De novo SOX6 variants cause a neurodevelopmental syndrome associated with ADHD, craniosynostosis, and osteochondromas

Tolchin, D.; Yeager, J.P.; Prasad, P.; Dorrani, N.; Russi, A.S.; Martinez-Agosto, J.A.; ... ; Genomics England Res Consortium

ARTICLE

De Novo SOX6 Variants Cause a Neurodevelopmental Syndrome Associated with ADHD, Craniosynostosis, and Osteochondromas

Dara Tolchin,[1] Jessica P. Yeager,[1] Priya Prasad,[2] Naghmeh Dorrani,[3] Alvaro Serrano Russi,[4] Julian A. Martinez-Agosto,[3,5] Abdul Haseeb,[1] Marco Angelozzi,[1] G.W.E. Santen,[6] Claudia Ruivenkamp,[6] Saadet Mercimek-Andrews,[7,8] Christel Depienne,[9] Alma Kuechler,[9] Barbara Mikat,[9] Hermann-Josef Ludecke,[9,10] Frederic Bilan,[11,12] Gwenael Le Guyader,[11,12] Brigitte Gilbert-Dussardier,[11,12] Boris Keren,[13] Solveig Heide,[13] Damien Haye,[14] Hilde Van Esch,[15] Liesbeth Keldermans,[16] Damara Ortiz,[17] Emily Lancaster,[17] Ian D. Krantz,[18] Bryan L. Krock,[19] Kieran B. Pechter,[19] Alexandre Arkader,[1] Livija Medne,[18] Elizabeth T. DeChene,[19] Eduardo Calpena,[20] Giada Melistaccio,[20] Andrew O.M. Wilkie,[20,21] Mohnish Suri,[21] Nicola Foulds,[22] Genomics England Research Consortium, Amber Begtrup,[23] Lindsay B. Henderson,[23] Cara Forster,[23] Patrick Reed,[23] Marie T. McDonald,[24] Allyn McConkie-Rosell,[24] Julien Thevenon,[25,26] Pauline Le Tanno,[25,26] Charles Coutton,[25,26] Anne C.H. Tsai,[27] Sarah Stewart,[27] Ales Maver,[28] Rudolf Gorazd,[28] Olivier Pichon,[29] Mathilde Nizon,[29,30] Benjamin Cogné,[29,30] Bertrand Isidor,[29,30] Dominique Martin-Coignard,[31] Radka Stoeva,[31] Véronique Lefebvre,[1,*] and Cédric Le Caignec[29,32,*]

SOX6 belongs to a family of 20 SRY-related HMG-box-containing (SOX) genes that encode transcription factors controlling cell fate and differentiation in many developmental and adult processes. For SOX6, these processes include, but are not limited to, neurogenesis and skeletogenesis. Variants in half of the SOX genes have been shown to cause severe developmental and adult syndromes, referred to as SOXopathies. We here provide evidence that SOX6 variants also cause a SOXopathy. Using clinical and genetic data, we identify 19 individuals harboring various types of SOX6 alterations and exhibiting developmental delay and/or intellectual disability; the individuals are from 17 unrelated families. Additional, inconstant features include attention-deficit/hyperactivity disorder (ADHD), autism, mild facial dysmorphism, craniosynostosis, and multiple osteochondromas. All variants are heterozygous. Fourteen are de novo, one is inherited from a mosaic father, and four offspring from two families have a paternally inherited variant. Intragenic microdeletions, balanced structural rearrangements, frameshifts, and nonsense variants are predicted to inactivate the SOX6 variant allele. Four missense variants occur in residues and protein regions highly conserved evolutionarily. These variants are not detected in the gnomAD control cohort, and the amino acid substitutions are predicted to be damaging. Two of these variants are located in the HMG domain and abolish SOX6 transcriptional activity in vitro. No clear genotype-phenotype correlations are found. Taken together, these findings concur that SOX6 haploinsufficiency leads to a neurodevelopmental SOXopathy that often includes ADHD and abnormal skeletal and other features.

[1]Department of Surgery, Division of Orthopaedic Surgery, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; [2]Department of Obstetrics and Gynecology, Wayne State University School of Medicine, Detroit, MI 48201, USA; [3]Division of Medical Genetics, Department of Pediatrics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA; [4]Division of Medical Genetics, Children's Hospital Los Angeles, Los Angeles, CA 90027, USA; [5]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA; [6]Department of Clinical Genetics, Leiden University Medical Centre, 2300 LC Leiden, the Netherlands; [7]Division of Clinical and Metabolic Genetics, Department of Pediatrics, The Hospital for Sick Children, University of Toronto, Toronto, ON M5G 1X8, Canada; [8]Institute of Medical Sciences, University of Toronto, Toronto, ON M5G 1X8, Canada; [9]Institute of Human Genetics, University Hospital Essen, University of Duisburg-Essen, 45147 Essen, Germany; [10]Institute für Humangenetik, Universitätsklinikum Düsseldorf, Heinrich-Heine-Universität, 40225 Düsseldorf, Germany; [11]Neurovascular Unit and Cognitive Disorders (EA-3808 NEUVACOD), Université de Poitiers, 86073 Poitiers, France; [12]Service de Génétique Clinique, Centre Hospitalier Universitaire de Poitiers, 86021 Poitiers, France; [13]Assistance Publique-Hôpitaux de Paris, Groupe Hospitalier Pitié-Salpêtrière, Département de Génétique, 75013 Paris, France; [14]Service de Génétique, Centre Hospitalier Universitaire de Nice Hôpital de l'Archet 2,151 route Saint Antoine de la Ginestière, 062002 Nice, France; [15]Center for Human Genetics, University Hospitals Leuven, 3000 Leuven, Belgium; [16]Laboratory for Molecular Diagnosis, Center for Human Genetics, University Hospitals Leuven, 3000 Leuven, Belgium; [17]University of Pittsburgh Medical Center, Children's Hospital of Pittsburgh, University of Pittsburgh School of Medicine, Pittsburgh, PA 15224, USA; [18]Roberts Individualized Medical Genetics Center, Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; [19]Division of Genomic Diagnostics, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; [20]MRC Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DS, UK; [21]Clinical Genetics Service, Nottingham University Hospitals NHS Trust, City Hospital Campus, Nottingham NG5 1PB, UK; [22]Wessex Clinical Genetics Services, University Hospital Southampton NHS Foundation Trust, Southampton SO16 5YA, UK; [23]GeneDx, Gaithersburg, MD 20877, USA; [24]Division of Medical Genetics, Department of Pediatrics, Duke University, Durham, NC 27707, USA; [25]Service de Génétique, Génomique, et Procréation, Centre Hospitalier Universitaire Grenoble Alpes, 38700 La Tronche, France; [26]INSERM 1209, CNRS UMR 5309, Institute for Advanced Biosciences, Université Grenoble Alpes, 38706 Grenoble, France; [27]Section of Genetics, Department of Pediatrics, University of Colorado School of Medicine and Children's Hospital Colorado, Aurora, CO 80045, USA; [28]Clinical Institute of Medical Genetics, University Medical Centre Ljubljana, 1000 Ljubljana, Slovenia; [29]Centre Hospitalier Universitaire Nantes, Service de Génétique Médicale, 44000 Nantes, France; [30]Université de Nantes, CNRS, INSERM, L'Institut du Thorax, 44000 Nantes, France; [31]Service de Cytogénétique, Centre Hospitalier Universitaire de Le Mans, 72037 Le Mans, France; [32]Centre Hospitalier Universitaire Toulouse, Service de Génétique Médicale, 31000 Toulouse, France

*Correspondence: lefebvrev1@email.chop.edu (V.L.), lecaignec.c@chu-toulouse.fr (C.L.C.)
https://doi.org/10.1016/j.ajhg.2020.04.015.
© 2020 American Society of Human Genetics.

Introduction

SRY-related high-mobility-group (HMG)-box-containing (SOX) genes code for transcription factors sharing at least 50% similarity with one another in a characteristic HMG-type DNA-binding domain.[1] The 20 SOX genes present in the human genome are distributed into eight groups (SOXA–SOXH) on the basis of the sequence conservation of their proteins within this domain. Further, SOX proteins belonging to the same group also feature significant conservation in sequences outside the HMG domain. SOX genes exhibit overlapping expression patterns and functions, especially among same-group members. Most SOX genes have been shown to participate pivotally in the control of cell fate and differentiation in one or multiple lineages, such that the SOX family as a whole has been implicated in almost every developmental, physiological, or pathological process. In line with these critical roles, to date, variants in half of the SOX genes have been associated with human developmental disorders, called SOXopathies.[2] The family founder, *SRY* (MIM: 480000), owes its name to its location in the sex-determining region of the Y chromosome. *SRY* is necessary for initiating male differentiation in the mammalian embryo,[3–5] and variants inactivating the gene cause disorders of sex development, including 46,XY sex reversal 1 (MIM: 400044). Besides *SRY* and *SOX3* (MIM: 313430), which are located on the Y and X chromosomes, respectively, all other SOX genes map to autosomal chromosomes, and disease-causing variants in these genes are most often heterozygous, *de novo*, and inactivating. For instance, the presence of such variants within and around *SOX9* (MIM: 608160) can cause campomelic dysplasia with or without XY sex reversal (CMPD1-SRA1 [MIM: 114290]); variants in *SOX10* (MIM: 602229) cause peripheral demyelinating neuropathy, central demyelination, Waardenburg syndrome and Hirschsprung disease (PCWH [MIM: 609136]), and the Waardenburg syndrome types 2E (MIM: 611584) and 4C (MIM: 613266); and variants in *SOX2* (MIM: 184429) cause anophthalmia, syndromic microphthalmia-3 (MCOPS3 [MIM: 206900]), optic nerve hypoplasia, and abnormalities of the central nervous system. Variants in several other SOX genes were recently shown to be responsible for intellectual disability and other features. Namely, the heterozygous loss of function of *SOX5* (MIM: 604975) causes Lamb-Shaffer syndrome (LAMSHF [MIM: 616803]), which is associated with global developmental delay, intellectual disability, and mild dysmorphic features;[6,7] *SOX11* (MIM: 600898) haploinsufficiency causes Coffin-Siris-like syndrome-9 (MIM: 615866), which is characterized by intellectual disability, microcephaly, and dysmorphic features;[8,9] and heterozygous *de novo* missense variants in *SOX4* (MIM: 184430) cause Coffin-Siris-like syndrome-10 (MIM: 618506), which features intellectual disability and mild facial and digital skeletal abnormalities.[10]

Human *SOX6* (MIM: 607257) is located at 11p15 and is expressed in a variety of tissues.[11,12] Its mouse ortholog was shown to share essential redundant functions with its closest SOXD relative, *Sox5*, in the differentiation of chondrocytes[13,14] but to be uniquely expressed and exert other important functions in skeletal myoblast differentiation,[15] cardiomyocyte proliferation,[16] erythroid cell maturation,[17] and regulation of insulin secretion.[18] Further, mouse *Sox6* is also pivotal in the developing central nervous system, where it regulates the development of cortical interneurons,[19,20] dopaminergic neurons in the substantia nigra pars compacta,[21] and oligodendrocytes.[22] Genome-wide association studies have linked human *SOX6* variants to a variety of adult clinical conditions, such as variability in bone mineral density,[23] carotid plaque formation,[24] high blood pressure,[25] and obesity.[26] In oncology, *SOX6* copy-number variants (CNVs) have been found in esophageal squamous cell carcinoma,[27] Ewing sarcoma,[28] and glioblastoma.[29] To date, however, no developmental disorder has been definitively associated with *SOX6* variants.

Here, we report 19 individuals who carry *SOX6* variants, share milestone delays and intellectual disability, and exhibit inconstant abnormalities, including mild dysmorphism, craniosynostosis, and osteochondromas; the individuals are from 17 unrelated families. The variants are private in each affected individual (apart from relatives), and all except two missense variants are predicted to abrogate or impair *SOX6* expression or protein activity. We thus propose that *SOX6* haploinsufficiency underlies a neurodevelopmental SOXopathy associated with other variable features.

Material and Methods

Recruitment of Subjects

The study cohort consisted of 19 individuals from 17 unrelated families originating from Belgium, Canada, France, Germany, the Netherlands, Slovenia, the UK, and the US. All individuals had molecular karyotyping, whole-exome sequencing (WES), or whole-genome sequencing (WGS) performed as part of local neurodevelopmental studies on developmental delay and intellectual disability or congenital abnormalities. Informed consent for participating in the genetic studies was obtained via protocols approved by institutional review boards of local hospitals. The parents or legal guardians of subjects provided consent for publication of all photographs shown in this study. No consanguinity was noted in the families. All affected individuals with a CNV or single-nucleotide variant (SNV) have been registered in the DECIPHER database (Table 1).

Genomic Variant Diagnosis

Genomic DNA was extracted from peripheral blood via standard protocols. Molecular karyotyping was conducted with various arrays according to the manufacturers' instructions (Table S1). Deletion breakpoints were mapped to the human genome assembly hg19 of the UCSC Genome Browser. Multiplex ligation-dependent probe amplification (MLPA) or qRT-PCR was used for confirming the genomic alteration in each individual and determining parental inheritance. WES was performed with target-enrichment designs according to the manufacturers' instructions. We filtered

Table 1.   Summary of Genetic and Clinical Data

Affected Individual	Internal ID; DECIPHER ID	Sex	Genomic Variant (GenBank: NM_ 033326.3; hg19)	Protein Change	Inheritance	Age	OFC	Weight	Height	ID	Facial Dysmorphism	Osteochondromas	Behavior	Other Features
1	PIT-1; 406933	male	deletion of exons 1 and 2	–	affected father	7 years	0 SD	+1 SD	+1 SD	mild	large ears, large nose	multiple osteochondromas	ADHD, autism, aggressive, emotional lability, tantrums, staring spells	submucous cleft palate
2	PIT-2	male	deletion of exons 1 and 2	–	affected father	3 years	−0.5 SD	−0.5 SD	−0.5 SD	mild	–	multiple osteochondromas	ADHD, autism, sleep disruption, staring spells, and left-sided weakness	bilateral nasolacrimal duct obstructions, food allergies
3	PIT-3	female	deletion of exons 1 and 2	–	affected father	12 months	+0.5 SD	+0.5 SD	−2.5 SD	normal	–	–	–	diastasis recti, umbilical hernia, cardiac rhabdomyoma
4	CHUN-1; 406934	male	deletion of exons 1–4	–	healthy father, mosaic 22%	10 years	+1 SD	+2 SD	+0.5 SD	mild	–	–	ADHD, anxiety	–
5	CHLM-1; 406929	female	deletion of exons 2 and 3	–	de novo	10 years	+1 SD	+3.5 SD	+4.5 SD	mild	–	–	–	arachnodactyly, precocious puberty
6	LJU-1; 406932	male	deletion of exons 2–13	–	affected father	7 years	−0.5 SD	0 SD	0 SD	mild to moderate	–	–	–	hypermetropia
7	IHG-1; 406930	male	deletion of exons 2–12	–	de novo	9 years, 4 months	+2 SD	+4 SD	+0.5 SD	mild	–	–	restlessness, short attention span, quick changes of mood, aggressive	–
8	CHUP-1; 406931	female	deletion of exons 5–7	–	de novo	12 years	−1 SD	−1 SD	−1 SD	moderate	oxycephaly (sagittal, metopic, and coronal craniosynostosis), micrognathia	–	attention deficit, anxiety	–
9	UK-1; 412103	male	c.242C>G	p.Ser81*	de novo	9 years, 7 months	−2 SD	−1 SD	0 SD	moderate	scaphocephaly (sagittal craniosynostosis), prominent occiput, hypertelorism	–	ADHD, lack of sense of danger, destructive, easily upset	short fifth fingers with clinodactyly, flat feet with valgus heels
10	UK-2; 412119	female	c.277C>T	p.Arg93*	unknown	6 years	−1 SD	−0.8 SD	+1 SD	mild	scaphocephaly (sagittal and left coronal craniosynostosis)	–	attention deficit, sleep disturbance, aggressive episodes	–
11	CHUN-2; 412120	male	c.293C>G	p.Ser98*	de novo	6 years	−1 SD	−1 SD	−1 SD	mild to moderate	high forehead	–	ADHD	ogival palate

(Continued on next page)

Table 1. *Continued*

Affected Individual	Internal ID; DECIPHER ID	Sex	Genomic Variant (GenBank: NM_033326.3; hg19)	Protein Change	Inheritance	Age	OFC	Weight	Height	ID	Facial Dysmorphism	Osteochondromas	Behavior	Other Features
12	GDX-3;[a] 412130	male	c.483G>C	p.Trp161Cys	*de novo*	13 years	+1 SD	+1 SD	+1.5 SD	severe	–	–	ADHD	obesity
13	GDX-1; 412121	female	c.718C>T	p.Gln240*	*de novo*	10 years	0 SD	+2 SD	+1 SD	mild	bitemporal narrowing, small mouth	–	ADHD, anxiety	bilateral inverted nipples, sensorineural hearing loss, vestibular dysfunction
14	GDX-2; 412122	female	c.878delC	p.Pro293Leufs*3	*de novo*	13 years	0 SD	+1.5 SD	−1 SD	mild	short palpebral fissures, hooded eyelids, hypertelorism, wide nasal bridge, low-set ears	–	ADHD	high arched palate; long, tapering fingers; hypotonia
15	PS-1; 412123	female	c.1728del	p.Glu577Argfs*29	*de novo*	27 years	0 SD	+4 SD	−1.5 SD	moderate	synophrys	–	–	short extremities, hirsutism, dental abnormalities, Hashimoto's thyroiditis
16	CHOP-1; 412124	female	c.1814T>C	p.Met605Thr	*de novo*	6 years	0 SD	+1 SD	+2 SD	mild	–	multiple osteochondromas	ADHD	dysgraphia, developmental delay
17	LEUV-1; 412126	female	c.1915T>A	p.Trp639Arg	*de novo*	3 years	−0.9 SD	+0.1 SD	+1.2 SD	mild to moderate	–	–	ADHD, autism	tremor hands
18	LEID-1;[b] 412127	male	c.2237C>T	p.Ser746Leu	*de novo*	27 years	−0.2 SD	0 SD	−0.5 SD	severe	long face, triangular; full eyebrows	–	autism, abnormal movements, automutilation, hyperventilation	–
19	CHUG-1	male	breakpoint in intron 13 of *SOX6*	–	*de novo*	2.5 years	+0.7 SD	+0.6 SD	+0.8 SD	mild to moderate	–	–	ADHD, oppositional disorder, self-endangerment, sleep disorders	gestural dyspraxia, graphomotor difficulties, troubles in neurosensory integration

Abbreviations are as follows: ID, intellectual disability; OFC, occipitofrontal circumference.
[a]Note that GDX-3 also harbored a hemizygous likely pathogenic *MECP2* variant that is consistent with his clinical presentation.
[b]In addition to trio WES, other targeted analyses were performed (e.g., *MECP2* and *TCF2* were analyzed by Sanger sequencing, and Smith-Magenis syndrome was analyzed by fluorescence *in situ* hybridization), but no other candidate variant was detected.

sequence variants in a stepwise manner to exclude synonymous variants, non-exonic SNVs, indels and variants with a minor allele frequency > 1% in gnomAD (version v2.1.1), the 1000 Genomes Project, and internal exome databases. WGS was performed for affected individual CHUG-1 at the Centre Hospitalier Universitaire de Grenoble and for affected individuals UK-1 and UK-2 by the Genomics England 100,000 Genomes Project. SNVs were confirmed by Sanger sequencing. The *SOX6* reference sequence GenBank: NM_033326.3 was used in naming these variants.

In Silico Assessment of SOX6 Variant Pathogenicity

SOX6 synonymous and missense variants present in control individuals were downloaded from gnomAD (version v2.1.1)[30] and analyzed with paired t tests. SOX sequences were downloaded from NCBI (Tables S2 and S3) and aligned with the ClustalW tool embedded in MacVector16 software (MacVector). The effects of missense variants on protein structure were predicted with HOPE,[31] SWISS-MODEL,[32] and PEP-FOLD3.[33] The best-scoring models were selected.

Functional Assessment of SOX6 Variants *In Vitro*

Expression plasmids for SOX6 missense variants were generated by PCR mutagenesis using appropriate primers (Table S4) and an expression plasmid for mouse wild-type (WT) SOX6 tagged with an N-terminal 3FLAG epitope.[12] Plasmid integrity was verified by Sanger sequencing. For reporter assays, HEK293 cells (CRL-1573; ATCC) were transfected in triplicate cultures with 3.5 μL ViaFect Transfection Reagent (Promega) containing 100 ng pNL1.1.TK[Nluc/TK] (NanoLuc *Renilla* luciferase control reporter plasmid), 400 ng mouse *Acan* [4xA1]-p89Luc (SOX6/9-dependent firefly luciferase reporter),[34] 75 ng 3FLAG-SOX9 plasmid,[35] and the indicated amounts of plasmids encoding no protein, 3FLAG-SOX6, and a 3FLAG-SOX6 variant for a total of 1,000 ng DNA. Cells were collected after 24 h in Tropix Lysis buffer (Applied Biosystems) supplemented with a protease inhibitor cocktail (Thermo Fisher Scientific), and extracts were subjected to the Nano-Glo Dual-Luciferase Reporter Assay (Thermo Fisher Scientific). Reporter activities were measured with a GloMax Explorer Multimode Microplate Reader (Promega). They are presented in the figures as means ± the standard deviation of firefly luciferase values measured for biological triplicates and normalized for transfection efficiency with NanoLuc values.

We tested SOX6 intracellular localization by transfecting HEK293 or COS-1 cells (CRL-1650; ATCC) with ViaFect (3.5 μL) and empty or SOX6 expression plasmid (1,000 ng per 10 cm² dish). Extracts were prepared the next day with NE-PER Nuclear and Cytoplasmic Extraction Reagents (Thermo Fisher Scientific). Immunoblots were carried out with FLAG M2-peroxidase-conjugated antibody (A8592, Sigma-Aldrich), P84 antibody (GTX70220-01, GeneTex), and β-actin antibody (sc-47778 [c4], Santa Cruz Biotechnology) as previously described.[10]

The abilities of SOX6 variants to homodimerize and to bind DNA were tested upon transfection of COS-1 cells with ViaFect (3.5 μL) and empty or SOX6 expression plasmid (1,000 ng per 10 cm² dish). Whole-cell extracts were prepared after 40 h in 14 mM HEPES buffer (pH 7.9) containing 1.5 mM MgCl₂, 6.0 mM KCl, 0.44 M NaCl, 0.08 mM EDTA, 2.3 mM DTT, 10% glycerol, and a protease inhibitor cocktail. Protein homodimerization was assessed by immunoblot after a 10 min incubation of cell extracts with varying amounts of glutaraldehyde, up to 0.01%. SOX6's binding to DNA was tested in an electrophoretic mobility shift assay (EMSA) with a DIG Gel Shift Kit, 2nd Generation (Sigma-Aldrich). The DNA probe contained a SOXD (SOX5 or SOX6) binding site that is located in the *Acan* enhancer used in the reporter assay described above (Agc1 K).[34] Mixtures contained 0.5 μL cell extract, 10 fmol DIG-labeled probe, and 25 ng poly [dG.dC] in a 20 mM HEPES buffer (pH 7.9) containing 40 mM KCl, 0.5 mM EDTA, 0.5 mM DTT, and 10% glycerol. After 30 min of incubation at 30°C, they were electrophoresed in 4% native polyacrylamide gels in TGE buffer (50 mM Tris, 0.4 M glycine, and 4.5 mM EDTA [pH 8.0]). We transferred DNA to a Zeta-Probe GT nylon membrane (Bio-Rad Laboratories) by electroblotting, and we detected signals according to the manufacturer's instructions by using a Chemi-Doc Imaging System (Bio-Rad Laboratories).

Evaluation of *SOX6* Expression in the Human Brain

We used RNA sequencing (RNA-seq) data available in the BrainSpan Atlas of the Developing Human Brain[36] to assess the expression of *SOX6* in various anatomical sites and at various developmental and adult ages.

Results

De Novo SOX6 Variants Associate with a Neurodevelopmental Syndrome

We identified 19 individuals sharing developmental milestone delays and/or intellectual disability and carrying various types of *SOX6* variants (Table 1). Ten individuals were males, and nine were females. At the most recent examination, their ages ranged from 12 months to 27 years. Pregnancies were uneventful, and postnatal growth parameters were generally in the normal range. Most individuals had intellectual disability (18/19) varying from mild (8/19) to mild-moderate (3/19), moderate (3/19), and severe (2/19). Abnormal behavior was observed in many affected individuals (14/19) and included attention-deficit/hyperactivity disorder (ADHD; 10/19) or autism spectrum disorder (4/19). Additional behavioral concerns included anxiety, aggressiveness, and automutilation. Dysmorphic features and skeletal anomalies were noted in many affected individuals. They included a mild facial dysmorphism, but this feature was nonspecific and thus not consistent with a clinically recognizable syndrome (Figure 1A). Of note, sagittal craniosynostosis with scaphocephaly was observed in two unrelated affected individuals, and oxycephaly with synostosis of the sagittal, metopic, and coronal sutures was observed in a third unrelated individual. Moreover, multiple diffuse osteochondromas were present in two brothers (7-year-old PIT-1 and 3-year-old PIT-2) and in one girl from an unrelated family (6-year-old CHOP-1) (Table S5). Variants in *EXT1* (MIM: 608177) and *EXT2* (MIM: 608210), which are deleted in a large percentage of individuals with exostoses (MIM: 133700 and MIM: 133701, respectively) and are not chromosomally linked to *SOX6*, were not detected in any of the individuals with osteochondromas. Interestingly, several of the osteochondromas that were initially detected in affected individual CHOP-1 at 4 years of age
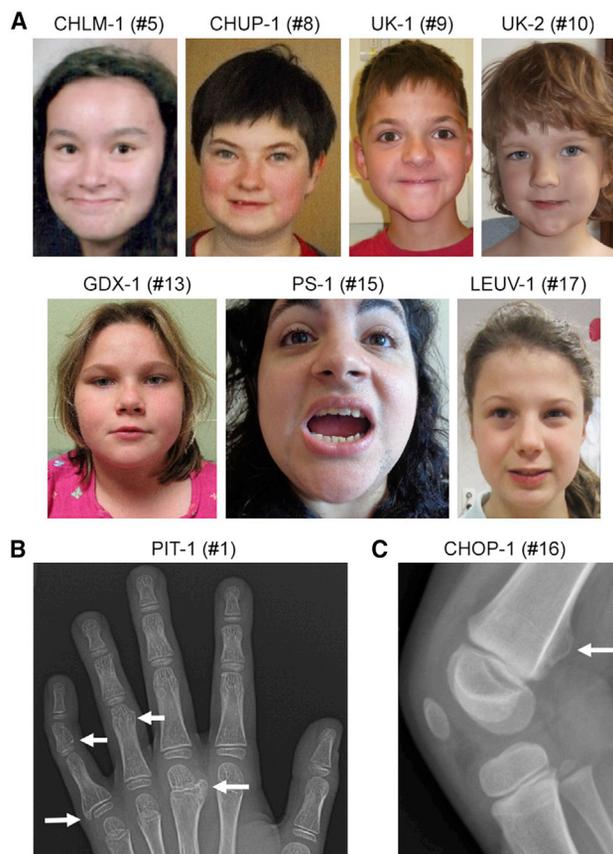
Figure 1. Clinical Findings in Subjects with SOX6 Variants
(A) Photos of seven subjects showing mild, nonspecific facial dysmorphism.
(B) X-ray showing multiple osteochondromas (marked by arrows) in the right hand of affected individual PIT-1.
(C) X-ray showing an osteochondroma (marked by arrow) at the right distal femur of affected individual CHOP-1.

were no longer detected on X-rays or were no longer palpable 2 years later (Table S5 and Figure 1B). The girl had no complaints of pain, and she ran and played without difficulty. Taken together, these data suggest that *SOX6* variants cause a unique form of human neurodevelopmental disorder often associated with mild dysmorphism and occasionally associated with craniosynostosis or osteochondromas.

SOX6 Translocation, CNVs, and SNVs Associate with Disease

The 19 affected individuals in our study displayed a spectrum of *SOX6* variants (Table 1). All individuals were heterozygous for the *SOX6* variant. Whereas 14 of them carried a *de novo* variant, four (including the PIT-1, PIT-2, and PIT-3 siblings) inherited the variant from their affected father, and one inherited it from an unaffected father who showed 22% mosaicism for the variant. The father of the PIT-1, PIT-2, and PIT-3 siblings presented with mild intellectual disability, but no further clinical evaluation was available, especially regarding the presence of osteochondromas.

Eight affected individuals (1–8), including the PIT-1–PIT-3 siblings, harbored a CNV, and one individual (19, CHUG-1) carried a balanced structural variant (Figure 2A). *SOX6* gives rise to four main transcripts that differ from one another in alternative promoter usage and alternative splicing of an exon encoding an internal, dispensable protein sequence. The most abundant transcript (GenBank: NM_033326.3) encodes a 5′ untranslated exon 1 and 15 coding exons. The reciprocal translocation detected in CHUG-1 occurred between the chromosomal regions 11p15.2 and 2p12 (Figure 2B). The 11p15.2 breakpoint was located in intron 13 of *SOX6*, thus separating exons 14–16, which encode the HMG domain and C terminus of the protein, from the more upstream exons.

The CNVs detected in all eight affected individuals (DECIPHER: 406929–406934) were different from one another (except for the those detected in the three siblings; Figure 2A and Table S6). All were partial deletions of *SOX6* and did not involve any other gene bodies. The PIT-1, PIT-2, and PIT-3 siblings' deletion involved exons 1 and 2 of all main *SOX6* transcripts. CHUN-1's deletion included exons 1–4 of all main *SOX6* transcripts. CHLM-1's deletion encompassed exons 2 and 3 of GenBank: NM_033326.3 and exons 1–3 of the other transcripts. IHG-1's and LJU-1's deletions comprised exons 2–12 and 2–13 of GenBank: NM_033326.3, respectively, and exons 1–11 and 1–12 of the other transcripts, respectively. CHUP-1's deletion removed exons 5–7, including most of the primary coiled-coil (CC1) domain sequence (discussed later in the manuscript). Splicing from exon 4 to exon 8 of GenBank: NM_033326.3 would not affect the coding sequence frame but would result in a SOX6 lacking a functional CC1 domain, and splicing of exons 4–9 would result in a frameshift with a stop codon after 27 residues. Thus, the balanced translocation and all deletions most likely inactivated the affected *SOX6* allele. Because no other abnormalities were detected in the microarray tests, the resulting diseases were postulated to reflect *SOX6* haploinsufficiency.

The other affected individuals in the study carried an SNV in *SOX6*. All SNVs were distinct and corresponded to four nonsense, two frameshift, and four missense mutations (Table 1). All, except one in an individual whose parents were unavailable, arose *de novo* and were absent in gnomAD control populations. The SOX6 encoded by GenBank: NM_033326.3 comprises 808 amino acids (Figure 3A). Its known functional domains are the HMG domain and two coiled-coil homodimerization domains, of which CC1 is longer and more critical than the secondary coiled-coil (CC2) domain.[12] The four nonsense variants caused SOX6 truncation upstream of CC1 (p.Ser81* [UK-1], p.Arg93* [UK-2], and p.Ser111* [CHUN-2]) or within CC1 (p.Gln240* [GDX-1]). In addition to encoding a short protein lacking all known functional domains, the mutant mRNAs were likely to undergo nonsense-mediated decay. The two frameshift variants (p.Pro293Lfs*3 [GDX-2] and p.Glu577Argfs*29 [PS-1]) caused SOX6 truncation

Figure 2. Structural Variants Detected in Affected Individuals 1–8 and 19

(A) Location of CNV variants. The main *SOX6* transcript isoforms are schematized; taller vertical lines correspond to coding sequences, smaller vertical lines correspond to 5′ and 3′ untranslated sequences, and arrowheads in introns point to the transcriptional direction. NCBI accession numbers are indicated. The GenBank: NM_033326.3 coding exons are labeled 1–16. Coordinates of chromosomal region 11p15 are shown above the schematics. Double-arrowed lines depict the microdeletions identified in subjects 1–8. CC1, primary coiled-coil domain; CC2, secondary coiled-coil domain; HMG, HMG domain.

(B) Location of the breakpoint of the 46,XY,t(2;11)(p11.2;p15.2)-balanced reciprocal translocation identified in affected individual 19. From top to bottom are schematics of chromosomes 2 and 11, in which a vertical dotted line indicates the breakpoint involving 11p15.2 and 2p11.2; reads (BAM file) aligned by the Integrative Genomics Viewer (IGV); the sequence of the breakpoint, in which red and blue lines show the normal sequences of chromosomes 2 and 11, respectively, and red sequences followed by blue sequences show the sequence junction; and the same representation of *SOX6* as in (A).

before CC2 and the HMG domain, respectively, and thus also most likely generated non-functional proteins. The missense variants affected different protein regions. p.Met605Thr (CHOP-1) and p.Trp639Arg (LEUV-1) occurred in the HMG domain, whereas p.Trp161Cys (GDX-3) and p.Ser746Leu (LEID-1) occurred in regions of unknown functions. p.Trp161Cys was located in the N terminus of SOX6, close to the CC1 domain, and p.Ser746Leu (LEID-1) was located in the C terminus of the protein. None of the missense variants were located at or near exon boundaries or in the alternatively spliced exon.

Thus, although we predicted the nonsense and frameshift SNVs to be pathogenic, we needed to undertake further analyses to determine whether the missense SNVs could be pathogenic too.

SOX6 Missense Variants Affect Residues Highly Conserved Evolutionarily and Mutated in Several SOXopathies

MetaDome analysis of mutation tolerance in the *SOX6* coding sequence in control human populations showed, not surprisingly, that the HMG domain is the most
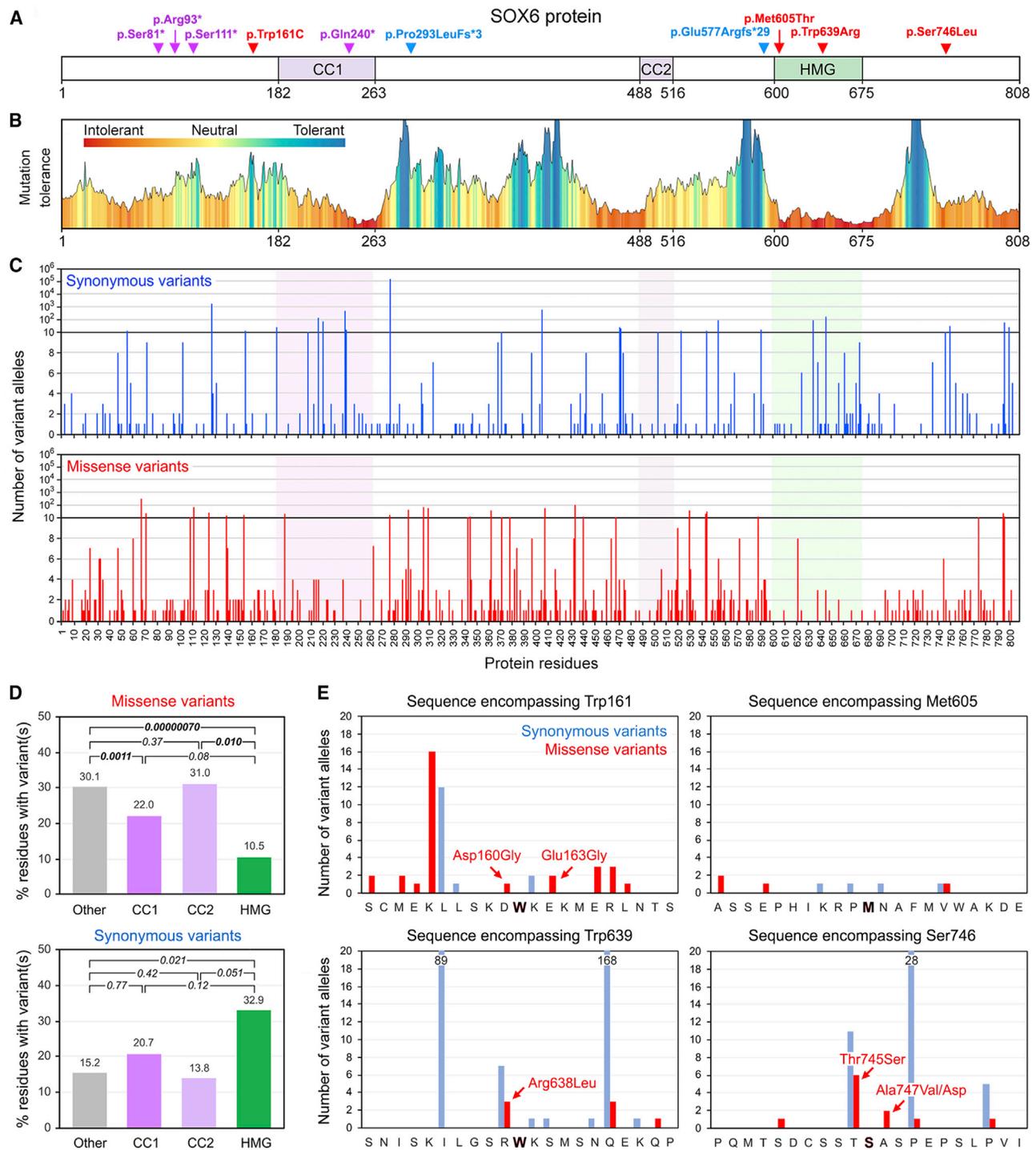
Figure 3. Analysis of SOX6 SNVs in Affected Individuals 9–18 and in gnomAD Individuals

(A) Location of study subjects' SNVs in the SOX6 isoform encoded by the GenBank: NM_033326.3 transcript. The protein and domain residue boundaries are indicated underneath the schematic. CC1, primary coiled-coil domain; CC2, secondary coiled-coil domain; HMG, HMG domain. Red represents missense variants, purple represents nonsense variants, and blue represents frameshift variants.

(B) Plot of the mutation tolerance of SOX6 residues downloaded from MetaDome.

(C) Counts and distribution of SOX6 synonymous and missense variants found in gnomAD individuals.

(D) Percentages of residues carrying at least one missense or synonymous variant in the functional and other domains of SOX6 in gnomAD individuals. We performed paired t tests to calculate the statistical significance of differences between domains. p values are indicated.

(E) Numbers of synonymous and missense variants detected in gnomAD individuals in the sequences encompassing the missense variants identified in four study subjects. The nature of the missense variants closest to the residues altered in the four affected individuals is indicated.

Figure 4. Evolutionary Conservation of SOX6 Residues Altered in Affected Individuals

(A) Alignment of human SOXD sequences encompassing Trp161. Asterisks represent fully conserved residues, and dots represent semi-conserved residues.

(B) Alignment of the HMG domain sequences of all human SOX proteins: residues altered in SOX6 in affected individuals in our study are shown in red, and residues altered in SOX5 in affected individuals with LAMSHF syndrome are shown in purple. Triangles represent residues mediating DNA binding (blue) and bending (green). Brackets represent H1, H2, and H3 α helices. Lines linked with dots represent key amino acids in nuclear localization signal sequences (NLSs) and nuclear export signal sequences (NESs).

(C) Alignment of human SOXD sequences encompassing Ser746 with indication of the position of SOX5 and SOX6 variants identified in affected individuals in our study.

intolerant region, but it also showed other intolerant regions, including CC1 (Figure 3B). On the basis of this finding, we analyzed the frequency and nature of *SOX6* variants in control individuals by using gnomAD, a database containing genomic and exonic sequences from more than 140,000 unrelated people of various genetic backgrounds.[35] Constraint metrics indicated that *SOX6* is under tight conservation constraint because 42.6 loss-of-function (nonsense) variants were expected, but only four were observed, resulting in a probability of loss-of-function intolerance of 1.00. Interestingly, whereas 169.7 synonymous variants were expected and 167 were observed (constraint Z score = 0.16), 454.7 missense variants were expected, but only 339 were observed (constraint Z score = 1.93). These findings thus suggested that *SOX6* missense variants might often be pathogenic.

Examination of the SNV distribution revealed that missense and synonymous variants occurred throughout the *SOX6* coding sequence (Figure 3C). However, although the three protein domains showed no statistically significant difference (t test, $p \leq 0.01$) in their percentages of residues carrying at least one synonymous variant, the HMG ($p = 7 \times 10^{-7}$) and CC1 ($p = 1.1 \times 10^{-3}$) domains had significantly fewer residues with missense variants than the CC2 domain and all other protein regions combined (Figure 3D). This result was consistent with the critical functions of the HMG and CC1 domains. Importantly, no missense variant was recorded in gnomAD individuals in the codons carrying a missense variant in our subjects (Figure 3E). Several gnomAD missense variants were located close to Trp161 and Ser746, and a few were close

to Met605 and Trp639. Thus, although *SOX6* missense variants are frequent in control individuals, those found in the affected individuals in our study have not been detected, leaving open the possibility that they are pathogenic.

Since residues important for protein function or regulation are generally conserved evolutionarily, we sought to determine whether Trp161, Met605, Trp639, Ser746, and their neighboring residues are conserved in SOX6 orthologs and other SOX proteins. The alignment of full-length SOX6 sequences from multiple vertebrate species revealed a high degree of evolutionary conservation both within and outside the known functional domains (Figure S1). This conservation included the residues altered in the affected individuals in our study and their neighboring residues. The alignment of SOX5, SOX6, and SOX13 sequences showed that Trp161 and its neighbors are also highly conserved among SOXD proteins (Figure 4A). The alignment of all SOX HMG domains showed full conservation at the positions corresponding to Met605 and Trp639 in SOX6 (Figure 4B). Furthermore, the Met605 neighbors are fully conserved too, and the Trp639 neighbors are conserved in SOXD and several other SOX proteins. Finally, Ser746 is conserved between SOX5 and SOX6, even though its neighbors are only weakly conserved (Figure 4C).

We next asked whether variants in other SOX genes at residues matching those found in *SOX6* in our subjects have been reported to cause disease. Although no SOX5 variant of the residue matching Trp161 or its neighbors has been associated with LAMSHF disease yet, a SOX5

Table 2. Pathological Missense Variants in Various SOX Proteins Match the Two SOX6 HMG Domain Variants Identified in Affected Individuals

Protein	Variant	Phenotype	Reference
SOX Variants Matching the SOX6 p.Met605Thr Variant			
SOX6	p.Met605Thr	ID with mild skeletal defects	this study
SRY	p.Met64Ile	XY sex reversal	Berta et al.[37]
SRY	p.Met64Arg	gonadal dysgenesis and XY sex reversal	Scherer et al.[38]
SOX5	p.Met560Val	Lamb-Shaffer syndrome	Zawerton et al.[7]
SOX10	p.Met108Thr	Kallmann syndrome	Pingault et al.[49]
SOX Variants Matching the SOX6 p.Trp639Arg Variant			
SOX6	p.Trp639Arg	ID with mild skeletal defects	this study
SRY	p.Trp98Arg	XY sex reversal with primary amenorrhea and low testosterone	Philibert et al.[50]
SRY	p.Trp98Cys	gonadal dysgenesis and XY sex reversal	Bastian et al.[51]
SOX2	p.Trp79Ser	bilateral anophthalmia with GH deficiency	Chassaing et al.[52]
SOX9	p.Trp143Arg	campomelic dysplasia with XY sex reversal	Meyer et al.[53]
SOX10	p.Trp142Arg	Kallman syndrome	Pingault et al.[49]
SOX17	p.Trp106Leu	idiopathic pulmonary arterial hypertension	Zhu et al.[54]

The following abbreviation is used: ID, intellectual disability.

p.Met560Val variant matching the SOX6 p.Met605Thr variant and a SOX5 p.Asn561His variant have caused LAMSHF syndrome[7] (Figure 4B and Table 2). In the SRY residue, two variants matching SOX6 p.Met605Thr (p.Met64Ile and p.Met64Arg) have caused gonadal dysgenesis, and a SOX10 p.Mer108Thr variant, identical to the SOX6 p.Met605Thr variant, has caused Kallmann disease (MIM: 147950).[37,38] Variants in the residue equivalent to SOX6 Trp639 were found to cause disease in SRY, SOX2, SOX9, SOX10, and SOX17; several of these variants are identical to the SOX6 variant (Trp → Arg substitution). Finally, an affected individual with a SOX5 p.Ser693Leu variant, located close to the SOX6 Ser746 residue, has LAMSHF syndrome[7] (Figure 4C). Together, sequence conservation and SOXopathy data strongly support the notion that the four missense variants found in the affected individuals in our study might be pathogenic.

Affected Individuals' Missense Variants Might Affect the Local Structure of SOX6

We used several software programs to get insights into the possible structural impact of the SOX6 missense variants detected in the affected individuals in our study (Figure S2). HOPE indicated that the p.Trp161Cys variant could alter SOX6 interactions with other molecules because cysteine is smaller than tryptophan. SWISS-MODEL, a template-based structure-prediction software, and PEP-FOLD3, a de novo software predicting protein

structures directly from amino acid sequences, predicted that residues 153–170, which are centered around Trp161, could form an α helix juxtaposed to CC1, and PEP-FOLD3 proposed that the replacement of tryptophan by cysteine would disrupt this structure. Regarding p.Met605Thr, HOPE indicated that the substitution of the sulfur-containing nonpolar methionine by a hydroxyl-containing, polar, and smaller threonine could affect the ability of the HMG domain to bind and bend DNA. For p.Trp639R, PEP-FOLD3 did not predict a change in the helical structure of the HMG domain, but HOPE indicated that the substitution of the bulky aromatic tryptophan with a positively charged arginine would most likely affect the interaction of the HMG domain with DNA. In regard to p.Ser746Leu, SWISS-MODEL could not predict a specific structure for the region encompassing Ser746, but PEP-FOLD3 predicted that this region forms a small α helix regardless of whether the 746 position is occupied by a serine or a leucine. Thus, a change in protein structure is unlikely for this variant. However, although both serine and leucine have a short side chain, this chain is polar and contains a hydroxyl group in serine, whereas it is aliphatic and nonpolar in leucine. It is thus possible that the p.Ser746Leu variant affects the ability of SOX6 to interact with other proteins. In conclusion, each SOX6 variant could destabilize the local structure or biochemical properties of SOX6 and thereby impair protein function or stability.

Affected Individuals' Missense Variants Affect the Nuclear Localization and Transcriptional Activity of SOX6 In Vitro

To test whether the missense variants detected in our subjects could affect SOX6 function, we transiently transfected expression plasmids for the WT and variant SOX6 proteins in COS-1 and HEK293 cells. Immunoblots of cytoplasmic and nuclear extracts showed that all variants were efficiently expressed and that the p.Met605Thr and p.Trp639Arg proteins were not translocated or retained into the nucleus as efficiently as WT SOX6 and the other two variants (Figure 5A). Because Trp161 could interact with CC1, we treated cell extracts with glutaraldehyde, an assay previously used to detect SOX6 homodimers[12] and observed that the p.Trp161Cys variant protein could homodimerize as efficiently as WT SOX6 and the p.Met605Thr variant (Figure 5B). Next, we tested cell extracts in EMSA with an Acan enhancer sequence previously shown to bind SOX5 and SOX6.[34] The proteins with variants outside the HMG domain behaved like WT SOX6, whereas the proteins harboring the p.Met605Thr and p.Trp639Arg variants failed to bind to the DNA probe (Figure 5C).

Finally, we tested the variants for transcriptional activity by co-transfecting cells with a reporter plasmid containing an Acan enhancer synergistically activated by SOX9 and SOX proteins (SOX5 or SOX6)[34] and with SOX6 and SOX9 expression plasmids. As expected, SOX6 alone,
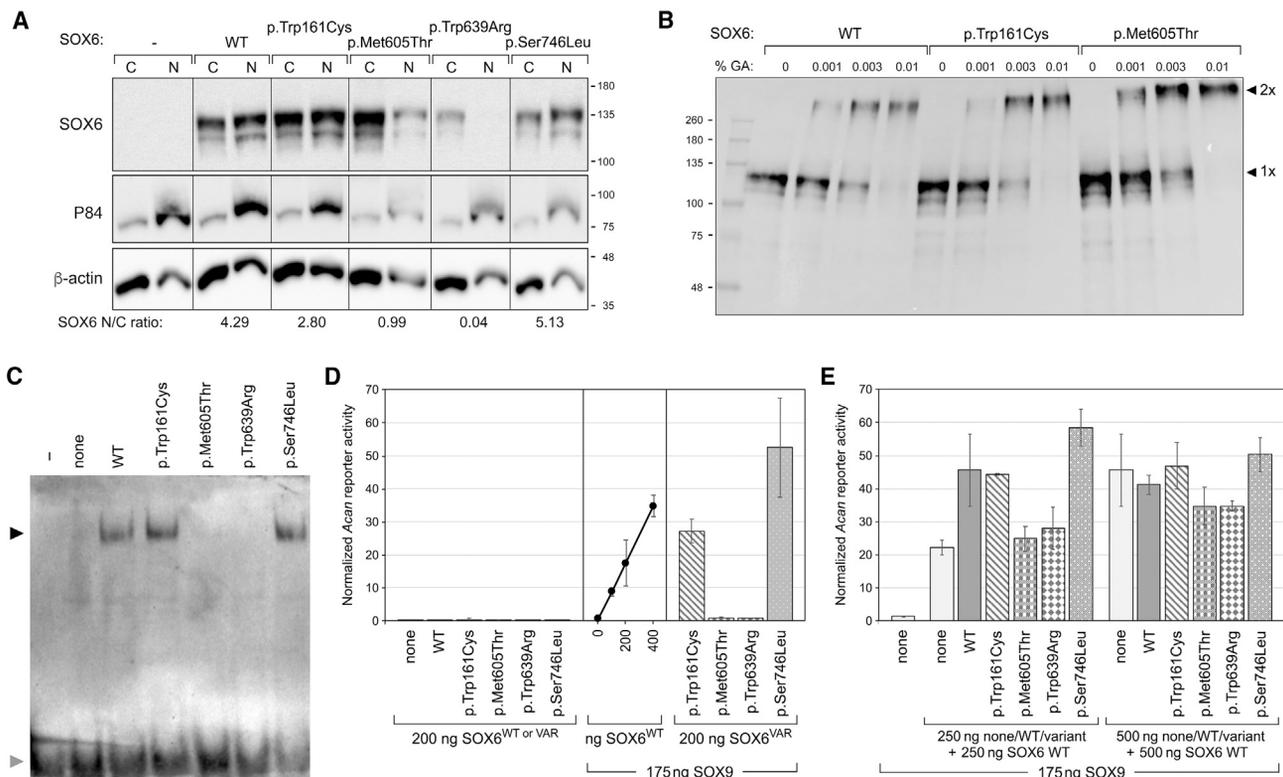
Figure 5. Functional Tests of SOX6 Missense Variants

(A) Stability and intracellular distribution of SOX6 variants. COS-1 cells were transfected with plasmids encoding 3FLAG-tagged SOX6 WT and variant proteins, as indicated. Cytoplasmic (C) and nuclear (N) extracts were tested via immunoblotting with a FLAG antibody. As expected, the P84 protein was enriched in nuclear extracts, and β-actin was enriched in cytoplasmic extracts. The distribution of SOX6 relative to β-actin in the nuclear versus cytoplasmic compartment is indicated underneath the blots. The migration of protein markers (Mr in k values) is indicated.

(B) Ability of SOX6 variants to homodimerize. Lysates of COS-1 cells, transfected as described in (A), were incubated without (0) or with 0.001%, 0.03%, or 0.01% glutaraldehyde. SOX6 was visualized via immunoblotting with a FLAG antibody. The migration of protein standards (Mr in k values) and SOX6 monomers (1×) and homodimers (2×) is indicated.

(C) Ability of SOX6 variants to bind DNA. An *Acan* enhancer probe was used in EMSA with no protein extract (−) or with extracts from COS-1 cells transfected with expression plasmids for no protein (none), SOX6 WT, or variant proteins, as indicated. SOX6-probe complexes (indicated with a black arrowhead) and free probes (indicated with a gray arrowhead) were resolved by electrophoresis.

(D) Ability of SOX6 variants to synergize with SOX9 in transactivation. HEK293 cells were transfected with an *Acan* reporter, a control reporter, and plasmids encoding no SOX (none), SOX9, SOX6 WT, or SOX6 variant proteins, as indicated. The amounts of SOX plasmids are indicated too. *Acan* reporter activities were normalized for transfection efficiency. They are presented as the mean ± standard deviation obtained for triplicates in an experiment representative of five independent ones. The arrowhead in the middle panel points to the amount of WT SOX6 plasmid (200 ng) that was also used for variant plasmids in the right panel.

(E) Interference of SOX6 variant proteins with WT SOX6 in transactivation. HEK293 cells were transfected as described in (D) with the indicated types and amounts of plasmids encoding no protein (none) or SOX proteins. *Acan* reporter activities were normalized for transfection efficiency. They are presented as the mean ± standard deviation obtained for triplicates in an experiment representative of four independent ones.

whether WT or mutant, and SOX9 alone hardly activated the reporter, whereas co-expression of WT SOX6 and SOX9 resulted in a SOX6 dose-dependent activation of the reporter (Figure 5D). In line with the EMSA results, the SOX6 variants located within the HMG domain were transcriptionally inactive. In contrast, the other variants were as potent as, or even more potent than, WT SOX6. Because all affected individuals in our study are heterozygous for their SOX6 variant, we next asked whether the variant proteins could have a dominant-negative influence on WT SOX6. When a dose of WT SOX6 within the linear dose-response range (250 ng plasmid) was doubled, the reported activity doubled as well. When this dose was sup-

plemented with the same dose of p.Trp161Cys or p.Ser746-Leu variant, a similar or slightly higher reporter activity was measured, whereas the same dose of p.Met605Thr or p.Trp639Arg variant failed to increase the reporter activity. These findings were consistent with the activities of the variant proteins when tested alone (Figure 5D) and thus suggested that none of the variant proteins negatively interfered with WT SOX6. When the SOX6 proteins were tested under competitive conditions, that is, at doses in the plateau of maximum reporter activity (500 ng plasmid), none of the variant proteins prevented WT SOX6 from reaching this plateau, again suggesting a lack of a dominant-negative effect of the variants.
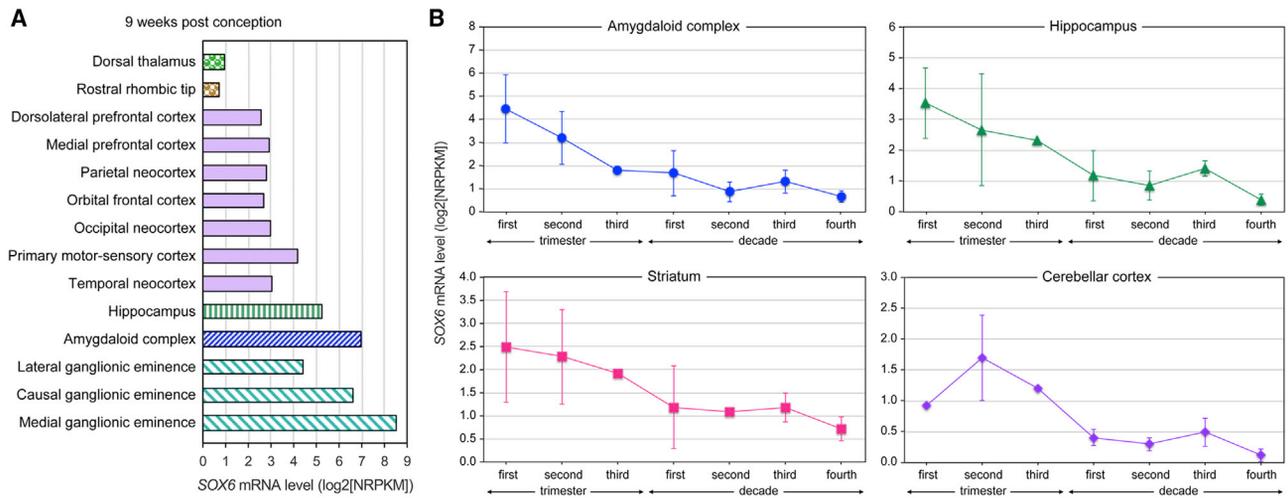
Figure 6. Analysis of *SOX6* Expression in the Human Brain
(A) *SOX6* RNA expression measured by RNA-seq in various regions of the developing brain of a representative human fetus 9 weeks after conception.
(B) *SOX6* RNA expression measured by RNA-seq in the amygdaloid complex, hippocampus, striatum, and cerebellar cortex of multiple individuals whose ages are within the range spanning the three trimesters of gestation *in utero* and the first four decades of life. Data are presented as averages with standard deviation for each age category (n = 1–10 per group). The four brain structures were selected because their *SOX6* RNA expression is higher than that of other structures.

Taken together, these functional assays indicated that the two missense variants in the HMG domain abolished the ability of SOX6 to function as a transcription factor. The two missense variants located outside this domain exhibited SOX6 WT activities in our assays, but we cannot rule out the possibility that they could impair critical activities of SOX6 *in vivo* and might thus be genuinely pathogenic.

SOX6 Is Highly Expressed in the Developing Human Brain

Whole-transcriptome profiling data deposited in the Brain-Span Atlas of the Developing Human Brain provided information on where and when *SOX6* is expressed in the brain of human fetuses, children, and adults. At 9 weeks after conception, fetuses were found to express *SOX6* in many prospective brain structures (Figure 6A). *SOX6* RNA expression was highest in the ganglionic eminence, the amygdaloid complex, and the hippocampus, all of which have central roles in brain development. Expression lower than that measured in the aforementioned regions were measured in sub-regions of the cortex and neocortex. The expression of *SOX6* declined in all brain structures in the final stages of gestation and in the neonatal period, such that it was already as low in infants as in adults (Figure 6B). These data thus provide support to the notion that SOX6 has important roles in several regions of the developing human brain.

Discussion

We reported here on 19 individuals, from 17 families, who exhibited genetic variants predicted in most cases to cause SOX6 haploinsufficiency. All subjects shared core clinical features of a neurodevelopmental syndrome that included developmental delay and intellectual disability. Inconstant features, such as ADHD (11 affected individuals), mild facial dysmorphism (9 affected individuals), craniosynostosis (3 affected individuals), and osteochondromas (3 affected individuals), were also shared by several subjects. This disease thus expands the currently known list of SOXopathies, i.e., developmental disorders due to pathogenic variants in SOX genes.

The spectrum of *SOX6* variants carried by the affected individuals in our study ranged from a balanced chromosomal translocation to partial or complete gene deletions and to nonsense, missense, and other SNVs. Although no individual affected by a *SOX6* SNV was reported prior to our study, two individuals were described as having a *SOX6* CNV. The first one was a 4-year-old girl with global developmental delay, spinal cord syrinx, and recurrent episodes of parkinsonian symptoms, including gait instability, tremors, and dysarthria.[39] Because molecular karyotyping showed a *de novo* 2.36 Mb deletion including *SOX6* and several other genes, no conclusion could be drawn regarding the implication of the *SOX6* deletion in the clinical phenotype. The second individual was a 15-year-old boy with significantly delayed speech development and ADHD.[40] This boy also had generalized dystonic and frequent athetoid movements of the arms, trunk, and neck. His gait was severely impaired secondarily to frequent dystonic postures. A whole-genome SNP array revealed a *de novo* 84 kb deletion encompassing *SOX6* exons 14–16, which encode the HMG domain and C terminus of SOX6. In combination with the affected individuals reported in our study, these additional subjects link *SOX6* variants to a complex neurodevelopmental syndrome.

The linking of *SOX6* variants with a neurodevelopmental syndrome fits with the evidence that *SOX6* is dynamically expressed in the human developing brain and that its mouse ortholog is also expressed in this structure and plays key roles in creating neuronal diversity.[19,20,41] *Sox6* is specifically expressed in neuronal progenitors in the developing dorsal telencephalon and induces differentiation of the cells.[19] This function might explain the neurodevelopmental disorders observed in the affected individuals in our study. In contrast to the individual with an 84 kb deletion, none of the affected individuals in our study had movement disorders. SOX6 is also strongly expressed in the medial ganglionic eminence, where it is necessary for the normal positioning and maturation of cortical interneurons. As a consequence, specific removal of *Sox6* from these cells results in a severe epileptic encephalopathy in mice.[42] None of the affected individuals reported here or previously, however, presented with epilepsy. A possible explanation is that the human individuals were heterozygous for the *SOX6* alteration, whereas the phenotype was seen in mice only when *Sox6* was inactivated homozygously.

Besides neurogenesis, chondrogenesis is another developmental process critically dependent upon *Sox6*. The gene is co-expressed with *Sox5* in chondrocytes, and the two genes act largely in redundancy to promote chondrocyte differentiation.[12–14] SOX5 and SOX6 form homodimers and heterodimers through their coiled-coil domains and cooperatively bind with SOX9 on enhancers driving hundreds of cartilage-specific genes. They thereby potentiate the ability of SOX9, which is a master chondrogenic factor, to transactivate these genes.[43] *Sox5*-null and *Sox6*-null mice are born with discrete skeletal malformations, and fetuses null for both *Sox5* and *Sox6* die *in utero* with severely underdeveloped cartilage structures and cartilage-derived bones. These chondrogenic roles of SOX5 and SOX6 could explain the mild facial dysmorphism and the shortness of fingers and extremities noted in some of the affected individuals in our study, as well as in some individuals with LAMSHF syndrome, who are heterozygous for *SOX5* pathogenic variants.[7] Dysmorphic features have not been described in *Sox5*- and *Sox6*-heterozygous-null mice but cannot be excluded because detailed morphometric measurements have not been performed.

Three unrelated affected individuals reported here presented with craniosynostosis. For two of them (UK-1 and UK-2), the sequences of genes known to be involved in craniosynostosis were analyzed in depth, but no variant likely to be pathogenic was identified. Of note, a 5-year-old boy with mild developmental speech delay and craniosynostosis was previously described.[44] This boy exhibited brachycephaly, proptosis, midfacial hypoplasia, and low-set ears. He carried a *de novo* balanced translocation, 46,XY,t(9;11)(q33;p15), whose breakpoint on chromosome 11 disrupted *SOX6*. *Sox5* and *Sox6* single- and double-mutant mice were not described to have defects in the intramembranous bones that form the skull vault, but such defects and the expression of *Sox5* and *Sox6* in skull progenitor cells and osteoblasts might have been overlooked in view of the strong expression of both genes in chondrocytes and the severity of cartilage and endochondral bone underdevelopment in mutant mice.[13]

Among the affected individuals in our study, two brothers and one unrelated girl presented with diffuse osteochondromas. The former each had a *SOX6* deletion, and the latter had a missense variant in the HMG domain. Thus, these individuals did not have SOX6 variant types that differed from those of other affected individuals in our study. Osteochondromas are benign tumors developing as ectopic bone and cartilage, most often in the perichondrium adjacent to cartilage growth plates. Because SOX6 promotes essential steps of the chondrocyte differentiation pathway, the finding that *SOX6* haploinsufficiency is associated with such tumors is intriguing; this calls for researchers to undertake future investigations to determine the importance of SOX6 in preventing ectopic cartilage and bone formation. At present, a possible explanation might come from evidence that SOX6 is a tumor suppressor in numerous types of cancer, including osteosarcoma, the most common type of malignant bone tumor (MIM: 259500).[25,27,45–47]

Most of the affected individuals in our study had *SOX6* variants that surely inactivated one *SOX6* allele and thus caused a disease that revealed *SOX6* haploinsufficiency. These variants were microdeletions and frameshift and nonsense variants. That missense variants in the HMG domain inactivated the protein made from the *SOX6* carrier allele was supported both by our *in vitro* functional assays and by evidence that many variants in the HMG domain, including those detected in affected individuals in our study, have been shown in other SOX genes to cause SOXopathies. One should, however, not deduce that any SOX variant in the HMG domain is pathogenic. gnomAD indeed contains multiple control individuals with variants in the HMG domain, and we showed in the case of SOX4 that these variants were retaining functions *in vitro*.[10] The identification of additional individuals with variants causing disease along with expansion of gnomAD should help future researchers establish algorithms to accurately make pathogenicity diagnoses for newly affected individuals.

Two of the affected individuals in our study carried a *SOX6* missense variant outside the HMG domain. Although these two individuals appeared to have more severe intellectual disability than all others and *in silico* tools based on amino acid structure and protein sequence predicted pathogenicity, *in vitro* functional assays did not consolidate this argument. A possible explanation is that these assays, which assessed the ability of SOX6 to synergize with SOX9 to activate a chondrocyte-specific enhancer, were not suitable for assessing the impact of the variants on SOX6 function or regulation in neurogenesis and other processes *in vivo*. A few individuals with LAMSHF disease whose variants were located outside the HMG domain (of SOX5),

were not present in gnomAD individuals, and were fully active in the functional assay *in vitro* used in the present study were also reported.[7] The pathogenicity of such variants is further supported by the extremely high degree of conservation observed between SOX5 and SOX6 throughout the proteins and among vertebrate orthologs. More studies are thus needed for determining how such variants might critically affect SOX5 and SOX6 function or regulation and thereby cause diseases. Of note, one of the two patients also carried a hemizygous *MECP2* (MIM: 300005) variant that was most likely pathogenic. Variants in this gene have been associated with X-linked autism susceptibility (MIM: 300496), mental retardation (MIM: 300260), and Rett syndrome (MIM: 312750).[48] It is highly likely that this variant has a major contribution to the clinical features of this patient.

In conclusion, the findings from this study concur that *SOX6* haploinsufficiency leads to a specific form of neurodevelopmental SOXopathy characterized by mild to severe intellectual disability and inconstantly associated with skeletal anomalies, such as mild facial dysmorphism, craniosynostosis, and osteochondromas.

Supplemental Data

Supplemental Data can be found online at https://doi.org/10.1016/j.ajhg.2020.04.015.

Web Resources

1000 Genomes Project, http://www.1000genomes.org/
ClinVar, https://www.ncbi.nlm.nih.gov/clinvar/
DECIPHER, https://decipher.sanger.ac.uk/
gnomAD Browser, http://gnomad.broadinstitute.org/
GenBank, https://www.ncbi.nlm.nih.gov/genbank/
Genomics England 100,000 Genomes Project, https://www.genomicsengland.co.uk/
HOPE, http://www.cmbi.ru.nl/hope/
NCBI, https://www.ncbi.nlm.nih.gov/
OMIM, http://www.omim.org/
PEP-FOLD-3, http://bioserv.rpbs.univ-paris-diderot.fr/services/PEP-FOLD3/
SWISS-MODEL, https://swissmodel.expasy.org/

References

1. Kamachi, Y., and Kondoh, H. (2013). Sox proteins: regulators of cell fate specification and differentiation. Development *140*, 4129–4144.

2. Angelozzi, M., and Lefebvre, V. (2019). SOXopathies: growing family of developmental disorders due to SOX mutations. Trends Genet. *35*, 658–671.

3. Gubbay, J., Collignon, J., Koopman, P., Capel, B., Economou, A., Münsterberg, A., Vivian, N., Goodfellow, P., and Lovell-Badge, R. (1990). A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. Nature *346*, 245–250.

4. Sinclair, A.H., Berta, P., Palmer, M.S., Hawkins, J.R., Griffiths, B.L., Smith, M.J., Foster, J.W., Frischauf, A.M., Lovell-Badge, R., and Goodfellow, P.N. (1990). A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature *346*, 240–244.

5. Kashimada, K., and Koopman, P. (2010). Sry: the master switch in mammalian sex determination. Development *137*, 3921–3930.

6. Lamb, A.N., Rosenfeld, J.A., Neill, N.J., Talkowski, M.E., Blumenthal, I., Girirajan, S., Keelean-Fuller, D., Fan, Z., Pouncey, J., Stevens, C., et al. (2012). Haploinsufficiency of SOX5 at 12p12.1 is associated with developmental delays with prominent language delay, behavior problems, and mild dysmorphic features. Hum. Mutat. *33*, 728–740.

7. Zawerton, A., Mignot, C., Sigafoos, A., Blackburn, P.R., Haseeb, A., McWalter, K., Ichikawa, S., Nava, C., Keren, B., Charles, P., et al.; Deciphering Developmental Disorder Study (2020). Widening of the genetic and clinical spectrum of Lamb-Shaffer syndrome, a neurodevelopmental disorder due to SOX5 haploinsufficiency. Genet. Med. *22*, 524–537.

8. Tsurusaki, Y., Koshimizu, E., Ohashi, H., Phadke, S., Kou, I., Shiina, M., Suzuki, T., Okamoto, N., Imamura, S., Yamashita, M., et al. (2014). De novo SOX11 mutations cause Coffin-Siris syndrome. Nat. Commun. *5*, 4011.

9. Hempel, A., Pagnamenta, A.T., Blyth, M., Mansour, S., McConnell, V., Kou, I., Ikegawa, S., Tsurusaki, Y., Matsumoto, N., Lo-Castro, A., et al.; DDD Collaboration (2016). Deletions and de novo mutations of SOX11 are associated with a neurodevelopmental disorder with features of Coffin-Siris syndrome. J. Med. Genet. *53*, 152–162.

10. Zawerton, A., Yao, B., Yeager, J.P., Pippucci, T., Haseeb, A., Smith, J.D., Wischmann, L., Kühl, S.J., Dean, J.C.S., Pilz, D.T., et al.; Deciphering Developmental Disorders Study; and University of Washington Center for Mendelian Genomics (2019). De novo SOX4 variants cause a neurodevelopmental disease associated with mild dysmorphism. Am. J. Hum. Genet. *104*, 777.

11. Cohen-Barak, O., Hagiwara, N., Arlt, M.F., Horton, J.P., and Brilliant, M.H. (2001). Cloning, characterization and chromosome mapping of the human SOX6 gene. Gene *265*, 157–164.

12. Lefebvre, V., Li, P., and de Crombrugghe, B. (1998). A new long form of Sox5 (L-Sox5), Sox6 and Sox9 are coexpressed in chondrogenesis and cooperatively activate the type II collagen gene. EMBO J. *17*, 5718–5733.

13. Smits, P., Li, P., Mandel, J., Zhang, Z., Deng, J.M., Behringer, R.R., de Crombrugghe, B., and Lefebvre, V. (2001). The transcription factors L-Sox5 and Sox6 are essential for cartilage formation. Dev. Cell *1*, 277–290.

14. Lefebvre, V. (2019). Roles and regulation of SOX transcription factors in skeletogenesis. Curr. Top. Dev. Biol. *133*, 171–193.

15. Quiat, D., Voelker, K.A., Pei, J., Grishin, N.V., Grange, R.W., Bassel-Duby, R., and Olson, E.N. (2011). Concerted regulation of myofiber-specific gene expression and muscle performance by the transcriptional repressor Sox6. Proc. Natl. Acad. Sci. USA *108*, 10196–10201.

16. Sluijter, J.P., van Mil, A., van Vliet, P., Metz, C.H., Liu, J., Doevendans, P.A., and Goumans, M.J. (2010). MicroRNA-1 and -499 regulate differentiation and proliferation in human-derived cardiomyocyte progenitor cells. Arterioscler. Thromb. Vasc. Biol. *30*, 859–868.

17. Dumitriu, B., Patrick, M.R., Petschek, J.P., Cherukuri, S., Klingmuller, U., Fox, P.L., and Lefebvre, V. (2006). Sox6 cell-autonomously stimulates erythroid cell survival, proliferation, and terminal maturation and is thereby an important enhancer of definitive erythropoiesis during mouse development. Blood *108*, 1198–1207.

18. Iguchi, H., Ikeda, Y., Okamura, M., Tanaka, T., Urashima, Y., Ohguchi, H., Takayasu, S., Kojima, N., Iwasaki, S., Ohashi, R., et al. (2005). SOX6 attenuates glucose-stimulated insulin secretion by repressing PDX1 transcriptional activity and is down-regulated in hyperinsulinemic obese mice. J. Biol. Chem. *280*, 37669–37680.

19. Azim, E., Jabaudon, D., Fame, R.M., and Macklis, J.D. (2009). SOX6 controls dorsal progenitor identity and interneuron diversity during neocortical development. Nat. Neurosci. *12*, 1238–1247.

20. Batista-Brito, R., Rossignol, E., Hjerling-Leffler, J., Denaxa, M., Wegner, M., Lefebvre, V., Pachnis, V., and Fishell, G. (2009). The cell-intrinsic requirement of Sox6 for cortical interneuron development. Neuron *63*, 466–481.

21. Panman, L., Papathanou, M., Laguna, A., Oosterveen, T., Volakakis, N., Acampora, D., Kurtsdotter, I., Yoshitake, T., Kehr, J., Joodmardi, E., et al. (2014). Sox6 and Otx2 control the specification of substantia nigra and ventral tegmental area dopamine neurons. Cell Rep. *8*, 1018–1025.

22. Stolt, C.C., Schlierf, A., Lommes, P., Hillgärtner, S., Werner, T., Kosian, T., Sock, E., Kessaris, N., Richardson, W.D., Lefebvre, V., and Wegner, M. (2006). SoxD proteins influence multiple stages of oligodendrocyte development and modulate SoxE protein function. Dev. Cell *11*, 697–709.

23. Yang, T.L., Guo, Y., Liu, Y.J., Shen, H., Liu, Y.Z., Lei, S.F., Li, J., Tian, Q., and Deng, H.W. (2012). Genetic variants in the SOX6 gene are associated with bone mineral density in both Caucasian and Chinese populations. Osteoporos. Int. *23*, 781–787.

24. Dong, C., Beecham, A., Slifer, S., Wang, L., Blanton, S.H., Wright, C.B., Rundek, T., and Sacco, R.L. (2010). Genomewide linkage and peakwide association analyses of carotid plaque in Caribbean Hispanics. Stroke *41*, 2750–2756.

25. Ganesh, S.K., Tragante, V., Guo, W., Guo, Y., Lanktree, M.B., Smith, E.N., Johnson, T., Castillo, B.A., Barnard, J., Baumert, J., et al.; CARDIOGRAM, METASTROKE; and LifeLines Cohort Study (2013). Loci influencing blood pressure identified using a cardiovascular gene-centric array. Hum. Mol. Genet. *22*, 1663–1678.

26. Liu, Y.Z., Pei, Y.F., Liu, J.F., Yang, F., Guo, Y., Zhang, L., Liu, X.G., Yan, H., Wang, L., Zhang, Y.P., et al. (2009). Powerful bivariate genome-wide association analyses suggest the SOX6 gene influencing both obesity and osteoporosis phenotypes in males. PLoS ONE *4*, e6827.

27. Qin, Y.R., Tang, H., Xie, F., Liu, H., Zhu, Y., Ai, J., Chen, L., Li, Y., Kwong, D.L., Fu, L., and Guan, X.Y. (2011). Characterization of tumor-suppressive function of SOX6 in human esophageal squamous cell carcinoma. Clin. Cancer Res. *17*, 46–55.

28. Lynn, M., Wang, Y., Slater, J., Shah, N., Conroy, J., Ennis, S., Morris, T., Betts, D.R., Fletcher, J.A., and O'Sullivan, M.J. (2013). High-resolution genome-wide copy-number analyses identify localized copy-number alterations in Ewing sarcoma. Diagn. Mol. Pathol. *22*, 76–84.

29. Schlierf, B., Friedrich, R.P., Roerig, P., Felsberg, J., Reifenberger, G., and Wegner, M. (2007). Expression of SoxE and SoxD genes in human gliomas. Neuropathol. Appl. Neurobiol. *33*, 621–630.

30. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al.; Exome Aggregation Consortium (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285–291.

31. Venselaar, H., Te Beek, T.A., Kuipers, R.K., Hekkelman, M.L., and Vriend, G. (2010). Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. BMC Bioinformatics *11*, 548.

32. Bienert, S., Waterhouse, A., de Beer, T.A., Tauriello, G., Studer, G., Bordoli, L., and Schwede, T. (2017). The SWISS-MODEL Repository-new features and functionality. Nucleic Acids Res. *45* (D1), D313–D319.

33. Lamiable, A., Thévenet, P., Rey, J., Vavrusa, M., Derreumaux, P., and Tufféry, P. (2016). PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex. Nucleic Acids Res. *44* (W1), W449-54.

34. Han, Y., and Lefebvre, V. (2008). L-Sox5 and Sox6 drive expression of the aggrecan gene in cartilage by securing binding of Sox9 to a far-upstream enhancer. Mol. Cell. Biol. *28*, 4999–5013.

35. Haseeb, A., and Lefebvre, V. (2019). The SOXE transcription factors-SOX8, SOX9 and SOX10-share a bi-partite transactivation mechanism. Nucleic Acids Res. *47*, 6917–6931.

36. Miller, J.A., Ding, S.L., Sunkin, S.M., Smith, K.A., Ng, L., Szafer, A., Ebbert, A., Riley, Z.L., Royall, J.J., Aiona, K., et al. (2014). Transcriptional landscape of the prenatal human brain. Nature *508*, 199–206.

37. Berta, P., Hawkins, J.R., Sinclair, A.H., Taylor, A., Griffiths, B.L., Goodfellow, P.N., and Fellous, M. (1990). Genetic evidence equating SRY and the testis-determining factor. Nature *348*, 448–450.

38. Scherer, G., Held, M., Erdel, M., Meschede, D., Horst, J., Lesniewicz, R., and Midro, A.T. (1998). Three novel SRY mutations in XY gonadal dysgenesis and the enigma of XY gonadal dysgenesis cases without SRY mutations. Cytogenet. Cell Genet. *80*, 188–192.

39. Scott, O., Pugh, J., Kiddoo, D., Sonnenberg, L.K., Bamforth, S., and Goez, H.R. (2014). Global developmental delay, progressive relapsing-remitting parkinsonism, and spinal syrinx in a child with SOX6 mutation. J. Child Neurol. *29*, NP164–NP167.

40. Ebrahimi-Fakhari, D., Maas, B., Haneke, C., Niehues, T., Hinderhofer, K., Assmann, B.E., and Runz, H. (2015). Disruption of SOX6 is associated with a rapid-onset dopa-responsive movement disorder, delayed development, and dysmorphic features. Pediatr. Neurol. *52*, 115–118.

41. Ji, E.H., and Kim, J. (2016). SoxD transcription factors: multifaceted players of neural development. Int. J. Stem Cells *9*, 3–8.

42. Rudrabhatla, P., Utreras, E., Jaffe, H., and Kulkarni, A.B. (2014). Regulation of Sox6 by cyclin dependent kinase 5 in brain. PLoS ONE *9*, e89310.

43. Liu, C.F., and Lefebvre, V. (2015). The transcription factors SOX9 and SOX5/SOX6 cooperate genome-wide through super-enhancers to drive chondrogenesis. Nucleic Acids Res. *43*, 8183–8203.

44. Tagariello, A., Heller, R., Greven, A., Kalscheuer, V.M., Molter, T., Rauch, A., Kress, W., and Winterpacht, A. (2006). Balanced translocation in a patient with craniosynostosis disrupts the SOX6 gene and an evolutionarily conserved non-transcribed region. J. Med. Genet. *43*, 534–540.

45. Wang, Z., Li, J., Li, K., and Xu, J. (2018). SOX6 is downregulated in osteosarcoma and suppresses the migration, invasion and epithelial-mesenchymal transition via TWIST1 regulation. Mol. Med. Rep. *17*, 6803–6811.

46. Xie, Q., Chen, X., Lu, F., Zhang, T., Hao, M., Wang, Y., Zhao, J., McCrae, M.A., and Zhuang, H. (2012). Aberrant expression of microRNA 155 may accelerate cell proliferation by targeting sex-determining region Y box 6 in hepatocellular carcinoma. Cancer *118*, 2431–2442.

47. Guo, X., Yang, M., Gu, H., Zhao, J., and Zou, L. (2013). Decreased expression of SOX6 confers a poor prognosis in hepatocellular carcinoma. Cancer Epidemiol. *37*, 732–736.

48. Qiu, Z. (2018). Deciphering MECP2-associated disorders: disrupted circuits and the hope for repair. Curr. Opin. Neurobiol. *48*, 30–36.

49. Pingault, V., Bodereau, V., Baral, V., Marcos, S., Watanabe, Y., Chaoui, A., Fouveaut, C., Leroy, C., Vérier-Mine, O., Francannet, C., et al. (2013). Loss-of-function mutations in SOX10 cause Kallmann syndrome with deafness. Am. J. Hum. Genet. *92*, 707–724.

50. Philibert, P., Leprieur, E., Zenaty, D., Thibaud, E., Polak, M., Frances, A.M., Lespinasse, J., Raingeard, I., Servant, N., Audran, F., et al. (2010). Steroidogenic factor-1 (SF-1) gene mutation as a frequent cause of primary amenorrhea in 46,XY female adolescents with low testosterone concentration. Reprod. Biol. Endocrinol. *8*, 28.

51. Bastian, C., Muller, J.B., Lortat-Jacob, S., Nihoul-Fékété, C., Bignon-Topalovic, J., McElreavey, K., Bashamboo, A., and Brauner, R. (2015). Genetic mutations and somatic anomalies in association with 46,XY gonadal dysgenesis. Fertil. Steril. *103*, 1297–1304.

52. Chassaing, N., Causse, A., Vigouroux, A., Delahaye, A., Alessandri, J.L., Boespflug-Tanguy, O., Boute-Benejean, O., Dollfus, H., Duban-Bedu, B., Gilbert-Dussardier, B., et al. (2014). Molecular findings and clinical data in a cohort of 150 patients with anophthalmia/microphthalmia. Clin. Genet. *86*, 326–334.

53. Meyer, J., Südbeck, P., Held, M., Wagner, T., Schmitz, M.L., Bricarelli, F.D., Eggermont, E., Friedrich, U., Haas, O.A., Kobelt, A., et al. (1997). Mutational analysis of the SOX9 gene in campomelic dysplasia and autosomal sex reversal: lack of genotype/phenotype correlations. Hum. Mol. Genet. *6*, 91–98.

54. Zhu, N., Welch, C.L., Wang, J., Allen, P.M., Gonzaga-Jauregui, C., Ma, L., King, A.K., Krishnan, U., Rosenzweig, E.B., Ivy, D.D., et al. (2018). Rare variants in SOX17 are associated with pulmonary arterial hypertension with congenital heart disease. Genome Med. *10*, 56.