The American Journal of Human Genetics, Volume *104*

Supplemental Data

De Novo SOX4 Variants Cause a Neurodevelopmental Disease Associated with Mild Dysmorphism

Ash Zawerton, Baojin Yao, J. Paige Yeager, Tommaso Pippucci, Abdul Haseeb, Joshua D. Smith, Lisa Wischmann, Susanne J. Kühl, John C.S. Dean, Daniela T. Pilz, Susan E. Holder, Deciphering Developmental Disorders Study, University of Washington Center for Mendelian Genomics, Alisdair McNeill, Claudio Graziano, and Véronique Lefebvre

Supplemental Note: Case Reports

Subject 1 was a 4-year-and-8-month-old male. Array comparative genomic hybridization (array-CGH) excluded pathogenic copy-number variants, and analysis for fragile-X syndrome was negative. Trio-based exome sequencing revealed no plausible pathogenic mutation beside a heterozygous *de novo SOX4* missense variant: hg19 chr6:21594963C>A, p.Phe66Leu. This subject was the son of healthy non-consanguineous Italian parents. He had two older healthy half-sisters (one paternal and one maternal). His mother was 33-year-old at conception and his father 41. Pregnancy was complicated by intrauterine growth retardation and he was born after 37 weeks of gestation by C-section. Birth weight was 1970 g (10^{th} percentile) and length was 43 cm (<3^{rd} percentile). On initial examination at 3 years and 2 months of age, his height was 85 cm (1^{st} percentile), his weight was 11.0 kg (1^{st} percentile) and his head circumference was 46.5 cm (3^{rd} percentile). At 4 years and 8 months of age, his height was 100.7 cm (10^{th} percentile), his weight was 15.0 kg (8^{th} percentile), and his head circumference was 47.6 cm (1^{st} percentile), indicating progressive microcephaly. He had global developmental delay: independent walking was only possible at 27 months and he spoke his first words at 4 years. He had seizures that started at 3.5 years of age. Treatment with valproic acid was started, but the response was unsatisfactory, with the child maintaining brief, daily crises. He was hypotonic and had a peri-membranous ventricular septal defect. He had feeding difficulties, was unable to eat solid food, and had dysphagia and severe constipation. Eye examination revealed strabismus and keratoconus. Brain MRI demonstrated a delay in myelination. His IQ score was 68 (mild ID). He had slight facial dysmorphism: microbrachycephaly, epicanthus of the left eye, a stellate iris pattern, a short nose with anteverted nares, a wide mouth with a cupid bow, and posteriorly rotated ears. He also had bilateral clinodactyly of the 5^{th} finger, but hands and feet were otherwise normal. Submission of *SOX4* to GeneMatcher[1] facilitated the present study by identifying Subjects 2 to 4 in the DDD database of *de novo* variants in 4296 children with developmental disorders.[2]

Subject 2 was an 11-year-and-10-month-old male of Scottish-Hungarian ancestry. He was found by trio-based exome sequencing to carry a heterozygous *de novo SOX4* missense variant: hg19 chr6:21595099G>C, p.Ala112Pro. No other plausible pathogenic variant was found. CGH revealed a duplication of Yp11.31-11.222 not present in his father and likely unrelated to his phenotype. This child was the fourth of five children of healthy non-consanguineous parents. His father was 66-year-old at conception and his mother 34. Pregnancy was complicated by a bleed in the first trimester which was thought to be associated with the loss of a twin sibling. Born at 40 weeks of gestation, the baby weighed 3.5 kg (46^{th} percentile), had bilateral, vertical talus, which was surgically corrected in infancy, and also had laryngomalacia. There were initial concerns about his weight gain and he had problems weaning. He otherwise reached his developmental milestones, but lost his ability to roll and bear weight almost overnight when he was 10-month-old. His parents associated this with a homeopathic treatment that he was receiving at the time. He developed a spastic quadriparesis, most prominent in the lower limbs. Brain MRI at age 15 months and 6 years demonstrated progressive cerebellar atrophy. There were also some patchy cerebral white matter changes. At 6 years and 8 months of age, his weight was 14.8 kg (1^{st} percentile) and his head circumference was 48.3 cm ($<0.4^{th}$ percentile). His OFC has remained at this size ever since. When he was last seen, at 11 years and 10 months of age, his neurodevelopmental disorder had been static for many years. He was still a slow eater, had problems chewing and requires mashed food. He had never walked, had no speech and had severe intellectual disability. However, he was very alert and aware of his surroundings. He demonstrated bruxism and drooling. Secondary dentition was markedly delayed as he still had several primary teeth. In addition to microcephaly, he had trigonocephaly with a mild metopic ridge, bilateral epicanthic folds, infra-orbital folds, a wide mouth and a defined cupid's bow of the upper lip. He also had mild 5^{th} finger clinodactyly.

Subject 3 was a 6-year-old female. Trio-based exome sequencing indicated a heterozygous *de novo SOX4* missense variant as sole pathogenic candidate variant: hg19 chr6:21594941 T>G, p.Ile59Ser. This child was small for her age (6th percentile for height, 9th for weight, and 12th for OFC). Her main problem at 3 years was delayed speech, with a limited vocabulary of single words. She responded well to speech therapy, and her speech was normal at 6 years. She had very mild learning disabilities and was not receiving extra help at mainstream school. The school were not concerned, but her mother reported that she was struggling to keep up with peers. She had characteristic facial features, with deep-set eyes, infra-orbital grooves, flat nasal bridge, upturned nares, wide cupid-bow mouth and full lips. She also had mild 5th finger clinodactyly and dysplastic 5th toenails.

Subject 4 was an 8-year-and-10-month-old female. Trio-exome sequencing showed a heterozygous *de novo* SOX4 missense variant: hg19 chr6:21595080 G>T, p.Lys105Asn. No other pathogenic variant candidate was found. She had a normal delivery after 39 weeks of gestation, but had a low birth weight (2.35 kg, 1st percentile). When initially assessed at 6 years and 10 months of age, she was small for her age, but had a normal head circumference (3rd percentile for height and weight, and 45th percentile for OFC). She did not walk until 21 months of age and pronounced her first words at 24 months. At 8 years and 10 months of age, her full-scale IQ (WISC-IV UK) was 52. She was in mainstream school and receiving educational support. She was sociable and had satisfactory speech, despite earlier expressive difficulties. She was yet unable to read or write, and had no sense of danger. She had a characteristic facial appearance, with deep-set eyes, infra-orbital creases, malar flattening, upturned nares, wide cupid-bow mouth and full lips. The ears were low set, and posteriorly rotated. There was mild 5th finger clinodactyly and mild camptodactyly.

Supplemental Figures

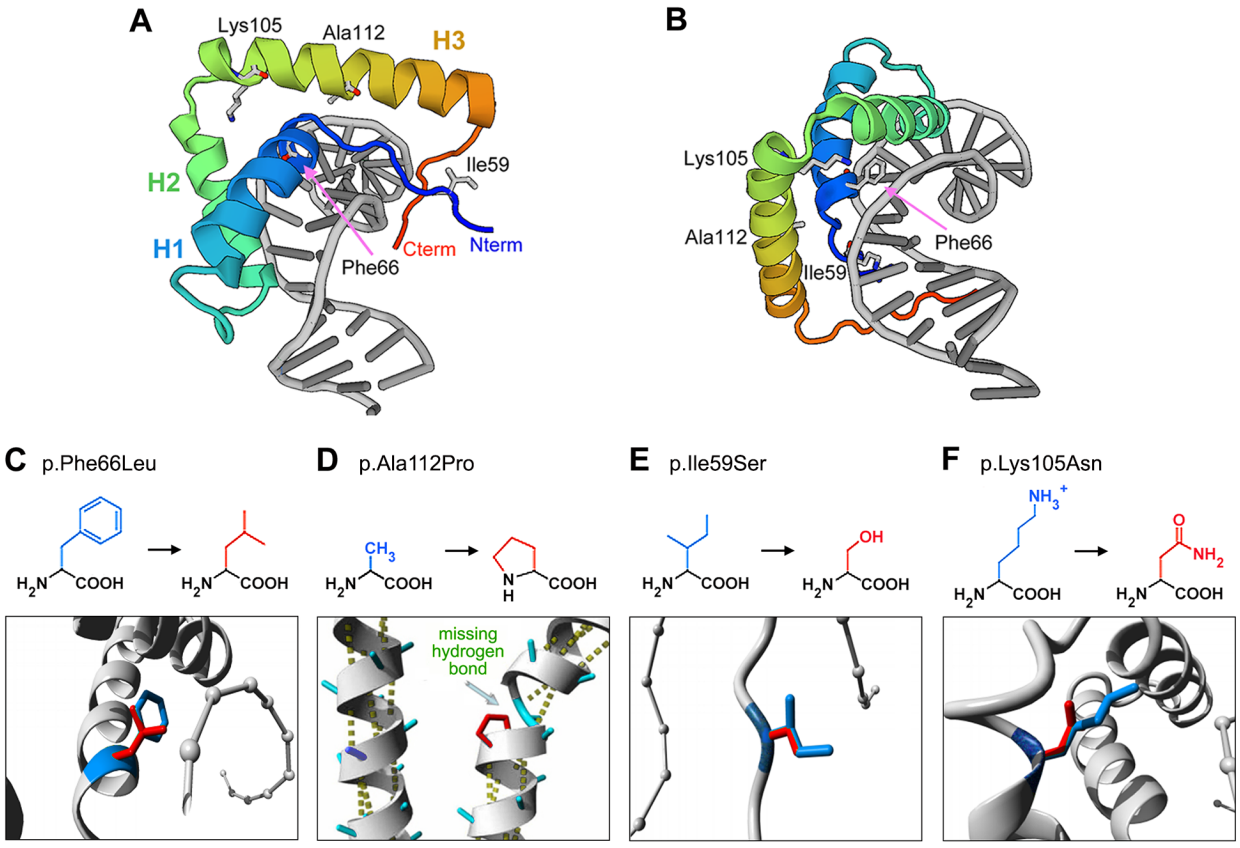
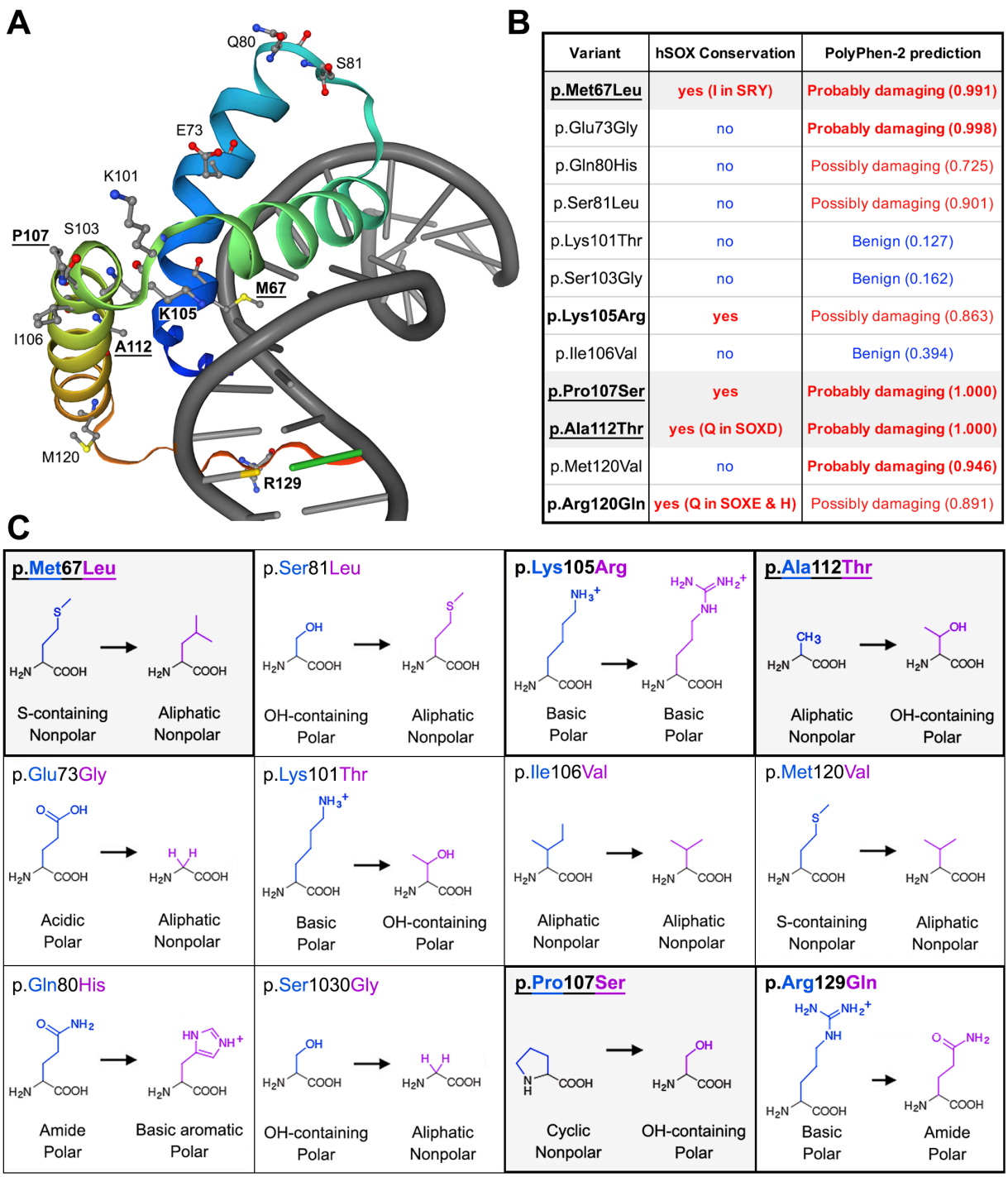


Figure S1. ***In Silico*** **Prediction of Changes in the HMG Domain Structure Induced by the Four Subjects' SOX4 Missense Variants.**

(A) Rendering of the wild-type SOX4 HMG domain–DNA complex. DNA is shown in grey and the SOX4 HMG domain is rainbow-colored. Its α-helices are labeled H1, H2, and H3, and its N- and C-termini are indicated. The residues mutated in our subjects are indicated and their side chains are depicted. The HMG domain wraps around the DNA helix and induces a strong bend. This cartoon was generated by SWISS-MODEL using published data.[3]

(B) Same protein-DNA model as in panel A, but oriented differently to show the intercalation of the Phe66 side chain into the minor groove of the DNA helix.

(C to F) Top, schematics of wild-type and variant amino acids, with side chains colored in blue and red, respectively. Bottom, ribbon presentation of the SOX4 protein structure generated by HOPE to show structural changes caused by the variants. The protein is presented in grey and the side chains of the wild-type and mutant amino acids are colored in blue and red, respectively.



Variant	hSOX Conservation	PolyPhen-2 prediction
p.Met67Leu	yes (I in SRY)	Probably damaging (0.991)
p.Glu73Gly	no	Probably damaging (0.998)
p.Gln80His	no	Possibly damaging (0.725)
p.Ser81Leu	no	Possibly damaging (0.901)
p.Lys101Thr	no	Benign (0.127)
p.Ser103Gly	no	Benign (0.162)
p.Lys105Arg	yes	Possibly damaging (0.863)
p.Ile106Val	no	Benign (0.394)
p.Pro107Ser	yes	Probably damaging (1.000)
p.Ala112Thr	yes (Q in SOXD)	Probably damaging (1.000)
p.Met120Val	no	Probably damaging (0.946)
p.Arg120Gln	yes (Q in SOXE & H)	Possibly damaging (0.891)

Figure S2. *In Silico* Prediction of Changes in the HMG Domain Structure Induced by SOX4 Missense Variants Identified in gnomAD.

(A) Rendering of the wild-type SOX4 HMG domain–DNA complex. DNA is shown in grey and the SOX4 HMG domain is rainbow-colored. Its α-helices are labeled H1, H2, and H3, and its N- and

C-termini are indicated. The residues showing variants in gnomAD are indicated and their side chains are depicted. Bold characters are used for five residues that are highly conserved in human SOX proteins. Three of these residues are also underlined because their mutation is predicted by PolyPhen-2 to be probably damaging. This cartoon was generated by SWISS-MODEL using published data.[3]

(B) Table summarizing for each SOX residue that features a variant in gnomAD (left column) the degree of conservation in human SOX proteins (middle column) and the prediction score attributed by PolyPhen-2 for the consequence of the variant (right column). In the middle column, “yes” means conservation in all 20 SOX proteins; “no” means poor conservation; parentheses indicate residues found in other SOX proteins. Bold characters are used for five residues that are highly conserved in human SOX proteins. Three of these residues are also underlined and their boxes shaded in light grey because their mutation is predicted by PolyPhen-2 to be probably damaging.

(C) Schematics of wild-type and gnomAD variant amino acids, with side chains colored in blue and red, respectively. The features of the side chains are indicated. Bold characters are used for five residues that are highly conserved in human SOX proteins. Three of these residues are also underlined and their boxes shaded in light grey because their mutation is predicted by PolyPhen-2 to be probably damaging.

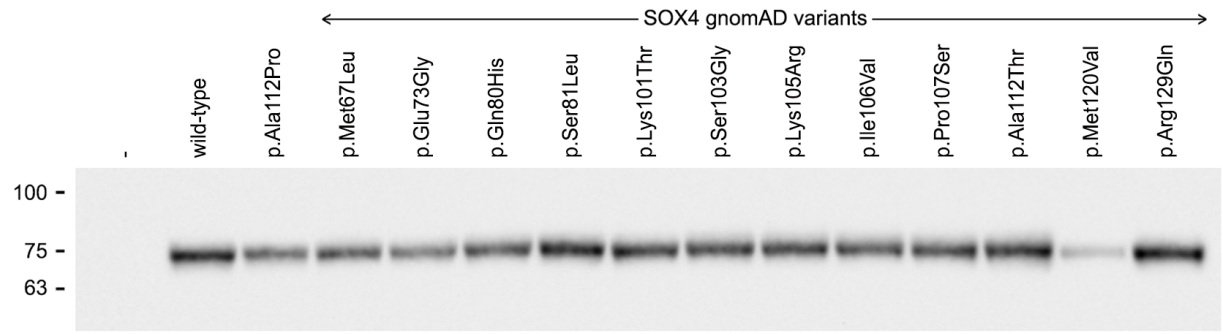


Figure S3. Assessment of SOX4 gnomAD Variant Levels Achieved in COS-1 Cells.

Western blot of lysates from COS-1 cells transiently transfected with expression plasmids for no protein (-), wild-type SOX4 (WT), the case-2 variant (p.Ala112Pro), and the twelve SOX4 gnomAD variants, as indicated. The blot was hybridized with an antibody recognizing the FLAG epitope added at the N-terminus of SOX4. The Mr of protein standards is indicated on the left (in k units). All SOX4 proteins migrate with an expected Mr of 70 k. This experiment complements the reporter assay shown in Figure 6D.

Supplemental Tables

Table S1. NCBI Accession Numbers for the HMG-Domain Sequences of Vertebrate SOX4 Orthologs Aligned in Figure 5B.

Vertebrate Species	Reference Sequence
Human *(Homo sapiens)*	NM_003107.2
Mouse *(Mus musculus)*	NM_009238.3
Cow *(Bos taurus)*	NM_001078128.1
Horse *(Equus caballus)*	XM_023624515.1
Bat *(Eptesicus fuscus)*	XM_008146131.1
Hedgehog *(Erinaceus europaeus)*	XM_007525152.2
Elephant *(Loxodonta africana)*	XM_010596185.2
Opossum *(Monodelphis domestica)*	XM_007487879.2
Platypus *(Ornithorhynchus anatinus)*	XM_007659031.1
Chicken *(Gallus gallus)*	AY249864.1
Frog *(Xenopus laevis)*	NM_001172201.1
Zebrafish *(Danio rerio)*	AY369082.1

Table S2. NCBI Accession Numbers for the HMG-Domain Sequences of All Human SOX Proteins Aligned in Figure 5C.

SOX protein	Reference Sequence
SOX1	NP_005977.2
SOX2	NP_003097.1
SOX3	NP_005625.2
SOX4	AAH72668.1
SOX5	AAH60773.1
SOX6	AAK26115.1
SOX7	CAC84226.1
SOX8	NP_055402.2
SOX9	CAA86598.1
SOX10	CAG38808.1
SOX11	BAA88122.1
SOX12	AAH67361.1
SOX13	AAD50120.1
SOX14	AAC95380.1
SOX15	AAH72003.1
SOX17	BAB83867.1
SOX18	BAA94874.1
SOX21	AAC95381.1
SOX30	BAA37146.1
SRY	CAA37790.1

Table S3. DNA Oligonucleotides Used to Generate SOX4 Variant Expression Plasmids. Wild-type and variant codons are shown in blue and red letters, respectively. Wild-type sequences are shown as references.

Subject Variants	Forward Primers	Reverse Primers
Wild-type	TGAACGCC**TTT**ATGGTGTGGTCGCAGA	ACCACACCAT**AAA**GGCGTTCATGGGCC
p.Phe66Leu	TGAACGCC**TTA**ATGGTGTGGTCGCAGA	ACCACACCAT**TAA**GGCGTTCATGGGCC
Wild-type	CATCCAGGAG**GCG**GAGCGGCTGCGCCT	AGGCGCAGCCGCTC**CGC**CTCCTGGATG
p.Ala112Pro	CATCCAGGAG**CCG**GAGCGGCTGCGCCT	AGGCGCAGCCGCTC**CGG**CTCCTGGATG
Wild-type	AGTGGCCAC**ATC**AAGCGGCCCATGAAC	GTTCATGGGCCGCTT**GAT**GTGGCCACT
p.Ile59Ser	AGTGGCCAC**AGC**AAGCGGCCCATGAAC	GTTCATGGGCCGCTT**GCT**GTGGCCACT
Wild-type	ACAGCGAC**AAG**ATTCCGTTCATCCAGG	CCTGGATGAACGGAAT**CTT**GTCGCTGT
p.Lys105Asn	ACAGCGAC**AAT**ATTCCGTTCATCCAGG	CCTGGATGAACGGAAT**ATT**GTCGCTGT
gnomAD Variants	**Forward Primers**	**Reverse Primers**
Wild-type	ACGCCTTT**ATG**GTGTGGTCGCAGATCG	CCACAC**CAT**AAAGGCGTTCATGGGCCG
p.Met67Leu	ACGCCTTT**CTG**GTGTGGTCGCAGATCG	CCACAC**CAG**AAAGGCGTTCATGGGCCG
Wild-type	GCAGATC**GAG**CGGCGCAAGATCATGGA	TGCGCCG**CTC**GATCTGCGACCACACCA
p.Glu73Gly	GCAGATC**GGG**CGGCGCAAGATCATGGA	TGCGCCG**CCC**GATCTGCGACCACACCA
Wild-type	ATGGAG**CAG**TCGCCCGACATGCACAAC	TCGGGCGA**CTG**CTCCATGATCTTGCGC
p.Gln80His	ATGGAG**CAC**TCGCCCGACATGCACAAC	TCGGGCGA**GTG**CTCCATGATCTTGCGC
Wild-type	GGAGCAG**TCG**CCCGACATGCACAACGC	TGTCGGG**CGA**CTGCTCCATGATCTTGC
p.Ser81Leu	GGAGCAG**TTG**CCCGACATGCACAACGC	TGTCGGG**CAA**CTGCTCCATGATCTTGC
Wild-type	GCTGCTC**AAG**GACAGCGACAAGATTCC	CGCTGTC**CTT**GAGCAGCTTCCAGCGTT
p.Lys101Thr	GCTGCTC**ACG**GACAGCGACAAGATTCC	CGCTGTC**CGT**GAGCAGCTTCCAGCGTT
Wild-type	TCAAGGAC**AGC**GACAAGATTCCGTTCA	CTTGTC**GCT**GTCCTTGAGCAGCTTCCA
p.Ser103Gly	TCAAGGAC**GGC**GACAAGATTCCGTTCA	CTTGTC**GCC**GTCCTTGAGCAGCTTCCA
Wild-type	CAGCGAC**AAG**ATTCCGTTCATCCAGGA	ACGGAAT**CTT**GTCGCTGTCCTTGAGCA
p.Lys105Arg	CAGCGAC**AGG**ATTCCGTTCATCCAGGA	ACGGAAT**CCT**GTCGCTGTCCTTGAGCA
Wild-type	GCGACAAG**ATT**CCGTTCATCCAGGAGG	GAACGG**AAT**CTTGTCGCTGTCCTTGAG
p.Ile106Val	GCGACAAG**GTT**CCGTTCATCCAGGAGG	GAACGG**AAC**CTTGTCGCTGTCCTTGAG
Wild-type	ACAAGATT**CCG**TTCATCCAGGAGGCGG	GATGAA**CGG**AATCTTGTCGCTGTCCTT
p.Pro107Ser	ACAAGATT**TCG**TTCATCCAGGAGGCGG	GATGAA**CGA**AATCTTGTCGCTGTCCTT
Wild-type	TCCAGGAG**GCG**GAGCGGCTGCGCCTCA	CCGCTC**CGC**CTCCTGGATGAACGGAAT
p.Ala112Thr	TCCAGGAG**ACG**GAGCGGCTGCGCCTCA	CCGCTC**CGT**CTCCTGGATGAACGGAAT
Wild-type	TCAAGCAC**ATG**GCTGACTACCCTGACT	TCAAG**CAT**GTGGCTGACTACCCTGACT
p.Met120Val	TCAAGCAC**GTG**GCTGACTACCCTGACT	TCAAG**CAC**GTGGCTGACTACCCTGACT
Wild-type	CAAGTAC**CGG**CCGCGAAAGAAGGTGAA	TTCGCGG**CCG**GTACTTGTAGTCAGGGT
p.Arg129Gln	CAAGTAC**CAG**CCGCGAAAGAAGGTGAA	TTCGCGG**CTG**GTACTTGTAGTCAGGGT

Table S4. Missense Variants in SRY and Other SOX Protein Residues Matching SOX4 HMG-domain Variants Identified in gnomAD. Mutations of interest, i.e., resulting in the same variants as in SOX4 gnomAD individuals, are written in blue if they were identified in gnomAD and in red if they were identified in disease. Note that the last SOX4 variant matches wild-type SOX10 and is therefore also written in blue.

Protein	Variant	Phenotype	Reference
SOX4	**p.Met67Leu**	**Unknown**	**gnomAD**
SRY	p.Ile68Thr	Gonadal dysgenesis	[4]
SOX9	p.Met113Thr	Campomelic dysplasia	[5]
SOX9	p.Met113Val	Acampomelic Dyslasia	[6]
SOX10	p.Met112Ile	Waardenburg syndrome	(Pingault, 2014}
SOX4	**p.Glu73Gly**	**Unknown**	**gnomAD**
SRY	p.Gln74His	XY sex reversal	[4]
SOX9	p.Ala119Val	Campomelic dysplasia	[7]
SOX11	p.Glu63Gln	Unknown	gnomAD
SOX4	**p.Gln80His**	**Unknown**	**gnomAD**
SOX9	p.Gln126His	Unknown	gnomAD
SOX4	**p.Ser81Leu**	**Unknown**	**gnomAD**
SOX10	p.Tyr126His	Unknown	gnomAD
SOX4	**p.Lys101Thr**	**Unknown**	**gnomAD**
SOX9	p.Asn147Thr	Unknown	gnomAD
SOX4	**p.Ser103Gly**	**Unknown**	**gnomAD**
SOX4	**p.Lys105Arg**	**Unknown**	**gnomAD**
SRY	p.Lys106Ile	Gonadal dysgenesis	[4]
SOX4	**p.Ile106Val**	**Unknown**	**gnomAD**
SOX9	p.Arg152Pro	Campomelic dysplasia	[7]
SOX11	p.Ile96Thr	Unknown	gnomAD

SOX4	**p.Pro107Ser**	**Unknown**	**gnomAD**
SRY	p.Pro108His	46, XY gonadal dysgenesis	[4]
SOX4	**p.Ala112Thr**	**Unknown**	**gnomAD**
SRY	p.Ala113Thr	Gonadal dysgenesis	[4]
SOX9	p.Ala158Val	Campomelic dysplasia	[7]
SOX9	p.Ala158Thr	Campomelic dysplasia, XY sex reversal	[7]
SOX10	p.Ala157Val	Waardenburg syndrome type IV	[8]
SOX11	p.Ala102Val	Coffin-Siris syndrome	[9]
SOX4	**p.Met120Val**	**Unknown**	**gnomAD**
SOX10	p.Lys165Arg	Unknown	gnomAD
SOX4	**p.Arg129Gln**	**Unknown**	**gnomAD**
SOX10	p.Gln174Pro	Waardenburg syndrome	[10]

Supplemental References

1. Sobreira, N., Schiettecatte, F., Valle, D., and Hamosh, A. (2015). GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. Hum Mutat 36, 928-930.
2. Wright, C.F., Fitzgerald, T.W., Jones, W.D., Clayton, S., McRae, J.F., van Kogelenberg, M., King, D.A., Ambridge, K., Barrett, D.M., Bayzetinova, T., et al. (2015). Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. Lancet 385, 1305-1314.
3. Jauch, R., Ng, C.K., Narasimhan, K., and Kolatkar, P.R. (2012). The crystal structure of the Sox4 HMG domain-DNA complex suggests a mechanism for positional interdependence in DNA recognition. Biochem J 443, 39-47.
4. Wang, X., Xue, M., Zhao, M., He, F., Li, C., and Li, X. (2018). Identification of a novel mutation (Ala66Thr) of SRY gene causes XY pure gonadal dysgenesis by affecting DNA binding activity and nuclear import. Gene 651, 143-151.
5. Wada, Y., Nishimura, G., Nagai, T., Sawai, H., Yoshikata, M., Miyagawa, S., Hanita, T., Sato, S., Hasegawa, T., Ishikawa, S., et al. (2009). Mutation analysis of SOX9 and single copy number variant analysis of the upstream region in eight patients with campomelic dysplasia and acampomelic campomelic dysplasia. Am J Med Genet A 149A, 2882-2885.
6. Staffler, A., Hammel, M., Wahlbuhl, M., Bidlingmaier, C., Flemmer, A.W., Pagel, P., Nicolai, T., Wegner, M., and Holzinger, A. (2010). Heterozygous SOX9 mutations allowing for residual DNA-binding and transcriptional activation lead to the acampomelic variant of campomelic dysplasia. Hum Mutat 31, E1436-1444.
7. Harley, V.R., Clarkson, M.J., and Argentaro, A. (2003). The molecular action and regulation of the testis-determining factors, SRY (sex-determining region on the Y chromosome) and SOX9 [SRY-related high-mobility group (HMG) box 9]. Endocr Rev 24, 466-487.
8. Morin, M., Vinuela, A., Rivera, T., Villamar, M., Moreno-Pelayo, M.A., Moreno, F., and del Castillo, I. (2008). A de novo missense mutation in the gene encoding the SOX10 transcription factor in a Spanish sporadic case of Waardenburg syndrome type IV. Am J Med Genet A 146A, 1032-1037.
9. Okamoto, N., Ehara, E., Tsurusaki, Y., Miyake, N., and Matsumoto, N. (2018). Coffin-Siris syndrome and cardiac anomaly with a novel SOX11 mutation. Congenit Anom (Kyoto) 58, 105-107.

10. Chaoui, A., Watanabe, Y., Touraine, R., Baral, V., Goossens, M., Pingault, V., and Bondurand, N. (2011). Identification and functional analysis of SOX10 missense mutations in different subtypes of Waardenburg syndrome. Hum Mutat 32, 1436-1449.