HHS Public Access

Author manuscript

Genet Med. Author manuscript; available in PMC 2022 May 03.

Published in final edited form as:

Genet Med. 2020 March ; 22(3): 524–537. doi:10.1038/s41436-019-0657-0.

Widening of the genetic and clinical spectrum of Lamb–Shaffer syndrome, a neurodevelopmental disorder due to *SOX5* haploinsufficiency

A full list of authors and affiliations appears at the end of the article.

Abstract

Purpose: Lamb–Shaffer syndrome (LAMSHF) is a neurodevelopmental disorder described in just over two dozen patients with heterozygous genetic alterations involving *SOX5*, a gene encoding a transcription factor regulating cell fate and differentiation in neurogenesis and other discrete developmental processes. The genetic alterations described so far are mainly microdeletions. The present study was aimed at increasing our understanding of LAMSHF, its clinical and genetic spectrum, and the pathophysiological mechanisms involved.

Methods: Clinical and genetic data were collected through GeneMatcher and clinical or genetic networks for 41 novel patients harboring various types of *SOX5* alterations. Functional consequences of selected substitutions were investigated.

Results: Microdeletions and truncating variants occurred throughout *SOX5*. In contrast, most missense variants clustered in the pivotal SOX-specific high-mobility-group domain. The latter variants prevented *SOX5* from binding DNA and promoting transactivation in vitro, whereas missense variants located outside the high-mobility-group domain did not. Clinical manifestations and severity varied among patients. No clear genotype–phenotype correlations were found, except that missense variants outside the high-mobility-group domain were generally better tolerated.

Conclusions: This study extends the clinical and genetic spectrum associated with LAMSHF and consolidates evidence that *SOX5* haploinsufficiency leads to variable degrees of intellectual disability, language delay, and other clinical features.

Keywords

autism; developmental delay; intellectual disability; epilepsy; missense variants

Correspondence: Véronique Lefebvre (lefebvrev1@email.chop.edu) or Christel Depienne (christel.depienne@uni-due.de).
These authors contributed equally: Ash Zawerton, Cyril Mignot, Ashley Sigafoos
co-first authors: Ash Zawerton, Cyril Mignot, Ashley Sigafoos
Joint senior authors: Véronique Lefebvre, Karl J. Clark, Christel Depienne

SUPPLEMENTARY INFORMATION

The online version of this article (https://doi.org/10.1038/s41436-019-0657-0) contains supplementary material, which is available to authorized users.

DISCLOSURE

M.J.G.S., K. McWalter, R.E.S., and Z.Z. are employees of GeneDx, Inc. S.I. is employed by Ambry Genetics, a company that provides testing for multigene panels and medical exome sequencing. The Department of Molecular and Human Genetics at Baylor College of Medicine receives revenue from clinical genetic testing performed at Baylor Genetics Laboratories. The other authors declare no conflicts of interest.

INTRODUCTION

The SOX protein family is made of transcription factors harboring a high-mobility-group (HMG) domain at least 50% similar to that of *SRY* (encoded by the sex-determining region on the Y chromosome).[1] This domain mediates DNA binding and bending, nuclear trafficking, and protein–protein interactions. The 20 SOX proteins existing in humans and other mammals fall into eight groups (SOXA to SOXH) based on sequence identity within and outside this domain.[2,3] Most have been shown in animal models to play pivotal roles in determining the lineage choice, differentiation program, and survival capacity of discrete cell types, such that as a whole the SOX family controls many crucial biological processes, including sex determination, neurogenesis, and skeletogenesis.[1] In humans, pathogenic variants in half of the SOX genes were shown to date to cause developmental disorders.[4] For example, *SRY* variants cause XY sex reversal (MIM 400045 and 400046);[5] *SOX9* variants cause campomelic dysplasia with or without XY sex reversal (MIM 114290);[5] *SOX18* variants cause hypotrichosis–lymphedema–telangiectasia syndrome (MIM 607823 and 137940);[6] and *SOX4* and *SOX11* (MIM 615866) variants cause Coffin–Siris syndrome–like syndromes.[7,8] Most pathogenic variants are de novo and, except for *SRY*, result in dominant disorders because of gene haploinsufficiency.

Lamb–Shaffer syndrome (LAMSHF, MIM 616803) was initially described as a condition caused by de novo deletions ranging from a few kilobases to several megabases and including at least part of *SOX5*.[9] LAMSHF is clinically characterized by developmental delays, language and motor deficits, intellectual disability, behavioral disturbances including autistic traits, and other, partially penetrant features.[9–12] *SOX5* is located on chromosome 12p12.1 and gives rise to at least five transcript isoforms through expression from different promoters, alternative start site usage, and alternative precursor messenger RNA (pre-mRNA) splicing. The longest isoform (NM_006940) encodes a 763–amino acid protein (originally referred to as L-*SOX5*, but more recently and henceforward called *SOX5*) and is the predominant brain isoform.[13] The shortest isoform (NM_178010) encodes a protein corresponding to the L-*SOX5* C-terminal half and is testis-specific. All long protein isoforms contain the same functional domains and are collectively critical in mouse development.[14] $SOX5^{-/-}$ mice are born with lethal skeletal malformations and with defective deep-layer cortical projection neurons, while $SOX5^{+/-}$ mice have a normal lifespan and no obvious abnormalities.[15–18]

To date, only a few *SOX5* point variants, mostly introducing premature termination codons, have been reported in LAMSHF patients[19–21] or in large genetic studies of developmental disorders without detailed clinical descriptions.[22–25] In this study, we describe 41 unpublished patients carrying various *SOX5* deletions and point variants, including 16 with missense variants. We delineate more precisely the clinical spectrum associated with *SOX5* alterations, aim at establishing genotype–phenotype correlations, and explore pathogenicity of selected variants using both in silico and functional approaches.

MATERIALS AND METHODS

Human subjects

We collected clinical and molecular data from patients with *SOX5* microdeletions or point variants through GeneMatcher,[26] DECIPHER[27] (patient IDs 333039, 340665, 271393, 264625), and clinical networks. Referring physicians used standard developmental scales and filled out a table with detailed developmental, neurological, and behavioral history, including imaging and electroencephalogram (EEG) data where available. The study was approved by INSERM (RBM C12–06). We obtained informed written consent for all genetic studies as well as for the use of photographs shown in Fig. 2g.

Genetic studies

Diagnostic laboratories performed genetic tests on blood samples using microarrays or next-generation sequencing (Supplementary Table 1). *SOX5* variants and deletions were validated and searched for in parents using Sanger sequencing and fluorescence in situ hybridization (FISH) or real-time polymerase chain reaction (PCR), respectively. *SOX5* variants were described based on the longest isoform (NM_006940.5) using Alamut 2.11 (Interactive Biosoftware, France) and Human Genome Variation Society guidelines (www.hgvs.org/mutnomen). The InterVarinterface was used to classify *SOX5* variants with adjusted criteria according to American College of Medical Genetics and Genomics (ACMG) recommendations.[28,29] Combined annotation dependent depletion (CADD) scores[30] were calculated for each variant (Supplementary Table 1). *SOX5* isoforms and promoters and other SOX sequences were retrieved from the National Center for Biotechnology Information (NCBI) and Fantom5databases and sequences were aligned using ClustalW (MacVector16 software). The effects of missense variants on protein structure and function were predicted using HOPE[31] and Swiss-Model.[32] *SOX5* variants were queried in human populations using gnomAD. Data were statistically analyzed using Fisher's exact and Wilcoxon–Mann–Whitney tests.

SOX5 plasmids

Expression plasmids for the longest *SOX5* isoform and variants thereof were generated in the pKTol2C-EGFP plasmid.[33] The EGFP sequence was replaced with custom-synthesized or PCR-amplified *SOX5* sequences (primers are available upon request). Plasmid integrity was verified using Sanger sequencing.

SOX5 immunolocalization

HEK-293 cells (ATCC® CRL-1573™) were plated on glass coverslips and transfected with pKTol2C-SOX5 plasmids (2 μg) and Lipofectamine 2000 Transfection Reagent (Thermo Fisher Scientific). Two days later, they were stained using Image-IT™ LIVE Plasma Membrane and Nuclear Labeling Kit (Thermo Fisher, I34406), fixed in 4% paraformaldehyde, permeabilized with 0.1% Triton X-100 in PBS (PBST), and blocked in PBST supplemented with 1% BSA and 22.5 mg/ml glycine. They were then incubated with rabbit polyclonal SOX5 antibody (1:200, Abcam, ab94396), followed by goat anti-rabbit antibody (1:500, Alexa Fluor 488, Invitrogen, A27034). After placing DAPI-containing

Vectashield antifade mounting medium (Vector Laboratories), cells were imaged by confocal laser scanning microscopy (Zeiss LSM 780; 100× objective).

Western blot, electrophoretic mobility shift, and dimerization assays

HEK-293 cells were plated in six-well dishes and transfected eight hours later with empty or SOX5 expression plasmid (1 μg) and FuGENE6 (3 μl, Promega). The next day, extracts were prepared using NE-PER Nuclear and Cytoplasmic Extraction Reagents (Thermo Fisher Scientific) and tested by western blotting using SOX5 antibody (1:1000) and horseradish peroxidase–conjugated goat anti-rabbit IgG (1:5000, Vector Biolabs). Signals were visualized using ECL Prime Western Blotting Detection Reagents (Amersham). Electrophoretic mobility shift assay (EMSA) was conducted using the same extracts, 10 fmoles [$\alpha-^{32}$P]-dCTP-labeled 2HMG probe and 1 μg poly(dG-dC).poly(dG-dC), as described.[34] Homodimerization was tested in western blots following cell extract incubation for 10 minutes with 0.01% glutaraldehyde.

Reporter assay

HEK-293 cells were transfected with FuGENE6 containing 150 ng pSV2βGal, 500 ng Acan [4xA1]-p89Luc reporter, 50 ng SOX9 expression plasmid, and 300 ng plasmid encoding no protein, wild-type (WT) SOX5, and/or variant SOX5, as previously described.[35] Forty hours later, cells were collected in Tropix Lysis buffer (Applied Biosystems) with protease inhibitor cocktail (Thermo Fisher Scientific) and tested using Dual-Light luciferase and *E. coli* β-galactosidase assays (Thermo Fisher Scientific). Reporter activities were calculated as means with standard deviation of luciferase values measured for triplicates and normalized for transfection efficiency using β-galactosidase values.

RESULTS

The *SOX5* variant spectrum associated with LAMSHF includes missense variants

We collected genetic and clinical information from 41 patients (Table 1, supplementary table 1). Eight patients (D1–D8), representing seven families, carried novel pathogenic microdeletions. These microdeletions ranged from 43.7 kb to 1.7 Mb and involved different breakpoints (Fig. 1a). While the largest deletion encompassed the entire *SOX5* gene and its 5' neighbor (*BCAT1*), the others were restricted to various segments of *SOX5*.

The other 33 patients belonged to 31 families and totaled 23 distinct point variants. Nineteen of these variants were classified by the InterVar interface as pathogenic or likely pathogenic (P1–P29) and the other four as variants of unknown significance (VUS) (V1–V4). Two patients had indels introducing frameshifts (P7 and P10) (Table 1, Fig. 1a, b). Two (P13 and P14) had variants altering the acceptor and donor splice sites of the coding exon 12, respectively. Thirteen patients (including a pair of dizygotic twins) totaled eight distinct nonsense variants (P1–P6, P8–P10, P11, P12, P15, P25, and P28). Truncating variants (i.e., nonsense, splice site, and frameshift variants) were scattered over the L-*SOX5* isoform from the N-terminus to the middle of the HMG domain. All truncating *SOX5* variants thus encode proteins lacking DNA-binding ability. Furthermore, since all variants spare the last exon, they likely trigger nonsense-mediated mRNA decay and thus prevent protein expression.

Sixteen patients (including a sib pair) had 11 different missense variants (P16–P24, P26, P27, P29, and V1–V4). Seven of these variants were clustered in the HMG domain, while the four VUS occurred in the first coiled coil (V1), between the coiled coils (V2), or after the HMG domain (V3 and V4).

Five identical nucleotide transitions were identified in several unrelated individuals: c.622C>T, p.Gln208* (P2 and P3); c.637C>T, p.Arg213* (P4/P5 and P6); c.1477C>T, p.Arg493* (P11 and P12); c.1678A>G, p.Met560Val (P16 and P17); and c.1711C>T, p.Arg571Trp (P19 to P23). Besides a few that were of unknown inheritance, these alterations were all de novo and thus suggested the presence of hot spots for nucleotide transitions. Of additional note, 17 of 22 single-nucleotide variants identified in patients are C>T and G>A transitions, suggesting that many *SOX5* point variants result from cytosine deamination, a prevailing mechanism of genetic alteration.[36]

High rate of parental mosaicism

Most microdeletions and variants predicted to be pathogenic or likely pathogenic were undetected in parental blood samples, suggesting de novo occurrence (25/34 families, 74%). However, in each of three families, the same alteration was found in two affected siblings (D6 and D7; P4 and P5; and P22 and P23), but not in their parents, and in two other patients (D3 and P9), the variant was present at low levels in maternal blood. In addition, one nonsense variant was transmitted to a patient (P1) from his affected mother, where it was de novo. Variant transmission could not be determined for four patients (D1, P7, P12, P19) due to unavailability of parental samples. These findings thus indicate that pathogenic LAMSHF variants are frequently inherited from a mosaic parent (5/34, 15%) and also occasionally from an affected parent (1/34, 3%).

Wide clinical spectrum associated with *SOX5* pathogenic alterations

Excluding the four patients with VUS, our patient series comprised 20 females and 17 males (Supplementary Table 1). The patients were 12.2 years of age on average at the time of examination (median: 8.0 years, range: 1.75–36), with 11 older than 15 and six younger than 4.

For most patients, pregnancy and delivery were unremarkable (21/36), birth measurements (weight, length, and head circumference) normal (15/18 for whom full information was available), and the neonatal period uneventful (25/36). Eight patients had mild growth retardation or a small head at birth, two were hypotonic, and three had feeding difficulties.

Developmental delay was present in all patients for whom information was available. Although more than half of the patients timely acquired the sitting position (≤9 months; n = 16/29), the age of walking was delayed in all but one (>18 months; n = 35/36), without clear timing differences among variant categories (Fig. 2a, b). The age of first words was delayed in 21/26 patients (>12 months; mean: 29.9 months, range: 10–60 months). The delay was significantly less pronounced in patients with missense variants (mean: 22.4 months, n = 11) than in those with deletions and truncating variants (mean: 35.2 months, n = 15; p value: 0.04, Wilcoxon rank sum test) (Fig. 2c). The levels of verbal expression were variable,

but most patients older than three years could make short or full sentences (Supplementary Tables 1 and 2).

Intellectual disability (ID) was reported in 30/33 patients, with 27 having mild-to-moderate ID and 3 having severe or moderate-to-severe ID. The three patients without ID had learning difficulties and either borderline functioning or discrepant verbal/performance IQ scores. No significant correlation was observed between degree of ID and variant type (Fig. 2d).

Of 25 patients evaluated for autism spectrum disorder (ASD), 6 (4 with truncating variants and 2 with missense variants) were positively diagnosed (24%) and 11 had other behavioral disturbances including stereotypies, isolation, tantrums, and hyperactivity (Fig. 2e). Of 36 patients, 8 experienced epileptic seizures (22%), but 5 of these had only one or two episodes and did not require medication. One of these patients (D6) had seizures triggered by environmental photosensitivity, an unusual finding in a “developmental delay plus seizures” syndrome (Supplementary Fig. 1). No correlation was found between the occurrence of seizures and the *SOX5* variant type (Fig. 2f).

Clinical examination revealed that stature and weight were within normal range for most patients. Head circumference of both males ($n = 14$) and females ($n = 15$) was in the low but normal range (~−1.5 SD) while two patients (P14 and P23) had microcephaly. Hypotonia was reported in 22 patients, and five had additional neurological features, including ataxia ($n = 2$) or pyramidal syndrome ($n = 3$). Thirty-one patients had mild dysmorphic facial features, including broad/full nasal tip ($n = 9$), thin upper lip or full lips ($n = 8$), small jaw or chin ($n = 5$), long face ($n = 3$), or epicanthus ($n = 3$). Strabismus was reported in 13 patients, optic atrophy in 5, and amblyopia or cortical visual impairment in 1 each. Except for thin optic nerves, brain magnetic resonance image (MRI) scans were normal or showed nonspecific anomalies. Besides dysmorphic facial features, other skeletal malformations included scoliosis in six patients, thoracic kyphosis and hip dysplasia in one patient each, and fused cervical vertebrae in two patients (Supplementary Table 1). Malformations of other organs were rare and restricted to individual patients. Again, no correlation was found between the occurrence of these features and the variant types. Moreover, patients with recurrent variants (e.g., P2–P3: p. Gln208*, P4-P6: p.Arg213*, P16–P17: p.Met560Val, and P19–P23: p.Arg571Trp) exhibited considerable clinical variability, indicating that factors other than the *SOX5* variants modulate the expression of the clinical phenotype.

SOX5 is tightly conserved in the general population

We used gnomAD, a genomic database for over 140,000 individuals who are theoretically unrelated and lacking severe pediatric disease, to investigate conservation constraints on *SOX5* in humans.[37] While 158 synonymous variants were predicted and 159 were observed (Z- score: −0.08), 42 loss-of-function variants were expected, but only 3 were observed (probability of loss-of-function intolerance [pLI] = 1). Moreover, 427 missense variants were predicted, but only 244 were observed (Z-score: 3.21). Thus, *SOX5* is under tight conservation constraint in control populations. Interestingly, gnomAD synonymous variants were found for 10–29% residues both within and outside functional domains, whereas missense variants altered significantly fewer residues in the HMG domain (six residues, i.e., 7.5%) than in other regions (21–33%) and significantly fewer than synonymous variants

(20%, $p = 0.017$) (Fig. 3a, b). The *SOX5* HMG domain is thus highly constrained within control populations, which is in contrast to the relatively high prevalence of HMG domain missense variants observed in our patient cohort. The first coiled-coil domain also had significantly fewer missense variants (20.7%) than the regions outside of known functional domains (33.2%; $p = 0.03$), suggesting that this domain, which is required for *SOX5* homodimerization and thereby for binding to pairs of recognition sites in target genes, is also under conservation constraint.

The six HMG domain missense variants found in gnomAD affected two of the same residues as in LAMSHF patients and four others, and all six occurred only once (Supplementary Table 5). In contrast, for missense variants located outside the HMG domain, we found four occurrences of the patient Arg235Cys variant (located in the first coiled coil) in gnomAD, and one for Ser693Leu. Other gnomAD variants affected the same residues as in patients, such as Arg235His, found in 11 individuals. These observations suggest that some *SOX5* variants, especially those located outside the HMG domain, may be better tolerated than others.

In silico prediction of variant pathogenicity

To predict pathogenicity of *SOX5* missense variants, we first examined the location and conservation of affected residues. Since all HMG domain residues are fully or semiconserved in *SOX5* vertebrate orthologs (Supplementary Fig. 2a), we focused on human SOX protein paralogs. All HMG domain residues altered in patients and gnomAD individuals affected residues involved in DNA binding or bending, α-helical configuration, or nuclear trafficking (Fig. 3c). Interestingly, 3 of the 5 residues altered in patients (Met560, Asn561 and Arg571) were among 23 residues identical in all protein paralogs, Tyr605 was among 13 semiconserved residues, and only Ala596 was among the 40 nonconserved residues. Conversely, only two of the six residues altered in gnomAD individuals were among the conserved and semiconserved ones. Outside the HMG domain, patient variants affected residues that are highly conserved in *SOX5* and its orthologs (Supplementary Fig. 2b). When the comparison was limited to human SOXD proteins (*SOX5*, SOX6, and SOX13), these conservation patterns held strongly for Arg235Cys, located in the first coiled-coil domain, and Thr632Asn, immediately flanking the HMG domain, but less strongly for residues located in functionally unknown regions (Fig. 3d). Together, these data suggested that all HMG domain variants and a few other patient variants might impact *SOX5* function.

We then asked whether the HMG domain residues altered in LAMSHF patients also cause disease when altered in other SOX genes. Interestingly, all residues affected in LAMSHF patients were shown to cause gonadal dysgenesis or XY sex reversal when altered in *SRY*, or campomelic dysplasia with or without XY sex reversal when altered in *SOX9* (Supplementary Table 6). In contrast, only two of the four variants found in gnomAD, but not in LAMSHF patients, were shown to cause disease when altered in *SRY*. These data further support pathogenicity of patient variants. They also suggest that some variants present in gnomAD individuals could be pathogenic, but clinical information was unavailable to validate this possibility.

Lastly, comparison of WT and variant residues using HOPE (Supplementary Fig. 3) showed that all variants differed from WT residues by at least one major structural feature: 16/18 differed in size, 13/18 differed in hydrophobicity, and 6/6 had a neutral instead of positive charge. All variants could thus affect the secondary structure and hence function of *SOX5*.

Overall, these analyses concurred that most missense variants identified in our patient series are likely pathogenic.

Truncating variants and missense variants located within or near nuclear import signals impair SOX5 translocation to the nucleus

We constructed expression plasmids for WT and variant forms of L-SOX5 and transiently transfected them in HEK-293 cells to explore the functional impacts of variants. Western blots of nuclear and cytoplasmic fractions (Fig. 4a) and cell immunostaining assays (Fig. 4b) showed that WT SOX5 localized primarily in the nucleus, as expected. On the contrary, expression of nonsense variants (Gln208*, Gln274*, Gly354*, and Arg493*) revealed that, if these variants were expressed in patients' cells (i.e., if their mRNAs were not subjected to nonsense-mediated decay), they would be primarily cytoplasmic. This result was expected since protein truncation occurs before the nuclear translocation signals. All proteins with a missense variant that we tested were able to translocate into the nucleus, except those in which the variant occurred within or near the N-terminal nuclear import signal. Accordingly, the Met560Val variant was localized to both the cytoplasm and nucleus, and the Asn561His and Arg571Trp variants were mainly cytoplasmic. Cytoplasmic retention of these missense variants may thus contribute to pathogenicity.

Missense variants in the HMG domain prevent SOX5 from participating in transactivation

We tested the transcriptional activity of SOX5 variants by transfecting HEK-293 cells with an Acan reporter whose enhancer is synergistically activated by SOX9 and SOXD proteins.[35] WT SOX5 increased transactivation by SOX9 in a dose-dependent manner (Fig. 4c). Nonsense and HMG domain missense variants exhibited little if any activity, whereas missense variants located outside the HMG domain had activity similar to WT. Since *SOX5* variants are heterozygous in our patients, we also tested whether they could interfere with the activity of WT SOX5. Nonsense and HMG domain missense variants did not affect the activity of WT SOX5, and missense variants located outside the HMG domain increased the reporter activity as much as WT SOX5 (Fig. 4d). Thus, none of the variants showed a dominant-negative effect.

We then tested the DNA-binding ability of SOX5 missense variants in EMSA using whole-cell extracts from HEK-293 cells transfected with SOX5 plasmids and a probe avidly binding SOXD homodimers.[38] HMG domain missense variants failed to bind DNA, whereas other missense variants efficiently bound DNA (Fig. 4e). This result also suggested that Arg235Cys, located in the main coiled-coil domain, can homodimerize effectively. Its ability to homodimerize was confirmed in an assay where closely interacting proteins were crosslinked with glutaraldehyde (Fig. 4f).

In conclusion, HMG domain missense variants prevented SOX5 from binding DNA and from participating in transcriptional activation, supporting their pathogenicity. On the

contrary, variants located outside the HMG domain had no deleterious impact in the assays used, but this finding does not rule out that they could be pathogenic and alter other, untested SOX5 activities.

DISCUSSION

LAMSHF syndrome was previously described in just over two dozen patients. Most patients had deletions of at least part of *SOX5*, and a few had either a chromosomal translocation involving *SOX5*, or *SOX5* nonsense or frameshift variants.[9–12,19–21] Our patient series more than doubles the number of cases described in the literature and demonstrates that *SOX5* missense variants clustering in the HMG domain can also cause LAMSHF syndrome. All variants were heterozygous, and most were predicted in silico and validated in vitro to be loss-of-function variants. This confirms that *SOX5* haploinsufficiency is deleterious for neurogenesis and a few other developmental processes. Our study also revealed that parental mosaicism, found in at least 14% of families in our series, is relatively frequent in LAMSHF syndrome. This finding is important for genetic counseling and in line with increasing evidence that somatic, gonosomal, or gonadal mosaicism in parents may cause recurrence of neurodevelopmental disorders, apparently due to de novo variants.[39] *SOX5* and LAMSHF syndrome thus expand the list of such genes and disorders.

Our extended study allowed further definition of the LAMSHF clinical features. ID is mostly within the mild-to-moderate range, and some cases have specific cognitive deficits rather than ID.[9] Delays in motor and language acquisition are observed in all patients and correlate with the level of ID. Behavioral disturbances are frequent and include ASD or autistic traits, as previously reported.[9,10,40] Microcephaly is infrequent; yet, brain growth seems frequently mildly altered. Hypotonia is common, whereas other neurological features are infrequent. Our findings also suggest that *SOX5* pathogenic variants predispose to epilepsy, with a prevalence of an order of magnitude higher than in the general population. Seizures in *SOX5* patients usually respond well to antiepileptic treatments and follow a benign course. Ophthalmologic features, including strabismus, optic nerve atrophy, amblyopia, and cortical visual impairment, are frequently observed[9,19,22] and, together with rare skeletal malformations (i.e., scoliosis and fused cervical vertebrae), constitute corroborating rather than defining features of LAMSHF syndrome.[9] The incomplete penetrance observed for some features suggests that *SOX5* haploinsufficiency manifests differently in distinct individual genetic backgrounds or that some variants retain partial activity. The investigation of clinical features according to variant types, however, did not reveal clear genotype–phenotype correlations. Patients with HMG domain missense variants tended to have milder language deficits, but this finding requires confirmation with larger patient cohorts. Based on the lack of obvious genotype–phenotype correlations and on the observation of variable phenotype severity in unrelated individuals with identical *SOX5* variants, we tentatively conclude that yet-unidentified factors significantly contribute to the penetrance and degree of disease severity.

We also describe in this study four patients with de novo variants located outside the HMG domain and altering amino acids conserved in SOX5 orthologs. However, the pathogenicity of these variants could not be established through functional assays, and it thus remains

unclear whether and how these variants contribute to disease in these patients. Three of these patients (V2–V4) had phenotypic features compatible with LAMSHF syndrome (although patient V2 was very young at the time of the study and patient V3 mainly had ASD), whereas the fourth patient (V1) had Tourette syndrome. The variant identified in the latter patient (Arg235Cys) was also present in four gnomAD individuals from different ethnicities. Although Tourette patients are included in gnomAD "neuro" cohorts, the individuals with Arg235Cys were not in these cohorts, suggesting that these individuals had no obvious neurological phenotype. Further investigations are therefore warranted to investigate whether missense variants outside the HMG domain could impair untested activities of SOX5 and whether these variants could predispose to LAMSHF or Tourette syndrome.

In conclusion, our study demonstrates that the genetic and clinical spectrum in LAMSHF syndrome is much larger than previously described, and extends to missense variants clustering in the HMG domain. In silico and in vitro functional data support the concept that these missense variants are pathogenic by causing loss of function of the SOX5 transcription factor, and thereby reflect gene haploinsufficiency during neurogenesis and occasionally during other developmental processes. The impacts of variants located outside the HMG domain remain to be determined.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Ash Zawerton, MS[1], Cyril Mignot, MD, PhD[2,3], Ashley Sigafoos, BSc[4], Patrick R. Blackburn, PhD[5], Abdul Haseeb, PhD[6], Kirsty McWalter, MS, CGC[7], Shoji Ichikawa, PhD[8], Caroline Nava, MD, PhD[2,3], Boris Keren, MD, PhD[2,3], Perrine Charles, MD, PhD[3], Isabelle Marey, MD[3], Anne-Claude Tabet, MD, PhD[9,10], Jonathan Levy, MD, PhD[9], Laurence Perrin, MD[9], Andreas Hartmann, MD[2,11], Gaetan Lesca, MD, PhD[12,13], Caroline Schluth-Bolard, MD, PhD[12,13], Pauline Monin, MD[12], Sophie Dupuis-Girod, MD, PhD[12,14], Maria J. Guillen Sacoto, MD[7], Rhonda E. Schnur, MD[7], Zehua Zhu, PhD[7], Alice Poisson, MD, PhD[15], Salima El Chehadeh, MD[16], Yves Alembik, MD[16], Ange-Line Bruel, PhD[17,18], Daphné Lehalle, MD, PhD[17,19], Sophie Nambot, MD[17,19], Sébastien Moutton, MD[17,19], Sylvie Odent, MD[20,21], Sylvie Jaillard, MD, PhD[22], Christèle Dubourg, PharmD, PhD[21,23], Yvonne Hilhorst-Hofstee, MD, PhD[24], Tina Barbaro-Dieber, MD[25], Lucia Ortega, MD[25], Elizabeth J. Bhoj, MD, PhD[26], Diane Masser-Frye, MS, MSW[27], Lynne M. Bird, MD[27,28], Kristin Lindstrom, MD[29], Keri M. Ramsey, RN[30], Vinodh Narayanan, MD[30], Emily Fassi, MS[31], Marcia Willing, MD, PhD[31], Trevor Cole, MBChB, FRCP[32], Claire G. Salter, BMBS, MRCPCH[32,33], Rhoda Akilapa, BMBS, MRCPCH[34], Anthony Vandersteen, MD, PhD[35], Natalie Canham, MBChB[36,37], Patrick Rump, MD, PhD[38], Erica H. Gerkes, MD[38], Klein Jolien S. Wassink-Ruiter, MD[38], Emilia Bijlsma, MD, PhD[38], Mariëtte J. V. Hoffer, PhD[24], Marcelo Vargas, MD[39,40], Antonina Wojcik, MS, CGC[39,40], Florian Cherik, MD[41], Christine Francannet, MD[41], Jill A. Rosenfeld, MS[42], Keren Machol, MD, PhD[42], Daryl A. Scott, MD, PhD[42,43], Carlos

A. Bacino, MD[42], Xia Wang, PhD[42], Gary D. Clark, MD[44], Marta Bertoli, MD[45], Simon Zwolinski, PhD[45], Rhys H. Thomas, MBChB, PhD[46,47], Ela Akay, MD[47], Richard C. Chang, MD[48], Rebekah Bressi, MS[48], Rossana Sanchez Russo, MD[49], Myriam Srour, MD, PhD[50], Laura Russell, MD[51], Anne-Marie E. Goyette, MD[52], Lucie Dupuis, MS[53], Roberto Mendoza-Londono, MD[53], Catherine Karimov, MD[54], Maries Joseph, MD[55], Mathilde Nizon, MD[56,57], Benjamin Cogné, PharmD[56,57], Alma Kuechler, MD[58], Amélie Piton, PhD[59,60], Deciphering Developmental Disorder Study[61], Eric W. Klee, PhD[5,62], Véronique Lefebvre, PhD[6], Karl J. Clark, PhD[4], Christel Depienne, PhD[2,58,60]

Affiliations

[1]Department of Cellular & Molecular Medicine, Cleveland Clinic Lerner Research Institute, Cleveland, OH, USA.

[2]INSERM U 1127, CNRS UMR 7225, Sorbonne Universités, UPMC Univ Paris 06 UMR S 1127, Institut du Cerveau et de la Moelle épinière, ICM, Paris, France.

[3]AP-HP, Hôpital Pitié-Salpêtrière, Département de Génétique et de Cytogénétique; Centre de Référence Déficiences Intellectuelles de Causes Rares, GRC UPMC « Déficience Intellectuelle et Autisme », Paris, France.

[4]Department of Biochemistry and Molecular Biology, Mayo Clinic, Rochester, MN, USA.

[5]Center for Individualized Medicine, Department of Health Science Research, and Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA.

[6]Department of Surgery, Division of Orthopaedic Surgery, The Children’s Hospital of Philadelphia, Philadelphia, PA, USA.

[7]GeneDx, Gaithersburg, MD, USA.

[8]Department of Clinical Diagnostics, Ambry Genetics, Aliso Viejo, CA, USA.

[9]Genetics Department, Robert Debré Hospital, APHP, Paris, France.

[10]Human Genetics and Cognitive Functions, Institut Pasteur, Paris, France.

[11]APHP, Department of Neurology, Hôpital de la Pitié-Salpêtrière, Paris, France.

[12]Service de Génétique, Hospices Civils de Lyon – GHE, Lyon, France.

[13]CNRS UMR 5292, INSERM U1028, CNRL, and Université Claude Bernard Lyon 1, GHE, Lyon, France.

[14]Centre de référence pour la maladie de Rendu-Osler, Bron, France.

[15]GénoPsy, Reference Center for Diagnosis and Management of Genetic Psychiatric Disorders, Centre Hospitalier le Vinatier and EDR-Psy Team (CNRS & Lyon 1 Claude Bernard University), Lyon, France.

[16]Département de Génétique Médicale, CHU de Hautepierre, Strasbourg, France.

[17]INSERM 1231 LNC, Génétique des Anomalies du Développement, Université de Bourgogne-Franche Comté, Dijon, France.

[18]FHU-TRANSLAD, Université de Bourgogne/CHU Dijon, Dijon, France.

[19]Centre de Génétique et Centre de Référence Maladies Rares «Anomalies du Développement de l'Interrégion Est», Hôpital d'Enfants, CHU Dijon Bourgogne, Dijon, France.

[20]CHU de Rennes, service de génétique clinique, Rennes, France.

[21]Univ Rennes, CNRS, IGDR, UMR 6290, Rennes, France.

[22]Univ Rennes, CHU Rennes, Inserm, EHESP, Irset (Institut de recherche en santé, environnement et travail) - UMR_S 1085, Rennes, France.

[23]Service de Génétique Moléculaire et Génomique, CHU, Rennes, France.

[24]Department of Clinical Genetics, Leiden University Medical Center, Leiden, Netherlands.

[25]Cook Childrens Medical Center, Fort Worth, TX, USA.

[26]Department of Clinical Genetics, Children's Hospital of Philadelphia, Philadelphia, PA, USA.

[27]Rady Children's Hospital San Diego, Division of Genetics and Dysmorphology, San Diego, CA, USA.

[28]Department of Pediatrics, University of California–San Diego, San Diego, CA, USA.

[29]Division of Genetics and Metabolism, Phoenix Children's Hospital, Phoenix, AZ, USA.

[30]Translational Genomics Research Institute (TGen), Center for Rare Childhood Disorders, Phoenix, AZ, USA.

[31]Division of Genetics and Genomic Medicine, Department of Pediatrics, Washington University School of Medicine, St. Louis, MO, USA.

[32]West Midlands Regional Genetics Service and Birmingham Health Partners, Birmingham Women's and Children's NHS Foundation Trust, Birmingham, UK.

[33]RILD Wellcome Wolfson Centre, Royal Devon and Exeter NHS Foundation Trust, Exeter, UK.

[34]North West Thames Regional Genetics Service, Northwick Park Hospital, Harrow, London, UK.

[35]IWK Health Centre, Dalhousie University, Halifax, NS, Canada.

[36]North West Thames Regional Genetics Service, Northwick Park Hospital, London, UK.

[37]Cheshire & Merseyside Regional Genetics Service, Liverpool Women's Hospital, Liverpool, UK.

[38]Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, Netherlands.

[39]Gillette Children's Specialty Healthcare, St. Paul, MN, USA.

[40]Children's Minnesota, Minneapolis, MN, USA.

[41]Service de génétique clinique, Centre de Référence Maladies Rares «Anomalies du Développement et syndromes malformatifs du Sud-Est", CHU de Clermont-Ferrand, Clermont-Ferrand, France.

[42]Department of Molecular & Human Genetics, Baylor College of Medicine, Houston, TX, USA.

[43]Department of Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, TX, USA.

[44]Pediatrics–Neurology, Baylor College of Medicine, Houston, TX, USA.

[45]Northern Genetics Service— Newcastle upon Tyne NHS Foundation Trust, Newcastle upon Tyne, UK.

[46]Institute of Neuroscience, Newcastle University, Framlington Place, Newcastle upon Tyne, UK.

[47]Department of Neurology, Royal Victoria Infirmary, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK.

[48]Division of Metabolic Disorders, Children's Hospital of Orange County (CHOC), Orange, CA, USA.

[49]Department of Human Genetics, Emory Universit, Atlanta, GA, USA.

[50]Division of Pediatric Neurology, Department of Pediatrics, Montreal Children's Hospital, McGill University Health Center, Montreal, QC, Canada.

[51]Division of Medical Genetics, Department of Specialized Medicine, McGill University, Montreal, QC, Canada.

[52]Child Development Program, Department of Pediatrics, Montreal Children's Hospital, McGill University Health Center, Montreal, QC, Canada.

[53]Division of Clinical and Metabolic Genetics, The Hospital for Sick Children and University of Toronto, Toronto, ON, Canada.

[54]Children's hospital of Los Angeles, Los Angeles, CA, USA.

[55]Medical Genetics and Metabolism, Valley Children's Hospital, Madera, CA, USA.

[56]CHU Nantes, Service de Génétique Médicale, Nantes, France.

[57]INSERM, CNRS, UNIV Nantes, l'institut du thorax, Nantes, France.

[58]Institut für Humangenetik, Universitätsklinikum Essen, Universität Duisburg-Essen, Essen, Germany.

[59]Laboratoire de Diagnostic Génétique, Hôpitaux Universitaires de Strasbourg, Strasbourg, France.

[60]IGBMC, CNRS UMR 7104/INSERM U964/Université de Strasbourg, Illkirch, France.

[61]DDD Study, Wellcome Sanger Institute, Hinxton, Cambridge, UK.

[62]Department of Clinical Genomics, Mayo Clinic, Rochester, MN, USA

ACKNOWLEDGEMENTS

We thank the patients and their families for their participation in this study, and the C4RCD Research Group (Newell Belnap, Amanda Courtright, Ana Claasen, David Craig, Matt Huentelman, Madison LaFleur, Sampathkumar Rangasamy, Ryan Richholt, Isabelle Schrauwen, Ashley L. Siniard, and Szabolics Szelinger) for providing clinical information on patient P18. This research was funded in part by the Agence Nationale de la Recherche and European High-Functioning Autism Network (ANR EUHFAUTISM), the Assistance Publique–Hôpitaux de Paris (AP-HP), the Institut National de la Santé et de la Recherche Médicale (INSERM), the BioPsy labex (to Christel Depienne and C.N.) and the Association Française du Syndrome Gilles de la Tourette (AFSGT) to Christel Depienne. It was also funded by the Cleveland Clinic Lerner Research Institute (LRI Chair's Innovative Research Award to V.L.), and by Harper's Quest and the LAMSHF Syndrome Research Fund (donations to V.L.) and the Center for Individualized Medicine, Mayo Clinic. This study makes use of data generated by the DECIPHER community and the Deciphering Developmental Disorders (DDD) Study, which is funded by the Wellcome Trust. The DDD study presents independent research commissioned by the Health Innovation Challenge Fund (grant number HICF-1009-003), a parallel funding partnership between Wellcome and the Department of Health, and the Wellcome Sanger Institute (grant number WT098051). The views expressed in this publication are those of the author(s) and not necessarily those of Wellcome or the Department of Health. The study has UK Research Ethics Committee approval (10/H0305/83, granted by the Cambridge South REC, and GEN/284/12 granted by the Republic of Ireland REC). The research team acknowledges the support of the National Institute for Health Research, through the Comprehensive Clinical Research Network.

REFERENCES

1. Kamachi Y, Kondoh H. Sox proteins: regulators of cell fate specification and differentiation. Development. 2013;140:4129–4144. [PubMed: 24086078]
2. Bowles J, Schepers G, Koopman P. Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators. Dev Biol. 2000;227:239–255. [PubMed: 11071752]
3. Ji EH, Kim J. SoxD transcription factors: multifaceted players of neural development. Int J Stem Cells. 2016;9:3–8. [PubMed: 27426080]
4. Angelozzi M, Lefebvre V. SOXopathies: growing family of developmental disorders due to SOX mutations. Trends Genet. 2019;35:658–671. [PubMed: 31288943]
5. Sim H, Argentaro A, Harley VR. Boys, girls and shuttling of SRY and SOX9. Trends Endocrinol Metab. 2008;19:213–222. [PubMed: 18585925]
6. Irrthum A, Devriendt K, Chitayat D, et al. Mutations in the transcription factor gene SOX18 underlie recessive and dominant forms of hypotrichosis–lymphedema–telangiectasia. Am J Hum Genet. 2003; 72:1470–1478. [PubMed: 12740761]
7. Zawerton A, Yao B, Yeager JP, et al. De novo SOX4 variants cause a neurodevelopmental disease associated with mild dysmorphism. Am J Hum Genet. 2019;104:246–259. [PubMed: 30661772]
8. Tsurusaki Y, Koshimizu E, Ohashi H, et al. De novo SOX11 mutations cause Coffin–Siris syndrome. Nat Commun. 2014;5:4011. [PubMed: 24886874]
9. Lamb AN, Rosenfeld JA, Neill NJ, et al. 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]

10. Rosenfeld JA, Ballif BC, Torchia BS, et al. Copy number variations associated with autism spectrum disorders contribute to a spectrum of neurodevelopmental disorders. Genet Med. 2010;12:694–702. [PubMed: 20808228]
11. Lee RW, Bodurtha J, Cohen J, Fatemi A, Batista D. Deletion 12p12 involving SOX5 in two children with developmental delay and dysmorphic features. Pediatr Neurol. 2013;48:317–320. [PubMed: 23498568]
12. Schanze I, Schanze D, Bacino CA, Douzgou S, Kerr B, Zenker M. Haploinsufficiency of SOX5, a member of the SOX (SRY-related HMG-box) family of transcription factors is a cause of intellectual disability. Eur J Med Genet. 2013;56:108–113. [PubMed: 23220431]
13. Ikeda T, Zhang J, Chano T, et al. Identification and characterization of the human long form of SOX5 (L-SOX5) gene. Gene. 2002;298:59–68. [PubMed: 12406576]
14. Kwan KY, Lam MM, Krsnik Z, Kawasawa YI, Lefebvre V, Sestan N. SOX5 postmitotically regulates migration, postmigratory differentiation, and projections of subplate and deep-layer neocortical neurons. Proc Natl Acad Sci U S A. 2008;105:16021–16026. [PubMed: 18840685]
15. Smits P, Li P, Mandel J, et al. The transcription factors L-SOX5 and Sox6 are essential for cartilage formation. Dev Cell. 2001;1:277–290. [PubMed: 11702786]
16. Smits P, Dy P, Mitra S, Lefebvre V. SOX5 and Sox6 are needed to develop and maintain source, columnar, and hypertrophic chondrocytes in the cartilage growth plate. J Cell Biol. 2004;164:747–758. [PubMed: 14993235]
17. Lai T, Jabaudon D, Molyneaux BJ, et al. SOX5 controls the sequential generation of distinct corticofugal neuron subtypes. Neuron. 2008; 57:232–247. [PubMed: 18215621]
18. Dy P, Han Y, Lefebvre V. Generation of mice harboring a SOX5 conditional null allele. Genesis. 2008;46:294–299. [PubMed: 18543318]
19. Nesbitt A, Bhoj EJ, McDonald Gibson K, et al. Exome sequencing expands the mechanism of SOX5-associated intellectual disability: a case presentation with review of sox-related disorders. Am J Med Genet A. 2015;167A:2548–2554. [PubMed: 26111154]
20. Zech M, Poustka K, Boesch S, et al. SOX5-null heterozygous mutation in a family with adult-onset hyperkinesia and behavioral abnormalities. Case Rep Genet. 2017;2017:2721615. https://www-hindawi-com.gate2.inist.fr/journals/crig/2017/2721615/. [PubMed: 29214085]
21. Lelieveld SH, Reijnders MR, Pfundt R, et al. Meta-analysis of 2,104 trios provides support for 10 new genes for intellectual disability. Nat Neurosci. 2016;19:1194–1196. [PubMed: 27479843]
22. Bosch DG, Boonstra FN, de Leeuw N, et al. Novel genetic causes for cerebral visual impairment. Eur J Hum Genet. 2016;24:660–665. [PubMed: 26350515]
23. Deciphering Developmental Disorders S. Prevalence and architecture of de novo mutations in developmental disorders. Nature. 2017; 542:433–438. [PubMed: 28135719]
24. Cherot E, Keren B, Dubourg C, et al. Using medical exome sequencing to identify the causes of neurodevelopmental disorders: experience of 2 clinical units and 216 patients. Clin Genet. 2018;93:567–576. [PubMed: 28708303]
25. Tumiene B, Maver A, Writzl K, et al. Diagnostic exome sequencing of syndromic epilepsy patients in clinical practice. Clin Genet. 2018; 93:1057–1062. [PubMed: 29286531]
26. 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]
27. Firth HV, Richards SM, Bevan AP, et al. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. Am J Hum Genet. 2009;84:524–533. [PubMed: 19344873]
28. Li Q, Wang K. InterVar: clinical interpretation of genetic variants by the 2015 ACMG-AMP guidelines. Am J Hum Genet. 2017;100: 267–280. [PubMed: 28132688]
29. Richards S, Aziz N, Bale S, et al. 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]

30. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res. 2019;47(D1):D886–D894. [PubMed: 30371827]

31. Venselaar H, Te Beek TA, Kuipers RK, Hekkelman ML, Vriend G. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. BMC Bioinformatics. 2010;11:548. [PubMed: 21059217]

32. Bienert S, Waterhouse A, de Beer TA, et al. The SWISS-MODEL Repository —new features and functionality. Nucleic Acids Res. 2017;45(D1): D313–D319. [PubMed: 27899672]

33. Hoeppner LH, Phoenix KN, Clark KJ, et al. Revealing the role of phospholipase Cbeta3 in the regulation of VEGF-induced vascular permeability. Blood. 2012;120:2167–2173. [PubMed: 22674805]

34. 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]

35. Han Y, Lefebvre V. 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. 2008;28:4999–5013. [PubMed: 18559420]

36. Besenbacher S, Sulem P, Helgason A, et al. Multi-nucleotide de novo mutations in humans. PLoS Genet. 2016;12:e1006315.

37. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536:285–291. [PubMed: 27535533]

38. Lefebvre V, Li P, de Crombrugghe B. 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. 1998; 17:5718–5733. [PubMed: 9755172]

39. D'Gama AM, Walsh CA. Somatic mosaicism and neurodevelopmental disease. Nat Neurosci. 2018;21:1504–1514. [PubMed: 30349109]

40. Coe BP, Witherspoon K, Rosenfeld JA, et al. Refining analyses of copy number variation identifies specific genes associated with developmental delay. Nat Genet. 2014;46:1063–1071. [PubMed: 25217958]

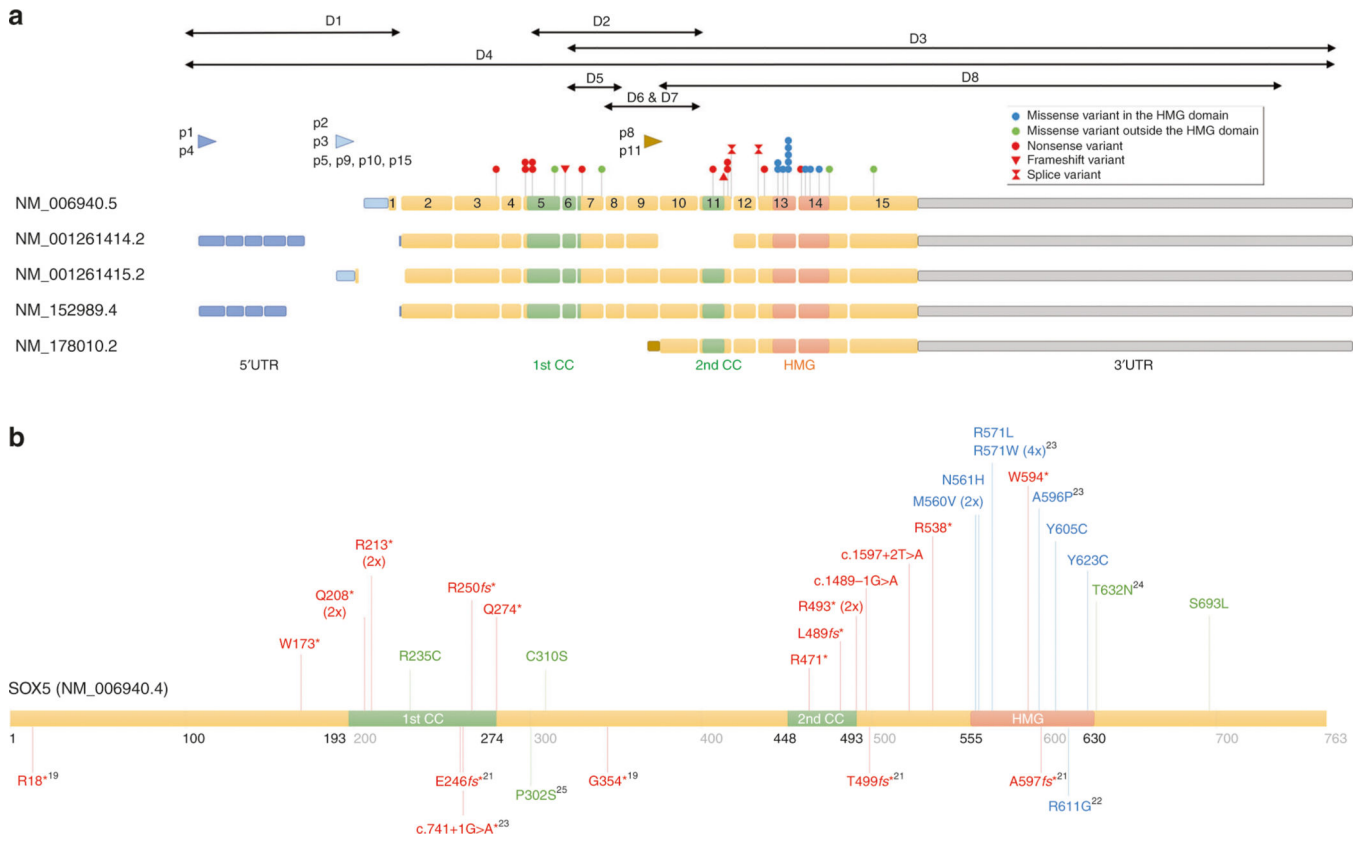

Fig. 1. *SOX5* variant spectrum associated with Lamb–Shaffer syndrome (LAMSHF).

(a) Location of genetic alterations identified in patients in this study. *SOX5* transcript isoforms are labeled with National Center for Biotechnology Information (NCBI) accession numbers. Boxes 1to 15, coding exons of isoform NM_006940. 5' and 3'UTR: 5' and 3' untranslated sequences. p1 to p11 represent *SOX5* promoters listed in the Fantom5 database; p1 and p2 (in bold) are the main promoters driving *SOX5* expression in brain. CC, coiled-coil domain. Double-arrowed lines, deletions in patients D1–D8. Point variants, labeled as indicated. **(b)** Location of point variants reported here (above) and previously (below) on the longest *SOX5* isoform. Protein and domain residue boundaries are indicated underneath the schematic. Red, nonsense and frameshift variants. Blue and green, missense variants within and outside the HMG domain, respectively. Superscripts, references.

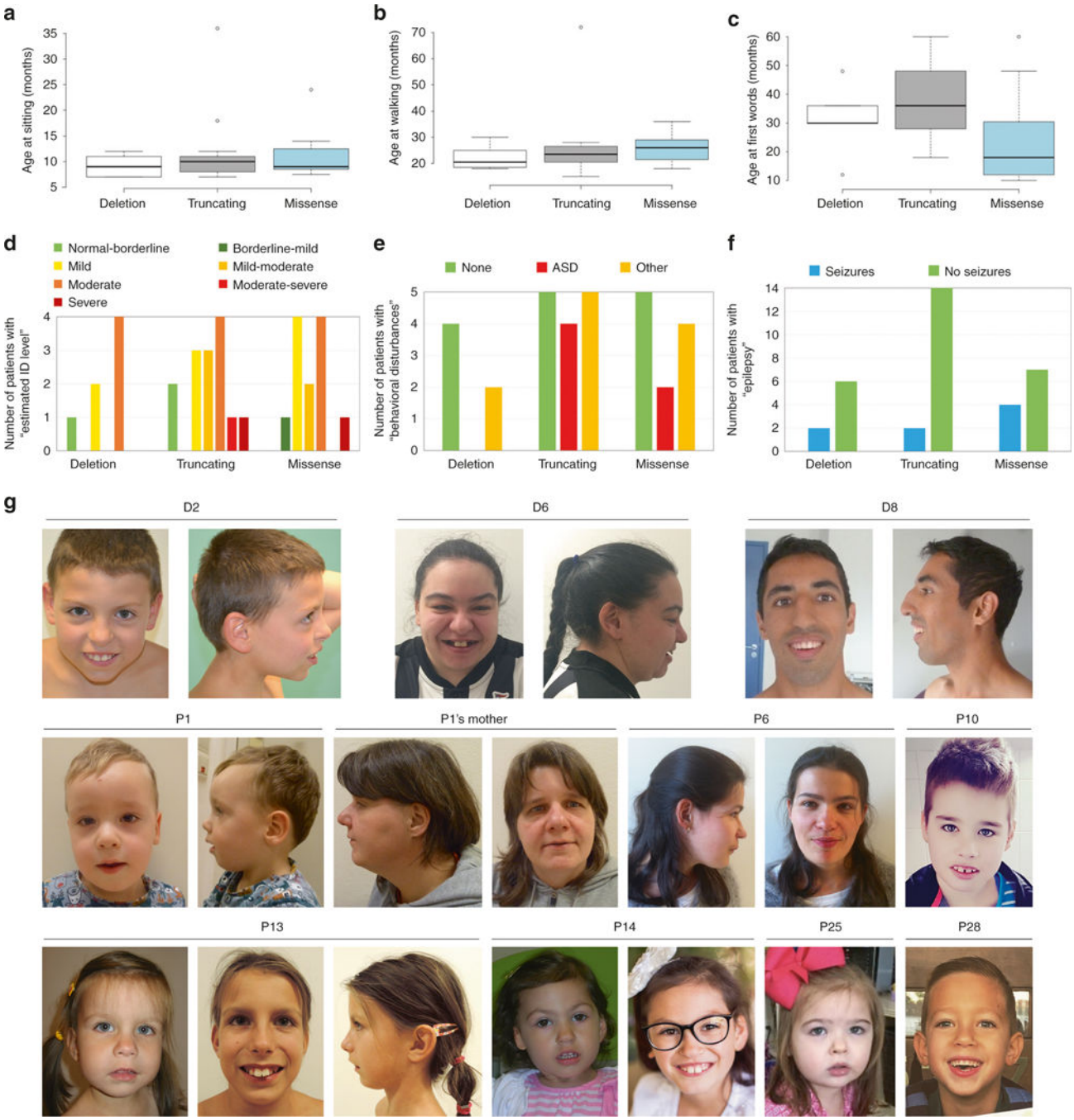

Fig. 2. Patients exhibit similar clinical features regardless of the *SOX5* alteration type.

Box plots showing comparative distribution of ages at **(a)** sitting unsupported, **(b)** walking unsupported, and **(c)** first words for patients with deletion, truncating, and missense variants. **(d)** Number of patients with normal to borderline cognitive abilities and various degrees of intellectual disability (ID). **(e)** Number of patients with autism spectrum disorder (ASD) or other behavioral disturbances. **(f)** Number of patients with seizures. **(g)** Facial profiles of individuals with de novo *SOX5* variants. Above: D2 at age 10 years; D6 at age 26 years; D8 at age 24 years. Center: P1 at age 2 years, and his mother (41 years old); P6 at age 19

years; P10 at age 8 years. Below: P13 at ages 2 years, 6 months, and 11 years, 4 months, respectively; P14 at ages 2 years, 4 months and 8 years, respectively; P25 at age 4 years; P28 at age 5 years. Common facial features include broad or full nasal tip, thin upper lip and/or full lower lips, small jaw or prominent chin, prominent upper incisors and epicanthus.

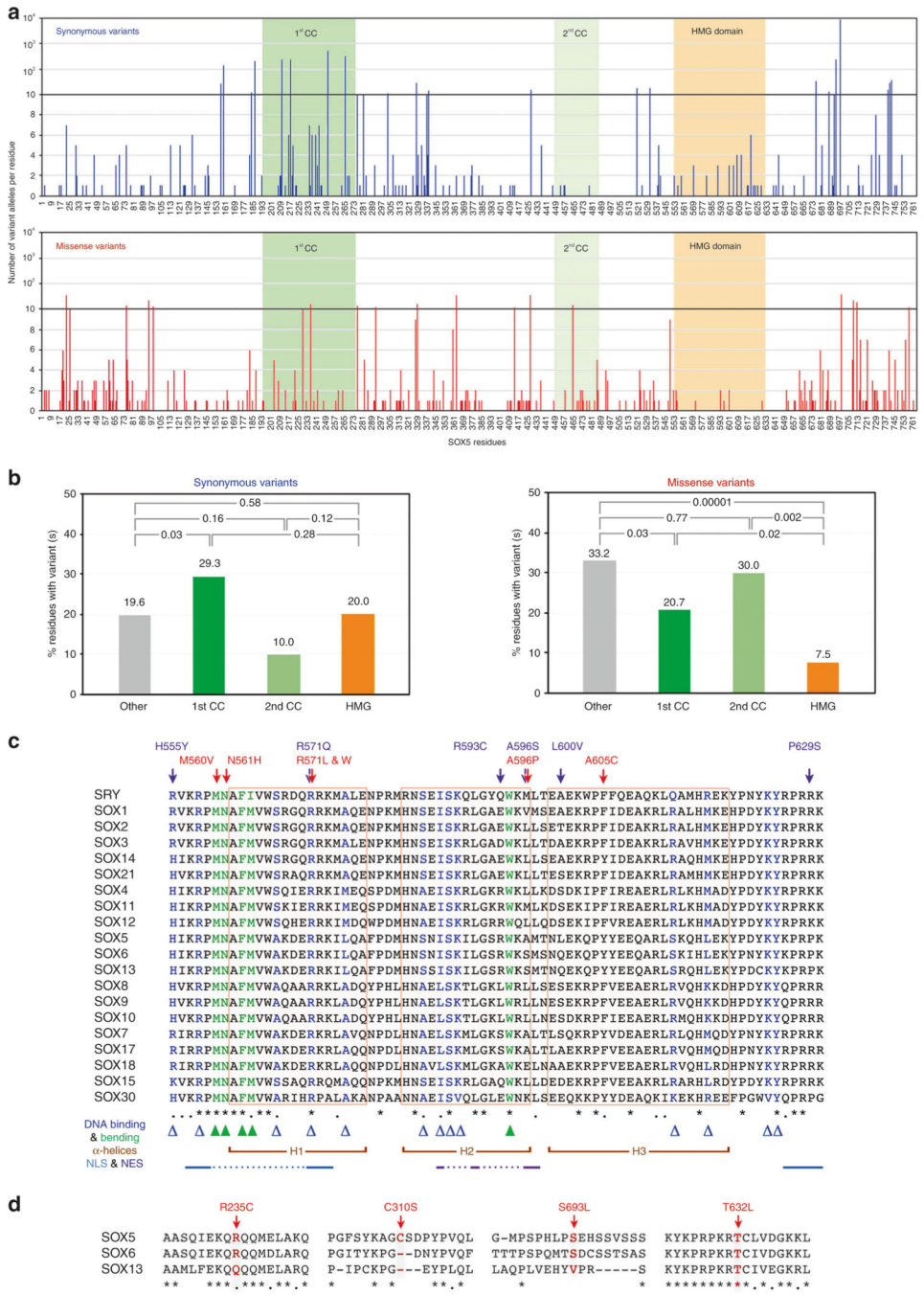

Fig. 3. Human *SOX5* is under tight conservation constraint.

(a) Distribution of synonymous and missense variants in *SOX5* in gnomAD individuals. CC, coiled coil. **(b)** Percentages of residues carrying at least one synonymous or missense variant in the functional and other domains of *SOX5* in gnomAD individuals. T-tests were performed to calculate the statistical significance of differences between protein domains. *P* values are indicated. **(c)** Alignment of all human SOX protein HMG domain sequences, with indication of residues altered in Lamb–Shaffer syndrome (LAMSHF) patients (red) and altered only in gnomAD individuals (purple). Asterisks, fully conserved residues.

Dots, semiconserved residues. Colored triangles, residues important for DNA binding and bending. Brackets, H1, H2, and H3 α-helices. Continued lines linked with dotted lines, key amino acids in nuclear localization signal sequences (NLS) and nuclear export signal sequence (NES). **(d)** Alignment of human SOXD protein sequences outside the HMG domain that encompass residues altered in LAMSHF patients.

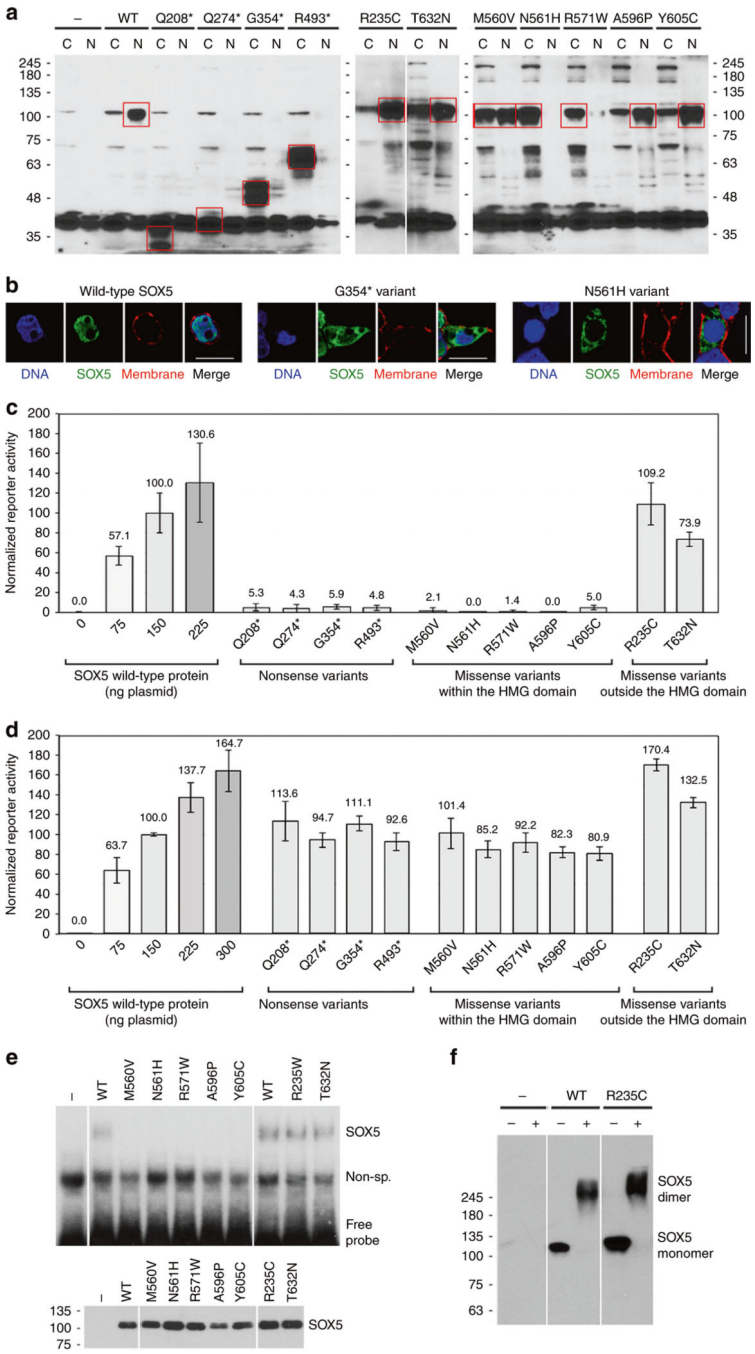

Fig. 4. Subcellular localization and activities of SOX5 variants.

(a) Western blots of cytoplasmic (C) and nuclear (N) extracts from HEK-293 cells transfected with plasmids encoding no protein (−), wild-type SOX5 (WT), or SOX5 variants. Blots were incubated with SOX5 antibody. Red boxes, SOX5specific protein signals. Numbers, Mr of protein standards. **(b)** Representative images of SOX5 immunostaining (green signal) in HEK-293 cells transfected with plasmids encoding wild-type SOX5 (WT) or the indicated variants. Nuclei are seen in blue and plasma membranes in red. Scale bars: 20 μm. **(c)** Test of the abilities of SOX5 variants to synergize with SOX9 in

transactivation. HEK-293 cells were transfected with *Acan* and pSV2βGal reporter plasmids and plasmids encoding no protein, SOX9, and/or SOX5. The WT SOX5 plasmid was used in the indicated amounts, and the variant plasmids at 150 ng. Reporter activities are presented as the mean ± standard deviation obtained for triplicates in one representative experiment. They were normalized for transfection efficiency and are reported as increase over the activity of SOX9 alone. **(d)** Test of the abilities of SOX5 variants to interfere with WT SOX5 in transactivation. HEK-293 cells were transfected essentially as described above. SOX5 variant plasmids were tested at 150 ng with 150 ng SOX5 WT plasmid. Reporter activities were calculated and are presented as described above. **(e)** Test of the abilities of SOX5 variants to bind DNA in electrophoretic mobility shift assay (EMSA). Extracts from HEK-293 cells transfected with empty, WT SOX5, or SOX5 variant plasmid were incubated with a 2HMG DNA probe. Top, X-ray film images. SOX5/DNA complexes migrated more slowly than nonspecific protein (non-sp.)/DNA complexes. Bottom, western blot showing similar amounts of all SOX5 proteins. **(f)** Dimerization assay with the same extracts as in (c) for no protein, WT SOX5, and the R235C variant. Western blots were performed using SOX5 antibody. SOX5 dimers ran in sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) with an apparent Mr twice as large as that of monomers.

Table 1

Summary of genetic and clinical data

ID	Family history	Age (years)	Sex	*SOX5* variant (NM_006940)	Variant type	Inheritance	ACMG class	Language delay	Language ability	ID (level)	Behavioral disturbances	Seizures	Other features
D1	Spo	6.5	M	Del ex1	Del	Unknown	5	Y	10 words	Y (mild)	NA	N	Hypotonia
D2	Spo	10	M	Del ex5–10	Del	DN	5	Y	NA	Y (moderate)	Intolerant to frustration, heteroaggressive	N	Strabismus, hypotonia, fused vertebrae
D3	Spo	14	F	Del ex6–15	Del	Mat mosaic	5	Y	Sentences	Y (mild)	NA	N	Strabismus
D4	Spo	5.5	F	Whole-gene del (+*BCAT1*)	Del	DN	5	Y	NA	Y (moderate)	NA	N	Hypotonia
D5	Spo	10	F	Del ex6–8	Del	DN	5	Y	Short sentences	Y (moderate)	NA	N	Strabismus
D6	Sib	26	F	Del ex8–10	Del	Mosaic	5	NA	NA	NA	N	Y	Fused vertebrae, subaortic ventricular septal defect, pulmonary stenosis
D7	Sib	36	M	Del ex8–10	Del	Mosaic	5	Y	Sentences	N (dyspraxia)	NA	Y	None
D8	Spo	24	M	Del ex10–15	Del	DN	5	Y	Sentences	Y (moderate)	Autistic features, blinking, phobia	N	Vesicoureteral reflux, strabismus, amblyopia, pyramidal syndrome, toe syndactyly
P1	Mat	2.5	M	c.518G>A, p.(Trp173*)	NS	Mat (DN in mother)	5	Y	20 words	Y (mild–moderate)	Stereotypies, intolerant to frustration	N	Strabismus, hypotonia
P2	Spo	7	F	c.622C>T, p.(Gln208*)	NS	DN	5	Y	Words	Y (mild)	NA	N	None
P3	Spo	3	F	c.622C>T, p.(Gln208*)	NS	DN	5	Y	Words	NA	NA	NA	Hypotonia, café au lait spots
P4	Sib	31	F	c.637C>T, p.(Arg213*)	NS	Mosaic	5	Y	Sentences	Y (mild)	Isolated, shy, anxious	N	Strabismus, optic atrophy, saccadic pursuit, tetrapyramidal syndrome, scoliosis
P5	Sib	31	M	c.637C>T, p.(Arg213*)	NS	Mosaic	5	Y	Sentences	Y (mild)	Aggressive, agitated	N	Saccadic pursuit, tetrapyramidal syndrome, scoliosis

ID	Family history	Age (years)	Sex	*SOX5* variant (NM_006940)	Variant type	Inheritance	ACMG class	Language delay	Language ability	ID (level)	Behavioral disturbances	Seizures	Other features
P6	Spo	17.8	F	c.637C>T, p.(Arg213*)	NS	DN	5	Y	Sentences	N (borderline IQ)	N	N	Strabismus, myopia, optic atrophy, hypotonia, gait instability, dysmetria
P7	Spo	2.5	F	c.747_748del, p.(Arg250Thr*fs**36)	FS	Unknown	5	Y	Nonverbal	NA	NA	N	Strabismus, hypotonia
P8	Spo	8	M	c.820C>T, p.(Gln274*)	NS	DN	5	Y	Short sentences	Y (mild–moderate)	ASD, intolerant to frustration	N	Strabismus, hypotonia, ear tubes, undescended testicles
P9	Spo	7	F		NS	Mat mosaic	5	Y	Associates words	Y (mild–moderate)	ADHD	N	Feeding difficulties, hypotonia
P10	Spo	8	M	c.1465dup, p.(Leu489Pro*fs**3)	FS	DN	5	Y	Words	NA	ASD	N	None
P11	Spo	3	F	c.1477C>T, p.(Arg493*)	NS	DN	5	Y	Words	Y (moderate)	NA	N	Hypotonia, cortical visual impairment, ear tubes, scoliosis, constipation, supernumerary nipple
P12	Spo	20	M	c.1477C>T, p.(Arg493*)	NS	Unknown	5	Y	Nonverbal	Y (severe)	NA	Y	Scoliosis
P13	Spo	12	F	c.1489–1G>A, p.?	Splice	DN	5	Y	Sentences	Y (moderate)	ASD, anxious	N	Hypotonia, constipation, thoracic kyphosis
P14	Spo	8	F	c.1597+2T>A, p.?	Splice	DN	5	Y	>100 words	Y (moderate)	NA	N (EEG anomalies)	Hypotonia, right microtia, microcephaly, myopia
P15	Spo	30	M	c.1613C>G, p.(Ser538*)	NS	DN	5	Y	Short sentences	Y (moderate)	Autistic features	Y	None
P16	Pat	13	M	c.1678A>G, p.(Met560Val)	MS	DN	5	Y	Sentences	Y (mild)	Temper tantrums	Y	None
P17	Spo	1.75	F	c.1678A>G, p.(Met560Val)	MS	DN	5	Y	Words	Y (mild)	NA	N	Hypotonia, joint hyperlaxity, hip dysplasia
P18	Sib	4	M	c.1681A>C, p.(Asn561His)	MS	DN	5	Y	Words	Y (severe)	Insomnia	N	Dysphagia with G-tube dependence, scoliosis, hypotonia, muscle weakness

ID	Family history	Age (years)	Sex	*SOX5* variant (NM_006940)	Variant type	Inheritance	ACMG class	Language delay	Language ability	ID (level)	Behavioral disturbances	Seizures	Other features
P19	Spo	27	F	c.1711C>T, p.(Arg571Trp)	MS	Unknown	5	Y	Sentences	Y (mild)	NA	N	Ovarian dystrophy
P20	Pat	17	M	c.1711C>T, p.(Arg571Trp)	MS	DN	5	Y	Sentences	Y (moderate)	Autistic features	N	None
P21	Mat	6	M	c.1711C>T, p.(Arg571Trp)	MS	DN	5	Y	Short sentences	Y (moderate)	Angry outbursts, stereotypies	Y	Joint hyperlaxity, strabismus, hypotonia
P22	Sib	3.3	F	c.1711C>T, p.(Arg571Trp)	MS	Mosaic	5	Y	Words	Y (mild–moderate)	NA	N	Hypotonia
P23	Sib	5.6	M	c.1711C>T, p.(Arg571Trp)	MS	Mosaic	5	Y	Short sentences	Y (mild–moderate)	ADHD	N	Hypotonia, microcephaly
P24	Spo	18	F	c.1712G>T, p.(Arg571Leu)	MS	DN	4	Y	Sentences	Y (moderate)	NA	Y	Hypotonia
P25	Spo	4	F	c.1782G>A, p.(Trp594*)	NS	DN	5	Y	Babbles	Y (moderate–severe)	ASD, sensory integration disorder	N	Bilateral optic nerve atrophy, hypotonia
P26	Spo	14	F	c.1786G>C, p. (Ala596Pro)	MS	DN	5	Y	Sentences	Y (moderate)	ASD, anxious	Y	Teeth anomalies, strabismus, hypotonia, muscle weakness
P27	Spo	6.5	F	c.1814A>G, p.(Tyr605Cys)	MS	DN	5	Y	Sentences	Y (borderline–mild)	ASD	N	Constipation, optic atrophy, joint hyperlaxity, hypotonia
P28	Spo	5.5	M	c.1819G>T, p.(Glu607*)	NS	DN	5	Y	Sentences	N (VIQ/PIQ discrepancy)	NA	N	Ear tubes, pallor of left optic nerve, strabismus, constipation, hypotonia
P29	Spo	4	M	c.1868A>G, p.(Tyr623Cys)	MS	DN	4	Y	Words	Y (mild)	NA	N	Gait ataxia
V1	Spo	18	M	c.703C>T, p.(Arg235Cys)	MS	DN	3	NA	NA	N	NA	N	Tourette syndrome
V2	Spo	0.7	M	c.928T>A, p.(Cys310Ser)	MS	DN	3	NA	NA	NA	NA	Y	Severe obstructive sleep apnea, severe gastroesophageal reflux, G-tube fed, hypotonia
V3	Spo	13	M	c.1895C>A, p.(Thr632Asn)	MS	DN	3	Y	Words	Y (moderate–severe)	ASD	N	None
V4	Spo	36	M	c.2078C>T, p.(Ser693Leu)	MS	DN	3	Y	Short sentences	Y (moderate)	NA	Y	None

IDs: D, pathogenic microdeletion; P, pathogenic or likely pathogenic variant; V, variant of unknown significance. American College of Medical Genetics and Genomics (ACMG) class: 5, pathogenic; 4, likely pathogenic; 3, variant of uncertain significance. See Supplementary Table 1 for more details.

ADHD attention deficit–hyperactivity disorder, *ASD* autism spectrum disorder, *Del* deletion, *DN* de novo, *EEG* electroencephalogram, *ex* exon(s), *F* female, *FS* frameshift, *G-tube* gastric tube, *M* male, *Mat* maternal, *MS* missense, *N* no, *NA* not available, *NS* nonsense, *Pat* paternal, *PIQ* performance *IQ*, *Sib* sibling (affected sib pairs), *Spo* sporadic, *VIQ* verbal *IQ*, *Y* yes.