The Role of Aristaless Related Homeobox *(ARX)* Gene Mutations in Intellectual Disability

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

Tod Fullston
B.Sc. (Hons)
Student No 1067053

Neurogenetics Laboratory
Discipline of Paediatrics
School of Paediatrics and Reproductive Health
Women's and Children's Hospital campus

Primary supervisor: Professor Jozef Gécz
Co-supervisor: Professor John Mulley

November 2011

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Intellectual disability (ID) affects ~1-3% of the population, profoundly impacting the lives of affected individuals and their families. An approximate 30% excess of males with ID implicates X-chromosome genes. The most common inherited form of ID is fragile-X syndrome, affecting ~1/5,000 live male births. Another X-linked gene, the aristaless related homeobox *(ARX)* gene, is also frequently mutated causing X-linked ID (XLID).

At least 50 pathogenic mutations spanning the *ARX* open reading frame (ORF) have been reported in 110 families. These mutations cause at least 10 clinically distinct pathologies, all of which include ID. These clinical entities range in severity from X-linked lissencephaly with ambiguous genitalia (XLAG) to mild ID with no other consistent clinical features.

Of the known *ARX* mutations 60% occur in the section of the ORF that encodes for the first two tracts of uninterrupted alanine, *ie* polyalanine (pA) tracts. This is likely due to the extraordinarily high GC content of these regions of the gene (>97%). Two recurrent mutations (c.304ins$(GCG)_7$ – pA1 and c.429_452dup – pA2) arise from expansion of their respective pA tracts. The c.429_452dup mutation alone accounts for ~40% of all reported *ARX* mutations.

To assess the frequency of *ARX* mutations among the intellectually disabled, genomic DNA from 613 individuals were screened for the most frequent *ARX* mutations. Of these, 500/613 samples were screened for mutations in the entire *ARX* ORF by either SSCP, dHPLC or direct Sanger sequencing. A subset of 94/500 patients were also screened for sequence variations in ultraconserved (uc) elements flanking the *ARX* gene, which likely act as *ARX* enhancers. Subsequently, using transient transfection studies we assessed the subcellular localisation of selected mutations and wildtype ARX proteins.

Six different *ARX* mutations were detected in eight individuals (8/613; 1.3%) and potentially pathogenic sequence variations were found in uc elements in three more individuals. A total of five duplication mutations were discovered in pA2, two larger than the recurrent c.429_452dup, confirming exon 2 of *ARX* as a mutation 'hot spot'. Increased aggregation was observed as a function of pA1 and pA2 length, aligning with the patient's phenotypic severity.

Three missense mutations were detected. A familial c.81G>C mutation caused a premature termination codon in exon 1, leading to Ohtahara syndrome (OS) and West syndrome (WS) in two male cousins. Although the c.81G>C mutation should truncate the ARX protein, re-initiation of translation at a down-stream methionine codon (c.121_123) likely occurs, 'rescuing' these patients from the otherwise severe XLAG phenotype.

Two point mutations (c.1074G>T/p.R358S; c.1136G>T/ p.R379L) that alter key residues within the homeodomain were found in two individuals with brain/genital malformations and led to increased ARX protein mislocalisation. These mutations impair vital properties of ARX's transcription factor function by perturbing its localisation into the nucleus (p.R379L) or DNA binding (p.R358S).

This study confirms that *ARX* mutations contribute significantly to XLID and that the majority of mutations occur within exon 2, specifically within the region of pA2. Moreover, there is a correlation between the subcellular localization of the mutant protein and the clinical severity in the patients.

STATEMENT AND DECLARATION

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

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library catalogue, the Australasian Digital Theses Program (ADTP) and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Signed:…….. Date:……..

ACKOWLEDGEMENTS

I sincerely thank members of the Gécz Neurogenetics Program and the Mulley epilepsy research team for their hospitality, assistance and mentoring. Specifically, I thank Bree Hodgson (SSCP), Xenia Iona and Merran Finniss (dHPLC) for their collaboration which I found invaluable for mutations screening. The people involved in the process can make the journey worthwhile and I want to thank the following people individually:

Dr Mark Corbett: For his companionship and always humouring me by not looking too bored during repeated conversations about aligning DNA sequences from mutant polyalanine tracts.

Marie Shaw: Thanks for getting me the job that led to my PhD studies and your friendship throughout my time in the Gécz Neurogenetics laboratory.

Dr Cheryl Shoubridge: Who acted as my surrogate supervisor and who kindly collaborated with me in developing *ARX* expression constructs.

The ghosts of ARX past: Marie Mengelsdorf, Olivia McKenzie and Desiree Cloostermann whose prior work laid the foundation for my project.

Prof. Jozef Gécz: I sincerely thank you for opportunity to undertake research in your lab and provide me with space/resources to do so, and the many valuable lessons you have taught me.

Prof. John Mulley: I may have managed to get lost along the way without your sage guidance and I have greatly appreciated your down to earth mentorship.

Both of my supervisors have managed to inspire me to work towards a career in research through their critical thinking, leadership, mentoring, guidance and expertise.

I would also like to express my gratitude toward my colleagues in my new adoptive lab, who have been very understanding and generously supportive of my situation.

Finally I must thank my out of lab support crew in the form of my long suffering wife and my two most precious daughters who have provided unconditional love and support for 'Team Tod'. My daughters have introduced a beautiful chaos into my life, which makes life so much more rich and rewarding.

'*Constant dripping hollows out a stone*.' Lucretius.

'*Long is the way, and hard, that out of hell leads up to light.*' John Milton.

TABLE OF ABBREVIATIONS

Abbreviation	Full description
A, C, G, T	adenosine, cytosine, guanine, thymine – nucleotides
aa	amino acid
ACC/AG	absence of the corpus callosum with abnormal genitalia
ADI-R	autism diagnostic interview – revised
AG	abnormal genitalia
AGRE	autism genetic research exchange
ANOVA	analysis of variance
ARID	autosomal recessive intellectual disability
aut	autism
BERA	brainstem evoked response audiometry
BLAST	basic local alignnment tool
BLAT	BLAST-like alignment tool
bp	base pairs
c.	coding sequence
CA	cytoplasmic positive with or without aggregates in either the nucleus or cytoplasm
cDNA	complimentary DNA
CGH	comparative genome hybridisation
CNS	central nervous system
CNV	copy number variation – deletion or duplication (> 1 kb)
CSF	cerebrospinal fluid
CT	computerised tomography (scan)
DAPI	4',6-diamidino-2-phenylindole
del	deletion
dHPLC	denaturing high pressure liquid chromatography
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
dup	duplication
EEG	electroencephalogram
EFMR	epilepsy and mental retardation limited to females
EIEE	early infantile epileptic encephalopathy
EMG	electromyogram
epi	epilepsy
exp	expansion
FosTeS	fork stalling and template switching
FXS	fragile-X syndrome
FXTAS	fragile-X associated tremor/ataxia syndrome
GABA	gamma-aminobutyric acid
HEK293T	human embryonic kidney; cell line 293T
hex	hexachlorofluorescein
HYD/AG	hydranencephaly with abnormal genitalia
ID	intellectual disability
ID/TS/Dys	intellectual disability with tonic seizures with dystonia
IEDE	infantile epileptic-dyskinetic encephalopathy
iGOLD	international genetics of learning disability
indels	insertions and deletions and (<1 kb)
ins	insertion
IQ	intelligence quotient
ISSX	infantile spasms syndrome, X linked
kb	kilo base pair

(*Cont. next page*)

Abbreviation	Full description	*Cont.*
LCL	lymphoblastoid cell line	
LGS	Lennox-Gastaut syndrome	
LOD	logarithm of the odds	
Mb	mega base pair	
MGB	minor groove binder	
milliQ	ddH_2O water from a Millipore milliQ system	
MIM	Mendelian inheritance in man – online reference	
miRNA	micro ribonucleic acid	
mis	missense	
MRI	magnetic resonance imaging	
mRNA	messenger ribonucleic acid	
MRS	magnetic resonance spectroscopy	
NGS	next generation sequencing	
NI	nuclear inclusions only	
NLS	nuclear localisation sequence	
NMD	nonsense mediated decay	
non	nonsense	
nsXLID	non syndromic X-linked intellectual disability	
OCF	occipital-frontal circumference	
ORF	open reading frame	
OS	Ohtahara syndrome	
p.	protein residue	
pA	polyalanine tract	
PAGE	polyacrylamide gel electrophoresis	
PBS	phosphate buffered saline	
PCR	polymerase chain reaction	
PRTS	Partington syndrome	
PS	Proud syndrome	
PTC	premature termination codon	
qPCR	quantitative real-time polymerase chain reaction	
RNA	ribose nucleic acid	
RNA	ribonucleic acid	
RT-PCR	reverse transcription polymerase chain reaction	
SDS	sodium dodecyl sulfate	
sil	silent	
SNP	single nucleotide polymorphism	
SSCP	single stranded conformational polymorphism	
STS	sequence-tagged site	
sXLID	syndromic X-linked intellectual disability	
Taq	Thermus aquaticus	
TBS	tris-buffered saline	
uc	ultraconserved elements	
UCSC	University of California, Southern California	
UTR	untranslated region	
UV	ultraviolet	
WS	West syndrome	
wt	wildtype	
XLAG	X-linked lissencephaly and ambiguous genitalia	
XLID	X-linked intellectual disability	
XMESID	X-linked myoclonic epilepsy with spasticity and intellectual disability	

INTRODUCTION

CHAPTER 1 CONTENTS

Parts of this chapter have been incorporated into Shoubridge *et al* (2010a) (Appendix A.1).

1 INTRODUCTION

1.1 WHAT IS INTELLECTUAL DISABILITY?

Intellectual disability (ID) is reported to affect approximately 1-3% of the population (reviewed in (Vasconcelos, 2004)). According to the World Health Organisation guidelines three criteria must be met for an individual to be diagnosed as having ID (WHO, 1996):

i. An intelligence quotient (IQ) below 70, which is two standard deviations below the population average of 100 (Figure 1.1a).

ii. Limitations in adaptive functioning in two of the following skills areas; communication, self care, home living, social skills, use of community resources, self direction, functional academic skills, work, leisure, health and safety.

iii. Age of onset before 18 years old.

The distribution of IQ in the population follows a normal distribution, or bell curve, with 100 set as the mean (Figure 1.1a) (Mackintosh, 1998). ID is categorised by the World Health Organisation (WHO, 1996) into the following five subclasses:

i. Borderline IQ between 70 – 85

ii. Mild IQ between 50 – 70

iii. Moderate IQ between 35 – 50

iv. Severe IQ between 20 – 35

v. Profound IQ $\leq$20

Borderline ID frequently remains undiagnosed in the population, therefore it is possible that prevalence of ID may be greater than the estimated 3% (reviewed in (Vasconcelos, 2004)).

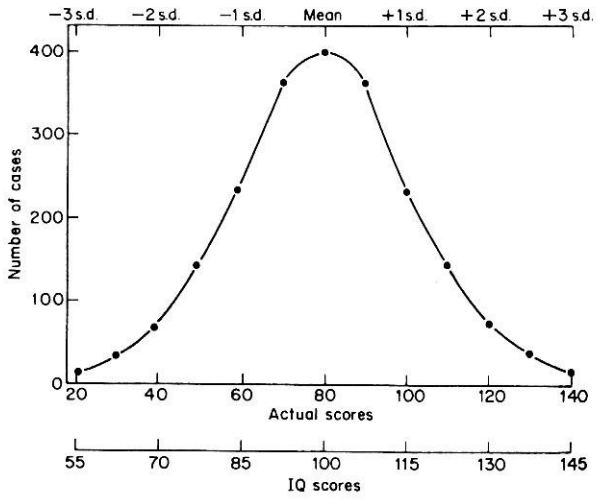

Figure 1.1a. The theoretical normal distribution of IQ scores across the population. The mean is set to 100, and the amount of IQ points that each standard deviation represents. (adapted from (Mackintosh, 1998).

IQ tests are not usually practical for children younger than five years of age, so developmental measures are used instead (Shevell *et al.*, 2003). The term global developmental delay is used when two or more developmental milestones are significantly delayed in the domains such as gross/fine motor skills, speech and language, cognition, personal and social development, or daily living activities (Majnemer and Shevell, 1995). Significant delays are usually defined as two or more standard deviations below the mean by developmental assessment (Shevell *et al.*, 2003). The extent of delay can also be classified by the amount of function retained in an individual's assessment compared to what is expected for their chronological age, *eg* mild <33% reduction; moderate 34-66% reduction; and severe if >66% reduction. A large proportion of individuals classified as having global development delay are subsequently classified as having ID later in life, when IQ testing can be performed (Shevell *et al.*, 2003).

There are two main causes of ID which each comprise approximately half of ID, environmental and genetic. The most common causes of environmental ID include physical brain trauma, prenatal exposure to toxins (including fetal alcohol exposure), malnutrition, and viral infection. There are also characterised genetic determinants that contribute to ID including chromosomal rearrangements (trisomy events, translocations, large duplications

and deletions visible microscopically), copy number variations (CNVs; sub-microscopic, but larger than 1 kb), insertions and deletions (indels; less than 1 kb), monogenic mutations, oligogenic disorders, and alterations to non-coding regulatory sequences.

1.2 GENETIC CAUSES OF INTELLECTUAL DISABILITY

Genetic factors contributing to ID will be expanded upon, focussing on the contribution of monogenic X-linked intellectual disability (XLID), specifically the contribution of the aristaless related homeobox gene (*ARX*) to ID.

The brain is one of the most complex human organs, therefore understanding its development and function is plagued with complexities. One approach to address this challenge has been to identify the plethora of genes that control brain development and function. It is easy to imagine an alteration in one of many genes involved in the development of such an intricate and complex structure as the brain, would affect its development or function. When such a delicate process is interrupted ID is likely to result. To date greater than 90 recessive causative mutations for ID (reviewed by (Gecz *et al.*, 2009)) have been discovered on the human X chromosome and at least a further 83 have been discovered in the autosomes (Najmabadi *et al.*, 2011).

1.2.1 Down syndrome

The most common genetic cause of ID is Down syndrome, with an average prevalence of 1 in 750 live births, although the probability dramatically increases with advanced maternal age (reviewed in (Antonarakis *et al.*, 2004)). The only common features amongst all individuals with Down syndrome are ID and dysmorphic features, which variably associate with other

non-specific pathologies (reviewed in (Gardiner *et al.*, 2010)). Approximately 95% of Down syndrome is due to an extra whole chromosome 21, with the remainder of cases due to 'partial trisomies' and 'microtrisomies' (reviewed in (Antonarakis *et al.*, 2004)). A minimal region has been identified on chromosome 21, that contain genes and non-coding conserved elements, which have been shown to be sufficient to induce Down syndrome (Korbel *et al.*, 2009; Lyle *et al.*, 2009). Although Down syndrome is technically heritable, virtually all cases result from a *de novo* trisomy 21 event and individuals with Down syndrome essentially do not reproduce.

1.2.2 Autosomal recessive intellectual disability (ARID)

As current estimates suggest that X chromosome mutations contribute to approximately 10-12% of ID (reviewed in (Ropers and Hamel, 2005; Ropers, 2008; Gecz *et al.*, 2009; Ropers, 2010)), there is a large scope for the discovery of genes involved in autosomal recessive intellectual disability (ARID). Given the predominantly recessive nature of XLID mutations, parallels can be drawn between XLID and ARID. Autosomal recessive intellectual disability (ARID) is therefore likely to contribute significantly to the understanding of heritable ID and further investigations may highlight underlying molecular mechanisms as targets for therapy. Until recently little was known about ARID owing to low rates of consanguity and small family sizes in Western countries, where the majority of research has been undertaken (Najmabadi *et al.*, 2011). Although, recently autosomal dominant mutations have been suggested to cause a significant amount of sporadic ID (Hamdan *et al.*, 2009; Vissers *et al.*, 2010; Hamdan *et al.*, 2011), individuals with autosomal dominant ID are unlikely to reproduce, thus hampering its heritability.

Large consanguineous families have been invaluable for discovering ARID genes, and investigations of these families has identified virtually all ARID genes identified to date (reviewed in (Ropers, 2010)). Homozygosity mapping within the large consanguineous families have identified many candidate linkage regions throughout the genome (Najmabadi *et al.*, 2007; Kuss *et al.*, 2011). To date mutations in at least 83 genes have been identified as to cause ARID (Najmabadi *et al.*, 2011). The majority of these causative mutations were recently identified in 50 novel ARID genes by targeted deep sequencing of exomes (Najmabadi *et al.*, 2011). The previously identification of homozygous linkage regions known to harbour homozygous defects was critical for this approach. The homozygous linkage regions were specifically targeted by enriching for exons from within the linkage regions by using them as bait prior to deep sequencing of the exome (Najmabadi *et al.*, 2011). It is likely that the discovery of novel ARID genes will be greatly advanced by further application of next generation sequencing approaches. Based on the number of XLID genes indentified (*ie* >90) and that XLID has been suggested to explain 10-12% of ID and that the X chromosome constitutes 4-5% of the genome (reviewed by (Gecz *et al.*, 2009)), the number of ARID genes has been estimated to be in the range of 800-850, with the maximum possible number approaching thousands (reviewed by (Ropers, 2010)).

ARID is an extremely heterogeneous disorder and as such common forms of ARID do not seem to exist (Kuss *et al.*, 2011; Najmabadi *et al.*, 2011). Although some ARID genes have clear brain and cognition functions (*eg PRSS12* and *CRBN* (Molinari *et al.*, 2002; Higgins *et al.*, 2004)), very few of the genes identified have synaptic or neuron specific functions. In fact the majority of genes are involved in essential ubiquitous cellular functions such as DNA transcription, translation, protein degradation, mRNA splicing, energy metabolism and fatty acid metabolism (Najmabadi *et al.*, 2011). It seems logical that genes involved in metabolism can also impact on brain development ultimately leading to ID, as mammalian

brains have a large energy requirement. This energy requirement is approximately 20–25% of resting metabolic rate and 16 times that of skeletal muscle (reviewed by (Leonard *et al.*, 2007)). Further, neuronal cells are enriched in mitochondria and not only are mitochondrial genomic variations involved in regulating brain function (reviewed in (Tuppen *et al.*, 2010)), but variation in mitochondria number is associated with various psychiatric diseases (reviewed by (Curley and Mashoodh, 2010)). This highlights the diverse range of cellular processes that are required for central nervous system development and function and suggests how vulnerable it is to damage.

1.3 X-LINKED INTELLECTUAL DISABILITY (XLID)

In 1938 Lionel Penrose observed a males excess of patients with ID, with a ratio of 1.25:1.0 within the population of institutionalised individuals (reviewed in (Raymond, 2005)). Further investigations confirmed a 30% male excess within individuals with ID in Australia, USA, Canada, and Europe (reviewed in (Raymond, 2005)).

The 30% male excess of ID implicated genes on the X chromosome, is consistent with males being affected more often than females. The excess was initially thought to be explained entirely by genes on the X chromosome, but it is now accepted that it is not explained by monogenic XLID alone, which accounts for approximately half of the 30% male excess (reviewed in (Mandel and Chelly, 2004; Ropers and Hamel, 2005; Gecz *et al.*, 2009)). Other causes of the male ID excess have been suggested, such as male perinatal susceptibility to brain damage and X-linked polymorphisms in regulatory regions with mild ID modifying effects (reviewed in (Mandel and Chelly, 2004; Ropers and Hamel, 2005; Gecz *et al.*, 2009)). Regardless, at least 90 genes on the X chromosome have been discovered with pathogenic mutations that cause XLID (reviewed by (Gecz *et al.*, 2009)). It is currently estimated that

mutations in X chromosome genes contributes to approximately 10-12% of all ID (reviewed in (Ropers and Hamel, 2005; Ropers, 2008)), which far exceeds the ~4% gene content of the X chromosome (reviewed in (Gecz *et al.*, 2009)).

XLID affects between 1/600 to 1/1,000 males. The cardinal feature of XLID is ID that exhibits an X-linked pattern of inheritance throughout a family, coincident with mapping of the causative gene to the X chromosome (reviewed in (Gecz *et al.*, 2009)). A LOD score of equal or greater than 2.0 is accepted as sufficient evidence for linkage to the X chromosome. Traditionally XLID is characterised by only males within the family being affected with ID, with no further transmission from affected males. The assumption that only males are affected by XLID is likely due to an ascertainment bias (Marsh *et al.*, 2009), and leaves many XLID disorders to yet be discovered in females. For example, epilepsy and mental retardation limited to females (EFMR) is caused mutations in the X-linked *PCDH19* gene that exclusively affects females with little or less severe phenotypes in males (Dibbens *et al.*, 2008). A strict set of exclusion criteria for the classification of XLID has been recommended by Shevell *et al* (2003), summarised in Table 1.3a. Collection of detailed clinical information is crucial for retrospective studies that might identify features in common between individuals with mutations in the same gene. These features may not have been originally apparent until their molecular diagnosis linked them.

Table 1.3a. Recommendations for exclusion criteria for investigation and diagnosis of a male child with XLID (derived from Shevell *et al.*, 2003).

Required information:
• Pedigree of 3 generations with details of all potentially affected individuals • Detailed information of clinical history and maternal health pre-pregnancy • Pregnancy history • Birth details, including height, weight, and head circumference • Developmental milestones and growth rates • Neonatal PKU and hypothyroidism • Educational record and IQ • Examination for dysmorphic features and neurological abnormalities • Karyotype (of 550 banded resolution) • Fragile X testing • Telomere screen • Brain MRI (if head circumference or neurological findings warrant it) • EEG to test for Epileptic seizures • Metabolic screen

1.3.1 Syndromic and non-syndromic XLID

Historically XLID has been classified as syndromic (sXLID; accounts for ~1/3 of XLID) or non-syndromic (nsXLID; accounts for ~2/3 of XLID) (reviewed (Mulley *et al.*, 1992; Gecz *et al.*)). sXLID is defined as XLID that consistently presents with other clinical phenotypes between affected individuals such as dysmorphic features, brain abnormalities, seizures, and obesity (reviewed in (Raymond, 2005; Gecz *et al.*, 2009)). nsXLID presents as ID and no other consistent phenotype (reviewed in (Mandel and Chelly, 2004; Raymond, 2005; Ropers and Hamel, 2005; Gecz *et al.*, 2009)). Once the same gene is implicated in different families, a subsequent revision of the clinical presentations of affected individuals can lead to a revised diagnosis from non-syndromal to syndromal XLID (*ie OPHN1* (Higgins and Topaloglu, 2005)). Indeed mutations in at least nine genes are known to cause both nsXLID and sXLID (Table 1.3.1a). In general, mutations in these genes that have the greatest impact on the resultant protein (*ie* truncation mutations) usually cause sXLID. Whereas more subtle

mutations, such as amino acid substitutions, tend to cause nsXLID (reviewed in (Frints *et al.*, 2002a; Gecz *et al.*, 2009)).

Table 1.3.1a. Genes that cause both nsXLID and sXLID when mutated.

Gene	Protein Function	References
ARX	Transcription factor; role in axonal guidance; role in neuronal proliferation/differentiation/ migration of GABAergic interneurons; role in differentiation of endocrine pancreas cells	Stromme *et al.*, 2002a,b Kitamura *et al.*, 2002 Collombat *et al.*, 2003
ATRX	DNA helicase; chromatin remodelling, regulator of cortical size	Gibbons *et al.*, 1995 Berube *et al.*, 2005
FGD1	RhoGEF; important role in regulating actin cytoskeleton organisation and subsequently cell shape and mobility	Pasteris *et al.*, 1994 Hou *et al.*, 2003
JARID1C	Transcription factor; role in chromatin remodelling	Jensen *et al.*, 2005
MECP2	Mediates transcriptional repression of neuronal genes	Amir *et al.*, 1999 Chen *et al.*, 2003
OPHN1	Negative regulator of RhoGTPases; role in stabilization of dendritic spine length	Billuart *et al.*, 1998 Govek *et al.*, 2004
PQBP1	Polyglutamine-binding; role in pre-mRNA processing	Kalscheuer *et al.*, 2003
RSK2	Growth factor regulated serine-threonine protein kinase, phosphorylates CREB and therefore plays a role in learning and memory, role in osteoblast differentiation and function	Trivier *et al.*, 1996 Xing *et al.*, 1996 Harum *et al.*, 2001 Yang *et al.*, 2004
SLC6A8	Creatine transporter	Salomons *et al.*, 2001 Salomons *et al.*, 2003

However, the distinction between nsXLID and sXLID has virtually disappeared. Through a variety of different mutations and interactions with genetic modifiers, mutations in one gene can underlie both syndromal and non-syndromal XLID; *ARX* mutations are a prime example of this phenomenon (reviewed in (Frints *et al.*, 2002a; Gecz *et al.*, 2009); see below). Further, if one mutation results in variable clinical expressivity between individuals, then perhaps multiple genetic inputs are at play, questioning whether these mutations are truly monogenic. Classification based on gene mutations, or other pathogenic DNA variations, has superseded the classification based on clinical presentations.

Structural variations in DNA such as deletions, duplications, translocations and inversions are estimated to contribute to approximately 15-20% to ID (Zahir and Friedman, 2007; Miller *et al.*, 2010). The contribution of pathogenic sub-microscopic copy number variations and indels has been shown to be approximately 12-17% within individuals with ID (Miller *et al.*, 2010). This compares to the rate of approximately 3% using G-banded karyotypes, excluding whole chromosome variations (*eg* Down syndrome; trisomy 21) (Miller *et al.*, 2010). The contribution of pathogenic sub-microscopic copy number variations and indels has been shown to be approximately 10% within a cohort of individuals with XLID (Whibley *et al.*, 2010, Appendix A.10). But this cohort of individuals was pre-screened for larger variations, and 10% is likely an underestimate (Whibley *et al.*, 2010, Appendix A.10). As sequencing technology advances it is possible that more sequence variations in X chromosome genes, regulatory regions, other non-coding regions and structural variations will be ascribed to XLID.

1.3.2 Fragile-X syndrome

Fragile-X syndrome (FXS, MIM 309550) is the most common form of heritable ID, occurring in approximately 1/4000 male births, and 1/8000 female births (reviewed in (Nussbaum and Ledbetter, 1986)). The initial discovery in 1943 by Martin and Bell that a 'mental defect' exhibited sex linkage (Martin, 1943) eventuated in the discovery of FXS (reviewed in (Raymond, 2005)). In 1969 Lubs discovered that affected individuals carried a 'marker chromosome' (reviewed in (Song *et al.*, 2003; Pandey *et al.*)), which was characterised as containing heritable microscopic fragile sites at Xq, inducible by alteration of cell culture conditions (Sutherland, 1979).

Subsequently, FXS was found to be caused by an expansion of a triplet repeat (CCG or CGG) in the 5'UTR of the *FMR1* gene (Oberle *et al.*, 1991; Verkerk *et al.*, 1991; Yu *et al.*, 1991) and this triplet repeat mapped to the precise point of the fragile sites (Xq27.3). If the triplet repeat is increased beyond 200 copies ('full' mutation) then methylation of the repeat region occurs. The cytogenetically visible fragile sites (caused by methylation) and the silencing of the *FMR1* gene are concurrent with methylation of the CCG/CGG triplet repeat region (reviewed in (Song *et al.*, 2003)).

FXS has been estimated to account for approximately 2-2.5% of males with ID and remains the most common cause of XLID (reviewed by (Biancalana and Macpherson, 2004; Rousseau *et al.*, 2011)). Female carriers of mutations and 'premutations' (55-200 CCG/CGG repeats) are eligible for prenatal diagnosis, and recent developments have allowed for pre-implantation genetic selection as a possible means of selecting embryos without the mutation to use for *in vitro* fertilisation (Burlet *et al.*, 2006; Malcov *et al.*, 2007). Whether these tests will impact FXS prevalence is yet to be determined as new expansions of the CCG/CGG repeat to premutation and mutation length are continually being discovered.

The *FMR1* protein (FMRP) has been shown to be involved in translational regulation, transport and localisation of mRNA in neurons, and interaction with micro-RNA (miRNA) molecules via the Dicer complex and the RNA induced silencing complex (reviewed in (Ropers and Hamel, 2005)). FMRP has been shown to play a role in controlling the expression levels of brain-specific miRNAs (Xu *et al.*, 2011). FMRP also interacts with FMR2P (FMR2 protein) in a complex that can interact with exon splice enhancer sequences and catalyse the maturation of mRNAs (reviewed in (Melko and Bardoni, 2010)).

One key effect caused by FMRP absence is the up-regulation of the group 1 metabotropic glutamate receptor (Gp1 mGluR) signalling pathway (Qin *et al.*, 2005; Dolen and Bear,

2009). Gp1 mGluR signalling triggers protein synthesis at synapses in neurons (Weiler and Greenough, 1993). A lack of FMRP increases protein synthesis downstream of Gp1 mGluR signalling and is thought to be a core pathogenic mechanism in FXS (reviewed by (Krueger and Bear, 2011)). It has been demonstrated that an inhibition of mGluR5, a specific sub-type of Gp1 mGluRs, ameliorates seven of eight mutant phenotypes measured in FXS animal models (Dolen and Bear, 2009). These discoveries have lead to the first 'open-label' trial of a mGluR5 antagonist (aripiprazole) to treat individuals with fragile-X syndrome, with some promising clinical outcomes (Erickson *et al.*, 2011).

Interestingly when the number of CCG/CGG repeats in the 5'UTR of the *FMR1* gene is expanded to 55-200 (premutation range) it can lead to fragile-X associated tremor/ataxia syndrome (FXTAS; MIM 300623) in both men and women (reviewed by (Leehey, 2009; Garcia-Arocena and Hagerman, 2010)). The main symptoms of FXTAS are progressive intention tremor, frontal executive dysfunction, global brain atrophy and gait ataxia, which may alternatively present as Parkinson's like tremors and ataxia (Jacquemont *et al.*, 2003). FXTAS was first reported to occur in premutation carrying males in 2001 (Hagerman *et al.*, 2001) and has also been subsequently discovered in female premutation carriers, although the phenotype is usually less severe (reviewed by (Leehey, 2009)). FXTAS has a late adulthood onset of pathology, typically diagnosed in the 7th decade of life, and affects approximately 1 in 3,000 men. FXTAS is less prevalent in women, who often present with the additional pathology of ovarian insufficiency.

In contrast to the full mutation where an absence of *FMR1* mRNA and protein cause FXS, premutation carrying *FMR1* mRNA is approximately 2-8 fold more abundant than normal (Tassone *et al.*, 2000), although FRMP abundance remains unaltered (Primerano *et al.*, 2002). Given that the premutation containing *FMR1* gene is actively transcribed and that the FXTAS

phenotype is restricted to premutation ranges, a toxic gain of function of the premutation carrying mRNA has been proposed (Handa *et al.*, 2003; Hashem *et al.*, 2009). These studies have been confirmed in animal models that demonstrate that the premutation *FMR1* mRNA alone can cause cellular toxicity in mouse (Willemsen *et al.*, 2003; Berman and Willemsen, 2009) and Drosphila (Jin *et al.*, 2003; Jin *et al.*, 2007). Treatment of FXTAS is limited to symptomatic therapy, as no disease modifying therapeutic interventions have been reported to date (reviewed by (Leehey, 2009)).

1.4 THE ARISTALESS RELATED HOMEOBOX (ARX) GENE

The genomic region, which contains the *ARX* gene, was initially implicated in XLID by two independent linkage studies. A 7 Mb region at Xq21.3-q22.1 co-segregated with ISSX (Bruyere *et al.*, 1999) and the same 7 Mb region was refined by locus mapping within a large nsXLID family from the European XLID consortium (MRX54 (Jemaa *et al.*, 1999)). In 2002 two groups independently discovered causative mutations in the *ARX* gene, in both nsXLID families (Bienvenu *et al.*; Stromme *et al.*, 2002b) and sXLID families (ISSX) (Stromme *et al.*, 2002b).

ARX is located on the human X chromosome at Xp22 6,715 bp from the *POLA* gene and is composed of five exons that code for a multi-domain protein (Stromme *et al.*, 2002b) (Figure 1.4a). The *ARX* gene is also flanked by 11 ultraconserved (uc) elements, some of which are the largest in the human genome (see section 4.5 for further details).

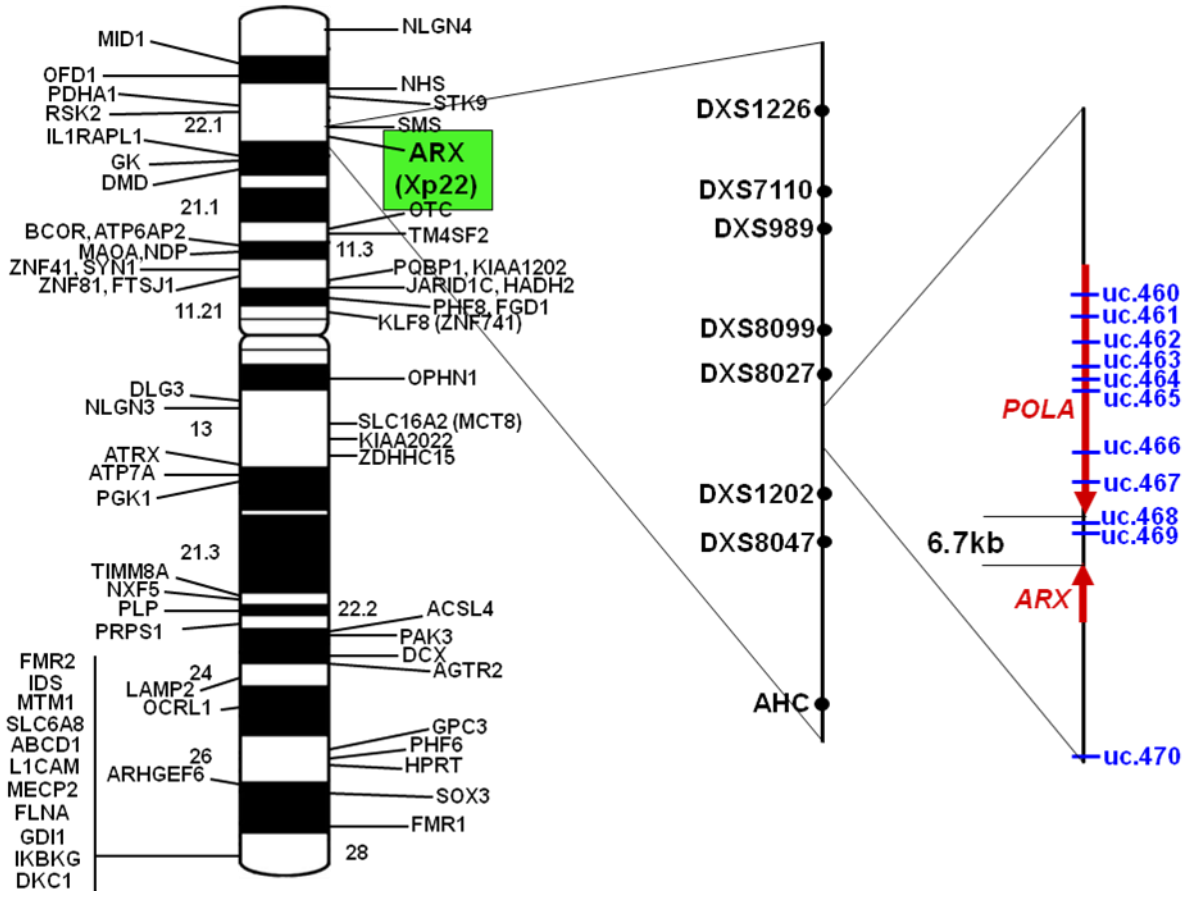

Figure 1.4a. The genomic context of *ARX* at Xp22. Other XLID genes and the relative direction of transcription for *ARX* and *POLA* are shown. Ultraconserved (uc) regions, located within non coding DNA, are shown in blue.

ARX belongs to a large family of 170 human homeobox genes that act in pathways that regulate development and differentiation processes (reviewed in (Shoubridge *et al.*, 2010a, Appendix A.1)). ARX is one of the 26 paired type homeodomain transcription factors. These transcription factors contain a conserved 60 amino acid DNA binding motif, the homeodomain, and play essential roles in neural development throughout development (Bienvenu *et al.*, 2002; Kitamura *et al.*, 2002; Stromme *et al.*, 2002b; Kato *et al.*, 2004). This specific sub-type of transcription factor often referred to as having Q_{50} paired type homeodomain, based on the invariant glutamine (Q) residue at position 50 of 60 amino acids that define the homeodomain. The ARX protein binds DNA via tightly conserved

homeodomain residues and subsequently acts as either a transcriptional activator or repressor, probably depending on available co-factors (Collombat *et al.*, 2005; McKenzie *et al.*, 2007; Colasante *et al.*, 2009; Quille *et al.*, 2011).

Extensive mouse and limited human studies show that *ARX* is important for proper forebrain and testis formation during development (Kitamura *et al.*, 2002; Colasante *et al.*, 2009). Specifically, functional ARX is required for the differentiation and tangential migration of GABAergic interneurons in the ventral telencephalon (Kitamura *et al.*, 2002; Colasante *et al.*, 2009) and for differentiation of alpha/beta islet cell fate in the pancreas (Collombat *et al.*, 2005; Collombat *et al.*, 2007).

The domains that constitute the ARX protein include: an octapeptide domain (oct); three nuclear localisation sequences (NLS); four polyalanie tracts (pA); an acidic domain; a homeodomain (hom); and an Aritaless domain (OAR; Figure 1.4b).

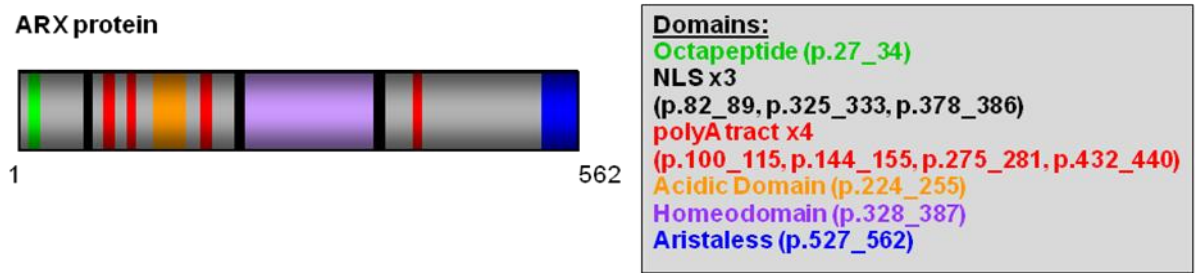

Figure 1.4b. Schematic of the full length ARX protein (562 amino acids). Domains are colour coded as per key.

1.4.1 Mutations in the *ARX* gene

Naturally arising mutations in *ARX* occur across the entire ORF, and are one of the most frequent causes of XLID, after *FMR1* (reviewed by (Gecz *et al.*, 2009)). A comprehensive list of published *ARX* mutations is attached as Appendix B and summarised in Figure 1.4.1a.

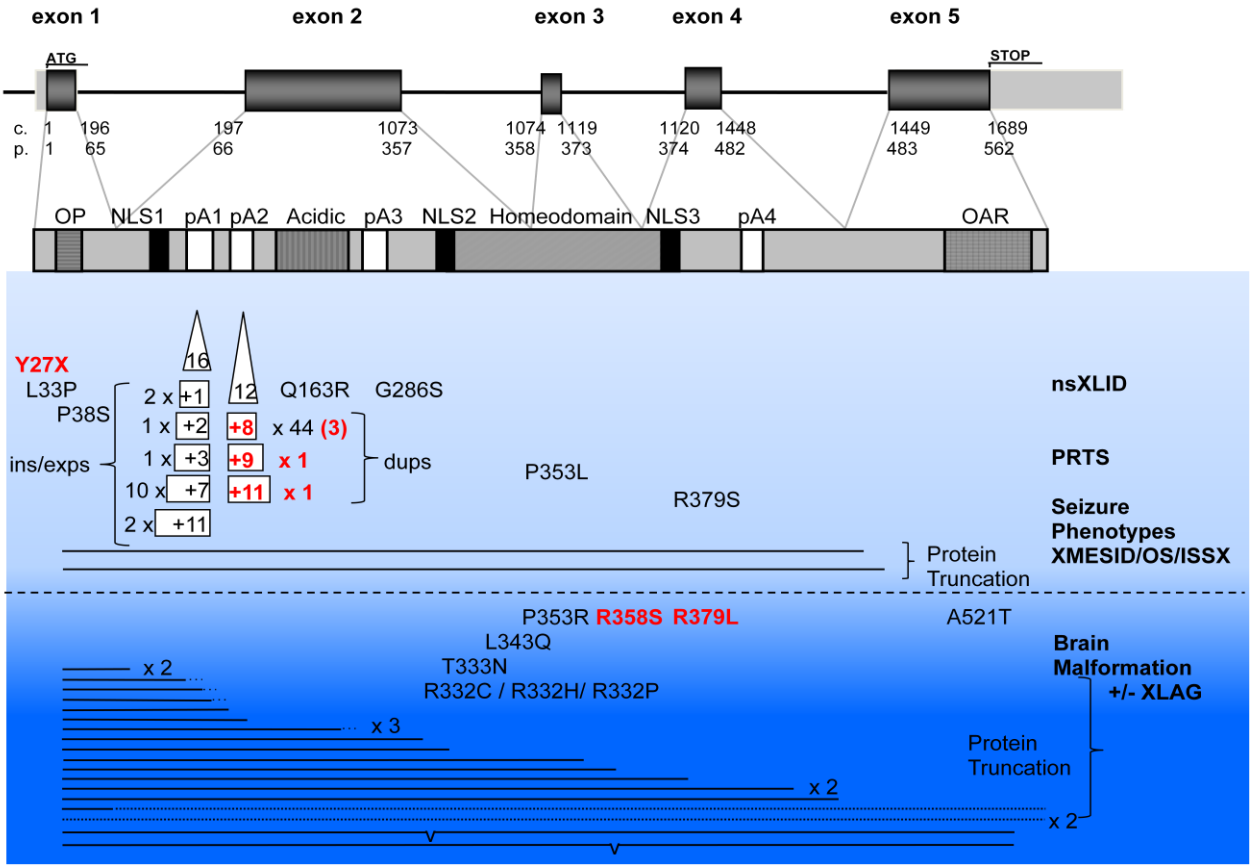

Figure 1.4.1a. Schematic summary of known mutations in ARX. Mutations are annotated as changes to the protein within the known functional domains they occur within: OP – octapeptide (p.27_34); NLS – nuclear localisation sequence (1 – p.82_89; 2 – p.325_333; 3 – p.378_386); pA – polyalanine tract (1 – p.100_115; 2 – p.144_155; 3 – p.275_281; 4 – p.432_440); Acidic domain (p.224_255); Homeodomain (328_387); OAR – aristaless domain (p.527_562). Protein truncating nonsense or frameshift mutations are represented by horizontal bars. The number of families known to carry the mutations is indicated when >1. Mutations that lead to non-malformation phenotypes are above the dotted line in the light blue shading, whilst brain and genital malformation inducing mutations appear below the dotted line in the dark blue shading. The mutations presented as part of this thesis are in red bold text. This figure is adapted from Shoubridge *et al.*, 2010a (Appendix A.1). A detailed list of published *ARX* mutations can be found in Appendix B.

Mutations impair the function of one of various domains within the ARX protein and contribute to at least 10 clinically distinct human disorders (see section 4.2; reviewed by (Gecz *et al.*, 2006; Shoubridge *et al.*, 2010a, Appendix A.1)). To date at least 50 different *ARX* mutations have been reported in 110 families worldwide (including mutations presented as part of this thesis) (reviewed by (Gecz *et al.*, 2006; Shoubridge *et al.*, 2010a, Appendix A.1) and reports by (Giordano *et al.*, 2010; Kato *et al.*, 2010; Conti *et al.*, 2011a; Cossee *et al.*, 2011; Eksioglu *et al.*, 2011; Fullston *et al.*, 2011, Appendix A.2)). The number of families and individuals with *ARX* mutations is likely to be greater than reported, as a large proportion often remain unreported when discovered in a diagnostic setting.

1.4.1.1 Mutations of the polyalanine tract 1

Mutations that expand polyalanine tract 1 (pA1) of ARX are the second most frequent cause of *ARX* related disorders and include the c.304ins$(GCG)_7$ (p.A111insA_7) and c.298_330dup (p.A111insA_{11}) mutations. Prior to the studies presented as part of this thesis 12/102 known families with *ARX* mutations, have been reported with the c.304ins$(GCG)_7$ mutation in pA1 (Stromme *et al.*, 2002b; Wohlrab *et al.*, 2005; Guerrini *et al.*, 2007; Poirier *et al.*, 2008; Wallerstein *et al.*, 2008; Shinozaki *et al.*, 2009; Cossee *et al.*, 2011).

The majority of affected males in 6/12 families with the c.304ins$(GCG)_7$ mutation display phenotypes of severe intellectual disability, generalized dystonia (increased muscle tone with dystonic posturing of limbs), infantile spasms, and a hypsarrhythmic EEG pattern (Stromme *et al.*, 2002b; Wohlrab *et al.*, 2005; Guerrini *et al.*, 2007; Poirier *et al.*, 2008; Wallerstein *et al.*, 2008; Absoud *et al.*, 2010; Cossee *et al.*, 2011). A further four of these families have infantile epileptic–dyskinetic encephalopathy (Guerrini *et al.*, 2007). Of the remaining two

families, one presented with psychomotor delay, epilepsy (generalized tonic–clonic seizures), and generalized dystonia but in the absence of infantile spasms (Shinozaki *et al.*, 2009). The remaining family presented with early infantile epileptic encephalopathy with suppression-burst pattern (EIEE) otherwise known as Ohtahara syndrome (OS) (Absoud *et al.*, 2010).

A second mutation that causes OS and expands pA1 tract length, c.298_330dup, was discovered in two separate families and was the first *ARX* mutation reported to cause OS (Kato *et al.*, 2007). Whilst the c.304ins$(GCG)_7$ mutation expands pA1 tract length by 7 alanines (A) from 16A (wt) to 23A, this c.298_330dup mutation expands pA1 by 11A from 16A to 27A, the longest expansion of any ARX pA tract (Kato *et al.*, 2007).

Smaller expansions of pA1 have been reported: c.304ins$(GCG)_1$ (p.A111insA_1) (Gronskov *et al.*, 2004); c.304ins$(GCG)_2$ (p.A111insA_2) (Bienvenu *et al.*, 2002); and c.304ins$(GCG)_3$ (p.A111insA_3) (Gronskov *et al.*, 2004). Although the c.304ins$(GCG)_1$ variant was discovered in two XLID families, it did not segregate with the disease in one family and was also found in 1/188 control samples, and was therefore reported as a benign rare polymorphism (Gronskov *et al.*, 2004). The c.304ins$(GCG)_2$ variant was found in a brother pair with moderate ID, one brother with epilepsy and the other with aggressive behaviour (Bienvenu *et al.*, 2002). But their mother's c.304ins$(GCG)_2$ carrier status was not reported and full clinical descriptions of family members remains unpublished (family 'T80' in (Bienvenu *et al.*, 2002)). Therefore, the c.304ins$(GCG)_2$ sequence variant is possibly pathogenic. Only the proband was tested for the presence of the c.304ins$(GCG)_3$ variant (Gronskov *et al.*, 2004), thus its pathogenicity is inconclusive. However, given that a 2A expansion of pA1 seems pathogenic (Bienvenu *et al.*, 2002), then a slightly larger 3A expansion is also likely to be pathogenic and, unlike the 1A expansion, neither the 2A or 3A expansion of pA1 has been reported in unaffected individuals of 200 (Bienvenu *et al.*, 2002) and 188 (Gronskov *et al.*,

2004) control X chromosomes tested, respectively). The milder phenotype (if any) observed for the smaller pA1 tract expansions suggests that the longer a mutant pA1 tract is, the more severe the resultant phenotype is likely to be.

No deletions in pA1 have been reported to date.

1.4.1.2 Mutations of the polyalanine tract 2

The c.429_452dup mutation expands pA2 of ARX and is the most common cause of pathology reported for *ARX* related disorders (reviewed by (Gecz *et al.*, 2006; Shoubridge *et al.*, 2010a, Appendix A.1)). Of 102 families with an *ARX* mutation at least 41 (~40%) carry the c.429_452dup mutation (reviewed in (Shoubridge *et al.*, 2010a, Appendix A.1)) and report by (Cossee *et al.*, 2011).

As with all *ARX* mutations, ID is a cardinal clinical feature of all patients with the c.429_452dup mutation. Patients are typically classified as having non-malformation phenotypes with or without additional features that are not consistent between affected males. Affected males display remarkably variable phenotypic expressivity both within and between families (Stromme *et al.*, 2002a; Turner *et al.*, 2002; de Souza Gestinari-Duarte *et al.*, 2006; Gestinari-Duarte Rde *et al.*, 2006; Szczaluba *et al.*, 2006; Laperuta *et al.*, 2007; Rujirabanjerd *et al.*, 2007; Cossee *et al.*, 2011).

The majority (29 of 41) of male probands with the c.429_452dup mutation, are described as having nsXLID (Partington *et al.*, 1988; Bienvenu *et al.*, 2002; Gronskov *et al.*, 2004; Partington *et al.*, 2004; Poirier *et al.*, 2005; de Souza Gestinari-Duarte *et al.*, 2006; Nawara *et al.*, 2006; Poirier *et al.*; Laperuta *et al.*, 2007; Rujirabanjerd *et al.*, 2007). A further 7 of 41

have moderate to severe ID concomitant with dystonic movement of the hands and dysarthria (Partington syndrome – PRTS (Partington *et al.*, 1988)) (Bienvenu *et al.*, 2002; Stromme *et al.*, 2002a; Partington *et al.*, 2004; Van Esch *et al.*, 2004; Stepp *et al.*, 2005; Poirier *et al.*, 2006; Szczaluba *et al.*, 2006; Rujirabanjerd *et al.*, 2007; Cossee *et al.*, 2011), originally named Partington syndrome in 1992 (Mulley *et al.*, 1992). Affected individuals from the remaining five families have an infantile seizure phenotype or X-linked infantile spasms (ISSX or West syndrome (Stromme *et al.*, 2002b; Kato *et al.*, 2003; Poirier *et al.*, 2008; Cossee *et al.*, 2011)). The modifiers that cause the variable clinical expressivity observed for the c.429_452dup mutation have not been identified.

Deletions of either 9 bp (Bienvenu *et al.*, 2002) or 24 bp (Bienvenu *et al.*, 2002; Gronskov *et al.*, 2004; Troester *et al.*, 2007; Conti *et al.*, 2011b) have been reported in pA2. These deletions have been found in healthy control individuals and are usually considered non-pathogenic (Bienvenu *et al.*, 2002; Gronskov *et al.*, 2004; Conti *et al.*, 2011b).

1.4.1.3 Mutations that alter homedomain residues

There are 16 *ARX* mutations reported that are known to either truncate or alter residues within the homeodomain of ARX (Table 1.4.1.3a). Of these 10/16 result in the most severe of *ARX* related disorders, X-linked lissencephaly with ambiguous genitalia (XLAG; MIM# 300215), and a further four result in brain and genital malformations other than XLAG. These mutations are thought to lead to a loss-of-function of the ARX protein as when *Arx* was first deleted in mouse (Kitamura *et al.*, 2002) the phenotype mimicked many clinical aspects of XLAG (Berry-Kravis and Israel, 1994; Dobyns *et al.*, 1999).

Table 1.4.1.3a. Truncation or missense mutations in the homeodomain of ARX. Mutations are listed in order that they occur in the ORF.

Mutation	Exon (Intron)	Change to protein	Phenotype	Reference
1 c.980delAACA	2	p.K327fs	XLAG +	Miyata *et al.*, 2009
2 c.994C>T	2	p.R332C	XLAG	Uyanik *et al.*, 2003
3 c.995G>A	2	p.R332H	XLAG	Kitamura *et al.*, 2002
4 c.995G>C	2	p.R332P	XLAG	Kato *et al.*, 2004
5 c.996_1004del(9bp)	2	p.T333_F335del	XLAG	Bhat *et al.*, 2005
6 c.998C>A	2	p.T333N	ACC/AG	Kato *et al.*, 2004
7 c.1013A>CinsC	2	p.Y338fs	XLAG	Okazaki *et al.*, 2008
8 c.1028T>A	2	p.L343Q	XLAG	Kitamura *et al.*, 2002
9 c.1058C>G	2	p.P353R	XLAG	Kato *et al.*, 2004
10 c.1058C>T	2	p.P353L	XMESID	Stromme *et al.*, 2002b
11 c.1072A>T	2	p.R358W	ACC/AG/ISSX	Conti *et al.*, 2011
12 c.1105G>T	3	p.E369X	HYD/AG	Kato *et al.*, 2004
13 c.1117C>T	3	p.Q373X	HYD/AG	Kitamura *et al.*, 2002
14 c.1119+1G>C	(3)	skipping of ex3	XLAG	Kato *et al.*, 2004
15 c.1135C>A	4	p.R379S	ISSX	Marsh *et al.*, 2009 Shoubridge *et al.*, 2010b
16 c.1187_1188insC	4	p.P396fs	XLAG	Kitamura *et al.*, 2002

ACC: agenesis of the corpus callosum, AG: abnormal genitalia, HYD: hydranenchaply, ISSX: infantile spasms (X-linked), XLAG: X-linked linssencepahly with ambiguous genitalia, +: additional features.

The remaining two mutations, c.1058 C>T (p.P353L) and c.1135C>A (p.R379L), result in either X-linked myoclonic epilepsy with spasticity and intellectual disability (XMESID; MIM# 300432 (Stromme *et al.*, 2002b)) or infantile spasms X linked (ISSX; MIM# 308350 (Marsh *et al.*, 2009; Shoubridge *et al.*, 2010b, Appendix A.9)), respectively. Interestingly, when amino acid residue p.P353 is altered to an arginine (R), XLAG results instead of XMESID. This suggests that the amino acid substitution that confers XMESID (p.P353L) retains some ARX function, but when altered to p.P353R it results in a phenotype akin to complete loss of ARX function.

Three of these point mutations which cause XLAG, occur in the DNA that codes for NLS2 (N-terminal of the homeodomain; Table 1.4.1.3a): c.994 C>T (p.R332C); c.995 G>A (p.R332H); and c.995 G>C (p.R332P). A fourth point mutation also alters a threonine

residue immediately adjacent to this arginine residue, T333N (c.998 C>A), and causes Proud syndrome (ACC/AG; MIM 30004). The basic residues within NLS2 and NLS3 are crucial for the nuclear import of ARX, via interactions with Importin 13 (IPO13) (Shoubridge *et al.*, 2007) and most likely in combination with other Importins (Lin *et al.*, 2009). IPO13 is part of the Importin-β superfamily, which import and export a variety of proteins into/out of the nucleus (Mingot *et al.*, 2001; Ploski *et al.*, 2004), a process catalysed by RanGTP (Lee *et al.*, 2005). This leads to insufficient accumulation and distribution of the ARX transcription factor within the nucleus (Shoubridge *et al.*, 2010b, Appendix A.9).

1.4.1.4 Mutations in the octapeptide domain

Two mutations occur in the DNA that codes for the octapeptide of ARX, c.98T>C (p.L33P (Bienvenu *et al.*, 2002)) and c.196+2T>C (skipping of exon 1 (Kato *et al.*, 2004)). The c.98T>C missense mutation alters an octapeptide domain residue (p.L33P) and was discovered in an individual with a mild nsXLID phenotype (Bienvenu *et al.*, 2002). Given the resultant mild phenotype it is likely that the c.98T>C (p.L33P) mutation retains partial ARX function. The c.196+2T>C mutation causes exon skipping of exon 1 (Kato *et al.*, 2004). An exon 1 lacking *ARX* mRNA would be out of frame, and would presumably be equivalent to a null mutation. The resultant XLAG phenotype supports the c.196+2T>C mutation results in a null allele.

1.4.1.5 Mutations in the aristaless domain

Of the eight mutations that alter the aristaless domain of ARX (Table 1.4.1.5a), two (c.1372delG and c.1419_1420insAC) cause a frameshift mutation ARX >50 amino acids prior to the aristaless domain, causing XLAG. An XLAG phenotype suggests these mutations are equivalent to a null allele. In contrast, the five mutations (c.1449(-816)_*460del, c.1465delG, c.1471_1472insC, c.1564_1568dup and c.1604_1605insT) that cause a frameshift mutation <50 amino acids prior to or within the aristaless domain, lead to less severe phenotypes (ISSX, WS, OS and LGS). A phenotype milder than XLAG suggests that partial ARX function is retained, despite harbouring these mutations. Further, it suggests that critical function of the ARX protein resides in amino acids residues 458 and 489, because when absent by artifact on an upstream frameshift mutation a null allele results.

Of the two missense mutations that occur within/adjacent the aristaless domain, one (c.1604T>A) results in OS, suggesting that the alteration of leucine (L) at residue 535 to a glutamine (Q) retains partial ARX function. The other (c.1561G>A) alters amino acid A521 to a threonine (T), which is actually located 6 amino acids prior to the aristaless domain (p.527_562) and causes XLAG. This phenotype is consistent with a null mutation, and suggests that alanine at residue A521 is essential to ARX function. Curiously, when the p.A521 residue is absent along with the remainder of the aristaless domain XLAG does not necessarily result. Perhaps the substitution to a threonine residue at p.521 is more detrimental to ARX function than its absence, or undetermined genetic or environmental factors have confounded the phenotype in this individual.

Table 1.4.1.5a. Mutations in DNA that codes for the aristaless domain of ARX.

Mutation	Exon (Intron)	Protein	Phenotype	Reference
1. c.1372delG	4	p.A458fs	XLAG	Kitamura *et al.*, 2002 Kato *et al.*, 2004
2. c.1419_1420insAC	4	p.T474fs	XLAG+	Hartmann *et al.*, 2004
3. c.1449(-816)_*460del	(4)	p.R483fs	ISSX (WS)	Stromme *et al.*, 2002b
4. c.1465delG	5	p.A489fs	ISSX (WS) F	Wallerstein *et al.*, 2008
5. c.1471_1472insC	5	p.L491fs	OS	Ekşioğlu *et al.*, 2011
*6. c.1561G>A	5	*p.A521T	XLAG+	Kato *et al.*, 2004
7. c.1564_1568dup	5	p.A524fs	OS/WS +	Kato *et al.*, 2010
8. c.1604T>A	5	p.L535Q	OS	Giordano *et al.*, 2010
9. c.1604_1605insT	5	p.E536fs	OS/WS/LGS	Kato *et al.*, 2010

*: mutation is adjacent to the homeodomain, LGS: Lennox-Gastaut syndrome.

1.4.1.6 Mutations outside of known domains

A total of 11 mutations occur outside of all known ARX domains (Table 1.4.1.6a). Three of these are missense mutations (c.112C>T, c.488A>G and c.856G>A) that lead to a nsXLID phenotype, suggesting that these residues (p.P38, p.Q163 and p.G286S), which are outside of known domains are not crucial to ARX function. However the remaining frameshift mutations cause XLAG (Table 1.4.1.6a), presumably due to the lack of downstream domains critical to ARX function (*ie* they all lack both the homeodomain and the aristaless domain).

Table 1.4.1.6a. Mutations outside of known domains within ARX.

Mutation	Exon (Intron)	Protein	Phenotype	Reference
1. Exon1_2del	1-2	p.M1_357del	XLAG	Kitamura *et al.*, 2002
2. c.112C>T	1	p.P38S	nsXLID	Poirier *et al.*, 2005b
3. Exon2_5del	2-5	p.G66_C562del	XLAG	Kato *et al.*, 2004
4. c.232G>T	2	p.E78X	XLAG	Kato *et al.*, 2004
5. c.392_452del	2	p.P131fs	XLAG	Kato *et al.*, 2004
6. c.420_451del	2	p.D140fs	XLAG	Kitamura *et al.*, 2002
7. c.488A>G	2	p.Q163R	nsXLID	Bienvenu *et al.* 2002
8. c.617delG	2	p.G206fs	XLAG/HYD	Kato *et al.* 2004
9. c.619_647del	2	p.A216delfs	XLAG	Kato *et al.* 2004
10. c.790delC	2	p.R264fs	XLAG	Kitamura *et al.* 2002 Uyanik *et al.* 2003 Hahn *et al.* 2004
11. c.856G>A	2	p.G286S	nsXLID	Bienvenu *et al.* 2002

HYD: hydranenchaply.

1.4.1.6 Copy number variations

Although mutations in many different X-linked genes cause XLID, approximately only 25% of XLID was directly explained by obvious mutations within coding DNA, using large-scale systematic re-sequencing of >7000 exons on the X chromosome in 208 families (Tarpey *et al.*, 2009, Appendix A.4). Many missense variations of uncertain pathogenicity were also discovered and further studies may subsequently confirm whether theses mutations are indeed pathogenic, *eg IQSEC2* (Shoubridge *et al.*, 2010c). If it turns out that a large proportion of the missense variations are proven pathogenic, then as much as ~70% XLID maybe be explained by mutations within X-chromosome coding regions. To date, the basis of XLID remains unexplained in the majority of families within this cohort. It is therefore possible that alterations within the X-chromosome that are not effectively detected by exon re-sequencing are candidates for pathogenicity.

Copy number variations (CNVs) fulfill this role, as they are a type of DNA mutation that are not effectively detected by direct Sanger sequencing of exons, but next generation sequencing technologies have the potential to do so. CNVs are genomic deletions or amplifications greater than 1 kb, which most commonly found as duplications. Deletions and duplications smaller than 1 kb are known as 'indels' (Scherer *et al.*, 2007). Duplications are not detected by conventional PCR/Sanger sequencing as doubling the amount of target DNA does not interfere with PCR amplification or the sequencing reaction. Although deletions in the X-chromosome in DNA from males would result in PCR failure, as one/both of the primer binding sites may be deleted. As PCR failure occurs for many reasons besides deletion of target sequences, this can confound the misinterpretation of such results. Again next generation sequencing technologies have the potential to overcome this type of complication, if sufficient coverage is achieved. Small deletions fortuitously flanked by the primer binding sites could be detected by the presence of smaller than expected PCR products.

CNVs can cause genetic diseases that include both neurological disorders (Lupski, 1998; Lee and Lupski, 2006) and ID (Qiao *et al.*, 2010). An average of 12.2% of patients with intellectual disability, developmental delay, dysmorphism and autism spectrum disorders have pathogenic CNVs, according to a recent meta-analysis of 33 original array comparative genome hybridisation (CGH) that reports on a total of 21,968 individuals (Miller *et al.*, 2010). Pathogenic CNVs that cause ID have been discovered using whole genome microarray, and as such this technique is currently incorporated into routine clinical practice as a first tier approach (Cheung *et al.*, 2007; Stankiewicz and Beaudet, 2007; Koolen *et al.*, 2009; Miller *et al.*, 2010). Traditionally CNVs that recurrently associate with disease are classified as pathogenic (Sharp *et al.*, 2008). But it remains possible that a significant number of novel non-recurrent or rare CNVs may well be identified as either causing or conferring susceptibility to ID (Ionita-Laza *et al.*, 2009; Miller *et al.*, 2010).

Previously, in the murine genome, a large duplication that flanked the contiguous *Arx* and *PolA* genes resulted in a clear eye and forebrain phenotype (exencephaly, microphthalmia and anophthalmia). This demonstrates that a large duplication incorporating the *ARX* locus can result in brain pathology (Cunningham *et al.*, 2002). But because other genes were also duplicated this pathology cannot be ascribed solely to the duplication of *ARX*. Two duplications at Xp21-22 that include *ARX* and also other known XLID genes have been associated with ID, but the relative proportion of the contribution of *ARX* in conjunction with the other genes remains unclear (Thorson *et al.*, 2010). An *ARX* locus deletion has been reported in a Yoruba Nigerian cohort, but sex or phenotype of the individual involved remains unreported (Matsuzaki *et al.*, 2009). Presumably an entire *ARX* locus deletion would results in XLAG in hemizygous males, consistent with other partial deletions (>1kb) of the *ARX* locus that lead to a loss of function (Kitamura *et al.*, 2002; Kato *et al.*, 2004).

1.4.2 Phenotypes of affected males with *ARX* mutations

An *ARX* mutation will affect the function of one of a number of different domains in the ARX protein (see Figure 1.4.1a). Various mutations therefore contribute to 10 clinically distinct disorders (Table 1.4.2a; reviewed by (Gecz *et al.*, 2006; Shoubridge *et al.*, 2010a, Appendix A.1).

Table 1.4.2a. Ten clinically distinct phenotypes are observed for mutations in the *ARX* gene, in order of discovery.

Phenotype	Acronym	MIM ID #[1]	Reference
X-linked infantile spasms (or West syndrome)	ISSX [2] (or WS)	308350	Strømme *et al.*, 2002a,b
Nonsyndromic X linked intellectual disability (ID)	nsXLID [2]		Strømme *et al.*, 2002b Bienvenu *et al.*, 2002
Partington syndrome	PRTS [2]	309510	Strømme *et al.*, 2002b Partington *et al.*, 2004
X-linked myoclonic epilepsy with ID and spasticity	XMESID [2]	300432	Strømme *et al.*, 2002b Scheffer *et al.*, 2002
X-linked lissencephaly with ambiguous genitalia	XLAG [3]	300215	Kitamura *et al.*, 2002 Kato *et al.*, 2004
Absence of the corpus callosum with abnormal genitalia (or Proud syndrome)	ACC/AG [3] (or PS)	300004	Kato *et al.*, 2004
Hydranencephaly with abnormal genitalia	HYD/AG [3]	300215	Kato *et al.*, 2004
Infantile epileptic-dyskinetic encephalopathy	IEDE [2]	308350	Guerrini *et al.*, 2007
Ohtahara syndrome	OS [2]	308350	Kato *et al.*, 2007
Intellectual disability with tonic seizures with dystonia	ID/TS/Dys [2]	309510	Shinozaki *et al.*, 2009

[1] Mendelian inheritance in man identification number
[2] non-malformation phenotype (shaded grey)
[3] malformation phenotype (unshaded)

All patients with *ARX* mutations have intellectual disability (ID) with and without additional features including epilepsy, infantile spasms, autism, dystonia, dysarthria, genital malformations and brain malformations (reviewed by (Gecz *et al.*, 2006; Shoubridge *et al.*, 2010a, Appendix A.1)). Clinical features of these patients can be grouped into either: (i) non-malformation phenotypes (shaded grey in Table 1.4.2a), or (ii) brain and genital malformation (unshaded in Table 1.4.2a).

X-linked infantile spasms (ISSX), or West syndrome (MIM 308350) (Stromme *et al.*, 2002a; Stromme *et al.*, 2002b) presents within the first year of life and is defined by infantile spasms, hypsarrhytmia ('noisy' EEG pattern), and moderate to severe ID. Infantile spasms

are brief involuntary flexions of muscles along the trunk, neck, and limbs, which usually precede a period of atonia (Bruyere *et al.*, 1999).

Partington syndrome (PRTS; MIM 309510) includes a phenotype of mild to moderate ID, facial muscle weakness, difficulty in articulating, and the hallmark characteristic dystonic movements of the hands (Partington *et al.*, 1988; Frints *et al.*, 2002b). The dystonic hand spasms include flexion of the wrist and extension of the joints that last for up to 30 seconds at a time.

X-linked myoclonic epilepsy with generalized spasticity and intellectual disability (XMESID; MIM 300432) is caused by a c.1058C>T mutation in *ARX*. Interestingly, carrier females were discovered to also have phenotypes (see section below).

Infantile epileptic-dyskinetic encephalopathy (IEDE; MIM 308350) is a form of infantile spasms that includes severe dyskinetic quadriparesis (Guerrini *et al.*, 2007).

Ohtahara syndrome (OS; MIM 308350) is an early infantile epileptic encephalopathy with the earliest known age of onset for any age-dependant epileptic syndrome. Epileptic seizures often present within days of birth, or possibly observed prenatally as violent fetal movement (Yamatogi and Ohtahara, 2002; Ohtahara and Yamatogi, 2006). OS manifests as frequent minor generalized seizures and distinct 'burst-suppression' activity on EEG readouts and is distinct from other neonatal epileptic encephalopathies (Ohtahara and Yamatogi, 2006). Seizures are intractable and occur concurrently with severe psychomotor disability.

Intellectual disability with tonic seizures with dystonia (ID/TS/Dys; MIM 309510) presented as ID, tonic seizures and dystonia but without infantile spasms (Shinozaki *et al.*, 2009).

The most frequent malformation phenotype reported to result from *ARX* mutations is X-linked lissencephaly with ambiguous genitalia (XLAG; MIM 300215). The XLAG syndrome

involves many features including: (i) lissencephaly with a posterior-to-anterior gradient and intermediate (5–7 mm) increase in cortical thickness; (ii) agenesis of the corpus callosum (ACC); (iii) intractable epilepsy with onset prenatally or during the first hours of life; (iv) hypothalamic dysfunction manifest by persistent poor temperature regulation; (v) severe hypotonia from birth; and (vi) ambiguous genitalia in affected males (Berry-Kravis and Israel, 1994; Dobyns *et al.*, 1999). XLAG has a distinct form of lissencephaly that displays an intermediate thickening of the cortex with a posterior to anterior gradient of gyral malformation (Dobyns *et al.*, 1999).

Other rarer brain and genital malformation phenotypes distinct from XLAG also result from mutations in *ARX*, such as: (i) absence of the corpus callosum with abnormal genitalia, also known as Proud Syndrome (MIM 300004) (Kato *et al.*, 2004); and (ii) Hydranencephaly with abnormal genitalia (MIM 300215) (Kato *et al.*, 2004).

1.4.2.1 Variable clinical expressivity of *ARX* mutations

The most common mutation in *ARX* is the c.429_452dup, which constitutes ~40% of all families reported with an *ARX* mutation to date (reviewed by (Gecz *et al.*, 2006; Shoubridge *et al.*, 2010a, Appendix A.1)). The c.429_452dup mutation gives rise to a number of different phenotypes and syndromic forms of XLID. The degree of ID seen within the c.429_452dup mutation cohort ranges from mild to severe. Different forms of sXLID are caused by the c.429_452dup mutation in *ARX* that include: (i) Partington syndrome (PRTS) (Partington *et al.*, 1988; Frints *et al.*, 2002b; Stromme *et al.*, 2002b); and (ii) X-linked Infantile spasms (ISSX) (Stromme *et al.*, 2002a,b).

Variable phenotypic expressivity is observed between patients with nsXLID and the c.429_452dup mutation, manifesting in different additional phenotypes within individuals, for example: epilepsy (Bienvenu *et al.*, 2002); autistic traits (Turner *et al.*, 2002); diabetes (Turner *et al.*, 2002); tonic clonic seizures (Stromme *et al.*, 2002a; Turner *et al.*, 2002); complex partial seizures (Partington *et al.*, 2004); dystonia (Bienvenu *et al.*, 2002; Stromme *et al.*, 2002a; Turner *et al.*, 2002); hypotonia (Kato *et al.*, 2004); brain cysts (Stromme *et al.*, 2003); agenesis of the corpus callosum (Van Esch *et al.*, 2004); and macrocephaly (Van Esch *et al.*, 2004)

Interfamily variability of clinical expressivity can be partly rationalised by different genetic backgrounds between different families. Curiously c.429_452dup mutation also shows phenotypic variability within the same family (Stromme *et al.*, 2002a; Turner *et al.*, 2002; Gestinari-Duarte Rde *et al.*, 2006; Szczaluba *et al.*, 2006; Laperuta *et al.*, 2007; Rujirabanjerd *et al.*, 2007; Cossee *et al.*, 2011). To date there has been no reported founder effect of the c.429_452dup mutation (Stromme *et al.*, 2002b).

Other mutations (*eg* c.304ins$(GCG)_7$ and c.81C>G) also display variable clinical expressivity within members of the same family. This implies that other as yet undetermined genetic and environmental modifiers are likely to have an impact on the resultant phenotype of an individual with an *ARX* mutation.

1.4.2.2 A genotype/phenotype correlation

Non-malformation phenotypes are generally caused by *ARX* mutations that expand the first two polyalanine (pA) tracts or alter residues outside the homeodomain (Kato *et al.*, 2004)

(reviewed by (Gecz *et al.*, 2006; Shoubridge *et al.*, 2010a, Appendix A.1)), whereas brain and genital malformation are normally associated with mutations that truncate the ARX protein or alter critical conserved homeodomain residues (Kato *et al.*, 2004) (reviewed by (Gecz *et al.*, 2006; Shoubridge *et al.*, 2010a, Appendix A.1)). The severe phenotypes observed in affected males with presumed null mutations in *ARX* (*ie* XLAG) are consistent with phenotypes observed for the *Arx* knockout mouse (Kitamura *et al.*, 2002; Marsh *et al.*, 2009).

ARX mutations are the only identified cause of XLAG (MIM 300215), which may partly explain the high frequencies of *ARX* mutations reported in XLAG patient cohorts (Kitamura *et al.*, 2002; Kato *et al.*, 2004). To date, 30 families have been reported to harbour both brain and genital malformations due to 26 different mutations in *ARX* (reviewed in (Shoubridge *et al.*, 2010a, Appendix A.1)). Most cases are classified as XLAG with or without additional features. In 23 of 30 (74%) of cases ARX protein truncation mutations are presumed to cause the loss-of-function of the mature protein (Kitamura *et al.*, 2002; Uyanik *et al.*, 2003; Hahn *et al.*, 2004; Hartmann *et al.*, 2004; Kato *et al.*, 2004; Bhat *et al.*, 2005; Jagla *et al.*, 2008; Okazaki *et al.*, 2008; Miyata *et al.*, 2009).

As *ARX* mutations that expand the first two polyalanine (pA) tracts or alter residues outside the homeodomain (Kato *et al.*, 2004) usually result in non-malformation phenotypes (reviewed by (Gecz *et al.*, 2006; Shoubridge *et al.*, 2010a, Appendix A.1)) it is likely that these mutations partially impair ARX function. Presumably some ARX function is retained since patient phenotypes can be considered milder than that caused by null mutations (*ie* brain and genital malformations; *eg* XLAG).

1.4.2.3 Phenotypes of female *ARX* mutation carriers

Female carriers heterozygous of *ARX* mutations are frequently reported with comorbidities such as epilepsy, infantile spasms and schizophrenia. In one family, three of four obligate female carriers of the c.429_452dup mutation had anecdotal accounts of cognitive dysfunction (Turner *et al.*, 2002). These carrier females had learning problems at school and below normal IQ, consistent with other studies (Frints *et al.*, 2002b). Another study reported the phenotype of 18 carrier females heterozygous for the c.429_452dup mutation (Partington *et al.*, 2004). Of the 18 carrier females three had below normal IQ and two had schizophrenia. Further, female carriers on the c.1058C>T mutation (causes XMESID in males) displayed spasticity and hyperflexia (Scheffer *et al.*, 2002). In 50% of 14 families where XLAG was observed in affected males, one or more carrier females had an absent corpus callosum (Kato *et al.*, 2003), consistent with a further report (Uyanik *et al.*, 2003). Interestingly, a c.1465delG mutation was identified in *ARX* from a female with ISSX (Wallerstein *et al.*, 2008).

A knockout mouse model discovered that ~50% of female mice heterozygous for a null *ARX* allele developed seizures (Marsh *et al.*, 2009), which lead to a critical revision of clinical data for carrier females. 14 mothers of genotypically affected males all had normal development and cognitive skills, whereas 67% (6/9) of other female relatives (sister, aunt, cousin) displayed neurological symptoms (ID, learning problems, and agenesis of the corpus callosum, but not infantile spasms) (Marsh *et al.*, 2009). Further, no correlation was discovered between the affected status of the heterozygous female and preferential X-inactivation of the normal *ARX* allele in blood (Marsh *et al.*, 2009), consistent with previous findings (Stromme *et al.*, 2002b). These findings highlight the unexpected complexity of investigating X-linked modes of inheritance (Marsh *et al.*, 2009).

These findings suggest that females heterozygous for *ARX* mutations do present with phenotypes in many cases, *eg* upon review only 33% of females carrying mutations that cause malformation phenotypes in males had completely normal development (Marsh *et al.*, 2009). Part of the phenotypic variance observed in carrier females mirrors the variable clinical expressivity observed for males and does not correlate with X-inactivation status (Stromme *et al.*, 2002b; Marsh *et al.*, 2009).

If X-Inactivation is random and neurons expressing mutant alleles are equally viable, then 50% of cells express mutant *ARX* in carrier females. Presumably 50% of neurons in carrier females mimic the aberrant mechanisms occurring as in affected males, possibly causing cognitive and structural defects. These defects are more subtle in carrier females than affected males probably due to random X-inactivation, *ie* due to 50% less cells expressing mutant *ARX*. Variable clinical expressivity is also observed among carrier females, again suggesting that other as yet undetermined genetic and environmental modifiers are likely to have an impact on carrier female phenotypes.

1.4.3 Function of ARX

1.4.3.1 *ARX* expression

The first indications of *ARX* function was determined by its expression throughout development, although expression does not necessarily belie function. The majority of these studies were performed in mice, but some human information has been reported. The *ARX* gene is expressed in human brain (fetal, infant, and adult), heart, pancreas and skeletal muscle (Bienvenu *et al.*, 2002; Stromme *et al.*, 2002b). Expression of *ARX* mRNA in the human

brain can be determined from the tissues of origin of ESTs used to generate the cluster (UniGene cluster Hs.300304) that the *ARX* refseq (NM_139058) is derived from, and by northern blot (Bienvenu *et al.*, 2002; Stromme *et al.*, 2002b). *ARX* expression is found in specific regions such as the adult cerebral cortex, amygdala, thalamus, corpus callosum, caudate nucleus, substantia nigra, and hippocampus. *ARX* expression and the ID phenotypes caused by *ARX* mutations suggest it is vital for proper human brain function and development.

Expression of murine *Arx* displays a pattern restricted to the central nervous system (CNS) during development (Bienvenu *et al.*, 2002). *Arx* CNS expression begins at embryonic day 8 (E8) in a structure that develops into the forebrain. Transient *Arx* expression in the telencephalic vesicles and the floor plate occurs at E9 to E10. At E10.5 *Arx* expression was only found in the telencephalon, which develops into the cerebral cortex. After E10.5 until neonatal stages, strong expression is limited to developing telencephalon and ventral thalamus. Adult mouse brain has high levels of *Arx* expression in the olfactory bulb, and a lower level of expression in all brain structures. Finer mapping determined that *Arx* expression is present in the ventricular and intermediate zones of the developing cortex, and in the cortical plate. These regions contain proliferating, migrating, and differentiating interneurons (Bienvenu *et al.*, 2002). *Arx* mRNA positive interneurons co-express gamma-aminobutyric acid (GABA) (Bienvenu *et al.*, 2002).

This pattern of murine *Arx* expression suggests it has a conserved role in brain development and function in mouse, specifically the development of the forebrain and cerebral cortex and proliferating, migrating and differentiating interneurons.

1.4.3.2 Mouse models of ARX mutations

The *Arx* knockout mouse phenotype (Kitamura *et al.*, 2002) mimicked the clinical aspects of XLAG (Berry-Kravis and Israel, 1994; Dobyns *et al.*, 1999), which mutations in human *ARX* cause (Kitamura *et al.*, 2002). Male mutant male mice died early in the perinatal period, had aberrant migration and differentiation of GABAergic interneurons in the ganglionic eminence and neocortex, coupled with abnormalities of testicular development (Kitamura *et al.*, 2002). Given the perinatal lethality of an *Arx* knockout, further investigation of *Arx* function in the developing brain utilised conditional knockout mouse models.

Region specific knockdown of *Arx* in the mouse brain identified aberrant proliferation and radial migration of cortical progenitor resulted from loss of *Arx* function (Friocourt *et al.*, 2008). Silencing of *Arx* expression in neural progenitors caused a rounded morphology (instead of the normal bipolar), causing cells to accumulate in the subventricular/intermediate zone instead of migrating to the cortical plate (Friocourt *et al.*, 2008). This suggests that *Arx* is not fully associated with expression of GABA in cortical and striatal neurons, as originally thought, and it was concluded that ARX plays a role in the specification of distinct subsets of GABAergic interneurons in the subpallium (Friocourt *et al.*, 2008). These findings are consistent with further studies using mice carrying conditional *Arx* knockout in neurons derived from the ganglionic eminence, including cortical interneurons (Marsh *et al.*, 2009). Initial characterisation indicated that hemizygous males mirror several clinical features seen in patients with *ARX* mutations, specifically infantile spasms. It was demonstrated that an interneuron-specific loss of ARX protein occurred, implicating interneurons in the pathogenesis of epilepsy in patients with infantile spasms.

Studies examining pathogenic mechanisms of expanded polyalanine tracts in diseases involving the transcription factors FOXL2 (Caburet *et al.*, 2004), HOXD13, SOX3, RUNX2,

HOXA13 (Albrecht *et al.*, 2004), PHOX2B (Bachetti *et al.*, 2005), and ARX (Nasrallah *et al.*, 2004; Shoubridge *et al.*, 2007) indicate expansions of polyalanine tracts beyond a certain threshold induce misfolding and aberrant protein interactions, degradation, mis-localization and/or aggregation. Wildtype polyalanine tracts do not exceed 20 alanines in length in any polyalanine containing protein (reviewed in (Albrecht and Mundlos, 2005)).

Whether *in vivo* aggregation of ARX protein with expanded polyalanine tract 1 or 2 occurs or not in human tissue has not yet been reported. However, two mouse models of the c.304ins$(GCG)_7$ polyalanine tract expansion mutation have been generated independently (Kitamura *et al.*, 2009; Price *et al.*, 2009). Both mouse models phenotypically mimic ISSX in humans, suggesting a partial loss of function of mutant ARX protein, specifically in the CNS. A selective reduction of ARX positive GABAergic interneurons in the striatum of mutant mice was demonstrated. Interestingly mutant ARX protein was mislocalised to the cytoplasm throughout the cortex in 55% of interneurons when examined in adult mice (Price *et al.*, 2009), consistent with previous *in vitro* over-expression findings (Nasrallah *et al.*, 2004; Shoubridge *et al.*, 2007). In contrast, no formation of intranuclear inclusions or cytoplasmic accumulation occurred in ARX positive migratory cells in the ganglionic eminence and in the cortical cells were detected during fetal development (Kitamura *et al.*, 2009). Further, no aggregation of mutant ARX protein was observed when the c.429_452dup mutation was modelled (Kitamura *et al.*, 2009). Therefore, aggregate formation is not an obvious pathogenic mechanism explaining the loss of ARX function due to mutant expanded polyalanine tracts. Mechanisms such as aberrant protein localisation, interaction and degradation of the mutant protein remain possible (reviewed in (Shoubridge *et al.*, 2010a, Appendix A.1)).

An *Arx* knockout in the subpallium and ganglionic eminences of the mouse brain identified targets of ARX transcriptional repression (Colasante *et al.*, 2008; Fulp *et al.*, 2008). The

genes *Lmo1*, *Ebf3*, and *Shox2* were identified as direct targets of ARX mediated repression, with ARX found to bind a specific transcription factor binding site (TAATTA) (Fulp *et al.*, 2008). Ectopic expression of *Ebf3* in ventral telencephalon arrested tangential migration of neurons. Subsequent silencing of *Ebf3* expression in *Arx* mutant mice partially rescued neuronal migration (Colasante *et al.*, 2008). These ARX transcriptional targets and consensus binding site sequence were recently confirmed as part of 1006 novel ARX targets discovered in cell based and mouse brain studies (Quille *et al.*, 2011). Curiously the consensus ARX binding site was confirmed in over-expression transfection studies but could not be confirmed *in vivo* in mouse fetal brain (Quille *et al.*, 2011). Further, of the genes bound by ARX 24% had altered expression, the majority of which were repressed (Quille *et al.*, 2011). Some ARX target genes showed evidence of being both repressed and induced by ARX binding, highlighting the likely modification of ARX function by specific biological context and molecular environment (Quille *et al.*, 2011).

Mutations that alter the same p.P353 residue of the ARX homeodomain to two different amino acids leads to different phenotypic outcomes in humans (*ie* p.P353R – XLAG (Kato *et al.*, 2004); p.P353L – XMESID (Stromme *et al.*, 2002b)). 'Knock-in' mouse models were used to investigate different mutant residues at p.P353 (Kitamura *et al.*, 2009). Male mice with the p.P353R mutation died early in the post-natal period, had significantly different *Arx* mRNA and protein levels, reduced GABAergic and cholinergic neuronal development and disturbed brain morphology. These phenotypic features were similar to the *Arx* knockout mouse model, equivalent to null mutations in humans that lead to XLAG. Conversely, male mice with the p.P353L mutation lived for greater than 6 months, exhibited mild seizures and slight learning impairment. The p.P353L mice had normal GABAergic progenitor cell migration, similar to the wildtype mice, but displayed impairments in tangential migration of interneurons (Kitamura *et al.*, 2009).

There have been 16 reports of mutations that alter residues within the homeodomain of ARX (listed above in Table 1.4.1.3a), 10 of which cause XLAG and a further four that cause other brain/genital malformations. XLAG results from a presumably null *ARX* allele. This highlights that these residues are critical to the function of ARX, because when mutated, often altering only one residue in the homeodomain results in a phenotype akin to a null mutation. Homeodomain residues that display high levels of conservation across all homeodomain containing proteins often cause pathology when mutated (reviewed by (Shoubridge *et al.*, 2010a, Appendix A.1))

1.4.3.3 Role of ARX in pancreas?

Curiously, findings of the critical functions of *Arx* in cell fate decisions in pancreas of the mouse with endocrine pancreatic insufficiency contradicts the lack of clinical outcomes associated reported for humans with *ARX* mutations.

A mouse *Arx* loss-of-function model had early onset hypoglycemia, dehydration and weakness, and early post-natal death (Collombat *et al.*, 2003). A critical role for *Arx* expression in α-cell, β-cell and δ-cell fates was discovered (Collombat *et al.*, 2003). Subsequently the underlying mechanism was found to be reciprocal transcriptional repression of *Arx* and *Pax4*. This provided the necessary regulatory balance that mediated the determination of pancreatic endocrine cell fate (Collombat *et al.*, 2005). Further studies, using a gain-of-function approach indicated that *Arx* expression was necessary and sufficient to promote an α-cell fate (Collombat *et al.*, 2007).

In the pancreas, *Pax4* and *Arx* are downstream targets of NEUROG3, which is coded for by the Neurogenin 3 gene (*Neurog3*) (Collombat *et al.*, 2003). *Neurog3*, is thought of as a

'master gene' that controls endocrine cell fate decision in the pancreatic endodermal progenitor cells (Gradwohl *et al.*, 2000; Schwitzgebel *et al.*, 2000) and primitive gastrointestinal tract (Jenny *et al.*, 2002). *Pax4* has also been implicated in the development of endocrine progenitors in the digestive tract (Wang *et al.*, 2008). The involvement of *Arx* in endocrine cells in the gastrointestinal tract remains to be determined, although a common feature reported as comorbid with XLAG from birth is intractable diarrhoea, which may implicate a role for *ARX* in the gastrointestinal tract (Hahn *et al.*, 2004; Hartmann *et al.*, 2004; Kato *et al.*, 2004; Kato and Dobyns, 2005; Jagla *et al.*, 2008; Okazaki *et al.*, 2008; Miyata *et al.*, 2009). Another study has examined the interaction of GROUCHO/TLE transcriptional co-repressors with transcription factors important in the endocrine pancreas (Hoffman *et al.*, 2008). GROUCHO/TLE co-repressors may act upon many transcription factors involved in endocrine pancreas development, including *Neurog3*. TLE2 was shown to interact with ARX and in β-cells this co-repressor modulates the repressive capacity of *Arx* transcriptional activity (Hoffman *et al.*, 2008), similar to ARX/TLE1 co-repression in humans (McKenzie *et al.*, 2007).

A small number of studies report XLAG patients with severe chronic hypoglycaemia, often within the first few weeks of life (Hahn *et al.*, 2004; Hartmann *et al.*, 2004) or rare occurrences of acute hypoglycaemia (Kato and Dobyns, 2005). But hypoglycaemia is rarely described as a phenotype of patients with *ARX* mutations, including XLAG patients. It remains possible that the glucose homeostasis regulating mechanisms in human are more robust than inbred laboratory mice, especially C57Bl6 mice which are prone to metabolic disorders. Despite human/mouse disparities, the function and regulation of *Arx* in the pancreas may also provide valuable insights into the molecular pathology of *ARX* mutations the brain.

1.4.4 Ultraconserved elements around the *ARX* locus

The *ARX* gene is flanked by 11 ultraconserved (uc) elements, some of which are the largest in the human genome (Table 1.4.4a; uc460-470); 10 are 3' of *ARX* (uc460-469); and 8 of these are within the 3' introns of *POLA1* (uc460-467) (Bejerano *et al.*, 2004). Some of these uc elements have been shown to act as enhancers of *ARX* expression (uc467-469) (Colasante *et al.*, 2008); however, no obvious phenotype resulted when these elements were deleted in mice (Ahituv *et al.*, 2007). Interestingly a sequence variation in uc221 on chromosome 7 (alternatively referred to as I56i) has been detected in an individual with autism (Poitras *et al.*, 2010).

Table 1.4.4a. The ultraconserved (uc) elements flanking the human *ARX* gene. Positions given are from UCSC genome browser (Feb 2009 assembly GRCh37/hg19).

uc	bp	5' start	3' end	bp from *ARX*	5'/3' (*ARX*)	Intronic (*POLA1*)
460	275	24,823,511	24,823,785	198,030	3'	√
461	397	24,864,797	24,865,193	156,622	3'	√
462	779	24,894,826	24,895,604	126,211	3'	√
463	275	24,915,882	24,916,156	105,659	3'	√
464	770	24,916,158	24,916,927	104,888	3'	√
465	310	24,917,481	24,917,790	104,025	3'	√
466	349	24,946,458	24,946,806	75,009	3'	√
467	731	25,008,354	25,009,084	12,731	3'	√
468	489	25,017,563	25,018,051	3,764	3'	x
469	222	25,018,053	25,018,274	3,541	3'	x
470	341	25,401,216	25,401,556	367,151	5'	x

Gene positions: *ARX* (NM_139058) 25,021,815 to 25,034,065
POLA1 (NM_016937) 24,712,056 to 25,015,100

1.5 RESEARCH AIMS

The three main aims of this project were:

i. To screen the entire *ARX* ORF and some non-coding 'ultraconserved' enhancer regions for sequence variations in a moderately sized cohort of patients.

ii. To investigate the effect of potential molecular mechanisms of novel sequence variations have on ARX function.

iii. To determine the effect of *ARX* variants on protein localisation in cell based studies.

CHAPTER 2 CONTENTS

Molecular protocols are based on those described in 'Molecular Cloning: A Laboratory Manual (*Third Edition*) (Sambrook and Russell, 2001).

2 SUBJECTS, MATERIALS AND METHODS

2.1 Patient cohort

This study was approved by the CYWHS research ethics committee. Clinical information and DNA or blood samples were collected with informed consent.

Following publication from our laboratory detailing among the first *ARX* mutations (Scheffer *et al.*, 2002; Stromme *et al.*, 2002b; Partington *et al.*, 2004), many clinical collaborators individually assessed patients and sent a moderate number of blood and DNA samples to our laboratory for *ARX* mutation analysis. In the cases where a mutation was discovered and further family investigations were required, the phenotypic details and specific clinical collaborators are acknowledged accordingly within the following chapters.

A heterogeneous cohort of DNA samples from 613 individuals were screened. All had been diagnosed with intellectual disability (ID) (Table 2.1a). They were initially screened for the recurrent mutations in *ARX* (c.429_452dup 'dup24bp' of pA2 and c.304ins$(GCG)_7$ 'expansion' of pA1). 500/613 samples without a pA1 or pA2 mutation and of sufficient DNA quantity and quality had the entire *ARX* ORF screened for mutations by SSCP (single stranded conformational polymorphism (Orita *et al.*, 1989a; Orita *et al.*, 1989b)) and/or dHPLC analysis (denaturing high pressure liquid chromatography (Oefner, 1995)) (reviewed by Xiao and Oefner, 2001]) (Table 2.1a). A further subset of 94/500, which were either one of multiple male siblings haplotype matched to Xp22 and/or had a milder phenotype (*eg* autism spectrum disorder, nsXLID) that may represent more subtle variation in *ARX*

expression but did not have an *ARX* ORF mutation, were also selected for variation testing of 'ultra-conserved' (uc) element sequence (Table 2.1a). The four ultraconserved elements in closest proximity to the *ARX* locus (3' of the gene) were selected for screening (ie uc466, 467, 468 and 469). Variations in uc element sequences have been associated with human disease such as multiple sclerosis (Ban *et al.*, 2005), familial breast cancer (Yang *et al.*, 2008) and autism (Poitras *et al.*, 2010) since they were first characterised (Bejerano *et al.*, 2004).

DNA was extracted from blood samples using a DNA blood midi kit (Qiagen). To ensure that no chaotropic salt contamination of the dHPLC column occurred, all DNA samples were ethanol precipitated prior to PCR.

Table 2.1a. Subsets of patients within the heterogeneous cohort screened for mutations in the *ARX* ORF and ultraconserved (uc) elements.

Sub cohort	Patient (#)
FraX A/E neg	147
Possibly X-linked	123 (14)[1]
'*ARX* testing' referrals	117 (24)[1] [+113][2]
Autism spectrum disorder	68 (55)[1,3]
Epilepsy/ISSX	37 (1)[1]
Brain malformations	8
Total	500 [613]

[1] In rounded brackets is the number of patients from this sub group within the cohort screened for uc element sequence variation.
[2] In squared brackets is the 113 extra patients that were only screened for pA2 dup/pA1 expansion mutations by PCR and agarose gel electrophoresis within the '*ARX* testing' referrals cohort.
[3] 55 of these families had multiple affected males Xp22 haplotype matched, sourced from the autism genetic research exchange (AGRE).

2.2 Tables of reagents

Table 2.2a. Details of reagents used in PCR and DNA electrophoresis protocols (sections 2.3-2.6).

Reagent	Conc/Amt	Catalogue #	Supplier	Location
[α-^{32}P]-dCTP (10mCi/ml)	1.0 mCi	BLU008H001MC	Perkin Elmer	Melbourne, Australia
Acrylamide/Bis-acrylamide Solution (37.5:1)	500 ml	161-0148	Bio-Rad	Sydney, Australia
Agarose, Molecular Biology Grade	500 g	15003-500	Mo Bio	Carlsbad, USA
Ammonium Persulfate (APS)	10 g	161-0700	Bio-Rad	Sydney, Australia
BigDye Terminator v3.1 Cycle Sequencing Kit	25,000 rxns	4337458	Applied Biosystems	Melbourne, Australia
dATP, dCTP, dGTP, dTTP (as a pack)	25 µmol	AB0315/A	Thermo Fisher Scientific	Melborune, Australia
dimethyl sulphoxide (DMSO)	100%	D9170	Sigma	St Louis, USA
DNA Load Buffer (Fermentas)	6.0 x	R0611	Quantum Scientific	Sydney, Australia
Ethidium bromide solution (10 ml)	10 mg/ml	E1510-10ML	Sigma	St Louis, USA
Expand DNA Polymerase (enzyme mix)	3.75 U/µl	117590600013	Roche	Sydney, Australia
FailSafe PCR premix J	2.0 x	FSP995J	Epicentre Biotechnologies	Madison, USA
milliQ (ddH_2O from a Millipore milliQ system)	-	-	Millipore	Billerica, USA
Oligos	40 nmol	(custom)	Geneworks	Adelaide, Australia
Power SYBR Green PCR Master Mix	2.0 x	4368708	Applied Biosystems	Melbourne, Australia
QIAamp DNA Blood Midi Kit	100 preps	51185	Qiagen	Melborune, Australia
RNaseP MGB taqman assay (VIC)	20.0 x	4316844	Applied Biosystems	Melbourne, Australia
Taq DNA Polymerase (10 x PCR buffer supplied)	5.0 U/µl	115965940012	Roche	Sydney, Australia
TaqMan Gene Expression Master Mix	2.0 x	4369542	Applied Biosystems	Melbourne, Australia
TEMED	5.0 ml	161-0800	Bio-Rad	Sydney, Australia

Table 2.2b. Details of reagents used for plasmid preparations, cloning and bacterial culture (sections 2.7, 2.8).

Reagent	Conc/Amt	Catalogue #	Supplier	Location
Ampicillin (powder)	5.0 g	A9393-5G	Sigma	St Louis, USA
Bacteriological agar	250 g	A5306-250G	Sigma	St Louis, USA
*Eco*NI restriction enzyme (supplied with 10 x buffer)	1,000 U	R0521S	New England Biolabs	Ipswich, USA
LB broth (powder)	1.0 kg	L3022-1KG	Sigma	St Louis, USA
*Nae*I restriction enzyme (supplied with 10 x buffer)	500 U	R0190S	New England Biolabs	Ipswich, USA
pCMV-myc vector	-	631604	Clontech	Mountain View, USA
pGEM-T Easy Vector System I	20 rxns	A1360	Promega	Sydney, Australia
Qiagen Plasmid Midi Kit	100 preps	12145	Qiagen	Melborune, Australia
QuikChange II XL Site-Directed Mutagenesis Kit	30 rxns	200522	Agilent	Santa Clara, USA
T4 DNA Ligase	100 U	M1801	Promega	Sydney, Australia
Wizard *Plus* Minipreps DNA Purification System	250 preps	A7510	Promega	Sydney, Australia
XL10-Gold Ultra competent cells	10 x 0.1 mL	200315	Agilent	Santa Clara, USA

Table 2.2c. Details of reagents used for mammalian cell based investigations and western immunoblotting (sections 2.9-1.12).

Reagent	Conc/Amt	Catalogue #	Supplier	Location
α-cMYC 9E10 (mouse monoclonal antibody)	200 µg/ml	sc-40	Santa Cruz Biotech	Santa Cruz, USA
α-mouse IgG-Cy3 (goat antibody)	-	115-166-062	Jackoson Laboratories	Bar Harbour, USA
α-mouse IgG-FITC (goat polyclonal antibody)	2.0 ml	F0479	Dako	Melborune, Australia
α-mouse IgG-HRP (goat polyclonal antibody)	2.0 ml	P0447	Dako	Melborune, Australia
α-Sheep IgG-HRP (Donkey)	2.0 ml	AB324P	Millipore	Billerica, USA
Dulbecco's modified Eagle's medium (DMEM)	10x 500 ml	10313-039	Invitrogen	Melborune, Australia
ECL Reagents (A&B; Amersham)	250 ml each	RPN2106	GE Healthcare	Uppsala, Sweden
Fetal Calf Serum (FCS)	500 ml	16140071	Invitrogen	Melborune, Australia
Formaldehyde (solution)	500 ml	F8775	Sigma	St Louis, USA
HEK293T cells (human embryonic kidney)	-	CRL-11268	ATCC	Manassas, USA
Lipofectamine 2000	1.5 ml	11668019	Invitrogen	Melborune, Australia
Na_3VO_4	10 g	S6508-10G	Sigma	St Louis, USA
NaCl	1.0 kg	S1679-1KG	Sigma	St Louis, USA
NaF	10 g	S1929	Sigma	St Louis, USA
NuPAGE Novex 4-12% Bis-Tris Gel (12 well, 1mm)	10x gels	NP0322BOX	Invitrogen	Melborune, Australia
Penicillin-Streptomycin (liquid)	20x 100 ml	15140163	Invitrogen	Melborune, Australia
PMSF (Phenylmethanesulfonyl fluoride)	50 ml	93482-50ML-F	Sigma	St Louis, USA
Protease Inhibitor Cocktail	1.0 ml	P8340-1ML	Sigma	St Louis, USA
Quick Start Bradford Protein Assay Kit 1	4,000 assays	500-0201	Bio-Rad	Sydney, Australia
Sodium dodecyl sulphate (SDS, powder)	500 g	L4390-500G	Sigma	St Louis, USA
Tris Base (Tris(hydroxymethyl)aminomethane)	1.0 kg	T87602-1KG	Sigma	St Louis, USA
Triton-X 100	1.0 l	X100-1L	Sigma	St Louis, USA
Trypan Blue Stain	100 ml	15250061	Invitrogen	Melborune, Australia
TWEEN-20	1.0 l	P1379-1L	Sigma	St Louis, USA
VectaShield	10 ml	H-1200	Vector Laboratories	Burlingame, USA

2.3 Tables of primer pairs

All primers were supplied by the custom oligo service of Geneworks (Adelaide, Australia).

Table 2.3a. PCR/Agarose gel electrophoresis screening and mutant polyalanine (pA) tract sub-cloning primers.

Primer	Exon	Sequence F primer	Sequence R primer	Annealing Tm (°C)	Length (wt - bp)	Conc. (final µM)
2J-1 [1]	2	CCCCGGCCGGCCGGCTGGCTGATA	GCGGCCCCTGCGCCGTCCGGCCGTTC	60	281	1.0
2J-2 [2]	2	CCCCTCCGCCGCCACCGCCAAC	TCCTCCTCGTCGTCCTCGGTGCCGGT	60	313	1.0

Primers were designed to either specifically detect variations in pA tract 1[1] or pA tract 2[2] length by PCR and agarose gel electrophoresis.

Table 2.3b. Hexachlorofluorescein (hex) labelled SSCP primers.

Primer	Exon	Sequence F primer	Sequence R primer	Annealing Tm (°C)	Length (wt - bp)	Conc. (final µM)
1-1 Ex1-1	1	GCTCACTACACTTGTTACCGC	TTACTTTTGCACTCGGGCCTC	60	288	1.5
1-2 Ex1-2	1	CAGCCATGAGCAATCAGTACC	AATTGACAATTCCAGGCCACTG	60	293	1.5
2-1 Ex2-1	2	CAGCAGCCCTGGCTGGGACT	CGTTCTCGCGGTACGACTTGC	60	400	1.0
2-1 Ex2-1 [1]	2	CCCCTCCGCCGCCACCGCCAAC	TCCTCCTCGTCGTCCTCGGTGCCGGT	60	313	1.0
2-2 Ex2-2	2	GCAGGTGAGCATCAGCCGCA	CAGCTCCTCCTTGGGTGACA	64	390	1.0
2-3 Ex2-3	2	AACTGCTGGAGGACGACGAGG	CGCGACCACCCTACGCGCAT	64	370	1.0
3 Ex3	3	GAAATAGCTGAGAGGGCATTGC	TCTCTTGGTTTTGTGAAGGGGAT	60	231	5.0
4-1 Ex4-1	4	GACGCGTCCGAAAACAACCTGAG	CAGTCCAAGCGGAGTCGAGCG	60	336	1.5
4-1 Ex4-2	4	ACCCACCCGCTCAGCCCCTA	CCCCAGCCTCTGTGTGTATG	60	291	1.5
5 Ex5	5	ACAGCTCCCGAGGCCATGAC	GAGTGGTGCTGAGTGAGGTGA	60	347	1.5

[1] Designed for the specific detection of c.429_452dup(24bp) mutation.

Table 2.3c. dHPLC primers.

Primer	Exon or uc	Sequence F primer	Sequence R primer	Annealing Tm (°C)	Length (wt - bp)	Conc. (final µM)
1-1 Ex1	1	GTCCACTACACTTGTTACCGC	TTACTTTTGCACTCGGGCCTC	60.0	288	0.5
1-2 Ex1/O	1	CAGCCATGAGCAATCAGTACC	AATTGACAATTCCAGGCCACTG	60.0	293	0.5
2-1 O/Ex2d	2	CAGCAGCCCTGGCTGGGACT	CGTTCTCGCGGTACGACTTGC	66.0	402	0.5
2-2 Ex2d	2	TGGGACACGCTCAAGATCAG	CAGCTCCTCCTTGGGTGACA	64.0	421	0.5
2-3 E2P2/O	2	AACTGCTGGAGGACGACGAGG	TGCGCTCTCTGCCGCTGCGA	67.5	392	0.5
3 O	3	GAAATAGCTGAGAGGGCATTGC	TCTCTTGGTTTTGTGAAGGGGAT	54.0	231	0.5
4-1 O/C	4	GACGCGTCCGAAAACAACCTGAG	CAGTCCAAGCGGAGTCGAGCG	60.0	336	2.5
4-2 Ex4/O	4	ACCCACCCGCTCAGCCCCTA	CCCCAGCCTCTGTGTGTATG	60.0	291	2.5
5 O	5	ACAGCTCCCGAGGCCATGGC	GAGTGGTGCTGAGTGAGGTGA	60.0	347	0.5
uc466	uc466	TTCATGTGGGTTGTCACTCG	GATTCAGCAGTAATTGCTGG	55.0	450	2.5
uc467-1	uc467	ACCAGCAATTTCAAACACAG	GCATGATGAAGTTAATGGAG	58.0	508	2.5
uc467-2	uc467	TAACAGCTATTCCAGCTGCC	ACGCTGTAATTCCAAGAACC	58.0	481	2.5
uc468	uc468	TGAACAGTTTGTGGTGGAGC	TCTCTGCACAAAACCTCTAG	55.0	581	2.5
uc469	uc469	GCAACCAGCCTTCAATTCAC	GATATGGGAGTGATTACCGG	60.0	373	2.5

Table 2.3d. Primers used for haplotype analysis of the region flanking the *ARX* locus (Xp22.11;Xp21.3 or chrX:24,419,207;27,645,778).

Marker	Chromosomal position	Sequence F primer	Sequence R primer	Length (bp)	NCBI UniSTS
↑(Telomere)					
DXS 8099	Xp22.11	AGCTGGTTTTGTGATTCTGC	GGCCTTGAGATGTAGCCA	105-119	14083
DXS 8027	Xp22.11	GTGAGACGCTGTCTTGG	AGCTGCTGTACTAATAACATAGG	232-242	57789
ARX	Xp21.3	chrX:25,021,815 (exon 5)	chrX:25,034,065 (exon 1)		
DXS 1202	Xp21.3	CCCAGCACAGGTGTCA	GGGTTGTTGCTCCAAAG	265-285	31334
DXS 8047	Xp21.3	GTGCCCGGCCCATATTA	CCAAAGTTTTGTTGCATCCC	130-146	58062
↓(Centromere)					

Markers are ordered down the table from the lowest nt position first and position of *ARX* as per UCSC genome build GRCh37/hg19 (Feb 2009)
All primers were used at a final concentration of 0.5µM and an annealing temperature of 60°C
All variation in marker size is due to a varying number of CA dinucleotide repeats.

Table 2.3e. Primers used for copy number variation qPCR.

Primer	Exon	Sequence F primer	Sequence R primer	Length (wt - bp)
POLA 35	35	ATTCTGACAAGTCCCTGTAC	GATCGGTAGTAAGTTTCTCC	87
371	-	TAGGGGAAATTATGAGTGGG	AGGGTGAGCATATTCTATCC	166
372	-	CTTAGGGGCACACTGTTTGG	AGATTCCTTAGCTTCTCAGG	189
373	-	AGTCCTCCAATGTTTGTGCC	ATGACAAAGGCCAAGATAGG	115
374	-	TCTTCCTGGGAGCCACTGAC	GGGTACCCAACTAGGAAAAC	106
375	-	ACTTCTACAGTGGACACAAG	GGGTTGATTCCAAGTCTTTG	89
376	-	TGTGCCAGTGCTGACGGAAT	GCCATTTTCCTCGTGTCTAG	100
377	-	AGGCTGGTCTTGAACTCCTG	GCCTCTGTTCCATATCAATATAGC	146
378	-	TGCTGCTCTAGAACTGTGAG	TGTATTCTGAGGCCTCCATC	122
POLA 36	36	AAAGTTCTGCAGGACTACAG	TAGGATTTCACGGCACAACC	116
ARX 5	5	GCTCCTGAGACAGCCCACAC	TTAGCACACCTCCTTGCCCG	193
ARX 4F 2/5R	4/5	CATCAGCCCGGCATTCGGCA	TTAGCACACCTCCTTGCCCG	262
ARX 4	4	CGCTCGACTCCGCTTGGACT	TGCCGAATGCCGGGCTGATG	174
ARX 3	3	ATTGCTGGGGCCTGCAGTGA	GTTGTGCAGCTCACCTGGAC	109
ARX 2	2	AACTGCTGGAGGACGACGAG	CAGCTCCTCCTTGGGTGACA	161
ARX 1	1	AGCCATGAGCAATCAGTACC	GGATGCTGTCGATGCAGTAG	102
876	-	TATCCTCTTCCTGCAGCTCC	CCGAGTGAGTCTGAACGTGA	170
877	-	AAGGGACAGGCGAGTATCAG	CTTAGATTTGGCTGCAGCCG	227

All primers were used with the standard qPCR thermocycle (section 2.4) with a 60.0°C annealing temperature at a final concentration of 1.0µM. Primers pairs ordered are in order from the lowest nt number as per UCSC genome build GRCh37/hg19 (Feb 2009)

Table 2.3f. Mutagenic primers used for site directed mutagenesis.

Primer	Exon	Sequence of primer	Conc. (final µM)	Reference
c.81 C>G (p.27 Y>X)	1	CAACTTTGCTCTCCTCCTA*g*TGCATCGACAGCATCCTGGG	4.0	Fullston *et al.*, 2010
c.1058 C>T (p.353P>L)	2,3	TTCCAGAAGACGCACTACC*t*GGACGTCTTCACCAGGGAGG	4.0	Strømme *et al.*, 2002b
c.1074 G>T (p.358 R>S)	2,3	GGACGTCTTCACCAG*t*GAGGAACTGGCCATGAGGCTGGAC	4.0	Fullston *et al.*, 2011
c.1135 C>A (p.379R>S)	4	GTCTGGTTCCAGAAC*a*GTCGGGCCAAGTGGCGCAAGCGGG	4.0	Shoubridge *et al.*, 2010b
c.1136 G>T (p.379R>L)	4	GTCTGGTTCCAGAACC*t*TCGGGCCAAGTGGCGCAAGCGGG	4.0	Shoubridge *et al.*, 2010b

n Denotes the nucleotide targeted for mutation by each primer.

Table 2.3g. Primers used for sequence confirmation of mutant *pCMV-myc-ARX* ORFs generated by site directed mutagenesis or sub-cloning.

Primer	Exon	Sequence of primer	Conc. (final µM)
ARX cR1	4	CAGTCCAAGCGGAGTCGAGCG	0.16
pCMV-myc F	-	GATCCGGTACTAGAGGAACTGAAAAAC	0.16
pCMV-myc R	-	GTTGTGGTTTGTCCAAACTCATCAATG	0.16

2.4 PCR formulae and thermocycle profiles

Details of the reagents used in the following PCRs can be found in Table 2.2a.

2.4.1 Standard PCR

Final concentrations per standard 50 µl PCR (Mullis, 1990) were as follows:

Genomic DNA	50-100 ng
PCR buffer (10 x)	1.0 x
dNTPs (10 mM)	0.2 mM
Forward primer (50 µM)	1.0 µM
Reverse primer (50 µM)	1.0 µM
Taq DNA Polymerase (5 U/µl)	1.0 U (total)
ddH2O (milliQ)	make up to 50 µl

The thermocycle profile used for standard PCR was as follows:

Initial denaturation		94 °C – 2 min
35x	Denaturation	94 °C – 30 sec
	Annealing	60 °C – 30 sec
	Extension	72 °C – 30 sec
Final extension		72 °C – 5 min

Note that annealing temperatures for specific primer pairs that deviate from the standard thermocycle profile below appear in the primer Tables 2.1a-c.

2.4.2 GC rich PCR (*ARX*)

Final concentrations per 50 µl PCR designed to amplify GC rich regions (Baskaran *et al.*, 1996) were as follows:

Genomic DNA	50-100 ng
FailSafe Buffer J (2.0 x)	1.0 x
Forward primer (50 µM)	1.0 µM*‡
Reverse primer (50 µM)	1.0 µM*
Expand DNA Polymerase (3.75 U/µl)	1.0 U (total)
ddH2O (milliQ)	make up to 50 µl*†‡

* dHPLC PCRs were of a volume of 20 µl using the above concentrations (except primers, as per Table 2.3c). Prior to being analysed by dHPLC all PCRs were diluted ½ with ddH_2O.

† Haplotype PCRs were performed using the above concentrations but in a 25 µl reaction volume and with the addition of 0.5 µl [α-^{32}P]-dCTP (5 µCi total, Perkin Elmer) per reaction.

‡ Fluorescent fragment analysis PCRs were also performed using the above concentrations in a 20 µl reaction but with one hex labelled primer (Table 2.3b).

The thermocycle profile used for GC rich PCR was as follows:

Initial denaturation		94 °C – 2 min
35x	Denaturation	94 °C – 30 sec
	Annealing	60 °C – 30 sec
	Extension	68 °C – 2 min
Final extension		68 °C – 10 min

Note that primer pairs with annealing temperatures and concentrations that deviate from the above generic conditions are listed in Tables 2.1b and c.

2.4.3 qPCR

Final concentrations per standard 20 µl SYBR green based qPCR (Woo *et al.*, 1998) were as follows:

Genomic DNA/cDNA	25 ng/ 2 µl
Power SYBR green PCR master mix (2.0 x)	1.0 x
Forward primer (50 µM)	1.0 µM
Reverse primer (50 µM)	1.0 µM
ddH2O (milliQ)	make up to 20 µl

Final concentrations per standard 20 µl *RNaseP* Taqman control qPCR were as follows:

Genomic DNA	25 ng
Taqman Gene expression master mix (2.0 x)	1.0x
RNaseP MGB taqman assay control (20 x)	0.75x
ddH2O (milliQ)	make up to 20 µl

The thermocycle profile used for qPCR was as follows:

5’ endonuclease incubation		50 °C – 5 min*
Initial denaturation		95 °C – 10 min†
40x	Denaturation	95 °C – 15 sec
	Annealing/Extension	60 °C – 30 sec
Melt curve		60 °C – 95 °C‡

* This step is required when using Taqman probes.

† Hot start step required to activate Taq in master mixes.
‡ Ramp rate is 1°C/min

2.4.4 Sanger sequencing reaction

Final concentrations per standard 20 µl Sanger sequencing reaction (Sanger *et al.*, 1977; Tracy and Mulcahy, 1991) were as follows:

Template DNA (PCR product/plasmid)	50-100 ng/500-1000 ng
Sequencing buffer (5.0 x)	0.75 x
Forward or Reverse primer (3.2 µM)	0.16 µM
Big Dye Terminator (v3.1)	1.0 µl
ddH2O (milliQ)	make up to 20 µl*

*If the template targeted for sequencing contained a high GC (*ie ARX*) content then 100% dimethyl sulphoxide (DMSO) was added to a final concentration of 5%.

The thermocycle profile used for sequencing was as follows:

Initial denaturation		96 °C – 1 min
25x	Denaturation	96 °C – 10 sec
	Annealing	50 °C – 5 sec
	Extension	60 °C – 4 min

2.4.5 *In vitro* mutagenesis PCR

This technique is used to introduce specific mutations via mutagenic primers (Table 2.3f) and was first developed using single stranded DNA inputs (Kunkel, 1985), and has been adapted to apply to double stranded templates. Final concentrations per standard 25 µl *in vitro* mutagenesis PCR were as follows (all reagents except for template DNA were sourced from the QuikChange Site-Directed Mutagenesis Kit, Stratagene/Agilent):

Template plasmid (*pCMVmycARXwt*) DNA	100 ng
Quick change multi-buffer (10 x)	1.0 x
Quick solution	0.75 µl
Mutagenic primer (50 µM)	4.0 µM
dNTPs	1.0 µl
PfuUltra DNA polymerase (2.5 U/µl)	1.25 U (total)
ddH2O (milliQ)	make up to 25 µl

The thermocycle profile used for sequencing was as follows:

Initial denaturation:		95 °C – 1 min
35x	Denaturation:	95 °C – 1 min
	Annealing:	55 °C – 1 min
	Extension:	65 °C – 12 min*

*>2 min extension per kb of plasmid.

2.4.6 PCR machines used

Eppendorf Mastercycler gradient PCR machines (Eppendorf, Sydney, Australia) were used for all standard, GC rich, sequencing and *in vitro* mutagenesis PCRs with single reactions in either 0.6 ml tubes or 0.2 ml wells of a 96 well plate.

All CNV and *ARX/POLA* expression studies (chapter 8) were performed on an Applied Biosystems StepOnePlus^{TM} real time PCR system and data analysed using the StepOne Software (v2.0) both from Applied Biosystems (Melbourne, Australia).

2.5 Gel electrophoresis

Reagents used for this method are listed in Table 2.2a.

A typical agarose gel consisted of 2% agarose in 1.0x TBE solution with 250 ng/ml (w/v) ethidium bromide (Sigma) and electrophoresis was performed using a Wide Mini-Sub Cell GT or Mini-Sub Cell GT gel tank (Bio-Rad) at 110V. Normally 5 µl of a PCR product was loaded per well with 2 µl of 6.0x DNA load buffer (Quantum Scientific) and visualised under UV light.

Haplotype analysis was performed using a vertical 6% polyacrylamide gel in a vertical Sequi-Gen GT cell get tank (Bio-Rad), the 25 µl radio labelled ([α-^{32}P]-dCTP) PCR product was added to 30 µl formamide loading dye and 5 µl was loaded per sample. Gels were used to

expose X-ray sensitive film and haplotypes were assigned arbitrarily from 1 for the smallest sized STS marker band observed within the all of the samples tested.

Fluorescent fragment analysis and sequencing analysis was performed using a resin based capillary as part of an ABI 3100 sequencer (Applied Biosystems).

2.6 SSCP/dHPLC screening for *ARX* mutations

2.6.1 SSCP mutation screening.

This screening was kindly performed in collaboration with research assistants Bree Hodgson and Merran Finnis.

A total of 314/500 samples were subjected to SSCP analysis of the *ARX* ORF. SSCP analysis was performed using a vertical 7.5% polacrylamide gel system as previously described PCR conditions (Stromme *et al.*, 2002b) using hexachlorofluorescein labelled primers (Table 2.3b) to generate 10 amplicons across the five coding exons of *ARX* including 30-50 bp flanking intronic sequence. This protocol detects approximately 84% of mutations (Jordanova *et al.*, 1997).

2.6.2 dHPLC screening

Overall 312/500 samples (126 previously SSCP screened) were screened using a DNAsep cartridge column as part of a Transgenomic WAVE dHPLC system (System Model 3500, Transgenomic, Santa Clara, USA) using a standard buffer gradient, with the exception of time shifts (Table 2.6a). A total of 9 amplicons were designed to cover all five coding exons of *ARX*, including 30-50 bp of intronic sequence, using established PCR conditions (Stromme *et al.*, 2002b) (see Chapter 2.4) using different primers and annealing temperatures (Table 2.3c). These fragments required a total of 21 different dHPLC denaturing temperatures (Table 2.6a). A typical work flow included PCR amplification of patient's DNA samples (20

µl in 96-well plate format, PCR as per section 2.4) and 2% agarose gel electrophoresis to confirm successful amplification of expected length amplicon. Subsequently all PCRs were diluted ½ with ddH_2O, to reduce Betaine concentration of buffer J (Epicentre Biotechnologies) below tolerance levels of the DNAsep cartridge, then an equivalent amount of wildtype control PCR product was added and heteroduplexed with patient PCR products followed by dHPLC analysis. This protocol has a detection rate of approximately 96– 100% of mutations (Xiao and Oefner, 2001).

In addition 4 ultraconserved elements 3' of *ARX* (uc466-469; (Bejerano *et al.*, 2004)) were screened for sequence variation in 94/500 samples (Table 2.1a), using 5 amplicons (Table 2.3c) screened at 11 dHPLC denaturing temperatures (Table 2.6a). Prior to screening, 20 control X-chromosomes were sequence confirmed not to have variations in uc sequences. Also of note is that no variations within these uc sequences have been reported to date.

Table 2.6a. dHPLC denaturation ('oven') temperatures and time-shifts used.

Amplicon	Temperature (°C)	Time shift (minutes)
1-1	63.5	
1-2	62.5, 65.0, 67.5	(67.5 °C) +1.0
2-1	66.5, 68.5, 72.0	(72.0 °C) +1.0
2-2	66.0, 70.0	(70.0 °C) +1.0
2-3	65.0, 66.0, 68.5	
3	63.0, 64.5	(64.5 °C) +1.0
4-1	67.1	
4-2	65.5, 68.5, 70.0	(70.0 °C) +1.0
5	64.5, 67.5, 69.0	
uc466	55.5, 58.3	
uc467-1	54.0, 55.5, 56.0, 59.0	(59.0 °C) +1.5
uc467-2	54.0, 55.0, 59.0, 61.0	(59.0 °C) +1.0, (61.0°C) +1.5
uc468	56.7, 58.0	
uc469	56.7	

Temperatures used for denaturation during dHPLC screening were determined using the Transgenomic Wavemaker software package (Transgenomic, Santa Clara, USA) to ensure all

coding (or conserved) base pairs, and some intronic (or non-conserved) sequence, were most likely analysed for single nucleotide variations.

2.7 *In vitro* mutagenesis of the *ARX* ORF

Reagents used in this section are listed in Table 2.2b

Point mutations discovered and/or previously reported (Table 2.3f) were introduced to the *ARX* ORF via site directed mutagenesis. All reagents used for this procedure were supplied as components of an Agilent (ex Stratagene) QuikChange II XL Site-Directed Mutagenesis Kit and were performed as per manufacturer's instructions with the exception of the final concentration of mutagenic primer (4 µM used instead of recommended 5 ng/µl). Sequences of the mutagenic primers used can be found in Table 2.1f. The template plasmid used was the *pCMV-Myc-ARX* wt vector (Shoubridge *et al.*, 2007) and details of the final PCR reagent concentrations and thermocycle can be found in section 2.4.

Resultant plasmids were initially prepared on a small scale (mini prep, Promega) and initially screened by sequencing (as per section 2.4), using a primer in proximity to the site of the intended mutation (c.81 C>G - 2J-R1; c.1058 C>T, c.1074 G>T, c.1135 C>A, c.1136 G>T - ARXcR1; see Table 2.3a and g for primers). Subsequently, larger plasmid preparations (midi prep, Qiagen) were generated from the original single colony and the entire *myc-ARX* ORF was sequenced using the primers ARX-cR1, pCMV-myc-F and pCMV-myc-R primers (Table 2.3g) to confirm no other mutations had occurred during the *in vitro* mutagenesis process prior to transfection studies.

2.8 Cloning strategy of polyAlanine tract mutations into *ARX* ORF

All reagents used for this method are listed in Table 2.2b.

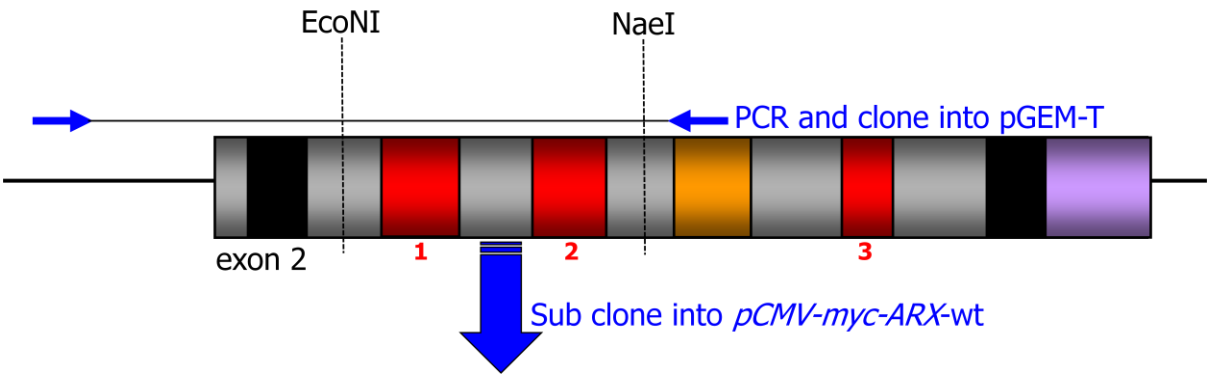

Figure 2.7a. Schematic of PCR, digest and sub-cloning strategy used to generate polyAlanine (pA) mutants from the *pCMV-myc-ARX* wt vector. Exon 2 of *ARX* is shown, which codes for NLSs (black), pA tracts (red), an acidic domain (orange) and part of the homeodomain (purple). Intronic DNA flanking exon 2 is represented by a black line. *Eco*NI and *Nae*I restriction enzymes were used to double digest both the PCR product (forward primer 2J-1, reverse primer 2J-2; as per Table 2.1a) and the *pCMV-myc-ARX* wt vector.

The initial cloning of the complete human wildtype (wt) *ARX* (562 aa), c.429_452dup ('dup24') and c.304ins$(GCG)_7$ ('expansion') mutant ORFs into the *pCMV-Myc* vector (Clontech) have been previously described (Shoubridge *et al.*, 2007) and all three constructs were kindly supplied by Dr Cheryl Shoubridge.

Genomic DNA from individuals with the relevant pA mutations was used for PCR amplification across both pA1 and pA2 using primers 2J-F1 and 2J-R2 (Figure 2.7; Table 2.1a). A DNA sample from an individual with the c.298_330dup (p.A111insA_{11}) mutation was kindly supplied by Shinji Saitoh (Kato *et al.*, 2007). The subsequent PCR products were cloned into the *pGEM-T* Easy Vector System I using manufacturer's instructions (Promega).

Both vectors, *pCMV-myc-ARX* wt and the *pGEM-T* with the desired pA1 or pA2 mutant PCR product, were double digested with *Eco*NI/*Nae*I in buffer 1 (New England Biolabs). 55 ng of linearised *pCMV-myc-ARX* wt was ligated to either 10 ng (1:3) or 20 ng (1:6) of purified and digested mutation containing pA1-pA2 PCR using T4 DNA Ligase (Promega). Subsequently, to increase efficiency of the cloning process, pA1-pA2 PCR products (75 ng)

were directly double digested and cloned into linearised *pCMV-myc-ARX* wt (500 ng), which also served to minimise mutations in the *ARX* ORF introduced by the bacteria while in culture.

Ligations were transformed into XL10-Gold Ultra competent cells (Agilent) and plated out. Colonies were selected on the basis of ampicillin resistance. Initially small overnight cultures were grown from single colonies and plasmid mini preparations (Promega) were double digested with *Eco*NI/*Nae*I (New England Biolabs). Constructs that gave bands of expected size with respect to the mutation being sub-cloned, were then re-grown from the original single colony and a larger volume plasmid preparation was performed (midi prep, Qiagen). The larger volume plasmid preparation had the entire *pCMV-myc-ARX* ORF sequenced using 3 primers as in section 2.7 (Table 2.3g) and were subsequently used for transfection experiments. *pCMV-myc-ARX* wt, 'dup24' and 'expansion' constructs previously generated (Shoubridge *et al.*, 2007) were also sequence confirmed across the entire ORF.

All bacteria were grown on LB agar plates or cultured in LB broth with 100 µg/ml (w/v) ampicillin added (LB broth, agar and ampicillin from Sigma) and incubated at 30 °C to minimise bacterial mutation of GC rich regions of the *ARX* ORF.

2.9 Cell culture

Reagents used in this section are listed in Table 2.2c.

Cell culture was performed as previously described (McKenzie *et al.*, 2007; Shoubridge *et al.*, 2007; Fullston *et al.*, 2010, Appendix A.6; Shoubridge *et al.*, 2010b, Appendix A.9). Briefly, HEK293T cells (ATCC) were maintained in Dulbecco's modified Eagle's medium (Invitrogen) supplemented with 10% (v/v) Fetal Calf Serum (FCS) (Invitrogen). All cells

were cultured using standard aseptic techniques in the presence of 100 U/ml sodium penicillin and 100 μg/ml of streptomycin sulphate (Penicillin-Streptomycin mixture, Invitrogen) in 5% CO_2 at 37 ^{o}C. Viable cells, as assessed by trypan blue (Invitrogen) exclusion, were counted by standard Haemocytometer techniques plated at $4x10^5$ per well in a 6 well plate the day before transfection in media lacking antibiotics.

Cells were transfected with a total of 1 μg of plasmid DNA and 4 μl Lipofectamine 2000 (Invitrogen) per well following manufacturer's instructions. The transfection protocol was scaled up by a factor of 7.5x to a T75 flask for use in western blot analysis. The scaling factor was calculated by direct comparison of the T75 (75 cm^2) flasks surface area compared to one well of a six well plate (10 cm^2), *ie* ≥7.5 μg of plasmid and 30 μl Lipofectamine 2000. 24 h post transfection cells were harvested, pelleted and snap frozen in liquid nitrogen and stored at -80 °C until required. Simultaneously a proportion of cells were briefly cultured on poly-lysine coated cover slips in 6 well plates until adherent and then fixed for subsequent immunohistochemistry.

2.10 Immunohistochemistry

Reagents used for the below method are listed in Table 2.2c.

Techniques used for ARX immunohistochemistry have been previously described (Shoubridge *et al.*, 2007; Shoubridge *et al.*, 2010b, Appendix A.9; Fullston *et al.*, 2011, Appendix A.2). Cells were harvested at 24 h post-transfection and fixed for 1 h in 3.7% formaldehyde (Sigma) in PBS (v/v) and permeabilised in 0.2% (v/v) triton-x PBS for 5 mins. Blocking was performed in 5% skim-milk powder (w/v) in Tris-buffered saline with 0.5% (v/v)Tween-20 (Sigma; TBS-T). Primary (2 h at room temp) and secondary (1 h at room

temp) antibodies were diluted in 1% skim-milk powder (w/v) in TBS-T. Removal of excess block and antibody was performed by 6 washes of TBS-T (3x 5-10 min, 3x 1 min). Nuclei were counterstained with a DAPI containing mounting media (VectaShield, Vector Laboratories).

Two primary antibodies were used to detect wt and mutant ARX proteins; mouse monoclonal α-ARX (1 μg/ml final) (Shoubridge *et al.*, 2007); and mouse monoclonal α-cMYC (Santa Cruz Biotech). The two secondary antibodies used were goat α-mouse-IgG conjugated to FITC (1:1000) (Dako) and goat α-mouse-IgG conjugated to Cy3 (1:1000) (Jackson Laboratories).

2.11 Fluorescence microscopy

For subcellular localisation studies between a total of 1,086-4,374 transfected cells were counted for each construct from at least three independent transfection reactions, using standard fluorescence microscopy (Shoubridge *et al.*, 2007; Shoubridge *et al.*, 2010b, Appendix A.9). Percent of abnormal localisation was determined as the proportion of cells exhibiting abnormal localisation (nuclear inclusions; partial or complete cytoplasmic localisation) relative to the number of ARX positive cells.

The total amount of abnormal localisation compared to that of the wildtype and other constructs was undertaken by general linear model and an unbalanced ANOVA analysis. A P value ≤ 0.05 was considered statistically significant.

2.12 Western immunoblotting

Reagents used for the following protocol are listed in Table 2.2c.

A previously described protocol was used for western immunoblotting of the ARX protein (Fullston *et al.*, 2010, Appendix A.6). Protein was extracted from cells harvested 24 h post transfection (Chapter 2.8) using RIPA buffer (50 mM Tris, 150 mM NaCl, 0.1% SDS, 0.5% Triton-X 100; 'activated' with 4% Protease cocktail inhibitor, 1 mM NaF, 1 mM Na_3VO_4, 0.5% PMSF; all components sourced from Sigma) and homogenised by repeated aspiration through a fine gauge needle (12x). Lysates were concentration tested by a Bradford protein assay (Quick Start kit, Bio-Rad) and were then subjected to standard SDS–PAGE using 4-12% Bis-Tris gels (Invitrogen), transferred to a nitrocellulose membrane, and analysed by immunoblotting using a mini cell electrophoresis and blot system (XCell, Invitrogen).

All primary and secondary antibodies used were diluted in 1% milk powder (w/v) in TBS-T and incubated with membrane for 2 h and 1 h (respectively) at room temperature. The membrane was initially blocked with 5% milk powder (w/v) and 5% horse sera (v/v) in TBS-T. Removal of excess block and antibody was performed by 6 washes of TBS-T (3x 5-10 min, 3x 1 min). The primary sheep polyclonal α-ARX antibody was diluted 1/2000 in and used for western blotting, the production of which has been previously described (Fullston *et al.*, 2010, Appendix A.6). The secondary antibody used was a Donkey α-Sheep IgG-HRP conjugated secondary antibody (Millipore) diluted 1/2000. Signal was detected via chemiluminescence (ECL, GE healthcare) and exposure of membrane to X-ray sensitive film. The membrane was stripped in 0.2 M NaOH for 5 minutes after ARX detection, to enable immunoblotting of the same membrane with the α-cMYC antibody. The primary mouse α-cMYC monoclonal antibody (Santa Cruz Biotechnology) was diluted 1/1000 and was detected by an α-Mouse IgG-HRP conjugated secondary antibody (Dako) diluted 1/2000.

2.13 List of URLs

Agilent	http://www.home.agilent.com/
AGRE:	http://www.agre.org/
Applied Biosystems	http://www.appliedbiosystems.com.au/
ATCC	http://www.atcc.org/
Bio-Rad	http://www3.bio-rad.com/
Clontech	http://www.clontech.com/
Dako	http://www.dako.com/
Epicentre Biotechnology	http://www.epibio.com/
Eppendorf	http://www.eppendorf.com/
GE Healthcare	http://www.gehealthcare.com/
Geneworks:	http://www.geneworks.com.au/
Invitrogen	http://www.invitrogen.com/
Jackson Laboratories	http://www.jax.org/
mfold web server	http://mfold.rna.albany.edu/?q=mfold/
Millipore	http://www.millipore.com/
Mo Bio	http://www.mobio.com/
NCBI:	http://www.ncbi.nlm.nih.gov/
New England Biolabs	http://www.neb.com/
Perkin Elmer	http://www.perkinelmer.com.au/
Promega	http://www.promega.com/
Qiagen	http://www.qiagen.com/
Quantum Scientific	http://www.quantum-scientific.com.au/
Sigma	http://www.sigmaaldrich.com/
Thermo Fisher Scientific	http://www.thermofisher.com/
Transgenomic	http://www.transgenomic.com/
UCSC	http://genome.ucsc.edu/
UCSC (uc element custom track)	http://users.soe.ucsc.edu/~jill/ultra.html
Vector Laboratories	http://www.vectorlabs.com/
Roche	http://www.roche.com/

Screening for Mutations in the *ARX* Gene and Putative Enhancer Regions

CHAPTER 3 CONTENTS

The majority of the work presented in this chapter has been published in Fullston *et al* (2011; Appendix A.2), and provided minor input for Brouwer *et al* (2004), Tarpey *et al* (2009), and White *et al* (2010) (Appendix A.3, A.4 and A.5 respectively).

3 SCREENING FOR MUTATIONS IN THE ARX GENE AND PUTATIVE ENHANCER REGIONS

Overall detail of our screening effort, which led to the discovery of various *ARX* mutations characterised in greater detail in Chapters 4 to 7, is presented here. Individual clinical assessment of patients suitable for *ARX* mutation screening, blood or DNA sample collection have been provided by collaborators listed in Fullston *et al* (2011) (Appendix A.2).

3.1 Introduction

More than 110 families and isolated cases have been reported to carry one of at least 50 *ARX* mutations (reviewed by (Gecz *et al.*, 2006; Shoubridge *et al.*, 2010a, Appendix A.1) reports by (Giordano *et al.*, 2010; Kato *et al.*, 2010; Conti *et al.*, 2011a; Cossee *et al.*, 2011; Eksioglu *et al.*, 2011)), including the four familial mutations and one novel mutation presented here and reported in Fullston *et al* 2011 (Appendix A.2). The most frequent *ARX* mutation (~39%; c.429_452dup(dup24bp) displays remarkable variation in clinical expressivity, conferring 10 clinical conditions (Table 3.1a) (Stromme *et al.*, 2002a; Turner *et al.*, 2002; Kato *et al.*, 2004; Gestinari-Duarte Rde *et al.*, 2006; Szczaluba *et al.*, 2006; Laperuta *et al.*, 2007; Rujirabanjerd *et al.*, 2007; Cossee *et al.*, 2011). Studies have identified *ARX* mutations in 9.5% of families with syndromic and non-syndromic X-linked intellectual disability (XLID) (Poirier *et al.*, 2006) and in 7.5% of such families with an asymptomatic obligate carrier female (de Brouwer *et al.*, 2007, Appendix A.3). Most *ARX* mutations are familial cases; however, a handful of *de novo* mutations have been identified (Bienvenu *et al.*, 2002; Kato *et al.*, 2003; Gronskov *et al.*, 2004). Mosaicism has been demonstrated in three families with the c.429_452dup (24bp) mutation (Partington *et al.*, 2004; Poirier *et al.*, 2005; Nawara

et al., 2006) and one with a c.430_456dup (27bp) mutation (Reish *et al.*, 2009, Appendix A.7) in the DNA coding for pA2.

Table 3.1a. Ten clinically distinct phenotypes are observed for mutations in the *ARX* gene, in order of discovery.

Phenotype	Acronym	MIM ID #[1]	Reference
X-linked infantile spasms (or West syndrome)	ISSX [2] (or WS)	308350	Strømme *et al.*, 2002a,b
Nonsyndromic X linked intellectual disability (ID)	nsXLID [2]		Strømme *et al.*, 2002b Bienvenu *et al.*, 2002
Partington syndrome	PRTS [2]	309510	Strømme *et al.*, 2002b Partington *et al.*, 2004
X-linked myoclonic epilepsy with ID and spasticity	XMESID [2]	300432	Strømme *et al.*, 2002b Scheffer *et al.*, 2002
X-linked lissencephaly with ambiguous genitalia	XLAG [3]	300215	Kitamura *et al.*, 2002 Kato *et al.*, 2004
Absence of the corpus callosum with abnormal genitalia (or Proud syndrome)	ACC/AG [3] (or PS)	300004	Kato *et al.*, 2004
Hydranencephaly with abnormal genitalia	HYD/AG [3]	300215	Kato *et al.*, 2004
Infantile epileptic-dyskinetic encephalopathy	IEDE [2]	308350	Guerrini *et al.*, 2007
Ohtahara syndrome	OS [2]	308350	Kato *et al.*, 2007
Intellectual disability with tonic seizures with dystonia	ID/TS/Dys [2]	309510	Shinozaki *et al.*, 2009

[1] Mendelian inheritance in man identification number
[2] non-malformation phenotype (shaded grey)
[3] malformation phenotype (unshaded)

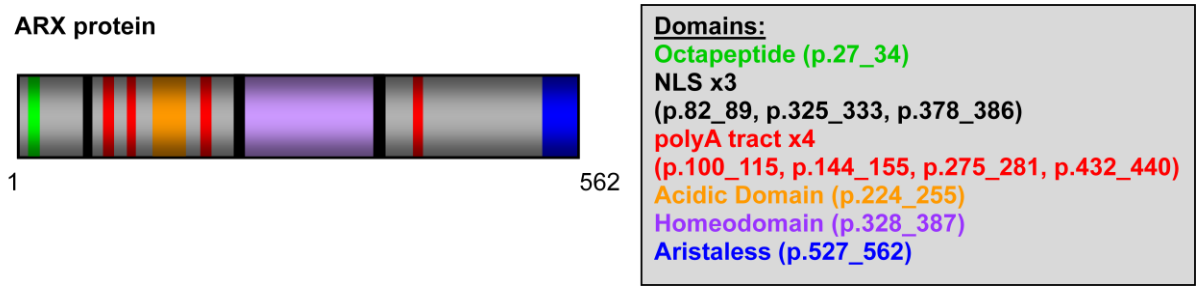

Figure 3.1a. Schematic of the 562 amino acid full length ARX protein. Domains are colour coded as per key.

Mutations in *ARX* potentially affect the function of one of a number of domains in the mature protein, contributing to the 10 known clinically distinct disorders (Table 3.1a). The domains that constitute the ARX protein include: an octapeptide domain (oct); three nuclear localisation sequences (NLS); four polyalanie tracts (pA); an acidic domain; a homeodomain (hom); and an Aritaless domain (OAR) (Figure 3.1a).

All patients have intellectual disability (ID) with and without additional features including epilepsy, infantile spasms, autism, dystonia, dysarthria, lissencephaly, hydranencephaly, abnormal genitalia and agenesis of the corpus callosum (reviewed by (Gecz *et al.*, 2006; Shoubridge *et al.*, 2010a, Appendix A.1)). Clinical features of these patients can be grouped into either: (i) non-malformation phenotypes (shaded grey in Table 3.1a), which are generally caused by *ARX* mutations that expand the first two polyalanine (pA) tracts or alter residues outside the homeodomain (reviewed by (Gecz *et al.*, 2006; Shoubridge *et al.*, 2010a, Appendix A.1) and report by (Kato *et al.*, 2004)); or (ii) brain and genital malformation (unshaded in Table 3.1a). Malformation pathologies are typically associated with mutations that truncate the ARX protein or alter critical conserved homeodomain residues (reviewed by (Gecz *et al.*, 2006; Shoubridge *et al.*, 2010a, Appendix A.1) and report by (Kato *et al.*, 2004)), consistent with mouse *Arx* knockout phenotypes (Kitamura *et al.*, 2002; Marsh *et al.*, 2009).

The *ARX* gene is flanked by some of the largest ultraconserved (uc) elements in the human genome: 11 in total (Table 3.1b; uc460-470); 10 are 3' of *ARX* (uc460-469); and eight of these are within the 3' introns of *POLA1* (uc460-467) (Bejerano *et al.*, 2004). Some of these uc elements have been shown to act as enhancers of *ARX* expression (uc467-469) (Colasante *et al.*, 2008); however, no obvious phenotype resulted when these elements were deleted in mice (Ahituv *et al.*, 2007). Interestingly a sequence variation in uc221 on chromosome 7

(alternatively referred to as I56i) has been detected in an individual with autism (Poitras *et al.*, 2010).

Table 3.1b. The ultraconserved (uc) elements flanking human the *ARX* gene. Positions given are from UCSC genome browser (Feb 2009 assembly GRCh37/hg19).

uc	bp	5' start	3' end	bp from *ARX*	5'/3' (*ARX*)	Intronic (*POLA1*)
460	275	24,823,511	24,823,785	198,030	3'	√
461	397	24,864,797	24,865,193	156,622	3'	√
462	779	24,894,826	24,895,604	126,211	3'	√
463	275	24,915,882	24,916,156	105,659	3'	√
464	770	24,916,158	24,916,927	104,888	3'	√
465	310	24,917,481	24,917,790	104,025	3'	√
466	349	24,946,458	24,946,806	75,009	3'	√
467	731	25,008,354	25,009,084	12,731	3'	√
468	489	25,017,563	25,018,051	3,764	3'	x
469	222	25,018,053	25,018,274	3,541	3'	x
470	341	25,401,216	25,401,556	367,151	5'	x

Gene positions: *ARX* (NM_139058) 25,021,815 to 25,034,065
POLA1 (NM_016937) 24,712,056 to 25,015,100

In this chapter findings are presented of screening a heterogeneous cohort of 613 patients referred to us by various colleagues from Australia and overseas (Table 2.1a). Of these, 500 had the entire open reading frame (ORF) screened and the results for these are presented. In 94/500 patients four distal ultraconserved elements from within the closest proximity to *ARX* (uc466-469) were also screened. More detailed results of our investigations into specific mutations and sequence alterations discovered are presented in Chapters 4 to7.

3.2 Strategy for *ARX* variation screening

Discovery of the *ARX* gene for XLID and epilepsy (Scheffer *et al.*, 2002; Stromme *et al.*, 2002b; Partington *et al.*, 2004) prompted the follow up study reported in this chapter. The referred patients had been tentatively diagnosed with an *ARX*-related pathology; typically infantile spasms, West syndrome, Partington syndrome or non-syndromic intellectual

disability (nsXLID). Progressively we accumulated a cohort of 613 individuals with ID as their common feature (as per Table 2.1a) and initially screened these for the most frequent *ARX* mutations (c.304ins$(GCG)_7$ and c.429_452dup), using previously described PCR conditions (Stromme *et al.*, 2002b) and standard 2% agarose gel electrophoresis (primers sequences in Table 2.3a). Samples generating PCR products of abnormal size (Figure 3.2a i) had the remainder of the PCR product purified and directly sequenced on an automated capillary sequencer (ABI 3100, Applied Biosystems, Foster City, USA).

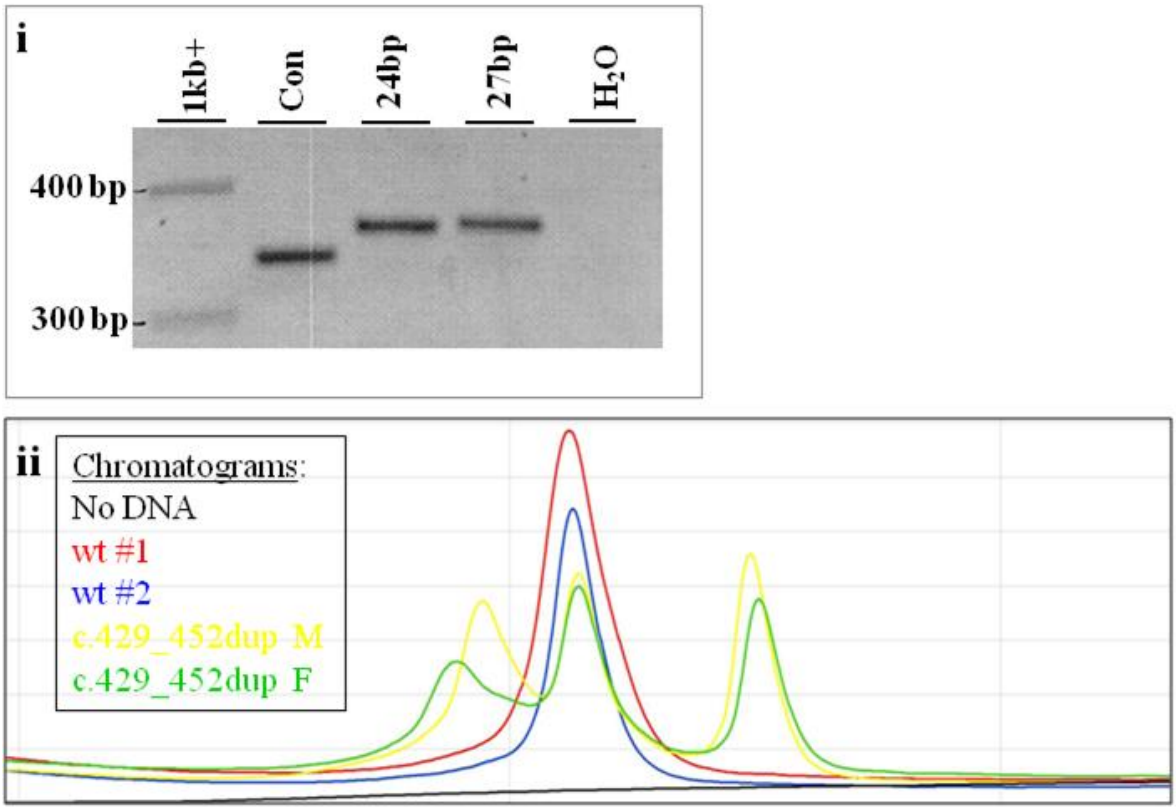

Figure 3.2a. Representative images of the *ARX* c.429_452dup(24 bp) mutation from **(i)** agarose gel electrophoresis or **(ii)** dHPLC analysis. **(i)** Products from a PCR that spans the DNA coding pA2 (2J-2 primers; Table 2.3a), separated in a 2% agarose gel with ethidium bromide: 1 kb+ marker DNA (1kb+ DNA ladder, Invitrogen, Melbourne, Australia); Con control DNA input (313 bp; wildtype); 24 bp DNA input from a male with the pA2 c.429_452dup(24bp) mutation (337 bp); 27 bp input from a male with the pA2 c.430_456dup(27 bp) mutation (340 bp); and H_2O is no DNA input. Relevant band sizes within the marker lane are indicated to the left of the gel. The gel was visualised under UV light and the negative image was cropped and resized to produce this figure. **(ii)** dHPLC chromatograms for DNA inputs of wildtype (wt), c.429_452dup(24 bp) from a hemizygous male (M) or heterozygous female (F) and no DNA input. Note that dHPLC analysis included heteroduplexing with wildtype DNA, and as such female and male DNA does not have vastly different chromatograms.

A subset of patient samples that were negative for the recurrent pA1/pA2 tract mutations that had sufficient DNA quality/quantity (500/613) were screened for mutations across the entire *ARX* ORF by single stranded conformation polymorphism (SSCP; 314/500 samples; Chapter 2.6.1) and/or denaturing high pressure liquid chromatography (dHPLC; 312/500 samples – 126 previously SSCP screened; Chapter 2.6.2). A selection of patients (94/500) negative for mutations in the *ARX* ORF (Table 2.1a) were subsequently screened for DNA variation in four 3' uc elements closest to the *ARX* locus (uc466-469) by dHPLC.

Genomic DNA from patients with known mutations was used as a positive control, whereas DNA with sequence confirmed wildtype *ARX* sequence was used as a negative control. Samples returning abnormal SSCP scans or dHPLC chromatograms (Figure 3.2a ii) were re-amplified, directly sequenced and compared to the *ARX* reference sequence (NM_139058). Once a DNA variant was discovered, segregation testing within the family was performed on DNA samples when available from other family members.

3.3 Mutations discovered in the *ARX* coding sequence.

A total of six mutations in eight families were identified (summarised in Figure 3.3a and Table 3.3a): c.423_455dup(33bp) (1 family (Demos *et al.*, 2009, Appendix A.8)); c.430_456dup(27bp) (1 family (Reish *et al.*, 2009, Appendix A.7)); c.81 C>G (1 family (Fullston *et al.*, 2010) , Appendix A.6); c.1136 G>T (1 family (Shoubridge *et al.*, 2010b, Appendix A.9)); c.1074 G>T (1 family (Fullston *et al.*, 2011) , Appendix A.2); and c.429_452dup(24bp) (3 families (Fullston *et al.*, 2011) , Appendix A.2). Segregation testing on available DNA samples from family members and cell-based assays were used to address the pathogenic effect of these and other selected mutations (presented in Chapters 4 to 7).

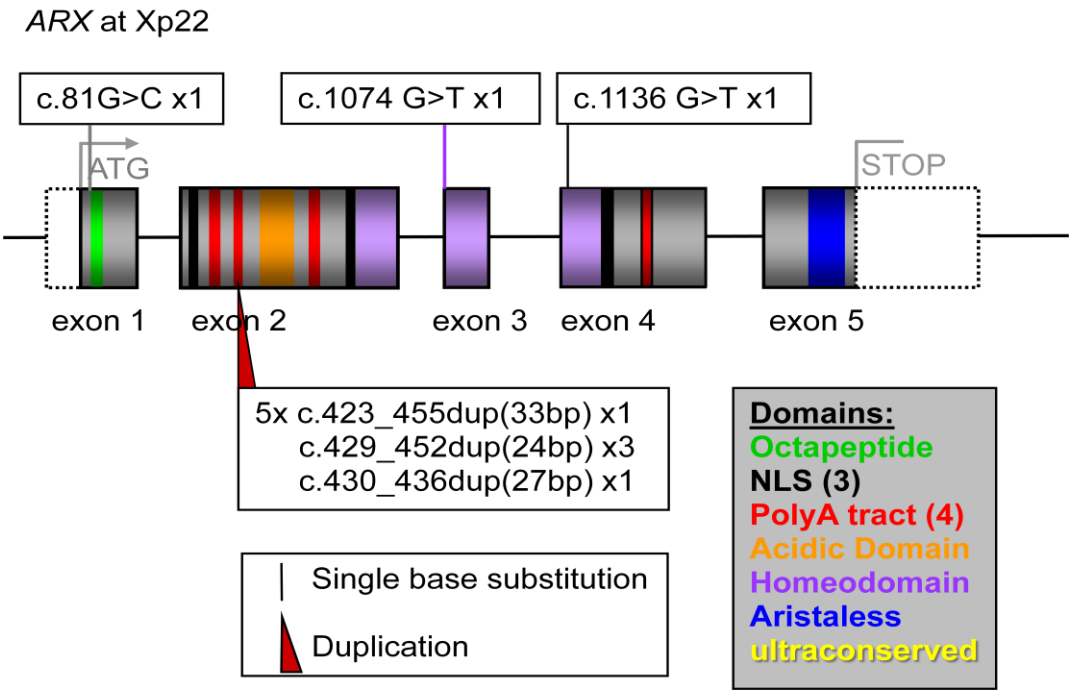

Figure 3.3a. Schematic summary of the six *ARX* mutations discovered in eight families. The number of families discovered for each mutation is shown.

Table 3.3a. Details of the six mutations identified.

Mutations (cDNA – bp)	Exon	Change (Prtoein – aa)	Domain	Type	Phenotype	MIM	Reference
c.81C>G	1	p.Y27X	oct	non	OS/ISSX (WS)	308350	Fullston *et al.*, 2010
c.423_455dup (33bp)	2	p.A151delAA--insG$(A)_{12}$	pA2	dup	PRTS +	309510	Demos *et al.*, 2009
c.429_452dup (24bp)	2	p.A151insA_8	pA2	dup	nsXLID ISSX (WS)	- 308350	Fullston *et al.*, 2011
c.430_456dup (27bp)	2	p.A151insA_9	pA2	dup	ISSX (WS) +	308350	Reish *et al.*, 2009
c.1074G>T	3	p.R358S	hom	mis	ISSX/ACC/AG	308350 300004	Fullston *et al.*, 2011
c.1136G>T	4	p.R379L	hom	mis	XLAG	300215	Shoubridge *et al.*, 2010

non: nonsense, dup: duplication, mis: missense, OS: Ohtahara syndrome, PRTS: Partington syndrome, +: additional phenotypic features, nsXLID: non syndromic X linked intellectual disability, ISSX: infantile spasms, X-linked, WS: West syndrome, ACC: agenesis of the corpus callosum, AG: abnormal genitalia, XLAG: X-linked linssencephaly with ambiguous genitalia.

3.4 Alterations discovered in ultraconserved (uc) element sequences

The 94/500 patients selected for subsequent dHPLC screening for uc element variation had previously been screened for mutations in the *ARX* ORF and had mild phenotypes, such as autism (55/94) and ID without dysmorphic features (39/94) (Table 2.1a). The phenotype selection for uc element screening was based on the hypothesis that mutations outside of the *ARX* ORF would result in a less severe patient phenotype.

The cohort of patients with autism were sourced from the autism genetics research exchange (AGRE, Los Angeles, USA). AGRE patients were selected based on matching Xp22 haplotypes from multiple affected male families. Autism was diagnosed using the standard autism diagnostic interview – revised (ADI-R) algorithm.

Interestingly, three unique DNA variants were discovered in or immediately adjacent to two ultraconserved (uc) elements (466 and 467; Table 3.4a and Figures 3.4a, 3.4b and 3.4c) within the 94 patients screened for uc element sequence variation.

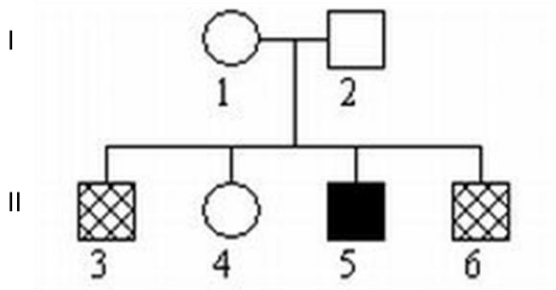

Figure 3.4a. Pedigree of AGRE family AU0432 with the uc466 54 G>A variation discovered in the proband (II-5). Solid black squares indicate a male with autism and cross-hatched squares indicate a male with broad spectrum autism disorder.
(Family E in Fullston *et al* 2011, Appendix A.2)

Clinical notes of family AU0432– provided by the AGRE

The DNA from the proband, II-5 (Figure 3.4a), was obtained from the Autism Genetic Research Exchange (AGRE, Los Angeles, USA). The diagnosis of Autism had been made for individual II-5 using the standard autism diagnostic interview – revised (ADI-R) algorithm, whilst his brothers (II-3 and II-6) were affected with broad spectrum disorder (*ie* autistic traits without scoring above the cut-off to be diagnosed as autistic).

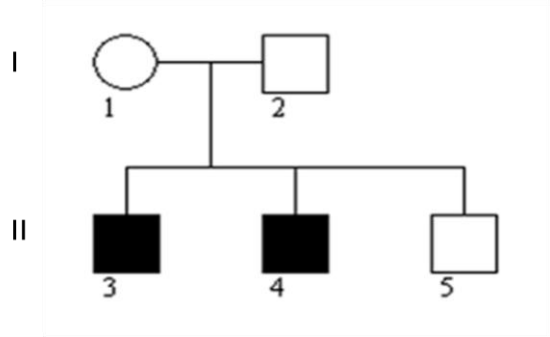

Figure 3.4b. Pedigree of AGRE family AU0598 with the uc466 -30 G>T variation discovered in the proband (II-3). Solid black squares indicate a male with autism.
(Family F in Fullston *et al* 2011, Appendix A.2)

Clinical notes of family AU0598– provided by the AGRE

The DNA from the proband, I-3 (Figure 3.4b), was obtained from the AGRE. The diagnosis of Autism had been made for individuals II-3 and II-4 using the standard ADI-R algorithm.

Clinical notes for family with uc 467 202 C>T variation – provided by Dr. Louise Brueton

Dr. Louise Brueton practices in the Clinical Genetics Unit, Birmingham Women's Hospital, Edgbaston, Birmingham, UK.

The mother was the only DNA sample that could be obtained and was screened as an obligate carrier. Her brother and son suffered from intractable infantile seizures, developmental delay, EEG abnormalities and minor brain abnormalities, which was suggestive of an X-linked mode of inheritance. This family was subsequently lost to follow up (Family G in Fullston *et al* 2011, Appendix A.2).

Table 3.4a. DNA sequence variation identified in the uc elements.

uc element bp variation (length)	Genomic co-ordinates of element (UCSC [1])	5' or 3' of *ARX* (dist to uc bp)	Phenotype	MIM	Reference
uc 466 54 G>A (349 bp)	Start 24 946 458 End 24 946 806	3' (75 358 bp) (Intron 2 of *POLA*)	Aut	209850	Fullston *et al.*, 2011
uc 466 -30 G>T (349 bp)	Start 24 946 458 End 24 946 806	3' (75 358 bp) (Intron 2 of *POLA*)	Aut	209850	Fullston *et al.*, 2011
uc 467 202 C>T (731 bp)	Start 25 008 354 End 25 009 084	3' (12 732 bp) (Intron 1 of *POLA*)	ISSX(WS)	308350	Fullston *et al.*, 2011

[1] UCSC build GRCh37
Aut: Autism, ISSX: infantile spasms, X-linked, WS: West syndrome.
ARX (NM_139058) 5'- 25,034,065 to 3'- 25,021,815; spanning 12,251 bp
POLA1 (NM_016937) 5'- 24,712,056 to 3'- 25,015,100; spanning 303,045 bp

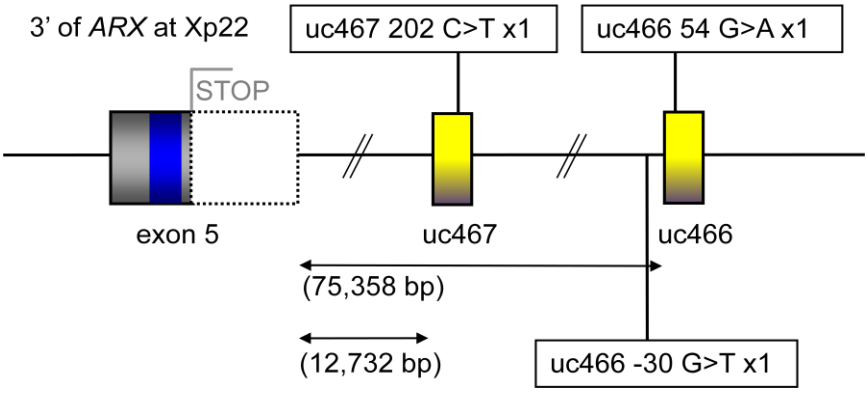

Figure 3.4c. Schematic of the 3 uc element variants identified in the context of *ARX* genomic location.

3.5 Non pathogenic variants identified

Naturally occurring *ARX* sequence variants were identified that either (a) did not co-segregate with disease in the family (*ie* found in unaffected males), (b) were silent nucleotide substitutions or (c) previously been reported as non-pathogenic or of questionable pathogenicity (Fullston *et al.*, 2011, Appendix A.2). These variants presented as either SNPs or small insertions/deletions in pA1 (Table 3.5a). The pA1 variations are investigated further in Chapter 5.

Table 3.5a. Details of presumed non-pathogenic variants identified.

Sequence variation	Exon (Intron)	Change (Prtoein)	Domain	Type	Fam (#)	Phenotype
1. c.304ins$(GCG)_1$ /c.441A>G [1]	2	p.A111delA_1 /p.A147A	pA1/pA2	ins/sil	1	nsXLID/none
2. c.305_313del(9bp)	2	p.A111delA_3	pA1	del	1	Epi/Seizures
3. c.327_341del(15bp)	2	p.A111delA_5	pA1	del	1	Epi/Seizures
4. c.447G>C [1]	2	p.A149A	pA2	sil	1	ISSX
5. c.1073+14G>T	(2)	-	-	sil	1	nsXLID
6. c.1347 C>T	4	p.G449G	-	sil	15	nsXLID, ISSX, Epi, Aut
7. c.1671 G>A	5	p.T557T	aristaless	sil	5	nsXLID/ISSX

[1] Sequence variation occurs within the sequence duplicated by the c.429_452dup (24bp) mutation.
ins: insertion, sil: silent, del: deletion, nsXLID: non syndromic X linked intellectual disability, Epi: Epilepsy, ISSX: infantile spasms, X-linked, Aut: Autism.

3.6 Discussion

Previous studies identified *ARX* mutations as a frequent cause of X-linked intellectual disability (XLID), especially in families with a clear X-linked mode of inheritance (9.5% (Poirier *et al.*, 2006) and in families with an obligate carrier female (7.5% (de Brouwer *et al.*, 2007, Appendix A.3). As expected, *ARX* mutation frequency is reduced in families with a questionable X-linked mode of inheritance, *ie* previous reports in brother pairs (or putative XLID families) between 1.5% (Poirier *et al.*, 2006) and 2.2% (de Brouwer *et al.*, 2007, Appendix A.3). The frequency diminishes even further to potentially very low detection in sporadic cases of ID (~0.15% (Gronskov *et al.*, 2004)).

Mutation Screening

From a mixed cohort of 613 individuals with ID, that included screening the entire *ARX* ORF for mutations in 500/613 samples, a total of six mutations were discovered in eight families, therefore 1.31% (8/613) of this cohort has an *ARX* mutation (Table 3.6a). This is the first time nucleotide substitutions in ultraconserved elements flanking *ARX* have been discovered in two families, one with autism and one with X-linked infantile spasms.

Table 3.6a. Subsets of patients within the heterogeneous cohort and mutations/sequence variations discovered within them.

Sub cohort	Patients (#)	Mutation (# Fam)	Freq (%)	uc variation (# Fam)	Freq (%)	Non pathogenic (# Fam)	Freq (%)
FRAXA/E neg	147	-	0.00	-	0.00	c.1347C>T (x6)	4.08
Possibly X-linked	123 (14)[1]	-	0.00	-	0.00	c.1347C>T (x2)	1.63
						c.1671G>A (x1)	0.81
						c.1073+14G>T (x1)	0.81
							(3.25)
'*ARX* testing' referrals[3]	117 (24)[1]	c.429_452dup24 (x3)	[1.30]	uc 467 202C>T (x1)	0.85	c.1347C>T (x3)	2.56
		c.423_455dup33 (x1)	[0.43]			c.1671G>A (x1)	0.85
						c.304ins$(GCG)_1$/c.441A>G (x1)	0.85
	[+113]		[1.74]				(4.27)
Autism spectrum disorder	68 (55)[1,2]	-	0.00	uc 466 54G>A (x1)	1.47	c.1347C>T (x2)	2.94
				uc 466 -30G>T (x1)	1.47	c.1671G>A (x1)	1.47
					(2.94)		(4.41)
Epilepsy/ infantile spasms, X-linked	37 (1)[1]	c.81C>G (x1)	2.70	-	0.00	c.1347C>T (x2)	5.41
		c.430_456dup27 (x1)	2.70			c.1671G>A (x2)	5.41
						c.447G>C (x1)	2.70
						c.327_341del (x1)	2.70
						c.304_312del (x1)	2.70
			(5.41)				(18.92)
Brain malformations	8	c.1074G>T(x1)	12.50	-	0.00	-	0.00
		c.1136G>T (x1)	12.50				
			(25.00)				
Total	500	pA2 dup24 bp (x3)	0.82[3]	uc 467 202C>T (x1)	1.06	c.1347C>T (x15)	3.00
		pA2 dup27 bp (x1)	0.20	uc 466 54G>A (x1)	1.06	c.1671G>A (x5)	1.00
		pA2 dup33 bp (x1)	0.20	uc 466 -30G>T (x1)	1.06	c.304ins$(GCG)_1$/c.441A>G (x1)	0.20
		c.81C>G (x1)	0.20			c.304_312del (x1)	0.20
		c.1074G>T(x1)				c.327_341del (x1)	0.20
		c.1136G>T (x1)				c.447G>C (x1)	0.20
						c.1073+14G>T (x1)	0.20
(overall frequency)	[613]	**6 mutations (x8)**	**(1.42 [1.31][3])**	**3 sequence variants(x3)**	**(3.19)**	**8 sequence variants (x25)**	**(5.00)**

Fam: The number of families discovered with this mutation/variation.
[1] Number of patients from this sub group within the cohort screened for uc element sequence variation
[2] 55 families with multiple affected males Xp22 haplotype matched from AGRE
[3] 613 used as a denominator (# screened for pA2 dup/pA1 expansion mutations only, extra 113 patients were '*ARX* testing' referrals [+113])

Given the majority of patients were isolated cases, a 1.31% mutation frequency is consistent with previous reports within brother pairs (or putative XLID families) between 1.5% (Poirier *et al.*, 2006) and 2.2% (de Brouwer *et al.*, 2007, Appendix A.3). As anticipated, this frequency was significantly lower than previous reports of screening within large families within a well characterised XLID cohort (9.5% (Poirier *et al.*, 2006)). Half (3/6) of all mutations discovered were outside exon 2 and a third (2/6) were outside exon 1 and 2 (Table 3.6a), therefore mutation screening limited to exon 1 and 2 would likely underestimate *ARX* mutation frequency.

Of the mutations found in eight families: 62.5% (5/8) were discovered in exon 2, all of which were duplications of various length within the pA2 coding DNA; and 37.5% (3/8) were c.429_452dup (24bp) mutations. This study confirms exon 2, especially pA2 coding DNA, as an *ARX* mutational 'hot spot' (reviewed by (Gecz *et al.*, 2006; Shoubridge *et al.*, 2010a, Appendix A.1)). However, mutations outside of this region also contribute appreciably to *ARX* mutations, especially for more severe brain and genital malformation phenotypes (*eg* c.1076 G>T and c.1136 G>T in exon 3 and 4, respectively).

No *ARX* ORF mutations were detected in the FraX A/E negative and autism cohorts, consistent with previous reports (Gronskov *et al.*, 2004; Chaste *et al.*, 2007). Nor were any mutations identified in families with possible X-linked inheritance of ID. The number of patients in this possible X-linked inheritance of ID subset (*ie* 123) was only expected to contain approximately two *ARX* mutations, so failure to detect any was not unexpected. Given previous studies have discovered 2/148 (Bienvenu *et al.*, 2002) and 2/76 (Poirier *et al.*, 2006) *ARX* mutations in families with a putative X-linked mode of inheritance, our negative finding in this subset of patients finding may reflect lower stringency in patient selection for testing compared with previously published cohorts. A low percentage of samples (1.74%; 4/230) referred for '*ARX* testing' due to symptoms within the '*ARX* spectrum' harboured *ARX*

mutations, a similar rate to what has been previously reported for brother pairs of 1.5-2.2% (Poirier *et al.*, 2006; de Brouwer *et al.*, 2007, Appendix A.3). A moderate percentage of patients (5.41%; 2/37) with epilepsy or infantile seizures had mutations in *ARX*, which is consistent with this patient sub group (Guerrini *et al.*, 2007), but may merely reflect small sample size. A quarter of patients (2/8) with brain malformations were identified with an *ARX* mutation. This frequency is bolstered by an XLAG sub-group (*ie* 1/2 families with XLAG had a mutation), which have been reported to have *ARX* mutations in the majority of cases examined (*ie* >90%) (Kitamura *et al.*, 2002).

These results suggest that a screening strategy based on clinical presentations should be employed. This would see the screening for polyalanine tract 1 and 2 or exon 2 mutations prioritised for milder ID pathologies with or without seizures, whereas the entire *ARX* ORF should be screened for brain/genital malformation phenotypes. The major caveat placed on this strategy is that for future screening to expand the clinical spectrum of *ARX* related disorders, new pathologies not currently related to *ARX* variations should also be considered.

Also screening efforts that encompass the non-coding regions at or surrounding the *ARX* locus should be considered. Examples of studies that have extended the phenotypic spectrum for *ARX* related disorders include Ohtahara syndrome (Kato *et al.*, 2007; Absoud *et al.*, 2010; Fullston *et al.*, 2010, Appendix A.6; Giordano *et al.*, 2010; Kato *et al.*, 2010; Eksioglu *et al.*, 2011) and the implication of uc element variations in autism presented here (Fullston *et al.*, 2011, Appendix A.2)

Variation in ultraconserved (uc) elements

Elements defined as ‘uc’ are DNA sequences longer than 200 bp that are invariant among human, rat, and mouse genomes, and display high levels of identity within chicken, dog and

fish (Bejerano *et al.*, 2004). They are often located near genes involved in transcriptional regulation and development, such as *ARX*, and display a rate of nucleotide exchange 20 times slower than the genome average (Bejerano *et al.*, 2004). Sequencing of uc466 and uc467 in 20 human X chromosomes did not display any sequence variation, consistent with constraints on variation of uc elements enforced by natural selection. Three unique changes were found in three separate families from a cohort of 94 patients (Table 3.4a and Figure 3.4c) in two uc elements distal to *ARX* (uc466, uc467).

The region containing and partly flanking uc467 acts as an enhancer for *Arx*, yielding an '*Arx*-like' pattern of expression, probably due to the binding of DLX2 to a minimal region within it (Colasante *et al.*, 2008) (Figure 3.6a). A large duplication of the *Arx* locus, including uc466 and uc467, causes eye and forebrain pathology in mouse (exencephaly and anopthalmia (Cunningham *et al.*, 2002)). As part of a recent study, the region encompassing the *ARX* locus, uc467 and the last exon of *POLA*, was duplicated and co-segregated with moderate ID (Chapter 8 and (Whibley *et al.*, 2010, Appendix A.10)). Although relevant patient material was not available for testing, this duplication presumably caused pathology due to an increase in *ARX* expression.

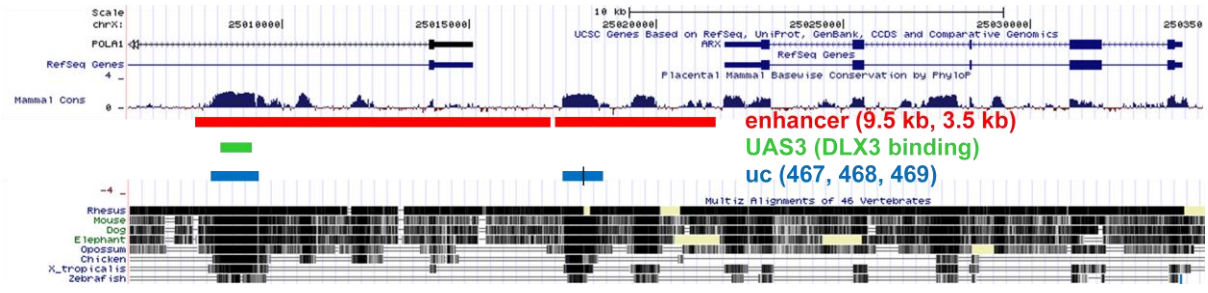

Figure 3.6a. Conservation and enhancers 3' of the *ARX* locus. The region between *ARX* and *POLA* is shown in the UCSC genome browser (Chapter 2.13) using the mammalian conservation track. Overall conservation is shown directly below the genes and specific species are shown individually at the bottom. Shown are the 3.5 kb/9.5 kb enhancer regions and the DLX2 binding region from Colosante *et al.*, 2008, together with uc 467, 468 and 469.

Other duplications at Xp21-22, that include *ARX* and other XLID genes, have also been discovered in two families with multiple affected individuals, including females, with developmental delay (Thorson *et al.*, 2010). In contrast, a knockout of uc467 showed no discernible phenotype in a mouse model (Ahituv *et al.*, 2007) suggesting that either redundancy exits between uc elements or that the resultant phenotype, on an inbred mouse background, is too subtle to detect or not present at all. Phenotypic differences between rodent models and humans harbouring identical mutations have been previously described. For example, in humans subtype three of maturity-onset diabetes of the young is caused by heterozygous mutations in the transcription factor hepatocyte nuclear factor (*HNF1a*) (Yamagata *et al.*, 1996). In contrast, mice with heterozygous mutations in *Hnf1a* gene are phenotypically normal (Pontoglio *et al.*, 1996). A reduction in *ARX* expression is likely to play a role in the observed heterozygous female carrier phenotypes (Turner *et al.*, 2002; Poirier *et al.*, 2005). The sequence variation discovered within uc467 at nt 202 C>T is 122 bp away from a minimal DLX2 binding site (HD-2) (Colasante *et al.*, 2008) and may impact on *ARX* expression.

Sequence variations observed in and immediately adjacent to uc466 (nt 54 G>A, nt -30 G>T) was found in 3.6% (2/55) patients with autism (II-5 Figure 3.4a and II-3 Figure 3.4b). The effect of these changes on *ARX* expression can only be speculated. A variation in uc221 element (alternatively referred to as I56i; located on chromosome 7) has been previously reported in an individual with autism (Poitras *et al.*, 2010). This pathology was suggested to be due to diminished enhancer activity shown for the element with altered DNA sequence, attributed to the observed reduction of Dlx protein binding in a mouse model (Poitras *et al.*, 2010). The 3' distance of uc466 from *ARX* is 75.4 kb (Table 3.4a) and the ~9.5 kb construct used for the mouse model lacked enhancers for testis and muscle *Arx* expression (Colasante *et al.*, 2008) (Figure 3.6a). It is therefore conceivable that uc466 may also act as an enhancer

of *ARX* expression, even though it is appreciably further 3' and remains to be tested for enhancer activity.

The closest uc element to *ARX* (uc467) was found to have an altered sequence in an individual with a relatively severe phenotype, whilst changes in a more distant uc element (uc466) were found in individuals with a milder phenotype. Human tissues with sufficient *ARX* expression are not accessible to test the effect that uc sequence variation might have on expression levels. Segregation testing and functional investigation of uc elements is warranted, now that these variations in ultraconserved elements have been associated with XLID.

These results reinforce previous observations of genotype-phenotype correlation for mutations in *ARX* (reviewed by (Gecz *et al.*, 2006; Shoubridge *et al.*, 2010a, Appendix A.1) and reported by (Kato *et al.*, 2004)). The discovery of 5 pA2 duplications of various sizes (24bp, 27bp and 33bp), 3 point mutations and particularly variations within ultraconserved non-coding putative enhancer regions 3' of *ARX* provide meaningful extensions. It underlines the importance of *ARX* mutation screening in various clinically defined patient groups. The contribution of non-coding and copy number variants to *ARX* spectrum disorders is open to further study in light of the results presented here.

An *ARX* Nonsense Mutation c.81 C>G (p.Y27X) Leads to Both Ohtahara and West Syndrome, but Not XLAG

CHAPTER 4 CONTENTS

The majority of the work presented in this chapter has been published in Fullston *et al* (2010; Appendix A.6) and part of Fullston *et al* (2011; Appendix A.2).

4. AN ARX NONSENSE MUTATION C.81 C>G (p.Y27X) LEADS TO BOTH OHTAHARA AND WEST SYNDROME, BUT NOT XLAG

The clinical descriptions of affected family members, segregation testing and mutant protein analysis of the c.81C>G (p.Y27X) *ARX* mutation are presented within this chapter. Clinical data and DNA sample collection have been provided by collaborators and co-authors as in Fullston *et al* (2010) (Appendix A.6).

4.1 Introduction

The first report of Ohtahara syndrome (OS) with mutations in the *ARX* gene was presented by Kato *et al* (2007). Subsequently six additional mutations have been reported as causative of OS in hemizygous males (Absoud *et al.*, 2010; Fullston *et al.*, 2010, Appendix A.6; Giordano *et al.*, 2010; Kato *et al.*, 2010; Eksioglu *et al.*, 2011) (summarised in Table 4.1a)

Table 4.1a. Summary of mutations in *ARX* that cause Ohtahara syndrome.

Mutations (DNA – bp)	**Exon**	**Change (Prtoein – aa)**	**Domain affected**	**Phenotype**	**Reference**
1. c.298_330dup	2	p.A111insA_{11}	pA1 (27A)	OS	Kato *et al.*, 2007
2. c.81C>G *	**1**	**p.Y27X**	**oct**	**OS/WS**	**Fullston *et al.*, 2010**
3. c.304ins$(GCG)_7$	2	p.A111insA_7	pA1 (23A)	OS +	Absoud *et al.*, 2010
4. c.1564_1568dup	5	p.A524fsX534	aristaless	OS/WS +	Kato *et al.*, 2010
5. c.1604_1605insT	5	p.E536fsX672	aristaless	OS/WS/LGS	Kato *et al.*, 2010
6. c.1604T>A	5	p.L535Q	aristaless	OS	Giordano *et al.*, 2010
7. c.1471_1472insC	5	p.L491fsX531	aristaless	OS +	Ekşioğlu *et al.*, 2011

OS: Ohtahara syndrome, WS: West syndrome, LGS: Lennox-Gastaut syndrome, +: additional features, *: mutation presented here.

Ohtahara syndrome is an early infantile epileptic encephalopathy (EIEE) with the earliest known age of onset for any age-dependant epileptic syndrome, often presenting within days of birth, or possibly observed prenatally as violent fetal movement (Yamatogi and Ohtahara, 2002; Ohtahara and Yamatogi, 2006). OS is characterized by frequent minor generalized

seizures and 'burst-suppression' on EEG readouts. 'Bursts' are high amplitude slow activities, often with sharp waves or spikes, interspersed with periods of 'suppression' with little or no activity (for an example see upper right panel in Figure 4.2b). The seizures are intractable and are co-morbid with severe psychomotor disability. Prognosis is poor with about one in three patients dying before the second year of life.

OS can be distinguished from other neonatal epileptic encephalopathies by the disease's progression during development (Ohtahara and Yamatogi, 2006). OS progresses into West syndrome (WS) in 75% of cases at approximately 3-4 months of age. Following this, 59% of WS cases develop into Lennox-Gastaut syndrome (LGS) (Yamatogi and Ohtahara, 2002), distinguishing it from early myoclonic encephalopathy, another EIEE, which does not progress into WS nor LGS (Ohtahara and Yamatogi, 2006).

The aetiology of OS is heterogeneous, and can be caused by structural brain abnormalities such as hemimegencephaly, porencephaly, hydrocephaly and lissencephaly and include symptoms such as those present in Aicardi syndrome (Yamatogi and Ohtahara, 2002; Ohtahara and Yamatogi, 2006). There is also a cryptogenic non-malformation form of OS. Similar to intellectual disability there is a 30% excess of males with OS, which is indicative of the involvement of genes located on the X chromosome (Djukic *et al.*, 2006). *ARX* is expressed in GABAergic neurons and Kato *et al* (2007) hypothesized that dysfunction of the GABAergic system is critical to the neuropathology of EIEE and West syndrome.

Curiously, the six other *ARX* mutations reported that cause OS (besides the one presented here) either expand polyalanine tract 1 (pA1) (Kato *et al.*, 2007; Absoud *et al.*, 2010), or mutate (Giordano *et al.*, 2010) or truncate (Kato *et al.*, 2010; Eksioglu *et al.*, 2011)) the aristaless domain. Given that a complete loss of ARX function results in an XLAG

phenotype (Kato *et al.*, 2004; Bhat *et al.*, 2005), it can be implied that a partial loss of ARX function precipitates OS.

The *ARX* mutation identified in this family (c.81C>G) is predicted to cause a premature termination codon and as a consequence ARX protein truncation (p.Y27X). Other *ARX* mutations predicted to cause a truncation of the ARX protein within the first 200 of 562 amino acids have been described (EX2-5del (Kato *et al.*, 2004), c.196+2T>C (Kato *et al.*, 2004), c.232G>T (Kato *et al.*, 2004); EX1-2.1del (Kitamura *et al.*, 2002); c.335_368del (Kato *et al.*, 2004); c.392_452del (Kato *et al.*, 2004); c.420_451del (Kitamura *et al.*, 2002)). Interestingly, all of these mutations cause the XLAG phenotype, which represents the severe brain and genital malformation end of the phenotypic spectrum of *ARX*-associated disorders (Kato *et al.*, 2004; Bhat *et al.*, 2005) and can be considered as an *ARX* null mutation. This is further supported by the original discovery of the human XLAG mutations that mimic the phenotype of the *Arx* knockout mouse (Kitamura *et al.*).

4.2 Patient and family investigations

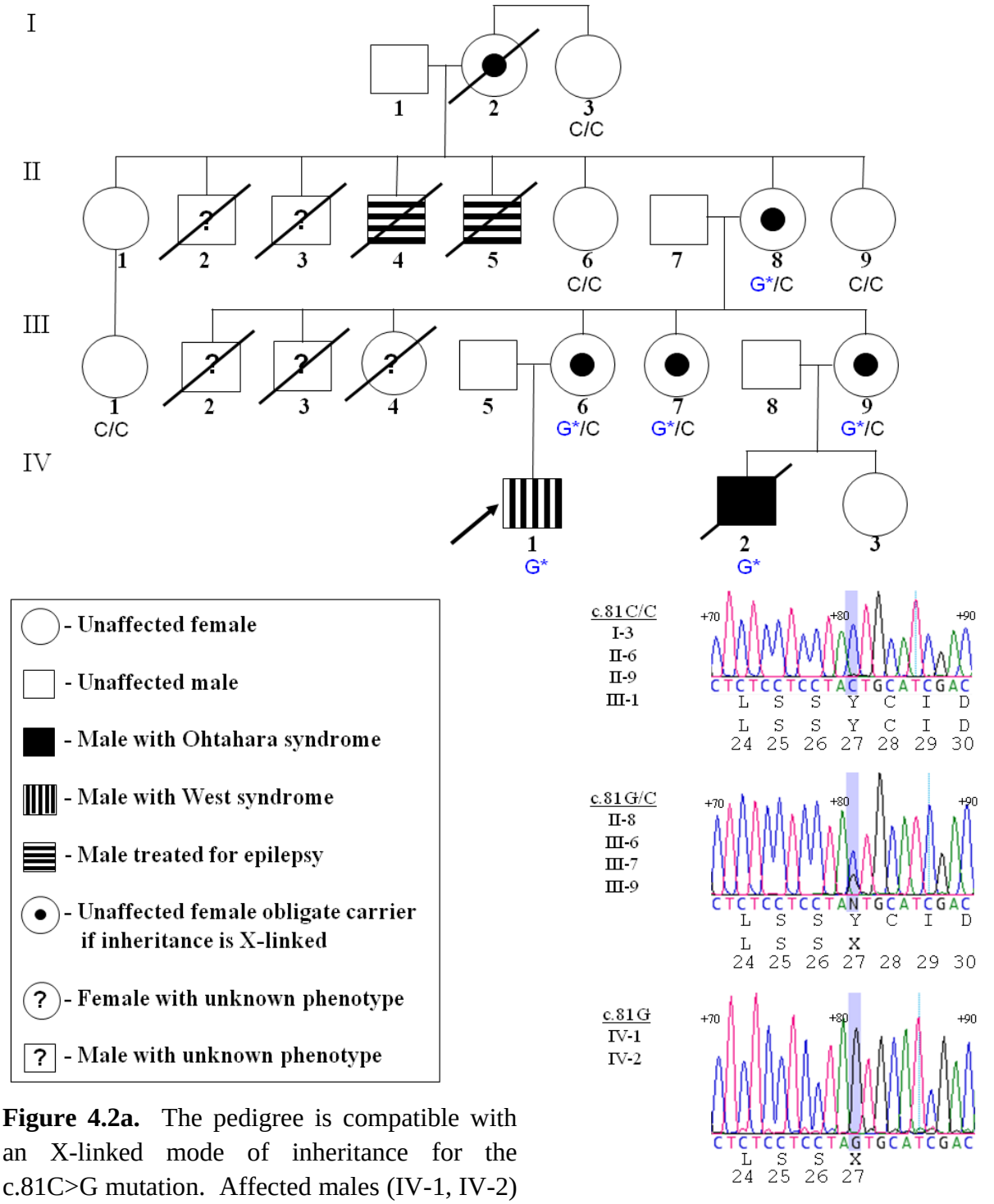

Figure 4.2a. The pedigree is compatible with an X-linked mode of inheritance for the c.81C>G mutation. Affected males (IV-1, IV-2) and obligate carrier females (I-2, II-8, III-6 and III-9) are shown together with the proband (IV-1) designated by an arrow. Individuals II-4 and II-5 are presumed to have been affected. Representative sequence traces are shown for all of the individuals tested, homozygous *ARX* c.81C/C wildtype females (I-3, II-6, II-9, III-1), heterozygous *ARX* c.81G/C carrier females (II-8, III-6, III-7, III-9), and hemizygous *ARX* c.81G males (IV-1, IV-2). Genotype at coding base 81 of *ARX* is shown below symbols.

Here we report a family with two male cousins (IV-1 and IV-2, Figure 4.2a), who were diagnosed with West syndrome and Ohtahara syndrome, respectively. There is an additional family history of stillbirth, perinatal death and epilepsy in males, compatible with X-linked inheritance.

Clinical notes – provided by Drs. Louise Brueton and Jenny Morton

Drs. Louise Brueton and Jenny Morton are from the Clinical Genetics Unit, Birmingham Women's Hospital, Edgbaston, Birmingham, UK.

The pedigree of this family is provided in Figure 4.2a. The proband IV-1 is the maternal first cousin of IV-2, as their mothers are sisters. Both mothers are of normal intelligence, in good health and have no history of seizures. There are no surviving males in their mother's or grandmother's generations. However, two maternal uncles (III-2 and III-3) died perinatally and the maternal grandmother had four male siblings, two of whom (II-2 and II-3) died shortly after birth. The other two (II-4 and II-5), who were epileptic, died in childhood. A maternal aunt (III-4) was also stillborn for unknown reasons.

The proband (IV-1) was born at term by normal vaginal delivery weighing 3.1 kg and was noted to have a left congenital cataract. He presented with focal seizures at the age of four weeks. On examination, no other abnormality was noted apart from the cataract. His genitalia were normal. At four months he developed infantile spasms and an electroencephalogram (EEG) confirmed hypsarrhythmia (Figure 4.2b). He was also found to have gallstones. At the age of five, he is profoundly delayed in all areas of development, is unable to feed orally and has remained refractory to anticonvulsant therapy. He continues to have daily seizures in the form of myoclonic jerks and occasional tonic-clonic fits. The most

recent EEG shows a severe epileptic encephalopathy with subclinical status epilepticus in sleep and frequent myoclonic jerks associated with general discharges when awake (data not shown).

Individual IV-2 was a male infant born at 42 weeks by emergency Caesarian section for failure to progress, but did not require resuscitation. Birth weight was 3.54 kg. He presented with tonic clonic seizures and myoclonic jerks on day five. An EEG at 3 weeks was grossly abnormal showing a burst suppression pattern consistent with a diagnosis of Ohtahara syndrome (Figure 4.2b), a severe early infantile epileptic encephalopathy. By 5 months of age he had axial hypotonia and hypertonia of all four limbs. His genitalia were normal. He made no developmental progress, was visually unaware and died aged eight months from intractable seizures. Cytogenetic analysis and extensive metabolic investigations were normal in both boys.

Magnetic Resonance Imaging (MRI) of the brain was performed on both patients (Figure 4.2b). The MRI brain scan of patient IV-1 at the age of 5 months showed mild cerebral atrophy, with no evidence of pachygyria or lissencephaly and normal myelination. A repeat MRI scan at the age of 21 months showed slight progression of the atrophy, an atrophic posterior corpus callosum (arrows on left image of Figure 4.2b), poor myelination and positional plagiocephaly. The MRI brain scan of patient IV-2 at the age of four weeks showed a normal gyral pattern with no evidence of pachygyria or lissencephaly and a hypoplastic corpus callosum (arrows on right image of Figure 4.2b).

Following the birth of IV-1 and having reviewed the family history (Figure 4.2a), X-linked recessive inheritance seemed likely. The possibility of an *ARX* related disorder was explored as X-linked infantile seizures and epilepsy form part of the phenotypic spectrum of *ARX*

related disorders. *ARX* gene mutation analysis was performed and an *ARX* nucleotide change was identified, c.81C>G in exon 1 of the gene in patient IV-1.

Subsequently, mutation analysis on DNA extracted from peripheral blood from family members I-3, II-6, II-8, II-9, III-1, III-6, III-7, III-9 and IV-2 (via Guthrie card) was performed (Figure 4.2a). Both affected males (IV-1 and IV-2) were found to be hemizygous for the c.81C>G mutation, while their mothers (III-6, III-9), aunt (III-7) and maternal grandmother (II-8) were shown to be heterozygous carriers. The two maternal great aunts (II-6 and II-9) were found to be homozygous for the wildtype allele. The great grandmother (I-2) was inferred to be a carrier, while her sister (I-3) and granddaughter (III-1) are homozygous for the wildtype allele.

The c.81C>G nucleotide change is predicted to cause a replacement of the tyrosine (p.Y27) codon (UAC) with a premature termination codon (PTC; UAG) leaving the ARX protein only 26 amino acids long (Figure 4.2a, 4.3b i). As such, this p.Y27X change was predicted to be deleterious and thus considered to be the mutation responsible for the phenotype of affected males within this family.

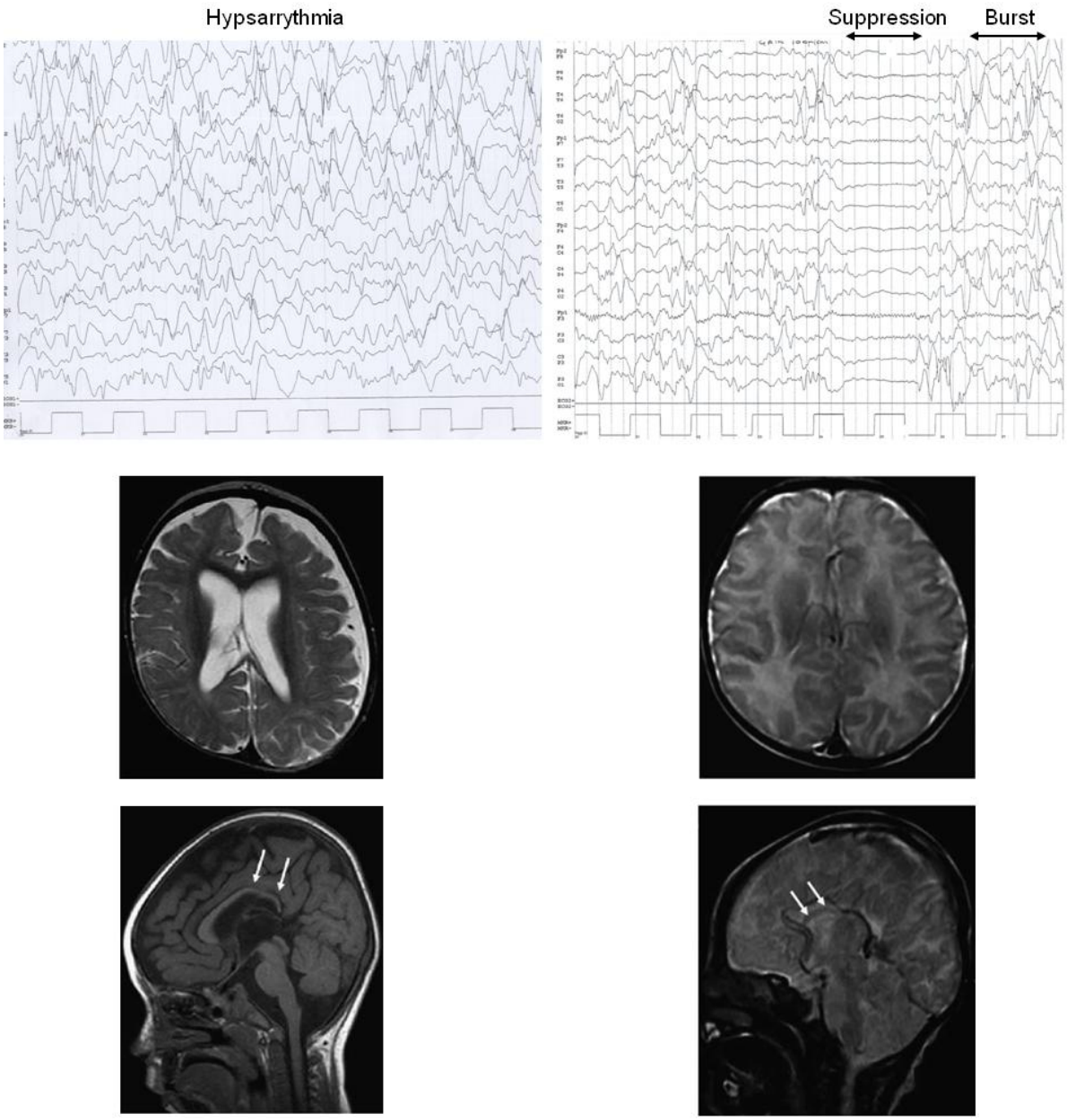

Figure 4.2b. EEG and MRI results for individuals IV-1 and IV-2 with the c.81C>G mutation. IV-1's EEG displays a hypsarrhythmia pattern and MRI brain scan at 21 months shows: axial T2 and sagittal T1 weighted images showing no lissencephaly, but cerebral atrophy, poor myelination, an atrophic posterior corpus callosum (arrows) and plagiocephaly (diagnosed with X-linked infantile spasms, ISSX; or West syndrome, WS). EEG of IV-2 is grossly abnormal showing a burst suppression pattern, consistent with a diagnosis of Ohtahara syndrome (OS). IV-2's MRI brain scan at 4 weeks displayed: axial and sagittal T2 weighted images showing normal gyral pattern and a hypoplastic corpus callosum (arrows).

4.3 Cell based studies

We initially predicted that the c.81C>G/p.Y27X mutation would cause an early protein truncation akin to a null mutation. The resultant p.Y27X ARX protein would lack all of its functional domains, including the octapeptide domain (aa 27-34), and would either not be made, be non-functional, or rapidly degraded. Other *ARX* mutations predicted to cause an early truncation of the ARX protein cause severe brain and genital malformation (XLAG) (Kitamura *et al.*, 2002; Kato *et al.*, 2004).

In this family, although minor brain abnormalities were visible on the MRI brain scans in the affected individuals (Figure 4.2b), there was a striking lack of the hallmark lissencephaly and ambiguous genitalia associated with the loss of ARX function, *ie* XLAG phenotype. As such we postulated that *ARX* mRNA translation might re-start at a more downstream AUG/Methionine codon. Only one such in-frame codon, at position c.121_123 of *ARX*, fulfils this role, suggesting that an N-terminally truncated ARX protein consisting of p.M41_C562 could in principle be made, although translational re-initiation is rarely reported phenomena (reviewed in (Kozak, 2002)). Re-initiation of translation has been established for the *ATP7A* mRNA, studied in the context of Menkes disease, where it was proven that it not only occurred but that the mutant transcript was preserved from degradation because of translational re-initiation (Paulsen *et al.*, 2006).

To test our hypothesis we used an over-expression vector with an N-terminal myc tag (*pCMV-myc-ARX-c.81C>G*). We transfected the HEK293T cells and investigated protein expression by immunfluoresence 24 h post transfection. A truncated ARX protein was detected by our polyclonal antibody raised against peptide covering residues p.161_179, of ARX, but not with an antibody against the N-terminal myc tag. This was the first evidence that the N-terminally truncated ARX protein (p.M41_C562) was indeed expressed and

localised normally (Fullston *et al.*, 2011, Appendix A.2), although it appeared to be at a greatly reduced level of expression by fluorescence intensity (data not shown). Initially to test if the N-terminally truncated ARX was less abundant than the wildtype, co-transfection studies with an empty GFP vector were performed.

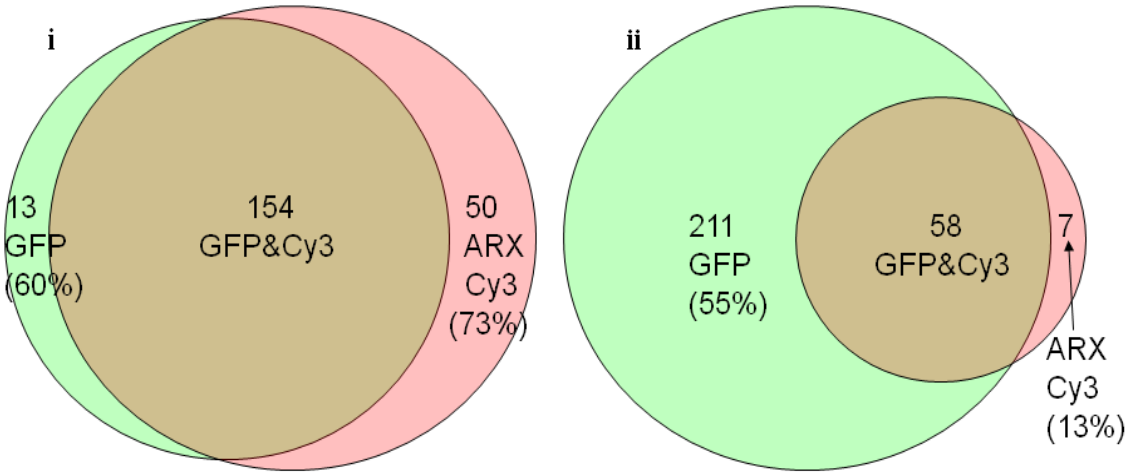

Figure 4.3a. The relative proportion of co-transfection between the *ARX* expression construct with an empty GFP vector. Transfection efficiencies for each plasmid given in parenthesis. **(i)** *pCMV-myc-ARX-wt* co-transfected with GFP; of a total of 217 cells positive for either ARX or GFP; 71% were positive for both. **(ii)** *pCMV-myc-ARX-c.81C>G* (PTC) co-transfected with GFP; of a total of 276 cells positive for either ARX (PTC) or GFP; 21% were positive for both.

We contend that low levels of ARX protein are not a result of reduced *pCMV-myc-ARX-c.81C>G* vector uptake by the HEK293T cells. This is because the co-transfected GFP vector and the *pCMV-myc-ARX-wt* construct used in parallel experiments had comparable levels of transfection of 55-73% (Figure 4.3a). The reduction of detectable protein from the c.121_123 codon translational re-initiation is consistent with previous reports (Paulsen *et al.*, 2006) of an approximate 50% reduction to that of wildtype, and likely reflects the inefficiency of translational re-initiation.

Further evidence of an N-terminally truncated ARX protein (p.M41_C562) was sought by means of western immunoblot performed on lysates of HEK293T cells transfected with either *pCMV-myc-ARX-wt* or *pCMV-myc-ARX-c.81C>G*.

An N-terminally truncated ARX was detected in lysates of HEK-293T cells transiently transfected with a vector (*pCMV-Myc*) that over-expresses the c.81C>G mutant form of *ARX* (Figure 4.3b). This truncated ARX, most likely ARX p.M41_C562, is detected by an ARX polyclonal antibody, but not by a cMYC antibody that detects the N-terminal MYC tag. Protein input was measured and 10 times more total protein extract was added to the c.81C>G lane (20 µg) compared to wt (2 µg). Given the truncated ARX protein is detected at approximately half the level of the wildtype (wt) and 10 times more protein was loaded for the c.81C>G mutant, there is approximately less than $1/20^{th}$ the amount of the re-initiation product compared to wt (by band intensity). The MYC tagged ARX p.S26 peptide could not be detected (~4kDa; data not shown). We predict that the MYC tagged ARX p.S26 peptide is either not translated or rapidly degraded, and as such could not be detected by the cMYC antibody.

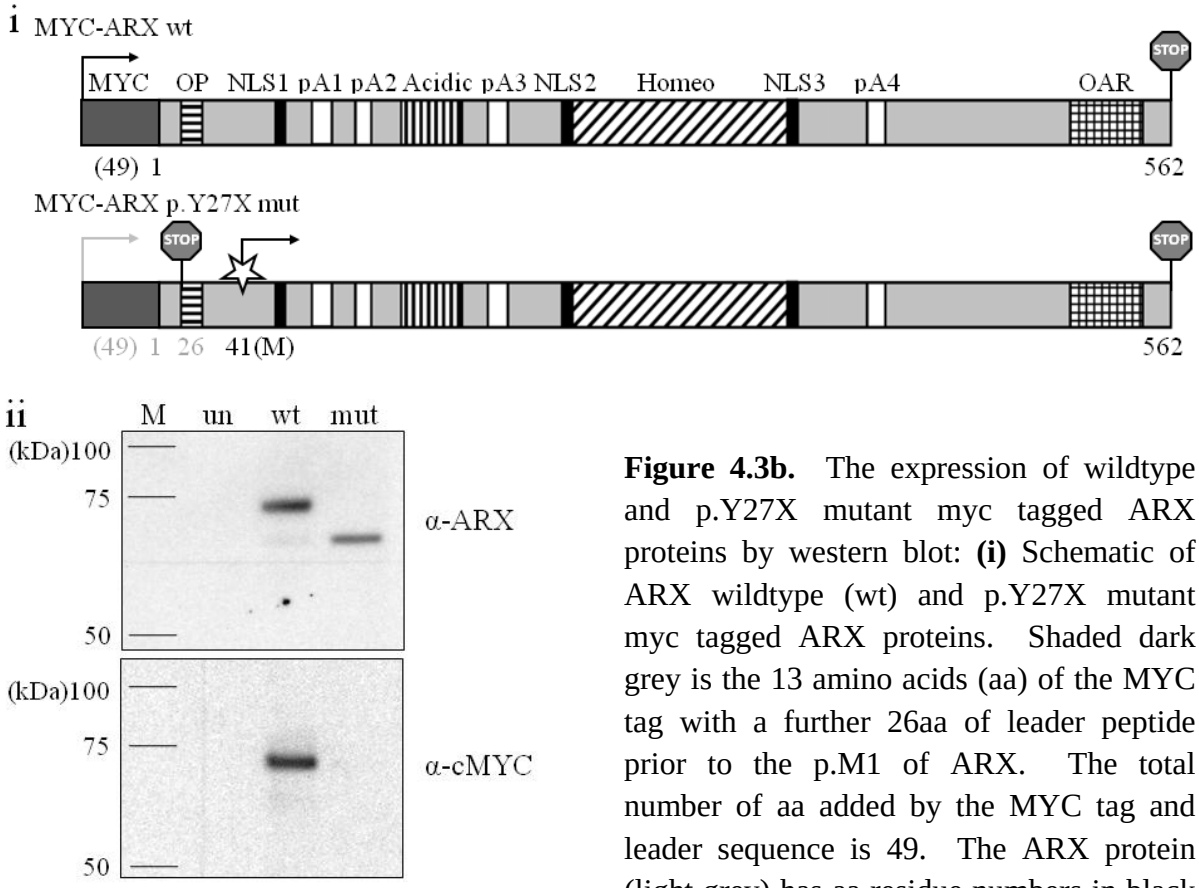

Figure 4.3b. The expression of wildtype and p.Y27X mutant myc tagged ARX proteins by western blot: **(i)** Schematic of ARX wildtype (wt) and p.Y27X mutant myc tagged ARX proteins. Shaded dark grey is the 13 amino acids (aa) of the MYC tag with a further 26aa of leader peptide prior to the p.M1 of ARX. The total number of aa added by the MYC tag and leader sequence is 49. The ARX protein (light grey) has aa residue numbers in black below each schematic and contains the following domains: OP, Octapeptide domain; NLS1-3, Nuclear localization sequence; pA1-4, polyalanine tracts; Acidic domain; Homeodomain and OAR Aristaless domain. Initiation of translation is indicated by an arrow above the protein schematics. A premature N-terminal termination codon is introduced by the c.81C>G mutation, as represented in the lower protein schematic. The star indicates the predicted residue (p.M41) at which re-initiation of translation most likely occurs. The predicted protein sizes that result from over expression from a CMV promoter in transiently transfected HEK-293T are: MYC-ARXwt – 611 aa (62 kDa), MYC-ARX(p.M1_S26) – 75 aa (4 kDa), ARX(p.M41_C562) – 522 aa (53.7 kDa). **(ii)** SDS-PAGE Western Immunoblot of *pCMV-Myc-ARX* transfected HEK-293T cell lysates. Lanes were loaded as follows: **M** - Marker 10 µL of Precision Plus protein standards, **un** – 20 µg of mock transfected cells, **wt** – 2 µg of *pCMV-Myc-ARXwt* transfected cells, **mut** – 20 µg of *pCMV-Myc-ARXc.81C>G* transfected cells. The same membrane was first probed with the ARX polyclonal antibody, stripped, then re-probed with the cMYC antibody.

4.4 Discussion

The c.81C>G/p.Y27X *ARX* mutation identified in this family is predicted to cause premature protein truncation, which should be akin to a null mutation. The resultant p.Y27X ARX protein would lack all of its functional domains, including the octapeptide domain (aa 27-34), and would either not be made, be non-functional, or rapidly degraded. Of the other *ARX* mutations that truncate the *ARX* ORF early, virtually all cause the XLAG phenotype (Kitamura *et al.*, 2002; Kato *et al.*, 2004). XLAG is most severe malformation end of the phenotypic spectrum of *ARX*-associated disorders (Kato *et al.*, 2004; Bhat *et al.*, 2005) resulting from *ARX* null mutations, that phenotypically mimic the *Arx* mouse knockout model (Kitamura *et al.*, 2002).

In the family studied there was no XLAG phenotype usually associated with *ARX* null mutations, as the c.81C>G (p.Y27X) mutation ought to be, although minor brain abnormalities were visible in affected individuals (Figure 4.2b). After being able to detect the N-terminally truncated ARX protein generated from the over-expression of the p.Y27X mutant, we speculated that *ARX* mRNA translation re-started at a more downstream AUG/Methionine codon. Only one such in-frame codon, at position c.121_123 of *ARX*, fulfilled this role, suggesting that the N-terminally truncated ARX protein consists of p.M41_C562. Re-initiation of mRNA translation is a known, although still poorly understood, phenomenon (for review see (Kozak, 2002)). Translation re-initiation may cause individuals with the c.81C>G (p.Y27X) mutation to have a relatively 'mild' non-malformation phenotype, as observed in this family, consistent with milder phenotypes observed for individuals with Menkes disease and re-initiation prone mutations in *ATP7A* (Paulsen *et al.*, 2006).

However, this explanation is subject to the proposition that the *in vivo* c.81C>G containing *ARX* mRNA escapes degradation by the nonsense mediated decay (NMD) pathway. We predict that NMD partially degrades the c.81C>G *ARX* mRNA, as is seen with other PTC containing mRNAs (Maquat, 2004), and is most likely protected from complete degradation due to the proximity of the PTC to the translation initiation codon (Silva *et al.*, 2008), or the reinitiation process itself (Paulsen *et al.*, 2006). Since tissues that express testable levels of *ARX* mRNA or protein cannot be ethically sampled from living patients we were unable to test *ARX* mRNA or protein expression *in vivo*.

Provided that *ARX* mRNA translation does re-initiate at codon c.121_123 (aa 41 of wildtype ARX), such an N-terminally truncated protein would completely lack the octapeptide domain (aa 27-34). This domain is an important transcription co-repression domain of ARX, and is involved in binding to the Groucho/transducin-like enhancer of split (TLE) co-factor proteins (McKenzie *et al.*, 2007). The p.L33P mutation of the octapeptide domain is known to give a nsXLID phenotype in the MRX54 family (Bienvenu *et al.*, 2002), but perhaps a more severe phenotype would have been observed if the entire octapeptide was absent. Combined with the absent octapeptide domain, the re-initiation peptide produced (p.M41_C562) is far less abundant. This ARX protein deficiency probably also contributes to the observed pathology in this family, *ie* OS/WS with minor brain abnormalities. A reduction of functional ARX protein has previously been suggested as a mechanism that precipitates phenotypes in female carriers of presumably null mutations, with an approximate 50% reduction of functional *ARX* locus due to random X-inactivation (Turner *et al.*, 2002; Poirier *et al.*, 2005).

In summary, we speculate that the lack of a severe malformation phenotype observed for these patients could be explained by a partially functioning, N-terminally truncated ARX protein (p.M41_C562). This truncated ARX protein appears to be less abundant than the

wildtype (Figure 4.3a & b), probably due to partial NMD degradation of the PTC containing *ARX* mRNA, and the inefficiency of translation re-initiation from the c.121_123 AUG codon.

Therefore it appears as though approximately ~5% of an N-terminally truncated ARX protein (p.M41_C562) is sufficient to 'rescue' these patients from a more severe XLAG phenotype expected for an early truncation of the ARX protein. However, this interpretation is based on over-expression of a mutant cDNA construct in a cellular environment somewhat removed from the genomic context of where the *ARX* gene is transcribed/translated *in vivo*.

Another possible mechanism behind the partial phenotypic rescue of these patients is RNA editing of the c.81C>G containing *ARX* mRNA, which would allow translation of a full length *ARX* mRNA (reviewed in (Mattick and Mehler, 2008)). However, this could not be tested due to inaccessibility of patient material that would be needed for such experiments.

Investigations did not reveal an underlying cause for the cataract or gallstones in patient IV-1, neither of which has been described previously in patients with an *ARX* gene mutation. In the absence of evidence to the contrary, although likely unrelated to the *ARX* mutation, it remains possible that these phenotypes are rare novel features of this *ARX* mutation (c.81C>G).

OS and XLAG have many features in common such as early onset during infancy, frequent intractable seizures, abnormal burst-suppression EEG patterns, hypotonia, impaired motor function and profound developmental delay. Mutations in *ARX* have been reported to account for up to 94% of XLAG (Kato *et al.*, 2004), and be responsible for X-linked West syndrome (Stromme *et al.*, 2002a,b). Clearly, impaired function of the ARX protein can also result in OS. Perhaps this finding is not surprising given the clinical overlap between XLAG and OS, and the progression of OS into WS.

In addition, this family demonstrates intrafamilial clinical heterogeneity, indicating the influence of different genetic background and environments on the clinical expression of a given *ARX* gene mutation. Specifically in this case, we suggest that subtle differences in genetic background and environment affected the extent to which NMD or translation re-initiation occurred in the two affected individuals, leading to the different clinical manifestations. This was only the second report of an *ARX* gene mutation causing Ohtahara syndrome and the first with an inherited mutation (Fullston *et al.*, 2010, Appendix A.6). Interestingly the cases reported by Kato *et al* (2007) both had a 33 bp duplication in exon 2 of *ARX* rather than a protein truncating mutation. Further, all other reported *ARX* mutations that cause OS either expand pA1 (Absoud *et al.*, 2010), mutate a residue in the aristaless domain (Giordano *et al.*, 2010) or truncate the ARX protein near the C-terminus (within the aristaless domain (Kato *et al.*, 2010; Eksioglu *et al.*, 2011)). This supports the conclusion that modified function rather than complete loss of function of the ARX protein is the cause of OS in our family.

This mutation represents the first ARX protein truncation mutation causing non-malformation phenotype and points towards the complexities of the interplay between the mutation, molecular mechanism, genetic background and possibly environmental influences. It also reminds us of the complexities involved in attempting to draw conclusions about genotype-phenotype correlation. Based on these and similar published findings we suggest that *ARX* screening should be considered in males affected with Ohtahara syndrome.

Affect of Variations in the Length of Polyalanine Tract 1 on ARX Function

CHAPTER 5 CONTENTS

Part of the work presented in this chapter has been published in Fullston *et al* (2011; Appendix A.2).

5 EFFECT OF VARIATIONS IN THE LENGTH OF POLYALANINE TRACT 1 ON ARX FUNCTION

This chapter presents segregation testing and cell based protein assessments. Segregation analysis was carried out for the c.305_307del(3bp), c.305_313del(9bp) and c.327_341del(15bp) *ARX* sequence variations. Cell based protein assessments were performed for the c.305_313del(9bp) (p.A111delA_3) sequence variation along with other selected polyalanine tract 1 (pA1) expansions. These included c.304ins$(GCG)_1$ (p.A111insA_1), c.304ins$(GCG)_7$ (p.A111insA_7) and c.298_330dup (p.A111insA_{11}). DNA samples were collected and provided by collaborators as listed in the following sub-sections. Cell based studies form a minor component of Fullston *et al* (2011) (Appendix A.2).

5.1 Introduction

Mutations that expand polyalanine tract 1 (pA1) of ARX are the second most frequent cause of *ARX* related disorders (see Figure 3.1a) and include the c.304ins$(GCG)_7$ (p.A111insA_7) and c.298_330dup (p.A111insA_{11}) mutations. The c.429_452dup(24bp) mutation is the most common and leads to an expansion of polyalanine tract 2 (pA2) (reviewed by (Shoubridge *et al.*, 2010a, Appendix A.1)). To date 12/110 known families with *ARX* mutations (Gecz *et al.*, 2006; Giordano *et al.*, 2010; Kato *et al.*, 2010; Shoubridge *et al.*, 2010a, Appendix A.1; Conti *et al.*, 2011a; Cossee *et al.*, 2011; Eksioglu *et al.*, 2011; Fullston *et al.*, 2011, Appendix A.2), have been reported with the c.304ins$(GCG)_7$ mutation in pA1 (Stromme *et al.*, 2002b; Wohlrab *et al.*, 2005; Guerrini *et al.*, 2007; Poirier *et al.*, 2008; Wallerstein *et al.*, 2008; Shinozaki *et al.*, 2009; Cossee *et al.*, 2011).

The majority of affected males in 6/12 families with the c.304ins$(GCG)_7$ mutation display phenotypes of severe intellectual disability, generalized dystonia (increased muscle tone with

dystonic posturing of limbs), infantile spasms, and a hypsarrhythmic EEG pattern (Stromme *et al.*, 2002a,b; Wohlrab *et al.*, 2005; Guerrini *et al.*, 2007; Poirier *et al.*, 2008; Wallerstein *et al.*, 2008; Absoud *et al.*, 2010; Cossee *et al.*, 2011). A further four of these families have infantile epileptic–dyskinetic encephalopathy (Guerrini *et al.*, 2007). Of the remaining two families, one presented with psychomotor delay, epilepsy (generalized tonic–clonic seizures), and generalized dystonia but in the absence of infantile spasms (Shinozaki *et al.*, 2009). The remaining family presented with early infantile epileptic encephalopathy with suppression-burst pattern (EIEE) otherwise known as Ohtahara syndrome (OS) (Absoud *et al.*, 2010). A second mutation that causes OS and expands pA1 tract length, c.298_330dup, was discovered in two separate families and was the first *ARX* mutation reported to cause OS (Kato *et al.*, 2007).

Whilst the c.304ins$(GCG)_7$ mutation expands pA1 tract length by 7 alanines (A) from 16A (wt) to 23A, the c.298_330dup mutation expands pA1 by 11A from 16A to 27A, the longest expansion of any ARX pA tract (Kato *et al.*, 2007). Although both mutations can cause OS, the c.304ins$(GCG)_7$ (p.A111insA_7) mutation does so in only 1/12 families, while causing a relatively more mild phenotype in the remaining 11/12 families. The c.298_330dup (p.A111insA_{11}) mutation consistently caused OS, albeit in only two families. Thus the c.298_330dup mutation has been suggested to cause a more severe phenotype (Shoubridge *et al.*, 2010a, Appendix A.1; Fullston *et al.*, 2011, Appendix A.2), which may be due to the greater pA1 tract length. This suggests a hypothesis that the longer a mutant pA1 tract is, the more severe the resultant phenotype is likely to be.

Other smaller expansions of pA1, such as c.304ins$(GCG)_1$ (p.A111insA_1) (Gronskov *et al.*, 2004), c.304ins$(GCG)_2$ (p.A111insA_2) (Bienvenu *et al.*, 2002) and c.304ins$(GCG)_3$ (p.A111insA_3) (Gronskov *et al.*, 2004) have been reported. The c.304ins$(GCG)_1$ variant was discovered in two XLID families, but did not segregate with the disease in one family and

was also found in 1/188 control samples, thus it was postulated to represent a benign rare polymorphism (Gronskov *et al.*, 2004), although variable penetrance remains as a possible explanation and as such this variation might be pathogenic or a susceptibility allele. Segregation testing within the family with the c.304ins$(GCG)_2$ showed the variant was in a brother pair with similar phenotypes of moderate ID, one brother with epilepsy and the other with aggressive behaviour (Bienvenu *et al.*, 2002). But whether their mother carried the c.304ins$(GCG)_2$ variant was not reported and full clinical descriptions of family members (family 'T80') remains unpublished (Bienvenu *et al.*, 2002). Therefore, the c.304ins$(GCG)_2$ sequence variant is possibly pathogenic. No family members, other than the proband, were available for testing the presence of the c.304ins$(GCG)_3$ variant (Gronskov *et al.*, 2004), so the pathogenicity of this variation is inconclusive. However, given that a 2A expansion of pA1 is pathogenic (Bienvenu *et al.*, 2002), than a slightly larger 3A expansion is also likely to be pathogenic. Further, neither the 2A or 3A expansion has been reported in unaffected individuals.

As part of our effort with *ARX* mutation screening (Chapter 3) we discovered deletions in the DNA that codes for pA1 of ARX; c.305_313del (p.A111delA_3) and c.327_341del (p.A111delA_5), which we reported as non-pathogenic (Fullston *et al.*, 2011, Appendix A.2). The sequence variants either did not segregate (c.305_313del) or other family members were not available for segregation testing (c.327_341del). The non-pathogenic nature of a 24bp deletion reported in pA2 (Bienvenu *et al.*, 2002; Gronskov *et al.*, 2004; Conti *et al.*, 2011b) also suggests that, in general, ARX pA tract contractions of less than 8A are likely to be non-pathogenic. Segregation testing of the c.305_313del (p.A111delA_3) and the c.305_307del (p.A111delA) sequence variation is presented here. Cell based recombinant protein subcellular localisation assays for one pA1 contraction c.305_313del (p.A111delA_3) and

selected expansions (c.304ins$(GCG)_1$ (p.A111insA_1), c.304ins$(GCG)_7$ (p.A111insA_7) and c.298_330dup (p.A111insA_{11})) were also undertaken.

5.2 Small in frame deletions in pA1 are rare variants

Two probands from separate families were identified to have either a 9 bp (Figure 5.2a) or 15 bp in frame deletions in the DNA that codes for pA1 (c.305_313del and c.327_341del) (Table 3.5a – Chapter 3 and (Fullston *et al.*, 2011, Appendix A.2)). These deletions are predicted to result in a pA1 contraction of 3A and 5A, respectively.

Limited clinical and family details were available for the female proband with infantile spasms discovered to carry the 15 bp deletion (c.327_341del) and male relatives were not available for clinical investigations. Affected female carriers of an *ARX* mutation have been reported (Turner *et al.*, 2002; Partington *et al.*, 2004; Wallerstein *et al.*, 2008; Eksioglu *et al.*), generally with mild or no clinical presentations when carrying mutations that cause non-malformation phenotypes in affected males (reviewed by (Gecz *et al.*, 2006; Shoubridge *et al.*, 2010a, Appendix A.1)). Skewed X-inactivation has been demonstrated to not correlate with the affected status of females with a polyalanine tract expansion (Stromme *et al.*, 2002b; Marsh *et al.*, 2009). Typically, affected carrier females are most frequently reported for mutations that present as severe malformations in affected males (Proud *et al.*, 1992; Bonneau *et al.*, 2002; Scheffer *et al.*, 2002; Kato *et al.*, 2004; Wallerstein *et al.*, 2008). Predominantly only males are reported to be affected by *ARX* mutations, which may be in part due to ascertainment bias (Marsh *et al.*, 2009). Regardless, little or no clinical description is reported for female carriers of variations that are of questionable pathogenicity. Taken together with the unavailability of DNA samples from further family members and that no male relatives were available for clinical investigations, the 15 bp deletion (c.327_341del) was not investigated further, despite being the largest pA1 deletion reported.

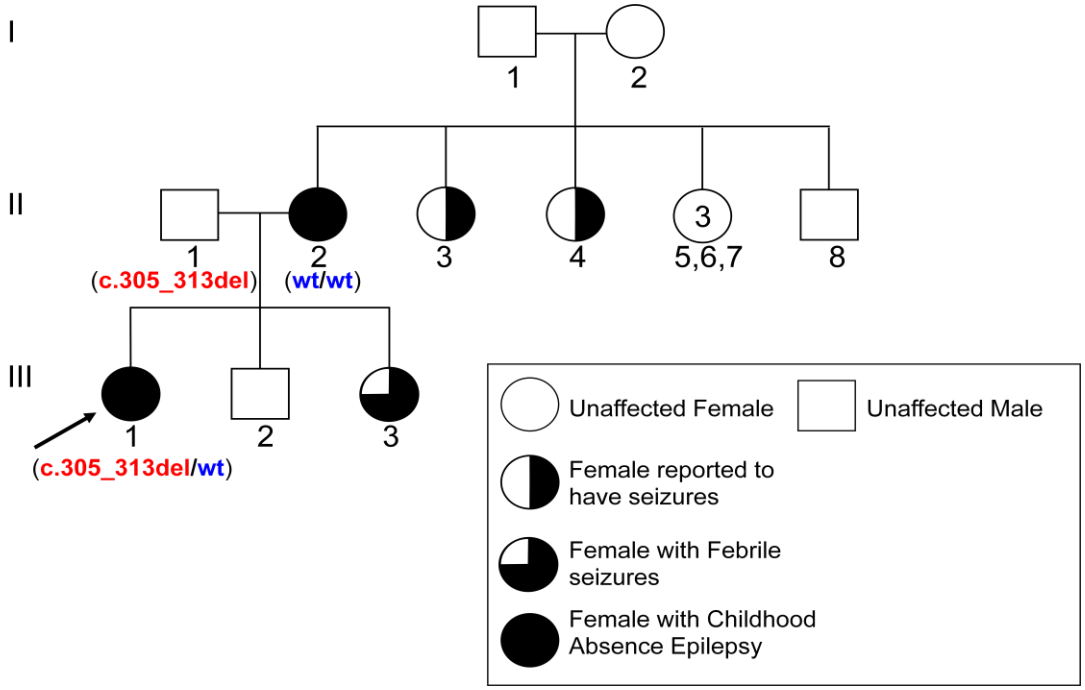

Figure 5.2a. The pedigree of the family with the 9 bp in frame deletion (c.305_313del) in the DNA that codes for pA1 of *ARX*. The proband (III-1) is highlighted by an arrow and the various classifications of seizures and epilepsy are delineated by symbols in the pedigree key. *ARX* genotype is given below pedigree symbols that represent individuals II-1, II-2 and III-1.

DNA samples and clinical information for the family with the c.305_313del sequence variation (Figure 5.2a) was supplied by Profs. Sam Berkovic and Ingrid Scheffer from the Epilepsy Research Centre, at the University of Melbourne. The father (II-1; Figure 5.2a) of the proband (III-1) had no history of epilepsy or seizures but had the 9 bp deletion (c.305_313del). The proband's mother (II-2), who was affected with the same childhood absence seizures as the proband, did not carry the deletion. The presence of the deletion in a healthy male (father; II-1), absence of it in an affected female (mother; II-2) and considering that female phenotypes are rare for *ARX* mutations that cause non-malformation pathologies in males. Taken together this suggests the 9 bp deletion (c.305_313del) is not the cause of the pathology observed within this family. Therefore, there is an unrelated history of seizures from the maternal side of the pedigree.

A third family with a 3 bp deletion (c.305_307del) in multiple males in a large unreported pedigree was referred by Prof. Charles Schwartz for segregation testing.

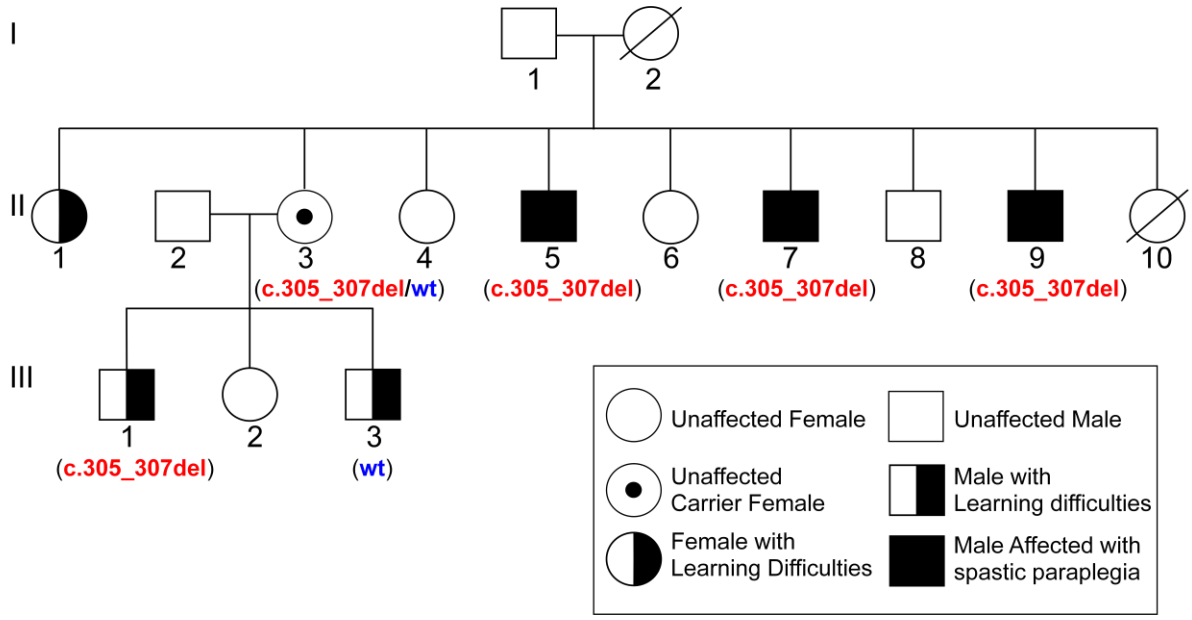

Figure 5.2b. The pedigree of the family with a 3 bp in frame deletion (c.305_307del) in the DNA that codes for pA1 of *ARX*. In the 2nd generation affected males have spastic paraplegia, whilst females were either unaffected or had learning difficulties. Males in the third generation also presented with learning difficulties. *ARX* genotype is given below pedigree symbols that represent individuals that were sequence tested (II-3, II-4, II-5, II-7, II-9, III-1 and III-3).

Although the 3 bp deletion (c.305_307del) segregates with spastic paraplegia in the 2nd generation of this family (Figure 5.2b), only one of two males with learning difficulties in the 3rd generation have the deletion. III-3 has wildtype *ARX* sequence and has learning difficulties akin to his brother (III-1) who is hemizygous for the 3 bp deletion. Neither individual III-1 or III-3 had spastic paraplegia associated with the 3 bp deletion (c.305_307del), as observed in the 2nd generation. The findings in the 3rd generation cannot rule out the pathogenic nature of this deletion, as III-3 could be explained as a phenocopy, given the frequency of learning difficulties in the population. It seems unlikely that a single

1A contraction of pA1 would result in the severe phenotype, *ie* spastic paraplegia, as other deletions of 9 bp in pA1 (above) and of 24 bp in pA2 are not pathogenic (Bienvenu *et al.*, 2002; Gronskov *et al.*, 2004; Conti *et al.*, 2011b). This suggests that the 3 bp (c.305_307del), 9 bp (c.305_313del) and 15 bp (c.327_341del) deletions are rare benign variants not previously reported and not likely pathogenic.

Given the lack of clinical and family data for the 15 bp deletion and the presumably non-pathogenic 1A retraction of pA1 caused by the 3 bp deletion, the 9 bp deletion was selected for further investigation by means of cell based localisation assays in section 5.5.

5.3 Expansion of pA1 tract

Besides the putative non-pathogenic deletions presented above, insertions in the DNA that code for pA1 of ARX have been reported. These in frame insertions are predicted to expand pA1 by 1A per extra inserted trinucleotide repeat. Five different size pA1 expansions have been described: (i) c.304ins$(GCG)_1$ (p.A111insA_1) – 17A pA1 (Gronskov *et al.*, 2004); (ii) c.304ins$(GCG)_2$ (p.A111insA_2) 18A pA1 (Bienvenu *et al.*, 2002); (iii) c.304ins$(GCG)_3$ (p.A111insA_3) 19A pA1 (Gronskov *et al.*, 2004); (iv) c.304ins$(GCG)_7$ (p.A111insA_7) 23A pA1 (Stromme *et al.*, 2002a,b; Wohlrab *et al.*, 2005; Guerrini *et al.*, 2007; Poirier *et al.*, 2008; Wallerstein *et al.*, 2008; Shinozaki *et al.*, 2009; Cossee *et al.*, 2011); and (v) c.298_330dup (p.A111insA_{11}) 27A pA1 (Kato *et al.*, 2007).

To address whether these naturally occurring pA1 tract expansions impact on protein localisation we interrogated the respective full length recombinant ARX proteins in cell based studies.

5.4 The length of pA1 tract aligns with phenotypic severity and degree of mutant ARX protein mislocalisation

Four different pA1 constructs (wildtype – 16A, c.305_313del – 13A, c.304ins$(GCG)_1$ – 17A, c.304ins$(GCG)_7$ – 23A and c.298_330dup – 27A) (derived from the *pCMV-myc-ARX-wt* vector; Chapter 2.8) were transfected into human embryo kidney cells (HEK-293T; product number CRL-11268, ATCC, Manassas, USA) (Chapter 2.9) and subjected to fluorescent immunohistochemistry (Chapter 2.10) and standard microscopy (2.10 and 2.11). An average of 492 ARX positive cells were scored per transfection across three separate transfection experiments. Localisation was classified as either normal, having nuclear inclusions, or cytoplasmic positive with or without aggregates in either the nucleus or cytoplasm (see representative images in Figure 5.4a) and compared to a wildtype *ARX* construct (16A; Figure 5.4b and Table 5.4a).

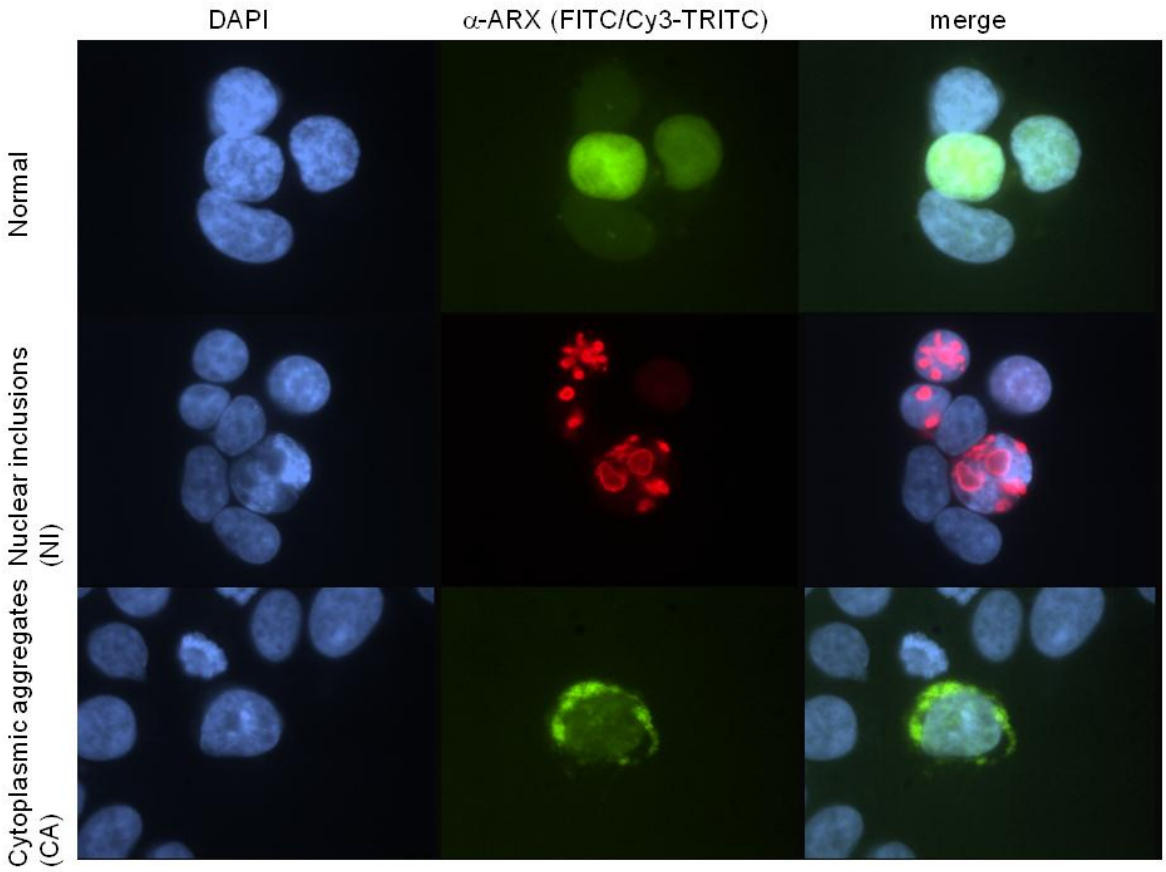

Figure 5.4a. Representative images for the sub-classification of ARX protein localisation. Normal ARX localisation is shown in the top three panels as a diffuse nuclear stain (*pCMV-Myc-ARX* wt vector). Sub types of mislocalisation observed for each ARX protein construct tested were classified as nuclear inclusions only (NI; middle three panels using the *pCMV-Myc-ARX* c.1136G>T/ p.R379L vector) or cytoplasmic positive with or without aggregates in either the nucleus or cytoplasm (CA; representative image of cytoplasmic aggregates shown in the bottom three panels using the *pCMV-Myc-ARX* c.304ins$(GCG)_7$ p.A111insA_7vector).

Only pA1 expansions reported as pathogenic (*ie* c.304ins$(GCG)_7$ - 23A and c.298_330dup - 27A) displayed a significant increase of ARX protein mislocalistion (Table 5.4a and Figure 5.4b). Expansions of non-pathogenic or those of uncertain pathogenic effect (*ie* c.305_313del – 13A and c.304ins$(GCG)_1$ – 17A) had no effect on ARX protein mislocalisation. Although the c.305_313del (13A) and c.304ins$(GCG)_1$ (17A) expression constructs repeatedly increased ARX protein misclocalisation, it was not significant by statistical analysis (as per Figure 5.4b and Table 5.4a).

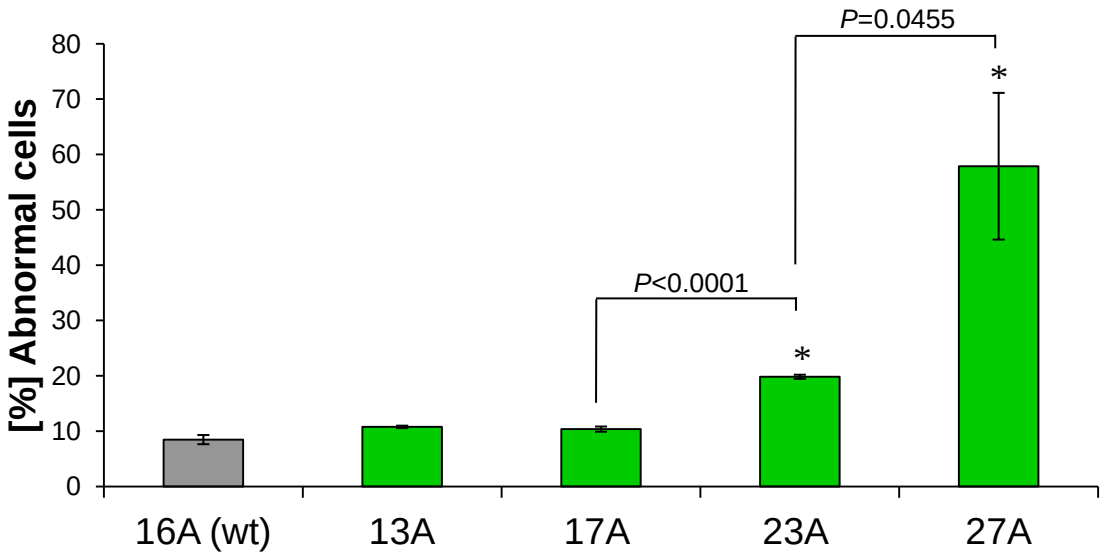

Figure 5.4b. Proportion of ARX positive cells displaying abnormal localisation (nuclear inclusions or cytoplasmic positive with or without aggregates in either the nucleus or cytoplasm) for four pA1 tract lengths (Green bars: c.305_313del – 13A; c.304ins$(GCG)_1$ – 17A; c.304ins$(GCG)_7$ – 23A; and c.298_330dup – 27A) compared to wildtype ARX (Grey bar; 16A). 492 ARX positive cells were scored per transfection for three separate transfection experiments 24 h post transfection. * Asterisks indicate which constructs had significantly different levels of protein mislocalisation compared to wildtype ($P \leq 0.05$; actual *P* values can be found in Table 5.4a) and between mutant *P* values are indicated (when $P \leq 0.05$), using a general linear model and an unbalanced ANOVA analysis.

The mislocalisation observed for the disease causing expansion mutations in pA1 was restricted to the cytoplasm with or without aggregates in either the nucleus or cytoplasm but not as nuclear inclusions alone (Table 5.4a), consistent with previous reports for the c.304ins$(GCG)_7$ mutation (Nasrallah *et al.*, 2004; Shoubridge *et al.*, 2007).

Although both of the c.304ins$(GCG)_7$ (23A) and c.298_330dup (27A) mutations can cause OS (Kato *et al.*, 2007; Absoud *et al.*, 2010), the c.304ins$(GCG)_7$ can be considered as causing a milder phenotype because in 11 out of 12 reported cases it does not cause OS and only in 1/12 cases it does. In contrast both reported cases with the c.298_330dup have OS. Therefore the magnitude of mislocalisation of mutant protein aligns with both phenotypic severity and pA1 tract length.

Table 5.4a. Amount of ARX protein mislocalisation per pA1 expression construct (*pCMVmycARX*) 24h post transfection (tfxn).

Construct	Protein	pA1 tract length	Abnormal localisation ±SEM (%)	*P* value (vs wt)[1]	NI [2] ±SEM (%)	*P* value (vs wt)[1]	CA [2] ±SEM (%)	*P* value (vs wt)[1]	*n*[3]	Average Cell count (per tfxn)	Phenotype
Wildtype (wt)	-	16A	8.47 ±0.64	-	6.03 ±0.59	-	2.44 ±0.05	-	9	486	None
c.305_313del	p.A111delA_3	13A	10.77 ±0.22	0.9929	6.06 ±2.58	0.9700	4.71 ±2.69	0.1480	3	458	Epilepsy/None
c.304ins$(GCG)_1$	p.A111insA_1	17A	10.36 ±0.47	0.9966	6.33 ±0.21	0.7258	4.03 ±0.78	0.0582	3	680	nsXLID/None
c.304ins$(GCG)_7$	p.A111insA_7	23A	**19.82 ±0.38**	**0.0281**	**2.02 ±0.26**	**0.0063**	**17.8 ±0.12**	**<0.0001**	3	362	ISSX+/OS
c.298_330dup	p.A111insA_{11}	27A	**57.87 ±13.26**	**<0.0001**	**0.22 ±0.14**	**0.0003**	**57.65 ± 13.1**	**<0.0001**	3	475	OS

[1] *P* value for comparison of total amount of abnormal localisation or for each sub-classification to that of the wildtype construct by general linear model and an unbalanced ANOVA analysis. A *P* value ≤0.05 was considered significant.
[2] Sub types of mislocalisation observed for each ARX protein construct tested, classified as nuclear inclusions only (NI) or cytoplasmic positive with or without aggregates in either the nucleus or cytoplasm (CA).
[3] The number of transfection replicates performed for each construct.
Bold font indicates a significant difference to wildtype.

5.5 Discussion

Expansions of pA1 are the second most frequent mutation reported in the *ARX* gene (Stromme *et al.*, 2002a,b; Wohlrab *et al.*, 2005; Guerrini *et al.*, 2007; Poirier *et al.*, 2008; Wallerstein *et al.*, 2008; Shinozaki *et al.*, 2009; Cossee *et al.*, 2011), but contractions are rare (Fullston *et al.*, 2011, Appendix A.2) and remain of questionable pathogenecity. Three families with DNA variations that contract pA1 length were presented. Two contractions were detected by *ARX* mutation screening and were reported as presumably non-pathogenic (Fullston *et al.*, 2011, Appendix A.2) (Chapter 3) and the third was referred for segregation testing.

The paucity of information and DNA samples from the family of a female with infantile spasms that carried the largest pA1 contraction (c.327_341del) did not allow for further investigations. Carrier females are rarely reported with clinical pathologies compared to affected males, especially for mutations that cause non-malformation phenotypes in affected males, although this may also represent an ascertainment bias (Marsh *et al.*, 2009). As such the infantile spasms observed in this female may partly result from this deletion. It remains inconclusive if the pA1 (c.327_341del – p.A111delA_5) retraction of 5A to 11A is pathogenic.

Testing in two further families with a 13A (c.305_313del) or 15A (c.305_307del) pA1 of ARX (wt – 16A) revealed that these variations did not segregate with disease. Together with the normal localisation of the 13A pA1 ARX protein (c.305_313del – p.A111delA_3), it suggests that contractions of pA1 are not pathogenic. Technically the 1A contraction of pA1 to 15A (c.305_307del) remains questionably pathogenic, as only one of five affected males did not carry the deletion (III-3; Figure 5.2b) and may be a phenocopy. Four of the five remaining affected males within this family did segregate with the deletion. Considering a larger 3A contraction of pA1 does not segregate with disease nor alter ARX localisation, it seems unlikely that a 1A retraction would be disease causing. Further circumstantial

evidence that a 1A contraction of pA1 is likely not pathogenic is the benign nature of an 8A contraction in pA2 (Bienvenu *et al.*, 2002; Gronskov *et al.*, 2004; Conti *et al.*, 2011b).

Mutant ARX protein mislocalised in over-expression studies as a function of pA1 tract length (Figure 5.4b), consistent with the phenotypic severity of affected males (Table 5.4a). Only disease causing pA1 expansion mutations caused significant increase in mislocalisation of mutant ARX protein. The caveat of these localisation studies is that the over-expression system used (*ie* the CMV promoter) caused ARX protein with mutant pA1 to aggregate, and may not detect subtle shifts in subcellular localisation. Although it was shown in c.304ins$(GCG)_7$ 'knock in' mice that the ARX protein mislocalises *in vivo* to the cytoplasm in 55% of interneurons throughout the cortex of adult mice, aggregates were not observed (Price *et al.*, 2009). Further, to date no evidence of mutant ARX aggregates have been found *in vivo* (reviewed in (Shoubridge and Gecz, 2011)).

Although the c.304ins$(GCG)_7$ (p.A111insA_7) and c.298_330dup (p.A111insA_{11}) mutations share OS as a phenotype, the c.304ins$(GCG)_7$ mutation does so in only one of 12 cases (Absoud *et al.*, 2010) and the c.298_330dup mutation does so in all reported cases. Given that the c.304ins$(GCG)_7$ mutation does not cause OS, but a milder phenotype in 11 of 12 cases, it can be considered as less severe. Therefore the magnitude of mislocalisation of mutant protein in an over-expression system was consistent with both phenotypic severity and pA1 tract length. Although function was not tested directly, this suggests that increasing pA1 tract length might confer increasing dysfunction to mutant ARX.

Although rare in frame deletions of questionable pathogenicity were discovered in the DNA that code for pA1 of ARX, we did not discover any new pA1 mutations as part of this study. Despite that a 3A contraction in pA1 (c.305_313del/p.A111delA_3) did not alter protein localisation in our over-expression system, other modes of ARX dysfunction and a hypothetical role for pA1 contractions in a two hit model of disease warrants further investigation.

Variations in the Length of ARX Polyalanine Tract 2

CHAPTER 6 CONTENTS

The family investigations presented in this chapter have been published in Fullston *et al* (2011), Reish *et al* (2009) and Demos *et al* (2009) (Appendix A.2, A.7 and A.8 respectively), as detailed below.

6 VARIATIONS IN THE LENGTH OF ARX POLYALANINE TRACT 2

Segregation testing, family investigations and cell based protein localisation studies of mutations in the DNA that code for ARX polyalanine tract 2 (pA2) that result in pA2 expansions are presented within this chapter. DNA samples were collected and provided by collaborators as listed in the following sub-sections and relevant publications. Segregation testing and clinical descriptions of the larger 'duplication' mutations have been published separately in Reish *et al* (2009; c.430_456dup(27bp); Appendix A.7) and Demos *et al* (2009; c.423_455dup(33bp); Appendix A.8). Three new families with the duplication mutation (c.429_452dup(24bp)) and cell based assessment of all mutant pA2 ARX proteins are published in Fullston *et al* (2011) (Appendix A.2).

6.1 Introduction

Mutations that expand pA2 tract length in ARX are the most frequent cause of pathology reported for *ARX* related disorders, all of which are caused by the c.429_452dup(24bp) mutation (reviewed by (Gecz *et al.*, 2006; Shoubridge *et al.*, 2010a, Appendix A.1)). To date at least 41/105 (39%) families with an *ARX* mutation harbour the c.429_452dup(24bp) mutation (reviewed in (Shoubridge *et al.*, 2010a, Appendix A.1)) (Cossee *et al.*, 2011), not including the five families presented here. It is likely that many more unreported individuals and families with the c.429_452dup(24bp) mutation have been discovered in a diagnostic setting. As for all *ARX* mutations reported, intellectual disability (ID) is a cardinal clinical feature of all patients with this mutation. Patients are typically classified as having non-malformation phenotypes with or without additional features that are not consistent, displaying remarkable variation in clinical expressivity both within and between families

(Stromme *et al.*, 2002a; Turner *et al.*, 2002; Gestinari-Duarte Rde *et al.*, 2006; Szczaluba *et al.*, 2006; Laperuta *et al.*, 2007; Rujirabanjerd *et al.*, 2007; Cossee *et al.*, 2011).

Of the male probands with the c.429_452dup(24bp) mutation, the majority (29 of 41) are described as having non-syndromic ID (nsXLID) (Partington *et al.*, 1988; Bienvenu *et al.*, 2002; Gronskov *et al.*, 2004; Partington *et al.*, 2004; Poirier *et al.*, 2005; de Souza Gestinari-Duarte *et al.*, 2006; Nawara *et al.*, 2006; Poirier *et al.*; Laperuta *et al.*, 2007; Rujirabanjerd *et al.*, 2007). A further 7 of 41 present with moderate to severe ID in conjunction with dystonic movement of the hands and dysarthria (Partington syndrome – PRTS (Partington *et al.*, 1988; Mulley *et al.*)) (Bienvenu *et al.*, 2002; Stromme *et al.*, 2002a; Partington *et al.*, 2004; Van Esch *et al.*, 2004; Stepp *et al.*, 2005; Poirier *et al.*, 2006; Szczaluba *et al.*, 2006; Rujirabanjerd *et al.*, 2007; Cossee *et al.*, 2011), originally named Partington syndrome in 1992 (PRTS (Mulley *et al.*, 1992)). The remaining 5 of 41 families have affected individuals with the infantile seizure phenotype or X-linked infantile spasms (ISSX or West syndrome (Stromme *et al.*, 2002a,b; Kato *et al.*, 2003; Poirier *et al.*, 2008; Cossee *et al.*, 2011). The modifiers that cause the variable clinical expressivity observed for the c.429_452dup(24bp) mutation have not been identified. Three new families that carry the c.429_452dup(24bp) mutation are described here (Chapter 6.2).

The discovery of two additional duplication mutations of larger size (27 bp – Chapter 6.3; 33 bp – Chapter 6.4) provided a unique opportunity to investigate the effects of the relevant mutant ARX proteins (Chapter 6.5). Three different size duplication mutations of 24 bp, 27 bp and 33 bp that were discovered to code for pA2 expansions of 12+8A, 12+9A and 12+10A interspersed by a glycine (G) residue, respectively (Table 6.5a). Subsequently, limited analysis of the primary and secondary structure of the DNA sequences that result from the three duplications was undertaken (Chapter 6.6). Whether the c.429_452dup(24bp)

duplication mutations are recurrent or identical by descent was also investigated (Chapter 6.7).

6.2 Three new families with the 24 bp duplication mutation

As part of the mutation screening of the *ARX* ORF (Chapter 3), three new families with the c.429_452(24bp) mutation were discovered. The mutations were identified by PCR (primers 2J-2 as per Table 2.3a) and standard agarose gel electrophoresis (Chapter 2.5). Where available, DNA samples from other family members were also tested.

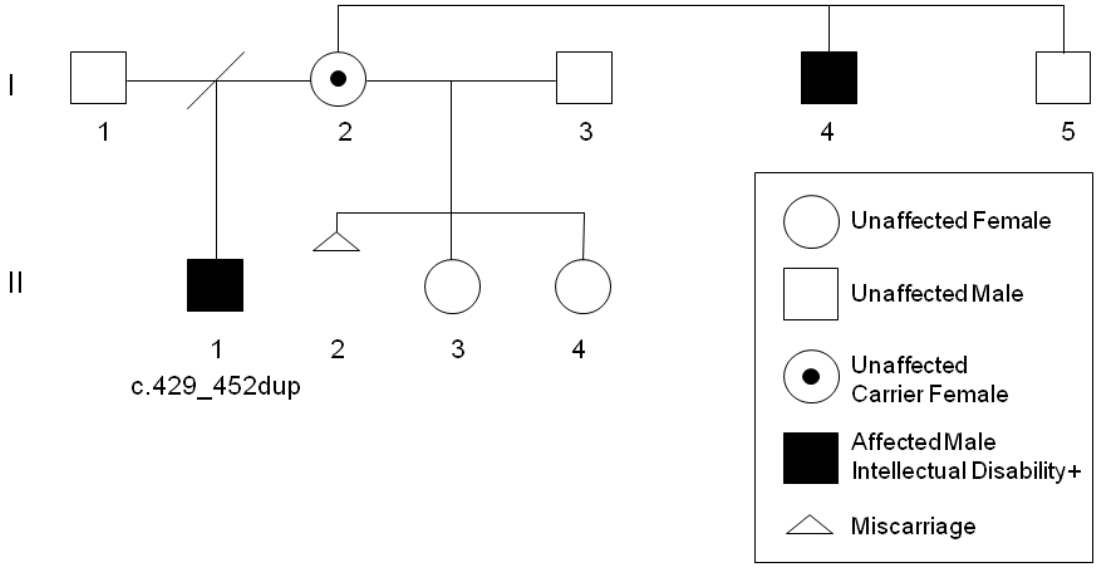

Figure 6.2a. Pedigree of family 6.2a with the c.429_452dup(24bp) duplication mutation discovered in the proband (II-1). The proband's mother (I-2) is a presumed carrier and his uncle is affected but DNA samples were not available for testing (Family A in Fullston *et al* 2011, Appendix A.2). Individuals with the c.429_452dup(24bp) mutation are indicated below the symbol that represents them.

Family 6.2a clinical notes – provided by Dr. Anna Hackett and Prof. Gillian Turner

Drs. Anna Hackett and Gillian Turner are part of the Genetics of Learning Disability Service of NSW, Hunter Genetics, Newcastle, Australia.

The proband, II-1 (Figure 6.2a), was induced at 41 weeks gestation after an uncomplicated pregnancy. His birth weight was 3020 g and head circumference was 33 cm. Hypotonia and poor feeding were present from the neonatal period. At six years of age, he was assessed as having dysarthria and moderate ID. He attended a mainstream class with support. He was a quiet and pleasant boy. A brain MRI was normal. He was seen aged 9 years of age. His height was 134.7 cm (50th centile) and head circumference was 52.2 cm (25^{th}-50^{th} centile). He was generally hypotonic, with normal deep tendon reflexes. Mild dysmorphic features included a high arched palate with dental crowding, a low set anterior hairline, a single palmar crease and inverted nipples. Dystonic movements of the elbows and wrists were noted during running, suggesting the possible diagnosis of '*ARX* syndrome' in the form of Partington syndrome (PRTS).

His maternal uncle, I-4 (Figure 6.2a), was a 'floppy baby' with delayed motor milestones. Seizures commenced at 2 years of age and temporal lobe epilepsy was diagnosed. He attended special school, lives with his parents and works in supervised employment. Examination revealed generalised hypotonia with normal reflexes and dystonic hand movements.

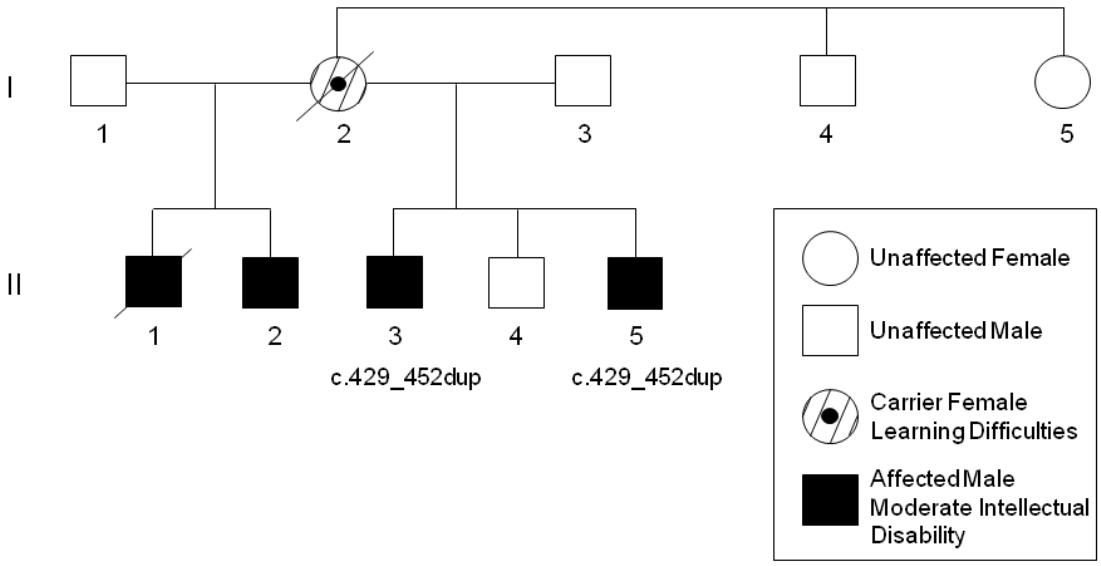

Figure 6.2b. Pedigree of family 6.2b with the c.429_452dup(24bp) duplication mutation identified in the proband (II-3) and his brother (II-5). The proband's mother (I-2) is an obligate carrier. Half brothers of the proband were also affected (II-1 and II-2) but DNA samples were not available for testing (Family B in Fullston *et al* 2011, Appendix A.2). Individuals with the c.429_452dup(24bp) mutation are indicated below the symbol that represents them.

Family 6.2b clinical notes – provided by Dr. Anna Hackett and Prof Gillian Turner

Two brothers, II-3 (proband) and II-5 (Figure 6.2b), and their two half-brothers (II-1, II-2) all had moderate ID. Their mother (I-2) was of borderline intelligence. Individual II-1 drowned as a child. None of the affected individuals suffered from seizures. The two brothers were referred to geneticists as adults, originally thought to have nsXLID. The proband, II-3 lived in a group home. He was tall and thin with normal growth parameters, not dysmorphic with minimal speech. His physical examination was otherwise normal. His brother (II-5) lives in an institute for the intellectually disabled. He was born with undescended testes and was tall with a 58.5 cm (98th centile) head circumference. *ARX* testing was performed as part of the clinical workup.

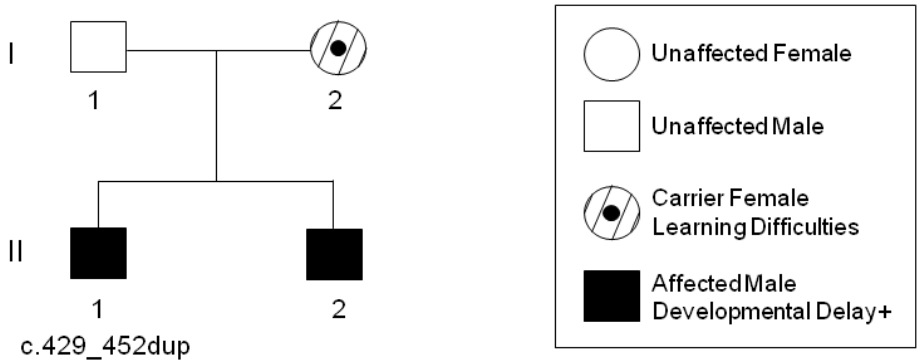

Figure 6.2c. Pedigree of family 6.2c with the c.429_452dup(24bp) duplication mutation discovered in the proband (II-1). The proband's mother (I-2) is an obligate carrier. The brother of the proband (II-2) was also affected but DNA sample was not available for testing (Family C in Fullston *et al* 2011, Appendix A.2). Individuals with the c.429_452dup(24bp) mutation are indicated below the symbol that represents them.

Family 6.2c clinical notes – provided by Dr. Gareth Baynam

Dr. Gareth Baynam works at the Genetic Services of Western Australia, Perth, Australia.

The proband, II-1 (Figure 6.2c), was born at term after prenatal onset large stature. At birth his head circumference was 38 cm, weight was 4465 g and length was 54 cm (all 98th centile). At 6 months of age he presented with developmental delay, infantile spasms and hypsarrthymia by EEG. Seizures responded well to vigabatrin. He exhibited intermittent right esotropia. A CT showed prominence of the lateral ventricles particularly surrounding the frontal horns and sylvian fissures bilaterally. At 16½ months of age he was assessed as having a global developmental quotient of 50. A review at 7 years of age revealed he had no speech capability, presenting with a long face, deep set eyes and large ears. He was required to be schooled in a full time education support unit.

Individual II-2 (Figure 6.2c) showed prenatal onset large stature, with a birth weight of 3840 g, length 55.5 cm and head circumference of 37 cm. At 23 months of age he was

developmentally delayed with a global developmental quotient of 50, the greatest delay observed in motor skills. He displayed central hypotonia with increased tone in the lower limbs and dystonic movements of the upper limbs. No overt seizures were observed and no neuroimaging was performed.

For individual I-2 (Figure 6.2c) there were concerns at school from Grade 2 about her motor abilities and word association. She had ongoing difficulties with reading and spelling, managing to complete four years of high school, but was approximately one year behind in some subjects.

In summary, three new families with four affected males that had a positive DNA test for the c.429_452dup(24bp) mutation were identified. The clinical presentations of the probands ranged from nsXLID (one proband and his brother; Family 6.2b), PRTS (Family 6.2a) to West syndrome (Family 6.2c). Two of the three families had an obligate carrier female with learning difficulties, but their carrier status was not confirmed due to refusal to participate in the study.

6.3 A 27 bp duplication mutation causes infantile spasms and early death

Mutation screening detected a duplication mutation in a proband (IV-2 in Figure 6.3a) of 27bp, c.430_456dup(27bp).

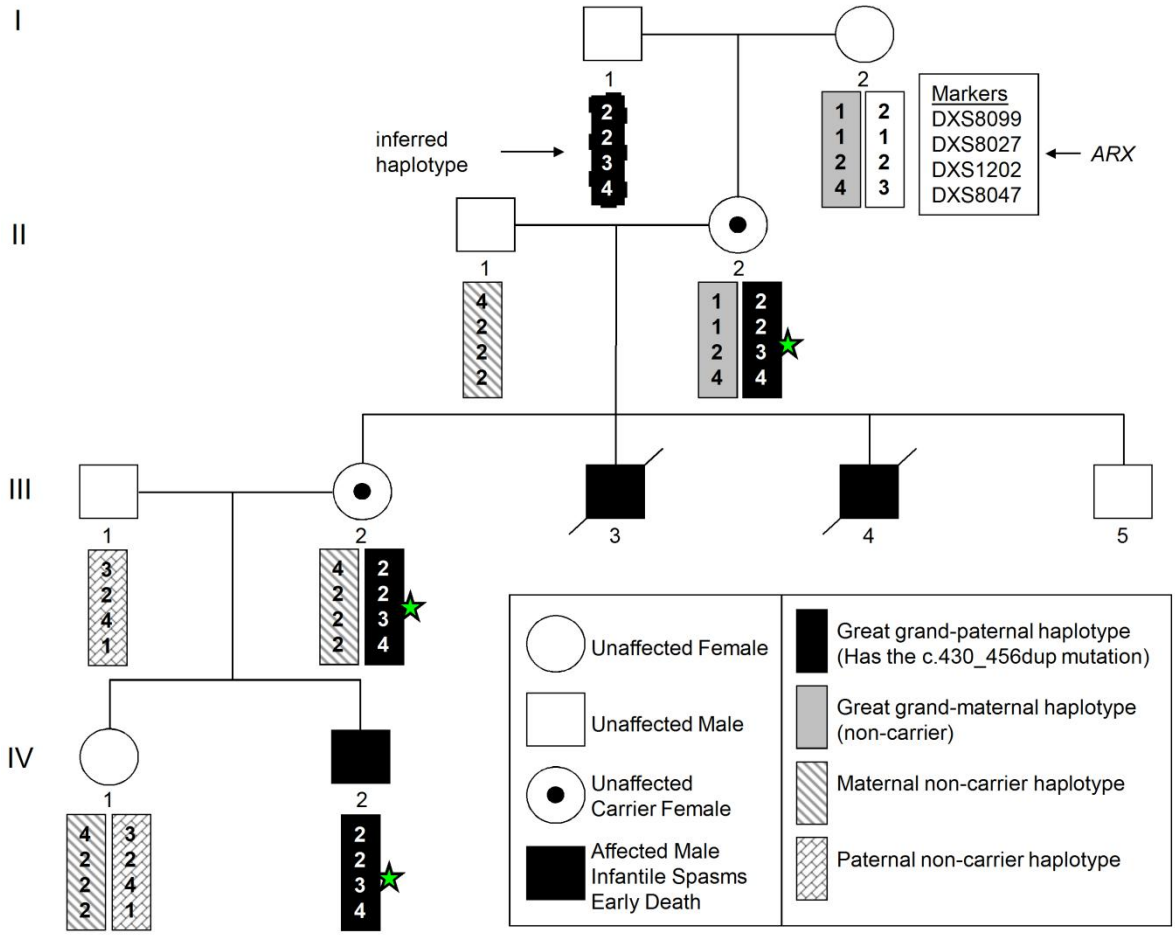

Figure 6.3a. A pedigree of family 6.3 with the c.430_456dup(27bp) mutation (indicated by a green star) within the great grand paternal haplotype. Haplotypes for microsatellite markers *DXS8099-DXS8027-ARX-DXS1202-DXS8047* are indicated by the numbered panels below each individual. Note: the haplotype for individual I-1 is inferred from his daughter (II-2) based on the haplotype that she did not inherit from her mother (I-2). Individual I-2 had three sisters that died during the Holocaust without any known offspring.

Family 6.3a – Clinical notes supplied by Drs. Orit Reish and Miriam Regev

Drs. Orit Reish and Miriam Regev practice at the Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel.

The proband (individual IV-2 in Figure 6.3a) was born to healthy parents (III-1, III-2) with Ashkenazi maternal lineage and mixed Ashkenazi/Egypt/Iraqi paternal origins. A hypsarrhythmic pattern was apparent on EEG analysis, and a brain biopsy showed 'atrophy'.

His EMG and eye findings were otherwise normal. Two maternal uncles (III-3, III-4) died about 30 years ago, at 18 months and 12 months of age respectively, yet another uncle is a healthy adult male (III-5). Both maternal uncles (III-3, III-4) had convulsions since age 2 months of age, and in one of them, hypotonia and poor eye contact were recorded.

The proband (IV-2) was born after an uneventful full pregnancy and delivery. His mother was 33 years old at conception. Triple test and elective amniocentesis were normal. Birth weight was normal and head circumference was 34 cm (50^{th} percentile). At 2 months of age infantile spasm-like convulsions were detected. Physical exam at 6 weeks of age revealed poor eye contact without nystagmus, hypotonia and foot clonus. No minor anomalies were seen. Results of a complete eye exam, BERA and echocardiography were normal. The first EEG exam at age 2 months of age was normal, but a subsequent hypsarrhythmic EEG pattern was detected at age 7 months of age. Brain MRI and MRS were normal at age two months. A metabolic work-up included blood gases, pyruvate, lactate, carnitine, blood amino acids, urine amino acids, urine organic acids and all returned normal results.

At 4 years of age the proband still suffers uncontrollable seizures intractable to variable medical interventions. He has a severe developmental delay comparable to an age of 1-2 months. He has no consistent eye contact, does not pull to sitting position, has inconsistent social smile and does not have any vocal capabilities.

Given the intractable nature of the infantile seizures and their early onset within the proband and two of his maternal uncles, whom also died at a young age (not normally seen for patients with the 24 bp duplication), these phenotypes appear to be a severe form of infantile seizures, relative to 24 bp duplication mutations.

The c.430_456dup(27bp) was detected in the proband (IV-2), his mother (III-2) and maternal grandmother (II-2; Figure 6.3a) by PCR, agarose gel electrophoresis and direct sequencing of

PCR products. Interestingly, the 27 bp that are duplicated in this family (c.430_456dup(27bp)) partially overlapped with the region duplicated in the recurrent dup24bp mutation, and is anchored by a single adenosine (A) nucleotide at c.441 within this GC rich sequence (Figure 6.3b). The maternal grandmother (II-2) had significantly lower levels of the dup27bp allele, when compared to the mother (III-2) by PCR product (PCR; Chapter 2.4.2, primers 2J-2; Table 2.3a) intensity when imaged by ethidium bromide staining in an agarose gel (data not shown). The maternal great-grandmother (I-2) did not carry this mutation in DNA obtained from either blood or buccal samples (Figure 6.2c) as determined by several PCR attempts (data not shown).

In order to identify the origin of this apparently *de novo* mutation in individual II-2, we performed haplotype analysis using flanking microsatellite markers on available family members. This analysis was performed using a radio-labelled PCR (Chapter 2.4.2) with primers specific to the STS markers *DXS8099*, *DXS8027*, *DXS12027* and *DXS804* (Table 2.3d). This analysis revealed that the haplotype carrying the dup27bp mutation was that of the maternal great-grandfather (I-1), for whom DNA samples were not available and therefore his haplotype had to be inferred (Figure 6.3a).

Table 6.3a. STS marker positions by UCSC genome browser (build Feb 2009; GRCh 37/hg19) for *DXS8099-DXS8027-ARX-DXS1202-DXS8047* at Xp22.11-Xp21.3, relative to the *ARX* locus.

STS Marker	Chromosomal location	Start	End
DXS8099	Xp22.11	24,419,207	24,419,647
DXS8027	Xp22.11	24,536,746	24,537,027
(ARX)	Xp21.3	25,021,815 (3')	25,034,065 (5')
DXS1202	Xp21.3	26,471,345	26,471,648
DXS8047	Xp21.3	27,645,467	27,645,778

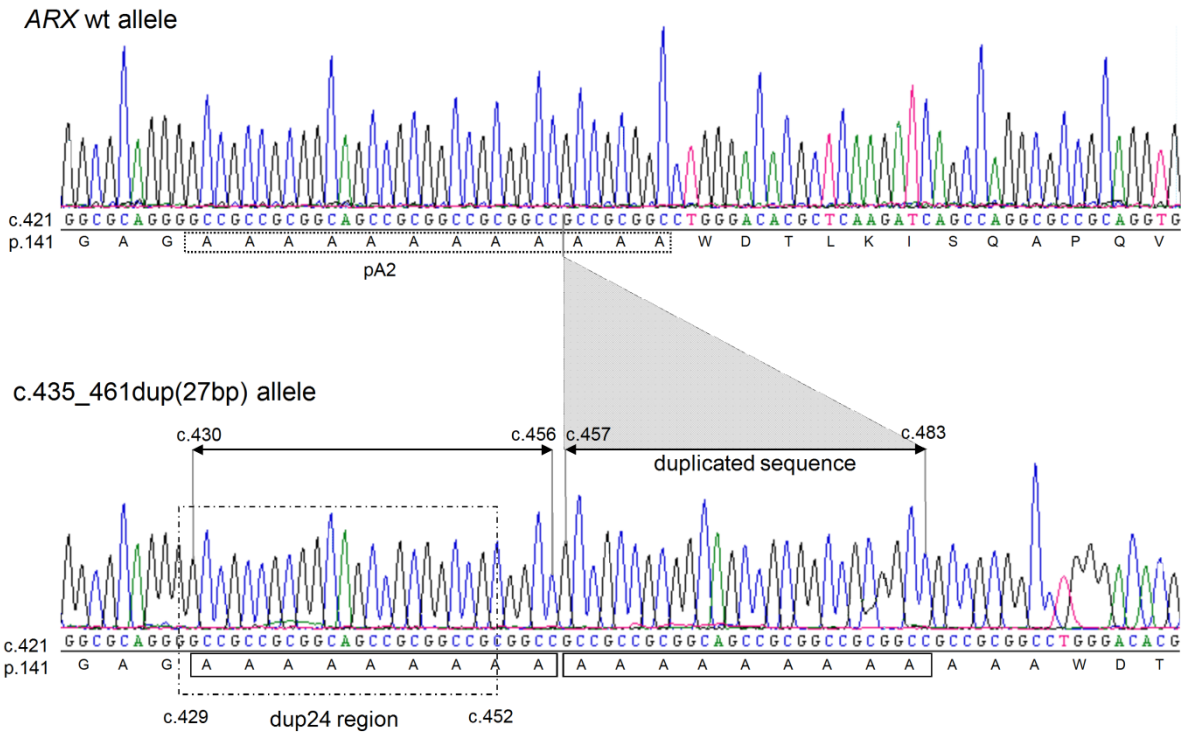

Figure 6.3b. Partial sequence chromatograms for the region surrounding the DNA that codes for polyalanine tract 2 (pA2) in exon 2 of the *ARX* gene. Chromatograms were generated from an unaffected individual (upper panel) and the proband (IV-2, lower panel). The 27 base pairs duplicated in the proband, c.430_456dup, are highlighted with solid line boxes (lower panel). The region of the most frequent *ARX* mutation, c.429_452dup(24bp), is boxed with a broken line (lower panel). Below each chromatogram are cDNA and amino acid sequences. The second polyalanine tract of wildtype ARX is boxed with a dotted line (upper panel). All chromatograms shown are derived from DNA extracted from whole blood.

To address the question of the origin of the dup27bp mutation, *ie* whether it originated in the great-grand paternal (I-1) sperm who would have to have been a gonadal mosaic or as an early somatic mutation in the grandmother (II-2; gonadal mosaicism not testable in the grandmother), we collected buccal swabs from the grandmother II-2 for DNA analyses of fibroblasts. Semiquantitative PCR using a combination of a standard primer (Table 2.3a) in conjunction with one hexachlorofluorescein (fluorescent) labelled primer (Table 2.3b) confirmed the presence of the dup27bp allele, as a proportion of total PCR product, was

significantly lower level in the grandmother (II-2: 18.3% in blood and 2.4% in buccal) when compared to the mother (III-2 40.4% in blood and 35.9% in buccal) (Figure 6.2c).

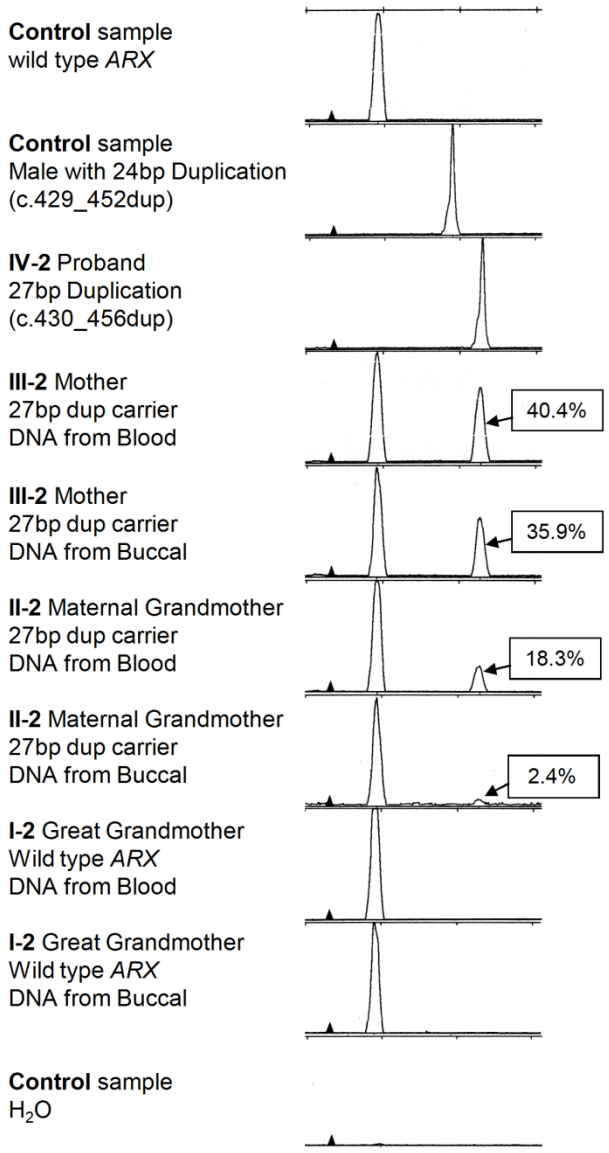

Figure 6.3c. Fluorescent fragment analysis of individuals from this family compared to a wildtype allele (top panel), the c.429_452dup (*ie* 24bp duplication) mutation (second panel) and a negative water control PCR (bottom panel). Input DNA derived from both blood and buccal sources are shown for individuals III-2, II-2, I-2. This clearly indicates that the 27 bp mutation is larger than the recurrent dup24bp mutation and is present in individuals II-2, III-2, IV-2. Boxed numbers to the right of the fragment analysis traces show the percent of PCR product derived from the c.430_456dup mutant allele (*ie* 27 bp duplication) as a proportion of total PCR product, based on peak areas.

The c.430_456dup(27bp) allele comprised 40.4% (blood) and 35.9% (buccal fibroblasts) of total PCR product when amplified from the mother's (III-2) DNA, resulting in a ratio of 1.125 between these tissues. Conversely, the grandmother's (II-2) DNA comprised of 18.3% (blood) and 2.4% (buccal fibroblasts) of the mutant allele resulting in a ratio of 7.6 (Figure 6.3c). This greatly increased difference in c.430_456dup(27bp) to wildtype allele ratio between blood and fibroblasts derived from individual II-2 is indicative of postzygotic

(somatic) mosaicism. We speculate that the c.430_456dup(27bp) *de novo* event likely occurred at an early embryonic stage in a multipotent cell, which gave rise predominantly to the mesoderm (blood) than the ectoderm (buccal fibroblasts).

A ratio of 1:1 is expected between the mutant and wildtype alleles in tissues examined. But it has been previously shown that in heterozygotes for the c.429_452dup(24bp) mutation that an underrepresentation of the mutant allele occurs in PCR products derived from blood (*ie* mutant allele/wildtype allele; ratio of 0.66 ((Poirier *et al.*, 2005); Table 6.3b). The above results from family 6.3a concur with this previous report as analyses of DNA from individual III-2, showed a c.430_456dup(27bp)/wildtype allele ratio of 0.67, and 0.55 in buccal fibroblasts (Table 6.3b). Conversely, the c.430_456dup(27bp)/wildtype allele ratio is in DNA from individual II-2 was 0.22 in blood, and 0.02 in buccal fibroblasts (Table 6.3b). This approximate ten fold difference in the mutant/wildtype allele ratio in blood vs buccal fibroblasts might reflect the presence of variable somatic mosaicism for the c.430_456dup(27bp) allele between the two tissues, exacerbated by the preferential amplification of the wildtype allele. Preferential amplification may occur for many reasons, but it is likely that the increased mutant allele size (*ie* 340 bp vs 313 bp for wt) and increased GC ratio of the mutant allele contribute to this phenomenon.

Table 6.3b. Amplification of wildtype and c.429_452dup(24bp) and c.430_456dup(27bp) mutant *ARX* alleles from female carrier's DNA sourced from various tissues.

Mutation	**Source**	**Tissue**	**mutant (%)**	**wildtype (%)**	**Ratio (mutant/wt)**
c.429_452dup(24bp)	Poirier *et al.*, 2005	Blood	60.0	40.0	0.66
c.430_456dup(27bp)	III-2	Blood	59.6	40.4	0.67
c.430_456dup(27bp)	III-2	Buccal	64.9	35.9	0.55
c.430_456dup(27bp)	II-2	Blood	81.7	18.3	0.22
c.430_456dup(27bp)	II-2	Buccal	97.6	2.4	0.02

6.4 A 33 bp duplication mutation causes Partington syndrome

Mutation screening and cascade testing detected another duplication mutation that was larger than the 24 bp duplication by 9 bp, *ie* c.423_455dup(33bp), in a proband (III-1), his brother (III-2) and his unaffected mother (II-1) (Figure 6.4a).

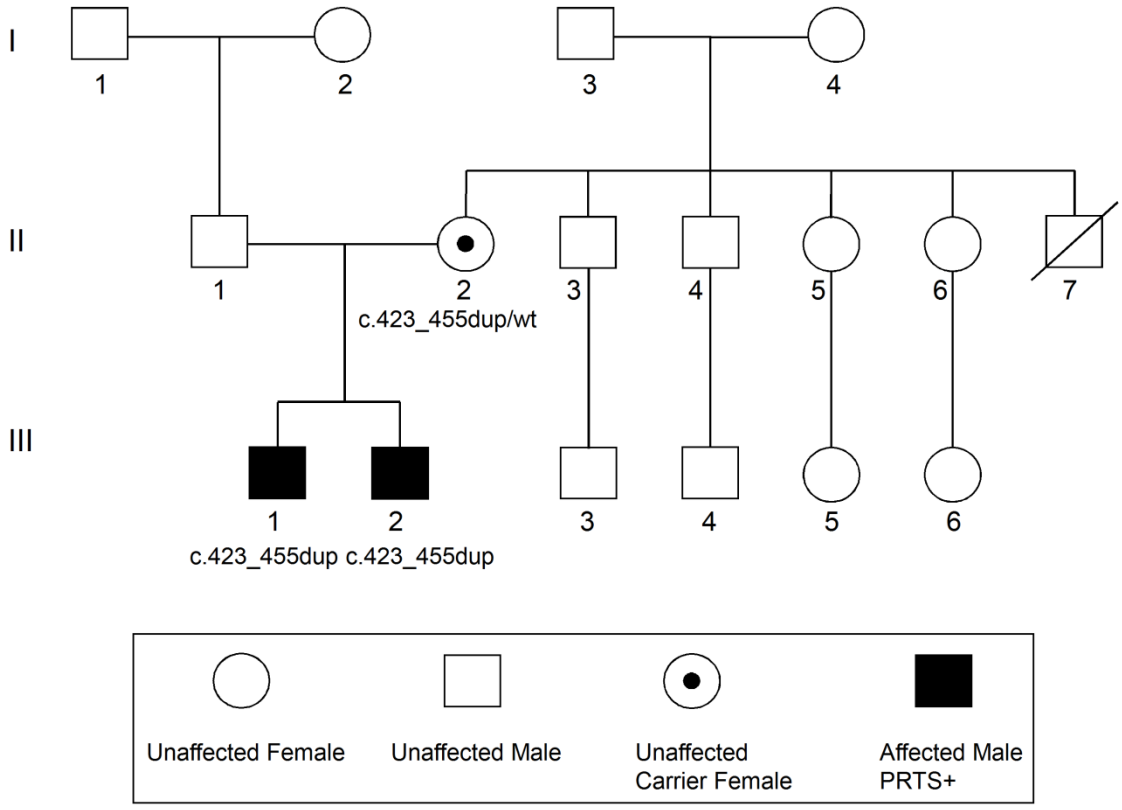

Figure 6.4a. A pedigree of family 6.4. Two brothers with atypical Partington Syndrome with additional phenotypic features (III-1 and III-2) have the c.423_455dup(33bp) mutation, whilst their mother (II-2) carries it, as indicated below the symbols that represent them.

Family 6.3a – Clinical notes supplied by Drs. Michelle Demos and William Gibson

Dr. Michelle Demos is from the Department of Paediatric Neurology, British Columbia's Children's Hospital, University of British Columbia, Vancouver, Canada.

Dr. William Gibson is from the Department of Medical Genetics and Provincial Medical Genetics Program, University of British Columbia, Vancouver, Canada.

The two affected male siblings (III-1 and III-2; Figure 6.4a) were brought to Hospital for psychomotor delay, dystonia, and epilepsy from infancy to their current age of 19 and 14 years of age, respectively. There was no family history of the above traits.

The proband (III-1) was the first child born to non-consanguineous parents of Northern European descent and was healthy at birth following a term pregnancy complicated by gestational hypertension. He was delivered by C-section for failure to progress. His birth weight was 4200 g (~91st centile). Birth head circumference was not recorded. Psychomotor delay was noted by late infancy. Infantile spasms were diagnosed at 5 months of age, and responded to vigabatrin, but not Adrenocorticotropic hormone. He has had no further seizures since 8 months of age and has not used anti-epileptic medication since 2 years of age. Hand dystonia was first noted at 18 months of age. At 15 years of age, he walks with support or mobilises on his knees. He never obtained a pincer grasp. His hands are non-functional, but his feet retain some function including turning pages of a book. His first word occurred at 5 years of age. Speech is mildly dysarthric and at the level of three to four word sentences. He is unable to read or write. He is estimated to have a moderate to severe intellectual disability and he is fully dependent on others to manage his activities of daily living.

Physical examination of the proband (III-1) at 14 years of age revealed height of 150 cm (2 cm< 5th centile), weight 49.3 kg (40th centile), and head circumference 58 cm (98th centile). He was not dysmorphic. Mild hypospadias was present. On neurological examination, his speech was mildly dysarthric. Eye movements appeared normal. He had generalized dystonia involving mouth, neck, trunk and limbs. He was unable to grasp objects with his hands. When walking with assistance, his neck and spine were severely extended. His lower extremities were also extended and turned in and he walked on his toes. Deep tendon reflexes were increased in the lower extremities and plantar responses were upgoing. Head MRIs at 5 months, 4 years and 6 years of age were normal. Electroencephalogram (EEG) at 5 months captured spasms, which electrographically were associated with electrodecremental events; an EEG at 4 years of age was normal. Chromosomal investigations, neurophysiological studies, muscle biopsy, plasma/cerebrospinal fluid (CSF) amino acids, CSF lactate and urine organic acids were all normal.

The proband's younger brother (III-2) was healthy at birth, following a 41 week pregnancy complicated by gestational hypertension and polyhydramnios in the 3rd trimester. He was delivered by elective C-section. Birth weight was 4300 g (95th centile) with a head circumference of 39.5 cm (>99.6th centile). At 2 weeks of age he developed clusters of brief episodes of generalized body jerks with arm stiffening and staring during sleep or drowsiness. These events were determined to be non-epileptic as they were not associated with EEG changes on a video-EEG recording, which captured several typical events. The same EEG did show some very mild background abnormalities, but no epileptiform activity. These episodes resolved by 7 months of age as shown by a normal EEG. Hand dystonia and psychomotor delay was evident by one year of age. Although his current hand function is limited, he is able to grasp a writing tool or utensil and transfer objects. He is able to walk without support, but he has poor lip control and requires assistance with drinking. At 7 years

of age he developed seizures, which are currently controlled on lamotrigine. These included frequent brief episodes of unresponsiveness, staring, oral automatisms and reduced tone. More rarely, he had prolonged episodes of generalized motor activity with unresponsiveness. At age 7 he also developed episodes of hyperventilation. These episodes occur on a multiple daily basis during wakefulness only and tend to occur more when he is anxious or excited. They are associated with mild cyanosis, but no loss of consciousness. They stop abruptly without intervention and last less than 2 minutes. He also has a history of poor eye contact and difficulty with social interaction. He has some repetitive behaviours and becomes anxious if his actions are interrupted.

Neurodevelopmentally, III-2 walked by 3 years of age and his first words were at 5 years of age. His vocabulary is limited to only 10 words, is unable to put two or more words together in a sentence and is unable to read or write. He is estimated to have a moderate-severe ID and is dependent on others for the majority of daily living activities.

A physical examination of III-2 at 9 years 9 months revealed height of 120 cm (7 cm < 5^{th} centile), weight of 22.1 kg (2 kg < 5^{th} centile) and head circumference of 56 cm (98^{th} centile). He had mild frontal bossing and mild hypospadias was present. On neurological exam, he had no speech and poor eye contact, but eye movements appeared normal. He had 2 to 3 episodes of hyperventilating lasting less than 1 minute. He had generalized dystonia involving mouth, neck, trunk and limbs. He also had choreoathetoid movements of his arms. Deep tendon reflexes were brisk in the lower extremities and plantar responses were extensor. While walking there was slight extension of cervical and thoracic spine and he also walked with his legs extended and turned in. Head MRI at age 9 was normal. His EEG at 9 years was dysrhythmic and demonstrated rare generalized spike and wave activity.

The unaffected mother of the proband (II-2) was normal for her examination at 43 years of age, with the exception of macrocephaly (OFC 58.5 cm - above 98th centile).

Given the additional features of hyperventilation, autistic like traits, severe dystonia in the hands and dystonia in other regions besides hands, these brothers appear to have a severe atypical form of PRTS, relative to 24 bp duplication mutations in *ARX*.

The duplicated 33 bp flank the region that is duplicated in the recurrent c.429_452dup(24bp) mutation and are inserted after base pair c.455 in the second polyalanine tract (Figure 6.4b). When translated, this mutation is predicted to result in a non-homogeneous expansion of pA2. The resultant tract would have the following structure – $AGA_{10}GA_{12}$, compared to that of wt (AGA_{12}) and the c.429_452dup(24bp) mutation (AGA_{20}) alleles. While the overall number of alanine residues is increased from 12 (wt) to 22, the resulting polyalanine tract is interrupted by a single glycine residue. Based on the previous data and mutations affecting this polyalanine tract of the ARX protein this duplication is likely causative of the pathogenesis seen in this family. Although this phenotype appears relatively mild, compared to what might be expected from the largest pA2 expansion reported, this may partly be due to the non-homogenous nature of the expansion.

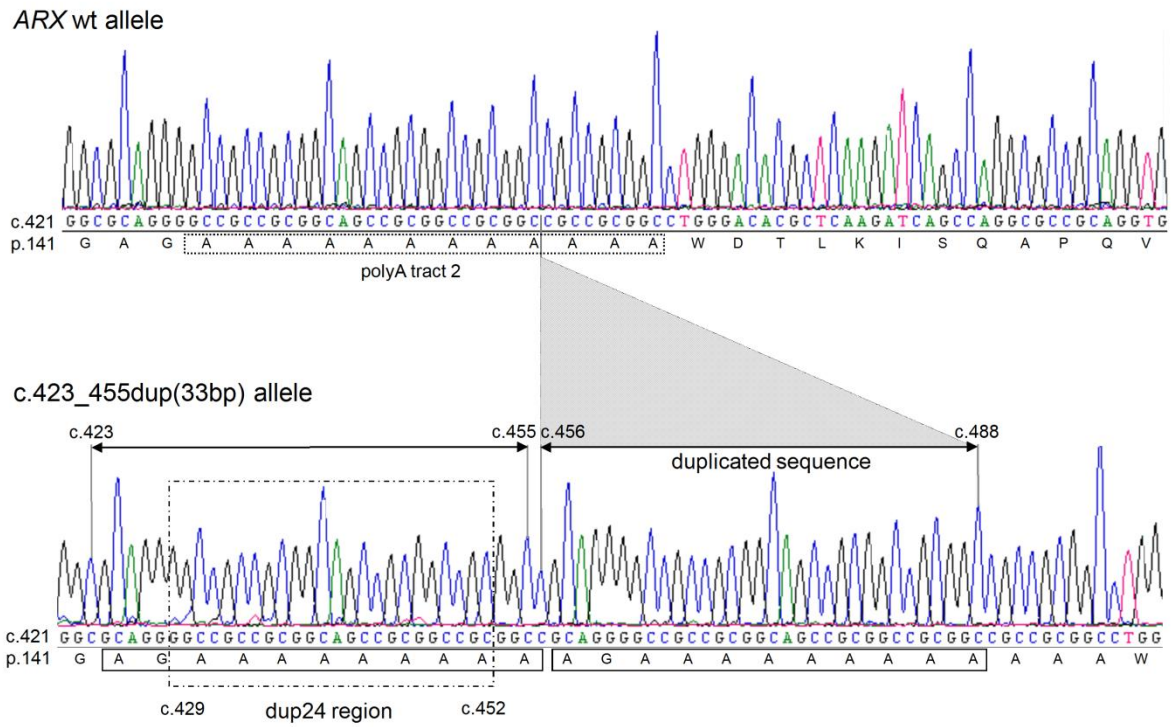

Figure 6.4b. DNA sequence chromatograms of partial sequence of exon 2 of the *ARX* gene from an unaffected individual (wildtype; upper panel) and the proband (lower panel). The 33 base pairs (bp) duplicated, c.423_455dup(33bp), are highlighted with solid line boxes (lower panel). The region of the most frequent *ARX* duplication, c.429_452dup(24bp), is boxed with a broken line (lower panel). Below each chromatogram are cDNA and amino acid sequences. The second polyalanine tract of ARX is boxed with a dotted line.

6.5 Mutant ARX protein mislocalises as a function of polyalanine tract 2 length

In general it appears that the two larger duplication mutations in *ARX*, c.429_456dup(27bp) and c.423_455(dup33), cause more severe phenotypes relative to the majority of c.429_452dup(24bp) mutations. This may be due to the larger expansions of pA2 presumably caused by these mutations, both of which breach the natural limit of 20A per wildtype pA tract (reviewed by (Albrecht and Mundlos, 2005)). This hypothesis was tested by cell based localisation studies of pA2 mutants in full length ARX proteins.

The different sized duplication mutations (24 bp, 27 bp and 33 bp) presumably expanded pA2 from 12A (wt) to 20A, 21A and 22A interspersed by a glycine (G) residue, respectively (See Table 6.5a). Individual *ARX* constructs with three different lengths of pA2 (c.429_452dup(24bp) – 20A pA2; c.430_456dup(27bp) – 21A pA2; c.423_455dup(33bp) – 10A-G-12A – pA2) were made from the *pCMV-myc-ARX-wt* vector (Chapter 2.8) and transfected into human embryo kidney cells (HEK-293T; product number CRL-11268**,** ATCC, Manassas, USA) (Chapter 2.9), subjected to fluorescent immunohistochemistry (Chapter 2.10) and standard microscopy (2.10 and 2.11). An average of 647 ARX positive cells were scored per transfection for three separate transfection experiments and localisation was classified as either normal, having nuclear inclusions, or cytoplasmic positive with or without aggregates in either the nucleus or cytoplasm (see representative images in Figure 5.4a; Chapter 5.4) compared to a wildtype *ARX* construct (12A pA2; Figure 6.5a and Table 6.5a).

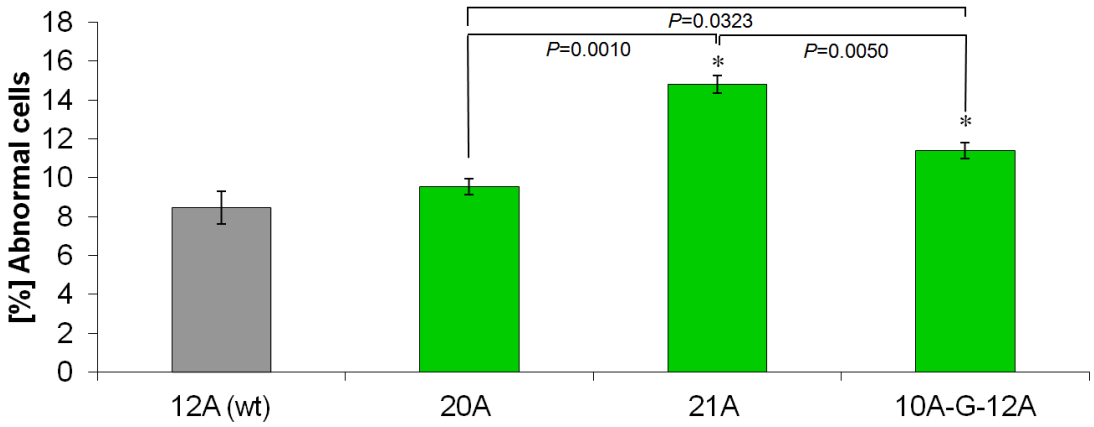

Figure 6.5a. Proportion of ARX positive cells displaying abnormal localisation (nuclear inclusions or cytoplasmic positive with or without aggregates in either the nucleus or cytoplasm) for three mutant pA2 tract lengths (green bars; c.429_452dup(24bp) – 20A; c.430_456dup(27bp) – 21A; c.423_455dup(33bp) – 10A-G-12A) compared to wildtype ARX (grey bar; 12A). 647 ARX positive cells were scored per transfection for three separate transfection experiments 24 h post transfection. * Asterisks indicate which constructs had significantly different levels of protein mislocalisation compared to wildtype ($P \leq 0.05$; actual *P* values can be found in Table 6.5a) and between mutant *P*-values are indicated (when $P \leq 0.05$), using a general linear model and an unbalanced ANOVA analysis.

Mutant ARX protein mislocalised as a function of pA2 tract (Figure 6.4a) and aligned with the phenotypic severity of males with the pA2 mutations investigated (Table 6.5a). But the alignment of increased protein mislocalisation with pA2 tract length only emerges when homogenous polyalanine tracts are considered.

The longest homogenous pA2 tract expansion to 21A (c.430_456dup(27bp)) causes significantly more ARX mislocalisation than all other pA2 mutations tested. The phenotype induced by the 21A-pA2 mutation is ISSX with early onset and death (Chapter 6.3; (Reish *et al.*, 2009, Appendix A.7)), more severe than the atypical PRTS with additional features resulting from the 10A-G-12A pA2 mutation (Chapter 6.4 (Demos *et al.*, 2009, Appendix A.8)). Although the 20A-pA2 mutation can cause PRTS (reviewed by (Gecz *et al.*, 2006; Shoubridge *et al.*, 2010a, Appendix A.1)), the 10A-G-12A-pA2 mutation caused atypical PRTS with a high degree of dystonia and additional features, suggesting a more severe phenotype than the majority of pA2 duplication mutations (Chapter 6.3; (Demos *et al.*, 2009, Appendix A.8)). This is mirrored by the 10A-G-12A-pA2 mutant protein mislocalising significantly more than the mutant 20A and wt (12A-pA2) ARX. The c.423_455dup(33bp) mutation has a 22A-pA2 tract interrupted by a single glycine residue (10A-G-12A) which was suspected to ameliorate the severity of the phenotype (Chapter 6.3; (Demos *et al.*, 2009, Appendix A.8)), and is supported by the protein localisation results.

The c.429_452dup(24bp) mutation did not significantly increase mislocalisation of the mutant protein compared to wt, consistent with previous over-expression (McKenzie *et al.*, 2007) and mouse (Kitamura *et al.*, 2009) studies. A 21A-pA2 causes significantly more protein mislocalisation than 20A, providing further evidence of a naturally occurring limit of 20A that has been previously suggested (reviewed by (Albrecht and Mundlos, 2005)), even when breached by a single alanine residue. As tract length increases over 20A in pA1 or pA2, a proportionate increase of mislocalised protein occurs.

Table 6.5a. Amount of ARX protein mislocalisation per pA2 mutant expression construct (*pCMVmycARX*) 24h post transfection (tfxn).

Construct	Protein	pA2 tract length	Abnormal localisation ±SEM (%)	*P* value (vs wt) [1]	NI [2] ±SEM (%)	*P* value (vs wt) [1]	CA [2] ±SEM (%)	*P* value (vs wt) [1]	*n* [3]	Average cell count [4]	Phenotype
Wildtype	-	12A	8.47 ±0.64	-	6.03 ±0.59	-	2.44 ±0.05	-	9	486	None
c.429_452dup (24bp)	p.A151insA_8	20A	9.5 ±0.41	0.7187	7.04 ±1.27	0.8019	2.46 ±0.42	0.7623	3	687	Variable [5]
c.423_455dup (33bp)	p.A151delAAinsG$(A)_{12}$	10AG12A	**11.41 ±0.41**	**0.0500**	**8.85 ±0.49**	**0.0164**	2.56 ±0.75	0.7171	3	591	PRTS+
c.430_456dup (27bp)	p.A151insA_9	21A	**14.80 ±0.45**	**0.0001**	**12.19 ±0.34**	**0.0066**	2.61 ±0.74	0.9257	3	824	ISSX+

[1] *P* value for comparison of total amount of abnormal localisation or for each sub-classification to that of the wildtype construct by general linear model and an unbalanced ANOVA analysis. A *P* value ≤0.05 was considered significant.
[2] Sub types of mislocalisation observed for each ARX protein construct tested, classified as nuclear inclusions only (NI) or cytoplasmic positive with or without aggregates in either the nucleus or cytoplasm (CA).
[3] The number of transfection replicates performed for each construct
[4] The average number of ARX positive cells counted for localisation per transfection (tfxn)
[5] Phenotype varies from nsXLID to PRTS and ISSX.
+ Additional phenotypic observations were made for patients with this type of mutation
Bold font indicates a significant difference to wildtype.

All of the mislocalisation of pA2 mutant ARX proteins was observed as nuclear inclusions (Table 6.5a), compared to the cytoplasmic positive with or without aggregates in either the nucleus or cytoplasm observed for pA1 mutants (Chapter 5). Despite the difference in the sub classification of the mislocalisation observed between pA1 and pA2 mutants, a clear trend of increasing mislocalisation as a function of increased pA tract (1 or 2) length emerges (Figure 6.5b). This trend is particularly evident when the resultant tract is greater than 20A.

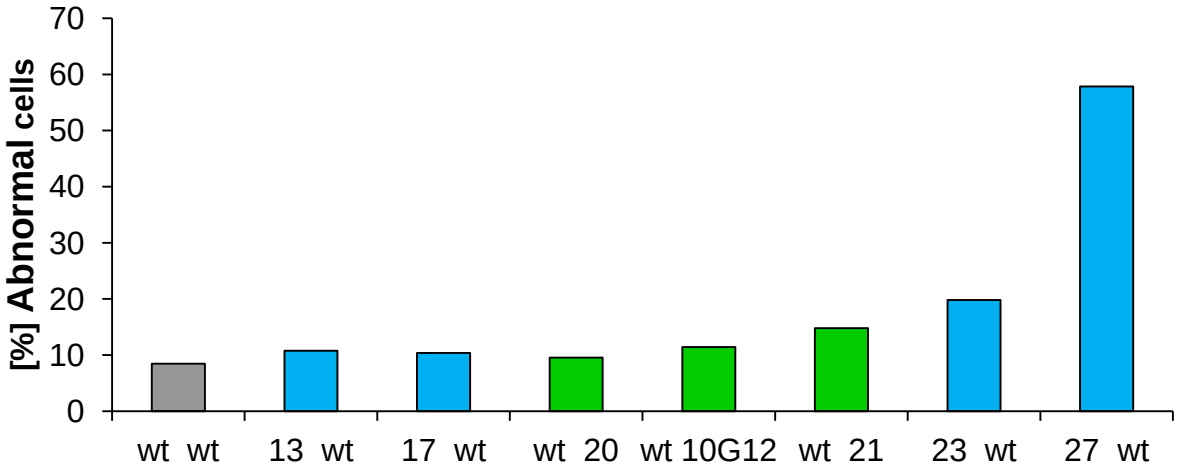

Figure 6.5b. Proportion of ARX positive cells displaying abnormal localisation for: four pA1 tract lengths (blue bars; c.305_313del - 13A, c.304ins$(GCG)_1$ - 17A, c.304ins$(GCG)_7$ - 23A and c.298_330dup - 27A); three pA2 tract lengths (green bars; c.429_452dup(24bp) – 20A; c.430_456dup(27bp) – 21A; c.423_455dup(33bp) – 10A-G-12A); compared to wildtype ARX (grey bar; 12A-16A). The X-axis labels indicate the length of the resultant pA1 and pA2 tracts, respectively. 570 ARX positive cells were scored per transfection for three separate transfection experiments 24 h post transfection.

6.6 DNA Sequence analysis of the three duplication mutations

The discovery of three duplication mutations in total (c.429_452dup(24bp), c.430_456dup(27bp) and c.423_455dup(33bp)) provided a unique opportunity to analyse the primary and secondary structure of the resultant mutant sequences. The pA2 duplication mutations are assigned nucleotide positions at the earliest possible nucleotide that duplicated

sequence can commence at. But the actual positions of the mutations are ambiguous and can be assigned to many different nucleotide positions within the resultant mutant sequence (Figure 6.6a).

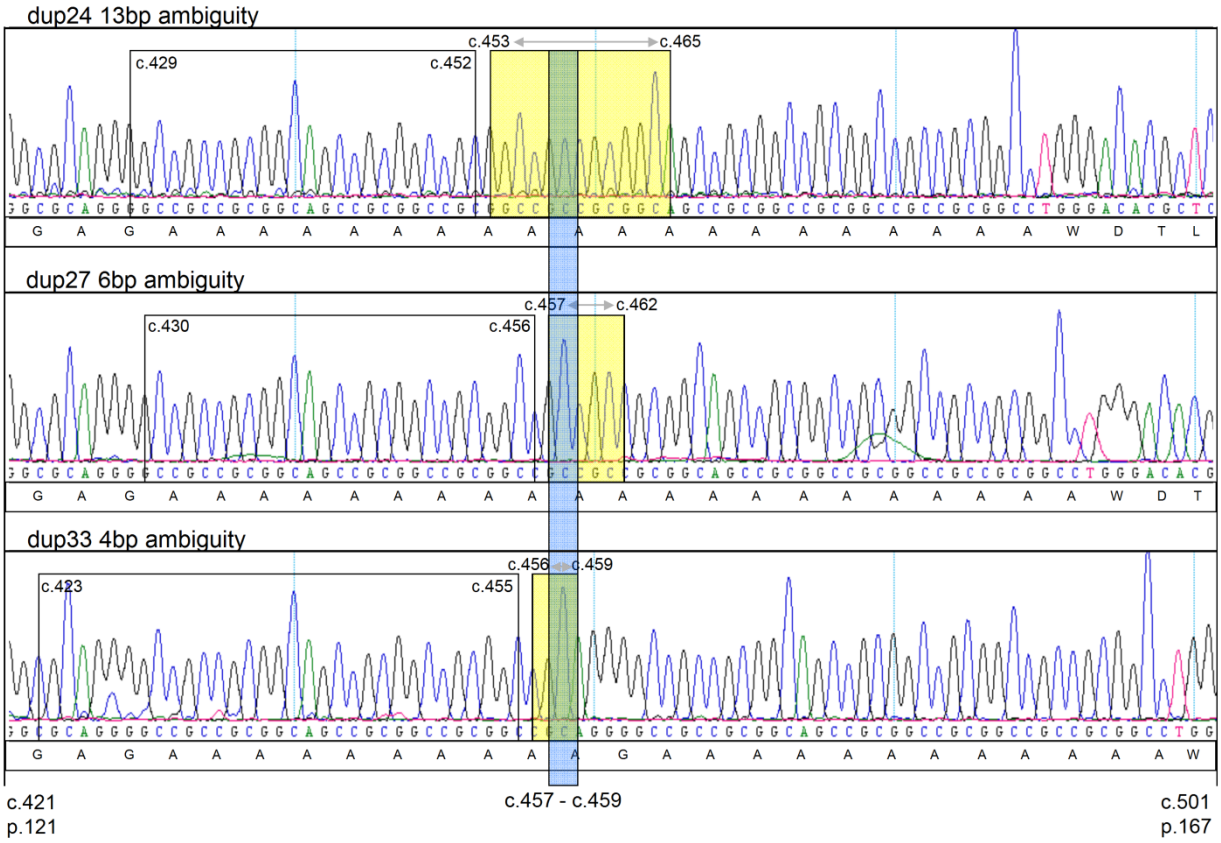

Figure 6.6a. The precise position where mutant sequence commences for each pA2 duplication mutation is ambiguous. Partial exon 2 sequence that flanks pA2 is given for each duplication mutation, aligned by coding position: (**i**) c.429_452dup (dup24); (**ii**) c.430_456dup (dup27); and (**iii**) c.423_455dup (dup33). The yellow highlights the base position at which mutant sequence can commence, indicating that this may occur at a range of nucleotide positions without altering the resultant mutant sequence. Blue highlight indicates the three base pairs (c.457_459) that are in common for mutant sequence commencement between all three duplication mutations.

Three nucleotides (c.457_459) are in common as possible start points of duplicated sequence for all three separate pA2 duplication mutations (c.429_452dup – dup24; c.430_456dup – dup27; and c.423_455dup – dup33). To generate the most frequent *ARX* mutation

(c.429_452dup (24bp)) 24 bp of sequence must be repeated, it is speculated that this probably occurs immediately prior or at nucleotides c.457_459. Therefore the secondary structure of the DNA that surrounds these nucleotides was subjected to secondary structure analysis and the most energetically favourable conformations were scrutinised (Figure 6.6b).

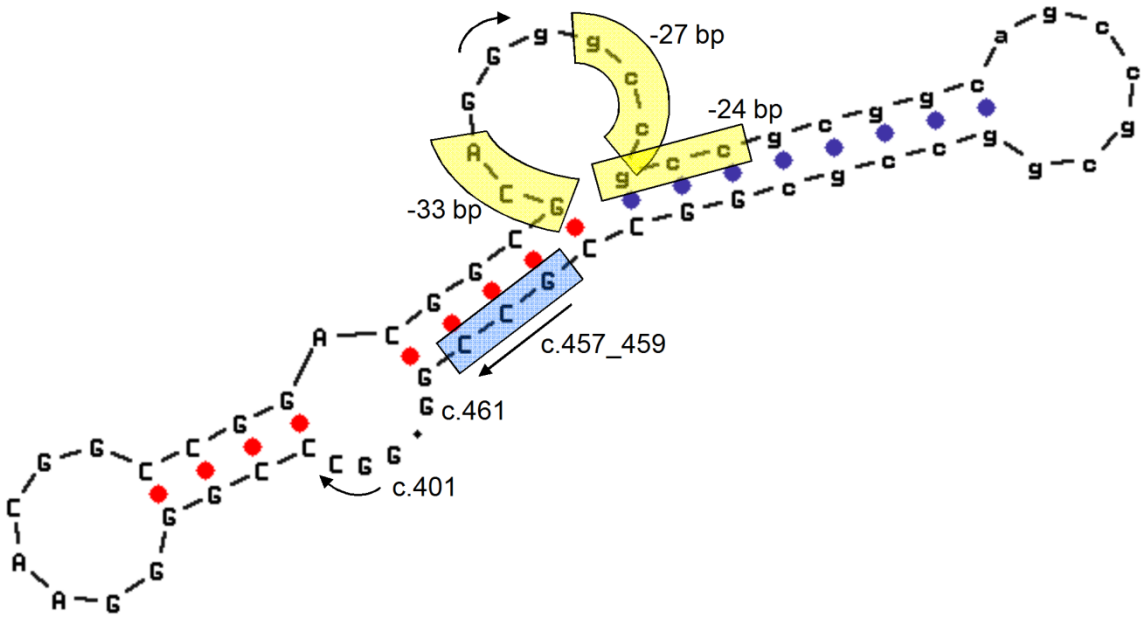

Figure 6.6b. The predicted conformation with the lowest kinetic energy by the mfold algorithm (Chapter 2.13) for the sequence surrounding the region involved in the pA2 duplication mutations (c.401_461). The blue box highlights c.457_459 as identified in Figure 6.6a, and the yellow boxes indicate where the sequence at the beginning of each of the three duplications would lie with respect to c.457_459 (*ie* -24bp, -27bp, -33bp from c.457_459). The top ranked energetically favoured secondary structures for this region of DNA, all have similar conformations surrounding the highlighted base pairs (data not shown).

It can be speculated that the secondary structure that surround these three nucleotides (c.457_459), is permissive of a mutation event that results in the duplication of most commonly 24 bp. Curiously the three nucleotides (c.457_459) form a GCC alanine codon of which is the exact sequence of the 3 nucleotides at -24 bp and -27 bp, and further match 2/3 nucleotides at -33 bp (Figure 6.6b). Perhaps this pair of GCC codons are involved in the

generation of both the 24 bp and 27 bp duplication, and the first 2 nucleotides may be the most important for a duplication event as they are preserved in the sequence at -33 bp.

When the DNA sequence that codes for pA2 of *ARX* is subjected to a BLAST search (NCBI website, Chapter 2.13) and restricted to sequences that are expanded in human disease, the most similar sequences originate from pA tracts within *HOXA13*, *HOXD13* and *FOXL2* (Muragaki *et al.*, 1996; Goodman *et al.*, 2000; Beysen *et al.*, 2008). Although no striking structural similarities are obvious between the DNA sequence of the pA tracts (Figure 6.6c), a similar phenomena exists as described above for the sequence of pA2 of *ARX*. That is within the DNA that codes for the pA tract in each case, two alanine codons with the same sequence are utilised at a spacing that matches the size of the expansion that occurs in the mutant pA tract. For example in *HOXA13*, which has the most similar pA sequence to *ARX* pA2, a 6A expansion might be caused by a two pairs of GCG alanine codons that are spaced 6 codons apart *(ie* codon numbers 2/7 and 3/8; Figure 6.6c). This holds true for all pA tract expansions with similarity to the sequence of pA2 of *ARX*, with the exception of a 14A expansion in *HOXD13*. The 14A *HOXD13* expansion might result from sequence elements outside of this particular pA tract, as occurs for the 33 bp duplication in *ARX* (Figure 6.6a).

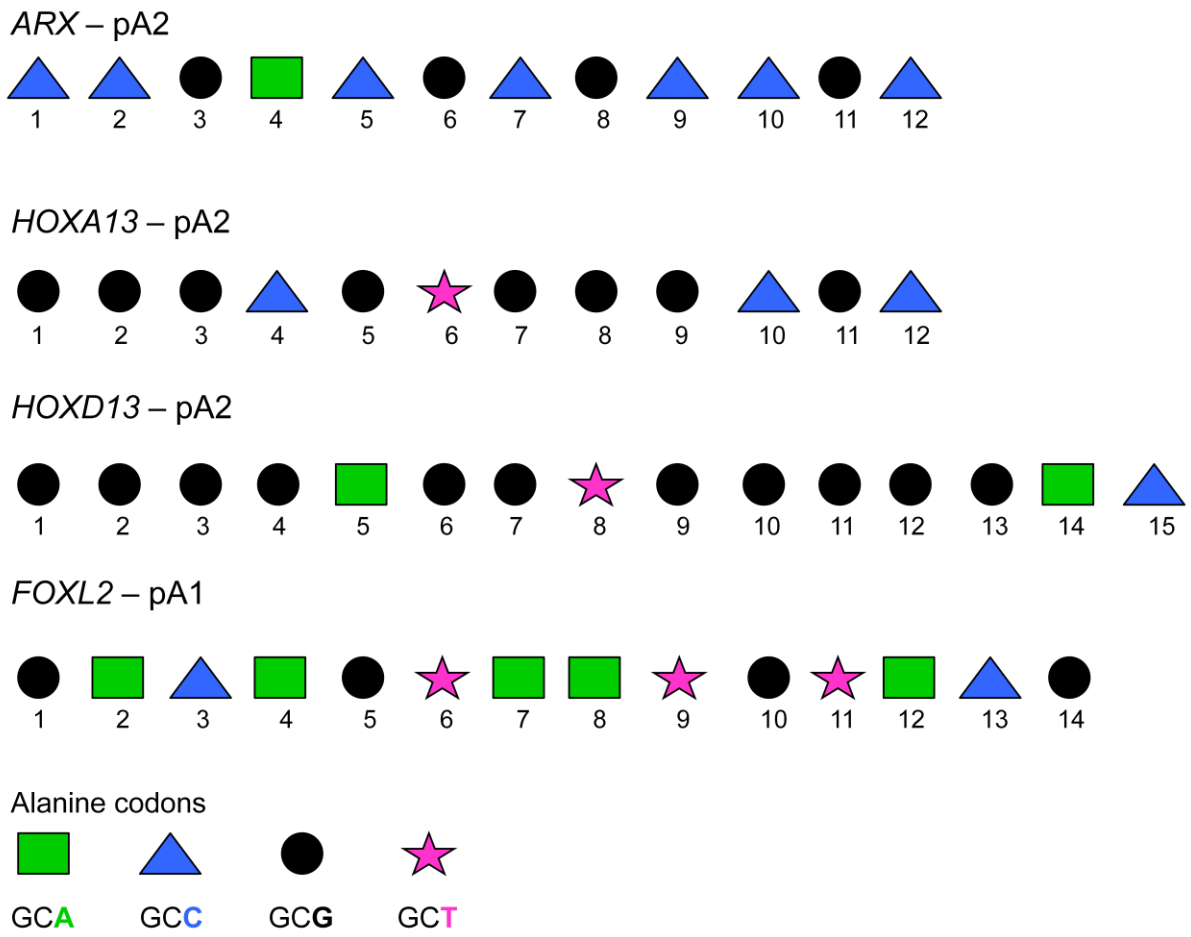

Figure 6.6c. Codons used within pA tracts with similar DNA sequences to that of *ARX* pA2. Expansions have been reported to cause human disease in all of the pA tracts shown: *ARX* pA2 (12A wt) +8A, +9A and -2A/+G+12A; *HOXA13* pA2 (12A wt) +6A; *HOXD13* pA2 (15A wt) +7A, +8A, +9A, +10A and + 14A; and *FOXL2* pA1 (14A wt) +5A and +10A. Reference sequences were obtained from NCBI (Chapter 2.13): *ARX* – NM_139058 ; *HOXA13* – NM_000522; *HOXD13* – NM_000523; *FOXL2* – NM_023067.

Once the duplication mutation in *ARX* is generated it is stably inherited, comprised of imperfect trinucleotide repeats and is formed from progenitor sequences much shorter than the 75–105 bp of perfect repeats that would be expected to cause replication slippage (Eichler *et al.*, 1994). Further, it has been suggested that DNA slippage does not suit the 'nature' of pA expansion mutations (Trochet *et al.*, 2007). It therefore seems unlikely that DNA replication slippage is the mechanism responsible for generating the duplication mutations in *ARX*.

Two mechanisms have been suggested to mutate the DNA that codes for pA tracts: (i) unequal crossing over between two mispaired wildtype alleles (Warren, 1997; Kato *et al.*, 2003); and more recently (ii) a version of fork stalling and template switching (FosTeS) was applied on a smaller scale to explain mutation events in the DNA that codes for pA tracts (Cocquempot *et al.*, 2009)). Mispairing of short repeats that surround pA2 and unequal crossing over events have been suggested as a single mutation event (Kato *et al.*, 2003) that can explain both 24 bp duplications and 24 bp deletions that have been reported (Bienvenu *et al.*, 2002; Gronskov *et al.*, 2004; Troester *et al.*, 2007; Conti *et al.*, 2011b). But this mispairing and crossing over mechanism does not explain the 9 bp deletion that has been reported in the same region (Bienvenu *et al.*, 2002).

FosTeS was first proposed as the mechanism responsible for generating nonrecurrent chromosomal rearrangements (Lee *et al.*, 2007). FosTeS utilises microhomologies of just a few base pairs that act as bridges for the DNA replication fork to skip (forward or backward) along the chromosome when encountering complex genomic architecture or DNA lesions (Gu *et al.*, 2008). Given that microhomology does occur in the all pA tracts that are of similar DNA sequence as pA2 of *ARX* (Figure 6.6c) and that the high GC content of pA coding sequences may form complex architecture, it seems plausible that expansions result from FosTeS. This mechanism of mutation has been suggested as the basis of a duplication of a short segment of the imperfect tri-nucleotide repeat within another gene prone to pA expansion mutations in mouse, *Hoxd13* (Cocquempot *et al.*, 2009).

Given the range of sizes and the sequences involved in expansion mutations reported in pA1of *ARX* (1, 2, 3, 7 and 11 additional tri-nucleotide repeats) and *HOXD13*, unequal cross over between mispaired wildtype alleles has been suggested as the most likely mechanism that generates mutations in these pA tracts (reviewed by (Shoubridge and Gecz, 2011)).

Once the pA1 mutations are formed they are stably inherited and are also generated from sequences shorter than those usually required for DNA slippage, therefore they are also not likely generated from DNA replication slippage.

Whether the duplication mutations in the DNA that codes for pA2 of *ARX* results from unequal crossing over between two mispaired wildtype alleles (Warren, 1997; Kato *et al.*, 2003) or FosTeS (Lee *et al.*, 2007; Cocquempot *et al.*, 2009) remains to be determined.

6.7 Is the 24 bp duplication mutation recurrent or identical by descent?

To investigate if the duplication mutation is recurrent or identical by descent, haplotype testing across the Xp.22 region was undertaken (radiolabelled PCR per Chapter 2.4.2; primers in Table 2.3d) for four STS markers (*DXS8099-DXS8027-ARX-DXS12027-DXS8047*). These markers span Xp21.3 – Xp22.11, flanking the *ARX* locus (genomic positions as Table 6.3a). The overall chance that the duplication is identical by descent is 3.4% (Table 6.7a), which makes this scenario unlikely. Further, given that the c.429_452dup mutation has been reported to occur *de novo* (Bienvenu *et al.*, 2002; Gronskov *et al.*, 2004; Partington *et al.*, 2004) or as gonadal (Partington *et al.*, 2004) or somatic (Poirier *et al.*, 2005) mosaics, it is most likely that the majority of new ascertainments arise from recurrent mutations and are not identical by descent.

Whether the most common haplotype associated with the duplication mutation (*ie* 2-5-3-2 for *DXS8099-DXS8027-ARX-DXS8047-DXS1202*) confers an increased risk of the duplication mutation event occurring *de novo* remains to be determined.

Table 6.7a. STS marker haplotype for *DXS8099-DXS8027-ARX-DXS12027-DXS8047* at Xp21.3-Xp22.11 for probands with an *ARX* duplication mutation from 12 separate families.

Family	**Reference**	**Haplotype**			
		DXS 8099	***DXS 8027***	***DXS 1202***	***DXS 8047***
MRX-M	Strømme *et al*., 2002b	1	1	2	7
MRX-B	Strømme *et al*., 2002b	3	1	3	2
Aust PRTS	Strømme *et al*., 2002b	1	3	2	3
1	Partington *et al*., 2004	2	4	1	2
2	Partington *et al*., 2004	1	1	1	6
3	Partington *et al*., 2004	3	5	3	2
MRX-29	Stepp *et al*., 2005	2	5	3	6
MRX-32	Stepp *et al*., 2005	4	4	2	2
MRX-33	Stepp *et al*., 2005	2	5	3	2
MRX-38	Stepp *et al*., 2005	2	5	3	6
1 [1]	Reish *et al*., 2009	2	4	3	6
A	Fullston *et al*., 2011	1	3	3	1
	Most common haplotype	'2' (5/12)	'5' (4/12)	'3' (7/12)	'2' (5/12)
	Proportion	0.42	0.33	0.58	0.42

[1] Proband with c.430_456dup(27bp) mutation, this is the only family in this reference.
Overall the most common haplotype for *DXS8099-DXS8027-ARX-DXS8047-DXS1202* that also contains an *ARX* duplication is 2-5-3-2. The chance that these haplotypes are identical by descent is 3.38% (*ie* 0.42 x 0.33 x 0.58 x 0.42) if only the haplotype data is considered.

6.8 Discussion

Five families with duplication mutations in the DNA that codes for pA2 of ARX were presented, with a total of 7 affected males. Interestingly two families had an obligate carrier female with learning difficulties. Although skewed X-inactivation may be a contributing factor, it has been previously shown not to correlate with affected female status for polyalanine expansion mutations (Stromme *et al.*, 2002b; Marsh *et al.*, 2009). Including the three families with the c.429_452dup(24bp) mutation, a total of 44/110 (40%) reported families have this mutation in *ARX*. A further two mutations larger than the 24 bp duplication were discovered (c.430_456dup(27bp) and c.423_455dup(33bp)), and lead to more severe phenotypes within affected males than the recurrent c.429_452dup(24bp)

mutation. If the larger duplication mutations are considered, 46/110 (41.8%) families have a pA2 duplication mutation. This unique set of three distinct duplication mutations allowed for investigation of both mutant *ARX* sequences and mutant ARX protein localisation.

The length of homogenous polyalanine tracts within pA2 of mutant ARX aligned with phenotypic severity, as previously described for pA1 (Chapter 5; Figure 5.4b and Table 5.4a). Once the natural limit of 20A is breached, even by 1A (*ie* 21A; see Figure 6.5a and Table 6.5a), appreciably more mutant protein is mislocalised (*ie* c.430_456dup(27bp) and c.423_455dup(33bp)). Whereas a 20A pA2 (c.429_452dup(24bp)) did not cause mislocalisation (see Figure 6.5a and Table 6.5a), as previously reported (Shoubridge *et al.* 2007). The over-expression studies also reflect *in vivo* studies of a mouse model using a 20A pA2 (c.429_452dup(24bp) that also did not detect aggregation of the mutant protein (Kitamura *et al.*, 2009). The 20A pA2 mutation may be at the 'tipping point' and subtle shifts in genetic or cellular environment might exacerbate or ameliorate protein dysfunction and partly explain the variable clinical expressivity observed for this mutation (Stromme *et al.*, 2002a; Turner *et al.*, 2002; Kato *et al.*, 2004; Gestinari-Duarte Rde *et al.*, 2006; Szczaluba *et al.*, 2006; Laperuta *et al.*, 2007; Rujirabanjerd *et al.*, 2007; Cossee *et al.*, 2011).

The proportionate increase of mislocalisation as a consequence of increased pA tract length occurs for either pA1 or pA2 when 20A is surpassed. Curiously, the sub-classification of mislocalisation is distinct for pA1 (cytoplasmic positive with or without aggregates in either the nucleus or cytoplasm) and pA2 (nuclear inclusions only). This distinct pattern of mislocalisation depending on which pA tract is expanded may result from the total number of alanines in the respective mutant tracts, *ie* pA1 tracts mutants are 23A and 27A versus pA2 tracts of 20A, 21A and 22A (interspersed by a G). But conclusions about distinct mechanism of pathogenesis for pA1 or pA2 mutants cannot be drawn from these over-expression studies.

Primary structure investigations of the DNA sequences of the three duplication mutations allowed for speculation about secondary structures that may result and underpin the

mechanism that generates them. FosTeS and unequal crossing over are candidate mechanisms that might generate these in frame tandem duplications of 24 bp, 27 bp and 33 bp. But primary and secondary structure investigations are of limited value given the ambiguity that surrounds the precise nucleotides (nt) involved in the mutation, *ie* up to 12 distinct sets of 24 nt can be duplicated in the case of the c.429_452dup(24bp). This suggests that within the 44 families reported with the 24 bp duplication to date, the possibility remains that they represent up to 12 distinct co-ordinates of the 24 bp duplication.

It is clear that pA2 duplication mutations are unlikely to be identical by descent, as markers did not reveal a common haplotype for the mutation in 12 families tested. But this does not rule out the possibility that other *cis* acting elements catalyse the mutation event at a scale finer than that investigated by the STS markers used here (*ie* –1,437,280 bp to +484,788 bp). Such sequence elements have been suggested as important the FosTeS mechanism (Cocquempot *et al.*, 2009) and appear to be present within pA tracts with DNA sequence similarity to pA2 of *ARX*. If a pA2 duplication mutation arises during the cellular divisions of gameteogenesis or in the gonadal progenitor cell(s) in mosaics, it is then possible that gametes that are formed which carry the duplication mutation and may subsequently be inherited. If a duplication mutation is inherited through multiple generations (*eg* family 6.3a), it can be speculated that if enough of the proband's maternal lineage is available for investigation, the ancestral individual in which the *de novo* mutation event occurred in is likely to be discovered.

This study confirms pA2 as a 'hotspot' for mutations in *ARX* and that duplications are responsible for ~42% of all mutations. The mechanism that duplication mutations arise from probably relies on secondary structure features of the GC rich DNA sequence that surrounds and constitutes the pA2 coding region, precipitating the duplication of 24 bp, 27 bp and 33 bp. The resultant pA2 is expanded and confers pathology of increasing severity dependent on the length of homogenous pA tract.

CHAPTER 7 CONTENTS

The family investigations presented in this chapter have been published in Shoubridge *et al* (2010b) (Appendix A.9) and Fullston *et al* (2011) (Appendix A.2), as detailed below.

7 MUTATIONS IN THE HOMEODOMAIN OF ARX

Segregation testing, family investigations and cell based protein localisation studies of mutations in the DNA that code for the ARX homeodomain are presented within this chapter. DNA samples were collected and provided by collaborators as listed in the following sub-sections and relevant publications. Segregation testing, clinical descriptions and mutant protein localisations derived from these mutations have been published separately in Shoubridge *et al* (2010b; c.1136 G>T – p.R379L; Appendix A.9) and Fullston *et al* (2011; c.1074 G>T – p.R358S; Appendix A.2).

7.1 Introduction

ARX belongs to a family of 170 human homeobox genes that act in pathways that regulate development and differentiation processes (reviewed in (Shoubridge *et al.*, 2010a, Appendix A.1)). ARX is one of the 26 paired type homeodomain transcription factors that contain a conserved 60 amino acid DNA binding motif (homeodomain) and play essential roles in embryogenesis and neural development (Bienvenu *et al.*, 2002; Kitamura *et al.*, 2002; Stromme *et al.*, 2002b; Kato *et al.*, 2004). These proteins are often referred to as having Q_{50} paired type homeodomain, based on the invariant glutamine (Q) residue at position 50 of 60 amino acids that define the homeodomain. The ARX protein binds DNA via highly conserved homeodomain residues and acts as a bi-functional transcriptional activator and repressor (Collombat *et al.*, 2005; McKenzie *et al.*, 2007; Colasante *et al.*, 2009; Quille *et al.*, 2011).

Mutations in *ARX* are the only identified genetic cause underlying the distinct syndrome of X-linked lissencephaly with ambiguous genitalia (XLAG; MIM 300215). This may partly

explain the high frequencies of *ARX* mutations reported in patients classified as having XLAG (Kitamura *et al.*, 2002; Kato *et al.*, 2004). XLAG involves: (a) lissencephaly with a posterior-to-anterior gradient and intermediate (5–7 mm) increase in cortical thickness; (b) agenesis of the corpus callosum (ACC); (c) intractable epilepsy with onset prenatally or during the first hours of life; (d) hypothalamic dysfunction manifest by persistent poor temperature regulation; (e) severe hypotonia from birth; and (f) ambiguous genitalia (AG) in affected males (Berry-Kravis and Israel, 1994; Dobyns *et al.*, 1999). This syndrome differs from the other forms of lissencephaly, displaying only an intermediate thickening of the cortex with a posterior to anterior gradient of gyral malformation, and occurs in conjunction with the other features as above (Dobyns *et al.*, 1999).

There are currently 30 families reported with both brain and genital malformations due to 26 different mutations in *ARX* (reviewed in (Shoubridge *et al.*, 2010a, Appendix A.1); not including the families presented here). The majority of these cases are classified as X-linked lissencephaly with ambiguous genitalia (XLAG; MIM 300215) with or without additional features. In 23 of 30 (74%) of these families protein truncation mutations are predicted to cause the loss-of-function of the mature ARX protein (Kitamura *et al.*, 2002; Uyanik *et al.*, 2003; Hahn *et al.*, 2004; Hartmann *et al.*, 2004; Kato *et al.*, 2004; Bhat *et al.*, 2005; Jagla *et al.*, 2008; Okazaki *et al.*, 2008; Miyata *et al.*, 2009). These mutations are thought to lead to a loss-of-function of the ARX protein as when *Arx* was first deleted in mouse (Kitamura *et al.*, 2002), the phenotype mimicked many clinical aspects of XLAG (Berry-Kravis and Israel, 1994; Dobyns *et al.*, 1999). In the remaining 7 families (26%) with XLAG, single nucleotide substitutions mutate highly conserved key residues of the homeodomain or NLS2/3 that flank it (Kitamura *et al.*, 2002; Uyanik *et al.*, 2003; Kato *et al.*, 2004), with the exception of one mutation that occurred just outside of the aristaless domain (c.1561 G>A – p.A521T (Kato *et al.*, 2004)).

In NLS2 (N-terminal of the homeodomain) there are three naturally occurring point mutations which cause XLAG: c.994 C>T (p.R332C (Uyanik *et al.*, 2003)); c.995 G>A (p.R332H (Kitamura *et al.*, 2002)); and c.995 G>C (p.R332P (Kato *et al.*, 2004)). Also a fourth point mutation alters a threonine residue adjacent to this arganine residue, p.T333N (c.998 C>A) (Kato *et al.*, 2004), and causes Proud syndrome (ACC/AG; MIM 30004).

The basic residues within NLS2 and NLS3 are important for the import of ARX into the nucleus, via interactions with Importin 13 (IPO13) (Shoubridge *et al.*, 2007) and potentially other Importins (Lin *et al.*, 2009). IPO13 is a member of the importin-β superfamily involved in nuclear import and export of a variety of proteins (Mingot *et al.*, 2001; Ploski *et al.*, 2004), a process catalysed by RanGTP (Lee *et al.*, 2005).

7.2 A c.1074 G>T mutation alters a residue of the homoedomain (p.R358S)

As part of *ARX* mutation screening (Chapter 3) we discovered a mutation that alters a key amino acid residue that is conserved in all transcription factors that contain a paired class homeodomain (c.1074 G>T – p.R358S). The p.R358 residue is important for contacting DNA (reviewed in (Gecz *et al.*, 2006; Shoubridge *et al.*, 2010a, Appendix A.1)).

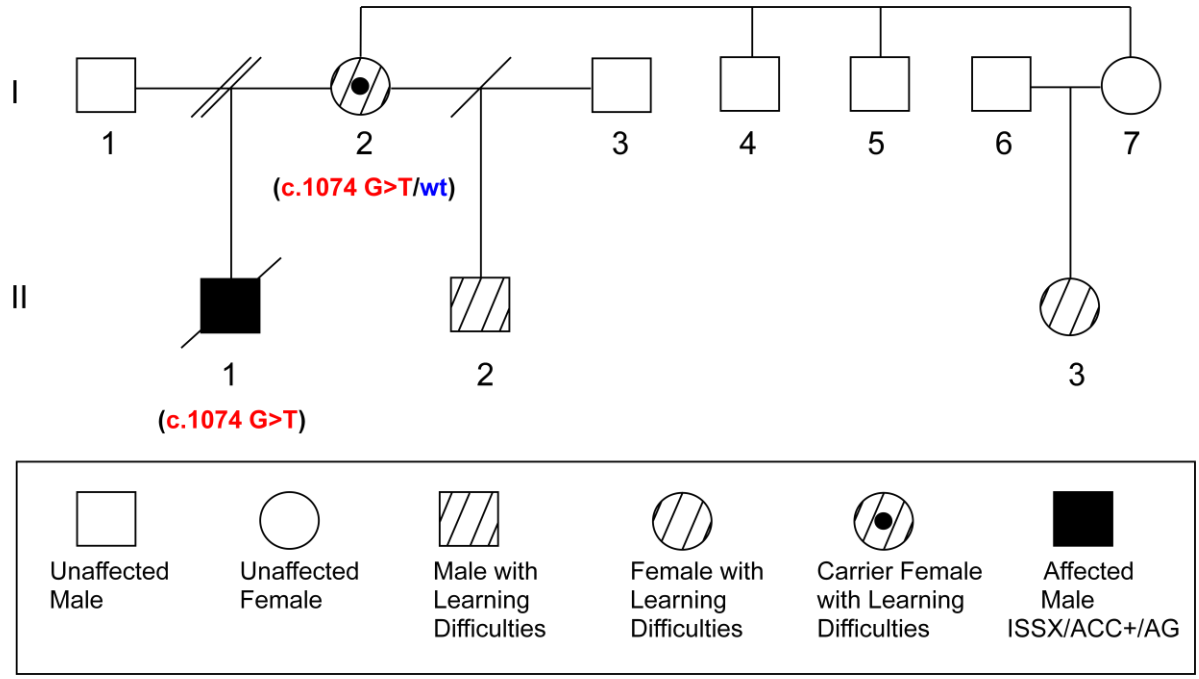

Figure 7.2a. Pedigree of family 7.2a with the c.1074 G>T (p.R358S) mutation discovered in the proband (II-1) and his mother (I-2) (Family D in Fullston *et al* 2011, Appendix A.2). DNA samples were not accessible for testing for the proband's half-brother (II-2), nor his maternal cousin (II-3). *ARX* genotype is indicated, for accessible samples, below the symbols that represent the proband (II-1) and his mother (I-2).

Family description supplied Dr Andrew Norman.

Dr Andrew Norman practices at the Clinical Genetics Unit in the Birmingham Women's Hospital, UK.

The proband, II-1 (Figure7.2a), was delivered by Caesarean section at approximately 34 weeks due to intrauterine growth restriction and poor fetal movement, with a birth weight of 2000 g and head circumference of 29cm (0.4^{th} centile). As an infant he was a poor feeder, reported as a 'floppy baby', suffering from severe physical disability and ID. He suffered congenital intractable diarrhoea, failed to thrive and had probable renal calculi. He was not grossly dysmorphic. A brain MRI revealed absence of the corpus callosum (ACC), widespread neuronal migration disorder, polymicrogyria and pachygyria. Infantile spasms (ISSX) and intractable epilepsy commenced at 3 months of age. An EEG displayed

hypsarrthymia. He was hypotonic and was severely developmentally delayed with weight and head circumference tracking at the 0.4th centile. He has micropenis and cryptorchidism (*ie* abnormal genitalia – AG). He had normal tone, could track and follow, and anecdotally would occasionally smile as a response. He died at 2 years and 3 months of age from intractable seizures and diarrhoea.

The mother, I-2 (Figure7.2a), had questionable parenting ability, learning difficulties and had concealed the proband's pregnancy. Both the half-brother (II-2; Figure7.2a) and female maternal cousin (II-3) have learning difficulties and attend a special school.

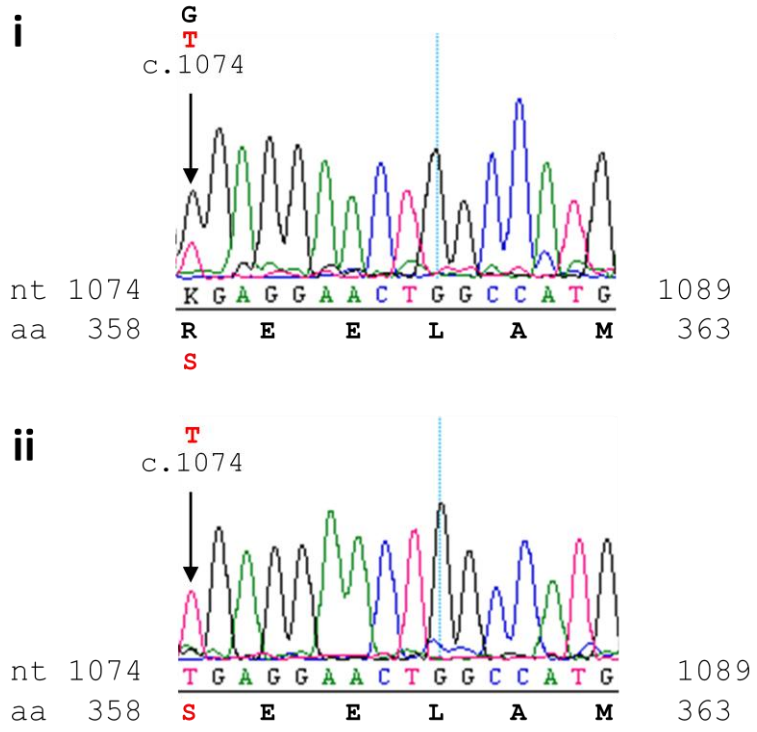

Figure 7.2b. DNA sequence chromatograms of partial sequence at the start of exon 3 of the *ARX* gene from (**i**) the heterozygous carrier mother (I-2) and (**ii**) the proband (II-1). Nucleotide (nt) c.1074 is highlighted above each chromatogram by an arrow, whilst the altered amino acid (aa) residue p.R358 is highlighted in red below each chromatogram with wildtype in black font and mutant in red. Amino acid residue p.R358 is important for DNA contact in proteins with paired class homeodomains.

Residue p.R358 is conserved in all paired class homeodomains and is involved in contacting DNA (reviewed in (Gecz *et al.*, 2006; Shoubridge *et al.*, 2010a, Appendix A.1)), which is essential for ARX to act as a transcriptional activator/repressor. Therefore an alteration of p.R358S is likely deleterious to ARX function and is the cause of pathology in the proabnd (II-1) and his mother (I-2).

7.3 A c.1136 G>T mutation alters a key residue of NLS3 (p.R379L)

ARX mutation screening (Chapter 3) discovered another mutation in the homedomain, this time it altered an amino acid residue within NLS3 (c.1136 G>T – p.R379L). The p.R379 amino acid, in combination with other basic residues, are key components of NLS3 that facilitates the nuclear import of the ARX protein (Shoubridge *et al.*, 2010b, Appendix A.9).

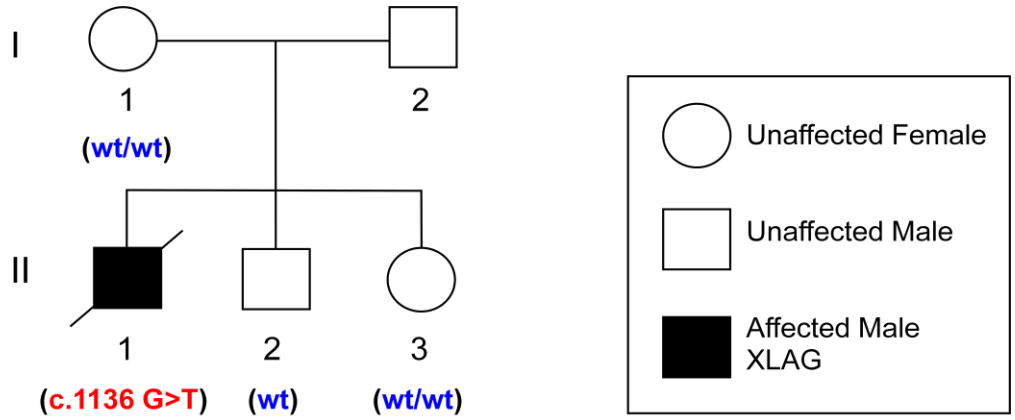

Figure 7.3a. Pedigree of family 7.3a with the c.1136 G>T (p.R379L) mutation discovered in the proband (II-1) but not in any other family members (Family 1 in Shoubridge *et al* 2010b, Appendix A.9). The proband's father's (I-2) DNA was not tested. *ARX* genotype is indicated below the symbols that represent the proband (II-1), his brother (II-2), sister (II-3) and mother (I-2).

Family description supplied Drs. David Coman and George McGillivray.

Drs. David Coman and George McGillivray are from the Murdoch Children's Research Institute, Melbourne, Australia.

The male proband (II-1; Figure 7.2a) presented with XLAG as a neonate. His clinical findings included dysmorphic features (high forehead, large anterior fontanelle, mild low set ears), ambiguous genitalia, hypotonia, a refractory seizure disorder, lissencephaly with posterior agyria and anterior pachygyria, agenesis of the corpus callosum, hypothalamic dysfunction (poor temperature control and hypothyroidism), hypophosphataemia, severe chronic diarrhoea, pancreatic insufficiency, small bowel malabsorption, megacolon and a nonspecific colitis. Magnetic resonance imaging (MRI) identified lissencephaly with agyria posteriorly transitioning abruptly to pachygyria anteriorly, agenesis of the corpus callosum, tiny fluid-signal intensity cystic structures in the putamen of the basal ganglia, and diffusely abnormal white matter (Figure 7.3b).

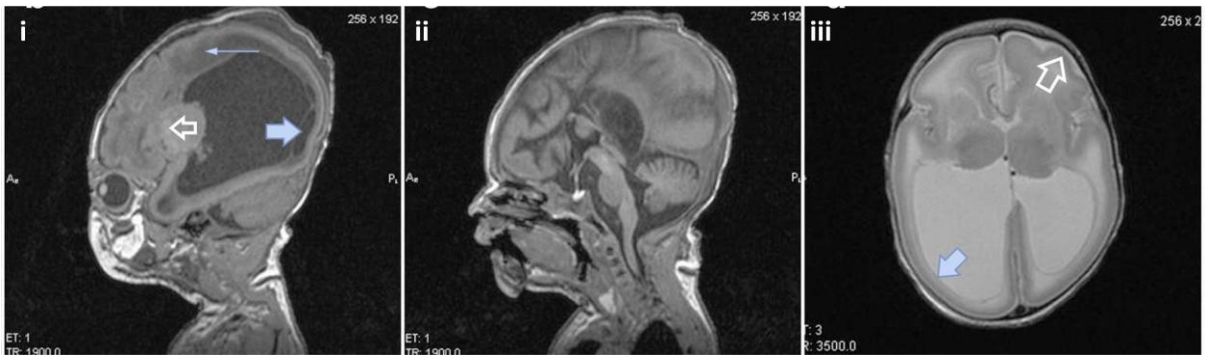

Figure 7.3b. Brain MRI findings from the individual from family 7.3a with XLAG and the c.1136 G>T mutation (II-1). Cerebrum at autopsy revealed posterior lissencephaly with transition to anterior pachygyria (not shown); (**i**) sagittal T1 MR sequence demonstrates lissencephaly, greater in the posterior than the anterior and marked lateral ventriculomegaly, note subependymal cysts (large solid arrow) abnormal periventricular white matter signal (small line arrow) and abnormal basal ganglia signal (large open arrow); (**ii**) sagittal T1 MRI sequence of the midline demonstrates agenesis of the corpus callosum; (**iii**) Axial T2 MR sequence demonstrates colpocephaly and lissencephaly with transition from posterior agyria (large solid arrow) to anterior pachygyria (large open arrow).

Chronic diarrhoea resulted in failure to thrive, dehydration, metabolic acidosis, and hypokalaemia, with presumptive evidence of small bowel malabsorption secondary to exocrine pancreatic insufficiency. Age appropriate developmental milestones were not attained and the patient died at 3 months of age.

Post mortem autopsy histological findings of the brain, pancreas and colon were performed. Most notably the pancreas showed nesidioblastosis with evidence of exocrine pancreatic insufficiency, and the colon revealed a patchy non-specific colitis with mild megacolon at the rectosigmoid junction and normal ganglion cells. Features consistent with XLAG include: (a) lissencephaly and agenesis of the corpus callosum (ACC); (b) perinatal encephalopathy with an intractable seizure disorder; (c) hypothalamic dysfunction; (d) ambiguous or underdeveloped genitalia; and (e) death early in infancy (Dobyns *et al.*, 1999).

Chronic diarrhoea has previously been reported in other XLAG patients (Hahn *et al.*, 2004; Kato *et al.*, 2004; Kato and Dobyns, 2005; Spinosa *et al.*, 2006; Nanba *et al.*, 2007), but the underlying cause is unknown. The persistence of faecal fat globules and the subsequent absence of tryptic activity in our patient suggests that small bowel malabsorption is secondary to exocrine pancreatic insufficiency. Exocrine pancreatic insufficiency has been described in 2 previous cases (Hahn *et al.*, 2004; Spinosa *et al.*, 2006), however the diarrhoea in this case remained intractable despite pancreatic enzyme replacement or total parenteral nutrition with bowel rest. Intractable diarrhoea has previously been reported comorbid with XLAG (Nanba *et al.*, 2007). In the proband vasoactive intestinal peptide was normal, and the administration of octreotide was effective in relieving the diarrhoea. We could not link the exocrine pancreatic insufficiency and the appearance of nesidioblastosis on pancreatic histology.

We identified a non-specific active colitis which may have been a secondary pathology. Mild megacolon was evident at the rectosigmoid junction in the presence of ganglion cells, consistent with previous reports (Ogata *et al.*, 2000; Spinosa *et al.*, 2006). In another XLAG patient in whom an intestinal biopsy was performed, results did not suggest a particular aetiology (Spinosa *et al.*, 2006).

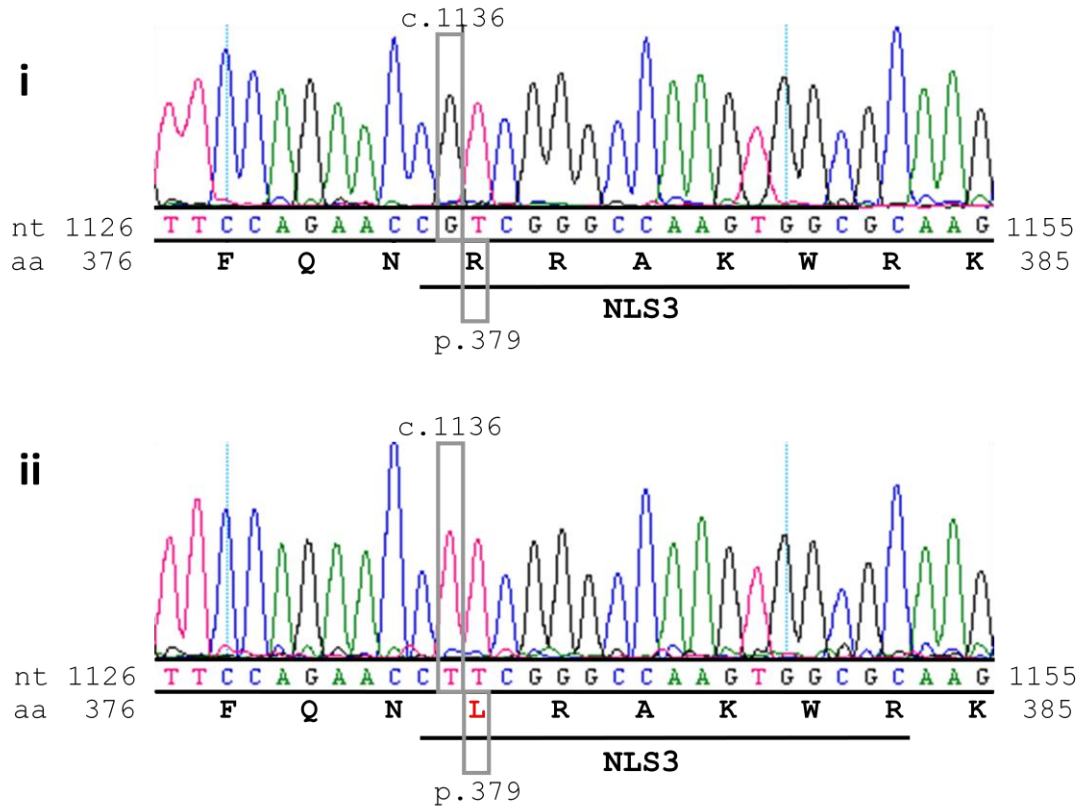

Figure 7.3c. DNA sequence chromatograms of partial sequence of exon 2 of the *ARX* gene from (**i**) the unaffected mother (I-1) and (**ii**) the proband with XLAG (II-1). Nucleotide (nt) c.1136 is highlighted above each chromatogram with a grey box, whilst the altered amino acid (aa) residue p.R379 is highlighted by a grey box below each chromatogram. The amino acid residues within NLS3 of ARX are underlined.

As the mutation is only present in the proband (II-1) and not his mother (I-1) or siblings (II-2 and II-3), it is likely due to a *de novo* mutation event. Residue p.R379 is conserved in all paired class homeodomains and is part of NLS3 that interacts with Importins and shuttle

ARX into the nucleus (reviewed in (Gecz *et al.*, 2006; Shoubridge *et al.*, 2010a, Appendix A.1)). Clearly nuclear import is essential for ARX to be able to bind target genes and transcriptionally activate/repress them. Therefore, an alteration to p.R379L is likely deleterious to ARX function and the pathology in this proband (II-1) is therefore most likely caused by this *de novo* mutation.

7.4 Homeodomain mutations cause ARX protein mislocalisation

Three different mutant homeodomain constructs (c.1058 C>T – p.P353L; c.1074 G>T – p.R358S; c.1136 G>T – p.R379L) were made from the *pCMV-myc-ARX-wt* vector by *in vitro* mutagenesis (Chapter 2.8) using mutagenic primers in Table 2.3f (PCR conditions; Chapter 2.4.5). Constructs were then transfected into human embryo kidney cells (HEK-293T; product number CRL-11268**,** ATCC, Manassas, USA) (Chapter 2.9) and subjected to fluorescent immunohistochemistry (Chapter 2.10) and standard microscopy (2.10 and 2.11). An expression vector was constructed for the c.1058 C>T (p.P353L) mutation and used as a control for a highly conserved homeodomain residue mutation, outside of NLS2/3, that causes a milder phenotype of X-linked myoclonic epilepsy with spasticity and intellectual disability (XEMSID (Scheffer *et al.*, 2002; Stromme *et al.*, 2002b)). An average of 725 ARX positive cells were scored per transfection across three separate transfection experiments. Localisation was classified as either normal, having nuclear inclusions, or cytoplasmic positive with or without aggregates in either the nucleus or cytoplasm (see representative images in Figure 5.4a) and compared to a wildtype *ARX* construct (Figure 7.4a and Table 7.4a).

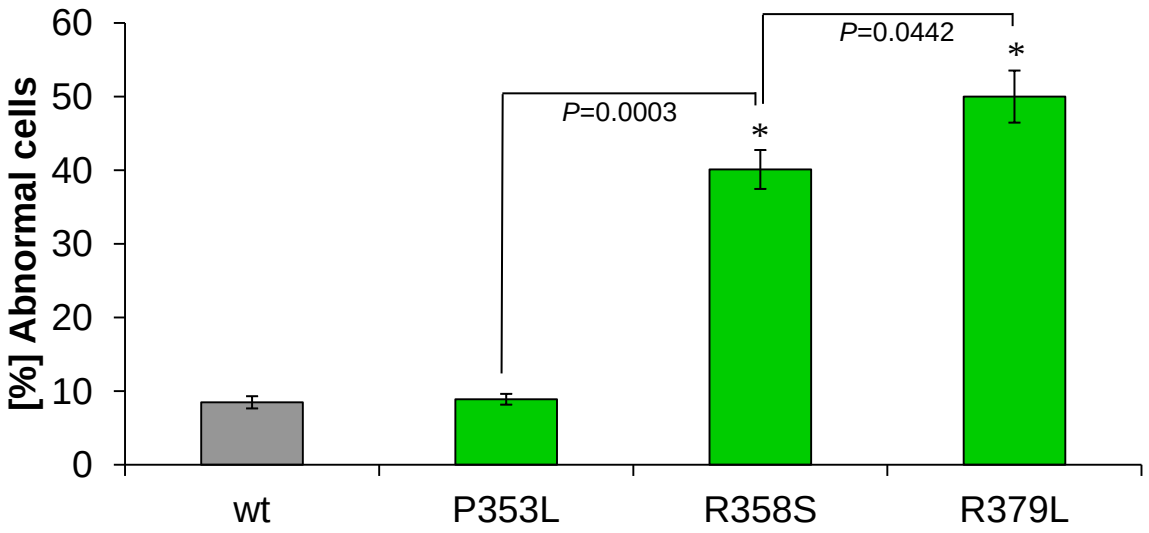

Figure 7.4a. Proportion of ARX positive cells displaying abnormal localisation (nuclear inclusions or cytoplasmic positive with or without aggregates in either the nucleus or cytoplasm) for three mutations in the homeodomain (green bars; c.1058 C>T – p.P353L; c.1074 G>T – p.R358S; c.1136 G>T – p.R379L) compared to wildtype ARX (wt, grey bar). An average of 725 ARX positive cells were scored per transfection for three separate transfection experiments 24 h post transfection. * Asterisks indicate which constructs had significantly different levels of protein mislocalisation compared to wildtype ($P \leq 0.05$; actual *P*-values can be found in Table 7.4a) and between mutant *P*-values are indicated (when $P \leq 0.05$), using a general linear model and an unbalanced ANOVA analysis.

The *ARX* homeodomain mutation, c.1136G>T (p.R379L), that leads to the most severe of *ARX* related disorders (XLAG) caused the most protein mislocalisation (Figure 7.4a and Table 7.4a). Whereas the homeodomain mutation, c.1058C>T (p.P353L (Stromme *et al.*, 2002b)) that causes a relatively milder phenotype (XMESID (Scheffer *et al.*, 2002)) did not alter protein localisation compared to wildtype. The c.1074G>T (p.R358S) mutation significantly increased the proportion of cells with mislocalised ARX protein compared to wildtype and the c.1058C>T (p.P353L) mutation, but significantly less than c.1136 G>T (p.R379L). The p.R358S mutation causes a relatively moderate phenotype (ISSX/ACC+/AG; Table 7.4a) compared to XLAG (c.1136G>T – p.R379L), but more severe than XMESID (c.1058C>T –p.P353L). Therefore mislocalisation of mutant homeodomain ARX protein aligns with phenotypic severity. Further, the major type of mislocalisation for these mutations was observed as nuclear inclusions (Figure 7.4b and Table 7.4a).

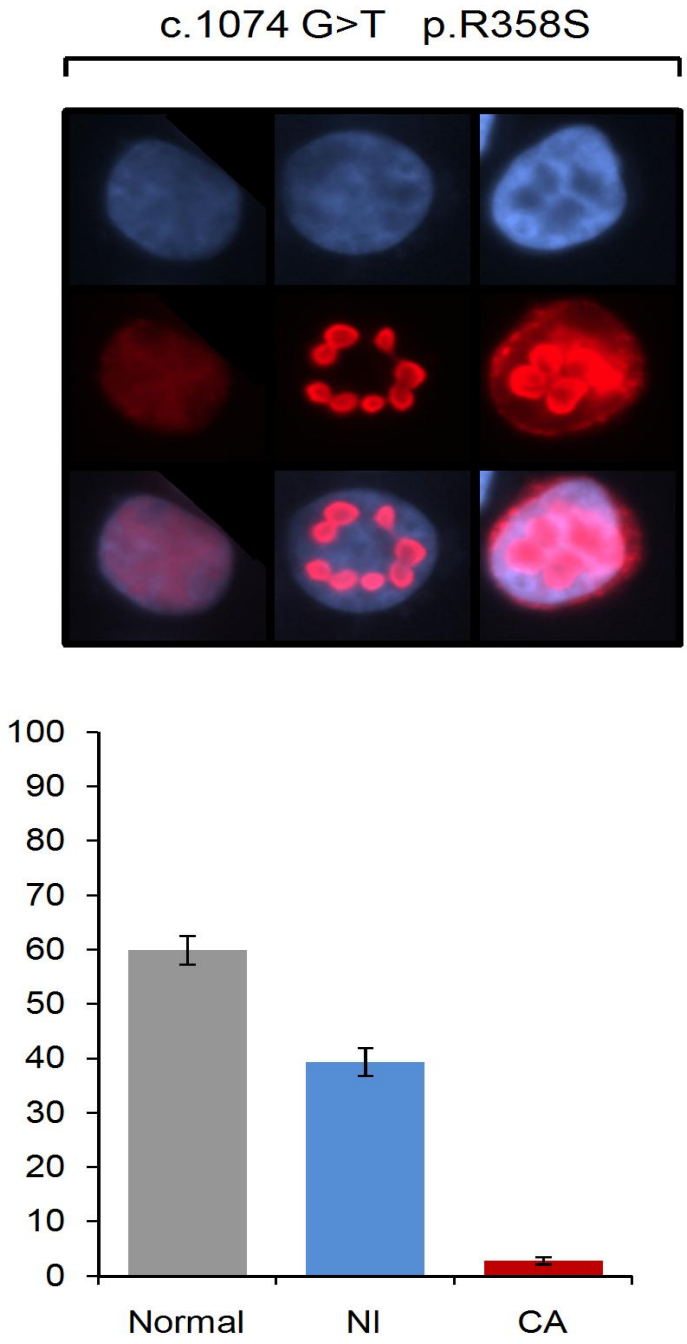

Figure 7.4b. Representative images of the localisation of the c.1074 G>T (p.R358S) mutant construct. Normal ARX localisation is shown in the leftmost three images of the upper panel (100x). Abnormal ARX localisation was classified as either nuclear inclusions only (NI; middle three images in the upper panel) or cytoplasmic positive with or without aggregates in either the nucleus or cytoplasm (CA; right most three images in upper panel). The images represent each classification (as labelled below the graph axis) and the proportion is shown in the graph below the representative images. Nuclear material (DNA) is stained with DAPI (top three images), ARX was detected by Cy3 in the TRITC channel (middle three images) and a merge of both is presented in the bottom three images. All images were produced from transfection of the pCMV-myc-ARX c.1074 G>T/p.R358S construct.

Table 7.4a. Amount of ARX protein mislocalisation per homeodomain mutant expression construct (*pCMVmycARX*) 24h post transfection.

Construct	Homeodomain residue altered	Abnormal localisation ±SEM (%)	*P* value (vs wt)[1]	NI [2] ±SEM (%)	*P* value (vs wt)[1]	CA [2] ±SEM (%)	*P* value (vs wt)[1]	*n* [3]	Average Cell count [4]	Phenotype
Wildtype	-	8.47 ±0.64	-	6.03 ±0.59	-	2.44 ±0.05	-	9	486	None
c.1058C>T	p.P353L	8.88 ±0.73	0.9995	6.04 ±0.36	0.9774	2.84±0.65	0.5258	3	627	XMESID [5]
c.1074G>T	p.R358S	**40.09 ±2.64**	**<0.0001**	**39.37 ±2.60**	**<0.0001**	**0.72 ±0.05**	**0.0080**	3	1032	ISSX/ACC+/AG
c.1136G>T	p.R379L	**49.99 ±3.54**	**<0.0001**	**49.19 ±3.81**	**<0.0001**	**0.03 ±0.28**	**0.0382**	3	757	XLAG

[1] *P* value for comparison of total amount of abnormal localisation or for each sub-classification to that of the wildtype construct by general linear model and an unbalanced ANOVA analysis. A *P* value ≤0.05 was considered significant.
[2] Sub types of mislocalisation observed for each ARX protein construct tested, classified as nuclear inclusions only (NI) or cytoplasmic positive with or without aggregates in either the nucleus or cytoplasm (CA).
[3] The number of transfection replicates performed for each construct.
[4] The average number of ARX positive cells counted for localisation per transfection (tfxn).
5 Reference; Strømme *et al* (2002b).
+ Additional phenotypic observations were made for patients with this type of mutation.
Bold font indicates a significant difference to wildtype.

7.5 Discussion

Homeodomain mutations displayed an association between phenotypic severity and degree of mutant protein mislocalisation. The non-malformation point mutation, c.1058 C>T (p.P353L), that causes XMESID (Scheffer *et al.*, 2002; Stromme *et al.*, 2002b) did not significantly alter the amount of protein mislocalisation compared to wildtype ARX. However the two point mutations that cause brain and genital malformation did significantly increase protein mislocalisation. The c.1136 G>T (p.R379L) mutation leads to the most severe phenotype (XLAG) and had the greatest amount of mislocalised protein, significantly more than c.1074 G>T (p.R358S).

The p.R358 residue is conserved in all paired type homeodomains and is important for contact with DNA targets (reviewed by (Gecz *et al.*, 2006; Shoubridge *et al.*, 2010a, Appendix A.1)). The pR358S (c.1074 G>T) mutant is probably reducing ARX's affinity for target sequences, impairing its transcriptional regulatory function. Because these localisation studies use an over-expression system (*ie* the CMV promoter) that drives a high level of expression, subtle shifts in subcellular localisation of proteins are not likely to be detected. As such, the lack of binding to DNA expected for the c.1058C>T (p.P353L) mutation, was not detected. This mutation did not alter protein localisation in this system, despite causing a phenotype similar to that of pA1 mutations that did alter mutant ARX protein localisation (Chapter 6). This limitation is further highlighted by the c.1136 G>T (p.R379L) mutation, which phenotypically mimics the *Arx* knockout mouse (Kitamura *et al.*, 2002) and is therefore presumably a null mutation, but does not exhibit 100% mislocalisation in this system.

The XLAG outcome the p.R379L is akin to a complete loss of function and is most likely due to aberrant cellular import, as previously described (Shoubridge *et al.*, 2007; Shoubridge *et al.*, 2010b, Appendix A.9). However, this mutation fails to abolish the binding of ARX to

IPO13, but diminishes the ability of the mutant ARX-IPO13 complex to uncouple (Shoubridge *et al.*, 2010b, Appendix A.9).

A partial loss of function due to a substitution of homeodomain proline (P) residue p.P353 with a similar hydrophobic non-polar neutral amino acid (*ie* leucine – L), as caused by the c.1058C>T (p.P353L) mutation (Stromme *et al.*, 2002b), results in a non-malformation phenotype (XMESID (Scheffer *et al.*, 2002)). Interestingly, when the highly conserved p.P353 residue is altered to a polar charged basic amino acid (*ie* arginine – R) by the c.1058 C>G (p.P353R) mutation, a complete loss of function underlies the XLAG phenotype (Kato *et al.*, 2004). This implies that although the p.P353 residue is highly conserved that some ARX function can be retained if the proline residue is altered to an amino acid with similar properties.

Conversely when the polar charged basic amino acid (*ie* arginine – R) at p.R379 is altered to a hydrophobic non-polar neutral amino acid (*ie* leucine – L) by the c.1136 G>T (pR379L) mutation XLAG results (as Chapter 7.3), *ie* presumed null mutation. But when p.R379 is changed to a polar uncharged neutral amino acid (*ie* serine – S) by the c.1135 C>A (p.R379S) mutation (Marsh *et al.*, 2009), a non-malformation phenotype of epilepsy with profound ID occurs in affected individuals (Marsh *et al.*, 2009; Shoubridge *et al.*, 2010b, Appendix A.9). This suggests if the amino acid at p.379 is polar that some ARX function is maintained, even if it is not basic, which is important for NLS2/3 interaction with IPO13 and subsequent nuclear import (Shoubridge *et al.*, 2007; Shoubridge *et al.*, 2010b, Appendix A.9).

This highlights the importance of highly conserved key residues in the homeodomain and the NLS2/3 that flank it, as the alteration of one residue can act as a null mutation and cause XLAG. But if the amino acid that is substituted is similar in molecular characteristics to the wildtype, some ARX function may be retained, and a relatively milder phenotype is likely to result.

Characterisation of a Duplication Involving the *ARX* Gene Locus (dup chrX:24,992,913_25,033,980)

CHAPTER 8 CONTENTS

The work presented in this chapter has been published as part of the Whibley *et al* (2010) publication (Appendix A.10).

8 CHARACTERISATION OF THE DUPLICATION

The descriptions of affected family members, array comparative genome hybridization summary data, qPCR confirmation, qPCR segregation testing, mRNA analysis, fine scale mapping and orientation of this *ARX* locus duplication (dup chrX:24,992,913_25,033,980; GRCh37 or hg19, Feb. 2009) are presented within this chapter. Clinical data and DNA sample collection and array comparative genome hybridisation analysis have been provided by collaborators listed in Whibley *et al* (2010) (Appendix A.10).

8.1 Introduction

Large-scale systematic re-sequencing of >7000 exons on the X chromosome in 208 families suggested that only ~25% of XLID was directly explained by obvious mutations within coding DNA (Tarpey *et al.*, 2009, Appendix A.4). Many missense variations of uncertain pathogenicity were also discovered and further familial studies based on individual missense mutations may subsequently confirm the pathogenetic nature of a proportion these, for example *IQSEC2* (Shoubridge *et al.*, 2010c). If a large proportion of the missense variations are indeed proven to be pathogenic in future reports, then as much as ~70% XLID maybe be explained by mutations within X-chromosome coding regions. Currently, XLID remains unexplained in the majority of families within this cohort. It remains possible that alterations to X-chromosome DNA which are not effectively detected by exon re-sequencing are likely candidates for pathogenicity.

One class of DNA mutation that are not effectively detected by direct Sanger exon sequencing is copy number variation (CNV). CNVs are defined as deletions or amplifications of genomic regions greater than 1kb in size and are more commonly found as duplications. Deletions and duplications smaller than 1kb are known as 'indels' (Scherer *et al.*, 2007). Duplications can not be detected by conventional PCR because they do not interfere with PCR amplification nor the sequencing reaction. X-chromosome deletions in DNA from males, where one or both of the primer binding sites correspond to deleted target sequences, would result in PCR failure. Whereas small deletions in between but not including primer binding sites would result in smaller than expected PCR products. PCR failure can occur for reasons besides deletion of target sequences and as such appropriate controls are needed to guard against misinterpretation of such results. But next generation sequencing techniques have the potential to overcome this misinterpretation if sufficient coverage can be achieved.

CNVs play a significant role in genetic disease, including neurological disorders (Lupski, 1998; Lee and Lupski, 2006) and ID (Qiao *et al.*, 2010). A recent meta-analysis of 33 original array comparative genome hybridisation (CGH) reports on a total of 21,968 individuals deduced an average diagnostic yield of 12.2% for cohorts including intellectual disability, developmental delay, dysmorphism and autism spectrum disorders (Miller *et al.*, 2010). Whole genome microarray has identified pathogenic CNVs in ID and is currently incorporated into routine clinical practice as a first tier approach (Cheung *et al.*, 2007; Stankiewicz and Beaudet, 2007; Koolen *et al.*, 2009; Miller *et al.*, 2010). Although CNVs recurrently associated with disease are known to be pathogenic (Sharp *et al.*, 2008), it is probable that significant numbers of novel non-recurrent or rare CNVs will be identified as either causing or conferring susceptibility to ID (Ionita-Laza *et al.*, 2009; Miller *et al.*, 2010).

A previous study discovered a large duplication in the murine genome that included the contiguous genes *Arx* and *PolA* that lead to a clear eye and forebrain phenotype (exencephaly, microphthalmia and anophthalmia). This provided evidence that a large CNV duplication incorporating the *ARX* locus can result in brain pathology (Cunningham *et al.*, 2002). Subsequently, two more duplications at Xp21-22 that include *ARX* and other known XLID genes have been associated with ID, but the relative contribution of *ARX* compared to the other genes remains unclear (Thorson *et al.*, 2010). Although there has been a report of an *ARX* locus deletion in a Yoruba Nigerian cohort, the report did not specify the sex or phenotype of the individual involved (Matsuzaki *et al.*, 2009). An XLAG phenotype most likely results in males hemizygous for an *ARX* deletion, as reported for other partial deletions (>1kb) of the *ARX* locus leading to a loss of function (Kitamura *et al.*, 2002; Kato *et al.*, 2004).

We used a custom 385K format high-density oligonucleotide X chromosome array (Catalogue # 05394538001, Roche-Nimblegen, Madison, Wisconsin, USA: 1 probe per 36 bp in targeted regions; median gap size of 463 bp) to investigate if CNVs caused X-linked intellectual disability in 251 families (Whibley *et al.*, 2010, Appendix A.10), the majority of which also had X chromosome coding exons sequenced by Tarpey *et al* (2009, Appendix A.4). A pathogenic CNV was discovered in 10% of the families (Whibley *et al.*, 2010, Appendix A.10). One such potentially pathogenic non-recurrent duplication of approximately 43 kb was discovered that involved the *ARX* locus and the most 3' end *POLA* exon 36 at Xp22 in a family with non syndromic XLID.

8.2 Family information and *ARX* gene locus duplication discovery

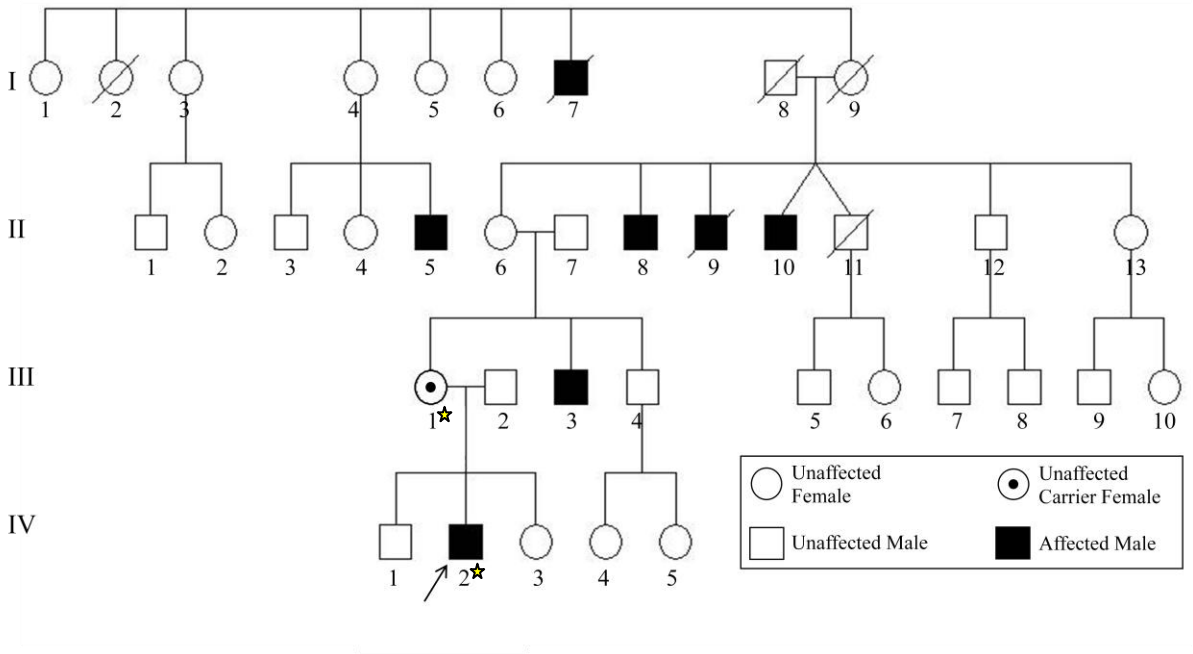

Figure 8.2a. Pedigree of the large XLID family investigated for the *ARX* gene locus duplication (dup chrX:24,992,913_25,033,980; GRCh37/hg19, Feb. 2009). The proband is indicated by an arrow, and individuals with the duplication are marked with a star (see section 5.3)

The family (Figure 5.2a; iGOLD # 505) is a large family displaying an X-linked mode of inheritance for ID. Overall the affected males displayed nsXLID, with no history of epilepsy, but some individuals displayed some gait disturbance and dystonia.

Clinical notes – provided by Drs. Jenny Morton and Louise Brueton

Drs. Louise Brueton and Jenny Morton are from the Clinical Genetics Unit, Birmingham Women's Hospital, Edgbaston, Birmingham, UK.

The proband, IV-2 (Figure 5.2a), was delivered by an uncomplicated birth with a head circumference at the 75th centile. He commenced walking at 7 years of age in an unsteady

manner that was not frankly ataxic. MRI and CT scans of his brain returned normal results. He has poor speech capabilities and has limited reading capacity, but cannot write.

The proband's mother (III-1 Figure 5.2a) and brother (IV-1) were both unaffected. A DNA sample from the unaffected brother was used as a presumably wildtype (wt) hemizygous *ARX* control (*ie* one copy of *ARX*). Whereas the mother served as a control sample containing 3 copies of the *ARX* locus, as she was presumably an obligate carrier of the duplication. These DNA samples served as internal copy number controls for qPCR confirmation, segregation testing and duplication orientation testing.

The uncle of the proband, III-3 (Figure 5.2a), has an IQ of 67 and lives in residential care. He walked at 3 years of age and was at the 10^{th} centile for height and the 50^{th} centile for head circumference. He was unable to read or write and displayed no spastic movements.

Anecdotally the proband's great uncle, II-8 (Figure 5.2a) was said to have speech problems, learning disability and a history of unusual gait (walking straight-legs, possibly spastic).

DNA and LCL samples were made available for the proband (IV-2, Figure 5.2a) along with DNA samples for the mother (III-1) and brother (IV-1). The proband's genomic DNA was interrogated by a comparative hybridisation using a custom 385K format Nimblegen oligonucleotide array (Whibley *et al.*, 2010, Appendix A.10). Resulting probe intensities surrounding the *ARX* locus at Xp22 are given for the proband (Figure 5.2b), normalised to the average of four control DNA samples. It is interesting to note that some probe samples actually returned a loss of DNA signal and is likely due to the high GC content of the DNA that codes for and surrounds the *ARX* locus confounding intensity signals (*eg ARX* exon 2 in Figure 5.2b).

.

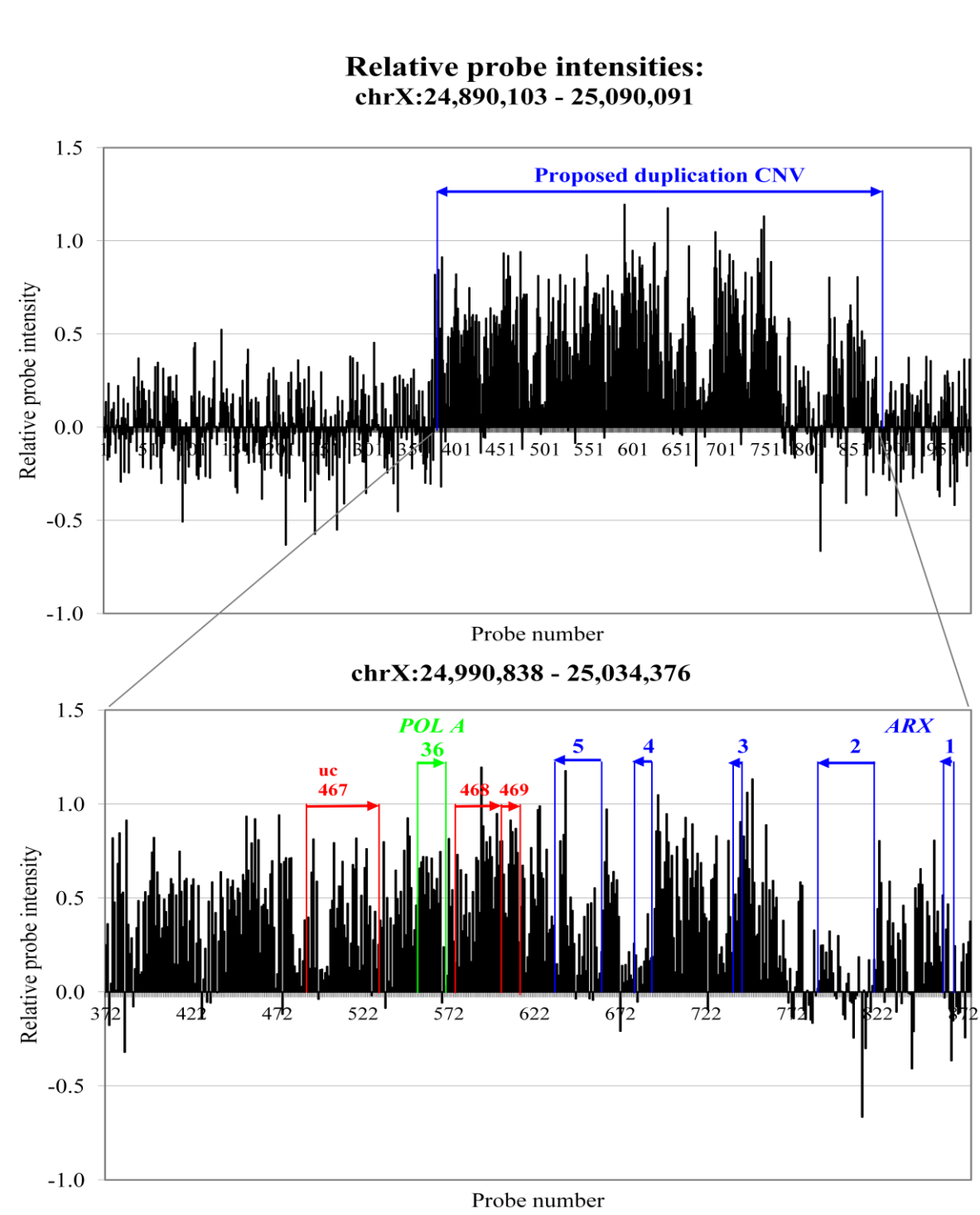

Figure 8.2b. Relative Xp22 probe intensities from the comparative genome hybridisation using the proband's (IV-2) DNA on a custom 385K format Nimblegen oligonucleotide array. Probe intensity is expressed as log base 2 values, and probe number was arbitrarily assigned from 1 along the X axis, starting at chrX: 24,890,103 (GRCh37 or hg19, Feb. 2009). Positions of *ARX* (blue arrows), *POLA* (green arrows) exons and ultraconserved elements (uc; red arrows) are indicated in the lower panel.

8.3 qPCR confirmation and segregation testing

Confirmation of array CGH results and segregation of the duplication were sought by qPCR on genomic DNA from the proband (IV-2), his unaffected brother (IV-1) and his mother (III-1, Figure 5.3a). qPCR amplicons were designed to probe sequences that flanked either end of the putative duplication (5' - 371/372, 3' - 876/877), all 5 *ARX* exons and exon 35/36 of *POLA* (primers in Table 2.3e). Once it was determined by qPCR that neither 5' region (probes 371/372) initially tested was duplicated in the proband (2x copies) or his mother (3x copies), qPCR amplicons more 3' were designed (*ie* probes 373, 374, 375, 376, 377 and 378) until a signal consistent with a duplication was observed for both proband (*ie* 2x copies) and his mother (*ie* 3x copies) (Figure 5.3a).

This strategy was not necessary for the 3' probes as 876 was sufficiently close (528 bp) to the qPCR amplicon of *ARX* exon 1 for standard PCR to span from one probe to the next. 25 ng genomic DNA from family members (IV-2, IV-1, III-1 Figure 5.3a) was used for qPCR input and amount of fluorescence was used to determine copy number relative to a standard curve derived from using varying amounts of DNA from an unrelated phenotypically normal male (2.5 ng, 25.0 ng, 50 ng, 75 ng, 100 ng inputs). All qPCR assays were normalised to a VIC labelled *RNaseP* MGB taqman assay (catalogue number 4316844, Applied Biosystems, Melbourne, Australia) and a graphical summary of all qPCR assays performed is given in Figure 8.3a.

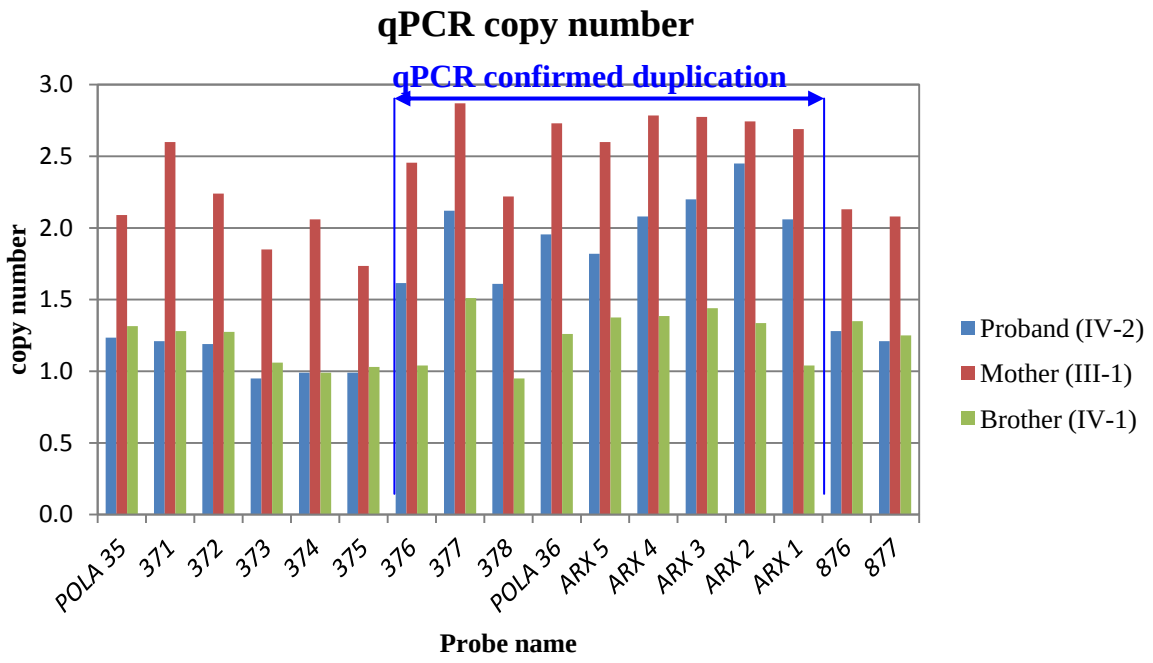

Figure 8.3a. Summary of qPCR assays to refine the size of the *ARX* gene duplication. 25 ng of genomic DNA was used normalised to a VIC labelled *RNaseP* MGB taqman assay (Applied Biosystems, Melbourne, Australia). qPCR amplicons spanned chrX:24,948,571-25,035,027 and the probable extent of the duplication was refined to a minimal region of chrX:24,993,012-25,033,858 (build GRCh37/hg19 Feb. 2009). Data shown is an average of 1-3 technical repeats per qPCR assay (average of 1.6x replicates).

qPCR data confirmed that the proband (IV-2) harboured a duplication, his unaffected mother (III-1) was a carrier and his unaffected brother (IV-1, Figure 5.2a) did not carry the duplication. The duplicated region is at most 41 kb in size (amplicons 376 to *ARX* 1, Figure 5.3a). Although the qPCR amplicons did not give perfect 2x or 3x copy number signals as expected, the ratios of proband or mother to the unaffected brother remains indicative of the extent on the duplication. This data suggested that all *ARX* exons were duplicated along with exon 36 of *POLA* (but not exon 35) and presumably uc467, uc468 and uc469. Also, the duplicated region was further refined from 43,539 bp (chrX:24,990,838-25,034,376) to a minimal region of 40,847 bp spanned by the 376 (5') and *ARX* exon 1 (3') amplicons (chrX:24,993,012-25,033,858).

8.4 Orientations and fine mapping of the duplication

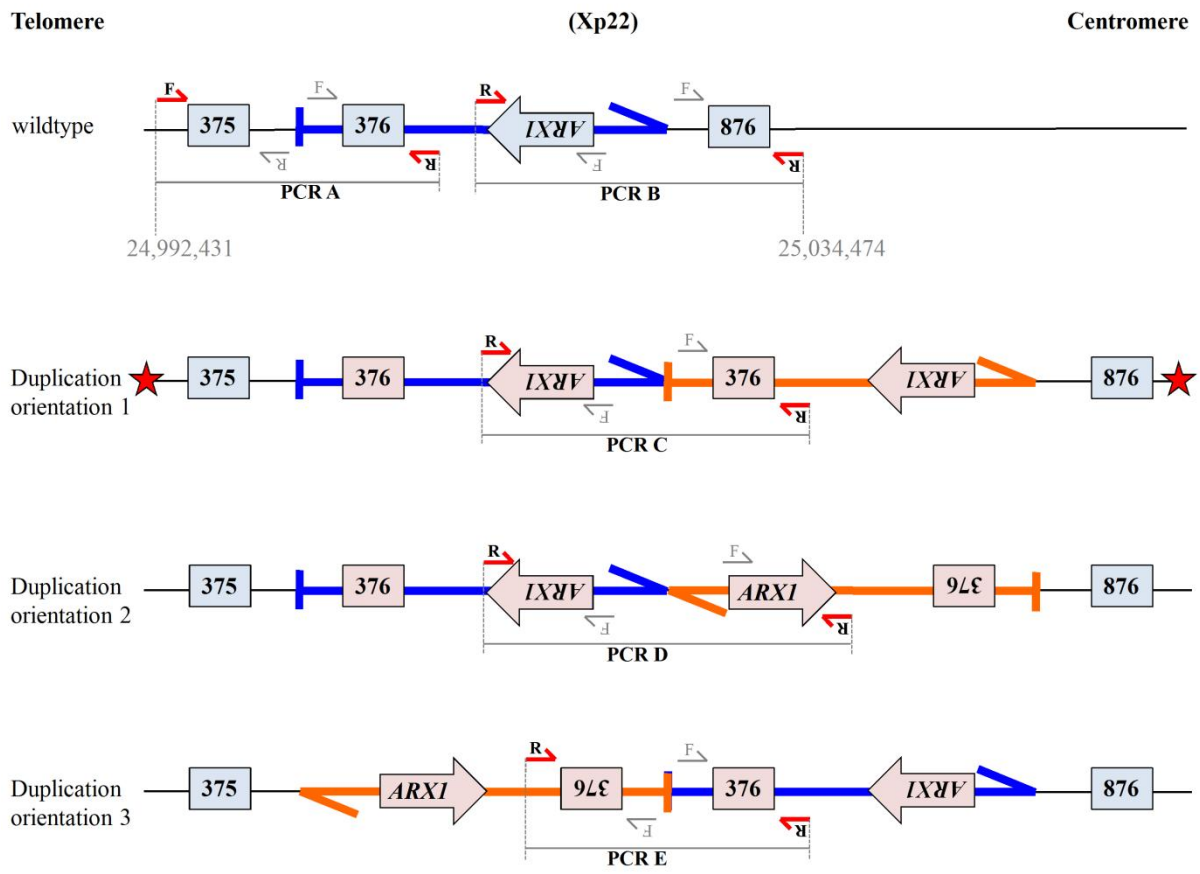

Figure 8.4a. Schematic of 3 theoretically possible orientations of the *ARX* duplication at Xp22 and PCR strategies used to resolve them. UCSC genome browser nt numbers are below the wildtype (wt) copy (build GRCh37/hg19 Feb. 2009). Solid blue arrows indicate the wt *ARX* locus, whilst solid orange arrows represent the duplicated copy (blunt end 5', arrow head end 3'). Shaded boxes/arrows represent qPCR amplicons (Blue 1x copy number, Pink 2x copy number; by qPCR). Small arrows above and below the boxes represent qPCR primers, where Red arrows represent the primer used for each PCR. PCRs C, D and E are specific to the proposed rearrangement, whilst PCR A and B will produce the same size PCR product as a wt DNA sample. Red stars indicate the orientation of the duplication.

The maximum estimated distance between wt and duplicated qPCR amplicons was less than 1 kb, therefore standard PCR was used. A PCR strategy was devised to investigate the orientation of the duplication, which would ultimately map the duplication location to a single nt (Figure 5.4a). PCR C, D and E were initially undertaken to determine if the duplicated region was in orientation 1 (tandem), 2 (inverted) or 3 (respectively; orientation 3

seemed the least likely (Hastings *et al.*, 2009)). These PCRs were specific to the proposed rearrangements (orientations 1, 2 and 3) capitalising on the unique proximity of terminal qPCR amplicons within the duplicated region. As PCR C, D and E were performed with either two reverse primers, or in the case of orientation 2, 3 a single reverse primer, the unaffected brother (IV-1) was used as a negative control.

Only PCR C yielded an amplification product, with a size of slightly larger than 400 bp (Figure 5.4b) from DNA templates containing the duplication, implying a contiguous tandem duplication (*ie* not duplicated and incorporated into a different region of the genome). This PCR product was subsequently purified and directly sequenced. To further ensure the duplication was confined to the original *ARX* locus PCR A and B were performed with the family trio and an unrelated control DNA. A control DNA was necessary because orientation 1 will produce the same size product as wt sequence for PCR A and B, only if the duplication lacked complex micro-rearrangements at the terminal ends. PCRs A (681 bp, albeit faint) and B (717 bp) gave products in the family trio that were the same size as a wt control DNA input, indicating that the duplication was in a contiguous tandem repeat without complex micro-arrangements at the terminal ends (orientation 1, Figure 5.4a).

The sequence from the product of PCR C was positioned using BLAT (UCSC genome browser) and aligned using the seqman software within the Lasergene package (DNASTAR, Madison, USA). The duplication extended over 41,068 bp and was positioned to chrX:24,992,913-25,033,980 (GRCh37/hg19, Feb. 2009) in wildtype sequence by UCSC.

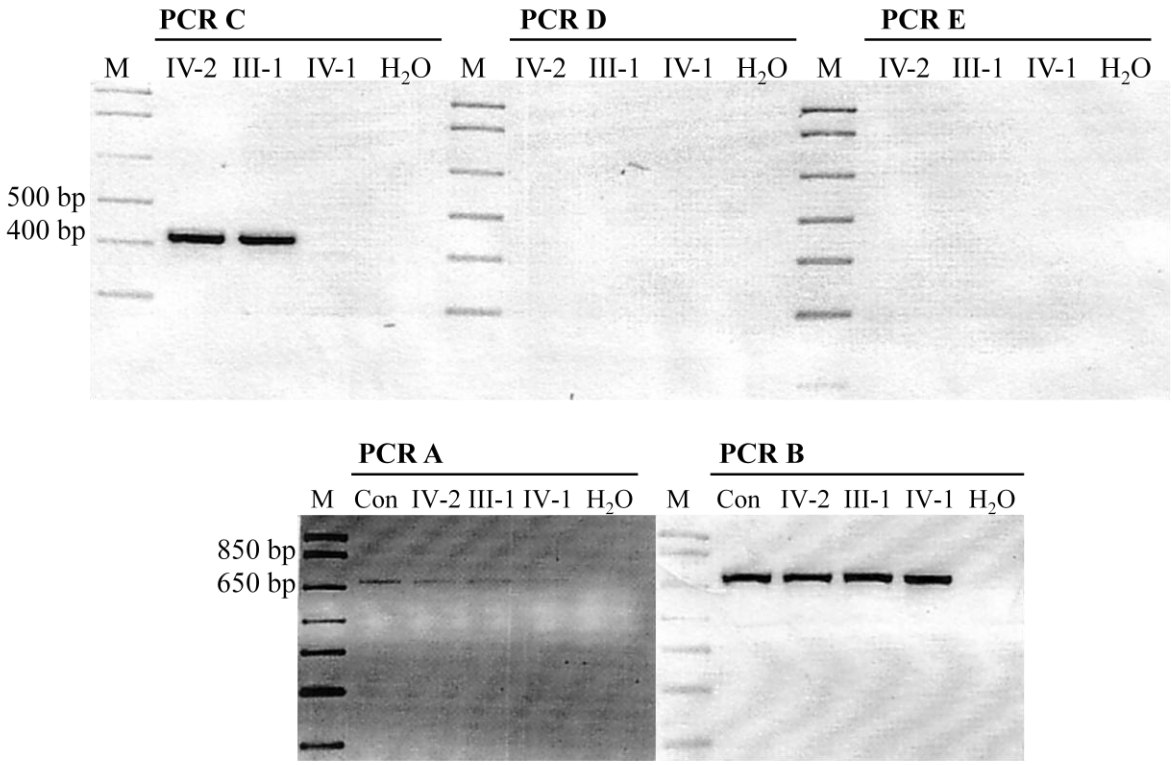

Figure 8.4b. Products from orientation specific PCRs (as per Figure5.4a), separated in a 2% agarose gel stained with ethidium bromide. M marker DNA (1kb+ DNA ladder, Invitrogen, Melbourne, Australia), Con control DNA input, IV-2 proband's DNA input, III-1 mother's DNA input, IV-1 brother's DNA input and H_2O is water input. Relevant band sizes within the marker lane are indicated to the left of the gels. PCRs C, D and E were run on a single gel, whilst PCRs A and B were run on a separate gels. Gels were visualised under UV light and negative images were cropped and resized to produce this figure.

8.5 Investigation of *ARX/POLA* mRNA in patient's LCLs

To determine if the expression of *ARX* or *POLA* was effected by the duplication (dup chrX:24,992,913_25,033,980) a combination of both qPCR and RT-PCR specific to scrambled *ARX* mRNA was employed on RNA extracted from two separate LCL pellets derived from the proband (IV-2), four unrelated control LCLs and fetal brain cDNA. Overall *ARX* qPCR was not successful, probably due to a combination of low mRNA abundance in LCLs and high GC content of the cDNA target. qPCR was attempted on four separate occasions with different *ARX* amplicons (to avoid GC rich regions), differing target cDNA dilutions, multiple RNA input amounts and multiple RNA extractions from separate LCL

pellets. Although the data repeatedly indicated an increase in *ARX* mRNA in the proband's LCLs relative to the average of four control LCLs, the assays failed quality control measures (*eg* such as poor reaction efficiency). One possible experimental option not exhausted was the use of Taqman probes, which are conventionally thought to allow more sensitive detection of mRNA of low abundance with guaranteed high reaction efficiency (assays Hs00292465_m1, Hs00992303_m1 and Hs00417686_m1 from Applied Biosystems, Melbourne, Australia). *POLA* expression, as tested by exon 35 and 36, normalised to *β-Actin*, was not affected by the duplication in RNA from the proband's LCLs (data not shown).

Given the contiguous tandem orientation of the duplication (orientation 1, Figure 5.4a) it remained possible that a scrambled *ARX* mRNA transcript resulted from exon splicing from the first (wt) copy of *ARX* to the second duplicated copy (*ie* exon 4/5 of the first copy to exon 2 of the duplicated copy). To test this hypothesis nested RT-PCR was performed by using two distinct pairs of primers sequentially; (a) the first set of primers from exon 4 (wt) to exon 2 (duplication) (forward, 4F; reverse, 2R; Table 2.3e) for 40x thermocyles of a standard GC rich PCR but with a 5 min extension step; (b) followed by a second round of PCR with a more 3' forward primer and a more 5' reverse primer (forward, 4F 2; Table 2.3e; reverse 2J R2; Table 2.3a) using the first PCR product as a template for a further 40x (same thermocycle).

Although a complex banding pattern was observed for fetal brain cDNA (lane B Figure 5.5a), probably an artifact of the abundance of *ARX* mRNA in this tissue, LCL samples gave no discernible bands. This was despite that RT-PCR of *ARX* exon 4 to 5 successfully amplified from all LCL cDNAs using only the initial PCR (as '(a)' above; 4F 2, 5R). This suggests that an *ARX* mRNA scrambled from exon 4 to 2 does not exist in the dup

chrX:24,992,913_25,033,980 proband. But this data is defined by an absence of a nested RT-PCR band, albeit after a total of 80x PCR thermocycles, and using patient material with notoriously low abundance of *ARX* mRNA.

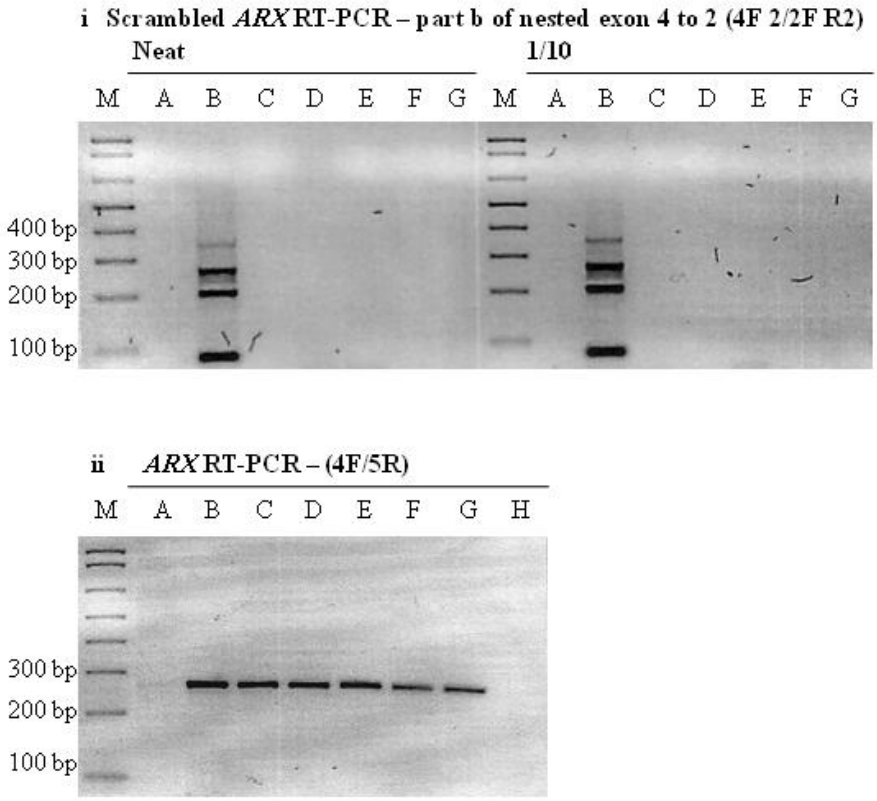

Figure 8.5a. Products from a RT-PCR specific to either **(i)** scrambled *ARX* mRNA (exon 4 to 2; nested) or **(ii)** *ARX* wt mRNA. PCR products are separated in a 2% agarose gel stained with ethidium bromide. **(i)** 1ul inputs are products from first tier of nested PCR either neat or diluted 1/10 PCR: M marker DNA (1kb+ DNA ladder, Invitrogen, Melbourne, Australia), A control genomic DNA, B fetal brain cDNA, C control LCL cDNA, D proband (IV-2) LCL cDNA, E proband (IV-2) LCL cDNA (separate LCL RNA extraction), F fetal brain cDNA (without RT enzyme added), G H_20 only. **(ii)** Standard *ARX* RT-PCR subjected to the same PCR conditions as only the first tier of the nested PCR: M marker DNA, A control genomic DNA, B fetal brain cDNA, C control LCL cDNA 1, D control LCL cDNA 2 E control LCL cDNA 3, F proband (IV-2) LCL cDNA 1, G proband (IV-2) LCL cDNA 2, H fetal brain cDNA (minus RT). M marker DNA Relevant band sizes within the marker lane are indicated to the left of the gel. The gel was visualised under UV light and the negative image was cropped to produce this figure.

8.6 Discussion

This is the first well characterised duplication associated with the *ARX* locus (Whibley *et al.*, 2010, Appendix A.10). The mutation segregates within the three members of the family tested (a mother and her two sons) and encompasses the entire *ARX* locus including exon 36 of *POLA* and a known *ARX* enhancer region (Colasante *et al.*, 2008) flanking uc467-469; Figure 5.2b). The proband was part of the Tarpey *et al* (2009, Appendix A.4) sequencing of the coding exons of 718 genes on the X chromosome. Although the average coverage was approximately 75% there was no evidence of a pathogenic sequence variation in the proband for any other potential XLID gene, and this initial approach was not expected to detect the duplication of the *ARX* locus. Given the incomplete coverage of the Tarpey *et al* (2009, Appendix A.4) approach, it is possible that a pathogenic variation in a known XLID gene may be detected within this family in the future.

Subsequently, array comparative genome hybridisation revealed a duplication at chrX:24,990,838 (GRCh37/hg19, Feb. 2009) in the proband. The duplicated region was refined to 41,068 bp and mapped to a single nucleotide resolution. The duplication was present in the mother of the proband and absent in an unaffected brother (Whibley *et al.*, 2010, Appendix A.10). The limited segregation testing and *ARX* expression data leaves open the possibility that this duplication occurred *de novo* in the mother.

Despite multiple approaches, whether *ARX* expression is altered due to this duplication remains inconclusive. Perhaps a combination of more sensitive qPCR technology such as Taqman probes and patient material with a greater abundance of *ARX* mRNA (epithelial, nasal epithelia, muscle or gonad) can address this issue. It is conceivable that *ARX* mRNA scrambling can occur but this could not be effectively detected due to low *ARX* mRNA abundance in LCLs, despite a total of 80 thermocycles of amplification.

The orientation of this duplication is in tandem, which is the most likely configuration with respect to the proposed mechanisms that cause CNVs (reviewed in (Hastings *et al.*, 2009)). This finding is consistent with the majority of duplications found on the X chromosome that

are in tandem repeats with very short regions of homology (Whibley *et al.*, 2010, Appendix A.10).

It can be speculated that a potential increase in *ARX* expression is due to the duplication caused by synergy between both an in tandem doubling of the *ARX* coding region and duplication enhancer regions that flank ultraconserved elements (uc467, uc 468 and uc 469, Figure 8.2b). Two fragments of DNA that have been shown to enhance *Arx* expression by the binding of the Distal-less 2 protein (DLX2) flank these uc elements in mouse (Colasante *et al.*, 2008). It is clear that mutations in *ARX* impact negatively on its function, presumably by either a partial or complete loss of transcriptional regulation of downstream targets (reviewed in (Gecz *et al.*, 2006; Shoubridge *et al.*, 2010a, Appendix A.1)). But whether a presumed increase in *ARX* expression can impart ID and a reduction in motor skills observed within this family remains to be determined. Nonetheless the observed phenotype falls within the spectrum of disorders reported for *ARX* mutations, most similar to that reported for the majority of individuals with the c.429_452dup(24bp) mutations, *ie* nsXLID with or without motor skill impairment (Dystonia; reviewed in (Gecz *et al.*, 2006; Shoubridge *et al.*, 2010a, Appendix A.1))

To date, less than half of XLID families are explained by mutations in known XLID genes on the X-chromosome (Tarpey *et al.*, 2009, Appendix A.4). Within the cohort of families with a clear X-linked mode of inheritance for ID without obvious mutations, we discovered approximately 10% have pathogenic CNV by array CGH (Whibley *et al.*, 2010, Appendix A.10). As CNV surveillance increases and further information pertaining to the variation of genomes within 'normal' individuals is gathered (*ie* 1,000 genome project (Durbin *et al.*, 2010)), more pathogenic CNVs associated with ID are likely to be discovered. There are 110 families with 50 different mutations reported within the *ARX* ORF (reviewed by (Gecz *et al.*, 2006; Shoubridge *et al.*, 2010a, Appendix A.1) and reports by (Giordano *et al.*, 2010; Kato *et al.*, 2010; Conti *et al.*, 2011a; Cossee *et al.*, 2011; Eksioglu *et al.*, 2011; Fullston *et al.*, 2011, Appendix A.2)). Given the amount of *ARX* ORF mutations that cause XLID, it is likely that additional disease causing CNVs apart from the duplication described here will be discovered, adding to the spectrum of *ARX* related pathologies.

CHAPTER 9 CONTENTS

9 FINAL DISCUSSION AND CONCLUSION

A total of eight families with six mutations were described here, comprising of approximately 1.3% of the heterogeneous cohort of 613 individuals screened. To date, including these families, a total of 110 have been reported to harbour one of 50 mutations in *ARX* (Gecz *et al.*, 2006; Giordano *et al.*, 2010; Kato *et al.*, 2010; Shoubridge *et al.*, 2010a, Appendix A.1; Conti *et al.*, 2011a; Cossee *et al.*, 2011; Eksioglu *et al.*, 2011; Fullston *et al.*, 2011, Appendix A.2). It is likely that more individuals/families have been identified by diagnostic services worldwide that remain unreported to the public domain.

9.1 *ARX* screening strategies and recommendations

An *ARX* screening strategy based on clinical presentations is recommended. This strategy would see milder ID pathologies with or without seizures prioritised for screening for pA1 and pA2 mutations and the entire *ARX* ORF screened for brain/genital malformation phenotypes. Regardless of the disorder the entire *ARX* ORF should be investigated for mutations, as an appreciable number of mutations occur outside of exon 2.

But for new pathologies to be discovered that are not currently related to *ARX* mutations screening of individuals with disorders outside of the typical *ARX* phenotypic spectrum is required. For example, Ohtahara syndrome was recently discovered to be caused by seven different mutations in *ARX* in some individuals (Kato *et al.*, 2007; Absoud *et al.*, 2010; Fullston *et al.*, 2010, Appendix A.6; Giordano *et al.*, 2010; Kato *et al.*, 2010; Eksioglu *et al.*, 2011). Interestingly all mutations, besides the c.81 C>G presented here, either expand pA1 or alter amino acids in the homeodomain of ARX. Further, autism was also associated with variations in ultraconserved elements and nsXLID and ataxia has been associated with a duplication of the *ARX* locus. This suggests that *ARX* screening should not be limited to the

ORF, but also extended to both non-coding regulatory regions and CNV analysis. Next generation sequencing (NGS) techniques are capable of sequencing entire loci, surrounding regulatory regions and CNV analysis in one procedure, perhaps even for the entire X chromosome for X-linked disorders. But, whole X-chromosome exome NGS re-sequencing data from the Neurogenetics laboratory (Figure 9.1a; J. Gécz, personal communication) indicates that sequence capture followed by NGS does not reliably cover regions of high GC content (*ie ARX*). Improved sequence coverage may well result from whole genome sequencing (WGS) instead of targeted sequencing as WGS does not require sequence enrichment by hybridisation. Given DNA with exceedingly high GC regions, such as exon 2 and 4 of *ARX*, require specific conditions for successful amplification and sequencing (Chapter 2.4.2 and 2.4.4) it is likely that mutations that occur in these regions could be missed using NGS technologies. Moreover, given the propensity for mutations in exon 2 of *ARX*, perhaps other GC rich regions within the human genome will similarly harbour mutations yet to be discovered.

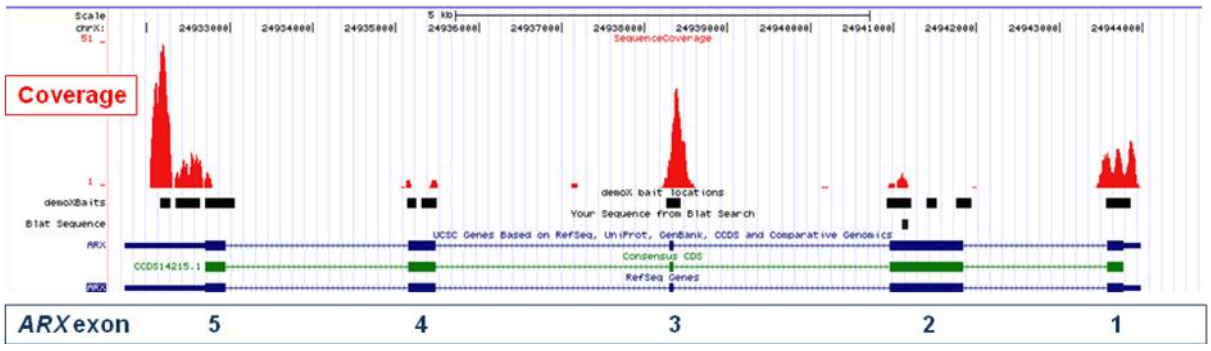

Figure 9.1a. UCSC genome browser image of human *ARX* locus with a custom exome sequencing coverage track enabled (chrX: 24,931,548 – 24,944,880; Mar 2006 build – GRCh36/hg18). Average sequencing coverage for whole exome sequences from >300 individuals is indicated above the exons of *ARX* in red and displays low coverage for high GC content exons 2 and 4.

During the studies presented here a number of *ARX* sequence variants identified were deemed non-pathogenic (Chapter 3.5). These variants were either silent nucleotide substitutions or contractions in pA1. It remains possible that although these variations alone do not confer obvious pathology, they may still act as modifiers. For example, the conservative c.1047 C>T polymorphism creates a strong binding site for the SRp55 protein (Gronskov *et al.*, 2004), which belongs to a family of proteins that bind to exonic splice enhancers (Screaton *et al.*, 1995). This may impact on *ARX* mRNA abundance and possibly contribute to ID. Contractions in pA1 may impact on ARX protein function, in a manner that does not consistently precipitate pathology in all individuals but may predispose some individuals to pathology.

9.2 A complex genotype-phenotype relationship exists

Generally a genotype-phenotype relationship exists, whereby mutations that do not truncate the ARX protein or are outside of the homeodomain (*eg* c.81 C>G, c.429_452dup(24b), c.430_456dup(27bp) and c.423_455dup(33bp)) lead to non-malformation phenotypes (reviewed by (Gecz *et al.*, 2006; Shoubridge *et al.*, 2010a, Appendix A.1)). In contrast mutations that do truncate the protein or alter residues in the highly conserved homeodomain (*eg* c.1074 G>T, c,1136 G>T), cause brain and genital malformation phenotypes (reviewed by (Gecz *et al.*, 2006; Shoubridge *et al.*, 2010a, Appendix A.1)). However the mutations discovered and characterised here highlight the complexities that exist between the mutations, molecular mechanism, genetic background and possibly environment.

For example a mutation that should cause an early truncation of the ARX protein (c.81 C>G – p.Y27X) does not cause the XLAG phenotype expected from a presumably null mutation. Males with the c.81 C>G (p.Y27X) mutation are likely somewhat 'rescued' by a rare poorly characterised mechanism of translation re-initiation. The efficiency of this mechanism likely relies on many factors outside of the *ARX* genotype, such as genetic background and possibly environmental cues. Further, the largest duplication in the DNA that codes for pA2 (c.423_455dup(33bp)) of ARX does not induce the most severe phenotype observed for the three duplication mutations reported in pA tract 2. Despite causing the largest expansion of pA2 reported to date, the resultant 22 alanine (A) tract is interspersed by a single glycine (G) residue that appears to ameliorate the dysfunction of this protein. The resultant phenotype is less severe than that caused by the mutant homogenous 21A tract induced by the c.430_456dup(27bp) mutation. Also, amino acid substitutions of key residues in the homeodomain by amino acids with similar molecular properties cause phenotypes less severe than the expected XLAG.

While there is substantial evidence for a genotype/phenotype correlation for *ARX* mutations, one should remain vigilant and cautious with respect to interpreting genotype-phenotype relationships, especially for predictive purposes. As in each genetic background the mutation may present differently, leading to different clinical outcomes in different individuals.

9.3 *ARX* related female phenotypes are complex

As part of these studies three carrier females were reported as having a discernible phenotype, *ie* learning difficulties (Families 6.2b, 6.2c and 7.2a). The carrier female

phenotype was less severe than in males with the same mutation in all cases. Interestingly the female phenotype was the similar in all three females, regardless of the severity of the affected male phenotype (*ie* c.1074G>T – brain and genital malformations; c.429_452dup(24bp) – ID). Female phenotypes are not reported as frequently as affected males, presumably due to the inconsistencies and complexities of carrier female phenotypes. For example mothers heterozygous for mutations that lead to of XLAG or Proud syndrome phenotypes in 14 probands are unaffected, whilst 66.7% (6/9) of other heterozygous female relatives were affected (Shoubridge *et al.*, 2010a, Appendix A.1). The under-detection of affected carrier females may also simply reflect an ascertainment bias (Marsh *et al.*, 2009). Previous studies suggest that no correlation exists between affected carrier females and unfavourable skewing of their X-inactivation (Stromme *et al.*, 2002b; Marsh *et al.*, 2009).

ARX is subject to X-inactivation, which results in the random inactivation of one X-chromosome in each cell of a female. *ARX* mutations that confer a partial loss of function do not consistently cause a carrier female phenotype. For example, due to X-inactivation 50% of cells will express the wildtype ARX whilst 50% will express mutant ARX. If the mutant retains 50% of normal function, overall 75% ARX function would be retained. But in the case of a mutation that results in a null *ARX* allele, an overall function of 50% is retained, and a carrier female phenotype is more likely. Perhaps brain tissue is more capable of overcoming a partial loss of ARX function in 50% of cells than a complete loss of function in 50% of cells. But carrier female phenotype remains more complicated as only half of female mice with a conditional knockout of *ARX* displayed a phenotype (Marsh *et al.*, 2009). Further complications arise when selection pressures on cells that express mutant *ARX* alleles, genetic modifying factors and environmental cues are considered. Although males are typically the focus of *ARX* mutation screening, further investigations into the pathogenesis of mutations in carrier females are warranted.

9.4 Abnormal protein localisation results from mutations

Abnormal protein localisation investigated in these studies, in general, paralleled phenotypic severity. This is despite the *in vitro* over-expression system used to investigate *ARX* mutations is far removed from an *in vivo* setting. The mirroring of phenotypic severity was observed for mutations that altered the same domain of the ARX protein, but were discordant for mutations impacting upon different domains. For example the c.1058 C>T (p.P353L) mutations causes XMESID, which is a similar phenotype to the infantile spasms and ID caused by the c.304ins$(GCG)_7$ (p.A111insA_7) mutation. But these two mutations have different impacts on mutant ARX protein localisation: *ie* c.1058 C>T (p.P353L) did not cause protein mislocalisation; whereas the c.304ins$(GCG)_7$ (p.A111insA_7) mutation did. Further, a presumably null mutation (c.1136G>T – p.R379L; due to inducing an XLAG phenotype) caused mislocalisation of mutant protein in 50% of cells, whereas a 100% reading would be expected from an ideal assay for a null mutation. The over-expression system also produces aggregates, even for wildtype ARX, when to date no evidence of mutant ARX aggregates have been found *in vivo* (Price *et al.*, 2009; Shoubridge and Gecz, 2011).

Polyalanine tracts have a natural limit of 20A (Albrecht and Mundlos, 2005) and the cell based protein localisation studies confirmed this limit. If either pA1 or pA2 of the ARX protein exceeded 20A in length the protein mislocalised significantly. Given that the c.429_452dup(24bp) duplication mutation expands pA2 to 20A, it may represent a 'tipping point' at which the effects of genetic background and environmental inputs can either exacerbate or ameliorate the mutant's molecular impact. As these inputs differ between individuals, perhaps this partly explains the vast variability observed for the clinical expressivity of this mutation (Stromme *et al.*, 2002a; Turner *et al.*, 2002; Gestinari-Duarte Rde *et al.*, 2006; Szczaluba *et al.*, 2006; Laperuta *et al.*, 2007; Rujirabanjerd *et al.*, 2007).

9.5 The DNA sequence of *ARX* is inherently prone to mutation

Here it is proposed that the primary sequence of the DNA that codes for pA1 and pA2 of ARX has inherent properties that may induce secondary structures that precipitate the most frequent of *ARX* mutations (*ie* c.304ins$(GCG)_7$ and c.429_452dup, respectively). A further consideration is although alanine (A) is coded by GC(N), the GCG codon is significantly over represented (Albrecht and Mundlos, 2005), which significantly increases the GC content of pA tracts, as highlighted in *ARX*.

For example within *ARX*: the 48 bp that code for pA1 has a 100% GC content, which is expanded to 69 bp of 100% GC content by the c.304ins$(GCG)_7$ mutation; and the 36 bp that codes for pA2 has a 97.2% GC content, which is extended to 60 bp of 96.7% GC content by the c.429_452dup duplication mutation. Both of these mutations increase the overall GC content of exon 2 from 72.5% to greater than 75%. These regions may well become increasingly difficult to replicate once the mutations have been incorporated and indeed poor or inconsistent replication plagues the amplification of such GC rich regions *in vitro*. Whether replication difficulties arising from high GC content of mutant *ARX* are sufficient to contribute to pathology *in vivo*, regardless of protein dysfunction, remains unclear. Perhaps, the increased GCG content of the *ARX* mRNA could cause toxicity, as is suggested for the triplet repeats in the premutations range for *FMR1* mRNA in FXTAS (Handa *et al.*, 2003; Hashem *et al.*, 2009). Further, given the extra CpG dinucleotides added to CpG island 130 (Figure 9.6a) by mutations expanding pA1/pA2, aberrant methylation impacting on *ARX* expression may also have an increased likelihood.

Indeed if an expanded number of CpG islands in exon 2 of *ARX* can impact upon its expression through hyper-methylation, there may be further parallels to the concurrent hyper-methylation and silencing of the *FMR1* gene that occurs in fragile-X syndrome (reviewed in (Song *et al.*, 2003)). Further, the DNA that codes for pA2 of HOXD13 has a sequence similar to that of pA2 of ARX (Chapter 6.6). An increase in methylation at an insulator region of *Hoxd13* associated with decreased expression in specific tissues (Yamagishi *et al.*, 2007), whether methylation occurs within the DNA that codes for pA2 of HOXD13 was not

determined. Moreover, epigenetics is an important consideration as altered epigenetic states are known to associate with impaired neurocognitive outcomes (reviewed in (Miller, 2010)).

9.6 Is *ARX* subject to regulation by environmental cues?

A variable clinical expressivity is observed both within and between families with the same *ARX* mutation (Stromme *et al.*, 2002a; Turner *et al.*, 2002; Kato *et al.*, 2004; Gestinari-Duarte Rde *et al.*, 2006; Szczaluba *et al.*, 2006; Laperuta *et al.*, 2007; Rujirabanjerd *et al.*, 2007; Cossee *et al.*, 2011). Although differences in genetic background may partially explain the phenotypic differences, variable phenotypic expressivity is also observed in an inbred mouse model where genetic background would have less impact (Kitamura *et al.*, 2009). A number of candidate CpG islands intersperse the *ARX* locus, including some within promoter and 3' enhancer elements, the majority of which are methylated in four different cell lines (Figure 9a). It is therefore possible that differential methylation of these regions could impact on *ARX* expression. Further, it remains possible that sequence variations of these methylated regions may also impact upon their methylation status and hence *ARX* expression.

Methylation of promoter elements 5' of *ARX* (CpG 74 and 28 in Figure 9a) has been correlated with *ARX* repression in pancreatic β–cells (Dhawan *et al.*, 2011). Methylation is sensitive to environmental cues specific to an individual even given similar genetic backgrounds, for example littermates (Michaud *et al.*, 1994) and identical twins (reviewed by (Bell and Spector, 2011)) raised in similar environments have different methylation profiles. Differential methylation may also impact on *ARX* expression during brain development between individuals and yield variable phenotypic outcomes dependant on an individual's specific environment both *in utero* and postnatally (reviewed by (Milosavljevic, 2011)). This remains a largely unexplored area for *ARX*, with the potential to explain at least some of the phenotypic heterogeneity.

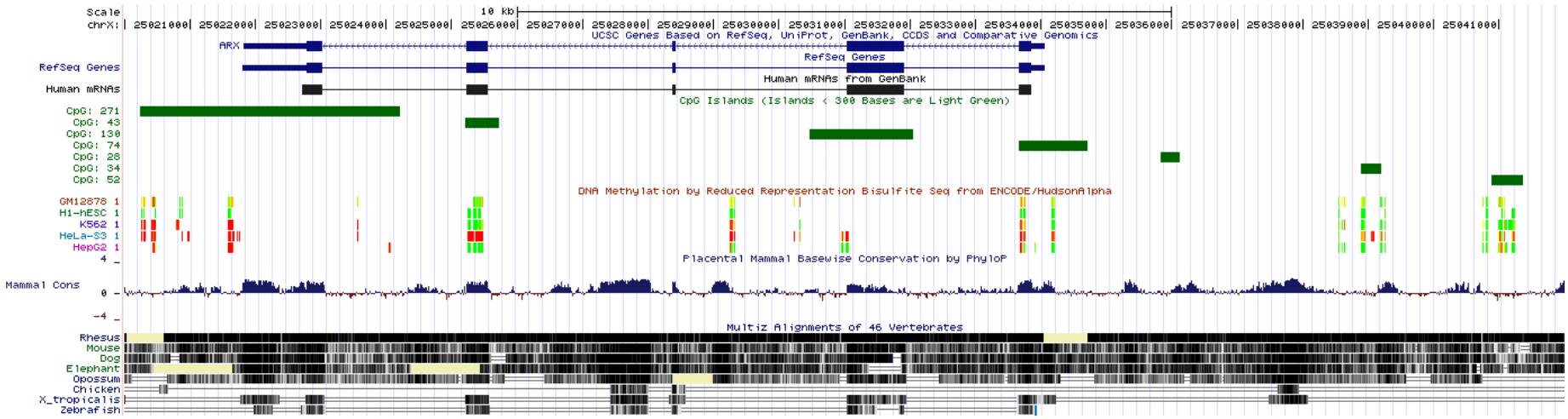

Figure 9.6a. UCSC genome browser image of human *ARX* locus (chrX: 25,020,000 – 25,042,000; Feb 2009 build - GRCh37/hg19). Known CpG islands are shown in dark green and sites with reduced representation by bisulphite sequencing (methylated Cytosine residues) from various cell lines are shown in light green, yellow and red. The conservation across mammalian genomes is shown in black, summarised for all mammals shown in blue.

9.7 Aberrant *ARX* over/under-expression may confer pathology

Both presumed under and over-expression of *ARX* are implicated in human disease and both have been demonstrated to impact on the expression or ARX targets (Quille *et al.*, 2011). For example, it is possible that the uc element variants presented here might reduce *ARX* expression through a potential reduction to enhancer activity, although expression studies were inconclusive. These variants and a speculated reduction in *ARX* expression were associated with patient's phenotypes of autism and infantile spasms. The reduction of *ARX* expression has been suggested as a mechanism responsible for phenotypes observed in female carriers and interestingly infantile spasms have also been previously observed in carrier females (Marsh *et al.*, 2009). Contrastingly, over-expression of *ARX* that presumably results from a duplication of the *ARX* locus is associated with an ID and ataxia phenotype. This suggests that an increase in ARX abundance, and presumed activity, may also lead to pathology.

If aberrant *ARX* expression is implicated in human disease, it would be preferable to investigate it using material from primary sources (*ie* not derived cell lines) with sufficiently abundant *ARX* expression (epithelium, muscle, gonad, brain). Such materials are only rarely collected from patients, with informed parental consent (often post mortem). If a 'bio-bank' could be developed for such materials to be deposited into and made available, *ARX* expression and methylation studies could potentially be performed on primary source material. As ethical and practical reasons make the collection and investigation of primary material challenging, alternative derivative sources could be utilised instead. Such materials could include neuronal cell lines derived from induced pluripotent stem (iPS) cells.

9.8 Concluding remarks

Although no treatment exists for *ARX* related pathologies as yet, further study of the fundamental molecular networks and pathways common to *ARX* pathogenesis may provide novel targets for such therapeutic interventions in the future. Such investigations have led to the first promising 'open-label' trial for treatment of individuals with fragile-X syndrome (Erickson *et al.*, 2011). This pharmaceutical intervention may be the first example of a bottom up (*ie* first gene discovery, then animal model pathophysiology and then human therapeutics) corrective treatment for a neurobehavioural disorder (reviewed by (Krueger and Bear, 2011)). The absence of FMRP protein leading to an excess of mGluR5 synthesis and signalling was crucial in using an mGluR5 antagonist as a therapeutic intervention. This avenue of treatment was only made available by the discovery of the underlying molecular pathology of fragile-X syndrome.

The role of methylation in the regulation of *ARX* expression throughout development and its potential role in variable clinical expressivity of *ARX* mutations remain to be investigated. The new generation of sequencing technologies provides the platform which can now be applied to detect CNV, sequence, methylation and expression variations on a genome wide scale that may expand the phenotypic spectrum of *ARX* related disorders further.

It is likely that new variations in the *ARX* ORF, regulatory and promoter regions are yet to be discovered. The identification of *ARX* mutations and the expanding knowledge of their molecular pathogenesis have been of great benefit for the patients and their families through the precision this knowledge now brings to their genetic counselling, even though no therapy currently exists. Further, knowledge of the complexities surrounding some of these mutations is helpful to the medical professionals who are involved with the care of these patients. Ultimately studies like those presented here contribute to the expanding knowledge of the spectrum of *ARX* disorders and are likely to provide a basis for future therapies to improve the quality of life for affected individuals.

APPENDIX A. PUBLICATIONS ARISING FROM THIS PROJECT

A.1 Shoubridge, C., **Fullston T.** and Gecz J. **2010a**. 'ARX spectrum disorders: making inroads into the molecular pathology' *Hum Mutat* **31**(8): 889-900.

A.2 **Fullston, T.**, Finnis M., Hackett A., Hodgson B., Brueton L., Baynam G., Norman A., Reish O., Shoubridge C. and Gecz J. **2011**. 'Screening and cell-based assessment of mutations in the Aristaless-related homeobox (ARX) gene' *Clin Genet* **80**(6): 510-522.

A.3 Brouwer A.P.D., Yntema H.G., Kleefstra T., Lugtenberg D., Oudakker A.R *et al.*, **Fullston T.**, *et al.* and Hamel B.C. **2007** 'Mutation frequencies of X-linked mental retardation genes in families from the EuroMRX consortium' *Hum Mutat* **28**(2): 207-208.

A.4 Tarpey, P. S., Smith R., Pleasance E., Whibley A., Edkins S. *et al.*, **Fullston T.** *et al.*, and Stratton M. R. **2009**. 'A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation' *Nat Genet* **41**(5): 535-543.

A.5 White R., Ho G., Schmidt S., Scheffer I.E., Fischer A., Yendle S.C., Bienvenu T., Nectoux J., Ellaway C.J., Darmanian A., Tong X., Cloosterman D., Bennetts B., Kalra V., **Fullston T.**, Gecz J., Cox T.C. and Christodoulou J. **2010.** 'Cyclin-dependent kinase-like 5 (CDKL5) mutation screening in Rett syndrome and related disorders' *Twin Res Hum Genet* **13**(2): 168-178.

A.6 *__Fullston, T.__, *Brueton L., Willis T., Philip S., MacPherson L., Finnis M., Gecz J.and Morton J. **2010**. 'Ohtahara syndrome in a family with an ARX protein truncation mutation (c.81C>G/p.Y27X)' *Eur J Hum Genet* **18**(2): 157-162.

A.7 *Reish, O., *__Fullston T.__, Regev M., Heyman E. and Gecz J. **2009**. 'A novel de novo 27 bp duplication of the ARX gene, resulting from postzygotic mosaicism and leading to three severely affected males in two generations' *Am J Med Genet A* **149A**(8): 1655-1660.

A.8 *Demos, M. K., *__Fullston T.__, Partington M. W., Gecz J. and Gibson W. T. **2009**. 'Clinical study of two brothers with a novel 33 bp duplication in the ARX gene' *Am J Med Genet A* **149A**(7): 1482-1486.

A.9 Shoubridge, C., Tan M. H., **Fullston T.**, Cloosterman D., Coman D., McGillivray G., Mancini G. M., Kleefstra T. and Gecz J. **2010b**. 'Mutations in the nuclear localization sequence of the Aristaless related homeobox; sequestration of mutant ARX with IPO13 disrupts normal subcellular distribution of the transcription factor and retards cell division' *Pathogenetics* **3**: 1.

A.10 Whibley, A. C., Plagnol V., Tarpey P. S., Abidi F., **Fullston T.**, *et al.*, and Raymond F. L. **2010**. 'Fine-scale survey of X chromosome copy number variants and indels underlying intellectual disability' *Am J Hum Genet* **87**(2): 173-188.

*Co-authroship

Shoubridge, C., Fullston, T. & Gecz, J. (2010). ARX spectrum disorders: making inroads into the molecular pathology.
Human Mutation, v. 31 (8), pp. 889-900

Fullston, T., Finnis, M., Hackett, A., Hodgson, B., Brueton, L., Baynam, G., Norman, A., Reish, O., Shoubridge, C. & Gecz, J. (2011). Screening and cell-based assessment of mutations in the Aristaless-related homeobox (ARX) gene.
Clinical Genetics, v. 80 (6), pp. 510-522

De Brouwer, A.P.D., Yntema, H.G., Kleefstra, T., Lugtenberg, D., Oudakker, A.R. *et al.*, Fullston, T. *et al*., & Hamel, B.C. (2007). Mutation frequencies of X-linked mental retardation genes in families from the EuroMRX consortium.
Human Mutation, v. 28 (2), pp. 207-208

Tarpey, P. S., Smith, R., Pleasance, E., Whibley, A., Edkins, S. *et al*., Fullston, T. *et al*. & Stratton, M.R. (2009). A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation.
Nature Genetics, v. 41 (5), pp. 535-543

White, R., Ho, G., Schmidt, S., Scheffer, I.E., Fischer, A., Yendle, S.C., Bienvenu, T., Nectoux, J., Ellaway, C.J., Darmanian, A., Tong, X., Cloosterman, D., Bennetts, B., Kalra, V., Fullston, T., Gecz, J., Cox, T.C. & Christodoulou, J. (2010). Cyclin-dependent kinase-like 5 (CDKL5) mutation screening in Rett syndrome and related disorders.
Twin Research and Human Genetics, v. 13 (2), pp. 168-178

Fullston, T., Brueton, L., Willis, T., Philip, S., MacPherson, L., Finnis, M., Gecz, J. & Morton, J. (2010).
Ohtahara syndrome in a family with an ARX protein truncation mutation (c.81C>G/p.Y27X).
European Journal of Human Genetics, v. 18 (2), pp. 157-162

Reish, O., Fullston, T., Regev, M., Heyman, E. & Gecz, J. (2009).
A novel de novo 27 bp duplication of the ARX gene, resulting from postzygotic mosaicism and leading to three severely affected males in two generations.
American Journal of Medical Genetics Part A, v. 149A (8), pp. 1655-1660

Demos, M.K., Fullston, T., Partington, M.W., Gecz, J. & Gibson, W.T. (2009).
Clinical study of two brothers with a novel 33 bp duplication in the ARX gene.
American Journal of Medical Genetics Part A, v. 149A (7), pp. 1482-1486

Shoubridge et al. PathoGenetics 2010, 3:1
http://www.pathogeneticsjournal.com/content/3/1/1

PathoGenetics

RESEARCH Open Access

Mutations in the nuclear localization sequence of the *Aristaless* related homeobox; sequestration of mutant ARX with IPO13 disrupts normal subcellular distribution of the transcription factor and retards cell division

Cheryl Shoubridge[1,2*], May Huey Tan[1,2], Tod Fullston[1,2], Desiree Cloosterman[1,2], David Coman[3], George McGillivray[4], Grazia M Mancini[5], Tjitske Kleefstra[6], Jozef Gécz[1,2]

Abstract

Background: *Aristaless* related homeobox *(ARX)* is a paired-type homeobox gene. ARX function is frequently affected by naturally occurring mutations. Nonsense mutations, polyalanine tract expansions and missense mutations in *ARX* cause a range of intellectual disability and epilepsy phenotypes with or without additional features including hand dystonia, lissencephaly, autism or dysarthria. Severe malformation phenotypes, such as X-linked lissencephaly with ambiguous genitalia (XLAG), are frequently observed in individuals with protein truncating or missense mutations clustered in the highly conserved paired-type homeodomain.

Results: We have identified two novel point mutations in the R379 residue of the ARX homeodomain; c.1135C>A, p.R379S in a patient with infantile spasms and intellectual disability and c.1136G>T, p.R379L in a patient with XLAG. We investigated these and other missense mutations (R332P, R332H, R332C, T333N: associated with XLAG and Proud syndrome) predicted to affect the nuclear localisation sequences (NLS) flanking either end of the ARX homeodomain. The NLS regions are required for correct nuclear import facilitated by Importin 13 (IPO13). We demonstrate that missense mutations in either the N- or C-terminal NLS regions of the homeodomain cause significant disruption to nuclear localisation of the ARX protein *in vitro*. Surprisingly, none of these mutations abolished the binding of ARX to IPO13. This was confirmed by co-immunoprecipitation and immmuno fluorescence studies. Instead, tagged and endogenous IPO13 remained bound to the mutant ARX proteins, even in the RanGTP rich nuclear environment. We also identify the microtubule protein TUBA1A as a novel interacting protein for ARX and show cells expressing mutant ARX protein accumulate in mitosis, indicating normal cell division may be disrupted.

Conclusions: We show that the most likely, common pathogenic mechanism of the missense mutations in NLS regions of the ARX homeodomain is inadequate accumulation and distribution of the ARX transcription factor within the nucleus due to sequestration of ARX with IPO13.

Background

The genes on the X-chromosome contribute significantly to genetic aetiology of intellectual disability [1]. The *Aristaless* related homeobox gene *(ARX)* [GenBank: NM_139058.2] is one of the more frequent contributors [1]. ARX belongs to a subset of *Aristaless*-related Paired-class (Prd-class) homeodomain proteins [2] and contains multiple domains, including the *aristaless* domain, homeodomain, the octapeptide and 4 polyalanine tracts [3]. *ARX* mutations cause intellectual

* Correspondence: cheryl.shoubridge@adelaide.edu.au
[1]Department of Genetics and Molecular Pathology, SA Pathology at the Women's and Children's Hospital, North Adelaide, South Australia 5006, Australia

disability with or without additional features including epilepsy, infantile spasms, dystonia, lissencephaly, autism and dysarthia [3-5]. To date, over 90 families and individual cases with 40 different types of mutations have been reported. These include missense mutations, protein truncations but most frequently, polyalanine tract expansions [6-19].

When *Arx* was first ablated in mouse [5], the phenotype recapitulated many clinical aspects of X-linked lissencephaly with ambiguous genitalia [XLAG; MIM 300215] [20,21]. Subsequently, this led to the identification of *ARX* mutations in patients with XLAG [5]. Lissencephaly is one of a heterogenous group of disorders arising from aberrant neuronal migration. The characteristic 'smooth brain', due to a paucity of normal gyri and sulci, is due to either the arrest of neuronal migration (classical) or an over-migration of neurons (cobblestone). Mutations in the X-linked gene *DCX* (MIM 300121) [22] and four autosomal genes: *LIS1* (MIM 601545) [23], *RELN* (MIM 600514) [24], *TUBA1A* (MIM 602529) [25,26] and *VLDLR* (MIM 192977) [27] have been associated with distinct lissencephaly syndromes. However, mutations in *ARX* are the only identified genetic cause underlying the distinct syndrome of XLAG. This syndrome differs from the other forms of lissencephaly, displaying a thickened cortex with posterior to anterior gradient of gyral malformation, agenesis of the corpus callosum and ambiguous genitalia [28].

There are currently 31 families reported with XLAG phenotypes, with and without additional features, due to 27 different mutations in *ARX*. The majority of these families are predicted to arise from protein truncation and the loss-of-function of the mature ARX protein [5,7,9,12,28-31]. In the remaining families with XLAG, single nucleotide substitutions clustered in the homeodomain or, in one case, just prior to the *aristaless* domain, are predicted to give rise to amino acid substitutions in the mature protein. Several of these single nucleotide substitutions occur in residues of nuclear localization sequences (NLS) that flank both ends of the ARX homeodomain. In the N-terminal NLS (NLS2) there are four naturally occurring point mutations - R332P, R332H and R332C which cause XLAG and a residue adjacent to this arginine, T333N, which causes Proud syndrome (agenesis of the corpus callosum with ambiguous genitalia) [ACC/AG; MIM 30004]. We have recently identified novel point mutations in a single residue of the NLS3 region flanking the C-terminal portion of the homeodomain, R379. In one case, a substitution of R379 with lysine (L) has been identified in a patient presenting a phenotype of XLAG (D Coman, unpublished data). In contrast, a second change in the same residue resulting in the substitution of R379 with serine (S) was identified in a proband and his female relatives [32]. Interestingly, the phenotype includes infantile spasms and severe mental retardation without obvious brain malformation.

The basic residues of ARX NLS2 and NLS3 are important for the correct localization of ARX in the nucleus through interaction with Importin 13 (IPO13) [33]. IPO13 is a member of the importin-β superfamily involved in nuclear import and export of a variety of proteins [34,35]. In addition to IPO13, a recent study suggests that multiple importins are utlized to import murine Arx into the nucleus, including importin β1 [36]. Regardless of the importin utilized, the driving force for nuclear protein import is provided by RanGTP and its interaction with the importin-cargo complex [37]. We show that missense mutations of NLS2 and NLS3 cause significant disruption to the nuclear localization of ARX *in vitro*. However, these mutations do not abolish the binding of ARX to IPO13. Interestingly, the binding of mutant ARX protein to endogenous IPO13 is indistinguishable from the binding to N-terminally truncated IPO13 lacking the RanGTPase binding domain. Hence, the ability of the mutant ARX-IPO13 complex to uncouple in the RanGTP rich nuclear environment appears to be compromised. As part of our investigation into the pathogenic mechanism underlying these mutations we identify TUBA1A, a component of the cytoskeleton important in mitosis, as a novel protein partner for ARX and examine the impact of sequestration of mutant ARX cargo with IPO13 on mitosis and cell division.

Materials and methods

Subjects and families

The relevant research ethics committees and institutional review boards of collaborating institutions approved this study. Clinical information and DNA samples were collected with informed consent.

Cloning of mutant full-length *ARX* constructs

The cloning of the full-length human *ARX* cDNA construct in pCMV-Myc vectors (ARX-Wt; 1-562 aa) and the V5-IPO13 (aa 217-963; minus the RanGTPase binding site) have been described previously [33]. Single nucleotide substitutions were introduced into the pCMV-Myc-ARX-Wt full-length construct via site directed mutagenesis (Stratagene) following the manufacturer's instructions. The mutations included c.995G>C leading to p.R332P (R332P) in the NLS2 region, c.998C>A leading to p.T333N (T333N) close to NLS2, c.1136G>T leading to p.R3479L (R379L) and c.1135C>A leading to p.R379S (R379S) in the NLS3 region. A double mutant construct was engineered with a mutation in each the NLS2 (c.995G>C) and NLS3 (c.1136G>T) regions in the same cDNA construct (R332P-R379L). In addition, a construct containing a c.1058C>T change

leading to p.P353L (P353L) was engineered to provide an ARX mutation located within the homeodomain but outside of the NLS regions. All clone preparations were verified by sequencing of the entire coding region.

Cell culture and transient transfection

HEK293T cells were maintained in Dulbecco's modified Eagle's medium supplemented with 10% (v/v) fetal calf serum. All cells were cultured in the presence of 100 U/ml sodium penicillin and 100 μg/ml of streptomycin sulphate in 5% CO_2 at 37°C. Cells plated at 4×10^5 per well in a six- well plate the day before transfection in media lacking antibiotics. Routinely, cells were transfected with a total of 1 μg of plasmid DNA using Lipofectamine 2000 (Invitrogen, CA, USA) following the manufacturer's instructions.

Antibodies

The following antibodies were used for immunofluorescence and/or western immunoblot analysis: mouse anti-ARX antibody (2 μg/ml final); goat anti-IPO13 (1:500) (Imgenex, CA, USA); rabbit-anti-alpha-tubulin (1:1500) (Rockland, PA, USA); rabbit anti-V5 antibody (1:5000) (Bethyl Laboratories, Tx, USA). The secondary antibodies were either fluorescent labelled; goat anti-mouse-IgG conjugated to CY3 (1:1000) (Jackson Laboratories, Maine, USA); goat anti-mouse-IgG conjugated to FITC (1:1000) (Dako, Glostrup, Sweden); goat anti-rabbit-IgG Alexa Fluor 488 (1:800) (Invitrogen); donkey anti-goat-IgG Alexa Fluor 555 (1:1000) (Invitrogen) or horseradish-peroxidase (HRP) conjugated; mouse anti-Myc HRP conjugated antibody (1:5000) (Invitrogen); mouse anti-V5 HRP conjugated antibody (1:5000) (Invitrogen); goat anti-mouse-HRP antibody (1:1000) (Dako); goat anti-rabbit HRP antibody (1:1000) (Dako); rabbit anti-goat HRP (1:2000).

Immunofluorescence and microscopy

Transfected cells were harvested at 24 and 48 h post-transfection by fixation in 3.7% formaldehyde-phosphate buffered saline (PBS) (v/v) and permeabilized in 0.2% (v/v) triton PBS for 5 min. Blocking of non-specific binding of the secondary antibody was achieved routinely by addition of 5% skim-milk powder (w/v) in Tris buffered saline and 0.5% Tween (v/v) (TBS-T) before incubation of the primary and secondary antibodies diluted in 1% milk (w/v) in TBS-T. Removal of excess block and antibody was achieved with multiple washes of TBS-T. Nuclei were counterstained with DAPI (Molecular probes, Invitrogen).

For subcellular localization studies: Between 1000-4000 transfected cells were counted for each construct from at least three different transfection reactions using standard fluorescence microscopy. The percent of transfected cells with abnormal localization or aggregates was determined as the number of cells containing abnormal localization or aggregates divided by the total number of ARX positive cells. Sub-cellular localization images were captured by Leica TCS SP5 spectral confocal microscope using a 100× Plan apochromat objective. Z-stacks were taken at 0.25 μm intervals and maximal projections were made for all cell images.

For mitosis studies: Between 1200 and 4000 cells were analysed at 24, 48 and 72 h post-transfection for: (i) percentage of cells transfected; (ii) percentage of mitotic cells in both transfected and un-transfected cells; and (iii) and the phase of mitosis for each cell undergoing division.

Yeast-2 hybrid screening

Plasmids encoding the *ARX* homeodomain (ARX-HD) were fused to the GAL4 DNA binding domain and the library protein were fused to GAL4-activation domain. The MaV203 yeast strain was co-transformed with the *ARX-HD* construct and human fetal brain cDNA library (ProQuest, Invitrogen). Colonies were selected on media lacking leucine, tryptophan and histidine and positive clones were analysed for expression of three reporter genes (*HIS3*, *URA3* and *LacZ*) by measuring β-galactosidase activity. Library inserts of positive clones that activated more than one reporter gene were amplified by polymerase chain reaction and sequenced to determine the identity of the library clone.

Co-immunoprecipitation

Cells transfected with *Myc-ARX*, both with and without *V5-IPO13* were harvested at 24 hrs post-transfection and cell lysates prepared using lysis buffer (120 mM NaCl, 50 mM Tris-HCl (pH 8.0), 0.5% NP-40 (v/v), 1× protease inhibitor cocktail (Sigma, AZ, USA), 1 mM Na_3VO_4, 1 mM NaF, 1 mM PMSF). Lysates were clarified by centrifugation (15 min, 13,000 *g* at 4°C). Aliquots of extracts were immunoprecipitated (IP) overnight at 4° C. Protein-A sepharose was pre-treated with un-transfected HEK293T cell lysate to ameliorate non-specific binding of cell proteins. The IP reactions were incubated with the pre-treated protein-A sepharose for 1 h at 4°C before removal of non-specifically bound proteins with four changes of high stringency wash buffer (500 mM NaCl, 20 mM Tris-HCl (pH 8.0), 1 mM EDTA, 0.5% NP-40 (v/v)) to ensure adequate removal of non-specific binding of alanine tract containing ARX protein. Proteins bound to the protein-A sepharose were eluted in SDS loading buffer (62.5 mM Tris-HCl (pH 6.8), 2% SDS (v/v), 10% glycerol (v/v), 5% β-Mercaptoethanol (v/v), 0.001% bromophenol blue (w/v), heated to 65°C before addition and incubated for 3 min). IP proteins were subjected to SDS-PAGE and transferred to nitrocellulose membrane. Lysates from HEK293T cells producing either Myc-ARX alone or V5-IPO13 alone were used as controls.

In co-transfected cells; Myc-ARX protein was IP with 0.5 μg of anti-Myc antibody (Santa Cruz Biotech, CA,

USA) or the reciprocal co-immunoprecipitation (Co-IP) with 0.5 μg of rabbit anti-V5 antibody immunoprecipitating V5-IPO13 protein. IP proteins were analysed for the presence of Myc-ARX and V5-IPO13 by western immunblotting. In cells transfected with ARX alone, endogenous IPO13 was IP with 1 μg of goat anti-IPO13 and Co-IP of Myc-ARX protein was identified using mouse anti-Myc HRP conjugated antibody.

To verify the interaction between Myc-ARX and TUBA1A, HEK293T cells transfected with Myc-ARX constructs were lysed and immunoprecipitated with mouse anti-Myc antibody. IP proteins were analysed for the presence of Myc-ARX and endogenous alpha-tubulin by western immunoblotting.

Statistical analysis

All data are reported as mean ± standard error of mean, determined from a minimum of three separate transfection reactions, with the number of cells counted indicated in each figure legend. Differences in the percentage of transfected cells with abnormal subcellular localization were analysed by Kruskal-Wallis Test and when significance was reached a one-tailed pair-wise comparison was achieved using Wilcoxon-Mann-Whitney U test. In the case of the mitosis data, differences in the proportions of cells in various phases of mitosis at increasing times post-transfection were compared by two-way ANOVA, with mutations and time as factors. $P < 0.05$ was considered significant.

Results

Novel mutations in ARX homeodomain

We have identified two novel point mutations in the same residue of the NLS3 region of ARX homeodomain in two unrelated families: (i) An apparently *de novo* and novel single base substitution c.1136G>T in exon 4 of *ARX*, resulting in the substitution of an arginine to a leucine amino acid substitution at position 379 (R379L); and (ii) A novel single base substitution c.1135C>A in exon 4 of *ARX* was identified in the proband and three female relatives [32], resulting in the substitution of an arginine to a serine amino acid substitution at position 379 (R379S) (Figure 1). In the first case a clinical diagnosis of XLAG phenotype included mildly dysmorphic features, ambiguous genitalia with a micropenis, fused scrotum, cryptorchidism and a severe refractory seizure disorder. Magnetic resonance imaging (MRI) identified lissencephaly with agyria posteriorly transitioning in the mid-parietal region to pachygyria anteriorly, agenesis of the corpus callosum and a severe encephalopathy (Figure 1a-c). Other clinical findings include hypothalamic dysfunction, hypophosphatasia and severe chronic diarrhoea with evidence of pancreatic insufficiency and small bowel malabsorption. No developmental milestones were attained and the disease was ultimately lethal. The c.1136G>T mutation was not found in the mother of the patient and her MRI brain scan showed a normally formed corpus callosum (data not shown). In the second case, the associated phenotype of carrier females of the c.1135C>A mutation from the family were reported [32]. We report here the clinical findings of the male proband, who had a history of infantile spasms and severe intellectual disability (Figure 1 Family 2, patient III-4). He was born after a pregnancy of 39 weeks by spontaneous delivery with a birth weight of 2600 g. In the first months eye contact was not adequate. At the age of 6 months he experienced seizures with tonic extension of the arms and turning of the eye globes, which progressively increased in frequency and were followed by apnoea. The electroencephalograph showed a disorganized background pattern compatible with hypsarrhythmia. On examination he was apathetic, made no eye contact and showed insufficient spontaneous movements. His head growth progressively slowed and at 6 years of age his occipitofrontal circumference was below -2SD. He had no significant psychomotor development and at 5 years of age had spastic tetraparesis with truncal hypotonia. At 8 years of age he is wheelchair bound, makes no eye contact, reacts to touch and can chew and swallow food. He has recurrent airway infections. His seizures are refractory to medical treatment, notwithstanding the use of several antiepileptic drug medications. Interestingly, this patient showed no XLAG or brain malformation when examined by MRI (data not shown). The patient had unilateral cryptorchidism. The presence of dry scaly skin prompted testing for steroid sulphatase deficiency and the diagnosis was confirmed by genomic DNA analysis that showed an STS gene deletion and leukocytes enzyme assay that showed reduced activity (data not shown). The borders of the STS deletion were fine mapped in another family member with the same deletion by 250 K SNP arrays and were comprised between SNP_A-2122915 and SNP_A-1933813 on chromosome Xp22.

Missense mutations in NLS2 and NLS3 disrupt nuclear localization of ARX

Transient transfection of full length Myc-tagged ARX-Wt and mutants were conducted to examine changes in subcellular localisation of the mutant proteins compared to the ARX-Wt protein. We tested R332P, T333N, R379L and R379S mutations (Figure 2A). We used confocal microscopy to analyse subcellular localisation, highlighted for the R379L mutation (Figure 2B). Aggregates are clearly distinguished in the peri-nuclear region of the cytoplasm, with a large aggregate mass in the top right quadrant. In addition to the cytoplasmic localisation, a series of aggregates in the nuclear space can be identified. The percentage of transfected cells with aberrant sub-cellular localisation of ARX protein was

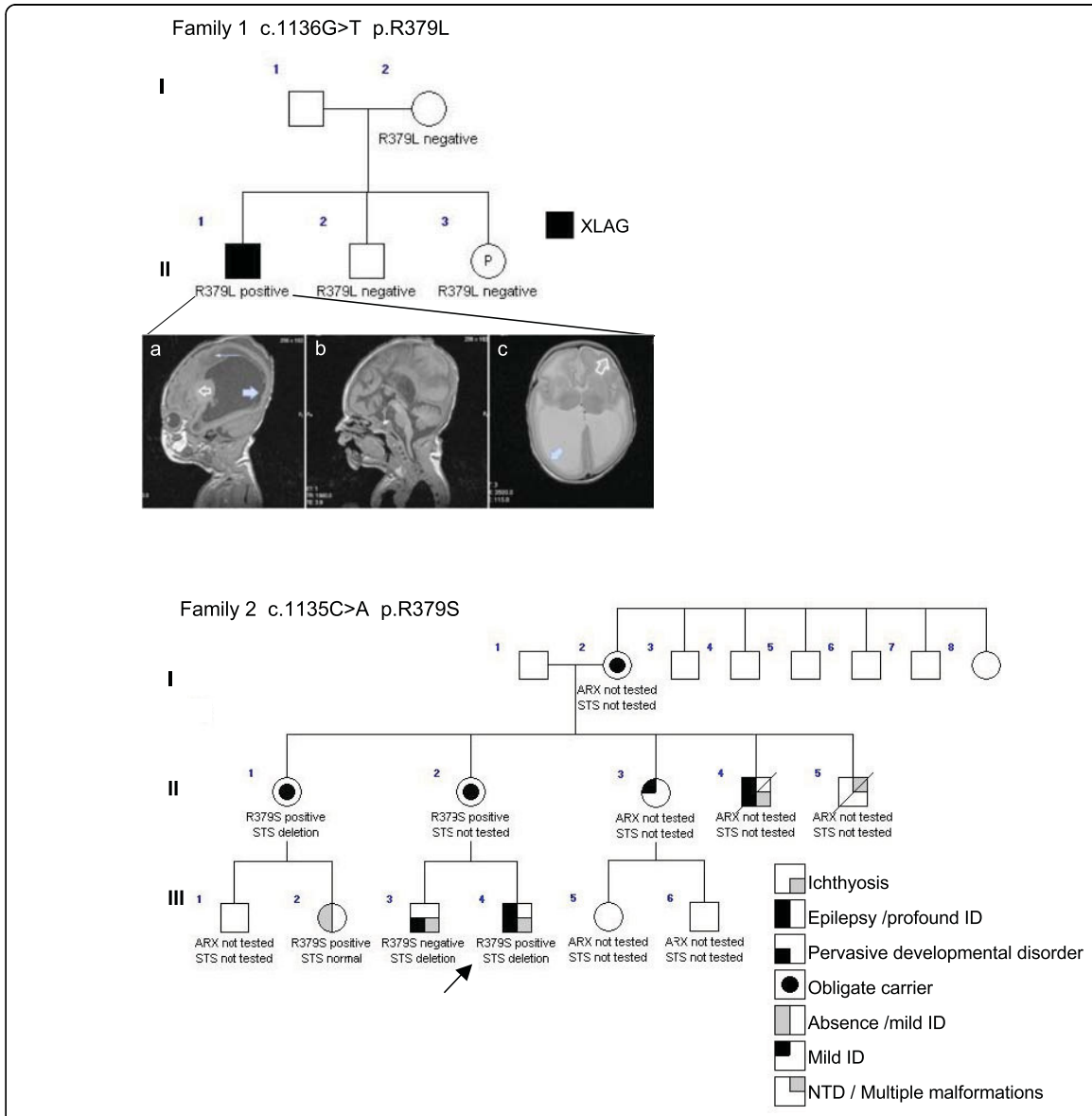

Figure 1 Pedigree of families with novel mutations in the ARX homeodomain. Open symbols represent normal individuals, filled symbols represent affected individuals as per the key. Proband indicated with arrow. Individual generations are numbered with Roman numerals. MRI findings in proband of family 1 (a) sagittal T1 MR sequence demonstrates lissencephaly P > A and marked lateral ventriculomegaly. Note subependymal cysts (large arrow) abnormal periventricular white matter signal (small arrow) and abnormal basal ganglia signal clear arrow, (b) sagittal T1 MR sequence of the midline demonstrates agenesis of the corpus callosum and (c) axial T2 MR sequence demonstrates colpocephaly and lissencephaly with transition from posterior agyria (solid arrow) to anterior pachygyria (open arrow).

assessed in at least 1000 transfected cells from a minimum of three separate transfection experiments. For counting purposes, transfected cells were categorised as (1) normal localization; diffuse staining restricted to the nucleus (Figure 2C, top panel) or (2) abnormal localization, aggregates of mutant protein in either the nucleus or in both the nucleus and cytoplasm (Figure 2C). Expression of the Myc-tagged ARX-Wt protein was detected in a non-homogenous pattern across the nucleus with little or no expression detectable in the cytoplasm. We observed this normal pattern of staining in over 87% of cells transfected with ARX-Wt at 24 hrs (Figure 2D) and 91% of cells at 48 h post-transfection (data not shown). Transient over-expression leads to aberrant localization of the ARX-Wt protein in approximately 13% of transfected cells (Figure 2D), generally in the form of a single bright spot in the nucleus in addition to the normal diffuse staining (data not shown). In contrast, ARX NLS mutant proteins were found within the nucleus as either multiple, small aggregates, or

longer continuous 'ribbon like structures' or larger aggregates in which the DAPI stained nuclear material had been excluded. The mutant protein was also found in aggregates in the cytoplasm, often forming large bodies to one side of the nucleus distorting the DAPI stained nuclear material (Figure 2C). Routinely, cells with abnormal localization of the mutant protein were a mixture of cells with aggregates in the nucleus and no detectable expression of ARX in the cytoplasm, in addition to cells with aggregates of mutant ARX in both the nucleus and cytoplasm. No cells with ARX protein exclusively in the cytoplasm were noted. The percentages of transfected cells with abnormal ARX protein localization were significantly elevated for all NLS mutations tested compared to ARX-Wt transfected cells (Figure 2D). Mutations leading to severe clinical outcomes resulted in the highest levels of cells with abnormal protein localization. Over-expression of R332P for 24 h resulted in 72% of transfected cells in which the mutant protein was incorrectly localized; an approx 5.5-fold increase compared to the 13% of ARX-Wt transfected cells. The T333N mutation, adjacent to the NLS2 region, mis-localized in 61% of transfected cells. Similarly, the R379L mutation resulted in 60% of all transfected cells with abnormal protein localization (Figure 2D). Unlike the other three mutations, the R379S mutaiton leads to a phenotype without malformation of the brain. In cells expressing the R379S mutation the increase in cells with abnormal localisation of mutant protein was the lowest of all the mutations tested at 54% (Figure 2D). In contrast to the NLS mutations, when another mutation the middle of the ARX homeodomain (P353L; XMESID phenotype) was expressed in HEK293T cells there were no increased levels of cells with mis-localized protein compared to ARX-Wt (Figure 2D). When a single construct containing two mutations (R332P and R379L) was transfected into HEK293T cells we found 90% of all of transfected cells displayed abnormal localisation of the mutant protein (Figure 2D). This increase was above the effect of either mutation alone and indicates the NLS regions flanking the homeodomain are not redundant but likely act co-operatively.

Missense mutations in NLS2 and NLS3 do not abolish ARX interaction with N-terminally truncated IPO13

In order to test if ARX with mutations in NLS2 or NLS3 were capable of binding to IPO13 in the cell environment we co-transfected V5-tagged-IPO13 lacking the N-terminal RanGTPase binding site. This truncated V5-IPO13 protein can bind and transport the ARX cargo but is unable to disassociate the cargo upon reaching the RanGTP rich nuclear environment. We IP over-expressed truncated V5-IPO13 by anti-V5 antibody (Figure 3A, top box, second panel) and Myc-ARX by anti-Myc antibody (Figure 3A, middle box, second panel) from whole cell extracts, with no protein detected in parallel experiments in which no IP antibody was added (Figure 3A, lane 1). In cells expressing both Myc-ARX and V5-IPO13 we were able to detect co-IP of Myc-ARX protein when immunoprecipitating with anti-V5 antibody (Figure 3A, top box, first panel) and in the reciprocal immunoprecipitation with anti-Myc antibody we detected Co-IP of V5-IPO13 protein (Figure 3A, middle box, first panel). Regardless of the antibody used for pull down we observed a stronger Co-IP signal for all ARX mutant samples, including the ARX protein containing the double NLS mutation compared to the ARX-Wt protein. In agreement with the Co-IP data, there was a shift in the truncated V5-IPO13 signal from diffuse cytoplasmic localisation when transfected alone (Figure 3B top panel) to complete overlap with the nuclear expression of ARX-Wt protein (Figure 3B second panel) and nuclear and cytoplasmic expression of NLS2 and NLS3 mutant ARX signal in co-transfected cells (Figure 3). These results were consistent across all of the mutations tested, including the R332P-R379L mutant. These results clearly indicate the mutant ARX proteins were able to bind truncated V5-IPO13, even when there were single amino acid substitutions in both NLS regions.

Endogenous IPO13 binds to ARX protein containing missense mutations in NLS2 and NLS3

HEK293T cells express detectable levels of endogenous IPO13 mRNA and protein. When we used an anti-IPO13 antibody the signal was localised as a diffuse punctate staining across the cytoplasm (Figure 4A top panel). In cells expressing mutant ARX, much of the endogenous IPO13 signal was sequestered with the ARX protein, with residual IPO13 signal in the cytoplasm (Figure 4A). This co-localization was observed in all cells expressing mutant ARX protein, including the protein with mutations in both NLS2 and NLS3 (Figure 4A bottom panel). In contrast, endogenous IPO13 protein did not co-localize with ARX-Wt (Figure 4A second panel), likely due to efficient recycling of IPO13 back to the cytoplasm after uncoupling from ARX in the RanGTP rich environment of the nucleus. We conclude that mutant ARX protein binding to endogenous full length IPO13 localizes in the same manner as if it were bound to the N-terminally truncated V5-IPO13 that cannot physically uncouple from its cargo. The interaction of endogenous IPO13 and mutant ARX protein was confirmed by Co-IP of Myc-ARX NLS mutant protein, but not the ARX-Wt protein from whole cell extracts using an antibody against endogenous IPO13 (Figure 4B, top panel).

The binding of N-terminally truncated V5-IPO13 to ARX-Wt increases the proportion of cells with abnormal ARX protein localization

When Myc-ARX-Wt is co-expressed with V5-IPO13, we measure an increased proportion of cells with abnormal localization of the ARX-Wt protein compared to cells

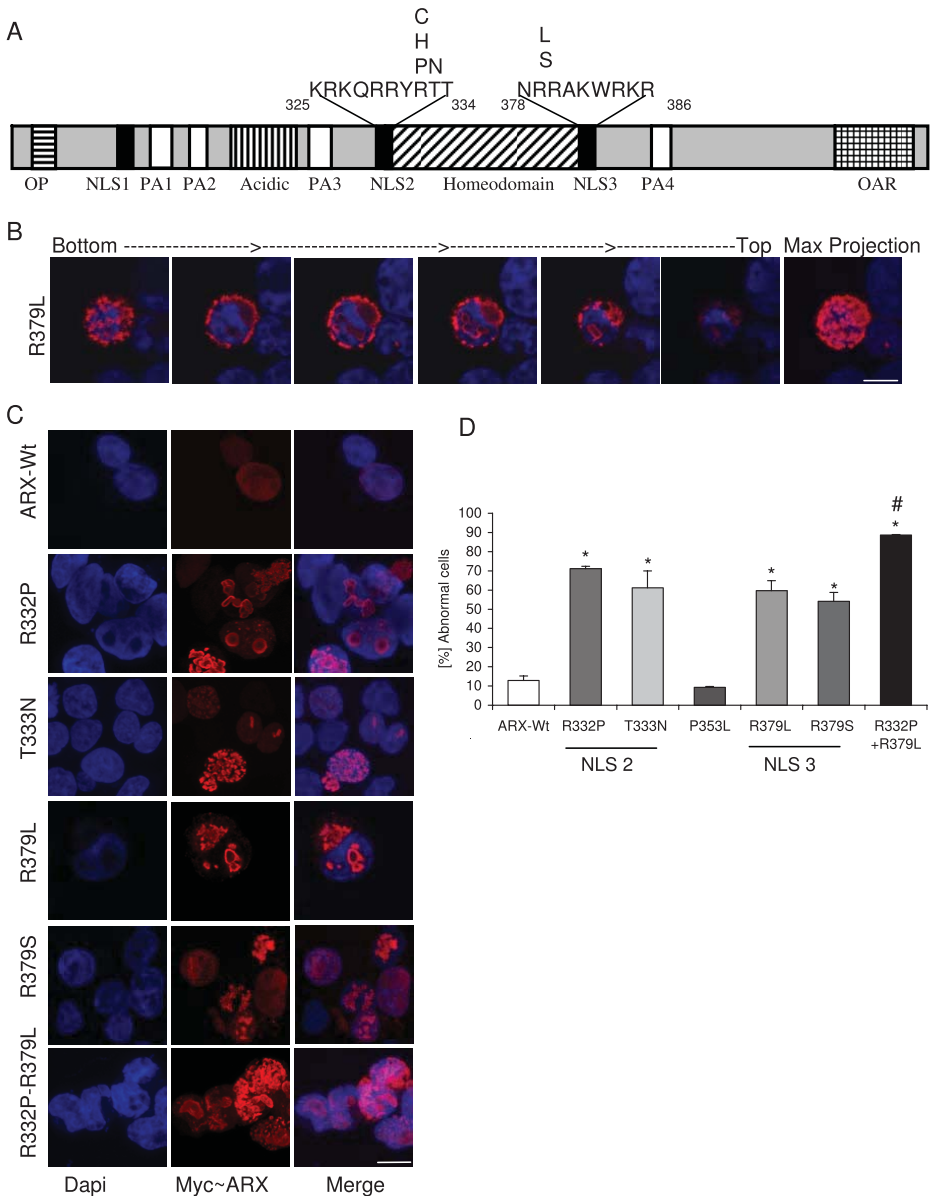

Figure 2 Missense mutations in nuclear localization sequence (NLS) regions of the *Aristaless related homeobox* (ARX) homeodomain disrupt the normal nuclear subcellular localization in HEK293T cells. (A) Schematic of the human ARX protein. Known functional domains are highlighted in the open reading frame: Octapeptide (OP) as horizontally hatched rectangle, NLS as three black rectangles, polyalanine tracts (PA) as four white rectangles, acidic domain as vertically hatched rectangle, homeodomain crosshatched and *Aristaless* domain (OAR) hatched. The sequences flanking the homeodomain are shown above the black rectangle, the locations indicated at either end, with the basic residues that are part of the NLS-like motifs in bold. (B) Individual focal planes of cell transfected with Myc-R379L (NLS3) mutation construct show aggregates of mutant ARX protein form both inside the nucleus and outside in the cytoplasm surrounding the nucleus. Panel on far left is at the top of the cells (focal plane 2) with subsequent images an additional four focal planes further through the cell, maximal projection of all images is shown on the far right. The aggregates often displace the nuclear material, shown by distortion and absence of the blue DAPI signal in location of the aggregates. Scale bar = 10 μM. (C) Representative pictomicrographs of the localization of the Wt and mutant ARX protein 24 hrs post transfection. My~ ARX detected by anti-ARX Ab and Cy3 conjugated secondary antibody (left panel) merged with DAPI stained nuclear material (right panel). Scale bar = 10 μM. (D) The percentage of transfected cells displaying abnormal localization as inclusions or aggregates in the nucleus, with or without aggregates in the cytoplasm, was determined from between 1000 and 4000 transfected cells per construct, from at least two separate transfection reactions, 24 h post-transfection. Full-length Myc-tagged constructs transfected are listed along the bottom of the graph; ARX Wt (open bar), R332P (light grey bar), T333N (grey bar), P353L mutant (cross hatched bar), R379L and R379S (both dark grey bars) and double mutant R332P-R379L (black bar). All groups comparison of the percentage of transfected cells with abnormal subcellular localization of expressed ARX protein was achieved by parametric Kruskal-Wallis test. * $P < 0.05$ versus ARX-Wt, # versus ARX-NLS mutants with single aa substitutions.

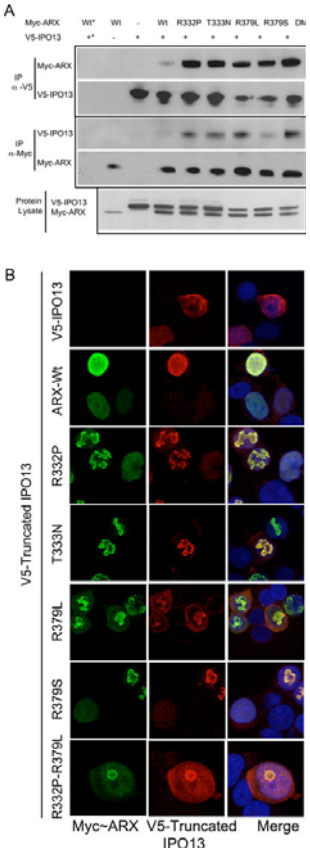

Figure 3 NLS mutant ARX co-immunoprecipitates and co-localizes with N-terminally truncated IPO13 in mammalian cells. (A) HEK293T cells transfected with Myc-ARX-Wt or NLS mutant constructs and V5-IPO13 were lysed and protein immunoprecipitated (IP) with antibodies against the V5 or Myc tags. Samples were loaded onto 4%-12% SDS-PAGE gels and analysed for the presence of co-IP proteins. Reciprocal Co-IPs was conducted on replicate samples; IP with rabbit anti-V5 antibody (top panel) and detection of Myc-ARX proteins bound to V5-IPO13 by immunoblotting with mouse anti-Myc HRP conjugated antibody. All samples transfected with V5-IPO13 showed a protein band of the correct size upon blotting with anti-V5 horseradish-peroxidase (HRP) conjugated antibody. IP with mouse anti-Myc antibody (middle panel) and detection of V5-IPO13 protein bound to Myc-ARX by immunblotting with mouse anti-V5 HRP conjugated antibody. All samples transfected with Myc-ARX showed a protein band of the correct size upon blotting with anti-Myc HRP conjugated antibody. Specific IP of each over-expressed protein was achieved with no band present in samples from cells transfected with both Myc-ARX and V5-IPO13 but no IP antibody added (*). Cells transfected with Myc-ARX alone or V5-IPO13 alone was used as negative controls. V5-IPO13 (~88.6 kDa) and Myc-ARX (~62.2 kDa) are both present in protein lysates (bottom panel). (B) Pictomicrographs showing V5-IPO13 co-localises with ARX-Wt and mutant ARX protein (indicated on left of panels) in both nuclear and cytoplasmic aggregates in cells co-transfected with Myc-ARX and V5-IPO13 constructs for 24 h. ARX detected by anti-ARX antibody and FITC conjugated secondary antibody (left panel), V5-IPO13 detected by anti-IPO13 antibody and Cy3 conjugated secondary antibody (middle panel) and merged image including blue DAPI stained nuclear material (right panel). Scale bar = 10 μM.

expressing the ARX-Wt alone (Figure 5A). In cells transfected with ARX alone we find 13% of cells with abnormal ARX-Wt localisation (Figure 5B; in agreement with previous result; Figure 2D), but when the cells are co-expressing V5-IPO13 we note 60% of cells with abnormal localization of ARX-Wt protein (Figure 5C). In addition, these cells often had large or multiple aggregates within the nucleus and in some cases in the cytoplasm, reminiscent of the results for ARX-NLS mutant proteins (data not shown). This data clearly indicates that the binding of ARX-Wt protein to V5-IPO13 contributes to aggregation and mis-localization of the ARX.

Potential pathogenic mechanism of NLS mutations in ARX

In many cells with mutant ARX protein expression that were undergoing mitosis we noted the mitotic spindles were messy or even disrupted by aggregates of mutant protein (Figure 6). This raised the question of whether the formation of microtubules during cell division is affected by sequestration of ARX.

Alpha Tubulin is a novel interacting protein partner of ARX

Interactions of ARX with cytoskeletal and other proteins were pursued using a GAL4-based yeast-2 hybrid screen. We identified alpha-tubulin (TUBA1A; NM_006009.2) and beta-tubulin (TUBB3; NM_006086.2) from a human fetal brain cDNA library. The bait protein used in this instance was the human ARX homeodomain (aa 303-431) fused to the GAL4 DNA binding domain (GAL4-DBD). This protein did not autoactivate the *HIS3, lacZ* or *URA3* reporter genes upon transformation of MaV203 yeast (data not shown). Given the involvement of mutations in TUBA1A in phenotypes with lissencephaly and other similar clinical findings to the severe XLAG phenotype due to mutations in ARX we wanted to investigate the novel interaction between ARX and TUBA1A further. As there is a high level of similarity between members of the alpha-tubulin family and we predicted that ARX may interact not only with the brain specific TUBA1A but also with other members of the family. We confirmed this interaction in mammalian cells by Co-IP. Overexpressed Myc-tagged ARX was immunoprecipitated from whole cell extracts using a monoclonal anti-Myc antibody. Endogenous alpha-tubulin immunoprecipiated with Myc-ARX was detected by rabbit anti-alpha-tubulin antibody (raised against a common immunogen; aa 417 to 441 of TUBA1A, NP_00600.2) (Figure 7). When we tested the mutations in the NLS regions of ARX we observed Co-IP of alpha-tubulin for all samples, but not in parallel samples in which no IP antibody was added (Figure 7). Our data indicates that alpha-tubulin interacts with the homeodomain of ARX and this interaction is not abolished by any of the NLS mutations tested. Despite the interaction

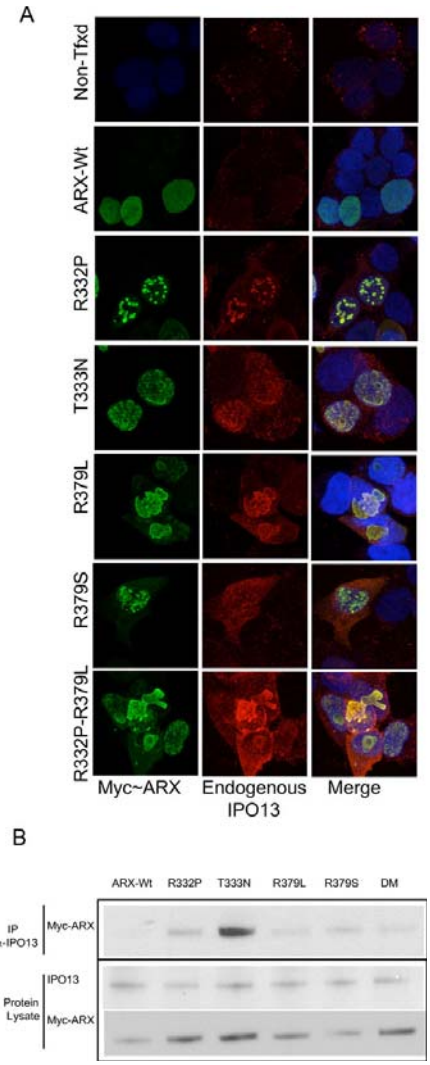

Figure 4 Endogenous IPO13 co-localizes and co-immunoprecipitates with mutant ARX protein. (A) Pictomicrographs showing sub-cellular localization of Wt and mutant Myc-ARX detected by anti-ARX antibody and FITC conjugated secondary antibody (left panel), endogenous IPO13 detected by anti-IPO13 antibody and Cy3 conjugated secondary antibody (middle panel) and merged image including blue DAPI stained nuclear material (right panel). Endogenous IPO13 co-localises with Myc-ARX mutant protein, including the double mutant R332P-R379L, in both nuclear and cytoplasmic aggregates, but does not co-localize with the ARX-Wt protein (lane 1). (B) HEK293T cells transfected with Myc-ARX-Wt or NLS mutant constructs were lysed and protein IP with antibodies against endogenous IPO13. Samples were loaded onto 4%-12% SDS-PAGE gels and analysed for the presence of co-IP proteins. IP with Goat anti-IPO13 antibody (top panel) and detection of Myc-ARX proteins bound to IPO13 by immunoblotting with mouse anti-Myc HRP conjugated antibody. All samples transfected with Myc-NLS mutant ARX showed a protein band of the correct size upon blotting with anti-Myc HRP conjugated antibody, with no band present in the ARX-Wt lane. Endogenous IPO13 (~108 kDa) and Myc-ARX (~62.2 kDa) are both present in protein lysates (bottom box). Scale bar = 10 μM.

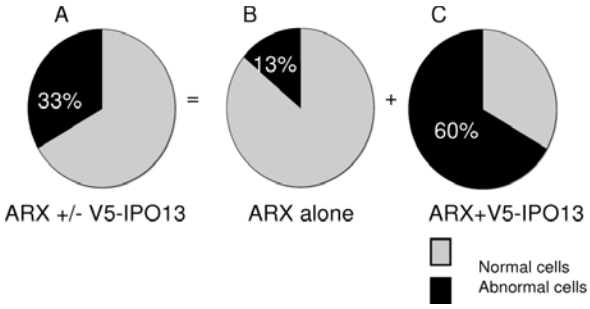

Figure 5 ARX-Wt protein is abnormally localized when co-transfected with N-terminally truncated V5-IPO13. HEK293T cells transfected with Myc-ARX-Wt and V5-IPO13 were assessed for localization of the resulting ARX protein, either normal nuclear localization (grey) or abnormal localization as inclusions or aggregates in the nucleus, with or without aggregates in the cytoplasm (black). (A) Combined count of cells transfected with ARX with or without V5-IPO13, (B) cells with only ARX expression and (C) cells with both ARX and V5-IPO13 expression.

of ARX and alpha-tubulin by yeast-2 hybrid studies and Co-IP, in cells co-stained for Myc-ARX and alpha-tubulin we did not see any apparent co-localization of the two proteins. However, in mitotic cells expressing the mutant ARX protein we noted the condensation of DNA and the formation of mitotic spindles were often localized around aggregates of the mutant protein (Figure 6).

Sequestration of mutant ARX and IPO13 may compromise mitosis

The effects of the sequestration of mutant ARX with IPO13 on mitosis and cell division were investigated at 24, 48 and 72 h post-transfection and compared with ARX-Wt. Co-staining of ARX (transfected cells expressing ARX protein), alpha-tublin (mitotic microtubules) and nuclear material (DAPI) was conducted to assist characterization of mitotic cells. At each time point cells

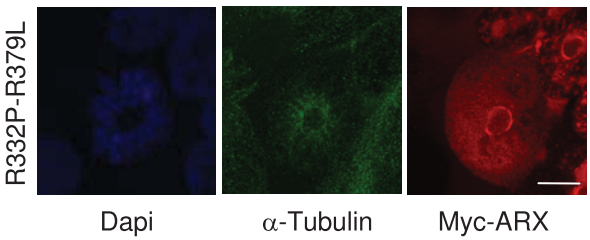

Figure 6 Mitotic spindles disrupted by aggregates of NLS mutant ARX protein. A representative pictomicrograph of a HEK293T cell transfected with mutant Myc-ARX undergoing DNA condensation stage of mitosis shows abnormal pattern of DNA condensation (DAPI) microtubule formation (alpha-tubulin) around a large aggregate of mutant ARX protein (Myc-ARX) with the remaining Myc-ARX signal localized diffusely throughout the cell interior. Scale bar = 10 μM.

were analysed for: (i) percentage of cells transfected; (ii) percentage of mitotic cells in both transfected and untransfected cells; and (iii) and the phase of mitosis for each cell undergoing division.

The percentage of cells expressing Myc-ARX were higher in cells transfected with mutant ARX compared to the ARX-Wt at all time points examined (Figure 8A). This difference was ~1.3-fold at 24 h, increasing to 1.6- to 2.2-fold by 48 h and between 2.3 (T333N) and 3.8 (R332P) fold by 72 h post-transfection. Although differences in the transfection efficiency may account for some of the initial difference at 24 h, it appears that the cells expressing mutant ARX protein persist in the population longer than cells expressing the ARX-Wt protein. The overall levels of mitosis (in both transfected and un-transfected cells) were consistently between 4%-5% of the total cell population, across all times and treatment groups (data not shown). However, there was a greater contribution to the mitotic pool in cells expressing mutant ARX protein compared to the ARX-Wt protein, increasingly prevalent with longer times post-transfection (Figure 8B). For example, although there were 1.3-fold more cells expressing R332P than ARX-Wt at 24 h, there were 1.6-fold more mitotic cells expressing the mutant protein compared to ARX-Wt protein. This difference was doubled at the 48 h time point, with 2.2-fold higher cells transfected but 4.4-fold more mitotic cells in cells with R332P expression compared to ARX-Wt. By 72 h post-transfection, although there were 3.9-fold more cells expressing R332P compared to ARX-Wt, the proportion undergoing mitosis was 4.7-fold higher in favour of the ARX mutant.

In order to examine if there were any gross disturbances to the progression of the mitosis we scored each mitotic cell as being either: (1) early phase: representative of DNA condensation during prophase and prometaphase; (2) middle phase - DNA lining up across the mitotic spindle representative of metaphase; or (3) late phase - representative of anaphase through to telophase and cytokinesis. Each of these characteristic phases of cell division was readily delineated using normal light microscopy. Each mitotic cell was also scored as transfected or un-transfected. The data for cells without ARX expression undergoing mitosis for each time point was pooled. Not surprisingly, these cells accounted for an increasing proportion of mitotic cells across time points

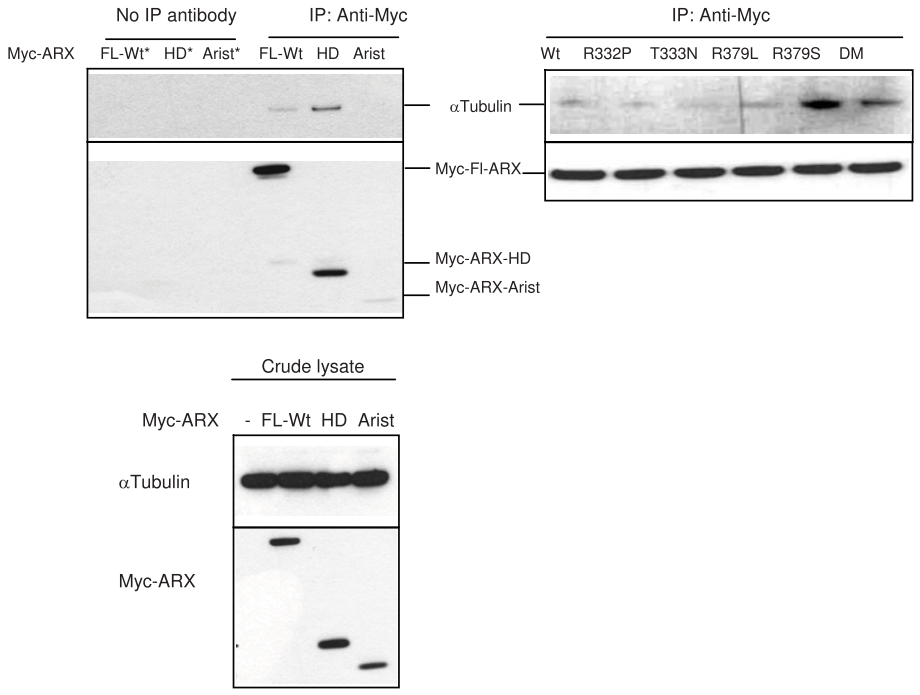

Figure 7 Alpha tubulin co-immunoprecipitates with ARX in mammalian cells. HEK293T cells transfected with Myc-ARX full length protein (FL-Wt), Myc-ARX-homeodomain (HD) or Myc-ARX-Aristaless (Arist) were lysed and proteins immunoprecipitated (IP) with mouse anti-Myc antibody. Samples were loaded on 4%-12% SDS-PAGE gels and analysed for the presence of endogenous alpha tubulin precipitated with Myc-ARX constructs by immunoblotting with rabbit anti-alpha tubulin antibody and goat anti-rabbit IgG conjugated to horseradish-peroxidase (HRP; top panel). Specific IP of Myc-ARX was achieved with no band present in samples from cells transfected with each of the Myc-ARX constructs but no IP antibody added (*). Myc-ARX full length (~62.2 kDa), Myc-ARX-HD (~19 kDa), Myc-ARX-Arist (~15 kDa) and endogenous alpha tubulin (~50 kDa) are present in crude protein lysates (bottom panel). HEK293T cells transfected with Myc-ARX NLS mutations were subjected to IP as above and alpha-tubulin was detected co-immunoprecipitating with all mutant ARX proteins (Left panel).

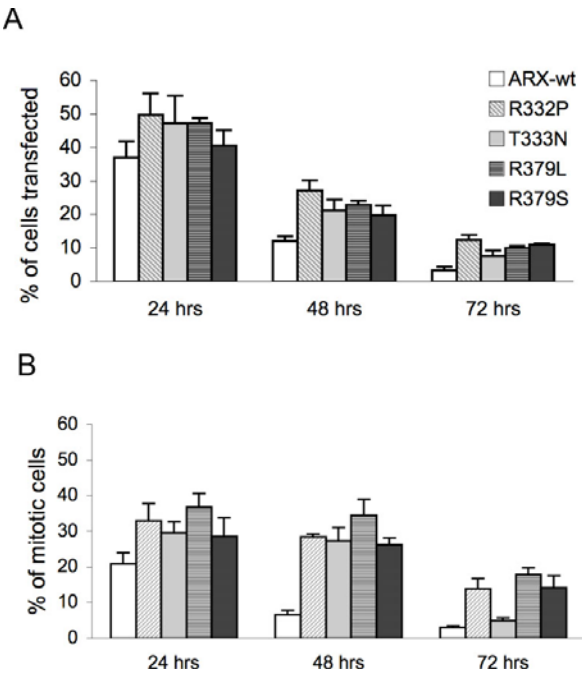

Figure 8 Increased proportion of mitotic cells transfected with nuclear localization sequence (NLS) mutant ARX. (A) The percentage of HEK293T cells transfected and (B) the percentage of mitotic cells transfected with Myc-ARX-Wt or NLS mutant constructs for 24, 48 and 72 h was determined from between 1200 and 4000 cells per sample, from at least three separate transfection reactions. Full-length Myc-tagged constructs transfected are indicated in the legend ARX Wt (open), R332P (vertical light grey stripes), T333N (grey), R379L (horizontal dark grey) and R379S (dark grey) with the time post-transfection indicated on the X-axis.

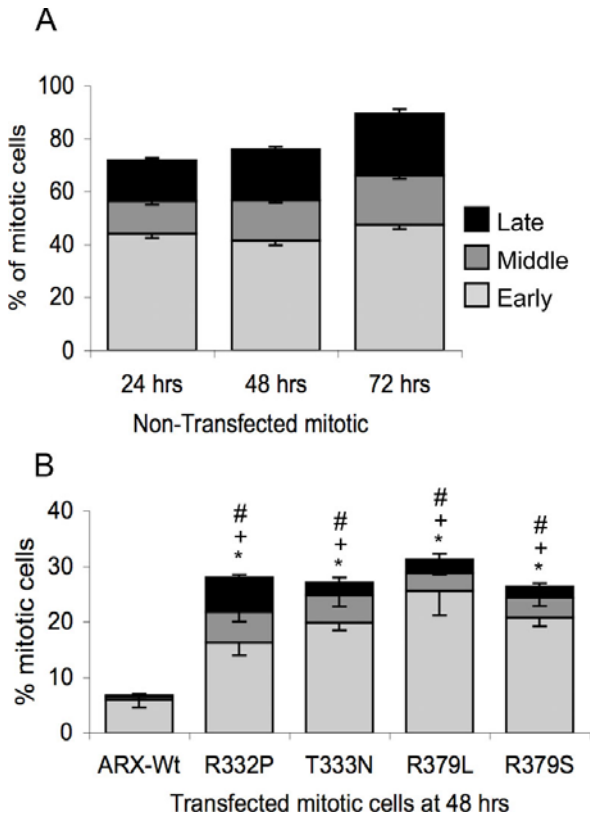

Figure 9 Cells transfected with nuclear localization sequence (NLS) mutant ARX accumulate in early mitosis. (A) The percentage of un-transfected HEK293T cells undergoing various phases of mitosis at 24, 48 and 72 h in culture (time indicated on the X-axis) were determined from between 1200 and 4000 cells per sample, from at least three separate transfection reactions. (B) The percentage of mitotic cells at the DNA condensation stage (early phase, light grey), mitotic spindle stage (middle phase, dark grey) or undergoing anaphase/telophase/cytokinesis (late phase, black) are shown for cells expressing Myc-ARX-Wt or NLS mutations 48 h post-transfection. Full-length Myc-tagged ARX used in each group is indicated on the X-axis. Differences in the percentage of cells undergoing each phase of mitosis at different times post-transfection were compared using a two-way ANOVA. $P < 0.05$ versus ARX-Wt for * early phase, + middle phase and # late phase.

post-transfection, from 71% to 76% and 89% at 24, 48 and 72 h, respectively. Within each time point, over half of the mitotic cells without ARX expression were found to exist in the early phase with the remaining cells split between the middle and late phases (Figure 9A). Although the proportion of these mitotic cells increased with time post-transfection, the distribution across these phases was consistent.

When mitotic cells expressing mutant ARX proteins were analysed we saw an accumulation of cells undergoing mitosis compared to the cells expressing ARX-Wt (Figure 9B). In particular, there was a significant increase in the proportion of cells in the early phase of mitosis at 24, 48 and 72 h post-transfection in cells expressing the ARX NLS mutant proteins compared to cells expressing ARX-Wt (with the exception of T333N at 72 h). In cells expressing the ARX NLS mutations there was also a significant increase in the proportion of cells in the middle phase of mitosis at both 24 and 48 h post-transfection compared to the ARX-Wt protein. Interestingly, the percentages of cells in the late phase of mitosis expressing NLS mutations were also significantly higher than ARX-Wt at 48 h post-transfection (Figure 9B) but not at 72 h (data not shown). This indicates that cells expressing the ARX NLS mutation proteins are progressing through cell division at a slower rate compared to the ARX-Wt protein.

Discussion

We have established that naturally occurring patient mutations leading to non-synonymous changes of single, specific residues of the NLS regions cause significant disruption to nuclear localization of the mutant ARX protein *in vitro*. These findings were consistent in all NLS2 and NLS3 mutations tested, indicating the

mechanism involved may be similar for missense mutations in both NLS2 and NLS3 regions of the ARX homeodomain. Consistent with the severity of clinical outcomes, mutations leading to severe malformation phenotypes had the highest percentages of cells with abnormal mutant protein localization while the mutation causing ISSX but not brain malformation gave the smallest, but still significant, increase in cells with abnormal protein localization compared to the ARX-Wt protein. In contrast to the NLS mutations, another ARX mutation in the middle of the homeodomain, but outside of the NLS regions, did not disrupt the nuclear localization of the mutant protein with the subcellular localization of the mutant protein the same as the ARX-Wt protein.

The NLS regions flanking the ARX homeodomain are predicted to be important in the bi-partite binding to IPO13 [33,34]. We wanted to investigate if changing single, specific residues of the NLS regions would diminish or potentially abolish binding of the mutant NLS with IPO13. The interaction of importins and cargo proteins are often difficult to visualize or measure, as these interactions are highly transient with efficient recycling of the importin proteins back to the cytoplasm after delivery of the cargo to the nucleus. In order to capture this interaction, N-terminal truncated V5-IPO13 was engineered to enable binding to cargo and transport across the nucleus coupled with an inability to interact with RanGTP and dissociate upon reaching the nucleus [33]. Both the Co-IP and co-localization results indicate single amino acid substitutions in either the NLS2 or NLS3 regions of the ARX homeodomain did not abolish binding of the mutant ARX protein to N-terminally truncated V5-IPO13. Moreover, protein containing both R332P and R379L mutations was still able to bind to V5-IPO13.

Interestingly, binding of mutant ARX protein to endogenous IPO13 was indistinguishable from the binding to the N-terminally truncated IPO13 lacking the RanGTPase binding domain. The interaction of endogenous IPO13 with mutant ARX protein implies that the binding of these two proteins not only occurs but that this complex is either unable to correctly transport across the nuclear pore or unable to discharge the ARX cargo once inside the nucleus, or both. We contend that the inability of mutant ARX to uncouple from endogenous IPO13 contributes to the overall increase in cells with abnormal localisation of ARX protein. In support of this suggestion, co-transfection of ARX-Wt protein with the N-terminally truncated IPO13 led to a marked increase in the proportion of transfected cells with abnormal localization compared to cells transfected with ARX-Wt alone. Hence, disruption of normal nuclear delivery and distribution of mutant ARX protein may be due to a compromised ability of the mutant ARX-IPO13 complex to uncouple in the presence of RanGTP. Inadequate nuclear accumulation of a transcription factor due to impaired interaction with importin-β has been suggested as the pathogenic mechanism behind mutations in the C-terminal NLS of the *SRY* gene leading to XY genotype developing as females [38]. A similar scenario of inadequate accumulation and distribution of the mutant ARX protein within the nucleus might ultimately mimic complete absence of ARX and contribute to the catastrophic phenotypic consequences routinely observed in patients with these mutations. This prediction fits with the emerging genotype-phenotype pattern for mutations in ARX. In particular, naturally occurring mutations such as insertions [5,7,9,12], deletions [5,28-31], nonsense changes [30] and splice mutations [30] in ARX that result in protein truncation and loss-of function of the mature ARX protein invariably lead to severe phenotypes, including XLAG.

Both arginine residues examined in this study, R332 in NLS2 and R379 in NLS3 of ARX, are invariant in all 26 paired-type homeodomain proteins. Within this family of proteins these two basic residues are frequent sites of missense and nonsense mutations associated with a range of diseases. A mutation in the homologous NLS2 residue has been reported for *ALX4* (MIM 605420) [39] and mutations in the homologous NLS3 residue have been identified in *CRX* (MIM 602225) [40], *OTX2* (MIM 600037) [41] and *SHOX* (MIM 312865) [42,43]. Mutations in both residues corresponding to the R332 and R379 of ARX have been identified in *PAX3* (MIM 606597) [44,45], *PAX6* (MIM 607108) [46,47], *PITX2* (MIM 601542) [48-50] and *PROP1* (MIM 601538) [51-54]. A mutation in the conserved residue at position 50 of the paired-type homeodomain *SHOX* results in the abolition of DNA binding [55]. This residue corresponds to R379 of ARX. However, this residue is not part of the NLS region of the SHOX protein. When a residue within the NLS is mutated (R173C of SHOX) this change does not affect the DNA binding but instead disrupts the nuclear transport of SHOX. This may also be the case with the mutations tested in our study. There is no doubt that some of the missense mutations in the ARX homeodomain are likely to be important in either specificity of binding or in the actual binding to DNA. For example, the P353L mutation did not disrupt nuclear location of the resulting protein. Perhaps this mutation leads to the severe XMESID phenotype due to changes in binding to DNA targets. The recent identification of a specific transcription factor-binding site for Arx in a study ablating *Arx* in the sub-palluim of the mouse brain [56] will provide an important tool to investigate the potential mechanisms underlying the pathogenesis of mutations in the ARX homeodomain in the future.

A recent examination of heterozygous females from families with known mutations in *ARX* has highlighted the fact that in some cases mutations that disrupt ARX in females may have pathogenic consequences [32]. In particular, a number of females with the T333N and R379S mutations display mental retardation or learning disabilities, although other female carriers of the same mutation are phenotypically normal [30,32]. Our data indicates that in addition to a predicted loss of normal transcription factor activity of ARX, sequestration of mutant ARX with IPO13 may lead to a disruption of interactions with other protein partners, contributing to the disease phenotype. An emerging feature of many forms of lissencephaly and pachygyria is the potential disruption of key elements of microtubule behaviour. For example, one of the key genes mutated in lissencephaly is doublecortin (*DCX*). Mutations in *DCX* are clustered in two tubulin binding domains and impair the polymerization of microtubules and as such correct neuronal migration [57]. Interestingly, DCX does not bind to the tubulin heterodimer itself but acts to nucleate the microtubule growth and to stabilize microtubules. More recently, mutations in *TUBA1A* have been reported to cause lissencephaly with a distinct clinical presentation, ranging from perisylvian pachygyria in the less severe form, to posteriorly predominant pachygyria in the most severe form, in association with dysgenesis of the anterior limb of the internal capsule and mild to severe cerebellar hypoplasia [58]. A recurrent mutation in *TUBA1A* compromises the efficiency of *de novo* alpha/beta-tubulin heterodimer formation [59]. We have shown that ARX interacts with TUBA1A by Co-IP, but the two proteins do not appear to co-localize during normal cell growth. However, in cells expressing mutant ARX we observed a disruption of the mitotic spindle structures and an accumulation of cells in the early stages of mitosis. Hence, neurons, which normally express ARX during the development of the sequestration of mutant ARX, might compromise the interaction with TUBA1A, disrupt normal microtubule formation and subsequently retard efficient cell division.

We cannot rule out the possibility that sequestration of IPO13 in these cells may also contribute to pathogenesis. The dynamic cycling of Ran between the GTP and GDP bound forms is exquisitely controlled by specific regulators differentially localized within the cellular environment [60]. The RanGAP, which accelerates GTP hydrolysis, is cytoplasmic. The Ran guanine nucleotide exchange factor, also known as RCC1 (regulator of chromatin condensation 1), is located in the nucleus bound to chromatin. Hence, the levels of RanGTP increase in proximity to chromatin. Importin proteins bind and inhibit spindle accessory factors (SAF) everywhere in the mitotic cytosol, except in the vicinity of the chromosome. The high levels of RanGTP close to the chromatin relieve the inhibition of SAFs by importins and subsequently allow local spindle assembly. RanGTP binding to importin protein generates conformation change in the Importin molecule that in turn alters the binding site of the importin for the cargo proteins [37]. Therefore, IPO13 trapped in complex with mutant ARX may not be able to adequately bind SAFs and as such compromise spindle assembly. Samples of the cortex taken at autopsy from the patient with the R379L mutation were stained with neurofilament immunoperoxidase and, interestingly, a paucity of neurons was noted (D Coman, unpublished data). The authors could not rule out the contribution of perinatal ischemia to the histological findings. However, in light of our findings, it is interesting to speculate that disruption of the mitotic spindle due to sequestration of mutant ARX with IPO13 retards cell division, potentially contributing to a disruption in neuronal migration and correct lamination patterns of the brain leading to lissencephaly and associated clinical outcomes.

Abbreviations

ARX: *Aristaless* related homeobox; HRP: horseradish-peroxidase; IP: Immunoprecipitation; IPO13: Importin 13; MRI: magnetic resonance imaging; NLS: Nuclear localization sequence; PBS: phosphate buffered saline; SAF: spindle accessory factor; TBS: tris buffered saline; TUBA1A: Tubulin A1A; XLAG: X-linked lissencephaly with ambiguous genitalia; XLID: X-linked intellectual disability; XMESID: X-linked myoclonic epilepsy with spasticity and intellectual disability.

Acknowledgements

The authors would like to thank the patients and their family members and Nancy Briggs (University of Adelaide, Australia) for assistance with statistical analysis and Dr Helger Yntema (Radboud University Nijmegen Medical Centre, The Netherlands) for performing ARX analysis in family 2. This work was supported by an Australian NHMRC programme grant (JG), a NHMRC Training Fellowship (CS) and a NHMRC Senior Research Fellowship (JG).

Author details

[1]Department of Genetics and Molecular Pathology, SA Pathology at the Women's and Children's Hospital, North Adelaide, South Australia 5006, Australia. [2]Department of Paediatrics, University of Adelaide, Adelaide, South Australia 5001, Australia. [3]Department of Metabolic Medicine, The Royal Children's Hospital, Brisbane, Queensland 4029, Australia. [4]Genetics Health Services Victoria, Murdoch Children's Research Institute, Melbourne, Victoria 3052, Australia. [5]Department of Clinical Genetics, Erasmus University Medical Centre, Rotterdam 3015GE, The Netherlands. [6]Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands.

Authors' contributions

GM, D Coman, TK and GMM contributed samples and clinical data from affected individuals. D Cloosterman performed yeast-2 hybrid studies. CS, MHT and TF performed cloning, cell culture studies and functional assays and analysed the data. CS and JG conceived and designed the study. CS directed the study and wrote the first draft of the manuscript. All authors contributed to discussion of the results and manuscript preparation. All authors have read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 17 October 2009
Accepted: 5 January 2010 Published: 5 January 2010

References

1. Gecz J, Shoubridge C, Corbett M: **The genetic landscape of intellectual disability arising from chromosome X.** *Trends Genet* 2009, **25(7)**:308-316.
2. Miura H, Yanazawa M, Kato K, Kitamura K: **Expression of a novel aristaless related homeobox gene 'Arx' in the vertebrate telencephalon, diencephalon and floor plate.** *Mech Dev* 1997, **65(1-2)**:99-109.
3. Stromme P, Mangelsdorf ME, Shaw MA, Lower KM, Lewis SM, Bruyere H, Lutcherath V, Gedeon AK, Wallace RH, Scheffer IE, Turner G, Partington M, Frints SG, Fryns JP, Sutherland GR, Mulley JC, Gécz J: **Mutations in the human ortholog of Aristaless cause X-linked mental retardation and epilepsy.** *Nat Genet* 2002, **30(4)**:441-445.
4. Partington MW, Turner G, Boyle J, Gecz J: **Three new families with X-linked mental retardation caused by the 428-451dup(24 bp) mutation in ARX.** *Clin Genet* 2004, **66(1)**:39-45.
5. 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 WB, Yokoyama M, Morohashi K: **Mutation of ARX causes abnormal development of forebrain and testes in mice and X-linked lissencephaly with abnormal genitalia in humans.** *Nat Genet* 2002, **32(3)**:359-369.
6. Gecz J, Cloosterman D, Partington M: **ARX: a gene for all seasons.** *Curr Opin Genet Dev* 2006, **16(3)**:308-316.
7. Hartmann H, Uyanik G, Gross C, Hehr U, Lucke T, Arslan-Kirchner M, Antosch B, Das AM, Winkler J: **Agenesis of the corpus callosum, abnormal genitalia and intractable epilepsy due to a novel familial mutation in the Aristaless-related homeobox gene.** *Neuropediatrics* 2004, **35(3)**:157-160.
8. Guerrini R, Moro F, Kato M, Barkovich AJ, Shiihara T, McShane MA, Hurst J, Loi M, Tohyama J, Norci V, Hayasaka K, Kang UJ, Das S, Dobyns WB: **Expansion of the first PolyA tract of ARX causes infantile spasms and status dystonicus.** *Neurology* 2007, **69(5)**:427-433.
9. Jagla M, Kruczek P, Kwinta P: **Association between X-linked lissencephaly with ambiguous genitalia syndrome and lenticulostriate vasculopathy in neonate.** *J Clin Ultrasound* 2008, **36(6)**:387-390.
10. Kato M, Saitoh S, Kamei A, Shiraishi H, Ueda Y, Akasaka M, Tohyama J, Akasaka N, Hayasaka K: **A longer polyalanine expansion mutation in the ARX gene causes early infantile epileptic encephalopathy with suppression-burst pattern (Ohtahara syndrome).** *Am J Hum Genet* 2007, **81(2)**:361-366.
11. Nawara M, Szczaluba K, Poirier K, Chrzanowska K, Pilch J, Bal J, Chelly J, Mazurczak T: **The ARX mutations: a frequent cause of X-linked mental retardation.** *Am J Med Genet A* 2006, **140(7)**:727-732.
12. Okazaki S, Ohsawa M, Kuki I, Kawawaki H, Koriyama T, Ri S, Ichiba H, Hai E, Inoue T, Nakamura H, Goto Y, Tomiwa K, Yamano T, Kitamura K, Itoh M: **Aristaless-related homeobox gene disruption leads to abnormal distribution of GABAergic interneurons in human neocortex: evidence based on a case of X-linked lissencephaly with abnormal genitalia (XLAG).** *Acta Neuropathol* 2008, **116(4)**:453-462.
13. Poirier K, Eisermann M, Caubel I, Kaminska A, Peudonnier S, Boddaert N, Saillour Y, Dulac O, Souville I, Beldjord C, Lascelles K, Plouin P, Chelly J, Bahi-Buisson N: **Combination of infantile spasms, non-epileptic seizures and complex movement disorder: a new case of ARX-related epilepsy.** *Epilepsy Res* 2008, **80(2-3)**:224-228.
14. Szczaluba K, Nawara M, Poirier K, Pilch J, Gajdulewicz M, Spodar K, Chelly J, Bal J, Mazurczak T: **Genotype-phenotype associations for ARX gene duplication in X-linked mental retardation.** *Neurology* 2006, **67(11)**:2073-2075.
15. Laperuta C, Spizzichino L, D'Adamo P, Monfregola J, Maiorino A, D'Eustacchio A, Ventruto V, Neri G, D'Urso M, Chiurazzi P, Ursini MV, Miano MG: **MRX87 family with Aristaless X dup24 bp mutation and implication for polyAlanine expansions.** *BMC Med Genet* 2007, **8**:25.
16. Rujirabanjerd S, Tongsippunyoo K, Sripo T, Limprasert P: **Mutation screening of the Aristaless-related homeobox (ARX) gene in Thai pediatric patients with delayed development: first report from Thailand.** *Eur J Med Genet* 2007, **50(5)**:346-354.
17. Shinozaki Y, Osawa M, Sakuma H, Komaki H, Nakagawa E, Sugai K, Sasaki M, Goto Y: **Expansion of the first polyalanine tract of the ARX gene in a boy presenting with generalized dystonia in the absence of infantile spasms.** *Brain Dev* 2008, **31(6)**:469-72.
18. Troester MM, Trachtenberg T, Narayanan V: **A novel mutation of the ARX gene in a male with nonsyndromic mental retardation.** *J Child Neurol* 2007, **22(6)**:744-748.
19. Wallerstein R, Sugalski R, Cohn L, Jawetz R, Friez M: **Expansion of the ARX spectrum.** *Clin Neurol Neurosurg* 2008, **110(6)**:631-634.
20. Berry-Kravis E, Israel J: **X-linked pachygyria and agenesis of the corpus callosum: evidence for an X chromosome lissencephaly locus.** *Ann Neurol* 1994, **36(2)**:229-233.
21. Dobyns WB, Berry-Kravis E, Havernick NJ, Holden KR, Viskochil D: **X-linked lissencephaly with absent corpus callosum and ambiguous genitalia.** *Am J Med Genet* 1999, **86(4)**:331-337.
22. des Portes V, Francis F, Pinard JM, Desguerre I, Moutard ML, Snoeck I, Meiners LC, Capron F, Cusmai R, Ricci S, Motte J, Echenne B, Ponsot G, Dulac O, Chelly J, Beldjord C: **doublecortin is the major gene causing X-linked subcortical laminar heterotopia (SCLH).** *Hum Mol Genet* 1998, **7(7)**:1063-1070.
23. Reiner O, Carrozzo R, Shen Y, Wehnert M, Faustinella F, Dobyns WB, Caskey CT, Ledbetter DH: **Isolation of a Miller-Dieker lissencephaly gene containing G protein beta-subunit-like repeats.** *Nature* 1993, **364(6439)**:717-721.
24. Hong SE, Shugart YY, Huang DT, Shahwan SA, Grant PE, Hourihane JO, Martin ND, Walsh CA: **Autosomal recessive lissencephaly with cerebellar hypoplasia is associated with human RELN mutations.** *Nat Genet* 2000, **26(1)**:93-96.
25. Poirier K, Keays DA, Francis F, Saillour Y, Bahi N, Manouvrier S, Fallet-Bianco C, Pasquier L, Toutain A, Tuy FP, Bienvenu T, Joriot S, Odent S, Ville D, Desguerre I, Goldenberg A, Moutard ML, Fryns JP, van Esch H, Harvey RJ, Siebold C, Flint J, Beldjord C, Chelly J: **Large spectrum of lissencephaly and pachygyria phenotypes resulting from *de novo* missense mutations in tubulin alpha 1A (TUBA1A).** *Hum Mutat* 2007, **28(11)**:1055-1064.
26. Keays DA, Tian G, Poirier K, Huang GJ, Siebold C, Cleak J, Oliver PL, Fray M, Harvey RJ, Molnar Z, Piñon MC, Dear N, Valdar W, Brown SD, Davies KE, Rawlins JN, Cowan NJ, Nolan P, Chelly J, Flint J: **Mutations in alpha-tubulin cause abnormal neuronal migration in mice and lissencephaly in humans.** *Cell* 2007, **128(1)**:45-57.
27. Boycott KM, Flavelle S, Bureau A, Glass HC, Fujiwara TM, Wirrell E, Davey K, Chudley AE, Scott JN, McLeod DR, Parboosingh JS: **Homozygous deletion of the very low density lipoprotein receptor gene causes autosomal recessive cerebellar hypoplasia with cerebral gyral simplification.** *Am J Hum Genet* 2005, **77(3)**:477-483.
28. Uyanik G, Aigner L, Martin P, Gross C, Neumann D, Marschner-Schafer H, Hehr U, Winkler J: **ARX mutations in X-linked lissencephaly with abnormal genitalia.** *Neurology* 2003, **61(2)**:232-235.
29. Hahn A, Gross C, Uyanik G, Hehr U, Hugens-Penzel M, Alzen G, Neubauer BA: **X-linked lissencephaly with abnormal genitalia associated with renal phosphate wasting.** *Neuropediatrics* 2004, **35(3)**:202-205.
30. Kato M, Das S, Petras K, Kitamura K, Morohashi K, Abuelo DN, Barr M, Tonneau D, Brady AF, Carpenter NJ, Cipero KL, Frisone F, Fukuda T, Guerrini R, Iida E, Itoh M, Lewanda AF, Nanba Y, Oka A, Proud VK, Saugier-Veber P, Schelley SL, Selicorni A, Shaner R, Silengo M, Stewart F, Sugiyama N, Toyama J, Toutain A, Vargas AL, Yanazawa M, Zackai EH, Dobyns WB: **Mutations of ARX are associated with striking pleiotropic and consistent genotype-phenotype correlation.** *Hum Mutat* 2004, **23(2)**:147-159.
31. Bhat SS, Rogers RC, Holden KR, Srivastava AK: **A novel in-frame deletion in ARX is associated with lissencephaly with absent corpus callosum and hypoplastic genitalia.** *Am J Med Genet A* 2005, **138(1)**:70-72.
32. Marsh E, Fulp C, Gomez E, Nasrallah I, Minarcik J, Sudi J, Christian SL, Mancini G, Labosky P, Dobyns W, Brooks-Kayal A, Golden JA: **Targeted loss of Arx results in a developmental epilepsy mouse model and recapitulates the human phenotype in heterozygous females.** *Brain* 2009, **132(Pt 6)**:1563-1576.
33. Shoubridge C, Cloosterman D, Parkinson-Lawerence E, Brooks D, Gecz J: **Molecular pathology of expanded polyalanine tract mutations in the Aristaless-related homeobox gene.** *Genomics* 2007, **90(1)**:59-71.
34. Ploski JE, Shamsher MK, Radu A: **Paired-type homeodomain transcription factors are imported into the nucleus by karyopherin 13.** *Mol Cell Biol* 2004, **24(11)**:4824-4834.
35. Mingot JM, Kostka S, Kraft R, Hartmann E, Gorlich D: **Importin 13: a novel mediator of nuclear import and export.** *EMBO J* 2001, **20(14)**:3685-3694.

36. Lin W, Ye W, Cai L, Meng X, Ke G, Huang C, Peng Z, Yu Y, Golden JA, Tartakoff AM, Tao T: **The roles of multiple importins for nuclear import of murine aristaless-related homeobox protein.** *J Biol Chem* 2009, **284(30)**:20428-20439.
37. Lee SJ, Matsuura Y, Liu SM, Stewart M: **Structural basis for nuclear import complex dissociation by RanGTP.** *Nature* 2005, **435(7042)**:693-696.
38. Harley VR, Layfield S, Mitchell CL, Forwood JK, John AP, Briggs LJ, McDowall SG, Jans DA: **Defective importin beta recognition and nuclear import of the sex-determining factor SRY are associated with XY sex-reversing mutations.** *Proc Natl Acad Sci USA* 2003, **100(12)**:7045-7050.
39. Mavrogiannis LA, Antonopoulou I, Baxova A, Kutilek S, Kim CA, Sugayama SM, Salamanca A, Wall SA, Morriss-Kay GM, Wilkie AO: **Haploinsufficiency of the human homeobox gene ALX4 causes skull ossification defects.** *Nat Genet* 2001, **27(1)**:17-18.
40. Swaroop A, Wang QL, Wu W, Cook J, Coats C, Xu S, Chen S, Zack DJ, Sieving PA: **Leber congenital amaurosis caused by a homozygous mutation (R90W) in the homeodomain of the retinal transcription factor CRX: direct evidence for the involvement of CRX in the development of photoreceptor function.** *Hum Mol Genet* 1999, **8(2)**:299-305.
41. Ragge NK, Brown AG, Poloschek CM, Lorenz B, Henderson RA, Clarke MP, Russell-Eggitt I, Fielder A, Gerrelli D, Martinez-Barbera JP, Ruddle P, Hurst J, Collin JR, Salt A, Cooper ST, Thompson PJ, Sisodiya SM, Williamson KA, Fitzpatrick DR, van Heyningen V, Hanson IM: **Heterozygous mutations of OTX2 cause severe ocular malformations.** *Am J Hum Genet* 2005, **76(6)**:1008-1022.
42. Rappold GA, Fukami M, Niesler B, Schiller S, Zumkeller W, Bettendorf M, Heinrich U, Vlachopapadoupoulou E, Reinehr T, Onigata K, Ogata T: **Deletions of the homeobox gene SHOX (short stature homeobox) are an important cause of growth failure in children with short stature.** *J Clin Endocrinol Metab* 2002, **87(3)**:1402-1406.
43. Binder G, Renz A, Martinez A, Keselman A, Hesse V, Riedl SW, Hausler G, Fricke-Otto S, Frisch H, Heinrich JJ, Ranke MB: **SHOX haploinsufficiency and Leri-Weill dyschondrosteosis: prevalence and growth failure in relation to mutation, sex, and degree of wrist deformity.** *J Clin Endocrinol Metab* 2004, **89(9)**:4403-4408.
44. DeStefano AL, Cupples LA, Arnos KS, Asher JH Jr, Baldwin CT, Blanton S, Carey ML, da Silva EO, Friedman TB, Greenberg J, Lalwani AK, Milunsky A, Nance WE, Pandya A, Ramesar RS, Read AP, Tassabejhi M, Wilcox ER, Farrer LA: **Correlation between Waardenburg syndrome phenotype and genotype in a population of individuals with identified PAX3 mutations.** *Hum Genet* 1998, **102(5)**:499-506.
45. Tassabehji M, Newton VE, Liu XZ, Brady A, Donnai D, Krajewska-Walasek M, Murday V, Norman A, Obersztyn E, Reardon W, *et al*: **The mutational spectrum in Waardenburg syndrome.** *Hum Mol Genet* 1995, **4(11)**:2131-2137.
46. Wolf M, Zabel B, Lorenz B, Blankenagel A, Ghorbani MB, Schwenn O, Wildhardt G: **Molecular genetic study of the PAX6 gene in aniridia patients.** *Ophthalmologe* 1998, **95(12)**:828-830.
47. Gronskov K, Rosenberg T, Sand A, Brondum-Nielsen K: **Mutational analysis of PAX6: 16 novel mutations including 5 missense mutations with a mild aniridia phenotype.** *Eur J Hum Genet* 1999, **7(3)**:274-286.
48. Chi YI: **Homeodomain revisited: a lesson from disease-causing mutations.** *Hum Genet* 2005, **116(6)**:433-444.
49. Perveen R, Lloyd IC, Clayton-Smith J, Churchill A, van Heyningen V, Hanson I, Taylor D, McKeown C, Super M, Kerr B, Winter R, Black GC: **Phenotypic variability and asymmetry of Rieger syndrome associated with PITX2 mutations.** *Invest Ophthalmol Vis Sci* 2000, **41(9)**:2456-2460.
50. Phillips JC: **Four novel mutations in the PITX2 gene in patients with Axenfeld-Rieger syndrome.** *Ophthalmic Res* 2002, **34(5)**:324-326.
51. Duquesnoy P, Roy A, Dastot F, Ghali I, Teinturier C, Netchine I, Cacheux V, Hafez M, Salah N, Chaussain JL, Goossens M, Bougnères P, Amselem S: **Human Prop-1: cloning, mapping, genomic structure. Mutations in familial combined pituitary hormone deficiency.** *FEBS Lett* 1998, **437(3)**:216-220.
52. Wu W, Cogan JD, Pfaffle RW, Dasen JS, Frisch H, O'Connell SM, Flynn SE, Brown MR, Mullis PE, Parks JS, Phillips JA, Rosenfeld MG: **Mutations in PROP1 cause familial combined pituitary hormone deficiency.** *Nat Genet* 1998, **18(2)**:147-149.
53. Vallette-Kasic S, Barlier A, Teinturier C, Diaz A, Manavela M, Berthezene F, Bouchard P, Chaussain JL, Brauner R, Pellegrini-Bouiller I, Jaquet P, Enjalbert A, Brue T: **PROP1 gene screening in patients with multiple pituitary hormone deficiency reveals two sites of hypermutability and a high incidence of corticotroph deficiency.** *J Clin Endocrinol Metab* 2001, **86(9)**:4529-4535.
54. Mody S, Brown MR, Parks JS: **The spectrum of hypopituitarism caused by PROP1 mutations.** *Best Pract Res Clin Endocrinol Metab* 2002, **16(3)**:421-431.
55. Schneider KU, Marchini A, Sabherwal N, Roth R, Niesler B, Marttila T, Blaschke RJ, Lawson M, Dumic M, Rappold G: **Alteration of DNA binding, dimerization, and nuclear translocation of SHOX homeodomain mutations identified in idiopathic short stature and Leri-Weill dyschondrosteosis.** *Hum Mutat* 2005, **26(1)**:44-52.
56. Fulp CT, Cho G, Marsh ED, Nasrallah IM, Labosky PA, Golden JA: **Identification of Arx transcriptional targets in the developing basal forebrain.** *Hum Mol Genet* 2008, **17(23)**:3740-3760.
57. Sapir T, Horesh D, Caspi M, Atlas R, Burgess HA, Wolf SG, Francis F, Chelly J, Elbaum M, Pietrokovski S, Reiner O: **Doublecortin mutations cluster in evolutionarily conserved functional domains.** *Hum Mol Genet* 2000, **9(5)**:703-712.
58. Bahi-Buisson N, Poirier K, Boddaert N, Saillour Y, Castelnau L, Philip N, Buyse G, Villard L, Joriot S, Marret S, Bourgeois M, Van Esch H, Lagae L, Amiel J, Hertz-Pannier L, Roubertie A, Rivier F, Pinard JM, Beldjord C, Chelly J: **Refinement of cortical dysgeneses spectrum associated with TUBA1A mutations.** *J Med Genet* 2008, **45(10)**:647-653.
59. Tian G, Kong XP, Jaglin XH, Chelly J, Keays D, Cowan NJ: **A pachygyria-causing alpha-tubulin mutation results in inefficient cycling with CCT and a deficient interaction with TBCB.** *Mol Biol Cell* 2008, **19(3)**:1152-1161.
60. Kalab P, Heald R: **The RanGTP gradient - a GPS for the mitotic spindle.** *J Cell Sci* 2008, **121(Pt 10)**:1577-1586.

doi:10.1186/1755-8417-3-1
Cite this article as: Shoubridge *et al.*: Mutations in the nuclear localization sequence of the *Aristaless* related homeobox; sequestration of mutant ARX with IPO13 disrupts normal subcellular distribution of the transcription factor and retards cell division. *PathoGenetics* 2010 **3**:1.

Whibley, A.C., Plagnol, V., Tarpey, P.S., Abidi, F., Fullston, T. & Raymond, F.L. (2010).
Fine-scale survey of X chromosome copy number variants and indels underlying intellectual disability.
American Journal of Human Genetics, v. 87 (2), pp. 173-188

APPENDIX B

Table of published *ARX* mutations, ordered by the position at which they occur in the ORF.

Change (cDNA – bp)	**Exon (Intron)**	**Change (Prtoein – aa)**	**domain**	**Type**	**# Fam total**	**Phenotype**	**MIM**	**REF (# Families in Ref)**
1. Exon1_2del	1-2	not PCR ex1&2		del	1	XLAG	300215	Kitamura *et al.*, 2002
***2. c.81C>G**	**1**	**p.Y27X**	**oct**	**non**	**1**	**OS/ISSX (WS)**	**308350**	**Fullston *et al.*, 2010**
3. c.98T>C	1	p.L33P	**oct**	mis	1	nsXLID		Bienvenu *et al.*, 2002
4. c.112C>T	1	p.P38S		mis	1	nsXLID		Poirier *et al.*, 2005b
5. c.196+2T>C	(1)	Skip exon 1?	**oct**	splice	1	XLAG	300215	Kato *et al.*, 2004
6. Exon2_5del	2-5	p.G66_C562del		del	1	XLAG	300215	Kato *et al.*, 2004
7. c.232G>T	2	p.E78X		non	2	XLAG	300215	(2) Kato *et al.*, 2004
8. c.298_330dup(33bp)	2	p.A111insA_{11}	**pA1**	dup	2	OS	308350	(2) Kato *et al.*, 2007
9. c.304ins$(GCG)_2$	2	p.A111insA_2	**pA1**	ins	1	nsXLID		Bienvenu *et al.*, 2002
10. c.304ins$(GCG)_3$	2	p.A111insA_3	**pA1**	ins	1	nsXLID		Groskov *et al.*, 2004
11. c.304ins$(GCG)_7$	2	p.A111insA_7	**pA1**	ins	12	ISSX (WS) +	308350	ISSX (WS): **(6)** (2) Stromme *et al.*, 2002 (1) Wohlrab *et al.*, 2005 (1) + Poirier *et al.*, 2008 (1) + Wallerstein *et al.*, 2008 (1) Cossée *et al.*, 2011
						IEDE ID, TS, dys OS OS		Other: **(6)** (4) Guerrini *et al.*, 2007 (1) Shinozaki *et al.*, 2008 (1) Absoud *et al.*, 2010
12. c.335_368del(34bp)	2	p.A112fs	**pA1**	del	1	XLAG	300215	Kato *et al.*, 2004
13. c.392_452del(61bp)	2	p.P131fs		del	1	XLAG	300215	Kato *et al.*, 2004
14. c.420_451del(32bp)	2	p.D140fs		del	1	XLAG	300215	Kitamura *et al.*, 2002
***15. c.423_455dup(33bp)**	**2**	**p.A151delAA insGA_{12}**	**pA2**	**dup**	**1**	**PRTS +**	**309510**	**Demos *et al.*, 2009**

(Cont. on next page)

Change (cDNA – bp)	Exon (Intron)	Change (Prtoein – aa)	domain	Type	# Fam total	Phenotype	MIM	REF (# Families in Ref) *Cont.*
*16. c.429_452dup(24bp) (Bienvenu 2002) c.428_451dup(24bp) (Stromme 2002) c.431_454dup(24bp) (Gronskov 2004)	2	p.A151insA_8	**pA2**	dup	44	ISSX (WS) PRTS nsXLID	308350 309510	ISSX: **(6)** (1) Stromme *et al.*, 2002 (1) Kato *et al.*, 2003 (1) Partington *et al.*, 2004 (3) Cossée *et al.*, 2011 PRTS: **(6)** (2) Stromme *et al.*, 2002 (1) Bienvenu *et al.*, 2002 (1) Partington *et al.*, 2004 (1) Poirier *et al.*, 2005b (1) Rujirabanjerd *et al.*, 2007 nsXLID: **(32)** (5) Bienvenu *et al.*, 2002 (2) Stromme *et al.*, 2002 (4) Stepp *et al.*, 2005 (1) Partington *et al.*, 2004 (1) Gronskov *et al.*, 2004 (1) Van Esch *et al.*, 2004 (3) Poirier *et al.*, 2005b (5) Nawara *et al.*, 2006 " Szczaluba *et al.*, 2006 (1) Gestinari-Duarte *et al.*, 2006 (1) Laperuta *et al.*, 2007 (1) Rujirabanjerd *et al.*, 2007 **(3) Fullston *et al.*, 2011**
***17. c.430_456dup(27bp)**	**2**	**p.A151insA_9**	**pA2**	**dup**	**1**	**ISSX (WS) +**	**308350**	**Reish *et al.*, 2009**
18. c.488A>G	2	p.Q163R		mis	1	nsXLID		Bienvenu *et al.*, 2002
19. c.617delG	2	p.G206fs		del	1	XLAG-HYD	300215	Kato *et al.*, 2004
20. c.619_647del	2	p.V207_A216delfs		del	1	XLAG	300215	Kato *et al.*, 2004
21. c.790delC	2	p.R264fs		del	3	XLAG	300215	Kitamura *et al.*, 2002 Uyanik *et al.*, 2003 Hahn *et al.*, 2004

(Cont next page)

Change (cDNA – bp)	Exon (Intron)	Change (Prtoein – aa)	domain	Type	# Fam total	Phenotype	MIM	REF (# Families in Ref)	*Cont.*
22. c.856G>A	2	p.G286S		mis	1	nsXLID		Bienvenu *et al.*, 2002	
23. c.980delAACA	2	p.K327fs	**hom**	del	1	XLAG +	300215	Miyata *et al.*, 2009	
24. c.994C>T	2	p.R332C	**hom**	mis	1	XLAG	300215	Uyanik *et al.*, 2003	
25. c.995G>A	2	p.R332H	**hom**	mis	1	XLAG	300215	Kitamura *et al.*, 2002	
26. c.995G>C	2	p.R332P	**hom**	mis	1	XLAG	300215	Kato *et al.*, 2004	
27. c.996_1004del(9bp)	2	p.T333_F335del	**hom**	del	1	XLAG	300215	Bhat *et al.*, 2005	
28. c.998C>A	2	p.T333N	**hom**	mis	1	ACC/AG	300215	Kato *et al.*, 2004	
29. c.1013A>CinsC	2	p.Y338fs	**hom**	ins	1	XLAG	300215	Okazaki *et al.*, 2008	
30. c.1028T>A	2	p.L343Q	**hom**	mis	1	XLAG	300215	Kitamura *et al.*, 2002	
31. c.1058C>G	2	p.P353R	**hom**	mis	1	XLAG	300215	Kato *et al.*, 2004	
32. c.1058C>T	2	p.P353L	**hom**	mis	1	XMESID	300432	Stromme *et al.*, 2002	
33. c.1072A>T	2	p.R358W	**hom**	Mis	1	**ACC/AG/ISSX**	308350 300004	Conti *et al.*, 2011	
***34. c.1074G>T**	**3**	**p.R358S**	**hom**	**mis**	**1**	**ISSX/ACC/AG**	**308350 300004**	**Fullston *et al.*, 2011**	
35. c.1105G>T	3	p.E369X	**hom**	non	1	HYD/AG	300215	Kato *et al.*, 2004	
36. c.1117C>T	3	p.Q373X	**hom**	non	1	HYD/AG	300215	Kitamura *et al.*, 2002	
37. c.1119+1G>C	(3)	skipping of ex3	**hom**	splice	1	XLAG	300215	Kato *et al.*, 2004	
38. c.1120_1448 exon4	4	*No details*	**hom?**	*?*	1	XLAG/LSV	300215	Jagla *et al.*, 2008	
39. c.1135C>A	4	p.R379S	**hom**	mis	1	ISSX	308350	Marsh *et al.*, 2009 Shoubridge *et al.*, 2010	
***40. c.1136G>T**	**4**	**p.R379L**	**hom**	**mis**	**1**	**XLAG**	**300215**	**Shoubridge *et al.*, 2010**	
41. c.1187_1188insC	4	p.P396fs	**hom**	ins	1	XLAG	300215	Kitamura *et al.*, 2002	
42. c.1372delG	4	p.A458fs		del	2	XLAG	300215	(1) Kitamura *et al.*, 2002 (1) Kato *et al.*, 2004	
43. c.1419_1420insAC	4	p.T474fs		ins	1	XLAG/ACC	300215	Hartmann *et al.*, 2004	
44. c.1449-816_*460del	(4)	p.R483fs	**arist**	del	1	ISSX (WS)	308350	Stromme *et al.*, 2002	
45. c.1465delG	5	p.A489fs		del	1	ISSX (WS) fem	308350	Wallerstein *et al.*, 2008	
46. c.1471_1472insC	5	p.L491fsX531	**arist**	ins	1	OS	308350	Ekşioğlu *et al.*, 2011	
47. c.1561G>A	5	p.A521T	**arist**	mis	1	XLAG/LCH	300215	Kato *et al.*, 2004	
48. c.1564_1568dup	5	p.A524fsX534	**arist**	dup	1	OS/WS +	308350	Kato *et al.*, 2010	
49. c.1604T>A	5	p.L535Q	**arist**	mis	1	OS	308350	Giordano *et al.*, 2010	
50. c.1604_1605insT	5	p.E536fsX672	**arist**	ins	1	OS/WS/LGS	308350	Kato *et al.*, 2010	

*** Mutations listed in red with an asterisk are the 6 mutations in separate 8 families presented throughout this thesis**

Total mutations:	50 distinct
Total families	110

Phenotypes

ACC – agenesis of the corpus callosum
AG – abnormal genitalia
Dys – dystonia
Fam – family
Female proband
HYD – hydranenchaply
IEDE – infantile epileptic-dyskinetic encephalopathy
ISSX – infantile spasms, X-linked
LCH – lissencephaly with cerebellar hypoplasia
LSV - lenticulostriate vasculopathy
nsXLID – non syndromic X-linked intellectual disability.
OS – Ohtahara syndrome
PRTS – Partington syndrome.
TS – tonic seizures
WS – West syndrome.
XLAG – X-linked linssencepahly with ambiguous genitalia.
XMESID – X-linked myclonic epilepsy with spasticity and intellectual disability.
+ – additional features.

Mutation types:
del – deletion.
ins – insertion.
mis – missense.
non – nonsense
splice – may alter exon splicing.
? – details not given in report.

BIBLIOGRAPHY

Absoud, M., J. R. Parr, D. Halliday, P. Pretorius, Z. Zaiwalla and S. Jayawant. **2010**. 'A novel ARX phenotype: rapid neurodegeneration with Ohtahara syndrome and a dyskinetic movement disorder' *Dev Med Child Neurol* **52**(3): 305-307.

Ahituv, N., Y. Zhu, A. Visel, A. Holt, V. Afzal, L. A. Pennacchio and E. M. Rubin. **2007**. 'Deletion of ultraconserved elements yields viable mice' *PLoS Biol* **5**(9): e234.

Albrecht, A. and S. Mundlos. **2005**. 'The other trinucleotide repeat: polyalanine expansion disorders' *Curr Opin Genet Dev* **15**(3): 285-293.

Albrecht, A. N., U. Kornak, A. Boddrich, K. Suring, P. N. Robinson, A. C. Stiege, R. Lurz, S. Stricker, E. E. Wanker and S. Mundlos. **2004**. 'A molecular pathogenesis for transcription factor associated poly-alanine tract expansions' *Hum Mol Genet* **13**(20): 2351-2359.

Antonarakis, S. E., R. Lyle, E. T. Dermitzakis, A. Reymond and S. Deutsch. **2004**. 'Chromosome 21 and down syndrome: from genomics to pathophysiology' *Nat Rev Genet* **5**(10): 725-738.

Bachetti, T., I. Matera, S. Borghini, M. Di Duca, R. Ravazzolo and I. Ceccherini. **2005**. 'Distinct pathogenetic mechanisms for PHOX2B associated polyalanine expansions and frameshift mutations in congenital central hypoventilation syndrome' *Hum Mol Genet* **14**(13): 1815-1824.

Ban, M., M. Maranian, T. W. Yeo, J. Gray, A. Compston and S. Sawcer. **2005**. 'Ultraconserved regions in multiple sclerosis' *Eur J Hum Genet* **13**(9): 998-999.

Baskaran, N., R. P. Kandpal, A. K. Bhargava, M. W. Glynn, A. Bale and S. M. Weissman. **1996**. 'Uniform amplification of a mixture of deoxyribonucleic acids with varying GC content' *Genome Res* **6**(7): 633-638.

Bejerano, G., M. Pheasant, I. Makunin, S. Stephen, W. J. Kent, J. S. Mattick and D. Haussler. **2004**. 'Ultraconserved elements in the human genome' *Science* **304**(5675): 1321-1325.

Bell, J. T. and T. D. Spector. **2011**. 'A twin approach to unraveling epigenetics' *Trends Genet* **27**(3): 116-125.

Berman, R. F. and R. Willemsen. **2009**. 'Mouse models of fragile X-associated tremor ataxia' *J Investig Med* **57**(8): 837-841.

Berry-Kravis, E. and J. Israel. **1994**. 'X-linked pachygyria and agenesis of the corpus callosum: evidence for an X chromosome lissencephaly locus' *Ann Neurol* **36**(2): 229-233.

Beysen, D., S. De Jaegere, D. Amor, P. Bouchard, S. Christin-Maitre, M. Fellous, P. Touraine, A. W. Grix, R. Hennekam, F. Meire, N. Oyen, L. C. Wilson, D. Barel, J. Clayton-Smith, T. de Ravel, C. Decock, P. Delbeke, R. Ensenauer, F. Ebinger, G. Gillessen-Kaesbach, Y. Hendriks, V. Kimonis, R. Laframboise, P. Laissue, K. Leppig, B. P. Leroy, D. T. Miller, D. Mowat, L. Neumann, A. Plomp, N. Van Regemorter, D. Wieczorek, R. A. Veitia, A. De Paepe and E. De Baere. **2008**. 'Identification of 34 novel and 56 known FOXL2 mutations in patients with Blepharophimosis syndrome' *Hum Mutat* **29**(11): E205-219.

Bhat, S. S., R. C. Rogers, K. R. Holden and A. K. Srivastava. **2005**. 'A novel in-frame deletion in ARX is associated with lissencephaly with absent corpus callosum and hypoplastic genitalia' *Am J Med Genet A* **138**(1): 70-72.

Biancalana, V. and J. Macpherson. **2004**. 'Fragile X disease' *Methods Mol Med* **92**: 157-182.

Bienvenu, T., K. Poirier, G. Friocourt, N. Bahi, D. Beaumont, F. Fauchereau, L. Ben Jeema, R. Zemni, M. C. Vinet, F. Francis, P. Couvert, M. Gomot, C. Moraine, H. van Bokhoven, V. Kalscheuer, S. Frints, J. Gecz, K. Ohzaki, H. Chaabouni, J. P. Fryns, V. Desportes, C. Beldjord and J. Chelly. **2002**. 'ARX, a novel Prd-class-homeobox gene highly expressed in the telencephalon, is mutated in X-linked mental retardation' *Hum Mol Genet* **11**(8): 981-991.

Bonneau, D., A. Toutain, A. Laquerriere, S. Marret, P. Saugier-Veber, M. A. Barthez, S. Radi, V. Biran-Mucignat, D. Rodriguez and A. Gelot. **2002**. 'X-linked lissencephaly with absent corpus callosum and ambiguous genitalia (XLAG): clinical, magnetic resonance imaging, and neuropathological findings' *Ann Neurol* **51**(3): 340-349.

Bruyere, H., S. Lewis, S. Wood, P. J. MacLeod and S. Langlois. **1999**. 'Confirmation of linkage in X-linked infantile spasms (West syndrome) and refinement of the disease locus to Xp21.3-Xp22.1' *Clin Genet* **55**(3): 173-181.

Burlet, P., N. Frydman, N. Gigarel, V. Kerbrat, G. Tachdjian, E. Feyereisen, J. P. Bonnefont, R. Frydman, A. Munnich and J. Steffann. **2006**. 'Multiple displacement amplification improves PGD for fragile X syndrome' *Mol Hum Reprod* **12**(10): 647-652.

Caburet, S., A. Demarez, L. Moumne, M. Fellous, E. De Baere and R. A. Veitia. **2004**. 'A recurrent polyalanine expansion in the transcription factor FOXL2 induces extensive nuclear and cytoplasmic protein aggregation' *J Med Genet* **41**(12): 932-936.

Chaste, P., G. Nygren, H. Anckarsater, M. Rastam, M. Coleman, M. Leboyer, C. Gillberg and C. Betancur. **2007**. 'Mutation screening of the ARX gene in patients with autism' *Am J Med Genet B Neuropsychiatr Genet* **144B**(2): 228-230.

Cheung, S. W., C. A. Shaw, D. A. Scott, A. Patel, T. Sahoo, C. A. Bacino, A. Pursley, J. Li, R. Erickson, A. L. Gropman, D. T. Miller, M. R. Seashore, A. M. Summers, P. Stankiewicz, A. C. Chinault, J. R. Lupski, A. L. Beaudet and V. R. Sutton. **2007**. 'Microarray-based CGH detects chromosomal mosaicism not revealed by conventional cytogenetics' *Am J Med Genet A* **143A**(15): 1679-1686.

Cocquempot, O., V. Brault, C. Babinet and Y. Herault. **2009**. 'Fork stalling and template switching as a mechanism for polyalanine tract expansion affecting the DYC mutant of HOXD13, a new murine model of synpolydactyly' *Genetics* **183**(1): 23-30.

Colasante, G., P. Collombat, V. Raimondi, D. Bonanomi, C. Ferrai, M. Maira, K. Yoshikawa, A. Mansouri, F. Valtorta, J. L. Rubenstein and V. Broccoli. **2008**. 'Arx is a direct target of Dlx2 and thereby contributes to the tangential migration of GABAergic interneurons' *J Neurosci* **28**(42): 10674-10686.

Colasante, G., A. Sessa, S. Crispi, R. Calogero, A. Mansouri, P. Collombat and V. Broccoli. **2009**. 'Arx acts as a regional key selector gene in the ventral telencephalon mainly through its transcriptional repression activity' *Dev Biol* **334**(1): 59-71.

Collombat, P., J. Hecksher-Sorensen, V. Broccoli, J. Krull, I. Ponte, T. Mundiger, J. Smith, P. Gruss, P. Serup and A. Mansouri. **2005**. 'The simultaneous loss of Arx and Pax4 genes promotes a somatostatin-producing cell fate specification at the expense of the alpha- and beta-cell lineages in the mouse endocrine pancreas' *Development* **132**(13): 2969-2980.

Collombat, P., J. Hecksher-Sorensen, J. Krull, J. Berger, D. Riedel, P. L. Herrera, P. Serup and A. Mansouri. **2007**. 'Embryonic endocrine pancreas and mature beta cells acquire alpha and PP cell phenotypes upon Arx misexpression' *J Clin Invest* **117**(4): 961-970.

Collombat, P., A. Mansouri, J. Hecksher-Sorensen, P. Serup, J. Krull, G. Gradwohl and P. Gruss. **2003**. 'Opposing actions of Arx and Pax4 in endocrine pancreas development' *Genes Dev* **17**(20): 2591-2603.

Conti, V., C. Marini, S. Gana, J. Sudi, W. B. Dobyns and R. Guerrini. **2011a**. 'Corpus callosum agenesis, severe mental retardation, epilepsy, and dyskinetic quadriparesis due to a novel mutation in the homeodomain of ARX' *Am J Med Genet A* **155**(4): 892-897.

Conti, V., C. Marini, D. Mei, M. Falchi, A. R. Ferrari and R. Guerrini. **2011b**. 'Contractions in the second polyA tract of ARX are rare, non-pathogenic polymorphisms' *Am J Med Genet A* **155A**(1): 164-167.

Cossee, M., L. Faivre, C. Philippe, H. Hichri, A. de Saint-Martin, V. Laugel, N. Bahi-Buisson, J. F. Lemaitre, B. Leheup, B. Delobel, B. Demeer, K. Poirier, V. Biancalana, J. M. Pinoit, S. Julia, J. Chelly, D. Devys and J. L. Mandel. **2011**. 'ARX polyalanine expansions are highly implicated in familial cases of mental retardation with infantile epilepsy and/or hand dystonia' *Am J Med Genet A* **155A**(1): 98-105.

Cunningham, D., Q. Xiao, A. Chatterjee, K. Sulik, D. Juriloff, F. Elder, W. Harrison, G. Schuster, P. A. Overbeek and G. E. Herman. **2002**. 'exma: an X-linked insertional mutation that disrupts forebrain and eye development' *Mamm Genome* **13**(4): 179-185.

Curley, J. P. and R. Mashoodh. **2010**. 'Parent-of-origin and trans-generational germline influences on behavioral development: the interacting roles of mothers, fathers, and grandparents' *Dev Psychobiol* **52**(4): 312-330.

de Brouwer, A. P., H. G. Yntema, T. Kleefstra, D. Lugtenberg, A. R. Oudakker, B. B. de Vries, H. van Bokhoven, H. Van Esch, S. G. Frints, G. Froyen, J. P. Fryns, M. Raynaud, M. P. Moizard, N. Ronce, A. Bensalem, C. Moraine, K. Poirier, L. Castelnau, Y. Saillour, T. Bienvenu, C. Beldjord, V. des Portes, J. Chelly, G. Turner, T. Fullston, J. Gecz, A. W. Kuss, A. Tzschach, L. R. Jensen, S. Lenzner, V. M. Kalscheuer, H. H. Ropers and B. C. Hamel. **2007**. 'Mutation frequencies of X-linked mental retardation genes in families from the EuroMRX consortium' *Hum Mutat* **28**(2): 207-208. **(Appendix A.3)**

de Souza Gestinari-Duarte, R., C. B. Santos-Reboucas and M. M. Pimentel. **2006**. 'Mutational screening of ARX gene in Brazilian males with mental retardation of unknown etiology' *J Hum Genet* **51**(8): 737-740.

Demos, M. K., T. Fullston, M. W. Partington, J. Gecz and W. T. Gibson. **2009**. 'Clinical study of two brothers with a novel 33 bp duplication in the ARX gene' *Am J Med Genet A* **149A**(7): 1482-1486. **(Appendix A.8)**

Dhawan, S., S. Georgia, S. I. Tschen, G. Fan and A. Bhushan. **2011**. 'Pancreatic beta cell identity is maintained by DNA methylation-mediated repression of Arx' *Dev Cell* **20**(4): 419-429.

Dibbens, L. M., P. S. Tarpey, K. Hynes, M. A. Bayly, I. E. Scheffer, R. Smith, J. Bomar, E. Sutton, L. Vandeleur, C. Shoubridge, S. Edkins, S. J. Turner, C. Stevens, S. O'Meara, C. Tofts, S. Barthorpe, G. Buck, J. Cole, K. Halliday, D. Jones, R. Lee, M. Madison, T. Mironenko, J. Varian, S. West, S. Widaa, P. Wray, J. Teague, E. Dicks, A. Butler, A. Menzies, A. Jenkinson, R. Shepherd, J. F. Gusella, Z. Afawi, A. Mazarib, M. Y. Neufeld, S. Kivity, D. Lev, T. Lerman-Sagie, A. D. Korczyn, C. P. Derry, G. R. Sutherland, K. Friend, M. Shaw, M. Corbett, H. G. Kim, D. H. Geschwind, P. Thomas, E. Haan, S. Ryan, S. McKee, S. F. Berkovic, P. A. Futreal, M. R. Stratton, J. C. Mulley and J. Gecz. **2008**. 'X-linked

protocadherin 19 mutations cause female-limited epilepsy and cognitive impairment' *Nat Genet* **40**(6): 776-781.

Djukic, A., F. A. Lado, S. Shinnar and S. L. Moshe. **2006**. 'Are early myoclonic encephalopathy (EME) and the Ohtahara syndrome (EIEE) independent of each other?' *Epilepsy Res* **70 Suppl 1**: S68-76.

Dobyns, W. B., E. Berry-Kravis, N. J. Havernick, K. R. Holden and D. Viskochil. **1999**. 'X-linked lissencephaly with absent corpus callosum and ambiguous genitalia' *Am J Med Genet* **86**(4): 331-337.

Dolen, G. and M. F. Bear. **2009**. 'Fragile x syndrome and autism: from disease model to therapeutic targets' *J Neurodev Disord* **1**(2): 133-140.

Durbin, R. M., G. R. Abecasis, D. L. Altshuler, A. Auton, L. D. Brooks, R. A. Gibbs, M. E. Hurles and G. A. McVean. **2010**. 'A map of human genome variation from population-scale sequencing' *Nature* **467**(7319): 1061-1073.

Eichler, E. E., J. J. Holden, B. W. Popovich, A. L. Reiss, K. Snow, S. N. Thibodeau, C. S. Richards, P. A. Ward and D. L. Nelson. **1994**. 'Length of uninterrupted CGG repeats determines instability in the FMR1 gene' *Nat Genet* **8**(1): 88-94.

Eksioglu, Y. Z., A. W. Pong and M. Takeoka. **2011**. 'A novel mutation in the aristaless domain of the ARX gene leads to Ohtahara syndrome, global developmental delay, and ambiguous genitalia in males and neuropsychiatric disorders in females' *Epilepsia*.

Erickson, C. A., K. A. Stigler, L. K. Wink, J. E. Mullett, A. Kohn, D. J. Posey and C. J. McDougle. **2011**. 'A prospective open-label study of aripiprazole in fragile X syndrome' *Psychopharmacology (Berl)* **216**(1): 85-90.

Frints, S. G., G. Froyen, P. Marynen and J. P. Fryns. **2002a**. 'X-linked mental retardation: vanishing boundaries between non-specific (MRX) and syndromic (MRXS) forms' *Clin Genet* **62**(6): 423-432.

Frints, S. G., G. Froyen, P. Marynen, D. Willekens, E. Legius and J. P. Fryns. **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**(4): 427-428.

Friocourt, G., S. Kanatani, H. Tabata, M. Yozu, T. Takahashi, M. Antypa, O. Raguenes, J. Chelly, C. Ferec, K. Nakajima and J. G. Parnavelas. **2008**. 'Cell-autonomous roles of ARX in cell proliferation and neuronal migration during corticogenesis' *J Neurosci* **28**(22): 5794-5805.

Fullston, T., L. Brueton, T. Willis, S. Philip, L. MacPherson, M. Finnis, J. Gecz and J. Morton. **2010**. 'Ohtahara syndrome in a family with an ARX protein truncation mutation (c.81C>G/p.Y27X)' *Eur J Hum Genet* **18**(2): 157-162. **(Appendix A.6)**

Fullston, T., M. Finnis, A. Hackett, B. Hodgson, L. Brueton, G. Baynam, A. Norman, O. Reish, C. Shoubridge and J. Gecz. **2011**. 'Screening and cell-based assessment of mutations in the Aristaless-related homeobox (ARX) gene' *Clin Genet*. **(Appendix A.2)**

Fulp, C. T., G. Cho, E. D. Marsh, I. M. Nasrallah, P. A. Labosky and J. A. Golden. **2008**. 'Identification of Arx transcriptional targets in the developing basal forebrain' *Hum Mol Genet* **17**(23): 3740-3760.

Garcia-Arocena, D. and P. J. Hagerman. **2010**. 'Advances in understanding the molecular basis of FXTAS' *Hum Mol Genet* **19**(R1): R83-89.

Gardiner, K., Y. Herault, I. T. Lott, S. E. Antonarakis, R. H. Reeves and M. Dierssen. **2010**. 'Down syndrome: from understanding the neurobiology to therapy' *J Neurosci* **30**(45): 14943-14945.

Gecz, J., D. Cloosterman and M. Partington. **2006**. 'ARX: a gene for all seasons' *Curr Opin Genet Dev* **16**(3): 308-316.

Gecz, J., C. Shoubridge and M. Corbett. **2009**. 'The genetic landscape of intellectual disability arising from chromosome X' *Trends Genet* **25**(7): 308-316.

Gestinari-Duarte Rde, S., C. B. Santos-Reboucas, R. T. Boy and M. M. Pimentel. **2006**. 'ARX mutation c.428-451dup (24bp) in a Brazilian family with X-linked mental retardation' *Eur J Med Genet* **49**(3): 269-275.

Giordano, L., S. Sartori, S. Russo, P. Accorsi, J. Galli, A. Tiberti, E. Bettella, M. Marchi, A. Vignoli, F. Darra, A. Murgia and B. D. Bernardina. **2010**. 'Familial Ohtahara syndrome due to a novel ARX gene mutation' *Am J Med Genet A* **152A**(12): 3133-3137.

Goodman, F. R., C. Bacchelli, A. F. Brady, L. A. Brueton, J. P. Fryns, D. P. Mortlock, J. W. Innis, L. B. Holmes, A. E. Donnenfeld, M. Feingold, F. A. Beemer, R. C. Hennekam and P. J. Scambler. **2000**. 'Novel HOXA13 mutations and the phenotypic spectrum of hand-foot-genital syndrome' *Am J Hum Genet* **67**(1): 197-202.

Gradwohl, G., A. Dierich, M. LeMeur and F. Guillemot. **2000**. 'neurogenin3 is required for the development of the four endocrine cell lineages of the pancreas' *Proc Natl Acad Sci U S A* **97**(4): 1607-1611.

Gronskov, K., H. Hjalgrim, I. M. Nielsen and K. Brondum-Nielsen. **2004**. 'Screening of the ARX gene in 682 retarded males' *Eur J Hum Genet* **12**(9): 701-705.

Gu, W., F. Zhang and J. R. Lupski. **2008**. 'Mechanisms for human genomic rearrangements' *Pathogenetics* **1**(1): 4.

Guerrini, R., F. Moro, M. Kato, A. J. Barkovich, T. Shiihara, M. A. McShane, J. Hurst, M. Loi, J. Tohyama, V. Norci, K. Hayasaka, U. J. Kang, S. Das and W. B. Dobyns. **2007**. 'Expansion of the first PolyA tract of ARX causes infantile spasms and status dystonicus' *Neurology* **69**(5): 427-433.

Hagerman, R. J., M. Leehey, W. Heinrichs, F. Tassone, R. Wilson, J. Hills, J. Grigsby, B. Gage and P. J. Hagerman. **2001**. 'Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X' *Neurology* **57**(1): 127-130.

Hahn, A., C. Gross, G. Uyanik, U. Hehr, M. Hugens-Penzel, G. Alzen and B. A. Neubauer. **2004**. 'X-linked lissencephaly with abnormal genitalia associated with renal phosphate wasting' *Neuropediatrics* **35**(3): 202-205.

Hamdan, F. F., J. Gauthier, Y. Araki, D. T. Lin, Y. Yoshizawa, K. Higashi, A. R. Park, D. Spiegelman, S. Dobrzeniecka, A. Piton, H. Tomitori, H. Daoud, C. Massicotte, E. Henrion, O. Diallo, M. Shekarabi, C. Marineau, M. Shevell, B. Maranda, G. Mitchell, A. Nadeau, G. D'Anjou, M. Vanasse, M. Srour, R. G. Lafreniere, P. Drapeau, J. C. Lacaille, E. Kim, J. R. Lee, K. Igarashi, R. L. Huganir, G. A. Rouleau and J. L. Michaud. **2011**. 'Excess of de novo deleterious mutations in genes associated with glutamatergic systems in nonsyndromic intellectual disability' *Am J Hum Genet* **88**(3): 306-316.

Hamdan, F. F., J. Gauthier, D. Spiegelman, A. Noreau, Y. Yang, S. Pellerin, S. Dobrzeniecka, M. Cote, E. Perreau-Linck, L. Carmant, G. D'Anjou, E. Fombonne, A. M. Addington, J. L. Rapoport, L. E. Delisi, M. O. Krebs, F. Mouaffak, R. Joober, L. Mottron, P. Drapeau, C.

Marineau, R. G. Lafreniere, J. C. Lacaille, G. A. Rouleau and J. L. Michaud. **2009**. 'Mutations in SYNGAP1 in autosomal nonsyndromic mental retardation' *N Engl J Med* **360**(6): 599-605.

Handa, V., T. Saha and K. Usdin. **2003**. 'The fragile X syndrome repeats form RNA hairpins that do not activate the interferon-inducible protein kinase, PKR, but are cut by Dicer' *Nucleic Acids Res* **31**(21): 6243-6248.

Hartmann, H., G. Uyanik, C. Gross, U. Hehr, T. Lucke, M. Arslan-Kirchner, B. Antosch, A. M. Das and J. Winkler. **2004**. 'Agenesis of the corpus callosum, abnormal genitalia and intractable epilepsy due to a novel familial mutation in the Aristaless-related homeobox gene' *Neuropediatrics* **35**(3): 157-160.

Hashem, V., J. N. Galloway, M. Mori, R. Willemsen, B. A. Oostra, R. Paylor and D. L. Nelson. **2009**. 'Ectopic expression of CGG containing mRNA is neurotoxic in mammals' *Hum Mol Genet* **18**(13): 2443-2451.

Hastings, P. J., J. R. Lupski, S. M. Rosenberg and G. Ira. **2009**. 'Mechanisms of change in gene copy number' *Nat Rev Genet* **10**(8): 551-564.

Higgins, J. J., J. Pucilowska, R. Q. Lombardi and J. P. Rooney. **2004**. 'A mutation in a novel ATP-dependent Lon protease gene in a kindred with mild mental retardation' *Neurology* **63**(10): 1927-1931.

Higgins, J. J. and H. Topaloglu. **2005**. 'X-linked oligophrenic vermian dysgenesis: syndromic vs non-syndromic X-linked mental retardation?' *Neurology* **65**(9): 1346-1347.

Hoffman, B. G., B. Zavaglia, M. Beach and C. D. Helgason. **2008**. 'Expression of Groucho/TLE proteins during pancreas development' *BMC Dev Biol* **8**: 81.

Ionita-Laza, I., A. J. Rogers, C. Lange, B. A. Raby and C. Lee. **2009**. 'Genetic association analysis of copy-number variation (CNV) in human disease pathogenesis' *Genomics* **93**(1): 22-26.

Jacquemont, S., R. J. Hagerman, M. Leehey, J. Grigsby, L. Zhang, J. A. Brunberg, C. Greco, V. Des Portes, T. Jardini, R. Levine, E. Berry-Kravis, W. T. Brown, S. Schaeffer, J. Kissel, F. Tassone and P. J. Hagerman. **2003**. 'Fragile X premutation tremor/ataxia syndrome: molecular, clinical, and neuroimaging correlates' *Am J Hum Genet* **72**(4): 869-878.

Jagla, M., P. Kruczek and P. Kwinta. **2008**. 'Association between X-linked lissencephaly with ambiguous genitalia syndrome and lenticulostriate vasculopathy in neonate' *J Clin Ultrasound* **36**(6): 387-390.

Jemaa, L. B., V. des Portes, R. Zemni, R. Mrad, F. Maazoul, C. Beldjord, H. Chaabouni and J. Chelly. **1999**. 'Refined 2.7 centimorgan locus in Xp21.3-22.1 for a nonspecific X-linked mental retardation gene (MRX54)' *Am J Med Genet* **85**(3): 276-282.

Jenny, M., C. Uhl, C. Roche, I. Duluc, V. Guillermin, F. Guillemot, J. Jensen, M. Kedinger and G. Gradwohl. **2002**. 'Neurogenin3 is differentially required for endocrine cell fate specification in the intestinal and gastric epithelium' *EMBO J* **21**(23): 6338-6347.

Jin, P., R. Duan, A. Qurashi, Y. Qin, D. Tian, T. C. Rosser, H. Liu, Y. Feng and S. T. Warren. **2007**. 'Pur alpha binds to rCGG repeats and modulates repeat-mediated neurodegeneration in a Drosophila model of fragile X tremor/ataxia syndrome' *Neuron* **55**(4): 556-564.

Jin, P., D. C. Zarnescu, F. Zhang, C. E. Pearson, J. C. Lucchesi, K. Moses and S. T. Warren. **2003**. 'RNA-mediated neurodegeneration caused by the fragile X premutation rCGG repeats in Drosophila' *Neuron* **39**(5): 739-747.

Jordanova A, Kalaydjieva L, Savov A, Claustres M, Schwarz M, Estivill X, Angelicheva D, Haworth A, Casals T, Kremensky I. **1997.** ' SSCP analysis: a blind sensitivity trial' *Hum Mutat* **10**(1): 65-70.

Kato, M., S. Das, K. Petras, K. Kitamura, K. Morohashi, D. N. Abuelo, M. Barr, D. Bonneau, A. F. Brady, N. J. Carpenter, K. L. Cipero, F. Frisone, T. Fukuda, R. Guerrini, E. Iida, M. Itoh, A. F. Lewanda, Y. Nanba, A. Oka, V. K. Proud, P. Saugier-Veber, S. L. Schelley, A. Selicorni, R. Shaner, M. Silengo, F. Stewart, N. Sugiyama, J. Toyama, A. Toutain, A. L. Vargas, M. Yanazawa, E. H. Zackai and W. B. Dobyns. **2004**. 'Mutations of ARX are associated with striking pleiotropy and consistent genotype-phenotype correlation' *Hum Mutat* **23**(2): 147-159.

Kato, M., S. Das, K. Petras, Y. Sawaishi and W. B. Dobyns. **2003**. 'Polyalanine expansion of ARX associated with cryptogenic West syndrome' *Neurology* **61**(2): 267-276.

Kato, M. and W. B. Dobyns. **2005**. 'X-linked lissencephaly with abnormal genitalia as a tangential migration disorder causing intractable epilepsy: proposal for a new term, "interneuronopathy"' *J Child Neurol* **20**(4): 392-397.

Kato, M., N. Koyama, M. Ohta, K. Miura and K. Hayasaka. **2010**. 'Frameshift mutations of the ARX gene in familial Ohtahara syndrome' *Epilepsia* **51**(9): 1679-1684.

Kato, M., S. Saitoh, A. Kamei, H. Shiraishi, Y. Ueda, M. Akasaka, J. Tohyama, N. Akasaka and K. Hayasaka. **2007**. 'A longer polyalanine expansion mutation in the ARX gene causes early infantile epileptic encephalopathy with suppression-burst pattern (Ohtahara syndrome)' *Am J Hum Genet* **81**(2): 361-366.

Kitamura, K., Y. Itou, M. Yanazawa, M. Ohsawa, R. Suzuki-Migishima, Y. Umeki, H. Hohjoh, Y. Yanagawa, T. Shinba, M. Itoh, K. Nakamura and Y. Goto. **2009**. 'Three human ARX mutations cause the lissencephaly-like and mental retardation with epilepsy-like pleiotropic phenotypes in mice' *Hum Mol Genet* **18**(19): 3708-3724.

Kitamura, K., M. Yanazawa, N. Sugiyama, H. Miura, A. Iizuka-Kogo, M. Kusaka, K. Omichi, R. Suzuki, Y. Kato-Fukui, K. Kamiirisa, M. Matsuo, S. Kamijo, M. Kasahara, H. Yoshioka, T. Ogata, T. Fukuda, I. Kondo, M. Kato, W. B. Dobyns, M. Yokoyama and K. Morohashi. **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**(3): 359-369.

Koolen, D. A., R. Pfundt, N. de Leeuw, J. Y. Hehir-Kwa, W. M. Nillesen, I. Neefs, I. Scheltinga, E. Sistermans, D. Smeets, H. G. Brunner, A. G. van Kessel, J. A. Veltman and B. B. de Vries. **2009**. 'Genomic microarrays in mental retardation: a practical workflow for diagnostic applications' *Hum Mutat* **30**(3): 283-292.

Korbel, J. O., T. Tirosh-Wagner, A. E. Urban, X. N. Chen, M. Kasowski, L. Dai, F. Grubert, C. Erdman, M. C. Gao, K. Lange, E. M. Sobel, G. M. Barlow, A. S. Aylsworth, N. J. Carpenter, R. D. Clark, M. Y. Cohen, E. Doran, T. Falik-Zaccai, S. O. Lewin, I. T. Lott, B. C. McGillivray, J. B. Moeschler, M. J. Pettenati, S. M. Pueschel, K. W. Rao, L. G. Shaffer, M. Shohat, A. J. Van Riper, D. Warburton, S. Weissman, M. B. Gerstein, M. Snyder and J. R. Korenberg. **2009**. 'The genetic architecture of Down syndrome phenotypes revealed by high-

resolution analysis of human segmental trisomies' *Proc Natl Acad Sci U S A* **106**(29): 12031-12036.

Kozak, M. **2002**. 'Pushing the limits of the scanning mechanism for initiation of translation' *Gene* **299**(1-2): 1-34.

Krueger, D. D. and M. F. Bear. **2011**. 'Toward fulfilling the promise of molecular medicine in fragile X syndrome' *Annu Rev Med* **62**: 411-429.

Kunkel, T. A. **1985**. 'The mutational specificity of DNA polymerase-beta during in vitro DNA synthesis. Production of frameshift, base substitution, and deletion mutations' *J Biol Chem* **260**(9): 5787-5796.

Kuss, A. W., M. Garshasbi, K. Kahrizi, A. Tzschach, F. Behjati, H. Darvish, L. Abbasi-Moheb, L. Puettmann, A. Zecha, R. Weissmann, H. Hu, M. Mohseni, S. S. Abedini, A. Rajab, C. Hertzberg, D. Wieczorek, R. Ullmann, S. Ghasemi-Firouzabadi, S. Banihashemi, S. Arzhangi, V. Hadavi, G. Bahrami-Monajemi, M. Kasiri, M. Falah, P. Nikuei, A. Dehghan, M. Sobhani, P. Jamali, H. H. Ropers and H. Najmabadi. **2011**. 'Autosomal recessive mental retardation: homozygosity mapping identifies 27 single linkage intervals, at least 14 novel loci and several mutation hotspots' *Hum Genet* **129**(2): 141-148.

Laperuta, C., L. Spizzichino, P. D'Adamo, J. Monfregola, A. Maiorino, A. D'Eustacchio, V. Ventruto, G. Neri, M. D'Urso, P. Chiurazzi, M. V. Ursini and M. G. Miano. **2007**. 'MRX87 family with Aristaless X dup24bp mutation and implication for polyAlanine expansions' *BMC Med Genet* **8**: 25.

Lee, J. A., C. M. Carvalho and J. R. Lupski. **2007**. 'A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders' *Cell* **131**(7): 1235-1247.

Lee, J. A. and J. R. Lupski. **2006**. 'Genomic rearrangements and gene copy-number alterations as a cause of nervous system disorders' *Neuron* **52**(1): 103-121.

Lee, S. J., Y. Matsuura, S. M. Liu and M. Stewart. **2005**. 'Structural basis for nuclear import complex dissociation by RanGTP' *Nature* **435**(7042): 693-696.

Leehey, M. A. **2009**. 'Fragile X-associated tremor/ataxia syndrome: clinical phenotype, diagnosis, and treatment' *J Investig Med* **57**(8): 830-836.

Leonard, W. R., J. J. Snodgrass and M. L. Robertson. **2007**. 'Effects of brain evolution on human nutrition and metabolism' *Annu Rev Nutr* **27**: 311-327.

Lin, W., W. Ye, L. Cai, X. Meng, G. Ke, C. Huang, Z. Peng, Y. Yu, J. A. Golden, A. M. Tartakoff and T. Tao. **2009**. 'The roles of multiple importins for nuclear import of murine aristaless-related homeobox protein' *J Biol Chem* **284**(30): 20428-20439.

Lupski, J. R. **1998**. 'Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits' *Trends Genet* **14**(10): 417-422.

Lyle, R., F. Bena, S. Gagos, C. Gehrig, G. Lopez, A. Schinzel, J. Lespinasse, A. Bottani, S. Dahoun, L. Taine, M. Doco-Fenzy, P. Cornillet-Lefebvre, A. Pelet, S. Lyonnet, A. Toutain, L. Colleaux, J. Horst, I. Kennerknecht, N. Wakamatsu, M. Descartes, J. C. Franklin, L. Florentin-Arar, S. Kitsiou, E. Ait Yahya-Graison, M. Costantine, P. M. Sinet, J. M. Delabar and S. E. Antonarakis. **2009**. 'Genotype-phenotype correlations in Down syndrome identified by array CGH in 30 cases of partial trisomy and partial monosomy chromosome 21' *Eur J Hum Genet* **17**(4): 454-466.

Mackintosh, N. J. **1998**. IQ and Human Intelligence, Oxford University Press.

Majnemer, A. and M. I. Shevell. **1995**. 'Diagnostic yield of the neurologic assessment of the developmentally delayed child' *J Pediatr* **127**(2): 193-199.

Malcov, M., T. Naiman, D. B. Yosef, A. Carmon, N. Mey-Raz, A. Amit, I. Vagman and Y. Yaron. **2007**. 'Preimplantation genetic diagnosis for fragile X syndrome using multiplex nested PCR' *Reprod Biomed Online* **14**(4): 515-521.

Mandel, J. L. and J. Chelly. **2004**. 'Monogenic X-linked mental retardation: is it as frequent as currently estimated? The paradox of the ARX (Aristaless X) mutations' *Eur J Hum Genet* **12**(9): 689-693.

Maquat, L. E. **2004**. 'Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics' *Nat Rev Mol Cell Biol* **5**(2): 89-99.

Marsh, E., C. Fulp, E. Gomez, I. Nasrallah, J. Minarcik, J. Sudi, S. L. Christian, G. Mancini, P. Labosky, W. Dobyns, A. Brooks-Kayal and J. A. Golden. **2009**. 'Targeted loss of Arx results in a developmental epilepsy mouse model and recapitulates the human phenotype in heterozygous females' *Brain* **132**(Pt 6): 1563-1576.

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

Matsuzaki, H., P. H. Wang, J. Hu, R. Rava and G. K. Fu. **2009**. 'High resolution discovery and confirmation of copy number variants in 90 Yoruba Nigerians' *Genome Biol* **10**(11): R125.

Mattick, J. S. and M. F. Mehler. **2008**. 'RNA editing, DNA recoding and the evolution of human cognition' *Trends Neurosci* **31**(5): 227-233.

McKenzie, O., I. Ponte, M. Mangelsdorf, M. Finnis, G. Colasante, C. Shoubridge, S. Stifani, J. Gecz and V. Broccoli. **2007**. 'Aristaless-related homeobox gene, the gene responsible for West syndrome and related disorders, is a Groucho/transducin-like enhancer of split dependent transcriptional repressor' *Neuroscience* **146**(1): 236-247.

Melko, M. and B. Bardoni. **2010**. 'The role of G-quadruplex in RNA metabolism: involvement of FMRP and FMR2P' *Biochimie* **92**(8): 919-926.

Michaud, E. J., M. J. van Vugt, S. J. Bultman, H. O. Sweet, M. T. Davisson and R. P. Woychik. **1994**. 'Differential expression of a new dominant agouti allele (Aiapy) is correlated with methylation state and is influenced by parental lineage' *Genes Dev* **8**(12): 1463-1472.

Miller, D. T., M. P. Adam, S. Aradhya, L. G. Biesecker, A. R. Brothman, N. P. Carter, D. M. Church, J. A. Crolla, E. E. Eichler, C. J. Epstein, W. A. Faucett, L. Feuk, J. M. Friedman, A. Hamosh, L. Jackson, E. B. Kaminsky, K. Kok, I. D. Krantz, R. M. Kuhn, C. Lee, J. M. Ostell, C. Rosenberg, S. W. Scherer, N. B. Spinner, D. J. Stavropoulos, J. H. Tepperberg, E. C. Thorland, J. R. Vermeesch, D. J. Waggoner, M. S. Watson, C. L. Martin and D. H. Ledbetter. **2010**. 'Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies' *Am J Hum Genet* **86**(5): 749-764.

Miller, G. **2010**. 'Epigenetics. A role for epigenetics in cognition' *Science* **329**(5987): 27.

Milosavljevic, A. **2011**. 'Emerging patterns of epigenomic variation' *Trends Genet* **27**(6): 242-250.

Mingot, J. M., S. Kostka, R. Kraft, E. Hartmann and D. Gorlich. **2001**. 'Importin 13: a novel mediator of nuclear import and export' *EMBO J* **20**(14): 3685-3694.

Miyata, R., M. Hayashi, K. Miyai, T. Akashi, M. Kato and J. Kohyama. **2009**. 'Analysis of the hypothalamus in a case of X-linked lissencephaly with abnormal genitalia (XLAG)' *Brain Dev* **31**(6): 456-460.

Molinari, F., M. Rio, V. Meskenaite, F. Encha-Razavi, J. Auge, D. Bacq, S. Briault, M. Vekemans, A. Munnich, T. Attie-Bitach, P. Sonderegger and L. Colleaux. **2002**. 'Truncating neurotrypsin mutation in autosomal recessive nonsyndromic mental retardation' *Science* **298**(5599): 1779-1781.

Mulley, J. C., B. Kerr, R. Stevenson and H. Lubs. **1992**. 'Nomenclature guidelines for X-linked mental retardation' *Am J Med Genet* **43**(1-2): 383-391.

Mullis, K. B. **1990**. 'Target amplification for DNA analysis by the polymerase chain reaction' *Ann Biol Clin (Paris)* **48**(8): 579-582.

Muragaki, Y., S. Mundlos, J. Upton and B. R. Olsen. **1996**. 'Altered growth and branching patterns in synpolydactyly caused by mutations in HOXD13' *Science* **272**(5261): 548-551.

Najmabadi, H., H. Hu, M. Garshasbi, T. Zemojtel, S. S. Abedini, W. Chen, M. Hosseini, F. Behjati, S. Haas, P. Jamali, A. Zecha, M. Mohseni, L. Puttmann, L. N. Vahid, C. Jensen, L. A. Moheb, M. Bienek, F. Larti, I. Mueller, R. Weissmann, H. Darvish, K. Wrogemann, V. Hadavi, B. Lipkowitz, S. Esmaeeli-Nieh, D. Wieczorek, R. Kariminejad, S. G. Firouzabadi, M. Cohen, Z. Fattahi, I. Rost, F. Mojahedi, C. Hertzberg, A. Dehghan, A. Rajab, M. J. Banavandi, J. Hoffer, M. Falah, L. Musante, V. Kalscheuer, R. Ullmann, A. W. Kuss, A. Tzschach, K. Kahrizi and H. H. Ropers. **2011**. 'Deep sequencing reveals 50 novel genes for recessive cognitive disorders' *Nature* **478**(7367): 57-63.

Najmabadi, H., M. M. Motazacker, M. Garshasbi, K. Kahrizi, A. Tzschach, W. Chen, F. Behjati, V. Hadavi, S. E. Nieh, S. S. Abedini, R. Vazifehmand, S. G. Firouzabadi, P. Jamali, M. Falah, S. M. Seifati, A. Gruters, S. Lenzner, L. R. Jensen, F. Ruschendorf, A. W. Kuss and H. H. Ropers. **2007**. 'Homozygosity mapping in consanguineous families reveals extreme heterogeneity of non-syndromic autosomal recessive mental retardation and identifies 8 novel gene loci' *Hum Genet* **121**(1): 43-48.

Nanba, Y., A. Oka and K. Ohno. **2007**. '[Severe diarrhea associated with X-linked lissencephaly with absent corpus callosum and abnormal genitalia: a case report of successful treatment with the somatostatin analogue octreotide]' *No To Hattatsu* **39**(5): 379-382.

Nasrallah, I. M., J. C. Minarcik and J. A. Golden. **2004**. 'A polyalanine tract expansion in Arx forms intranuclear inclusions and results in increased cell death' *J Cell Biol* **167**(3): 411-416.

Nawara, M., K. Szczaluba, K. Poirier, K. Chrzanowska, J. Pilch, J. Bal, J. Chelly and T. Mazurczak. **2006**. 'The ARX mutations: a frequent cause of X-linked mental retardation' *Am J Med Genet A* **140**(7): 727-732.

Nussbaum, R. L. and D. H. Ledbetter. **1986**. 'Fragile X syndrome: a unique mutation in man' *Annu Rev Genet* **20**: 109-145.

Oberle, I., F. Rousseau, D. Heitz, C. Kretz, D. Devys, A. Hanauer, J. Boue, M. F. Bertheas and J. L. Mandel. **1991**. 'Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome' *Science* **252**(5010): 1097-1102.

Oefner, P. J. **1995**. 'Surface-charge reversed capillary zone electrophoresis of inorganic and organic anions' *Electrophoresis* **16**(1): 46-56.

Ogata, T., N. Matsuo, N. Hiraoka and J. I. Hata. **2000**. 'X-linked lissencephaly with ambiguous genitalia: delineation of further case' *Am J Med Genet* **94**(2): 174-176.

Ohtahara, S. and Y. Yamatogi. **2006**. 'Ohtahara syndrome: with special reference to its developmental aspects for differentiating from early myoclonic encephalopathy' *Epilepsy Res* **70 Suppl 1**: S58-67.

Okazaki, S., M. Ohsawa, I. Kuki, H. Kawawaki, T. Koriyama, S. Ri, H. Ichiba, E. Hai, T. Inoue, H. Nakamura, Y. Goto, K. Tomiwa, T. Yamano, K. Kitamura and M. Itoh. **2008**. 'Aristaless-related homeobox gene disruption leads to abnormal distribution of GABAergic interneurons in human neocortex: evidence based on a case of X-linked lissencephaly with abnormal genitalia (XLAG)' *Acta Neuropathol* **116**(4): 453-462.

Orita, M., H. Iwahana, H. Kanazawa, K. Hayashi and T. Sekiya. **1989a**. 'Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms' *Proc Natl Acad Sci U S A* **86**(8): 2766-2770.

Orita, M., Y. Suzuki, T. Sekiya and K. Hayashi. **1989b**. 'Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction' *Genomics* **5**(4): 874-879.

Pandey, U. B., S. R. Phadke and B. Mittal. **2004**. 'Molecular diagnosis and genetic counseling for fragile X mental retardation' *Neurol India* **52**(1): 36-42.

Partington, M. W., J. C. Mulley, G. R. Sutherland, A. Thode and G. Turner. **1988**. 'A family with the Coffin Lowry syndrome revisited: localization of CLS to Xp21-pter' *Am J Med Genet* **30**(1-2): 509-521.

Partington, M. W., G. Turner, J. Boyle and J. Gecz. **2004**. 'Three new families with X-linked mental retardation caused by the 428-451dup(24bp) mutation in ARX' *Clin Genet* **66**(1): 39-45.

Paulsen, M., C. Lund, Z. Akram, J. R. Winther, N. Horn and L. B. Moller. **2006**. 'Evidence that translation reinitiation leads to a partially functional Menkes protein containing two copper-binding sites' *Am J Hum Genet* **79**(2): 214-229.

Ploski, J. E., M. K. Shamsher and A. Radu. **2004**. 'Paired-type homeodomain transcription factors are imported into the nucleus by karyopherin 13' *Mol Cell Biol* **24**(11): 4824-4834.

Poirier, K., J. Abriol, I. Souville, C. Laroche-Raynaud, C. Beldjord, B. Gilbert, J. Chelly and T. Bienvenu. **2005**. 'Maternal mosaicism for mutations in the ARX gene in a family with X linked mental retardation' *Hum Genet* **118**(1): 45-48.

Poirier, K., M. Eisermann, I. Caubel, A. Kaminska, S. Peudonnier, N. Boddaert, Y. Saillour, O. Dulac, I. Souville, C. Beldjord, K. Lascelles, P. Plouin, J. Chelly and N. Bahi-Buisson. **2008**. 'Combination of infantile spasms, non-epileptic seizures and complex movement disorder: a new case of ARX-related epilepsy' *Epilepsy Res* **80**(2-3): 224-228.

Poirier, K., D. Lacombe, B. Gilbert-Dussardier, M. Raynaud, V. Desportes, A. P. de Brouwer, C. Moraine, J. P. Fryns, H. H. Ropers, C. Beldjord, J. Chelly and T. Bienvenu. **2006**. 'Screening of ARX in mental retardation families: Consequences for the strategy of molecular diagnosis' *Neurogenetics* **7**(1): 39-46.

Poitras, L., M. Yu, C. Lesage-Pelletier, R. B. Macdonald, J. P. Gagne, G. Hatch, I. Kelly, S. P. Hamilton, J. L. Rubenstein, G. G. Poirier and M. Ekker. **2010**. 'An SNP in an ultraconserved regulatory element affects Dlx5/Dlx6 regulation in the forebrain' *Development* **137**(18): 3089-3097.

Pontoglio, M., J. Barra, M. Hadchouel, A. Doyen, C. Kress, J. P. Bach, C. Babinet and M. Yaniv. **1996**. 'Hepatocyte nuclear factor 1 inactivation results in hepatic dysfunction, phenylketonuria, and renal Fanconi syndrome' *Cell* **84**(4): 575-585.

Price, M. G., J. W. Yoo, D. L. Burgess, F. Deng, R. A. Hrachovy, J. D. Frost, Jr. and J. L. Noebels. **2009**. 'A triplet repeat expansion genetic mouse model of infantile spasms syndrome, Arx(GCG)10+7, with interneuronopathy, spasms in infancy, persistent seizures, and adult cognitive and behavioral impairment' *J Neurosci* **29**(27): 8752-8763.

Primerano, B., F. Tassone, R. J. Hagerman, P. Hagerman, F. Amaldi and C. Bagni. **2002**. 'Reduced FMR1 mRNA translation efficiency in fragile X patients with premutations' *RNA* **8**(12): 1482-1488.

Proud, V. K., C. Levine and N. J. Carpenter. **1992**. 'New X-linked syndrome with seizures, acquired micrencephaly, and agenesis of the corpus callosum' *Am J Med Genet* **43**(1-2): 458-466.

Qiao, Y., C. Harvard, C. Tyson, X. Liu, C. Fawcett, P. Pavlidis, J. J. Holden, M. E. Lewis and E. Rajcan-Separovic. **2010**. 'Outcome of array CGH analysis for 255 subjects with intellectual disability and search for candidate genes using bioinformatics' *Hum Genet* **128**(2): 179-194.

Qin, M., J. Kang, T. V. Burlin, C. Jiang and C. B. Smith. **2005**. 'Postadolescent changes in regional cerebral protein synthesis: an in vivo study in the FMR1 null mouse' *J Neurosci* **25**(20): 5087-5095.

Quille, M. L., S. Carat, S. Quemener-Redon, E. Hirchaud, D. Baron, C. Benech, J. Guihot, M. Placet, O. Mignen, C. Ferec, R. Houlgatte and G. Friocourt. **2011**. 'High-throughput analysis of promoter occupancy reveals new targets for Arx, a gene mutated in mental retardation and interneuronopathies' *PLoS One* **6**(9): e25181.

Raymond, F. **2005**. 'X-linked Mental Retardation: a Clinical guide' *J Med Genet* **In press**.

Reish, O., T. Fullston, M. Regev, E. Heyman and J. Gecz. **2009**. 'A novel de novo 27 bp duplication of the ARX gene, resulting from postzygotic mosaicism and leading to three severely affected males in two generations' *Am J Med Genet A* **149A**(8): 1655-1660. **(Appendix A.7)**

Ropers, H. H. **2008**. 'Genetics of intellectual disability' *Curr Opin Genet Dev* **18**(3): 241-250.

Ropers, H. H. **2010**. 'Genetics of early onset cognitive impairment' *Annu Rev Genomics Hum Genet* **11**: 161-187.

Ropers, H. H. and B. C. Hamel. **2005**. 'X-linked mental retardation' *Nat Rev Genet* **6**(1): 46-57.

Rousseau, F., Y. Labelle, J. Bussieres and C. Lindsay. **2011**. 'The fragile x mental retardation syndrome 20 years after the FMR1 gene discovery: an expanding universe of knowledge' *Clin Biochem Rev* **32**(3): 135-162.

Rujirabanjerd, S., K. Tongsippunyoo, T. Sripo and P. Limprasert. **2007**. 'Mutation screening of the Aristaless-related homeobox (ARX) gene in Thai pediatric patients with delayed development: first report from Thailand' *Eur J Med Genet* **50**(5): 346-354.

Sambrook, J. and D. W. Russell. **2001**. Molecular cloning: a laboratory manual (3rd Edition), Cold Spring Harbor Laboratory Press.

Sanger, F., S. Nicklen and A. R. Coulson. **1977**. 'DNA sequencing with chain-terminating inhibitors' *Proc Natl Acad Sci U S A* **74**(12): 5463-5467.

Scheffer, I. E., R. H. Wallace, F. L. Phillips, P. Hewson, K. Reardon, G. Parasivam, P. Stromme, S. F. Berkovic, J. Gecz and J. C. Mulley. **2002**. 'X-linked myoclonic epilepsy with spasticity and intellectual disability: mutation in the homeobox gene ARX' *Neurology* **59**(3): 348-356.

Scherer, S. W., C. Lee, E. Birney, D. M. Altshuler, E. E. Eichler, N. P. Carter, M. E. Hurles and L. Feuk. **2007**. 'Challenges and standards in integrating surveys of structural variation' *Nat Genet* **39**(7 Suppl): S7-15.

Schwitzgebel, V. M., D. W. Scheel, J. R. Conners, J. Kalamaras, J. E. Lee, D. J. Anderson, L. Sussel, J. D. Johnson and M. S. German. **2000**. 'Expression of neurogenin3 reveals an islet cell precursor population in the pancreas' *Development* **127**(16): 3533-3542.

Screaton, G. R., J. F. Caceres, A. Mayeda, M. V. Bell, M. Plebanski, D. G. Jackson, J. I. Bell and A. R. Krainer. **1995**. 'Identification and characterization of three members of the human SR family of pre-mRNA splicing factors' *Embo J* **14**(17): 4336-4349.

Sharp, A. J., H. C. Mefford, K. Li, C. Baker, C. Skinner, R. E. Stevenson, R. J. Schroer, F. Novara, M. De Gregori, R. Ciccone, A. Broomer, I. Casuga, Y. Wang, C. Xiao, C. Barbacioru, G. Gimelli, B. D. Bernardina, C. Torniero, R. Giorda, R. Regan, V. Murday, S. Mansour, M. Fichera, L. Castiglia, P. Failla, M. Ventura, Z. Jiang, G. M. Cooper, S. J. Knight, C. Romano, O. Zuffardi, C. Chen, C. E. Schwartz and E. E. Eichler. **2008**. 'A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures' *Nat Genet* **40**(3): 322-328.

Shevell, M., S. Ashwal, D. Donley, J. Flint, M. Gingold, D. Hirtz, A. Majnemer, M. Noetzel and R. D. Sheth. **2003**. 'Practice parameter: evaluation of the child with global developmental delay: report of the Quality Standards Subcommittee of the American Academy of Neurology and The Practice Committee of the Child Neurology Society' *Neurology* **60**(3): 367-380.

Shinozaki, Y., M. Osawa, H. Sakuma, H. Komaki, E. Nakagawa, K. Sugai, M. Sasaki and Y. Goto. **2009**. 'Expansion of the first polyalanine tract of the ARX gene in a boy presenting with generalized dystonia in the absence of infantile spasms' *Brain Dev* **31**(6): 469-472.

Shoubridge, C., D. Cloosterman, E. Parkinson-Lawerence, D. Brooks and J. Gecz. **2007**. 'Molecular pathology of expanded polyalanine tract mutations in the Aristaless-related homeobox gene' *Genomics* **90**(1): 59-71.

Shoubridge, C., T. Fullston and J. Gecz. **2010a**. 'ARX spectrum disorders: making inroads into the molecular pathology' *Hum Mutat* **31**(8): 889-900. **(Appendix A.1)**

Shoubridge, C. and J. Gecz **2011**. Polyalanine Tract Disorders and Neurocognitive Phenotypes. Madame Curie Bioscience Database. A. J. Hannan. Austin, Texas, USA, Landes Bioscience and Springer Science+Business Media. **Cell Biology:** 19.

Shoubridge, C., M. H. Tan, T. Fullston, D. Cloosterman, D. Coman, G. McGillivray, G. M. Mancini, T. Kleefstra and J. Gecz. **2010b**. 'Mutations in the nuclear localization sequence of the Aristaless related homeobox; sequestration of mutant ARX with IPO13 disrupts normal subcellular distribution of the transcription factor and retards cell division' *Pathogenetics* **3**: 1. **(Appendix A.9)**

Shoubridge, C., P. S. Tarpey, F. Abidi, S. L. Ramsden, S. Rujirabanjerd, J. A. Murphy, J. Boyle, M. Shaw, A. Gardner, A. Proos, H. Puusepp, F. L. Raymond, C. E. Schwartz, R. E. Stevenson, G. Turner, M. Field, R. S. Walikonis, R. J. Harvey, A. Hackett, P. A. Futreal, M. R. Stratton and J. Gecz. **2010c**. 'Mutations in the guanine nucleotide exchange factor gene IQSEC2 cause nonsyndromic intellectual disability' *Nat Genet* **42**(6): 486-488.

Silva, A. L., P. Ribeiro, A. Inacio, S. A. Liebhaber and L. Romao. **2008**. 'Proximity of the poly(A)-binding protein to a premature termination codon inhibits mammalian nonsense-mediated mRNA decay' *RNA* **14**(3): 563-576.

Song, F. J., P. Barton, V. Sleightholme, G. L. Yao and A. Fry-Smith. **2003**. 'Screening for fragile X syndrome: a literature review and modelling study' *Health Technol Assess* **7**(16): 1-106.

Spinosa, M. J., P. B. Liberalesso, S. C. Vieira, A. S. Olmos and A. Lohr, Jr. **2006**. 'Lissencephaly, abnormal genitalia and refractory epilepsy: case report of XLAG syndrome' *Arq Neuropsiquiatr* **64**(4): 1023-1026.

Stankiewicz, P. and A. L. Beaudet. **2007**. 'Use of array CGH in the evaluation of dysmorphology, malformations, developmental delay, and idiopathic mental retardation' *Curr Opin Genet Dev* **17**(3): 182-192.

Stepp, M. L., A. L. Cason, M. Finnis, M. Mangelsdorf, E. Holinski-Feder, D. Macgregor, A. MacMillan, J. J. Holden, J. Gecz, R. E. Stevenson and C. E. Schwartz. **2005**. 'XLMR in MRX families 29, 32, 33 and 38 results from the dup24 mutation in the ARX (Aristaless related homeobox) gene' *BMC Med Genet* **6**: 16.

Stromme, P., S. J. Bakke, A. Dahl and J. Gecz. **2003**. 'Brain cysts associated with mutation in the Aristaless related homeobox gene, ARX' *J Neurol Neurosurg Psychiatry* **74**(4): 536-538.

Stromme, P., M. E. Mangelsdorf, I. E. Scheffer and J. Gecz. **2002a**. 'Infantile spasms, dystonia, and other X-linked phenotypes caused by mutations in Aristaless related homeobox gene, ARX' *Brain Dev* **24**(5): 266-268.

Stromme, P., M. E. Mangelsdorf, M. A. Shaw, K. M. Lower, S. M. Lewis, H. Bruyere, V. Lutcherath, A. K. Gedeon, R. H. Wallace, I. E. Scheffer, G. Turner, M. Partington, S. G. Frints, J. P. Fryns, G. R. Sutherland, J. C. Mulley and J. Gecz. **2002b**. 'Mutations in the human ortholog of Aristaless cause X-linked mental retardation and epilepsy' *Nat Genet* **30**(4): 441-445.

Sutherland, G. R. **1979**. 'Heritable fragile sites on human chromosomes I. Factors affecting expression in lymphocyte culture' *Am J Hum Genet* **31**(2): 125-135.

Szczaluba, K., M. Nawara, K. Poirier, J. Pilch, M. Gajdulewicz, K. Spodar, J. Chelly, J. Bal and T. Mazurczak. **2006**. 'Genotype-phenotype associations for ARX gene duplication in X-linked mental retardation' *Neurology* **67**(11): 2073-2075.

Tarpey, P. S., R. Smith, E. Pleasance, A. Whibley, S. Edkins, C. Hardy, S. O'Meara, C. Latimer, E. Dicks, A. Menzies, P. Stephens, M. Blow, C. Greenman, Y. Xue, C. Tyler-Smith, D. Thompson, K. Gray, J. Andrews, S. Barthorpe, G. Buck, J. Cole, R. Dunmore, D. Jones, M. Maddison, T. Mironenko, R. Turner, K. Turrell, J. Varian, S. West, S. Widaa, P. Wray, J. Teague, A. Butler, A. Jenkinson, M. Jia, D. Richardson, R. Shepherd, R. Wooster, M. I. Tejada, F. Martinez, G. Carvill, R. Goliath, A. P. de Brouwer, H. van Bokhoven, H. Van Esch, J. Chelly, M. Raynaud, H. H. Ropers, F. E. Abidi, A. K. Srivastava, J. Cox, Y. Luo, U. Mallya, J. Moon, J. Parnau, S. Mohammed, J. L. Tolmie, C. Shoubridge, M. Corbett, A. Gardner, E. Haan, S. Rujirabanjerd, M. Shaw, L. Vandeleur, T. Fullston, D. F. Easton, J. Boyle, M. Partington, A. Hackett, M. Field, C. Skinner, R. E. Stevenson, M. Bobrow, G. Turner, C. E. Schwartz, J. Gecz, F. L. Raymond, P. A. Futreal and M. R. Stratton. **2009**. 'A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation' *Nat Genet* **41**(5): 535-543. **(Appendix A.4)**

Tassone, F., R. J. Hagerman, A. K. Taylor, L. W. Gane, T. E. Godfrey and P. J. Hagerman. **2000**. 'Elevated levels of FMR1 mRNA in carrier males: a new mechanism of involvement in the fragile-X syndrome' *Am J Hum Genet* **66**(1): 6-15.

Thorson, L., C. Bryke, G. Rice, A. Artzer, C. Schilz, J. Israel, S. Huber, J. Laffin and G. Raca. **2010**. 'Clinical and molecular characterization of overlapping interstitial Xp21-p22 duplications in two unrelated individuals' *Am J Med Genet A* **152A**(4): 904-915.

Tracy, T. E. and L. S. Mulcahy. **1991**. 'A simple method for direct automated sequencing of PCR fragments' *Biotechniques* **11**(1): 68-75.

Trochet, D., L. de Pontual, B. Keren, A. Munnich, M. Vekemans, S. Lyonnet and J. Amiel. **2007**. 'Polyalanine expansions might not result from unequal crossing-over' *Hum Mutat* **28**(10): 1043-1044.

Troester, M. M., T. Trachtenberg and V. Narayanan. **2007**. 'A novel mutation of the ARX gene in a male with nonsyndromic mental retardation' *J Child Neurol* **22**(6): 744-748.

Tuppen, H. A., E. L. Blakely, D. M. Turnbull and R. W. Taylor. **2010**. 'Mitochondrial DNA mutations and human disease' *Biochim Biophys Acta* **1797**(2): 113-128.

Turner, G., M. Partington, B. Kerr, M. Mangelsdorf and J. Gecz. **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**(4): 405-411.

Uyanik, G., L. Aigner, P. Martin, C. Gross, D. Neumann, H. Marschner-Schafer, U. Hehr and J. Winkler. **2003**. 'ARX mutations in X-linked lissencephaly with abnormal genitalia' *Neurology* **61**(2): 232-235.

Van Esch, H., K. Poirier, F. de Zegher, M. Holvoet, T. Bienvenu, J. Chelly, K. Devriendt and J. P. Fryns. **2004**. 'ARX mutation in a boy with transsphenoidal encephalocele and hypopituitarism' *Clin Genet* **65**(6): 503-505.

Vasconcelos, M. M. **2004**. '[Mental retardation]' *J Pediatr (Rio J)* **80**(2 Suppl): S71-82.

Verkerk, A. J. M. H., M. Pieretti, J. S. Sutcliffe, Y.-H. Fu, D. P. A. Kuhl, A. Pizzuti, O. Reiner, S. Richards, M. F. Victoria and F. Zhang. **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**(5): 905-914.

Vissers, L. E., J. de Ligt, C. Gilissen, I. Janssen, M. Steehouwer, P. de Vries, B. van Lier, P. Arts, N. Wieskamp, M. del Rosario, B. W. van Bon, A. Hoischen, B. B. de Vries, H. G. Brunner and J. A. Veltman. **2010**. 'A de novo paradigm for mental retardation' *Nat Genet* **42**(12): 1109-1112.

Wallerstein, R., R. Sugalski, L. Cohn, R. Jawetz and M. Friez. **2008**. 'Expansion of the ARX spectrum' *Clin Neurol Neurosurg* **110**(6): 631-634.

Wang, Q., L. Elghazi, S. Martin, I. Martins, R. S. Srinivasan, X. Geng, M. Sleeman, P. Collombat, J. Houghton and B. Sosa-Pineda. **2008**. 'Ghrelin is a novel target of Pax4 in endocrine progenitors of the pancreas and duodenum' *Dev Dyn* **237**(1): 51-61.

Warren, S. T. **1997**. 'Polyalanine expansion in synpolydactyly might result from unequal crossing-over of HOXD13' *Science* **275**(5298): 408-409.

Weiler, I. J. and W. T. Greenough. **1993**. 'Metabotropic glutamate receptors trigger postsynaptic protein synthesis' *Proc Natl Acad Sci U S A* **90**(15): 7168-7171.

Whibley, A. C., V. Plagnol, P. S. Tarpey, F. Abidi, T. Fullston, M. K. Choma, C. A. Boucher, L. Shepherd, L. Willatt, G. Parkin, R. Smith, P. A. Futreal, M. Shaw, J. Boyle, A. Licata, C. Skinner, R. E. Stevenson, G. Turner, M. Field, A. Hackett, C. E. Schwartz, J. Gecz, M. R. Stratton and F. L. Raymond. **2010**. 'Fine-scale survey of X chromosome copy number variants and indels underlying intellectual disability' *Am J Hum Genet* **87**(2): 173-188. **(Appendix A.10)**

WHO. **1996**. 'ICD-10 Guide For Mental Retardation' *World Health Organisation* **10th edition**.

Willemsen, R., M. Hoogeveen-Westerveld, S. Reis, J. Holstege, L. A. Severijnen, I. M. Nieuwenhuizen, M. Schrier, L. van Unen, F. Tassone, A. T. Hoogeveen, P. J. Hagerman, E. J. Mientjes and B. A. Oostra. **2003**. 'The FMR1 CGG repeat mouse displays ubiquitin-positive intranuclear neuronal inclusions; implications for the cerebellar tremor/ataxia syndrome' *Hum Mol Genet* **12**(9): 949-959.

Wohlrab, G., G. Uyanik, C. Gross, U. Hehr, J. Winkler, B. Schmitt and E. Boltshauser. **2005**. 'Familial West syndrome and dystonia caused by an Aristaless related homeobox gene mutation' *Eur J Pediatr* **164**(5): 326-328.

Woo, T. H., B. K. Patel, M. Cinco, L. D. Smythe, M. L. Symonds, M. A. Norris and M. F. Dohnt. **1998**. 'Real-time homogeneous assay of rapid cycle polymerase chain reaction product for identification of Leptonema illini' *Anal Biochem* **259**(1): 112-117.

Xiao W, Oefner PJ. **2001**. 'Denaturing high-performance liquid chromatography: A review' *Hum Mutat* **17**(6): 439-474.

Xu, X. L., R. Zong, Z. Li, M. H. Biswas, Z. Fang, D. L. Nelson and F. B. Gao. **2011**. 'FXR1P But Not FMRP Regulates the Levels of Mammalian Brain-Specific microRNA-9 and microRNA-124' *J Neurosci* **31**(39): 13705-13709.

Yamagata, K., N. Oda, P. J. Kaisaki, S. Menzel, H. Furuta, M. Vaxillaire, L. Southam, R. D. Cox, G. M. Lathrop, V. V. Boriraj, X. Chen, N. J. Cox, Y. Oda, H. Yano, M. M. Le Beau, S. Yamada, H. Nishigori, J. Takeda, S. S. Fajans, A. T. Hattersley, N. Iwasaki, T. Hansen, O. Pedersen, K. S. Polonsky, G. I. Bell and et al. **1996**. 'Mutations in the hepatocyte nuclear factor-1alpha gene in maturity-onset diabetes of the young (MODY3)' *Nature* **384**(6608): 455-458.

Yamagishi, T., M. Ozawa, C. Ohtsuka, R. Ohyama-Goto and T. Kondo. **2007**. 'Evx2-Hoxd13 intergenic region restricts enhancer association to Hoxd13 promoter' *PLoS One* **2**(1): e175.

Yamatogi, Y. and S. Ohtahara. **2002**. 'Early-infantile epileptic encephalopathy with suppression-bursts, Ohtahara syndrome; its overview referring to our 16 cases' *Brain Dev* **24**(1): 13-23.

Yang, R., B. Frank, K. Hemminki, C. R. Bartram, B. Wappenschmidt, C. Sutter, M. Kiechle, P. Bugert, R. K. Schmutzler, N. Arnold, B. H. Weber, D. Niederacher, A. Meindl and B. Burwinkel. **2008**. 'SNPs in ultraconserved elements and familial breast cancer risk' *Carcinogenesis* **29**(2): 351-355.

Yu, S., M. Pritchard, E. Kremer, M. Lynch, J. Nancarrow, E. Baker, K. Holman, J. C. Mulley, S. T. Warren, D. Schlessinger and et al. **1991**. 'Fragile X genotype characterized by an unstable region of DNA' *Science* **252**(5010): 1179-1181.

Zahir, F. and J. M. Friedman. **2007**. 'The impact of array genomic hybridization on mental retardation research: a review of current technologies and their clinical utility' *Clin Genet* **72**(4): 271-287.