UNIVERSITY OF ADELAIDE LIBRARY
20.10.03

Identification of Novel Genes for X-linked Mental Retardation

A thesis submitted for the degree of Doctor of Philosophy to the University of Adelaide

by

Marie Mangelsdorf

BSc (Hons)

School of Medicine
Department of Paediatrics, Women's and Children's Hospital

May 2003

Corrections

The following references should be referred to in the text as:

Page 2, line 2: (Birch *et al.*, 1970)
Page 2, line 2: (Moser *et al.*, 1983)
Page 3, line 15: (Martin and Bell, 1943)
Page 3, line 4 and line 9: (Stevenson *et al.*, 2000)
Page 17, line 5: (Monaco *et al.*, 1986)

And in the reference list as:

Birch H. G., Richardson S. A., Baird D., Horobin, G. and Ilsley, R. (1970) Mental Subnormality in the Community: A Clinical and Epidemiological Study. Williams and Wilkins, Baltimore.

Martin J. P. and Bell J. (1943). A pedigree of mental defect showing sex-linkage. *J. Neurol. Psychiatry* **6:** 154.

Monaco A.P., Nerve R.L., Colletti-Feener C., Bertelson C.J., Kurnit D.M. and Kunkel L.M. (1986) Isolation of candidate cDNAs for portions of the Duchenne muscular dystrophy gene. *Nature* **323:** 646-650.

Moser H.W., Ramey C.T. and Leonard C.O. (1983) In Principles and Practice of Medical Genetics (Emery A.E.H. and Rimoin D.L., Eds). Churchill Livingstone, Edinburgh UK

Penrose L. (1938) A clinical and genetic study of 1280 cases of mental defect. (The Colchester survey). Medical Research Council, London, UK.

Stevenson R.E., Schwartz C.E. and Schroer R.J. (2000) X-linked Mental Retardation. Oxford University Press.

Page 7 line 1: 'better known an better studied' should read 'better known and better studied'.
Page 13 line 8: 'in vitro' should be '*in vitro*'.
Page 16 line 2: '(*eg.*)' should read '(*eg.* William's syndrome)'
Page 22 line 20: *causing* should not be italicised.
Page 24 line 2: should read 'The human genome sequence is expected to be fully completed soon, with the maps of individual chromosomes being published by the end of 2003'.
Page 34, line 4: 'was' should read 'were'.
Page 58, line 6: (company) should read (Multi Cel (one L), Thermo Trace Ltd).
Page 77 legend Figure 3.8: + and – refer to RT-PCR with and without the addition of reverse transcriptase respectively.
Page 84 line 15: 'these results' should read 'the spanning signals for these clones'
Page 100 line 16: The ORF of *PRRG1* was screened by direct sequencing.
Page 139 line 17: 'Panay and island of the Phillipines' should read 'Panay an island of the Phillipines'.
Page 153 line 5: 'lysine at position (K_{50} paired-type) at position 50' should read 'lysine (K_{50} paired-type) at position 50'.
Page 160 line 1: 'evalutaion' should be 'evaluation'.

"Almost all aspects of life are engineered at the molecular level, and without understanding molecules we can only have a very sketchy understanding of life itself."

FRANCIS CRICK (1988)

Statement and Declaration

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. I give my consent to this copy of my thesis, when deposited in the University library, being available for loan and photocopying

Date 22/05/2003

CONTENTS

Statement and Declaration		ii
Contents		iii
List of Abbreviation		iv
List of Gene Names and Abbreviations		vi
Summary		viii
Acknowledgements		xii
CHAPTER 1	Introduction	1
CHAPTER 2	Materials and Methods	29
CHAPTER 3	Molecular characterisation of X chromosome rearrangements involving and Xq28 breakpoint in three patients with mental retardation	66
CHAPTER 4	Fine mapping and candidate gene screening in a family with non-syndromic X-linked mental retardation mapping to either Xq13.1 or Xq23	112
CHAPTER 5	Identification of mutations in the homeobox gene *ARX* in patients with syndromic and non-syndromic XLMR	141
CHAPTER 6	Conclusions	182
References		186
Appendices		

LIST OF ABBREVIATIONS

A,C,G,T	-	nucleotides: adenine, cytosine, guanine, thymine
BAC	-	bacterial artificial chromosome
BLAST	-	basic local alignment search tool
bp	-	base pairs
cDNA	-	complementary DNA
cM	-	centimorgans
CNS	-	central nervous system
cos	-	cosmid
dNTP	-	deoxynucleoside triphosphate
ddNTP	-	dideoxynucleoside triphosphate
DNA	-	deoxyribonucleic acid
dup	-	duplicaion
EST	-	expressed sequence tag
FISH	-	fluorescence *in situ* hybridisation
FRAXA	-	rare, folic acid type, fragile site, fra(X)(q27.3)
FRAXE	-	rare, folic acid type, fragile site, fra(X)(q28)
gDNA	-	genomic DNA
GDP	-	guanine diphosphate
GLA	-	γ-carboxy glutamic acid
GTP	-	guanosine triphosphate
IL-1	-	interleukin 1
Ins	-	insertion
inv	-	inversion
IQ	-	intelligence quotient
kb	-	kilo base pairs
LCR	-	locus control region
lod	-	log of the odds
LR-PCR	-	long range PCR
LRR	-	leucine-rich repeat
Mb	-	megabase
MR	-	mental retardation
mRNA	-	messenger RNA
MRX	-	Linkage mapped non-syndromic X-linked mental retardation
MRXS	-	Linkage mapped X-linked mental retardation syndrome
NCBI	-	National Centre for Biotechnology Information
NSXLMR	-	non-syndromic X-linked mental retardation
OD_{260}	-	optical density at a wavelength of 260 nm
OD_{600}	-	optical density at a wavelength of 600 nm
OMIM	-	online Mendelian Inheritance in Man
ORF	-	open reading frame
PAC	-	P1 artificial chromsome
PCR	-	polymerase chain reaction
rec	-	recombinant
RNA	-	ribonucleic acid
RT-PCR	-	reverse transcribe PCR

SLRP	-	small leucine rich protein
SNP	-	single nucleotide polymorphism
SSCA	-	single strand conformation analysis
STS	-	sequence tagged site
UTR	-	untranslated region
v/v	-	volume per volume
w/v	-	weight per volume
WCH	-	Women's and Children's Hospital
XLMR	-	X-linked mental retardation
YAC	-	yeast artificial chromosome

LIST OF GENE AND PROTEIN NAME ABBREVIATIONS

ABCD1	-	ATP-binding cassette, sub-family D (ALD), member 1
AGTR2	-	angiotensin II receptor, type 2
ARHGEF6	-	Rac/Cdc42 guanine nucleotide exchange factor (GEF) 6
ARX	-	*aristaless* related homeobox
ATP2B3	-	plasma membrane, calcium transporting, ATPase 3
ATRX	-	alpha thalassemia/mental retardation syndrome
BGN	-	biglycan
CBP	-	CREB binding protein
Cdc42	-	cell division cycle 42 (GTP binding protein)
CNG2	-	cyclic nucleotide gated channel 2
CYFIP1	-	cytoplasmic FMR1 interacting protein 1
DNCH1	-	dynein, cytoplasmic heavy polypeptide 1
DUSP9	-	dual specificity phosphatase 9
ELK1	-	ELK1, member of ETS oncogene family
ERCC6	-	excision repair cross-complementing rodent repair deficiency, complementation group 6
ESD	-	esterase D
F2	-	coagulation factor II (thrombin)
F7	-	coagulation factor VII (serum prothrombin conversion accelerator)
F9	-	coagulation factor IX (plasma thromboplastic component, haemophila B)
FACL4	-	fatty-acid-Coenzyme A ligase, long-chain 4
FATE	-	fetal and adult testis expressed transcript
FGD1	-	faciogenital dysplasia (Aarskog-Scott syndrome)
FMR1	-	fragile mental retardation 1 gene, associated with the fragile site *FRAXA*
FMR2	-	fragile mental retardation 2 gene, associated with the fragile site *FRAXE*
FMRP	-	protein product of the *FMR1* gene
GABRA3	-	gamma-aminobutyric acid (GABA) A receptor, alpha 3 receptor
GABRE	-	gamma-aminobutyric acid (GABA) A receptor, epsilon
GABRQ	-	gamma-aminobutyric acid (GABA) receptor, theta
GAP	-	GTP-ase activating protein
GAS6	-	growth arrest-specific 6
GDI1	-	GDP dissociation inhibitor 1
GEF	-	guanine nucleotide exchange factor
GPR50	-	G protein-coupled receptor 50
GRIA3	-	glutamate receptor, ionotrophic, AMPA 3
IL1RAPL1	-	interleukin 1 receptor accessory protein-like 1
KLF8	-	kruppel-like factor 8
MECP2	-	methyl CpG binding protein 2 (Rett syndrome)
NSDHL	-	NAD(P) dependent steroid dehydrogenase-like
NXF5	-	nuclear RNA export factor 5
OPHN1	-	oligophrenin 1
PAK3	-	p21 (CDKN1A)-activated kinase 3
PAX6	-	paired box gene 6 (aniridia, keratitis)

PIN4/hPAR14 - protein (peptidyl-prolyl cis/trans isomerase) NIMA-interacting
PRRG1 - proline-rich Gla (G-carboxyglutamic acid) polypeptide 1
PRRG2 - proline-rich Gla (G-carboxyglutamic acid) polypeptide 2
PRSS12 - serine protease12 (neurotrypsin)
RAC1 - ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein)
RLGP - Ras-like GTPase
RPS6KA3 - ribosomal protein S6 kinase, 90kDa, polypeptide 3
RPS6KA6 - ribosomal protein S6 kinase, 90kDa, polypeptide 6
SHH - sonic hedgehog homolog (Drosophila)
SLC6A8 - solute carrier family 6 (creatine neurotransmitter transporter),
TM4SF2 - transmembrane 4 superfamily member 2
TMG3 - transmembrane gamma-carboxyglutamic acid gene 3,
TMG4 - transmembrane gamma-carboxyglutamic acid gene4
TREX2 - three prime repair exonuclease 2
VCX-A - variable charge protein on X with eight repeats
ZNF185 - zinc finger gene 185 (LIM domain)
ZNF261 - zinc finger gene 261
ZNF275 - zinc finger gene 275
ZXDA/ZXDB - duplicated zinc finger, X-linked, A and B

SUMMARY

Mental retardation (MR) is estimated to affect 1-2% of the population, and is due to both environment and improperly functioning genes. A higher incidence of MR in males suggests a significant proportion of MR with a genetic cause is due to mutations in genes on the X-chromosome. The aim of this thesis is to identify novel genes involved in X-linked mental retardation (XLMR). Understanding the genetic causes of MR will result in better diagnosis, and as a result, will lead to improved patient management and counselling of family members. It will also contribute to the understanding of the genes and mechanisms required for normal cognitive function.

The first part of the project involved molecular characterisation of the breakpoints of three X chromosome rearrangements in three unrelated patients with MR, to identify candidate genes for familial XLMR.

For Patient 1 an inversion breakpoint was found to lie within the 3' untranslated region (3'UTR) of the biglycan gene (*BGN*), such that *BGN* in this patient obtained a new 3'UTR and polyadenylation signal from Xq13.1. The open reading frame of *BGN* remained intact, and apparently normal levels of mRNA transcribed. 3'UTRs have been shown to contain elements, important for mRNA localization, stability and translation efficiency. Therefore, in this patient disruption of elements in the 3'UTR may affect the levels of BGN protein produced and thereby cause MR. Characterisation of this inversion has identified *BGN* as a candidate gene for XLMR.

For Patient 2, who also has an inversion of the X chromosome, all BAC clones derived from Xq28 that were used as probes for FISH gave spanning signal. This suggested that this patient

had a previously undetected duplication of Xq28 as well as the initially detected inversion. This duplication covers at least 2.7 Mb of Xq28, an extremely gene rich region. Thus the MR in this patient is likely to be the result of functional disomy of many genes from Xq28. This work suggests that some familial cases of XLMR may be due to submicroscopic duplication of many genes, a mechanism that would not be detected by current PCR based methods of gene screening.

For Patient 3 a translocation of the X chromosome associated with MR and cardiomyopathy was characterised. In this patient, the normal X chromosome is preferentially inactivated suggesting that disruption of a gene on the X chromosome may be responsible for the phenotype. The break at Xq28 was located within a 75 kb BAC clone. Although not physically disrupted by the translocation breakpoint one gene, *TMG3a*, whose expression was affected by the breakpoint was identified. This gene will therefore also be a candidate gene for XLMR as well as for X-linked cardiomyopathy.

The second part of the thesis involved positional candidate gene screening in a family that was localised to two regions of the X chromosome by linkage analysis. Additional family members for a linkage study were gathered and analysed and Xq13 became the most likely localisation. This region was gene rich and contained many good candidate genes for XLMR. A mutation screen of these genes has begun; as yet no disease causing changes have been identified, and this project remains ongoing.

The final part of this project has been the identification of the homeobox gene, *ARX*, and its role in both syndromic and non-syndromic XLMR. Initially, candidate gene screening within the minimal linkage interval for X-linked infantile spasms syndrome (ISSX) was undertaken,

in order to identify the gene responsible for this disorder, which is characterised by infantile spasms, hypsarrhythmia and severe to profound MR. Expansions of two polyalanine tracts in the *ARX* gene were detected in three families with ISSX, as well as a truncating mutation in a more severe case of ISSX. Subsequent screening of other XLMR families has shown that mutations in *ARX* account for a significant proportion of NSXLMR as well as other syndromic cases of XLMR. Mutations in ARX account for approximately 8% of MR where there is a clear X-linked inheritance and is now the most significant cause of MR in mapped MRX families.

Polyalanine tract expansions have been detected in other genes. In the case of *PABP2,* mutations in which cause oculopharyngeal muscular dystrophy, the polyalanine expansion results in the formation of protein aggregates in cells of patients. Transfection of HeLa cells with ARX fused in frame with green fluorescent protein (GFP) has shown that ARX with an expanded polyalanine tract does not form aggregates. These were preliminary studies and further confirmation of this result is required before conclusions can be drawn about the mechanism by which these ARX polyalanine expansion cause MR.

This thesis has therefore addressed the identification of candidate genes for XLMR using two positional cloning approaches. Firstly positional cloning of X chromosome rearrangements in patients with MR, and secondly by positional candidate gene screening of genes within minimal linkage intervals of mapped XLMR in large families. This work has identified two candidate genes *BGN* and *TMG3a*, which can now be screened in other XLMR cases. It has also resulted in a refined linkage interval in a large NSXLMR family, in which screening of candidate genes has begun. Finally, mutations in the gene *ARX* have been identified as a significant contributor to the aetiology of XLMR.

ACKNOWLEDGMENTS

I would like to acknowledge the Department of Cytogenetics and Molecular Genetics (WCH) for their financial support and for use of equipment throughout my PhD. I would especially like to thank Dr Jozef Gécz and Associate Professor John Mulley for their support, encouragement and advice. I would like to thank Professor Grant Sutherland and Dr Sui Yu for allowing me to spend this time in their department.

I would also like to thank those people that have made contributions to the work presented in this thesis. In particular Erica Woollatt for her FISH expertise; Profs Gillian Turner and Michael Partington for their many years of gathering clinical information, and collecting samples for laboratory use in the study of X-linked mental retardation; the families who participate in the research; Cathy Derwas and Sarah McDonnell for cell establishment and maintenance of cell lines.

I would also like to thank my colleagues in the Department of Cytogenetics and Molecular Genetics who have made PhD years enjoyable: Dr Kathie Gécz, Marie Shaw, Karen Lower, Merran Finnis, Lynne Hobson, Joanna Crawford, Helen Eyre and Rosalie Smith.

I gratefully acknowledge the financial support throughout these studies provided by the University of Adelaide Benjamin Poulton Medical Research Scholarship, and The Faculty of Health Science for The Travelling Fellowship for Health Science Research.

Finally I would like to thank my family for their support throughout my studies.

Chapter 1

Introduction

1.1. INTRODUCTION **2**

1.2. GENETIC CAUSES OF MR **2**

1.3. X-LINKED MENTAL RETARDATION **3**

1.3.1 Syndromic XLMR 4
1.3.2 Non Syndromic XLMR 7
1.3.3 Genetic Heterogeneity of NSXLMR 10

1.4. FUNCTIONS OF NSXLMR GENES **12**

1.4.1 Intracellular signalling pathways 12
1.4.2 Chromatin Remodelling 14

1.5. AUTOSOMAL MENTAL RETARDATION **16**

1.6. IDENTIFICATION OF DISEASE GENES **17**

1.6.1 Positional cloning of chromosome rearrangement breakpoints 17
1.6.2 Linkage analysis of XLMR families 18
1.6.3 Identification of candidate genes 19
1.6.3 The 'tour de force' approach to disease gene identification 20
1.6.4 Mutation detection in genes 21

1.7. HUMAN GENOME PROJECT **23**

1.8. BENEFITS OF IDENTIFYING MENTAL RETARDATION GENES **25**

1.9. AIMS **26**

1.1. INTRODUCTION

Mental retardation (MR) or intellectual disability is a large clinically and genetically diverse group which affects 1-2% of the population (Birch HG, 1970; Moser HW, 1983). MR can be a component of a more complex syndrome (e.g. Down syndrome, fragile X syndrome), metabolic disorder (e.g. adrenoleukodystrophy) or neuromuscular disorder (e.g. Duchenne muscular dystrophy) or can be the only detectable phenotype (non-syndromic MR). A person is diagnosed as mentally retarded if they have an intelligence quotient (IQ) of less than 70. This is further subclassified into profound (IQ <19), severe (IQ 20-34), mild (IQ 50-69) and borderline (IQ 70-80). It is now well accepted that both environmental factors and improperly functioning genes play a role in the aetiology of MR. Genes may act separately or in combination with other genes, or in combination with the environment, of which head trauma is a significant component. MR encompassing all causes is a common disorder, and patient management can be costly both financially and emotionally for family members and carers. Finding genes involved in MR for a subset of cases will assist in patient diagnosis, management of family planning through carrier testing and prenatal diagnosis as well as contributing to the understanding of pathways that are required for normal brain development.

1.2. GENETIC CAUSES OF MR

There are many different factors that are known to cause MR. Environmental (or non-genetic) causes include brain injury during prenatal and postnatal periods associated with, for example, premature birth and viral infections. Genetic causes may be heritable or spontaneous either through gene mutation or cytogenetic abnormality. These disorders may be primarily monogenic, polygenic or multifactorial. Many of the spontaneous causes are due to

chromosomal aberrations with trisomy 21 being by far the most common. For the sporadic cases, where the cause cannot be determined it is often difficult to know if MR is due to a genetic or an environmental cause.

There are many chromosomal aberrations associated with MR (Stevenson, 2000). These include aberrations in chromosome number, for example Kleinfelter syndrome (47,XXY), trisomy such as with trisomy of chromosome 21 causing Down's syndrome, deletion as in cri du chat syndrome due to deletion of 5p and uniparental disomy or deletion associated with Prader-Willi and Angelman syndromes. For both mild and severe MR in approximately half of the cases a cause cannot be determined (Stevenson, 2000).

1.3. X-LINKED MENTAL RETARDATION

It has long been recognised that there is an excess of males over females with MR with an approximately 30% higher incidence of MR in males (Penrose, 1938). This has been taken to suggest that a significant number of mutations causing MR are in a gene or genes residing on the X chromosome. The correctly functioning gene product from their other X chromosome would compensate for these same mutations in females. In 1943, the first pedigree with MR that was segregating in an X-linked manner was described (Martin, 1943). Since this time a large number of families with XLMR syndromes and non-syndromic XLMR (NSXLMR) have been described. The incidence of MR showing X-linked inheritance has been estimated at 1.66/1000 males (Glass, 1991; Turner, 1996)

A search of the Online Mendelian Inheritance in Man database (http://www3.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM) reveals 1,074 entries containing

the term mental retardation. Of these 170 have been linked to the X chromosome and 412 to an autosome. There are 486 that have not been assigned to any particular chromosome.

XLMR can be divided into syndromic (MRXS), where the phenotype is associated with other dysmorphic, neuromuscular or metabolic features, or non-syndromic (NSXLMR) where MR is the sole phenotype, and where affected members of the same family do not display any other consistent features which would allow classification as an existing syndrome, or delineation of a new syndrome. MRXS nomenclature was proposed as a temporary label until the syndrome could be named after peer review and publication. MRX nomenclature was introduced to label that subset of NSXLMR that could be mapped by linkage to a specific region of the X chromosome (Mulley *et al.*, 1992).

1.3.1 Syndromic XLMR

At the beginning of the work presented in this thesis, in March 2000, there were 134 XLMR syndromes described. Of these 127 display recessive inheritance and 7 show dominant inheritance: in 23 and two cases respectively, the genes responsible have been identified; a further 57 XLMR syndromes had been mapped by linkage to a region of the X chromosome (Hamel *et al.*, 2000) (Figure 1.1). Since this time mutations in genes causing several others have also been discovered.

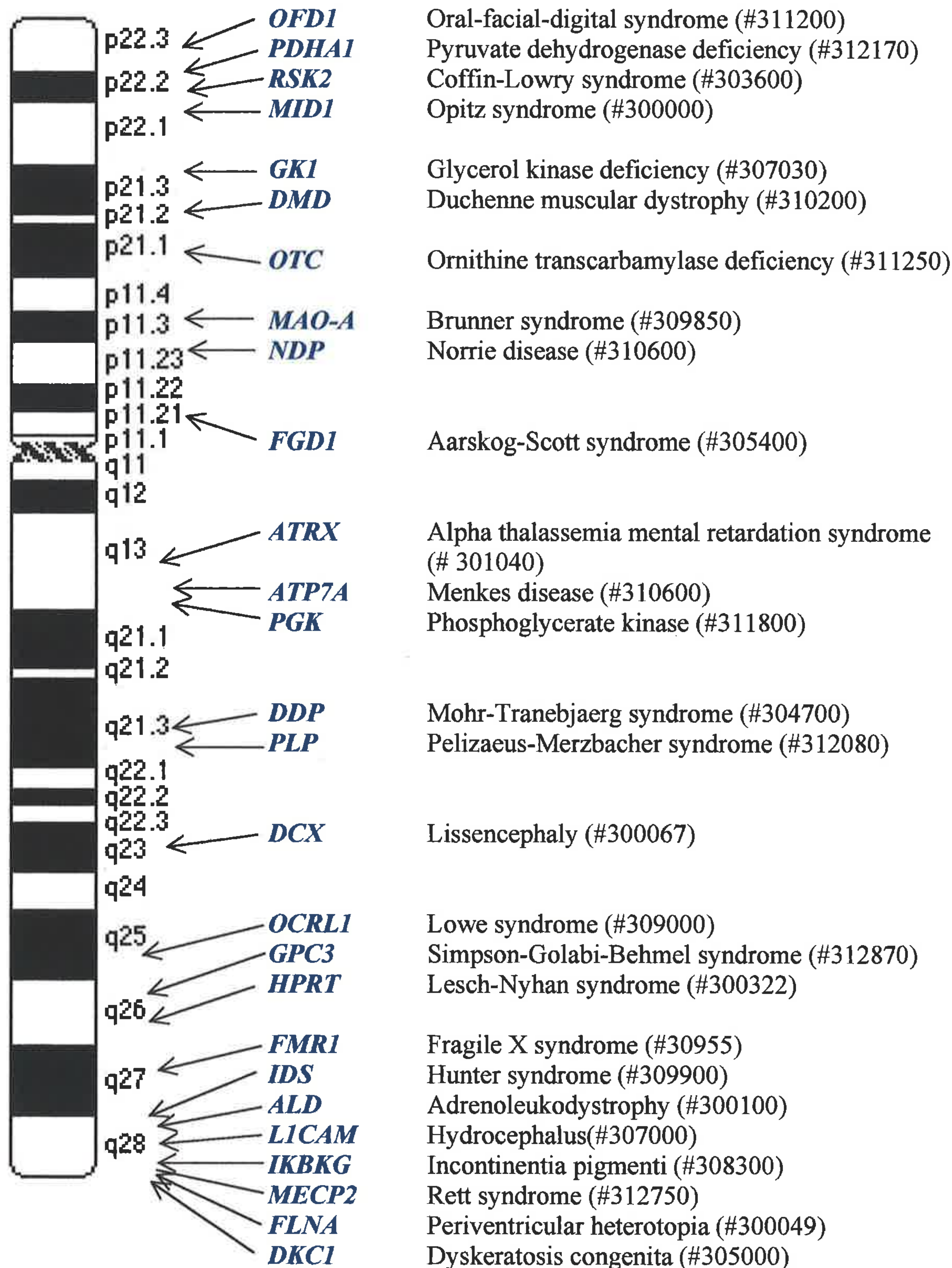


Figure 1.1: Ideogram of the X chromosome showing XLMR syndromes for which the genes have been discovered. OMIM numbers are given in brackets.

The most common form of XLMR is fragile X syndrome (*FRAX*A), with a frequency of 0.22/1000 males (Turner *et al.*, 1996), which is the second most common form of MR after Down's Syndrome, and the most common form of heritable MR. The folate sensitive fragile site in Xq27 is associated with the fragile X syndrome (Lubs, 1969). At the molecular level the fragile site is due to the unstable expansion of a $(CCG)_n$ repeat and hypermethylation of the CpG island (Yu *et al.*, 1991; Oberle *et al.*, 1991; Verkerk *et al.*, 1991). In the normal population the $(CCG)_n$ repeat is polymorphic (with 5-50 CCGs), while *FRAX*A patients have expansions of greater than two hundred CCGs. This expansion and hypermethylation results in a lack of transcription of the *FMR1* gene located downstream. The *FMR1* gene encodes an RNA binding protein (FMRP) (Ashley *et al.*, 1993) that is thought to regulate transport or translation of mRNAs.

Rett syndrome (OMIM# 312750) is another form of syndromic XLMR that has been extensively studied. This disorder has been estimated to affect 0.44/10,000 females (Kozinetz *et al.*, 1993), and is due to mutations in the *MECP2* gene, located within Xq28 (Amir *et al.*, 1999). Absence of affected males and the presence of affected half sisters in families, suggested that this disorder was X-linked dominant, with lethality in males (Hagberg *et al.*, 1983). Different mutations in this gene can cause variable severity of disease. In some families males are viable though such mutations are milder than those with male lethality. Males with these mutations still have a more severe phenotype than their affected sisters. Additionally, milder mutations have recently been found in families with NSXLMR (Orrico *et al.*, 2000).

These are just two of the better known an better studied examples of syndromic XLMR. It is beyond the scope of this thesis to provide too much detail on others; however, more information can be found in Hamel *et al.,* (2000) and in Stevenson *et al.,* (2002).

1.3.2 Non Syndromic XLMR

A number of families in which NSXLMR is segregating have been described, with more than 70 of these being published as MRX families (Chiurazzi *et al.*, 2001) (Figure 1.2). That is, families in which the gene has been localised to a region of the X chromosome by linkage analysis, whereby a lod score of at least +2 is achieved which, in the case of X-linkage, demonstrates linkage (Ott, 1991). These families were assigned consecutive numbers to identify each family (MRX1-78). The genes in many other unpublished NSXLMR families have also been localised and by linkage analysis by groups working on XLMR.

In March 2000, mutations in eight genes had been found to be the cause of NSXLMR (Figure 1.2). The first of these was the *FMR2* gene located in Xq28, which is associated with the folate sensitive fragile site, *FRAXE* (Gécz *et al.*, 1996; Gu *et al.*, 1996). The characterisation of the other seven genes from NSXLMR families soon followed. These include *PAK*3 (Allen *et al.*, 1998), *OPHN1* (Billuart *et al.*, 1998), *GDI*1 (D'Adamo *et al.*, 1998), *RPS6KA3* (Merienne *et al.*, 1999), *IL1RAPL1* (Carrie *et al.*, 1999), *TM4SF2* (Zemni *et al.*, 2000) and *ARHGEF6* (Kutsche *et al.*, 2000). Since this time several others have been implicated by the discovery of mutations in NSXLMR families including *FACL4* (Meloni *et al.*, 2002), *AGTR2* (Vervoort *et al.*, 2002), *MECP2* (Orrico *et al.*, 2000), *FGD1* (Lebel *et al.*, 2002), *ATRX* (Yntema *et al.*, 2002) and *SLC6A8* (Hahn *et al.*, 2002).

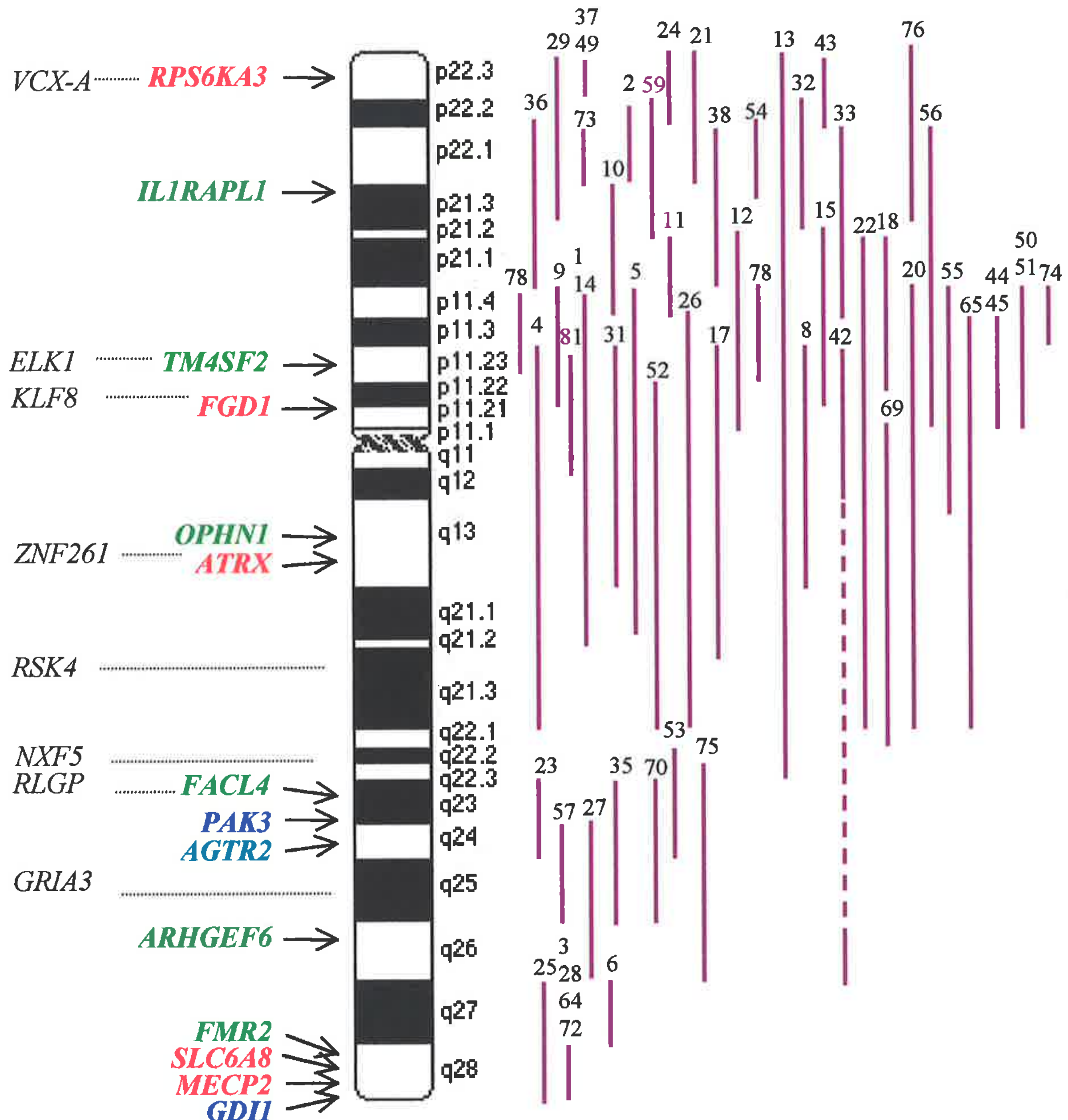


Figure 1.2. Idiogram on the X chromosome showing NSXLMR genes. On the left in green, blue and red are genes in which mutations have been found in MRX families. Green indicates those that were first identified by X chromosome rearrangements. Blue indicates those that were identified in MRX families by positional candidate gene screening. Red indicates genes that were initially found to be MRXS genes prior to the identification of mutations in MRX families. In black are candidate genes that have been shown to be disrupted by X rearrangements in patients with MR, but screening of MRX families has not yet revealed disease causing mutations. On the right are MRX families in which a disease causing mutation is yet to be found. Each bar represents the linkage interval for the family with the MRX number indicated above.

As a result of mapping of chromosome aberrations in patients with MR, such as deletions and translocations, other genes have also been identified as potential candidates for XLMR (Figure 1.2). These are *VCX-A* (Fukami *et al.*, 2000), *RPS6KA6* (Yntema *et al.*, 1999), *ZNF261* (van der Maarel *et al.*, 1996), *GRIA3* (Gécz *et al.*, 1999), *KLF8* (Lossi *et al.*, 2002), *NXF5* (Jun *et al.*, 2001), and *RLGP* (Saito-Ohara *et al.*, 2002). The involvement of these genes in familial forms of XLMR has not been confirmed by the identification of mutations segregating with MR in MRX families. A mutation in the gene *ELK1* (Schroer *et al.*, 2000) in a small family with NSXLMR has also implicated this gene as important for the development of normal cognitive function. This may represent a true NSXLMR mutation, or may be due to a rare polymorphism linked to the MR locus in this family. Again *ELK1* involvement in NSXLMR would be supported by the identification of additional families with mutations in this gene. Alternatively, functional analysis of these candidate genes is needed to show what affect disruption of the gene has on the normal function of the protein product and how this may affect the normal development of the brain.

Different mutations in the same gene can be the cause of both syndromic and non-syndromic forms of MR. Mutations in the gene *RPS6KA3*, as well as resulting in NSXLMR in MRX families, more commonly cause of Coffin-Lowry syndrome (Trivier *et al.*, 1996). Similarly, although most mutations in the gene *MECP2* cause Rett syndrome (Amir *et al.*, 1999) and are lethal to males, Orrico *et al.,* (2000) found a mutation in a family with severe NSXLMR in affected males and mild MR in carrier females. In the *ATRX* gene, mutations usually cause ATR-X syndrome (Gibbons *et al.*, 1995), less frequently Juberg-Marsidi syndrome (Villard *et al.*, 1996), Smith-Fineman-Myers syndrome (Villard *et al.*, 2000) and occasionally NSXLMR (Yntema *et al.*, 2002). Similarly mutations in *FGD1* generally cause Aarskog-Scott syndrome (Pasteris *et al.*, 1994) and mutations in *SLC6A8* cause X-linked creatine deficiency syndrome

(Salomons *et al.*, 2001), but in both cases some mutations lead to NSXLMR (Lebel *et al.*, 2002; Hahn *et al.*, 2002). For *MECP2*, *RPS6KA3*, *ATRX*, *SLC6A8* and *FGD1*, they each became candidate genes for NSXLMR after they were first implicated in XLMR syndromes.

There are now five XLMR genes where syndromic MR is allelic with non-syndromic forms. (Table 1.1). This has long been recognised for *FRAX*A where there are large numbers of affected family members with MR, many of whom don't have the syndromic characteristics of Fragile X syndrome. They cannot all be detected by clinical examination alone and rely on the laboratory examination of *FRAX*A $(GCG)_n$ expansions. Hence as more XLMR genes are being discovered the overlap between NSXLMR and MRXS is becoming apparent.

1.3.3 Genetic Heterogeneity of NSXLMR

Of the 81 MRX families that had been mapped on the X-chromosome only twelve have been shown to display a mutation in one of the 14 known NSXLMR genes. Unpublished families and smaller, unmapped families make up the remainder (Table 1.1). This means that there are still many more NSXLMR genes to be discovered in order to account for the remaining 66 published families, not to mention those that have not been published. Initial estimates of the number of MRX genes based on non-overlapping linkage regions, gave a minimal number of 8 (Gedeon *et al.*, 1996). Expansion of the *FRAX*E $(CCG)_n$ repeat associated with the *FMR2* gene accounts for the greatest number of patients with non-syndromic MR, with an incidence of at least 1/50,000 males (Brown, 1996). All other NSXLMR genes have only been found to contain mutations in one to a few families. The initial estimate of the number of genes for non-syndromic MR on the X-chromosome has now been revised to approximately 22 (Gécz

Table 1.1: **NSXLMR is a genetically heterogeneous group**. The genes causing NSXLMR have only been shown to contain mutations in 1 or a few families (with the exception of *FRAXE* $(CCG)_n$ repeat expansions). In at least 66 MRX families the gene remains unknown. Mutations in the same gene can cause both NSXLMR and MRXS. ('Other' refers to unpublished and unmapped XLMR in which mutations were found).

Gene Name	Gene Symbol	NSXLMR	MRXS	Reference
Fragile X mental retardation 2	*FMR2*	NSXLMR, *FRAXE*		(Gécz *et al.*, 1996; Gu *et al.*, 1996)
p21-activated kinase	*PAK3*	MRX30, MRX47		(Allen *et al.*, 1998)
Oligophrenin 1	*OPHN1*	MRX60		(Billuart *et al.*, 1998)
Guanine-dissociation inhibitor 1	*GDI1*	MRX41, MRX48, 1x other		(Bienvenu *et al.*, 1998; D'Adamo *et al.*, 1998)
Interleukin 1 receptor accessory protein like 1	*IL1RAPL1*	MRX34, 1x other		(Carrie *et al.*, 1999)
Tetraspanin	*TM4SF2*	MRX58, 2x other		(Zemni *et al.*, 2000)
Rho guanine nucleotide exchange factor 6	*ARHGEF6*	MRX46		(Kutsche *et al.*, 2000)
Fatty Acid Co-A ligase 4	*FACL4*	MRX63, MRX68, 1x other		(Meloni *et al.*, 2002)
Angiotensin receptor	*AGTR2*	9x other		(Vervoort *et al.*, 2002)
Ribosomal S6 Kinase	*RPS6KA3*	MRX19,	Coffin Lowry Syndrome	(Merienne *et al.*, 1999)
Methyl CpG binding protein 2	*MECP2*	MRX16, MRX79, 1x other	Rett Syndrome	(Orrico *et al.*, 2000), Winnepenninckx *et al.*, 2002)
Faciogenital Dysplasia	*FGD1*	1x other	Aarskog syndrome	(Lebel *et al.*, 2002)
X-linked helicase 2	*ATRX*	1x other	ATR-X, Juberg-Marsidi, and Smith-Fineman-Myers syndromes	(Yntema *et al.*, 2002)
Solute Carrier 6	*SLC6A8*		X-linked creatine deficiency syndrome	(Hahn *et al.*, 2002)

and Mulley, 2000) although it is likely to be even higher than this. The question still to be answered is whether there is any one gene that accounts for a significant proportion of NSXLMR (other than *FMR2*), or whether the trend will continue, with a large number of genes responsible and mutations in each being a rare cause of NSXLMR.

1.4. FUNCTIONS OF NSXLMR GENES

1.4.1 Intracellular signalling pathways

Several of the genes involved in XLMR, especially NSXLMR, are part of intracellular signal transduction pathways, in particular the Rho signalling pathways. Rho proteins are known to modulate neurite outgrowth in response to extracellular signals, as well as directing the outgrowth of axons and dendrites and the formation of dendritic spines (Jalink *et al.*, 1994; Hall, 1998). RhoA, Rac1 and Cdc42 are the Rho-GTPases which have been studied the most extensively.

Rho-GTPases and their signalling networks are reviewed extensively in (Van Aelst & D'Souza-Schorey, 1997). Briefly, Rho-GTPases are guanine nucleotide binding proteins, that cycle between active GTP-bound and inactive GDP-bound forms. Rho GDP dissociation inhibitors (GDIs), stabilise the GTPase bound to GDP, and sequester them to the cytoplasm where they are inactive. GDP/GTP exchange factors (GEFs) exchange the bound GDP for GTP to change the GTPase to its active form. Once activated the GTPase initiates signalling complexes with one of several effector proteins. GTPase activating proteins (GAPs) catalyse the reaction to hydrolyse the bound GTP to GDP, thus inactivating the effector complexes and

signalling activity is ceased. Mutations in regulators and effectors of Rho-GTPases cause a number of disorders, including XLMR.

GDI1 is a guanine dissociation inhibitor for Rab3a and Rab3c, which are related to Rho-GTPases (Sasaki *et al.*, 1990). *PAK3* is a serine/threonine protein kinase which are known to mediate effects downstream of Rac and Cdc42 on the actin cytoskeleton and gene expression (Manser *et al.*, 1995). *ARHGEF6* is a GEF for RacI and Cdc42 and interacts with PAKs (Daniels *et al.*, 1999). *OPHN1* is a putative RhoGap protein and has GAP activity for RhoA, RacI and Cdc42 in vitro (Billuart *et al.*, 1998) and *FGD1* is a GEF for Cdc42 (Zheng *et al.*, 1996). Therefore there are already several examples of NSXLMR that are members of the Rho signalling pathways.

IL1RAPL1 and *TM4SF2* have been indirectly implicated in Rho-GTPase signalling pathways. *IL1RAPL1* is a member of the interleukin 1 (IL-1) receptor family. IL-1 can stimulate Cdc42 via IL-1 receptors in fibroblasts (Puls *et al.*, 1999). Rac and Rho-GTPases are also activated. *TM4SF2* is a member of the tetraspanin membrane protein family that participate in membrane associated complexes which are focal points of Rho signalling (Stipp & Hemler, 2000).

Additionally, although the *FMR1* gene is involved in MRXS, the protein FMRP has also been indirectly implicated in Rho-GTPase signalling pathways, as it was shown to associate with CYFIP1, which is a known Rac1 interacting protein (Schenck *et al.*, 2001). Hence, the protein products of NSXLMR and MRXS genes may play roles the same pathways.

Therefore correct signalling in these pathways is essential for the development of normal cognition. These are intricate pathways that have many regulatory and effector protein, many of which remain unknown. As the genes and the proteins they encode are discovered, they will become good candidate genes for MR conditions both X-linked or autosomal depending of chromosomal location.

1.4.2 Chromatin Remodelling

ATRX, *MeCP2* and *RPS6KA3* have all been implicated as having a role in chromatin remodelling and transcription regulation. In all three of these cases the majority of mutations in the genes cause more severe syndromic XLMR; however, mutations have been found which cause the milder NSXLMR phenotype. Chromatin remodelling mechanisms play an important role in controlling the accessibility of genetic regions.

ATRX encodes a protein that contains a PHD zinc finger motif and an ATPase/helicase domain of the SWI/SNF type, that are known to be involved in chromatin remodelling. Mutations in this gene cause X-linked alpha thalassemia syndrome (OMIM #301040), which is characterised by severe MR, characteristic dysmorphic facies, genital abnormalities, and an unusual, mild form of haemoglobin H (Hb H) disease. Mutations were also shown to cause Juberg-Marsidi (OMIM #309590), and Smith-Fineman-Myers (OMIM #309580) syndromes, as well as NSXLMR. The ATRX protein associates with pericentromeric heterochromatin, and with the short arm of acrocentric chromosomes which contain ribosomal DNA arrays, suggesting that ATRX might exert chromatin-mediated effects in the nucleus (Gibbons *et al.*, 2000).

RPS6KA3 encodes a member of the ribosomal protein S6 serine/threonine kinase family of protein kinases. Mutations in this gene were initially found to be the cause of Coffin Lowry syndrome (OMIM #303600) which is characterised by MR with peculiar pugilistic nose, large ears, tapered fingers, drumstick terminal phalanges by x-ray, and pectus carinatum. *RPS6KA3* is required for EGF-induced phosphorylation of histone H3, which is important for chromatin remodelling (He *et al.*, 2003). Additionally interaction between RPS6KA3 and CBP (CREB-binding protein) was shown to regulate acetylation of histone H3 (Merienne *et al.*, 2001).

Mutations in *MECP2* were originally found to cause Rett syndrome (OMIM #312750), a severe MR syndrome found to affect mostly females, probably due to lethality in males. Milder mutations were later shown to cause NSXLMR in males and females (Orrico *et al.*, 2000; Winnepenninckx *et al.*, 2002). In a mouse model for Rett syndrome with truncated MeCP2, the protein localises normally however histone H3 is hyperacetylated. This provided evidence that the chromatin structure is affected thereby affecting regulation of gene expression (Shahbazian *et al.*, 2002).

Therefore, these two processes, intracellular signalling and chromatin remodelling, seem to be important in the development of the brain and in achieving normal cognitive function. It can be speculated that other genes involved in these processes also play a role, and that mutations in these may also result in MR. When searching for candidate genes on the X chromosome to be screened in XLMR patients, these genes present as the best candidates.

1.5. AUTOSOMAL MENTAL RETARDATION

The majority of autosomal MR where the cause is known is due to chromosomal aberrations such as translocations, deletions (*eg*.), abnormal copy number (*eg*. Down's Syndrome), or in the case of Prader-Willi and Angelman Syndromes uniparental disomy or incorrect imprinting. Subtelomeric fluorescence *in situ* hybridisation (FISH) detects submicroscopic rearrangements involving the telomeres of autosomes in a proportion of patients with idiopathic MR (Flint *et al.*, 1995). In a group of patients with moderate to severe MR, chromosome rearrangements were detected in 5-7% (Knight & Flint, 2000). In these patients MR is typically associated with other dysmorphic features which is not surprising as the rearrangements may disrupt the function of more than one gene through deletion, physical interruption by breakpoints or by positional effects.

There is only one reported instance where a mutation in an autosomal gene results in non-syndromic MR. In the neurotrypsin gene *PRSS12* on chromosome 4q24, an homozygous 4 bp deletion was found in an inbred eastern Algerian family with four mentally retarded children (Molinari *et al.*, 2002). There are no other reports of large families with autosomally inherited non-syndromic MR, which could be subject to linkage analysis. This may be because mutations in autosomal genes result in mild to borderline MR which may remain undiagnosed and undocumented, or familial inheritance may not be recognised due to incomplete penetrance.

1.6. IDENTIFICATION OF DISEASE GENES

As with other genetic disorders, identification of genes involved in XLMR has been achieved by positional cloning and positional candidate strategies: that is, searching for the gene for a genetic disorder, without prior knowledge of the function of the gene. One of the earliest positional cloning applications in molecular genetics was the discovery of the dystrophin gene, mutations in which cause X-linked Duchenne muscular dystrophy (Monaco, 1986). In this case, large deletions in affected males as well as X;autosome translocations in affected females tagged the region in which genes were to be identified. Positional candidate approaches are also used for identifying the genes in large families segregating a genetic disorder. In this case, linkage analysis is used to identify a region on a chromosome that is shared by all the affected members of the family, but not by the unaffected members. Previously localised genes from within this region are collected and prioritised as candidate genes for mutation screening, based on knowledge of the tissue expression and/or functional information. One of the earliest applications of the positional candidate gene approach was in the discovery of the rhodopsin gene for retinitis pigmentosa (Dryja *et al.*, 1990). In the case of NSXLMR, the positional candidate gene approach was used to identify seven genes, and positional cloning of the breakpoints of X chromosome rearrangements was used to discover the remaining seven genes.

1.6.1 Positional cloning of chromosome rearrangement breakpoints

Even though linkage analysis and candidate gene screening has been successfully used in the identification of genes involved in MRX, this approach has many disadvantages. Characterisation of X chromosome rearrangements associated with MR represents an alternative method for finding MR genes. Deletion (e.g. *FMR*2, *ILRAPL*1) and translocation

(e.g. *ARHGEF6*, *TM4SF2*) breakpoints have provided a specific site on the X chromosome to search for XLMR genes. Typically, a physical map across the region of the rearrangement breakpoint(s) is constructed, such that a contig of clones with human inserts (YAC, BAC, PAC or cosmid clones) is generated. These clones are then used as probes for FISH to narrow down the region in which to search for the breakpoint and therefore genes that may be disrupted. For example, BAC clone inserts are usually between 50-250 kb in length providing a relatively small region in which to search for genes. These rearrangements are usually *de novo* and therefore unlike linkage analysis approaches, DNA and RNA from the person in a family who carries the X chromosome rearrangement is sufficient to find the disrupted gene/s that result in MR. This means that the long and difficult process of collecting samples from a large number of people is not necessary.

1.6.2 Linkage analysis of XLMR families

Linkage analysis involves the use of large families that show an X-linked mode of inheritance of MR. A panel of polymorphic markers along the X chromosome (Dib *et al.*, 1996) is used and linkage is determined by a minimum lod score of +2 between the MR locus and one or more of these markers (unlike autosomal disorders which require a minimum lod score of +3) (Ott, 1991). It is desirable that this analysis also excludes the remainder of the X chromosome (negative lod scores). The outer boundaries of the localisation are defined by recombination events. The limits to narrowing down this regional localisation are influenced by the size of the family being studied, the number of people from which DNA can be obtained, the distribution of recombination events and the informativeness of markers. Correct diagnosis of affected and unaffected males is also vital. This can be difficult especially in families with mild MR where other factors such as genetic background and environment may affect IQ

levels. Due to the high incidence of MR in the population the occurrence of phenocopies within the family can also hinder gene localisation.

Alternatively, assuming phenocopies can also create problems as in the case of one MRX family where a mutation was found in the *TM4SF2* gene (Zemni *et al.*, 2000). One affected male in the pedigree was shown to not carry the same mutation as his affected male relatives. When this male was included in lod score calculations the whole of the X-chromosome was excluded by linkage analysis. However once this male was left out of the analysis a lod score of greater than two was found in the region of the *TM4SF2* gene. It was however later discovered that the mutation found in the family is unlikely to be the MR causing mutation, as linkage was later found at another location (Gomot *et al.*, 2002). Other MRX families however exist with mutations in this gene, and *TM4SF2* remains a NSXLMR gene.

The extreme degree of genetic heterogeneity observed for NSXLMR means that families cannot be grouped together for linkage analysis to minimise the size of the linkage region. Hence each family must be viewed as a separate entity and it must be assumed that each contains mutations in a different gene. This means that some localisations may be very large, making candidate gene screening an enormous task, as even small regions may contain many genes that present as good candidates for MRX.

1.6.3 Identification of candidate genes

Once a region containing the XLMR gene has been delineated by linkage analysis or by localisation of rearrangement breakpoints, candidate genes within this region must be identified. This task has been made easier by progression of the human genome project

leading to the availability of large regions of completed genomic sequence (e.g. National Centre for Biotechnology Information and Sanger Centre). BLAST analysis can be used to identify known genes within a sequenced region, regions that show similarity to other known genes and may represent new members of gene families, and genes that have been already identified in other organisms. As well as this the EST database contains sequences from parts of genes which are yet to be fully characterised or assigned a function. Good candidate genes for MR are those that show similarity to other known XLMR genes or are involved in the same signalling pathways. Genes of unknown function that are expressed in the brain become candidates if they map to the region and no other good candidates are identified. Genes previously shown to be involved in MRXS and which are located within a MRX linkage interval, are also good candidate genes for mutation screening in MRX families. Genes that are shown to be disrupted by the breakpoints of X chromosome rearrangements in patients with MR are also excellent candidates for NSXLMR.

1.6.3 The 'tour de force' approach to disease gene identification

As more and more genes are being discovered and technologies for high throughput screening of genes are being improved, the hunt for genes causing familial XLMR can be attempted without the need for the linkage analysis. A large number of genes are screened in a large group of patients. This approach will be especially helpful in the case of NSXLMR where there is extreme genetic heterogeneity, and a large number of genes remain to be identified. This will also be of great assistance in the identification of genes affected in sporadic cases of MR where it is not even clear that the cause of MR is genetic or environmental, and where there is no indication which region of the genome needs to be studied. An example of this approach is in the identification of the *AGTR2* gene as an XLMR gene. Characterisation of an

X chromosome rearrangement initially identified this gene as a candidate; however, high throughput screening was used to screen for mutations in more than 550 patients with MR of unknown cause (Vervoort *et al.*, 2002). These approaches will become more common, especially when high throughput methods for mutation screening improve.

1.6.4 Mutation detection in genes

After candidate genes have been identified they are screened by any of various mutation detection systems (e.g. direct sequencing, single stranded conformation analysis, denaturing gradient gel electrophoresis (for review see Nollau & Wagener, 1997)). Success rates of mutation detection vary with the different screening methods. Ultimately the mutation must be characterised by sequencing irrespective of which mutation screening method is used. RT-PCR can be used for screening genes more rapidly; however, this requires that the gene of interest is expressed in the available tissue source of the RNA, which is not always the case.

SSCA (single stranded conformation analysis) (Orita *et al.*, 1989) is one method which has been successfully used to detect single base mutations. This technique is based on the fact that single stranded DNA can fold back on itself in a sequence dependant manner. Altering the base composition will produce a new conformation of the single strand. These different conformations migrate differently when electrophoresed on non-denaturing polyacrylamide gels. The sensitivity of SSCA depends on how the mutation affects the folding of the single strand, and how this folding affects its mobility through the gel. This method has been estimated to detect 80-90% of single base mutations using optimal conditions (i.e. analysing fragments of <200 bp) (Sheffield *et al.*, 1993), however this figure may be different for DNA sequences with different characteristics (i.e. high GC content).

Current methods for mutation detection are primarily PCR based, and as such there is the danger that some types of mutations may remain undetected. X chromosome deletions are always detectable in males who only have one X chromosome. However with females a product will be obtained from the normal X chromosome and the deletion would be missed. Similarly with submicroscopic duplications, using PCR based techniques, a product will always be obtained and unless quantitative methods are used the mutation would not be detected.

Additionally mutations can occur that lie outside of the region being screened. Typically genomic DNA is screened by PCR amplification of amplicons that cover the exons containing the open reading frame (ORF), and in some cases the 5' promoter regions. If an RNA source in which the gene of interest is expressed is available the entire ORF may be screened in fewer reactions. Although screening cDNA assists in finding mutations that may affect the transcript that would not be identified from genomic DNA screening (i.e. mutations in the promoter sequences or in the 3'UTR regions including the polyadenylation signals which may result in no transcript being identified by PCR) mutations may still be missed.

Mutations may also lie some distance from the affected gene. Several examples have now been put forward of translocations breakpoints that have been shown to cause a genetic disorder by affecting expression of genes that are located several hundreds of kilobases away. Examples of such positional effects include *PAX6* causing aniridia (Fantes *et al.*, 1995) and *SHH causing* preaxial polydactyly (Lettice *et al.*, 2002) and also in the *AGTR2* where a translocation 150 kb away from the gene effects transcription and causes NSXLMR (Vervoort *et al.*, 2002). Therefore honing in on the breakpoint may not always lead to the identification

of a candidate gene, and sequences further from the breakpoint may need to be examined more closely.

1.7. HUMAN GENOME PROJECT

The Human Genome Project is the work of an international consortium whose broad aim was to map and sequence the entire human genome. The initial stage of the project resulted in a 1 cM genetic map of the human genome, a physical map with 52,000 mapped STSs, 180 Mb of human genomic DNA sequence, improved technologies for sequencing, 30,000 mapped ESTs, as well as progress in sequencing of the genomes of model organisms, and construction of the mouse physical map. The five year goals for 1998-2003 (Collins *et al.*, 1998) were to continue on with this work and were, briefly to:

1 complete the sequence of the human genome;
2. further advances in sequencing technologies;
3. study human genome sequence variation and create a SNP map of at least 100,000 markers;
4 develop technologies for functional genomics, including full length cDNA sequences;
5 comparative genomic projects which include sequencing of model organism;
6. examination of the ethical, legal and social implications of the project;
7. improve bioinformatics and computational biology, by improving databases, and improving tools for functional studies;
8. train scientist and encourage the establishment of academic career paths for genomic scientists.

The draft sequence of the human genome was released by two groups in 2001 (Lander *et al.*, 2001; Venter *et al.*, 2001). The human genome sequence is expected to be fully completed in soon, with the maps of individual chromosomes being published by the end of 2003. Along with this the draft sequence of the mouse genome was released in December 2002 (Waterston *et al.*, 2002) and there is now a waiting list for the sequencing of the genomes of additional model organisms.

Such sequencing efforts are of immeasurable value for the identification of disease genes. In the case of positional cloning of chromosome rearrangements, BAC and PAC human clones can be easily identified from the region of interest. Once a clone is identified that spans a breakpoint the sequence contained within a spanning human clone can be readily searched for known genes or expressed sequence tagged sites (ESTs) (Lennon *et al.*, 1996). The sequences of more and more full length cDNA clones are also becoming available due to the efforts of the Human Genome Project. This can quickly identify potential genes for further analysis. The type of tissue from which the cDNA was derived also gives some indication as to the expression of the gene, and can assist in choosing candidate genes for screening (for example, in the case of XLMR, ESTs from a linkage region that were derived from a brain cDNA present as good candidates). In the case of linkage analysis the genomic sequence between the two outer recombinants is readily available, and a transcript map of the region can quickly be constructed that includes both known and putative genes. Genes that present as good candidate genes can then be screened for mutations. This, of course, has greatly increased the speed of positional cloning approaches to disease gene identification, where human clone contigs once had to be constructed from screening cosmid, YAC, PAC and BAC libraries, and gene identification required a number of techniques such as exon trapping or cDNA selection. All of this can now be done *in silico* in a relatively much shorter amount of time.

1.8. BENEFITS OF IDENTIFYING MENTAL RETARDATION GENES

Identification of genes from familial mental retardation will result in precise diagnostic tests, which in turn will assist in managing suitable patient treatment and counselling. In the case of large families where the causative gene remains unknown, female relatives of affected boys can only get a percentage risk figure for their carrier status based on linked markers and hence of the risk of a male foetus being affected. In some cases polymorphic markers used for linkage analysis can be uninformative in all or some parts of the pedigree, and hence gathering of data for risk calculations can be unsatisfactory in cost and labour. Discovery of causative genes in these families will result in definitive carrier status determination, allowing carrier females to make informed choices when planning their families, and allowing definitive prenatal diagnosis of male foetuses.

Risk determination based on linked markers cannot be used in the case of small NSXLMR families. It can be used with caution for syndromic XLMR if genetic homogeneity can be assumed. However, finding causative genes in large families will result in diagnostic tests for small families and for sporadic cases with MR in addition to the large families from where these genes were first identified. As mutations in large number of genes are ultimately going to be the cause of MR, high throughput diagnostic tests will have to be developed in order to make practical the screening of a large number of genes in the large number of patients with MR.

1.9. AIMS

The aim of the research presented in this thesis is to increase our understanding of what is required for normal cognitive abilities, by identifying some of the genes in familial and sporadic cases of MR. To do this two approaches will be used. Firstly characterisation of breakpoints of X chromosome rearrangements in patients with MR, will lead to the identification of genes in sporadic cases which become candidate genes for screening in familial XLMR. Secondly, candidate gene screening from within minimal linkage intervals in mapped XLMR families will be applied to identify MR causing mutations.

Specific Aims

1) To characterise at the molecular level, the inversion breakpoints from two unrelated males with MR with the karyotypes 46,Y,inv(X)(q13.1q28), and 46,Y,inv(X)(p11.2q28). The steps to achieve this aim are:

- Construction of a contig of human clones across part of Xq28
- These human clones will be used as probes for FISH in order to narrow down the breakpoint localisations
- Genes will be identified in the region of the breakpoint, and it will be determined if the breakpoint disrupts the gene
- The other breakpoint will be determined
- Candidate genes identified will be screened in available XLMR families mapping to the region in which the new candidate gene is located

Hypothesis

i. The X chromosome inversions in the patients disrupt a gene or genes which results in MR

2) To characterise the Xq28 breakpoint in a female patient with MR and cardiomyopathy associated with the karyotype 46, X, t(X;10)(q28;q11.1). Steps to achieve this aim are:

- Human clones from Xq28 will be used to narrow down the localisation of the breakpoint
- Genes in the region of the breakpoint will be characterised and assessed as to wether the breakpoint disrupts the gene
- The gene will then become a candidate gene for screening in families with both MR and cardiomyopathy
- The breakpoint on chromosome 10 will be assessed to determine whether autosomal genes contribute to the phenotype

Hypotheses

i. The patient has skewed X-inactivation such that only genes from the translocated X chromosome will be expressed

ii. The translocation breakpoint in Xq28 is the main contributor to the phenotype

iii. More than one gene may be disrupted, at least one accounting for MR alone

3) To determine the region of linkage and to define the minimal linkage interval in a previously unresolved family where linkage was consistent two regions of the X-

chromosome, and to screen candidate genes in the redefined region. Steps to achieve this aim are:

- Additional family members will be sought for inclusion in the linkage analysis
- Clinical details of affected males will be re-examined
- A transcription map of the defined region will be constructed, *in silico*
- Good candidate genes for XLMR will be identified and screened by direct sequencing in affected members from the family

Hypothesis

i. MR in the family is segregating as a monogenic disorder

4) To identify the gene causing X-linked Infantile Spasms Syndrome (ISSX). Steps to achieve this aim are:

- A transcript map will be constructed across the minimal linkage interval
- Candidate genes will be screened by direct sequencing
- Other XLMR families mapping to the same region will be screened for mutations in the ISSX gene

Hypotheses

i. MR in the ISSX families being studied is monogenic

ii. MRXS genes also contain mutations causing NSXLMR

Chapter 2

Materials and Methods

2.1. PATIENT SAMPLES **31**

2.2. DNA EXTRACTION **32**

2.2.1 LYMPHOCYTE DNA ISOLATION 32
2.2.1.1 Proteinase K Treatment *32*
2.2.1.2 Phenol Extraction *33*
2.2.1.3 Ethanol Precipitation *33*
2.2.2 PLASMID DNA ISOLATION 34
2.2.3 COSMID DNA ISOLATION 35
2.2.4 BAC/PAC DNA ISOLATION 36
2.2.5 PLASMID DNA ISOLATION FOR MAMMALIAN CELL TRANSFECTION 38

2.3. RNA EXTRACTION **39**

2.3.1 DNASE TREATMENT 40

2.4. DNA AND RNA QUANTITATION **40**

2.5. POLYMERASE CHAIN REACTION **41**

2.5.1 PCR PRIMERS 41
2.5.2 PCR CONDITIONS 41
2.5.3 PCR AMPLIFICATION OF THE *ARX* GENE FOR SEQUENCING 42

2.5.4 Cycling Conditions 42
2.5.5 PCR amplification of the *ARX* gene for Single Strand Conformational Analysis 43
Exons 1, 4 and 5 .... *43*
Exon 2 (parts 1, 2 and 3) *44*
Exon 3 .... *45*
2.5.6 Long range PCR 45

2.6. GENOTYPING 46

2.6.1 Polyacrylamide Gel Electrophoresis 46
2.6.2 Linkage analysis 46

2.7. PREPARATION OF SOUTHERN BLOTS 47

2.7.1 Restriction Digestion and Agarose Gel Electrophoresis 47
2.7.2 Southern Transfer 48

2.8. PROBE LABELLING AND HYBRIDISATION 49

2.8.1 Labelling Reaction 49
2.8.2 Southern Blot Hybrisation 50
2.8.3 Northern Blot Hybridisation 51

2.9. CLONING 52

2.9.1 Insert Preparation 52
2.9.1.1 PCR Products *52*
2.9.1.2 Plasmid DNA *53*
2.9.2 Vector Preparation 53

2.10. LIGATION 54

2.10.1 In Solution 54
2.10.2 In Low Melt Temperature Agarose 54

2.11. TRANSFORMATION 55

2.11.1 Preparation of Competent Cells 55
2.11.2 Transformation 55

2.12. SEQUENCING 56

2.12.1 BigDye Terminator Cycle Sequencing 56
2.12.2 Purifying Extension Products 56

2.13. FIRST STRAND cDNA SYNTHESIS 57

2.14. MAMMALIAN CELL TRANSFECTION 58

2.15. FLUORESCENCE IN SITU HYBRIDISATION 60

2.15.1 Interphase FISH 60

2.16. 3'RACE 60

2.16.1 Synthesis of first stand cDNA 61
2.16.2 Amplification of Target cDNA 61

2.17. SSCA 62

2.17.1 Fluorescent SSCA 63

2.18. SEQUENCE ANALYSIS 63

2.18.1 DNA and protein sequence analysis programmes 63
2.18.2 Online databases for sequence analysis 64

2.1. PATIENT SAMPLES

Samples were collected with informed consent from affected and normal individuals.

Patients samples with X chromosome rearrangements (Chapter 3) were collected by Drs David Pincus [(Patient 1) Canada Allamanda Medical Centre, Southport, Australia], Steven Bamforth [(Patient 2) Medical Genetics Clinic, Edmonton) and John Christodoulou [(Patient 3) Children's Hospital at Westmead, Australia) in the form of peripheral blood and fibroblast biopsies. Fibroblast cell lines were established and maintained by Cathy Derwas and Sarah McDonnell (WCH, Australia).

Patient samples for linkage analysis and gene screening in the OH family (Chapter 4) were collected by Professors Michael Partington and Gillian Turner (Hunter Genetics, New South Wales, Australia).

DNA samples from ISSX probands (Chapter 5) were provided by Suzanne Lewis (Children's and Women's Health Centre of British Columbia, Canada), Drs Suzanna Frints and Jean-Pierre Fryns (University Hospital, Leuven, Belgium). XMESID family material was provided by Dr Ingrid Scheffer (University of Melbourne, Australia). Patient DNA samples for subsequent screening of the *ARX* gene were collected by Professors Gillian Turner and Michael Partington from the Hunter Genetics, NSW, Australia. Probands from additional XLMR families were kindly provided by Dr Charles Schwartz from Greenwood Genetic Center, Greenwood, USA.

2.2. DNA EXTRACTION

2.2.1 Lymphocyte DNA Isolation

DNA was extracted from 5-20 ml of whole blood collected in tubes containing EDTA using a method modified from (Wyman and White, 1980).

2.2.1.1 Proteinase K Treatment

I. The blood was transferred to a 50 ml tube and made up to 30 ml with cell lysis buffer (0.32 M sucrose, 10 mM Tris-HCL (pH 7.5), 5 mM $MgCl_2$, 1% (v/v) triton-X 100). The samples were incubated on ice for 30 minutes and then centrifuged for 15 minutes at 2,000 x g at 4°C. The supernatant was discarded, and the pellet gently resuspended in a further 30 ml of lysis buffer. The samples were centrifuged at 2,000 x g for 15 minutes at 4°C.

II. Proteinase K buffer was then added (3.25 ml of 10 mM Tris-HCl (pH 7.4), 10 mM NaCl, 10 mM EDTA (pH8.0)). The samples were vortexed, and 500 μl of 10% SDS added.

III. Next, 200 μl of Proteinase K (10mg/ml) was added, and the samples incubated at 37°C overnight with constant agitation.

2.2.1.2 Phenol Extraction

I. To the cell lysate 5 ml of buffer saturated phenol was added (1:1, cell lysate:phenol) (the phenol had previously been equilibrated with 10 mM Tris-HCl, 1 mM EDTA). The samples were placed on a rotating wheel (10rpm) for 15 minutes.

II. The samples were centrifuged at 2,000 x g for 10 minutes at 15°C and the upper aqueous phase was transferred to a clean 10 ml tube.

III. Steps I and II were repeated and the sample was then made up to 10 ml with chloroform:isoamyl alcohol (24:1) and mixed gently before being centrifuged at 400 x g for 10 minutes at room temperature. The upper aqueous phase was removed and transferred to a clean 10 ml tube.

2.2.1.3 Ethanol Precipitation

I. To the aqueous phase 1/10 v/v (500 μl) of 3 M sodium acetate (pH 5.2) and 2 volumes of cold ethanol (-20°C) (10 ml) were added. The samples were mixed gently followed by centrifugation at 400 x g for 10 minutes at room temperature. The precipitated DNA was then transferred to an eppendorf tube and was washed in 1 ml of 70% ethanol.

II. The samples were centrifuged at maximum speed in a microfuge for 15 minutes. The pellets were washed once more with 70% ethanol and were vacuum dried for approximately 15 minutes or until the DNA pellets were dry.

III. The dried DNA pellet was gently resuspended overnight in 100 µl of sterile TE (10 mM Tris-HCl, 0.1 mM EDTA (pH 8.0)) ready for DNA quantitation.

2.2.2 Plasmid DNA Isolation

All plasmid DNA purification were using the alkaline lysis method (Birnboim and Doly, 1979). Plasmid DNA for all uses except mammalian cell transfection, was extracted using the Rapid Pure Miniprep (RPM™) (BIO 101). The method was modified slightly to obtain DNA from a starting culture of 10 ml (the method starts from 1.5 ml).

I. An overnight culture of plasmid was grown in 10 ml of LB. The culture was centrifuged at 2,000 x g for 10 minutes at room temperature. The supernatant was discarded and the cell pellet resuspended in 50 µl of kit-supplied pre-lysis buffer and transferred to an eppendorf tube. To the resuspended pellet 100 µl of kit-supplied alkaline lysis buffer was added and the cell lysate was mixed well by vigorous pipetting until the cell lysate was clear and viscous. To the cell lysate 100 µl of kit-supplied neutralising solution was then added and the sample mixed by vortexing before centrifugation in a microfuge at maximum speed for 2 minutes. The supernatant was then transferred to a kit-supplied spin filter containing 250 µl of glassmilk. The solution was mixed by pipetting before centrifugation in a microfuge at maximum speed for 1 minute.

II. The solution in the collection vial was removed, 350 µl of kit supplied wash solution was added to the spin filter before centrifugation for 1 minute in a microfuge at

maximum speed. The collection vial was emptied and the tube centrifuged for a further 1 minute.

III. The spin filter was transferred to a clean collection vial. Plasmid DNA was eluted by the addition of 50 µl of sterile dH_20 and mixed by vortexing before centrifugation in a microfuge at maximum speed for 30 seconds.

2.2.3 Cosmid DNA Isolation

Cosmid DNA was isolated using QIAGEN Plasmid Midi Kit.

I. Single colonies of cosmids were grown overnight at 37°C in a shaking incubator, in 100 ml of LB media containing the appropriate antibiotic. Cells were harvested by centrifugation at 3,000rpm for 10 minutes at 4°C.

II. The bacterial pellets were resuspended in 4 ml of kit supplied Buffer P1, and were transferred to 40 ml oakridge tubes. This was followed by the addition of 4 ml of kit supplied buffer P2 (cell lysis buffer). The sample was mixed thoroughly by inverting the tube until the cell lysates were clear and viscous. Kit supplied buffer P3 (neutralisation buffer) was then added (4 ml), and the samples mixed thoroughly by inverting the tubes, and were incubated on ice for 15 minutes.

III. The samples were centrifuged at 18,000 x g for 30 minutes at 4°C. The supernatant was removed and transferred to a new tube. The centrifugation was repeated, and again the supernatant transferred to a new tube.

IV. QIAGEN-tip 100 columns were then equilibrated with 4 ml of kit-supplied buffer QBT. The resuspended pellets from step III were then applied to the columns, and once the samples had flowed through, each was washed two times with 10 ml of the kit-supplied wash buffer.

V. Cosmid DNA was then eluted from the columns by the addition of 4 ml of kit-supplied elution buffer QF. The sample was then transferred into 5 eppendorf tubes (800 μl in each eppendorf). DNA was precipitated by the addition of 560 μl (0.7 volumes) of isopropanol and centrifuged in a microfuge at maximum speed for 30 minutes. The supernatant was removed and the DNA pellets washed by the addition of 1 ml of 70% ethanol, before further centrifugation for 5 minutes at maximum speed. The 70% ethanol was removed and the DNA pellets dried under vacuum. The DNA from each eppendorf was resuspended in 20 μl of sterile dH_20, before all DNA from the same clone were combined in one tube ready for quantitation and further use.

2.2.4 BAC/PAC DNA Isolation

Both BAC and PAC DNA were isolated using QIAGEN Plasmid Midi Kit.

I. Single colonies of BACs and PACs were grown overnight at 37°C in a shaking incubator, in 200 ml of LB media supplemented with the appropriate antibiotic (170 μg/ml of chloramphenicol and 50 μg/ml of kanamycin respectively). Bacterial cells were harvested by centrifugation at 3,000 x g for 10 minutes at 4°C.

II. Pellets were resuspended in 10ml of kit-supplied P1 buffer and were transferred to 40 ml oakridge tubes. To this, 10 ml of kit-supplied buffer P2 was then added, and the samples gently mixed by inverting until the cell lysates became clear and viscous. Kit-supplied buffer P3 was added (10 ml), the samples mixed by inverting, and were incubated on ice for 30 minutes.

III. The samples were centrifuged at 18,000 x g for 25 minutes at 4°C. The supernatant was transferred to fresh tubes and re-centrifuged at 18,000 x g for 10 minutes at 4°C. The supernatant was split and transferred into 2 new tubes. DNA was precipitated by the additions of 0.7 volumes of room-temperature isopropanol, followed by centrifugation at 15,000 x g for 30 minutes at 4°C. The supernatant was decanted and the pellets resuspended in 250 µl of dH_2O. The two suspensions of each sample were recombined before the addition of 450 µl of kit-supplied buffer QBT.

IV. QIAGEN Tip-100 columns were then equilibrated with 4ml of kit-supplied buffer QBT. The resuspended pellets from step III were then applied to the columns, and once the samples had flowed through, each was washed two times with 10 ml of the kit-supplied wash buffer.

V. PAC and BAC DNA was then eluted from the columns by the addition of 4 ml of kit-supplied elution buffer QF. The sample was then transferred into 5 eppendorf tubes (800 µl in each eppendorf). DNA was precipitated by the addition of 560 µl (0.7 volumes) of isopropanol and centrifuged in a microfuge at maximum speed for 30 minutes. The supernatant was removed and the DNA pellets washed by the addition of 1 ml of 70% ethanol, before further centrifugation for 5 minutes at maximum speed.

The 70% ethanol was removed and the DNA pellets dried under vacuum. The DNA from each eppendorf was resuspended in 20 µl of sterile dH_2O, before all DNA from the same clone were combined in one tube ready for quantitation and further use.

2.2.5 Plasmid DNA Isolation for Mammalian Cell Transfection

Plasmid DNA for transfection into mammalian cells was prepared using the QIAGEN Plasmid Midi kit and QIAfilters as follows.

I. Single colonies were grown in 50 ml cultures of LB media supplemented with 100 µg/ml of ampicillin and grown overnight, shaking at 37°C. The cells were harvested by centrifugation for 10 minutes at 3,000rpm at 4°C.

II. The bacterial pellet was resuspended in 4 ml of kit supplied buffer P1, and then 4 ml of kit supplied buffer P2 was added. The solution was mixed thoroughly by inverting and incubated at room temperature for 5 minutes. To the lysate 4 ml of kit-supplied buffer P3 was added and mixed several times by inverting before being poured into the barrel of the QIAfilter midi cartridge and incubated at room temperature for 10 minutes.

III. A QIAGEN tip-100 column was equilibrated for each sample by loading 4 ml of kit-supplied buffer QBT onto the column. A plunger was inserted into the QIAfilter cartridge and the cell lysate filtered into the prepared column. Once the solution had run through the column, the DNA was washed by the addition of 10 ml of kit-supplied buffer QC; this wash was repeated.

IV. The DNA was then eluted in 4 ml of kit supplied buffer QF. The sample was then transferred into 5 eppendorf tubes (800 μl in each eppendorf). DNA was precipitated by the addition of 560 μl (0.7 volumes) of isopropanol and centrifuged in a microfuge at maximum speed for 30 minutes. The supernatant was removed and the DNA pellets washed by the addition of 1 ml of 70% ethanol, before further centrifugation for 5 minutes at maximum speed. The 70% ethanol was removed and the DNA pellets dried under vacuum. The DNA from each eppendorf was resuspended in 20 μl of sterile dH_2O, before all DNA from the same clone were combined in one tube ready for quantitation and further use.

2.3. RNA EXTRACTION

RNA was extracted from cultured fibroblast and lymphoblast cells using the TRIZOL ® LS reagent. Cells were washed two times in PBS before proceeding as by the manufacturers instructions as follows.

I. The cell pellet was resuspended in sterile dH_20 to make a final volume of 250 μl. To this, 750 μl of TRIZOL LS reagent was added and mixed by pippetting to lyse the cells. The solution was incubated at room temperature for 5 minutes before the addition of 200 μl of chloroform. This was mixed by vigorous shaking for 15 seconds and then incubated at room temperature for 10 minutes. The samples were centrifuged at 12,000 x g for 15 minutes at 4°C.

II. The aqueous phase was transferred to a clean tube and the RNA was precipitated by the addition of 500 µl of isopropanol. The solution was incubated at room temperature for 10 minutes and then centrifuged at 12,000 x g for 10 minutes at 4°.

III. The supernatant was removed and the RNA pellet washed by the addition of 1 ml of 75% ethanol. After vortexing the sample was centrifuged at 7,500 x g for 5 minutes at 4°C. The supernatant was removed and the RNA was allowed to air-dry.

IV. The RNA was resuspended in 50 µl of DEPC treated sterile dH_2O.

2.3.1 DNase Treatment

DNA was removed from the RNA sample by the use of DNA-free™ (Ambion).

To the RNA sample, 6 µl of kit-supplied 10x DNase I buffer and 1 µl of DNase I (2 U) were added, mixed and incubated at 37°C for 30 minutes. The DNase I was inactivated by addition of 5 µl of kit-supplied DNase Inactivation Reagent and was mixed well before incubating at room temperature for 2 minutes. This was centrifuged at 10,000 x g for 1 minute, and the RNA solution was removed from the pelleted DNase Inactivation Reagent and transferred to a clean tube, ready for quantitation and further use.

2.4. DNA AND RNA QUANTITATION

DNA concentration was determined by measuring the optical density on a spectrophotometer (Pharmacia Biotech UltraSpec 3000) at a wavelength of 260 nm. One OD unit is equal to 50

μg/ml of double stranded DNA. For RNA one OD unit is equal to 40 μg/ml. DNA or RNA was diluted 1 in 100 and the optical density measured at 260 nm. Using Beer's law and the known molar extinction coefficients above, the concentration was calculated using the formula:

DNA μg/ml = absorbance at OD_{260} x dilution factor (100) x molar extinction coefficient/1,000

2.5. POLYMERASE CHAIN REACTION

2.5.1 PCR primers

Oligonucleotides for PCR were designed such that they contained as close to 50% GC content. Oligonucleotides were usually 20 bp in length. All oligonucleotides were checked to make sure they did not contain repetitive sequence by using the BLAST programme at NCBI. Most sequences are given throughout the text in the sections for which they are relevant. Oligonucleotides for gene screening are listed in Appendix 1.

2.5.2 PCR conditions

All PCRs except those for the amplification of the *ARX* gene were amplified in 50 μl reactions using the 10x buffer supplied (100mM Tris-HCl, 15 mM $MgCl_2$, 500 mM KCl (pH8.3 [20°C]) with Taq DNA Polymerase (Roche) as per the manufacturers instructions. Each reaction contained the following

	Vol per 50 µl	Final Concentration
DNA (100 ng/µl)	1.0	2 ng/µl
10x reaction buffer	5.0	1x (1.5 mM $MgCl_2$)
10mM dNTP mix	1.0	200 µM
forward primer (50 µM)	1.0	1 µM
reverse primer(50 µM)	1.0	1 µM
Taq polymerase (5U/ul)	0.5	0.5 U/µl
dH_2O	to 50 µl	

2.5.3 PCR amplification of the ARX gene for sequencing

Due to the difficulty in amplifying the *ARX* gene using these conditions (because of the high GC content of the sequences being amplified) the Failsafe 2x Buffer system (Epicentre Technologies) was used. 2x Buffer J was found to amplify *ARX* exons for sequencing, using the DNA polymerase mix from the Expand Long Template PCR System (Roche). The 2x buffer contains 100 mM Tris-HCl (pH 8.3), 100 mM KCl, 400 µM of each dNTP, 3 mM $MgCl_2$ and 8x Failsafe PCR Enhancer. Each 50 µl reaction contained 100 ng of total human DNA, 50 pmol of each oligonucleotide (final concentration of 1 µM), 1x J Buffer and 1.75 U of DNA polymerase mix.

2.5.4 Cycling Conditions

All reactions were carried out using the same cycling conditions as follows:

94°C	30 sec	
60°C	30 sec	x 30 cycles
72°C	30 sec	

This was altered if expected PCR product sizes were greater than 500 bp, by adding 30 sec to the extension time per 500 bp. If the size of the amplicon was unknown, the extension time was 3 minutes.

2.5.5 PCR amplification of the ARX gene for Single Strand Conformational Analysis

Exons 1, 4 and 5

Exons 1, 4 and 5 were amplified using the same conditions. A 2x PCR buffer was used that consisted of: 33 mM $(NH_4)_2SO_4$, 133 mM Tris-HCl (pH 8.8), 13 μM EDTA, 0.34 mg/ml BSA, 20% DMSO, 400 μM dNTPs.

Each reaction consisted of::

	μl per 10 μl reaction	Final Concentration
DNA (30 ng/μl)	1	3 ng/μl
2x PCR buffer	5	1x
15 mM $MgCl_2$	1.35	2.0 mM
1 M βME	0.1	10 mM
Primer mix (100 ng/μl of forward and reverse primer)	1.0	5 ng/μl
Taq Polymerase (5 U/μl)	0.2	0.1 U/μl
dH_2O	to 10 μl	

Cycling conditions were as follows:

94°C	60 sec	x 10 cycles
60°C	90 sec	
72°C	90 sec	

followed by:

94°C	60 sec	
55°C	90 sec	x 25 cycles
72°C	90 sec	

Exon 2 (parts 1, 2 and 3)

The three products using the different sets of primers for exon 2 were all amplified using the same conditions. The reaction buffer is 2x Buffer J mix (Epicentre), and the DNA polymerase mix is from the Expand Long Template PCR System (Roche). Each reaction consisted of::

	μl per 10 μl reaction	Final Concentration
DNA (30 ng/μl)	1	3 ng/μl
Forward Primer (10 μM)	0.5	0.5 μM
Reverse primer (10 μM)	0.5	μM
2x Premix J	5	1x
DNA polymerase mix (3.5 U/μl)	0.15	0.5 U/μl
dH_2O	to 10 μl	

Cycling conditions were as follows:

94°C	2 min	x 1 cycle
94°C	30 sec	
60°C	30 sec	x 35 cycles
72°C	68 sec	
68°C	10 min	x 1 cycle

Exon 3

Exon 3 amplification was the same as for exon 2 above, however, instead of 2x Buffer J mix, the 2x Buffer G mix (Epicentre) was used. Cycling conditions were also the same as for exon 2.

2.5.6 Long range PCR

Expand Long Template PCR System (Roche) system 2 was used for amplifying products greater than 3 kb as per the manufacturers instructions. This system used a 10x buffer that contained 22.5 mM $MgCl_2$. Each 50 µl reaction consisted of:

	µl per 50 µl reaction	Final Concentration
DNA (250 ng/µl)	1	5 ng/µl
10x Buffer 2	5	1x
10mM dNTP mix	1.75	350 µM
Forward primer (10 µM)	1	1 µM
Reverse primer (10 µM)	1	1 µM
DNA polymerase mix (3.5 U/µl)	0.75	0.5 U/µl
dH_2O	to 50 µl	

Cycling conditions were as follows:

94°C	30 sec	
60°C	30 sec	x 35 cycles
68°C	10 minutes	

2.6. GENOTYPING

2.6.1 Polyacrylamide Gel Electrophoresis

Primers for amplification of polymorphic repeat loci were from (Dib *et al.*, 1996), and were amplified as in Section 2.5, but were in 10 μl reactions, with the addition of 1 μCi of α^{32}P-dCTP per reaction. PCR products of $(AC)_n$ repeats used for linkage and generating haplotypes were separated using a 38 x 50 cm BioRad Sequigen™ sequencing cell with 0.4 mm spacers. The gels consisted of 5% polyacrylamide (19:1 ratio of acrylamide:bisacrylamide) containing 7 M urea and 1xTBE. Gels were pre-electrophoresed at 2,200 V in order to allow the gel to warm up to 50°C. Prior to electrophoresis PCR products were mixed with 6x formamide loading buffer (98% formamide, 0.05% xylene cyanol, 0.05% bromophenol blue), and heat denatured at 94°C for 5 minutes, and 5 μl of this was loaded onto the gel. Gels were then electrophoresed at 1,800-2,000 V at constant temperature of 50°C. Gels were dried under vacuum and PCR products were visualised by exposing the dried gel to X-ray film at -70°C for 2-24 hours.

2.6.2 Linkage analysis

Two point lod scores were calculated from genotyping data using the program MLINK, from the LINKAGE package version 5.1 (Lathrop and Lalouel, 1984).

The LINKAGE package of programs (version 5.1) comprise four analytical programmes MLINK, LINKMAP, LODSCORE and ILINK (Lathrop and Lalouel, 1984; Lathrop *et al.*,

1984). Only the first one was used in the analysis of the family OH (Chapter 4) to generate lod scores in a two point analysis.

To run linkage programmes the user must first input two different files. The INFILE contains all of the information regarding the pedigree in terms of pedigree structure, affection status and marker genotypes for each locus on each individual. The DATAFILE contains the disease gene frequency (usually 0.001) and mode of inheritance. The allele frequencies were determined by dividing one by the number of alleles observed. The linkage support programmes (MAKEPED, PREPLINK, LCP, PEDIN and LRP) allow for easier construction and running of linkage input files.

MLINK was used for gathering evidence of linkage by preparing lod score tables of polymorphic markers against the disease phenotype. The programme calculated pairwise lod scores at given recombination fractions for each marker.

2.7. PREPARATION OF SOUTHERN BLOTS

The method used for preparation of Southern Blots as well as hybridisation (section 2.8.2) is derived from (Haan *et al.*, 1988)

2.7.1 Restriction Digestion and Agarose Gel Electrophoresis

For Southern Blots, 8 μg of total human DNA was digested overnight at the appropriate temperature (usually 37°C) with 20 U of restriction endonuclease in the user supplied recommended buffer. Restriction enzymes were supplied by New England Biolabs and MBI

Fermentas. The digested DNA was separated on a 1% (w/v) agarose gel containing 1xTBE. Gels were electrophoresed in 1xTBE at 15mA for 16 hours. DNA was visualised by staining with ethidium bromide in 1xTBE (2.5 μg/ml) for 30 minutes and UV transillumination.

2.7.2 Southern Transfer

I. Following photography the gel was submerged in denature solution (2.5 M NaCl, 0.5 M NaOH) for 30 minutes with gentle agitation. The solution was then removed and replaced with neutralising solution (1.5 M NaCl, 0.5 M Tris-HCl (pH 7.5)) and gently agitated at room temperature for a further 30 minutes.

II. The gel was inverted onto a Southern Transfer tray containing 10xSSC. A piece of Gene Screen Plus (Dupont) membrane was soaked in 10xSSC and placed on top of the gel. This was topped with absorbent paper, and the transfer was allowed to proceed overnight.

III. The absorbent paper was discarded, and the membrane was denatured by soaking in 0.5 M NaOH with gentle agitation for 1 minute, followed by neutralising in a solution containing 100 mM Tris-HCl (pH 7.5), 2xSSC for a further 1 minute. The membrane was allowed to air dry at room temperature before hybridisation.

2.8. PROBE LABELLING AND HYBRIDISATION

2.8.1 Labelling Reaction

DNA probes were labelled using the random priming method derived (Feinberg and Leahy, 1983) as follows.

For each probe, 50 ng of DNA to be used as a probe was labelled in a 50 μl reaction. A 10x buffer containing 500 mM Tris-HCl (pH 7.5), 100 mM $MgCl_2$ 10 mM DTT, 500 mg/ml BSA, 0.0005 A_{260}/ul pdN(6) (random hexamers)] was used (Pharmacia), as well as a 10x dNTP solution contained 10 mM each of dATP, dGTP and dTTP (Pharmacia). The probes were labelled with α^{32}P-dCTP (NEB). Each reaction contained the following:

	Vol per 50 μl	Final Concentration
DNA (50 ng/μl)	1.0	2 ng/μl
10x reaction buffer	5.0	1x
10x dNTP mix	5.0	1x
α^{32}P-dCTP (10 μCi/μl)	1.0	0.2 μCi/μl
Klenow fragment (5U/μl)	1.0	0.5 U/μl
dH_2O	to 50μl	

The sample was heated to 100°C and allowed to cool before the addition of 1 μl (10 U) of Klenow fragment (MBI Fermentas). The labelling reaction was allowed to proceed at 37°C for 30 minutes.

Prior to hybridisation, single copy probes were denatured at 94°C for 5 minutes before being added to the hybridisation bottle. If the probe to be used contained repetitive DNA sequences a prereassociation reaction was carried out to block these repeats as follows. The probe (50 μl) was mixed with 100 μl of 10 mg/ml human placental DNA, 50 μl of 20xSSC and 1 μl of 0.5 M EDTA (pH 8.0). The sample was heated to 94°C for 15 minutes, and then incubated at 65°C for at least 1 hour. The probe was then added directly to the hybridisation bottle when required.

2.8.2 Southern Blot Hybrisation

Southern blots were hybridised in a hybridisation mix containing 5xSSPE, 1% SDS, 50% deionised formamide, 10% w/v dextran sulphate MW500,000 as follows.

I. The membranes were prewet in 5xSSC and placed in hybridisation bottles with 10 ml of hybridisation mix. The membrane was prehybridised for 1 hour at 42°C for 1 hour.

II. The probe was heat denatured and added to the bottle. If the probe had been prereassociated then the heat denature step was omitted. Hybridisation was allowed to proceed overnight at 42°C was constant agitation.

III. Following hybridisation the membranes were washed twice in 500 ml of 2xSSC, 0.5% SDS with constant agitation at 65°C for 5 minutes. A second, more stringent wash followed using a solution containing 0.1xSSC, 0.1% SDS for 10 minutes at 65°C.

IV. The membrane was then blotted to dry, wrapped in plastic and X-ray film was exposed for 24 hours (or as required) at -70°C with an intensifying screen before developing.

2.8.3 Northern Blot Hybridisation

Northern blots were purchased from Clontech and were hybridised in ExpressHyb™ Hybridisation solution (Clontech) by the manufacturers instructions. Briefly,

I. The ExpressHyb Solution was warmed to 68°C before 5 ml was added to a hybridisation bottle containing the Northern blot. The membrane was prehybridised for 30 minutes at 68°C. The denatured probe was then added to the bottles and hybridisation was allowed to proceed overnight at 68°C with continuous shaking.

II. The membrane was rinsed several times in a solution containing 2xSSC, 0.05% SDS. It was then washed for 30 minutes at room temperature with constant agitation. The membrane was then washed for a further 30 minutes in a solution containing 0.1xSSC, 0.1% SDS.

III. The membrane was then blotted to dry, wrapped in plastic and X-ray film was exposed for 24 hours (or as required) at -70°C with an intensifying screen before developing.

2.9. CLONING

2.9.1 Insert Preparation

2.9.1.1 PCR Products

For cloning of PCR products a 50 µl PCR reaction was purified using the QIAquick PCR Purification kit (QIAGEN) to remove primers, polymerase and unincorporated nucleotides. All centrifugation steps were carried out at maximum speed in a microfuge.

I. The sample to be purified was mixed with 5x the volume (250 µl) of the QIAquick PB buffer. The sample was transferred into a kit supplied spin column and placed in a 2 ml collection tube. The sample was centrifuged for 1 minute and the solution in the collection tube discarded.

II. To the column, 750 µl of QIAquick PE buffer was added, and the sample centrifuged for 1 minute. The solution was removed from the collection tube, and the tube was recentrifuged. The spin column was transferred to a clean eppendorf an allowed to air dry for 5 minutes.

III. To elute the DNA from the column, 50 µl of sterile dH_2O was added to the spin column and the eluted DNA was collected in the eppendorf by centrifugation for 1 minute.

2.9.1.2 Plasmid DNA

Plasmid DNA was digested with the appropriate restriction endonuclease in the user supplied recommended buffer, and at the recommended temperature. Restriction enzymes were from New England Biolabs and MBI Fermentas.

The digested DNA was separated on 1.5 % low melt temperature agarose gels in 1xTBE electrophoresed at 100V for 1 hour. The fragment of correct size for further cloning was excised from the gel using a scalpel. This was melted at 70°C for 2 minutes prior to use in the ligation reaction.

2.9.2 Vector Preparation

For cloning of PCR products, prepared pGEM-T vector was used (Promega).

For cloning of full length *ARX* open reading frame, vectors used were pGEM-T, pUC19 (New England Biolabs) and pSPORT1 (GIBCOBRL). For localisation of ARX protein, the insert was cloned into the phrGFP-N1 vector (Stratagene) with the GFP fused in frame at the N-terminus of ARX.

For vector preparation, 1 μg of vector DNA was digested in with the appropriate restriction endonuclease in a volume of 20 μl. The vector was purified using the QIAquick PCR purification kit (see section 11.1.2). The linearised vector was then dephosphorylated to prevent religation, using 2.5 U of calf intestinal alkaline phosphatase (CIAP) (Roche) and the provided buffer in a volume of 60 μl at 37°C for 1 hour. After CIAP treatment the vector was once again purified using the QIAquick PCR purification kit.

2.10. LIGATION

2.10.1 In Solution

Vector and inserts were ligated in a range of vector:insert ratios from 1:1 to 1:3, in a 20 μl reaction. The reaction contained 100 ng of vector and 100-600 ng of insert, 3 U of T4 DNA ligase Promega) and 1x Rapid Ligation buffer (Promega). Ligations were at room temperature for 2 hours or overnight.

Ligation of *EcoN1* overhangs (1 bp 5' overhang), which are more difficult to ligate than ends of other restriction sites, was essentially the same, however highly concentrated ligase was used (NEB). The ligations were carried out using 1 μl (2000 U) of the enzyme, the supplied buffer and the addition of 10% (v/v) PEG_{8000}. Ligations were at room temperature overnight.

Efficacy of ligations was tested before transformation by running 5 μl of the ligation on 1% (w/v) agarose gel, in 1xTBE at 100V for 1 hour.

2.10.2 In Low Melt Temperature Agarose

The concentration of DNA in the agarose gel slice was estimated, and an appropriate volume was ligated to vector as follows. The agarose gel slice was melted at 70°C for 2 minutes. The appropriate volume was then mixed with 100ng of prepared vector, and ligation buffer. This was mixed thoroughly by pipetting. Once the solution was cool and the agarose had started to thicken, 1 μl (3U) of T4 DNA ligase was added and mixed by pipetting. The reaction was incubated at room temperature for 2 hours or overnight.

2.11. TRANSFORMATION

2.11.1 Preparation of Competent Cells

Ligation reactions which used vectors carrying resistance to ampicillin were transformed into SURE® 2 cells (Stratagene) using a method modified from (Chung *et al.*, 1989). SURE® 2 cells are an *E.coli* strain that have mutations in DNA repair pathways, and thus make it easier to clone hard to clone fragments (such as those containing repetitive DNA sequences).

I. SURE® 2 cells were grown on LB agar plates containing tetracycline, and a single colony was inoculated into 10 ml of LB media and grown at 37°C shaking overnight. The next day the cells were diluted 1/50 into fresh media (400 µl of overnight culture into 20 ml of LB) and incubated at 37°C with shaking until the OD_{600} was near 0.3 (~2.5 hours). The cells were harvested by centrifugation at 4,000rpm for 10 minutes at 4°C, and the supernatant discarded.

II. The cells were resuspended in 2 ml of PEG/DMSO solution (10% (w/v) PEG_{3000}, 5% (v/v) DMSO, 10 mM $MgSO_4$, mM $MgCl_2$ in LB media) and placed on ice for 10-30 minutes prior to transformation.

2.11.2 Transformation

To each ligation 50 µl of prepared competent cells was added, mixed gently and incubated on ice for 15 minutes. Then 20 µl of 1 M glucose and 950 µl of PEG/DMSO solution was added. The samples were incubated at 37°C with vigorous shaking for 1 hour, before plating onto LB agar plates containing the appropriate antibiotic.

2.12. SEQUENCING

All sequencing samples were electrophoresed by the ABI automated DNA sequencers at the IMVS Molecular Pathology Unit (Adelaide, Aust) as part of a paid service.

2.12.1 BigDye Terminator Cycle Sequencing

PCR products and cloned DNA were both sequenced using the BigDye terminator cycle sequencing kit as per the instructions of the supplier (Perkin Elmer). PCR products were cleaned using the QIAquick PCR purification kit (see section 2.9.1.1) and plasmid DNA was made using the Rapid Pure Miniprep (RPM™) (BIO 101) (see section 2.2.2). Each 20 µl reaction contained 500 ng of plasmid DNA or 50 ng of PCR product, 3.2 pmol of primer, 2 µl ABI BigDye™ sequencing mix and 6 µl of BetterBuffer (MicrozoneLtd).

PCR conditions were as follows:

96°C	30 sec	x 25 cycles
50°C	15 sec	
60°C	4 min	

2.12.2 Purifying Extension Products

PCR reactions were transferred to eppendorf tubes and mixed with 60 µl of isopropanol and 20 µl of sterile dH_20, vortexed to mix, and were incubated at room temperature for 15 minutes. The extension products were then pelleted by centrifugation in a microfuge at full speed for 20 minutes. The supernatant was removed and the pellets were washed with 1 ml of

75% isopropanol, mixed by vortexing, and centrifuged in a microfuge at full speed for 10 minutes. The isopropanol was removed and the pellets allowed to air-dry.

2.13. FIRST STRAND cDNA SYNTHESIS

First strand cDNA synthesis was performed using SUPERSCRIPT™RNase H $^-$ Reverse Transcriptase (GIBCOBRL®) as follows.

In a 0.5 ml PCR tube 1 µg of total RNA was mixed with 1 µl of random hexamers and 10 µl of sterile dH_20 (11 µl of dH_20 for the negative reaction that would not contain reverse transcriptase). This was incubated at 70°C for 10 minutes and then place on ice.

To each tube 4 µl of 5x First Strand Buffer, 2 µl of 0.1 M DTT (both provided with the enzyme) and 1 µl of 10 mM dNTP mix were added. This was placed at 42°C to equilibrate the temperature before the addition of 1 µl (200 U) SUPERSCRIPT RT (excluding negative controls) and subsequent incubation for 1 hour.

Following first strand cDNA synthesis, 1 µl of the resulting products were used as template for PCR using standard conditions described in (see section 2.5). All were tested with a control reaction using oligonucleotides for the Esterase D (*ESD*) gene, to confirm that the first strand cDNA synthesis was successful. The oligonucleotide sequences were as follows:

Forward 5' – GGAGCTTCCCCAACTCATAAATGCC – 3'

Reverse 5' – GCATGATGTCTGATGTGGTCAGTAA – 3'.

2.14. MAMMALIAN CELL TRANSFECTION

HeLa cells were grown in DMEM growth media in 5% CO_2 at 37°C. The day before transfection, the cells were trypsinized, counted and $5x10^5$ cells plated out on 35 mm plates as follows.

I. The cells were washed twice with PBS. For T75 (75 ml) culture flasks 2 ml of trypsin (company) was added to the flask. The cells were incubated in the trypsin for 5 minutes at room temperature. The cells were removed from the culture flask by hitting the side of the flask. 8 ml of DMEM (minus antibiotic) was added and the cells were removed from the sides of the flask and mixed by pippetting. The solution was transferred to a 10 ml tube and centrifuged at 1100rpm for 5 minutes. The culture media was removed and the cell pellet was resuspended in 10 ml of PBS. The cells were mixed thoroughly to obtain a single cell suspension.

II. Trypan blue was diluted 1 in 5 in 0.9 M NaCl. For cell counting, 100 µl of the single cell suspension was added to 100 µl of the diluted trypan blue, and the cells were loaded onto a haemocytometer, the cells counted and the concentration of the cell suspension determined.

III. $5x10^5$ cells were added to 35mm culture dishes containing 2 ml of DMEM (minus antibiotic) and containing a sterilised coverslip. The cells were grown overnight at 37°C with CO_2, and the following day they were observed to ensure they were at 90-95 confluency. Before transfection, the media was replaced with DMEM (minus foetal calf serum, and minus antibiotic).

HeLa cells were transfected with plasmid DNA containing the entire open reading frame of the *ARX* gene in frame with the GFP protein (phrGFP-N1 vector – Promega), using the LIPOFECTAMINE™ 2000 Reagent (LF2000) as follows.

I. For each dish of HeLa cells to be transfected, 1 μg of plasmid DNA was diluted in 200 μl of DMEM (without foetal calf serum, and without antibiotic). Likewise for each dish 3 μl of LF2000 was diluted in 200 μl of media. This was incubated at room temperature for 5 minutes. The DNA mix was then combined with the LF2000 mix and incubated at room temperature for 20 minutes.

II. The DNA/LF2000 solution was added to the culture dishes containing HeLa cells at a confluency of 90-95%. The cells were incubated at 37°C with CO_2 until ready to be fixed. For growth of longer than 24 hours the media was replaced each day.

III. Coverslips were removed from the culture dishes and the cells were washed twice with PBS and then treated in 0.37% (v/v) formaldehyde for 30 minutes at room temperature. Following two further washes in PBS, the coverslips were placed onto slides containing 40 μl of DAPI antifade and sealed to the slides with rubber cement.

IV. GFP was viewed by FITC filter, and nuclear staining by DAPI filter on an Olympus BX40 microscope. Images of cells were captured by a cooled CCD camera using the CytoVision Ultra image collection and enhancement system (Applied Imaging Int Ltd).

2.15. FLUORESCENCE IN SITU HYBRIDISATION

All FISH was carried out by Erica Woollatt (WCH). DNA from BAC clones to be used as probes for FISH were nick translated with biotin-14-dUTP and hybridised *in situ* at a final concentration of 20 ng/µl to metaphases chromosomes. The FISH method was modified from that described by Callen *et al.,* (1990) in that chromosomes were stained before analysis with both propidium iodine (as counterstain) and DAPI (for chromosome identification). Images of metaphase preparations were captured by a cooled CCD camera using the CytoVision Ultra image collection and enhancement system (Applied Imaging Int Ltd). FISH signals and the DAPI banding pattern were merged for figure preparation.

2.15.1 Interphase FISH

BAC 3007F1 and BAC 54I20 were directly labelled with SpectrumOrange and SpectrumGreen (Vysis) respectively, according to the manufacturers instructions and co-hybridised at a final concentration of 20 ng/ul to interphase nuclei of Patient 2 (Chapter 3).

2.16. 3'RACE

3'RACE was performed using the 3'RACE System for Rapid Amplification of cDNA ends (GIBCO BRL) as per the manufacturers instructions as follows.

2.16.1 Synthesis of first stand cDNA

I. Each RNA sample (1 µg of total RNA) was mixed with sterile dH_2O to a final volume of 11 µl in a 0.5 ml tube. The oligo(dT) adapter primer (AP) was added (1 µl from a 10 µM stock) before the sample was heated to 70°C for 10 minutes. The sample was chilled on ice for 1 minute.

II. To the chilled sample 2 µl of 10x PCR buffer, 2 µl of 25 mM $MgCl_2$, 1 µl of 10 mM dNTP mix and 2 µl 0.1 M DTT (all kit supplied) were added. The final composition of the reaction was 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 2.5 mM $MgCl_2$, 10 mM DTT, 500 nM of AP, 500 µM of each dNTP and 1 µg of RNA. The sample was mixed and equilibrated to 42˚C for 5 minutes.

III. To each sample, 1 µl of SuperScript™ II Reverse Transcriptase was added. The tubes were incubated at 42˚C for 40 minutes. The reaction was terminated by incubation at 70˚C for 15 minutes.

IV. To each sample, 1 µl of RNase H was added, and incubated for 20 minutes at 37˚C.

2.16.2 Amplification of Target cDNA

A reaction in a fresh 0.5 ml tube was set up which included the following:

First strand cDNA synthesis product	2 µl
10x PCR buffer	5 µl
25 mM $MgCl_2$	3 µl
sterile dH_2O	36.5 µl
10 mM dNTP mix	1 µl
Gene specific primer (10 µM)	1 µl
AUAP (10 µM)	1 µl
Taq DNA polymerase (5 U/µl)	0.5 µl
dH_2O	to 50 µl

All components were kit supplied except for the Gene specific primer. This primer was designed from the gene whose 3' end was being examined. AUAP was a primer designed to prime off of the AP introduced by the first strand cDNA synthesis using AP.

The samples were amplified by 35 cycles of 94˚C 20 sec, 60˚C 30 sec, 72˚C 3 minutes.

3' RACE products were analysed by electrophoresis on 1.5% agarose gels. Products were purified using the QIAquick PCR purification kit (see section 2.9.1.1) and sequenced using the BigDye terminator cycle sequencing kit as per the instructions of the supplier (Perkin Elmer).

2.17. SSCA

For analysis of *ZXDA* and *ZXDB* (Chapter 3), PCR products were amplified as by section 2.5, but in a final volume of 10 µl, and with the addition of 1 µCi of $\alpha^{32}P$-dCTP. After completion

of the PCR, the samples were mixed with 10 μl of formamide loading buffer (98% formamide, 0.05% bromophenol blue, 0.05% xylene cyanol). Each sample was heated to 94°C for 10 minutes and was placed on ice before 3 μl was loaded on non-denaturing polyacrylamide gel [4.5% (49:1, acrylamide:bisacrylamide) in 1xTBE]. The gels were electrophoresed at 600 V overnight. Gels were dried under vacuum and the single strand conformation of the PCR products were visualised by exposing the dried gel to X-ray film at -70°C for 2-24 hours.

2.17.1 Fluorescent SSCA

Fluorescent SSCA was used for screening large numbers of samples for mutations in the *ARX* gene (Chapter 5).

Fluorescently (HEX) labelled oligonucleotides were used for the amplification of *ARX* exons in 9 separate reactions. Oligonucleotides are listed in Appendix 1. Products were separated on non-denaturing 4% polyacrylamide gels containing 2% glycerol on a GelScan 2000 (Corbett research) (by Bree Hodgson, WCH).

2.18. SEQUENCE ANALYSIS

2.18.1 DNA and protein sequence analysis programmes

The LaserGene sequence analysis programme (DNASTAR) was used for analysing both DNA and protein sequences. This package consists of several programmes for sequence analysis, EditSeq, SeqMan, MegAlign and MapDraw were used.

DNA and protein sequences were imported into EditSeq where they could be saved as files that could be used in other programmes. SeqMan was used to align DNA sequences. This included both aligning ESTs from UniGene clusters, as well as aligning sequence chromatograms with normal sequence in order to identify DNA sequence changes. MegAlign was used for the alignment of multiple protein sequences. MapDraw was used to identify recognition sites of restriction endonucleases in a known DNA sequence.

2.18.2 Online databases for sequence analysis

Most of the searching was done within the NCBI web pages. All genomic sequences and characterised gene sequences were downloaded from the nucleotide database division and were imported into the EditSeq programme (LaserGene) for further analysis. DNA sequences were searched for similarity to other sequences within GenBank by using the BLAST search (nr for genomic sequences and characterised genes; dbEST for EST sequences indicating the presence of both characterised and uncharacterised genes; htgs for genomic sequence in draft form). The BLAST 2 sequences option was used for comparison of mRNA and genomic sequences in order to determine intron/exon boundaries. Protein sequences were searched against the protein databases at NCBI in order to find orthologs and other similar proteins. The address for NCBI and the BLAST page are as follows:

http://www.ncbi.nlm.nih.gov

http://www.ncbi.nlm.nih.gov/BLAST/

Genomic sequences were searched against the BAC end database maintained by The Institute of Genome Research (TIGR) in order to find overlapping clones for contig construction. The address for the TIGR BAC end database is as follows:

http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html

Repeats were identified within genomic sequences using the Repeat Masker web site. This site also provides the user with their submitted sequence with repeats masked as Ns, to allow further database searching. The Repeat Masker address is as follows:
http://ftp.genome.washington.edu/cgi-bin/RepeatMasker

Once large segments of genomic sequence were available and were included contigs the Ensembl Genome Browser was used. This web site allowed confirmation of sequence contigs, gene localisations. The address for the Ensembl Genome Browser is as follows:
http://www.ensembl.org

Chapter 3

Molecular Characterisation of X chromosome Rearrangements Involving an Xq28 Breakpoint in Three Patients with Mental Retardation

3.1. INTRODUCTION **67**

3.2. RESULTS **69**

3.2.1 PHYSICAL MAP, Xq28 69

3.3. RESULTS PATIENT 1: 46,Y,inv(X)(q13.1q28) **72**

3.3.1 FLUORESCENCE IN SITU HYBRIDISATION 72
3.3.2 MOLECULAR CHARACTERISATION OF THE Xq28 BREAKPOINT 72
3.3.3 THE Xq13 BREAKPOINT 78
3.3.4 DISCUSSION 79

3.4. RESULTS PATIENT 2: 46,Y,inv(X)(q13.1q28) **82**

3.4.1 Xp11.2 BREAKPOINT 82
3.4.2 Xq28 FISH AND INTERPHASE FISH 84
3.4.3 DISCUSSION 86

3.5. RESULTS PATIENT 3: t(X;10)(q28;q11.2) **89**

3.5.1 Xq28 FISH 89
3.5.2 THE GENE *TMG3a* IS DISRUPTED BY THE Xq28 TRANSLOCATION BREAKPOINT. 91
3.5.3 MUTATION SCREENING OF THE PROLINE-RICH GLA PROTEIN GENE, *PRRG1*, LOCATED IN Xp21.1 100
3.5.4 ANALYSIS OF THE CHROMOSOME 10 BREAKPOINT 101
3.5.5 DISCUSSION 103

3.6. CONCLUSION **108**

3.1. INTRODUCTION

Characterisation of chromosome rearrangement breakpoints has been a successful and relatively straightforward approach for the identification of disease genes. In the case of XLMR several genes have been identified using this positional cloning approach (see Chapter 1). Cloning of genes at or near the breakpoints is especially useful to identify candidate genes to be screened in familial XLMR where linkage intervals are large and contain many genes. In this chapter, three unrelated patients with different X chromosome rearrangements, and all three having MR, were studied with the aim of identifying genes whose expression was disrupted by breakpoints (Figure 3.1). All three patients have one of their breakpoints within Xq28 and characterisation of all three Xq28 breakpoints was performed. This meant that a physical map of Xq28 and BAC contigs constructed would be relevant to all three.

Patient 1 is a male who has a paracentric inversion of the X chromosome [46, Y, inv(X) (q13.1q28)] and has severe learning difficulties. This patient had been described in detail in (Villard *et al.*, 1999). Briefly, he also has poor speech, reduced muscle tone and mild dysmorphic features. This inversion is also carried by his mother who is intellectually normal. Interestingly there is a history of severe psychiatric illness among her male relatives; however, this was not further investigated on the request of the family. Fluorescence *in situ* hybrisation (FISH) and long range restriction mapping had previously localised the Xq13.1 breakpoint to a 250 kb fragment near *DXS131* and *DXS162* (Villard et al., 1999). However, there was no gene found to be disrupted by the breakpoint. At the time genomic sequence for this region was unavailable. The patient was seen and samples collected by Dr David Pincus from the Allamanda Medical Centre, Southport, Australia.

Patient 2 has a pericentric inversion of the X chromosome [46, Y, inv(X)(p11.2q28)], and has non-syndromic MR with no dysmorphic or other features. This patient was seen by Dr Steven Bamforth (Medical Genetics Clinic, Edmonton, Canada).

Patient 3 is a female who has a balanced translocation between the X chromosome and chromosome 10 [46, X, t(X;10)(q28;q11.2)]. This patient presented with MR associated with cardiomyopathy (clinical investigations and samples collected by Dr John Christodoulou Childrens Hospital at Westmead, Australia). In this girl the normal X was preferentially inactivated suggesting that the phenotype is caused by the disruption of a gene located within Xq28. It is possible, however, that all or part of the phenotype in this patient is attributable to a gene on chromosome 10.

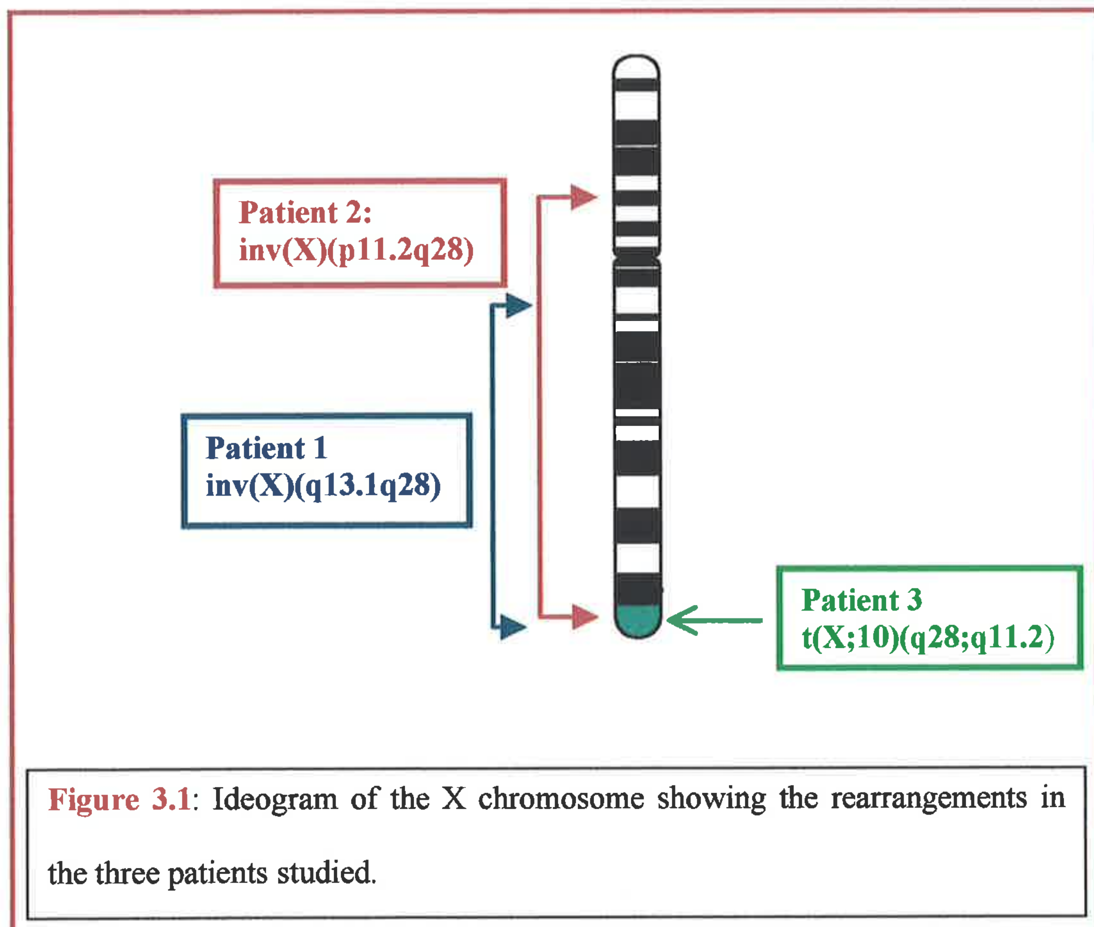


Figure 3.1: Ideogram of the X chromosome showing the rearrangements in the three patients studied.

In all three a positional cloning approach was used to characterise the breakpoints at the molecular level. This was carried out in order to identify a gene at each breakpoint that was either physically interrupted by the break, or whose normal expression was effected. As all patients have a breakpoint within Xq28, this region was concentrated on for the initial part of the study.

3.2. RESULTS

3.2.1 Physical map, Xq28

A cosmid clone, cos E0166 from Xq28 (provided by Daniela Toniolo) had previously been used as a probe for FISH for both Patients 1 and 2. Patient 1 gave distal signal using this probe, while Patient 2 gave spanning signal. Additionally the YAC 46B2 had been used for FISH for Patient 1 and had given proximal signal. A BAC and sequence contig were constructed *in silico*, between this cosmid and YAC that spanned a large part of Xq28 (Figure 3.2). Initially, previously localised STSs and genes from Xq28 (Rogner *et al.*, 1994; Dib *et al.*, 1996) were used to identify genomic sequences from within the region by BLAST searches of the nr and htgs divisions of GenBank. At the time, some of this region was fully sequenced, parts were working draft sequences and some regions were not represented by any genomic sequences in GenBank. This resulted in a physical map and BAC contig across Xq28 using the publicly available sequences and by searching the BAC end database (TIGR) (Figure 3.2). Searching the BAC end database allowed gaps to be filled in. If two different sequences contained one end each of a BAC clone then they could be linked and orientated with respect to each other. The final BAC contig contained several gaps where there was lack of genomic sequence and it was not possible to walk further *in silico*.

The BACs from the constructed contig were used as probes for FISH (carried out by Erica Woollatt, WCH, Adelaide, Australia) for Patient 1 as well as for Patient 3 to try to find a BAC clone spanning the inversion and translocation breakpoints respectively. The contig was regularly updated once new sequences were deposited in GenBank. The nr and dbEST divisions of GenBank were searched with the sequence in order to determine the location of both characterised and uncharacterised genes within the contig. The sequence and BAC contigs as well as some of the genes from the region are shown in Figure 3.2. The region contains more than 30 known genes including the creatine-transporter gene, *SLC6A8,* that has recently been found to contain mutations in patients with NSXLMR (Hahn *et al.*, 2002). This region is now almost completely sequenced and is represented by positions 144.6-147.3 Mb of the X chromosome (Ensembl release 8.30.1). BAC clones were purchased from Genome Research and DNA was prepared and used as a probe to metaphase chromosomes in order to refine the breakpoint localisations.

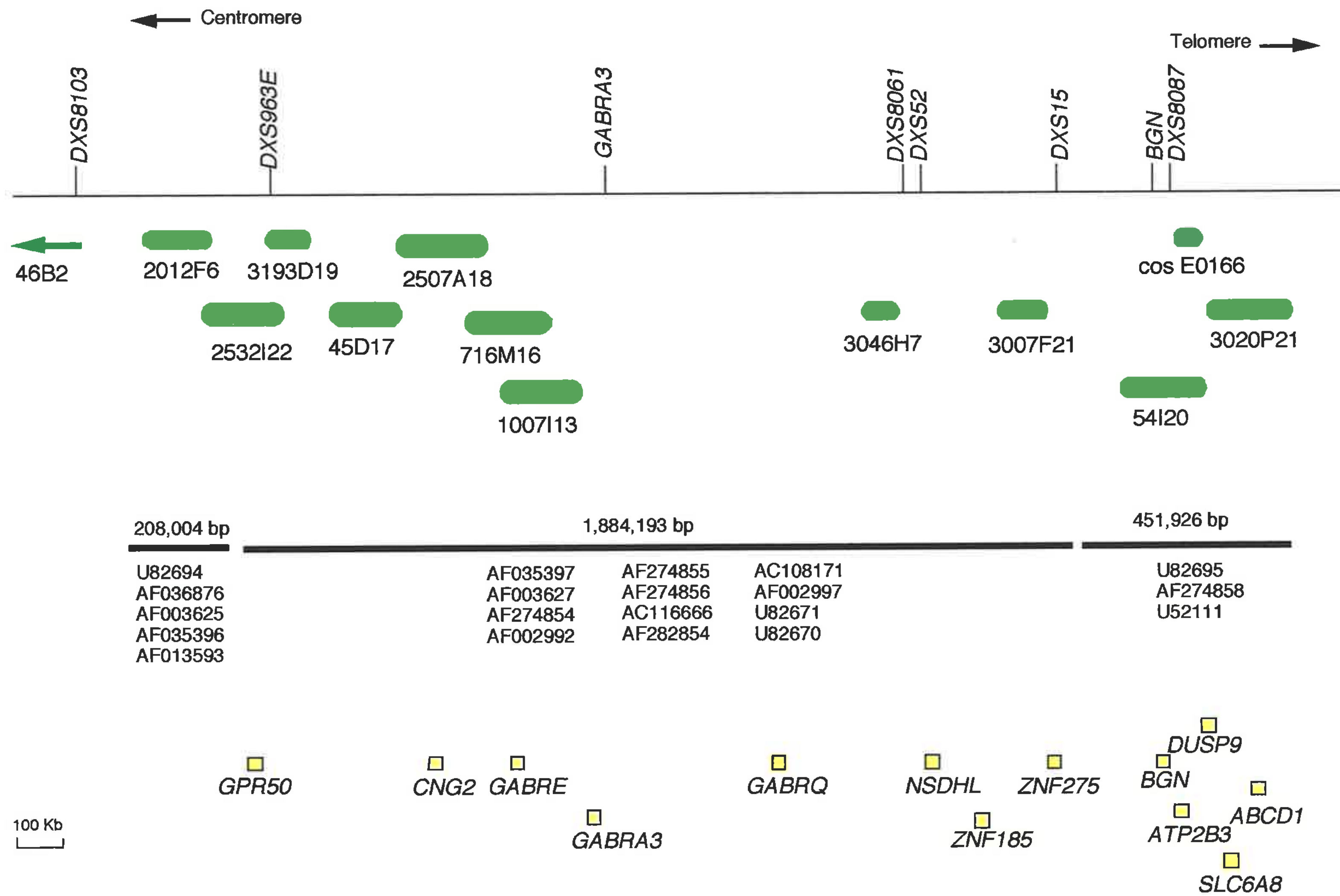


Figure 3.2: Physical map of Xq28 showing STSs, BAC contigs, sequence contigs (and the accession numbers of the sequences within each contig) and some of the characterised genes from across this region.

3.3. RESULTS PATIENT 1: 46,Y,inv(X)(q13.1q28)

3.3.1 Fluorescence in situ hybridisation

The BAC clone 54I20 was shown to span the Xq28 breakpoint by FISH, with probes located telomeric and centromeric to this giving distal and proximal signal respectively (Figure 3.3). This was consistent with the inversion in this patient occurring without any other rearrangements and with the breakpoint being located within the 54I20 sequence. 54I20 is a large clone of ~250 kb, however, further searching of the BAC end database revealed another BAC (3011F21) located within. This BAC, which has an insert of only 60 kb, also showed spanning signal by FISH (Figure 3.3) and hence the breakpoint was refined to within 60 kb.

3.3.2 Molecular Characterisation of the Xq28 breakpoint

The 60 kb BAC, 3011F21, contained the entire biglycan gene (*BGN*), the 5' end of the brain specific plasma membrane calcium ATPase (*ATP2B3*) and part of the exonuclease *TREX2* (Figure 3.4). Given that Patient 1 is male and therefore only has one X chromosome, PCR was used to determine the location of the breakpoint. Long range (LR)-PCR was used such that up to 6 kb segments were initially amplified (using the oligonucleotides listed in Table 3.1) in order to exclude most of the sequence in the vicinity of these genes from being interrupted by the inversion breakpoint. The InvPCR 2 product was obtained from control DNA, however no product was obtained from Patient 1. New primers were designed such that this segment was amplified in two reactions (InvPCR 6 and 7). InvPCR 7 resulted in a product in both patient and control DNA, however the product InvPCR 6 was only obtained

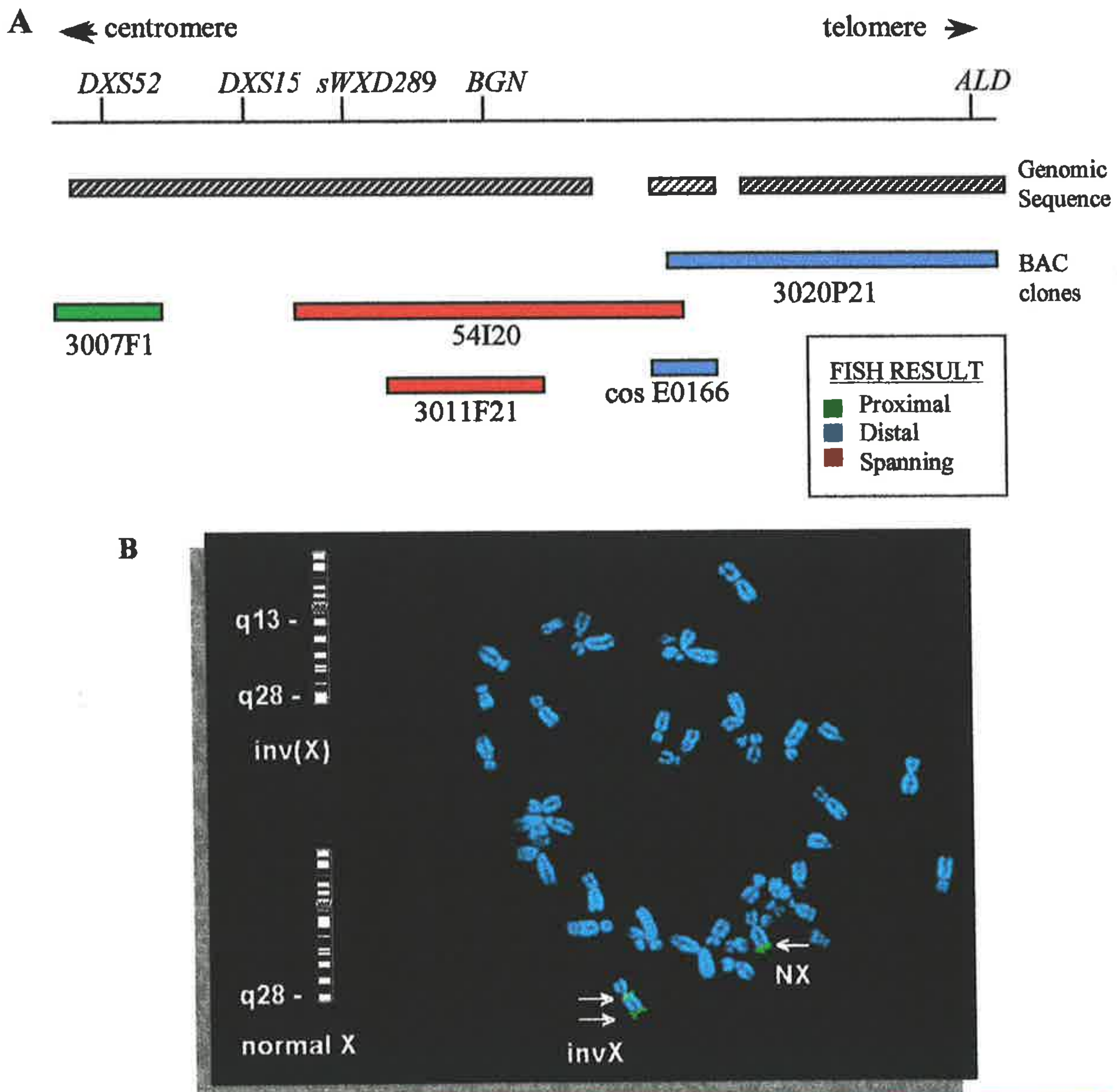


Figure 3.3: **A** Physical map of the inversion breakpoint region in Patient 1. STS markers are shown across the top, regions for which genomic sequence was available is shown is hatched boxes and genomic clones used as FISH probes are shown below (not to scale). **B** The BAC clone 3011F21 was hybridised to metaphase chromosomes of the mother of patient 1 who carries the inversion. Signal was seen on her normal X chromosome (NX) from which the probe was derived, and was also seen on in inverted X chromosome [inv(X)] on both sides of the inversion at Xq28 and also Xq13.1

from control DNA. This placed the breakpoint within the ~3 kb defined by InvPCR 6 (Figure 3.4B) within the BAC 3011F21. This ~3 kb product was hybridised to Southern blots containing *Pst*I and *Hind*III digested patient and control DNA (Figure 3.4C). In the patient, a 1 kb *Pst*I fragment was absent and was replaced by two junction fragments. These junction fragments were also seen in the mother of the patient who is a carrier of the inverted X. The sequence and the position of the restriction from this 3 kb was known and as a result the breakpoint was localised within the 3' UTR of the *BGN* gene.

Table 3.1: Oligonucleotides sequences used for LR-PCR amplification of segments of the sequence of the BAC 3011F21 (listed in the 5'→ 3' orientation). InvPCR 2 did not provide an amplification product in Patient 1, and was divided into two reactions, InvPCR 6 and InvPCR 7 (blue).

	Size	Forward	Reverse
InvPCR 1	3,346 bp	CCAGGTCCATCCGCCATGTG	AGGTGGAGGAAAGGTGGCCT
InvPCR 2	6,210 bp	AGGCCACCTTTCCTCCACCT	ACTGTCTCCCTGGCTGAGGT
InvPCR 6	3,079 bp	AGGCCACCTTTCCTCCACCT	CACCACGACCCAGAAGCCCT
InvPCR 7	3,149 bp	AGGGCTTCTGGGTCGTGGTG	ACTGTCTCCCTGGCTGAGGT
InvPCR 3	3,340 bp	ACCTCAGCCAGGGAGACAGT	CTTCCTGGCCTGGCATTGCC
InvPCR 4	3,450 bp	GGCAATGCCAGGCCAGGAAG	GGTGGCTCAGCTTCTGAGCT
InvPCR 5	4,283 bp	AGCTCAGAAGCTGAGCCACC	ATCTTCTGCCGGCCCACAGT

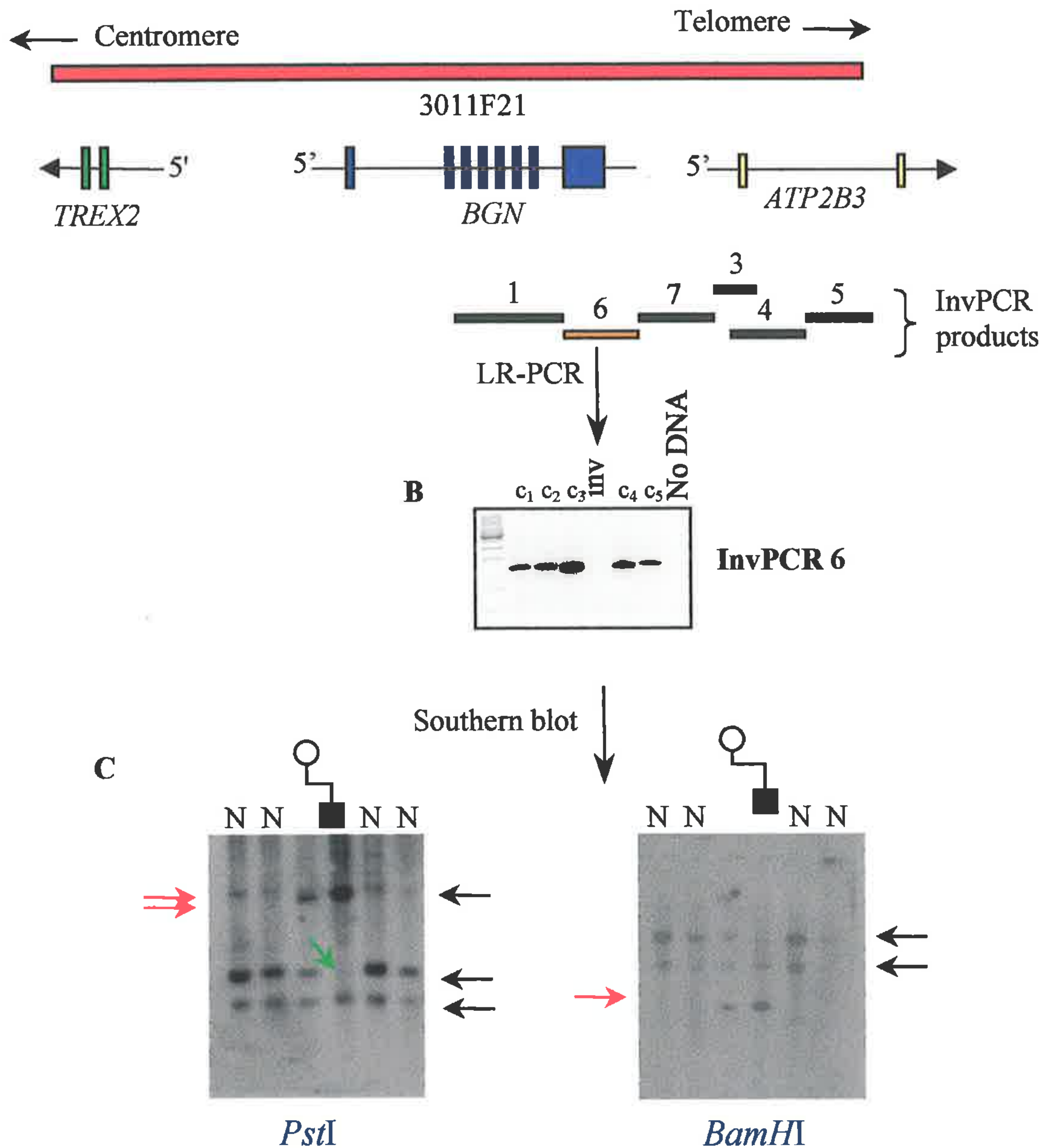


Figure 3.4: The sequence from the spanning BAC contained the *BGN* gene and the 5' end of the *ATP2B3* gene. Oligonucleotides were designed such that these regions could be screened in the Patient 1. **(A)**. LR-PCR identified a product (InvPCR 6) not obtained in the patient but obtained in normal controls **(B)**. **C** DNA from Patient 1 (black box), his mother (open circle), and four normal controls (C1-C5), digested with *Pst*I and *BamH*I and hybridised with the ~3 kb InvPCR 6 fragment. Normal bands are indicated by black arrows. The *Pst*I digest shows an ~1 kb band missing in Patient 1 (green arrow) as well as junction fragments (red arrows) appearing as a doublet at ~2.5 kb. The *BamH*I digest shows one junction fragment more clearly, however the resolution of the gel is insufficient to detect the small change in size of the other junction fragment.

To assess whether the Xq28 breakpoint affected the normal expression of *BGN*, RT-PCR was carried out using RNA from cultured fibroblasts of Patient 1 and a control as template (oligonucleotide sequences are listed in Table 3.2) (Figure 3.5). Expected products were not obtained from the patient when using primers flanking the inversion breakpoint. *BGN* specific primers 5' to the breakpoint, however, gave products of normal size which were comparable in intensity with the control samples. This suggested that in Patient 1 *BGN* is still expressed at apparently normal levels with the ORF remaining intact, however a novel 3' UTR is present.

As *BGN* was expressed in Patient 1 fibroblasts, 3'RACE was used to find the new 3'UTR of the *BGN* gene and thus to obtain the sequence across the inversion breakpoint. The reaction was carried out using the oligonucleotide e7-e8 Forward (Table 3.2). Indeed there was a new 3'UTR of approximately 650 bp (Figure 3.6A). This product was sequenced and revealed sequence homology to *BGN* up until position 1,139 bp (after the start codon) at which point the sequence matched that of a genomic sequence from Xq13.1 (accession number AL135749), at position 15,896 bp from the T7 end (Figure 3.6B). As well as this there was an extra 23 bp present between the two, which does not have significant homology to any sequence in GenBank. The sequence revealed usage of a novel cryptic polyadenlyation signal from Xq13.1.

Table 3.2: Oligonucleotide sequences used for RT-PCR analysis of the *BGN* gene. E6 was used for 3' RACE.

	Size	**Forward**	**Reverse**
e6-e7	217 bp	CTGAGACCCTGAATGAACTC	TTGAGGTCTGGGAGCCCTGA
e7-e8	1,025 bp	CCTAGGCCACAACCAGATCA	ACCAAGCTCAGAAGCGAGAA

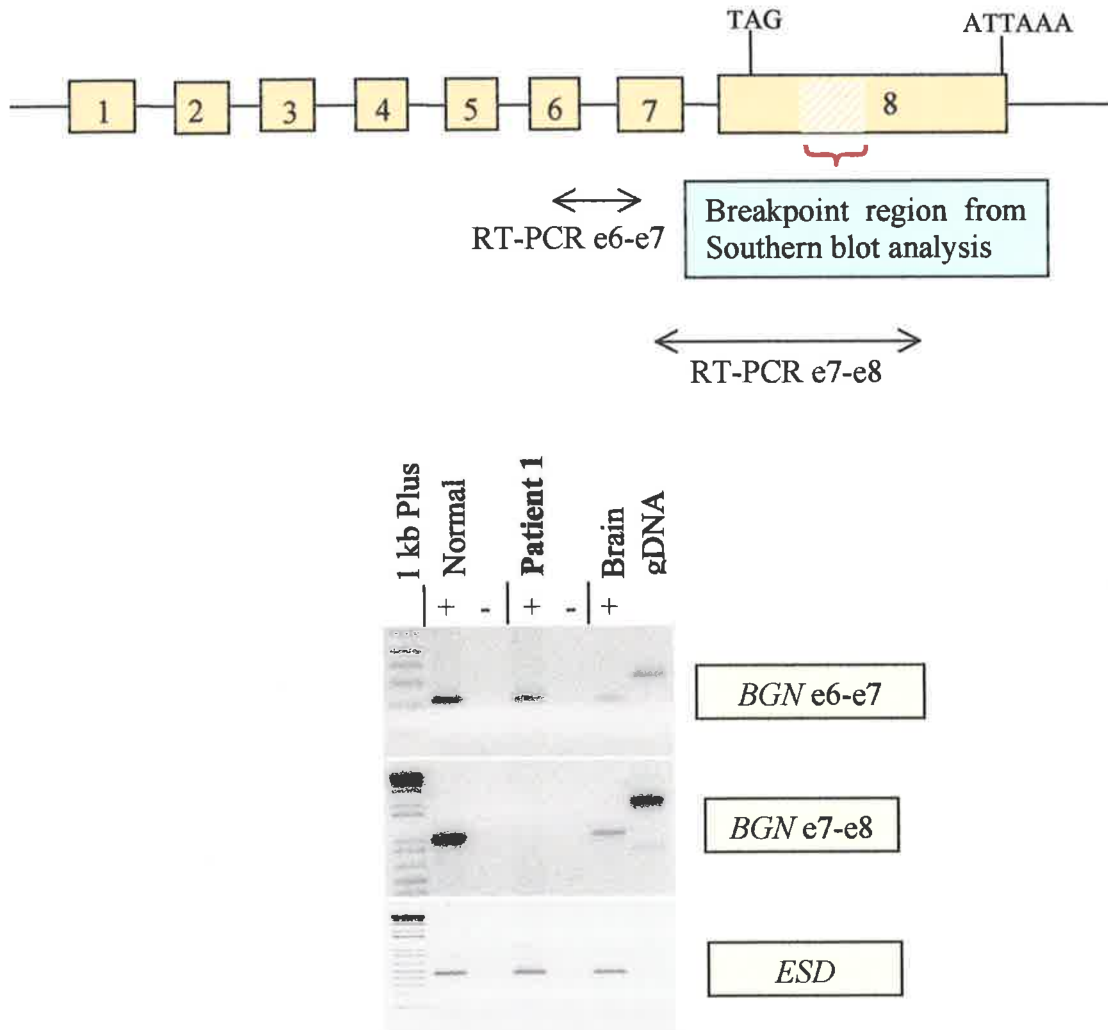


Figure 3.5: Southern blot hybridisations suggested that the inversion breakpoint lay within exon 8 of *BGN*, after the stop codon but before the polyadenlyation signal. RT-PCR was used to confirm this. Fibroblast RNA from Patient 1 and from normal control, as well as brain RNA positive control and gDNA control for genomic DNA contamination was used as template with primers from before the breakpoint and flanking the breakpoint. Although *BGN* is expressed in the patient (top panel), the breakpoint disrupts the 3' end and no product is obtained (middle panel). The bottom panel is RT-PCR using oligonucleotides to the *ESD* gene, showing that the RNA from all samples is intact.

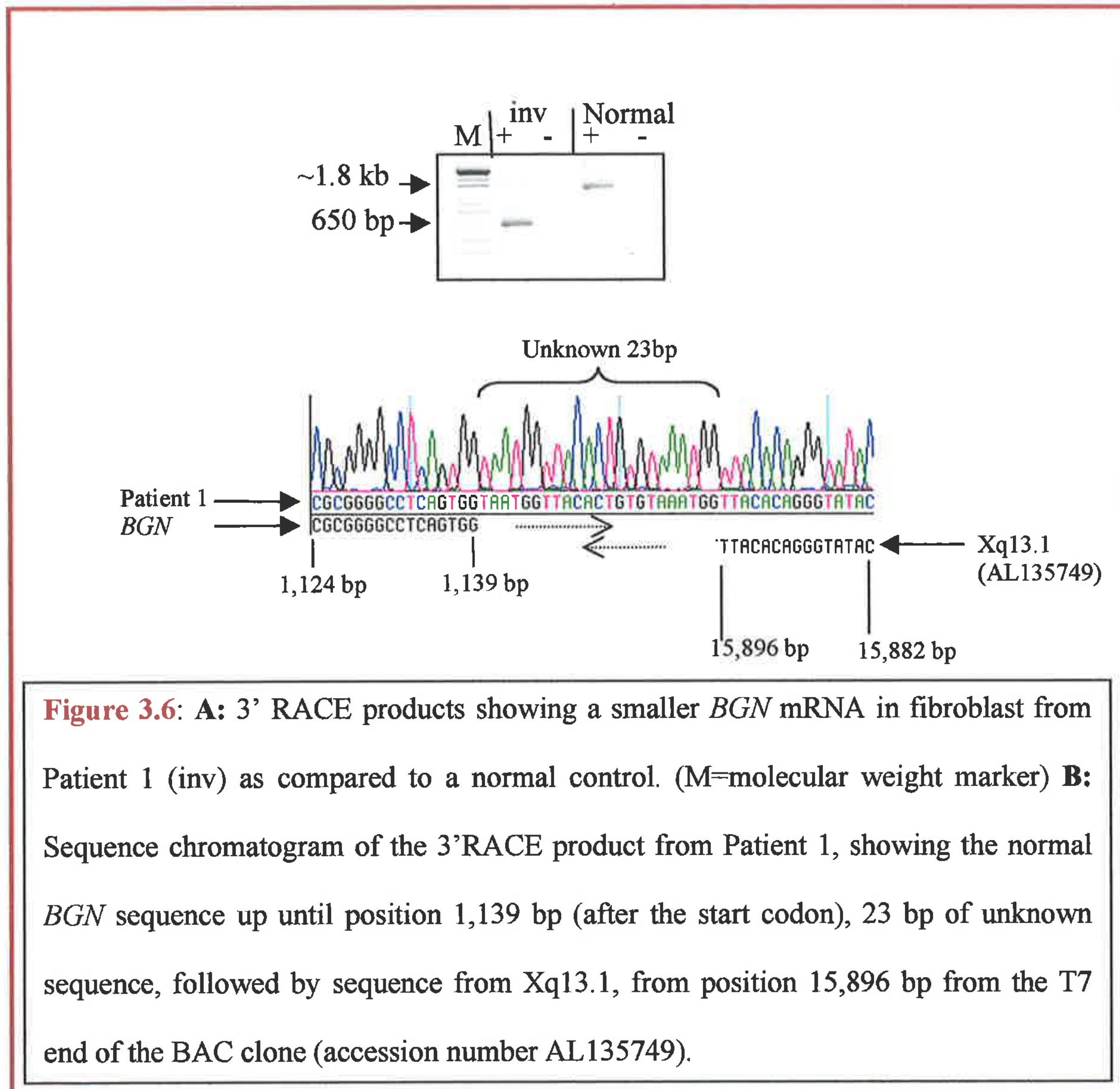


Figure 3.6: **A:** 3' RACE products showing a smaller *BGN* mRNA in fibroblast from Patient 1 (inv) as compared to a normal control. (M=molecular weight marker) **B:** Sequence chromatogram of the 3'RACE product from Patient 1, showing the normal *BGN* sequence up until position 1,139 bp (after the start codon), 23 bp of unknown sequence, followed by sequence from Xq13.1, from position 15,896 bp from the T7 end of the BAC clone (accession number AL135749).

3.3.3 The Xq13 breakpoint

The breakpoint at Xq13 is within the sequence of the BAC clone CEPHB197N14 (accession number AL135749) at position 15,896 bp from the T7 end. This is located between the NIMA-interacting peptidyl-prolyl cis/trans isomerase 4 gene (*PIN4)* or parvulin 14 (*hPar14*) (Uchida *et al.*, 1999) and a cDNA sequence which is weakly similar to the excision repair

protein *ERCC6* (accession number AK056494); however neither is physically disrupted by the inversion breakpoint.

3.3.4 Discussion

Both breakpoints of the X chromosome inversion carried by Patient 1 have been characterised at the level of the DNA sequence. No genes are physically interrupted by the breakpoint at Xq13.1, however disruption of the normal expression of the nearby genes *PIN4* and *ERCC6-like* by a positional effect cannot be ruled out. RT-PCR of these two genes was attempted using fibroblast RNA as template. However, no product was obtained even from the normal control, suggesting that neither of these genes is expressed in fibroblast, and hence their expression in Patient 1 could not be examined. At Xq28 the translocation break is within the 3'UTR of the gene *BGN* (Figure 3.7). This does not affect the ORF and a novel 3' UTR and polyadenylation signal from Xq13.1 results in apparently normal transcription of the gene.

BGN is a member of the small leucine-rich proteoglycans (SLRP) which belong to the leucine-rich repeat (LRR) superfamily of proteins (for review see Iozzo, 1998). Class 1 SLRPs (into which *BGN* falls) are characterised by 10 LRRs, a unique N-terminal cysteine sequence (CX_3CXCX_6C) and are all encoded by 8 exons. BGN is known to bind to transforming growth factor beta (TGFβ) a cytokine known to be involved in cell proliferation and differentiation (Hildebrand *et al.*, 1994). A role for BGN has also been implicated in cell adhesion (Bidanset *et al.*, 1992), and cell migration (Kinsella *et al.*, 1997). It is ubiquitously expressed with a high level of expression in bone, consistent with the fact that it has long been thought to be involved in bone growth.

A mouse knockout for *Bgn* has been produced (Xu *et al.*, 1998). These mice are apparently normal at birth, but develop an osteoporosis like phenotype with reduced bone mass and decreased growth. This study highlighted the importance of Bgn in bone growth, however effect on the development of the mouse brain has not been examined. Studies in rats, however have shown that Bgn enhances the survival of cultured neurons from the neocortex of embryonic rats (Junghans *et al.*, 1995). Additionally, when injected into the brain it can increase memory and learning in a dose dependant manner (Huston *et al.*, 2000). Therefore, although not well studied in brain, it appears that *BGN* has neurochemical effects and that disruption of this gene could cause the mild MR phenotype seen in Patient 1. Thus, BGN defects could conceivably be a cause of familial cases of mild MR.

The 3'UTRs of some genes are known to contain sequences that are involved in mRNA localisation, stability and translation efficiency (Grzybowska *et al.*, 2001). The 3'UTR of *BGN*, which is disrupted by the inversion breakpoint in Patient 1, contains a CT-rich sequence followed by a polymorphic $(GT)_n$ repeat. Searching the 3' UTRs of orthologs of *BGN* using the Repeat Masker web server (http://ftp.genome.washington.edu/cgi-bin/RepeatMasker) shows that repeats are also present in some other species. In rat a CT-rich sequence is followed by a $(CA)_n$ repeat, while in mouse the $(AC)_n$ repeat is preceded by a $(TCTCTG)_n$ CT-rich simple repeat. In horse a CT-rich sequence is followed by an AT-rich sequence. However, the 3' UTR of cow and dog *Bgn*, do not contain any repeats. 3' UTRs of many genes contain microsatellite repeats, however it is not known if they play any role in gene regulation.

Normal X

Xq13.1

Xq28

PIN4

3'

ERCC6- like

TREX2

5'

BGN

ATP2B3

Breakpoint

Breakpoint

Inverted X

Xq13.1 breakpoint

Xq28 breakpoint

PIN4

5'

BGN

TREX2

3'

ERCC-like

3

BGN

ATP2B3

New 3' end
From Xq13

Figure 3.7: Diagramatic representation of the inversion in Patient 1. The breakpoint at Xq28 is within the 3' UTR of the *BGN* gene. The 5' end of *BGN* is relocated to Xq13.1 and gains a new polyadenylation signal from this side of the breakpoint. The breakpoint at Xq13.1 is located downstream of the gene *PIN4*, however no gene at this location is physically disrupted by the break.

It is therefore possible that disruption of regulatory sequences within the 3' UTR effects BGN in this patient at the post-transcription level. It would be of interest to test the level of BGN protein produced in Patient 1 compared to normal controls to see if this is the case. As the mother of the patient is a carrier of the inverted chromosome there may be other males in the family that have this X chromosome inversion. The mother reports a history of psychiatric problems among her relatives however material from these family members is unavailable for further study.

3.4. RESULTS PATIENT 2: 46,Y,inv(X)(q13.1q28)

3.4.1 Xp11.2 breakpoint

BAC clones from Xp11.2 were used as probes for FISH and the clones 966K21 and 1158B12 were shown to span the breakpoint (Figure 3.8) (J. Gècz unpublished results). The sequence contained within these BACs is highly repetitive in nature and BLAST searches of GenBank revealed that it contains only 1 gene, *ZXDA*, which is an intronless zinc finger gene. Located ~30 kb telomeric is *ZXDB* which shares 98.7% DNA sequence identity to *ZXDA* (Greig *et al.*, 1993). To determine if *ZXDA* is disrupted by the breakpoint RT-PCR was used. As primers for *ZXDA* also amplify the same sized product from *ZXDB*, the products were analysed by SSCA on non-denaturing polyacrylamide gels (Figure 3.8B). In Patient 2 the expression of both *ZXDA* and *ZXDB* were as normal, and thus the inversion breakpoint does not significantly affect expression of either of these genes.

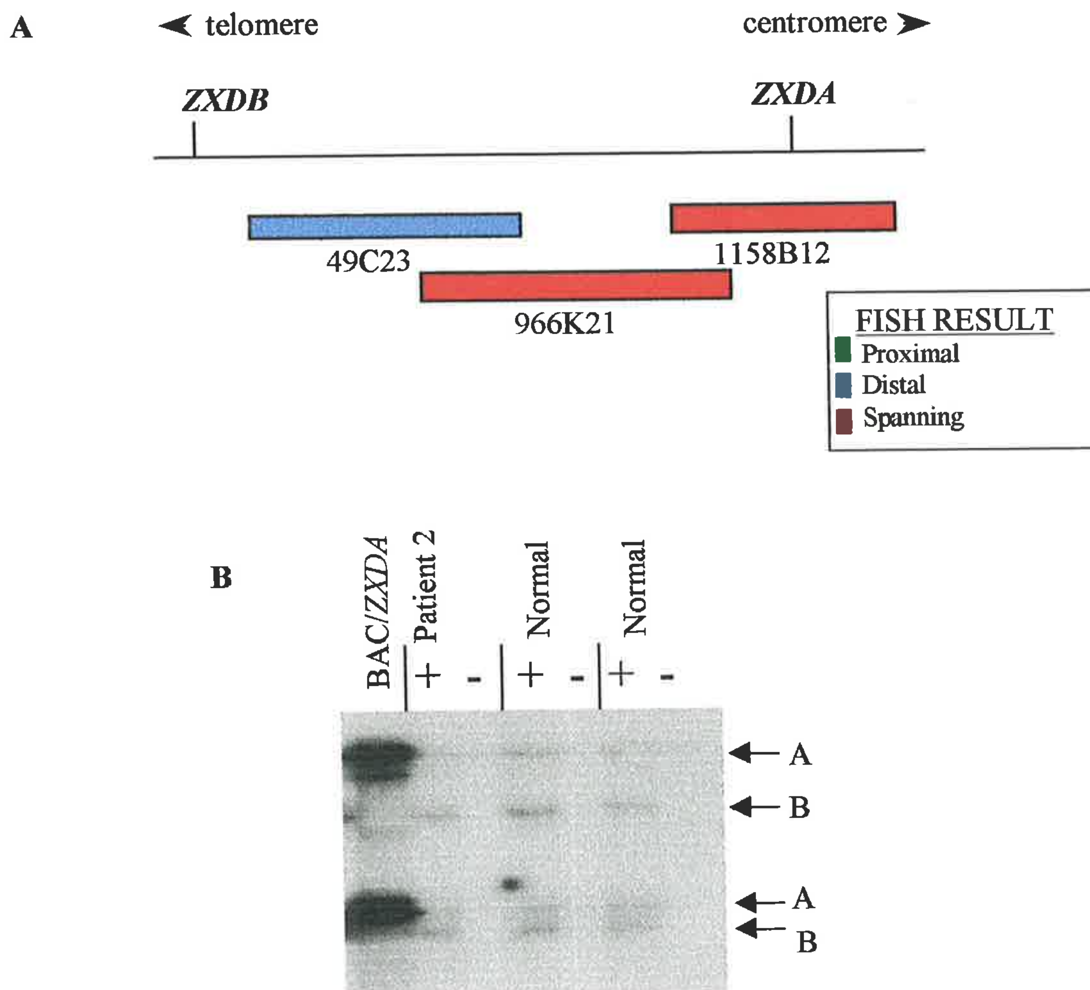


Figure 3.8: **A** BAC contig from Xp11.2 showing the BAC clones giving spanning signal when used as probes for FISH. The only gene within these spanning BACs is *ZXDA* which also has a duplicated copy, *ZXDB* that is located telomeric to the breakpoint. **B** SSCA analysis of the *ZXDA* and *ZXDB* genes from Xp11.2. The BAC clone 1158B12 which contained *ZXDA* (but not *ZXDB*) was amplified so that it was possible to tell the difference between *ZXDA* and *ZXDB* in human samples. Both genes were amplified with the same oligonucleotides (due to the high degree of sequence similarity between the two). Template for PCR was RNA from cultured fibroblasts of both Patient 2 and two normal controls as template. (+ denotes RT reaction with reverse transcriptase added; - denotes RT reaction with no reverse transcriptase added; A and B indicate the bands corresponding to *ZXDA* and *ZXDB* respectively). The oligonucleotide sequences are as follows F: CTC TTA CAA GCT CAA GAG GC and R: ACA TGA ACC TCC GGT CAT CG (written in the 5'→3' orientation)

3.4.2 Xq28 FISH and Interphase FISH

Initially the cosmid clone cos E0166 from Xq28 was used as a probe for hybridisation to metaphase chromosomes of Patient 2 and was shown to span the breakpoint (unpublished results). As sequence for this region was unavailable at the time a working draft sequence of this cosmid was produced (by Gabriel Kremmidiotis and Alison Gardener, WCH, Adelaide, Australia). Searches of the provided sequence against GenBank database using the nr and dbEST divisions of BLAST revealed the presence of two genes within this sequence. These are the mitogen activated protein kinase (*MKP4* or *DUSP9*) and two overlapping ESTs (accession numbers AI422876 and AI424701) which are now part of the UniGene cluster Hs.161358 for an uncharacterised gene. LR-PCR and Southern blotting were used in an attempt to identify the location of the inversion breakpoint within this sequence and to determine if either of these genes were disrupted. However, these techniques failed to detect any region which would contain an inversion breakpoint. This led to further FISH analysis using additional BAC clones from Xq28 (the same as those used for FISH for Patient 1) (Figure 3.1). All BAC clones used were shown to span the inversion breakpoint at Xq28.

There was no obvious reason to speculate that these results were due to normally duplicated regions in Xq28 or low copy X chromosome repeats which might result in multiple signals by FISH, as these clones had been used successfully for Patient 1. Rather these results suggested that the inverted chromosome also contained a duplication of Xq28. The BAC clones used cover a region of at least 2.7 Mb of Xq28. This duplication was further confirmed using interphase FISH (Figure 3.9), where two BAC probes located greater than 160 kb apart were hybridised to interphase nuclei of the Patient 2. From 100 nuclei scored, 92 contained two signals from each probe. This was consistent with an Xq28 submicroscopic duplication which

A

centromere
telomere

DXS52 *DXS15* *sWXD289* *BGN* *ALD*

Genomic Sequence

3020P21
3007F1
54I20
3011F21
cosE0166

FISH RESULT

Proximal
Distal
Spanning

TREX2 *BGN* *ATP2B3*

B

BAC3007F1
BAC54I20

Figure 3.9: **A** All of the BAC clones used as probes for FISH from Xq28 gave spanning signal on the inverted X chromosome in Patient 2 (not to scale). **B** Interphase FISH of nuclei of Patient 2 hybridised with BAC clones 3007F1 (labelled with red fluorescence) and 54I20 (labelled with green fluorescence). Nuclei contain two signals from each probe confirming that Patient 2 has a duplication of Xq28, with a copy located on each side of the inversion.

was previously undetected and hence the karyotype of this patient was redefined as 46, Y, rec(X)dup(Xq)inv(X)(p11q28).

3.4.3 Discussion

For Patient 2, both inversion breakpoints have been studied. At Xp11.2 the breakpoint lies within a region of DNA close to the centromere that is highly repetitive in nature. Searches of GenBank revealed that this region is also gene poor. The expression of the only gene in the region, *ZXDA*, is not disrupted by the inversion in this patient, either physically or in its expression.

At Xq28 this study has revealed that in addition to the inversion, there is a duplication of a large segment of Xq28, hence the rearrangement has proven to be more complex than initially thought (Figure 3.10). No BAC clones were identified that when used as probes for FISH gave only distal or proximal signal. This means that the outer limits of the duplication are yet to be determined. The duplication is, however, greater than 2.7 Mb, and encompasses a region containing more than 30 known genes.

Disomy of clusters of genes within Xq28 genes has previously been described in boys with XYX_q syndrome (Lahn *et al*, 1994). This syndrome was characterised by a study of boys with a 46XYq- karyotype, but who had additional clinical features not shared by most of the affected males with this karyotype. It was revealed that in these boys the deleted Y chromosome also contained some Xq28 sequence. Disomy of Xq28 genes was shown to cause phenotypes other than that observed in boys with the 46XYq- karyotype alone, including severe mental retardation, hypotonia and microcephaly. In these cases, and in the

case of patient 2 described here, the phenotypes are likely to be the result of abnormal gene dosage that would normally be compensated for by X-inactivation in females. The region duplicated in Patient 2 overlaps with that of the patients with XYX_q syndrome, a gene rich region containing more than 30 genes.

A female patient with MR has recently been reported where duplication of Xq28 genes was also concluded to be the cause of MR (Cox *et al.*, 2003). In this case a translocation with one break in Xq28 was initially discovered in the patient; however, molecular characterisation of the breakpoint showed that there was duplication of 650 kb of Xq28. The duplication in this patient is located ~ 500 kb telomeric to the most telomeric clone used as a probe for FISH for Patient 2 (PAC clone 3020P21). As the outer limits of the duplication of Patient 2 remain unknown it is possible that these two regions of duplication overlap.

Thus it is likely that the intellectual disability in Patient 2 is due to this large duplication and functional disomy of many genes, rather than disruption of any one gene in Xq28 or Xp11. This shows that the use of positional cloning to characterise the inversion breakpoints in Patient 2 with the aim of identifying candidate genes for XLMR is not viable, as the MR in Patient 2 is not the result of a single gene defect. This does however show that the duplication of genes on the X chromosome can be a cause of XLMR, and these duplications may not always be visible at the level of light microscopy. If they do not occur with another accompanying X chromosome rearrangement such as the inversion in Patient 2, such duplications would remain undetected by light microscopy. Current PCR based techniques for mutation detection will not detect these mutations and a high throughput method of screening genes base on the detection of normal levels of transcription is required.

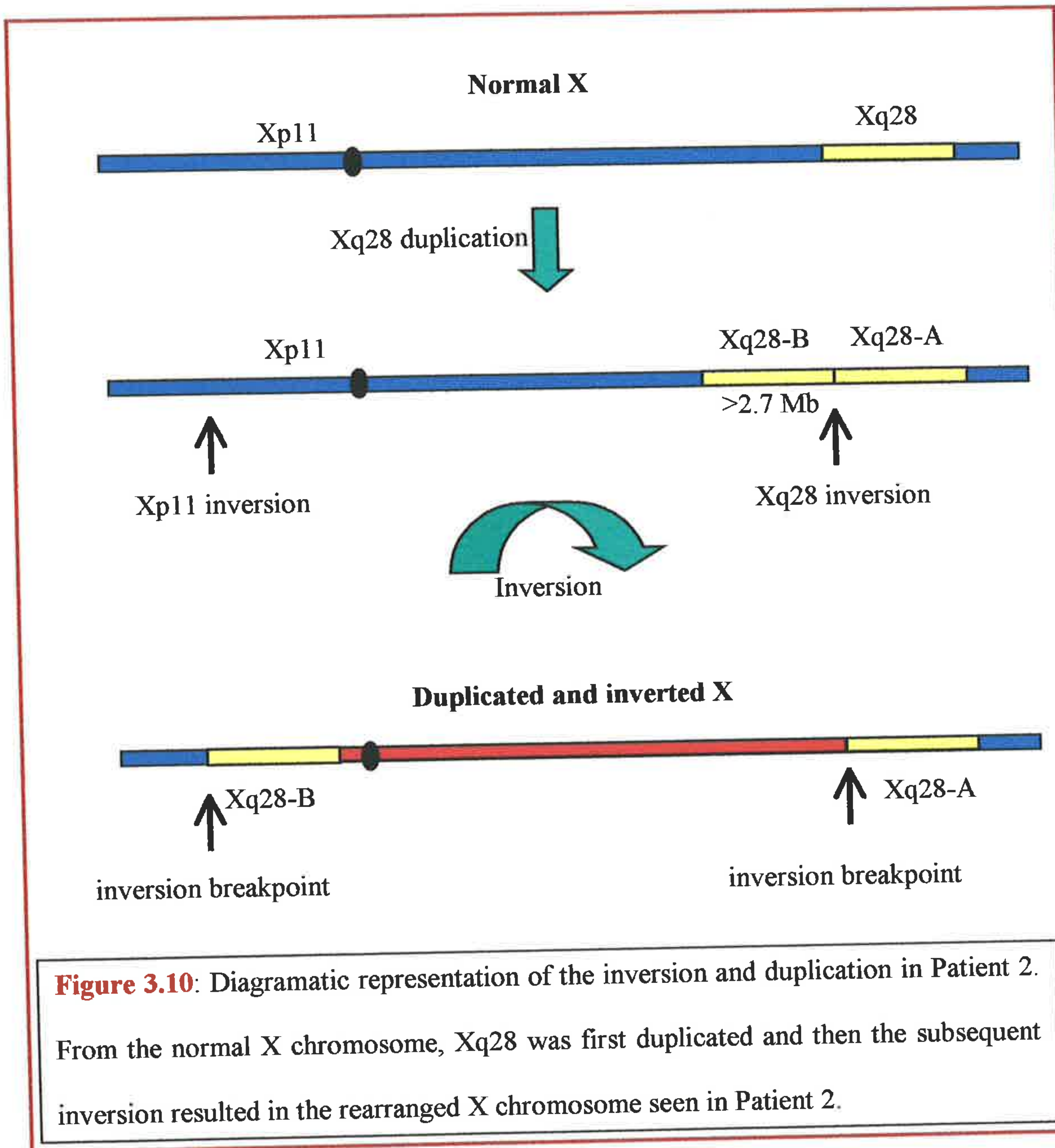


Figure 3.10: Diagramatic representation of the inversion and duplication in Patient 2. From the normal X chromosome, Xq28 was first duplicated and then the subsequent inversion resulted in the rearranged X chromosome seen in Patient 2.

3.5. RESULTS PATIENT 3: t(X;10)(q28;q11.2)

3.5.1 Xq28 FISH

BAC clones from Xq28 (from the contig shown in Figure 3.2) were used as probes for FISH against metaphase chromosomes of Patient 3. The BAC clone 3193D19 gave proximal signal, while the BAC 45D17 gave distal signal. These two clones were separated by a gap of approximately 65 kb. The clones 2532F24 and 2514I12 were then used for FISH to try to cover this gap in the contig, and gave distal and spanning signal (Figure 3.11) respectively. The end of the spanning BAC 2514I12 lay approximately 10 kb away from the end of 3193D19. At the time the sequence across this region was not complete. The sequence AF003627 accounted for some of the spanning BAC 2514I12, however the sequence from the other end (the centromeric end) was only in working draft form (BAC clone 45D17, accession number AF274854), and the exact length of the BAC was unknown. The finishing of this sequence during the course of the project has revealed that 2514I12 is ~75 kb in length.

BLAST searches using the sequence (AF003627 that contained most of the sequence of the BAC 2514I12 as well as ~117 kb of sequence centromeric to the BAC) did not reveal any obvious candidate genes for either the MR or the cardiomyopathy seen in Patient 3. In fact this region proved to be a gene poor region. Only two ESTs from testis cDNA libraries were within the sequence in the region of the breakpoint. These sequences were similar to a region of X chromosome low copy repeats and it was unclear as to whether these ESTs represented a real gene. The overlapping working draft sequence AF274854 (BAC clone 45D17) was then searched for the presence of genes.

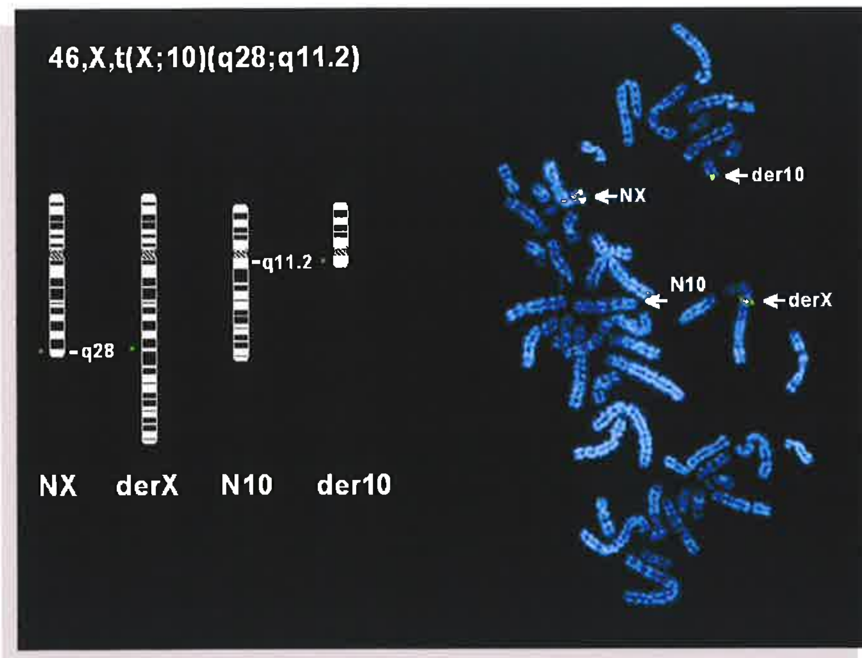


Figure 3.11: The BAC clone 2514I12 was used as a probe for FISH against the chromosomes of Patient 3 and showed signal on the normal X chromosome as well as on the derivative X and derivative chromosome 10, indicating that this clone contained the translocation breakpoint.

3.5.2 The gene TMG3a is disrupted by the Xq28 translocation breakpoint.

The BLAST searches of the genomic sequence AF274854 (BAC clone 45D17) revealed the presence of several significant matches. These are described below and Figure 3.12

1. Firstly, a small region of similarity was found to the rabbit ortholog of the cyclic nucleotide gated channel (*Cng2)* gene (accession number X59668). This region of similarity is in fact not part of the human *CNG2* gene, which is located ~40 kb telomeric to the region of similarity, within the genomic sequence AF002992. Nor do any orthologs from other species share this extended 5' end of the *CNG2* gene. An attempt was made to link the two by RT-PCR using fibroblast RNA as template, however no product was obtained. It has been noted that this extra 5' end from the rabbit gene has similarity to Vitamin K-dependent carboxylation/γ-carboxyglutamic acid (GLA) domain proteins but that this domain is out of frame with the rest of the rabbit *Cng2* (Kulman *et al.*, 1997). It is therefore likely that the extra 5' sequence in the rabbit gene is an artefact.

2. Searches of the dbEST division of GenBank revealed human ESTs that corresponded to the extra 5' sequence in the rabbit *Cng2* gene, suggesting that, even though this was not part of *CNG2*, it was part of another gene. These ESTs belonged to the Unigene cluster Hs.209253.

3. During the course of this project, the gene *TMG3* was reported (Kulman *et al.*, 2001). The gene was identified by this group by a search of sequences within GenBank that contained the GLA domain. The sequence provided for this gene (accession number

NM_024082) was based on the UniGene cluster Hs.209253 (from 2 above). The predicted mRNA from alignment of these ESTs, however, resulted in a transcript that lacked a polyadenlyation signal and is probably not the complete mRNA sequence.

4. Located telomeric to the UniGene cluster Hs.209253, is a second UniGene cluster, Hs.356692. Hs.356692 was originally joined with the ESTs for *dynein* (*DNCH1*) located on chromosome 4. The ESTs from Xq28 were placed in Hs.356692 due to the presence of several ESTs (all from the same cDNA library) that were chimeric clones, with sequence of *DNCH1* at one end and sequence matching the gene from Xq28 at the other end. The two groups of ESTs were separated and ESTs from Xq28 were aligned using the SeqMan sequence alignment software (DNASTAR). This alignment revealed that the ESTs represented in Hs.356692 were from the 3' end of a transcript, as most were primed from the polyA tail. Additionally there was a predicted polyadenylation signal located close to the 3' end. One of the ESTs from this cluster (accession number R89637) appeared to be derived from an alternatively spliced mRNA. Additionally the 5' end of R89637 overlapped with the consensus sequence derived from the ESTs from the UniGene cluster Hs.209253 (from 3 above).

5. Two cDNA sequences in GenBank, one monkey and one mouse cDNA (accession numbers AB046660 and AK079502 respectively) overlapped with the ESTs from the UniGene cluster Hs. 356692 (from 4 above).

6. When the finished genomic sequence of the region became available, two other possible groups of ESTs appeared. The first group consisted of only two ESTs (accession numbers

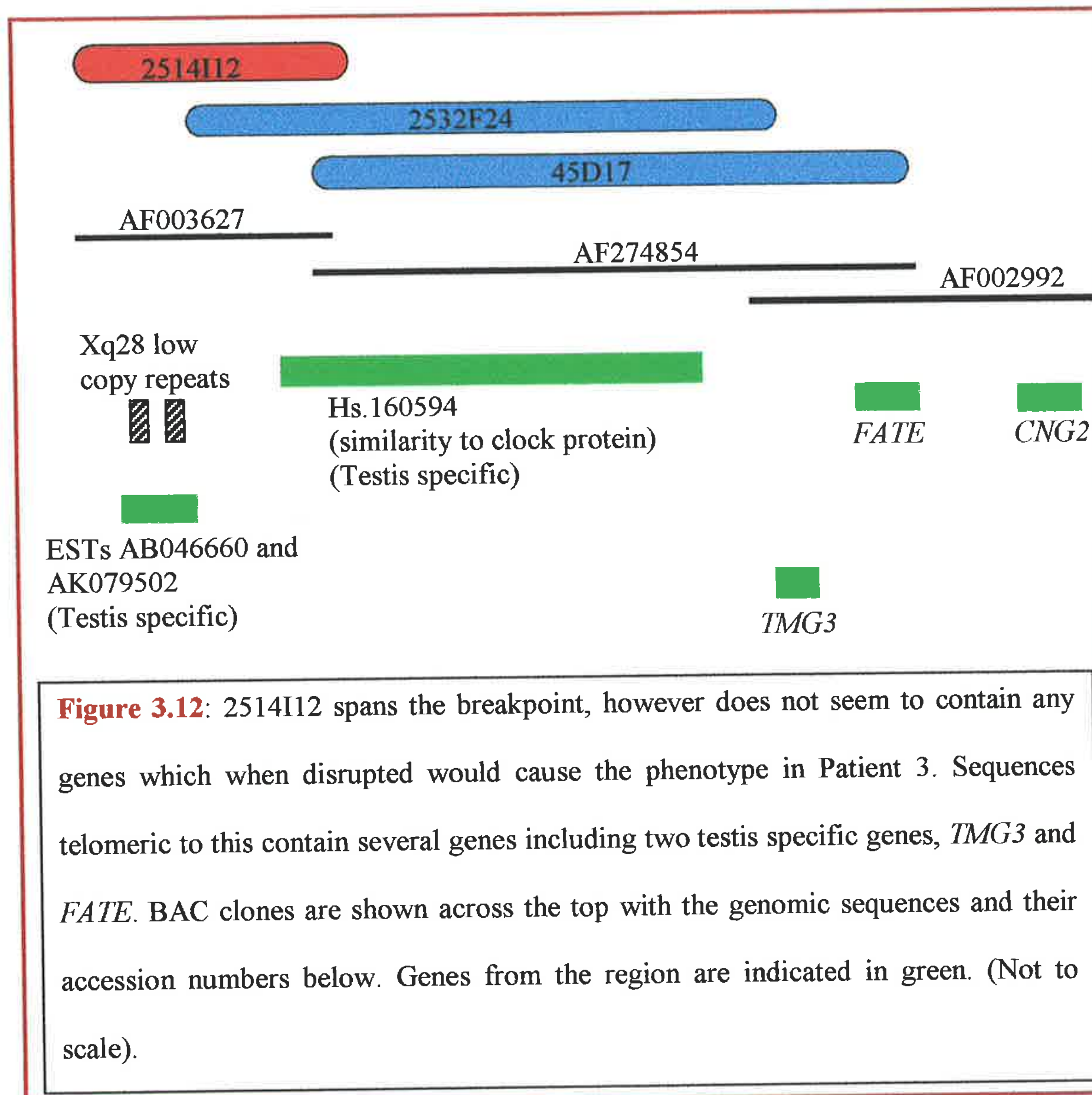


Figure 3.12: 2514I12 spans the breakpoint, however does not seem to contain any genes which when disrupted would cause the phenotype in Patient 3. Sequences telomeric to this contain several genes including two testis specific genes, *TMG3* and *FATE*. BAC clones are shown across the top with the genomic sequences and their accession numbers below. Genes from the region are indicated in green. (Not to scale).

BC044778 and BG724103) around the region of an Xq28-specific low copy repeat. This is the only gene that could be physically disrupted by the translocation breakpoint in Patient 3. LOC139135, a predicted gene based on the sequence of ESTs from the UniGene cluster Hs.160594 was also present. The predicted protein has homology to circadian locomoter output cycle kaput (clock) protein from mouse. All ESTs from both of these groups were from testis cDNA libraries. Hs.160594 and the two ESTs

are ~20 kb apart and it is possible that they could be part of the same gene. RT-PCR on a multiple tissue RNA panel (Clontech) confirmed that this gene/ these two genes are indeed only expressed in testis, and not in the other tissues tested which included brain and heart (results not shown).

The sequences in GenBank described in points 2→5 above all appear to originate from the same gene (and its orthologs in mouse and macaque monkey) with the overlap based on only one EST (R89367) from Unigene cluster Hs.356692. RT-PCR using primers from the consensus sequences of alignments of the ESTs from Hs.209253 and Hs.356692 and using fibroblast RNA as template was carried out, in order to confirm that these two groups were part of the same gene (oligonucleotides 2F and Hs.356692 R listed in Table 3.3). A product was obtained which when sequenced was the same as the alternatively spliced form predicted by R89367. No other products were obtained.

The analysis of the ESTs is consistent with the presence of a gene containing at least three exons, which overlaps with but is longer than the published sequence for *TMG3*, and will be referred to as *TMG3a*. This includes a small exon 1, which contains the start of the predicted ORF, exon 2 and at least part of exon 3. These three exons overlap with the rabbit *Cng2* and the UniGene cluster Hs.209253. The GLA domain is encoded by the sequence of exon 2.

Other GLA domain containing proteins include coagulation factors (for example *factor IX*, mutations in which cause haemophilia B, OMIM #306900), the bone related protein osteocalcin *(BGLAP)* (Celeste *et al.*, 1986) and the growth arrest protein *GAS6* (Manfioletti *et al.*, 1993). In invertebrates a GLA domain was found in conotoxins, a neurotoxin from the marine cone snail (McIntosh *et al.*, 1984). These proteins all contain γ-carboxyglutamic acid

(GLA) residues that are produced by post translational modification of glutamic acid residues by a vitamin K-dependent carboxylase (Stenflo *et al.*, 1974). The GLA domain is made up of approximately 48 amino acids which contains between 9 and 13 carboxylated glutamic acid residues. Calcium ions bind to these residues allowing the vitamin K-dependent proteins to interact with membranes in a calcium-dependent manner that has not been fully elucidated (Mann *et al.*, 1990).

As this was the only gene in the vicinity of the translocation breakpoint in Xq28 which presented as a candidate for both MR and cardiomyopathy, RT-PCR between exons two and three was carried out (using oligonucleotides 2F and 3R from Table 3.3) to determine if expression of this gene was effected in Patient 3. This revealed that although this gene was expressed in fibroblasts from a normal control, it was not expressed in the fibroblasts of Patient 3 (Figure 3.13). Therefore the expression of this gene appeared to be silenced by the translocation.

Table 3.3: Oligonucleotide sequences used for RT-PCR analysis of the *TMG3a* gene.

	Size	**Forward**	**Reverse**
2F + 3R	210 bp	AGGATGCCCATTCGGTCCTG	CAGGACCGAATGGGCATCCT
2F + Hs.356692 R	426 bp	AGGATGCCCATTCGGTCCTG	GACTCGCTGCCCACAGTCTG

To determine the tissue expression patterns of *TMG3a*, and to determine the size of the full length transcript, the RT-PCR product between exons 2 and 3 was used as a probe for hybridisation to a human multiple tissue Northern blot (Clontech). This probe, however, contains a region that has a high degree of similarity to the gene for coagulation factor II (*F2*)

which is located on chromosome 11 (with 87% similarity to *F2* over 66 bp) (Figure 3.14). This probe was therefore likely to also detect *F2* mRNA for which a 2 kb liver specific band was expected. This region contains the GLA domain, and also has lesser similarity to other GLA containing proteins; protein C (*PROC*) from chromosome 2 (mRNA of 1.8 kb) as well as coagulation factor VII (*F7*) from chromosome 13 (2.5 kb mRNA). However it was unlikely that the probe would detect these two mRNAs. As this was the only region consistently present in a large proportion of the sequences used to construct *TMG3a* sequence, it was deemed the only region to use as a reliable probe to detect a *TMG3a* transcript.

The Northern blot did indeed result in a band of 2 kb specific to liver, which was due to cross hybridisation to *F2*. There are no other bands present that would correspond to either *PROC* or *F7* mRNAs. Apart from the *F2* mRNA there are also faint bands at approximately 6 kb, which were detected in heart and brain, as well as an approximately 4.5 kb transcript which was barely visible in heart, liver, skeletal muscle, kidney and pancreas (Figure 3.13). Both of these bands were faint and were only detected after a long exposure to film. It is possible that low levels of expression are undetected in the other lanes of the blot. One or both of the 6 kb and 4.5 kb bands are likely to correspond to *TMG3a*. Using the same probe a human brain Northern blot containing mRNA from different regions of the brain (Clontech) was hybridised (Figure 3.13). The ~6 kb transcript was detected in cerebral cortex, medulla, spinal cord, frontal lobe, temporal lobe and putamen. The ~4.5 kb transcript is also expressed at a lower level, although due to the technical quality of the blot it is unclear as to which parts of the brain this is expressed in apart from cerebellum, cerebral cortex and putamen. RT-PCR from *TMG3* (Kulman *et al.*, 2001) showed expression in adult brain, lung, kidney, heart, skeletal muscle, pancreas and plancenta.

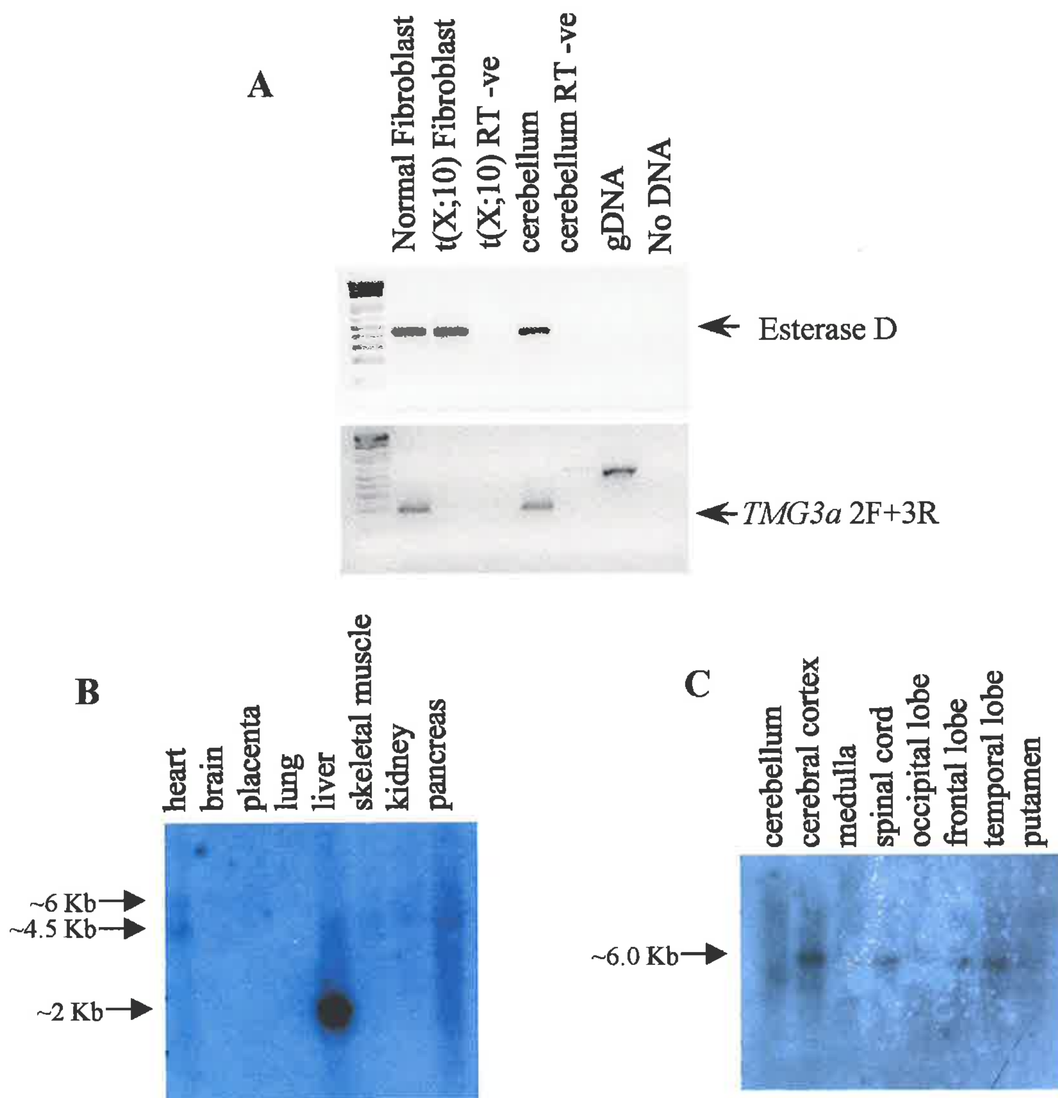


Figure 3.13: **A**: RT-PCR using primers designed within exons 2 and 3 of *TMG3* using RNA from cultured fibroblast of Patient 3 and control as template. This gene appears to be disrupted by the translocation breakpoint in Patient 3. **B** and **C**: Northern blot hybridisation of the product from **A** to a human multiple tissue Northern blot (**B**) and human brain northern blot (**C**) (Clontech). The probe has similarity to *F2* a 2 kb liver specific gene, and hence the band in B in liver represents the *F2* transcript. Faint signal of 4.5 kb (heart, liver, skeletal muscle, kidney and pancreas) and 6 kb (heart and brain) are also detected. In C the 6 kb band is predominant in several parts of the brain while the 4.5 kb band barely detectable.

```
TMG3 probe G T T C C T G G A G G A G C T G C G C C A G G G C A C C A T C G A G C G A G A G
F2         - - - - - T G G A G G A G G T G C G C A A G G G C A A C C T A G A G C G A G A G
F7         G T T C C T G G A G G A G C T G C G G C C G G G C T C C C T G G A G A G G G A G
PROC       - T T C C T G G A G G A G C T C C G T C A C A G C A G C C T G G A G C G G G A G

TMG3 probe T G C A T G G A G G A G A T C T G C A G C T A C G A G G A G G
F2         T G C G T G G A G G A G A C G T G C A G C T A C G A G G A G G
F7         T G C A A G G A G G A G
PROC       T G C A T A G A G G A G A T C T G T G A C T T C G A G G A G G C C A A G G A A
```

Figure 3.14: DNA sequence alignment of part of the probe used for Northern Blot hybridisation that shows similarity to other GLA domain containing proteins. Shading indicates sequence differing from *TMG3a*. The boxed region shows the sequence from the probe that is also likely to cross hybridise to F2. Although there is high degree of similarity with the other sequences shown, there are no large stretches of identical sequence which would cross hybridise.

The information gathered from BLAST searches that was used to make the mRNA sequence for *TMG3a* was reviewed and is summarised in Figure 3.15. It is possible that the region connecting the two UniGene clusters, Hs.209253 and Hs.356692, does not contain any introns and is in fact part of one larger exon (Figure 3.15). This sequence contains several repetitive sequences, and hence this may be the reason for a lack of representation in the sequenced ESTs. The ESTs from the 3' end of the published *TMG3* sequence are adjacent to an adenine rich sequence and it is likely that these are due to mispriming of the oligo(dT) oligonucleotides from here. If the region between the two UniGene clusters is a large intronless final exon then a predicted transcript of 5.6 kb can be derived from the ESTs. The RT-PCR performed originally to link the two UniGene clusters, would not have revealed this

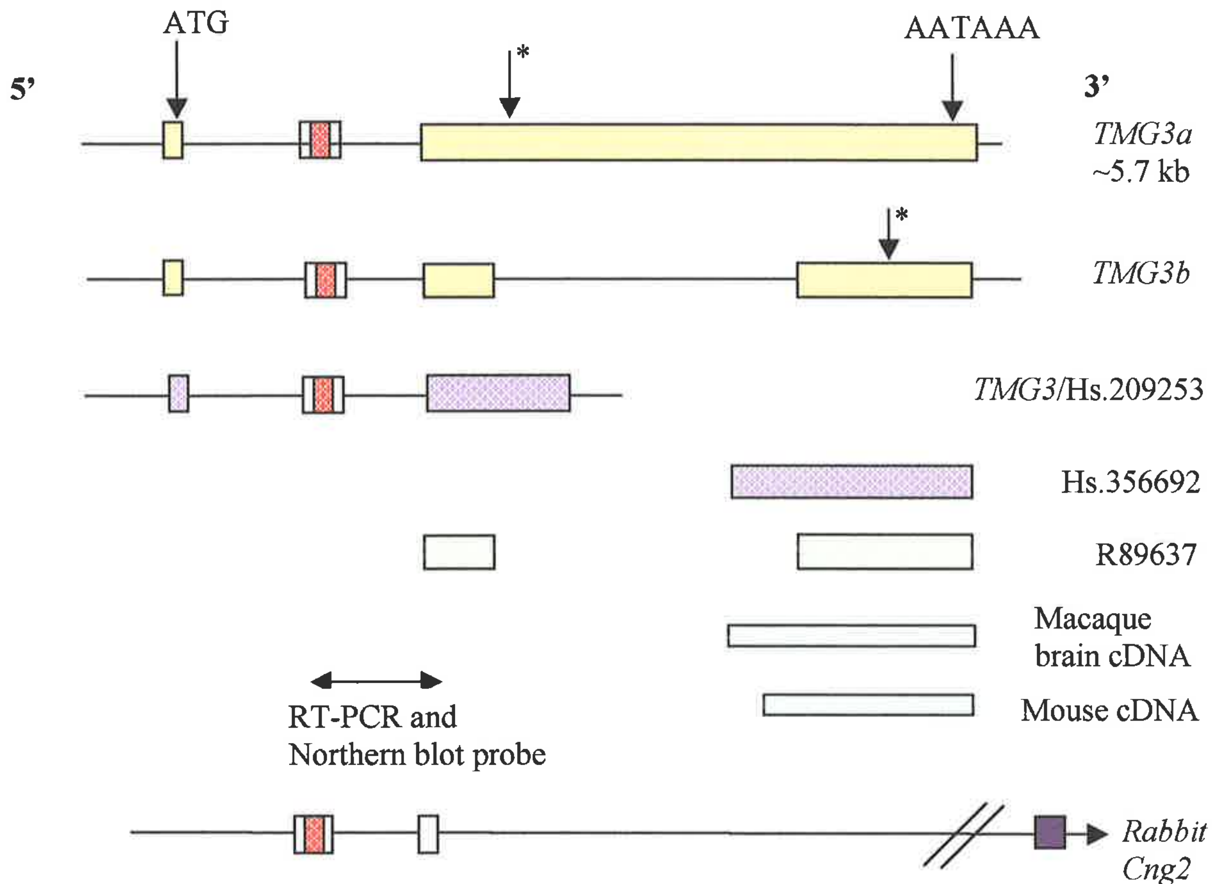


Figure 3.15: The gene *TMG3a* is in the region of the translocation breakpoint. The transcript was constructed based on two UniGene clusters (purple). The linking of these two resulted in *TMG3a*, a transcript of ~5.7 kb, (top). One EST (R89637 shown in green) suggested there was a second isoform (*TMG3b*) which is predicted to use the same translation start site, but has an alternative carboxy terminus which uses an alternative stop codon (*). The sequence of *TMG3* from GenBank is a truncated version of the gene that lacks most of the 3' UTR and polyadenylation signal. cDNA sequences from other species with similarity to the predicted transcript are in blue. The Rabbit *Cng*2 sequence overlaps with this gene for all of exon 2 and part of exon 3, however this is likely to be an artefact. The location of human *CNG2* is ~ 40 kb telomeric to this region. The GLA domain is in red. An arrow indicates the region consistent with all of these, and shows the region tested by RT-PCR in Patient 3 (Figure 14).

large isoform as the expected amplification product would be ~5 kb, and would have been too large for RT-PCR amplification.

3.5.3 Mutation Screening of the Proline-rich GLA protein gene, PRRG1, located in Xp21.1

The characterisation the breakpoint in Patient 3 led to the hypothesis that other genes containing a GLA domain residing on the X chromosome would be candidates for XLMR. One gene containing a GLA domain, which is also located on the X chromosome at Xp21.1, is *PRRG1* (Kulman *et al.*, 1997). Like *TMG3a* this gene also lacks the signal peptide characteristic of other GLA domain containing proteins. *PRRG*1 has a broad tissue expression including brain. The similarity between these two proteins suggests that they may have a similar function, and thus because *TMG3a* has been identified as a candidate gene for NSXLMR, *PRRG1* is also a candidate. The minimal linkage interval of several MRX families spans Xp21.1 in which *PRRG1* resides, and therefore it was a candidate gene in which to look for mutations in these families.

The genomic structure of this gene was determined by comparison of the mRNA sequence (accession number NM_000950) and the genomic sequence from this region (accession number AL356858). The intron/exon boundaries are listed in Table 3.2. The ORF of *PRRG1* (which starts in exon 3) was screened using the oligonucleotides listed in Appendix 1. Eight families mapping to this region of the chromosome were screened. These were MRX9, MRX10, MRX11, MRX12, MRX18, (Willems *et al.*, 1993; Kerr *et al.*, 1992; Gedeon *et al.*, 1994), two unpublished NSXLMR families, and one unpublished family with a syndrome of MR and macrocephaly (linkage analysis performed by Dr A. Gedeon, WCH, Adelaide, Australia). No changes were identified.

Table 3.2: Intron exon boundaries of the 4 exons of the *PRRG1* gene. Intronic sequence is in italics, exonic sequences in bold, and the polyadenylation signal is underlined.

Exon	Size	Left splicing sequence	Right splicing sequence
1			**GGGACCCGCT***GTGAGTGTGG*
2	91 bp	*TAAATTGTAG***GATGCTCACC**	**GATTACCTAG***GTAAGAATCT*
3	51 bp	*TTTTGTACAG***GGAATCATCA**	**ATGGGGAGGG***GTAAGTTTCT*
4	161 bp	*TTCTTTTTAG***TTTTCCTCAC**	**TGAAAAAACT***GTAAGTATGT*
5	4,154 bp	GTATTTTCAG**AAGGAGTTTT**	**GCTAAATAAAGCTATTTTTAA**

3.5.4 Analysis of the Chromosome 10 breakpoint

GenBank was searched for genomic sequence that was known to be located within 10q11.2. The BAC clones from which these sequences were derived were used as probes for FISH against metaphase chromosomes of Patient 3. At the time there was not a lot of genomic sequence from the region with all of these BACs only being working draft sequence and their location within 10q22.1 and with respect to each other was unknown. This region is now largely finished sequence and shows that the BAC clones used are arranged as shown in Figure 3.15. All probes gave distal signal except for 124O11 that gave spanning signal. This spanning signal is inconsistent with the other results as this BAC is located between two other BACs which were distal to the breakpoint. This sequence is very close to the centromere of chromosome 10 and hence contains highly repetitive DNA sequences and this may be the cause of the misleading spanning signal. This indicates that the breakpoint lies proximal to all of the BAC clones used. This would place the breakpoint in the pericentromeric heterochromatin region that contains satellites and other tandem repeats and is extremely gene poor (Guy *et al.*, 2000). Therefore it is unlikely that a gene from chromosome 10 is physically disrupted by this translocation.

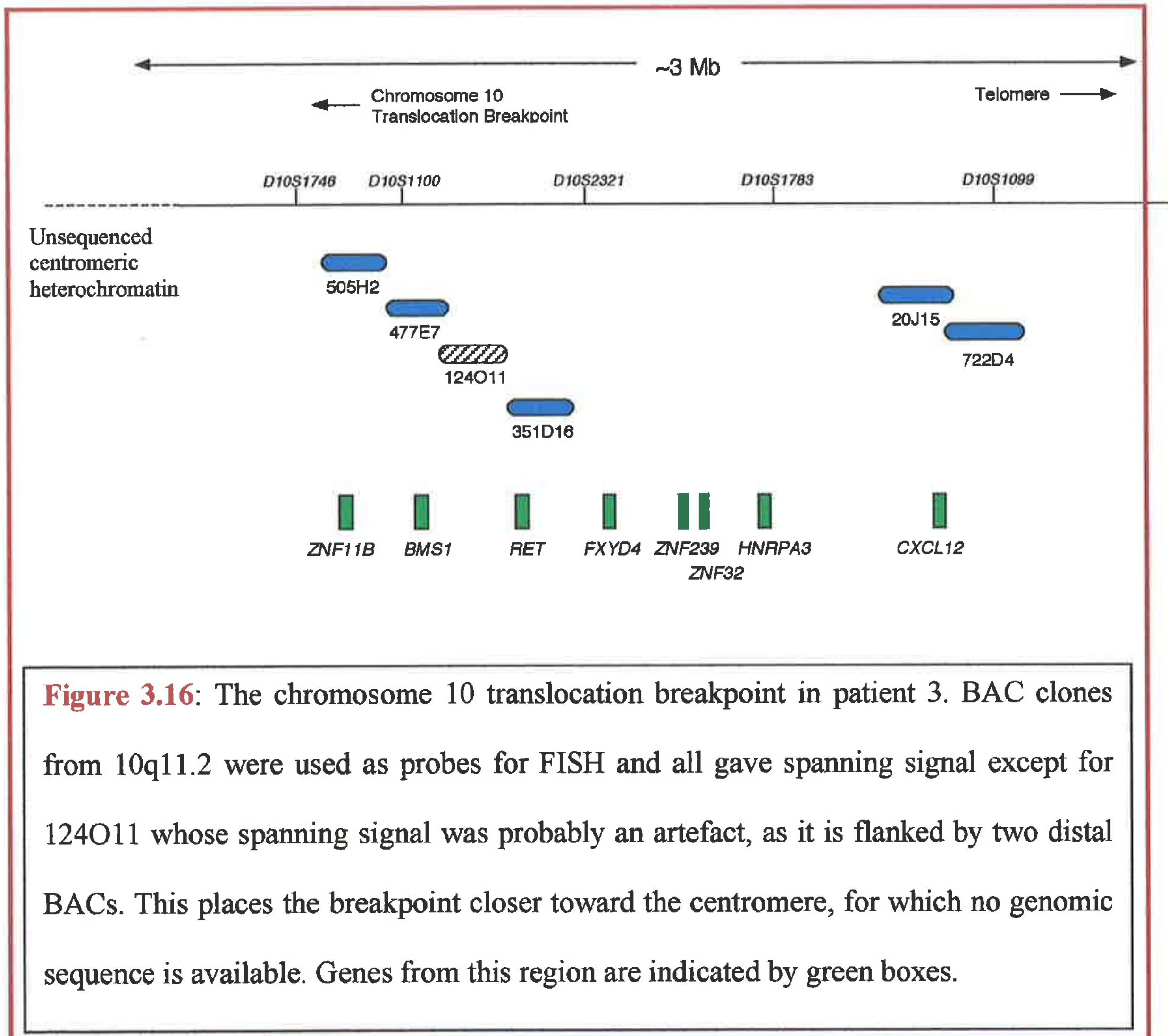


Figure 3.16: The chromosome 10 translocation breakpoint in patient 3. BAC clones from 10q11.2 were used as probes for FISH and all gave spanning signal except for 124O11 whose spanning signal was probably an artefact, as it is flanked by two distal BACs. This places the breakpoint closer toward the centromere, for which no genomic sequence is available. Genes from this region are indicated by green boxes.

3.5.5 Discussion

Both breakpoints in Patient 3 have been examined for the presence of genes that are disrupted by the translocation. For the chromosome 10 breakpoint it appears that the break lies within the pericentromeric heterochromatin. This region is known to contain very repetitive DNA sequences and are known to be very gene poor. It is therefore unlikely that there will be a gene disrupted by this breakpoint; however, this still needs to be confirmed by further FISH analysis so that a sequence spanning the breakpoint is identified and analysed for the presence of genes. This may not be possible due to the saturation of this region with repetitive sequences.

For the breakpoint at Xq28, a BAC clone of only ~75 kb has been identified that spans the breakpoint. The only gene from this region that could be physically disrupted by the breakpoint is expressed exclusively in testis (the gene represented by the ESTs BC044778 and BG724103). It seems unlikely that lack of expression of this gene would result in a phenotype of MR and cardiomyopathy as seen in Patient 3. However it is possible that there is very low expression of this gene in heart and brain, that was undetected by RT-PCR, or alternatively that there is embryonic expression of this gene important for development of heart and brain.

Analysis of other genes in the region has identified the gene *TMG3a*, the partial mRNA sequence of which was reported during the course of this project (Kulman *et al.*, 2001). This project has resulted in further study of this gene, and has revealed the identification of a large ~6 kb *TMG3a* isoform that is expressed in heart and brain, as well as a smaller *TMG3b* isoform that is more widely expressed. The discrepancy in size between the predicted mRNA from alignment of overlapping EST sequences and that of the transcripts identified by

Northern blot hybridisations, suggests that there may be additional sequence at the 5' end of the gene, that at present remain unknown. 5' RACE was attempted in order to identify additional 5' sequence using fibroblast RNA as template, however no product was obtained. This could be due to the low level of expression of this gene, as indicated by the low level of signal from the Northern blot hybridisations.

The predicted mRNA of *TMG3a* contains an ORF of 696 bp that would correspond to a predicted protein of 231 amino acids. In the case of the alternatively spliced isoform a predicted mRNA of only ~ 1 kb can be determined from alignment of the overlapping ESTs. This would contain an ORF of 489 bp and would encode a protein of 162 amino acids that has an alternative carboxy terminus to the larger isoform. Both isoforms contain the GLA domain which has up to 13 GLA residues (Kulman *et al.*, 2001). The 5.6 kb sequence derived from the alignment of ESTs may therefore correspond to the approximately 6 kb transcript seen on the Northern blot in brain and heart. The 1 kb alternatively spliced isoform may correspond to the ~4.5 kb transcript on the Northern blot, however if this is the case a large portion of this transcript remains unknown.

The only known biological function of the essential vitamin, vitamin K, is in the carboxylation of glutamic acid to γ-carboxyglutamic acid (Furie *et al.*, 1999). It has long been known that warfarin interferes with this vitamin K dependent activity which has led to its use as an anticoagulant, where it works by eliminating the effects of the GLA domain containing blood coagulation factors. It has also been observed that warfarin exposure *in utero* results in bone and neurological brain abnormalities (warfarin embryopathy syndrome) (Pati & Helmbrecht, 1994). This suggests that as well as interfering with the GLA domain containing proteins in blood, there may also be GLA domain containing proteins whose primary function

is in development of the CNS. This is supported by the identification of the protein *GAS6* that is highly expressed in the CNS (Manfioletti *et al.*, 1993) and has been shown to be involved in signal transduction pathways whereby it acts as a ligand for members of the Axl family of receptor tyrosine kinases (Nagata *et al.*, 1996).

When compared to other GLA domain containing proteins *TMG3a* lacks the signal peptide which is a characteristic of the other GLA domain proteins, except for *PRRG1* (Kulman *et al.*, 1997). The GLA domain is preceded by a propeptide sequence that is thought to be the recognition sequence of γ-glutamyl carboxylase. *TMG3a* is predicted to be a single pass transmembrane protein, the cytoplasmic domain of which contains two PPXY domains (Kulman *et al.*, 2001). This is the minimal binding motif of WW domains, found in a wide range of regulatory, cytoskeletal and signalling molecules (for review see (Zarrinpar & Lim, 2000)).

TMG3a does not seem to be physically interrupted by the breakpoint however the breakpoint does affect the expression of this gene. Therefore *TMG3a* now becomes a candidate for screening in other XLMR families mapping to Xq28. Due to the difficulty of determining the sequence of *TMG3a*, and because the full structure is yet to be determined, this gene has not been screened for mutations in XLMR families at this stage.

```
TMG3a  MA-V------------------------------------  3
PRRG1  MGRV------------------------------------  4
TMG4   MFTLLVLLS------QLPTVTLGFPHCARGPKASKHAGEE  34
PRRG2  MRGHPSLL----------LLYMALTTCL-DTSPSEETDQE  29
F2     MAHVRGLQLPGCLALAALCSLV------------------  22
F7     MV---------SQALRLLCLLLGLQGCLAAGGVAKASGGE  31
F9     MQRVN-MIMAESPGLITICLL----GYLLSAECT------  29
PROC   M-------------WQLTSLLLFVA---------------  12
GAS6   MAPS---LSPGPAALRRAPQLLLL---LLAAECALAA---  31

                     propeptide
              *     *
TMG3a  --------------FLEAKDAHSVL---KRFPR-ANEFLE  25
PRRG1  --------------FLTGEKANSIL---KRYPR-ANGFFE  26
TMG4   -------------VFTSKEEANFFI--HRRLL-YNRFDLE  58
PRRG2  -------------VFLGPPEAQSFLSSHTRIPRANHWDLE  56
F2     ---------HSQHVFLAPQQARSLL---QRVRR-ANTFLE  49
F7     TRDMPWKPGP-HRVFVTQEEAHGVL---HRRRR-ANAFLE  66
F9     -------------VFLDHENANKIL---NRPKRYNSGKLE  53
PROC   TWGISGTPAPLDSVFSSSERAHQVL---RIRKR-ANSFLE  48
GAS6   --------------LLPAREATQFLRPRQRR---AFQVFE  54

                     GLA domain
TMG3a  ELRQGTIERECMEEICSYEEVKEVFENKEKTMEFWKGYPN  65
PRRG1  EIRQGNIERECKEEFCTFEEAREAFENNEKTKEFWSTYTK  66
TMG4   LFTPGNLERECNEELCNYEEAREIFVDEDKTIAFWQEYSA  98
PRRG2  LLTPGNLERECLEERCSWEEAREYFEDNTLTERFWESYIY  96
F2     EVRKGNLERECVEETCSYEEAFEALESSTATDVFWAKYT-  88
F7     ELRPGSLERECKEEQCSFEEAREIFKDAERTKLFWISYSD  106
F9     EFVQGNLERECMEEKCSFEEAREVFENTERTTEFWKQYVD  93
PROC   ELRHSSLERECIEEICDFEEAKEIFQNVDDTLAFWSKHVD  88
GAS6   EAKQGHLERECVEELCSREEAREVFENDPETDYFYPRYLD  94
```

Figure 3.17: Protein sequence alignment of the amino terminal end of TMG3 and other GLA domain containing proteins. Conserved residues are highlighted in yellow. The GLA domain (red arrow) and propeptide (black bar) sequences are indicated. Asterisks indicate residues thought to be involved in the recognition by γ-glutamyl carboxylase. All but TMG3 and PRRG1 have additional sequence of a signal peptide located prior to the propeptide.

Although *TMG3a* is located ~114 kb telomeric to the end of the spanning BAC clone it, RT-PCR has shown that this gene is not expressed in Patient 3. Several genes have now been identified where chromosome rearrangement breakpoints have disrupted their expression even though the breakpoints are located some distance from the gene. The most studied gene is the β-globin locus where five regulatory elements are located 5' to the gene, that control its expression (for review see Li *et al.*, 1999). Such regions are termed locus control regions (LCRs) and can be located either 3' or 5' to the gene. 3' LCRs have now been identified for several genes where their disruption causes a number of disorders. These include aniridia (*PAX6*) (Fantes *et al.*, 1995), Greig cephalopolysyndactyly (*GLI3*) (Vortkamp *et al.*, 1991), Saethre-Chotzen syndrome (*TWIST*) (Rose *et al.*, 1997) and fascioscapulohumural dystrophy (*FSHD*) (van Deutekom *et al.*, 1996). Similarly LCRs have been identified 50-900 kb 5' to genes causing campomelic dysplasia (*SOX9*) (Wunderle *et al*, 1998), X-linked deafness (*POU3F4*) (de Kok *et al.*, 1996), Rieger syndrome (*PITX4*) (Flomen *et al.*, 1998), preaxial poyldactyly (*SHH*) (Lettice *et al.*, 2002), NSXLMR (*AGTR2*) (Vervoort *et al.*, 2002) and blepharophimosis/ptosis/epicanthus inversus syndrome (*FOXL2*) (Crisponi *et al.*, 2001). *TMG3a* can now be added to this growing list.

3.6. CONCLUSION

Three X chromosome rearrangements have been studied at the molecular level in order to identify candidate genes for familial XLMR. The sequencing associated with the HGP has assisted greatly in the approach used here. Even at the beginning of these studies in 2000 when much of the sequence was either not available or was in working draft form, construction of BAC contigs was made relatively simple by the available sequence at the time. As the sequencing of the HGP progressed any gaps could be easily filled in. This has resulted in rapid isolation of BAC clones from the region of interest in Xq28.

These BAC clones were used as probes for FISH, refining the location of the Xq28 breakpoints in each of the three cases. In each case a different mechanism for disruption of normal gene function, that results in MR, has been identified.

In Patient 1 the sequence across both of the breakpoints was determined. The 3' UTR of the gene *BGN*, is disrupted by the breakpoint but the ORF remains intact. It appears that there is a normal level of transcription in the cultured fibroblasts tested, however, this would need to be further supported by quantitative RT-PCR. 3' UTRs however, are known to be involved in the stability of mRNA as well as in the regulation of translation. Although the role of the 3' UTR of *BGN* is unknown, it is possible that such regulatory sequences are disrupted, and therefore *BGN* translation is affected. BGN is a well studied protein in terms of its role in bone growth, however, although it is expressed in brain its function here remains unclear.

In the case of Patient 2 the rearrangement proved to be more complex than originally thought. The analysis at Xp11.2 proved straightforward and showed that there were no genes in this

region that were physically disrupted, and that the two genes in the region were still expressed. At the Xq28 breakpoint this study has shown that the MR in this patient is likely to be due to a duplication of a relatively small region (submicroscopic by conventional cytogenetics); a region that is very gene dense. The functional disomy of these genes is therefore the likely cause of MR. Although the cause of MR was discovered, this did not identify candidate genes for familial XLMR.

For Patient 3 yet another mechanism for MR has been detected. In this case the translocation breakpoint from chromosome 10 does not appear to disrupt any genes. At Xq28 the gene *TMG3a* has been identified and was found not to be expressed in cultured fibroblasts of Patient 3. *TMG3a* is located more than 100 kb away from the breakpoint and therefore is not physically disrupted by the translocation. More and more cases are being found where the function of LCRs are affected by chromosome rearrangements. These regions can be located up to hundreds of kb away from the genes whose expression they control. Completion of the human genome sequence as well as comparison of non-coding sequences of other organisms may help in identifying exactly what these sequences are as well as how they work.

Two isoforms of *TMG3a* have been identified. The larger of these is expressed in heart and in brain and the smaller in several other tissues as well. It is therefore likely that silencing of this gene causes the phenotype in Patient 3.

Positional cloning of X chromosome rearrangements in patients with MR does not always lead to identification of a candidate gene, and even less often to a gene which will be found to contain mutations in familial NSXLMR (Figure 3.18). In many cases no gene is identified at or near the breakpoint. Or, as with the case of one patient described here, the rearrangement

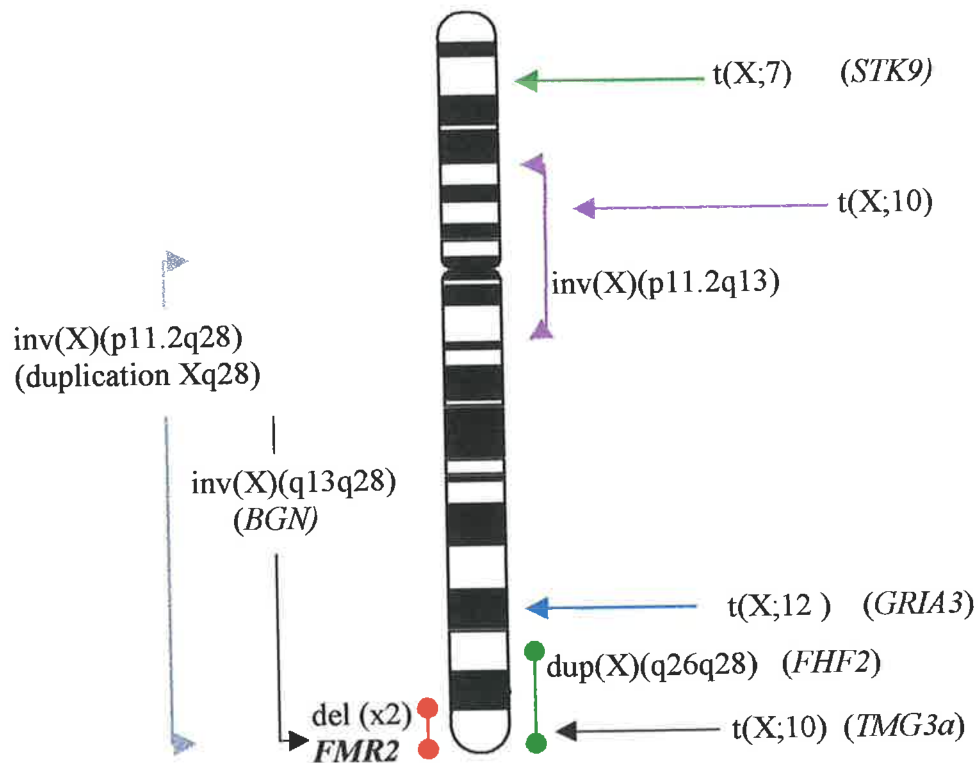


Figure 3.18: Ideogram of the X chromosome showing characterised X chromosome rearrangements from the laboratory of the Department of Cytogenetics and Molecular Genetics, WCH, or through collaboration. Several scenarios have resulted; the NSXLMR was identified (*FMR2*); no gene was identified near the breakpoint (purple); a gene was identified at the breakpoint but no mutations were found in familial cases (blue); a gene was found at the breakpoint, but subsequently mutations in other genes have been shown to cause the syndrome being studied (green); the rearrangement was more complex than initially thought (grey). *TMG3a* and *BGN* have been identified as candidates but are yet to be screened in familial XLMR.

proved to be more complex than was initially thought, shows that in some cases NSXLMR is not monogenic. Additionally, as with *TMG3a,* genes affected may be located up to hundreds of kilobases away from the breakpoint, whereby locus control regions are disrupted. As well as this, although a gene can be shown to be physically interrupted in a patient with MR, these same genes may not contain mutations in familial XLMR. For example, although breakpoints have been found to disrupt the *STK9* gene in patients with a phenotype of ISSX, no mutations were found in familial ISSX (Kalscheuer *et al.*, 2003). Familial ISSX has now been shown to be the result of mutations in the *ARX* gene (Chapter 5 and Strømme *et al*, 2002). Similarly, characterisation of a duplication of the X chromosome in a patient with symptoms mimicking Borjeson-Forssman-Lehmann syndrome (BFLS) identified *FHF2* as a candidate (Gécz *et al.*, 1999); however, mutations in another gene *PHF6* were recently shown to cause this XLMR syndrome (Lower *et al.*, 2002). Hence in these cases the characterisation of the sporadic rearrangement was misleading in the context of presenting candidate genes for screening in familial cases.

This project has identified *BGN* and *TMG3a* as candidates for screening in familial XLMR. As well as this, other genes with similar functions also become good candidates. *TMG3a* is also uncovered as a candidate for X-linked cardiomyopathy, however pursuit of this hypothesis is beyond the scope of this project.

Chapter 4

Fine Mapping and Candidate Gene Screening in a Family with Non-Syndromic X-linked Mental Retardation Mapping to either Xq13.1 or Xq23

4.1. INTRODUCTION **113**

4.2. RESULTS **118**

4.2.1 FINE MAPPING AND LINKAGE ANALYSIS - Xq13.1 118
4.2.2 FINE MAPPING AND LINKAGE ANALYSIS - Xq23 122
4.2.3 TRANSCRIPTION MAP OF Xq13.1 124
4.2.4 GENE SCREENING 128

4.3. CONCLUSIONS **138**

4.1. INTRODUCTION

Large families in which a Mendelian disorder is segregating are a powerful resource in the search for disease causing mutations in genes. Once the appropriate institutional ethics approval and DNA samples have been obtained, linkage analysis is used in order to determine the region of the genome to which the gene localises in the family being investigated. However, ambiguity in the gene localisation can arise when a statistically significant lod score of greater than +3 for autosomal disorders or +2 for X-linked disorders is not achieved. This may occur even though SLINK analysis (Weeks *et al.*, 1990) shows that a family is potentially large enough to reach this statistical value. This may be due to one or more of the reasons outlined below with specific reference to MR.

1. Phenocopies

Mental impairment is not uncommon in the population, so there may be males with MR in the family whose phenotype is due to environmental or genetic factors other than those causing MR in the other affected family members. Haplotype analysis based on markers encompassing the true gene localisation will show that they have the same genotype as the normal males in the pedigree. This will erroneously exclude that region which carries the gene containing the mutation of interest unless some allowance is made in the linkage analysis for the presence of phenocopies.

2. Incomplete penetrance

Conversely, incomplete penetrance of the mutation in a male member of the family who is not clinically assessed as affected but carries the same mutation as the affected males in the family, will also exclude the region that is the true gene localisation. For linkage analysis he

will be coded as unaffected, but would have the same haplotype as his affected relatives. Where incomplete penetrance is expected, it must be incorporated into the linkage analysis.

Incomplete penetrance is not generally an issue for X-linked recessive disorders. However in the case of a mutation resulting in mild to borderline MR, the clinical diagnosis may be difficult when the genetic background would otherwise confer a higher than average IQ. Males coded as unaffected for linkage analysis may indeed carry the same mutation as their affected relatives. This would falsely exclude the true localisation of the gene containing a mutation in the family being studied, unless allowances for incomplete penetrance were built into the linkage analysis.

An example of such incomplete penetrance of mild to borderline MR is seen in males with the *FRAXE* fragile site. In some cases the only clue that males carry the defective *FMR2* gene is that they are positive for the associated molecular marker of expansion of the $(GCG)_n$ repeat at the fragile site (Gécz *et al.*, 1997). Mutations in other genes that are related to cognitive function, but are not associated with a molecular marker such as a fragile site, would remain cryptic. While they may lower IQ, they would not necessarily lower it to the extent that it reaches the mild to borderline range.

3. Skewed X-inactivation

In the case of X-linked disorders, skewed X-inactivation can also create problems for diagnosing females. If all females carrying the mutations have the same X-inactivation patterns (i.e. random or skewed) then the appropriate diagnosis for linkage analysis is likely to be made. However, this is not always the case and some carriers may have skewed X-inactivation while others may have random X-inactivation. This has the potential to make

heterozygous carriers of a mutation present with different phenotypes in the same family, and hence misdiagnosis may prevent reaching a statistically significant lod score. Here, the safest option is to code all females as unaffected with penetrance of carrier females set at zero, relying on obligate carrier females and affected males for gene localisation.

4. Uninformative markers

The distribution of recombination events along the chromosome is of importance in order to find a minimal linkage interval that is small, and suitable for candidate gene screening. It is desirable to obtain a large family, with as many members of the family as possible participating in the study, as this will increase the chance of finding recombination events that will reduce the minimal linkage interval. Ideally, markers used need to be informative in all parts of the family, so that each X chromosome can be fully tracked. Where markers are not informative, other nearby markers need to be genotyped so that a haplotype covering the region becomes informative.

Once a region of linkage has been identified, the analysis of additional markers located between the marker for which there is a recombination event and the first marker for which there are no recombinants can further refine the linkage interval. Creation of extended haplotypes can reduce the region by a significant amount and hence reduce the number of putative candidate genes from within the region that need to be screened by mutation detection methods.

The family O'H is represented by those family members willing to co-operate in a linkage study; these family members along with other key members are shown in the pedigree in

Figure 4.1: Partial pedigree of the O'H family.

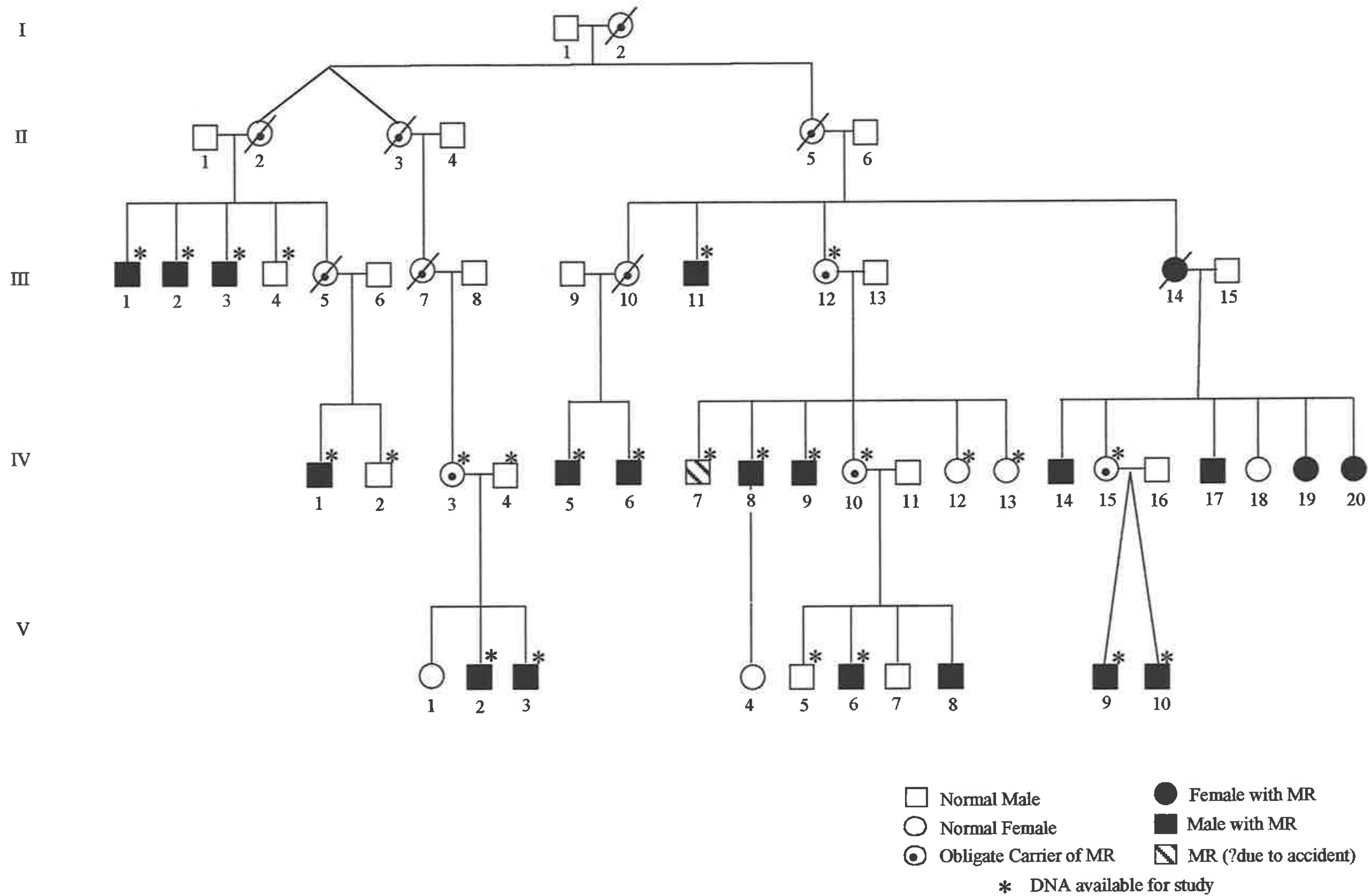


Figure 4.1. This is a large family with mild to moderate mental retardation segregating in an X-linked recessive manner. This family was first investigated by Dr John Ferguson in 1987, and follow-up clinical investigations were performed by Professors Michael Partington and Gillian Turner of Hunter Genetics, New South Wales, Australia. The large size of this family promised the possibility of defined localisation and the chance of a relatively small minimal linkage interval in which to search for candidate genes. Initial linkage analysis for this family was carried out by Dr Andrew Donnelly during his PhD candidature in 1997 (Donnelly, 1997). Two point lod score analysis indicated linkage to either of two possible regions of the X chromosome, while excluding the remainder. The Xq13 region lay between markers *DXS1275* and *DXS559* with a lod score of 2.14 at *DXS453*. At Xq23 the linked region was between markers *DXS1059* and *DXS1212*, with a lod score of 2.55 at *DXS424*. These regions were both relatively small regions, estimated to be approximately 5 Mb and 6 Mb respectively.

Although clear linkage to one region was not achieved, this family still remains a promising family in which to look for NSXLMR genes as such small linkage intervals are rarely obtained even for families of this size. Initially, DNA from an additional three affected males (III-1, III-2 and III-3), their unaffected brother (III-4) and one obligate carrier female (III-12) (Figure 4.1), was obtained, in the hope that linkage analysis would rule out one of these regions and reinforce the other

4.2. RESULTS

Markers used for the initial linkage analysis were retested in the newly obtained family members. The new lod scores are shown in Table 4.1. Unfortunately neither Xq13 nor Xq23 was clearly excluded. Individual IV-7 (Figure 4.1) was coded as unknown for linkage analysis as he had an accident at a young age, which could be the cause of his MR. Extended haplotypes of both regions were then constructed, by typing additional $(AC)_n$ markers from both regions. At both localisations it is clear that there are a pair of brothers that do not carry the same haplotype as the rest of the affected in males in the family (Figures 4.2 and 4.3 discussed below).

4.2.1 Fine Mapping and Linkage Analysis - Xq13.1

The new lod scores revealed that at Xq13.1 there was no lod score of >+2 at a recombination fraction of 0.0. At 0.1 however, a lod score of 2.40 was reached for the marker *DXS8107*. This could be explained if the gene truly mapped to Xq13.1, but was flanked by close crossover events.

Haplotype analysis at Xq13.1 (Figure 4.2) revealed two affected boys (V-9 and V-10) that did not share the same haplotype as the other affected males in the family. It was also possible to infer haplotypes for some of the females at the top of the pedigree for which no DNA was available for the linkage study. For this region this suggested that the positive lod score of 2.14 obtained for the marker *DXS453* could be an artefact resulting from absence of actual genotypes from critical parts of the family. Extended haplotypes also show that apart from *DXS453*, V-9 and V-10 have entirely different haplotypes to the other affected males and that

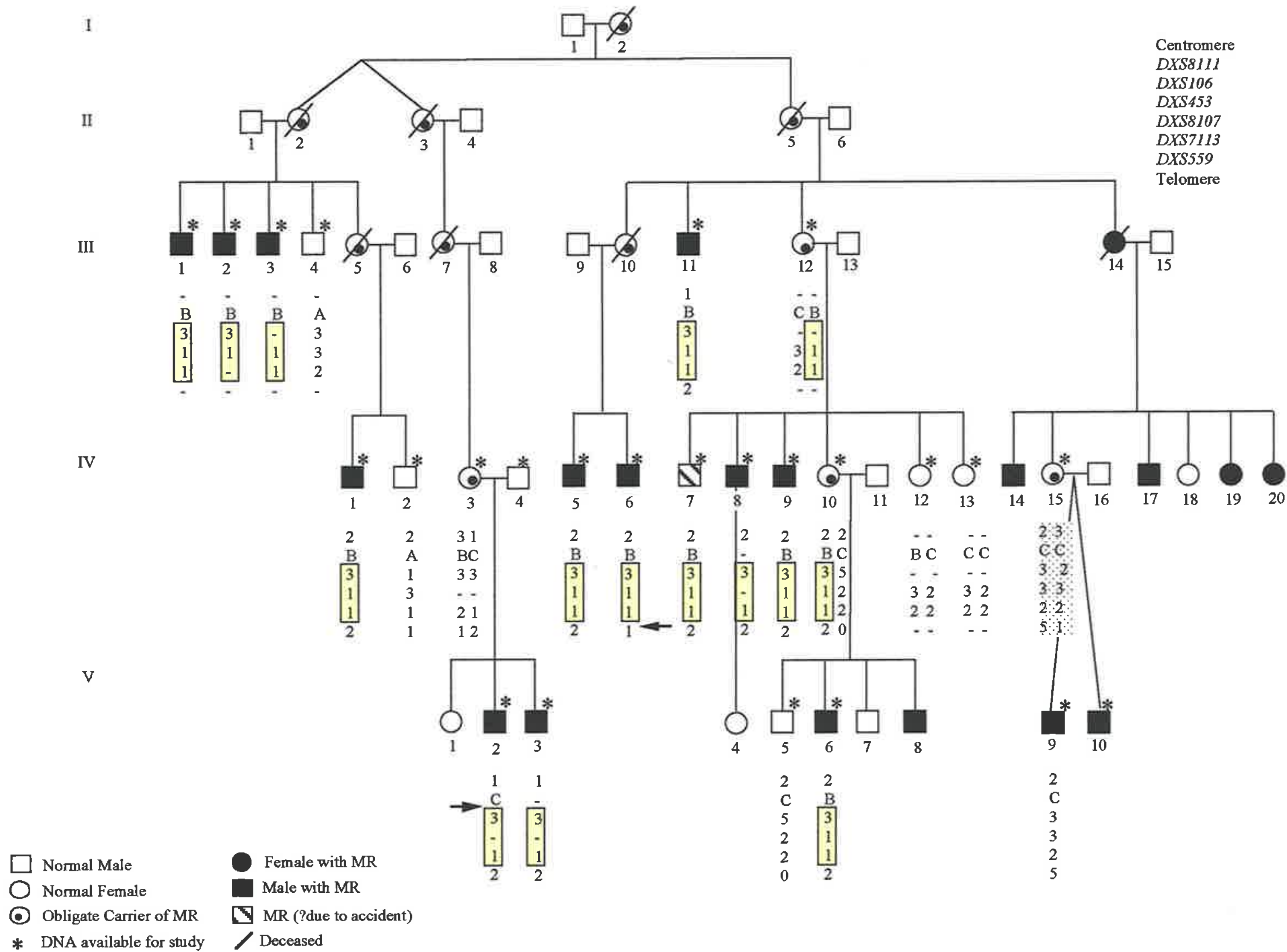
Table 4.1: Lod scores for the X chromosome for the O'H family. Markers from Xq13 and Xq23 are labelled.

	Order	0.0	0.01	0.05	0.1	0.2	0.3	0.4
	DXS996	-infini	-7.60	-2.98	-1.23	0.11	0.49	0.41
	DXS16	-infini	-5.36	-2.66	-1.58	-0.65	-0.25	-0.05
	DXS999	-infini	-6.79	-3.43	-2.08	-0.89	-0.33	-0.06
	DXS365	-infini	-7.63	-4.12	-2.66	-1.35	-0.73	-0.34
	DXS989	-infini	-8.14	-3.48	-1.69	-0.25	0.24	0.28
	DXS451	-infini	-4.36	-2.36	-1.58	-0.87	-0.48	-0.21
	DXS992	-infini	-7.76	-3.76	-2.20	-0.88	-0.30	-0.04
	5'DYS	-infini	-2.05	-0.76	-0.29	0.04	0.11	0.09
	DXS1068	-infini	-6.53	-3.21	-1.93	-0.86	-0.39	-0.14
	MAOA	-infini	-0.63	-0.04	0.13	0.21	0.19	0.12
	DXS1003	-infini	-3.21	-0.63	0.27	0.78	0.70	0.38
	DXS991	-infini	-0.52	0.04	0.18	0.17	0.1	0.03
	ALAS2	-infini	-3.12	-0.59	0.26	0.69	0.55	0.24
	DXS1125	-infini	-0.04	0.47	0.54	0.41	0.21	0.06
	DXS1275	-infini	0.33	1.41	1.60	1.33	0.78	0.20
	DXS106	-infini	0.61	1.66	1.81	1.46	0.80	0.13
Xq13	*DXS8107*	-infini	2.40	2.76	2.61	1.96	1.15	0.36
	DXS7113	-infini	1.96	2.33	2.20	1.60	0.85	0.22
	DXS559	-infini	-1.96	-0.12	0.46	0.71	0.59	0.33
	DXS1124	-infini	-0.66	0.54	0.88	0.90	0.66	0.33
	DXS986	-infini	0.31	0.81	0.88	0.72	0.48	0.23
	DXS995	-infini	-3.46	-0.94	-0.09	0.37	0.31	0.10
	DXS990	-infini	-1.07	0.16	0.54	0.68	0.56	0.33
	DXS454	-infini	1.56	2.06	2.08	1.76	1.27	0.68
	DXS1120	-infini	-1.03	0.69	1.10	1.00	0.58	0.22
	DXS456	-infini	-1.76	0.10	0.69	0.89	0.66	0.29
	DXS1059	-infini	-0.95	0.83	1.31	1.35	0.96	0.45
	DXS1072	-infini	0.05	1.13	1.32	1.07	0.58	0.15
Xq23	*DXS11*	-infini	2.05	2.41	2.27	1.67	0.93	0.26
	DXS8088	1.24	1.20	1.07	0.91	0.64	0.41	0.21
	DXS1220	-infini	1.07	2.07	2.16	1.68	0.93	0.29
	DXS424	-infini	1.22	1.64	1.59	1.19	0.72	0.32
	DXS8081	-infini	0.01	0.53	0.60	0.47	0.27	0.12
	DXS8064	-infini	1.40	1.78	1.68	1.21	0.71	0.31
	DXS8067	-infini	-0.25	0.28	0.37	0.28	0.15	0.05
	DXS8059	-infini	-4.45	-1.87	-0.94	-0.28	-0.09	-0.03

they are likely to have inherited either the grandpaternal or great-grandpaternal haplotype. It is possible that there is a double recombination event in this region, but this is unlikely as it is a very small region.

If the gene does in fact map to this region, then these two males (V-9 and V-10) would represent phenocopies. These two males are twins and have identical genotypes for all markers tested (including several autosomal markers tested by A. Donnelly) and therefore they are assumed to be monozygous twins. As such they were counted as only one meiotic event for linkage analysis. It is also of note that MR in this branch of the family occurs in two aunts of V-9 and V-10, as well as in their maternal grandmother. Furthermore additional Xq13 markers used for haplotype analysis has reduced this possible location to between *DXS106* and *DXS559*, by recombination events in V-2 (and probably V-3) and IV-6 respectively. Unfortunately there was limited DNA available from V-2 and V-3 and therefore this interval could not be reduced from the *DXS106* side.

Figure 4.2: O'H family with haplotype data from Xq13.1



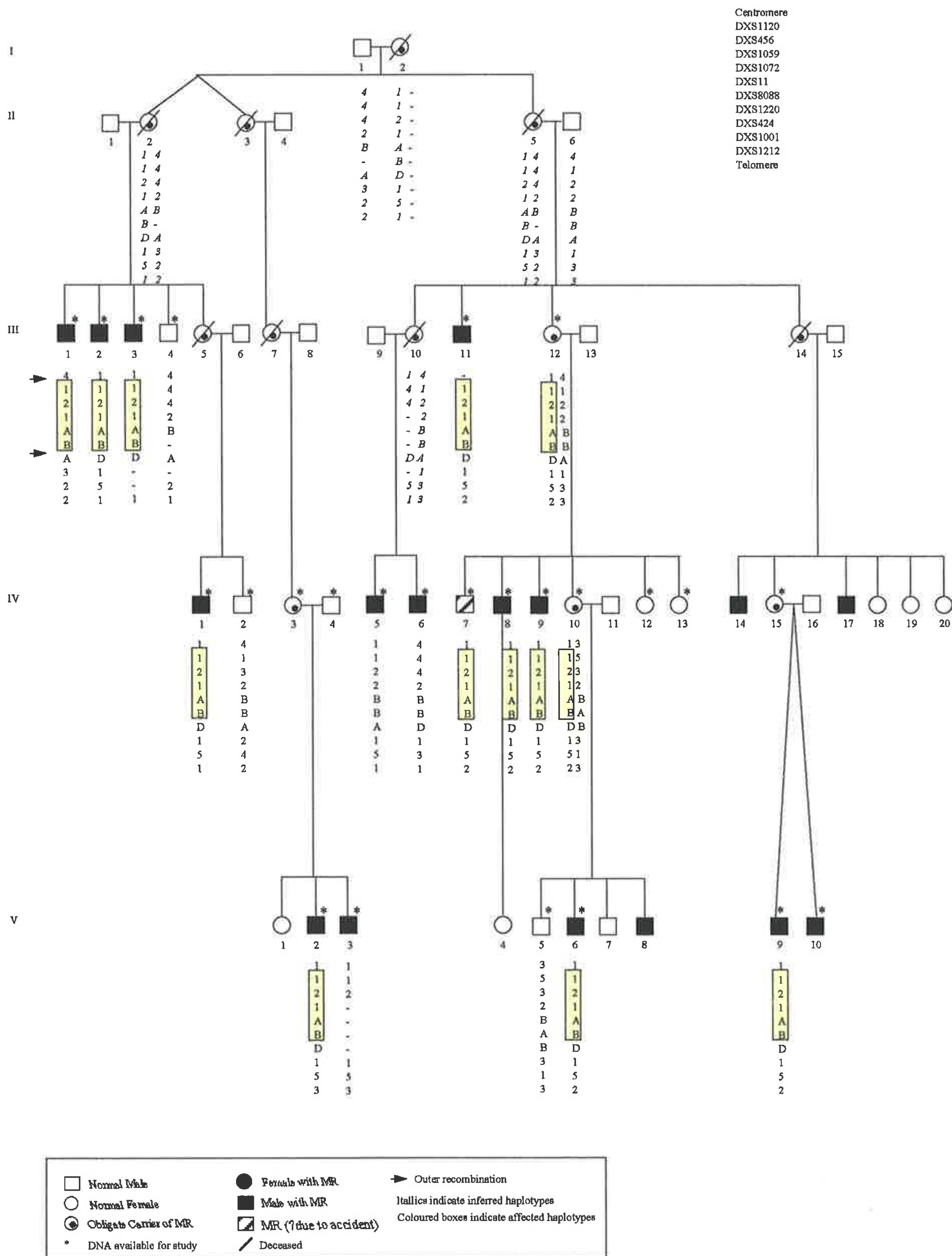
4.2.2 Fine Mapping and Linkage Analysis - Xq23

The haplotypes from the new family members, and extended haplotype analysis using markers from Xq23 also modified the original the positive lod score of 2.55 obtained for *DXS424*. Firstly III-12 (Figure 4.3) is uninformative for *DXS424* and therefore it is likely that her sister (III-10 for whom no DNA is available for the study) is also uninformative. Hence the positive lod score could have been an artefact of not knowing the exact genotype for at least part of this family. Additionally, analysis of microsatellites in the newly acquired DNA samples shows that the affected male III-1 (Figure 4.3) does not have the same allele as other affected males in the family including his two affected brothers, but has the same allele as his unaffected brother. All markers were fully informative in this branch of the family.

There is a hint of linkage at *DXS8088* with a lod score of 1.24 at a recombination fraction of 0.0, however this marker is largely uninformative in this pedigree. Individuals IV-5 and IV-6 (Figure 4.3) have the same allele for this marker as the other affected males in the family. However analysis of the Xq23 haplotype shows that this is the only marker in the region that IV-6 shares with the affected males. His brother, IV-5, also shares *DXS456* and *DXS1059* alleles, but it is likely that in this case these alleles came from his maternal grandfather (II-6) and can therefore not be the chromosome carrying the same mutation as the other affected males in this family.

The only locus in this region for which all affected males share the same genotype, is now *DXS8088*. This marker, however, is again uninformative in most parts of the family. Therefore in order for the gene to be linked to this region it has to be assumed that one or both of IV-5 and IV-6 are phenocopies. This would then place the gene between *DXS1120* and *DXS1220*.

Figure 4.3: O'H family and haplotypes from Xq23



Further haplotyping has therefore assisted in understanding the lod scores from the original study. This has also enabled haplotypes for some of the key obligate carriers for which no DNA is available for the study to be inferred. These inferred haplotypes suggest that the positive lod scores at both markers from the original linkage study, were due to the markers not being informative in some parts of the family. As the rest of the X chromosome has been excluded, it is still likely that one of these regions does contain the NSXLMR gene. For linkage to either Xq13 or Xq23, there has to be at least one phenocopy for MR in this family. As there are affected females in the branch of the family containing V-9 and V-10, it seems more likely that at least some of the MR in this branch of the pedigree is due to other causes. If this is the case then linkage to Xq13 would be suggested, although linkage to Xq23 cannot be discounted. It seems likely that for definitive localisation the gene defect needs to be established.

4.2.3 Transcription Map of Xq13.1

As both of the regions associated with either putative localisation are relatively small when compared to most localisations from a single family, both were examined for the presence of candidate genes. The Xq13 region was the most likely location and it was mainly concentrated on for further study. Initially, the publicly available genomic DNA sequences that were available at the time of this study were searched for STSs that were known to lie within the minimal interval. This revealed a number of BAC clones which were either fully sequenced or that were in working draft form. The ends of these clones were used to search for overlapping sequences to create a contig across the regional localisation. This resulted in a contig that contained some gaps. These were filled in by a search of the BAC end database (TIGR), whereby two sequences were linked and orientated with respect to each other if they contained the sequences of the opposite ends of the same BAC clone. Over the course of this

project, draft sequences became finished sequences and gaps in the contig were filled in, and the resulting contig is shown in Figure 4.4.

At the *DXS106* (centromeric) end of the minimal linkage interval it was not possible to reduce the interval further using polymorphic markers, as there was not sufficient DNA available from the key recombinants (V-2 and V-3) (Figure 4.2). From the *DXS559* (telomeric) end there was sufficient DNA available for study from the individual (III-11) who was recombinant at this marker. Assessing other polymorphic markers in the region had the potential to rule out some of the genes as positional candidates by placing them outside the minimal linkage interval. The genomic DNA sequence was scanned for $(AC)_n$ stretches with more than fifteen copies of the repeat, in a search for novel polymorphic markers that would be potentially informative in the family and minimise the regional localisation. Three loci were identified and oligionucleotides designed to test if they were polymorphic (listed in Table 4.2). One repeat was found at the 5' end of the *OGT* gene (*OGT-(AC)*$_n$, Figure 4.4) Typing of this marker showed that it was informative in the OH family and showed that III-11 was still a recombinant at this locus. This further reduced the minimal linkage interval such that the genes *OGT* and *GPR9* were now excluded given that the orientation of the gene is centromeric to telomeric.

Table 4.2: Oligonucleotide sequences for amplification of $(AC)_n$ repeats.

	Forward	**Reverse**
AC-1	GGCTGCTTTGAAGTTTGGTCT	CTTAGTCATTGCCTGGCCAG
AC-2	TTGATATAGTACAGCTTGAATG	GGCACGAAGATAAGCTGCCA
AC-3	CTGTGTACATTGGCTTGTCC	GGCGCACCACATCAATCAC

A transcription map was then generated *in silico* by using the sequence contig from the region. BLAST searches against the nr division of GenBank identified already characterised genes, and the dbEST division was searched in order to identify uncharacterised genes. The transcription map of the minimal linkage interval at Xq13 is shown in Figure 4.4. This region is now almost completely sequenced and is approximately 2 Mb in length.

The searches revealed a gene rich region that included genes such as those for X-linked severe immunodeficiency syndrome (*IL2RG*) (Noguchi *et al.*, 1993), and X-linked Charcot-Marie-Tooth disease (*GJB1*) (Bergoffen *et al.*, 1993). The genes in the region were analysed and were determined whether they were good candidates for NSXLMR based on the following criteria:

1. For characterised genes, if the gene was expressed in brain as assessed by published Northern blots.

2. For uncharacterised genes, EST sequences were examined to see if they were derived from brain cDNA libraries.

3. Genes were excluded if they were not expressed in brain from one of the previous two criteria (e.g. *ITGB1BP2* is muscle specific, and *ARR3* is retina specific).

4. Genes were excluded if they were already found to contain disease causing mutations. This was especially the case if a large number of different mutations along the length of the gene had already been described and the phenotype did not include MR (i.e.

Figure 4.4: Transcription map of the minimal linkage interval of the NSXLMR gene in the O'H family. The outer recombinants occur at DXS106 and DXS559. This region is approximately 2 Mb in size.

Centromere
Telomere
~ 2 Mb
DXS106
DXS7625
sWXD513
DXS453
DXS7111
DXS7112
DXS8107
DXS8052
DXS7118
DXS7117
DXS7119
OGT-(AC)n
DXS559
AC002416
AL158141
AL357752
AL139109
AL627390
AL590764
AL590763
AL158069
AF003528
AL139111
AL139398
AL627071
AL590762 (305,000bp)
AL392090
AL450448
ED1 3'
Similar to RAS related protein RAB
KIF4
TEX11
SLC7A3
ZNF261
NE-DLG
AFX1
HOPA
ITGB1BP1
TAF1
OGT
ARR3
IL2RG
P2RY4
GJB1
SNX12
NONO
GPR9
IGBP1 3'
NLG3
Genes involved in other disorders or not expressed in brain
Genes expressed in brain
50Kb

mutations along the length of the *IL2RG* gene cause SCID with no apparent brain involvement.

5. Genes were considered functional candidates if the gene product had a proposed function in the normal development of the brain, or had a proposed function in the same pathways as other previously identified XLMR genes.

Table 4.3 shows a summary of the genes from the region and the information available from published Northern blots and the UniGene clusters. This table highlights that in fact most of the genes in this region can be classified as good candidates for XLMR based on the above criteria.

4.2.4 Gene Screening

Direct sequencing of candidate genes from affected members of the family was used to try to identify disease causing mutations. Affected males selected for sequencing were those that shared the same haplotype as the affected members from the rest of the family. This specifically excluded V-9 and V-10. Screening of two affected males, excluding these two males, would reduce the risk of missing disease causing mutations due to phenocopies, the presence of which has been suggested by the linkage analysis.

Table 4.3 Genes from the minimal linkage interval of the O'H family. (N/A – expression results from brain not published; Yes/No indicates expression in brain; genes in blue are those deemed to be good candidates for NSXLMR, and an asterisk (*) indicates those genes that were screened during this study).

Gene	Disease	Expressed in brain		
		UniGene cluster/ESTs	Tissue expression (Published Northern blot, RT-PCR)	
ED1	Anhydrotic ectodermal dysplasia 1 (OMIM #305100)	Hs.105407 **No**	**Yes**	(Kere *et al.*, 1996)
*IGBP1**	Immunoglobulin binding protein	Hs.3631 **Yes**	Ubiquitous **Yes**	(Onda *et al.*, 1997)
P2RY4	G-protein coupled pyrimidinergic receptor	Hs.248157 **No**	placenta, pancreas, ?	(Communi *et al.*, 1995; Stam *et al.*, 1996)
ARR3	Regulator of G protein-coupled receptors	Hs.308 **No**	**No**	(Murakami *et al.*, 1993)
*Ras related protein Rab**	Novel gene with similarity to *RAB6*	Hs.276327 No (only 2)	N/A	N/A
*KIF4**	Microtubule based motor protein	Hs.279766 **Yes**	**Yes**	(Oh *et al.*, 2000),
*NE-DLG**	Membrane associated guanylate kinase	Hs.11101 **Yes**	**Yes**	(Makino *et al.*, 1997)
TEX11	Unknown function	Hs.121776 **No**	**No**	(Wang *et al.*, 2001)
*SLC7A3**	High-affinity cationic amino acid transporter	Hs.175220 **Yes**	**Yes**	(Vekony *et al.*, 2001)
*SNX12**	Sorting nexin	Hs.287867 **Yes**	N/A	N/A
*AFX1**	Forkhead domain transcription factor	Hs.239663 **No**	**Yes**	(Peters *et al.*, 1997)
IL2RG	X-linked severe combined immunodeficiency syndrome (OMIM# 300400)	Hs.84 **Yes**	N/A	N/A
HOPA	Possible association with X-linked mental retardation/hypothyroidism syndrome	Hs.211607 **Yes**	**Yes**	(Philibert *et al.*, 2002)
*NLGN3**	Cell-cell interactions	Hs.47320 **Yes**	**Yes**	(Philibert *et al.*, 2000)
GJB1	X-linked Charcot-Marie-Tooth disease (OMIM# 302800)	Hs.333303 **Yes**	**Yes**	(Scherer *et al.*, 1995)
*ZNF261**	Disrupted by translocation in XLMR	Hs.9568 **Yes**	**Yes**	(van der Maarel *et al.*, 1996)
ITGB1BP2	Integrin beta(1) - interacting protein	Hs.109999 **Yes**	**No**	(Brancaccio *et al.*, 1999)
NONO	Nuclear RNA binding protein	Hs.172202 **Yes**	**Yes**	(Peters *et al.*, 1997)
TAF1	TATA binding protein associated factor/cell cycle regulatory protein	Hs.1179 **No**	N/A	N/A

ZNF261

One of the genes within the minimal linkage interval was the zinc finger gene *ZNF261,* which was previously shown to be disrupted by a translocation involving the X chromosome, [t(X;13)], in a female patient with MR, scoliosis and abdominal hypopigmentation (van der Maarel *et al.*, 1996). However no cases of familial NSXLMR with mutations in this gene have been described. This gene is highly expressed in foetal and adult brain and therefore presented as the best candidate gene from this region, for NSXLMR in the O'H family. Primers were kindly provided by Dr van der Maarel (MPIMG) (see Appendix 1 for primer sequences) and were used to PCR amplify all exons of the gene that were subsequently sequenced. No sequence changes, either disease causing or otherwise, were identified.

As part of this project the genomic structure of several of these genes that presented as good candidate genes was determined. They were *KIF4*, *NE-DLG3*, *SNX12*, *SLC7A3*, *AFX1*, a novel gene with similarity to the *RAS related protein RAB* and *NLGN3*. The order in which they were screened was determined in part, by the availability of the full genomic sequence of the gene, such that intron/exon boundaries could be established, and oligonucleotides for PCR amplification and sequencing could be easily designed.

NLGN3

Neuroligins are a protein family with a putative function in cell-cell interactions between neurons (Missler *et al.*, 1998). The human *NLGN3* gene is expressed as three transcripts of approximately 7.5 kb, 4 kb and 2.4 kb, with the 4 kb transcript being widely expressed in adult brain (Philibert *et al.*, 2000). The gene consists of eight exons that span approximately

25 kb of genomic sequence. Interestingly the putative promoter region of *NLGN3* overlaps with the *HOPA* gene. As the exon/intron boundaries were already described (Philibert *et al.*, 2000), oligonucleotides were designed (Appendix 1) within the intron sequences, and the entire ORF in affected members from the OH family was PCR amplified and sequenced. No sequence changes were identified.

KIF4

The kinesin superfamily of proteins, are a group of motor proteins that are involved in the transport of cargo (e.g. organelles, protein complexes, and mRNAs) along microtubules (Goldstein, 2001). *KIF4* is one member of this large family of proteins, and was previously shown to encode an mRNA of ~5 kb, however, the genomic structure of this gene had not been determined. This gene was represented by the UniGene cluster Hs.279766. The EST sequences and the mRNA sequence were aligned and the consensus sequence obtained was compared to genomic sequences from the region (Accession numbers AL139398 and AL357752). The exon and intron boundaries were thus determined (Table 4.3). *KIF4* is comprised of 30 exons that reside within a genomic region of approximately 130 kb. The mRNA is 4,342 bp and has an ORF of 3,699 bp encoding a protein of 1, 232 amino acids. This is in agreement with the mRNA sequence published (Oh *et al.*, 2000) (accession number NM_012310). The alignment of the ESTs revealed single nucleotide polymorphisms (SNPs) at positions 1265 (G or T) and 1656 (C or T).

Oligonucleotides were again designed such that PCR amplification and direct sequencing would allow detection of any sequence changes in the ORF. Oligonucleotide sequences are presented in Appendix 1. Sequencing affected males identified one change at position 27

(C→T); however, this change did not result in an amino acid change. This is not a disease causing change as it is remote from sites that would be expected to be involved in splicing.

Table 4.3: Intron exon boundaries of the 30 exons of the *KIF4* gene. Intronic sequences are in italics, exonic sequences in bold, and the polyadenylation signal is underlined.

Exon	Size (bp)	3' acceptor splice site	5' donor splice site
1			
2	115	*CCTTCTCTAG***GTGGTGGTTG**	**GTATTTAAAG***GTAAGGCGAT*
3	190	*ATTTTGTCAG***GATATAATGC**	**TTACTTAGAG***GTAAGCAATT*
4	90	*ACAATATCAG***ATTTACAATG**	**AGGCATAAAG***GTGTGTTTGT*
5	167	*TTGCAATTAG***ATTGTGGGAC**	**AAAGTGACAA***GTAAGTTACA*
6	95	*ACAGGAATAG***CAGCTTTCGC**	**CTAAAAGAGG***GTAAGAGAGT*
7	117	*CTGCTATTAG***GTATTAATAT**	**CTGCTTCAAG***GTAAGCCCAA*
8	176	*TCTTTTTCAG***ATTCTCTAGG**	**AAAGCAACAG***GTATAAGGGA*
9	62	*CTTCTCCCAG***GTACAACAGC**	**GATCTATAAC***GTAAGAGTCA*
10	133	*TTTGTTTTAG***TGTGGAACCA**	**GATCATTTTG***GTAAGCCCCC*
11	59	*TTTGTTGCAG***ACAGAGCAAG**	**AGCATGCGGC***GTAAGTTGCC*
12	106	*GAATTTTCAG***CTGCAAACTG**	**CCAGTTATCG***GTAAGCCAAG*
13	57	*GTTACTGCAG***GATGAAACTG**	**GCAAGAAGCC***GTAAGTAATT*
14	186	*TTTTTTCCAG***CAAGTAGAAA**	**CCAATACCAG***GTAAACTATT*
15	104	*ACCTTTCTAG***GATAACATAA**	**ACCAAGCCAA***GTAAGAATAA*
16	145	*CTGTGTTTAG***GTTGAGTGAG**	**GGAGATACGG***GTAATAAATT*
17	111	*GTTTTTATAG***ATGATGAAAA**	**AAAAGAACGA***GTAAGTAACT*
18	84	*TTCTTGTCAG***GACCGTAAGA**	**AACGGAGGAG***GTAAGAAAAT*
19	114	*TTTTATTTAG***GCAGCAGCTG**	**TCGAGTGAAG***GTATGAACAA*
20	156	*TCTCTTCCAG***AATTGGCTTG**	**TAAACTCCGG***GTAAGTACGC*
21	101	*TTATTTACAG***AGGCGTACAT**	**TGGAATTCAG***GTAACAGGGA*
22	130	*TTCCCTGAAG***GAGTGCTCAG**	**GATTGGAGAG***GTAAACATCA*
23	171	*TTCCTTCCAG***CTGGTCTCCT**	**CCAAGAGAAG***GTAAACTCTA*
24	108	*CTTTCTCTAG***GTGCTGTACC**	**GAAGTGTCAG***GTATGATCAC*
25	78	*CTGGAAATAG***GATGAAGAAC**	**CATCAAGCAG***GTAATACAGT*
26	96	*TTTCAAATAG***AAACTGACCC**	**CCCACCTAAG***GTAAAATGAT*
27	183	*TTTTTCCTAG***CCAAAACCTT**	**CATCCAAGGG***GTAGGAGCCA*
28	117	*TGTTTTGCAG***TGTTCCTGCA**	**GCAAGGCAAG***GTAGGATCAG*
29	123	*TGATTTTCAG***GATAGCTTGG**	**CAATAGCAAG***GTAGGTGGGC*
30	765	*TCCCTCCTAG***ATCCTGAAAG**	**ATGTTAAGTAAGTA<u>AATAAA</u>**

NE-DLG

The neuroendrocine - discs large related (*NE-DLG*) gene is a member of the membrane associated guanylate kinase (MAGUK) family of proteins that are involved in membrane and cytoskeleton associated signalling (Anderson, 1996). *NE-DLG* is the human ortholog of the mouse gene *SAP102* (102 kDa synapse associated protein) that has been shown to interact with NMDA receptors at neuronal synapses, suggesting that it may link ion channels to the submembranous cytoskeleton associated with synapses (Muller *et al.*, 1996). *NE-DLG* therefore presented as a good positional as well as a good functional candidate gene. The ORF has been partially screened by PCR amplification and direct sequencing of affected members from the O'H family. No sequence changes were identified.

SNX12

Sorting nexins are a family of proteins involved in the regulation of membrane trafficking (Haft *et al.*, 1998). The *SNX12* gene was first identified by a search of databases for similarity matches to *SNX1* (Teasdale *et al.*, 2001). Alignment of the ESTs from the UniGene cluster (Hs.287867 and Hs.5076) revealed that the mRNA sequence in GenBank (Accession number AF171229) is only a partial sequence of this gene. The mRNA sequence as determined by the alignment of ESTs is now ~3.6 kb and has an ORF of 489 bp encoding a predicted protein of 162 amino acids. The intron/exon boundaries were determined by comparison of the ORF with the corresponding genomic sequence (Accession number AL590764), and are listed in Table 4.4. Oligonucleotides were designed such that the entire ORF could be sequenced. No sequence changes were identified after sequencing the ORF.

Table 4.4. Oligonucleotides were designed such that the entire ORF could be sequenced. No sequence changes were identified after sequencing the ORF.

Table 4.4: Intron exon boundaries of the 4 exons of the *SNX12* gene. Intronic sequence is in italics, exonic sequences in bold, and the polyadenylation signal is underlined.

Exon	Size	3' acceptor splice site	5' donor splice site
1	282 bp		**TCGCATGCGG***GTGAGTCACG*
2	96 bp	*TTCCATGTAG***ACAAACCTAC**	**AGATAGCAAG***GTATGGCTTT*
3	125 bp	*GTGCCTGCAG***ATTGTAGTAC**	**TTATTAACAA***GTAAGCCAAG*
4	1870 bp	*TTCTTTATA***GAATTGCTGGG**	**TCAACTCTGTTCATATTAAA**

SLC7A3

The cationic amino acid transporters are membrane associated proteins involved in amino acid transport, some of which act as neurotransmitters, synaptic modulators, or neurotransmitter precursors (Palacin *et al.*, 1998). The *SLC7A3* gene presented as a good candidate for NSXLMR based on this as well as on the knowledge that mutations in the gene *SLC6A8* were recently found to contain mutations in XLMR patients (Hahn *et al.*, 2002; Salomons *et al.*, 2001).

The *SLC7A3* mRNA sequence was previously determined (Vekony *et al.*, 2001). Alignment of ESTs from the UniGene cluster hs.175220 was in agreement with this sequence, except that the 5'UTR was extended by an additional 205 bp, a sequence that contains a (GT) repeat. The exon/intron boundaries were determined by comparison of the predicted mRNA sequence with the genomic sequence AL627071 (Table 4.5). This showed that the gene is comprised of

twelve exons, with the predicted start codon located within exon 2, and the stop codon within the last exon. The mRNA sequence is ~2.4 kb with an ORF of 1,860 bp encoding a protein of 619 amino acids. Oligonucleotides were designed in order to sequence the entire ORF from genomic DNA. No sequence changes were identified.

Table 4.5: Intron exon boundaries of the 12 exons of the *SLC7A3* gene. Intronic sequence is in italics, exonic sequences in bold, and the polyadenylation signal is underlined.

Exon	Size	3' acceptor splice site	5' donor splice site
1			**CAGCTCTCTG***GTAAGCGTTG*
2	395 bp	*TTTTCACCAG***ATCATCTTCT**	**TATGTCATTG***GTGAGATATG*
3	159 bp	*GCTACACCAG***GTACAGCCAG**	**CTGCTCACTG***GTGAGGCAAA*
4	178 bp	*TCTCCCACAG***GATTGTTGGC**	**ACACCTATAG***GTTAGATGGT*
5	113 bp	*TCCACTCTAG***CTTGGGTCCT**	**GCTACCACTG***GTAACACAGT*
6	223 bp	*CTTCCTTCAG***GAGAAGAAGC**	**TTTCTACCAG***GTCAGTATCA*
7	140 bp	*TCCTCTCCAG***CCTCCTGGGC**	**ATTATTGCAG***GTAACAGACC*
8	103 bp	*TTAATTTTAG***CATTCATGGC**	**TCATCCTCAG***GTGAGACTCT*
9	167 bp	*TTTGCCTCAG***GTATCAACCT**	**TCATTGCTTG***GTGAGCAGTG*
10	167 bp	*ACCTTTGAAG***CTGTCCTGCT**	**TCACTTTAAG***GTAAATGAAC*
11	109 bp	*TTGTTTTCAG***GTGCCTGCTT**	**ATGCTGATTG***GTAGGTATCC*
12	365 bp	*TCCCCACTAG***GCTTTGCTAT**	**AAAGAAAGTAGAAAAATAAA**

AFX1

AFX1 is a member of the forkhead family of transcription factors, a family of proteins with a wide range of biological functions. Transcription factors have a role in cell fate determination and development and thus this gene presented as a good functional candidate. The genomic structure of *AFX1* was previously described (Peters *et al.*, 1997). Oligonucleotides were designed such that the entire ORF could be sequenced. This gene consists of 3 coding exons,

were designed in order to sequence the ORF in the affected males from the O'H family. No sequence changes were identified.

RAS like GTP-ase RAB

A search of the sequence between *DXS106* and *DXS559* on the Ensembl Genome Browser (www.ensembl.org) revealed the presence of predicted gene (Ensembl Transcript ID ENST00000276066). This gene has similarity to the *RAS*-related *RAB* gene family. RAB GTPases are the largest group of the RAS superfamily of proteins, and have been implicated in vesicle trafficking. Therefore this gene presented as a good functional candidate.

A BLAST search of the predicted gene sequence showed an overlap with two ESTs in the dbEST database. These ESTs (accession numbers AI126463 and BM931690) are from cDNA libraries derived from testis, and a pooled library of eye components respectively. An additional two ESTs (accession numbers BQ877160 and BU679329) also overlap with these two, but they do not have any introns and it is possible that these clones are genomic contaminants from the cDNA libraries. This is especially likely for BU679329 that reads in the opposite direction from all the other clones. There is no indication that this gene is expressed in brain, however, as there are so few ESTs available, it cannot be assumed that it is not. Together this made a gene that consisted of 943 bp of mRNA, with eight exons covering approximately 8 kb of genomic DNA (accession number AL357752). This gives a predicted ORF of 516 bp encoding a predicted protein of 172 amino acids. Comparison of the mRNA sequence with the genomic sequence revealed 8 exons, whose intron exon boundaries are listed in Table 4.6.

Table 4.6: Exon/intron boundaries of the *RAS like GTPase RAB* gene. Intron sequences are in italics, and the polyadenylation signal at the end of exon 8 is in bold.

Exon	Size	3' acceptor splice site	5' donor splice site
1	41 bp		GAGCAGAGCG*GTGTGGGTCT*
2	59 bp	*TCCCTCACAG*TAGGGAAGAC	CGCCTGCCAG*GTAAGACCAC*
3	54 bp	*TCTTTTGCAG*GCAACTGTTG	GGACCAAATA*GTGAGTGTTA*
4	106 bp	*TTTTATGCAG*GTTCAGCTGC	GACATTACAA*GTGGGTGTTA*
5	112 bp	*TCTCTTGCAG*ACATCAATTC	TAACAAAAG*GTAAAGTATAA*
6	94 bp	*TTCGTCCAGA*CAAGTCACTG	CGTGAAAAAG*GTAATACTTG*
7	64 bp	*ACACACACAG*CTGTTCCGGC	AAAGAGGGGA*GTATCCTTTT*
8	399 bp	*CACACTTCAG*CGGTTGAAAT	TTACCTACACCCTC**AATAAA**

Oligonucleotides were designed to PCR amplify all coding exons and the products obtained were sequenced in affected members of the O'H family and are listed in Appendix 1. No sequence changes were identified.

AGTR2

During the course of the project *AGTR2* located within the linkage interval for Xq23 was identified as a gene for XLMR/NSXLMR (Vervoort, 2002). The 1,092 bp ORF was located within a single exon and was directly sequenced in affected males from the O'H family using oligonucleotides listed in Appendix 1. No sequence changes were identified.

4.3. CONCLUSIONS

Additional family members for linkage analysis were invaluable for further refinement of the region of the X chromosome the gene in this family most likely to maps to. Haplotypes for obligate carrier females, for which there is no material available to study, could be inferred through the haplotypes of these additional family members. This suggested that for markers at Xq23, the positive lod scores achieved in the initial linkage analysis (Donnelly, 1997) were probably due to uninformative markers. In fact the additional males studied III-1, III-2, III-3 and III-4 (Figure 4.3) rule out linkage to the marker *DXS424* where the maximum lod score was achieved, as III-1 has a different allele to his affected brothers (III-2 and III-3) and the same allele as his normal brother (III-4). This moves the haplotype shared by most of the affected males to between *DXS456* and *DXS1220*; however, this region is not shared by affected brothers IV- 5 and IV-6. *DXS8088* is the only marker at which these males have an allele in common with the other affected males. This is likely to be because this marker is uninformative in their mother, as suggested by inferred haplotypes. Although it is still not clear whether the gene is located within Xq13.1 or Xq23 the available data suggest that Xq23 may be ruled out as a likely localisation.

By exclusion, Xq13 then remained the best region in which to search for mutations in candidate genes. V-9 and V-10 do not share the same haplotypes as other affected males for this region. In fact it is likely that they have the grandpaternal haplotype. The presence of affected females in this part of the pedigree, suggests that there may be a second cause of MR in this part of the family.

The linkage region at Xq13 was thus localised to between *DXS106* and *DXS559*, a region covering approximately 2 Mb. This is a fairly gene rich region and contains many genes that presented as good candidate genes. Screening was initiated on some of these genes; however, a disease causing change is yet to be identified. Recently a mutation has been identified within the *NLGN3* gene in two brothers with autism (Jamain *et al.*, 2003). It is of note that although no mutations were found in this gene in the O'H family, that the criteria used here for the selection of candidate genes is reliable.

Several good candidates have not yet been screened including *HOPA* and *NONO.* In particular the *HOPA* gene is one that stands out as a good candidate. This is a large gene containing several repetitive sequences including a $(CTG)_n$ repeat as well as a dodecamer repeat. One allele of this exonic dodecamer repeat has been implicated in various disorders including depression, schizophrenia and MR by some groups (Beyer *et al.*, 2002; Philibert *et al.*, 2002). However other groups do not agree with these findings (Michaelis *et al.*, 2000).

The genes described here will also present as candidates for other disorders that have also been mapped to Xq13.1. Of note are X-linked recessive torsion dystonia 3 (*DYT3*) (OMIM# 314250). This disorder was mapped between *DXS7117* and *DXS7119* based on linkage disequilibrium studies in a population study in Panay and island of the Philippines in which this disorder was present in an unusually high frequency. FG syndrome (OMIM #305450) is a genetically heterogeneous XLMR syndrome where one form has been mapped across this region (Xq13.1-Xq21.31) encompassing the minimal linkage interval of MR in the O'H family. The minimal linkage interval for Wieacker-Wolff syndrome, one feature of which is mild MR, also includes this region of Xq13.

The gene structures presented here were generally determined by sequence alignment of the ESTs representing each gene. It is possible that some isoforms of these genes are not represented within these sequences and therefore additional isoforms containing extra exons or coding sequences may have not been identified. The full characterisation of the genes will identify such isoforms and extra sequences can be screened for mutations in affected members from this family. This project is ongoing and the remaining genes are now being screened by high throughput methods by a group headed by Dr Lucy Raymond at the Sanger Centre and future screening of these genes may identify the mutation in this family.

Chapter 5

Identification of mutations in the homeobox gene *ARX* in patients with syndromic and non-syndromic XLMR

5.1. INTRODUCTION **142**

5.2. RESULTS **143**

5.2.1 TRANSCRIPTION MAP AND IDENTIFICATION OF CANDIDATE GENES FOR ISSX 143
5.2.2 *ARX* GENE CHARACTERISATION AND IDENTIFICATION OF MUTATIONS IN ISSX FAMILIES 145
5.2.3 *ARX* SCREENING OF XLMR FAMILIES 151
5.2.4 *ARX* GENE EXPRESSION, GENE FUNCTION AND CONSERVATION 153
5.2.5 *ARX* SCREENING IN SMALL FAMILIES WITH PUTATIVE XLMR 159
5.2.6 *ARX* SCREENING OF PATIENTS WITH DEVELOPMENTAL DELAY THAT TESTED NEGATIVE FOR *FRAXA* 160
5.2.7 MAMMALIAN CELL TRANSFECTION OF NORMAL ARX AND ARX WITH AN EXPANDED POLYALANINE TRACT 162

5.3. DISCUSSION **167**

5.3.1 INSERTION IN THE FIRST POLYALANINE TRACT OF THE *ARX* GENE 168
5.3.2 DUPLICATION AND DELETION OF THE SECOND POLYALANINE TRACT OF THE *ARX* GENE. 170
5.3.3 MISSENSE MUTATIONS IN THE *ARX* GENE 172
5.3.4 TRUNCATION MUTATIONS IN THE *ARX* GENE 174
5.3.5 OTHER POLYALANINE EXPANSION DISORDERS 175
5.3.6 MUTATIONAL MECHANISM OF SEQUENCES ENCODING POLYALANINE TRACTS 177
5.3.7 DISEASE MECHANISMS OF ARX POLYALANINE EXPANSION 179
5.3.8 ANIMAL MODELS OF *ARX* MUTATIONS 180

5.4. CONCLUSION **181**

5.1. INTRODUCTION

X-linked infantile spasms syndrome (ISSX or West Syndrome) (OMIM# 308350) is a syndromic form of XLMR. Affected males have normal intrauterine and early postnatal growth with infantile spasms occurring within the first year of life, which is followed by developmental regression. MR is severe to profound and death usually occurs within the first decade. Hypsarrhythmia, a chaotic electroencephalogram pattern, is also seen in affected males. W.J. West first described this syndrome in 1841 in his own son. X-linked inheritance was suggested by two reports of families with only affected males (Feinberg & Leahy, 1977) (Rugtveit, 1986). Claes, *et al* (1997) reported two families with ISSX and showed by linkage analysis that in both families the gene mapped between Xpter and Xp11.4. It was also noted that in these families obligate carrier females were asymptomatic. Linkage was reinforced in a third family and the localisation refined to a 7 cM region in Xp21.3 - Xp22.1 (Bruyere *et al.*, 1999a).

Previously two translocations between Xp22 and different autosomes were identified in unrelated female patients with a phenotype characteristic of ISSX (Kalscheuer *et al.*, 2003). The breakpoints were characterised in order to find a gene disrupted by the breakpoint, which would then become a candidate gene for familial ISSX. Both translocation breakpoints were found to disrupt the serine threonine kinase gene *STK9.* Direct sequencing of this gene in ISSX families did not reveal any disease causing mutations (Dr J. Gécz unpublished results).

Linkage intervals, such as the one for ISSX, make positional candidate gene screening approaches feasible, especially with the release of the draft of the human genome sequence. The linkage region for ISSX also overlapped with several other MRXS linkage intervals, as

well as nine MRX families. Therefore, it was a promising region in which to search for XLMR genes.

This study has included the contributions of many others, which have been acknowledged where these contributions are described. The main role I have played in this project was in the screening of probands from XLMR families (sections 5.2.3-6, and in the creation of the GFP-ARX construct (section 5.2.7). My contribution to this project earned equal first authorship on the seminal publication in Nature Genetics (Strømme *et al.*, 2002b), as well as co-authorship on two other related publications (Turner *et al.*, 2002; Strømme *et al.*, 2002a). The contributions of all are included in this chapter, so that my work could be put into perspective.

5.2. RESULTS

5.2.1 Transcription map and identification of candidate genes for ISSX

(Carried out by Dr Jozef Gécz)

Linkage analysis had previously localised the gene for ISSX between the loci *DXS1226* and *AHC* (adrenal hypoplasia, congenital) in Xp21-p22. The genetic length of this region is approximately 7 cM. Transcription mapping of the corresponding physical interval indicated the presence of a relatively small number of genes. The partial transcription map of this region is shown in Figure 5.1. These genes included *IL1RAPL1*, previously shown to contain mutations that result in NSXLMR (Carrie *et al.*, 1999). Direct sequencing of this gene was carried out on four previously described and mapped ISSX families (Bruyere *et al.*, 1999b), (Strømme *et al.*, 1999; Claes *et al.*, 1997), however no disease causing mutations were found.

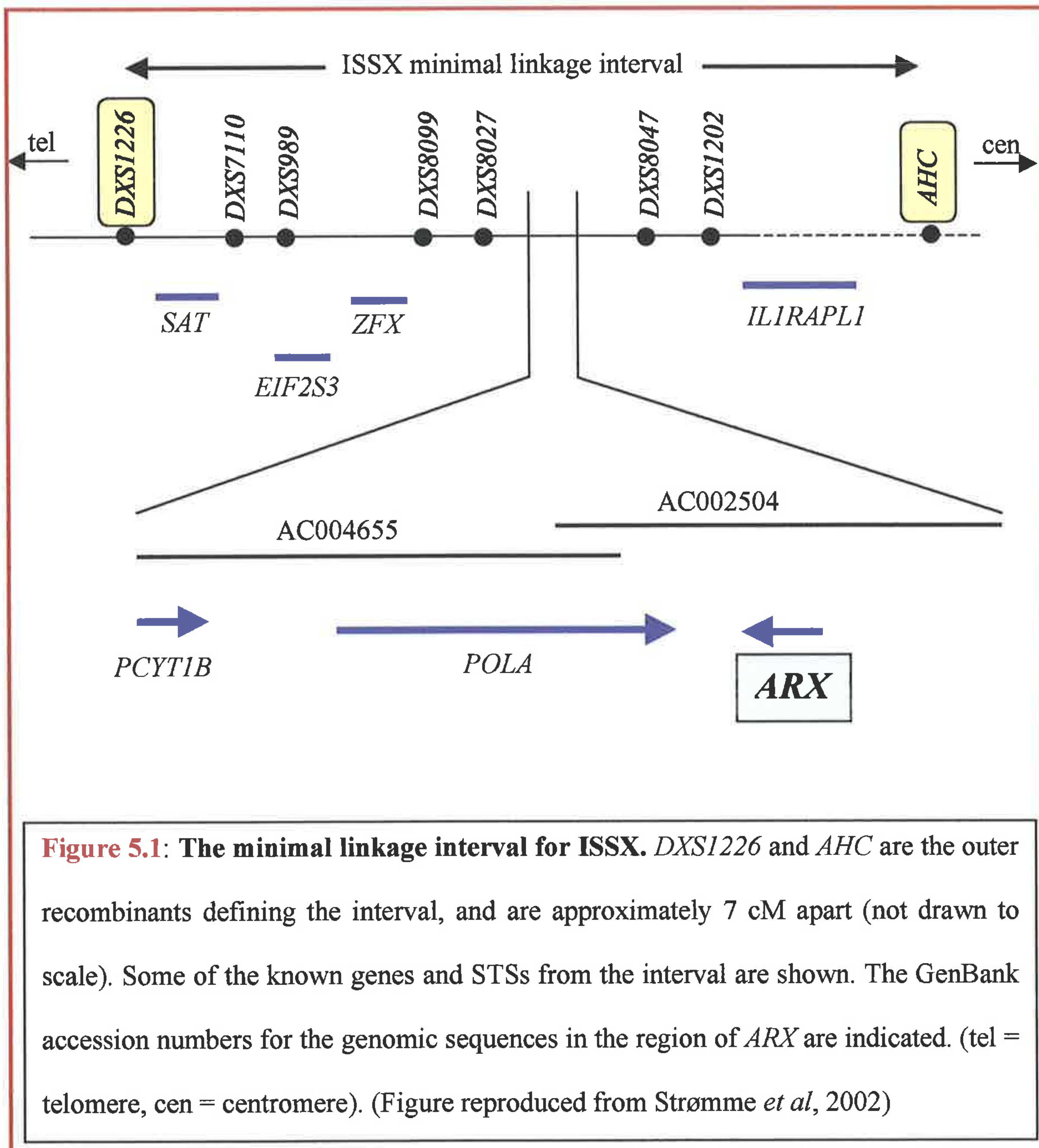


Figure 5.1: **The minimal linkage interval for ISSX.** *DXS1226* and *AHC* are the outer recombinants defining the interval, and are approximately 7 cM apart (not drawn to scale). Some of the known genes and STSs from the interval are shown. The GenBank accession numbers for the genomic sequences in the region of *ARX* are indicated. (tel = telomere, cen = centromere). (Figure reproduced from Strømme *et al*, 2002)

The *ARX* (*aristaless*-related homeobox) gene was also within this linkage interval, and presented as a good candidate gene based on its expression in foetal, infant and adult brain as indicated by ESTs (Unigene cluster Hs. 157208). *ARX* is a member of the paired-type homeobox gene family, and was shown to play a major role in the developing forebrain of mouse and zebrafish (Miura *et al.*, 1997). Additionally, *aristaless* related homeobox genes are involved in vertebrate embryogenesis, in particular head development (Galliot & Miller, 2000).

5.2.2 *ARX gene characterisation and identification of mutations in ISSX families*

(Carried out by Drs Jozef Gécz, Petter Strømme, Marie Mangelsdorf).

The *ARX* gene covers a region of about 12.5 kb of genomic DNA. It consists of five exons and is transcribed into a 2.8 kb mRNA, with an open reading frame (ORF) of 1,686 bp encoding a protein of 562 amino acids (Figure 5.2). It is an extremely GC-rich gene, with a GC content of roughly 72%. There are also four polyalanine tracts within the protein, which are encoded by GCN triplets.

Oligonucleotides were designed that amplified all five exons. Oligonucleotides for amplification of exons three and four were designed within the flanking introns so that the entire exon was amplified. This was similar with exon two, but due to the size and high GC content of this exon it was necessary to amplify this in two overlapping parts. Exon one was amplified from the region prior to the start codon, and extended into intron one. Similarly the exon five PCR product was amplified from intron four and extended into the 3' UTR after the stop codon. Oligonucleotide sequences are given in Table 5.1.

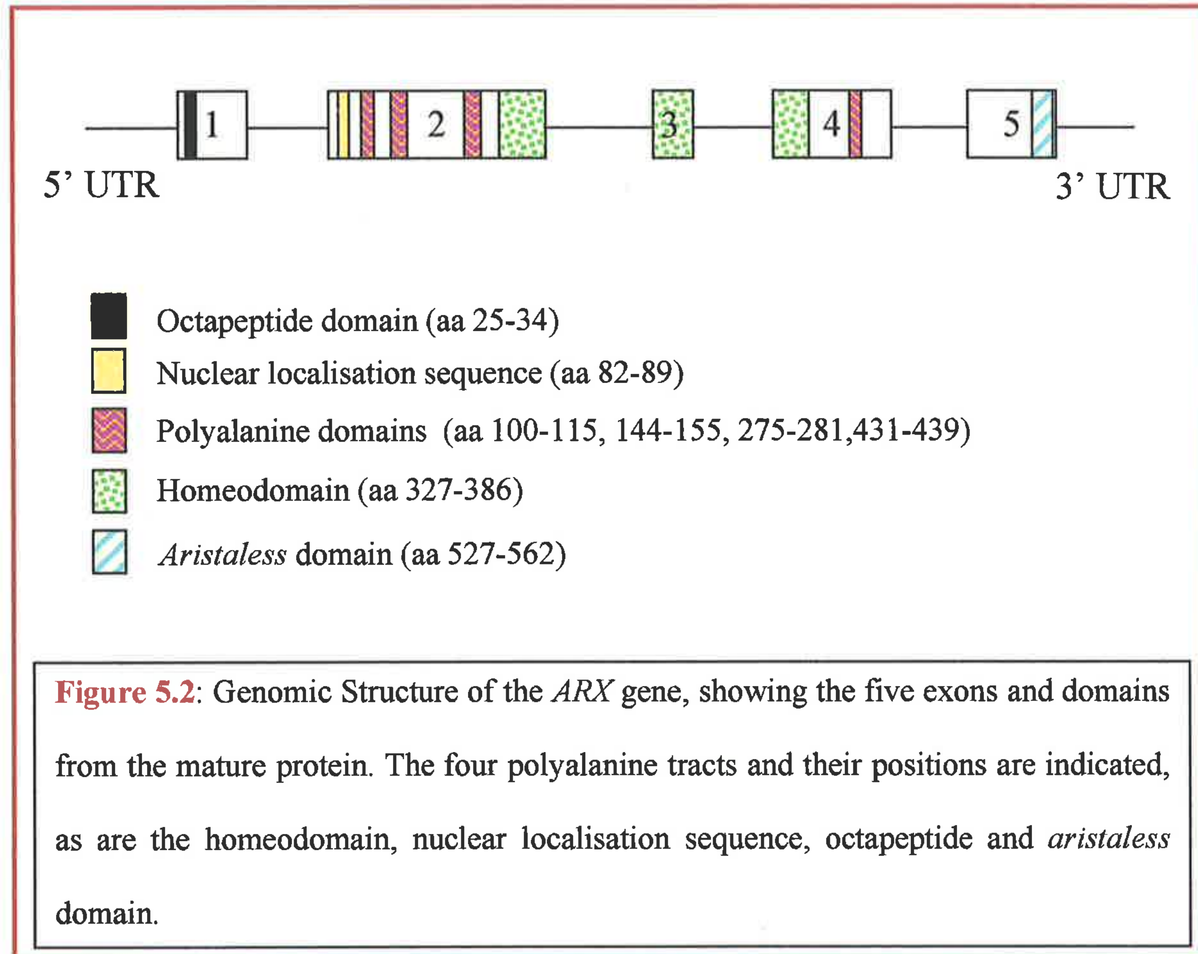


Figure 5.2: Genomic Structure of the *ARX* gene, showing the five exons and domains from the mature protein. The four polyalanine tracts and their positions are indicated, as are the homeodomain, nuclear localisation sequence, octapeptide and *aristaless* domain.

Table 5.1: Oligonucleotides used for PCR amplification and sequencing of the *ARX* gene from patient genomic DNA. All primers are given in the 5' to 3' orientation.

Exon 1	Forward Reverse	GTC CAC TAC ACT TGT TAC CGC AAT TGA CAA TTC CAG GCC ACT G
Exon 2	Part 1 Forward Part 1 Reverse Part 2 Forward Part 2 Reverse	ACG CCT GGG CCTA GGC ACT G CTC GGT GCC GGT GCC ACC AC GCA AGT CGT ACC GCG AGA ACG TGC GCT CTC TGC CGC TGC GA
Exon 3	Forward Reverse	GAA ATA GCT GAG AGG GCA TTG C TCT CTT GGT TTT GTG AAG GGG AT
Exon 4	Forward Reverse	GAC GCG TCC GAA AAC AAC CTG AG CCC CAG CCT CTG TGT GTA TG
Exon 5	Forward Reverse	ACA GCT CCC GAG GCC ATG GC GAG TGG TGC TGA GTG AGG TGA

The *ARX* gene was initially screened in the four ISSX families previously described (Bruyere *et al.*, 1999b; Strømme *et al.*, 1999; Claes *et al.*, 1997). Disease casing mutations were detected in three of them. The first mutation found was an expansion of a $(GCG)_n$ repeat, that is normally present in ten copies and was expanded to 17 copies $[(GCG)_{10+7})]$ in the patients with ISSX (Figure 5.3). This $(GCG)_{10}$ encodes a stretch of alanine residues that makes up part of a polyalanine tract normally consisting of 16 alanine residues. The insertion of the extra 21 bp in the (GCG) repeat results in a stretch of 23 alanine residues in the ISSX patient. This identical mutation was found in one other unrelated ISSX family.

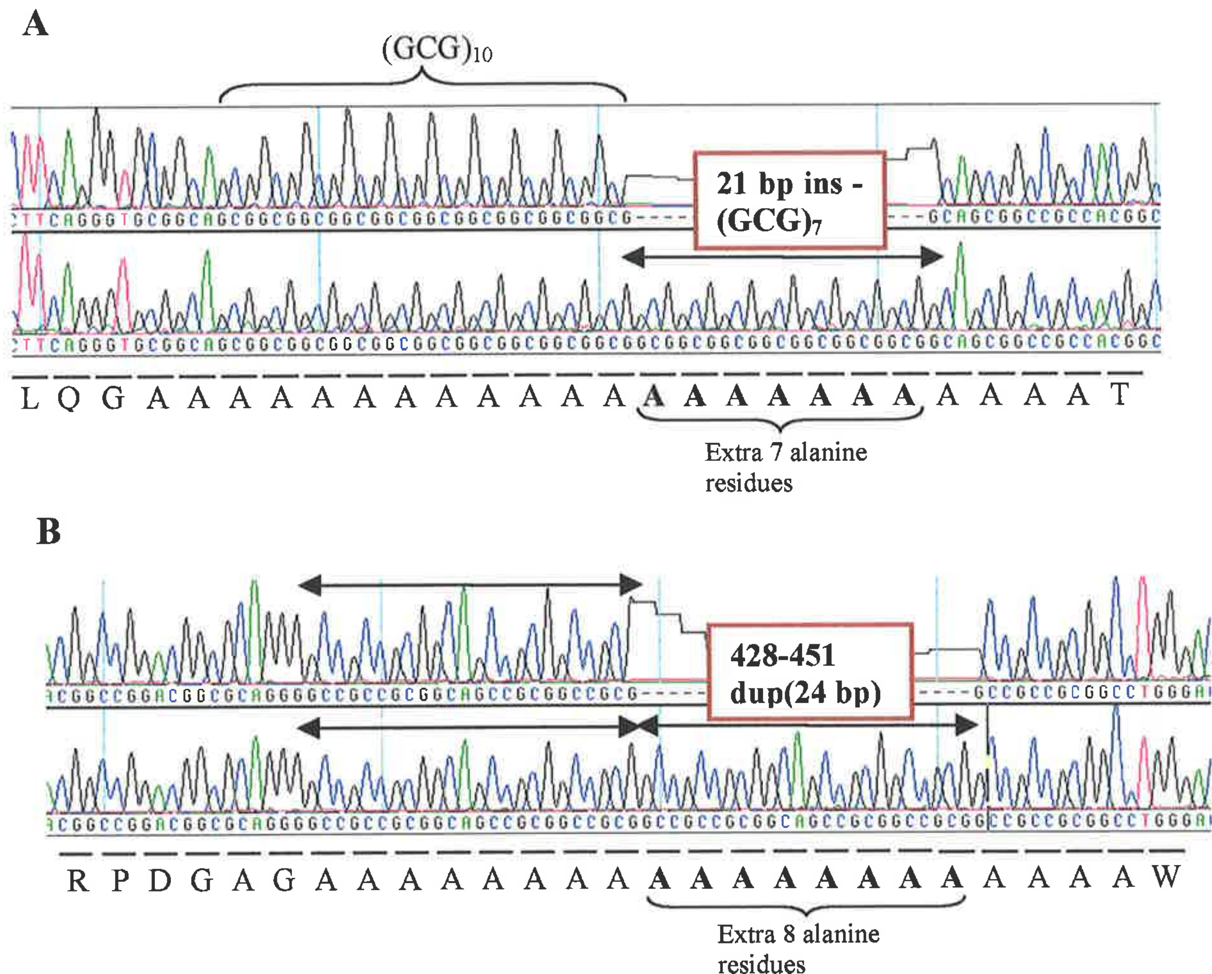


Figure 5.3: Partial sequence chromatograms aligned by the Seqman DNA analysis programme (Lasergene, DNA Star) from normal and ISSX patient DNA from exon two of the *ARX* gene. **A:** In two unrelated ISSX families comparison of normal (top) and patient (bottom) DNA sequence revealed an insertion of 21 bp (indicated by an arrow). This mutation alters a $(GCG)_{10}$ repeat to $(GCG)_{10+7}$. This sequence encodes a polyalanine tract and the mutation changed the normal stretch of 16 alanine residues to 23. **B:** In a third ISSX family comparison of normal (top) and patient (ISSX) sequence revealed a 24 bp duplication within a second sequence which also encodes a polyalanine tract. This duplication changed the normal stretch of 12 alanine residues to 20 in the patient. Arrows show the duplicated 24 bp.

A second mutation was found in a third ISSX family. In this case an in frame duplication of 24 bp [428-451dup(24 bp)] was identified within a second polyalanine tract in exon two (Figure 5.3). This 24 bp also encodes alanine residues with the normal protein containing a stretch of 12 alanine residues and the 24 bp duplication resulting in 20 alanine residues in the protein of the ISSX patients.

Neither of these changes was found in over 300 normal X chromosomes screened. The $(GCG)_{10}$ repeat is not polymorphic within the normal population. Hence it was concluded that these changes were the cause of ISSX. In the fourth ISSX family (Claes *et al*, 1997, family B) no change was identified, suggesting genetic heterogeneity, or a mutations in a non coding region of *ARX* which had not been sequenced.

A fifth patient from a small unmapped family was then screened. The proband from this family had a more severe phenotype with severely delayed development, and had commencement of seizures at 4.5 weeks, which progressed to infantile spasms and hypsarrhythmia. MRI showed unilateral flattening of the skull, right microphthalmia and delayed myelenation. A maternal uncle died at 20 months with infantile spasms and severe developmental delay (Strømme *et al.*, 2002a). PCR amplification and sequencing of exons from this patient showed normal sequence from exons one to four. A PCR product from exon five, however, was never obtained. This suggested that all or part of this exon was deleted, which was confirmed by DNA digestion and Southern blot hybridisation (Figure 5.4D). Using oligonucleotides located downstream from exon five a PCR product was obtained that crossed the deletion. Sequencing revealed that the deletion was 1,517 bp (IVS4-816_EX5701del) (Figure 5.4E) that removes 816 bp of intron four and 701 bp of exon five that is

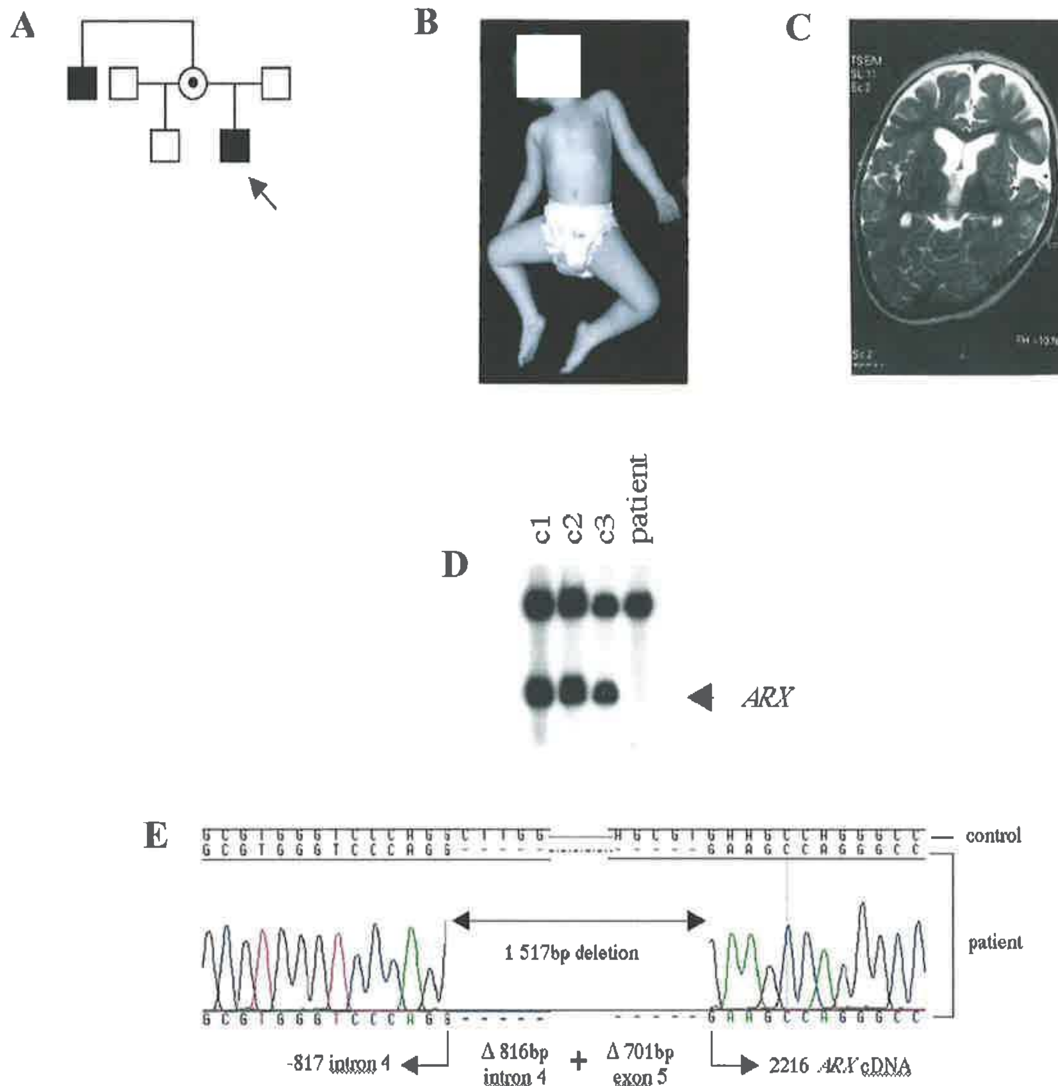


Figure 5.4: Screening of a fifth patient with a more severe form of ISSX. **A:** The proband is from a small family, and has a maternal uncle who died at 18 months with severe developmental delay and infantile spasms. **B:** The proband has some dysmorphic features and MRI (**C**) shows unilateral flattening of the skull, right microphthalmia and delayed myelination (courtesy of Dr P Strømme). A PCR product was not obtained from exon 5 in this patient. Southern Blot hybridisation (**D**) of digested patient and normal (c1, 2 and 3) DNA confirmed that this was due to a deletion. **E:** Sequence chromatogram across the deletion in this patient revealed a 1,517 bp deletion from position 816 in intron 4 to position 701 of the mRNA sequence in exon five.

predicted to result in an alternative carboxy terminus of the ARX protein that lacks the *aristaless* domain.

5.2.3 *ARX screening of XLMR families*

Mutations that cause MRXS, have also been shown to cause both NSXLMR and other forms of MRXS. Therefore a mutation screen was carried out on other available families that mapped to the same region as *ARX*. These include

1. Twelve families with NSXLMR. These include MRX-M and MRX-E from which blood samples of family members were collected by Prof Gillian Turner (Hunter Genetics, New South Wales, Australia) and where linkage analysis had been carried out by Dr Agi Gedeon (Department of Cytogenetics and Molecular Genetics, WCH). Probands from an additional ten mapped NSXLMR families were kindly provided by Dr Charles Schwartz (Greenwood Genetic Center, Greenwood, USA).

2. Two families with Partington syndrome (OMIM # 309510), which is characterised by MR and dystonic movements of the hands (Partington *et al.*, 1988; Frints *et al.*, 2002a).

3. One family with myoclonic epilepsy, intellectual disability and spasticity (XMESID) (screening performed by Dr Petter Strømme) (Scheffer *et al.*, 2002).

The same 428-451dup(24 bp) mutation that was first found in the ISSX family of Strømme *et al* (1999) was identified in MRX-M, MRX29, MRX32 and MRX33 and both families with Partington syndrome (Strømme *et al*, 2002 and unpublished data).

In the XMESID family a C→T transition was identified at position 1058 (Figure 5.5), which is predicted to change a proline at position 353 to leucine (P353L). This residue is highly conserved and is one of the six invariant residues found in the paired-type homeodomain. This mutation destroys a *Hpa*II site so that PCR amplification across the change and restriction digestion is able to differentiate between normal and mutation bearing chromosomes. The 1058C→T was shown to segregate with the phenotype, and was not found in over 100 normal chromosomes screened.

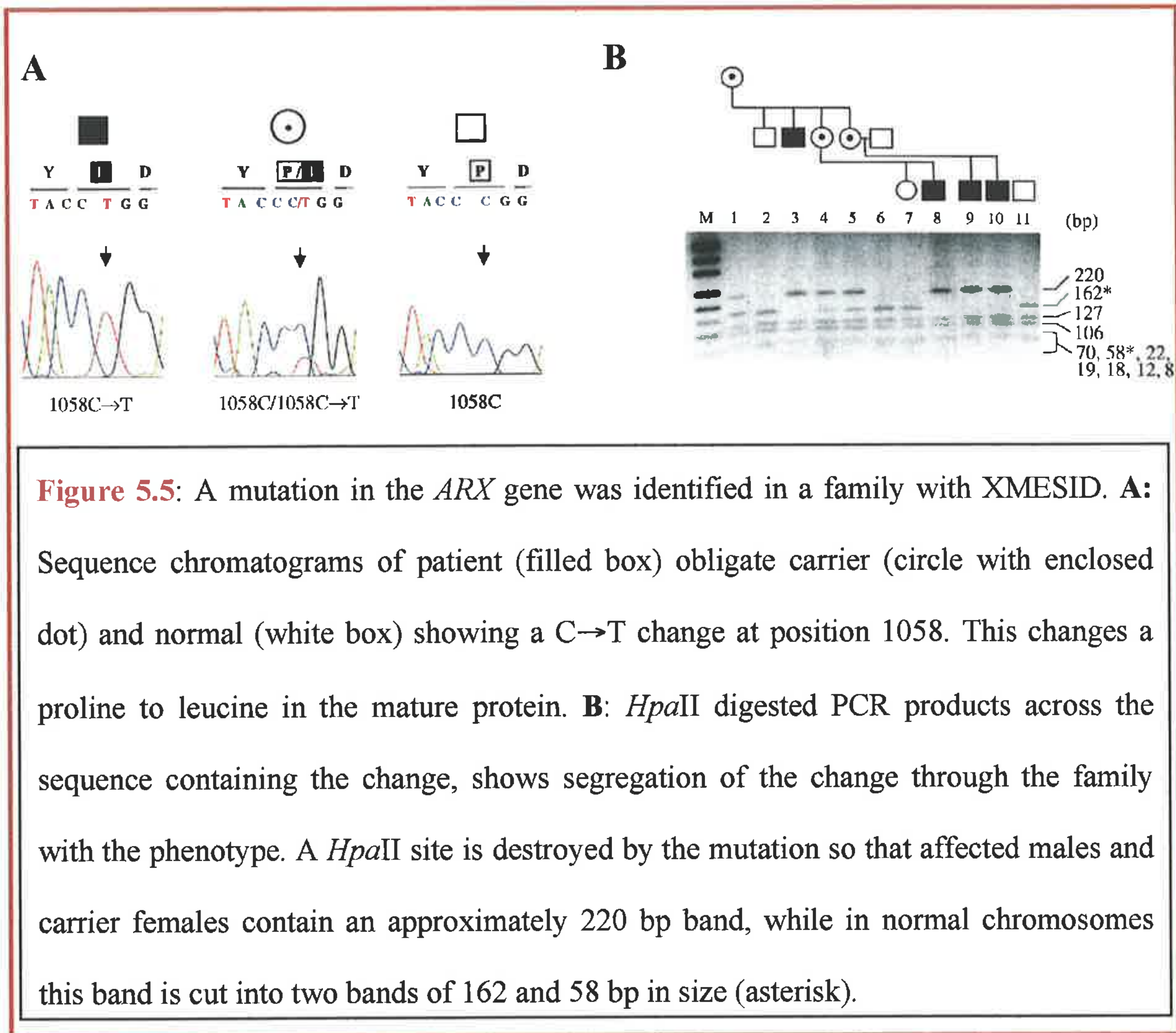


Figure 5.5: A mutation in the *ARX* gene was identified in a family with XMESID. **A:** Sequence chromatograms of patient (filled box) obligate carrier (circle with enclosed dot) and normal (white box) showing a C→T change at position 1058. This changes a proline to leucine in the mature protein. **B**: *Hpa*II digested PCR products across the sequence containing the change, shows segregation of the change through the family with the phenotype. A *Hpa*II site is destroyed by the mutation so that affected males and carrier females contain an approximately 220 bp band, while in normal chromosomes this band is cut into two bands of 162 and 58 bp in size (asterisk).

5.2.4 ARX gene expression, gene function and conservation

ARX is a member of the *aristaless*-related paired type homeodomain proteins. The homeodomain is the DNA-binding domain found in a large variety of transcription factors involved in controlling cell fate decisions and development. The paired class is characterised by six invariant amino acids in the homeodomain, with a serine (PAX type), glutamine (Q_{50} paired-type) or lysine at position (K_{50} paired-type) at position 50, with ARX falling into the Q_{50} class. The *aristaless*-related proteins also have a C-terminal domain referred to as the

aristaless domain (also known as the OAR domain, C-peptide or paired tail). This domain has been proposed to suppress activity of the transcription factor via an effect on DNA binding (Brouwer *et al.*, 2003).

The human ARX protein has 94.3% and 57.2% similarity to its mouse and zebrafish orthologs respectively (Figure 5.6). The octapeptide, nuclear localisation signal, and homeodomain are identical in all three species. The *aristaless* domain is identical between human and mouse and 87% similar between human and zebrafish. The polyalanine tracts are not conserved between the orthologs, with the human and mouse having different numbers of alanine residues in both polyalanine stretches, while both are completely absent in zebrafish. These polyalanine tracts have been proposed to have a function in transcriptional repression (Han & Manley, 1993).

```
                                                      OP
H. sapiens     MSNQYQEEGC SERPECKSKS PTLLSSYCID SILGRRSPCK MRLLGAAQSL PAPLTSRADP  60
M. musculus    MSNQYQEEGC SERPECKSKS PTLLSSYCID SILGRRSPCK MRLLGAAQSL PAPLASRADQ  60
D. rerio       MSSQYDDDS- RDRSECKSKS PTVLSSYCID SILGRRSPCK VRQLGA-QSL PAPV--RPDH  56

                 1▼2                   NLS
H. sapiens     EKAVQGSPKS SSAPFEAELH LPPKLRRLYG PGGGRLLQGA AAAAAAAAAA AAAA-ATATA  119
M. musculus    EKAMQGSPKS SSAPFEAELH LPPKLRRLYG PGGGRLLQGA AAAAAAAAAA AAAATATGTA  120
D. rerio       EMTTEVTSKE NS--FDSDMH LPPKLRRLYG PGG-KYLDSG RGFHEHLEK- ----------  102

H. sapiens     GPRGEAPPPP PPTARPGERP DGAGAAAAAA AAAAAAWDTL KISQAPQVSI SRSKSYRENG  179
M. musculus    GPRGEVPPPP PPAARPGERQ DSAGAVAAAA AAA--AWDTL KISQAPQVSI SRSKSYRENG  178
D. rerio       ---------- ------GERE RLLDQAC--- -------ESL KISQAPQVSI SRSKSYREN-  135

                                                               acidic domain
H. sapiens     APFVPPPPAL DELGGPGGVT HPEERLGVAG GPGSAPAAGG GTGTEDDEEE LLEDEEDEDE  239
M. musculus    APFVPPPPAL DELSGPGGVA HPEERLSAAS GPGSAPAAGG GTGAEDDEEE LLEDEEDEEE  238
D. rerio       APF------- ---------- ---------- -----SQSDE GQSEEHMAQE LVELSTLKFF  163

               acidic domain
H. sapiens     EEELLEDDEE ELLEDDARAL LKEPRRCPVA ATGAVAAAAA AA---VATEG GELSPKEELL  296
M. musculus    EEELLEDDDE ELLEDDARAL LKEPRRCSVA TTGTVAAAAA AAAAAVATEG GELSPKEELL  298
D. rerio       EDVVKEEA-- ---------- ------CG-- ---------- ---------D NSLSPKDEES  184

                                                       Homeodomain
H. sapiens     LHPEDAEGKD GEDSVCLSAG SDSEEGLLKR KQRRYRTTFT SYQLEELERA FQKTHYPDVF  356
M. musculus    LHPEDAEGKD GEDSVCLSAG SDSEEGLLKR KQRRYRTTFT SYQLEELERA FQKTHYPDVF  358
D. rerio       LH-NDGDVKD GEDSVCLSAG SDSEEGMLKR KQRRYRTTFT SYQLEELERA FQKTHYPDVF  243

              2▼3          3▼4 Homeodomain
H. sapiens     TREELAMRLD LTEARVQVWF QNRRAKWRKR EKAGAQTHPP GLPFPGPLSA THPLSPYLDA  416
M. musculus    TREELAMRLD LTEARVQVWF QNRRAKWRKR EKRGAQTHPP GLPFPGPLSA THPLSPYLDA  418
D. rerio       TREELAMRLD LTEARVQVWF QNRRAKWRKR EKAGVQAHPT GLPFPGPLAA AHPLSHYLEG  303

H. sapiens     SPFPPH-HPA LDSAWTAAAA AAAAAFPSLP PPPGSASLPP SGAPLGLSTF LGAAVFRHPA  425
M. musculus    SPFPPH-HPA LDSAWTAAAA AAAAAFPSLP PPPGSASLPP SGAPLGLSTF LGAAVFRHPA  427
D. rerio       GPFPPHPHPA LESAWTAAAA AAAA-FPGLA PPPNSSALPP A-TPLGLGTF LGTAMFRHPA  361

                      4▼5
H. sapiens     FISPAFGRLF STMAPLTSAS TAAALLRQPT PAVEGAV-AS GALADPATA- ---AADRRAS  530
M. musculus    FISPAFGRLF STMAPLTSAS TAAALLRQPT PAVEGAV-AS GALADPATA- ---AADRRAS  532
D. rerio       FIGPTFGRLF SSMGPLTSAS TAAALLRQTA PPVESPVQPS AALPEPPSSS SSTAADRRAS  421

                        aristaless
H. sapiens     SIAALRLKAK EHAAQLTQLN ILPGTSTGKE VC                               562
M. musculus    SIAALRLKAK EHAAQLTQLN ILPGTSTGKE VC                               564
D. rerio       SIAALRLKAK EHSAQLTQLN ILPSGTAGKE VC                               453
```

Figure 5.6: Protein sequence alignment of human ARX with the mouse (*Mus musculus*) and zebrafish (*Danio rerio*) orthologs, with the amino acid positions of each shown on the left. The two polyalanine tracts in the human protein in which mutations have been identified are in bold. Mismatches are highlighted by shading and the conserved domains are circled. (OP = octapeptide, NLS = nuclear localisation sequence, exon-exon boundaries are shown by bold arrows).

Hybridisation of *ARX* probes to a human multiple tissue Northern blot, and a Northern blot containing RNA from brain sections (carried out by Marie Shaw), and EST analyses indicate that the human ortholog of *ARX* is expressed in foetal and adult brain (occipital, frontal and temporal lobes of the cerebral cortex, amygdala, corpus callosum, caudate nucleus and hippocampus). Expression was also detected in skeletal muscle where two additional smaller isoforms are also observed (Figure 5.7). The origins of these extra isoforms is yet to be determined, but as they are detected by probes from the 3' end of the gene and not the 5' end it is possible that they are due to alternative 5' ends of *ARX*. 5'RACE has been used to try to identify any alternative 5' end, but has not revealed any new transcript sequences (carried out by Lynne Hobson). This may be due to the high GC content of the gene, which may reduce the success of the RACE method. It is also possible that the extra bands seen on the Northern blot are the result of cross hybridisation to other genes with a high similarity to *ARX,* or to *ARX* pseudogenes. Searches of GenBank have not revealed any such sequences. It is therefore possible that the genomic structure of *ARX* presented here is not complete.

The mouse and zebrafish *Arx* genes were first described by (Miura *et al.*, 1997), and were shown to be expressed in the forebrain (cerebral cortex) and floorplate in the developing central nervous system. A probe was made by cloning part of the 3'UTR of the mouse *Arx* gene into pGEM-T (Promega), which contains SP6 and T7 promoters from which riboprobes could be made. The probe was hybridised to brain sections of foetal (11.5 dpc) and adult mouse (carried out by Dr Paul Thomas, Murdoch Institute, Melbourne, Australia). In the mouse embryo, expression was detected in the lateral and ventral telencephalic vesicles, as well as in the floor plate of the neural tube extending to the midbrain region (Figure 5.8). In adult mouse, there was an interesting pattern of expression within the cortical neurons, with expression of *Arx* in a subset of unknown nuclei.

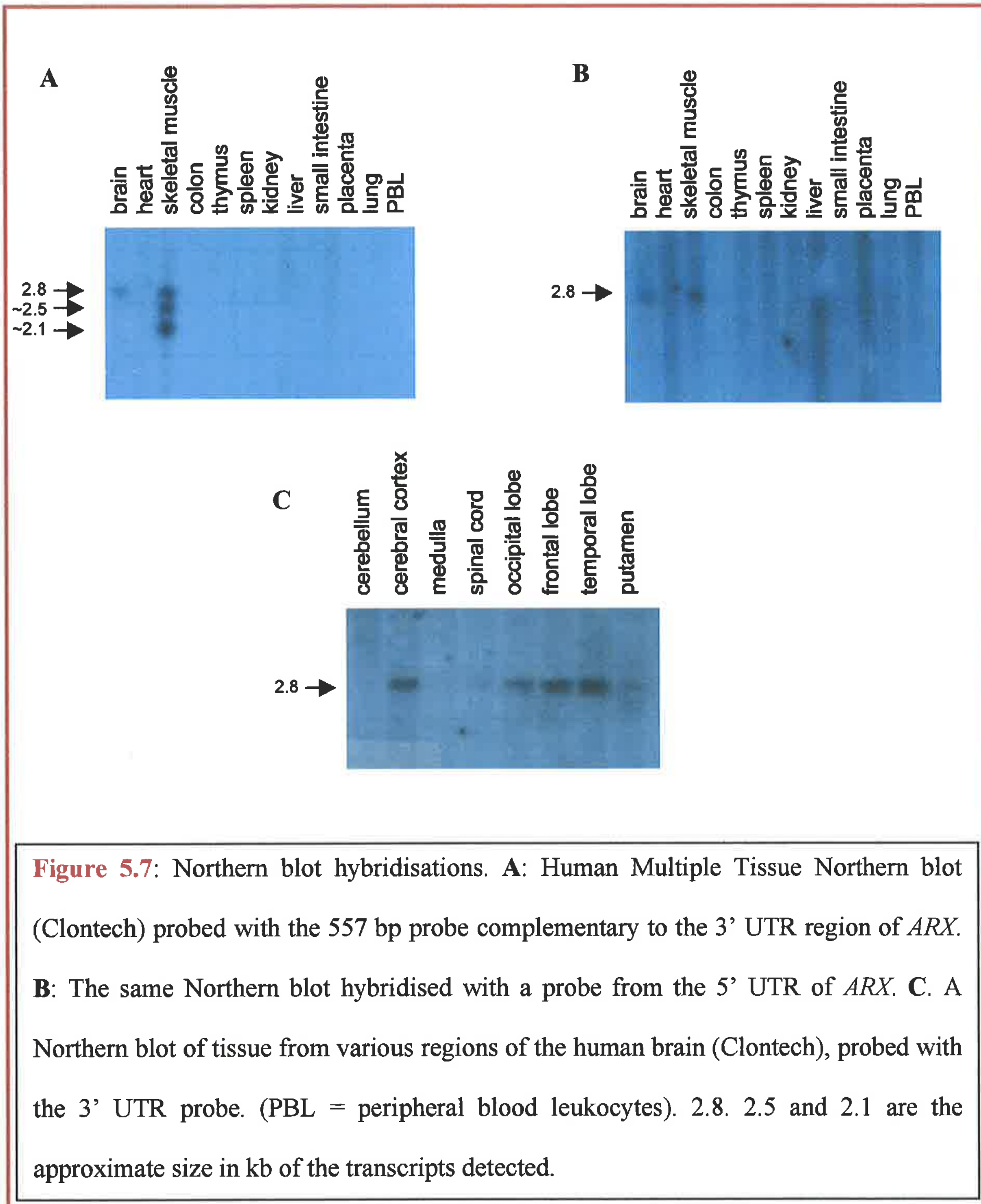


Figure 5.7: Northern blot hybridisations. **A**: Human Multiple Tissue Northern blot (Clontech) probed with the 557 bp probe complementary to the 3' UTR region of *ARX*. **B**: The same Northern blot hybridised with a probe from the 5' UTR of *ARX*. **C**. A Northern blot of tissue from various regions of the human brain (Clontech), probed with the 3' UTR probe. (PBL = peripheral blood leukocytes). 2.8. 2.5 and 2.1 are the approximate size in kb of the transcripts detected.

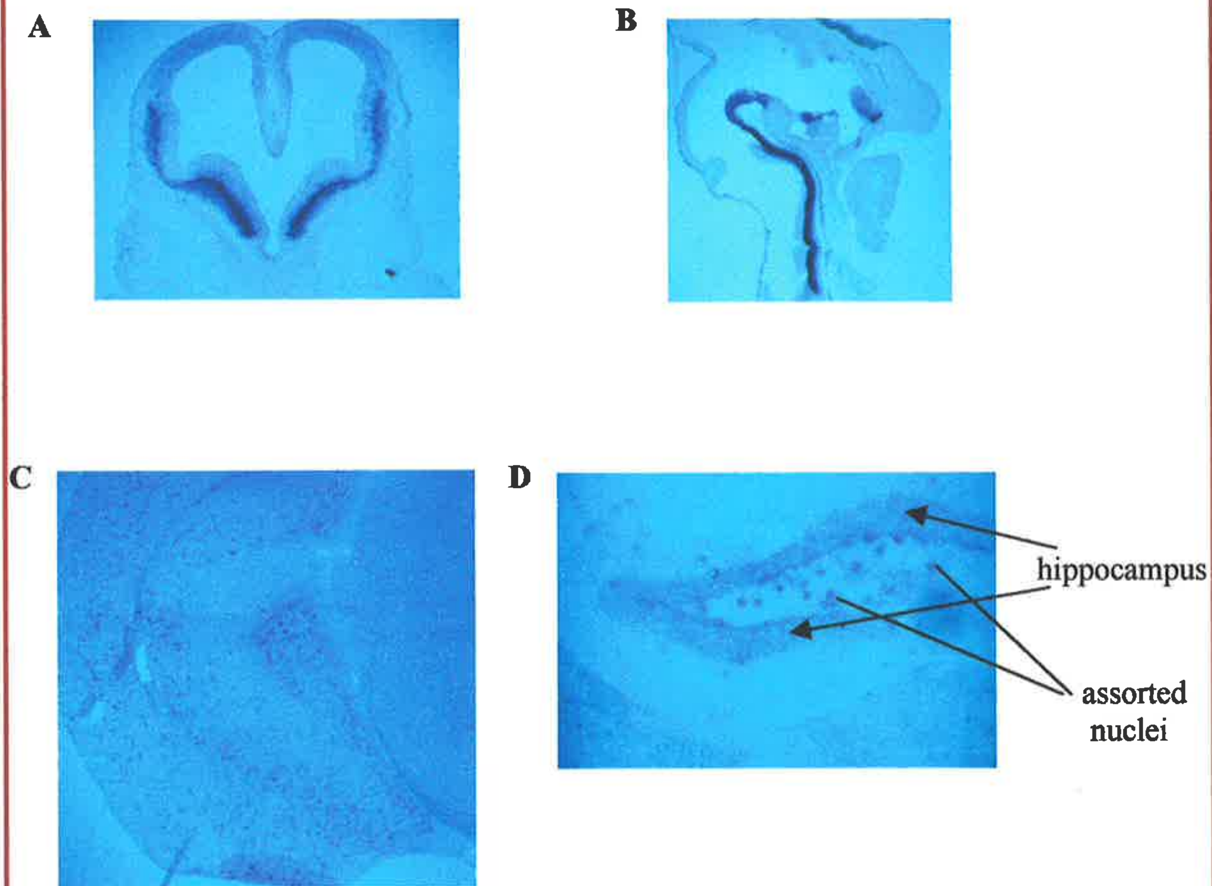


Figure 5.8: Mouse *in situ* hybridisations. A probe was made by PCR amplification and cloning of part of the 3' UTR of the mouse *Arx* gene (oligonucleotides used were F: `5'- AAGGCTGCCCTCCGCCCTTG - 3'` and R: `5'- GAAGAGAACAAGAACGAGACGC 3'`). **A** and **B** are mouse embryos at 11.5 dpc. **A** Is a sagital section and shows high *Arx* expression in the lateral and ventral telencephalic vesicles. **B** is a transverse section and shows *Arx* expression in the floor plate of the neural tube, extending to the midbrain region. **C** Adult mouse brain transverse section showing *Arx* expression in a subset of cortical neurons. **D**. The same section as C under high power showing expression in the hippocampus and in a subset of unknown nuclei.

5.2.5 *ARX screening in small families with putative XLMR*

(Assisted by Merran Finnis and Bree Hodgson)

The initial mutation screening in mapped XLMR families uncovered mutations in eleven out of twenty families, indicating that mutations in this gene account for a significant proportion of XLMR. This is especially the case with NSXLMR where until now, mutations in MRX genes have only been found in one or a few families. Of the more than 70 published MRX families, mutations in *ARX* now account for 5 (results presented here and by (Bienvenu *et al.*, 2002). This prompted the screening of other XLMR. These included

1. Families in which MR was segregating in an X-linked manner, where a significant lod score of >+2 could not be reached. This was either due to the family not having enough members to reach a significant lod score as predicted by SLINK, or where the family was large enough but DNA was not available from enough members of the family for a definitive linkage study.

2. Small families with X-linkage suggested by at least uncle/nephew pairs of affected males.

3. Families with two or more affected brothers where X-linkage was uncertain.

In total 48 families were screened by PCR amplification of exons, followed by analysis by SSCA. From this screen four additional families with the 428-452dup(24 bp) mutation were identified. The first (MRX-B) was a large family with MR associated with hypsarrhythmia (Turner *et al.*, 2002). The last three mutations were in small families with mild MR (families

MRX-D, MRX-R and MRX-Mo; these families were seen for clinical evalutaion and DNA samples provided by Prof G. Turner and M. Partington).

In the MRX-D family, the mutation was initially detected in the proband and was subsequently shown to also be carried by his affected brother. Testing of DNA from blood leukocytes from the mother showed that she did not carry the mutation. Haplotype analysis showed that both boys had inherited the same X chromosome for this region from their mother. This suggests that rather than each occurring *de novo*, that the mother is gonadal mosaic for the 428-452dup(24 bp) mutation. This is the first case described of gonadal mosaicism in a carrier female in the *ARX* gene.

In the sample of small, unmapped XLMR families mutations in *ARX* are the underlying cause of MR in 8% of cases. Again this highlights that mutations in *ARX* are the cause of a significant proportion of XLMR, and at this time mutations in *ARX* appears to be the second most common cause of XLMR after fragile X syndrome.

5.2.6 *ARX screening of patients with developmental delay that tested negative for FRAXA*

(Assisted by Karen Lower and Rachael Bennett)

ARX was then screened for mutations in patients with MR who had previously tested negative for expansion of the $(CCG)_n$ repeat at the *FRAXA* fragile site (WCH, Australia). A panel of five hundred DNA samples was collected for screening by SSCA. Initially the region of exon two containing the $(GCG)_{10}$ repeat and the 24 bp duplicated region was screened by electrophoresis of PCR amplicons on denaturing acrylamide gels, as this was a simple and rapid way to detect changes in the copy number of alanines in both of these polyalanine tracts.

This identified one change in alanine copy number in one patient (family MR-N) described below. The screening of these 500 samples by SSCA in ongoing, however, as of yet no single base changes have been identified.

In one case a male (family MR-N) who presented with mild MR was shown to contain the change $(GCG)_{10+1}$, that resulted in one extra alanine residue in the first polyalanine tract (a change of sixteen to seventeen alanine residues). DNA was collected from the proband's siblings, mother and grandmother. Analysis of the $(GCG)_{10}$ repeat in this family revealed that the $(GCG)_{10+1}$ allele was also carried by three of the proband's unaffected brothers, but not by his sister who is more severely affected than the proband (Figure 5.9).

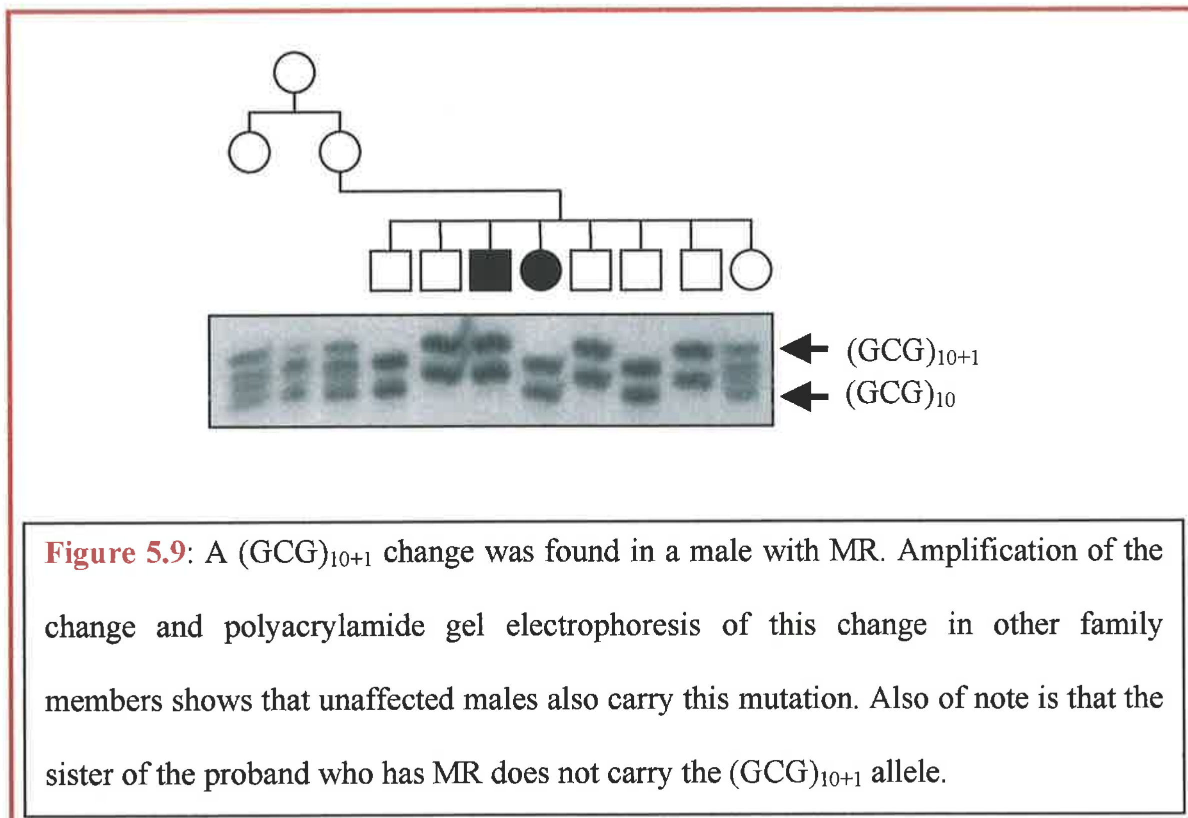


Figure 5.9: A $(GCG)_{10+1}$ change was found in a male with MR. Amplification of the change and polyacrylamide gel electrophoresis of this change in other family members shows that unaffected males also carry this mutation. Also of note is that the sister of the proband who has MR does not carry the $(GCG)_{10+1}$ allele.

5.2.7 Mammalian cell transfection of normal ARX and ARX with an expanded polyalanine tract

Similar to ARX, polyalanine tract expansions have been observed in a number of other disorders. In the case of oculopharyngeal muscular dystrophy (OPMD) it had been shown that expanded alanines tracts form intranuclear aggregation of PABP2 protein both in patient cells as well as in transfected cultured cells (Calado *et al.*, 2000). Another study has shown that when alanine residues were fused in frame with GFP (7, 19, 25 and 37 alanine residues) cytoplasmic inclusions formed after 24 hours when 19, 25 and 37 polyalanine stretches were present (Rankin *et al.*, 2000). Cells transfected with these constructs were also more prone to death.

Therefore, it is possible that the polyalanine expansions in ARX may also result in the aggregation of ARX protein. If this were the case it may be part of the mechanism by which these polyalanine expansions result in MR. To determine if the expansions of the polyalanine tracts were also causing ARX protein aggregation, the ORF was fused in frame with GFP and transfected into HeLa cells using the phrGFP vector (Promega).

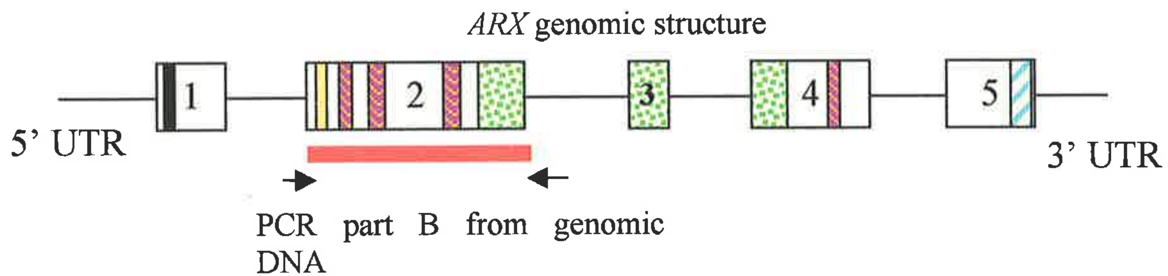
It was not possible to amplify the entire ORF in one reaction, probably due to the high GC content of the gene. Therefore the construct was cloned in three separate parts, which were later joined via unique restriction sites within the *ARX* ORF (Figure 5.10). Each part was first cloned into the PCR vector, pGEM-T (Promega) and was sequence verified before they were linked together. The first part was amplified from brain cDNA, and extended from the start codon into exon two. For the second part, exon two was amplified from normal and patient genomic DNA. As *ARX* was not expressed in the available patient material (cultured

lymphoblastoid cell lines, and fibroblasts) insertion of this segment would allow cloning of mutations found in XLMR patients, as it contained both regions encoding polyalanine tracts that mutations had been found in. The third part was amplified from brain cDNA, and extended from exon two up to the stop codon.

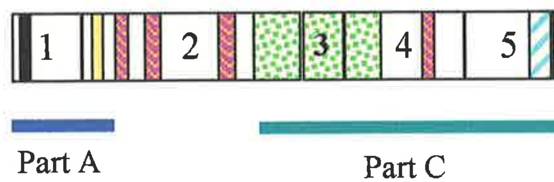
Initially plasmids were transformed into XL1Blue *E. coli* cells. The $(GCG)_{10+7}$ and the duplication from exon two, were both deleted in these cells. A cell line deficient in recombination pathways (SURE 2 - Stratagene) and thus less likely to delete difficult to clone sequences, was then used. The 428-451dup(24 bp) was still deleted from the clones obtained; however, both the normal ORF and that containing the $(GCG)_{10+7}$ mutation were successfully cloned.

The outermost primers used to amplify the insert were constructed with the addition of *Bgl*II and *Eco*R1 sites at the 5' and 3' ends of the *ARX* insert respectively. This was so that the full length inserts could be easily cloned into the multiple cloning site of the phrGFP-N1 vector (Stratagene). *EcoN*I and *Pst*I sites were used to join the three parts together (Figure 5.10). This resulted in two constructs containing the entire ORF in frame with the GFP. Both constructs were sequence verified before commencing with mammalian cell transfection. The constructs were then transfected into HeLa cells, were mounted with DAPI and were viewed by fluorescent microscopy.

Step 1: Amplify part B from genomic DNA from both normal DNA and DNA containing polyalanine expansion mutations (ISSX patients).



Step 2: Amplify parts A and C from normal brain cDNA



Step 3: Join parts A, B and C together using introduced restriction sites to clone into vector (*Eco*RI and *Bgl*II) and using unique restriction sites to join parts together (*Eco*NI and *Pst*I).

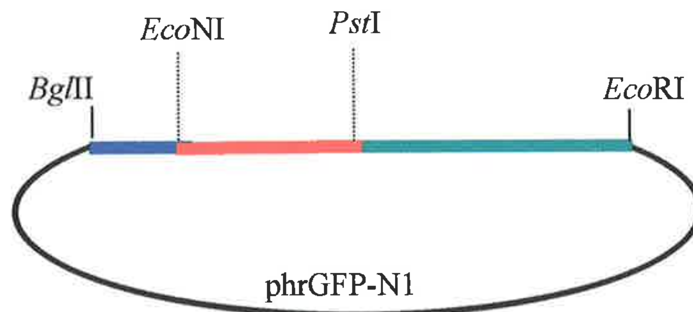


Figure 5.10: The method used for cloning the entire ORF of *ARX* fused in frame with GFP.

Table 5.2: Sequence of PCR primers used for cloning the ORF of *ARX*.

Part A	Forward Reverse	*BglII* (gaa gat ctA) gaa gat ctA TGA GCA ATC AGT ACC AGG TGC CGC ACC CTG AAG GAG GCG GCC C
Part B	Forward Reverse	ccg gaa ttc CAG CAG CGC CCC GTT CGA GG gcg gga tcc CGG GTA GTG CGT CTT CTG G
Part C	Forward Reverse	AAC TGC TGG AGG ACG ACG AGG ccg gaa ttc TTA GCA CAC CTC CTT GCC C *EcoRI* (gaa ttc)

The transfection efficiency of both constructs was very low when compared with empty vector. However, in the case of the normal construct, the protein localised tightly to the nucleus, and appeared to avoid the nucleoli. In the case of the construct containing the $(GCG)_{10+7}$ mutation, most of the protein was still localised within the nucleus, but there was signal apparent within the cytoplasm. In some cells, this signal appeared to be somewhat grainy (Figure 5.11). The transfection was then repeated, and the transfected cells were grown for 72 hours. The extra time made no difference on either the appearance of the cells or the GFP fluorescence. These were initial experiments and the results are therefore not conclusive.

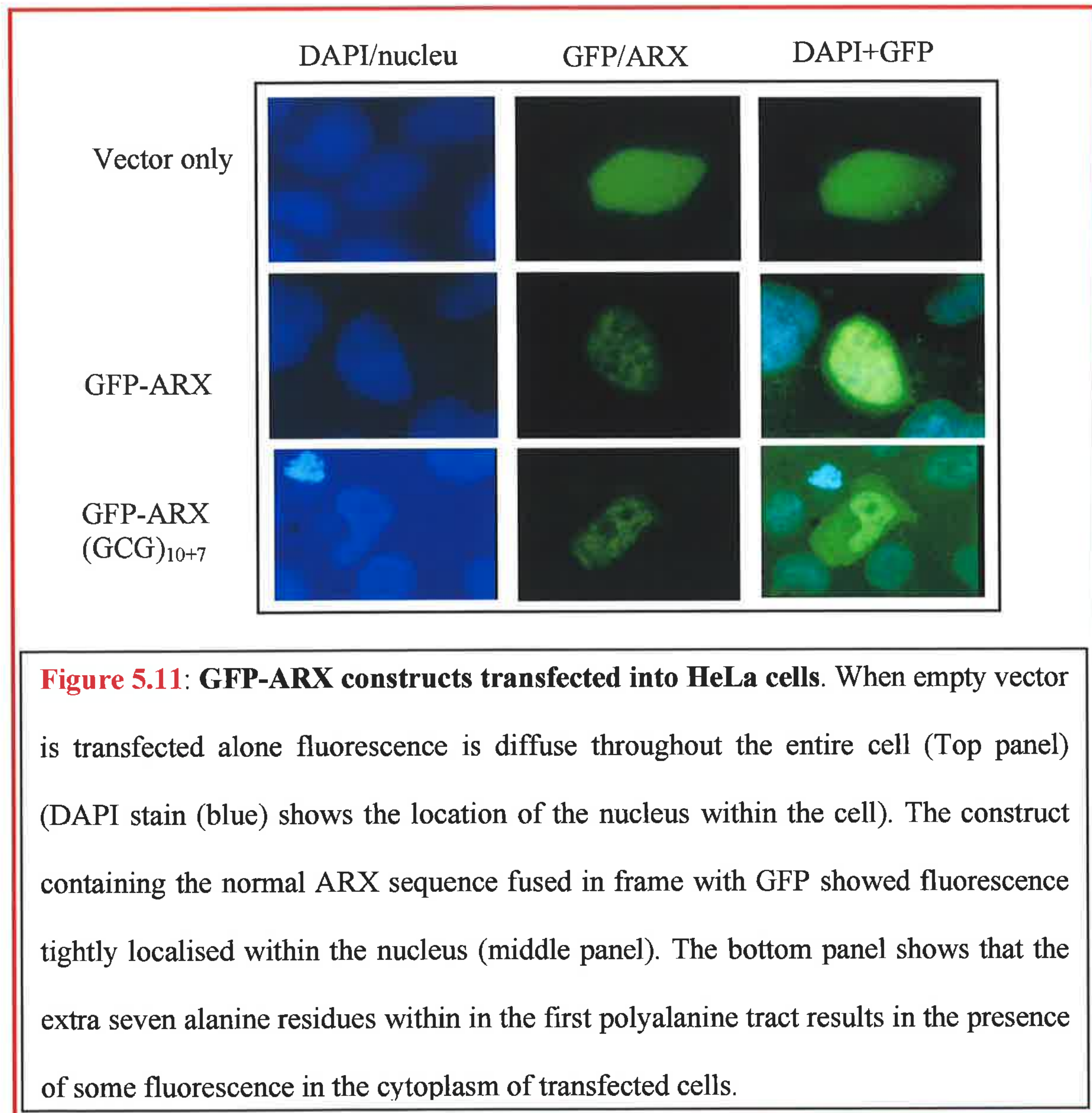


Figure 5.11: **GFP-ARX constructs transfected into HeLa cells**. When empty vector is transfected alone fluorescence is diffuse throughout the entire cell (Top panel) (DAPI stain (blue) shows the location of the nucleus within the cell). The construct containing the normal ARX sequence fused in frame with GFP showed fluorescence tightly localised within the nucleus (middle panel). The bottom panel shows that the extra seven alanine residues within in the first polyalanine tract results in the presence of some fluorescence in the cytoplasm of transfected cells.

5.3. DISCUSSION

Mutations in the *ARX* gene have been shown to cause ISSX, NSXLMR, PRTS and XMESID (Summary Table 5.3). The 16 mutations found include expansion of two different polyalanine tracts, one missense mutation, and one deletion resulting in a truncated protein lacking the *aristaless* domain. One change $(GCG)_{10+1}$ was also found but it remains unclear as to whether this has a role in disease susceptibility or if it is an innocent rare variant.

One ISSX family that had previously been localised to Xp21, was not found to contain any mutations with the *ARX* gene. This may be because the mRNA sequence of the gene that is so far known, is incomplete. This is suggested by the alternative transcripts seen in skeletal muscle on the Northern blot, which are yet to be accounted for. Alternatively, a mutation may reside within regulatory regions in the promoter or parts of the 5' and 3' UTRs that were not covered by the mutation screen. It is already known that disruption of a second gene (*STK9*) located close to the region containing *ARX* can cause a similar phenotype of MR with infantile spasms (Kalscheuer *et al.*, 2003). However, affected members of this family do not contain a mutation in this gene either indicating the possible presence of a third locus for MR and infantile spasms in Xp21, other than *ARX* and *STK9*.

Since the *ARX* gene and mutations causing XLMR was described (Strømme *et al.*, 2002b), other groups have also identified mutations in *ARX* that cause XLMR. Several mutations in patients with NSXMLR were found by Bienvenu *et al.*, (2002), and mutations causing XLAG (X-linked lissencephaly with abnormal genitalia) were found by Kitamura *et al.,* (2002). XLAG is a severe XLMR syndrome characterised by lissencephaly, agenesis of the corpus callosum, intractable epilepsy of neonatal onset, hypothalamic dysfunction and ambiguous

genitalia (OMIM# 300215). Additionally, female relatives of affected boys may have mental retardation and epilepsy, and they often display agenesis of the corpus callosum (Bonneau *et al.*, 2002). The additional mutations in *ARX* found by other groups are also summarised in Table 5.3, and are discussed below.

5.3.1 Insertion in the first polyalanine tract of the ARX gene

This study has identified two families with ISSX with the $(GCG)_{10+7}$ mutation in the first polyalanine tract. This mutation results in the largest polyalanine tract reported thus far in the *ARX* gene, in terms of the number of alanine residues in the tract that result from the mutation.

Additionally, insertion of only one extra (GCG) within this tract was seen in one family (MRX-N). The proband of this family was initially tested for expansion of the $(CCG)_n$ repeat at the *FRAXA* locus, and proved to be negative for this mutation. Screening of *ARX* in the proband and his family members has revealed that this mutation is also carried by the proband's unaffected brothers, but not by his sister who also has MR. It is therefore unlikely that this mutation is causing MR in the proband.

One family with a similar mutation in the first polyalanine tract has been described in which a $(GCG)_{10+2}$ mutation segregates with MR. This results in an extra two alanine residues within this polyalanine tract (Bienvenu *et al.*, 2002). In over three hundred normal X chromosomes tested during the course of this project the $(GCG)_{10}$ repeat in which these insertions occur, is invariant. This therefore raises the question of whether or not the $(GCG)_{10+1}$ allele in the MR-N family should be regarded as a disease susceptibility change or a neutral rare variant. If it is

Table 5.3: Summary of mutations found in the *ARX* gene in patients with XLMR. Blue and red writing indicates mutations in the first and second polyalanine tracts respectively.

Mutation/DNA	**Mutation/protein**	**Family**	**Reference**
	This study		
$(GCG)_{10+7}$	A16→23/polyalanine expansion	ISSX	(Bruyere *et al.*, 1999b)
		ISSX	(Claes *et al.*, 1997)
$(GCG)_{10+1}$	A16→A17/ polyalanine expansion	MR-N	E. Hahn (unpublished)
428-451dup(24 bp)	A12→A20/ polyalanine expansion	ISSX	(Strømme *et al.*, 1999)
		PRTS1	(Partington *et al.*, 1988)
		PRTS2	(Frints *et al.*, 2002a)
		MRX-M	(Turner *et al.*, 2002)
		MRX-B	(Turner *et al.*, 2002)
		MRX-D	G. Turner (unpublished)
		MRX-Mo	G. Turner (unpublished)
		MRX-R	G. Turner (unpublished)
		MRX29	(Hane *et al.*, 1996)
		MRX32	(Hane *et al.*, 1999)
		MRX33	(Holinski-Feder *et al.*, 1996)
1058C→T	P353L/missense	XMESID	(Scheffer *et al.*, 2002)
IVS4-816_EX5701del	R483fs/truncation	ISSX	(Strømme *et al.*, 2002b)

NSXLMR (Bienvenu *et al*, 2002)

98C→T	L33P	MRX54
$GCG_{(10+2)}$	A16→A18 polyalanine expansion	T80
428-451dup(24 bp)	A12→A20/polyalanine expansion	*P49*
		P73
		MRX36
		N52
		MRX43
		SC
448del9	A12→A9 polyalanine deletion	T6
429del24	A12→A4 polyalanine deletion	L45
	G286S	P25
490A→G	Q163R	T4

XLAG (Kitamura *et al*, 2002)

420-451del32	frameshift/truncation	P1
790delC	frameshift/truncation	P2
1372delG	frameshift/truncation	*P7*
995G→A	R332H (missense)	P3
1028T→A	L343Q (missense)	P8/9
1188insC	frameshift/truncation	P5
1117C→T	nonsense	P4
del exons 1 and 2	unknown	P6

a susceptibility allele, it may be that other factors such as genetic background and environment may need to act in combination with the $(GCG)_{10+1}$ allele to create an affected phenotype. In the case of family MR-N, it has been suggested that there are members of the family on the grandmaternal side, who have learning difficulties, and it would be of interest to see if these individuals also carry the $(GCG)_{10+1}$ allele. If one extra alanine residue in this tract does not cause disease, then is it also possible that the $(GCG)_{10+2}$ change is also not the true cause of MR in the family described, and that this is also just a rare variant which is segregating with MR in the family. A study of the effect of these two changes have on the normal function of the protein will be needed in order to determine if they are true disease causing changes.

5.3.2 Duplication and deletion of the second polyalanine tract of the ARX gene

The 428-451dup(24 bp) mutation was found in a large proportion of the XLMR tested. It is interesting to note that this single mutation can result in such a wide range of phenotypes. Although all individuals with this mutation have some degree of MR, the association with other features is variable. This includes one family with ISSX (infantile spasms, hypsarrhythmia, MR), two families with PRTS (MR with dystonic movements of the hands), and several cases of NSXLMR. In the case of the latter, further clinical investigation of some of these families has revealed intrafamilial variation, with different members of the family also displaying some of the features of ISSX or PRTS (Turner *et al.*, 2002). These families were originally classified as NSXLMR, as the presence of the extra clinical features was not consistent among affected males in the pedigrees.

Family MRX-M highlights the intrafamilial phenotypic variation in families with the 428-451dup(24 bp) mutation (Figure 5.13). In this family affected males II-6 and III-5 presented with NSXLMR, individual II-5 had infantile spasms associated with MR, while individual III-1 has the most severe phenotype and was originally diagnosed as autistic and has dystonic movements of the hands (Turner *et al.*, 2002). Hence, based on the fact that the additional clinical features were not consistent among affected males in this family, the MR was classified as NSXLMR.

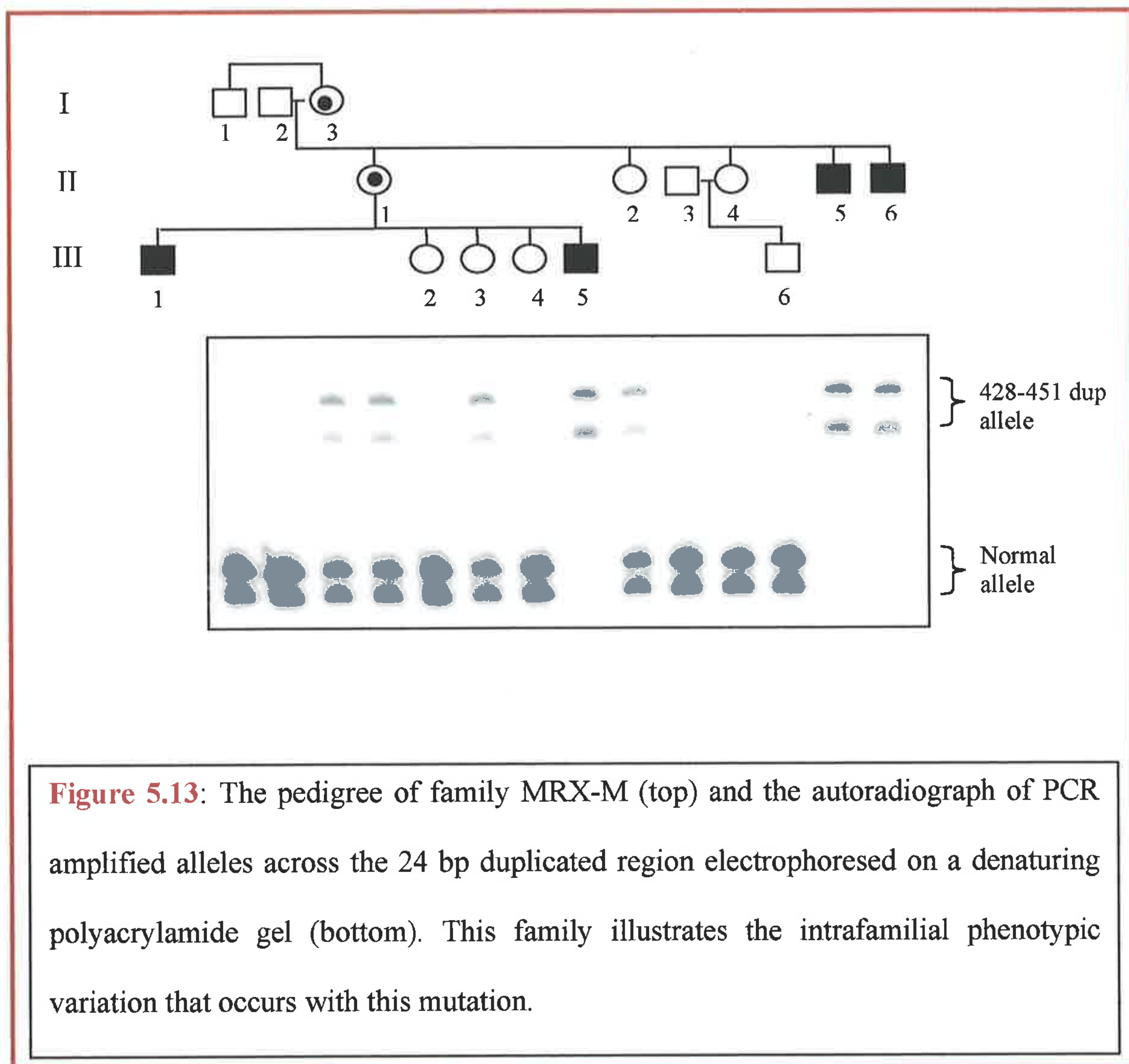


Figure 5.13: The pedigree of family MRX-M (top) and the autoradiograph of PCR amplified alleles across the 24 bp duplicated region electrophoresed on a denaturing polyacrylamide gel (bottom). This family illustrates the intrafamilial phenotypic variation that occurs with this mutation.

Two in frame deletions within the second polyalanine tract of *ARX* have also been identified by Bienvenu *et al.,* (2002), resulting in four and nine alanine residues instead of the normal twelve. In the family with nine alanine residues, the mutation segregated with MR in the family. However in the family with four alanine residues, one unaffected boy was identified who carried the deletion. This again raises the problem of proving variants observed in MR families are in fact disease causing mutations. Such deletions were not found in the hundreds of normal chromosomes that have been screened. It may be that genetic background and/or environment again needs to be involved as a possible explanation for the observed phenotypic anomaly. Alternatively, it may be that in fact they are not disease causing changes. The effect these mutations have on the normal function of the protein needs to be studied in order to elucidate this. Meanwhile, ascertainment of additional MR families segregating the same variant may provide clarification.

5.3.3 Missense mutations in the ARX gene

The missense mutation P353L identified in homeodomain in affected males from the family with XMESID, changes a highly conserved proline residue in the paired-type homeodomain protein to a leucine (Figure 5.12). In some homeodomain proteins however, a leucine residue appears at this position. This change was not found in 100 normal X chromosomes screened. Although on the basis of cross species conservation it is likely that this is the disease causing mutation in this family, functional studies on the effect of this change on the ARX protein is needed for final proof.

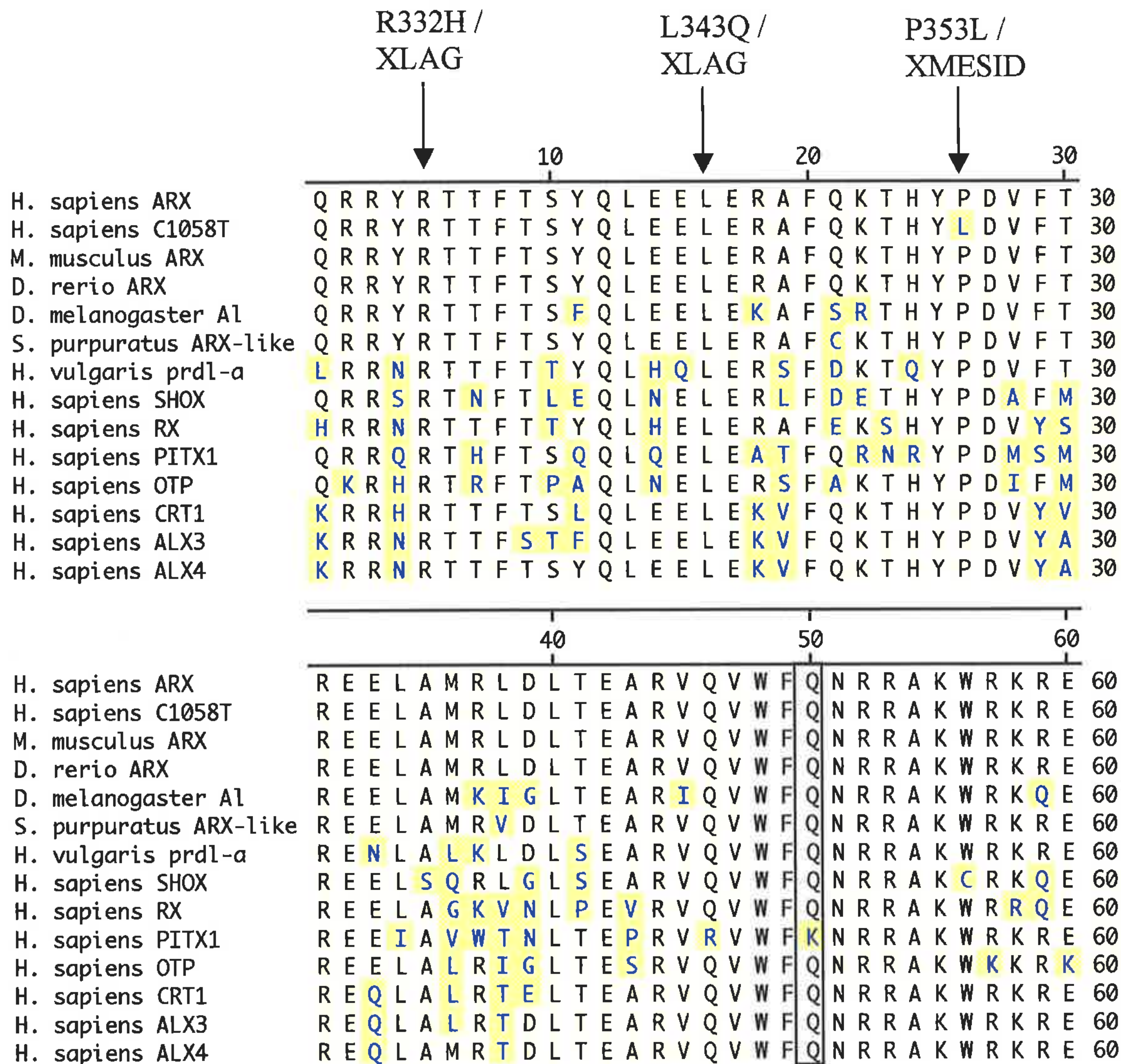


Figure 5.12 Alignment of the homeodomains of human, mouse and zebrafish ARX proteins with other members of the paired-type homeodomain proteins, and that of the XMESID P353L change. This changes a residue that is conserved in this group of proteins. Differences in proteins sequence are highlighted in yellow. Arrows show the locations of the missense mutations found in patients with XLAG (Kitamura *et al.*, 2002). The amino acid at position 50 is boxed and shows that all but PITX1 are members of the Q_{50} paired type domain.

Three missense mutations have also been described by Bienvenu *et al.,* (2002) and are listed in Table 5.3. These mutations all lie outside of the homeodomain. The first that results in a predicted L33P change in the protein sequence is in the octapeptide sequence. The second (Q163R) is located 8 residues away from the second polyalanine tract, and the third G286S falls between the acidic domain and the homeodomain. These mutations all cause the less severe phenotype of NSXLMR, however one family (MRX36) has recently been reviewed and has been found to have overlapping clinical features with PRTS (Frints *et al.*, 2002b). Additionally missense mutations have been found that change amino acid residues that are highly conserved within the paired-type homeodomain proteins (Figure 5.12). These changes (R332H and L343Q) cause the more severe phenotype of XLAG and suggest that these amino acids are highly important for the normal function of the ARX protein.

In the large subset of patients screened by SSCA presented in this thesis, no missense mutations were identified. This may be either because there are no missense mutations to be found in this group or it may reflect the inability of SSCA to identify all single base changes (Jordanova *et al.*, 1997). This may especially be a problem for genes such as *ARX* that have a high GC content, where single base changes might not be enough to alter the confirmation of the DNA such that it will be detected by non denaturing gel electrophoresis.

5.3.4 Truncation Mutations in the ARX gene

Six mutations in *ARX* that result in truncation of the protein have been identified. In all cases the phenotype is more severe than that seen in any other mutation. The first, which is described in this thesis was in a patient with severe, early onset ISSX and has a deletion from

within intron 4 into exon 5. This is predicted to produce a protein with an alternative carboxy terminus that lacks the *aristaless* domain.

Five truncation mutations (1 nonsense and four frameshift mutations), and one large deletion including exons 1 and 2 with an unknown affect on the protein, have been found in patients with XLAG (Kitamura *et al.*, 2002). This represents the most severe phenotype found in patients with mutations in *ARX*. All of the truncated proteins lack the entire *aristaless* domain, with two also lacking the whole of the homeodomain.

5.3.5 Other Polyalanine Expansion disorders

Several other disorders have been described in which expansions of polyalanine tracts within transcription factor genes causes the disorder (Table 5.4). These include synpolydactyly (*HOXD13*), cleidocranial dysplasia (*RUNX2*) holoprosencephaly (*ZIC2*), hand-foot genital syndrome (*HOXA13*), type II BPES (*FOXL2*), MR with growth hormone deficiency (*SOX3*) and congenital central hypoventilation syndrome (*PHOX2B*). The only expansion of a polyalanine tract in a gene other than a transcription factor is in *PABP2* causing OPMD. As with mutations in *ARX*, expansion of normal polyalanine tracts by one to 14 alanine residues causes disease in these cases.

It has been noted for synpolydactyly, that the size of the polyalanine expansion in HOXD13 has a relation to the penetrance and severity of the disorder (Goodman *et al.*, 1997). The longer expansion with 29 alanine residues has the most severe phenotype, while shorter expansions of 22 alanine residues had a less severe phenotype and in a number of cases is

Table 5.4: Polyalanine expansion disorders and alanine copy numbers of normal and expanded alleles.

Polyalanine Expansion Disorders				
Gene	Disorder	Normal Ala copy number	Expansion Ala copy number	Reference
HOXD13	Synpolydactyly	15	22-25, 29	(Akarsu *et al.*, 1996; Goodman *et al.*, 1997).
RUNX2	Cleidocranial dysplasia	17	27	(Mundlos *et al.*, 1997)
PABP2	Oculopharyngeal muscular dystrophy (OPMD)	10	12-17	(Brais *et al.*, 1998)
	Autosomal recessive OPMD	10	11	(Brais *et al.*, 1998)
ZIC2	Holoprosencephaly	15	25	(Brown *et al.*, 1998)
HOXA13	Hand-foot genital syndrome	14	24	(Goodman *et al.*, 2000; Utsch *et al.*, 2002)
FOXL2	Type II blepharphimosis/ epicanthus/ptosis	14	24	(Crisponi *et al.*, 2001)
SOX3	XLMR with growth hormone deficiency	15	26	(Laumonnier *et al.*, 2002)
PHOX2B	Congenital central hypoventilation	20	25-29	(Amiel *et al.*, 2003)
ARX	ISSX/NSXLMR	16	23 (17,18)	This thesis; Strømme *et al.*, 2002b; Bienvenu *et al.*, 2002)
	ISSX/NSXLMR	12	20	

non-penetrant. It was also noted that intrafamilial variation is observed more in families with the shorter expansions, and it was suggested that this reflected an interaction with other genetic and environmental factors.

This seems to be similar to what is observed with polyalanine expansions in the *ARX* gene. The 16 to 23 alanine change in the first polyalanine tract found in two families with ISSX where the phenotype is relatively uniform in these family members. However, a greater number of families with this same mutation would have to be discovered in order to draw any conclusions about the phenotypic variability caused by this mutation. In the case of the 12 to 20 alanine change in the second polyalanine tract, a considerable range of interfamilial phenotypic variability has been observed, ranging from ISSX to NSXLMR. Although this mutation is fully penetrant in the sense that all males that carry the mutation have some degree of MR with or without the other features of ISSX, intrafamilial variation is evident.

The non-penetrance seen in synpolydactyly, may explain the case of MR-N who has only one extra alanine (sixteen to seventeen in the first polyalanine tract), and may support the idea that the intellectually normal brothers carrying the $(GCG)_{10+1}$ allele are normal due to non-penetrance of the phenotype.

5.3.6 Mutational mechanism of sequences encoding polyalanine tracts

In both cases of polyalanine expansions observed in the *ARX* gene, the number of repeats remains stable when transmitted from generation to generation. This is not surprising in the case of the second polyalanine [position of the 428-451dup(24 bp)] as this region is not encoded by a perfect triplet repeat. Thus the mechanism of mutation is unlikely to be due to

replication slippage, as has been hypothesised for the larger expansions of $(CCG)_n$ and $(CAG)_n$ repeats at fragile sites and in polyglutamine expansion disorders respectively (Kunst & Warren, 1994). Rather, unequal crossing over is more likely to be the cause of the initial mutation in each family. This is similar to what has been observed for other disorders with short expansions of polyalanine tracts. OPMD is caused by similar mutations in the *PABP2* gene (Brais *et al.*, 1998). Sequencing of the expanded alleles has shown that they are duplications or insertions within an existing sequences encoding polyalanine tracts that do not consist of perfect triplet repeats (Nakamoto *et al.*, 2002). Additionally, the mutations in some cases, have been tracked through several generations (over approximately 150 years), and unlike the expansions associated with fragile sites and the $(CAG)_n$ expansions in the polyglutamine tracts, the size of the alleles encoding expanded polyalanine tracts in *PABP2* remains stable (van der Sluijs *et al.*, 2003).

The locus of the 24 bp duplication in the *ARX* gene is a mutational hotspot, accounting for more than 50% (17 families) of the mutations found in XLMR patients. It is not surprising that this is the case when one looks at this 24 bp sequence. It is an almost perfect palindrome, which could form a hairpin structure of 8 complementary G and C pairs around a central adenine residue. There are two cases of this as a *de novo* mutation (family MRX-D this thesis and in (Bienvenu *et al.*, 2002)). Similarity can be seen with mutations in the *FOXL2* gene which cause Type II blepharphimosis/epicanthus/ptosis (BPES) (Crisponi *et al.*, 2001), where 30% of mutations causing BPES result from expansion of a polyalanine tract (De Baere *et al.*, 2003). The sequence encoding the polyalanine tract in the *FOXL2* gene is also an almost perfect palindrome, and replication errors caused by formation of hairpin structures during replication has been proposed as the mutational mechanism (De Baere *et al.*, 2001).

5.3.7 Disease mechanisms of ARX polyalanine expansion

The mechanism by which these polyalanine expansions result in disease remains unclear. In the case of OPMD the cells of affected individuals as well as transiently transfected Cos-7 cells with protein containing the expanded polyalanine tract, have been shown to contain intranuclear inclusions/aggregates (INIs) of mutant *PABP2* protein. In all cases of polyalanine expansions, with the exception of *ARX* and *SOX3* mutations, the inheritance is autosomal dominant. *ARX* mutations segregate as X-linked recessive, with carrier females showing random X-inactivation patterns and not displaying any phenotype. This suggests a loss rather than a gain of function, which is suggested for the dominant disorders, and suggests that the mechanism whereby polyalanine tract expansion causes misfunction of the protein may be different. The results presented in this thesis of transiently transfected HeLa cells with mutant protein containing 23 alanine residues within the first polyalanine tract of ARX, suggest that in fact the protein is not forming INIs. These results, however, are preliminary and further data is required before firm conclusions can be drawn.

5.3.8 Animal models of ARX mutations

Kitamura *et al.,* (2002) identified *ARX* as a candidate for XLAG based on the phenotype of mice deficient in *Arx*. These mice had many features in common with human XLAG patients including small brains, shown to be due to aberrant migration and differentiation of interneurons in the ganglionic eminence and neocortex, as well as abnormal testicular differentiation. Although this does not mimic the polyalanine expansions, the most common change found in *ARX* in XLMR patients, it has provided clues to the normal function of ARX and its importance in normal brain development.

As *ARX* is expressed almost exclusively in brain, the effects the mutations have on the normal function of the human protein will be difficult to elucidate. For disorders of the brain, model organisms are a valuable resource for this work. Orthologs of *ARX* are found in a wide range of organisms (including zebrafish) which will help in the future study of the ARX protein, its normal function, the other proteins it interacts with, and the effect the mutations presented in this thesis have on the normal function.

5.4. CONCLUSION

Mutations in *ARX* have been implicated in a significant number of XLMR families and with varying phenotypes associated with MR. The majority of mutations found have been expansions of one of the two polyalanine tracts of the ARX protein. The second polyalanine tract especially, appears to be a hotspot for mutations with nearly 50% (17 of 35 mutations; numbers from Table 5.3) of reported mutations being of the 428-451dup(24 bp) type. Additionally there are two cases where this is a *de novo* change (family MRX-D and Bienvenu *et al.*, 2002). It still remains unclear if small increases in alanine copy number or deletions of alanines are disease susceptibility mutations or are simply rare variants. The effect these variations have on the function of the transcription factor is also unknown. The knockout mouse for *Arx*, resembles the more severe phenotype seen in XLAG patients. It would be of interest to observe the phenotype of the mouse, and the effect on the normal development of the mouse brain that would result from a knock-in for both of the $(GCG)_{10+7}$ and the 428-451dup(24 bp) polyalanine expansion mutations.

Chapter 6

Conclusions

At the outset of this study in March 2000, there were only 8 genes for non-syndromic X-linked mental retardation known. Since then another 7 such genes have been identified. While the number of these genes rose to 15 (including *ARX* described in Chapter 5), mutations have only been found in 19 of the 81 published families where the gene has been mapped by linkage. This low mutation rate highlights the unexpected genetic heterogeneity now characteristic of X-linked mental retardation and also shows that a large number of genes are still to be identified. The aim of this thesis therefore, was to extend knowledge of the genes involved in non-syndromic X-linked mental retardation.

The summary of the main contributions toward this goal, generated as part of this thesis are:

- Identification of mutations in the novel paired-type homeobox gene *ARX* as the cause of ISSX, as well as other syndromic and non-syndromic forms of mental retardation (Partington syndrome, XMESID syndrome, and MRX; Strømme P., Mangelsdorf M. E. *et al.*, (2002) *Nat Genet* **30:** 441-5)
- Discovery and characterisation of the considerable clinical heterogeneity of the *ARX* mutations (Turner G. *et al.*, (2002) *Am J Med Genet* **112:** 405-11; Strømme P. *et al.*, (2002a) *Brain Dev* **24:** 266-8)
- Initiation of a large scale screening for mutations in the *ARX* gene in unmapped familial and sporadic cases with mental retardation, and showing that for familial cases, mutations in *ARX* account for ~8% of those screened
- Identification of the candidate gene, *BGN*, for non-syndromic mental retardation, by characterisation of the inversion breakpoints from a patient with mental retardation and pericentric inversion of X-chromosome [(46, Y, inv(X) (q13.1q28); Mangelsdorf *et al.*, manuscript in preparation - Appendix VI]

- Identification of disomy of Xq28 genes as a likely cause of mental retardation, by characterisation of a inversion breakpoints from a patient whose original karytype of 46, Y, inv(X)(p11.2q28) has been redefined as 46, Y, rec(X)dup(Xq)inv(X)(p11q28) (Mangelsdorf *et al.*, manuscript in preparation - Appendix VI)
- Identification of the gene *TMG3a* as a candidate for X-linked mental retardation, as well as cardiomyopathy, by the characteristion of translocation breakpoints in a patient with a balanced translocation t(X;10) [46, X, t(X;10)(q28;q11.2)]
- Localisation of a novel gene by redefining the minimal linkage to an ~2 Mb interval in a family with non-syndromic mental retardation, and initiating a systematic transcription mapping and candidate gene screen of genes from within this localisation

The Human Genome Project has played a phenomenal role in the facilitation of gene identification in genetic disorders. The rate of discovery of disease causing genes almost approached one gene a day. This is because the gradual release of the human genome sequence over the last 5 years has shifted the emphasis for disease gene identification to *in silico* technology. The availability of such a resource has dramatically reduced the time required for many positional and candidate gene approaches. The ultimate completion of the human sequence, as announced in April 2003, as well as that of the other organisms, will perhaps make the identification of genes for disease more routine, but also high throughput and almost automated.

Identification of genes for mental retardation has been very successful during the past decade. The knowledge about these genes not only helps the families and individuals involved, but also greatly improves our understanding of what is required for normal brain development. The brain is an extremely complex organ, and therefore it is not surprising that a large number

of genes as well as a plethora of environmental factors play a role in its development and functioning. Unfortunately, there is no foreseeable cure for mental retardation at this time and it is unlikely on the basis of present knowledge that drugs can be developed to correct the problems that occur during early development of the brain. However identification of genes causing mental retardation will further our understanding of this complex organ, as well as facilitating more precise diagnosis and thus family counselling which is of considerable and immediate practical value to the affected families.

REFERENCES

Akarsu A. N., Stoilov I., Yilmaz E., Sayli B. S., and Sarfarazi M. (1996). Genomic structure of HOXD13 gene: a nine polyalanine duplication causes synpolydactyly in two unrelated families. *Hum Mol Genet* **5:** 945-52.

Allen K. M., Gleeson J. G., Bagrodia S., Partington M. W., MacMillan J. C., Cerione R. A., Mulley J. C., and Walsh C. A. (1998). PAK3 mutation in nonsyndromic X-linked mental retardation. *Nat Genet* **20:** 25-30.

Amiel J., Laudier B., Attie-Bitach T., Trang H., De Pontual L., Gener B., Trochet D., Etchevers H., Ray P., Simonneau M., Vekemans M., Munnich A., Gaultier C., and Lyonnet S. (2003). Polyalanine expansion and frameshift mutations of the paired-like homeobox gene PHOX2B in congenital central hypoventilation syndrome. *Nat Genet* **33:** 459-61.

Amir R. E., Van den Veyver I. B., Wan M., Tran C. Q., Francke U., and Zoghbi H. Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat Genet* **23:** 185-8.

Anderson J. M. (1996). Cell signalling: MAGUK magic. *Curr Biol* **6:**382-4.

Ashley C. T., Jr., Wilkinson K. D., Reines D., and Warren S. T. (1993). FMR1 protein: conserved RNP family domains and selective RNA binding. *Science* **262:** 563-6.

Bergoffen J., Scherer S. S., Wang S., Scott M. O., Bone L. J., Paul D. L., Chen K., Lensch M. W., Chance P. F., and Fischbeck K. H. (1993). Connexin mutations in X-linked Charcot-Marie-Tooth disease. *Science* **262:** 2039-42.

Beyer K. S., Klauck S. M., Benner A., Poustka F., and Poustka A. (2002). Association studies of the HOPA dodecamer duplication variant in different subtypes of autism. *Am J Med Genet* **114:** 110-5.

Bidanset D. J., LeBaron R., Rosenberg L., Murphy-Ullrich J. E., and Hook M. (1992). Regulation of cell substrate adhesion: effects of small galactosaminoglycan-containing proteoglycans. *J Cell Biol* **118:** 1523-31.

Bienvenu T., des Portes V., Saint Martin A., McDonell N., Billuart P., Carrie A., Vinet M. C., Couvert P., Toniolo D., Ropers H. H., Moraine C., van Bokhoven H., Fryns J. P., Kahn A., Beldjord C., and Chelly J. (1998). Non-specific X-linked semidominant mental retardation by mutations in a Rab GDP-dissociation inhibitor. *Hum Mol Genet* **7:** 1311-5.

Bienvenu T., Poirier K., Friocourt G., Bahi N., Beaumont D., Fauchereau F., Ben Jeema L., Zemni R., Vinet M. C., Francis F., Couvert P., Gomot M., Moraine C., van Bokhoven H., Kalscheuer V., Frints S., Gécz J., Ohzaki K., Chaabouni H., Fryns J. P., Desportes V., Beldjord C., and Chelly J. (2002). ARX, a novel Prd-class-homeobox gene highly expressed in the telencephalon, is mutated in X-linked mental retardation. *Hum Mol Genet* **11:** 981-91.

Billuart P., Bienvenu T., Ronce N., des Portes V., Vinet M. C., Zemni R., Roest Crollius H., Carrie A., Fauchereau F., Cherry M., Briault S., Hamel B., Fryns J. P., Beldjord C., Kahn A., Moraine C., and Chelly J. (1998). Oligophrenin-1 encodes a rhoGAP protein involved in X-linked mental retardation. *Nature* **392:** 923-6.

Birch HG R. S., Baird D, Horobin G, Ilsley R. (1970). "Mental subnormality in the community: a clinical and epidemiological study.,", Baltimore.

Birnboim H. C., and Doly J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* **7:** 1513-23.

Bonneau D., Toutain A., Laquerriere A., Marret S., Saugier-Veber P., Barthez M. A., Radi S., Biran-Mucignat V., Rodriguez D., and Gelot A. (2002). X-linked lissencephaly with absent corpus callosum and ambiguous genitalia (XLAG): clinical, magnetic resonance imaging, and neuropathological findings. *Ann Neurol* **51:** 340-9.

Brais B., Bouchard J. P., Xie Y. G., Rochefort D. L., Chretien N., Tome F. M., Lafreniere R. G., Rommens J. M., Uyama E., Nohira O., Blumen S., Korczyn A. D., Heutink P., Mathieu J., Duranceau A., Codere F., Fardeau M., Rouleau G. A., and Korcyn A. D.

(1998). Short GCG expansions in the PABP2 gene cause oculopharyngeal muscular dystrophy. *Nat Genet* **18:** 164-7.

Brancaccio M., Guazzone S., Menini N., Sibona E., Hirsch E., De Andrea M., Rocchi M., Altruda F., Tarone G., and Silengo L. (1999). Melusin is a new muscle-specific interactor for beta(1) integrin cytoplasmic domain. *J Biol Chem* **274:** 29282-8.

Brouwer A., ten Berge D., Wiegerinck R., and Meijlink F. (2003). The OAR/aristaless domain of the homeodomain protein Cart1 has an attenuating role in vivo. *Mech Dev* **120:** 241-52.

Brown S. A., Warburton D., Brown L. Y., Yu C. Y., Roeder E. R., Stengel-Rutkowski S., Hennekam R. C., and Muenke M. (1998). Holoprosencephaly due to mutations in ZIC2, a homologue of Drosophila odd-paired. *Nat Genet* **20:** 180-3.

Brown W. T. (1996). The FRAXE Syndrome: is it time for routine screening? *Am J Hum Genet* **58:** 903.

Bruyere H., Lewis M. E., Wood S., MacLeod P., and Langlois S. (1999a). Increasing evidence for a new X-linked mental retardation/epilepsy gene localized to Xp21.3-Xp22.1. *Am J Med Genet* **86:** 401.

Bruyere H., Lewis S., Wood S., MacLeod P. J., and Langlois S. (1999b). Confirmation of linkage in X-linked infantile spasms (West syndrome) and refinement of the disease locus to Xp21.3-Xp22.1. *Clin Genet* **55:** 173-81.

Calado A., Tome F. M., Brais B., Rouleau G. A., Kuhn U., Wahle E., and Carmo-Fonseca M. (2000). Nuclear inclusions in oculopharyngeal muscular dystrophy consist of poly(A) binding protein 2 aggregates which sequester poly(A) RNA. *Hum Mol Genet* **9:** 2321-8.

Callen D.F., Baker E., Eyre H.J., Chernos J.E., Bell J.A. and Sutherland G.R. (1990). Reassessment of two apparent deletions of chromosome 16p to an ins(11;16) and a t(1;16) by chromosome painting. *Annals in Genetics* **33:** 219-21.

Carrie A., Jun L., Bienvenu T., Vinet M. C., McDonell N., Couvert P., Zemni R., Cardona A., Van Buggenhout G., Frints S., Hamel B., Moraine C., Ropers H. H., Strom T., Howell G. R., Whittaker A., Ross M. T., Kahn A., Fryns J. P., Beldjord C., Marynen P., and Chelly J. (1999). A new member of the IL-1 receptor family highly expressed in hippocampus and involved in X-linked mental retardation. *Nat Genet* **23:** 25-31.

Celeste A. J., Rosen V., Buecker J. L., Kriz R., Wang E. A., and Wozney J. M. (1986). Isolation of the human gene for bone gla protein utilizing mouse and rat cDNA clones. *Embo J* **5:** 1885-90.

Chiurazzi P., Hamel B. C., and Neri G. (2001). XLMR genes: update 2000. *Eur J Hum Genet* **9:** 71-81.

Chung C. T., Niemela S. L., and Miller R. H. (1989). One-step preparation of competent Escherichia coli: transformation and storage of bacterial cells in the same solution. *Proc Natl Acad Sci U S A* **86:** 2172-5.

Claes S., Devriendt K., Lagae L., Ceulemans B., Dom L., Casaer P., Raeymaekers P., Cassiman J. J., and Fryns J. P. (1997). The X-linked infantile spasms syndrome (MIM 308350) maps to Xp11.4-Xpter in two pedigrees. *Ann Neurol* **42:** 360-4.

Collins F. S., Patrinos A., Jordan E., Chakravarti A., Gesteland R., and Walters L. (1998). New goals for the U.S. Human Genome Project: 1998-2003. *Science* **282:** 682-9.

Communi D., Pirotton S., Parmentier M., and Boeynaems J. M. (1995). Cloning and functional expression of a human uridine nucleotide receptor. *J Biol Chem* **270:** 30849-52.

Cox J. J., Holden S. T., Dee S., Burbridge J. I., and Raymond F. L. (2003). Identification of a 650 kb duplication at the X chromosome breakpoint in a patient with 46,X,t(X;8)(q28;q12) and non-syndromic mental retardation. *J Med Genet* **40:** 169-74.

Crisponi L., Deiana M., Loi A., Chiappe F., Uda M., Amati P., Bisceglia L., Zelante L., Nagaraja R., Porcu S., Ristaldi M. S., Marzella R., Rocchi M., Nicolino M., Lienhardt-Roussie A., Nivelon A., Verloes A., Schlessinger D., Gasparini P., Bonneau D., Cao A., and Pilia G. (2001). The putative forkhead transcription factor FOXL2 is mutated in blepharophimosis/ptosis/epicanthus inversus syndrome. *Nat Genet* **27:** 159-66.

D'Adamo P., Menegon A., Lo Nigro C., Grasso M., Gulisano M., Tamanini F., Bienvenu T., Gedeon A. K., Oostra B., Wu S. K., Tandon A., Valtorta F., Balch W. E., Chelly J., and Toniolo D. (1998). Mutations in GDI1 are responsible for X-linked non-specific mental retardation. *Nat Genet* **19:** 134-9.

Daniels R. H., Zenke F. T., and Bokoch G. M. (1999). alphaPix stimulates p21-activated kinase activity through exchange factor-dependent and -independent mechanisms. *J Biol Chem* **274:** 6047-50.

De Baere E., Beysen D., Oley C., Lorenz B., Cocquet J., De Sutter P., Devriendt K., Dixon M., Fellous M., Fryns J. P., Garza A., Jonsrud C., Koivisto P. A., Krause A., Leroy B. P., Meire F., Plomp A., Van Maldergem L., De Paepe A., Veitia R., and Messiaen L. (2003). FOXL2 and BPES: Mutational Hotspots, Phenotypic Variability, and Revision of the Genotype-Phenotype Correlation. *Am J Hum Genet* **72:** 478-87.

De Baere E., Dixon M. J., Small K. W., Jabs E. W., Leroy B. P., Devriendt K., Gillerot Y., Mortier G., Meire F., Van Maldergem L., Courtens W., Hjalgrim H., Huang S., Liebaers I., Van Regemorter N., Touraine P., Praphanphoj V., Verloes A., Udar N., Yellore V., Chalukya M., Yelchits S., De Paepe A., Kuttenn F., Fellous M., Veitia R., and Messiaen L. (2001). Spectrum of FOXL2 gene mutations in blepharophimosis-ptosis-epicanthus inversus (BPES) families demonstrates a genotype--phenotype correlation. *Hum Mol Genet* **10:** 1591-600.

de Kok Y. J., Vossenaar E. R., Cremers C. W., Dahl N., Laporte J., Hu L. J., Lacombe D., Fischel-Ghodsian N., Friedman R. A., Parnes L. S., Thorpe P., Bitner-Glindzicz M., Pander H. J., Heilbronner H., Graveline J., den Dunnen J. T., Brunner H. G., Ropers H. H., and Cremers F. P. (1996). Identification of a hot spot for microdeletions in

patients with X-linked deafness type 3 (DFN3) 900 kb proximal to the DFN3 gene POU3F4. *Hum Mol Genet* **5:** 1229-35.

Dib C., Faure S., Fizames C., Samson D., Drouot N., Vignal A., Millasseau P., Marc S., Hazan J., Seboun E., Lathrop M., Gyapay G., Morissette J., and Weissenbach J. (1996). A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* **380:** 152-4.

Donnelly A. J. (1997). The Characterisation of human X-linked polymorphic markers and their use in disease gene localisation and identification. PhD Thesis, Department of Genetics, University of Adelaide, Australia.

Dryja T. P., McGee T. L., Hahn L. B., Cowley G. S., Olsson J. E., Reichel E., Sandberg M. A., and Berson E. L. (1990). Mutations within the rhodopsin gene in patients with autosomal dominant retinitis pigmentosa. *N Engl J Med* **323:** 1302-7.

Fantes J., Redeker B., Breen M., Boyle S., Brown J., Fletcher J., Jones S., Bickmore W., Fukushima Y., Mannens M., and et al. (1995). Aniridia-associated cytogenetic rearrangements suggest that a position effect may cause the mutant phenotype. *Hum Mol Genet* **4:** 415-22.

Feinberg A. P., and Leahy W. R. (1977). Infantile spasms: case report of sex-linked inheritance. *Dev Med Child Neurol* **19:** 524-6.

Feinberg A. P., and Vogelstein B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem* **132:** 6-13.

Flint J., Wilkie A. O., Buckle V. J., Winter R. M., Holland A. J., and McDermid H. E. (1995). The detection of subtelomeric chromosomal rearrangements in idiopathic mental retardation. *Nat Genet* **9:** 132-40.

Flomen R. H., Vatcheva R., Gorman P. A., Baptista P. R., Groet J., Barisic I., Ligutic I., and Nizetic D. (1998). Construction and analysis of a sequence-ready map in 4q25: Rieger

syndrome can be caused by haploinsufficiency of RIEG, but also by chromosome breaks approximately 90 kb upstream of this gene. *Genomics* **47:** 409-13.

Frints S. G., Borghgraef M., Froyen G., Marynen P., and Fryns J. P. (2002a). Clinical study and haplotype analysis in two brothers with Partington syndrome. *Am J Med Genet* **112:** 361-8.

Frints S. G., Froyen G., Marynen P., Willekens D., Legius E., and Fryns J. P. (2002b). Re-evaluation of MRX36 family after discovery of an ARX gene mutation reveals mild neurological features of Partington syndrome. *Am J Med Genet* **112:** 427-8.

Fukami M., Kirsch S., Schiller S., Richter A., Benes V., Franco B., Muroya K., Rao E., Merker S., Niesler B., Ballabio A., Ansorge W., Ogata T., and Rappold G. A. (2000). A member of a gene family on Xp22.3, VCX-A, is deleted in patients with X-linked nonspecific mental retardation. *Am J Hum Genet* **67:** 563-73.

Furie B., Bouchard B. A., and Furie B. C. (1999). Vitamin K-dependent biosynthesis of gamma-carboxyglutamic acid. *Blood* **93:** 1798-808.

Galliot B., and Miller D. (2000). Origin of anterior patterning. How old is our head? *Trends Genet* **16:** 1-5.

Gécz J., Barnett S., Liu J., Hollway G., Donnelly A., Eyre H., Eshkevari H. S., Baltazar R., Grunn A., Nagaraja R., Gilliam C., Peltonen L., Sutherland G. R., Baron M., and Mulley J. C. (1999). Characterization of the human glutamate receptor subunit 3 gene (GRIA3), a candidate for bipolar disorder and nonspecific X-linked mental retardation. *Genomics* **62:** 356-68.

Gécz J., Gedeon A. K., Sutherland G. R., and Mulley J. C. (1996). Identification of the gene FMR2, associated with FRAXE mental retardation. *Nat Genet* **13:** 105-8.

Gécz J., Oostra B. A., Hockey A., Carbonell P., Turner G., Haan E. A., Sutherland G. R., and Mulley J. C. (1997). FMR2 expression in families with FRAXE mental retardation. *Hum Mol Genet* **6:** 435-41.

Gécz J., Baker E., Donnelly A., Ming J. E., McDonald-McGinn D. M., Spinner N. B., Zackai E. H., Sutherland G. R., and Mulley J. C. (1999). Fibroblast growth factor homologous factor 2 (FHF2): gene structure, expression and mapping to the Borjeson-Forssman-Lehmann syndrome region in Xq26 delineated by a duplication breakpoint in a BFLS-like patient. *Hum Genet* **104:** 56-63.

Gécz J., and Mulley J. (2000). Genes for cognitive function: developments on the X. *Genome Res* **10:** 157-63.

Gedeon A. K., Donnelly A. J., Mulley J. C., Kerr B., and Turner G. (1996). How many X-linked genes for non-specific mental retardation (MRX) are there? *Am J Med Genet* **64:** 158-62.

Gedeon A., Kerr B., Mulley J., and Turner G. (1994). Pericentromeric genes for non-specific X-linked mental retardation (MRX). *Am J Med Genet* **51:** 553-64.

Gibbons R. J., McDowell T. L., Raman S., O'Rourke D. M., Garrick D., Ayyub H., and Higgs D. R. (2000). Mutations in ATRX, encoding a SWI/SNF-like protein, cause diverse changes in the pattern of DNA methylation. *Nat Genet* **24:** 368-71.

Gibbons R. J., Picketts D. J., Villard L., and Higgs D. R. (1995). Mutations in a putative global transcriptional regulator cause X-linked mental retardation with alpha-thalassemia (ATR-X syndrome). *Cell* **80:** 837-45.

Glass I. A. (1991). X linked mental retardation. *J Med Genet* **28:** 361-71.

Goldstein L. S. (2001). Kinesin molecular motors: transport pathways, receptors, and human disease. *Proc Natl Acad Sci U S A* **98:** 6999-7003.

Gomot M., Ronce N., Dessay S., Zemni R., Ayrault A. D., Moizard M. P., Nivelon A., Gilgenkrantz S., Dourlens J., Des Portes V., Chelly J., and Moraine C. (2002).

TM4SF2 gene involvement reconsidered in an XLMR family after neuropsychological assessment. *Am J Med Genet* **112:** 400-4.

Goodman F. R., Bacchelli C., Brady A. F., Brueton L. A., Fryns J. P., Mortlock D. P., Innis J. W., Holmes L. B., Donnenfeld A. E., Feingold M., Beemer F. A., Hennekam R. C., and Scambler P. J. (2000). Novel HOXA13 mutations and the phenotypic spectrum of hand-foot-genital syndrome. *Am J Hum Genet* **67:** 197-202.

Goodman F. R., Mundlos S., Muragaki Y., Donnai D., Giovannucci-Uzielli M. L., Lapi E., Majewski F., McGaughran J., McKeown C., Reardon W., Upton J., Winter R. M., Olsen B. R., and Scambler P. J. (1997). Synpolydactyly phenotypes correlate with size of expansions in HOXD13 polyalanine tract. *Proc Natl Acad Sci U S A* **94:** 7458-63.

Greig G. M., Sharp C. B., Carrel L., and Willard H. F. (1993). Duplicated zinc finger protein genes on the proximal short arm of the human X chromosome: isolation, characterization and X-inactivation studies. *Hum Mol Genet* **2:** 1611-8.

Grzybowska E. A., Wilczynska A., and Siedlecki J. A. (2001). Regulatory functions of 3'UTRs. *Biochem Biophys Res Commun* **288:** 291-5.

Gu Y., Shen Y., Gibbs R. A., and Nelson D. L. (1996). Identification of FMR2, a novel gene associated with the FRAXE CCG repeat and CpG island. *Nat Genet* **13:** 109-13.

Guy J., Spalluto C., McMurray A., Hearn T., Crosier M., Viggiano L., Miolla V., Archidiacono N., Rocchi M., Scott C., Lee P. A., Sulston J., Rogers J., Bentley D., and Jackson M. S. (2000). Genomic sequence and transcriptional profile of the boundary between pericentromeric satellites and genes on human chromosome arm 10q. *Hum Mol Genet* **9:** 2029-42.

Haan E. A., Mulley J. C., Gedeon A. K., Sheffield L. J., and Sutherland G. R. (1988). Presymptomatic testing for myotonic dystrophy by means of the linked DNA marker APOC2. *Med J Aust* **149:** 326-9.

Haft C. R., de la Luz Sierra M., Barr V. A., Haft D. H., and Taylor S. I. (1998). Identification of a family of sorting nexin molecules and characterization of their association with receptors. *Mol Cell Biol* **18:** 7278-87.

Hagberg B., Aicardi J., Dias K., and Ramos O. (1983). A progressive syndrome of autism, dementia, ataxia, and loss of purposeful hand use in girls: Rett's syndrome: report of 35 cases. *Ann Neurol* **14:** 471-9.

Hahn K. A., Salomons G. S., Tackels-Horne D., Wood T. C., Taylor H. A., Schroer R. J., Lubs H. A., Jakobs C., Olson R. L., Holden K. R., Stevenson R. E., and Schwartz C. E. (2002). X-linked mental retardation with seizures and carrier manifestations is caused by a mutation in the creatine-transporter gene (SLC6A8) located in Xq28. *Am J Hum Genet* **70:** 1349-56.

Hall A. (1998). Rho GTPases and the actin cytoskeleton. *Science* **279:** 509-14.

Hamel B. C., Chiurazzi P., and Lubs H. A. (2000). Syndromic XLMR genes (MRXS): update 2000. *Am J Med Genet* **94:** 361-3.

Han K., and Manley J. L. (1993). Functional domains of the Drosophila Engrailed protein. *Embo J* **12:** 2723-33.

Hane B., Schroer R. J., Arena J. F., Lubs H. A., Schwartz C. E., and Stevenson R. E. (1996). Nonsyndromic X-linked mental retardation: review and mapping of MRX29 to Xp21. *Clin Genet* **50:** 176-83.

Hane B., Stevenson R. E., Arena J. F., Lubs H. A., Simensen R. J., and Schwartz C. E. (1999). Gene for apparently nonsyndromic X-linked mental retardation (MRX32) maps to an 18-Mb region of Xp21.2-p22. *Am J Med Genet* **85:** 271-5.

He Z., Ma W. Y., Liu G., Zhang Y., Bode A. M., and Dong Z. (2003). Arsenite-induced phosphorylation of histone H3 at serine 10 is mediated by Akt1, ERK2, and RSK2 but not MSK1. *J Biol Chem* **14:** 14.

Hildebrand A., Romaris M., Rasmussen L. M., Heinegard D., Twardzik D. R., Border W. A., and Ruoslahti E. (1994). Interaction of the small interstitial proteoglycans biglycan, decorin and fibromodulin with transforming growth factor beta. *Biochem J* **302:** 527-34.

Holinski-Feder E., Golla A., Rost I., Seidel H., Rittinger O., and Meindl A. (1996). Regional localization of two MRX genes to Xq28 (MRX28) and to Xp11.4-Xp22.12 (MRX33). *Am J Med Genet* **64:** 125-30.

Huston J. P., Weth K., De Souza Silva A., Junghans U., Muller H. W., and Hasenohrl R. U. (2000). Facilitation of learning and long-term ventral pallidal-cortical cholinergic activation by proteoglycan biglycan and chondroitin sulfate C. *Neuroscience* **100:** 355-61.

Iozzo R. V. (1998). Matrix proteoglycans: from molecular design to cellular function. *Annu Rev Biochem* **67:** 609-52.

Jalink K., van Corven E. J., Hengeveld T., Morii N., Narumiya S., and Moolenaar W. H. (1994). Inhibition of lysophosphatidate- and thrombin-induced neurite retraction and neuronal cell rounding by ADP ribosylation of the small GTP-binding protein Rho. *J Cell Biol* **126:** 801-10.

Jamain S., Quach H., Betancur C., Rastam M., Colineaux C., Gillberg I. C., Soderstrom H., Giros B., Leboyer M., Gillberg C., Bourgeron T., Nyden A., Philippe A., Cohen D., Chabane N., Mouren-Simeoni M. C., Brice A., Sponheim E., Spurkland I., Skjeldal O. H., Coleman M., Pearl P. L., Cohen I. L., Tsiouris J., Zappella M., Menchetti G., Pompella A., Aschauer H., and Van Maldergem L. (2003). Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nat Genet* **31:** 31.

Jordanova A., Kalaydjieva L., Savov A., Claustres M., Schwarz M., Estivill X., Angelicheva D., Haworth A., Casals T., and Kremensky I. (1997). SSCP analysis: a blind sensitivity trial. *Hum Mutat* **10:** 65-70.

Jun L., Frints S., Duhamel H., Herold A., Abad-Rodrigues J., Dotti C., Izaurralde E., Marynen P., and Froyen G. (2001). NXF5, a novel member of the nuclear RNA export factor family, is lost in a male patient with a syndromic form of mental retardation. *Curr Biol* **11:** 1381-91.

Junghans U., Koops A., Westmeyer A., Kappler J., Meyer H. E., and Muller H. W. (1995). Purification of a meningeal cell-derived chondroitin sulphate proteoglycan with neurotrophic activity for brain neurons and its identification as biglycan. *Eur J Neurosci* **7:** 2341-50.

Kalscheuer V., Tao J., Donnelly A., Hollway G., Schwinger E., Kubart S., Menzel C., Hoeltzenbein M., Tommerup N., Eyre H., Harbord M., Haan E., Sutherland G. R., Ropers H. H., and Gécz J. (2003). Disruption of the Serine/Threonine Kinase 9 gene causes severe X-linked infantile spasms and mental retardation. *Am J Hum Genet* (May 7 published online).

Kere J., Srivastava A. K., Montonen O., Zonana J., Thomas N., Ferguson B., Munoz F., Morgan D., Clarke A., Baybayan P., Chen E. Y., Ezer S., Saarialho-Kere U., de la Chapelle A., and Schlessinger D. (1996). X-linked anhidrotic (hypohidrotic) ectodermal dysplasia is caused by mutation in a novel transmembrane protein. *Nat Genet* **13:** 409-16.

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

Kinsella M. G., Tsoi C. K., Jarvelainen H. T., and Wight T. N. (1997). Selective expression and processing of biglycan during migration of bovine aortic endothelial cells. The role of endogenous basic fibroblast growth factor. *J Biol Chem* **272:** 318-25.

Kitamura K., Yanazawa M., Sugiyama N., Miura H., Iizuka-Kogo A., Kusaka M., Omichi K., Suzuki R., Kato-Fukui Y., Kamiirisa K., Matsuo M., Kamijo S., Kasahara M., Yoshioka H., Ogata T., Fukuda T., Kondo I., Kato M., Dobyns W. B., Yokoyama M., and Morohashi K. (2002). Mutation of ARX causes abnormal development of

forebrain and testes in mice and X-linked lissencephaly with abnormal genitalia in humans. *Nat Genet* **32:** 359-69.

Knight S. J., and Flint J. (2000). Perfect endings: a review of subtelomeric probes and their use in clinical diagnosis. *J Med Genet* **37:** 401-9.

Kozinetz C. A., Skender M. L., MacNaughton N., Almes M. J., Schultz R. J., Percy A. K., and Glaze D. G. (1993). Epidemiology of Rett syndrome: a population-based registry. *Pediatrics* **91:** 445-50.

Kulman J. D., Harris J. E., Haldeman B. A., and Davie E. W. (1997). Primary structure and tissue distribution of two novel proline-rich gamma-carboxyglutamic acid proteins. *Proc Natl Acad Sci U S A* **94:** 9058-62.

Kulman J. D., Harris J. E., Xie L., and Davie E. W. (2001). Identification of two novel transmembrane gamma-carboxyglutamic acid proteins expressed broadly in fetal and adult tissues. *Proc Natl Acad Sci U S A* **98:** 1370-5.

Kunst C. B., and Warren S. T. (1994). Cryptic and polar variation of the fragile X repeat could result in predisposing normal alleles. *Cell* **77:** 853-61.

Kutsche K., Yntema H., Brandt A., Jantke I., Nothwang H. G., Orth U., Boavida M. G., David D., Chelly J., Fryns J. P., Moraine C., Ropers H. H., Hamel B. C., van Bokhoven H., and Gal A. (2000). Mutations in ARHGEF6, encoding a guanine nucleotide exchange factor for Rho GTPases, in patients with X-linked mental retardation. *Nat Genet* **26:** 247-50.

Lahn B.T., Ma N., Breg W.R., Stratton R., Surti U. and Page D.C. (1994) Xq-Yq interchange resulting in supernormal X-linked gene expression in severely retarded males with 46,XYq-karyotype. *Nat Genet.* **8:** 243-50.

Lander E. S., Linton L. M., Birren B., Nusbaum C., Zody M. C., Baldwin J., Devon K., Dewar K., Doyle M., FitzHugh W., Funke R., Gage D., Harris K., Heaford A., Howland J., Kann L., Lehoczky J., LeVine R., McEwan P., McKernan K., Meldrim J.,

Mesirov J. P., Miranda C., Morris W., Naylor J., Raymond C., Rosetti M., Santos R., Sheridan A., Sougnez C., Stange-Thomann N., Stojanovic N., Subramanian A., Wyman D., Rogers J., Sulston J., Ainscough R., Beck S., Bentley D., Burton J., Clee C., Carter N., Coulson A., Deadman R., Deloukas P., Dunham A., Dunham I., Durbin R., French L., Grafham D., Gregory S., Hubbard T., Humphray S., Hunt A., Jones M., Lloyd C., McMurray A., Matthews L., Mercer S., Milne S., Mullikin J. C., Mungall A., Plumb R., Ross M., Shownkeen R., Sims S., Waterston R. H., Wilson R. K., Hillier L. W., McPherson J. D., Marra M. A., Mardis E. R., Fulton L. A., Chinwalla A. T., Pepin K. H., Gish W. R., Chissoe S. L., Wendl M. C., Delehaunty K. D., Miner T. L., Delehaunty A., Kramer J. B., Cook L. L., Fulton R. S., Johnson D. L., Minx P. J., Clifton S. W., Hawkins T., Branscomb E., Predki P., Richardson P., Wenning S., Slezak T., Doggett N., Cheng J. F., Olsen A., Lucas S., Elkin C., Uberbacher E., Frazier M., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* **409:** 860-921.

Lathrop G. M. and Lalouel J. M. (1984). Easy calculations of lod scores and genetic risks on small computers. *Am J. Hum. Genet* **36:** 460-5.

Lathrop G. M., Lalouel J. M., Julien C. and Ott J. (1984). Strategies for multilocus linkage analysis in humans. *Proc Natl Acad Sci* **81:** 3443-6.

Laumonnier F., Ronce N., Hamel B. C., Thomas P., Lespinasse J., Raynaud M., Paringaux C., Van Bokhoven H., Kalscheuer V., Fryns J. P., Chelly J., Moraine C., and Briault S. (2002). Transcription factor SOX3 is involved in X-linked mental retardation with growth hormone deficiency. *Am J Hum Genet* **71:** 1450-5.

Lebel R. R., May M., Pouls S., Lubs H. A., Stevenson R. E., and Schwartz C. E. (2002). Non-syndromic X-linked mental retardation associated with a missense mutation (P312L) in the FGD1 gene. *Clin Genet* **61:** 139-45.

Lennon G., Auffray C., Polymeropoulos M., and Soares M. B. (1996). The I.M.A.G.E. Consortium: an integrated molecular analysis of genomes and their expression. *Genomics* **33:** 151-2.

Lettice L. A., Horikoshi T., Heaney S. J., van Baren M. J., van der Linde H. C., Breedveld G. J., Joosse M., Akarsu N., Oostra B. A., Endo N., Shibata M., Suzuki M., Takahashi E., Shinka T., Nakahori Y., Ayusawa D., Nakabayashi K., Scherer S. W., Heutink P., Hill R. E., and Noji S. (2002). Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. *Proc Natl Acad Sci U S A* **99:** 7548-53.

Li Q., Harju S., and Peterson K. R. (1999). Locus control regions: coming of age at a decade plus. *Trends Genet* **15:** 403-8.

Lossi A. M., Laugier-Anfossi F., Depetris D., Gécz J., Gedeon A., Kooy F., Schwartz C., Mattei M. G., Croquette M. F., and Villard L. (2002). Abnormal expression of the KLF8 (ZNF741) gene in a female patient with an X;autosome translocation t(X;21)(p11.2;q22.3) and non-syndromic mental retardation. *J Med Genet* **39:** 113-7.

Lower K. M., Turner G., Kerr B. A., Mathews K. D., Shaw M. A., Gedeon A. K., Schelley S., Hoyme H. E., White S. M., Delatycki M. B., Lampe A. K., Clayton-Smith J., Stewart H., van Ravenswaay C. M., de Vries B. B., Cox B., Grompe M., Ross S., Thomas P., Mulley J. C., and Gécz J. (2002). Mutations in PHF6 are associated with Borjeson-Forssman-Lehmann syndrome. *Nat Genet* **32:** 661-5.

Lubs H. A. (1969). A marker X chromosome. *Am J Hum Genet* **21:** 231-44.

Makino K., Kuwahara H., Masuko N., Nishiyama Y., Morisaki T., Sasaki J., Nakao M., Kuwano A., Nakata M., Ushio Y., and Saya H. (1997). Cloning and characterization of NE-dlg: a novel human homolog of the Drosophila discs large (dlg) tumor suppressor protein interacts with the APC protein. *Oncogene* **14:** 2425-33.

Manfioletti G., Brancolini C., Avanzi G., and Schneider C. (1993). The protein encoded by a growth arrest-specific gene (gas6) is a new member of the vitamin K-dependent proteins related to protein S, a negative coregulator in the blood coagulation cascade. *Mol Cell Biol* **13:** 4976-85.

Mann K. G., Nesheim M. E., Church W. R., Haley P., and Krishnaswamy S. (1990). Surface-dependent reactions of the vitamin K-dependent enzyme complexes. *Blood* **76:** 1-16.

Manser E., Chong C., Zhao Z. S., Leung T., Michael G., Hall C., and Lim L. (1995). Molecular cloning of a new member of the p21-Cdc42/Rac-activated kinase (PAK) family. *J Biol Chem* **270:** 25070-8.

Martin J. P. a. B., J. (1943). A pedigree of mental deficit showing sex-linkage. *J. Neurol. psychiatry* **6:** 154.

McIntosh J. M., Olivera B. M., Cruz L. J., and Gray W. R. (1984). Gamma-carboxyglutamate in a neuroactive toxin. *J Biol Chem* **259:** 14343-6.

Meloni I., Muscettola M., Raynaud M., Longo I., Bruttini M., Moizard M. P., Gomot M., Chelly J., des Portes V., Fryns J. P., Ropers H. H., Magi B., Bellan C., Volpi N., Yntema H. G., Lewis S. E., Schaffer J. E., and Renieri A. (2002). FACL4, encoding fatty acid-CoA ligase 4, is mutated in nonspecific X-linked mental retardation. *Nat Genet* **30:** 436-40.

Merienne K., Jacquot S., Pannetier S., Zeniou M., Bankier A., Gécz J., Mandel J. L., Mulley J., Sassone-Corsi P., and Hanauer A. (1999). A missense mutation in RPS6KA3 (RSK2) responsible for non-specific mental retardation. *Nat Genet* **22:** 13-4.

Merienne K., Pannetier S., Harel-Bellan A., and Sassone-Corsi P. (2001). Mitogen-regulated RSK2-CBP interaction controls their kinase and acetylase activities. *Mol Cell Biol* **21:** 7089-96.

Michaelis R. C., Copeland-Yates S. A., Sossey-Alaoui K., Skinner C., Friez M. J., Longshore J. W., Simensen R. J., Schroer R. J., and Stevenson R. E. (2000). The HOPA gene dodecamer duplication is not a significant etiological factor in autism. *J Autism Dev Disord* **30:** 355-8.

Missler M., Fernandez-Chacon R., and Sudhof T. C. (1998). The making of neurexins. *J Neurochem* **71:** 1339-47.

Miura H., Yanazawa M., Kato K., and Kitamura K. (1997). Expression of a novel aristaless related homeobox gene 'Arx' in the vertebrate telencephalon, diencephalon and floor plate. *Mech Dev* **65:** 99-109.

Molinari F., Rio M., Meskenaite V., Encha-Razavi F., Auge J., Bacq D., Briault S., Vekemans M., Munnich A., Attie-Bitach T., Sonderegger P., and Colleaux L. (2002). Truncating neurotrypsin mutation in autosomal recessive nonsyndromic mental retardation. *Science* **298:** 1779-81.

Monaco A. P. N., R L; Colletti-Feener, C; Bertelson, C J; Kurnit, D M; Kunkel, L M (1986). Isolation of candidate cDNAs for portions of the Duchenne muscular dystrophy gene. *Nature* **323:** 646-650.

Moser HW R. C. a. L. C. (1983). Principles and practice of medical genetics (R. D. Emery AEH, Ed.), Churchill Livingstone., Edinburgh.

Muller B. M., Kistner U., Kindler S., Chung W. J., Kuhlendahl S., Fenster S. D., Lau L. F., Veh R. W., Huganir R. L., Gundelfinger E. D., and Garner C. C. (1996). SAP102, a novel postsynaptic protein that interacts with NMDA receptor complexes in vivo. *Neuron* **17:** 255-65.

Mulley J. C., Kerr B., Stevenson R., and Lubs H. (1992). Nomenclature guidelines for X-linked mental retardation. *Am J Med Genet* **43:** 383-91.

Mundlos S., Otto F., Mundlos C., Mulliken J. B., Aylsworth A. S., Albright S., Lindhout D., Cole W. G., Henn W., Knoll J. H., Owen M. J., Mertelsmann R., Zabel B. U., and Olsen B. R. (1997). Mutations involving the transcription factor CBFA1 cause cleidocranial dysplasia. *Cell* **89:** 773-9.

Murakami A., Yajima T., Sakuma H., McLaren M. J., and Inana G. (1993). X-arrestin: a new retinal arrestin mapping to the X chromosome. *FEBS Lett* **334:** 203-9.

Nagata K., Ohashi K., Nakano T., Arita H., Zong C., Hanafusa H., and Mizuno K. (1996). Identification of the product of growth arrest-specific gene 6 as a common ligand for Axl, Sky, and Mer receptor tyrosine kinases. *J Biol Chem* **271:** 30022-7.

Nakamoto M., Nakano S., Kawashima S., Ihara M., Nishimura Y., Shinde A., and Kakizuka A. (2002). Unequal crossing-over in unique PABP2 mutations in Japanese patients: a possible cause of oculopharyngeal muscular dystrophy. *Arch Neurol* **59:** 474-7.

Noguchi M., Yi H., Rosenblatt H. M., Filipovich A. H., Adelstein S., Modi W. S., McBride O. W., and Leonard W. J. (1993). Interleukin-2 receptor gamma chain mutation results in X-linked severe combined immunodeficiency in humans. *Cell* **73:** 147-57.

Nollau P., and Wagener C. (1997). Methods for detection of point mutations: performance and quality assessment. IFCC Scientific Division, Committee on Molecular Biology Techniques. *Clin Chem* **43:** 1114-28.

Oberle I., Rousseau F., Heitz D., Kretz C., Devys D., Hanauer A., Boue J., Bertheas M. F., and Mandel J. L. (1991). Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. *Science* **252:** 1097-102.

Oh S., Hahn H., Torrey T. A., Shin H., Choi W., Lee Y. M., Morse H. C., and Kim W. (2000). Identification of the human homologue of mouse KIF4, a kinesin superfamily motor protein. *Biochim Biophys Acta* **1493:** 219-24.

Onda M., Inui S., Maeda K., Suzuki M., Takahashi E., and Sakaguchi N. (1997). Expression and chromosomal localization of the human alpha 4/IGBP1 gene, the structure of which is closely related to the yeast TAP42 protein of the rapamycin-sensitive signal transduction pathway. *Genomics* **46:** 373-8.

Orita M., Iwahana H., Kanazawa H., Hayashi K., and Sekiya T. (1989). Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc Natl Acad Sci U S A* **86:** 2766-70.

Orrico A., Lam C., Galli L., Dotti M. T., Hayek G., Tong S. F., Poon P. M., Zappella M., Federico A., and Sorrentino V. (2000). MECP2 mutation in male patients with non-specific X-linked mental retardation. *FEBS Lett* **481:** 285-8.

Ott J. (1991). "Analysis of Human Genetic Linkage (rev ed)," John Hopkins University Press, Baltimore.

Palacin M., Estevez R., Bertran J., and Zorzano A. (1998). Molecular biology of mammalian plasma membrane amino acid transporters. *Physiol Rev* **78:** 969-1054.

Partington M. W., Mulley J. C., Sutherland G. R., Hockey A., Thode A., and Turner G. (1988). X-linked mental retardation with dystonic movements of the hands. *Am J Med Genet* **30:** 251-62.

Pasteris N. G., Cadle A., Logie L. J., Porteous M. E., Schwartz C. E., Stevenson R. E., Glover T. W., Wilroy R. S., and Gorski J. L. (1994). Isolation and characterization of the faciogenital dysplasia (Aarskog-Scott syndrome) gene: a putative Rho/Rac guanine nucleotide exchange factor. *Cell* **79:** 669-78.

Pati S., and Helmbrecht G. D. (1994). Congenital schizencephaly associated with in utero warfarin exposure. *Reprod Toxicol* **8:** 115-20.

Penrose L. (1938). " A clinical and genetic study of 1280 cases of mental defect. (The Colchester survey).,", London.

Peters U., Haberhausen G., Kostrzewa M., Nolte D., and Muller U. (1997). AFX1 and p54nrb: fine mapping, genomic structure, and exclusion as candidate genes of X-linked dystonia parkinsonism. *Hum Genet* **100:** 569-72.

Philibert R. A., Winfield S. L., Sandhu H. K., Martin B. M., and Ginns E. I. (2000). The structure and expression of the human neuroligin-3 gene. *Gene* **246:** 303-10.

Philibert R., Caspers K., Langbehn D., Troughton E. P., Yucuis R., Sandhu H. K., and Cadoret R. J. (2002). The association of a HOPA polymorphism with major depression and phobia. *Compr Psychiatry* **43:** 404-10.

Puls A., Eliopoulos A. G., Nobes C. D., Bridges T., Young L. S., and Hall A. (1999). Activation of the small GTPase Cdc42 by the inflammatory cytokines TNF(alpha) and IL-1, and by the Epstein-Barr virus transforming protein LMP1. *J Cell Sci* **112:** 2983-92.

Rankin J., Wyttenbach A., and Rubinsztein D. C. (2000). Intracellular green fluorescent protein-polyalanine aggregates are associated with cell death. *Biochem J* **348:** 15-9.

Rogner U. C., Kioschis P., Wilke K., Gong W., Pick E., Dietrich A., Zechner U., Hameister H., Pragliola A., Herman G. E., and et al. (1994). A YAC clone map spanning 7.5 megabases of human chromosome band Xq28. *Hum Mol Genet* **3:** 2137-46.

Rose C. S., Patel P., Reardon W., Malcolm S., and Winter R. M. (1997). The TWIST gene, although not disrupted in Saethre-Chotzen patients with apparently balanced translocations of 7p21, is mutated in familial and sporadic cases. *Hum Mol Genet* **6:** 1369-73.

Rugtveit J. (1986). X-linked mental retardation and infantile spasms in two brothers. *Dev Med Child Neurol* **28:** 544-6.

Saito-Ohara F., Fukuda Y., Ito M., Agarwala K. L., Hayashi M., Matsuo M., Imoto I., Yamakawa K., Nakamura Y., and Inazawa J. (2002). The Xq22 inversion breakpoint interrupted a novel Ras-like GTPase gene in a patient with Duchenne muscular dystrophy and profound mental retardation. *Am J Hum Genet* **71:** 637-45.

Salomons G. S., van Dooren S. J., Verhoeven N. M., Cecil K. M., Ball W. S., Degrauw T. J., and Jakobs C. (2001). X-linked creatine-transporter gene (SLC6A8) defect: a new creatine-deficiency syndrome. *Am J Hum Genet* **68:** 1497-500.

Sasaki T., Kikuchi A., Araki S., Hata Y., Isomura M., Kuroda S., and Takai Y. (1990). Purification and characterization from bovine brain cytosol of a protein that inhibits the dissociation of GDP from and the subsequent binding of GTP to smg p25A, a ras p21-like GTP-binding protein. *J Biol Chem* **265:** 2333-7.

Scheffer I. E., Wallace R. H., Phillips F. L., Hewson P., Reardon K., Parasivam G., Strømme P., Berkovic S. F., Gécz J., and Mulley J. C. (2002). X-linked myoclonic epilepsy with spasticity and intellectual disability: mutation in the homeobox gene ARX. *Neurology* **59:** 348-56.

Schenck A., Bardoni B., Moro A., Bagni C., and Mandel J. L. (2001). A highly conserved protein family interacting with the fragile X mental retardation protein (FMRP) and displaying selective interactions with FMRP-related proteins FXR1P and FXR2P. *Proc Natl Acad Sci U S A* **98:** 8844-9.

Scherer S. S., Deschenes S. M., Xu Y. T., Grinspan J. B., Fischbeck K. H., and Paul D. L. (1995). Connexin32 is a myelin-related protein in the PNS and CNS. *J Neurosci* **15:** 8281-94.

Schroer A., Scheer M. P., Zacharias S., Schneider S., Ropers H. H., Nothwang H. G., Chelly J., Hamel B., Fryns J. P., Shaw P., and Moraine C. (2000). Cosegregation of T108A Elk-1 with mental retardation. *Am J Med Genet* **95:** 404-5.

Shahbazian M., Young J., Yuva-Paylor L., Spencer C., Antalffy B., Noebels J., Armstrong D., Paylor R., and Zoghbi H. (2002). Mice with truncated MeCP2 recapitulate many Rett syndrome features and display hyperacetylation of histone H3. *Neuron* **35:** 243-54.

Sheffield V. C., Beck J. S., Kwitek A. E., Sandstrom D. W., and Stone E. M. (1993). The sensitivity of single-strand conformation polymorphism analysis for the detection of single base substitutions. *Genomics* **16:** 325-32.

Stam N. J., Klomp J., Van de Heuvel N., and Olijve W. (1996). Molecular cloning and characterization of a novel orphan receptor (P2P) expressed in human pancreas that shows high structural homology to the P2U purinoceptor. *FEBS Lett* **384:** 260-4.

Stenflo J., Fernlund P., Egan W., and Roepstorff P. (1974). Vitamin K dependent modifications of glutamic acid residues in prothrombin. *Proc Natl Acad Sci U S A* **71:** 2730-3.

Stevenson R., Schwartz, CE, Schroer, RJ. (2000). "X-linked Mental Retardation," Oxford University Press, New York Oxford.

Stipp C. S., and Hemler M. E. (2000). Transmembrane-4-superfamily proteins CD151 and CD81 associate with alpha 3 beta 1 integrin, and selectively contribute to alpha 3 beta 1-dependent neurite outgrowth. *J Cell Sci* **113:** 1871-82.

Strømme P., Mangelsdorf M. E., Scheffer I. E., and Gécz J. (2002a). Infantile spasms, dystonia, and other X-linked phenotypes caused by mutations in Aristaless related homeobox gene, ARX. *Brain Dev* **24:** 266-8.

Strømme P., Mangelsdorf M. E., Shaw M. A., Lower K. M., Lewis S. M., Bruyere H., Lutcherath V., Gedeon A. K., Wallace R. H., Scheffer I. E., Turner G., Partington M., Frints S. G., Fryns J. P., Sutherland G. R., Mulley J. C., and Gécz J. (2002b). Mutations in the human ortholog of Aristaless cause X-linked mental retardation and epilepsy. *Nat Genet* **30:** 441-5.

Strømme P., Sundet K., Mork C., Cassiman J. J., Fryns J. P., and Claes S. (1999). X linked mental retardation and infantile spasms in a family: new clinical data and linkage to Xp11.4-Xp22.11. *J Med Genet* **36:** 374-8.

Teasdale R. D., Loci D., Houghton F., Karlsson L., and Gleeson P. A. (2001). A large family of endosome-localized proteins related to sorting nexin 1. *Biochem J* **358:** 7-16.

Trivier E., De Cesare D., Jacquot S., Pannetier S., Zackai E., Young I., Mandel J. L., Sassone-Corsi P., and Hanauer A. (1996). Mutations in the kinase Rsk-2 associated with Coffin-Lowry syndrome. *Nature* **384:** 567-70.

Turner G. (1996). Intelligence and the X chromosome. *Lancet* **347:** 1814-1815.

Turner G., Partington M., Kerr B., Mangelsdorf M., and Gécz J. (2002). Variable expression of mental retardation, autism, seizures, and dystonic hand movements in two families with an identical ARX gene mutation. *Am J Med Genet* **112:** 405-11.

Turner G., Webb T., Wake S., and Robinson H. (1996). Prevalence of fragile X syndrome. *Am J Med Genet* **64:** 196-7.

Uchida T., Fujimori F., Tradler T., Fischer G., and Rahfeld J. U. (1999). Identification and characterization of a 14 kDa human protein as a novel parvulin-like peptidyl prolyl cis/trans isomerase. *FEBS Lett* **446:** 278-82.

Utsch B., Becker K., Brock D., Lentze M. J., Bidlingmaier F., and Ludwig M. (2002). A novel stable polyalanine [poly(A)] expansion in the HOXA13 gene associated with hand-foot-genital syndrome: proper function of poly(A)-harbouring transcription factors depends on a critical repeat length? *Hum Genet* **110:** 488-94.

Van Aelst L., and D'Souza-Schorey C. (1997). Rho GTPases and signaling networks. *Genes Dev* **11:** 2295-322.

van der Maarel S. M., Scholten I. H., Huber I., Philippe C., Suijkerbuijk R. F., Gilgenkrantz S., Kere J., Cremers F. P., and Ropers H. H. (1996). Cloning and characterization of DXS6673E, a candidate gene for X-linked mental retardation in Xq13.1. *Hum Mol Genet* **5:** 887-97.

van der Sluijs B. M., van Engelen B. G., and Hoefsloot L. H. (2003). Oculopharyngeal muscular dystrophy (OPMD) due to a small duplication in the PABPN1 gene. *Hum Mutat* **21:** 553.

van Deutekom J. C., Lemmers R. J., Grewal P. K., van Geel M., Romberg S., Dauwerse H. G., Wright T. J., Padberg G. W., Hofker M. H., Hewitt J. E., and Frants R. R. (1996). Identification of the first gene (FRG1) from the FSHD region on human chromosome 4q35. *Hum Mol Genet* **5:** 581-90.

Vekony N., Wolf S., Boissel J. P., Gnauert K., and Closs E. I. (2001). Human cationic amino acid transporter hCAT-3 is preferentially expressed in peripheral tissues. *Biochemistry* **40:** 12387-94.

Venter J. C., Adams M. D., Myers E. W., Li P. W., Mural R. J., Sutton G. G., Smith H. O., Yandell M., Evans C. A., Holt R. A., Gocayne J. D., Amanatides P., Ballew R. M., Huson D. H., Wortman J. R., Zhang Q., Kodira C. D., Zheng X. H., Chen L., Skupski M., Subramanian G., Thomas P. D., Zhang J., Gabor Miklos G. L., Nelson C., Broder S., Clark A. G., Nadeau J., McKusick V. A., Zinder N., Levine A. J., Roberts R. J., Simon M., Slayman C., Hunkapiller M., Bolanos R., Delcher A., Dew I., Fasulo D., Flanigan M., Florea L., Halpern A., Hannenhalli S., Kravitz S., Levy S., Mobarry C., Reinert K., Remington K., Abu-Threideh J., Beasley E., Biddick K., Bonazzi V., Brandon R., Cargill M., Chandramouliswaran I., Charlab R., Chaturvedi K., Deng Z., Di Francesco V., Dunn P., Eilbeck K., Evangelista C., Gabrielian A. E., Gan W., Ge W., Gong F., Gu Z., Guan P., Heiman T. J., Higgins M. E., Ji R. R., Ke Z., Ketchum K. A., Lai Z., Lei Y., Li Z., Li J., Liang Y., Lin X., Lu F., Merkulov G. V., Milshina N., Moore H. M., Naik A. K., Narayan V. A., Neelam B., Nusskern D., Rusch D. B., Salzberg S., Shao W., Shue B., Sun J., Wang Z., Wang A., Wang X., Wang J., Wei M., Wides R., Xiao C., Yan C., et al. (2001). The sequence of the human genome. *Science* **291:** 1304-51.

Verkerk A. J., Pieretti M., Sutcliffe J. S., Fu Y. H., Kuhl D. P., Pizzuti A., Reiner O., Richards S., Victoria M. F., Zhang F. P., and et al. (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:** 905-14.

Vervoort V. S., Beachem M. A., Edwards P. S., Ladd S., Miller K. E., de Mollerat X., Clarkson K., DuPont B., Schwartz C. E., Stevenson R. E., Boyd E., and Srivastava A. K. (2002). AGTR2 mutations in X-linked mental retardation. *Science* **296:** 2401-3.

Villard L., Briault S., Lossi A. M., Paringaux C., Belougne J., Colleaux L., Pincus D. R., Woollatt E., Lespinasse J., Munnich A., Moraine C., Fontes M., and Gécz J. (1999). Two unrelated patients with inversions of the X chromosome and non-specific mental

retardation: physical and transcriptional mapping of their common breakpoint region in Xq13.1. *J Med Genet* **36:** 754-8.

Villard L., Fontes M., Ades L. C., and Gécz J. (2000). Identification of a mutation in the XNP/ATR-X gene in a family reported as Smith-Fineman-Myers syndrome. *Am J Med Genet* **91:** 83-5.

Villard L., Gécz J., Mattei J. F., Fontes M., Saugier-Veber P., Munnich A., and Lyonnet S. (1996). XNP mutation in a large family with Juberg-Marsidi syndrome. *Nat Genet* **12:** 359-60.

Vortkamp A., Gessler M., and Grzeschik K. H. (1991). GLI3 zinc-finger gene interrupted by translocations in Greig syndrome families. *Nature* **352:** 539-40.

Wang P. J., McCarrey J. R., Yang F., and Page D. C. (2001). An abundance of X-linked genes expressed in spermatogonia. *Nat Genet* **27:** 422-6.

Waterston R. H., Lindblad-Toh K., Birney E., Rogers J., Abril J. F., Agarwal P., Agarwala R., Ainscough R., Alexandersson M., An P., Antonarakis S. E., Attwood J., Baertsch R., Bailey J., Barlow K., Beck S., Berry E., Birren B., Bloom T., Bork P., Botcherby M., Bray N., Brent M. R., Brown D. G., Brown S. D., Bult C., Burton J., Butler J., Campbell R. D., Carninci P., Cawley S., Chiaromonte F., Chinwalla A. T., Church D. M., Clamp M., Clee C., Collins F. S., Cook L. L., Copley R. R., Coulson A., Couronne O., Cuff J., Curwen V., Cutts T., Daly M., David R., Davies J., Delehaunty K. D., Deri J., Dermitzakis E. T., Dewey C., Dickens N. J., Diekhans M., Dodge S., Dubchak I., Dunn D. M., Eddy S. R., Elnitski L., Emes R. D., Eswara P., Eyras E., Felsenfeld A., Fewell G. A., Flicek P., Foley K., Frankel W. N., Fulton L. A., Fulton R. S., Furey T. S., Gage D., Gibbs R. A., Glusman G., Gnerre S., Goldman N., Goodstadt L., Grafham D., Graves T. A., Green E. D., Gregory S., Guigo R., Guyer M., Hardison R. C., Haussler D., Hayashizaki Y., Hillier L. W., Hinrichs A., Hlavina W., Holzer T., Hsu F., Hua A., Hubbard T., Hunt A., Jackson I., Jaffe D. B., Johnson L. S., Jones M., Jones T. A., Joy A., Kamal M., Karlsson E. K., et al. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature* **420:** 520-62.

Weeks D. E., Ott J., and Lathrop G. M. (1990). SLINK: general simulation program for linkage analysis. *Am J Hum Genet* **47:** A204.

Willems P., Vits L., Buntinx I., Raeymaekers P., Van Broeckhoven C., and Ceulemans B. (1993). Localization of a gene responsible for nonspecific mental retardation (MRX9) to the pericentromeric region of the X chromosome. *Genomics* **18:** 290-4.

Winnepenninckx B., Errijgers V., Hayez-Delatte F., Reyniers E., and Frank Kooy R. (2002). Identification of a family with nonspecific mental retardation (MRX79) with the A140V mutation in the MECP2 gene: is there a need for routine screening? *Hum Mutat* **20:** 249-52.

Wunderle V.M., Critcher R., Hastie N., Goodfellow P.N. and Schedl A. (1998). Deletion of long-range regulatory elements upstream of SOX9 causes campomelic dysplasia. *Proc Natl Acad Sci U S A.* **95:** 10649-54.

Wyman A. R., and White R. (1980). A highly polymorphic locus in human DNA. *Proc Natl Acad Sci U S A* **77:** 6754-8.

Xu T., Bianco P., Fisher L. W., Longenecker G., Smith E., Goldstein S., Bonadio J., Boskey A., Heegaard A. M., Sommer B., Satomura K., Dominguez P., Zhao C., Kulkarni A. B., Robey P. G., and Young M. F. (1998). Targeted disruption of the biglycan gene leads to an osteoporosis-like phenotype in mice. *Nat Genet* **20:** 78-82.

Yntema H. G., Poppelaars F. A., Derksen E., Oudakker A. R., van Roosmalen T., Jacobs A., Obbema H., Brunner H. G., Hamel B. C., and van Bokhoven H. (2002). Expanding phenotype of XNP mutations: mild to moderate mental retardation. *Am J Med Genet* **110:** 243-7.

Yntema H. G., van den Helm B., Kissing J., van Duijnhoven G., Poppelaars F., Chelly J., Moraine C., Fryns J. P., Hamel B. C., Heilbronner H., Pander H. J., Brunner H. G., Ropers H. H., Cremers F. P., and van Bokhoven H. (1999). A novel ribosomal S6-kinase (RSK4; RPS6KA6) is commonly deleted in patients with complex X-linked mental retardation. *Genomics* **62:** 332-43.

Yu S., Pritchard M., Kremer E., Lynch M., Nancarrow J., Baker E., Holman K., Mulley J. C., Warren S. T., Schlessinger D., and et al. (1991). Fragile X genotype characterized by an unstable region of DNA. *Science* **252:** 1179-81.

Zarrinpar A., and Lim W. A. (2000). Converging on proline: the mechanism of WW domain peptide recognition. *Nat Struct Biol* **7:** 611-3.

Zemni R., Bienvenu T., Vinet M. C., Sefiani A., Carrie A., Billuart P., McDonell N., Couvert P., Francis F., Chafey P., Fauchereau F., Friocourt G., Portes V., Cardona A., Frints S., Meindl A., Brandau O., Ronce N., Moraine C., Bokhoven H., Ropers H. H., Sudbrak R., Kahn A., Fryns J. P., Beldjord C., and et al. (2000). A new gene involved in X-linked mental retardation identified by analysis of an X;2 balanced translocation. *Nat Genet* **24:** 167-70.

Zheng Y., Fischer D. J., Santos M. F., Tigyi G., Pasteris N. G., Gorski J. L., and Xu Y. (1996). The faciogenital dysplasia gene product FGD1 functions as a Cdc42Hs-specific guanine-nucleotide exchange factor. *J Biol Chem* **271:** 33169-72.

APPENDICES

- ***Appendix I.*** Oligonucleotide sequences for PCR amplification and direct sequencing of genes.

- ***Appendix II.*** Strømme P., Mangelsdorf M.E *et al.,* (2002). Mutations in the human ortholog of Aristaless cause X-linked mental retardation and epilepsy. *Nat Genet* **30**: 441-5.

- ***Appendix III.*** Turner G., Partington M., Kerr B., Mangelsdorf M., and Gécz J. (2002). Variable expression of mental retardation, autism, seizures, and dystonic hand movements in two families with an identical ARX gene mutation. *Am J Med Genet* **112**: 405-11.

- ***Appendix IV.*** Strømme P., Mangelsdorf M.E., Scheffer I. E., and Gécz J. (2002). Infantile spasms, dystonia, and other X-linked phenotypes caused by mutations in Aristaless related homeobox gene, ARX. *Brain Dev* **24**: 266-8.

- ***Appendix V.*** Lower, K., Mangelsdorf, M.E. and Gécz, J. (2001). Molecular genetics of X-linked mental retardation: a complex picture emerging. *Expert Rev. Mol. Diagn.* **1**: 220-5.

- ***Appendix VI.*** Mangelsdorf, M., Woollatt, E., Pincus, D., Villard, L., Toniolo, D., Bamforth, J.S., and Gécz, J. Characterisation of X chromosome inversion breakpoints in two patients with mental retardation (Manuscript in Preparation).

Appendix I. Oligonucleotide sequences for PCR amplification and direct sequencing of genes.

Primers were designed from intronic sequences such that PCR amplified the whole exon. All sequences orientated 5'→3'. Large exons were screened in two or more parts, and exons separated by small introns were amplified in one reaction.

Gene	Exon	Forward	Reverse
Ras-related Rab	2 and 3	TGAGGGCGACGATTGGCCTGC	GATGGTGCTACCCAACTGTGATG
	4 and 5	CAGGGATCCAGTCTACCTGACC	GACCCTCCCCATAGAGACTGAAT
	6 and 7	CCTAGGTAAGCATCATTACACCTCAG	CCCTTAGGTACCACTCAGGC
	8	GACCTTGGGTGTTAAGATTTGAC	ACCCAGTAGCTGTGAGATGCG
KIF4	1 and 2	TGCTGACCTACTAACATTCATC	GGTACAGAGACGTTGACAAATG
	3	GAAGGAAATGGAGTATATTGATTAG	GGATTTATTTTTCGACCCAAAGAC
	4	GATTACCTATAGGGTTATGGTCC	GGCAGATAAAAGGGATGTGAAC
	5	GGTTCTACTAGACTATCAAGGC	ATCAATTGAAGAAGCAGTCTAATG
	6	AGTTACTCTCAAGCTTGTGTACC	CAACAATACCAACAAGAGTACTAG
	7	TATATGTTGGATGCATTACCAAGC	CTTTGCTGTCTCAGCACTATC
	8	CGGCTTTGCTTATATCCTACTTG	CTCTATCATTCCACTCACATAC
	9	TACTGCCAAGGGTTAAGTAGC	GGAAAAATAGCCTCTATTAGAATAC
	10	CACTTTTCCCAGACTTTTGTGG	CTGTCTCTCATAATGAGATACATG
	11 and 12	GAGCATAATTATGTGAGGCTGTC	GAAAGGCTTGTATGTGTCAAGAG
	13	ACCAGTAAGCTTTACTTCAGAC	GGAAACGGCAGTTAGAGCTC
	14	GTGCCTAAGGAACTCCATTTGC	CCATTCAGTTTAAGAGACTGTAG
	15	GAGACATGGAATGCTGAAGTG	TTCTTGCAAAGCTCCTCACAC
	16	CTCTGGGAGTCTAGTATTATGTC	CTTGAAATGATTGCTCTGCCTG
	17	ATCCATCCGCACCAGCCTAG	GCTCAGAGAAGGGTTTATTACC
	18	TGTCTTGAGAGACACCTTATAG	GTTTCTCAGGACGCTTGCTAG
	19	TAAACCTGGTTCCTATAGGAAC	CTGTTCCCAGTGTCCACAAGAG
	20 and 21	GAGCTGAATCTAGCCCATGAG	TGTTTTTCCACCATCACAAGGC
	22	TGATATTTAACTGATAGCTCCATC	TTCAAGGCTAACATCCCATTTC
	23	CAAGAAAGCTGACATTATCCTGC	TGCAAAATTTAGAAAAGGCCACG
	24	GGATCACTGCTTGAACTAAGTC	GACAAGATTCTCAGTTAGTACTG
	25	GGTAACCCACATGTAGTGCC	AGCTGGCCTCAGTAGCAACG
	26	CTAGCCTAGGTCTGGTATCTG	TTAGCACACTCACCCCAAACC
	27	CTTTGCTCTGAGAATATAGCTGG	GAAAACTATCCCAGTGGCAGTG
	28	TAAACTAGTCTTCTAACCCTTTC	GTGGAAACTAGAGGAAGGTCTG
	29	CAGACTGATTTTGTAAACCTGGC	CAGAGAAGGACGCTAATCCAG
	30	CATCTATAGCAGCTCGGCTGG	AAGCATCTCCCAAGCCAGAC

Gene	Exon	Forward	Reverse
NLGN3	2	GACCAACTCTGTTGCCCTCC	CATGCTGCAGGGCACACAAC
	3	AGCAAGCCTGGTGGGTGATG	GCTACTTGGAGGAAGGCTGG
	4	TGGGCCACACTGCAGTCATG	GCTTCTCCCCATCGAGAAGG
	5	ACCTGGGATAGCTTTGCTGC	CACCAGCTAGAGAAGCAAGG
	6	GCGTGCTCATTCTCTATTCC	AGAGCTGGCCGATTCCCTTG
	7	CATGGGGCAGCCTCAGTGAC	ATGGAAGAGGTTTAGCTAGAG
	8 (part 1)	GTGGTGACCCCAGATTTCCAT	CGGCTCCCCGCTGAGGGCTAG
	8 (part 2)	CCTCGTGACTACTCCACTGAA	GGAGCCTGGAGATTGGCTGTG
SLC7A3	2	CCTCATTTCCCAATAGCTCCTTTG	GGACGGAGATGGATAGAATG
	3 and 4	GACTTGTGTGAGCCTGGGGTC	TTAGCTAGGTCTTCAAAGCCC
	5 and 6	GGGCTTTGAAGACCTAGCTAA	ACCTGAGACTCTGACAGCAG
	7 and 8	CATACAGTCTTGGGAAGAGAGATC	CTCCATCATTCCTCATCCAAGA
	9 and 10	GAATCTGGGTATCTCACCCTG	GACCTCAAGATATCTGGAGCCTA
	11 and 12	CTGGCCTGCATAGGATAGGG	CACTTCTCTGAGGCTCTTCACT
AFX1	1	AGAACTGTGTGAAGGGACAGC	CAGGGGGTAGATGGTGTTGGG
	2 (part 1)	GACCGCTGGCCACTGACCTCC	CCGGTAACCCCAGGATGCT
	2 (part 2)	CTCGGAGTGGTCTCTCTGG	CTCCCCTTCGCCCAGATCATG
	3	AGAATGTTGGCAAGCCCCAGG	CTTGGCCCTTCCAGATCCCCT
ZNF261	1		
	2 (part 1)	GGAATTCAGGCTAAGACCACC	CCAGCAACTCAGTGGCTCCA
	2 (part 2)	TTGATACCCCTGCTGGCCTG	CTCTAGTAGCCCCTCAGGTG
	2 (part 3)	GGCTGGGGCAAATTCCTGTT	CACCCTCTCCCAGCCCTG
	3	TTCTTCCCCAGGAATGGAGG	GTCCTTCCCAGAGCTAAGGG
	4	CTGGCCAGGGTGTCAGTTTTC	GGACCTTTTCCACACCCAGAAC
	5	CTGTTGGTCAGGGGATGGTTG	CTCATACTCCTGCCCAGATC
	6	CATTGATCCTGGAGGAGGGG	GAGGGGCAGGTCAGGGCTG
	7 (part 1)	CTGTGAAGATCTAGGGTGAGG	ACACTGGTCACAACAGTTGG
	7 (part 2)	CCGGGCCAACAAGGGACTG	AACCCAAGGAACCACACCCC
	8	AATAAAAGTGAAGGGACAAATCC	GGGAGTATGGGAGGTAGAGAG
	9	CCCTCTCTCTTTCTCTTTTCC	TTCTCCAACCCTCTTTATGTT
	10	GCTGCACACATCTTGCCCCT	GACAGTCTGGTCGTGTGGC
	11	GAAGGGGTCCTTCCACTTGG	CTGGCTGGGAGTTATAGTGTC
	12	CTCTTGCCCTTGGGTATCTGTGT	GCTCCCCACTGCCCTCTTTC
	13	GGGGAAAGAAAGGGCCCAAC	AGGATTAGGAGAGTCCAGGC
	14	AAATGGAGAGCTAGTCCCTTG	TGCCCCGACCTTGCCCAAAT
	15	AGGAGGGTACGGCGGTTGG	GGCAGGCCTAGGACTCTGG
	16	AGATGTGACCAGTCCAGCAG	TGATGAGAGTAAGGGGAAAGG
	17	CAACTGCTGGTGTGGTTCTTG	GGGCAGAGACATGCTGGTAG
	18	CTGAAAGCTTGTGAGGGTGC	CCTGATGGCACAAATGCTACC
	19	GGAGGGCTGCAGGGTTGAG	CCTTTGGGCCTCCATTCTTC

Gene	Exon	Forward	Reverse
ZNF261 (cont)	20	CTCTGGCCTTACCTCTGCCA	CTCCTCCTCCCTGTCCCTG
	21	ATGGAACTAAGTGAAATTCCCCAT	CCTGCTGTGGTAAGAGTGACAG
	22	CTTTCTTCTAGGACCTACAGTC	GATGGTGGATTGGAAGAATGG
	23	GGTTGGGTTTGGGTAACCAC	TGGGTTATGGCAGGGCATGG
	24	TGTGGCTCATGGAGCCTGTG	ATCCCTCGGGTCCCCATGC
	25	TGGTACTGATCTTCTGTCCTG	TGGCCACAGGACAGACATTG
SNX12	1		
	2	GTGGCTACCACAGGCTGAACA	ACTCAGCCTATGCAGAGCTCC
	3	GGACAGGATGGACCCTATGC	GACTACGATGGCAAAAGTAGCAA
	4 (part 1)	CCCTCTGCTAATTCTCTCTGAT	CTGGCCCATACTGACTCACC
	4 (part 2)	GAAAGCCATAGGGCAGTAGG	CTGCATAATACCCAGCATATCC
AGTR2	3 (part 1)	GGATGTCCTCAGCTCTGTATGTG	TAAGGCAATCCCAGCTGACCAT
	3 (part 2)	CACCTGCATGAGTGTTGATAGG	GGTGAGCCTCAAAGCAAGTAGCC
PRRG1		TGTAGCCACTCAGCCTCTATC	CAAGAAATACAATTTAGGTGTCCAC
		GTCACTGTCCCTGCCCAGGG	GTGAACCTAGCTCTCCAACCC
		GTTCATTGAACCCAGACATTTATT C	GTGGTATGCGGTAAGAGTCCAA

Fluorescently labelled oligonucleotides for SSCA analysis of the *ARX* gene

Gene	Exon	Forward	Reverse
ARX	1 (part 1)	GCTCACTACACTTGTTACCGC	TTACTTTTGCACTCGGGCCTC
	1(part 2)	CAGCCATGAGCAATCAGTACC	AATTGACAATTCCAGGCCACTG
	2 (part 1)	CAGCAGCCCTGGCTGGGACT	CGTTCTCGCGGTACGACTTGC
	2 (part 2)	GCAGGTGAGCATCAGCCGCA	CAGCTCCTCCTTGGGTGACA
	2 (part 3)	AACTGCTGGAGGACGACGAGG	CGCGACCACCCTACGCGCAT
	3	GAAATAGCTGAGAGGGCATTGC	TCTCTTGGTTTTGTGAAGGGGAT
	4 (part 1)	GACGCGTCCGAAAACAACCTGAG	CAGTCCAAGCGGAGTCGAGCG
	4 (part 2)	GCAGGTGAGCATCAGCCGCA	CCCCAGCCTCTGTGTGTATG
	5	ACAGCTCCCGAGGCCATGAC	GAGTGGTGCTGAGTGAGGTGA

- ***Appendix II.*** Strømme P., Mangelsdorf M.E *et al.,* (2002). Mutations in the human ortholog of Aristaless cause X-linked mental retardation and epilepsy. *Nat Genet* **30**: 441-5.

Strømme, P., Mangelsdorf, M.E., Shaw, M.A., et al., (2002) Mutations in the human ortholog of *Aristaless* cause X-linked mental retardation and epilepsy.
Nature Genetics, v. 30 (4), pp. 441-445.

NOTE:
This publication is included in the print copy of the thesis held in the University of Adelaide Library.

It is also available online to authorised users at:

http://dx.doi.org/10.1038/ng862

- ***Appendix III.*** Turner G., Partington M., Kerr B., Mangelsdorf M., and Gécz J. (2002). Variable expression of mental retardation, autism, seizures, and dystonic hand movements in two families with an identical ARX gene mutation. *Am J Med Genet* **112**: 405-11.

Turner, G., Partington, G., Kerr, B., Mangelsdorf, M., and Gecz, J., (2002) Variable expression of mental retardation, autism, seizures, and dystonic hand movements in two families with an identical *ARX* gene mutation.
American Journal of Medical Genetics, v. 112 (4), pp. 405-411.

NOTE:
This publication is included in the print copy of the thesis held in the University of Adelaide Library.

It is also available online to authorised users at:

http://dx.doi.org/10.1002/ajmg.10714

- ***Appendix IV.*** Strømme P., Mangelsdorf M.E., Scheffer I. E., and Gécz J. (2002). Infantile spasms, dystonia, and other X-linked phenotypes caused by mutations in Aristaless related homeobox gene, ARX. *Brain Dev* **24**: 266-8.

Strømme, P., Mangelsdorf, M. E., Scheffer, I.E., and Gecz, J., (2002) Infantile spasms, dystonia, and other X-linked phenotypes caused by mutations in Aristaless related homeobox gene, *ARX.*
Brain and Development, v. 24 (5), pp. 266-268.

NOTE:
This publication is included in the print copy of the thesis held in the University of Adelaide Library.

It is also available online to authorised users at:

http://dx.doi.org/10.1016/S0387-7604(02)00079-7

- ***Appendix V.*** Lower, K., Mangelsdorf, M.E. and Gécz, J. (2001). Molecular genetics of X-linked mental retardation: a complex picture emerging. *Expert Rev. Mol. Diagn.* **1**: 220-5.

Lower, K., Mangelsdorf, M., and Gecz, J., (2001) Molecular genetics of X-linked mental retardation: a complex picture emerging.
Expert Review of Molecular Diagnostics, v. 1 (2), pp. 220-225.

NOTE:
This publication is included in the print copy of the thesis held in the University of Adelaide Library.

It is also available online to authorised users at:

http://dx.doi.org/10.1586/14737159.1.2.220

Appendix VI: Mangelsdorf, M., Woollatt, E., Pincus, D., Villard, L., Toniolo, D., Bamforth, J.S., and Gécz, J. Characterisation of X chromosome inversion breakpoints in two patients with mental retardation (Manuscript in Preparation).

Characterisation of X chromosome inversion breakpoints in two patients with mental retardation

M Mangelsdorf[1,2], E Woollatt[1], D Pincus[3], L Villard[4], D Toniolo[5], JS Bamforth[6], J Gécz[1,2 #]

[1]*Centre for Medical Genetics, Department of Laboratory Genetics, Women's and Children's Hospital, North Adelaide, SA 5006, Australia.*
[2]*Department of Paediatrics, University of Adelaide, Adelaide, Australia.*
[3] *Allamanda Medical Centre, Southport, Australia*
[4]*Inserm U491, Faculte de Medecine La Timone, 27 Bd Jean Moulin, 13385 Marseille Cedex 5, France*
[5]*IGM-CNR, Via Abbiategrasso 207, 27100 Pavia, Italy*
[6]*Section of Clinical Genetics, Department of Medical Genetics, 853 Medical Sciences Building, University of Alberta, Edmonton, Alberta, T6G 2H7*

X chromosome rearrangements have been a valuable tool for the identification of candidate genes for X-linked mental retardation (XLMR). We have localised the breakpoints of one paracentric inversion and one pericentric inversion of the X chromosome in two unrelated patients with mental retardation. For Patient 1 [46,Y,inv(X)(q13.1q28)] the Xq13.1 breakpoint was previously localised by fluorescence *in situ* hybridisation (FISH) and long range restriction mapping to a 250 kb genomic fragment near *DXS131-DXS162* [1] We have now localised the breakpoint at Xq28, and subsequently shown that the break lies within the 3' untranslated (UTR) region of the biglycan gene (*BGN*). In this patient *BGN* acquires a novel 3' UTR and polyadenylation signal from Xq13.1, but retains apparently normal levels of expression. The sequence across the breakpoints has been determined, and shows that there are no genes at Xq13.1 that are disrupted by the inversion. In Patient 2, (46,Y,inv(X)(p11.2q28), the Xp11.2 breakpoint has been localised by FISH. *ZXDA*, the only gene in the region, is unaltered by the inversion. At Xq28 all BAC clones probed to the patient's chromosomes were found to span the breakpoint, indicating that the patient has a previously unidentified submicroscopic duplication of Xq28 as well as the initially detected inversion [46,Y,rec(X)dup(Xq)inv(X)(p11q28)]. The duplication covers of region of at least 2.7 Mb and thus it is likely that in this patient the phenotype is not due to disruption of a single gene but rather the result of the functional disomy of many Xq28 genes.

Introduction

Intellectual disability is a common disorder estimated to affect 1-2% of the population [2,3]. The proportion of these which are due to alteration of genes on the X chromosome is referred to as X-linked mental retardation (XLMR) and affects 1.66/1000 males [4,5]. XLMR is subdivided into syndromic, where the mental retardation (MR) phenotype is associated with other distinctive clinical features, and non-syndromic (NSXLMR) where MR is the only consistent phenotype within a family. More than 75 families have been published which segregate non-syndromic MR and which have been shown by linkage to map to a region of the X chromosome (based on a lod

score of >+2) (termed MRX) [6]. The regions they are linked to are scattered all along the length of the chromosome. To date there have been 15 genes implicated in NSXLMR however mutations have been found in only 12 of the MRX families. Thus NSXLMR has proven to be a highly genetically heterogeneous condition and there are likely to be many more genes on the X chromosome yet to be associated with the phenotype.

FMR2 was the first gene for NSXLMR identified after patients were found with deletions near the FRAXE fragile. [7]. Of the remaining genes so far identified, 6 were discovered at breakpoints of X chromosome rearrangements, or within deletions in patients with MR, and were subsequently shown to contain mutations in one or more families with MRX (*OPHN1* [8], *IL1RAPL1* [9], *TM4SF2* [10], *ARHGEF6* [11], *FACL4* [12] and *AGTR2* [13]). Two were identified by positional candidate gene screening of MRX families (*GDI1* [14] and *PAK3* [15]). Six became candidates for MRX after they had previously been shown to cause syndromic XLMR (*ATRX* [16], *RSK2* [17], *MECP2* [18], *SLC6A8* [19], *ARX* [20] and *FGD1* [21]) An additional 7 have been implicated in XLMR only by their disruption by breakpoints of X chromosome rearrangements in patients with MR, but have not been shown to contain mutations in any familial cases (*ZNF261* [22], *VCX-A* [23], *RSK4*.[24], *GRIA3* [25], *KLF8* [26], *NXF5* [27] and *RLGP* [28]). History therefore shows that X chromosome rearrangements have proven to be a valuable tool for the identification of candidate genes for non-syndromic XLMR.

We have previously identified a patient with non-syndromic MR who has an inversion of the X chromosome; 46,Y,inv(X)(q13.1q28). In this patient the Xq13 breakpoint was localised to a 250 kb region but no disrupted genes were discovered [1]. A second patient has now been identified that also has an inversion breakpoint at Xq28; inv(X)(p11,q28). Mutations in four genes that cause non-syndromic MR are already located in Xq28 (*FMR2, GDI1*, *MECP2* and *SLC6A8*), however 5 MRX families that map to this region do not have mutations in any of these genes. Many families also map to Xp11 and Xq13, and identification of candidate genes in these regions would also be beneficial. In an attempt to identify one or more candidate genes for XLMR we have characterised the inversion breakpoints at Xq28, Xq13.1 and Xp11.2 in these two patients.

Materials and Methods

CLINICAL SUMMARY

Patient 1

This patient presented with severe learning difficulties and has the karyotype 46,Y,inv(X)(q13.1q28). A more detailed description has been published in [1]

Patient 2

46,Y,inv(X)(p11.2q28)

SEQUENCE DATA, BAC AND COSMID CLONES

Genomic sequence of Xq28 was available from the nr and htgs databases at the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov). BAC contigs were constructed by searching the TIGR BAC end database (http://www.tigr.org/tdb/humgen/bac_end_search). Genes were identified by searching both nr and EST divisions of BLAST. BAC clones were obtained from Research Genetics. DNA was prepared from liquid cultures using QIAGEN plasmid DNA extraction kit (tip 100).

FLUORESCENCE IN SITU HYBRIDISATION

BAC clone DNA for probes for FISH were nick translated with biotin-14-dUTP and hybridised *in situ* at a final concentration of 20 ng/μl to metaphases from the mother of Patient 1 who carries the inversion, and to metaphases from Patient 2. The FISH method was modified from that described by [29] in that chromosomes were stained before analysis with both propidium iodine (as counterstain) and DAPI (for chromosome identification). Images of metaphase preparations were captured by a cooled CCD camera using the CytoVision Ultra image collection and enhancement system (Applied Imaging Int Ltd). FISH signals and the DAPI banding pattern were merged for figure preparation.

INTERPHASE FISH

BAC 3007F1 and BAC 54I20 were directly labelled with SpectrumOrange and SpectrumGreen (Vysis) respectively, according to the manufacturers instructions and co-hybridised at a final concentration of 20ng/ul to interphase nuclei. 100 nuclei were
counted.

RT-PCR, 3'RACE AND SEQUENCING

Total RNA was extracted from patient and control fibroblast cell lines using Trizol reagent (Life Technologies). RT-PCR was performed using SuperscriptII Rnase H^- Reverse Transcriptase (GIBCO BRL) and with different combinations of primers using standard PCR conditions. Primer sequences for amplification of *BGN* were as follows: 6F, 5'- CTG AGA CCC TGA ATG AAC TC -3'; 7R,5'- TTG AGG TCT GGG AGC CCT GA -3'; 7F, 5'- CCT AGG CCA CAA CCA GAT CA -3' and 8R, 5'- ACC AAG CTC AGA AGC GAG AA -3'.

6F was used for 3'RACE using the 3'RACE System for Rapid Amplification of cDNA ends (GIBCO BRL) as per the manufacturers instructions. 3' RACE products were purified using the Ultraclean™ PCR clean-up DNA purification kit (MOBIO laboratories inc) and sequenced using the BigDye terminator cycle sequencing kit as per the instructions of the supplier (Perkin Elmer).

Primers for *ZXDA* screening were as follows: F 5'- ctc tta caa gct caa gag gc 3'; R 5'act aga acc tcc ggt cat cg – 3'. Products were amplified as above with the addition of α^{32}P-dCTP. Products were diluted 1:1 in loading buffer (96% formamide) and loaded on a 10% (49:1) acrylamide gel in 1XTBE and run at 600V overnight. Gels were dried under vacuum and exposed to X-ray film for 3 hours before developing.

LONG RANGE (LR)-PCR AND SOUTHERN BLOTS

Expand Long Template PCR System (Roche) was used for amplifying products greater than 3 kb as per the manufacturers instructions. Products from control samples that did not amplify in the patient was with the following primers F: 5' AGG CCA CCT TTC CTC CAC CT 3' and R: 5' CAC CAC GAC CCA GAA GCC CT 3'. This PCR product was probed to Southern blots containing 8ug of DNA from patient and control samples digested overnight with *Pst*1 and *Hind*III (NEB). The probe was labelled using a random priming reaction with [α^{32}P]-dCTP (NEN) and prereassociated with human placental DNA before hybridisation. Filters were hybridised overnight in 1*M* sodium phosphate (pH7.0) and 7% SDS at 65°C, and were washed under standard conditions.

Results and Discussion

PHYSICAL MAPPING AND FLUORESCENCE IN SITU HYBRIDISATION (FISH)

The cosmid clone E0166, was originally used for FISH in Patient 1 and gave distal signal. An in silico chromosome walk from this point towards the centromere was undertaken. This resulted in physical map and BAC contig across Xq28 between positions 144.6-147.3 Mb (ensembl release 8.30.1) using the publicly available sequences at the time and by searching the BAC end database (TIGR) (Figure 1). This region contains more than 30 known genes including the creatine-transporter gene, SLC6A8 which has recently been found to contain mutations in patients with NSXLMR (Hahn et al, 2002). BAC clones were used as probes to metaphase chromosomes in order to refine breakpoint localizations.

Patient 1

The BAC clone 54I20 was shown to span the Xq28 breakpoint by FISH (Figure 2a), with probes located telomeric and centromeric to this giving distal and proximal signal respectively. This is consistent with the inversion in this patient occurring without any other rearrangements and with the breakpoint being located within this sequence. 54I20 is a large clone of ~250 kb, however located within is the BAC 3011F21 which also showed spanning signal by FISH and hence the breakpoint was localised within this 60 kb.

This 60 kb contains the entire biglycan gene (*BGN*), the 5' end of the brain specific plasma membrane calcium ATPase (*PMCA3*) and part of the exonuclease *TREX2* (Figure 1). LR PCR was used to exclude most of the sequence in the vicinity of these genes from being interrupted by the inversion breakpoint. One LR PCR product was obtained from control DNA, however no product was obtained from Patient 1 (results not shown). This ~3 kb product was hybridised to Southern blots containing *Pst*I and *Hind*III digested patient and control DNA (Figure 3a). In the patient, a 1 kb *Pst*I fragment was absent and was replaced by two junction fragments. These junction fragments were also seen in the mother of the patient who is a carrier of the inverted X (Figure 3a). These experiments placed the breakpoint within the 3'UTR of the *BGN* gene.

To assess whether the Xq28 breakpoint affected the normal expression of *BGN*, RT-PCR was carried out using RNA from cultured fibroblasts of the patient and a control as template (Figure

3b). Expected products were not obtained from the patient when using primers flanking the inversion breakpoint. *BGN* specific primers 5' to the breakpoint, however, gave products of normal size which were comparable in intensity with the control samples. This suggested that in this patient, *BGN* acquired a new 3' UTR and a novel polyadenylation signal from the Xq13 side of the breakpoint.

3'RACE was used to find the new 3'UTR of the *BGN* gene and to obtain the sequence across the breakpoint (Figure 3C and 3D). Indeed there was a new smaller 3'UTR and a new polyadenylation signal identified. The sequence at the breakpoint also contained 23bp of DNA of unknown origin; BLAST searches of GenBank found no significant hits. 3' UTRs are known to contain elements that control mRNA localisation, stability and translation efficiency. Although the ORF of *BGN* is highly conserved between species the 3'UTR is not. There is however, a $(GT)_n$ repeat within the 3'UTR of *BGN* which is present in , however in mouse and rat is an $(AC)_n$ repeat. Whether this has a role in post-transcriptional regulation of *BGN* is unknown. A mouse knockout for this gene has also been described [30]. This knockout has an osteoporosis like phenotype, consistent with the postulated role of *BGN* as a bone matrix proteoglycan. BGN is expressed in brain, and although its function here remains unclear BGN injected into rat brains has been shown to facilitate learning [31].

The breakpoint at Xq13 is within the sequence of the BAC clone CEPHB197N14 (Accession number AL135749) at position 15,884 bp from the T7 end. This is located between the NIMA-interacting peptidyl-prolyl cis/trans isomerase 4 gene (*PIN4)* and a cDNA clone which is weakly similar to the excision repair protein *ERCC-6* (Accession number AK056494), however neither is disrupted by the inversion breakpoint. A positional effect on these genes is a possibility, although seems unlikely as the breakpoint lies downstream of the 3' end of both of these genes.

As the mother of the patient is a carrier of the inverted chromosome there may be other males in the family that have the inversion. The mother reports a history of psychiatric problems among her relatives, however she is unwilling to discuss this genetic problem with them. Therefore, material from other males in the family is unavailable to determine if they do indeed have the inverted X chromosome.

Patient 2

FISH was initially performed using the same BAC clones as for Patient 1 as probes. All probes tested gave spanning signal. There was no obvious reason to speculate that these results were due to normally duplicated regions in Xq28 or low copy X chromosome repeats which might result in multiple signal by FISH. Rather we have interpreted this as the inverted chromosome also containing a duplication of Xq28. BAC clones used cover a region of at least 2.7 Mb of Xq28. This duplication was further confirmed using interphase FISH (figure 2B), where two probes located ~XX kb apart were hybridised to interphase nuclei of the Patient 2. 92/100 nuclei contained two signals from each probe. This was consistent with an Xq28 submicroscopic duplication which had been previously undetected and hence the patient had the karyotype 46,Y,rec(X)dup(Xq)inv(X)(p11q28).

From this same patient FISH localised the Xp11.2 breakpoint within the BAC 966K21. This region is highly repetitive in nature and contains only 1 gene, *ZXDA*, which is an intronless zinc finger gene. Located ~30 kb telomeric is *ZXDB* which share 98.7% DNA sequence identity [32].

To determine if *ZXDA* is disrupted by the breakpoint both PCR and RT-PCR were used. As primers for *ZXDA* also amplify the same sized product from *ZXDA*, the products were separated using SSCA (results not shown). In the patient the expression of both *ZXDA* and *ZXDB* were as normal, and thus the inversion breakpoint does not affect these genes.

Therefore in this patient, there are no genes at Xp11.2 which are affected by the breakpoint. The phenotype in this patient is likely to be due to duplication of many genes in Xq28 rather than disruption of a single gene.

Disomy of Xq28 genes has previously been described in boys with XY_{Xq} syndrome, characterised by deletion of the long arm of the Y chromosome, along with Xq-Yq interchange [33]. Disomy of Xq28 was shown to cause phenotypes other than those associated with boys with 46XYq- karyotype alone, including severe mental retardation, hypotonia and microcephaly. In these cases, and in the case of Patient 2 described here, the phenotypes are likely to be the result of lack of dosage compensation that is normally accounted for by X-inactivation in females. The region duplicated in Patient 2 overlaps with that of the patients with XY_{Xq} syndrome (figure 4), a region containing many genes. Thus it is likely that the intellectual disability in this patient is due to this large duplication and functional disomy of many genes, rather than disruption of any one gene in Xq28.

Conclusion

We have characterised inversions of the X chromosome in 2 patients with MR. In the first case, the 3'UTR of *BGN* is disrupted. The 3'UTRs of some genes are known to contain sequences that are involved in post-transcriptional gene regulation. The 3'UTR of *BGN* houses a polymorphic $(GT)_n$ repeat. Interestingly in mouse and rat, the 3'UTR of BGN contains and (AC)n repeat. 3' UTRs of many genes contain microsatellite repeats, however it is not known if they play any role in gene regulation. Although this gene is expressed in brain its function here is not well understood, however this work presents *BGN* as a candidate gene for screening XLMR families.

For the second patient the rearrangement has proven to be more complex than initially thought, with a large duplication in Xq28. This is a gene rich region and contains several genes already known to be involved in NSXLMR and syndromic XLMR. Disomy of these and many other genes in Xq28 is likely to be the cause of MR in this patient.

ACKNOWLEDGEMENTS

We wish to thank S McDonnell and C Derwas for maintenance of cell lines. MM is funded by the Benjamin Poulton Medical Research Scholarship from the University of Adelaide. This work was supported by the National Health and Medical Research Council of Australia.

References

1. Villard L, Briault S, Lossi AM, Paringaux C, Belougne J, Colleaux L, Pincus DR, Woollatt E, Lespinasse J, Munnich A, Moraine C, Fontes M, Gecz J. Two unrelated patients with inversion of the X chromosome and non-specific mental retardation: physical and transcriptional mapping of their common breakpoint region in Xq13.1. *J Med Genet* 1999;**36**: 754-8.

2. Birch HG, Richardson SA, Baird D, Horobin G, Ilsley R. (1970) Mental subnormality in the community: a clinical and epidemiological study. Baltimore: Williams and Wilkins

3. Moser HW, Ramey CT, Leonard CO. (1983) In: Emery AEH, Rimoin DL (editors) Principles and practice of medical genetics. Edinburgh: Churchill Livingstone.

4. Glass IA. X-linked mental retardation. *J Med Genet* 1991;**28**:361-71.

5. Turner G, Webb T, Wake S, Robinson H. Prevalence of fragile X syndrome. *Am J Med Genet* 1996;**64**:196-7.

6. Mulley JC, Kerr B, Stevenson R, Lubbs H. Nomenclature guidelines for X-linked mental retardation. *Am J Med Genet* 1992;**43**:383-91.

7. Gecz J, Gedeon AK, Sutherland GR, Mulley JC. Identification of the gene *FMR2*, associated with FRAXE mental retardation. *Nat Genet* 1996;**13**:105-8.

8. Billuart P, Bienvenu T, Ronce N, des Portes V, Vinet MC, Zemni R, Crollius HR, Carrie A, Fauchereau F, Cherry M, Briault S, Hamel B, Fryns JP, Beldjord C, Kahn A, Moraine C, Chelly J. *Oligophrenin-1* encodes a rhoGAP protein involved in X-linked mental retardation. *Nature* 1998;**392**:923-6.

9. Carrie A, Jun L, Bienvenu T, Vinet MC, McDonell N, Couvert P, Zemni R, Cardona A, van Buggenhout G, Frints S, Hamel B, Moraine C, Ropers HH, Strom T, Howell GR, Whittaker A, Ross MT, Kahn A, Fryns JP, Beldjord C, Marynen P, Chelly J. A new member of the IL-1 receptor family highly expressed in hippocampus and involved in X-linked mental retardation. *Nat Genet* 1999;**23**: 25-31.

10. Zemni R, Bienvenu T, Vinet MC, Sefiani A, Carrie A, Billuart P, McDonell N, Couvert P, Francis F, Chafey P, Fauchereau F, Friocourt G, des Portes V, Cardona A, Frints S, Meindl A, Brandau O, Ronce N, Moraine C, van Bokhoven H, Ropers HH, Sudbrak R, Kahn A, Fryns JP, Beldjord C, Chelly J. A new gene involved in X-linked mental retardation identified by analysis of an X;2 balanced translocation. *Nat Genet* 2000;**24**:167-70.

11. Kutsche K, Yntema H, Brandt A, Jantke I, Nothwang HG, Orth U, Boavida MG, David D, Chelly J, Fryns JP, Moraine C, Ropers HH, Hamel BCJ, van Bokhoven H, Gal A. Mutations in *ARHGEF6*, encoding a guanine nucleotide exchange factor for Rho GTPases, in patients with X-linked mental retardation. *Nat Genet* 2000;**26**:247-50.

12. Melonie I, Muscettola M, Raynaud M, Longo I, Bruttini M, Moizard MP, Gomot M, Chelly J, des Portes V, Fryns JP, Ropers HH, Magi B, Bellan C, Volpi N, Yntema HG, Lewis SE, Schaffer JE, Renieri A. *FACL4*, encoding fatty acid-CoA ligase 4, is mutated in nonspecific X-linked mental retardation. *Nat Genet* 2002;**30**:436-40.

13. Vervoort VS, Beachem MA, Edwards PS, Ladd S, Miller KE, de Mollerat X, Clarkson K, DuPont B, Schwartz CE, Stevenson RE, Boyd E, Srivastava AK. *AGTR2* mutations in X-linked mental retardation. *Science* 2002;**28**:2401-3.

14. D'Adamo P, Menegon A, Nigro CL, Grasso M, Gulisano M, Tamanini F, Bienvenu T, Gedeon AK, Oostra B, Wu SK, Tandon A, Valtorta F, Balch WE, Chelly J, Toniolo D. Mutations in *GDI1* are responsible for X-linked non-specific mental retardation. *Nat Genet* 1998;**19**:134-9.

15. Allen KM, Gleeson JG, Bagrodia S, Partington MW, MacMillan JC, Cerione RA, Mulley JC, Walsh CA. *PAK3* mutation in non syndromic X-linked mental retardation. *Nat Genet* 1998;**20**:25-30.

16. Yntema HG, Poppelaars FA, Derksen E, Oudakker AR, Van Roosmalen T, Jacobs A, Obbema H, Brunner HG, Hamel BC, van Bokhoven H. Expanding phenotype of XNP muations: Mild to moderate mental retardation. *Am J Med Genet* 2002; **110**;243-7.

17. Merriene K, Jacquot S, Pannetier S, Zeniou M, Bankier A, Gecz J, Mandel JL, Mulley J, Sassone-Corsi P, Hanauer A. A missense mutation in *RPS6KA3* (*RSK2*) responsible for non-specific mental retardation. *Nat Genet* 1999;**22**:13-14.

18. Orrico A, Lam CW, Galli L, Dotti M, Hayek G, Tong SF, Poon PMK, Zappella M, Frederico A, Sorrentino V. *MECP2* mutation in male patients with non-specific X-linked mental retardation. *FEBS Lett.* 2000;**481**:285-8.

19. Hahn KA, Salomons GS, Tackels-Horne D, Wood TC, Taylor HA, Schroer RJ, Lubs HA, Jakobs C, Olson RL, Holden KR, Stevenson RE, Schwartz CE. X-linked mental retardation with seizures and carrier manifestation is caused by a mutation in the creatine-transporter gene (*SLC6A8*) located in Xq28. *Am J Hum Genet* 2002;**70**:1349-56.

20. Stromme P, Mangelsdorf ME, Shaw MA, Lower KM, Lewis SME, Bruyere H, Lutcherath V, Gedeon AK, Wallace RH, Scheffer IE, Turner G, Partington M, Frints SGM, Fryns JP, Sutherland GR, Mulley JC, Gecz J. Mutations in the human ortholog of *Aristaless* cause X-linked mental retardation and epilepsy. *Nat Genet* 2002;**30**: 441-5.

21. Lebel RR, May M, Pouls S, Lubs HA, Stevenson RE, Schwartz CE. Non-syndromic X-linked mental retardation associated with a missense mutation (P312L) in the *FGD1* gene. *Clin Genet* 2002;**61**:139-45.

22. Van der Maarel S, Scholten IHJM, Huber I, Philippe C, Suijkerbuijk RF, Gilgenkrantz S, Kere J, Cremers FP, Ropers HH. Cloning and characterization of *DXS6673E*, a candidate gene for X-linked mental retardation in Xq13.1. *Hum Mol Genet* 1996;**5**:887-97.

23. Fukami M, Kirsch S, Schiller S, Richter A, Benes V, Franco B, Muroya K, Rao E, Merker S, Niesler B, Ballabio A, Ansorge W, Ogata T, Rappold GA. A member of a gene family on Xp22.3, *VCX-A*, is deleted in patients with X-linked nonspecific mental retardation. *Am J Hum Genet* 2000;**67**:563-73.

24. Yntema HG, van den Helm B, Kissing J, van Duijnhoven G, Poppelaars F, Chelly J, Moraine C, Fryns JP, Hamel BCJ, Heilbronner H, Pander H, Brunner H, Ropers HH, Cremers FPM, van Bokhoven H. A novel ribosomal S6-kinase (*RSK4*: *RPS6KA6*) is commonly deleted in patients with complex X-linked mental retardation. *Genomics* 1999;**62**:332-43.

25. Gecz J, Barnett S, Liu J, Hollway G, Donnelly A, Eyre H, Eshkevari HS, Baltazar R, Grunn A, Nagaraja R, Gilliam C, Peltonen L, Sutherland GR, Baron M, Mulley JC. Characterisation of the human glutamate receptor subunit 3 gene (*GRIA3*), a candidate gene for bipolar disorder and non-specific X-linked mental retardation. *Genomics* 1999;**62**:356-68.

26. Lossi AM, Laugier-Anfossi F, Depetris D, Gecz J, Gedeon A, Kooy F, Schwartz C, Mattei MG, Croquette MF, Villard L. Abnormal expression of the *KLF8* (*ZNF741*) gene in a female patient with an X;autosome translocation t(X;21)(p11.2;q22.3) and non-syndromic mental retardation. *J Med Genet* 2002;**39**:113-17.

27. Jun L, Frints S, Duhamel H, Herold A, Abad-Rodrigues J, Dotti C, Izaurralde E, Marynen P, Froyen G. *NXF5*, a novel member of the nuclear RNA export factor family, is lost in a male patient with a syndromic form of mental retardation. *Curr Biol* 2001;**18**:1381-91.

28. Saito-Ohara, F, Fukuda Y, Ito M, Agarwala KL, Hayashi M, Matsuo M, Imoto I, Yamakawa K, Nakamura Y, Inazawa J. The Xq22 inversion breakpoint interrupted a novel Ras-like GTPase gene in a patient with Duchenne Muscular Dystrophy and profound mental retardation. *Am J Hum Genet*. 2002;**71**:

29. Callen DF, Baker E, Eyre HJ, Chernos JE, Bell JA, Sutherland GR. Reassessment of two apparent deletions of chromosome 16p to an ins(11;16) and a t(1;16) by chromosome painting. *Annals in Genetics* 1990;**33**:219-21.

30. Xu T, Bianco P, Fisher LW, Longenecker G, Smith E, Goldstein S, Bonadio J, Boskey A, Heegaard A, Sommer B, Satomura K, Dominguez P, Zhao C, Kulkarni AB, Robey PG, Young MF. Targeted disruption of the *biglycan* gene leads to an osteoporosis-like phenotype in mice. *Nat Genet* 1998;**20**:78-82.

31. Hasenöhrl RU, Frisch C, Junghans U, Müller HW, Huston JP. Facilitation of learning following injection of the chondroitin sulfate proteoglycan biglycan into the vicinity of the nucleus basalis magnocellularis. *Behav Brain Res* 1995;**70**;59-67.

32. Greig GM, Sharp CB, Carrel L, Willard HF. Duplicated zinc finger protein genes on the proximal short arm of the human X chromosome: isolation, characterization and X-inactivation studies. *Hum Mol Genet* 1993;**2**:1611-18.

33. Lahn BT, Ma N, Breg WR, Stratton R, Surti U, Page DC. Xq-Yq interchange resulting in supernormal X-linked gene expression in severely retarded males with 46,XYq- karyotype. *Nat Genet* 1994;**8**:243-50.

Figure legends

Figure 1 a. Physical map and BAC contig of Xq28. STS markers from Xq28 were used to identify genomic sequence from the publicly available database, GenBank. An attempt was made to join sequences with overlapping sequences and constructing a contig. Where a sequence did not overlap with another, the TIGR BAC end database was searched in order to link sequences by the presence of each of a BACs ends. In this way a BAC contig was also constructed across the region. NT numbers represent the current sequence contigs from the Reference Sequence project. 33 known genes are within the sequences available for this region of 2.7 Mb of Xq28. Some regions were not covered by the BAC contig, but later FISH experiments ruled out the need to fill these in. All dark boxes represent BAC clones except 3020P21, which is a PAC clone and E0166 which is a cosmid. 54I20 and 3011F21 were shown to span the inversion breakpoint in Patient 1. 3011F21 is a 60 kb BAC clone which contains the whole of the *BGN* gene, the 3' end of *TREX2,* and a region of homology to the 5' end of the rat *PMCA3* gene, the human ortholog of which is located downstream.

Figure 2 **a**. FISH results for Patient 1.Metaphase chromosomes from the mother of Patient 1 hybridised with BAC 3011F21. On the normal X the probe hybridises to Xq28 only, but shows signal at both Xq28 and Xq13.1 on the inverted X chromosome, indicating that this clones contains the inversion breakpoint. **b** Interphase FISH: Patient 2. Interphase nuclei of Patient 2 were hybridised with BAC clones 3007F1 and 54I20 labelled with SpectrumOrange and SpectrumGreen respectively. 92 of 100 nuclei observed show two orange signals and two green signals, indicating that both of these sequences are represented twice on the inverted X chromosome. This confirms the presence of a submicroscopic duplication of Xq28.

Figure 3. Patient 1: The 3' UTR of BGN is disrupted by the inversion breakpoint. **a**. *Pst* I and *Hind* III Southern blots containing digested DNA from Patient 1 and his mother along with normal controls. The Southern blot was probed with a PCR fragment which was not present in the patient in order to confirm that the breakpoint lay within this fragment. A 1 kb *Pst*I fragment is absent in the patient, with 2 junction fragments, appearing as a doublet, being present in both Patient 1 and his mother. *Hind*III digests also show a junction fragment, however there is no noticeable change in the sizes of the normal bands to see a missing fragment in the patient DNA. The long range PCR and Southern blot placed the breakpoint in the 3' UTR of BGN. **b.**BGN RT-PCR. PCR from exons 6-7 of BGN show that there is transcript present in fibroblast RNA from

patient and normal controls. The intensity of the bands is comparable between patient and control samples (top panel). No product was obtained from patient RNA between exons 7 and 8 which contains the inversion breakpoint (middle panel). (Bottom Panel) Esterase D control reaction. **c.** 3'RACE was performed using primer 6F located 5' to the breakpoint. Patient 1 RNA template results in a transcript of ~650bp compared to normal of 1.8 kb. **d.** Sequence of 3' RACE product in Patient 1. The sequence of the 3' RACE product from **c** matched *BGN* sequence up until position 1139 after the start codon. This is 32 bp after the stop codon. This is followed by 23bp of unidentified sequence, which in turn is followed by sequence from Xq13 (Accession #AL135749).

Figure 4: Xq28 duplications. Patients with XYq- syndrome were shown to be disomic for Xqter (Lahn et al, 1994). The duplication in Patient 2 covers at least 2.7 Mb of genomic DNA which overlaps with the regions of disomy in the XYq- syndrome patients with Xq-Yq interchange.

Figure 1

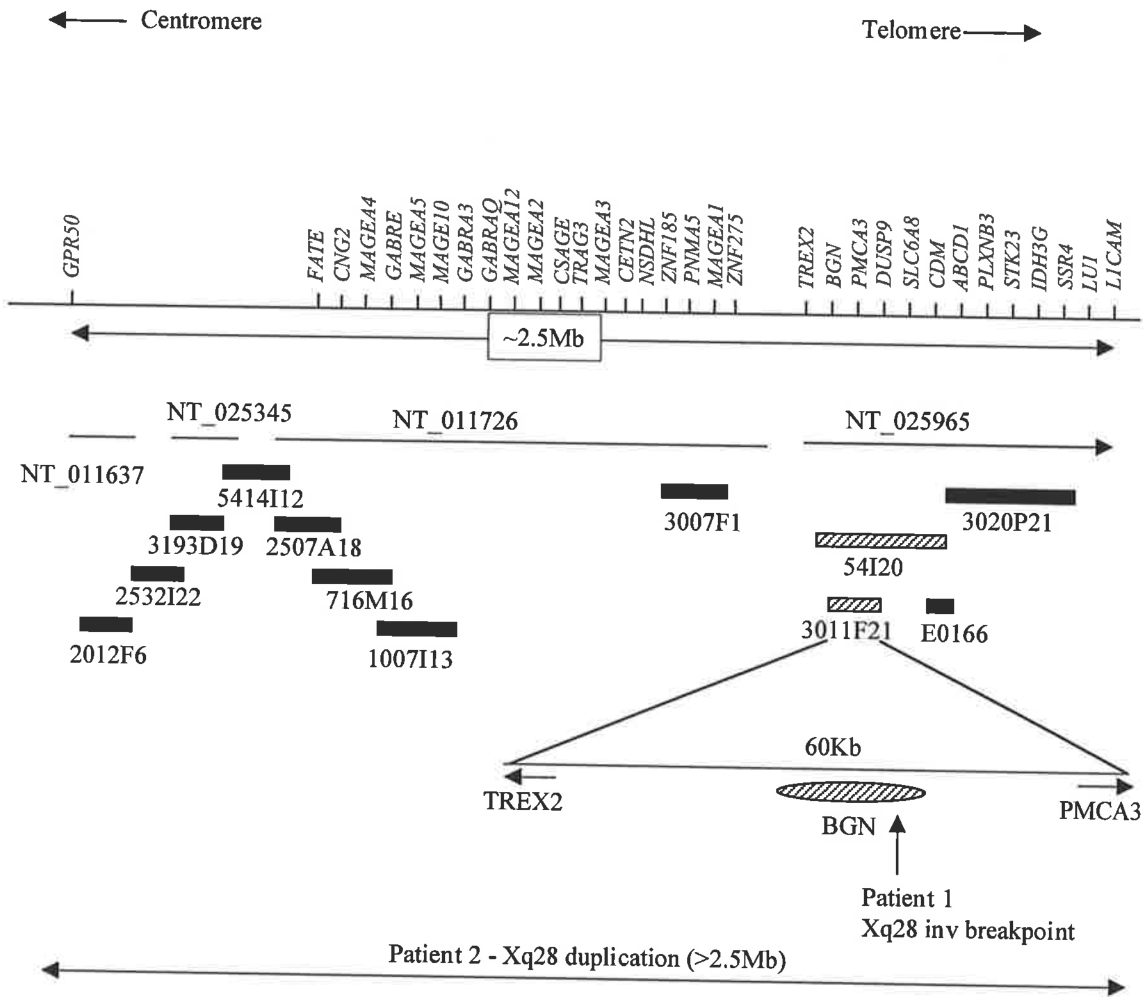
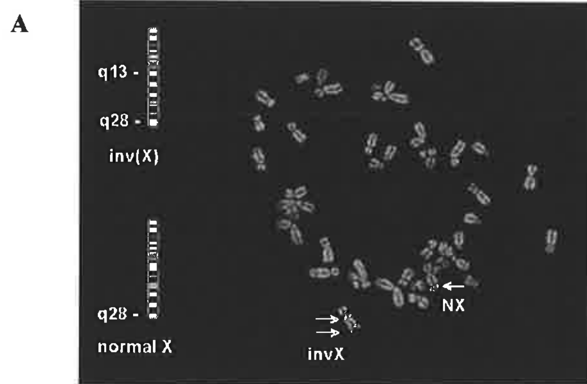


Figure 2

A



B

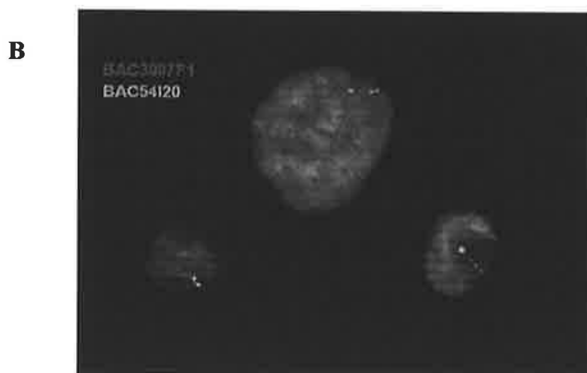


Figure 3

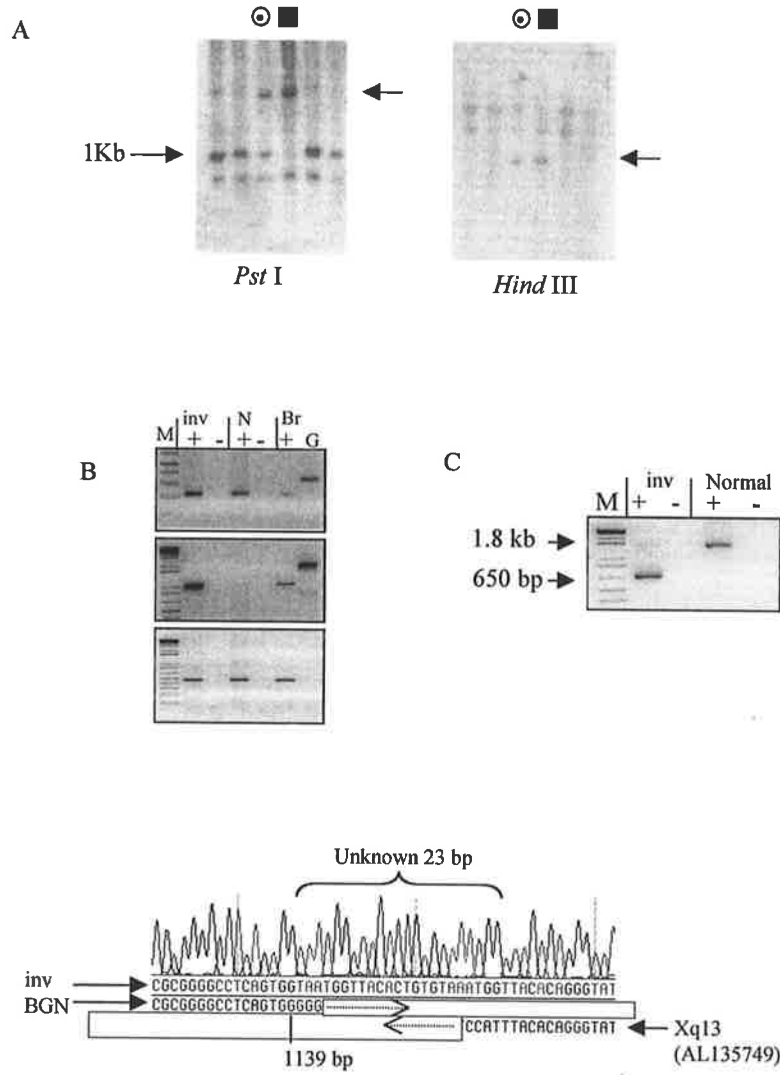


Figure 4