Constitutional *SOX4* Variation in Human Disorders

Subjects: Genetics & Heredity

Contributor: Mina Grippa, Claudio Graziano

SOX proteins are transcription factors which play a role in regulating the development of progenitor cells and tissue differentiation. Twenty members are known, clustered in eight groups named A through H and sharing a common DNA-binding domain called the HMG (high-mobility-group) box. Eleven of the SOX genes have been associated with genetic disorders so far, covering a broad spectrum of developmental diseases. *SOX4* is a single-exon gene and belongs to the SOXC group, together with *SOX11* and *SOX12*. *SOX4* variants have been recently described to cause a highly penetrant but heterogeneous disorder, with a phenotypic spectrum ranging from mild developmental delays and learning difficulties to intellectual disabilities with congenital anomalies. Nineteen pathogenic variants have been reported to date, generally de novo, heterozygous, and inactivating, either stop–gain or missense, the latter ones primarily targeting the HMG domain.

Keywords: SOX4 ; SOXopathy ; intellectual disability ; developmental delay

1. Introduction

Members of the SOX family of transcription factors are defined as based on the presence of a DNA-binding domain with homology to the high-mobility-group (HMG) box of *SRY* (sex-determining region Y) [1]. It is well recognized that the SOX family plays pivotal roles in many developmental and pathological processes, including male differentiation [2][3], eye development [4], skeletogenesis [5] and neurogenesis [6][7]. Twenty members are known, clustered in eight groups (SOXA to SOXH), and half of them are associated with human genetic disorders, termed “SOXopathies” [8].

SOX4, *SOX11*, and *SOX12* belong to the SOXC group. They are expressed in many progenitor cell types and have redundant roles in controlling cell survival and fate determination in response to various signalling pathways [9]. SOXC proteins have almost identical DNA-binding domains and show a high degree of conservation in their other known functional region, a transactivation domain located at the C terminus. Animal models first showed that *SOX4* and *SOX11* are essential developmental genes, since both *sox11*-null and *sox4*-null mice die in utero or at birth with multiple abnormalities [10][11]. *sox12*-null mice do not show any apparent phenotype, thanks to functional compensation by *sox4* and *sox11* [12]. Two de novo heterozygous missense variants in the *SOX11* HMG box have been linked to a human disorder characterized by intellectual disability (ID), growth deficiency, facial dysmorphism, and hypoplasia of the fifth digit [13]. The disease was classified as a “Coffin–Siris syndrome-like syndrome”. More de novo heterozygous mutations were later reported in patients with similar features, including *SOX11*-containing 2p25 deletions, a nonsense variant, and additional HMG-domain missense variants [14], but deeper phenotyping and analysis of DNA-methylation profiles proved that this condition is distinct from Coffin–Siris syndrome (CSS) [15].

2. SOX4 Structure and Function

SOX4 is a single-exon gene, and the open reading frame encodes a 474-amino acid protein, which includes two functionally important domains: namely, an HMG box and a C-terminal transactivation domain (TAD). The HMG box facilitates DNA binding, bending, and nuclear trafficking, whereas TAD mediates the interaction with different cofactors [16], although knowledge of critical residues within TAD is still missing. By binding to the minor groove of DNA, SOX4 alters chromatin architecture, leading to changes in transcriptional activities of downstream genes. SOX4 (and SOX11) has pleiotropic functions, which are likely to be mediated by distinct regulatory elements and downstream target genes involved in multiple developmental processes, including neurogenesis [17][18], heart development [10] and skeletal patterning [19], but also lymphocyte maturation [20] and more recently development of the inner ear [21]. The specific target genes of SOX4 have not yet been fully defined, but, among the genes that have been shown to be regulated by SOX4, some are important for neurodevelopment and disease (e.g., *RELN*, *DCX*, and *WDR45*) [18][22]. In the human brain, *SOX4* expression was found to be high in several regions (the dorsolateral prefrontal cortex, striatum, and cerebellar cortex)

during the first two trimesters of embryonic gestation, and then to decrease progressively to reach a very low level by the 4th decade of postnatal life; the expression was higher in areas of active neurogenesis [23].

This widespread involvement in developmental processes is consistent with the heterogeneous set of anomalies which can be observed in individuals carrying *SOX4* pathogenic variants.

3. SOX4 Single Nucleotide Variants

3.1. Heterozygous Pathogenic Variants

In line with other SOXopathies, most *SOX4* causative variants reported to date are dominant, either loss of function or missense variants which target the HMG box [23][24][25]. Globally, 13 pathogenic missense and six stop–gain (one frameshift, five nonsense) variants have been described (**Figure 1**). These variants were all identified through exome or genome sequencing studies and were defined as pathogenic/likely pathogenic based on (a) in silico assessments (modification of highly conserved amino acids in the HMG box, in different vertebrates orthologos, and in other human SOX proteins; generation of premature stop codons; absence in gnomAD database (https://gnomad.broadinstitute.org/ accessed on 30 December 2023); predictions on the effects of missense variants on protein structure and function) and (b) functional assessments (weak or absent DNA binding and loss of transactivation activity in reporter gene assays).

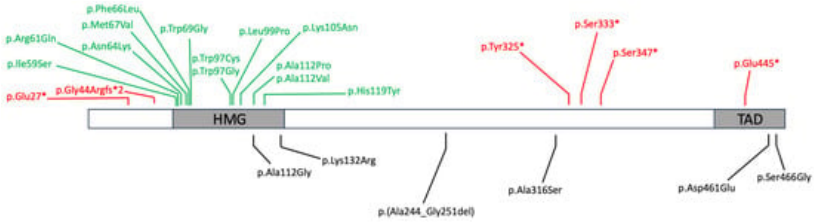

Figure 1. Representation of SOX4 protein with the two known domains (HMG: high-mobility group; TAD: transactivation domain). Known causative variants are reported above the protein (missense variants in green, stop–gain variants in red); variants of uncertain significance and the bi-allelic variant below.

Missense variants were defined as causative mainly based on their inability to bind DNA and activate transcription. Nonetheless, differences in functional tests were reported among distinct variants. For instance, most HMG variants were severely underrepresented in the nucleus when compared with SOX4 wild-type protein and did not bind DNA, whereas p.Leu99Pro was only slightly underrepresented, and p.Leu99Pro and p.Ala112Thr bound DNA weakly. Furthermore, p.Leu99Pro and p.Ala112Thr were weak in transactivation assays, whereas other HMG variants were fully inactive [24].

These differences may be a source of phenotype heterogeneity, but many more variants will need to be assessed to draw any conclusions.

Since *SOX4* is a single-exon gene, mutations that result in premature stop codons are expected to escape nonsense-mediated m-RNA decay (NMD) and produce a shorter, truncated protein, which may exert partial functions and/or impair functioning of the wild-type SOX4 copy through a dominant-negative effect. p.Gly44Argfs*2 and p.Glu27* variants fall very early in the open reading frame and encode a peptide which has an unknown function but lacks both functional DNA-binding and transactivation domains. The p.Tyr325*, p.Ser333*, and p.Ser347* variants removed 128–150 residues, including the functionally essential TAD, and p.Glu445* deprived TAD of its C-terminal segment. While the p.Gly44Argfs*2 peptide was undetectable in the cytoplasm and nuclei of transfected cells with SOX expression plasmids, p.Tyr325* and p.Glu445* were over-represented, especially in the nucleus, and p.Glu445*, although weak, was not fully inactive in the transactivation assay [24].

Pathogenic missense and truncating variants were also found to interfere with the activity of wild-type SOX4, suggesting a dominant-negative effect.

Parents were available to be tested in most (18/19) cases: 14 variants were de novo (4 stop–gain mutations, 10 missense), 4 transmitted (2 stop–gain, 2 missense). *SOX4* variants may thus be not fully penetrant, although it is not possible to draw definite conclusions. Only DNA samples from peripheral blood were tested; p.Gly44Argfs*2 was present with low-level somatic mosaicism in the patient's mother, and this woman was reported to suffer neurocognitive issues. Also, the p.Asn64Lys variant was present in the patient's father, a man with learning difficulties, mild facial dysmorphism, and limited extension of 5th finger. Two parents (the mother of a patient with p.His119Tyr and the father of a patient with p.Glu445*) carried the variant in non-mosaic state in blood and did not have ID or other relevant health issues.

3.2. Variants of Uncertain Significance and Possible Bi-Allelic Inheritance

Five missense variants are reported in the medical literature as variants of uncertain significance (VUS) [24], but many more are present in the ClinVar database (https://www.ncbi.nlm.nih.gov/clinvar/ accessed on 30 December 2023) [26] (**Figure 1**).

Of the five heterozygous variants, two are in the HMG box domain, whereas three involve different regions of the protein. Again, both in silico and functional assessments were implemented for variant classification. These five variants are extremely rare (absent in gnomAD with the exception of p.Asp461Glu—one single allele reported). Regarding p.Lys132Arg (last position of the HMG box), it is interesting to note that Lys132 is conserved in SOXC proteins, but replaced by Arg in other SOX proteins including SRY. p.Ala316Ser, outside the HMG box, lies in an unstructured and functionally uncharacterised region, and Ala316 is poorly conserved. Two variants (p.Asp461Glu and p.Ser466Gly) targeted the SOX4 TAD. Neither of these five variants was found to be less (or more) abundant in transfected cells than wild-type SOX4, and all of them showed normal activity in the transactivation assay. Although they affect the HMG box, p.Ala112Gly and p.Lys132Arg could bind DNA.

The Ala112 position deserves a specific comment: four distinct variants have been described involving this residue. p.Ala112Pro and p.Ala112Val were observed in affected individuals and were shown to alter SOX4 function, whereas p.Ala112Gly (affected individual) and p.Ala112Thr (in gnomAD, thus presumably unaffected or mildly affected) showed normal activity on functional tests. This is an interesting proof of the limits of in silico assessment and the need to perform functional studies to ascertain variant pathogenicity.

Parents were available to be tested in 3/5 cases (p.Ala112Gly, p.Asp461Glu, and p.Ser466Gly) and the variant was always transmitted. The mother carrying the p.Asp461Glu variant had epilepsy, while the mother carrying the p.Ser466Gly had learning difficulties, childhood epilepsy, and a mood disorder. The mother with the p.Ala112Gly variant is also reported to be mildly affected.

On average, the clinical features associated with VUS were reported to be milder when compared with those defined as (likely) pathogenic, but the number of cases is still too limited. It is possible that these variants, although rare, are irrelevant with respect to the patients' clinical features; otherwise, some/all of these variants may exert a pathogenic effect, possibly milder, and could be defined as "hypomorphic". No functional tests are currently available to prove this hypothesis.

Further, a bi-allelic variant, an in-frame microdeletion [c.730_753del; p.(Ala244_Gly251del)], was reported in a single consanguineous family [26]. Two affected siblings had global developmental delay, moderate to severe ID, hypotonia, and mild facial dysmorphism. The parents were heterozygous for the p.(Ala244_Gly251del) variant and had normal IQ levels with no history of hypotonia or facial dysmorphism. This variant removes eight evolutionarily conserved amino acids within a functionally unknown SOX4 domain, suggesting that sequences outside the DNA-binding and transactivation domains could modulate SOX4 activity and thus undergo pathogenic alterations. Again, the same authors hypothesize that the p.(Ala244_Gly251del) variant might be a hypomorphic allele of *SOX4*.

Functional studies were not performed, but it is tempting to think about the possibility that SOX4 activity needs to reach a threshold not to compromise normal development. Either monoallelic loss of function variants or bi-allelic hypomorphic variants may thus be disease-causing. Bi-allelic inheritance is very rare in SOXopathies. A homozygous *SOX10* deletion was found to cause a severe form of four-limb arthrogryposis but, rather than representing hypomorphic variants, it was a co-dominant occurrence and parents were both affected by Waardenburg syndrome [27]. Hypotrichosis–lymphedema–telangiectasia is a very rare disorder caused by heterozygous loss of function variants affecting *SOX18*, but the first report also described unrelated individuals with homozygous missense variants (p.Ala104Pro and p.Trp95Arg) and healthy heterozygous parents [28]; functional studies were not performed. However, these may represent hypomorphic variants.

3.3. Co-Occurrence of Variants in Other Genes

The presence of pathogenic variations at two distinct loci that lead to the expression of two Mendelian conditions, which segregate independently, has been appreciated as a relatively frequent phenomenon after the introduction of large-scale sequencing studies. It is termed "dual molecular diagnosis", and several reports have demonstrated that this scenario occurs in a percentage of approximately 5% among patients who received a molecular diagnosis [29]. The two diagnoses can be "distinct", when the two conditions have different phenotypes, or "overlapping", when at least some of the clinical features are common to the two disorders [30].

At least three *SOX4* patients have a second causative variant: (a) the proband carrying p.Trp69Gly has a distinct molecular diagnosis, a bleeding disorder caused by a stop–gain mutation (p.Gln106*) in *F11*; (b) the patient with p.Gly44Argfs*2 has a *TTN*-related cardiomyopathy caused by a stop–gain variant (p.Arg18985*), and, since heart defects are frequently associated with *SOX4*, the presence of cardiomyopathy may be considered an overlapping phenotype; (c) the proband carrying the p.Arg466Gly VUS has an associated autosomal dominant myopia caused by a frameshift variant (p.Arg179Valfs*224) in *SLC39A5*, and it is possible that a liability to developmental delay was worsened or unveiled by an associated disorder causing poor eye-sight.

Furthermore, two VUS of potential interest were reported, p.Pro639Arg in *PHF8* in the patient carrying the p.Tyr325* variant, and p.R1406H in *CHD4* in the patient carrying p.Ala112Gly. Both *PHF8* and *CHD4* encode chromatin remodellers, associated with human developmental disorders [31][32], which might interact with SOX4 and contribute to the clinical phenotype.

References

1. Schepers, G.E.; Teasdale, R.D.; Koopman, P. Twenty Pairs of Sox: Extent, Homology, and Nomenclature of the Mouse and Human Sox Transcription Factor Gene Families. Dev. Cell 2002, 3, 167–170.
2. Harley, V.R.; Clarkson, M.J.; Argentaro, A. The Molecular Action and Regulation of the Testis-Determining Factors, SRY (Sex-Determining Region on the Y Chromosome) and SOX9 . Endocr. Rev. 2003, 24, 466–487.
3. Sutton, E.; Hughes, J.; White, S.; Sekido, R.; Tan, J.; Arboleda, V.; Rogers, N.; Knower, K.; Rowley, L.; Eyre, H.; et al. Identification of SOX3 as an XX Male Sex Reversal Gene in Mice and Humans. J. Clin. Investig. 2011, 121, 328–341.
4. Fantes, J.; Ragge, N.K.; Lynch, S.-A.; McGill, N.I.; Collin, J.R.O.; Howard-Peebles, P.N.; Hayward, C.; Vivian, A.J.; Williamson, K.; van Heyningen, V.; et al. Mutations in SOX2 Cause Anophthalmia. Nat. Genet. 2003, 33, 462–463.
5. Wagner, T.; Wirth, J.; Meyer, J.; Zabel, B.; Held, M.; Zimmer, J.; Pasantes, J.; Bricarelli, F.D.; Keutel, J.; Hustert, E.; et al. Autosomal Sex Reversal and Campomelic Dysplasia Are Caused by Mutations in and around the SRY-Related Gene SOX9. Cell 1994, 79, 1111–1120.
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. 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.
7. Pingault, V.; Zerad, L.; Bertani-Torres, W.; Bondurand, N. SOX10: 20 Years of Phenotypic Plurality and Current Understanding of Its Developmental Function. J. Med. Genet. 2022, 59, 105–114.
8. Angelozzi, M.; Lefebvre, V. SOXopathies: Growing Family of Developmental Disorders Due to SOX Mutations. Trends Genet. 2019, 35, 658–671.
9. Kavyanifar, A.; Turan, S.; Lie, D.C. SoxC Transcription Factors: Multifunctional Regulators of Neurodevelopment. Cell Tissue Res. 2018, 371, 91–103.
10. Schilham, M.W.; Oosterwegel, M.A.; Moerer, P.; Ya, J.; de Boer, P.A.J.; van de Wetering, M.; Verbeek, S.; Lamers, W.H.; Kruisbeek, A.M.; Cumano, A.; et al. Defects in Cardiac Outflow Tract Formation and Pro-B-Lymphocyte Expansion in Mice Lacking Sox-4. Nature 1996, 380, 711–714.
11. Sock, E.; Rettig, S.D.; Enderich, J.; Bösl, M.R.; Tamm, E.R.; Wegner, M. Gene Targeting Reveals a Widespread Role for the High-Mobility-Group Transcription Factor Sox11 in Tissue Remodeling. Mol. Cell. Biol. 2004, 24, 6635–6644.
12. Hoser, M.; Potzner, M.R.; Koch, J.M.C.; Bösl, M.R.; Wegner, M.; Sock, E. Sox12 Deletion in the Mouse Reveals Nonreciprocal Redundancy with the Related Sox4 and Sox11 Transcription Factors. Mol. Cell. Biol. 2008, 28, 4675–4687.
13. Tsurusaki, Y.; Koshimizu, E.; Ohashi, H.; Phadke, S.; Kou, I.; Shiina, M.; Suzuki, T.; Okamoto, N.; Imamura, S.; Yamashita, M.; et al. De Novo SOX11 Mutations Cause Coffin–Siris Syndrome. Nat. Commun. 2014, 5, 4011.
14. Hempel, A.; Pagnamenta, A.T.; Blyth, M.; Mansour, S.; McConnell, V.; Kou, I.; Ikegawa, S.; Tsurusaki, Y.; Matsumoto, N.; Lo-Castro, A.; et al. Deletions and de Novo Mutations of SOX11 Are Associated with a Neurodevelopmental Disorder with Features of Coffin–Siris Syndrome. J. Med. Genet. 2016, 53, 152–162.
15. Al-Jawahiri, R.; Foroutan, A.; Kerkhof, J.; McConkey, H.; Levy, M.; Haghshenas, S.; Rooney, K.; Turner, J.; Shears, D.; Holder, M.; et al. SOX11 Variants Cause a Neurodevelopmental Disorder with Infrequent Ocular Malformations and Hypogonadotropic Hypogonadism and with Distinct DNA Methylation Profile. Genet. Med. 2022, 24, 1261–1273.
16. Van de Wetering, M.; Oosterwegel, M.; van Norren, K.; Clevers, H. Sox-4, an Sry-like HMG Box Protein, Is a Transcriptional Activator in Lymphocytes. EMBO J. 1993, 12, 3847–3854.

17. Bhattaram, P.; Penzo-Méndez, A.; Sock, E.; Colmenares, C.; Kaneko, K.J.; Vassilev, A.; DePamphilis, M.L.; Wegner, M.; Lefebvre, V. Organogenesis Relies on SoxC Transcription Factors for the Survival of Neural and Mesenchymal Progenitors. Nat. Commun. 2010, 1, 9.
18. Shim, S.; Kwan, K.Y.; Li, M.; Lefebvre, V.; Šestan, N. Cis-Regulatory Control of Corticospinal System Development and Evolution. Nature 2012, 486, 74–79.
19. Bhattaram, P.; Penzo-Méndez, A.; Kato, K.; Bandyopadhyay, K.; Gadi, A.; Taketo, M.M.; Lefebvre, V. SOXC Proteins Amplify Canonical WNT Signaling to Secure Nonchondrocytic Fates in Skeletogenesis. J. Cell Biol. 2014, 207, 657–671.
20. Sun, B.; Mallampati, S.; Gong, Y.; Wang, D.; Lefebvre, V.; Sun, X. Sox4 Is Required for the Survival of Pro-B Cells. J. Immunol. 2013, 190, 2080–2089.
21. Wang, X.; Llamas, J.; Trecek, T.; Shi, T.; Tao, L.; Makmura, W.; Crump, J.G.; Segil, N.; Gnedeva, K. SoxC Transcription Factors Shape the Epigenetic Landscape to Establish Competence for Sensory Differentiation in the Mammalian Organ of Corti. Proc. Natl. Acad. Sci. USA 2023, 120, e2301301120.
22. Scharer, C.D.; McCabe, C.D.; Ali-Seyed, M.; Berger, M.F.; Bulyk, M.L.; Moreno, C.S. Genome-Wide Promoter Analysis of the SOX4 Transcriptional Network in Prostate Cancer Cells. Cancer Res. 2009, 69, 709–717.
23. 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. De Novo SOX4 Variants Cause a Neurodevelopmental Disease Associated with Mild Dysmorphism. Am. J. Hum. Genet. 2019, 104, 246–259.
24. Angelozzi, M.; Karvande, A.; Molin, A.N.; Ritter, A.L.; Leonard, J.M.M.; Savatt, J.M.; Douglass, K.; Myers, S.M.; Grippa, M.; Tolchin, D.; et al. Consolidation of the Clinical and Genetic Definition of a SOX4-Related Neurodevelopmental Syndrome. J. Med. Genet. 2022, 59, 1058–1068.
25. Grosse, M.; Kuechler, A.; Dabir, T.; Spranger, S.; Beck-Wödl, S.; Bertrand, M.; Haack, T.B.; Grasemann, C.; Manka, E.; Depienne, C.; et al. Novel Variants of SOX4 in Patients with Intellectual Disability. Int. J. Mol. Sci. 2023, 24, 3519.
26. Ghaffar, A.; Rasheed, F.; Rashid, M.; van Bokhoven, H.; Ahmed, Z.M.; Riazuddin, S.; Riazuddin, S. Biallelic In-Frame Deletion of SOX4 Is Associated with Developmental Delay, Hypotonia and Intellectual Disability. Eur. J. Hum. Genet. 2022, 30, 243–247.
27. Stevenson, R.E.; Vincent, V.; Spellicy, C.J.; Friez, M.J.; Chaubey, A. Biallelic Deletions of the Waardenburg II Syndrome Gene, SOX10, Cause a Recognizable Arthrogryposis Syndrome. Am. J. Med. Genet. Part A 2018, 176, 1968–1971.
28. Irrthum, A.; Devriendt, K.; Chitayat, D.; Matthijs, G.; Glade, C.; Steijlen, P.M.; Fryns, J.-P.; Van Steensel, M.A.M.; Vikkula, M. Mutations in the Transcription Factor Gene SOX18 Underlie Recessive and Dominant Forms of Hypotrichosis-Lymphedema-Telangiectasia. Am. J. Hum. Genet. 2003, 72, 1470–1478.
29. Posey, J.E.; Harel, T.; Liu, P.; Rosenfeld, J.A.; James, R.A.; Coban Akdemir, Z.H.; Walkiewicz, M.; Bi, W.; Xiao, R.; Ding, Y.; et al. Resolution of Disease Phenotypes Resulting from Multilocus Genomic Variation. N. Engl. J. Med. 2017, 376, 21–31.
30. Smith, E.D.; Blanco, K.; Sajan, S.A.; Hunter, J.M.; Shinde, D.N.; Wayburn, B.; Rossi, M.; Huang, J.; Stevens, C.A.; Muss, C.; et al. A Retrospective Review of Multiple Findings in Diagnostic Exome Sequencing: Half Are Distinct and Half Are Overlapping Diagnoses. Genet. Med. 2019, 21, 2199–2207.
31. Sobering, A.K.; Bryant, L.M.; Li, D.; McGaughran, J.; Maystadt, I.; Moortgat, S.; Graham, J.M.; van Haeringen, A.; Ruivenkamp, C.; Cuperus, R.; et al. Variants in PHF8 Cause a Spectrum of X-Linked Neurodevelopmental Disorders and Facial Dysmorphology. HGG Adv. 2022, 3, 100102.
32. Weiss, K.; Terhal, P.A.; Cohen, L.; Bruccoleri, M.; Irving, M.; Martinez, A.F.; Rosenfeld, J.A.; Machol, K.; Yang, Y.; Liu, P.; et al. De Novo Mutations in CHD4, an ATP-Dependent Chromatin Remodeler Gene, Cause an Intellectual Disability Syndrome with Distinctive Dysmorphisms. Am. J. Hum. Genet. 2016, 99, 934–941.