ATRX regulates H3.3 incorporation and gene expression at G-rich ancestral pseudoautosomal genes

Michael A. Levy
The University of Western Ontario
Supervisor
Dr. Nathalie Berube
The University of Western Ontario

Graduate Program in Developmental Biology
A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy
© Michael A. Levy 2013

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Molecular Genetics Commons

Recommended Citation
ATRX REGULATES H3.3 INCORPORATION AND GENE EXPRESSION AT G-RICH ANCESTRAL PSEUDOAUTOSOMAL GENES

(Thesis format: Integrated Article)

by

Michael Aaron Levy

Graduate Program in Biochemistry

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

© Michael Aaron Levy 2013
Abstract

Mutations in the human ATRX gene cause the ‘alpha thalassemia mental retardation X linked’ syndrome or can enable cancer progression. ATRX encodes a Swi2/Snf2 chromatin remodeling protein involved in deposition of the histone variant H3.3 at telomeres and pericentromeric heterochromatin. Loss of ATRX leads to genomic instability, mitotic defects, and increased apoptosis in the developing mouse brain. The aim of this study was to determine the role of ATRX in the regulation of gene expression. I identified the ancestral pseudoautosomal region (aPAR) genes as some of the most downregulated genes throughout mouse forebrain development in the absence of ATRX. The pseudoautosomal regions (PARs) are areas of homology between the ends of the otherwise dissimilar X and Y chromosomes, and are exceptional in that they are rich in repetitive sequences and GC content. During the evolutionary divergence between mice and humans, mouse PAR homologs have translocated to various autosomes and are now called ancestral PAR genes. Remarkably, mouse aPAR genes regulated by ATRX are located near telomeres. To investigate the mechanism by which ATRX promotes aPAR gene expression, we focused on the mouse aPAR gene ‘dehydrogenase/reductase (short-chain) X chromosome’ (Dhrsx). Chromatin immunoprecipitation showed that ATRX and histone H3.3 occupy the Dhrsx gene body in a guanine-rich DNA segment predicted to form secondary DNA structures called G-quadruplexes. In the absence of ATRX, I observed a significant decrease in H3.3 levels at Dhrsx and at the other downregulated aPAR genes. Several other epigenetic marks are not altered in and around the Dhrsx gene in the ATRX-null forebrain, and thus cannot provide an explanation for transcriptional dysregulation. However, increased RNA polymerase II occupancy at the ATRX/H3.3/G-rich region of Dhrsx indicates stalling of the polymerase in the absence of ATRX, and suggests that ATRX normally promotes transcriptional elongation. I conclude that ATRX facilitates the passage of the transcription machinery at G-quadruplex forming regions of a gene in a process that involves incorporation of the histone variant H3.3. The identification of this mechanism of gene regulation by ATRX may lead to a better understanding of the consequences of ATRX mutations in human patients.
Keywords

ATRX, H3.3, Dhrsx, pseudoautosomal region, transcription elongation, G-quadruplex, telomere, brain development.
Co-Authorship Statement

I participated in the design and execution of all experiments, performed data analysis, and prepared all written material, with the following exceptions:

In chapter two, RNA for P0.5 microarray analysis was isolated by Deanna Tremblay, and for P17 microarray analysis by Corinna Zogel. Phylogenetic trees in Figure 2-3 and 2-5 were generated by Andrew Fernandes. In section 2.3.6, shATRX HeLa cell lines were created by Kieran Ritchie, and RNA purification and cDNA synthesis were performed by Kristin Kernohan. Some mouse husbandry was conducted by Claudia Seah.

In chapter three, in section 3.3.2, E14 mouse forebrains for RNA-seq analysis were dissected by Ashley Watson. For section 3.3.6, Yan Jiang performed bisulfite treatment on gDNA and did two out of the three H3K27Me3 ChIP reactions. Most of the mouse husbandry and some mouse dissections were done by Yan Jiang.

Nathalie Bérubé assisted with overall experimental design and direction, and in the preparation of manuscripts.
Acknowledgments

This tale grew in the telling, and there were many people who have helped me along the way. First and foremost, I would like to thank my supervisor, Dr. Nathalie Bérubé, for her years of guidance and support. I would also like to thank my advisors, Dr. Gabe DiMattia and Dr. Susan Meakin, for their advice throughout my project, and the other professors at the Victoria Research Laboratories whose guidance I have sought over the years.

I have been fortunate to work alongside some wonderful colleagues and make great friends in the Bérubé lab and throughout the VRL. I wouldn’t have enjoyed my time in grad school near as much, or been able to maintain my sanity without you. To my friends outside of grad school and outside of London, thank you for the opportunities to take some breaks from my work, and for being so understanding when I turned into somewhat of a hermit over the last year or two.
# Table of Contents

Abstract ................................................................................................................................. ii  
Co-Authorship Statement ...................................................................................................... iv  
Acknowledgments ................................................................................................................... v  
Table of Contents .................................................................................................................. vi  
List of Tables .......................................................................................................................... x  
List of Figures ........................................................................................................................ xi  
List of Appendices .................................................................................................................. xiii  
List of Abbreviations ............................................................................................................. xiv  
Chapter 1 ................................................................................................................................ 1  
1 Introduction .......................................................................................................................... 1  
   1.1 General introduction ........................................................................................................ 1  
      1.1.1 The history and discovery of ATRX ................................................................. 1  
   1.2 The ATRX gene and protein ......................................................................................... 3  
      1.2.1 ATRX encodes a SWI2/SNF2 chromatin remodeling protein ......................... 5  
      1.2.2 ATRX is part of a multi-protein complex ......................................................... 6  
      1.2.3 Cellular and genomic localization of ATRX .................................................. 7  
   1.3 ATRX in development and disease ............................................................................ 11  
      1.3.1 ATR-X and related syndromes ........................................................................... 11  
      1.3.2 ATRX, cancer, and alternative lengthening of telomeres ................................. 13  
      1.3.3 ATRX in development ....................................................................................... 15  
   1.4 Cellular functions of ATRX ....................................................................................... 16  
      1.4.1 Mitosis and meiosis ......................................................................................... 16  
      1.4.2 DNA replication ............................................................................................... 16  
      1.4.3 Gene regulation ............................................................................................... 17
Chapter 1: Regulating the Genome

1.5 Regulating the genome

1.5.1 DNA methylation

1.5.2 Chromatin remodeling proteins

1.5.3 The histone variant H3.3 and transcriptional elongation

1.6 Chromosome ends: the pseudoautosomal regions and telomeres

1.6.1 The pseudoautosomal region: ancient regions of homology between the X and Y chromosomes

1.6.2 Telomeres

1.6.3 G-quadruplexes

1.7 Thesis overview

1.8 References

Chapter 2: The SWI/SNF protein ATRX co-regulates pseudoautosomal genes that have translocated to autosomes in the mouse genome

2.1 Introduction

2.2 Materials and Methods

2.2.1 Mouse husbandry

2.2.2 Microarray analysis

2.2.3 Quantitative reverse transcriptase PCR

2.2.4 Bioinformatics analysis of novel ancestral PAR genes

2.2.5 Cell culture and RNA interference

2.2.6 Immunofluorescence

2.3 Results

2.3.1 Effects of ATRX deletion on forebrain gene expression

2.3.2 Ancestral pseudoautosomal genes are downregulated in the ATRX-null mouse forebrain

2.3.3 Verification of gene expression changes

References
2.3.4 Identification of a novel arylsulfatase family mouse homolog .................. 63
2.3.5 Identification of an ASMTL-like gene ........................................... 67
2.3.6 Expression of PAR genes regulated by ATRX in the mouse is unchanged upon depletion of ATRX in non-murine cell lines ......................... 69

2.4 Discussion ..................................................................................... 71
2.5 Supplementary Figures .................................................................. 75
2.6 Supplementary tables ...................................................................... 80
2.7 References .................................................................................... 85

Chapter 3 .......................................................................................... 90

3 ATRX aids transcription elongation through G-rich gene segments in a process involving histone H3.3 incorporation ........................................... 90
3.1 Introduction ................................................................................... 90
3.2 Materials and Methods ................................................................. 92
  3.2.1 Mouse husbandry and genotyping ............................................ 92
  3.2.2 Mouse embryonic fibroblast isolation, culture and viral infection ...... 92
  3.2.3 RNA isolation and transcriptional assays .................................... 93
  3.2.4 RNA dot blots ........................................................................... 93
  3.2.5 Bisulfite mutagenesis and sequencing ....................................... 94
  3.2.6 Chromatin immunoprecipitation .............................................. 94
  3.2.7 Next generation sequencing and analysis .................................... 95
3.3 Results ............................................................................................ 96
  3.3.1 ATRX is required for the normal expression of Dhrsx and flanking non-coding RNAs in mouse cells ....................................................... 96
  3.3.2 ATRX deficiency in the mouse forebrain or in MEFs does not affect TERRA levels ................................................................. 98
  3.3.3 ATRX and H3.3 are enriched within the gene body of Dhrsx .......... 101
  3.3.4 Reduced levels of H3.3 at Dhrsx and other aPAR genes in the absence of ATRX correlates with decreased gene expression ....................... 101
3.3.5 H3.3 is enriched at telomeres and towards the ends of chromosomes, and depleted from telomeres in the absence of ATRX.................. 106

3.3.6 DNA methylation and histone modifications are not altered at Dhrsx in the absence of ATRX.................................................. 109

3.3.7 Increased RNA polymerase II occupancy at the ATRX/H3.3/G-rich region of Dhrsx in the absence of ATRX................................. 113

3.4 Discussion.................................................................................. 115

3.5 Supplementary figures ................................................................ 120

3.6 Supplementary tables .................................................................. 124

3.7 References.................................................................................... 125

Chapter 4......................................................................................... 132

4 General Discussion and Future Directions........................................ 132

4.1 Thesis summary ........................................................................... 132

4.2 A role for ATRX in gene regulation............................................... 133

4.3 ATRX and structured DNA .......................................................... 136

4.4 Histone chaperones and the genomic localization of histone H3.3 .... 138

4.5 H3.3 the and regulation of aPAR genes........................................ 140

4.6 A model for gene regulation by ATRX ........................................ 142

4.7 Concluding remarks...................................................................... 144

4.8 References.................................................................................... 145

Appendices....................................................................................... 152
List of Tables

Table 2-1: Downregulated genes in the ATRX-null forebrain at E13.5 and P0.5.............. 60

Table 2-2 (supplementary): Conditions for quantitative real-time PCR............................ 80

Table 2-3 (supplementary): Significantly misregulated GO categories. ......................... 81

Table 2-4 (supplementary): Pairwise comparisons of arylsulfatase family members.......... 84

Table 3-1 (supplementary): Primer sequences and annealing temperatures.................... 124
List of Figures

Figure 1-1: Protein domains and binding partners of ATRX. .......................................................... 4
Figure 1-2: G-quadruplexes. ............................................................................................................. 9
Figure 1-3: Gene regulation by ATRX. .............................................................................................. 21
Figure 1-4: Evolution of the pseudoautosomal region..................................................................... 27
Figure 2-1: Evolution of PAR genes in humans and mice................................................................. 52
Figure 2-2: Relative expression of ancestral PAR genes in ATRX-null mouse forebrains.... 62
Figure 2-3: Phylogenetic tree of arylsulfatase proteins. ................................................................. 64
Figure 2-4: Arsd/e transcriptional downregulation is recapitulated in ATRX-depleted cells. 66
Figure 2-5: Phylogenetic tree of ASMTL proteins................................................................. 68
Figure 2-6: PAR gene expression is unchanged in human and bovine cells. ......................... 70
Figure 2-7 (supplementary): Summary of microarray results.............................................. 77
Figure 2-8 (supplementary): Amino acid alignment of a small portion of ARSD/E between multiple species.......................................................................................................................... 78
Figure 2-9 (supplementary): Amino acid alignment of the N terminal of ASMTL between multiple species.......................................................................................................................... 79
Figure 3-1: Expression of non-coding RNAs flanking Dhrsx is decreased in the Atrx-null mouse forebrain. ....................................................................................................................................... 97
Figure 3-2: TERRA expression is not altered in forebrain tissue or mouse embryonic fibroblasts lacking ATRX............................................................... 100
Figure 3-3: ATRX and H3.3 enrichment at Dhrsx. ...................................................................... 103
Figure 3-4: ATRX and H3.3 enrichment at additional aPAR genes.......................... 105

Figure 3-5: H3.3 distribution along chromosomes and at telomeres. ..................... 107

Figure 3-6: DNA methylation at Dhrsx CpG islands is not changed in the P0.5 ATRX-null mouse forebrain. ......................................................................................................................... 110

Figure 3-7: Histone modifications are unchanged in the P17 ATRX-null mouse forebrain. 112

Figure 3-8: RNA Pol-II stalling in the absence of ATRX. ........................................... 114

Figure 3-9: A model for the regulation of transcription by ATRX................................. 119

Figure 3-10 (supplementary): Adenoviral treatment of Atrx-floxed MEFs..................... 120

Figure 3-11 (supplementary): Comparison of H3.3 enrichment at Dhrsx in the mouse brain at P0.5 and P17. ......................................................................................................................... 121

Figure 3-12 (supplementary): ATRX and H3.3 ChIP-Seq input tracks ......................... 122

Figure 3-13 (supplementary): DNA methylation at the Csf2ra promoter is not altered in the neonatal ATRX-null mouse forebrain. ..................................................................................................... 123
List of Appendices

Appendix A: Permission to reproduce previously published work .............................................. 152

Appendix B: Statement of permission for the use of animals for experimental research............. 152
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADD</td>
<td>ATRX-DNMT3A/B-DNMT3L</td>
</tr>
<tr>
<td>µg</td>
<td>micrograms</td>
</tr>
<tr>
<td>µL</td>
<td>microlitres</td>
</tr>
<tr>
<td>32P</td>
<td>Phosphorus-32</td>
</tr>
<tr>
<td>Ad</td>
<td>adenovirus</td>
</tr>
<tr>
<td>ALT</td>
<td>alternative lengthening of telomeres</td>
</tr>
<tr>
<td>AMELX</td>
<td>amelogenin, X-linked (gene)</td>
</tr>
<tr>
<td>APB</td>
<td>ALT-associated PML nuclear body</td>
</tr>
<tr>
<td>aPAR</td>
<td>ancestral pseudoautosomal region</td>
</tr>
<tr>
<td>AR</td>
<td>androgen receptor</td>
</tr>
<tr>
<td>ARSD/E/F/H</td>
<td>arylsulfatase d/e/f/g</td>
</tr>
<tr>
<td>ASMTL</td>
<td>acetylserotonin O-methyltransferase-like</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine triphosphate</td>
</tr>
<tr>
<td>ATR-16</td>
<td>alpha-thalasemia mental retardation, chromosome 16</td>
</tr>
<tr>
<td>ATRX</td>
<td>alpha-thalasemia mental retardation, X linked protein</td>
</tr>
<tr>
<td>ATR-X</td>
<td>alpha-thalasemia mental retardation, X linked protein syndrome</td>
</tr>
<tr>
<td>ATRXt</td>
<td>alpha-thalasemia mental retardation, X linked protein, truncated isoform</td>
</tr>
<tr>
<td>B-actin</td>
<td>Beta-actin</td>
</tr>
<tr>
<td>BLASTn</td>
<td>Basic Local Alignment Search, nucleotide</td>
</tr>
<tr>
<td>BLAT</td>
<td>BLAST-Like Alignment Tool</td>
</tr>
<tr>
<td>BLM</td>
<td>Bloom's syndrome helicase</td>
</tr>
<tr>
<td>bp</td>
<td>base pairs</td>
</tr>
<tr>
<td>BRCA1</td>
<td>breast cancer type 1</td>
</tr>
<tr>
<td>BRG1</td>
<td>Brahma-related Gene 1</td>
</tr>
<tr>
<td>BRIP1</td>
<td>BRCA1 interacting protein</td>
</tr>
<tr>
<td>BRM</td>
<td>Brahma</td>
</tr>
<tr>
<td>C</td>
<td>celcius</td>
</tr>
<tr>
<td>CD99</td>
<td>CD99 antigen</td>
</tr>
<tr>
<td>cDNA</td>
<td>complementary DNA</td>
</tr>
</tbody>
</table>
CenH3  centromeric histone H3
CENP-A  centromere protein A
CHD  chromodomain, helicase, DNA binding
ChIP  chromatin immunoprecipitation
ChIP-seq  chromatin immunoprecipitation sequencing
Ci  Curie
CMV  cytomegalovirus
CO2  carbon dioxide
Cre  cyclization recombinase
CRLF2  cytokine receptor-like factor 2
cRNA  complementary RNA
Csf2ra  colony stimulating factor 2 receptor, alpha
CTCF  CCCTC-binding factor
CTD  C-terminal domain (or RNA PolII)
dATRX  Drosophila ATRX
DAXX  death-domain associated protein
DHRXSX  dehydrogenase/reductase short-chain dehydrogenase/reductase family, X chromosome
DHX36  DEAH (Asp-Glu-Ala-His) box polypeptide 36
DMEM  Dulbecco's Modified Eagle Medium
DNA  deoxyribonucleic acid
DNMT  DNA methyltransferase
dsDNA  double stranded DNA
DYZ2  DNA Y-chromosome Z repeats 2
E13.5  embryonic day 13.5
ERCC6  excision repair cross-complementing rodent repair deficiency, complementation group 6
ESC  embryonic stem cells
EST  expressed sequence tag
EYFP  enhanced yellow fluorescent protein
EZH2  enhancer of zest
Frog1  forkhead box G1
Gapdh   glyceraldehyde-3-phosphate dehydrogenase
gDNA    genomic DNA
GFP     green fluorescence protein
GO      gene ontology
GQN1    G quartet nuclease 1
G-rich  guanine rich
GTPBP6  GTP binding protein 6
h       hour
H3K27Me3 histone 3 lysine 27 trimethylation
H3K36   histone 3 lysing 36
H3K4Me0 histone 3 lysine 4 unmethylated
H3K4Me3 histone 3 lysine 4 trimethylation
H4Ac    histone 4 acetylation
Hb h    Hemoglobin h
HIRA    histone cell cycle regulator
HP1     heterochromatin protein 1
IgG     immunoglobulin G
IL3RA   interleukin 3 receptor, alpha (low affinity)
INO80   inositol requiring 80
IP      immunoprecipitation
ISWI    imitation switch
ITS     interstitial telomere sequences
kb      kilobases
kD      kilodaltons
LiCl    lithium chloride
loxP    locus of X-over P1
MAF     musculoaponeurotic fibrosarcoma
MEF     mouse embryonic fibroblasts
mH2A    macro H2A
mm9/mm10 mouse genome version 9/10
mmol    millimole
NCBI    National Center for Biotechnology Information
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>V</td>
<td>volts</td>
</tr>
<tr>
<td>WAG</td>
<td>Whelan And Goldman</td>
</tr>
<tr>
<td>WRN</td>
<td>Werner syndrome helicase</td>
</tr>
<tr>
<td>XCI</td>
<td>X chromosome inactivation</td>
</tr>
<tr>
<td>XH2</td>
<td>X-linked helicase 2</td>
</tr>
<tr>
<td>XNP</td>
<td>X-linked nuclear protein</td>
</tr>
</tbody>
</table>
Chapter 1

1 Introduction

1.1 General introduction

Studying the interplay between proteins linked to disease and the genes they regulate is essential to understanding disease mechanisms, with the ultimate goal of developing future treatments. The gene *ATRX* has been linked to two major disease conditions in humans: its namesake that led to its discovery, alpha-thalassemia mental retardation X-linked syndrome (ATR-X), and more recently, brain and pancreatic cancers. *ATRX* is involved in multiple cellular processes: mitosis and meiosis, DNA replication, nucleosome remodeling, and gene expression. It is a chromatin remodeling protein initially found associated with heterochromatin, which suggests that *ATRX* might be a negative regulator of gene expression. More recent studies have shown *ATRX* also localizes to guanine-rich and repetitive regions of the genome, and the work presented in this thesis aims to identify the mechanism by which *ATRX* acts as a positive regulator of expression for a group of genes connected to these unique regions of the genome.

1.1.1 The history and discovery of *ATRX*

“Hemoglobin H disease and mental retardation: a new syndrome or a remarkable coincidence?” This was the question asked by Weatherall *et al* in 1981 when they published a report describing three families, each having a son with intellectual disability and non-Mendelian (not inherited) haemoglobin H (Hb H) disease (Weatherall *et al.*, 1981). Hb H disease is caused when decreased α-globin expression leads to an excess of β-globin. The excess β-globin forms β₄ tetramers (called Hb H) instead of the usual α₂/β₂ tetramers. Hb H reduces the oxygen carrying capacity of blood cells, causing damage and forming intracellular precipitates called Hb H inclusions which are visible under the microscope (Chui *et al.*, 2003). The cause of this new Hb H/intellectual disability syndrome was found to be mutations within the α-globin gene cluster near the telomere of chromosome 16, explaining the cause of the α-thalassemia and suggesting that this region may also play a role in mental development (Weatherall *et al.*, 1981).
Nine years later two studies published together described additional cases and features of this same syndrome. One study described patients with large deletions of the α-globin cluster on chromosome 16, and this syndrome was termed ‘α-thalassemia mental retardation, chromosome 16’ (ATR-16). These patients had more severe Hb H disease and less severe intellectual disability (Wilkie et al., 1990a). The second paper described the more mysterious situation in which no mutations or deletions in the α-globin cluster could be identified. These patients had milder Hb H disease but more severe intellectual disability, along with a consistent spectrum of other symptoms such as microcephaly, genital abnormalities, and a distinct facial appearance. It was suggested that a trans-acting factor may be responsible for this syndrome and that it may be X linked as all the described cases were genotypically male (Wilkie et al., 1990b). Subsequent cases supported the X-linkage hypothesis (Cole et al., 1991; Donnai et al., 1991; Harvey et al., 1990) and the syndrome came to be called ‘α-thalassemia mental retardation, X-linked’ (ATR-X).

The X linked nature of the syndrome was confirmed with linkage analysis placing the disease causing locus at Xq12-q21.31. In addition to the male patients, phenotypically normal female carriers were identified by the presence of rare Hb H inclusions and an extremely skewed pattern of X chromosome inactivation (XCI), whereby the mutated X chromosome was preferentially inactivated (Gibbons et al., 1992). The location of the gene responsible for this condition was subsequently narrowed to chromosome Xq13 and the was cloned in humans (XH2) (Stayton et al., 1994) and mice (Xnp) (Gecz et al., 1994). XH2, a SNF2 helicase gene located at Xq13.3 was ultimately confirmed as being responsible for the ATR-X syndrome by the identification of several mutations in this gene in ATR-X patients (Gibbons et al., 1995), and XH2 was therefore renamed ATRX.

The initial 15 years of study into the ATR-X syndrome therefore identified an X-linked transcription factor capable of affecting gene expression at the α-globin locus near the telomere on chromosome 16, and showed that while some of the developmental symptoms of the ATR-X syndrome arise from disruptions on chromosome 16 (based on the moderate mental retardation seen in ATR-16), effects elsewhere in the genome are also likely involved.
1.2 The ATRX gene and protein

The ATRX gene is located on the X chromosome at Xq13.3 (Gibbons et al., 1995) and undergoes X chromosome inactivation (Gibbons et al., 1992; Stayton et al., 1994). ATRX consists of 36 exons spanning 300 kb (Picketts et al., 1996). Two full length transcripts (NM_000489 and NM_138270) generated by alternative splicing of exon 6 differ by 117 bp and are approximately 11 kb in length. They produce proteins of either 280 (NP_000480) or 265 kD (NP_612114) (Villard et al., 1997). A truncated form of ATRX (Bérubé et al., 2000; McDowell et al., 1999) arises from a failure to splice intron 11 from the primary transcript and use of a proximal intronic poly(A) signal, generating a 200 kD protein from a 7 kb transcript and is called ATRXt (Garrick et al., 2004) (Figure 1-1A).

ATRX is ubiquitously expressed (Gecz et al., 1994; Stayton et al., 1994), while ATRXt shows some variability in tissue expression. In particular, while ATRX is highly expressed in human fetal brain ATRXt is not. In adults, ATRX is expressed at lower levels in the brain but is high in adult skeletal muscle, heart, and somewhat higher in the pancreas. In general, a relationship was found in which tissues expressing higher levels of full length ATRX had lower levels of ATRXt (Garrick et al., 2004).
Figure 1-1: Protein domains and binding partners of ATRX.

(A) ATRX contains two highly conserved domains, the ADD and Swi2/Snf2 domains. The ADD domain is subdivided into three smaller structural units. The bars and labels above ATRX indicate domains that target ATRX to the specified factors. Nuc, nucleus. ATRXt is a truncated form of ATRX lacking the Swi2/Snf2 domain. Numbers represent amino acid positions in the human ATRX protein.

(B) Binding partners of ATRX. ATRX directly binds DAXX and HP1. Targeting to PCH is by direct recognition of H3K9Me3/H3K4Me0 and enhanced indirectly through HP1. DAXX facilitates ATRX localization to promyelocytic leukemia nuclear bodies (PML-NBs). ATRX binds H3.3-H4 indirectly through DAXX.
1.2.1 ATRX encodes a SWI2/SNF2 chromatin remodeling protein

ATRX contains several protein domains, most notably two highly conserved regions (Picketts et al., 1998): the ATRX-DNMT3-DNMT3L (ADD) domain (Aapola et al., 2000) and the SWI2/SNF2 helicase domain (Picketts et al., 1996; Stayton et al., 1994) (Figure 1-1A). SWItching defective/Sucrose NonFermenter (SWI/SNF) complexes are powered by an ATPase catalytic domain (SNF2/SWI2 in yeast, BRG1/SMARCA4 or BRM/SMARCA2 in humans) and can remodel nucleosomes to regulate gene expression (Cote et al., 1994; Liu et al., 2011; Tang et al., 2010). DNA assembled into a nucleosome generally displays a characteristic pattern of DNase I digestion, with the enzyme cutting only the most exposed bases to generate a ladder of 10-11 bp intervals (Noll, 1974), and in this conformation is said to be “rotationally phased”. When exposed to the SWI/SNF complex this digestion pattern is largely abolished, demonstrating that the DNA-histone interactions have been disrupted and the DNA phasing around the nucleosome has been randomized (Cote et al., 1998). The ATRX complex is able to moderately enhance this digestion pattern in an ATP-dependent manner, in particular at the DNA entry site into the nucleosome, demonstrating that it has the ability to alter DNA-histone interactions but not to randomize DNA phasing (Xue et al., 2003). DNA translocases are proteins that use ATP hydrolysis to move along the DNA double helix. One way to demonstrate this activity is to form a triple-helical DNA structure then assay whether the protein in question is able to dissociate the third DNA strand as it moves down the DNA helix. ATRX exhibits this so-called triple-helix displacement activity but not double-helix displacement activity, demonstrating that ATRX is able to translocate across but not unwind double stranded DNA (dsDNA), and is therefore a translocase but not a helicase (Xue et al., 2003). In addition to translocating itself across naked dsDNA, ATRX is able to assemble and mobilize nucleosomes (Lewis et al., 2010). Taken together, these studies demonstrate that ATRX is able to remodel chromatin by altering the interactions between DNA and histones, allowing it to deposit histones and reposition nucleosomes along DNA.

Approximately 30% of ATR-X syndrome patient mutations are in the conserved ATRX SWI2/SNF2 domain (Gibbons et al., 2008). When 21 different mutations in this domain
were tested, most caused decreased protein stability, but a few mutations specifically disrupted the ATPase activity, helicase activity, or both. This demonstrates that patient mutations in this region can cause the ATR-X syndrome by disrupting either the structure or function of ATRX (Mitson et al., 2011). ATRXt lacks the SWI2/SNF2 chromatin remodeling domain, suggesting that it likely has a different function from full length ATRX (Garrick et al., 2004) (Figure 1-1A).

The ADD domain is named as such because of its significant similarity with de novo methyltransferases DNMT3A/B and DNMT3L (Aapola et al., 2000). It encompasses several smaller functional units: a C2C2-GATA-like zinc finger (Argentaro et al., 2007; Picketts et al., 1998; Villard et al., 1997), a C4C4-plant homeodomain (PHD)-like zinc-finger (Gibbons et al., 1997), and an α-helix domain (Argentaro et al., 2007; Picketts et al., 1998) (Figure 1-1A). The ADD domain can bind naked dsDNA (Cardoso et al., 2000), but more importantly it targets ATRX to heterochromatin by recognizing the double histone marks of H3K9Me3/H3K4Me0 (Dhayalan et al., 2011; Eustermann et al., 2011; Iwase et al., 2011).

1.2.2 ATRX is part of a multi-protein complex

Mammalian SWI/SNF complexes contain nine to 12 proteins with subunit composition dependent on cell type and specific complex function (Euskirchen et al., 2012; Wang et al., 1996). While ATRX is a SWI2/SNF2-like protein, it is not a member of a typical SWI/SNF complex. It is however found in complexes with several other proteins, the most well studied being heterochromatin protein 1 (HP1), death-domain associated protein (DAXX), and the histone variant H3.3 (Figure 1-1B). HP1 was the first interacting partner of ATRX to be identified (Bérubé et al., 2000; Le Douarin et al., 1996). HP1 is able to directly bind ATRX and it enhances the localization of ATRX to pericentromeric heterochromatin (PCH) by recognizing additional H3K9Me3 residues (Eustermann et al., 2011; Kourmouli et al., 2005). The chromatin remodeling ability of ATRX was first recognized through analysis of a protein complex containing ATRX and DAXX (Xue et al., 2003), a protein previously identified as both a suppressor and activator of apoptosis (Salomoni and Khelifi, 2006). DAXX plays a dual role with ATRX: it targets ATRX to promyelocytic leukemia nuclear bodies (PML-NBs) (Ishov et
al., 2004; Tang et al., 2004), and acts with ATRX as a histone chaperone complex to deposit the histone variant H3.3 at PCH (Drane et al., 2010) and telomeres (Lewis et al., 2010). DAXX binds H3.3 through the histone’s globular structure (Lewis et al., 2010), and while ATRX recognizes H3.3’s tail domain, this secondary interaction is dispensable for targeting H3.3 to telomeres (Wong et al., 2010). As opposed to the PML targeting by DAXX, localization of the ATRX-DAXX-H3.3 complex to telomeres is directed by ATRX itself (Goldberg et al., 2010; Lewis et al., 2010), demonstrating that the combination of ATRX and DAXX allows chromatin remodeling at multiple cellular and genomic regions. The interaction of ATRX with H3.3 is an unexpected finding given that ATRX usually associates with repressive chromatin domains while H3.3 is traditionally a marker of active transcription (Ahmad and Henikoff, 2002; Schwartz and Ahmad, 2005).

1.2.3 Cellular and genomic localization of ATRX

ATRX is a nuclear protein due to a centrally located nuclear localization signal domain (Berube et al., 2007). Within the nucleus, ATRX localizes to specific cellular and genomic regions where it is found with its various protein partners. ATRX frequently localizes to heterochromatin with HP1, including PCH (Kourmouli et al., 2005; McDowell et al., 1999), telomeres (Wong et al., 2010), and the condensed chromosomes during mitosis (Bérubé et al., 2000). ATRX is found at telomeres in mouse ESCs (Law et al., 2010; Wong et al., 2010), mouse neuroprogenitors (Watson et al., 2013) and human erythroid cells (Law et al., 2010). It is responsible for depositing H3.3 at both PCH and telomeres (Drane et al., 2010; Goldberg et al., 2010; Wong et al., 2010). Interestingly, despite the localization of ATRX at telomeres in the terminally differentiated human erythroid cells (Law et al., 2010), it is lost from telomeres of mouse ESCs after neuronal differentiation (Wong et al., 2010). ATRX is found on the inactive X chromosome after initiation of inactivation (Baumann and De La Fuente, 2008) and along the Y chromosome (which is largely heterochromatic) in spermatogonial cells (Baumann et al., 2008). Targeting of ATRX to heterochromatin is mediated by binding of the ADD domain to the heterochromatic histone mark H3K9 tri-methylation (and to a lesser extent di-methylation) in the absence of H3K4 methylation and this targeting is enhanced by
HP1 which binds ATRX and recognizes additional H3K9Me3 marks (Dhayalan et al., 2011; Eustermann et al., 2011; Iwase et al., 2011). However, ATRX can also be found at euchromatin. It is found in decondensed chromatin of growing oocytes (De La Fuente et al., 2004), and *Drosophila* ATRX (dATRX), which lacks the ADD domain, localizes to sites of active transcription (Schneiderman et al., 2009).

ATRX also has a preference for repetitive G-rich sequences including ribosomal DNA (rDNA) repeats (Gibbons et al., 2000; Law et al., 2010; McDowell et al., 1999), G-rich tandem repeats, the G-rich strand of telomeres (comprised of TTAGGG repeats), and half of all ATRX binding sites in humans and mice overlap with CpG islands (Law et al., 2010). ATRX is enriched at subtelomeres in humans but not mice due to the presence of high GC levels in this region on human but not mouse chromosomes, demonstrating that ATRX targets high GC regions rather than subtelomeres in particular (Law et al., 2010). Enrichment of ATRX at these repetitive G-rich repeats may be mediated by the ability of ATRX to bind G-quadruplexes (Law et al., 2010). G-quadruplexes are short sequences containing four guanine triplicates. They form a four stranded secondary structure upon DNA denaturation during replication or transcription and are particularly enriched within telomeres (Biffi et al., 2013; Duquette et al., 2004) (Figure 1-2) (See section 1.6.3 for more details).
Figure 1-2: G-quadruplexes.

(A) Typical pattern of potential G-quadruplex forming sequences (top), and the G-rich telomeric strand sequence, which has a propensity to form G-quadruplexes (bottom). (B) G-quadruplexes forming during DNA replication or transcription can inhibit passage of the DNA polymerase (blue triangles) or RNA polymerase (green triangle), respectively. Newly synthesized DNA is indicated in purple and newly synthesized RNA is in red.
PML-NBs are nuclear structures associated with up to 100 proteins. They are implicated in diverse cellular functions such as protein sequestration, transcription activation and repression, DNA repair and recombination, cancer, and response to viral infection (reviewed in (Lallemand-Breitenbach and de The, 2010)). PML-NBs form spherical, highly organized structures during interphase (Lang et al., 2010). The outer layers consist of the PML protein itself and ‘nuclear antigen speckled 100 kDa’ (Sp100) (Lang et al., 2010; Luciani et al., 2006). ‘Small ubiquitin-like modifier’ (SUMO) is found mostly in these outer layers (Lang et al., 2010; Luciani et al., 2006) and is responsible for sumoylation of PML which is necessary for the formation of the spherical PML-NB organization (Ishov et al., 1999; Zhong et al., 2000). Major components of the inner layers include DAXX, ATRX, and HP1 (Lang et al., 2010; Luciani et al., 2006).

Centromeric DNA is associated with PML-NBs during the G2 phase of the cell cycle (Everett et al., 1999; Luciani et al., 2006), and telomeres are associated with PML-NBs during S phase (Chang et al., 2013).

ATRX is targeted to PML-NBs largely by DAXX (Ishov et al., 2004; Tang et al., 2004; Xue et al., 2003), but also through a C terminal PML-targeting domain (Berube et al., 2007). The localization of ATRX to PML-NBs appears to be important for the ATRX-mediated deposition of H3.3 onto telomeres. Deposition occurs during S phase which is when both ATRX and telomeres have been shown to associate with PML-NBs. In addition, depletion of PML causes a reduction of telomeric ATRX and H3.3 (Chang et al., 2013; Wong et al., 2010). Movement of ATRX between cellular regions may depend on phosphorylation, as unphosphorylated ATRX displays a speckled nuclear pattern during interphase (presumably while at PML-NBs) then is phosphorylated and found at condensed chromatin during mitosis (Bérubé et al., 2000).

ATRXt displays different cellular localization than full-length ATRX; it co-localizes with ATRX at PCH but not at PML-NBs (Garrick et al., 2004), likely on account of a loss of the DAXX interacting domain and/or the C terminal PML targeting domain in ATRXt. Disruptions to the N terminal of ATRX, such as ADD domain mutations seen in some ATR-X patients, leads to a more diffuse cellular localization (Cardoso et al., 2000). ATRX localization differs between male and female germ cells; it is associated with
chromosomes throughout meiosis in oocytes but not in spermatocytes (Baumann et al., 2008).

1.3 ATRX in development and disease

Mutations in the *ATRX* gene were initially identified as the cause of a condition causing both α-thalassemia and intellectual disability (Gibbons et al., 1995; Weatherall et al., 1981) but have since been implicated in several human conditions.

1.3.1 ATR-X and related syndromes

Mutations in *ATRX* result in a diversity of clinical abnormalities collectively known as the ATR-X syndrome (clinical phenotype reviewed in (Gibbons, 2006; Gibbons and Higgs, 2000)). A moderate to profound cognitive deficit is the most common feature, with 95% of patients exhibiting profound intellectual disability. Related common neurological symptoms include delayed and limited expressive language skills, microcephaly, and seizures. A distinctive facial appearance is seen in almost all cases and includes telecanthus (wide-set eyes), epicanthic folds (skin folds of the upper eyelid covering the inner corner of the eye), flat nasal bridge and small triangular upturned nose, and a tented upper lip. A wide range of relatively mild musculoskeletal abnormalities include hypotonia (low muscle tone), fixed flexion deformity (inability to properly straighten or bend) of the fingers, short, bent or conjoined fingers, and flat or clubbed feet. Short stature is seen in 65% of cases. Gastrointestinal difficulties such as recurrent vomiting, regurgitation, and gut dysmotility are seen in 75% of cases. Genital abnormalities are seen in 80% of patients and range from undescended testes to pseudohermaphrodisism. Lastly, α-thalassemia (decreased α-globin production) is seen in nearly 90% of patients and is diagnosed by the presence of Hb H inclusions.

At the molecular level ATR-X patients show aberrant DNA methylation at repetitive sequences: loss of rDNA methylation, greatly increased methylation of the heterochromatic Y-chromosome DYZ2 repeats, and slightly altered methylation at subtelomeric sequences. Interestingly, no changes in methylation were found at other repetitive sequences including telomeres (Gibbons et al., 2000) despite ATRX being found at both rDNA and telomeres. Approximately 150 mutations and 200 cases of
ATR-X syndrome have been identified. Most mutations are found within either the ADD or helicase-SWI2/SNF2 domain, reiterating the importance of these conserved regions (Gibbons, 2006; Gibbons et al., 2008).

Given the X-linked nature of ATRX, the syndrome generally affects males while females exhibit little or no symptoms due to extremely skewed inactivation of the X chromosome containing the mutated ATRX allele (Gibbons et al., 1992; Yntema et al., 2002). Mutations inherited from carrier mothers are the cause of 85% of ATR-X syndrome cases (Badens et al., 2006a; Gibbons and Higgs, 2000). If the mutated ATRX allele is not properly inactivated then female carriers will exhibit ATR-X syndrome phenotypes (Badens et al., 2006b; Wada et al., 2005). The exact cause of skewed XCI in female carriers is unclear. When Atrx is deleted from various types of mouse embryonic progenitor cells XCI skewing occurs, while postnatal deletion (even in replicative cell types) does not cause XCI, demonstrating that in progenitor cells, those expressing wild-type ATRX may outcompete mutant cells (Muers et al., 2007). Upon mouse embryonic deletion of Atrx, skewed XCI appears only after the earliest stages of development (Muers et al., 2007), and ATRX is associated with the inactive X chromosome only after onset of random X inactivation, suggesting a role for ATRX in maintenance but not initiation of XCI (Baumann and De La Fuente, 2008). Extraembryonic tissue exhibits imprinted XCI of the paternal X chromosome (Takagi and Sasaki, 1975). ATRX is associated with the paternally inactive mouse X chromosome in this tissue (Baumann and De La Fuente, 2008). When Atrx is deleted from early mouse embryos in a parent-of-origin manner so that only the mutated (maternal) Atrx should be active in these tissues, some embryos are able to escape this imprinted XCI allowing expression of wild-type Atrx and rescue of embryonic development (Garrick et al., 2006). It has been proposed that reactivation of the paternal allele could be due to either incomplete initial paternal inactivation allowing the few remaining wild-type cells to overtake the population (Garrick et al., 2006), or that paternal XCI is not maintained due to loss of ATRX, given the potential role for ATRX in maintaining XCI (Baumann and De La Fuente, 2008). Taken together, these studies show that ATRX may be playing a dual role in XCI: expression of the mutated Atrx allele may lead to both a decrease in cell viability and a failure to maintain the XCI status of the given cell.
ATRX mutations that lead to intellectual disability in the absence of α-thalassemia have been given several syndrome names (e.g. Juberg-Marsidi (Villard et al., 1996), Carpenter-Waziri (Abidi et al., 1999), Smith-Fineman-Myers (Villard et al., 2000), Chudley-Lowry (Abidi et al., 2005)) and are collectively known as ‘mental retardation-hypotonic facies syndrome, X-linked’ (OMIM 309580). Alternatively, somatic mutations leading to defects in erythropoiesis in the absence of cognitive or other symptoms cause α-thalassemia myelodysplasia syndrome (Gibbons et al., 2003; Steensma et al., 2005).

Lastly, ATRX overexpression was recently identified in a Drosophila model of Huntington’s disease, a neurodegenerative disorder caused by mutations in the Huntington gene. Increased levels of ATRX led to larger PML-NBs and increased condensation of PCH (Lee et al., 2012).

1.3.2 ATRX, cancer, and alternative lengthening of telomeres

Over the last few years, genomic sequencing and immunohistochemistry have identified ATRX mutations in pancreatic neuroendocrine tumors (de Wilde et al., 2012; Heaphy et al., 2011; Jiao et al., 2011) and brain cancer (Cheung Nv and et al., 2012; Heaphy et al., 2011; Jiao et al., 2012; Khuong-Quang et al., 2012; Molenaar et al., 2012; Schwartzentruber et al., 2012). Tumorigenic cells lacking ATRX expression typically exhibit alternative lengthening of telomeres (ALT) (Cheung Nv and et al., 2012; de Wilde et al., 2012; Heaphy et al., 2011; Jiao et al., 2012; Schwartzentruber et al., 2012), a recombination process that maintains telomere length and proliferative capacity in the absence of the telomerase enzyme (reviewed in (Cesare and Reddel, 2010)). The cause of the excessive recombination leading to ALT is uncertain, but seems to require proteins involved in recombination (e.g. the meiotic recombination complex (Jiang et al., 2005)) along with the loss of proteins involved in telomere maintenance (e.g. the Werner Syndrome protein WRN, which is required for proper replication of telomeres (Crabbe et al., 2004; Laud et al., 2005)). It has been hypothesized that ATRX’s association with telomeric G-quadruplexes may assist in telomere replication, and that loss of ATRX could therefore lead to stalled replication forks which trigger DNA damage and homologous recombination (Clynes and Gibbons, 2013).
Loss of ATRX is frequently associated with ALT in cell lines (Bower et al., 2012; Lovejoy et al., 2012), but artificial depletion of ATRX does not itself lead to ALT (Lovejoy et al., 2012), suggesting that ATRX helps repress ALT and that additional events must occur for its initiation. One of the additional events may be loss of p53, as p53 can inhibit telomere recombination (Razak et al., 2004) and ALT occurs more readily in p53-null cells (Laud et al., 2005). Correspondingly, many ATRX-null ALT-positive tumors also have p53 mutations (Jiao et al., 2012; Kannan et al., 2012). Therefore, loss of ATRX may cause telomere replication fork stalling leading to the telomeric DNA damage seen in ATRX-null muscle (Huh et al., 2012) and brain (Watson et al., 2013), while the addition of a p53 mutation may be needed to activate ALT as seen in ATRX-null tumors. Mutations in DAXX and H3.3 are also associated with ATRX-null and ALT-positive tumors, further demonstrating a role for the ATRX-DAXX-H3.3 axis in telomere maintenance (Schwartzentruber et al., 2012). Interestingly, cells that exhibit ALT have larger than normal PML-NBs, called ALT-associated PML-NBs (APBs) (Chung et al., 2012). These enlarged PML-NBs may play a role in ALT by acting as sites of telomere accumulation allowing for increased recombination (Draskovic et al., 2009). APBs can be artificially induced by treating cells with the viral protein ICP0 (Draskovic et al., 2009), a protein that can inhibit DAXX function (Lukashchuk and Everett, 2010) and H3.3 deposition (Newhart et al., 2012). DAXX normally targets ATRX to PML-NBs (Ishov et al., 2004; Tang et al., 2004), where ATRX and DAXX then associate with telomeres to deposit H3.3 to maintain telomere integrity (Chang et al., 2013; Delbarre et al., 2012). These studies reinforce the idea that the ATRX-DAXX-H3.3 axis is essential to telomere maintenance, and that PML-NBs are a key mediator of this process.

Mutations in DAXX have been found in pancreatic neuroendocrine tumors (Heaphy et al., 2011; Jiao et al., 2012), and mutations in either DAXX or H3.3 have been found in brain cancers (Khuong-Quang et al., 2012; Liu et al., 2012; Schwartzentruber et al., 2012). No ATRX mutations or disruptions of ATRX expression were found in gastric, colorectal or prostate cancers (Je et al., 2012), suggesting that ATRX may be a required tumor suppressor in some tissues but not others.
1.3.3 ATRX in development

Given the myriad of disease states caused by ATRX mutations, several mouse models have been used to identify the role ATRX plays in normal development. It was demonstrated early on that Atrx has widespread expression during early embryogenesis but a more restricted expression pattern at later timepoints (Gecz et al., 1994; Stayton et al., 1994). By E13, expression in the brain is mostly restricted to the telencephalon (primitive forebrain), and in newborn mice is highest in the olfactory bulb and hippocampus, regions of ongoing adult neurogenesis, with lower levels in the cortex. This suggested a role for ATRX in early development and neurogenesis (Gecz et al., 1994; Stayton et al., 1994). The expression pattern seen in the newborn brain is maintained in the adult mice (Bérubé et al., 2005).

Deletion of Atrx in mouse ESCs causes reduced proliferation, and GATA1-Cre-mediated conditional deletion at the 8- to 16-cell stage is embryonic lethal in most cases due to defects in trophoblast development, demonstrating the importance of ATRX in the earliest stages of development. A small number of female embryos are able to survive by selectively inactivating the X chromosome containing the mutated Atrx allele, similar to what is seen in human female carriers of ATRX mutations (Garrick et al., 2006). Several roles for ATRX in brain development have been identified. Constitutive overexpression causes neural tube defects, disorganization of the neuroepithelial cell layer, and increased embryonic and perinatal lethality (Bérubé et al., 2002). Constitutive deletion was embryonic lethal, while Foxg1-Cre-mediated conditional deletion in the embryonic forebrain beginning at E8.5 caused a significant increase in p53 mediated apoptosis, reduced neuronal migration, and mitotic defects in cortical neuroprogenitor cells. These defects likely combine to cause the smaller forebrain size, disruption of the hippocampus with loss of the dentate gyrus, and perinatal lethality seen in these mice (Bérubé et al., 2005; Ritchie et al., 2008; Seah et al., 2008).

Several patients with a milder form of ATR-X syndrome have been identified with a mutation in exon 2 leading to reduced expression of a truncated form of ATRX from an alternative downstream initiation site (Abidi et al., 2005; Guerrini et al., 2000; Howard et al., 2004). Deletion of exon 2 in mice mirrored this ATRX expression pattern, and
caused no noticeable anatomical or reproductive symptoms. However, it led to 
behavioural deficits associated with memory and learning (functions of the hippocampus) 
due to reduced phosphorylation of a glutamate receptor leading to reduced neuronal 
signaling (Nogami et al., 2011).

ATRX is also involved in terminal differentiation of neurons in the retina (Medina et al., 
2008), muscle cell development and regeneration (Huh et al., 2012), and gonad 
development in both males and females (Bagheri-Fam et al., 2011; Huyhn et al., 2011).

1.4  Cellular functions of ATRX

1.4.1  Mitosis and meiosis

The first study to look at the cellular localization of ATRX found it associated with both 
interphase and metaphase chromosomes (McDowell et al., 1999). The finding that 
ATRX is phosphorylated as it transitions from its interphase to mitotic locations 
suggested distinct roles for ATRX at different phases of the cell cycle (Bérubé et al., 
2000). A role for ATRX in cell division has been confirmed by demonstrating that loss 
of ATRX leads to improper chromosome congression, cohesion, and segregation during 
mitosis, with misaligned chromosomes seen at the metaphase plate in both HeLa cells 
and embryonic forebrain. Loss of ATRX therefore leads to slower mitosis because 
chromosomes take longer to align to the metaphase plate leading to activation of the 
mitotic spindle checkpoint (Ritchie et al., 2008). Similar defects are seen during meiosis. 
Loss of ATRX leads to improper metaphase II chromosome alignment (Baumann et al., 
2010; De La Fuente et al., 2004) leading to aneuploidy and decreased fertility (Baumann 
et al., 2010). Loss of ATRX was also shown to cause genome instability leading to 
delayed mitotic progression through S phase in muscle (Huh et al., 2012).

1.4.2  DNA replication

A role for ATRX in DNA replication is mainly believed to involve the replication of 
difficult to process DNA, such as regions that are repetitive and/or contain secondary 
DNA structures (reviewed in (Clynes and Gibbons, 2013)). ATRX is enriched at 
repetitive G-rich sequences including rDNA repeats (Gibbons et al., 2000; Law et al.,
2010; McDowell et al., 1999), the G-rich strand of telomeres (comprised of TTAGGG repeats), and G-rich tandem repeats throughout the genome (Law et al., 2010), all regions that can form G-quadruplexes (Law et al., 2010; Lipps and Rhodes, 2009). As a secondary DNA structure, G-quadruplexes present a barrier to replication requiring specialized proteins for their resolution and bypass (Lopes et al., 2011; Paeschke et al., 2011; Schwab et al., 2013). ATRX may therefore assist in DNA replication by resolving these structures to allow passage of the DNA replication machinery. Supporting this, loss of ATRX causes DNA damage and dysfunction at telomeres (Huh et al., 2012; Watson et al., 2013; Wong et al., 2010), and delayed progression through the S phase of the cell cycle (Huh et al., 2012; Watson et al., 2013). Resolution of G-quadruplexes by ATRX may involve deposition of H3.3, as ATRX associates with telomeres and deposits H3.3 during S phase (Wong et al., 2010).

1.4.3 Gene regulation

SWI/SNF proteins have long been known to be regulators of gene expression and were initially identified as transcriptional activators (Winston and Carlson, 1992). As a member of the SWI2/SNF2 subfamily of proteins ATRX has therefore been speculated to be a regulator of gene expression ever since its sequence was originally determined (Stayton et al., 1994). The presence of a PHD domain added to this early speculation, as PHD fingers were thought to regulate transcription through interactions with chromatin (Aasland et al., 1995). The PHD finger of ATRX was later found to be a subcomponent of the ADD domain, a domain shared with the DNMT3A/B/L de novo methyltransferase genes (Aapola et al., 2000). DNMT3A and DNMT3B can repress transcription through DNA methylation (Bachman et al., 2001) and ATR-X syndrome patients have aberrant methylation at several repetitive regions of the genome (Gibbons et al., 2000), suggesting a role for ATRX in DNA methylation, a common epigenetic method of regulating gene expression (see section 1.5.1 for more details). However, a direct mechanism relating ATRX to DNA methylation has not yet been identified.

Besides having particular protein domains, ATRX is also implicated in gene regulation on account of its protein binding partners. ATRX was shown to bind EZH2 by yeast two-hybrid and in vitro binding assays (Cardoso et al., 1998), although no subsequent
studies confirming in vivo binding have yet to be published. EZH2 is the catalytic subunit of the polycomb repressive complex 2 that represses chromatin through histone methylation (O'Meara and Simon, 2012). ATRX’s binding to HP1 supported its role as a potential regulator of heterochromatin. HP1 proteins (HP1α, HP1β, HP1γ) have traditionally been associated with heterochromatin and gene silencing, but mounting evidence shows additional roles in gene activation (Kwon and Workman, 2011). ATRX binds DAXX at PML-NBs. Both PML-NBs and DAXX have been implicated in numerous cellular functions, including gene regulation (Dundr, 2012; Salomoni and Khelifi, 2006) but it is unclear exactly how or if ATRX could play a role in general gene regulation through these proteins.

While the above indirect evidence is highly suggestive of a role for ATRX in regulating gene expression, the identification of specific genes regulated by ATRX has cemented this position. The recognition that ATR-X syndrome patients have decreased α-globin but not β-globin expression suggested that ATRX may regulate genes in particular genomic regions. The two globin genes exist in very different chromatin environments, with α-globin being near a telomere in a region of high GC content and constitutively open chromatin (Figure 1-3A) (Gibbons et al., 1995). Evidence from studies of viral infection has demonstrated the ability for ATRX and DAXX working at PML to repress expression of viral genes by blocking recruitment of RNA polymerase II (RNA PolII) to the promoter (Newhart et al., 2012; Schreiner et al., 2013; Tsai et al., 2011). An additional study of promoter regulation showed that if ATRX is targeted to a promoter it can inhibit transcription (Tang et al., 2004; Valadez-Graham et al., 2012), and that DAXX can alleviate this repression likely by sequestering ATRX to PML-NBs (Tang et al., 2004). Alternatively, in a mouse testis cell line it was shown that ATRX interacts with the androgen receptor (AR) protein, binds the promoter of the AR target gene Rhox5, and promotes instead of represses expression (Bagheri-Fam et al., 2011), demonstrating that ATRX can both activate or repress gene expression by binding to promoters depending on context, such as its binding partner and cell type (Figure 1-3B). ATRX can also regulate gene expression by binding to regulatory regions. It binds to imprinting control regions where it regulates the recruitment of CTCF and the cohesin
complex to silence nearby imprinted genes, possibly by facilitating chromosome looping (Kernohan et al., 2010) (Figure 1-3C).

Several lines of evidence have shown interactions of ATRX with PCH or telomeres, and it has also been seen that, along with regulating transcription of specific genes, as already discussed, ATRX can regulate expression of transcripts from these repetitive regions. ATRX deposits the histone variant H3.3 at both PCH and telomeres, and in the absence of ATRX, H3.3 is lost from both of these sites. It is interesting to note then that while expression from PCH is decreased upon ATRX (and therefore H3.3) loss (Drane et al., 2010), expression of telomere repeat-containing RNA (TERRA) increases (Figure 1-3D) (Goldberg et al., 2010). Increasing evidence over the last few years has continued to demonstrate this dual nature of H3.3’s role in gene expression (Elsaesser et al., 2010; Szenker et al., 2011).

Drosophila ATRX can assemble heterochromatin to repress nearby genes (Figure 1-3E) (Bassett et al., 2008; Emelyanov et al., 2010), while a different study found that dATRX is found mostly at sites of active transcription and that both overexpression and deletion of dAtrx can cause derepression of gene silencing (Schneiderman et al., 2009). dATRX lacks the ATRX ADD domain, and Drosophila have no PML protein, so ATRX and dATRX likely play similar but different roles in regulating gene expression (Schneiderman et al., 2009). For example, dATRX may lack some of the targeting ability of ADD-containing ATRX, while maintaining the same chromatin remodeling abilities.

Taken together, it is clear that ATRX can act as either a repressor or activator of gene expression through several different mechanisms.
Figure 1-3: Gene regulation by ATRX.

(A) The α-globin gene cluster contains intergenic potential G-quadruplex forming tandem repeats (TRs) thought to inhibit DNA replication. In the absence of ATRX (top), this may lead to a downregulation of nearby genes. This repression is enhanced by proximity to TRs or increased TR length. In the presence of ATRX (bottom), ATRX facilitates replication through the G-quadruplex forming regions, allowing proper gene expression. (B) ATRX can act at promoters. ATRX can bind the androgen receptor (AR) to cooperatively activate the Rhox5 promoter (top), or it can inhibit expression by blocking RNA polymerase II (PolII) (bottom left). In the latter scenario, DAXX can sequester ATRX to PML-NBs (PML) to remove the ATRX promoter blocking (bottom right). (C) Binding of ATRX to the H19 imprinting control region recruits CTCF and cohesin to silence genes by altering DNA looping. (D) ATRX recruits histone H3.3 to activate transcription from repetitive pericentromeric heterochromatin (PCH) or to repress transcription of telomere repeat-containing RNA (TERRA). (E) In Drosophila, ATRX binds to HP1 and is required for heterochromatin gene silencing, presumably by assembling heterochromatin through ATRX’s and HP1’s ability to bind the repressive chromatin mark H3K9Me3. ATRX is purple throughout, and green arrows represent relative levels of transcription within each section of the figure.
1.5 Regulating the genome

1.5.1 DNA methylation

Methylation of cytosines within eukaryotic DNA, and the idea that it could act as a cellular regulatory mechanism, was proposed over 35 years ago. In particular, it was suggested that DNA methylation could change the affinity of DNA binding proteins, and that DNA methylation played a role in X chromosome inactivation. At the time, however, DNA methylation was thought to activate rather than repress genes (Riggs, 1975). It was subsequently determined that DNA methylation does in fact affect protein binding and gene expression, when MeCP2 was found to bind methylated DNA, leading to transcriptional repression through recruitment of a histone deacetylase (Jones et al., 1998; Nan et al., 1998). More recently, it has been seen that MeCP2 can bind both methylated and unmethylated DNA, and either repress or activate gene expression, depending on context (Hansen et al., 2010). DNA methylation, however, seems to be largely associated with transcriptional repression (DNA methylation and its regulation of transcription reviewed in (Deaton and Bird, 2011)). DNA methylation occurs throughout the genome at CpG dinucleotides. While most individual CpG’s are methylated, areas of high CpG concentration called CpG islands are often found within gene promoters and are generally unmethylated. When methylation of CpG island promoters does occur, it acts as a stable repressor of transcription. This de novo CpG methylation is mediated by the DNA methyltransferases DNMT1, DNMT3A, and DNMT3B (reviewed in (Klose and Bird, 2006)). Promoter methylation can repress transcription either by directly blocking transcriptional activators from binding, or by attracting methyl binding proteins to inhibit transcription, such as with MeCP2 as described above. CpG islands found outside promoters, in particular those found within gene bodies, are more often methylated compared to those found in promoters. Changes in intragenic CpC island methylation may have roles in transcription elongation or splicing (Deaton and Bird, 2011), for example, increased intragenic DNA methylation was shown to inhibit transcription elongation (Lorincz et al., 2004).
1.5.2 Chromatin remodeling proteins

DNA is packaged into chromatin which allows organization and regulation of the nearly two meters of DNA in each human cell. Chromatin remodeling complexes act on the basic unit of chromatin, the nucleosome, to alter the interactions between DNA and histones allowing these complexes to reposition, assemble, eject, or restructure the composition of nucleosomes (reviewed in (Clapier and Cairns, 2009)). For example, histones are repositioned or ejected to allow or inhibit access to the DNA by transcription factors, assembled after new DNA is synthesized during DNA replication, or restructured to replace canonical histones with histone variants. Chromatin remodeling complexes are divided into four families: SWI/SNF; imitation switch (ISWI); chromodomain, helicase, DNA binding (CHD); and inositol requiring 80 (INO80). Each complex contains a catalytic subunit with a SWI2/SNF2 or SWI2/SNF2-like ATPase domain, but differing complex composition allows for distinctive targeting and overall function. Targeting of these complexes to particular regions of the genome relies on multiple nucleosome recognition domains. Bromodomains found in SWI/SNF family members recognizes acetylated lysines. Chromodomains of the CHD proteins recognize methylated lysine. Plant homeodomains found in various proteins of the chromatin remodeling complexes also recognize methylated lysine. SANT-SLIDE domains of ISWI proteins binds unmodified histone tails through the SANT domain, and bind nucleosomal DNA through the SLIDE domain. As a SWI2/SNF2, ATPase-containing protein, ATRX therefore resembles the catalytic subunits seen in a wide variety of chromatin remodeling complexes, while its PHD-like domain allows nucleosomal targeting.

1.5.3 The histone variant H3.3 and transcriptional elongation

Nucleosomes consist of about 147 bp of DNA wrapped around a histone octamer complex consisting of an (H3-H4)$_2$ histone tetramer with two H2A-H2B histone dimers (Luger et al., 1997). Besides these canonical histones several variant histones exist. The H3 family of histones contains several variants which perform different functions and differ across species (Szenker et al., 2011). Firstly, all organisms have a centromere-specific CenH3 (CENP-A in mammals). Besides CenH3, yeast has a single H3 variant most similar to human H3.3. *Drosophila* has H3.3 and H3.2, mice have H3.3, H3.2,
H3.1, and the testis-specific H3t, while humans have seven histone H3 variants: the five found in mice, plus the recently identified primate-specific H3.X and H3.Y (Wiedemann et al., 2010). Most chromatin is established during DNA replication with the assembly of nucleosomes that contain replication-dependent histones—H3.1/2, H2A/B, and H4—which are expressed most highly during S phase. On the contrary, replication-independent variants such as H2A.Z and H3.3 are synthesized continuously throughout the cell cycle and can be incorporated in a replication-independent manner (Henikoff and Ahmad, 2005). In yeast and organisms with only a single H3 variant, this “universal” H3 fulfills both replication-dependent and -independent incorporation (Szenker et al., 2011). H3.3 differs from H3.1 by five amino acids and from H3.2 by only four. Three amino acids specific to the H3.3 core (an “AIG” motif) determine its replication-independent deposition (Ahmad and Henikoff, 2002) and genomic localization patterns (Goldberg et al., 2010), and an alanine to serine change on the H3.3 histone tail adds an additional site for potential phosphorylation (Hake et al., 2005).

Transcription is a multistep process involving different combinations of histone modifications and transcription factors at different steps in the process (reviewed in (Buratowski, 2009)). RNA PolII contains a C-terminal domain (CTD) which exhibits different states of phosphorylation at different stages during transcription (Komarnitsky et al., 2000). These phosphorylation states are involved in recruiting other factors necessary at each stage. RNA PolII is unphosphorylated when initially bound to promoters then is phosphorylated on CTD serine 5 (Ser5) for the first few hundred nucleotides. Ser5 phosphorylation attracts methyltransferases which leads to enrichment of H3K4Me3 at the promoters of active and recently active genes (Ng et al., 2003). Ser5 phosphorylation is gradually replaced with Ser2 phosphorylation on the elongating RNA PolII and the regions of double Ser2/5 phosphorylation attract H3K36 methylation. H3K36 methylation therefore marks actively elongating transcription.

Nucleosomes act as barriers to transcription (Bondarenko et al., 2006; Knezetic and Luse, 1986). In order for the transcriptional machinery to proceed down DNA, the interactions between the DNA and histones must be disrupted. Therefore, the nature of these interactions affects how easily transcription can occur. Different combinations of
histones exhibit different levels of stability, referring to how strong the histone-DNA interactions are and how easily the histones can be evicted from DNA. Nucleosomes containing H3.3 are inherently unstable. H2A.Z contributes to this instability but is relatively stable itself if paired with H3.1/2 (Jin and Felsenfeld, 2007). Besides the core nucleosome particles, histone tail modifications can also affect stability. Acetylation inhibits the positive charge on the histone tail, reducing the strength of DNA-histone interaction (Brower-Toland et al., 2005). Histone tail methylation attracts secondary remodelers. For example, H3K9Me3 attracts the repressive protein HP1 (Fischle et al., 2003; Lachner et al., 2001) which may bridge adjacent nucleosomes to stabilize them (Canzio et al., 2011), and H3K27Me3 attracts repressive polycomb group proteins (Fischle et al., 2003). H3.3 acts doubly to promote transcription due to its inherent instability and high propensity to have activating modifications (McKittrick et al., 2004).

Elongating RNA PolII leads to eviction of one of the H2A-H2B histone dimers, while elongation at higher rates can lead to eviction of the entire nucleosome (Petesch and Lis, 2012). Re-assembling chromatin after the passage of RNA PolII (or of DNA polymerase during replication) is largely the work of various histone chaperones (reviewed in (Avvakumov et al., 2011; Duina, 2011)). ‘Facilitates chromatin transcription’ (FACT) is largely responsible for depositing H2A-H2B, though it can also bind H3-H4 (Orphanides et al., 1998; Xin et al., 2009), while SPT6 acts on H3-H4 (Bortvin and Winston, 1996; Ivanovska et al., 2011). As a replication-independent histone H3 variant, one of the main functions of H3.3 is incorporation at sites of active transcription after passage of the transcription machinery and eviction of the original histones (Ahmad and Henikoff, 2002; Schwartz and Ahmad, 2005). Incorporation of H3.3 into transcribing gene bodies is thought to be reliant on the histone chaperone HIRA (Goldberg et al., 2010; Schwartz and Ahmad, 2005). H3.3 is therefore both a mark of active transcription and a facilitator of future transcription.
1.6 Chromosome ends: the pseudoautosomal regions and telomeres

Several lines of evidence have now associated ATRX with G-rich and repetitive genomic targets. Two regions that exhibit these characteristics are the pseudoautosomal regions (PARs) and telomeres.

1.6.1 The pseudoautosomal region: ancient regions of homology between the X and Y chromosomes

The X and Y sex chromosomes in modern placental mammals are highly dimorphic but initially evolved from a homologous pair of autosomes (origin and evolution of the PAR reviewed in (Graves et al., 1998; Helena Mangs and Morris, 2007; Katsura et al., 2012)). Over millions of years of mammalian evolution the sex chromosomes have lost most of their homology due to chromosome Y attrition. However, the additions of genetic material from other autosomes onto the ends of the diverging X and Y chromosomes at several points in time have added new homology in regions known as pseudoautosomal regions. The X and Y chromosomes have therefore gone through several rounds of addition and attrition resulting in the modern X and Y chromosomes and PARs. Because species diverged at different points in history as this process continued, the PARs show various levels of similarity between species; therefore, the human and chimpanzee PAR1’s are very similar, but are different from species such as cattle and sheep which themselves have more closely related PAR1’s. Rodents exhibit an entirely different PAR organization. It is believed that at some point after the evolutionary divergence between mice and humans, nearly the entire ancestral PAR1 was lost from the rodent lineage while a new, smaller PAR1 was gained (Graves et al., 1998; Helena Mangs and Morris, 2007; Katsura et al., 2012) (Figure 1-4).

It is not clear exactly what the purpose of the PAR is. While they are used for pairing and crossing over in most mammals, marsupials and some rodents have no PARs demonstrating that they are not a universal requirement for mammalian meiosis (Graves et al., 1998; Helena Mangs and Morris, 2007; Katsura et al., 2012).
Figure 1-4: Evolution of the pseudoautosomal region.

(A) The pseudoautosomal region (PAR) originated when DNA from a pair of autosomes translocated onto the ends of the ancient X and Y chromosomes. (B) Loss of homology has gradually shrunk the PAR. Genes just outside the X chromosome PAR (light pink) often contain pseudogenes on the Y chromosome. (C) The human PAR has continued to shrink, but the modern human PAR is largely the same as the predicted ancestral PAR. (D) Several rodents, including mice and rats, have completely lost the original PAR but gained a new smaller, unique PAR. Genes that were in the ancestral PAR have translocated to autosomes, with some now located near telomeres and others not yet definitively placed in the mouse genome. Black circles represent centromeres. mya, millions of years ago.
By comparing sequences between the X and Y chromosomes, and between species, it has been estimated that the boundary of the ancestral PAR1, before attrition, was at the AMELX gene (Iwase et al., 2003) giving the original PAR1 a size of 11.3 Mb. The modern cattle PAR1 is estimated at 5-9 Mb (Das et al., 2009), the human PAR1 is 2.5 Mb, while rodents, who have lost the original PAR1, have gained a small region of about 0.7 Mb (Perry et al., 2001). The differences in PAR size are reflected in their gene content. Fewer than half of the 24 PAR1 genes identified so far in humans have also been found in the mouse genome, and all have diverged considerably (Perry et al., 2001). This divergence is largely due to the increased recombination rates in the PARs during male meiosis (Lien et al., 2000). Many genes located on the human X chromosome between 2.5 and 11.3 Mb (i.e. within the original PAR but outside the modern one) have degenerative, non-functional copies (pseudogenes) on the Y chromosome.

The divergence between PAR orthologs makes the identification of human PAR1 orthologs difficult. Interestingly, in the mouse, all human PAR1 orthologs identified to date are located on autosomes. For example, ‘colony stimulating factor 2 receptor, alpha’ (Csf2ra) is located on mouse chromosome 19 (Disteche et al., 1992) and ‘CD99 antigen’ (Cd99) and ‘dehydrogenase/reductase short-chain dehydrogenase/reductase family, X chromosome’ (Dhrsx) are located on chromosome 4 (Bixel et al., 2004; Gianfrancesco et al., 2001). Human orthologs of ‘acetylserotonin O-methyltransferase-like’ (ASMTL) and several members of the arylsulfatase (ARS) family of genes (ARSD, ARSE, ARSF, and ARSH) are located just outside the human PAR1, and have not yet been reported in the mouse. The human X chromosome arylsulfatases are an example of genes that have pseudogenes on the Y chromosome. Due to the location of these mouse genes in the PAR region of evolutionary ancestors (before rodents lost the original PAR), and their current autosomal location, I will refer to these genes in the mouse as “ancestral PAR genes” (aPAR genes).

Equal gene dosage between XX females and XY males is usually achieved by the silencing of one X chromosome in every female cell, a process known as X chromosome inactivation (Lyon, 1989). Because both males and females have two copies of all PAR genes, there is no requirement for dosage compensation and these genes therefore escape
this inactivation process in humans (Carrel and Willard, 2005) (and being on autosomes
are not subject to XCI in mice).

In addition to being subtelomeric, PARs comprise a unique chromosomal environment
that is rich in repetitive sequences (Bacolla et al., 2006; Gianfrancesco et al., 2001) and
GC-rich, making PARs and PAR genes potential targets for ATRX.

1.6.2 Telomeres

Telomeres are nucleoprotein structures that protect chromosome ends from degradation
upon DNA replication (the “end-replication problem”) and from unwanted double strand
break repair (the “end-protection problem”) (reviewed in (Stewart et al., 2012)). The free
ends of DNA are usually recognized by the cell as double stranded breaks which activate
DNA repair mechanisms. To avoid this process the six protein shelterin complex
specifically recognizes telomeres and acts as a protective cap on the ends of
chromosomes (de Lange, 2005). The end-replication problem exists because DNA
polymerase is unable to fully replicate the 5’ end of telomeres, meaning that after each
replication the telomere would get shorter. To alleviate this shortening the riboprotein
telomerase uses its RNA template to reverse transcribe new single-stranded DNA
consisting of “TTAGGG” repeats onto telomeres. DNA polymerase then fills in the
complementary strand. Telomeres therefore consist of many copies of short repeats, with
an overhanging G-rich strand (5’-TTAGGG-3’) and the complementary C-rich strand (5’-
CCCTAA-3’) (Gilson and Geli, 2007). The single-stranded G-rich strand is then able to
form G-quadruplexes.

Besides a role in chromosome end protection, it was discovered just a few years ago that
telomeres are not transcriptionally silent, as previously thought. Starting in the
subtelomere and proceeding through the telomeric repeats, the non-coding telomeric
repeat-containing RNA (TERRA) is transcribed, and these transcripts then associate with
the telomeres (Azzalin et al., 2007). The complete role of TERRA is uncertain, but
TERRA is able to alter telomere length by inhibiting telomerase (Schoeftner and Blasco,
2008). A role for ATRX at telomeres was identified when it was shown that deletion of
ATRX causes a moderate (~1.7 fold) increase in TERRA expression (Goldberg et al., 2010).

Telomere-like repeats found outside telomeres are called interstitial telomere repeats/sequences (ITS) (Azzalin et al., 1997; Meyne et al., 1990). ITS’s are often found within subtelomeres, where they are likely the product of recombination events. They are also found within pericentromeric heterochromatin, and at intrachromosomal locations where they may be the product of chromosomal fusions or the remnants of ancient double stranded break repair mechanisms that may have involved telomerase (Azzalin et al., 2001; Flint et al., 1994; Ruiz-Herrera et al., 2008). While ITS’s are often associated with fragile sites, it is unclear whether ITS’s themselves can lead to chromosome breaks, or whether they have simply been inserted into already fragile sites during DNA repair (Azzalin et al., 2001; Bolzan, 2012). Unlike traditional telomeres, intrachromosomal ITS’s do not constitute constitutive heterochromatin, but similar to telomeres, there is evidence that they may be at least moderately transcribed (Svetlova et al., 2007). In addition, ITS’s share at least some of their bound protein partners with those normally found at telomeres as part of the shelterin complex (Simonet et al., 2011; Yang et al., 2011). Depletion of at least two of these proteins (RAP1 or TRF2) can lead to altered transcription of genes proximal to RAP1- or TRF2-targeted ITS’s (Yang et al., 2011). While the complete function of ITS’s is unclear, they may contribute to genome instability, and appear to have the potential to affect the regulation of nearby genes.

1.6.3 G-quadruplexes

As described earlier, G-quadruplexes are short sequences containing four guanine triplicates that form a four stranded secondary structure upon DNA denaturation such as during replication or transcription ((Biffi et al., 2013; Duquette et al., 2004) and reviewed in (Bochman et al., 2012)) (Figure 1-2). G-quadruplexes can therefore form a barrier to processing ssDNA, requiring specific proteins to bypass them ((Belotserkovskii et al., 2010) and reviewed in (Tornaletti, 2009)). If not properly bypassed, stalling of RNA PolIII during transcription, or of DNA polymerase during replication, triggers DNA repair mechanisms which attempt to repair these lesions by nucleotide excision repair. This superfluous DNA repair can introduce DNA damage and mutations (Vasquez and Wang,
G-quadruplexes form throughout the genome in regions of high guanine concentration but are particularly enriched at telomeres due to the telomeric “TTAGGG” repeats. ATRX can bind directly to G-quadruplexes and is highly enriched at telomeres (Law et al., 2010) and loss of ATRX causes telomere-specific DNA damage (Huh et al., 2012; Watson et al., 2013; Wong et al., 2010). ATRX may therefore be involved in processing or otherwise bypassing G-quadruplexes (Clynes and Gibbons, 2013).

1.7 Thesis overview

The overall aim of this work was to identify genes regulated by ATRX and determine a mechanism by which this regulation could be achieved. At the beginning of this study, genes that are directly regulated by the ATRX protein had not yet been identified, and even in subsequent studies the focus has largely been on the regulation of individual genes. To identify potential genes that are controlled by ATRX, we used microarrays to perform wide-scale screens of expression changes in developing mouse brains with and without ATRX. We found that a subset of ancestral PAR1 genes is consistently downregulated in the absence of ATRX. Among them are two potentially novel mouse orthologs of Arsd/e and Asmtl. Remarkably, murine PAR1 homologs have translocated to various autosomes, reflecting the complex recombination history during the evolution of the mammalian X chromosome. Despite the common ancestral location of these genes, we found that ATRX does not influence their expression in cells from species in which these genes remain in the PAR. Therefore, we proposed that the aPAR genes misregulated in mice share conserved sequences and/or chromatin features targeted by ATRX, which are nevertheless unique to their autosomal location in mice (Levy et al., 2008).

Having identified a unique class of genes regulated by ATRX, we next sought to determine a possible mechanism by which they could be co-regulated. A direct mechanism for the regulation of gene transcription by ATRX had not yet been described and could be relevant to the function of ATRX in the central nervous system and for its tumor suppressive activities. Here we report that ATRX and the histone variant H3.3 are enriched at G-rich regions within the gene bodies of Dhrsx and other aPAR genes in the mouse brain. Loss of ATRX causes decreased H3.3 occupancy within the gene body.
without affecting TERRA levels, histone modifications or DNA methylation. Importantly we provide evidence that in the absence of ATRX, RNA polymerase II progression is impeded at the G-rich region of \textit{Dhrsx}. We propose a model whereby ATRX facilitates transcription elongation of particular target genes by assisting the passage of the transcription machinery through G-rich templates (Levy et al., Nucleic Acid Research, in revision). Taken together, the findings presented here identify a novel function for ATRX in the regulation of aPAR genes, and provide insight into how defects in ATRX could lead to problems in development and disease.
1.8 References


DNA binding properties of the XNP/ATR-X protein. Journal of medical genetics 37, 746-751.


42
WRN-deficient, telomere dysfunctional cells promotes escape from senescence and engagement of the ALT pathway. Genes & development 19, 2560-2570.


to a common mechanism underlying the ATR-X syndrome. Human molecular genetics 5, 1899-1907.


Chapter 2

2 The SWI/SNF protein ATRX co-regulates pseudoautosomal genes that have translocated to autosomes in the mouse genome

The work described in this chapter represents the first genome-wide analysis of gene expression changes that result from loss of ATRX and identifies the aPAR genes as a unique class of genes regulated by ATRX. This chapter, except for data and commentary related to section 3.6, has been published in BMC Genomics (Levy et al., 2008). The conclusions have been modified to accommodate the new data from section 3.6.

2.1 Introduction

The sex chromosomes in modern placental mammals (eutharians) are highly dimorphic but initially evolved from a homologous pair of autosomes (Ohno, 1967). Over millions of years of mammalian evolution, the sex chromosomes have lost most of their homology due to chromosome Y attrition (Charlesworth and Charlesworth, 2000). The remaining homology between the sex chromosomes exists in the pseudoautosomal regions (PARs), located at the ends of the X and Y chromosomes (Graves, 2006), and was generated when genetic material from the tips of autosomes translocated onto the ancient sex chromosomes (Graves et al., 1998). Gene dosage between XX females and XY males is usually achieved by the silencing of one X chromosome in every female cell, a process known as X chromosome inactivation (XCI) (Lyon, 1989). Because both males and females have two copies of all PAR genes, there is no requirement for dosage compensation and these genes therefore escape this inactivation process (Carrel and Willard, 2005).

Comparison of human PARs with those of other primates and carnivores (dogs and cats) and artiodactyls (cattle, sheep, pigs; representing the common evolutionary ancestor between humans and mice) has revealed that gene content is mostly conserved in eutherians, including the existence of PARs at both ends of the X and Y chromosomes (Graves et al., 1998). However, rodents are strikingly different in that they have a single,
dissimilar and considerably shorter PAR region (Perry et al., 2001). Fewer than half of the 24 PAR1 genes identified so far in humans have also been found in the mouse genome, and all have diverged considerably (Perry et al., 2001). This divergence is largely due to the increased recombination rates in the PARs during male meiosis (Lien et al., 2000). In addition, the PARs comprise a unique chromosomal environment that is rich in repetitive sequences (Bacolla et al., 2006; Gianfrancesco et al., 2001). For these reasons, the identification of human PAR genes and orthologs has been difficult. Interestingly, in the mouse, all human PAR1 orthologs identified to date are located on autosomes. For example, Csf2ra is located on mouse chromosome 19 (Disteche et al., 1992) and Cd99 and Dhrsx are located on chromosome 4 (Bixel et al., 2004; Gianfrancesco et al., 2001). Human orthologs of ASMTL and the arylsulfatase (ARS) family of genes (ARSE, ARSD, ARSF, and ARSH) located just outside the PAR1, have not yet been reported in the mouse (Figure 2-1). Due to their location in the PAR region of evolutionary ancestors, and their current autosomal location, we will refer to these genes in the mouse as “ancestral PAR genes”.
Figure 2-1: Evolution of PAR genes in humans and mice.

PAR genes that are downregulated in the ATRX-null mouse forebrain are clustered together within the PAR1 region of common evolutionary ancestors of humans and mice but have translocated to autosomes in the mouse. Vertical lines and arrows represent individual genes. The position of the first nucleotide for each gene is as follows: SHOX (505,079), CSF2RA (1,347,701), ASMTL (1,482,032), DHRSXY (2,147,553), CD99 (2,619,553), ARSD (2,848,421), ARSE (2,832,011) (Human reference sequence NCBI Build 36.1). PAR1, pseudoautosomal region 1; PAR2, pseudoautosomal region 2. PAR regions are highlighted in orange.
The α-thalassemia mental retardation, X linked (ATRX) protein, transcribed from Xq13.3 belongs to the sucrose non-fermenting 2 (Snf2) family of enzymes that use the energy of adenosine tri-phosphate (ATP) hydrolysis to disrupt nucleosome stability (Eisen et al., 1995; Picketts et al., 1996). Mutations in ATRX result in moderate to profound cognitive deficits, facial dysmorphisms, as well as skeletal and urogenital abnormalities, among other symptoms (Gibbons and Higgs, 2000). The chromatin remodeling properties of ATRX have been demonstrated *in vitro* (Xue et al., 2003). In addition to a conserved ATPase/helicase domain, ATRX has an N-terminal zinc finger ATRX-DNMT3A/B-DNMT3L (ADD) domain that is shared with *de novo* methyltransferases. Several lines of evidence have also linked ATRX to highly repetitive genomic regions including pericentromeric heterochromatin in mouse and human cells (McDowell et al., 1999). Moreover, ATRX mutations in humans result in aberrant DNA methylation patterns at several repetitive elements, including ribosomal DNA (rDNA) repeats, subtelomeric repeats and Y-specific satellite repeats (Gibbons et al., 2000). These repetitive sequences usually form heterochromatic structures and seem to be specifically targeted by the ATRX protein.

To assess the role of ATRX in brain development, we previously used Cre-*lox*P recombination to remove *Atrx* specifically in the mouse forebrain beginning at E8.5. Loss of ATRX in the embryonic forebrain caused hypocellularity, a reduction in forebrain size, and loss of the dentate gyrus (Bérubé et al., 2005).

Genes that are directly regulated by ATRX have not yet been identified in either humans or mice. To identify potential ATRX target genes we performed a screen of gene expression in control and ATRX-null mouse forebrain tissue. We found that a subset of ancestral PAR1 genes is consistently downregulated in the absence of ATRX in the developing mouse brain, where they are located on autosomes, but not in non-murine species, where they are located in the X/Y PAR1. Among these genes are two potentially novel mouse orthologs of *Arsd/e* and *Asmtl*. The only common link between aPAR genes is their adjacent location and shared chromatin environment in the ancestral PAR region. We propose that conserved sequences and/or chromatin features were maintained upon translocation from the PAR1 on the ancestral X chromosome to their current location on
mouse autosomes, and that these sequences and/or features allow ATRX to modulate their expression.

2.2 Materials and Methods

2.2.1 Mouse husbandry

Mice conditionally deficient for ATRX in the forebrain were generated by crossing Atrx$^{loxP}$ females with heterozygous Foxg1Cre male mice, as previously described (Bérubé et al., 2005). Pregnant females were sacrificed at E13.5, embryos were recovered and yolk sac DNA was genotyped by PCR using the primers 17F, 18R and neo' as described previously (Bérubé et al., 2005). For newborns (P0.5) and juveniles (P17), pups were sacrificed and tail DNA was used for genotyping as previously described (Bérubé et al., 2005).

2.2.2 Microarray analysis

Total forebrain RNA (10 µg) was isolated from three pairs of littermate-matched ATRX-null and control embryos using the RNeasy Mini kit (Qiagen). Complementary RNA was generated and hybridized to an Affymetrix Mouse Genome 430 2.0 Array at the London Regional genomics Center (London, Canada). For the analysis at E13.5, RNA from two forebrains was pooled for each array. Probe signal intensities were generated using GCOS1.4 (Affymetrix Inc., Santa Clara, CA) using default values for the Statistical Expression algorithm parameters and a Target Signal of 150 for all probe sets and a Normalization Value of 1. Gene level data were generated using the RMA preprocessor in GeneSpring GX 7.3.1 (Agilent Technologies Inc., Palo Alto, CA). Data were then transformed (measurements less than 0.01 set to 0.01), normalized per chip to the 50th percentile, and per gene to control samples. Probe sets representing Atrx transcripts were removed (10 sets). Remaining probe sets were filtered by fold change of either ≥1.5 or 2 between control and ATRX-null samples, and by confidence level of P<0.05. Heatmaps were generated using the GeneSpring hierarchical clustering gene tree function. Significantly overrepresented GO categories were determined using GeneSpring: at E13.5 and P0.5, probesets were filtered by 1.5 fold change, P<0.05 and categorized as
either up or downregulated. Where there were multiple probesets for a gene, duplicates were removed. P<0.001 was used as the significance cutoff.

2.2.3 Quantitative reverse transcriptase PCR

Total RNA was isolated using the RNeasy Mini kit (Qiagen). First-strand cDNA was synthesized from 3 µg of total RNA using the SuperScript™ II Reverse Transcriptase kit (Invitrogen) with 25 mM dNTPs (GE Healthcare), porcine RNAguard (GE Healthcare) and random primers (GE Healthcare). PCR reactions were performed in triplicate on a Chromo4 Continuous Fluorescence Detector in the presence of iQ™ SYBR Green Supermix and recorded using the Opticon Monitor 3 software (all Bio-Rad Laboratories, Inc.). Samples were amplified as follows: 95°C for 10 sec, annealed for 20 sec, 72°C for 30 sec (See Supplementary table 2-2 for primer sequences and annealing temperatures). After amplification, a melting curve was generated, and samples were run on a 1.5% agarose gel (75V for 1h) to visualize amplicon purity. Standard curves were generated for each primer pair using five-fold serial dilutions of control cDNA. Primer efficiency was calculated as E = \([10^{(-1/slope)} - 1]*100\), where a desirable slope is ~3.32 and \(r^2>0.99\). Samples were normalized to β-actin expression and relative gene expression levels were calculated using GeneX software (Bio-Rad Laboratories, Inc.).

For Arsd/e and Asmtl, the PCR products were gel extracted using the QIAquick Gel Extraction Kit (Qiagen) according to the manufacturer’s instructions and sequenced at the DNA Sequencing Facility at Robarts Research Institute (London, Canada).

2.2.4 Bioinformatics analysis of novel ancestral PAR genes

Probeset sequences were obtained from the Netaffx website (Liu et al., 2003) and used for BLASTn searches (www.ncbi.nlm.nih.gov/BLAST). For calculation of interspecies similarity, sequences were obtained from NCBI RefSeq (www.ncbi.nlm.nih.gov/RefSeq) or Ensemble (www.ensembl.org) where RefSeq sequences were not available, and pairwise comparisons made using Jalview (Clamp et al., 2004).

For generation of trees and sequence alignments, human arylsulfatase E precursor (ARSE, SwissProt P51690, RefSeq NP_000038) and human N-acetylserotonin O-
methyltransferase-like protein (ASMTL, SwissProt O95671, RefSeq NP_004183) were used as seeds and the GenBank NR database was searched for high-similarity, full-length orthologs and paralogs. Fifty-nine ARSE and twenty-two ASMTL sequences met or exceeded the similarity cutoff, with resultant species spanning the metazoa from anemone and urchin to a diverse set of vertebrates. Sequences were aligned using T-Coffee 5.56 (Notredame et al., 2000) using default parameters. Alignments were manually adjusted via inspection prior to further analysis. Approximate maximum-likelihood trees were built using PHYML 2.4.5 (Guindon and Gascuel, 2003) using the WAG model of protein evolution (Whelan and Goldman, 2001) and a seven-category Gamma-plus-invariant model of rate heterogeneity. All rate parameters were estimated from the data. One hundred bootstrap replicates were performed to assess support for the inferred tree topology. All trees are presented as midpoint-rooted phylograms. Since the given mouse sequences were quite short compared to the full protein length, two sets of trees were built for each family to assess if the mouse sequences were long enough to definitively support their taxonomic clustering. One set utilized a “trimmed” alignment where all alignment columns outside the mouse sequence domain were removed. The trees produced with this trimmed alignment were compared with the set of trees produced from the alignment of the mouse sequences to their respective full-length proteins. For both ARSD/E and ASMTL, very little difference was observed between full-length and trimmed-alignment trees. The trimmed alignments tended to exaggerate sequence divergence and modestly lower bootstrap support levels. Overall topology did not appear significantly different, however, and the text references the full-length sequence phylogeny exclusively.

2.2.5 Cell culture and RNA interference

Neuro-2a, IMR-90 and bovine fetal fibroblast cells were grown at 37°C with 5% CO₂ in DMEM supplemented with 10% fetal bovine serum (Sigma-Aldrich). For siRNA treatment, 1.5x10⁴ cells were plated in a plastic six well dish (Corning Incorporated) on glass coverslips and allowed to grow to 25% confluency (approximately 24 hours). Cultures were transfected using Lipofectamine 2000 (Invitrogen) with 8 nM siATRX (Dharmacon), a non-specific control siRNA (Sigma-Aldrich), or with no siRNA
(“Mock”) according to the manufacturers’ instructions (for siRNA sequences refer to (Ritchie et al., 2008)). Total RNA was extracted from cells after 72 hours, cDNA was generated and qPCR analysis performed as described above. Alternatively, cells were processed for immunofluorescence staining as described below.

2.2.6 Immunofluorescence

Cells were fixed using 4% formaldehyde, incubated for 1 h with the primary antibody (H300 anti-ATRX, 1:100 dilution; Santa Cruz) followed by the secondary antibody (goat-anti rabbit Alexa 594, 1:1500 dilution; Molecular Probes), then counterstained with 4’,6-diamidino-2-phenylindole (DAPI) (Sigma-Aldrich) for 5 min. Coverslips were mounted with Vectashield (Vector Laboratories), Z-stack images were captured using a Leica DMI6000b inverted microscope and Openlab software (v5.0, Improvision) and processed using Velocity software (v4.0, Improvision); deconvolution was performed using iterative restoration set with a confidence limit of 95%.

2.3 Results

2.3.1 Effects of ATRX deletion on forebrain gene expression

The ability of ATRX to remodel chromatin (Xue et al., 2003) suggests that ATRX can regulate gene expression. To identify possible gene targets of the ATRX protein in the developing mouse brain, we used the previously described Atrx^Foxg1Cre mice that lack ATRX in the forebrain (Bérubé et al., 2005). In this model system, Atrx deletion is achieved by crossing Atrx^loxP “floxed” mice to mice that express cyclization recombinase (Cre) under the control of the forebrain-specific forkhead box G1 (Foxg1) promoter (Hebert and McConnell, 2000). We performed microarray analysis to compare the expression profiles of the Atrx^Foxg1Cre and control telencephalon at embryonic day 13.5 (E13.5) (n = 3 pairs) using an Affymetrix mouse genome expression array representing approximately 39,000 transcripts (Affymetrix, 2008). Only probe sets showing a significant difference (p<0.05) were included in all subsequent studies. By setting a threshold of 1.5 fold change we identified 202 misregulated probesets, and at a threshold of 2 fold change we identified only 22 altered probe sets. Approximately two-thirds of
the probe sets demonstrating altered expression were upregulated (Supplementary Figure 2-7).

We next compared gene expression patterns in control and ATRX-null forebrain tissue at postnatal day 0.5 (P0.5) (n = 4 pairs). At a threshold of 1.5 fold change we identified 304 probe sets, and at a threshold of 2 fold change we identified 57 probe sets showing altered transcript levels. When we compared results between the two timepoints, we identified 14 probesets commonly upregulated and 13 commonly downregulated more than 1.5 fold, and one increased and three decreased more than 2 fold (Supplementary Figure 2-7). We used GeneSpring software to identify significantly overrepresented Gene Ontology (GO) categories in the ATRX-null mouse forebrain. Several statistically and biologically significant categories of upregulated genes were related to immune response. This could be an indirect response to the increased apoptosis that characterizes the ATRX-null forebrain at E13.5 in the developing cortex and to a lesser extent at P0.5 in the hippocampus (Bérubé et al., 2005). In particular, categories and genes involved in phagocytotic clearing of apoptotic cells, such as complement activation (Trouw et al., 2008), were enriched at both E13.5 and P0.5. Several genes involved in cell adhesion processes were upregulated at P0.5 and, consistent with the abnormal forebrain development described in the ATRX-null forebrain (Bérubé et al., 2005), genes involved in neurogenesis and nervous system development were downregulated at both timepoints (Supplementary table 2-3).

2.3.2 Ancestral pseudoautosomal genes are downregulated in the ATRX-null mouse forebrain

Five of the most downregulated transcripts identified in the microarray analysis were unidentified cDNA clones (Affymetrix IDs 1436320_at, 1448057_at, 1443755_at, 1429730_at and 1453066_at; GenBank Accessions W45978, BI202412, BE457721, AK007409 and BI320076, respectively). To further investigate these probe sets, their NCBI nucleotide sequences were used for a Basic Local Alignment Search Tool nucleotide (BLASTn) search of the nr database. The expressed sequence tag (EST) W45978 has similarity to Mus musculus Dhrsx (NM_001033326, score = 120, E value 5e-24). The EST BI202412 displayed similarity to several unidentified mouse cDNA
clones. Interestingly, a BLAST-like Alignment Tool (BLAT) search of this clone showed similarity to intron 1 of mouse Dhrsx and it could represent an unknown splice variant of Dhrsx. The EST BE457721 is annotated as similar to human ARSE and a BLASTn search revealed high similarity to Rattus norvegicus Arse (NM_001047885, score 197, E value 6e-28). BLASTn of AK007409 showed high similarity to Asmtl in cow (BT02626, score = 248, E value = 6e-62) as well as dog, human, the putative rat Asmtl, and numerous other species. The EST BI320076 displayed no significant hits to any sequences by either BLASTn or BLAT.

Interestingly, while Dhrsx, Arse, and Asmtl do not display an obvious functional connection, they do share a common link in that they are all pseudoautosomal genes in nearly all eutherians (placental mammals). In addition, the microarray data showed decreased expression of Cd99, Shox2 and Csf2ra, genes that also lie within the eutherian pseudoautosomal region. Therefore, while GO analysis identified a subset of downregulated genes involved in brain development at both timepoints, a more in depth analysis of downregulated targets revealed that many are orthologs of PAR1 genes residing on the tip of the X and Y chromosomes in most placental mammals. Overall, our differential gene expression analysis identified six of these genes, constituting approximately half of all PAR1 orthologs discovered in the mouse genome so far. The more intriguing aspect of this finding is that in the mouse, these genes no longer reside within the PAR1 region but have translocated to autosomes (Figure 2-1). It also identified two potential novel PAR1 orthologs—Arse and Asmtl—not previously identified in the mouse genome. At E13.5, these genes represent 6 of the 15 most downregulated transcripts identified by microarray analysis. Strikingly, they constitute 4 of the top 5 most downregulated genes in the microarray performed on P0.5 forebrain tissue (Arse and Shox2 were not significantly decreased in the microarray at P0.5) (Table 2-1, Supplementary figure 2-7). These results suggest that ATRX normally participates in the transcription of these genes during the proliferative (E13.5) and more differentiated (P0.5) stages of forebrain development.
Table 2-1: Downregulated genes in the ATRX-null forebrain at E13.5 and P0.5.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Description</th>
<th>Chromosome</th>
<th>Fold Change</th>
<th>Genbank</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>E13.5 Downregulated Genes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMAGE:354942</td>
<td>Similar to dehydrogenase/reductase (SDR family) X chromosome (Dhrsxy)¹</td>
<td>4 X/Y PAR</td>
<td>-4.81</td>
<td>W45978</td>
</tr>
<tr>
<td>Csf2ra</td>
<td>Colony stimulating factor 2 receptor, alpha, low-affinity (granulocyte-macrophage)</td>
<td>19 X/Y PAR</td>
<td>-3.05</td>
<td>BM941868</td>
</tr>
<tr>
<td>Vit</td>
<td>Vitrin</td>
<td>17 X/Y PAR</td>
<td>-2.74</td>
<td>AF454755</td>
</tr>
<tr>
<td>Shox2</td>
<td>Short stature homeobox 2</td>
<td>3 X/Y PAR</td>
<td>-2.73</td>
<td>AV332957</td>
</tr>
<tr>
<td>Tcf7l2</td>
<td>Transcription factor 7-like 2, T-cell specific, HMG-box</td>
<td>19 X/Y PAR</td>
<td>-2.72</td>
<td>BB175494</td>
</tr>
<tr>
<td>Gbx2</td>
<td>Gastrulation brain homeobox 2</td>
<td>1 X/Y PAR</td>
<td>-2.55</td>
<td>L39770</td>
</tr>
<tr>
<td>IMAGE:3326212</td>
<td>Similar to Arylsulfatase E (Arse)¹</td>
<td>1 X/Y PAR</td>
<td>-2.21</td>
<td>BE457721</td>
</tr>
<tr>
<td>Syt13</td>
<td>Synaptotagmin 13</td>
<td>2 X/Y PAR</td>
<td>-2.19</td>
<td>BB244585</td>
</tr>
<tr>
<td>Cd99</td>
<td>CD99 antigen</td>
<td>4 X/Y PAR</td>
<td>-2.09</td>
<td>AK004342</td>
</tr>
<tr>
<td>Nxph1</td>
<td>Neurexophilin 1</td>
<td>6 X/Y PAR</td>
<td>-1.92</td>
<td>BB274960</td>
</tr>
<tr>
<td>Neurod4</td>
<td>Neurogenic differentiation 4</td>
<td>10 X/Y PAR</td>
<td>-1.86</td>
<td>NM_007501</td>
</tr>
<tr>
<td>Nxph2</td>
<td>Neurexophilin 2</td>
<td>2 X/Y PAR</td>
<td>-1.86</td>
<td>BB169128</td>
</tr>
<tr>
<td>Peg10</td>
<td>Paternally expressed 10</td>
<td>6 X/Y PAR</td>
<td>-1.86</td>
<td>BG076799</td>
</tr>
<tr>
<td>RIKEN:1810009N02</td>
<td>Similar to Asmtl (acetylserotonin O-methyltransferase-like)¹</td>
<td>- X/Y PAR</td>
<td>-1.81</td>
<td>AK007409</td>
</tr>
<tr>
<td>Wif1</td>
<td>Wnt inhibitory factor 1</td>
<td>10 X/Y PAR</td>
<td>-1.81</td>
<td>BC004048</td>
</tr>
<tr>
<td><strong>P0.5 Downregulated Genes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Csf2ra</td>
<td>Colony stimulating factor 2 receptor, alpha, low-affinity (granulocyte-macrophage)</td>
<td>19 X/Y PAR</td>
<td>-7.14</td>
<td>BM941868</td>
</tr>
<tr>
<td>Nr4a2</td>
<td>Nuclear receptor subfamily 4, group A, member 2</td>
<td>2 X/Y PAR</td>
<td>-3.33</td>
<td>NM_013613</td>
</tr>
<tr>
<td>IMAGE:354942</td>
<td>Similar to Dhrsxy (dehydrogenase/reductase (SDR family) X chromosome)¹</td>
<td>4 X/Y PAR</td>
<td>-3.33</td>
<td>W45978</td>
</tr>
<tr>
<td>IMAGE:5656844</td>
<td>Unknown EST</td>
<td>-</td>
<td>-2.86</td>
<td>BI320076</td>
</tr>
<tr>
<td>Met</td>
<td>Met proto-oncogene</td>
<td>6 X/Y PAR</td>
<td>-2.78</td>
<td>BG060788</td>
</tr>
<tr>
<td>Cd99</td>
<td>CD99 antigen</td>
<td>4 X/Y PAR</td>
<td>-2.22</td>
<td>AK004342</td>
</tr>
<tr>
<td>Dsc3</td>
<td>Desmocollin 3</td>
<td>18 X/Y PAR</td>
<td>-2.22</td>
<td>NM_007882</td>
</tr>
<tr>
<td>Mbp</td>
<td>Myelin basic protein</td>
<td>18 X/Y PAR</td>
<td>-2.17</td>
<td>AI323506</td>
</tr>
<tr>
<td>Cbln4</td>
<td>Cerebellin 4 precursor protein</td>
<td>2 X/Y PAR</td>
<td>-2.08</td>
<td>AV343573</td>
</tr>
<tr>
<td>EST</td>
<td>Unknown EST</td>
<td>-</td>
<td>-2.08</td>
<td>BI202412</td>
</tr>
<tr>
<td>Trpc4</td>
<td>Tansient receptor potential cation channel, subfamily C, member 4</td>
<td>3 X/Y PAR</td>
<td>-2.04</td>
<td>BB271442</td>
</tr>
<tr>
<td>RIKEN:1810009N02</td>
<td>Similar to Asmtl (acetylserotonin O-methyltransferase-like)¹</td>
<td>- X/Y PAR</td>
<td>-2.00</td>
<td>AK007409</td>
</tr>
</tbody>
</table>

¹ By BLASTn
² -/- indicates unknown chromosome
2.3.3 Verification of gene expression changes

To validate the microarray results, we performed real-time quantitative reverse transcriptase polymerase chain reaction (qRT-PCR) analysis of \textit{Dhrsx, Cd99, Csf2ra, Shox2} and also of the putative new orthologs of \textit{Asmtl} and \textit{Arse} in ATRX-null and control E13.5 and P0.5 forebrain (n=3 at each time point). Since \textit{Arse} and \textit{Asmtl} have not yet been identified in the mouse, we sequenced the PCR products to ensure they corresponded to the transcripts identified on the microarray, and not to other contaminating sequences. The qRT-PCR results confirmed that five of the six genes exhibit decreased expression in the ATRX-null forebrain at E13.5, and that these genes remain downregulated at P0.5 (Figure 2-2A). In addition, analysis at P17 demonstrated decreased expression of ancestral PAR genes at this later time point as well (Figure 2-2A). One exception was \textit{Shox2} which exhibited highly variable expression differences between the ATRX-null and control tissue at E13.5, P0.5 and P17, ranging from a 170 fold decrease to a 90 fold increase (Figure 2-2B). Therefore, while the expression of \textit{Shox2} is clearly affected by the loss of ATRX protein, the outcome on expression levels appears to be highly variable and does not validate the consistent downregulation observed by microarray analysis.

Our discovery that the expression of several ancestral PAR1 genes is controlled by ATRX throughout the early developmental period of the mouse brain reveals an unexpected association between the levels of ATRX protein and the expression of these ancestral PAR1 genes.
Figure 2-2: Relative expression of ancestral PAR genes in ATRX-null mouse forebrains.

(A) Real-time quantitative RT-PCR of the indicated genes was performed on RNA isolated from the forebrains of littermate-matched Atrx-null and control embryos/mice at E13.5, P0.5 and P17. Results were normalized to β-actin expression levels. Error bars represent standard error of the mean between biological replicates for n=3. (** p<0.05, * p<0.1) (B) Expression of Shox2 in seven (E13.5) or three (P0.5 and P17) littermate-matched pairs. Error bars represent standard error of the mean for three technical replicates. In (A) and (B) expression levels for the control forebrains were set to one for each reaction.
2.3.4 Identification of a novel arylsulfatase family mouse homolog

In humans, a cluster of arylsulfatase genes is located approximately 115 kb centromeric to the PAR1 region on the X chromosome, but still possesses the ability to escape XCI in females (Franco et al., 1995). Located outside the PAR1, these genes do not have an identical homolog on the Y chromosome but have pseudogenes, and in the evolutionary past it is believed that they were true pseudoautosomal genes with identical copies on both the X and Y chromosome (Meroni et al., 1996).

An alignment of the amino acid sequence coded for by BE45772 suggested that it is a fragment of the full length ARSE protein, aligning in the middle of the approximately 600 amino acid ARSE proteins of multiple other species (Supplementary figure 2-8). The putative mouse ARSE is 65% identical to rat and 47% identical to human.

Phylogenetic analysis demonstrated that the mouse ARSE sequence clusters with near certainty with the rat ARSE, however, this putative ARSE clustered within the ARSD proteins, not ARSE as expected (Figure 2-3). Therefore, we propose that we have identified a member of the PAR1 ARS family but at this time cannot determine the exact identity and will refer to this sequence as *Arsd/e*. We note that the long branch-length between the rodent ARS sequences and the remaining ARSD clade may be an artifact due to the short mouse sequence and its high similarity to the rat sequence, which has undergone seemingly accelerated evolutionary change.
Figure 2-3: Phylogenetic tree of arylsulfatase proteins.

Human ARSE (NP_000038) was used as a seed to search the GenBank NR database for orthologs and an approximate maximum-likelihood tree was generated. The putative ARS family gene downregulated in the ATRX-null mouse forebrain clusters closely with rat ARSE, but within the ARSD rather than ARSE protein family (boxed). Entries are annotated with species, chromosome (where known) and GenBank Accession number.
Comparisons to available mouse *Ars* gene family members shows that BE457721 is more similar to *Arse* genes in rat than to other mouse arylsulfatase family members (Supplementary table 2-4), suggesting that we have identified an *Arse* gene. These data, combined with our ability to specifically amplify this transcript from mouse brain cDNA and also from a commercially available E15 cDNA library (data not shown), indicates that we have likely identified the mouse homologue of a previously unidentified mouse *Ars* gene rather than a gene fragment from a known mouse family member.

To further confirm the identity of BE457721, we assessed the outcome of ATRX depletion on *Arsd/e* expression by RNA interference in the Neuro-2a cultured neuroblastoma cell line. Small interfering RNAs (siRNAs) were used to transiently deplete ATRX, as was done previously (Ritchie et al., 2008). Cells transfected with a non-specific siRNA or no siRNA (“Mock”) were used as controls. At 72 hours following siRNA transfection, we monitored the effectiveness of ATRX depletion by indirect immunofluorescence using an ATRX-specific antibody (H300) and qRT-PCR analysis of *Atrx* expression levels using primers that simultaneously amplify both the full length isoform and the reported truncated isoform (Garrick et al., 2004). In the siATRX-treated samples, approximately 95% of cells were negative for ATRX (Figure 2-4A) and *Atrx* transcript levels were depleted by approximately 5 fold (Figure 2-4B). We then used qRT-PCR to determine the outcome of ATRX silencing on the expression level of the *Arsd/e*. Similar to the results obtained in the ATRX-null forebrain, the expression of *Arsd/e* was decreased two fold (Figure 2-4B). These findings support the conclusion that we have identified the mouse *Arsd/e* gene, confirm the regulation of this ancestral PAR gene by ATRX, and show that this outcome on gene expression can be recapitulated in two different systems: *in vivo* in the ATRX-null developing forebrain and *in vitro* in ATRX-depleted cultured neuronal cells.
Figure 2-4: *Arsd/e* transcriptional downregulation is recapitulated in ATRX-depleted cells.

(A) RNA interference was used to deplete ATRX in Neuro-2a neuroblastoma cells. Cells were transfected with 8 nM siRNA, fixed after 72 h and processed for immunofluorescence staining using an anti-ATRX primary antibody (H300) and anti-rabbit Alexa 488 secondary antibody, then counterstained with DAPI to detect nuclei. In the siATRX treated samples, approximately 95% of cells were negative for ATRX. Scale bar = 20 μM. (B) Total RNA was isolated for quantitative real-time PCR of *Atrx* and *Arsd/e* gene expression at 72 hours post-transfection. Mock (transfection reagent only) expression levels were set to one and a non-specific siRNA was used as a control. Results were normalized to β-actin expression levels. Error bars represent standard error of the mean for n = 3. Numbers on siATRX bars indicate p values.
2.3.5 Identification of an ASMTL-like gene

AK007409 is the “RIKEN cDNA 1810009N02” gene and contains a musculoaponeurotic fibrosarcoma (MAF) domain. A multiple sequence alignment of amino acid sequences was used to further determine the identity of AK007409 (Supplementary figure 2-9). AK007409 aligns to the N terminus of ASMTL from multiple other species. The N terminal portion of ASMTL also contains a MAF domain. Human ASMTL was generated by a fusion of a duplicated acetylserotonin O-methyltransferase (ASMT) with the bacterial maf gene (Ried et al., 1998). While AK007409 contains a MAF domain, it lacks the ASMT domain. However, this is similar to the putative rat ASMTL (Accession NP_001099385) which also lacks the ASMT domain. The putative mouse ASMTL is 54% identical to rat, and 51% identical to the human protein.

In contrast to ARSD/E, ASMTL has fewer discernible high-similarity full-length orthologs, and its evolution appears tied to the pseudoautosomal region (Ried et al., 1998). Therefore fewer sequences were available for analysis. Figure 2-5 shows the inferred phylogeny of the ASMTL family, with the primate branches collapsed for clarity. With fairly high bootstrap support, the tree mirrors the known branching of the placental mammals, marsupials, monotremes, birds, amphibians, and fish. The mouse sequence displays the only anomalous placement in the tree, clustering well outside the mammalian clade. Both the placement and the branch-length of the mouse sequence indicate that it is of considerable evolutionarily derived character compared to the putative ancestor, and it appears to have followed an evolutionary path quite distinct from its paralogs. The lack of the ASMT domain in the mouse sequence may also be responsible for the placement of the mouse sequence in the tree. As with ARSD/E, some of this divergence may be due to the availability of a partial mouse sequence, but the sequence remains quite unique, nonetheless.
Figure 2-5: Phylogenetic tree of ASMTL proteins.

Human ASMTL (NP_004183) was used as a seed to search the GenBank NR database for orthologs and an approximate maximum-likelihood tree was generated. The putative mouse ASMTL lacks the ASMT domain and clusters well outside the mammalian clade (at bottom), indicating that it has considerably diverged compared to the putative ancestor. Entries are annotated with species, chromosome (where known) and GenBank Accession number.
2.3.6 Expression of PAR genes regulated by ATRX in the mouse is unchanged upon depletion of ATRX in non-murine cell lines

Our finding of common regulation of a group of genes now dispersed throughout the mouse genome, but connected by their common ancestral location, suggested that ATRX may be able to regulate the expression of these genes regardless of their chromosomal location. To study this, we examined expression of these genes in two species where the PAR genes are located together on the X chromosome. In humans, *ASMTL, CD99, CSF2RA* and *DHRSX* are located in the PAR1 region. *ARSD* and *ARSE* are located just outside the human PAR1 but expression of these genes was not detected in the IMR-90 lung fibroblast cells used in this study (data not shown). In cattle (*Bos taurus*), *Arse, Asmtl, Cd99, and Csf2ra* are all within the PAR1 region, as the PAR1 is larger than in humans, but no *Dhrsx* gene has been identified. Bovine fetal fibroblasts were provided by Dr. Jim Petrik (University of Guelph). *Atrx* was depleted in the two fibroblast cell lines using siRNA that targeted both the full length and truncated forms of *Atrx* at a region conserved in mouse, humans, and cattle. Depletion was assessed using qRT-PCR and immunofluorescence. qRT-PCR showed that 72 hours after siRNA transfection overall transcript levels were depleted to 30% of that seen in mock treated (transfection reagent only) (Figure 2-6B). Immunofluorescence with an antibody that recognizes full length ATRX protein showed that while some cells continued to express ATRX, the majority (>85%) had no or very little detectable ATRX protein (Figure 2-6A).

Expression of the PAR genes was then measured using qRT-PCR, but no change was seen in either human or bovine *Atrx*-depleted fibroblasts (Figure 2-6B). Additionally, in shRNA-mediated ATRX-depleted HeLa cells (Ritchie et al., 2008), PAR gene expression was also unchanged (data not shown). These data suggest that the regulation of PAR genes by ATRX depends on their genomic location, contrary to our initial hypothesis. Therefore, both the current autosomal locations as well as common ancestry of these genes appear essential for their regulation by ATRX.
Figure 2-6: PAR gene expression is unchanged in human and bovine cells.

(A) RNA interference was used to deplete ATRX in human (IMR90) and bovine (fetal fibroblast) cells. Cells were transfected with 10 nM siRNA, fixed after 72 h, and processed for immunofluorescence using an anti-ATRX antibody (H300) and anti-rabbit Alexa 488 secondary antibody, then counterstained with DAPI to detect nuclei. In siATRX samples, approximately 85% of cells were negative for ATRX. Scale bar = 20 μM. (B) Total RNA was isolated for quantitative real-time PCR of *Atrx* and the aPAR genes 72 hours post-transfection. Mock (transfection reagent only) expression levels were set to one. Results were normalized to β-actin expression levels. Error bars represent standard error of the mean for n=3 (* p<0.05).
2.4 Discussion

Mutations in the ATRX gene result in profound cognitive deficits, facial dysmorphisms, as well as skeletal and urogenital abnormalities (Gibbons and Higgs, 2000). Global deletion of Atrx in mouse embryonic stem cells resulted in a growth disadvantage (Garrick et al., 2006), and conditional loss of Atrx beginning at the 8-16 cell stage leads to embryonic lethality by E9.5 (Garrick et al., 2006). To bypass early embryonic lethality, we have previously used a conditional approach to delete Atrx in the mouse forebrain beginning at E8.5. These mice had significantly increased cortical progenitor cell apoptosis causing a reduction in forebrain size and hypocellularity in the neocortex and hippocampus (Bérubé et al., 2005). ATRX is a chromatin remodeling protein (Xue et al., 2003) and has been proposed to regulate gene expression by modulating chromatin structure, but gene targets of ATRX have not yet been reported. We used a microarray approach to perform large-scale analysis of gene expression changes in the ATRX-null versus wild type mouse forebrain at E13.5 and P0.5. The fact that relatively few genes display altered expression indicates that ATRX is not a global regulator of gene expression but likely controls specific gene loci. It is not clear at this point if ATRX acts by binding directly to DNA or through other unidentified factors to upregulate the aPAR genes identified in our study. The only target of ATRX identified to date is α-globin which is downregulated in patients with germline or somatic ATRX mutations (Gibbons et al., 1995), including α-thalassemia myelodysplastic syndrome (Steensma et al., 2005), although at the time of this study, evidence that ATRX directly binds to the α-globin locus is was still lacking.

Through global transcriptional profiling we have now identified a distinct group of genes, the ancestral PAR genes, which are controlled by ATRX in the mouse brain. The human PAR1 contains 24 genes, but only 10 of these have been reported in the mouse genome. Arsd/e, Asmtl, Cd99, Csf2ra, Dhrsx and Shox2 were among the most downregulated genes identified in the ATRX-null embryonic forebrain. Although these genes are unrelated in function, they share a common ancestral location in the PAR1 of the X chromosome millions of years ago. Our findings demonstrate that they have maintained a mechanism of co-regulation that was conserved in evolution and that requires ATRX,
even after their dispersal to autosomes in the mouse genome. However, the absence of an effect on expression in human and bovine cells suggests that common ancestry may be necessary but is not sufficient for these genes to be regulated by ATRX.

The PAR1 region exhibits recombination rates approximately 10 times higher than the rest of the human genome (Lien et al., 2000). Consequently, genes in this region undergo rapid evolution leading to high interspecies divergence (Ellison et al., 1996; Gianfrancesco et al., 2001) making positive identification of homologs difficult. Using multiple sequence alignments and phylogenetic analysis we have identified *Arsd/e* and *Asmtl* as putative novel mouse ancestral PAR transcripts. Identity between mouse and human sequences are 47%, 40% and 51% for *ARSE* (NM_000047), *ARSD* (NM_001669) and *ASMTL* (NM_004192), respectively, which is similar to what was reported for other PAR1 genes. For example, *DHRSX* exhibits 59% protein identity between humans and mice (Gianfrancesco et al., 2001), *CD99* 46% identity (Park et al., 2005), and 35% for *CSF2RA* (Park et al., 1992).

*ARSD* and *ARSE* are members of the arylsulfatase gene family and are located just outside the human PAR1 in a cluster of four arylsulfatase gene family members (Meroni et al., 1996). *ARSE* gene mutations cause X-linked chondrodysplasia punctata, a disorder characterized by abnormalities in cartilage and bone development (Daniele et al., 1998). ARSE may therefore play a role in the skeletal defects seen in patients with the ATR-X syndrome if it is also regulated by ATRX in humans. The role of *ARSD* is unknown and it has no demonstrated sulfatase activity despite its high conservation of the N-terminal domain important for catalytic sulfatase activity (Urbitsch et al., 2000). *ARSE* exhibits a restricted pattern of expression (Franco et al., 1995) while *ARSD* is ubiquitously expressed (Dooley et al., 2000).

The function of human *ASMTL* is unknown. The gene was generated by the duplication of the PAR1 gene *Asmt* which then fused with the bacterial orfE/maf gene (Ried et al., 1998). While other *ASMT* genes involved in the serotonin/N-acetylserotonin/melatonin pathway are expressed specifically in the human brain, pineal gland and retina (Gauer and Craft, 1996), *ASMTL* has a wider expression pattern and may not be involved in this
pathway but could still have methyltransferase activity since it retains the necessary domain (Ried et al., 1998).

We have also identified the mouse Shox2 gene as a potential target of ATRX, and we observed that Shox2 expression levels are highly sensitive to ATRX deficiency in the developing mouse brain. Two SHOX genes, SHOX and SHOX2 have been identified in the human genome, on chromosomes X and 3, respectively. Only one mouse homolog has been identified and is mapped to chromosome 3. Like ARSE, SHOX genes are involved in skeletal development: mutations and deletions in SHOX lead to Leri-Weill dyschondrosteosis (Belin et al., 1998; Shears et al., 1998) and mutations cause the short stature phenotype seen in Turner syndrome (Clement-Jones et al., 2000). SHOX2 is involved in craniofacial and limb development (Blaschke et al., 1998) and SHOX2 mutations lead to cleft palate (Yu et al., 2005). Along with ARSE, the SHOX genes provide an intriguing correlation with the skeletal phenotype of ATR-X patients, and future work should address whether these genes are regulated by ATRX in humans.

Collectively, our findings suggest that even though they are now located on different chromosomes, a large subset of ancestral PAR genes might share a common sequence or factor that was conserved upon translocation from the pseudoautosomal region on the X chromosome to their current autosomal locations in the mouse genome. Uniform regulation of gene expression may be due to similar regulatory features such as common sequences or epigenetic modifications (e.g. CpG islands). Despite the sequencing of the human X chromosome, gaps remain, most notably in the PAR1 region (Ross et al., 2005). The repetitive nature of the PARs likely explains the paucity of sequence data for these regions, and the lack of genomic sequence data for the PAR1 genes that have translocated to autosomes in the mouse. However, we speculate that ATRX could be targeted to repetitive sequences surrounding these genes. One indication that ATRX would preferentially target repetitive sequences comes from studies done in human ATR-X syndrome patients. The analysis of blood samples revealed altered DNA methylation of several highly repeated sequences including ribosomal DNA arrays, the Y-specific repeat DYZ2 and subtelomeric repeats (Gibbons et al., 2000). Conservation of repetitive elements in the PAR1 region of eutherians may have been maintained with the PAR1
genes as they moved to autosomes, and perhaps allow ATRX to target these genes in their modern chromosomal locations.

Future work should focus on identifying the molecular mechanisms by which ATRX can co-regulate this diverse set of genes linked by their ancestral localization in the PAR1 region. This will lead to a better understanding of ATRX function in the regulation of chromatin structure and its effects on gene expression in general.
2.5 Supplementary Figures
A

Increased 1.5 fold
Decreased 1.5 fold
Increased 2 fold
Decreased 2 fold

B

1.5 fold

Probe Set ID

2 fold

Expression

0 0.5 1 2.5 5
Figure 2-7 (supplementary): Summary of microarray results.

Complementary RNA was generated from total forebrain RNA from three pairs of littermate-matched ATRX-null and wild type forebrain tissue and hybridized to an Affymetrix Mouse Genome 430 2.0 Array. Data was analyzed using GeneSpring. Probesets were filtered by fold change (1.5 and 2 fold at both E13.5 and P0.5) and confidence (P<0.05) and duplicate genes were removed. (A) Venn diagrams to categorize probesets according to developmental timepoint and fold change in expression levels. (B) Hierarchical clustering of differentially expressed probesets. Approximately two-thirds of the misregulated genes are upregulated. Ancestral PAR genes are consistently downregulated at both timepoints and are indicated by blue text. Probesets were filtered by 1.5 fold or 2 fold change, P<0.05, at either E13.5 or P0.5. Normalized expression levels are displayed.
Figure 2-8 (supplementary): Amino acid alignment of a small portion of ARSD/E between multiple species.

Sequences were aligned using T-Coffee 5.56 (Notredame et al., 2000) using default parameters, edited using JalView (Clamp et al., 2004) and shaded using Boxshade (Hofmann and Baron, 2007). Mouse ARSD/E has highest identity to rat ARSE (65%).

GenBank accession numbers are ARSE: chicken (XP_416856), cow (ABS45001), dog (NP_001041587), horse (XP_001495573), macaque (Q60HH5), human (CAA58556), platypus (XP_001514429), opossum (XP_001362844), pufferfish (CAG09268), rat (CAI84983). ARSD: dog (XP_548838), horse (XP_001495553), human (CAA58555), macaque (XP_001092405), opossum, (XP_001362931), platypus (XP_001507106), chicken (XP_416855), zebrafish (XP_700386). Putative mouse ARSD/E translated from BE457721.
Figure 2-9 (supplementary): Amino acid alignment of the N terminal of ASMTL between multiple species.

Sequences were aligned using T-Coffee 5.56 (Notredame et al., 2000) using default parameters, edited using JalView (Clamp et al., 2004) and shaded using Boxshade (Hofmann and Baron, 2007). The putative mouse ASMTL aligns within the N terminal MAF domain and is most similar to rat ASMTL (54% identity) which also contains only the MAF domain. GenBank accession numbers are: human (XP_001133965), orangutan (CAH90398), chimpanzee (XP_001137696), cow (AAI03000), dog (XP_851655), frog (NP_001085814), chicken (XP_001231914), zebrafish (NP_998676), platypus (XP_001506357), mouse (NP_081215).
### 2.6 Supplementary tables

**Table 2-2 (supplementary): Conditions for quantitative real-time PCR.**

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Annealing Temp.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arse Forward</td>
<td>GTCGAAGGTCGGGCTACGAAGT</td>
<td>68</td>
</tr>
<tr>
<td>Arse Reverse</td>
<td>GCCCACGAGGTCCTGCTCCACTA</td>
<td>68</td>
</tr>
<tr>
<td>Asmtl Forward</td>
<td>CTTCCACGAGGACGAGCAGG</td>
<td>60</td>
</tr>
<tr>
<td>Asmtl Reverse</td>
<td>CCACAGCGTTGAGGACATC</td>
<td>60</td>
</tr>
<tr>
<td>Cd99 Forward</td>
<td>AGCTTCGTGGCCTATCGC</td>
<td>60</td>
</tr>
<tr>
<td>Cd99 Reverse</td>
<td>ATGAGCAGCGTCAGTGATGT</td>
<td>60</td>
</tr>
<tr>
<td>Csf2ra Forward</td>
<td>ATCAGTACTCGTGCCATCC</td>
<td>60</td>
</tr>
<tr>
<td>Csf2ra Reverse</td>
<td>TTGATCATGAAGGCCAGTTG</td>
<td>60</td>
</tr>
<tr>
<td>DhrsX Forward</td>
<td>GACCTGTGACCTCAAACAT</td>
<td>63</td>
</tr>
<tr>
<td>DhrsX Reverse</td>
<td>CCTCGCACATTCTAGCTC</td>
<td>63</td>
</tr>
<tr>
<td>Shox2 Forward</td>
<td>TTGGTTTCAAATCGAAGAGC</td>
<td>60</td>
</tr>
<tr>
<td>Shox2 Reverse</td>
<td>TGGCATCCTTAAGCACCTAC</td>
<td>60</td>
</tr>
</tbody>
</table>
Table 2-3 (supplementary): Significantly misregulated GO categories.

**E13.5 Upregulated Biological Process GO:0008150**

<table>
<thead>
<tr>
<th>Category</th>
<th>Genes</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:6955: immune response</td>
<td>12</td>
<td>8.80E-07</td>
</tr>
<tr>
<td>GO:6952: defense response</td>
<td>13</td>
<td>1.19E-06</td>
</tr>
<tr>
<td>GO:16064: humoral defense mechanism (sensu Vertebrata)</td>
<td>5</td>
<td>1.61E-06</td>
</tr>
<tr>
<td>GO:9607: response to biotic stimulus</td>
<td>13</td>
<td>1.66E-06</td>
</tr>
<tr>
<td>GO:9613: response to pest, pathogen or parasite</td>
<td>8</td>
<td>1.48E-05</td>
</tr>
<tr>
<td>GO:43207: response to external biotic stimulus</td>
<td>8</td>
<td>2.07E-05</td>
</tr>
<tr>
<td>GO:30888: regulation of B cell proliferation</td>
<td>3</td>
<td>2.34E-05</td>
</tr>
<tr>
<td>GO:6959: humoral immune response</td>
<td>5</td>
<td>3.17E-05</td>
</tr>
<tr>
<td>GO:42100: B cell proliferation</td>
<td>3</td>
<td>3.54E-05</td>
</tr>
<tr>
<td>GO:50864: regulation of B cell activation</td>
<td>3</td>
<td>5.08E-05</td>
</tr>
<tr>
<td>GO:42113: B cell activation</td>
<td>4</td>
<td>5.13E-05</td>
</tr>
<tr>
<td>GO:50670: regulation of lymphocyte proliferation</td>
<td>3</td>
<td>5.68E-05</td>
</tr>
<tr>
<td>GO:6958: complement activation, classical pathway</td>
<td>3</td>
<td>0.000144</td>
</tr>
<tr>
<td>GO:51249: regulation of lymphocyte activation</td>
<td>3</td>
<td>0.000168</td>
</tr>
<tr>
<td>GO:50865: regulation of cell activation</td>
<td>3</td>
<td>0.000168</td>
</tr>
<tr>
<td>GO:50874: organismal physiological process</td>
<td>14</td>
<td>0.000168</td>
</tr>
<tr>
<td>GO:46651: lymphocyte proliferation</td>
<td>3</td>
<td>0.000254</td>
</tr>
<tr>
<td>GO:30889: negative regulation of B cell proliferation</td>
<td>2</td>
<td>0.000312</td>
</tr>
<tr>
<td>GO:50869: negative regulation of B cell activation</td>
<td>2</td>
<td>0.000312</td>
</tr>
<tr>
<td>GO:9605: response to external stimulus</td>
<td>8</td>
<td>0.000344</td>
</tr>
<tr>
<td>GO:6909: phagocytosis</td>
<td>3</td>
<td>0.000432</td>
</tr>
<tr>
<td>GO:6956: complement activation</td>
<td>3</td>
<td>0.000506</td>
</tr>
<tr>
<td>GO:45576: mast cell activation</td>
<td>2</td>
<td>0.000539</td>
</tr>
<tr>
<td>GO:46649: lymphocyte activation</td>
<td>4</td>
<td>0.000616</td>
</tr>
<tr>
<td>GO:50672: negative regulation of lymphocyte proliferation</td>
<td>2</td>
<td>0.000627</td>
</tr>
<tr>
<td>GO:6897: endocytosis</td>
<td>5</td>
<td>0.000723</td>
</tr>
<tr>
<td>GO:51250: negative regulation of lymphocyte activation</td>
<td>2</td>
<td>0.000825</td>
</tr>
<tr>
<td>GO:50866: negative regulation of cell activation</td>
<td>2</td>
<td>0.000825</td>
</tr>
<tr>
<td>GO:45321: immune cell activation</td>
<td>4</td>
<td>0.00087</td>
</tr>
<tr>
<td>GO:1775: cell activation</td>
<td>4</td>
<td>0.000911</td>
</tr>
</tbody>
</table>

**E13.5 Upregulated Molecular Function GO:0003674**

<table>
<thead>
<tr>
<th>Category</th>
<th>Genes</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:19864: IgG binding</td>
<td>2</td>
<td>5.48E-04</td>
</tr>
</tbody>
</table>

**E13.5 Upregulated Cellular Component GO:0005575**

<table>
<thead>
<tr>
<th>Category</th>
<th>Genes</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:5576: extracellular region</td>
<td>21</td>
<td>1.55E-03</td>
</tr>
<tr>
<td>GO:9986: cell surface</td>
<td>3</td>
<td>5.98E-03</td>
</tr>
<tr>
<td>GO:17177: alpha-glucosidase II complex</td>
<td>1</td>
<td>0.0077</td>
</tr>
<tr>
<td>GO:5615: extracellular space</td>
<td>17</td>
<td>0.0161</td>
</tr>
<tr>
<td>GO:9897: external side of plasma membrane</td>
<td>2</td>
<td>0.0194</td>
</tr>
<tr>
<td>GO:5901: caveola</td>
<td>1</td>
<td>0.0204</td>
</tr>
<tr>
<td>GO:16599: caveolar membrane</td>
<td>1</td>
<td>0.0204</td>
</tr>
<tr>
<td>GO:5923: tight junction</td>
<td>2</td>
<td>0.0221</td>
</tr>
<tr>
<td>GO:16327: apicolateral plasma membrane</td>
<td>2</td>
<td>0.0349</td>
</tr>
<tr>
<td>GO:43296: apical junction complex</td>
<td>2</td>
<td>0.0349</td>
</tr>
<tr>
<td>GO:307: cyclin-dependent protein kinase holoenzyme complex</td>
<td>1</td>
<td>0.0355</td>
</tr>
<tr>
<td>GO:5641: nuclear membrane lumen</td>
<td>1</td>
<td>0.0478</td>
</tr>
</tbody>
</table>
### E13.5 Downregulated Biological Process GO:0008150

<table>
<thead>
<tr>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:6353: regulation of transcription, DNA-dependent</td>
</tr>
<tr>
<td>GO:6354: transcription, DNA-dependent</td>
</tr>
<tr>
<td>GO:45449: regulation of transcription</td>
</tr>
<tr>
<td>GO:6139: nucleobase, nucleoside, nucleotide and nucleic acid metabolism</td>
</tr>
<tr>
<td>GO:19219: regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolism</td>
</tr>
<tr>
<td>GO:6350: transcription</td>
</tr>
<tr>
<td>GO:50794: regulation of cellular process</td>
</tr>
<tr>
<td>GO:50794: regulation of cellular process</td>
</tr>
<tr>
<td>GO:7399: nervous system development</td>
</tr>
<tr>
<td>GO:30182: neuron differentiation</td>
</tr>
<tr>
<td>GO:48699: neurogenesis</td>
</tr>
<tr>
<td>GO:51244: regulation of cellular physiological process</td>
</tr>
<tr>
<td>GO:7275: development</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gene</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>7.82E-06</td>
</tr>
<tr>
<td>16</td>
<td>9.98E-06</td>
</tr>
<tr>
<td>16</td>
<td>1.32E-05</td>
</tr>
<tr>
<td>20</td>
<td>1.34E-05</td>
</tr>
<tr>
<td>16</td>
<td>1.43E-05</td>
</tr>
<tr>
<td>16</td>
<td>2.25E-05</td>
</tr>
<tr>
<td>16</td>
<td>2.94E-05</td>
</tr>
<tr>
<td>16</td>
<td>4.17E-05</td>
</tr>
<tr>
<td>7</td>
<td>7.39E-05</td>
</tr>
<tr>
<td>17</td>
<td>0.000397</td>
</tr>
<tr>
<td>6</td>
<td>0.000413</td>
</tr>
<tr>
<td>4</td>
<td>0.000648</td>
</tr>
<tr>
<td>3</td>
<td>0.000753</td>
</tr>
<tr>
<td>16</td>
<td>0.00083</td>
</tr>
<tr>
<td>12</td>
<td>0.000933</td>
</tr>
</tbody>
</table>

### E13.5 Downregulated Molecular Function GO:0003674

<table>
<thead>
<tr>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:3700: transcription factor activity</td>
</tr>
<tr>
<td>GO:30528: transcription regulator activity</td>
</tr>
<tr>
<td>GO:3677: DNA binding</td>
</tr>
<tr>
<td>GO:3676: nucleic acid binding</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gene</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>2.79E-09</td>
</tr>
<tr>
<td>15</td>
<td>9.01E-08</td>
</tr>
<tr>
<td>17</td>
<td>3.64E-06</td>
</tr>
<tr>
<td>20</td>
<td>1.70E-05</td>
</tr>
</tbody>
</table>

### E13.5 Downregulated Cellular Component GO:0005575

<table>
<thead>
<tr>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:5654: nucleoplasm</td>
</tr>
<tr>
<td>GO:5667: transcription factor complex</td>
</tr>
<tr>
<td>GO:43234: protein complex</td>
</tr>
<tr>
<td>GO:5634: nucleus</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gene</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>1.01E-08</td>
</tr>
<tr>
<td>11</td>
<td>1.11E-08</td>
</tr>
<tr>
<td>14</td>
<td>0.000428</td>
</tr>
<tr>
<td>21</td>
<td>0.000469</td>
</tr>
</tbody>
</table>

### P0.5 Upregulated Biological Process GO:0008150

<table>
<thead>
<tr>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:7155: cell adhesion</td>
</tr>
<tr>
<td>GO:6956: complement activation</td>
</tr>
<tr>
<td>GO:7178: transmembrane receptor protein serine/threonine kinase signaling pathway</td>
</tr>
<tr>
<td>GO:30509: BMP signaling pathway</td>
</tr>
<tr>
<td>GO:7167: enzyme linked receptor protein signaling pathway</td>
</tr>
<tr>
<td>GO:30199: collagen fibril organization</td>
</tr>
<tr>
<td>GO:6958: complement activation, classical pathway</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gene</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>1.05E-06</td>
</tr>
<tr>
<td>4</td>
<td>0.000181</td>
</tr>
<tr>
<td>5</td>
<td>0.00043</td>
</tr>
<tr>
<td>3</td>
<td>0.000495</td>
</tr>
<tr>
<td>8</td>
<td>0.000605</td>
</tr>
<tr>
<td>2</td>
<td>0.000639</td>
</tr>
<tr>
<td>3</td>
<td>0.000824</td>
</tr>
</tbody>
</table>

### P0.5 Upregulated Molecular Function GO:0003674

<table>
<thead>
<tr>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:30020: extracellular matrix structural constituent conferring tensile strength</td>
</tr>
<tr>
<td>GO:5201: extracellular matrix structural constituent</td>
</tr>
<tr>
<td>GO:8289: lipid binding</td>
</tr>
<tr>
<td>GO:4528: phosphodiesterase I activity</td>
</tr>
<tr>
<td>GO:30332: cyclin binding</td>
</tr>
<tr>
<td>GO:5198: structural molecule activity</td>
</tr>
<tr>
<td>GO:8083: growth factor activity</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gene</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>3.64E-06</td>
</tr>
<tr>
<td>6</td>
<td>7.99E-06</td>
</tr>
<tr>
<td>8</td>
<td>0.000353</td>
</tr>
<tr>
<td>2</td>
<td>0.000571</td>
</tr>
<tr>
<td>2</td>
<td>0.000571</td>
</tr>
<tr>
<td>14</td>
<td>0.000596</td>
</tr>
<tr>
<td>6</td>
<td>0.000689</td>
</tr>
<tr>
<td>Category</td>
<td>Genes</td>
</tr>
<tr>
<td>-------------------------------------------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>GO:16722: oxidoreductase activity, oxidizing metal ions</td>
<td>2</td>
</tr>
<tr>
<td>GO:16724: oxidoreductase activity, oxidizing metal ions, oxygen as acceptor</td>
<td>2</td>
</tr>
<tr>
<td>GO:4322: ferroxidase activity</td>
<td>2</td>
</tr>
<tr>
<td>GO:4551: nucleotide diphosphatase activity</td>
<td>2</td>
</tr>
</tbody>
</table>

**P0.5 Upregulated Cellular Component GO:0005575**

<table>
<thead>
<tr>
<th>Category</th>
<th>Genes</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:5615: extracellular space</td>
<td>58</td>
<td>1.71E-16</td>
</tr>
<tr>
<td>GO:5576: extracellular region</td>
<td>60</td>
<td>6.76E-16</td>
</tr>
<tr>
<td>GO:5578: extracellular matrix (sensu Metazoa)</td>
<td>15</td>
<td>1.17E-07</td>
</tr>
<tr>
<td>GO:31012: extracellular matrix</td>
<td>15</td>
<td>1.38E-07</td>
</tr>
<tr>
<td>GO:5581: collagen</td>
<td>5</td>
<td>8.60E-06</td>
</tr>
<tr>
<td>GO:5911: intercellular junction</td>
<td>7</td>
<td>6.96E-05</td>
</tr>
<tr>
<td>GO:30054: cell junction</td>
<td>8</td>
<td>7.97E-05</td>
</tr>
<tr>
<td>GO:5604: basement membrane</td>
<td>4</td>
<td>0.000663</td>
</tr>
</tbody>
</table>

**P0.5 Downregulated Biological Process GO:0008150**

<table>
<thead>
<tr>
<th>Category</th>
<th>Genes</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:48666: neuron development</td>
<td>4</td>
<td>0.000148</td>
</tr>
<tr>
<td>GO:30182: neuron differentiation</td>
<td>4</td>
<td>0.000262</td>
</tr>
<tr>
<td>GO:48468: cell development</td>
<td>4</td>
<td>0.000702</td>
</tr>
<tr>
<td>GO:7399: nervous system development</td>
<td>5</td>
<td>0.001</td>
</tr>
</tbody>
</table>

**P0.5 Downregulated Molecular Function GO:0003674**

<table>
<thead>
<tr>
<th>Category</th>
<th>Genes</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:5509: calcium ion binding</td>
<td>8</td>
<td>2.13E-05</td>
</tr>
<tr>
<td>GO:5102: receptor binding</td>
<td>6</td>
<td>0.000193</td>
</tr>
<tr>
<td>GO:4871: signal transducer activity</td>
<td>12</td>
<td>0.000287</td>
</tr>
<tr>
<td>GO:4970: ionotropic glutamate receptor activity</td>
<td>2</td>
<td>0.000539</td>
</tr>
<tr>
<td>GO:5179: hormone activity</td>
<td>3</td>
<td>0.000651</td>
</tr>
<tr>
<td>GO:4972: N-methyl-D-aspartate selective glutamate receptor activity</td>
<td>1</td>
<td>0.00118</td>
</tr>
<tr>
<td>GO:8066: glutamate receptor activity</td>
<td>2</td>
<td>0.0023</td>
</tr>
<tr>
<td>GO:1664: G-protein-coupled receptor binding</td>
<td>2</td>
<td>0.00253</td>
</tr>
<tr>
<td>GO:16623: oxidoreductase activity, acting on the aldehyde or oxo group of donors, oxygen as acceptor</td>
<td>1</td>
<td>0.00473</td>
</tr>
<tr>
<td>GO:4031: aldehyde oxidase activity</td>
<td>1</td>
<td>0.00473</td>
</tr>
<tr>
<td>GO:4723: calcium-dependent protein serine/threonine phosphatase activity</td>
<td>1</td>
<td>0.00473</td>
</tr>
<tr>
<td>GO:5262: calcium channel activity</td>
<td>2</td>
<td>0.00498</td>
</tr>
</tbody>
</table>

**P0.5 Downregulated Cellular Component GO:0005575**

<table>
<thead>
<tr>
<th>Category</th>
<th>Genes</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Table 2-4 (supplementary): Pairwise comparisons of arylsulfatase family members.**

<table>
<thead>
<tr>
<th></th>
<th>Arse</th>
<th>Arsa</th>
<th>Arsb</th>
<th>Arsc/Sts</th>
<th>Arsg</th>
<th>Arsi</th>
<th>Arsj</th>
<th>Arsk</th>
<th>Arse</th>
<th>Arsc/Sts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arse</td>
<td>-</td>
<td>50</td>
<td>54</td>
<td>60</td>
<td>50</td>
<td>51</td>
<td>51</td>
<td>49</td>
<td>66</td>
<td>54</td>
</tr>
<tr>
<td>Arsa</td>
<td>50</td>
<td>-</td>
<td>49</td>
<td>50</td>
<td>51</td>
<td>53</td>
<td>49</td>
<td>48</td>
<td>51</td>
<td>51</td>
</tr>
<tr>
<td>Arsb</td>
<td>54</td>
<td>49</td>
<td>-</td>
<td>48</td>
<td>48</td>
<td>59</td>
<td>53</td>
<td>49</td>
<td>49</td>
<td>48</td>
</tr>
<tr>
<td>Arsc/Sts</td>
<td>60</td>
<td>50</td>
<td>48</td>
<td>-</td>
<td>47</td>
<td>49</td>
<td>51</td>
<td>46</td>
<td>63</td>
<td>67</td>
</tr>
<tr>
<td>Arsg</td>
<td>50</td>
<td>51</td>
<td>48</td>
<td>47</td>
<td>-</td>
<td>49</td>
<td>48</td>
<td>47</td>
<td>49</td>
<td>50</td>
</tr>
<tr>
<td>Arsi</td>
<td>51</td>
<td>53</td>
<td>59</td>
<td>49</td>
<td>49</td>
<td>-</td>
<td>63</td>
<td>49</td>
<td>50</td>
<td>51</td>
</tr>
<tr>
<td>Arsj</td>
<td>51</td>
<td>49</td>
<td>53</td>
<td>51</td>
<td>48</td>
<td>63</td>
<td>-</td>
<td>49</td>
<td>49</td>
<td>48</td>
</tr>
<tr>
<td>Arsk</td>
<td>49</td>
<td>48</td>
<td>49</td>
<td>46</td>
<td>47</td>
<td>49</td>
<td>49</td>
<td>-</td>
<td>51</td>
<td>51</td>
</tr>
<tr>
<td>Rat Arse</td>
<td>66</td>
<td>51</td>
<td>49</td>
<td>63</td>
<td>49</td>
<td>50</td>
<td>49</td>
<td>51</td>
<td>-</td>
<td>62</td>
</tr>
<tr>
<td>Rat Arsc/Sts</td>
<td>54</td>
<td>51</td>
<td>48</td>
<td>67</td>
<td>50</td>
<td>51</td>
<td>48</td>
<td>51</td>
<td>62</td>
<td>-</td>
</tr>
</tbody>
</table>

*Arse* [GenBank:BE457721], *Arsa* [GenBank:NM_009713], *Arsb* [GenBank:NM_009712.3], *Arsc/Sts* [GenBank:NM_009293.1], *Arsg* [GenBank:NM_028710.2], *Arsi* [GenBank:NM_001038499.1], *Arsj* [GenBank:NM_173451.2], *Arsk* [GenBank:NM_028847.4], *rat Arse* [GenBank:NM_001047885.1], *rat Arsc/Sts* [GenBank:NM_012661.1]
2.7 References


Chapter 3

3 ATRX aids transcription elongation through G-rich gene segments in a process involving histone H3.3 incorporation

The aim of the work outlined in this chapter was to identify mechanisms by which ATRX modulates gene expression. The aPAR gene Dhrsxx was used as a model gene for this purpose. My findings suggest that ATRX can regulate the expression of genes by influencing transcriptional elongation. This represents the first evidence of such a function for ATRX and may have important implications in understanding how ATRX mutations lead to cognitive deficits and brain tumors in humans.

3.1 Introduction

Decreased activity of the SWI2/SNF2 chromatin remodeling protein ATRX, as a result of mutations in the corresponding gene, causes a rare syndrome called alpha thalassemia mental retardation, X linked (the ATR-X syndrome) (Gibbons et al., 1995). Affected individuals exhibit intellectual disabilities, facial dysmorphisms, genital abnormalities and alpha thalassemia, among other less common phenotypes (Gibbons, 2006). At a molecular level, ATR-X patients show aberrant DNA methylation of repetitive DNA regions, including ribosomal DNA, subtelomeric regions, and DYZ2 repeats (Gibbons et al., 2000). More recently, ATRX mutations have been identified in pancreatic neuroendocrine tumors (de Wilde et al., 2012; Heaphy et al., 2011; Jiao et al., 2011) and brain cancer (Cheung Nv and et al., 2012; Heaphy et al., 2011; Jiao et al., 2012; Khuong-Quang et al., 2012; Molenaar et al., 2012; Schwartzentruber et al., 2012). Tumorigenic cells lacking ATRX expression typically exhibit alternative lengthening of telomeres (ALT), a recombination process that maintains telomere length and proliferative capacity in the absence of the telomerase enzyme (Cheung Nv and et al., 2012; de Wilde et al., 2012; Heaphy et al., 2011; Jiao et al., 2012; Schwartzentruber et al., 2012). ALT-positive cell lines frequently lack ATRX, although artificial depletion of ATRX does not itself lead to ALT (Lovejoy et al., 2012).
ATRX contains a conserved ATPase/helicase domain and a zinc finger ATRX-DNMT3A/B-DNMT3L (ADD) domain that is shared with de novo methyltransferases (Aapola et al., 2000). It is enriched at heterochromatic regions including the inactive X chromosome (Baumann and De La Fuente, 2008), pericentromeric heterochromatin (McDowell et al., 1999), and telomeres (Law et al., 2010). At pericentromeric heterochromatin and telomeres, ATRX forms a complex with the death-associated protein (DAXX) and deposits the replication-independent histone variant H3.3 (Goldberg et al., 2010; Lewis et al., 2010; Wong et al., 2010). However, the deposition of H3.3 at transcribed genes and regulatory elements is believed to occur independently of ATRX (Goldberg et al., 2010). Targeting of ATRX to heterochromatin is mediated by binding of the ADD domain to the heterochromatic histone mark H3K9me3 in the absence of H3K4 methylation (Dhayalan et al., 2011; Eustermann et al., 2011; Iwase et al., 2011). Enrichment of ATRX at telomeric ‘TTAGGG’ tandem repeats may be mediated by the ability of ATRX to bind G-quadruplexes (Law et al., 2010). G-quadruplexes are short sequences containing four G triplicates that form a four stranded secondary structure upon DNA denaturation during replication or transcription (Duquette et al., 2004). Expression of the non-coding telomeric repeat-containing RNA (TERRA) (Azzalin et al., 2007; Schoeftner and Blasco, 2008) from telomeric tandem repeats is increased in the absence of ATRX in mouse ES cells (Goldberg et al., 2010).

We previously reported that loss of ATRX in the mouse forebrain causes reduced expression of ancestral pseudoautosomal (aPAR) genes (Levy et al., 2008). These genes are located on autosomes in the mouse, but are pseudoautosomal (located on the X and Y chromosomes and expressed biallelically in males and females) in humans and in the common evolutionary ancestors between humans and mice. These aPAR genes are amongst the most downregulated genes in the ATRX-null mouse forebrain regardless of the developmental time point (Levy et al., 2008). The most affected genes are dehydrogenase/reductase (SDR family) X-linked (Dhrxs) and colony stimulating factor 2 receptor, alpha (Csf2ra). These genes are located at the subtelomeric region of mouse chromosomes 4 and 19, respectively. Other downregulated aPAR genes Asmit and Cd99 have not yet been mapped to the mouse genome, indicating that they are likely positioned
within highly repetitive regions that are notoriously difficult to sequence and place within existing genomic assemblies.

A direct mechanism for the regulation of gene transcription by ATRX has not yet been identified and could be relevant to the function of ATRX in the central nervous system and for its tumor suppressive activities. Here we report that ATRX and H3.3 are enriched at G-rich regions within the gene bodies of Dhrsx and other aPAR genes in the mouse brain. Loss of ATRX causes decreased H3.3 occupancy within the gene body, without affecting TERRA levels, histone modifications or DNA methylation. Importantly we provide evidence that in the absence of ATRX, RNA polymerase II progression is impeded at the G-rich region of Dhrsx. We propose a model whereby ATRX facilitates transcription elongation of particular target genes by assisting the passage of the transcription machinery through G-rich templates.

3.2 Materials and Methods

3.2.1 Mouse husbandry and genotyping

The Atrx gene was conditionally deleted in the mouse forebrain and mice were genotyped as previously described (Bérubé et al., 2005; Seah et al., 2008). To obtain E13.5 embryos, midday of vaginal plug discovery was considered E0.5. All animal studies were conducted in compliance with the regulations of The Animals for Research Act of the province of Ontario, the guidelines of the Canadian Council on Animal Care, and the policies and procedures approved by the University of Western Ontario Council on Animal Care.

3.2.2 Mouse embryonic fibroblast isolation, culture and viral infection

Primary mouse embryonic fibroblasts (MEFs) were isolated from male E13.5 AtrxloxP embryos. Internal organs and heads were removed, embryos were minced, digested in trypsin at 37°C for 45 minutes, then cultured in DMEM (Sigma-Aldrich) with 10% FBS (Sigma-Aldrich), Pen Strep (Gibco), and GlutaMAX (Invitrogen). Experiments were performed at passages 3 to 6. To delete Atrx, MEFs at 30-40% confluence were
incubated with 50 MOI Ad-Cre-GFP, or Ad-CMV-GFP (Vector Biolabs) and assayed 72 hours post-infection.

### 3.2.3 RNA isolation and transcriptional assays

Forebrains were dissected, flash frozen, and stored at -80°C and MEFs were grown in culture and scraped off the plates. Total RNA was isolated using the RNeasy Mini Kit (Qiagen) and cytoplasmic and nuclear RNA was extracted using the RNeasy Plus Mini Kit (Qiagen) in conjunction with the supplementary protocol, “Purification of cytoplasmic RNA from animal cells using the RNeasy Mini Kit” (Qiagen). For expression analysis of \( \text{Atrx, Csf2ra, and Dhrs} \), cDNA was synthesized from 1 µg of RNA using the SuperScript II Reverse Transcriptase kit (Invitrogen) according to the manufacturer’s instructions. For non-coding RNA expression analysis, cDNA was synthesized from 0.5 µg of cytoplasmic or nuclear RNA using the SuperScript III Reverse Transcriptase kit (Invitrogen). PCR of cDNA was performed as follows: 95°C for 10 seconds, anneal for 20 seconds, 72°C for 30 seconds, for 30-40 cycles to ensure amplification in the linear range. Real-time PCR was performed in duplicate on a Chromo4 Continuous Fluorescence Detector (Bio-Rad) in the presence of iQ SYBR Green Supermix and recorded using the Opticon Monitor 3 software (Bio-Rad). Samples were amplified for 35-40 cycles as follows: 95°C for 10 seconds, annealed for 20 seconds, 72°C for 30 seconds. To ensure amplicon purity, a melting curve was generated and samples were resolved on an agarose gel. Gene expression was normalized to expression of \( \beta\)-actin and/or \( \text{Gapdh} \). Normalization and relative gene expression levels were calculated using the \( \Delta \Delta \text{Ct} \) method with the Excel Gene Expression macro (Bio-Rad). Primer sequences are listed in Supplementary Table 3-1.

### 3.2.4 RNA dot blots

RNA was manually spotted onto Amersham Hybond-XL nylon membrane (GE Healthcare Life Sciences) according to the manufacturer’s instructions. \( \text{P} \) end-labelled probes were generated using T4 polynucleotide kinase (NEB) along with ATP[\( \gamma \text{P} \) 3000Ci/mmol (PerkinElmer) and oligos complementary to TERRA or \( \text{Gapdh} \) (Zhang et al., 2009), then purified using illustra MicroSpin G-25 columns (GE Healthcare Life Sciences).
Membranes were pre-hybridized in Amersham Rapid-Hyb Buffer (GE Healthcare Life Sciences) for 20 minutes at 42°C, hybridized with 25 µL labelled TERRA probe at 42°C for 1 hour, washed according to the Hybond-XL instructions, then exposed to Amersham Hyperfilm MP (GE Healthcare Life Sciences) at -80°C for 6 hours to 14 days to generate multiple exposures. Blots were stripped with boiling 0.1% SDS then washed in 2X SSC, exposed to film to ensure removal of probes, hybridized with the Gapdh probe, and re-exposed. Dots on non-saturated exposures were quantified using the integrated density method with ImageJ (Rasband, 2012), and normalized to Gapdh levels.

3.2.5 Bisulfite mutagenesis and sequencing

Genomic DNA was isolated from neonatal mouse cortices using the DNeasy Blood & Tissue Kit (Qiagen). Bisulfite treatment was performed essentially as described (Market-Velker et al., 2010), except the starting material used was purified gDNA, and the final DNA was purified using the QIAquick Gel Extraction Kit (Qiagen). Nested PCR (see Supplementary Table 3-1 for primer sequence) was performed, including a 72°C, 10 min step at the end of the second round of PCR to add adenine overhangs. Purified amplicons were ligated into a pGEM-T Easy vector using T4 DNA Ligase (both Promega) and clone inserts were sequenced at the London Regional Genomics Centre, London, Canada, or Bio Basic Inc., Markham, Canada.

3.2.6 Chromatin immunoprecipitation

All reagents are from Sigma-Aldrich unless otherwise noted. ChIP was performed as previously described (Kernohan et al., 2010) except that the cells were fixed at 37°C, the LiCl wash was omitted following the IP step and all tissues were fixed fresh. For ChIP-sequencing, samples were pre-cleared for 1 hour with 20 µL of ChIP-Grade Protein G Magnetic Beads (Cell Signaling). 1/25th the amount of chromatin used for IP was kept for input. 3 µg of antibody (15 µg for H300 and D5, or 3 µL where concentration was not specified) was added and immunoprecipitation reactions were incubated 16-20 hours at 4°C. Antibodies (from Millipore unless noted) used were: rabbit IgG (PP64B), H4Ac (06-866), H3K4Me3 (04-745), H3K9Me3 (07-442), H3K27Me3 (07-449), Histone H3.3
(17-10245), RNA PolII (05-623), mouse IgG (Santa Cruz sc-2025), ATRX D5 (Santa Cruz, sc-55584), ATRX H300 (Santa Cruz, sc-15408). Immunoprecipitated samples were incubated with 20 µL Magnetic Beads for 2 hours to collect the chromatin/antibody complexes, beads were washed, chromatin/antibody complexes eluted, and DNA purified as previously described (Kernohan et al., 2010). Real-time PCR reactions were performed as described above, and percent input was calculated as previously described (Kernohan et al., 2010).

3.2.7 Next generation sequencing and analysis

For ChIP-seq, DNA was sent to The Centre for Applied Genomics at the Hospital for Sick Children, Toronto, Canada, and 30-40 million 100 base pair, paired-end reads were generated for each sample using an Illumina HiSeq 2000. Sequences (cDNA, since genomic DNA sequences are unavailable) for Asmtl (4713 bp; AK084779 and AK010990), and Cd99 (1976 bp; NW_016967) were added to the mouse genome version mm9 (including ‘random’ chromosomes) and a Bowtie index (Langmead et al., 2009) was created. Raw sequencing data for ATRX and H3.3 ESC ChIP-seq was downloaded from the NCBI Sequence Read Archive (Accession numbers GSE22162 and GSE16893). H3.3 ChIP-seq data for control and ATRX-null forebrain tissue was generated as described above. Quality control was done using FastQC (Andrews, 2012) and reads were aligned to the mouse genome using Bowtie version 0.12.8 in the -n alignment mode. During alignment up to 3 mismatches were allowed and reads that aligned to more than one location were discarded. Duplicate sequences were removed. Genome-wide coverage tracks (.wig files) were generated using custom Perl scripts to extend reads to their fragment lengths (for single end reads, average fragment length was determined using SISSRS (Jothi et al., 2008)) and to normalize coverage to 20 million reads. Data was then viewed using the UCSC Genome Browser (Kent et al., 2002). G-quadruplexes were identified using Quadfinder (Scaria et al., 2006) with default settings.

Custom Perl scripts were used to perform the following analysis. Genes enriched for ATRX and H3.3 were identified by dividing each gene in the control sample into 100 bp bins, then looking for bins that had a minimum average coverage of 10 reads and a minimum enrichment of five-fold over input for both ATRX and H3.3. Sites found
enriched in the control were then compared with the corresponding location in the ATRX-null sample to look for sites with a 1.5 fold or greater decrease in H3.3 enrichment. Genes were considered positive hits if they had at least two bins within 500 bp fulfilling the above criteria. Genomic DNA sequences within the identified regions were then screened for high GC content (70% G or C) overlapping the ATRX/H3.3 binding site. Telomere enrichment was determined by looking for sequence reads that consisted of at least 80% telomere repeats (TTAGGG or CCCTAA). Genome-wide distribution of H3.3 was determined by dividing each chromosome into 500 segments, summing the number of reads within each segment, then calculating average enrichment within each segment across mouse chromosomes 1-19 and X.

3.3 Results

3.3.1 ATRX is required for the normal expression of Dhrsx and flanking non-coding RNAs in mouse cells

We have previously shown that ATRX regulates the expression of Dhrsx, Csf2ra and other aPAR genes in the mouse brain (Levy et al., 2008). To determine the genomic location of Dhrsx, we examined the mouse genome (Waterston et al., 2002) using the UCSC Genome Browser (Kent et al., 2002). Dhrsx is most likely subtelomeric on chromosome 4 (Gianfrancesco et al., 2001), and while mouse genomes up to and including version mm10 place it on an unassembled fragment of chromosome 4, the alternate assembly Mm_Celera (Mural et al., 2002) places Dhrsx approximately 110 kb from the end of an incomplete chromosome 4 (accession AC_000026.1). While the precise distance from the telomere is unknown, there are high concentrations of telomere sequence repeats (TTAGGG) flanking the gene (Figure 3-1A). These likely represent interstitial telomeric sequences (ITSs) (Lin and Yan, 2008) as opposed to actual telomeres, since the repeats are generally not in tandem and do not continue to the end of the fragments. Along with the ITSs, the UCSC Genome Browser shows non-coding ESTs and mRNAs flanking Dhrsx, in particular in the 5’ region overlapping the larger concentration of ITSs (Figure 3-1A). We therefore wondered if non-coding RNAs are transcribed from the regions flanking Dhrsx in the mouse brain, and if so, whether this expression correlates with expression of the Dhrsx gene itself in the absence of ATRX.
Figure 3-1: Expression of non-coding RNAs flanking Dhrsx is decreased in the Atrx-null mouse forebrain.

(A) Analysis of 50 kb of genomic DNA around Dhrsx reveals several mRNAs, expressed sequence tags (ESTs), and interstitial telomere sequences (ITS) both upstream and downstream of the Dhrsx gene. (B) RT-qPCR of RNA isolated from the P17 mouse forebrain showed the presence of transcripts from approximately 7 kb upstream of Dhrsx to 5 kb downstream, with higher levels 5'. Highest expression was seen with primers that also detect exon 1 of Dhrsx. (C) Semi-quantitative RT-PCR showed nuclear enrichment of transcripts flanking Dhrsx. Cyto, cytoplasmic fraction. Nuc, nuclear fraction. (D) RT-qPCR of RNA isolated from control and ATRX-null P17 forebrains showed decreased expression of all sites analyzed at Dhrsx in the absence of ATRX. In (B) and (D), error bars represent standard error of the mean for n=3. *p-value < 0.05.
To look for expression from regions surrounding *Dhrsx*, we designed primers to sites upstream and downstream of the gene (locations indicated in Figure 3-1A), and used quantitative reverse transcription polymerase chain reaction (qRT-PCR) to measure transcript levels. Expression was normalized to amplification from gDNA to account for primer efficiency. We observed moderate expression from up to 7 kb upstream of the *Dhrsx* start site and low expression up to 5 kb downstream of the gene (Figure 3-1B). As a control, primers were designed to the first exon of *Dhrsx*. This site exhibited highest expression, but it is not clear whether these primers are detecting just the *Dhrsx* transcript, or the *Dhrsx* transcript along with transcripts initiating upstream of *Dhrsx* which overlap with the gene. Non-quantitative RT-PCR showed that transcripts from sites flanking *Dhrsx* are generally enriched in the nucleus (Figure 3-1C), suggesting that these transcripts may be non-coding RNAs (Forrest et al., 2009). When *Atrx* was deleted, we found a decrease in expression at most sites, although not all decreases were statistically significant (Figure 3-1D). These results demonstrate that transcripts are generated from within some of the ITSs, and that loss of ATRX affects a region larger than just the *Dhrsx* gene itself.

### 3.3.2 ATRX deficiency in the mouse forebrain or in MEFs does not affect TERRA levels

Loss of ATRX in mouse embryonic stem cells (mESCs) was reported to induce a ~1.7 fold increase in levels of non-coding telomeric repeat-containing RNA (TERRA) (Goldberg et al., 2010). TERRA transcription starts in the subtelomere (Azzalin et al., 2007) and in human cells is driven by repetitive CpG-island promoters within 1 kb of the start of the telomeric hexamers (Nergadze et al., 2009). Increased TERRA expression can promote heterochromatin formation at telomeres by binding to HP1α and H3K9Me3 (Deng et al., 2009), proteins that interact with ATRX (Dhayalan et al., 2011; McDowell et al., 1999). The downregulation of subtelomeric genes in the ATRX-null brain may therefore be due to their proximity to telomeres, perhaps as part of the telomere position effect (TPE) (Baur et al., 2001; Gottschling et al., 1990; Pedram et al., 2006).

A series of radioactive RNA dot blots were performed to assess TERRA transcript levels (Figure 3-2A). The results did not show changes in the levels of TERRA transcripts in
the P17 ATRX-null forebrain (Figure 3-2B). To confirm these findings, we examined this phenomenon in mouse embryonic fibroblasts (MEFs). We used adenoviral vectors expressing Cre recombinase to delete Atrx in MEFs isolated from Atrx-floxed mice (Bérubé et al., 2005). qRT-PCR and immunofluorescence assays revealed nearly undetectable levels of Atrx RNA and protein upon Cre recombinase expression (Supplementary Figure 3-10 A,B). In addition, both Csf2ra and Dhrsx showed a 2.7 fold decrease in transcript levels in ATRX-null MEFs (Supplementary Figure 3-10B), a result analogous to that seen in the mouse forebrain (Levy et al., 2008). Thus loss of ATRX in MEFs recapitulates the gene expression effects observed in vivo in brain tissue. The expression of TERRA in wildtype MEFs was considerably lower than in forebrain tissue and MEF blots had to be exposed for approximately four times longer than the forebrain RNA blots to generate comparable signal intensity. Again, we observed that the levels of TERRA transcripts were comparable between control and Ad-Cre treated cells (Figure 3-2C,D). To completely rule out an effect of ATRX loss on TERRA levels, we analyzed RNA sequencing data of control (109,539,088 sequence reads) and Atrx-null (84,639,248 sequence reads) E14 forebrains (A. Watson, unpublished data). TERRA RNA (sequences made up of (TTAGGG)$_{10}$ or (CCCTAA)$_{10}$) constituted 2.54x10$^{-4}$% of all sequence reads in control and 2.95x10$^{-4}$% of all sequence reads in the Atrx-null forebrain. Therefore, we detect only a 1.16 fold increase in TERRA levels in the Atrx-null embryonic forebrain, strongly suggesting that TERRA levels are not affected by ATRX, and that TERRA is regulated differently in the mouse brain and fibroblasts compared to ESCs. Moreover, Dhrsx gene silencing upon loss of ATRX cannot be explained by a change in TERRA transcript levels.
Figure 3-2: TERRA expression is not altered in forebrain tissue or mouse embryonic fibroblasts lacking ATRX.

(A) $^{32}$P-labelled probes for TERRA and Gapdh were hybridized to membranes spotted with cytoplasmic and nuclear RNA from control or ATRX-null P17 mouse forebrains. A representative experiment is shown from n=4. The non-coding TERRA was enriched in the nuclear fraction while Gapdh was equal between fractions. (B) Quantification of spot intensity showed no change in TERRA expression between control and ATRX-null forebrains. Error bars represent standard error of the mean for n=4. (C) $^{32}$P-probes for TERRA and Gapdh were hybridized to membranes spotted with cytoplasmic and nuclear RNA from uninfected MEFs, MEFs infected with a control adenovirus expressing GFP, or MEFs infected with an adenovirus expressing Cre recombinase to delete Atrx. A representative experiment is shown from n=2. The non-coding TERRA was enriched in the nuclear fraction while Gapdh was equal between fractions. (D) Quantification showed no change in TERRA expression between control and ATRX-null forebrains. Error bars represent the range for n=2.
3.3.3 ATRX and H3.3 are enriched within the gene body of *Dhrsx*

We next investigated the possibility that ATRX localizes near or within the aPAR genes in the mouse brain. To study this, we first analyzed data from previously published ATRX ChIP sequencing (ChIP-seq) done in ESCs (Law et al., 2010). Our analysis demonstrated that ATRX is highly enriched across a 5 kb region of the *Dhrsx* gene body (Figure 3-3A). Importantly, this segment of the gene is G-rich and has a high potential of forming G-quadruplex structures as predicted by Quadfinder (Scaife, 2005) (Figure 3-3A). To determine whether ATRX also binds within the *Dhrsx* gene in the mouse brain we performed ChIP-qPCR for ATRX in the neonatal forebrain. ChIP with two different antibodies showed that ATRX is bound to *Dhrsx*, similar to the results in ESCs (Figure 3-3B).

It was previously established that ATRX is required for the deposition of H3.3 at telomeres (Goldberg et al., 2010). We therefore examined H3.3 distribution at *Dhrsx*. To achieve this, we analyzed the available ChIP-seq data from mouse ESCs which used EYFP-tagged H3.3 (Goldberg et al., 2010) as well as our own H3.3 ChIP-seq data in the neonatal mouse forebrain done with an H3.3-specific antibody. We were surprised to find that H3.3 enrichment mirrored that of ATRX at *Dhrsx* (Figure 3-3A). We confirmed enrichment of H3.3 across the 5 kb G-rich region of *Dhrsx* by ChIP-qPCR (Supplementary Figure 3-11A). These findings show that H3.3 incorporation within gene body nucleosomes corresponds to the presence of ATRX at the same chromatin region.

3.3.4 Reduced levels of H3.3 at *Dhrsx* and other aPAR genes in the absence of ATRX correlates with decreased gene expression

Since ATRX and H3.3 co-localize within a gene that exhibits decreased expression in the absence of ATRX, we examined whether loss of ATRX affects H3.3 deposition at that site. Indeed, we observed a substantial loss of H3.3 at *Dhrsx* in ATRX-null mouse ESCs and forebrain (Figure 3-3A). ChIP-qPCR for H3.3 at *Dhrsx* showed loss of H3.3 at both P0.5 and P17 (Figures 3-3C and Supplementary Figure 11-3B). Results obtained in ATRX-null neonatal brain showed an approximately 50% decrease in H3.3, but this effect was more modest at P17 (Supplementary Figure 11-3B). These results suggest that
H3.3 is not maintained or incorporated as efficiently in the ATRX/H3.3/G-rich region of Dhrsx when ATRX is absent, potentially implicating H3.3 deposition in the control of transcription at specific ATRX target genes. To determine whether the binding of ATRX and H3.3 is specific to Dhrsx or is a common feature of all the aPAR genes regulated by ATRX, we examined the ChIP-seq data at three other genes: Asmtl, Cd99 and Csf2ra. We found that ATRX and H3.3 are also highly enriched in the gene body of these additional aPAR genes and that H3.3 enrichment is decreased in the absence of ATRX (Figure 3-4A). Moreover, these three aPAR genes are enriched for putative G-quadruplexes (Figure 3-4A). Thus, the correlation between ATRX-dependent enrichment of H3.3 within the gene body and gene transcript levels is not specific to Dhrsx, as it is also observed at several other ATRX-regulated genes.
Figure 3-3: ATRX and H3.3 enrichment at Dhrsx.

(A) ChIP-seq raw sequence data from previously published studies of ATRX (GEO accession GSE22162) and H3.3 (GEO accession GSE16893) in mouse embryonic stem cells (ESCs) and novel ChIP-seq for H3.3 in the P0.5 mouse forebrain were aligned to the mouse genome and visualized at Dhrsx. ATRX and H3.3 are highly enriched across 5 kb at the 3’ end of Dhrsx, and H3.3 was lost in the absence of ATRX. Enrichment overlaps with high GC content and putative G quadruplexes (as predicted by Quadfinder (Scaif, 2005)). Input tracks showed only background levels of enrichment (see Supplementary figure 3.12). (B) ChIP-PCR for ATRX in the P0.5 mouse forebrain showed high enrichment for ATRX within the binding region identified by ChIP-sequencing. ChIP was performed with anti-ATRX antibodies H300 and D5. Error bars represent standard error of the mean. (C) ChIP-qPCR for H3.3 at in the P0.5 mouse forebrain showed a significant loss of H3.3 throughout its binding domain. Error bars represent standard error of the mean for n=5. * p<0.05.
To identify additional genes potentially regulated by ATRX in a manner similar to the previously identified aPAR genes, we analyzed genome-wide expression and ChIP data. Downregulated genes were first identified from the previously published wild-type versus ATRX-null P0.5 microarray dataset (Levy et al., 2008). Not including probesets for Atrx, there were 463 probesets with an average downregulation of 1.3 fold or greater, p<0.05, across 3 pairs. These probesets were then curated to filter the probesets to a list of 366 unique genes that could be reliably mapped to specific locations in the mouse genome. Each gene was then analyzed in the ATRX ESC (Law et al., 2010) and H3.3 P0.5 ChIP-Seq data sets to identify sites that were enriched for both ATRX and H3.3, that showed a 1.5 fold or greater loss of H3.3 in the ATRX-null sample, and that had high (70%) GC content overlapping the ATRX/H3.3 binding region. Ultimately, this analysis identified Csf2ra, Dhrsx and two additional genes, Ppp2r3d and 4930526I15Rik. The latter two genes overlap each other but are expressed in opposite directions and are located on an unassembled fragment of chromosome 9 in mouse genome version mm9 (Figure 3-4B). Gene expression changes as shown by microarray analysis (Levy et al., 2008) is not as drastic as with previous aPAR genes; expression at E13.5, P0.5, and P17 for Ppp2r3d is -1.2, -1.4, and -1.3 fold, respectively, and for 4930526I15Rik is -1.1, -1.5, and -1.3 fold, respectively. Analysis of the Ppp2r3d/4930526I15Rik region using Quadfinder (Scaria et al., 2006) identified potential G-quadruplexes, and an examination of the more recent mouse genome version mm10 showed that this region is located approximately 140 kb from the telomere of chromosome 9 and is the most telomeric gene on the chromosome.
Figure 3-4: ATRX and H3.3 enrichment at additional aPAR genes.

(A) ChIP-seq raw sequence data from previously published studies of ATRX (GEO accession GSE22162) and H3.3 (GEO accession GSE16893) in mouse embryonic stem cells (ESCs) and novel ChIP-seq for H3.3 in the P0.5 mouse forebrain were aligned to the mouse genome. Enrichment patterns of ATRX and H3.3 at Asmtl, Cd99, and Csf2ra are similar to that seen at Dhrsx, with high ATRX and H3.3, and loss of H3.3 in ATRX-null samples. Enrichment corresponds to putative G quadruplexes (as predicted by Quadfinder (37)). * Peak artifacts not seen in the paired-end samples due to their higher alignment specificity. (B) Analysis of expression microarray and ChIP-seq data identified Ppp2r3d (and an overlapping non-coding RNA 4930526l15Rik) as an additional aPAR gene enriched for ATRX and H3.3 and that exhibits decreased H3.3 enrichment in the absence of ATRX. See Supplementary Figure 3-12 for ChIP-seq input tracks.
There is no direct human ortholog of the mouse gene *Ppp2r3d*, but it is very similar to, and may have evolved from a common ancestor of the closely related family member *PPP2R3B* (Zwaenepoel et al., 2008). *PPP2R3B* codes for a subunit of protein phosphatase 2A and is responsible for determining localization, activity, and substrate specificity (Stevens et al., 2003; Yan et al., 2000; Zhou et al., 2003). Perhaps unsurprisingly, the gene is location within the human pseudoautosomal region 1 (PAR1). Therefore, based on the criteria of ATRX binding, H3.3 binding and loss, GC richness/presence of G-quadruplexes, proximity to the mouse telomere and presence in the human PAR, *Ppp2r3d/4930526I15Rik* is an additional region where gene expression is likely regulated by ATRX through deposition of H3.3.

### 3.3.5 H3.3 is enriched at telomeres and towards the ends of chromosomes, and depleted from telomeres in the absence of ATRX

Previous studies have reported H3.3 enrichment at telomeres in mouse ESCs and neuroprogenitors and that this enrichment in ESCs is dependent on ATRX (Goldberg et al., 2010; Wong et al., 2010). Analysis of the P0.5 ChIP-seq dataset revealed that telomeric sequences were 5.8 fold enriched over input in the control sample and 2.5 fold enriched in the ATRX-null tissue. This compares to 7.7 and 1.6 fold enriched over input in wild-type and ATRX-null mouse ESCs, respectively (Figure 3-5A) (Goldberg et al., 2010). Therefore, in the absence of ATRX, there was a 2.3 fold decrease in telomeric enrichment of H3.3 in the P0.5 mouse forebrain and a 3.9 fold decrease in mouse ESCs, demonstrating that ATRX assists H3.3 deposition at telomeres in the differentiating forebrain, as well as undifferentiated ESCs as previously described (Goldberg et al., 2010).
Figure 3-5: H3.3 distribution along chromosomes and at telomeres.

(A) The percent of total reads representing telomeres was calculated from ESC (Goldberg et al., 2010) and P0.5 forebrain from H3.3 ChIP-seq, with or without ATRX. ESCs showed a 2.5 fold decrease and P0.5 showed a 1.6 fold decrease in telomeric sequences upon loss of ATRX. (B) H3.3 ChIP-seq sequence-read chromosomal enrichment profile. Each chromosome was divided into 500 “bins” and the number of reads in each bin was calculated then averaged across all mouse chromosomes.
ATRX is also enriched in mouse subtelomeres and towards the ends of chromosomes, with a high concentration of ATRX binding sites in the most telomeric 15-20% of chromosomes (Law et al., 2010). H3.3 enrichment profiles along whole chromosomes were generated using the P0.5 H3.3 ChIP-seq data. In the P0.5 forebrain, H3.3 enrichment was highest towards the telomeres, in the most telomeric 15-20% of the chromosomes (Figure 3-5B). This corresponded to the same pattern of ATRX binding sites across chromosomes seen in mouse ESCs (Law et al., 2010). There was no obvious change in H3.3 enrichment when ATRX was deleted (Figure 3-5B), demonstrating that while ATRX and H3.3 co-localize at multiple sites throughout the mouse genome, a regulatory role for ATRX in H3.3 enrichment is only seen at telomeres and a small number of specific genes.
3.3.6 DNA methylation and histone modifications are not altered at *Dhrsx* in the absence of ATRX

To determine whether the regulation of *Dhrsx* by the chromatin remodeling factor ATRX was solely correlated with a reduction of H3.3 within the gene, we studied additional chromatin features. The PAR region in humans is GC-rich, and *Dhrsx*, despite having translocated to mouse autosomes, maintains this feature. Given the altered patterns of DNA methylation in ATR-X patients (Gibbons et al., 2000), and presence of CpG islands at the subtelomeric target genes in the mouse, we first investigated whether loss of ATRX causes changes in DNA methylation. *Dhrsx* has a 5’ CpG island in the promoter region and a second at the 3’ end encompassing the last three exons. Bisulfite mutagenesis and sequencing of neonatal control forebrain tissue revealed that the *Dhrsx* 5’ CpG island was largely unmethylated (12%) and the 3’ island was almost fully methylated (98%). However, DNA methylation at these sites was similar in the ATRX-null forebrain tissue (Figure 3-6). We then extended this analysis to *Csf2ra*, which has a small CpG island at the promoter (Supplementary Figure 3-13). We found that the *Csf2ra* 5’ island was approximately 65% methylated in two littermate-matched ATRX-null and control mice (Supplementary Figure 3-13). Thus, ATRX does not appear to influence gene expression at these sites by altering DNA methylation states.
Figure 3-6: DNA methylation at Dhrsx CpG islands is not changed in the P0.5 ATRX-null mouse forebrain.

Bisulfite mutagenesis and sequencing at Dhrsx showed low DNA methylation (12%) within the CpG island overlapping a putative promoter region (CpG1), and high DNA methylation (98%) within the 3’ CpG island (CpG2). No change was seen between P0.5 control and Atrx-null forebrains (n=2). Empty circles represent unmethylated cytosines, while filled circles represent methylated cytosines.
Next, we analyzed various histone post-translational modifications within and outside the Dhrsx gene (Figure 3-7A). Promoter regions including the transcription start site of active genes show high levels of acetylation of histone H4 (H4Ac) and of trimethylation on lysine 4 of histone H3 (H3K4me3). These active sites are also marked by low levels of histone H3 tri-methylation one lysine 9 (H3K9me3) and lysine 27 (H3K27me3) (Barski et al., 2007; Li et al., 2007; Wang et al., 2008). Enrichment of these four histone modifications was examined across the 20 kb Dhrsx gene, revealing the expected profile for an active gene. High peaks of H4Ac and H3K4Me3 occur in the promoter region that overlaps with the 5’ CpG island of Dhrsx. The repressive marks H3K9me3 and H3K27me3 both showed low levels of enrichment specifically at this site (Figure 3-7B-E). In the absence of ATRX no significant changes were observed in the enrichment of these histone modifications (Figures 3-7B-E). Thus, the effects of ATRX deficiency on gene expression cannot be explained by changes in CpG island DNA methylation or in key histone post-translational modifications.
Figure 3-7: Histone modifications are unchanged in the P17 ATRX-null mouse forebrain.

(A) Location of PCR amplicons (numbered) used for ChIP analysis at Dhrsx. Primers 1 and 18 are located 15 kb upstream and downstream, respectively, from the indicated locations. ChIP for the active histone modifications H4Ac (B) and H3K4me3 (C) showed enrichment at the Dhrsx putative promoter region and transcription start site (TSS) (primer sites 2 and 3) but no change between control and ATRX-null mice. The repressive modification H3K9me3 (D) and H3K27me3 (E) showed an opposite pattern of enrichment with low levels at the promoter/TSS region, but also had no significant changes between in the absence of ATRX. To control for variability between experiments, in all cases raw percent input was normalized to enrichment at control sites for the Myod1 promoter and Gapdh promoter. Error bars represent standard error of the mean for n=3 (n=2 for H3K9Me3).
3.3.7 Increased RNA polymerase II occupancy at the ATRX/H3.3/G-rich region of Dhrsx in the absence of ATRX

So far, we have determined that in the absence of ATRX, H3.3 deposition is decreased in the G-rich segment of the Dhrsx gene body, whereas several other epigenetic marks within the gene body and at the promoter remain unchanged. We therefore hypothesized that the decrease in Dhrsx transcripts consistently observed when ATRX is absent could be due to a problem in transcriptional elongation rather than initiation. To study this possibility, we performed ChIP-qPCR for RNA polymerase II (PolII) across Dhrsx in the presence and absence of ATRX in the mouse forebrain. We found moderate levels of PolII at the Dhrsx promoter/transcriptional start site region and lower levels across the gene (Figure 3-8). In the absence of ATRX, PolII occupancy was significantly increased within the ATRX/H3.3-binding/G-rich region of Dhrsx (Figure 3-8), suggesting that the enzyme is more prone to stalling. This is predicted to either decrease the rate of elongation or to stop elongation altogether, which could provide an explanation for decreased levels of gene transcripts. Taken together, these data suggest that ATRX may be required for proper transcription through certain G-rich gene regions.
Figure 3-8: RNA Pol-II stalling in the absence of ATRX.

(A) ChIP-qPCR for RNA polymerase II (PolII) at Dhrsx in control and ATRX-null P0.5 mouse forebrain. PolII is enriched at the promoter and TSS region followed by lower levels throughout the gene body. In the absence of ATRX, PolII enrichment is increased at sites 9 to 13, corresponding to the ATRX/H3.3 binding region, with three sites showing statistically significant change. Error bars represent standard error of the mean for n=5. * p<0.05.
3.4 Discussion

This work explored possible epigenetic mechanisms that could explain how ATRX regulates gene expression. We focused on a gene, Dhrsx, which we previously identified as one of the most downregulated genes in the ATRX-deficient brain (Levy et al., 2008). Our findings demonstrate that transcriptional regulation of Dhrsx does not occur through interactions of ATRX with the gene promoter, but rather with the gene body in a G-rich region containing many predicted G-quadruplexes. We show that loss of ATRX results in decreased levels of histone H3.3 within this region of the gene and at four other aPAR genes, showing that ATRX is able to regulate H3.3 deposition at a subset of genes in the genome. In fact we see decreased H3.3 deposition at all of the aPAR genes previously identified as downregulated in the ATRX-null brain (Levy et al., 2008). In addition, we show increased occupancy of RNA PolII within this G-rich region of Dhrsx in the absence of ATRX, suggesting that loss of ATRX and H3.3 at this G-rich transcribed region results in the stalling of RNA polymerase.

While a small fraction of the mammalian genome codes for proteins, transcription of non-coding RNAs is likely a widespread phenomenon (Clark et al., 2011; Kapranov and St Laurent, 2012). We detected transcription from regions flanking Dhrsx, including from within an ITS. This transcription was decreased in the absence of ATRX, similar to the decrease seen in the Dhrsx gene itself. Transcription of DNA flanking the Dhrsx gene may indicate that the annotated start and end sites are incorrect. However, the levels of transcription and degree of repression are inconsistent across the region, making it unlikely that a single long transcript spans the entire domain. Rather, non-coding transcripts are likely expressed in addition to, and overlapping with, the protein-coding Dhrsx transcript. The direction of transcription for these transcripts in comparison to Dhrsx transcription is unknown. Given the location of ATRX and H3.3 binding at the 3’ end of Dhrsx, we speculate that the non-coding transcripts, regardless of start sites and direction of transcription, pass through the ATRX/H3.3/G-rich region, and are thus co-regulated with Dhrsx. A more detailed analysis will be required to determine the number, length and direction of transcripts and to examine the effects of ATRX at these sites and provide a complete picture of the transcriptional architecture within this domain.
Loss of ATRX in mouse ESCs has been associated with increased TERRA expression (Goldberg et al., 2010) while a series of human ALT cell lines showed no correlation between loss of ATRX and levels of TERRA (Lovejoy et al., 2012). We found no change in TERRA expression in either ATRX-null MEFs or mouse forebrain tissue. We do note however, that expression of Dhrsx and Csf2ra was decreased in ATRX-null MEFs, similar to the effect observed in the mouse forebrain. We also see strong downregulation of these genes in the ATRX-null embryonic limb bud (our unpublished data). Together, our data demonstrates that ATRX regulates the expression of these genes in many cell types, and that this regulation is independent of TERRA transcripts.

ATRX is a chromatin remodeling protein that exhibits DNA translocase and nucleosome remodeling activity (Xue et al., 2003). Loss of ATRX also correlates with altered patterns of DNA methylation at several repetitive regions in ATR-X syndrome patients (Gibbons et al., 2000) and in mice (Garrick et al., 2006; Kernohan et al., 2010). Conversely, loss of ATRX did not affect DNA methylation at imprinted domains (Kernohan et al., 2010). In this present study we find no differences in DNA methylation at Dhrsx and Csf2ra CpG islands, and no change in active or repressive histone modifications at Dhrsx. The epigenetic profile of the Dhrsx promoter region indicates an active gene, and this profile remains unchanged in the absence of ATRX. These data suggest that the mechanism of regulation might not involve transcriptional initiation, a supposition reinforced by the position of ATRX and H3.3 further along the gene.

One third of ATRX binding sites in ESCs are within genes bodies, yet intragenic H3.3 binding was shown to be largely unchanged upon loss of ATRX (Goldberg et al., 2010; Law et al., 2010). Instead, histone regulator A (HIRA) is thought to be the histone chaperone responsible for depositing H3.3 within gene bodies (Goldberg et al., 2010). In contrast, our findings clearly show that ATRX and H3.3 are co-localized within the gene bodies of all four aPAR genes and that the level of H3.3 at these sites depends on ATRX. H3.3 is traditionally associated with active transcription (Ahmad and Henikoff, 2002), in part because its rapid turnover rate maintains more easily accessible DNA which allows binding of transcription factors and passage of transcription machinery (Henikoff, 2008; Schwartz and Ahmad, 2005). Without ATRX to deposit H3.3, the more stable H3.1 or
H3.2 may be deposited instead, as is seen in *Drosophila* where in the absence of H3.3, replication independent incorporation of H3 is observed (Sakai et al., 2009). Incorporation of H3.3 by ATRX may assist passage of the transcription machinery by helping to resolve inhibitory DNA structures such as G-quadruplexes, as previously hypothesized (Law et al., 2010), since putative G-quadruplex-forming sequences were identified within the ATRX/H3.3/G-rich binding domains of the aPAR genes.

Supporting the idea that ATRX assists transcriptional elongation via H3.3 deposition, RNA polymerase II showed increased enrichment at the ATRX/H3.3/G-rich region of *Dhrsx* in the absence of ATRX indicating pausing of the transcription machinery. The P17 ATRX-null forebrain exhibited a more subtle decrease in H3.3 enrichment compared to neonatal brain tissue. Post-mitotic neurons are enriched for H3.3 and have reduced H3.1 and H3.2 (Bosch and Suau, 1995; Pina and Suau, 1987) therefore the ATRX-null cells may have little choice but to incorporate H3.3, perhaps at a slower rate using an ATRX-independent mechanism.

From a genome-wide perspective, the binding pattern of H3.3 in the mouse brain reflected that previously reported for H3.3 and ATRX in mouse ESCs (Goldberg et al., 2010; Law et al., 2010), including high enrichment at telomeres and loss of H3.3 in ATRX-null cells. The higher enrichment of ATRX and H3.3 towards telomeres suggests that they may be more involved in gene regulation in this region. The fact that all aPAR genes regulated by ATRX and that can be definitively positioned in the mouse genome are located within this region supports this assessment.

Based on the known functions of the four studied aPAR genes, it is not immediately clear how decreased expression of the aPAR genes may contribute to the ATRX-null phenotypes in mice and humans. However, the regulation of *Csf2ra* might shed some light into the placental defects seen in the ATRX-null concepti (Garrick et al., 2006; Sferruzzi-Perri et al., 2009). *Csf2ra* codes for a receptor subunit of the cytokine granulocyte-macrophage colony-stimulating factor. Mutations in this factor can lead to adverse fetal and placental health and viability (Robertson, 2007). In addition, *Csf2ra*
was identified as a schizophrenia susceptibility gene, which suggests a potential role in the functioning of the central nervous system (Lencz et al., 2007; Loe-Mie et al., 2010).

In conclusion, our study provides mechanistic insight of ATRX-mediated transcriptional regulation. We propose a model whereby incorporation of H3.3, as directed by ATRX, promotes transcriptional elongation of a G-rich region by organizing chromatin into a more accessible state (Figure 3-9). Future studies should seek to identify other components of the regulatory system by which loss of H3.3 affects expression of DhrsX and the other target genes. It will be interesting to see whether DAXX is present with ATRX, as DAXX has been shown to be required for the ATRX-mediated deposition of H3.3 at telomeres (Lewis et al., 2010) and pericentromeric heterochromatin (Drane et al., 2010). Further analyses of genome wide protein binding and expression data may identify additional genes besides the four aPAR genes which are regulated in the manner described here.
Figure 3-9: A model for the regulation of transcription by ATRX.

Transcription proceeds from the transcription start site (TSS). In the presence of ATRX (top, purple), ATRX resolves G-quadruplex structures in G-rich regions allowing passage of RNA polymerase (PolII, green triangles). After passage of the transcription machinery ATRX directs the incorporation of H3.3-containing histones (blue), which facilitates subsequent transcription. In the absence of ATR (bottom), G-quadruplexes that form in the single-stranded region of the transcription bubble impede the passage of the transcription machinery leading to RNA polymerase stalling and decreased transcription. In the absence of ATRX, H3.3 deposition is decreased, further inhibiting transcription.
3.5 Supplementary figures

Figure 3-10 (supplementary): Adenoviral treatment of Atrx-floxed MEFs.

(A) MEFs were infected with adenovirus expressing Cre recombinase with GFP, or just GFP, and ATRX expression was assessed by immunofluorescence 72 hours post infection. Scale bar = 500 µm. (B) Atrx, Csf2ra, and Dhrsx expression was measured by qRT-PCR 72 hours after Adeno-Cre-mediated Atrx deletion. Error bars represent standard error of the mean for n=3. * p<0.05.
Figure 3-11 (supplementary): Comparison of H3.3 enrichment at Dhrsx in the mouse brain at P0.5 and P17.

(A) ChIP-qPCR of histone H3.3 at Dhrsx in the P0.5 and P17 forebrain confirmed the P0.5 ChIP-seq data and showed a similar pattern of enrichment at P17, with highest levels at the ATRX/H3.3/G region. Enrichment was normalized to H3.3 binding at B-actin. (B) ChIP-qPCR of histone H3.3 at P17 showed decrease H3.3 binding at amplicons 8 and 11 of the ATRX/H3.3/G region of Dhrsx. Error bars represent standard error of the mean for n=4. * p<0.05.
Figure 3-12 (supplementary): ATRX and H3.3 ChIP-Seq input tracks.

No antibody control (input) tracks for *Dhrsx* (A) and the additional aPAR genes *Asmtl*, *Cd99*, and *Csf2ra* (B) and *Ppp2r3d* (C). For ChIP-seq in the P0.5 forebrain, control and ATRX-null ChIP DNA was pooled prior to sequencing.
Figure 3-13 (supplementary): DNA methylation at the Csf2ra promoter is not altered in the neonatal ATRX-null mouse forebrain.

Bisulfite mutagenesis and sequencing showed 65% DNA methylation within the Csf2ra putative promoter region. No change was seen between control and ATRX-null forebrains. Open circles represent unmethylated cytosines, closed circles represent methylated cytosines.
## 3.6 Supplementary tables

### Table 3-1 (supplementary): Primer sequences and annealing temperatures.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Forward</th>
<th>Reverse</th>
<th>Anneal. temp.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Genotyping</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atrx</td>
<td>AGAAATTGAGGATGCTTCACC</td>
<td>TGAACCTGGGGACTCTTTTGTG</td>
<td>58</td>
</tr>
<tr>
<td>NeoR</td>
<td>Atrx forward used as forward primer</td>
<td>CCACCATGATATTCGGCAAG</td>
<td>58</td>
</tr>
<tr>
<td>Cre</td>
<td>TGACCGAGATCATCCTTACGCGG</td>
<td>AATGCTTCTGGCGTTTGGCC</td>
<td>58</td>
</tr>
<tr>
<td>Sry</td>
<td>GCAGGTGGGAAGGCCCTTACA</td>
<td>AAGCTTTCTGTGGTTTGGGA</td>
<td>58</td>
</tr>
<tr>
<td><strong>Gene expression</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gapdh</td>
<td>CAACGACCCCTCTCCTGGACT</td>
<td>ATCCAGCAGCAGACATGGA</td>
<td>55</td>
</tr>
<tr>
<td>Bactin</td>
<td>CTGCGGATGCGGGCTACACC</td>
<td>ACATGCGAGCCGGTCTTGTG</td>
<td>60</td>
</tr>
<tr>
<td>Atrx</td>
<td>AGAAATTGAGGATGCTTCACC</td>
<td>TGAACCTGGGGACTCTTTTGTG</td>
<td>58</td>
</tr>
<tr>
<td>Cre</td>
<td>TGACCGAGATCATCCTTACGCGG</td>
<td>AATGCTTCTGGCGTTTGGCC</td>
<td>58</td>
</tr>
<tr>
<td>Sry</td>
<td>GCAGGTGGGAAGGCCCTTACA</td>
<td>AAGCTTTCTGTGGTTTGGGA</td>
<td>58</td>
</tr>
<tr>
<td><strong>Dhrsx</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genotyping</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atrx</td>
<td>AGAAATTGAGGATGCTTCACC</td>
<td>TGAACCTGGGGACTCTTTTGTG</td>
<td>58</td>
</tr>
<tr>
<td>Csf2ra</td>
<td>ATCAGTACTCGTGCGTTTGTCTTG</td>
<td>TTAGCTCAGGTTTTTGTCTTG</td>
<td>58</td>
</tr>
<tr>
<td>Dhrsx</td>
<td>GCAGGTGGGAAGGCCCTTACA</td>
<td>AAGCTTTCTGTGGTTTGGGA</td>
<td>58</td>
</tr>
<tr>
<td><strong>Bisulfite sequencing</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dhrsx CpG1 outside</td>
<td>GAAATGTATATAAATATTTTTTATTTT</td>
<td>TATTCTAACCAACATCTTAATTT</td>
<td>50</td>
</tr>
<tr>
<td>Dhrsx CpG1 inside</td>
<td>TTTTGTGTTTGAGAATGTTTGGTT</td>
<td>AAAACCTCCCAAACACCTACCA</td>
<td>55</td>
</tr>
<tr>
<td>Dhrsx CpG2 outside</td>
<td>TTGTTTTTTTTGTATTGTTT</td>
<td>TTTTCTGTGGCTCTTGGTTTATCT</td>
<td>60</td>
</tr>
<tr>
<td>Dhrsx CpG2 inside</td>
<td>TTGTTTTTTTTGTATTGTTT</td>
<td>TTTTCTGTGGCTCTTGGTTTATCT</td>
<td>60</td>
</tr>
<tr>
<td>Csf2ra outside</td>
<td>ATGTTTTTTTTGTATTGTTT</td>
<td>TTTTCTGTGGCTCTTGGTTTATCT</td>
<td>60</td>
</tr>
<tr>
<td>Csf2ra inside</td>
<td>ATGTTTTTTTTGTATTGTTT</td>
<td>TTTTCTGTGGCTCTTGGTTTATCT</td>
<td>60</td>
</tr>
<tr>
<td>M13</td>
<td>GTAAAAGACGACGCCCAG</td>
<td>CAGGAAACAGCTTACAG</td>
<td>55</td>
</tr>
</tbody>
</table>
3.7 References


telomeres phenotype are late events in a small subset of MEN-1 syndrome pancreatic neuroendocrine tumors. Mod Pathol 25, 1033-1039.


Chapter 4

4 General Discussion and Future Directions

4.1 Thesis summary

The overall goal of my thesis was to identify gene targets of ATRX and determine how they are regulated by ATRX. To this end, I mainly used a mouse model of ATRX loss to identify these genes and study how ATRX interacts with them. I focused on epigenetic changes associated with ATRX loss given that ATRX is a chromatin remodeling protein.

The first part of my thesis, presented in chapter two, describes the identification of a unique group of genes regulated by ATRX, the aPAR genes. A connection between these genes was not immediately recognized, and indeed we initially took note of genes involved in brain development and neurogenesis before changing direction to address a novel question: how could a group of genes, seemingly only connected by their genomic location in the evolutionary past, be regulated by the same protein? These genes, which are GC rich and originate from a highly repetitive region (the PAR) were identified before it was known that ATRX is in fact highly enriched at these types of regions throughout the genome (Law et al., 2010). In the process of identifying aPAR genes regulated by ATRX we found two new mouse aPAR genes, Arsd/e and Asmtl. We hypothesized that the target genes could be regulated by ATRX due to an intrinsic sequence within, or epigenetic environment surrounding these genes, regardless of their genomic location. However, we found that when these genes are in the PAR region itself they are not regulated by ATRX, suggesting that their genomic location is also important for their targeting and regulation by ATRX.

The second part of my thesis, presented in chapter three, attempted to find a mechanism by which ATRX regulates gene expression. To study this, I focused on a role for ATRX at the previously identified aPAR genes. We first found that ATRX binds directly at these genes and that the binding pattern corresponds to high GC regions and the presence of potential secondary DNA structures known as G-quadruplexes. This suggested a direct mechanism of gene regulation by ATRX. Several epigenetic factors were
examined and found to be unchanged: DNA methylation, histone modifications, and TERRA expression. We then investigated the histone variant H3.3 after several reports described a link between ATRX and H3.3 at telomeres (Goldberg et al., 2010; Wong et al., 2010). I made the novel discovery that H3.3 binds in a pattern mirroring ATRX within the bodies of the aPAR genes, and that H3.3 is consistently depleted at all aPAR genes misregulated in the absence of ATRX. I further demonstrated that there is an increase in binding of RNA PolII within the ATRX/H3.3/G-rich binding region, suggesting stalled transcription machinery. Therefore, I believe that we have demonstrated one potential mechanism by which ATRX can regulate gene expression: facilitating transcription though regions prone to forming inhibitory secondary DNA structures.

4.2 A role for ATRX in gene regulation

When I started this work, ATRX had a suspected but not confirmed role in regulating gene expression. This was largely based on two observations. First, conserved domains in ATRX had been implicated in gene regulation in other proteins, such as the PHD DNA binding domain (Aasland et al., 1995) and SWI2/SNF2 chromatin remodeling domain (Pazin and Kadonaga, 1997). Second, ATRX binding partners HP1, EZH2 and DAXX were themselves known to be involved in gene regulation (Cao et al., 2002; Cao and Zhang, 2004; Kourmouli et al., 2005; Salomoni and Khelifi, 2006; Stewart et al., 2005). The repressive nature of HP1 (Grewal and Jia, 2007; Stewart et al., 2005) and EZH2 (Cao and Zhang, 2004), the frequent association of ATRX with condensed PCH (McDowell et al., 1999), the ability of ATRX to assemble heterochromatin and suppress nearby gene expression in Drosophila (Bassett et al., 2008; Emelyanov et al., 2010), and the finding that ATRX tethered to a promoter can suppress transcription (Tang et al., 2004), suggested a predominantly repressive role for ATRX. However, the decreased expression of α-globin in ATR-X syndrome patients (Gibbons et al., 1995) hinted at an additional activating role, although whether it was direct or indirect regulation was unknown. As my work progressed, additional small-scale studies found roles for ATRX regulation at specific genes (Bagheri-Fam et al., 2011; Kernohan et al., 2010; Newhart et al., 2012; Schreiner et al., 2013; Tang et al., 2011; Tsai et al., 2011), but my work was the
first to look at genome-wide changes in gene expression upon loss of ATRX (Levy et al., 2008). When ATRX is lost in the developing mouse forebrain, two-thirds of misregulated genes had increased transcript levels. This observation confirmed a mostly suppressive role for ATRX in gene expression. However, with approximately 100 genes downregulated in the absence of ATRX there is a clear activating role as well. Given that the presumed mechanism for gene activation seen at aPAR genes accounts for a small number of these targets, there is at least one and possibly several other mechanisms by which ATRX can regulate gene expression. While many of the genes identified in our microarray screen may be secondary effects of ATRX loss, I believe it is reasonable to assume that there are at least some additional genes besides the aPAR genes that are directly regulated by ATRX. Indeed, a mechanism of gene regulation similar to the one proposed here has been identified for the α-globin gene cluster in humans (Law et al., 2010). The α-globin gene cluster is near the telomere of chromosome 16 and contains several globin genes, regulatory elements, and other genes unrelated to globin production (Higgs et al., 2005). It was found that ATRX binds directly to intergenic G-rich tandem repeats predicted to form G-quadruplexes, and that alleles with more repeats had a stronger downregulation of α-globin (Law et al., 2010). In light of the known roles for ATRX in replication (Huh et al., 2012; Wong et al., 2010), it was proposed that the presence of G-quadruplexes hinders DNA replication through the α-globin region (Figure 1-3A). This hindrance could lead to DNA damage, replication fork collapse, and loss of activating epigenetic marks which could all contribute to the decreased expression of a number of genes surrounding the α-globin locus (Clynes and Gibbons, 2013; Gibbons and Higgs, 2010; Law et al., 2010). It is important to note that the α-globin G-quadruplexes are intergenic, necessitating an indirect method of regulation (i.e. secondary to a replication defect), while the aPAR G-quadruplexes are within gene bodies, potentiating a direct inhibition of transcription.

The impact of ATRX loss on the α-globin locus has been examined in a second, more recent study (Ratnakumar et al., 2012). It was found that a unique ATRX complex regulates the repressive histone variant macroH2A (mH2A), and that this is distinct from the previously identified ATRX-DAXX-H3.3 complex. In the absence of ATRX, increased enrichment of mH2A was found at telomeres and across the α-globin gene
cluster (Ratnakumar et al., 2012). This is in contrast to H3.3 enrichment which is decreased at telomeres in the absence of ATRX (Goldberg et al., 2010). It would therefore be interesting to measure mH2A levels at the aPAR genes to see if they show an inverse relation with H3.3 enrichment.

We found that the aPAR genes misregulated in the mouse are not affected in human cells upon loss of ATRX. Why could this be? Like the individual mouse aPAR genes, the overall human PAR is rich in potential G-quadruplex forming sequences. The aPAR genes in mice have diverged considerably from their counterparts in the human PAR, and along with the coding sequences, the overall genomic structure between the human PAR and mouse aPAR genes are quite different. For example, the mouse Dhrsx gene spans approximately 20 kb of genomic DNA, while the human DHRSX is over 200 kb long. Nevertheless, some of the human PAR genes also contain potential G-quadruplex forming regions, suggesting that these genes in the human PAR require the ability to transcribe through G-quadruplexes similar to the autosomal aPAR genes in mice. However, preliminary analysis of a previously published ChIP-seq dataset from human erythroid cells (Law et al., 2010) showed that while ATRX binding sites are present within the overall human PAR, none of the sites overlap the G-quadruplex regions within the studied PAR genes themselves, and indeed there are no ATRX binding sites that directly overlap the studied PAR genes. It is therefore possible that these human PAR genes utilize a different mechanism to bypass their G-quadruplexes, or that G-quadruplexes fail to form at the predicted sites. It is interesting to note that while the genes in the human PAR are near the end of the X and Y chromosomes, they are between 10 and 30 times farther from their respective telomeres than the aPAR genes on mouse autosomes (one to three megabases for human PAR genes, compared to 100 to 150 kb for mouse aPAR genes). A role for ATRX in H3.3 deposition and G-quadruplex bypass may therefore be restricted to the telomeres and to regions within a certain distance of the telomeres. This restriction may be mediated by PML-NBs. It has been shown that the ATRX-DAXX complex co-localizes with telomeres within PML-NBs, and that this is where the H3.3 deposition occurs (Chang et al., 2013; Delbarre et al., 2012). It is possible that this co-localization may bring the mouse aPAR genes within the PML-NBs
along with the telomeres themselves, allowing ATRX-DAXX to deposit H3.3, and that the human orthologs may be too far from the telomeres for this to occur.

Despite the lack of changes in human PAR gene expression, we can still speculate on a role for ATRX in gene regulation in humans based on our findings and those of others. Only 5% of ATRX binding sites in mice overlap CpG islands (Law et al., 2010). The aPAR genes are therefore members of the small category of mouse ATRX targets that contain CpG islands, further suggesting that they represent a unique subset of genes regulated by ATRX. In humans, nearly half overlap CpG islands compared to the 5% in mice (Law et al., 2010). In addition, gene density and CpG content near telomeres is higher in humans than mice (Law et al., 2010; Murmann et al., 2005). It is therefore possible that ATRX regulates more genes in humans than in mice through the mechanism proposed in the present study. When ATRX binds a promoter, it represses transcription (Newhart et al., 2012; Valadez-Graham et al., 2012). While approximately one-third of ATRX binding sites are in gene bodies in both humans and mice, ATRX bound 78 promoters (6% of all sites) in mouse ESCs and 326 promoters (36% of all sites) in human erythrocytes (Law et al., 2010), possibly due to the generally higher CpG density of human CpG islands (Illingworth et al., 2010) at promoters. Therefore, loss of ATRX may cause more genes to be upregulated in humans than in mice.

4.3 ATRX and structured DNA

Clearly, ATRX plays a role in managing secondary DNA structures like G-quadruplexes, but what exactly could it be doing at these sites? G-quadruplexes form on single stranded DNA, such as during replication and transcription. It has been shown that DNA damage increases in the absence of ATRX, and that this damage may be the result of the inability to replicate telomeric G-quadruplexes (Huh et al., 2012; Watson et al., 2013; Wong et al., 2010). Transcription through G-quadruplexes can also signal additional DNA damage. Just as the replication fork can stall at these structures, so too can RNA PolII and the transcription machinery. Without ATRX to assist in the transcriptional bypass of these structures, paused or stopped RNA PolII may be detected by transcription-coupled DNA repair factors, even though no actual damage has occurred (Tornaletti, 2009). Superfluous DNA repair through the nucleotide excision repair pathway can increase the
chances of mutations leading to DNA instability and possibly cancer (Bochman et al., 2012; Tornaletti, 2009). Mutations in the Fanconia anemia gene known as BRCA1 interacting protein (BRIP1) is associated with breast cancer and the bone marrow disease Fanconi anemia. BRIP1 is a helicase that can unwind G-quadruplexes, and when depleted, sensitizes cells to DNA damage caused by stabilization of G-quadruplexes (Wu et al., 2008). In addition, mutations in ATRX have recently been found in pancreatic and several types of brain cancers (Heaphy et al., 2011; Jiao et al., 2011; Molenaar et al., 2012; Schwartzentruber et al., 2012). Looking for markers of transcription-coupled DNA repair, such as the excision repair protein ERCC6 and ubiquitination of RNA PolII (Fousteri and Mullenders, 2008), associated with the increased RNA PolII at Dhrsxx would clarify this question.

It is unclear exactly how ATRX assists in the transcriptional bypass of G-quadruplexes, but several other proteins that act at these structures have previously been identified. For example, the G4 resolvase DEAH box polypeptide 36 (DHX36), and the RecQ helicases involved in Werner syndrome (WRN) (Fry and Loeb, 1999) and Bloom’s syndrome (BLM) (Sun et al., 1998) can unwind G-quadruplexes; protection of telomeres 1 (POT1) (Zaug et al., 2005) and replication protein A (RPA) (Salas et al., 2006) can unfold telomere G-quadruplexes; and G quartet nuclease 1 (GQN1) can cut within them (Sun et al., 2001). While ATRX has a helicase domain, it has no helicase activity (Xue et al., 2003). ATRX may therefore act through an as yet undiscovered mechanism, or through an additional protein partner, to resolve G-quadruplexes to assist transcriptional elongation. In addition, the mere act of depositing H3.3 may help to suppress G-quadruplex formation as G-quadruplexes are enriched in nucleosome-free regions (Halder et al., 2009; Hershman et al., 2008; Wong and Huppert, 2009).

Transcribing RNA PolII generates positive supercoiling in front of, and negative supercoiling behind the passing transcription machinery (Liu and Wang, 1987). Negative supercoiling facilitates G-quadruplex formation (Sun and Hurley, 2009) and transcription blockage in potential G-quadruplex forming regions is enhanced by negative supercoiling (Belotserkovskii et al., 2010). Nucleosome assembly on negatively supercoiled DNA can relieve superhelical stress (Clark and Felsenfeld, 1991; Worcel et al., 1981). Therefore,
ATRX may suppress G-quadruplexes in DNA undergoing transcription by assisting in nucleosome deposition after passage of the transcription machinery, creating a helical state less conducive to G-quadruplex formation.

G-rich regions have the ability to form additional non-B structures besides G-quadruplexes. While DNA G-quadruplexes can form when the G-rich strand is either the coding or template strand, RNA G-quadruplexes and R-loops can form only when the G-rich strand is the coding strand (and is therefore also reflected in the RNA itself). R-loops can form during transcription when the single stranded coding strand binds with an unusually stable RNA:DNA hybrid, due to the G:C-rich nature of the sequences (Aguilera and Garcia-Muse, 2012). R-loops have been shown to inhibit transcription within genes and at telomeres (Belotserkovskii et al., 2010). Since Dhrsx and the other aPAR genes have G-rich stretches on both their coding and template strands, it is unclear exactly which secondary structure may be forming. Given that ATRX is known to target DNA G-quadruplexes (Law et al., 2010), it is most likely that it is these structures that form and interfere with transcription. However, possible interactions between ATRX and other non-B DNA have not yet been studied.

4.4 Histone chaperones and the genomic localization of histone H3.3

Involvement of H3.3 in regulation of transcription is mediated by different chaperones in different regions of the genome. H3.3 is recruited to PCH by ATRX and DAXX where it is required to maintain expression from the repetitive PCH DNA (Drane et al., 2010). It is also recruited to telomeres by ATRX and DAXX but here loss of ATRX (and therefore H3.3) causes an increase instead of decrease in transcription from the repetitive telomeric DNA (Goldberg et al., 2010; Lewis et al., 2010; Wong et al., 2010). DAXX and HIRA are responsible for localizing H3.3 to at least a small number of regulatory elements (Goldberg et al., 2010; Michod et al., 2012). Lastly, HIRA targets H3.3 to intergenic regions and the majority of gene bodies, but loss of ATRX was found to cause minimal change at these sites (Goldberg et al., 2010; Law et al., 2010). Nevertheless, we have now identified a small class of genes that require ATRX for H3.3 localization within gene bodies. Given that the aPAR genes do not fit neatly into one of the previously
described categories of H3.3 deposition regions, it would be interesting to study other chaperones involved in H3.3 regulation at these genes. Their location near telomeres may mean that H3.3 at these genes is solely deposited by the ATRX-DAXX complex. However, since the deposition is occurring in gene bodies, HIRA may be able to partly compensate for loss of ATRX-mediated deposition. Examining DAXX and HIRA at the aPAR genes would rule in or out these two options.

PML-NBs may be involved in deciding which chaperones will associate with H3.3 and therefore where in the genome it will be deposited. In ESCs, the ATRX-DAXXX complex localizes to PML-NBs and directs the deposition of H3.3 to telomeres (Chang et al., 2013; Delbarre et al., 2012), while a separate complex containing HIRA also localizes to PML-NBs where it presumably performs its role in depositing H3.3 at intergenic regions and gene bodies (Delbarre et al., 2012). In somatic and fully differentiated cells, ATRX-DAXX still localizes to PML-NBs (Drane et al., 2010; Ishov et al., 2004; Xue et al., 2003) but telomeres do not (Chang et al., 2013), and ATRX is no longer found at telomeres (Wong et al., 2010). The proximity of the specific aPAR genes studied in this work to telomeres may mean they are within PML-NBs and therefore within the vicinity of ATRX. The subtelomeric location would also explain why they are the only aPAR genes affected by loss of ATRX. At least four other aPAR genes have been identified in mice (Crlf2, Plcd1, Gtphbp6, Il3ra) that are not located near telomeres. Analysis of our microarray data revealed that expression of these genes is not altered upon loss of Atrx.

The newborn mouse forebrain contains a mix of differentiated and undifferentiated cells. We showed that in the P0.5 mouse forebrain, H3.3 is decreased at telomeres and aPAR genes in the absence of ATRX, but this decrease is not as severe as seen in ESCs (Goldberg et al., 2010). In addition, the loss of H3.3 in the ATRX-null P17 forebrain was less than that seen at P0.5. Therefore, ATRX may play a continually less important role in H3.3 deposition as cells differentiate. Nevertheless, the small loss of H3.3 at P17 appears to be enough to inhibit Dhrsx transcription. Given that H3.3 is incorporated in a replication independent manner, it is the predominant H3 variant in the post-mitotic adult brain (Bosch and Suau, 1995; Pina and Suau, 1987). While there is an overall increase in H3.3 levels in this tissue, the specific genomic distribution of H3.3 has not been studied.
beyond the neuroprogenitor stage (Goldberg et al., 2010). By ChIP-seq, we found that H3.3 was enriched at telomeres in the newborn forebrain, but that this enrichment was not as high as seen in ESCs (Goldberg et al., 2010). Given the requirement for ATRX in telomeric deposition of H3.3, and the loss of ATRX from telomeres in differentiated cells (Xue et al., 2003), it is possible that while H3.3 may be enriched throughout most of the genome (through HIRA-mediated deposition), it may be depleted from post-mitotic telomeres. This supports the idea that ATRX is required at telomeres specifically in replicating cells. Indeed, H3.3 is incorporated into telomeres during S phase (Wong et al., 2010; Wong et al., 2009), and loss of ATRX results in telomere dysfunction in ESCs (Wong et al., 2010), replicating muscle cells (Huh et al., 2012), and neuroprogenitors (Watson et al., 2013).

The dentate gyrus is one of only two regions where adult neurogenesis occurs (Altman and Das, 1965; Ming and Song, 2011). When Atrx is deleted from the mouse forebrain, the most severe effects are seen in this region (Bérubé et al., 2005), and this is in part caused by defective migration of dentate gyrus precursor cells (Seah et al., 2008). It is possible that an additional reason the dentate gyrus is so severely affected upon deletion of ATRX is because the ongoing neurogenesis and cell division in this structure makes it particularly sensitive to ATRX loss. Future studies to investigate the association of ATRX and H3.3 with telomeres in the dentate gyrus versus other regions of the forebrain, in the adult brain and throughout hippocampal development, would help clarify the role for ATRX in neurogenesis and brain development.

### 4.5 H3.3 the and regulation of aPAR genes

H3.3 is a multi-faceted histone variant. Firstly, it was originally associated with active chromatin, but more recently found to also play a role at repressive regions (Ahmad and Henikoff, 2002; Elsaesser et al., 2010; Szenker et al., 2011). Secondly, it may play dual roles in enhancing gene expression: H3.3 is deposited in transcribed regions to replace nucleosomes that are evicted upon passage of the transcriptional machinery (Schwartz and Ahmad, 2005, 2006), but it can also facilitate future gene expression by maintaining a chromatin state more permissive to transcription (compared to canonical H3) that persists through cell division (Henikoff, 2008; Ng and Gurdon, 2008). Therefore, it is
possible that the change in H3.3 enrichment at aPAR genes upon loss of ATRX is both a cause and effect of the decreased transcription.

The question of what happens when transcribing RNA PolIII and the transcription machinery encounters a nucleosome is the matter of much study (reviewed recently in (Kulaeva et al., 2013; Subtil-Rodriguez and Reyes, 2011)). A picture is emerging where RNA PolII loops around nucleosomes causing transient disruptions between the DNA and the various histones. Upon passage of an initial RNA PolII, a single H2A/H2B dimer is displaced from the nucleosome octamer leaving behind a histone hexamer (the hexasome). For genes expressed at low levels the H2A/H2B dimer is replaced before the next RNA PolII encounters the nucleosome, meaning that the original hexasome and its modifications are maintained. If the next RNA PolII encounters the hexasome before H2A/H2B is replaced, such as in a high expressing gene, the entire nucleosome is displaced from the DNA and will need to be replaced, likely with an H3.3-containing nucleosome through replication independent assembly. Based on our informal observations of gene expression levels and given that H3.3 is not enriched across the entirety of the Dhrsx gene body, it appears that Dhrsx is not a very highly expressing gene. However, several studies have found that extremes of either AT or GC are enriched at generally nucleosome free regions (Fenouil et al., 2012; Schwarzbauer et al., 2012), and that CpG promoters can be enriched for H3.3 (Delbarre et al., 2010). The GC-rich nature this region could therefore increase histone instability, enhancing nucleosome octamer removal, and allowing for the incorporation of histone H3.3 after passage of the transcription machinery.

It is unclear exactly what the nucleosome state is when H3.3 is not deposited due to ATRX loss. Either the remaining nucleosomes could be redistributed across the depleted region, or some of the canonical histones could be reinserted. In Drosophila, cells are able to upregulate and insert H3 when H3.3 is deleted, and except for infertility are otherwise unaffected (Hodl and Basler, 2009; Sakai et al., 2009). Mice with deleted H3.3 exhibit neonatal lethality, stunted growth, neuromuscular deficits, and male sub-fertility, suggesting that mammals may not be able to overcome the loss of H3.3 as easily (Couldrey et al., 1999). In either case (redistribution or reinsertion), it is likely to lead to
at least a partial decrease in overall nucleosome density. A decrease in nucleosome density due to the inability to deposit H3.3 could itself lead to G-quadruplex formation, as G-quadruplexes are enriched in nucleosome-free regions (Halder et al., 2009; Wong and Huppert, 2009). Studying the overall composition of nucleosomes at DhrsX in the absence of ATRX would help shed light on this question.

### 4.6 A model for gene regulation by ATRX

By integrating the results of this work with the body of previously known knowledge related to ATRX, I suggest a model for gene regulation by ATRX, and speculate on what may happen when ATRX is absent (Figure 4-1). On newly synthesized DNA, canonical H3 is deposited, while the presence of G-quadruplexes that may form at the aPAR genes and at telomeres may attract ATRX to deposit H3.3. At this point, it is possible that nucleosome density is lower at the high GC regions (Figure 4-1A). Upon initial passage of RNA PolII, and possibly facilitated by a lower initial nucleosome density, single stranded DNA is generated and the presence of guanine triplicates leads to G-quadruplex formation. ATRX is targeted to the G-quadruplexes, perhaps facilitated by the common localization between telomeres, aPAR genes, and ATRX within PML-NBs. ATRX mitigates G-quadruplex formation using one or both of the following mechanisms: ATRX directly suppresses or resolves G-quadruplexes, and/or ATRX, in a complex with DAXX, deposits H3.3. Subsequent RNA PolII movement through the G-rich region is facilitated due to lack of G-quadruplexes and labile nature H3.3-containing nucleosomes (Figure 4-1B). In the absence of ATRX, loss of H3.3 deposition may lead to decreased nucleosome density which could be somewhat mitigated by deposition of H3. Transcription elongation by RNA PolII would be hindered by decreased nucleosome density which would promote the formation of G-quadruplexes and/or by more stable H3-containing nucleosomes (Figure 4-1C). H3.3 is the only remaining H3 variant in post mitotic brain cells, causing H3.3 levels to partially recover over time, even in the absence of ATRX, perhaps through a less efficient ATRX-independent salvage pathway. However, H3.3 levels remain below that necessary to fully facilitate aPAR gene expression.
Figure 4-1: A model for regulation of transcription by ATRX.

(A) H3.3 is deposited at newly synthesized DNA by ATRX targeting G-quadruplexes at telomeres and aPAR genes. (B) ATRX facilitates transcription at aPAR genes by directly binding and suppressing G-quadruplexes, and/or by acting with DAXX to deposit H3.3 after passage of the transcription machinery. (C) In the absence of ATRX, H3.3 is not deposited and G-quadruplexes inhibit transcriptional elongation at aPAR genes. Despite decreased telomeric H3.3, TERRA expression is unaffected. TSS, transcription start site.
4.7 Concluding remarks

The identification of ancestral pseudoautosomal region genes as ATRX targets was an unexpected finding, but pursuing this avenue of investigation has proven a valuable basis to study a model of gene regulation by ATRX. The characteristics of these genes – GC-rich, presence of G-quadruplexes, proximity to the mouse telomere and presence in the human pseudoautosomal region – enables their unique regulation by ATRX, and we have shown that this mechanism of regulation involves H3.3 deposition and facilitation of transcription. This role for ATRX in transcription at intragenic G-quadruplexes complements its role in replication at intergenic G-quadruplexes, and these two mechanisms likely work in concert to maintain proper cellular function. Certainly, a number of details remain to be addressed before we fully understand how ATRX regulates gene expression, such as what other proteins ATRX may be working with at these sites, and how a role for ATRX in transcription and replication may change throughout brain development. ATRX loss of function was originally linked to developmental defects, but has since also been linked to cancer. Other biological roles may be discovered in the future, and by understanding the molecular mechanisms by which ATRX functions, we will develop a better understanding of how genetic changes in ATRX lead to problems in development and disease.
4.8 References


Appendices

Appendix A: Permission to reproduce previously published work

The data and associated text in Chapter 2, sections 2.3.1 - 2.3.5 represents a previously published manuscript:


This material is re-used here in accordance with the policy of *BMC Genomics* that authors retain copyright to their work, and illustrations, figures, or tables may be reproduced provided that BioMed Central is duly identified as the original publisher, and that proper attribution of authorship and correct citation details are given.

Appendix B: Statement of permission for the use of animals for experimental research.

All animal studies were conducted in compliance with the regulations of The Animals for Research Act of the province of Ontario, the guidelines of the Canadian Council on Animal care, and the policies and procedures approved by the University of Western Ontario Council on Animal Care. This work was conducted in compliance with the animal use protocol 2008-041-02 held by Dr. Nathalie Bérubé, principal investigator at the Schulich School of Medicine and Dentistry and the department of Paediatrics at the University of Western Ontario in London, Ontario, Canada.
Curriculum Vitae

Name: Michael Levy

Post-secondary Education and Degrees:

2001-2005 B.Sc. (Honors) genetics, University of Western Ontario, London, Canada.

Honours and Awards:

2012 CIHR Inst. of Human Development, Child and Youth Health travel award.
2009-2011 NSERC CGS Ph.D. scholarship.
2009-2010 OGS Ph.D. scholarship (declined).
2010 UWO/LHRI Dept. of Paediatrics Research Day, best poster.
2007 VRL Research Day, 2nd place, platform competition.
2005-2006 CHRI Curtis Cadman Graduate Studentship.
2002-2005 Deans Honor List, Faculty of Science, UWO.
2001 Western Scholarship of Excellence admission scholarship.

Relevant Work Experience

2005 (summer) Student researcher, CHRI, London, Canada.
2004 (summer) Let’s Talk Science camp co-ordinator, London, Canada.

Relevant Volunteer Experience

2006-2009 Lawson Association of Fellows and Students, executive committee

Publications:


Levy MA, Fernandes AD, Tremblay DC, Seah C, Bérubé NG. The SWI/SNF protein ATRX co-regulates pseudoautosomal genes that have translocated to autosomes in the mouse genome. BMC Genomics 9, 486. 2008 Oct 8.

Selected Posters & Presentations


