THE UNIVERSITY of ADELAIDE

SUB CRUCE LUMEN

Investigating the Pathogenic Mechanism of Expanded Polyalanine Tract Mutations in the ARX Homeobox Transcription Factor causing Intellectual Disability

Tessa Mattiske

B.Sc, Biomedical Science (Hons)

Intellectual Disability Laboratory

Primary Supervisor: A/Prof. Cheryl Shoubridge

Co-supervisors: Dr Kristie Lee and Prof. Jozef Gecz

Thesis submitted for the degree of

Doctor of Philosophy

in

Discipline of Paediatrics, Adelaide Medical School
Faculty of Health and Medical Sciences
The University of Adelaide

2017

II

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Abstract

Intellectual disability (ID) is a highly prevalent disorder that affects 1-3% of the population, with ~ 100 causative genes mapped to the X-chromosome. The *Aristaless-related homeobox gene* (*ARX*) is an important transcription factor with critical roles in development, particularly in the developing brain. Variants are not well tolerated within *ARX,* with missense mutations resulting in phenotypes that always involve ID with frequent comorbidities of epilepsy, infantile spasms, hand dystonia, autism or dysarthria. Historically, it was thought that only males were affected by mutations in *ARX* due to their single X-chromosome. In this thesis, we describe a family with multiple affected individuals, including two females with a novel insertion mutation within *ARX*. We furthermore review the reported phenotype of females with mutations in *ARX* and highlight the importance of screening *ARX* in both male and female patients with ID and seizures.

The majority of patients with *ARX* mutations are affected by expansion mutations in polyalanine tract 1 and 2 within the protein, giving rise to ID with or without epilepsy and movement disorders of varying severity. Mice modelling the two most frequent polyalanine expansion mutations ($Arx^{(GCG)7}$ referred to as PA1 and $Arx^{432-455dup24}$ referred to as PA2) recapitulate many of the clinical features seen in humans (Kitamura et al. 2009, Jackson et al. Submitted 2017). To dissect the molecular basis of different polyalanine expansions *in vivo*, we used 12.5 dpc brain samples from PA1 and PA2 mice to analyse disruptions in gene expression in the developing forebrain to capture the primary disruption leading to the developmental phenotypes caused by these mutations. A greater number of genes deregulated in the more severe PA1 mice was shown, with the majority of genes also

perturbed in the milder PA2 mice, but failed to reach significance compared to WT at this early stage of development. We saw a significant overlap with a number of known direct targets of ARX (5%) and genes implicated in ID, epilepsy and autism (12%). From my analysis, I predict a core pathway of transcription regulators as potential drivers of the ID and infantile spasms in patients with ARX polyalanine expansion mutations.

Next, I investigated the mechanisms driving the partial loss of function. My data indicates this reduced function does not occur through disruptions of binding to DNA or protein interactors in relation to the region of ARX spanning both polyalanine tract 1 and 2. However, in this thesis, I demonstrate a marked reduction in polyalanine mutant protein may be the contributing factor to disease manifestation. Transcription activity assays indicated ARX responds in a dose-depend manner, and greater reduction in protein leads to an increased severity of the disease. Investigations into the molecular mechanism contributing to this reduction in protein level show no significant change in protein stability (*in vitro*). Instead, initial studies indicate inefficiency of translation resulting in reduced protein abundance. From my data and other previous studies, I discuss the likelihood of a multiple hit model contributing to the partial loss of function and leading to the variability of clinical presentations.

Thesis Structure Format

This thesis is presented in a conventional format consisting of an introduction covering the background of the research conducted in this thesis (Chapter 1), followed by a chapter containing material and methods (Chapter 2), 3 results chapters (Chapters 3-5) and a final discussion and conclusions (Chapter 6). Chapter 3 and 4 provide the basis for published papers, but are presented here in the conventional format and include in some cases additional data. The accepted publications included in the appendix.

Thesis Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide.

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 acknowledge that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

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 Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Tessa Mattiske

Bachelor of Science (Honours)

Student Number: a1160579

DATE ...03/...07./.2017

Acknowledgements

While undertaking my PhD many people have aided me with their support and expertise. Most importantly, I'd like to thank my supervisors A/Prof Cheryl Shoubridge, Dr Kristie Lee and Prof. Jozef Gecz for their guidance, motivation and patience. Cheryl's knowledge in the area of neurogenetics, in particular, ARX, polyalanine tracts and functional assays, Kristie's expertise in the field of animal models and developmental biology and Jozef's long term involvement in the field of neurogenetics provided me with a great foundation from which to learn. Each has willingly provided their time in meetings, offering valuable advice and editing my work, in particular, Cheryl. I will forever be grateful for this opportunity to have been mentored by such amazing individuals academically and personally.

Thank you to the members of the Intellectual Disability Laboratory, past and present: May Tan, Kristie Lee, Susan Hinze, Ching Moey, Megan Bleeze, Matilda Jackson, Karagh Loring and Oliver Dearsley. Special thanks must go to Kristie Lee, Megan Bleeze, and Matilda Jackson for maintaining the animal breeding colonies, helping in the setup of timed matings and euthanasia of animals. Kristie Lee for mentoring me on animal dissections and *in situ* hybridization. Additionally all members of The Neurogenetics Laboratory who have offered valuable feedback, guidance and resources when required. Everyone has contributed in their own way to make me a better scientist.

I am grateful to the clinicians and genetic counsellors for coordinating the genetic material needed for patient screening of ARX mutations. I am also deeply thankful to the patients

and their families for giving consents and for donating their genetic material which helps in our understanding of this important gene.

Finally, I would like to thank my family and friends. I would not be where I am today if it wasn't for my mum, Christine and dad, Brenton for forever instilling the importance of seeking knowledge. My grandmothers, Simona and Margaret, thank you for your support and encouragement. My sister, Danielle for always offering an ear to listen and eyes to edit. My brother, Alexander for trying to pretend to understand what is included in this thesis. Lastly but by no means least, Jason you have done everything and more in supporting and contributing to my personal growth, and a constant source of motivation. However, most importantly thank you for always putting a smile on my face ☺

Abbreviations

Abbreviation	Full Description
ACC	agenesis of the corpus callosum
AG	abnormal genitalia
ARX	*aristaless*-related homeobox
Bp	base-pair
CGH	comparative genomic hybridization
CNS	central nervous system
CNV	copy number variant
DD	developmental delay
DE	deregulated genes
DNA	deoxyribonucleic acid
Dpc	days post coitum
EIEE	early infantile epileptic encephalopathies
GE	ganglionic eminences
GO	gene ontology
HD	homeodomain
HYD-AG	hydranencephaly and abnormal genitalia
ID	intellectual disability
IEDE	infantile epileptic encephalopathy
IQ	intelligence quotient
ISSX	infantile spasms X-linked
MR/TS/Dys	mental retardation with tonic seizures with dystonia

mRNA	messenger RNA
NDDs	neurodevelopmental disorders
NGS	next generation sequencing
NLS	nuclear localisation sequence
NSID	non-syndromic intellectual disability
NS-XLID	non-syndromic X-linked intellectual disability
OAR	*aristaless* domain
OP	octapeptide
ORF	open reading frame
OS	Ohtahara syndrome
PA	polyalanine tract
PA1	mouse modelling the Arx(GCG)7 expansion mutation in polyalanine tract 1
PA1-23A	(GCG)7 insertion mutation (c.306GGC[17]) in polyalanine tract 1 in humans
PA2	mouse modelling the 24 bp duplication mutation in polyalanine tract 2
PA2-20A	24 bp duplication mutation (Arx432-455dup24) in polyalanine tract 2 in humans
PolyA	polyalanine expansion
PolyAPool	combined data from PA1 and PA2 mice
PRTS	Partington syndrome
RNA	ribonucleic acid

VOUS	variant of unknown significance
VZ	ventricular zone
WES	whole exome sequencing
WT	wild-type
XCI	X chromosome inactivation
XLAG	X-linked lissencephaly associated with abnormal genitalia
XLID	X-linked intellectual disability
XLMESID	X-linked myoclonic epilepsy with spasticity and ID

Chapter One:

Introduction

1   Introduction

The human brain is a highly complex structure, and its normal development and functioning is critically dependent on the proper and tightly regulated activity of a large number of genes. Consequently, there are more than 1,000 Mendelian disorders listed in OMIM for which intellectual disability is one or the only hallmark of the condition. Hundreds of causative genes have already been identified. However, it is predicted many more genes remain to be identified due to patients remaining undiagnosed.

1.1   Intellectual disability

Intellectual disability (ID) encompasses a range of clinically and genetically variable disorders defined by defects in the central nervous system and involving a reduced capacity for intellectual functioning and adaptive behaviour (Schalock et al. 2011). The defining feature of intellectual disability is characterised by an IQ (intelligence quotient) of less than 70, impairment in at least two adaptive skills, and disease onset before 18 years of age. With a prevalence of 1-3% worldwide, intellectual disability has a high impact on public health systems and families of affected individuals alike (Larson et al. 2001, Maulik et al. 2011). In Australia, it is conservatively estimated that intellectual disability costs $14,720 billion annually (Doran et al. 2012).

Clinical manifestations distinguish intellectual disability into a syndromic form with additional consistent biological, physiological or radiological abnormalities, or a nonsyndromic ID with cognitive impairment the only consistent clinical feature (Ropers

2010). The level of intellectual functioning is assessed by the status of the IQ value, and this value subclasses the disease condition as mild (IQ level: 50-55 to 70), moderate (IQ level: 35-40 to 50-55) severe (IQ level: 20-25 to 35-40), or profound (IQ level: below 20-25) (Puzynski 1992).

1.2   Genetic etiology of intellectual disability

In developed countries, intellectual disability is caused by both genetic and environmental factors, with equal rates of incidence. The environmental factors may involve malnutrition/undernutrition, prolonged depressing conditions due to cultural deprivation, exposure to toxic substances, physical trauma or head injury, severe infection, pregnancy complications and childhood illness. Genetic forms of intellectual disability prevail in 50% of ID cases, and this number might be expected to be even higher in those countries where consanguineous marriages are commonly practised. However severe intellectual disability occurs in 0.5% of newborns and is thought to be largely genetic in origin (Ropers 2010, Mefford et al. 2012) and can arise due to chromosomal anomalies, mitochondrial disorders, epigenetic defects, repeat expansion diseases and single gene impairments (Winnepenninckx et al. 2002).

Next-generation sequencing (NGS) has revolutionised gene discovery in patients with intellectual disability and has enabled an unprecedented expansion in the number of genes implicated in this disorder. Now a genetic basis for ID is well established. Cytogenetic alterations including aneuploidies, duplications, deletions, translocations, and inversions

visible on standard karyotype account for approximately 5% of cases (Stankiewicz and Beaudet 2007), and smaller copy number variants (>400 kb) (CNVs) detected by array comparative genomic hybridization (CGH) explain a further 14% of cases (Cooper et al. 2011). Among the monogenic causes, there are more than 100 genes on the X chromosome (Lubs et al. 2012) and more than 600 autosomal recessive genes in which mutations of which cause ID (Vissers et al. 2016). With the use of high throughput technologies, the detection of autosomal genes has not only increased but has become instrumental in the detection of *de novo* mutations.

1.3 X-linked intellectual disability

X-linked gene defects have long been considered to be substantial causes of ID. ID is significantly more common in males than in females, with a reported excess of 30% to 40% of males versus female patients (Lehrke 1972, Lehrke 1974, Penrose 1983, Leonard and Wen 2002). However, only in part due to mutations in genes on the X-chromosome (10%) (Mandel and Chelly 2004). X-linked inheritance can be recognised in families, even cases with only two affected male patients and an obligate female carrier can be considered for X-linked genes. Additionally, the hemizygous status of males means that variant identification is often more straightforward to detect and validate than for autosomal conditions. These reasons lead the initial interest of investigators and clinicians into X-linked intellectual disability (XLID) and consequently, more than 150 XLID syndromes have been characterized and have led to the discovery of more than 100 genes known to date which are responsible for 81 of the known XLID syndromes and over 50 families with

4

nonsyndromic XLID (Lubs et al. 2012, Tzschach et al. 2015). Additionally, 30 XLID syndromes and 48 families with nonsyndromic XLID have been regionally mapped, however, the genes have not yet been identified (Lubs et al. 2012).

1.4   The *Aristaless*-related homeobox gene

Of particular interest to our area of research focus, is that impact of mutations in genes involved in early neurodevelopment and embryogenesis. One of the main genes frequently mutated on the X-chromosome and resulting in ID, the *Aristaless*-related homeobox (*ARX*), was first identified in 2002 (Bienvenu et al. 2002, Stromme et al. 2002). Since then, over 60 different mutations have been described and implicate *ARX* in a wide spectrum of disorders extending from phenotypes with severe neuronal migration defects such as X-linked lissencephaly with abnormal genitalia (XLAG) to milder forms of intellectual disability without apparent brain abnormalities but with associated features of dystonia and epilepsy (Shoubridge et al. 2016).

The *ARX* gene (Genbank: NM_139058.2) (MIM# 300382) spans a 12.25-kb region at Xp21.3 (Human Genome Build: GRCh38/hg38) and has a 1689 bp open reading frame (ORF), comprising of five coding exons, and encoding for a 562 amino acid protein (Figure 1.1) (Ohira et al. 2002). Designated *ARX* in humans and *Arx* in mice, with significant homology between human and mouse (89.3% sequence and 95% amino acid homology). ARX encodes a transcription factor which belongs to the family of Aristaless-related genes, a subset of the Paired-related homeobox genes. These genes are important regulators of

essential events during vertebrate embryogenesis, including the development of the central and peripheral nervous systems (Meijlink et al. 1999). ARX has several conserved domains: the octapeptide domain, three nuclear localisation sequences, four polyalanine tracts, the DNA-binding homeodomain and the *Aristaless* domain (Figure 1.1).
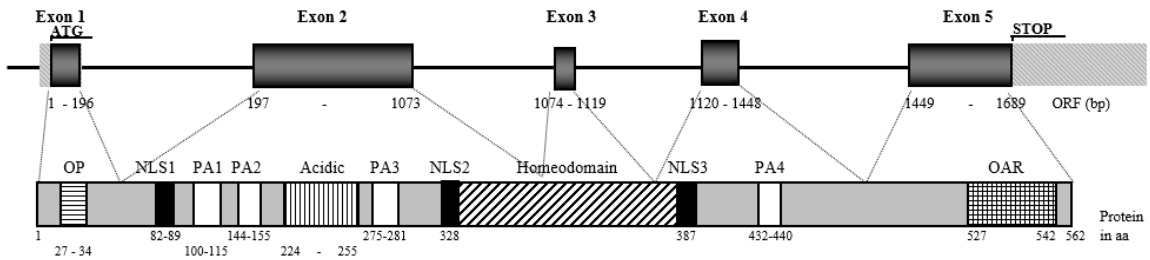
Figure 1.1 Gene and protein structure of *ARX*/ARX homeobox transcription factor.
Exon-intron structure of *ARX* gene with the five exons boxed, and the open reading frame in dark grey, ATG and STOP codon positions, 5' and 3' untranslated regions depicted with diagonal light grey stripes, base pair sequence span of individual exons shown below each exon (top panel). Predicted protein structure of ARX with known functional domain highlighted; Octapeptide (OP, horizontal stripes), three nuclear localisation sequences (NLS, black), four polyalanine tracts (PA, white), acidic domain (vertical stripes), homeodomain (diagonal stripes), and the Aristaless domain (OAR, crosshatched) with the corresponding amino acid below each domain (middle panel).

1.5   Mutations within *ARX* leads to a spectrum of phenotypes

Mutations in the Aristaless-related homeobox gene are identified in a broad spectrum of neurocognitive disorders. Since its discovery, 138 cases (individuals/families) have been reported in *ARX* that lead to 10 distinct clinical phenotypes that all feature ID (Table 1) (Shoubridge et al. 2010a). The number of cases, however, is likely to be an underestimation as many are diagnosed in a diagnostic setting and not reported. The finding of a single gene having significant clinical and genetic heterogeneity has led to a considerable amount of research on this gene, both at the clinical and basic science levels.

Patients with mutations in *ARX* present with ID alone, or in conjunction with an array of additional co-morbidities including lissencephaly, infantile spasms, epilepsy, autism, testicular malformation, dysarthria, and hand dystonia (Shoubridge et al. 2010a). Due to its location on the X-chromosome, hemizygous males that carry the mutation are symptomatic and heterozygous females transmit the mutant allele but are often asymptomatic (Bonneau et al. 2002). The clinical spectrum of phenotypes due to *ARX* mutations has been recently reviewed by our group and others (Kato et al. 2004, Wallerstein et al. 2008, Shoubridge et al. 2010a, Marques et al. 2015, Shoubridge et al. 2016). For this introduction, I will briefly summarise *ARX* disorders which can generally be divided into malformation and non-malformation groups.

1.5.1   Malformation disorders

The most severe clinical phenotypes cause brain malformation and include X-linked lissencephaly associated with abnormal genitalia (XLAG) (MIM# 300215) (Kitamura et al. 2002), hydranencephaly and abnormal genitalia (HYD-AG) (MIM# 300215) (Kato et al. 2004) and Proud syndrome (MIM# 300004) (Proud et al. 1992). These phenotypes are a result of premature termination mutations, missense mutations within the homeobox domain and the nuclear localisation sequences flanking this domain (Table 1 and Figure 1.2) (Kitamura et al. 2002, Kato et al. 2004, Shoubridge et al. 2010a, Shoubridge et al. 2010b).

1.5.2   Non-malformation disorders

In the non-malformation patients, the clinical severity ranges from ID as a sole consistent clinical presentation (NS-XLID) (MIM# 300419), to disorders with additional symptoms, including seizures and movement disorders. The seizure types include infantile spasms X-linked (ISSX/West syndrome) (MIM# 308350), X-linked myoclonic epilepsy with spasticity and ID (XMESID) (MIM# 308350), and Ohtahara syndrome / early infantile epileptic encephalopathies (EIEE) (MIM# 308350). In addition, the movement disorder Partington syndrome (PRTS) (MIM# 309510) describes dystonia, particularly of the hands. These non-malformation phenotypes are frequently caused by expansions of the first or second polyalanine tracts and missense mutations outside the homeodomain (Table 1 and Figure 1.2) (Stromme et al. 2002).

1.6   Emerging heterozygous female phenotype

As *ARX* is on the X-chromosome, mutations in this gene typically result in families with affected males across multiple generations transmitted via (usually) asymptomatic carrier females. With decreasing costs of new sequencing technologies, increased access and numbers of samples have been screened for mutations, including singleton females screened for X-chromosome gene mutations. Removing the ascertainment bias has revealed *ARX,* and a growing number of X-chromosome genes, including *IQSEC2*, *USP9X* and *PHF6,* that have phenotypic effects in males and females that are distinct, depending on the functional severity of the mutation (Zweier et al. 2013, Reijnders et al. 2016, Zerem et al. 2016). Examination of heterozygous females from families with known mutations in *ARX* supports that the female carrier phenotype might be under-ascertained. A number of carrier females display intellectual or learning disabilities, although other carriers of the same mutation are phenotypically normal. In females, the only CNS malformation currently described is agenesis of the corpus callosum (ACC) with or without intellectual disability and seizures (Kato et al. 2004, Marsh et al. 2009). Females have also been reported to have infantile spasms, epilepsy, and varying degrees of cognitive dysfunction without obvious brain malformation (Figure 1.2) (Kato et al. 2004, Wallerstein et al. 2008, Marsh et al. 2009, Eksioglu et al. 2011, Bettella et al. 2013, Kwong et al. 2015).

Variation within *ARX* is not well tolerated. This is demonstrated from the single largest aggregation of coding variants in the world known as the ExAC database (Lek et al. 2016). The ExAC database consists of exome sequencing data from 60,706 individuals from more than 20 research studies and a resource that can be used to look at the level of selective

constraint against variation across genes. Highly pathogenic variants with a large effect should be selected against and seen with a lower frequency in the general population. For *ARX*, there are no loss of function variants reported emphasising the critical requirement for this gene. As many as 40 missense variants are reported at a very low frequency (<0.03%), but no incidence of homozygote variants, 11 (27.5%) hemizygotes (likely males) and the remaining (72.5%) are likely heterozygous females, hence, they cannot be ruled out as potentially pathogenic in males. These variants occur across *ARX* however, the very few that fall within functional domains are in heterozygous female non-symptomatic carriers (Figure 1.2). The variable penetrance in females, which is further discussed in chapter 3, makes the disease impact of *ARX* mutations in females a difficult issue to quantitate.
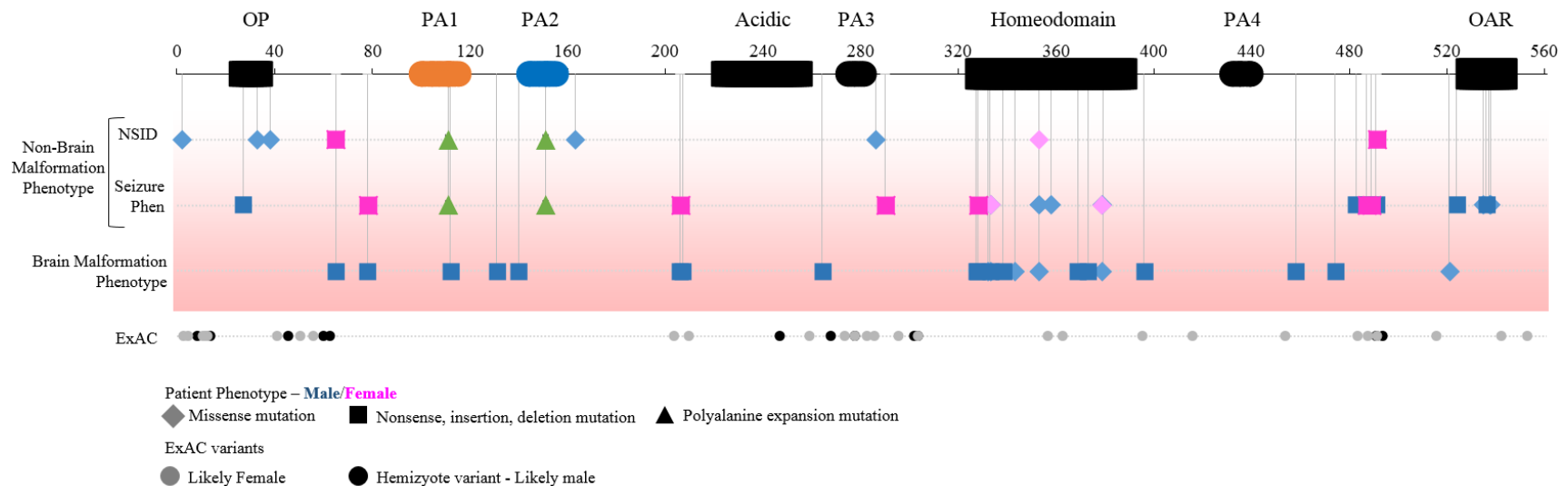
Figure 1.2 Identified ARX mutation in Male and Females leading to arrange of phenotypes.
Overview of ARX with the protein domain locations highlight in black and the polyalanine tract 1 (orange) and polyalanine tract 2 (blue). All reported mutations are shown according to their relative position at the protein level in both males and females. Different mutation types are indicated by a change in symbol while missense mutation is shown by a diamond and all other nonsense, insertion or deletion mutation by a square. Polyalanine expansion mutations are shown by green triangles. Phenotype severity is indicated on the y-axis. Along the bottom of the graph are the reported ExAC variants in *ARX* indicated by a circle symbol, with hemizygote variants (likely male) shown in black and the remaining variants (likely heterozygous females) in grey.

Table 1 Clinical phenotypes of patients with mutations in *ARX*.
(Adapted and updated from Shoubridge et al. 2010)

Pathogenic features	Phenotype	Families affected	Families per mutation type and location	
	NS-XLID	48	37 Duplication	Expanded PA 2
			5 Insertion	Expanded PA 1 (4)/ PA 2(1)
			4 Missense	
	AG	1	1 Missense	OAR
Seizures	PRTS	12	12 Duplication	Expanded PA 2
	MR/TS/Dys	1	1 Insertion	Expanded PA 1
	XMESID	1	1 Missense	HD
	ISSX (WS)	22	2 Deletion	Loss of OAR
			8 Duplication	Expanded PA 2
			8 Insertion	Expanded PA 1
			4 Missense	
	IEDE	10	10 Insertion	Expanded PA 1
	OS / EIEE	11	1 Nonsense	Protein restart after Oct domain
			5 Insertion	OAR (3)/ Expanded PA1 (2)
			4 Missense	OAR (3 male/1 female)
			2 Deletion	OAR (1 male/ 1 female)
Brain Malformation	Proud (ACC/AG)	1	1 Missense	HD
	HYD/AG	2	2 Nonsense	HD
	ISSX, PMG, PVNH	1	1 Insertion	Expanded PA 1
	XLAG + other features	28	14 Deletion	Protein truncation
			4 Insertion	Protein truncation
			7 Missense	HD and OAR
			2 Nonsense	Protein truncation
			2 Splice	Skip exons/protein truncation

AG: abnormal genitalia. HD: Homeodomain. HYD/AG: hydranencephaly with abnormal genitalia. IEDE: infantile epileptic encephalopathy. ISSX, PMG, PVHG: X-linked infantile spasms with periventricular nodular heterotopia and polymicrogyria. ISSX (WS): infantile spasms X-linked (West syndrome). MR/TS/Dys: Mental retardation with tonic seizures with dystonia. NS-XLID: nonsyndromic X-linked intellectual disability. OAR: *Aristaless* domain. Oct: Octapeptide domain. OS: Ohtahara syndrome – early infantile epileptic encephalopathy. PA: Polyalanine tract. Proud (ACC/AG): Proud syndrome with agenesis of the corpus callosum and abnormal genitalia. PRTS: Partington syndrome – intellectual disability with dystonic movements, ataxia and seizures. XLAG: X-linked lissencephaly with abnormal genitalia. XMESID: X-linked myoclonic epilepsy with severe intellectual disability.

1.7   Identification of mutations within *ARX*

Although fifteen years has passed since the first disease-causing mutations were identified in *ARX*, novel mutations and new clinical phenotypes are still being characterised. Genetic diagnose of an *ARX*-related neurodevelopmental (NDD) disorder can be problematic due to the number of phenotypes associated with *ARX* mutations and variable penetrance of the severity of the overlapping phenotypes with other neurodevelopment disorders. These subtleties result in initial genetic screening not always focused on *ARX*. This is particularly relevant in the case of affected females. With the improvement in technology and increased diagnostic services, we have seen the implementation of whole genome sequencing, whole exome sequencing, and ID gene panels. Despite improvements in diagnosis with these tools, the high GC content of *ARX* means that even these sophisticated sequencing approaches often fail to reach adequate coverage across this gene. For example, exome sequencing using benchtop ion proton machines result in poor coverage of the *ARX* gene, with mean coverage reported at 43% (Lacoste et al. 2016). Although disease-causing variants occur across all five coding exons, most are located in the largest exon, exon 2, resulting in expansion mutations of the first or second polyalanine tracts. As most mutations are in the largest exon *ARX* mutation analysis is routinely confined to screening exon 2 using Sanger sequencing or PCR size analysis. To continue characterising this complex genotype-phenotype relation researchers (Poirier et al. 2006, Fullston et al. 2011, Tan et al. 2013, Marques et al. 2015) have designed specific sequencing conditions to provide comprehensive screening of the whole gene. The continued characterization of *ARX* variants not only contributes to the identification of specific phenotype features but also assists in unravelling the pathogenicity of additional rare variants. This type of

information and correct diagnosis has particular implications for accurate genetic counselling for families and furthering our understanding of the underlying biology of ARX.

Result Chapter 3 Summary: An emerging female phenotype with loss of function mutations in the Aristaless-related homeodomain transcription factor.

In my first result chapter (chapter 3) I present the outcomes of ARX screening in a family presenting ARX NDD.

Hypothesis: Patients with negative results using gene panels or WES/WGS and clinical presentation compatible with ARX NDD will have an ARX mutation when tested by an optimised Sanger sequencing protocol.

Aim: To provide comprehensive Sanger sequencing of the entire *ARX* gene in patients presenting with key clinical features associated with *ARX* mutations which may have also been missed by high-throughput sequencing technologies and targeted sequencing of the polyalanine tract 1 and 2 diagnosed screening.

Expected Outcomes: Collate data of ARX mutations to highlight the relationships between genotype and phenotype 'severity'. This information is required to be able to provide evidence-based recommendations for the consideration of pathogenicity in affected individuals with novel variants identified in *ARX*.

1.8 ARX is highly expressed during brain development

ARX is expressed in several structures including the brain, pancreas, developing testes, heart, skeletal muscle and liver. However, the most striking consequences of loss of function of ARX concern the function and development of the brain and testis in both mouse and human (Kitamura et al. 2002, Colombo et al. 2004, Poirier et al. 2004). Considering the association between *ARX* mutations and congenital disorders involving ID, it is not surprising that the highest levels of ARX expression is found within the human fetal brain. Given the difficulties in mapping ARX expression during development in the human setting, researchers rely on surrogate models. In mice, *Arx* expression is first detectable at the three-somite stage. After the 10-somite stage, *Arx* expression is confined to a specific area in the anterior neural plate (Colombo et al. 2004). By embryonic day 8 (E8) Arx is expressed in the developing hypothalamus, thalamus, basal ganglia and cerebral cortex (Bienvenu et al. 2002, Colombo et al. 2004). By E10.5, the expression is clearly observed in the ventricular zone (VZ) of the dorsal cortex and begins to be expressed in the mantle zone of the emerging ganglionic eminences (GE) (Colombo et al. 2004). Ventral telencephalic Arx expression is clearly present in the mantle zone of the GE at E12.5 (Figure 1.3 and Figure 1.4) (Colombo et al. 2004). At later stages, it is widespread throughout telencephalic structures such as the ganglionic eminences, the cerebral cortex and the hippocampus (Bienvenu et al. 2002, Colombo et al. 2004, Poirier et al. 2004). This expression pattern remains until shortly after birth when the dorsal VZ and GE is no longer present, and ARX expression in the forebrain is observed only in scattered cells in the cortex and basal ganglia (Figure 1.3).
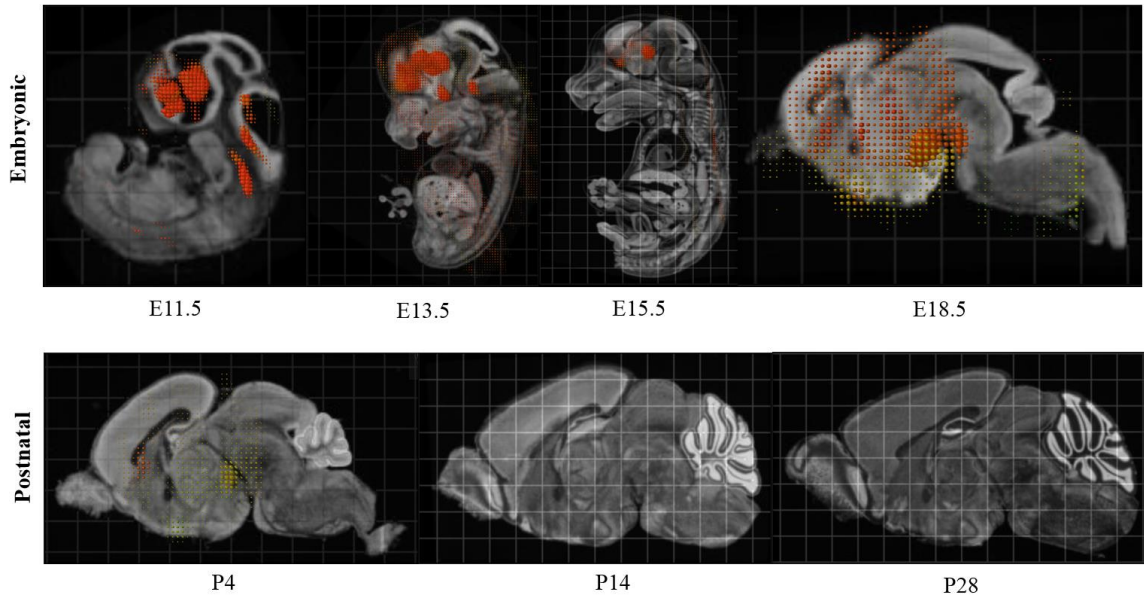
Figure 1.3 ARX expression throughout embryo development and in the brain across postnatal life.
Images are taken from the Allen Mouse Brain Atlas show *Arx* expression at four embryonic developmental time points and three postnatal time points with the log expression (red = high and yellow – low) shown. Image credit: Allen Institute for Brain Science (http://mouse.brain-map.org).

1.9   Modelling *ARX* mutations

The functional role of ARX in the telencephalon is an active area of research by our laboratory and others. To understand the role of ARX in brain development, investigators have performed overexpression or loss-of-expression experiments using different models including *X. laevis, C. elegans* and *M. musculus*. Since the first description of ARX-related disorders, a number of Arx mouse models have been developed and have led to various insights into the disorder.

Kitamura et al. (2002) generated a knock-out mouse replacing *Arx* exon 2 with *Lac*Z gene, producing an *Arx* allele with a premature stop codon and resulting in a truncated protein.

These mice died at P0-P2 but were found to have many of the features of XLAG in humans. These mice have been instrumental in the study of the downstream targets of Arx and the alterations in interneurons, basal ganglia, and cholinergic neurons during brain development. Similarly, Collombat and colleagues created a mouse line with target knock-out of exon 1 and 2, producing a mutant protein with the loss of the first 360 N-terminal amino acids (Collombat et al. 2003). As these Arx-deficient mice died 2 days after birth due to developing severe hypoglycemia, attributable to the loss of glucagon-producing alpha cells, no phenotypic or physiological analysis could be performed. Therefore, a series of other Arx mice lines were developed and published in 2009.

A conditional knockout model was established using the Cre-*Lox* system to stop expression of Arx in cells expressing Dlx5/6, a transcription factor that is expressed in the majority of developing interneurons within the GE (Marsh et al. 2009). The male mice developed convulsive seizures at P14 and as indicated disruption to ARX function impacts development of interneuron subpopulations. In contrast to the mice modelling knock-out of Arx, models with known human mutations associated with disease were next to be generated. These include two mutations disrupting the homeodomain, one associated with the XLAG phenotype and the other with the XLID phenotype (Kitamura et al. 2009). Two mice models exist for a knock-in *Arx* mouse with the expansion of the first polyalanine tract, increasing the repeat size by 7 alanines (Kitamura et al. 2009, Price et al. 2009). Although it is essential to consider the limitations of each model when interpreting the various mutation-specific results, particularly when relating back to the human setting, the investigation of how Arx impacts normal brain formation and contributes to gene

regulatory networks of telencephalon development requires the continued use of these various Arx mutant mouse models. It is my hope by furthering our understanding of how Arx dysfunction contributes to varying degrees of cognitive dysfunction with and without epilepsy phenotypes, will contribute to the knowledge that will be vital to be able to generate potential therapeutic interventions for patients with Arx mutations. I predict that by identifying key regulatory events in brain development may provide mechanistic insights into other genetic disorders with overlapping phenotypes of ID and epilepsy.

1.10 Understanding the functional role of ARX in interneuron development

The loss of interneurons in patients with *ARX* mutations are believed to be a major contributor to the epilepsy and infantile spasms phenotype (Kato and Dobyns 2005). Interneurons make up 25% of human cortical neurons with almost all originating from the ganglionic eminences (Ma et al. 2013). Their dynamic modulation of cortical activity is necessary for normal cognition and underlies multiple aspects of learning and memory. Cortical interneurons collectively function to maintain the excitatory-inhibitory balance in the cortex by dampening hyperexcitability and synchronising activity of projection neurons, primarily through the use of the inhibitory neurotransmitter gamma-aminobutyric acid (GABA). Disruption of the excitatory-inhibitory balance is a common pathophysiological feature of multiple seizure and neuropsychiatric disorders, including epilepsy, schizophrenia, and autism. The mice modelling mutations or knock-out of Arx all indicate a pivotal role in tangential and radial migration (Figure 1.4) of a broad range of GABAergic interneuron precursors (Kitamura et al. 2002, Friocourt et al. 2008). The

18

extensive cellular co-localization of Arx with GABA in mouse (and human) brains, as well as the absence of interneurons documented in the cortex of XLAG patients and Arx mice, have led the field to propose 'interneuronopathy' as the term to describe the group of pathologies ARX are responsible for (Kato and Dobyns 2005). This 'interneuronopathy hypothesis' proposes that a developmental deficiency of inhibitory cortical interneurons caused by defects in proliferation, specification, differentiation and/or migration, result in epilepsy (Kato and Dobyns 2005, Marsh et al. 2009). Complete loss of Arx within the developing subpallium is detrimental for tangential migration of GABAergic interneuron precursors into the developing cortex such that a significant loss of a broad range of GABAergic interneurons was noted in mutant mouse models, with specific subtypes affected including calbindin (Cb), calretinin (Cr), acetylcholine (ChAT) positive cells in both cortex and striatum (Kitamura et al. 2002, Collombat et al. 2003, Fulp et al. 2008, Marsh et al. 2009).

While a number of transcription factors have been identified as critical to establishing the diversity of interneuron subtypes, the specific role of these transcription factors, including Arx, in delineating interneuron subpopulations is incompletely understood. Early *Mash1* expression is known to set ventral forebrain identity (Rallu et al. 2002), and subsequent expression of Dlx1/2 is vital for interneuronal fate determination (Anderson et al. 1997a, Anderson et al. 1997b). The Dlx transcription factors, expressed early in the interneuron lineage, appear to be essential in controlling interneuron fate. Multiple other transcription factors (Nkx2.1, Lhx6, and Sox6) have also been identified as necessary for interneuron subtype determination, refinement of the lateral, medial and caudal ganglionic eminence

populations, and migration to the olfactory bulb, striatum, and cortex (Butt et al. 2005, Wonders and Anderson 2006, Hebert and Fishell 2008). Though it is clear that loss of Arx alters interneuron development, and that Arx function is primarily, though not exclusively, downstream of the Dlx transcription factors, the ultimate role of Arx in this process needs to be further elucidated (Colasante et al. 2008).
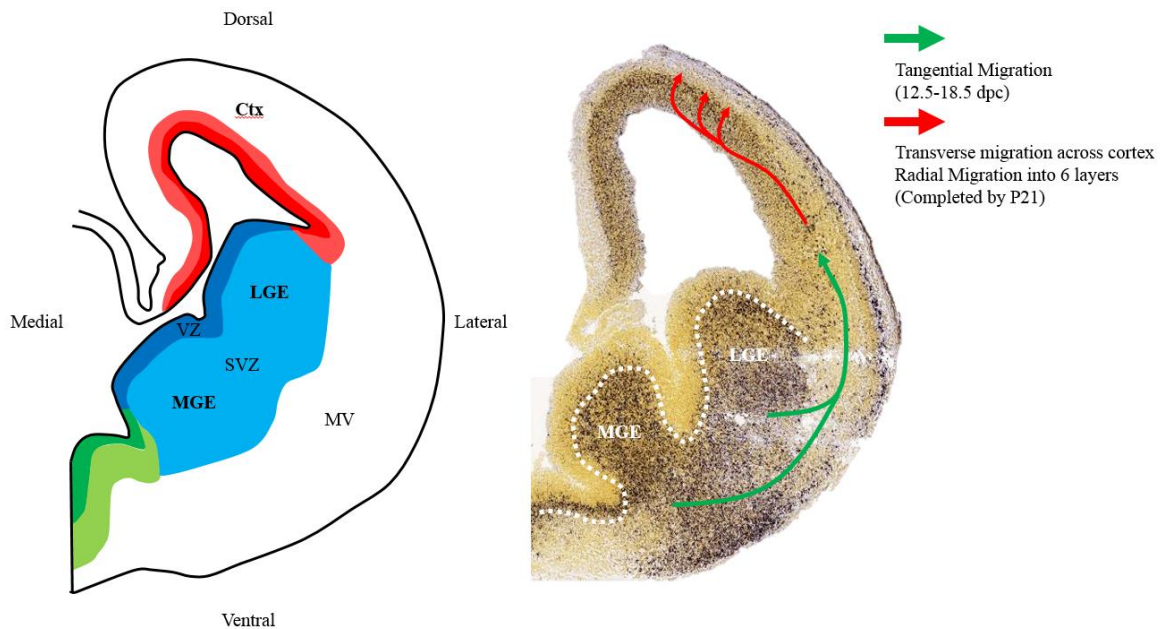
Figure 1.4 ARX expression profile in key neuronal proliferative zones.
A representation of a coronal section of the brain from an embryonic day 14 mouse and the major subdivisions of the telencephalic zones (left panel). Ctx, cortex; LGE, lateral ganglionic eminence; MGE, medial ganglionic eminence; VZ, ventricular zone; SVZ subventricular zone, MZ, Mantel zone. The ARX expression profile in a coronal section of an embryonic day 14 mouse with the major migration pathways followed by neurons from the proliferative zones. Green arrows follow tangential migration and red arrows represent the transverse migration across the cortex and radial migration into the 6 layer of the cortex.

1.11 ARX as a transcription factor

ARX transcription activity is defined by the presence of a homeodomain consisting of a conserved 60 amino acid DNA binding domain. The homeodomain is comprised of three α-helices that are preceded by a flexible N-terminal arm. Helix II and III form a helix-turn-helix major structural motif that folds to make contact with the major groove of DNA. Further contact with DNA is made in the minor groove as mediated by the N-terminal arm (Kissinger et al. 1990, Gehring et al. 1994). Because of this feature, homeodomain proteins are identified as DNA-binding transcription factors that recognise specific DNA sequences to access their target genes in the genome and to control their expression. The molecular mechanisms employed by homeodomain proteins to regulate transcription, such as ARX, are still poorly understood. A central issue is how homeodomain transcription factors select their target genes in the genome using a DNA binding domain with limited sequence specificity, which recognises short (four to six nucleotides), AT-rich sequences. A study in 1988 identified a conserved 5'-TAAT-3' motif as a binding site of high-affinity for numerous homeodomain proteins (Desplan et al. 1988), which is now accepted as the canonical homeodomain binding sequence. It was later confirmed by mutational studies that illustrated the preferred binding motif for engrailed is 5'-TAATTA-3' (Ades and Sauer 1994). Microarray and Chromatin Immunoprecipitation (ChIP) approaches have added to our general understanding of how ARX regulates gene expression *in vivo* and further defined possible targets and binding preferences.

In addition to the homeodomain, the ARX protein contains a number of conserved domains, providing some insight into its potential role(s). The aristaless domain, which is

present in more than 40 other members of the paired class homeobox family is located off the carboxyl-tail of ARX. The function of this domain is not well understood, although a number of studies have shown it is involved in transcription activation (Figure 1.5) (Norris et al. 2000, Norris and Kern 2001, Collombat et al. 2003, Collombat et al. 2005, McKenzie et al. 2007). Four hydrophobic polyalanine tracts, located throughout the ARX protein, are suggested to be involved in protein-protein and protein-DNA interaction and as such to stabilise the interaction between transcription regulators and/or DNA (Brown and Brown 2004). A study identifying 95 amino acids within ARX that includes the fourth polyalanine tract, showed this region is largely responsible for the transcription repression activity of ARX (McKenzie et al. 2007) (Figure 1.5). Also contributing to the repressive nature of ARX is a highly conserved octapeptide domain located near the N-terminus. This sequence shares high similarity with the Engrailed homology repressor domain (eh1) known to be involved in transcriptional repression both *in vitro* and *in vivo*. This domain recruits Grouch/transducing-like enhancer of split (TLE) co-factor protein (TLE1-4), which modulate transcription repression activity.
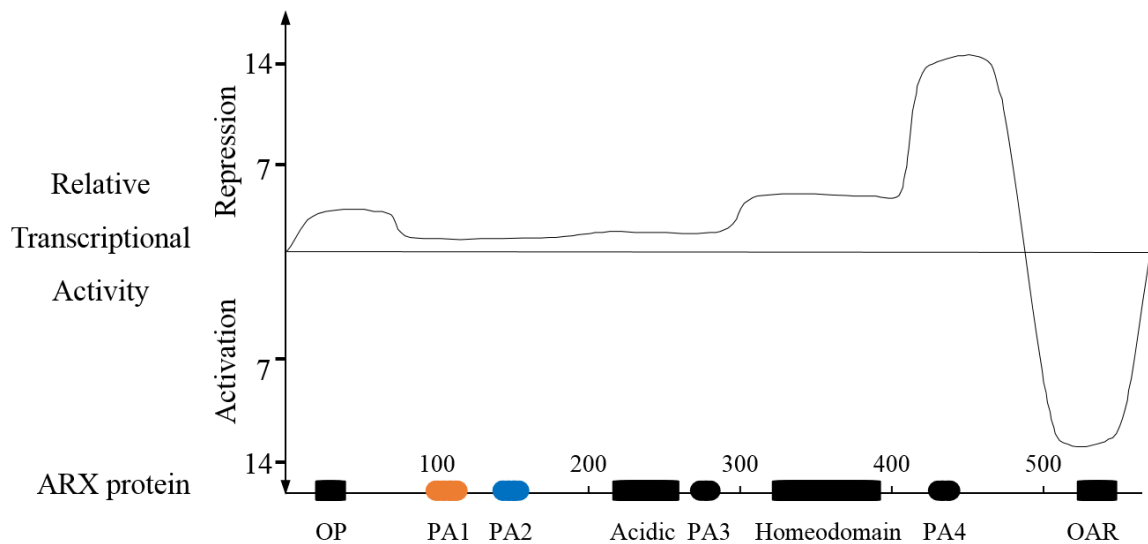
Figure 1.5 Mapping of the ARX transcription regulatory domains.
A schematic diagram showing the distribution and relative strength of the transcription activator and repressor domains along the ARX protein shown belong (adapted from McKenzie et al. 2007).

1.12 Investigation of ARX-dependent transcriptional networks.

Given that ARX is a transcription factor, defining the transcriptional targets and gene regulatory networks in which ARX participates has been essential in further elucidating its role in normal brain development and how disruptions in ARX lead to human disease. Due to the fundamental role that ARX plays in the brain, multiple studies have been conducted with the majority using complete knock-out of ARX to determine its regulatory function and specific impact on brain development. Although two gene expression profiles comparing E14.5 wild-type and Arx mutant ventral telencephalic tissues have recently been published in mouse, very few targets for this transcriptional factor have been described, and only three have been confirmed to be direct (Fulp et al. 2008, Colasante et al. 2009). Evidence indicates that Arx is primarily a transcriptional repressor, which binds

to the 5' promoter region of *Shox2*, *Ebf3*, and *Lmo1* and represses their expression. However, these studies focused on the ventral telencephalon for their gene expression experiments, thus possibly overlooking genes involved in neuroblast proliferation and/or radial migration. Using chromatin immunoprecipitation in Arx-transfected neuroblastoma cells (N2a) or E15.5 mouse embryonic brain, followed by hybridization to mouse promoter arrays (ChIP-chip), Quille et al. identified new direct targets of Arx (mouse) (Quille et al. 2011). A total of 1006 Arx-bound genes were found, with a significant proportion of these promoters enriched for a sequence very similar to a motif previously identified as an ARX-binding motif (6-mer 5'-TAATTA-3'). Similarly, Quille selected only one stage of development in the mouse for their ChIP experiment, and it is, therefore, likely that they have not identified genes involved in earlier steps such as brain patterning or later steps such as synaptogenesis and connectivity. A limitation with overexpression studies is that Arx is not normally expressed in N2a cells and thus may lack binding partners and/or co-factors necessary to regulate the expression of certain genes. Hence, while expression studies using Arx-transfected cell lines may provide important insights for further research, these studies are likely to underappreciate the full profile of ARX function.

The question of how the loss of ARX in humans (and mice) results in the catastrophic XLAG phenotypes is unknown but must reflect significant alterations in ARX-dependent transcriptional networks. Gene expression studies have shown genes regulated (directly or indirectly) by ARX are involved in the patterning of the central nervous system, axonal guidance, neurodevelopment, and neurotransmission and neurite outgrowth (Quille et al. 2011). Interestingly, these studies also suggest new possible functions for Arx, for example

in osteoblast differentiation or mesenchymal cell proliferation and differentiation (Quille et al. 2011).

1.13 Polyalanine expansion mutations leads to partial loss of function

Strikingly, mutations leading to the expansion of the first or second polyalanine tracts account for over half of the mutations reported in *ARX* (Shoubridge et al. 2010a, Marques et al. 2015). It remains challenging to assess the true prevalence of *ARX* mutations given the potential ascertainment bias and also due to routine diagnostic screening of polyalanine 1 and 2 findings often not reported. The in-frame 24 duplication (previously reported as c.429_452dup, now following HGVS nomenclature reported as c.441-464dup; referred to as PA220A mutation in this thesis) of the second polyalanine tract, and expansion of the first polyalanine tract (previously reported as c.304ins(GCG)[7], now following HGVS nomenclature reported as c.306GGC[17] ; referred to as PA123A mutation in this thesis) are the most common. These two mutations alone are causative of an array of clinical presentations (Stromme et al. 2002, Absoud et al. 2010, Mirzaa et al. 2013, Marques et al. 2015). It was commonly thought that there was a correlation between increasing expansion length and an increase in phenotype severity. However, the growing number of phenotypic reports for these mutations, have highlighted the expansion of this genotype-phenotype relationship. ARX polyalanine tract expansion mutations of the same size and composition can cause a range of overlapping, but distinct clinical phenotype and mutations in different tracts can result in the same clinical phenotype. Figure 1.6 captures the unusual span of variation in the clinical presentation, intra and interfamilial variability, associated with the

frequent expansion mutations in polyalanine tract 1 and 2, in particular, that has been noticed, but not adequately explained, and thus research is continuing into the molecular mechanism driving the pathogenesis associated with these mutations (Shoubridge et al. 2007, Price et al. 2009, Fullston et al. 2011, Nasrallah et al. 2012, Beguin et al. 2013, Lee, K. et al. 2014, Olivetti et al. 2014).

Figure 1.6 Clinical variability associated with expanded in polyalanine tract mutations in *ARX*.
The panel above the graph shows the normal length of the polyalanine tract 1 (PA1) as 16 alanine residues and polyalanine tract 2 (PA2) as 12 alanine residues. The graph shows the relationship between the number of alanines and phenotype severity in published families with polyalanine tract 1 (orange) and polyalanine tract 2 families (blue). The size of each circle is proportional to the number of reported cases. With increasing length of residues in the polyalanine tracts, the clinical presentation becomes more severe (adapted from Shoubridge et al. 2016).

1.14 Polyalanine tract expansion mutations

1.14.1 Role of polyalanine tracts

Almost 500 proteins have been identified in humans that contain polyalanine segments, making polyalanine tracts the third-most prevalent homopeptide repeat in eukaryotes, behind polyglutamine and polyasparagine (Bernacki and Murphy 2011). In contrast to polyglutamine tracts, which can tolerate a broad number of residues with expansions in the hundreds often required to produce disease, polyalanine tracts never consist of >20 alanines (Albrecht and Mundlos 2005) (Figure 1.7). Furthermore, they do not undergo the dynamic expansion that leads to the significant intergenerational increase of repeat length encoding glutamines.
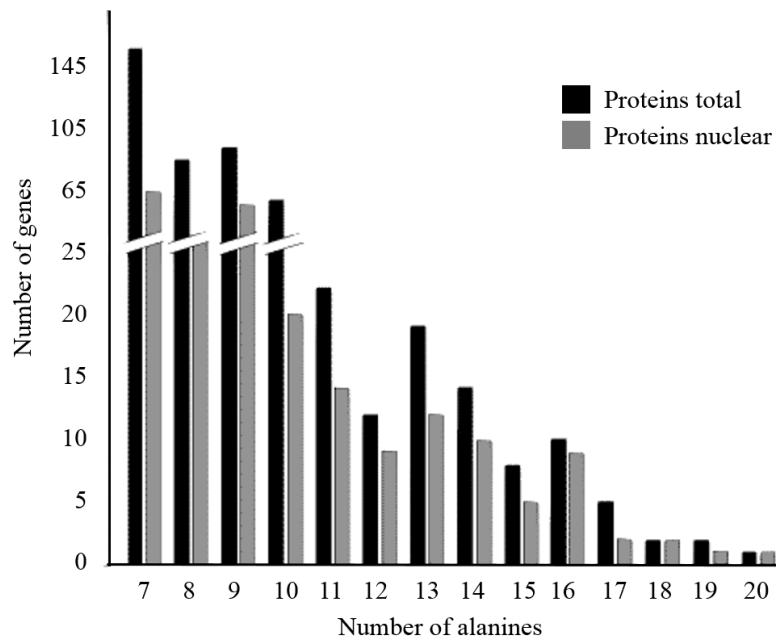
Figure 1.7 Distribution of polyalanine repeat-containing proteins in the genome, and the frequency of nuclear localisation.
A genome-wide search based on Swiss-Prot and Trembl revealed a total of 461 proteins that contain polyalanine repeats between 7 to 20 alanines. Almost 50% (218) of these are known to have a nuclear localisation. (Albrecht and Mundlos 2005)

The function of polyalanine tracts within proteins has not been unequivocally confirmed. However, the interdomain or terminal location of the repeat tract within genes suggests that they may play a role as flexible space elements between individually folded domains in molecules, mediating protein-protein or protein-nucleic acid interaction. These tracts may also serve a function in mediating the assembly of protein complexes. Repeats of seven or more alanines are particularly common in transcription factors (Bernacki and Murphy 2011) and in several transcription factors, alanine-rich regions have been shown to be responsible for activation/repression of target genes (Han and Manley 1993, McKenzie et al. 2007). Therefore it has also been proposed that polymeric runs might serve to fine-tune the activity of these transcription factors (Karlin and Burge 1995). Although the precise molecular mechanism of this activity via polyalanine tracts remains to be determined.

1.14.2 Expansions to polyalanine tracts cause disease

Unlike polyglutamine, expanded alanine tracts are stably transmitted across multiple generations, most giving rise to a number of congenital disorders (Figure 1.8) (Albrecht and Mundlos 2005). The ubiquitous RNA-binding protein PABPN1 has disease causing expansions to polyalanine tracts and is associated with the late onset disease oculopharyngeal muscular dystrophy (OPMD). In contrast, all other expansions occur in developmentally important transcription factors including SOX3 (X-linked Hypopituitarism), HOXA13 (hand–foot–genital syndrome), HOXD13 (synpolydactyly type II), PHOX2B (congenital central hypoventilation syndrome), FOXL2 (blepharophimosis, ptosis and epicanthus inversus), ZIC2 (holoprosencephaly) and

RUNX2 (cleidocranial dysplasia) and ARX (Figure 1.8). The absence of any wild-type protein with a polyalanine tract exceeding 20 residues indicates that there is a critical threshold above which polyalanine expansion are not tolerated. The number of tracts within these proteins ranges between 1-4 while also varying in length and with no clear preference in location, extending right across the protein (Figure 1.8). The normal length of polyalanine tracts tends to be between 9-20 alanines with a sharp threshold for expansion (+1-14A) resulting in a disease phenotype (Figure 1.9). Wild-type proteins with tracts as low as 12A are prone to expansion (HOXA13 and ARX) with an addition of even a single alanine to tract 1 in ARX (16A to 17A) resulting in disease. The largest reported expansion is an addition of 14A in tract 3 of HOXA13 (18A to 32A) however the longest tract resulting is disease in an addition of 13A in tract 2 of PHOX2B resulting is a total tract length of 33A. Both replication slippage and non-homologous recombination have been proposed to explain the increase in tract length and reviewed in Shoubridge and Gecz, 2012 (Shoubridge and Gecz 2012). Despite considerable functional analyses, the mechanism by which polyalanine expansion mutations cause disease remains to be accurately determined.
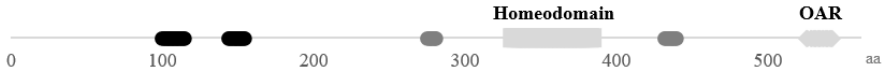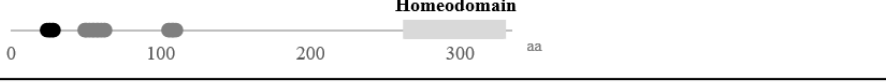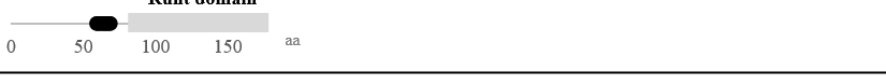
Gene (MIM)	Polyalanine diseases	Disease phenotype	Protein function	Protein Structure
ARX (300382)	Syndromic and non-syndromic X-linked mental retardation (XLMR)	NS-XLID; PRT; ID/PRTS/dys; ISSX/WS; IEDE; OS	Cerebral development and patterning	Homeodomain, OAR
ZIC2 (603073)	Holoprosencephaly (HPE)	HPE5	Central nervous system development	DNA-Binding Domain
SOX3 (313430)	X-linked hypopituitarism (XH)	XH	Neural development and the correct function of the hypothalamic–pituitary axis	HMG Domain
HOXA13 (142959)	Hand–foot–genital syndrome (HFGS)	HFGS	Genitourinary tract and limb development	Homeodomain
HOXD13 (142959)	Synpolydactyly type II (SPD)	Syndactylyl type 5 SPD1	Genitorurinary tract and limb development	Homeodomain
PHOX2B (603851)	Congenital central hypoventilation syndrome (CCHS)	CCHS	Regulation of the autonomic nervous system (ANS)	Homeodomain
FOXL2 (110100)	Blepharophimosis, ptosis and epicanthus inversus (BPES)	BPES I and II -/+ POF	Ovarian maintenance and eyelid development	Fork-head Domain
RUNX2 (600211)	Cleidocranial dysplasia (CCD)	Brachydactyly and minor clinical features CCD	Osteoblast differentiation and skeletal development	Runt domain

Figure 1.8 Polyalanine expansion mutations linked to eight developmentally important transcription factors.
A brief description of the disease resulting from polyalanine expansions in the transcription factors listed. The protein structure highlights the main functional domain/s (light grey) and the number/location of polyalanine tract within each protein (grey). Tracts with reported expansions are in black.

Figure 1.9 Polyalanine tract expansions resulting in disease.
Polyalanine tract rarely exceeds 16 alanine within a single tract. For each transcription factor with reported polyalanine expansion mutations, the number of alanine in each tract of wild type is shown in grey and reported expansions result in disease shown in orange.

1.15 Mechanisms of polyalanine expansion disease

Protein aggregation occurs in the related polyglutamine disorders, where polyglutamine expansion confers a toxic gain of function (Davies et al. 1997, La Spada and Taylor 2010). In the case of polyalanine disease, nuclear inclusions that contain mutant PABPN1 protein are a hallmark of OPMD (MIM# 164300) (Calado et al. 2000). Together, this type of data suggests that aggregate formation might have a pathogenic role in polyalanine disease alleles with a gain of function activity. In support of this prediction, initial studies from multiple groups indicate that expansion of polyalanine tracts above a certain threshold may induce misfolding and aberrant protein interactions, degradation, mislocalisation and/or aggregation (Albrecht et al. 2004, Caburet et al. 2004, Bachetti et al. 2005, Utsch et al. 2007, Moumne et al. 2008), including studies from our laboratory in regards to ARX (Shoubridge et al. 2007, Fullston et al. 2011). The thresholds at which these events occur differs between proteins, but all polyalanine tract proteins behaved similarly *in vitro* such that over-expression in cell culture results in the generation of cytoplasmic and/or nuclear aggregates (Albrecht et al. 2004, Caburet et al. 2004, Nasrallah et al. 2004, Bachetti et al. 2005, Shoubridge et al. 2007, Fullston et al. 2011). Out of a cellular context, filter precipitation assays have demonstrated that polyalanine expansion proteins are extremely prone to spontaneously from aggregates compared to wildtype proteins (Albrecht et al. 2004). This indicates that the polyalanine expansion is necessary and sufficient for aggregation of the protein. All polyalanine expansion mutations that have been modelled in cell assays show a clear correlation between increasing expansion length and the propensity to aggregate and mislocalise (Albrecht and Mundlos 2005, Fullston et al. 2011), contributing to the prediction that the length of the tract contributes to phenotypic severity.

Apart from OPMD, for all diseases due to expanded polyalanine tract mutations, the human data on *in vivo* aggregation of mutant proteins is not available or not supportive of aggregation as a mechanism. Hence, the critical question of whether aggregates form *in vivo* and, if so, how they may be implicated in the pathogenesis of loss of function polyalanine diseases remains to be answered. The growing evidence, however, does indicate that aggregation formation is likely to be an inherent artefact due to the overproduction and/or inefficient degradation of the mutant proteins due to *in vitro* overexpression. As such, when mutant protein is studied at physiological levels, aggregation does not seem to occur. Retroviral infection systems permitting incorporation of a single gene copy per cells (Albrecht et al. 2004) and targeted mutagenesis of embryonic stem cell lines (Hughes et al. 2013) have been used to test the effect of low levels of polyalanine expansion mutant proteins. Both these systems expressing either HOXD13 (Albrecht et al. 2004) or SOX3 (Hughes et al. 2013) at a physiological dose revealed a dramatic reduction of the mutant protein.

Over the last decade, a number of polyalanine expansion mutation mouse models have been generated to study the phenotypic outcomes of these common mutations seen in humans and include Hoxd13 (15Ala+9Ala) (spdh) (Bruneau et al. 2001), Hoxa13 (18Ala+10Ala) (Innis et al. 2004), ARX (Kitamura et al. 2002, Price et al. 2009), Phox2B (20Ala+7Ala)(Dubreuil et al. 2008) and Sox3 (15Ala+11A)(Hughes et al. 2013). Whilst limited molecular work has been completed on these models, a lack of *in vivo* aggregation and instead a reduction of protein is a consistent finding across all models (Bruneau et al. 2001, Albrecht et al. 2004, Innis et al. 2004, Goridis et al. 2010, Hughes et al. 2013, Lee,

K. et al. 2014). Immunohistochemistry and immunoblot analysis show the differing extent of the reduction to mutant protein levels in the heterozygous and homozygous states for SOX3 (Hughes et al. 2013). Hoxd13 protein in spdh/+ showed a reduction by ~50% in the heterozygous and by ~90% in the homozygous spdh/spdh animals. This was mirrored in the Hoxa13 model, and even though no reduction of protein is noted from phox2b27ala/+ (Dubreuil et al. 2008), protein cannot be visualised in the phox2b27ala/*laz*C model (Goridis et al. 2010).

It is unclear if polyalanine expansion causes partial loss of function, a toxic gain of function or a dominant negative effect. Of the nine genes reported with polyalanine expansion mutations, seven are located on the autosomes. The predicted mechanism of protein dysfunction for these autosomal genes becomes complicated by the fact heterozygotes carry one wild-type allele and one mutant allele. Cells specifically expressing the genes in question will, therefore, have both wild-type and mutant protein present in the same cell leading to this heterozygotes states with a more severe phenotype than a heterozygous null mutation. This dominant negative or toxic gain of function is thought to arise from the mutant protein sequestrating the wild-type protein and/or other factors. The remaining two genes, ARX and SOX3 are located on the X-chromosome and due to X-inactivation in females, do not result in expression of both mutant and wild-type protein in a single cell. Thus, characterisation with X-linked models of the effect of polyalanine expansion mutations on protein function and the molecular mechanism behind disease can be elucidated without the added complexity of the presence of wild-type protein.

1.15.1 *ARX* polyalanine expansion mutations modelled in mice

Currently, two mouse models of the most common ARX polyalanine expansion mutations have been generated. There are two independently generated mouse models for the expansion in the first polyalanine tract from 16 to 23Ala ($Arx^{(GCG)7}$ referred to as PA1 throughout this thesis), (Kitamura et al. 2009, Price et al. 2009); and one mouse model for the expansion of the second tract from 12 to 20Ala ($Arx^{432-455dup24}$ referred to as PA2 throughout this thesis) (Kitamura et al. 2009). As mentioned previously, patients with these mutations display a wide spectrum of clinical presentations, ranging from mild intellectual disability only to intellectual disability with early onset of epileptic seizures (Ohtahara syndrome). Similarly, these mouse models recapitulate many of the phenotypic features (Kitamura et al. 2009, Price et al. 2009). Kitamura et al. (2009) reported nuclear localisation in 12.5dpc GE-originated migrating cells and in 18.5dpc cortical interneurons (18.5dpc) in the two mouse models, PA1 and PA2, that recapitulate the respective human mutations (Kitamura et al. 2009). In contrast, Price et al. (2009) reported an increased Arx cytoplasmic localisation in the cerebral neurons from adult brains of their independent mouse line modelling the PA1 mutation. Our subsequent investigations of both PA1 and PA2 mice (originally from Kitamura 2009) confirm that aggregate formation is an unlikely pathogenic hallmark for ARX polyalanine expansion mutations, and as such, alternative mechanisms require investigation.

1.16 Alternative mechanisms of polyalanine expansion pathogenesis

Although a reduction of mutant polyalanine expansion proteins has been shown, no change in mRNA across multiple genes has been detected (SOX3 (Hughes et al. 2013), HOXA13 (Innis et al. 2004), ARX (Lee, K. et al. 2014), HOXD13 (Bruneau et al. 2001, Villavicencio-Lorini et al. 2010)). Therefore, the reduction in mutant protein is hypothesised to result from either a disruption in translation efficiency of the RNA encoding the expanded protein or degradation of the mutant protein *in vivo*. At present, the evidence indicates that post-translational etiology may be involved. In support, *in vitro* transcription/translation of SOX3-26ala and HOXA13-28ala shows no difference in the translation efficiency of polyalanine expansion proteins compared to wild type (Innis et al. 2004, Hughes et al. 2013). However, this has not been confirmed *in vivo*.

1.16.1 Mutant polyalanine expansion proteins are degraded via the proteasome.

Inhibition of the proteasome results in increased aggregation indicating the proteasome plays a significant role in the clearance of polyalanine expansion mutant proteins (Albrecht et al. 2004, Di Zanni et al. 2012, Parodi et al. 2012). On the other hand, up-regulation of hsp40 and hsp70 (two heat shock proteins involved in recognition and refolding denatured molecules and disassembly intracellular protein aggregates) drastically reduces aggregate formation in overexpression studies and leads to an increase in nuclear localisation of the transcription factor (Albrecht et al. 2004, Bachetti et al. 2005, Utsch et al. 2007). Taken together, this data suggests polyalanine expansion mutations do not intrinsically disrupt

nucleus translocation of the protein and can sustain nuclear localisation provided aggregation does not occur. Furthermore, indicates that mutant protein is normally degraded by the proteasome.

Expansion of polyalanine tracts has also been reported to increase the turnover of the mutant protein in comparison to wildtype (Parodi et al. 2012). Cycloheximide treatment showed the half-life of the polyalanine expansion protein was half that of wildtype indicating that the presence of an expanded polyalanine tract increases the turnover rate of the protein (Parodi et al. 2012). Therefore, it is hypothesised that the polyalanine expansion modifies the stability of mutant proteins. Degradation of expanded polyalanine tract proteins by the ubiquitin-proteasome system are indicated by mutant proteins having a higher degree of ubiquitination and proteasome inhibition resulting in increased accumulation of mutant protein (Parodi et al. 2012). Thus, at low (physiological levels) of production, the mutant protein can be efficiently degraded via the proteasome, and we can speculate that *in vivo* the mouse has a protective mechanism against the formation of aggregates.

1.16.2 Functional consequences of polyalanine expansion mutations.

The regulation of transcription by these developmentally important transcription factors is complex, involving many interactions that are tissue and developmental stage-specific. *In vitro* assays of transcriptional activity are highly simplified and are likely to detect only severe alterations in transcription factor function. Some constraints of *in vitro* assays include the interpretation of result as the finding that a mutation results in near-total loss

of activity in such as assay implies that the protein must be altered such that very basic components of the transcriptional machinery do not function correctly. Transactivation assays using firefly luciferase reporter constructs with various proximal promoters saw a correlation with the length the alanine expansion and decrease in activity (Brown et al. 2005, Woods et al. 2005, Bachetti et al. 2007). However determining the effect of increasing alanine tract length on activity of the protein as a transcription factor seems to be problematic as a reduction of similar magnitude to the reduced luciferase outputs is also observed with lower nuclear protein levels of the mutant protein due to aggregation which is an inherent defect of *in vitro* overexpression systems (Hughes et al. 2013). Therefore, this would suggest that the mutant protein that is present in the nucleus has similar activity to WT or is there still an inherent defect in the mutant protein transactivation activity. To combat this, various treatments shown to reduce aggregation have been added to cell culture to alleviate the impact of aggregation on the transactivation assay system (Bachetti et al. 2007, Di Zanni et al. 2012, Parodi et al. 2012). Successful treatments showed to enhance the refolding which prompts the correct localisation of the remaining protein to the nuclear with a decrease in aggregate formation (Bachetti et al. 2007, Di Zanni et al. 2012). PHOX2B polyalanine expansion proteins were shown to recover their transactivation activity when two of the shorter expansion were tested (20Ala+5Ala & +9Ala) however the largest expanded alanine tract mutation (20Ala + 13Ala) could only regain part of the transcriptional activity.

Any loss of activity does not appear to be due to the inability of the protein to bind to DNA (Woods et al. 2005, Takagi et al. 2014) indicating it could be due to an alteration in some

other function of the protein. Many of these transcription factors harbouring polyalanine expansion can function as both activators and repressors. In the case of ARX, recent data indicates, polyalanine expansion mutations have loss of function at only selective targets *in vivo* (Lee, K. et al. 2014) indicating partial loss of function is in a context-dependent manner. It still needs to be established whether the transcriptional impact of the polyalanine expansion mutations depends on the nature of the promoter. Furthermore, it is unclear whether different target promoters have different sensitivities to decreased amount of available mutant protein if selective protein-protein interactions are perturbed, or DNA binding is affected.

1.17 Summary of understanding the molecular mechanism of polyalanine expansion mutations in ARX causing ID and seizures

Result Chapter 4 Summary: Embryonic forebrain transcriptome of mice with polyalanine expansion mutation in the *Arx* homeobox gene

While *in vitro* methods have permitted a rapid growth in our understanding of Arx-driven molecular pathology and transcriptional defects, the use of rodent models engineered with Arx mutations remains necessary. It highlights the importance to investigate how endogenously expressed mutant Arx proteins alter specific steps in interneuron development leading to the neural circuit deficits and clinical features. The recent studies investigating how Arx mutations alter function, the identification of novel transcriptional targets (Fulp et al. 2008, Colasante et al. 2009, Quille et al. 2011), and the creation of clinically relevant models of ARX, represent a significant advance towards understanding the pathogenesis of human ARX-mediated disease. In my research, a major aim was to examine the early consequences of altered transcriptional activity of ARX due to the polyalanine expansion mutations during embryonic development in mice modelling the most frequent mutations in polyalanine tract 1 and 2 (Kitamura et al. 2009). This approach will enable us to examine the expanded polyalanine tract containing protein in an endogenous setting to determine if expanded polyalanine tract in *Arx* have a functional impact on binding to transcriptional targets and transcriptional activity in a relevant neuronal cell environment. Together with the genome-wide expression data, we aim to identify affected gene targets that contribute to the manifestation of clinical features of the associated disease outcomes. Gene expression patterns driven by ARX mutations can also

be examined to highlight genes whose precise roles in neurodevelopment have yet to be described. The mouse models used will be valuable tools to validate downstream gene changes in the context of the development brain, and an opportunity to link these with the human neurological phenotypes. These models are also likely to provide useful tools for investigating novel therapies for ARX-related diseases, for which only limited treatment options exist.

Hypothesis: Expanded polyalanine tract mutations in Arx alter the capacity of Arx as a transcription factor *in vivo* leading to the phenotype observed in patients.

Aim: By transcriptome-wide analysis determine the consequences of polyalanine tract expansions mutations on Arx ability to act as a transcription factor *in vivo*.

Aim: To determine the transcription changes caused by expansion polyalanine tract mutations in Arx contributing to intellectual disability and seizures.

Expected outcomes: To identify transcriptional targets and differences in deregulated targets due to the more severe PA12-3A compared to the milder PA2-20A expanded polyalanine tract mutations contributing to the spectrum of phenotype severity.

Result Chapter 5 Summary: Investigating the molecular mechanism of how polyalanine expansion mutations in ARX lead to a partial loss of function.

Years of *in vitro* data indicate the poly-A tract expansions result in nuclear inclusions most likely caused by misfolding of the mutant protein. However, the lack of inclusions or aggregates in neurons in the brains of our mouse models suggests that aggregation is not the likely pathogenic mechanism contributing to disease. Hence, another major aim of my research was to consider and investigate alternative mechanisms. As the reduction in mutant protein levels is a consistent feature across multiple polyalanine expansion models determining the molecular mechanism driving this reduction may provide important insights for treatment across all polyalanine expansion disorders.

Hypothesis: Reduction of ARX polyalanine expansion mutant protein abundance drives disease in a dose depend manner.

Aim: To determine whether the mutation or reduced protein contributes to the partial loss of function and the phenotype severity between PA1-23A and PA2-20A.

Aim: To establish if inefficiency of translation or increase in degradation underpins the reduction of the mutant protein level compared to WT.

Chapter Two:

Material and Methods

2 Material and Methods

2.1 General solutions

PBS: Phosphate Buffered Saline; 10X: 1.37M NaCl, 27mM KCl, 100mM Na$_2$HPO$_4$, 20mM KH$_2$PO$_4$, adjusted to pH7.4

TBS: Tris Buffered Saline: 50mM Tris-Cl, 150mM NaCl, adjusted to pH7.4

TBE: Tris/Borate/EDTA: 89mM Tris, 89mM boric acid, 2mM EDTA, pH7.6

2.2 Patient screening

2.2.1 DNA extraction

The extraction was set up in a 1.5 ml Eppendorf screw cap tube (containing the cell pellet) with 400 µl of proteinase K buffer, 50 µl of 10% SDS and 50 µl of 10 mg/ml proteinase K (final concentration 1 mg/ml (Proteinase K, recombinant, PCR Grade Cat#03115836001, Roche). The tube was sealed off with laboratory film (Parafilm, Cat# MNP52858-000, VWR) to prevent leakage and placed on rotating wheel overnight at 37°C. DNA was isolated from water insoluble cell components by extracting twice with Phenol/chloroform (1:1 ratio). Aqueous and organic phases were separated by centrifugation for 5 minutes at 10,000 x g between each extraction. To remove RNA contamination, 2 µl of RNase A (RNase A (17,500 U) made up to 2.5 ml (100 mg/ml; 7000 units/ml, solution), Cat#19101, QIAGEN) was added to the DNA mixture and incubated for 1 hour at 37°C. DNA was precipitated by adding ice cold absolute ethanol (EtOH) and incubated at -20°C for 30 minutes, then centrifuged at 10,000 x g for 10 minutes. All supernatant was discarded, and

the DNA pellet was dried under vacuum. The dried DNA pellet was resuspended in 100-300 µl of H_2O depending on the size of the original sample. DNA concentration was determined by measuring absorbance at 260 nm using a UV spectrophotometer (Nanodrop, Cat#ND-1000).

2.2.2 Patient screening

The screening protocol were approved by the appropriate institution review board, and informed consent was obtained from the parents of patients via clinical genetics colleagues. Genomic DNA from the proband was extracted from whole blood using standard techniques. Each of the five exons of *ARX* was amplified by PCR using primers designed to amplify coding and flanking non-coding sequence. The exception to this was exon 2, for which four overlapping amplicons were used to achieve robust amplification of GC-rich regions coding for three polyalanine tracts. The PCR conditions and primer sequences (Table 2.1) and are described in detail previously (Tan et al. 2013). All primer sequences are based on the human sequence of the *ARX* gene [NG_008281.1]. Sequencing reactions were performed using ABI Big Dye terminator chemistry version 3.1 and purified products subjected to an automated capillary sequencing on ABI 3100 sequencer (Applied Biosystems, Foster City, CA, USA) and sequence was compared to the *ARX* reference sequence (NM_139058) using SeqMan module of the Lasergene DNA and protein analysis software package (DNAStar, Inc., Madison, WI).

Table 2.1 ARX –specific primers used for amplification and sequencing of human *ARX* coding regions.

ARX Exon Coverage	Primer Name	Direction	Primer Sequence	Length (bp)	PCR Annealing Tm (°C)	PCR Product (bp)
1	ARXe1-F	Forward	GTCCACTACACTTGTTACCGC	21	60	520
1	ARXe1-R	Reverse	AATTGACAATTCCAGGCCACTG	22		
2.1	CH-ARX2F	Forward	CTGATAGCTCTCCCTTGCCC	20	60	262
2.1	CH-ARX2JR	Reverse	GCGGCCCCTGCGCCGTCCGGCCGTTC	26		
2.2	CH-ARX2J-F2	Forward	CCCCTCCGCCGCCACCGCCAAC	22	60	313
2.2	CH-ARX2J-R2	Reverse	TCCTCCTCGTCGTCCTCGGTGCCGGT	26		
2.3	ARX02-2F	Forward	GCAAGTCGTACCGCGAGAACG	21	60	371
2.3	ARXE2-R7	Reverse	CAGCTCCTCCTTGGGTCACA	20		
2.4	ARXE2P2-F5	Forward	AACTGCTGGAGGACGACGAGG	21	60	392
2.4	ARX02-R2	Reverse	TGCGCTCTCTGCCGCTGCGA	20		
3	ARXe3-F	Forward	GAAATAGCTGAGAGGGCATTGC	22	60	231
3	ARXe3-R	Reverse	TCTCTTGGTTTTGTGAAGGGGAT	23		
4	ARXe4-F	Forward	GACGCGTCCGAAAACAACCTGAG	23	60	551
4	ARXe4-R	Reverse	CCCCAGCCTCTGTGTGTATG	20		
5	ARXe5-F	Forward	ACAGCTCCCGAGGCCATGAC	20	60	347
5	ARXe5-R	Reverse	GAGTGGTGCTGAGTGAGGTGA	20		

2.2.3   PCR product purification

The QIAquick PCR purification kit (Cat#28106, QIAGEN) was used to remove components of the PCR reaction (unused primers) that may interfere with subsequent sequencing reactions. Whole PCR reaction volume mixed with five-time volume of buffer PB. Up to 600 µl of the PCR and buffer PB was transferred into the QIAquick spin column, and centrifuged in a conventional tabletop microcentrifuge at room temperature (15–25°C) at 10,000 x g to bind PCR products onto the membrane. The membrane was washed once with 700 µl of buffer PE. All flow through was discarded, and the column was centrifuged again at 15, 000 x g for 5 minutes to evaporate all leftover alcohol. PCR bands were eluted in 30 µl of H_2O. Purified PCR product concentration was determined using the UV spectrophotometer (Nanodrop, Cat#ND-1000) and stored at 4°C until further use.

2.2.4   Sanger sequencing reaction

Purified DNA was sequenced using BigDye™ Terminator sequencing kit (Applied Biosystem™ BigDye® Terminator v3.1 Cycle Sequencing Kit, Cat#4337455, ThermoFisher). The reaction was set up in 20 µl volume containing DNA (~100 ng of genomic DNA or 400 ng of plasmid DNA), 1 µl of primer at 10µM, 1 µl of BigDye™ V3.1, 3 µl of BigDye™ Buffer v3.1 and distilled H_2O. Sequencing conditions were: 15 cycles of denaturation at 96°C for 30 seconds, annealing at 50°C for 15 seconds and extension at 60°C for 4 minutes. Precipitation of sequencing products was achieved by the addition of 20 µl of distilled H_2O and 60 µl of 100% isopropanol, incubated at room temperature for 15 minutes and centrifuged for 20 minutes at 15,000 x g. A second wash

with 250 µl of 70% isopropanol and centrifuged for 10 minutes at 15, 000 x *g*, prior to the precipitated DNA being air dried for 10 minutes to remove any residual solvents. Dried sequencing products were size fractionated by capillary electrophoresis using an ABI automated DNA sequences by a service provided by the Molecular Genetics Laboratory, Genetic Medicine, SA Pathology (Women's and Children's Hospital, Adelaide, Australia).

2.2.5   X-inactivation testing

Genomic DNA (2 µl total with a final DNA concentration of 100ng/µl) was digested with 1.5 µl (15 Units) of *Hpa*II enzyme (Cat#R0171S, NEB), 2 µl of 10x CutSmart Buffer (Cat#B7204S, NEB) and diluted in H_2O to a final volume of 20 µl and incubated overnight at 37°C (minimum incubation period is 3 hours). An enzyme minus sample was also conducted for comparison of an uncut DNA sample. The enzyme was heat inactivated at 80°C for 20 minutes. Routinely we undertake PCRs targeting the highly polymorphic CAG repeat region in exon 1 of the Androgen Receptor (AR) gene and the CGG repeat region in the 5' UTR of the fragile X gene (FRAXA), amplified with fluorescently labelled primers. PCR reaction set-up and cycle conditions listed in Table 2.2 for AR and Table 2.3 for FRAXA and primer details in Table 2.4.

Table 2.2 Androgen Receptor PCR reaction and cycle conditions.

PCR Reaction		PCR Cycle Conditions	
H_2O	1.6 µl	94°C – 3 mins	
2x Homemade PCR mix	5 µl	94°C – 30 secs	X10 cycles
15mM $MgCl_2$	1 µl	60°C – 30 secs	
AR mix (150ng ea. Oligo)	1 µl	72°C – 30 secs	
1M BME	0.2 µl	94°C – 30 secs	X 12 cycles
Roche Taq	0.2 µl	55°C – 30 secs	
DNA (100ng/ µl)	1 µl	72°C – 30 secs	
Total	10 µl	72°C – 10 mins	
		14°C – Hold	

Table 2.3 FRAXA PCR reaction and cycle conditions.

PCR Reaction		PCR Cycle Conditions	
H_2O	3.4 µl	94°C – 3 mins	
10x PCR Buffer(+15mM MgCl)*	1 µl	94°C – 30 secs	X10 cycles
5x Q solution*	2 µl	60°C – 35 secs	
FRAXA mix (75ng ea. Oligo)	1 µl	72°C – 60 secs	
7deaza G dNTP mix	1.2 µl	94°C – 30 secs	X 12 cycles
Hot Star Taq (1/10 diln in pure H_2O)*	0.2 µl	55°C – 35 secs	
Roche Taq (neat)	0.2 µl	72°C – 60 secs	
DNA (100ng/ µl)	1 µl	72°C – 10 mins	
Total	10 µl	14°C – Hold	
*Items from HotStarTaq DNA Polymerase Kit (Cat#203203, QIAGEN)			

Table 2.4 X-inactivation Primers.

Primer Name	Label	Primer Sequence	Length (bp)	PCR Product (bp)
FRAXA-Forward	Hex	GCTCAGCTCCGTTTCGGTTTCACTTCCGGT	30	309
FRAXA-Reverse	Hex	AGCCCCGCACTTCCACCACCAGCTCCTCCA	30	
Androgen Receptor-Forward	Fam	TCCAGAATCTGTTCCAGAGCGTGC	24	279
Androgen Receptor-Reverse	Fam	GCTGTGAGGGTTGCTGTTCCTCAT	24	

2.3   Animal model – *in vivo* analysis

2.3.1   Animals and tissue collection

Heterozygous female Arx$^{GCG7/+}$ (PA1) and Arx$^{432-455dup/+}$ (PA2) mice were obtained as described in (Lee, K. et al. 2014) and were maintained in the C57BL background. We will refer to these mice as PA1 and PA2 throughout this thesis. All animal procedures were approved by the relevant Animal Ethics committees of the University of Adelaide, SA Pathology and the Women's and Children's Hospital, Adelaide. Mice were housed in the Women's and Children's Hospital Animal House (North Adelaide, Australia) from 2011-2014 and in the Laboratory Animal Services (LAS), the University of Adelaide (Adelaide, Australia) from 2014, onwards. Mice were kept in a 12 hour day: night light cycle with ad libitum water and food. When mice were re-located to the LAS facility, the mice were housed behind the barrier under a specific pathogen free environment. For tissue collection, the animals were sacrificed by cervical dislocation (2 weeks or older) or decapitation (neonates and embryos) to achieve rapid death and to minimise distress to animals.

2.3.2   Embryo collection and tissue extraction

Pregnant dams were euthanised by cervical dislocation and embryos were harvested from the uterus. As embryos of the same gestational age may differ in their stage of development, each embryo was staged to allow for correct age matching of samples. For younger embryos (12.5 dpc) somites were counted, and for older embryos (14.5 dpc) the development of the eye, limb and tail formation details were taken into account to classify the theiler stage of an embryo. For detailed descriptions of staging criteria and theiler

classifications see (Theiler 1989). Skin and ectodermal layers were removed to isolate the telencephalic vesicles. Samples were snap frozen in liquid nitrogen and stored at −80°C until analysed.

2.3.3 Tissue lyse/genotyping

Genomic DNA was extracted as per Maxwell® 16 Tissue DNA purification Kit manual (Promega™ Maxwell™ 16 Purification Kits, Cat#PRAS1030, Fisher Scientific) or as per High Pure PCR Template Preparation Kit manual for tissue samples (Cat#11796828001, Roche). PCR was performed using FailSafe™ PCR 2X PreMix J (Epicentre, Cat#FSP995J_INCL, Gene Target Solutions) for 35 cycles of 30 seconds of 94°C for denaturation, 30 seconds of 60°C for annealing and 40 seconds of 72°C for elongation. Primers to amplify the Arx knock-in region were described (Kitamura et al. 2009) and are listed in Table 2.5. We also included a Sry sexing PCR as part of our genotyping pipeline as described in (Lee, K. et al. 2014) and listed in Table 2.5.

Table 2.5 Mouse Genotyping PCR Primers.

Detection	Primer Name	Direction	Primer Sequence	Length (bp)	PCR Annealing Tm (°C)	PCR Product (bp)
Sex	Sry-F	Forward	CACTGGCCTTTTCTCCTACC	20	60	349
	Sry-R	Reverse	CATGGCATGCTGTATTGACC	20		
Knock-In	pMC1neo ATGr	Reverse	TGTTCAATGGCCGATCCCAT	20	60	
	mArx jjr	Reverse	CTTTAGCTCCCCTTCCTGGCACAC	24	60	
	mArx kkf	Forward	AAAGGCGAAAAGGACGAGGAAAGG	24		
PA1 mutation	mARX-GCG	Forward	GCGCTGACCACTTTTCCTT	19	60	208
	mARX-GCG v2	Reverse	ACCTCTCCACGGGGACCT	18		
PA2 mutation	mARX-Dp24	Forward	AGGGGAGCGTCAGGACAG	19	60	282
	mARX-Dp24	Reverse	AACAGCTCCTCCTCGTCGT	19		

2.4 Transcriptome analysis

2.4.1 RNA extraction

RNA was extracted from the isolated telencephalon of hemizygous male mice of each strain (PA1 and PA2) and stage matched male wild-type littermates (WT) using Trizol (TRIzol® Reagent, Cat#1559026, ThermoFisher) and RNeasy Mini Kit (Cat#74104, QIAGEN) and RNase-Free DNase set (Cat#79254, QIAGEN). Frozen tissue samples were thoroughly homogenised in 500 µl of TRIzol® Reagent, and the sample progressively passed through a 1 ml pipette tip, 200 µl pipette tip and a 25G needle until completely homogenised and left at RT for no longer than 5 minutes. 200 µl of chloroform was added to the tube. The tube was shaken vigorously for 1 minute and left for 2-3 minutes at RT for the layers to separate, then centrifuged at 10,000 x g for 15 minutes at 4°C. The upper aqueous phase containing RNA was transferred to a fresh tube (usually about 350 to 600 µl). A similar amount of 70% EtOH was added to the tube and mixed well by pipetting. Up to 700 µl of the supernatant was transferred to an RNeasy mini column and centrifuged at 10,000 x g for 1 minute to bind RNA to the membrane. The flow-through was discarded, and the process was repeated until all aqueous phase and EtOH mixture has been used up. The column was washed once with 350 µl of buffer RW1. DNA was removed by adding 80 µl of DNase I from RNase-Free DNase Set (Cat#79254, QIAGEN) incubation mix (70 µ of buffer RDD and 10 µl of DNase I Stock) right onto the membrane. The DNase reaction was left at RT for 15 minutes. The membrane was washed again with 350 µl of buffer RW1, then twice with 500 µl of buffer RPE. RNA was eluted in 30 µl of RNase-Free H_2O.

RNA concentration was determined us the UV spectrophotometer (Nanodrop, Cat#ND-1000).

2.4.2   RNASeq

Library preparation using the TruSeq RNA Sample Preparation Kit v2 was performed on 4.5 µg of RNA at the ACRF South Australia Cancer Genomics Facility (Adelaide, Australia). Samples (n=4 each from WT, PA1, PA2) were sequenced on the Illumina (San Diego, CA, USA) HISeq 2000 platform. 100 bp paired-end reads were obtained and were mapped to the mouse genome reference sequence (UCSC mm10). The number of reads mapped to each gene was obtained using htseq-count (Anders et al. 2015). In order to correct for variation between lanes/samples, the count data was normalised to library size. Genes with low count data were excluded, the minimum required at least 2 sample having > 60-177 reads. Differential gene expression was calculated using the R package, edge R (Robinson et al. 2010). Transcripts significantly altered compared to WT were selected by applying a log2 fold change-cutoff of 1.1 and p-value cutoff of ≤0.05.

2.4.3   RNASeq validation – RT-PCR

RNA-Seq results were validated using TaqMan RT-qPCR on two pools of RNA, a technical validation pool using RNA from the same samples used the RNASeq and a separate biological validation pool of RNA from 4 different samples of each genotype. RNA was extracted as per 2.4.1. cDNA was prepared as described in SuperScript III reverse transcriptase (Invitrogen™ SuperScript® III Reverse Transcriptase, Cat#18080093, ThermoFisher) manual with 1µg of RNA primed by random

hexanucleotides. Along with samples, template negative and reverse transcriptase negative controls were included to determine product specificity. Genes selected for validation studies were assayed as described in TaqMan® PreAmp Master Mix Kit user guide (Applied Biosystem). For each validation gene quantified with a Taqman® probe labelled with FAM, the expression values were normalised to the reference gene *Gapdh* assayed in the same well using the Taqman probe labelled with VIC. RNA was also extracted from 12.5, 14.5 and 18.5 dpc telencephalon and pooled before cDNA was prepared (as described previously). Expression of genes was determined using TaqMan® PreAmp Master Mix with gene specific Taqman probes labelled with FAM. The expression values were normalised to the reference gene *Tbp* which we confirmed was stably expressed across the chosen time points (data not shown). Taqman probes used in this study are listed in Table 2.6.

Table 2.6 Taqman assays probes.

Gene ID	Taqman assay ID	Label
Alx1	Mm00553295_m1	FAM
Arx	Mm00545903_m1	FAM
Asphd1	Mm01278674_m1	FAM
Aspm	Mm00486659_m1	FAM
Barx1	Mm01353100_m1	FAM
Cdkl5	Mm01156815_m1	FAM
Cxcr5	Mm00432086_m1	FAM
Ebf1	Mm00432948_m1	FAM
Egr3	Mm00516979_m1	FAM
Erbb4	Mm01256793_m1	FAM
Esrp1	Mm00839264_m1	FAM
Gabbr2	Mm01352554_m1	FAM
Gpr26	Mm01165717_m1	FAM
Kcna3	Mm00434599_s1	FAM
Lmo1	Mm00475438_m1	FAM
Myt1l	Mm00485408_m1	FAM
Nrp	Mm02391703_s1	FAM
Pitx2	Mm01316994_m1	FAM
Plcxd3	Mm01307828_m1	FAM
Sacs	Mm0131311_mH	FAM
Six2	Mm03003557_s1	FAM
Six6	Mm00488257_m1	FAM
Sorl1	Mm01169526_m1	FAM
Tenm1	Mm00600053_m1	FAM
Tnr	Mm00659075_m1	FAM
Twist1	Mm04208233_g1	FAM
Zfp536	Mm00552423_m1	FAM
Hprt	Mm03024075_m1	VIC
Gapdh	Mm99999915_g1	VIC
Tbp	Mm01277042_m1	VIC

2.4.4 Functional annotations

Statistical analysis of the enrichment of Gene Ontology (GO) categories was performed using EnrichR, a bioinformatics tool that retrieves molecular information from transcription factor databases and defines transcription factors statistically enriched from gene lists (Chen et al. 2013). To rank the enrichment results we used the score calculated by EnrichR using the P-value and Z-score. The top 10 results are reported with an adjusted p-value of $p<0.05$.

2.4.5 Pathway analysis

Ingenuity Pathway Analysis (Ingenuity Systems) was used to assess connectivity of deregulated proteins. The requirements for assessing protein-protein interconnectivity included direct interactions observed experimentally. The permitted interaction types were: protein-protein, protein-DNA, activation, inhibition, phosphorylation, and ubiquitination. ARX/Arx (Fulp et al. 2008, Quille et al. 2011) and TWIST1/Twist1 (Lee, M.P. et al. 2014) ChIP interaction data was manually superimposed onto this pathway.

2.4.6 Statistics

Statistical significance of the overlap between two groups of genes was calculated using exact hypergeometric probability (nemates.org/MA/progs/overlap_stats.html). Total genes, in this case, equalled 13821 genes with detected reads.

2.5   In-Situ hybridization analysis

2.5.1   Tissue embedding

Upon collection, embryos were fixed in RNase-free 4% paraformaldehyde in PBS overnight at 4°C, washed 3×10 min in PBS, cryoprotected in 30% sucrose in PBS until tissues sink to the bottom of the tube. Samples were embedded in OCT (TissueTEK OCT Compound, cryostat specimen matrix, Cat#IA018, ProSciTech) and stored at −80°C until sectioned. Tissue sections (10 µm) were prepared using a Leica CM1900 cryostat.

2.5.2   Riboprobe production

MGC fully Sequenced Mouse cDNA clones were obtained from Millennium Science for *Hdac4* (Clone ID: 6827645, Cat#MMM1013-202859554), *Cdkl5* (Clone ID: 4013904, Cat#MMM1013-202704375) and *Twist1* (Clone ID: 4935230, Cat#MMM4769-202766995). To linearize the cDNA 3 µg of the plasmid template was digested with specific restriction enzymes (Table 2.7) for 2 hours at the required incubation temperature and checked on a 1% agarose gel before continuing. In order to purify linearized plasmids, each digested sample was mixed with 1 volume of phenol/chloroform and centrifuged at 18,000 x *g* for 2 minutes. The top aqueous layer was removed and precipitated with 1/10 volume of 3M NaAc (pH 5.2) and 2 volumes of 100% EtOH at -20°C overnight. The sample was then centrifuged at 18,000 x *g* for 30 minutes and washed with 70% v/v EtOH before resuspension in ultrapure H_2O. DIG labelling RNA antisense probes were transcribed *in vitro* from digested plasmid templates using a DIG RNA labelling Kit (Cat#11175025910, Roche) and T7, Sp6 or T3 RNA polymerase (Table 2.7) (T7 and Sp6

provided in Kit, T3 RNA polymerase, Cat#110311630014, Sigma-Aldrich). Template was

eliminated with the addition of 2 µl of RNase-free DNase I to the reaction and incubated

for 30 minutes at 37°C. Removal of unincorporated DIG label and reaction components

and concentration of DIG-labelled probes was achieved by passing the dilute reaction

volume through a Quick Spin Columns (CHROMA SPIN-100+DEPC-H_2O Columns,

Cat#636089, Scientifix).   Columns were pre-centrifuged to remove all contents

immediately prior to loading probes onto the columns to prevent drying of the matrix.

Riboprobes were eluted off the column into a fresh Eppendorf tube at 700 x g for 5 minutes,

and integrity and size of labelled probes were confirmed on a 2% agarose gel before use.

Table 2.7 Antisense Riboprobe Production.

cDNA	Plasmid Backbone	Restriction Enzyme	RNA pol
Hdac4	pXY-Asc	*Nae*I	T3
Cdkl5	pCMV-SPORT6	*Sma*I	T7
Twist1	pCMV-SPORT6	*Sap*I	T7
ARX (Lee, K. et al.)	pGEMT	*Spe*I	T7

2.5.3   In Situ hybridization

All details of solutions used in 0 are listed in Table 2.8. Slides were defrosted and dried at room temperature for at least 1 hour. Probes were diluted in RNase-free H_2O to a total volume of 7.5 µl for each slide and denatured at 80°C for 2 minutes. Hybridization mix was added to each slide consisting of 7.5 µl of diluted probe and 69 µl of hybridization buffer (Table 2.8) covered with a coverslip and placed in a hybridization chamber filled with 1:1 formamide: ultrapure H_2O and incubated overnight at 65°C. Slides were washed for 15 minutes in preheated coplin jars filled with wash solution (Table 2.8) and placed in a waterbath at 65°C to allow coverslips to fall off. Washing with wash solution was repeated two additional times for 30 minutes each at 65°C. Subsequent washes with Maleic Acid Buffer + Tween-20 (MABT, Table 2.8) at room temperature for 30 minutes each was repeated two times. After blocking with 700 µl of blocking solution (Table 2.8) for 2 hours in a humidified chamber to limit fluid evaporation, 75 µl anti-DIG antibody mix was added to each slide (0.25 µl anti-DIG antibody (Cat#11093274910, Sigma) to 1ml of Blocking Solution) and left overnight at room temperature in a humified chamber. Slides were washed in MABT four times for 20 minutes each at room temperature. Followed by two washes in alkaline phosphatase staining buffer (APB, Table 2.8) for 10 minutes each. To complete the staining reaction 4.5 µl (450 µg) of NBT (Roche Nitrotetrazolium Blue Chloride solution (100mg/ml), Cat#11383213001, Sigma-Aldrich) and 3.2 µl (160 µg) of BCIP (Roche 5-Bromo-4-chloro-3-indolyl phosphate disodium salt (50 mg/ml), Cat#11383221001, Sigma-Aldrich) to 1ml of fresh APB and 95 µl of the staining mix added to each slide to be incubated in a clean humified chamber, sealed and in the dark for

> 4 hours. Slides were washed three times for 5 minutes each in PBS in the dark and fixed

with 4% PFA to inactivate the alkaline phosphatase enzyme.

Table 2.8 In Situ Hyribisation Solutions.

10x Salt Buffer	2M NaCl, 100 mM Tris-HCL pH 7.5, 50 mM sodium phosphate monobasic dehydrate, 50mM sodium phosphate dibasic, 50 mM EDTA
Hybridization Buffer	50% formamide, 1% dextran sulphate, 1x Salt Buffer, 2x Denharts and 1 mg/ml tRNA
20x SSC	3M NaCl, 300mM Sodium Citrate, pH 7.0
Wash Solution	50% Formamide, 1x SSC diluted in ultrapure H_2O
Maleic Acid Buffer	100mM Maleic acid, 100mM NaCl, pH to 7.5 with NaOH
MABT	Maleic Acid Buffer, 0.1% Tween-20
Blocking Solution	20% Heat inactivated sheep serum, 2% Blocking reagent (Roche Blocking Reagent, Cat#11096176001, Sigma) diluted in MABT
APB	100mM NaCl, 50mM $MgCl_2$, 100mM Tris pH 9.5

2.6   *In Vitro* analysis

2.6.1   Plasmid generation

All plasmids readily available within the Shoubridge Laboratory (Table 2.9). Plasmids were initially prepared on a small scale (Wizard Plus SV Minipreps DNA Purification System, Cat#A1330, Promega) and sequenced confirmed (as per section 2.2.4), using primers listed in Table 2.10. High volume purity stocks were generated from the original single colony using Qiagen Plasmid Midi Kit (Cat#12143).

Table 2.9 Plasmid Information.

Construct	Vector	Tag	Ref
empty	pCMV-Myc	N-terminal Myc tag	(Shoubridge et al. 2012)
ARX WT	pCMV-Myc	N-terminal Myc tag	(Shoubridge et al. 2007)
ARX PA1 21A	pCMV-Myc	N-terminal Myc tag	
ARX PA1 23A	pCMV-Myc	N-terminal Myc tag	(Shoubridge et al. 2007)
ARX PA1 27A	pCMV-Myc	N-terminal Myc tag	(Fullston et al. 2011)
ARX PA2 20A	pCMV-Myc	N-terminal Myc tag	(Shoubridge et al. 2007)
ARX PA2 21A	pCMV-Myc	N-terminal Myc tag	(Fullston et al. 2011)
ARX PA2 23A	pCMV-Myc	N-terminal Myc tag	(Fullston et al. 2011)
empty	pcDNA3.1/nV5-DEST	N-terminal V5 tag	
UBQLN4 WT	pcDNA3.1/nV5-DEST	N-terminal V5 tag	
3xLMO1 TFBS	pGL4.13[luc2/SV40]	Luciferase reporter	(Shoubridge et al. 2012)
	pGL4.74[hRluc/TK]	Renilla reporter	(Shoubridge et al. 2012)

Table 2.10 Plasmid Sequencing Primers.

Use	Name	Direction	Species	Primer Sequence (5'-3')	Length
Vector Primer	pCMV-myc-F	Forward	Vector	GATCCGGTACTAGAGGAACTGAAAAAC	27
Vector Primer	pCMV-myc R	Reverse	Vector	GTTGTGGTTTGTCCAAACTCATCAATG	27
Vector Primer	pDEST-F	Forward	Vector	TAATACGACTCACTATAGGG	20
Vector Primer	pDEST-R	Reverse	Vector	GGAAAGGACAGTGGGAGTGG	20
ARX ORF Plasmid Sequencing	ARXc ex1/2 F	Forward	homo sapiens	TGCAAGGCTCCCCTAAGAGCA	21
	ARXc ex2/3 F	Forward	homo sapiens	CGTCTTCACCAGGGAGGAACT	21
	ARXc ex3/4 F	Forward	homo sapiens	CCCAGTCCAGGTCTGGTTCCA	21
	ARXcR6 (Exon 2)	Reverse	homo sapiens	CGCTGCTGCTCTTAGGGGAGC	21
	ARXc ex2-R (Exon 2)	Reverse	homo sapiens	GTACGACTTGCTGCGGCTGAT	21
	ARXE2P2F5	Forward	homo sapiens	AACTGCTGGAGGACGACGAGG	21
	ARXE2R7	Reverse	homo sapiens	CAGCTCCTCCTTGGGTGACA	20
	ARXcF2 (Exon 4)	Forward	homo sapiens	CGCTCGACTCCGCTTGGACTG	21
	ARXc-R1	Reverse	homo sapiens	CAGTCCAAGCGGAGTCGAGCG	21
A1UP ORF Plasmid Sequencing	A1UpE2F1	Forward	homo sapiens	CCGGAGGTTTAAGGCTCAG	19
	A1UpE2R1	Reverse	homo sapiens	CTGAGCCTTAAACCTCCGG	19
	A1UpE5F2	Forward	homo sapiens	GAGATGATGCGGAACCAGG	19
	A1UpE5R2	Reverse	homo sapiens	CCTGGTTCCGCATCATCTC	19
	A1UpE6F4	Forward	homo sapiens	GATCAATGCGGCTAGCCTG	19
	A1UpE6R4	Reverse	homo sapiens	CAGGCTAGCCGCATTGATC	19
	A1UpE9F3	Forward	homo sapiens	GGCATTGCTGCAGATCCAG	19
	A1UpE9R3	Reverse	homo sapiens	CTGGATGTGCAGCAATGCC	19

2.6.2 Maintaining human embryonic kidney 293T cell line

Human Embryonic Kidney 293T cell line (HEK293T) was cultured in Dulbecco's Modified Eagle's Medium (Gibco™ DMEM, high glucose, pyruvate, no glutamine, Cat #10313021 , ThermoFisher), supplemented with 10% Fetal Calf Serum, 100U/ml sodium penicillin, 100ug/ml strestreptomycin sulfate (Gibco™ Penicillin-Streptomycin (10,000U/ml), Cat#15140122, ThermoFisher) and 2mM L-glutamine (Gibco™ L-Glutamine (200mM), Cat#25030081, ThermoFisher). Cells were incubated in 5% CO_2 at 37°C. Cells were passaged or collected at 90% confluency. Cell were treated with 1 ml of Trypsin (Gibco™ Trypsin-EDTA (0.05%), phenol red, Cat#25300054, ThermoFisher) for 5 minutes to allow attached cells to detach from the flask surface, and then resuspended in 9 ml of DMEM. Cell are counted by staining with Trypan Blue (Gibco™ Trypan Blue Solution, 0.4%, Cat#15250061, ThermoFisher) using a haemocytometer to determine cell number prior to plating.

2.6.3 Transient transfection

One day before transfection, appropriate number of cells (see Table 2.11) were counted and plated in a 6 or 12 well plate (Corning® Costar® cell culture 6/12 well, flat bottom plates, Cat#CLS3516-50EA/CLS3513-50EA, Sigma-Aldrich) in Dulbecco's Modified Eagle's Medium (Gibco™ DMEM, high glucose, pyruvate, no glutamine, Cat #10313021 , ThermoFisher) supplemented with 10% Fetal Cal Serum (FCS), and 2mM L-glutamine (Gibco™ L-Glutamine (200mM), Cat#25030081, ThermoFisher). On the day of transfection, two separate solutions were prepared: (1) Plasmid DNA (variable amount)

was diluted in 200 µl of DMEM with 2mM L-glutamine; and (2) 4 µl of Lipofectamine™ (Invitrogen™ Lipofectamine® 2000 Transfection Reagent, Cat#11668019, ThermoFisher) was diluted in 200 µl of DNA with 2mM L-glutamine. The two solutions were mixed together and incubated for 20 minutes at room temperature to allow for DNA-lipid complexes to form. Media was removed from cells and replaced with 600 µl of DMEM with 2mM L-glutamine, prior to 400 µl of the plasmid: Lipofectamine mixture aliquoted into appropriate wells. The plates were kept in 5% CO_2 in 37°C incubator for 3 hours. Each well was then supplemented up with 1 ml 20% FCS and 2mM L-glutamine DMEM. Cells were collected at time points indicated in each study.

Table 2.11 Seeding Density for Transient Transfection.

Flask/Plate	Area	Cell Density	Vol. of Medium (ml)
6 well plate	$9cm^2$	8×10^5	2 ml
12 well plate	$4cm^2$	4×10^5	2 ml

2.6.4 Immunofluorescence

Cells were fixed using 4% formaldehyde for 1 hour at room temperature. Cells were permeabilized with a solution of PBS containing 0.2% Triton (Triton™ X-100, Cat#X100, Sigma-Aldrich) for 5 minutes and washed with 3 x PBS to remove detergent. Cells were blocked with 5% skim milk. Primary and fluorescently tagged secondary antibodies were diluted in PBST containing 3% normal horse serum (NHS). Primary antibodies were incubated overnight at 4°C. Secondary antibodies were incubated for 1 hour at room temperature in the dark. All antibody dilutions are listed in Table 2.12 and Table 2.13. Prolong Diamond Antifade Mountant with DAPI (Molecular Probes™ ProLong®

Diamond Antifade Mountant with DAPI, Cat#P36962, ThermoFisher Scientific) as per manufacturer protocols was used to stain the cell Nuclei and to mount coverslips.

Table 2.12 Primary Antibody List.

Protein	Species	Affinity	Cat #	Company	Dilution
ARX (N-15)	Goat	Polyclonal	SC-48845	Santa Cruz	1:1000
ARX	Mouse	Monoclonal	-	Neurogenetics	1-2 µg/ml final
ARX	Sheep	Polyclonal	-	Neurogenetics	1-2 µg/ml final
ARX	Rabbit	Polyclonal	-	Gifted by Kitamura	1:500
Myc-HRP	Mouse	Monoclonal	46-0709	Invitrogen	1:5000
Myc (9E10)	Mouse	Monoclonal	SC-40	Santa Cruz	1:1000
V5	Rabbit	Polyclonal	A190-120A	Bethyl	1:5000
V5	Mouse	Monoclonal	46-0705	Invitrogen	1:5000
V5-HRP	Mouse	Monoclonal	46-0708	Invitrogen	1:5000

Table 2.13 Secondary Antibody List.

Host	Target	Conjugate	Clonality	Conc^n	Cat #	Company
Donkey	α Mouse	Alexa 488	Polyclonal	2mg/mL	A21202	Invitrogen
Goat	α Mouse	FITC	Monoclonal	-	F0479	Dako
Goat	α Rabbit	Cy3	Polyclonal		111-165-144	Jackson Lab
Goat	α Mouse	HRP	Polyclonal	1 g/L	P0447	Dako
Goat	α Rabbit	HRP	Polyclonal	0.25 g/L	P0448	Dako
Rabbit	α Goat	HRP	Polyclonal	0.65 g/L	P0160	Dako
Donkey	α Sheep/Goat	HRP	Polyclonal		AB321P	Millipore

2.6.5   Microscopy

All immunofluorescence images were captured using Zeiss Axio Imager.M2 microscope equipped with Axio Vision 5.1 Software. Immunofluorescence images were acquired by Zeiss AxioCam mRM black and white camera. All captured images were processed by the Axio Vision 5.1 Software and Image J for analysis.

2.7 Gene expression analysis

2.7.1 RNA extraction

Total RNA from cell pellets were extracted using a combined method with TRIzol®
Reagent (Cat#1559026, ThermoFisher), RNeasy Mini Kit (Cat#74104, QIAGEN) and
RNase-Free DNase set (Cat#79254, QIAGEN) following the protocol described in 2.4.1.
Depending on the size of the cell pellets collected RNA was eluted in 30 µl - 50 µl of
RNase-Free H_2O.

2.7.2 Reverse transcription cDNA synthesis

cDNA was prepared as described in section 2.4.3 with 2 µg of RNA per sample. For
negative controls, SuperScript™ III RT was replaced with 1 µl of H_2O. The newly
synthesised cDNA was diluted by adding 20 µl of H_2O and stored at -20°C for further use.
The efficiency of the RT-PCR was determined by PCR using primers specific to
ubiquitously expressed housekeeping genes Esterase D (human samples) or Beta Actin
(mouse samples) (Primers listed in Table 2.14) (results not shown).

Table 2.14 House Keeper Primer Sets.

Name	Species	Primer Sequence (5'-3')	Length	PCR annealing Tm (°C)	PCR Product (bp)
Esterase D-Forward	homo sapiens	GGAGCTTCCCCAACTCATAAATGCC	25	60	453
Esterase D-Reverse	homo sapiens	GCATGATGTCTGATGTGGTCAGTAA	25		
Beta Actin–Forward	mus musculus	GATATCGCTGCGCTGGTCGTC	21	60	177
Beta Actin–Reverse	mus musculus	TCTCTTGCTCTGGGCCTCGTCAC	23		

2.7.3 Polymerase chain reaction (PCR)

cDNA was amplified with 1 µl (5 U) of Taq DNA Polymerase (Cat# 11146173001, Roche), 1x PCR buffer with $MgCl_2$ (stock at 10x, Cat# 11699121001, Roche), specific single-stranded DNA primers and H_2O to 50 µl volume. The PCR cycle condition was as follows: initial denaturation at 94°C for 5 minutes, then 35 cycles of denaturation at 94°C for 30 seconds, annealing for 30 seconds at 60°C (or specific annealing Tm of each primer pair), extension at 72°C for 30 seconds, followed by final extension at 72°C for 10 minutes.

2.7.4 Gel electrophoresis

PCR products were visualised on a 1-2% (w/v) agarose gel in TBE buffer (45mM Tris-borate, 1mM EDTA, pH 8.5) with the addition of Ethidium Bromide (0.2 µg/ml) for 40 minutes at 100V alongside pUC19/*Hpa*II or 1kb+ molecular weight markers. Samples were premixed with loading buffer (20% (v/v) Ficoll 400, 0.1M Na_2EDTA, 0.25% (w/v) bromophenol blue, 1.0% (w.v) SDS) before being loaded onto the gel. Gels were then visualised under UV (Sygene INGENIUS LHR: Gel Documentation System, LabGear Australia)

2.7.5 Quantitative real-time PCR (RT-PCR)

Pre-designed TaqMan® Gene Expression Assays were selected from ThermoFisher. Reactions were set up on 96-well plate. Each well contains 2 µl of cDNA template (of a 1ng/ µl to 50ng/ µl stock), 1 µl of the 20x TaqMan® Gene Expression Assay (FAM™ dye-labelled MGB probe) and 1 µl of 20xTaqMan® Endogenous Control Assay (VIC® dye-

labelled MGB probe), 10 µl of the 2x TaqMan® Gene Expression Master Mix (Applied

Biosystems™, Cat#4369016, ThermoFisher Scientific) and RNase-free H_2O. Reactions

were run on the Applied Biosystems StepOnePlus™ Real-Time PCR System using a

standard run with the following conditions: activation at 50°C for 2 minutes followed by

95°C incubation for 10 minutes, 40 cycles of denaturation at 95°C for 15 seconds and

extension at 60°C for 1 minute. Signal emitted from the dye reporter was recorded at the

end of each cycle. All sample were analysis in triplicate and efficiency of the assay was

determine by the amplication of the standard curve of diluted control cDNA. Expression

values were calculated using comparative ΔCt method (Bookout and Mangelsdorf 2003).

Refer to Table 2.6 all probes used in thesis.

2.8 Protein analysis

2.8.1 Protein extraction

RIPA activated buffer (65.3 mM Tris, 150 mM NaCl, 1% (v/v) Nonidet P-40) was made up by adding 80 µl of 25 x protein inhibitor cocktail (Protease Inhibitor Cocktail, Cat#P8340, Sigma), 10 µl of 200mM Na_2VO_4, 200mM NaF and 200mM phenylmethylsulfonyl fluoride (PMSF) to 2 ml of RIPA buffer. Cell pellets were kept frozen on dry ice until they were homogenised with 100-200 µl of RIPA activated buffer to minimise protein degradation. The tube containing the homogenised cell lysate was placed horizontally in normal ice with shaking for 15 minutes. To shear DNA, the lysates were passed through an 18G needle and 1 ml syringe 10 times. Lysates were clarified by centrifugation (15 minutes, 13,000 x g at 4°C). The supernatant containing the soluble proteins was transferred to a clean 1.5 ml screw cap tube and stored at -80°C until further use.

2.8.2 Protein quantification

Diluted sample (1/10 and 1/20 with H_2O) of each protein aliquot were made from the original stock. In a 98 well plate (Corning® 96 well plates PVC flat bottom, Cat#CLS2595-100EA, Sigma-Aldrich), the diluted samples were aliquoted in triplicates of 10 µl per well. Pre-made Bovine Serum Albumin (BSA) standards were used to construct a standard curve at 0, 0.2, 0.4, 0.6, 0.8 and 1.0 mg/ml. 200 µl of a 1 in 4 dilution of Bradford reagent in H_2O was added to each well. The OD of each well was read immediately at 570nm using automated plate reader (Cat#MR5000, Dynateach). Relative starting protein was

established for each protein dilution utilising the pre-determined BSA standard curve. The

final protein concentration was determined by averaging the values of the two dilutions.

2.8.3   Running SDS-PAGE

Protein samples (usually 10 – 20 µg) were prepared in 1 x DTT reducing agent (10x 1M

stock), 1 x loading buffer (Novex™ NuPAGE® LDS Sample Buffer (4x), Cat#NP0008,

Thermofisher) and H_2O to 16 µl volume. The protein was heat denatured at 95°C for 5

minutes. To detect protein smaller than 100 kDa, samples were loaded onto 4-12% Bis-

Tris Protein Gel 1.0 mm (Invitrogen™ NuPAGE™ Novex™ 4-12% Bis-Tris Protein

Gels, 1.0 mm, Cat#NP0321BOX, Thermofisher) as part of the Xcell Surelock® Mini-Cell

(Novex™ Cat#E10002, Thermofisher). The inner chamber was filled with cold 1x MOPS

running buffer (Novex™ NuPAGE® MOPS SDS Running Buffer (20X), Cat#NP0001,

Thermofisher) and 500 µl of NuPAGE® Antioxidant (Novex™ NuPAGE Antioxidant,

Cat#NP0005, Thermofisher). The outer chamber was filled with only the cold 1x MOPS

running buffer.

2.8.4   Membrane transfer

The separated proteins in the gel were transferred onto Pure Nitrocellulose Blotting

Membrane (Pall Laboratory BioTrace™ NT Nitrocellulose Transfer Membrane, Cat#732-

3031, VWR) using the XCell II™ Blot Module (Novex™, Cat#EL9051, Thermofisher

Scientific). The inner chamber was filled with freshly made cold transfer buffer (1xTowbin

Buffer (0.025M Tris, 0.192M Glycine) and 20% Methanol diluted in deionized H_2O); the

outer chamber was filled with ice cold H₂O. Proteins were transferred at 30 volts powered by PowerPac HV High-Voltage Power Supply (Cat#1645056, Bio-Rad) for 1 hour.

2.8.5 Immunoblot

Membranes were rinsed with ultrapure H_2O to remove residual methanol prior to blocking with a solution of Tris-buffered saline with Tween-20 (TBST) and 5% skim milk powder and incubated with shaking for 1 and a half hours at room temperature. Primary antibody (Ab) was diluted in a solution of TBST, and 1% skim milk. The membrane was incubated with primary Ab for 4 hours at room temperature or at 4°C overnight in the fridge on the shaker. The membrane was washed 3 x 10 minutes with shaking using TBST before incubated with Horse Radish Peroxidase (HRP)-conjugated secondary antibody diluted in a solution of TBST and 1% skim milk. The membrane was washed 3 x 5 minutes with shaking using TBST to remove all unbound secondary antibody.

Table 2.12 and Table 2.13 describes all primary and secondary antibodies used in this thesis. Signal was detected via chemiluminescence (Amersham ECL Western Blotting Detection Reagent, Cat#RPN2106, GE Healthcare Life Science) and exposure of the membrane to X-ray sensitive film (CURIX ORTHO HT-G AGFA Medical X-Ray Film (18x24cm), Cat#E7FYQ, Total Medical Imaging Solutions or Amersham Hyperfilm ECL, Cat#28906839, GE Healthcare Life Science).

2.8.6 Co-immunoprecipitation (Co-IP)

Cells transfected with Myc-ARX, both with and without V5-UBQLN4 were harvested at 24 hours post-transfection and cell lysates prepared using lysis buffer (120mM NaCl,

50mM Tris-HCl (pH 8.0), 0.5% NP-40 (v/v), 1× protease inhibitor cocktail (Cat#P8340, Sigma-Aldrich), 1mM Na3VO4, 1mM NaF, 1mM PMSF). Lysates were clarified by centrifugation (15 minutes, 13,000 x g at 4°C). Aliquots of extracts were immunoprecipitated (IP) overnight at 4°C. Protein-A sepharose (Cat#17-0780-01, VWR) was pre-treated with non-transfected HEK293T cell lysate to reduce 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 (250 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; V5-UBQLN4 was IP with 0.5 µg of rabbit anti-V5 antibody (Cat#A190-120A, Bethyl Laboratories). IP proteins were analysed for the presence of V5-UBQLN4 and Myc-ARX by immunoblotting using mouse anti-V5 HRP conjugated antibody, and mouse anti-Myc HRP conjugated antibody respectively (listed in Table 2.13).

2.8.7 In silico protein modelling

The software package I-TASSER suite was used (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) to model protein structures. The predicted secondary structures are calculated from the target amino acid sequence in which the prediction contain three states: alpha helix, beta strand, and coil. The secondary structure with the highest confidence score is provided with the B-factor profile. B-factor is associated with the inherent thermal mobility of local atoms and residues, which is essential for proteins to fold and function in the physiological environment. The predicted secondary structure can be used for estimating the number of secondary structure elements and the tertiary structure class of the query protein.

2.9   Biochemical assays

2.9.1   Pulse-chase

For pulse-chase experiments, cells were transfected with 0.5 µg of cMyc-ARX WT or cMyc-ARX PA123A plasmid. The day following transfection, cells were rinsed once with phosphate-buffered saline (PBS) and incubated for 15 minutes in Met/Cys-free DMEM (Cat#21013024, Thermofisher Scientific) and then radiolabeled in the same medium containing 50 µCi/ml of EasyTag™ EPRESS35S Protein Labeling Mix (Cat#NEG772014MC, PerkinElmer) for 1 hour. Cells were rinsed with PBS and incubated for chase intervals of 0, 1 and 3 hours in non-radioactive complete DMEM (Cat#11960051, Thermofisher Scientific) supplemented with 10% fetal bovine serum. At the end of each time point the cells were rinsed in cold PBS 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 (Cat#P8340, Sigma-Aldrich), 1 mM Na3VO4, 1 mM NaF, 1 mM PMSF). Lysates were passed through a 25G needle and clarified by centrifugation (15 minutes, 13,000 x *g* at 4°C). Aliquots of extracts were immunoprecipitated (IP) overnight at 4°C with anti-cMyc (c-Myc Antibody (9E10), Cat#sc-40, Santa Cruz Biotechnology). Protein-A sepharose (Cat#P3391, Sigma-Aldrich) was pre-treated with untransfected HEK293T cell lysate to reduce non-specific binding of cell proteins (Mattiske et al. 2013). The IP reactions were incubated with the pre-treated protein-A sepharose for 1 hour at 4°C. Samples were then loaded on a MultiScreenHTS-DV Plate (Cat#MSDVN6B50, Millipore) and washed under suction for removal of non-specifically bound proteins with 10 changes of high stringency wash buffer (250 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. The plastic webbing on the MultiScreenHTS-DV Plate was removed and placed in a heated drying rack until samples/filter paper was dry. 50μl of optiphase supermix was added to each well containing a sample, and 35S level was measured using beta scintillation plate counter.

2.9.2   Luciferase assay

We used the orthologous sequence upstream of the human LMO1 gene to generate luciferase reporter constructs containing specific TFBSs for ARX. As previously described in (Shoubridge et al. 2012) an oligo containing three TFBS sequences in tandem 5′-(gaattgattTAATTAacaggggaa) ×3-3′ and *Bam*H1 and *Bgl*II sites flanking the end of the oligo was directionally cloned upstream of either the SV40 promoter driving the luciferase reporter gene (*luc2*) in the pGL4.13[*luc2*/SV40] vector (Cat#E668A/#E5061, Promega). HEK293T cells (4×10^5 per well) were plated into 12-well plates, and 24 h later were transfected with 500 ng of luciferase reporter plasmid DNA, 10 ng of pGL4.74[*h*R*luc*/TK] plasmid DNA (Cat#E692A, Promega) and 500 ng of Myc-expression plasmid DNA containing full-length ARX-WT or homeodomain mutation sequence using Lipofectamine 2000 (Cat#11668019, Thermofisher Scientific). Cells were lysed 24 h posttransfection and the measurement of both Firefly and renilla luciferase was achieved using Dual-Glo Luciferase Assay System (Cat#E2920, Promega) on a Veritus Microplate Luminometer (Turner BioSystems, Sunnyvale, CA, USA). In four independent transfections, each sample was carried out in replicate, with triplicates of each replicate measured in the reporter assay. The Firefly luciferase activity was normalised according to the

corresponding renilla luciferase activity, and the ratio of luciferase activity was reported

relative to pCMV-Myc empty vector for each transfection.

Chapter Three:

An emerging female phenotype with loss of function mutations in the *Aristaless*- related homeodomain transcription factor.

3 An emerging female phenotype with loss of function mutations in the Aristaless-related homeodomain transcription factor.

Publications, Presentations and Published Abstracts from this work:

<u>**Accepted first author publication**</u>

- Mattiske, T., C. Moey, L. E. Vissers, N. Thorne, P. Georgeson, M. Bakshi and C. Shoubridge (2017). "An Emerging Female Phenotype with Loss-of-Function Mutations in the Aristaless-Related Homeodomain Transcription Factor ARX." Hum Mutat. (DOI: 10.1002/humu.23190)

In addition to the patient screening featured in this chapter, a number of other individuals/families were screened for *ARX* mutations. This data was combined with a larger study and resulted in the following publication, with my inclusion as a co-author for this work.

- Marques, I., M. J. Sa, G. Soares, C. Mota Mdo, C. Pinheiro, L. Aguiar, M. Amado, C. Soares, A. Calado, P. Dias, A. B. Sousa, A. M. Fortuna, R. Santos, K. B. Howell, M. M. Ryan, R. J. Leventer, R. Sachdev, R. Catford, K. Friend, T. R. Mattiske, C. Shoubridge and P. Jorge (2015). "Unraveling the pathogenesis of ARX polyalanine tract variants using a clinical and molecular interfacing approach." <u>Mol Genet Genomic Med</u> **3**(3): 203-214. (DOI: 10.1002/mgg3.133)

3.1 Abstract

The devastating clinical presentation of X-linked lissencephaly with abnormal genitalia (XLAG) is invariably caused by loss of function mutations in the *Aristaless*- related homeobox (*ARX)* gene. Mutations in this X-chromosome gene contribute to intellectual disability (ID) with co-morbidities including seizures and movement disorders such as dystonia in affected males. The detection of affected females with mutations in *ARX* is increasing. We present a family with multiple affected individuals, including two females. Two male siblings presenting with XLAG were deceased prior to full term gestation or within the first few weeks of life. Of the two female siblings, one presented with behavioural disturbances, mild ID, a seizure disorder and complete agenesis of the corpus callosum, similar to the mother's phenotype. A novel insertion mutation in Exon 2 of *ARX* was identified, c.982delCinsTTT predicted to cause a frameshift at p.(Q328Ffs*37). Our finding is consistent with loss-of-function mutations in *ARX* causing XLAG in hemizygous males and extends the findings of ID and seizures in heterozygous females. We review the reported phenotypes of females with mutations in *ARX* and highlight the importance of screening *ARX* in male and female patients with ID, seizures and in particular with complete agenesis of the corpus callosum.

3.2   Introduction

The *Aristaless*-related homeobox gene (*ARX*) (NM_139058.2) (MIM# 300382) (Shoubridge et al. 2010a), is critical for early development and formation of a normal brain (Kitamura et al. 2002, Ohira et al. 2002, Scheffer et al. 2002). This paired-type homeodomain transcription factor plays a vital role in telencephalic development specifically in tangential migration and differentiation of GABAergic and cholinergic neurons (Kitamura et al. 2002, Colombo et al. 2007, Friocourt et al. 2008, Colasante et al. 2009, Lee, K. et al. 2014). Mutations in *ARX* result in a range of phenotypes with intellectual disability (ID) as a consistent feature. Mutations leading to loss of function of the ARX protein typically lead to brain malformation phenotypes, including X-linked lissencephaly with abnormal genitalia (XLAG, also known as LISX2) (MIM# 300215) (Kitamura et al. 2002, Shoubridge et al. 2010a). XLAG is a developmental disorder characterised by structural brain anomalies leading to agyria (absent cerebral grooves brain) or pachygyria (reduced folds or grooves) and agenesis of the corpus callosum. In addition, patients commonly have early-onset intractable seizures, severe psychomotor retardation, and ambiguous genitalia (Dobyns et al. 1999, Kitamura et al. 2002). Males are severely affected and often die within the first days or months of life.

As the *ARX* gene is on the X-chromosome, mutations in this gene typically result in families with affected males across multiple generations transmitted via (usually) asymptomatic carrier females. However, there is an increasing prevalence of reported mutations in *ARX* that result in the severe phenotypic outcomes of XLAG in male patients that, with variable penetrance, affect females within the families resulting in a generally milder phenotype

than affected males (Kitamura et al. 2002, Scheffer et al. 2002, Stromme et al. 2002, Kato et al. 2004, Marsh et al. 2009, Eksioglu et al. 2011). Here we report a novel mutation in *ARX* in a family ascertained by a female proband displaying a phenotype of mild learning disabilities, seizure disorder and agenesis of the corpus callosum. As part of this work, we have reviewed reported phenotype of females with mutations in *ARX* and recommend screening of the *ARX* gene in female patients with suitable intellectual disability, seizure phenotypes and corpus callosum agenesis, particularly if there is evidence of X-linkage and no surviving males. *ARX* adds to a growing list of genes on the X chromosome including genes such as *USP9X, PHF6* and *IQSEC2* that have phenotypic effects in males and females that are distinct depending on the functional severity of the mutation (Zweier et al. 2013, Reijnders et al. 2016, Zerem et al. 2016).

3.3 Materials and Methods

3.3.1 Clinical description of patient and family

The proband presented to a Genetics Clinic at 10 years of age with learning difficulties, mild ID and seizures. The proband was born at term weighing 3.2kg, with no admission into the neonatal intensive care unit nor special-care nursery required. She was reported as sitting with support at 9 months of age with single words spoken at 9-10 months of age and walking at 22 months of age. Seizure onset was around 5 years of age with the first seizure classified as a prolonged generalised tonic-clonic which required intubation. At this time she was assessed and was able to draw a triangle with help, able to hop, dress herself and talk in simple sentences, knowing a few colours and numbers. No unusual grasping is reported. She was evaluated using the WISC-IV Australian at age 8 years and 10 months. She scored in the "Extremely Low" range for verbal comprehension, perceptual reasoning index, working memory index, processing speed index with a resulting full-scale IQ in the "Extremely Low" range. She was assessed as operating in the mild range of intellectual disability. She subsequently had a number of complex partial seizures, which were reasonably well controlled on Tegretol. Brain MRI revealed complete agenesis of the corpus callosum with mild dilated ventricles and colpocephaly. No lissencephaly, dysmorphic features or behaviour problems were reported.

The mother of the proband was aged 46 years old at the time of this report and presented with mild ID, seizures and mental health challenges. After admission to the public health hospital she was diagnosed with borderline personality disorder and major depressive

disorder however, no neuropsychiatric assessment is currently available. She was non-dysmorphic. History revealed that her first seizure was around 7-8 years of age and classified as complex partial seizures. Brain MRI scan done at the age of 38 showed complete agenesis of the corpus callosum, with no grey matter heterotopia. Small white matter lesions were noted which comprised of tiny FLAIR (fluid attenuated inversion recovery) hyperintensities involving the left centrum semiovale bilaterally and in the frontal region. The treating neurologist at the time felt these were specific white matter hyperintensities however the actual films are no longer available for review. In addition, EEG showed intermittent spike, and wave discharge maximal over left hemispheres, which were at a frequency of 2.5-3.5 Hz.

Two male offspring of II-4 were affected and died in early infancy, or were terminated during pregnancy. A maternal half-brother (III-1) of the proband was born at 36 weeks of gestation with an onset of seizures 20 minutes after birth. III-1 died at 26 days of age. Investigation of brain morphology identified agenesis of the corpus callosum, lissencephaly, grey matter heterotopias and bilateral optic nerve hypoplasia. Genitalia were ambiguous with labioscrotal folds, bilateral gonads and microphallus. Pelvis ultrasonography revealed the presence of a bicornuate uterus, while a male urethra was confirmed with a genitogram. Karyotype analysis showed a normal male chromosomal constitution (46,XY). Facial dysmorphic features included a large head, large anterior fontanels, low set ears, and abnormal nails. III-3 eventuated with a medical termination of pregnancy due to identification at 18 weeks gestation via ultrasonography of ventriculomegaly and abnormal genitalia. Brain malformations were confirmed at post

mortem showing lissencephaly, absent corpus callosum, wide-open sylvian fissure and dilated ventricles. Facial dysmorphic features included flattened facies, mild macrocephaly and wide open sutures. Abnormal genitalia consisted of rudimentary genitalia with a small phallus. Karyotype analysis established the presence of male chromosomes.

Another pregnancy, III-5 was terminated when chorionic villus sampling revealed 45XO after an ultrasound showed fetal hydrops. The remaining sibling (III-2) is a healthy female with normal intelligence. A maternal uncle's (II-1) death at one month of age was attributed to sudden infant death syndrome. The maternal grandmother (I-2) is reported to be healthy with no seizures.

3.3.2 Molecular analysis of *ARX* gene

The screening protocols were approved by the appropriate institution review boards, and informed consent was obtained from the parents of patients. Genomic DNA from the proband was extracted from whole blood using standard techniques. Each of the five exons of *ARX* was amplified by PCR using primers designed to amplify coding and flanking non-coding sequence. The exception to this was exon 2, for which four overlapping amplicons were used to achieve robust amplification of GC-rich regions coding for three polyalanine tracts. The PCR conditions and primer sequences are described in detail previously (Tan et al. 2013). Sequencing reactions were performed using ABI Big Dye terminator chemistry version 3.1 and purified products subjected to an automated capillary sequencing on ABI 3100 sequencer (Applied Biosystems, Foster City, CA, USA) and sequence was compared

to the *ARX* reference sequence (NM_139058.2) using SeqMan module of the Lasergene

DNA and protein analysis software package (DNAStar, Inc., Madison, WI).

3.4 Results

3.4.1 Molecular Analysis of the *ARX* variant

The diagnoses of intellectual disability and seizures in the female proband in conjunction with XLAG in her male siblings, from two different fathers, prompted evaluation of the *ARX* gene. Sequence analysis demonstrated a novel indel mutation, c.982delCinsTTT in exon 2 of the *ARX* gene, that is predicted to result in the creation of a premature stop codon, p.(Q328Ffs*37) (LOVD ID 0000128956) (http://www.lovd.nl/ARX). The child's mother presented with a similar phenotype and was confirmed to also be a symptomatic carrier of this novel *ARX* mutation. The father (II-5) and unaffected sister (III-2) do not carry the mutation (Figure 3.1). The *ARX* mutation was confirmed in III-4 in genomic DNA extracted from amniocytes. The amino acid affected by the mutation p.328Q is located at the start of the region containing the second nuclear localisation signal (NLS2) preceding the conserved paired-type homeodomain. The predicted stop codon arising from this variant occurs within 29 nucleotides of the exon 3-4 junction and is likely to escape nonsense-mediated decay (NMD). Due to restricted levels of *ARX* expression in the patient-derived materials we are unable to confirm if this truncated protein is produced. Despite this, even if the predicted C-terminal truncated protein was produced and subsisted at appreciable levels, the severe XLAG phenotype in affected male patients is expected with this variant resulting in a nonsense peptide being transcribed after residue p.328 and complete loss of the paired-type homeodomain residues (Figure 3.1C).

a)

I

II

III
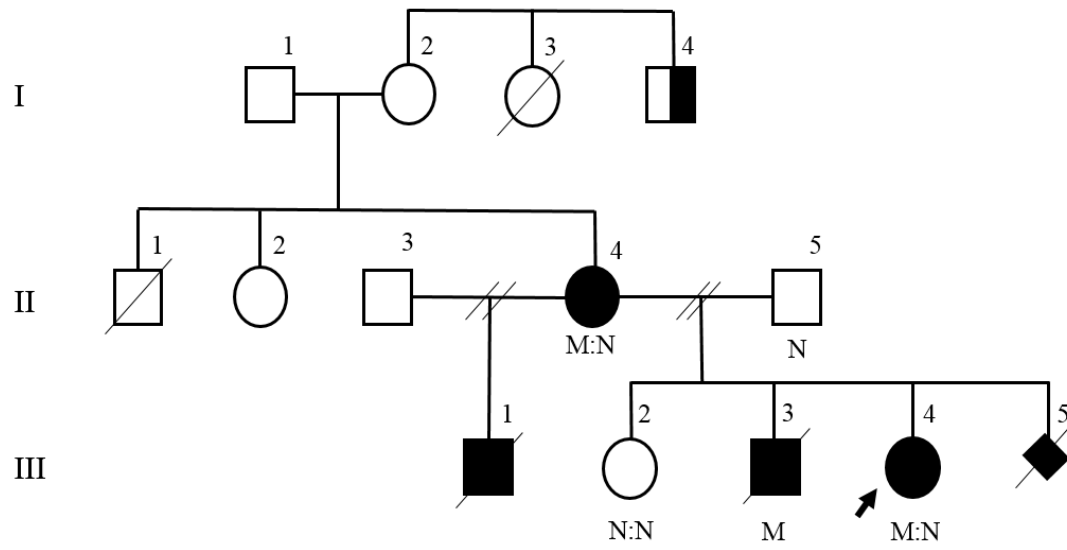
= Lissencephaly, hydrocephalus, profound
developmental delay and ambiguous genitals

= Seizures

= ID + Seizures

N = Normal Allele
M = Mutant Allele

b)

II-5

:T NGG AGGA GGG GT TGA TGGA G CGC AAA CA GA GGC GC TA C CSCCC CAC GTT

150   145   140   135   130   125   120   115   110   105

III-2

TC GGA GGA GGGGC TG NTGAAA CGC AAA CA GA GGC GCTA CC GCA CCA CGTT

III-3

Forward

T AG GAG GAG GGGC TGC TGAAA CGCAAA TTTA GA GGC GC TAC CGC ACCACG TT

Reverse

:TC GGA GGAGGGGC TKY T AAACGC AAA TTTA GAGG CGC TAC CGCA CC ACG TY

III-4

Forward

:TC GRAGG AG GGGC T GC TGAAA CGCAAA TTTA GAGG CKC TM CC GCM CCR YK TT

Reverse

TC GGA GGA GGGG STGY KGAAA MRC AAA TTTA GAGG CGC TAC CGC ACC AC GTT

II-4

Forward

TC GGA GGA GGGGC TGC TGAAA CGC AAA YWKA GGS GC TM CC SCR CCM CG TT

Reverse

:T SGGAGGA GGGG STGG TGAAA CGC AA WYA GA GGC GC TAC C GC ACC ACG TT

c)

Figure 3.1 Identification of a c.982delCinsTTT mutation resulting in (p.(Q328Ffs*37) in ARX

a) Pedigree of immediate family members tested. Open symbols represent normal individuals, filled black circles represent females with moderate intellectual disability and seizures, filled squares represent males with Lissencephaly, hydrocephalus, profound developmental delay and ambiguous genitalia. Individual generations are numbered with Roman numerals on the left-hand side of the pedigree. Individuals who were sequence confirmed to carry either the normal (N) or mutant (M) are shown b): DNA sequence electropherograms for the chrX: 25,031,130 (GRCh37/hg19 assembly) deletion of a C and insertion of TTT mutation in exon 2 of 5 of ARX reported in this study. A normal sequence was confirmed in unaffected father (II-5) and sister (III-2) with the normal sequence shown. The mutation in the heterozygous state is shown in both the affected proband (III-4) and mother (II-4). The dotted line highlights the disrupted heterozygous trace present in females caused by the insertion (solid black underline). Additional mutation sequence change was seen in the hemizygous state of an affected brother (III-3). c) The predicted functional consequence of a novel nonsense mutation in ARX. Schematic of the human ARX protein (top panel). Human ARX domains and regions are indicated above the schematic. Known functional domain are highlighted, octapeptide (OP) as horizontally hatched rectangle, nuclear localisation sequences (NLS) as three black rectangles, polyalanine tracts (PA) as four white rectangles, acidic domain as vertically hatched rectangle, homeodomain as crosshatched, and aristaless domain (OAR) as hatched. A schematic of the homeodomain and the flaking NLS domains (middle panel) with the amino acid sequence below (homeodomain sequence underlined). The amino acid change indicated by the first black arrow resulting in a nonsense peptide and a stop codon indicated by the second red arrow.

A methylation-specific PCR at the human FraxA and Androgen Receptor genes was performed on genomic DNA from blood as previously described (Plenge et al. 1997). X-inactivation studies for both the proband (III-3) and the healthy sister (III-2) detected no significant deviation.

As part of initial investigations, both II-4 and III-3 were identified to carry a duplication on chromosome 5 at 5p15.5 (Chr5:10,907,608-11,363,857). This duplication contains only part of one gene, *Catenin Delta 2* (*CTNND2*). This duplication was classified as a variant of unknown significance. This duplication was later also identified in the female sibling III-2, who is healthy and has no learning issues, indicating this copy number variant is unlikely to contribute to the phenotype of the proband, her mother or affected brothers.

3.5 Discussion

Here we report a family with a novel truncating mutation (c.982delCinsTTT/p.(Q328Ffs*37)) in *ARX*. The mutation is predicted to yield a non-functional protein product after p.328 due to the nonsense peptide prior to truncation of the protein at amino acid 364. This variant was not found in either the ExAC (http://exac.broadinstitute.org/) or 1000 genome project databases (http://www.internationalgenome.org/). Although this mutation may escape nonsense-mediated decay, the resulting protein will have functional loss of the homeodomain and Aristaless domains. The catastrophic phenotype of XLAG reported in 2 males in this family is consistent with the predicted disruption of the ARX homeodomain function. The female proband and mother have a milder phenotype consisting of ID, seizures and agenesis of the corpus callosum. The investigation of *ARX* as a cause of the phenotype in the female proband was due largely to the distinctive phenotypic presentation and early deaths of the affected male siblings.

Our report underscores that a carrier female phenotype is likely to be under-ascertained for *ARX*. This is supported by a recent examination of heterozygous females from families identified with *ARX* mutations (Marsh et al. 2009) and examples of gender bias (92% male: 8% female) in a recent cohort of patients referred for molecular analysis of *ARX* (Marques et al. 2015). Under-ascertainment may be occurring due to several contributing factors. Patients with phenotypes such as intellectual disability and infantile spasms have been typically screened for mutations in known genes such as *STXBP1*, *CDKL5*, *KCNQ2*, *GRIN2A*, *MAGI2* and *ARX*. However, in the case of *ARX*,

screening is commonly only considered in affected males. Moreover, as the majority of all mutations reported to date in *ARX* lead to an expansion of the first or second polyalanine tracts, both located in exon 2, *ARX* mutation analysis is routinely limited to screening exon 2 and often using size variant analysis approaches (Marques et al. 2015). The expanding use of next generation sequencing approaches on cohorts of individuals with intellectual disability and or seizure phenotypes are likely to provide a platform to potentially overcome some of these biases. However, even these types of approaches have constraints that need to be considered, particularly for genes with high GC content or near regions of low gene density, such as *ARX*. For example, sequence coverage across the *ARX* gene in 50 representative whole exome sequencing (WES) experiments undertaken at the Radboud University Medical Centre showed the median coverage for *ARX* was 40-fold less than the coverage of all ID genes; with the median percentages of *ARX* bases covered at 20x only being 73% compared to 97% for all ID genes (Figure 3.2). However, experience at this, and other centres indicates that the coverage differences although lower for *ARX* generally, may also depend on the region of the gene being considered (Figure 3.3). Exome sequencing using benchtop ion proton machines also result in poor coverage of the *ARX* gene, with mean coverage reported at 43% (Lacoste et al. 2016). It remains to be determined how the increasing use of whole genome sequencing approaches as well as improvements to WES technologies and analysis pipelines address some of these coverage issues.
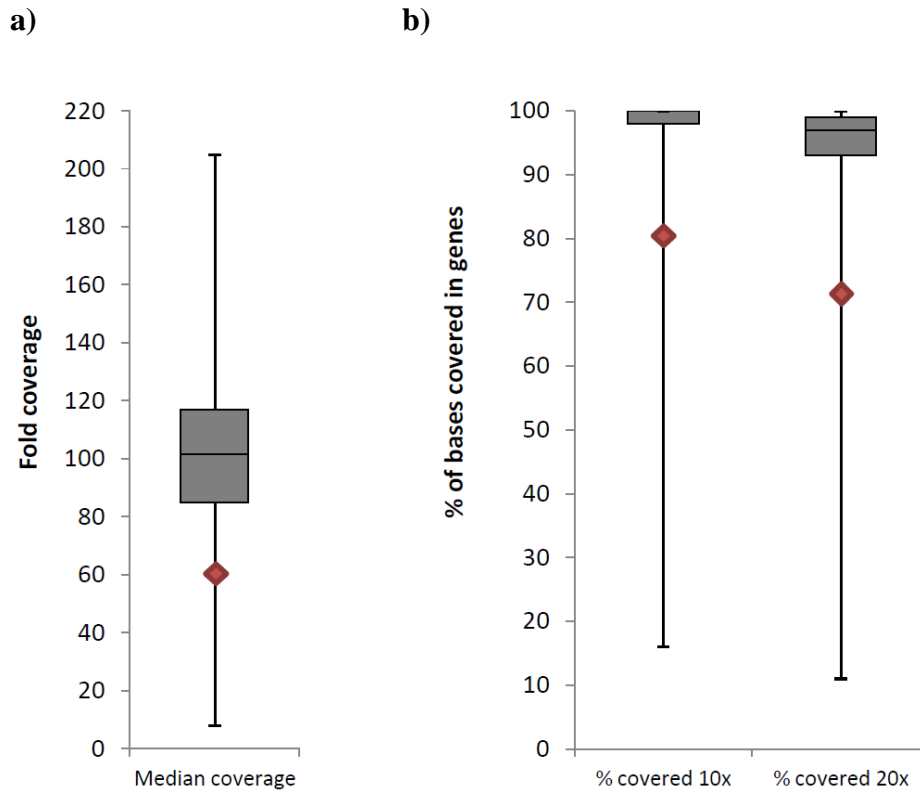
Figure 3.2 Radboud University Medical Center coverage analysis of 50 representative WES experiments of an ID gene panel consisting of 749 genes.
Boxplots representing a) the median coverage of all ID genes in 101-fold, whereas for ARX this is 60-fold. b) The median percentage of bases covered at 10x for all ID gene is 100%, whereas this is 80% of bases for ARX. Similarly, the average percentage of bases covered at 20x for all ID genes is 97%, whereas this is 73% for ARX. (Gene Panel DG2.5.x; https://www.radboudumc.nl/Informatievoorverwijzers/Genoomdiagnostiek/en/Pages/Intellectualdisability.aspx)

a)
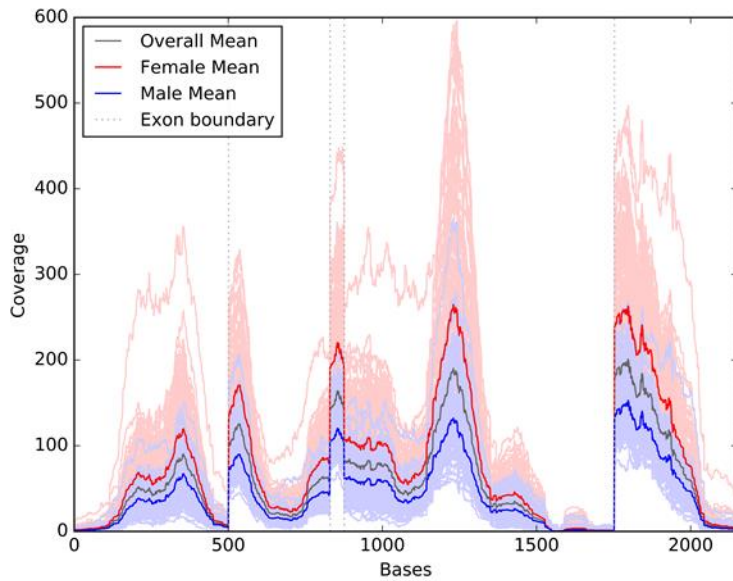
b)
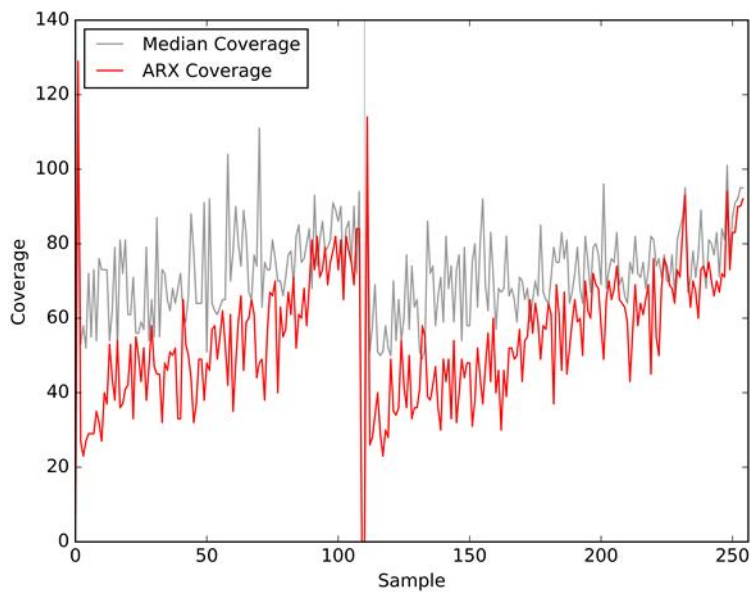
Figure 3.3 Melbourne Genomics Health Alliance cohort of 250 WES samples analysed with valid coverage
a) The coverage for each sample across the 5 ARX exons, as well as the mean coverage for males, females and the overall means. b) The median coverage of ARX compared to the mean median coverage of all genes. Females are on the left of the grey vertical line.

To date there have been both familial and *de novo* cases of affected females due to *ARX* mutations (Scheffer et al. 2002, Kato et al. 2004, Wallerstein et al. 2008, Marsh et al. 2009, Eksioglu et al. 2011, Bettella et al. 2013, Kwong et al. 2015) (Table 3.1). Penetrance is variable both within and across families, with 55% of carrier females in these families presenting with a phenotype of intellectual disability to some degree with and without seizures (Table 3.1). Intellectual disability and/or developmental delay are prominent features in affected females across all families. A seizure phenotype was reported in 64% of the females with ID which equates to ~ 35% of all females in these families (Table 3.1).

Table 3.1 Clinical Summary of Females with *ARX* mutations

	Familial	This report	*De novo*	Total
Females	25	2	4	31
Females with ID, with and without Seizures	11	2	4	17
ID or DD	11	2	4	17
Seizures	5	2	4	11
Other clinical features	number (symptomatic : non-symptomatic)			
MRI reported	10 (5:5)	2 (2:0)	3	15 .
Brain malformation	9 (4:5)	2 (2:0)	3	14 (9:5)
ACC	8 (4:4)	2 (2:0)	1	11 (7:4)
Other	1 (0:1)	0 .	2	3 (3:0)
Movement disorder	4 (3:1)	0 .	4	8 (7:1)
Psychiatric features	4 (3:1)	1 (1:0)	0	5 (4:1)
Behaviour disturbance	2 (2:0)	0 .	0	2 (2:0)

Similar to the novel case we report here, affected females have been reported in families in which the male proband presents with severe XLAG or seizure phenotypes (Scheffer et al. 2002, Kato et al. 2004, Marsh et al. 2009, Eksioglu et al. 2011). In these familial cases ascertained by the male proband, the mutations are either missense mutations of residues in the homeodomain or nonsense/ deletion mutations resulting in a loss of function of the ARX homeodomain and/or aristaless domain activity (Table 3.2). Similarly, a smaller number of *de novo* cases also result in truncation and loss of *ARX* function (Wallerstein et al. 2008, Marsh et al. 2009, Bettella et al. 2013, Kwong et al. 2015). Across these cases, there is a consistent phenotype of intellectual disability and/or developmental delay, infantile seizures and hypotonia/ dystonia/ ataxia (Table 3.3). The type and location of mutations in affected females are restricted compared to those contributing to affected males (Figure 3.4). Brain MRI imaging is reported in approximately 35% of all females in these cases/ families, including the two females from this current report. Interestingly, of those individuals screened, 73% had features consistent with agenesis of the corpus callosum (ACC), but only 64% of these patients were those classed as symptomatic based on intellectual disability and/or developmental delay and seizure phenotypes. Hence, there are a number of asymptomatic carrier females that do not display these key clinical features but still have the brain abnormality of ACC. There is also variable penetrance of both psychiatric and behavioural phenotypes across the symptomatic and asymptomatic females within these families. Movement disorders including ataxia and hypotonia are noted, particularly prevalent in the cases of *de novo* mutations (Table 3.3).

Table 3.2 Clinical features of females in familial cases of *ARX* mutations

Mutation cDNA / protein AA	Exon / Domain	Male Phenotype	Relationship to male proband	ID/DD	Seizures	Brain malformation	Additional Features	References
					Truncation mutations			
Exon1_2del	1, 2	XLAG	Mother	N	onset 12y (GTCS)			(Kitamura, et al., 2002), (Marsh, et al., 2009)
Exon2_5del p.G66_C562del	2–5 HD + OAR	XLAG	Mother	N				(Kato, et al., 2004)
			Sister	ID		ACC + CVH	ADD	
c.232G>T p.E78X	2 HD + OAR	XLAG	Mother	N		ACC		(Kato, et al., 2004)
			Sister	N		ACC	Duane anomaly	
			Mother	N		ACC-p		(Kato, et al., 2004)
			Aunt	Mod ID +DD	onset 1y (GTCS)	ACC		
c.617delG p.G206Afs*119	2 HD + OAR	XLAG-HYD	Mother	N				(Kato, et al., 2004)
			Sister	ID	Unilateral ~ 9 weeks	ACC-p	hypotonia	
c.982delCinsTTT p.Q328Ffs*37	2 HD + OAR	XLAG	Proband	Mild ID	onset 5y	ACC		this report
			Mother	Mild ID	onset 7-8y	ACC	Depression	
c.1471_1472insC p.L491Pfs*41	5 OAR	OS, AG, ID	Mother	ID			Anxiety, depression & personality disorder	(Eksioglu et al., 2011)
			Mat aunt 1	N			Depression, Bicornuate uterus	
			Mat aunt 2	ID			Schizophrenia. Small Genitals & small bladder with pockets	
			Mat half aunt 1	ID +DD	Generalized		Enlarged Clitoris	
			Mat half aunt 2	N				
			Mat grandmother	MD			Anxiety & depression	
					Missense mutations in the homeodomain			
c.998C>A p.T333N	2 HD	ACC/ AG	Mother	N		ACC-p		(Kato, et al., 2004)
			Cousin	Mod ID				
			Aunt	N				
			Aunt	Sev ID	onset 3mo	ACC-p	Spasticity, scoliosis, contractures	
c.1058C>T p.P353L	2 HD	XMESID	Mat grandmother	N		small vessel ischemic changes	Progressive spastic ataxia	(Stromme, et al., 2002) , (Scheffer, et al., 2002)
			Mat Aunt	ID			subtle hyperreflexia	
			Mother	N				
c.1135C>A p.R379S	4 HD	ISSX	Mother	N				(Marsh, et al., 2009)
			Aunt	N				
			Cousin	ID	onset 5y (absence)	N	PDD	

ACC = agenesis of the corpus callosum; ACC-p = partial ACC; / = not reported; ADD = attention deficit disorder; AG= ambiguous genitalia; CVH = cerebellar vermis hypoplasia; DD = developmental delay; HD = Homeodomain; ID = Intellectual Disability; ISSX = X-linked infantile spasm syndrome; Mat = Maternal; MD= Mild Delay; mo = months; Mod ID = Moderate ID; N=Normal; OAR = Aristaless; OS = Ohtahara Syndrome; PDD = Pervasive developmental disorder; Sev ID = Severe ID; y=years; XMESID = X-linked myoclonic epilepsy with generalized spasticity and ID. Nucleotide numbering reflects cDNA numbering with +1 corresponding to the A of the ATG translation initiation codon in the reference sequence for ARX gene [GenBank: NM_139058.2]

Table 3.3 Clinical features of females with *de novo* mutations in *ARX*

Mutation cDNA / protein AA	Exon / Domain	Mutation type	Development	ID	Seizures	Brain Malformation	Muscle Phenotype	Additional Features	References
c.869C>A p.S290X	2 HD + OAR	NS	Sev DD		IEE onset 1m	microcephaly	dystonia	UMN Signs, CVI	(Kwong, et al., 2015)
c.1459delA p.T487Qfs*5	5 OAR	Del	GPD	ID	EIEE	N	significantly ataxic	Divergent Strabismus. Congenital hip dislocation	(Bettella, et al., 2013)
c.1465delG p.A489Pfs*3	5 OAR	Del	DD	ID	ISSX (WS)	Cysts	hypotonia + torticollis	Poor visual tracking. Hydronephrosis. Plagiocephaly + small bilateral epicanthal fold with mildly low-set ear with slightly overfolded helices.	(Wallerstein, et al., 2008)
Inversion	interruption of the whole gene	Inversion	DD	ID	ISS onset in utero	ACC + HYD	Mild truncal hypotonia	Prominent forehead with mild frontal bossing, wide nasal bridge, slightly up slant palpebral fissures, wildly downturned corners of the mouth, slightly low-set ears with normal ear shape, and high palate	(Marsh, et al., 2009)

ACC-t = total agenesis of the corpus callosum; CVI = cortical visual impairment; DD = developmental delay; Del = deletion; EIEE – Early infantile epileptic encephalopathy; GPD = global psychomotor delay; HD = homeodomain; HYD = hydranencephaly; ID = Intellectual Disability; ISS = infantile spasm syndrome; ISSX = X-linked infantile spasm syndrome; N=Normal; NS = nonsense; OAR = aristaless domain; Sev Delay = Severe DD; UMN = upper motor neuron syndrome; WS = West Syndrome. Nucleotide numbering reflects cDNA numbering with +1 corresponding to the A of the ATG translation initiation codon in the reference sequence for *ARX* gene [GenBank: NM_139058.2]
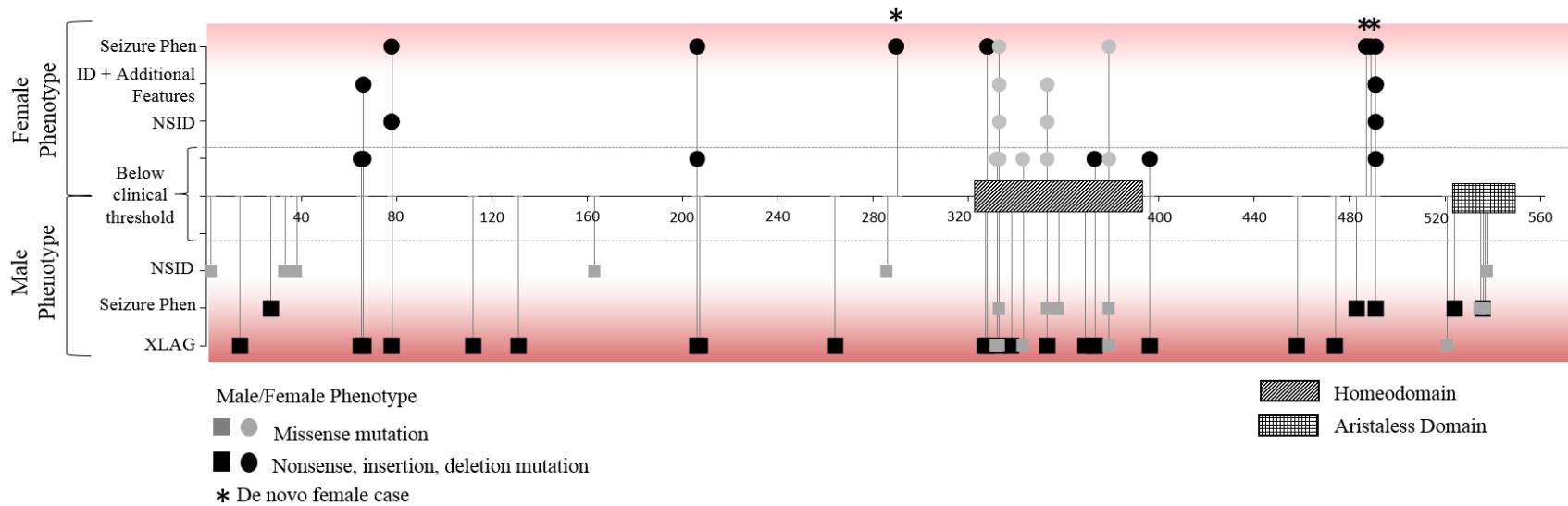
Figure 3.4 Identified ARX mutations in Females and Males leading to a range of phenotypes

a) *ARX* exon structure in relation to the ARX protein. b) Overview of ARX including the homeodomain (crosshatched) and aristaless domain (hatched) and the location of reported mutations according to their relative position at the protein level in both males and females. Differences in mutation type are indicated by a change in colour while missense mutation is shown in grey and all other nonsense, insertion or deletion mutation in black. Phenotype severity is indicated on the y-axis with unaffected carrier females (below clinical threshold) shown below the dotted line.

Given that the *ARX* gene is located on the X chromosome and is subject to X chromosome inactivation (XCI), the contribution of skewed X-inactivation (80:20) to the phenotype in heterozygous females is always a consideration. However, in the female proband of this current report we detected no skewing of X-inactivation in samples collected from blood. X-inactivation in previous studies has been inconclusive with little correlation of the affected status in heterozygous females to preferential inactivation of the normal *ARX* allele (measured in blood) (Marsh et al. 2009). Indeed, skewing of X-inactivation that is not consistent with disease severity has been reported for other X-linked genes that were originally thought to affect males but have had affected female cases described, including mutations in genes such as *USP9X*, *MTM1*, *SLC9A6* and *MED12* (Prontera et al. 2016) (Reijnders et al. 2016, Savarese et al. 2016, Sinajon et al. 2016). This highlights the complexity of X-linked inheritance. Recently, the variability and complexity of X-inactivation status within the brain has been demonstrated in an *Arx* knockout mouse model (Marsh et al. 2016). Heterozygous female mice only have one functional copy of *Arx* and consistent with human data, female mice showed great variation of phenotype manifestations. These mice display striations of radially oriented streams of *Arx* positive neurons within and emerging from the ganglionic eminence (GE) ventricular zone, which was noted to vary dramatically between embryos. This is consistent with the site of random X-inactivation suggested to occur in the ventral forebrain, followed by clonal proliferation of Arx positive or negative cells migrate radially away during early stages of development. Even this incomplete loss of *Arx* in female mice was detrimental and resulted in a change in the profile of interneurons in the adult female mice, consistent with a loss to a greater extent in hemizygous male mice.

We have identified a novel truncating mutation (c.982delCinsTTT/p.(Q328Ffs*37)) in *ARX* in a family ascertained by a female proband displaying a phenotype of mild intellectual disability, seizure disorder and agenesis of the corpus callosum, in conjunction with a phenotype of XLAG in her deceased male siblings. Review of the phenotypes of affected females with published mutations in *ARX* indicates that screening of the *ARX* gene in female patients with intellectual disability, seizure phenotypes and corpus callosum agenesis, particularly if there is evidence of X-linkage and no surviving males is warranted. Moreover, recent *de novo* mutations reported in females recommends scrutiny of *ARX* in all cases with suitable phenotypic presentation with and without other indications of X-linked inheritance. The emerging appreciation of phenotypic consequences in females with loss-of-function mutations in *ARX* will be important in counselling of affected families.

Chapter Four:

Embryonic forebrain transcriptome of mice with polyalanine expansion mutations in the *Arx* homeobox gene

4 Embryonic forebrain transcriptome of mice with polyalanine expansion mutations in the Arx homeobox gene

Publications, Presentations and Published Abstracts from this work:

<u>**Accepted first author publication**</u>

- Mattiske, T., K. Lee, J. Gecz, G. Friocourt and C. Shoubridge (2016). "Embryonic forebrain transcriptome of mice with polyalanine expansion mutations in the ARX homeobox gene." <u>Hum Mol Genet</u>. (DOI: 10.1093/hmg/ddw360)

Accepted publication

- Lee, K., T. Mattiske, K. Kitamura, J. Gecz and C. Shoubridge (2014). "Reduced polyalanine-expanded Arx mutant protein in developing mouse subpallium alters Lmo1 transcriptional regulation." <u>Hum Mol Genet</u> **23**(4): 1084-1094. (DOI: 10.1093/hmg/ddt503)

Conference Abstracts

Talks presented

- Mattiske, T.R, Lee, K., Gécz J and Shoubridge, C (Nov 2016). Investigating the molecular and cellular disruptions in the developing brain resulting in intellectual disability and seizures caused by polyalanine expansion mutations in ARX. Proceedings of the 6th Adelaide ANZSCDB Meeting 2016, Adelaide, Australia.

- Mattiske, T.R, Lee, K., Gécz J and Shoubridge, C (April 2016). Global gene disruption resulting in ISSX caused by polyalanine expansion mutations in ARX. Proceedings of the International Congress Meeting of Human Genetics (ICHG) 2016 Kyoto, Japan.

- Mattiske, T.R, Lee, K., Gécz J and Shoubridge, C (July 2015). Investigating the transcriptome-wide impact of expanded polyalanine tract mutations in ARX contributing to intellectual disability and seizures. Proceedings of the Genetics Society of Australasia 2015, Adelaide, Australia.

Posters presented

- Mattiske, T.R, Lee, K., Gécz J and Shoubridge, C (Feb 2015). Investigating the transcriptome-wide impact of expanded polyalanine tract mutations in *ARX* contributing to intellectual disability and seizures. Proceedings of the 2015 Lorne Genome Conference, Lorne, VIC.

- Mattiske, T.R, Lee, K., Gécz J and Shoubridge, C (Aug 2014). Casting the net transcriptome-wide: How do expanded polyalanine tract mutations in *ARX* contribute to intellectual disability and seizures? Proceedings of the 38th Human Genetics Society Annual Scientific Meeting, Adelaide, SA.

4.1 Abstract

The *Aristaless* related homeobox *(ARX)* gene encodes a paired-type homeodomain transcription factor with critical roles in embryonic development. Mutations in *ARX* give rise to intellectual disability (ID), epilepsy and brain malformation syndromes. To capture the genetic and molecular disruptions that underpin the *ARX*-associated clinical phenotypes, we undertook a transcriptome-wide RNASeq approach to analyse developing (12.5 dpc) telencephalon of mice modelling two recurrent polyalanine expansion mutations with different phenotypic severities in the *ARX* gene. Here we report 238 genes significantly deregulated (Log2FC >+/-1.1, P-value <0.05) when both mutations are compared to wild-type (WT) animals. When each mutation is considered separately, a greater number of genes were deregulated in the severe PA1 mice (825) than in the PA2 animals (78). Analysing genes deregulated in either or both mutant strains, we identified 12% as implicated in ID, epilepsy and autism (99/858), with ~5% as putative or known direct targets of ARX transcriptional regulation. We propose a core pathway of transcription regulators, including *Hdac4,* involved in chromatin condensation and transcriptional repression, and one of its targets, the transcription factor *Twist1*, as potential drivers of the ID and infantile spasms in patients with *ARX* polyalanine expansion mutations. We predict that the subsequent disturbance to this pathway is a consequence of ARX protein reduction with a broader and more significant level of disruption in the PA1 in comparison to the PA2 mice. Identifying early triggers of *ARX*-associated phenotypes contributes to our understanding of particular clusters/pathways underpinning comorbid phenotypes that are shared by many neurodevelopmental disorders.

4.2 Introduction

Neurodevelopmental disorders (NDDs), which include intellectual disability, seizure disorders and autism spectrum disorders are prevalent in the population. Large-scale sequencing efforts have highlighted the genetic heterogeneity contributing to each of these disorders (Epi et al. 2013, Chen et al. 2014, Euro et al. 2014, Krumm et al. 2014, Pinto et al. 2014). Understanding how such divergent etiologies produce similar clinical features remains a major challenge. Despite this, recent studies indicate that many of the pathophysiological mechanisms might be shared, opening the possibility that more than one condition may be amenable to a treatment or disease modification that exploits a common mechanism (Chen et al. 2014, Krumm et al. 2014, Pinto et al. 2014). Here we investigate the molecular mechanisms and functional impact of mutations in the disease-causing gene contributing to intellectual disability and infantile spasms, the *Aristaless* related homeobox gene (*ARX*) [NM_139058.2] (MIM# 300382).

ARX is a member of the paired-type homeodomain transcription factor family with critical roles in development, particularly in the developing brain (Kitamura et al. 2002, Ohira et al. 2002) Depositario-Cabacar, 2010 #78}. ARX is indispensable for telencephalic morphogenesis particularly involved in radial and tangential migration of GABAergic interneuron progenitors, early commitment of cholinergic neurons and is emerging as a selector gene important in preserving identity of specific brain regions (Kitamura et al. 2002, Colombo et al. 2007, Friocourt et al. 2008, Colasante et al. 2009, Lee, K. et al. 2014). In accordance with the essential function of ARX during early brain development, *Arx* expression is detected in mice at embryonic day 8 in a restricted area of the neuroepithelium

corresponding to the prospective forebrain (Bienvenu et al. 2002). During peak neural proliferation and neurogenesis expression of *Arx* within the subpallium peaks between 12.5 to 14.5 days post coitum (dpc), persisting during embryogenesis and is down regulated during postnatal life (Miura et al. 1997, Kitamura et al. 2002).

ARX is an X-chromosome gene. As such, patients are generally affected males with carrier females being asymptomatic. Over 60% of all mutations in *ARX* expand the first or second polyalanine tract, and affected males with these mutations invariably present with intellectual disability with and without infantile spasms and epilepsy (Shoubridge et al. 2010a, Marques et al. 2015). In particular, patients with expansions to the first polyalanine tract (previously reported as c.304ins(GCG)[7], now following HGVS nomenclature reported as c.306GGC[17] ; referred to as PA1 mutation in this study) invariably display seizures, with infantile spasms in 85% of these PA1 patients (Marques et al. 2015). The key phenotypic features seen in these patients are recapitulated in well characterised mutant mouse models, including infantile spasm-like movements, electrodecremetal discharges, and multifocal EEG spikes, with seizures in juvenile and older mice (Kitamura et al. 2009, Price et al. 2009, Beguin et al. 2013). The most frequent mutation in *ARX* in patients results in an expansion to the second polyalanine tract (previously reported as c.429_452dup, now following HGVS nomenclature reported as c.441-464dup; referred to as PA2 mutation in this study) with at least 10-15% of these PA2 patients presenting with infantile spasms in addition to intellectual disability (Marques et al. 2015). Although there is a mouse model of the most common PA2 mutation (Kitamura et al. 2009), phenotypic data for this strain, including the prevalence of seizures is limited.

Our recent investigations in these mice modelling to two most frequent polyalanine expansion mutations in human patients demonstrated aggregation of mutant Arx protein does not occur in the embryonic brain. Instead, we identified a marked reduction in mutant Arx protein expression in the developing forebrain (Kitamura et al. 2009, Lee, K. et al. 2014). Interestingly, this data indicates that both PA1 and PA2 mutations give rise to similar molecular outcomes. Despite recent studies identifying genes regulated by ARX (Fulp et al. 2008, Colasante et al. 2009, Friocourt and Parnavelas 2011) there is limited understanding of what impact the expanded polyalanine tract mutations in *Arx* may have on the transcriptional activity (Nasrallah et al. 2012, Lee, K. et al. 2014) and how this may contribute to phenotypic outcomes. In this study, we use RNASeq on brain tissue at 12.5 dpc during embryonic development to capture early disruptions of Arx function. We show that the gene expression consequences of the polyalanine expansion mutations of *Arx* do overlap but the PA1 mutation leads to a greater and broader disturbance than the PA2 mutation. From our analysis of the deregulated genes, we propose a pathway involving Histone Deacetylase 4 (*Hdac4*) and Twist Family BHLH Transcription Factor 1 (*Twist1*) that when deregulated by either *Arx* mutation contributes to the comorbid phenotypes of intellectual disability and epilepsy.

4.3   Results

4.3.1   PA1 and PA2 mice deregulated transcriptomes overlap.

Arx is indispensable for brain development with expression detected as early as 8 dpc (Miura et al. 1997, Bienvenu et al. 2002). To capture early changes to the transcriptome due to mutations in *Arx* we collected and investigated the 12.5 dpc telencephalon of mice modelling two mutations in *Arx*, PA1 and PA2 (Figure 4.1). We compared four males from each strain with stage-matched WT male littermates.   Sequences were aligned using TopHat and count data generated from HTSeq was used as the input for EdgeR to identify genes with differential expression between samples.
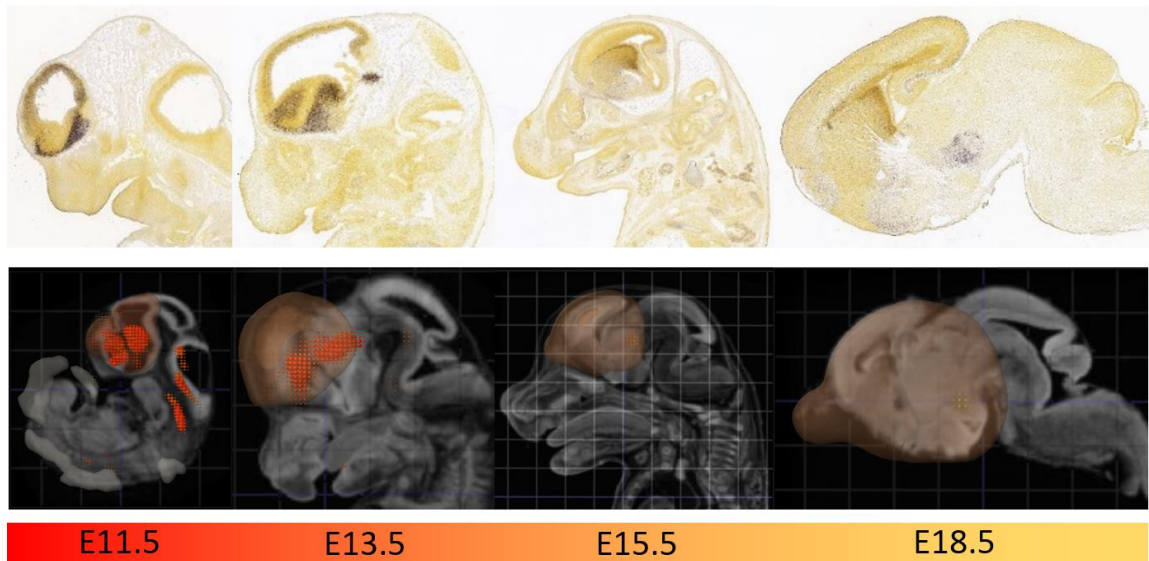
Figure 4.1 Arx expression during embryonic development.
In the top panel *in situ* hybridization images taken from the Allen Mouse Brain Atlas show *Arx* expression at four embryonic developmental time points with the log expression (red = high and yellow = low) shown in the bottom panel. The telencephalon structure is indicated on the 3D reference image (orange shading). Image credit: Allen Institute for Brain Science (http://mouse.brain-map.org).

Analysis of the PA1 mice revealed 825 genes deregulated by Log2 fold change greater than +/- 1.1 with a P-value of less than 0.05, with 54% found at higher levels of expression than WT (Figure 4.2a). In contrast, the PA2 mutation resulted in 78 genes deregulated using the same fold cutoff, with 72% of these found at higher levels of expression than the WT animals (Figure 4.2a). Despite the large difference in the number of genes significantly deregulated between the two PolyA strains, we noted that expression of many of the genes deregulated in the PA1 mice shared the same trend of deregulation in the PA2 mice although the lower log2FC values did not reach the required significance (Figure 4.2b - within dotted lines). Given the similarities of the transcriptome changes between PA1 and PA2, we speculate that the disrupted pathway may be shared between PA1 and PA2. This is supported by previous studies suggesting both PA mutations result in a reduction of the ARX protein together with the overlap of clinical phenotypes in human patients with either PA1 or PA2 mutations. With this in mind, we analysed the genome-wide expression data of both the PA1 and PA2 mice as a single mutation group (referred to as PolyA^{pool}) compared to the WT samples to capture genes deregulated in both mutant strains. Lists of deregulated genes for each analysis are provided in Appendix Table 1-Appendix Table 3. From this analysis, a total of 238 genes were identified (Log2 fold change greater than +/- 1.1 with a P-value < 0.05) (Figure 4.2c). The majority of genes deregulated in the PolyA^{pool} (89%) were at higher levels of expression in mutant mice (Figure 4.2a). Not surprisingly, all genes identified as deregulated by this analysis had already been identified as deregulated in either PA1 or PA2 mice. When we consider the 238 genes identified as deregulated in the PolyA^{pool} analysis, the mean log2FC value of the more severe PA1 group on its own is at 1.3, above the 1.1 cutoff value. In contrast, the milder PA2 group when

considered on its own is below the 1.1 cutoff at 0.7. This means that of the total 238

significantly deregulated genes in the PolyA^{pool} data, 94% (224/238) of genes in PA1 and

24% (56/238) of genes in PA2 met the cutoff values, respectively (Figure 4.2d). Of the 238

genes that were significantly deregulated in the PolyA^{pool} analysis, 42 genes (18%) were

significantly deregulated in both PA1 and PA2 animals when each mutant strain was

considered independently compared to WT (P>3.753e-31). When considered

independently, 182 genes (76%) were significantly deregulated in PA1 animals only, and

14 genes (6%) were deregulated specifically in the PA2 animals. Of the 42 significantly

deregulated genes compared to WT in both PA1 and PA2, 79% (33/42) of the genes had

increased expression (Figure 1E). The overall disruption to the transcriptome observed

supports the notion that PA1 and PA2 disrupt overlapping pathways, with PA2 mice being

a 'milder' form of transcriptome deregulation of the PA1 mice.

a)

	PA1	PA2	PolyA^{pool}
# genes deregulated	825	78	238
Increased expression	446 (54%)	56 (72%)	212 (89%)
Decreased expression	379 (46%)	22 (28%)	26 (11%)

b)
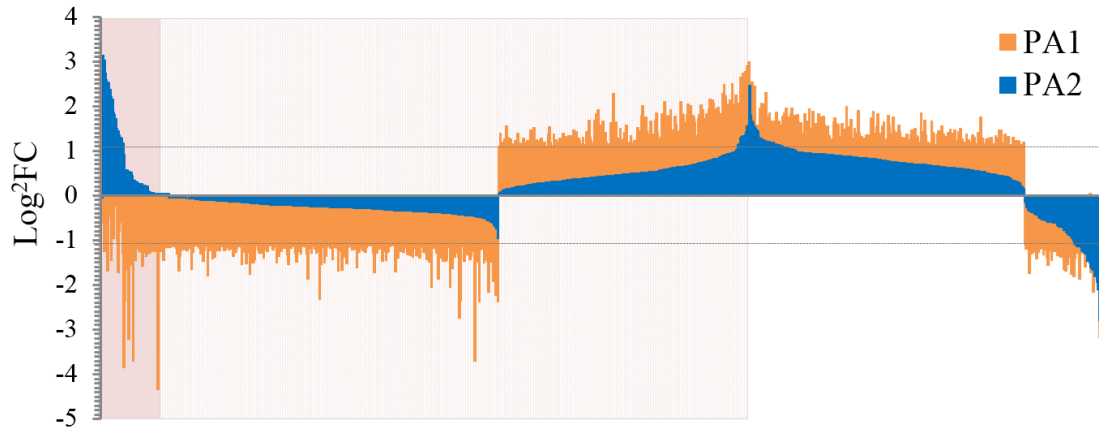
c)
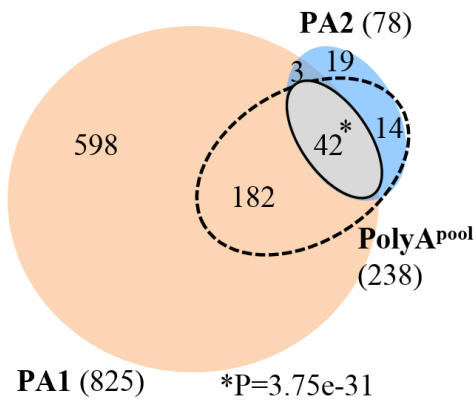
d)
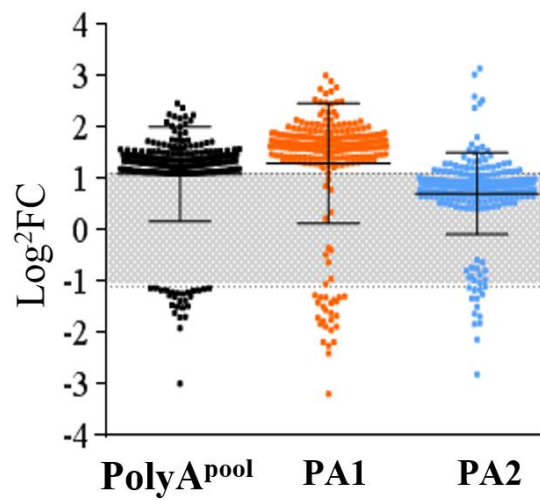
e)
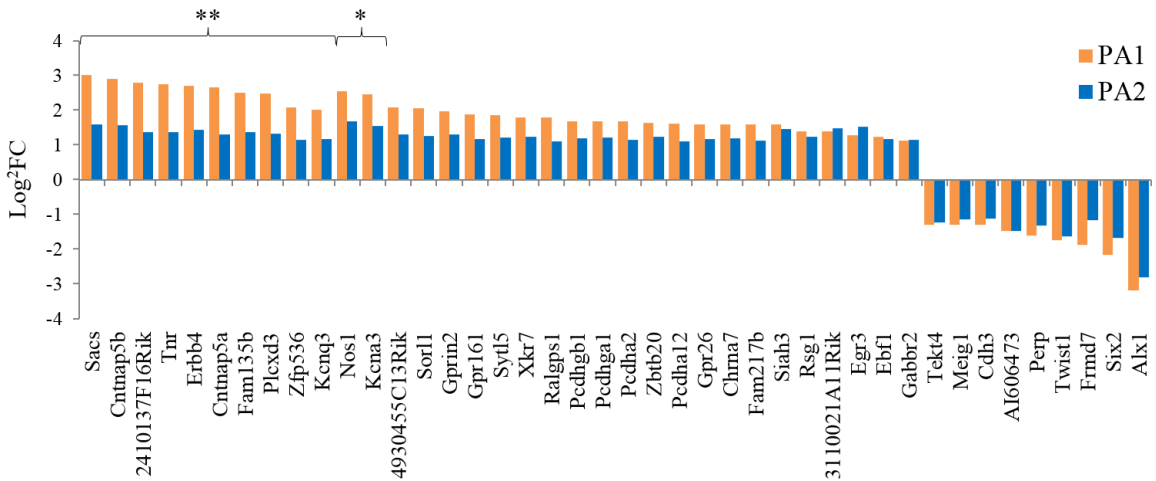
Figure 4.2 Transcriptome analysis of embryonic brains of PolyA Arx mutant mice.
Differential expression of genes from 12.5 dpc mice was determined using EdgeR and
selected based on a Log2 fold change greater than +/- 1.1 with a P-value < 0.05. **(a)** Total
number of deregulated genes and the proportion of either increased or decreased expression
from our analysis from both mutant mice strains independently, PA1 and PA2, and analysis
of pooled PA1 and PA2 strains, PolyA[pool] when compared to WT animals. **(b)** Log2FC was
plotted for all deregulated genes for both PA1 (orange) and PA2 (blue) (total = 858). Genes
within the darker red shaded area are deregulated in different directions compared to WT
with expression of genes significantly different between PA1 and PA2 (58/858, 6.75%).
Genes within the lighter red area are deregulated in the same direction in both PA mutation
groups compared to WT, but are still significantly different between PA1 and PA2
(498/858, 58.05%). The remaining genes are deregulated in the same direction compared
to WT, but are not significantly different between PA1 and PA2 (35.2%). The dotted line
indicates the log2FC +/- 1.1 cut off. **(c)** Gene lists from PA1 (orange), PA2 (blue) and
PolyA[pool] analysis (dotted outline) was used to determine the overlap of genes deregulated
in each group and visualised as a Venn diagram. Overlapping genes in PA1 and PA2
samples with log2FC>+/-1.1 in both lists and a P-value <0.05 are highlighted in grey (solid
outline) with logFC values for individual genes shown in **(e).** *significantly different
between groups with a log2FC<1.1 and P-values <0.05, **significantly different between
groups with a log2FC>+1.1 and P-values <0.05.

4.3.2 PA1 and PA2 mutations disturb overlapping biological processes in the developing brain.

Functional enrichment analysis is a common tool to understand global changes in phenotypes in cells and tissues. EnrichR (Chen et al. 2013) analysis was used to investigate the enrichment of groups of genes with overlapping gene ontology (GO) terms representing gene properties with the focus on biological process. To interrogate the types of genes disturbed due to polyalanine tract expansion mutations in *Arx* we focused on the deregulated genes identified in our analysis with higher levels of expression compared to WT (+1.1 log2FC). This focus was due mainly to the very small numbers of deregulated genes at lower levels of expression in PolyA mutant groups compared to WT in the core and PolyA^{pool} subgroups. GO Terms were ranked using the EnrichR method of combining the p-value computed using the Fisher exact test with the z-score of the deviation from the expected rank by multiplying these two numbers as follows: c = log(p) * z (Chen et al. 2013). Enrichment of GO biological process identified 12 biological processes (adjusted *P*-value <0.05) enriched in the differentially expressed genes (with higher levels of expression compared to WT) relative to all expressed genes (Figure 4.3). These biological processes fall into three main categories; those implicated in cell-cell adhesion (GO:0007156, GO:0098609, GO:0098742); regulation of neuron differentiation/nervous system development (GO:0045664, GO:0010975, GO:0007399); and neuron membrane properties such as synaptic transmission (GO:0007268), regulation of membrane potential (GO:0042391) and potassium ion transmembrane transport (GO:0071805) (Figure 4.3). The levels to which each GO biological process is enriched in each group is displayed as a heat map on the right of Figure 4.3. This analysis indicates that the same categories of GO

processes were enriched whether the deregulated genes considered were from the core overlap group, the PolyA^{pool} group or the broadest PA1 group of deregulated genes, particularly for the cell-cell adhesion process (Figure 4.3). Overlap was also seen in the terms identified by Panther pathway analysis when comparing the top 10 enriched pathways in genes deregulated in PA1 mice and genes deregulated in PA2 mice (Table 4.1-Table 4.2).

	GO number	GO Biological Process	PolyA pool	PA1 Only	All DE Genes	Combined score
cell-cell adhesion	0007156	homophilic cell adhesion via plasma membrane adhesion molecules	██	██	██	125 / 110 / 70 / 50 / 30 / 15 / 10 / 5
	0098609	cell-cell adhesion	██	██	██	
	0098742	cell-cell adhesion via plasma-membrane adhesion molecules	██	██	██	
neuron development	0045664	regulation of neuron differentiation		██	██	
	0010975	regulation of neuron projection development		██	██	
	0007399	nervous system development	██		██	
neuron membrane properties	0007268	synaptic transmission	██		██	
	0031344	regulation of cell projection organization			██	
	0042391	regulation of membrane potential			██	
	0071804	cellular potassium ion transport	██		██	
	0071805	potassium ion transmembrane transport	██		██	
	0060059	embryonic retina morphogenesis in camera-type eye		██	██	
	0006366	transcription from RNA polymerase II promoter		██		
	0042787	protein ubiquitination in ubiquitin-dependent protein catabolic process		██		
	0010921	regulation of phosphatase activity		██		

Figure 4.3 Gene ontology classification of deregulated genes.
Functional enrichment analysis of gene ontology (GO) terms for biological processes shows the differentially expressed genes within subgroups, PolyA[pool], PA1 only and all deregulated genes from both PA1 and PA2 (All deregulated (DE) genes) with a log2 fold change with +>1.1 cutoff value used as the input into EnrichR. The GO terms were ranked based on the combined EnrichR score, and all had a p-value <0.05.

Table 4.1 Top 10 enriched Panther pathways for deregulated genes in PA1.

Term	Overlap	P-value	Adjusted P-value	Z-score	Combined Score	Genes
Cadherin signalling pathway_P00012	34/150	5.99E-15	3.83E-13	-1.6	45.74	PCDHGB6;PCDHGB4;PCDH11X;PCDHGB2;CDH6;CDH3;ERBB4;PCDHA5;PCDHA3;PCDHA2;PCDHA9;PCDHAC2;PCDHA7;PCDHAC1;PCDHA6;PCDHGA8;PCDHGA7;PCDHGA6;PCDHGA5;PCDH9;PCDHGA3;PCDHGA2;PCDHGC4;PCDHGA1;PCDHGC3;PCDHA12;PCDHGA10;PCDHGA11;PCDHGA12;PCDHB2;FAT1;PCDHGB1;CDH13;FAT3
Wnt signalling pathway_P00057	38/278	1.56E-10	4.99E-09	-1.7	32.16	PCDHGB6;PCDHGB4;PCDH11X;PCDHGB2;LRP6;CDH6;CDH3;PCDHA5;GNG8;PCDHA3;TBL1X;PCDHA2;PCDHA9;PCDHA7;PCDHAC2;PCDHA6;PCDHAC1;PCDHGA8;PCDHGA7;CREBBP;PCDHGA6;PCDHGA5;PCDH9;PCDHGA3;PCDHGA2;PCDHGC4;PCDHGA1;PCDHGC3;PCDHA12;PCDHGA10;PYGO1;PCDHGA11;PCDHGA12;PCDHB2;FAT1;PCDHGB1;CDH13;FAT3
GABA-B receptor II signaling_P05731	5/35	0.026	0.55	-1.6	0.96	GABBR2;ADCY9;CACNA1B;GNG8;KCNJ3
Metabotropic glutamate receptor group II pathway_P00040	4/30	0.054	0.87	-1.3	0.19	VAMP8;STX1B;CACNA1B;CACNA1E
p53 pathway_P00059	6/71	0.107	0.98	-0.92	0.019	CREBBP;PERP;BAX;MDM4;SFN;GADD45G
Synaptic vesicle trafficking_P05734	3/23	0.096	0.98	-0.35	0.007	UNC13C;RIMS1;STX1B
Cortocotropin releasing factor receptor signalling pathway_P04380	3/30	0.164	0.99	-0.65	0.004	POMC;VAMP8;GNG8
Metabotropic glutamate receptor group III pathway_P00039	4/54	0.233	0.99	-0.6	0.004	VAMP8;STX1B;CACNA1B;CACNA1E
Thyrotropin-releasing hormone receptor signalling pathway_P04394	3/41	0.289	0.99	-0.36	0.002	VAMP8;CACNA1B;CACNA1E
Muscarinic acetylcholine receptor 2 and 4 signalling pathway_P00043	3/39	0.265	0.99	-0.32	0.002	VAMP8;STX1B;KCNJ3

Table 4.2 Top 10 enriched Panther pathways for deregulated genes in PA2 mice.

Term	Overlap	P-value	Adjusted P-value	Z-score	Combined Score	Genes
Cadherin signalling pathway_P00012	6/150	5.9E-05	0.00059	-1.6	11.9	CDH3;ERBB4;PCDHGA1;PCDHGB1;PCDHA2;PCDHA12
Wnt signalling pathway_P00057	5/278	0.011	0.055	-1.68	4.89	CDH3;PCDHGA1;PCDHGB1;PCDHA2;PCDHA12
Alzheimer disease-presenilin pathway_P00004	2/99	0.099	0.33	-1.39	1.54	CDH3;ERBB4
GABA-B receptor II signaling_P05731	1/35	0.177	0.44	-1.45	1.178	GABBR2
Alzheimer disease-amyloid secretase pathway_P00003	1/56	0.267	0.465	-1.13	0.86	CHRNA7
Nicotinic acetylcholine receptor signalling pathway_P00044	1/68	0.314	0.47	-1.04	0.8	CHRNA7
p53 pathway_P00059	1/71	0.326	0.47	-0.92	0.7	PERP
EGF receptor signalling pathway_P00018	1/109	0.455	0.57	-1.08	0.6	ERBB4
Integrin signalling pathway_P00034	1/156	0.58	0.6	-1.02	0.51	COL8A1
CCKR signalling map ST_P06959	1/165	0.6	0.61	-0.94	0.47	NOS1

4.3.3 Identifying early triggers of *ARX* associated phenotypes

To ascertain which of the deregulated genes were likely to be direct transcriptional targets of Arx, the data from *Arx* knockout expression analysis (Fulp et al. 2008, Colasante et al. 2009) and ChIP studies (Quille et al. 2011) were used to identify that 46 genes deregulated in this studies data set are either known or putative direct targets of Arx. These targets accounted for 5% of all 858 deregulated genes and were detected across all groups considered, with a small level of enrichment in the PolyA^{pool} group (Figure 4.4). Consistent with the fact that ARX is a transcriptional repressor, the majority of these targets were detected at higher levels of expression in the PolyA mutant animals compared to WT. We have previously demonstrated a marked reduction of mutant Arx protein abundance within the developing forebrain of both PA1 and PA2 (Lee, K. et al. 2014), indicating that the expanded polyalanine tract mutations in our mouse models represent a partial loss of *Arx* function. In Figure 4.4c we captured the response of these target genes to PolyA mutation in *Arx* in our study compared to the response in the previously reported studies modelling knocked-out or ablated *Arx* expression in mouse brain (Fulp et al. 2008, Colasante et al. 2009) or in response to exogenous *Arx* overexpression in N2a cells (Quille et al. 2011). This analysis indicates that less than half (43%) of the direct gene targets of Arx identified as deregulated in the brains of 12.5 dpc mice with *Arx* PolyA mutations were also deregulated in *Arx* deficient mice. In general, the direction of deregulation of the target genes was in agreement between the loss of function studies and the partial loss of function in our PolyA mice. Not unexpectedly, there was more variation in the direction of

deregulation between our partial loss of function in the PolyA mice when compared to the

overexpression of exogenous *Arx* in Na2 cells (Figure 4.4c).

a)

	Core Overlap	PolyA^{pool}	PA1 Only	All DE Genes
Genes per Group	42	182	598	858
Direct Targets	3 (7%)	13 (7%)	28 (4.7%)	46 (5%)
Increased expression	3 (100%)	11 (85%)	13 (46%)	27 (59%)
Decreased expression		2 (15%)	15 (54%)	19 (41%)

b)

c)

Figure 4.4 Disruption to putative and known ARX target genes.
a) The number of putative and known direct targets of ARX and the direction of deregulation in the PolyA mutant mice is shown for each subgroup of the RNASeq analysis. b) The putative and known direct gene targets of ARX are spread across all subgroups of the RNASeq analysis as illustrated on the Venn diagram. c) For each of these putative or direct gene targets of Arx identified in the PolyA mutant mice the change in expression compared to WT from this study (PA1 and PA2) with comparison of the changes to expression in studies of *Arx* loss of function (ARX KO) (Fulp et al. 2008, Colasante et al. 2009), *Arx* overexpression in N2a cells (ARX OE) (Quille et al. 2011) and ChIP studies (Quille et al. 2011). Red indicates an increase in expression with Green a decrease in expression compared to WT.

4.3.4 De-regulation of early triggers of *ARX* associated phenotypes persists across embryonic development

Considering the spatial expression of *Arx* in the subpallium is restricted to the in both lateral and medial ganglionic eminence, we were interested if there was any obvious relationship of spatial localisation in regard to the genes deregulated in our PA mice. We compared expression of a number of deregulated genes that had available expression data from the Allen Developing Mouse Brain Atlas and EURExpress at suitable embryonic stages. Genes with a similar spatial expression profiles as determined by *in situ* hybridization images to that of *Arx* include *Ebf1*, *Rapgef5*, *Myt1l*, *Erbb4*, *Al606473*, *Zfp536* and *Gpr26*. In contrast, several genes had an opposite spatial expression profiles compared to *Arx* including *Zbtb20, Ptch1* and *Hdac4*, which are expressed in the proliferating cells in the ventricular zone (Figure 4.5). This analysis highlights that the early triggers of *Arx* associated phenotypes identified in this study includes both directly regulated and indirectly regulated gene targets contributing to disease outcomes.
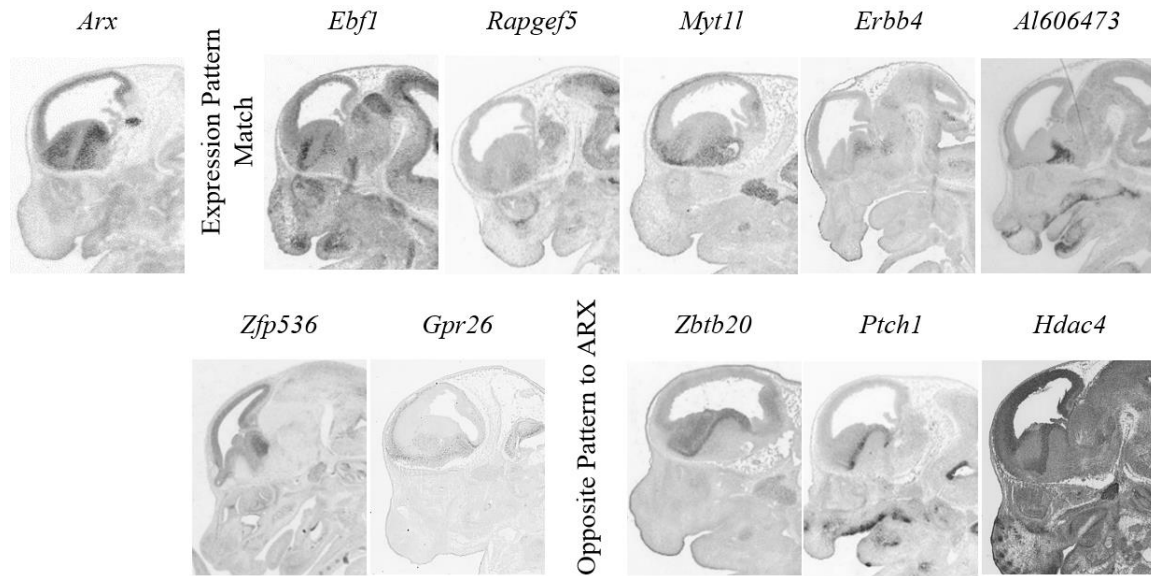
Figure 4.5 Spatial expression profiles of selected genes deregulated in expanded polyA mutant mice

Arx expression is present in the subpallium (in both lateral and medial ganglionic eminence) with similar spatial expression profiles seen for *in situ* hybridization images of *Ebf1*, *Rapgef5*, *Myt1l*, *Erbb4*, *Al606473*, *Zfp536* and *Gpr26*. In contrast, *Zbtb20*, *Ptch1* and *Hdac4* are expressed within the proliferating cells in the ventricular zone. The *in situ* hybridization images for *Arx, Erbb4, Ebf1, Rapgef5, Myt1l, Zbtb20,* and *Ptch1* are from the Allen Mouse Brain Atlas at developmental time point E13.5 dpc. The *in situ* hybridization images of Zfp536, Gpr26, Al606473 and Hdac4 are from EURExpress at developmental time point 14.5 dpc. Image credit: Allen Institute for Brain Science (http://mouse.brain-map.org) and EURExpress (http://www.eurexpress.org/).

To determine whether disruption to the transcriptome was constrained to very early development in the time point tested by RNASeq or continued throughout embryonic life we chose to test fifteen genes by quantitative PCR in the WT compared to PA1 mice across three developmental time points, 12.5, 14.5 and 18.5 dpc. The genes selected included a number a putative targets of Arx previously reported, *Plcxd3*, *Gpr26*, *Ebf1*, *Myt1l*, *Aspm*, *Hdac4* and *Ptch1*. Of interest also were deregulated genes that were known neurodevelopment genes (*Cdkl5*, *Erbb4*, *Sor1l*, *Twist1*, *Myt1l*, *Aspm*, *Hdac4* and *Ptch1*) or displayed spatial expression patterns indicated above (*Erbb4*, *Zfp536*, *Zbtb20*, *Gpr26*, *Ebf1*, *Myt1l*, *Hdac4* and *Ptch1*), or were deregulated in both PA1 and PA2 (*Kcna3* and *Six2*). Our analysis indicates that 11/15 genes validated with deregulated gene expression at 12.5 dpc by quantitative PCR, but also had deregulated expression in PA1 mutant mice across more advanced stages of embryonic development, with *Myt1l, Plcxd3,* and *Six2* consistently significantly deregulated across all time points examined (Figure 4.6). For several genes with validated deregulation by quantitative PCR in PA1 mice we confirmed that the deregulation was also validated in PA2 samples at 12.5 dpc by this analysis, and moreover identified that deregulation in these genes persisted in later embryonic developmental time points (Figure 4.6). Considering that the expression of *Arx* during development significantly diminishes at 18.5 dpc it was not surprising that we noted less consistent deregulation of genes persisting in the 18.5 dpc time point compared to earlier 14.5 dpc. This variability at 18.5 dpc could be due to a number of factors, but is likely influenced by the increased complexity and size of the brain (tissue sample collected) at this time point and underscores the strategy for targeting the early developmental stages where *Arx* expression is appreciable.

a)

| | PA1 | | | | | PA2 | | | | | |
| | % relative to WT | | | | | % relative to WT | | | | | |
	12.5 dpc		14.5 dpc		18.5 dpc		12.5 dpc		14.5 dpc		18.5 dpc	
Erbb4	11.70	*	13.03	*	1.37		NT		NT		NT	
Plcxd3	33.25	**	30.18	**	13.04	*	NT		NT		NT	
Zfp536	51.00	**	31.13	**	-26.26	*	59.22	**	30.21	*	-17.81	*
Kcna3	42.60	**	27.22	**	-23.36	*	43.50	**	16.35	*	-19.24	*
Sorl1	81.88	**	33.73	**	-30.33	*	72.44	**	24.85	**	-22.02	*
Zbtb20	-7.69		23.87	*	24.29	**	*NT*		*NT*		*NT*	
Gpr26	20.79	*	1.51		4.62	*	*NT*		*NT*		*NT*	
Ebf1	19.04	**	-12.62		39.93	**	*NT*		*NT*		*NT*	
Cdkl5	7.38		15.89	*	24.61		*NT*		*NT*		*NT*	
Aspm	26.27		-38.37		-50.65	*	*NT*		*NT*		*NT*	
Myt1l	27.98	**	34.18	**	50.31	*	*NT*		*NT*		*NT*	
Hdac4	4.97	*	1.47		-2.56		*NT*		*NT*		*NT*	
Ptch1	13.09	*	-13.58	*	-29.57	**	*NT*		*NT*		*NT*	
Twist1	-30.78	*	-44.28	*	-9.01		-6.75		-45.90	*	5.38	
Six2	-35.05	**	-31.96	**	-4.63	*	-21.14	*	-35.49	**	10.69	*

Up-regulated

Down-regulated

* p<0.05 **p<0.001

Opposite level of deregulation

NT = Not Tested

b)

Erbb4

Plcxd3

Zfp536

128
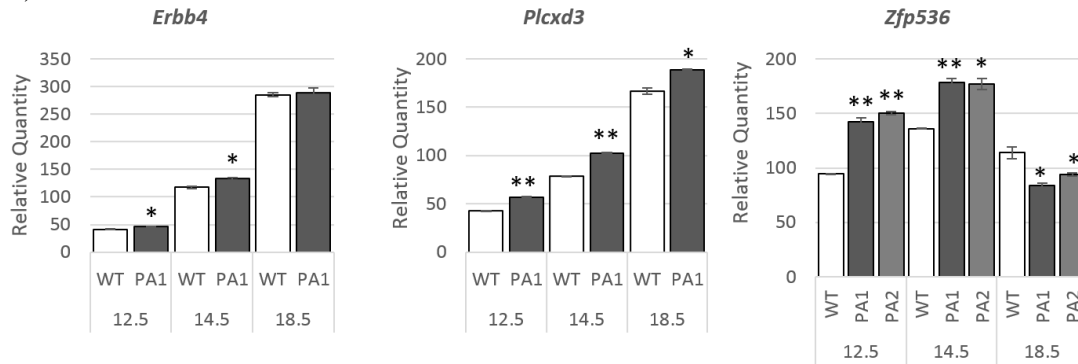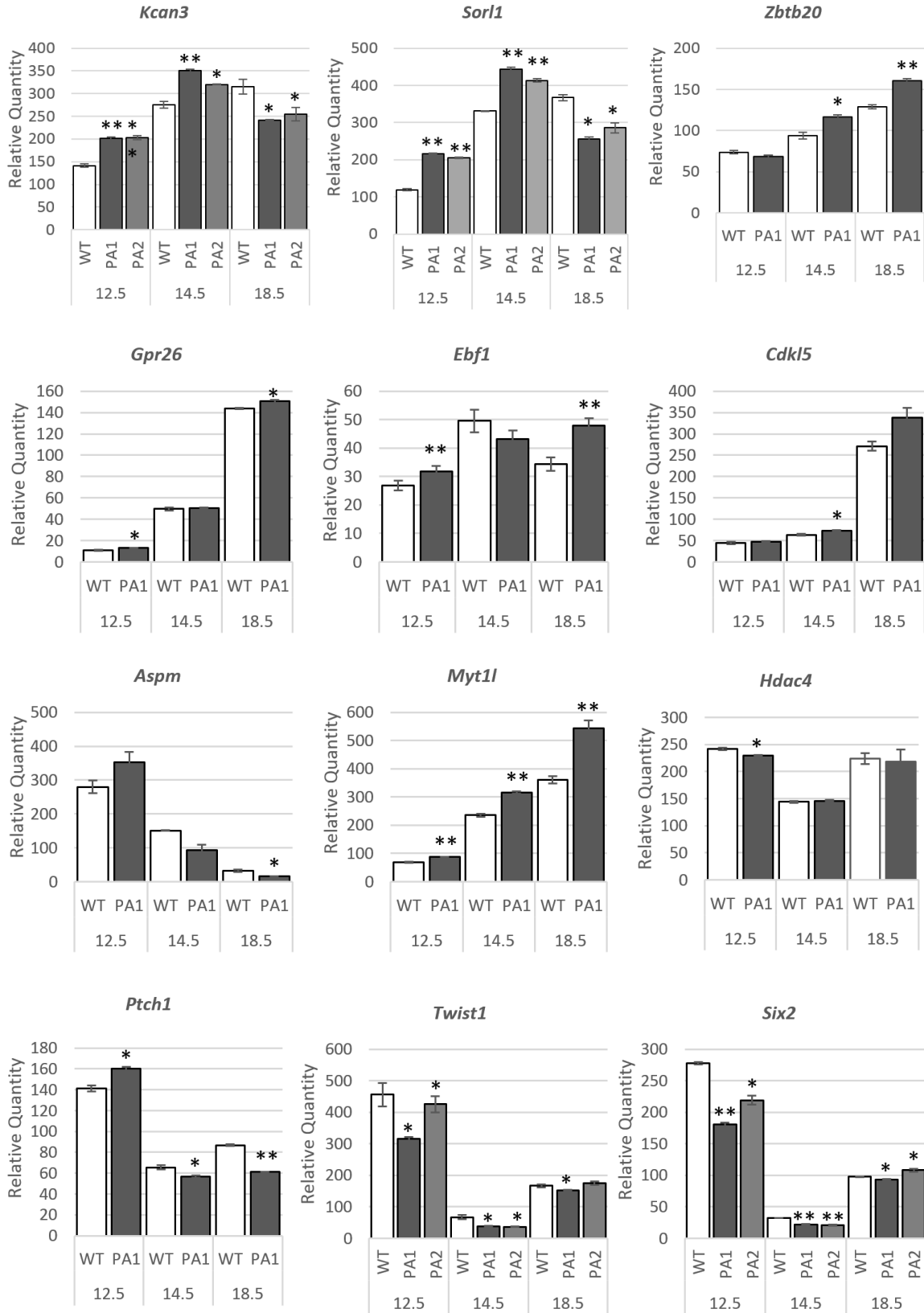
Figure 4.6 Real-time Analysis of deregulated genes across embryonic development due to polyalanine tract mutations in Arx

Sample tested were pooled RNA samples isolated from the telencephalon of embryos for each genotype across the three time points and compared to pooled WT samples with the same number of individual samples (12.5 and 14.5 dpc n=8, 18.5 dpc n=4). Expression values were normalised to the reference gene *Tbp*. A) Summary of the genes analysed for PA1 and selected genes analysed for PA2. B) Individual graphs of relative quantity for each gene across three embryonic time points for WT (white bars), PA1 (dark grey bars) and PA2 expression (pale grey bars). *$P < 0.05$, **$P < 0.01$ (one-tailed t-test of PA1 or PA2 compared to WT at each time point).

4.3.5 Deregulation of neurodevelopment disorder genes contribute to the polyalanine expansion mutation phenotype

Deregulated genes that overlap with lists of genes implicated in epilepsy, ID and autism (Pinto et al. 2014) showed in the total group of 858 deregulated genes, 99 genes (12%) are known disease-causing genes for Epilepsy (55 genes), ID (29 genes) and Autism (51 genes) (Figure 4.7a). This overlap is significant when considering the total group of deregulated genes (P<0.014), the PolyA[pool] group ($29/182 = 16\%$, P<0.003) and the core overlap group ($9/42 = 21\%$, P<0.014) but not in the PA1 only group (P<0.292) (Figure 4.7a and b). Given the incidence of comorbidity of phenotypic features of these particular neurodevelopmental disorders (NDD), we were not surprised to see that many of the ID genes ($22/29 = 76\%$) identified in our data also contribute to epilepsy and autism, with 8 genes contributing to all three comorbidities of epilepsy, ID and autism (Figure 4.7c). The distribution of these NDD genes that were deregulated in our PolyA mice are shown in Figure 4.7d. Of these 99 NDD genes, only 7 are currently identified as direct targets of ARX (bold entries on Figure 4.7d) (Quille et al. 2011). Movement disorders including dystonia and dyskinesia

are a frequent co-morbidity in patients with expanded polyA mutations in *ARX*. Known

genes for these disorders are highlighted in the 99 NDD genes (underlined entries on

Figure 4.7d) with a full list of these genes deregulated in the PolyA mice in Table 4.3.

a)

	Core Overlap	PolyA^{pool}	PA1 Only	All DE Genes
Total Genes deregulated	42	182	598	858
ID Genes	3 (7%)	6 (3%)	20 (3%)	29 (3%)
Epilepsy Genes	4 (10%)	15 (8%)	35 (6%)	55 (6%)
Autism	6 (14%)	16 (9%)	29 (5%)	51 (6%)
Total NDD genes	**9 (21%)**	**29 (16%)**	**60 (10%)**	**99 (12%)**
P value	p < 0.014	p < 0.003	p < 0.292	p < 0.014

b)

c)

132

d)

	ID Epilepsy Autism	ID Epilepsy	Epilepsy Autism	ID Autism	ID	Epilepsy	Autism
Core Overlap	*Kcnq3*		*Chrna7*		*Twist1*	*Gabbr2, Sorl1*	*Fam135b, Pcdha2, Pcdha12*
PolyA pool	*Cdkl5 Hcfc1* **Hdac4**	*Fktn* **Ptch1**	*Lrrc7*	**Aff2** *Erbb4*	*Pou3f4*	*Cbl, Csmd1, Csmd3, Fos, Gatad2b, Kcnh5, Lyst, Mib1, Mrs2, Syt14*	*Chrm3, Gabrq, Gan, Grid2, Pcdha6, Pcdha7, Pcdha9, Pcdhac1, **Pcdhac2**, Pcdhga11, Plxna4, Ptprt*
PA1 Only	*Crebbp Disc1 Lrp2 Vps13b*	**Aspm** *Cdk5rap Cep290 Chkb Kcna1 Nhs Shroom4*	*Dlgap2 Herc2 Kcnma1 Reln*	*Huwe1* **Myt1l** *Shank2*	*Cdon Gp1bb Nfix Rnaseh2c Tfap2a*	*Alg10b, Aqp4, Casc5, Ccdc88c, Cit, Dtnbp1, Epg5, Nalcn, Nanos3, Ndufa2, Ndufa8, Npy, Pet100, Pomc, Rest, Scarb2, Sepsecs, Vps13a, Xpnpep3*	*Aff4, Cacna1b, Cib2, Cntn5, Cntnap3, Fat1, Grcc10, Myo16, Nbea, Ndufa5, Nos1ap, Pcdh9, Pcdha3, Pcdha5, Rims3, Tbl1x, Wnk3*
PA2 Only							**Msx2**
Total	8	9	6	5	7	32	32

Figure 4.7 Genes deregulated in expanded polyA mice include known ID, epilepsy and autism genes.
a) Total number of genes associated with overlapping phenotypes of ID, epilepsy and autism and the proportion (%) per mutation subgroup. P-value indicates the significance of overlap between each PolyA mutant subgroup and the known neurodevelopment disorder genes. b) A combined graph showing the total number of genes per subgroup (grey bars) on the right-hand y-axis compared to the proportion of overlap with known neurodevelopmental disorder genes across RNASeq analysis subgroups shown on the left-hand y-axis (legend for line graph shown below the graph). b) Venn diagram showing the overlap of genes associated with ID, Epilepsy and Autism comorbid phenotypes. d) List of known neurodevelopmental disorder genes within each PolyA mutant subgroup, grouped by the associated disease phenotypes. Known movement disorder genes are underlined. The putative and known direct targets of ARX / Arx are highlighted in bold.

Table 4.3 Muscle disorder genes found in polyalanine expansion mutant mice.
The following gene panels were interrogated: Blueprint Genetics & The University of Chicago Ataxia panels, CGC genetics Dystonia Panel, Radboud University Medical Centre muscle disorder gene panel (DGD20062014).

	Ataxia	Dystonia	Muscle Disorder
Core Group	Sacs		
PolyA^{Pool}	Atp2b3, Cdkl5, Gan, Grid2, Lars2, Syt14		Fktn
PA1 Only	Atp8a2, Ccdc88c, Cep290, Dmxl2, Herc1, Kcna1, Ndufa2, Reln, Scarb2, Sepsecs, Ttbk2	Atp7b, Cacna1b, Kcnma1, Vps13a	Chkb, Dpm3, Eno3, Syne2

4.3.6   Early triggers of *ARX* associated Overlap *Arx-Hdac4-Twist1* pathway

The top 10 enriched terms by KEGG and Reactome pathway analysis for genes deregulated in the PolyA pool group highlight the breadth of neurological processes impacted by the genes deregulated in these mutant mice (Table 4.4 and Table 4.5). Interestingly, from our analyses, we propose one particular pathway of transcriptional regulators implicated in NDD and deregulated in the PolyA mutant mice that may act as potential drivers of the phenotypes, particularly ID and epilepsy (Figure 4.8). The pathway highlights direct interactions based on Ingenuity Pathway analysis and includes manually curated data from the literature based on ChIP studies to identify additional direct interactions across the pathway. We propose that the reduced protein expression in the developing brains of the PolyA mice (Lee, K. et al. 2014) leads to inadequate regulation of *Hdac4*, involved in chromatin condensation and transcriptional repression (Fischer et al. 2010), which in turn has a flow on effects directly on targets such as *Mef2c* and *Twist1*. *Twist1* is a basic-helix-loop-helix transcription factor involved in cell lineage determination & differentiation (Nieto 2013). The expression of *Twist1* was significantly decreased in both the PA1 and PA2 mutant mice when analysed independently and confirmed by qRT-PCR (Figure 4.6b).

A recent study using a ChIP approach to identify targets of TWIST1 (Lee, M.P. et al. 2014) indicates that 302 genes deregulated in the current study overlap with TWIST1 targets, 23/302 of which were deregulated in both PA1 and PA2 mutant mice, 17/302 being known ID genes and 32/302 being known epilepsy genes. We recently demonstrated that there was a consistently greater level of reduction of mutant Arx protein in the developing brains of the PA1 mice with a more variable reduction of mutant protein in the embryonic brains

of PA2 mice (Lee, K. et al. 2014). Hence, we propose that adequate levels of Arx protein are required to regulate direct targets such as *Hdac4* for normal brain development. When this regulation is not achieved the subsequent disturbance is relative to the reduction in Arx mutant protein expression, with a broader and more significant level of disruption observed in the PA1 mice as compared to the PA2 mice.

We performed in situ hybridization of *Arx*, *Cdkl5*, *Hdac4* and *Twist1* in 14.5 dpc brain sections from WT and PA1 embryos to test whether there is a change in the ectopic expression of these genes in cells that would not normally express *Cdkl5*, *Hdac4* and *Twist1*. *In situ* hybridization confirmed no overall loss of *Arx* transcript (Lee, K. et al. 2014) and showed normal *Arx* spatial expression pattern within the 14.5 dpc forebrain in PA1 (Figure 4.10 b, c, k and l). Riboprobes for *Cdkl5*, *Hdac4* and *Twist1* were tested over multiple *in situ* series to determine the riboprobe efficiency and specificity with the optimal chosen dilutions shown in Adult WT samples in Figure 4.9. The efficiency of probes was low compared to *Arx*, used at a 1/50 dilution, with used at *Cdkl5* at 1/10, *Hdac4* at 1/5 and *Twist1* was used neat. In an effort to improve probes, they were remade however no increases in strength was seen. *Cdkl5* shows robust expression in the adult tissue in the cortex and hippocampus with specificity confirmed with previously published work (Rusconi et al. 2008). *Twist1* and *Hdac4* also showed expression in the cortex and hippocampus however staining was not robust. Specificity of *Twist1* and *Hdac4* was in agreement with Allen Brain Atlas for Adult Mice (http://mouse.brain-map.org/). No change in ectopic expression was present in PA1 when compared to WT embryos at 14.5 dpc for *Cdkl5*, *Hdac4* and *Twist1* (Figure 4.10 d-I and m-r). Due to poor efficiency of in

situ hybridization of these probes in both WT and PA1 embryos, it would need to be confirmed in a repeat experiment with probes possible directed towards alternative mRNA sequence or through alternative methods.

Table 4.4 Top 10 enriched KEGG pathways for PolyAPool genes (238).

Term	Overlap	P-value	Adjusted P-value	Z-score	Combined Score	Genes
Cholinergic synapse 04725	7/111	5.E-05	0.007138	-2.06	10.19	CHRM3;ADCY9;CHRNA7;KCNQ3;FOS;CREB5;KCNJ3
Estrogen signalling pathway 04915	6/99	0.0002	0.016	-1.896	7.87	GABBR2;ADCY9;SHC3;FOS;CREB5;KCNJ3
Calcium signalling pathway 04020	7/180	0.0009	0.04	-1.91	6.098	CHRM3;ADCY9;ERBB4;CHRNA7;ATP2B3;NOS1;CACNA1E
cAMP signalling pathway 04024	6/199	0.007	0.14	-1.84	3.606	GABBR2;ADCY9;PTCH1;ATP2B3;FOS;CREB5
Insulin secretion 04911	4/85	0.006	0.141	-1.81	3.549	CHRM3;ADCY9;KCNN3;CREB5
Viral carcinogenesis 05203	6/205	0.008	0.141	-1.80	3.534	HDAC4;EGR3;CDK6;HIST1H4K;HIST1H4F;CREB5
Salivary secretion 04970	4/89	0.007	0.141	-1.71	3.34	CHRM3;ADCY9;ATP2B3;NOS1
Circadian entrainment 04713	4/95	0.009	0.144	-1.71	3.317	ADCY9;FOS;NOS1;KCNJ3
Morphine addiction 05032	4/91	0.008	0.141	-1.69	3.305	GABRQ;GABBR2;ADCY9;KCNJ3
Retrograde endocannabinoid signalling 04723	4/101	0.01	0.159	-1.68	3.09	GABRQ;RIMS1;ADCY9;KCNJ3

Table 4.5 Top 10 enriched Reactome pathways for PolyAPool genes (238).

Term	Overlap	P-value	Adjusted P-value	Z-score	Combined Score	Genes
Neuronal System-112316	13/301	1.5E-06	0.00053	-2.23	16.79	GABRQ;GABBR2;KCNH5;KCNH7;KCNB2;CHRNA7;KCNA3;CACNA1E;RIMS1;ADCY9;KCNQ3;KCNN3;KCNJ3
Potassium Channels-1296071	8/99	2.6E-06	0.00053	-1.95	14.72	GABBR2;KCNH5;KCNH7;KCNB2;KCNQ3;KCNA3;KCNN3;KCNJ3
Voltage gated Potassium channels-1296072	5/43	4.4E-05	0.006091	-1.99	10.13	KCNH5;KCNH7;KCNB2;KCNQ3;KCNA3
GABA receptor activation-977443	4/55	0.0014	0.144624	-1.98	3.82	GABRQ;GABBR2;ADCY9;KCNJ3
Transmission across Chemical Synapses-112315	7/211	0.0021	0.174529	-2.17	3.79	GABRQ;GABBR2;RIMS1;ADCY9;CHRNA7;CACNA1E;KCNJ3
GABA B receptor activation-977444	3/39	0.00498	0.293	-1.99	2.45	GABBR2;ADCY9;KCNJ3
Activation of GABAB receptors-991365	3/39	0.00498	0.293	-1.98	2.43	GABBR2;ADCY9;KCNJ3
Neurotransmitter Receptor Binding And Downstream Transmission In The Postsynaptic Cell-112314	5/142	0.0073	0.378	-2.047	1.99	GABRQ;GABBR2;ADCY9;CHRNA7;KCNJ3
Signaling by FGFR1-5654736	7/336	0.023	0.513	-2.52	1.68	ADCY9;SHC3;ERBB4;FLRT1;KSR2;FOXO3;CBL
Signaling by FGFR-190236	7/366	0.034	0.513	-2.5	1.67	ADCY9;SHC3;ERBB4;FLRT1;KSR2;FOXO3;CBL
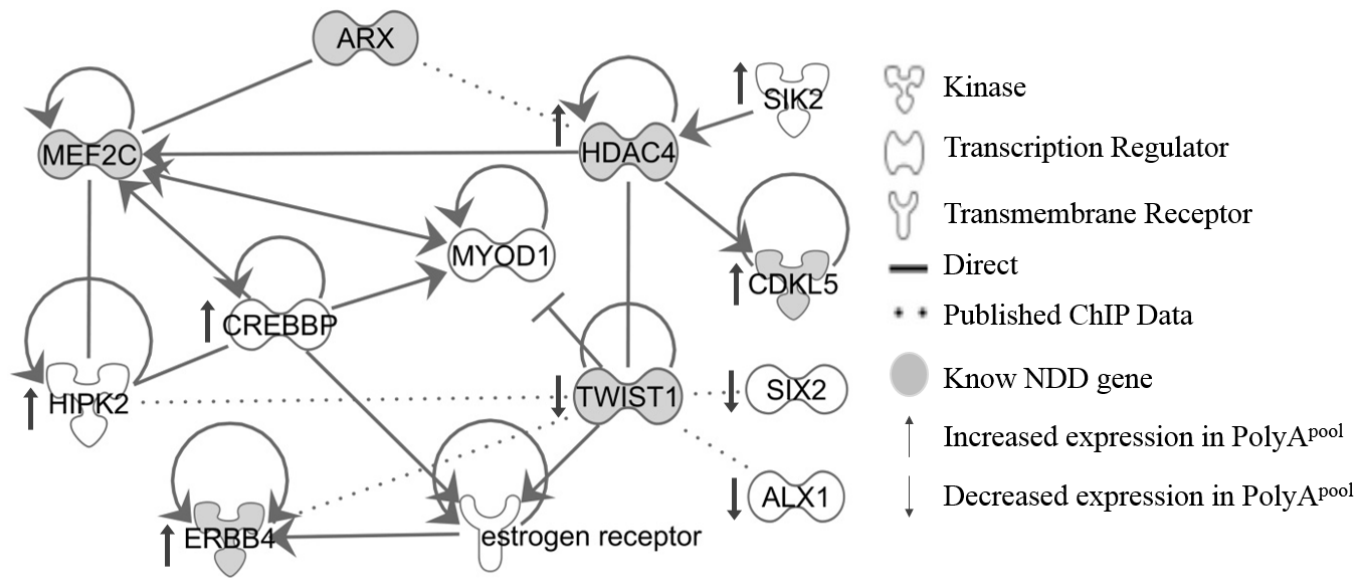
Figure 4.8 *Arx-Hdac4-Twist1* **PolyA-deregulated pathway**
Pathway analysis was used to assess connectivity of deregulated genes of PolyA-*Arx* loss-of-function embryonic brains. The geometric shapes reflect differing types of proteins as defined by ingenuity analysis (Ingenuity Systems) (see legend) with direct connections shown as solid lines. Published ChIP data highlighting relationships between genes and proteins are shown as dotted lines. The direction of deregulation in PolyA[pool] is indicated with arrows (↑ = increased expression, ↓ = decreased expression). Known NDD genes are shaded in grey.
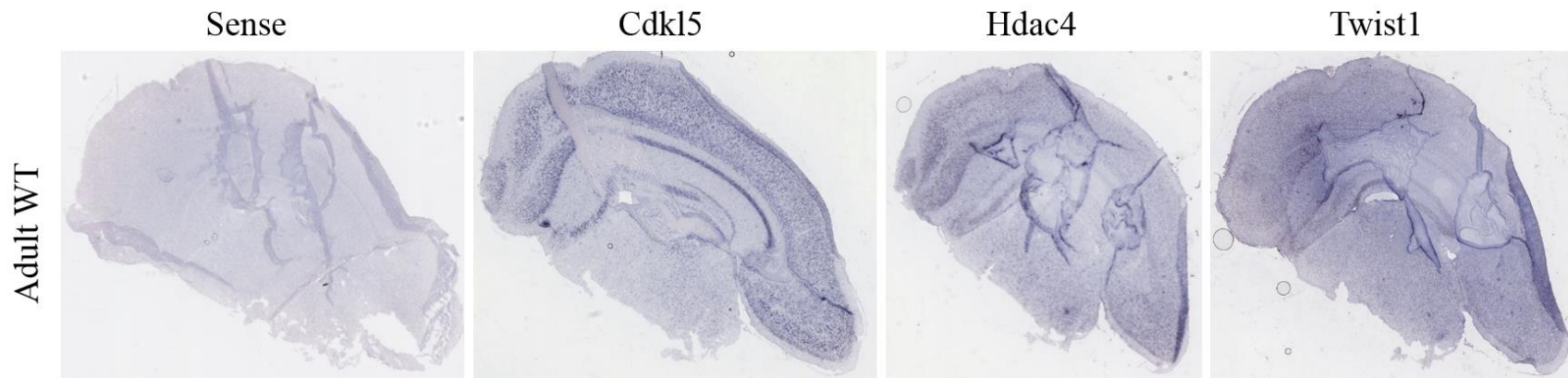
Figure 4.9 Confirmation of *in situ* riboprobes directed towards *Cdkl5*, *Hdac4* and *Twist1* validity in Adult Brain Tissue.
Coronal adult brain sections (16 µm) probed with *in situ* hybridization confirms the performance of riboprobes. *In situ* hybridization of *Cdkl5* (probe dilution 1/10), *Hdac4* (probe dilution 1/5) and *Twist1* (neat probe) with all three showing expression in the cortex and hippocampus. Arx Sense probe was used as a negative control.
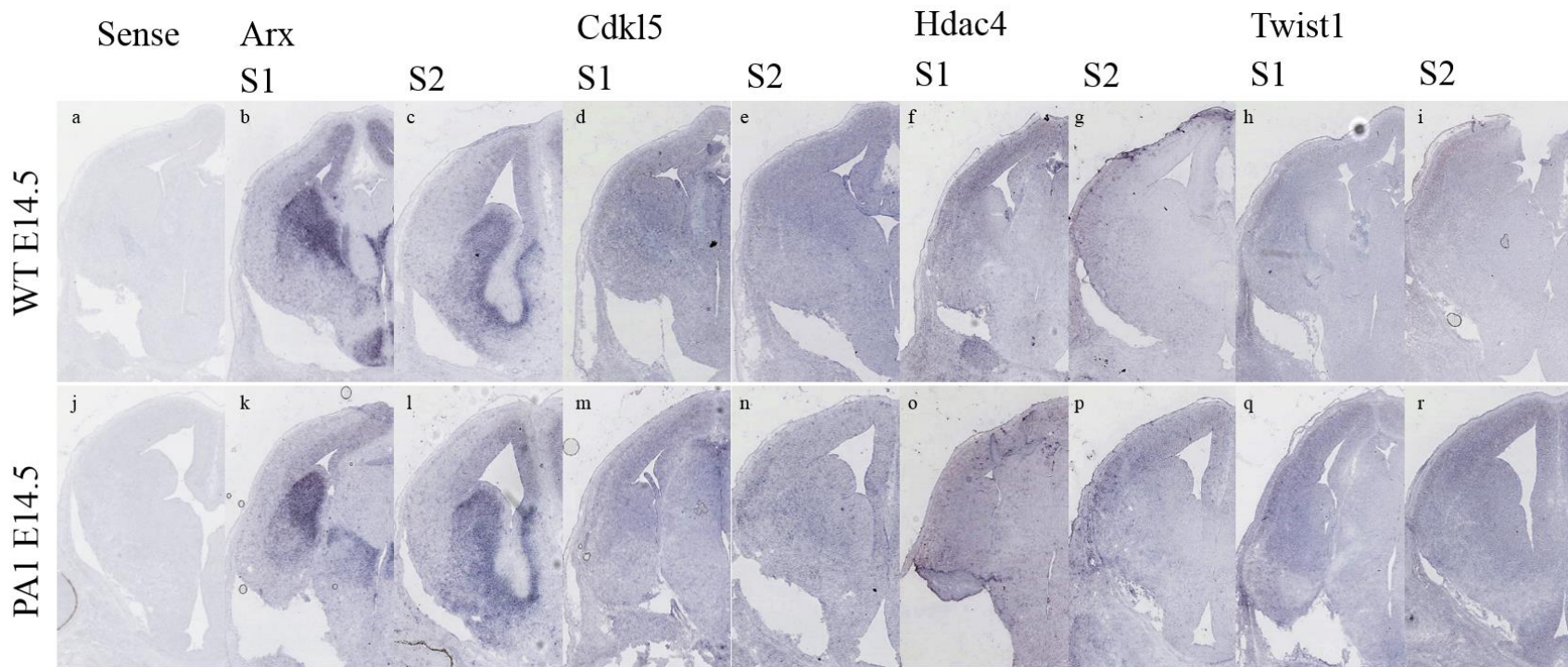
Figure 4.10 *In situ* analysis of deregulated targets *Cdkl5*, *Hdac4* and *Twist1* in E14.5 Tissue
Coronal embryonic (E14.5) brain sections (10 µm) probed with *in situ* hybridization in both WT (a-i) and PA1 (j-r) samples across two section series (S1 and S2). *In situ* hybridization of *Cdkl5* (d, e, m, n) (probe dilution 1/10), *Hdac4* (f, g, o, p) (probe dilution 1/5) and *Twist1* (h, o, q, r) (neat probe). *Arx* probe (b, c, k, l) was used as a positive control, and *Arx* Sense probe (a, j) was used as a negative control.

4.4   Discussion

Expanded polyalanine tracts are the most frequent mutations reported in the *ARX* gene (Shoubridge et al. 2010a) with the patients' phenotypic features invariably including ID with early onset seizures as a frequent comorbidity. To gauge the early events of compromised ARX function, we assessed transcriptome-wide outcomes of the two most frequent *ARX* polyalanine tract expansion mutations within the telencephalon of their respective mouse models. RNASeq analysis of the forebrain of 12.5 dpc telencephalon highlights early alterations that are common to both PolyA mutants, with a greater impact in mice modelling the more severe PA1 mutation. Our validation analysis indicates that this deregulation identified at 12.5 dpc persists across embryonic development. Many of the genes deregulated are known ID, epilepsy or autism disease genes. The enrichment of genes with synaptic function supports the convergence of common pathways contributing to these disorders. Interrogating the profile of deregulated genes from the developing brains of *Arx* PolyA mutant mice we have identified a number of pathways impacted by genes deregulated in the mutant mice including a 'core' pathway of transcriptional regulators that we suggest may represent early triggers of ARX-associated phenotypes.

The molecular pathogenesis of expanded polyalanine tracts is not well understood. In the case of *Arx/*ARX, we have demonstrated a marked reduction of the Arx protein expression in the developing forebrain of mutant PA1 and PA2 mice, occurring as early as embryonic stage 12.5 dpc (Lee, K. et al. 2014). Reduced protein expression of the mutant protein *in vivo* may be a common molecular mechanism contributing to the disease associated with

expanded polyalanine tracts (Innis et al. 2004, Kuss et al. 2009, Krumm et al. 2014, Lee, K. et al. 2014). At this early embryonic stage, *Arx/*ARX is highly expressed in the telencephalon during peak cellular proliferation/neurogenesis and the first wave of tangential migration of neurons from the ganglionic eminence to the developing cortex. There is evidence that polyalanine expansion mutations in *Arx* retain transcription regulation capacity for some gene targets, but fail to repress a subset of predicted or known targets (Quille et al. 2011, Nasrallah et al. 2012, Lee, K. et al. 2014). Our RNASeq approach offers an unbiased and comprehensive interrogation of deregulation of gene expression due to these mutations. We show a substantial number of deregulated genes as a consequence of PolyA expansion mutations in *Arx* during early embryonic brain development. Interestingly, most of the significantly deregulated genes in the PA1 mutant mice were also perturbed in PA2 mice, but failed to reach significance by our analysis. We cannot rule out the possibility that the structure and location of the PA2 mutation within Arx contributes to the milder phenotypic outcomes in both our mice and affected patients. However, a combination of data from this study and our previous work leads us to suggest an alternative hypothesis. We have shown that the reduction in Arx mutant protein in the embryonic brain appears more variable in the PA2 mice compared to the PA1 mice (Lee, K. et al. 2014). Hence, the variability in Arx protein abundance in the PA2 mouse brains may diminish the power with which to detect significant deregulation of gene expression in some genes significantly altered in the PA1 mice. Our previous report identified expression of a direct transcriptional target of Arx, *Lmo1* in the PA1 mutant mice was deregulated significantly across all stages of embryonic development tested. In contrast, the deregulated expression of this gene did not reach significance in the PA2 mutant mice

compared to wild-type until much later in embryonic development (Lee, K. et al. 2014). Hence, in the subsequent analysis undertaken in this study, we included not only genes significantly deregulated in both mutant animals but also those genes deregulated in either PA1 and PA2 animals.

The deregulated expression of several genes tested was maintained across increasing developmental time points. It is attractive to speculate that perhaps the cells with altered gene expression at E12.5 dpc maintain this altered state as they migrate and develop, despite the reduction in Arx expression with embryonic development. From our current data, we cannot speculate on alterations to the migration of particular neurons impacted by expanded polyA mutant Arx. However, we have conducted our analysis on a very spatially discrete region of the brain, namely the telencephalon at E12.5. We know from a recent elegant study (Marsh et al. 2016) that only a small subset of cells is affected even when there is a complete loss of *Arx* function. Hence, to address if there is migration deficit in these mutant animals extensive and painstaking cellular analysis of interneuron migration will be required. In addition, other approaches such as single cell expression analysis on the different interneuron subtypes would be very powerful and with the increasing technologies available, perhaps increasingly possible.

To date, there are only a handful of known or putative direct targets of ARX. In our study, approximately 5% of the overall deregulated genes in the PA1 and PA2 mice identified as direct ARX targets (known or putative). Knock-out studies (Fulp et al. 2008, Colasante et al. 2009) include phosphatidylinositol-specific phospholipase C, X domain containing 3

gene (*PLCXD3*) which is primarily expressed in the brain and associated with lipid catabolism and signal transduction, along with AF4/FMR2 Family, Member 2 (*Aff2*), Cellular Retinoic Acid Binding Protein 1 *(Crabp1)* and 5'-3' Exoribonuclease 1 *(Xrn1)*. Other direct targets of Arx regulation identified using ChIP based approaches (Quille et al. 2011) that were deregulated in the PolyA Pool group included Histone Deacetylase 4 (*Hdac4*). However, we hypothesise that ARX drives key transcriptional events not only through the impact on specific gene targets but in combination with the regulation of key transcription factors and transcriptional regulators. Hence, we predict that deregulation of other transcription factors and regulators are likely to be contributing to the global changes observed in the PolyA mutant mice.

A number of transcription factors are present in the list of significantly deregulated genes in PA1 and PA2 mice, including *Twist1*, ALX Homeobox 1 (*Alx1)*, SIX Homeobox 2 (*Six2)*, Early B-Cell Factor 1 (*Ebf1)*, Early Growth Response 3 (*Egr3)*, Zinc Finger And BTB Domain Containing 20 (*Zbtb20)* and Zinc Finger Protein 536 (*Zfp536)* which are likely contributing to the downstream deregulation of the transcriptome captured in this study. From our analysis, we propose a 'core' *Arx-Hdac4-Twist1* pathway is likely contributing to the downstream phenotypic outcomes. The pathway indicates that Arx directly represses *Hdac4*, which in turn is involved in chromatin condensation and transcriptional repression (Fischer et al. 2010). Overexpression of *HDAC4* has been shown to impair long-term memory in a drosophila model (Fitzsimons et al. 2013), and overexpression has also been found in patients with autism (Nardone et al. 2014). Salt-inducible kinase 2 (SIK2) an AMP-activated protein kinase, shown to regulate HDAC4 via

phosphorylation, is significantly up-regulated in PA1 and PA2 (although the change in transcript level is subtle). This modification of HDAC4 disrupts the MEF2C-HDAC4 complexes and mediates the activation of MEF2-dependent transcription. *MEF2C* is highly expressed during embryo development and involved in neurogenesis and synaptic function. In mice, it has been indicated to have an essential role in hippocampal-dependent learning and memory by suppressing the number of excitatory synapses and thus regulating basal and evoked synaptic transmission (Barbosa et al. 2008). The hyperexcitability of hippocampal and neocortical networks found in *Arx* PA1 mice is thought to be most likely due to an increase in excitatory drive rather than an inhibitory failure (Beguin et al. 2013). Pathway analysis indicates Arx has a direct interaction with *Mef2c*, although we did not detect significant disturbances in transcript expression in our PolyA mutant mice at 12.5 dpc.

As we travel down the *Arx-Hdac4-Twist1* pathway, likely outside of direct regulation by Arx, we encounter Twist1, a basic-helix-loop-helix transcription factor involved in cell lineage determination & differentiation (Nieto 2013). A wide range of mutations have been reported in *TWIST1* resulting in a variable phenotype from mild to severe intellectual disability as a consequence of large gene deletions however, the majority of missense and nonsense mutations have been associated with the craniosynostosis disorder Saethre-Chotzen syndrome (el Ghouzzi et al. 1997). A recent study using a ChIP approach has identified targets of TWIST1, and comparison with our deregulated gene lists indicates that over one-third of all deregulated genes in PolyA mutant mice identified in our current study are direct or likely targets of TWIST1 regulation, including genes known to cause ID and

epilepsy (Lee, M.P. et al. 2014). The influence of TWIST1 is further demonstrated with the down-regulation of a known target Dnm30s (mir199a/214 cluster) normally upregulated via TWIST1 during development (Lee et al. 2009). Research has focused on determining the role of *Twist1* in cancer development with little emphasis to date of the contribution of *Twist1* in brain development.

The *Arx-Hdac4-Twist1* pathway we propose is one that may reflect very early consequences of ARX loss (or partial loss) of function. For multiple genes deregulated at 12.5 dpc, we have confirmed that the deregulation persists across later stages of embryonic development. Interestingly, genes that are enriched in the proposed pathway contain estrogen response elements (EREs) in their promoter regions. More broadly, when we examine our list of deregulated genes, we find 30% contain EREs (Bourdeau et al. 2004, Tang et al. 2004, Kamalakaran et al. 2005). This is of significant interest given the recent findings that early postnatal treatment with 17 β-estradiol (E2) prevents spasms, restored depleted interneuron populations without increasing GABAergic synaptic density and altered mRNA levels of three downstream targets of ARX (*Ebf3*, *Shox2*, *Lgi1*) in an independent PA1 mutant mouse model (Olivetti et al. 2014). It remains to be established if administration of E2 during early stages of postnatal life, leading to seizure ablation, is impacting upon the transcriptome specifically at the ERE-containing genes we have identified as deregulated in the PolyA mice at early stages of embryonic development.

In summary, we have interrogated the transcriptome of mice modelling the most frequent expansion mutations of PolyA tracts of ARX. Based on our analysis we propose the *Arx-*

Hdac4-Twist1 as the 'core' pathway, which is contributing significantly to the *ARX* mutation-associated phenotypes, namely ID and epilepsy. While other gene targets and factors are likely at play, the *Arx-Hdac4-Twist1* pathway offers a plausible target for future interventions.

Chapter Five:

Investigating the molecular mechanism of

how polyalanine expansion mutations in ARX

lead to a partial loss of function

5 Investigating the molecular mechanism of how polyalanine expansion mutations in ARX lead to partial loss of function

Publications, Presentations and Published Abstracts from this work:

<u>**Accepted first author publication**</u>

- Mattiske, T.R, Tan M.H, Gécz, J, and Shoubridge, C. (2013) 'Challenges of "Sticky" Co-immunoprecipitation: Polyalanine Tract Protein–Protein Interactions', in Hatters, D.M. & Hannan, A.J. (eds), Tandem Repeats in Genes, Proteins, and Disease: Methods and Protocols, Methods in Molecular Biology, vol. 1017, Springer Science+Business Media New York. (DOI: 10.1007/978-1-62703-438-8_9)

Work from this chapter has contributed to the following publication

- Jackson, M.R, Lee, K, Mattiske, T, Jaehne, E.J, Ozturk, E, Baune, B.T, O'Brien, T.J, Jones, N, and Shoubridge, C. (Submitted Jan 2017). Extensive phenotypic evaluation of mouse models recapitulating two common ARX polyalanine expansion mutations which span the clinical spectrum of ID and epilepsy. Neurobiology of Disease.

Conference Abstracts

Talks presented

- Mattiske, T.R, Gécz J and Shoubridge, C (Sept 2013). Expanded polyalanine tract mutations in arx alter interaction with the novel protein partner ubqln4 and increase degradation by the ubiquitin-proteasome pathway. Proceedings of the 16th International workshop on Fragile X & other Early Onset Cognitive Disorders 2013, Barossa, SA.

Posters presented

- Mattiske, T.R, Gécz J and Shoubridge, C (Sept 2012). Do expanded polyalanine tract mutations in ARX alter interaction with the novel protein partner UBQLN4 and disrupt degradation by Ubiquitin-proteasome pathway? Proceedings of the annual ComBio Conference 2012, Adelaide, SA.

5.1 Abstract

Disease-causing polyalanine expansion mutations have been identified in nine genes, eight of which encode transcription factors with important roles in development. *In vitro* studies have shown that expanded polyalanine tracts result in protein misfolding and aggregation. In recent years, significant advances have been made through the analysis of a number of engineered (knock-in) and spontaneous polyalanine expansion mouse models, indicating reduced levels of mutant protein as a contributing factor of disease. As loss of function mutations in *ARX* are associated with a more severe disorder (XLAG), the phenotype in patients with polyalanine expansion mutations suggests a partial loss of function. Further evidence of partial loss of function is provided in chapter 4 with only selective target genes deregulated in our polyalanine expansion mutant mouse models. However it is still not clear what causes this defect. My data indicates this reduced function does not occur through disruptions of binding to DNA or protein interactors in relation to the region of ARX spanning both polyalanine tract 1 and 2. However, I demonstrate a marked reduction in polyalanine mutant protein may be the contributing factor to disease manifestation, reported in both strains at 12.5 dpc (Lee, K. et al. 2014) and confirmed here in adult testes samples. Transcription activity assays indicate ARX may respond in a dose-depend manner and suggests a greater reduction in protein leads to an increase in the severity of the disease. Investigations into the molecular mechanism contributing to this reduction in protein level show no significant change in protein stability (*in vitro*), instead, initial studies indicate inefficiency of translation resulting in reduced protein abundance. This has relevance to other polyalanine tract disorders.

154

5.2   Introduction

Polyalanine tract disorders are the result of an expansion of trinucleotide sequence and give rise to a number of congenital disorders. Expansion of polyalanine tracts results in at least 9 inherited human diseases with eight of these nine diseases due to expansions in transcription factors. In the human proteome, approximately 500 proteins contain polyalanine tracts of greater than 4 residues, with a quarter of these containing a tract of seven or more uninterrupted alanines but not exceeding 20 residues (Lavoie et al. 2003, Albrecht and Mundlos 2005). Thus, it appears that strong structural or physicochemical constraints are imposed on the length of the polyalanine tract. It has been proposed that polyalanine tracts act as flexible spacer elements between functional domains, and lengthening such a spacer could readily perturb the interaction between these domains and their targets either by affecting their orientation or by steric hindrance.

Expansion of polyglutamine tracts also forms the basis of several human diseases, including Huntington's disease and six spinocerebellar ataxias (SCA) types 1, 2, 6, 7, and 17. These diseases are characterised by the expansion of unstable trinucleotide repeats and generally result in neuronal dysfunction from mid-life progressing to severe neurodegeneration. In contrast, neither the penetrance nor severity of the phenotype of the congenital disorders due to expanded polyalanine tracts increases in successive generations, consistent with the observed stability of the expansion leading to alanines. The often large expansions of polyglutamine tracts are thought to contribute to the generation of misfolded protein intermediates that eventually lead to aggregates in susceptible neuronal sub-types, a hallmark of the associated disease states. Therefore, for many years

aggregation of expanded polyalanine tract mutant proteins was predicted to be the mechanism leading to disease. Despite this prediction, the human data for *in vivo* aggregation is not supported. The exception is the non-transcription factor with expanded polyalanine tracts, PABPN1. This is a protein involved in mRNA polyadenylation, in which expanded polyalanine tract mutations cause later onset oculopharyngeal myotonic dystrophy (OPMD) (Brais 2003).

ARX is one of eight transcription factors in which expansions of the polyalanine tracts cause hereditary disease. Over half of all mutations in ARX lead to an expansion of the first and second polyalanine tracts (Marques et al. 2015). The evidence to date leads us to predict that polyalanine expansion mutations in ARX are likely to result in a partial loss of function. In support of this prediction, in chapter 4 of this thesis, analysis of direct and candidate Arx target genes in Arx-WT and polyalanine expansion mutant animals during embryonic brain development demonstrated that there was no blanket loss of transcriptional activity due to the expansion of either PA1 and PA2. Instead, there was a selective loss of transcriptional activity that we propose contributes to the manifestation of disease features. However, the mechanism underpinning this outcome remains unknown. Interestingly, although there was significant overlap between the genes deregulated in the PA1 and PA2 mutant mice the level of deregulation mirrored the phenotypic severity generally noted due to these mutations, as highlighted by the greater magnitude of deregulation in the severe PA1 mice. The increased level of deregulation was correlated with the amount of residual Arx protein observed in embryonic brain samples (16-17% in PA1 compared to 8-50% in PA2 12.5 dpc brains) (Lee, K. et al. 2014).

The correlation between phenotypic severity and expansion size in polyalanine expansion disease suggests that the length of the expansion in the polyalanine tract confers a proportional loss (or gain in the case of autosomal genes) of function on the proteins affected. In the case of ARX, differences in the length of the expansion size are likely to account for variation in severity of the phenotypes. However, what remains less clear is the occurrence of intrafamilial variation seen particularly with the most common PA2 expansion mutation (Turner et al. 2002). As the expansion to polyalanine tracts is stable once expanded, the striking phenotypic variation of affected individuals between families and even within families is difficult to reconcile. To elucidate the pathophysiologic mechanism underpinning the disease phenotype, let alone the variation among affected individuals, a number of fundamental cellular processes need to be explored. ARX has expansions of differing lengths in both polyalanine tract 1 and tract 2 of the protein, and this provides us with the opportunity to investigate aspects of the phenotype severity being impacted by i) the length of the expansion and ii) the position within the protein of the tract in which the expansion occurs. In addition to mice modelling the common ARX polyalanine expansions (PA1 and PA2), we have a range of naturally occurring disease mutations leading to different length tracts in both PA1 and PA2 cloned into a mammalian expression vector with an N-terminal Myc-tag including; 16A tract of PA1 is expanded to PA1-19A, PA1-23A, and PA1-27A; 12A tract of PA2 is expanded to PA2-20A, PA2-21A and PA2-23A (but with a glycine at position 11 of 23) (Figure 5.1 and Table 5.1).

Some of the key questions to be examined include how polyalanine expansion mutations affect the structure of the protein and if these expansions themselves influence the normal transcriptional regulation of target genes. The partial loss of ARX function due to polyalanine expansion mutations may rely upon changes in the protein conformation that could lead to modified interactions of a functional protein domain. In studies of other proteins mutated with expanded polyalanine tracts, the presence of the Hsp90 inhibitor geldanamycin during *in vitro* transcriptional activation assays lead to the recovery of mutant protein activity comparable to the levels measured with the WT protein (Bachetti et al. 2007). Given that Hsp90 inhibitor geldanamycin is known to up-regulate heat shock proteins HSP40 and HSP70, involved in disassembly of intracellular protein aggregates, this finding indicates that aggregation of the mutant protein in the *in vitro* setting was the likely major cause for loss of transcriptional activity. This indicates that the expanded polyalanine tract themselves do not negatively impact the transcriptional regulation. Further support of this retained functionality comes from EMSA studies confirming that binding of the multiple polyalanine mutant proteins to DNA does not seem to be affected (Woods et al. 2005, Nasrallah et al. 2012). Interestingly, in these cases, the functionality of polyalanine expansion mutant protein seems to be suboptimal. We cannot rule out that this could be a consequence of the mutation structurally affecting the interaction with other interactors. However, from our previous data (Lee, K. et al. 2014), we predict there may be a dose-dependent consideration as a consistent feature due to a reduced abundance of mutant protein *in vivo*.

We hypothesise that a reduction in mutant protein abundance may be a common disease mechanism of polyalanine expansion disease. Evidence indicates that the reduction in mutant protein is not a consequence of cell death (Bruneau et al. 2001, Lee, K. et al. 2014), nor is this due to reductions in mRNA levels detected (Innis et al. 2004, Cocquempot et al. 2009, Hughes et al. 2013, Lee, K. et al. 2014). Hence, we propose that two molecular processes may underpin the reduction of the mutant protein level compared to WT; namely lower translation efficiency and/or increased degradation of the expanded polyalanine tract mutant protein.

A consideration when undertaking investigations to address these questions is that studies into the molecular pathogenesis of ARX disorders have been restricted due to the limited capacity to robustly detect endogenous ARX protein, due to the lack of suitable antibodies, coupled with a tightly regulated expression profile *in vivo*. So, although multiple antibodies have been generated against ARX, including several from our laboratory, that are robustly successful in the detection of overexpressed ARX *in vitro*, the detection and manipulation of endogenous ARX protein remains difficult. Hence, we (and others) have chosen to utilise an *in vitro* (overexpression) conditions to undertake selected investigations, mindful of the limitations of type of approach.

Summary

It remains unclear if the pathogenesis of expanded polyalanine tracts is due to the mutation disrupting protein function directly, or if the mutation itself leads to a reduction of the protein driving the disease. It is also unclear if the contribution to the severity of phenotype is due to the polyalanine tract length and/or its location within the protein. Our investigations address if the transcriptional activity of ARX is altered due to the mutations themselves, or is a product of the marked reduction in the mutant protein expression.

Hypothesis: Reduction of ARX polyalanine expansion mutant protein abundance drives disease in a dose depend manner.

Aim: To determine whether the mutation or reduced protein contributes to the partial loss of function and the phenotype severity between PA1-23A and PA2-20A.

Aim: To establish if inefficiency of translation or increase in degradation underpins the reduction of the mutant protein level compared to WT.
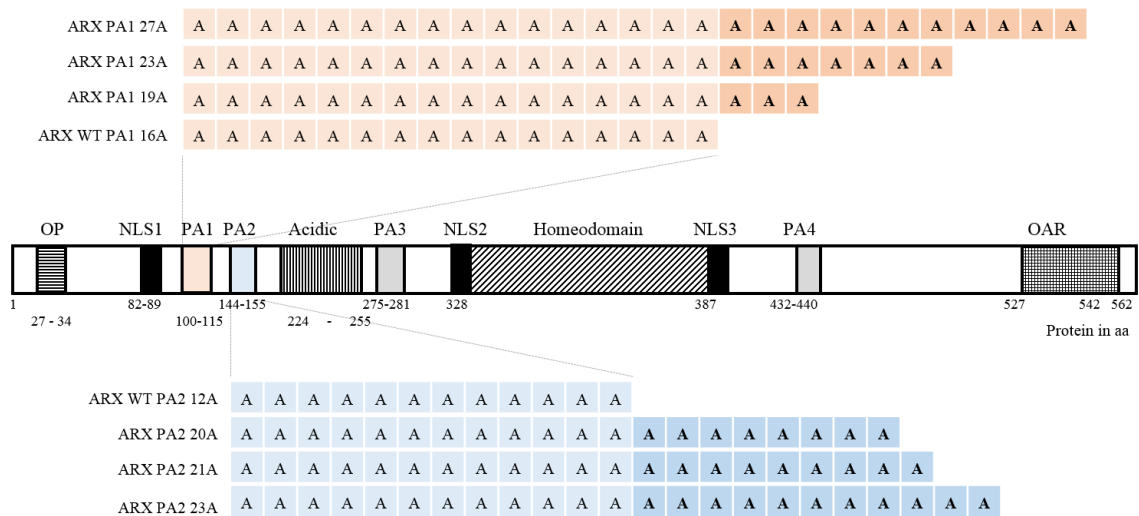
Figure 5.1 Common mutations reported in ARX in polyalanine tract 1 and 2.
Expanded view of the polyalanine tract 1 (orange) and tract 2 (blue) and the addition of alanines (A) due to expansion mutations.

Table 5.1 Summary of the Nomenclature for the Polyalanine Tract Mutation tested in this study.

Polyalanine content variation	HGVS Nomenclature		Referred to in thesis
	cDNA: NM_139058.2	Protein-level: NP_620689.1	
ARX WT PA1 16A			ARX WT
+3A 16>19	c.306GGC[13]	p.(Ala113_Ala115dup)	PA1-19A
+ 4A 16>20	c.306GGC[14]	p.(Ala112_Ala115dup)	PA1-23A
+7A 16>23	c.306GGC[17]	p.(Ala109_Ala115dup)	PA1-27A
ARX WT PA2 12A			ARX WT
+8A 12>20	c.441_464dup	p.(Ala148_Ala155dup)	PA2-20A
+9A 12>21	c.435_461dup	p.(Ala147_Ala155dup)	PA2-21A
+11A 12>23	c.426_458dup	p.(Gly143_Ala153dup)	PA2-23A

161

5.3   Results

5.3.1   Structural impact of polyalanine expansion mutations on the ARX protein

To determine if polyalanine expansion mutations have a possible effect on protein structure, and in particular result in disruption of a functional domain a software package I-TASSER suite was used (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) comparing ARX-WT to the two most frequent polyalanine expansions PA1-23A and PA2-20A. The predicted secondary structure of ARX-WT contains 7 alpha helix structures spread across the protein with 6 corresponding to the location of the known major domains of ARX, the octapeptide domain, acidic domain, homeodomain and the aristaless domain (Table 5.2). Across the homeobox, the main functional domain, the predicted secondary structure is unchanged in the PA2-20A model and a slight deviation seen in the PA1-23A models with the longest alpha helix broken into two shorter ones with a break of 2 coil residues in-between. However, as the location is markedly similar to the WT model, it is not clear if this would have any detrimental effect on the function of the homeodomain. The most noticeable difference between WT and the polyalanine mutations is the second predicted alpha helix (40-43aa) prior to the polyalanine tract 1 and 2 where PA1-23A has a loss of an alpha helix and is replaced with a beta strand. However, for PA2-20A, there is a shift in location from 40-43aa to 83-86aa. As the location of this change does not correspond to any known domain, it is difficult to predict the effect on the protein. Hence, this analysis indicates there is no consistent change in both PA mutants that may account for a similar loss of function.
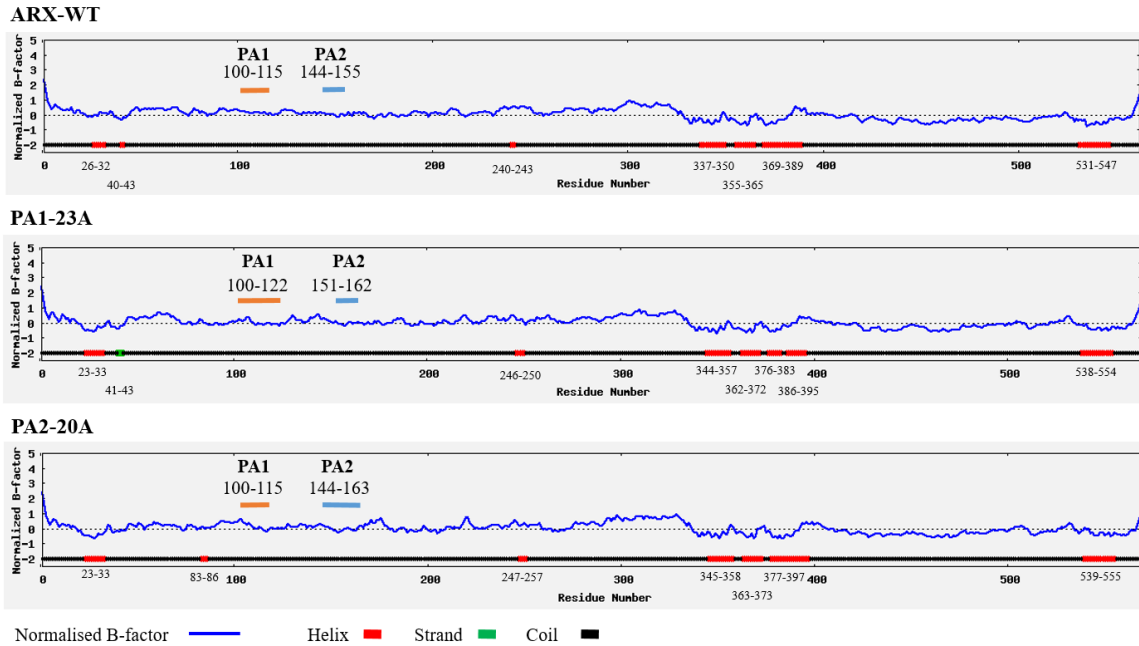
Figure 5.2 I-TASSER predicted secondary structure for ARX–WT and polyalanine tract mutants.
The predicted secondary structure is indicated along the bottom of each graph for ARX WT, PA1-23A and PA2-20A. A black line represents a coil structure with a helix structure shown in red and strand in green. The accuracy of the prediction is demonstrated by the normalised b-factor in blue.

Table 5.2 Summary of the I-TASSER predicted secondary structure for ARX-WT and polyalanine tract mutants and the corresponding protein domains within ARX.
Cells highlight in red=alpha helix, green = beta strand. Numbers italicised indicate the corrected aa number for the length of the polyalanine expansion to correspond to WT.

Location of Functional Domains	WT		PA1-23A		PA2-20A	
	aa	Length (aa)	aa	Length (aa)	aa	Length (aa)
OP domain: 27-34	26-32	7	23-33	11	23-33	11
	40-43	4	41-43	3		
NLS1: 82-89					83-86	4
PA1	100-115	16	100-122	23	100-115	16
PA2	144-155	12	151-162	12	144-163	20
Acidic Domain: 224-255	240-243	4	*239-243*	5	*239-243*	5
Homeodomain: 329-387	337-350	14	*337-350*	14	*337-350*	14
	355-365	11	*355-365*	11	*355-365*	11
	369-389	21	*369-376*	8	*369-389*	21
			379-388	10		
Aristaless domain: 527-542	531-547	17	*531-547*	17	*531-547*	17

163

For each target, I-TASSER simulations generate a large ensemble of structural conformations, called decoys. In this analysis, a C-score is typically in the range of [-5, 2], where C-score of a higher value signifies a model with higher confidence. TM-score and RMSD are estimated based on C-score, and protein length following the correlation observed between these qualities. Similar to the secondary structure predictions, ARX-WT (-0.53) and PA2-20A (-0.52) resulted in similar model predictions and C-score. In contrast, as a consequence of the changes seen in the PA1-23A predicted secondary structure a divergent 3D model to ARX-WT and ARX-PA2-20A is evident (Figure 5.3). In particular, we were interested if any structural impact occurred within the main functional homeodomain critical for DNA binding. Conformational stability of the recognition helix as well as the stability of the entire homeodomain is required for optimal DNA binding. Nucleotide binding sites were reported using S-site (Yang et al. 2013) (http://zhanglab.ccmb.med.umich.edu/COACH/) and listed within the homeodomain WT. With the addition of the polyalanine expansion mutations (+7 for PA1 and +8 for PA2), these sites were not disrupted in either mutant protein (Figure 5.3 and Table 5.3). Taken together, this analysis predicts that the differences in the predicted secondary and tertiary structures of PA1 and PA2 would not affect the binding capacity of the ARX homeodomain in these mutants, but may disrupt the conformational tertiary structure of parts of the protein.
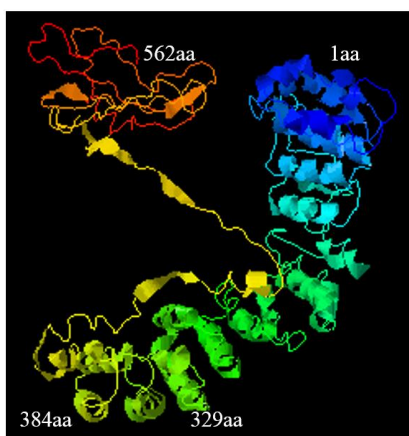
ARX-Wt
Predicted Model 1
C-score = -0.53
Estimated TM-score = 0.65±0.13
Estimated RMSD = 8.8±4.6Å
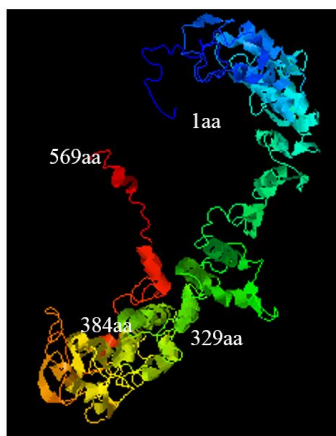
ARX-PA123A
Predicted Model 1
C-score = -1.70
Estimated TM-score = 0.51±0.15
Estimated RMSD = 11.7±4.5Å
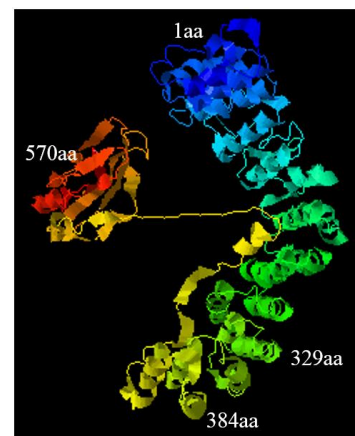
ARX-PA220A
Predicted Model 1
C-score = -0.52
Estimated TM-score = 0.65±0.13
Estimated RMSD = 8.8±4.6Å

Figure 5.3 3D Model Predictions by I-TASSER.
Images of the best-ranked model prediction by I-TASSER for ARX-WT, PA123A, and PA2-20A. The confidence score for estimating the model quality is reported as C-score.

Table 5.3 S-Site Predicted Nucleotide Binding Sites for ARX-WT and Polyalanine Expansion Mutants.

S-Site Results	Predicted Nucleotide Binding Site (aa)
Wildtype 562aa	329-335, 352, 355, 358, 371, 373-375, 377, 378, 380, 382, 384
PA1-23A 569aa	*Corrected (-7aa)* *329- 335, 352, 355, 358, 371, 373- 375, 377, 378, 380, 382, 384*
PA2-20A 570aa	*Corrected (-8aa)* *329- 335, 352, 355, 358, 371, 373- 375, 377, 378, 380, 382, 384*

5.3.2 Functional impact of polyalanine expansion mutations on ARX transcriptional activity.

To correlate the changes to polyalanine tract length and the functional impact on the transcriptional activity of polyalanine expansion mutations in polyalanine tract 1 and 2 of the ARX protein, a cell-based transcriptional assay using luciferase reporter constructs were conducted. This reporter assay models an ARX-binding site in the enhancer region of Lmo1 (Fulp et al. 2008) by cloning three copies of this orthologous region upstream, of the SV40-luciferase promoter (Shoubridge et al. 2012) (Figure 5.4a). Co-transfection of a renilla expression vector was used to normalise the expression of luciferase. Reporter constructs were transiently transfected into HEK293T cells along with a Myc-expression vector (Myc-empty) or fused to full-length ARX vector. The ratio of luciferase to renilla expression for the empty Myc-vector is set to 100% (Figure 5.4b). Compared with this value, the ARX-WT repressed the expression of luciferase to 71% (SEM±9.9), a 30% reduction. A similar level of repression was measured across all PA1 and PA2 mutations. The most common PA1-23A and PA2-20A mutations reflecting the expansions modelled in the mice, both gave very similar levels of repression as ARX-WT, with 75% (SEM±7.9) and 73% (SEM±7.7), respectively. Hence, expansion of either tract did not result in loss of transcriptional capacity of the ARX protein. Using this approach, we were able to assess not only PA1-23A and PA2-20A but a full range of naturally occurring expansions to both tracts to ascertain if length and position of the expansion impacts on the transcriptional activity of the ARX protein. PA1-27A showed the largest loss of repression capacity reducing the expression of luciferase to by only 7% (SEM±11.5) which was not significant compared to Myc-vector. This reflects a 21% difference between PA1-27A and WT. The

remaining mutations showed minor changes in repression compared to WT, ranging from a 1-10% difference and were not significantly different to WT. This result suggests that ARX polyalanine expansion mutant proteins are able to bind to DNA with a similar capacity to ARX-WT and retain the ability to function at similar levels in the context of this assay. This is in agreement with the outcomes of the I-TASSER modelling. Moreover, our analysis indicates that there are no significant differences in transcriptional activity due to which tract harboured the expansion and no significant differences due to the length of the expansion in the context of this cell-based assay.

a)
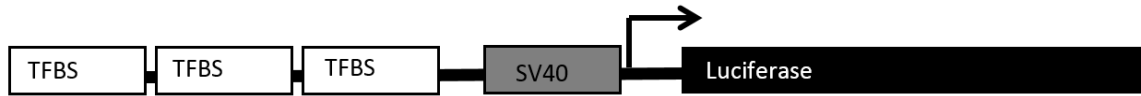
b)
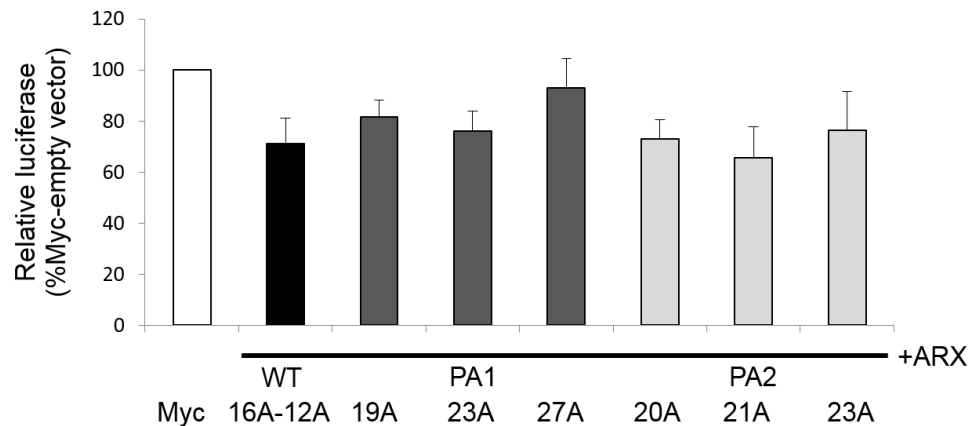
Figure 5.4 Polyalanine expansions mutations in the ARX polyalanine tract 1 & 2 do not affect the repression activity of ARX.
Repression of luciferase expression by ARX is not affected by expansion mutations in polyalanine tract 1 and tract 2. a) The Luciferase reporter assay utilised three copies of the ARX-binding site identified in the enhancer region of *LMO1*, cloned upstream of SV40-luciferase. Polyalanine expansion mutations in the ARX polyalanine tract 1 & 2 have similar levels of repression of *LMO1*, a direct target of ARX transcription repression. b) Luciferase data were normalised to renilla expression and expressed as a percentage relative to empty Myc-vector-transfected cells. Full-length Myc-tagged constructs are listed along the bottom of the graph: ARX WT (PA1-16A, PA2-12A) (white), Polyalanine expansion mutations in Tract1 PA1-19A, PA1-23A, and PA1-27A (dark grey) and in Tract 2, PA2-20A, PA2-21A and PA2-23A (light grey).

To determine if protein function in the context of the reduction in protein amount seen in our *in vivo* studies is causing a decreased level of transcription activity of the ARX protein we repeated the luciferase reporter assay with diluted amounts of plasmid concentration from 500 ng to 125 ng and 30 ng (Figure 5.5a). When the plasmid concentration was decreased to 125ng, ARX-WT repression capacity also decreased from 77% with 500 ng of plasmid to 88%. All repressive capacity was lost (110%) at the lowest concentration of plasmid tested of 30 ng. PA1-23A and PA2-20A exhibited an equivalent loss of repression capacity in a dose-dependent manner to 91% and 84% respectively at 125 ng and with all/most repression lost at 30 ng of plasmid to 121% and 95% respectively. Protein analysis (Figure 5.5b) indicates there was a consistent level of protein for ARX-WT and polyalanine mutants at all concentrations tested. This result supports the idea that polyalanine expansion mutant proteins are still functionally competent with no difference between which tract is mutated. This data also supports that the functional capacity ARX as a transcription factor is dose-dependent (Figure 5.5c).

a)
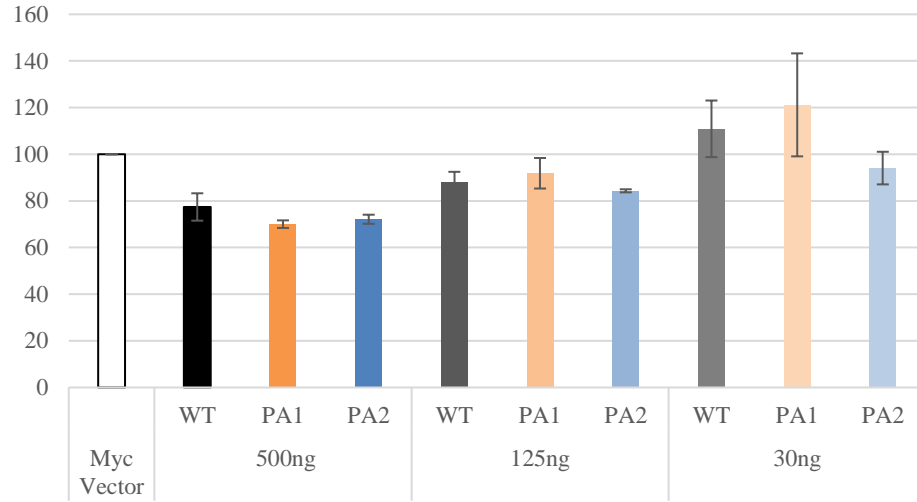
b)
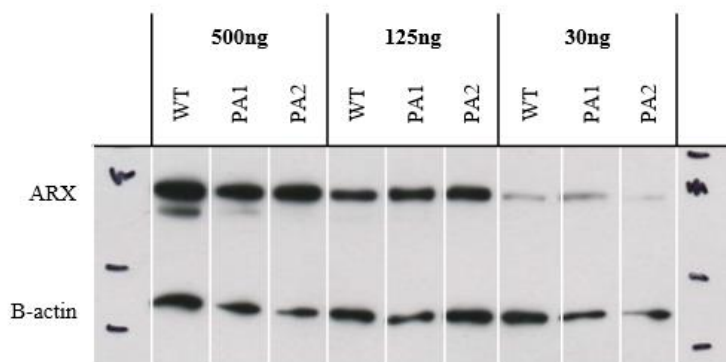
c)
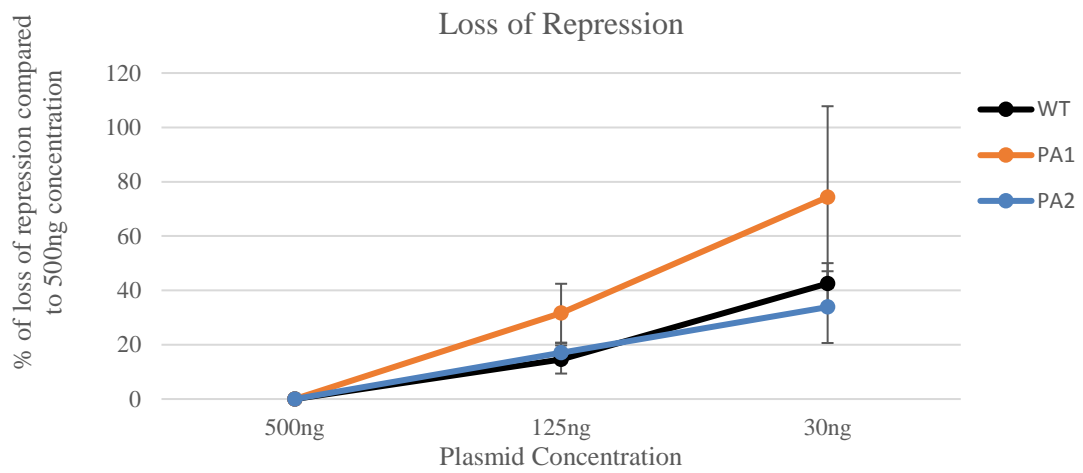
Figure 5.5 Loss of repression is dose dependent in both ARX-WT and polyalanine tract mutant proteins

a) Luciferase data were normalised to renilla expression and expressed as a percentage relative to empty Myc-vector-transfected cells (white). Full-length ARX Myc-tagged WT (Black/grey) and polyalanine expansion mutations in tract 1 PA1-23A (PA1: orange) and in tract 2, PA2-20A (PA2: blue) transfected at 500 ng, 125 ng, and 30 ng. n=4 for all constructs across all concentrations b) Representative image of the level of ARX protein expression between WT and polyalanine expansion mutants across decreased plasmid concentration that were used as inputs for the luciferase assay. c) Loss of repression capacity was calculated as a percentage relative to 500ng of plasmid concentration of each plasmid tested.

5.3.3 Identifying possible protein interactions disrupted by polyalanine tract expansions

As ARX with polyalanine tract expansion mutations retains DNA binding capacity and functions at a comparable level to WT when tested in an *in vitro* context, we next wanted to test if the mutation might structurally interfere the interaction with protein partners. To identify proteins interacting with the polyalanine tract regions that are expanded in disease, we used a GAL4-based yeast-2 hybrid screen. The bait protein contained amino acids 60 to 174 of the mature ARX protein, to specifically focus upon the first two polyalanine tracts (PA1: 100-115 aa) and PA2: 144-155 aa) (Figure 5.6a). This protein did not autoactivate the *HIS3*, *LacZ* or *URA3* reporter genes upon transformation of MaV203 yeast (data not shown). We identified ubiquilin 4 (Figure 5.6b) (UBQLN4: NM_02121.3, NP_064516.2) interacting with this region of ARX. UBQLN4 is also known as A1U; A1Up; UBIN; CIP75; C1orf6 and plays a role in the regulation of protein degradation via the ubiquitin-proteasome system. UBQLN4 has been reported to mediate the proteasomal targeting of misfolded or accumulated proteins for degradation by binding (via UBA domain) to their polyubiquitin chains and by interacting (via ubiquitin-like domain) with the subunits of the proteasome. As the interaction of ARX with UBQLN4 may potentially link the production of ARX to efficient degradation of the proteasome to ensure tight control of the transcription factor, we wanted to investigate the novel interaction between ARX and UBQLN4 further. The yeast 2 hybrid work was undertaken by a previous member of the laboratory with all subsequent follow-up reported here completed by me as part of my PhD studies.
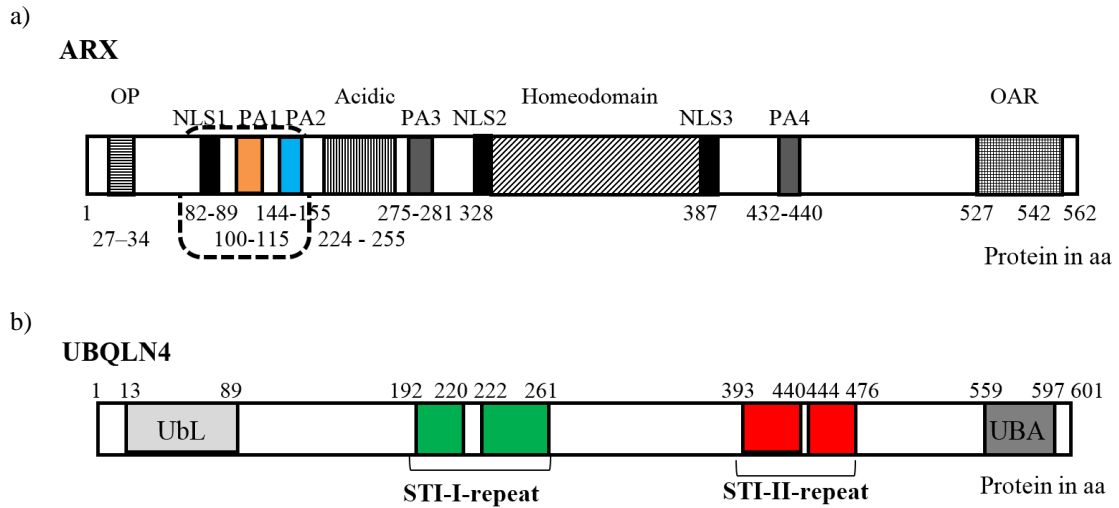
Figure 5.6 UBQLN4 interacts with a region spanning polyalanine tract 1 and 2 of ARX

a) Schematic of human ARX protein. Known functional domains are highlighted in open-reading frame (ORF): octapeptide (OP) horizontally hatched; NLS black rectangles; polyalanine tract 1 (PA1) orange rectangle, polyalanine tract 2 (PA2) blue rectangle, polyalanine tracts 3 & 4 (PA) as shades of grey; acid domain vertically hatched; homeodomain crosshatched and Aristaless domain (OAR) hatched. The dashed box highlights the bait construct spanning the first two PA tracts of ARX (60-174 aa) used in the GAL-4 based yeast-2 hybrid assay. b) Schematic of human UBQLN4 protein with known functional domains highlighted in the ORF: Ubiquilin-like domain (UbL), four stress-inducible, heat shock chaperoning-binding motifs (STI1), and a UBA domain at the C-terminus of the protein.

5.3.4 Confirmation of the novel interaction between ARX-UBQLN4

The interaction between ARX-UBQLN4 was confirmed in mammalian cells by Co-IP. To obtain usable amounts of UBQLN4 for Co-IP, V5-tagged UBQLN4 was transiently transfected into HEK 293T cells along with Myc-tagged ARX-WT or ARX polyalanine expansion mutations. A monoclonal anti-V5 antibody was used to immunoprecipitate V5-UBQLN4 from whole-cell extracts. Myc-tagged ARX co-immunoprecipitating with V5-UBQLN4 was detected using an anti-Myc antibody directly conjugated to horseradish peroxidase (HRP). Polyalanine tracts in the ARX protein were particularly 'sticky' and can cause the protein to bind to the protein-A sepharose in the absence of an immunoprecipitating antibody. Hence, several controls were run in parallel with the test samples to ensure interaction of the proteins was specific. One control to confirm that the binding and stringency of the washing buffers are optimised to detect protein binding specifically to the precipitated UBQLN4 and not binding to protein-A sepharose directly, we include a 'no antibody' control using a cell lysate know to contain ARX protein. Pre-incubation of the protein-A sepharose with non-transfected cell lysate blocks binding of polyalanine tract containing proteins to the protein-A sepharose. When pre-blocked protein-A sepharose is added to the protein-antibody immune complex sample, we get consistent and robust enrichment of our target protein with negligible pull down in our 'No Antibody' control (Figure 5.7a). Cells transfected with Myc-ARX or V5-UBQLN4 alone act as negative controls. V5-UBQLN4 was detected in the V5-UBQLN4 alone control and no pull down was detected in Myc-ARX alone confirming specific co-immunoprecipitation of Myc-ARX only occurs in the presence of UBQLN4 (Figure 5.7a).

As this investigation was carried out *in vitro* we are able to examine a wide range of ARX polyalanine expansion mutations to include 3 polyalanine tract 1 mutations with the addition of +3A (PA1-19A), +7A (PA1-23A) and +11A (PA1-27A), and 3 polyalanine tract 2 mutations with the addition of +8A (PA2-20A), +9A (PA2-21A) and +11A (PA2-23A) (Figure 5.1 and Table 5.1). This allows us to determine if ARX protein with expanded tracts in general binds or does not bind to UBQLN4 and assess if there are any differences due to the tract or length of the expansion in either tract 1 or tract 2. When we tested the interaction of ARX with expanded polyalanine tracts we observed Co-IP with UBQLN4 for all mutations tested, but not in parallel samples in which no IP antibody was added (Figure 5.7a). Our data indicates that UBQLN4 interacts with the region containing the first two polyalanine tracts in ARX and that this interaction is not abolished by any of the polyalanine expansion mutations. However, relative to ARX-WT, we measured lower levels of all mutant proteins interacting with UBQLN4, suggesting expanded polyalanine mutations may reduce the interaction with the novel partner, UBLNQ4 (Figure 5.7b).

We investigated the subcellular localisation of these interacting proteins in mammalian cells. Myc-ARX WT demonstrates a discreet localisation to the nucleus (Figure 5.8), as expected for a transcription factor and in agreement with our previous data (Fullston et al. 2011). Transfected V5-UBQLN4 was detected as diffuse signals across the cell cytoplasm and nucleus in non-transfected cells (data not shown) and was not altered when co-expressed with ARX-WT. This likely reflects the dynamic and transient nature of this interaction. A similar outcome is seen for all polyalanine tract expansion mutations when localised correctly to the nucleus as seen for WT (data not shown). However, considering

UBQLN4 seems to localise to aggregates containing mutant ARX protein *in vitro* (Figure 5.8), the amount of soluble, correctly localised ARX protein will be reduced compared to wild-type ARX protein. In this context, the reduced amount of mutant protein (by Co-IP) being pulled down by UBQLN4 may actually reflect an increased interaction with the lower amount of soluble mutant protein available.

a)

b)
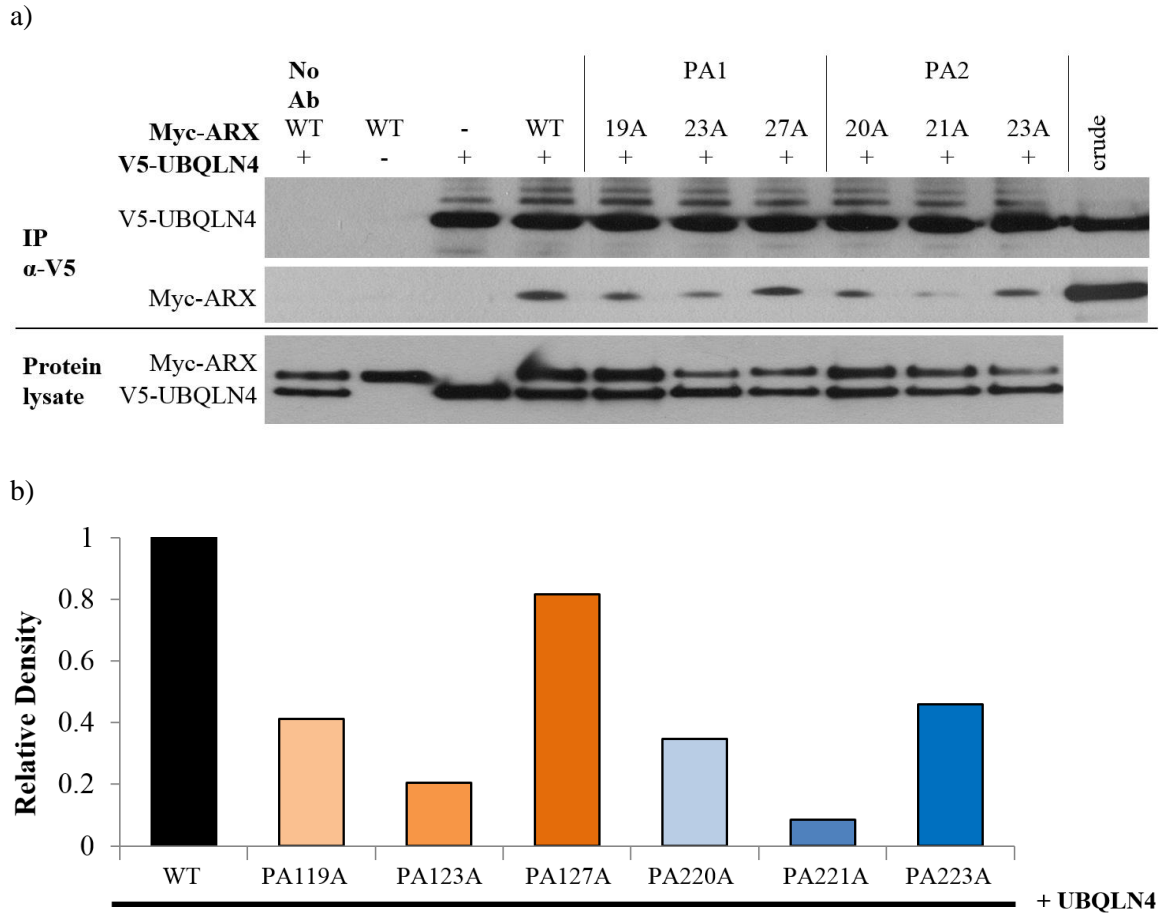
Figure 5.7 Confirmation of interaction between ARX and UBQLN4 which is not loss due to polyalanine expansion mutations

a) HEK293T cells transfected with Myc-ARX-WT or PA mutant constructs and V5-UBQLN4 were lysed and protein immunoprecipitated (IP) with antibodies against the V5. Samples were loaded onto 4%-12% SDS-PAGE gels and analysed for the presence of co-IP proteins. Detection of Myc-ARX proteins bound to V5-UBQLN4 by immunoblotting with mouse anti-Myc HRP conjugated antibody. All samples transfected with V5-UBQLN4 showed a protein band of the predicted size upon blotting with anti-V5 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-UBQLN4, but no IP antibody added (*). Cells transfected with Myc-ARX alone or V5-UBQLN4 alone was used as negative controls. V5-UBQLN4 (~64 kDa) and Myc-ARX (~ 62 kDa) are both present in protein lysates (bottom panel). b) Densitometry of Myc-ARX bands using NIH ImageJ program. The density of the target protein was normalised to the density of ARX WT and is expressed in arbitrary units. A representative image from 2 experiments.

a)
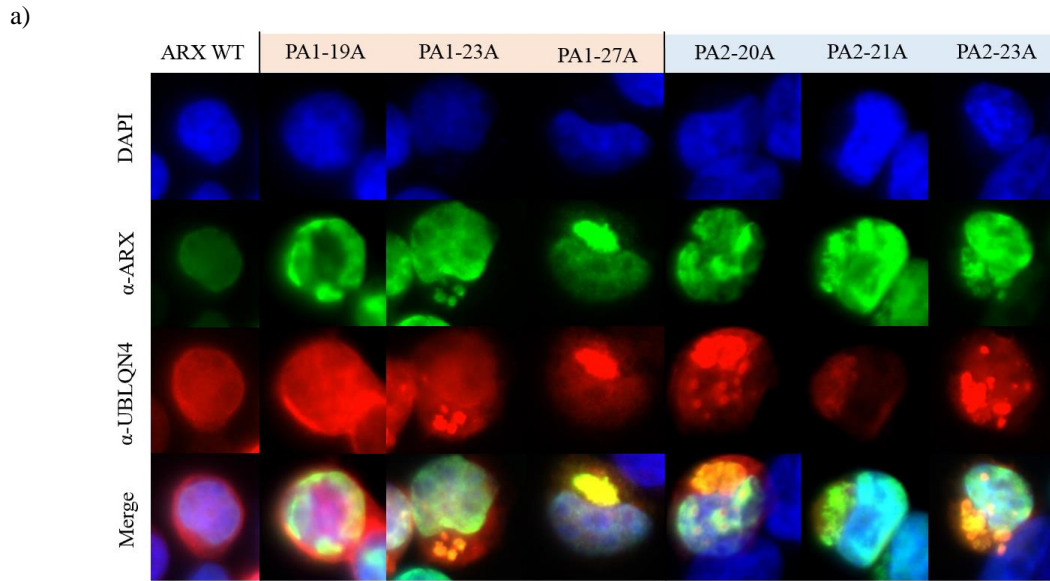
Figure 5.8 Subcellular location and co-localisation of ARX and UBQLN4.
a) Representative images of cells transfected with Myc-ARX-WT with normal ARX nuclear staining and Myc-tagged polyalanine expansion constructs, PA1-19A, PA1-23A, PA1-27A, PA2-20A, PA2-21A and PA2-23A show aggregates of mutant ARX form both inside the nucleus and outside in the cytoplasm. The nuclear material is indicated by the blue DAPI signal. Nuclear localisation of ARX protein overexpressed in HEK293T cells was detected using a mouse monoclonal anti-ARX and DonkeyαMouse IgG Alexa 488 (green), and UBQLN4 localised to the cytoplasm was stained using a rabbit anti-V5 and GoatαRabbit Cy3 (Red). ARX mutant protein formed aggregates and mislocalised to the cytoplasm. UBQLN4 was detected to co-localised with the outside of the ARX aggregates in the cytoplasm.

5.3.5   Alanine expansion mutations result in a reduction of mutant protein *in vivo*.

Polyalanine expansion mutations in tract 1 and 2 modelled in mice have been shown to result in a reduction in mutant protein in the embryonic brain (Lee, K. et al. 2014). However, it has not been established if this reduction in mutant protein is a consistent feature across other tissues nor if this persists into adult life. Arx is highly expressed in the brain during embryonic development and then markedly downregulated in post-natal life, and persists in a small number of neurons spread throughout the cortex during postnatal and adult life (Poirier et al. 2004). Arx is also expressed in skeletal muscle, pancreas and testis (Kitamura et al. 2002, Ohira et al. 2002, Collombat et al. 2003). To assess if the mutant protein is reduced compared to WT as seen in the brain in an alternative tissue, testis samples were used as Arx expression is known to continue into adult life (Yu et al. 2014). Moreover, this analysis would allow us to better examine the relationship in the reduction of protein expression between the two PA tract mutations.

We wanted to initially establish the level of Arx-WT protein in the testes of our mice during adulthood, and ensure there was no change in the protein levels across developmental ages of interest. We tested samples at postnatal day 40 as this was a significant time that we see the onset of seizures and day 70 representative of young adult life, at which time we have undertaken behavioural testing (Jackson et al. Submitted 2017). Samples were analysed by immunoblot (Figure 5.9a) and subsequent densitometry of resulting Arx and B-actin bands were determined using LICOR western blot analysis software (Image Studio Lite) (Figure 5.9b). The density of Arx protein was normalised to the loading control B-actin and is expressed in arbitrary units. To compare across a number of immunoblots, all analysis was

normalised to sample #1322 (Table 5.4). Our data indicates that ARX is robustly expressed in adult testes samples and the abundance of Arx protein was not significantly different between the testes of Arx-WT mice collected at day 40 (n=3) compared to samples collected at day 70 (n=10), namely 94.2% (SD±7.8) and 114.6% (SD±27.2), respectively (Table 5.5 and Figure 5.10).

a)

b)
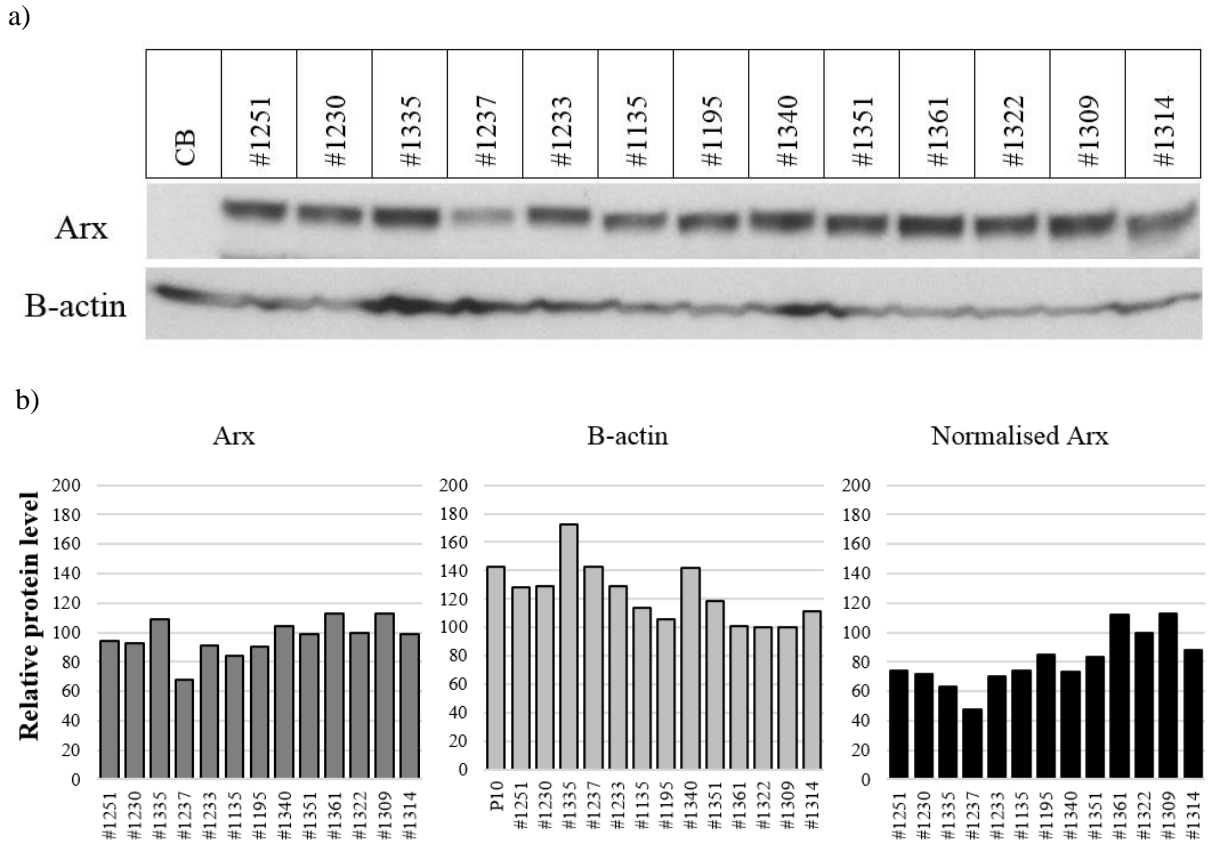
Figure 5.9 Robust protein abundance in Arx-WT testes samples during adult life.
a) Western immunoblots showing robust Arx protein abundance across Arx-WT testes samples from adult mice ranging in age from day 40 and day 70. Samples were loaded onto 4-12% SDS-PAGE gels, and Arx was detected by immunoblotting with rabbit anti-Arx. B-actin was used as a loading control. CB: P10 cerebellum was used as a negative control as this region lacks any Arx protein. b) Densitometry of Arx and B-actin bands were determined using LICOR western blot analysis software, Image Studio Lite. The density of Arx protein was normalised to the density of the loading control B-actin and is expressed in arbitrary units and all normalised to sample #1322.

Table 5.4 Densitometry and normalisation results for the level of Arx protein in Arx-WT mice testes samples.

Sample ID	Age (Days)	Arx Signal	NORM Arx	B-actin Signal	NORM B-actin	Arx/ B-actin	NORM Arx/ B-actin
P10	10		236000	143.0			
#1251	40	94.5	211000	127.9	1.46	73.9	94.5
#1230	42	93.0	213000	129.1	1.42	72.0	92.9
#1335	42	108.6	284000	172.1	1.25	63.1	108.6
#1237	69	67.8	235000	142.4	0.94	47.6	67.8
#1233	69	90.8	213000	129.1	1.39	70.3	90.8
#1135	70	84.1	188000	113.9	1.46	73.8	84.1
#1195	70	90.2	175000	106.1	1.68	85.0	90.2
#1340	70	104.3	234000	141.8	1.45	73.5	104.3
#1351	70	98.5	195000	118.2	1.65	83.3	98.5
#1361	70	112.9	166000	100.6	2.22	112.2	112.9
#1322	71	100.0	165000	100.0	1.98	100.0	100.0
#1309	73	113.2	165000	100.0	2.24	113.2	113.2
#1314	73	98.5	184000	111.5	1.74	88.3	98.5

Table 5.5 Average level of Arx protein in day 40 and day 70 WT mice show no significant difference.

	No. Samples	Ave	SD
D40	3	94.3	7.8
D70	10	114.7	27.2
All	13	110.0	25.4
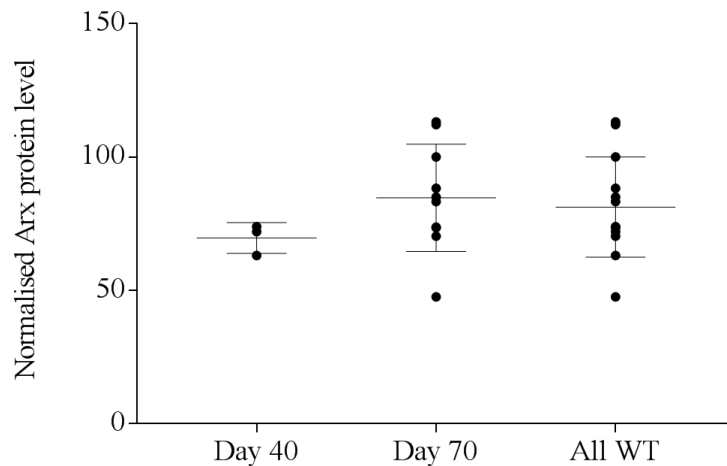
Figure 5.10 Arx protein abundance across day 40 and day 70 age groups.
Normalised densitometry of Arx protein levels detected via western immunoblot analysis in Arx-WT testes samples.

Much of the research into ARX function, and the impact of the PA mutations, has been focused on embryonic stages of development with only limited studies and knowledge of phenotype during postnatal life. Indeed, the phenotype of the PA2 mouse model (Kitamura 2009) that we are using has not been reported yet. Our laboratory has undertaken extensive phenotyping studies comparing PA1 and PA2 mice across postnatal development. Although this work is outside the scope of my thesis, in brief, we find that the PA2 mice have extensive seizure phenotypic and behavioural deficits similar, albeit less severe than the PA1 mouse (Jackson et al. Submitted 2017). The morbidity and mortality of the mice mean that collecting large cohorts at extended postnatal ages in challenging. However, as part of this larger study, I was able to collect testes material from animals from the time we see the onset of seizures ($\sim$ day 40) through to the time points of early adult life at which behavioural testing and end points were collected $\sim$ day 70. Testes protein was extracted from Arx-WT (n=13), Arx-PA1 (n=11) and Arx-PA2 (n=10) mice and the level of protein abundance were measured by immunoblot analysis. Testes from PA1 mice showed a marked reduction in Arx protein levels across all samples tested (Figure 5.11). When the level of protein was normalised and compared to Arx-WT levels, a reduction of between 30-60% was determined (Table 5.6). The reduction in Arx protein abundance wasn't as marked in PA2 testes samples (Figure 5.12) with a reduction of between 3-40% measured (Table 5.7). The average of all the Arx-WT samples was 96.63% (SEM±3.44). Both mutant groups show a clear reduction in the level of Arx protein abundance with the average level in PA1 testes at 59.99% (SEM±3.97) and PA2 at 79.78% (SEM±3.17), respectively compared to Arx-WT protein levels. (Table 5.8 and Figure 5.13a). The 40% reduction between Arx-WT and PA1 at $P<0.0001$ was highly significant, with the 20%

reduction between Arx-WT and PA2 also significant at P=0.0062. The 20% difference in Arx protein abundance was also significant between PA1 and PA2 at P=0.002. To ensure that we did not have a difference in the total protein abundance between genotypes due to different proportions of animals at varying ages of postnatal life, we also examined the protein levels for each animal relative to postnatal age. By this analysis, the level of reduction was consistent across all genotypes at differing ages ranging from day 40 to day 70 (Figure 5.13b). This result confirms the reduction in protein in a different tissue and suggests it would persist across the lifetime of the animal. In accordance with our previous study (Lee, K. et al. 2014), protein reduction corresponds to the phenotype severity and deregulation of targets seen in chapter four of this thesis. As no functional difference between the two polyalanine tract mutations has been found, it highlights reduced levels of polyalanine tract mutation protein to be the likely driver of the disease.

a)
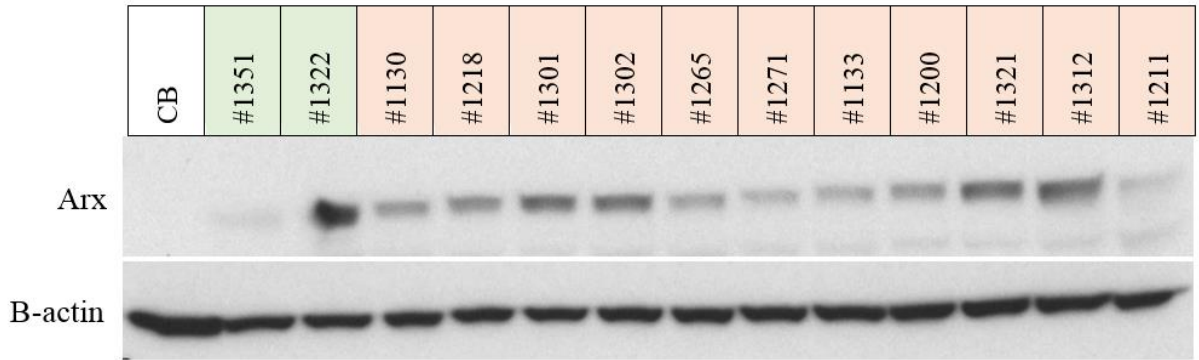
b)
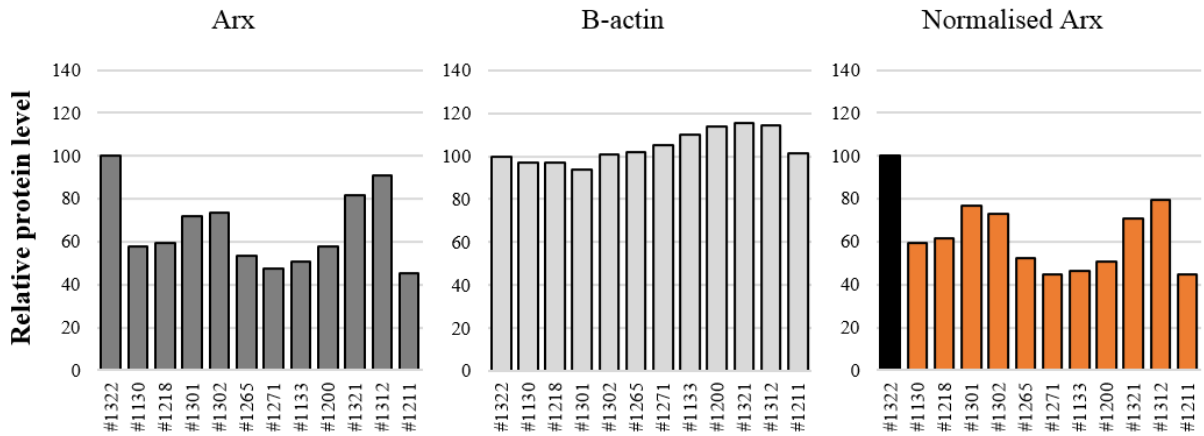
Figure 5.11 Reduced Arx protein abundance in PA1 testes samples during adult life compared to Arx-WT mice

a) Western immunoblots showing reduced Arx protein abundance across PA1 testes samples (highlight in orange) from adult mice ranging in age from day 40-70 compared to the Arx-WT sample #1322 (highlight in green). Samples were loaded onto 4-12% SDS-PAGE gels, and Arx was detected by immunoblotting with rabbit anti-Arx. B-actin was used as a loading control. CB: P10 cerebellum was used as a negative control as this region lacks any Arx protein. b) Densitometry of Arx and B-actin bands were determined using LICOR western blot analysis software, Image Studio Lite. The density of Arx protein was normalised to the density of the loading control B-actin and is expressed in arbitrary units and all normalised to sample #1322 (black).

Table 5.6 Densitometry and normalisation results for the protein abundance of Arx in PA1 mice testes samples.

Genotype	Sample ID	Age (Days)	Arx Signal	NORM Arx	B-actin Signal	NORM actin	B- Arx/ B-actin	NORM Arx/ B-actin
WT	#1322	71	630000	100.0	698000	100	0.90	100.0
PA1	#1130	43	363000	57.6	677000	97.0	0.54	59.4
PA1	#1218	49	376000	59.7	678000	97.1	0.55	61.4
PA1	#1301	59	454000	72.1	656000	94.0	0.69	76.7
PA1	#1302	59	463000	73.5	702000	100.6	0.66	73.1
PA1	#1265	61	337000	53.5	712000	102.0	0.47	52.4
PA1	#1271	61	298000	47.3	734000	105.2	0.41	45.0
PA1	#1133	63	320000	50.8	768000	110.0	0.42	46.2
PA1	#1200	70	363000	57.6	793000	113.6	0.46	50.7
PA1	#1321	71	514000	81.6	807000	115.6	0.64	70.6
PA1	#1312	73	573000	91.0	798000	114.3	0.72	79.6
PA1	#1211	75	287000	45.6	709000	101.6	0.40	44.8

a)

b)
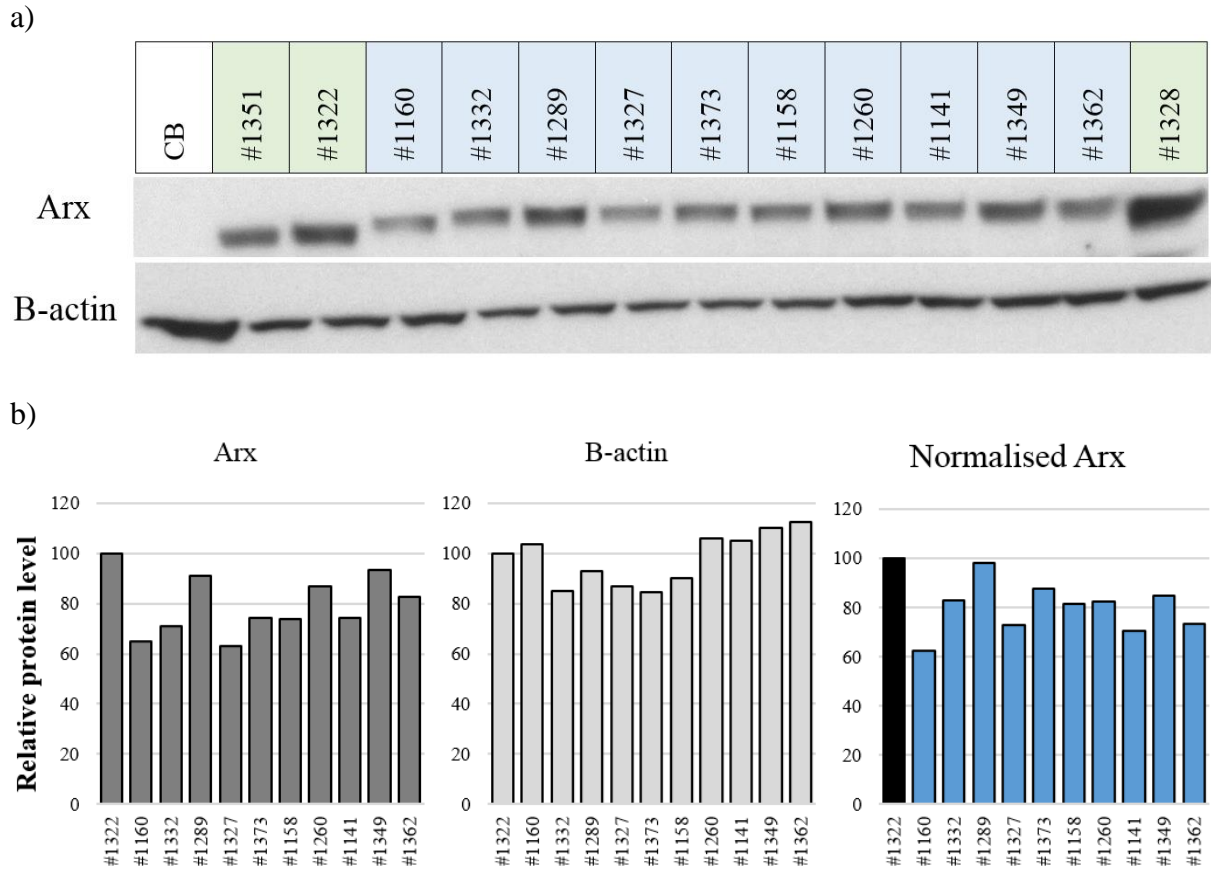
Figure 5.12 Reduced Arx protein levels in PA2 testes samples during adult life compared to Arx-WT mice.

a) Western immunoblots showing reduced Arx protein abundance across PA2 testes samples (highlight in orange) from adult mice ranging in age from day 40-70 compared to the Arx-WT sample #1322 (highlight in green). Samples were loaded onto 4-12% SDS-PAGE gels, and Arx was detected by immunoblotting with rabbit anti-Arx. B-actin was used as a loading control. CB: P10 cerebellum was used as a negative control as this region lacks any Arx protein. b) Densitometry of Arx and B-actin bands were determined using LICOR western blot analysis software, Image Studio Lite. The density of Arx protein was normalised to the density of the loading control B-actin and is expressed in arbitrary units and all normalised to sample #1322 (black).

Table 5.7 Densitometry and normalisation results for protein abundance of Arx in PA2 mice testes samples.

Genotype	Sample ID	Age (Days)	Arx Signal	NORM Arx	B-actin Signal	NORM actin	B- Arx/ B-actin	NORM Arx/ B-actin
WT	#1322	71	653000	100.0	548000	100	1.19	100.0
PA2	#1160	40	423000	64.8	567000	103.5	0.75	62.6
PA2	#1332	42	462000	70.8	466000	85.0	0.99	83.2
PA2	#1289	44	593000	90.8	508000	92.7	1.17	98.0
PA2	#1327	49	412000	63.1	475000	86.7	0.87	72.8
PA2	#1373	49	486000	74.4	464000	84.7	1.05	87.9
PA2	#1158	53	481000	73.7	494000	90.1	0.97	81.7
PA2	#1260	67	568000	87.0	579000	105.7	0.98	82.3
PA2	#1141	70	486000	74.4	576000	105.1	0.84	70.8
PA2	#1349	70	609000	93.3	603000	110.0	1.01	84.8
PA2	#1362	70	540000	82.7	615000	112.2	0.88	73.7
WT	#1328	71	978000	149.8	637000	116.2	1.54	128.8

Table 5.8 Summary of normalised protein abundance of Arx protein levels in testes samples from WT and polyalanine expansion mice

	WT	PA1	PA2
	94.5	59.4	62.6
	93.0	61.4	83.2
	108.6	76.7	98.0
	67.8	73.1	72.8
	90.8	52.4	87.9
	84.1	45.0	81.7
	90.2	46.2	82.3
	104.3	50.7	70.8
	98.4	70.6	84.8
	112.9	79.6	73.7
	100.0	44.9	
	113.2		
	98.5		
Mean	96.6	60.0	79.8
Std. Error of Mean	3.44	3.97	3.2
Adj P Value (WTvPA1/PA2)		<0.0001	0.0062
Adj P Value (PA1vPA2)			0.002
Normalised Mean	100	62.1	82.6

a)
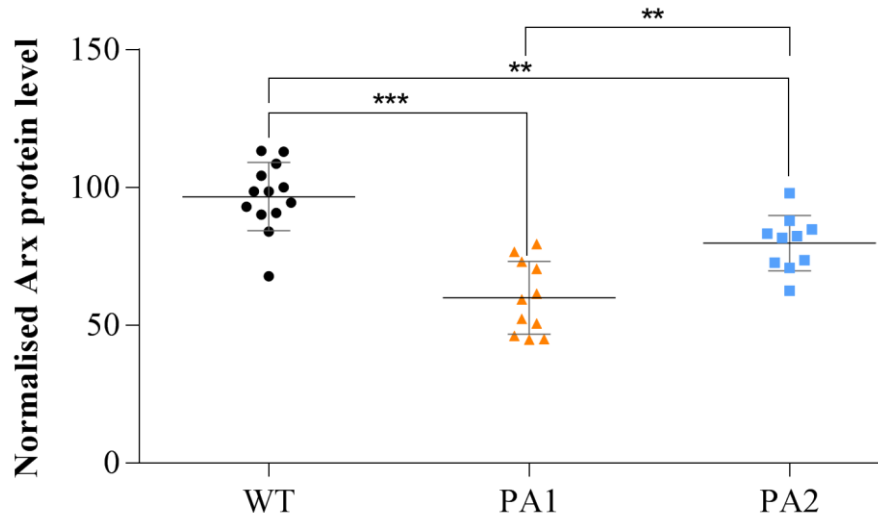
b)
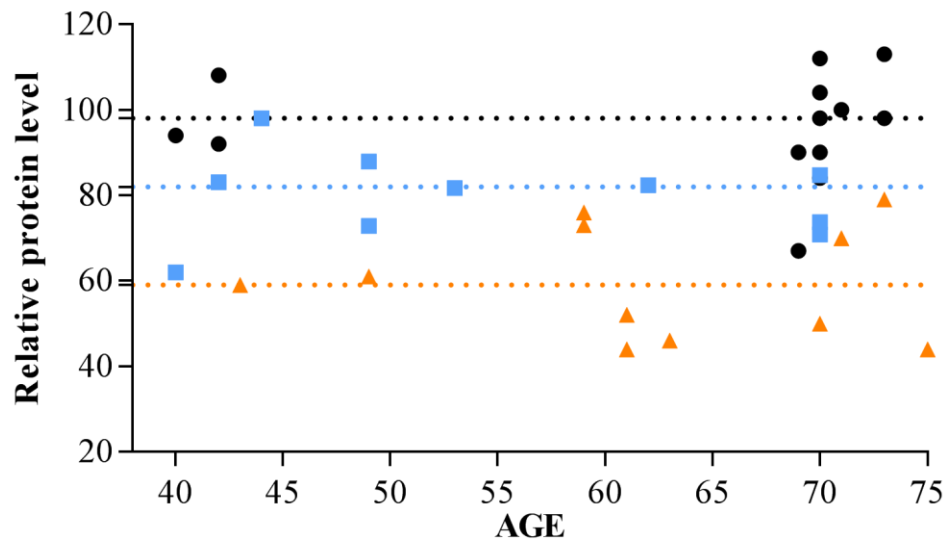
Figure 5.13 Reduction of Arx polyalanine expansions mutant protein in adult testes samples across adult life
a) Summary of normalised protein abundance of Arx protein levels in testes samples from WT (black circles) and polyalanine expansion mice, PA1 (orange triangle) and PA2 (blue square). *** = P <0.0001, ** = P<0.005 (Tukey's multiple comparison test). b) The level of relative Arx protein abundance across different ages ranging from day 40 to day 70 in Arx-WT, Arx-PA1, and Arx-PA2 samples.

5.3.6 Polyalanine expansion mutation effect on degradation of the mutant protein.

Reduction of mutant protein has now been established as a consistent feature across a number of other polyalanine tract mutations whilst the mRNA levels remain unaffected (Hughes et al. 2013, Lee, K. et al. 2014). Possible mechanisms to account for the reduction of mutant protein could be due to decrease stability or increased degradation of the mutant protein compared to the wild-type protein. Hence, we have investigated this possibility using a pulse-chase assay (Li et al. 2008). This experimental procedure allows newly made protein over a short period of time to be labelled within the cell, then sequentially exposed to the same compound, this time unlabeled, allowing the protein of interest to be followed from synthesis to degradation in its cellular environment. HEK293T cells were transfected with 500 ng of either Myc-tagged ARX-WT, PA1-23A or PA2-20A and labelled with ^{35}S methionine/cysteine for 1 hour, followed by a chase of 1 or 3 hours without the radiolabel before harvesting the cells. The two most common mutations were chosen to represent mutations in tract 1 and 2 as we see reasonably consistent data between the differing lengths in luciferase transcription studies. After 1 hour, ARX-WT had 72% of protein remaining when compared to the respective 0 hours sample which is set to 100%. The half-life of ARX-WT protein was reached at 3 hours with 49.5% measured (Figure 5.14a). At 1 and 3 hours, PA2-20A showed a similar trend with 85% and 53% of protein remaining. Over the first hour, PA1-23A showed an increased rate of degradation compared to ARX-WT with 57% of protein remaining. However, by 3 hours the rate of degradation plateaued with a similar level of measured protein at 55% remaining (Figure 5.14a). When the rate of degradation of PA1-23A and PA2-20A is directly compared to ARX-WT as the change in proportion of remaining protein when ARX-WT is set to 100% at each time point there is

no significant change of the rate of clearance for PA2-20A however at 1 hour PA1-23A is significantly different with ~30% less protein then ARX-WT indicating increase clearance (Figure 5.14b). However, by 3 hours ARX-WT and mutant proteins have plateaued and showed similar levels. This could be due to the inherent difficulties using an overexpression *in vitro* system where overexpression of ARX polyalanine expansion mutants increased the propensity to form aggregates (Figure 5.8) which can also be seen in ARX-WT to a lesser extent. It is likely the formation of aggregates renders the proteins resistant to proteasomal degradation as no further degradation was detected past 3 hours (data not shown – 6h time point).

a)

b)
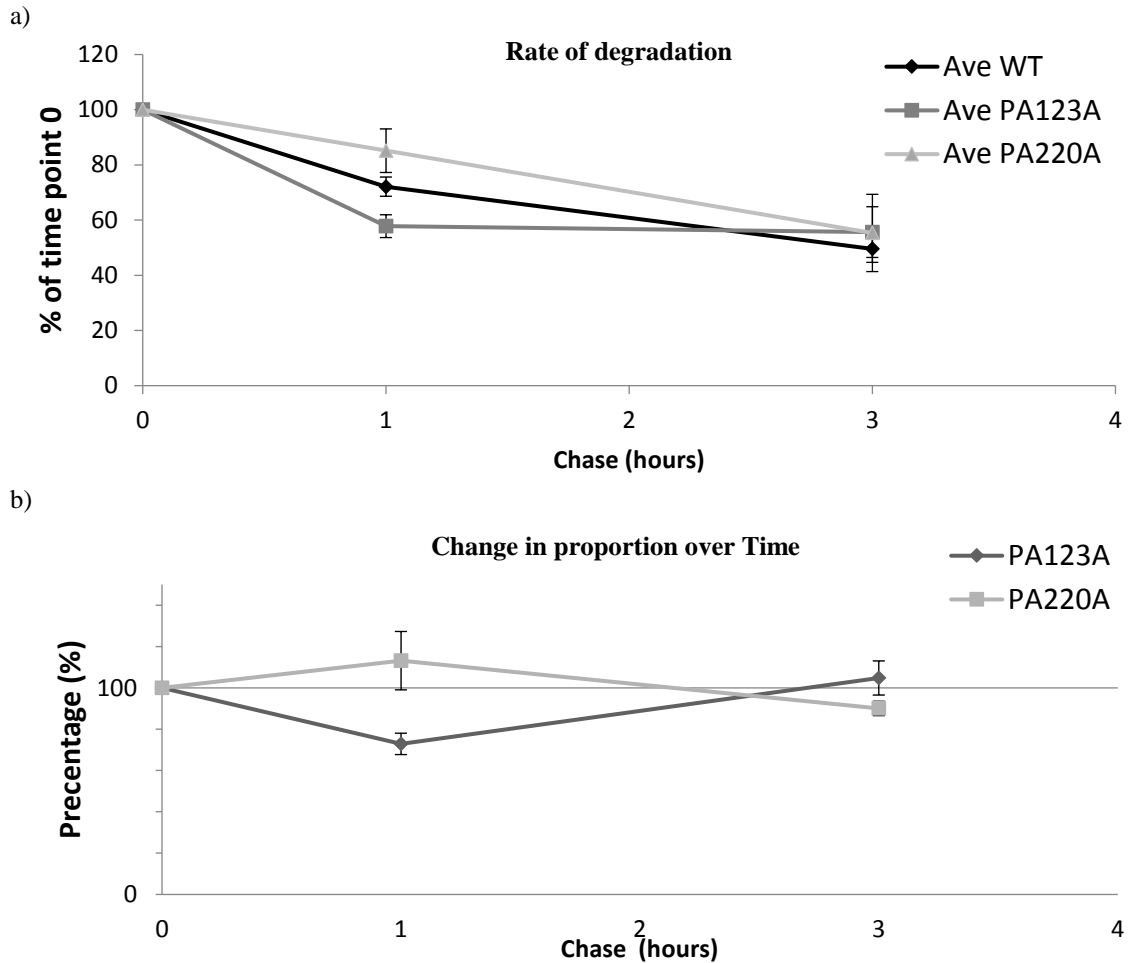
Figure 5.14 Protein stability of ARX polyalanine expansion mutant proteins.
HEK293T cells were transiently transfected with ARX-WT, PA1-23A, and PA2-20A, incubated for 24 h, labelled with [35S]methionine/cysteine for 1 h, followed by a chase of 1 or 3 h. Cell lysates were immunoprecipitated with anti-Myc, and ^{35}S levels were measured using a beta scintillation plate counter. a) The rate of degradation of ARX-WT, PA1-23A, and PA2-20A when each sample is compared to their respective 0hr sample which is set to 100%. B) The change in the proportion of mutant protein when compared to ARX-WT at the same time point. The data represent the means of at least two independent experiments ± S.E. The y-axis represents the percentage of the radiolabelled ARX remaining after the 1- or 3-h chase. (WT n=6, PA1-23A n=6, PA2-20A n=2)

192

5.3.7 Decreased rate of translation efficiency may be contributing to the reduction of polyalanine expansion mutant protein.

Considering that there was no significant change in the stability of polyalanine mutant protein as determined in the pulse-chase assay (Figure 5.14), we predict that stretches of GC-rich regions due to expanded polyalanine tract mutations in ARX may impede translation of the mRNA to protein. This may be the mechanism to account for the reduced levels of mutant protein seen *in vivo* (Lee, K. et al. 2014). Stretches of GCG sequences are known to be difficult to be transcribed and have an increased propensity to form highly stable mRNA hairpin loops. We predict that these factors may lead to decreased translation efficiency via ribosomal stalling. As an initial investigation, we conducted an *in vitro* time course experiment monitoring the rate of protein production in transfected HEK293T cells with PA1-23A compared to ARX-WT. We demonstrate that when cells were transfected with 0.5 µg of plasmid, protein production was first detected via immunoblot after 8-hours post-transfection with Myc-tagged ARX-WT plasmid (data not shown). Therefore, in subsequent tests, ARX-WT and polyalanine mutant samples were collected at 8, 10, 12, 16 and 22-hours post transfection. HEK293T cells were transfected with 0.5ug of plasmid and confirmed to have a similar transfection efficiency of around 50% (data not shown). Protein was analysed by immunoblot, and subsequent densitometry (Image Studio Lite) was used to calculate the rate of protein production across time. Samples were compared to the initial time point of 8-hours which was given the arbitrary unit of 1. Using this approach, by 12-hours post-transfection, ARX-WT protein had increased 4-fold. ARX-WT continued to increase from 12-hours across to 16 and 22-hours by 1.5 and 1-fold respectively. This was compared to a much lower increase in protein production for the PA1-23A which showed

little or no increase in the rate of production across 8-12 hours (Figure 5.15). By 16-hours

post-transfection PA1-23A started to increase with a 2.5-fold increase, with a similar level

also detected at 22-hours. Therefore, if ARX-WT is considered as 100% maximal

production rate at each time point, the PA1-23A mutant displayed a 67% lower production

levels with an efficiency of only 37.8%, which may be consistent with a disturbance

leading to decreased efficiency of translation. Hence, this analysis reveals another possible

mechanism contributing to the reduction in polyalanine expansion mutant protein *in vivo*.

Hours	ARX-WT	PA1-23A
8	1	1
10	1.35	1.01
12	4.15	0.95
16	6.87	2.52
22	7.10	2.69
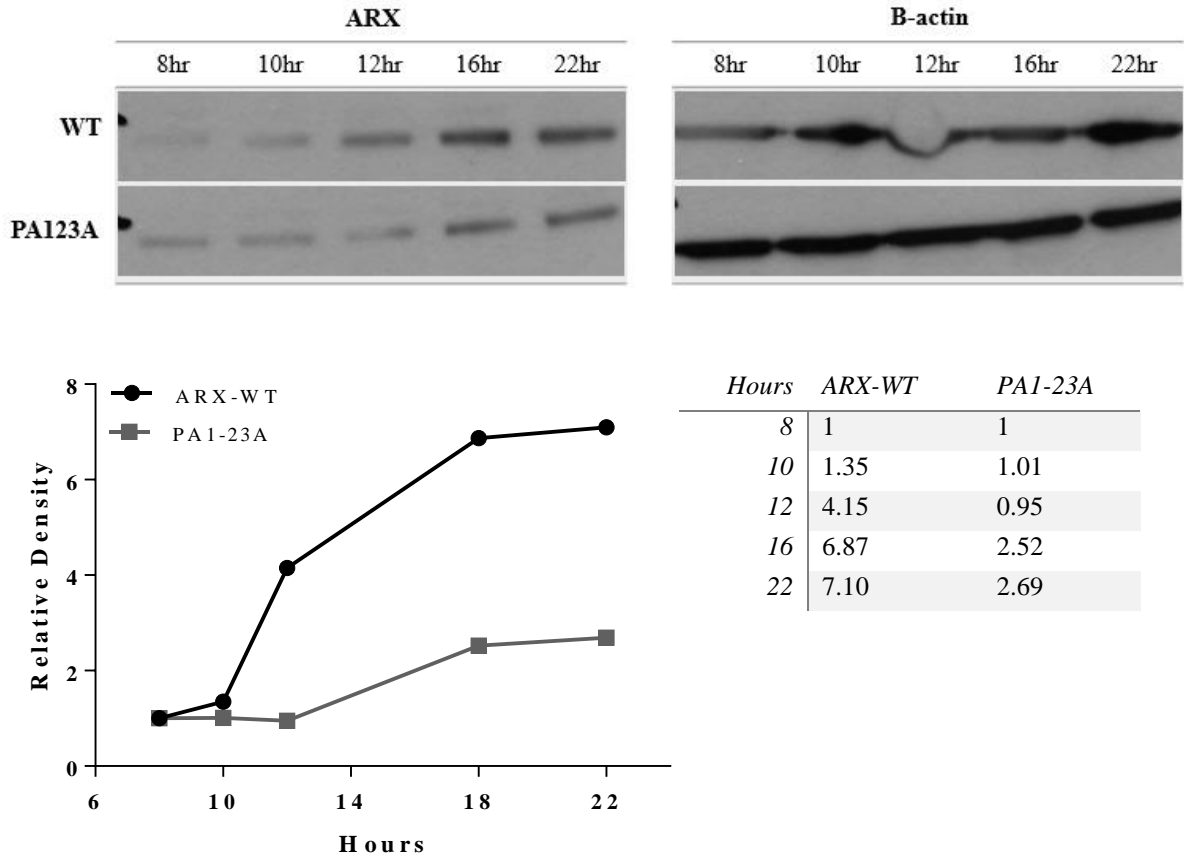
Figure 5.15 Decrease rate of protein production in polyalanine tract mutant sample.
Western immunoblots showing the rate of transfected ARX protein production across time
in ARX-WT and PA1-23A. Samples were loaded onto 4-12% SDS-PAGE gels, and ARX
was detected by immunoblotting with Myc-HRP (Invitrogen). B-actin was used as a
loading control. b) Densitometry of ARX bands was determined using LICOR western blot
analysis software, Image Studio Lite. The density of ARX protein was normalised to the
8-hour sample and plotted across time. ARX WT = black circle and PA1-23A = gray
square.

5.4   Discussion

We have previously demonstrated a marked reduction of mutant Arx protein abundance within the developing forebrain of both PA1 and PA2 (Lee, K. et al. 2014). Assessment of the transcriptional consequence indicates a selective effect of these mutations impacting on some but not all Arx downstream targets. Therefore expanded polyalanine tract mutations represent a partial loss of *Arx* function. However, the molecular mechanism driving disease is still unknown. In this study, *in silico* tools predict that expansions to polyalanine tracts do not cause any major structural impact on the ARX protein. Nor do these mutations disrupt the functional capacity or the novel interaction with UBQLN4 when tested *in vitro*. Reduction in protein due to the mutation is emerging as a major contributor to disease *in vivo.* We show this reduction in protein levels due to the mutation persists into adult life and is also present in another tissue, the testes. The mechanism resulting in the reduction in protein is still under investigation. Our studies to date have shown there is no change in the ARX protein stability between the WT and polyalanine mutant proteins. However, our data suggests there may be a decreased efficiency of protein production evident in an *in vitro* context. Taken together, our data supports a mechanism of inefficient translation of the *ARX* mRNA with expanded GCG sequences.

Polyalanine tracts are present in several transcription factors, and expansions in these sequences are responsible for several human genetics diseases. Structural studies have revealed that while short polyalanine peptides form alpha-helices (Giri et al. 2003), the secondary structure of longer polyalanine tracts is predicted to predominantly form beta-strand structures (Giri et al. 2003). This suggests that longer alanine-tracts found in

transcription factors like ARX possess an enhanced tendency to form beta-sheet rich fibrillary structures. This prediction is in agreement with *in silico* structure prediction of ARX-WT and two of the most common polyalanine expansion PA1-23A and PA2-20A. The I-TASSER predicted secondary structure showed several alpha helix structures spanning the ARX-WT protein corresponding to known functional domains with the exception of one located toward the C-terminal side (40-43aa) of the polyalanine tracts. In particular, this alpha helix appears prone to changes in secondary structure due to polyalanine tract mutation. We predict a shift in location (83-86aa) as a result of the addition of 8 alanines in polyalanine tract 2 (PA2-20A) and a more substantial disruption with a change in structure from an alpha helix to a beta strand (41-43aa) as a result of the addition of 7 alanines in polyalanine tract 1 (PA1-23A). This change in the model of the secondary structure is predicted to have an effect on the 3D structure, namely PA1-23A showing some change in protein folding. However, this is unlikely to affect the main nucleotide binding sites in either mutant protein. Even though the expansion itself in polyalanine tract 2 is longer at 8 alanines, the overall size of the tract at 20 alanines remains smaller than the expanded polyalanine tract 1 at 23 alanines in total. Therefore, the greater disturbances seen in the PA1-23A structure indicate that the resulting total size of the tract may play a greater role in disease than the size of the expansion itself. This may be the reason expansion mutations in polyalanine tract 3 and 4 of the ARX protein have not been reported as the smaller start size of 7 and 9 alanine, respectively may be able to tolerate expansions without causing disease.

By way of *in silco* predictions, the DNA binding of polyalanine mutant proteins is not impacted. It is well documented that mutations leading to the loss of ARX function cause catastrophic brain malformation disorders in addition to ID and seizures, including XLAG (Kitamura et al. 2002, Kato et al. 2004). Given the absence of these brain malformation phenotypes with polyalanine expansion mutations, it is likely these mutations cause a partial loss of function and retain some functional capacity (Lee, K. et al. 2014). When we interrogated the impact of expanded polyalanine tracts on the capacity of ARX transcriptional capacity using a cell-based assay, we show no loss of repressive activity with either mutation tested. This luciferase assay is useful in detecting changes in transcriptional activity when mutation impact on homeodomain binding (Shoubridge et al. 2012). Therefore, this suggests there is no major alteration in DNA binding and functionality of the mutant protein in this context. However, when we test activity in terms of the difference in protein levels seen *in vivo*, we see a concomitant loss of repression activity. This loss of repression due to lower levels of protein even with ARX-WT indicates that some targets of ARX repression are likely to be dosage sensitive. Similar results have been shown for SOX3, another X-linked polyalanine tract containing protein with reported polyalanine tract expansion mutations (Hughes et al. 2013). The effect of an expansion of a polyalanine tract by 11 alanines from 15 to 26 alanines in SOX3 resulted in reduced activity compared to SOX3-WT however a reduction of similar magnitude to the reduced luciferase output was observed suggesting that the mutant protein that is present in the nucleus has similar activity to WT (Hughes et al. 2013). Additionally, different effects of polyalanine mutant protein have been reported using different reporters and different cell types suggesting pathogenesis might be promoter and cell type specific (Di Lascio et al.

2013). This agrees with *in vivo* ARX polyalanine mutant data of selective deregulation of known targets (Lee, K. et al. 2014).

It has been previously proposed that polymeric runs might serve to fine-tune the activity of the transcription factor (Karlin and Burge 1995). Hence, one of the possible mechanisms by which polyalanine repeats could regulate the functions of proteins that contain them is by mediating the interaction with other proteins. The region spanning polyalanine tract 1 and 2 of the ARX protein was screened for novel protein binding partners, and UBQLN4 was identified as a novel interactor of this region (Figure 5.6). Interestingly, UBQLN4 is not a known co-factor of transcription factors. Despite this, ubiquilins, a family of ubiquitin-binding proteins, are involved in several protein degradation pathways and have been implicated in various neurodegenerative diseases. UBQLN4 belongs to the UBL-UBA family of proteins which play a role in the regulation of protein degradation via the ubiquitin-proteasome system (UPS). They mediate the proteasomal targeting of misfolded or accumulated proteins for degradation by binding (via the UBA domain) to their polyubiquitin chains and by interacting (via ubiquitin-like domain) with the subunit of the proteasome (Riley et al. 2004). Studies involving polyglutamine expansions have shown a role for Ubiquilin in the regulation of aggregate formations and suppressed polyglutamine-induced cell death.

At the time when the interaction between ARX and UBQLN4 was first identified, aggregation of the mutant protein was still considered to be a possible mechanism for disease. Therefore, we expected there might be a loss of interaction between ARX and

UBQLN4 resulting in the accumulation of mutant protein. However, in more recent years this was disproven as a likely artefact of *in vitro* systems and a reduction in mutant protein is a more likely factor contributing to disease (Innis et al. 2004, Kuss et al. 2009, Lee, K. et al. 2014). As the interaction between ARX and UBQLN4 is maintained with ARX polyalanine expansion mutant proteins, this interaction might contribute to the tight regulation of the ARX protein and may be implicated in the increased turnover of the mutant ARX protein. As the original portion of the ARX protein used to identify the interaction with UBQLN4 was only focused on the region containing polyalanine tract 1 and 2 only, this interaction could represent a common relationship across other polyalanine expansion containing proteins leading to tight regulation of transcription factor protein levels *in vivo*. Therefore, this may have broader implications in addressing the pathogenic mechanism behind expanded polyalanine tract disorders in general. Studies of several polyalanine expanded proteins have shown that nuclear inclusion or aggregates are often ubiquitinated, and the degree of ubiquitination of expanded polyalanine tracts mutant protein was higher than that of WT proteins (Parodi et al. 2012). This interaction with UBQLN4 may suggest alanine runs might serve a general function, such as regulation of the intracellular/intranuclear concentrations of the polyalanine tract containing transcription factors by establishing a chaperone-dependent equilibrium and resulting in an increased affinity for expanded polyalanine tracts.

Transcription factors affected by polyalanine expansion mutations are important for embryonic development and *in vivo* studies have demonstrated a reduction of the mutant protein with a focus on embryonic time points (Bruneau et al. 2001, Innis et al. 2004,

Hughes et al. 2013, Lee, K. et al. 2014). Even though reduction of mutant protein has been shown not to be stage dependent during embryonic life (Hughes et al. 2013), it has not been established whether the reduction of mutant protein persists or if it is tissue specific. ARX is most notably expressed in the brain during development (Kitamura et al. 2002, Ohira et al. 2002). Despite this, ARX-WT protein expression persists at a robust level in the adult testis and we demonstrate a significantly lower level of mutant ARX protein in this tissue in both PA1 and PA2 animals. Indicating the reduction in mutant protein continued across adult life and multiple sites.

Similar to *Hoxa13*, *Hoxd13* and *Sox3*, polyalanine expansion in *Arx* does not have a detectable impact on steady-state mRNA levels (Innis et al. 2004, Cocquempot et al. 2009, Hughes et al. 2013, Lee, K. et al. 2014). Therefore, reduced protein abundance could be secondary to a reduction in efficiency of translation or to degradation or combined effects. Moreover, this could be a common mechanism contributing to the disease of expanded polyalanine tracts more broadly. Given the interaction between ARX and UBQLN4, other polyalanine proteins co-localising with chaperones *in vitro* and that aggregation is promoted by pharmacological inhibition of the proteasome, it seems likely that mutant polyalanine proteins are cleared from the cell at an increased rate. Investigating the mechanisms that potentially contribute to this marked reduction in the expression of the mutant protein, our data indicates the stability of the mutant protein is not significantly compromised and exhibits the same rate of clearance as the WT protein. A primary analysis of translation efficiency showed slower protein production of polyalanine expansion mutant protein *in vitro* and suggested the structure of RNA carrying the expanded triplet

repeats impedes translation. Further investigations identifying the extent of ribosomal stalling is an active area of research and may prove to be the mechanism that drives the reduction in protein expression seen in other genes with disease-causing expansion in polyalanine tracts. However, these studies fall outside the capacity of my thesis to complete.

Hence, from my studies in conjunction with existing data, we contend that polyalanine expansion mutations in *ARX* result in a partial loss of function with the regulation of the steady-state levels of mutant protein representing a crucial component in the pathogenesis of these human disorders. Interestingly this reduction is detected in other tissues including the testes which have no reported gross morphological deficit in patients with polyalanine expansion mutations. It is of interest to note that there are no reports of paternal transmission of polyalanine tract disorders in ARX. We cannot rule out that this may be due, at least in part, to reduced functionality of the testes. This could be addressed using our knock-in mouse models. However, the evidence to date highlights that brain development is sensitive to the level of ARX protein present. Additional studies into understanding the molecular details of this sensitivity to gene dosage and identify the process causing the reduction in mutant protein will be essential in further understand the driver of polyalanine expansion disease.

Chapter Six:

Final Discussion: Contributing factors leading to polyalanine expansion disorders, particularly in *ARX*

6 Final Discussion: Contributing factors leading to polyalanine expansion disorders, particularly in *ARX*

Missense mutations in *ARX* result in phenotypes that involve ID with frequent comorbidities of epilepsy, infantile spasms, hand dystonia, autism or dysarthria. Historically, it was thought that only males were affected by mutations in *ARX* due to their single X-chromosome. We identified a novel mutation in an affected female. Our subsequent review of female phenotypes due to mutations in *ARX* in the literature highlights that *ARX* could be considered a syndromic form of X-chromosome-linked ID with associated co-morbidities with complete penetrance but variable expressivity in males, with incomplete penetrance and variable expressivity in females. The disparities in clinical phenotypes in females compared to males may be attributed in part to differences in genomic background and X-inactivation patterns in the brain (Marsh et al. 2016). The distinct phenotypes in males and females depending on the functional severity of the variant with *ARX*-related disorders is also noted in several other X-linked conditions, such as *PHF6*, *IQSEC2* and *USP9X* (Zweier et al. 2013, Reijnders et al. 2016, Zerem et al. 2016). Further investigation into the mechanism(s) underlying the female phenotypes in X-linked conditions is required, however, is likely due to a combination of both cellular mosaicism and reduced functional protein expression. Our study illustrates the importance of screening *ARX* in both male and female patients with ID and seizures.

The increasing awareness of female X-linked conditions in the past few years has been aided by unbiased approaches of next generation sequencing (NGS), allowing an

unprecedented increase in the ability to sequence large numbers of genes. The increased availability and affordability of NGS technology have been particularly important in the diagnosis of neurodevelopmental disorders given the broad spectrum of overlapping phenotypes resulting from a large number of potential genetic causes. Molecular diagnosis achieved by a genome-first approach using high-throughput NGS will not only resolve many undiagnosed ID cases, but this unbiased approach will provide accurate estimates of the prevalence of genetic mutations. A significant obstacle that remains to be overcome is the interpretation of pathogenicity of novel variants. Bioinformatics pipelines use stringent criteria to classify sequence variants as pathogenic, and functional data are often required to support the accuracy of these predictions. However, the false negative rate is not known making the continued development of functional studies for validation critical, especially for novel *de novo* cases. As we continue to unravel the full spectrum of genotype-phenotype relationships associated with mutations in *ARX*, this knowledge will enhance our ability to interpret and predict more accurately the phenotypic consequence of variants of unknown significance.

Diseases caused by *ARX* mutations can be considered under the umbrella of neurodevelopment disorders (NDDs), classified as any disorder caused by an alteration in how the brain develops. These include autism spectrum disorders, intellectual disability, seizure disorders, schizophrenia and attention deficit disorder. NDDs are a complex array of conditions arising due to high genetic heterogeneity (Chen et al. 2014). It is estimated approximately 20-45% of all NDDs are associated with variation in specific genes (Krumm et al. 2014). However, the small numbers of individual cases for the majority of causative

genes coupled with widespread clinical outcomes has limited the research into identifying the pathogenic mechanisms and restricted the development and implementation of therapeutic strategies for each genetic case and mutation type. *ARX* is a good example of the challenges faced by researchers and clinicians in the NDDs field. *ARX* has over 60 mutations reported to date, in over 138 cases with as many as 10 clinically distinct phenotypes (Shoubridge et al. 2016). Despite this daunting outlook, overlapping comorbid outcomes between NDDs arising from different genetic causes suggests the possibility that diseases giving rise to similar phenotypes, even due to vast genetic heterogeneity, may converge on similar molecular pathways (Chen et al. 2014). Hence, we have approached identifying molecular pathways involved in these disease outcomes using the expanded polyalanine tract mutations in *ARX* as a model. Assessment of the transcriptome-wide outcomes in mice modelling PA1 and PA2 mutations showed many of the genes deregulated were known ID, epilepsy or autism disease genes (Mattiske et al. 2016). The enrichment of genes with synaptic function supports the convergence of common pathways contributing to these disorders. Clinically children with severe infantile spasms (seen with expansion to both tracts, but more frequently in PA1) respond poorly to anticonvulsants. Short-term reduction in spasms is achieved with the treatment of ACTH (Hrachovy and Frost 2013). However, ACTH can have significant side effects (Shields 2006), and better therapeutic alternatives are required. Our investigations into the pathways disturbed due to *ARX* mutations aim to provide insights that may underpin the development of effective treatment options for seizure disorders due to *ARX* dysfunction and in turn perhaps other genetic causes of epilepsy. Arising from the work in which we interrogated the profile of deregulated genes from the developing brain of PA1 and PA2 mice, I have identified a

206

number of pathways enriched by genes deregulated in the mutant mice. One 'core' pathway of transcriptional regulators we suggest may represent early drivers of ARX-associated phenotypes, namely ID and epilepsy. We are building on this within our laboratory treating *ARX* polyalanine expansion mice with 17B-estradiol in early postnatal life. ACTH stimulates the synthesis of adrenocortical steroid hormones, including androgens. This has focused investigations on the role that oestrogen receptor expression and 17B-estradiol (E2) during the development of interneurons. Treatment of a different genetic strain of PA1 mutant mice with E2 during postnatal life resulted in sustained ablation of seizures in young adult mice (Olivetti et al. 2014). The treatment restored some depleted interneuron populations and altered mRNA levels of three downstream targets of Arx (*Ebf3*, *Shox2*, *Lgi1*) (Olivetti et al. 2014). However, the underlying molecular drivers of these responses have not been established. Of the 858 deregulated genes from our E12.5 dpc mutant mice, there were 299 genes harbouring an estrogen response element. It remains to be determined which genes are deregulated in early postnatal life when estrogen treatment has a beneficial impact on reducing seizures (Olivetti et al. 2014). The data that I have generated as part of my PhD studies can now be used as a baseline of early drivers contributing the ARX disease phenotype. We can build upon this knowledge to investigate the transcriptional changes that underpin long-term ablation of seizures after treatment E2 in both PA1 and PA2 mutant mice. Furthermore, it is unclear if this treatment has any beneficial impact on the cognitive impairments in this model. It is my hope that this data will enhance our knowledge of the pathways and clusters of deregulated genes that may be amenable to therapeutic manipulation to improve clinical outcomes not only disease caused by *ARX* mutations but by other ID and epilepsy disorders.

My research has focused on understanding the mechanism(s) by which mutations in *ARX* contribute to disease outcomes, with a particular focus on understanding the pleiotropy present in the patients and mouse models. When I examined Arx protein abundance in WT mouse testes, I found a variable expressivity occurs against the uniform genetic and environmental background. We predict that the protein level within individuals will lead to the variability of the disease outcome. Larger numbers of mice with corresponding phenotyping will be needed to provide statistical support for this prediction. What still needs to be determined is if the reduced level of the mutant protein is due to variable levels of *ARX* produced or if individuals have different efficiencies of degradation of the mutant protein. Although much remains to be understood about how cells degrade aberrant proteins, it has become clear that other components besides the mutant protein get sequestered and thus at least temporarily inactivated during the degradation process, resulting in compromised cellular function (Selkoe 2003). The extent to which other proteins are compromised by mutant ARX and the influence on general cellular functions is another potential avenue that may contribute to the variability in phenotype severity across affected individuals. The impact of epigenetic effects contributing to the variability in clinical severity observed in families with ARX expansion mutations has not been studied to date, and remains an area of interest for further investigation in our laboratory.

Understanding the general function of ARX during brain development has been of great interest in my examination of the transcriptome data I have generated. From the RNA-Seq analysis, we can conclude that partial loss of function of Arx is selective, and does not

seem to be due to gross defects in DNA binding. This is supported by the majority of known direct targets being unaffected in the mutant mice, with only a subset of targets deregulated in the developing brain. Why this specificity occurs is perplexing, and may indicate that the effect of alanine-tract length on DNA binding and presumably on transcriptional activity depends on, to some extent, of the DNA sequence being queried. Assessing whether the transcriptional impact of polyalanine expansions depends on the nature of the promoter may lead to additional insights. Indeed, we predict that different target promoters will have different sensitivities to decreased amount of available/'soluble' ARX. A computational analysis of promoters displaying high vs. low-sensitivity to ARX concentration might shed light on the molecular basis of these differences (i.e. the difference in the number of ARX binding sites or sites for potential partners?). Further transcriptome-wide expression and ChIP experiments in relevant tissue samples will allow for the identification of true or direct gene targets in the relevant cellular context. The genes that are expected to display altered expression patterns will be (i) direct and indirect ARX targets and (ii) possibly genes whose products are involved in the adequate removal of the mutant protein (i.e. proteasome subunits, ubiquitination pathways members, chaperones, etc.). These latter genes, even though not directly regulated by ARX, are likely to be modifiers of an individual's phenotype.

The frequent occurrence of reduced levels of mutant protein in transcription factors with disease causing polyalanine expansion mutations (Bruneau et al. 2001, Innis et al. 2004, Goridis et al. 2010, Hughes et al. 2013, Lee, K. et al. 2014) indicates there may be a shared mechanism leading to the development of disease. Throughout my studies, I have

considered two fundamental cellular processes, which when perturbed could lead to a reduction in protein; namely decreased translation and / or increased degradation of the mutant protein. Determining the cause of this reduced protein abundance will help to focus further research on rescuing the deficit produced by polyalanine expansion mutations. It has long been speculated that expansion mutations result in secondary hairpin structures, stalling ribosome progression, leading to inefficient translation. mRNA hairpin structures are prevalent in human transcripts and are suggested to play a role in gene expression regulation. The increased GC content due to expansion mutations is predicted to result in decreased translation (Jasinska et al. 2003). Recently, it was established that repetitive RNA motifs can support translation initiation in the absence of an AUG start codon across a wide variety of sequence contexts (Zu et al. 2011, Zu et al. 2013), and the products of these atypical translation initiation events contribute to neuronal toxicity (Sellier et al. 2017). Studies have demonstrated that polyglutamine repeats in *FMR1* elicit production of a toxic polyglycine protein FMRpolyG, which is now known as repeat-associated non-AUG (RAN)-initiated translation (Todd et al. 2013). After cap-dependent initiation of translation and subsequent scanning through the 5' UTR ribosome stalling at secondary structures formed at repeats lead to aberrant translation initiation at non-AUG codons, resulting in toxic repeat peptides. RNA secondary and tertiary structures contribute significantly to the dynamics and regulation of translation. The efficiency of RAN translation is dependent on repeat length and the imperfect RNA hairpin structures formed by the expanded repeats. Studies suggest that the ribosome engages with the mRNA as established but the presence of the structured repeats promotes reduced specificity of translation initiation which is a consequence of an activated 40S subunit stalled by a hairpin

structure (Kearse et al. 2016). RAN translation represents a new and provocative mechanism by which protein translation can occur in the setting of nucleotide repeat expansion to produce a novel set of toxic proteins which due to out of frame products may not have been detected due to antibody specificity. Establishing if the triplet repeat mutation leading to expansion polyalanine tract in Arx alters the translation of mRNA and leads to a reduced protein expression is an area of active research in our laboratory. A ribosome profiling approach will be undertaken. In this approach, a translating ribosome protects a footprint approximately 30 nucleotides of the mRNA temples. Deep-sequencing ribosome-protected mRNA fragments will yield subcodon resolution of translation (Ingolia et al. 2009). If a ribosome stalls at a particular location, an excess of a particular footprint will occur.

A posttranscriptional reduction in protein levels is suggested, partly due to the ability of cells to recognise and degrade misfolded mutant protein, thus circumventing the formation of overt aggregates. Even though it is now accepted that polyalanine expansion protein does not form amyloids under normal cellular conditions, instead we and our collaborators have shown these expansions form α-helical clusters, thus providing another level for how polyalanine mutations may disrupt normal function (Polling et al. 2015). Under physiological expression condition using tissue slices from PA1 and PA2 mice, focusing on the ventrolateral mantle zone of the telencephalon of embryonic mice at 12.5 dpc, where Arx expression in most pronounced, Arx mutants displayed elevated granularity of nuclear-localised protein compared to Arx WT (Polling et al. 2015). This result suggests that the abnormal clustering patterns caused by polyalanine expansion underlie a general

mechanism for the displacement of normal protein functionality and may lead to the disruption of normal ligand interactions. Such forms may be expected to be targeted for clearance by the quality control machinery, and indeed polyalanine expanded proteins have been shown to colocalise with cellular makers of degradation (Nasrallah et al. 2004, Trochet et al. 2005, Utsch et al. 2007, Parodi et al. 2012). In a polyglutamine model, these soluble oligomers are more pathogenic than insoluble aggregates but engage in more aberrant interactions with the potential to disrupt basic cellular function involving low-complex-repeat containing proteins, such as RNA processing and ribosome biogenesis (Kim et al. 2016). Neuronal dysfunctional may result from a culmination of multiple defects that are also caused by the presumably dynamic interaction of soluble aggregates.

Many of the drug treatments being investigated for polyalanine expansion disorders are effective at reducing aggregate formation *in vitro*. The prevailing model for drug action is that increased chaperone activity leads to a reduction in aggregate formation. The alternative hypothesis for drug action is that increased chaperone activity increases the efficiency of primary protein folding to increase the amount of functional protein. For the future it would be interesting to assess the effect of these drugs in a polyalanine expansion mouse model where protein aggregates are not seen *in vivo*, thus testing the toxicity of aggregates/oligomers in the correct cellular and physiological context. Finally, another avenue of investigation that has not yet been explored is the use of patient-derived induced pluripotent stem (iPS) cells. This approach is likely to yield further insights into the cellular pathology of polyalanine expansion disorders and provide a useful vehicle for pharmacological screening including the role of the genetic background contributing to

severity. However, as Arx is particularly important in interneuron development, this neuron population remains difficult to achieve by differentiation of iPS cells. Possible treatments could also be considered to target mRNA secondary structures such as TMPyP4, shown to reduce translation delay due to disease causing CGG repeats in FMR1 (Weisman-Shomer et al. 2003, Ofer et al. 2009).

The molecular mechanism(s) underpinning polyalanine expansion mutations are complex. We predict it is likely that many factors contribute to disease manifestation and variability of phenotypes. From the investigations during my PhD studies, my current working hypothesis is that multiple defects occur to the resulting mutant protein leading to a partial loss of function. Polyalanine expansion mutation proteins are functional but are likely to be sub-optimal, due to an altered capacity to bind to interacting proteins, oligomers engaging in aberrant interactions, and from the reduction in the mutant protein itself leading to dose-dependent changes to transcriptional capacity. The process leading to the reduction in mutant protein is an exciting area of research as insights may lead to viable treatment options not only for the disorders caused by polyalanine expansion in ARX but the additional seven other transcription factors affected by polyalanine expansion mutations. In addition, modelling the molecular disruptions caused by expansion in polyalanine tract 1 and tract 2 of ARX has enabled us to begin to assess the degree of molecular convergence between these two mutations and establish the drivers of disease outcomes contributing to the shared phenotype that may also apply to other genetic causes of neurodevelopmental disorders.

Appendix

7 Appendix

7.1 Result Tables from Chapter 4

Appendix Table 1: PA1 deregulated gene list

Gene ID	logFC	logCPM	LR	PValue	FDR
0610010B08Rik	-1.27201	0.376094	6.617006	0.010101	0.042781
0610012G03Rik	-1.18657	4.7496	10.86284	0.000981	0.011248
0610040B10Rik	-1.38869	1.006032	12.29983	0.000453	0.007748
1110001J03Rik	-1.11495	4.630258	14.51541	0.000139	0.004625
1190005I06Rik	-1.71375	-0.33152	12.19626	0.000479	0.007835
1500016L03Rik	-1.31547	1.72895	8.127712	0.004359	0.025371
1500032L24Rik	-1.12949	6.673542	15.71578	7.36E-05	0.003888
1600002K03Rik	-1.76795	2.050941	14.5542	0.000136	0.00459
1700007J10Rik	1.339379	0.655445	7.738763	0.005405	0.028974
1700008O03Rik	-2.38033	0.510312	13.60667	0.000225	0.005773
1700011J10Rik	-1.25951	2.070304	13.00246	0.000311	0.006716
1700016K19Rik	-1.40064	-0.40003	6.368842	0.011614	0.047028
1700020I14Rik	1.185394	4.864457	19.2199	1.16E-05	0.002311
1700029I15Rik	-1.5964	-0.82634	8.252613	0.004069	0.024272
1700048O20Rik	-1.19607	3.471648	10.83008	0.000999	0.011322
1700067K01Rik	-1.14392	0.039629	7.041731	0.007963	0.036771
1700113A16Rik	-1.29603	3.752218	12.43809	0.000421	0.00757
2010320M18Rik	-1.15201	2.83816	11.50988	0.000692	0.009316
2310002D06Rik	1.675014	-0.23603	9.319089	0.002268	0.017522
2410066E13Rik	1.522695	4.163534	18.60847	1.61E-05	0.00238
2410137F16Rik	2.784826	1.125381	21.95337	2.79E-06	0.00099
2610005L07Rik	1.296677	5.943825	12.07005	0.000512	0.008049
2610017I09Rik	-1.18238	6.544142	12.15462	0.00049	0.00786
2610019E17Rik	-1.12904	3.221525	11.40037	0.000734	0.009683
2810405F15Rik	-1.48493	0.820805	11.00392	0.000909	0.010795
2810428I15Rik	-1.31538	4.008788	14.40609	0.000147	0.004746
3110021A11Rik	1.377668	-0.36248	9.948215	0.00161	0.014515
3110021N24Rik	1.722986	1.306118	10.4136	0.001251	0.012665
3110047P20Rik	1.249986	1.236017	10.74658	0.001045	0.011545
3300002I08Rik	1.388177	0.294927	12.61777	0.000382	0.007294
4632427E13Rik	1.24632	0.542107	13.50952	0.000237	0.005953
4921525B02Rik	1.761232	0.023865	10.3895	0.001267	0.012701
4930404N11Rik	-1.14609	1.630962	8.345079	0.003867	0.023722

4930444P10Rik	-1.94316	0.312289	16.523	4.81E-05	0.003301
4930455C13Rik	2.07035	-0.73006	14.01351	0.000182	0.005226
4930506C21Rik	1.297217	0.138302	7.481899	0.006232	0.031546
4930565N06Rik	2.300628	-0.47566	8.532889	0.003488	0.022357
4930578M01Rik	1.155101	0.358787	11.10447	0.000861	0.01044
4931417G12Rik	-1.25245	0.337899	9.4277	0.002137	0.017034
4932411E22Rik	1.15665	0.644623	4.106265	0.042725	0.112467
4932418E24Rik	1.707146	0.720514	11.04525	0.000889	0.010639
4932438A13Rik	1.385145	5.781775	14.75029	0.000123	0.004535
4933436C20Rik	-1.37594	0.81167	11.10946	0.000859	0.01043
5031434O11Rik	-1.80118	-0.95457	7.755146	0.005356	0.028824
5330426P16Rik	1.242911	2.053445	12.97902	0.000315	0.00676
5730408K05Rik	-1.31294	1.139762	15.56091	7.99E-05	0.003957
5830454E08Rik	-1.50039	-0.35417	10.09787	0.001484	0.013857
6330403M23Rik	-1.11854	1.649876	9.94876	0.00161	0.014515
6330415B21Rik	-3.86228	-0.91273	11.36401	0.000749	0.0098
6330418K02Rik	-1.26533	-0.05641	9.799575	0.001746	0.015153
6330549D23Rik	-1.2752	1.676895	4.405447	0.035824	0.099741
8030462N17Rik	1.156862	3.99937	7.714813	0.005477	0.029225
9530082P21Rik	-1.2943	4.809901	11.11751	0.000855	0.010417
9930013L23Rik	1.381232	2.013537	6.968165	0.008297	0.03779
A330023F24Rik	-1.42171	1.604096	12.74817	0.000356	0.007147
A630001G21Rik	-1.225	-0.24736	11.81838	0.000586	0.008528
A630007B06Rik	1.704247	4.012707	11.15399	0.000839	0.010304
A630089N07Rik	1.175358	2.741932	7.179838	0.007373	0.035232
A730008H23Rik	1.672879	0.698125	8.771926	0.003059	0.020711
Abca8b	1.132348	1.718968	8.507562	0.003537	0.022493
Abhd14a	-1.11569	4.410861	11.49221	0.000699	0.009368
Abhd2	1.62879	4.480231	11.56001	0.000674	0.009191
Acvr2b	1.156761	4.829312	14.70336	0.000126	0.004552
Adamts12	1.402052	2.094112	4.833304	0.027915	0.084251
Adamts6	1.347386	3.243249	7.464686	0.006292	0.031749
Adamtsl3	1.458963	-0.64298	11.24414	0.000799	0.010072
Adarb2	1.413084	2.594551	23.19265	1.47E-06	0.000693
Adcy9	1.541791	2.108944	15.16114	9.87E-05	0.004289
Aff2	1.991401	4.248189	12.89637	0.000329	0.006894
Aff4	1.13488	7.632438	15.99581	6.35E-05	0.003671
Ager	-1.14503	2.669702	7.416238	0.006464	0.032284
AI413582	-1.19663	2.891611	10.45587	0.001223	0.012484
AI462493	-1.1917	3.261246	12.70084	0.000365	0.007212
AI506816	1.213334	1.817627	10.37741	0.001276	0.012719

AI606473	-1.48263	2.387366	16.70756	4.36E-05	0.003172
Akap6	1.111425	6.334586	24.66974	6.8E-07	0.000522
Alg10b	1.448524	5.881762	7.920775	0.004887	0.027157
Alkbh2	-1.33662	2.343902	15.29217	9.21E-05	0.004146
Alkbh8	1.297815	4.206992	10.29657	0.001333	0.013037
Alx1	-3.19149	0.446978	12.26699	0.000461	0.007788
Amh	-1.22756	-0.11543	9.116098	0.002534	0.018596
Anapc13	-1.31005	6.058474	18.79981	1.45E-05	0.00238
Ankfn1	1.845155	0.634943	6.728277	0.00949	0.041147
Ankrd26	1.133347	4.84469	7.370827	0.006629	0.032801
Ap1s3	-1.27234	2.419265	5.403281	0.020099	0.066948
Apom	-2.17905	0.162485	11.90452	0.00056	0.008259
Aqp4	1.148759	-0.41002	6.735173	0.009453	0.041017
Arhgap32	1.165808	5.658843	15.7437	7.25E-05	0.003886
Arhgap5	1.241946	6.039809	7.578624	0.005906	0.030597
Arhgdig	-1.21335	-0.17272	7.956896	0.00479	0.026814
Arid5b	1.576979	3.064004	10.54304	0.001166	0.012178
Arl5b	1.354838	4.970345	9.291975	0.002302	0.017705
Ascc1	-1.15884	4.016392	12.75923	0.000354	0.007147
Ash1l	1.107099	6.602328	14.5791	0.000134	0.004552
Asphd1	-2.05662	0.98562	11.94868	0.000547	0.008204
Aspm	1.191281	6.90871	14.63206	0.000131	0.004552
Asprv1	-1.22835	-0.52238	6.401175	0.011404	0.046378
Atf7	1.606527	2.891744	13.46606	0.000243	0.006016
Atf7ip	1.251472	6.215053	18.84171	1.42E-05	0.00238
Atp2b3	1.436523	1.238266	12.34304	0.000443	0.007704
Atp5e	-1.51157	5.737868	16.50519	4.85E-05	0.003301
Atp5g1	-1.15614	6.780373	14.89333	0.000114	0.00441
Atp5g2	-1.1906	7.687985	19.20017	1.18E-05	0.002311
Atp5h	-1.18433	6.501443	15.713	7.37E-05	0.003888
Atp5k	-1.12528	5.996161	12.39453	0.000431	0.007639
Atp7b	1.127652	2.259245	7.289666	0.006935	0.033802
Atp8a2	1.427292	2.899832	25.02705	5.65E-07	0.000499
B230119M05Rik	1.1055	0.569271	9.940591	0.001617	0.014534
Barx1	-3.71234	-1.68519	7.263453	0.007037	0.034136
Bax	-1.1018	5.490625	12.17445	0.000484	0.007835
BC002163	-1.18351	1.653945	5.981706	0.014455	0.053802
BC005561	1.836793	3.930392	13.54843	0.000232	0.005895
BC017612	-1.25054	2.761067	16.42664	5.06E-05	0.003339
Beta-s	-1.18488	5.485102	14.90694	0.000113	0.00441
Bex2	-1.13413	6.11868	15.39141	8.74E-05	0.004101

Bfsp2	-1.27277	-0.21646	7.32199	0.006812	0.03337
Bicd1	1.380977	3.516111	15.21719	9.58E-05	0.004191
Birc6	1.483883	7.161296	23.7506	1.1E-06	0.00062
Bmpr2	1.78605	4.514409	13.47674	0.000242	0.005994
Bola2	-1.55065	5.230313	18.57273	1.64E-05	0.00238
Brwd3	1.716507	4.337986	22.85745	1.74E-06	0.000707
Btbd7	1.589851	5.44782	12.42903	0.000423	0.007587
Btbd8	1.900242	-0.22009	17.35943	3.09E-05	0.00284
C130071C03Rik	1.358523	2.38598	19.85818	8.34E-06	0.001984
C1galt1	1.155016	2.439774	6.516916	0.010685	0.044359
C5ar2	-1.45341	-0.55005	4.383712	0.036284	0.100645
C77370	1.184783	3.822507	23.70593	1.12E-06	0.00062
C78339	1.293569	4.230786	7.492418	0.006196	0.031439
Cabp1	-1.19931	1.855533	7.412505	0.006477	0.032328
Cacfd1	-1.13806	5.707027	11.03697	0.000893	0.010668
Cacna1b	1.111241	4.402469	19.42999	1.04E-05	0.002289
Cacna1e	1.454316	3.722204	9.613919	0.001931	0.016029
Carf	1.198872	1.690007	7.235797	0.007146	0.03446
Cartpt	-1.75389	1.311517	5.666944	0.017288	0.060408
Casc5	1.309591	5.328364	8.20905	0.004168	0.024586
Cbl	2.159575	5.222165	12.68322	0.000369	0.007212
Ccdc107	-1.13487	3.125673	10.64234	0.001105	0.011872
Ccdc12	-1.48329	5.162624	16.18382	5.75E-05	0.00347
Ccdc124	-1.32578	5.077542	13.65023	0.00022	0.005743
Ccdc171	1.145685	2.510222	10.4567	0.001222	0.012484
Ccdc23	-1.34044	5.309399	17.65208	2.65E-05	0.002831
Ccdc24	-1.16876	4.150018	10.13796	0.001452	0.013692
Ccdc38	-1.24509	0.510328	10.69736	0.001073	0.01166
Ccdc88c	1.105373	6.872172	11.60048	0.000659	0.00904
Ccnt1	1.66108	4.03689	10.53944	0.001169	0.012178
Cdh13	1.19145	2.921916	15.05466	0.000104	0.004361
Cdh3	-1.30886	0.357543	5.578471	0.018183	0.062287
Cdh6	1.838818	3.535725	11.31515	0.000769	0.009906
Cdk2ap2	-1.18606	4.514507	12.38186	0.000434	0.007652
Cdk5rap2	1.281559	5.620158	24.86792	6.14E-07	0.000499
Cdk6	1.938026	1.628986	25.52619	4.36E-07	0.000496
Cdkl5	2.074945	2.76514	23.13058	1.51E-06	0.000693
Cdon	1.164197	7.367636	11.72897	0.000615	0.008758
Cep290	1.118912	3.694128	13.28278	0.000268	0.006285
Cep350	1.252393	6.069266	18.02059	2.19E-05	0.002643
Cep85l	1.777539	1.765974	7.455213	0.006325	0.031845

Cers6	1.521702	4.327928	11.08373	0.000871	0.010521
Chchd1	-1.26702	5.039863	19.0912	1.25E-05	0.002311
Chd9	1.100631	6.24118	16.99857	3.74E-05	0.003133
Chkb	-1.69613	2.229955	11.92687	0.000553	0.008233
Chmp2a	-1.22972	5.368229	16.65939	4.47E-05	0.003186
Chrm3	1.612218	-0.24518	8.395219	0.003762	0.023388
Chrna7	1.591034	1.546752	10.46006	0.00122	0.012484
Cib2	-1.16433	2.480952	12.00956	0.000529	0.008127
Cisd3	-1.22773	2.026566	8.729338	0.003131	0.021028
Cit	1.112135	4.514056	10.07484	0.001503	0.013932
Clcn5	1.798172	4.278019	11.25253	0.000795	0.010052
Cldn3	-1.3225	0.209263	9.986951	0.001577	0.014325
Cldn6	-1.60797	0.06983	6.275664	0.012241	0.048486
Clock	1.2197	4.939799	14.10575	0.000173	0.005127
Cnksr2	1.126403	3.268661	12.43041	0.000422	0.007587
Cnot6	1.13126	6.302926	6.897375	0.008632	0.038796
Cntn5	1.235768	1.297549	12.66709	0.000372	0.007217
Cntnap3	1.222721	0.461133	5.822004	0.015827	0.057079
Cntnap5a	2.661611	-0.86058	22.80869	1.79E-06	0.000707
Cntnap5b	2.90496	0.237802	20.0605	7.5E-06	0.001863
Coa3	-1.11717	4.764819	12.87013	0.000334	0.006926
Cox17	-1.31313	4.501685	17.31555	3.17E-05	0.002841
Cox4i1	-1.11823	7.376168	16.76486	4.23E-05	0.003172
Cox6b2	-1.18812	2.658062	11.29844	0.000776	0.009957
Cox7a1	-1.29641	-0.69029	5.734541	0.016634	0.058976
Cpa2	-1.15581	0.630104	12.30583	0.000452	0.007742
Cpeb4	1.898035	3.513677	10.39174	0.001266	0.012701
Crabp1	-1.55693	2.168367	16.21224	5.66E-05	0.00347
Crabp2	-1.18677	5.8457	14.48493	0.000141	0.004649
Creb5	1.549474	2.474859	10.11387	0.001472	0.013788
Crebbp	1.124841	5.3643	13.79009	0.000204	0.005601
Crip1	-1.10637	0.975628	7.679131	0.005586	0.02958
Crip3	-1.44234	-0.3104	8.353353	0.00385	0.023702
Cript	-1.10848	5.219816	14.67162	0.000128	0.004552
Csmd1	1.752413	1.917879	18.59261	1.62E-05	0.00238
Csmd3	1.74923	1.819934	17.33898	3.13E-05	0.00284
Csrnp3	1.175899	5.953111	30.4826	3.37E-08	0.000266
Cuedc2	-1.15617	6.666162	10.18597	0.001415	0.013494
Cxcl14	-1.38701	1.865674	13.47635	0.000242	0.005994
Cxcr5	-2.38523	-0.40142	10.61695	0.001121	0.011931
Cxx1a	-1.10696	4.296978	10.58776	0.001138	0.012028

Cxx1b	-1.44784	4.776053	12.73898	0.000358	0.007147
Cyp4x1	1.438721	-0.36712	6.342831	0.011786	0.047348
D030047H15Rik	1.144295	2.08087	8.986846	0.002719	0.019308
D10Bwg1379e	1.648543	2.468903	21.0052	4.58E-06	0.001456
D130040H23Rik	1.679909	1.626747	24.39814	7.83E-07	0.00057
D330041H03Rik	1.2467	0.925094	10.05105	0.001523	0.014028
D430020J02Rik	1.144343	3.461164	8.028441	0.004605	0.026265
D630041G03Rik	1.301865	-0.90309	7.781315	0.005279	0.028542
D830031N03Rik	2.162125	4.479222	17.39447	3.04E-05	0.00284
D8Ertd738e	-1.39783	5.375905	15.40168	8.69E-05	0.004099
D930028M14Rik	-1.16723	3.000448	7.897393	0.004951	0.027404
Dand5	-1.11053	2.510285	7.519289	0.006104	0.031129
Darc	-1.29929	3.813836	14.16624	0.000167	0.005061
Dbpht2	-1.38791	0.079628	8.462988	0.003624	0.022841
Dcdc2b	-1.46372	4.290363	8.619996	0.003325	0.021777
Ddi2	1.368096	4.021388	15.34271	8.97E-05	0.004106
Ddt	-1.17155	4.890628	13.67627	0.000217	0.005695
Dennd1c	-1.29311	3.535207	12.33139	0.000445	0.007704
Dgkh	1.951137	1.244332	21.82597	2.99E-06	0.001031
Dgki	1.459666	2.345633	17.39905	3.03E-05	0.00284
Dhrs3	-1.16167	-0.07412	9.110471	0.002542	0.018637
Disc1	1.194491	0.18588	5.639336	0.017562	0.06095
Dkkl1	-1.87888	-0.61914	10.31232	0.001321	0.01298
Dleu2	-1.26751	2.600216	13.88671	0.000194	0.005441
Dlgap2	1.320066	1.602044	13.74388	0.00021	0.005661
Dlk2	-1.36854	1.939745	8.95075	0.002774	0.019536
Dmxl2	1.306287	5.854497	23.62977	1.17E-06	0.000621
Dnahc5	1.178599	0.115558	8.35959	0.003837	0.02367
Dnajb14	1.692645	1.905508	17.5718	2.77E-05	0.002832
Dnm3os	-1.82239	0.309098	6.433796	0.011197	0.045863
Doc2g	-1.28157	3.296994	8.51812	0.003516	0.022426
Dock4	1.352059	4.834045	18.56236	1.64E-05	0.00238
Dok3	-1.41428	2.806783	13.35186	0.000258	0.006218
Dok6	2.016688	-0.1702	12.68318	0.000369	0.007212
Dopey1	1.166896	5.116068	17.95801	2.26E-05	0.002668
Dpm3	-1.28697	4.680548	16.76423	4.23E-05	0.003172
Dpy19l4	1.819367	3.728988	12.2791	0.000458	0.007768
Dpysl2	-1.88363	7.340254	18.73005	1.51E-05	0.00238
Dpysl3	1.647775	7.90096	17.4544	2.94E-05	0.00284
Drap1	-1.26395	6.545533	13.53404	0.000234	0.005919
Dtnbp1	-1.13287	4.872933	13.33585	0.00026	0.006226

Dusp4	1.136163	6.405928	8.414931	0.003722	0.02323
Dync2h1	1.211622	5.651332	22.41176	2.2E-06	0.000822
Dynlrb2	-1.35015	0.496206	8.096235	0.004436	0.025692
Dynlt1f	-1.44417	3.446717	16.74762	4.27E-05	0.003172
Dyrk2	1.31672	3.959067	10.99479	0.000914	0.010829
E330009J07Rik	1.921575	3.17155	15.86023	6.82E-05	0.003785
Ebf1	1.221055	3.546801	25.10638	5.43E-07	0.000499
Egfl8	-1.12351	1.139903	9.760403	0.001783	0.015335
Egr3	1.279613	0.059908	6.916402	0.008541	0.038599
Eif1	-1.21454	8.020039	12.55505	0.000395	0.007408
Eif2c2	1.841938	6.112434	15.04328	0.000105	0.004361
Eif2c3	1.805612	3.342115	24.21347	8.62E-07	0.000588
Eif3f	-1.17132	6.577129	11.11682	0.000855	0.010417
Eif4ebp3	-1.2868	3.321373	12.26502	0.000462	0.007788
Elfn2	1.263931	1.944207	10.59271	0.001135	0.012023
Elk4	1.613643	2.421925	14.43186	0.000145	0.004708
Ell3	-1.27029	0.621425	9.978791	0.001584	0.014369
Eml6	1.205643	1.981921	10.32219	0.001314	0.012938
Endog	-1.23812	-0.02571	9.094785	0.002563	0.018734
Eno3	-1.18568	3.169377	8.804551	0.003005	0.02052
Epg5	1.102587	4.678974	16.7109	4.35E-05	0.003172
Epha3	1.182253	6.493735	6.011059	0.014216	0.053386
Eppk1	-1.10637	0.440119	4.024695	0.044839	0.116501
Erbb4	2.705876	2.252102	26.40038	2.77E-07	0.000496
Ern1	1.661821	3.475774	9.913466	0.001641	0.014633
Esrp1	-4.3481	-0.94951	7.773064	0.005303	0.02864
Exoc4	1.180107	6.3829	13.50081	0.000238	0.005968
F730043M19Rik	1.51018	1.562642	12.35083	0.000441	0.007701
Fabp3	-1.16962	1.966222	8.72442	0.00314	0.021074
Fam135b	2.489531	-0.61144	19.92473	8.06E-06	0.001953
Fam159a	-1.62957	0.024012	11.76483	0.000604	0.00868
Fam171b	1.4813	5.867235	7.879969	0.004999	0.027577
Fam183b	-1.20212	-0.21987	4.99846	0.02537	0.078828
Fam217b	1.58671	0.661525	14.66283	0.000129	0.004552
Fam57b	-1.2082	5.753549	11.85277	0.000576	0.008438
Fam92b	-1.21151	-0.47659	5.143639	0.023332	0.074347
Fat1	1.150387	7.75576	10.67784	0.001084	0.011734
Fat3	1.358352	6.461513	17.40787	3.02E-05	0.00284
Fau	-1.17683	7.603006	12.17408	0.000485	0.007835
Fbxl18	1.2412	3.735558	16.19535	5.71E-05	0.00347
Fbxo48	1.589908	2.350058	8.23336	0.004113	0.0244

Fgd4	1.121176	4.09403	12.78222	0.00035	0.007112
Fis1	-1.13259	5.323704	10.97909	0.000921	0.010884
Fkbp11	-1.20783	0.170898	9.315606	0.002272	0.017522
Fkbp2	-1.52625	3.93137	13.00386	0.000311	0.006716
Fktn	1.652059	4.198069	11.64549	0.000644	0.008962
Flrt1	1.423708	3.240581	15.23127	9.51E-05	0.004191
Flywch2	-1.2646	2.845391	9.027757	0.002659	0.019113
Fos	1.663942	-0.18269	11.63908	0.000646	0.008962
Foxo3	1.576477	3.214509	10.6712	0.001088	0.011758
Frmd7	-1.87996	0.164286	19.09159	1.25E-05	0.002311
Frrs1l	1.358951	3.410007	17.39304	3.04E-05	0.00284
Fry	1.129039	4.67241	25.39851	4.66E-07	0.000496
Ftl1	-1.16516	9.687198	11.826	0.000584	0.008516
Fxyd7	-1.48877	1.676048	8.254028	0.004066	0.024265
G0s2	-1.29387	1.174337	10.27512	0.001348	0.013104
Gabarap	-1.10753	7.393981	10.28635	0.00134	0.01308
Gabbr2	1.122174	1.606292	10.19619	0.001407	0.013469
Gabrq	1.695123	0.729874	7.510021	0.006136	0.031221
Gadd45g	-1.15004	6.019118	17.6955	2.59E-05	0.002831
Gan	2.081829	2.495791	18.32957	1.86E-05	0.00238
Gar1	-1.15496	5.161983	13.76108	0.000208	0.005659
Garem	1.392916	3.744325	12.28228	0.000457	0.007768
Gatad2b	1.718537	4.535137	12.83202	0.000341	0.006995
Gatsl2	1.186366	5.17935	18.57336	1.63E-05	0.00238
Gdap10	1.869155	0.625074	6.026622	0.014092	0.053122
Gfod1	1.216736	3.13254	16.07025	6.1E-05	0.003574
Ggnbp1	-1.22645	1.939697	9.831265	0.001716	0.014997
Ghrh	-1.31996	-0.19504	5.915035	0.015012	0.055075
Gk5	1.387933	0.727131	7.883772	0.004988	0.02753
Glg1	1.811977	6.615133	13.63102	0.000222	0.005773
Glis3	1.316205	1.510247	4.25271	0.039188	0.10617
Glrx3	-1.27071	4.823459	12.0409	0.00052	0.008063
Gm10406	-1.16878	3.75967	14.38754	0.000149	0.004782
Gm11974	-1.14292	2.390353	14.86915	0.000115	0.004436
Gm12060	-2.37978	-0.7179	7.098266	0.007716	0.036184
Gm12070	-1.13468	3.858969	8.451994	0.003646	0.022927
Gm12709	-1.20181	-0.58965	7.133074	0.007567	0.035743
Gm13826	-1.58267	-0.48614	7.674701	0.0056	0.029607
Gm14827	-1.24191	2.822259	10.59849	0.001132	0.012023
Gm15421	-1.66808	1.439452	6.607449	0.010155	0.04288
Gm16119	1.294645	0.381587	7.415498	0.006466	0.032285

Gm16386	1.763171	1.820124	14.92643	0.000112	0.00441
Gm16576	1.267653	1.492862	9.761958	0.001782	0.015331
Gm16617	1.236076	-0.42242	4.820024	0.028131	0.084771
Gm16982	1.364941	-0.37423	7.139162	0.007542	0.035646
Gm19557	1.429528	-0.59605	9.348784	0.002231	0.017412
Gm19757	1.207851	0.235278	10.85736	0.000984	0.011248
Gm2694	-1.23277	3.066031	8.071482	0.004497	0.025904
Gm3414	1.285991	3.104024	15.74208	7.26E-05	0.003886
Gm3500	-2.39987	0.662582	18.31771	1.87E-05	0.00238
Gm5176	-1.51736	-0.20076	13.52917	0.000235	0.005924
Gm5415	-3.72281	1.526026	19.8059	8.57E-06	0.001984
Gm5506	-1.46309	3.04718	10.7559	0.001039	0.01152
Gm5617	-1.38393	1.80405	10.9628	0.00093	0.010934
Gm5796	-1.1111	2.172572	6.5872	0.010271	0.043225
Gm6402	-1.6636	-0.19558	4.972315	0.025756	0.079542
Gm9839	-3.23116	-0.2531	15.54144	8.07E-05	0.003984
Gng8	-1.48347	0.681217	14.05705	0.000177	0.005128
Gnmt	-1.11027	-0.42261	5.227069	0.022238	0.072008
Gp1bb	-1.17673	2.496783	10.78249	0.001025	0.011448
Gpam	1.425383	4.417006	10.79574	0.001017	0.011421
Gpr161	1.872469	4.680483	12.03669	0.000522	0.008064
Gpr165	1.254482	0.970513	4.448607	0.03493	0.097897
Gpr26	1.593178	2.198451	19.12836	1.22E-05	0.002311
Gpr63	1.240192	1.004311	6.773167	0.009254	0.040506
Gprin2	1.972105	0.482702	16.15881	5.82E-05	0.00347
Gps2	-1.12255	5.721627	17.08677	3.57E-05	0.003084
Gramd1b	1.326243	2.85005	12.16464	0.000487	0.007835
Grcc10	-1.17061	6.368554	16.89721	3.95E-05	0.003172
Grid2	1.43718	-0.51294	9.433115	0.002131	0.016994
H2-DMa	-1.21982	1.349857	11.20067	0.000818	0.01018
H2-Ke2	-1.12513	5.852906	13.40553	0.000251	0.0061
H2-T24	1.483173	1.681903	6.757224	0.009337	0.040748
Hcfc1	1.146818	8.578247	14.29874	0.000156	0.004874
Hdac4	1.480865	3.730581	13.29332	0.000266	0.006279
Hdx	1.813528	1.604231	11.6166	0.000654	0.008999
Hectd2	1.325376	3.47498	8.292421	0.003981	0.023984
Hecw1	1.109564	5.202361	26.27891	2.95E-07	0.000496
Heg1	1.410635	5.168437	9.66639	0.001877	0.015734
Helz	1.381706	4.602012	24.14401	8.94E-07	0.000588
Herc1	1.210039	7.60909	22.47284	2.13E-06	0.000818
Herc2	1.249837	7.148423	20.12297	7.26E-06	0.001863

Hes6	-1.13073	6.184411	13.40357	0.000251	0.0061
Hhipl1	1.367953	1.866488	7.997865	0.004683	0.026469
Hipk2	1.897061	5.428741	19.38681	1.07E-05	0.002305
Hist1h2ag	-1.11255	-0.52362	4.958179	0.025968	0.080016
Hist1h2bj	-1.15257	0.406593	6.758923	0.009328	0.040744
Hist1h4b	-1.54904	0.80467	14.15036	0.000169	0.005092
Hist1h4h	-1.30464	-0.6033	5.772031	0.016283	0.058182
Hist2h2bb	-1.69987	-0.79728	7.032898	0.008003	0.036902
Hivep3	1.40581	1.237817	10.56199	0.001154	0.01211
Hmbox1	1.791085	4.705443	15.47347	8.37E-05	0.004074
Hspb1	-1.77301	-0.46615	12.03886	0.000521	0.008063
Huwe1	1.113472	9.146273	14.67867	0.000127	0.004552
Hyi	-1.14709	1.913277	6.229442	0.012564	0.04939
Igflr1	-1.12571	3.526278	9.553425	0.001996	0.016301
Igsf6	1.405341	0.343773	11.91737	0.000556	0.008238
Igsf9b	2.149806	2.203568	20.93637	4.75E-06	0.001458
Il3ra	-1.18774	0.459467	7.362416	0.00666	0.032919
Imp3	-1.20643	4.447446	14.67705	0.000128	0.004552
Ipw	1.379733	3.891881	18.30978	1.88E-05	0.00238
Isg15	-1.2765	-0.22718	10.8049	0.001012	0.011402
Itga6	1.236612	5.418824	6.586521	0.010275	0.043229
Itgb8	1.906099	3.687262	8.126188	0.004363	0.025378
Josd2	-1.19263	4.076653	16.19985	5.7E-05	0.00347
Kat6a	1.172227	7.409175	17.6155	2.7E-05	0.002831
Kcna1	1.161347	0.591073	7.574711	0.005919	0.030615
Kcna3	2.458217	2.516219	10.69615	0.001074	0.01166
Kcnb2	1.424241	1.116385	13.79405	0.000204	0.005601
Kcnh5	1.479921	-0.03215	10.86082	0.000982	0.011248
Kcnh7	1.436404	2.703022	12.05275	0.000517	0.008049
Kcnj3	1.71057	0.00667	8.549311	0.003457	0.022239
Kcnma1	1.273384	1.499007	15.89211	6.71E-05	0.003773
Kcnn3	1.408146	2.389437	18.57564	1.63E-05	0.00238
Kcnq3	1.999707	1.674859	28.87847	7.71E-08	0.000266
Kidins220	1.207609	7.751495	23.07979	1.55E-06	0.000693
Kif13b	1.637094	3.920098	13.74697	0.000209	0.005661
Kif26b	1.30313	3.715713	13.64982	0.00022	0.005743
Klf12	1.733765	5.306331	14.61418	0.000132	0.004552
Klf7	1.832694	3.116543	17.63327	2.68E-05	0.002831
Klhdc9	-1.39084	3.419942	12.51098	0.000405	0.007471
Klhl11	2.068372	2.931601	13.62902	0.000223	0.005773
Klhl28	1.482887	3.098623	6.712809	0.009572	0.041359

Klhl3	1.272765	1.009257	8.907264	0.00284	0.019887
Krt1	-1.17372	0.738685	7.106629	0.00768	0.036064
Krt8	-1.28466	0.132339	8.914609	0.002829	0.019848
Ksr2	1.550044	1.548651	15.49602	8.27E-05	0.004052
Lancl3	1.405629	0.864257	11.61883	0.000653	0.008999
Lcat	-1.24639	-0.11221	11.39449	0.000737	0.009705
Lcor	2.510665	3.013175	16.82127	4.11E-05	0.003172
Lgals4	-1.71009	-0.64516	8.388654	0.003776	0.023452
Lgals7	-1.24983	-0.31403	6.48079	0.010905	0.044959
Lmbrd2	2.120035	3.42089	12.52697	0.000401	0.007449
Lmln	1.37385	3.580508	9.20623	0.002412	0.018138
Lnpep	2.029171	3.823017	15.04471	0.000105	0.004361
Lphn3	1.111953	4.87382	14.66399	0.000128	0.004552
Lrch3	1.110127	5.712017	8.80024	0.003012	0.020538
Lrp1b	1.643148	0.781819	10.56271	0.001154	0.01211
Lrp2	1.119804	4.054476	12.76235	0.000354	0.007147
Lrp6	1.125296	6.185103	13.73825	0.00021	0.005661
Lrrc23	-1.33042	1.929652	10.96992	0.000926	0.010912
Lrrc46	-1.29193	0.909635	11.25874	0.000792	0.01003
Lrrc7	1.425171	3.099876	22.04919	2.66E-06	0.000966
Lrrc8b	1.597778	3.813988	8.383021	0.003787	0.023503
Lrrc9	1.380559	1.963974	17.3869	3.05E-05	0.00284
Lsm3	-1.45032	5.237332	18.45975	1.74E-05	0.00238
Lsm7	-1.28446	5.574476	18.3894	1.8E-05	0.00238
Ly6g5b	-1.17136	0.058314	7.685904	0.005565	0.02952
Ly75	1.200611	0.316598	3.878385	0.048912	0.123872
Lyst	1.733182	4.560856	14.55562	0.000136	0.00459
Lzts1	1.338874	3.351188	12.48363	0.000411	0.007535
Map1lc3a	-1.14365	4.164551	9.476904	0.002081	0.016715
Map3k13	1.357633	3.906818	16.724	4.32E-05	0.003172
MARCH11	-1.21399	0.235336	6.217533	0.012649	0.04955
Masp1	1.267381	5.061064	10.80526	0.001012	0.011402
Mdk	-1.26986	7.958931	17.43973	2.97E-05	0.00284
Mdm4	1.457911	6.282018	10.46495	0.001217	0.012482
Mdn1	1.269315	6.517645	18.32571	1.86E-05	0.00238
Med12l	1.429714	4.481823	12.8912	0.00033	0.006902
Med13l	1.104999	6.188017	14.60045	0.000133	0.004552
Mef2a	1.377023	4.115216	9.414019	0.002153	0.017113
Megf9	1.677178	4.943735	9.915367	0.001639	0.014633
Meig1	-1.30773	-0.89289	5.759572	0.016399	0.058407
Mgat5	2.054247	3.691716	18.02829	2.18E-05	0.002643

Mia	-1.32962	-0.29919	9.314613	0.002273	0.017522
Mib1	1.777975	5.186639	12.14978	0.000491	0.007871
Minos1	-1.12376	6.16726	16.91842	3.9E-05	0.003172
Mir1191	1.160674	0.401392	6.707094	0.009603	0.041395
Mir124a-2	1.810541	0.653584	16.22673	5.62E-05	0.00347
Mir16-1	1.589412	-0.50152	7.544	0.006021	0.03091
Mir186	1.501772	0.275197	5.052274	0.024594	0.077055
Mir25	-1.16552	0.341529	10.77183	0.001031	0.011489
Mir703	-1.29146	3.982785	5.119814	0.023654	0.074891
Mirg	-1.27367	4.598748	11.77483	0.0006	0.008643
Mnf1	-1.15677	5.648107	9.370417	0.002205	0.017316
Mob1a	1.143443	4.440417	5.436584	0.019719	0.066126
Mob1b	1.234803	4.089354	9.651758	0.001892	0.015787
Mospd2	1.204634	3.370512	5.672952	0.017228	0.060339
Moxd1	-1.15541	0.10669	5.93748	0.014822	0.054727
Mpv17l2	-1.17976	3.004434	10.43847	0.001234	0.012569
Mrpl14	-1.12729	4.283761	14.43806	0.000145	0.004708
Mrpl23	-1.30622	4.960087	17.01869	3.7E-05	0.003133
Mrpl33	-1.23066	5.237939	17.33078	3.14E-05	0.00284
Mrpl52	-1.12458	5.180932	14.85022	0.000116	0.004444
Mrpl54	-1.16661	4.276868	11.15581	0.000838	0.010304
Mrps16	-1.27127	3.831537	13.16157	0.000286	0.006483
Mrs2	1.602955	4.029131	14.74355	0.000123	0.004535
Mt1	-1.10082	3.118867	13.12349	0.000292	0.006564
Mt2	-1.20842	3.235242	11.79038	0.000595	0.008625
Mt3	-1.30652	6.070244	12.65641	0.000374	0.007217
Myeov2	-1.32004	5.220457	15.57294	7.94E-05	0.003946
Myl6	-1.1173	6.575545	14.76288	0.000122	0.004535
Mylpf	-1.2575	0.588535	9.379017	0.002195	0.017284
Myo16	1.149898	3.305477	14.66174	0.000129	0.004552
Myt1l	1.188984	5.511371	18.38108	1.81E-05	0.00238
N4bp2	1.935574	5.923515	11.60554	0.000658	0.009024
Naa25	1.298703	5.367776	13.37736	0.000255	0.006165
Nalcn	1.150572	3.716196	18.62974	1.59E-05	0.00238
Nanos3	-1.14956	1.028507	5.228279	0.022223	0.071991
Nat9	-1.1698	2.664176	8.964674	0.002752	0.019418
Nav3	1.818862	3.822291	15.50997	8.21E-05	0.004036
Nbea	1.264953	6.989596	24.89171	6.06E-07	0.000499
Nbeal1	1.681763	3.488032	11.13636	0.000847	0.010352
Ncam2	1.535911	2.349431	25.58544	4.23E-07	0.000496
Ndst1	1.28626	6.756393	10.8862	0.000969	0.011176

Ndufa13	-1.39598	5.850663	20.07195	7.46E-06	0.001863
Ndufa2	-1.22491	5.099771	14.5267	0.000138	0.004614
Ndufa5	-1.22103	4.632714	16.33208	5.32E-05	0.003449
Ndufa8	-1.18995	5.782637	17.9921	2.22E-05	0.002643
Nedd8	-1.38103	6.400413	20.9828	4.63E-06	0.001456
Nenf	-1.17021	2.698738	9.030647	0.002655	0.01911
Nfix	1.41006	4.437181	13.33916	0.00026	0.006225
Nhlrc2	1.192563	4.618805	14.53468	0.000138	0.004614
Nhlrc3	1.296349	3.060894	6.144446	0.013183	0.050937
Nhs	1.529382	2.876867	9.433699	0.00213	0.016994
Nkain3	1.314596	0.916492	12.3246	0.000447	0.007713
Nlrp1a	-2.33663	1.89007	14.58102	0.000134	0.004552
Nmb	-1.2349	1.659926	10.16327	0.001433	0.013603
Nos1	2.545938	-0.30619	18.8559	1.41E-05	0.00238
Nos1ap	1.115253	2.80339	9.404801	0.002164	0.017157
Npff	-2.05923	1.465159	12.47687	0.000412	0.007552
Npm2	-1.27953	0.666283	7.384298	0.00658	0.032638
Npnt	1.191027	2.772113	6.400848	0.011407	0.046378
Npy	-1.64504	1.640132	15.10359	0.000102	0.004341
Nr2c2	1.84137	5.721044	15.36262	8.87E-05	0.004106
Nr2c2ap	-1.55044	1.645056	12.91802	0.000325	0.006866
Nr3c2	1.223806	-0.6245	4.839496	0.027815	0.084041
Nrip1	1.77807	3.327623	9.276231	0.002321	0.017745
Nrp	-1.72905	2.624318	4.31719	0.037729	0.103559
Nrsn2	-1.15711	3.220179	8.674816	0.003226	0.021344
Nup98	1.439777	5.364788	14.91118	0.000113	0.00441
Nwd1	1.823648	2.044129	17.71194	2.57E-05	0.002831
Nyap2	1.888723	0.234556	11.19436	0.00082	0.010186
Olfr856-ps1	-1.16534	4.309481	5.184375	0.022791	0.073232
Onecut1	1.836227	-0.14937	6.041871	0.01397	0.052795
Pcbd2	-1.3497	1.553811	12.86509	0.000335	0.006926
Pcdh11x	1.836253	2.051288	29.19523	6.54E-08	0.000266
Pcdh9	1.296309	4.233492	12.44493	0.000419	0.00757
Pcdha12	1.603198	0.064415	8.399088	0.003754	0.02336
Pcdha2	1.670318	-0.43926	16.67523	4.44E-05	0.003186
Pcdha3	1.196008	0.129927	8.898984	0.002853	0.019935
Pcdha5	1.294614	-0.53924	10.24143	0.001373	0.013259
Pcdha6	1.752185	-0.49797	8.593534	0.003374	0.021951
Pcdha7	1.560573	-0.00012	6.983533	0.008226	0.037521
Pcdha9	1.647339	-0.07637	10.99622	0.000913	0.010829
Pcdhac1	1.272348	0.500408	5.956007	0.014667	0.054285

Pcdhac2	1.861691	2.591087	10.08942	0.001491	0.013878
Pcdhb2	1.721621	-0.41813	7.923551	0.00488	0.027142
Pcdhga1	1.681244	1.09205	12.07633	0.000511	0.008049
Pcdhga10	1.130685	2.428365	7.284323	0.006956	0.033825
Pcdhga11	1.399907	3.897024	11.54665	0.000679	0.009205
Pcdhga12	1.19932	3.137803	13.16846	0.000285	0.006482
Pcdhga2	1.423584	2.010739	9.772981	0.001771	0.015287
Pcdhga3	1.301439	2.760975	10.29638	0.001333	0.013037
Pcdhga4	1.353822	3.115732	12.96187	0.000318	0.00679
Pcdhga5	1.111251	2.214229	7.868254	0.005031	0.027668
Pcdhga6	1.276124	1.936222	8.004593	0.004666	0.026461
Pcdhga7	1.361761	2.591387	9.941427	0.001616	0.014534
Pcdhga8	1.444029	1.808635	11.33444	0.000761	0.009864
Pcdhga9	1.326916	2.032254	8.460653	0.003629	0.02286
Pcdhgb1	1.683327	1.569055	12.11114	0.000501	0.007971
Pcdhgb2	1.145134	2.258245	14.23057	0.000162	0.004951
Pcdhgb4	1.314889	1.677036	9.000419	0.002699	0.019266
Pcdhgb5	1.291509	2.075513	8.259044	0.004055	0.024228
Pcdhgb6	1.371498	3.687627	8.909803	0.002836	0.019878
Pcdhgc3	1.319643	6.21387	11.98807	0.000535	0.008188
Pcdhgc4	1.19926	2.712505	9.907059	0.001646	0.014671
Pcsk1n	-1.26798	3.44369	10.88802	0.000968	0.011175
Pdap1	-1.29663	7.352556	15.4722	8.37E-05	0.004074
Pdcd5	-1.10796	5.554279	16.66555	4.46E-05	0.003186
Pde3a	1.183225	0.767918	5.066656	0.02439	0.076591
Pdpr	2.01147	3.833233	9.365477	0.002211	0.017333
Pds5a	1.199183	6.565692	8.685871	0.003207	0.021269
Peg3	1.3963	7.10989	8.695665	0.00319	0.021248
Perp	-1.60746	0.034744	13.7537	0.000208	0.005661
Pet100	-1.30681	3.179673	11.37838	0.000743	0.00975
Phf16	1.170001	3.510929	6.610355	0.010139	0.042836
Plag1	1.310721	3.443369	5.655541	0.0174	0.060618
Plcxd3	2.481108	1.543862	19.18477	1.19E-05	0.002311
Plxna2	1.232451	7.909549	15.15176	9.92E-05	0.004289
Plxna4	1.275236	6.116857	19.79704	8.61E-06	0.001984
Polr2l	-1.21193	4.744317	14.62774	0.000131	0.004552
Pomc	-1.56513	2.407797	10.52445	0.001178	0.012223
Pop5	-1.49903	3.865034	17.2227	3.32E-05	0.002942
Pou3f4	1.304893	3.923369	9.024847	0.002663	0.019113
Ppm1l	1.634264	4.839566	20.59684	5.67E-06	0.001599
Ppp1r12a	1.169785	5.318492	10.1872	0.001414	0.013494

Ppp1r12b	1.833575	3.832999	19.75625	8.8E-06	0.001993
Prdm10	1.496236	4.294735	9.317701	0.00227	0.017522
Prdm11	1.236128	1.511784	11.02056	0.000901	0.010707
Prdx5	-1.41131	5.199049	19.28169	1.13E-05	0.002311
Prg4	-1.33647	-0.24477	11.32707	0.000764	0.009866
Prox1	1.230552	4.088017	22.86384	1.74E-06	0.000707
Prr19	-1.15801	-0.4213	7.349738	0.006707	0.032987
Prr22	-1.52885	0.438722	11.02892	0.000897	0.010696
Prrc2c	1.238801	7.974996	14.74945	0.000123	0.004535
Psmb3	-1.13011	7.147476	15.13756	1E-04	0.00429
Psmg3	-1.28337	3.367266	14.06657	0.000176	0.005127
Psmg4	-1.17771	5.038896	11.92411	0.000554	0.008234
Ptar1	2.236005	2.783682	23.09303	1.54E-06	0.000693
Ptch1	1.398424	5.458806	8.582893	0.003393	0.021989
Ptp4a1	-1.47924	4.008398	20.6714	5.45E-06	0.00157
Ptpn4	1.940394	3.850047	12.30597	0.000452	0.007742
Ptprt	1.946253	1.625616	17.61719	2.7E-05	0.002831
Pygo1	1.343313	4.079279	16.87181	4E-05	0.003172
Rab11fip4	1.333375	5.007179	14.14346	0.000169	0.0051
Rab24	-1.12156	5.247995	15.57564	7.93E-05	0.003946
Ralgps1	1.779761	4.691633	18.10476	2.09E-05	0.00258
Ramp3	-1.16876	-0.36197	7.615208	0.005788	0.030264
Rangrf	-1.21548	3.899681	17.37882	3.06E-05	0.00284
Rapgef5	1.233415	3.958123	12.45843	0.000416	0.007567
Rassf8	1.249635	1.660934	6.094884	0.013557	0.051815
Rcor1	1.353459	4.634013	8.827985	0.002966	0.02038
Reln	1.472696	6.718251	26.09471	3.25E-07	0.000496
Rest	1.302787	4.149753	6.279117	0.012217	0.048475
Rims1	1.369012	2.168199	13.87568	0.000195	0.005442
Rims3	1.259465	2.343427	12.88444	0.000331	0.006907
Rnaseh2c	-1.11268	4.856157	16.82026	4.11E-05	0.003172
Rnf223	-1.13358	1.865779	10.75678	0.001039	0.01152
Romo1	-1.24031	5.585554	13.68349	0.000216	0.005688
Rpl12	-1.16425	8.960003	15.65513	7.6E-05	0.00394
Rpl13	-1.22318	8.425992	10.69211	0.001076	0.011671
Rpl13a	-1.10852	9.489128	11.73664	0.000613	0.00874
Rpl18	-1.11307	8.808979	14.32809	0.000154	0.004838
Rpl18a	-1.40194	8.489068	12.03981	0.000521	0.008063
Rpl21	-1.3758	9.012828	17.78312	2.48E-05	0.002804
Rpl23	-1.14278	9.300492	16.26594	5.5E-05	0.00347
Rpl28	-1.27541	8.947792	10.45932	0.00122	0.012484

Rpl29	-1.19517	8.863916	14.34367	0.000152	0.004838
Rpl31	-1.14649	10.41732	12.90201	0.000328	0.006894
Rpl31-ps12	-1.65946	1.324112	7.892775	0.004963	0.027459
Rpl35	-1.17815	8.012264	7.613779	0.005792	0.030276
Rpl35a	-1.12775	8.411085	14.61547	0.000132	0.004552
Rpl36	-1.4548	7.984897	15.23834	9.48E-05	0.004191
Rpl36al	-1.41427	6.314296	18.57054	1.64E-05	0.00238
Rpl37	-1.22312	8.883172	15.81665	6.98E-05	0.003812
Rpl37a	-1.47984	8.327329	16.015	6.28E-05	0.003649
Rpl38	-1.25676	8.029534	15.4348	8.54E-05	0.004084
Rpl39	-1.11548	8.329423	12.17192	0.000485	0.007835
Rpl41	-1.52298	8.705834	13.06982	0.0003	0.006573
Rpl8	-1.10724	8.343902	12.05333	0.000517	0.008049
Rplp1	-1.20102	8.963204	14.60781	0.000132	0.004552
Rplp2	-1.72984	7.901284	20.04857	7.55E-06	0.001863
Rpp21	-1.14411	3.920895	9.705886	0.001837	0.015573
Rpph1	-1.46847	0.701161	6.207379	0.012722	0.049711
Rps12	-1.48421	8.658737	16.49169	4.89E-05	0.003301
Rps14	-1.49352	9.759923	20.88558	4.88E-06	0.001465
Rps15	-1.53287	8.341066	12.10286	0.000503	0.007994
Rps15a-ps6	-1.13712	1.56967	5.403884	0.020092	0.066941
Rps16	-1.10038	8.644803	13.24092	0.000274	0.006351
Rps18	-1.11812	9.522147	13.44025	0.000246	0.006078
Rps19	-1.44562	9.129901	14.23121	0.000162	0.004951
Rps19-ps3	-1.70455	1.290623	9.844551	0.001703	0.014946
Rps20	-1.15051	8.633077	13.10565	0.000294	0.006567
Rps21	-1.30495	8.116402	17.40582	3.02E-05	0.00284
Rps24	-1.17232	8.729518	13.92305	0.00019	0.005382
Rps28	-1.50523	8.96497	15.14991	9.93E-05	0.004289
Rps4y2	-1.19564	5.206366	16.84993	4.05E-05	0.003172
Rps5	-1.27681	8.665955	14.85836	0.000116	0.004444
Rps8	-1.16038	9.242084	14.53974	0.000137	0.004614
Rps9	-1.12392	9.209017	12.59489	0.000387	0.007343
Rsbn1	1.298926	4.757458	8.499733	0.003552	0.022569
Rsg1	1.3923	1.144316	8.618606	0.003327	0.021783
Rsph1	-1.13585	0.341127	5.644446	0.017511	0.060848
S100a1	-1.11718	0.174764	8.177599	0.004241	0.024909
S100a10	-1.30315	4.126349	13.72466	0.000212	0.005665
S100a16	-1.15039	2.472981	14.07409	0.000176	0.005127
S100a6	-1.29306	0.578899	9.998013	0.001567	0.014267
S100pbp	1.155094	3.933805	6.353302	0.011716	0.047226

Sacs	3.013629	1.850426	27.43793	1.62E-07	0.000448
Samd5	1.418856	1.950325	9.269057	0.002331	0.017756
Scarb2	1.139259	5.31506	9.913503	0.001641	0.014633
Sctr	-1.2676	-0.33816	5.663828	0.017318	0.0605
Scube2	1.177667	0.038291	7.155522	0.007473	0.035443
Sdk2	1.554259	3.110314	12.04629	0.000519	0.008058
Selm	-1.23989	2.969057	11.29611	0.000777	0.009957
Sema3e	1.133561	1.498317	14.95044	0.00011	0.00441
Sepsecs	1.166678	2.627471	8.979469	0.00273	0.01933
Sept1	-1.32343	2.463707	11.41453	0.000729	0.009655
Sepw1	-1.28939	7.735396	11.79924	0.000593	0.008593
Serf2	-1.20293	7.788494	16.74968	4.26E-05	0.003172
Serp2	-1.28989	3.758316	13.85014	0.000198	0.005483
Sertad1	-1.11955	1.546584	12.57746	0.00039	0.007384
Sesn3	1.397673	4.337472	18.46782	1.73E-05	0.00238
Sfn	-1.87924	-0.08908	14.20537	0.000164	0.004978
Sh2d7	-1.30147	0.079088	7.575296	0.005917	0.030615
Sh3bgrl3	-1.22236	5.512054	12.03033	0.000523	0.008081
Shank2	1.14513	4.520646	12.83038	0.000341	0.006995
Shc3	1.451855	2.647133	12.33736	0.000444	0.007704
Shfm1	-1.24813	7.093693	18.76589	1.48E-05	0.00238
Shroom4	1.506759	2.459186	5.15301	0.023206	0.074118
Siah3	1.581912	3.368374	12.96256	0.000318	0.00679
Sik2	1.766665	3.840019	13.7323	0.000211	0.005661
Six2	-2.17558	1.722135	8.024594	0.004615	0.026273
Slc16a14	1.657316	2.718963	17.58744	2.74E-05	0.002832
Slc26a2	1.558917	2.15725	8.652935	0.003265	0.02154
Slc30a3	-1.40172	-0.03052	9.218894	0.002395	0.018079
Slc36a4	1.642954	3.91521	10.26309	0.001357	0.01316
Slc39a5	-2.76336	0.690177	17.54822	2.8E-05	0.00284
Slc44a1	1.141832	3.380938	6.117588	0.013384	0.051467
Slc7a10	-1.39212	1.836285	8.233632	0.004112	0.0244
Slc9a7	1.842957	0.356801	12.8451	0.000338	0.006969
Slfn9	1.201542	3.917959	8.08294	0.004468	0.025799
Slit1	1.152341	5.572597	18.58417	1.63E-05	0.00238
Smim11	-1.21889	5.312722	14.40717	0.000147	0.004746
Smim4	-1.14779	3.500216	12.56267	0.000394	0.007389
Snapc5	-1.347	4.484627	15.89233	6.7E-05	0.003773
Snhg8	-1.17717	4.448497	13.5245	0.000235	0.005927
Snora28	1.144035	0.535526	7.856469	0.005064	0.027804
Snora31	-1.17035	0.131996	10.45508	0.001223	0.012484

Snora68	-1.46519	-0.77671	6.256943	0.012371	0.048917
Snord104	-1.47156	3.18886	14.93607	0.000111	0.00441
Snord2	1.191475	0.308479	4.154328	0.041528	0.110413
Snord64	1.966932	0.685151	6.969355	0.008292	0.037782
Snord99	-1.5133	0.037224	12.30708	0.000451	0.007742
Snrpd2	-1.19237	6.736997	16.19071	5.73E-05	0.00347
Snrpg	-1.22597	6.767004	17.58327	2.75E-05	0.002832
Sntb2	1.221416	2.867029	10.35598	0.001291	0.012799
Snx22	-1.6808	3.690697	11.17324	0.00083	0.01024
Snx29	1.38522	3.106858	15.70425	7.41E-05	0.003892
Socs4	1.19008	3.584108	6.572915	0.010354	0.043463
Sod1	-1.13495	6.685787	15.26107	9.36E-05	0.004174
Sorl1	2.061359	5.063529	17.09926	3.55E-05	0.003084
Spa17	-1.36817	2.477321	8.02539	0.004613	0.026273
Sphkap	1.417268	2.710074	8.518618	0.003515	0.022426
Spred2	1.253892	5.286657	7.582688	0.005893	0.030572
Sprtn	1.215571	3.559658	6.221961	0.012618	0.049478
Ssh1	1.293482	3.602075	14.06134	0.000177	0.005127
St6gal2	1.966639	2.561494	10.20259	0.001402	0.01345
Stac3	-1.1648	1.63911	12.0558	0.000516	0.008049
Stard10	-1.27453	2.487239	10.71086	0.001065	0.011647
Stmn1	-1.17549	9.338249	12.46181	0.000415	0.007564
Stra13	-1.14783	4.902291	16.85987	4.02E-05	0.003172
Strn	1.45796	3.829379	9.404024	0.002165	0.017157
Stx1b	1.162585	3.208823	9.686431	0.001856	0.015632
Supt4a	-1.1595	5.640387	12.40202	0.000429	0.007628
Syne2	1.180858	8.085965	18.69783	1.53E-05	0.00238
Syt14	1.94391	2.670329	11.69779	0.000626	0.008824
Sytl5	1.854586	0.711377	14.62691	0.000131	0.004552
Taf10	-1.14029	5.157479	9.448082	0.002114	0.016905
Tbck	1.394137	3.094197	10.71301	0.001064	0.011647
Tbl1x	1.164124	5.694238	8.606471	0.00335	0.021867
Tceal3	-1.13768	3.715123	12.26833	0.000461	0.007788
Tceb2	-1.3187	7.410333	16.87432	3.99E-05	0.003172
Tekt4	-1.30354	-0.9023	7.153693	0.007481	0.035467
Tenm1	2.487991	1.818748	15.97473	6.42E-05	0.003696
Tenm3	1.1427	6.823606	13.84647	0.000198	0.005483
Tenm4	2.044544	6.938024	15.80475	7.02E-05	0.003821
Tfap2a	-1.24103	0.794686	8.343258	0.003871	0.023729
Thpo	-1.36866	-0.44443	7.202757	0.007279	0.034918
Timm13	-1.37156	5.714693	14.89606	0.000114	0.00441

Timm8b	-1.11772	6.188553	14.28689	0.000157	0.004874
Tle6	-1.13262	0.048383	8.352983	0.003851	0.023702
Tmem132d	1.321475	-0.12146	7.595236	0.005852	0.03044
Tmem178b	2.322611	3.252894	9.622811	0.001922	0.01597
Tmem179b	-1.10867	3.815001	17.32871	3.14E-05	0.00284
Tmem191c	-1.17152	2.701899	10.9359	0.000943	0.011
Tmem196	1.101025	-0.62559	5.003095	0.025302	0.07872
Tmem245	1.588336	3.993604	15.91148	6.64E-05	0.003773
Tmem256	-1.19997	5.208917	12.81675	0.000344	0.007033
Tmem42	-1.13249	2.301867	8.497191	0.003557	0.022589
Tmem53	-1.11112	1.183673	11.74817	0.000609	0.008729
Tmem74	1.356384	1.247409	9.160885	0.002472	0.018333
Tmem88	-1.23922	2.019339	12.07467	0.000511	0.008049
Tnfrsf18	1.414745	0.948684	9.76505	0.001779	0.015324
Tnfrsf4	-1.19469	0.473686	7.97639	0.004739	0.02662
Tnks	1.836309	6.414188	12.23063	0.00047	0.007835
Tnni3	-1.6977	0.375094	8.12236	0.004372	0.02541
Tnpo1	1.412169	7.389997	13.34428	0.000259	0.006219
Tnr	2.748304	1.265542	26.33612	2.87E-07	0.000496
Tomm6	-1.23804	5.663237	16.04381	6.19E-05	0.003609
Tomm7	-1.1919	6.913723	12.41898	0.000425	0.007608
Tppp3	-1.19855	2.945247	9.557881	0.001991	0.01629
Trappc6a	-1.26225	3.021136	12.98892	0.000313	0.006745
Trim43c	1.38657	-0.73269	9.736001	0.001807	0.015453
Trim56	1.520683	2.201869	9.814558	0.001731	0.015064
Trip11	1.693491	4.056964	13.99529	0.000183	0.005255
Tssc4	-1.13219	5.038514	19.11746	1.23E-05	0.002311
Tstd1	-1.47665	1.037365	10.84209	0.000992	0.011276
Ttbk2	1.124858	6.34403	22.95537	1.66E-06	0.000707
Ttc28	1.263798	6.940998	18.44953	1.74E-05	0.00238
Ttc37	1.296672	4.435353	10.71405	0.001063	0.011647
Ttc9b	-1.14466	3.21161	6.798684	0.009123	0.0402
Ttll4	1.378866	5.786668	11.20735	0.000815	0.010153
Twist1	-1.74519	1.893949	7.960734	0.00478	0.02679
Uba52	-1.1065	9.980958	11.51187	0.000692	0.009316
Ubald2	-2.24485	1.190599	11.97488	0.000539	0.008204
Ubr4	1.146124	7.635513	20.15271	7.15E-06	0.001863
Ubxn1	-1.10044	5.763993	14.80264	0.000119	0.004475
Uhmk1	1.698495	3.152363	9.769525	0.001774	0.015306
Unc13c	1.24532	1.009152	13.46151	0.000244	0.00602
Upf3a	-1.14064	4.551259	15.28056	9.27E-05	0.004158

Upk1a	-1.11722	-0.10755	5.916471	0.015	0.055045
Uprt	1.750611	1.533759	6.494647	0.01082	0.044719
Use1	-1.1527	5.232062	9.723279	0.001819	0.015491
Usp34	1.325834	6.677439	12.7098	0.000364	0.007212
Usp53	1.310515	2.403172	9.351775	0.002228	0.017394
Utp14b	1.259525	2.328579	9.702893	0.00184	0.015589
Vamp5	-1.12459	0.118988	7.030339	0.008014	0.036936
Vamp8	-1.29423	2.049835	14.69151	0.000127	0.004552
Vgll2	-1.43573	-0.14709	4.260491	0.039009	0.105893
Vkorc1	-1.38034	3.427133	18.45419	1.74E-05	0.00238
Vps13a	1.164539	4.479069	18.60366	1.61E-05	0.00238
Vps13b	1.141857	5.362688	17.61829	2.7E-05	0.002831
Vps13c	1.801612	4.657138	15.64318	7.65E-05	0.00394
Vps13d	1.267797	5.645336	23.48283	1.26E-06	0.000645
Wasf3	1.101944	3.481533	6.475384	0.010938	0.045056
Wdfy2	1.796163	3.62964	12.45252	0.000417	0.00757
Wdfy3	1.148897	7.611394	18.90747	1.37E-05	0.00238
Wdr52	1.111759	2.630508	7.923222	0.00488	0.027142
Wnk3	1.111342	5.53259	15.83964	6.89E-05	0.003803
Xkr4	2.2932	1.707996	18.30979	1.88E-05	0.00238
Xkr7	1.790512	2.869899	11.71202	0.000621	0.008802
Xpnpep3	1.269781	3.115469	7.898549	0.004947	0.027404
Xpo4	1.815103	4.196897	16.16405	5.81E-05	0.00347
Xrn1	2.034946	4.756355	14.8396	0.000117	0.004444
Xylt1	1.85225	3.456239	12.5195	0.000403	0.007451
Yod1	1.196445	2.540133	7.382707	0.006585	0.032644
Zbed6	2.111239	5.57454	8.70167	0.003179	0.021236
Zbtb20	1.632369	2.069767	20.78255	5.14E-06	0.001513
Zbtb37	1.859659	2.921653	25.90464	3.59E-07	0.000496
Zbtb38	1.663504	2.227117	17.03431	3.67E-05	0.003132
Zc3h12b	1.629038	1.162345	11.19391	0.000821	0.010186
Zc3h12c	1.172969	4.963085	11.90486	0.00056	0.008259
Zfand2b	-1.14756	4.102624	11.04846	0.000888	0.010634
Zfp109	1.330035	2.228677	6.676152	0.009771	0.041887
Zfp169	1.340342	2.981307	13.42686	0.000248	0.006089
Zfp296	-1.41611	-0.29979	9.777212	0.001767	0.015261
Zfp369	2.256045	3.049385	17.79116	2.47E-05	0.002804
Zfp382	1.370267	2.750015	13.20407	0.000279	0.006423
Zfp398	1.176324	3.297424	14.88692	0.000114	0.00441
Zfp428	-1.23791	5.036374	14.07116	0.000176	0.005127
Zfp442	1.576232	1.675122	7.987409	0.00471	0.026569

Zfp458	1.108168	1.212729	10.28811	0.001339	0.013077
Zfp536	2.070484	4.284665	28.89152	7.65E-08	0.000266
Zfp551	1.389092	1.711071	8.129135	0.004356	0.025371
Zfp579	-1.27126	5.437566	14.08308	0.000175	0.005127
Zfp580	-1.12729	3.95784	8.658926	0.003255	0.02149
Zfp618	1.82913	4.676228	14.07745	0.000175	0.005127
Zfp619	1.103324	3.17769	13.36059	0.000257	0.006209
Zfp81	1.218513	3.929203	10.07896	0.0015	0.013913
Zfpm2	1.133419	3.120055	12.55284	0.000396	0.007408
Zglp1	-1.21891	1.13933	8.280878	0.004006	0.024095
Zkscan16	1.319344	1.175036	4.523565	0.033431	0.095301
Zkscan2	1.231335	4.460519	14.68828	0.000127	0.004552
Zkscan7	1.263579	2.563544	8.347364	0.003862	0.023722

Appendix Table 2: PA2 deregulated genes

Gene ID	logFC	logCPM	LR	PValue	FDR
2410137F16Rik	1.375975	1.125381	5.839117	0.015674	0.92136
3110021A11Rik	1.472895	-0.36248	11.36839	0.000747	0.397084
4930455C13Rik	1.289068	-0.73006	5.586199	0.018103	0.92136
AI606473	-1.48794	2.387366	16.79578	4.16E-05	0.063919
Alx1	-2.80159	0.446978	9.800454	0.001745	0.560735
C130021I20Rik	-1.51719	0.788954	8.720906	0.003146	0.663921
Cdh3	-1.11624	0.357543	4.100674	0.042866	0.92136
Chrna7	1.176552	1.546752	5.831131	0.015745	0.92136
Cntnap5a	1.297273	-0.86058	5.694204	0.017021	0.92136
Cntnap5b	1.56071	0.237802	6.289386	0.012146	0.92136
Col8a1	-2.12645	-0.49369	4.060336	0.043902	0.92136
Csf2rb2	-1.49059	-1.03465	6.029445	0.014069	0.92136
Dbx1	1.199766	4.125421	24.32489	8.14E-07	0.00592
Dct	-1.28738	2.20927	12.00753	0.00053	0.318377
Ebf1	1.154484	3.546801	22.50073	2.1E-06	0.009677
Egr3	1.514941	0.059908	9.622646	0.001922	0.565128
Erbb4	1.439934	2.252102	8.114557	0.004391	0.766187
Fam135b	1.354867	-0.61144	6.179474	0.012924	0.92136
Fam217b	1.130457	0.661525	7.569203	0.005937	0.811513
Frmd7	-1.16082	0.164286	7.620011	0.005772	0.811513
Gabbr2	1.135102	1.606292	10.43529	0.001236	0.471115
Gjb2	-1.8318	0.387615	5.514841	0.018856	0.92136
Gpr161	1.157476	4.680483	4.781493	0.028767	0.92136
Gpr26	1.166205	2.198451	10.44708	0.001228	0.471115
Gprin2	1.298518	0.482702	7.220252	0.007209	0.877049
Hbb-bh1	-1.32465	7.118271	6.655916	0.009883	0.92136
Hemgn	-1.96168	1.543236	4.420255	0.035515	0.92136
Hist1h1a	2.167643	-1.05249	13.50862	0.000237	0.234417
Hist1h1b	2.602779	-0.72899	17.70724	2.58E-05	0.057077
Hist1h1d	1.302365	-0.24869	8.096436	0.004435	0.766187
Hist1h1e	1.818167	0.373351	12.43931	0.00042	0.284351
Hist1h2af	1.508723	2.346636	6.530975	0.010601	0.92136
Hist1h2bg	1.314979	-0.15295	9.178374	0.002449	0.573627
Hist1h2bn	1.740887	-0.99176	7.136678	0.007552	0.884514
Hist1h4a	1.377632	-1.1202	6.405658	0.011376	0.92136
Hist1h4f	2.380765	-0.96538	15.96788	6.44E-05	0.089037
Hist1h4h	1.186827	-0.6033	5.04129	0.02475	0.92136
Hist1h4k	1.621516	0.05712	10.60877	0.001126	0.471115
Hist2h2bb	2.553215	-0.79728	17.1239	0.000035	0.060495

Hist2h4	2.751735	-0.72786	13.37484	0.000255	0.234961
Irx5	-1.1692	0.253105	4.10825	0.042674	0.92136
Kcna3	1.54769	2.516219	4.503321	0.033829	0.92136
Kcnq3	1.170605	1.674859	10.28775	0.001339	0.474545
Lars2	2.533208	8.854135	21.01736	4.55E-06	0.015725
Meig1	-1.1549	-0.89289	4.523722	0.033428	0.92136
Msx2	-1.26761	1.357208	10.37367	0.001278	0.471115
Nos1	1.67979	-0.30619	8.585586	0.003388	0.682594
Pcdha12	1.10128	0.064415	4.043772	0.044335	0.92136
Pcdha2	1.144361	-0.43926	7.910157	0.004916	0.799251
Pcdhga1	1.206927	1.09205	6.360461	0.011669	0.92136
Pcdhgb1	1.197129	1.569055	6.268258	0.012292	0.92136
Perp	-1.32018	0.034744	9.456184	0.002104	0.573339
Pitx2	-4.7446	0.631186	9.687094	0.001856	0.563992
Plcxd3	1.330921	1.543862	5.912378	0.015035	0.92136
Ralgps1	1.107682	4.691633	7.261628	0.007044	0.877049
Raver1-fdx1l	1.260868	3.162548	9.665846	0.001877	0.563992
Rmrp	3.037495	2.033332	17.48817	2.89E-05	0.057077
Rn45s	3.15516	12.21427	24.22584	8.57E-07	0.00592
Rpph1	2.228112	0.701161	14.1144	0.000172	0.198112
Rsg1	1.229102	1.144316	6.765745	0.009292	0.92136
Sacs	1.593493	1.850426	8.454038	0.003642	0.708973
Scarna9	2.471549	-0.70672	19.10659	1.24E-05	0.034168
Serpinb1b	-1.6577	-0.93456	7.459233	0.006311	0.828174
Siah3	1.450913	3.368374	10.98219	0.00092	0.454044
Six2	-1.67497	1.722135	4.922298	0.026512	0.92136
Six6	1.881747	1.439062	7.610531	0.005803	0.811513
Slc14a1	-1.15212	-0.43964	6.484926	0.010879	0.92136
Snord17	1.808382	-0.72806	12.54137	0.000398	0.284351
Sorl1	1.261155	5.063529	6.707069	0.009603	0.92136
Sv2c	1.314702	0.337716	9.231957	0.002378	0.573339
Sytl5	1.211748	0.711377	6.426157	0.011245	0.92136
Tbx15	-1.80958	0.573199	10.14754	0.001445	0.48705
Tekt4	-1.24037	-0.9023	6.492119	0.010835	0.92136
Tnr	1.354219	1.265542	6.951823	0.008373	0.918414
Twist1	-1.62539	1.893949	6.957485	0.008347	0.918414
Xkr7	1.222064	2.869899	5.622179	0.017735	0.92136
Zbtb20	1.237337	2.069767	12.14382	0.000493	0.309384
Zfp536	1.144457	4.284665	9.293894	0.002299	0.573339

Appendix Table 3: PolyA Pool deregulated genes

Gene ID	logFC	logCPM	LR	PValue	FDR
2310002D06Rik	1.3561239	-0.2360265	6.8605015	0.0088122	0.1851118
2410066E13Rik	1.2675779	4.16353375	14.121423	0.0001714	0.0815076
2410137F16Rik	2.2442006	1.1253809	13.467701	0.0002427	0.0882681
2610005L07Rik	1.1008607	5.94382492	10.249338	0.0013673	0.1242324
3110021A11Rik	1.4262511	-0.3624843	12.492613	0.0004086	0.095701
3110021N24Rik	1.2671154	1.30611769	5.8616267	0.0154745	0.1998302
3300002I08Rik	1.133717	0.29492701	9.2071332	0.0024107	0.1429889
4921525B02Rik	1.4563935	0.02386542	8.0024897	0.0046713	0.1634366
4930444P10Rik	-1.145201	0.31228925	6.480351	0.0109073	0.1881628
4930455C13Rik	1.7289492	-0.7300606	10.471531	0.0012123	0.1219646
4930506C21Rik	1.1027136	0.13830233	6.4100966	0.0113473	0.1881628
4930565N06Rik	1.6518598	-0.4756614	4.3784136	0.0363969	0.2449998
4932418E24Rik	1.4129412	0.72051371	8.6388123	0.0032908	0.1538584
A630007B06Rik	1.2400253	4.01270748	6.0784247	0.0136843	0.1940084
A730008H23Rik	1.3625055	0.69812514	6.7507036	0.0093711	0.1876902
Abhd2	1.3297558	4.48023113	8.8128057	0.0029912	0.1517881
Adamtsl3	1.1538691	-0.6429793	7.4733082	0.006262	0.1729126
Adarb2	1.1262135	2.59455063	14.369209	0.0001502	0.0798578
Adcy9	1.2851365	2.10894415	11.793681	0.0005943	0.1064152
Aff2	1.5310844	4.24818856	8.0190285	0.0046288	0.1634366
AI606473	-1.485263	2.38736639	24.897352	6.05E-07	0.0041782
Alx1	-2.982993	0.44697804	18.194085	1.99E-05	0.0335573
Ankfn1	1.4739203	0.63494283	5.0269185	0.0249563	0.223378
Apom	-1.167087	0.16248503	4.1627255	0.0413229	0.2526136
Arid5b	1.190682	3.06400427	6.5194808	0.0106699	0.1881628
Atf7	1.3124589	2.8917443	10.095082	0.0014867	0.1249683
Atp2b3	1.2144098	1.23826606	10.190616	0.0014116	0.1242324
BC005561	1.2989352	3.93039233	6.3217723	0.0119264	0.188537
Bmpr2	1.324022	4.51440901	7.4429646	0.0063685	0.1729126
Brwd3	1.2685316	4.33798559	10.635464	0.0011094	0.1176564
Btbd7	1.1456785	5.44781968	6.4485061	0.0111046	0.1881628
Btbd8	1.5300055	-0.2200918	11.348644	0.000755	0.1146651
Cacna1e	1.1151647	3.72220386	6.3359227	0.0118317	0.1881628
Cbl	1.6435622	5.22216452	7.5792149	0.0059045	0.1717113
Ccnt1	1.2938552	4.03688978	7.1266275	0.0075947	0.180484
Cdh3	-1.209288	0.35754329	7.012202	0.0080956	0.1810886
Cdh6	1.4615827	3.5357246	7.9807352	0.0047278	0.1645789
Cdk6	1.4627727	1.62898609	12.203993	0.0004769	0.1028121
Cdkl5	1.6271567	2.76514046	13.273384	0.0002692	0.0930089

Cep85l	1.2929263	1.76597427	4.3427432	0.037167	0.2464186
Cers6	1.2339961	4.32792844	8.3612326	0.0038331	0.156264
Chrm3	1.3203625	-0.2451778	6.3985214	0.0114215	0.1881628
Chrna7	1.3984376	1.54675243	9.6790738	0.0018638	0.1343934
Clcn5	1.376978	4.27801912	7.1427949	0.0075266	0.1802726
Cntnap5a	2.1374694	-0.8605756	13.113226	0.0002932	0.0940169
Cntnap5b	2.3833674	0.2378017	13.168133	0.0002848	0.0940169
Col8a1	-1.693433	-0.4936914	4.0161821	0.0450656	0.2581048
Cpeb4	1.4630496	3.51367747	6.7824586	0.0092058	0.1854598
Crabp1	-1.169186	2.16836668	11.796491	0.0005934	0.1064152
Creb5	1.2320205	2.47485887	7.290717	0.0069312	0.1760452
Csmd1	1.4411809	1.91787868	13.413759	0.0002498	0.0885137
Csmd3	1.319581	1.81993404	9.4540025	0.0021069	0.1373453
Cxcl14	-1.124257	1.86567352	12.047497	0.0005186	0.1054018
Cyp4x1	1.1939767	-0.367116	5.16647	0.023027	0.2177747
D10Bwg1379e	1.2792087	2.46890327	12.149545	0.000491	0.1028121
D130040H23Rik	1.2194447	1.62674688	10.144556	0.0014473	0.1242324
D830031N03Rik	1.6390382	4.47922221	9.508402	0.0020453	0.1359519
Dgkh	1.487572	1.24433179	11.398733	0.0007349	0.1139708
Dnajb14	1.2447785	1.90550765	8.7363789	0.0031192	0.1517881
Dnm3os	-1.371207	0.30909778	5.3611701	0.02059	0.2110929
Dok6	1.496639	-0.170196	6.7838773	0.0091985	0.1854598
Dpy19l4	1.3794931	3.72898839	7.4205904	0.0064482	0.1733728
Dpysl3	1.3274025	7.90096013	11.988768	0.0005352	0.1064152
E330009J07Rik	1.5137445	3.17154964	10.243104	0.001372	0.1242324
Ebf1	1.1880799	3.54680125	27.542738	1.54E-07	0.0021236
Egr3	1.4021273	0.05990763	9.8383048	0.0017091	0.1309183
Eif2c2	1.3488753	6.11243439	7.710042	0.0054914	0.1703473
Eif2c3	1.3022736	3.34211472	9.8845878	0.0016667	0.1309183
Elfn2	1.1379369	1.94420735	10.368031	0.0012822	0.123052
Elk4	1.2084196	2.42192488	8.1253693	0.004365	0.1611608
Erbb4	2.2067006	2.25210207	16.452833	4.99E-05	0.0428617
Ern1	1.3481517	3.47577446	7.5407683	0.0060318	0.1729126
F730043M19Rik	1.168694	1.56264178	7.9534566	0.0047996	0.1658249
Fam135b	2.0320269	-0.6114364	12.305926	0.0004515	0.1006458
Fam171b	1.118847	5.86723463	5.1571344	0.0231511	0.2177747
Fam217b	1.3756237	0.66152493	12.569766	0.000392	0.094991
Fbxo48	1.2329618	2.35005849	5.675227	0.0172061	0.2038935
Fktn	1.2554153	4.19806868	7.2065746	0.0072637	0.1778994
Flrt1	1.1806204	3.24058121	11.721406	0.0006179	0.1064152
Fos	1.1917105	-0.182691	5.7425044	0.0165592	0.203082

Foxo3	1.2406022	3.21450928	7.4246342	0.0064337	0.1733728
Frmd7	-1.476607	0.16428633	15.67158	7.53E-05	0.0548057
Gabbr2	1.1286179	1.6062922	12.584457	0.000389	0.094991
Gabrq	1.2858098	0.72987385	4.8478426	0.0276807	0.2290527
Gan	1.5889967	2.49579066	10.088439	0.001492	0.1249683
Gatad2b	1.3144241	4.53513657	7.9443189	0.0048239	0.1662486
Gdap10	1.4044553	0.62507397	3.8519919	0.049687	0.2646947
Gfod1	1.1100188	3.13254029	15.840406	6.89E-05	0.0529114
Gjb2	-1.609652	0.38761485	6.5607252	0.0104254	0.1881628
Glg1	1.3600142	6.61513281	7.8161315	0.0051782	0.1703473
Gm15421	-1.162578	1.43945242	4.5461502	0.0329928	0.2398151
Gm16386	1.3854196	1.82012431	9.6638426	0.0018793	0.1343934
Gm16982	1.1225415	-0.3742334	5.628065	0.0176751	0.2049241
Gm19557	1.2576237	-0.5960542	8.3439726	0.0038697	0.1568302
Gm3500	-1.51304	0.66258206	9.0039925	0.0026939	0.1491236
Gpr161	1.5587569	4.68048305	9.6190551	0.0019257	0.1343934
Gpr26	1.395007	2.19845097	16.678569	4.43E-05	0.0428617
Gprin2	1.6724334	0.48270246	12.793622	0.0003478	0.0948142
Gramd1b	1.1103153	2.85005035	9.8628126	0.0016865	0.1309183
Grid2	1.1328186	-0.5129373	6.4273089	0.0112379	0.1881628
Hbb-bh1	-1.126332	7.1182712	7.0553853	0.0079027	0.1810886
Hdac4	1.11877	3.73058092	7.9069272	0.0049246	0.1672181
Hdx	1.3316674	1.60423134	6.3479378	0.0117518	0.1881628
Hhipl1	1.1790083	1.86648761	7.2147855	0.0072305	0.1778967
Hipk2	1.5142718	5.4287414	12.624408	0.0003807	0.094991
Hist1h1b	1.7573182	-0.7289896	5.0746768	0.0242778	0.2202976
Hist1h4f	1.5683893	-0.9653768	4.5837253	0.032277	0.2380886
Hist1h4k	1.1216218	0.05711985	4.8544875	0.0275743	0.228875
Hivep3	1.2059305	1.23781664	9.1849325	0.0024402	0.1441152
Hmbox1	1.3177839	4.70544322	8.0312028	0.0045978	0.1634366
Hspb1	-1.285022	-0.466154	8.3906174	0.0037716	0.156264
Igsf6	1.1587897	0.34377262	9.0287186	0.0026577	0.1491236
Igsf9b	1.6695684	2.20356771	11.804103	0.000591	0.1064152
Itgb8	1.4783435	3.68726197	5.5924601	0.0180379	0.2049241
Kcna3	2.0734388	2.51621881	8.7993448	0.0030134	0.1517881
Kcnb2	1.1017598	1.11638483	8.6326365	0.0033019	0.1538584
Kcnh5	1.1499019	-0.0321468	7.0103018	0.0081042	0.1810886
Kcnh7	1.1199508	2.70302243	8.0196187	0.0046273	0.1634366
Kcnj3	1.338798	0.00666991	5.8191424	0.0158527	0.2001393
Kcnn3	1.1627882	2.38943716	13.596267	0.0002266	0.0882681
Kcnq3	1.6428116	1.67485854	18.699055	1.53E-05	0.0331369

Kif13b	1.240713	3.92009847	8.1601827	0.004282	0.1610859
Klf12	1.330891	5.30633138	8.905197	0.0028436	0.1501806
Klf7	1.4532961	3.1165427	11.427627	0.0007236	0.1136381
Klhl11	1.5347816	2.93160066	7.3626893	0.0066591	0.1743286
Ksr2	1.2580217	1.54865137	11.055269	0.0008844	0.116398
Lars2	1.7019722	8.85413543	5.3527782	0.0206893	0.2111712
Lcor	1.9382317	3.01317476	9.6634634	0.0018797	0.1343934
Lmbrd2	1.5663355	3.4208898	6.7979695	0.0091262	0.1854598
Lmln	1.1344013	3.58050769	7.4548396	0.0063266	0.1729126
Lnpep	1.4653698	3.82301676	7.2106644	0.0072472	0.1778967
Lrp1b	1.1682923	0.78181917	5.3360813	0.0208884	0.2119152
Lrrc7	1.2017351	3.09987641	16.839048	4.07E-05	0.0428617
Lrrc8b	1.2954932	3.81398802	6.4865244	0.0108695	0.1881628
Lyst	1.3201609	4.56085623	8.6497288	0.0032711	0.1538584
Med12l	1.1037723	4.48182261	8.2804209	0.0040075	0.1586918
Megf9	1.2638878	4.94373456	6.1760378	0.0129492	0.1922092
Meig1	-1.229587	-0.8928936	7.3292583	0.0067841	0.1758148
Mgat5	1.6043859	3.69171551	10.947469	0.0009373	0.1176564
Mib1	1.2861947	5.18663926	6.3421818	0.01179	0.1881628
Mir124a-2	1.4697082	0.65358426	11.290546	0.000779	0.114946
Mir16-1	1.2392291	-0.501519	5.1205097	0.0236447	0.2185748
Mir703	-1.126489	3.9827852	5.7706762	0.0162958	0.2025247
Mrs2	1.1636046	4.02913111	7.4523771	0.0063353	0.1729126
N4bp2	1.4251061	5.92351548	6.420417	0.0112816	0.1881628
Nav3	1.4140944	3.82229061	9.6799366	0.0018629	0.1343934
Nbeal1	1.1911758	3.48803199	5.5702455	0.0182682	0.2049241
Ncam2	1.1153518	2.34943125	10.590525	0.0011367	0.1181129
Nos1	2.1748865	-0.3061876	14.520486	0.0001386	0.0766422
Nr2c2	1.4062175	5.72104369	9.072255	0.0025952	0.1491236
Nrip1	1.3620282	3.32762282	6.0670097	0.013773	0.1946245
Nwd1	1.4332099	2.04412908	11.13296	0.0008481	0.11543
Nyap2	1.494347	0.23455559	7.6209433	0.0057694	0.1704832
Onecut1	1.41704	-0.1493713	4.1618754	0.0413436	0.2526136
Pcdh11x	1.3646996	2.05128825	12.44176	0.0004198	0.0967031
Pcdha12	1.3737504	0.06441496	7.2589242	0.007055	0.1772718
Pcdha2	1.430592	-0.4392593	13.135868	0.0002897	0.0940169
Pcdha6	1.3699191	-0.4979672	5.673356	0.0172245	0.2038935
Pcdha7	1.2820794	-0.0001198	5.4843851	0.0191871	0.2082727
Pcdha9	1.3901522	-0.0763711	8.8936739	0.0028616	0.1501806
Pcdhac1	1.129472	0.50040776	5.7483547	0.0165041	0.2029242
Pcdhac2	1.527899	2.59108694	7.8343024	0.0051264	0.1703473

Pcdhb2	1.3610621	-0.4181272	5.5448225	0.0185354	0.206414
Pcdhga1	1.4632207	1.09204977	10.745195	0.0010455	0.1176564
Pcdhga11	1.1454788	3.89702419	8.9341526	0.0027989	0.1501806
Pcdhga2	1.1306602	2.01073866	7.0363875	0.007987	0.1810886
Pcdhga4	1.1593769	3.1157316	11.099299	0.0008636	0.11543
Pcdhga9	1.1331009	2.03225381	7.4443052	0.0063637	0.1729126
Pcdhgb1	1.4604063	1.56905457	10.716565	0.0010618	0.1176564
Pcdhgb4	1.1579528	1.67703648	8.446198	0.0036581	0.156264
Pcdhgc3	1.122205	6.21386955	10.221658	0.001388	0.1242324
Pdpr	1.559337	3.83323274	6.2949039	0.0121086	0.1895133
Perp	-1.456543	0.03474365	16.347552	5.27E-05	0.0428617
Plcxd3	2.01632	1.54386228	12.879131	0.0003323	0.0940169
Plxna4	1.1109436	6.11685725	16.92282	3.89E-05	0.0428617
Pou3f4	1.1385536	3.92336887	8.3909334	0.003771	0.156264
Ppm1l	1.2298516	4.83956573	10.731471	0.0010533	0.1176564
Ppp1r12b	1.4323803	3.83299853	11.8488	0.000577	0.1064152
Prdm10	1.2276944	4.29473514	7.3742104	0.0066166	0.1743286
Ptar1	1.7259241	2.78368152	12.260936	0.0004625	0.101465
Ptch1	1.1204745	5.45880645	6.4599501	0.0110333	0.1881628
Ptpn4	1.376638	3.85004703	5.9142641	0.0150188	0.1992042
Ptprt	1.4969137	1.62561562	10.159856	0.0014353	0.1242324
Ralgps1	1.4823965	4.69163288	13.79346	0.000204	0.0854517
Raver1-fdx1l	1.1319182	3.1625481	9.4947486	0.0020606	0.1359519
Rims1	1.1333199	2.16819871	10.693796	0.001075	0.1176564
Rmrp	2.1047888	2.03333224	4.7361075	0.0295358	0.2324191
Rn45s	2.2566418	12.2142734	6.3499444	0.0117385	0.1881628
Rps19-ps3	-1.176953	1.29062292	6.3497268	0.0117399	0.1881628
Rsg1	1.3129682	1.14431611	9.4077231	0.0021607	0.1378491
Sacs	2.4702528	1.85042552	17.040812	3.66E-05	0.0428617
Scarna9	1.7430059	-0.7067179	6.8277507	0.0089752	0.1854598
Sdk2	1.279603	3.11031449	9.3961089	0.0021745	0.1378491
Sfn	-1.209569	-0.0890799	7.2132257	0.0072368	0.1778967
Shc3	1.1066244	2.64713288	7.6483741	0.0056824	0.1703473
Siah3	1.5179208	3.36837413	14.610111	0.0001322	0.0761271
Sik2	1.414189	3.84001919	9.6373162	0.0019066	0.1343934
Six2	-1.903528	1.72213509	9.7273728	0.0018154	0.1343934
Slc16a14	1.3635437	2.71896346	12.887139	0.0003308	0.0940169
Slc30a3	-1.239792	-0.0305236	10.323052	0.0013138	0.1242324
Slc36a4	1.3519272	3.9152104	8.085089	0.0044631	0.1617702
Slc9a7	1.5006256	0.35680094	9.3093072	0.0022799	0.1399853
Snord17	1.3837372	-0.7280553	7.3960909	0.0065366	0.1741388

Snord64	1.593969	0.6851508	5.3289398	0.0209741	0.2120429
Sorl1	1.7159686	5.06352851	13.016992	0.0003087	0.0940169
St6gal2	1.5807506	2.56149411	7.479596	0.0062402	0.1729126
Sv2c	1.1058896	0.33771556	7.6155869	0.0057866	0.1704832
Syt14	1.5351642	2.67032908	8.0159495	0.0046367	0.1634366
Sytl5	1.5674831	0.71137704	11.687943	0.0006291	0.1064152
Tbx15	-1.378061	0.57319855	8.3627592	0.0038299	0.156264
Tekt4	-1.271828	-0.9022976	9.6777822	0.0018651	0.1343934
Tenm1	1.9046239	1.81874756	8.8821453	0.0028797	0.1501806
Tenm4	1.6114086	6.93802408	10.241648	0.001373	0.1242324
Tmem178b	1.877436	3.25289359	7.1347205	0.0075605	0.180484
Tmem245	1.1259243	3.99360364	7.2109243	0.0072461	0.1778967
Tnfrsf18	1.147034	0.94868407	7.3656513	0.0066482	0.1743286
Tnks	1.4134339	6.41418773	7.7911856	0.0052502	0.1703473
Tnr	2.2114871	1.2655418	15.347712	8.94E-05	0.0617956
Trip11	1.1957881	4.05696392	6.4962348	0.0108103	0.1881628
Twist1	-1.684039	1.89394871	11.517571	0.0006894	0.1107873
Ubald2	-1.294633	1.19059931	5.1020037	0.0238982	0.2190142
Uhmk1	1.194381	3.15236305	4.925755	0.0264592	0.2261091
Uprt	1.3940521	1.53375923	4.8705725	0.0273184	0.2281213
Vps13c	1.3416624	4.65713801	8.4406591	0.0036692	0.156264
Wdfy2	1.4321034	3.62963951	8.7548145	0.0030879	0.1517881
Xkr4	1.7598324	1.70799631	10.092395	0.0014888	0.1249683
Xkr7	1.5339784	2.86989879	10.090024	0.0014907	0.1249683
Xpo4	1.3539973	4.19689655	8.6784169	0.00322	0.1538584
Xrn1	1.5298196	4.7563551	8.2677347	0.0040356	0.1587374
Xylt1	1.4323649	3.456239	8.0032885	0.0046692	0.1634366
Zbed6	1.5445581	5.57454046	4.961358	0.0259198	0.225624
Zbtb20	1.4480771	2.06976695	18.634564	1.58E-05	0.0331369
Zbtb37	1.3482909	2.92165295	10.463019	0.0012179	0.1219646
Zbtb38	1.338636	2.227117	11.682571	0.0006309	0.1064152
Zc3h12b	1.3105452	1.16234478	8.1500622	0.004306	0.1610859
Zfp369	1.7621415	3.0493847	10.696683	0.0010733	0.1176564
Zfp536	1.6800347	4.28466461	18.020531	2.19E-05	0.0335573
Zfp618	1.4720347	4.67622833	9.9973506	0.0015677	0.1274245

7.2   Published Papers

Chapter 3

- Mattiske T, Moey C, Vissers LE, Thorne N, Georgeson P, Bakshi M and Shoubridge C (2017) An emerging female phenotype with loss of function mutations in the Aristaless- related homeodomain transcription factor, Human Mutation.

Chapter 4

- Mattiske, T, Lee, K., Gécz J and Shoubridge, C.   (2016) Embryonic forebrain transcriptome from mice modelling the two most frequent polyalanine expansion mutations in the intellectual disability ARX homeobox gene, *Human Molecular Genetics*.

Statement of Authorship

Title of Paper	An Emerging Female Phenotype with Loss-of-Function Mutations in the Aristaless-Related Homeodomain Transcription Factor ARX
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Published in Human Mutation, 15 February 2017 DOI: 10.1002/humu.23190

Principal Author

Name of Principal Author (Candidate)	Tessa Mattiske
Contribution to the Paper	Screening of addition family members and performed X-inactivation studies on females. Complied and review reported phenotypes. Convinced manuscript idea and main contributor to manuscript preparation.
Overall percentage (%)	70
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date 17/03/17

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Ching Moey
Contribution to the Paper	Screening of proband in the reported family.
Signature	Date 22/3/17

Name of Co-Author	Lisenka Vissers
Contribution to the Paper	Analysis of whole exon sequencing coverage of ARX
Signature	Date 25/3/17

Name of Co-Author	Natalie Thorne
Contribution to the Paper	Analysis of whole exon sequencing coverage of ARX
Signature	Date 28/3/17

Name of Co-Author	Peter Georgeson
Contribution to the Paper	Analysis of whole exon sequencing coverage of ARX
Signature	Date 22/3/17

Name of Co-Author	Madhura Bakshi
Contribution to the Paper	Provided clinical information and contributor to manuscript preparation
Signature	Date 22/3/17

Name of Co-Author	Cheryl Shoubridge
Contribution to the Paper	Supervised the project and provided critical feedback and revision of the manuscript
Signature	Date 17/3/17

An Emerging Female Phenotype with Loss-of-Function Mutations in the *Aristaless*-Related Homeodomain Transcription Factor *ARX*

Tessa Mattiske,[1,2] Ching Moey,[1,2] Lisenka E. Vissers,[3] Natalie Thorne,[4,5,6] Peter Georgeson,[6,7] Madhura Bakshi,[8] and Cheryl Shoubridge[1,2]*

[1]*Department of Paediatrics, School of Medicine, University of Adelaide, Adelaide, SA, Australia; [2]Robinson Research Institute, University of Adelaide, Adelaide, SA, Australia; [3]Department of Human Genetics, Donders Institute for Brain, Cognition, and Behaviour, Radboud University Medical Center, Nijmegen, the Netherlands; [4]Murdoch Children's Research Institute, Melbourne, Australia; [5]University of Melbourne, Melbourne, Australia; [6]Melbourne Genomics Health Alliance, Melbourne, Australia; [7]Victorian Life Sciences Computation Initiative, The University of Melbourne, Melbourne, Australia; [8]Department of Clinical Genetics, Liverpool Hospital, Liverpool, NSW, Australia*

ABSTRACT: The devastating clinical presentation of X-linked lissencephaly with abnormal genitalia (XLAG) is invariably caused by loss-of-function mutations in the *Aristaless*-related homeobox (*ARX*) gene. Mutations in this X-chromosome gene contribute to intellectual disability (ID) with co-morbidities including seizures and movement disorders such as dystonia in affected males. The detection of affected females with mutations in *ARX* is increasing. We present a family with multiple affected individuals, including two females. Two male siblings presenting with XLAG were deceased prior to full-term gestation or within the first few weeks of life. Of the two female siblings, one presented with behavioral disturbances, mild ID, a seizure disorder, and complete agenesis of the corpus callosum (ACC), similar to the mother's phenotype. A novel insertion mutation in Exon 2 of *ARX* was identified, c.982delCinsTTT predicted to cause a frameshift at p.(Q328Ffs*37). Our finding is consistent with loss-of-function mutations in *ARX* causing XLAG in hemizygous males and extends the findings of ID and seizures in heterozygous females. We review the reported phenotypes of females with mutations in *ARX* and highlight the importance of screening *ARX* in male and female patients with ID, seizures, and in particular with complete ACC.
Hum Mutat 00:1–8, 2017. © 2017 Wiley Periodicals, Inc.

KEY WORDS: X-linked lissencephaly-2; X-linked lissencephaly; ARX; *Aristaless*-related homeobox; intellectual disability; seizure; LISX2; XLAG

Introduction

The *Aristaless*-related homeobox gene (*ARX*; NM_139058.2; MIM# 300382) [Shoubridge, et al., 2010] is critical for early development and formation of a normal brain [Kitamura, et al., 2002], [Ohira, et al., 2002], [Stromme, et al., 2002]. This paired-type homeodomain transcription factor plays a vital role in telencephalic development specifically in tangential migration and differentiation of GABAergic and cholinergic neurons [Kitamura, et al., 2002], [Colombo, et al., 2007], [Friocourt, et al., 2008], [Colasante, et al., 2009], [Lee, et al., 2014]. Mutations in *ARX* result in a range of phenotypes with intellectual disability (ID) as a consistent feature. Mutations leading to loss-of-function of the ARX protein typically lead to brain malformation phenotypes, including X-linked lissencephaly with abnormal genitalia (XLAG, also known as LISX2) (MIM# 300215) [Kitamura, et al., 2002], [Shoubridge, et al., 2010]. XLAG is a developmental disorder characterized by structural brain anomalies leading to agyria (absent cerebral grooves brain) or pachygygria (reduced folds or grooves) and agenesis of the corpus callosum (ACC). In addition, patients commonly have early-onset intractable seizures, severe psychomotor retardation, and ambiguous genitalia [Dobyns, et al., 1999; Kitamura, et al., 2002]. Males are severely affected and often die within the first days or months of life.

As the *ARX* gene is on the X-chromosome, mutations in this gene typically result in families with affected males across multiple generations transmitted via (usually) asymptomatic carrier females. However, there is an increasing prevalence of reported mutations in *ARX* that result in the severe phenotypic outcomes of XLAG in male patients that, with variable penetrance, affect females within the families resulting in a generally milder phenotype than affected males [Eksioglu, et al., 2011; Kato, et al., 2004; Kitamura, et al., 2002; Marsh, et al., 2009; Scheffer, et al., 2002; Stromme, et al., 2002]. Here, we report a novel mutation in *ARX* in a family ascertained by a female proband displaying a phenotype of mild learning disabilities, seizure disorder and ACC. As part of this work, we have reviewed reported phenotype of females with mutations in *ARX* and recommend screening of the *ARX* gene in female patients with suitable ID, seizure phenotypes, and corpus callosum agenesis, particularly if there is evidence of X-linkage and no surviving males. *ARX* adds to a growing list of genes on the X chromosome including genes such as *USP9X*, *PHF6*, and *IQSEC2* that have phenotypic effects

in males and females that are distinct depending on the functional severity of the mutation [Zweier, et al., 2013; Reijnders, et al., 2016; Zarem et al., 2016].

Materials and Methods

Clinical Description of Patient and Family

The proband presented to a Genetics Clinic at 10 years of age with learning difficulties, mild ID, and seizures. The proband was born at term weighing 3.2 kg, with no admission into the neonatal intensive care unit nor special-care nursery required. She was reported as sitting with support at 9 months of age with single words spoken at 9–10 months of age and walking at 22 months of age. Seizure onset was around 5 years of age with the first seizure classified as a prolonged generalized tonic-clonic which required intubation. At this time, she was assessed and was able to draw a triangle with help, able to hop, dress herself, and talk in simple sentences, knowing a few colors and numbers. No unusual grasping is reported. She was evaluated using the WISC-IV Australian at the age of 8 years and 10 months. She scored in the "Extremely Low" range for verbal comprehension, perceptual reasoning index, working memory index, processing speed index with a resulting full scale IQ in the "Extremely Low" range. She was assessed as operating in the mild range of ID. She subsequently had a number of complex partial seizures, which were reasonably well controlled on Tegretol. Brain MRI revealed complete ACC with mild dilated ventricles and colpocephaly. No lissencephaly, dysmorphic features, or behavior problems were reported.

The mother of the proband was aged 46 years old at the time of this report and presented with mild ID, seizures, and mental health challenges. After admission to the public health hospital she was diagnosed with borderline personality disorder and major depressive disorder, however no neuropsychiatric assessment is currently available. She was non dysmorphic. History revealed that her first seizure was around 7–8 years of age and classified as complex partial seizures. Brain MRI scan done at the age of 38 showed complete ACC, with no gray matter heterotopia. Small white matter lesions were noted which comprised of tiny fluid-attenuated inversion recovery hyperintensities involving the left centrum semiovale bilaterally and in the frontal region. The treating neurologist at the time felt these were specific white matter hyperintensities however the actual films are no longer available for review. In addition, EEG showed intermittent spike and wave discharge maximal over left hemispheres, which were at frequency of 2.5–3.5 Hz.

Two male offspring of II-4 were affected and died in early infancy, or were terminated during pregnancy. A maternal half-brother (III-1) of the proband was born at 36 weeks of gestation with an onset of seizures 20 min after birth. III-1 died at 26 days of age. Investigation of brain morphology identified ACC, lissencephaly, gray matter heterotopias, and bilateral optic nerve hypoplasia. Genitalia were ambiguous with labioscrotal folds, bilateral gonads and microphallus. Pelvis ultrasonography revealed the presence of a bicornuate uterus, while a male urethra was confirmed with a genitogram. Karyotype analysis showed a normal male chromosomal constitution (46,XY). Facial dysmorphic features included a large head, large anterior fontanels, low set ears, and abnormal nails. III-3 eventuated with a medical termination of pregnancy due to identification at 18 weeks gestation via ultrasonography of ventriculomegaly and abnormal genitalia. Brain malformations were confirmed at post mortem showing lissencephaly, absent corpus callosum, wide-open sylvian fissure, and dilated ventricles. Facial dysmorphic features included flattened facies, mild macrocephaly, and wide open sutures. Abnormal genitalia consisted of rudimentary genitalia with a small phallus. Karyotype analysis established the presence of male chromosomes.

Another pregnancy, III-5 was terminated when chorionic villus sampling revealed 45XO after ultrasound showed fetal hydrops. The remaining sibling (III-2) is a healthy female with normal intelligence. A maternal uncle's (II-1) death at 1 month of age was attributed to sudden infant death syndrome. The maternal grandmother (I-2) is reported to be healthy with no seizures.

Molecular Analysis of *ARX* Gene

The screening protocols were approved by the appropriate institution review boards and informed consent was obtained from the parents of patients. Genomic DNA from the proband was extracted from whole blood using standard techniques. Each of the five exons of *ARX* was amplified by PCR using primers designed to amplify coding and flanking non-coding sequence. The exception to this was exon 2, for which four overlapping amplicons were used to achieve robust amplification of GC-rich regions coding for three polyalanine tracts. The PCR conditions and primer sequences are described in detail previously [Tan, et al., 2013]. Sequencing reactions were performed using ABI Big Dye terminator chemistry version 3.1 and purified products subjected to an automated capillary sequencing on ABI 3100 sequencer (Applied Biosystems, Foster City, CA) and sequence was compared to the *ARX* reference sequence (NM_139058.2) using SeqMan module of the Lasergene DNA and protein analysis software package (DNAStar, Inc., Madison, WI).

Results

Molecular Analysis of the *ARX* Variant

The diagnoses of ID and seizures in the female proband in conjunction with XLAG in her male siblings, from two different fathers, prompted evaluation of the *ARX* gene. Sequence analysis demonstrated a novel indel mutation, c.982delCinsTTT in exon 2 of the *ARX* gene, that is predicted to result in the creation of a premature stop codon, p.(Q328Ffs*37) (LOVD ID 0000128956) (http://www.lovd.nl/ARX). The child's mother presented with a similar phenotype and was confirmed to also be a symptomatic carrier of this novel *ARX* mutation. The father (II-5) and unaffected sister (III-2) do not carry the mutation (Fig. 1). The *ARX* mutation was confirmed in III-4 in genomic DNA extracted from amniocytes. The amino acid affected by the mutation p.328Q is located at the start of the region containing the second nuclear localization signal (NLS2) preceding the conserved paired-type homeodomain. The predicted stop codon arising from this variant occurs within 29 nucleotides of the exon 3–4 junction, and is likely to escape nonsense-mediated decay (NMD). Due to restricted levels of *ARX* expression in the patient derived materials we are unable to confirm if this truncated protein is produced. Despite this, even if the predicted C-terminal truncated protein was produced and subsisted at appreciable levels, the severe XLAG phenotype in affected male patients is expected with this variant resulting in a nonsense peptide being transcribed after residue p.328 and complete loss of the paired-type homeodomain residues (Fig. 1C).

A methylation-specific PCR at the human FraxA and Androgen Receptor genes was performed on genomic DNA from blood as previously described [Plenge, et al., 1997]. X-inactivation studies
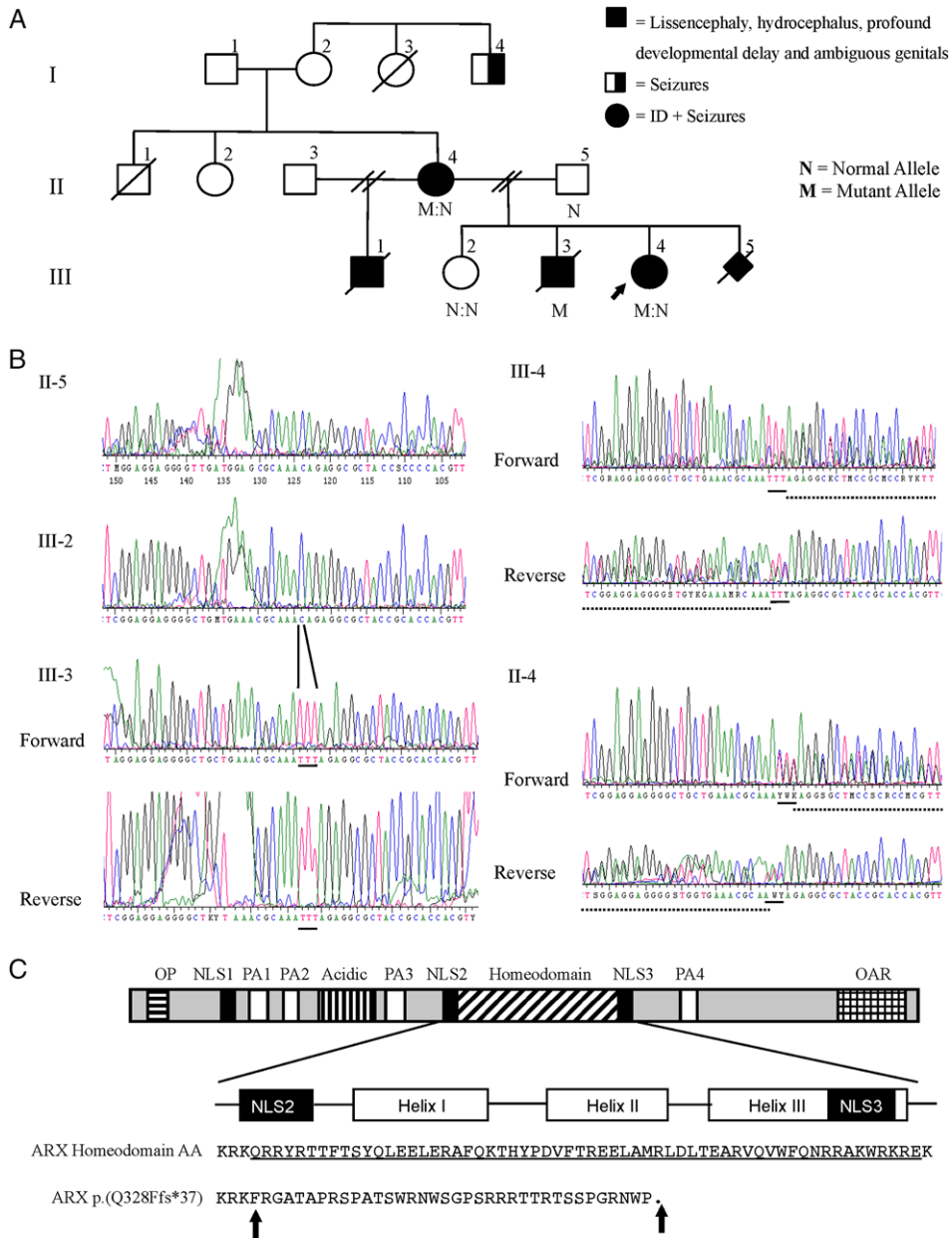
Figure 1. Identification of a c.982delCinsTTT mutation resulting in (p.(Q328Ffs*37) in ARX. **A**: Pedigree of immediate family members tested. Open symbols represent normal individuals, filled black circles represent females with moderate intellectual disability and seizures, filled squares represent males with lissencephaly, hydrocephalus, profound developmental delay, and ambiguous genitalia. Individual generations are numbered with Roman numerals on the left hand side of the pedigree. Individuals which were sequence confirmed to carry either the normal (N) or mutant (M) are shown. **B**: DNA sequence electropherograms for the chrX: 25,031,130 (GRCh37/hg19 assembly) deletion of a C and insertion of TTT mutation in exon 2 of 5 of *ARX* reported in this study. A normal sequence was confirmed in unaffected father (II-5) and sister (III-2) with normal sequence shown. The mutation in the heterozygous state is shown in both the affected proband (III-4) and mother (II-4). The dotted line highlights the disrupted heterozygous trace present in females caused by the insertion (solid black underline). Additional mutation sequence change seen in the hemizygous state of an affected brother (III-3). **C**: Predicted functional consequence of a novel nonsense mutation in ARX. Schematic representation of the human ARX protein (top panel). Human ARX domains and regions are indicated above the schematic. Known functional domains are highlighted, octapeptide (OP) as horizontally hatched rectangle, nuclear localization sequences (NLS) as three black rectangles, polyalanine tracts (PA) as four white rectangles, acidic domain as vertically hatched rectangle, homeodomain as crosshatched, and aristaless domain (OAR) as hatched. A schematic representation of the homeodomain and the flaking NLS domains (middle panel) with the amino acid sequence below (homeodomain sequence underlined). The amino acid change indicated by the first black arrow resulting in a nonsense peptide and a stop codon indicated by the second black arrow.

Table 1. Clinical Summary of Females with *ARX* mutations

	Familial	This report	De novo	Total
Females	25	2	4	31
Females with ID, with and without seizures	11	2	4	17
ID or DD	11	2	4	17
Seizures	5	2	4	11
Other clinical features	Number (symptomatic:non-symptomatic)			
MRI reported	10 (5:5)	2 (2:0)	3	15
Brain malformation	9 (4:5)	2 (2:0)	3	14 (9:5)
ACC	8 (4:4)	2 (2:0)	1	11 (7:4)
Other	1 (0:1)	0	2	3 (3:0)
Movement disorder	4 (3:1)	0	4	8 (7:1)
Psychiatric features	4 (3:1)	1 (1:0)	0	5 (4:1)
Behavior disturbance	2 (2:0)	0	0	2 (2:0)

for both the proband (III-3) and the healthy sister (III-2) detected no significant deviation.

As part of initial investigations, both II-4 and III-3 were identified to carry a duplication on chromosome 5 at 5p15.5 (Chr5:10,907,608-11,363,857). This duplication contains only part of one gene, *Catenin Delta 2* (*CTNND2*). This duplication was classified as a variant of unknown significance. This duplication was later also identified in the female sibling III-2, who is healthy and has no learning issues, indicating this copy number variant is unlikely to contribute to the phenotype of the proband, her mother or affected brothers.

Discussion

Here, we report a family with a novel truncating mutation (c.982delCinsTTT/p.(Q328Ffs*37)) in *ARX*. The mutation is predicted to yield a non-functional protein product after p.328 due to the nonsense peptide prior to truncation of the protein at amino acid 364. This variant was not found in either the ExAC or 1000 genome project databases. Although this mutation may escape NMD, the resulting protein will have functional loss of the homeodomain and Aristaless domains. The catastrophic phenotype of XLAG reported in two males in this family is consistent with the predicted disruption of the ARX homeodomain function. The female proband and mother have a milder phenotype consisting of ID, seizures, and ACC. The investigation of *ARX* as a cause of the phenotype in the female proband was due largely to the distinctive phenotypic presentation and early deaths of the affected male siblings.

Our report underscores that a carrier female phenotype is likely to be under ascertained for *ARX*. This is supported by a recent examination of heterozygous females from families identified with *ARX* mutations [Marsh, et al., 2009] and examples of gender bias (92% male:8% female) in a recent cohort of patients referred for molecular analysis of *ARX* [Marques, et al., 2015]. Under ascertainment may be occurring due to several contributing factors. Patients with phenotypes such as ID and infantile spasms have been typically screened for mutations in known genes such as *STXBP1*, *CDKL5*, *KCNQ2*, *GRIN2A*, *MAGI2*, and *ARX*. However, in the case of *ARX,* screening is commonly only considered in affected males. Moreover, as the majority of all mutations reported to date in *ARX* lead to expansion of the first or second polyalanine tracts, both located in exon 2, *ARX* mutation analysis is routinely limited to screening exon 2 and often using size variant analysis approaches [Marques, et al., 2015]. The expanding use of next generation sequencing approaches on cohorts of individuals with ID and or seizure phenotypes are likely to provide a platform to potentially overcome some of these biases. However, even these types of approaches have constraints that need to be considered, particularly

for genes with high GC content or near regions of low gene density, such as *ARX*. For example, sequence coverage across the *ARX* gene in 50 representative whole exome sequencing (WES) experiments undertaken at the Radboud University Medical Centre showed the median coverage for *ARX* was 40-fold less than the coverage of all ID genes; with the median percentages of *ARX* bases covered at $20\times$ only being 73% compared with 97% for all ID genes (Supp. Fig. S1). However, experience at this, and other centers, indicates that the coverage differences although lower for *ARX* generally, may also depend upon the region of the gene being considered (Supp. Fig. S2). Exome sequencing using benchtop ion proton machines also result in poor coverage of the *ARX* gene, with mean coverage reported at 43% [Lacoste, et al., 2016]. It remains to be determined how the increasing use of whole genome sequencing approaches as well as improvements to WES technologies and analysis pipelines address some of these coverage issues.

To date there have been both familial and de novo cases of affected females due to *ARX* mutations [Bettella, et al., 2013; Eksioglu, et al., 2011; Kato, et al., 2004; Kwong, et al., 2015; Marsh, et al., 2009; Scheffer, et al., 2002; Wallerstein, et al., 2008] (Table 1). Penetrance is variable both within and across families, with 55% of carrier females in these families presenting with a phenotype of ID to some degree with and without seizures (Table 1). ID and/or developmental delay are prominent features in affected females across all families. A seizure phenotype was reported in 64% of the females with ID which equates to ∼35% of all females in these families (Table 1).

Similar to the novel case we report here, affected females have been reported in families in which the male proband presents with severe XLAG or seizure phenotypes [Eksioglu, et al., 2011; Kato, et al., 2004; Marsh, et al., 2009; Scheffer, et al., 2002]. In these familial cases ascertained by the male proband, the mutations are either missense mutations of residues in the homeodomain or nonsense/deletion mutations resulting in a loss-of-function of the ARX homeodomain and/or aristaless domain activity (Table 2). Similarly, a smaller number of de novo cases also result in truncation and loss of *ARX* function [Bettella, et al., 2013; Kwong, et al., 2015; Marsh, et al., 2009; Wallerstein, et al., 2008]. Across these cases, there is a consistent phenotype of ID and/or developmental delay, infantile seizures, and hypotonia/dystonia/ataxia (Table 3). The type and location of mutations in affected females is restricted compared with those contributing to affected males (Fig. 2).

Brain MRI imaging is reported in approximately 35% of all females in these cases/families, including the two females from this current report. Interestingly, of those individuals screened, 73% had features consistent with ACC but only 64% of these patients were those classed as symptomatic based on ID and/or developmental delay and seizure phenotypes. Hence, there are a number of asymptomatic carrier females that do not display these key clinical features

Table 2.   Clinical Features of Females in Familial Cases of *ARX* Mutations

Mutation cDNA/protein AA	Exon/domain	Male phenotype	Relationship to male proband	ID/DD	Seizures	Brain malformation	Additional features	References
Truncation mutations								
Exon1_2del	1, 2	XLAG	Mother	N	Onset 12y (GTCS)			Kitamura et al. (2002), Marsh et al. (2009)
Exon2_5del p.G66_C562del	2–5 HD + OAR	XLAG	Mother	N				Kato et al. (2004)
c.232G>T p.E78X	2 HD + OAR	XLAG	Sister	ID		ACC + CVH	ADD	Kato et al. (2004)
			Mother	N		ACC		
			Sister	N		ACC	Duane anomaly	Kato et al. (2004)
			Mother	N		ACC-p		
c.617delG p.G206Afs* 119	2 HD + OAR	XLAG-HYD	Aunt	Mod ID +DD	Onset 1y (GTCS)	ACC		Kato et al. (2004)
			Mother	N				
c.982delCinsTTT p.Q328Fs*37	2 HD + OAR	XLAG	Sister	ID	Unilateral ~ 9 weeks	ACC-p	Hypotonia	This report
			Proband	Mild ID	Onset 5y	ACC		
c.1471_1472insC p.L491Pfs*41	5 OAR	OS, AG, ID	Mother	Mild ID	Onset 7–8y	ACC	Depression	Eksioglu, et al. (2011)
			Mother	ID			Anxiety, depression, and personality disorder	
			Mat aunt 1	N			Depression, bicornuate uterus	
			Mat aunt 2	ID			Schizophrenia. Small genitals and small bladder with pockets	
			Mat half aunt 1	ID +DD	Generalized		Enlarged clitoris	
			Mat half aunt 2	N				
			Mat grandmother	MD			Anxiety and depression	
Missense mutations in the homeodomain								
c.998C>A p.T333N	2 HD	ACC/AG	Mother	N		ACC-p		Kato et al. (2004)
			Cousin	Mod ID				
			Aunt	N				
c.1058C>T p.P353L	2 HD	XMESID	Aunt	Sev ID	Onset 3mo	ACC-p	Spasticity, scoliosis, and contractures	Stromme et al. (2002), Scheffer et al. (2002)
			Mat grandmother	N		Small vessel ischemic changes	Progressive spastic ataxia	
			Mat Aunt	ID			Subtle hyperreflexia	
			Mother	N				
c.1135C>A p.R379S	4 HD	ISSX	Mother	N				Marsh et al. (2009)
			Aunt	N				
			Cousin	ID	Onset 5y (absence)	N	PDD	

ACC, agenesis of the corpus callosum; ACC-p, partial ACC; /, not reported; ADD, attention deficit disorder; AG, ambiguous genitalia; CVH, cerebellar vermis hypoplasia; DD, developmental delay; HD, homeodomain; ID, intellectual disability; ISSX, X-linked infantile spasm syndrome; Mat, maternal; MD, mild delay; mo, months; Mod ID, moderate ID; N, normal; OAR, aristaless; OS, Ohtahara Syndrome; PDD, pervasive developmental disorder; Sev ID, severe ID; y, years; XMESID, X-linked myoclonic epilepsy with generalized spasticity and ID.
Nucleotide numbering reflects cDNA numbering with +1 corresponding to the A of the ATG translation initiation codon in the reference sequence for *ARX* gene [GenBank: NM_139058.2].

Table 3. Clinical Features of Females with De Novo Mutations in *ARX*

Mutation cDNA/ protein AA	Exon/Domain	Mutation type	Development	ID	Seizures	Brain malformation	Muscle phenotype	Additional features	References
c.869C>A p.S290X	2 HD + OAR	NS	Sev DD		IEE onset 1m	microcephaly	Dystonia	UMN Signs, CVI	Kwong et al. (2015)
c.1459delA p.T487Qfs*5	5 OAR	Del	GPD	ID	EIEE	N	Significantly ataxic	Divergent Strabismus. Congenital hip dislocation	Bettella et al. (2013)
c.1465delG p.A489Pfs*3	5 OAR	Del	DD	ID	ISSX (WS)	Cysts	Hypotonia + torticollis	Poor visual tracking. Hydronephrosis. Plagiocephaly + small bilateral epicanthal fold with mildly low set ear with slightly over folded helices.	Wallerstein et al. (2008)
Inversion	Interruption of whole gene	Inversion	DD	ID	ISS onset in utero	ACC + HYD	Mild truncal hypotonia	Prominent forehead with mild frontal bossing, wide nasal bridge, slightly up slant palpebral fissures, mildly downturned corners of the mouth, slightly low-set ears with normal ear shape, and high palate	Marsh et al. (2009)

ACC, agenesis of the corpus callosum; CVI, cortical visual impairment; DD, developmental delay; Del, deletion; EIEE, early infantile epileptic encephalopathy; GPD, global psychomotor delay; HD, homeodomain; HYD, hydranencephaly; ID, intellectual disability; ISS, infantile spasm syndrome; ISSX, X-linked infantile spasm syndrome; N, normal; NS, nonsense; OAR, aristaless domain; Sev Delay, Severe DD; UMN, upper motor neuron syndrome; WS, West Syndrome. Nucleotide numbering reflects cDNA numbering with +1 corresponding to the A of the ATG translation initiation codon in the reference sequence for *ARX* gene [GenBank: NM_139058.2].

Figure 2. Identified ARX mutations in females and males leading to a range of phenotypes. **A**: *ARX* exon structure in relation to the ARX protein. **B**: Overview of ARX including the homeodomain (crosshatched) and aristaless domain (hatched) and the location of reported mutations according to their relative position at the protein level in both males and females. Differences in mutation type are indicated by a change in color, whereas missense mutation shown in gray and all other nonsense, insertion, or deletion mutation in black. Phenotype severity is indicated on the *y*-axis with unaffected carrier females (below clinical threshold) shown below the dotted line.

but still have the brain abnormality of ACC. There is also variable penetrance of both psychiatric and behavioral phenotypes across the symptomatic and asymptomatic females within these families. Movement disorders including ataxia and hypotonia are noted, particularly prevalent in the cases of de novo mutations (Table 3).

Given that the *ARX* gene is located on the X chromosome and is subject to X chromosome inactivation, the contribution of skewed X-inactivation (80:20) to the phenotype in heterozygous females is always a consideration. However, the female proband in this current report detected no skewing of X-inactivation in samples collected from blood. X-inactivation in previous studies has been inconclusive with little correlation of the affected status in heterozygous females to preferential inactivation of the normal *ARX* allele (measured in blood) [Marsh, et al., 2009]. Indeed, skewing of X-inactivation that is not consistent with disease severity has been reported for other X-linked genes that were originally thought to affect males but have had affected female cases described, including mutations in genes such as *USP9X*, *MTM1*, *SLC9A6*, and *MED12* [Prontera, et al., 2016; Reijnders, et al., 2016; Savarese, et al., 2016; Sinajon, et al., 2016]. This highlights the complexity of X-linked inheritance. Recently, the variability and complex X-inactivation status within the brain has been demonstrated in an *Arx* knockout mouse model [Marsh, et al., 2016]. Heterozygous female mice only have one functional copy of *Arx* and consistent with human data, female mice showed great variation of phenotype manifestations. These mice display striations of radially oriented streams of *Arx* positive neurons within and emerging from the ganglionic eminence ventricular zone, which were noted to vary dramatically between embryos. This is consistent with the site of random X-inactivation suggested to occur in the ventral forebrain, followed by clonal proliferation of Arx positive or negative cells migrate radially away during early stages of development. Even this incomplete loss of *Arx* in female mice was detrimental and resulted in a change in the profile of interneurons in the adult female mice, consistent with a loss to a greater extent in hemizygous male mice.

We have identified a novel truncating mutation (c.982delCins TTT/p.(Q328Ffs*37)) in *ARX* in a family ascertained by a female proband displaying a phenotype of mild ID, seizure disorder and ACC, in conjunction with a phenotype of XLAG in her deceased male siblings. Review of the phenotypes of affected females with

published mutations in *ARX* indicates that screening of the *ARX* gene in female patients with ID, seizure phenotypes, and ACC, particularly if there is evidence of X-linkage and no surviving males is warranted. Moreover, recent de novo mutations reported in females recommends scrutiny of *ARX* in all cases with suitable phenotypic presentation with and without other indications of X-linked inheritance. The emerging appreciation of phenotypic consequences in females with loss-of-function mutations in *ARX* will be important in counseling of affected families.

References

Bettella E, Di Rosa G, Polli R, Leonardi E, Tortorella G, Sartori S, Murgia A. 2013. Early-onset epileptic encephalopathy in a girl carrying a truncating mutation of the ARX gene: Rethinking the ARX phenotype in females. Clin Genet 841:82–85.

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

Colombo E, Collombat P, Colasante G, Bianchi M, Long J, Mansouri A, Rubenstein JL, Broccoli V. 2007. Inactivation of Arx, the murine ortholog of the X-linked lissencephaly with ambiguous genitalia gene, leads to severe disorganization of the ventral telencephalon with impaired neuronal migration and differentiation. J Neurosci 27:4786–4798.

Dobyns WB, Berry-Kravis E, Havernick NJ, Holden KR, Viskochil D. 1999. X-linked lissencephaly with absent corpus callosum and ambiguous genitalia. Am J Med Genet 86:331–337.

Eksioglu YZ, Pong AW, Takeoka M. 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 52:984–992.

Friocourt G, Kanatani S, Tabata H, Yozu M, Takahashi T, Antypa M, Raguenes O, Chelly J, Ferec C, Nakajima K, Parnavelas JG. 2008. Cell-autonomous roles of ARX in cell proliferation and neuronal migration during corticogenesis. J Neurosci 28:5794–5805.

Kato M, Das S, Petras K, Kitamura K, Morohashi K, Abuelo DN, Barr M, Bonneau D, Brady AF, Carpenter NJ, Cipero KL, Frisone F, et al. 2004. Mutations of ARX are associated with striking pleiotropy and consistent genotype-phenotype correlation. Hum Mutat 23:147–159.

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, et al. 2002. Mutation of ARX causes abnormal development of forebrain and testes in mice and X-linked lissencephaly with abnormal genitalia in humans. Nat Genet 32:359–369.

Kwong AK, Ho AC, Fung CW, Wong VC. 2015. Analysis of mutations in 7 genes associated with neuronal excitability and synaptic transmission in a cohort of children with non-syndromic infantile epileptic encephalopathy. PLoS One 10:e0126446.

Lacoste C, Desvignes JP, Salgado D, Pecheux C, Villard L, Bartoli M, Beroud C, Levy N, Badens C, Krahn M. 2016. Coverage analysis of lists of genes involved in heterogeneous genetic diseases following benchtop exome sequencing using the ion proton. J Genet 95:203–208.

Lee K, Mattiske T, Kitamura K, Gecz J, Shoubridge C. 2014. Reduced polyalanine-expanded Arx mutant protein in developing mouse subpallium alters Lmo1 transcriptional regulation. Hum Mol Genet 23:1084–1094.

Marques I, Sa MJ, Soares G, Mota Mdo C, Pinheiro C, Aguiar L, Amado M, Soares C, Calado A, Dias P, Sousa AB, Fortuna AM, et al. 2015. Unraveling the pathogenesis of ARX polyalanine tract variants using a clinical and molecular interfacing approach. Mol Genet Genomic Med 3:203–214.

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. 2009. Targeted loss of Arx results in a developmental epilepsy mouse model and recapitulates the human phenotype in heterozygous females. Brain 132:1563–1576.

Marsh ED, Nasrallah MP, Walsh C, Murray KA, Nicole Sunnen C, McCoy A, Golden JA. 2016. Developmental interneuron subtype deficits after targeted loss of Arx. BMC Neurosci 17:35.

Ohira R, Zhang YH, Guo W, Dipple K, Shih SL, Doerr J, Huang BL, Fu LJ, Abu-Khalil A, Geschwind D, McCabe ERB. 2002. Human ARX gene: Genomic characterization and expression. Mol Genet Metab 77:179–188.

Plenge RM, Hendrich BD, Schwartz C, Arena JF, Naumova A, Sapienza C, Winter RM, Willard HF. 1997. A promoter mutation in the XIST gene in two unrelated families with skewed X-chromosome inactivation. Nat Genet 17:353–356.

Prontera P, Ottaviani V, Rogaia D, Isidori I, Mencarelli A, Malerba N, Cocciadiferro D, Rolph P, Stangoni G, Vulto-van Silfhout A, Merla G. 2016. A novel MED12 mutation: Evidence for a fourth phenotype. Am J Med Genet A 170:2377–2382.

Reijnders MR, Zachariadis V, Latour B, Jolly L, Mancini GM, Pfundt R, Wu KM, van Ravenswaaij-Arts CM, Veenstra-Knol HE, Anderlid BM, Wood SA, Cheung SW. 2016. De novo loss-of-function mutations in USP9X cause a female-specific recognizable syndrome with developmental delay and congenital malformations. Am J Hum Genet 98:373–381.

Savarese M, Musumeci O, Giugliano T, Rubegni A, Fiorillo C, Fattori F, Torella A, Battini R, Rodolico C, Pugliese A, Piluso G, Maggi L, et al. 2016. Novel findings associated with MTM1 suggest a higher number of female symptomatic carriers. Neuromuscul Disord 26:292–299.

Scheffer IE, Wallace RH, Phillips FL, Hewson P, Reardon K, Parasivam G, Stromme P, Berkovic SF, Gecz J, Mulley JC. 2002. X-linked myoclonic epilepsy with spasticity and intellectual disability: Mutation in the homeobox gene ARX. Neurology 59:348–356.

Shoubridge C, Fullston T, Gecz J. 2010. ARX spectrum disorders: Making inroads into the molecular pathology. Hum Mutat 31:889–900.

Sinajon P, Verbaan D, So J. 2016. The expanding phenotypic spectrum of female SLC9A6 mutation carriers: A case series and review of the literature. Hum Genet 135:841–850.

Stromme P, Mangelsdorf ME, Scheffer IE, Gecz J. 2002. Infantile spasms, dystonia, and other X-linked phenotypes caused by mutations in Aristaless related homeobox gene, ARX. Brain Dev 24:266–268.

Tan MH, Gecz J, Shoubridge C. 2013. PCR amplification and sequence analysis of GC-rich sequences: Aristaless-related homeobox example. Methods Mol Biol 1017:105–120.

Wallerstein R, Sugalski R, Cohn L, Jawetz R, Friez M. 2008. Expansion of the ARX spectrum. Clin Neurol Neurosurg 110:631–634.

Zarem A, Haginoya K, Lev D, Blumkin L, Kivity S, Linder I, Shoubridge C, Palmer EE, Field M, Boyle J, Chitayat D, Gaillard WD, et al. 2016. The molecular and phenotypic spectrum of IQSEC2 related epilepsy. Epilepsia 57:1858–1869.

Zweier C, Kraus C, Brueton L, Cole T, Degenhardt F, Engels H, Gillessen-Kaesbach G, Graul-Neumann L, Horn D, Hoyer J, Just W, Rauch A, et al. 2013. A new face of Borjeson-Forssman-Lehmann syndrome? De novo mutations in PHF6 in seven females with a distinct phenotype. J Med Genet 50:838–847.

Statement of Authorship

Title of Paper	Embryonic forebrain transcriptome of mice with polyalanine expansion mutations in the ARX homeobox gene
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Published in Human Mutations, 25 October 2016 DOI: 10.1093/hmg/ddw360

Principal Author

Name of Principal Author (Candidate)	Tessa Mattiske
Contribution to the Paper	All experimental work and Analysis of Data. Main contributor to manuscript preparation.
Overall percentage (%)	75%
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date 17/03/17

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Kristie Lee
Contribution to the Paper	Contributor to manuscript preparation
Signature	Date 28/3/17

Name of Co-Author	Jozef Gecz
Contribution to the Paper	Contributor to manuscript preparation
Signature	Date 29/3/17

Name of Co-Author	Gaelle Friocourt			
Contribution to the Paper	Contributor to manuscript preparation			
Signature			Date	17/3/17

Name of Co-Author	Cheryl Shoubridge			
Contribution to the Paper	Supervised the project and provided critical feedback and revision of the manuscript			
Signature			Date	17/3/17

ORIGINAL ARTICLE

Embryonic forebrain transcriptome of mice with polyalanine expansion mutations in the *ARX* homeobox gene

Tessa Mattiske[1,2], Kristie Lee[1,2], Jozef Gecz[1,2], Gaelle Friocourt[3,4] and Cheryl Shoubridge[1,2,*]

[1]Department of Paediatrics, Adelaide Medical School, [2]Robinson Research Institute, University of Adelaide, SA, Australia, [3]Inserm, UMR1078, Brest, France and [4]Brest University, Faculté de Médecine et des Sciences de la Santé, Sfr ScInBioS, Brest, France

*To whom correspondence should be addressed at: Cheryl Shoubridge, Discipline of Paediatrics, Adelaide Medical School, Faculty of Health Sciences, The University of Adelaide, Adelaide, SA 5005, Australia. Tel: 61 8 8313 2355; Fax: 61 8 8313 4099; Email: cheryl.shoubridge@adelaide.edu.au

Abstract

The *Aristaless*-related homeobox (*ARX*) gene encodes a paired-type homeodomain transcription factor with critical roles in embryonic development. Mutations in *ARX* give rise to intellectual disability (ID), epilepsy and brain malformation syndromes. To capture the genetics and molecular disruptions that underpin the *ARX*-associated clinical phenotypes, we undertook a transcriptome wide RNASeq approach to analyse developing (12.5 dpc) telencephalon of mice modelling two recurrent polyalanine expansion mutations with different phenotypic severities in the *ARX* gene. Here we report 238 genes significantly deregulated (Log2FC > +/−1.1, P-value <0.05) when both mutations are compared to wild-type (WT) animals. When each mutation is considered separately, a greater number of genes were deregulated in the severe PA1 mice (825) than in the PA2 animals (78). Analysing genes deregulated in either or both mutant strains, we identified 12% as implicated in ID, epilepsy and autism (99/858), with ∼5% of them as putative or known direct targets of ARX transcriptional regulation. We propose a core pathway of transcription regulators, including *Hdac4*, involved in chromatin condensation and transcriptional repression, and one of its targets, the transcription factor *Twist1*, as potential drivers of the ID and infantile spasms in patients with *ARX* polyalanine expansion mutations. We predict that the subsequent disturbance to this pathway is a consequence of ARX protein reduction with a broader and more significant level of disruption in the PA1 in comparison to the PA2 mice. Identifying early triggers of *ARX*-associated phenotypes contributes to our understanding of particular clusters/pathways underpinning comorbid phenotypes that are shared by many neurodevelopmental disorders.

Introduction

Neurodevelopmental disorders (NDDs), which include intellectual disability, seizure disorders and autism spectrum disorders are prevalent in the population. Large-scale sequencing efforts have highlighted the genetic heterogeneity contributing to each of these disorders (1–5). Understanding how such divergent aetiologies produce similar clinical features remains a major challenge. Despite this, recent studies indicate that many of the pathophysiological mechanisms might be shared, opening the possibility that more than one condition may be amenable to a treatment or disease modification that exploits a common mechanism (3–5). Here we investigate the molecular mechanisms and functional impact of mutations in the disease-causing gene contributing to intellectual disability and infantile spasms, the

Aristaless-related homeobox gene (*ARX*) [NM_139058.2] (MIM 300382).

ARX is a member of the paired-type homeodomain transcription factor family with critical roles in development, particularly in the developing brain (6–8). ARX is indispensable for telencephalic morphogenesis particularly involved in radial and tangential migration of GABAergic interneuron progenitors, early commitment of cholinergic neurons and is emerging as a selector gene important in preserving the identity of specific brain regions (6,9–12). In accordance with the essential function of ARX during early brain development, *Arx* expression is detected in mice at embryonic day 8 in a restricted area of the neuroepithelium corresponding to the prospective forebrain (13). During the peak neural proliferation and neurogenesis expression of *Arx* within the subpallium peaks between 12.5 to 14.5 days post coitum (dpc), persisting during embryogenesis and is down regulated during postnatal life (6,14).

ARX is an X-chromosome gene. As such, patients are generally affected males with carrier females being asymptomatic. Over 60% of all mutations in ARX expand the first or second polyalanine tract, and affected males with these mutations invariably present with intellectual disability with and without infantile spasms and epilepsy (15,16). In particular, patients with expansions to the first polyalanine tract (previously reported as c.304ins(GCG)[7], now following HGVS nomenclature reported as c.306GGC[17]; referred to as a PA1 mutation in this study) invariably display seizures, with infantile spasms in 85% of these PA1 patients (16). The key phenotypic features seen in these patients are recapitulated in well characterised mutant mouse models, including infantile spasm-like movements, electrodecremetal discharges, and multifocal EEG spikes, with seizures in juvenile and older mice (17–19). The most frequent mutation in ARX in patients results in an expansion to the second polyalanine tract (previously reported as c.429_452dup, now following HGVS nomenclature reported as c.441-464dup; referred to as PA2 mutation in this study) with at least 10–15% of these PA2 patients presenting with infantile spasms in addition to intellectual disability (16). Although there is a mouse model of the most common PA2 mutation (17), phenotypic data for this strain, including prevalence of seizures is limited.

Our recent investigations in these mice modelling to two most frequent polyalanine expansion mutations in human patients demonstrated aggregation of mutant Arx protein does not occur in the embryonic brain. Instead, we identified a marked reduction in mutant Arx protein expression in the developing forebrain (12,17). Interestingly, this data indicates that both PA1 and PA2 mutations give rise to similar molecular outcomes. Despite recent studies identifying genes regulated by ARX (11,20,21) there is limited understanding of what impact the expanded polyalanine tract mutations in *Arx* may have on the transcriptional activity (12,22) and how this may contribute to phenotypic outcomes. In this study, we use RNASeq on brain tissue at 12.5 dpc during embryonic development to capture early disruptions of Arx function. We show that the gene expression consequences of the polyalanine expansion mutations of *Arx* do overlap but the PA1 mutation leads to a greater and broader disturbance than the PA2 mutation. From our analysis of the deregulated genes we propose a pathway involving Histone Deacetylase 4 (*Hdac*4) and Twist Family BHLH Transcription Factor 1 (*Twist*1) that when deregulated by either *Arx* mutation contributes to the comorbid phenotypes of intellectual disability and epilepsy.

Results

PA1 and PA2 mice deregulated transcriptomes overlap

Arx is indispensable for brain development with expression detected as early as 8 dpc (13,14). To capture early changes to the transcriptome due to mutations in *Arx* we collected and investigated the 12.5 dpc telencephalon of mice modelling two mutations in *Arx*, PA1 and PA2 (Supplementary Material, Fig. S1). We compared four males from each strain with stage-matched WT male littermates. Sequences were aligned using TopHat and count data generated from HTSeq was used as the input for EdgeR to identify genes with differential expression between samples.

Analysis of the PA1 mice revealed 825 genes deregulated by Log2 fold change greater than +/- 1.1 with a *P*-value of less than 0.05, with 54% found at higher levels of expression than WT (Fig. 1A). In contrast, the PA2 mutation resulted in 78 genes deregulated using the same fold cutoff, with 72% of these found at higher levels of expression than the WT animals (Fig. 1A). Despite the large difference in the number of genes significantly deregulated between the two PolyA strains, we noted that expression of many of the genes deregulated in the PA1 mice shared the same trend of deregulation in the PA2 mice although the lower log2FC values did not reach the required significance (Fig. 1B -within dotted lines). Given the similarities of the transcriptome changes between PA1 and PA2 we speculate that the disrupted pathway may be shared between PA1 and PA2. This is supported by previous studies suggesting both PA mutations result in a reduction of the ARX protein together with the overlap of clinical phenotypes in human patients with either PA1 or PA2 mutations. With this in mind, we analysed the genome wide expression data of both the PA1 and PA2 mice as a single mutation group (referred to as PolyA[pool]) compared to the WT samples to capture genes deregulated in both mutant strains. Lists of deregulated genes for each analysis are provided in Supplementary Materials, Tables S1–S3. From this analysis a total of 238 genes were identified (Log2 fold change greater than +/- 1.1 with a *P*-value < 0.05) (Fig. 1C). The majority of genes deregulated in the PolyA[pool] (89%) were at higher levels of expression in mutant mice (Fig. 1A). Not surprisingly, all genes identified as deregulated by this analysis had already been identified as deregulated in either PA1 or PA2 mice. When we consider the 238 genes identified as deregulated in the PolyA[pool] analysis, the mean log2FC value of the more severe PA1 group on its own is at 1.3, above the 1.1 cutoff value. In contrast, the milder PA2 group when considered on its own is below the 1.1 cutoff at 0.7. This means that of the total 238 significantly deregulated genes in the PolyA[pool] data, 94% (224/238) of genes in PA1 and 24% (56/238) of genes in PA2 met the cut off values, respectively (Fig. 1D). Of the 238 genes that were significantly deregulated in the PolyA[pool] analysis, 42 genes (18%) were significantly deregulated in both PA1 and PA2 animals when each mutant strain was considered independently compared to WT (*P* > 3.753e-31). When considered independently, 182 genes (76%) were significantly deregulated in PA1 animals only and 14 genes (6%) were deregulated specifically in the PA2 animals. Of the 42 significantly deregulated genes compared to WT in both PA1 and PA2, 79% (33/42) of the genes had increased expression (Fig. 1E). The overall disruption to the transcriptome observed supports the notion that PA1 and PA2 disrupt overlapping pathways, with PA2 mice being a 'milder' form of transcriptome deregulation of the PA1 mice.
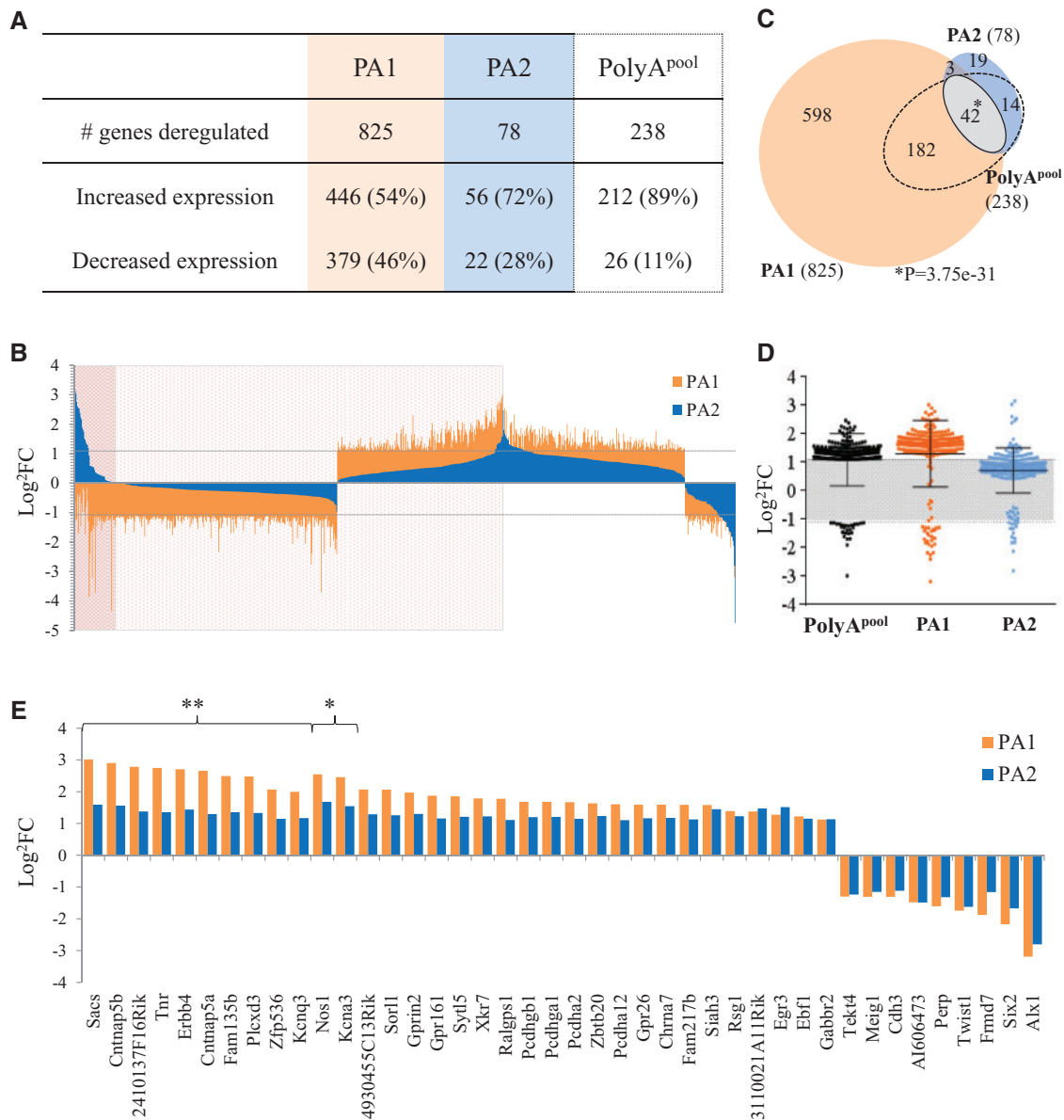
Figure 1. Transcriptome analysis of embryonic brains of PolyA Arx mutant mice. Differential expression of genes from 12.5 dpc mice was determined using EdgeR and selected based on a Log2 fold change greater than +/- 1.1 with a P-value < 0.05. **(A)** Total number of deregulated genes and the proportion of either increased or decreased expression from our analysis from both mutant mice strains independently, PA1 and PA2, and analysis of pooled PA1 and PA2 strains, PolyA[pool] when compared to WT animals. **(B)** Log2FC was plotted for all deregulated genes for both PA1 (orange) and PA2 (blue) (total = 858). Genes within the darker red shaded area are deregulated in different directions compared to WT with expression of genes significantly different between PA1 and PA2 (58/858, 6.75%). Genes within the lighter red area are deregulated in the same direction in both PA mutation groups compared to WT, but are still significantly different between PA1 and PA2 (498/858, 58.05%). The remaining genes are deregulated in the same direction compared to WT, but are not significantly different between PA1 and PA2 (35.2%). The dotted line indicates the log2FC +/- 1.1 cut off. **(C)** Gene lists from PA1 (orange), PA2 (blue) and PolyA[pool] analysis (dotted outline) was used to determine the overlap of genes deregulated in each group and visualised as a Venn diagram. Overlapping genes in PA1 and PA2 samples with log2FC > +/-1.1 in both lists and a P-value <0.05 are highlighted in grey (solid outline) with logFC values for individual genes shown in **(E)**. *significantly different between groups with a log2FC<1.1 and P-values <0.05, **significantly different between groups with a log2FC > +1.1 and P-values <0.05.

PA1 and PA2 mutations disturb overlapping biological processes in the developing brain

Functional enrichment analysis is a common tool to understand global changes in phenotypes in cells and tissues. EnrichR (23) analysis was used to investigate the enrichment of groups of genes with overlapping gene ontology (GO) terms representing gene properties with the focus of biological process. To interrogate the types of genes disturbed due to polyalanine tract expansion mutations in *Arx* we focused on the deregulated genes

identified in our analysis with higher levels of expression compared to WT (+1.1 log2FC). This focus was due mainly to the very small numbers of deregulated genes at lower levels of expression in PolyA mutant groups compared to WT in the core and PolyA[pool] subgroups. GO Terms were ranked using the EnrichR method of combining the P-value computed using the Fisher exact test with the z-score of the deviation from the expected rank by multiplying these two numbers as follows: $c = \log(p) * z$ (23). Enrichment of GO biological process identified

	GO number	GO Biological Process	PolyA pool	PA1 Only	All DE Genes	Combined score
cell-cell adhesion	0007156	homophilic cell adhesion via plasma membrane adhesion molecules				125
	0098609	cell-cell adhesion				110
	0098742	cell-cell adhesion via plasma-membrane adhesion molecules				70
neuron development	0045664	regulation of neuron differentiation				50
	0010975	regulation of neuron projection development				30
	0007399	nervous system development				15
neuron membrane properties	0007268	synaptic transmission				10
	0031344	regulation of cell projection organization				5
	0042391	regulation of membrane potential				
	0071804	cellular potassium ion transport				
	0071805	potassium ion transmembrane transport				
	0060059	embryonic retina morphogenesis in camera-type eye				
	0006366	transcription from RNA polymerase II promoter				
	0042787	protein ubiquitination in ubiquitin-dependent protein catabolic process				
	0010921	regulation of phosphatase activity				

Figure 2. Gene ontology classification of deregulated genes. Functional enrichment analysis of gene ontology (GO) terms for biological processes shows the differentially expressed genes within subgroups, PolyA^pool, PA1 only and all deregulated genes from both PA1 and PA2 (All deregulated (DE) genes) with a log2 fold change with +>1.1 cutoff value used as the input into EnrichR. The GO terms were ranked based on the combined EnrichR score and all had a P-value <0.05.

12 biological processes (adjusted *P*-value <0.05) enriched in the differentially expressed genes (with higher levels of expression compared to WT) relative to all expressed genes (Fig. 2). These biological processes fall into three main categories; those implicated in cell-cell adhesion (GO:0007156, GO:0098609, GO:0098742); regulation of neuron differentiation/nervous system development (GO:0045664, GO:0010975, GO:0007399); and neuron membrane properties such as synaptic transmission (GO:0007268), regulation of membrane potential (GO:0042391) and potassium ion transmembrane transport (GO:0071805) (Fig. 2). The levels to which each GO biological process is enriched in each group is displayed as a heat map on the right of Figure 2. This analysis indicates that the same categories of GO processes were enriched whether the deregulated genes considered were from the core overlap group, the PolyA^pool group or the broadest PA1 group of deregulated genes, particularly for the cell-cell adhesion process (Fig. 2). Overlap was also seen in the terms identified by Panther pathway analysis when comparing the top 10 enriched pathways in genes deregulated in PA1 mice and genes deregulated in PA2 mice (Supplementary Materials, Tables S4 and S5).

Identifying early triggers of ARX associated phenotypes

To ascertain which of the deregulated genes were likely to be direct transcriptional targets of Arx, the data from *Arx* knockout expression analysis (11,21) and ChIP studies (24) were used to identify that 46 genes deregulated in this study's data set are either known or putative direct targets of Arx. These targets accounted for 5% of all 858 deregulated genes and were detected across all groups considered, with a small level of enrichment in the PolyA^pool group (Fig. 3). Consistent with the fact that ARX is a transcriptional repressor, the majority of these targets were detected at higher levels of expression in the PolyA mutant animals compared to WT. We have previously demonstrated a marked reduction of mutant Arx protein abundance within the

developing forebrain of both PA1 and PA2 (12), indicating that the expanded polyalanine tract mutations in our mouse models represent a partial loss of *Arx* function. In Figure 3C, we captured the response of these target genes to PolyA mutation in *Arx* in our study compared to the response in the previously reported studies modelling knocked-out or ablated *Arx* expression in mouse brain (11,21) or in response to exogenous *Arx* overexpression in N2a cells (24). This analysis indicates that less than half (43%) of the direct gene targets of Arx identified as deregulated in the brains of 12.5 dpc mice with *Arx* PolyA mutations were also deregulated in *Arx* deficient mice. In general, the direction of deregulation of the target genes was in agreement between the loss of function studies and the partial loss of function in our PolyA mice. Not unexpectedly, there was more variation in the direction of deregulation between our partial loss of function in the PolyA mice when compared to the overexpression of exogenous *Arx* in Na2 cells (Fig. 3C).

De-regulation of early triggers of ARX-associated phenotypes persists across embryonic development

Considering the spatial expression of *Arx* in the subpallium is restricted to the in both lateral and medial ganglionic eminence, we were interested if there was any obvious relationship of spatial localisation in regard to the genes de-regulated in our PA mice. We compared expression of a number of deregulated genes that had available expression data from the Allen Developing Mouse Brain Atlas and EURExpress at suitable embryonic stages. Genes with a similar spatial expression profiles as determined by *in situ* hybridization images to that of *Arx* include *Ebf1*, *Rapgef5*, *Myt1l*, *Erbb4*, *Al606473*, *Zfp536* and *Gpr26*. In contrast, several genes had an opposite spatial expression profiles compared to *Arx* including *Zbtb20*, *Ptch1* and *Hdac4*, which are expressed within the proliferating cells in the ventricular zone (Supplementary Material, Fig. S2). This analysis highlights that the early triggers of *Arx* associated phenotypes identified in

A

	Core Overlap	PolyA^pool	PA1 Only	All DE Genes
Genes per Group	42	182	598	858
Direct Targets	3 (7%)	13 (7%)	28 (4.7%)	46 (5%)
Increased expression	3 (100%)	11 (85%)	13 (46%)	27 (59%)
Decreased expression		2 (15%)	15 (54%)	19 (41%)

B

C
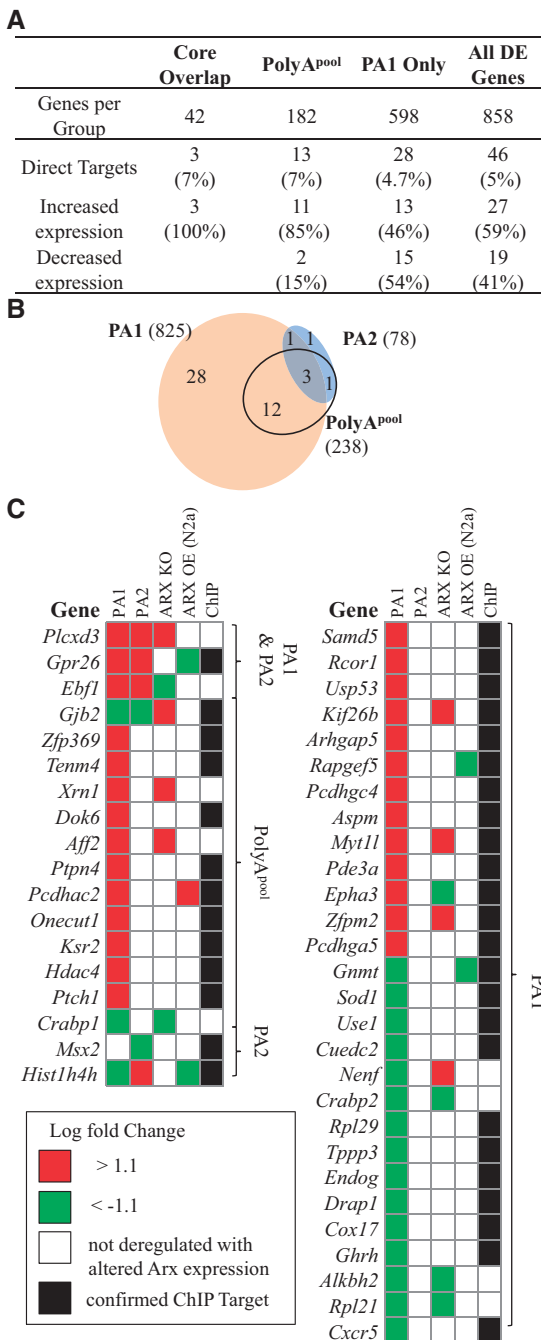
Figure 3. Disruption to putative and known ARX target genes. **(A)** The number of putative and known direct targets of ARX and the direction of deregulation in the PolyA mutant mice is shown for each subgroup of the RNASeq analysis. **(B)** The putative and known direct gene targets of ARX are spread across all subgroups of the RNASeq analysis as illustrated on the Venn diagram. **(C)** For each of these putative or direct gene targets of Arx identified in the PolyA mutant mice the change in expression compared to WT from this study (PA1 and PA2) with a comparison of the changes to expression in studies of *Arx* loss of function (ARX KO) (11,21), *Arx* overexpression in N2a cells (ARX OE) (24) and ChIP studies (24). Red indicates an increase in expression with Green a decrease in expression compared to WT.

this study includes both directly regulated and indirectly regulated gene targets contributing to disease outcomes.

To determine whether disruption to the transcriptome was constrained to the very early development in the time point tested by RNASeq or continued throughout embryonic life we chose to test fifteen genes by quantitative PCR in the Wt compared to PA1 mice across three developmental time points, 12.5, 14.5 and 18.5 dpc. The genes selected included a number a putative targets of Arx previously reported, *Plcxd3, Gpr26, Ebf1, Myt1l, Aspm, Hdac4* and *Ptch1*. Of interest also were deregulated genes that were known neurodevelopment genes (*Cdkl5, Erbb4, Sor1l, Twist1, Myt1l, Aspm, Hdac4* and *Ptch1*) or displayed spatial expression patterns indicated above (*Erbb4, Zfp536, Zbtb20, Gpr26, Ebf1, Myt1l, Hdac4* and *Ptch1*), or were deregulated in both PA1 and PA2 (*Kcna3* and *Six2*). Our analysis indicates that 11/15 genes validated with deregulated gene expression at 12.5 dpc by quantitative PCR, but also had deregulated expression in PA1 mutant mice across more advanced stages of embryonic development, with *Myt1l, Plcxd3,* and *Six2* consistently significantly deregulated across all time points examined (Supplementary Material, Fig. S3). For several genes with validated deregulation by quantitative PCR in PA1 mice we confirmed that the deregulation was also validated in PA2 samples at 12.5 dpc by this analysis, and moreover identified that deregulation in these genes persisted in later embryonic developmental time points (Supplementary Material, Fig. S3). Considering that the expression of *Arx* during development significantly diminishes at 18.5 dpc it was not surprising that we noted less consistent deregulation of genes persisting in the 18.5 dpc time point compared to earlier 14.5 dpc. This variability at 18.5 dpc could be due to a number of factors, but is likely influenced by the increased complexity and size of the brain (tissue sample collected) at this time point and underscores the strategy for targeting the early developmental stages where *Arx* expression is appreciable.

Deregulation of neurodevelopment disorder genes contributes to the polyalanine expansion mutation phenotype

Deregulated genes that overlap with lists of genes implicated in epilepsy, ID and autism (4) showed in the total group of 858 deregulated genes, 99 genes (12%) are known disease-causing genes for Epilepsy (55 genes), ID (29 genes) and Autism (51 genes) (Fig. 4A). This overlap is significant when considering the total group of deregulated genes ($P < 0.014$), the PolyA^pool group (29/182 = 16%, $P < 0.003$) and the core overlap group (9/42 = 21%, $P < 0.014$) but not in the PA1 only group ($P < 0.292$) (Fig. 4A and B). Given the incidence of comorbidity of the phenotypic features of these particular neurodevelopmental disorders (NDD), we were not surprised to see that many of the ID genes (22/29 = 76%) identified in our data also contribute to epilepsy and autism, with 8 genes contributing to all three comorbidities of epilepsy, ID and autism (Fig. 4C). The distribution of these NDD genes that were deregulated in our PolyA mice are shown in Fig. 4D. Of these 99 NDD genes, only 7 are currently identified as direct targets of ARX (bold entries on Fig. 4D) (24). Movement disorders including dystonia and dyskinesia are a frequent co-morbidity in patients with expanded polyA mutations in ARX. Known genes for these disorders are highlighted in the 99 NDD genes (underlined entries on Fig. 4D) with a full list of these genes deregulated in the PolyA mice in Supplementary Material, Table S6.

Early triggers of ARX associated phenotypes de-regulates the *arx-Hdac4-Twist1* pathway

The top 10 enriched terms by KEGG and Reactome pathway analysis for genes deregulated in the PolyA pool group highlight

A

	Core Overlap	PolyA^{pool}	PA1 Only	All DE Genes
Total Genes deregulated	42	182	598	858
ID Genes	3 (7%)	6 (3%)	20 (3%)	29 (3%)
Epilepsy Genes	4 (10%)	15 (8%)	35 (6%)	55 (6%)
Autism	6 (14%)	16 (9%)	29 (5%)	51 (6%)
Total NDD genes	**9 (21%)**	**29 (16%)**	**60 (10%)**	**99 (12%)**
P value	p < 0.014	p < 0.003	p < 0.292	p < 0.014

B

Combined graph (grey bars = Total NDD genes; right-hand y-axis = total number of NDD per group, 0–70; left-hand y-axis = % of total genes per group, 0–25). Categories: Overlap, PolyA^{pool}, PA1 Only.

Legend: Total NDD genes; ·····●····· % NDD Genes; —▲— % ID Genes; —◆— % Epilepsy Genes; —■— % Autism Genes

C

Venn diagram of overlap of genes associated with ID, Epilepsy and Autism comorbid phenotypes:
- ID (29)
- Epilepsy (55)
- Autism (51)
- ID only: 7; Epilepsy only: 32; Autism only: 32
- ID∩Epilepsy: 9; ID∩Autism: 5; Epilepsy∩Autism: 6; ID∩Epilepsy∩Autism: 8

D

	ID Epilepsy Autism	ID Epilepsy	Epilepsy Autism	ID Autism	ID	Epilepsy	Autism
Core Overlap	*Kcnq3*		*Chrna7*		*Twist1*	*Gabbr2, Sorl1*	*Fam135b, Pcdha2, Pcdha12*
PolyA pool	*Cdkl5* *Hcfc1* **Hdac4**	*Fktn* **Ptch1**	*Lrrc7*	**Aff2** *Erbb4*	*Pou3f4*	*Cbl, Csmd1, Csmd3, Fos, Gatad2b, Kcnh5, Lyst, Mib1, Mrs2, Syt14*	*Chrm3, Gabrq, Gan, Grid2, Pcdha6, Pcdha7, Pcdha9, Pcdhac1, Pcdhac2, Pcdhga11, Plxna4, Ptprt*
PA1 Only	*Crebbp Disc1 Lrp2 Vps13b*	**Aspm** *Cdk5rap Cep290 Chkb Kcna1 Nhs Shroom4*	*Dlgap2 Herc2 Kcnma1 Reln*	*Huwe1* **Myt1l** *Shank2*	*Cdon Gp1bb Nfix Rnaseh2c Tfap2a*	*Alg10b, Aqp4, Casc5, Ccdc88c, Cit, Dtnbp1, Epg5, Nalcn, Nanos3, Ndufa2, Ndufa8, Npy, Pet100, Pomc, Rest, Scarb2, Sepsecs, Vps13a, Xpnpep3*	*Aff4, Cacna1b, Cib2, Cntn5, Cntnap3, Fat1, Grcc10, Myo16, Nbea, Ndufa5, Nos1ap, Pcdh9, Pcdha3, Pcdha5, Rims3, Tbl1x, Wnk3*
PA2 Only							**Msx2**
Total	8	9	6	5	7	32	32

Figure 4. Genes deregulated in expanded polyA mice include known ID, epilepsy and autism genes. **(A)** Total number of genes associated with overlapping phenotypes of ID, epilepsy and autism and the proportion (%) per mutation subgroup. P-value indicates the significance of overlap between each PolyA mutant subgroup and the known neurodevelopment disorder genes. **(B)** A combined graph showing the total number of genes per subgroup (grey bars) on the right-hand y-axis compared to the proportion of overlap with known neurodevelopmental disorder genes across RNASeq analysis subgroups shown on the left-hand y-axis (legend for line graph shown below the graph). **(C)** Venn diagram showing the overlap of genes associated with ID, Epilepsy and Autism comorbid phenotypes. **(D)** List of known neurodevelopmental disorder genes within each PolyA mutant subgroup, grouped by the associated disease phenotypes. Known movement disorder genes are underlined. The putative and known direct targets of ARX/Arx are highlighted in bold.

the breadth of neurological processes impacted by the genes deregulated in these mutant mice (Supplementary Materials, Tables S7 and S8). Interestingly, from our analyses, we propose one particular pathway of transcriptional regulators implicated in NDD and deregulated in the PolyA mutant mice that may act as potential drivers of the phenotypes, particularly ID and epilepsy (Fig. 5). The pathway highlights direct interactions based on Ingenuity Pathway analysis and includes manually curated data from the literature based on ChIP studies to identify additional direct interactions across the pathway. We propose that the reduced protein expression in the developing brains of the

PolyA mice (12) leads to inadequate regulation of *Hdac4*, involved in chromatin condensation and transcriptional repression (25), which in turn has a flow on effects directly on targets such as *Mef2c* and *Twist1*. *Twist1* is a basic-helix-loop-helix transcription factor involved in cell lineage determination & differentiation (26). The expression of *Twist1* was significantly decreased in both the PA1 and PA2 mutant mice when analysed independently and confirmed by qRT-PCR (Supplementary Material, Fig. S3B).

A recent study using a ChIP approach to identify targets of TWIST1 (27) indicates that 302 genes deregulated in the current
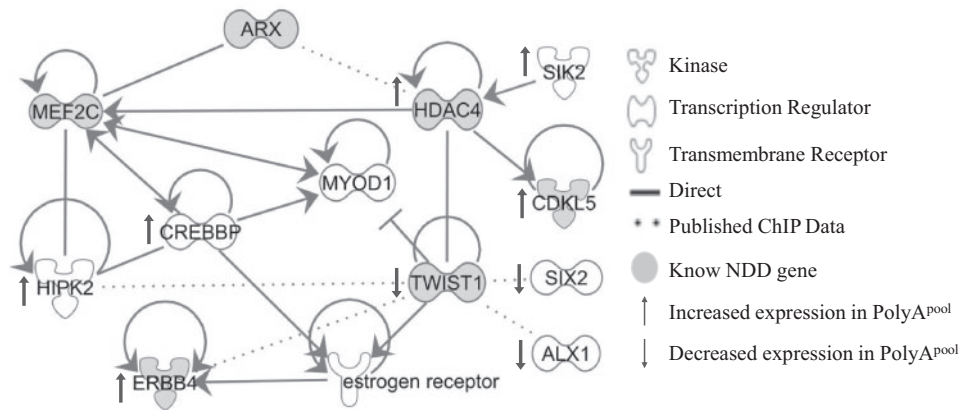
Figure 5. *Arx-Hdac4-Twist1* PolyA-deregulated pathway. Pathway analysis was used to assess connectivity of deregulated genes of PolyA-*Arx* loss-of-function embryonic brains. The geometric shapes reflect differing types of proteins as defined by ingenuity analysis (Ingenuity Systems) (see legend) with direct connections shown as solid lines. Published ChIP data highlighting relationships between genes and proteins are shown as dotted lines. The direction of deregulation in PolyA[pool] is indicated with arrows (↑ = increased expression, ↓ = decreased expression). Known NDD genes are shaded in grey.

study overlap with TWIST1 targets, 23/302 of which were deregulated in both PA1 and PA2 mutant mice, 17/302 being known ID genes and 32/302 being known epilepsy genes. We recently demonstrated that there was a consistently greater level of reduction of mutant Arx protein in the developing brains of the PA1 mice with a more variable reduction of mutant protein in the embryonic brains of PA2 mice (12). Hence, we propose that adequate levels of Arx protein are required to regulate direct targets such as *Hdac4* for normal brain development. When this regulation is not achieved the subsequent disturbance is relative to the reduction in Arx mutant protein expression, with a broader and more significant level of disruption observed in the PA1 mice as compared to the PA2 mice.

Discussion

Expanded polyalanine tracts are the most frequent mutations reported in the *ARX* gene (15) with the patients' phenotypic features invariably including ID with early onset seizures as a frequent comorbidity. To gauge the early events of the compromised ARX function, we assessed transcriptome-wide outcomes of the two most frequent *ARX* polyalanine tract expansion mutations within the telencephalon of their respective mouse models. RNASeq analysis of the forebrain of 12.5 dpc telencephalon highlights early alterations that are common to both PolyA mutants, with a greater impact in mice modelling the more severe PA1 mutation. Our validation analysis indicates that this deregulation identified at 12.5 dpc persists across embryonic development. Many of the genes deregulated are known ID, epilepsy or autism disease genes. The enrichment of genes with synaptic function supports convergence of common pathways contributing to these disorders. Interrogating the profile of deregulated genes from the developing brains of *Arx* PolyA mutant mice we have identified a number of pathways impacted by genes deregulated in the mutant mice including a 'core' pathway of transcriptional regulators that we suggest may represent early triggers of ARX-associated phenotypes.

The molecular pathogenesis of expanded polyalanine tracts is not well understood. In the case of *Arx*/ARX, we have demonstrated a marked reduction of the Arx protein expression in the developing forebrain of mutant PA1 and PA2 mice, occurring as early as embryonic stage 12.5 dpc (12). Reduced protein expression of mutant protein *in vivo* may be a common molecular

mechanism contributing to the disease associated with expanded polyalanine tracts (3,12,28,29). At this early embryonic stage *Arx*/ARX is highly expressed in the telencephalon during peak cellular proliferation/neurogenesis and the first wave of tangential migration of neurons from the ganglionic eminence to the developing cortex. There is evidence that polyalanine expansion mutations in *Arx* retain transcription regulation capacity for some gene targets, but fails to repress a subset of predicted or known targets (12,22,24). Our RNASeq approach offers an unbiased and comprehensive interrogation of deregulation of gene expression due to these mutations. We show a substantial number of deregulated genes as a consequence of PolyA expansion mutations in *Arx* during early embryonic brain development. Interestingly, most of the significantly deregulated genes in the PA1 mutant mice were also perturbed in PA2 mice, but failed to reach significance by our analysis. We cannot rule out the possibility that the structure and location of the PA2 mutation within Arx contributes to the milder phenotypic outcomes in both our mice and affected patients. However, a combination of data from this study and our previous work leads us to suggest an alternative hypothesis. We have shown that the reduction in Arx mutant protein in the embryonic brain appears more variable in the PA2 mice compared to the PA1 mice (12). Hence, the variability in Arx protein abundance in the PA2 mouse brains may diminish the power with which to detect significant deregulation of gene expression in some genes significantly altered in the PA1 mice. Our previous report identified expression of a direct transcriptional target of Arx, *Lmo1* in the PA1 mutant mice was deregulated significantly across all stages of embryonic development tested. In contrast, the deregulated expression of this gene did not reach significance in the PA2 mutant mice compared to wild-type until much later in embryonic development (12). Hence, in the subsequent analysis undertaken in this study we included not only genes significantly deregulated in both mutant animals but also those genes deregulated in either PA1 and PA2 animals.

The deregulated expression of several genes tested was maintained across increasing developmental time points. It is attractive to speculate that perhaps the cells with altered gene expression at E12.5 dpc maintain this altered state as they migrate and develop, despite the reduction in Arx expression with embryonic development. From our current data, we cannot speculate on alterations to migration of particular neurons

impacted by expanded polyA mutant Arx. However, we have conducted our analysis on a very spatially discrete region of the brain, namely the telencephalon at E12.5. We know from a recent elegant study (30) that only a small subset of cells is affected even when there is a complete loss of *Arx* function. Hence, to address if there is a migration deficit in these mutant animals extensive and painstaking cellular analysis of interneuron migration will be required. In addition, other approaches such as single cell expression analysis on the different interneuron subtypes would be very powerful and with the increasing technologies available, perhaps increasingly possible.

To date, there are only a handful of known or putative direct targets of ARX. In our study, approximately 5% of the overall deregulated genes in the PA1 and PA2 mice identified as direct ARX targets (known or putative). Knock-out studies (11,21) include phosphatidylinositol-specific phospholipase C, X domain containing 3 gene (*PLCXD3*) which is primarily expressed in the brain and associated with lipid catabolism and signal transduction, along with AF4/FMR2 Family, Member 2 (*Aff2*), Cellular Retinoic Acid Binding Protein 1 (*Crabp1*) and 5'-3' Exoribonuclease 1 (*Xrn1*). Other direct targets of Arx regulation identified using ChIP based approaches (24) that were deregulated in the PolyA Pool group included Histone Deacetylase 4 (*Hdac4*). However, we hypothesize that ARX drives key transcriptional events not only through impact on specific gene targets but in combination with regulation of key transcription factors and transcriptional regulators. Hence, we predict that deregulation of other transcription factors and regulators are likely to be contributing to the global changes observed in the PolyA mutant mice.

A number of transcription factors are present in the list of significantly deregulated genes in PA1 and PA2 mice, including *Twist1*, ALX Homeobox 1 (*Alx1*), SIX Homeobox 2 (*Six2*), Early B-Cell Factor 1 (*Ebf1*), Early Growth Response 3 (*Egr3*), Zinc Finger And BTB Domain Containing 20 (*Zbtb20*) and Zinc Finger Protein 536 (*Zfp536*) which are likely contributing to the downstream deregulation of the transcriptome captured in this study. From our analysis, we propose a 'core' *Arx-Hdac4-Twist1* pathway likely contributing to the downstream phenotypic outcomes. The pathway indicates that Arx directly represses *Hdac4*, which in turn is involved in chromatin condensation and transcriptional repression (25). Overexpression of *HDAC4* has been shown to impair long-term memory in a drosophila model (31) and overexpression has also been found in patients with autism (32). Salt-inducible kinase 2 (SIK2) an AMP-activated protein kinase, shown to regulate HDAC4 via phosphorylation, is significantly up regulated in PA1 and PA2 (although the change in transcript level is subtle). This modification of HDAC4 disrupts the MEF2C-HDAC4 complexes and mediates the activation of MEF2-dependent transcription. *MEF2C* is highly expressed during embryo development and involved in neurogenesis and synaptic function. In mice, it has been indicated to have an essential role in hippocampal-dependent learning and memory by suppressing the number of excitatory synapses and thus regulating basal and evoked synaptic transmission (33). The hyperexcitability of hippocampal and neocortical networks found in *Arx* PA1 mice is thought to be most likely due to an increase in excitatory drive rather than an inhibitory failure (19). Pathway analysis indicates Arx has a direct interaction with *Mef2c*, although we did not detect significant disturbances in transcript expression in our PolyA mutant mice at 12.5 dpc.

As we travel down the *Arx-Hdac4-Twist1* pathway, likely outside of direct regulation by Arx, we encounter Twist1, a basic-helix-loop-helix transcription factor involved in the cell linage

determination & differentiation (26). A wide range of mutations have been reported in *TWIST1* resulting in a variable phenotype from mild to severe intellectual disability as a consequence of large gene deletions, however the majority of missense and nonsense mutations have been associated with the craniosynostosis disorder Saethre-Chotzen syndrome (34). A recent study using a ChIP approach has identified targets of TWIST1, and comparison with our deregulated gene lists indicates that over one third of all deregulated genes in PolyA mutant mice identified in our current study are direct or likely targets of TWIST1 regulation, including genes known to cause ID and epilepsy (27). The influence of TWIST1 is further demonstrated with the down regulation of a known target Dnm30s (mir199a/214 cluster) normally upregulated via TWIST1 during development (35). Research has focused on determining the role of *Twist1* in cancer development with little emphasis to date of the contribution of *Twist1* in brain development.

The *Arx-Hdac4-Twist1* pathway we propose is one that may reflect very early consequences of ARX loss (or partial loss) of the function. For multiple genes deregulated at 12.5 dpc we have confirmed that the deregulation persists across later stages of embryonic development. Interestingly, genes that are enriched in the proposed pathway contain oestrogen response elements (EREs) in their promoter regions. More broadly, when we examine our list of deregulated genes we find 30% contain EREs (36–38). This is of significant interest given the recent findings that early postnatal treatment with 17 β-estradiol (E2) prevents spasms, restored depleted interneuron populations without increasing GABAergic synaptic density and altered mRNA levels of three downstream targets of ARX (*Ebf3*, *Shox2*, *Lgi1*) in an independent PA1 mutant mouse model (39). It remains to be established if administration of E2 during early stages of postnatal life, leading to seizure ablation, is impacting upon the transcriptome specifically at the ERE containing genes we have identified as deregulated in the PolyA mice at early stages of embryonic development.

In summary, we have interrogated the transcriptome of mice modelling the most frequent expansion mutations of PolyA tracts of ARX. Based on our analysis, we propose the *Arx-Hdac4-Twist1* as the 'core' pathway, which is contributing significantly to the *ARX* mutation-associated phenotypes, namely ID and epilepsy. While other gene targets and factors are likely at play, the *Arx-Hdac4-Twist1* pathway offers a plausible target for future interventions.

Materials and Methods

Sample collection

Animals and tissue collection
All animal procedures were approved by the relevant Animal Ethics committees of the University of Adelaide, SA Pathology and the Women's and Children's Hospital, Adelaide. $Arx^{GCG7/+}$ (PA1) and $Arx^{432–455dup/+}$ (PA2) mice were obtained as described in (12) and were maintained in the C57BL/6 background. Pregnant dams were euthanized by cervical dislocation followed by harvesting of the 12.5 dpc embryos from the uterus. Skin and ectodermal layers were removed to isolate the telencephalic vesicles. Samples for RNA extraction were snap frozen at −80°C until the time of assay.

Genotyping
Genomic DNA was extracted as per Maxwell® 16 Tissue DNA purification Kit manual (Promega). PCR was performed using

FailSafe™ PCR 2X PreMix J (Epicentre) for 35 cycles of 30 s of 94°C for denaturation, 30 s of 60°C for annealing and 40 s of 72°C for elongation. Primers to amplify the *Arx* knock-in region were described (17). We also included *Sry* sexing PCR as part of our genotyping pipeline (12).

RNA-sequencing

RNA was extracted from the isolated telencephalon of hemizygous male mice of each strain (PA1 and PA2) and stage matched male wild-type littermates (WT) using Trizol (Invitrogen, Grand Island, NY, USA) and RNeasy Mini Kit (QIAGEN, Hilden, Germany), according to the manufacturer's instructions. Library preparation using the TruSeq RNA Sample Preparation Kit v2 was performed on 4.5 µg of RNA at the ACRF South Australia Cancer Genomics Facility (Adelaide, Australia). Samples (n = 4 each from WT, PA1 and PA2) were sequenced on the Illumina (San Diego, CA, USA) HISeq 2000 platform. RNA-SEQ reads were mapped to the reference sequences, which includes the latest build of the mouse genome (mm10). Reads were mapped at the original 100 bp length. The number of reads mapped to each gene was obtained using htseq-count (40). In order to correct for variation between lanes/sample, the count data was normalized to library size. Genes with low count data were excluded, the minimum required at least two samples having >60–177 reads. Differential gene expression was calculated using the R package edgeR (41). Transcripts significantly altered compared to WT were selected by applying a log2 fold change-cutoff of 1.1 and P-value cutoff of ≤ 0.05.

RT-qPCR – RNASeq validation

RNASeq results were validated using Taqman RT-qPCR on two pools of RNA, a technical validation pool using RNA from the same samples used for the RNASeq and a separate biological validation pool of RNA from four different samples of each genotype (Supplementary Material, Fig. S1). Collected tissues were homogenized with a 21G needle and total RNA was extracted using Trizol (Invitrogen) and RNeasy Mini Kit (Qiagen) and treated with DNase I (Qiagen) according to the manufacturer's instruction. cDNA was prepared as described in SuperScript III reverse transcriptase (Invitrogen) manual with 1 µg of RNA primed by random hexanucleotides. Along with samples, template negative and reverse transcriptase negative controls were included to determine product specificity. Genes selected for validation studies were assayed as described in Taqman® PreAmp Master Mix Kit user guide (Applied Biosystem). For each validation gene quantified with a Taqman probe labelled with FAM, the expression values were normalized to the reference gene *Gapdh* assayed in the same well using the Taqman probe labelled with VIC. For further analysis, RNA was extracted from 12.5, 14.5 and 18.5 dpc telencephalon and pooled before cDNA was prepared (as described previously). Expression of genes was determined using Taqman® PreAmp Master Mix with gene specific Taqman probes labelled with FAM. The expression values were normalized to the reference gene *Tbp* which was shown to be stably expressed across the chosen time points. Taqman probes used in this study are listed in Supplementary Material, Table S9.

Functional annotations

Statistical analysis of the enrichment of Gene Ontology (GO) categories was performed using EnrichR, a bioinformatics tool that retrieves molecular information from transcription factor databases and defines transcription factors statistically enriched from gene lists (23). To rank the enrichment results we used the score calculated by EnrichR using the P-value and Z-score. The top 10 results are shown with adjusted P-value of $P < 0.05$.

Pathway analysis

Ingenuity Pathway Analysis (Ingenuity Systems) was used to assess connectivity of deregulated proteins. The requirements for assessing protein-protein interconnectivity included direct interactions observed experimentally. The permitted interaction types were: protein-protein, protein-DNA, activation, inhibition, phosphorylation, and ubiquitination. ARX (21,24) and TWIST1 (27) ChIP interaction data was manually superimposed onto this pathway.

Statistics

The statistical significance of the overlap between two groups of genes was calculated using exact hypergeometric probability (http://nemates.org/MA/progs/overlap_stats.html; date last accessed September 27, 2016). Total genes in this case equalled 13821 genes with detected reads.

Supplementary Material

Supplementary Material is available at *HMG* online.

Conflict of Interest statement. None declared.

Funding

References

1. Epi, K.C., Epilepsy Phenome/Genome, P., Allen, A.S., Berkovic, S.F., Cossette, P., Delanty, N., Dlugos, D., Eichler, E.E., Epstein, M.P., Glauser, T., *et al.* (2013) De novo mutations in epileptic encephalopathies. *Nature*, **501**, 217–221.
2. Euro, E.R.E.S.C., Epilepsy Phenome/Genome, P. and Epi, K.C. (2014) De novo mutations in synaptic transmission genes including DNM1 cause epileptic encephalopathies. *Am. J. Hum. Genet.*, **95**, 360–370.
3. Krumm, N., O'Roak, B.J., Shendure, J. and Eichler, E.E. (2014) A de novo convergence of autism genetics and molecular neuroscience. *Trends Neurosci.*, **37**, 95–105.
4. Pinto, D., Delaby, E., Merico, D., Barbosa, M., Merikangas, A., Klei, L., Thiruvahindrapuram, B., Xu, X., Ziman, R., Wang, Z., *et al.* (2014) Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *Am. J. Hum. Genet.*, **94**, 677–694.
5. Chen, E.S., Gigek, C.O., Rosenfeld, J.A., Diallo, A.B., Maussion, G., Chen, G.G., Vaillancourt, K., Lopez, J.P., Crapper, L., Poujol, R., *et al.* (2014) Molecular convergence of neurodevelopmental disorders. *Am. J. Hum. Genet.*, **95**, 490–508.
6. Kitamura, K., Yanazawa, M., Sugiyama, N., Miura, H., Iizuka-Kogo, A., Kusaka, M., Omichi, K., Suzuki, R., Kato-Fukui, Y., Kamiirisa, K., *et al.* (2002) Mutation of ARX causes abnormal development of forebrain and testes in mice and X-linked

lissencephaly with abnormal genitalia in humans. *Nat. Genet.*, **32**, 359–369.

7. Ohira, R., Zhang, Y.H., Guo, W., Dipple, K., Shih, S.L., Doerr, J., Huang, B.L., Fu, L.J., Abu-Khalil, A., Geschwind, D., *et al.* (2002) Human ARX gene: genomic characterization and expression. *Mol. Genet. Metab.*, **77**, 179–188.

8. Depositario-Cabacar, D.F. and Zelleke, T.G. (2010) Treatment of epilepsy in children with developmental disabilities. *Dev. Disabil. Res. Rev.*, **16**, 239–247.

9. Colombo, E., Collombat, P., Colasante, G., Bianchi, M., Long, J., Mansouri, A., Rubenstein, J.L. and Broccoli, V. (2007) Inactivation of Arx, the murine ortholog of the X-linked lissencephaly with ambiguous genitalia gene, leads to severe disorganization of the ventral telencephalon with impaired neuronal migration and differentiation. *J. Neurosci.*, **27**, 4786–4798.

10. Friocourt, G., Kanatani, S., Tabata, H., Yozu, M., Takahashi, T., Antypa, M., Raguenes, O., Chelly, J., Ferec, C., Nakajima, K., *et al.* (2008) Cell-autonomous roles of ARX in cell proliferation and neuronal migration during corticogenesis. *J. Neurosci.*, **28**, 5794–5805.

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

12. Lee, K., Mattiske, T., Kitamura, K., Gecz, J. and Shoubridge, C. (2014) Reduced polyalanine-expanded Arx mutant protein in developing mouse subpallium alters Lmo1 transcriptional regulation. *Hum. Mol. Genet.*, **23**, 1084–1094.

13. Bienvenu, T., Poirier, K., Friocourt, G., Bahi, N., Beaumont, D., Fauchereau, F., Ben Jeema, L., Zemni, R., Vinet, M.C., Francis, F., *et al.* (2002) ARX, a novel Prd-class-homeobox gene highly expressed in the telencephalon, is mutated in X-linked mental retardation. *Hum. Mol. Genet.*, **11**, 981–991.

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

15. Shoubridge, C., Fullston, T. and Gecz, J. (2010) ARX spectrum disorders: making inroads into the molecular pathology. *Hum. Mutat.*, **31**, 889–900.

16. Marques, I., Sa, M.J., Soares, G., Mota Mdo, C., Pinheiro, C., Aguiar, L., Amado, M., Soares, C., Calado, A., Dias, P., *et al.* (2015) Unraveling the pathogenesis of ARX polyalanine tract variants using a clinical and molecular interfacing approach. *Mol. Genet. Genomic Med.*, **3**, 203–214.

17. Kitamura, K., Itou, Y., Yanazawa, M., Ohsawa, M., Suzuki-Migishima, R., Umeki, Y., Hohjoh, H., Yanagawa, Y., Shinba, T., Itoh, M., *et al.* (2009) Three human ARX mutations cause the lissencephaly-like and mental retardation with epilepsy-like pleiotropic phenotypes in mice. *Hum. Mol. Genet.*, **18**, 3708–3724.

18. Price, M.G., Yoo, J.W., Burgess, D.L., Deng, F., Hrachovy, R.A., Frost, J.D., Jr. and Noebels, J.L. (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**, 8752–8763.

19. Beguin, S., Crepel, V., Aniksztejn, L., Becq, H., Pelosi, B., Pallesi-Pocachard, E., Bouamrane, L., Pasqualetti, M., Kitamura, K., Cardoso, C., *et al.* (2013) An epilepsy-related ARX polyalanine expansion modifies glutamatergic neurons excitability and morphology without affecting GABAergic neurons development. *Cereb. Cortex*, **23**, 1484–1494.

20. Friocourt, G. and Parnavelas, J.G. (2011) Identification of Arx targets unveils new candidates for controlling cortical interneuron migration and differentiation. *Front. Cell Neurosci.*, **5**, 28.

21. Fulp, C.T., Cho, G., Marsh, E.D., Nasrallah, I.M., Labosky, P.A. and Golden, J.A. (2008) Identification of Arx transcriptional targets in the developing basal forebrain. *Hum. Mol. Genet.*, **17**, 3740–3760.

22. Nasrallah, M.P., Cho, G., Simonet, J.C., Putt, M.E., Kitamura, K. and Golden, J.A. (2012) Differential effects of a polyalanine tract expansion in Arx on neural development and gene expression. *Hum. Mol. Genet.*, **21**, 1090–1098.

23. Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R. and Ma'ayan, A. (2013) Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*, **14**, 128.

24. Quille, M.L., Carat, S., Quemener-Redon, S., Hirchaud, E., Baron, D., Benech, C., Guihot, J., Placet, M., Mignen, O., Ferec, C., *et al.* (2011) High-throughput analysis of promoter occupancy reveals new targets for Arx, a gene mutated in mental retardation and interneuronopathies. *PLoS One*, **6**, e25181.

25. Fischer, A., Sananbenesi, F., Mungenast, A. and Tsai, L.H. (2010) Targeting the correct HDAC(s) to treat cognitive disorders. *Trends Pharmacol. Sci.*, **31**, 605–617.

26. Nieto, M.A. (2013) Epithelial plasticity: a common theme in embryonic and cancer cells. *Science*, **342**, 1234850.

27. Lee, M.P., Ratner, N. and Yutzey, K.E. (2014) Genome-wide Twist1 occupancy in endocardial cushion cells, embryonic limb buds, and peripheral nerve sheath tumor cells. *BMC Genomics*, **15**, 821.

28. Innis, J.W., Mortlock, D., Chen, Z., Ludwig, M., Williams, M.E., Williams, T.M., Doyle, C.D., Shao, Z., Glynn, M., Mikulic, D., *et al.* (2004) Polyalanine expansion in HOXA13: three new affected families and the molecular consequences in a mouse model. *Hum. Mol. Genet.*, **13**, 2841–2851.

29. Kuss, P., Villavicencio-Lorini, P., Witte, F., Klose, J., Albrecht, A.N., Seemann, P., Hecht, J. and Mundlos, S. (2009) Mutant Hoxd13 induces extra digits in a mouse model of synpolydactyly directly and by decreasing retinoic acid synthesis. *J. Clin. Invest.*, **119**, 146–156.

30. Marsh, E.D., Nasrallah, M.P., Walsh, C., Murray, K.A., Nicole Sunnen, C., McCoy, A. and Golden, J.A. (2016) Developmental interneuron subtype deficits after targeted loss of Arx. *BMC Neurosci.*, **17**, 35.

31. Fitzsimons, H.L., Schwartz, S., Given, F.M. and Scott, M.J. (2013) The histone deacetylase HDAC4 regulates long-term memory in Drosophila. *PLoS One*, **8**, e83903.

32. Nardone, S., Sams, D.S., Reuveni, E., Getselter, D., Oron, O., Karpuj, M. and Elliott, E. (2014) DNA methylation analysis of the autistic brain reveals multiple dysregulated biological pathways. *Transl. Psychiatry*, **4**, e433.

33. Barbosa, A.C., Kim, M.S., Ertunc, M., Adachi, M., Nelson, E.D., McAnally, J., Richardson, J.A., Kavalali, E.T., Monteggia, L.M., Bassel-Duby, R., *et al.* (2008) MEF2C, a transcription factor that facilitates learning and memory by negative regulation of synapse numbers and function. *Proc. Natl Acad. Sci. U S A*, **105**, 9391–9396.

34. el Ghouzzi, V., Le Merrer, M., Perrin-Schmitt, F., Lajeunie, E., Benit, P., Renier, D., Bourgeois, P., Bolcato-Bellemin, A.L., Munnich, A. and Bonaventure, J. (1997) Mutations of the TWIST gene in the Saethre-Chotzen syndrome. *Nat. Genet.*, **15**, 42–46.

35. Lee, Y.B., Bantounas, I., Lee, D.Y., Phylactou, L., Caldwell, M.A. and Uney, J.B. (2009) Twist-1 regulates the miR-199a/214 cluster during development. *Nucleic Acids Res.*, **37**, 123–128.

36. Tang, S., Han, H. and Bajic, V.B. (2004) ERGDB: Estrogen Responsive Genes Database. *Nucleic Acids Res.*, **32**, D533–D536.

37. Bourdeau, V., Deschenes, J., Metivier, R., Nagai, Y., Nguyen, D., Bretschneider, N., Gannon, F., White, J.H. and Mader, S. (2004) Genome-wide identification of high-affinity estrogen response elements in human and mouse. *Mol. Endocrinol.*, **18**, 1411–1427.

38. Kamalakaran, S., Radhakrishnan, S.K. and Beck, W.T. (2005) Identification of estrogen-responsive genes using a genome-wide analysis of promoter elements for transcription factor binding sites. *J. Biol. Chem.*, **280**, 21491–21497.

39. Olivetti, P.R., Maheshwari, A. and Noebels, J.L. (2014) Neonatal estradiol stimulation prevents epilepsy in Arx model of X-linked infantile spasms syndrome. *Sci. Transl. Med.*, **6**, 220ra212.

40. Anders, S., Pyl, P.T. and Huber, W. (2015) HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166–169.

41. Robinson, M.D., McCarthy, D.J. and Smyth, G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.

References

8   References

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." <u>Dev Med Child Neurol</u> **52**(3): 305-307.

Ades, S. E. and R. T. Sauer (1994). "Differential DNA-binding specificity of the engrailed homeodomain: the role of residue 50." <u>Biochemistry</u> **33**(31): 9187-9194.

Albrecht, A., 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." <u>Hum Mol Genet</u> **13**(20): 2351-2359.

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

Anders, S., P. T. Pyl and W. Huber (2015). "HTSeq--a Python framework to work with high-throughput sequencing data." <u>Bioinformatics</u> **31**(2): 166-169.

Anderson, S. A., D. D. Eisenstat, L. Shi and J. L. Rubenstein (1997a). "Interneuron migration from basal forebrain to neocortex: dependence on Dlx genes." <u>Science</u> **278**(5337): 474-476.

Anderson, S. A., M. Qiu, A. Bulfone, D. D. Eisenstat, J. Meneses, R. Pedersen and J. L. Rubenstein (1997b). "Mutations of the homeobox genes Dlx-1 and Dlx-2 disrupt the striatal subventricular zone and differentiation of late born striatal neurons." <u>Neuron</u> **19**(1): 27-37.

Bachetti, T., P. Bocca, S. Borghini, I. Matera, I. Prigione, R. Ravazzolo and I. Ceccherini (2007). "Geldanamycin promotes nuclear localisation and clearance of PHOX2B misfolded proteins containing polyalanine expansions." <u>Int J Biochem Cell Biol</u> **39**(2): 327-339.

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." <u>Hum Mol Genet</u> **14**(13): 1815-1824.

Barbosa, A. C., M. S. Kim, M. Ertunc, M. Adachi, E. D. Nelson, J. McAnally, J. A. Richardson, E. T. Kavalali, L. M. Monteggia, R. Bassel-Duby and E. N. Olson (2008). "MEF2C, a transcription factor that facilitates learning and memory by negative regulation of synapse numbers and function." <u>Proc Natl Acad Sci U S A</u> **105**(27): 9391-9396.

Beguin, S., V. Crepel, L. Aniksztejn, H. Becq, B. Pelosi, E. Pallesi-Pocachard, L. Bouamrane, M. Pasqualetti, K. Kitamura, C. Cardoso and A. Represa (2013). "An epilepsy-

related ARX polyalanine expansion modifies glutamatergic neurons excitability and morphology without affecting GABAergic neurons development." <u>Cereb Cortex</u> **23**(6): 1484-1494.

Bernacki, J. P. and R. M. Murphy (2011). "Length-dependent aggregation of uninterrupted polyalanine peptides." <u>Biochemistry</u> **50**(43): 9200-9211.

Bettella, E., G. Di Rosa, R. Polli, E. Leonardi, G. Tortorella, S. Sartori and A. Murgia (2013). "Early-onset epileptic encephalopathy in a girl carrying a truncating mutation of the ARX gene: rethinking the ARX phenotype in females." <u>Clin Genet</u> **84**(1): 82-85.

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." <u>Hum Mol Genet</u> **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." <u>Ann Neurol</u> **51**(3): 340-349.

Bookout, A. L. and D. J. Mangelsdorf (2003). "Quantitative real-time PCR protocol for analysis of nuclear receptor signaling pathways." <u>Nucl Recept Signal</u> **1**: e012.

Bourdeau, V., J. Deschenes, R. Metivier, Y. Nagai, D. Nguyen, N. Bretschneider, F. Gannon, J. H. White and S. Mader (2004). "Genome-wide identification of high-affinity estrogen response elements in human and mouse." <u>Mol Endocrinol</u> **18**(6): 1411-1427.

Brais, B. (2003). "Oculopharyngeal muscular dystrophy: a late-onset polyalanine disease." <u>Cytogenet Genome Res</u> **100**(1-4): 252-260.

Brown, L., M. Paraso, R. Arkell and S. Brown (2005). "In vitro analysis of partial loss-of-function ZIC2 mutations in holoprosencephaly: alanine tract expansion modulates DNA binding and transactivation." <u>Hum Mol Genet</u> **14**(3): 411-420.

Brown, L. Y. and S. A. Brown (2004). "Alanine tracts: the expanding story of human illness and trinucleotide repeats." <u>Trends Genet</u> **20**(1): 51-58.

Bruneau, S., K. R. Johnson, M. Yamamoto, A. Kuroiwa and D. Duboule (2001). "The mouse Hoxd13(spdh) mutation, a polyalanine expansion similar to human type II synpolydactyly (SPD), disrupts the function but not the expression of other Hoxd genes." <u>Dev Biol</u> **237**(2): 345-353.

Butt, S. J., M. Fuccillo, S. Nery, S. Noctor, A. Kriegstein, J. G. Corbin and G. Fishell (2005). "The temporal and spatial origins of cortical interneurons predict their physiological subtype." Neuron **48**(4): 591-604.

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.

Calado, A., F. M. Tome, B. Brais, G. A. Rouleau, U. Kuhn, E. Wahle and M. Carmo-Fonseca (2000). "Nuclear inclusions in oculopharyngeal muscular dystrophy consist of poly(A) binding protein 2 aggregates which sequester poly(A) RNA." Hum Mol Genet **9**(15): 2321-2328.

Chen, E. S., C. O. Gigek, J. A. Rosenfeld, A. B. Diallo, G. Maussion, G. G. Chen, K. Vaillancourt, J. P. Lopez, L. Crapper, R. Poujol, L. G. Shaffer, G. Bourque and C. Ernst (2014). "Molecular convergence of neurodevelopmental disorders." Am J Hum Genet **95**(5): 490-508.

Chen, E. Y., C. M. Tan, Y. Kou, Q. Duan, Z. Wang, G. V. Meirelles, N. R. Clark and A. Ma'ayan (2013). "Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool." BMC Bioinformatics **14**: 128.

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., 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.

Colombo, E., P. Collombat, G. Colasante, M. Bianchi, J. Long, A. Mansouri, J. L. Rubenstein and V. Broccoli (2007). "Inactivation of Arx, the murine ortholog of the X-

linked lissencephaly with ambiguous genitalia gene, leads to severe disorganization of the ventral telencephalon with impaired neuronal migration and differentiation." <u>J Neurosci</u> **27**(17): 4786-4798.

Colombo, E., R. Galli, G. Cossu, J. Gecz and V. Broccoli (2004). "Mouse orthologue of ARX, a gene mutated in several X-linked forms of mental retardation and epilepsy, is a marker of adult neural stem cells and forebrain GABAergic neurons." <u>Dev Dyn</u> **231**(3): 631-639.

Cooper, G. M., B. P. Coe, S. Girirajan, J. A. Rosenfeld, T. H. Vu, C. Baker, C. Williams, H. Stalker, R. Hamid, V. Hannig, H. Abdel-Hamid, P. Bader, E. McCracken, D. Niyazov, K. Leppig, H. Thiese, M. Hummel, N. Alexander, J. Gorski, J. Kussmann, V. Shashi, K. Johnson, C. Rehder, B. C. Ballif, L. G. Shaffer and E. E. Eichler (2011). "A copy number variation morbidity map of developmental delay." <u>Nat Genet</u> **43**(9): 838-846.

Davies, S. W., M. Turmaine, B. A. Cozens, M. DiFiglia, A. H. Sharp, C. A. Ross, E. Scherzinger, E. E. Wanker, L. Mangiarini and G. P. Bates (1997). "Formation of neuronal intranuclear inclusions underlies the neurological dysfunction in mice transgenic for the HD mutation." <u>Cell</u> **90**(3): 537-548.

Desplan, C., J. Theis and P. H. O'Farrell (1988). "The sequence specificity of homeodomain-DNA interaction." <u>Cell</u> **54**(7): 1081-1090.

Di Lascio, S., T. Bachetti, E. Saba, I. Ceccherini, R. Benfante and D. Fornasari (2013). "Transcriptional dysregulation and impairment of PHOX2B auto-regulatory mechanism induced by polyalanine expansion mutations associated with congenital central hypoventilation syndrome." <u>Neurobiol Dis</u> **50**: 187-200.

Di Zanni, E., T. Bachetti, S. Parodi, P. Bocca, I. Prigione, S. Di Lascio, D. Fornasari, R. Ravazzolo and I. Ceccherini (2012). "In vitro drug treatments reduce the deleterious effects of aggregates containing polyAla expanded PHOX2B proteins." <u>Neurobiol Dis</u> **45**(1): 508-518.

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." <u>Am J Med Genet</u> **86**(4): 331-337.

Doran, C. M., S. L. Einfeld, R. H. Madden, M. Otim, S. K. Horstead, L. A. Ellis and E. Emerson (2012). "How much does intellectual disability really cost? First estimates for Australia." <u>J Intellect Dev Disabil</u> **37**(1): 42-49.

Dubreuil, V., N. Ramanantsoa, D. Trochet, V. Vaubourg, J. Amiel, J. Gallego, J. F. Brunet and C. Goridis (2008). "A human mutation in Phox2b causes lack of CO2 chemosensitivity, fatal central apnea, and specific loss of parafacial neurons." <u>Proc Natl Acad Sci U S A</u> **105**(3): 1067-1072.

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 **52**(5): 984-992.

el Ghouzzi, V., M. Le Merrer, F. Perrin-Schmitt, E. Lajeunie, P. Benit, D. Renier, P. Bourgeois, A. L. Bolcato-Bellemin, A. Munnich and J. Bonaventure (1997). "Mutations of the TWIST gene in the Saethre-Chotzen syndrome." Nat Genet **15**(1): 42-46.

Epi, K. C., P. Epilepsy Phenome/Genome, A. S. Allen, S. F. Berkovic, P. Cossette, N. Delanty, D. Dlugos, E. E. Eichler, M. P. Epstein, T. Glauser, D. B. Goldstein, Y. Han, E. L. Heinzen, Y. Hitomi, K. B. Howell, M. R. Johnson, R. Kuzniecky, D. H. Lowenstein, Y. F. Lu, M. R. Madou, A. G. Marson, H. C. Mefford, S. Esmaeeli Nieh, T. J. O'Brien, R. Ottman, S. Petrovski, A. Poduri, E. K. Ruzzo, I. E. Scheffer, E. H. Sherr, C. J. Yuskaitis, B. Abou-Khalil, B. K. Alldredge, J. F. Bautista, S. F. Berkovic, A. Boro, G. D. Cascino, D. Consalvo, P. Crumrine, O. Devinsky, D. Dlugos, M. P. Epstein, M. Fiol, N. B. Fountain, J. French, D. Friedman, E. B. Geller, T. Glauser, S. Glynn, S. R. Haut, J. Hayward, S. L. Helmers, S. Joshi, A. Kanner, H. E. Kirsch, R. C. Knowlton, E. H. Kossoff, R. Kuperman, R. Kuzniecky, D. H. Lowenstein, S. M. McGuire, P. V. Motika, E. J. Novotny, R. Ottman, J. M. Paolicchi, J. M. Parent, K. Park, A. Poduri, I. E. Scheffer, R. A. Shellhaas, E. H. Sherr, J. J. Shih, R. Singh, J. Sirven, M. C. Smith, J. Sullivan, L. Lin Thio, A. Venkat, E. P. Vining, G. K. Von Allmen, J. L. Weisenberg, P. Widdess-Walsh and M. R. Winawer (2013). "De novo mutations in epileptic encephalopathies." Nature **501**(7466): 217-221.

Euro, E.-R. E. S. C., P. Epilepsy Phenome/Genome and K. C. Epi (2014). "De novo mutations in synaptic transmission genes including DNM1 cause epileptic encephalopathies." Am J Hum Genet **95**(4): 360-370.

Fischer, A., F. Sananbenesi, A. Mungenast and L. H. Tsai (2010). "Targeting the correct HDAC(s) to treat cognitive disorders." Trends Pharmacol Sci **31**(12): 605-617.

Fitzsimons, H. L., S. Schwartz, F. M. Given and M. J. Scott (2013). "The histone deacetylase HDAC4 regulates long-term memory in Drosophila." PLoS One **8**(12): e83903.

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.

Friocourt, G. and J. G. Parnavelas (2011). "Identification of Arx targets unveils new candidates for controlling cortical interneuron migration and differentiation." Front Cell Neurosci **5**: 28.

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." <u>Clin Genet</u> **80**(6): 510-522.

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." <u>Hum Mol Genet</u> **17**(23): 3740-3760.

Gehring, W. J., Y. Q. Qian, M. Billeter, K. Furukubo-Tokunaga, A. F. Schier, D. Resendez-Perez, M. Affolter, G. Otting and K. Wuthrich (1994). "Homeodomain-DNA recognition." <u>Cell</u> **78**(2): 211-223.

Giri, K., U. Ghosh, N. P. Bhattacharyya and S. Basak (2003). "Caspase 8 mediated apoptotic cell death induced by beta-sheet forming polyalanine peptides." <u>FEBS Lett</u> **555**(2): 380-384.

Goridis, C., V. Dubreuil, M. Thoby-Brisson, G. Fortin and J. F. Brunet (2010). "Phox2b, congenital central hypoventilation syndrome and the control of respiration." <u>Semin Cell Dev Biol</u> **21**(8): 814-822.

Han, K. and J. L. Manley (1993). "Functional domains of the Drosophila Engrailed protein." <u>EMBO J</u> **12**(7): 2723-2733.

Hebert, J. M. and G. Fishell (2008). "The genetics of early telencephalon patterning: some assembly required." <u>Nat Rev Neurosci</u> **9**(9): 678-685.

Hrachovy, R. A. and J. D. Frost, Jr. (2013). "Infantile spasms." <u>Handb Clin Neurol</u> **111**: 611-618.

Hughes, J., S. Piltz, N. Rogers, D. McAninch, L. Rowley and P. Thomas (2013). "Mechanistic insight into the pathology of polyalanine expansion disorders revealed by a mouse model for X linked hypopituitarism." <u>PLoS Genet</u> **9**(3): e1003290.

Ingolia, N. T., S. Ghaemmaghami, J. R. Newman and J. S. Weissman (2009). "Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling." <u>Science</u> **324**(5924): 218-223.

Innis, J. W., D. Mortlock, Z. Chen, M. Ludwig, M. E. Williams, T. M. Williams, C. D. Doyle, Z. Shao, M. Glynn, D. Mikulic, K. Lehmann, S. Mundlos and B. Utsch (2004). "Polyalanine expansion in HOXA13: three new affected families and the molecular consequences in a mouse model." <u>Hum Mol Genet</u> **13**(22): 2841-2851.

Jackson, M. R., K. Lee, T. Mattiske, E. J. Jaehne, E. Ozturk, B. T. Baune, T. J. O'Brien, N. Jones and C. Shoubridge (Submitted 2017). "Extensive phenotypic evaluation of mouse models recapitulating two common ARX polyalanine expansion mutations which span the clinical spectrum of ID and epilepsy." <u>Neurobiology of Disease</u>.

Jasinska, A., G. Michlewski, M. de Mezer, K. Sobczak, P. Kozlowski, M. Napierala and W. J. Krzyzosiak (2003). "Structures of trinucleotide repeats in human transcripts and their functional implications." Nucleic Acids Res **31**(19): 5463-5468.

Kamalakaran, S., S. K. Radhakrishnan and W. T. Beck (2005). "Identification of estrogen-responsive genes using a genome-wide analysis of promoter elements for transcription factor binding sites." J Biol Chem **280**(22): 21491-21497.

Karlin, S. and C. Burge (1995). "Dinucleotide relative abundance extremes: a genomic signature." Trends Genet **11**(7): 283-290.

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. 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.

Kearse, M. G., K. M. Green, A. Krans, C. M. Rodriguez, A. E. Linsalata, A. C. Goldstrohm and P. K. Todd (2016). "CGG Repeat-Associated Non-AUG Translation Utilizes a Cap-Dependent Scanning Mechanism of Initiation to Produce Toxic Proteins." Mol Cell **62**(2): 314-322.

Kim, Y. E., F. Hosp, F. Frottin, H. Ge, M. Mann, M. Hayer-Hartl and F. U. Hartl (2016). "Soluble Oligomers of PolyQ-Expanded Huntingtin Target a Multiplicity of Key Cellular Factors." Mol Cell **63**(6): 951-964.

Kissinger, C. R., B. S. Liu, E. Martin-Blanco, T. B. Kornberg and C. O. Pabo (1990). "Crystal structure of an engrailed homeodomain-DNA complex at 2.8 A resolution: a framework for understanding homeodomain-DNA interactions." Cell **63**(3): 579-590.

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." <u>Nat Genet</u> **32**(3): 359-369.

Krumm, N., B. J. O'Roak, J. Shendure and E. E. Eichler (2014). "A de novo convergence of autism genetics and molecular neuroscience." <u>Trends Neurosci</u> **37**(2): 95-105.

Kuss, P., P. Villavicencio-Lorini, F. Witte, J. Klose, A. N. Albrecht, P. Seemann, J. Hecht and S. Mundlos (2009). "Mutant Hoxd13 induces extra digits in a mouse model of synpolydactyly directly and by decreasing retinoic acid synthesis." <u>J Clin Invest</u> **119**(1): 146-156.

Kwong, A. K., A. C. Ho, C. W. Fung and V. C. Wong (2015). "Analysis of mutations in 7 genes associated with neuronal excitability and synaptic transmission in a cohort of children with non-syndromic infantile epileptic encephalopathy." <u>PLoS One</u> **10**(5): e0126446.

La Spada, A. R. and J. P. Taylor (2010). "Repeat expansion disease: progress and puzzles in disease pathogenesis." <u>Nat Rev Genet</u> **11**(4): 247-258.

Lacoste, C., J. P. Desvignes, D. Salgado, C. Pecheux, L. Villard, M. Bartoli, C. Beroud, N. Levy, C. Badens and M. Krahn (2016). "Coverage Analysis of Lists of Genes involved in Heterogeneous Genetic Diseases following Benchtop Exome Sequencing using the Ion Proton." <u>J Genet</u> **95**(1): 203-208.

Larson, S. A., K. C. Lakin, L. Anderson, N. Kwak Lee and D. Anderson (2001). "Prevalence of mental retardation and developmental disabilities: estimates from the 1994/1995 national health interview survey disability supplements." <u>Am J Ment Retard</u> **106**(3): 231-252.

Lavoie, H., F. Debeane, Q. D. Trinh, J. F. Turcotte, L. P. Corbeil-Girard, M. J. Dicaire, A. Saint-Denis, M. Page, G. A. Rouleau and B. Brais (2003). "Polymorphism, shared functions and convergent evolution of genes with sequences coding for polyalanine domains." <u>Hum Mol Genet</u> **12**(22): 2967-2979.

Lee, K., T. Mattiske, K. Kitamura, J. Gecz and C. Shoubridge (2014). "Reduced polyalanine-expanded Arx mutant protein in developing mouse subpallium alters Lmo1 transcriptional regulation." <u>Hum Mol Genet</u> **23**(4): 1084-1094.

Lee, M. P., N. Ratner and K. E. Yutzey (2014). "Genome-wide Twist1 occupancy in endocardial cushion cells, embryonic limb buds, and peripheral nerve sheath tumor cells." <u>BMC Genomics</u> **15**: 821.

Lee, Y. B., I. Bantounas, D. Y. Lee, L. Phylactou, M. A. Caldwell and J. B. Uney (2009). "Twist-1 regulates the miR-199a/214 cluster during development." <u>Nucleic Acids Res</u> **37**(1): 123-128.

Lehrke, R. (1972). "Theory of X-linkage of major intellectual traits." <u>Am J Ment Defic</u> **76**(6): 611-619.

Lehrke, R. G. (1974). "X-linked mental retardation and verbal disability." <u>Birth Defects Orig Artic Ser</u> **10**(1): 1-100.

Lek, M., K. J. Karczewski, E. V. Minikel, K. E. Samocha, E. Banks, T. Fennell, A. H. O'Donnell-Luria, J. S. Ware, A. J. Hill, B. B. Cummings, T. Tukiainen, D. P. Birnbaum, J. A. Kosmicki, L. E. Duncan, K. Estrada, F. Zhao, J. Zou, E. Pierce-Hoffman, J. Berghout, D. N. Cooper, N. Deflaux, M. DePristo, R. Do, J. Flannick, M. Fromer, L. Gauthier, J. Goldstein, N. Gupta, D. Howrigan, A. Kiezun, M. I. Kurki, A. L. Moonshine, P. Natarajan, L. Orozco, G. M. Peloso, R. Poplin, M. A. Rivas, V. Ruano-Rubio, S. A. Rose, D. M. Ruderfer, K. Shakir, P. D. Stenson, C. Stevens, B. P. Thomas, G. Tiao, M. T. Tusie-Luna, B. Weisburd, H. H. Won, D. Yu, D. M. Altshuler, D. Ardissino, M. Boehnke, J. Danesh, S. Donnelly, R. Elosua, J. C. Florez, S. B. Gabriel, G. Getz, S. J. Glatt, C. M. Hultman, S. Kathiresan, M. Laakso, S. McCarroll, M. I. McCarthy, D. McGovern, R. McPherson, B. M. Neale, A. Palotie, S. M. Purcell, D. Saleheen, J. M. Scharf, P. Sklar, P. F. Sullivan, J. Tuomilehto, M. T. Tsuang, H. C. Watkins, J. G. Wilson, M. J. Daly, D. G. MacArthur and C. Exome Aggregation (2016). "Analysis of protein-coding genetic variation in 60,706 humans." <u>Nature</u> **536**(7616): 285-291.

Leonard, H. and X. Wen (2002). "The epidemiology of mental retardation: challenges and opportunities in the new millennium." <u>Ment Retard Dev Disabil Res Rev</u> **8**(3): 117-134.

Li, X., V. Su, W. E. Kurata, C. Jin and A. F. Lau (2008). "A novel connexin43-interacting protein, CIP75, which belongs to the UbL-UBA protein family, regulates the turnover of connexin43." <u>J Biol Chem</u> **283**(9): 5748-5759.

Lubs, H. A., R. E. Stevenson and C. E. Schwartz (2012). "Fragile X and X-linked intellectual disability: four decades of discovery." <u>Am J Hum Genet</u> **90**(4): 579-590.

Ma, T., C. Wang, L. Wang, X. Zhou, M. Tian, Q. Zhang, Y. Zhang, J. Li, Z. Liu, Y. Cai, F. Liu, Y. You, C. Chen, K. Campbell, H. Song, L. Ma, J. L. Rubenstein and Z. Yang (2013). "Subcortical origins of human and monkey neocortical interneurons." <u>Nat Neurosci</u> **16**(11): 1588-1597.

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." <u>Eur J Hum Genet</u> **12**(9): 689-693.

Marques, I., M. J. Sa, G. Soares, C. Mota Mdo, C. Pinheiro, L. Aguiar, M. Amado, C. Soares, A. Calado, P. Dias, A. B. Sousa, A. M. Fortuna, R. Santos, K. B. Howell, M. M. Ryan, R. J. Leventer, R. Sachdev, R. Catford, K. Friend, T. R. Mattiske, C. Shoubridge and P. Jorge (2015). "Unraveling the pathogenesis of ARX polyalanine tract variants using a clinical and molecular interfacing approach." <u>Mol Genet Genomic Med</u> **3**(3): 203-214.

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.

Marsh, E., M. P. Nasrallah, C. Walsh, K. A. Murray, C. Nicole Sunnen, A. McCoy and J. A. Golden (2016). "Developmental interneuron subtype deficits after targeted loss of Arx." BMC Neurosci **17**(1): 35.

Mattiske, T., K. Lee, J. Gecz, G. Friocourt and C. Shoubridge (2016). "Embryonic forebrain transcriptome of mice with polyalanine expansion mutations in the ARX homeobox gene." Hum Mol Genet.

Mattiske, T. R., M. H. Tan, J. Gecz and C. Shoubridge (2013). "Challenges of "sticky" co-immunoprecipitation: polyalanine tract protein-protein interactions." Methods Mol Biol **1017**: 121-133.

Maulik, P. K., M. N. Mascarenhas, C. D. Mathers, T. Dua and S. Saxena (2011). "Prevalence of intellectual disability: a meta-analysis of population-based studies." Res Dev Disabil **32**(2): 419-436.

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.

Mefford, H. C., M. L. Batshaw and E. P. Hoffman (2012). "Genomics, intellectual disability, and autism." N Engl J Med **366**(8): 733-743.

Meijlink, F., A. Beverdam, A. Brouwer, T. C. Oosterveen and D. T. Berge (1999). "Vertebrate aristaless-related genes." Int J Dev Biol **43**(7): 651-663.

Mirzaa, G. M., A. R. Paciorkowski, E. D. Marsh, E. M. Berry-Kravis, L. Medne, A. Alkhateeb, A. Grix, E. C. Wirrell, B. R. Powell, K. C. Nickels, B. Burton, A. Paras, K. Kim, W. Chung, W. B. Dobyns and S. Das (2013). "CDKL5 and ARX mutations in males with early-onset epilepsy." Pediatr Neurol **48**(5): 367-377.

Miura, H., M. Yanazawa, K. Kato and K. Kitamura (1997). "Expression of a novel aristaless related homeobox gene 'Arx' in the vertebrate telencephalon, diencephalon and floor plate." Mech Dev **65**(1-2): 99-109.

Moumne, L., A. Dipietromaria, F. Batista, A. Kocer, M. Fellous, E. Pailhoux and R. A. Veitia (2008). "Differential aggregation and functional impairment induced by polyalanine expansions in FOXL2, a transcription factor involved in cranio-facial and ovarian development." Hum Mol Genet **17**(7): 1010-1019.

Nardone, S., D. S. Sams, E. Reuveni, D. Getselter, O. Oron, M. Karpuj and E. Elliott (2014). "DNA methylation analysis of the autistic brain reveals multiple dysregulated biological pathways." Transl Psychiatry **4**: e433.

Nasrallah, I., 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.

Nasrallah, M., G. Cho, J. C. Simonet, M. E. Putt, K. Kitamura and J. A. Golden (2012). "Differential effects of a polyalanine tract expansion in Arx on neural development and gene expression." Hum Mol Genet **21**(5): 1090-1098.

Nieto, M. A. (2013). "Epithelial plasticity: a common theme in embryonic and cancer cells." Science **342**(6159): 1234850.

Norris, R. A. and M. J. Kern (2001). "The identification of Prx1 transcription regulatory domains provides a mechanism for unequal compensation by the Prx1 and Prx2 loci." J Biol Chem **276**(29): 26829-26837.

Norris, R. A., K. K. Scott, C. S. Moore, G. Stetten, C. R. Brown, E. W. Jabs, E. A. Wulfsberg, J. Yu and M. J. Kern (2000). "Human PRRX1 and PRRX2 genes: cloning, expression, genomic localization, and exclusion as disease genes for Nager syndrome." Mamm Genome **11**(11): 1000-1005.

Ofer, N., P. Weisman-Shomer, J. Shklover and M. Fry (2009). "The quadruplex r(CGG)n destabilizing cationic porphyrin TMPyP4 cooperates with hnRNPs to increase the translation efficiency of fragile X premutation mRNA." Nucleic Acids Res **37**(8): 2712-2722.

Ohira, R., Y. H. Zhang, W. Guo, K. Dipple, S. L. Shih, J. Doerr, B. L. Huang, L. J. Fu, A. Abu-Khalil, D. Geschwind and E. R. McCabe (2002). "Human ARX gene: genomic characterization and expression." Mol Genet Metab **77**(1-2): 179-188.

Olivetti, P. R., A. Maheshwari and J. L. Noebels (2014). "Neonatal estradiol stimulation prevents epilepsy in Arx model of X-linked infantile spasms syndrome." Sci Transl Med **6**(220): 220ra212.

Parodi, S., E. Di Zanni, S. Di Lascio, P. Bocca, I. Prigione, D. Fornasari, M. Pennuto, T. Bachetti and I. Ceccherini (2012). "The E3 ubiquitin ligase TRIM11 mediates the degradation of congenital central hypoventilation syndrome-associated polyalanine-expanded PHOX2B." J Mol Med (Berl) **90**(9): 1025-1035.

Penrose, L. S. (1983). A clinical and genetic study of 1280 cases of mental defect. Medical Research Council Special Report Series. London, Institute for Research into Mental and Multiple Handicap.

Pinto, D., E. Delaby, D. Merico, M. Barbosa, A. Merikangas, L. Klei, B. Thiruvahindrapuram, X. Xu, R. Ziman, Z. Wang, J. A. Vorstman, A. Thompson, R. Regan, M. Pilorge, G. Pellecchia, A. T. Pagnamenta, B. Oliveira, C. R. Marshall, T. R. Magalhaes, J. K. Lowe, J. L. Howe, A. J. Griswold, J. Gilbert, E. Duketis, B. A. Dombroski, M. V. De Jonge, M. Cuccaro, E. L. Crawford, C. T. Correia, J. Conroy, I. C. Conceicao, A. G. Chiocchetti, J. P. Casey, G. Cai, C. Cabrol, N. Bolshakova, E. Bacchelli, R. Anney, S. Gallinger, M. Cotterchio, G. Casey, L. Zwaigenbaum, K. Wittemeyer, K. Wing, S. Wallace, H. van Engeland, A. Tryfon, S. Thomson, L. Soorya, B. Roge, W. Roberts, F. Poustka, S. Mouga, N. Minshew, L. A. McInnes, S. G. McGrew, C. Lord, M. Leboyer, A. S. Le Couteur, A. Kolevzon, P. Jimenez Gonzalez, S. Jacob, R. Holt, S. Guter, J. Green, A. Green, C. Gillberg, B. A. Fernandez, F. Duque, R. Delorme, G. Dawson, P. Chaste, C. Cafe, S. Brennan, T. Bourgeron, P. F. Bolton, S. Bolte, R. Bernier, G. Baird, A. J. Bailey, E. Anagnostou, J. Almeida, E. M. Wijsman, V. J. Vieland, A. M. Vicente, G. D. Schellenberg, M. Pericak-Vance, A. D. Paterson, J. R. Parr, G. Oliveira, J. I. Nurnberger, A. P. Monaco, E. Maestrini, S. M. Klauck, H. Hakonarson, J. L. Haines, D. H. Geschwind, C. M. Freitag, S. E. Folstein, S. Ennis, H. Coon, A. Battaglia, P. Szatmari, J. S. Sutcliffe, J. Hallmayer, M. Gill, E. H. Cook, J. D. Buxbaum, B. Devlin, L. Gallagher, C. Betancur and S. W. Scherer (2014). "Convergence of genes and cellular pathways dysregulated in autism spectrum disorders." Am J Hum Genet **94**(5): 677-694.

Plenge, R. M., B. D. Hendrich, C. Schwartz, J. F. Arena, A. Naumova, C. Sapienza, R. M. Winter and H. F. Willard (1997). "A promoter mutation in the XIST gene in two unrelated families with skewed X-chromosome inactivation." Nat Genet **17**(3): 353-356.

Poirier, K., D. Lacombe, B. Gilbert-Dussardier, M. Raynaud, V. Desportes, A. P. M. 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.

Poirier, K., H. Van Esch, G. Friocourt, Y. Saillour, N. Bahi, S. Backer, E. Souil, L. Castelnau-Ptakhine, C. Beldjord, F. Francis, T. Bienvenu and J. Chelly (2004). "Neuroanatomical distribution of ARX in brain and its localisation in GABAergic neurons." Brain Res Mol Brain Res **122**(1): 35-46.

Polling, S., A. R. Ormsby, R. J. Wood, K. Lee, C. Shoubridge, J. N. Hughes, P. Q. Thomas, M. D. Griffin, A. F. Hill, Q. Bowden, T. Bocking and D. M. Hatters (2015). "Polyalanine expansions drive a shift into alpha-helical clusters without amyloid-fibril formation." Nat Struct Mol Biol **22**(12): 1008-1015.

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.

Prontera, P., V. Ottaviani, D. Rogaia, I. Isidori, A. Mencarelli, N. Malerba, D. Cocciadiferro, P. Rolph, G. Stangoni, A. Vulto-van Silfhout and G. Merla (2016). "A novel

MED12 mutation: Evidence for a fourth phenotype." <u>Am J Med Genet A</u> **170**(9): 2377-2382.

Proud, V. K., C. Levine and N. J. Carpenter (1992). "New X-linked syndrome with seizures, acquired micrencephaly, and agenesis of the corpus callosum." <u>Am J Med Genet</u> **43**(1-2): 458-466.

Puzynski, S. (1992). "[Affective disorders in international classification of disorders ICD-10: classification of mental and behavioral disorders]." <u>Psychiatr Pol</u> **26**(3-4): 175-183.

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." <u>PLoS One</u> **6**(9): e25181.

Rallu, M., J. G. Corbin and G. Fishell (2002). "Parsing the prosencephalon." <u>Nat Rev Neurosci</u> **3**(12): 943-951.

Reijnders, M. R., V. Zachariadis, B. Latour, L. Jolly, G. M. Mancini, R. Pfundt, K. M. Wu, C. M. van Ravenswaaij-Arts, H. E. Veenstra-Knol, B. M. Anderlid, S. A. Wood, S. W. Cheung, A. Barnicoat, F. Probst, P. Magoulas, A. S. Brooks, H. Malmgren, A. Harila-Saari, C. M. Marcelis, M. Vreeburg, E. Hobson, V. R. Sutton, Z. Stark, J. Vogt, N. Cooper, J. Y. Lim, S. Price, A. H. Lai, D. Domingo, B. Reversade, D. D. D. Study, J. Gecz, C. Gilissen, H. G. Brunner, U. Kini, R. Roepman, A. Nordgren and T. Kleefstra (2016). "De Novo Loss-of-Function Mutations in USP9X Cause a Female-Specific Recognizable Syndrome with Developmental Delay and Congenital Malformations." <u>Am J Hum Genet</u> **98**(2): 373-381.

Riley, B. E., Y. Xu, H. Y. Zoghbi and H. T. Orr (2004). "The effects of the polyglutamine repeat protein ataxin-1 on the UbL-UBA protein A1Up." <u>J Biol Chem</u> **279**(40): 42290-42301.

Robinson, M. D., D. J. McCarthy and G. K. Smyth (2010). "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data." <u>Bioinformatics</u> **26**(1): 139-140.

Ropers, H. H. (2010). "Genetics of early onset cognitive impairment." <u>Annu Rev Genomics Hum Genet</u> **11**: 161-187.

Rusconi, L., L. Salvatoni, L. Giudici, I. Bertani, C. Kilstrup-Nielsen, V. Broccoli and N. Landsberger (2008). "CDKL5 expression is modulated during neuronal development and its subcellular distribution is tightly regulated by the C-terminal tail." <u>J Biol Chem</u> **283**(44): 30101-30111.

Savarese, M., O. Musumeci, T. Giugliano, A. Rubegni, C. Fiorillo, F. Fattori, A. Torella, R. Battini, C. Rodolico, A. Pugliese, G. Piluso, L. Maggi, A. D'Amico, C. Bruno, E. Bertini,

F. M. Santorelli, M. Mora, A. Toscano, C. Minetti and V. Nigro (2016). "Novel findings associated with MTM1 suggest a higher number of female symptomatic carriers." Neuromuscul Disord **26**(4-5): 292-299.

Schalock, R. L., M. A. Verdugo and L. E. Gomez (2011). "Evidence-based practices in the field of intellectual and developmental disabilities: an international consensus approach." Eval Program Plann **34**(3): 273-282.

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.

Selkoe, D. J. (2003). "Folding proteins in fatal ways." Nature **426**(6968): 900-904.

Sellier, C., R. A. Buijsen, F. He, S. Natla, L. Jung, P. Tropel, A. Gaucherot, H. Jacobs, H. Meziane, A. Vincent, M. F. Champy, T. Sorg, G. Pavlovic, M. Wattenhofer-Donze, M. C. Birling, M. Oulad-Abdelghani, P. Eberling, F. Ruffenach, M. Joint, M. Anheim, V. Martinez-Cerdeno, F. Tassone, R. Willemsen, R. K. Hukema, S. Viville, C. Martinat, P. K. Todd and N. Charlet-Berguerand (2017). "Translation of Expanded CGG Repeats into FMRpolyG Is Pathogenic and May Contribute to Fragile X Tremor Ataxia Syndrome." Neuron **93**(2): 331-347.

Shields, W. D. (2006). "Infantile spasms: little seizures, BIG consequences." Epilepsy Curr **6**(3): 63-69.

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., M. Field, R. J. Leventer and J. Gecz (2016). Developmental Abnormalities Due to Mutations in the Aristaless-Related Homeobox Gene. Epstein's Inborn Errors of Development: The Molecular Basis of Clinical Disorders of Morphogenesis (3 ed.). R. P. Erickson and A. J. Wynshaw-Boris. New York, United States of America, Oxford University Press.

Shoubridge, C., T. Fullston and J. Gecz (2010a). "ARX spectrum disorders: making inroads into the molecular pathology." Hum Mutat **31**(8): 889-900.

Shoubridge, C. and J. Gecz (2012). "Polyalanine tract disorders and neurocognitive phenotypes." Adv Exp Med Biol **769**: 185-203.

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.

Shoubridge, C., M. H. Tan, G. Seiboth and J. Gecz (2012). "ARX homeodomain mutations abolish DNA binding and lead to a loss of transcriptional repression." Hum Mol Genet **21**(7): 1639-1647.

Sinajon, P., D. Verbaan and J. So (2016). "The expanding phenotypic spectrum of female SLC9A6 mutation carriers: a case series and review of the literature." Hum Genet **135**(8): 841-850.

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.

Stromme, P., M. E. Mangelsdorf, I. E. Scheffer and J. Gecz (2002). "Infantile spasms, dystonia, and other X-linked phenotypes caused by mutations in Aristaless related homeobox gene, ARX." Brain Dev **24**(5): 266-268.

Takagi, M., T. Ishii, C. Torii, K. Kosaki and T. Hasegawa (2014). "A novel mutation in SOX3 polyalanine tract: a case of Kabuki syndrome with combined pituitary hormone deficiency harboring double mutations in MLL2 and SOX3." Pituitary **17**(6): 569-574.

Tan, M. H., J. Gecz and C. Shoubridge (2013). "PCR amplification and sequence analysis of GC-rich sequences: Aristaless-related homeobox example." Methods Mol Biol **1017**: 105-120.

Tang, S., H. Han and V. B. Bajic (2004). "ERGDB: Estrogen Responsive Genes Database." Nucleic Acids Res **32**(Database issue): D533-536.

Theiler, K. (1989). The House Mouse, Springer-Verlag Berlin Heidelberg.

Todd, P. K., S. Y. Oh, A. Krans, F. He, C. Sellier, M. Frazer, A. J. Renoux, K. C. Chen, K. M. Scaglione, V. Basrur, K. Elenitoba-Johnson, J. P. Vonsattel, E. D. Louis, M. A. Sutton, J. P. Taylor, R. E. Mills, N. Charlet-Berguerand and H. L. Paulson (2013). "CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome." Neuron **78**(3): 440-455.

Trochet, D., S. J. Hong, J. K. Lim, J. F. Brunet, A. Munnich, K. S. Kim, S. Lyonnet, C. Goridis and J. Amiel (2005). "Molecular consequences of PHOX2B missense, frameshift and alanine expansion mutations leading to autonomic dysfunction." Hum Mol Genet **14**(23): 3697-3708.

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.

Tzschach, A., U. Grasshoff, S. Beck-Woedl, C. Dufke, C. Bauer, M. Kehrer, C. Evers, U. Moog, B. Oehl-Jaschkowitz, N. Di Donato, R. Maiwald, C. Jung, A. Kuechler, S. Schulz, P. Meinecke, S. Spranger, J. Kohlhase, J. Seidel, S. Reif, M. Rieger, A. Riess, M. Sturm, J. Bickmann, C. Schroeder, A. Dufke, O. Riess and P. Bauer (2015). "Next-generation sequencing in X-linked intellectual disability." Eur J Hum Genet **23**(11): 1513-1518.

Utsch, B., C. D. McCabe, K. Galbraith, R. Gonzalez, M. Born, J. Dotsch, M. Ludwig, H. Reutter and J. W. Innis (2007). "Molecular characterization of HOXA13 polyalanine expansion proteins in hand-foot-genital syndrome." Am J Med Genet A **143A**(24): 3161-3168.

Villavicencio-Lorini, P., P. Kuss, J. Friedrich, J. Haupt, M. Farooq, S. Turkmen, D. Duboule, J. Hecht and S. Mundlos (2010). "Homeobox genes d11-d13 and a13 control mouse autopod cortical bone and joint formation." J Clin Invest **120**(6): 1994-2004.

Vissers, L. E., C. Gilissen and J. A. Veltman (2016). "Genetic studies in intellectual disability and related disorders." Nat Rev Genet **17**(1): 9-18.

Wallerstein, R., R. Sugalski, L. Cohn, R. Jawetz and M. Friez (2008). "Expansion of the ARX spectrum." Clin Neurol Neurosurg **110**(6): 631-634.

Weisman-Shomer, P., E. Cohen, I. Hershco, S. Khateb, O. Wolfovitz-Barchad, L. H. Hurley and M. Fry (2003). "The cationic porphyrin TMPyP4 destabilizes the tetraplex form of the fragile X syndrome expanded sequence d(CGG)n." Nucleic Acids Res **31**(14): 3963-3970.

Winnepenninckx, B., V. Errijgers, E. Reyniers, P. P. De Deyn, F. E. Abidi, C. E. Schwartz and R. F. Kooy (2002). "Family MRX9 revisited: further evidence for locus heterogeneity in MRX." Am J Med Genet **112**(1): 17-22.

Wonders, C. P. and S. A. Anderson (2006). "The origin and specification of cortical interneurons." Nat Rev Neurosci **7**(9): 687-696.

Woods, K. S., M. Cundall, J. Turton, K. Rizotti, A. Mehta, R. Palmer, J. Wong, W. K. Chong, M. Al-Zyoud, M. El-Ali, T. Otonkoski, J. P. Martinez-Barbera, P. Q. Thomas, I. C. Robinson, R. Lovell-Badge, K. J. Woodward and M. T. Dattani (2005). "Over- and underdosage of SOX3 is associated with infundibular hypoplasia and hypopituitarism." Am J Hum Genet **76**(5): 833-849.

Yang, J., A. Roy and Y. Zhang (2013). "Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment." Bioinformatics **29**(20): 2588-2595.

Yu, H., A. J. Pask, Y. Hu, G. Shaw and M. B. Renfree (2014). "ARX/Arx is expressed in germ cells during spermatogenesis in both marsupial and mouse." Reproduction **147**(3): 279-289.

Zerem, A., K. Haginoya, D. Lev, L. Blumkin, S. Kivity, I. Linder, C. Shoubridge, E. E. Palmer, M. Field, J. Boyle, D. Chitayat, W. D. Gaillard, E. H. Kossoff, M. Willems, D. Genevieve, F. Tran-Mau-Them, O. Epstein, E. Heyman, S. Dugan, A. Masurel-Paulet, A. Piton, T. Kleefstra, R. Pfundt, R. Sato, A. Tzschach, N. Matsumoto, H. Saitsu, E. Leshinsky-Silver and T. Lerman-Sagie (2016). "The molecular and phenotypic spectrum of IQSEC2-related epilepsy." Epilepsia **57**(11): 1858-1869.

Zu, T., B. Gibbens, N. S. Doty, M. Gomes-Pereira, A. Huguet, M. D. Stone, J. Margolis, M. Peterson, T. W. Markowski, M. A. Ingram, Z. Nan, C. Forster, W. C. Low, B. Schoser, N. V. Somia, H. B. Clark, S. Schmechel, P. B. Bitterman, G. Gourdon, M. S. Swanson, M. Moseley and L. P. Ranum (2011). "Non-ATG-initiated translation directed by microsatellite expansions." Proc Natl Acad Sci U S A **108**(1): 260-265.

Zu, T., Y. Liu, M. Banez-Coronel, T. Reid, O. Pletnikova, J. Lewis, T. M. Miller, M. B. Harms, A. E. Falchook, S. H. Subramony, L. W. Ostrow, J. D. Rothstein, J. C. Troncoso and L. P. Ranum (2013). "RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia." Proc Natl Acad Sci U S A **110**(51): E4968-4977.

Zweier, C., C. Kraus, L. Brueton, T. Cole, F. Degenhardt, H. Engels, G. Gillessen-Kaesbach, L. Graul-Neumann, D. Horn, J. Hoyer, W. Just, A. Rauch, A. Reis, B. Wollnik, M. Zeschnigk, H. J. Ludecke and D. Wieczorek (2013). "A new face of Borjeson-Forssman-Lehmann syndrome? De novo mutations in PHF6 in seven females with a distinct phenotype." J Med Genet **50**(12): 838-847.