BMC
Research Notes

RESEARCH ARTICLE   Open Access

Comparative analysis of the *ATRX* promoter and 5' regulatory region reveals conserved regulatory elements which are linked to roles in neurodevelopment, alpha-globin regulation and testicular function

Paisu Tang[1,2], Stephen Frankenberg[1], Anthony Argentaro[2], Jennifer M Graves[1,3] and Mary Familari[1*]

Abstract

Background: ATRX is a tightly-regulated multifunctional protein with crucial roles in mammalian development. Mutations in the *ATRX* gene cause ATR-X syndrome, an X-linked recessive developmental disorder resulting in severe mental retardation and mild alpha-thalassemia with facial, skeletal and genital abnormalities. Although ubiquitously expressed the clinical features of the syndrome indicate that ATRX is not likely to be a global regulator of gene expression but involved in regulating specific target genes. The regulation of *ATRX* expression is not well understood and this is reflected by the current lack of identified upstream regulators. The availability of genomic data from a range of species and the very highly conserved 5' regulatory regions of the *ATRX* gene has allowed us to investigate putative transcription factor binding sites (TFBSs) in evolutionarily conserved regions of the mammalian *ATRX* promoter.

Results: We identified 12 highly conserved TFBSs of key gene regulators involved in biologically relevant processes such as neural and testis development and alpha-globin regulation.

Conclusions: Our results reveal potentially important regulatory elements in the *ATRX* gene which may lead to the identification of upstream regulators of *ATRX* and aid in the understanding of the molecular mechanisms that underlie ATR-X syndrome.

Background

ATR- X (<u>a</u>lpha <u>t</u>halassemia, mental <u>r</u>etardation, <u>X</u>-linked) syndrome is an X-linked recessive developmental disorder affecting males. Clinical features include severe mental retardation, mild alpha-thalassemia, microcephaly, short stature, and facial, skeletal and genital abnormalities [1-3]. The ATRX protein is large (280 kDa) and contains two highly conserved domains, a PHD-like finger which interacts with chromatin, and a SWI//SNF-like ATPase domain which displays nucleosome remodeling activity, implying that ATRX functions as a chromatin remodeling protein [4-6].

While ATRX is widely expressed throughout development, studies in mice have revealed that ATRX has specific tissue/cell type functions [7-9]. For example, while loss of *ATRX* in chondrocytes has minimal effects on bone growth, mice lacking *ATRX* in the forebrain show apoptosis of cortical neurons leading to a reduction in forebrain size [9,10]. Moreover targeted overexpression of *Atrx* gives rise to phenotypic features that are common to ATR-X syndrome such as severe neural defects and facial dysmorphology [11].Thus, ATRX expression and function must rely on temporal-spatial regulators and/or cofactors for tissue-specific gene regulation and function. While a number of cofactors have been discovered for ATRX including Heterochromatin protein 1 (HP1) [12], Enhancer of zeste (EZH2) [13], Death domain-associated protein (DAXX) [6], methyl-CpG binding protein

* Correspondence: m.familari@unimelb.edu.au
[1]Department of Zoology, University of Melbourne, Victoria 3010, Australia
Full list of author information is available at the end of the article

(MECP2) [14] and Cohesin [15], the upstream regulators of *ATRX* transcription remain unknown and the boundaries of the *ATRX* promoter remain undefined.

In an earlier study focused on the description of the *ATRX* gene and protein sequence, the putative promoter region of human *ATRX* was reported to contain *'multiple CCAAT boxes and binding sites for the CTF family of transcription factors'* [16]. Thereafter, studies have established that the CCAAT motif is not only recognized by the Nuclear Factor 1/CCAAT transcription factor (NF1/CTF) family but also several other transcription factors/complexes which include; NF-E3, GATA1, NF-Y and C/EBP [17]. Thus, relatively little is known about the *ATRX* promoter and the upstream regulatory regions of the gene. The aim of this study was to establish a starting point for experimental studies on the human ATRX promoter by: (i) identifying evolutionarily conserved regions (ECRs) in the mammalian *ATRX* promoter and 5' regulatory regions using phylogenetic footprinting and (ii) identifying putative transcription factor binding sites (TFBSs).

Results

Conservation analysis of the *ATRX* promoter and 5' upstream sequence

The degree of conservation between the regulatory regions of *ATRX* from different mammals, compared with human, were examined across a region spanning -13 kbp to +300 bp, which includes the 215 bp 5'UTR of human *ATRX* sequence [16] since some regulatory elements can lie within the 5'UTR of a gene [17,18]. The molecular timescale of divergence between these species are indicated in Figure 1.

Conservation analysis revealed a clear loss of homology beyond (-500) between human and the most distantly related elephant and tammar wallaby. In addition, three additional dispersed regions of homology were identified by pair-wise alignment of human, mouse, dog, horse, elephant and armadillo regulatory regions. These are located at (-1069 to -693), (-1623 to -1457) and
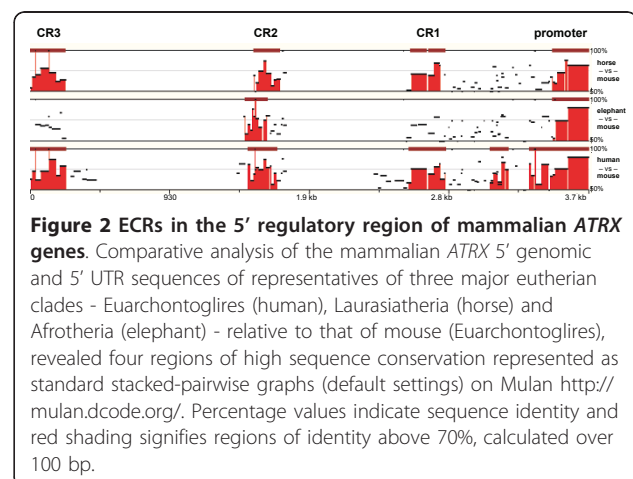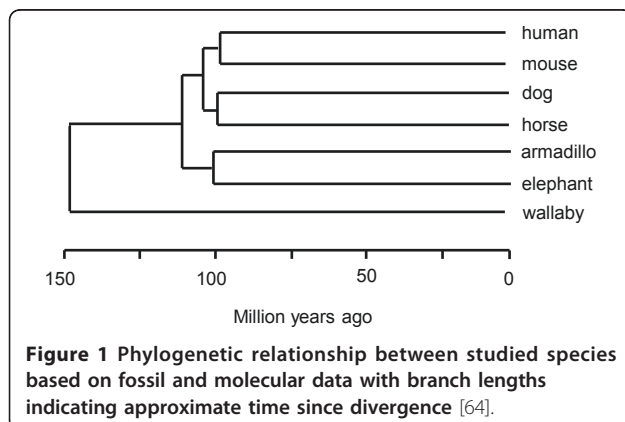
(-12759 to -12669) of the human sequence, and were labelled conserved region 1 (CR1), CR2 and CR3 where CR3 lies most distal to the Transcription Start Site (TSS). Thus, four dispersed regions of high conservation were identified in the upstream regulatory region of the mammalian *ATRX* promoter. These are shown in Figure 2 in a comparison of four distantly-related species representing a broad range of eutherian divergence. Armadillo, representing Xenarthra was not included in Figure 2 and Figure 3 due to the sequence being incomplete. The conserved regions were then examined computationally for conserved, putative transcription factor binding sites.
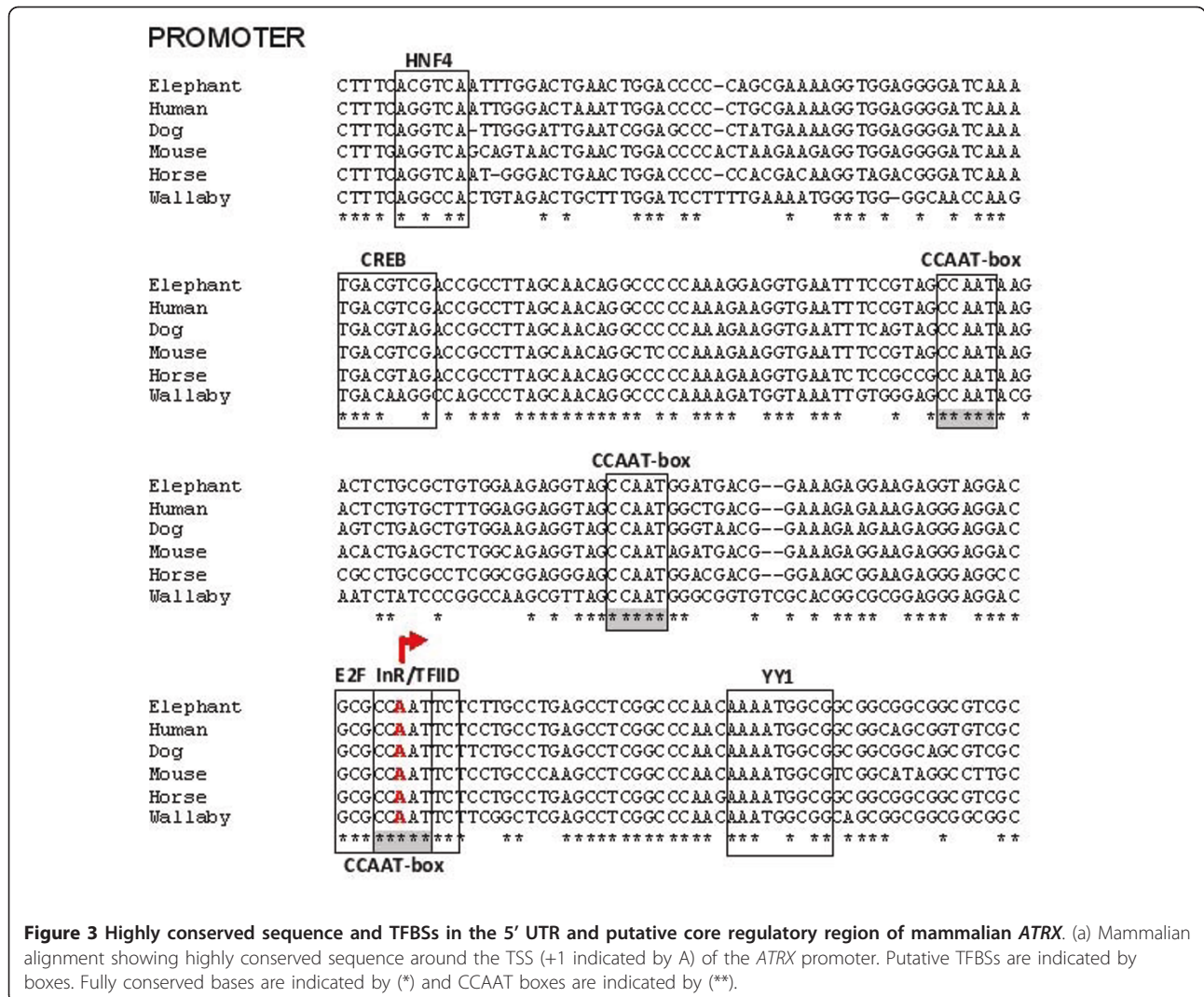
Putative TFBSs within evolutionarily conserved regions

Putative TFBSs within the putative promoter of *ATRX* (-500 to +300) were investigated using the online program, Mulan http://mulan.dcode.org/. Since bioinformatic analyses of DNA sequences typically produce a large number of potential TFBSs and because there is such high sequence conservation in this region, only TFBSs with 85%-100% sequence conservation and consensus identity, as defined by the program matrices, were considered for discussion. These are indicated in Figures 3 and 4.

In addition, a manual examination of the core promoter region for *ATRX* revealed a putative recognition site for TFIID known as the initiator (Inr) element (YYAN-WYY)[18,19] that is not represented in the TRANSFAC database, and which is fully conserved from human to wallaby (Figure 3). Within the core promoter region, there are also two fully conserved CCAAT-box domains as well as binding sites for HNF4 and CREB.

The putative TFBSs within CR1-3 were analysed by processing different paired combinations of mouse, human, horse, dog, elephant and armadillo sequences with Mulan http://mulan.dcode.org/. This is due to the

Figure 1 Phylogenetic relationship between studied species based on fossil and molecular data with branch lengths indicating approximate time since divergence [64].

Figure 2 ECRs in the 5' regulatory region of mammalian *ATRX* genes. Comparative analysis of the mammalian *ATRX* 5' genomic and 5' UTR sequences of representatives of three major eutherian clades - Euarchontoglires (human), Laurasiatheria (horse) and Afrotheria (elephant) - relative to that of mouse (Euarchontoglires), revealed four regions of high sequence conservation represented as standard stacked-pairwise graphs (default settings) on Mulan http://mulan.dcode.org/. Percentage values indicate sequence identity and red shading signifies regions of identity above 70%, calculated over 100 bp.

Figure 3 Highly conserved sequence and TFBSs in the 5' UTR and putative core regulatory region of mammalian *ATRX*. (a) Mammalian alignment showing highly conserved sequence around the TSS (+1 indicated by A) of the *ATRX* promoter. Putative TFBSs are indicated by boxes. Fully conserved bases are indicated by (*) and CCAAT boxes are indicated by (**).

fact that Mulan requires a relatively high conservation of sequence in order to identify TFBSs. Wallaby was not included in this figure due to low conservation of sequence in this region. The TFBSs found in CR1-3 are indicated in Figure 4 and Table 1 contains a summary of all highly conserved TFBSs in the mammalian *ATRX* 5'regulatory regions, in addition to the biological role of the corresponding transcription factors (TFs).

The first intron of some eukaryotic genes, including human, has been shown to contain regulatory elements which influence gene expression [20]. An attempt to align intron 1 of human, mouse, horse and dog *ATRX* revealed little sequence conservation. Pair-wise alignment of these species using both Mulan and ClustalW2 was only achievable for the combinations of human/mouse and human/horse and these alignments revealed only small regions of conservation that did not overlap indicating that it is unlikely that conserved TFBSs are located within intron 1.

Discussion

A key obstacle in the study of ATRX function has been the absence of identified upstream regulators. The availability of genome data from a wide range of species and the recent development of powerful bioinformatic tools has provided a means of studying the mammalian *ATRX* promoter and regulatory regions. In this study we attempted to derive information on the evolution and conservation of the *ATRX* 5'regulatory region for the purpose of identifying putative regulators of *ATRX*, and generate a foundation upon which to design future experiments. In our approach, we make the assumption that the transcriptional regulation of *ATRX* is conserved.

Analysis of the *ATRX* gene revealed high conservation of synteny from humans to birds (data not shown) and high sequence conservation between eutherian mammals and a marsupial, the tammar wallaby. Out of the (-13 kbp to +300 bp) studied, the region corresponding to (-500 to +300) of the human *ATRX* promoter region was found

```
CR1                                                                      CP2
Mouse     AACCTAGTCTGATTGTTGGAATGCATTGTTTGAAAAGAGACTCTTCTCTTGTCATTCTAGT
Human     AGCCTA-TCAGATTTTCAAAGTGTTTCATTCAAAAA-AGGCTCCCCTCTTGCTATCCCAGT
Horse     AGTCTA-TCAGATTATCAAAGTGTTCCATTCAAGAA-AGGCTCATCACTTGCTATTCCAGT
Dog       TGTCTA-TCAGATTGTCAAAATTTTCCATTGAAAAA-AGGCTCACTATTTGCGATTTCAGT
Elephant  AGTCCA-TCAGATTATCAAGGTGACTCATTCAAAAA-AGGCTAATTTCTTGCTATTCCAGT
Armadillo TTTCTA-TCAGATTATCAAAGTGATCCATTTTAAAA-AGGCTCATCTCCTGCTTTTGCAGT
              * ***** *      *     **    * ** **    **  *  * ***

CCAAAACTGCAACCCTGACATCTAAGCATGACCAAAACTGTTGCATATTTCAGGAAAAATAAACA------
AGAAAGCAGCAACCCTGTCACCTAACCCCGACAGGAGCTCTTGCAAATTTCAAGAAAAAGTAAATG------
AGAAAGCAGCAACCCTGTCACCTAAACACAACAGGAGGTCTTGCAAATTTCAAGAAAAGTAAATG------
GGAAAGCAGTAACTCTTTCACCTAAATACAATGGGAGCATTTTCAAATTTCAAGAAAATTAAATG------
AGAAAGCAGCAAC-------CTAACCACAACCGGGGTTCTCGCAAATTTCAAGAAAAACAAATA------
AGATAGCAGTAACCCTATCACCTAACCACAATAGGAGCTCTTGAAAATTTCAAGAAAAACAAATGTATTGA
      * * * ***   ****   ****    *  * *********** ***

----GAATGAACAAAATAAAGGGAATGAACATGTGTTCCAGAATAAATGTGGGCACAAAGTTGGAAAC--T
----GAATGAAC----TAAAG--------CATGCATGGCAGAATAAATTTGGAAATAAATTTCGAAGCGAT
----GAATGGAC----TAAAA--------CATGTATGGCGGAATAAAATTGGAAATAAAATTTGAAGCCAT
---------AAC----TAAAC--------TATATATGTCAGAATAAATTTGGAAATAAAATTAGAAACTAT
----AAATGAAC----TAAAG--------CACGCATGGTAGAATAAATTTCGAAATAAAAATTGAAGCCAT
ATAGGATTGAAG----TAAAG--------TATACATGGAAGAATAAATTTGGAAATAAAATTAGATGCTAT
        ****          ****       *   ******* *  *  * ***  **  *  *

TCAAATTTAGAAAACTTATTTGTGATTTGCAGA----------------------------GATTATGAA-
TAGATTTTAGAGGACTTGTTTGTGATTAGCAGA---ATTACCAATGTAAAATCAAATTATAGATCATAAAA
TCCATTTTAGAGAACTGATTTGTGACTTGCAGAG---ATTACCAATCGACAATTAAATTATAGAACATGAAA
TC-AATTTTGAAAACTAATTTGTGACTTGCAGA---CATTAGTAGTGGACAATTAAATTAGAGAACATGAAA
TTTATTTTAGAGAGCTGATTTGTGAGTTGCAGATGTAGTTACCAACGTAGAAACAAATTATAGAATATGAAA
AG-ATCTTAGAGGACTGGTTT-TCACTTGCAAAACTTGCTACTAATGTAGAATATAATTTTAGAACATGAAA
        ** **   **  *** * *   ***         **  *    * **   ****  ***  ** ** **

                                  E2F
--TGAAAAA-------------------------------TGAAATGAGGGGGAAAC-GATTTGGATCAGT
GCGGGAAAGGCATTT-ACAGATCATCCAGCCAAGCTTCCTCTGTAATGAGAAGGAAGCAGATCTAAATTGAT
GTTGGGGAGGCTTTT-AGAGCTCATCCAGCCAAA---CCTCTGTAATGAGGAGAAAATAGATGTAGATTGGT
ACTGGAAAGGCTTTTTAGAGATCATTCAGCCAAA---CTTCTGTAATGAGGAGAAAATAGACCTTGATTGGT
GCTGGAAAGACTTTC-AGATTCATCCAGCCAAACCTCCTCCGTAATGAGGTGGAAACAGATCTAGAATGGT
GCTGTAAATACTTTG-AGAGATCATACAGCCAAAACTCCTCTGTATTGAGGAAGAAACAGACCTAGAATGCT
    *     *     **  * *  *  ****** *      *  * * * ****   **  *** *  *     *

CR2
                                        HNF6
Mouse     CTTGAACATTTAAAAAT--TATTTGAAATGT-TGCAAGTTTCTGTTCATCAAAAA
Human     CTCAAACATTTGAAAGATATAATTGCAATGTGTACAGGCGTCTGTTGATTAAAAA
Horse     CTTGAATACTTGACAGA-GGACTTGAAATGTGTAAAAGCTTCTCTTGATTAAAAA
Elephant  CTTCAACATTTGAAAAT--GACTTCAAATCTGTACAGGCTTCAGTTGCTTAAAAA
Armadillo -TCCAACATTTAAAAGA-TGAGTTGAAATGTGTACAGGGTTCAGCTGATTAAAAA
              *   ** * **        **   ***    *      **    *  * *****

CR3
Mouse     TATTAG-AATTTAATGAGAACATCTGTTAAAATAAT---GAGCTGAGGTTTATGTCAGCAT
Human     TATTGA-AATCTAG-GAAAGTACCTTTTAAAATAATATTGAGCTAAGATTTGTGTGGAGAT
Horse     TATTTG-AATTTTGGTAGAGTACTTTTTAAAATAATATTGAGTTAGGGTTTATGTTGGCAT
Elephant  CCACCAGTAA--TTATCAA-TGAACCTTTTCA-TATT--GAACTAAGGTTTAGGTTGTCAT
Armadillo CTTCAAAAAAATAATAAAATAAAATGAAACAATATT---GATCTATGGTTTATGTTGCCAT
              *        *        ** *   * ** *   *** **   **

FOXJ2/NKX25
AATCCCTACT------TGC-TTCATATTTGTTGCTAGAAAATGGATAATCCTTTTGTTACTTGCAACAATTA
AATTGCTCTC------TGAATTCATATATGCTGGCAGAAATAGGTAGCCCTGTTACTGCTTACAACGATTA
AATTGATGCT------TGGATTCATACTTGCTAGCAGAAAATGGATGGTCCTATTGTTGCTTACCACAATTA
GATTGCTACT------TGAATTCATATCTACTGACAGAA----AATAGGGTTGTTATTGCTTACAACAGGTA
GATTGCTATTAATGTTTGGATTCATATGTGCTGGCAAAA--TGGATAGCCTTAT-GTTTCTTACAACAGTTA
 **   *         **  ******   *   * **      *    *** * **   **
```

Figure 4 Highly conserved sequence and TFBSs in the 5' proximal and distal regulatory regions of mammalian *ATRX*. Fully conserved bases are indicated by (*). The coordinates for Conserved Regions 1-3 (CR1-3) corresponding to the human sequence are (-1069 to -693) (-1623 to -1457) and (-12759 to -12669) respectively; however, only the regions containing putative TFBSs are indicated in this figure.

to be the most conserved with 82-90% identity between humans and other mammals studied. The high conservation between human and mouse coding sequences (88%), in addition to high conservation between the human and murine ATRX proteins [21], further supports the extrapolation of data generated from studies in mouse.

The core promoter is the fundamental part of the eukaryotic promoter that provides a platform for the assembly

Table 1 A summary of the 5' regulatory elements and corresponding TFs found in the *ATRX* upstream regulatory region and 5'UTR

	DNA binding consensus sequence	Putative binding site found in putative ATRX promoter	Biological roles	References
5'UTR of ATRX (+215 to +1)				
YY1 (+38 to +29)	SNCCATNTT	CGCCATTTT	Activator/repressor involved in neural development with roles in the testis and meiosis	[35,36]
ATRX promoter region				
Inr element (-2 to +3)	CTTNACC	CTTAACC	Binds to basal transcription factor TFIID	[19]
CCAAT boxes (-2 to +3) (-38 to -34) (-69 to -65)	CCAAT	CCAAT	Binds NF-Y, C/EBP, GATA1, NF-E3	[42]
E2F (-6 to -1)	TTTGGCGC	ATTGGCGC	Family of TFs that regulate mitosis, apoptosis, DNA repair, development and differentiation	[52]
CREB (-122 to -117)	TGACGTCA	CGACGTCA	Mediation of basal and PKA-inducible transcription	[65]
HNF4 (-173 to -168)	TGACCT	TGACCT	Liver factor involved in liver development and digestive function	[66]
CR1				
STAT5B (-963 to -956)	$TTCN_3GAA$	TTCAAGAA	Involved in immune response, erythropoesis and mammary development	[67]
CP2 (-1017 to -1003)	A/TCTGG-$CNRGN_6CNRS$	CTATT-$CCAGN_6GCAG$	Regulates transcription of *SRY*. Initially shown to bind murine alpha-globin gene	[50]
E2F (-796 to -791)	TTTGGCGC	TTTGGCTG	See E2F above	[52]
CR2				
HNF6 (-1413 to -1408)	GAATCA	GAAGCY	Liver factor involved in regulation of HNF4 and pancreatic precursor cells	[68]
CR3				
FOXJ2/NKX2-5 (-12720 to -12714)	TNAAGTG	TTAAGTA	Involved in regulation of cardiogenesis	[69]

of the RNA polymerase II initiation complex (reviewed in [22]). This region is usually within 80-100 bp surrounding the TSS and is typically a focused region, as opposed to dispersed, containing elements such as a TATA box, BRE, Inr, MTE, DPE and DCE. However, these elements are not universal and not all are present in the same promoters. Moreover, some core promoters do not contain any of the known core promoter elements. Bioinformatic analysis of the very highly conserved sequence around the TSS, which likely comprises the core promoter of *ATRX*, did not reveal any typical core promoter elements; however, manual examination of the core promoter region for *ATRX* revealed a potential initiator (Inr) element (YYAN-WYY) [19] which is a recognition site for TFIID, a key basal transcription factor (reviewed by [23]).

Analysis of the 5' regulatory region of mammalian *ATRX* revealed several highly conserved, potential regulatory motifs. Only those sites that could be verified in the literature in terms of sequence identity to an experimentally proven consensus binding sequence are summarised in Table 1.

In ATR-X syndrome, mental retardation manifests in approximately 95% of patients and alpha-thalassaemia, one of the defining symptoms of the syndrome, is present in 90% of cases while 80% of patients have genital abnormalities [1]. Significantly, the TFBSs identified in our study correlate to TFs with overlapping roles in neural, globin and testicular development, as well as roles in tissues that are affected to a lesser extent in ATR-X syndrome, suggesting that our findings are not likely to be the result of chance-occurrence of motifs in a DNA sequence. For example, both YY1 and CP2 are involved in brain development [24,25], NF-Y and CP2 have a role in globin regulation [17,26], YY1, NF-Y and CP2 are implicated in the development and function of the testis [27-29], and FOXJ2/NKX2-5 is one of the master transcriptional regulators of cardiogenesis [30] where cardiac defects are observed in 18% of *ATRX* patients.

Furthermore, like ATRX, most of these TFs are crucial for development (YY1, NF-Y, CP2, HNF4, FOXJ2) and ubiquitously expressed (YY1, NF-Y, CP2, E2F), indicating

spatial and functional compatibility with ATRX expression.

Interestingly, STAT5B, HNF4 and HNF6 play important roles in liver function [34-36] which do not feature in typical clinical descriptions of ATR-X syndrome yet raises the question of whether *ATRX*, which appears to be quite highly expressed in the fetal liver [31], has a role in liver development and/or function which is not clinically obvious in ATR-X syndrome. Alternatively, HNF4 expression is restricted to liver, kidney and intestines [32] and a regulatory pathway involving HNF4 and ATRX is possible given that renal-urinary abnormalities manifest in approximately 14% of ATR-X patients and gastro-intestinal problems arising from gut dismotility occur in 75% of clinical cases [1]. Furthermore, HNF4 binding sites in many genes expressed in the liver have been found in the vicinity of other liver-enriched TFs such as NF-Y and CREB [33] which is the case in the 5' regulatory region of *ATRX*. Thus, our investigation has revealed a handful of potential upstream regulators for *ATRX* expression which are of biological and clinical relevance. Some of these are discussed further.

The transcription factor Yin Yang (YY1) is a fully conserved, multifunctional protein that can function both as an activator or repressor of transcription (for a review, see [34]). Human YY1 is a sequence-specific DNA binding protein with consensus binding sequence: SNCCATNTT [35,36]. It was later reported that a similar consensus; CCATCTT with flanking nucleotides of G or C constitutes an 'activator' motif for YY1 function with the original consensus being characteristic of a 'repressor' site [37]. The consensus sequence identified in the eutherian alignment corresponds to a YY1 binding site for transcription repression, suggesting that YY1 may be a negative regulator of *ATRX* transcription.

Like *ATRX*, *YY1* is ubiquitously expressed and genetic ablation of *Atrx* and *Yy1* in separate mouse studies lead to lethality shortly after implantation, suggesting both proteins play crucial roles in development [6,27]. Furthermore, both ATRX and YY1 have important roles in the development of the brain and nervous system. For example, overexpression of *Atrx* in mouse embryos lead to similar neurological phenotypes of growth retardation, exencephaly and neural tube defects [11] that are also seen in heterozygous mutants of *Yy1* [24,38], suggesting that ATRX and YY1 could theoretically operate in the same pathway during mammalian neurodevelopment. An obvious hypothesis for the ATRX-YY1 relationship would involve YY1 as a negative upstream regulator of *ATRX* transcription. Such a model may seem spurious in light of the fact that ATR-X syndrome in humans is well documented to arise from reduced levels of functional ATRX protein [1]. However, overexpression of *Atrx* in transgenic mice gave rise to cranio-facial dysmorphology

and seizures which are reminiscent of ATR-X syndrome [11], indicating that expression levels of ATRX require strict regulation for the survival of cells and organisms. In addition, deletion of the 5' UTR of *ATRX*, where the putative YY1 binding site is located, has been reported in an *ATR*-X patient [31] and while the precise mechanism of disease arising from this clinical mutation is unclear, we speculate that YY1 could constitute the negative feedback loop for *ATRX* expression while a different transcription factor may be involved in the upregulation of *ATRX*.

Additionally, separate immunostaining of YY1 and ATRX in rodent testis sections reveal a strikingly similar pattern of expression as both proteins are expressed in Sertoli and Leydig cells and in spermatogonia and spermatocytes but not round spermatids and spermatozoa [27,39,40]. Moreover, both YY1 and ATRX have been shown to cause defects in meiosis when ablated in spermatocytes and oocytes respectively [27,41]. Thus, it is possible that YY1 may also regulate *ATRX* in the mammalian testis.

Unrelated promoters for eukaryotic RNA polymerase II have several common elements including the cap signal, TATA box, CCAAT box and GC boxes [42]. We identified three CCAAT motifs at (+3), (-34) and (-65) which are described in one of the earliest publications on the *ATRX* gene and protein [16]. In addition, the spacing between all three CCAAT boxes and between the CCAAT boxes and other regulatory motifs is conserved across all eutherian promoters studied, consistent with findings which suggest that in higher eukaryotic promoters, sequences around the CCAAT box and the spacing between the CCAAT box and other regulatory motifs are highly conserved for specific genes in different species [43,44].

The CCAAT box is estimated to be present in one third of eukaryotic house-keeping and lineage-specific genes [16] and multiple copies are found in promoters of genes regulated during cell proliferation [42]. Notably, ATRX is known to be regulated by phosphorylation in a cell-cycle dependant manner [12] and the *ATRX* promoter with its multiple CCAAT boxes is typical of a gene that is regulated during cell proliferation.

Of all the transcription factors that bind CCAAT boxes, NF-Y (also known as CBF and CP1) can be distinguished from other proteins on the basis of DNA sequence requirements since NF-E3, GATA1 and C/EBP often bind sites containing an incomplete CCAAT motif [42]. High affinity binding sites for NF-Y contain YRR at the 5' flanking end of the CCAAT box and CA at the 3' adjacent sequence. Changes in three or more nucleotides decreases or abolishes NF-Y binding, while one or two nucleotide changes lead to a modest decrease in NF-Y binding [42]. The CCAAT boxes we identify in the *ATRX*

promoter contain highly conserved GCG (+2) and TAG (-34, -65) at the 5' flanking sequence and TG (+2), GG (-34) and AA (-65) at the 3' flanking sequence. Thus, by definition, the highly conserved CCAAT boxes at (-34 and -65) would be able to bind NF-Y.

NF-Y is an evolutionarily conserved TF present from yeast to human [45]. A double mutation leads to embryonic lethality in mice as early as 8.5 *dpc* while heterozygous mutations are similar to wild-type [45]. NF-Y has been shown to regulate genes in cooperation with members of the Sp (specificity protein) family of TFs. For example, NF-Y and Sp3 together regulate the *MAP kinase* (mitogen-activated protein kinase) gene [46], and NF-Y and Sp1 have known roles in the regulation of *topo II* (topoisomerase II) [47], *SF-1* (steroidogenic factor 1) [48] and *SOX9* (SRY-related HMG box 9) [28]. ATRX, SF-1 and SOX9 play crucial roles in the testicular development pathway and thus it is possible that ATRX, SF-1 and SOX9 are regulated by a similar set of upstream factors - including NF-Y.

The CP2 TF family exists as six isoforms in human (LBP-1a-d, -9 and -32) and four in mouse (CP2a-c and CRTR). CP2, also known as CP2c, has important roles in hematopoeisis, immune response, cell cycle and neural development [25]. These biological processes advocate CP2 to be a temporal, spatial and functionally compatible regulator of *ATRX*. While immune response has not been typically linked to ATRX function, a recent study implicated ATRX and one of its cofactors, DAXX [6], in intrinsic antiviral resistance to the HSV-1 virus [49].

In addition, CP2c was initially identified as an activator of mouse alpha-globin via the consensus sequence CNRGN6CNRS [50] and thus, the putative regulatory pathway between CP2c and ATRX may represent the first direct biochemical connection between ATRX and the regulation of alpha-globin expression. Such a link has been elusive despite the fact that clinical mutations of the *ATRX* gene lead to a downregulation of alpha-globin resulting in alpha-thalassaemia [31]. One possible scenario for alpha-globin regulation is a cascade that involves CP2c activation of ATRX expression which results in ATRX-mediated repression of a negative regulator of alpha-globin expression.

A recent study identified putative CP2 sites in the promoter of testis-determining factor, *SRY*, and showed that CP2 activates transcription of the *SRY* through these sites in a cell-based assay [30]. Binding of CP2 to the *SRY* promoter was confirmed by ChIP analysis in the human testis cell line, NT2/D1 [30]. Within the currently understood casade of genes involved in testicular development, ATRX function lies downstream of SRY and SOX9 ([51], A. Argentaro pers.comm). Thus, in the context of testicular differentiation, CP2 may initiate the pathway for testicular development via activation of *SRY*

which eventually leads to ATRX expression. No putative SRY binding sites were identified in our eutherian alignments suggesting SRY is not a direct regulator of *ATRX* transcription.

The E2F family of TFs consists of 8 genes whose products bind to the same consensus sequence: TTTSSCGC (reviewed by [52,53]). Mutational studies of putative E2F binding sites suggest that E2F TFs can function both as activators or repressors of transcription depending upon the promoter context. This is because similar mutations can result in upregulation or downregulation of a gene depending on which gene promoter is being mutated (reviewed by [54]). Interestingly, only a very small number (~23%) of *in vivo* E2F sites actually contain the consensus motif and an analysis of conforming and non-conforming *in vivo* sites suggest that (i) a *bona fide* site must be in the core promoter (+1 to -100) and (ii) the region must be utilised as a promoter in that cell type [55]. One of the two E2F consensus motifs identified in this study (-1 to -6) conforms to these conditions and thus may be a biologically active regulatory element within the *ATRX* promoter.

Members of the E2F family are implicated in a myriad of functions including the regulation of mitosis, apoptosis, DNA repair, differentiation, development and tumourigenesis, with no clear division of labour between members [54]. These roles are very similar to that of the SNF2 family of TFs of which ATRX is a member [16] and recent studies indicate a role for ATRX in apoptosis as inactivation of *Atrx* in mouse forebrain resulted in p53-dependant apoptosis of neuroprogenitors in the developing mouse [56]. In addition, E2F binding sites have been identified in the promoters of many genes intimately involved in the regulation of cell cycle progression [54] and *ATRX* could be one of these genes since *ATRX* null mutations can affect cell-cycle progression and interfere with proper chromosome segregation in both somatic and germ cells [41,57].

Conclusion

We identified twelve highly conserved TFBSs in the 5' regulatory region and 5' UTR of mammalian *ATRX* that are likely candidates for regulating *ATRX* expression. Future experiments involving DNA binding, mutational analysis, ChIP and transactivation assays will verify whether these sites are *bona fide* regulatory motifs. Furthermore, the identification of ECRs in the *ATRX* promoter region could provide a starting point for experimental studies to define the minimal promoter of *ATRX*. While a literature search did not reveal any idiopathic cases of ATR-X syndrome which are could be attributable to mutations in the 5' regulatory regions, an understanding of how *ATRX* is regulated is an important component of understanding its functions during development and other important biological processes.

Table 2 References to the sequence data used in this study

Species	Reference sequence	Source
human	NCBI36/hg18 chrX: 76928075-76941134	UCSC Genome Browser http://genome.ucsc.edu/
mouse	NCBI37/mm9 chrX:103124406-103128433	UCSC Genome Browser http://genome.ucsc.edu/
dog	Broad/canFam2 chrX:63091021-63104321	UCSC Genome Browser http://genome.ucsc.edu/
horse	Broad/equCab2 chrX:57708104-57721404	UCSC Genome Browser http://genome.ucsc.edu/
elephant	Broad/loxAfr3 scaffold24:23535243-23535990	UCSC Genome Browser http://genome.ucsc.edu/
armadillo	AAGV020428528	GenBank
wallaby	ABQO010522248	GenBank

Methods

To compare the human *ATRX* 5' regulatory region to that of an evolutionarily divergent range of species, input sequences were retrieved with BLAT searches [58] at the UCSC Genome Browser http://genome.ucsc.edu/ or by BLAST searches of whole genome shotgun databases, using human sequence corresponding to the 5' regulatory region of the *ATRX* gene (-13 kbp to +300 bp). Only species for which an *ATRX* orthologue was definitively identified and for which sufficient 5' sequence was retrievable were included in the generation of each results figure. These belonged to human (*Homo sapiens*), mouse (*Mus musculus*), dog (*Canis familiaris*), horse (*Equus caballus*), elephant (*Loxodonta africana*), armadillo (*Dasypus novemcinctus*) and tammar wallaby (*Macropus eugenii*) (Table 2).

Multiple sequence alignments were compiled using Mulan in 'TBA' mode http://mulan.dcode.org/[59,60]. Candidate TFBSs within ECRs were identified using multiTF [60][61] from the Mulan website, selecting the TRANSFAC professional V10.2 TFBS database for vertebrates, with matrix similarity set at 0.85, 0.95 or 1.0, and using only high-specificity matrices. The analysis was performed on large regions of upstream genomic sequences as well as on individual, short ECRs identified from the initial alignments. Identification of TFBSs was performed on alignments of two-to-several species from diverse taxonomic groups and only those conserved in at least two species are reported here. In addition, Map Viewer http://www.ncbi.nlm.nih.gov/mapview/[62] was used to investigate the conservation of synteny around the *ATRX* gene between species, Readseq http://www-bimas.cit.nih.gov/molbio/readseq/ was used to convert sequences to FASTA format, and ClustalW2 http://www.ebi.ac.uk/Tools/clustalw2/index.html[63] was used as an additional sequence alignment tool.

See Additional file 1: Input sequences, for the original data used to perform our analyses.

Additional material

Additional file 1: Input sequences. The input sequences used in this study are provided in FASTA format

List of Abbreviations

ECR: evolutionarily conserved region; TFBS(s): transcription factor binding site (s); 5'UTR: 5' untranslated region; TSS: transcription start site; CR: conserved region; TF: transcription factor; DNA: deoxyribonucleic acid; bp: base pairs; kb: kilobase pair

Author details
¹Department of Zoology, University of Melbourne, Victoria 3010, Australia. ²Prince Henry's Institute of Medical Research, PO Box 5152, Clayton, Victoria 3168, Australia. ³Research School of Biological Sciences, the Australian National University, Canberra, ACT, 2601, Australia.

Authors' contributions
PT conceived of the study, PT and SF designed the study and carried out the experiments, PT, SF, TA, JMG, MF wrote and drafted the manuscript. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

References
1. Gibbons RJ, Higgs DR: **Molecular-clinical spectrum of the ATR-X syndrome.** *Am J Med Genet* 2000, **97(3)**:204-212.
2. Gibbons RJ, Wilkie AO, Weatherall DJ, Higgs DR: **A newly defined × linked mental retardation syndrome associated with alpha thalassaemia.** *J Med Genet* 1991, **28(11)**:729-733.
3. Wilkie AO, Zeitlin HC, Lindenbaum RH, Buckle VJ, Fischel-Ghodsian N, Chui DH, Gardner-Medwin D, MacGillivray MH, Weatherall DJ, Higgs DR: **Clinical features and molecular analysis of the alpha thalassemia/mental retardation syndromes. II. Cases without detectable abnormality of the alpha globin complex.** *Am J Hum Genet* 1990, **46(6)**:1127-1140.
4. Argentaro A, Yang JC, Chapman L, Kowalczyk MS, Gibbons RJ, Higgs DR, Neuhaus D, Rhodes D: **Structural consequences of disease-causing mutations in the ATRX-DNMT3-DNMT3L (ADD) domain of the chromatin-associated protein ATRX.** *Proc Natl Acad Sci USA* 2007, **104(29)**:11939-11944.
5. Tang J, Wu S, Liu H, Stratt R, Barak OG, Shiekhattar R, Picketts DJ, Yang X: **A novel transcription regulatory complex containing death domain-associated protein and the ATR-X syndrome protein.** *J Biol Chem* 2004, **279(19)**:20369-20377.
6. Xue Y, Gibbons R, Yan Z, Yang D, McDowell TL, Sechi S, Qin J, Zhou S, Higgs D, Wang W: **The ATRX syndrome protein forms a chromatin-remodeling complex with Daxx and localizes in promyelocytic leukemia nuclear bodies.** *Proc Natl Acad Sci USA* 2003, **100(19)**:10635-10640.
7. Garrick D, Sharpe JA, Arkell R, Dobbie L, Smith AJ, Wood WG, Higgs DR, Gibbons RJ: **Loss of Atrx affects trophoblast development and the pattern of X-inactivation in extraembryonic tissues.** *PLoS Genet* 2006, **2(4)**:e58.
8. Medina CF, Mazerolle C, Wang Y, Berube NG, Coupland S, Gibbons RJ, Wallace VA, Picketts DJ: **Altered visual function and interneuron survival**

in Atrx knockout mice: Inference for the human syndrome. *Hum Mol Genet* 2008.

9. Berube NG, Mangelsdorf M, Jagla M, Vanderluit J, Garrick D, Gibbons RJ, Higgs DR, Slack RS, Picketts DJ: **The chromatin-remodeling protein ATRX is critical for neuronal survival during corticogenesis.** *J Clin Invest* 2005, **115(2)**:258-267.

10. Solomon LA, Li JR, Berube NG, Beier F: **Loss of ATRX in chondrocytes has minimal effects on skeletal development.** *PLoS One* 2009, **4(9)**:e7106.

11. Berube NG, Jagla M, Smeenk C, De Repentigny Y, Kothary R, Picketts DJ: **Neurodevelopmental defects resulting from ATRX overexpression in transgenic mice.** *Hum Mol Genet* 2002, **11(3)**:253-261.

12. Berube NG, Smeenk C, Picketts DJ: **Cell cycle-dependent phosphorylation of the ATRX protein correlates with changes in nuclear matrix and chromatin association.** *Hum Mol Genet* 2000, **9(4)**:539-547.

13. Cardoso C, Timsit S, Villard L, Khrestchatisky M, Fontes M, Colleaux L: **Specific interaction between the XNP/ATR-X gene product and the SET domain of the human EZH2 protein.** *Hum Mol Genet* 1998, **7(4)**:679-684.

14. Nan X, Hou J, Maclean A, Nasir J, Lafuente MJ, Shu X, Kriaucionis S, Bird A: **Interaction between chromatin proteins MECP2 and ATRX is disrupted by mutations that cause inherited mental retardation.** *Proc Natl Acad Sci USA* 2007, **104(8)**:2709-2714.

15. Kernohan KD, Jiang Y, Tremblay DC, Bonvissuto AC, Eubanks JH, Mann MR, Berube NG: **ATRX partners with cohesin and MeCP2 and contributes to developmental silencing of imprinted genes in the brain.** *Dev Cell* 2010, **18(2)**:191-202.

16. Picketts DJ, Higgs DR, Bachoo S, Blake DJ, Quarrell OW, Gibbons RJ: **ATRX encodes a novel member of the SNF2 family of proteins: mutations point to a common mechanism underlying the ATR-X syndrome.** *Hum Mol Genet* 1996, **5(12)**:1899-1907.

17. Fang X, Han H, Stamatoyannopoulos G, Li Q: **Developmentally specific role of the CCAAT box in regulation of human gamma-globin gene expression.** *J Biol Chem* 2004, **279(7)**:5444-5449.

18. Javahery R, Khachi A, Lo K, Zenzie-Gregory B, Smale ST: **DNA sequence requirements for transcriptional initiator activity in mammalian cells.** *Mol Cell Biol* 1994, **14(1)**:116-127.

19. Smale ST, Baltimore D: **The "initiator" as a transcription control element.** *Cell* 1989, **57(1)**:103-113.

20. Bornstein P, McKay J, Morishima JK, Devarayalu S, Gelinas RE: **Regulatory elements in the first intron contribute to transcriptional control of the human alpha 1(I) collagen gene.** *Proc Natl Acad Sci USA* 1987, **84(24)**:8869-8873.

21. Picketts DJ, Tastan AO, Higgs DR, Gibbons RJ: **Comparison of the human and murine ATRX gene identifies highly conserved, functionally important domains.** *Mamm Genome* 1998, **9(5)**:400-403.

22. Juven-Gershon T, Hsu JY, Theisen JW, Kadonaga JT: **The RNA polymerase II core promoter - the gateway to transcription.** *Curr Opin Cell Biol* 2008, **20(3)**:253-259.

23. Juven-Gershon T, Kadonaga JT: **Regulation of gene expression via the core promoter and the basal transcriptional machinery.** *Dev Biol* 2009, **339(2)**:225-229.

24. Donohoe ME, Zhang X, McGinnis L, Biggers J, Li E, Shi Y: **Targeted disruption of mouse Yin Yang 1 transcription factor results in peri-implantation lethality.** *Mol Cell Biol* 1999, **19(10)**:7237-7244.

25. Veljkovic J, Hansen U: **Lineage-specific and ubiquitous biological roles of the mammalian transcription factor LSF.** *Gene* 2004, **343(1)**:23-40.

26. Barnhart KM, Kim CG, Banerji SS, Sheffery M: **Identification and characterization of multiple erythroid cell proteins that interact with the promoter of the murine alpha-globin gene.** *Mol Cell Biol* 1988, **8(8)**:3215-3226.

27. Wu S, Hu YC, Liu H, Shi Y: **Loss of YY1 impacts the heterochromatic state and meiotic double-strand breaks during mouse spermatogenesis.** *Mol Cell Biol* 2009, **29(23)**:6245-6256.

28. Piera-Velazquez S, Hawkins DF, Whitecavage MK, Colter DC, Stokes DG, Jimenez SA: **Regulation of the human SOX9 promoter by Sp1 and CREB.** *Exp Cell Res* 2007, **313(6)**:1069-1079.

29. Sato Y, Shinka T, Sakamoto K, Ewis AA, Nakahori Y: **The male-determining gene SRY is a hybrid of DGCR8 and SOX3, and is regulated by the transcription factor CP2.** *Mol Cell Biochem* 2010, **337(1-2)**:267-275.

30. Riazi AM, Takeuchi JK, Hornberger LK, Zaidi SH, Amini F, Coles J, Bruneau BG, Van Arsdell GS: **NKX2-5 regulates the expression of beta-catenin and GATA4 in ventricular myocytes.** *PLoS One* 2009, **4(5)**:e5698.

31. Gibbons RJ, Wada T, Fisher CA, Malik N, Mitson MJ, Steensma DP, Fryer A, Goudie DR, Krantz ID, Traeger-Synodinos J: **Mutations in the chromatin-associated protein ATRX.** *Hum Mutat* 2008, **29(6)**:796-802.

32. Zhong W, Mirkovitch J, Darnell JE Jr: **Tissue-specific regulation of mouse hepatocyte nuclear factor 4 expression.** *Mol Cell Biol* 1994, **14(11)**:7276-7284.

33. Viollet B, Kahn A, Raymondjean M: **Protein kinase A-dependent phosphorylation modulates DNA-binding activity of hepatocyte nuclear factor 4.** *Mol Cell Biol* 1997, **17(8)**:4208-4219.

34. He Y, Casaccia-Bonnefil P: **The Yin and Yang of YY1 in the nervous system.** *J Neurochem* 2008, **106(4)**:1493-1502.

35. Hyde-DeRuyscher RP, Jennings E, Shenk T: **DNA binding sites for the transcriptional activator/repressor YY1.** *Nucleic Acids Res* 1995, **23(21)**:4457-4465.

36. Yant SR, Zhu W, Millinoff D, Slightom JL, Goodman M, Gumucio DL: **High affinity YY1 binding motifs: identification of two core types (ACAT and CCAT) and distribution of potential binding sites within the human beta globin cluster.** *Nucleic Acids Res* 1995, **23(21)**:4353-4362.

37. Shrivastava A, Calame K: **An analysis of genes regulated by the multi-functional transcriptional regulator Yin Yang-1.** *Nucleic Acids Res* 1994, **22(24)**:5151-5155.

38. Kwon HJ, Chung HM: **Yin Yang 1, a vertebrate polycomb group gene, regulates antero-posterior neural patterning.** *Biochem Biophys Res Commun* 2003, **306(4)**:1008-1013.

39. Baumann C, Schmidtmann A, Muegge K, De La Fuente R: **Association of ATRX with pericentric heterochromatin and the Y chromosome of neonatal mouse spermatogonia.** *BMC Mol Biol* 2008, **9**:29.

40. Tang P, Argentaro A, Pask AJ, O'Donnell L, Marshall-Graves J, Familari M, Harley VR: **Localisation of the Chromatin Remodelling Protein, ATRX in the Adult Testis.** *J Reprod Dev* 2009.

41. De La Fuente R, Viveiros MM, Wigglesworth K, Eppig JJ: **ATRX, a member of the SNF2 family of helicase/ATPases, is required for chromosome alignment and meiotic spindle organization in metaphase II stage mouse oocytes.** *Dev Biol* 2004, **272(1)**:1-14.

42. Maity SN, de Crombrugghe B: **Role of the CCAAT-binding protein CBF/NF-Y in transcription.** *Trends Biochem Sci* 1998, **23(5)**:174-178.

43. Chodosh LA, Baldwin AS, Carthew RW, Sharp PA: **Human CCAAT-binding proteins have heterologous subunits.** *Cell* 1988, **53(1)**:11-24.

44. Takahashi Y, Ohoka N, Hayashi H, Sato R: **TRB3 suppresses adipocyte differentiation by negatively regulating PPARgamma transcriptional activity.** *J Lipid Res* 2008, **49(4)**:880-892.

45. Bhattacharya A, Deng JM, Zhang Z, Behringer R, de Crombrugghe B, Maity SN: **The B subunit of the CCAAT box binding transcription factor complex (CBF/NF-Y) is essential for early mouse development and cell proliferation.** *Cancer Res* 2003, **63(23)**:8167-8172.

46. Sugiura N, Takishima K: **Regulation of the gene promoter for extracellular signal-regulated protein kinase 2 by transcription factors NF-Y and Sp3.** *Biochem J* 2000, **347(Pt 1)**:155-161.

47. Magan N, Szremska AP, Isaacs RJ, Stowell KM: **Modulation of DNA topoisomerase II alpha promoter activity by members of the Sp (specificity protein) and NF-Y (nuclear factor Y) families of transcription factors.** *Biochem J* 2003, **374(Pt 3)**:723-729.

48. Daggett MA, Rice DA, Heckert LL: **Expression of steroidogenic factor 1 in the testis requires an E box and CCAAT box in its promoter proximal region.** *Biol Reprod* 2000, **62(3)**:670-679.

49. Lukashchuk V, Everett RD: **Regulation of ICP0-null mutant herpes simplex virus type 1 infection by ND10 components ATRX and hDaxx.** *J Virol* 2010, **84(8)**:4026-4040.

50. Chae JH, Kang HC, Kim CG: **The relative cellular levels of CP2a and CP2b potentiates erythroid cell-specific expression of the alpha-globin gene by regulating the nuclear localization of CP2c.** *Biochem Biophys Res Commun* 2009, **380(4)**:813-817.

51. Tang P, Park DJ, Marshall Graves JA, Harley VR: **ATRX and sex differentiation.** *Trends Endocrinol Metab* 2004, **15(7)**:339-344.

52. DeGregori J, Johnson DG: **Distinct and Overlapping Roles for E2F Family Members in Transcription, Proliferation and Apoptosis.** *Curr Mol Med* 2006, **6(7)**:739-748.

53. Dimova DK, Dyson NJ: **The E2F transcriptional network: old acquaintances with new faces.** *Oncogene* 2005, **24(17)**:2810-2826.

54. Sahin F, Sladek TL: **E2F-1 has dual roles depending on the cell cycle.** *Int J Biol Sci* 2010, **6(2)**:116-128.

55. Rabinovich A, Jin VX, Rabinovich R, Xu X, Farnham PJ: **E2F in vivo binding specificity: comparison of consensus versus nonconsensus binding sites.** *Genome Res* 2008, **18(11)**:1763-1777.
56. Seah C, Levy MA, Jiang Y, Mokhtarzada S, Higgs DR, Gibbons RJ, Berube NG: **Neuronal death resulting from targeted disruption of the Snf2 protein ATRX is mediated by p53.** *J Neurosci* 2008, **28(47)**:12570-12580.
57. Ritchie K, Seah C, Moulin J, Isaac C, Dick F, Berube NG: **Loss of ATRX leads to chromosome cohesion and congression defects.** *J Cell Biol* 2008, **180(2)**:315-324.
58. Kent WJ: **BLAT - the BLAST-like alignment tool.** *Genome Res* 2002, **12(4)**:656-664.
59. Ovcharenko I, Loots GG, Giardine BM, Hou M, Ma J, Hardison RC, Stubbs L, Miller W: **Mulan: multiple-sequence local alignment and visualization for studying function and evolution.** *Genome Res* 2005, **15(1)**:184-194.
60. Loots GG, Ovcharenko I: **Mulan: multiple-sequence alignment to predict functional elements in genomic sequences.** *Methods Mol Biol* 2007, **395**:237-254.
61. Loots GG, Ovcharenko I: **Dcode.org anthology of comparative genomic tools.** *Nucleic Acids Res* 2005, **33(Web Server issue)**:W56-64.
62. Dombrowski SM, Maglott D: **Using the Map Viewer to Explore Genomes.** In *The NCBI Handbook.* Edited by: McEntyre J, Ostell J Bethesda. National Center for Biotechnology Information; 2002:.
63. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD: **Multiple sequence alignment with the Clustal series of programs.** *Nucleic Acids Res* 2003, **31(13)**:3497-3500.
64. Bininda-Emonds OR, Cardillo M, Jones KE, MacPhee RD, Beck RM, Grenyer R, Price SA, Vos RA, Gittleman JL, Purvis A: **The delayed rise of present-day mammals.** *Nature* 2007, **446(7135)**:507-512.
65. Montminy MR, Sevarino KA, Wagner JA, Mandel G, Goodman RH: **Identification of a cyclic-AMP-responsive element within the rat somatostatin gene.** *Proc Natl Acad Sci USA* 1986, **83(18)**:6682-6686.
66. Jiang G, Nepomuceno L, Hopkins K, Sladek FM: **Exclusive homodimerization of the orphan receptor hepatocyte nuclear factor 4 defines a new subclass of nuclear receptors.** *Mol Cell Biol* 1995, **15(9)**:5131-5143.
67. Hennighausen L, Robinson GW: **Interpretation of cytokine signaling through the transcription factors STAT5A and STAT5B.** *Genes Dev* 2008, **22(6)**:711-721.
68. Briancon N, Bailly A, Clotman F, Jacquemin P, Lemaigre FP, Weiss MC: **Expression of the alpha7 isoform of hepatocyte nuclear factor (HNF) 4 is activated by HNF6/OC-2 and HNF1 and repressed by HNF4alpha1 in the liver.** *J Biol Chem* 2004, **279(32)**:33398-33408.
69. Chen CY, Schwartz RJ: **Identification of novel DNA binding targets and regulatory domains of a murine tinman homeodomain factor, nkx-2.5.** *J Biol Chem* 1995, **270(26)**:15628-15633.