Title: Functional significance of mutations in the Snf2 domain of ATRX

Authors: Matthew Mitson[1], Lawrence A. Kelley[2], Michael J.E. Sternberg[2], Douglas R. Higgs[1] and Richard J. Gibbons[1*]

[1]MRC Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, Headington, Oxford OX3 9DS, UK

[2]Structural Bioinformatics Group, Division of Molecular Biosciences, Department of Life Sciences, Imperial College London, London SW7 2AZ, UK

*Correspondence should be addressed to: richard.gibbons@imm.ox.ac.uk
Dr Richard J Gibbons, MRC Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, Headington, Oxford OX3 9DS, UK
Phone +44 1865 222632
Fax +44 1865 222424

Words in abstract: 174

Key Words: ATR-X syndrome; Snf2 domain; mutations

Abstract

ATRX is a member of the Snf2 family of chromatin remodelling proteins and is mutated in an X-linked mental retardation syndrome associated with alpha-thalassaemia (ATR-X syndrome). We have carried out an analysis of 21 disease causing mutations within the Snf2 domain of ATRX by quantifying the expression of ATRX protein and placing all missense mutations in their structural context by homology modelling. While demonstrating the importance of protein dosage to the development of ATR-X syndrome we also identified three mutations which primarily affect function rather than protein structure. We show that all three of these mutant proteins are defective in translocating along DNA while one mutant, uniquely for a human disease causing mutation, partially uncouples ATP hydrolysis from DNA binding. Our results highlight important mechanistic aspects in the development of ATR-X syndrome and identify crucial functional residues within the Snf2 domain of ATRX. These findings are important for furthering our understanding of how ATP hydrolysis is harnessed as useful work in chromatin remodelling proteins and the wider family of nucleic acid translocating motors.

Introduction

Over the past 20 years it has become increasingly clear that a wide range of nuclear processes (recombination, replication, repair and transcription) are influenced by the manner in which DNA is packaged into chromatin. Members of the large Snf2 family of proteins (e.g. Swi/Snf, ISWI, Mot1, Mi-2) act as molecular motors and some (but not all) use the movement between nucleic acid and protein to alter the accessibility of DNA wrapped around a nucleosome (1). However, the role of many Snf2 proteins and the mechanisms by which they exert their effects *in vivo* are unknown.

Some important clues come from structure-function studies based on sequence conservation and/or mutagenesis. The Snf2 family belongs to the SF2 helicase superfamily which together with helicase SF1 share 7 conserved motifs (I, Ia, II, III, IV, V and VI) (2). In addition the Snf2 family can be subdivided into 24 Snf2 subfamilies with additional conserved motifs (3). These motifs are involved in ATP hydrolysis, the motor function of the molecule and communicating the motor state through the protein. However much remains to be learnt as to how the energy of ATP hydrolysis is used by these proteins to drive their molecular motors and how the different motifs co-ordinate their activities.

ATRX in complex with the histone chaperone DAXX can assemble H3.3 containing nucleosomes on DNA *in vitro* (4, 5) and *in vivo* is required for histone H3.3 deposition at telomeres and pericentromeric repeats (5-7). Furthermore recent

studies locate ATRX at G-rich tandem repeats (TRs) in telomeres and euchromatin where it may recognise unusual DNA structures (8). When ATRX is mutated it gives rise to a syndromal form of X-linked mental retardation (ATR-X syndrome). In affected individuals the expression of genes in the proximity of these G-rich TRs may be perturbed helping to explain one aspect of the condition, alpha thalassaemia seen in many cases and associated with down regulation of the alpha globin genes. (9).

The 82 missense mutations found in patients with ATR-X syndrome are almost exclusively confined to its N-terminal PHD-like domain (known as the ADD domain) and the conserved Snf2 domain (10). Here, we analyse 21 mutations identified in the Snf2 domain of individuals with ATR-X syndrome. Many mutations lead to protein instability demonstrating the importance of ATRX dosage in the development of ATR-X syndrome; others primarily affect protein function. By building an homology model of the Snf2 domain we are able to explain the structural consequences of most of the mutations. We have also expressed recombinant wild-type and mutant ATRX for three functional mutations and analysed their behaviour in ATPase and DNA translocation assays showing that one novel mutation uncouples ATPase activity from DNA binding thus highlighting an intra-molecular interaction that may link DNA binding and ATP hydrolysis to the Snf2 molecular motor.

Results

Some ATRX mutations produce relatively normal levels of protein

Naturally occurring missense mutations in ATRX are largely confined to the ADD and helicase domains in regions where the sequence is conserved and the structure is predicted to be ordered (Fig. 1). We sought to distinguish mutations in the Snf2 domain of *ATRX* that primarily affect stability and those that primarily affect function. Four different mutation types (splicing, deletion, nonsense and missense) were analysed in 27 different individuals with ATR-X syndrome: (Tables S1, S2 and Figures S1, S2) (10). Using EBV transformed cell lines derived from these individuals we measured *ATRX* mRNA expression and ATRX protein levels reasoning that functional mutations would have nearer normal levels of protein.

All the splicing mutations introduce premature stop codons by shifting the reading frame. They show reduced levels of ATRX mRNA relative to normal lymphoblastoid cells (Fig. 1b), presumably because the mRNA is subject to nonsense-mediated decay. Consequently, they all have much reduced levels of ATRX protein (Fig. 1c and Fig. S2). The western blot data indicate this is full-length ATRX protein (Fig. S2) and not the predicted premature termination product, showing that a proportion of the RNA is correctly spliced. By contrast, the predicted nonsense mutation (6003G>A; W2001X) has normal levels of mRNA. This mutation generates a cryptic splice site removing nucleotides 5957 to 6022 including the nonsense mutation, and thus behaves as a deletion (p.1986_2007del) ensuring nearly full-length ATRX protein is produced (Fig. S2 and S3).

The deletion mutations and many missense mutations have normal levels of ATRX mRNA but reduced levels of protein (7-27% of normal levels, Figs. 1c, S2 and Table

S1); these mutations probably destabilise the protein. However, three missense mutations, V1552F, K1650N and L1746S accumulate greater amounts of ATRX protein (39-51% of normal levels) (Table S1). It therefore seemed unlikely that they were solely protein destabilising mutations, particularly as we have analysed a normal cell line in which ATRX is as low as 64% of the average normal levels (Fig. 1c). It seemed more likely that these mutations influence ATRX activity.

Homology modelling of ATRX and its mutations

It is known that most protein molecules are only marginally stable and mutations that cause only minor changes in protein stability can cause large changes in the ratio of folded to unfolded protein (11). By placing the mutations in a structural context we sought to understand whether these mutations were of a type that could destabilise ATRX and conversely why our suspected functional mutations were less destabilising. Two structures have been solved for Snf2 family proteins; both for Rad54. One is from zebrafish (12) and the other from archaebacteria (*Sulfolobus solfataricus,* (13)). These structures show that the Snf2 ATPase domain closely resembles that of other helicase family members, and is formed from two RecA-like domains. Given the high degree of structural conservation not only between the Rad54 proteins but also within the helicase superfamily generally, we considered a model of the Snf2 domain based on the zebrafish Rad54 structure (with which ATRX has 29% sequence identity) could be highly informative.

We therefore generated homology models for the Snf2 domain of ATRX using Phyre (Fig. 2, (14)) and HHpred (data not shown, (15)). The resulting models are highly similar with 90% of the residues superposable within 1.2 Angstroms. Two patches of clustered mutations were observed. Cluster I contains mutations H1609R, C1614R, A1622T, L1623S and I1680T while cluster II contains Y1833C and Y2163C. These clusters have been previously noted and it was suggested that cluster II may lie within a domain involved with ATP driven conformational change (13). As all these mutations appear to have significant destabilising affects on the ATRX protein, the clusters may represent no more than structurally sensitive regions. However, we cannot rule out these mutations also exert some functional impact.

Looking at individual mutations, it appears that destabilisation is mainly caused by either introducing a polar residue into an otherwise hydrophobic environment (A1622T, L1623S, I1680T, Y2084H; Table S3) or disrupting an aromatic interaction (Y1833C, Y1847C, Y2163C; Table S3). Of the three mutations that are significantly less destabilising, V1552F falls outside the modelled structure. K1650N is a surface exposed residue neighbouring motif Ib and L1746S is found very close to the ATPase active site within conserved helicase motif III (Fig. 2). To determine whether these residues might alter the function of ATRX rather than simply altering stability we expressed recombinant and mutated ATRX protein.

Analysis of DNA and nucleosome stimulated ATPase activity

Wild type and mutant ATRX proteins were expressed in Sf9 insect cells using baculovirus and purified using a C-terminal HA-tag. The recombinant proteins (rATRX) were mostly full-length (~300kD) but some smaller molecular weight products were noted (Fig. S4a). These are derived from ATRX, most probably C-terminal degradation products as they were identified in western blots (Fig. S4b) using an ATRX C-terminal specific antibody (H300) but not an ATRX N-terminal antibody (fxnp5). As a negative control for all experiments, we generated a mutated sequence encoding a substitution of the highly conserved Lys1600 within helicase motif I by Arg which is a commonly used means to abolish ATPase activity in SWI/SNF proteins.

It has previously been shown that different Snf2 family members show distinctive enzymatic characteristics. For example, in the Snf2 subfamily ATPase activity is equally stimulated by DNA and nucleosomes (16, 17); in contrast, the Mi-2 subfamily ATPase is maximally stimulated by nucleosomes (17, 18). The ISWI subfamily is also maximally stimulated by nucleosomes but has the additional requirement that the H4 tail should be intact (16, 17, 19, 20). ATRX is a member of the Rad54 subfamily. Consistent with the finding that Rad54 and immunopurified ATRX/Daxx complex are maximally stimulated by DNA rather than nucleosomes (21)(22), when we analysed the ATPase activity of wild-type rATRX we found the ATPase activity was stimulated approximately 9 fold in the presence of DNA and only 4 fold by nucleosomes (Fig. 3a). Little activity was seen in the absence of effector. This is in contrast to the findings of Tang et al. (23) who found rATRX was stimulated two fold by nucleosomes but not at all by naked DNA. The ATPase mutant K1600R showed

very little ATPase activity indicating that ATP hydrolysis in this preparation is solely due to wild-type ATRX and not contaminating proteins.

Having shown the ATPase activity of rATRX is maximally stimulated by DNA, we next tested each of the mutant proteins that least perturb ATRX expression but were predicted to affect function. V1552F and K1650N behave in a similar manner to the wild-type protein. The ATPase activity of both of these mutants was stimulated by DNA but to a lesser degree than the wild-type enzyme (Fig. 3b). This might be due to a direct affect on ATPase activity or an indirect affect via binding of DNA.

Unexpectedly, we found that L1746S showed significant ATPase activity in the absence of DNA, producing more than 5x greater than normal ATP hydrolysis after 5 minutes incubation with ATP (Fig. 3b). When DNA is added to the assay, L1746S shows a similar level of DNA stimulation as wild-type protein. It appears that this mutation has partially uncoupled ATPase activity from DNA binding, an intriguing finding given our modelled structure which places this mutation close to the ATPase binding site (Fig. 2).

Analysis of DNA translocation

The Snf2 family have the ability to translocate along DNA. This has been shown both by single molecule studies (24, 25) and by a triple helix displacement assay in which a third DNA strand wrapped around duplex DNA is displaced by the DNA

translocating enzyme (13, 21, 26, 27). The triple helix displacement assay was previously used to show translocation by the ATRX/Daxx complex (22). As the L1746S mutation partially uncouples DNA stimulated ATPase activity, we investigated the mutant enzyme's ability to do useful work by testing if it could remove a barrier to DNA translocation on double stranded DNA (triple helix assay).

The time course of third strand displacement was followed over one hour and we found that each of the mutant enzymes removed less of the third strand than wild-type ATRX (Fig. 3c). These results show that the reduced ATPase activity of V1552F and K1650N proteins no doubt contributes to their weakened displacement activity. Importantly despite its normal level of DNA stimulated activity, the L1746S mutation has reduced translocation consistent with the conclusion that the ATPase activity is uncoupled from its ability to perform useful work.

Discussion

ATRX is a large protein (2492 residues) in which almost all non-truncating mutations associated with ATR-X syndrome fall within 97 conserved residues of the ADD domain or 733 conserved residues of the Snf2 domain. Such pathogenic mutations are not seen in the remaining poorly conserved, structurally disordered 1662 residues of ATRX. It seems likely that mutations in these regions are not observed in ATR-X syndrome because they act as neutral polymorphisms rather than lethal mutations. It is striking that the ADD and Snf2 domains are separated by over 1300

residues which are structurally disordered and this separation is conserved in chicken and frog (Fig. S5). This arrangement of a long, potentially flexible, linker between these domains may be an important aspect of the molecular structure of the ATRX protein.

This study and a previous analysis of the ADD domain (28) have characterised 68 disease-causing mutations in ATRX. In all cases some ATRX protein was expressed. No expression of ATRX would probably cause early embryonic lethality as observed in a mouse ATRX knockout (29). It also seems likely that dosage of ATRX is critical to the development of ATR-X syndrome since the level of ATRX protein is significantly reduced (3-55%) in all disease cases (this report and (28)). Even the splicing mutations produce full-length protein and in three of these mutants (4317G>A (2 cases) and 6218-12574G>A) the full-length protein would be structurally normal. As we can detect no mutant protein in any of these cases, it is unlikely they act via a dominant negative effect and low levels of ATRX alone (between 3 – 10% of normal levels) account for syndrome development.

It appears therefore that, in most cases, a significant deficiency of ATRX protein results in ATR-X syndrome underlying the importance of protein instability in mongenic disease (30). Placing the mutants in a structural model has given insights as to why they destabilise the protein, knowledge that will prove useful in analysing newly identified ATR-X syndrome mutations.

Three patient mutations, V1552F, K1650N and L1746S are expressed at relatively normal levels (39-53%). Although it is not clear whether this is due to comparatively normal protein stability or through another mechanism affecting protein turnover, it was of interest to determine whether these mutations might primarily affect ATRX function rather than via low dosage. Like other members of the Rad54 subfamily, we found that the ATPase activity of rATRX is maximally stimulated by DNA rather than nucleosomes. Recombinant V1552F and K1650N proteins were defective in DNA stimulated ATPase activity. This might be a consequence of a defect in ATP binding/hydrolysis but given that these mutations are found away from the catalytic core we think this unlikely. Instead, these mutations might affect DNA binding or have a role in communicating the occurrence of DNA binding to the catalytic site. Consistent with this, K1650 is found in a region of the protein that in the crystal structures of the SF2 helicases VASA and *Sulfolobus solfataricus* Rad54 makes contact with the DNA backbone (13, 31). Attempts to confirm this hypothesis with the mutant proteins using standard electrophoretic mobility shift assay conditions has however been hampered by the protein precipitating in the wells of the gel.

In contrast, L1746S showed normal DNA stimulated ATP hydrolysis and remarkably generated a high level of ATPase activity in the absence of DNA. This uncoupling of ATPase activity from DNA stimulation has not, so far as we are aware, been previously observed in a human mutation.

One measureable activity of Snf2 proteins is their ablility to translocate across DNA and this study confirms that rATRX translocates across DNA in an ATP dependent

manner (22). Snf2 proteins may convert this movement into various useful applications in a wide variety of nuclear functions, most often to remodel or remove nucleosomes. Recently an ATRX-DAXX complex has been shown to assemble H3.3 containing nucleosomes, an activity that may be used to deposit H3.3 at telomeric repeats (4-6).

All three functional ATRX mutations identified in this study are defective in DNA translocation. Given their severe phenotypic consequences, they provide evidence that DNA translocation is required for *in vivo* function. The V1552F mutation shows both impaired ATPase and DNA translocase activities. This is an interesting mutation as it lies upstream of the most highly conserved sequences and potentially represents a region not previously recognised as functionally important.

The L1746S mutation identified here addresses a key question associated with this group of proteins; how is the activity of ATP hydrolysis converted, via intramolecular interactions, into useful activity? L1746S lies within conserved helicase motif III (Fig. S6) and both structural and mutagenesis studies suggest that this motif makes intra-molecular interactions with motifs II and VI as part of an extended series of contacts around the ATPase site that also includes motifs Ia and V (31, 32). In so doing it has been suggested that motif III co-ordinates nucleic acid binding with ATP hydrolysis (33). The L1746S mutation has uncoupled that co-ordination allowing ATP hydrolysis to proceed in the absence of DNA. How the mutation disrupts the co-ordinating role of motif III is unclear but it identifies an area of motif III that has not previously been targeted for mutagenesis as being important for activity.

The mutation replaces a hydrophobic residue with a polar OH group that has the potential to interact with residues in both motif I and motif II, 4-5Å distant. One possible explanation for the behaviour of this mutation is that motif III not only communicates the state of ATPase activity but also DNA binding with the L1746S mutation perhaps switching motif III into a permanently "on" state signalling that DNA is bound even in its absence, causing the continued futile hydrolysis of ATP. A connection with nucleic acid binding and motif III has recently been shown in the DEAD-box RNA helicase Ded1 (33). It was suggested that binding of ATP promotes the creation of a high affinity nucleic acid binding site mediated by motif III and binding of nucleic acid in turn promotes ATP hydrolysis. It is possible that the L1746S mutation mimics this high affinity nucleic acid bound state.

This analysis has identified a rare and informative mutation (L1746S) in which ATP hydrolysis is not only uncoupled from DNA translocation but also from DNA binding generating a molecular motor that continues to turn over ATP but is defective in driving along DNA.

Materials and Methods

Patients

The study was approved by the Multi-centre Research Ethics Committee (ref MREC 07/MRE00/70). All the individuals studied had ATR-X syndrome with severe to profound mental retardation as defined in the ICD-10 classification (34). The presence of genital abnormalities and alpha thalassaemia are shown in Table S2.

Materials

Nucleosomes 243 and DNA 243 were kindly provided by Lynda Chapman and Daniela Rhodes (MRC Laboratory of Molecular Biology, Cambridge). The histone octamer was positioned on the Widom 601 nucleosome positioning sequence (35). Nucleosome 243 had 48 nucleotides extending beyond either end of the nucleosome core.

Preparation and quantification of ATRX mRNA and protein

This was done as previously described (28). ATRX was detected using 39F as the primary antibody(36). In every western blot the same normal extract was used for comparison. Levels of full length ATRX are expressed relative to the average for normal individuals. Protein loading was checked and corrected by re-probing with antibody to Sin3A (Abcam, Cambridge, UK).

ATRX expression and purification

ATRX was over-expressed and purified from insect cells using the Bac-to-Bac method (Invitrogen, Paisley, UK) as follows. The full length ATRX expression vector was a gift from David Picketts (23). *ATRX* was subsequently cloned into pFastBacHTB (Invitrogen, Paisley, UK) with a 3′ HA tag sequence and transformed into DH10Bac to generate recombinant bacmids. Successful transposition was confirmed by PCR. Infectious viral particles were generated by transfection of the bacmid into Sf9 cells and virus was further amplified to generate high titre virus. For ATRX protein expression, 2L of insect cells at 1 x 10^6 cells per ml were infected with virus and cells harvested after 48 hours. After washing in PBS, cells were resuspended in Hepes lysis buffer (20mM HEPES pH7.5, 0.5M NaCl, 10mM KCl,

2mM $MgCl_2$, 20% glycerol, 0.1% NP-40, 0.2mM EDTA and 0.2mM DTT together with protease inhibitors). Cells were lysed with a Dounce homogeniser, clarified by centrifugation before incubation with Ezview red anti-HA affinity gel (Sigma, St. Louis, MO). The gel was washed four times with 20mM HEPES pH7.5, 0.9M NaCl, 10mM KCl, 2mM $MgCl_2$, 20% glycerol, 0.01% NP-40, 0.1% Tween, 0.2mM EDTA and 0.2mM DTT together with protease inhibitors. ATRX was eluted with HA peptide at 100µg/ml in Hepes lysis buffer, frozen in aliquots on dry ice and stored at -70^{O}C.

ATRX mutagenesis

Mutations were introduced into *ATRX* by site directed mutagenesis using the QuikChange site directed mutagenesis kit (Stratagene, La Jolla, CA). Primer sequences are in the supporting information. Expression and purification of mutant proteins was carried out as above.

ATPase assay

Reactions were performed in a buffer containing 25mM HEPES pH7.5, 5% glycerol, 10mM KCl, 5mM $MgCl_2$, 50µM ATP, 0.005% NP-40, 0.2mM DTT, 100µg/ml BSA and 2.5µCi [γ-^{32}P]-ATP at 3000Ci/mmol. The concentration of added ATRX, nucleic acid or nucleosome is indicated in the relevant figure. The reaction was incubated at 30^{O}C for 30 minutes and stopped by spotting 1µl onto polyethyleneimine-cellulose plates. Spots were allowed to dry and products separated by thin layer chromatography (TLC) in 0.75M KH_2PO_4 pH3.5. After drying, the TLC plate was exposed to a phosphor screen, developed on Typhoon (GE Healthcare, Amersham, UK) and spot intensity quantified using Image Quant TL v2005 (GE Healthcare, Amersham, UK). Background intensity was deducted from the lane profiles using the

rolling ball method. The results are expressed as a percentage of hydrolysed [γ-^{32}P]-ATP of total $^{32}P_i$ (ATP plus P_i) added to the reaction. Intrinsic hydrolysed ATP within the [γ-^{32}P]-ATP reactant was determined by carrying out a control reaction in the absence of enzyme and deducting the percentage of intrinsically hydrolysed [γ-^{32}P]-ATP (typically 2-5%) from the percentage of hydrolysed ATP in the presence of enzyme. Results are expressed as the average of at least two independent experiments and error bars represent one standard deviation.

Homology modelling

Structure prediction was performed using the PHYRE program (http://www.sbg.bio.ic.ac.uk/~phyre/). Phyre is a fully-automated protein modelling resource. It uses a sensitive profile-profile alignment algorithm, comparing both sequence and secondary structure profiles, to reliably detect homologues with mutual sequence identity down to 10-15%.

Models were built for the wild-type structure and each of the previously described missense mutations with the exception of V1538G and V1552F that lie outside the region of homology. The modelled region stretched from residues 1560 to 2259. Because of a lack of sequence homology the following residues were omitted from the structure: 1577-1580, 1688-1702, 1879-1881, 1910-2005, 2057-2078, 2106, 2249-2253. Structures were viewed using VMD 1.8.3 (37).

Triple helix displacement assay

Triple helix displacement experiments were performed based on previously published methods (38-40). The triplex forming oligo (TFO)

(AAAGAAGAAAGAGAAAAGAA) was end labelled with ^{33}P and annealed to a 2-fold excess of EcoRI cut pSV20 in 25mM MES pH 5.5, 10mM $MgCl_2$ by heating to 60^{o}C for 15 minutes and slowly cooling. Displacement reactions were performed at 30^{o}C using 1-2nM triple helix, 100nM protein in a buffer of 50mM Hepes pH7.25, 50mM NaCl, 10mM $MgCl_2$, 10mM KCl, 5% glycerol, 100µg/ml BSA, 1mM DTT. The reaction was followed at 15 minute intervals and stopped using 40mM tris acetate pH7, 0.1% SDS, 3% glycerol, 10mM $MgCl_2$, 0.8mM bromophenol blue. Products were resolved on a 6% native polyacrylamide gel at 4^{o}C, the gel dried and exposed to phosphor screen, developed on Typhoon (GE Healthcare, Amersham, UK) and band intensity quantified using Image Quant TL v2005 (GE Healthcare, Amersham, UK). Background intensity was deducted from the lane profiles using the rolling ball method. The percentage of displaced TFO = 100 X TFO/(TFO+Triple helix). Results are expressed as the average of at least three independent experiments and error bars represent one standard deviation.

Acknowledgements

This work was supported by the MRC and by the National Institute for Health Research Biomedical Research Centre Programme. We thank Sue Butler for the growth of cell lines.

Conflict of Interest Statement. None declared.

References

1. Clapier, C.R. and Cairns, B.R. (2009) The biology of chromatin remodeling complexes. *Annu Rev Biochem,* **78,** 273-304.
2. Singleton, M.R., Dillingham, M.S. and Wigley, D.B. (2007) Structure and mechanism of helicases and nucleic acid translocases. *Annu Rev Biochem,* **76,** 23-50.
3. Flaus, A., Martin, D.M., Barton, G.J. and Owen-Hughes, T. (2006) Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. *Nucleic Acids Res,* **34,** 2887-905.
4. Lewis, P.W., Elsaesser, S.J., Noh, K.M., Stadler, S.C. and Allis, C.D. (2010) Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. *Proc Natl Acad Sci U S A,* **107,** 14075-80.
5. Drane, P., Ouararhni, K., Depaux, A., Shuaib, M. and Hamiche, A. (2010) The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3.3. *Genes Dev,* **24,** 1253-65.
6. Goldberg, A.D., Banaszynski, L.A., Noh, K.M., Lewis, P.W., Elsaesser, S.J., Stadler, S., Dewell, S., Law, M., Guo, X., Li, X. *et al.* (2010) Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell,* **140,** 678-91.
7. Wong, L.H., McGhie, J.D., Sim, M., Anderson, M.A., Ahn, S., Hannan, R.D., George, A.J., Morgan, K.A., Mann, J.R. and Choo, K.H. (2010) ATRX interacts with H3.3 in maintaining telomere structural integrity in pluripotent embryonic stem cells. *Genome Res,* **20,** 351-60.

8. Law, M.J., Lower, K.M., Voon, H.P., Hughes, J.R., Garrick, D., Viprakasit, V., Mitson, M., De Gobbi, M., Marra, M., Morris, A. *et al.* (2010) ATR-X Syndrome Protein Targets Tandem Repeats and Influences Allele-Specific Expression in a Size-Dependent Manner. *Cell,* **143,** 367-78.

9. Gibbons, R.J., Picketts, D.J., Villard, L. and Higgs, D.R. (1995) Mutations in a putative global transcriptional regulator cause X-linked mental retardation with a-thalassemia (ATR-X syndrome). *Cell,* **80,** 837-845.

10. Gibbons, R.J., Wada, T., Fisher, C., Malik, N., Mitson, M., Steensma, D., Goudie, D., Fryer, A., Krantz, I. and Traeger-Synodinos, J. (2008) Mutations in the chromatin associated protein ATRX. *Human Mutation***,** in press.

11. DePristo, M.A., Weinreich, D.M. and Hartl, D.L. (2005) Missense meanderings in sequence space: a biophysical view of protein evolution. *Nat Rev Genet,* **6,** 678-87.

12. Thoma, N.H., Czyzewski, B.K., Alexeev, A.A., Mazin, A.V., Kowalczykowski, S.C. and Pavletich, N.P. (2005) Structure of the SWI2/SNF2 chromatin-remodeling domain of eukaryotic Rad54. *Nat Struct Mol Biol,* **12,** 350-6.

13. Durr, H., Korner, C., Muller, M., Hickmann, V. and Hopfner, K.P. (2005) X-ray structures of the Sulfolobus solfataricus SWI2/SNF2 ATPase core and its complex with DNA. *Cell,* **121,** 363-73.

14. Kelley, L.A. and Sternberg, M.J. (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc,* **4,** 363-71.

15. Hildebrand, A., Remmert, M., Biegert, A. and Soding, J. (2009) Fast and accurate automatic structure prediction with HHpred. *Proteins,* **77 Suppl 9,** 128-32.

16. Aalfs, J.D., Narlikar, G.J. and Kingston, R.E. (2001) Functional differences between the human ATP-dependent nucleosome remodeling proteins BRG1 and SNF2H. *J Biol Chem,* **276,** 34270-8.

17. Boyer, L.A., Logie, C., Bonte, E., Becker, P.B., Wade, P.A., Wolffe, A.P., Wu, C., Imbalzano, A.N. and Peterson, C.L. (2000) Functional delineation of three groups of the ATP-dependent family of chromatin remodeling enzymes. *J Biol Chem,* **275,** 18864-70.

18. Brehm, A., Langst, G., Kehle, J., Clapier, C.R., Imhof, A., Eberharter, A., Muller, J. and Becker, P.B. (2000) dMi-2 and ISWI chromatin remodelling factors have distinct nucleosome binding and mobilization properties. *EMBO J,* **19,** 4332-41.

19. Clapier, C.R., Langst, G., Corona, D.F., Becker, P.B. and Nightingale, K.P. (2001) Critical role for the histone H4 N terminus in nucleosome remodeling by ISWI. *Mol Cell Biol,* **21,** 875-83.

20. Hamiche, A., Kang, J.G., Dennis, C., Xiao, H. and Wu, C. (2001) Histone tails modulate nucleosome mobility and regulate ATP-dependent nucleosome sliding by NURF. *Proc Natl Acad Sci U S A,* **98,** 14316-21.

21. Jaskelioff, M., Van Komen, S., Krebs, J.E., Sung, P. and Peterson, C.L. (2003) Rad54p is a chromatin remodeling enzyme required for heteroduplex DNA joint formation with chromatin. *J Biol Chem,* **278,** 9212-8.

22. Xue, Y., Gibbons, R., Yan, Z., Yang, D., McDowell, T.L., Sechi, S., Qin, J., Zhou, S., Higgs, D. and Wang, W. (2003) The ATRX syndrome protein forms a chromatin-remodeling complex with Daxx and localizes in promyelocytic leukemia nuclear bodies. *Proc Natl Acad Sci U S A,* **100,** 10635-40.

23. Tang, J., Wu, S., Liu, H., Stratt, R., Barak, O.G., Shiekhattar, R., Picketts, D.J. and Yang, X. (2004) A novel transcription regulatory complex containing death domain-associated protein and the ATR-X syndrome protein. *J Biol Chem,* **279,** 20369-77.

24. Amitani, I., Baskin, R.J. and Kowalczykowski, S.C. (2006) Visualization of Rad54, a chromatin remodeling protein, translocating on single DNA molecules. *Mol Cell,* **23,** 143-8.

25. Lia, G., Praly, E., Ferreira, H., Stockdale, C., Tse-Dinh, Y.C., Dunlap, D., Croquette, V., Bensimon, D. and Owen-Hughes, T. (2006) Direct observation of DNA distortion by the RSC complex. *Mol Cell,* **21,** 417-25.

26. Saha, A., Wittmeyer, J. and Cairns, B.R. (2002) Chromatin remodeling by RSC involves ATP-dependent DNA translocation. *Genes Dev,* **16,** 2120-34.

27. Whitehouse, I., Stockdale, C., Flaus, A., Szczelkun, M.D. and Owen-Hughes, T. (2003) Evidence for DNA translocation by the ISWI chromatin-remodeling enzyme. *Mol Cell Biol,* **23,** 1935-45.

28. Argentaro, A., Yang, J.C., Chapman, L., Kowalczyk, M.S., Gibbons, R.J., Higgs, D.R., Neuhaus, D. and Rhodes, D. (2007) Structural consequences of disease-causing mutations in the ATRX-DNMT3-DNMT3L (ADD) domain of the chromatin-associated protein ATRX. *Proc Natl Acad Sci U S A,* **104,** 11939-44.

29. Garrick, D., Sharpe, J.A., Arkell, R., Dobbie, L., Smith, A.J., Wood, W.G., Higgs, D.R. and Gibbons, R.J. (2006) Loss of Atrx affects trophoblast development and the pattern of X-inactivation in extraembryonic tissues. *PLoS Genet,* **2,** e58.

30. Yue, P., Li, Z. and Moult, J. (2005) Loss of protein structure stability as a major causative factor in monogenic disease. *J Mol Biol,* **353,** 459-73.

31. Sengoku, T., Nureki, O., Nakamura, A., Kobayashi, S. and Yokoyama, S. (2006) Structural basis for RNA unwinding by the DEAD-box protein Drosophila Vasa. *Cell,* **125,** 287-300.

32. Schwer, B. and Meszaros, T. (2000) RNA helicase dynamics in pre-mRNA splicing. *EMBO J,* **19,** 6582-91.

33. Banroques, J., Doere, M., Dreyfus, M., Linder, P. and Tanner, N.K. (2010) Motif III in superfamily 2 "helicases" helps convert the binding energy of ATP into a high-affinity RNA binding site in the yeast DEAD-box protein Ded1. *J Mol Biol,* **396,** 949-66.

34. ICD-10 (1992) *The ICD-10 Classification of Mental and Behavioural Disorders. Clinical description and diagnostic guidelines*. WHO, Geneva.

35. Lowary, P.T. and Widom, J. (1998) New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J Mol Biol,* **276,** 19-42.

36. McDowell, T.L., Gibbons, R.J., Sutherland, H., O'Rourke, D.M., Bickmore, W.A., Pombo, A., Turley, H., Gatter, K., Picketts, D.J., Buckle, V.J. *et al.* (1999) Localization of a putative transcriptional regulator (ATRX) at pericentromeric heterochromatin and the short arms of acrocentric chromosomes. *Proc Natl Acad Sci U S A,* **96,** 13983-8.

37. Humphrey, W., Dalke, A. and Schulten, K. (1996) VMD: visual molecular dynamics. *J Mol Graph,* **14,** 33-8, 27-8.

38. Durr, H. and Hopfner, K.P. (2006) Structure-function analysis of SWI2/SNF2 enzymes. *Methods Enzymol,* **409,** 375-88.

39. Firman, K. and Szczelkun, M.D. (2000) Measuring motion on DNA by the type I restriction endonuclease EcoR124I using triplex displacement. *EMBO J,* **19,** 2094-102.

40. Sivanathan, V., Allen, M.D., de Bekker, C., Baker, R., Arciszewska, L.K., Freund, S.M., Bycroft, M., Lowe, J. and Sherratt, D.J. (2006) The FtsK gamma domain directs oriented DNA translocation by interacting with KOPS. *Nat Struct Mol Biol,* **13,** 965-72.

41. Yang, Z.R., Thomson, R., McNeil, P. and Esnouf, R.M. (2005) RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics,* **21,** 3369-76.

42. Picketts, D.J., Tastan, A.O., Higgs, D.R. and Gibbons, R.J. (1998) Comparison of the human and murine ATRX gene identifies highly conserved, functionally important domains. *Mammalian Genome,* **9,** 400-403.

Figure Legends

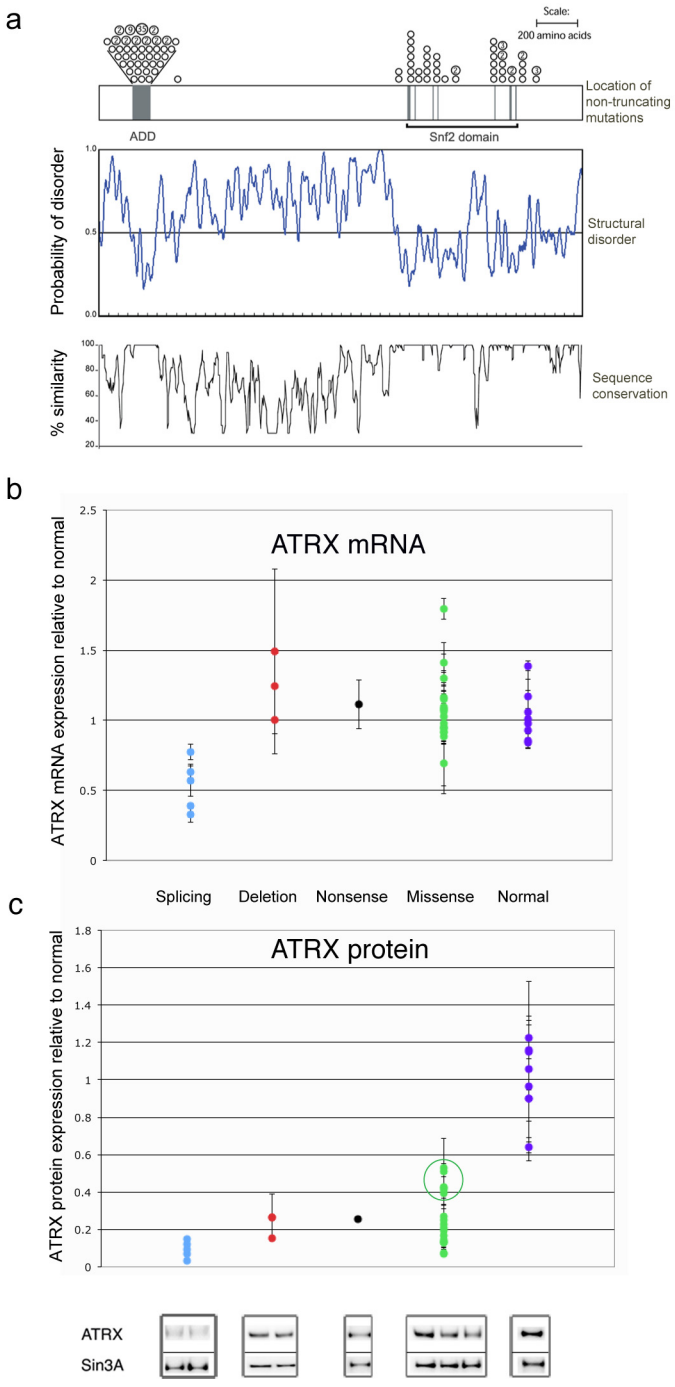
Figure 1 Most non-truncating mutations in ATRX occur in structurally ordered regions and lead to protein instability. (a) Analysis of regions of structural disorder within ATRX and their correlation with non-truncating mutations. The RONN program (41) was used to determine predicted regions of structural disorder (middle figure) and compared with the location of non-truncating mutations (top figure, adapted from

(10)) and the degree of sequence conservation between mouse and human ATRX (bottom figure, adapted from (42)). Regions of structural disorder are both devoid of mutations and show low levels of sequence conservation. (b) Real-time PCR analysis of the levels of ATRX mRNA for patient and normal cell lines (n=8) for the Snf2 domain by mutation type. Error bars represent 1 S.D.**.** ATRX mRNA levels are normal except for the splicing mutations which have reduced levels of mRNA. (c) Levels of ATRX protein in normal (n=8) and patient cells lines as determined by quantitative western blots. Error bars represent 1 S.D.. All patient cell lines have reduced ATRX protein levels. Circled are the results for three pairs of siblings with their respective missense mutations, V1552F, K1650N and L1746S whose relatively higher level of ATRX protein (compared to other patients) may indicate a functional mutation. Representative western blots for each of the mutation types are shown at the bottom of the figure together with the loading control (Sin3A). ATRX mRNA and protein data points are the average of at least four independent experiments from two biological replicates (see also Table S1 for mean values and standard deviations and Figure S2 for representative western blots).

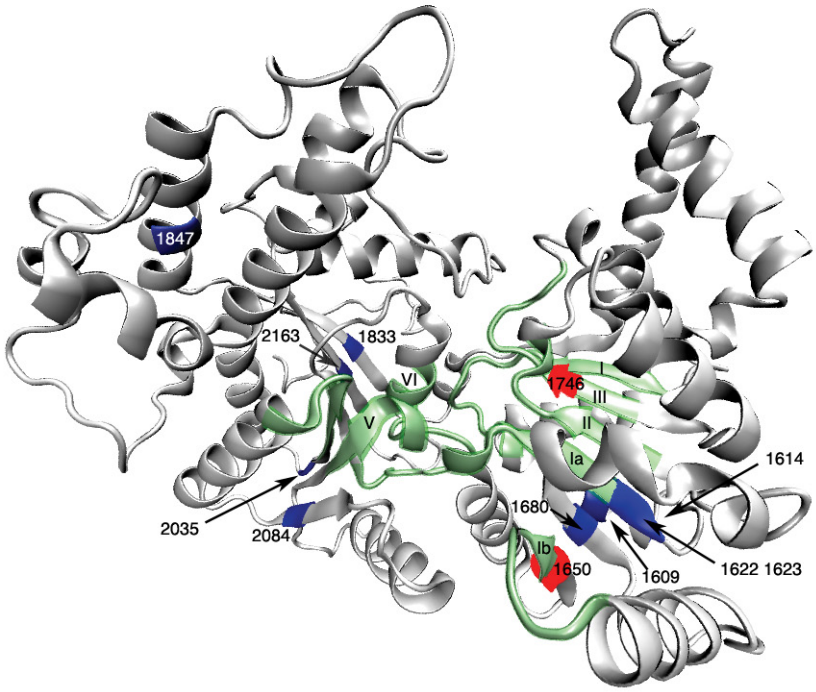
Figure 2 Structure of ATRX. (a) Homology model of the Snf2 domain of ATRX. The location of mutations and highly conserved sequences are shown in the model of ATRX. Mutations that particularly destabilise the ATRX protein are in blue, while in red are the K1650N and L1746S mutations that have a lesser impact on stability. The helicase conserved sequences (I-VI) are in green. (b) Sequence alignment of the Snf2 domains of ATRX and the two Rad54 proteins for which crystal structures have been determined from zebrafish and *Sulfolobus solfataricus* (Sso). Residue

and domain colours as in (a) with the Snf2 conserved domains A-D boxed. Arrowed in domain Ib are residues involved in DNA binding in SsoRad54.

Figure 3 Functional analysis of ATRX. (a) The ATPase activity of ATRX is stimulated more by DNA than nucleosomes. An equal mass of either naked dsDNA 243 or dsDNA wrapped around the histone octamer (nucleosome 243) was added in each case to 10nM of ATRX or the ATPase mutant K1600R. Samples were incubated for 10 minutes at 30^{O}C and the extent of ATP hydrolysis determined. (b) ATPase activity of mutant ATRX proteins. ATPase activity was measured using 50nM of protein with or without an excess of plasmid DNA for the patient mutant proteins V1552F, K1650N and L1746S, wild-type ATRX and as a negative control ATRX K1600R. In the absence of DNA L1746S shows an elevated level of ATPase activity. (c) DNA translocase activity of mutant ATRX proteins. Triple helix displacement assay in which translocation of 100nM ATRX or mutant ATRX along DNA is followed over 60 minutes through the displacement of a triple helix forming oligo (TFO) from dsDNA. DNA translocation is impaired in the patient mutant proteins.

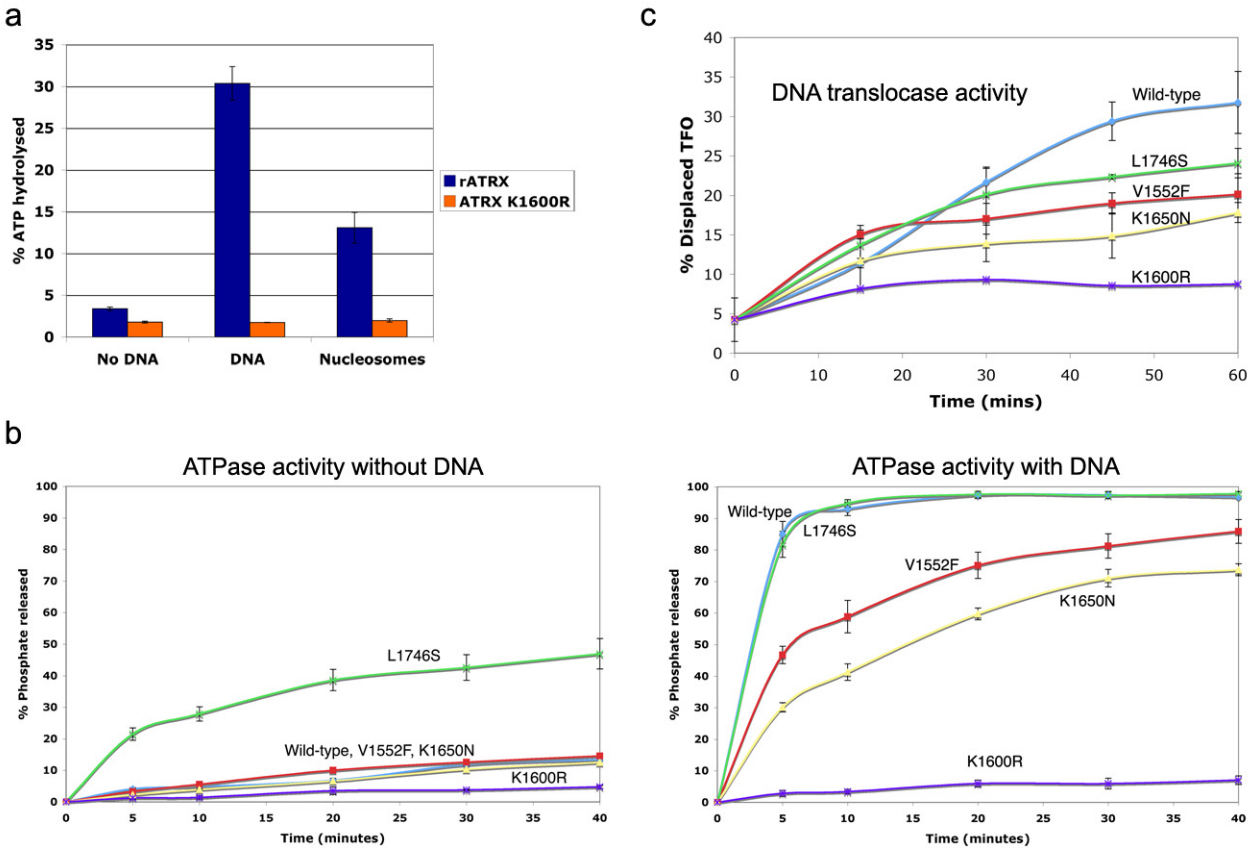

a
Scale:
200 amino acids
Location of non-truncating mutations
ADD
Snf2 domain
Probability of disorder
Structural disorder
% similarity
Sequence conservation
b
ATRX mRNA
ATRX mRNA expression relative to normal
Splicing
Deletion
Nonsense
Missense
Normal
c
ATRX protein
ATRX protein expression relative to normal
ATRX
Sin3A

a


1847
2163
1833
VI
V
1746
I
III
II
Ia
1614
1680
2035
2084
Ib
1650
1609
1622 1623

b


ATRX 1546 KEPLVQVHRNMVIKLKPHQVDGVQFMWDCCCESVKKTKKSPGSGCILAHCMGLGKTLQVVSFLHTVLLCD---KLDFSTALVVCPLNTALNWWNEFEKWQ
Zebrafish Rad54 133 LPVHVVVDPVLSKVLRPHQREGVKFLWDCVTG----RRIENSYGCIMADEMGLGKTLQCITLIWTLLKQSPDCKPEIDKVIVVSPSSLVRNWVNEVGKWL
Sso Rad54 430 KSFQLLEPYNIKANLRPYQIKG--FSWXRFXN-------KLGFGICLADDXGLGKTLQTIAVFSDAKKEN-----ELTPSLVICPLSVLKNWEEELSKFA
I
Ia
ATRX 1643 EGLKDDEKLEVSELATVKRPQERSYMLQRWQEDGGVMIIIGYEMYRNLAQGRNVKSRKLKEIFNKALVDPGPDFVVCDEGHILKNEASAVSKAMNSIRSRR
Zebrafish Rad54 229 GGRVQPVAIDGGSKDEIDSKLVNFISQQGMRIPTPILIISYETFR----------LHAEVLHKGKVG----LVICDEGHRLKNSDNQTYLALNSMNAQR
Sso Rad54 516 PHLRFAVFHEDRSKIKLEDYDIILTTYAVLLRDTRLKEVEWKYIV----------IDEAQNIKNPQTKIFKAVKELKSKY
Ib
A
II
ATRX 1743 RIILTGTPLQNNLIEYHCMVNFIKENLLGSIKEFRNRFINPIQNGQCADSTMVDVRVMKKRAHILYEMLAGCVQRKD--YTALTKFLPPKHEYVLAVRMT
Zebrafish Rad54 314 RVLISGTPIQNDLLEYFSLVHFVNSGILGTAQEFKKRFEIPILKGRDADASDKDRAAGEQKLQELISIVNRCLIRRT--SDILSKYLPVKIEQVVCCNLT
Sso Rad54 586 RIALTGTPIENKVDDLWSIXTFLNPGLLGSYSEFKSKFATPIKKGDNX----------AKEELKAIISPFILRRTKYDKAIINDLPDKIETNVYCNLT
III
B
ATRX 1841 SIQCKLYQYYLDHLTGVGNNSEGGRGKAGAKLFQDFQMLSRIWTHPWCLQLDYISKENKGYFDEDSMDEFIASDSDETSMSLSSDDYTKKKKKGKKGKKD
Zebrafish Rad54 412 PLQKELYKLFLKQAKPVESLQTG---KISVSSLSSITSLKKLCNHPALIYEKCLTGE----------EGFDGALDLFPQNYSTKAVEP------
Sso Rad54 674 PEQAAXYKAEVENLFNNIDSVTG--IKRKGXILSTLLKLKQIVDHPALLKGGEQSVR----------
C
ATRX 1941 SSSSGSGSDNDVEVIKVWNSRSRGGGEGNVDETGNNPSVSLKLEESKATSSSNPSSPAPDWYKDFVTDADAEVLEHSGKMVLLFEILRMAEEIG-DKVLV
Zebrafish Rad54 ----------QLSGKMLVLDYILAMTRTTTSDKVVL
Sso Rad54 ----------RSGKXIRTXEIIEEALDEG-DKIAI
IV
ATRX 2040 FSQSLISLDLIEDFLELASREKTEDKDKPLIYKGEGKWLRNIDYYRLDGSTTAQSRKKVAEEFNDETNVRGRLFIISTKAGSLGINLVAANRVIIFDASW
Zebrafish Rad54 513 VSNYTQTLDLFEKLCRNR----------RYLYVRLDGTMSIKKRAKIVERFNNPSSPE-FIFMLSSKAGGCGLNLIGANRLVMFDPDW
Sso Rad54 753 FTQFVDXGKIIRNIIEKELN----------TEVPFLYGELSKKERDDIISKFQNNPSVK--FIVLSVKAGGFGINLTSANRVIHFDRWW
D
V
ATRX 2140 NPSYDIQSIFRVYRFGQTKPVYVYRFLAQGTMEDKIYDRQVTKQSLSFRVVDQQQ-VERHFTMNELTELY
Zebrafish Rad54 590 NPANDEQAMARVWRDGQKKTCYIYRLLSTGTIEEKILQRQAHKKALSSCVVDEEQDVERHFSLGELRELF
Sso Rad54 830 NPAVEDQATDRVYRIGQTRNVIVHKLISVGTLEEKIDQLLAFKRSLFKDIISSGDSWITELSTEELRKVI
VI


a
% ATP hydrolysed
No DNA
DNA
Nucleosomes
rATRX
ATRX K1600R
c
DNA translocase activity
% Displaced TFO
Time (mins)
Wild-type
L1746S
V1552F
K1650N
K1600R
b
ATPase activity without DNA
% Phosphate released
Time (minutes)
L1746S
Wild-type, V1552F, K1650N
K1600R
ATPase activity with DNA
Wild-type
L1746S
V1552F
K1650N
K1600R

SI Figure legends

Figure S1 Location of non-truncating mutations within the Snf2 domain found in ATR-X syndrome. The position of mutations (in blue) relative to the conserved helicase (I-VI in yellow) and Snf2 motifs (A-D in red) are shown.

Figure S2 Representative western blots of ATRX protein extracted from EBV cell lines derived from ATR-X syndrome patients grouped by different mutation type. All produce full-length ATRX but in reduced amounts. ATRXt is the truncated form of ATRX of molecular weight 180kD which terminates upstream of the Snf2 domain (1). Its expression is therefore unaffected by these mutations. Note that the presumed nonsense mutation was found in fact to give rise to a splicing mutation - see Fig S3.

Figure S3 Splicing out of the premature stop codon in W2001X. (a) PCR on cDNA made from W2001X EBV cell line using primers either side of the stop codon produces a smaller fragment than the expected 869bp. (b) Sequence of W2001X PCR product reveals that the premature stop codon is removed by alternative splicing. Highlighted in yellow is the final sequence of exon 26, which ends with nucleotide 5956. This has been joined with nucleotide 6023 within exon 27 (highlighted in green) indicating that residues 5957 to 6022 have been spliced out.

Figure S4 Expression of recombinant wild-type and mutant ATRX. (a) Purification of rATRX and mutant ATRX. SDS-PAGE stained with colloidal blue of recombinantly expressed wild type ATRX, three patient mutations V1552F, K1650N and L1746S

and the ATPase mutant for use as a negative control, K1600R. The full-length ATRX protein and probable degradation products (*) are indicated. 500ng of protein was loaded on each lane. (b) Western blot of purified rATRX and mutant ATRX. The western blot was sequentially probed with the N-terminal ATRX specific antibody fxnp5, the C-terminal ATRX specific antibody H300 and an antibody to the C-terminal HA tag. These demonstrate the additional starred bands in (a) are most likely C-terminal degradation products of ATRX.

Figure S5 Spacing between the ADD and Snf2 domains in human, chicken and frog ATRX proteins.

Figure S6 Proposed relationship of motif III with motifs I and II from the homology model of L1746S. Helicase conserved motifs I, II and III are labelled. Residue 1746 is predicted to lie at the end of the β-sheet before the loop containing the highly conserved T1747-G1748-T1749 and in a position to make contacts with residues in both motifs I and II that lie within 5Å. For the sake of clarity hydrogen atoms are not shown.

SI Table legends

Table S1 Expression levels of ATRX mRNA and protein for each of the mutations. Where more than one expression level is quoted for a mutation data has been derived from the cell line of a different patient with the same mutation (usually a sibling). Values for 1 standard deviation (S.D.) are indicated.

Table S2 Summary of clinical phenotype associated with the *ATRX* mutations.

Table S3 Summary of structural consequences of mutations in the chromatin remodelling domain of ATRX.

Supporting Information

Materials and Methods

Mutagenic oligos

V1552F

CCAAAGAACCTTTAGTGCAGTTTCATAGAAATATGGTTATC
GATAACCATATTTCTATGAAACTGCACTAAAGGTTCTTTGG

K1650N

GGATTAAAAGATGATGAGAATCTTGAGGTTTCTGAATTAGC
GCTAATTCAGAAACCTCAAGATTCTCATCATCTTTTAATCC

L1746S

GGAGGAGGATTATTTCAACAGGAACACCACTTC
GAAGTGGTGTTCCTGTTGAAATAATCCTCCTCC

K1600R

GTATGGGCCTTGGTAGGACTTTACAGGTGGTAAG
CTTACCACCTGTAAAGTCCTACCAAGGCCCATAC

SI References

1. Garrick D, Samara V, McDowell TL, Smith AJH, Dobbie L, Higgs DR, Gibbons RJ (2004) *Gene* 326:23-34.

Figure S1: Non-truncating mutations

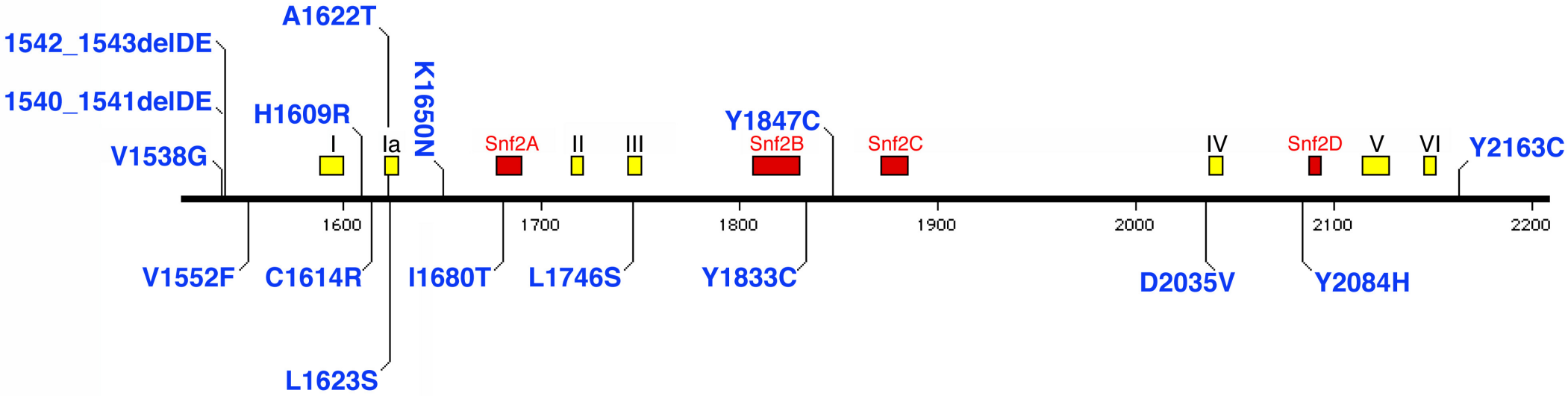
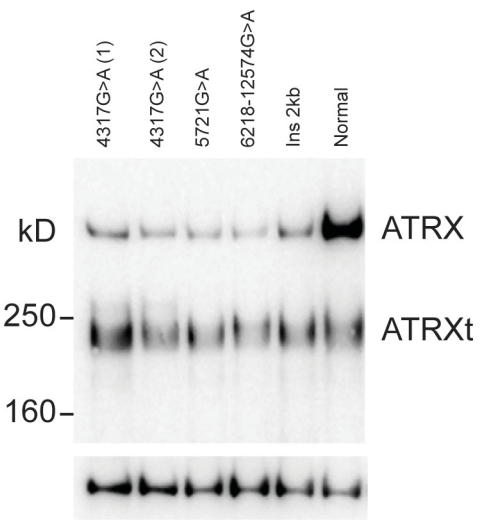
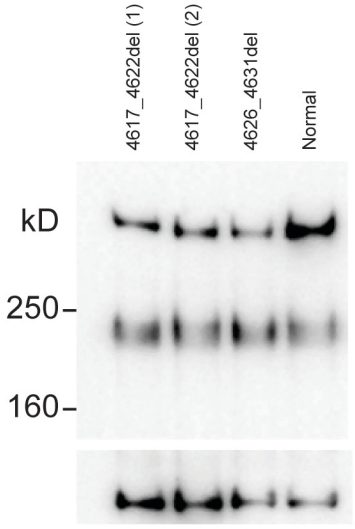


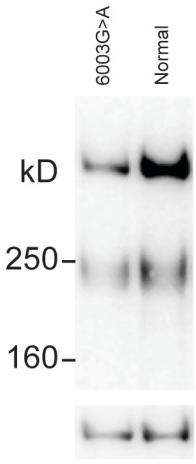
Figure S2: Representative Western Blots



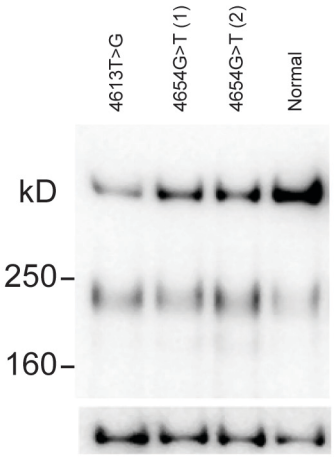
Splicing



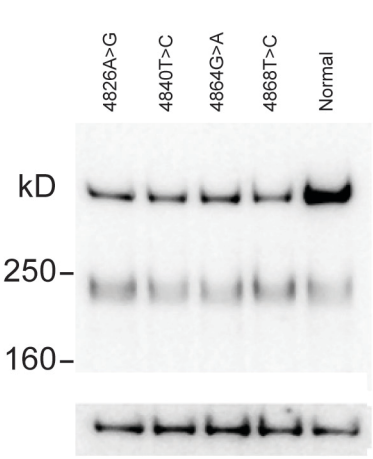
Deletion



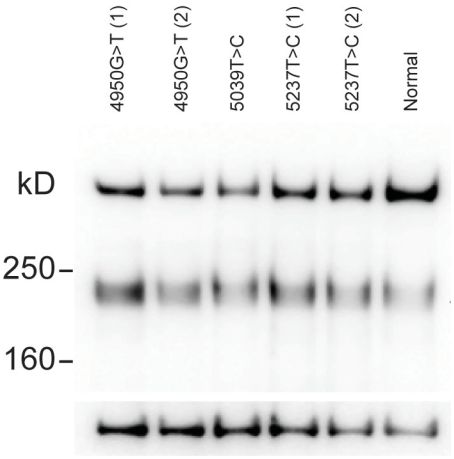
Nonsense



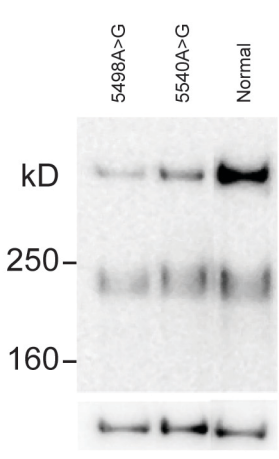
Missense



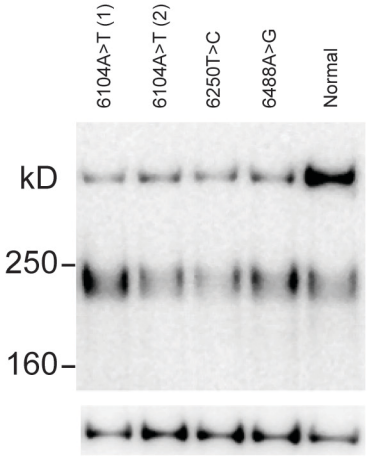
Missense



Missense

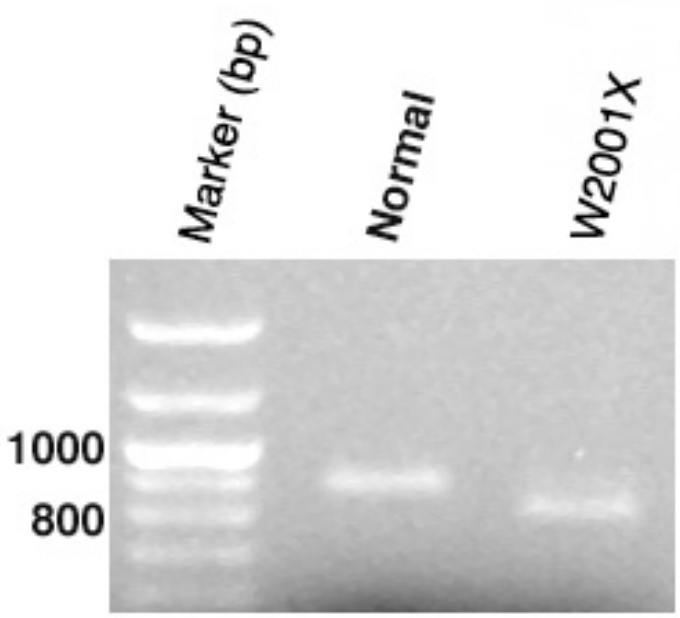


Missense

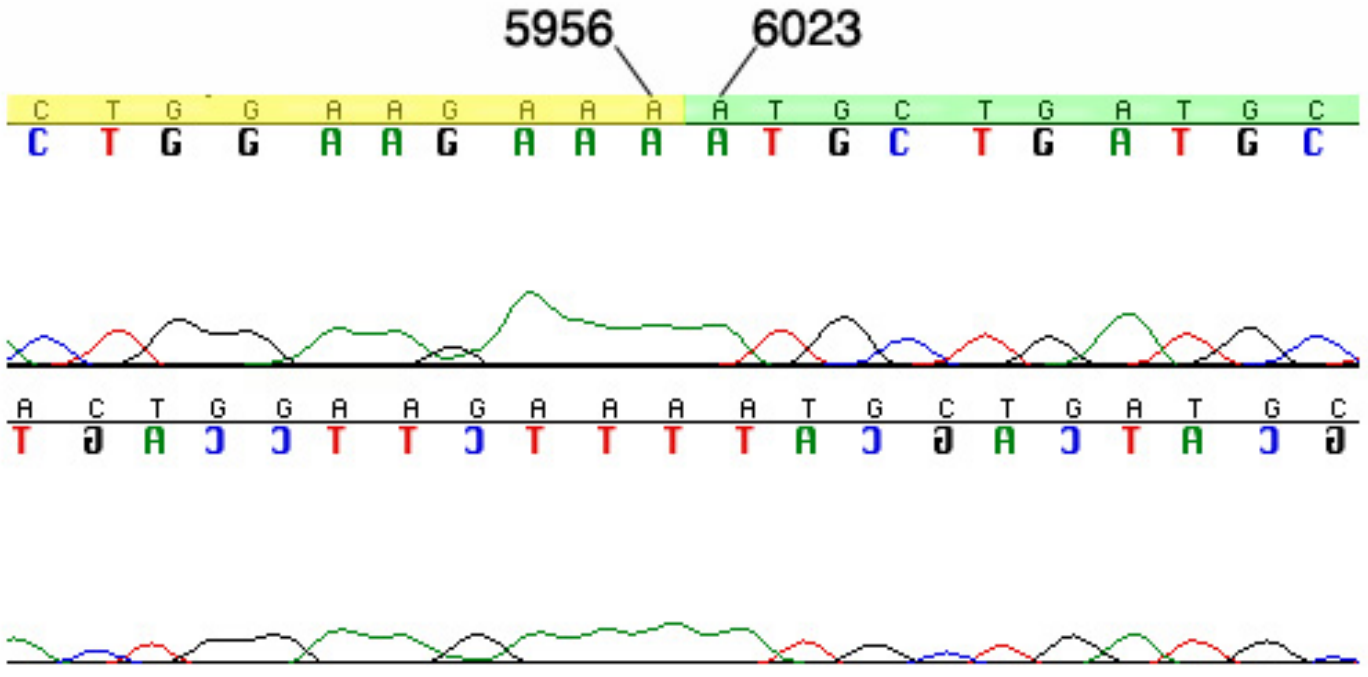


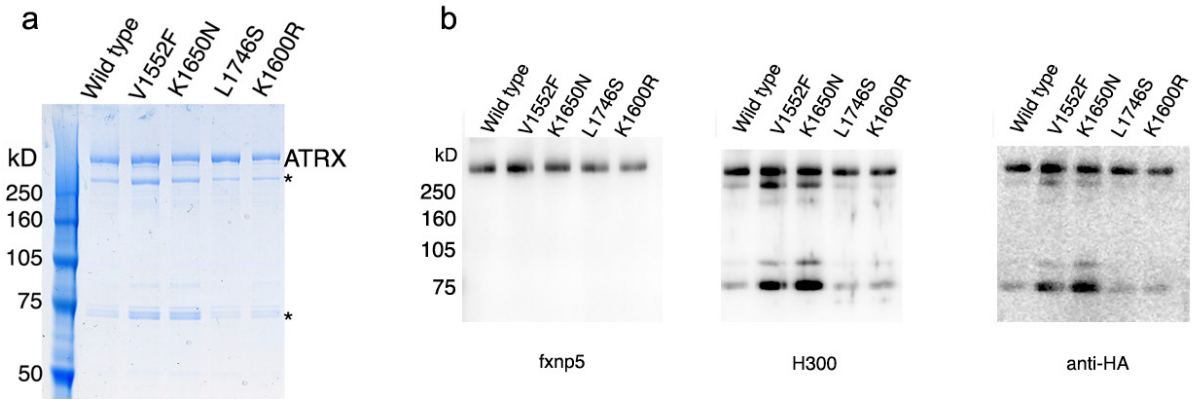
Missense

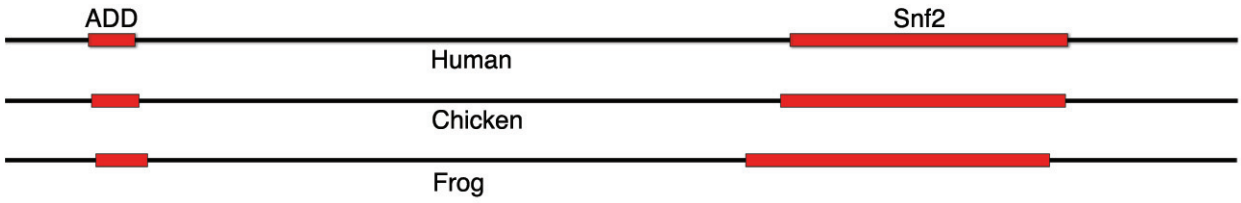
a


Marker (bp)
Normal
W2001X
1000
800

b


5956
6023
C T G G A A G A A A A T G C T G A T G C


a
Wild type
V1552F
K1650N
L1746S
K1600R
kD
250
160
105
75
50
ATRX
*
*
b
Wild type
V1552F
K1650N
L1746S
K1600R
kD
250
160
105
75
fxnp5
H300
anti-HA


ADD
Snf2
Human
Chicken
Frog

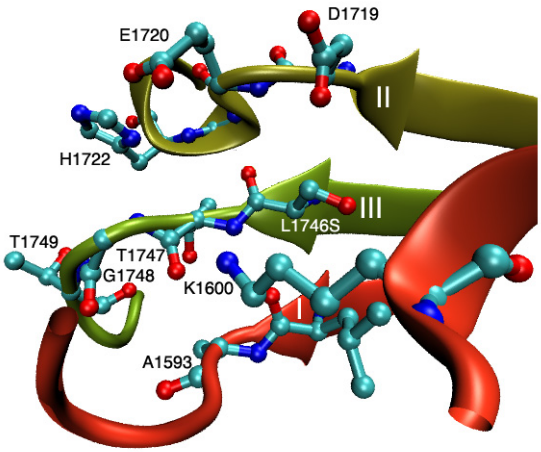

D1719
E1720
II
H1722
III
L1746S
T1749
T1747
G1748
K1600
I
A1593

Table S1

Mutation type	Mutation	mRNA level (relative to normal)	mRNA level 1 S.D.	Protein level (relative to normal)	Protein level 1 S.D.
Splicing	4317G>A; S insertion 53bp (fs)	0.63	0.04	0.10	0.03
		0.78	0.06	0.07	0.01
	5721G>A; S c5698_5786del; G1900fs	0.33	0.06	0.15	0.00
	6218-12574G>A; S insertion 124bp (fs)	0.39	0.02	0.03	0.01
	Ins 2kb from chrom 2	0.57	0.11	0.12	0.03
Deletion	4617_4622del; 1540_1541delDE	1.25	0.25	0.26	0.01
		1.50	0.59	0.27	0.12
	4626_4631del; 1542_1543delDE	1.00	0.24	0.15	0.01
Nonsense	6003G>A; W2001X	1.11	0.17	0.25	0.01
Missense	4613T>G; V1538G	0.70	0.16	0.13	0.03
	4654G>T; V1552F	1.06	0.15	0.41	0.14
		1.80	0.07	0.51	0.18
	4826A>G; H1609R	1.08	0.10	0.25	0.06
	4840T>C; C1614R	1.10	0.03	0.27	0.06
	4864G>A; A1622T	1.30	0.05	0.21	0.02
	4868T>C; L1623S	1.41	0.14	0.20	0.02
	4950G>T; K1650N	1.02	0.18	0.53	0.00
		0.95	0.12	0.43	0.06
	5039T>C; I1680T	0.91	0.01	0.23	0.04
	5237T>C; L1746S	0.98	0.00	0.39	0.02
		1.15	0.32	0.42	0.00
	5498A>G; Y1833C	0.91	0.44	0.07	0.01
	5540A>G; Y1847C	0.88	0.20	0.13	0.01
	6104A>T; D2035V	0.94	0.09	0.17	0.04
		0.95	0.26	0.13	0.03
	6250T>C; Y2084H	1.08	0.10	0.07	0.00
	6488A>G; Y2163C	1.16	0.08	0.15	0.05

Expression levels of ATRX mRNA and protein for each of the mutations. Where more than one expression level is quoted for a mutation, data has been derived from the cell line of a different patient with the same mutation (usually a sibling). Values for 1 standard deviation (S.D.) are indicated.

Table S2 : Summary of clinical phenotype associated with the *ATRX* mutations

Mutation type	Nucleotide change	Amino acid change	Genital abnormality[b]	HbH inc cells %[c]
Splicing	c.4317G>A; r[insertion 53bp]	p.S1440fs	++	0.9
			+++	3.6
	c.5721G>A; r[5698_5786del]	p.G1900fs	+++	+
	c.6218-12574G>A; r[insertion 124bp]	p.G2073fs	++	1.4
	Insertion of 2kb from chromosome 2; splicing abnormality		U	0
Deletion	c.4617_4622del	p.1540_1541delDE	+++	1.0
			N	0.9
	c.4626_4631del	p.1542_1543delDE	++	0.6
Nonsense [a]	c.6003G>A;	W2001X	U	0.06
Missense	c.4613T>G	p.V1538G	+++	11
	c.4654G>T	p.V1552F	+++	56
			+	41
	c.4826A>G	p.H1609R	+++	1.6
	c.4840T>C	p.C1614R	+++++	>5
	c.4864G>A	p.A1622T	++	18.2
	c.4868T>C	p.L1623S	+	1.5
	c.4950G>T	p.K1650N	N	0.4
			N	0.6
	c.5039T>C	p.I1680T	+++++	+
	c.5237T>C	p.L1746S	N	0
			N	3.2
	c.5498A>G	p.Y1833C	++	0.2
	c.5540A>G	p.Y1847C	+	+
	c.6104A>T	p.D2035V	+++	7
			+++	27
	c.6250T>C	p.Y2084H	++	>5
	c.6488A>G	p.Y2163C	+	12

Table adapted from (1). All patients had severe to profound mental retardation as defined in the ICD-10 classification (2).

a Subsequent analysis shows this to be a splicing mutation r[5957_6022del]; p.1986_2007del.

b For a given mutation the range of genital abnormality present is shown: U not determined, N normal, + very mild eg high lying testes, ++ cryptorchidism, +++ hypospadias,, ++++ micropenis, +++++ ambiguous genitalia or male pseudohermaphrodite.

c For a given mutation the proportion of cells with HbH inclusions is shown. Where values from more than one individual are available, a range is shown. + indicates when inclusions were observed but not quantified

References

1. Gibbons, R.J., Wada, T., Fisher, C., Malik, N., Mitson, M., Steensma, D., Goudie, D., Fryer, A., Krantz, I. and Traeger-Synodinos, J. (2008) Mutations in the chromatin associated protein ATRX. *Human Mutation,* **29,** 796-802.
2. ICD-10 (1992) The ICD-10 Classification of Mental and Behavioural Disorders. Clinical description and diagnostic guidelines. WHO, Geneva

Table S3

Mutation	Protein level (% relative to normal)	Solvent accessibility	Consequence
A1622T L1623S I1680T Y2084H	25 20 23 7	55 3 0 43	Polar residue in hydrophobic β-sheet
Y1833C Y1847C Y2163C	7 13 15	0 0 0	Disrupt aromatic interactions
H1609R	25	55	Buried charge
D2035V	17, 13	11	Loss of salt bridge
C1614R	27	95	Unexplained
K1650N L1746S	53, 43 39, 42	82 0	Minor affect

Summary of structural consequences of mutations in the chromatin remodelling domain of ATRX.