



Regulatory variants of *FOXG1* in the context of its topological domain organisation

Mana M Mehrjouy¹ · Ana Carolina S. Fonseca² · Nadja Ehmke^{3,4} · Giorgio Paskulin⁵ · Antonio Novelli⁶ · Francesco Benedicenti⁷ · Maria Antonietta Mencarelli⁸ · Alessandra Renieri⁸ · Tiffany Busa⁹ · Chantal Missirlian⁹ · Claus Hansen¹ · Kikue Terada Abe¹⁰ · Carlos Eduardo Speck-Martins¹⁰ · Angela M. Vianna-Morgante² · Mads Bak¹ · Niels Tommerup¹

Received: 19 January 2017 / Revised: 28 August 2017 / Accepted: 31 August 2017 / Published online: 30 December 2017
© European Society of Human Genetics 2018

Abstract

FOXG1 syndrome is caused by *FOXG1* intragenic point mutations, or by long-range position effects (LRPE) of intergenic structural variants. However, the size of the *FOXG1* regulatory landscape is uncertain, because the associated topologically associating domain (TAD) in fibroblasts is split into two domains in embryonic stem cells (hESC). Indeed, it has been suggested that the pathogenetic mechanism of deletions that remove the stem-cell-specific TAD boundary may be enhancer adoption due to ectopic activity of enhancer(s) located in the distal hESC-TAD. Herein we map three de novo translocation breakpoints to the proximal regulatory domain of *FOXG1*. The classical FOXG1 syndrome in these and in other translocation patients, and in a patient with an intergenic deletion that removes the hESC-specific TAD boundary, do not support the hypothesised enhancer adoption as a main contributor to the FOXG1 syndrome. Also, virtual 4C and HiC-interaction data suggest that the hESC-specific TAD boundary may not be critical for *FOXG1* regulation in a majority of human cells and tissues, including brain tissues and a neuronal progenitor cell line. Our data support the importance of a critical regulatory region (SRO) proximal to the hESC-specific TAD boundary. We further narrow this critical region by a deletion distal to the hESC-specific boundary, associated with a milder clinical phenotype. The distance from *FOXG1* to the SRO (> 500 kb) highlight a limitation of ENCODE DNase hypersensitivity data for functional prediction of LRPE. Moreover, the SRO has little overlap with a cluster of frequently associating regions (FIREs) located in the proximal hESC-TAD.

Introduction

Massive sequence-based analyses of genomic all-versus-all DNA-interactions (HiC-chromosome conformation capture) have revealed that the human genome is organised into megabase-sized topologically associating domains (TADs) [1]. *Cis*-regulatory interactions predominantly occur within

Electronic supplementary material The online version of this article (<https://doi.org/10.1038/s41431-017-0011-4>) contains supplementary material, which is available to authorised users.

✉ Niels Tommerup
ntommerup@sund.ku.dk

¹ Wilhelm Johannsen Centre for Functional Genome Research, Department of Cellular and Molecular Medicine, University of Copenhagen, Copenhagen, Denmark

² Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil

³ Institut für Medizinische Genetik und Humangenetik, Charité-Universitätsmedizin Berlin, Berlin, Germany

⁴ Berlin Institute of Health (BIH), Berlin, Germany

⁵ GENEX Instituto de Exames Genéticos Ltda, Porto Alegre, Brazil

⁶ Laboratory of Medical Genetics, Bambino Gesù Children's Hospital, IRCCS, Rome, Italy

⁷ Genetic Counselling Service, Department of Paediatrics, Regional Hospital of Bolzano, Bolzano, Italy

⁸ Medical Genetics Unit, Azienda Ospedaliera Universitaria Senese, Siena, Italy

⁹ Département de Génétique, Hôpital Timone Enfants, Marseille, France

¹⁰ Rede Sarah de Hospitais de Reabilitação, Brasília, Brazil

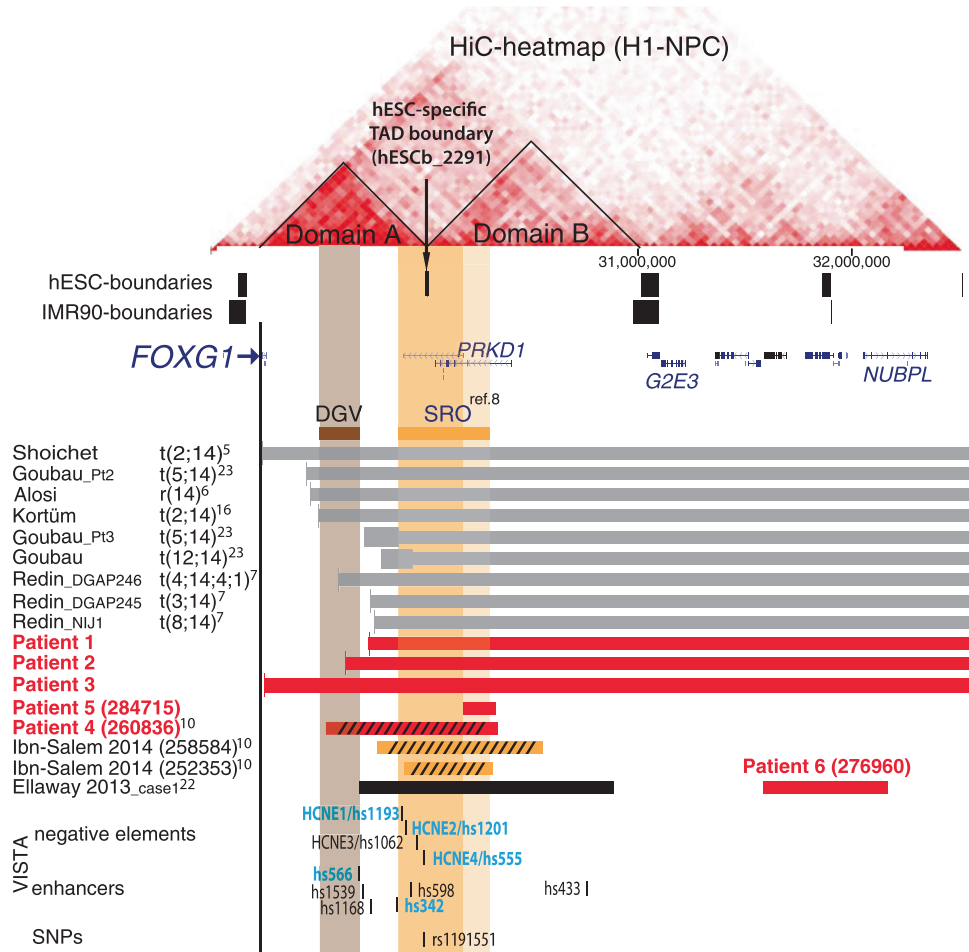


Fig. 1 Schematic view of structural variants located 3' to *FOXG1*, obtained from the literature, DECIPHER and the present study. The Hi-C profile around the *FOXG1* locus in human neuronal progenitor cells H1-NPC is shown on top. The tents indicate Domain A and Domain B, separated by the embryonic cell-specific boundary (hESCb_2291). *FOXG1* is located in the extreme proximal region of Domain A (vertical line). Grey bars indicate regions removed by previously published translocations [5–7, 16, 23] and red bars indicate patients in the present study (P1–P6). DGV: deletions/duplications present in the Database of Genomic Variants in healthy individuals

indicated by brown area. Striped deletions are those that were used in ref. [10] to exemplify enhancer adoption by deletion of hESCb_2291 (vertical arrow between Domains A and B). The black bar exemplifies one of several additional reported deletions that remove the distal enhancer hs433 and also hESCb_2291. The orange region indicates the SRO-region as defined by ref. [8], encompassing enhancer hs598 and the VISTA [34] negative elements HCNE1–4. The SRO is further refined (darker shade) by the deletion in Patient 5 with only mild ID, but no other features of the *FOXG1* syndrome. The elements coloured in blue overlap with DNase I hypersensitive sites

a TAD, and these interactions are mediated by chromatin looping [1, 2]. TADs are separated by boundaries, enriched in transcriptional start sites, housekeeping genes, binding sites for CCCTC-binding factor (CTCF) insulator factor, and short interspersed nuclear elements (SINE) elements. The majority of TADs are identical in human embryonic stem cells (hESC) and terminally differentiated cell types. However, the Forkhead box G1 (*FOXG1*) gene on chromosome 14 is located within a 2 Mb TAD in IMR90-fibroblasts, which is split into two TADs (Domains A and B) by a hESC-specific TAD boundary (hESCb_2291) (Fig. 1) [1]. *FOXG1* encodes a winged-helix transcriptional

repressor critical for early telencephalon development, and *FOXG1* intragenic point mutations cause a congenital variant of Rett syndrome [3].

Chromosomal structural variants (e.g., translocations, inversions, insertions, deletions and duplications) may interfere with TAD structure by disrupting/fusing domains or boundaries, which may dysregulate genes with potential pathogenic consequence [4]. Due to long-range position effects (LRPE), chromosomal breakpoints 3 to *FOXG1* lead to a congenital variant of Rett syndrome, similar to *FOXG1* intragenic point mutations (Fig. 1) [5–7]. Intergenic deletions have defined an ~430 kb (chr14:29,875,672–

30,303,083) smallest region of deletion overlap (SRO), more than 600 kb distal to *FOXG1* in the distal part of Domain A, which defines a critical region with several conserved regulatory elements, deletion of which causes the FOXG1 phenotype (Fig. 1) [8].

However, this SRO also spans hESCb_2291 (Fig. 1), and it has been proposed that deletion of hESCb_2291 may lead to enhancer adoption [9], whereby regulatory elements in Domain B (e.g., enhancer hs433) are brought into ectopic physical contact with *FOXG1* [10]. Whether the phenotypic effects of *FOXG1*-LRPE is due to loss of enhancers or silencers, gain of enhancer(s), deletion of the hESC-specific boundary, or a combination of these mechanisms, is yet to be answered.

Herein we have mapped three additional translocations associated with the *FOXG1*-LRPE, with breakpoints in the distal part of Domain A, providing further support for the proximal boundary of the SRO. This is also supported by deletions and duplications proximal to the SRO region in presumed healthy individuals. In addition, we identified a new deletion within Domain B associated with a different clinical phenotype, which narrows the distal SRO-boundary towards the highly conserved non-coding elements within this region, the loss of which has been proposed to be critical for the FOXG1 syndrome [8]. Moreover, we provide clinical evidence against the role of enhancer adoption as a major contributor to the FOXG1 syndrome. We also included small deletions that remove either neighbouring TADs or TAD boundaries.

Furthermore, we used the available virtual 4C and HiC DNA–DNA interaction data sets to show that the *FOXG1* regulatory landscape in most human cell lines and tissues involves both Domains A and B. This may be even more evident in the mouse cell lines, where the human pattern of two (embryonic) TADs is fused into a single domain.

The proximal part of the *FOXG1* regulatory domain is also enriched for the recently described frequently associating regions (FIREs), including hippocampus, prefrontal cortex and neuronal progenitor cell specific FIREs [11]. However, the *FOXG1*-critical regulatory region does not overlap with FIREs, and apparently normal individuals may have deletions/duplications of FIREs within the *FOXG1* TAD.

Patients and methods

A German (P1) and two Brazilian patients (P2 and P3) with de novo translocations t(9;14), t(4;14) and t(2;14), respectively, were obtained via the International Breakpoint Mapping Consortium (IBMC, in preparation). All had classical features associated with the FOXG1 syndrome: severe intellectual disability, corpus callosum hypoplasia, seizures, severe speech impairment, and postnatal

microcephaly. Genomic imbalances at the translocation breakpoints were excluded by array CGH and by mate-pair sequencing (M-PS) (Fig. 2a, c). The chromosome 14 breakpoints were mapped by M-PS within the *FOXG1* Domain A and defined to bp-level by Sanger sequencing (P1, P2) (Figs. 1 and 2). To define further the *FOXG1* critical region, three patients (P4–P6) with deletions within and distal to the *FOXG1* TADs were obtained from DECIPHER (<https://decipher.sanger.ac.uk/> - 260836, 284715, 276960) (Fig. 1).

We submitted P1–P3 to ClinVar and the accession numbers are: SCV000576475 = P1, SCV000576474 = P2, SCV000576473 = P3.

Karyotyping

GTG-banded metaphase chromosome analyses were performed on cultured peripheral blood lymphocytes (Supplementary Fig. 1).

Array CGH

The presence of cryptic imbalances was investigated by genome-wide array-comparative genomic hybridisation (aCGH) using different aCGH platforms (Supplementary Table 1).

Mate-pair Sequencing

Mate-pair sequencing (M-PS) (Illumina San Diego, CA, USA) was performed on peripheral blood DNA from the three translocation patients. Each DNA library was created from 1 µg of DNA using Illumina's Nextera Mate-pair sample preparation kit, following the instructions for a gel-free preparation of a 2 kb effective insert size library (Mate Pair Library v2, Illumina). The final libraries were quantified using Pico Green (Quant-iT, Invitrogen) and the Agilent Bioanalyzer DNA 1000 kit (Quant-iT, Invitrogen). Samples were pooled together with other samples into groups of four. Each pool was sequenced on a single lane of the HiSeq2500 (Illumina, San Diego, CA, USA) (2 × 100 bp). Raw sequence reads were trimmed and adapter sequences were removed using cutadapt. Trimmed reads were aligned to the human reference genome (all genomic coordinates in this study are hg19), using Burrows-Wheeler-Aligner (BWA). Reads with mapping qualities < 25 were discarded from further analysis. Reads with unexpected strand orientation or reads aligning to different chromosomes were extracted and SVDetect (<http://svdetect.sourceforge.net/>) and Delly (www.korbel.embl.de/software.html) were used to identify potential rearrangements. Structural variants (SVs) from the patients were filtered against ~200 in-house mate-pair sequencing data sets, and

with the Database of Genomic Variants (<http://dgv.tcag.ca/>) to identify unique SVs. Bed-files from SVDetect were visualised in Integrative Genome Viewer (IGV) (<http://www.broadinstitute.org/igv/>) to find the approximate location of breakpoints (BP) at chromosome 14q12.

Breakpoint mapping by Sanger Sequencing

The precise sites of the BPs were determined by Sanger sequencing. To design the primers, genomic sequences harbouring the BPs, based on M-PS data, were extracted from the UCSC Genome Browser. To avoid designing primers located within segmental duplications, the concatemers resulting from merging the sequences constituting each BP were masked for repetitive sequences. Primer3plus was used to design the primers for each BP. PCR was performed by standard methods using genomic DNA from the patients mixing with High Fidelity Master Mix (Phusion® High-Fidelity DNA Polymerase Products). A normal control was used as a template. PCR products were run in 1% agarose gels, and specific fragments not present in the control were excised, purified and Sanger sequenced with BigDye Terminator chemistry (Applied Biosystems) on an automated capillary sequencer (ABI 3130XL) according to the manufacturer's instruction. To find the BPs, the Sanger sequencing data was uploaded to the UCSC Genome Browser. New primer sets were designed when the sequences did not span individual BPs, until spanning PCR products were obtained. Finished junction sequences were split up at the BP and aligned to genomic sequence of the BP-region to visualise possible indels.

Topological domains and interaction loops

Topological domain boundaries in hESC-cells and IMR90 cells, and in 21 additional tissues and cell types were taken from references [1, 2, 11]. Chromatin interaction loops were taken from ref. [2].

HiC and virtual 4C

4C-seq [12] can be used to experimentally visualise cell line and tissue-specific DNA–DNA interactions between a specific locus (e.g., the *FOXG1* promoter) and the rest of the genome and to study alterations of these interactions in chromosomal rearrangements [4]. However, normal 4C-seq data can also be extracted from the vast amount of HiC-data that has been generated by HiC-sequencing (Virtual 4C) [4, 12]. We generated HiC and Virtual 4C-maps of the 4–5 Mb region spanning the *FOXG1* TADs from neuronal progenitor cells H1-NPC [13] and from mouse cortex, using the 3D Genome Browser (<http://promoter.bx.psu.edu/hi-c/virtual4c.php>) (Supplementary Figs. 2 and 3).

Results

Clinical presentations

Patient 1 (P1), a four-year-old boy, is the first born child of German and Thai non-consanguineous parents. The pregnancy resulted from in vitro fertilisation. He was delivered at term after an uncomplicated pregnancy, with a birth weight of 4100 g (+1 SD), body length of 56 cm (+1 SD), and head circumference (HC) of 36 cm (+0.4 SD). Apgar score was 10/10. At the age of 3 months, he developed generalised seizures. Brain magnetic resonance imaging (MRI) at the age of eight months was normal. At nine months, he was noted to have severe developmental delay, postnatal microcephaly and muscular hypotonia, especially in the upper part of his body. Thus, the regression after a period of normal development which is one of the differential features between Rett Syndrome and the *FOXG1* syndrome was not present. Ophthalmological examination revealed convergent strabismus and hypermetropia. He had tongue protrusion, constipation and hypersalivation. He had undescended testes on the left side; therefore, at the age of one year orchidopexy was performed. At the age of 19 months, he had poor head control, could not sit independently, and had no speech development. He showed behavioural anomalies including inappropriate laughter, bruxism, stereotypic hand and head movements, high pain tolerance and increased sensitivity to noise. He had a 2 cm café-au-lait spot on his trunk. He did not have any dysmorphic features. Chest X-ray at the age of 2 years showed broad ribs. At the age of 4 years, his height was 103 cm (−0.5 SD), HC was 45 cm (−4.5 SD), and weight was 15 kg (−0.8 SD). He still had severe muscular hypotonia but better head control. He did not show any breathing abnormalities. Chromosomal analysis revealed a de novo apparently balanced translocation, 46,XY,t(9;14)(q22.3;q11.2)dn (Supplementary Fig. 1). No gain or loss of chromosomal material was detected at the translocation breakpoints by aCGH, but maternally inherited duplications at 15q21.12(47292250–47309868)x3 and Xp22.31(6451676–8115124)x3 of unknown clinical significance were revealed.

Patient 2 (P2), a 5 year-old-girl, was the second child of healthy non-consanguineous Brazilian parents. She was born at term after an uneventful pregnancy; she developed respiratory distress, and was admitted into the intensive care unit, where she remained hospitalised and intubated for the first nine months of age, with nutrition being provided through a nasogastric tube. She presented with cystic adenomatoid malformation involving part of the left lung, laryngotracheomalacia and several episodes of pneumonia. At the age of 2 months she had two convulsive episodes, and since then has been on anticonvulsant therapy. At three

months of age, the patient had a cardiorespiratory arrest due to accidental endotracheal extubation. Tracheostomy was performed at the age of 4 months. Her brain MRI demonstrated agenesis of the corpus callosum, with parallel appearance of bilateral lateral ventricles, dilated superficial blood vessels and occipital and temporal horns. She was diagnosed with quadriplegic cerebral palsy with axial hypotonia and distal spasticity. At 1 year and 10 months of age, her length was 84 cm (-0.19 SD), her weight was 11 kg (-0.83 SD), and HC was 40 cm (-4.9 SD). Clinical evaluation at the age of 5 years revealed severe developmental delay (not able to sit up without support) and intellectual disability, with no speech development, stereotypic movements, inappropriate laughter and eye contact. Ophthalmological examination revealed strabismus and astigmatism. Due to oropharyngeal dysphagia, gastrostomy and tracheostomy were still necessary. Chromosome analysis revealed a de novo apparently balanced reciprocal translocation 46,XX,t(4;14)(q27;q13)dn (Supplementary Fig. 1). No potentially pathogenic chromosomal imbalances were detected by aCGH. Her parents had normal karyotypes. Her older sister was clinically normal.

Patient 3 (P3), a one year-old-boy was the second child of healthy non-consanguineous Brazilian parents. He was born after caesarean section, at 33 weeks of gestation, due to placental abruption. The Apgar score was normal and the amniotic fluid was non-meconial. Despite a normal length, weight and HC at birth, he developed microcephaly within the first year of his life. At the age of one year, his length and weight were within standard deviations, but the cephalic perimeter was below the 3rd centile. He has developmental delay and hypotonia and is not able to sit without assistance. Ophthalmologic investigation showed convergent strabismus without retinal abnormalities. Central Nervous System (CNS) MRI showed a volumetric reduction of the cerebral parenchyma, hypomyelination, and hypoplasia of corpus callosum mainly involving the anterior part and prominent lateral ventricles. Chromosome analysis revealed a de novo apparently balanced reciprocal translocation 46,XY,t(2;14)(q37;q11.2)dn.

Patient 4 (P4, DECIPHER 260836) is one of three deletion carriers used by Ibn Salem et al. [10] to exemplify enhancer adoption (Fig. 5 in ref. [8]). The patient was the second child of healthy consanguineous parents. The boy was born at 40 weeks gestation after an uneventful pregnancy. At birth, his weight was 3890 g ($+0.4$ SD), length 50 cm (± 0 SD) and HC 35 cm ($+0.2$ SD). He was referred at 24 months due to developmental delay without regression. Clinical evaluation showed an abnormal facial appearance with microcephaly, an open mouth with an averted lower lip, and stereotypic hand movements. His weight was 10.4 kg (-1.5 SD), length 83 cm (-1 SD), and HC 45 cm (-3 SD). He was unable to sit but was able to

interact. He presented with developmental delay, could sit at 4 years and crawl at 6 years. At the last evaluation at the age of 7 years, language was absent. He was unable to walk alone. His height was 112 cm (-1.5 SD), weight 16 kg (-2.5 SD) and HC 47.5 cm (-4 SD). Brain MRI showed cerebral ventricular dilatation, frontal lobe hypoplasia and abnormal hippocampus morphology. Chromosome analysis and methylation screening of 15q11q13 were normal. aCGH revealed an ~ 800 kb deletion, including exons 2 to 18 of *PRKDI*. The deletion was confirmed by Fluorescence In Situ Hybridization (FISH) with the RP11-419C10 BAC clone (data not shown), 46,XY.ish 14q12(RP11-419C10 $\times 1$).arr[hg19] 14q12(29538004_30335064)x1. No other clinically significant copy number variations were detected.

Patient 5 (P5, DECIPHER 284715) shows a deletion at position 46,XY.arr[hg19] 14q12(30173942_30325824)x1 that removes part of *PRKDI*, leaving both hESCb_2291 and hs433 intact (Fig. 1).

The proband is a 9-year-old boy. He is the second-born child of healthy unrelated parents. He was born at 38 weeks gestation through caesarean delivery after a pregnancy complicated by threatened preterm labour. At birth, he was 2650 g in weight (-1.2 SD), 49 cm in length ($+1$ SD) and his HC was 34.5 cm (-0.4 SD). The Apgar score was 7/1' and 8/5'. Postnatal growth was normal until 6 years of age, when the body weight increased as a result of marked hyperphagia. Global development was normal with an IQ in the lower normal range (88). However, speech and language development was severely delayed with first words at the age of 6 years and the first short sentences at the age of 7 years. When seen last, significant speech delay was present. The boy also showed clumsiness with reduced fine motor skills, and severe impairment of social functioning with low self-esteem, behavioural problems including low frustration tolerance associated with violent outbursts and episodes of self-injurious and heteroaggressive behaviour. Hyperactivity, hyperphagia, attraction to water and sleep disturbances were also present. His facial dysmorphism included straight and downslanting eyebrows, underdeveloped tragus and antitragus, prominent nasal bridge, broad nasal tip, and mild micrognathia. He also showed mild hyperextensibility of elbows and of finger joints, bilateral sandal gaps and bilateral tibial deviation of the distal phalanges of the 4th toes. There was no evidence of other skeletal abnormalities, microcephaly, genital anomalies, ocular abnormalities, gastrointestinal or respiratory disturbance, dyskinesias, epilepsy or other neurological symptoms. Brain imaging was normal.

Patient 6 (P6, DECIPHER 276960) carries a deletion removing the next TAD boundary distal to Domain B, 46,XX.arr[hg19] 14q12(31562289_32142560)x1 (Fig. 1). This female proband was born by induced labour at 38 weeks of

gestation. Starting from the third trimester of gestation, growth retardation was observed. Birth weight was 2160 g (-2 SD), length 41 cm (-3.2 SD), HC 30 cm (-2.6 SD), Apgar indices 8-1' and 9-5'. She presented with poor sucking in the first weeks of life. Head control was acquired at the age of 4 months, sitting at 5 months, crawling at 11 months and walking with support at 12 months. Language was acquired at about 1 year, and at the time of genetic counselling, 16 months of age, it was limited to about 10 words. Further evaluation showed a HC of 43.5 cm (-2.7 SD) and her height 70.5 cm (-4 SD). She was noted to have a high frontal hairline, periorbital succulence, large pointed ears with hypoplastic lobes, a flat face, depressed nasal bridge, small mouth held open with thin lips and protruding tongue, joint hyperlaxity, and minimal clinodactyly of the fourth and fifth fingers.

Characterisation of the translocation breakpoints by mate-pair sequencing

Patient 1

The total number of M-PS read pairs passing the alignment score were 57,072,837. The translocation breakpoints were refined by a cluster of 24 BP-spanning reads to 9q22.31:94,269,066~94,269,642 and 14q12:29,738,715~29,739,465 (Fig. 2a). The BP on chromosome 9 does not truncate any gene or predicted regulatory domain. The breakpoint on the chromosome 14 truncates a highly conserved regulatory landscape 502 kb downstream of *FOXG1*. Sanger sequencing defined the exact breakpoint position to:

46,XY,t(9;14)(q22;q12)dn.arr[GRCh37] 15q21.12
(47292250-47309868)x3 mat,Xp22.31(6451676-8115124)
x3 mat.seq[GRCh37] t(9;14)(q22.31;q12)
g.[chr9:pter_cen_94269450::chr14:29738735_qter]
g.[chr14:pter_cen_29738725::chr9:94269454_qter].

Nine base pairs (CACAGCTC) were deleted on the der(14), and three base pairs (CAT) were deleted on the der(9) (Fig. 2a).

Patient 2

The total number of M-PS read pairs passing the alignment score were 44,220,256. The translocation breakpoints were defined by a cluster of 19 BP-spanning reads to 4q26:114,869,960~114,871,254 and 14q12:29,630,531~29,631,032 (Fig. 2b). The breakpoint on chromosome 4 truncates the *ARSJ* (arylsulfatase family, member J) gene, whereas the breakpoint on chromosome 14 was located 392 kb downstream of *FOXG1*. The breakpoints were refined by Sanger sequencing to:

46,XX,t(4;14)(q27;q13)dn.arr[GRCh37] (1-22,X)x2.seq
[GRCh37] t(4;14)(q26;q12)

g.[chr4:pter_cen_114870849::chr14:29630929_qter]
g.[chr14:pter_cen_29630919::chr14:114870855_qter]

Five base pairs (GACTT) were deleted at the der(4) breakpoint, while nine base pairs (GCCCTGGCT) were deleted at the der(14) breakpoint.

Patient 3

The total number of M-PS read pairs passing the alignment score were 43,101,205. A cluster of 34 BP-spanning reads identified the translocation breakpoints to 2q36.1:224,764,222~224,765,719 and 14q12:29,263,288~29,263,828 (Fig. 2c). The breakpoint on chr2 truncates *WDFY1* and the breakpoint on chr14 is located only 27 kb downstream of *FOXG1* truncating a non-coding RNA.

The coverage of normal reads in all three patients did not reveal gross duplications or deletions surrounding the breakpoint regions (Fig. 2a–c). The karyotype is defined to 46,XY,t(2;14)(q37;q11.2)dn. arr[GRCh37] (1-22)x2,(XY)x1.seq[GRCh37] t(2;14)(q36.1;q12). The breakpoint region on chromosome 14 overlaps with SINE elements, and Sanger Sequencing failed to delineate the breakpoint to single-nucleotide level.

Discussion

Many developmental genes including *FOXG1* are located in evolutionarily conserved genomic landscapes, which are enriched in conserved non-coding elements (CNEs). Some of these CNEs have been shown to be regulatory elements located hundreds to more than a thousand kilobases from the target gene [14]. These long-range regulatory elements can be brought into close proximity with the promoter through the formation of chromatin loops. This mechanism provides an attractive model for how a promoter can be physically linked to an array of regulatory elements, and specifies the spatiotemporal expression of the specific target gene. SVs can interfere with this interaction by deleting, truncating or translocating these *cis*-acting regulators [15], and/or by introducing ectopic regulators into the vicinity of the developmental gene [4].

Herein we dissected the *FOXG1* regulatory domain by evaluating three novel translocations in addition to reviewing published cases, with a focus on those patients with deletions and LRPE-associated breakpoints that might deregulate *FOXG1* [5–7, 16–24].

Deletions and inactivating point mutations of *FOXG1* itself clearly support that haploinsufficiency is the pathogenetic mechanism behind the FOXG1 syndrome [5–7], and, therefore, that haploinsufficiency of regulatory elements, e.g., enhancers, could be the pathogenetic mechanism of the breakpoints in the *FOXG1* regulatory domain.

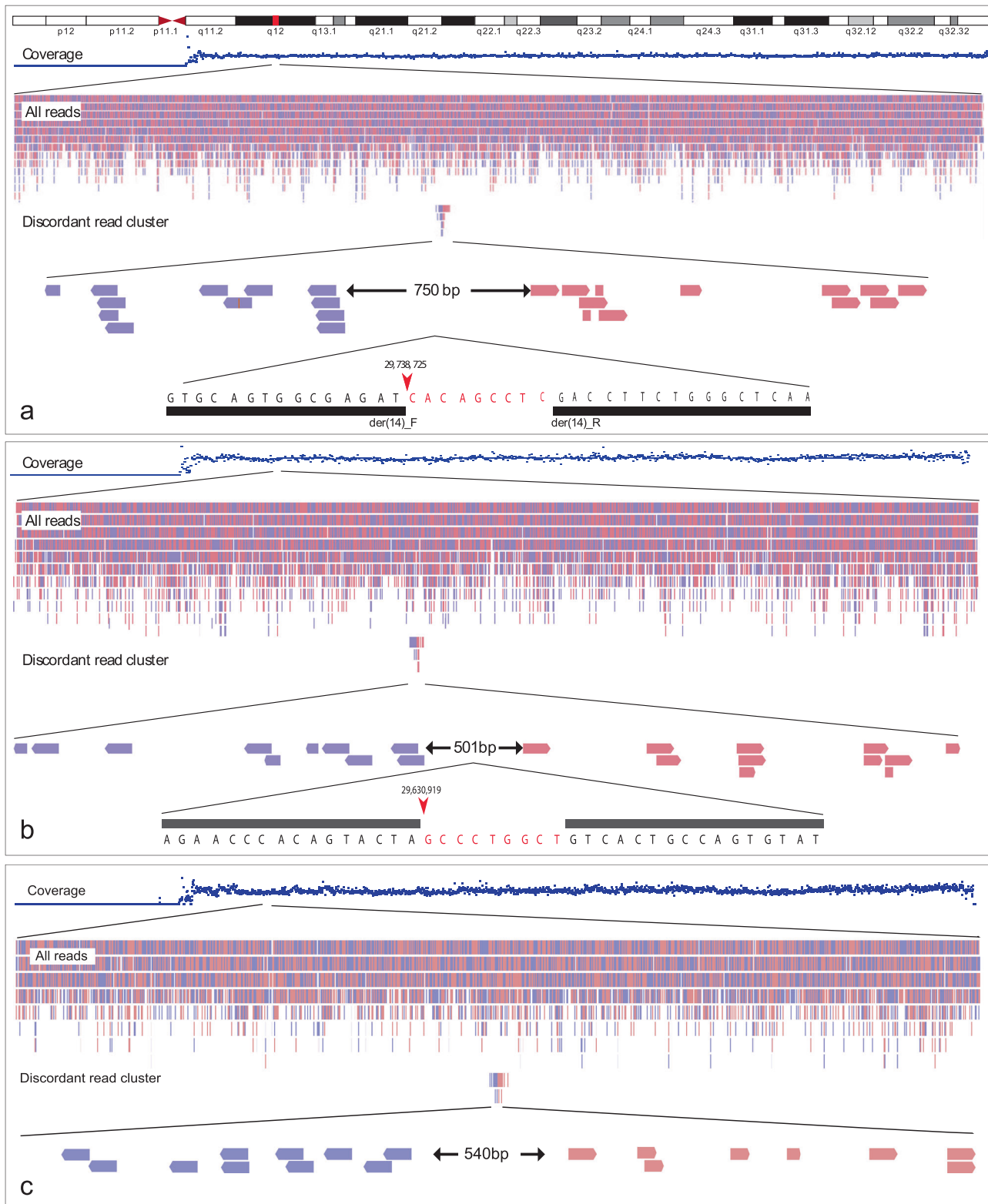


Fig. 2 Mapping of translocation breakpoints by M-PS. The MP-S read coverage does not reveal imbalances at the breakpoints sites on chromosome 14 (upper tracks in P1-P3). **a** $t(9;14)(q22;q12)dn$ in P1. The chr14 breakpoint is refined to a 750 bp-region by two oppositely oriented clusters of mate-pair reads. The red arrow shows the breakpoint obtained from Sanger sequencing data right after nt 29,738,725. The bases coloured in red are the deleted ones. The black bars illustrate

the BLAT results of the forward and reverse sequences in the UCSC Genome Browser. **b** $t(4;14)(q26;q12)dn$ in P2. The chr14 breakpoint is refined to a 501 bp-region. Sanger sequencing with forward and reverse primer pairs reveals a 9 bp deletion (red bases) associated with the breakpoint right after nt 29,630,919 (red arrow). **c** $t(2;14)(q36.1;q12)dn$ in P3. The breakpoint is refined to 540 bp-region. Sanger sequencing data is not available for this patient

On the other hand, a previous expression study of cultured fibroblast cells derived from FOXG1 syndrome patients showed that the expression of *FOXG1* at the transcription level was elevated in patients with a deletion of *cis*-regulatory elements. The authors inferred that the clinical relevant regulatory element(s) downstream of *FOXG1* within the SRO are silencer elements (HCNE1-4) (Fig. 1) [8]. However, *FOXG1* expression is very low in transformed fibroblast cell lines (<http://www.gtexportal.org/home/>) emphasising that this cell line may not be a reliable model to study *FOXG1* expression. Indeed, expression studies of whole blood from two patients with overlapping intergenic deletions (Ellaway case 2,3) revealed diminished *FOXG1* expression [22], whereas there was normal *FOXG1* expression in a lymphoblastoid cell line from a third deletion patient (Ellaway case 1²², Fig. 1). These deviating results might reflect that cell lines are poor substitutes for studying *FOXG1* expression since it is a tissue-specific transcription factor involved in early embryonic brain development.

Moreover, over-expression of *FOXG1* via retroviral expression in avian embryos leads to extreme outgrowth of the telencephalon and mesencephalon, and thickening of the neuroepithelium due to suppressed normal programmed cell death as well as inducing a persistent undifferentiated state [25, 26]. This resembles the phenotype of duplications involving *FOXG1*, ranging from a normal phenotype to severe developmental delay, where the majority are normocephalic with a normal brain imaging and intact corpus callosum morphology (Table 1) [27, 28]. Thus, removal of silencers may not be the best explanation for the generation of the FOXG1 syndrome.

The SRO identified by Allou et al. [8], with at least five *cis*-acting regulatory elements including one neural tube-specific enhancer (hs598), and four elements (HCNE1-4) proposed to be silencers [8], is located in the distal part of Domain A. However, this SRO also involves the hESC-specific boundary (hESCb_2291) between Domains A and B (Fig. 1). The deleted region in Patient 5, who does not present features of classic FOXG1 syndrome (Table 1), refines the SRO region to chr14:29,875,672-30,173,942 (darker shade in Fig. 1). Importantly, the five regulatory elements and rs1191551 (a SNP that physically interacts with *FOXG1* and regulates its expression [29]), are still retained within this refined SRO. The proximal limit of the SRO is supported by two of our de novo translocation breakpoints (P1, P2) as well as by two recently mapped BPs (DGAP245, NIJ1) [7] (Fig. 1, Supplementary Fig. 4). The proximal limit of the SRO is also in line with a deleted/duplicated region observed in several healthy individuals (Database of Genomic Variants, brown bar, Fig. 1), which removes an active enhancer (element_566). Thus, removal of at least one experimentally proven enhancer is not

clinically critical (assuming that these DGV-deletions are of meiotic origin and present in the brain). Kortüm et al. [16] proposed two putative evolutionary highly conserved *cis*-regulatory elements (hs556 and hs342) as main elements regulating *FOXG1* expression. Whereas hs566 is within the region deleted in healthy individuals, the forebrain-specific enhancer hs342 is at the proximal border of the SRO and might thus also be a good candidate for a phenotype relevant *cis*-regulatory element.

Patient 4 was one of the patients carrying a deletion used to exemplify the enhancer adoption hypothesis [10]. According to this hypothesis, regulatory elements in Domain B (i.e., hs433) are brought into physical contact with *FOXG1*, leading to ectopic regulation of the gene and consequently the FOXG1 syndrome [10]. Accordingly, topological domain boundary disruption (TDBD) would be a better explanation for misexpression of *FOXG1* than the removal of silencer elements. However, this deletion and the two other described deletions remove the SRO and hence the five aforementioned regulatory elements. Importantly, both Patient 4 and the translocation carriers (Patients 1–3) have clinical features similar to *FOXG1* intragenic point mutation carriers. Moreover, there is no clinical difference between Patient 4, invoked to represent enhancer adoption (of enhancer hs433), and a patient (case 1) reported by Ellaway et al. [22] in which a microdeletion removed both the hESC-specific boundary and hs433 (Table 1) (Fig. 1).

These genotype–phenotype correlations are further supported by DNA–DNA interaction data. Although Virtual 4C-profiles and HiC-heatmaps indicate that Domains A and B are two separate domains, with a stronger interaction between *FOXG1* and Domain A, chromatin loops that link *FOXG1* with Domain B are evident in several cell lines, supporting that Domain B is also a part of the normal *FOXG1* regulatory landscape (Supplementary Fig. 4). Indeed, the splitting into Domains A and B is only present in two (embryonic) out of 14 tissues and 7 cell lines (Supplementary Fig. 2). This pattern is even more evident in the mouse cell lines. The heterozygous *Foxg1*^{+/-} mouse model resembles the human FOXG1 syndrome, with post-natal microcephaly, hyperlocomotion and severe impairment of visual function [30, 31]. The four regulatory elements in the critical SRO region are conserved in the mouse [8], whereas the mouse HiC-data support a fused domain with little evidence of the human hESC-specific TAD boundary (Supplementary Fig. 3).

Finally, the deletion in P6, which is associated with a different phenotype than the FOXG1 syndrome, fits with the notion that TAD-boundaries are LRPE-boundaries as well (Table 1).

Recently the enhancer-promoter interactions within a range of 500 kb reported by the ENCODE project [32] were used for the prediction of LRPE-associated chromosomal

Table 1 Clinical features associated with different *FOXP1* variants

| Removal of | <i>FOXP1</i> intragenic mutations [6] | | Balanced breakpoints including Patient 1/Patient 2/ Patient 3 | | Patients used for enhancer adoption hypothesis [10] Patient 4/DECIPHER 258584/252353 | | Case 1 [22] | | Patient 5 | Patient 6 | Duplications [27, 28] |
|-----------------------------------|---------------------------------------|-----------------------------|---|-----------------|--|----------------------|------------------------|---------------------|-----------|-----------|-----------------------|
| | <i>FOXP1</i> codons | Region distal to breakpoint | Region distal to breakpoint | hESCb_2291, SRO | hESCb_2291, SRO and hs433 | Part of <i>PRKDI</i> | TAD distal to Domain B | <i>FOXP1</i> codons | | | |
| Normal pregnancy/ Neonatal period | + | (+) 7/8 | (+) 1/3, NA 2/3 | + | - | - | (+) 6/9 | | | | |
| Postnatal growth deficiency | + | (-)/8, NA7/8 | (+) 1/3, NA 2/3 | NA | - | + | NA 9/9 | | | | |
| Severe intellectual disability | + | (+) 8/8 | (+) 3/3 | + | - | - | (+) 1/9 | | | | |
| Developmental delay (severe) | + | (+) 8/8 | (+) 1/3, NA 2/3 | + | - | - | (+) 7/9 | | | | |
| Postnatal microcephaly | + | (+) 8/8 | (+) 3/3 | + | - | - | (+) 2/9 | | | | |
| Hypogenesis of corpus callosum | + | (+) 7/8 | (+) 2/3, NA 1/1 | + | - | NA | NA 9/9 | | | | |
| Epilepsy | + | (+) 7/8 | (+) 2/3, NA 1/3 | + | - | - | (+) 5/9 | | | | |
| Absent language | + | (+) 8/8 | (+) 3/3 | + | + | + | (+) 7/9 | | | | |
| Poor social interaction | - | (-) 6/8, NA 2/8 | (-) 1/3, NA 2/3 | NA | - | NA | NA 9/9 | | | | |
| Walking | - | (-) 8/8 | (-) 3/3 | - | + | With support | Delayed 7/9 | | | | |
| Facial dysmorphism | - | (+) 4/8 | (+) 1/3, NA 2/3 | + | + | + | (+) 5/9 | | | | |
| Poor sleep pattern | + | (+) 2/8, (-) 1/8, NA 5/8 | (+) 1/3, NA 2/3 | + | + | NA | (+) 2/9 | | | | |
| Inappropriate laughing | + | (+) 3/8, NA 5/8 | (+) 1/3, NA 2/3 | + | + | NA | (+) 1/9 | | | | |
| Strabismus | + | (+) 6/8, (-) 1/8, NA 1/8 | (+) 1/3, NA 2/3 | + | - | - | (+) 1/9 | | | | |
| Hypotonia | + | (+) 6/8, NA 2/7 | (+) 2/3 | + | - | - | NA 9/9 | | | | |
| Stereotypic movements | + | (+) 6/8, NA 2/8 | (+) 1/3, NA 2/3 | + | - | - | (+) 2/9 | | | | |
| Bruxism | + | (+) 5/8, (-) 1/8, NA 2/8 | (+) 3/3 | + | + | - | NA 9/9 | | | | |
| Feeding difficulty | + | 08-Aug | (+) 3/3 | + | - | - | NA 9/9 | | | | |
| Genital abnormalities | (+) 2/50 | (+) 1/8, (-) 1/8, NA 6/8 | (-) 1/3, NA 2/3 | NA | - | - | NA 9/9 | | | | |

breakpoints [33]. Our study illustrates a limitation of the ENCODE interaction data for LRPE-prediction in very large regulatory landscapes such as the *FOXG1* locus, where the SRO and several of the LRPE-associated breakpoints are located > 500 kb from *FOXG1* (Supplementary Fig. 4).

Also, factors outside of the SRO could be important for proper cis-acting regulation over long distances. Recently, small regions with tissue-specific increased interactions (frequently associating regions (FIREs) were described, that are enriched for disease-associated SNPs and overlap with active enhancers in a tissue-specific manner [11]. We note that the proximal part of the *FOXG1* regulatory domain is enriched for FIREs, including hippocampus, prefrontal cortex and neuronal progenitor cell-specific FIREs, which are embedded within the chromatin loops (Supplementary Fig. 4). The *FOXG1* critical regulatory region does not overlap significantly with FIREs, whereas the DGV-deletions/duplications in assumed healthy individuals involve several brain-specific FIREs (Supplementary Fig. 4). Thus, the consequences of mutating FIREs remain to be shown, e.g., by functional studies.

In conclusion, we provide clinical and molecular evidence against the enhancer adoption hypothesis suggested by Ibn Salem et al. [10]. Indeed, both human and mouse virtual 4C and HiC-heatmaps indicate that the regulatory domain of *FOXG1* includes both Domains A and B, arguing against the clinical importance of a selective loss of the hESC-specific TAD boundary. The combined clinical and molecular evidence support that the loss of regulatory elements within the refined SRO is the main cause of the FOXG1 syndrome in patients with structural rearrangements associated with LRPE.

Acknowledgements We thank the patients and family members for participating in this study.

Funding This work was supported by The Lundbeck Foundation [2013–14290]; the University of Copenhagen's Programme for Interdisciplinary Research (Global Genes, Local Concerns); The Danish Council for Independent Research—Medical Sciences [4183-00482B]; FAPESP—Fundação de Amparo à Pesquisa do Estado de São Paulo (Grants: CEPID-Human Genome and Stem Cell Research Centre 2013/08028-1; student fellowships 2011/14293-4 and 2013/01146-9); the BIH-Charité Clinical Scientist Programme funded by the Charité—Universitätsmedizin Berlin and the Berlin Institute of Health.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interests.

References

- Dixon JR, Selvaraj S, Yue F, *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*. 2012;485:376–80.
- Rao SSP, Huntley MH, Durand NC, *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. 2014;159:1665–80.
- Miyoshi G, Fishell G. Dynamic FoxG1 expression coordinates the integration of multipolar pyramidal neuron precursors into the cortical plate. *Neuron*. 2012;74:1045–58.
- Lupiáñez DG, Kraft K, Heinrich V, *et al.* HHS public access. 2016;161:1012–25.
- Shoichet SA, Kunde SA, Viertel P, *et al.* Haploinsufficiency of novel FOXG1B variants in a patient with severe mental retardation, brain malformations and microcephaly. *Hum Genet*. 2005;117:536–44.
- Alosi D, Klitten LL, Bak M, Hjalgrim H, Møller RS, Tommerup N. Dysregulation of FOXG1 by ring chromosome 14. *Mol Cytogenet*. 2015;8:24.
- Redin C, Brand H, Collins RL, *et al.* The genomic landscape of balanced cytogenetic abnormalities associated with human congenital anomalies. *Nat Genet*. 2016;49:36–45.
- Allou L, Lambert L, Amsellem D, *et al.* 14q12 and severe Rett-like phenotypes: new clinical insights and physical mapping of FOXG1-regulatory elements. *Eur J Hum Genet*. 2012;20:1216–23.
- Lettice LA, Daniels S, Sweeney E, *et al.* Enhancer-adoption as a mechanism of human developmental disease. *Hum Mutat*. 2011;32:1492–9.
- Ibn-Salem J, Köhler S, Love MI, *et al.* Deletions of chromosomal regulatory boundaries are associated with congenital disease. *Genome Biol*. 2014;15:423.
- Schmitt AD, Hu M, Jung I, *et al.* A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Rep*. 2016;17:2042–59.
- Dekker J, Marti-Renom MA, Mirny LA. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nat Rev Genet*. 2013;14:390–403.
- Dixon JR, Jung I, Selvaraj S, *et al.* Chromatin architecture reorganization during stem cell differentiation. *Nature*. 2015;518:331–6.
- Kimura-Yoshida C, Kitajima K, Oda-Ishii I, *et al.* Characterization of the pufferfish *Otx2* cis-regulators reveals evolutionarily conserved genetic mechanisms for vertebrate head specification. *Development*. 2004;131:57–71.
- Kleinjan DA, Van Heyningen V. Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am J Hum Genet*. 2005;76:8–32.
- Kortüm F, Das S, Flindt M, *et al.* The core FOXG1 syndrome phenotype consists of postnatal microcephaly, severe mental retardation, absent language, dyskinesia, and corpus callosum hypogenesis. *J Med Genet*. 2011;48:396–406.
- Papa FT, Mencarelli MA, Caselli R, *et al.* A 3 Mb deletion in 14q12 causes severe mental retardation, mild facial dysmorphism and rett-like features. *Am J Med Genet Part A*. 2008;146:1994–8.
- Ariani F, Hayek G, Rondinella D, *et al.* FOXG1 is responsible for the congenital variant of Rett syndrome. *Am J Hum Genet*. 2008;83:89–93.
- Mencarelli MA, Spanhol-Rosseto A, Artuso R, *et al.* Novel FOXG1 mutations associated with the congenital variant of Rett syndrome. *J Med Genet*. 2010;47:49–53.
- Yeung A, Bruno D, Scheffer IE, *et al.* 4.45 Mb microduplication in chromosome band 14q12 including FOXG1 in a girl with refractory epilepsy and intellectual impairment. *Eur J Med Genet*. 2009;52:440–2.
- Takagi M, Sasaki G, Mitsui T, Honda M, Tanaka Y, Hasegawa T. A 2.0Mb microdeletion in proximal chromosome 14q12, involving regulatory elements of FOXG1, with the coding region of FOXG1 being unaffected, results in severe developmental delay, microcephaly, and hypoplasia of the corpus callosum. *Eur J Med Genet*. 2013;56:526–8.

- 22 Ellaway CJ, Ho G, Bettella E, *et al.* 14q12 microdeletions excluding FOXP1 give rise to a congenital variant Rett syndrome-like phenotype. *Eur J Hum Genet.* 2012;21:522–7.
- 23 Goubau C, Devriendt K, Van der Aa N, *et al.* Platelet defects in congenital variant of Rett syndrome patients with FOXP1 mutations or reduced expression due to a position effect at 14q12. *Eur J Hum Genet.* 2013;21:1349–55.
- 24 Bisgaard A, Kirchhoff M, Tumer Z, *et al.* Additional chromosomal abnormalities in patients with a previously detected abnormal karyotype, mental retardation, and dysmorphic features. *Am J Med Genet.* 2006;140:2180–7.
- 25 Adesina MD, Veo BL, Girard Courteau G, *et al.* FOXP1 expression shows correlation with neuronal differentiation in cerebellar development, aggressive phenotype in medulloblastomas, and survival in a xenograft model of medulloblastom. *Hum Pathol.* 2015;46:1859–71.
- 26 Ahlgren S, Vogt P, Bronner-Fraser M. Excess FoxG1 causes overgrowth of the neural tube. *J Neurobiol.* 2003;57:337–49.
- 27 Brunetti-Pierri N, Paciorkowski A, Ciccone R, *et al.* Duplications of FOXP1 in 14q12 are associated with developmental epilepsy, mental retardation, and severe speech impairment. *Eur J Hum Genet.* 2011;19:102–7.
- 28 Amor D, Burgess T, Tan T, Pertile M. Questionable pathogenicity of FOXP1 duplication. *Eur J Hum Genet.* 2012;20:595–6.
- 29 Won H, Torre-Ubieta L, Stein JL, *et al.* Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature.* 2016;538:1–20.
- 30 Boggio EM, Pancrazi L, Gennaro M, *et al.* Visual impairment in FOXP1-mutated individuals and mice. *Neuroscience.* 2016;324:496–508.
- 31 Shen L, Nam H, Song P, Moore HAS. Results in impaired neurogenesis in the postnatal hippocampus and contextual memory deficits. *Hippocampus.* 2006;16:875–90.
- 32 Thurman R, Rynes E, Humbert H, *et al.* The accessible chromatin landscape of the human genome. *Nature.* 2012;489:75–82.
- 33 Zepeda-Mendoza CJ, Ibn-Salem J, Kammin T, *et al.* Computational prediction of position effects of apparently balanced human chromosomal rearrangements. *Am J Hum Genet.* 2017;101:206–17.
- 34 Visel A, Minovitsky S, Dubchak I, Pennacchio LA. VISTA enhancer browser—a database of tissue-specific human enhancers. *Nucleic Acids Res.* 2007;35:D88–92.