as AR, (PGK1P1, DXS559), (DXS566/441, PGK1) (Lindsay et al, 1993), and recombination fractions of DXS255-0.16-PGK1P1-0.1-DXS566 have been assumed although variation had little effect on the lod score with Aarskog syndrome (Porteous et al, 1992). DXS441 and PGK1 are in a YAC contig of 2.1Mb that contains up to 10 genes (Willard et al, 1994). The microsatellite at DXYS1X at Xq21.31, can increase heterozygosity when haplotyped with the RFLP pDP34/TaqI at the same locus (Browne et al, 1991b). At DXS3 the dinucleotide repeat was in a cosmid positive for the probe p19-2 localised to Xq21.3 (Stanier et al, 1991). Both DXS458-Mfd72 and DXS454-Mfd79 were assigned to Xq21.1-q23 on hybrid panels (Weber et al, 1990), with DXS458 more proximal, which corresponds with genetic order (Huang et al, 1992; Parkar et al, 1994). DXS456 (Fain et al, 1991) was localised on hybrids to Xq21-22 (Luty et al, 1990) and mapped between breakpoints as distal to DXS458 and proximal to DXS571 (Willard et al, 1994). A CA_{18} repeat which maps to Xq21.3-q22 was selected from a 240kb YAC with the probe p212 at DXS178 (Allen and Belmont, 1992), while at DXS101 a $CTT_{13}ATT_{10}$ trinucleotide repeat element from a 350kb YAC clone maps to Xq13-24 on an X hybrid panel (Allen and Belmont, 1993). The DXS101 RFLP marker has been mapped to Xq21.3-22 (Willard et al, 1985). The clone 2B6 in Xq22 contained a dinucleotide in the COL4A5 gene (Barker et al, 1992). The DXS571/1105 locus maps to Xq21.1-21.3 on hybrids and by FISH (Curtis et al, 1992), but fine-mapping by PCR localisation against chromosomal breakpoints suggests it lies distal to Xq22.2, ie distal to DXS3 and DXS990 (Barker and Fain, 1993). XCW5 corroborates placement of this locus distal to COL4A5 at Xq22.3, on the physical map (Willard et al, 1994).

At loci from Xq24-qter there is a discrepancy between the physical (XCW5) and genetic maps for order of DXS1001, DXS424 and DXS425. DXS1001 lies between DXS424 and DXS425 on genetic map with good support (Fain et al, 1995). DXS425 (XL90A3) is localised to Xq26-q27.1 and DXS424 (XL5A) to Xq24-26 on a hybrid panel (Luty et al, 1990). DXS692 maps to Xq25-q26.2 between DXS100 and DXS144E (the latter physically mapped proximal to DXS300) or between DXS425 and DXS51 in broad terms. The order with respect to HPRT and therefore to DXS1114 was not determined (Lasser et al, 1993) until physical data at XCW5 mapped DXS692 between DXS1047 and DXS79 (Willard et al, 1994), the same interval as DXS1122 on the genetic map (Chapter 5). The $AC_{14}AG_{14}$ repeat at DXS1114 is about 350kb proximal to the AGATn tetranucleotide repeat at HPRT (Weber et al, 1993a; Edwards et al, 1991). Both DXS300 and DXS294 map to Xq26 on a hybrid panel Appendix I, Paper 1). The RFLP at DXS102 has been placed into

the same physical interval as F9 (Suthers et al, 1990), and ordered against other loci in the region on a large YAC contig (Little et al, 1989). DXS691 physically within the interval Xq26.2-q27, has been genetically confirmed as distal to F9 (Lasser et al, 1993). The physical order of RFLPs around FRAXA have been established on cell lines (Suthers et al, 1990) and 3cM defined between DXS292 and DXS297 (Richards et al, 1991). DXS548 is approximately 150kb proximal to the fragile site (Riggins et al, 1992). At DXS1113 a $TC_{10}AC_{21}$ repeat lies about 250kb distal to IDS and 200kb proximal to DXS29 in Xq27.3-q28 (Weber et al, 1993b). The DXS1684 source cosmid was assigned between DXS304 and DXS305/374 by hybridisation to a physical map of cosmids (Gong et al, 1994). Distally lies an AC near an exon of the neuronally-expressed human $GABA_A$ receptor α3-subunit gene GABRA3 (Hicks et al, 1991). YAC XY845 represents 500kb containing DXS52, DXS134 and DXS15, with p26 at 220kb and p39 at 10kb proximal to DXS15 respectively (Wehnert et al, 1993). DXS707(2-55) and DXS605(2-19) are *Msp*I and *Eco*RI RFLPs respectively that map in the G6PD-RCP/GCP gene region, 2-19 is about 16kb from the 3'end of G6PD (Maestrini et al, 1992). Primer sequences were supplied by personal communication with Dr Daniela Toniolo (Pavia, Italy). The gene for coagulation factor VIII, causes haemophilia A when defective, but has a CAn in intron 13 (Lalloz et al, 1991). The sDF-2 source cosmid for $(GT)_5GC(GT)_{13}$ at DXS1108, contains an X-specific sequence that has sequence homology on the Y chromosome, therefore it is likely to contain the physical boundary between the X-specific and XY homology regions; cosmids from two overlapping YACs carrying the DXYS64 locus (which maps 350kb from the terminal repeats at the telomere and identifies the XY homology region in Xq28 of less than 500kb) were screened for AC and represent the most distal segment in Xq28 (Freije et al, 1992). In sDF-1 at DXYS154, the most distal dinucleotide $(CA)_{18}$ marker lies within the XY homology region and demonstrates a high rate of recombination in male meiosis 2% (Freije et al, 1992).

3.5. The composite or integrated normal background map of the X

One of the difficulties in mapping and diagnosis of disease genes is that within individual families, markers are often uninformative by chance. When a marker critical to the definition of a gene is not informative, knowledge of marker order facilitates selection of the next available adjacent marker. By integrating several sources of evidence it has been possible to construct a background map ordering markers in the WCH polymorphism library, to be used for gene mapping and linkage analysis in the family studies presented in this thesis. Integration relied upon evidence from published physical order and subsets of the genetic map in specific regions preferentially using maps derived from CEPH families

where possible. The consensus orders given at XCW4, then XCW5 (Schlessinger et al, 1993; Willard et al, 1994), have been used to order additional markers on the comprehensive background map spanning over 236cM derived from CEPH families. The weakness of XCW maps is that evidence for locus locations are usually not given and placement of unfamiliar loci may be unsupported by more than one contributor leading to a potentially high error rate (Wang et al, 1994).

For this consolidated map, any pair of loci were considered ordered relative to one another if i) a physical breakpoint or other physical evidence separated them into distinct intervals or orders, or if ii) recombination events had been detected between the loci and a recombination fraction given on the comprehensive or other available regional CEPH maps. Marker pairs with very few or no recombinations detected between them could sometimes be ordered on the physical evidence, but linkage distances were left near zero and markers were either haplotyped or the most informative one was selected to represent that locus for multipoint analysis (the LINKMAP program does not allow 0.00 recombination between markers). Conversely, markers with physical data available but not included in the comprehensive genetic map, were ordered physically and arbitrary genetic distances were estimated from adjacent markers flanking the interval on the comprehensive map where the likelihood support for order was greater than 10:1. Where the odds for inversion between adjacent loci were approaching or equal to 1:1, the order and distance between markers was arbitrarily estimated given the order on the physical map. Although it is not possible to resolve definitely the order of all markers because either no recombinations were observed (between the loci in each pair) or because the odds for inversion were negligible, it is reasonable to assume that those loci are not identical but are separated by small but unknown distances. Based on this assumption, recombination fractions of 0.001 were inserted between these markers in keeping with the physical order given in XCW5. The final map incorporates both physical mapping data and genetic data from several sources as detailed above, however confidence in the order is based largely on the consensus physical supporting data.

The dinucleotide at DXS538 though not on the physical map, is located within a linkage cluster including DYSII and DXS84 but is telomeric to DXS1068, while CYBB is not on the genetic map (3.4.1.2.) but placed between DXS84 and OTC, DXS1068 on the consensus map (3.4.2.3.). The order of DXS538 and CYBB have not been resolved with respect to one another but for mapping purposes they were placed between 5'DMD and DXS1068 with map distances 0.7 and 1.6cM taken from the comprehensive map. Similarly DXS1125 was mapped between PGK1P1 and DXS453/983 on the comprehensive map, while DXS106 lies physically between these same loci on the consensus map. For the

purposes of linkage analyses DXS1125 and DXS106 came to be considered indistinguishably close. Recombinant events in MRX1 and MRX17 (Chapter 4) however, are consistent with DXS1125 being proximal to DXS106. Fine-mapping against hybrids has ordered the following pericentromeric loci with respect to one another on the long arm: (AR, PGK1P1, DXS106), (DXS453/983, DXS559), (DXS227), (DXS441/566, PGK1), (DXS986), (DXS995), (DXYS1), (DXS3, DXS990) (Barker and Fain, 1993) and the DXS571/DXS1105 locus lies distal to the COL4A5 locus (Willard et al, 1994). For the ensuing mapping studies the order cen-AR-0.0-PGK1P1-2.4-DXS1125/DXS106-4.5-DXS453/DXS559-1.0-DXS227-1.0-DXS566/441-1.0-DXS986-3.5-DXYS1X-4.5-DXS3-tel was utilised. Map distances were taken from the comprehensive map between PGK1P1-DXS1125-DXS453-DXS986-DXYS1, except that the distance between DXS453 and DXS986 was arbitrarily divided given that the two intervening loci had been mapped to separate physical intervals.

From accumulated physical data the Xq22 order appears to be [DXS3,DXS990], [DXS178,DXS101], [COL4A5,DXS1105], [DXS1001,DXS424], DXS425. New markers DXS1153 and DXS1120 mapped closely around DXS1105 (and COL4A5) with low odds against alternative orders, while DXS456 in a broad physical consensus interval is a genetic framework marker between DXS1120 and DXS424 (Appendix I, Paper 3). Genetic data inserts DXS458, DXS454 proximal to COL4A5 (Huang et al, 1992) and to DXS178 and DXS101 (Parkar et al, 1994). Since DXS178 and DXS101 had not been included in the comprehensive map, and DXS458 and DXS454 had not been on the XCW4 consensus map, an interim order was used as a working map to select markers for gene mapping; DXS3, DXS990, [458=178, 454=101], [COL4A5, 1105, 1153, and 1120 all roughly the same locus], DXS456, DXS424, DXS1001, DXS425.

It was assumed that data derived using the same sample of families, the CEPH reference families, would be additive - therefore order established in these families should not change where there is high support, but the distances may depend on the number of loci included in the analysis. Use of different algorithms for map construction have altered distances, though the total map distance may be comparable (Matise et al, 1994; Wang et al, 1994). Length estimates may be inflated when gaps between markers are large (Fain et al, 1995). The locus order derived here (Figure 3.1) did not present evidence in conflict with an extensive integrated map or with a high density 2D crossover map using different criteria for order (Wang et al, 1994; Fain et al, 1995). Given that the orders correspond, only distance will be varied which has less effect on mapping.

Integration of maps is a high priority for gene mapping and diagnosis. When the available physical and genetic evidence from normal families and published in the literature

has been considered and an apparent discrepancy occurs in a disease family, examination of multifactor crosses following Drayna and White, 1985 (in exclusion of the phenotype) may help determine order. The method defines order based on the least number of crossovers within a tight linkage map, for example if AXB and AXC but B-C, then the order is A-B-C since double crossovers are least likely to occur, also BXC and AXC but A-B reveals the same order.

Given the exponential increase in the availability, quantity and quality of genetic markers over the last few years (Weissenbach et al, 1993; Gyapay et al, 1994) this map cannot hope to represent more than a fleeting moment in the development of the full cognisance of the X. It has been used in the ensuing chapters as the basis for the mapping studies and to determine approximate gene intervals so that genetic counselling might be offered. It is not anticipated that any changes to the map brought about by full integration will affect the conclusions in the following disease gene localisations.

CHAPTER 4

Non-Specific X-Linked Mental Retardation (MRX)

Note to Publications

Much of the work summarised in this chapter has been published, with many of the clinical assessments carried out by Drs. Bronwyn Kerr and Gillian Turner (POWCH), while all molecular studies were carried out and prepared for publication by the candidate. Reprints are bound in Appendix II in the following order:

1. **Gedeon A**, Kerr B, Turner G, Mulley J, (1991) Localisation of the MRX3 gene for non-specific X linked mental retardation. J Med Genet 28:372-377.

2. Kerr B, Turner G, Mulley J, **Gedeon A**, Partington M, (1991) Non-specific X linked mental retardation. J Med Genet 28:378-382.

3. Kerr B, **Gedeon A**, Mulley J, Turner G (1992) Localisation of non-specific X-linked mental retardation genes. Am J Med Genet 43:392-401.

4. **Gedeon A**, Kerr B, Mulley J, Turner G, (1994) Pericentromeric genes for non-specific X-linked mental retardation (MRX). Am J Med Genet 51:553-564.

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

6. **Gedeon AK**, Glass IA, Connor JM, Mulley JC (1996) Genetic localisation of MRX27 to Xq24-26 defines another discrete gene for non-specific X-linked mental retardation. Am J Med Genet (in press).

7. **Gedeon AK**, Donnelly A, Kerr B, Turner G, Mulley JC (1996) How many X-linked genes for non-specific mental retardation (MRX) are there? Am J Med Genet letter to the editor (in press).

It is recommended that these reprints be read prior to reading this chapter since there is frequent reference from the chapter to the publications and many figures of pedigrees, lod tables and clinical photographs are not reproduced herein.

4.1. Introduction

Mental retardation has a wide range of etiologies involving both environmental and genetic factors. Of those arising from known prenatal factors, Down syndrome is the most common congenital form and fragile X syndrome the most common hereditary form. Formal testing of intellectual function under standardised conditions yields a normal distribution (bell-shaped curve) of intelligence quotient (IQ) with a mean of 100 in the population. Approximately 3% of the population have an IQ less than 2 standard deviations below the mean and are considered mentally retarded. The proportionate excess of retardation in males, versus that in females, has at least in part been accounted for by the theory of X-linked mental retardation proposed by Lehrke (1974) (Chapter 1.6.). X-linked mental retardations (XLMR) represent up to one quarter of all mental retardation in humans (Opitz, 1986) and the majority are subdivided into two main categories, syndromal (MRXS) and non-syndromal or non-specific (MRX) on the basis of dysmorphic clinical findings or lack thereof (Chapter 1.6.).

Mental retardation is classified as mild (50-70), moderate (35-49), severe (20-34) or profound (less than 20) based on IQ ranges established by the World Health Organisation (WHO) in 1980 (Glass, 1991a; Sutherland and Hecht, 1985). Both the classification terms and the IQ ranges for X-linked mental retardation conditions have been only loosely applied by authors in the past. Males with a full-scale IQ score of 85 have been classified with borderline mental retardation (Brunner et al, 1993), at the limit of the WHO range for this category which is IQ 70-85. Some earlier studies describe affected males merely as retarded without specifying IQ, or as feeble-minded, idiot, imbecile and moron (Lehrke, 1974) or in grades as dull (70-84), simpleton (50-69), imbecile (20-49) and idiot (<20) (Morton et al, 1977). Others studied individuals in the broad range IQ 30-55 (Turner and Turner, 1974), later propagated by other authors as defining retardation in the moderate group. In some cases psychometric data were estimated from medical history and interview, and classified as educable (IQ ~50-70), trainable (IQ ~30-50) or severely multiply handicapped (IQ <30) (Proops et al, 1983).

The major clinical feature of MRX is the non-progressive idiopathic retardation itself, without other dysmorphic, metabolic or neurological signs. The lack of specific characteristics of phenotype makes this a clinically homogenous disorder that is not clinically classifiable. Nevertheless until 1988, when the first MRX genes were localised by linkage (Arveiler et al, 1988; Suthers et al, 1988), only clinical descriptions of families were being reported, many with the marker X chromosome seen in fragile X syndrome (Herbst, 1980; Proops et al, 1983). Later, the fragile X syndrome was recognised as a syndromal

disorder delineated by characteristic cytogenetic and clinical features. This most common XLMR is caused by mutation in a single major X-linked gene associated with mental retardation in humans, the FMR1 gene (Chapter 1.6.1.). Genetic mapping of two MRX loci distinct from the fragile site and from one another (Arveiler et al, 1988; Suthers et al, 1988) provided the first conclusive evidence for genetic heterogeneity of MRX. Theoretical estimates of the number of X-linked loci causing mental retardation have ranged from 7-19 (Herbst and Miller, 1980; Morton, 1977) to an upper limit of 25 (Morton, 1992). Given the apparent clinical homogeneity of MRX, further classification of this condition became dependant on determination of the extent of genetic heterogeneity and on defining any clinical correlates.

The prevalence of MRX is difficult to estimate since individual cases cannot be recognised without evidence of an X-linked pedigree. Inheritance of the gene is mostly X-linked recessive and thus always expressed in males that carry the gene. Except for mild retardation evident in some females, carriers of MRX can only be identified once they have an affected son or grandson. Ascertainment of families depends on thorough assessment of the clinical and family history of the presenting male patient. Mildly affected heterozygous females in a family can conceal X-linkage and may cause such families to be less frequently ascertained. Indirect estimates of MRX prevalence can be based on epidemiology studies of the fragile X syndrome and from the study of etiology in pairs of retarded brothers. The ratio of fragile X to MRX is at least 12:9 and may be as much as 12:16 (Appendix II, Paper 2). Since the fragile X syndrome accounts for the greatest proportion of MRXS and up to 40% of all XLMR, MRX may represent an equally large portion of XLMR, while other syndromes occur as rare private disorders (Chapter 5).

In 1990, a clinical review of 19 published pedigrees, with the fragile X excluded, was performed by clinical colleagues to determine any subtle clinical characteristics in MRX families (Appendix II, Paper 2). The level of intellectual handicap was consistent within only 9 sibships and varied from mild to severe within the rest. In two families, obligate carriers, were described as mildly affected heterozygotes. Testicular enlargement was a feature in 6 of 17 families where this was measured. This study demonstrated the need to place criteria on families described as MRX and on the method of clinical evaluation required to establish a basis for comparison of families in whom the gene was to be localised. Clinical criteria established and implemented in the families described below included; X-linked inheritance of the retardation demonstrated by affected males in at least two generations or two maternally related sibships; cytogenetic exclusion of fragile site expression in at least one affected male from each family; information where available on

developmental history and birthweight; anthropometric measurements including head and testicular size; assessment of somatic features, IQ and behaviour.

Variability in the severity of expression of the primary mental retardation phenotype has been shown in affected individuals both within and between families, and in carrier females (Appendix II, Paper 2). Other 'soft' syndromal features or specific adaptive behavioural difficulties, while not evident in more than one affected male, do not prevent a family being regarded as having non-syndromal or non-specific mental retardation. This interfamilial variability supports the possibility of a number of MRX loci on the X chromosome. In view of the lack of clear clinical correlates to establish nosology, this study was conceived as an endeavour to map each gene and work towards the molecular classification of MRX.

The investigations described in this chapter have contributed to the establishment of a system of nomenclature, to the delineation of the number of genes for idiopathic X-linked mental retardation and their distribution along the X chromosome. In families where the MRX gene was localised between flanking markers with significant evidence for linkage, carrier risks were made available when requested by the medical genetics consultant. The correlation of gene location with clinical findings was examined as a means of recognising new syndromes or predicting allelism with known syndromes. A number of families were analysed for mutations in a positional candidate gene, leading to characterisation of a base substitution.

4.2. Nomenclature

Once localisations for MRX genes began to coincide, a need to differentiate between the genes mapped in different families became apparent. Since overlapping regional localisations may or may not represent mutations in the same gene, the MRX gene segregating in each family must maintain an individual identity until mutation screening in MRX associated genes can define them into groups according to a more precise classification. It is worth noting that before a system of nomenclature to differentiate between MRX genes and syndromal genes segregating in different families was established, some confusion arose as to the nomenclature leaving some misclassifications to be rectified. The MRX tag introduced by Suthers (1988), caused the gene in that family to become known as MRX1. Initially those genes mapping to the same region of the X chromosome were referred to by the same number, so that for example all those mapping around the centromere were called MRX1. Since this did not differentiate between families it was suggested (at the Fifth International Fragile X and XLMR Conference) that the name of the first author be appended to each published localisation. The alternative suggestion that each

family be given a consecutive number has now been adopted (Mulley et al, 1992). The classification was based on genetic localisation, such that MRX genes achieving a peak lod score of +2 or greater were assigned a consecutive serial number ie. MRX1, MRX2 ..etc, so that each family (or the gene segregating in each family) has a distinctive label. Consecutive gene numbers are assigned independently by Phyllis McAlpine of the Nomenclature Committee and once the corresponding data are published, are appropriately referenced and entered onto the Genome DataBase (GDB). Syndromal forms of XLMR are represented by the interim symbols MRXS1, MRXS2 ..etc until formal gene symbols can be assigned. Adoption of a third series of symbols; MRXC1, MRXC2 ..etc has been proposed to define those MRX genes defined by contiguous gene syndromes (personal communication J-L Mandel). This nomenclature has been applied throughout this thesis.

In the interim, however some misnomers have arisen in the literature. The two families with lod scores greater than +2 described by Arveiler et al, (1988) have been retrospectively named MRX2 and MRX4 (Mulley et al, 1992). The MIM catalogue has designated these loci MRX1 and MRX2 respectively, although the latter locus overlapping with MRX1 (Suthers et al, 1988) has been referred to on occasion as MRX1 (McKusick, 1994; OMIM, 1995). Samanns et al, (1991) refer to the disease locus in another family as MRX1 since it too maps in the pericentromeric region, and a further family earlier referred to as MRX1 had also been mapped to Xp11-q21 (Willems et al, 1993). The latter two have since been assigned the numbers MRX5 and MRX9 respectively (section 4.6.).

The MRX2 symbol had been previously applied to the Sutherland-Haan syndrome (Chapter 5) but was later withdrawn and reclassified as MRXS3 to establish a consistent nomenclature that also differentiated between the syndromal and non-syndromal forms. Another MRX gene identified in Xp22 by deletion mapping in contiguous gene syndrome patients with complex phenotypes was thought to overlap the regional localisation of MRX2 (Ballabio et al, 1989). This locus was referred to also as MRX2 and is entered in GDB as such, yet recent refinement of the original MRX2 gene interval excludes a region of overlap (Hu et al, 1994). Henceforth this MRX occurring in association with complex phenotypes ought to be referred to as MRXC1.

Self-assignment of MRX numbers resulted in simultaneous adoption of the same gene number, for example MRX3, initially assigned by two groups have been subsequently renamed MRX3 and MRX8 (Appendix II, Paper 1; Schwartz et al, 1992). Symbol assignments were thenceforth designated by the Nomenclature committee. The system for requesting MRX numbers was established so that for families where the lod score of +2 had been attained this number could be included into the publication of data from that family. To date 32 MRX numbers have been issued however some of these remain 'unassigned' in that

the numbers were obtained as personal communication from Dr McAlpine but the corresponding journal article has not yet appeared. Recent publications that reduce former localisations, other MRX loci published without numbers and those mapped by physical means eg the MRX mapped in the contiguous gene syndromes at Xp22 (MRXC1) and that involved in the choroideremia-MR-DFN3 complex (MRXC2) should be added to update the MRX listing in GDB.

4.3. Mapping MRX genes

Clinical homogeneity within the MRX proportion of XLMR means that the only approach to differentiating between them is through molecular characterisation. Linkage studies that assign MRX genes to different parts of the chromosome demonstrate locus heterogeneity through non-overlapping regional localisations. Analysis of families separately will 'split' MRX genes such that each gene localised represents the disorder in a single family. Where more than one disease gene is assigned to the same chromosomal region it cannot be determined by linkage analysis whether only a single gene is involved, just as variable clinical indicators do not necessarily exclude allelic heterogeneity at a single locus. Recognition of subtle clinical and behavioural or adaptive characteristics on their own are unlikely to differentiate co-localised subtypes as specific entities, particularly as inter and intra-familial variation in expression of phenotype is not uncommon. The possibility of misdiagnoses or phenocopies (males who are mentally retarded but not carrying the MRX mutation) may need to be addressed. Suspected allelic disorders mapping to the same chromosomal band can in future be substantiated by molecular cloning and identification of the mutation(s) in the gene. The basis for differences in XLMR phenotype or severity may also be demonstrated through 'lumping' of allelic disorders. Some of the MRX genes may be found to be milder allelic forms of syndromal conditions. For now progress in gene identification is based on linkage mapping to refine candidate gene regions. In this manner the contributions of laboratory based molecular genetic mapping are essential to clinical nosology.

Genetic heterogeneity of MRX was initially demonstrated in the families now known as MRX1, MRX2 and MRX4, by genetic evidence of X-linkage to widely separated chromosomal localisations (Suthers et al, 1988; Arveiler et al, 1988; Mulley et al, 1992). These regional localisations were distinct from the marker X (FRAXA) associated with the most common cause of XLMR, the fragile X syndrome (Chapter 1.6.1.). At commencement of this study therefore, only two broadly defined MRX genes had been reported in the literature. Given earlier speculations that between 7 and 25 loci could account for the prevalence of non-specific XLMR (section 4.1), gene mapping was the only way to

accurately determine the minimum number of X-linked genes associated with mental retardation and differentiate between the causative genes in MRX families.

4.3.1. Materials and Methods

Over a period of five years, peripheral blood samples were collected by clinical colleagues from individuals in eleven large MRX families in accordance with the criteria for clinical evaluation as established (Appendix II, Paper 2). Nuclear DNA was extracted as described (Chapter 2.3.). X-linked transmission was assumed from the presence of at least one affected male either in the maternal family of the propositus or among his male siblings, absence of equivalently affected females and absence of consanguinity. Affected men did not have any recognised disorder of metabolism, major dysmorphic features or other consistent physical abnormalities.

Several MRX families including MRX1, MRX3, MRX10, MRX11, MRX12, MRX13, MRX17, MRX18 and Family E came from regional centres and outlying areas of NSW and were supplied to this linkage study by Dr. B. Kerr and Dr. G. Turner. They performed all clinical assessments and collected or coordinated collection of blood or tissue samples from participating family members. All evaluations including any subjective assessments in this group of families were thus standardised, potentially allowing clinical comparisons between them. Cell lines, fibroblasts or lymphoblasts, were established from at least one affected male and his mother from each family.

At least one affected male from each family was screened for cytogenetic abnormalities by karyotype analysis. Structural abnormalities such as translocations, duplications, inversions and large deletions or insertions can be detected cytogenetically. A cytogenetic abnormality would have been immediately useful in pinpointing the location of a candidate disease gene. Specific attention was given to the analysis of expression of a fragile site on the X chromosome, to exclude the fragile X syndrome associated with FRAXA. Lack of expression of a fragile site in Xq27-28 can now retrospectively be extended to include the cytogenetically indistinguishable FRAXE associated with non-specific mental retardation (Sutherland and Baker, 1992; Knight et al, 1993; Knight et al, 1994; and Appendix V, Papers 6 and 8) and FRAXF with no proven disease association (Hirst et al, 1993; Parrish et al, 1994; Ritchie et al, 1994).

Once fragile site expression was excluded, the rest of the chromosome was analysed for molecular evidence of linkage or exclusion of the MRX gene to polymorphic markers along its length (Chapters 2 and 3). The demonstrated genetic heterogeneity of MRX precluded the combination of data from families, thus each family was treated separately for linkage analysis (Chapter 2.7.). Genotyping of key family members according to the strategy

outlined in Chapter 3.3.3. determined whether a marker was concordant, discordant or not informative. In the search for linkage, discordant findings reflect recombination as different alleles present in carriers and affected males, or the same allele in both affecteds and unaffecteds. Discordant markers exclude linkage to the MRX gene, therefore complete genotyping of concordant informative markers was a priority.

For the purposes of linkage analysis the frequency of the disease allele was fixed at one in ten thousand though the gene segregating in each family was analysed independently. Two-point lod scores were calculated on LIPED or MLINK (Chapter 2.7.). By convention, for X-linked loci, the test for linkage was declared significant when the maximum lod score exceeded Z=+2 (Ott, 1991) and was accepted as sufficient support for linkage (Chapter 1.3.1.). Linkage data were generated under the model of X-linked recessive inheritance except in those families with manifesting females. Special schooling required by females with mild mental retardation was the criterion regarded as evidence for partial dominance. In those families where incomplete penetrance was a factor, penetrance in females was coded at an arbitrary 0.7, under the assumption that these females carry the same mutation as the affected males. Complete penetrance was assumed in males. Once recombinants were detected at flanking markers, a minimum of two adjacent loci were required to independently confirm the crossover and define the interval of the disease gene.

Recombination events have been detected in females in some families (MRX10), but only recombinants in males and obligate carrier females have been used as determinants of the gene interval to provide information on genetic risks. The diagnosis in affected non-obligate carrier females, generally with milder symptoms than affected males from the same family, cannot be confirmed as arising from the same X-linked mutation; hence, male and obligate carrier determinants ensure conservative estimates of risk, while recombination in other females may define a tighter interval for selection and mutation screening of candidate genes. In all families where grandparents were deceased or genotypes were inferred from offspring, the disease status was coded as normal in both.

Multipoint analyses were used to maximise the linkage data when families were not fully informative at any single locus. In families where a marker was fully informative in all meioses, the peak lod score by pairwise analysis was identical to the multipoint lod score. Multipoint analysis using LINKMAP on the FASTLINK system (Chapter 2.7.3.) was therefore particularly applied in families where none of the markers were fully informative over the gene interval. When linked markers are relatively close together or equally informative within a family, and there is no recombination detected the multipoint curve may sweep across several loci and only plunge (antimode) at a given marker locus with evidence of recombination (Lathrop et al, 1984).

The size of each family segregating MRX and unambiguous clinical classification of all family members were critical factors in the ability to demonstrate linkage. The equivalent of eight potentially informative meioses are sufficient to establish linkage to the X chromosome by LOD2 criteria. The ability to determine phase is an important requirement since phase unknown meioses give only a limited amount of information to the analysis. Usual pedigrees comprise a mixture of phase known and unknown with many family members unattainable or unwilling to participate until beneficial effects of the research became apparent. The difficulty is that in many cases the benefits to given individuals cannot be made fully apparent until sufficient numbers of family members come forward. One of the benefits of the time taken for sample collection was that the technology used for molecular genetic analysis was undergoing constant improvement (Chapter 3). Families remaining too small to achieve LOD2 for linkage mapping of the gene do not contribute to enumeration of MRX loci and were omitted from this study except for one (Family E - Appendix II, Paper 4) included in screening of a positional candidate gene (section 4.6.). These samples will be useful once more family members come forward or for screening known MRX mutations.

4.3.2. MRX gene localisations

X-linked segregation of mental retardation was substantiated in these families by evidence of linkage to polymorphic markers on the X chromosome. The regional localisations of ten MRX gene loci have been reported in peer-reviewed scientific literature (Appendix II, Papers 1, 3, 4, 5, 6 and 7). The following section therefore only summarises and up-dates these published findings, with individual identifiers corresponding to the published pedigree numbers. MRX numbers were assigned to those families in which peak lod scores exceeded +2 in accord with the guidelines for nomenclature (section 4.2). Regions demonstrating significant linkage become the sites for future investigation of candidate genes. An immediate advantage to members of families in which the gene has been localised is the availability of genetic risk analysis. Determination of individual-specific risks will be particularly valuable in enabling consultands to make informed reproductive choices - see for example MRX1.

MRX 1 (Family H/L)

The clinical features of this large family with non-specific mental retardation were described by Turner et al (1971). Affected men had moderate retardation, but no other distinguishing features. The MRX1 gene was initially mapped, in this laboratory, by linkage

to the RFLP marker DXS14(p58.1) (Suthers et al, 1988). The gene localisation was limited by recombination events, to the interval between the markers DXS7(L1.28) and DXYS1(pDP34). The estimated 90% confidence interval for the recombination fraction was 0.00-0.22, a total interval of approximately 44cM. Members of the extended pedigree (Figure 4.1), including 13 women of unknown carrier status, had been genetically counselled based on multipoint risk estimations derived by a resampling method of 'bootstrapping' (Suthers and Wilson, 1990). As polymorphic markers within the interval increased in numbers and PIC (Chapter 3) it was considered worthwhile to attempt to reduce the localisation in this family. This was achieved and published (Appendix II, Paper 3) as a reduced localisation between the dinucleotide repeat markers DXS426 and DXYS1X spanning approximately 30cM. Refinement of the MRX1 gene limits separated this locus from the MRX10 locus and defined at least two distinct MRX loci in the pericentromeric region (Appendix II, Paper 3).

Continued analyses with microsatellite markers as they appeared in the literature (Chapter 3) has further reduced the localisation of MRX1 within limits defined by the recombinant III-21 at SYN1 and DXS426 on Xp, and recombination detected in II-4 and II-8 at DXS106 and DXS1125 on Xq (Appendix II, Paper 7). Several markers within this interval including PFC, DXS1126, DXS255 and ALAS2 were not informative in the family. Haplotypes were constructed at 7 loci flanking and across the MRX1 interval (Figure 4.1). Pairwise lod scores for informative markers flanking and within the original interval (Table 4.1) were calculated on MLINK (Chapter 2.7.). The peak lod score of 5.07 (θ=0.0) at AR is the maximum achievable with fully informative genotype data on all participating family members. The markers DXS991, AR and PGK1P1 within the interval show no evidence of double crossover on haplotype analysis. Recombination between the DXS106 and DXS1125 loci in III-16 is consistent with DXS1125 lying proximal to DXS106 and confirms the order of these loci suggested by recombinants in Family R at MRX17. At the DXS453 and DXS559 loci, the genotype of the matriarch I-2 could not be inferred from her offspring, hence with phase not known, recombinants could not be detected. This apparent discrepancy in the data given the marker order in proximal Xq as AR-PGK1P1-DXS106/DXS1125-DXS453/DXS559-DXS227-DXS566-DXS986-DXYS1X is overruled by the evidence that the same recombinants (and more) are detected at distal loci where phase could be inferred.

Multipoint analyses were not necessary to indicate the most likely position of the MRX1 locus with respect to the background map since the peak location score 23.33 at AR is equivalent to a peak multipoint lod score of 5.07, as expected. The RFLP marker DXS14 is known from physical data (XCW5) to lie very near the centromere, flanked by DXS991

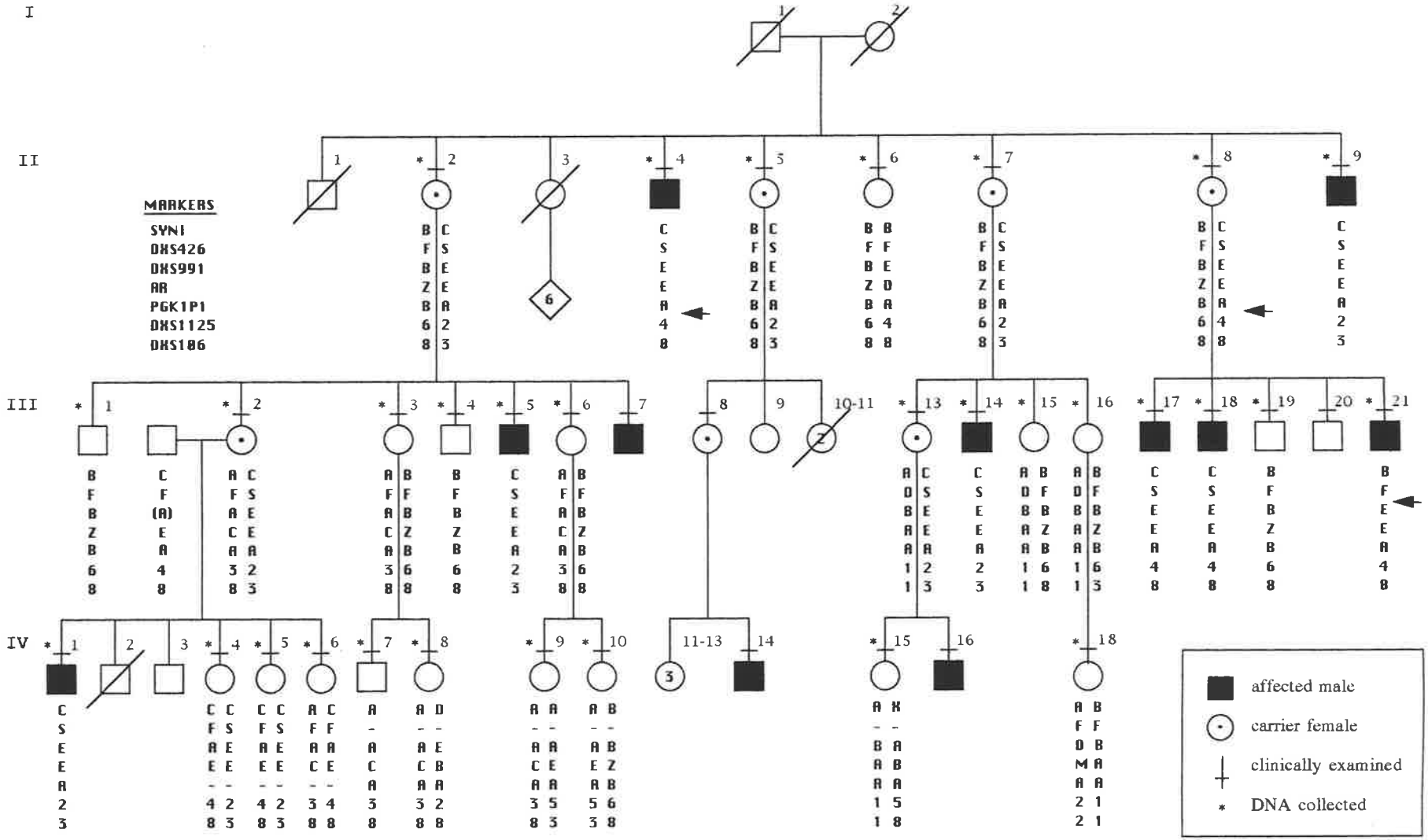


Figure 4.1. Extended MRX1 pedigree. Haplotypes have been constructed from genotypes at seven marker loci flanking and across the gene interval. Arrows indicate detected recombination events.

TABLE 4.1 : PAIRWISE ANALYSIS BETWEEN MRX1 AND MARKERS ON THE X CHROMOSOME

	θ								
	0.001	0.01	0.05	0.1	0.2	0.3	0.4	Zmax	θ
DXS538	-4.24	-1.30	0.54	1.11	1.29	1.02	0.54	1.30	0.18
DXS7*	-1.24	0.70	1.82	2.06	1.89	1.38	0.68	2.07	0.12
DXS1003	-1.07	0.87	1.99	2.23	2.04	1.50	0.75	2.24	0.12
SYN1	2.06	2.99	3.40	3.31	2.75	1.96	0.98	3.40	0.06
DXS426	1.76	2.70	3.11	3.03	2.52	1.78	0.88	3.11	0.06
DXS991	3.71	3.65	3.41	3.10	2.45	1.73	0.92	3.71	0.00
DXS14*	2.89	2.85	2.67	2.43	1.93	1.37	0.74	2.90	0.00
AR	5.06	4.99	4.68	4.26	3.36	2.32	1.15	5.07	0.00
PGK1P1	3.24	3.20	3.00	2.74	2.18	1.55	0.84	3.25	0.00
DXS1125	-0.94	1.00	2.12	2.36	2.17	1.64	0.91	2.36	0.12
DXS106	-2.14	-0.19	0.99	1.32	1.32	1.03	0.57	1.38	0.15
DXS453	3.33	3.28	3.07	2.80	2.21	1.57	0.84	3.34	0.00
DXS559	3.82	3.76	3.51	3.19	2.50	1.75	0.93	3.82	0.00
DXS227	-0.94	1.00	2.12	2.36	2.17	1.64	0.91	2.36	0.12
DXS566	-0.94	1.00	2.12	2.36	2.17	1.64	0.91	2.36	0.12
DXS986	-3.17	-1.20	0.04	0.45	0.63	0.54	0.31	0.63	0.20
DXYSIX*	-11.27	-6.30	-2.93	-1.61	-0.48	-0.03	0.10	0.10	0.41

* RFLP markers genotyped earlier (Suthers et al, 1988).
Lod calculated with published allele frequencies.

Table 4.2. DETERMINATION OF CARRIER RISK IN MRX1 FEMALES

INDIVIDUAL	Pedigree alone	Multi-point*	Flanking markers
II-6	33%	1-16%	0.10%†
III-3	33%	<1%	0.19%†
III-6	50%	<2%	0.20%†
III-15	50%	<2%	0.20%†
III-16	50%	<2%	0.20%†
IV-4	50%	81-100%	99.8%†
IV-5	50%	81-100%	99.8%†
IV-6	50%	0-19%	0.2%†
IV-8	17%	<1%	0.002%‡
IV-9	25%	<1%	0.002%‡
IV-10	25%	<2%	0.46%‡
IV-15	50%	32-39%	0.46%‡
IV-18	25%	0-4%	0.10%†

* Suthers and Wilson (1990).
† Distance between informative flanking markers DXS426-DXS1125/DXS106 is 8.6cM.
‡ Distance between informative flanking markers SYNI-DXS1125/DXS106 = 12.8cM (DXS426 locus was not informative in mother or not typed in daughter ie IV-15).

on Xp and AR on Xq, this marker has been arbitrarily inserted equidistant to these loci in the background map. The MRX1 gene interval DXS426 - DXS106/DXS1125 now spans a genetic distance of 8.6cM (Chapter 3).

Based on pedigree information alone the daughters of an obligate carrier all have an *a priori* risk of 50% of being a carrier (Table 4.2). Genotyping of several flanking markers provides greater flexibility and precision in determining carrier and prenatal risk estimates. Gene mapping in 1988 had reduced the risk that IV-15 was a carrier from 50% to 32-39% on multipoint risk analysis (Suthers and Wilson, 1990). Reduction of the distance spanned, now improves the risk calculated on MLINK using the flanking markers SYNI and DXS1125/DXS106 to a negligible 0.46%. In this example the mother of IV-15 (III-13) was uninformative at the closest flanking marker DXS426. The distance between the next nearest informative marker SYNI, and DXS426 has been estimated from a genetic map based on CEPH database ver 3 (Kirchgessner et al, 1991), and may in fact be less than the 4.2cM used here. Similarly, the risk that IV-4 carries the MRX1 gene is now greater than 99.8%, and is information that is more useful to the consultand than the 81-100% range available in 1990 (Suthers and Wilson). The genetic risk of MRX1 is less than one percent in eleven of the thirteen women (Table 4.2).

MRX 3 (Family I)

The MRX3 gene in this family (Figure 1; Appendix II, Paper 1) was initially localised distal to the fragile X using RFLP markers spanning the X chromosome. The gene was mapped to Xq28 by support of linkage to the fully informative VNTR marker DXS52 (St14) and exclusion of the rest of the X. A recombinant event detected in VI-13 with the RFLP marker at DXS304 (U6.2) defined the proximal limit. The distal limit was physically defined by the telomere of the chromosome (Appendix II, Paper 1). No consistent clinical features besides mild to moderate mental retardation were noted in the 5 living affected males examined. Slightly short stature with relative macrocephaly and large limb spans in some affected males may represent soft syndromal signs. Four of the five men were described as aggressive despite evidence of self-sufficiency (Appendix II, Paper 1).

Following localisation of MRX3 to Xq28, a member of the family, VI-2, presented with a request for prenatal diagnosis. Flanking marker analysis could not be applied since no recombinants defined the distal limit of MRX3. The distance between the proximal limit, DXS304, and DXS52 was estimated to be 12% from multipoint analysis in CEPH individuals (Suthers et al, 1991). The distance from DXS52 to qter was also assumed to be 12%, and risk could be calculated as an interval where the recombination fraction varied

between 0% and 12%. MRX3 segregates with allele 2 of DXS52 in this family, VI-2 is heterozygous at this locus, while her normal son, VII-3, carries allele 8. The carrier risk of VI-2 was in the range of 66-93% with recombination frequency at the maximum of 12%. Without the genotype of VII-3, her risk interval was 53-86%. PCR-based sexing (Kogan et al, 1987) of the CVS sample determined that the fetus was male. Genotype data at DXS52 gave a risk of 58-93% that the fetus carrying allele 2 was affected. Had the fetus carried allele 8, his risk under 12% recombination would have been 7.9%. Carrier risks less than 12% were similarly determined for other women in this family, VI-1, VI-5, VI-11 and VI-12, none of whom carry allele 2 at DXS52. Risks for VI-7, VI-8 and VI-9 could not be determined since their mother, V-15, was uninformative at this locus.

Further refinement to the localisation was attempted with dinucleotide repeat markers mapped within the gene interval in Xq28 (Chapter 3). All family members from whom DNA had been stored were genotyped and included in the linkage analysis (V-13 and V-15 had not been included in the published analysis because of suspected non-paternity and uncertain carrier status). Detection of non-paternity at DXS52, confirmed at the dinucleotide repeat marker p26, was addressed by coding separate fathers for V-13 and V-14 as well as V-15 and V-16 for the purposes of linkage analysis. This reduced the peak lod at DXS52 from the published value Zmax=2.84 (θ=0.0) to 2.14 (θ=0.0). Pairwise lod scores for DXS304 and markers distal are given in Table 4.3. No recombinants were detected at any locus distal to DXS304, therefore the localisation cannot be improved. PCR formatted RFLPs at DXS707 and DXS605 were uninformative and only partly informative respectively. The most distal polymorphic marker on the X chromosome was DXYS154 in the pseudoautosomal region.

The clinical relevance of the recently cloned FRAXF, also in Xq28 (section 1.6.1), has been shown to lack clear direct correlation with mental impairment (Parrish et al, 1994). Refinement of the FRAXF localisation proximal to DXS304 (U6.2) - onto YAC yWXD873, a minimum of 100kb distal to IDS (Ritchie et al, 1994) - excludes involvement of FRAXF in the etiology of MRX3. A number of other candidate loci have been described in the gene rich Xq28 region (Maestrini et al, 1992). Three cDNAs expressed in brain have been sublocalised on a radiation hybrid panel within Xq28, two of these lie distal to DXS304 (Smahi et al, 1994). The complex combined disorders of spastic paraplegia, MASA syndrome and hydrocephalus (Fryns et al, 1991) together with X-linked corpus callosum agenesis comprise the clinical spectrum of CRASH syndrome (Fransen et al, 1995). These neurologic anomalies are allelic mutations in the gene for the L1 neural cell adhesion molecule (L1CAM) in Xq28 (Jouet et al, 1994). The MRX3 gene may represent the allelic

TABLE 4.3 : PAIRWISE ANALYSIS BETWEEN MRX3 AND POLYMORPHIC MARKERS IN Xq28

	θ								
Loci	0.001	0.01	0.05	0.1	0.2	0.3	0.4	Z max	θ
DXS1113	-9.37	-5.38	-2.65	-1.54	-0.57	-0.15	0.01	0.02	0.45
DXS304	-2.39	-1.39	-0.72	-0.44	-0.20	-0.08	-0.02	0.00	0.50
DXS374	-0.72	0.71	0.65	0.58	0.43	0.28	0.14	0.72	0.00
GABRA 3	1.43	1.41	1.30	1.16	0.88	0.60	0.31	1.43	0.00
DXS52	2.14	2.11	1.96	1.76	1.33	0.88	0.44	2.14	0.00
p26	2.48	2.43	2.22	1.96	1.46	0.97	0.50	2.49	0.00
p39	1.29	1.26	1.12	0.95	0.65	0.40	0.19	1.29	0.00
DXS15	0.69	0.68	0.62	0.55	0.40	0.26	0.13	0.70	0.00
DXS605	1.04	1.02	0.91	0.77	0.50	0.25	0.08	1.05	0.00
F8C	1.96	1.93	1.78	1.60	1.22	0.83	0.43	1.97	0.00
DXS1108	2.14	2.10	1.94	1.73	1.31	0.90	0.46	2.14	0.00
DXYS154	1.30	1.27	1.13	0.96	0.66	0.41	0.19	1.31	0.00

variant associated solely with MR in this complex disorder since none of the males had any features or radiologic signs of MASA or skeletal dysplasia. Allelic mutation in L1CAM, 1-2cM distal to DXS52, is a possible candidate that has not yet been excluded. Since it is difficult to further reduce the genetic interval of the gene, further candidate loci are being more efficiently screened for potential disease causing mutations through international collaborations established with Drs. Niklas Dahl (Norway), Daniela Toniolo (Italy) and Aïda Metzenberg (USA) who are independently preparing transcription maps and actively cloning genes within Xq28.

MRX 10 (Family Q)

The MRX10 gene segregating in family Q (Figure 1a; Appendix II, Paper 3) was initially localised between DXS28 and DXS426, overlapping the entire dystrophin gene locus This localisation did not overlap with that of MRX1, defining a distinct MRX locus (Appendix II, Paper 3). Three women of unknown carrier status in the family (IV-12, IV-16 and V-10) had unexplained mild/moderate intellectual handicap although obligate heterozygotes were intellectually normal. Assuming variable heterozygote expression of MRX, penetrance was set at 0.75. Pairwise lod scores between MRX10 and several subsequently genotyped markers within the interval are given in Table 4.4. A peak lod score of 4.14 (θ=0.0) was achieved with the STR49 (DXS1236) dinucleotide repeat marker in intron 49 of the dystrophin gene in Xp21.3. The markers DXS992, STR50 (DXS1235), STR45 (DXS1237), DYSII, 5'DMD, DXS84, DXS538, DXS1068, DXS7 (both the dinucleotide repeat and the RFLP), OTC, MAOB and MAOA3 were uninformative. The distal boundary remains unchanged, defined by the recombinant IV-11 at DXS28. Recombination detected in an affected male (IV-10) at DXS228 however, redefines the conservative proximal boundary for MRX10 (Appendix II, Paper 7).

Recombination was also detected in an affected female (IV-16) at STR44 (DXS1238) and at DYSIII, DYSI and CYBB more proximally (Figure 4.2). Inability to infer the parental genotypes in generation II at the intervening DXS164 locus results in a phase unknown analysis so that recombination cannot be detected in this individual at this locus. If her mental impairment has been correctly diagnosed and is due to mutation at the same locus as in the affected males in the pedigree, this event defines the most proximal boundary of the MRX10 gene at STR44. It is not altogether surprising that this recombination event occurred between STR44 and DXS997, since this is a recognised hotspot of recombination in normal families (Oudet et al, 1992; section 3.4.1.2.). All three retarded females carry the same haplotype between DXS28 and STR44 as the one segregating with the disease

TABLE 4.4 : PAIRWISE ANALYSIS BETWEEN MRX10 AND X CHROMOSOME MARKERS

	θ								
Loci	0.001	0.01	0.05	0.1	0.2	0.3	0.4	Zmax	θ
DXS418	-1.48	0.46	1.57	1.81	1.63	1.15	0.55	1.81	0.12
DXS443	1.01	1.95	2.37	2.31	1.86	1.23	0.57	2.38	0.07
DXS41	0.19	1.13	1.59	1.57	1.22	0.74	0.30	1.61	0.09
DXS28	-0.29	0.67	1.18	1.24	1.04	0.68	0.26	1.25	0.09
3'DMD	2.59	2.56	2.41	2.19	1.69	1.13	0.55	2.59	0.00
STR49	4.13	4.07	3.80	3.44	2.65	1.76	0.84	4.14	0.00
DXS997	3.32	3.28	3.09	2.81	2.16	1.42	0.66	3.32	0.00
STR44	0.83	1.78	2.23	2.21	1.82	1.26	0.64	2.25	0.07
DXS164	1.67	1.64	1.52	1.35	1.00	0.63	0.23	1.67	0.00
DYSIII	1.10	2.04	2.49	2.46	2.02	1.38	0.66	2.51	0.07
DYSI	-0.54	0.49	1.19	1.34	1.23	0.94	0.53	1.35	0.12
CYBB	-2.18	-0.21	1.03	1.40	1.43	1.10	0.60	1.48	0.16
DXS228	-2.21	-0.25	0.96	1.31	1.34	1.03	0.55	1.39	0.15
MAOB	-6.08	-3.11	-1.14	-0.41	0.11	0.20	0.11	0.20	0.29
DXS1003	-4.91	-1.95	-0.04	0.61	0.95	0.84	0.50	0.96	0.22
TIMP	-2.26	-0.30	0.89	1.23	1.24	0.93	0.49	1.29	0.15
PFC	-4.47	-2.49	-1.16	-0.68	-0.36	-0.32	-0.27	0.00	0.50
DXS426	-6.17	-3.19	-1.18	-0.40	0.18	0.32	0.23	0.32	0.30
DXS573	-5.38	-2.42	-0.51	0.15	0.54	0.51	0.31	0.56	0.24

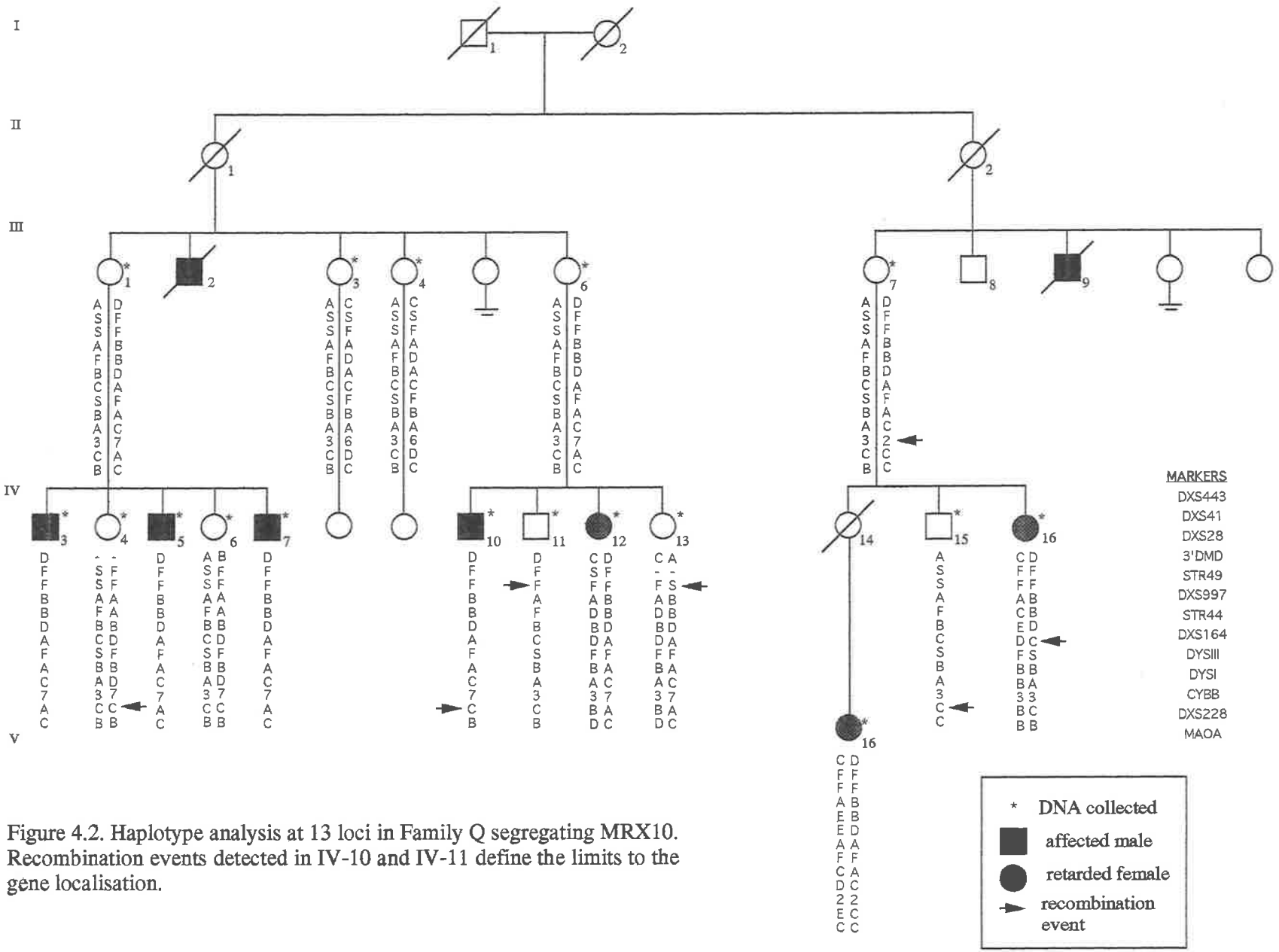


Figure 4.2. Haplotype analysis at 13 loci in Family Q segregating MRX10. Recombination events detected in IV-10 and IV-11 define the limits to the gene localisation.

phenotype in affected males. Probable recombination in IV-13, an apparently mentally normal female, cannot refine the interval since her affection status cannot be determined as normal or asymptomatic carrier. It is worthy of note however, that if she were assumed normal, the recombination event between DXS28 and 3'DMD results in the reverse complementary haplotype seen in IV-11 (Figure 4.2). This unusual circumstance may point to localisation of MRX10 between the independent recombinant events in the 1.1cM separating these loci.

Multipoint analysis resulted in a peak location score of 19.812 between DXS28 and 3'DMD. The peak multipoint lod score equivalent is 4.31 between these loci. The gene interval spans 22.4cM between DXS28 and DXS228, but only 6.2cM between DXS28 and STR44. The effect of the female recombinant can be seen to reduce the likelihood of the gene being located in the region proximally, between STR44 and DXS228. If the proximal limit were accepted at STR44, given the assumption that the retarded females are carriers of MRX10, this redefinition of MRX10 would separate this gene from overlap with the regional localisations of MRX11 and MRX18.

Refinement of this regional localisation is valuable for provision of risk to women of unknown carrier status. In 1991, a risk estimate could not be calculated for one consultand, IV-4, because a crossover had occurred between 3'DMD and DXS426 making it impossible to determine her status since the gene could lie anywhere in the interval. The conservatively reduced localisation of MRX10 between DXS28 and DXS228, now demonstrates that she does not inherit the haplotype associated with mental retardation in her siblings, since the crossover occurred between DXS1003 and TIMP, outside the gene region. Risks estimated in this family were based on the gene interval defined in individuals of known status with respect to the disease gene (ie. males and obligate carrier females) rather than the assumption that affected females carry the same mutation as affected males.

MRX 11 (Family H)

The gene localisation in family H, with moderate to severe mental retardation in males spanning three generations, was also established by the candidate and reported in 1992 (Appendix II, Paper 3). One (V-21) of the eight affected males examined had minor anomalies including microcephaly, brachycephaly, micrognathia, a triangular face and was severely retarded (individual numbers are as given in the published pedigree Figure 1b; Appendix II, Paper 3). Two of the women, V-15 and V-19, can be described as manifesting heterozygotes since they required special schooling. The MRX11 locus was initially mapped between the dystrophin locus at Xp21.3 and DXS255 near the centromere. Further analysis

TABLE 4.5 : PAIRWISE ANALYSIS BETWEEN MRX11 AND X CHROMOSOME MARKERS

	θ								
Loci	0.001	0.01	0.05	0.1	0.2	0.3	0.4	Zmax	θ
DXS28	-8.49	-4.53	-1.92	-0.96	-0.25	-0.04	0.00	0.01	0.44
STR49	-1.50	-0.49	0.20	0.46	0.61	0.55	0.34	0.62	0.22
DXS997	-4.42	-1.48	0.35	0.92	1.10	0.87	0.48	1.11	0.18
STR45	-3.64	-1.67	-0.38	0.07	0.32	0.29	0.16	0.33	0.23
DXS269	-3.89	-1.92	-0.67	-0.24	-0.01	0.01	-0.01	0.01	0.27
DXS164	-0.85	0.14	0.78	0.96	0.93	0.66	0.29	0.99	0.14
5'DMD	2.88	2.83	2.60	2.29	1.65	0.96	0.32	2.89	0.00
DYSIII	2.73	2.71	2.59	2.42	1.97	1.41	0.74	2.73	0.00
DYSII	2.14	2.13	2.05	1.91	1.53	1.02	0.45	2.15	0.00
DXS84	3.06	3.01	2.78	2.48	1.83	1.14	0.46	3.07	0.00
DXS538	2.36	2.32	2.14	1.91	1.41	0.87	0.34	2.37	0.00
CYBB	2.74	2.70	2.50	2.23	1.67	1.07	0.45	2.75	0.00
DXS1068	3.86	3.80	3.54	3.18	2.40	1.52	0.64	3.86	0.00
DXS7	-0.74	0.24	0.84	0.98	0.89	0.61	0.27	0.99	0.12
MAOB	-0.42	0.56	1.15	1.29	1.18	0.87	0.46	1.30	0.12
PFC	0.99	1.94	2.40	2.38	1.98	1.39	0.72	2.42	0.07
DXS573	-3.83	-1.84	-0.48	0.04	0.37	0.37	0.22	0.40	0.25

of this family with markers within this interval have been effective in reducing the original localisation. The new interval is defined by recombination events detected at flanking markers DXS164 distally and the DXS7 dinucleotide proximally (Appendix II, Paper 7). Two point lod scores in the region are presented in Table 4.5. A peak lod score of 3.86 (θ=0.0) was obtained with the DXS1068 dinucleotide repeat marker. Several markers including 3'DMD, DMDC2, DXS1003, OTC, DXS7 (L1.28 RFLP), SYNI, TIMP and DXS426 were uninformative in key obligate carriers. Multipoint analysis was not performed. The earlier localisation encompassed a genetic distance of approximately 36cM, this has now been reduced to approximately 16cM (Chapter 3.4.1.2.). Six women of unknown carrier status who participated in the mapping study may benefit from risk estimations based on this refinement of the MRX11 gene localisation.

The MRX11 regional localisation encompasses the 5' end of the dystrophin gene. Some patients with muscular dystrophy are mentally retarded suggesting that specific mutations in the brain specific promoter or the dystrophin gene itself may be involved (see section 4.6.2.). Seven of the dystrophin gene exons (4, 12, 17, 19, 44, 45, and 51) and the brain promoter were screened by PCR in four of the affected males (IV-5, IV-19, V-6, VI-16). No deletions were detected in these exons nor conformational polymorphisms in the promoter.

MRX 12 (Family M)

Notable clinical findings in this family with mild to moderate mental retardation included low birth weights (<1.8kg) and short stature (usually defined as being below the 3rd percentile) in all affected males, hypotelorism and large ears in 5 out of 6 examined (Appendix II, Paper 3). Individuals IV-5 and V-2 (Figure 1c; Appendix II, Paper 3) were not available for examination however their vulnerability in early life and development conformed to the family pattern (personal communication Dr. B. Kerr). The MRX12 gene was localised between recombinants detected at the 3'DMD and DXS453 loci (Appendix II, Paper 3).

The localisation of the MRX12 gene has been subsequently significantly reduced by genotyping of dinucleotide repeats between flanking markers at the 3'DMD and DXS453 loci. Pairwise lod scores in this interval reveal the peak lod score of z=3.31 (θ=0.0) at 5 fully informative loci (Table 4.6). The key obligate carrier (III-7) was homozygous rendering the family uninformative at KAL, DXS997, DXS228, MAOA, DXS1003, SYN1, PFC, DXS426, DXS573 and COL4A5 dinucleotide repeats and DXS43, DXS84, TIMP, DXS1, DXS14, DXYSI and DXS10 RFLP loci. A number of affected males were genotyped

TABLE 4.6 : PAIRWISE ANALYSIS BETWEEN MRX12 AND X CHROMOSOME MARKERS

	θ								
Loci	0.001	0.01	0.05	0.1	0.2	0.3	0.4	Zmax	θ
3'DMD	-6.29	-3.32	-1.35	-0.61	-0.06	0.07	0.04	0.07	0.32
DXS992	-5.69	-2.72	-0.77	-0.05	0.44	0.50	0.34	0.51	0.27
STR50	-0.97	0.01	0.63	0.81	0.82	0.64	0.36	0.85	0.15
STR49	-5.69	-2.72	-0.77	-0.05	0.44	0.50	0.34	0.51	0.27
STR45	-5.99	-3.02	-1.05	-0.31	0.24	0.36	0.26	0.36	0.30
STR44	-5.69	-2.72	-0.77	-0.05	0.44	0.50	0.34	0.51	0.27
DXS164	1.20	1.18	1.09	0.98	0.72	0.44	0.15	1.20	0.00
5'DMD	3.01	2.96	2.77	2.51	1.94	1.31	0.62	3.01	0.00
DYSII	3.31	3.26	3.04	2.76	2.15	1.46	0.70	3.31	0.00
DYSI	2.91	2.87	2.66	2.39	1.81	1.16	0.46	2.91	0.00
DXS538	3.31	3.26	3.04	2.76	2.15	1.46	0.70	3.31	0.00
DXS1068	3.01	2.96	2.74	2.46	1.85	1.16	0.42	3.01	0.00
CYBB	3.31	3.26	3.04	2.76	2.15	1.46	0.70	3.31	0.00
DXS7	2.07	2.04	1.92	1.76	1.41	1.00	0.53	2.07	0.00
MAOB	3.31	3.26	3.04	2.76	2.15	1.46	0.70	3.31	0.00
DXS1126	3.01	2.96	2.74	2.46	1.85	1.16	0.42	3.01	0.00
DXS255	3.31	3.26	3.04	2.76	2.15	1.46	0.70	3.31	0.00
ALAS2	-3.99	-2.00	-0.65	-0.14	0.22	0.28	0.19	0.29	0.28
DXS991	-1.27	-0.28	0.35	0.56	0.61	0.50	0.28	0.62	0.17
AR	-3.97	-1.98	-0.65	-0.14	0.22	0.28	0.19	0.29	0.28
DXS453	-6.99	-4.00	-1.93	-1.09	-0.37	-0.08	0.02	0.03	0.43
DXS105	-7.82	-4.84	-2.81	-1.94	-1.01	-0.47	-0.15	0.00	0.50
DXS52	-6.59	-3.62	-1.63	-0.87	-0.26	-0.06	0.00	0.00	0.50

at the loci DXS1003, SYN1, TIMP, PFC and DXS426 since they occur in a contiguous string along the chromosome. Deletion mutation was excluded by amplification and detection of the appropriate products in the individuals tested. Recombinants (III-11, III-12 and IV-13) detected at STR loci (44, 45, 49 and 50) within introns of the dystrophin gene define the new distal limit to the MRX12 gene with STR44 being the most proximal, while recombinants at ALAS2 (IV-3 and IV-13) define the new proximal limit (Appendix II, Paper 7). Refinement of the MRX12 gene interval excludes the q arm altogether and spans over 35cM of the p arm. Multipoint analysis was performed, but since all meioses were informative at five equally spaced loci, the gene is equally likely to be found at any point between flanking markers (not shown). Since the localisation spans the 5' end of the dystrophin gene, the dystrophin brain promoter was PCR amplified in affected individuals, but no deletion or conformational polymorphism was detected (section 4.6.2.1.).

The commonly made assumption of equal allele frequencies (Chapter 2.6.2.1.) has previously been shown to lead to false positive evidence for linkage or biased estimation of the recombination fraction (Ott, 1992). The effects of such inaccuracy are suggested to be most pronounced in small pedigrees, however correct detection of true linkage ought not be affected (Freimer et al, 1993). Allele sizes were determined at four loci to investigate the contribution of correct allele frequencies to the lod score. At DYSII and MAOB, where the genotype of the matriarch could be inferred unequivocally, the frequencies were irrelevant as expected, but at DYSI and DXS7 the correct allele frequencies (Table 4.6) increased the peak lod score from that gained using equal 1/n frequencies ($z=2.71$ and $z=2.01$ at $\theta=0.0$ respectively). If 1/n were to increase false positive linkage as suggested, using the correct allele frequencies should have lowered the lod scores. The direction and magnitude of the effect on the lod score depends on whether the disease gene is associated with a common or rare marker allele. One may conclude that the effect of allele frequencies on the lod score is highly dependent on the family structure and that, at least in Family M, the lod scores are conservative rather than artificially high because of using equal allele frequencies.

One of the women of unknown carrier status (IV-6) requested prenatal diagnosis in early 1992. The broader gene localisation available at the time showed her to be a recombinant within the gene interval. Although her genotypes were that of a non-carrier at DYS and DXS538, they were consistent with the carrier state at DXS255 and DXS453. This evidence of a crossover within the interval prevented the prediction of risk since the exact location of the MRX12 gene in relation to the crossover point was not known. Reduction of the localisation (Appendix II, Paper 7) has not improved informativeness of the diagnosis for this woman since the crossover remains within the gene interval. The crossover is known

to have occurred between MAOB and DXS1126, since genotypes of the latter are also consistent with her being a carrier. Further reduction at the proximal limit to eliminate the crossover from the gene interval is not possible since DXS255 is fully informative in all meioses contributing to gene mapping. Two more women from whom DNA was collected, IV-12 and IV-14, may benefit from risk estimation based on the gene mapping data.

MRX 13 (Family O'N)

Hyperactivity and severe behavioural problems in childhood were described in two of the affected males in family O'N. The MRX13 gene was localised to a broad interval between DXS85 (Xp22.3) and DXS456 (Xq21.22) in 1991, largely by RFLP and early PCR formatted microsatellite markers (Appendix II, Paper 3). The participating family was only just large enough to detect linkage, with three affected males having no normal brothers in two generations (Figure 1d; Appendix II, Paper 3). Two manifesting females, IV-7 and IV-10, were included for pairwise lod score analysis under the assumption of incomplete penetrance at 0.75. The status of their mother, III-5, was accordingly that of obligate carrier. Consideration of their contribution to delimiting the localisation by recombination events should be mindful that their mild retardation may be due to an unrelated cause. Genotypes from six females of unknown carrier status and a normal male sibling (II-3) to the matriarch (II-1) were used to infer the most likely genotypes of the parents in generation I where possible. Non-paternity was not an evident issue.

Attempted further reduction to the gene localisation required that numerous polymorphic loci genotyped within the interval be fully informative. The most proximal recombinant on the short arm was detected in IV-9, the normal male son of III-5, at DXS989. Both IV-9 and III-2 were recombinants at the more distal loci DXS365, DXS999 and DXS987. Individual III-5 was recombinant at DXS228 and at several loci in the interval to DXS426. Pericentromeric markers DXS573, DXS255 and ALAS2 were uninformative, but markers at AR, PGK1P1, DXS453, DXS566, DXYS1X and DXS3 no longer detected recombination near the centromere on the long arm. At the DXS178, DXS101, DXS1120, DXS456 and COL4A5 loci, III-5, was again recombinant. More distal markers were uninformative until DXS1114, DXS51 and DXS105 detected recombination events in the affected male DNA samples.

The gene in this family has been localised by bimodal peaks of lod score maxima, each greater than +2, separated by a large antimodal region where recombination was detected. If the mental retardation in the two females were not assumed to arise from the same mutation as in the affected males, the supposed recombination observed in their

mother would not occur. In such a case, however there are insufficient meiotic events to map the MRX gene to any interval smaller than between DXS365 to DXS1114 where recombination events were observed in males. The lod score would not reach the threshold of +2 required for evidence of X-linkage. Screening for deletion or point mutation in the dystrophin brain promoter, did not reveal any abnormalities (section 4.6.2.1.).

MRX 17 (Family R)

The MRX17 gene for moderate intellectual handicap segregating in family R, has been localised between recombinants at DXS255 on Xp and DXS990 on Xq (Appendix II, Paper 4). The pedigree and clinical evaluations were given in the publication (Figure 1 and Table I; Appendix II, Paper 4); speech delay and brachycephaly together with minor facial anomalies including large mouth and thick lips were described. Significant linkage was recorded at the AR, DXS1125 and DXS571 loci and multipoint analysis against ten loci demonstrated that MRX17 maps to an interval of 30cM with a peak location score between AR and DXS1125 (Table III and Figure 12; Appendix II, Paper 4). Eleven further dinucleotide repeat markers in the region have been genotyped, however seven of these (DXS573, DXS1126, ALAS2, PGK1P1, DXYS1X, DXS178 and DXS1001) were uninformative. The *Taq*I RFLP at DXS1 near the centromere on Xq was also uninformative. Pairwise lod scores at 16 loci encompassing the original regional localisation are given in Table 4.7.

Additional information on the genetic background map (Chapter 3) has modified the MRX17 gene interval since publication of the original localisation against an early genetic map (Appendix II, Paper 4). The final published version of the CEPH background map reduced the map distance between the markers DXS255 and DXS990, conservatively flanking the MRX17 gene localisation, from nearly 30cM to 22.5cM (Appendix I, Paper 3). The consensus physical map places DXS106 proximal to DXS453 and distal to PGK1P1 (XCW5) into which interval DXS1125 has been placed on the genetic map (Chapter 3.5.). Refined mapping of the markers DXS559 and DXS566 (the latter detecting the same repeat as DXS441) has placed both of these markers between DXS453 and DXS986 in the order: cen-AR-PGK1P1-DXS106/DXS1125-DXS453/DXS559-DXS227-DXS566/441-DXS986-DXYS1X-DXS3-tel (Lindsay et al, 1993; Schlessinger et al, 1993; Chapter 3). The DXS571 locus lies distal to the COL4A5 locus at Xq22.3, on the physical map (Willard et al, 1994). These findings contribute to the more accurate localisation of the gene for MRX17, since the recombinant (II-2) detected at DXS990, DXS559 and DXS566 is now also found at DXS227 and at DXS106 the closest marker flanking the interval in Xq (Appendix II, Paper

TABLE 4.7: PAIRWISE ANALYSIS BETWEEN MRX17(R) AND MARKERS ON THE X CHROMOSOME

	θ								
Loci	0.001	0.01	0.05	0.1	0.2	0.3	0.4	Zmax	θ
SYN1	-0.60	0.37	0.93	1.04	0.93	0.65	0.29	1.04	0.12
DXS426	-0.60	0.37	0.93	1.04	0.93	0.65	0.29	1.04	0.12
DXS255	-0.60	0.37	0.93	1.04	0.93	0.65	0.29	1.04	0.12
DXS991	0.88	0.87	0.83	0.76	0.61	0.44	0.23	0.88	0.00
AR	2.40	2.37	2.21	2.00	1.54	1.02	0.47	2.41	0.00
DXS1125	2.10	2.07	1.93	1.74	1.33	0.87	0.39	2.11	0.00
DXS106	-0.60	0.37	0.93	1.04	0.94	0.68	0.35	1.05	0.11
DXS453	0.88	0.87	0.82	0.76	0.61	0.44	0.23	0.88	0.00
DXS559	-0.60	0.37	0.93	1.04	0.94	0.68	0.35	1.05	0.11
DXS227	-0.60	0.37	0.93	1.04	0.94	0.68	0.35	1.05	0.11
DXS566	-0.60	0.37	0.93	1.04	0.94	0.68	0.35	1.05	0.11
DXS986	1.17	1.16	1.10	1.01	0.81	0.58	0.31	1.17	0.00
DXS995	1.17	1.16	1.10	1.01	0.81	0.58	0.31	1.17	0.00
DXS3	1.17	1.16	1.09	1.00	0.80	0.57	0.31	1.17	0.00
DXS990	-0.90	0.07	0.63	0.75	0.67	0.47	0.24	0.75	0.12
DXS1120	-0.90	0.07	0.63	0.75	0.67	0.47	0.24	0.75	0.12

7). There were no recombination events at DXS1125, a locus in close proximity to DXS106. This suggests the order cen-DXS1125-DXS106-tel for this pair of markers. The Xp boundary remains defined by recombination in III-1 at DXS255. Recombinants could not be detected at DXS991 on the proximal short arm, and at a number of loci distal to DXS106 (DXS453, DXS986, DXS995 and DXS3) on the long arm, because phase could not be determined when the genotype of I-2 could not be inferred from her offspring, or was uninformative. The consistency of detectable recombinant events however, at informative markers outside the closest flanking markers, provide reliable evidence in support of this localisation.

Multipoint analysis was not repeated in this family as DXS106 was positioned equivalent to DXS1125 and physical evidence (Chapter 3.5.) placed DXS559, DXS227 and DXS566 in the next interval between DXS453 and DXS986. Since DXS559 to DXS566 were the more distal loci detecting the same recombinant as DXS106, it would be expected that the shape of the multipoint curve would be altered in width only. Manipulations of the background map distances in multipoint analyses in a large single family, have demonstrated that the shape of the curve is not affected, in that the height remains the same and only the width is stretched or contracted to suit (Suthers and Wilson, 1990). The same effect is expected from the refinements to the CEPH background map and the addition of the recombinant detected at DXS106, DXS559 and DXS227. Recombination at DXS106, just distal to DXS1125, would alter the shape of the curve on the right hand side to drop away towards minus infinity between DXS1125 and DXS453 thereby reducing the width of the multipoint curve. The localisation of MRX17, between DXS255 and DXS106, has been reduced to less than a maximum of 10.9cM (DXS255 to DXS453). This reduction has greatest significance in application to risk calculations based on the mapping data, but also reduces the physical interval between flanking markers to be searched for candidate genes and expressed sequences. Another MRX gene has been localised in a family with severe retardation and aphasia as well as brachycephaly and facial anomalies, to Xp11-q11 (Wilson et al, 1992). Similarities in these conditions bear consideration when candidate genes for MRX17 are screened.

MRX 18 (Family K)

Localisation of the MRX18 gene segregating in this family was published recently together with that of MRX17 (Appendix II, Paper 4). The pedigree is given in Figure 6 of this publication in Appendix II. Again, alterations to map distances in the CEPH background map have had a minor effect on estimation of the region of interest. The resultant negligible

difference reduces the estimated span of MRX18 to 26.9cM, instead of the 28.3cM reported. It must be noted however that these distances represent a snapshot in time and are dependent on the method and numbers of markers included into the construction of the background map. They remain subject to change with refinement of the map of the X chromosome. Risk figures calculated from these localisations for genetic counselling purposes, should be based on the most conservative estimates of the interval spanned.

Clinical findings include a spectrum of mental retardation segregating within this family, ranging from mild to severe, with at least two females requiring special schooling. Other variable features include brachycephaly and seizures in IV-4, IV-5 and IV-7, but not in III-9 or the mildly affected IV-8. Only III-9 and IV-4 with severe handicap, had difficult or aggressive behaviour. All four individuals examined in generation IV have limited elbow extension laterally, which is present in and may have been inherited from their father III-4, as this was not evident in III-9. The mildly retarded male V-6, had no speech by the age of two and is considered to be affected. His disease status confirms the non-penetrant carrier status of his mother as determined from haplotype analysis although no DNA was available from him to confirm his genotypes.

Few further loci have been examined within the gene interval, since in the interval between 5'DMD and DXS1126, nine loci (DXS84, OTC, DXS1068, DXS228, MAOA, MAOA3, SYNI, TIMP and DXS426) were uninformative. The CYBB locus physically located between the 5' end of the dystrophin gene and DXS1068 proximally, was the only informative marker but is not on the genetic map. The current distal limit to the gene localisation is defined by recombination in IV-7 and IV-9 at the DXS538 locus. This locus has been genetically mapped in the same interval as CYBB, but the relative order of the pair to one another has not been determined. CYBB detects the same recombinants as DXS538 and generates the same pairwise lod score against MRX18 (Table 4.8). Proximally the DXS573 locus has recently been mapped between DXS722 and DXS255 (Thiselton et al, 1995) in which interval the DXS1126 and SYP loci have also been physically placed (Willard et al, 1994). The relative order of DXS573 and DXS1126 is not currently established therefore the recombinant IV-8 defines the proximal limit to the MRX18 gene localisation at either of these loci. Two minor typographical errors are noted in the published table of two-point lod scores (Appendix II, Paper 4), at DXS538 z=-3.60 at θ=0.001 and at DXS7 zmax=2.11 at θ=0.0. Corrected pairwise lod scores between MRX18 and 11 informative markers are given in Table 4.8. with penetrance set at 0.75. Multipoint analyses were not repeated since the genetic location of CYBB is not known and that of DXS573 has not been determined in CEPH families (Thiselton et al, 1995).

TABLE 4.8 : PAIRWISE ANALYSIS BETWEEN MRX18(K) AND MARKERS ON THE X CHROMOSOME

	θ								
Loci	0.001	0.01	0.05	0.1	0.2	0.3	0.4	Zmax	θ
5'DMD	-3.59	-1.62	-0.35	0.09	0.34	0.31	0.17	0.35	0.23
CYBB	-3.60	-1.63	-0.37	0.04	0.24	0.18	0.05	0.24	0.21
DXS538	-3.60	-1.63	-0.37	0.04	0.24	0.18	0.05	0.24	0.21
DXS7	2.10	2.08	1.95	1.79	1.43	1.02	0.55	2.11	0.00
MAOB	2.40	2.36	2.19	1.95	1.45	0.89	0.32	2.41	0.00
DXS1003	2.10	2.08	1.95	1.79	1.43	1.02	0.55	2.11	0.00
PFC	2.70	2.67	2.49	2.25	1.75	1.19	0.60	2.71	0.00
DXS1126	-0.29	0.67	1.21	1.30	1.14	0.83	0.43	1.30	0.10
DXS573	-0.60	0.37	0.93	1.04	0.94	0.68	0.35	1.05	0.11
DXS255	-0.29	0.67	1.21	1.30	1.14	0.83	0.43	1.30	0.10
ALAS2	-0.29	0.67	1.21	1.30	1.14	0.83	0.43	1.30	0.10

MRX 19 (Family ML)

The gene in this family was localised by A. Donnelly, WCH (Appendix II, Paper 5), using the established library of markers and genotyping strategies as described in Chapter 3. Previously studied by Choo (1984), affected males had moderate retardation with no distinctive characteristics. Mild retardation was described in two females, one of them an obligate carrier.

The MRX19 gene was initially localised to Xp22 between the KAL and DXS989 loci (Appendix II, Paper 5), and was recently redefined between DXS1043 and DXS1052 (Appendix II, Paper 7). The localisation differentiates this MRX locus from that described in the contiguous gene syndromes in the distal short arm (Ballabio and Andria, 1992). Although MRX19 overlaps with the MRX2 gene between DXS999 and DXS989 (Figure 1 and Table 1; Appendix II, Paper 7) and with the MRX21 gene (Kozak et al, 1993), the clinical findings in MRX2 and MRX19 have not been reviewed since 1984. These overlapping localisations represent at least one MRX locus in Xp22.

MRX 27 (Family GLA 2617)

This MRX family (GLA 2617) was first described in 1991 (Glass et al, 1991b). Delayed language acquisition and behaviour difficulties were the only associated manifestations to the mild to moderate mental retardation in affected males. Two of the carrier females (III-4 and IV-3) had borderline intellectual handicap and required special schooling. Linkage analysis with 14 RFLP loci spanning the X chromosome suggested linkage of the MRX gene segregating in this family to Xq26-27 with a maximum lod score of 1.6 ($\theta = 0.10$) at the F9 gene locus. The fragile-X was excluded on chromosome analysis (Glass et al, 1991b) which retrospectively excludes expression of the FRAXE site in Xq28 also (section 4.3.1.). The localisation for this MRX gene has now been defined using dinucleotide repeat markers and has been designated MRX27 (Appendix II, Paper 6).

DNA samples from 24 members of the family (Figure 1; Appendix II, Paper 6) were genotyped at several microsatellite markers within an interval up to 50cM either side from F9. A refined genetic background map prepared from the CEPH database (Appendix III, Paper 7) established the order and distance (in cM) between loci in the region as COL4A5-20.0-DXS424,DXS1001-14.6-DXS994,DXS692-4.3-DXS1114,HPRT-7.3-DXS102-1.7-F9-18.1-DXS292-4.4-DXS297. Two point lod scores were generated at ten loci spanning Xq22.3-q27.3 (Table 4.9). Peak pairwise lod scores of 4.02 and 4.01 ($\theta = 0.00$) were calculated between MRX27 and DXS1114 and DXS994 loci respectively, under the

TABLE 4.9 : PAIRWISE ANALYSIS BETWEEN MRX27 vs MARKER LOCI SPANNING Xq22.3-q27.4

	θ								
Loci	0.001	0.01	0.05	0.1	0.2	0.3	0.4	Z max	θ
COL4A5	-6.22	-3.26	-1.30	-0.57	-0.01	0.16	0.15	0.17	0.35
DXS424	0.20	1.16	1.66	1.70	1.42	0.95	0.43	1.71	0.09
DXS1001	1.07	2.02	2.46	2.42	1.99	1.35	0.65	2.47	0.07
DXS994	4.00	3.94	3.66	3.31	2.55	1.71	0.81	4.01	0.00
DXS692	2.42	2.39	2.23	2.03	1.61	1.14	0.52	2.43	0.00
DXS1114	4.01	3.95	3.66	3.28	2.47	1.58	0.62	4.02	0.00
HPRT	2.44	2.40	2.26	2.06	1.64	1.18	0.64	2.44	0.00
DXS102	-0.56	0.41	0.98	1.11	1.04	0.81	0.46	1.12	0.13
DXS292	-2.96	-0.99	0.26	0.66	0.84	0.72	0.44	0.84	0.20
DXS297	-8.63	-4.65	-1.96	-0.89	-0.02	0.27	0.25	0.29	0.35

assumption of zero penetrance in carrier females. The markers at DXS1122, DXS300, DXS294 and DXS984 loci in this region were uninformative. The MRX27 gene was localised to Xq24-26 in an interval defined by recombination events detected in IV-5 at DXS102 distally, and in the affected male IV-9 at DXS424 proximally. The recombination detected in IV-9 was sustained at the more proximal locus COL4A5 where IV-7 and IV-8 were additionally recombinants. Recombination detected at DXS1001 in a female of unknown carrier status, III-7, with two normal sons, may affect the determination of her carrier risk since the genetic background map places DXS1001 in the immediate vicinity of DXS424 with 1:1 odds for inversion. No male recombinants were detected at the DXS1001 and DXS994 loci placing them both distal to DXS424. The MRX27 gene lies within an interval spanning 26.4cM and does not overlap with any other known MRX gene except perhaps the ill-defined MRX6 for which no boundaries have ever been formally described (Kondo et al, 1991). Genetic localisation of the MRX27 gene defines a distinct MRX gene interval on the X chromosome (Appendix II, Paper 7).

An RFLP marker, DXS51, within this interval was previously reported with high negative lod scores consistent with exclusion of linkage at this marker (Glass et al, 1991b). Recombinants were inconsistent with haplotype analysis and recombinants defining flanking markers. Sample confusion uncovered by the microsatellite polymorphisms, may account for possible errors in genotyping at this and possibly other RFLP loci which were largely uninformative and therefore left out of the current analyses.

4.3.3. Corollary

The gene localisations in the ten MRX families described above confirm and extend the genetic heterogeneity of MRX. The genes at MRX3 (Appendix II, Paper 1), MRX19 (Appendix II, Paper 5) and MRX27 (Appendix II, Paper 6) contribute three clearly distinct loci towards enumerating the genes for MRX (section 4.4) (Appendix II, Paper 7). The clustering of loci at the centromere had been noted by 1992, however refined localisation of MRX1 delineated this locus from MRX10 and MRX11 by excluding a region of overlap and thereby dividing the apparent pericentromeric cluster (Appendix II, Paper 3). Mapping of MRX17 in an interval distinct from MRX18 further reinforced the conclusion that there are at least two discrete pericentromeric genes for MRX (Appendix II, Paper 4).

Under the assumption that manifesting females in Family Q carry the same mutation, designated MRX10, as do their affected male relatives, recombination detected in an affected female redefines and splits MRX10 from overlap with MRX11. For the enumeration of MRX loci, the potential refinement of these genes differentiates an additional locus responsible for MRX. Given the reservations for accepting the assumed

homogeneity of retardation, if MRX10 were regarded as defining a distinct MRX locus, the regional localisations within this cohort of families permit the conclusion that there are six discrete MRX genes represented by the loci MRX1, MRX3, MRX10, MRX11, MRX19 and MRX27. Conservative localisation of MRX10, particularly for the purposes of risk estimation, was not based on this assumption however, since all obligate heterozygotes were intellectually normal. Therefore, at least five MRX loci have been delineated here.

The genetic heterogeneity of MRX has been demonstrated through mapping of these loci by linkage to different chromosomal regions. The linkage approach necessitates 'splitting' of the condition since each family must be considered individually given the known heterogeneity. Overlapping localisations may however reflect allelic mutations in one or several genes involved in brain development and function. Speculative 'lumping' of overlapping genes, for example those defined by MRX11, MRX12, MRX13 and MRX18 which may represent a single locus, serves little purpose until mutations in candidate genes are to be examined. Identification of clinical correlates within such a subgroup (section 4.5.) may suggest 'soft' syndromal features to define specific entities. Clinical grounds for lumping and splitting of MRX are by definition not convincing, but minor features or qualities may ultimately lead to recognition of distinguishing traits or implicate the shortest region of overlap within which to investigate candidate genes.

Further refinements to 'split' overlapping localisations and enumerate MRX genes are limited by family sizes and structures as well as the availability of untested polymorphic loci within the gene intervals. A finite set of polymorphic markers (Chapter 3) were applied to the families above, such that these localisations have reached the limits of resolution within these parametric confines. Saturation mapping through genotyping of the available microsatellite markers within each gene interval did not detect microdeletion, however this mode of mutation cannot be dismissed from consideration as a cause of mental retardation in any of these families.

4.4. Enumerating MRX genes

Efforts in determining the molecular basis of X-linked mental retardation elsewhere have also focussed on gene mapping in large families with MRX. In addition to the ten MRX families above, the genes in twenty-one more families with MRX have been localised by independent linkage studies to markers on the X chromosome. The genetic localisations in all MRX families large enough to fill the criteria for MRX number assignment have been summarised (Table 1; Appendix II, Paper 7). Of the 32 MRX numbers assigned to date however, data are incomplete or not yet fully published for some (MRX6, 23, 26, 28, 29 and 32). Further families have been reported in brief with lod scores greater than or equal to +2,

but have not yet been assigned MRX numbers (Charlton et al, 1994; Moraine et al, 1994; Schwartz et al, 1994). Some large clinically described MRX families, have not been analysed for linkage (Yarbrough and Howard-Peebles, 1976) or not been accepted as 'non-specific' (Wilson et al, 1992). Representation of 31 known MRX loci against markers placed on an X chromosome ideogram give a diagrammatic impression of the distribution of MRX genes on the chromosome (Figure 1; Appendix II, Paper 7). MRX loci appear to cluster around the centromere and proximal short arm of the X chromosome; over 40% (13 of 31) cross the centromere, while nearly 50% (15 of 31) cross an interval between DXS255 and DXS988 in Xp11.22-p11.23. This clustering of entities might indicate a cluster of genes involved in brain function or a few major loci expressed at high frequency.

The minimum number of discrete MRX loci are determined from the maximum number of non-overlapping genetic localisations. Reductions to previously published localisations based on RFLP markers (eg. MRX5) may be possible using PCR-based microsatellite markers spanning the chromosome. In this manner Hu et al (1994) restudied two of the families described by Arveiler et al (1988), and defined the limits to the gene localisation in each. Through redefinition of the gene limits in seven families available to us and those described by Hu et al (1994), a more accurate assessment of the number of discrete MRX genes spanning the chromosome has been established (Appendix II, Paper 7). The MRX24 gene near pter is distinct from MRX2 in Xp22, while in Xq28 the MRX3 locus is coincident with MRX16, MRX25 and MRX28. Genetic localisations of MRX27 and MRX30 to discrete intervals in Xq22-q26 define two further distinct MRX loci. Taking the loci with the lowest numbers first MRX1 and MRX10 define the loci around the centromere, though alternative combinations such as MRX4 and MRX11, or MRX11 and MRX14, or indeed MRX17 and MRX18 are equally valid. Conservatively, a minimum of 7 discrete genes (MRX24, MRX2, MRX10, MRX1, MRX30, MRX27 and MRX3 from pter to qter) have been defined by linkage mapping (Figure 1; Appendix II, Paper 7).

The locations of putative MRX loci have been suggested through physical events such as X:autosome translocations in mentally impaired girls and small deletions in retarded boys. Within the 10Mb of Xpter, a gene for MRX, has been identified in patients with complex disease entities known as contiguous gene syndromes caused by deletions, translocations or duplications of the region (Petit et al, 1990; Ballabio and Andria, 1992; Herrell et al, 1995). This gene, herein designated MRXC1, in Xp22.3 has been listed under the MIM #309530 (McKusick, 1994; Ballabio and Andria, 1992), which number describes MRX1, so is an erroneous reference to the gene location. Analysis of male patients with interstitial deletions in Xp22.3, show that the MRXC1 gene lies distal to the steroid sulphatase gene (STS) and to a breakpoint between the marker 40BT and the GS1 gene

(Herrell et al, 1995), while MRX24 maps proximal to DXS278 locus (Martinez et al, 1995). The complex polymorphism detected by CRI-S232 (DXS278) involves four copies of homologous sequences partly distal and partly proximal to STS (Lee et al, 1993), therefore MRX24 cannot be clearly differentiated from the MRX locus in MRXC1 patients. Similarly MRX21 with no recombinants in p22.3, cannot be distinguished from the localisation of MRXC1 (Kozak et al, 1993) and therefore cannot be concluded to be the same, or a different gene. Chromosome walking from a YAC contig around STS, towards the telomere and beyond GS1, may help in the isolation of the adjacent MRXC1 gene and position it with respect to the S232 related sequences (Carrozzo et al, 1992).

Deletion mapping in patients with contiguous gene syndromes has defined another MRX gene, herein designated MRXC2, on the proximal long arm in close association with deafness (DFN3), choroideremia (CHM), cleft lip and palate and immunodeficiency in various combinations (Cremers et al, 1989). These genes and three microsatellites DXS986, DXS995 and DXS1002 have been ordered against deletion cases in Xq21 (Clark et al, 1994), with DXS995 found to be associated with X-linked mixed deafness (DFN3) with or without other features (Huber et al, 1994). The putative MRX gene involved in the complex phenotype has been located by phenotype analysis of patients with deletions to the same interval as DFN3 (Bach et al, 1992). Recent refinement of the MRX critical region in Xq21 proposes ZNF6, a zinc finger transcription factor highly expressed in fetal brain, as a candidate gene just proximal to CHM (May et al, 1995). None of the seven MRX loci defined through linkage studies in single families crosses this interval, therefore physical identification of this locus reveals a further MRX gene.

The breakpoint of a balanced translocation t(X:13) in a female with profound mental retardation and minor anomalies is located in Xq13.1. The phenotype is presumed to arise due to disruption of the putative disease gene at the site of the translocation breakpoint on the X in subsequent inactivation of the normal X. The t(X:13) breakpoint has been cloned and a candidate gene (DXS6673E) predominantly expressed in brain encodes a 7.5kb message showing no homology with others in databases (Van der Maarel et al, 1994). This new gene may have a role in the etiology of MRX genes in families mapping across Xq13. The exact molecular location of the breakpoint has not yet been published, however MRX1, representing the locus nearest to the centromere, does not extend distal to Xq12 nor therefore impinge on this putative MRX gene.

The FRAXE fragile site in Xq28 is proximal to and distinct from the MRX3 localisation although the phenotype is clinically non-specific with primary manifestations of learning problems. The discovery of CCGn repeat expansion in families with FRAXE mental retardation has permitted direct detection of the mutation in families with mild

mental handicap (Knight et al, 1993; Knight et al, 1994). The possibility of phenocopies confound establishment of a clear phenotype-genotype relationship (Appendix V, Paper 6). Families with FRAXE add another MRX locus distinct from MRX3 in Xq28 (Appendix II, Paper 7). Developmental delay due to deletion of a region adjacent to FRAXE (Chapter 7) has been shown to contain elements of the FMR2 gene associated with FRAXE (Appendix V, Papers 7 and 8). A total of ten discrete MRX loci have been ascertained from the data presently available, with eight of these known to be familial.

4.5. Clinical correlates

Since by definition MRX families have no consistent clinical findings, even the variable parameters within and between families are often overlooked. Once gene locations are established it may be possible to find subtle clinical features more common in a subgroup with overlapping localisations. Correlation of clinical findings associated with MRX genes already related by position, may enable diagnosis of new syndromes in the same manner in which the Renpenning and fragile X syndromes have been recognised (Appendix II, Paper 2). A summary of the clinical findings reported in 26 families are tabulated in Table 4.10. Few of the criteria established in Appendix II, Paper 2 had been examined in each family. Many families have been reported with scanty information, while some published in abstract form or not at all, lack clinical details (MRX20, MRX23, MRX26, MRX28, MRX29 and MRX32). Even the ethnic backgrounds of MRX families were not always reported although there was much variation including Japanese, Spanish and German origins permitting no suggestion that a significant proportion were of similar ethnic origin.

Speech delay was noted in twelve families (Table 4.10), gene assignment in eight of these (MRX4, MRX5, MRX12, MRX13, MRX14, MRX17, MRX18, MRX22) was close to or across the centromere. Language delay may be due to mental retardation, aphasia or deafness and requires thorough assessment of auditory function to exclude deafness (Gordon, 1993). Hearing tests were normal in MRX14 and MRX22 and audition was apparently normal in the other families although this was not specifically stated. Another family with severe mental retardation and aphasia has also been mapped to Xp11-q11, though affected males were brachycephalic with minor facial anomalies, growth failure, frequent infections and seizures (Wilson et al, 1992). All three patients described by Wilson (1992) had 'normal' hearing, and the degree of speech problems and cognition were dissociated, implying a specific problem with language (Wilson, 1993). Although speech impairment is regarded as a common and non-specific feature in mentally retarded individuals, such strong correlation with gene location may suggest clustering of genes required for development of the language centres in the brain and needs be considered as a

Table 4.10. Summary of variable clinical findings in MRX.

MRX locus*	Mental Retardation	Speech delay	Motor delay	Behavioural	Facies and/or features	Seizures	Manifesting heterozygotes
MRX1	moderate	-	-	-	-	-	-
MRX2	mild-moderate	+	-	-	+	-	+
MRX3	mild-moderate	-	-	+	-	-	-
MRX4	moderate	+	+	-	-	-	-
MRX5	moderate	+	-	+	-	-	-
MRX6	moderate	+	-	-	+	-	+
MRX7	mild	-	-	+/-	-	-	-
MRX8	mild-severe	-	-	-	+	-	-
MRX9	moderate-severe	-	+	-	+	-	-
MRX10	mild-severe	+/-	+/-	-	-	-	+
MRX11	moderate	+/-	+/-	-	-	-	+
MRX12	mild-moderate	+	+	-	+	-	-
MRX13	mild-moderate	+	-	+	-	-	+
MRX14	moderate-severe	+	+	-	+	+	+
MRX15	severe	-	-	+	-	-	-
MRX16	moderate	-	-	-	-	-	-
MRX17	moderate	+	-	-	-	-	-
MRX18	mild/mod/severe	+/-	-	+/-	-	+/-	+
MRX19	moderate	-	-	-	-	-	+
MRX21	moderate	-	-	-	-	-	+
MRX22	severe	+	+	-	-	-	-
MRX24	moderate-severe	-	-	+/-	+	-	+
MRX25	mod-profound	-	-	-	-	-	-
MRX27	mild-moderate	-	-	-	-	-	-
MRX30	mild	-	-	-	+	-	-
MRX31	moderate	-	-	-	-	-	-

*no clinical information available on MRX20, MRX23, MRX26, MRX28, MRX29 and MRX32.

specific trait indicating lesion of the central nervous system. Motor delay occurred in only half as many families and was mostly concurrent with speech delay. Genes mapping to p21-p22 appear to be variable in the extent of associated mental retardation however several families have mildly affected males and a proportion of manifesting females. The occurrence of manifesting heterozygotes in ten families (Table 4.10) is at a greater fraction than might be expected and are most apparent in MRX loci mapped to Xp21 and more distally. All families with localisations extending into Xp22 (MRX2, MRX13, MRX19, MRX21 and MRX24) comprise mildly expressing unexplained retardation in obligate carriers and sometimes other females possibly due to unfavourable inactivation of the wild type gene. A cluster of genes in the Xp22.3 region are known to escape inactivation (Mohandas et al, 1979; Goodfellow et al, 1984), although other evidence suggests that inactivation may be regulated in a locus-autonomous manner (Brown and Willard, 1990). Skewed inactivation may be a mechanism that exacerbates MRX gene expression in females segregating MRX10, MRX11, MRX18 and MRX19, rather than partial dominance of a specific gene.

Within the cluster of overlapping pericentromeric loci, those defined by MRX1, MRX4, MRX5, MRX9, MRX13, MRX14, MRX17 and MRX31 have moderate retardation. Of these MRX4, MRX5 and MRX13 have speech delay, MRX14 has a lack of development after normal early acquisition; and MRX4, MRX9 and MRX14 have psychomotor delay. Of the eight families described with severe retardation in at least one male, all but MRX10 and MRX24 share the common region of overlap in p11.2-p11.4. The gene causing the severe MRX of MRX22 (with severe hypotonia without cerebellar abnormalities), also overlaps with the cluster mapped to Xp11-cen with mild to moderate phenotypes (Passos-Bueno et al, 1993) as well as with the gene in a family segregating severe mental retardation with aphasia and facial dysmorphism (Wilson et al, 1992). This Brazilian family may represent intragenic or locus heterogeneity of MRX genes at the centromere. The MRX7, MRX8 and MRX12 loci are also pericentromeric but with mild retardation. Although MRX8 has some features similar to the Renpenning phenotype and subsequent mapping of the original Renpenning pedigree and another family (K8240) with clinical similarities all map in the pericentromeric region, heterogeneity cannot be excluded. A recombinant with the AR locus in the Renpenning family may separate this disorder from that in K8240 and possibly in MRX8 (Schwartz et al, 1992 and 1994).

Of the six families described with behavioural problems (Table 4.10), the loci at MRX5, MRX7, MRX13 and MRX18 share a large overlapping interval. Some of the affected males in each of these families demonstrated difficult or deviant behaviour compounded by aggression and hyperactivity. The common localisation spans the

monoamine oxidase (MAOA) gene known to be responsible for borderline mental retardation (IQ of ~85) associated with a behavioural disturbance dominated by aggression (Brunner et al, 1993a). Affected males had a point mutation in the MAOA gene causing isolated MAOA deficiency (Brunner et al, 1993b). Abnormal MAOA activity has been implicated in subtypes of familial alcoholism with antisocial personality disorder, in bipolar affective disorder, depression and schizophrenia (Devor et al, 1993; Lim et al, 1994; Hinds et al, 1992). The relationship between lowered MAO activity and family history of psychiatric illness had been documented as early as 1977 (Devor et al, 1993). Allelic mutation in this gene may account for one or all of these MRX families (see section 4.6.1.). The family defining MRX3 in Xq28 had 4 of 5 males exhibiting aggressive behaviour, while in the overlapping loci MRX16 and MRX25 there was no mention of aggression or adaptive behaviour. The family described by Brunner et al, (1993) is of particular importance as an example of the value of gene localisation in determining a candidate gene by position, but also of the importance of clinical descriptions of behavioural phenomena because it was this that lead to MAO- being a suggested candidate.

Divergent phenotypes of Norrie disease (NDP) and retinitis pigmentosa (RP2) also show strong linkage to MAO. Classic NDP have retinal dysplasia suggestive of specific early developmental failure. About two-thirds of NDP patients manifest mental retardation often with dementia or psychosis, while atypical patients have been described with hypogenitalism and/or microcephaly (Sims et al, 1992). A kindred with RP segregating with mild non-specific mental retardation, had 3 of 4 affected males examined having microcephaly (Aldred et al, 1994). The possibility of a contiguous gene syndrome involving an MRX and the RP2 loci has been raised. The differential diagnosis includes the MRX loci in eleven families (MRX5, 7, 8, 9, 12, 13, 15, 18, 20, 22, 26), none of which are associated with signs of RP although some males in the MRX8 kindred have microcephaly (Chapter 5; SHS).

Of the three families described by Arveiler et al, (1988) Family 1 (now MRX2) had been clinically described by Proops (1983), with affected males consistently demonstrating the fragile X phenotype including large heads, prominent ears and macroorchidism but no fragile site. Arveiler demonstrated negative lod scores to exclude all three families from DXS52, the closest marker at the time to the fragile X, and concluded that at least two further loci complicate the analysis of non-specific XLMR. Macroorchidism and large ears, described in MRX2, MRX12 and MRX14, are not unusual in that detailed anthropometric measurements in mentally retarded males, with and without the fragile X, revealed these features as commonly occurring in both groups, however more frequently in the fragile X group (Butler et al, 1991). MRX6 affecteds also had characteristic manifestations when

compared with normal siblings, including short stature, coarse face, stubby hands and toes, and females manifesting mild mental retardation (Kondo et al, 1991). The facial features were likened to those described by Atkin-Flaitz which are also similar to the XLMR with macrocephaly, described in Chapter 5, although head circumferences in MRX6 patients were normal. The mutant gene in the Atkin-Flaitz family has not yet been regionally localised however the locations of MRX6 and XLMR with macrocephaly do not coincide (Chapter 5). Definition of the MRX6 gene interval between flanking markers may differentiate the locus segregating in this family from those mapping in Xq28. Consistent phenotypical findings such as these are now considered specific clinical features that ought to exclude these families from the MRX classification though subsequently they may represent allelic heterogeneity at an MRX gene.

4.6. Candidate genes for MRX

Linkage analyses in 31 MRX families have identified the most likely gene interval in each case. The minimum interval to which genes in single families can be localised is delimited by the size of the family, occurrence of recombination and the informativeness of the markers in the region, but is often physically large. Since no gross chromosomal or DNA rearrangements or deletions have become apparent through all the technical manipulations of patient samples, a number of avenues towards disease gene identification can be followed. Each gene interval can be physically dissected and sequenced, exon trapped, or directly selected for cDNA clones, however these procedures are heavily reliant on time and labour and would be ineffective in the short term for characterising MRX genes given the number and extent of the localisations. The candidate gene approach of direct mutation detection can potentially circumvent the time and difficulty factor.

Most MRX genes localised in single families, are mapped at best to intervals of 10-30cM. Theoretically all the genes mapped into an MRX gene interval are positional candidates for the disease gene. If there are 2500-5000 genes on the 210cM X chromosome (Chapter 1.2.3.) and 1Mb/1cM, by extrapolation, an MRX gene mapped to an interval of say 30cM has a possible 30Mb (30,000kb) or over 700 possible genes to be screened. The smallest gene interval between flanking markers may contain several known genes, however some may be implicated as higher probability candidates for the disease-causing mutation based on position, tissue expression, abundance and function. Given the lack of syndromal features in MRX it is easy to assume that the defective gene acts on a single target tissue, namely brain. Roughly 30,000 genes are expressed in brain (Chapter 1.6.2.), with 20,000 of these being uniquely expressed in brain (Crowe, 1993). Up to 1500 'brain' genes may be on the X chromosome which represents 5% of the genome. Further, if they were randomly

distributed, there would be at least 7 genes per centiMorgan, therefore on average MRX intervals potentially encompass over 200 positional candidates expressed in brain. Narrowing down the number of potential candidates will be vital to approaches to gene identification through mutation detection.

4.6.1. Searching for MRX candidate genes

The cerebral cortex is most associated with cognitive function, while the cerebellum is involved in generating smooth coordinated movements. Most MRX may be considered to be due to a specific defect in the pre-frontal portions of the cerebral hemispheres since motor, somatic and sensory functions are normal. Since there is no evidence that MRX is progressive, this suggests that the disorder is essentially a defect of development, which exists from birth and remains unchanged throughout life. Candidate genes for MRX may be expressed at detectable levels only in specialised areas and/or critically involved at specific times during brain folding and development, but need not necessarily be expressed only in brain (Chapter 1.6.3.). Ultimately classification of genes and ESTs derived from brain cDNA libraries (Chapter 1.6.2.) according to expression in specialised areas or ubiquitously, will aid in the selection of candidates for further analysis.

Several of the MRX localisations in Xp11 overlap sequences either expressed in neuronal tissue or known to affect neuronal function including OTC, MAOA, MAOB and synapsin 1. The MAO structural genes regulate metabolism of neurotransmitters in the brain., the two forms differ in that MAOA has higher affinity to endogenous neurotransmitters, while dietary amines are the preferred substrate for MAOB (Vanyukov et al, 1995). Expression of MAOA and MAOB genes in the brain is developmentally regulated and may have a role in neuronal development (Sims et al, 1992). Complete absence of MAOA due to a point mutation results in a mild disorder of retardation characterised by aggressive behaviour, deletion of both MAOA and MAOB result in severe mental retardation, while isolated disruption of MAOB on the other hand does not appear to affect intelligence or behaviour (Brunner et al, 1993). Three dinucleotide repeats have been described in the introns of MAOA, MAOB and MAOA3, the last alongside a novel 23bp VNTR motif (Chapter 3). An association has been observed between the length of the dinucleotide repeat at MAOA, MAOA enzyme activity and propensity for substance abuse (Vanyukov et al, 1995). Closer examination of polymorphic MAOA activity may construe normal behavioural variation. The identification of a metabolic defect related to behavioural phenomena and mental illness is important as it opens the possibility for therapeutic intervention. If however mental illness is a developmental process, then gene families such as MAO may have a role in the pathogenic process rather than directly in the end-point

phenotype, therefore it is important to study the processes in the brain that involve MAO. It is already known that MAO is the primary enzyme in degradation of neurotransmitters like dopamine, noradrenaline and serotonin as well as a neurotoxin precursor (Konradi et al, 1992). Examination of MAOA activity in MRX families with behavioural disorders and overlapping gene localisation is warranted. Proximally, the synapsin I gene codes for a neuron specific phosphoprotein that may regulate neuro-transmitter release (Kirchgessner et al, 1991). Microsatellites close to and within these candidate genes have been genotyped in several of the families in this study and no unusual findings were discernible. Although deletion mutations were not detected, the disease-causing mutation cannot be excluded at these loci.

A secondary cluster of MRX gene localisations is apparent near Xqter (section 4.4.). The region from FRAXA/FMR1 to qter encompassing Xq28 is physically an estimated 9Mb (XCW5) and contains upwards of 20 identifiable CpG islands (Maestrini et al, 1992). The Xq28 chromosomal band is not uniformly gene rich however, with contigs showing moderate then high and finally low GC and CpG content extending from GABRA3 to G6PD to Xqter (Willard et al, 1994). Over 30 genes have been placed between DXS304 (U6.2) and the telomere (Korn et al, 1994). With a high density of genes in an area there is a greater chance of contiguous gene syndromes such as at Xpter, where by deletion mapping different genes and their order can be assessed and around NDP where MAOA defect leads to Brunner-type MRX but deletion of the entire MAO- locus leads to complex phenotype. Mutations in the neural cell adhesion molecule L1 (L1CAM) in Xq28, have also been associated with several phenotypes; hydrocephalus with stenosis of the aquaduct of Sylvius (HSAS), MASA syndrome (mental retardation, adducted thumbs, shuffling gait, aphasia), and spastic paraplegia (SPG1) (Jouet et al, 1994; Vits et al, 1994; Fransen et al, 1994). Each of these conditions are generally associated with moderate to severe mental retardation. This spectrum of clinical findings, now referred to as CRASH syndrome (a mnemonic for the clinical findings) (Fransen et al, 1995), was originally thought to represent separate entities in a contiguous gene arrangement. Private mutations in this single gene have enabled the 'lumping' of these disorders, now known to be allelic or evidence of variable expression, since all three have occurred in one family (Fryns et al, 1991; Ruiz et al, 1995). L1 has been implicated in neural cell migration during development, mediation of neuronal interaction and regeneration of damaged nerve tissue (Fransen et al, 1995). This gene may be a candidate for the MRX3, MRX16, MRX25 and MRX28 genes mapping to Xq28.

The implication that mental retardation genes are expressed only in the central nervous system is due to the premise of one gene - one phenotype. Ample precedent now exists to demonstrate that allelic mutations in a single gene can cause distinct clinical

phenotypes, some well known examples are; CFTR mutations causing cystic fibrosis and congenital absence of the vas deferens (Gervais et al, 1993); androgen receptor mutations that cause spinal bulbar muscular atrophy and androgen insensitivity (LaSpada et al, 1991); mutations in the RET protooncogene on 10q11.2 causing three phenotypically distinguishable cancer syndromes and Hirschsprung's disease, a noncancerous developmental anomaly (van Heyningen, 1994); other oncogenes are also involved in developmental disorders and cause phenotypic diversity manifesting as distinct clinical disorders (Romeo and McKusick, 1994). Candidate genes for central nervous system disorders may therefore have pleiotropic effects and be expressed in unexpected tissues or be expressed in different tissues at different stages of development.

In the absence of deletions and rearrangements that pinpoint the site of the mutation, refinement of MRX gene localisations will reduce the number of candidate genes within any given interval. The critical region can be considerably narrowed by examining recombination events in females assuming homogeneity within families. Reduction of the MRX10 regional localisation to 6.2cM and possibly to as small an interval as 1.1cM by acceptance of the validity of recombination in affected females (where affection status is difficult to define) has been discussed (section 4.3.2.). Based on estimations of gene number (section 4.6.), screening of positional candidates for MRX10 would involve up to 7 genes expressed in brain out of nearly 25 genes mapping to the smallest interval.

4.6.2. A candidate gene in Xp21

The dystrophin gene in Xp21 encodes a cell membrane protein transcribed in smooth, skeletal and cardiac muscle, and in brain. Duchenne (DMD) and Becker (BMD) are allelic muscular dystrophies caused by large deletions in the dystrophin gene of ~65% of patients (Koenig et al, 1987). The remainder may be due to small mutations, however the entire coding region has not been screened exhaustively for point mutations or small deletions (involving less than an exon) (Roberts et al, 1994). Slight to moderate, non-progressive mental retardation with a mean IQ of 80, is associated with physical handicap in one-third to one-half of all DMD patients (Moser, 1984; Emery, 1987; Bies et al, 1992). All DMD patients are affected at the level of verbal skills (Górecki et al, 1994), while reduced verbal IQ is also found in some cases of BMD (Zatz et al, 1993).

The 2.4Mb dystrophin gene has been very well characterised and several isoforms and promoters have been identified. Transcription of dystrophin in the brain is initiated from a different promoter than that in muscle (Boyce et al, 1991). Functionally distinct alternatively spliced transcripts display tissue and developmental-stage specific expression (Bies et al, 1992). At the 5' end of the gene, and within 400kb of the common second exon,

separate brain and muscle promoters have been described that lead to full length 14kb transcripts (Boyce et al, 1991). Cortical and hippocampal neurons and glial cells express both isoforms (Feener et al, 1989; Górecki et al, 1992).

Altered expression of the dystrophin gene in brain as well as in muscle may be associated with retardation in DMD, however specific defects affecting only brain dystrophin transcription or regulation could cause central nervous system disease without the skeletal muscle abnormalities. By analogy, frequent cardiac involvement in DMD (>80%), may be ascribed to specific dystrophin mutations (Muntoni et al, 1993). The closely linked form of X-linked dilated cardiomyopathy has reduced cardiac dystrophin, with normal skeletal muscle dystrophin, suggesting dysfunction in a cardiac specific promoter (Towbin et al, 1993).

Mental retardation in DMD has been associated with specific patterns of deletion detected with the cDNA probe cf56a in exon 51 (Rappaport et al, 1991), though no exons were deleted in chronic schizophrenia-like psychiatric disorders (Zatz et al, 1993; Lindor et al, 1994). The specific distribution of dystrophin at the cerebral cortical synapses is in keeping with the possibility that a defect may produce a cognitive deficit (Lidov et al, 1990). The dystrophin brain promoter was a conceivable candidate for those MRX genes mapping near to the 5' end or across the dystrophin gene. Promoters are not translated and they are not always and not completely transcribed. A mutation within the promoter, whether in the transcribed region or not, may affect the regulatory function or the binding recognition sequence of a transcription factor. It was reasoned that even a single point mutation in the promoter may be critical in producing inadequate levels of the brain isoform without affecting skeletal muscle dystrophin.

4.6.2.1. Mutation detection

The dystrophin brain promoter (DBP) region was PCR amplified using the published primers, DBP- F: 5'- GAA GAT CTA TAT TTT ACA ACG CAG AAA TGT GG -3' and DBP- R: 5'- CTT CCA TGC CAG CTG TTT TTC CTG TCA CTC -3' (Boyce et al, 1991), that yielded a 630bp fragment (Hugnot et al, 1993). At least one affected male from each of the families Q, H, M, O'N (section 4.3.2.) and E (Figure 10 and Table V; Appendix II, Paper 4), mapping across the 5' end of the dystrophin gene, was screened for gross deletion on agarose gels stained with ethidium bromide. No deletions or truncated fragments were observed in any of the males tested.

Single strand conformational analysis (SSCA) was used to detect point mutations and microdeletions in these DNAs. Changes in conformation of the DNA molecule are recognised through aberrant migration of one or both of the single strands on a non-

denaturing gel (Orita et al, 1989) (Chapter 2.5.5.2.). PCR amplification and SSCA of the DBP in several affected individuals and normal members of each family detected a change in one family. The band shift observed in Family E occurred in a normal male used as a control. It could be demonstrated to have been transmitted from his heterozygous mother and formerly from her father who also transmitted it to his other daughter (Figure 4.4). A total of 40 additional chromosomes from a control population and from other unrelated MRX family members were screened for the conformational change. All controls carried the more common form suggesting that this shift in the SSCA band is a normal sequence variant[1] not associated with the retardation phenotype.

DNA aliquots from the normal boy carrying the altered SSCA band, and an affected sibling were directly sequenced by two alternate methods (Chapter 2.5.5.1.) to determine the base substitution that was suspected. The base change A to G (Figure 4.5), is a transitional mutation replacing a purine with a purine within the promoter region. This base A, in the most common situation, is part of an exonic ATG codon upstream from the putative ATG translation initiation codon of dystrophin. This site is also upstream of two major and several minor transcription initiation sites (Boyce et al, 1991). This single point mutation cannot be linked to the disease-causing mutation in this family since it occurs within the promoter of the dystrophin gene in a normal boy and did not occur in any of the affected males. It can be concluded that neither transcription nor translation of the protein have been adversely affected ie. it is a conservative substitution with no pathophysiological effect. This non-pathogenic base change in the promoter may have occurred outside the functional domains involved in regulation. If there is any effect on the full-length brain isoform, production of other brain specific isoforms and muscle dystrophin may compensate. Dystrophin expression would have to be measured in the boy to determine products from the different isoforms.

The sequence of this promoter or alternative first exon, matched that given in GenBank accession number M59228. The altered base is 203 of 433 (Boyce et al, 1991). It is remarkable that in several species sequenced for this promoter region (located on the GenBank Database) this base remains the same as the most common allele in the human population examined here. Sequence comparison of the brain specific exon has shown an extraordinary 99-100% homology between mammals and 95% between mammals and birds (Hugnot et al, 1993). This extreme conservation suggests an essential role underlying the preservation of this sequence intact through evolution and leads to the expectation that the

---

[1]A variant has an allele frequency < 0.01, occurring at very low frequency in the population. (while a polymorphism has at least two alleles with freq ≥ 0.01).

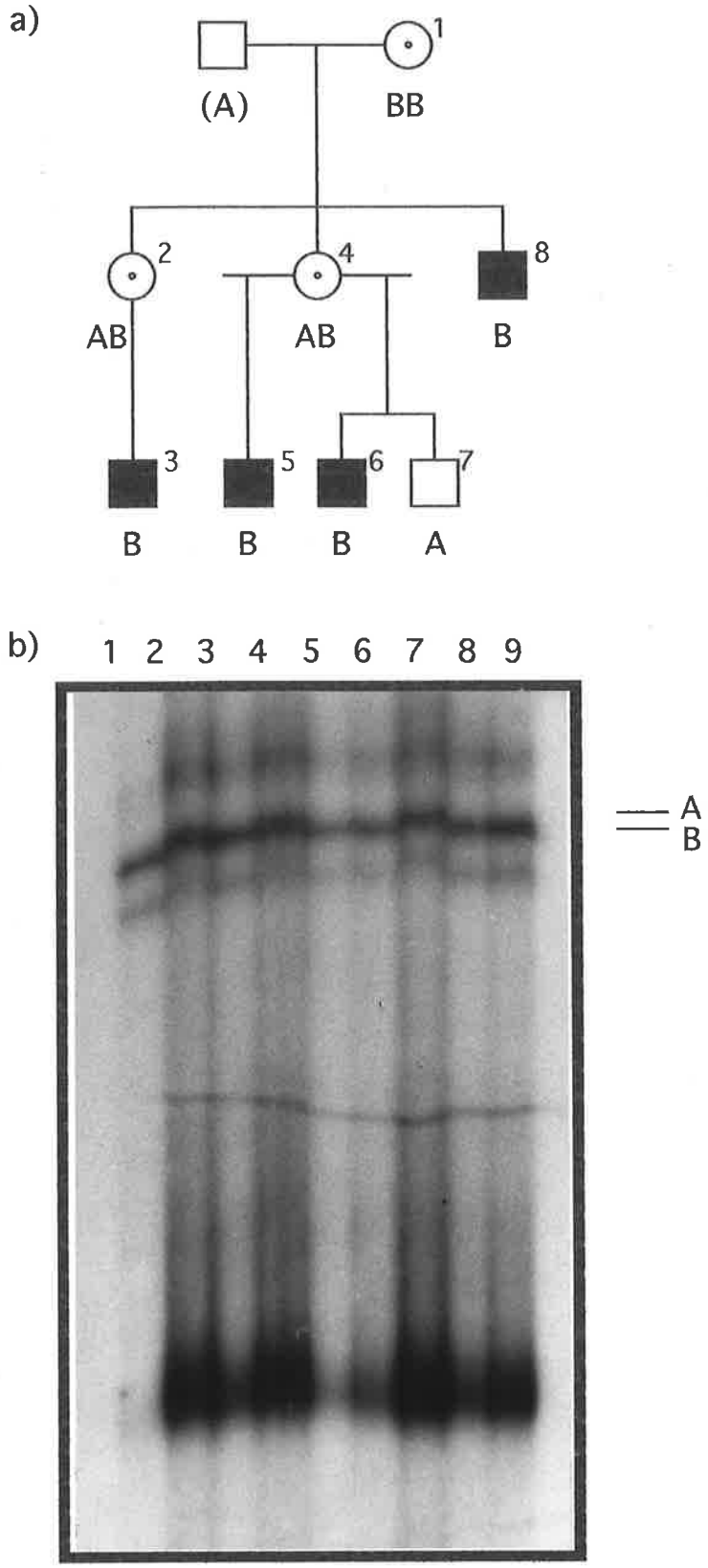


Figure 4.4. Single strand conformation analysis of the dystrophin brain promoter in family E (Appendix II, Paper 4). a) pedigree with numbers corresponding to lanes in b, and genotypes as detected by SSCA. b) conformational change seen in lanes 2, 4 and 7. Lane 9 is representative of the control population.

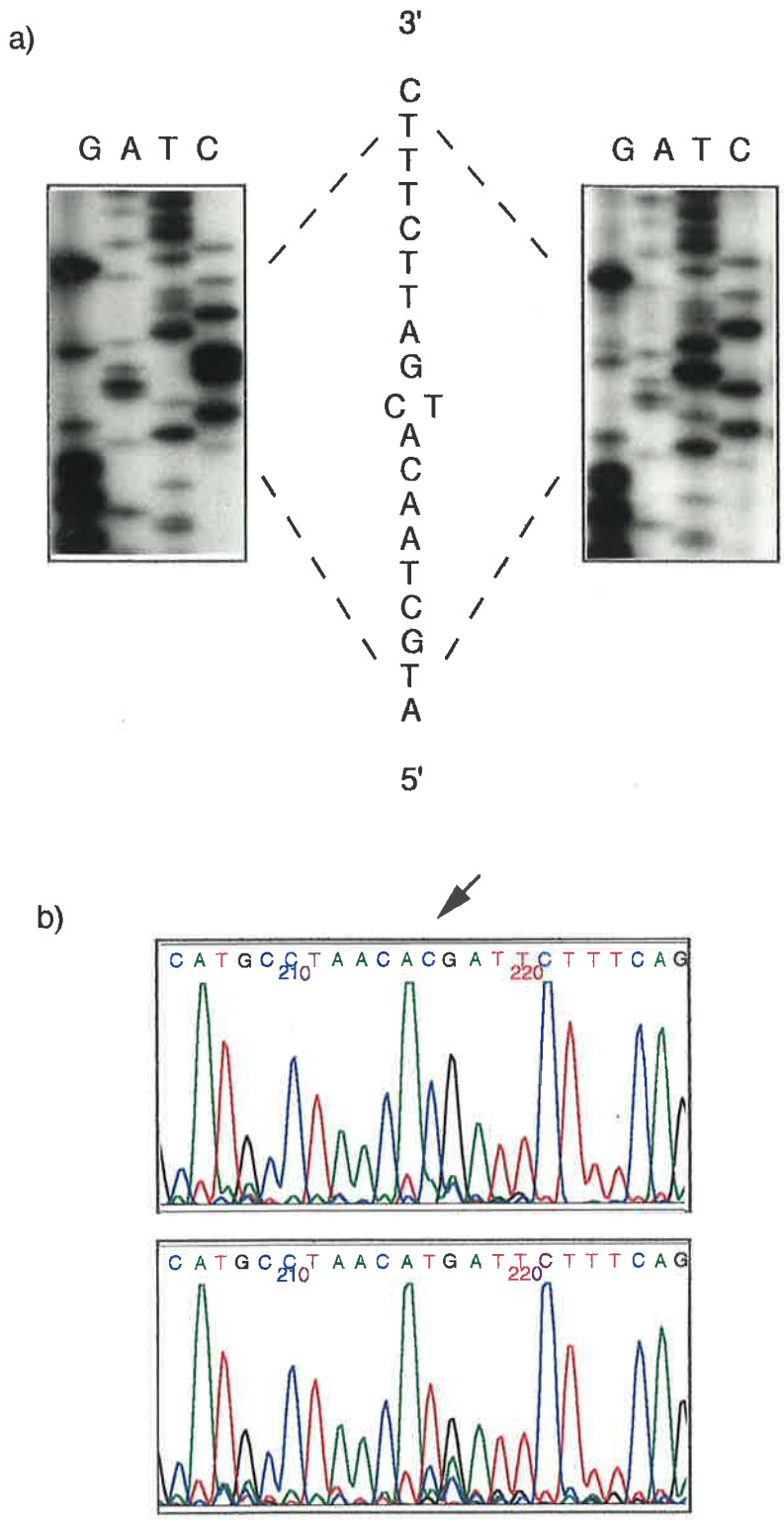


Figure 4.5. Samples were subjected to DNA sequencing. a) The left panel shows a portion of the sequence identified by SSCA as containing a mutation and the right panel shows the wild-type sequence. b) The corresponding region sequenced by the automated fluorescent technique. A different peak can be seen on the chromatogram at the position of the base change. On the other strand the variant possessed a single A to G base change confirmed by both methods in both directions.

region should be sensitive to point mutation. Parts of the dystrophin gene, however, seem tolerant to a certain amount of non-pathogenic sequence variation (Roberts et al, 1992). Only seven bases upstream of the substituted base described here, an A in the human corresponds to a highly conserved G in dog, rat, horse and chicken (Hugnot et al, 1993).

Deletions of this brain promoter in DMD patients have not been associated with mental retardation (Hugnot et al, 1993) suggesting that the brain promoter is not solely responsible for dystrophin in the brain. The brain and muscle promoters have been shown to be tissue selective rather than strictly tissue-specific (Bies et al, 1992) so that other non-muscle isoforms may have more specific developmental or functional roles. At least four alternative transcripts, expressed as distinct mRNAs in specific brain areas associated with cognitive function, are regulated by different promoters (Górecki et al, 1992). Promoters of these shorter transcripts in the brain and of the Purkinje isoform may provide insight into the functional differences of these isoforms in specific tissues and into the mechanisms of mental retardation in DMD. In the C-terminal region at the 3' end for example, the protein Dp71 is the most abundant dystrophin transcript in brain (and is not detected in differentiated skeletal muscle). Disruption (point mutation) of the translational reading frame of this non-muscle isoform has recently been characterised by high association with cognitive dysfunction in severely affected DMD patients (Lenk et al, 1993). Although this correlation may be due to specific mutations or deletions of the dystrophin gene, another gene could be present within the extremely large DMD locus (Koenig et al, 1987).

In conclusion the brain promoter at the 5' end of the dystrophin gene has been excluded from involvement in the phenotype of mental retardation in these MRX families. The detection by SSCA and sequencing of a point mutation in the dystrophin brain promoter demonstrates one rapid approach for the detection of mutation in candidate genes. The sensitivity of SSCA has proven to be effective for detection of single base substitution in a large (>600bp) DNA fragment. This method can be applied to screening the Dp71 candidate in those MRX families where the localisation extends to the 3' end of dystrophin (ie. MRX10). Specific defects in functional isoforms of dystrophin during brain development may be associated with expression of the non-specific MRX phenotype.

4.7. Conclusion

Non-specific X-linked mental retardations (MRX) are a genetically heterogeneous group of conditions that cannot be clinically subdivided. The method of gene localisation by linkage therefore necessarily 'split' the disorder into individual families representing specific mutations. The mapping studies outlined in this chapter have made a significant contribution toward defining genetic heterogeneity for MRX by determining the gene

localisations in ten large families and compiling data available on 21 more. Overlapping localisations may represent allelic mutations at some loci. Only identification of the gene may resolve some of these, through 'lumping' families with mutations in the same gene. The short-term implications of the localisations presented are advantageous - firstly to determine the number and distribution of genes contributing to idiopathic MRX, secondly to the nosology of MRX and lastly but perhaps most righteously to the direct and practical application of genetic counselling to participating families.

Linkage studies in these 31 large families have defined a minimum of 7 discrete loci responsible for familial MRX. The possibility of phenocopies occurring within these families at about 1-2% (Mandel, 1994) may result in mapping errors being subsequently discovered in some MRX families. Nevertheless current data, together with the MRX expressed at the FRAXE fragile site at Xq28 (Chapter 7), support the conclusion of 8 distinct genes for familial MRX cumulatively spanning virtually the entire X chromosome. Two further MRX loci defined by physical methods on Xq, are not overlapped by other MRX gene localisations defining discrete intervals. There are at least ten genes or specific gene intervals involved in this extremely heterogeneous disorder. Earlier theoretical estimates of the number of X-linked genes involved in non-specific intellectual disability were based on several assumptions of unknown parameters (section 4.1.) and included the fragile X and Renpenning syndromes which have since been delineated. This study has defined about half of the maximum expected number of MRX loci lending some support to the accuracy of the original premises upon which theoretical estimates were based.

The establishment of gene locations has been shown to be of benefit in provision of carrier risk in a number of families (section 4.3.2.). Risk values calculated from the observed inheritance of the MRX associated haplotype provide more absolute estimates than figures based on pedigree structure alone. Progressive refinement of localisations has been particularly valuable in determination of risk for a woman in Family Q with a recombination within the early MRX10 gene interval. Further refinements of the genes segregating in these ten families depend on extending the parameters defined in this study, particularly increasing family size and marker number.

The expression of genetic disease is never absolutely uniform between individuals, and may depend on many modifying factors including environmental and genetic interactions. The severity of retardation in MRX families is sometimes variable between sibships and even within sibships of a family (section 4.1.). Against such a background, similar clinical findings in overlapping genes cannot provide hard evidence to substantiate the possibility that at least some of the MRX genes described to date may be allelic. The most clear clinical discriminator that has been uncovered above is the presence of

manifesting heterozygotes in some families. This finding may have significance with relation to the position of the gene(s) and skewed inactivation. Variable expression of other features such as stature, head size, gait and behaviour may help to associate the disorder in one MRX family to another, or to one of more consistent or severe expression that qualifies as a syndrome. Objective anthropometric measurements will in future permit comparisons with the regional localisation of MRX genes between families to define specific subtypes and may reveal subtle but specific clinical features in common between families at a given locus.

It can not be proposed that MRX involves only brain development or function, since there is much evidence that metabolic disorders can cause mental retardation as well (Chapter 1.6.3.). Consideration of MAOA deficiency as a mild inborn error of metabolism suggests that metabolic defects need not have extensive phenotypic effects. MRX may represent milder phenotypic variants due to allelic mutations at genes causing syndromal XLMR or integral parts of contiguous gene complexes such as MRX at Xq21 frequently occurring in association with deafness (DFN3) and/or choroideremia (CHM). An apparent cluster of MRX genes at Xq28 may be due to mutations in a candidate gene such as the L1CAM gene, encoding a neuronal cell adhesion molecule, known to give rise to the broad clinical spectrum of CRASH syndrome. The observed clustering of MRX genes, for example at the centromere, could reflect a number or family of genes essential to the developing brain or may reflect the gene frequency at a single relatively more common locus. The number and extent of pericentromeric gene localisations may however reflect the reduced recombination around the centromere and difficulties with refinement of mapping, rather than imply a greater number of MRX genes in the region.

Given the numerous examples of allelism where very different clinical patterns arise from mutations in the same gene, the real number of genes causing XLMR will become more difficult and perhaps unnecessary to determine. Enumeration of MRX genes gives merely an indication of the number of genes involved in retardation but it cannot be concluded that these genes are necessarily determinants of brain function only, until allelic conditions can be excluded. The concept that MRX genes represent 'pure' determinants of mental function naively disregards the possibility of allelic mutations possibly expressed at different developmental stages, causing variation in phenotype.

Further linkage mapping in families may delineate only a few more loci before it is impossible to distinguish involvement of non-overlapping regions. Once the limit of resolution by linkage has been reached, and without discovery of gross chromosomal rearrangements or translocations, direct screening of candidate genes for plausible mutations expressed in the appropriate tissue may be the most straightforward for elucidation of these

genes. It has been proposed that an international collaborative lymphoblastoid cell line repository be established to enable rapid screening of a gene of known position through probands from families exhibiting overlapping gene localisations (Mandel, 1994). In this manner samples from families too small to achieve a lod score of +2 or greater, and families with syndromal forms of X-linked mental retardation can be included into mutation screening protocols. Eventually it is envisaged that overlapping regional localisations in MRX families will be lumped or split by these means and the true number of X-linked genes for mental retardation determined.

CHAPTER 5

Syndromal X-linked Mental Retardation (MRXS)

Note to Publications

The work summarised in this chapter has been published, with clinical assessments carried out by Dr. E.A. Haan, Dr. G. Turner, Dr. M.W. Partington and Dr. M. Wilson. All molecular studies were carried out and prepared for publication by the candidate. Reprints are bound in Appendix III in the following order:

1. Sutherland GR, **Gedeon AK**, Haan EA, Woodroffe P, Mulley JC (1988) Linkage studies with the gene for an X-linked syndrome of mental retardation, microcephaly and spastic diplegia (MRX2). Am J Med Genet 30:493-508.
2. Turner G, **Gedeon A**, Mulley J, Sutherland G, Rae J, Power K, Arthur I (1989) Börjeson-Forssman-Lehmann Syndrome: Clinical manifestations and gene localisation to Xq26-27. Am J Med Genet 34:463-469.
3. Wilson M, Mulley J, **Gedeon A**, Robinson H, Turner G (1991) New X-linked syndrome of mental retardation, gynecomastia, and obesity is linked to *DXS255*. Am J Med Genet 40:406-413.
4. Mulley JC, **Gedeon AK**, Wilson S, Haan EA (1992) Use of linkage data obtained in single families: Prenatal diagnosis of a new X-linked mental retardation syndrome. Am J Med Genet 43:415-419.
5. **Gedeon A**, Partington M, Mulley J (1994) X-linked mental retardation with dystonic movements of the hands (PRTS): Revisited. Am J Med Genet 51:565-568.
6. 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.
7. **Gedeon AK**, Kozman HM, Robinson H, Pilia G, Schlessinger D, Turner G, Mulley JC (1996) Refinement of the background genetic map of Xq26-q27 and gene localisation for Börjeson-Forssman-Lehmann syndrome. Am J Med Genet 64:63-68.
8. **Gedeon AK**, Haan E, Mulley JC (1996) Gene localisation for Sutherland-Haan syndrome (SHS:MIM 309470). Am J Med Genet letter to the editor 64:78-79.
9. **Gedeon AK**, Turner G, Mulley JC (1996) Gene localisation for Wilson-Turner syndrome (WTS:MIM 309585). Am J Med Genet letter to the editor 64:80-81.

It is recommended that these reprints be read prior to reading this chapter since there is frequent reference from the chapter.

5.1. Introduction

The classification of X-linked mental retardations (XLMR) into non-specific (MRX) or syndromal (MRXS) forms based on evident clinical features has been discussed in the previous chapter. Whilst MRX genes represent the 'simple' clinically homogenous fraction of XLMRs, MRXS are the clinically heterogeneous or 'complicated' forms largely classified on the basis of the pattern of anomalies contributing to the complex clinical phenotype (Chapter 1.6.3.). Over 127 conditions subdivided into five categories have been described as XLMR, with 57 (44%) of these identified by a recognizable pattern of physical anomalies (Neri et al, 1994). Mental retardation syndromes also occur as inborn errors of metabolism such as MAOA deficiency, Hunter and Lesch-Nyhan syndromes, as dominant male lethal disorders and secondary to neuromuscular defects such as DMD and Pelizaeus-Merzbacher. Families with a primary 'idiopathic' mental retardation concomitant with dysmorphic features or characteristic pattern of physical phenotype are the subject of this chapter.

Over the past 50 years numerous XLMR conditions have been reported with the implicit suggestion that diverse phenotypes represent distinct disorders. Follow-up studies have indeed delineated some specific entities as well as broadened the clinical spectrum toward recognition of others (Stevenson et al, 1994). Clinical differences alone are no longer recognised as sufficient evidence that the disorders in different families are due to different loci (Romeo and McKusick, 1994). In Chapter 4 it was explained that clinical homogeneity (which in MRX is an implied lack of consistent phenotypic findings) was insufficient evidence for assuming disorders to be the same, even in the light of co-localisation of disease genes. Lumping of disorders should be delayed at least until mutations in the same gene can demonstrate allelism. Clinical heterogeneity can be defined as phenotypic variation arising as a consequence of; a) genetic heterogeneity, b) allelic heterogeneity or c) variable expression of the same mutant allele because of i) genetic and/or ii) environmental background. Genetic or locus heterogeneity is evident when more than one locus is responsible for the disease, as demonstrated by linkage mapping of non-overlapping MRX gene localisations (Chapter 4). Conversely, clinically different entities have been shown to be allelic variants due to specific mutations at a single locus such as for example the L1CAM gene (Fransen et al, 1995). Ideally then, distinct syndromes ought to be defined as both clinically and genetically heterogeneous.

The occurrence of X-linked mental retardation at an estimated rate of 1 in 600 male births (Sutherland and Hecht, 1985) and the theoretical expectation of there being at least 7 to 19 genes for mental retardation (Herbst and Miller, 1980) mean that in determining the

number of genes mapping to the X, consideration must also be given to syndromal XLMR. Given the possibility of clinical heterogeneity determined by allelic rather than locus heterogeneity, the precise mapping of MRXS genes is increasingly of value as an adjunct to the clinical definition of specific syndromal entities. Accurate genetic mapping and clinical descriptions of these loci may ultimately have significance for the delineation of genes for MRX, MRXS and the nosology of syndromes, though more immediately may benefit genetic counselling and prenatal diagnosis in individual families.

The MRXS disorders are often private syndromes in that they are generally rare and often described only in one or a few families. The characteristic dysmorphic features or physiological signs, believed to be causally related to the mental retardation, are often used to describe and name the disorder. Many are also identified by an eponym utilising the name(s) of the person(s) first characterising the syndrome in a formal publication. Introduction of a system of nomenclature to differentiate molecular studies in individual families enumerates mapped loci as MRXS1, MRXS2...etc (Mulley et al, 1992), until formal gene symbols can be assigned (Chapter 4.2.). A syndrome may thus have more than one name by which it is known (section 5.3.).

For gene mapping of any disorder, including MRX (Chapter 4) and MRXS, establishment of the clinical definition of the phenotype provides the basis for classifying individuals within a family as affected or unaffected. Reliability of the clinical diagnosis may be complicated by phenocopies and variation of the phenotype within the dataset. The first is usually suspected by the clinician and may be detected during gene mapping as repeatedly recombinant when the phenotypes of everyone else in the family are apparently linked to a marker allele. The assessment of individuals within a family demonstrating a range of phenotypic variation, is dependant on the skill and experience of the clinician, although where a diagnosis is in doubt that individual is omitted from the analysis.

Many mental retardation syndromes occuring in small families cannot be localised by molecular genetic means but have been described as distinct clinical entities that purport to delineate a new syndrome (for example Sklower Brooks et al, 1994). The eventual molecular characterisation of the X chromosome and the isolation of genes involved in large families with mental retardation as a prominent feature, may ultimately group many of these as allelic mutations within a finite number of genes on the X. While clinical differentiation of the disorder segregating in families contributes to the delineation of XLMRs, gene localization and identification by molecular approaches is required to provide evidence to define distinct genetic disorders.

The main goal of this study was to map, or refine the mapping of MRXS entities. These included BFLS, SHS, WTS and PRTS that had been previously localised using only

RFLP markers, and a new syndrome characterised by macrocephaly along with mental retardation in affected males. During the course of this work a high resolution background map of dinucleotide repeat markers was developed (Chapter 3) and a specific regional map around BFLS was constructed including three newly characterised microsatellite markers. The reduction of each of these localisations reduces the number of candidate genes within the gene interval and in the interim improves the informativeness of risks provided for genetic counselling. Reduced regional localisations will also be useful in the nosology of XLMR by excluding allelism for X-linked disorders with distinct gene locations.

5.2. Materials and Methods

Peripheral blood samples were gathered from participating individuals in each of five large families segregating a specific syndrome of XLMR (section 5.3.) by the collaborating clinician involved with each family. The eponyms by which three of the families are known, at least in part reflect the contribution of each clinician to the definition of the syndrome ie. Drs. Partington (PRTS), Haan (SHS), and Wilson and Turner (WTS). The clinical delineation of affecteds from normals was determined before mapping studies were commenced. In those cases where the phenotype of an individual posed any doubt, the genotypes were either excluded from the analysis or the affection status was coded as unknown.

All methods applied, from DNA extractions to genotyping strategies, were as described in Chapters 2 and 4 and/or as specified in publications for each family. Gene localisations were performed chiefly using PCR-based polymorphisms against the background map (Chapter 3). The genes in the SHS and BFLS families had been mapped previously by the candidate, so that DNA samples from many individuals had been stored. In the family with SHS, a number of individuals had to be rebled when some of the original samples did not observe Mendelian laws of inheritance. These errors may have arisen in the field (during bleeding), or in the laboratory (during DNA extraction) and only became apparent with the use of the multiallelic microsatellite markers. The possibility of further such errors was minimised by careful DNA extraction procedures that treated each tube of blood from any individual independently and with due care of transcription errors (Chapter 2).

Estimates of carrier risk were determined in individuals on request only, using the best though most conservative data available (Chapter 2). Prenatal diagnoses were only provided if carrier risks were high and informative microsatellite markers closely flanked the gene interval with one exception (see WTS). Informative markers within the gene interval were used to establish the haplotype associated with MRXS. These genotypes could

exclude double recombination events between flanking markers though for loci separated by less than 10cM this chance is less than 1 in 400.

5.3. Family studies

Partington syndrome (PRTS)

An XLMR syndrome with dystonic movements of the hands, dysarthria and seizures was originally described and mapped in a large family, to the distal short arm of the chromosome (Partington et al, 1988). Affected family members had mild to moderate MR with no signs of the disorder in obligate carrier females. The disorder is known by the eponym Partington syndrome (PRTS) and may be found in (O)MIM (MIM#309510) and listed in GDB under the symbol MRXS1 in keeping with more recently established nomenclature (Mulley et al, 1992).

The original localisation established with RFLPs encompassed the entire distal half of the short arm of the X chromosome, with the peak lod score of 2.1 at DXS41 (Partington et al, 1988). Refinement of this expansive gene interval was one objective undertaken by this project. Utilisation of dinucleotide repeat markers (Chapter 3) facilitated this goal. The pedigree, lod score table and multipoint analysis are reproduced in Appendix III, Paper 5. The PRTS gene interval was refined to 14cM in Xp22.1 between DXS365 and DXS28, by recombination events detected in VI-2 and V-30 respectively. Three further dinucleotide repeat markers mapping into this interval have since been genotyped in the family. DXS1052 and DXS274 were not informative, but DXS1683 which lies physically proximal to DXS365 (Econs et al, 1994), detected the recombination event in VI-2, thereby further reducing the gene interval. Without placement of DXS1683 in the current genetic background map (Willard et al, 1994), however, the map distance cannot be estimated, therefore this marker was not used for determinations of risk.

Marker genotypes at 12 loci spanning the PRTS interval, from DXS207 (most distal) to STR44 (proximally) were haplotyped to minimise recombinants (Figure 5.1). The haplotype segregating with the PRTS mental retardation phenotype is G,2,A,A,H,S,10,F,A,B,B,D. The reduced localization permitted carrier and prenatal diagnoses using informative flanking markers to identify the haplotype. The closest informative flanking markers in the matriarch, III-5, were DXS365 and STR50, separated by 25cM. Only two (IV-10, IV-16) of her six participating daughters were not obligate carriers, their carrier risks were calculated as 2%. The risk to the daughter of IV-10 (V-15), was further reduced to less than 2%, while the risk to V-32 (daughter of IV-16) remains 2%

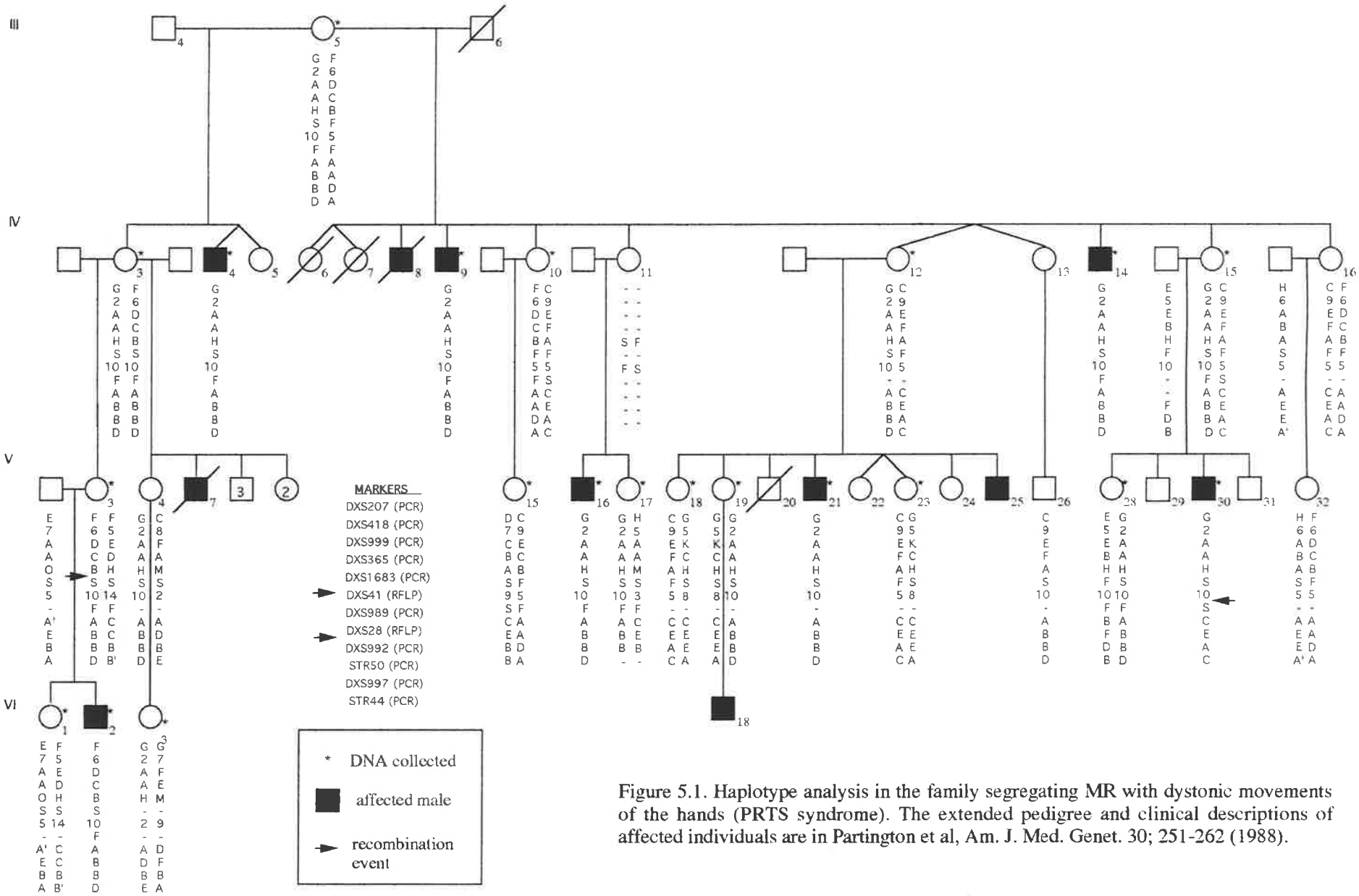


Figure 5.1. Haplotype analysis in the family segregating MR with dystonic movements of the hands (PRTS syndrome). The extended pedigree and clinical descriptions of affected individuals are in Partington et al, Am. J. Med. Genet. 30; 251-262 (1988).

reflecting the chance of double crossover within the PRTS gene interval. Determination of carrier risk for V-4 was complicated because her mother, IV-3, was uninformative at each of five proximal loci up to 13cM from the gene interval. The chance that V-4 is a carrier remains 50:50 as estimated from the pedigree. Without the paternal genotype to aid in haplotyping, the likely location of the meiotic recombination detected in VI-3 cannot be determined.

Prenatal diagnosis was requested for the potential carrier V-28. The carrier status of V-28 was first established with a risk greater than 99.5%, allowing for the possibility of undetected double crossover. Informative PCR-based flanking markers DXS365 and DXS992, separated by 21cM, were assessed with 21% chance that a single crossover would render the diagnosis uninformative. PCR amplification of a Y-specific repeat (Kogan et al, 1987) in DNA extracted from the CVS, determined the fetus to be male and excluded maternal contamination of the sample. The DXS999 and DXS365 loci were informative distally and DXS992, STR50 and STR44 proximally. The affected haplotype at these loci is A,A,A,B,D while the haplotype found in the fetus was E,B,B,F,B for these markers. This showed that the fetus inherited the normal grandpaternal haplotype. The risk that the fetus was predestined to be affected, using the closest flanking markers DXS365 and DXS992, was 1.3%. The error rate of 1.3% takes into account the chance of double crossover between the closest flanking markers.

The male, V-26, has mild mental retardation but no dystonic movements and is clinically quite unlike the other affected males. In the describing clinicians experience, although 9 of 9 males had dystonia in the rest if the family, the dystonic movements appeared to be progressive with age-dependent expression. Genotype analysis of V-26 assessed the likelihood that he carries the same gene mutation as other affected family members. A single recombination event was detected between the markers DXS1683 and DXS41, disrupting the haplotype and rendering these markers uninformative for risk determination. Further reduction of the gene interval excluding the recombinant region or isolation of the PRTS gene, may be the only means of determining the involvement of the PRTS gene in his phenotype.

Other mental retardation genes located in Xp22, including the MRX2 and MRX19 loci and male-lethal Goltz syndrome (MIM#305600), cannot be excluded as examples of allelic heterogeneity. The Coffin-Lowry (CLS) syndrome (MIM#303600) of mental retardation also mapped to Xp22.3-22.1 has specific diagnostic features not shared by individuals with PRTS. Other disorders with dystonic movements include X-linked dystonia-parkinsonism mapped to Xq12-q22; Waisman syndrome, a neuromuscular disorder associated with mental retardation mapped at Xq27.3-qter, and MRXS5 with Dandy-Walker

malformation and seizures mapped to Xq26 (Pettigrew et al, 1991; Huang et al, 1991). Such genetic heterogeneity suggests that dystonia can occur as a progressive non-specific feature in several X-linked conditions. This feature may be overlooked in clinical assessments of children and should be critically evaluated in affected males from families where the gene localization overlaps that of PRTS (Appendix III, Paper 5).

Sutherland-Haan Syndrome (SHS)

An XLMR of borderline/mild to moderate retardation associated with short stature, microcephaly, brachycephaly, small testes and spastic diplegia was initially designated MRX2 (Appendix III, Paper 1). Growth retardation, probably beginning *in utero*, and delayed puberty were notably common features among affected males, with spastic diplegia appearing to be more severe in early life than adulthood. The MRX2 symbol was later withdrawn (Appendix II, Paper 2) when classification guidelines were suggested reserving the MRX nomenclature for non-specific, and introducing MRXS for syndromal mental retardations (Mulley et al, 1992). This syndrome, now referred to by the eponym Sutherland-Haan syndrome (SHS), by MIM#309470, or as MRXS3, was initially mapped between Xp21 and Xq21 with a peak lod score of 2.10 (θ=0.11) at DXYS1 (Appendix III, Paper 1). The RFLP markers at DXS84 and DXS87 were the closest flanking the SHS gene.

Application of multiallelic PCR-based microsatellite markers (Chapter 3) to the regional localisation of this gene detected sample mismatches between the first and second samples from a number of cases. Seven individuals were re-collected when the genotypes of one of their original samples did not obey Mendelian inheritance. Genotyping of the new samples restored Mendelian inheritance at all but one locus (AR) which demonstrated a new mutation in the microsatellite repeat (see below). DNA was available from 8 affected males in three generations, with pedigree identifiers as given (Appendix III, Paper 1). The patient IV-16, was very young at the time of the original study and his affection status was uncertain although he had some features in common only with the affected males in the family. Reassessment of this boy at age 4yrs 2mths, indicated that the features including spastic diplegia, anal stenosis and borderline/mild mental retardation were consistent with affection status (Appendix III, Paper 9). He was coded for linkage as an affected member of the pedigree.

The mildly retarded 47,XXX female, III-5, caused a problem to the 1988 linkage analysis since the biallelic RFLP markers could not be used to determine whether she had one or two copies of the mental retardation gene. She was coded then as unknown at each marker genotype and it was noted that at DXYS1, where one of her sons was a recombinant,

the program would designate her as a heterozygote. Genotyping at 20 informative markers between DXS538 and DXYS1X permitted construction of the haplotypes segregating in the family. Recombination events detected in III-10 and III-12 at MAOA and in IV-4 at DXS106/DXS1125 represent the closest loci flanking the SHS gene within a 21.7cM interval. It was observed that III-5 carried three alleles, both maternal and the inferred paternal, at the DXS538, CYBB, DXS1125 and DXS566 loci. At those intervening markers, where all three parental alleles were known or could be inferred, only a single maternal and the normal paternal alleles were present in III-5. This phenomenon can be explained by invoking maternal non-disjunction at the second meiotic division, coupled with recombinations between the homologues at Xp and Xq. This combination of events would lead to maternal iso-disomy in the region between chiasma. These chiasma occurred between the CYBB and DXS7 loci on Xp, and proximal to DXS1125 on Xq. Since this region overlaps the gene interval defined by male recombinants, it is likely that III-5 carries two copies of the SHS gene. The phenotype of III-5 can not be wholly attributed to the double dose of the SHS gene since XXX females can be mentally retarded.

The linkage analysis programs do not allow for individuals having more than the normal diploid complement of chromosomes. Omission of this section of the pedigree would falsely increase the interval of gene localisation, since the recombinant defining the q arm limit is detected in her son IV-4. Two alternative methods for dealing with this situation were devised to limit loss of information. Alternative A is based on the finding that III-5 has only two apparent alleles in the gene interval, the two copies of the maternal homologue being identical, thus recombination between these sister chromosomes cannot be detected. In this interval then, she can be coded as a normal 46,XX female. Genotypes could not be coded at loci where three alleles and therefore XXX were detected since these would imply non-paternity when the paternal genotype would be inferred from the genotype of her sister III-4. Alternative B removes the 47,XXX female from the analysis and places her two sons (IV-3 and IV-4) up one generation as sons of their grandmother (II-4) whose chromosomes they inherit. The grandpaternal haplotype had not been transmitted to either boy. This second alternative reduces the information by one meiosis but provides a more conservative lod score estimate at all loci including those flanking the SHS gene. Pairwise lod scores between SHS and 20 polymorphic markers, with both alternatives at the 13 markers within the gene interval, are given in Table 5.1. The peak lod scores at seven loci within the interval were Zmax=4.63 at θ=0.0 under alternative A and Zmax=4.33 at θ=0.0 under alternative B. Multipoint lod scores were not calculated since all potentially informative meioses were informative at several markers across the interval. Mapping the disease locus

TABLE 5.1 : PAIRWISE ANALYSIS BETWEEN SHS AND 20 MARKER LOCI

Loci	θ 0.001	0.01	0.05	0.1	0.2	0.3	0.4	Z max	θ
DXS538	-10.66	-5.70	-2.39	-1.12	-0.12	0.21	0.22	0.24	0.35
CYBB	-10.67	-5.71	-2.40	-1.14	-0.14	0.18	0.19	0.22	0.36
DXS7 B	-2.57	-0.61	0.61	0.96	1.00	0.70	0.26	1.04	0.15
A	-2.57	-0.61	0.59	0.93	0.96	0.66	0.24	1.00	0.16
MAOA3 B	-1.97	-0.01	1.19	1.52	1.50	1.15	0.61	1.57	0.15
A	-1.97	-0.01	1.17	1.48	1.46	1.11	0.58	1.54	0.14
MAOA B	-2.27	-0.32	0.87	1.18	1.16	0.81	0.31	1.24	0.14
A	-1.97	-0.02	1.15	1.44	1.36	0.95	0.40	1.47	0.13
DXS1003B	4.33	4.27	3.99	3.63	2.84	1.96	0.99	4.34	0.00
A	4.63	4.57	4.27	3.89	3.05	2.11	1.07	4.64	0.00
SYN1 B	4.33	4.27	4.00	3.65	2.87	1.99	1.01	4.34	0.00
A	4.63	4.57	4.28	3.90	3.07	2.14	1.09	4.64	0.00
PFC B	4.33	4.27	3.99	3.63	2.84	1.96	0.99	4.34	0.00
A	4.63	4.57	4.27	3.89	3.05	2.11	1.07	4.64	0.00
DXS426 B	4.33	4.27	4.00	3.65	2.87	1.99	1.01	4.34	0.00
A	4.63	4.57	4.28	3.90	3.07	2.14	1.09	4.64	0.00
DXS573 B	4.33	4.27	3.98	3.60	2.77	1.83	0.80	4.34	0.00
A	4.03	3.97	3.70	3.35	2.57	1.69	0.72	4.04	0.00
DXS991 B	3.73	3.68	3.46	3.17	2.50	1.74	0.88	3.74	0.00
A	3.73	3.68	3.45	3.14	2.46	1.70	0.86	3.74	0.00
ALAS2 B	4.33	4.27	4.00	3.65	2.87	1.99	1.01	4.34	0.00
A	4.63	4.57	4.28	3.90	3.07	2.14	1.09	4.64	0.00
AR B	4.33	4.27	4.00	3.65	2.87	1.99	1.01	4.34	0.00
A	4.63	4.57	4.28	3.90	3.07	2.14	1.09	4.64	0.00
PGK1P1 B	4.33	4.27	3.99	3.63	2.84	1.96	0.99	4.34	0.00
A	4.63	4.57	4.27	3.89	3.05	2.11	1.07	4.64	0.00
DXS106 B	1.33	2.28	2.72	2.68	2.24	1.59	0.82	2.73	0.07
A	1.63	2.57	2.99	2.93	2.44	1.74	0.90	3.00	0.07
DXS1125	1.03	1.98	2.45	2.44	2.06	1.48	0.76	2.47	0.07
DXS453	-4.97	-2.02	-0.14	0.47	0.73	0.56	0.19	0.73	0.19
DXS559	1.03	1.99	2.46	2.47	2.10	1.51	0.78	2.50	0.07
DXS566	0.43	1.39	1.91	1.97	1.75	1.32	0.74	1.98	0.09
DXYSIX	0.43	1.39	1.89	1.91	1.60	1.07	0.42	1.93	0.08

against the fixed background map of markers would not increase the likelihood for gene localisation to this region.

Another anomaly detected in this family was allelic mutation at the trinucleotide repeat of the AR locus. The allele 2 segregating with the disease in all affected males became one repeat unit shorter in the female II-5. This change possibly arose due to slippage during replication but was reliably transmitted to two of her sons, III-10 and III-11, confirming that the mutation is both somatic and germline. These affected males carry the haplotype consistently segregating with mental retardation in this family. The mutant allele in these three individuals was therefore coded as allele 2 for the linkage analysis given that the aberration was not associated with the disease state. Similar new mutations have been shown to arise at this and other microsatellite loci in CEPH families (Appendix I, Paper 3) and although clonal evolution and artefacts of cell culturing account for some, a proportion (mutation rate estimated around $1x10^{-4}$) may arise *in vivo* as in the SHS family (Chapter 1.2.2.).

Wilson-Turner Syndrome (WTS)

The gene for a private syndrome with mental retardation, gynaecomastia, obesity and minor anomalies including speech difficulties (MIM#309585) was initially localised using RFLPs spanning the chromosome (Appendix III, Paper 3). The disorder is now known by the eponym WTS (Wilson-Turner Syndrome) but previously by the acronym MRGO (mental retardation with gynaecomastia and obesity), or the interim MRXS6 symbol. Mapped between the closest flanking markers DXS84 and DXS94, the WTS phenotype was significantly linked with a peak lod score ($z=4.82$, $\theta=0.0$) to the DXS255 (M27β) locus at Xp11.

At this time an urgent request for prenatal diagnosis in an obligate carrier of this syndrome, IV-24 (pedigree identifier from Figure 1; Appendix III, Paper 4), posed a quandary given that the closest flanking markers were separated by over 50cM (with an unacceptable chance that a crossover would render the meiosis uninformative or that double recombination could occur and not be detected) and that she was uninformative for both. Without sufficiently close informative flanking markers, an alternative approach to risk assessment relied on the accuracy of the estimated two-point recombination frequency between WTS and DXS255 (Appendix III, Paper 4). Additional relatives genotyped at DXS255 contributed to establishment of a 90% support interval determined by the one-LOD-down method (Chapter 2.8.). The diagnosis was offered with a recombination frequency estimated between 0.0-0.09 (0-9%), with a 9% maximum risk of diagnostic error.

Reduction of the original regional localisation which spans approximately 66cM on the current map (Chapter 3) was subsequently attempted using dinucleotide repeats mapping between the established flanking markers. Pairwise lod scores at 22 informative loci are given in Paper 8 of Appendix III. The peak lod score of 6.07 (θ=0.0) was reached with the fully informative trinucleotide repeat at the androgen receptor (AR) gene locus. Recombinant events detected at loci internal to the original localisation refine the reduced limits to the gene. The closest flanking markers, defining the smallest gene interval for WTS, are DXS426 at Xp11.3 and DXS990 at Xq21.3. The genetic distance between these flanking markers is 24.7cM, a reduction of over 41cM. This reduction to the original gene interval has significance for genetic risk analyses and future attempts at gene identification.

The androgen receptor gene (AR) gene is involved in male sex differentiation such that androgen insensitivity syndrome patients develop breast hypertrophy after puberty (gynecomastia) and have reduced fertility (Biancalana et al, 1992). Signs of androgen insensitivity in X-linked spinal and bulbar muscular atrophy (SBMA or Kennedy disease) lead to investigations of AR as a candidate for this adult onset form of motor neuron disease and showed an absolute association of amplified CAG repeat in the coding regions of patients (La Spada et al, 1991). The endocrine effects in SBMA affected males and involvement in neuron growth provide valid reasons to evaluate the AR gene as a candidate for the overlapping WTS gene. The possibility that WTS arises as an allelic defect in androgen receptor function is yet to be determined.

An MRX gene associated with aphasia has been localised to the same pericentromeric region of the chromosome (Wilson et al, 1993) and should ultimately be investigated for allelism to WTS given the speech difficulties or mutism apparent in affected males. Syndromes with similar clinical features must be considered in the differential diagnosis of WTS. Both the specific clinical and linkage findings distinguish WTS from BFLS (see below) and from Simpson-Golabi-Behmel syndrome with some clinical similarities. Three other syndromes are characterised by obesity, hypogonadism and short stature (Chudley-Lowry MIM#309490; Vasquez syndrome and Young-Hughes syndrome) but none have been regionally localised (Neri et al, 1994). One could speculate that at least some of these represent the same syndrome and that discriminating clinical features may be allelic mutations in the same gene.

Börjeson-Forssman-Lehmann Syndrome (BFLS)

The earliest description of this syndrome (MIM#301900) was of a family with severe mental defect, epilepsy, hypogonadism, obesity and large ears (Börjeson et al, 1962).

Many years later a family with mild to moderate mental retardation, long ears, deep-set eyes, small testes and gynecomastia was presumed to have a milder form of BFL (Appendix III, Paper 2). Short, widely spaced and flexed toes with a broad forefoot were noted in affected males and variably expressed in heterozygous females. The gene was located by linkage to Xq26-27 with a peak two-point lod score at DXS51 (Appendix III, Paper 2). In another family with more classic BFL features (Ardinger et al, 1984), the gene was mapped also to Xq26 (Mathews et al, 1989). Combination of these data assuming genetic homogeneity had, until recently, defined the BFLS gene interval between HPRT and DXS105. The gene localisation in a third family, again with only mild/moderate retardation, overlapped both these limits (Berg et al, 1992). There is no evidence for genetic heterogeneity since both severe and mild forms map to the same region, although one further family described with similar phenotype but distinct facial appearance, has been mapped between discordant DNA markers at Xp21-q25 (Chudley et al, 1988). This molecular differentiation defined the clinically similar entity as a 'new' syndrome (Chudley, 1991) rather than an allelic mutation. Recent analyses have shown, however, that the recombinant defining the closest proximal limit of BFLS at HPRT was not recombinant at a more proximal marker (Mathews et al, 1993), therefore the proximal limit to the BFLS gene was undefined.

A high resolution, comprehensive background genetic map of markers surrounding and within the BFLS interval was constructed from CEPH family data (Appendix III, Paper 7). The map was developed first from a framework linkage map constructed by H Kozman, WCH (Chapter 3.4.1.1.) with order supported by the physical map of the region established on a YAC contig (by G.Pilia and D.Schlessinger, Washington USA) spanning over 8Mb (Little et al, 1992). Markers that could not be ordered by linkage with odds 1000:1 were then placed according to the physical map. Genetic distances between all markers were thus determined with integration of physical support for order (Chapter 3.4.1.2.). Map distances thus derived were applied to the definition of the BFLS gene interval, determination of genetic risk and provision of prenatal diagnoses as required by the family (Appendix III, Paper 7).

Dinucleotide repeat markers were applied to an extended version of the family first described by Turner et al (Appendix III, Paper 2) to define a proximal limit to the BFLS gene localisation. Pairwise lod scores at 16 loci spanning the interval and identification of the BFLS haplotype segregating in the pedigree are given in Paper 7 of Appendix III. In providing prenatal diagnoses and carrier assessments for family members, the most conservative estimates of map distance and mapping data were used to define the gene interval. The closest proximal recombinant accepted for this purpose was DXS425 in individual IV-4. The distal endpoint of the regional localisation was resistant to further

reduction and remained defined by recombination at DXS105 in individual V-3. Markers from within the regional localisation were also genotyped to guard against undetected double crossover.

Haplotype analysis of the pedigree demonstrated segregation of the disease haplotype that could predict some and exclude others as carriers. Prenatal diagnoses were performed in seven pregnancies of two women. Individual IV-5 requested prenatal diagnosis on two occasions (CVS3 and CVS5). Each time, the fetus received the grandpaternal haplotype and was male as determined by PCR sexing and confirmed cytogenetically. The carrier status of III-12 remains uncertain since II-8 may be a new mutation, thus predictive testing could only exclude the possibility of a male fetus being affected. Haplotype data in the first two male fetuses (CVS1 and CVS2) did not exclude affection status. The third fetus of III-12, CVS4, was female and carried a recombinant maternal chromosome. The recombination event between HPRT and DXS294 confirmed the order of the regional background map with placement of the DXS994 locus distal to the crossover point. The proximal loci carry the grandmaternal alleles consistent with affected males in the rest of the pedigree. The inherited allele at the DXS994 locus is grandpaternal and is consistent with grandpaternal alleles inherited at loci distal to and including DXS294. The unusual feet of CVS4 (at 2yrs of age), characteristic of affected males and some obligate carriers, suggests that she may carry the BFLS gene (although this feature alone is not a reliable indicator of carrier status). Given the recombination, the distal limit for the regional localisation of BFLS might therefore reasonably be reduced to DXS294 for mutation analysis but remains DXS105 for genetic counselling purposes. The risk to the male fetus (CVS6) was the same as that to CVS1 and CVS2, however their brother CVS7 was excluded from having BFLS since he carried the grandpaternal haplotype at nine informative marker loci.

A candidate gene in the Xq26-27 region, SOX3 (SRY-related HMG-box gene), maps to about 5Mb between DXS51 (52A) and DXS98 (4D-8). The murine equivalent Sox genes are embryonically expressed and may have a role in regulating the development of the nervous system (Stevanovic et al, 1993). SOX3 expression in human neuronal tissues implicates it as a candidate for BFLS. It is a simple locus with only one large exon (2.3kb transcript) and homology to SRY on the Y chromosome. Deletion of F9 and SOX3 in a mentally retarded male with haemophilia B suggests that SOX3 is not essential for sex determination/testis formation (Stevanovic et al, 1993). A similar deletion in a lymphoblastoid cell line LL556 derived from a mentally retardated haemophiliac (Rousseau et al, 1991) may suggest SOX3 is distal to F9. The HMG portion of SOX3 (Stevanovic et al, 1993) was examined for possible disease causing mutations by PCR, SSCA and sequencing in genomic DNA (PCR product of 283bp) by the methods outlined in Chapter 4.6.2.1.

Involvement of the HMG domain of this gene was excluded in one BFLS patient (Appendix III, Paper 7). Examination of the HMG domain alone, however, cannot rule out the SOX3 gene as a candidate for BFLS but exclusion by position would be more definitive. Given the recombination in CVS4 together with the possibility of carrier status from the phenotype of the feet, the BFLS gene may lie proximal to DXS294 and may thus be physically excluded from overlapping with the SOX3 gene. This gene may be a candidate for other mental retardation conditions in the region.

The BFLS gene in this family spans 24.6cM between DXS425 and DXS105, but may lie within the 15.5cM between DXS425 and DXS294. Regional localisation of BFLS on the background map of polymorphic markers will be useful to refine the mapping of overlapping syndromes and to elucidate possibly allelic disorders. BFLS at least partly overlaps the localisations of MRX6 and MRX27 (Chapter 4.3.2.). Others have mapped syndromal XLMR genes into the same region. An XLMR with severe mental retardation and associated microcephaly, optical atrophy, severe hearing defect, characteristic facies, epileptic seizures, spasticity and restricted joint mobility is mapped between DXS424 and DXS297 (Malmgren et al, 1993). Affecteds demonstrate pre- and post-natal growth retardation and early death in infancy or childhood in a pedigree suggestive of X-linked recessive inheritance (Gustavson et al, 1993). Another XLMR syndrome, assigned the symbol MRXS5, is localised between DXS425 and F9 (Huang et al, 1991). This syndrome of severe mental retardation is associated with early hypotonia, spasticity, seizures, characteristic facies and Dandy Walker malformation (Pettigrew et al, 1991). Comparison between syndromes with this range of clinical manifestations associated with mental retardation (usually severe) suggests the possibility of a contiguous gene disorder. This would require the presence of a series of overlapping submicroscopic deletions or some other genetic mechanism affecting more than one locus. If this were the case then the MRX6 and/or the MRX27 gene(s) may be the contributor of mental retardation to the phenotype. No deletions associated with BFLS have yet been detected. Variable expression between families with BFLS, originally described with severe mental retardation, may be accounted for by allelic heterogeneity given a lack of evidence for locus heterogeneity.

XLMR with heterozygote expression and macrocephaly

Regional localisation of the gene and clinical description in this large family with variably retarded affected males in two generations and mild expression in heterozygotes has been recently published (Appendix III, Paper 6). The cardinal findings suggest a clinically distinct entity with macrocephaly evident in both hemizygous and heterozygous individuals but not in their unaffected relatives. The condition in this family bears some

similarities to the Atkin-Flaitz syndrome with macrocephaly, coarse facies and hypertelorism (Atkin et al, 1985). It remains to be determined by linkage mapping of the latter whether the regional localisations overlap or exclude the possibility that the disorder in both families might be caused by mutation in the same gene. Since the possibility of allelism of Coffin-Lowry syndrome (CLS) and Atkin-Flaitz syndrome had been raised earlier (Turner et al, 1986), markers in Xp22-pter were also analysed.

Pairwise lod scores were conservatively estimated by exclusion of individuals of uncertain affection status based on clinical findings. Both IV-12 and IV-14 are mildly retarded and slow at school (Figure 1; Appendix III, Paper 6). The girl, IV-12, has a head circumference on the 95th percentile like the obligate carriers in the family, while the boy has a normal head circumference and is unlike the other affected males. Penetrance was arbitrarily set at 0.75 as a proportion of the females express features of the phenotype. Peak lod scores of 3.31 were acheived at MAOB, DXS991 and AR. Gene localisation is pericentromeric, with recombination events consistent with the placement of the gene in the 32cM interval between the DXS1068 locus on Xp and DXS1125 locus on Xq. The CLS region near Xpter was excluded, as was the region linked to hydrocephalus in Xq28.

The greatest chance of single crossover disabling a diagnosis is 32% in this family while double crossover might occur in less than 2.5% of cases. Haplotype construction within the DXS1068 to DXS1125 interval defined the haplotype common to all affected males (III-2, III-14 and IV-16). Affected and/or obligate carrier females (III-13, III-16 and IV-12) also shared the same haplotype and no double recombinants were detected. In the absence of DNA from family members in generations I and II, the haplotypes of common ancestors and structure at the top of the pedigree cannot be verified. Based on the present gene localisation and difficulties in inferring derivation of haplotypes, potential female carriers and their daughters requesting information were counselled regarding only exclusion of the affected haplotype and hence carrier status.

5.4. Discussion

Refined genetic localisation of the specific MRXS gene in each of five families with clinically different XLMR entities has been described. The localisations have reached the limits of resolution by linkage given the pedigree structures and application of a finite set of polymorphic markers. In the four families where the genes had originally been localised by RFLP, refinement into gene intervals of less than 25cM each was accomplished by subsequent application of microsatellite markers. In three families defining distinct clinical entities (WTS, SHS and XLMR with macrocephaly) the gene intervals are pericentromeric and at least partially overlap one another. The apparent clustering of MRX loci at the

centromere has also been noted (Chapter 4), raising the possibility of a region containing one or more genes critical for normal cognitive function and brain development. The phenotypic differences between MRX and MRXS may represent allelic mutations causing differential splicing or differential expression of the gene product. Examples of allelic heterogeneity have been demonstrated elsewhere by a series of private mutations in the L1 gene at Xq28 in association with seemingly unrelated clinical entities (section 5.1.). Therefore genes with overlapping localisations that may once have been considered as defining clinically distinct entities, (not necessarily only those defining XLMR syndromes), cannot be excluded from allelism.

Recent linkage mapping in the original family described by Renpenning in 1962, has demonstrated pericentric localisation of this gene between the DMD and AR loci (Schwartz et al, 1994). The phenotype of the Renpenning syndrome (MIM#309500), has some similarities to the SHS family regionally localised above, including short stature and microcephaly but with the exception of spastic diplegia. Overlap of these gene localisations are suggestive that these conditions may be allelic mutations of the same gene. Pericentric localisations have also been described in two further families; MRX8 (Schwartz et al, 1992) and K8240 with retarded males having small head size, small testes and shorter than average stature (Schwartz et al, 1994). Retardation was severe in the Renpenning and K8240 families, ranged from mild to severe in MRX8 and from borderline to moderate in SHS.

Another syndromal XLMR condition with associated short stature, deafness, hypogonadism and ocular abnormalties known as the Juberg-Marsidi syndrome (MIM#309590) had been clinically differentiated from the SHS syndrome on the basis of deafness and midface hypoplasia not seen in the latter (Appendix III, Paper 1). The recent mapping of the Juberg-Marsidi syndrome (JM) to Xq12-q21 (Saugier-Veber et al, 1993) overlaps SHS at less than the 3cM between DXS159 and DXS453, not sufficient to genetically delineate these disorders. Mutations in a putative global transcription regulator XH2, have been demonstrated in an XLMR syndrome with α-thalassemia (ATR-X, MIM#301040) and present a candidate for overlapping XLMR syndromes (Gibbons et al, 1995). XH2 (distal to DXS453) is physically excluded as a candidate for SHS. Certainly the gene mapping and phenotypic overlap between JM and ATR-X suggest lumping of these syndromes that are worthy of further investigation (Saugier-Veber et al, 1995).

Many Norrie (NDP) patients manifest mental retardation and atypical forms have been described with hypogenitalism and/or microcephaly (Sims et al, 1992). The NDP gene has been mapped to a 150kb region centromeric to the MAOB microsatellite within intron 2 of the MAOB gene at Xp11.3. The SHS gene overlaps this region (as does MRX8) and has clinical features in common with NDP including microcephaly and small testes. Reduction

of the WTS gene localisation however, no longer overlaps this region. Some of the clinical features of WTS were originally believed to resemble features of BFLS. Gene localisations have since delineated both these syndromes as separate conditions (Appendix III, Papers 2 and 3). The best known mental retardation syndrome mapping in the vicinity of BFLS is the Lesch-Nyhan syndrome (MIM#308000) caused by deficiency of the HPRT enzyme. Lesch-Nyhan individuals have associated spastic cerebral palsy as well as self-destructive behavioural anomalies and are clinically clearly distinct to BFLS. Other XLMR syndromes that overlap the BFLS gene interval may be involved in a contiguous gene syndrome or represent allelic variants (see BFLS in section 5.3.). Screening of the neuronally expressed SOX3 HMG domain has excluded mutations in this candidate locus from causative effects in BFLS.

During the 1980's genes were localised using several available RFLP markers, though the most valuable were the more heterozygous. Genes were frequently localised by positive lod scores to the VNTRs; DXS255(M27B) in Xp11, DXS52(St-14) Xq28;. or to DXYS1 at Xq21. Early localisations to the latter, for example (see SHS in section 5.3.), have been demonstrated to be useful indicators of linkage that may be adjacent though quite distant to the true localisation. Carrier risks and prenatal diagnosis can now be given to individuals in these families based on informative markers flanking the gene localisation. The mapping information established in this chapter has already been applied to carrier detection and prenatal diagnosis in these families and made a significant impact in the lives of several individuals faced with the dilemma of reproductive choice.

5.5. Conclusion

Separate locations of clinically similar conditions are sufficient evidence to say that they are not allelic mutations of the same gene, but clinically different disorders with common localisations do not necessarily exclude allelism. Whether a syndrome represents a contiguous gene defect or pleiotropy in a single gene cannot be determined by the mapping of the gene in a single family. Evidence that widely expressed genes (such as XH2) can have pleiotropic effects throughout development and may have a role in development of the CNS suggest that candidate genes for XLMR syndromes can be ubiquitously expressed.

The regional localisations of these syndromal XLMR genes define specific regions of interest for further analysis. Overlapping localisations in clinically similar conditions, though not representing evidence of locus homogeneity, may be useful in refining the candidate loci. Gross structural rearrangements of this region of the chromosome through translocation or deletion, particularly in similarly afflicted individuals, will expedite positional cloning of the gene as was the case for the identification of FMR2, the first MRX

gene to be isolated (Chapter 7). The markers used in these studies (Chapter 3), failed to detect submicroscopic deletions, which might be the basis for the clinical defect in at least a proportion of the MRXS families studied. None of these gene localisations by linkage have yet led to the identification of the gene by the candidate gene approach, however only the HMG portion of the SOX3 gene can be excluded from involvement in BFLS. The refinement of each MRXS gene interval within the parameters defined by this thesis, still suggest up to 700 possible candidate genes to be screened for each disease causing mutation (Chapter 4.6.). This seemingly daunting approach has successfully led to the identification of the gene for Barth syndrome (Chapter 6), ATR-X syndrome (Gibbons et al, 1995) and most recently demonstrated allelic mutation of XH2 in Juberg-Marsidi syndrome (Villard et al, 1996).

CHAPTER 6

Other X-linked disorders

Note to Publications

Much of the work summarised in this chapter has been published, with clinical assessments carried out by Dr. L.C. Ades, Dr. M.W. Partington and Dr. M. Wilson. All linkage studies were carried out and prepared for publication by the candidate. Reprints are bound in Appendix IV in the following order:

1. Adès LC, **Gedeon AK**, Wilson MJ, Latham M, Partington MW, Mulley JC, Nelson J, Lui K, Sillence DO (1993) Barth Syndrome: Clinical features and confirmation of gene localisation to distal Xq28. Am J Med Genet 45:327-334.

2. **Gedeon AK**, Mulley JC, Kozman H, Donnelly A, Partington MW (1994) Localisation of the gene for X-linked reticulate pigmentary disorder with systemic manifestations (PDR), previously known as X-linked cutaneous amyloidosis. Am J Med Genet 52:75-78.

3. **Gedeon AK**, Wilson M, Colley AC, Sillence DO, Mulley JC (1995) X-linked fatal infantile cardiomyopathy maps to Xq28 and is possibly allelic to Barth syndrome. J Med Genet 32:383-388.

4. Bione S, D'Adamo P, Maestrini E, **Gedeon AK,** Bolhuis PA, Toniolo D (1996) A novel X-linked gene, G4.5, is responsible for Barth syndrome. Nature Genetics 12:385-389.

It is recommended that Papers 1, 3 and 4 be read conjoint with the section on cardiomyopathy (section 6.4.) and Paper 2 with reference to reticulate pigmentary disorder (section 6.3.).

6.1. Introduction

Mapping and refinement of gene localisations for all X-linked disorders follows the same methodology described in Chapters 4 and 5. Four large families segregating dissimilar X-linked conditions were available for gene mapping studies. Mapping of the genes in these families was aimed at contribution to the map of genes on the X chromosome and development of a reduced gene interval toward identification of the gene for each disorder by the positional candidate approach or ultimately by positional cloning.

This chapter is subdivided into three sections representing the major clinical involvement in these families. Mapping was undertaken through collaboration with the medical geneticists counselling each family. Dr Lesley Ades for Barth syndrome (BTHS), Dr Meredith Wilson for fatal infantile cardiomyopathy in Sydney, Australia and Dr Michael Partington for an X-linked pigmentary disorder diagnosed in Kingston, Canada. Publication of the gene localisations in the cardiomyopathy families opened the way to collaboration with groups creating transcription maps of Xq28; Dr Daniela Toniolo of Pavia, Italy and Dr Aïda Metzenberg of San Francisco, USA and has resulted in the isolation of the putative gene responsible for Barth syndrome (section 6.4.3.).

6.2. Lutheran suppressor gene (XS)

Mulley et al (1988), localised an X-linked suppressor gene (XS) for the Lutheran blood group antigens in the only known family segregating the rare suppressor allele XS2. The maximum pairwise lod score was Zmax=1.96 (θ=0.0) suggesting linkage of XS with DXS14 close to the centromere. Genotypes of common ancestors in generation I were inferred from those of generation II (Figure 6.1). Three obligate carriers in generation II were homozygous at this locus and therefore uninformative for determining phase in generation III. The XS gene was located between DXS84 and DXYS1, the closest markers flanking DXS14. Reduction of this localisation spanning over 45cM on the background map (Appendix I, Paper 3), is now presented.

Twenty loci mapping within the boundaries of the XS gene interval were genotyped in this family. Three of these at DXS426, PGK1P1 and the VNTR DXS255 were not informative in females in generation II. Pairwise lod scores were calculated using MLINK between XS and the remaining 17 markers and at the linked RFLP marker DXS14 (Table 6.1). The obligate heterozygotes were homozygous at DXS1126, DXS227, DXS566 and DXS986 rendering the pedigree phase unknown at these loci. The peak lod score was Zmax=3.69 (θ=0.0) at DXS573. The critical region defined by recombinants is demarcated

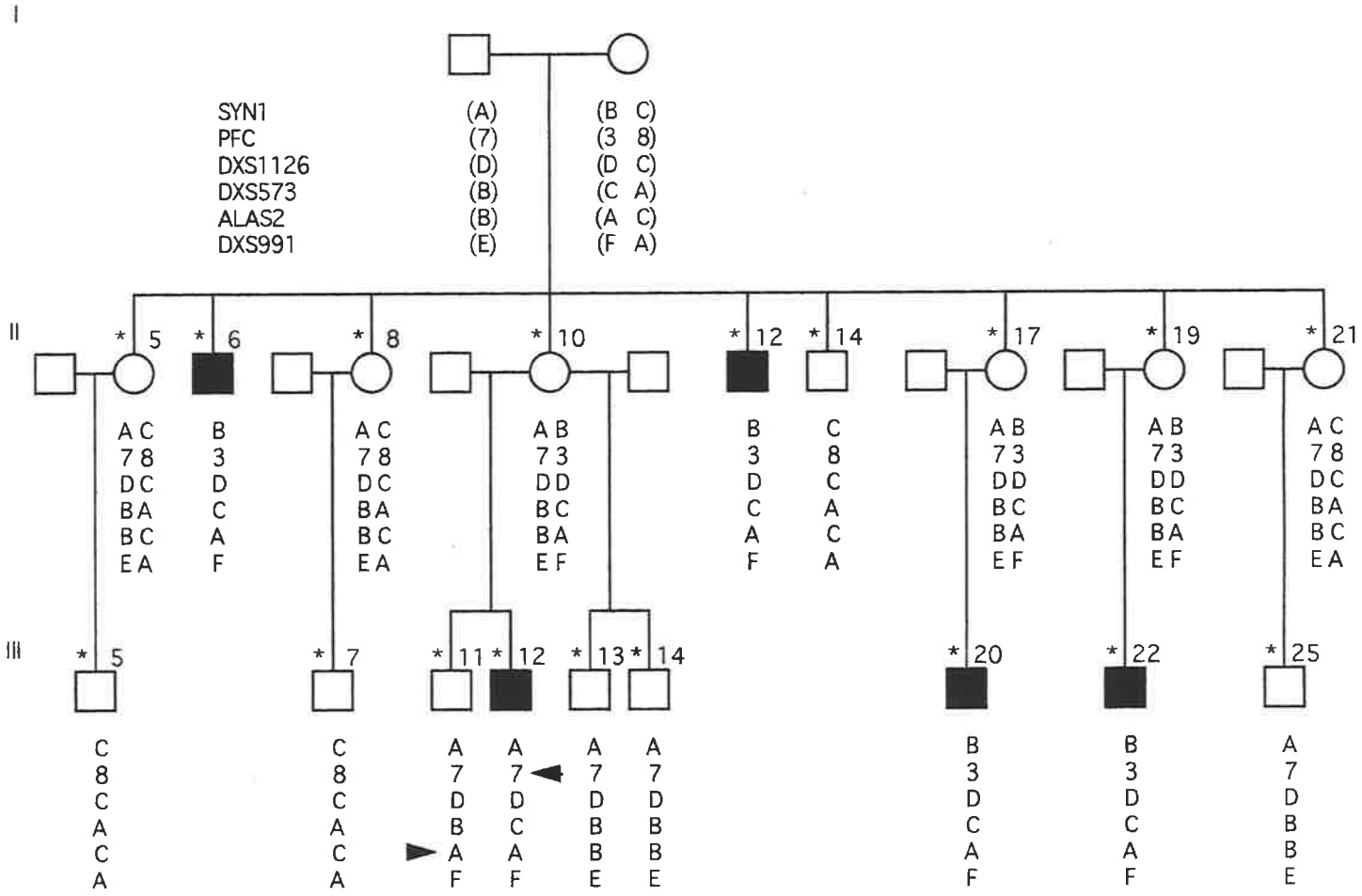


Figure 6.1. Pedigree of family segregating *Lu* suppressor gene (XS).

TABLE 6.1: PAIRWISE ANALYSIS OF XS WITH X CHROMOSOME MARKERS WITHIN THE ORIGINAL GENE LOCALISATION

	θ								
Loci	0.001	0.01	0.05	0.1	0.2	0.3	0.4	Zmax	θ
DXS1068	-6.52	-3.55	-1.56	-0.80	-0.19	0.03	0.08	0.08	0.40
DXS228	-1.43	-0.44	0.19	0.41	0.51	0.44	0.27	0.51	0.20
MAOB	-6.22	-3.25	-1.28	-0.54	0.01	0.14	0.11	0.14	0.32
DXS1003	0.68	1.63	2.10	2.10	1.75	1.21	0.56	2.12	0.07
SYN	0.68	1.63	2.10	2.10	1.75	1.21	0.56	2.12	0.07
PFC	0.68	1.63	2.10	2.10	1.75	1.21	0.56	2.12	0.07
DXS1126	1.88	1.85	1.70	1.52	1.13	0.71	0.26	1.88	0.00
DXS573	3.68	3.63	3.38	3.05	2.35	1.58	0.73	3.69	0.00
ALAS2	0.68	1.63	2.10	2.10	1.75	1.21	0.56	2.12	0.07
DXS991	0.68	1.63	2.10	2.10	1.75	1.21	0.56	2.12	0.07
DXS14	1.88	1.85	1.70	1.52	1.13	0.71	0.26	1.88	0.00
AR	0.68	1.63	2.10	2.10	1.75	1.21	0.56	2.12	0.07
DXS1125	0.68	1.63	2.10	2.10	1.75	1.21	0.56	2.12	0.07
DXS106	0.68	1.63	2.11	2.12	1.78	1.25	0.58	2.14	0.08
DXS453	0.68	1.63	2.10	2.10	1.75	1.21	0.56	2.12	0.07
DXS559	-2.32	-0.37	0.82	1.14	1.15	0.85	0.41	1.20	0.15
DXS227	-1.67	-0.69	-0.07	0.12	0.20	0.14	0.05	0.20	0.19
DXS566	-1.12	-0.15	0.42	0.56	0.53	0.34	0.11	0.58	0.13
DXS986	-1.12	-0.15	0.43	0.58	0.56	0.37	0.13	0.60	0.14

by PFC distally and ALAS2 proximally. Since the recombinant at PFC was detected in the Lu deficient III-8, and that at ALAS2 in his brother III-7, the heterozygous genotype of their mother became a crucial factor that contributed to the refined mapping of the XS gene. This interval spans the cytogenetic bands Xp11.22 and Xp11.23 in their entirety and a genetic distance of 5.8cM, a significant reduction from the original localisation. Based on the premise that there are 2500-5000 genes on the 210cM X chromosome (Chapter 1.2.3.), between 70 and 140 genes might be estimated to map to this interval if gene distribution were uniform. A systematic database search for positional candidate genes may identify the XS gene.

6.3. Reticulate Pigmentary Disorder (PDR)

A large family described with X-linked cutaneous amyloidosis (MIM#301220) was first presented in 1981 (Partington et al, 1981) then clinically reviewed in 1989 (Partington and Prentice, 1989). The characteristic brown pigmentation of the skin follows the lines of Blaschko in females but is generalised and reticulate in males. Affected males in earlier generations suffered severe gastrointestinal disorders in infancy with failure to thrive and pneumonia causing early death. Improved management has improved survival and revealed other manifestations such as corneal dystrophy with severe photophobia or chronic respiratory disease. Amyloid deposits found in the pigmented skin of adults named the disorder, however later studies failed to show amyloid deposition in children. Another family, believed to have the same disorder, was described recently (Ades et al, 1993a) and has caused the condition to be renamed X-linked reticulate pigmentary disorder with systemic manifestations and with the gene symbol PDR.

DNA samples from the original family of Partington (1981) were acquired from Kingston, Ontario through collaboration with Dr. Michael Partington. Linkage analysis in this large family localised the PDR gene to Xp21-p22 (Appendix IV, Paper 2). An affected female was found to be 47,XXX and was omitted from the analysis. Genotype data from the remaining 17 individuals in the pedigree were analysed under an assumed penetrance of 0.75. Two-point lod scores exceeded +2 at 8 marker loci between DXS999 and DXS228. This interval of more than 40cM (Chapter 3) spans the entire dystrophin locus but is potentially amenable to reduction by up to 11cM proximally. The extent of clinical variation and clarification of allelic versus genetic heterogeneity will only be resolved when the gene is cloned. The size of the family described by Ades (1993a) was not sufficient for gene localisation by linkage analysis. Stored samples will be directly tested once the PDR gene is cloned. Until then genetic heterogeneity cannot be excluded.

Other X-linked pigmentary disorders include the two genetically distinct loci for incontinentia pigmenti (IP) Type I and II (Davies et al, 1991). A potentially more interesting microdeletion syndrome has been described in Xp22.3, with reticulolinear skin defects and micropthalmia. All five known patients have cytogenetically detectable translocations or deletions of Xp22.3, and none had a normal Y chromosome (Lindor et al, 1992). A clinician may be able to determine from the clinical description of these initially erythematous linear and reticular lesions, that heal to brown streaks of pigmentation on the face and neck of these patients, whether they can be distinguished from the pigmentation in PDR. The marker at the distal limit (DXS999) of the PDR gene (above) is at the boundary of the Xp22.13 and Xp22.2 chromosomal bands. Resolution of the chromosomal preparations in Lindor et al (1992) are not of sufficient quality to confirm that the breakpoint was distal to Xp22.2 in their patient. This illustrates the advantages of a multidisciplinary approach required for reliable gene mapping. These cases with chromosomal rearrangements may provide a means for cloning of the gene involved and differentiation or identity with the PDR gene.

6.4. X-linked cardiomyopathy/endocardial fibroelastosis (EFE)

Cardiomyopathy strictly means 'disease of the heart muscle' and is involved in many diseases as a secondary effect. Family studies have shown, however, that there is a genetic component to the high rate of idiopathic myocardial diseases. Endocardial fibroelastosis (EFE) is a non-specific histological finding in both contracted and dilated forms of X-linked cardiomyopathy documented under MIM#305300 (OMIM). The etiology of cardiomyopathies, particularly those that are autosomally inherited, is heterogeneous despite the common phenotype. Identification of some of the underlying gene defects has shown that cardiomyopathy is caused by disorders of cardiac energy metabolism or by abnormalities of myocardial contractile and structural proteins (Kelly and Strauss, 1994). Defects in mitochondrial oxidative phosphorylation have been demonstrated in several myocardial disorders due to deletions and point mutations of the mitochondrial genome. Several mtDNA and tRNA mutations have been described (Silvestri et al, 1994) that can be identified by maternal inheritance and can occur alone, as fatal infantile cardiopathies, or associated with skeletal myopathy.

The description of families with a distinct X-linked inheritance of cardiomyopathy has lead to the clinical delineation of Barth syndrome (BTHS) (Barth et al, 1983). BTHS (MIM#302060) is characterised by skeletal myopathy, short stature, abnormal mitochondria and recurrent neutropenia in association with cardiomyopathy of variable severity that is often fatal in childhood. Affected males in the family described by Barth, died between 3 days and 31 months of age from cardiac failure or infection due to neutropenia. Cardiac

pathology showed ventricular hypertrophy and dilatation. Mitochondria with abnormal cristae, decreased plasma carnitine and growth retardation were other described features. Progression of the disease is variable, but with standard anti-cardiac failure regimens a child may gradually stabilise or improve, as does the severity of infections (Christodoulou et al, 1994). Mitochondrial abnormalities suggest that the primary defect may lie in cardiac energy metabolism and imply nuclear-encoded defects in mitochondrial oxidative phosphorylation (Kelly and Strauss, 1994). On the other hand, X-linked dilated cardiomyopathy (designated XLCM) is a rapidly progressive primary myocardial disorder presenting in teenage males as congestive heart failure usually with no skeletal disease but decreased concentration of dystrophin in cardiac muscle (Towbin et al, 1993). It has been proposed that this disorder is due to a defective structural protein, namely dystrophin or a specific cardiac isoform of dystrophin since cardiac involvement occurs frequently in patients with DMD, although only 10% are estimated to die of primary heart failure (Moser, 1984). The clear X-linked inheritance pattern demonstrated in such families, has encouraged attempts at gene mapping by linkage.

6.4.1. Mapping X-linked cardiomyopathy

X-linked cardiomyopathy has been recently shown to be heterogeneous with the localisations of two distinct gene regions at Xp21 and at Xq28. BTHS was independently localised in two large families (one of these by this candidate) to a cluster of markers in Xq28 (Bolhuis et al, 1991; Appendix IV Paper 1). Mapping data were consistent with genetic homogeneity of Barth syndrome. The XLCM gene has been recently delineated from BTHS by linkage to the proximal 5' portion of the dystrophin locus at Xp21 (Towbin et al, 1993). Deletion of the muscle promoter and first exon of dystrophin have been described in a family with XLCM but no skeletal muscle weakness (Muntoni et al, 1993). Although some clinical differences do exist, including the age at clinical presentation and the variable severity within and between families, the XLCM and BTHS loci are most clearly delineated by linkage analysis.

Two large families segregating X-linked cardiomyopathy were studied: one (Fam T) with clinical features consistent with BTHS, the other (Fam H) an apparently more severe fatal infantile form of cardiomyopathy. The genetic localisation, in Fam T, was carried out and described by the candidate (Paper 1, Appendix IV) and confirmed an initial report of linkage of BTHS to Xq28 (Bolhuis et al, 1991). The closest proximal recombination in Fam T was established at DXS374 (1A1.1) and established a reduced proximal limit for the BTHS gene assuming both families have the same syndrome. No distal limit could be determined defining the BTHS gene interval between DXS374 and qter with a peak lod

score Zmax=2.78, θ=0.0 at DXS52 (St14). Since the publication of Family T in 1993, further PCR based markers in Xq28 have been genotyped for refinement of the localisation. The markers distal to DXS374 are: cen - GABRA3- DXS52- p26, p39, DXS15- DXS707- DXS605- F8C- DXS1108- DXYS154- tel (Chapter 3). No recombinant meioses were identified to define the localisation in Xq28 as distinct from the telomere or to further reduce the extent of the gene interval proximally. Another family with clinical signs of BTHS but only **two** surviving males, Fam W, was too small for independent linkage analysis. No recombinants were detected with markers within the BTHS interval, thus although linkage cannot be demonstrated locus homogeneity cannot be excluded. Members of this family were included in collaborative efforts to identify the BTHS gene (section 6.4.3.).

The second large family, Fam H, segregates a fatal infantile form of cardiomyopathy. Linkage analysis was undertaken to determine the gene location and determine evidence for heterogeneity of this locus. Clinical and linkage findings have been reported in Paper 3, (Appendix IV). Analysis of recombinants and peak lod scores showed that the cardiomyopathy gene in this family was not linked to dystrophin (as is XLCM), but segregated with telomeric markers in Xq28. A recombinant with DXS296 defined the proximal limit of the cardiomyopathy gene of Fam H which lies distal to the DXS296 locus. The distal limit to the localisation is the telomere since recombination has not been observed nor can be inferred with any markers distal to DXS52. Map distances between loci in Xq28 were:DXS296-0.001-DXS1113-0.10-DXS52-0.008-p26,p39,DXS15-0.001-DXS707-0.001-DXS605-0.001-F8C-0.005-DXS1108-0.0001-DXYS154. Although linkage analysis in Xq28 did not give any evidence for locus heterogeneity of BTHS and the infantile form of X-linked cardiomyopathy, the existence of two different genes close to each other cannot be excluded since this region of Xq28 is 5-6Mb long (~11cM) and is extremely gene rich. To determine whether the genes in these two families are phenotypic variants of mutations within the same gene or caused by different genes requires the identification of at least one of the genes involved.

6.4.2. Clinical findings

The detailed clinical descriptions of both large families were given by the clinician involved in each case, Family T (Dr.Ades) and Family H (Dr.Wilson), in the respective references (Ades et al, 1993b and Gedeon et al, 1995; Papers 1 and 3, Appendix IV). The phenotypic expression of each disorder varies widely. Family T, had many features of Barth syndrome - all affected males had dilated cardiomyopathy, some with EFE. A spectrum of

age of onset and severity of symptoms were described with a maximum survival age of 10yrs 3mths currently. All had normal CPK, some mild myopathic changes were evident, neutropenia was present from birth, mitochondrial structure was normal but there was a possibility of mitochondrial respiratory chain dysfunction. The clinical data suggest that growth retardation is a secondary feature with postnatal onset which may explain why this was not apparent in Fam H.

Early neonatal death in Family H, made the retrospective assessment of earlier generations difficult. Six male infants (IV-22, IV-23, V-1, V-3, V-5 and V-9), were known to have died with hypertrophic cardiomyopathy however the very early death in each case did not allow for extensive documentation of features such as short stature, skeletal myopathy and neutropenia consistent with BTHS. There is a strong family history of unexplained death of infant males, all less than 6 months old, occurring over at least four generations. Affected males are related through apparently healthy women (IV-1, IV-8, III-1, III-8 and III-10) in a pattern consistent with X-linked recessive inheritance. In 3 infants with neonatal onset of cardiac failure, echocardiography documented non-obstructive ventricular hypertrophy and dilatation. Associated features included a mild talipes equinovarus in two infants and slight neutropenia in one, but there was no obvious skeletal myopathy in the neonatal period.

The age of onset and severity of the disease course are the most apparent clinical differences in these two families. The linkage data have been unable to detect genetic heterogeneity, since the gene in each family has been localised to Xq28. The possibility of allelism to BTHS, will be addressed once mutation in the causative gene(s) are demonstrated conclusively (section 6.4.3.). Assessment of different therapies is difficult when the underlying biochemical defect is not known, but once the gene and mutations have been revealed a suitable therapy may be devised to address the specific nature of the defect. The candidate genes will most likely be involved in mitochondrial respiratory chain function, but structural protein and myosin mutations should not be overlooked. Another recently cloned neuromuscular disease mapping to Xq28 is the Emery-Dreifuss muscular dystrophy (EMD) (Yates et al, 1993; Bione et al, 1994). A characteristic of EMD (MIM#310300-X-linked recessive and MIM#181350-autosomal dominant) is cardiac involvement with conduction defects, however both of these disorders are clinically and histologically distinguishable from BTHS-type cardiomyopathy.

6.4.3. Identification of the BTHS gene

Over two dozen genetic diseases are known to lie in Xq28, most of them at low frequency. The responsible gene has not been identified in most due to small family sizes

and lack of cytogenetic aberrations useful for positional cloning. Two of the research groups involved in construction of transcription maps of Xq28, endeavour to find the disease-causing mutations by examining candidate genes by virtue of their map position. Initially, fibroblasts from an affected male and representative DNA samples with BTHS (Fam T) were tested for abnormalities of genes in Xq28. The San Francisco group of Dr Aïda Metzenberg had isolated genes and CpG-rich islands in and around the F8 gene near the telomere. One candidate gene (MPP-1), encoding a protein (p55) with potential involvement in cytoskeletal stability (Metzenberg and Gischier, 1992) was not involved in BTHS, EMD or dyskeratosis congenita (Metzenberg et al, 1994).

The group of Dr Daniela Toniolo of Pavia had cDNAs proximal to the G6PD gene. The rationale involved searching for mutation in one of five genes expressed at high levels in cardiac and skeletal muscle selected from the transcription map of distal Xq28 (Bione et al, 1993). The filamin gene which encodes a structural protein (Patrosso et al, 1994; Gariboldi et al, 1994) did not detect variation from the normal patterns on RT-PCR and Northern blot. Later, Northern blot analysis of total RNA extracted from the fibroblast line showed complex 'differences' in the pattern of expression of two adjacent genes when compared to normals. Fibroblast cell lines were sent from two affecteds from families T and H, and from one surviving adult male with a heart transplant, GW, from family W (with insufficient meioses for gene mapping by linkage). Lymphoblast cell lines were also established from one of the surviving boys in family T and in GW.

Mutations in the gene G4.5, encoding the putative proteins termed tafazzins, introduce stop codons into the open reading frame disrupting protein production in BTHS patients (Appendix IV, Paper 4). In patient GW, a G to A base substitution caused incorrect splicing of intron 2 with a one base shift and introduced a new stop codon downstream, while in OAT (from Family T) insertion of an additional base within exon 7 also introduced a frameshift and a stop codon (Appendix IV, Paper 4). Similar mutations were found in two Dutch families one of which is the original family defining BTHS. Segregation of these mutations has been demonstrated within each family confirming association with the disorder and enabling future diagnostic predictions to be made by direct mutation analysis. The immediate practical consequence of identifying the genetic 'lesion' in each family is that it becomes possible to devise prenatal DNA tests to determine whether an individual has inherited the gene defect and is therefore at risk for the disease. This is particularly beneficial to members of Family W that could not be diagnosed by linkage.

Different phenotypes or severity of phenotypes can be caused by mutations in different parts of a gene because of disruption of specific functional domains of the encoded protein. Further analyses of the G4.5 gene in Family H with severe infantile cardiomyopathy

have also shown mutational changes (personal communication, D Toniolo). This confirms that these phenotypically different conditions are allelic mutations at the same locus and initiates speculation on the functional domains of the protein.

6.5. Conclusion

Genetic localisation of PDR between the markers DXS228 proximally and DXS999 distally currently spans a broad interval (section 6.3.). Progression toward gene identification may involve refinement of the map position that will reduce the numbers of potential positional candidates to be screened. The rare condition of Lutheran deficiency with lack of expression of Lu blood group antigens is not clinically serious. Nevertheless, refined mapping of the XS gene confirmed and established the evidence for linkage in Xp11 near the centromere and contributes to the map of genes on the X chromosome.

The efforts in this and previous chapters have been directed at localisation and refinement of gene intervals causing X-linked disorders to the smallest regions possible within constraints established by individual family sizes and a finite set of markers. Although these gene localisations have been applied to provide carrier risk estimations for genetic counselling (Chapter 4 and 5), none had as yet led to identification of the gene until now. Identification of the causative gene for BTHS permits diagnosis with the direct test for the specific mutation in these families (T and W) with BTHS (6.4.3.). Further work is underway to determine the mutation in the more severe form of cardiomyopathy mapped to Xq28 (Fam H) which is allelic to BTHS. Ultimately, functional studies of the tafazzin gene may reveal appropriate therapies for survival of these patients but also elucidate the normal function. This work is an example *quid pro quo* of the progression from genes localised by linkage to the identification of these genes by the positional candidate approach. It is also a practical demonstration of the use of molecular methods to contribute to lumping (and nosology) of clinically distinguishable syndromes. This aim can also be achieved through isolation of the gene by positional cloning (Chapter 7).

CHAPTER 7

Submicroscopic Deletions Of The X Chromosome Associated With Mental Impairment

Note to Publications

Much of the work summarised in this chapter has been published. Research and development of reagents for direct molecular diagnosis of the fragile X lead to the detection of submicroscopic deletions in the region. These cases were analysed and prepared for publication by the candidate. Reprints are bound in Appendix V in the following order:

1. Sutherland GR, **Gedeon AK**, Kornman L, Donnelly A, Byard RW, Mulley JC, Kremer E, Lynch M, Pritchard M, Yu S, Richards RI (1991) Prenatal diagnosis of fragile X syndrome by direct detection of the unstable DNA sequence. New Engl J Med 325:1720-1722.

2. Yu S, Mulley J, Loesch D, Turner G, Donnelly A, **Gedeon A**, Hillen D, Kremer E, Lynch M, Pritchard M, Sutherland GR, Richards RI (1992) Fragile-X syndrome: Unique genetics of the heritable unstable element. Am J Hum Genet 50:968-980.

3. Mulley JC, Yu S, **Gedeon AK**, Donnelly A, Turner G, Loesch D, Chapman CJ, Gardner RJM, Richards RI, Sutherland GR (1992) Experience with direct molecular diagnosis of fragile X. J Med Genet 29:368-374.

4. **Gedeon AK**, Baker E, Robinson H, Partington MW, Gross B, Manca A, Korn B, Poustka A, Yu S, Sutherland GR, Mulley JC (1992) Fragile X syndrome without CCG amplification has an FMR1 deletion. Nature Genetics 1:341-344.

5. Loesch DZ, Huggins R, Hay DA, **Gedeon AK**, Mulley JC, Sutherland GR (1993) Genotype-phenotype relationships in Fragile X syndrome: A family study. Am J Hum Genet 53:1064-1073.

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

7. **Gedeon ÁK**, Keinänen M, Adès LC, Kääriäinen H, Gécz J, Baker E, Sutherland GR, Mulley JC (1995) Overlapping submicroscopic deletions in Xq28 in two unrelated boys with developmental disorders: identification of a gene near FRAXE. Am J Hum Genet 56: 907-914.

8. Gécz J, **Gedeon ÁK**, Sutherland GR, Mulley JC (1996) Identification of the gene FMR2, associated with FRAXE mental retardation. Nature Genetics 13:105-108.

7.1. Introduction

Naturally occuring chromosomal rearrangements such as deletions, translocations or inversions, when associated with a disease entity, may pinpoint the location of the defective gene (Chapter 1.3.3). The gene may be within the deletion or at the cytogenetic breakpoint of the inversion or translocation and can be exploited as a most useful resource for gene isolation by positional cloning. Detection of a deletion will in most cases facilitate gene identification. The clinical features manifested are presumed to be a direct consequence of the deletion of a specific gene or genes contributing to the phenotype. When individuals with a chromosomal abnormality show no major associated phenotype however, the sequences within the disrupted region can be assumed to be relatively void of vital genes. Based on the theoretical number of genes in the human genome (Chapter 1.2.3.), if genes were uniformly distributed they would occur on average every 30 - 60kb. It is known however that gene distribution is not uniform, even within gene dense chromosomal subbands such as Xq28 (Willard et al, 1994), therefore the number of putative genes in a deleted region cannot be accurately predicted.

Deletion mutations of >20bp are most commonly observed in genes with highly repetitive sequence motifs or large relative gene length. In human α-globin and growth hormone gene clusters, large deletions occur by homologous recombination between elements (segments) of greater than 200bp of high (>95%) sequence homology or between Alu sequences with 20-40bp of identical sequence (Ketterling et al, 1994). On the X chromosome a family of low copy number repeats detected by the probe CRI-S232 at Xp22.3 have been demonstrated to be frequently involved in interstitial deletions of physically close genes for single mendelian traits. Cloning of deletion breakpoints in patients with STS (steroid sulphatase) deficiency has shown abnormal pairing and homologous recombination between different repeat units (Ballabio and Andria, 1992). Since most genes however, do not contain long highly homologous segments and are not excessively large, deletion mutations only account for 5-10% of all mutations that cause disease (Wicking and Williamson, 1991). Short regions of sequence identity have been sequenced at the deletion junction in a patient with Lesch-Nyhan syndrome, but with no evidence for homologous recombination (Fuscoe and Nelsen, 1994). Slipped mispairing and transient misalignment of broken ends are mechanisms that have been suggested to cause deletions. The use of polymorphic repeat sequences for the genetic mapping of several X-linked disorders in this study however, did not reveal any deletions as had been hoped at the outset.

The phenotype of a patient carrying a deletion on the X chromosome depends on the gene involved and the sex of the affected individual. In males, nullisomy for any functionally important gene results in an abnormal phenotype, whereas in females an abnormal phenotype would only be evident when the deletion causes monosomy for a dominant trait. Deletions on the X in nullisomic males are very useful for gene mapping since the phenotype reflects the missing gene, or genes in a contiguous gene syndrome, and their extent can be delineated as a prelude to gene identification by positional cloning.

A number of X-linked disorders are commonly caused by deletions. Deletions in the dystrophin gene spanning 2.4Mb in Xp21, account for approximately 60% of patients with DMD and BMD. Typical Norrie disease syndrome (NDP) patients have been shown with partial deletions in the MAO gene region (Chen et al, 1993). Deficiency of HPRT due to large genomic alterations account for up to 15% of Lesch-Nyhan syndrome (LN) mutations (Fuscoe and Nelsen, 1994). At Xp22.3 panels of deletion cases have been collected and used to help physically map genes and loci in this pter region (Ballabio and Andria, 1992). Identification of a patient with a terminal deletion of 10Mb in Xp22.3 resulted in the definition of a contiguous gene syndrome including 6 monogenic disorders (Meindl et al, 1993). The presence of position effects were ruled out by a viable isolated deletion resulting in a single mendelian trait (Bassi et al, 1994). Microscopically visible male-viable deletions of Xq21 also cause contiguous gene syndromes with complex phenotypes including MRX, DFN3 and choroideremia (CHM). The order of the three genes involved in have been revealed by phenotype analysis of patients with deletions (Huber et al, 1994: Bach et al, 1992). In many other disorders the discovery of a deletion can occur after disease gene identification, for example the majority of mutations causing haemophilia B are single-base substitutions or microdeletions in the factor IX gene (Ketterling et al, 1993 and 1994), while single base pair to large deletions in exons of the COL4A5 gene cause Alport syndrome (Renieri et al, 1994). The detection of deletion mutations thus contribute primarily to the localisation and identification of genes, but also to elucidation of their components vital to normal function.

The following chapter describes discoveries of deletion mutations detected by the diagnostic reagents used for direct molecular diagnosis of the fragile X syndrome. The two probes used routinely for this testing were subcloned and described by Dr. Sui Yu (WCH) as part of her doctoral thesis. The candidate had some involvement in establishment of these probes for this purpose in fragile X families although this was not an integral part of this thesis. The characterisation of deletion mutations at FMR1 and at FMR2 are summarised in this chapter, detailed reports have been published (Appendix V).

7.2. The fragile X syndrome

Fragile X syndrome is the most common form of familial mental retardation in males and is associated with a rare fragile site (FRAXA) at Xq27.3 (Chapter 1.6.1.). In 1991, cloning of the fragile X mental retardation 1 (FMR1) gene and disclosure of an unstable CCG trinucleotide repeat within the 5' UTR, marked a turning point in the diagnosis of the fragile X syndrome. Abnormal expansion of the CCG repeat was found to be responsible for the overwhelming majority of fragile X patients (Fu et al, 1991) although the phenotype reflects methylation-reduced transcription of the FMR1 gene (Pieretti et al, 1991). Methylation of the CCG repeat has been shown to prevent binding of a specific nuclear protein (CCG-BP1) and may be a basis of the fragile X phenotype (Richards et al, 1993). Cytogenetic expression of the folate-sensitive fragile site or genetic risk analysis using closely linked flanking markers had been the most reliable laboratory means to confirm a clinical diagnosis. A clinical scoring method (rated out of 10) had also been described to assess the expression of the Martin-Bell phenotype (Laing et al, 1991).

Several laboratories developed reagents for a direct molecular diagnostic test for the fragile X genotype. The probes StB12.3 (Oberle et al, 1991), Ox1.9 (Hirst et al, 1991) and pfxa3 (Yu et al, 1991) all detect abnormal amplification within the same 5.2kb EcoRI fragment. Isolation of the 520bp pfxa3 fragment immediately distal to the unstable element, as a diagnostic reagent, followed from demonstration of variability between normal and affected individuals and strong hybridisation to a single PstI restriction fragment of genomic sequences (Yu et al, 1991). Direct molecular diagnosis of the fragile X could now be performed by estimation of the CCG amplification on Southern analysis and hybridisation with the pfxa3 probe. Another anonymous fragment, pS8, was simultaneously included in the diagnostic protocol as an internal positive control to identify partial digestion of patient samples. The pS8 probe was derived from a subclone of YAC XY539 in Xq28 containing the VK21 sequence of the locus DXS296 (Appendix V, Paper 1).

Digestion of DNA samples with PstI gives the necessary resolution for diagnosis by double hybridisation with pfxa3 and pS8. The pfxa3 probe detects a 1.0kb fragment in non-fragile X individuals and/or a fragment of higher molecular weight commensurate with the size of the amplification in affected and carrier patients. The 800bp pS8 fragment excised with PstI, detects only itself on Southern analyses and does not overlap the size distribution of fragments detected by pfxa3. Diagnostic reliability of this probe combination was established through studies correlating size of the amplified fragment with fragile X phenotype (Appendix V, Papers 2, 3 and 5). Prenatal diagnosis in a male fetus showed that the normal 1.0kb was missing and replaced by one of 2.3kb (Appendix V, Paper 1). Analysis

of methylation showed that all tissues were variably methylated in the fetus which was not reflected by the lack of methylation in the villus sample. Direct molecular diagnosis superseded cytogenetic or linked flanking marker analysis of the fragile X syndrome (Sutherland and Mulley, 1990).

7.2.1. Deletion of the FMR1 gene

A retarded boy, IV-4 (Figure 1; Appendix V, Paper 4), assessed on clinical screening for the fragile X, with a high score of 10 by up to 3 independent clinicians (section 7.2.), did not express the fragile X cytogenetically in direct contrast with the high clinical score. Direct molecular analysis using the pfxa3 and pS8 probe combination was used to determine whether any expansion of the CCG repeat could be detected. The normal pS8 fragment was present, but the 1.0kb pfxa3 fragment was missing with no evidence of a higher molecular weight fragment or smear. Deletion of the pfxa3 probe was validated, and data published (Appendix V, Paper 4).

The extent of the deletion (close to 2.5Mb) was determined by presence or absence of markers flanking the FMR1 gene. Proximally DXS369 (RN1) was present but DXS297 (VK23) was deleted, while distally 45B and DXS296 (VK21A, pS8) were present. This deletion encompassing the entire FMR1 gene accounted for the lack of expression of the FRAXA fragile site in this patient. The deletion was submicroscopic and therefore not possible to detect cytogenetically. This mechanism is likely to account for a proportion of cases with Martin-Bell phenotype but no cytogenetic expression (Chapter 1.6.1.). The characteristic fragile X phenotype of this patient (Figure 2; Appendix V, Paper 4) confirmed that complete deletion of the gene had the same phenotypic effect as the more common CCG repeat expansion. Although it was previously thought that expression of adjacent genes could be affected by methylation spreading from the CCG repeat, the extent of the deletion suggested that the FMR1 gene was solely responsible for the phenotype. One might conclude that other genes flanking FMR1 are not vital to the phenotype or that genes in this region are scarce due to non-random distribution.

The mildly retarded mother of IV-4 (III-2), was demonstrated by dosage analysis to carry the deletion (Appendix V, Paper 4). FISH analysis to determine the pattern of inactivation of her X chromosomes was performed by E. Baker, WCH. Usually a deleted X is preferentially inactivated, although cases involving deletion of the region around fragile X have shown the deleted X to be preferentially active (Clarke et al, 1992; Schmidt et al, 1990). It was hypothesised that if the deleted X of III-2 were preferentially active the family history of mental impairment among females might be explained. Random inactivation was evidenced by hybridisation of the FMR1a probe to both early replicating (active) and late

replicating (inactive) X chromosomes in lymphocytes from III-2. From this result it could only be concluded that any putative sequences altering the pattern of inactivation near Xqter are outside the deleted interval. Earlier suggestions of skewed inactivation may therefore reflect a bias of ascertainment.

Another *de novo* deletion spanning 250kb, was described involving only the first exons of the FMR1 gene and sequences 200kb upstream (Wöhrle et al, 1992). Subsequently, a missense point mutation was found within a presumed functional domain of the gene (De Boulle et al, 1993). These findings supported the conclusion that loss of function of FMR1 is sufficient for expression of the fragile X phenotype. Since then several other deletions have been described; a *de novo* deletion spanning 3Mb (Tarleton et al, 1993), another of 200bp (De Vries et al, 1993), a small deletion of part of FMR1 along with 70-100kb upstream sequence (Trottier et al, 1994), a familial 1.6kb deletion proximal to the CGG that includes the promoter region of the gene (Meijer et al, 1994) and another *de novo* deletion of up to 100kb including the first 6 FMR1 exons (Gu et al, 1994). The most recent *de novo* cases include a mosaic male individual carrying a deletion of 660bp in ~40% of his lymphocytes and a normal chromosome in the remainder and a second case involving more than 35kb and the first ten exons of FMR1 (Hirst et al, 1995). Intragenic deletion of a single nucleotide of FMR1 has been shown to cause a loss of function mutation and absence of FMRP in a male with typical features (Lugenbeel et al, 1995). The largest deletion in a male spans 3Mb and includes the FRAXE as well as FRAXA but the phenotype was no more severe than fragile X syndrome (Albright et al, 1994). All deletions involving the FMR1 gene point to the conclusion that the fragile X phenotype is a monogenic disorder caused by the FMR1 gene and that no other gene in the maximum interval of up to 2.5Mb encompassing the gene is implicated in this disorder.

The reported occurrence of fragile X syndrome due to a deletion mutation is still relatively rare compared with those resulting from expansion of the CCG trinucleotide repeat. Most of the deletions described to date have been discovered by Southern analysis during routine molecular investigations of males with idiopathic mental retardation or developmental delay. The specific probe/enzyme systems used for such diagnostic application vary with the laboratory although all are based on detection of expansion around the 5' end of FMR1. Deletion and point mutations involving up to two intragenic nucleotides have been reported (De Boulle et al, 1993; Lugenbeel et al, 1995), though these cannot be detected by routine diagnostic protocols. Deletion detection has therefore been limited by the diagnostic protocols established for the detection of repeat expansion accounting for the majority of cases. Other deletions elsewhere in the gene may account for

a proportion of retarded individuals negative for fragile site expression and CCG expansion, particularly those described clinically as typical fragile X.

7.3. FRAXE mental retardation

Two other folate sensitive fragile sites, FRAXE and FRAXF, distal to FRAXA have been described on the X chromosome (Sutherland and Baker, 1992; Hirst et al, 1993). The fragile site at FRAXE is the second fragile site on the X chromosome to be implicated in the etiology of X-linked mental retardation (Knight et al, 1994; Appendix V, Paper 6). Initially the association of FRAXE with mental impairment was not clear (Sutherland and Baker, 1992; Knight et al, 1994) and was posed to be an effect of biased ascertainment through putative fragile X families (Flynn et al, 1993). The FRAXE associated phenotype is still not well delineated though mild mental retardation can be identified with expansion and subsequent hypermethylation of the CCG trinucleotide repeat similarly to the fragile X (Knight et al, 1994; Appendix V, Paper 6). The phenotype would be best classified as non-specific mental retardation (Chapter 4), although speech defects appear to be more common than in the fragile X syndrome (FRAXA) and the possibility of phenocopies confound establishment of a clear phenotype-genotype relationship (Appendix V, Paper 6).

Direct molecular diagnosis of FRAXE by detection of the CCG amplification relies on the OxE20 probe (Knight et al, 1993). The OxE20 sequence lies 600bp distal to the FRAXE trinucleotide repeat and detects a 5.2kb HindIII fragment in DNA from normal individuals corresponding to 6-25 copies of the repeat. Over 200 copies of the repeat are found in individuals expressing the FRAXE fragile site. The FRAXE locus lies approximately 600kb distal to FMR1 and 800kb proximal to the IDS gene (Appendix V, Paper 7).

Molecular characterisation of two other folate sensitive fragile sites conforming to Mendelian inheritance, FRAXF and FRA16A, show CCG expansion and methylation near CpG islands (Parrish et al, 1994; Nancarrow et al, 1994). The lack of association with an obvious clinical phenotype at both loci suggests that these genes are nonessential or that expression of nearby genes is not affected. Until 1995 no gene had been identified at FRAXE (Appendix V, Papers 7 and 8), although the FRAXE site had been shown to be adjacent to a CpG island methylated in fragile site positive individuals (Knight et al, 1993).

7.3.1. Deletion in Xq28 leads to isolation of the FMR2 gene

Developmental delay patients are frequently referred for direct molecular diagnosis to exclude the fragile X syndrome (FRAXA) and MRX at FRAXE. Two unrelated boys with

developmental disorders were found to be deleted for the anonymous sequence pS8 at DXS296 (Appendix V, Paper 7). The probe pS8 is commonly used as a control in conjunction with the probe pfxa3 in the molecular diagnosis of the fragile X syndrome at FRAXA (section 7.2.). Neither boy had cytogenetic expression of a fragile site nor evidence of abnormal expansion of the CCG repeat at FRAXA or FRAXE. The extent of the deletion in each boy was determined by analysis of presence or absence of the VK21 subclones of DXS296, PCR amplification of VK18 at DXS295 and the ends of the subcloned YAC XY539 (Figure 4; Appendix V, Paper 7). The results of clinical and molecular examination of these boys have been published (Appendix V, Paper 7). The deletion in the Finnish patient, MK with speech delay, was between 800bp to 100kb, and was completely overlapped by the deletion of up to 200kb in the Australian boy CB, with global developmental delay. The deletion in MK removed only the pS8 sequence of all the sequences tested. The maximum extent of the deletion was therefore defined by the closest markers flanking pS8 (VK21A and XY539R). In CB the deletion of these flanking markers defined a larger region deleted to within ~150kb distal to the FRAXE CpG island. The mothers of both boys were demonstrated to carry the deletions, with random inactivation demonstrated by FISH in lymphocytes from the mother of CB.

Proximity of the deletions to the FRAXE site, suggested possible involvement of the gene for mental retardation at FRAXE which had not yet been characterised. The working hypothesis was that the CCG trinucleotide repeat was within the 5' UTR (as was the case for FMR-1), and that the gene for FRAXE would be transcribed towards the telomere and extend as far or even beyond the DXS296 locus. This premise was attractive particularly to account for the phenotype in the Australian case (CB). Conservation through species, of the sequences detected by VK21A and pS8, and detection of a transcript with VK21A, all pointed to the presence of a gene with clinical significance in development. Expression of a 9.5kb or a 2.5kb transcript in various tissues was determined by hybridisation of a Northern blot with VK21A detecting the deletion in CB (Appendix V, Paper 7 - Northern hybridised by Dr J Gecz, WCH).

Sequencing of pS8 was carried out for comparison with gene sequences in public databases such as GenBank (Chapter 1.3.3.). The genomic clone of approximately 800bp had been subcloned into a pUC vector (Yu, 1991), so that using pUC primers it was now possible to amplify the template and sequence by DyeDeoxy™ Terminator PCR (Chapter 2.5.5.1.). The resultant sequence was incomplete and when compared with the GenBank database did not yield high homologies with any known gene sequence, although the sequence had been shown to be conserved across several species by zooblot. Given that the

patient CB had mental retardation, it was speculated that the deleted gene might be expressed in the developing brain. Several of the cDNA clones isolated with the VK21A probe (Chapter 2.6.) were assembled by Dr. Gecz into a contig spanning the deletions. This VK21A associated gene was named FMR2 once CCGn expansion at FRAXE correlated with loss of FMR2 expression, demonstrating that FMR2 is the gene associated with the CpG island at FRAXE (Appendix V, Paper 8).

The pS8 probe lies between exons 3 and 4 of FMR2, while VK21A corresponds to exon 3. Both are downstream to OxE20 which detects CCG expansion at the FRAXE locus. No deletions have been reported to date with the OxE20 probe. These distal probes revealed the first deletions within the gene for FRAXE mental retardation, FMR2. The FMR2 gene represents the first MRX gene to be characterised.

7.4. Conclusion

Deletions of the FMR1 gene are detected as rare fragile X negative patients with the typical clinical characteristics of fragile X syndrome but cytogenetically normal (section 7.2.1.). They are now known to account for a proportion of retarded males with high clinical scores for the fragile X or Martin-Bell phenotype on clinical assessment. Occurrence of the Martin-Bell phenotype in the absence of fragile site expression provides evidence to revive the proposed XLMR category of fra(X) negative Martin-Bell syndrome (Chapter 1.6.1.). Deletions involving FMR1 may range in size from 1bp to 3Mb with the size of the deletion having no apparent bearing on the severity of the phenotype, while a single point mutation can have a severe atypical phenotypic effect (section 7.2.1). Although the mechanisms and pathophysiology are not yet understood, these cases have contributed to understanding that the FMR1 gene expression is critical, such that mutations resulting in a lack of expression have a pathologic effect on the individual.

The deletions involving the pS8 anonymous clone in Xq28 may fuel speculation about the relative differences in phenotype and the involvement of adjacent genes in these two boys (section 7.3.1.). The larger deletion of up to 200kb described in the Australian boy CB, involves the probe VK21A at the DXS296 locus, now known to contain part of coding sequence of the FMR2 gene which accounts for his FRAXE phenotype. This case was instrumental to the identification of the FMR2 gene. The smaller deletion of less than 100kb in the Finnish boy MK does not involve known exons of the FMR2 gene but appears to lie within an intron. Since the deletion in MK does not span DXS296, the differences in phenotype may be explained by the removal of intronic sequence that may have some functional significance, however the sequence represented by the pS8 probe may code for another gene.

The 2.5Mb FMR1 deletion (section 7.2.1.) does not involve DXS296 but may still encompass part of the 5' end of the FMR2 gene at FRAXE. Only one other deletion is known to remove the FRAXE locus (Tarleton et al, 1993) however the deletion in this patient spans 3Mb and also includes the FMR1 gene, therefore the two causes of mental retardation cannot be separated. It is interesting to note, however, that the latter boy had bilateral epicanthic folds and delays in psychomotor and language skill development, as well as hand-flapping and other typical behaviours (Albright et al, 1994). A number of these features are in common with both boys deleted for the pS8 sequence, particularly with CB. Phenotypic variability is common amongst fragile X patients and may also vary at this non-specific disorder. Mutation screening of FMR2, beyond detection of CCG amplification, should be considered in any MRX families mapping across the FMR2 interval.

Typically X inactivation is random on each of the chromosomes of a female somatic cell. Structural abnormalities may skew inactivation such that the chromosome carrying a deletion for example may be more usually inactivated. Deletions in Xq28 have been shown with the reverse however, so that the deleted X was preferentially active, leading to speculations concerning the role of putative genes in the region in determining X inactivation. The opportunity to test this premise was presented in the mother of the boy deleted for FMR1 and the mother of CB deleted for FMR2, since neither boy was a *de novo* mutation. FISH analyses demonstrated random inactivation in both cases, providing no support for the hypothesis of preferential activation due to deletion in Xq28.

Although deletion events causing FRAXA and FRAXE are apparently infrequent, their ascertainment has been through routine molecular investigations limited to the fragment carrying the CCG repeat. The pfxa3 probe will routinely only detect deletions of FMR1 within the 1.0kb *Pst*I target sequence. Probing of *Eco*RI digests for example will increase detection of FMR1 deletions affecting a greater part of the gene. Deletions of FMR2 at FRAXE have not yet been detected with the routine OxE20 probe on *Hind*III digests, however probing with pS8 has detected at least one. The proportion of FRAXA and FRAXE cases arising due to a deletion mutation cannot be quantitated until thorough screening studies examining the whole FMR1 and FMR2 genes have been completed. Given the numerous examples of phenotypic variability arising from allelic heterogeneity, the phenotype need not be expected to be of a 'typical' form.

CHAPTER 8

Conclusion

The objectives of this project (Chapter 1.7) have been realised. In this study, disease genes were predicted to map to the X chromosome by their characteristic segregation pattern and relied on the genetic map of the X, based on polymorphic microsatellite repeats, to provide the basis for mapping and refinement of X-linked genes. Improved molecular techniques were implemented to enable more rapid gene localisation by linkage as well as for construction of genetic maps of the X chromosome. The development of extensive maps of polymorphic markers has enabled the localisation of several X-linked disease genes and has provided the basis for development of specific tests for carrier and prenatal diagnosis of the disease segregating in each family.

Contributions to the map of the X chromosome have included new polymorphic markers, a regional genetic map of Xq25-27, integration of these loci with the physical map of the region, establishment of both framework and comprehensive genetic maps of microsatellite markers spanning the X chromosome and a composite background map established as a basis for gene mapping studies. The genes segregating in fifteen families with X-linked mental retardation have been localised by linkage to specific intervals on the X chromosome. MRX gene localisations in ten families defined five discrete loci for MRX by linkage (Chapter 4). Currently the MRX genes in 31 families have been localised by linkage on an international scale with molecular classification based on regional localisation of the gene borne in each family. At least 7 distinct genes for MRX have now been localised by linkage mapping (Chapter 4), and a further 3 have been physically defined.

Genetic mapping of the disease genes for X-linked disorders segregating in four other families has lead to new localisation of a pigmentary disorder (PDR), confirmation of mapping of BTHS and through collaboration the identification of the gene responsible for the phenotypically variable BTHS and fatal infantile cardiomyopathies (Chapter 6). The gene for BTHS was successfully identified by the positional candidate approach (Chapter 6) and mutations such as altered splice junctions, frameshifts and premature stop codons underlie the genetic pathology. Significant reductions to earlier localisation in several families (XS, PRTS, MRGO, MRX1) promote them for studies of candidate genes in these refined intervals, where before the intervals were rather large for consideration of this approach.

Naturally occurring rearrangements such as deletions associated with a disease entity, inherently pinpoint the location of the defective gene (Chapter 7). It had been anticipated at commencement of this project that a submicroscopic deletion might be found in one of the MRX families studied. Extensive genotyping of markers spanning the chromosome, initially to define the gene interval then within the interval to reduce localisations, did not yield any deletions. Since most localisations span large physical

distances, deletion mutations cannot necessarily be excluded. The integration of genetic and physical maps is important in working towards the positional cloning of mapped genes.

Characterisation of overlapping submicroscopic deletions in two unrelated boys with developmental delay, one with an MRX phenotype, revealed a common region including a fragment at the DXS296 locus conserved in evolution. Detection of a 9.5kb transcript expressed in placenta and brain provided the basis for identification of the FMR2 gene involved in mental retardation at FRAXE in Xq28. This is the first MRX gene to be cloned and may provide insight into a possible family of genes causing MRX or into the functional pathway by which such a gene adversely affects intelligence. Elsewhere, characterisation of the genes involved in contiguous gene syndromes at three other locations on the chromosome (two of which do not overlap with MRX intervals defined by linkage) will provide candidate genes with an increased likelihood of involvement in the familial forms of XLMR.

Enumeration of loci for non-specific XLMR mapped by genetic and physical means invite speculation on the density of genes for intelligence on the X chromosome and discussion of the genetic heterogeneity of this clinically apparently homogenous disorder. There is no evidence to suggest that there are more genes on the X chromosome than on autosomes relative to size (Chapter 1.5.2.), although it would appear from the excess of retarded males in the population that genes on the X chromosome do contribute to a greater proportion of the mentally retarded population than would be expected relative to chromosome size (Chapter 1.6). Theoretical estimations of the likely number of genes for XLMR required to account for the excess of male retardates have been proved to be surprisingly accurate to date. Given the estimated high frequency of XLMR, identification of the underlying gene mutations are of considerable medical importance. These genes may be fundamental in development of the normal brain or in development of cognitive processes. This study has contributed to the delineation of 11 discrete genes primarily responsible for X-linked mental retardation. Of these, expansion of a CCG trinucleotide within the FMR1 gene at FRAXA causes the most common form of XLMR accounting for nearly 25% of all retardation. Deletions of FMR1 also contribute a very small number of cases of fragile X syndrome (Chapter 7).

Refinement of earlier MRX localisations and newer additional localisations may increase the minimum number of MRX loci, however the next phase towards elucidation of these genes needs implementation of a new approach. Under the premise that the X chromosome is equivalent to 5% of the total genome, up to 1500 of the 30000 genes expressed in brain would map to the X chromosome and there would be at least 6/cM or 9/Mb given random distribution. The deletion of 2Mb around the FMR-1 locus (Chapter 7)

caused only fragile X syndrome without signs of involvement from other genes, while MRX and MRXS genes demonstrate a greater density around the centromere (Chapters 4 and 5). The obvious conclusion is that genes expressed in brain are not randomly distributed therefore mutation screening in positional candidates should not be excluded from studies to identify the disease causing mutation. A simplistic cause and effect cannot be predicted for any genes affecting intelligence since the processes involved in normal development and maintenance of brain function are too many and ill understood.

Tens of thousands of genes are known to be expressed in the adult brain, and an inestimable number are likely to be involved in the development and differentiation of the brain and its neurons. Kidd (1993) states that the choice of a candidate gene for neuropsychiatric disorders is based less on knowledge and more on ignorance of the complex structure and function relationships of the central nervous system. He adds that for every candidate proposed hundreds of equally likely unidentified genes exist. Crowe (1993) suggests that most positive findings of mutations in psychiatric candidate genes will be false positives, and that adequate procedures for minimising and disproving these results will have to be established since the genotype can only be confirmed as the pathogenic cause of the phenotype by reproducible population associations. For MRX genes the positional candidate approach has merit in that a vast proportion of potential candidates are already excluded by their physical location. Advances in the construction of physical and transcriptional maps of the X chromosome will also contribute to selection criteria for candidacy. The characterisation and identification of implicated genes may be approached by a number of techniques proven to be successful but will be most efficiently directed by an international collaboration with this common goal. Constructive progress now requires greater scientific collaboration to share established resources and expertise in place elsewhere. The findings of collaborating groups will be faster and more valuable than the sum of its parts. The establishment and nurturing of fruitful collaborations is in itself an important step towards achieving common goals.

This final decade of the 20th century is a most exciting time to be a scientist, with the unravelling of the mysteries of the genome and new mysteries being created. The breathtaking rapidity with which genes are being cloned and analysed has virtually overtaken the ethical issues. Mapping of the genes causing X-linked disorders is an important contribution to the map of the complete human genome, to the unravelling of complex mechanisms eg. a reason for some genes to escape X inactivation and aids in determining the role of the X chromosome in defining human diversity. It is predicted that most of the genes localised in this thesis will be isolated by the commencement of the next millennium. Given the ephemeral quality of families and the modern tendency towards

fewer children it is recommended that in the interim, where possible, DNA samples or preferably cell lines be archived from older members of a family. Once the mutations causing defective gene function have been elucidated, however, the need for family material will be reduced as direct mutation detection will prevail over linkage analysis.

References

AAMR (1992) Introduction to mental retardation: definition by the American Association on Mental Retardation. Website; http://thearc.org/faqs/mrqa.html.

Abbs S, Roberts RG, Mathew CG, Bentley DR, 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, Merril CR, Wu A, Olde B, Moreno RF, Kerlavage AR, McCombie WR, Venter JC (1991) Complementary DNA sequencing: sequence tags and human genome project. Science 252:1651-1656.

Adams MD, Kerlavage AR, Fields C, Venter JC (1993a) 3,400 new expressed sequence tags identify diversity of transcripts in human brain. Nature Genet 4:256-267.

Adams MD, Bento Soares M, Kerlavage AR, Fields C, Venter JC (1993b) Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library. Nature Genet 4:373-380.

Ades LC, Rogers M, Sillence DO (1993a) An X-linked reticulate pigmentary disorder with systemic manifestations: report of a second family. Pediatr Dermatol 10/4:344-351.

Ades LC, Gedeon AK, Wilson MJ, Latham M, Partington MW, Mulley JC, Nelson J, Lui K, Sillence DO, (1993b) Barth syndrome: Clinical features and confirmation of gene localisation to distal Xq28. Am J Med Genet 45:327-334.

Albright SG, Lachiewicz AM, Tarleton JC, Rao KW, Schwartz CE, Richie R, Tennison MB, Aylsworth AS, (1994) Fragile X phenotype in a patient with a large de novo deletion in Xq27-q28. Am J Med Genet 51:294-297 (XLMR6).

Aldred MA, Dry KL, Knight-Jones EB, Hardwick LJ, Teague PW, Lester DH, Brown J, Spowart G, Carothers AD, Raeburn JA, Bird AC, Fielder AR, Wright AF (1994) Genetic analysis of a kindred with X-linked mental handicap and retinitis pigmentosa. Am J Hum Genet 55:916-922.

Allen RC and Belmont JW, (1992) Dinucleotide repeat polymorphism at the DXS178 locus. Hum Mol Genet 1/3:216.

Allen RC and Belmont JW, (1993) Trinucleotide repeat polymorphism at DXS101. Hum Mol Genet 2/9:1508.

Antequera F, Bird A (1993) Number of CpG islands and genes in human and mouse. Proc Natl Acad Sci USA 90:11995-11999.

Ardinger HH, Hanson JW, Zellweger HU (1984) Börjeson-Forssman-Lehmann syndrome: Further delineation in five cases. Am J Med Genet 19:653-664.

Arveiler B, Alembik Y, Hanauer A, Jacobs P, Tranebjaerg L, Mikkelsen M, Puissant H, Larget Piet L, Mandel JL (1988) Linkage analysis suggests at least two loci for X-linked non-specific mental retardation. Am J Med Genet 30:473-483.

Atkin JF, Flaitz K, Patil S, Smith W (1985) A new X-linked mental retardation syndrome. Am J Med Genet 21:697-705.

Bach I, Robinson D, Thomas HH, 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, 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, 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, Andria G (1992) Deletions and translocations involving the distal short arm of the human X chromosome: review and hypotheses. Hum Mol Genet 1/4:221-227.

Banchs I, Bosch A, Guimerà J, Lázaro C, Puig A, Estivill X (1994) New alleles at microsatellite loci in CEPH families mainly arise from somatic mutations in the lymphoblastoid cell lines. Hum Mutation 3:365-372.

Barker DF, Cleverly J, Fain PR, (1992) Two CA-dinucleotide polymorphisms at the COL4A5 (Alport syndrome) gene in Xq22. Nucl Acids Res 20/4:929.

Barker DF and Fain PR, (1993) Definition and Mapping of STSs at STR and RFLP Loci in Xp11-Xq22. Genomics 18:712-716.

Barth PG, Scholte HR, Berden JA, Moorsel VDK, Luyt-Houwen IEM, Van'T Veer Korthof ET, Van der Harten JJ, Sobotka-Plojhar MA (1983) An X-linked mitochondrial disease affecting cardiac muscle, skeletal muscle and neutrophil leukocytes. J Neurol Sci 62:327-355.

Bassi MT, Bergen AAB, Wapenaar MC, Schiaffino M, Yates JRW, Charles SJ, Meitinger T, Ballabio A (1994) A submicroscopic deletion in a patient with isolated X-linked ocular albinism (OA1). Hum Mol Genet 3/4:647-648.

Beckmann JS, Weber JL (1992) Survey of human and rat microsatellites. Genomics 12:627-631.

Bell MV, Hirst MC, Nakahori Y, MacKinnon RN, Roche A, Flint TJ, Jacobs PA, Tommerup N, Tranebjaerg L, Froster-Iskenius U, Kerr B, Turner G, Lindenbaum RH, Winter R, Pembrey M, Thibodeau S, Davies KE (1991) Physical mapping across the fragile X: hypermethylation and clinical expression of the fragile X syndrome. Cell 64:861-866.

Berg MA, Mathews K, Kobori J, Wiles K, Schelley S, Johnston K (1992) New Borjeson-Forssman-Lehmann family: Mildly affected males, clinically identifiable female carriers and linkage to Xq26-27. Abstract A4.

Biancalana V, Serville F, Pommier J, Julien J, Hanauer A, Mandel JL (1992) Moderate instability of the trinucleotide repeat in spino bulbar muscular atrophy. Hum Mol Genet 1:255-258.

Biancalana V, Trivier E, Weber C, Weissenbach J, Rowe PSN, O'Riordan JLH, Partington MW, Heyberger S, Oudet C, Hanauer A (1994) Construction of a high-resolution linkage map for Xp22.1-p22.2 and refinement of the genetic localisation of the Coffin-Lowry syndrome gene. Genomics 22:617-625.

Bies RD, Phelps SF, Cortez MD, Roberts R, Caskey CT, Chamberlain J (1992) Human and murine dystrophin mRNA transcripts are differentially expressed during skeletal muscle, heart, and brain development. Nucl Acids Res 20/7:1725-1731.

Bione S, Maestrini E, Rivella S, Mancini M, Regis S, Romeo G, Toniolo D (1994) Identification of a novel X-linked gene responsible for Emery-Dreifuss muscular dystrophy. Nature Genet 8/4:323-327.

Bird AP (1987) CpG islands as gene markers in the vertebrate nucleus. Trends Genet 3:342-347.

Black GCM, Chen ZY, Craig IW, Powell JF, (1991) Dinucleotide repeat polymorphism at the MAOA locus. Nucl Acids Res 19/3:689.

Black GM, Chen ZY, Coleman M, Davies KE, Craig IW (1994) A bidirectional walk from L1.28 (DXS7) embracing 19 YAC clones. Cytogenet Cell Genet 67:334 Abs 19.

Boehnke M (1994) Limits of resolution of genetic linkage studies: Implications for the positional cloning of human disease genes. Am J Hum Genet 55:379-390.

Bolhuis PA, Hensels GW, Hulsebos TJM, Baas F, Barth PG, (1991) Mapping of the locus for X-linked cardioskeletal myopathy with neutropenia and abnormal mitochondria (Barth syndrome) to Xq28. Am J Hum Genet 48:481-485.

Börjeson M, Forssman H, Lehmann O (1962) An X-linked recessively inherited syndrome characterised by grave mental deficiency, epilepsy and endocrine disorder. Acta Med Scand 171:12-21.

Botstein D, White RL, Skolnick M, Davis RW, (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314-331.

Bouloux PMG, Hardelin JP, Munroe P, Kirk JMW, Legouis R, Levilliers J, Hazan J, Weissenbach J, Petit C, (1991) A dinucleotide repeat polymorphism at the Kallmann locus (Xp22.3). Nucl Acids Res 19/19:5453.

Bowcock A, Cavalli-Sforza LC (1991) The study of variation in the human genome. Genomics 11:491-498.

Boyce FM, Beggs AH, Feener C, Kunkel LM (1991) Dystrophin is transcribed in brain from a distant upstream promoter. PNAS *USA* 88:1276-1280.

Brook JD, McCurrach ME, Harley HG, Buckler AJ, Church D, Aburatani H, Hunter K, Stanton VP, Thirion JP, Hudson T, Sohn R, Zemelman B, Snell RG, Rundle SA, Crow S, Davies J, Shelbourne P, Buxton J, Jones C, Juvonen V, Johnson K, Harper PS, Shaw DJ, Housman DE (1992) Molecular basis of myotonic dystrophy: Expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. Cell 69:784-789.

Brown CJ, Willard HF (1990) Localization of a gene that escapes inactivation to the X chromosome proximal short arm: implications for X inactivation. Am J Hum Genet 46:273-279.

Brown, SDM (1992) The Mouse Genome Project and human genetics. Genomics 13:490-492.

Browne DL, Luty JA, Litt M, (1991a) Dinucleotide repeat polymorphism at the DXS538 locus. Nucl Acids Res 19/5:1161.

Browne DL, Zonana J, Litt M, (1991b) Dinucleotide repeat polymorphism at the DXYS1X locus. Nucl Acids Res 19/7:1721.

Browne DL, Zonana J, Litt M, (1991c) Dinucleotide repeat polymorphism at the PGK1 locus. Nucl Acids Res 19/7:1721.

Browne DL, Zonana J, Litt M, (1992a) Dinucleotide repeat polymorphism at the PGK1P1 locus. Nucl Acids Res 20/5:1169.

Browne D, Barker D, Litt M, (1992b) Dinucleotide repeat polymorphisms at the DXS365, DXS443 and DXS451 loci. Hum Mol Genet 1/3:213.

Brunner HG, Nelen MR, van Zandvoort P, Abeling NGGM, van Gennip AH, Wolters EC, Kuiper MA, Ropers HH, van Oost BA (1993a) X-linked borderline mental retardation with prominent behavioural disturbance: phenotype, genetic localization, and evidence for disturbed monoamine metabolism. Am J Hum Genet 52/6:1032-1039.

Brunner HG, Nelen MR, Breakefield XO, Ropers HH, van Oost BA (1993b) Abnormal behaviour associated with a point mutation in the structural gene for monoamine oxidase A. Science 262:578-580.

Buetow KH, Weber JL, Ludwigsen S, Scherpbier-Heddema T, Duyk GM, Sheffield VC, Wang Z, Murray JC (1994) Integrated human genome-wide maps constructed using the CEPH reference panel. Nature Genet 6:391-393.

Burks C, Cassidy M, Cinkovsky MJ, Cumella KE, Gilna P, Hayden JED, Keen GM, Kelley TA, Kelly M, Kristofferson D, Ryals J (1991) GenBank. Nucl Acids Res 19 Suppl:2221-2225.

Butler MG, Allen GA, Haynes JL, Singh DN, Watson MS, Breg WR (1991) Anthropometric comparison of mentally retarded males with and without the fragile X syndrome. Am J Med Genet 38:260-268.

Cann HM, (1992) CEPH maps. Current Opinion of Genet and Dev 2:393-399.

Carlson EA (1991) Defining the gene: an evolving concept. Am J Hum Genet 49:475-487.

Carrozzo R, Ellison J, Yen P, Taillon-Miller P, Brownstein BH, Persico G, Ballabio A, Shapiro L (1992) Isolation and characterisation of a yeast artificial chromosome (YAC) contig around the human steroid sulfatase gene. Genomics 12:7-12.

Cavalli-Sforza LL, Wilson AC, Cantor CR, Cook-Deegan RM, King MC (1991) Call for a worldwide survey of human genetic diversity: A vanishing opportunity for the human genome project. Genomics 11:490-491.

Chang YPC, Smith KD, Dover GJ (1994) Dinucleotide repeat polymorphisms at the DXS85, DXS16 and DXS43 loci. Hum Mol Genet 3/6:1029.

Charlton R, Slaney S, Roberts A, Cheetham CM, Huson SM, Seller A (1994) Linkage of a non-specific X-linked mental retardation gene to Xq26.1. J Med Genet 31:172.

Chen Z-Y, Battinelli EM, Hendriks RW, Powell JF, Middleton-Price H, Sims KB, Breakefield XO, Craig IW (1993) Norrie disease gene: characterisation of deletions and possible function. Genomics 16:533.

Christodoulou J, McInnes RR, Jay V, Wilson G, Becker LE, Lehotay DC, Platt BA, Bridge PJ, Robinson BH, Clarke JTR, (1994) Barth syndrome: Clinical observations and genetic linkage studies. Am J Med Genet 50:255-264.

Chudley AE, Lowry RB, Hoar DI (1988) Mental retardation, distinct facial changes, short stature, obesity, and hypogonadism: a new X-linked retardation syndrome. Am J Med Genet 31:741-751.

Chudley AE (1991) Gene localization in Börjeson-Forssman-Lehmann Syndrome and other X-linked mental retardation syndromes. Am J Med Genet 38:248.

Clark PA, Lester T, Villard L, Fontes M, Kinnon C (1994) Deletion mapping of the DXS986, DXS995 and DXS1002 loci defines their order within Xq21. J Med Genet 31:344-345.

Clarke JTR, Wilson PJ, Morris CP, Hopwood JJ, Richards RI, Sutherland GR, Ray PN (1992) Characterization of a deletion at Xq27-q28 associated with unbalanced inactivation of the nonmutant X chromosome. Am J Hum Genet 51:316-322.

Clemens PR, Fenwick RG, Camberlain JS, Gibbs RA, de Andrade M, Chakraborty R, Caskey CT (1991) Carrier detection and prenatal diagnosis in Duchenne and Becker muscular dystrophy families, using dinucleotide repeat polymorphisms. Am J Hum Genet 49:951-960.

Clépet C, Schafer AJ, Sinclair AH, Palmer MS, Lovell-Badge R, Goodfellow PN (1993) The human SRY transcript. Hum Mol Genet 2/12:2007-2012.

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

Coleman MP, Bhattacharya S, Lindsay S, Wright A, Jay M, Litt M, Craig I, Davies K, (1990) Localization of the microsatellite probe DXS426 between DXS7 and DXS255 on Xp and linkage to X-linked retinitis pigmentosa. Am J Hum Genet 47:935-940.

Coleman MP, Murray JC, Willard HF, Nolan KF, Reid KBM, Blake DJ, Lindsay S, Bhattacharya SS, Wright A, Davies KE, (1991) Genetic and physical mapping around the Properdin P gene. Genomics 11:991-996.

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

Collins F, Galas D (1993) A new five-year plan for the US Human Genome Project. Science 262:43-46.

Conneally PM, Edwards JH, Kidd KK, Lalouel JM, Morton NE, Ott J, White R (1985) Report of the committee on methods of linkage analysis and reporting: Human Gene Mapping 8. Cytogenet Cell Genet 40:356-359.

Cox TC, Kozman HM, Raskind WH, May BK, 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, Human Molecular Genetics 1/8:639-641.

Craig JM, Bickmore WA (1994) The distribution of CpG islands in mammalian chromosomes. Nature Genetics 7:376-381.

Cremers FPM, van de Pol DJR, Diergaarde PJ, Wieringa B, Nussbaum RL, Schwartz M, Ropers HH (1989) Physical fine mapping of the choroideremia locus using Xq21 deletions associated with complex syndromes. Genomics 4:41-46.

Crowe RR (1993) Candidate genes in psychiatry; An epidemiological perspective. Am J Med Genet (Neuropsychiatric Genetics) 48:74-77.

Curtis ARJ, Roustan P, Kamakari S, Thiselton D, Lindsay S, Bhattacharya SS, (1992) Two dinucleotide repeat polymorphisms at the DXS571 locus. Hum Mol Genet 1/9:776.

Dausset J, Cann H, Cohen D, Lathrop M, Lalouel J-M, White R, (1990) Centre d'Etude du Polymorphisme Humain (CEPH): collaborative genetic mapping of the human genome. Genomics 6:575-577.

Davies KE, Craig IW (1991) Report of the second X chromosome workshop. Genomics 10:842-846.

Davies K, Mandel JL, Monaco AP, Nussbaum RL, Willard HF (1991) Report of the committee on the genetic constitution of the X chromosome (HGM11). Cytogenet Cell Genet 58:853-966.

De Boulle K, Verkerk AJMH, Reyniers E, Vits L, Hendrickx J, Van Roy B, Van Den Bos F, de Graaff E, Oostra BA, Willems PJ (1993) A point mutation in the FMR-1 gene associated with fragile X mental retardation. Nature Genetics 3:31-35.

de la Chapelle A (1988) Invited editorial: The complicated issue of human sex determination. Am J Hum Genet 43:1-3.

Devor EJ, Cloninger CR, Hoffman PL, Tabakoff B, (1993) Association of monoamine oxidase (MAO) activity with alcoholism and alcoholis subtypes. Am J Med Genet (Neuropsychiatric Genetics) 48:209-213.

Drayna D, Davies K, Hartley D, Mandel J-L, Camerino G, Williamson R, White R, (1984) Genetic mapping of the human X chromosome by using restriction fragment length polymorphisms. Proc Natl Acad Sci 81:2836-2839.

Drayna D, White R (1985) The genetic linkage map of the human X chromosome. Science 230:753-758.

Duyk GM, Kim SW, Myers RM, Cox DR (1990) Exon trapping: a genetic screen to identify candidate transcribed sequences in cloned mammalian genomic DNA. Proc Natl Acad Sci USA 87:8995-8999.

Econs MJ, Francis F, Rowe PSN, Speer MC, Riordan JLHO, Lehrach H, Becker PA (1994) Dinucleotide repeat polymorphism at the DXS1683 locus. Hum Mol Genet 3:680.

Edwards A, Civitello A, Hammond HA, Caskey CT (1991) DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Am J Hum Genet 49:746-756.

Edwards JH (1987) Exclusion mapping. J Med Genet 24:539-543.

Ellison KA, Fill CP, Terwilliger J, DeGennaro LJ, Martin-Gallardo A, Anvret M, Percy AK, Ott J, Zoghbi H (1992) Examination of X chromosome markers in Rett syndrome: Exclusion mapping with a novel variation on multilocus linkage analysis. Am J Hum Genet 50:278-287.

Emery AEH (1987) Am J Med Genet 26:235-236.

Fain PR, Luty JA, Guo Z, Nguyen K, Barker DF, Litt M, (1991) Localization of the highly polymorphic microsatellite DXS456 on the genetic linkage map of the human X chromosome. Genomics 11:1155-1157.

Fain PR, Kort EN, Chance PF, Nguyen K, Redd DF, Econs MJ, Barker DF (1995) A 2D crossover-based map of the human X chromosome as a model for map integration. Nature Genet 9:261-266.

Fairweather N, Chelly J, Monaco AP, (1993) Dinucleotide repeat polymorphisms from DXS106 and DXS227 YACs using a two stage approach. Hum Mol Genet 2/5:607-608.

Falk CT (1991) A simple method for ordering loci using data from radiation hybrids. Genomics 9:120-123.

Feener CS, Koenig M, Kunkel LM (1989) Alternative splicing of human dystrophin mRNA generates isoforms at the carboxy terminus. Nature (London) 338(6215):509-511.

Feener CA, Boyce FM, 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.

Feinberg AP, Vogelstein B (1983) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Analyt Biochem 132:6-13.

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, Ballabio A (1995) An integrated physical and genetic map of a 35Mb region on chromosome Xp22.3-Xp21.3. Hum Mol Genet 4/10:1821-1827.

Fields C, Adams MD, White O, Venter JC (1994) How many genes in the human genome? Nature Genetics 7:345-346.

Flynn GA, Hirst MC, Knight SJL, Macpherson JN, Barber JCK, Flannery AV, Davies KE et al (1993) Identification of the FRAXE fragile site in two families ascertained for X linked mental retardation. J Med Genet 30:97-100.

Fox P, Fox D, Gerrard JW (1980) X-linked mental retardation: Renpenning revisited. Am J Med Genet 7:491-495.

Francis F, Rowe PSN, Econs MJ, See CG, Benham F, O'Riordan JLH, Drezner MK, Hamvas RMJ, Lehrach H (1994) A YAC contig spanning the hypophosphatemic rickets disease gene (HYP) candidate region. Genomics 21:229-237.

Fransen E, Schrander-Stumpel C, Vits L, Coucke P, van Camp G, Willems PJ (1994) X-linked hydrocephalus and MASA syndrome present in one family are due to a single missense mutation in exon 28 of the L1CAM gene. Hum Mol Genet 3/12:2255-2256.

Fransen E, Lemmon V, van Camp G, Vits L, Coucke P, Willems PJ (1995) CRASH syndrome: clinical spectrum of corpus callosum hypoplasia, retardation, adducted thumbs, spastic paraperesis and hydrocephalus due to mutations in one single gene, L1. Eur J Hum Genet 3:273-284.

Freije D, Helms C, Watson MS, Donis-Keller H (1992) Identification of a second pseudoautosomal region near the Xq and Yq telomeres. Science 258:1784-1787.

Freimer NB, Sankuijl LA, Blower SM (1993) Incorrect specification of marker allele frequencies: effects on linkage analysis. Am J Hum Genet 52:1102-1110.

Frézal J, Schinzel A (1991) Report of the committee on clinical disorders and chromosomal deletion syndromes (HGM11). Cytogenet Cell Genet 58:986-1052.

Fryns JP, Spaepen A, Cassiman JJ, van den Berghe H (1991) X linked complicated spastic paraplegia, MASA syndrome, and X linked hydrocephalus owing to congenital stenosis of the aqueduct of Sylvius: variable expression of the same mutation at Xq28. J Med Genet 28:429-431.

Fu YH, Kuhl DPA, Pizzuti A, Pieretti M, Sutcliffe J, Richards S, Verkerk AJMH, Holden JJA, Fenwick RG, Warren ST, Oostra BA, Nelson DL, Caskey CT (1991) Variation of the CGG repeat at the fragile X site results in instability: resolution of the Sherman paradox. Cell 67:1047-1058.

Fu YH, Pizzuti A, Fenwick RG, King J, Rajnarayan S, Dunne PW, Dubel J, Nasser GA, Ashizawa T, de Jong P, Wieringa B, Korneluk R, Perryman MB, Epstein HF, Caskey CT (1992) An unstable triplet repeat in a gene related to myotonic muscular dystrophy. Science 255:1256-1258.

Fuscoe JC, Nelsen AJ (1994) Molecular description of a hypoxanthine phosphoribosyltransferase gene deletion in Lesch-Nyhan syndrome. Hum Mol Genet 3/1:199-200.

Gariboldi M, Maestrini E, Canzian F, Manenti G, De Gregorio L, Rivella S, Chatterjee A, Herman GE,Archidiacono N, Antonacci R (1994)Comparative mapping of the actin-binding protein 280 genes in human and mouse. Genomics 21/2:428-430.

Gendrot C, Ronce N, Toutain A, Moizard M-P, Müh J-P, Raynaud M, Dourlens J, Briault S, Moraine C (1994) X-linked mental retardation exhibiting linkage to DXS255 and PGK1P1: A new MRX family (MRX14) with localization in the pericentromeric region. Clin Genet 45:145-153.

Généthon Microsatellite Map Catalogue, (1992) Human Genome Research Centre.

Gervais R, Lafitte JJ, Dumur V, Kesteloof M, Lalau G, Houdret N, Roussel P (1993) Sweat chloride and delta F508 mutation in cronic bronchiectasis (letter). Lancet 342(8877):997.

Gibbons RJ, Picketts DJ, Villard L, Higgs DR (1995) Mutations in a putative global transcriptional regulator cause X-linked mental retardation with α-thalassemia (ATR-X syndrome). Cell 80:837-845.

Glass I (1991) X-linked mental retardation. J Med Genet 28:361-371.

Glass IA, White EM, Pope MJ, Pirrit LA, Cockburn F, Connor JM (1991b) Linkage analysis in a large family with nospecific X-linked mental retardation. Am J Med Genet 38:240-243.

Gong W, Hu L, Kioschis P, Poutska A, Dahl N (1994) Dinucleotide repeat polymorphism at the DXS1684 locus. Hum Mol Genet 3/8:1442.

Goodfellow P, Pym B, Mohandas T, Shapiro LJ (1984) The cell surface antigen locus, MIC2X, escapes X-inactivation. Am J Hum Genet 36:777-782.

Gordon AG (1993) Aphasia, deafness, or mental retardation. J Med Genet 30:262-264.

Górecki DC, Monaco AP, Derry JMJ, Walker AP, Barnard EA, Barnard PJ (1992) Expression of four alternative dystrophin transcripts in brain regions regulated by different promoters. Hum Mol Genet 1/7:505-510.

Górecki DC, Derry JMJ, Barnard EA (1994) Dystroglycan: brain localisation and chromosome mapping in the mouse. Hum Mol Genet 3/9:1589-1597.

Greenberg DA (1989) Inferring mode of inheritance by comparison of lod scores. Am J Med Genet 34:480-486.

Grimsby J, Chen K, Devor EJ, Cloninger CR, Shih JC, (1992) Dinucleotide repeat $(TG)_{23}$ polymorphism in the MAOB gene. Nucl Acids Res 20/4:924.

Grompe M (1993) The rapid detection of unknown mutations in nucleic acids. Nature Genetics 5:111-117.

Gu Y, Lugenbeel KA, Vockley JG, Grody WW, Nelson DL (1994) A de novo deletion in FMR1 in a patient with developmental delay. Human Mol Genet 3/9:1705-1706.

Gusella JF, Wexler NS, Conneally PN, Naylar SL, Anderson MA, Tanzi RE, Watkins PC, Ottina K, Wallace MR, Sakaguchi AY, Young AB, Shoulson I, Bonilla E, Martin JB, (1983) A polymorphic DNA marker genetically linked to Huntington's disease. Nature 306:234-238.

Gusella JF (1986) DNA polymorphism and human disease. Ann Rev Biochem 55:831-854.

Gustavson K-H, Annerén G, Malmgren H, Dahl N, Ljunggren C-G, Bäckman 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.

Gyapay G, Morisette J, Vignal A, Dib C, Fizames C, Millasseau P, Marc S, Bernardi G, Lathrop M, Weissenbach J (1994) The 1993-94 Généthon human genetic linkage map. Nature Genet 7:246-339.

Hamada H, Kakunaga T (1982) Potential Z-DNA forming sequences are highly dispersed in the human genome. Nature 298:396-398.

Hamel BCJ, Smits APT, de Graaf E et al (1994) Segregation of FRAXE in a large family: clinical, psychometric, cytogenetic and molecular data. Am J Hum Genet 55:923-931.

Hardwick LJ, Brown J, Wright AF, (1993) An STR polymorphism at the CYBB locus. Hum Mol Genet 2/10:1755.

Hauge XY, Litt M (1993) A study of the origin of 'shadow bands' seen when typing dinucleotide repeat polymorphisms by the PCR. Hum Mol Genet 2/4:411-415.

Hawkins JR (1994) Sex determination. Hum Mol Genet 3:1463-1467.

Heard E, Avner P (1994) Role play in X-inactivation. Hum Mol Genet 3:1481-1485.

Herbst DS (1980) Nonspecific X-linked mental retardation I: a review with information from 24 new families. Am J Med Genet 7:443-460.

Herbst DS, Miller JR (1980) Non-specific X-linked mental retardation. II. The frequency in British Columbia. Am J Med Genet 7:461-469.

Herbst DS, Dunn HG, Dill FJ, Kalousek DK, Krywaniuk LW (1981) Further delineation of X-linked mental retardation. Hum Genet 58:366-372.

Herrell S, Novo FJ, Charlton R, Affara NA (1995) Development and physical analysis of YAC contigs covering 7 Mb of Xp22.3-p22.2. Genomics 25:526-537.

Hicks AA, Johnson KJ, Barnard EA, Darlison MG, (1991) Dinucleotide repeat polymorphism in the human X-linked $GABA_A$ receptor α3-subunit gene. Nucl Acids Res 19/14:4016.

Hinds HL, Hendriks RW, Craig IW, Chen ZY, (1992) Characterization of a highly polymorphic region near the first exon of the human MAOA gene containing a GT dinucleotide and a novel VNTR motif. Genomics 13:896-897.

Hirst M, Nakahori Y, Knight SJL, Schwartz C, Thibodeau SN, Roche A, Flint TJ, Connor JM, Fryns JP, Davies KE (1991) Genotype prediction in the fragile X syndrome. J Med Genet 28:824-829.

Hirst MC, Barnicoat A, Flynn G, Wang Q, Daker M, Buckle VJ, Davies KE, Bobrow M (1993) The identification of a third fragile site, FRAXF, in Xq27-28 distal to both FRAXA and FRAXE. Hum Mol Genet 2:197-200.

Hirst M, Grewal P, Flannery A, Slatter R, Maher E, Barton D, Fryns JP, Davies K (1995) Two new cases of FMR1 deletion associated with mental impairment. Am J Hum Genet 56:67-74.

Hochgeschwender U (1992) Toward a transcriptional map of the human genome. Trends Genet 8/2:41-44.

Hu LJ, Blumenfeld-Heyberger S, Hanauer A, Weissenbach J, Mandel JL (1994) Non-specific X-linked mental retardation: Linkage analysis in MRX2 and MRX4 families revisited. Am J Med Genet 51:569-574.

Huang THM, Hejtmacik JF, Edwards A, Pattigrew AL, Herrera CA, Hammond HA, Caskey CT, Zoghbi HY, Ledbetter DH (1991) Linkage of the gene for an X-linked mental

retardation disorder to a hypervariable (AGAT)n repeat motif within the human HPRT locus (Xq26). Am J Hum Genet 49:1312-1319.

Huang THM, Cottingham RW, Ledbetter DH, Zoghbi HY (1992) Genetic mapping of four dinucleotide repeat loci, DXS453, DXS458, DXS454, and DXS424, on the X chromosome using multiplex polymerase chain reaction. Genomics 13:375-380.

Huber I, Bitner-Glindzicz M, de Kok YJM, van der Maarel SM, Ishikawa-Brush Y, Monaco AP, Robinson D, Malcolm S, Pembrey ME, Brunner HG, Cremers FPM, Ropers HH (1994) X-linked mixed deafness (DFN3): cloning and characterization of the critical region allows the identification of novel microdeletions. Hum Mol Genet 3/7:1151-1154.

Hughes AE (1993) Optimization of microsatellite analysis for genetic mapping. Genomics:15 433-434.

Hugnot JP, Gilgenkrantz H, Jeanpierre M, Chelly J, Kaplan JC, Kahn A (1993) Striking conservation of the brain-specific region of the dystrophin gene. Mammalian Genome 4:393-396.

Huntington Disease Collaborative research group (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable in Huntington's disease chromosomes. Cell 72:971-983.

Jacobs P (1979) More on Marker X chromosomes, mental retardation and macro-orchidism. N Engl J Med 300:739.

Jedele KB, Michels VV, Schaid DJ, Schowalter KV, Thibodeau SN (1992) Linkage of non-specific mental retardation to Xq21.31. Am J Med Genet 43:436-442.

Jeffreys AJ, Wilson V, Thein SL (1985) Hypervariable 'minisatellite regions in human DNA. Nature 314:67-73.

Jennings C (1995) How trinucleotide repeats may function. Nature 378:127.

Jouet M, Rosenthal A, Armstrong G, MacFarlane J, Stevenson R, Paterson J, Metzenberg A, Ionasescu V, Temple K, 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.

Kawaguchi Y, Okamoto T, Taniwaki M, Aizawa M, Inoue M, Katayama S, Kawakami H, Nakamura S, Nishimura M, Akiguchi I, Kimura J, Narumiya S, Kikizuka A (1994) CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. Nature Genetics 8:221-228.

Keats BJB, Ott J, Conneally M (1989) Report of the committee on linkage and gene order. Cytogenet Cell Genet 51:459-502.

Keats BJB, Sherman SL, Morton NE, Robson EB, Buetow KH, Cartwright PE, Chakravarti A, Francke U, Green PP, Ott J (1991) Guidelines for human linkage maps: an international system for human linkage maps (ISLM, 1990). Genomics 9:557-560.

Kelly DP and Strauss AW, (1994) Inherited Cardiomyopathies - Mechanisms of Disease Review Article. New England J Med 330/13:913-919.

Ketterling RP, Ricke DO, Wurster MW, Sommer SS (1993) Deletions with inversions: report of a mutation and review of the literature. Hum Mut 2:53-57.

Ketterling RP, Vielhaber EL, Lind TJ, Thorland EC, Sommer SS (1994) The rates and patterns of deletions in the human Factor IX gene. Am J Hum Genet 54:201-213.

Khan AS, Wilcox AS, Polymeropoulos MH, Hopkins JA, Stevens TJ, Robinson M, Orpana AK, Sikela JM (1992) Single pass sequencing and physical and genetic mapping of human brain cDNAs. Nature Genetics 2:180-185.

Kidd KK (1993) Associations of disease with genetic markers: Déjà vu all over again. Am J Med Genet (Neuropsychiatric Genetics) 48:71-3.

Kirchgessner CU, Trofatter JA, Mahtani MM, Willard HF, DeGennaro LJ, (1991) A highly polymorphic dinucleotide repeat on the proximal short arm of the human X chromosome: Linkage mapping of the SynapsinI/ArafI genes. Am J Hum Genet 49:184-191.

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.

Knight SJL, Voelckel MA, Hirst MC, Flannery AV, Moncla A, Davies KE (1994) Triplet repeat expansion at the FRAXE locus and X-linked mild mental handicap. Am J Hum Genet 55:81-86.

Knowles JA, Vieland VJ, Gilliam TC (1992) Perils of gene mapping with microsatellite markers. Am J Hum Genet 51:905-909.

Koenig M, Hoffman EP, Bertelson CJ, Monaco AP, Feener C, Kunkel LM (1987) Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organisation of the DMD gene in normal and affected individuals. Cell 50:509-517.

Kogan SC, Doherty BS, Gitschier J (1987) An improved method for prenatal diagnosis of genetic disease by analysis of amplified DNA sequences. New Engl J Med 317:985-990.

Koide R,Ikeuchi T, Onodera O, Tanaka H, Igarashi S, Endo K, Takahashi H, Kondo R, Ishikawa A, Hayashi T (1994) Unstable expansion of CAG repeat in hereditary dentatorubral-pallidoluysian atrophy (DRPLA) Nature Genetics 6:9-13.

Kondo I, Tsukamoto K, Okano K, Kanazawa I, Hupkes PE, Niikawa N (1991) A new form of X-linked mental retardation linked to DXS369 at Xq27. Cytogenet Cell Genet 58:2071 (abs only).

Konradi C, Ozelius L, Breakefield XO, (1992) Highly polymorphic (GT)n repeat sequence in intron II of the human MAOB gene. Genomics 12:176-177.

Korn B, Sedlacek Z, Kioschis P, Konecki D, Poustka A (1994) Establishment of an Xq27.3-Xqter specific cDNA library using cDNA selection. Cytogenet Cell Genet 67:356 Abstract 88.

Kozak L, Chiurazzi P, Genuardi M, Pomponi MG, Zollino M, 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, Mulley JC (1993) Integration of the cytogenetic and genetic linkage maps of hman chromosome 16 using 50 physical intervals and 50 polymorphic loci. Cytogenet Cell Genet 62:194-198.

Kremer EJ, Pritchard M, Lynch M, Yu S, Holman K, Baker E, Warren ST, Schlessinger D, Sutherland GR, Richards RI (1991) Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)n. Science 252:1711-1714.

Kvaløy K, Galvagni F, Brown WRA (1994) The sequence organization of the long arm pseudoautosomal region of the human sex chromosomes. Hum Mol Genet 3/5:771-778.

Lafrenière RG, Brown CJ, Powers VE, Carrel L, Davies KE, Barker DF, Willard HF (1991) Physical mapping of 60 DNA markers in the p21.1-q21.3 region of the human X chromosome. Genomics 11:352-363.

Lahn BT, Ma N, Breg WR, Stratton R, Surti U, Page DC (1994) Xq-Yq interchange resulting in supernormal X-linked gene expression in severely retarded males with 46,XYq- karyotype. Nature Genet 8:243-250.

Laing S, Partington M, Robinson H, Turner G (1991) Clinical screening score for the fragile X (Martin-Bell) syndrome. Am J Med Genet 38:256-259.

Lalloz MRA, McVey JH, Pattinson JK, Tuddenham EGD, (1991) Haemophilia A diagnosis by analysis of a hypervariable dinucleotide repeat within the factor VIII gene. The Lancet 338:207-211.

Lalouel JM, Lathrop GM, White R (1986) Construction of human genetic linkage maps: II. Methodological issues. Cold Spring Harbor Symp Quant Biol 51:39-47.

Larsen F, Gundersen G, Lopez R, Prydz H (1992) CpG islands as gene markers in the human genome. Genomics 13:1095-1107.

La Spada AR, Wilson EM, Lubahn DB, Harding AE, Fischbeck KH (1991) Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. Nature 352:77-79.

Lasser DM, Wilhelmsen KC, Nygaard TG, Tantravahi U, (1993) Characterization of mocrosatellite polymorphisms DXS691 and DXS692: genetic mapping to Xq26.2-Xq27 and Xq25-Xq26.2. Genomics 16:785-786.

Lathrop GM, Lalouel JM (1984) Easy calculations of lod scores and genetic risks on small computers. Am J Hum Genet 36:460-465.

Lathrop GM, Lalouel JM, Julier C, Ott J (1984) Strategies for multilocus linkage analysis in humans. Proc Natl Acad Sci USA 81:3443-3446.

Lathrop M, Nakamura Y, Cartwright P, O'Connell P, Leppert M, Jones C, Tateishi H, Bragg T, Lalouel JM, White R (1988) A primary genetic map of markers of human chromosome 10. Genomics 2:157-164.

Laval SH, Boyd Y (1993) Novel sequences conserved on the human and mouse X chromosomes. Genomics 15:483-491.

Lazzarini A, Stenroos ES, Lehner T, McKoy V, Gold B, McKormack MK, Reid CS, Ott J, Johnson WG (1994) STRP linkage studies in a new family with X-linked mental retardation: Tight linkage to DXS458. ASHG meeting AJHG55/3:A192 abstract 1118.

Leal SM, Ott J (1994) A likelihood approach to calculating risk support intervals. Am J Hum Genet 54:913-917.

Lee WC, Ferrero GB, Chinault AC, Yen PH, Ballabio A (1993) A yeast artificial chromosome contig linking the steroid sulphatase and Kallmann syndrome loci on the human X chromosome short arm. Genomics 18:1-6.

Lehrke RG (1974) X-linked mental retardation and verbal disability. The National Foundation - March of Dimes 10/1:1-100.

Lenk U, Hanke R, Thiele H, Speer A (1993) Point mutations at the carboxy terminus of the human dystrophin gene: implications for an association with mental retardation in DMD patients. Hum Molec Genet 2/11:1877-1881.

Li SH, McInnes MG, Margolis RL, Antonarakis SE, Ross CA (1993) Novel tripet repeat containing genes in human brain: cloning, expression and length polymorphisms. Genomics 16:572-579.

Lidov HGW, Byers TJ, Watkins SC, Kunkel LM (1990) Localization of dystrophin to postsynaptic regions of central nervous system cortical neurons. Nature 348:725-728.

Lim LCC, Powell JF, Murray R, Gill M (1994) Monoamine oxidase A gene and bipolar affective disorder. Am J Hum Genet 54:1122-1124.

Lincoln SE, Lander ES (1992) Systematic detection of errors in genetic linkage data. Genomics 14:604-610.

Lindor NM, Michels VV, Hoppe DA, Driscoll DJ, Leavitt JA, Dewald GW, (1992) Xp22.3 microdeletion syndrome with micropthalmia, sclerocornea, linear skin defects and congenital heart defects. Am J Med Genet 44:61-65.

Lindor NM, Sobell JL, Heston LL, Thibodeau SN, Sommer SS (1994) Screening the dystrophin gene suggests a high rate of polymorphism in general but no exonic deletions in schizophrenics. Am J Med Genet (Neuropsych Gen) 54:1-4.

Lindsay S, Bird AP (1987) Use of restriction enzymes to detect potential gene sequences in mammalian DNA. Nature 327:336-338.

Lindsay S, Curtis ARJ, Roustan P, Kamakari S, Thiselton DL, Stephenson A, Bhattacharya SS, (1993) Isolation and Characterizaton of three microsatellite markers in the proximal long arm of the human X chromosome. Genomics 17:208-210.

Litt M, Luty JA (1989) A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am J Hum Genet 44:397-401.

Little RD, Pilia G, Johnson S, d'Urso M, Schlessinger D (1992) Yeast artificial chromosomes spanning 8 megabases and 10-15 centimorgans of human cytogenetic band Xq26. Proc Natl Acad Sci USA 89:177-181.

Lovett M, Kere J, Hinton LM (1991) Direct selection: a method for the isolation of cDNAs encoded by large genomic regions. Proc Natl Acad Sci USA 88:9628-9632.

Lubs HA (1969) A marker X chromosome. Am J Hum Genet 21:231.

Lugenbeel KA, Peier AM, Carson NL, Chudley AE, Nelson DL (1995) Intragenic loss of function mutations demonstrate the primary role of FMR1 in fragile X syndrome. Nature Genetics 10:483-485.

Luty JA, Guo Z, Willard HF, Ledbetter DH, Ledbetter S, Litt M, (1990) Five polymorphic microsatellite VNTRs on the human X chromosome. Am J Hum Genet 46:776-783.

Maestrini E, Rivella S, Tribioli C, Rocchi M, Camerino G, Santachiara-Benerecetti S, Parolini O, Notarangelo LD, Toniolo D (1992) Identification of Novel RFLPs in the Vicinity of CpG Islands in Xq28: Application to the analysis of the Pattern of X Chromosome Inactivation. Am J Hum Genet 50:156-163.

Mahtani MM, Lafrenière RG, Kruse TA, Willard HF (1991) An 18-locus linkage map of the pericentromeric region of the human X chromosome: genetic framework for mapping X-linked disorders. Genomics 10:849-857.

Malmgren H, Sundvall M, Dahl N, Gustavson KH, Annerén G, Wadelius C, Steén-Bondeson ML, Pettersson U (1993) Linkage mapping of a severe X-linked mental retardation syndrome. Am J Hum Genet 52:1046-1052.

Mandel JL (1994) Towards identification of X-linked mental retardation genes: A proposal. Am J Med Genet 51:550-552.

Maniatis T, Fritsh EF, Sanbrook J (1982) Molecular cloning: A laboratory manual. First edition, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Marshall Graves JA, Watson JM (1991) Mammalian sex chromosomes: Evolution of organization and function. Chromosoma 101:63-68.

Martin JP, Bell J (1943) A pedigree of mental defect showing sex-linkage. J Neurol Psychiatry 6:154-157.

Martinez F, Gal A, Palau F, Prieto F (1994) Locus for X-linked non-specific mental retardation is closely linked to DXS85 in a Spanish family. Cytogenet Cell Genet 67:338 Abstract 29. (XCW5)

Martinez F, Gal A, Palau F, Prieto F (1995) Localization of a gene for X-linked nonspecific mental retardation (MRX24) in Xp22.2-p22.3. Am J Med Genet 55:387-390.

Mathews KD, Ardinger HH, Nishimura DY, Buetow KH, Murray JC, Bartley JA (1989) Linkage localization of Börjeson-Forssman-Lehmann syndrome. Am J Med Genet 34:470-474.

Mathews K, Wiles K, Murray J (1993) Mapping of Börjeson-Forssman-Lehmann syndrome in Xq25-Xq26. Proc Fourth X Chromosome Workshop, St. Louis, Missouri, abstract 53.

Matise TC, Perlin M, Chakravarti A (1994) Automated construction of genetic linkage maps using an expert system (MULTIMAP): a human genome linkage map. Nature Genet 6:384-390.

May M, Colleaux L, Murgia A, Aylsworth A, Nussbaum R, Fontes M, 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 4/8:1465-1466.

Mazzarella R, Srivastava AK (1994) Physical linkage of expressed sequence tags (ESTs) to polymorphic markers on the X chromosome. Hum Mol Genet 3/7:1095-1101.

McKusick VA (1992) Mendelian Inheritance in Man: A catalog of human genes and genetic disorders. Tenth edition, The Johns Hopkins University Press, Baltimore and London.

McKusick VA (1994) Mendelian Inheritance in Man: A catalog of human genes and genetic disorders. Eleventh edition, The Johns Hopkins University Press, Baltimore and London.

Meijer H, de Graaff E, Merckx DML, Jongbloed RJE, de Die-Smulders CEM, Engelen JJM, Fryns J-P, Curfs PMG, Oostra BA (1994) A deletion of 1.6kb proximal to the CGG repeat of the FMR1 gene causes the clinical phenotype of the fragile X syndrome. Hum Mol Genet 3/4:615-620.

Meindl A, Hosenfeld D, Brückl W, Schuffenhauer S, Jenderny J, Bacskulin A, Oppermann H-C, Swensson O, Bouloux P, Meitinger T (1993) Analysis of a terminal Xp22.3 deletion in a patient with six monogenic disorders: implications for the mapping of X linked ocular albinism. J Med Genet 30:838-842.

Metzenberg AB, Gitschier J (1992) The gene encoding the palmitoylated erytrocyte membrane protein, p55, originates at the CpG island 3' to the factor VIII gene. Hum Mol Genet 1:97-101.

Metzenberg AB, Pan Y, Das S, Shashidar Pai G, Gitschier J (1994) Molecular evidence that the p55 gene is not responsible for either of two Xq28-linked disorders:Emery-Dreifuss muscular dystrophy and dyskeratosis congenita. Am J Hum Genet 54:920-922.

Miesfield R, Krystal M, Arnhein N (1981) A member of a new repeated sequence family which is conserved throughout eukaryotic evolution is found between the human δ and β globin genes. Nucl Acids Res 9:5931-5947.

Mohandas T, Shapiro LJ, Sparkes RS, Sparkes MC (1979) Regional assignment of the steroid sulfatase-X-linked ichthyosis locus: implications for a noninactivated region on the short arm of human X chromosme. Proc Natl Acad Sci USA 76:5779-5783.

Monaco AP (1994) Isolation of genes from cloned DNA. Curr Op in Genet and Devel 4:360-365.

Monckton DG, Caskey T (1995) Unstable triplet repeat diseases. Circulation 91:513-520.

Moore BJ, Kwan SP, Bech-Hansen NT, (1992) A polymorphic dinucleotide repeat at the DXS7 locus. Nucl Acids Res 20/4:929.

Moraine C, Dessay B, Toutain A, Briault S, Gendrot C, Ronce N (1994) X-linked mental retardation: linkage results in five unrelated families. ASHG meeting AJHG55/3:A196 abstract 1141.

Morton NE (1955) Sequential tests for the detection of linkage. Am J Hum Genet 7:277-318.

Morton NE, Rao DC, Lang-Brown H, Maclean CJ, Bart RD, Lew R (1977) Colchester revisited: A genetic study of mental defect. J Med Genet 14:1-9.

Morton NE (1991) Parameters of the human genome. Proc Natl Acad Sci USA 88:7474-7476.

Morton NE (1992) Genes for intelligence on the X chromosome. J Med Genet 29:71-72.

Moser H (1984) Duchenne muscular dystrophy: Pathogenetic aspects and genetic prevention. Hum Genet 66:17-40.

Mulley JC, Norman PC, Tippett P, Beal RW (1988) A regional localisation for an X-linked suppressor gene (XS) for the Lutheran blood group. Hum Genet 78:127-129.

Mulley JC, Yu S, Gedeon AK, Donnelly A, Turner G, Loesch D, Chapman CJ, Gardner RJM, Richards RI, Sutherland GR (1992a) Experience with direct molecular diagnosis of fragile X. J Med Genet 29:368-374.

Mulley J, Kerr B, Stevenson R, Lubs H (1992) Nomenclature guidelines for X-linked mental retardation. Am J Med Genet 43:383-391.

Mulley JC, Gedeon AK, Wilson S, Haan EA (1992) Use of linkage data obtained in single families. Prenatal diagnosis of a new X-linked mental retardation syndrome. Am J Hum Genet

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

Muntoni F, Cau M, Ganau A, Congiu R, Arvedi G, Mateddu A, Marrosu MG, Cianchetti C, Realdi G, Cao A, Melis MA (1993) Brief report: Deletion of the dystrophin muscle-promoter region associated with X-linked dilated cardiomyopathy. New Engl J Med 329/13:921-925.

Murray JC, Ludwigsen SJ, Buetow KH, (1992) Chromosome X - A Comprehensive Genetic Linkage Map of the Human Genome, NIH/CEPH Collaborative Mapping Group. Science 258: 67-86.

Murray JC, Duyk GM, Sheffield VC, Weber JL, Buetow KH, Weir RF, Newkirk N (1993a) CHLC Report. Cooperative Human Linkage Center vol1/1.

Murray J, Wiles K, Mathews K, Schlessinger D, Ludwigsen S, Buetow K, Weissenbach J (1993b) Short-tandem-repeat-polymorphism-based map of the human X chromosome. Abs 52 XCW4. Cytogenet Cell Genet 64:183.

Nakamura Y, Leppert M, O'Connell P, Wolff TH, Culver M, Martin C, Fujimoto E, Hoff M, Kumlin E, White R, (1987) Variable number of tandem repeat (VNTR) markers for human mapping. Science 235:1616-1622.

Nancarrow JK, Kremer E, Holman K, Eyre H, Doggett NA, Le Paslier D, Callen DF, Sutherland GR, Richards RI (1994) Implications of FRA16A structure for the mechanism of chromosomal fragile site genesis. Science 264:1938-1941.

Neri G, Opitz JM, Mikkelsen M, Jacobs PA, Davies K, Turner G (1988) Conference report: Third International workshop on the fragile X and X-linked mental retardation. Am J Med Genet 30:1-29 plus erratum 31:727-730.

Neri G, Gurrieri F, Gal A, Lubs HA (1991) XLMR genes: Update 1990. Am J Med Genet 38:186-189.

Neri G, Chiurazzi P, Arena JF, Lubs HA, Glass IA (1992) XLMR genes: Update 1992. Am J Med Genet 43:373-382.

Neri G, Chiurazzi P, Arena JF, Lubs HA (1994) XLMR genes: Update 1994. Am J Med Genet 51:542-549.

NIH/CEPH Collaborative Mapping Group (1992) A comprehensive genetic linkage map of the human genome. Science 258:67-86.

Nordqvist K, Lovell-Badge R (1994) Setbacks on the road to sexual fulfilment. Nature Genet 7:7-9.

Nordström A-M, Penttinen M, von Koskull H (1992) Linkage to Xq28 in a family with nonspecific X-linked mental retardation. Hum Genet 90:263-266.

Obérle I, Rousseau F, Heitz D, Kretz C, Devys D, Hanauer A, Boué J, Bertheas MF, Mandel JL (1991) Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. Science 252:1097-1102.

Ogata T, Matsuo N (1994) Testis determining gene(s) on the X chromosome short arm:chromosomal localisation and possible role in testis determination. J Med Genet 31:349-350.

Okubo K, Hori H, Matoba R, Hiiyama T, Fukushima A, Kojima Y, Matsubara K (1992) Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression. Nature Genetics 2:173-179.

OMIM (1995) Online mendelian inheritance in man (TM). Center for Medical Genetics, Johns Hopkins University (Baltimore, MD) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, MD). World wide web URL: http://www3.ncbi.nlm.nih.gov/omim/

Opitz JM, Sutherland GR (1984) International workshop on the fragile X and X-linked mental retardation. Am J Med Genet 17:5-94.

Opitz JM (1986) Editorial comment: On the gates of hell and a most unusual gene. Am J Med Genet 23:1-10.

Opitz JM (1994) Associations and Syndromes: Terminology in clinical genetics and birth defects epidemiology: Comments on Khoury,Moore, and Evans. Am J M Genet 49:14-20.

Orita M, Iwahana H, Kanazawa H, Hayashi K, Sekiya T (1989) Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. Proc Natl Acad Sci USA 86:2766-2770

Orr HT, Chung MY, Banfi S, Kwiatkowski TJ, Servadio A, Beaudet AL, McCall AE, Duvick LA, Ranum LPW, Zoghbi HY (1993) Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type I. Nature Genet 4:221-226.

Ott J (1976) A computer program for linkage analysis of general human pedigrees. Am J Hum Genet 28:528-529.

Ott J (1978) A simple scheme for the analysis of HLA linkages in pedigrees. Ann Hum Genet 42:255-257.

Ott J (1991) Analysis of human genetic linkage. Revised edition, The Johns Hopkins University Press, Baltimore and London.

Ott J (1992) Strategies for characterizing highly polymorphic markers in human gene mapping. Am J Hum Genet 51:283-290.

Oudet C, Heilig R, Mandel JL, (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, Mandel J-L, (1991) Nonradioactive 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, Mandel JL (1992) Two hot spots of recombination in the DMD gene correlate with the deletion prone regions. Hum Mol Genet 1/8:599-603.

Oudet C, Weber C, Kaplan J, Segues B, Croquette MF, Ollagnon Roman E, Hanauer A, (1992) Characterisation of a highly polymorphic microsatellite at the DXS207 locus: confirmation of very close linkage to the retinoschisis disease gene. J Med Genet 30:300-303.

Parkar M, Lovering R, Levinsky RJ, Kinnon C, (1994) Genetic mapping of two loci, DXS454 and DXS458, with respect to the X-linked agammaglobulinemia gene locus. Hum Genet 93:89-90.

Parrish JE, Nelson DL (1993) Regional assignment of 19 X-linked ESTs. Hum Mol Genet 2/11:1901-1905.

Parrish JE, Oostra BA, Verkerk AMJH, Richards CS, Reynolds J, Spikes AS, Shaffer LG, Nelson DL (1994) Isolation of a GCC repeat showing expansion in FRAXF, a fragile site distal to FRAXA and FRAXE. Nature Genet 8:229-235.

Partington MW, Marriott PJ, Prentice RSA, Cavaglia A, Simpson NE (1981) Familial cutaneous amyloidosis with systemic manifestations in males. Am J Med Genet 10:65-75.

Partington MW, Mulley JC, Sutherland GR, Hockey A, Thode A, Turner G (1988) X-linked mental retardation with dystonic movements of the hands. Am J Med Genet 30:251-262.

Partington MW, Prentice RSA (1989) X-linked cutaneous amyloidosis: further clinical and pathological observations. Am J Med Genet 32:115-119.

Partington MW, Ryan AK, Donnelly AJ, Mulley JC (1995) Two contrasting families with X-linked mental retardation (XLMR) from Queensland. 7th International Workshop on the fragile X and X-linked mental retardation. Abs 16.

Passos-Bueno MR, Byth BC, Rosenberg S, Takata RI, Bakker E, Beggs AH, Pavanello RC, Vainzof M, Davies KE, Zatz M (1993) Severe non-specific 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.

Patrosso MC, Repetto M, Villa A, Milanesi L, Frattini A, Faranda S, Mancini M, Maestrini E, Toniolo D, Vezzoni P (1994) The exon-intron organization of the human X-linked gene (FLN1) encoding actin-binding protein 280 Genomics 21/1:71-76.

Pearson PL (1991) The genome data base (GDB)- a human gene mapping repository. Nucl Acids Res 19 Suppl:2237-2239.

Petit C, Melki J, Levilliers J, Serville F, Weissenbach J, Maroteaux P (1990) An interstitial deletion in Xp22.3 in a family with X-linked recessive chondrodysplasia punctata and short stature. Hum Genet 85:247-250.

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.

Philippe C, Cremers FPM, Chery M, Bach I, Abbadi N, Ropers HH, Gilgenkrantz S (1993) Physical mapping of DNA markers in the q13-q22 region of the human X chromosome. Genomics 17:147-152.

Pieretti M, Zhang F, Fu YH, Warren ST, Oostra BA, Caskey CT, Nelson DL (1991) Absence of expression of the FMR1 gene in fragile X syndrome. Cell 66:817-822.

Polymeropoulos MH, Xiao H, Sikela JM, Adams M, Venter CJ, Merril CR (1993) Chromosomal distribution of 320 genes from a brain cDNA library. Nature Genet 4:381-386.

Porteous MEM, Curtis A, Lindsay S, Williams O, Goudie D, Kamakari S, Bhattacharya SS, (1992) The gene for Aarskog syndrome is located between DXS255 and DXS566 (Xp11.2-q13). Genomics 14:298-301.

Powell JF, Fodor FH, Cockburn DJ, Monaco AP, Craig IW, (1991) A dinucleotide repeat polymorphism at the DMD locus. Nucl Acids Res 19/5:1159.

Proops R, Webb T (1981) The 'fragile' X chromosome in the Martin-Bell-Renpenning syndrome and in males with other forms of familial mental retardation. J Med Genet 18:366-373.

Proops R, Mayer M, Jacobs PA (1983) A study of mental retardation in children in the Island of Hawaii. Clin Genet 23:81-96.

Ram KT, Barker DF, Puck JM, (1992) Dinucleotide repeat polymorphism at the DXS441 locus. Nucl Acids Res 20/6:1428.

Rappold GA, Klink A, Weiss B, Fischer C (1994) Double crossover in the human Xp/Yp pseudoautosomal region and its bearing on interference. Hum Mol Genet 3/8:1337-1340.

Rappaport (1991) MR in DMD due to deletion of cf56a in exon 51. Am J Med Genet 39:437-41.

Reed EW, Reed SC (1965) Mental retardation: A family study. Philadelphia: WB Saunders.

Renieri A, Galli L, De Marchi M, Li Volti S, Mollica F, Lupo A, Maschio G, Peissel B, Rossetti S, Pignatti P, Turco AE (1994) Single base pair deletions in exons 39 and 42 of the COL4A5 gene in Alport syndrome. Hum Mol Genet 3/1:201-202.

Renwick JH (1969) Progress in mapping human autosomes. Br Med Bull 25/1:65-73.

Richards BW, Sylvester PE, Brooker C (1981) Fragile X-linked mental retardation: The Martin-Bell syndrome. J Ment Defic Res 25/4:253-256.

Richards RI, Shen Y, Holman K, Kozman H, Hyland VJ, Mulley JC, Sutherland GR, (1991) Fragile X syndrome: diagnosis using highly polymorphic microsatellite markers. Am J Hum Genet 48:1051-1057.

Richards RI, Holman K, Yu S, Sutherland GR (1993) Fragile X syndrome unstable element, $p(CCG)_n$, and other simple tandem repeat sequences are binding sites for specific nuclear binding proteins. Hum Mol Genet 2:1429-1435.

Richards RI, Sutherland GR (1994) Simple repeat DNA is not replicated simply. Nature Genet 6:114-116.

Rider SH, Monaco AP, (1993) Primers for the dinucleotide repeat at the DXS453 locus also recognizes the DXS983 locus. Human Molecular Genetics 2/9:1510.

Riggins GJ, Sherman SL, Oostra BA, Sutcliffe JS, Feitell D, Nelson DL, van Oost BA, Smits APT, Ramos FJ, Pfender E, Kuhl DPA, Caskey T, Warren ST, (1992) Characterization of a highly polymorphic dinucleotide repeat 150kb proximal to the fragile X site. Am J Med Genet 43:237-243.

Ritchie RJ, Knight SJL, Hirst MC, Grewal PK, Bobrow M, Cross GS, Davies KE (1994) The cloning of FRAXF: trinucleotide repeat expansion and methylation at a third fragile site in distal Xqter. Hum Mol Genet 3/12:2115-2121.

Roberts RG, Cole CG, Hart KA, Bobrow M, Bentley DR, (1989) Rapid carrier and prenatal diagnosis of Duchenne and Becker muscular dystropy. Nucl Acids Res 17/2:811.

Roberts RG, Bobrow M, Bentley DR (1992a) Point mutations in the dsytrophin gene. Pro Natl Acad Sci USA 89:2331-2335.

Roberts RG, Gardner RJ, Bobrow M (1994) Searching for the 1 in 2,400,000: A review of dystrophin gene point mutations. Hum Mutation 4:1-11.

Romeo G, McKusick VA (1994) Phenotypic diversity, allelic series and modifier genes. Nature Genet 7:451-453.

Rousseau F, Vincent A, Rivella S, Heitz D, Triboli C, Maestrini E, Warren ST, Suthers GK, Oberlé I (1991) Four chromosomal breakpoints and four new probes mark out a 10-cM region encompassing the fragile-X locus (FRAXA). Am J Hum Genet 48:108-116.

Rousseau F, Heitz D, Mandel JL (1992) The unstable and methylatable mutations causing the fragile X syndrome. Hum Mutation 1:91-96.

Roustan P, Curtis ARJ, Kamakari S, Thiselton D, Lindsay S, Bhattacharya SS, (1992) Dinucleotide repeat polymorphism at the DXS559 locus. Human Molecular Genetics 1/9:778.

Roustan P, Curtis ARJ, Kamakari S, Thiselton D, Lindsay S, Bhattacharya SS, (1993) Dinucleotide repeat polymorphism at the DXS573 locus. Human Molecular Genetics 2/1:92.

Royer-Pokora B, Kunkel LM, Monaco AP, Goff SC, Newburger PE, Bachner RL, Cole FS, Curnutte JT, Orkin SH (1986) Cloning the gene for an inherited human disorder - chronic granulomatous disease - on the basis of its chromosomal location. Nature 322:32-38.

Royle NJ, Armour JAL, Crosier M, Jeffreys AJ (1993) Abnormal segregation of alleles in CEPH pedigree DNAs arising from allele loss in lymphoblastoid DNA. Genomics 15:119-122.

Ruiz JC, Cuppens H, Leguis E, Fryns JP, Glover T, Marynen P, Cassiman JJ (1995) Mutations in L1-CAM in two families with X-linked complicated spastic paraplegia, MASA syndrome, and HSAS. J Med Genet 32:549-552.

Saad FA, Busque C, Vitiello L, Danieli GA, (1993) DXS997 localized to intron 48 of dystrophin. Hum Mol Genet 2/12:2199.

Saccone S, De Sario A, Della Valle G, Bernardi G (1992) The highest gene concentrations in the human genome are in telomeric bands of metaphase chromosomes. Proc Natl Acad Sci USA 89:4913-4917.

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487-491.

Saiki RK, Chang CA, Levenson CH, Warren TC, Boehm CD, Kazazian HH Jr, Erlich HA (1988) Diagnosis of sickle cell anemia and beta-thalassemia with enzymatically amplified DNA and nonradioactive allele-specific oligonucleotide probes. N Engl J Med 319:537-541.

Samanns C, Albrecht R, Neugebauer M, 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.

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463-5467.

Saugier-Veber P, Abadie V, Moncla A, Mathieu M, Piussan C, Turleau C, Mattei JF, Munnich A, Lyonnet S (1993) The Juberg-Marsidi syndrome maps to the proximal long arm of the X chromosome (Xq12-q21). Am J Hum Genet 52:1040-1045.

Saugier-Veber P, Munnich A, Lyonnet S, Toutain A, Moraine C, Piussan C, Mathieu M, Gibbons R (1995) Lumping Juberg-Marsidi syndrome and X-linked α-thalassemia/ mental retardation syndrome. Am J Med Genet 55:300-301.

Schlessinger D, Little R, Freije D, Abidi F, Zucchi I, Porta G, Pilia G, Nagaraja R, Johnson SK, Yoon JY, Srivastava A, Kere J, Palmieri G, Ciccodicola A, Montanaro V, Romano G, Casamassimi A, D'Urso M (1991) Yeast artificial chromosome-based genome mapping: some lessons from Xq24-28. Genomics 11:783-793.

Schlessinger D, Mandel J-L, Monaco AP, Nelson DL, Willard HF (1993) Report of the Fourth International Workshop on Human X Chromosome Mapping 1993, Cytogenet Cell Genet 64:147-194.

Schmidt M, Certoma A, Du Sart D, Kalitsis P, Leversha M, Fowler K, Sheffield L et al (1990) Unusual X chromosome inactivation in a mentally retarded girl with an interstitial deletion Xq27: implications for the fragile X syndrome. Hum Genet 84:347-352.

Schmidtke J, Cooper DN (1990) A comprehensive list of cloned human DNA sequences. Nucl Acids Res 18/Suppl:2413-2414.

Schmidtke J, Cooper DN (1991) A comprehensive list of cloned human DNA sequences - 1990 update. Nucl Acids Res 19/Suppl:2111.

Schmitt K, Lazzeroni LC, Foote S, Vollrath D, Fisher EMC, Goradia TM, Lange K, Page DC, Arnheim N (1994) Multipoint linkage map of the human pseudoautosomal region, based on single-sperm typing: Do double crossovers occur during male meiosis? Am J Hum Genet 55:423-430.

Schmitt K, Goodfellow PN (1994) Predicting the future. Nature Genetics 7:219.

Schwartz CE, May M, Huang T, Ledbetter D, Anderson G, Barker DF, Lubs HA, Arena F, Stevenson RE (1992) MRX8: An X-linked mental retardation condition with linkage to Xq21. Am J Med Genet 43:467-474.

Schwartz CE (1993) Invited editorial: X-linked mental retardation: in pursuit of a gene map. Am J Hum Genet 52:1025-1031.

Schwartz CE, Ouzts L, Gibson A, Cadle R, Arena JF, Boyd E, Hall B, Lubs HA, Stevenson RE (1994) Renpenning Syndrome: Evidence for pericentric location of the gene in two families, including the original Renpenning family. Am J Med Genet 51/4:611 Abstract 26 (XLMR6).

Schwartz CE, et al, (1994) Cytogenet Cell Genet abs74 XCW5.

Sherman SL, Morton NE, Jacobs PA, Turner G (1984) The marker(X) syndroem: a cytogenetic and genetic analysis. Ann Hum Genet 48: 21-37.

Sherman SL, Jacobs PA, Morton NE et al (1985) Further segregtion analysis of the fragile X syndrome with special reference to transmitting males. Hum Genet 69:289-299.

Sherman SL (1991) Workshop report: Combining genetic and physical maps. Cytogenet Cell Genet 58:1842-1843.

Silvestri G, Santorelli FM, Shankse S et al (1994) A new mtDNA mutation in the $tRNA^{Leu(UUR)}$ gene associated with maternally inherited cardiomyopathy. Hum Mutat 3:37-43.

Sims KB, Lebo RV, Benson G, Shalish C, Schuback D, Chen ZY, Bruns G, Craig IW, Golbus MS, Breakefield XO (1992) The Norrie disease gene maps to a 150kb region on chromosome Xp11.3. Hum Mol Genet 1/2:83-89.

Sinclair AH, Berta P, Palmer MS, Hawkins JR, Griffiths B, Smith M, Foster J, Frischauf AM, Lovell-Badge R, Goodfellow PN (1990) A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature 346:240-244.

Sklower Brooks S, Wisniewski K, Brown WT (1994) New X-linked mental retardation (XLMR) Syndrome with distinct facial appearance and growth retardation. Am J Med Genet 51:586-590.

Smahi A, Peterlin B, Onodera O, Tsuji S, Fulchignoni-Lataud MC (1994) Use of a radiation-reduced hybrid panel for the localisation of seven markers in the Xq28 region of the human genome. Ann Génét 37/1:11-13.

Smith CAB (1986) The development of human linkage analysis. Ann Hum Genet 50:293-311.

Southern EM (1975) Detection of specific seqences among DNA fragments separated by gel elctrophoresis. J Mol.Biol. 98: 503-517.

Spurr NK (1992) Standards for reporting alleles at highly polymorphic loci: a proposal. Hum Mol Genet 1/3:211-212.

Stallings RL, Ford AF, Nelson D, Torney DC, Hildebrand CE, Moyzis RK (1991) Evolution and distribution of $(GT)_n$ repetitive sequences in mammalian genomes. Genomics 10:807-815.

Stanier P, Newton R, Forbes SA, Ivens A, Moore GE (1991) Polymorphic dinucleotide repeat at the DXS3 locus. Nucl Acids Res 19/17:4793.

Stevanovic M, Lovell-Badge R, Collignon J, Goodfellow PN (1993) SOX3 is an X-linked gene related to SRY. Hum Mol Genet 2/12:2013-8.

Stevenson RE, Schwartz CH, Arena JF, Lubs HA (1994) X-Linked Mental Retardation: The early era from 1943-1969. Am J Med Genet 51:538-541.

Sutherland GR (1977) Fragile sites on human chromosomes: Demonstration of their dependence on type of tissue culture medium. Science 197:265-266.

Sutherland GR, Ashforth PLC (1979) X-linked mental retardation with macro-orchidism and the fragile site at Xq27 or 28. Hum Genet 48:117-128.

Sutherland GR, Hecht F (1985) Fragile sites on human chromosomes. Oxford monographs on medical genetics no.13. Oxford University Press.

Sutherland GR, Gedeon AK, Kornman L, Donnelly A, Byard RW, Mulley JC, Kremer E, Lynch M, Pritchard M, Yu S, Richards R (1991) Prenatal diagnosis of fragile X syndrome by direct detection of the unstable DNA sequence. New Engl J Med 325:1720-1722.

Sutherland GR, Baker E (1992) Characterisation of a new rare fragile site easily confused with the fragile X. Hum Mol Genet 1:111-113.

Sutherland GR, Richards RI (1995) Simple tandem DNA repeats and human genetic disease. Proc Natl Acad Sci USA 92:3636-3641.

Suthers GK, Turner G, Mulley JC (1988) A non-syndromal form of X-linked mental retardation (XLMR) is linked to DXS14. Am J Med Genet 30:485-491.

Suthers GK, Hyland VJ, Callen DF, Oberle I, Rocchi M, Thomas NS, Morris CP, Schwartz CE, Schmidt M, Ropers HH, Baker E, Oostra BA, Dahl N, Wilson PJ, Hopgood JJ, Sutherland GR (1990) Physical mapping of new DNA probes near the fragile X mutation (FRAXA) by using a panel of cell lines. Am J Hum Genet 47: 187-195.

Suthers GK, Wilson SR (1990) Genetic counselling in rare syndromes: A resampling method for determining an approximate confidence interval for gene location with linkage data from a single pedigree. Am J Hum Genet 47:53-61.

Suthers GK, Hyland VJ, Callen DF, Oberlé I, Rocchi M, Thomas NS, Morris CP, Schwartz CE, Schmidt M, Ropers HH, Baker E, Oostra BA, Dahl N, Wilson PJ, Hopwood JJ, Sutherland GR (1990) Physical mapping of new DNA probes near the fragile X (FRAXA) by using a panel of cell lines. Am J Hum Genet 47:187-195.

Suthers GK (1990) The human gene map near the fragile X. PhD thesis, The University of Adelaide, Adelaide, South Australia.

Suthers GK, Oberle I, Nancarrow J, Mulley JC, Hyland VJ, Wilson PJ, McCure J, Morris CP, Hopwood JJ, Mandel JL, Sutherland GR (1991) Genetic mapping of new RFLPs at Xq27-q28 Genomics 9:37-43 - Xq28.

Tarleton J, Richie R, Schwartz C, Rao K, Aylsworth AS, Lachiewicz A (1993) An extensive de novo deletion removing FMR1 in a patient with mental retardation and the fragile X syndrome phenotype. Hum Mol Genet 2:1973-1974.

Tariverdian G, Weck B (1982) Nonspecific X-linked mental retardation - a review. Hum Genet 62:95-109.

Tautz D, Renz M (1984) Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucl Acids Res 12:4127-4138.

Thiselton DL, Lindsay S, Kamakari S, Hardcastle AJ, Roustan P, Bhattacharya SS (1995) Genetic and physical mapping of five novel microsatellite markers on human Xp21.1-p11.22. Genomics 25:279-281.

Thode A, Laing S, Partington MW, Turner G (1988) Is there a fragile(X) negative Martin-Bell syndrome? Am J Med Genet 30:459-471.

Todd JA (1992) La carte des microsatellites est arrivée! Hum Mol Genet 1/9:663-666.

Tommerup N (1993) Mendelian cytogenetics. Chromosome rearrangements associated with mendelian disorders. J Med Genet 30:713-727.

Towbin JA, Hejtmancik F, Brink P, Gelb B, Zhu XM, Camberlain JS, McCabe ERB, Swift M, (1993) X-linked Dilated Cardiomyopathy: Molecular genetic evidence of linkage to the Duchenne Muscular Dystrophy (dystrophin) gene at the Xp21 locus. Circulation 87/6:1854-1864.

Tribioli C, Mancini M, Plassart E, Bione S, Rivella S, Sala C, Torri G, Toniolo D (1994) Isolation of new genes in distal Xq28: transcriptional map and identification of a human homologue of the ARD1 N-acetyl transferase of *Saccharomyces cerevisiae*. Hum Mol Genet 3/7:1061-1067.

Trottier Y, Imbert G, Poustka A, Fryns J-P, Mandel J-L, (1994) A male with typical fragile X phenotype is deleted for part of the FMR1 gene and for about 100kb of upstream region. Am J Med Genet 51:454-457.

Tugendreich S, Bassett DE Jr, McKusick VA, Boguski MS, Hieter P (1994) Genes conserved in yeast and humans. Hum Mol Genet 3:1509-1517.

Turner G, Turner B, Collins E (1971) X linked mental retardation without physical abnormality: Renpenning's syndrome. Develop Med Child Neurol 13:71-78.

Turner G and Turner B (1974) X linked mental retardation. J Med Genet 11:109-113.

Turner G, Daniel A, Frost M (1980) X-linked mental retardation, macro-orchidism, and the fragile site. J Pediatrics 96/5:837-841.

Turner G, Opitz J, Brown WT, Davies KE, Jacobs PA, Jenkins EC, Mikkelsen M, Partington M, Sutherland GR (1986) Conference report: Second international workshop on the fragile X and X-linked mental retardation. Am J Med Genet 23:11-67.

Turner G, Partington MW (1991) Genes for intelligence on the X chromosome. J Med Genet 28:429-432.

Van der Maarel SM, Rohme D, Bach I, Gilgenkranz S, Olde Weghius D, Kere J, Cremers FPM, Ropers HH (1994) Cloning of a candidate gene for X-linked mental retardation in Xq13.1. Cytogenet Cell Genet 67:349 Abstract 68.

Van de Vosse E, Booms PFM, Vossen RHAM, Wapenaar MC, Van Ommen GJB, Den Dunnen JT, (1993) A CA-repeat polymorphism near DXS418 (P122). Hum Mol Genet 2/12:2202.

Van Heyningen V (1994) One gene - four syndromes. Nature 367:319-320.

Vanyukov MM, Moss HB, Yu LM, Tarter RE, Deka R (1995) Preliminary evidence for an association of a dinucleotide repeat polymorphism at the MAOA gene with early onset alcoholism/substance abuse. Am J Med Genet (Neuropsychiatric Genetics) 60:122-126.

Verkerk AJMH, Pieretti M, Sutcliffe JS, Fu YH, Kuhl DPA, Reiner O, Richards S, Victoria MF, Zhang F, Eussen BE, van Ommen GJB, Blonden LAJ, Riggins GJ, Chastain JL, Kunst CB, Galjaard H, Caskey CT, Nelson DL, Oostra BA, Warren ST (1991) Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell 65:905-914.

Villard L, Gecz J, Mattéi JF, Fontés M (1996) XNP mutation in a large family with Juberg-Marsidi syndrome. Nature Genet 12:359-360.

Vits L, van Camp G, Coucke P, Fransen E, De Boulle K, Reyniers E, Korn B, Poustka A, Wilson G, Schrander-Stumpel C, Winter RM, Schwartz C, Willems PJ (1994) MASA

syndrome is due to mutations in the neural cell adhesion gene L1CAM. Nature Genet 7:402-413.

Wang LH, Collins A, Lawrence S, Keats BJ, Morton NE (1994) Integration of gene maps: Chromosome X. Genomics 22: 590-604.

Watson JM, Spencer JA, Riggs AD, Marshall Graves JA (1991) Sex chromosome evolution: Platypus gene mapping suggests that part of the human X chromosome was originally autosomal. Prc Natl Acad Sci USA 88:11256-11260.

Webb TP, Bundey S, Thake A, Todd J (1986) The frequency of the fragile X chromosome among school children in Coventry. J Med Genet 23:396-399.

Weber JL, 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 (1990) Informativeness of human $(dC\text{-}dA)_n$ $(dG\text{-}dT)_n$ polymorphisms. Genomics 7:524-530.

Weber JL, Kwitek AE, May PE, Polymeropoulos MH, Ledbetter S, (1990) Dinucleotide repeat polymorphisms at the DXS453, DXS454 and DXS458 loci. Nucl Acids Res 18/13:4037.

Weber C, Oudet C, Johnson S, Pilia G, Schlessinger D, Hanauer A, (1993a) Dinucleotide repeat polymorphism at Xq26.1 (DXS1114). Hum Mol Genet 2/5:612.

Weber C, Oudet C, Johnson S, Pilia G, Schlessinger D, Hanauer A, (1993b) Dinucleotide repeat polymorphism close to IDS gene in Xq27.3-q28 (DXS1113). Hum Mol Genet 2/5:612.

Weber JL, Wong C (1993) Mutation of human short tandem repeats. Hum Mol Genet 2/8:1123-1128.

Wehnert M, Reiner O, Caskey CT, (1993) Four STR polymorphisms map to a 500kb region between DXS15 and DXS134. Hum Mol Genet 2/9:1503.

Weissenbach J, Gyapay G, Dib C, Vignal A, Morisette J, Millasseau P, Vaysseix G, Lathrop M (1992) A second generation linkage map of the human genome. Nature 359:794-801.

White RL, Lalouel JM, Nakamura Y, Donis-Keller H, Green P, Bowden DW, Mathew CGP, Easton DF, Robson EB, Morton NE (1990) The CEPH consortium primary linkage map of chromosome 10.Genomics 6:393-412.

Wicking C, Williamson B (1991) From linked marker to gene. Trends Genet 7:288-293.

Willard HF, Skolnick MH, Pearson PL, Mandel JL (1985) Report of the committee on human gene mapping by recombinant DNA techniques.Cytogenet Cell Genet 40:360-489

Willard HF, Cremers F, Mandel J-L, Monaco AP, Nelson DL, Schlessinger D (1994) Report of the Fifth International Workshop on Human X Chromosome Mapping 1994, Cytogenet Cell Genet 67:296-358.

Willems P, Vits L, Buntinx I, Raeymaekers P, Van Broeckhoven C, Ceulemans B (1993) Localization 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 P, Chipperfield M, Cooper DN, Coutelle C, Hewitt J, Klinger K, Langley K, Beckmann J, Tolley M, Maidak B (1991) HGM11-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 GN, Richards CS, Katz K, Brookshire GS (1992) Non-specific X-linked mental retardation with aphasia exhibiting genetic linkage to chromosomal region Xp11. J Med Genet 29:629-634.

Wilson GN (1993) Reply to Gordon. J Med Genet 30:262-264.

Wöhrle D, Kotzot D, Hirst MC, Manca A, Korn B, Schmidt A, Barbi G, Rott H-D, Poustka A, Davies KE, Steinbach P, (1992) A microdeletion of less than 250kb, including the proximal part of the FMR1 gene and the fragile X site, in a male with the clinical phenotype of Fragile X syndrome. Am J Hum Genet 51:299-306.

Wooster R, Cleton-Jansen AM, Collins N, Mangion J, Cornelis RS, Cooper CS, Gusterson BA, Ponder BAJ, von Deimling A, Wiestler OD, Cornelisse CJ, Devilee P, Stratton MR (1994) Instability of short tandem repeats (microsatellites) in human cancers. Nature Genet 6:152-156.

Wortis H, Pollack M, Wortis J (1966) Families with 2 or more mentally retarded or mentally disturbed siblings:the preponderance of males. Am J Ment Defic 70/5:745-752.

Wu BL, Milunsky A, Nelson D, Schmeckpeper B, Porta G, Schlessinger D, Skare J (1993) High-resolution mapping of probes near the X-linked lymphoproliferative disease (XLP) locus. Genomics 17:163-170.

Wyman AR, White R (1980) A highly polymorphic locus in human DNA. Proc Natl Acad Sci USA 77:6754-6758.

Yarbrough KM, Howard-Peebles PN (1976) X-linked non-specific mental retardation. Clin Genet 9:125-130.

Yates JRW, Warner JP, Smith JA, Deymeer F, Azulay J-P, Hausmanowa-Petrusewicz, Zaremba J, Borkowska J, Affara NA, Ferguson-Smith MA, (1993) Emery-Dreifuss muscular dystrophy: linkage to markers in distal Xq28. J Med Genet 30:108-111.

Yu S (1992) Molecular basis of fragile X syndrome. PhD thesis, The University of Adelaide, Adelaide, South Australia.

Yu S, Pritchard M, Kremer E, Lynch M, Nancarrow J, Baker E, Holman K, Mulley J, Warren ST, Schlessinger D, Sutherland GR, Richards RI (1991) Fragile X genotype characterized by an unstable region of DNA. Science 252:1179-1181.

Yu S, Mulley J, Loesch D, Turner G, Donnelly A, Gedeon A, Hillen D, Kremer E, Lynch M, Pritchard M, Sutherland GR, Richards RI (1992) Fragile-X syndrome: Unique genetics of the heritable unstable element. Am J Hum Genet 50:968-980.

Zatz M, Vallada H, Melo MS, Passos-Bueno MR, Vieira AHG, Vainzof M, Gill M, Gentil V (1993) Cosegregation of schizophrenia with Becker muscular dystrophy: susceptibility locus for schizophrenia at Xp21 or an effect of the dystrophin gene in the brain? J Med Genet 30:131-134.

Zhang Y, McMahon R, Charles SJ, Green JS, Moore AT, Barton DE, Yates JR (1993) Genetic mapping of the Kallmann syndrome and X-linked ocular albinism gene loci. J Med Genet 30/11 923-925.

APPENDIX I

Publications from Chapter 3

The following publications arose directly from the research described in Chapter 3 of this thesis. The role of the candidate is specified, and copies of the manuscripts follow.

1. **Gedeon AK**, Richards RI, Mulley JC (1991) Dinucleotide repeat polymorphisms at the DXS294 and DXS300 loci in Xq26. Nucl Acids Res 19/18:5087.

The candidate characterised two novel polymorphic markers from ACn positive lambda clones and designed primers flanking the repeats. Allele sizes and allele frequencies were determined by the candidate. The report was prepared and submitted by the candidate. These polymorphic markers comprise part of the background map of the X chromosome (Chapter 3) and lie within the interval of the BFLS gene (Chapter 5).

2. **Gedeon AK**, Holman K, Richards RI, Mulley JC (1992) Characterization 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.

The candidate was wholly responsible for the design and completion of this work. Primer design and optimisation, as well as genotyping of CEPH and unrelated blood bank samples to determine allele sizes and population frequencies were carried out by the candidate. The manuscript was prepared and submitted by the candidate. These polymorphic markers comprise part of the background map of the X chromosome (Chapter 3), the latter lies within the interval of the BFLS gene (Chapter 5) and is an important diagnostic tool.

3. Donnelly A, Kozman H, **Gedeon AK**, Webb S, Lynch M, Sutherland GR, Richards RI, Mulley JC (1994) A linkage map of microsatellite markers on the human X chromosome. Genomics 20:363-370.

The candidate genotyped polymorphic microsatellites through 40 CEPH families at 5 loci (including the 4 described in Papers 1 and 2 above) to increase resolution of the PCR-based genetic map. This was part of a team effort to improve the comprehensive background map of the X chromosome to facilitate mapping of genes in families segregating an X-linked disorder.

APPENDIX I

Paper 1

Dinucleotide repeat polymorphisms at the DXS294 and DXS300 loci in Xq26

A.K.Gedeon, R.I.Richards and J.C.Mulley
Department of Cytogenetics and Molecular Genetics, Adelaide Children's Hospital, North Adelaide, South Australia 5006, Australia

Source/Description: The human genomic DNA phage preparations VK17 (DXS294) and VK29 (DXS300) were positive when hybridised with poly(dC-dA).poly(dG-dT) (1). Fragments from Sau3AI digests were subcloned into M13m19 and positive clones selected for sequencing. Primers were designed from the sequences flanking the repeats.
DXS294 locus: Clone designation VK17AC, predicted length 142 bp. The dinculeotide repeat was of the form $(GT)_2AC(GT)_{23}$.

Primer Sequences:
forward: 5'GCC TGT GCC ACA TTT CTG TCA TTA C 3'
reverse: 5'GGC CTA AAT GAG CTT TTA GTC CAG G 3'

Frequency: Estimated in 104 chromosomes of unrelated individuals. PIC = 0.75.

Allele (bp)	Frequency	Allele (bp)	Frequency
B1 148	0.01	B6 138	0.03
B2 146	0.08	B7 136	0.14
B3 144	0.23	B8 134	0.02
B4 142	0.07	B9 128	0.01
B5 140	0.40	B10 122	0.01

DXS300 locus: Clone designation VK29AC, predicted length 206 bp. The dinucleotide repeat sequence was of the form $(GT)_{15}$.

Primer Sequences:
forward: 5'GAC TCA GGA TGT TGA TAG TAC CAG C 3'
reverse: 5'AAG TAA ATG GCT ACC ACT 3'

Frequency: Estimated in 83 chromosomes of unrelated individuals. PIC = 0.40

Allele	Frequency
A1 208	0.01
A2 206	0.26
A3 204	0.73

Chromosomal Localization: Both DXS294 and DXS300 have been mapped to the interval between somatic cell hybrids LL556p and CY2 with breakpoints in Xq26 (2).

Mendelian Inheritance: X-linked codominant inheritance was observed in 16 CEPH families informative at DXS300 and in 28 CEPH pedigrees informative at the DXS294 locus.

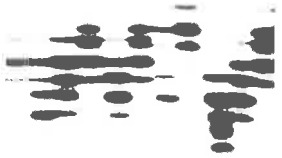
PCR Conditions: PCR was performed according to (3) except that 100 ng genomic DNA samples were processed through 35 cycles of amplification. The first 10 cycles consisting of 1 min at 94°C, 1.5 min at 60°C and 1.5 min at 72°C, followed by 25 cycles with the annealing temperature reduced to 55°C. 10 μCi α-P^{32} dCTP were incorporated during PCR. $MgCl_2$ concentrations were optimised at 4.5 mM for DXS294 and 6.0 mM for DXS300. The respective sizes of the products allow for multiplexing the reactions in the same tube at a final $MgCl_2$ concentration of 6.0 mM.

Other Comments: The use of multiplex conditions allows rapid typing of two closely linked loci while conserving sample DNA. Inclusion of these markers as well as those in (4) into the CEPH linkage map at Xq26 may render these markers useful for diagnosis of Borjeson-Forssman-Lehmann syndrome.

Acknowledgements: This work was supported by the Channel 7 Children's Medical Research Foundation of South Australia. We thank K.Holman for synthesis of oligodeoxyribonucleotides.

References: 1) Richards,R.I., Shen,Y., Holman,K., Hyland,V., Mulley,J.C. and Sutherland,G.R. (1991) *Am. J. Hum. Genet.* **48**, 1051–1057. 2) Suthers,G.K., Hyland,V.J., Callen,D.F., Oberle,I., Rocchi,M. *et al.* (1990) *Am. J. Hum. Genet.* **47**, 187–195. 3) Mulley,J.C., Gedeon,A.K., White,S.J., Haan,E.A. and Richards,R.I. (1991) *J. Med. Genet.* **28**, 448–452. 4) Gedeon,A.K., Holman,K., Richards,R.I. and Mulley,J.C. (1991) Submitted.

DXS300
(VK29AC)



DXS294
(VK17AC)

The butyrylcholinesterase gene (BCHE) at 3q26.2 shows two RFLPs

P.J.McAlpine, M.Dixon, P.W.Allderdice[1], O.Lockridge[2] and B.N.La Du[2]
Department of Human Genetics, University of Manitoba, Winnipeg, MB, R3E OW3, [1]Faculty of Medicine, Memorial University of Newfoundland, St John's, NF, A1B 3V6, Canada and [2]Department of Pharmacology, University of Michigan Medical School, Ann Arbor, MI 48109-0626, USA

Source and Description: Exon 1 probe has a 0.8 kb PstI-HindIII insert and exon 3 has a 1.7 kb EcoRI-XbaI insert (Arpagaus *et al.*, 1990).

Polymorphism: The exon 1 probe identifies two alleles on PstI digests at 23.1 kb (C1) and 4.4 kb (C2). The exon 3 probe identifies two alleles on MspI digests at 10.5 kb (D1) and 5.4 kb (D2). To date complete correspondence in phenotypes with both enzyme-probe combinations found.

Frequency:

MspI:	C1: 0.87
	C2: 0.13
PstI:	D1: 0.87
	D2: 0.13

Studied in 19 Caucasian Canadians, 11 females, 8 males.

Not Polymorphic For: Not polymorphic with EcoRI, PvuII, TaqI.

Mendelian Inheritance: Codominant inheritance demonstrated in several two and three generation families.

Probe Availability: From ATCC.

Other Comments: No problems with RFLP analysis under normal stringency.

Acknowledgement: MRC Canada (MT6112:PJM); CEIC Canada (A03142−6:PWA); NIH grant GM27028 (BNL).

Reference: Arpagaus,M. *et al.* (1990) *Biochem.* **29**, 124.

VNTR polymorphism in the hepatic lipase gene (LIPC)

S.Bhattacharya, D.Ameis[1], P.Cullen, T.M.Narcisi, J.Bayliss, H.Greten[1], M.C.Schotz[2] and J.Scott*
Division of Molecular Medicine, Clinical Research Centre, Harrow, Middlesex HA1 3UJ, UK, [1]Medizinische Klinik, Universitats-Krankenhaus Eppendorf, D-2000 Hamburg 20, FRG and [2]VA Wadsworth Medical Center, Los Angeles, CA 90073 and Dept. of Medicine, University of California, Los Angeles, CA 90073, USA

Source/Description: A CT repeat is located in intron 8 of the human hepatic lipase gene at the 3′ end of an Alu sequence (1). Two oligonucleotides flanking the CT repeat and including the poly-T tail of the Alu sequence, HLIP1 and HLIP2, were used to selectively amplify the sequence from genomic DNA by the polymerase chain reaction (PCR).

Primer Sequences:
HLIP1 = ATGTGATGTCAGTGCTGCCAGTCCA
HLIP2 = ACTGACATTTGAAAGATACGACCAC

Frequency: This was estimated in 38 unrelated Caucasian English individuals:

Allele (nt)	Frequency
174	0.013
171	0.184
169	0.013
167	0.645
165	0.145

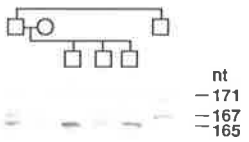
The heterozygosity index was 0.53 and polymorphic information content 0.42.

Chromosomal Localization: Hepatic lipase gene has been assigned to chromosome 15q21-q23 (2).

Mendelian Inheritance: Codominant segregation was observed in 5 families with 28 meioses.

Other Comments: The PCR reaction was performed on genomic DNA (1) using end labelled oligo HLIP1 and unlabelled HLIP2. DNA was denatured at 94°C for 5 minutes, followed by 30 one minute cycles of denaturing at 94°C, annealing at 48°C, and extension at 72°C, with a final extension step of 9 minutes. The PCR buffer was made up to 14 mmol $MgCl_2$, and 0.25 μl of PerfectMatch™ (Stratagene) was added to the 25 μl reaction volume. The PCR products were sized on a 6% denaturing polyacrylamide gel by simultaneously running the dideoxy chain termination reaction products of phage M13mp18 (Sequenase 2.0). The odd size of allele 1 may be due to polymorphism in the poly-T region of the Alu-sequence.

References: 1) Ameis,D. *et al.* (1990) *JBC* **265**, 6552−6555. 2) Sparkes,R.S. *et al.* (1987) *Cytogenet. Cell Genet.* **46**, 697. 3) Saiki,R.K. *et al.* (1988) *Science* **230**, 487−491.



* To whom correspondence should be addressed

APPENDIX I

Paper 2

A.K. Gedeon, K. Holman, R.I. Richards and J.C. Mulley (1992) Characterization of new PCR based markers for mapping and diagnosis: AC dinucleotide repeat markers at the DXS237 (GMGX9) and DXS102 (cX38.1) loci.
American Journal of Medical Genetics, v. 43(1/2), pp. 255–260, April/May 1992

APPENDIX I

Paper 3

A. Donnelly, H. Kozman, A.K. Gedeon, S. Webb, M. Lynch, G.R. Sutherland, R.I. Richards and J.C. Mulley (1994) A Linkage Map of Microsatellite Markers on the Human X Chromosome.
Genomics, v. 20 (3), pp. 363-370, April 1994

APPENDIX II

Publications from Chapter 4

This appendix contains papers prepared during candidature relevant to non-specific X-linked mental retardation (MRX). These publications arose directly from the research described in Chapter 4 of this thesis. The role of the candidate is specified, and reprints/copies of the manuscripts follow.

1. **Gedeon A**, Kerr B, Turner G, Mulley J, (1991) Localisation of the MRX3 gene for non-specific X linked mental retardation. J Med Genet 28:372-377.

The candidate performed the DNA extraction, molecular genotyping, two-point linkage analysis and gene localisation in this family (Chapter 4 - MRX3). The paper was written and submitted by the candidate.

2. Kerr B, Turner G, Mulley J, **Gedeon A**, Partington M, (1991) Non-specific X linked mental retardation. J Med Genet 28:378-382.

The candidate contributed to discussion concerning classification of MRX genes.

3. Kerr B, **Gedeon A**, Mulley J, Turner G (1992) Localization of non-specific X-linked mental retardation genes. Am J Med Genet 43:392-401.

The candidate performed all aspects of molecular and linkage analyses for gene localisation in four families (Chapter 4 - MRX10, MRX11, MRX12 and MRX13) and examined a candidate gene for involvement. The gene interval in a fifth family (Chapter 4 - MRX1) was reduced by the candidate and thus resolved pericentromeric MRX genes into two distinct intervals. The candidate wrote the molecular and linkage portions of the paper and was extensively involved in preparation of the entire manuscript.

4. **Gedeon A**, Kerr B, Mulley J, Turner G, (1994) Pericentromeric genes for non-specific X-linked mental retardation (MRX). Am J Med Genet 51:553-564.

The candidate performed the molecular and linkage analyses, and defined the gene localisations in three families (Chapter 4 - MRX17, MRX18 and Family E/E). The paper was prepared and submitted by the candidate.

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

The candidate established an extensive library of primers as a resource for gene mapping (Chapter 3) and provided tabulated data on marker order, source, optimal $MgCl_2$ concentration, allele size and frequency and optimal cycling conditions for each primer pair, and simultaneous multiplex performance in pairs and sets of loci. Availability of this information vastly hastened the localisation of this gene for MRX.

6. **Gedeon AK**, Glass IA, Connor JM, Mulley JC (1996) Genetic localisation of MRX27 to Xq24-26 defines another discrete gene for non-specific X-linked mental retardation. Am J Med Genet 64:121-124.

The candidate established the collaboration and performed the molecular and linkage analyses to map the gene in this family (Chapter 4 - MRX27). The manuscript was written by the candidate.

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

The candidate performed the molecular and linkage analyses for reduction of gene localisations in several families (Chapter 4) and collated data from published regional localisations of MRX genes. The manuscript was prepared by the candidate.

APPENDIX II

Paper 1

A. Gedeon, B. Kerr, J. Mulley and G. Turner (1991) Localisation of the MRX3 gene for non-specific X linked mental retardation.
Journal of Medical Genetics, v. 28 (6), pp. 372-377, June 1991

APPENDIX II

Paper 2

B. Kerr, G. Turner, J. Mulley, A. Gedeon and M. Partington (1991) Non-specific X linked mental retardation.
Journal of Medical Genetics, v. 28 (6), pp. 378-382, June 1991

APPENDIX II

Paper 3

B. Kerr, A.Gedeon, J. Mulley and G.Turner (1992) Localization of non-specific X-linked mental retardation genes.
American Journal of Medical Genetics, v. 43 (1/2), pp. 392–401, April/May 1992

APPENDIX II

Paper 4

A.Gedeon, B. Kerr, J. Mulley and G.Turner (1994) Pericentromeric Genes for Non-Specific X-Linked Mental Retardation (MRX).
American Journal of Medical Genetics, v. 51 (4), pp. 553–564, July 1994

APPENDIX II

Paper 5

A.J. Donnelly, K.H.A. Choo, H.M. Kozman, A.K. Gedeon, D.M. Danks and J.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), pp. 581–585, July 1994

APPENDIX II

Paper 6

A.K. Gedeon, I.A. Glass, J.M. Connor and J.C. Mulley (1996) Genetic localisation of MRX27 to Xq24-26 defines another discrete gene for non-specific X-linked mental retardation.
American Journal of Medical Genetics, v. 64 (1), pp. 121–124, July 1996

APPENDIX II

Paper 7

A.K. Gedeon, A.J. Donnelly, J.C. Mulley, B. Kerr and G. Turner (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

APPENDIX III

Publications from Chapter 5

This appendix lists relevant papers that were published immediately prior to or during candidature. These publications arose directly from the research described in Chapter 5 of this thesis. The role of the candidate is specified, and copies of the manuscripts follow.

1. Sutherland GR, **Gedeon AK**, Haan EA, Woodroffe P, Mulley JC (1988) Linkage studies with the gene for an X-linked syndrome of mental retardation, microcephaly and spastic diplegia (MRX2). Am J Med Genet 30:493-508.

The candidate performed the gene mapping for this manuscript prior to commencement of candidature. Further work on this family (Chapter 5 - SHS) including reduction of the initial gene localisation is described in Paper 9 (below).

2. Turner G, **Gedeon A**, Mulley J, Sutherland G, Rae J, Power K, Arthur I (1989) Börjeson-Forssman-Lehmann Syndrome: Clinical manifestations and gene localisation to Xq26-27. Am J Med Genet 34:463-469.

The candidate performed the gene mapping for this manuscript prior to commencement of candidature. Initial gene localisation involved Southern analyses and pairwise lod scores were calculated using the LIPED computer program. Further work on this family (Chapter 5 - BFLS) including prenatal diagnoses and refinement of the background map of the region are described in Paper 7 (below).

3. Wilson M, Mulley J, **Gedeon A**, Robinson H, Turner G (1991) New X-linked syndrome of mental retardation, gynecomastia, and obesity is linked to *DXS255*. Am J Med Genet 40:406-413.

The candidate performed the gene localisation, carried out pairwise linkage analyses and lod table preparation for the manuscript. Further reduction to the localisation is described in Paper 8 (below) and Chapter 5.

4. Mulley JC, **Gedeon AK**, Wilson S, Haan EA (1992) Use of linkage data obtained in single families: Prenatal diagnosis of a new X-linked mental retardation syndrome. Am J Med Genet 43:415-419.

The candidate performed prenatal diagnosis by linkage for a carrier of this private syndrome of XLMR (Chapter 5 - WTS).

5. **Gedeon A**, Partington M, Mulley J (1994) X-linked mental retardation with dystonic movements of the hands (PRTS): Revisited. Am J Med Genet 51:565-568.

The candidate performed the molecular and linkage analyses for refinement of the gene localisation in this family (Chapter 5 - PRTS). The paper and figures were prepared, written and submitted by the candidate.

6. 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.

The candidate performed the molecular genotyping and linkage analyses to map the gene in this family. The candidate wrote the molecular and linkage portions of the paper and was extensively involved in preparation and submission of the entire manuscript for publication.

7. **Gedeon AK**, Kozman HM, Robinson H, Pilia G, Schlessinger D, Turner G, Mulley JC (1996) Refinement of the background genetic map of Xq26-q27 and gene localisation for Börjeson-Forssman-Lehmann syndrome. Am J Med Genet 64:63-68.

The candidate performed the molecular and linkage analyses for establishment of the boundaries of the gene localisation in this family (Chapter 5 - BFLS). Prenatal diagnoses were carried out by the candidate. The manuscript was prepared by the candidate.

8. **Gedeon AK**, Haan E, Mulley JC (1996) Gene localisation for Sutherland-Haan syndrome (SHS:MIM 309470). Am J Med Genet letter to the editor 64:78-79.

The candidate performed the molecular and linkage analyses for reduction of the gene localisation in this large family (Chapter 5 - SHS). Data updating the published regional localisation was presented to provide a refined gene interval (see paper 1). The paper was written by the candidate.

9. **Gedeon AK**, Turner G, Mulley JC (1996) Gene localisation for Wilson-Turner syndrome (WTS:MIM 309585). Am J Med Genet letter to the editor 64:80-81.

The candidate performed the molecular and linkage analyses for reduction of the gene localisation in this large family (Chapter 5 - WTS). Data updating the published regional localisation (see Paper 3) was presented to provide a refined gene interval. The paper was written by the candidate.

APPENDIX III

Paper 1

G. Sutherland, A. Gedeon, E. Haan, P. Woodrooffe and J. Mulley (1988) Linkage studies with the gene for an X-linked syndrome of mental retardation, microcephaly and spastic diplegia (MRX2).
American Journal of Medical Genetics, v. 30 (1/2), pp. 493–508, May/June 1988

APPENDIX III

Paper 2

G. Turner, A. Gedeon, J. Mulley, G. Sutherland, J. Rae, K. Power and I. Arthur (1989) Börjeson-Forssman-Lehmann syndrome: Clinical manifestations and gene localization to Xq26-27.
American Journal of Medical Genetics, v. 34 (4), pp. 463–469, December 1989

APPENDIX III

Paper 3

M. Wilson, J. Mulley, A. Gedeon, H. Robinson and G. Turner (1991) New X-linked syndrome of mental retardation, gynecomastia, and obesity is linked to DXS255. *American Journal of Medical Genetics, v. 40 (4), pp. 406–413, September 1991*

APPENDIX III

Paper 4

J.C. Mulley, A.K. Gedeon, S. Wilson and E.A. Haan (1992) Use of linkage data obtained in single families: Prenatal diagnosis of a new X-linked mental retardation syndrome.
American Journal of Medical Genetics, v. 43 (1/2), pp. 415–419, April/May 1992

APPENDIX III

Paper 5

Gedeon, A., Partington, M. and Mulley, J. (1994) X-linked mental retardation with dystonic movements of the hands (PRTS): Revisited.
American Journal of Medical Genetics, v. 51 (4), pp. 565–568, July 1994

APPENDIX III

Paper 6

G. Turner, A. Gedeon and J. Mulley (1994) X-linked mental retardation with heterozygous expression and macrocephaly: Pericentromeric gene localization. *American Journal of Medical Genetics, v. 51 (4), pp. 575–580, July 1994*

APPENDIX III

Paper 7

A.K. Gedeon, H.M. Kozman, H. Robinson, G. Pilia, D. Schlessinger, G. Turner and J. C. Mulley (1996) Refinement of the background genetic map of Xq26-q27 and gene localisation for Börjeson-Forssman-Lehmann syndrome.
American Journal of Medical Genetics, v. 64 (1), pp. 63–68, July 1996

APPENDIX III

Paper 8

Agi Gedeon, John Mulley and Eric Haan (1996) Letter to the editor: Gene localisation for Sutherland-Haan syndrome (SHS:MIM 309470).
American Journal of Medical Genetics, v. 64 (1), pp. 78–79, July 1996

APPENDIX III

Paper 9

Agi Gedeon, John Mulley and Gillian Turner (1996) Letter to the editor: Gene localisation for Wilson-Turner syndrome (WTS:MIM 309585).
American Journal of Medical Genetics, v. 64 (1), pp. 80–81, July 1996

APPENDIX IV

Publications from Chapter 6

The publications listed in this appendix arose directly from the research described in Chapter 6 of this thesis. The role of the candidate is specified, and copies of the manuscripts follow in chronological sequence.

1. Adès LC, **Gedeon AK**, Wilson MJ, Latham M, Partington MW, Mulley JC, Nelson J, Lui K, Sillence DO (1993) Barth Syndrome: Clinical features and confirmation of gene localisation to distal Xq28. Am J Med Genet 45:327-334.

The candidate performed the molecular and linkage analyses for gene localisation in this family (Chapter 6 - BTHS). The candidate wrote the molecular and linkage portions of the paper and prepared Tables I and II and Figure 4.

2. **Gedeon AK**, Mulley JC, Kozman H, Donnelly A, Partington MW (1994) Localisation of the gene for X-linked reticulate pigmentary disorder with systemic manifestations (PDR), previously known as X-linked cutaneous amyloidosis. Am J Med Genet 52:75-78.

The candidate performed the gene localisation in this family (Chapter 6 - PDR). The paper was prepared, written and submitted by the candidate.

3. **Gedeon AK**, Wilson M, Colley AC, Sillence DO, Mulley JC (1995) X-linked fatal infantile cardiomyopathy maps to Xq28 and is possibly allelic to Barth syndrome. J Med Genet 32:383-388.

The candidate performed the molecular and linkage analyses for gene localisation in this family (Chapter 6). Apart from the clinical descriptions, the paper was written and submitted by the candidate.

4. Bione S, D'Adamo P, Maestrini E, **Gedeon AK,** Bolhuis PA, Toniolo D (1996) A novel X-linked gene, G4.5, is responsible for Barth syndrome. Nature Genetics 12:385-389.

The candidate entered into collaboration with Dr D Toniolo and provided DNA samples from families with Barth syndrome. The candidate gathered stored cell lines and organised establishment of new cell lines from affected individuals.

APPENDIX IV

Paper 1

L.C. Adès, A.K. Gedeon, M.J. Wilson, M. Latham, M.W. Partington, J.C. Mulley, J. Nelson, K. Lui and D.O. Sillence (1993) Barth syndrome: Clinical features and confirmation of gene localisation to distal Xq28.
American Journal of Medical Genetics, v. 45 (3), pp. 327–334, February 1993

APPENDIX IV

Paper 2

A.K. Gedeon, J.C. Mulley, H. Kozman, A. Donnelly and M.W. Partington (1994) Localisation of the gene for X-linked reticulate pigmentary disorder with systemic manifestations (PDR), previously known as X-linked cutaneous amyloidosis. *American Journal of Medical Genetics, v. 52 (1), pp. 75–78, August 1994*

APPENDIX IV

Paper 3

A.K. Gedeon, M.J. Wilson, A.C. Colley, D.O. Sillence, J.C. Mulley (1995) X linked fatal infantile cardiomyopathy maps to Xq28 and is possibly allelic to Barth syndrome.
Journal of Medical Genetics, v. 32 (5), pp. 383-388, May 1995

APPENDIX IV

Paper 4

Silvia Bione, Patrizia D'Adamo, Elena Maestrini, Agi K. Gedeon, Pieter A. Bolhuis and Daniela Toniolo (1996) A novel X-linked gene, G4.5. is responsible for Barth syndrome.
Nature Genetics, v. 12 (4), pp. 385-389, April 1996

APPENDIX V

Publications from Chapter 7

This appendix contains papers describing early experience with the direct molecular diagnosis of FRAXA. Publications describing the deletion of FMR1 associated with the fragile X(A) syndrome and deletions of an anonymous sequence (pS8) in Xq28 associated with non-specific delay arose directly from this research and development and are described in chapter 7 of this thesis. The role of the candidate is specified, and copies of the manuscripts follow.

1. Sutherland GR, **Gedeon AK**, Kornman L, Donnelly A, Byard RW, Mulley JC, Kremer E, Lynch M, Pritchard M, Yu S, Richards RI (1991) Prenatal diagnosis of Fragile X syndrome by direct detection of the unstable DNA sequence. New Engl J Med 325:1720-1722.

The candidate performed the molecular diagnosis for the fragile X syndrome (FRAXA) in this case.

2. Yu S, Mulley J, Loesch D, Turner G, Donnelly A, **Gedeon AK**, Hillen D, Kremer E, Lynch M, Pritchard M, Sutherland GR, Richards RI (1992) Fragile X syndrome: Unique genetics of the heritable unstable element. Am J Hum Genet 50:968-980.

The candidate performed molecular analysis of families segregating the fragile X syndrome as part of a team effort to assess the correlation of phenotype to genotype.

3. Mulley JC, Yu S, **Gedeon AK**, Donnelly A, Turner G, Loesch D, Chapman CJ, Gardner RJM, Richards RI, Sutherland GR (1992) Experience with direct molecular diagnosis of fragile X. J Med Genet 29:368-374.

The candidate performed molecular diagnosis in families segregating the fragile X syndrome.

4. **Gedeon AK**, Baker E, Robinson H, Partington MW, Gross B, Manca A, Korn B, Poustka A, Yu S, Sutherland GR, Mulley JC (1992) Fragile X syndrome without CCG amplification has an *FMR1* deletion. Nature Genetics 1:341-344.

The candidate performed the molecular diagnosis of fragile X syndrome and determined the molecular extent of the deletion in this case (Chapter 7). Southern filters were prepared and supplied to the collaborators by the candidate to determine more accurately the distal extent of the deletion. The manuscript was prepared by the candidate.

5. Loesch DZ, Huggins R, Hay DA, **Gedeon AK**, Mulley JC, Sutherland GR (1993) Genotype-phenotype relationships in Fragile X syndrome: A family study. Am J Hum Genet 53:1064-1073.

The candidate performed the molecular diagnosis of fragile X syndrome in these families by determination of the size of amplification and methylation status.

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

The candidate was responsible for the molecular diagnosis of FRAXA and FRAXE in the families reported in this manuscript, to establish whether any genotype-phenotype relationship in FRAXE was due to biased ascertainment.

7. **Gedeon ÁK**, Keinänen M, Adès LC, Kääriäinen H, Gécz J, Baker E, Sutherland GR, Mulley JC (1995) Overlapping submicroscopic deletions in Xq28 in two unrelated boys with developmental disorders: Identification of a gene near FRAXE. Am J Hum Genet 56:907-914.

The deletion in the Australian boy was detected by the candidate, while deletion in the Finnish boy was purported elsewhere and referred for investigation. Characterisation of the molecular extent of both deletions (Chapter 7), was completed by the candidate. The paper was written and submitted by the candidate.

8. Gécz J, **Gedeon ÁK**, Sutherland GR, Mulley JC (1996) Identification of the gene FMR2, associated with FRAXE mental retardation. Nature Genetics 13:105-108.

The candidate commenced the screening of a human fetal brain cDNA library with the probes VK21A and pS8 (DXS296).

APPENDIX V

Paper 1

BRIEF REPORT

PRENATAL DIAGNOSIS OF FRAGILE X SYNDROME BY DIRECT DETECTION OF THE UNSTABLE DNA SEQUENCE

GRANT R. SUTHERLAND, PH.D., D.SC., AGI GEDEON, B.SC., LOUISE KORNMAN, M.B., B.S., ANDREW DONNELLY, B.SC., ROGER W. BYARD, M.B., B.S., JOHN C. MULLEY, PH.D., ERIC KREMER, PH.D., MICHAEL LYNCH, PH.D., MELANIE PRITCHARD, PH.D., SUI YU, M.B., M.M., AND ROBERT I. RICHARDS, PH.D.

FRAGILE X syndrome is the most common form of familial mental retardation.[1] Its prenatal diagnosis has relied on cytogenetic detection of the fragile X chromosome in cultured amniotic fluid, chorionic-villus cells, or fetal blood obtained by cordocentesis. The rate of misdiagnosis is about 5 percent and is due to rare false positive and more frequent false negative diagnoses.[2] A molecular-genetic approach using DNA polymorphisms linked to the fragile site is feasible for diagnosis, but the results are probabilistic rather than absolutely diagnostic.[3,4] The fragile X syndrome has recently been shown to be characterized by an unstable DNA sequence that can be detected by Southern blot analysis.[5,6] We report the use of this approach to establish the carrier status of a cytogenetically normal woman in a family with the fragile X syndrome and to diagnose the mutation in her male fetus by detecting the unstable sequence in DNA obtained by chorionic-villus sampling.

CASE REPORT

The family with the fragile X syndrome (Fig. 1) had been known for some years to be affected, but the women in generation II had not requested studies to determine their carrier status. The index patient (Subject II-5) presented for such studies in January 1991, before starting her own family. Genotyping of the family with the highly polymorphic microsatellite marker at locus DXS297[7] (an AC repeat) showed that this woman had a 96 percent risk of carrying the fragile X genotype.[4] Cytogenetic investigation failed to demonstrate the presence of the fragile X chromosome, as it fails to do in more than half the women of normal mental status who are carriers. In March 1991 the patient presented at nine weeks' gestation. Southern blot analysis of her DNA with the pfxa3 probe, which detects the unstable DNA sequence $p(CCG)_n$, the amplification of which is characteristic of the fragile X mutation,[5] confirmed that she was a carrier of the mutation (Fig. 2). A transabdominal chorionic-villus biopsy was performed at 10 weeks' gestation, and about 30 mg of villi were obtained. A polymerase chain reaction using primers specific for the Y chromosome[8] showed that the fetus was male; this was confirmed by short-term culture of the biopsy specimen, which revealed a normal male karyotype. Southern blot analysis with probe pfxa3 showed the fetus to have a 2.3-kb allele (Fig. 2). Alleles of this size are typical of symptomatic males with the fragile X genotype. Normal males have a 1.0-kb allele, whereas asymptomatic male carriers have alleles ranging in size from 1.1 to 1.6 kb (unpublished data). The parents were informed that the fetus was male and had a greater than 99 percent chance[4] of having the fragile X genotype, on the basis of studies with linked DNA polymorphisms at loci DXS297 and DXS296, which flank the fragile site. In addition, the results obtained with pfxa3[6] were explained to the parents. They decided to terminate the pregnancy because of the results of the two DNA tests, with the awareness that in their case the result of a normal chromosome test could not exclude a diagnosis of fragile X syndrome. The pregnancy was terminated by suction evacuation of the uterus. Cultured fetal tissue and chorionic-villus cells in long-term culture expressed the fragile X chromosome.

From the Department of Cytogenetics and Molecular Genetics (G.R.S., A.G., A.D., J.C.M., E.K., M.L., M.P., S.Y., R.I.R.) and the Department of Histopathology (R.W.B.), Adelaide Children's Hospital; and the Queen Victoria Hospital (L.K.); all in Adelaide, Australia. Address reprint requests to Dr. Sutherland at the Department of Cytogenetics and Molecular Genetics, Adelaide Children's Hospital, SA 5006, Australia.

METHODS

Cytogenetic Analysis

Direct three-day culture of the chorionic-villus sample was carried out as described elsewhere.[9] Long-term cultures of this tissue and of fetal skin and muscle tissue obtained at the time of the termination of the pregnancy were established in Chang medium and transferred to Ham's F10 medium 48 hours before the induction of the fragile X chromosome by exposure to 0.05 μM fluorodeoxyuridine or 300 mg of thymidine per liter of solution as described elsewhere.[9]

Polymerase Chain Reaction

The sex of the fetus was determined with use of primers for a repeat specific to the Y chromosome (Y1.1 and Y1.2) and reaction conditions as described elsewhere[8] in a polymerase chain reaction. The reaction products were subjected to electrophoresis on 1.4 percent agarose gels and visualized by staining with ethidium bromide. The genotypes at the DXS297 locus were determined as described elsewhere.[7]

Southern Blot Analysis

DNA was extracted from blood leukocytes obtained from family members, from the chorionic-villus sample, and from the fetal tissue obtained after the termination of the pregnancy. Eight-microgram aliquots of DNA were digested to completion with the restriction enzymes *Pst*I and *Taq*I. Electrophoresis and Southern blotting were performed as described elsewhere (unpublished data). The *Taq*I blot was probed with VK21A[4] to determine polymorphisms at the DXS296 locus. The *Pst*I blot was simultaneously probed with pfxa3[5] and pS8. The pfxa3 probe (Oncor, Gaithersburg, Md.) detects a constant 1.0-kb band in all noncarriers; this probe is a 520-bp fragment released by *Pst*I digestion from a larger genomic fragment cloned into pUC18. The pS8 control probe detects a constant 800-bp band in all subjects that does not overlap with the size distribution of the *Pst*I fragments detectable with pfxa3. The pS8 probe is an 800-bp *Pst*I fragment from a subclone of YAC XY539 in the region of locus DXS296. These inserts were labeled by random priming.

Methylation of the DNA in the region of the unstable sequence has been shown to correlate with the presence of mental handicap in males with the fragile X genotype.[10] Methylation was assessed in the fetal tissue obtained after the termination of the pregnancy. The products of conception were separated, and samples were collected for histologic confirmation of tissue type and DNA extraction. DNA

Reprinted from the *New England Journal of Medicine*
325:1720-1722 (December 12), 1991

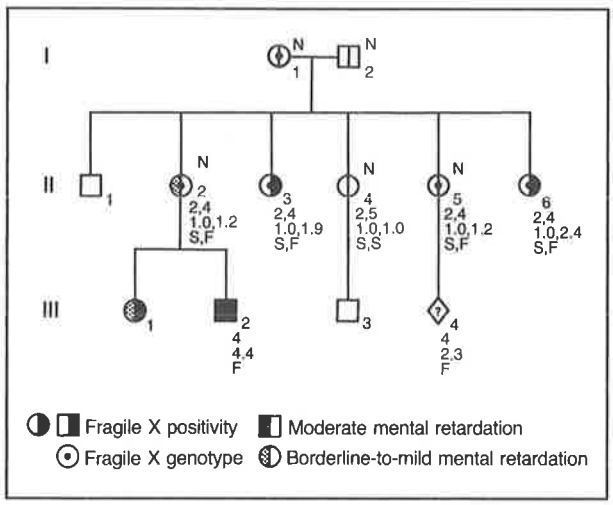


Figure 1. Pedigree of the Family with the Fragile X Syndrome.

Data shown below the symbols representing the subjects studied are as follows: top row, subject number; second row, genotypes at the DXS297 locus (numbered arbitrarily in order of increasing size); third row, band sizes (in kilobases) detected with the pfxa3 probe; bottom row, genotypes at the DXS296 locus, with S (slow) denoting the 10.9-kb allele, and F (fast) the 9.9-kb allele. N denotes cytogenetically negative for the fragile X chromosome. Circles represent females, squares males, and the diamond shape the fetus studied. Open symbols represent subjects not studied.

was extracted from various tissues, including the chorionic villi; digested with the restriction endonucleases *Eco*RI and *Sac*II; and subjected to Southern blot analysis with probe pfxa3. In normal males, fully methylated DNA yields a 5.0-kb band, and fully unmethylated DNA a 2.8-kb band. In males with the fragile X syndrome, the size of each of these bands is increased by the length of the $p(CCG)_n$ amplification.

Results

Cytogenetic study of the chorionic-villus sample did not show the fragile X genotype in 100 metaphases after induction with thymidine, and it was present in only 1 metaphase of another 100 examined after induction with fluorodeoxyuridine. Culture of fetal skin and muscle tissue detected 6 cells expressing the fragile X genotype among 260 cells examined after induction with fluorodeoxyuridine, and 1 cell among 40 cells after thymidine induction. Without other information, the results of the cytogenetic analysis of the chorionic villi would have been regarded as inconclusive. More metaphases would have been examined, but it is unlikely that a firm diagnosis would have been made. The fetus had been shown to be male by polymerase chain reaction, which detected a male amplification pattern; by direct chromosomal examination of short-term cultures of chorionic-villus samples, which are not suitable to study for the fragile X genotype; and by probing of the *Pst*I blot with pfxa3, which showed only an affected allele.

Southern blot analysis using the pfxa3 probe (Fig. 2) showed that Subject II-5 had the pattern typical of a carrier, with her normal X chromosome giving a 1.0-kb band and her fragile X chromosome giving a 1.2-kb band. The fetal DNA from the chorionic-villus samples (and the DNA obtained after the termination of the pregnancy from villi and skin and muscle tissue) showed that the fetus lacked the normal 1-kb band, which was replaced by a band at 2.3 kb — one of the patterns typically seen in males with the fragile X syndrome. Polymorphisms flanking the fragile X locus (DXS297 proximally and DXS296 distally) were informative, indicating with virtual certainty that the fetus had the fragile X genotype.

Analysis of the DNA methylation patterns (Fig. 3) showed that the *Sac*II restriction site was methylated to varying degrees in all the tissues studied, except for the samples of villi, which were unmethylated.

Discussion

Prenatal diagnosis of the fragile X syndrome has depended largely on cytogenetic demonstration of the presence of the fragile X chromosome. Although it is reasonably reliable, cytogenetic analysis of the fragile

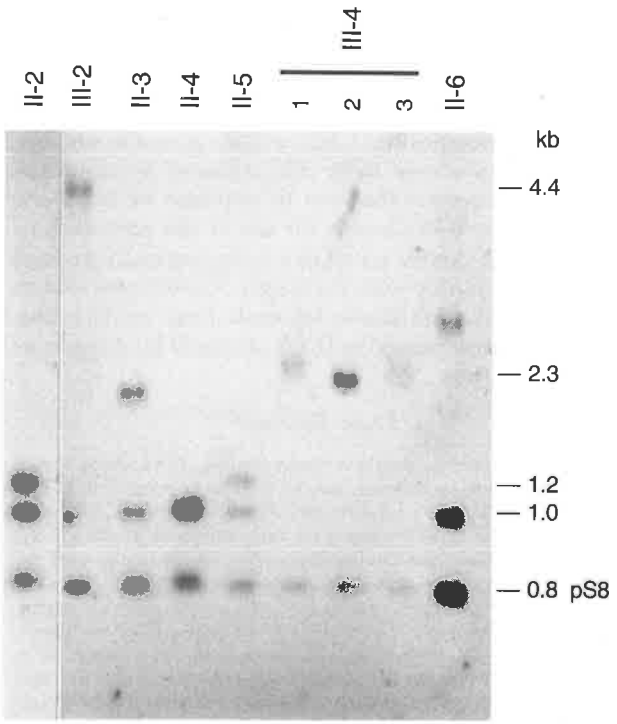


Figure 2. Southern Blotting, after Digestion by *Pst*I, of Leukocyte DNA from the Family Members, the Chorionic-Villus Sample, and the Tissues Obtained after the Termination of the Pregnancy.

The blot was probed with pfxa3 and pS8 simultaneously. Lanes are marked with the numbers used to identify the members of the pedigree in Figure 1. For the fetus (III-4), three samples were studied: the chorionic-villus-biopsy specimen in lane 1, specimens of skin and muscle tissue in lane 2, and villi obtained after the termination of pregnancy in lane 3.

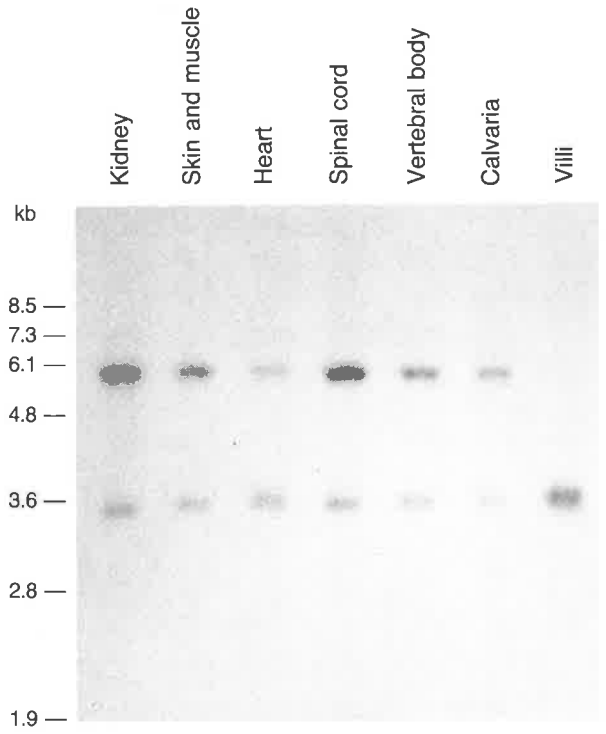


Figure 3. Southern Blot Assay Showing the Absence of Methylation in DNA from Chorionic Villi.

DNA was isolated from fetal tissues after histologic examination and subjected to restriction-endonuclease digestion with the enzymes *Eco*RI and *Sac*II. After electrophoresis and transfer to a nylon membrane, the DNA was hybridized to radiolabeled pfxa3. DNA size markers were SPP-1 bacteriophage DNA digested with *Eco*RI (Bresatec, Adelaide, Australia).

X genotype has always been difficult, and both false positive and false negative diagnoses have been reported.[2]

Until now, molecular-genetic diagnoses have relied on the use of linked DNA polymorphisms.[3,4,7] If closely linked flanking polymorphisms were used, the fetal genotype could be predicted with a very high probability. In the past, discrepancies between cytogenetic and linkage-based diagnoses have been very difficult to interpret, and consequently molecular-based diagnosis has not been widely used. If, for example, the fragile X genotype could be detected in a fetus by linkage but not by cytogenetic analysis, the discrepancy could have been due either to a failure to detect the genotype cytogenetically or to the fact that the fetus was a phenotypically normal carrier, male or female. Risk analysis based on linked DNA markers has not been used widely in prenatal diagnosis and has been largely restricted to the detection of carriers.[3]

The use of pfxa3 provides a direct assay for the fragile X genotype. A male fetus with this genotype who has a mentally normal mother has a 76 percent risk of being mentally handicapped.[11] However, since the size of the unstable sequence appears to correlate with the phenotype[10] (and unpublished data), it should be possible to modify this risk according to the pattern of hybridization of the pfxa3 probe. Data on which to base such modified risk estimates are still being assembled, but normal male carriers usually have a fragile X band that is between 1.1 and 1.6 kb in size, with larger bands and multiple bands being associated with the fragile X syndrome. The fragile X band in the prenatally diagnosed fetus was 2.3 kb in size.

It has been shown that methylation of the cytidine phosphate guanosine (CpG)–rich DNA adjacent to the unstable region is a marker of the phenotype of males with the fragile X syndrome.[10] In the fetus we studied, DNA from all the tissues except the villi samples was methylated to a varying extent. This methylation accords with the other evidence indicating that this fetus had the fragile X syndrome. The lack of methylation of DNA in villi indicates that on the basis of this case, the methylation status of the villi does not reflect that of the embryonic tissues, and it should not be used as a diagnostic measure. Others[10] have shown that the unstable sequence that characterizes the fragile X genotype can be detected in DNA extracted from chorionic villi and that the DNA methylation patterns in this tissue are not the same as those in other tissues.

The pfxa3 probe is a direct diagnostic reagent for the fragile X genotype. Because this probe can make the primary diagnosis of the fragile X syndrome and can identify carriers within families bearing the genotype, it offers a major advance in the prenatal diagnosis of this common form of mental handicap.

We are indebted to Maria Digenis and Rosalie Smith for assistance with the cytogenetics analysis, to Dr. E.A. Haan for helpful comments on the manuscript, and to Dr. D. Schlessinger for isolating YAC XY539, from which the pS8 probe was obtained.

References

1. Sutherland GR, Hecht F. Fragile sites on human chromosomes. New York: Oxford University Press, 1985.
2. Jenkins EC, Krawczun MS, Stark-Houck SL, et al. Improved prenatal detection of fra(X)(q27.3): methods for prevention of false negatives in chorionic villus and amniotic fluid cell cultures. Am J Med Genet 1991;38:447-52.
3. Sutherland GR, Mulley JC. Diagnostic molecular genetics of the fragile X. Clin Genet 1990;37:2-11.
4. Suthers GK, Mulley JC, Voelckel MA, et al. Genetic mapping of new DNA probes at Xq27 defines a strategy for DNA studies in the fragile X syndrome. Am J Hum Genet 1991;48:460-7.
5. Yu S, Pritchard M, Kremer E, et al. Fragile X genotype characterized by an unstable region of DNA. Science 1991;252:1179-81.
6. Kremer EJ, Pritchard M, Lynch M, et al. Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)*n*. Science 1991;252: 1711-4.
7. Richards RI, Shen Y, Holman K, et al. Fragile X syndrome: diagnosis using highly polymorphic microsatellite markers. Am J Hum Genet 1991;48: 1051-7.
8. Kogan SC, Doherty M, Gitschier J. An improved method for prenatal diagnosis of genetic diseases by analysis of amplified DNA sequences: application to hemophilia A. N Engl J Med 1987;317:985-90.
9. Purvis-Smith SG, Laing S, Sutherland GR, Baker E. Prenatal diagnosis of the fragile X — the Australasian experience. Am J Med Genet 1988;30:337-45.
10. Oberlé I, Rousseau F, Heitz D, et al. Instability of a 550–base pair DNA segment and abnormal methylation in fragile X syndrome. Science 1991; 252:1097-102.
11. Sherman SL, Morton NE, Jacobs PA, Turner G. The marker (X) syndrome: a cytogenetic and genetic analysis. Ann Hum Genet 1984;48:21-37.

THE NEW ENGLAND JOURNAL OF MEDICINE is published weekly in the English language by the Massachusetts Medical Society (Waltham, MA, USA). Editorial office: 10 Shattuck Street, Boston, MA 02115, USA. **SUBSCRIPTIONS:** Subscription offices: 1440 Main Street, Waltham, MA 02154-1649, USA; 1800 Ironstone Manor, Pickering, Ontario, L1W 3J9, Canada; and Saxon Way, Melbourn, Royston, Herts SG8 6NJ, UK. Subscription prices: USA: $89 per year (interns, residents $53 per year; students $45 per year). Outside USA: Canadian dollars drawn on a Canadian bank: C$128 per year (interns, residents C$92 per year; students C$77 per year). Pounds sterling drawn on a UK bank: £80 per year (interns, residents, and students £52 per year). US dollar payments accepted at current rate of exchange pro rata to the pound. Payments and correspondence for all subscriptions delivered to Japan should be sent directly to Nankodo Co., Ltd., 42-6, Hongo 3-chome, Bunkyo-ku, Tokyo 113, Japan. Rates are subject to change without notice. Sample copies available on request.

APPENDIX V

Paper 2

Yu, S. et al. (1992) Fragile-X Syndrome - Unique Genetics of the Heritable Unstable Element.
American Journal of Human Genetics, v. 50 (5), pp. 968-980, May 1992

APPENDIX V

Paper 3

Mulley, J.C., Yu, S., Gedeon, A., Donnelly, A., Turner, G., Loesch, D., Chapman, C.J., Gardner, R.J.M., Richards, R.I. and Sutherland, G.R. (1992) Experience with Direct Molecular Diagnosis of Fragile-X.
Journal of Medical Genetics, v. 29 (6), pp. 368-374, June 1992

APPENDIX V

Paper 4

A.K. Gedeon, E. Baker, H. Robinson, M.W. Partington, B. Gross, A. Manca, B. Korn, A. Poustka, S. Yu, G.R. Sutherland and J.C. Mulley (1992) Fragile X syndrome without CCG amplification has an FMR1 deletion.
Nature Genetics, v. 1 (5), pp. 34 -344, August 1992

APPENDIX V

Paper 5

Loesch, D.Z., Huggins, R., Hay, D.A., Gedeon, A.K., Mulley, J.C. and Sutherland, G.R. (1993) Genotype-Phenotype Relationships in Fragile-X Syndrome - a Family Study.
American Journal of Human Genetics, v. 53 (5), pp. 1064-1073, November 1993

APPENDIX V

Paper 6

J.C. Mulley, S. Yu, D.Z. Loesch, D.A. Hay, A. Donnelly, A.K. Gedeon, P. Carbonell, I. López, G. Glover, I. Gabarrón (1995) FRAXE and mental retardation.
Journal of Medical Genetics, v. 32 (3), pp. 162-169, March 1995

APPENDIX V

Paper 7

Gedeon, A.K., Keinanen, M., Ades, L.C., Kaariainen, H., Gecz, J., Baker, E., Sutherland, G.R. and Mulley, J.C. (1995) Overlapping Submicroscopic Deletions in Xq28 in 2 Unrelated Boys with Developmental Disorders - Identification of a Gene Near Fraxe.
American Journal of Human Genetics, v. 56 (4), pp. 907-914, April 1995

APPENDIX V

Paper 8

J. Gecz, A.K. Gedeon, G.R. Sutherland and J.C. Mulley (1996) Identification of the gene FMR2, associated with FRAXE mental retardation.
Nature Genetics, v. 13 (1), pp. 105-108, May 1996