SUPPLEMENTARY DATA

SUPPLEMENTARY TABLES

Table S1. NCBI accession numbers for SOX protein sequences used in this study

Protein	Animal Species	Reference Sequence
SOX8	Human (*Homo sapiens*)	NP_055402.2
	Mouse (*Mus musculus*)	NP_035577.1
	Dog (*Canis familiaris*)	XP_022275986
	Chicken (*Gallus gallus*)	AAF73917.1
	African clawed frog (*Xenopus laevis*)	AAQ67212.1
	Coelacanth (*Latimeria menadoensis*)	XP_005993175.1
	Zebrafish (*Danio rerio*)	AAX73357.1
SOX9	Human (*Homo sapiens*)	CAA86598.1
	Mouse (*Mus musculus*)	NP_035578.3
	Dog (*Canis familiaris*)	AAP69840.1
	Chicken (*Gallus gallus*)	NP_989612.1
	African claw frog (*Xenopus laevis*)	AAI70060.1
	Coelacanth (*Latimeria menadoensis*)	CCP19141
	Zebrafish (*Danio rerio*)	NP_571718.1
	Lamprey [SOXE3] (*Petromyzon marinus*)	ABC58685.1
SOX10	Human (*Homo sapiens*)	CAG38808.1
	Mouse (*Mus musculus*)	AAH25171.1
	Dog (*Canis familiaris*)	XP_538379.3
	Chicken (*Gallus gallus*)	NP_990123.1
	African claw frog (*Xenopus laevis*)	NP_001082358.1
	Coelacanth (*Latimeria menadoensis*)	XP_006002010.1
	Zebrafish (*Danio rerio*)	AAK84872.1
SOXE	Sea squirt (*Ciona intestinalis*)	CAD58841
	Green sea urchin (*Lytechinus variegatus*)	ALG35687.1
	Common fruit fly (*Drosophila melanogaster*)	CAB63903.1
	Velvet worm (*Euperipatoides kanangrensis*)	SOB55490.1
	Common cuttlefish (*Sepia officinalis*)	AGL08099
	Starlet sea anemone [SOXE1] (*Nematostella vectensis*)	ABA02365.1
SOX7	Human (*Homo sapiens*)	CAC84226.1
SOX17	Human (*Homo sapiens*)	BAB83867.1
	African clawed frog {SOX17B] (*Xenopus laevis*)	NP_001081633.1
SOX18	Human (*Homo sapiens*)	BAA94874.1
	Chicken (*Gallus gallus*)	AAK71352.1
SOXF	Lamprey (*Petromyzon marinus*)	AAW34333.1
	Velvet worm (*Euperipatoides kanangrensis*)	SOB55491.1
	Starlet sea anemone [SOXE1] (*Nematostella vectensis*)	ABA02366.1

Table S2. Primers used to generate plasmids encoding SOX proteins and variants thereof

Plasmid name	Forward primer	Reverse primer
pCDNA-3xFLAG-hSOX9 WT	TAGTGGATCCATGAATCTCCTGGACC	CGAGGTCGACGGTATCGATAAGCT
pCDNA-3xFLAG-hSOX9$^{\Delta TAC}$	TAGTGGATCCATGAATCTCCTGGACC	GCAGAATTCTCAGCCCGGCTCGCTGCTCAGCGT
pCDNA-3xFLAG-hSOX9$^{[1-338]}$	TAGTGGATCCATGAATCTCCTGGACC	GCAGAATTCTCACTTGGACATCCACACGTGGCC
pCDNA-3xFLAG-hSOX9$^{[1-307]}$	TAGTGGATCCATGAATCTCCTGGACC	GCAGAATTCTCACGGCACCCCCGGGTGGCCGT
pCDNA-3xFLAG-hSOX9$^{[1-224]}$	TAGTGGATCCATGAATCTCCTGGACC	GCAGAATCCTCAGTGCTCGCCGGGGGAGTGCAC
pCDNA-3xFLAG-hSOX9$^{\Delta TAM}$	GAGCACCCGGGCAAGGCTGACCTGAAGCGAGA	GCCCGGGTGCTCGCCGGGGGAGTGCACCTC
pCDNA-3xFLAG-hSOX9$^{\Delta TAM-A}$	GAGCACCCGGGCAAGGCGAC CTGAAGCGAGA	GCCCGGGTGCTCGCCGGGGGAGTGCACCTC
pCDNA-3xFLAG-hSOX9$^{\Delta TAM-B}$	CAGCCG GGCCCCCCTATCGACTTCCGC	GATAGGGGGGCCCGGCTGCACGTCGGT
pCDNA-3xFLAG-hSOX9$^{\Delta TAM-C}$	CAGCCCCCTTCCAACATCGAGACCTTC	GATGTTGGAAGGGGGCTGTCTGCCCCC
pCDNA-3xFLAG-hSOX9$^{\Delta TAM-D}$	ATCTCCAACCCGGGGGTGCCGGCCACG	CACCCCCGGGTTGGAGATGACGTCGCT
pCDNA-3xFLAG-hSOX9$^{\Delta PQA}$	CAGCGTGTGCTTGGACATCCACACGTG	ATGTCCAAGCACACGCTGACCACGCTG
pCDNA-3xFLAG-hSOX9$^{\Delta E-L}$	GATGTCAACCCGCCCAACGGCCACCCGGGGGT	GTTGGGCGGGTTGACATCGAAGGTCTCGATGT
pCDNA-3xFLAG-hSOX8	GACGGGATCCATGCTGGACATGAGCGAGGCC	CGTGAATTCTCAGGGCCTGGTCAGGGTGGT
pCDNA-3xFLAG-hSOX8$^{\Delta TAC}$	GACGGGATCCATGCTGGACATGAGCGAGGCC	GCAGAATTCTCACGTCTCGGTGGGCGACGCGGAG
pCDNA-3xFLAG-hSOX8$^{\Delta TAM}$	CATGGCGACCACCCGGGCCAGGCCTATGGG	TAGGCCTGGCCCGGGTGGTCGCCATGGTGGTG
pCDNA-3xFLAG-hSOX10	GACGGGATCCATGGCGGAGGAGCAGGACCTA	CGTGAATTCTTAGGGCCGGGACAGTGTCGT
pCDNA-3xFLAG-hSOX10$^{\Delta TAC}$	GACGGGATCCATGGCGGAGGAGCAGGACCTA	GCAGAATTCTTAACCAGGTGGTGAGACCGTG
pCDNA-3xFLAG-hSOX10$^{\Delta TAM}$	CCCGAGCACCCCAGCAGCTACTCAGCAGCC	TGAGTAGCTGCTGGGGTGCTCGGGGTTCCCAT
pCDNA-3xFLAG-hSOX9 [SOX8TAM]	CAGTCCCGGGGACCACACAGGGCAGAC	CAGTCCCGGGGGGCCGCCCAGGGGCAGGTA
pCDNA-3xFLAG-hSOX9 [SOX10TAM]	CAGTCCCGGG CACCCCTCAGGCCAGAGC	CAGTCCCGGGTGCCCATTGGGCGGCAGGTA
pCDNA-3xFLAG-hSOX9 [SOX8TAC]	GGCCAGTCCCAGCGGCCGCACATCAAGACG	GATGTGCGGCCGCTGGGACTGGCCCGGCTC
pCDNA-3xFLAG-hSOX9 [SOX10TAC]	GGCCAGTCCCAGAAAGCCCAGGTGAAGACA	CACCTGGGCTTTCTGGGACTGGCCCGGCTC
pCDNA-3xFLAG-hSOX17	AACGGATCCATGAGCAGCCCGGATGCGGGATAC	TAGCGAATTCTCACACGTCAGGATAGTTGCAGTA
pCDNA-3xFLAG-hSOX17$^{\Delta E-L}$	GACCGCACGTTCGTGTGCAAGCCTGAGAT	GCACACGAACGTGCGGTCCACCTCCCCGA

Table S3. Primers used for the generation of plasmids encoding GAL4^{DBD}/SOXE fusion proteins

Plasmid name	Forward primer	Reverse primer
pBind-hSOX9^{TAM}	CGGGATCCGTCCCGGCGAGCACTCGGGGCAA	CGGAGATCTCACGGCACCCCCGGGTGGCCGT
pBind-hSOX9^{TAC}	CGGGATCCGTCAGTCCCAGCGAACGCACATCA	CGGAGATCTCAAGGTCGAGTGAGCTGTGTGTA
pBind-hSOX9$^{TAM>TAC}$	CGGGATCCGTCCCGGCGAGCACTCGGGGCAA	CGGAGATCTCAAGGTCGAGTGAGCTGTGTGTA
pBind-hSOX9^{PQA}	ATGGGATCCAGCAGGCGCCGCCGCCA	CGGTCTAGATCACAGCGTGGTCAGCGTGTGCGCCTG
pBind-hSOX9^{TAM-AB}	CGGGATCCGTCCCGGCGAGCACTCGGGGCAA	CGGAGATCTCAGGGCTGTCTGCCCCCCTCTGG
pBind-hSOX9^{TAM-CD}	CGGGATCCGTCCTATCGACTTCCGCGACGTG	CGGAGATCTCACGGCACCCCCGGGTGGCCGT
pBind-hSOX9^{TAM-C}	CGGGATCCGTCCTATCGACTTCCGCGACGTG	CGGTCTAGATCAGGTCTCGATGTTGGAGATGAC
pBind-hSOX9^{TAM-D}	CGGGATCCGTAACATCGAGACCTTCGATGTCAACG	CGGTCTAGATCACGGCACCCCCGGGTGGCCGT

pBind-hSOX8^{TAM}	CGGGATCCGTCATGCGA CCACACAGGGCAG	CGGTCTAGATCACTCGGGTGGGGCGGGGCCGC
pBind-hSOX8^{TAC}	CGGGATCCGTGGTCCCCCACGGCCGCACATC	CGGTCTAGATCAGGGCCTGGTCAGGGTGGTG
pBind-hSOX8$^{TAM>TAC}$	CGGGATCCGTCATGGCGACCACACAGGGCAG	CGGTCTAGATCAGGGCCTGGTCAGGGTGGTG
pBind-hSOX10^{TAM}	CGGGATCCGTCCCGAGCACCCCTCAGGCCAG	CGGTCTAGATCACACATGGCCTGGGTGCCCATTG
pBind-hSOX10^{TAC}	CGGGATCCGTGTGGATGCCAAAGCCCAGGTG	CGGTCTAGATCAGGGCCGGGACAGTGTCGTATATAC
pBind-hSOX10$^{TAM>TAC}$	CGGGATCCGTCCCGAGCACCCCTCAGGCCAG	CGGTCTAGATCAGGGCCGGGACAGTGTCGTATATAC

Table S4. Primers used to generate plasmids encoding SOX9 and GAL4^{DBD}/SOX9 proteins with missense variants. Wild-type sequences are shown as references. Wild-type and variant codons are shown in blue and red, respectively, and mutations in bold letters.

Variant	Forward primer	Reverse primer
Wild-type	TGGACATC**GGC**GAGCTGAGCAGCGAC	TCAGCTC**GCC**GATGTCCACGTCGCGG
G276S	TGGACATC**AGC**GAGCTGAGCAGCGAC	TCAGCTC**GCT**GATGTCCACGTCGCGG
Wild-type	ACATCGGC**GAG**CTGAGCAGCGACGTCATC	TCGCTGCTCAG**CTC**GCCGATGTCCACG
E277A	ACATCGGC**GCC**CTGAGCAGCGACGTCATC	TCGCTGCTCAG**GGC**GCCGATGTCCACG
Wild-type	ACATCGGCGAG**CTG**AGCAGCGACGTCATCTCC	CGTCGCTGCT**CAG**CTCGCCGATGTCCACG
L278S	ACATCGGCGAG**AGC**AGCAGCGACGTCATCTCC	CGTCGCTGCT**GCT**CTCGCCGATGTCCACG
Wild-type	ACATCGGCGAG**CTG**AGCAGCGACGTCATC	CGTCGCTGCT**CAG**CTCGCCGATGTCCACG
L278F	ACATCGGCGAG**TTT**AGCAGCGACGTCATC	CGTCGCTGCT**AAA**CTCGCCGATGTCCACG
Wild-type	TGAGCAGC**GAC**GTCATCTCCAACATCGAG	GAGATGAC**GTC**GCTGCTCAGCTCGCC
D281A	TGAGCAGC**GCC**GTCATCTCCAACATCGAG	GAGATGAC**GGC**GCTGCTCAGCTCGCC
Wild-type	TGAGCAGC**GAC**GTCATCTCCAACATCGAG	CTCGATGTTGGAGATGAC**GTC**GCTGCTCAG
D281Y	TGAGCAGC**TAC**GTCATCTCCAACATCGAG	CTCGATGTTGGAGATGAC**GTA**GCTGCTCAG
Wild-type	AGCTGAGCAGC**GAC**GTCATCTCCAACATCG	TGGAGATGAC**GTC**GCTGCTCAGCTCG
D281L	AGCTGAGCAGC**CTC**GTCATCTCCAACATCG	TGGAGATGAC**GAG**GCTGCTCAGCTCG
Wild-type	TGAGCAGCGAC**GTC**ATCTCCAACATCGAG	CTCGATGTTGGAGAT**GAC**GTCGCTGCTCAG
V282D	TGAGCAGCGAC**GAC**ATCTCCAACATCGAG	CTCGATGTTGGAGAT**GTC**GTCGCTGCTCAG
Wild-type	TGAGCAGCGACGTC**ATC**TCCAACATCGAG	CTCGATGTTGGAGAT**GAC**GTCGCTGCTCAG
I283F	TGAGCAGCGAC**TTC**TCCAACATCGAG	CTCGATGTTGGAGAT**GTC**GTCGCTGCTCAG
Wild-type	TGAGCAGCGACGTCATC**TCC**AACATCGAG	CTCGATGTTGGAGAT**GAC**GTCGCTGCTCAG
S284F	TGAGCAGCGACGACATC**TTC**AACATCGAG	CTCGATGTTGGAGAT**GTC**GTCGCTGCTCAG
Wild-type	AGACCTTC**GAT**GTCAACGAGTTTGACCAG	TCGTTGAC**ATC**GAAGGTCTCGATGTTG
D290A	AGACCTTC**GCT**GTCAACGAGTTTGACCAG	TCGTTGAC**AGC**GAAGGTCTCGATGTTG
Wild-type	GAGACCTTCGATGTCAAC**GAG**TTTGACCAGTACCTGCC	GGCAGGTACTGGTCAAA**CTC**GTTGACATCGAAGGTCTC
E293G	GAGACCTTCGATGTCAAC**AGG**TTTGACCAGTACCTGCC	GGCAGGTACTGGTCAAA**CAT**GTTGACATCGAAGGTCTC
Wild-type	GAGACCTTCGATGTCAAC**GAG**TTTGACCAGTACCTGCC	GGCAGGTACTGGTCAAA**CTC**GTTGACATCGAAGGTCTC
E293K	GAGACCTTCGATGTCAAC**AAG**TTTGACCAGTACCTGCC	GGCAGGTACTGGTCAAA**CAT**GTTGACATCGAAGGTCTC

Wild-type	GAGACCTTCGATGTCAACGAGTTTGACCAGTACCTGCC	GGCAGGTACTGGTCAAACTCGTTGACATCGAAGGTCTC
E293M	GAGACCTTCGATGTCAACATGTTTGACCAGTACCTGCC	GGCAGGTACTGGTCAAACATGTTGACATCGAAGGTCTC
Wild-type	GAGACCTTCGATGTCAACGAGTTTGACCAGTACCTGCC	GGCAGGTACTGGTCAAACTCGTTGACATCGAAGGTCTC
E293T	GAGACCTTCGATGTCAACACGTTTGACCAGTACCTGCC	GGCAGGTACTGGTCAAACGGTGTTGACATCGAAGGTCTC
Wild-type	ATGTCAACGAGTTTGACCAGTACCTGCCGCCC	AGGTACTGGTCAAACTCGTTGACATCGAAGG
F294L	ATGTCAACGAGCTGGACCAGTACCTGCCGCCC	AGGTACTGGTCCAGCTCGTTGACATCGAAGG
Wild-type	ATGTCAACGAGTTTGACCAGTACCTGCCGCCC	AGGTACTGGTCAAACTCGTTGACATCGAAGG
F294S	ATGTCAACGAGTCTGACCAGTACCTGCCGCCC	AGGTACTGGTCAGACTCGTTGACATCGAAGG
Wild-type	ATGTCAACGAGTTTGACCAGTACCTGCCGCCC	AGGTACTGGTCAAACTCGTTGACATCGAAGG
D295Y	ATGTCAACGAGTCTTACCAGTACCTGCCGCCC	AGGTACTGGTCAGACTCGTTGACATCGAAGG
Wild-type	AGTTTGACCAGTACGATCCGCCCAAC	TTGGGCGGATCGTACTGGTCAAACTCG
Q296R	AGTTTGACCGTTACGATCCGCCCAAC	TTGGGCGGATCGTAACGGTCAAACTCG
Wild-type	AGTTTGACCAGTACCTGCCGCCC	TGGGCGGCAGGTACTGGTCAAAC
Y297A	AGTTTGACCAGGCCCTGCCGCCC	TGGGCGGCAGGGCCTGGTCAAAC
Wild-type	AGTTTGACCAGTACCTGCCGCCC	TGGGCGGCAGGTACTGGTCAAAC
Y297S	AGTTTGACCAGTCCCTGCCGCCC	TGGGCGGCAGGGACTGGTCAAAC
Wild-type	AGTTTGACCAGTACCTGCCGCCC	TGGGCGGCAGGTACTGGTCAAAC
Y297D	AGTTTGACCAGGATCTGCCGCCC	TGGGCGGCAGATCCTGGTCAAAC
Wild-type	AGTTTGACCAGTACCTGCCGCCCAACG	TGGGCGGCAGGTACTGGTCAAACTCGTTG
Y297L	AGTTTGACCAGTTGCTGCCGCCCAACG	TGGGCGGCAGCAACTGGTCAAACTCGTTG
Wild-type	AGTTTGACCAGTACCTGCCGCCCAACG	TGGGCGGCAGGTACTGGTCAAACTCGTTG
Y297F	AGTTTGACCAGTTCCTGCCGCCCAACG	TGGGCGGCAGGAACTGGTCAAACTCGTTG
Wild-type	AGTTTGACCAGTACCTGCCGCCCAAC	GTTGGGCGGCAGGTACTGGTCAAAC
L298D	AGTTTGACCAGTACGATCCGCCCAAC	GTTGGGCGGATCGTACTGGTCAAAC

Table S5. Primers used for mRNA expression analysis by qRT-PCR

Gene	Forward primer	Reverse primer
Acan	GATCTACCGCTGTGAAGTGATG	GGGTGTAGCGTGTGGAAATAG
Col2a1	ACATAGGGCCTGTCTGCTTCTTGT	TGACTGCGGTTGGAAAGTGTTTGG
Hprt	CCTCATGGACTGATTATGGACAG	TCAGCAAAGAACTTATAGCCCC

SUPPLEMENTARY FIGURES

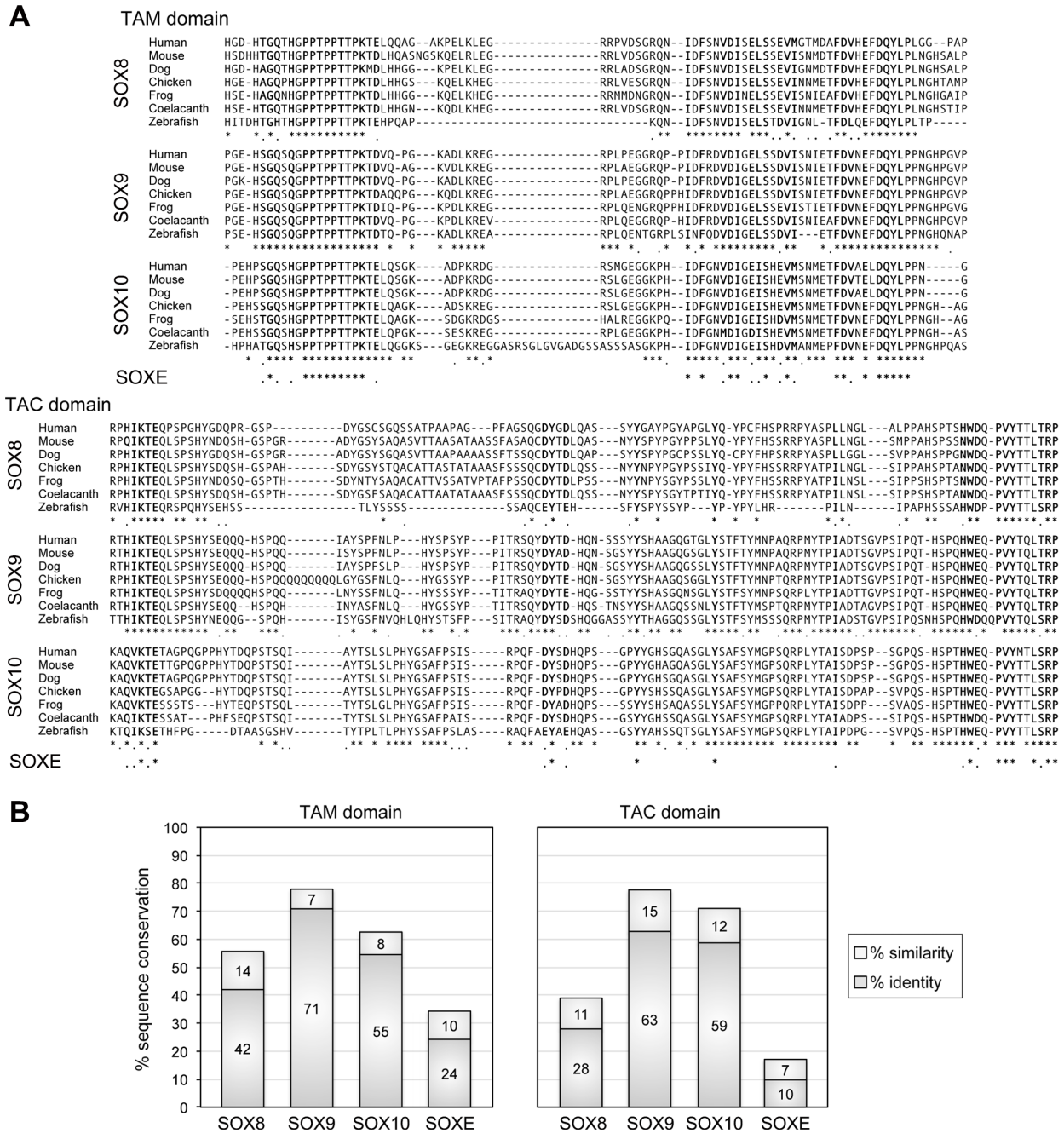
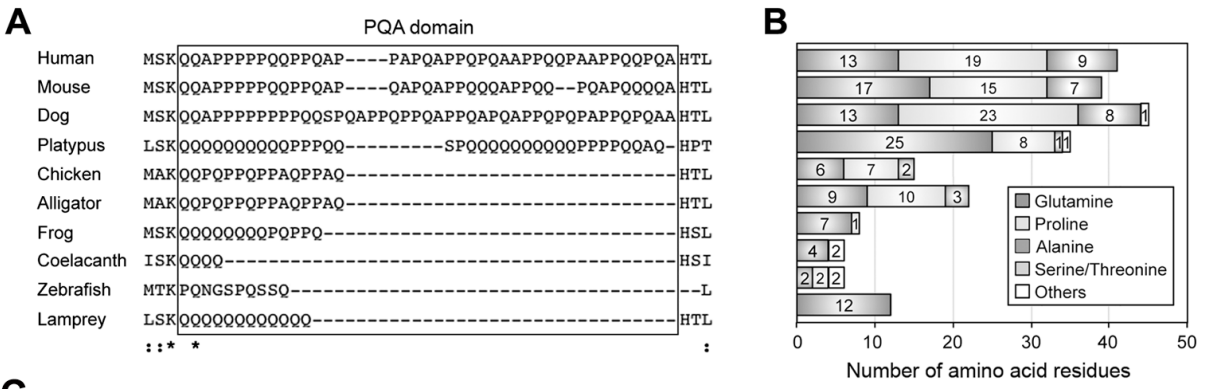


Figure S1. TAM and TAC conservation in vertebrate SOXE proteins. (**A**) ClustalW alignment of TAM and TAC in SOXE proteins from human (*Homo sapiens*), mouse (*Mus musculus*), dog (*Canis lupus familiaris*), chicken (*Gallus gallus*), African clawed frog (*Xenopus laevis*), coelacanth (*Latimeria chalumnae*) and zebrafish (*Danio rerio*). sox9a, sox8b, and sox10b were used for zebrafish. Conservation is shown with stars for identical residues and dots for similar residues. (**B**) Quantification of conservation in TAM and TAC in SOXE orthologs and all SOXE proteins. Percentages of sequence identity and similarity were calculated based on ClustalW alignment.



C

Human (*Homo sapiens*)

MNLLDPFMKMTDEQEKGLSGAPSPTMSEDSAGSPCPSGSGSDTENTRPQENTFPKGEPDLKKESEEDKFPVCIREAVSQVLKGYDWTLVPMPVRVNGSSKNK
PHVKRPMNAFMVWAQAARRKLADQYPHLHNAELSKTLGKLWRLLNESEKRPFVEEAERLRVQHKKDHPDYKYQPRRRKSVKNGQAEAEEATEQTHISPNAIF
KALQADSPHSSSGMSEVHSPGEHSGQSQGPPTPPTTPKTDVQPGKADLKREGRPLPEGGRQPPIDFRDVDIGELSSDVISNIETFDVNEFDQYLPPNGHPGV
PATHGQVTYTGSYGISSTAATPASAGHVW**MSKQQAPPPPPQQPPQAPPAPQAPPQPQAAPPQQPAAPPQQPQAHTL**TTLSSEPGQSQRTHIKTEQLSPSHYS
EQQQHSPQQIAYSPFNLPHYSPSYPPITRSQYDYTDHQNSSSYYSHAAGQGTGLYSTFTYMNPAQRPMYTPIADTSGVPSIPQTHSPQHWEQPVYTQLTRP

Sea urchin (*Lytechinus variegatus*)

MSSPESLELHHSLSEGSSPRTPGSDSDDSSSECSREDLAILPGRVDPSALVVGHEGAATQFSPSIKDAVSRVLNGYDWSVVAIPTRTGPNGKRK**PHIKRPMN**
AFMVWAQAARKKLGNQYPQLHNAELSKTLGKLWRLLSDKEKQPFIEEAERLRQQHKKDYPDYKYQPRRRNKNDNSNTKKPCPPNNRSTLVPSPDSSNHVSTK
ALLSAMVGEEITEANMKERTEKLGMMMGGAGGQGPPTPPTTPKNDLDCTRPNKRQKYSLKVKTEMPVDFAGVDVRDFGGDIMGMEEFSSE**ELDQYI**VQTIAS
VTASQPMPCQQGMVRQTCAMPPFTTHSSYP**MS**NVNT**Q**SSNGR**Q**WMGGRHH**P**SGGNTS**P**L**QA**TVLDNVNSKLE**H**DMMSPPPQYPSSQ**QLHQMQAFHFAAMQQA**
QEQQPPQQQPYDFRQSQCEYPAQQHSPQQQAQMDFYNANAGATPVQNMPPAYQYPHTSPQRSPAYVDLTEPATTMIPESRPWDSFAGTVRS

Velvet worm (*Euperipatoides kanangrensis*)

GYDWTLVPLPTRQNGSEKRK**PHVKRPMNAFMVWAQAARRKLADQYPHLHNAELSKTLGKLWRLLNDDEKKPFIEEAERLRVVHKKEHPDYKYQPRRRK**PLKG
AANSSDLVGQSSPTVIFRTLKPQIENSASQDSESSSVISAKTSPSGSTHGPPTPPTTPNHQDRLSGKESSCLKMSHSNRTGRNETAPPIDFSHIDIGQLSSD
VLNPIENFDES**ELDQYL**PPNGQPGLSRDHHPYSVNYSSSISVPTTATAPSW**MSK**YCL**A**TVVSGSTYV**P**SSVTNNKDCS**Q**NMDNYSSSTYHTTDESRF**HE**L**Q**P
SPS**VKLEHLPSRQHSSSQDNFAQSRLLQHHYSGSNYYSSNQASMMPSYTCMMASRGTIFPS**

Common fruit fly (*Drosophila melanogaster*)

MSDSSSSNCSKDRAKPVETLVLANYALKAEQKKAQGQGGRKEDERITTAVMKVLEGYDWNLVQASAKAPTDRKK**EHIKRPMNAFMVWAQAARRVMSKQYPHL**
QNSELSKSLGKLWKNLKDSDKKPFMEFAEKLRMTHKQEHPDYKYQPRRKKARVLPSQQSGEGGSPGPEMTLSATMGSSGKPRSSNSNGQRRAGKGNAAADLG
SCASTISHANVGSNSSDVFSNEAFMKSLNSACAASLMEQSLIETGLDSPCSTASSMSSLTPPATPYNVAPSNAKASAANNPSLLLRQLSEPVANAGDGYGVL
LEAGREYVAIGEVNYQGQSAGVQSGAEGGGAGQEMDFLENINGYGGYTGSRVSYPAYSYPANGGHFATEEQQQQQALQASEALNYKPAAADIDPK**EIDQYF**M
DQMLPMTQHHHPHHTHPLHHPLHHSPPLNSSA**SLSSACSSASSQQPVAEYYEHLGYSPAASSASQNPNFGPQQPYANGA**ASMTPTLGDP**APQQELQSQQQEQ**
QHQNPSQHHLWGTYTYVNP

Starlet sea anemona (*Nematostella vectensis*)

MDKKVTEQQVQAVLGLDTDGSQVRNHQLSNAIASAVNHVLDGYDWSLIPLPVRVNGIKTQK**PHVKRPMNAFMVWAQAVRRKLADQYPHLHNAELSKTLGKLW**
KLLNDSEKKPFIEEAERLRIKHKREHPDYKYQPRKKKQKGNGNGDAGDATISADDLLKVLKGDSKLVPNNGDASASCASPESVSDGEVSSSESCSVPSPETP
TAVPVKNEDVKNDEALSAQPGFPSCSKKDDSNSHAIDFDVGDLTTDLMAMGDVDST**EFDQYL**PTYSQALLDSTLTKAINTTQINTQSLSNSR**FTTSQAVQSP**
PPLPSSYREFMVQLQKLPSEGSFPPNRVAPSATMQQRNQPDSNSFFPFSESEVNVCSSLAATRQPAFLSSPSTSGTLSSSSNSGRTHTLIWK

Color code: HMG domain, EΦDQYΦ motif, PQA domain & flanking residues, TAC domain

D

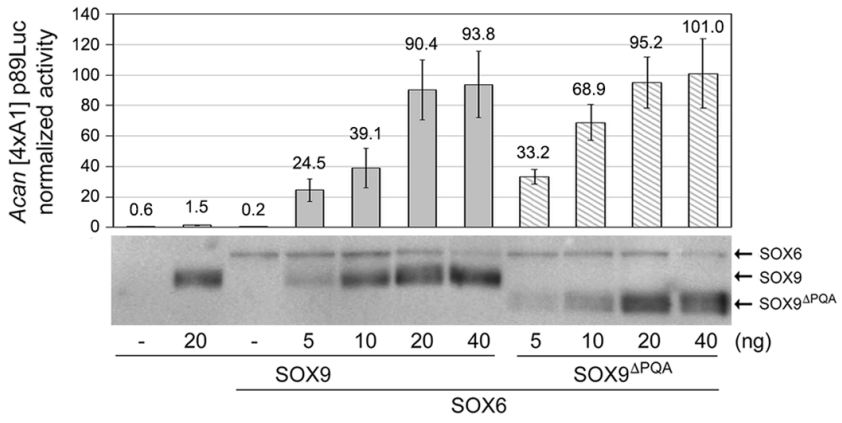


Figure S2. Analysis the SOX9 PQA domain. (**A**) ClustalW alignment of the PQA domains (boxed) and flanking residues in SOX9 orthologs from representative species in the vertebrate subphylum: human (*Homo sapiens*), mouse (*Mus musculus*), dog (*Canis lupus familiaris*), platypus (*Ornithorhynchus anatinus*), chicken (Gallus gallus), American alligator (*Alligator mississippiensis*), African clawed frog (*Xenopus laevis*), coelacanth (*Latimeria chalumnae*), zebrafish (*Danio rerio*; the protein encoded by *sox9a* was used), and lamprey (SOXE3; *Petromyzon marinus*). (**B**) Length of the PQA domains shown in panel A and numbers of specific residues. (**C**) Comparison of human SOX9 and invertebrate SOXE sequences. Protein domains are colored, as indicated. P, Q, and A residues are bolded in human SOX9 PQA and in invertebrate SOXE segments that were aligned with human SOX9 PQA by the ClustalW tool. (**D**) Test of the ability of SOX9 lacking PQA to transactivate an *Acan* [4xA1]-p89Luc reporter in synergy with SOX6. The reporter contains 4 copies of a cartilage-specific *Acan* enhancer (A1) upstream of a *Col2a1* minimal promoter and the firefly luciferase gene. SW-1353 cells were transfected with the reporter, a control plasmid to assess transfection efficiency, a SOX6 or empty expression plasmid, and increasing amounts of SOX9 or $SOX^{\Delta PQA}$ plasmid, as indicated. The western blot of cell lysates shows that deletion of PQA did not affect the production or stability of the SOX9 protein.

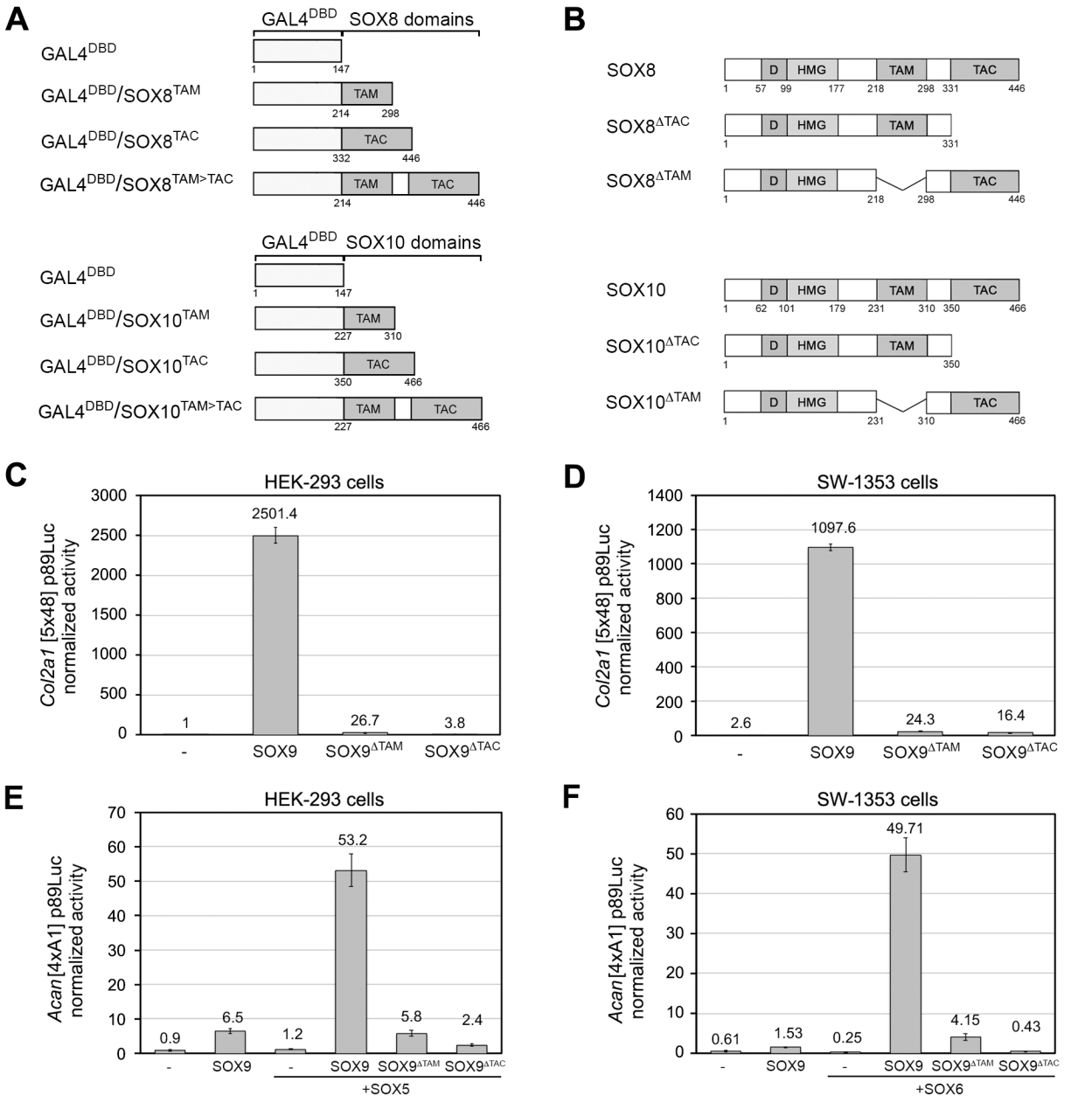


Figure S3. Analysis of the activities of SOXE TAM and TAC domains in transactivation. (**A**) Schematics showing fusion proteins of GAL4^{DBD} and the TAM, TAC and TAM-to-TAC domains of SOX8 and SOX10 used in Figure 3B and C. (**B**) Schematics of SOX8 and SOX10 wild-type proteins and deletion mutants lacking TAM or TAC used in Figure 3D and E. (**C** to **F**) Test of the ability of SOX9 deletion mutants to activate a *Col2a1* or *Acan* reporter in HEK-293 and SW-1353 cells. The SOX9 expression plasmids encoded the proteins shown in Figure 3D. The data presented in panel C are the same as in the middle panel of Figure 3E. They are shown to facilitate their comparison to those shown in panels D to F. Activation of the *Acan* reporter was tested in the presence of an expression plasmid for SOX5 or SOX6 (panels E and F).

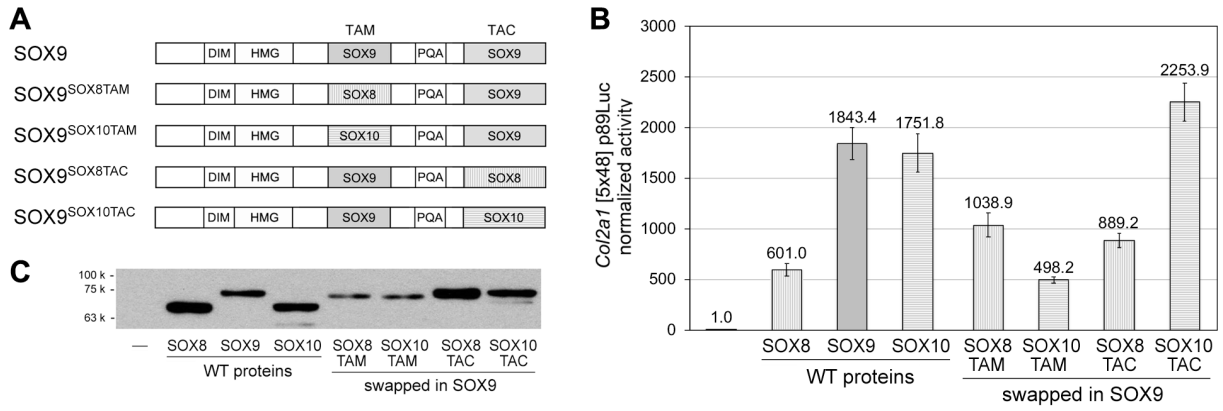


Figure S4. Effect of swapping SOX9^{TAM} and SOX9^{TAC} with the corresponding SOX8 and SOX10 domains on SOX9 activity. (**A**) Schematics of the SOX9 wild-type protein and proteins in which the TAM or TAC domain of SOX9 was exchanged with the equivalent domain from SOX8 or SOX10. (**B**) Reporter assay comparing the abilities of the wild-type SOXE proteins and SOX9 chimeric proteins to activate the *Col2a1* reporter in HEK-293 cells. Reporter activities are presented as the mean ± standard deviation obtained for triplicate cultures per condition. Data were normalized for transfection efficiency and are reported as fold increase relative to the activity of the reporter in the presence of empty expression plasmids. These results were reproduced in multiple experiments. (**C**) Western blot showing the relative amounts of SOX proteins present in the cells at the end of the culture period. The blots were made with lysate amounts normalized for transfection efficiency. Note that differences in the relative amounts of proteins may contribute to explain the lower activities of the SOX9 proteins harboring the SOX8 and SOX10 TAM domains compared to wild-type SOX9, but not the difference in activities between these two mutants.

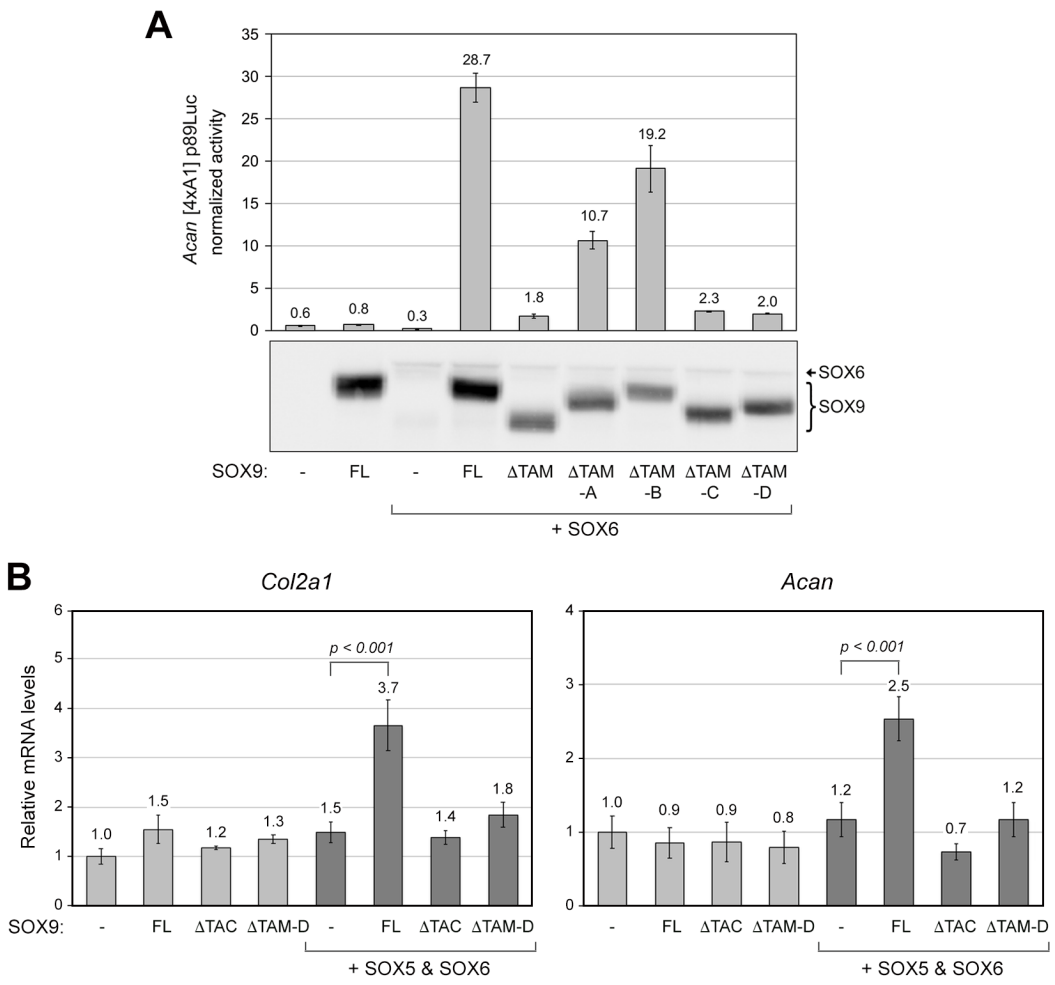


Figure S5. Importance of TAM subdomains for SOX9 transactivation. (**A**) HEK-293 cells were transiently transfected with the *Acan* reporter and with expression plasmids encoding no SOX protein (-), wild-type SOX9 (FL), SOX9 deletion mutants (schematized in Figure 3D), and SOX6, as indicated. Reporter activities are presented for a representative experiment as described in other figures. The western blot of cell lysates shows that the lower activities of the SOX9 proteins lacking TAM-A and TAM-B may be explained at least in part by lower amounts of these proteins compared to wild-type SOX9. (**B**) Importance of TAC and TAM-D in SOX9 transactivation of the endogenous *Col2a1* and *Acan* genes. ATDC5 cells were transiently transfected with expression plasmids for no SOX protein (-), wild-type SOX9 (FL), SOX9 lacking TAC (ΔTAC) or SOX9 lacking TAM-D (ΔTAM-D), and SOX5 and SOX6. The relative levels of *Col2a1* and *Acan* mRNA, as assessed by qRT-PCR 24 h after the start of transfection, are presented as fold increases over the levels present in cells transfected with the empty expression plasmid. Each value is presented as the mean ± standard deviation obtained for four cell culture replicates generated in two separate experiments. The Student's T-test was used to assess the statistical significance of differences in mRNA levels among all conditions. The only differences that reached a high degree of significance ($p < 0.05$) are indicated. The actual p value for these differences was < 0.001.

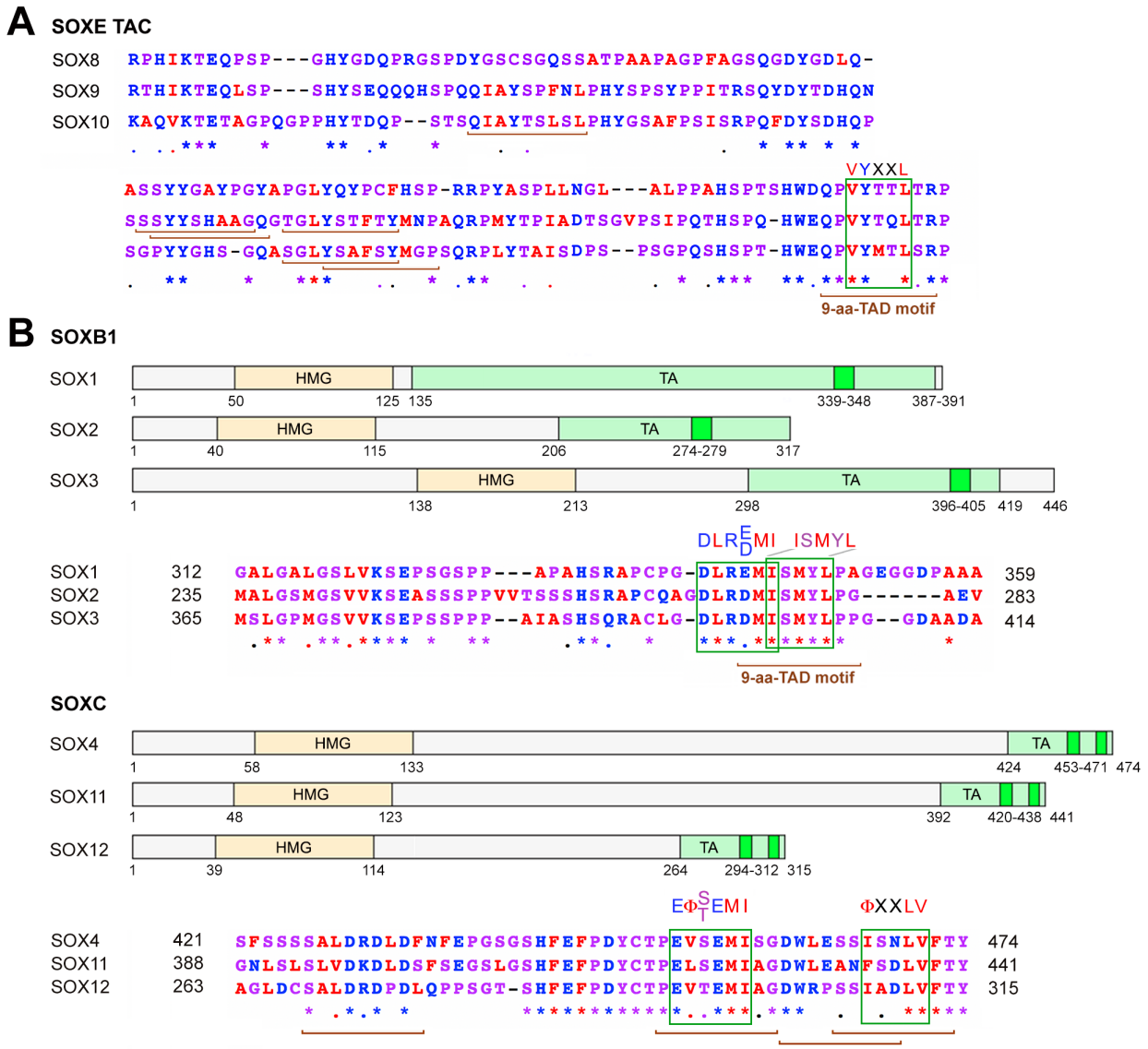


Figure S6. Identification of 9-aa-TAD, ΦXXΦΦ-like, and EΦ[D/E]QYΦ-like motifs in SOX protein transactivation domains. (**A**) Alignment of the human SOXE TAC sequences showing residue conservation and the position of 9-aa-TAD motifs (brown brackets) and an ΦXXΦΦ-like motif (green box). The consensus sequence for this ΦXXΦΦ-like motif is written above the sequence alignment. (**B**) Schematics showing the locations of the HMG and transactivation regions (TA, pale green boxes) of SOXB1 and SOXC proteins and alignments of TA regions that show a high degree of conservation among group members. 9-aa-TAD motifs and ΦXXΦΦ/EΦ[D/E]QYΦ-like motifs are indicated. The TA regions are shown as previously delineated for SOX1 (1), SOX2 (2), SOX3 (3) and SOXC (4).

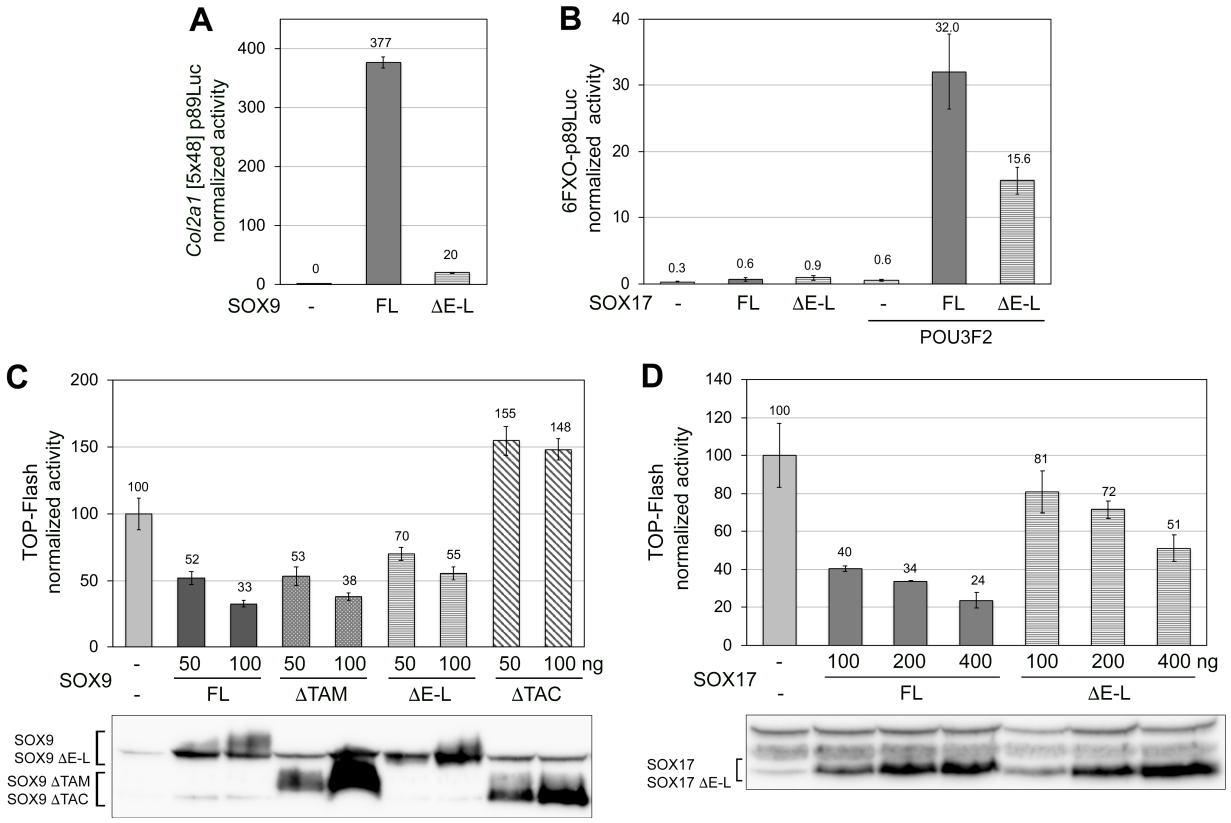


Figure S7. Importance of the EΦ[D/E]QYΦ motif in SOX9 and SOX17 activities. (**A**) HEK-293 cells were transfected with the *Col2a1* reporter and with expression plasmids for no SOX protein (-), wild-type SOX9 (FL), or SOX9 lacking the EFDQYL motif (ΔE-L). Reporter activities are presented for a representative experiment as described in other figures. (**B**) HEK-293 cells were transfected with a *6FXO-p89Luc* reporter and with plasmids encoding no protein (-), wild-type SOX17 (FL), SOX17 lacking the EFEQYL motif (ΔE-L), or POU3F2. The *6FXO-p89Luc* reporter contains 6 copies of an *Fgf4* enhancer featuring adjacent binding sites for SOX and POU-domain proteins. Reporter activities are presented for a representative experiment as described in other figures. (**C**) HEK-293 cells were transfected with the TopFlash reporter, a plasmid encoding constitutively stabilized β-catenin, and plasmids encoding no protein, wild-type SOX9 or SOX9 lacking TAM, EFDQYL or TAC. SOX9 plasmids were tested at 50 and 100 ng. Reporter activities are presented in percentages of the activity of β-catenin in the absence of SOX protein. Values are the mean ± standard deviation obtained for triplicate cultures in one experiment representative of three independent ones. The western blot shows that all SOX9 proteins were made and that the slightly stronger ability of SOX9$^{\Delta TAM}$ than SOX9$^{\Delta E\text{-}L}$ to inhibit β-catenin is likely due to its larger amount. (**D**) HEK-293 cells were transfected with the TOP-Flash reporter, an expression plasmid for constitutively stabilized β-catenin, and expression plasmids for no SOX protein (-), wild-type SOX17 (FL) or SOX17 lacking EFEQYL. The SOX17 plasmids were tested at 100, 200 and 400 ng. Reporter activities are presented as in panel C. The western blot shows that the SOX17 wild-type and mutant proteins were made in similar amounts and thus that the weaker ability of the mutant to inhibit β-catenin is genuinely due to the EFEQYL deletion.

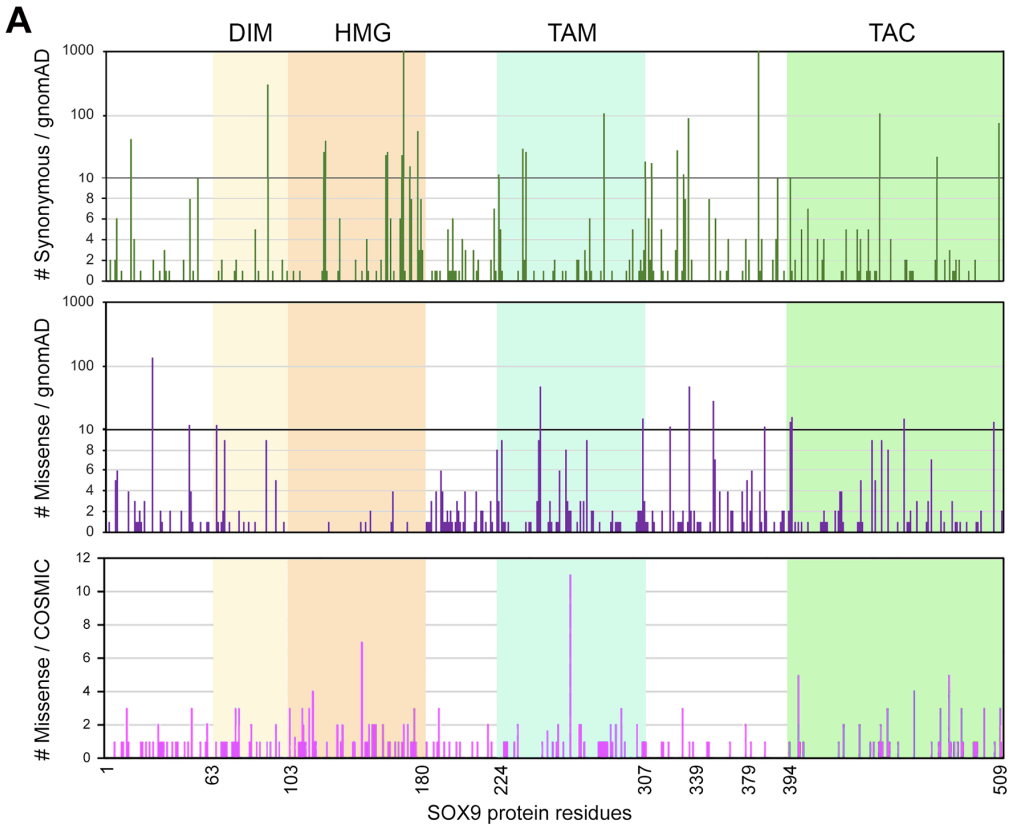

A
DIM
HMG
TAM
TAC
Synonymous / gnomAD
Missense / gnomAD
Missense / COSMIC
SOX9 protein residues
B
% Residues with at least one variant
p < 0.01
p < 0.01
p < 0.01
DIM
HMG
TAM
TAC
Other
Synonymous / gnomAD
Missense / gnomAD
Missense / COSMIC
C
Missense / COSMIC
α-helix 1
α-helix 2
TAM-C
TAM-D
D
Col2a1 [5x48] p89Luc normalized activity
100.0
176.1
178.5
31.3
12.2
63.9
WT
I283F
S284F
E293G
E293K
D295Y

Figure S8. Analysis of missense variants detected in SOX9 in control individuals and in cancer samples and effect of missense variants reported in cancers on the transcriptional activity of SOX9. (**A**) Bar graphs showing the numbers of synonymous variants (top) and missense variants (middle) reported in control individuals in the gnomAD database, and missense variants reported in cancers in the COSMIC database (bottom) throughout the human SOX9 protein. (**B**) Percentages of residues with at least one variant in each SOX9 domain. "Other" refers to the SOX9 sequences outside the known functional domains. This graph used the same data as in panel A. Statistically significant differences between datasets were calculated using the Student's T test. Brackets link data for which the p value was < 0.01. (**C**) Numbers and types of missense variants reported for the SOX9 TAM-CD region in COSMIC. The five variants tested in panel D are boxed. Stars mark residues that have missense variants in the gnomAD cohort. (**D**) *Col2a1* reporter activities achieved in HEK-293 cells by SOX9 proteins harboring a subset of COSMIC missense variants. Values are presented as percentages of the activities measured for wild-type SOX9. Each value is the mean ± standard deviation obtained for triplicate cultures in one representative experiment. Similar results were obtained in multiple experiments.

SUPPLEMENTARY REFERENCES

1. Kamachi, Y., Cheah, K.S. and Kondoh, H. (1999) Mechanism of regulatory target selection by the SOX high-mobility-group domain proteins as revealed by comparison of SOX1/2/3 and SOX9. *Mol Cell Biol*, **19**, 107-120.
2. Nowling, T.K., Johnson, L.R., Wiebe, M.S. and Rizzino, A. (2000) Identification of the transactivation domain of the transcription factor Sox-2 and an associated co-activator. *J Biol Chem*, **275**, 3810-3818.
3. Xia, Y., Papalopulu, N., Vogt, P.K. and Li, J. (2000) The oncogenic potential of the high mobility group box protein Sox3. *Cancer Res*, **60**, 6303-6306.
4. Dy, P., Penzo-Mendez, A., Wang, H., Pedraza, C.E., Macklin, W.B. and Lefebvre, V. (2008) The three SoxC proteins--Sox4, Sox11 and Sox12--exhibit overlapping expression patterns and molecular properties. *Nucleic Acids Res*, **36**, 3101-3117.