Author Manuscript

HHS Public Access

Author manuscript

J Med Genet. Author manuscript; available in PMC 2022 November 01.

Published in final edited form as:

J Med Genet. 2022 November ; 59(11): 1058–1068. doi:10.1136/jmedgenet-2021-108375.

Consolidation of the clinical and genetic definition of a *SOX4*-related neurodevelopmental syndrome

Marco Angelozzi[1], Anirudha Karvande[1], Arnaud N Molin[1], Alyssa L Ritter[2], Jacqueline M M Leonard[2], Juliann M Savatt[3], Kristen Douglass[3], Scott M Myers[3], Mina Grippa[4], Dara Tolchin[1], Elaine Zackai[2], Sarah Donoghue[2], Anna C E Hurst[5], Maria Descartes[5], Kirstin Smith[5], Danita Velasco[6], Andrew Schmanski[6], Amy Crunk[7], Mari J Tokita[7], Iris M de Lange[8], Koen van Gassen[8], Hannah Robinson[9], Katie Guegan[9], Mohnish Suri[10], Chirag Patel[11], Marie Bournez[12], Laurence Faivre[13], Frédéric Tran-Mau-Them[14,15], Janice Baker[16], Noelle Fabie[16], K Weaver[17], Amelle Shillington[17], Robert J Hopkin[18], Daniela Q C.M Barge-Schaapveld[19], Claudia AL Ruivenkamp[20], Regina Bökenkamp[21], Samantha Vergano[22], Maria Noelia Seco Moro[23], Aranzazu Díaz de Bustamante[24], Vinod K Misra[25,26], Kelly Kennelly[27], Caleb Rogers[28], Jennifer Friedman[29,30,31,32], Kristen M Wigby[29,30], Jerica Lenberg[30], Claudio Graziano[29,30,31,32,33], Rebecca C Ahrens-Nicklas[2], Veronique Lefebvre[1]

[1]Surgery/Division of Orthopaedic Surgery, The Children's Hospital of Philadelphia Research Institute, Philadelphia, Pennsylvania, USA

[2]Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA

[3]Autism & Developmental Medicine Institute, Geisinger, Danville, Pennsylvania, USA

[4]U.O. Genetica Medica, Universita di Bologna, Bologna, Italy

[5]Department of Genetics, The University of Alabama at Birmingham School of Medicine, Birmingham, Alabama, USA

[6]Department of Pediatrics, University of Nebraska Medical Center, Omaha, Nebraska, USA

[7]GeneDx Inc, Gaithersburg, Maryland, USA

[8]Department of Medical Genetics, University Medical Centre Utrecht Brain Centre, Utrecht, The Netherlands

[#]To whom correspondence should be addressed (AhrensNicklasR@chop.edu; lefebvrev1@chop.edu).

[§]Joint senior authors

Contributors. MA, CG, RCA and VL designed this study. MA, DT and AK performed most of the in silico and in vitro functional tests. MA, AK, ANM, ALR, JL, CG, RAN and VL analyzed the data and wrote the manuscript. All other authors helped recruit patients and provided clinical information. All authors read, provided feedback, and approved the manuscript.

Competing interests. AC and MJT are employees of GeneDx, Inc. Other authors have no competing interests.

Patient consent for publication. Parental/guardian consent was obtained for all patients.

Ethics approval. Individuals were identified in different centers worldwide in diagnostic or research settings. All subjects provided informed consent for publication as approved by each individual institution's local institutional review board. This includes the IRB of the corresponding authors' institution, The Children's Hospital of Philadelphia (IRB 20-018394).

Data availability statement. De-identified individual participant data and *in vitro* study materials, such as *SOX4* variant expression plasmids, will be made available upon request to the corresponding authors, subject to the participant data sharing plan and consent provided.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[9]Exeter Genomics Laboratory, Royal Devon and Exeter NHS Foundation Trust, Exeter, UK

[10]Clinical Genetics, Nottingham University Hospitals NHS Trust, Nottingham, UK

[11]Genetic Health Queensland, Royal Brisbane and Women's Hospital, Herston, Queensland, Australia

[12]Centres de référence Anomalies du Développement et Syndrome Malformatifs, Centre Hospitalier Universitaire de Dijon, Dijon, France

[13]Centre de Génétique, Centre Hospitalier Universitaire de Dijon Hôpital d'Enfants, Dijon, France

[14]Genetics of Developmental Disorders, INSERM - Bourgogne Franche-Comté University, UMR 1231 GAD Team, Dijon, France

[15]Functional Unit 6254 Innovation in Genomic Diagnosis of Rare Diseases, CHU Dijon Bourgogne, Dijon, France

[16]Genomics and Genetic Medicine, Children's Minnesota, Minneapolis, Minnesota, USA

[17]Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA

[18]Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA

[19]Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands

[20]Laboratory for Diagnostic Genome Analyses, Department of Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands

[21]Department of Pediatric Cardiology, Leiden University Medical Center, Leiden, The Netherlands

[22]Division of Medical Genetics and Metabolism, Children's Hospital of The King's Daughters, Norfolk, Virginia, USA

[23]Clinical Analysis Department, Hospital Universitario de Móstoles, Mostoles, Spain

[24]Department of Genetics, Hospital Universitario de Móstoles, Mostoles, Spain

[25]Department of Pediatrics, Division of Genetic, Genomic, and Metabolic Disorders, Children's Hospital of Michigan, Detroit, Michigan, USA

[26]Discipline of Pediatrics, Central Michigan University, Mount Pleasant, Michigan, USA

[27]Department of Pediatrics, Children's Hospital of Michigan, Detroit, Michigan, USA

[28]Department of Molecular and Medical Genetics, Oregon Health & Science University School of Medicine, Portland, Oregon, USA

[29]Department of Pediatrics, University of California San Diego, San Diego, California, USA

[30]Division of Genetics/Dysmorphology and Institute for Genomic Medicine, Rady Children's Hospital, San Diego, California, USA

[31]Department of Neurosciences, University of California San Diego, San Diego, California, USA

[32]Division of Neurology, Rady Children's Hospital, San Diego, California, USA

[33]U.O. Genetica Medica, AUSL della Romagna Rimini, Cesena, Italy

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Abstract

Background: A neurodevelopmental syndrome was recently reported in four patients with *SOX4* heterozygous missense variants in the high-mobility-group (HMG) DNA-binding domain. The present study aimed to consolidate clinical and genetic knowledge of this syndrome.

Methods: We newly identified seventeen patients with *SOX4* variants, predicted variant pathogenicity using *in silico* tests and *in vitro* functional assays, and analyzed the patients' phenotypes.

Results: All variants were novel, distinct and heterozygous. Seven HMG-domain missense and five stop-gain variants were classified as pathogenic or likely pathogenic (L/PV) as they precluded SOX4 transcriptional activity *in vitro*. Five HMG- and non-HMG-domain missense variants were classified as of uncertain significance (VUS) due to negative results from functional tests. When known, inheritance was *de novo* or from a mosaic unaffected or non-mosaic affected parent for patients with L/PV, and from a non-mosaic asymptomatic or affected parent for patients with VUS. All patients had neurodevelopmental, neurological, and dysmorphic features, and at least one cardiovascular, ophthalmologic, musculoskeletal or other somatic anomaly. Patients with L/PV were overall more affected than patients with VUS. They resembled patients with other neurodevelopmental diseases, including the *SOX11*-related and Coffin-Siris (CSS) syndromes, but lacked the most specific features of CSS.

Discussion: These findings consolidate evidence of a fairly non-specific neurodevelopmental syndrome due to *SOX4* haploinsufficiency in neurogenesis and multiple other developmental processes.

INTRODUCTION

SOX4 belongs to a family of transcription factors harboring a Sex-determining Region on the Y chromosome (SRY)-related high-mobility-group (HMG) DNA-binding domain[1]. In humans and most mammals, this SOX family comprises twenty members distributed into eight groups (SOXA to SOXH) based on sequence conservation. As a whole, the family specifies cell fate and differentiation in most lineages and controls nearly every biological process. Loss-offunction variants in twelve SOX genes have been linked to date to developmental disorders termed SOXopathies[2]. Variants include missense changes, most often located within the HMG domain, stop-gain alterations and microdeletions. For instance, *SRY* (SOXA) variants cause disorders of sex development (MIM 400044 and 400045)[3,4]; *SOX9* (SOXE) variants cause campomelic dysplasia and XY sex reversal (MIM 114290)[5,6]; and *SOX5* and *SOX6* (SOXD) variants cause the neurodevelopmental Lamb-Shaffer (MIM 616803) and Tolchin-Le Caignec (MIM 618971) syndromes, respectively[7,8].

SOX4 and *SOX11* (SOXC) heterozygous variants have been described in patients with neurodevelopmental syndromes, mild dysmorphisms and various other inconstant anomalies[9–17]. The two disease phenotypes are similar, an observation consistent with findings in animal models that *SOX4* and *SOX11* are co-expressed in various progenitor cell types and have additive or redundant roles in many developing organs, including the brain, skeleton, heart, and eye[18–24]. These SOXC-related diseases share features with Coffin-Siris syndrome (CSS, MIM 135900), often called BAFopathy because causal

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

variants occur in genes encoding chromatinremodeling BAF-complex components, such as *ARID1B* and *SMARCB1.* CSS is characterized by hypotonia, global developmental delay, intellectual disability, craniofacial dysmorphism, hypoplastic or absent fifth distal phalanges, hypertrichosis and sparse scalp hair[25]. Inconstant features include poor growth and various somatic defects. Based on phenotype overlap with CSS and evidence that *SOX4* and *SOX11* are BAF-complex targets[26], the *SOX4*-related and *SOX11*-related syndromes have been assigned CSS family acronyms (CSS10, MIM 618506, and CSS9, MIM 615866, respectively). The diseases, however, remain incompletely defined, especially the *SOX4*-related syndrome, which was described in only four patients, each with an HMG-domain missense variant. Recently, a similar disease was associated with a bi-allelic in-frame microdeletion within a functionally unknown SOX4 domain, suggesting that sequences outside the DNA-binding and transactivation domains could modulate SOX4 activity and thus undergo pathogenic alterations[27].

Here, we aimed to consolidate and expand the definition of the *SOX4*-related syndrome. We describe seventeen unrelated patients with novel *SOX4* heterozygous variants within and outside the HMG domain, predict variant pathogenicity *in silico* and *in vitro*, and discuss findings relative to other neurodevelopmental syndromes.

PATIENTS AND METHODS

Human subjects and genetic studies

This is a retrospective analysis of individuals with syndromic intellectual disability, whose exome or genome sequencing revealed a *SOX4* heterozygous variant. Two children were enrolled in the CSS/BAF pathway international registry[28]. Others were referred by their physicians or GeneMatcher[29]. Patients were assessed using standard developmental scales for neurological, behavioral and other systems' anomalies, and familial, gestational and developmental history. Data were de-identified and consent for publication was obtained per local institutional review board policies. *SOX4* variants are described based on the GRCh37/Hg19 genome build and ENSG00000124766.4 Ensembl *SOX4* transcript. They were classified using InterVar and criteria adjusted according to American College of Medical Genetics and Genomics (ACMG)Association of Molecular Pathologists (AMP) guidelines[30] and functional test results. Combined annotation-dependent depletion (CADD) scores were calculated.

In silico assessment of variant pathogenicity

Evolutionary conservation of *SOX4* sequences was assessed with ClustalW (MacVector16 software) using NCBI sequences (online supplemental tables S1 and S2). The mutation tolerance of SOX4 residues was assessed using MetaDome[31]. *SOX4* variants were queried in control individuals using gnomAD (versions v2.1.1 and v3.1)[32]. Effects of missense variants on protein structure and function were predicted using SWISS-MODEL[33], PEP-FOLD3[34] and PolyPhen-2[35]. HMG-domain variants are numbered in sequence analyses as previously described[36].

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

In vitro assessment of variant functionality

The stability, intracellular distribution and transcriptional ability of SOX4 variant proteins were tested in Neuro-2a mouse neuroblasts (ATCC CCL-131) and COS-1 monkey kidney fibroblasts (ATCC CRL-1650) following transfection with plasmids encoding human SOX4 wild-type and variant proteins. Details are provided in supplemental methods.

RESULTS

Patient recruitment

We collected genetic and clinical information from 17 unrelated patients, aged 28 months to 20 years, and having a neurodevelopmental syndrome and distinct *SOX4* heterozygous variants (online supplemental tables S4). Nine missense variants within the HMG domain, three missense variants outside this domain, and five stop-gain variants were identified (figure 1A). All missense variants affected residues highly intolerant to mutations, except p.Ala316Ser (figure 1B). All were different from those reported in previous patients[17](figure 1D–F). One – p.Asp461Glu – was reported in a gnomAD individual, for which mild disease cannot be excluded.

In silico analyses predict pathogenicity of most *SOX4* variants

As explained below, *in silico* analyses supported pathogenicity for eight of the twelve SOX4 missense variants and for all truncating variants.

Pathogenic and likely pathogenic variants (L/PV)—Five missense variants - p.Asn64Lys, p.Met67Val, p.Trp69Gly, and p.Trp97Gly or Cys - occurred at HMG-domain positions that are directly involved in DNA binding (8, 11, 13 and 41, respectively)[37–38] and are fully conserved in SOX4 vertebrate orthologues and other human SOX proteins, except that isoleucine fills the position 11 in SRY (figure 1C,D and G). Asn64 precedes the H1 to H3 α-helices that give the domain its characteristic L-shape (figure 1C,D). Its carboxyamide group facilitates DNA binding and bending by establishing a H bond with the fourth base of the CATTGT SOX motif[37] (online supplemental figure S1A). Having a longer side chain than asparagine and no carboxyamide group, lysine would likely dislodge the HMG domain from the minor groove. Met67, located in H1, and Trp97, located in H2, belong to a hydrophobic core of residues whose side chains insert into the DNA minor groove and pry it open[37]. The replacement of the long, sulfur-containing chain of Met67 by the shorter hydrophobic chain of valine should thus be detrimental (online supplemental figure S1B), as should be the replacement of the large aromatic Trp97 chain with short, aliphatic (glycine) or sulfur-containing (cysteine) chains (online supplemental c). As Trp69 serves a buttress role for the hydrophobic wedge[38] and is conserved among all SOX proteins, its replacement with glycine should be highly destabilizing (online supplemental figure S1C). According to these considerations, PolyPhen-2 predicted that p.Asn64Lys, p.Met67Val, p.Trp69Gly, p.Trp97Gly and p.Trp97Cys were probably damaging (online supplemental figure S1A–D). Noticeably, variants in these residues in other SOX proteins were linked to SOXopathies (online supplemental table S3). They include asparagine-to-lysine and tryptophan-to-cysteine substitutions, as in our patients.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Other HMG missense variants affected residues not involved in DNA binding. A p.Leu99Pro variant (HMG-domain position 43) affected a residue conserved in many SOX4 vertebrate orthologues and human SOX proteins (figure 1D,G). Leu99 has a solvent-exposed aliphatic chain, is located at the H2 end, and lies in a nuclear export signal (online supplemental figure S1D). Lacking a similar side chain, proline would unlikely functionally replace Leu99[40]. PolyPhen-2 predicted that p.Leu99Pro was probably damaging. Accordingly, an equivalent SOX5 variant was linked to Lamb-Shaffer syndrome (online supplemental table S3).

Two other HMG missense variants - p.Ala112Gly and p.Ala112Val (position 56) – impacted the same residue as in a previous patient[17] (p.Ala112Pro) and gnomAD individual (p.Ala112Thr). Ala112 stabilizes H3 through hydrogen bonds (online supplemental figure S1E) and is adjacent to a Glu residue interacting with POU domain-containing partners of SOX proteins in neuronal cells[41]. The one-hydrogen side chain of glycine has poor helix-forming propensity, and valine has a longer aliphatic chain than alanine. PolyPhen-2 predicted that both variants were probably damaging. Accordingly, Ala-to-Val substitutions in equivalent SOX9, SOX10 and SOX11 locations caused SOXopathies (online supplemental table S3), but no Alato-Gly substitution has yet been linked to a SOXopathy.

The five stop-gain variants truncated SOX4 in distinct regions (figure 1A). The p.Gly44Argfs*2 variant encoded a peptide of unknown function. The p.Tyr325*, p.Ser333* and p.Ser347* variants removed 128–150 residues, including the functionally essential transactivation domain (TAD), and p.Glu445* deprived TAD of its pivotal C-terminal segment[42].

Variants of uncertain significance (VUS)—A p.Lys132Arg variant occurred in the HMG-domain C-terminus (position 76), which stabilizes the L-shaped structure and features a nuclear import signal (figure 1A and online supplemental figure S1F). The lysine residue is conserved in SOX4 vertebrate orthologues and human SOXC relatives, but replaced by proline or arginine, as in this variant, in other human SOX proteins (figure 1D,G). Both lysine and arginine are positively charged, but lysine is slightly shorter and less hydrophilic. It undergoes more post-translational modifications than arginine, but it is unknown whether either residue is modified in SOX proteins. Lysine and arginine might thus have differential functions and regulations. While PolyPhen-2 predicted that p.Lys132Arg was possibly damaging, it is noteworthy that Arg-to-Trp and Arg-to-Gly substitutions in SRY caused sex reversal (online supplemental table S3), suggesting that arginine is functional, at least in SRY.

A p.Ala316Ser variant occurred in a SOX4 region neutral to missense mutations (figure 1B), unstructured and functionally uncharacterized (figure 1A and online supplemental figure S1G). Ala316 and flanking residues are well conserved in SOX4 placental orthologues, but not in lower vertebrates and human SOX11 and SOX12 (figure 1E,H). Two p.Ala316Gly and 1494 p.Ala316Val variants occurred in gnomAD individuals (figure 1I). Alanine is short and hydrophobic, while serine is slightly longer and polar. Thus, although p.Ala316Ser was not predicted by PEP-FOLD3 to affect the protein structure, a harmful effect cannot be

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

excluded, and PolyPhen-2 foresaw p.Ala316Ser as probably damaging (online supplemental figure S1G).

Two variants – p.Asp461Glu and p.Ser466Gly – targeted the SOX4 transactivation domain (TAD) (figure 1A). Asp461, Ser466 and most neighboring residues are conserved or semi-conserved in SOX4 vertebrate orthologues (figure 1F), and while Asp461 and neighbors are well conserved in human SOX11 and SOX12, Ser466 and neighbors are not (figure 1J). A gnomAD individual had a p.Asp461Glu variant, but none had a Ser466 variant (figure 1K). PEP-FOLD3 predicted that an α-helix harboring Asp461 and Ser466 could be partially destabilized by p.Asp461Glu, but not by p.Ser466Gly (online supplemental figure S1H,I). Both aspartate and glutamate are negatively charged. Their reciprocal substitutions are common and generally nonpathogenic, even though the shorter aspartate chain generates more rigid interactions with positively charged residues than the glutamate chain. Serine and glycine are small but, unlike glycine, serine has a polar, hydroxyl-containing chain allowing hydrogen-bond formation. PolyPhen-2 predicted that p.Asp461Glu was possibly damaging or benign, and p.Ser466Gly was probably or possibly damaging.

Functional assays consolidate pathogenicity predictions of most *SOX4* variants

Since *in silico* predictions of pathogenicity rely on knowledge of proteins and similar, but not necessarily identical, variants to those of interest, we further investigated the pathogenic likelihood of *SOX4* variants through functional tests. For this, we constructed mammalian expression plasmids for 3FLAG-tagged human SOX4 wild-type and variant proteins. We included all missense variants, except p.Met67Val (identified late in the study). We also included two previously identified Ala112 variants to compare them to those also impacting Ala112 in our cohort, and selected p.Gly44Argfs*2, p.Tyr325* and p.Glu445* as representative truncating variants.

Protein stability and intracellular distribution assays—To assess variant protein stability and intracellular distribution, we transfected Neuro-2a cells with SOX expression plasmids, treated them with or without lactacystin, and analyzed SOX4 cytoplasmic and nuclear levels. Lactacystin is a highly specific inhibitor of proteasome activity[43], a prevalent mode of intracellular protein degradation. Without lactacystin, all HMG missense variants predicted *in silico* to be pathogenic, i.e., p.Asn64Lys, p.Trp69Gly, p.Trp97Gly, p.Trp97Cys, p.Ala112Val, and the previously reported p.Ala112Pro, were severely under-represented in the nucleus compared to wild-type SOX4, and the p.Leu99Pro, p.Ala112Gly and p.Lys132Arg variants were slightly or insignificantly under-represented (figure 2A,B). As all variants were or tended to be over-represented in the cytoplasm, the nuclear/cytoplasmic ratios of all variants, but p.Lys132Arg, were significantly reduced compared to that of wild-type SOX4. While lactacystin treatment insignificantly affected wild-type SOX4, it allowed most variants to reach wild-type nuclear levels. The non-HMG missense variants were as abundant as wild-type SOX4, whereas the p.Gly44Argfs*2 peptide was undetectable, and the p.Tyr325* and p.Glu445* proteins were several times over-represented, especially in the nucleus. These results strongly suggested that the pathogenic HMG-missense variant proteins were highly susceptible to proteasomal degradation in the nucleus, whereas truncation facilitated SOX4 accumulation.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

DNA binding assays—We tested whether HMG missense variants could bind DNA in an electrophoretic mobility shift assay (EMSA). As DNA probe, we used a *Tubb3* (tubulin-β3) promoter sequence containing a validated SOXC motif[44]. As protein source, we used extracts from COS1 cells transfected with plasmids encoding SOX4 proteins truncated at residue 284 to increase their stability. While the wild-type, p.Ala112Gly and p.Lys132Arg proteins avidly bound DNA, the p.Leu99Pro and p.Ala112Thr variants weakly did, and the other variants did not (figure 3A and online supplemental figure S2A). Thus, most HMG missense variants were both unstable and functionally impaired.

Transactivation assays—To functionally test the variants in intact cells, we transfected Neuro-2a cells with SOX4 wild-type and variants plasmids and a pTubb3-Luc reporter featuring a −215/+54 *Tubb3* promoter containing several SOXC sites[44], or a 6FXO-p89-Luc reporter featuring six copies of an *Fgf4* enhancer containing SOX and POU-domain binding sites. This enhancer is directly targeted by SOX2 and OCT4 in pluripotent stem cells[45] and is synergistically activated by SOXC proteins and the POU-domain protein BRN2 *in vitro*[42]. No HMG-domain variants activated pTubb3-Luc, except p.Ala112Gly and p.Lys132Arg, which were partially and fully active, respectively (figure 3B and online supplemental figure S2B). Wild-type SOX4, p.Ala112Gly and p.Lys132Arg powerfully synergized with BRN2 to activate 6FXO-p89-Luc, whereas p.Leu99Pro and p.Ala112Thr were weak, and the other variants inactive (figure 3C). The non-HMG missense variants showed wild-type activity on either reporter (figure 3D,E), whereas p.Gly44Argfs*2 and p.Tyr325* were inactive, and p.Glu445* weakly activated pTubb3-Luc (figure 3F,G). These results asserted pathogenicity to all HMG missense variants, but p.Ala112Gly and p.Lys132Arg, and to all nonsense variants, but not to the non-HMG missense variants.

Since several HMG missense and nonsense variants were not just inactive, but reduced the pTubb3-Luc intrinsic activity and 6FXO-p89-Luc activation by BRN2, we presumed that they were perhaps interfering with the activity of endogenous proteins. Since patients were heterozygous for their *SOX4* variant, we explored the possibility of dominant-negative activity of variants by transfecting Neuro-2a cells with 6FXO-p89-Luc, one dose of near-saturating SOX4 wild-type plasmid and an additional dose of SOX4 wild-type or variant plasmid (figure 3H). Two doses of wild-type SOX4 were slightly more potent than one dose in activating 6FXO-p89-Luc. One dose of p.Asn64Lys, p.Trp69Gly, p.Trp97Gly, pAla112Val, pTyr325* or p.Glu445* significantly reduced the activity of wild-type SOX4, whereas one dose of p.Ala112Gly and p.Lys132Arg did not. Thus, pathogenicity of some variants may result from lack of intrinsic activity and from interference with wild-type SOX4 activity.

Finally, we tested the ability of representative variants to activate *Tubb3* and *Hes5* (HESfamily transcription factor-5) in Neuro-2a cells, as both genes are SOXC targets in neuronal cells[44,46]. In agreement with other findings, both genes were upregulated by wild-type SOX4, p.Ala112Gly and p.Lys132Arg, but not by p.Trp97Gly and pAla112Val, and were downregulated by p.Tyr325* and p.Glu445* (figure 3I).

Together, functional and *in silico* assays supported pathogenicity of many novel *SOX4* variants, allowing upgrading of their ACMG scores to 4 or 5. However, they left

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

the p.Ala112Gly and p.Lys132Arg HMG variants, and p.Ala316Ser, p.Asp461Glu and p.Ser466Gly non-HMG variants as VUS, with ACMG scores of 3 (table 1).

Patients with *SOX4* L/PVs presented a phenotype overlapping with CSS and other neurodevelopmental syndromes

Five new patients with *SOX4* L/PVs (patients$^{L/PV}$) were female and the seven others were male (online supplemental table S4). While the previous cohort had two females and two males, all with a *de novo* L/PV[17], L/PVs of known inheritance in our new cohort were *de novo* (8/11), transmitted by an unaffected mosaic parent (2/11) or transmitted by a mildly affected father (1/11). Two patients$^{L/PV}$ had other genetic conditions beside a neurodevelopmental disease (*F11*-related bleeding disorder and *TTN*-related cardiomyopathy), and three others carried a VUS in another gene (*DYNC1H1*; *PHF8*, previously associated with neurodevelopmental disease; or *UBR5*).

Most patients$^{L/PV}$ had unremarkable gestation, except two, born prematurely, and two others, born at term following intrauterine growth retardation. The main phenotype of all patients$^{L/PV}$ included a syndromic neurodevelopmental disorder with hypotonia (7/12), borderline-to-mild intellectual disability (8/8 old enough for assessment), behavioral issues (12/12), speech delay (12/12), and fine (9/12) and gross (9/11) motor delay (table 2 and online supplemental table S4). Occipitofrontal circumference and postnatal growth were mostly normal (11/12 and 9/12, respectively). All patients$^{L/PV}$ exhibited dysmorphic features, such as a tall forehead possibly associated with metopic suture anomalies (5/12), epicanthal folds (4/12), and a wide mouth with full lips and marked philtrum (9/12). Various other malformations were reported, including palatal anomalies (5/12), retrognathia (3/12), and cardiac defects (6/12, excluding the patient with a *TTN* variant). The latter included ventricular septal defects (4) and vascular anomalies (6). Visual impairment was frequent (10/12, mainly myopia and strabismus) and hearing impairment occasional (3/12). Most of these features were fairly non-specific as they are common in Coffin-Siris syndrome, the *SOX11*-related syndrome, and various other neurodevelopmental syndromes. The majority of the more specific features of CSS (fifth-finger nail hypoplasia, corpus callosum agenesis and hypertrichosis and hirsutism) were not reported in this cohort, whereas some reported features (mouth shape (9/12), fifth ray anomalies (4/12) and corpus callosum dysgenesis (2/5)) were evocative of this disease.

Patients with *SOX4* VUS exhibited a milder phenotype

Given that *in silico* and *in vitro* tests may not consider or recapitulate the environment of the SOX4 protein in human development, we also analyzed the phenotype of patients^{VUS}. All were male, and when known (3/5), variant inheritance was from a non-mosaic, asymptomatic or mildly affected mother (table 2 and online supplemental table S4). One patient^{VUS} had a heterozygous L/PV in *SLC39A5*, explaining his severe myopia. Three others acquired a heterozygous variant in another gene, independently of their *SOX4* variant: a VUS in *CHD4*, linked to intellectual disability with variable cardiac, skeletal and urogenital defects; a microdeletion in *MACROD2*, not linked to a disease or developmental process; and a microdeletion including *SOBP*, linked to an autosomal recessive form of intellectual disability. Most patients^{VUS} exhibited the most constant features of the

Author Manuscript

SOX4-related syndrome, i.e., global developmental delay, intellectual disability and facial dysmorphisms, and all had at least one inconstant feature of this syndrome. Most defects were or tended to be less frequent in these patients than in patients$^{L/PV}$, such that patients^{VUS} had an overall significantly milder phenotype (table 2 and online supplemental figure S3). Together, clinical and biochemical findings left open the question of whether the phenotype of patients^{VUS} was contributed by their *SOX4* VUS.

DISCUSSION

This study significantly consolidated the clinical and genetic definition of a disease previously described in only four patients. It revealed that both missense and truncating variants abolishing SOX4 transcriptional activity *in vitro* likely cause global developmental delay, intellectual disability, facial dysmorphisms and inconstant neurological and somatic anomalies in heterozygous patients. The syndrome overlaps with many neurodevelopmental syndromes, including CSS. However, features specific to CSS were either not reported or mild in our patients, such that these patients would not prompt an obvious CSS diagnosis. In contrast, the the spectrum of our patients' features was consistent with evidence from animal models that SOX4 is critical in many developmental processes. These observations should help manage future patients carrying rare *SOX4* variants in a concept of reverse phenotyping and personalized medicine. Further, our work emphasizes the usefulness of functional studies to properly interpret genetic variations for diagnostic purposes and to provide new insights into human pathology.

The *SOX4*-related phenotype closely resembles the *SOX11*-related phenotype both genetically and in terms of developmental, neurological and somatic findings. This resemblance corroborates in humans the observations made in animal models that *Sox4* and *Sox11* have greatly overlapping expression patterns and encode proteins that are closely related and have essential, additive and redundant roles in such key developmental processes as cardiac outflow tract septation and skeleton patterning, beside neurogenesis. The two syndromes were named CSS10 and CSS9, respectively, to reflect a possible link to CSS. This link combines evidence that SOXC genes are targets of the BAF complex, whose components are encoded by genes mutated in CSS, and the observation that patients display some CSS features. However, the legitimacy of this CSS-related classification of the SOXC-related diseases is questioned by evidence from the current study and previous ones that most patients with *SOX4* or *SOX11* variants did not exhibit the most specific characteristics of CSS, by the sharing of many features between SOXC-related, Coffin-Siris and various other syndromes, and by the fact that SOXC may be targets of, but not BAF components. This argument is however softened when one acknowledges that CSS is a clinically and genetically heterogenous disease, of which even the most specific features (sparse scalp hair, fifth-digit hypoplasia, hypotonia, hypertrichosis and corpus callosum agenesis) are not always present and, therefore, the CSS diagnosis is often made retrospectively, based on whole-genome or whole-exome sequencing. Similarly, the identification of *SOX4* variants in this cohort was also made retrospectively. The globally non-specific nature of these syndromes often raises the fundamental question of their right interpretation in terms of pathogenicity and clinical impact of the identified variants.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

In silico assessment of variant pathogenicity relies on structure/function and disease knowledge. For SOX proteins, such knowledge is only extensive for the family hallmark, the HMG DNA-binding domain. Thus, pathogenicity could be predicted with high confidence for most *SOX4* missense variants occurring within, but not outside, this domain. It is well established that the C-terminus of SOX4 is an essential transactivation domain, but knowledge of critical residues within this domain is still missing. Therefore, stop-gain variants occurring upstream or within this domain could be classified as L/PV, whereas missense variants within this domain could only be classified as VUS at the present time. Our *in vitro* functional tests, complemented by *in silico* predictions, facilitated upgrading of ACMG scores for 11 variants (from four to five for nine L/PV variants and from three [VUS] to four or five [L/PV] for two others). In contrast, they led to downgrading of the p.Ala112Gly HMG-missense variant from four to three. This variant was transmitted by an apparently healthy mother to a male with a mild non-specific neurodevelopmental disease. Previous structure/function data and the existence of three other Ala112 L/PV were supporting pathogenicity, but our *in vitro* tests detected mild or no functional consequence. We therefore classified this variant as VUS. It is nevertheless important to consider that failure to support pathogenicity does not exclude pathogenicity. Indeed, *in vitro* assays may not reproduce the human developmental environment *in vivo*, where SOX4 may undergo important regulatory post-translational modifications, interact with different partners, and target different sequences than in our *in vitro* tests.

Five variants whose pathogenicity was not supported by *in vitro* tests were associated with an overall milder phenotype than in patients$^{L/PV}$. While it is tempting to conclude that this phenotype was caused by other variants, one should also consider an oligogenic model. These patients could have genetic modifiers of the *SOX4*-related disease. Such modifiers could include the *CHD4*, *DYNC1H1*, *PHF8* or *UBR5* VUS detected in some patients, as L/PV variants in these genes cause neurodevelopmental diseases. Particularly, as *CHD4* and *PHF8* encode chromatin remodelers[47], genetic interaction between the *SOX4* and *PHF8/CHD4* variants might have caused or aggravated the *SOX4*-related disease. Thus, even if their direct clinical utilization remains questionable, *SOX4* VUS should not be dismissed, but should rather motivate investigations on patients' genetic make-up and possibly critical, yet-unknown *in vivo* mechanisms of SOX4 protein activity or regulation involving the variants.

Finally, our functional data suggested that the *SOX4*-related disease originates from *SOX4* haploinsufficiency, but possibly also from dominant-negative interference, e.g., by competition between the wild-type and variant proteins for DNA or protein interaction. It remains unknown, however, whether SOX4 is present *in vivo* in limiting or excess amount relative to DNA targets and functional partners, under which *SOX4* haploinsufficiency or dominant-negative interference would cause disease, respectively. Both situations may exist in different cells and developmental stages and explain why some processes are more affected than others.

In conclusion, this study corroborated the existence of a *SOX4*-related syndrome reflecting dependence of multiple developmental processes on tightly controlled SOX4 activity. It should stimulate comprehensive screening of patients with similar phenotypes and new

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

investigations to fully uncover pathogenetic mechanisms and identify treatments for affected patients.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments.

We thank the patients and their families who contributed to this study. We thank Alberic Gan for technical help in functional assays.

Funding. This work was funded by the Children's Hospital of Philadelphia (VL and RAN), National Institute of Health NIAMS R01-AR68308 grant (VL), National Institute of Health NINDS K08-NS105865 grant (RAN), and National Institute of Health NIMH R01MH074090 and U01MH119705 grants (SMM). MB and LF are members of the European Reference Network for Developmental Anomalies and Intellectual Disability (ERN-ITHACA). Genome sequencing was funded for one patient by an internal grant from Children's of Alabama and the HudsonAlpha Institute for Biotechnology Clinical Services Laboratory (ACEH, MD, KS). The inclusion of one patient was made possible through access to the France Genomic Medicine Plan 2025 database.

REFERENCES

1. Kamachi Y, Kondoh H. Sox proteins: regulators of cell fate specification and differentiation. Development 2013, 140:4129–4144. [PubMed: 24086078]
2. Angelozzi M, Lefebvre V. SOXopathies. Growing Family of Developmental Disorders Due to SOX Mutations. Trends Genet 2019, 35:658–671. [PubMed: 31288943]
3. Berta P, Hawkins JR, Sinclair AH, Taylor A, Griffiths BL, Goodfellow PN, Fellous M. Genetic evidence equating SRY and the testis-determining factor. Nature 1990, 348:448–450 [PubMed: 2247149]
4. Jager RJ, Anvret M, Hall K, Scherer G. A human XY female with a frame shift mutation in the candidate testis-determining gene SRY. Nature 1990, 348:452–454. [PubMed: 2247151]
5. Foster JW, Dominguez-Steglich MA, Guioli S, Kwok C, Weller PA, Stevanovic M, Weissenbach J, Mansour S, Young ID, Goodfellow PN, Brook JD, Schafer AJ. Campomelic dysplasia and autosomal sex reversal caused by mutations in an SRY-related gene. Nature 1994, 372:525–530. [PubMed: 7990924]
6. Wagner T, Wirth J, Meyer J, Zabel B, Held M, Zimmer J, Pasantes J, Bricarelli FD, Keutel J, Hustert E, Wolf U, Tommerup N, Schempp W, Scherer G. Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the SRY-related gene SOX9. Cell 1994, 79:1111–1120. [PubMed: 8001137]
7. Lamb AN, Rosenfeld JA, Neill NJ, Talkowski ME, Blumenthal I, Girirajan S, Keelean-Fuller D, Fan Z, Pouncey J, Stevens C, Mackay-Loder L, Terespolsky D, Bader PI, Rosenbaum K, Vallee SE, Moeschler JB, Ladda R, Sell S, Martin J, Ryan S, Jones MC, Moran R, Shealy A, Madan-Khetarpal S, McConnell J, Surti U, Delahaye A, Heron-Longe B, Pipiras E, Benzacken B, Passemard S, Verloes A, Isidor B, Le Caignec C, Glew GM, Opheim KE, Descartes M, Eichler EE, Morton CC, Gusella JF, Schultz RA, Ballif BC, Shaffer LG Haploinsufficiency of SOX5 at 12p12.1 is associated with developmental delays with prominent language delay, behavior problems, and mild dysmorphic features. Hum Mutat 2012, 33:728–740. [PubMed: 22290657]
8. Tolchin D, Yeager JP, Prasad P, Dorrani N, Russi AS, Martinez-Agosto JA, Haseeb A, Angelozzi M, Santen GWE, Ruivenkamp C, Mercimek-Andrews S, Depienne C, Kuechler A, Mikat B, Ludecke, H.J., Bilan F, Le Guyader G, Gilbert-Dussardier B, Keren B, Heide S, Haye D, Van Esch H, Keldermans L, Ortiz D, Lancaster E, Krantz ID, Krock BL, Pechter KB, Arkader A, Medne L, DeChene ET, Calpena E, Melistaccio G, Wilkie AOM, Suri M, Foulds N, Genomics England Research C, Begtrup A, Henderson LB, Forster C, Reed P, McDonald MT, McConkie-Rosell A, Thevenon J, Le Tanno P, Coutton C, Tsai ACH, Stewart S, Maver A, Gorazd R, Pichon O, Nizon M, Cogne B, Isidor B, Martin-Coignard D, Stoeva R, Lefebvre V, Le Caignec C. De Novo SOX6

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Variants Cause a Neurodevelopmental Syndrome Associated with ADHD, Craniosynostosis, and Osteochondromas. Am J Hum Genet 2020, 106:830–845. [PubMed: 32442410]

9. Tsurusaki Y, Koshimizu E, Ohashi H, Phadke S, Kou I, Shiina M, Suzuki T, Okamoto N, Imamura S, Yamashita M, Watanabe S, Yoshiura K, Kodera H, Miyatake S, Nakashima M, Saitsu H, Ogata K, Ikegawa S, Miyake N, Matsumoto N. De novo SOX11 mutations cause Coffin-Siris syndrome. Nat Commun 2014, 5:4011. [PubMed: 24886874]
10. Khan U, Study D, Baker E, Clayton-Smith J. Observation of Cleft Palate in an Individual with SOX11 Mutation: Indication of a Role for SOX11 in Human Palatogenesis. Cleft Palate Craniofac J 2018, 55:456–461. [PubMed: 29437512]
11. Timberlake AT, Jin SC, Nelson-Williams C, Wu R, Furey CG, Islam B, Haider S, Loring E, Galm A, Yale Center for Genome A, Steinbacher DM, Larysz D, Staffenberg DA, Flores RL, Rodriguez ED, Boggon TJ, Persing JA, Lifton RP. Mutations in TFAP2B and previously unimplicated genes of the BMP, Wnt, and Hedgehog pathways in syndromic craniosynostosis. Proc Natl Acad Sci U S A 2019, 116:15116–15121.
12. Diel H, Ding C, Grehn F, Chronopoulos P, Bartsch O, Hoffmann EM. First observation of secondary childhood glaucoma in Coffin-Siris syndrome: a case report and literature review. BMC Ophthalmol 2021, 21:28. [PubMed: 33430815]
13. Okamoto N, Ehara E, Tsurusaki Y, Miyake N, Matsumoto N. Coffin-Siris syndrome and cardiac anomaly with a novel SOX11 mutation. Congenit Anom (Kyoto) 2018, 58:105–107. [PubMed: 28787104]
14. Hempel A, Pagnamenta AT, Blyth M, Mansour S, McConnell V, Kou I, Ikegawa S, Tsurusaki Y, Matsumoto N, Lo-Castro A, Plessis G, Albrecht B, Battaglia A, Taylor JC, Howard MF, Keays D, Sohal AS, Collaboration DDD, Kuhl SJ, Kini U, McNeill A. Deletions and de novo mutations of SOX11 are associated with a neurodevelopmental disorder with features of Coffin-Siris syndrome. J Med Genet 2016, 53.152–162. [PubMed: 26543203]
15. Sekiguchi F, Tsurusaki Y, Okamoto N, Teik KW, Mizuno S, Suzumura H, Isidor B, Ong WP, Haniffa M, White SM, Matsuo M, Saito K, Phadke S, Kosho T, Yap P, Goyal M, Clarke LA, Sachdev R, McGillivray G, Leventer RJ, Patel C, Yamagata T, Osaka H, Hisaeda Y, Ohashi H, Shimizu K, Nagasaki K, Hamada J, Dateki S, Sato T, Chinen Y, Awaya T, Kato T, Iwanaga K, Kawai M, Matsuoka T, Shimoji Y, Tan TY., Kapoor S, Gregersen N, Rossi M, Marie-Laure M, McGregor L, Oishi K, Mehta L, Gillies G, Lockhart PJ,Pope K, Shukla A, Girisha KM, Abdel-Salam GMH, Mowat D, Coman D, Kim OH, Cordier MP, Gibson K, Milunsky J, Liebelt J, Cox H, El Chehadeh S, Toutain A, Saida K, Aoi H, Minase G, Tsuchida N, Iwama K, Uchiyama Y, Suzuki T, Hamanaka K, Azuma Y, Fujita A, Imagawa E, Koshimizu E, Takata A, Mitsuhashi S, Miyatake S, Mizuguchi T, Miyake N, Matsumoto N. Genetic abnormalities in a large cohort of Coffin-Siris syndrome patients. J Hum Genet 2019, 64:1173–1186. [PubMed: 31530938]
16. Wakim V, Nair P, Delague V, Bizzari S, Al-Ali MT, Castro C, Gambarini A, El-Hayek S, Megarbane A. SOX11-related syndrome: report on a new case and review. Clin Dysmorphol 2021, 30:44–49. [PubMed: 33086258]
17. Zawerton A, Yao B, Yeager JP, Pippucci T, Haseeb A, Smith JD, Wischmann L, Kuhl SJ, Dean JCS, Pilz DT, Holder SE. De Novo SOX4 Variants Cause a Neurodevelopmental Disease Associated with Mild Dysmorphism. Am J Hum Genet 2019, 104:246–259. [PubMed: 30661772]
18. Schilham MW, Oosterwegel MA, Moerer P, Ya J, de Boer PA, van de Wetering M, Verbeek S, Lamers WH, Kruisbeek AM, Cumano A, Clevers H. Defects in cardiac outflow tract formation and pro-B-lymphocyte expansion in mice lacking Sox-4. Nature 1996, 380:711–714. [PubMed: 8614465]
19. Sock E, Rettig SD, Enderich J, Bosl MR, Tamm ER, Wegner M. Gene targeting reveals a widespread role for the high-mobility-group transcription factor Sox11 in tissue remodeling. Mol Cell Biol 2004, 24:6635–6644. [PubMed: 15254231]
20. Bhattaram P, Penzo-Mendez A, Sock E, Colmenares C, Kaneko KJ, Vassilev A, Depamphilis ML, Wegner M, Lefebvre V. Organogenesis relies on SoxC transcription factors for the survival of neural and mesenchymal progenitors. Nat Commun 2010, 1:9. [PubMed: 20596238]
21. Potzner MR, Tsarovina K, Binder E, Penzo-Mendez A, Lefebvre V, Rohrer H, Wegner M, Sock E. Sequential requirement of Sox4 and Sox11 during development of the sympathetic nervous system. Development 2010, 137:775–784. [PubMed: 20147379]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

22. Bergsland M, Ramskold D, Zaouter C, Klum S, Sandberg R, Muhr J. Sequentially acting Sox transcription factors in neural lineage development. Genes Dev 2011, 25:2453–2464. [PubMed: 22085726]
23. Bhattaram P, Penzo-Mendez A, Kato K, Bandyopadhyay K, Gadi A, Taketo MM, Lefebvre V. SOXC proteins amplify canonical WNT signaling to secure nonchondrocytic fates in skeletogenesis. J Cell Biol 2014, 207:657–671. [PubMed: 25452386]
24. Pillai-Kastoori L, Wen W, Morris AC. Keeping an eye on SOXC proteins. Dev Dyn 2015, 244:367–376. [PubMed: 25476579]
25. Vasko A, Drivas TG, Schrier Vergano SA. Genotype-Phenotype Correlations in 208 Individuals with Coffin-Siris Syndrome. Genes (Basel) 2021, 12.
26. Feng W, Khan MA, Bellvis P, Zhu Z, Bernhardt O, Herold-Mende C, Liu HK. The chromatin remodeler CHD7 regulates adult neurogenesis via activation of SoxC transcription factors. Cell Stem Cell 2013, 13:62–72. [PubMed: 23827709]
27. Ghaffar A, Rasheed F, Rashid M, van Bokhoven H, Ahmed ZM, Riazuddin S, Riazuddin S. Biallelic in-frame deletion of SOX4 is associated with developmental delay, hypotonia and intellectual disability. Eur J Hum Genet 2021.
28. McCague EA, Lamichhane R, Holt N, Schrier Vergano SA. Growth charts for individuals with Coffin-Siris syndrome. Am J Med Genet A 2020, 182:2253–2262. [PubMed: 32851773]
29. Sobreira N, Schiettecatte F, Valle D, Hamosh A. GeneMatcher. a matching tool for connecting investigators with an interest in the same gene. Hum Mutat 2015, 36:928–930. [PubMed: 26220891]
30. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL, Committee ALQA. Standards and guidelines for the interpretation of sequence variants. a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 2015, 17:405–424. [PubMed: 25741868]
31. Wiel L, Baakman C, Gilissen D, Veltman JA, Vriend G, Gilissen C. MetaDome: Pathogenicity analysis of genetic variants through aggregation of homologous human protein domains. Hum Mutat 2019, 40:1030–1038. [PubMed: 31116477]
32. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, Gauthier LD, Brand H, Solomonson M, Watts NA, Rhodes D, Singer-Berk M, England EM, Seaby EG, Kosmicki JA, Walters RK, Tashman K, Farjoun Y, Banks E, Poterba T, Wang A, Seed C, Whiffin N, Chong JX, Samocha KE, Pierce-Hoffman E, Zappala Z, O'Donnell-Luria AH, Minikel EV, Weisburd B, Lek M, Ware JS, Vittal C, Armean IM, Bergelson L, Cibulskis K, Connolly KM, Covarrubias M, Donnelly S, Ferriera S, Gabriel S, Gentry J, Gupta N, Jeandet T, Kaplan D, Llanwarne C, Munshi R, Novod S, Petrillo N, Roazen D, Ruano-Rubio V, Saltzman A, Schleicher M, Soto J, Tibbetts K, Tolonen C, Wade G, Talkowski ME, Genome Aggregation Database C, Neale BM, Daly MJ, MacArthur DG. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 2020, 581:434–443. [PubMed: 32461654]
33. Bienert S, Waterhouse A, de Beer TA, Tauriello G, Studer G, Bordoli L, Schwede T. The SWISS-MODEL Repository-new features and functionality. Nucleic Acids Res 2017, 45:D313–D319. [PubMed: 27899672]
34. Lamiable A, Thevenet P, Rey J, Vavrusa M, Derreumaux P, Tuffery P. PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex. Nucleic Acids Res 2016, 44:W449–454. [PubMed: 27131374]
35. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. Nat Methods 2010, 7:248–249. [PubMed: 20354512]
36. Bowles J, Schepers G, Koopman P. Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators. Dev Biol 2000, 227:239255.
37. Werner MH, Huth JR, Gronenborn AM, Clore GM. Molecular basis of human 46X,Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex. Cell 1995, 81:705–714. [PubMed: 7774012]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

38. Racca JD, Chen YS, Maloy JD, Wickramasinghe N, Phillips NB, Weiss MA. Structure-function relationships in human testis-determining factor SRY: an aromatic buttress underlies the specific DNA-bending surface of a high mobility group (HMG) box. J Biol Chem 2014, 289:32410–32429.
39. Hou L, Srivastava Y, Jauch R. Molecular basis for the genome engagement by Sox proteins. Semin Cell Dev Biol 2017, 63:2–12. [PubMed: 27521520]
40. Pace CN, Scholtz JM. A helix propensity scale based on experimental studies of peptides and proteins. Biophys J 1998, 75:422–427. [PubMed: 9649402]
41. Merino F, Ng CKL, Veerapandian V, Scholer HR, Jauch R, Cojocaru V. Structural basis for the SOX-dependent genomic redistribution of OCT4 in stem cell differentiation. Structure 2014, 22:1274–1286. [PubMed: 25126959]
42. Dy P, Penzo-Mendez A, Wang H, Pedraza CE, Macklin WB, Lefebvre V. The three SoxC proteins--Sox4, Sox11 and Sox12--exhibit overlapping expression patterns and molecular properties. Nucleic Acids Res 2008, 36:3101–3117. [PubMed: 18403418]
43. Omura S, Crump A. Lactacystin: first-in-class proteasome inhibitor still excelling and an exemplar for future antibiotic research. J Antibiot (Tokyo) 2019, 72:189–201. [PubMed: 30755736]
44. Bergsland M, Werme M, Malewicz M, Perlmann T, Muhr J. The establishment of neuronal properties is controlled by Sox4 and Sox11. Genes Dev 2006, 20:3475–3486. [PubMed: 17182872]
45. Yuan H, Corbi N, Basilico C, Dailey L. Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. Genes Dev 1995, 9:2635–2645. [PubMed: 7590241]
46. Braccioli L, Vervoort SJ, Puma G, Nijboer CH, Coffer PJ. SOX4 inhibits oligodendrocyte differentiation of embryonic neural stem cells in vitro by inducing Hes5 expression. Stem Cell Res 2018, 33:110–119. [PubMed: 30343100]
47. Varga J, Kube M, Luck K, Schick S. The BAF chromatin remodeling complexes: structure, function, and synthetic lethalities. Biochem Soc Trans 2021, 49:1489–1503. [PubMed: 34431497]
48. Jauch R, Ng CK, Narasimhan K, Kolatkar PR. The crystal structure of the Sox4 HMG domain-DNA complex suggests a mechanism for positional interdependence in DNA recognition. Biochem J 2012, 443:39–47. [PubMed: 22181698]

Author Manuscript

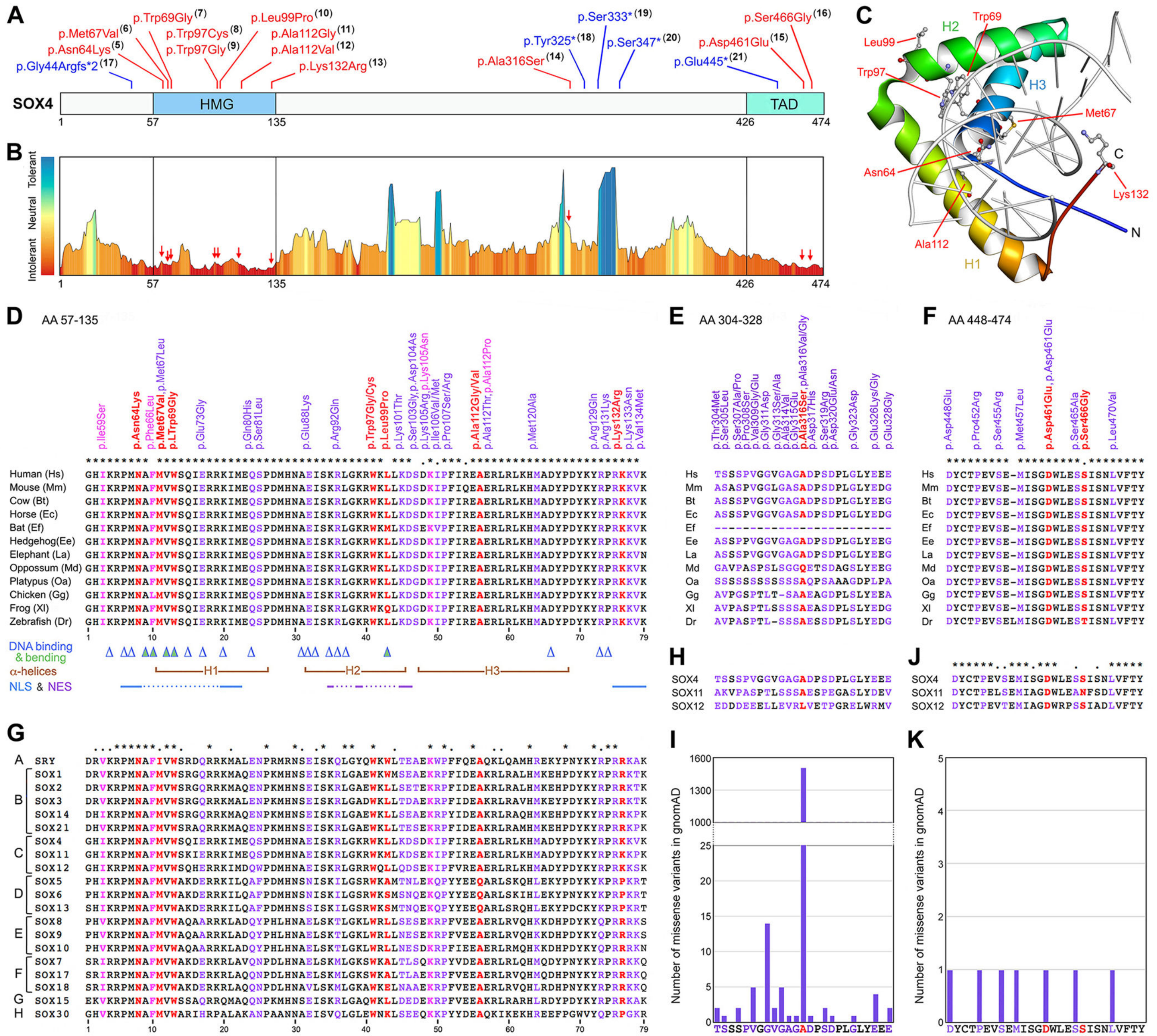

Figure 1. *In silico* analysis of SOX4 variants at the protein level.

(A) Schematic of the human SOX4 protein showing the location of the patients' variants. Numbers below the schematic indicate amino acid numbers. Frameshift and nonsense variants (blue) and missense variants (red) are indicated with superscripts referring to the numbers assigned to patients in the Supplemental Table 4.

(B) MetaDome plot showing the tolerance of SOX4 protein residues to missense mutations. Red arrows, location of the patients' missense variants.

(C) Swiss-Model rendering of the human SOX4 HMG domain–DNA complex (template 3u2b)[48]. The HMG domain (colored) forms three α-helices (H1, H2 and H3) that interact with DNA (grey shade) in the minor groove and induce a strong bend of the DNA helix. The N- and C-termini of the HMG domain are indicated (N and C, respectively). The

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

residues altered in patients are indicated, with their side chains depicted in grey (carbons), red (oxygen) and blue (amine group).

(D) Alignment of the HMG-domain sequences of various SOX4 vertebrate orthologues. SOX4 missense variants described in this study (red), described previously[17] (pink) and reported in gnomAD (purple) are shown above sequences. Amino acids matching variants are similarly colored in the sequences. Above the sequences, symbols denote fully conserved (asterisks) and semi-conserved (dots) amino acids. Below the sequences, residues important for DNA binding and bending are shown with open blue triangles and green-colored triangles, respectively. Brown brackets demarcate the H1, H2, and H3 α-helices. Key amino acids in the N-terminal and C-terminal nuclear localization signal sequences (NLS) and nuclear export signal sequence (NES) are shown with continued lines and linked with dotted lines.

(E) Alignment of the sequences encompassing Ala316 in SOX4 vertebrate orthologues.

(F) Alignment of the C-terminal TAD sequences of SOX4 vertebrate orthologues.

(G) Alignment of the HMG-domain sequences of all human SOX proteins. Proteins are listed based on group (A to H) classification.

(H) Alignment of the sequences encompassing Ala316 in the human SOXC proteins.

(I) Number of missense variants in the Ala316 region reported in gnomAD.

(J) Alignment of the C-terminal TAD sequences of the human SOXC proteins.

(K) Number of missense variants in the TAD C-terminus reported in gnomAD.

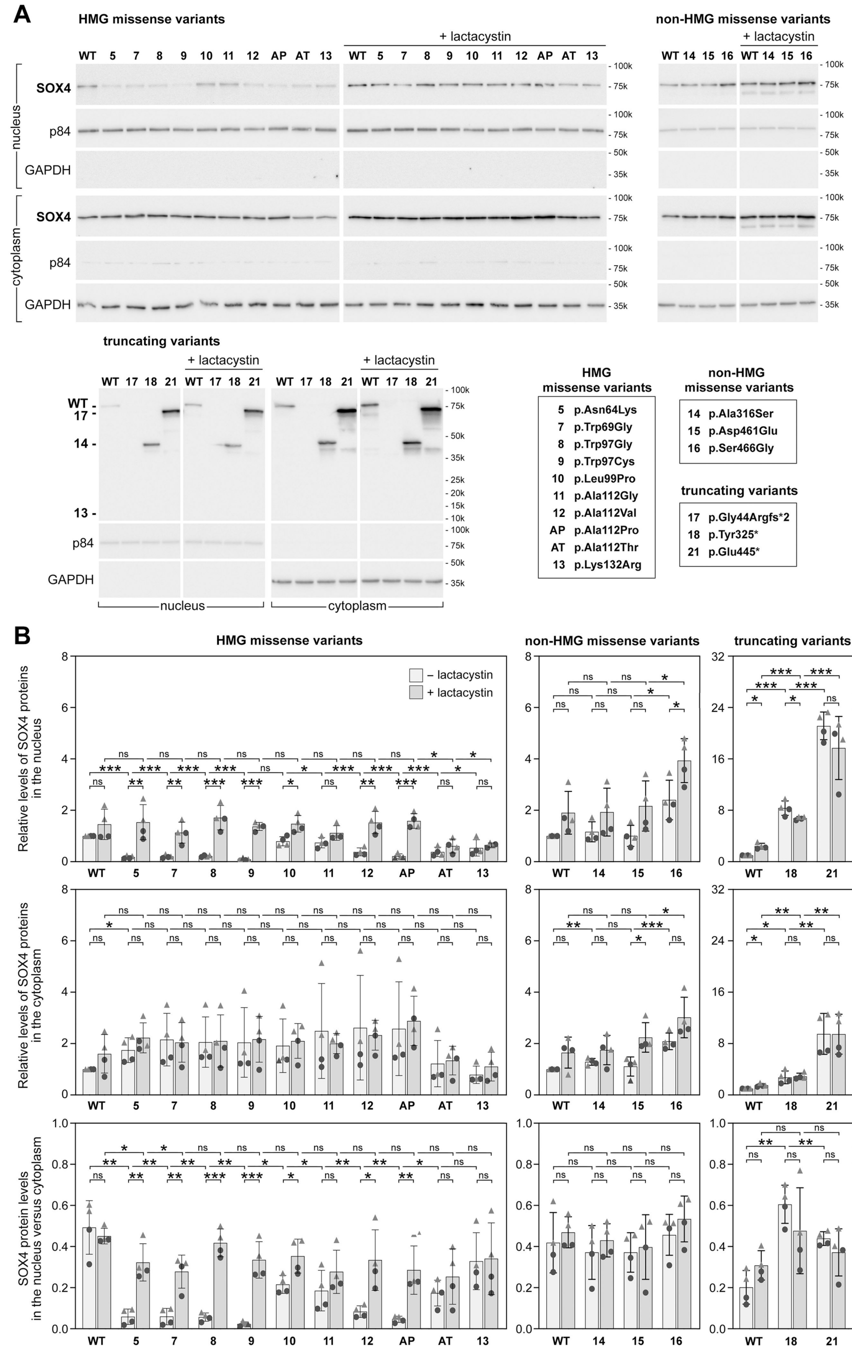

Figure 2. Tests of the stability and intracellular distribution of SOX4 variant proteins.

(A) Western blots of cytoplasmic and nuclear extracts from Neuro-2a cells transfected with plasmids encoding 3FLAG-tagged SOX4 wild-type (WT) and variant (1 to 17) proteins. Cells were treated without or with lactacystin for the last 6 h of a 20–24 h transfection period. SOX4 proteins were detected using a FLAG antibody. Signals obtained for P84 (nuclear) and GAPDH (cytoplasmic) demonstrate the quality of samples and their fairly even loading amounts. The Mr of protein standards is indicated. Images are representative of duplicate samples tested in each of at least two independent experiments. Variants are

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

numbered according to the patients in which they were detected (online supplemental table S4).

(B) Quantification of SOX4 protein levels visualized in western blots prepared as in panel a. The top and middle graphs show the nuclear and cytoplasmic levels of the various SOX4 protein types relative to wild-type SOX4 in the absence of lactacystin. Bottom graphs show the ratios of nuclear versus cytoplasmic SOX4 protein levels. Individual values are shown for four samples corresponding to duplicate cultures (same symbols) in two independent experiments (circles and triangles). The bars show the mean ± standard deviation for all four values. Two-tailed paired Student's T-tests were used to calculate the statistical significance of differences recorded between wild-type SOX4 and each variant, and between values obtained without and with lactacystin treatment for each protein type, as indicated with brackets (*, $p \le 0.05$; **, $p \le 0.01$; and ***, $p \le 0.001$; ns, not significant).

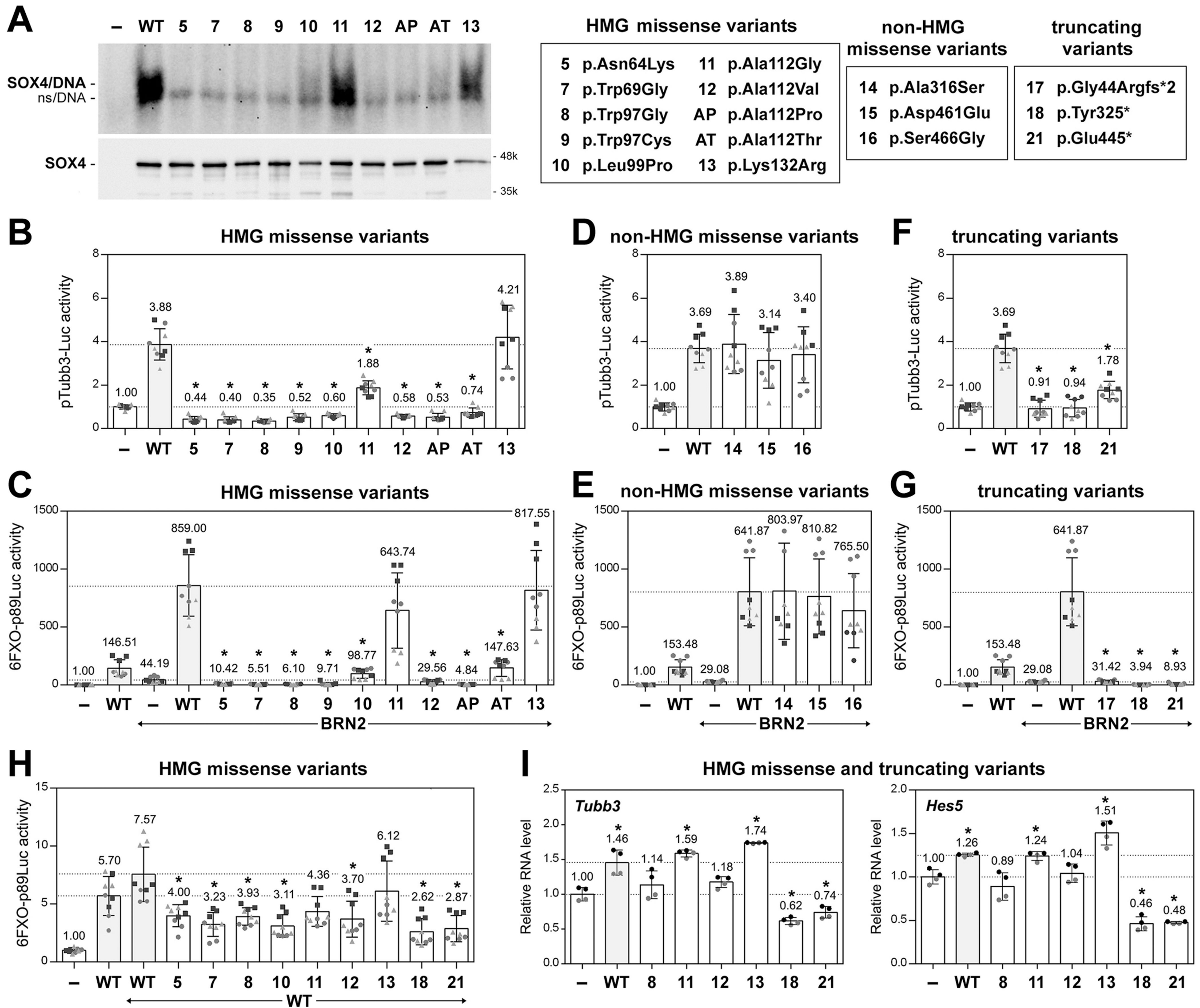

Figure 3. Functional tests of SOX4 variant proteins

(A) EMSA using a *Tubb3* DNA probe and whole-cell extracts from COS-1 cells transfected with expression plasmids for SOX4 wild-type and HMG missense variants. Top panel, picture of the gel showing the formation of complexes between the probe and SOX4 or a non-specific (ns) protein (the entire gel is shown in the online supplemental figure S2A). Bottom panel, western blot showing that the extracts contained fairly even amounts of SOX4 wild-type (WT) and variant proteins. The proteins were truncated at residue 214 and thus ran with an apparent Mr of 45k. Images are representative of data obtained with protein extracts from three independent experiments. Variants are numbered according to the patients in which they were detected (online supplemental table S4).

(B-G) Tests of the abilities of SOX4 wild-type and variant proteins to transactivate pTubb3-Luc or 6FXO-p89-Luc reporters. Neuro-2a cells were transfected with either reporter, a Nanoluc control reporter, and empty (–), SOX4 and/or BRN2 expression plasmids, as indicated. Reporter activities were normalized for transfection efficiency and are presented

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

as fold change increase relative to the activity of the reporter tested with the empty expression plasmid. Individual values are shown for triplicate cultures in three experiments (with squares, circles and triangles used to differentiate the experiments). The mean ± standard deviation of all data points are shown as bars and brackets. Average values are written. Asterisks indicate that a p value 0.05 was obtained in a Student's T-test comparing values for the SOX4 wild-type and variant proteins.

(H) Test of the abilities of SOX4 variants to interfere with wild-type SOX4 activity. Neuro-2a cells were transfected with 6FXO-p89-Luc, Nanoluc, and either 200 ng of empty expression plasmid (–) or 100 ng of SOX4 expression plasmid and 100 ng of SOX4 WT or variant expression plasmid, as indicated. Reporter activities were calculated and are presented as in (b to g). Asterisks indicate p values 0.05 in a Student's T-test comparing values obtained when the SOX4 wild-type plasmid was tested in one dose and when it was tested along with another dose of itself or a SOX4 variant expression plasmid.

(I) Test of the abilities of SOX4 wild-type and variant proteins to enhance the expression of neuronal markers in Neuro-2a cells. Cells were transfected with empty or SOX4 expression plasmids, as indicated. Total RNA was extracted the next day and used in qRT-PCR to measure the expression levels of *Tubb3* and *Hes5* relative to that of *Gapdh*. Data are presented as fold change increases obtained in cells transfected with a SOX4 plasmid compared to cells transfected with an empty plasmid. Individual values are shown for duplicate cultures in two independent experiments (with light and dark circles differentiating the experiments). The mean ± standard deviation obtained for all four values are shown as bars and brackets. Average values are written. Asterisks indicate p values 0.05 in a Student's T-test comparing values obtained for SOX4 wild-type or variant proteins to values obtained without protein.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1.

Summary of findings from in silico and functional tests of *SOX4* variant pathogenicity

Patients	Variants	PolyPhen-2/ domain	Stability	DNA binding	Transactivation	DN effect	SOXo-pathies	Pathogenicity (ACMG)
1	p.Asn64Lys	probably damaging/HMG	Very low	No	No	Yes	Same	L/PV (5)
2	p.Met67Val	probably damaging/HMG	ND	ND	ND	ND	Same	L/PV (5)
3	p.Trp69Gly	probably damaging/HMG	Very low	No	No	Yes	Yes	L/PV (5)
4	p.Trp69Gly	probably damaging/HMG	Very low	No	No	No	Yes	L/PV (5)
5	p.Trp97Cys	probably damaging/HMG	Very low	No	No	ND	Same	L/PV (5)
6	p.Leu99Pro	probably damaging/HMG	Low	Weak	Weak	Yes	Yes	L/PV (5)
7	p.Ala112Gly	probably damaging/HMG	Low	Yes	Yes	No	Yes	VUS (3)
8	p.Ala112Val	probably damaging/HMG	Very low	No	No	Yes	Same	L/PV (4)
9	p.Lys132Arg	possibly damaging/HMG	Unchanged	Yes	Yes	No	No	VUS (3)
10	p.Ala316Ser	probably damaging/ unknown	Unchanged	ND	Yes	ND	No	VUS (3)
11	p.Asp461Glu	Benign/unknown	Unchanged	ND	Yes	ND	No	VUS (3)
12	p.Ser466Gly	possibly damaging/ unknown	Unchanged	ND	Yes	ND	No	VUS (3)
13	p.Gly44Argfs*2	ND/HMG & TAD deleted	Unchanged	ND	no	ND	No	L/PV (5)
14	p.Tyr325*	ND/TAD deleted	High	ND	No	Yes	No	L/PV (5)
15	p.Ser333*	ND/TAD deleted	ND	ND	ND	ND	No	L/PV (5)
16	p.Ser347*	ND/TAD deleted	ND	ND	ND	ND	No	L/PV (5)
17	p.Glu445*	ND/TAD truncated	High	ND	No	Yes	No	L/PV (5)

DN, dominant-negative; ND = not determined; TAD = transactivation domain; Yes, missense change causing disease in the equivalent residue in another SOX gene; Same, identical missense change causing disease in the equivalent residue in another SOX gene; L/PV, likely pathogenic variant; VUS, variant of uncertain significance

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 2.

Summary of findings in individuals carrying *SOX4* variants in this and a previous[17] study

Clinical manifestations	HMG missense L/PV (previous study)	HMG missense L/PV (this study)	HMG missense L/PV (both studies)	Truncating L/PV (this study)	All L/PV (both studies)	All VUS (this study)
Known inheritance; [unknown inheritance]	4, *de novo;* [0]	5, *de novo* or affected parent; [2]	9, *de novo;* [2]	5, *de novo* or mosaic parent; [0]	14, *de novo* or mosaic parent; [2]	3, asymptomatic or mildly affected mother; [2][#]
Gross/fine motor delay[^]	3/4 (75%)	5/6[A] (83%)	8/10 (80%)	4/5[A] (80%)	12/15 (80%)	2/5 (40%)[#]
Speech delay[^]	4/4 (100%)	7/7 (100%)	11/11 (100%)	5/5 (100%)	16/16 (100%)	4/5 (80%)
Global developmental delay[^]	4/4 (100%)	6/6 (100%)	10/10 (100%)	4/5 (80%)	14/15 (93%)	4/5 (80%)
Intellectual disability[^]	4/4 (100%)	4/4 (100%)	8/8 (100%)	4/4 (100%)	12/12 (100%)	4/4 (100%)
Behavioral concerns[^]	0/0 (0%)	7/7 (100%)	7/7 (100%)	5/5 (100%)	12/12 (100%)	3/5 (60%)
Sensory concerns[^]	0/0 (0%)	1/7 (14%)	1/7 (14%)	2/5 (40%)	3/12 (25%)	1/5 (20%)
Microcephaly[^]	2/4 (50%)	3/7 (43%)	5/11 (45%)	0/4 (0%)	5/15 (33%)	0/5 (0%)
Brain anomalies on MRI[^]	2/2 (100%)	4/4 (100%)	6/6 (100%)	1/1 (100%)	7/7 (100%)	2/4 (50%)[#]
Seizures[^]	1/4 (25%)	2/7 (29%)	3/11 (27%)	0/5 (0%)	3/16 (19%)	1/3 (33%)
Hypotonia[^]	2/4 (50%)	4/7 (55%)	6/11 (55%)	3/5 (60%)	9/16 (56%)	4/4 (100%)
Ophthalmologic findings[^]	1/4 (25%)	6/7 (86%)	7/11 (64%)	4/5 (80%)	10/16 (63%)	2/5[B] (40%)
Ear-nose-throat findings[^]	1/4 (25%)	6/7 (86%)	7/11 (64%)	0/5 (0%)[*]	7/16 (44%)	0/4 (0%)
Gastrointestinal findings[^]	2/4 (50%)	3/5 (60%)	5/9 (56%)	3/5 (60%)	8/14 (57%)	1/4 (25%)
Genitourinary findings[^]	0/4 (0%)	3/6 (43%)	3/10 (30%)	2/5 (40%)	5/15 (33%)	1/5 (20%)
Musculoskeletal findings[^]	4/4 (100%)	5/6 (83%)	9/10 (90%)	3/5 (60%)	12/15 (80%)	3/5 (60%)
5th finger/toe malformations[^]	4/4 (100%)	2/6 (33%)	6/10 (60%)	1/5 (20%)	7/15 (47%)	2/5 (40%)
Facial dysmorphisms[^]	4/4 (100%)	6/7 (86%)	10/11 (91%)	5/5 (100%)	15/16 (94%)	5/5 (100%)
Cardiac anomalies[^]	1/4 (25%)	4/7 (57%)	5/11 (45%)	3/4[C] (75%)	8/15 (53%)	1/4 (25%)

Data are shown as numbers of subjects with feature over number of patients with available data. Percentages of patients with unknown inheritance (first data row) or with finding (other rows) are shown in brackets.

[^] Feature shared with CSS.

[A] One patient presented fine, but no gross motor delay.

[B] The patient with *SLC39A5*-related myopia was excluded.

[C] The patient with *TTN*-related cardiomyopathy was excluded.

[*]P 0.05 for differences between HMG-missense and truncating L/PV;

[#]P 0.05 for differences between L/PV and VUS (Student's *t*-test).

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript