Hum Genet (2015) 134:1163–1182
DOI 10.1007/s00439-015-1594-x

ORIGINAL INVESTIGATION

Copy number variants in patients with intellectual disability affect the regulation of *ARX* transcription factor gene

Minaka Ishibashi[1] · Elizabeth Manning[1] · Cheryl Shoubridge[2] · Monika Krecsmarik[3] · Thomas A. Hawkins[4] · Jean Giacomotto[1] · Ting Zhao[1,5] · Thomas Mueller[6] · Patricia I. Bader[7] · Sau W. Cheung[8] · Pawel Stankiewicz[8] · Nicole L. Bain[9] · Anna Hackett[10,11] · Chilamakuri C. S. Reddy[12,13] · Alejandro S. Mechaly[1,14] · Bernard Peers[15] · Stephen W. Wilson[4] · Boris Lenhard[12,16,17] · Laure Bally-Cuif[3] · Jozef Gecz[2,18] · Thomas S. Becker[1] · Silke Rinkwitz[1]

Received: 16 April 2015 / Accepted: 16 August 2015 / Published online: 4 September 2015

Abstract Protein-coding mutations in the transcription factor-encoding gene *ARX* cause various forms of intellectual disability (ID) and epilepsy. In contrast, variations in surrounding non-coding sequences are correlated with milder forms of non-syndromic ID and autism and had suggested the importance of *ARX* gene regulation in the etiology of these disorders. We compile data on several novel and some already identified patients with or without ID that carry duplications of *ARX* genomic region and consider likely genetic mechanisms underlying the neurodevelopmental defects. We establish the long-range regulatory domain of *ARX* and identify its brain region-specific autoregulation. We conclude that neurodevelopmental disturbances in the patients may not simply arise from increased dosage due to *ARX* duplication. This is further exemplified by a small duplication involving a non-functional *ARX* copy, but with duplicated enhancers. *ARX* enhancers are located within a 504-kb region and regulate expression specifically in the forebrain in developing and adult zebrafish. Transgenic enhancer-reporter lines were used as in vivo tools to delineate a brain region-specific

Electronic supplementary material The online version of this article (doi:10.1007/s00439-015-1594-x) contains supplementary material, which is available to authorized users.

✉ Silke Rinkwitz
silke.rinkwitz@sydney.edu.au

1 Department of Physiology, Brain and Mind Research Institute, Sydney Medical School, University of Sydney, 94 Mallet Street, Camperdown, NSW 2050, Australia

2 School of Paediatrics and Reproductive Health, University of Adelaide, Adelaide, SA 5006, Australia

3 Laboratory of Neurobiology and Development, CNRS, Institute of Neurobiology Alfred Fessard, 91198 Gif-sur-Yvette Cédex, France

4 Department of Cell and Developmental Biology, UCL, London WC1E 6BT, UK

5 Department of Neurology, The People's hospital of Zhengzhou University, Henan 450003, China

6 Division of Biology, Kansas State University, Manhattan, KS 66502, USA

7 Parkview Cytogenetics and Northeast Indiana Genetics, Fort Wayne, IN, USA

8 Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, NAB 2015, Houston, USA

9 Department of Molecular Medicine, Pathology North, John Hunter Hospital, Newcastle, NSW 2305, Australia

10 Genetics of Learning Disability Service, Hunter Genetics, John Hunter Hospital, Newcastle, NSW 2305, Australia

11 School of Medicine and Public Health, University of Newcastle, Newcastle, NSW 2305, Australia

12 Department of Biology and Bergen Center for Computational Science, University of Bergen, 5008 Bergen, Norway

13 Department of Tumor Biology, Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway

14 Laboratorio de Ictiofisiología y Acuicultura, Instituto Tecnológico de Chascomús (IIB-INTECH) (CONICET-UNSAM), Buenos Aires, Argentina

15 GIGA Research, Universite de Liege, 4000 Liege, Belgium

16 Institute of Clinical Sciences, Faculty of Medicine, Imperial College London, London, UK

17 MRC Clinical Sciences Centre, London, UK

18 Neurogenetics Laboratory, Genetics and Molecular Pathology, SA Pathology at the Women's and Children's Hospital, Adelaide, SA 5006, Australia

negative and positive autoregulation of *ARX*. We find autorepression of *ARX* in the telencephalon and autoactivation in the ventral thalamus. Fluorescently labeled brain regions in the transgenic lines facilitated the identification of neuronal outgrowth and pathfinding disturbances in the ventral thalamus and telencephalon that occur when *arxa* dosage is diminished. In summary, we have established a model for how breakpoints in long-range gene regulation alter the expression levels of a target gene brain region-specifically, and how this can cause subtle neuronal phenotypes relating to the etiology of associated neuropsychiatric disease.

Introduction

Approximately 5–10 % of X-chromosome linked intellectual disability (XLID) cases are caused by mutations in *ARX*, encoding the aristaless-related homeobox transcription factor (Bienvenu et al. 2002; Partington et al. 2004). Both polyalanine expansions and coding sequence mutations of *ARX* lead to ID accompanied by other clinical comorbidities including lissencephaly, epilepsy and infantile spasms, and hand dystonia with varying severity (Stromme et al. 2002; Bienvenu et al. 2002; Gecz et al. 2006; Kitamura et al. 2009). Whereas disruptions in *ARX* coding sequence cause developmental defects, polyalanine expansions have been associated with non-syndromic ID as well as epilepsy (Price et al. 2009; Shoubridge et al. 2010).

Mouse models of *Arx* deficiency as well as polyalanine expansions identified Arx as an important regulator of telencephalic development, showing that defects in tangential migration and differentiation of GABAergic and cholinergic neurons are potentially cardinal developmental defects leading to ID and lissencephaly (Kitamura et al. 2002, 2009). Recent mutation screening in families with ID and/or autism identified three distinct point mutations in ultraconserved non-coding sequences in the vicinity of *ARX* that were correlated to autism and infantile spasms (Fullston et al. 2011). While the pathogenicity of these mutations was not assessed, their predicted cis-regulatory function suggested functionality in these patients. The transcriptional cascade modulating *Arx* is connected to Dlx transcription factor function and promotes GABAergic neuron development in the telencephalon and other parts of the forebrain (Colasante et al. 2008). It regulates diverse aspects of neurogenesis such as neuronal stem cell proliferation, migration and differentiation (Colombo et al. 2004, 2007; Friocourt et al. 2006, 2008). Consequently, its target genes function in neuronal differentiation and axonal guidance, but also in synaptic activity (Seufert et al. 2005; Fulp et al. 2008). Many of these are correlated with schizophrenia, intellectual disability, autism(s) and depressive disorders, stressing the importance of Arx-regulated signaling pathways for development and functioning of the forebrain (Quillé et al. 2011).

Arx functions in multiple organs during vertebrate development (Miura et al. 1997; Biressi et al. 2008; Collombat et al. 2009; Djiotsa et al. 2012). Usually, genes with pleiotropic functions have multiple enhancers that orchestrate their regulation in time and space (Komisarczuk et al. 2009). Such regulatory domains can be very large and can encompass several genes of which one is the target gene of the enhancers as this was revealed by testing of human and zebrafish non-coding sequences in transgenic zebrafish (Navratilova et al. 2009, 2010; Fredman et al. 2009). Defined as genomic regulatory block (GRB), this contains the target gene, unrelated non-target genes and/or a gene desert harboring multiple regulatory sequences (Kikuta et al. 2007a, b; Becker and Lenhard 2007). A breakage within this region would disturb the regulation of the target gene and consequently a GRB is evolutionary conserved in all vertebrate genomes. About 300 of such large regulatory domains have been discovered, with the *Arx* gene locus being one of these (Engstrom et al. 2008).

A 41-kb tandem duplication of *ARX* and surrounding sequences was discovered in a patient with moderate non-syndromic XLID (Whibley et al. 2010). The upstream breakpoint in this patient is within the 5′untranslated UTR of *ARX*, leaving the duplicated copy without promoter. In consequence, the only genetic abnormality in this patient is a duplication of enhancers acting on *ARX*. A similar *ARX* duplication with a breakpoint within or near the 5′UTR of *ARX* has recently been described in an affected boy and his normal grandfather (Popovici et al. 2014). We identified four more ID patients with duplications including *ARX* and describe six cases that are available in DECIPHER database. The duplicated genomic regions vary in sizes and often include or break further genes. In order to shed light on the disturbed genetic mechanisms in these patients, we have analyzed the regulation of *ARX* in zebrafish.

Zebrafish and mammalian forebrains have many regional anatomical homologies. These were established by expression pattern similarities of evolutionary conserved genes, which function in the production of molecularly defined neural phenotypes both during development and adulthood (Osório et al. 2010; Mueller et al. 2011; Mueller 2012). For instance, regions such as the subpallium, the preoptic region, prethalamus and hypothalamus are defined by the activity of genes important for the production of GABAergic neurons (Mueller and Wullimann 2002; Filippi et al. 2014).

Using zebrafish transgenesis we established the long-range regulatory domain of *ARX* and identified enhancers

to characterize their brain region-specific activity. Three of the enhancers mediate autoregulatory mechanisms that are likely critical to maintain proper expression levels of the gene in specific regions of the developing forebrain. The transgenic GFP lines served as in vivo markers to establish loss of function phenotypes within the labeled regions of Arx activity. Remarkably, inhibition of *arxa* function in embryos of these lines has revealed that *arxa*/*ARX* disturbances have very subtle effects on neuronal outgrowth and guidance that likely affect the formation of neuronal circuits. Without suitable markers and tools such delicate defects are usually overlooked and will remain mostly undiscovered in patients.

Our nomenclature for human, zebrafish and pan-vertebrate orthologs is *ARX*, *arxa* and *Arx*, respectively.

With this study, we exemplify how genomic rearrangements and duplications can cause brain region-specific expression level changes that result in relatively subtle neurological disturbances finding their appearance in a neuropsychiatric phenotype.

Materials and methods

Patients

The screening protocols were approved by the appropriate institution review boards and informed consent was obtained from the parents of patients. Genomic DNA from patients and healthy controls was isolated from peripheral blood according to standard procedures and stored at 4 °C.

Array cGH

Genomic DNA from patient and control samples was extracted from blood collected using EDTA as the anticoagulant. High-resolution chromosomal microarray and analysis was conducted by contributing cytogenetic diagnostic laboratories. UCSC Genome Browser (GRCh37/hg19 build) Human Genome coordinates were used for the annotation of the copy number variants identified.

Zebrafish maintenance and staging

Zebrafish were maintained at 14 h light/10 h dark cycle at 28.7 °C. Embryos and larvae were staged according to established zebrafish protocols (Kimmel et al. 1995). Zebrafish maintenance and experiments had been in accordance with terms and conditions of the University of Sydney Animal Ethics Committee.

Computational analyses

The regulatory domain of *ARX* was defined by analyzing conserved gene order in vertebrate genomes. Synorth genome browser (Dong et al. 2009) was used to establish an overview, while the actual size and coordinates of the human synteny block was established by a method where smoothened HCNE densities were used to estimate the edges followed by visual inspection on Ancora browser (Engstrom et al. 2008). Conserved non-coding sequences were extracted using the following thresholds: human–mouse >96 % over 50 bp, human–mouse >90 % over 50 bp and human–frog >70 % over 50 bp. In addition, we determined the most conserved sequences in the *ARX* locus by utilizing blastn search with human (GRCh37) and mouse (mm9, chrX:89473103-91500000) genome sequences (Berman et al. 2000). We also compared to ENCODE data, as DNAse hypersensitive sites (DHS) and ChIP sequencing signals predict enhancers. However, almost all available datasets had been generated in non-neural cell lines, and most elements did not reveal specificities in public datasets other than DHS. Some were identified as enhancers through p300 ChIP sequencing and subsequent validation in the mouse (Vista Enhancer Browser; http://enhancer.lbl.gov/). The coordinates and lengths of all tested elements as well as reference data are listed in Supplement Table 1.

The JASPAR CORE database (http://jaspar.cgb.ki.se/) (Sandelin et al. 2004a) and ConSite (http://consite.genereg.net/cgi-bin/consite) (Sandelin et al. 2004b) were used to identify the putative transcription factor binding sites (TFBSs). The TFBSs were selected using human and mouse matrices with thresholds above 95 %.

Zebrafish reporter gene transgenesis and constructs

Sequences chosen for testing were PCR amplified from human genomic DNA. The fragments were cloned upstream of a *gata2* promoter into a Tol2–GFP reporter gene vector and purified construct DNA was injected into fertilized zebrafish eggs. Cloning procedure, microinjection, raising and screening of fish as well as imaging have been described in detail (Ishibashi et al. 2013). For each construct at least four positive founders were isolated to give rise to independent transgenic lines and F1 embryos were used for GFP expression pattern analyses. For the expression of Gal4, the GFP was exchanged with a 'Kal4' (Distel et al. 2009) in the Tol2 reporter gene construct. For overexpression of *arxa*, a Tol2 reporter gene construct was modified in two steps to co-express the cDNA and a red fluorescent mCherry as a marker. First, a bidirectional 5 × UAS–E1b promoter sequence was inserted in between

mCherry and a Gateway® recombination cassette, second, an SV40 sequence was inserted upstream of the GW cassette. Zebrafish *arxa* cDNA was amplified from a cDNA clone (IMAGE:9038800, gene bank ID: BC163872) with following primer sequences: forward CGAGCACGACTGAGGACGAT, reverse GGATCAGCACACTTCTTTCCCT and then subcloned into pCR8 plasmid to create an entry vector. A clone containing *arxa* cDNA in reverse orientation was used for Gateway recombination into the Tol2 destination vector. The construct was injected into wild-type TAB fish and fish grown to adulthood were crossed with ARX_enhancer:Kal4 fish to be screened for red expression and to establish a F1 generation. F2 embryos resulting from incrossed F1 fish were used for analysis.

Whole mount anti-GFP staining

Antibody stainings were carried out according to a protocol described previously (Punnamoottil et al. 2008; Turner et al. 2014). Primary antibodies: rabbit polyclonal anti-GFP antibody (AMS Biotechnology cat # TP401) at 1:1000; mouse monoclonal anti acetylated tubulin (Sigma cat # T7451) 1:400. Secondary antibodies were used from Life Technologies: Alexa Fluor® 488 goat anti-rabbit (A-11034) and Alexa Fluor® 594 Chicken Anti-Mouse (A-21201). Cell nuclei were visualized using DAPI (10 µg/ml in PBT at room temperature for 1 h).

Whole mount in situ hybridization

Embryos were fixed in 4 % paraformaldehyde overnight at 4 °C, then subsequently dehydrated in a series of methanol/PBT (0.1 % Tween-20) and then stored at −20 °C. The hybridization followed a standard zebrafish protocol to be stained by an alkaline phosphatase-mediated NBT/BCIP staining reaction (Oxtoby and Jowett 1993). Sectioning was described in Punnamoottil et al. (2008). The DNA template for *arxa* probe synthesis was kindly provided by the Kitamura laboratory and linearized with *Sac*I to be transcribed with T3 polymerase (Miura et al. 1997). DIG-labeled RNA probe was used at a concentration of 1 ng/µl hybridization buffer.

Fluorescent in situ protocol was modified from Machluf and Levkowitz (2011). Hybridization solution contained 2 % dextran sulphate and probe was used at 2-4 ng/ml. Probe was detected using DIG POD antibody (Roche 11207733910) and color was developed using Tyramide Signal Amplification (TSA) kit#25 (Molecular Probes by Life Technologies # T-20935). For co-labeling with GFP antibody, the embryos were washed 5 times for 15 min with PBT (0.15 % Triton-X) and processed according to standard immunostaining protocol as described above. Anatomical nomenclature followed the 'Atlas of early zebrafish brain development' by Mueller and Wullimann (2005).

Morpholino injections

Arxa translation was inhibited using two morpholinos (MOs) that had been described in our previous study (Djiotsa et al. 2012). The first targets the intron 2–exon 2 splice junction (5′GCGTCATATTTACCTGGTGAACACA) and the second targets the ATG (5′TCGTCGTCGTACTGACTGCTCATGA). The MOs were diluted with 0.05 % phenol red and 2 ng of the splice-MO compared to 3 ng of the ATG-MO were injected into the yolk at the 1-cell stage.

Morpholino information for *dlx2* genes is available through zfin.org (MO1-dlx2a, MO1-dlx2b; Jackman and Stock 2006) and target the 5UTR/ATG. *dlx2a*: 5-TGAGGCTGTCAAAAACTCCAGTCAT-3; *Dlx2b*: 5-GCTGTACTTCCTAACAGTTAATAGT-3. The Morpholinos had been prepared as 250 µM solution and were co-injected in the amount of 7.6 ng (3.8 ng each).

For injection, zebrafish eggs were collected from a transgenic parent outcrossed against wild-type fish. 100 fertilized eggs were injected with the antisense-morpholino and another 100 with a standard-control-Morpholino, further 100 were kept non-injected. Embryos were analyzed at 2 and 3 dpf by fluorescent microscopy using an inverted microscope (Leica DMIL LED). 15–20 embryos each batch were embedded into 2 % methylcellulose in water and imaged. Ten of these were selected randomly for fluorescent intensity measurements by confocal microscopy on day 3.

Confocal imaging and quantification

Imaging was performed with a Zeiss LSM 710 confocal microscope using Zen software. Embryos were mounted in 1–1.5 % low melting point agarose (Sigma A9414) in a glass-bottom petri dish. For fluorescent intensity measurements each embryo was quickly scanned to define the appropriate parameters, which allowed signal acquisition for all samples without pixel saturation. All embryos (*arxa* inhibited and controls) were then scanned using the same parameters including laser intensity, pinhole, detector gain, scan speed, temperature and objective. An average of 90 slices per embryos were generated. Integrated fluorescent density minus background was then evaluated for each projection using image J and analyzed using Microsoft Excel. Data obtained were compared using a *t* test.

Adult brain analyses

In situ hybridization

3- to 9-month-old wild-type fish (AB strain) were culled in ice-water, decapitated and the dissected brains were fixed in 4 % formaldehyde (wt/vol) overnight at 4 °C. Brains were dehydrated in ascending MeOH series (25–100 %, 5 min

each) and kept at −20 °C until use. The brains were rehydrated in descending MeOH series and were permeabilized by a 30-min treatment with proteinase K (10 μg/ml; P6556, Sigma) and prehybridized for 3 h at 65 °C. Hybridization was performed at 65 °C for 18 h in hybridization buffer containing DIG-labeled *arxa* probe (Miura et al. 1997) at a concentration of 2 ng/μl. After hybridization, the brains were embedded in 3 % agarose and 80-μm-thick cross sections were cut using a vibratome. The sections were blocked in blocking buffer (2 % normal goat serum, 2 mg/ml bovine serum albumin), incubated with anti-DIG AP Fab fragments (sheep, Roche, 1:5000) and the signal was developed with NBT/BCIP. For double in situ hybridization/immunohistochemistry staining on transgenic brains, in situ hybridization was performed first and developed with SIGMAFAST Fast red/TR Naphthol (Sigma), and then the sections were rinsed in PBT and directly processed for immunohistochemistry.

Immunohistochemistry on free-floating sections

50-μm-thick vibratome sections were blocked in blocking buffer (10 % goat serum, 0.5 % Triton X-100 in phosphate-buffered saline) for 1 h at RT. The following primary antibodies were used: GFP (chicken, 1:500, Aves Laboratories), HuC/D (mouse, 1:600, Invitrogen), S100β (rabbit, 1:2000, Dako) and TH (mouse, 1:500, Chemicon MAB318, recognizes TH1) diluted in blocking buffer. Secondary antibodies raised in goat coupled to AlexaFluor dyes (Invitrogen) were used in 1:1000 dilutions for 2 h at RT. At least two brains were analyzed for each transgenic line.

Image acquisition

Pictures of the ISH on sections were taken on a Nikon AZ100 microscope equipped with a Nikon DS Ri1 camera. Fluorescent images were taken on a Zeiss LSM700 confocal microscope using 20× air, 40× oil or 63× oil immersion objectives. Images were processed using the ZEN software (Zeiss). Composite images were automatically stitched upon acquisition using 'Tilescan' mode on the Zeiss ZEN software. Anatomical nomenclature followed Wullimann et al. 1996 (Neuroanatomy of the Zebrafish brain, A topological atlas).

Results

Genomic dosage imbalances in patients with XLID

Patients with various neurodevelopmental problems, intellectual disability and autism were identified to have duplications encompassing the *ARX* gene locus. Table 1 lists the patients (P) in the order as they appear in Fig. 1 and provides specificities. P505 is a previously reported patient, who displays moderate non-syndromic intellectual disability and has a 41-kb tandem duplication that includes *ARX* and part of *POLA1* (Whibley et al. 2010). P1 presents with speech delay mild intellectual disability and hypotonia, but has a 400-kb duplication with breakage of *POLA1*. P2, the mother of P1 has only mild learning difficulties. P3 and his sister P4 have poor growth and delayed speech. Their maternally inherited duplication encompasses all of *POLA1* and 19 kb sequence upstream of *ARX*, which is in antisense direction. DECIPHER has recorded six further male patients (DP1–5), who presented with phenotypes including intellectual disability with speech and/or language delay. Other comorbidities include autism, ADHD and poor growth. All duplications encompass *ARX* and have breakpoints in *POLA1*, except DP1 and DP2, with a full duplication of *POLA1* and disruption of *PCYT1B*.

One more case was mapped, P5, who presented with autism and mild developmental delay. The 0.549 Mb duplication is approximately 100 kb downstream of *ARX* and includes *PCYT1B*, *PDK3*, while *POLA1* appears disrupted. In contrast, DECIPHER patient DP6, who inherited from a normal parent a 9-kb small deletion 70 kb upstream of *ARX*, does not present any particular phenotype.

Although some of the duplications are larger and include duplications of *POLA1* (P3, P4, DP1, DP2) as well as breakage of *POLA1* (P1/P2, DP3, DP4, DP5) and *PCYT1B* (DP1, DP2), these genes are not associated with central nervous system disorders.

The long-range regulatory domain of *ARX* and search for enhancers

Conserved synteny of human *ARX* and surrounding HCNEs with the zebrafish genome was used as criterion to identify the regulatory domain of the gene. *ARX* is in conserved gene order with *PDK3–PCYT1B–POLA1* in most vertebrates. In zebrafish the synteny block is split into *pdk3a–pcyt1ba* and *pola1–arxa* [illustrated in Synorth genome browser (Dong et al. 2009)], suggesting that the majority of the enhancers regulating *ARX* are within the *POLA1–ARX* region. This locus contains 88 sequences with more than 96 % conservation over 50 bp between human and mouse (red bars in Fig. 2a). The density of such highly conserved non-coding sequences (HCNEs) is highest around *ARX* and extends into *POLA1*. Fourteen of these sequences with highest conservation score in blastn searches and a further 12 sequences conserved more than 70 % over 50 bp between human and frog genomes (green bars in Fig. 2a) were tested for enhancer function by reporter gene transgenesis in zebrafish (hs1–hs26, coordinates and primers listed in Suppl. Table 1). In addition, we chose

Table 1 CNVs in Xp22.11 in *ARX* long-range regulatory domain

Global	Gender	Clinical description	Inheritance	Gain	Variant coordinates (hg19) and size	Genes involved	Duplicated enhancers	Other chromosomal abnormalities/comments	References
P505	Male	Moderate non-syndromic intellectual disability	Familial	GAIN of 41 kb	chrX:24992914–25033979	**ARX**	hs8a, hs8b, hs9		Whibley et al. (2010)
P1	Male	Mild intellectual disability, speech delay and hypotonia	Familial	GAIN of 0.4 Mb	chrX:24887676–25325777	**ARX,** POLA1	hs2, hs8a, hs8b, hs9	–	This study
P2	Female	Learning difficulties	Familial (mother of P1)	GAIN of 0.4 Mb	chrX:24887676–25325777	**ARX,** POLA1	hs2, hs8a, hs8b, hs9	–	This study
P3	Male	Developmental delay, growth retardation, delayed speech	Familial	GAIN of 0.377 Mb	chrX: 24677441–25054698	**ARX**, POLA1,	hs2, hs8a, hs8b, hs9	GAIN of Xp11.4 region spanning a min 0.071 Mb and max 0.161 Mb. Duplicates part of TSPAN7 gene, familial, unclear clinical significance	This study
P4	Female	Short stature and speech delay	Familial (sister of P3)	Gain of 0.377 Mb	chrX: 24677441–25054698	**ARX**, POLA1,	hs2, hs8a, hs8b, hs9	–	This study
Decipher 253498 DP1	Male	Severe intellectual disability, microphthalmia, growth retardation, bisuspic aortic valve No phenotype reported in Decipher	Familial	Gain of 0.45 Mb	chrX:24593306–25033815	**ARX**, POLA1, PCYT1B	hs2, hs8a, hs8b, hs9		Popovici et al. (2014) (P2)
Decipher P265145 DP2	Male	Autism, hyperactivity, delayed speech and language development, ADHD, other physical abnormalities	Unknown	Gain of 0.58 Mb	chrX:24650157–25230368	**ARX**, POLA1, PCYTB1	hs2, hs8a, hs8b, hs9	–	Popovici et al. (2014) (P4)

Table 1 continued

Global	Gender	Clinical description	Inheritance	Gain	Variant coordinates (hg19) and size	Genes involved	Duplicated enhancers	Other chromosomal abnormalities/comments	References
Decipher 266096 DP3	Male	Generalized hypotonia, microcephaly, other physical abnormalities	Unknown	Gain of 0.81 Mb	chrX:24741372–25554818	**ARX**, POLA1	hs2, hs8a, hs8b, hs9, hs13	–	
Decipher P250183 DP4	Male	Delayed speech and language development, psychiatric abnormalities, hyperactivity, intellectual disability, dysarthria	De novo	Gain of 0.72 Mb	chrX:24815107–25542728	**ARX**, POLA1	hs2, hs8a, hs8b, hs9, hs13	Gain of Xq25 of 0.41 Mb including STAG2 and XIAP—inherited from normal parent—unknown pathogenicity	Popovici et al. (2014) (P5)
Decipher P277835 DP5	Male	Moderate intellectual disability	Familial	Gain of 0.3 Mb	chrX:24861601–25163763	**ARX**, POLA1	hs2, hs8a, hs8b, hs9	–	
P5	Male	Autism and mild developmental delay	Unknown	Gain of 0.549 Mb	ChrX:24351258–24900393	POLA1, PCYT1B, PDK3	hs2	GAIN of 3q11.1q11.2 of 0.929, incl. PROS1 and ARL13B. No clear clinical relevance for duplication of these genes. GAIN 4q35.2 of 0.597 Mb; GAIN 11.2q12 of 0.439 Mb. No disease associated with gains	This study
Decipher P267095 DP6	Male	No phenotype recorded in Decipher	Familial	Loss of 0.09 Mb	Xp21.3	–	–	–	

Patients (P) are listed in the order as they appear in Fig. 1. Clinical description, duplication details and other specificities are provided

two sequences that were identified as poised enhancers in human embryonic stem cells with the prediction that these would be active in differentiated cells and tissues (Rada-Iglesias et al. 2011). hs8 was separated into hs8a and hs8b.

Mouse orthologous sequences of hs2, hs3, hs8a and hs8b were found to bind enhancer-associated protein p300 in the developing forebrain (Visel et al. 2009).

In our assay 19 sequences drove GFP expression in defined tissues in at least two out of four independent transgenic zebrafish lines (Fig. 2b, c; Suppl. Table 1). As defined by the major *arxa* expression domains in zebrafish (telencephalon, diencephalon and floor plate), we further established if these regulate *ARX* specifically (Fig. 2c). Sequences hs2, hs8a, hs8b, hs9 and hs13 had specific activity and were identified as enhancers of *ARX* (Fig. 2d, e). The other sequences may influence *ARX* gene regulation, but were not conclusive.

Brain region-specific enhancer activity of *ARX*

We found that zebrafish *arxa* is strongly expressed in GABAergic forebrain regions, i.e., the subpallium, the preoptic region, the ventral thalamus, the (caudal) hypothalamus and the ventral posterior tubercular region (Fig. 2e; Suppl. Figure 1). Weak expression was also detected in the developing dorsal telencephalon (pallium) and the thalamic eminence, both of which at this stage are thought to consist of primarily glutamatergic neurons. These expression domains are highly consistent with findings in mice (Colombo et al. 2004), suggesting that *arxa* and *Arx* have a conserved role in GABAergic neuron specification and migration from fish to mammals. In addition *arxa* and *Arx* may function in the development of glutamatergic territories such as the thalamic eminence and pallium. Figure 2f illustrates the gross similarities between the zebrafish and the mammalian telencephalon (modified from Folgueira et al. (2012).

Fig. 2 Mapping of *ARX* cis-regulatory activity in the developing ▶ zebrafish brain. **a** Custom-modified UCSC genome browser window comprising 1.86 Mb of the human X-chromosome (hg19). The *ARX* transcription factor gene is located downstream of *POLA1* and transcribed in antisense direction. The *colored little bars* represent non-coding sequences conserved in the genomes of mouse (*dark* and *bright red*), opossum (*blue*), frog (*green*), and zebrafish (*dark blue*). The *red hill above* reflects their density. Highly conserved non-coding element (HCNE) density is highest around *ARX* and *POLA1*. The *grey bars at the bottom* show elements that we tested in zebrafish. **b** The expanded window of 623 kb shows detailed location of tested sequences (excluding hs16, which lies almost 300 kb downstream of hs13) that were taken into analysis. Elements were categorized according to their regulatory activity. *Red-colored* enhancers acted highly consistent in isolation from the genomic locus, whereas *green-colored* sequences expressed variable activity. **c** Percentage of recurrence of GFP reporter gene expression in *arx* transcript positive tissues in at least four different transgenic zebrafish lines for each of the tested sequence. **d** Representative images of transgenic 2.5 dpf embryos with *ARX* brain enhancers regulated GFP. *Arrows point* to the expression domains in a *color code* that corresponds to the anatomical descriptions in **e**. **e** Simplified zones of *arx* expression in 2–3 dpf zebrafish embryos in lateral view. **f** The telencephalon compared between zebrafish and mammals in illustrations of transverse sections. The *red-colored lines* point to different morphogenesis (eversion in zebrafish versus evagination in mammals). These schematics are modified after Folgueira et al. (2012)

The specific regulatory activities of the brain enhancers were analyzed in more detail by confocal imaging of transgenic embryos (Fig. 3a–e) and colocalization of the GFP with *arxa* mRNA transcripts was confirmed (Suppl. Figure 2). hs2 regulates a broad pattern within the GFP expression domain covering clusters of neurons in the dorsomedial and ventral telencephalon, the dorsolateral ventral thalamus, the preoptic region, the caudal hypothalamus, the midbrain basal plate and hindbrain floorplate (Fig. 3a). In contrast to hs2, hs8b regulates expression in

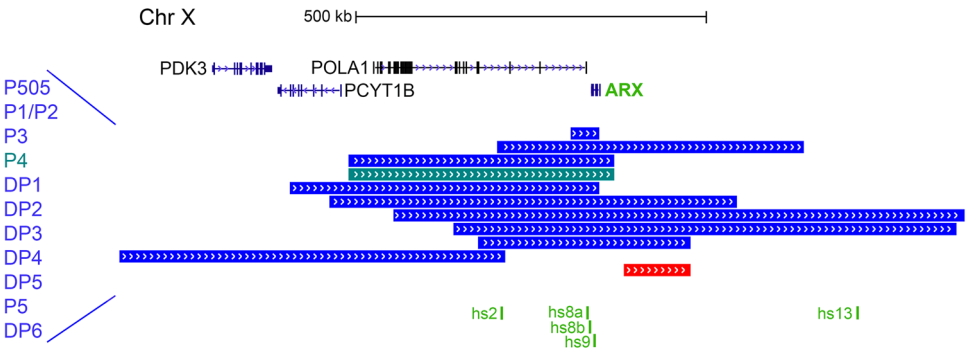
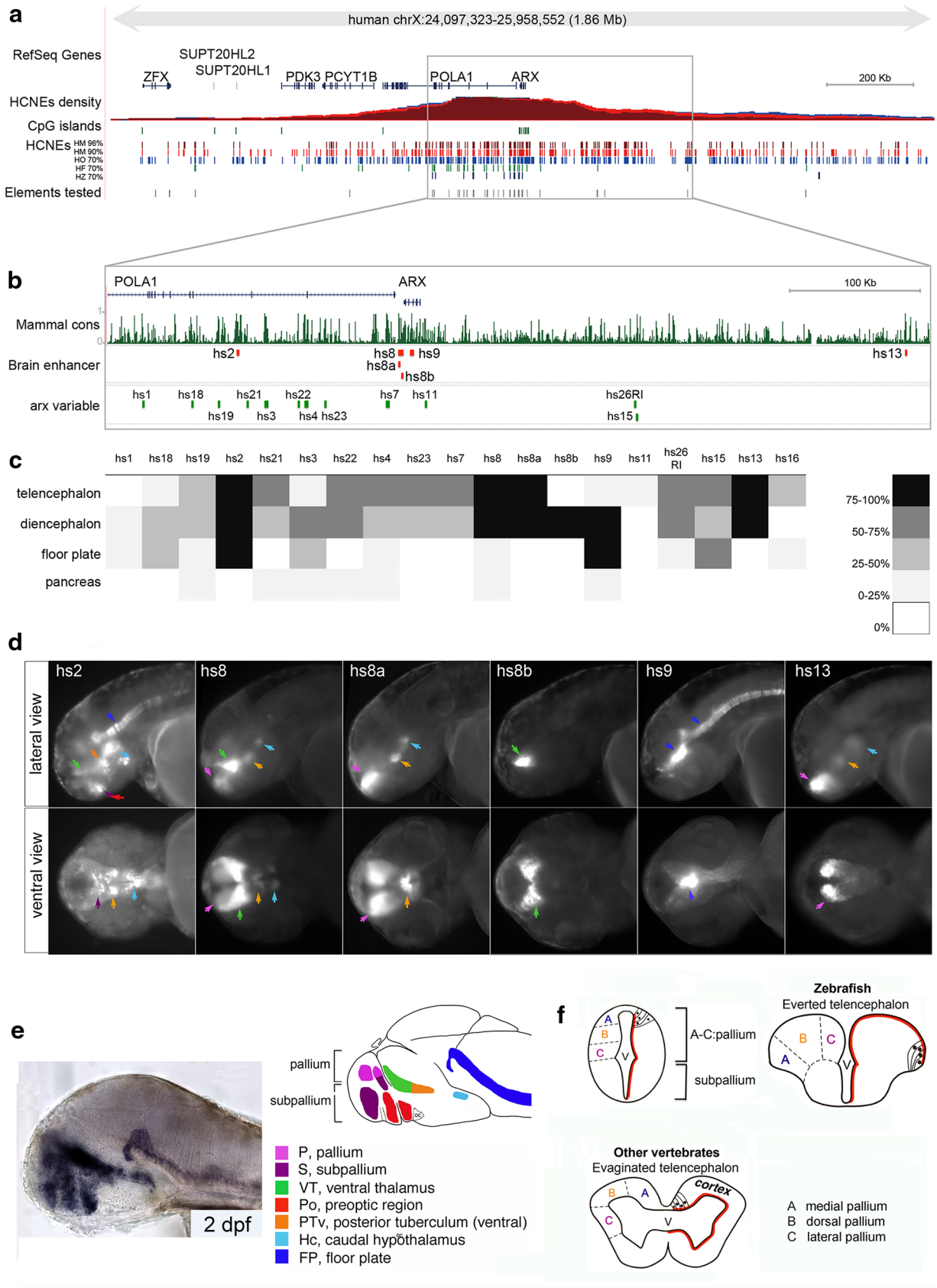


Fig. 1 Duplications of *ARX* and its regulatory domain in patients with ID. UCSC browser (hg19) window with view on *ARX*, flanking UCSC genes and genomic rearrangements in patients (P) 505 and one to five and six additional cases identified from DECIPHER (DP1-6). The *blue bars* indicate the extent of the duplications in males, and *teal bar* in a female, with the *red bar* indicating a deletion in a male. The location of *ARX* enhancers as identified in this study are indicated below


a
human chrX:24,097,323-25,958,552 (1.86 Mb)
RefSeq Genes
SUPT20HL2
SUPT20HL1
ZFX
PDK3 PCYT1B
POLA1
ARX
200 Kb
HCNEs density
CpG islands
HCNEs HM 96%
HM 90%
HO 70%
HF 70%
HZ 70%
Elements tested
b
POLA1
ARX
100 Kb
Mammal cons
Brain enhancer
arx variable
hs2
hs8
hs9
hs8a
hs8b
hs13
hs1
hs18
hs21
hs22
hs7
hs11
hs26RI
hs19
hs3
hs4 hs23
hs15
c
hs1 hs18 hs19 hs2 hs21 hs3 hs22 hs4 hs23 hs7 hs8 hs8a hs8b hs9 hs11 hs26 RI hs15 hs13 hs16
telencephalon
diencephalon
floor plate
pancreas
75-100%
50-75%
25-50%
0-25%
0%
d
lateral view
ventral view
hs2
hs8
hs8a
hs8b
hs9
hs13
e
2 dpf
pallium
subpallium
P, pallium
S, subpallium
VT, ventral thalamus
Po, preoptic region
PTv, posterior tuberculum (ventral)
Hc, caudal hypothalamus
FP, floor plate
f
A
B
C
V
A-C:pallium
subpallium
Zebrafish
Everted telencephalon
Other vertebrates
Evaginated telencephalon
cortex
A medial pallium
B dorsal pallium
C lateral pallium

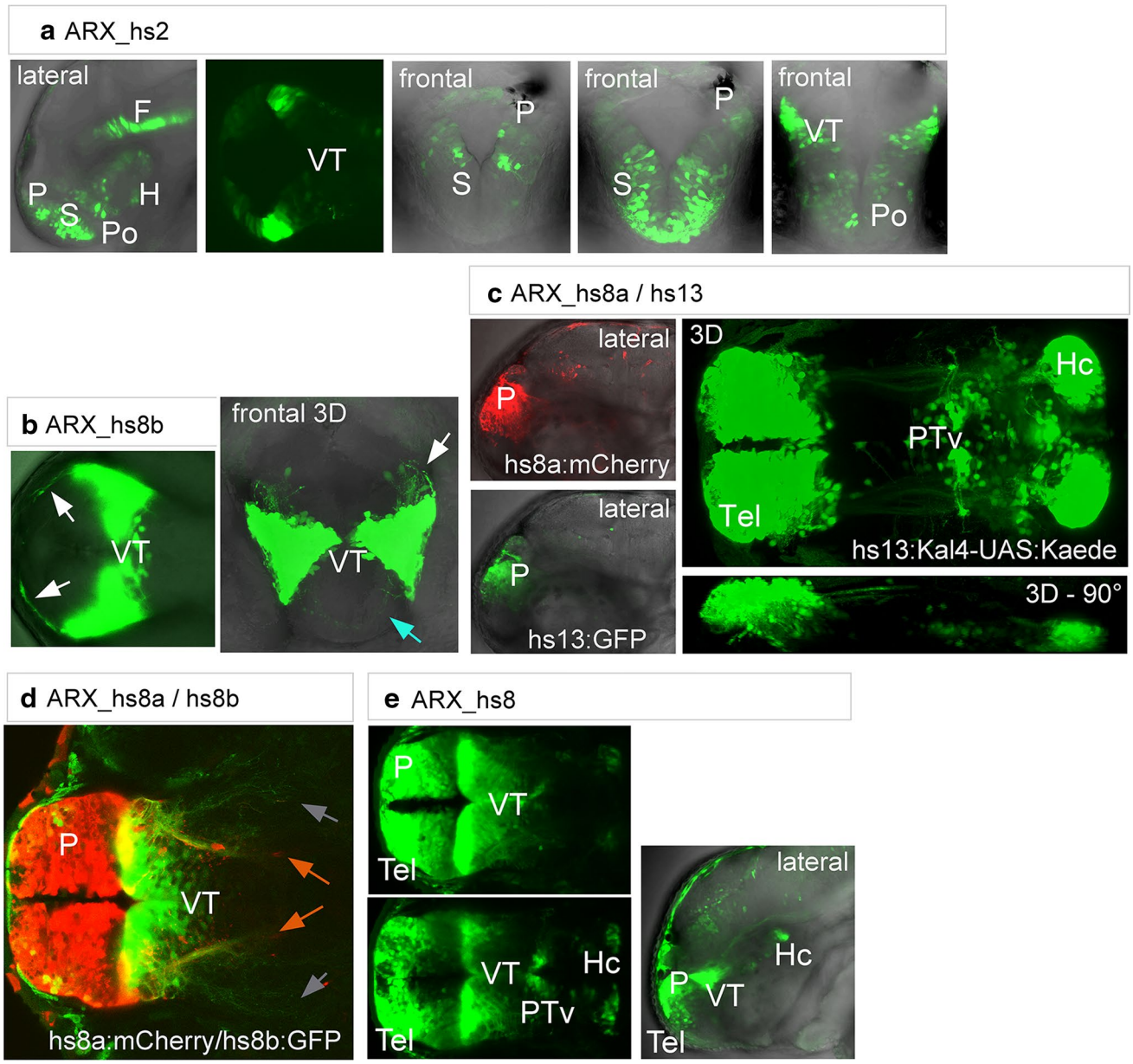


Fig. 3 *ARX*-enhancer regulated reporter expression in the brain. Confocal microscopy images of representative *ARX*-enhancer regulated fluorescent reporter expression. Horizontal sections of 2 and 3 dpf embryos are shown if not labeled otherwise. **a** hs2:GFP embryos at 2 dpf in three orientations. The horizontal optical section is comparable to the horizontal section in **b** and shows that hs2 regulated GFP is only in the dorsolateral region of the VT, whereas hs8b labels all VT. The series of transverse sections leads from anterior to posterior through the forebrain and provides a frontal view. **b** hs8b regulated GFP in the ventral thalamus (VT) at 2 dpf. *White arrows* point to the habenular commissure, the *green arrow* marks fibers that project ventrally. **c** Enhancers hs8a and hs13 drive very similar expression in the dorsal telencephalon at 3 dpf, images of a double transgenic embryo. hs13 regulated Gal4-UAS:Kaede shown as 3D image with 90° rotation below, visualizes projections between the telencephalon and the ventral diencephalon. **d** Co-regulated hs8a:mCherry and hs8b:GFP at 3 dpf shows almost complementary activity. *Orange arrows point* to neuronal projections originating dorsolateral and projecting posterior, and *grey arrows* mark fibers originating medial and projecting dorsolateral. **e** Combined activity of hs8a and hs8b in element hs8. Transverse optical sections through the forebrain are at two levels with the *upper image* being further dorsal. *AC* anterior commissure, *F* floor plate, *H* hypothalamus, *Hc* caudate hypothalamus, *Hi* intermediate hypothalamus, *OT* optic tectum, *P* pallium, *Po* preoptic region, *S* subpallium, *Tel* telencephalon (dorsal), *VT* ventral thalamus

the whole ventral thalamus and the GFP distributes in projections leading into the habenular commissure (Fig. 3b) (Wilson et al. 1990; Barresi et al. 2005). hs8a has regulatory functions in the dorsomedial telencephalon, the anterodorsal ventral thalamus, the ventral posterior tuberculum and the caudal hypothalamus (Fig. 3c). The

strong activity of hs8a in the dorsal telencephalon (i.e., pallium) was found in contrast to *arxa* expression, which is normally only weak in this region (compare to Fig. 2f; Suppl. Figure 1). Despite the small overlap in the anterodorsal ventral thalamus, hs8a and hs8b fragments have almost mutually exclusive regulatory activity (see double labeling image in Fig. 3d), which is combined when both sequences are tested together in one construct (Fig. 3e). Neuronal projections in hs8a and hs8b:GFP transgenic larvae were strongly labeled and were analyzed in detail (Suppl. Figure 3). hs13 drove GFP in a very similar pattern to hs8a in the dorsal telencephalon, posterior tuberculum and caudal hypothalamus at 3 dpf (Fig. 3c); however, at 1 dpf and at later stages the activity of the enhancers differed (Suppl. Figure 4). In summary, all enhancers have distinct, but also overlapping functions in regulating *ARX* expression in the developing forebrain.

ARX enhancer activity in the adult brain

arxa expression pattern was established in the adult zebrafish brain and the activity of the identified enhancers was compared. Only hs2 was found to act specifically in adults. *arxa* transcripts were detected in the dorsal nucleus of the ventral telencephalon (Fig. 4a, Vd), which corresponds to the dorsal subpallium and is equivalent to the striatum in mammals (Rink and Wullimann 2004). hs2 was found to be active in this region, but the expression was mosaic (Fig. 4a). Posterior of the Vd, *arxa* transcripts were detected in the supracommissural nucleus of the ventral telencephalic area and in the ventromedial thalamic nucleus (data not shown). *arxa* mRNA was further identified in the caudo-ventral diencephalon, including in the periventricular nucleus of the posterior tuberculum (TPp), where a small subpopulation of neurons was positive for tyrosine hydroxylase 1 (TH) (Fig. 4b). The posterior tuberculum contains dopaminergic neurons and its axonal projections into the dorsal subpallium had identified both structures as the equivalent of the nigro-striatal system in mammals (Rink and Wullimann 2001; Wullimann and Mueller 2004; Forlano and Bass 2011). hs2-driven GFP overlapped *arxa* mRNA in the TPp, but again was mosaic (Fig. 4b). GFP was also seen in areas where *arxa* transcripts were either below detection levels or absent: in the ventral nucleus of the ventral telencephalon (Vv), dorsally to neurons of the medial zone of dorsal telencephalon (Dm) and in the posterior zone of the dorsal telencephalon (Dp) (Fig. 4a).

Since hs2 extends its regulatory function into adult stages, it might be possible that this enhancer may control expression of *ARX* in processes relating to neuronal function or maintenance.

Negative autoregulation of *ARX* in the developing brain

Arx gene regulation is known to respond to Dlx2 (Colasante et al. 2008) and we wondered if ARX-transgenic lines would function as sensors to establish which of the enhancers would respond to dlx2. Morpholino antisense oligonucleotides (MO) were used to inhibit *dlx2a* and *dlx2b* translation and it was subsequently established whether the enhancers had altered activity as evaluated by increased or decreased GFP expression. No significant changes in GFP expression and on *arxa* transcript level could be detected. We next asked if identified *ARX* enhancers would mediate autoregulatory functions in the brain (Fig. 5a). The knockdown of *arxa* in hs9:GFP and hs13:GFP embryos had no significant effect on enhancer activity (Suppl. Figure 5a), while the GFP in hs2 and hs8a transgenic embryos was significantly upregulated and indicated that Arxa normally has a suppressive effect on these enhancers (Suppl. Figures 5b, c). Knockdown of *arxa* in line hs8b:GFP altered the GFP expression domain. The GFP disappeared from the medial ventral thalamus, while expression in the lateral ventral thalamus became stronger (red asterisks in Fig. 5a, d).

We next asked if a negative autoregulation was reflected at the *arxa* transcript level. In situ hybridization analysis of arxa morphants compared to controls revealed an accumulation of *arxa* transcripts in the developing brain (Fig. 5b). This identified domains of low-level transcription in wild-type embryos (Fig. 5b). One example is the normally very weak expression of *arxa* in the pallium of wild-type embryos (compare section 1 in Fig. 5c to section 1 in Suppl. Figure 1). Another example is the telencephalic migrated area that normally had only very faint staining (compare section 2 in Fig. 5c to sections 2, 3 in Suppl. Figure 1).

In summary, we have discovered a negative autoregulatory feedback mechanism of *ARX* that is mediated through hs2, hs8a and hs8b enhancers. In addition, hs8b may require activation through ARX in the medial ventral thalamus.

Brain region-specific positive autoregulation of *ARX*

In order to test a putative positive feedback mechanism in *ARX* regulation, we overexpressed zebrafish *arxa* cDNA using hs8a, hs8b and hs13 enhancers. In this experimental design increased fluorescent expression should be seen when Arxa transcription factor would signal back to the enhancer (either directly or indirectly). Transgenic lines in which hs8a, hs8b and hs13 drive a Gal4 were crossed into transgenic lines carrying a bidirectional UAS:arxa/:mCherry cassette. F2 larvae expressed mCherry and *arxa* cDNA in enhancer-specific brain regions and were compared to specimen not overexpressing *arxa* (Fig. 6).

A definitive effect of Arxa on hs13 could not be established (data not shown). hs8a-driven overexpression of *arxa* increased activation of the enhancer in the medial ventral thalamus (Fig. 6a), which is a strong regulatory domain only at 1 dpf (Suppl. Figure 4). This was in contrast to hs8a activity in the telencephalon, which appeared unaltered or perhaps weaker. hs8b responded to increased Arxa with enhanced activity in the medial part of the ventral thalamus where neurons projecting laterally were strongly labeled (Fig. 6b). The complex positive and negative autoregulation through hs8a and hs8b is illustrated in Fig. 6c. Of note, hs8a-driven overexpression of *arxa* in the telencephalon caused variable malformations becoming obvious by displacement of the eyes towards the midline (Suppl. Figure 6).

We concluded from our experiments that *ARX* is positively autoregulated in the ventral thalamus, which occurs at sequences that are within hs8a and hs8b enhancers.

Knockdown of *arxa* disturbs ventral thalamic and telencephalic development

Abrogated *arxa* function led to disturbances in ventral forebrain development as visualized by hs8a and hs8b-driven GFP. GFP served as marker for neurons with *arxa* identity and since this distributes uniformly in the cell, neuronal projections are also visible as described in Suppl. Figure 3a–d. Arxa antisense morpholinos were injected into fertilized eggs of hs8a and hs8b transgenic lines and GFP-labeled neuronal projections were analyzed at 3 and 6 dpf and compared to those of control embryos (Fig. 7a, b). Prominently and as described above (Fig. 4), hs8b driven GFP was missing from the medial domain of the ventral thalamus (marked by the red asterisk in Fig. 7a). In contrast, the lateral expression domain was slightly expanded, which resulted in a more prominent fiber bundle originating in this region (orange arrows in Fig. 7a). Neuronal projections that originate in the medial ventral thalamus and that project into the tectum (blue arrows in Fig. 7a) were strongly reduced (blue asterisks). In addition to the phenotype observed in the ventral thalamus at 3 dpf, we detected increased fiber outgrowth towards the telencephalon in the morphants at 6 dpf (pink arrows in Fig. 7b) and caudally directed projections (orange arrows) appeared disorganized at that stage.

hs8a:GFP specimen label the anterior commissure, which connects the left and right hemispheres and crosses in the ventral telencephalon (Suppl. Figure 3). This appeared diffuse and with fewer projections in arxa knockdown specimen as shown in representative horizontal optical sections in Fig. 7c (yellow arrows). Using GFP-labeled anterior commissure and ventral posterior tuberculum as markers, the horizontal views also illustrate that the anterior–posterior axis of the ventral telencephalon and diencephalon was shortened (red and yellow bars in Fig. 7c). This points to more severe disturbances in the ventral forebrain after knockdown of *arxa*.

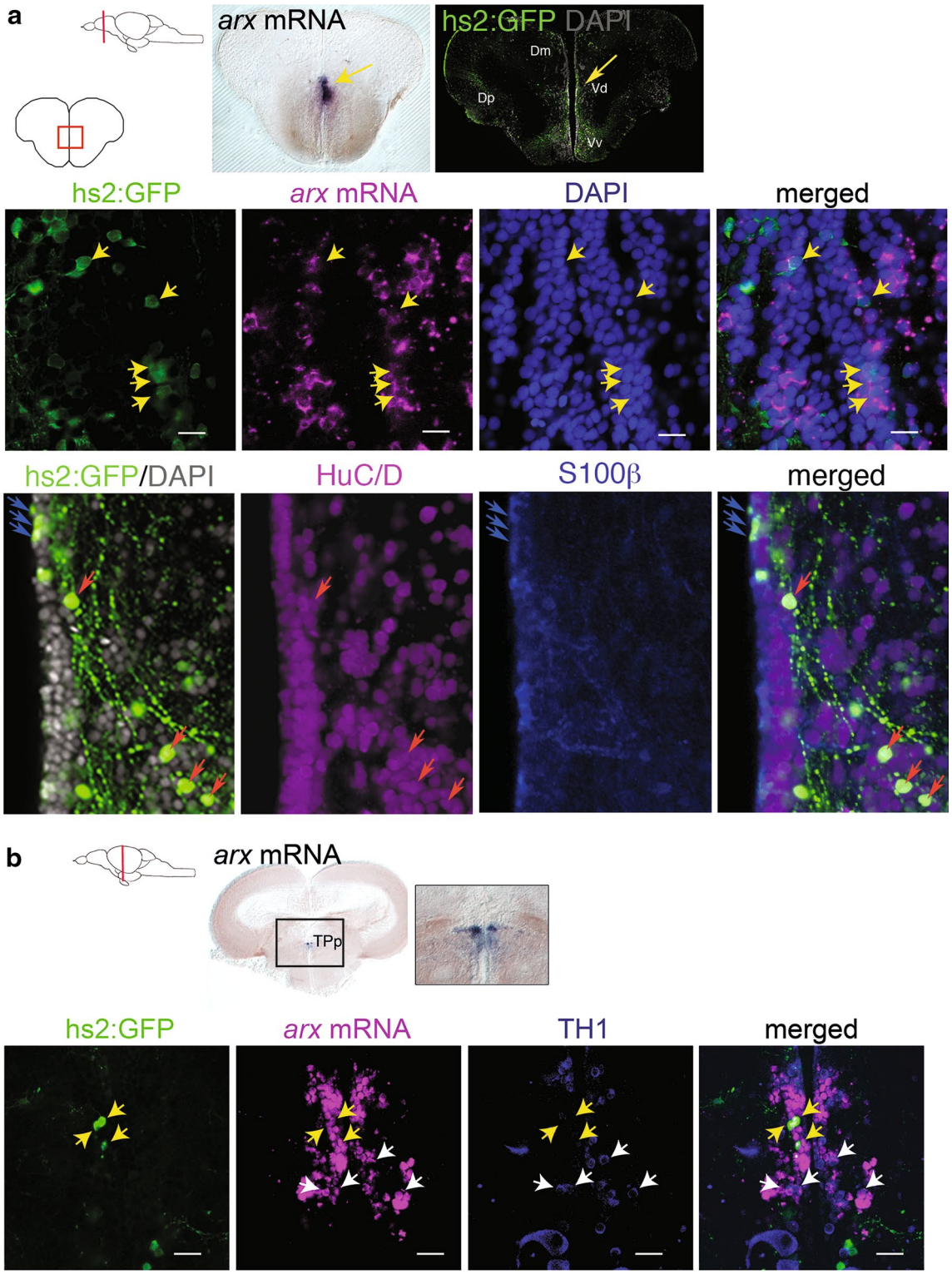
In conclusion, the specific reporter lines facilitated the detection of morphological and neuronal wiring defects in the ventral forebrain when *arxa* function is disturbed and will provide useful in vivo tools for future functional studies on *arxa*.

Fig. 4 Adult brain analysis. Expression of *arxa* as revealed by chromogenic in situ hybridization and expression of hs2-enhancer-regulated GFP in the adult zebrafish brain. Representative transverse sections document *arxa* mRNA distribution and are compared to confocal optical sections of immunostained hs2:GFP brains. The levels of sections are indicated in the schematics. **a** The *yellow arrow points* to *arxa* mRNA transcripts in the Vd. A corresponding section labels GFP (*green*) with a DAPI counterstain in hs2 transgenic line and indicates the localization of four nuclei that carry abbreviations (see below). The *rectangle in the schematic* indicates the photographed area shown in the magnified images below: *First row* an overlap of fluorescent *arxa* hybridization signal and hs2:GFP in Vd is marked by the *yellow arrows*; *second row* triple immunostaining for GFP (*green*), HuC/D (labeling neurons, *magenta*) and S100β (labeling radial glial cells, *blue*) shows that GFP mostly labels neurons (*red arrows*) and occasional radial glial cells (*blue arrows*). **b** Focus is on *arxa* expression domain in the periventricular posterior tuberculum. The *rectangle* indicates the region that is magnified in the images below. Images of a single confocal optical section show co-expression of hs2:GFP and *arxa* mRNA (marked by the *yellow arrows*). A small subpopulation of *arxa*-positive cells also expressed TH1 (*white arrows*). *Dm* medial zone of the dorsal telencephalon, *Dp* posterior zone of the dorsal telencephalon, *PTv* ventral posterior tuberculum, *TPp* periventricular nucleus of the posterior tuberculum, *Vd* dorsal nucleus of the ventral telencephalon, *Vv* ventral nucleus of the ventral telencephalon

Discussion

This study has identified several novel patients and collected information on existing patients with moderate intellectual disability presenting delay of speech development as their common symptoms, with all of them having genomic duplications that include the transcription factor gene *ARX*. The duplications vary in sizes and include further genes, of which none are correlated with brain diseases.

Having mapped the regulatory domain of *ARX* the duplicated genomic fragments, breakpoints and cis-regulation can now be put into context (for overview see Fig. 1 and Table 1 lists duplicated enhancers). Five *ARX* brain enhancers are located within 504 kb around the gene and reveal that this genomic region is highly functional and interactive and that duplications, dependent on the specific breakpoints, will disturb gene regulatory mechanisms by enhancer displacement. The identified enhancers are specifically active in the telencephalon, ventral thalamus, preoptic region, posterior tuberculum, caudal hypothalamus and floorplate with partially overlapping



functions. Overlapping enhancer functions are known to regulate genes with complex spatio-temporal expression patterns and it is thought that these are critical to fine tune transcript dose (Komisarczuk et al. 2009; McBride et al. 2011). The function of redundant enhancers is also to provide robust buffering in differing environmental conditions (Hong et al. 2008; Frankel et al. 2010; Perry et al. 2010).

The duplication of patient 505, which displays a non-syndromic intellectual disability, exemplifies critical *ARX*

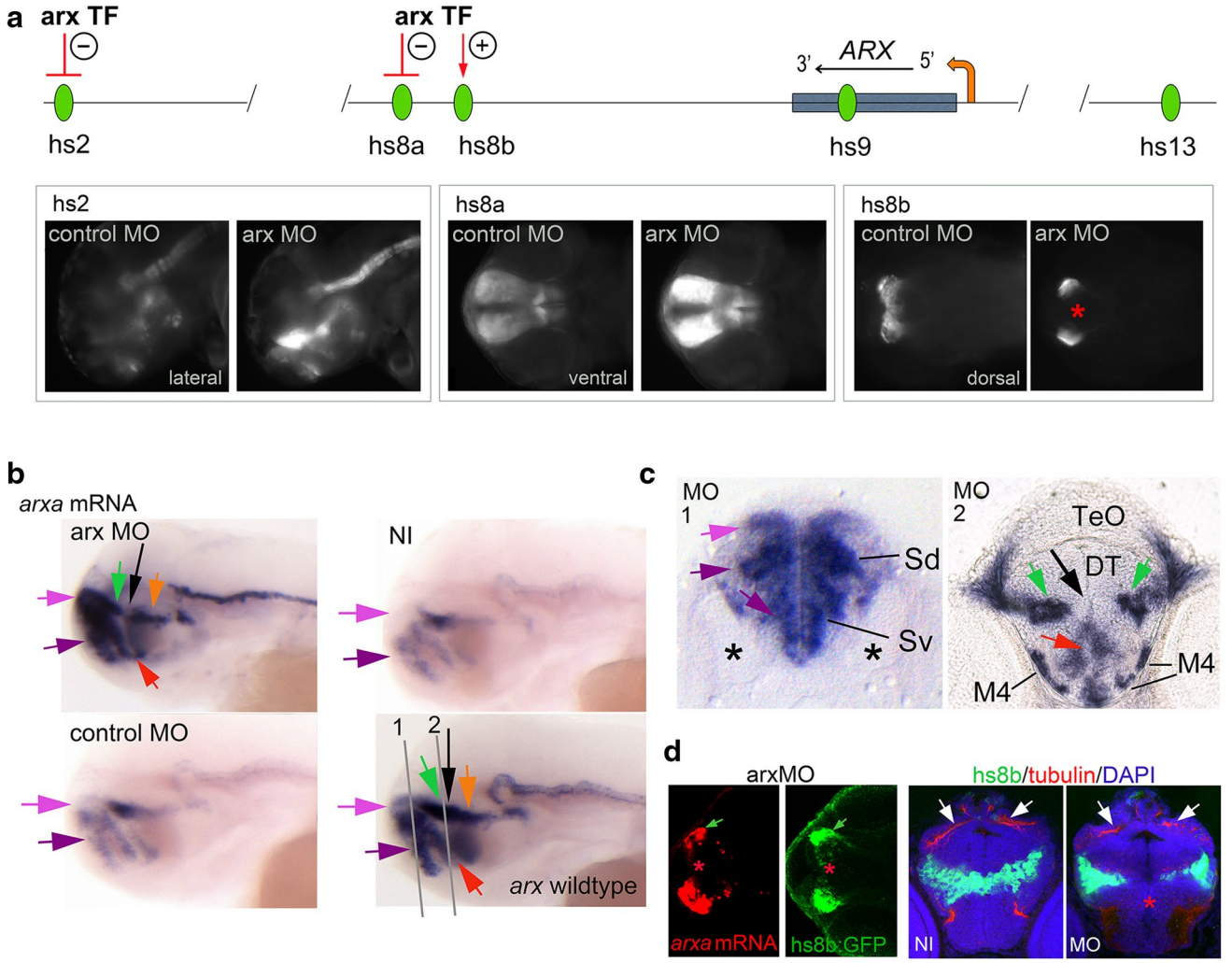


Fig. 5 *ARX* negative autoregulation in the forebrain. arxa function was inhibited in ARX_enhancer:GFP lines by injection of antisense morpholino oligonucleotides (MO). GFP intensity was used as a measure to assess the activity of the enhancers (**a**). **a** 2 dpf embryos are shown in different orientations. hs2 and hs8a regulated GFP was significantly enhanced in arxa-inhibited embryos. hs8b-regulated domain missed GFP in the medial part of the ventral thalamus (marked by a *red asterisk*). Differences in hs9 and hs13-regulated GFP expression could not be established. See also Suppl. Figures 5 and 6. **b** *arxa* in situ hybridizations of 3 dpf embryos. The regions expressing *arxa* are labeled with *arrows* in a color-code described in Figs. 1f and 7c. *Fine lines* carry further descriptions that use nomenclature established in the 'Atlas of the developing zebrafish brain', Mueller and Wullimann (2005) (Elsevier). arxa MO injected and control embryos (control MO and non-injected, NI) had same staining times, but the morphants developed much quicker a very strong signal indicating increased *arxa* transcript levels. A wild-type embryo with longer staining time was added to the panel for proper expression pattern comparison. Black arrows point to the medial ventral thalamus (VT) where transcripts were missing in the morphants. **c** Two transverse sections through arxa morphants provide further details. **d** The area in the medial VT that missed transcripts (around the proliferative zone flanking the 3rd ventricle, see in **c**) also missed hs8b regulated GFP (horizontal and transverse confocal optical sections). *DT* dorsal thalamus, *M4* telencephalic migrated area, *Sd* dorsal division of subpallium, *Sv* ventral division of subpallium, *TeO* tectum opticum

gene regulatory functions (Fig. 8). P505 has a relatively small tandem duplication including *ARX* and downstream enhancers (except hs2). We found the breakpoint in 5′untranslated UTR, which implicates that the promoter is missing from the duplicated copy. Consequently, the additional coding sequence cannot be transcribed, but the duplicated enhancers, in proximity to the original promoter, have most likely additional regulatory influence on *ARX*. In contrast, hs2, normally located 112 kb downstream of *ARX*, is brought further away and its regulatory influence is likely diminished. The gene regulatory deficits in this patient could be exacerbated by disturbed autoregulatory mechanisms as enhancers hs2, hs8a and hs8b respond differentially to decreased and increased Arxa levels in the zebrafish model (Figs. 5, 6; Suppl. Figure 5).

Decipher patient 1 (DP1), presenting with severe intellectual disability, could be exemplary for a duplicated

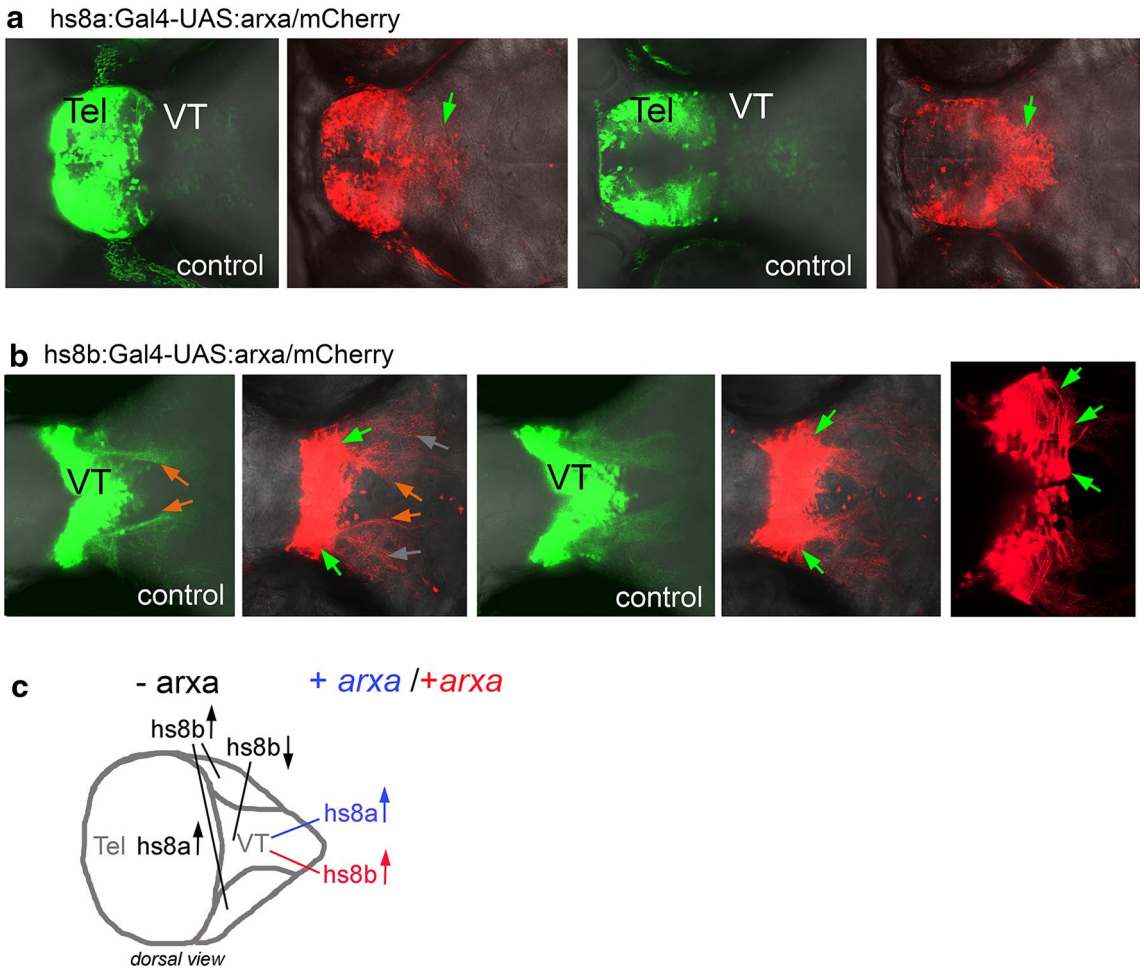


Fig. 6 *ARX* positive autoregulation in the ventral thalamus. Confocal imaging of larvae expressing *arxa* cDNA and co-regulated mCherry. Shown are representative transverse optical sections at two different levels from dorsal to further ventral. Control images are from UAS:Kaede expressing larvae and carry morphological descriptions. **a** hs8a-regulated overexpression of *arxa* in the dorsomedial telencephalon as well as weakly in the medial region of the ventral thalamus (VT). The weak activity of hs8a in the VT triggered a positive autoregulation that led to activation of the enhancer as indicated by strong expression of the mCherry (marked by the *green arrows*). **b** hs8b activated overexpression of *arxa* in the ventral thalamus. Neuronal projections leading dorsolateral (*grey arrows*) appeared increased. The expression domain appeared widened and changed the triangle into a square shape. This is due to increased labeling of fibers growing from medially located neurons towards lateral in the ventral half of the expression domain (pointed out by the *green arrows*). **c** Summary of *ARX* positive and negative autoregulatory mechanisms through hs8a and hs8b enhancers as established through overexpression of *arxa* and arxa morpholino inhibition. Basically, *ARX* enhancers respond brain region-specific to Arxa

ARX copy that may also lack the promoter (the inner breakpoint of the 14 kb mapped interval is in exon 1 at chrX:25,033,815), but since the breakpoint downstream of *ARX* is far beyond hs2, the gene regulatory domain of the original gene (as defined through the location of the five brain enhancers) appears non-disturbed. This patient was described in Popovici et al. (2014) as Patient 2, whereas his grandfather (Patient 3—duplication of unknown inheritance pattern) from whom he had inherited the duplication appeared normal. Another patient presented in this study (Patient 1) with a de novo *ARX* duplication covering a region 32 kb downstream of hs2 to 1.7 kb downstream of hs13 (chrX:24,861,402-25,398,496) also presented without ID. It has to be mentioned that the two patients presenting without ID were not tested for mosaicism (Popovici et al. 2014). Somatic mosaicism for CNVs has been underappreciated (Campbell et al. 2015), and in these individuals it could potentially explain why they are not affected. The absence of *ARX* duplications in the Database of Genomic Variants and subcellular mislocalization of overexpressed *ARX* (Shoubridge et al. 2010) as well as the effect of *ARX* overexpression on its known targets, *LMO1* and *SHOX2* (Shoubridge et al. 2012), suggest that high ARX protein dosage is likely detrimental to the cell. Alternatively, the non-ID phenotypes of patients with *ARX* duplications may point to a variable phenotype penetrance, which may also

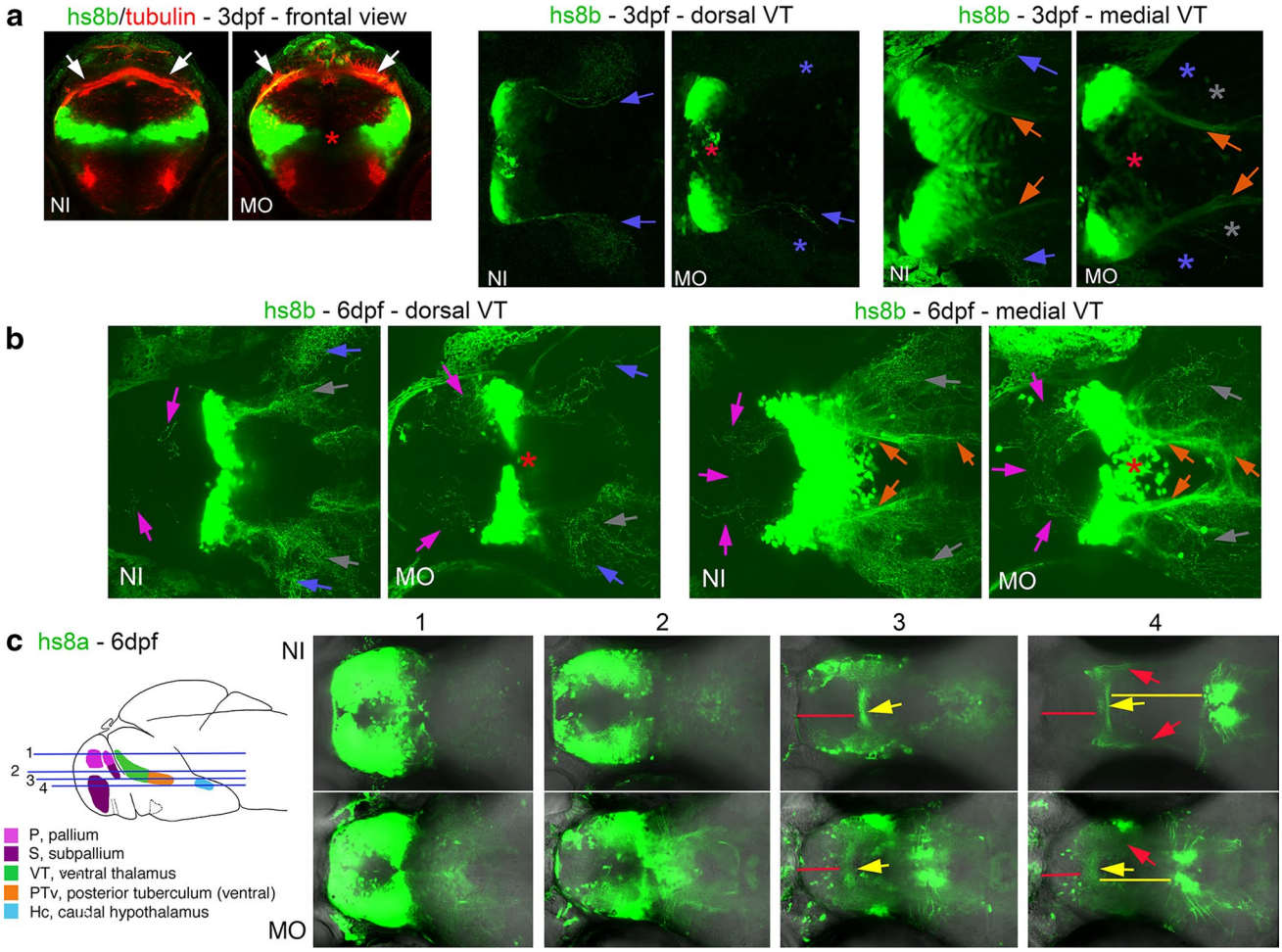


Fig. 7 Morphological changes as visualized by hs8b and hs8a regulated GFP. **a** hs8b regulated GFP-labeled VT. Transverse optical sections of anti-GFP/anti-acetylated tubulin co-stained embryos provide frontal views and show that the habenular commissure (*white arrows*) developed normal. Horizontal z-stack images lead through the dorsal and medial VT expression domain in live imaged specimen. Projections from VT into OT (*blue arrows*) were reduced and missing (*blue asterisks*). GFP appeared enhanced and expanded in the lateral domain with originating axonal projections leading posteriorly appearing stronger (*orange arrows*). *Grey asterisks* mark the area where VT projections towards dorsolateral were strongly reduced. **b** Horizontal z-stack images of live imaged larvae on two different levels of the VT expression domain. *Pink arrows* point to fine neuronal projections growing towards the telencephalon—these were increased as well as disorganized in the morphants, while neuronal projections projecting dorsolateral (*blue* and *grey arrows*) were reduced. Fiber bundles projecting posteriorly (*orange arrows*) had stronger label in the morphants and appeared misrouted. **c** hs8a regulated GFP labeled the telencephalon, posterior tuberculum and hypothalamus. Representative horizontal optical sections of 6 dpf larvae show that the anterior commissure appeared diffuse and weakened (marked by *yellow arrows*). Also the length of the ventral telencephalon (*red bars*) and the distance between the anterior commissure and the ventral posterior tuberculum (*yellow bars*) appeared reduced

be reflected in variable physical abnormalities that often accompany the ID ('syndromic ID'). In this study, we focus on ID to relate disturbances in a gene regulatory domain to a neuronal phenotype, but it has to be kept in mind that *ARX* is also active outside the nervous system. Expression studies in zebrafish and mouse show that *arxa*/*Arx* is for example expressed in the myotome, notochord/backbone, limbs and craniofacial skeleton (Miura et al. 1997; Collombat et al. 2003; Norton et al. 2005; Eurexpress Transcriptome Atlas, http://www.eurexpress.org). Since we have tested many more elements than we present here as enhancers, with many of these having activity for example in the fins and notochord, we included the screening results of the transgenic zebrafish lines in Suppl. Table 1 (for interpretation of screening results see Ishibashi et al. 2013). It would be too speculative at this point to invoke putatively disturbed cis-regulatory mechanisms in different 'syndromic' patients.

Morpholino antisense oligonucleotide-mediated inhibition of *arxa* translation served to identify autoregulatory mechanisms of the gene (Fig. 5). While morpholinos partially loose their activity starting at 3 dpf and sometimes

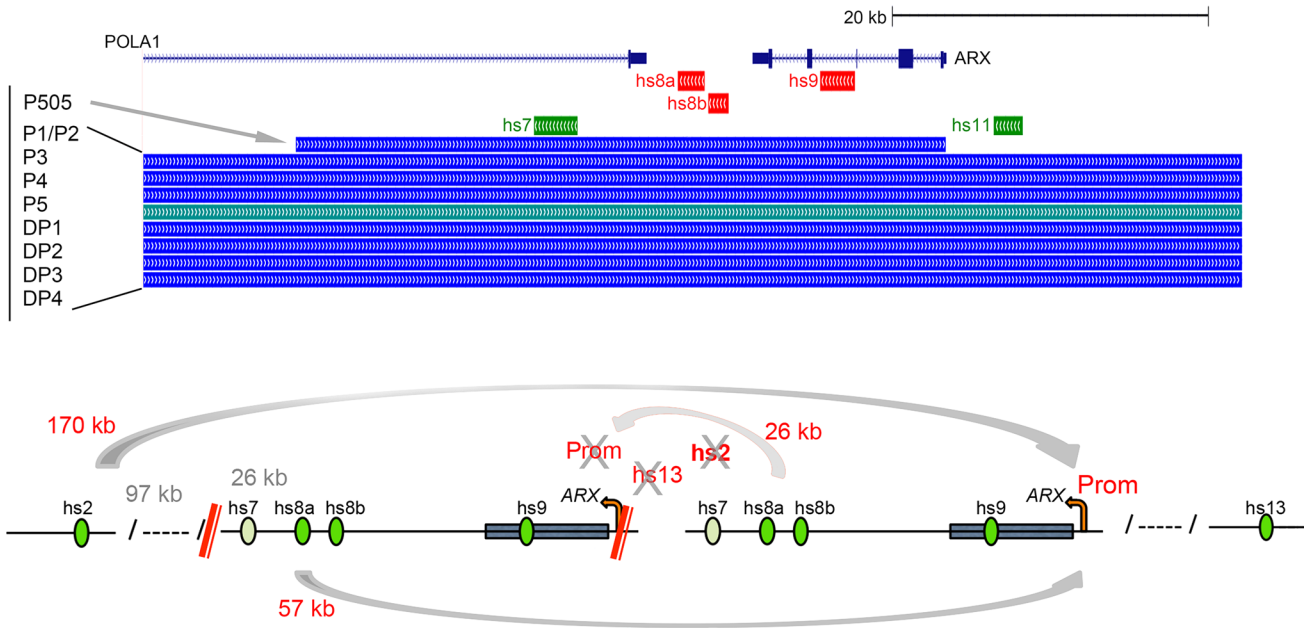


Fig. 8 Illustration of *ARX* gene regulatory abnormalities in P505. A zoom into the *POLA-ARX* region with focus on the 41 kb sequence that is tandem duplicated in P505, modified from Fig. 1. *ARX* coding sequence is in reverse orientation mirroring the genomic arrangement. Tested sequences appear in *red* (enhancers) and *green* (inconclusive cis-regulatory activity). hs7, approximately 9 kb distant of hs8a, was added to the illustration because this harbors an ultraconserved sequence carrying a mutation in a patient with autism and was identified as specific enhancer in a mouse model (Colasante et al. 2008; Fullston et al. 2011). Upstream breakpoint is at the transcription start site of *ARX* indicating that the duplicated copy misses a functional promoter. *Numbers in grey* give the distances between the breakpoint 3′ of the gene and hs2, hs7 respectively. Abnormal regulatory influences are indicated by *grey arrows with numbers in red* giving distances between enhancers and functional *ARX* promoter. Autoregulatory mechanisms in the telencephalon and ventral thalamus are not emphasized in this illustration. In essence, with the duplicated copy being non-functional, hs2 enhancer moved further away from the gene (170 kb instead of 112 kb) subsequently loosing influence on *ARX*, while duplicated enhancers hs7–hs9 have increased regulatory influence. Since DNA folding is unknown for this region, the model remains hypothetical

cause unspecific effects, we have compared *arxa* morpholino injected specimen to specimens injected with control morpholinos in all experiments. Hs8a and hs8b enhancers act negatively autoregulatory in the pallium and in the lateral ventral thalamus (summarized in Fig. 6c). While hs8a has strong GFP regulatory activity in the pallium, Arxa suppresses the activity of this enhancer, which could be a mechanism to reduce *arxa* transcript levels in this brain region (compare Fig. 5b, c and Suppl. Figure 1). From prevented transcription of *arxa* (Fig. 5c) and missing activation of hs8b-regulated GFP in the medial ventral thalamus of arxa morpholino injected specimen (Fig. 5d) we concluded that autoactivation of hs8b enhancer might be essential for *ARX* expression in the medial region of the ventral thalamus. This positive autoregulatory feedback loop was confirmed through hs8b enhancer_Gal4-UAS mediated overexpression of *arxa* in the zebrafish model (Fig. 6a, b).

To identify potential upstream factors and signaling pathways that regulate the enhancers, we have used transcription factor binding site (TFBS) prediction tools ConSite and Jasper (Sandelin et al. 2004a, b; see method section). Selected TFBSs that bind transcription factors with expression domains (Allen Brain Atlas, www.brain-map.org) overlapping that of *Arx* are presented in Suppl. Figure 7. For example, Sox5 and Sox17 signaling may act on hs2, hs9 and hs13 enhancers. While Sox5 is highly expressed in the developing cortex and the dorsomedial diencephalon, *Sox17* is not, but was found to control cell cycle exit of oligodendrocyte progenitor cells in the ventricular zone in the ventral telencephalon (ganglionic eminences; Sohn et al. 2006).

The autoregulatory mechanisms identified in this study suggest that an overall transcript increase, as expected in patients having the whole *ARX* regulatory region duplicated, would lead to brain region-specific imbalances in *ARX* expression. DECIPHER patients DP3 and DP4 have such duplications, while duplicated regions in the other patients lack hs13. Our experiments in the zebrafish model suggest that this would result in enhanced activation of *ARX* in the ventral thalamus, while expression in the cortex would be expected to decrease. In general, dependent on the distance of the enhancers to the breakpoints and the orientation of the duplicated fragments in the genome, enhancers could act on both gene copies thereby 'over'-regulating *ARX* brain region-specifically.

The autoregulatory mechanisms explain how a general increase of *ARX* transcripts would exacerbate expression level imbalances in the critical brain areas, including the ventral thalamus, which was previously not associated with ID. The ventral thalamus functions in memory processing and thus may have a role in speech learning, which is a major deficit in all described patients (Table 1) (Cholvin et al. 2013). Using the transparency of the zebrafish model and in vivo labeling of *arxa* neurons, we have shown that diminished *arxa* function results in disturbances in neuronal projections in the ventral thalamus and in the ventral telencephalon (Fig. 7). Since most of the ID patients exhibit only a delay, but not the inability, of speech learning, underdeveloped neuronal circuits may provide a reasonable explanation for the rather subtle phenotype in patients carrying *ARX* genomic rearrangements.

So far, *ARX* mutations have been mostly associated with a reduced number of GABAergic and cholinergic neurons in the cortex, occurring through inhibited proliferation of intermediate progenitor cells and disturbed migration of GABAergic neurons in the ganglionic eminences (Colasante et al. 2009, 2013; Kitamura et al. 2009). This ventral telencephalic expression domain is controlled by hs2, not only during development but also in the adult brain, where *ARX* may modulate neuronal function (Beguin et al. 2013). An orthologous sequence of hs7, which had only inconclusive activity in our assay (Fig. 2b/c; Suppl. Table 1), was previously identified to have enhancer function in this area (Suppl. Table 1; Colasante et al. 2008). This is of special interest since an hs7 core-conserved sequence has a single base pair change in a family with ID and had been correlated to infantile spasm [uc467, (Fullston et al. 2011)]. The same study reported another conserved sequence (uc466) with single nucleotide exchanges in two patients with autism, which is included in hs4. This sequence had also only variable activity in zebrafish (Suppl. Table 1) and a similar fragment was inconclusive in the mouse (Visel et al. 2009, 2013). The observed abnormalities in axon outgrowth and pathfinding in specimen with diminished Arxa function may provide an explanation for the spectrum of neuropsychiatric phenotypes observed in patients with copy number variations and *ARX* gene regulatory abnormalities.

Conclusions

We have shown that enhancers of *ARX* are located far upstream and downstream of the gene, in a 504-kb regulatory domain, which makes the regulation of the gene vulnerable to genomic rearrangements. A complex autoregulation of *ARX* further implies that increased expression levels due to duplicated coding regions will feed back into brain region-specific transcriptional control and will amplify the effects of genome dosage. Genomic rearrangements as observed in patients with ID would therefore alter *ARX* transcript levels brain region-specifically. In addition to formerly reported disturbances in the ventral telencephalon, this would also affect neurons and neuronal circuits of the ventral thalamus.

Acknowledgments This work was funded by the Medical Foundation of the University of Sydney, the European Union (NeuroXsys, FP7/2007-2013; Grant No. 223262 to TB, LBC and SWW), and the Australian National Health and Medical Research Council (Grant No. 1008077 and Grant No. 628952 to JG). CS is supported by the Australian Research Council (Future Fellowship FT120100086). JoG is supported by NHMRC research fellowship 1041920). TAH was supported by BBSRC grant BB/H012516/1 and Wellcome trust funding to the laboratory of SWW. We thank Reinhard Koester for providing Gal4 and bidirectional 5 × UAS cassettes.

References

Barresi MJF, Hutson LD, Chien C-B, Karlstrom RO (2005) Hedgehog regulated Slit expression determines commissure and glial cell position in the zebrafish forebrain. Development 132:3643–3656. doi:10.1242/dev.01929

Becker TS, Lenhard B (2007) The random versus fragile breakage models of chromosome evolution: a matter of resolution. Mol Genet Genomics 278:487–491. doi:10.1007/s00438-007-0287-0

Beguin S, Crépel V, Aniksztejn L et al (2013) An epilepsy-related ARX polyalanine expansion modifies glutamatergic neurons excitability and morphology without affecting GABAergic neurons development. Cereb Cortex 23:1484–1494. doi:10.1093/cercor/bhs138

Berman P, Zhang Z, Wolf YI et al (2000) Winnowing sequences from a database search. J Comput Biol 7:293–302. doi:10.1089/10665270050081531

Bienvenu T, Poirier K, Friocourt G et al (2002) ARX, a novel Prd-class-homeobox gene highly expressed in the telencephalon, is mutated in X-linked mental retardation. Hum Mol Genet 11:981–991

Biressi S, Messina G, Collombat P et al (2008) The homeobox gene Arx is a novel positive regulator of embryonic myogenesis. Cell Death Differ 15:94–104. doi:10.1038/sj.cdd.4402230

Cholvin T, Loureiro M, Cassel R et al (2013) The ventral midline thalamus contributes to strategy shifting in a memory task requiring both prefrontal cortical and hippocampal functions. J Neurosci 33:8772–8783. doi:10.1523/JNEUROSCI.0771-13.2013

Colasante G, Collombat P, Raimondi V et al (2008) Arx is a direct target of Dlx2 and thereby contributes to the tangential migration of GABAergic interneurons. J Neurosci 28:10674–10686. doi:10.1523/JNEUROSCI.1283-08.2008

Colasante G, Sessa A, Crispi S et al (2009) Arx acts as a regional key selector gene in the ventral telencephalon mainly through its transcriptional repression activity. Dev Biol 334:59–71. doi:10.1016/j.ydbio.2009.07.014

Colasante G, Simonet JC, Calogero R et al (2013) ARX regulates cortical intermediate progenitor cell expansion and upper layer neuron formation through repression of Cdkn1c. Cereb Cortex. doi:10.1093/cercor/bht222

Collombat P, Mansouri A, Hecksher-Sorensen J et al (2003) Opposing actions of Arx and Pax4 in endocrine pancreas development. Genes Dev 17:2591–2603. doi:10.1101/gad.269003

Collombat P, Xu X, Ravassard P et al (2009) The ectopic expression of Pax4 in the mouse pancreas converts progenitor cells

into alpha and subsequently beta cells. Cell 138:449–462. doi:10.1016/j.cell.2009.05.035

Colombo E, Galli R, Cossu G et al (2004) Mouse orthologue of ARX, a gene mutated in several X-linked forms of mental retardation and epilepsy, is a marker of adult neural stem cells and forebrain GABAergic neurons. Dev Dyn 231:631–639. doi:10.1002/dvdy.20164

Colombo E, Collombat P, Colasante G et al (2007) Inactivation of Arx, the murine ortholog of the X-linked lissencephaly with ambiguous genitalia gene, leads to severe disorganization of the ventral telencephalon with impaired neuronal migration and differentiation. J Neurosci 27:4786–4798. doi:10.1523/JNEUROSCI.0417-07.2007

Distel M, Wullimann MF, Koster RW (2009) Optimized Gal4 genetics for permanent gene expression mapping in zebrafish. Proc Natl Acad Sci USA 106:13365–13370. doi:10.1073/pnas.0903060106

Djiotsa J, Verbruggen V, Giacomotto J et al (2012) Pax4 is not essential for beta-cell differentiation in zebrafish embryos but modulates alpha-cell generation by repressing arx gene expression. BMC Dev Biol 12:37. doi:10.1186/1471-213X-12-37

Dong X, Fredman D, Lenhard B (2009) Synorth: exploring the evolution of synteny and long-range regulatory interactions in vertebrate genomes. Genome Biol 10:R86. doi:10.1186/gb-2009-10-8-r86

Engstrom PG, Fredman D, Lenhard B (2008) Ancora: a web resource for exploring highly conserved noncoding elements and their association with developmental regulatory genes. Genome Biol 9:R34. doi:10.1186/gb-2008-9-2-r34

Filippi A, Mueller T, Driever W (2014) vglut2 and gad expression reveal distinct patterns of dual GABAergic versus glutamatergic cotransmitter phenotypes of dopaminergic and noradrenergic neurons in the zebrafish brain. J Comp Neurol 522:2019–2037. doi:10.1002/cne.23524

Folgueira M, Bayley P, Navratilova P et al (2012) Morphogenesis underlying the development of the everted teleost telencephalon. Neural Dev 7:32. doi:10.1186/1749-8104-7-32

Forlano PM, Bass AH (2011) Neural and hormonal mechanisms of reproductive-related arousal in fishes. Horm Behav 59:616–629. doi:10.1016/j.yhbeh.2010.10.006

Frankel N, Davis GK, Vargas D et al (2010) Phenotypic robustness conferred by apparently redundant transcriptional enhancers. Nature 466:490–493

Fredman D, Engstrom PG, Lenhard B (2009) Web-based tools and approaches to study long-range gene regulation in Metazoa. Brief Funct Genomic Proteomic 8:231–242. doi:10.1093/bfgp/elp023

Friocourt G, Poirier K, Rakić S et al (2006) The role of ARX in cortical development. Eur J Neurosci 23:869–876. doi:10.1111/j.1460-9568.2006.04629.x

Friocourt G, Kanatani S, Tabata H et al (2008) Cell-autonomous roles of ARX in cell proliferation and neuronal migration during corticogenesis. J Neurosci 28:5794–5805. doi:10.1523/JNEUROSCI.1067-08.2008

Fullston T, Finnis M, Hackett A et al (2011) Screening and cell-based assessment of mutations in the Aristaless-related homeobox (ARX) gene. Clin Genet 80:510–522. doi:10.1111/j.1399-0004.2011.01685.x

Fulp CT, Cho G, Marsh ED et al (2008) Identification of Arx transcriptional targets in the developing basal forebrain. Hum Mol Genet 17:3740–3760. doi:10.1093/hmg/ddn271

Gecz J, Cloosterman D, Partington M (2006) ARX: a gene for all seasons. Curr Opin Genet Dev 16:308–316. doi:10.1016/j.gde.2006.04.003

Hong JW, Hendrix DA, Levine MS (2008) Shadow enhancers as a source of evolutionary novelty. Science 321:1314

Ishibashi M, Mechaly AS, Becker TS, Rinkwitz S (2013) Using zebrafish transgenesis to test human genomic sequences for specific enhancer activity. Methods 62:216–225. doi:10.1016/j.ymeth.2013.03.018

Jackman WR, Stock DW (2006) Transgenic analysis of Dlx regulation in fish tooth development reveals evolutionary retention of enhancer function despite organ loss. Proc Natl Acad Sci USA 103:19390–19395

Kikuta H, Fredman D, Rinkwitz S et al (2007a) Retroviral enhancer detection insertions in zebrafish combined with comparative genomics reveal genomic regulatory blocks—a fundamental feature of vertebrate genomes. Genome Biol 8(Suppl 1):S4. doi:10.1186/gb-2007-8-s1-s4

Kikuta H, Laplante M, Navratilova P et al (2007b) Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. Genome Res 17:545–555. doi:10.1101/gr.6086307

Kimmel CB, Ballard WW, Kimmel SR et al (1995) Stages of embryonic development of the zebrafish. Dev Dyn 203:253–310. doi:10.1002/aja.1002030302

Kitamura K, Yanazawa M, Sugiyama N et al (2002) Mutation of ARX causes abnormal development of forebrain and testes in mice and X-linked lissencephaly with abnormal genitalia in humans. Nat Genet 32:359–369. doi:10.1038/ng1009

Kitamura K, Itou Y, Yanazawa M et al (2009) Three human ARX mutations cause the lissencephaly-like and mental retardation with epilepsy-like pleiotropic phenotypes in mice. Hum Mol Genet 18:3708–3724. doi:10.1093/hmg/ddp318

Komisarczuk AZ, Kawakami K, Becker TS (2009) Cis-regulation and chromosomal rearrangement of the fgf8 locus after the teleost/tetrapod split. Dev Biol 336:301–312. doi:10.1016/j.ydbio.2009.09.029

Machluf Y, Levkowitz G (2011) Visualization of mRNA expression in the zebrafish embryo. Methods Mol Biol 714:83–102. doi:10.1007/978-1-61779-005-8_6

McBride DJ, Buckle A, van Heyningen V, Kleinjan DA (2011) DNaseI hypersensitivity and ultraconservation reveal novel, interdependent long-range enhancers at the complex Pax6 cis-regulatory region. PLoS One 6:e28616. doi:10.1371/journal.pone.0028616

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

Mueller T (2012) What is the thalamus in zebrafish? Front Neurosci 6:64. doi:10.3389/fnins.2012.00064

Mueller T, Wullimann MF (2002) BrdU-, neuroD (nrd)- and Hu-studies reveal unusual non-ventricular neurogenesis in the postembryonic zebrafish forebrain. Mech Dev 117:123–135

Mueller T, Wullimann MF (2005) Atlas of early zebrafish brain development: a tool for molecular neurogenetics. Elsevier, Amsterdam

Mueller T, Dong Z, Berberoglu MA, Guo S (2011) The dorsal pallium in zebrafish, *Danio rerio* (Cyprinidae, Teleostei). Brain Res 1381:95–105. doi:10.1016/j.brainres.2010.12.089

Navratilova P, Fredman D, Hawkins TA et al (2009) Systematic human/zebrafish comparative identification of cis-regulatory activity around vertebrate developmental transcription factor genes. Dev Biol 327:526–540. doi:10.1016/j.ydbio.2008.10.044

Navratilova P, Fredman D, Lenhard B, Becker TS (2010) Regulatory divergence of the duplicated chromosomal loci sox11a/b by subpartitioning and sequence evolution of enhancers in zebrafish. Mol Genet Genomics 283:171–184. doi:10.1007/s00438-009-0503-1

Norton WH, Mangoli M, Lele Z et al (2005) Monorail/Foxa2 regulates floorplate differentiation and specification of oligodendrocytes, serotonergic raphé neurones and cranial motoneurones. Development 132:645–658. doi:10.1242/dev.01611

Osório J, Mueller T, Rétaux S et al (2010) Phylotypic expression of the bHLH genes Neurogenin2, Neurod, and Mash1 in the mouse

embryonic forebrain. J Comp Neurol 518:851–871. doi:10.1002/cne.22247

Oxtoby E, Jowett T (1993) Cloning of the zebrafish krox-20 gene (krx-20) and its expression during hindbrain development. Nucleic Acids Res 21:1087–1095

Partington MW, Turner G, Boyle J, Gecz J (2004) Three new families with X-linked mental retardation caused by the 428-451dup(24 bp) mutation in ARX. Clin Genet 66:39–45. doi:10.1111/j.0009-9163.2004.00268.x

Perry MW, Boettiger AN, Bothma JP, Levine M (2010) Shadow enhancers foster robustness of drosophila gastrulation. Curr Biol. doi:10.1016/j.cub.2010.07.043

Popovici C, Busa T, Boute O, Thuresson AC, Perret O, Sigaudy S, Södergren T, Andrieux J, Moncla A, Philip N (2014) Whole ARX gene duplication is compatible with normal intellectual development. Am J Med Genet A 164A(9):2324–2327

Price MG, Yoo JW, Burgess DL et al (2009) A triplet repeat expansion genetic mouse model of infantile spasms syndrome, Arx(GCG)10 + 7, with interneuronopathy, spasms in infancy, persistent seizures, and adult cognitive and behavioral impairment. J Neurosci 29:8752–8763. doi:10.1523/JNEUROSCI.0915-09.2009

Punnamoottil B, Kikuta H, Pezeron G et al (2008) Enhancer detection in zebrafish permits the identification of neuronal subtypes that express Hox4 paralogs. Dev Dyn 237:2195–2208. doi:10.1002/dvdy.21618

Quillé M-L, Carat S, Quéméner-Redon S et al (2011) High-throughput analysis of promoter occupancy reveals new targets for Arx, a gene mutated in mental retardation and interneuronopathies. PLoS One 6:e25181. doi:10.1371/journal.pone.0025181

Rada-Iglesias A, Bajpai R, Swigut T et al (2011) A unique chromatin signature uncovers early developmental enhancers in humans. Nature 470:279–283. doi:10.1038/nature09692

Rink E, Wullimann MF (2001) The teleostean (zebrafish) dopaminergic system ascending to the subpallium (striatum) is located in the basal diencephalon (posterior tuberculum). Brain Res 889:316–330

Rink E, Wullimann MF (2004) Connections of the ventral telencephalon (subpallium) in the zebrafish (*Danio rerio*). Brain Res 1011:206–220. doi:10.1016/j.brainres.2004.03.027

Sandelin A, Alkema W, Engstrom P, Wasserman WW, Lenhard B (2004a) JASPAR: an open-access database for eukaryotic transcription factor binding pro- files. Nucleic Acids Res 32D:91–94

Sandelin A, Wasserman WW, Lenhard B (2004b) ConSite: web-based prediction of regulatory elements using cross-species comparison. Nucleic Acids Res 32:W249–W252

Seufert DW, Prescott NL, El-Hodiri HM (2005) Xenopus aristaless-related homeobox (xARX) gene product functions as both a transcriptional activator and repressor in forebrain development. Dev Dyn 232:313–324. doi:10.1002/dvdy.20234

Shoubridge C, Fullston T, Gecz J (2010) ARX spectrum disorders: making inroads into the molecular pathology. Hum Mutat 31:889–900. doi:10.1002/humu.21288

Sohn J, Natale J, Chew LJ, Belachew S, Cheng Y, Aguirre A, Lytle J, Nait-Oumesmar B, Kerninon C, Kanai-Azuma M, Kanai Y, Gallo V (2006) Identification of Sox17 as a transcription factor that regulates oligodendrocyte development. J Neurosci 26:9722–9735

Stromme P, Mangelsdorf ME, Shaw MA et al (2002) Mutations in the human ortholog of Aristaless cause X-linked mental retardation and epilepsy. Nat Genet 30:441–445. doi:10.1038/ng862

Turner KJ, Bracewell TG, Hawkins TA (2014) Anatomical dissection of zebrafish brain development. Methods Mol Biol 1082:197–214. doi:10.1007/978-1-62703-655-9_14

Visel A, Blow MJ, Li Z et al (2009) ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature 457:854–858

Visel A, Taher L, Girgis H et al (2013) A high-resolution enhancer atlas of the developing telencephalon. Cell 152:895–908. doi:10.1016/j.cell.2012.12.041

Whibley AC, Plagnol V, Tarpey PS et al (2010) Fine-scale survey of X chromosome copy number variants and indels underlying intellectual disability. Am J Hum Genet 87:173–188. doi:10.1016/j.ajhg.2010.06.017

Wilson SW, Ross LS, Parrett T, Easter SS (1990) The development of a simple scaffold of axon tracts in the brain of the embryonic zebrafish, *Brachydanio rerio*. Development 108:121–145

Wullimann MF, Mueller T (2004) Identification and morphogenesis of the eminentia thalami in the zebrafish. J Comp Neurol 471:37–48. doi:10.1002/cne.20011

Wullimann MF, Rupp B, Reichert H (1996) Neuroanatomy of the zebrafish brain. A topological atlas. Birkhäuser, Basel