ATRX promotes gene expression by facilitating transcriptional elongation through guanine-rich coding regions

Michael A. Levy1,2,3, Kristin D. Kernohan1,2,3, Yan Jiang1,2,3, and Nathalie G. Bérubé1,2,3,*

1Department of Paediatrics, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada N6A 5C1, 2Department of Biochemistry, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada N6A 5C1, and 3Children’s Health Research Institute, London, Ontario, Canada N6C 2V5

*To whom correspondence should be addressed. Tel: +1 5196858500, Ext. 55066; Fax: +1 5196858616; Email: nberube@uwo.ca

Abstract

ATRX is a chromatin remodeling protein involved in deposition of the histone variant H3.3 at telomeres and pericentromeric heterochromatin. It also influences the expression level of specific genes; however, deposition of H3.3 at transcribed genes is currently thought to occur independently of ATRX. We focused on a set of genes, including the autism susceptibility gene Neuroligin 4 (Nlgn4), that exhibit decreased expression in ATRX-null cells to investigate the mechanisms used by ATRX to promote gene transcription. Overall TERRA levels, as well as DNA methylation and histone modifications at ATRX target genes are not altered and thus cannot explain transcriptional dysregulation. We found that ATRX does not associate with the promoter of these genes, but rather binds within regions of the gene body corresponding to high H3.3 occupancy. These intragenic regions consist of guanine-rich DNA sequences predicted to form non-B DNA structures called G-quadruplexes during transcriptional elongation. We demonstrate that ATRX deficiency corresponds to reduced H3.3 incorporation and stalling of RNA polymerase II at these G-rich intragenic sites. These findings suggest that ATRX promotes the incorporation of histone H3.3 at particular transcribed genes and facilitates transcriptional elongation through G-rich sequences. The inability to transcribe genes such as Nlgn4 could cause deficits in neuronal connectivity and cognition associated with ATRX mutations in humans.

Introduction

ATRX is an SNF2-type chromatin remodeler enriched at heterochromatic regions including the inactive X chromosome (1), pericentromeric heterochromatin (2) and telomeres (3,4). ATRX forms a complex with the death-associated protein DAXX at pericentromeric heterochromatin and telomeres where it deposits the replication-independent histone variant H3.3 (5–7). However, the deposition of H3.3 at transcribed genes and regulatory elements is believed to occur independently of ATRX (6). Enrichment of ATRX at telomeric ‘TTAGGG’ tandem repeats may be mediated by the ability of ATRX to bind short sequences containing four guanine triplicates that form four stranded secondary structures called G-quadruplexes upon DNA denaturation during replication or transcription (8,9). It was proposed that ATRX might facilitate DNA replication through these structures by an unknown mechanism (4,10).

Decreased ATRX protein levels and/or activity resulting from mutations in the corresponding gene cause an intellectual disability syndrome called alpha thalassemia mental retardation, X-linked (ATR-X syndrome) (11). At a molecular level, ATR-X patients exhibit aberrant DNA methylation of repetitive DNA regions including ribosomal DNA, subtelomeric repeats and...
Y-chromosome DY22 repeats (12). More recently, ATRX mutations have been identified in cancers including pancreatic neuroendocrine tumors (13,14), brain cancer (13,15,16), osteosarcoma (17) and melanoma (18). Tumor 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 (13,16,19–21).

Several roles for ATRX in gene regulation have been suggested. ATRX chromatin immunoprecipitation (ChIP) sequencing (ChIP-seq) using human erythroblasts indicated that ATRX tends to bind G-rich sequences with high potential to form G-quadruplexes near the alpha-globin gene (3). Based on this, it was proposed that reduced expression of alpha-globin in ATR-X syndrome patients might be a consequence of stalled replication leading to DNA damage, replication fork collapse and/or loss of activating epigenetic marks, all of which could interfere with proper expression of nearby genes (10,22). The link between ATRX and DNA replication was indeed demonstrated in vivo. Conditional inactivation of Atrx in muscle or brain was shown to cause DNA replication stress (4,23). Treatment with telomestatin, a G-quadruplex ligand, exacerbated this effect in ATRX-null neuroprogenitors, suggesting that replication stress is linked to the stability of these DNA structures (4). ATRX can also influence gene expression through the repressive histone variant macroH2A (mH2A). In the absence of ATRX, increased enrichment of mH2A was observed across the alpha-globin gene cluster, suggesting that ATRX inhibits mH2A deposition thereby facilitating gene expression within this region (24). In the testes, ATRX was shown to interact with the androgen receptor protein to regulate the RhoX5 gene (25). Finally, we reported the regulation of imprinted gene expression by ATRX through the modulation of nucleosome positioning, CTCF binding and long-range chromatin interactions (26,27).

We previously reported that ancestral pseudoautosomal region (aPAR) genes are among the most downregulated genes upon loss of ATRX in the mouse forebrain throughout development (28). Mouse aPAR genes are located on autosomes while the equivalent genes in humans are found at the ends of the X and Y chromosomes (reviewed in 29). The most affected genes are dehydrogenase/reductase (SDR family) X-linked (Dhrsx) and colony stimulating factor 2 receptor, alpha (Csf2ra), which are located near the subtelomeric regions of mouse chromosomes 4 and 19, respectively. Other downregulated aPAR genes Asmtl and Cd99 have not yet been mapped to the mouse genome, indicating that they may lie within highly repetitive regions that are difficult to sequence and position within the assembled genome.

The aPAR genes are regulated by ATRX in both the developing and mature forebrain (28), suggesting a direct mechanism of regulation that is independent of DNA replication. Identifying direct mechanisms for the activation of gene transcription by ATRX could be relevant to the brain abnormalities and propensity to develop tumors upon mutation of the gene. Our investigations show that in the mouse brain, ATRX and H3.3 are enriched at the aPAR gene bodies (33). We identified 111 643 H3.3 binding sites in these ATRX-null samples and they displayed a highly similar distribution to that seen in wild-type brain (Fig. 1C and Supplementary Material, Fig. S2B). Using a threshold of 2-fold change, we found that 16 845 H3.3 sites displayed increased enrichment and 10 748 sites showed decreased enrichment in ATRX-null samples. The distribution of altered H3.3 peaks was similar to that of unaltered peaks, with a few minor differences. In ATRX-null cells, H3.3 occupancy was altered in intergenic regions at a higher frequency than expected, based on the distribution of unchanged sites, while peaks in promoters and 5′ UTRs were less likely to be changed (Fig. 1D). Average GC content was similar between increased and decreased peaks (57.3 and 55.8%). The distribution of increased or decreased peaks across chromosomes was also similar and generally mirrored the distribution of genes. Interestingly, H3.3 peaks (whether changed or unchanged) showed an enrichment toward telomeres similar to that seen for ATRX (3) (Supplementary Material, Fig. S2A). Lastly, there was no strong correlation between altered H3.3 enrichment and changes in gene expression in the ATRX-null forebrain (28), although genes exhibiting increased expression sometimes displayed higher H3.3 enrichment (Supplementary Material, Fig. S2C). Overall, these findings indicate that ATRX regulates histone H3.3 occupancy in the forebrain at many sites across the genome. However, our observation that H3.3 enrichment is increased at multiple sites suggests that the relationship between ATRX and H3.3 is more complex than previously appreciated.

We noted that the enrichment of histone H3.3 mirrors that of ATRX at Dhrsx, Csf2ra, Asmtl and Cd99. H3.3 enrichment was decreased at all four of the above genes in the absence of ATRX, with peak heights being one-half to one-third compared with controls (Fig. 1A and E). ChIP-qPCR for H3.3 at Dhrsx and Csf2ra confirmed reduced occupancy of the histone variant in forebrain tissue (Supplementary Material, Fig. S3). Thus, high levels of H3.3

Results

ATRX-mediated enrichment of H3.3 at G-rich segments of aPAR gene bodies

We previously reported that ATRX regulates the expression of Dhrsx, Csf2ra and other aPAR genes in the mouse brain (28). Csf2ra is the most telomeric gene on mouse chromosome 19, located ∼205 kb away from the telomere in mouse genome version mm10. Dhrsx is most likely subtelomeric on chromosome 4 as demonstrated by FISH (30) and by its placement ∼110 kb from the end of an incomplete chromosome 4 (Mm Celera) (31). Given that ATRX regulates H3.3 enrichment and transcription at telomeres (6), we investigated the possibility that it might directly bind aPAR genes in the mouse brain. We analyzed data from published ATRX ChIP-seq performed in mouse ESCs (3). Our analysis revealed that ATRX is highly enriched in the gene bodies of Dhrsx, Asmtl, Cd99, and Csf2ra, indicating that these genes are direct targets of ATRX (Fig. 1A and E). We noted that the segments of the genes bound by ATRX are G-rich and have a high potential of forming G-quadruplex structures as predicted by Quadfinder (23) (Fig. 1A and E). We confirmed that ATRX binds the 5 kb G-rich region of Dhrsx in neonatal mouse brain using ChIP-qPCR with two different anti-ATRX antibodies (Fig. 1B).

The involvement of ATRX in histone H3.3 incorporation (6) led us to examine H3.3 distribution across the P0.5 mouse forebrain genome using ChIP-seq. In wild-type control mice, we identified 222 246 sites of enrichment, the largest number of sites being found in gene bodies (Fig. 1C). To determine the effect of ATRX loss, we used mice with conditional inactivation of Atrx in the forebrain (33). We identified 111 643 H3.3 binding sites in these ATRX-null samples and they displayed a highly similar distribution to that seen in wild-type brain (Fig. 1C and Supplementary Material, Fig. S2B). Using a threshold of 2-fold change, we found that 16 845 H3.3 sites displayed increased enrichment and 10 748 sites showed decreased enrichment in ATRX-null samples. The distribution of altered H3.3 peaks was similar to that of unaltered peaks, with a few minor differences. In ATRX-null cells, H3.3 occupancy was altered in intergenic regions at a higher frequency than expected, based on the distribution of unchanged sites, while peaks in promoters and 5′ UTRs were less likely to be changed (Fig. 1D). Average GC content was similar between increased and decreased peaks (57.3 and 55.8%). The distribution of increased or decreased peaks across chromosomes was also similar and generally mirrored the distribution of genes. Interestingly, H3.3 peaks (whether changed or unchanged) showed an enrichment toward telomeres similar to that seen for ATRX (3) (Supplementary Material, Fig. S2A). Lastly, there was no strong correlation between altered H3.3 enrichment and changes in gene expression in the ATRX-null forebrain (28), although genes exhibiting increased expression sometimes displayed higher H3.3 enrichment (Supplementary Material, Fig. S2C). Overall, these findings indicate that ATRX regulates histone H3.3 occupancy in the forebrain at many sites across the genome. However, our observation that H3.3 enrichment is increased at multiple sites suggests that the relationship between ATRX and H3.3 is more complex than previously appreciated.

We noted that the enrichment of histone H3.3 mirrors that of ATRX at Dhrsx, Csf2ra, Asmtl and Cd99. H3.3 enrichment was decreased at all four of the above genes in the absence of ATRX, with peak heights being one-half to one-third compared with controls (Fig. 1A and E). ChIP-qPCR for H3.3 at Dhrsx and Csf2ra confirmed reduced occupancy of the histone variant in forebrain tissue (Supplementary Material, Fig. S3). Thus, high levels of H3.3
incorporation in gene body nucleosomes correlate with ATRX occupancy at a subset of sites, while genome-wide patterns of H3.3 occupancy at transcription start site (TSS), gene bodies and TES are not altered in ATRX-null cells. We next analyzed H3.3-EYFP occupancy in control and ATRX-null mouse ESCs from published ChIP-seq data (6). The pattern of H3.3-EYFP enrichment in ESCs closely mirrored that of the mouse forebrain. H3.3 occupancy was observed at the four aPAR genes and was reduced in the

---

**Figure 1.** ATRX and H3.3 enrichment at Dhrsx and other aPAR genes. (A) Re-aligned ChIP-seq data from previously published studies of ATRX (3) and H3.3 (6) in mouse embryonic stem cells (ESCs), and novel ChIP-seq for H3.3 in the P0.5 mouse forebrain visualized at Dhrsx. ATRX and H3.3 are highly enriched across 5 kb at the 3′ end of Dhrsx and H3.3 is reduced in the absence of ATRX. Enrichment overlaps with high GC content and putative G-quadruplexes. Input tracks show negligible levels of enrichment (see Supplementary Material, Fig. S1). (B) ChIP-qPCR for ATRX using antibodies H300 and D5 in the P0.5 mouse forebrain showed high enrichment of ATRX within the binding region identified by ChIP-seq. Error bars represent standard error of the mean. (C) Distribution of H3.3 ChIP-seq peaks in the mouse forebrain. The random distribution was generated by plotting the location of 100 000 randomly chosen reads from the input sample. Promoter and 3′ flank represent 1 kb up- and downstream, respectively, from the gene body. (D) Distribution of changed (2-fold or greater) versus unchanged peaks in the ATRX-null forebrain. (E) ChIP-seq data at Asmtl, Cd99 and Csf2ra show overlapping enrichment of ATRX and H3.3 and corresponding decrease in H3.3 levels in the absence of ATRX. Grey arrows in Csf2ra tracks indicate probable non-specific enrichment only seen when single-end short reads are used (36/40 bp) but not when longer (100 bp) paired-end reads are used. Input tracks show negligible enrichment (see Supplementary Material, Fig. S1).
absence of ATRX (Fig. 1A and E). Since histone regulator A (HIRA) is the primary histone chaperone involved in depositing H3.3 at transcribed genes, we analyzed H3.3-EYFP in HIRA-null ESCs (6). H3.3 levels at the aPAR genes were unchanged between control and HIRA-null ESCs, indicating that it is not the chaperone involved at these genes (Supplementary Material, Fig. S4). Our results suggest that H3.3 is not incorporated or maintained as efficiently in the ATRX/H3.3/G-rich region of aPAR genes when ATRX is absent, potentially implicating H3.3 deposition in the regulation of transcription at specific ATRX target genes. Alternatively, decreased H3.3 occupancy at these sites might be a secondary effect of decreased transcription of the gene in the absence of ATRX.

ATRX deficiency in the mouse forebrain or embryonic fibroblasts does not affect TERRA levels

H3.3 occupancy at telomeres in mouse ESCs was reported to be ATRX dependent (6,7). To determine whether this is also the case in the newborn mouse forebrain, we analyzed H3.3 enrichment at repetitive sequences in control and ATRX-null tissue using ChIP-seq data. H3.3 occupancy at telomere sequences (‘TTAGGG’ and ‘CCCTAA’) was enriched 3.5-fold over input in the control forebrain and was decreased 2.8-fold compared with control in the absence of ATRX (Fig. 2A). This result is similar to previously published data showing an ~4-fold decrease in H3.3 enrichment at telomere sequences in ATRX-null mouse ESCs (6). These findings support the idea that ATRX assists H3.3 deposition at telomeres not only in undifferentiated ESCs as previously described (6), but also in post-mitotic brain cells. H3.3 was also detected at rDNA repeats and was moderately decreased in the absence of ATRX (Fig. 2A), as described in mouse ESCs (6).

Expression of TERRA non-coding RNA (34,35) from telomeric tandem repeats is increased ~1.7-fold in ATRX-deficient mouse ESCs and correlates with reduced H3.3 enrichment at telomeres (6). Given that increased TERRA expression can promote heterochromatin formation at telomeres (36), we evaluated whether altered expression of subtelomeric aPAR genes in the ATRX-null brain may be related in part to their proximity to telomeres due to a telomere position effect (37–39). We performed RNA-seq of control and ATRX-null E14 forebrains. TERRA RNA [(TTAGGG)₁₀ and (CCCTAA)₁₀] showed only a 1.16-fold increase in the ATRX-null embryonic forebrain (Fig. 2B). Confirmation was obtained from a series of radioactive RNA dot blots probed for TERRA.

Figure 2. TERRA expression is not altered in forebrain tissue or mouse embryonic fibroblasts lacking ATRX. (A) ChIP-seq for H3.3 shows that it is enriched at telomeres in the mouse forebrain and decreased in the absence of ATRX. Simple repeats represent all simple repeats minus telomere repeats. (B) RNA-seq of control and ATRX-null E14 forebrains. TERRA RNA [(TTAGGG)₁₀ and (CCCTAA)₁₀] showed only a 1.16-fold increase in the ATRX-null embryonic forebrain (Fig. 2B). Confirmation was obtained from a series of radioactive RNA dot blots probed for TERRA.
transcripts (Fig. 2C). To verify these findings in a different system, we examined the effects of ATRX deficiency on TERRA expression in Atrx-floxed MEFs (33) infected with adenoviral vectors expressing Cre recombinase. qRT-PCR and immunofluorescence assays show nearly undetectable levels of Atrx RNA and protein upon Cre recombinase expression in infected cells (Supplementary Material, Fig. S5A and B). Both Csf2ra and Dhrsx showed a 2.7-fold decrease in transcript levels in ATRX-null MEFs (Supplementary Material, Fig. S5B), a result analogous to that seen in the mouse forebrain (28). Thus, loss of ATRX in vitro in MEFs recapitulates the aPAR gene expression defects observed in vivo in brain tissue. Again, we observed that TERRA transcript levels were not significantly different between control MEFs and those lacking ATRX (Fig. 2D), suggesting that TERRA levels are not substantially affected by ATRX deficiency and that TERRA transcription is regulated in a cell-type-specific manner. We conclude that aPAR gene silencing upon loss of ATRX is likely not influenced by an increase in TERRA transcript levels.

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

The PAR region on the human X chromosome is GC-rich, and Dhrsx, despite having translocated to mouse autosomes, maintains this feature. Given the altered pattern of DNA methylation in ATR-X patients (12), and the presence of CpG islands at the subtelomeric target genes in the mouse, we examined 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 highly methylated (98%) and methylation status was maintained in the absence of ATRX (Fig. 3A). We extended this analysis to Csf2ra, which has a small CpG island at the promoter. The Csf2ra CpG island was ~65% methylated in two littermate-matched ATRX-null and control pairs (Supplementary Material, Fig. S6), indicating that ATRX does not influence expression of these particular genes by altering DNA methylation, at least at the sites tested here.

We next analyzed a subset of histone post-translational modifications within and outside the Dhrsx gene. Promoter regions including the TSS of active genes typically exhibit elevated acetylation of histone H4 (H4Ac) and trimethylation of lysine 4 of histone H3 (H3K4me3). These active sites are also marked by low levels of histone H3 trimethylation on lysine 9 (H3K9me3) and lysine 27 (H3K27me3) (40–42). Enrichment of these histone modifications was examined across the 20 kb Dhrsx gene region, 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 (Fig. 3B). In the absence of ATRX, no significant changes were observed in the pattern of enrichment of these histone modifications, except for a small increase in H3K9me3 upstream of the ATRX/H3.3 binding region (Fig. 3B). Thus, altered CpG island DNA methylation or key histone post-translational modifications examined here cannot explain the effects of ATRX deficiency on gene expression.

Increased RNA PolII occupancy at the ATRX/H3.3/G-rich region of Dhrsx in the absence of ATRX

Our observation that ATRX binds the gene body of the aPAR genes suggested that decreased expression might be caused by a problem in transcriptional elongation. To test this possibility, we performed RNA polymerase II (PolII) ChIP-seq of PolII control and ATRX-null forebrain. PolII in both the presence and absence of ATRX was over-represented in promoters and 5′ UTRs (Supplementary Material, Fig. S4A). Using a threshold of 2-fold change, we identified 3310 peaks showing increased PolII enrichment and 795 showing decreased enrichment in the ATRX-null forebrain. Altered PolII occupancy was more frequent at intergenic sites and in coding sequences, while promoters and 5′ UTR PolII peaks were less likely to be changed in the absence of ATRX (Fig. 4B). Average GC content was similar at 51.3 and 49.2% for increased and decreased peaks, respectively, as was the distribution of changed peaks across chromosomes (Supplementary Material, Fig. S2A). Genes with increased expression (28) had a slightly higher enrichment of PolII (Supplementary Material, Fig. S2B and C), suggesting that many genes exhibit increased PolII occupancy upon loss of ATRX.

At Dhrsx, we observed moderate levels of PolII at the promoter/TSS and background levels across the gene. In the absence of ATRX, PolII occupancy was increased within the ATRX/H3.3-binding/G-rich region of Dhrsx (Fig. 4C) and ChIP-qPCR across Dhrsx confirmed this result (Supplementary Material, Fig. S8A). ChIP-seq analysis also revealed increased occupancy of PolII at Asnl and Cd99, suggesting a common mechanism between these affected genes (Fig. 4D). PolII accumulation suggests that the enzyme might be more prone to stalling in the absence of ATRX, which is predicted to either decrease the rate of elongation or to stop elongation altogether, resulting in lower levels of gene transcripts (Fig. 4D). Csf2ra was an exception and had background levels of PolII occupancy by ChIP-seq and ChIP-qPCR without notable changes in the absence of ATRX (Supplementary Material, Fig. S8A and B). We therefore cannot yet explain how the interactions between ATRX, H3.3 and Csf2ra contribute to the decreased expression of this gene. However, we note that the overall level of ATRX and H3.3 and magnitude of H3.3’s decrease upon loss of ATRX is less at Csf2ra than that seen at the other aPAR genes (Fig. 1E). The mechanism may therefore be different for this gene and would need to be further investigated.

Evidence of decreased H3.3 and stalling of RNA PolII at Nlgn4 in the ATRX-null forebrain

Having identified a pattern of ATRX, H3.3 and PolII enrichment at the downregulated aPAR genes, we next wanted to see whether any other genes were regulated in a similar manner. We re-examined our microarrays of control and ATRX-null mouse brain tissue (28) and identified one candidate probe set (1429245_at) that showed decreased expression at E13.5 and P0.5 of 1.7- and 2.0-fold, respectively. This probe set is not associated with any other genes were regulated in a similar manner. We re-examined our microarrays of control and ATRX-null mouse brain tissue (28) and identified one candidate probe set (1429245_at) that showed decreased expression at E13.5 and P0.5 of 1.7- and 2.0-fold, respectively. This probe set is not associated with any particular gene and its chromosomal location is unknown. BLASTn found highest similarity to a 20 kb mouse gDNA fragment (accession EU350930.1) containing the gene neurexin 4 (Nlgn4). NLGN4 is a neuronal postsynaptic cell-adhesion molecule, and mutations in this gene were identified in several cases of autism (43–45). In mice, Nlgn4 is located on an unknown autosome (46). In humans, it is located on the X-chromosome (NLGN4X) 3 MB outside the PAR boundary and a nearly identical gene (NLGN4Y) is located centrally on the Y chromosome (46), suggesting that it is in fact pseudautosomal in humans. It has been believed to have been within the ancestral PAR (29) and is in the PAR of several current mammalian species (29). We first confirmed that the expression of Nlgn4 is decreased in the ATRX-null mouse forebrain by qPCR (Fig. 5A). To determine the status of
Figure 3. DNA methylation and histone modifications at Dhrsx in the absence of ATRX. (A) Bisulfite mutagenesis and sequencing at Dhrsx reveal low levels of DNA methylation (12%) within the CpG island overlapping a putative promoter region (C1), while the 3′ CpG island (C2) is highly methylated (98%). Levels of DNA methylation were similar between P0.5 control and ATRX-null forebrains (n = 2). Empty circles represent unmethylated cytosines, while filled circles represent methylated cytosines. TSS, transcription start site. (B) ChIP-qPCR results for the specified histone modifications at sites indicated by the numbered amplicons. Amplicons 1 and 17 are located 15 kb upstream and downstream, respectively, from the indicated locations. ChIP for the histone modifications H4Ac and H3K4me3 were normalized to the Gapdh promoter and showed enrichment at the Dhrsx putative promoter region and TSS (primer sites 2 and 3). The repressive modifications H3K9me3 and H3K27me3 were normalized to the Myod1 promoter. Error bars represent standard error of the mean for n = 3 (n = 2 for H3K9Me3). ChIP-qPCR at the Gapdh and Myod1 promoters are shown in Supplementary Material, Figure S7.
ATRX, H3.3 and PolII enrichment at the unmapped Nlgn4 gene, we added the 20 kb mouse gDNA fragment to the mouse genome sequence and realigned all ChIP-seq data sets. As with other aPAR genes, ATRX and H3.3 were bound to the gene body. In the ATRX-null forebrain, H3.3 occupancy was decreased and PolII increased. We also noted that Nlgn4 has high GC content and potential G-quadruplex forming sequences (Fig. 5B). Overall, Nlgn4 expression depends on ATRX and the gene exhibits all the features of the other aPAR genes that indicate a problem in transcriptional elongation.

Discussion

In this study, we investigated mechanisms of gene activation by the ATRX chromatin remodeling protein. We found that both ATRX and the histone variant H3.3 are enriched at G-rich segments of the aPAR gene bodies. In the absence of ATRX, the levels of H3.3 are reduced at these sites, demonstrating for the first time that ATRX can influence H3.3 deposition not only at telomeres, but also at a subset of active genes. We demonstrate that RNA PolII accumulates at these G-rich DNA elements, which suggests stalling in the absence of ATRX, thus providing an explanation for reduced transcript levels.

Increased TERRA telomeric transcripts and H3.3 depletion at telomeres were previously detected in ATRX-null mouse ESCs (6) which could conceivably affect the expression of genes proximal to the telomere. Nevertheless, we failed to detect changes in TERRA expression in either ATRX-null MEFs or mouse forebrain tissue, indicating that regulation of Dhrs and Csf2ra expression by ATRX is independent of TERRA. The association of ATRX with telomeres, and between ATRX loss and TERRA expression, is variable across studies and biological systems. ATRX is enriched at telomeres in mouse ESCs (3,7). ChIP-seq showed that ATRX is present at telomere repeats in human erythroid cells, but the level of enrichment (16-fold over input) was less compared with mouse ESCs (40-fold over input) (3). ATRX is released from telomeres after neuronal differentiation of mouse ESCs (7) and detected at only a subset of telomeres in mouse neuroprogenitors (4). Furthermore, no correlation between loss of ATRX and levels of TERRA was detected in a series of ATRX-null human ALT cell lines (21), indicating that there is not a strict relationship between ATRX loss and increased TERRA expression. Taken together, the absence of a change in TERRA levels, and the smaller decrease in telomeric H3.3 seen in the forebrain compared with ESCs, indicates that ATRX has a more modest effect at telomeres or that ATRX only acts at a subset telomeres in the forebrain. Rather, the reduction of telomeric H3.3 likely contributes to the general telomere instability we previously detected in the ATRX-null brain (4,47). In addition, effects may be muted in the mixed population of cells in the forebrain compared with a pure population of ESCs. Further studies will be needed to determine whether diverse cell types in the brain respond differently to the loss of ATRX.

Reduced ATRX protein levels or activity correlate with altered patterns of DNA methylation of several repetitive regions in ATRX syndrome patients (12) and in ATRX-null mice (26,48). However, loss of ATRX does not affect DNA methylation at ATRX-regulated imprinted domains (26) or at Dhrs and Csf2ra CpG islands. Additionally, histone modification profiles across Dhrs appear to be unaffected. The histone modifications at the Dhrs promoter are typical of an active gene and are not changed in the absence of ATRX. This observation indicated that the mechanism of
regulation probably does 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 located within gene bodies, yet intragenic H3.3 binding was reported to be overall intact upon loss of ATRX (3,6). Instead, HIRA is thought to be the histone chaperone responsible for depositing H3.3 within gene bodies (6). In contrast, our findings show that ATRX and H3.3 are co-localized in gene bodies of several genes studied here and that the level of H3.3 at these sites depends on ATRX. H3.3 is traditionally associated with active transcription (49), in part due to its rapid turnover rate, which promotes increased DNA accessibility to transcription factors and passage of the transcription machinery (50,51). We previously reported that increased G-quadruplex stability by treatment with telomestatin exacerbated DNA replication stress in the absence of ATRX, suggesting a role for ATRX in DNA replication of G-quadruplex forming structures (4). In the current study, RNA polymerase II ChIP revealed a build-up of the enzyme at the ATRX/H3.3/G-rich region of Dhrx and other aPAR genes in the absence of ATRX. PolII accumulation can indicate pausing of the transcriptional machinery and supports the idea that ATRX assists transcriptional elongation of G-quadruplex-forming DNA in a process involving H3.3 deposition.

A role for G-quadruplexes in gene expression is usually associated with inhibition at promoters, one of the best studied being the MYC oncogene promoter where stabilizing the structures represses transcription (52). G-quadruplexes also play a role in inhibiting transcription as part of normal cellular processes. They are often found downstream of TSSs and are correlated with promoter-proximal pausing (53). They also occur downstream of poly(A) signals where they may be involved in PolII pausing leading to transcription termination at some genes (54). There are a few examples in the literature of transcriptional pausing in G-rich gene bodies. Stabilization of G-quadruplexes within the HIV-1 nef coding region can attenuate transcription (55), and formation of G-quadruplexes within a construct containing the c-Myb proto-oncogene was shown to arrest its T7 RNA polymerase-mediated transcription (56).

Since putative G-quadruplex-forming sequences were identified within the ATRX/H3.3/G-rich binding domains of the aPAR genes, one attractive possibility is that ATRX binds to G-quadruplexes that form during transcription elongation and helps to unwind these structures. However, a recent report demonstrated that ATRX itself cannot unwind G-quadruplexes (57). Alternatively, the incorporation of H3.3 into chromatin might prevent the formation of complex DNA structures such as G-quadruplexes. Further studies are required to elucidate the full mechanism of ATRX action during transcription elongation.

The mouse Nlgn4 gene has undergone rapid evolution, a hallmark of PAR/aPAR genes (30,58,59), and has therefore diverged considerably from its human counterpart. Nevertheless, the mouse NLGN4 protein was demonstrated to retain functionality such as its ability to bind neurexins (46), and its importance for proper synapse maturation and neurological function has been demonstrated (46,60,61). In humans, mutations in NLGN4 are associated with autism spectrum disorder (43,44,62). The identification of Nlgn4 as a direct target of ATRX provides a potential link between ATRX mutations and some of the phenotypic outcomes, including autistic-like behaviors, seizures and intellectual disability.

Our study provides mechanistic insight of ATRX-mediated transcriptional elongation. We propose a model whereby incorporation of H3.3, as directed by ATRX, promotes transcriptional elongation through G-rich intragenic sequences by organizing chromatin into a state that limits the formation or allows the bypass of G-quadruplex DNA structures (Fig. 4D). DAXX is the H3.3 chaperone identified for ATRX-mediated deposition of H3.3 at telomeres (5) and pericentromeric heterochromatin (63). An important future goal will be to determine whether DAXX is present with ATRX at these genes, or whether a different histone chaperone is involved.

Materials and Methods

Mouse husbandry and genotyping

Atrx was conditionally deleted in the mouse forebrain and mice were genotyped as previously described (33,64). Animal studies were in accordance with the policies and regulations of The Animals for Research Act of the province of Ontario, the Canadian Council on Animal Care and the University of Western Ontario Council on Animal Care.

**Figure 5.** The transcription of Nlgn4 is controlled by ATRX. (A) RT-qPCR for Atrx and Nlgn4 in the P0.5 control and ATRX-null mouse forebrain, normalized to Gapdh and B-actin. The locations of Nlgn4 primers are indicated in (B). Error bars represent standard error of the mean for n = 4. *p < 0.005. (B) ChIP-seq data from previously published studies of ATRX (3) and H3.3 (6) in mouse ESCs, and novel ChIP-seq for H3.3 in the P0.5 mouse forebrain, aligned to mouse genome mm9 with the addition of 20 kb gDNA fragment (accession EU350930) containing Nlgn4. ATRX and H3.3 are enriched in the gene body. H3.3 levels are lower in the ATRX-null brain while PolII occupancy appears to be increased. Grey arrows indicate peaks that also show enrichment in the input and may therefore be all or in part artifacts (see Supplementary Material, Fig. S1A).
Mouse embryonic fibroblast isolation, culture and infection

Mouse embryonic fibroblasts (MEFs) were isolated from male E13.5 AtrloxP embryos (33) and cultured in Dulbecco’s modified Eagle’s medium with 10% FBS (Sigma-Aldrich), Pen Strep (Gibco) and GlutaMAX (Invitrogen). 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 h post-infection.

RNA isolation and transcriptional assays

Whole-cell RNA was isolated using the RNeasy Mini Kit (Qiagen) while cytoplasmic and nuclear RNA was extracted using the RNeasy Plus Mini Kit (Qiagen) with the supplemental protocol, ‘Purification of cytoplasmic RNA from animal cells using the RNeasy Mini Kit’ (Qiagen). cDNA was prepared from 1 µg of RNA using SuperScript II Reverse Transcriptase (Invitrogen). Real-time PCR was performed as previously described (26). Primer sequences are listed in Supplementary Material, Table S1.

DNA dot blots

DNA was spotted onto Amersham Hybond-XL membranes (GE Healthcare Life Sciences,GE). Probes were generated using T4 polynucleotide kinase (NEB), ATP[y-32P] 3000 Ci/mmol (Perkin-Elmer) and oligos complementary to TERRA or Gapdh (85), then purified using illustra MicroSpin G-25 columns (GE). Membranes were pre-hybridized in Amersham Rapid-Hyb Buffer (GE) for 20 min at 42°C, hybridized with 25 µl labelled TERRA probe for 1 h at 42°C, washed according to the Hybond-XL instructions and then exposed to Amersham Hyperfilm MP (GE) at ~80°C. Blots were stripped, hybridized with the Gapdh probe and re-exposed. Dots were quantified using the integrated density method with ImageJ (66) and normalized to Gapdh levels.

Bisulphite mutagenesis and sequencing

Bisulphite treatment was performed as described (67) except the starting material was gDNA purified from neonatal mouse forebrains using the DNeasy Blood & Tissue Kit (Qiagen) and the final DNA product was purified using the QIAquick Gel Extraction Kit (Qiagen). Nested PCR (see Supplementary Material, Table S1 for primer sequence) was performed, and purified amplicons were ligated into a pGEM-T Easy cloning vector using T4 DNA Ligase (both Promega) and clone inserts were sequenced at the London Regional Genomics Centre, London, Canada.

Chromatin immunoprecipitation

All reagents are from Sigma-Aldrich unless noted. ChIP was performed as previously described (26) except fixation was performed at 37°C, and the LiCl wash was omitted. For ChIP-seqencing, samples were pre-cleared with 20 µl of ChIP-Grade Protein G Magnetic Beads (Cell Signalling). Antibodies were added and immunoprecipitation reactions were incubated 16–20 h at 4°C. Amounts of antibody were 15 µg for H300 and D5, 3 µl where concentration unknown and 3 µg for the remaining antibodies. Antibodies (from Millipore unless noted) used were rabbit IgG (PP64B), H4Ac (06–866), H3K4Me3 (04–745), H3K9Me3 (07–442), Histone H3.3 (17–10245), RNA PolII (05–623), mouse IgG (Santa Cruz sc–2025), ATRX D5 (Santa Cruz, sc–55584) and ATRX H300 (Santa Cruz, sc–15408). Chromatin/antibody complexes were collected with 20 µl magnetic beads, then washed, eluted and DNA purified as previously described [EZ-ChIP protocol (Upstate) and (26)]. Real-time PCR reactions were performed, and percent input was calculated as previously described (26).

Next-generation sequencing and analysis

For H3.3 and RNA PolII ChIP-seq in the mouse forebrain, purified ChIP DNA form control and ATRX-null forebrain tissue was sent to The Centre for Applied Genomics at the Hospital for Sick Children, Toronto, Canada. Thirty to forty million 100 base pair, paired-end reads were generated for each sample using an Illumina HiSeq 2000. For ATRX and H3.3 ChIP-seq in ESCs (3,6), sequence reads were downloaded from the NCBI Sequence Read Archive (Accession numbers GSE22162 and GSE16893). Sequences added to the mouse genome mm9 were Asmtl (AK084779), Cdf99 (NW_016967) and Nlgn4 (EU1350930). Reads were aligned using Bowtie version 1.0 (68). Custom Perl scripts were used to remove duplicates and generate genome-wide coverage tracks (wig files) normalized to 10 million reads by multiplying each data set by a constant normalization factor. These were viewed using the UCSC Genome Browser (69). Peaks were called using MACS2 (70) and peaks within 200 bp (PolII) or 500 bp (H3.3) were merged to generate a list of enriched sites. Custom Perl scripts were used to identify called peaks changed 2-fold or greater between control and ATRX-null samples, annotate sites with gene features and generate density profiles. To identity enrichment at repeat sequences, reads were aligned to a Bowtie index created as previously described (71). G-quadruplexes in the Asmtl, Cdf99, Csfr2a, Dhrsx and Nlgn4 regions were identified using Quadfinder (72). For RNA-seq, frozen P0.5 mouse forebrains were sent to Oto-genetics Corporation, Norcross, GA, USA, for processing and sequencing, and telomere enrichment was determined by counting reads that consisted of at least 80% telomere repeats (TTAGGG or CCCTAA).

Supplementary Material

Supplementary Material is available at HMG online.

Acknowledgements

We thank Drs Douglas Higgs and Richard Gibbons (Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, Oxford, UK) for kindly providing the AtrloxP mice used in this study, and Ashley Watson (Western University, Canada) for providing frozen mouse forebrain samples for RNA sequencing.

Conflict of Interest statement. None declared.

Funding

This work was supported by the Canadian Institutes for Health Research (MOP93697) to N.G.B.; and Natural Sciences and Engineering Research Council of Canada Alexander Graham Bell Canada Graduate Scholarships to M.A.L. and K.D.K.

References


