

# The three SoxC proteins—Sox4, Sox11 and Sox12—exhibit overlapping expression patterns and molecular properties

Peter Dy<sup>1</sup>, Alfredo Penzo-Méndez<sup>1</sup>, Hongzhe Wang<sup>1</sup>, Carlos E. Pedraza<sup>2</sup>,  
Wendy B. Macklin<sup>2</sup> and Véronique Lefebvre<sup>1,\*</sup>

<sup>1</sup>Department of Cell Biology and Orthopaedic Research Center and <sup>2</sup>Department of Neurosciences, Lerner Research Institute, Cleveland Clinic, Cleveland, OH, USA

Received December 6, 2007; Revised March 1, 2008; Accepted March 20, 2008

## ABSTRACT

The group C of Sry-related high-mobility group (HMG) box (Sox) transcription factors has three members in most vertebrates: Sox4, Sox11 and Sox12. Sox4 and Sox11 have key roles in cardiac, neuronal and other major developmental processes, but their molecular roles in many lineages and the roles of Sox12 remain largely unknown. We show here that the three genes are co-expressed at high levels in neuronal and mesenchymal tissues in the developing mouse, and at variable relative levels in many other tissues. The three proteins have conserved remarkable identity through evolution in the HMG box DNA-binding domain and in the C-terminal 33 residues, and we demonstrate that the latter residues constitute their transactivation domain (TAD). Sox11 activates transcription several times more efficiently than Sox4 and up to one order of magnitude more efficiently than Sox12, owing to a more stable  $\alpha$ -helical structure of its TAD. This domain and acidic domains interfere with DNA binding, Sox11 being most affected and Sox4 least affected. The proteins are nevertheless capable of competing with one another in reporter gene transactivation. We conclude that the three SoxC proteins have conserved overlapping expression patterns and molecular properties, and might therefore act in concert to fulfill essential roles *in vivo*.

## INTRODUCTION

The family of Sry-related high-mobility group (HMG) box (Sox) transcription factors counts 20 members in most vertebrates (1,2). Sry, which was first discovered, owns its acronym to its critical role as the sex-determining region on the Y chromosome (3). Its main

feature is a DNA-binding domain similar to that originally identified in a HMG of chromatin-associated nonhistone proteins (4,5). This domain is therefore referred to as the HMG box. Sox proteins distinguish themselves in the super-family of HMG box proteins by sharing 46% or more identity to Sry in the HMG box. The family is subdivided into eight groups, A–H (6). Members of the same group are highly similar to each other within and outside the HMG box, whereas members of different groups share a lower degree of identity in the HMG box and no significant identity outside of this domain. All Sox genes characterized so far have a specific expression pattern and most have critical roles in the determination of cell fate and differentiation into specific lineages. These lineages include, but are not limited to, embryonic stem cells, neuronal and glial cells, Sertoli cells, chondrocytes, endothelial cells and erythroid cells. Mutations in *SOX* genes cause severe congenital diseases in humans, such as XY sex reversal (*SRY*), campomelic dysplasia (*SOX9*), Waardenburg–Hirschsprung syndrome (*SOX8*) and anophthalmia–esophageal–genital syndrome (*SOX2*). The Sox domain contacts the minor groove of DNA to AACAAAG, AACAAAT and related sequences, and induces a sharp bend of DNA. It is believed thereby to allow the proteins to contribute an architectural key role in the assembly of transcriptional enhancer complexes. It also interacts with various types of transcription factors to increase its efficiency and specificity of action. In addition to the HMG box domain, most Sox proteins feature one or two other functional domains, such as a transactivation, transrepression or homodimerization domain, and thus control transcription in several ways.

The SoxC group is composed of a single member in *Drosophila melanogaster* (7) and lower animal species, but has three members in mice, humans and most other vertebrates: Sox4, Sox11 and Sox12. The three proteins share a high degree of identity both in the HMG box domain and in the C-terminal region (8–11). Sox4 preferentially binds the AACAAAG sequence in

\*To whom correspondence should be addressed. Tel: +1 216 445 0762; Fax: +1 216 444 9404; Email: lefebvv@ccf.org

electrophoretic mobility shift assay (EMSA) and is capable of transactivating a reporter construct harboring a multimer of this sequence (8). The HMG box is located in the N-terminal third and the transactivation domain (TAD) is located within the protein C-terminal third, but has not been precisely mapped yet. Sox11 has similar structural and functional properties, but is a more potent transactivator than Sox4 in transient transfection assays (9). It also harbors an internal acidic domain that appears responsible for blocking its ability to bind DNA in EMSAs and for reducing it in cells (12). Sox12, reported to as SOX22 in humans, is structurally similar to Sox4 and Sox11, but its functional properties have not been reported (13). Altogether, these data are still rudimentary but already suggest that the three SoxC proteins could exhibit similar functional properties. A precise delineation and detailed characterization of their functional domains and a comparison of their relative activities are needed to test this possibility.

The expression pattern of each SoxC gene has been described in several studies, but always independently of the other group members. Sox4 is expressed in adult mouse thymus, heart, lung, pancreas and gonads (8), in embryonic neural progenitors (14,15), pancreatic islets (16), endocardial cushions (17), terminal chondrocytes and bone primary spongiosa (18). Sox11 is widely expressed during organogenesis in the mouse embryo, and is highly expressed in the human and mouse fetus in the central and peripheral nervous system, lung, gastrointestinal tract, pancreas, spleen, kidneys, gonads, branchial arches and mesenchyme (10,19,20). Similarly, SOX12 is expressed in neural tissue and mesenchyme in human embryos, and in brain, lung, heart, liver, spleen, thymus, pancreas and kidneys in adults (13). Expression of Sox4 and Sox11 was also shown to be upregulated in various types of cancers (21–23). These independent studies strongly suggest major overlap in expression of the three SoxC genes, but a direct comparison of their expression patterns is needed not only to confirm this possibility, but also to assess their relative expression levels in sites of co-expression.

While no studies have yet been reported on the roles of Sox12 *in vivo*, several studies have focused on characterizing roles for either Sox4 or Sox11 *in vivo* and have started to uncover important roles for these two genes in various developmental, physiological and pathological processes. Sox4 was shown upon gene inactivation or misexpression in mice or chickens to have essential roles in heart outflow tract and valve development (17), T and B lymphocyte differentiation (24), endocrine islet formation (25), osteoblast development (26) and neural and glial cell development (15,27,28). Upregulation of SOX4 was shown to promote survival and proliferation of several types of cancer cells (22). Sox11 inactivation in the mouse was shown to cause heart arterial outflow tract malformation, lung hypoplasia, asplenia, cleft lip and palate, open eyelids and skeletal malformations (20). Only one study so far has addressed the possibility of functional interaction between the SoxC genes *in vivo*. Using RNA interference and cDNA misexpression in the chicken embryo, this study has demonstrated essential interaction (likely redundancy)

between Sox4 and Sox11 in the differentiation of neuronal progenitors (15). It has shown that Sox4 and Sox11 are needed for the expression of pan-neuronal genes, including the class-III  $\beta$ -tubulin gene *Tubb3* (also named *Tuj1*). This gene was proposed to be a direct target of the SoxC proteins by showing that forced expression of Sox4 and Sox11 in Cos1 cells resulted in upregulation of *Tubb3* reporter constructs, and that this effect required *cis*-acting elements that the proteins were able to bind to in EMSAs. *Tubb3* is thereby the first and still only gene to be identified as a potential direct target of SoxC proteins.

Collectively, the expression pattern and functional data currently available thus strongly suggest that the SoxC proteins are important transcription factors in many processes. Further studies are needed to be able to enumerate the processes that they control, identify their exact roles and evaluate their degree of possible redundancy or other form of interaction. The present study was designed to address several of these questions. We have directly compared the expression patterns of the three SoxC genes in the mouse and we have analyzed structural and functional properties of their protein products. We show that the three genes are co-expressed in many tissues in the developing mouse, but that their relative expression levels are highly variable. The three proteins are remarkably conserved in the HMG box domain and in the C-terminus, and the latter domain is their TAD. The three proteins nevertheless exhibit differential DNA-binding and transactivation capabilities, and are thereby able to functionally interact with each other both positively and negatively. Our data thus strongly support the notion that the three genes may work in concert to fulfill critical roles in a multitude of *in vivo* processes.

## MATERIALS AND METHODS

### RNA assays

Total RNA was extracted from E12.5 whole mouse embryos, newborn mouse tissues and cultured cells using TriZol (Invitrogen, Carlsbad, CA), as recommended by the manufacturer. Cos1 and MC3T3-E1 cells were cultured in monolayer under standard conditions in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum (FCS). Mouse primary neuronal precursors were prepared from E14.5 mouse embryo cerebrum and cultured in neurospheres as previously described (29,30). Northern blots were prepared according to a standard protocol. Transcript sizes were estimated by comparison with a ladder of DNA markers (Invitrogen). The Sox4 probe was a BglII/SmaI 634-bp fragment of the coding sequence (positions 925–1558 in the NM\_009238 cDNA sequence) (31). The Sox11 probe was a PstI/XhoI 591-bp fragment of the coding sequence (positions 611–1198 in the NM\_009234 cDNA sequence). The Sox12 probe was a 685-bp fragment of the 3' untranslated region (positions 1585–2270 in the NM\_011438 cDNA sequence). *In situ* hybridization was carried out on mouse embryo sections with [<sup>35</sup>S]-rUTP-labeled RNA probes (32). Cell nuclei were counterstained with the Hoechst 33258 DNA-binding dye. RNA probe

signals were photographed under dark-field exposure using a red filter and blue fluorescence. Staining of adjacent sections with hematoxylin and eosin was performed according to a standard procedure.

### Protein sequence analysis

Sequences of SoxC proteins from various vertebrate species were downloaded from the NCBI website (<http://www.ncbi.nlm.nih.gov>). Accession numbers were as follows: *Homo sapiens* (human) SOX4: NP003098, SOX11: NP003099, SOX12: NM006943; *Mus musculus* (mouse) Sox4: NP033264, Sox11: NP033260, Sox12: NM 011438; *Bos taurus* (cow) Sox4: NP001071596, Sox11: NP 990518; *Monodelphis domestica* (opossum) Sox4: hmm10497, Sox11: XP001372169, Sox12: XP001365568; *Gallus gallus* (chicken) Sox4: NP989815, Sox11: NP990518; *Xenopus tropicalis* (clawed frog) Sox4: NP001039084, Sox11: NP001008053; *Takifugu rubripes* (pufferfish) Sox4: AAQ18501, Sox11: AAQ18502, Sox12: AAQ18530 and *Canis familiaris* (dog) Sox12: XP542944. These sequences were aligned with ClustalX software (<http://www-igbmc.u-strasbg.fr/BioInfo/ClustalX>) (33) using default settings. Secondary structure predictions were obtained using the MLRC webware tool (<http://pbil.ibcp.fr>) (34).

### Expression plasmids

Mouse SoxC and Brn2 expression plasmids were generated essentially as previously described for human SOX9 and mouse Sox6 (35,36). Briefly, the full-length coding sequence of each protein was amplified by PCR using mouse 129xB6 genomic DNA. The PCR products were cloned in the pCR4-TOPO vector (Invitrogen) and verified by DNA sequencing. They were then cloned in the pcDNA3.1+ mammalian expression plasmid (Invitrogen) with or without an in-frame N-terminal FLAG epitope sequence. SoxC mutant sequences were generated by PCR using appropriate primers. Plasmids encoding GAL4/SoxC fusion proteins were generated by cloning SoxC segments downstream and in frame with the GAL4 DNA-binding domain in the pBIND plasmid (Promega, Madison, WI).

### Reporter plasmids

The 6FXO-p89Luc reporter was constructed by cloning six tandem copies of the FXO sequence (37) in the previously described p89Luc plasmid (38). This plasmid featured a minimum (89 bp) *Col2a1* promoter driving the firefly luciferase gene. Tandem copies of the FXO sequence were generated by ligating oligonucleotides corresponding to the FXO sequence and flanked with BamHI and BglII restriction sites at the 5' and 3' end, respectively, with each other. Multimers were cloned in p89Luc and verified by sequencing. The 4x48-p89Luc reporter was as described (38). The 2HMG-p89Luc reporter was constructed by multimerizing the 2HMG probe and as described 6FXO-p89Luc (36). The *Tubb3*-pLuc reporter was constructed by cloning a PCR-amplified -215/+54 mouse *Tubb3* promoter into the promoter-less pA3Luc plasmid (38). The GAL4 reporter pG5Luc plasmid contained five copies of the GAL4

binding site upstream of a minimal promoter driving the firefly luciferase gene (Promega). The pSV2-lacZ plasmid, in which the SV40 promoter and enhancer drove *lacZ* expression, was used to determine transfection efficiency (39).

### Transient transfection

Cos1 and MC3T3-E1 cells were plated at a density of  $3.10^5$  cells/10 cm<sup>2</sup> dish in 2 ml DMEM supplemented with 10% FCS and antibiotics. Transfection mixtures were added to the culture medium 3–6 h later. These mixtures consisted of 1 µg total plasmid, 3 µl FuGENE6 (Roche, Indianapolis, IN) and 100 µl DMEM. Neurospheres were transfected at the passage 2–6 by electroporation with the Amaxa system according to the manufacturer's instructions for mouse neural stem cells (Amaxa, Gaithersburg, MD). Unless otherwise indicated, the plasmid mixtures of all transfections consisted of 300 ng pSV2-lacZ control plasmid, 600 ng luciferase reporter plasmid and 200 ng expression plasmid. Expression plasmids contained various proportions of empty pcDNA3.1+ vector and Sox or Brn2 expression vectors, as indicated in figures. Cells were harvested 24 h after transfection and whole-cell extracts were made in 150 µl buffer containing 14 mM HEPES (pH 7.9), 1.5 mM MgCl<sub>2</sub>, 6.0 mM KCl, 0.44 M NaCl, 0.08 mM EDTA, 2.3 mM DTT, 0.5 mM PMSF, 10 µg/ml leupeptin, 10 µg/ml pepstatin A, 4.6% Ipegal and 10% glycerol. Luciferase and β-galactosidase activities were assayed in these extracts using the Dual-Light combined reporter gene assay system (Applied Biosystems, Foster City, CA) and a Wallac Victor 1420 microplate reader (Perkin-Elmer, Waltham, MS). Reporter gene activities are presented in figures in relative luciferase units (RLU). They were obtained by multiplying luciferase activity units by 10<sup>3</sup> and dividing them by β-galactosidase activity units. Figures show averages with standard deviation of three independent assays in one representative experiment.

### Western blot analysis

Western blot of whole-cell extracts (prepared as described earlier) were performed according to a standard protocol. Dual-Color standards were used as protein Mr markers (Amersham, Piscataway, NJ). Proteins harboring a FLAG epitope were detected using a mouse monoclonal anti-FLAG M2-peroxidase HRP antibody (Sigma, St. Louis, MO, A8592). As indicated in figure legends, SoxC proteins were also detected using rabbit polyclonal antibodies directed against a 14-amino-acid sequence fully conserved in all SoxC HMG box domains (Sigma, S8068; sold as Sox11 antibody), and goat HRP-linked anti-rabbit IgG secondary antibodies. GAL4/SoxC fusion proteins were detected using a mouse monoclonal HRP conjugate antibody against the GAL4 DNA-binding domain (Santa Cruz, Santa Cruz, CA, sc-510). Antibody hybridization signals were detected using the ECL Western Blotting Analysis System (Amersham).

### EMSA

Double-stranded oligonucleotide probes were synthesized with the addition of three Gs in 5' overhang on



each strand and they were labeled with  $\alpha$ - $^{32}\text{P}$ -dCTP by Klenow filling-in. Assay mixtures were made in DNA-binding buffer (20 mM HEPES, pH 7.9, 40 mM KCl, 0.5 mM EDTA, 0.5 mM DTT, 1 mM PMSF, 10% glycerol) with 10 fmol labeled probe, 1–3  $\mu\text{l}$  whole-cell extracts and 0.25–1  $\mu\text{g}$  poly(dG-dC)-poly(dG-dC). They were incubated for 30 min at 30°C followed by electrophoresis in 4% polyacrylamide native gels in 0.5 $\times$  TGE buffer (50 mM Tris, 0.4 M glycine, 4.5 mM EDTA, pH 8.0). Dried gels were exposed to X-ray films for 1 h to overnight.

## RESULTS

### Overlapping expression of the *SoxC* genes *in vivo*

We compared the expression patterns of the three *SoxC* genes in the mouse by northern blot and RNA *in situ* hybridization. To estimate the relative expression levels of the genes, we used RNA probes that exhibited similar annealing temperatures and labeling specificities. Northern blot of total RNA from whole embryos demonstrated that the probes recognized specific transcripts (Figure 1A). *Sox4* was expressed predominantly as a ~5 kb RNA, *Sox11* as a ~7 kb RNA and *Sox12* as ~3 and 3.5 kb RNAs. The comparison of tissue RNAs from 2-day-old mice showed that all three genes were expressed at a very high and similar level in the brain (Figure 1B). *Sox4* and *Sox12* were also expressed at a high level in a distinct set of organs, including the heart and the lung, and at a low or insignificant level in most other tissues, such as the liver, pancreas and small intestine. *Sox11*, in contrast, was hardly detected in nonneural tissues at that age. Hybridization of RNA from cultured cells showed high expression of the three genes in neuronal precursor cells isolated from mouse embryo cerebrum and maintained in primary culture as neurospheres (40,41). *Sox4* and *Sox11* but not *Sox12* were also expressed in kidney Cos1 fibroblasts and in mesenchymal preosteoblastic MC3T3-E1 cells (42).

*In situ* hybridization of RNA in mouse embryo sections from E10.5 to E18.5 showed co-expression of the three *SoxC* genes in most areas of the central and peripheral nervous system, including the brain, neural tube, retina, olfactory epithelium, cochlear epithelium and dorsal root ganglia (Figure 1C and D; data not shown). The three genes were also co-expressed in the mesenchyme of many presumptive and developing organs, such as skeletogenic primordia and genital tubercle, and in both the epithelium and the mesenchyme of the developing lung, kidney and midgut. Several sites of differential expression were observed. For instance, the eyelid primordium and palatal shelf mesenchyme expressed *Sox11* and *Sox12*, but not *Sox4*; the developing heart endocardial cushions expressed *Sox4* and *Sox12*, but not *Sox11*; and the developing teeth, spleen, thymus and hair follicles predominantly expressed *Sox4* (Figure 1C–E).

The three *SoxC* genes are thus co-expressed at high levels in neuronal and mesenchymal tissues and cells *in vivo* and *in vitro*, and at distinct, variable levels in many other tissues *in vivo*. These results prompted us to

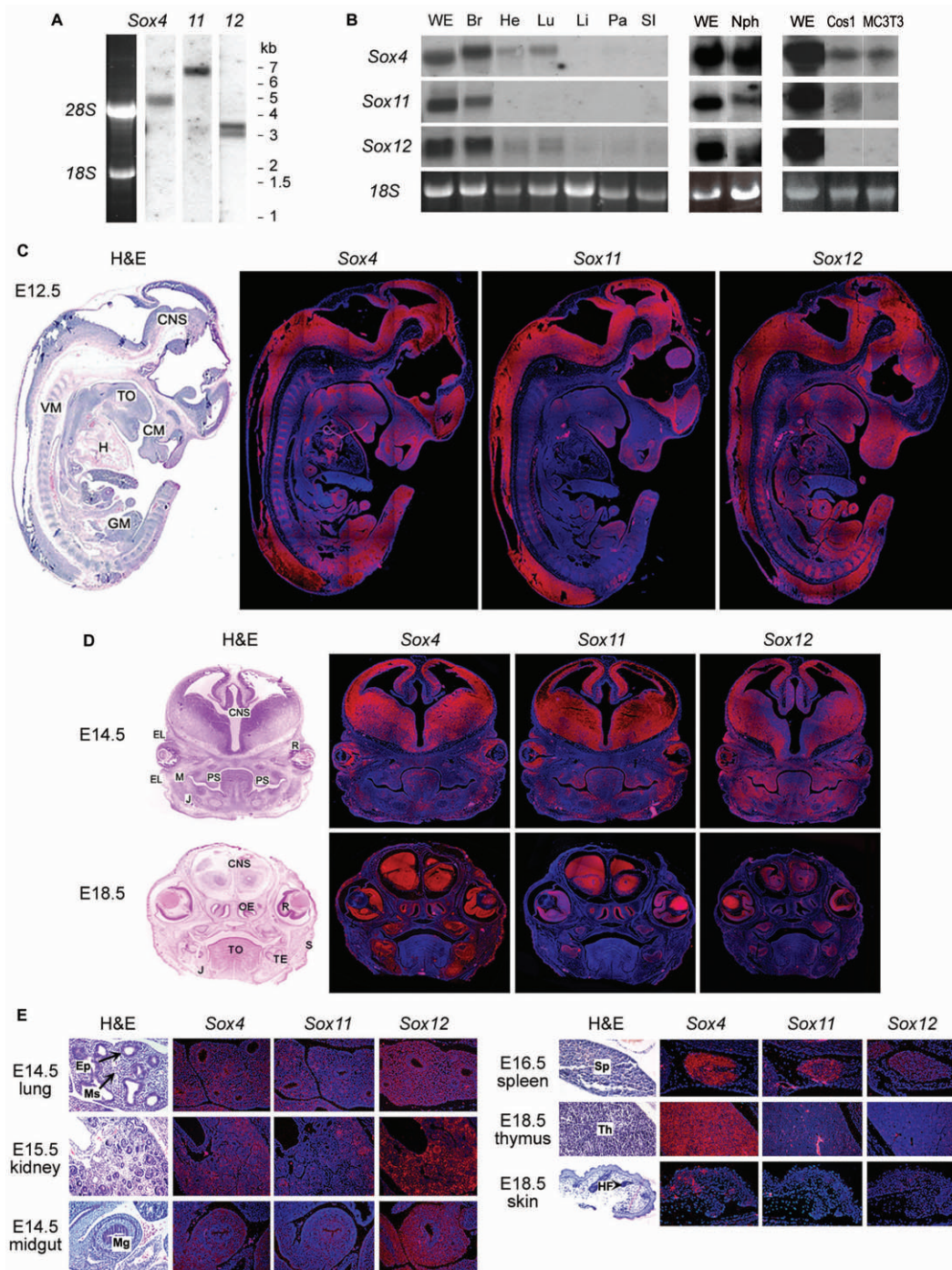
determine to which extent the three *SoxC* proteins share molecular properties, a prerequisite to understand their mode of possible functional interaction in sites of co-expression.

### The *SoxC* proteins share a highly conserved C-terminus

We downloaded 18 *SoxC* protein sequences from seven vertebrate species, from fish to human, and aligned them with ClustalX software to identify evolutionary conservation (Figure 2A). The results were clear-cut: the three *SoxC* proteins share the three N-terminal residues, most residues constituting and located immediately upstream of the HMG box domain, most of the 33 C-terminal residues and a few upstream residues, but no other stretch of contiguous residues. Orthologous proteins, nonetheless, share significant conservation in other regions, such as acidic and serine-rich domains (Supplementary Figure 1). Close inspection of the HMG box domains revealed 84% identity and 95% similarity between all *SoxC* proteins (Figure 2B), but only 45–67% identity to Sry and the other *Sox* proteins (unpublished results). All *SoxC* proteins share 67% identity and 94% similarity in the C-terminal 33 residues, hereafter referred to as the C33 domain. This domain is thus almost as highly conserved as the HMG box domain. It has no significant identity, however, with sequences in other proteins (unpublished results). Interestingly, the HMG box and C33 domains are more highly conserved among *Sox4* orthologues and among *Sox11* orthologues than among *Sox12* orthologues. They are also more highly conserved between *Sox4* and *Sox11* orthologues than between these two types of orthologues and the *Sox12* orthologues. Almost half of the C33 domain sequence is constituted of proline, serine, and acidic residues, as typically seen in TADs (Figure 2C). Protein secondary structure models predict that 20 residues in the *Sox11* C33 domain (C23–C3) form a noninterrupted  $\alpha$ -helix. A similar helix is predicted to form in *Sox4* and *Sox12*. However, the *Sox4* helix is predicted to be interrupted in the middle of its sequence by three randomly coiled residues (C16–C14), and to have the last two residues (C4–C3) in extended conformation rather than in helical conformation. The *Sox12* helix is predicted to be even more discontinuous, with seven randomly coiled residues (C15–C7) in the middle of its sequence. The vertebrate *SoxC* proteins have thus tightly conserved both the HMG box domain and the C33 domain. Each orthologue has nevertheless evolved to exhibit specific features, which suggest important, protein-specific functions.

### The *SoxC* proteins transactivate with different efficiencies

We compared the transactivation potential of the mouse *SoxC* proteins by transiently transfecting Cos1 cells with expression plasmids for these proteins and with several reporter plasmids (Figure 3A). Three reporters featured a minimal 89 bp *Col2a1* promoter (p89) driving the firefly luciferase gene (38). This promoter contains a TATA box and GC-rich elements, but no other binding sites for known transcription factors, and it was previously shown to be virtually inactive in transient



**Figure 1.** Comparison of the *SoxC* gene expression pattern in the mouse. (A) Northern blots of total RNA from E12.5 mouse embryos hybridized with Sox4, Sox11 and Sox12 probes, as indicated. A picture of the RNA stained with ethidium bromide before blotting is shown in the left lane. The migration level and size of DNA markers run on the same gel are indicated on the right. The three probes had the same labeling specificity and the three blots were exposed for the length of time to allow direct comparison of RNA levels. (B) Northern blots showing the relative expression levels of the *SoxC* RNAs in various tissues of 2-day-old mice and in cultured cells. WE, whole embryo at E12.5; Br, brain; He, heart; Lu, lung; Li, liver; Pa, pancreas; SI, small intestine; Nph, mouse embryo primary neurospheres; Cos1, monkey kidney fibroblastic cell line; MC3T3, preosteoblastic MC3T3-E1 cell line. The three probes had the same labeling specificity and the blots in each of the three sub-panels were exposed for the same length of time to allow direct comparison of RNA levels. The size of the *Sox11* RNA in monkey Cos1 cells was ~5 kb, instead of ~7 kb in mouse tissues and cells (unpublished results). (C) RNA *in situ* hybridization of adjacent mid-sagittal sections of E12.5 mouse embryos with *Sox4*, *Sox11* and *Sox12* probes. The left-side section was stained with hematoxylin and eosin (H&E). CM, craniofacial (skeletogenic) mesenchyme; CNS, central nervous system; GM, genital tubercle mesenchyme; H, heart; TO, tongue; VM, vertebral column (skeletogenic) mesenchyme. (D) RNA *in situ* hybridization of adjacent coronal sections of mouse embryo heads at E14.5 and E18.5 with *SoxC* probes. EL, eyelid mesenchyme; J, presumptive jaw mesenchyme; M, mesenchyme; OE, olfactory epithelium; PS, palatal shelves; R, retina; S, skin primordium TO, tongue; TE, teeth. (E) RNA *in situ* hybridization of various developing organs in E14.5 to E18.5 embryos with *SoxC* probes, as indicated. Ep, epithelium; HF, hair follicle; Mg, midgut; Ms, mesenchyme; Sp, spleen; Th, thymus.





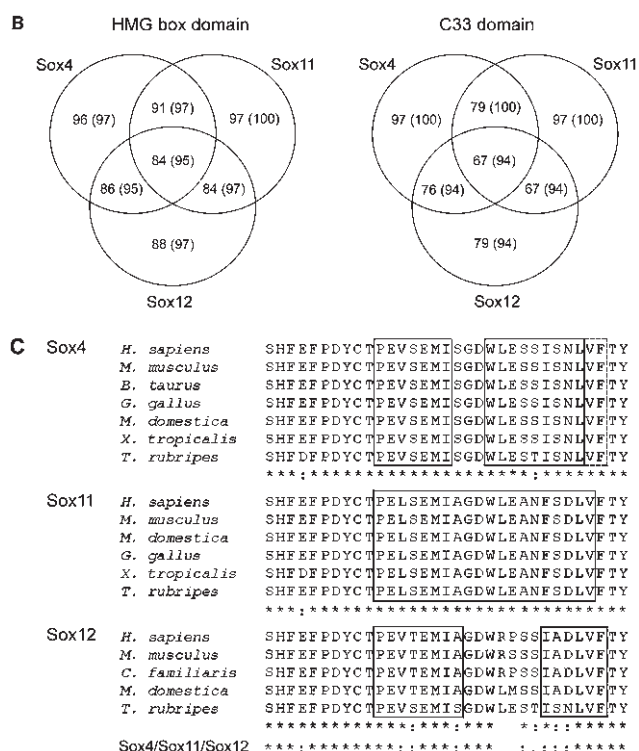


Figure 2. Continued.

transfection of any cell type (36). The 6FXO-p89Luc reporter harbored six tandem copies of a minimum *Fgf4* enhancer directly upstream of the p89 promoter. This enhancer is a direct target of Sox2 and the POU domain protein Oct3 in embryonic stem cells (37). It contains a binding site for Sox proteins immediately adjacent to a POU domain protein-binding site, and it is synergistically activated by several types of Sox and POU proteins, including Sox4, Sox11 and the neuronal cell-specific POU domain protein Brn2 (9,37). The 2HMG-p89Luc reporter harbored two tandem copies of a *CD3ε* minimal enhancer (43). This enhancer was originally shown to contain a binding site for Sry and has been frequently used as a reference to evaluate and compare the DNA-binding and transactivation properties of various members of the Sox family. The 4x48-p89Luc reporter featured four tandem copies of a 48 bp *Col2a1* intron-1 sequence (35,38). This sequence is a cartilage-specific enhancer directly targeted by the SoxE protein Sox9 and SoxD proteins L-Sox5 and Sox6 in chondrocytes. A fourth reporter featured the mouse *Tubb3* proximal promoter from -215 to +54 directly driving the firefly luciferase gene. As described in the Introduction section, this promoter segment contains three Sox-like binding sites needed for activity and is believed to be directly targeted by Sox4 and Sox11 (15). Using the 6FXO-p89Luc reporter first, we found, as previously reported (9), that Sox11 was a more potent transactivator than Sox4 (Figure 3B). Sox12 was also able to transactivate this reporter but less efficiently. Interestingly, although Sox11 was a better transactivator, we reproducibly observed that less Sox4 and Sox12

protein was needed than Sox11 protein to reach a plateau of maximum transactivation. The same relative differences were observed when the proteins were allowed to synergize with Brn2 (Figure 3C). They were also observed when the proteins harbored an N-terminal FLAG epitope (Figure 3D), and we therefore pursued our studies with FLAG-tagged proteins. The same relative differences were also observed using the 2HMG-p89Luc and 4x48-p89Luc reporters (Figure 3E). Of note, the SoxC proteins activated 6FXO-p89Luc and 2HMG-p89Luc as efficiently as Sox9 did, but activated 4x48-p89Luc much less efficiently. This result is consistent with the notion that the SoxC proteins are functionally more closely related to each other than to other Sox proteins. Finally, the same relative differences were also observed when the 6FXO-p89Luc and *Tubb3*-pLuc reporters were transfected into Cos1, MC3T3-E1 and neuronal precursor cells (Figure 3F). Together, these data demonstrate for the first time that Sox12 is a transactivator and that it is several fold less potent than Sox4, and up to an order of magnitude less potent than Sox11.

**The SoxC C33 domain is needed for transactivation**

The TAD of Sox4 and Sox11 was previously shown to be located in the protein 140- and 123-residue-long C-terminal third, respectively, but it was not precisely mapped (9,12). We postulated that it could correspond to the most conserved region of the C-terminus. To test whether this region is needed for transactivation, we constructed expression plasmids for SoxC proteins lacking the C-terminal third (Sox4-C140 and Sox11-C123), the C-terminal 52 residues (SoxC-C52), the C-terminal 33 residues (SoxC-C33) or the entire C-terminal third except the last 52 residues (Sox4-C140/53 and Sox11-C123/53) (Figure 4A). The SoxC-C33 and SoxC-C52 proteins were virtually unable to transactivate 6FXO-p89Luc in transient transfection of Cos1 cells (Figure 4B). Sox4-C140 and Sox11-C123 were also unable to transactivate, as expected, but Sox4-C140/53 and Sox11-C123/53 could transactivate as efficiently as the full-length proteins (Figure 4C). Interestingly, deletion of the C33 domain completely blocked the ability of the SoxC proteins to cooperate with Brn2 (Figure 4D). These data thus demonstrated that the SoxC C33 domain is absolutely needed for transactivation, and that upstream sequences in the C-terminal third are not.

**The SoxC C33 domain is sufficient for transactivation**

To determine which segment in the C-terminus of the SoxC proteins is sufficient for transactivation, we constructed plasmids encoding the GAL4 DNA-binding domain fused to the Sox4 or Sox11 C-terminal third (GAL4/C140[4] or GAL4/C140[11]), the SoxC C52 domain (for Sox4: GAL4/C52[4]) or the C33 domain (for Sox4: GAL4/C33[4]), and we tested the ability of these fusion proteins to transactivate pG5Luc, a reporter gene featuring five GAL4 binding sites (Figure 5A). GAL4/C140[4] and GAL4/C123[11] were able to transactivate the reporter, as previously shown (9,12). Interestingly, GAL4/C123[11] was more potent than





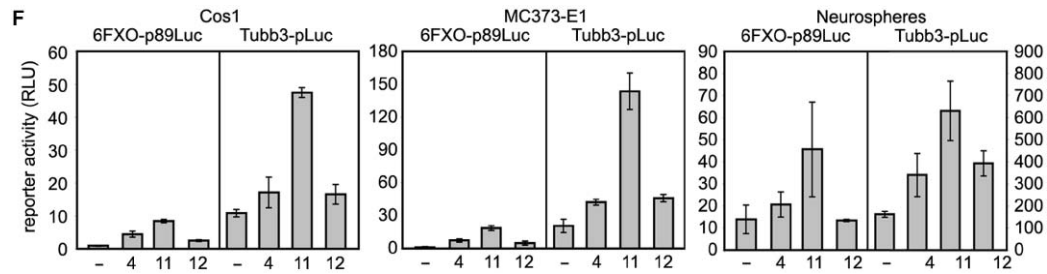
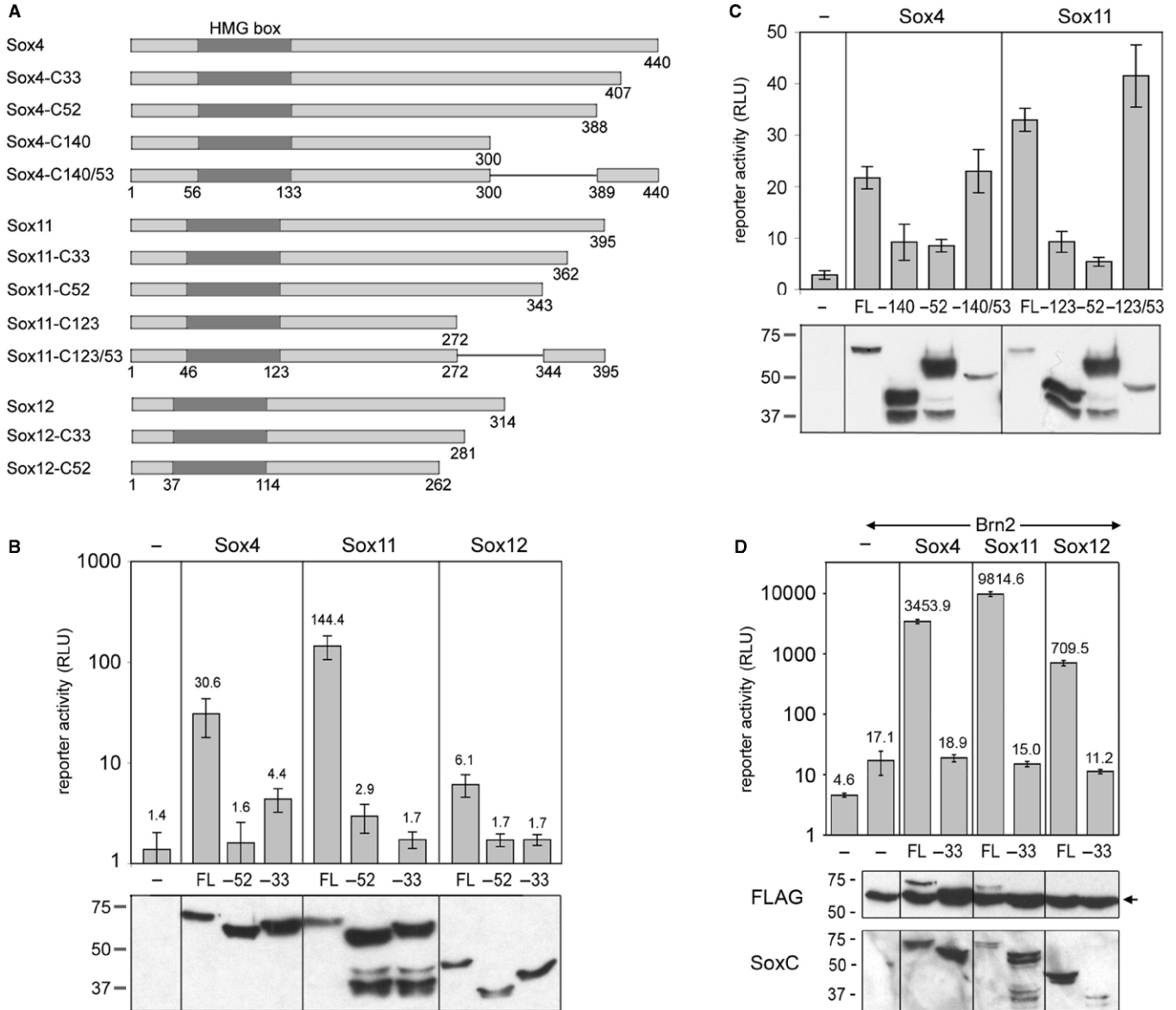
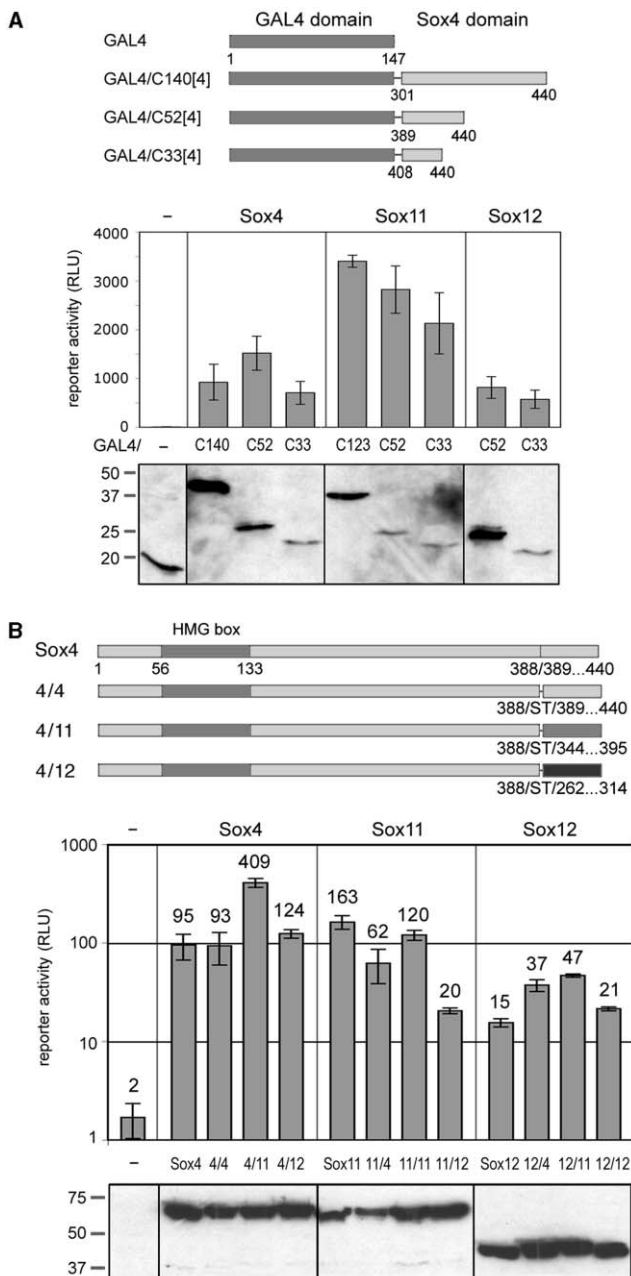


Figure 3. Continued.



**Figure 4.** The SoxC C33 domain is required for transactivation. **(A)** Schematic of SoxC proteins truncated in the C-terminus. The name of each protein is indicated on the left of the rectangle representing it. Numbers designate the first and last protein residues and relevant domain boundaries. A thin line denotes internal deletions. **(B)** Effect of deleting the C52 or C33 region on the ability of SoxC proteins to transactivate 6FXO-p89Luc. Cos1 cells were transfected with 200 ng expression plasmid encoding a full-length SoxC protein (FL) or a SoxC protein lacking the C52 (-52) or C33 (-33) domain. Reporter activities are indicated. SoxC proteins were detected by western blot with anti-FLAG antibody. Note partial degradation of the Sox11-C33 and Sox11-C52 proteins. **(C)** Effect of deleting the SoxC C-terminal third or an internal region on the ability of Sox4 and Sox11 to transactivate 6FXO-p89Luc. Cos1 cells were transfected with 200 ng expression plasmid encoding a full-length SoxC protein (FL), a SoxC protein lacking the C-terminal third (-140 and -123) or a Sox protein lacking an internal segment (-140/53 and -123/53). SoxC proteins were detected by western blot with anti-FLAG antibody. **(D)** Effect of deleting the C33 domain on the ability of SoxC proteins to synergize with Brn2 in transactivating 6FXO-p89Luc. Cos1 cells were transfected with 100 ng expression plasmid encoding Brn2 and 100 ng of expression plasmid encoding a full-length (FL) SoxC protein or a SoxC protein lacking C33 (-C33). The western blot with a FLAG antibody shows the amount of Brn2 protein (arrow) made in each culture and some of the SoxC proteins. The western blot with the SoxC antibody shows the relative amounts of SoxC proteins made in each culture. These proteins appear as doublet, an electrophoresis artifact.



**Figure 5.** The SoxC C33 domain is sufficient for transactivation. (A) Schematic of GAL4-SoxC fusion proteins and transactivation of pG5Luc by these proteins. Cos1 cells were transfected with 100 ng GAL4/SoxC expression plasmid. Reporter activities are presented in comparison with the amount of GAL4/SoxC fusion proteins present in each culture at the end of the experiment and detected with anti-GAL4 antibody. (B) Schematic of Sox4 proteins with swapped SoxC C52 regions and transactivation of 6FXO-p89Luc by these swapped SoxC proteins. The swapped Sox11 and Sox12 proteins were made as shown for Sox4. Cos1 cells were transfected with 100 ng swapped SoxC expression plasmid. SoxC proteins were detected with an anti-FLAG antibody.

GAL4/C140[4], but GAL4/C52[4] and GAL4/C33[4] were as potent as GAL4/C140[4], and GAL4/C52[11] and GAL4/C33[11] were as potent as GAL4/C123[11]. GAL4/C52[12] and GAL4/C33[12] were able to transactivate, but were less potent than the GAL4/Sox4 and GAL4/Sox11 fusion proteins. These data thus indicate

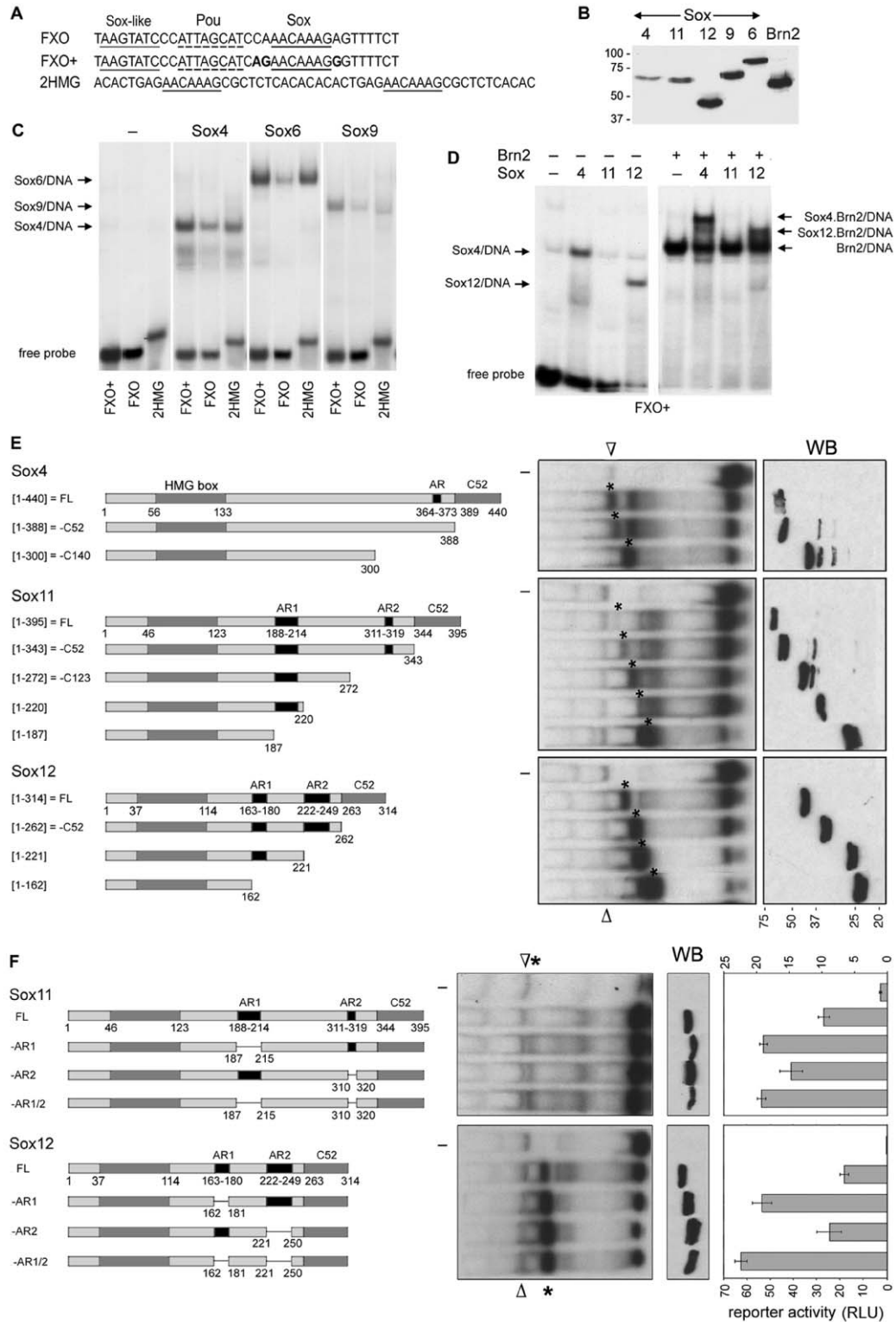
that the C33 domain is sufficient for transactivation, and they suggest that sequence differences between the three proteins in this domain contribute, at least in part, to the differences in transactivation efficiency of the full-length proteins.

To test the latter possibility, we exchanged the SoxC C52 domains (Figure 5B). In doing so, we introduced a serine and a threonine residue (encoded by the ScaI restriction enzyme site) between the C53 and C52 residues, but verified that swapping the C52 domain of either protein with itself (i.e. inserting two new amino acids) did not significantly affect the activity of the proteins. Swapping the Sox4 C-terminus with C52[11] resulted in increasing the activity of Sox4, whereas swapping it with C52[12] had no effect. Swapping the Sox11 C-terminus with C52[4] or C52[12] decreased the activity of Sox11, and swapping the Sox12 C-terminus for C52[4] or C52[11] resulted in increasing the activity of Sox12. We concluded that the C33 domain is the SoxC TAD, and that strength differences between the three SoxC domains explain, at least in part, differences in strength between the full-length proteins.

### The SoxC proteins bind DNA with different efficiencies *in vitro*

We next compared the DNA-binding properties of the three SoxC proteins. We used several probes (Figure 6A). FXO corresponded to the *Fgf4* DNA element described earlier. FXO+ was an FXO variant, in which the Sox binding site was optimized for SoxC proteins by changing flanking nucleotides according to a previous report (44). 2HMG featured two tandem copies of the HMG sequence described earlier. As a source of protein, we used crude extracts from Cos1 cells transiently transfected with expression plasmids for Sox proteins harboring an N-terminal FLAG epitope (Figure 6B). We found that Sox4 bound all probes as efficiently as Sox6 and Sox9 (Figure 6C). Each protein bound FXO+ slightly more efficiently than the other two probes. We therefore conducted further assays with FXO+. Binding of Sox11 to this probe was hardly if ever detectable (Figure 6D). Sox12 binding was readily detectable, but always less efficient than Sox4 binding. Sox4 formed a ternary complex with Brn2 and DNA, and Sox12 did too, but even in this condition, Sox11 did not bind DNA. The same results were obtained using probes corresponding to SoxC binding sites in the *Tubb3* promoter (unpublished results).

Based on the fact that the HMG box domains of the three SoxC proteins are almost identical, we postulated that specific regions outside this domain were responsible for controlling DNA binding. The deletion of the N-terminal region, up to six residues upstream of the HMG box, blocked the ability of each protein to bind DNA (unpublished results). Deletion of the C52 or C140 domain did not significantly affect the ability of Sox4 to bind DNA (Figure 6E). In contrast, the deletion of the C52 domain was sufficient to allow Sox11 to bind DNA and to increase the ability of Sox12 to bind DNA. Larger deletions in the C-terminal half further increased the



**Figure 6.** Comparison of the DNA-binding efficiency of the SoxC proteins. (A) Sequences of FXO, FXO+ and 2HMG EMSA probes. Sox, Sox-like and POU domain protein binding sites are underlined. (B) Western blot with anti-FLAG antibody demonstrating the relative amounts of proteins used in EMSA in (C) and (D). (C) EMSA of Sox4, Sox6 and Sox9 with the FXO, FXO+ and 2HMG probes. The Sox protein/DNA complexes are identified with arrows. (D) EMSA of SoxC proteins with FXO+ in the presence (+) or absence (-) of Brn2. Specific protein/DNA complexes are identified with arrows. Note that the Sox4/DNA complex migrates with the same mobility as a less abundant complex present in all samples and that the free probe ran off the right part of the gel. (E) Effect of deleting an increasing segment of the SoxC C-terminus on the ability of the proteins to bind DNA. SoxC proteins were made in Cos1 cells as schematized and used in EMSA with the FXO+ probe. The SoxC/DNA complexes are identified in the EMSA with asterisks and a nonspecific complex with an open triangle. A western blot (WB) hybridized with an anti-FLAG antibody shows the relative amounts of the various proteins. (F) Effect of deleting the acidic regions of Sox11 and Sox12 on the protein's ability to bind DNA and transactivate. Cos1 cells were transfected with 200 ng of expression plasmid encoding full-length or deletion mutant proteins, as schematized, and with the 6FXO-p89Luc reporter and control plasmids. SoxC proteins were tested in EMSA with the FXO+ probe and detected in western blot with an anti-FLAG antibody.



DNA-binding efficiency of Sox11 and Sox12. Deletion of the main acidic region (AR1) of Sox11 resulted in increasing about 2-fold the ability of the protein to bind DNA in EMSA and to transactivate 6FXO-p89Luc in cultured cells, as previously reported (12). Deletion of a smaller acidic region (AR2) had a similar, but milder effect, and deletion of both regions was as effective as deleting AR1 only. Similar effects were observed upon deleting similar acidic regions in Sox12. The data therefore indicated that several domains in Sox11 and Sox12, including the acidic regions and TAD, interfere with the protein's ability to bind DNA *in vitro* and possibly also in cells.

### The SoxC proteins functionally interfere with each other

Finally, we asked whether the SoxC proteins are capable of functionally interacting with each other when co-expressed in the same cells. We did so by testing the effect of co-transfecting full-length SoxC proteins with SoxC proteins truncated in the C-terminus. We observed that the activity of either full-length protein was dramatically inhibited by expressing equivalent amounts of any of the SoxC proteins lacking the C33 domain (Figure 7A). Similar results were obtained in the presence (Figure 7A) and absence of Brn2 (unpublished results) and with larger C-terminal truncations in Sox11 (Figure 7B). Furthermore, each SoxC protein lacking the C33 domain was also able to interfere with the activity of the *Tubb3* reporter, supporting the notion that these mutant proteins were able to interfere with the activity of the endogenous SoxC proteins (Figure 7C). These data thus strongly suggest that, when co-expressed *in vivo*, the SoxC proteins compete with each other to bind and activate their target genes and that gene mutations or protein modifications that would either delete or inactivate the TAD of a SoxC protein could not only block the activity of this protein, but could also affect the activity of all co-expressed, intact SoxC proteins.

## DISCUSSION

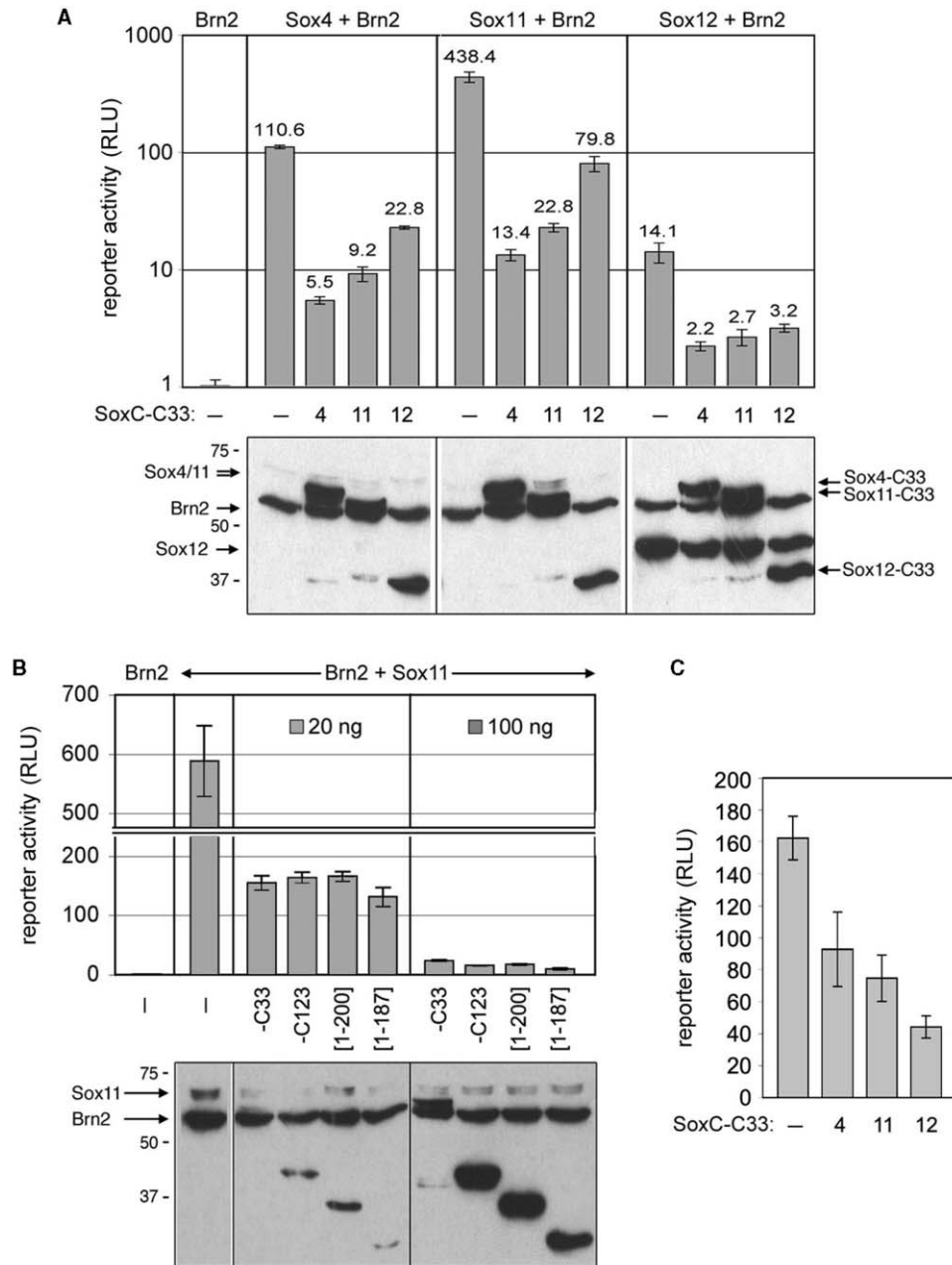
Previous studies have started to reveal critical roles for either or both Sox4 and Sox11 in developmental, physiological and possibly pathological processes, but the extent of their individual and interacting roles *in vivo*, their exact molecular roles, and their modes of regulation remain poorly known. There also remains virtually nothing known about the third member of SoxC group, Sox12. This study has provided several pieces of evidence that strongly support the notion that the three SoxC proteins are likely to functionally interact with each other in many lineages and thereby to fulfill a multitude of critical roles in developmental, physiological and possibly pathological processes. We have shown that the three genes are co-expressed in a number of important cell lineages. Their tight conservation in the HMG box DNA-binding domain and in the TAD allows them to compete for DNA binding and transactivation. However, sequence variations between the three proteins in the TAD and in

other domains are responsible for differential efficiencies in DNA-binding and transactivation, such that each protein may have specific activities and modes of regulation, and may thereby interact either positively or negatively with its relatives.

Several research groups previously analyzed the expression pattern of the SoxC genes, but each group analyzed only one gene at a time, and often used a different species, developmental stages, probe or method than the other groups (8,10,13,18,20). The results of these studies therefore did not indicate how the three genes compare to each other in their pattern and level of expression. To obtain such information, which is important to determine whether the genes may functionally interact, we prepared embryos sections and northern blots with various mouse tissues and cultured cells, and hybridized them with SoxC probes that had similar length, nucleotide composition and labeling efficiencies. Both approaches revealed co-expression of the three genes in many tissues. Neuronal and mesenchymal tissues expressed the three genes at similarly high levels, whereas many other tissues, such as the heart, thymus, spleen and hair follicles, expressed the three genes at generally lower and variable relative levels. These expression data thus constitute a first piece of evidence that the three SoxC genes might functionally interact in many processes. They imply that the three genes may have conserved a high degree of identity in critical promoter elements or in other regulatory regions that specify co-expression in multiple tissues. A combination of bioinformatics and experimental approaches will be needed to test this possibility and uncover the shared and distinct modes of transcriptional regulation of the three genes.

Functional studies *in vivo* have already revealed important roles for Sox4 and Sox11 in several developmental processes, but one can predict that the roles of the genes will only be fully revealed upon inactivation of the entire SoxC group. Only in neuronal cells has the consequence of inactivating two genes (*Sox4* and *Sox11*) been assessed so far (15). This study, however, did not demonstrate whether redundancy exists between the two genes and did not assess the role of *Sox12*. Co-expression of group members was shown for several other Sox groups, including the B, D, E and F groups, and was demonstrated to result in gene redundancy in most cases (2). The SoxB2 genes (*Sox14* and *Sox21*), however, were shown to counteract the action of the SoxB1 genes (*Sox1*, *Sox2* and *Sox3*) in neuronal cells (45). Counteraction was explained by the fact that the SoxB1 and SoxB2 proteins have almost identical HMG box DNA-binding domains, but the SoxB1 proteins have a TAD, whereas the SoxB2 proteins have a transrepression domain. Since Sox4 and Sox11 were previously shown to have a TAD, but Sox12 was not previously characterized, these data prompted us to further characterize and compare the functional properties of the three SoxC proteins.

The alignment of SoxC protein sequences from multiple species revealed that all proteins are 94–100% similar in two domains: the HMG box and the C-terminal 33-residue segment. The former domain has long been known to mediate DNA binding, but the function of the C-terminal



**Figure 7.** Dominant-negative interference of SoxC proteins truncated in the C-terminus. (A) Test of the ability of SoxC proteins lacking the C33 domain to interfere with the activity of SoxC full-length proteins. Cos1 cells were transfected with 100 ng Brn2 expression plasmid, 100 ng SoxC full-length expression plasmid and 100 ng truncated SoxC protein expression plasmid, and with the 6FXO-p89Luc reporter and control plasmids. SoxC and Brn2 proteins were detected by western blot using an anti-FLAG antibody. (B) Test of the ability of Sox11 proteins lacking various lengths of the C-terminus to interfere with the activity of full-length Sox11. Cos1 cells were transfected with 100 ng Sox11 full-length expression plasmid and either 20 or 100 ng truncated Sox11 expression plasmid, as indicated, and with the 6FXO-p89Luc reporter and control plasmids. The Sox11 proteins were detected by western blot using an anti-FLAG antibody. (C) Inhibition of the activity of the *Tubb3* promoter by SoxC proteins lacking the C33 domain. Cos1 cells were transiently transfected with the *Tubb3*-pLuc reporter and with 600 ng of the expression plasmids for the mutant SoxC proteins.

segment had not been previously demonstrated. Given that the TAD of Sox4 and Sox11 was previously located within the C-terminal third of the proteins, but not precisely mapped, we postulated that the most conserved segment of this C-terminal third was the TAD. We demonstrated that this is indeed the case by performing standard, complementary assays. We first proved that this domain

is needed for transactivation by showing that its deletion abrogates transactivation. We then demonstrated that it is sufficient for transactivation by showing that its fusion to the GAL4 DNA-binding domain results in a protein capable of transactivating a GAL4 reporter. Supporting the notion that the C33 region is solely responsible for transactivation, we showed that the GAL4/C33 fusions

were as potent as fusions harboring a longer C-terminal SoxC segment, and that deletion of the region located directly upstream of C33 had no effect on the ability of Sox4 and Sox11 to transactivate a Sox reporter. With only 33 residues, the SoxC TAD is strikingly short, the shortest mapped in the Sox family (2). It has nevertheless a typical TAD amino acid composition, with acidic, serine and proline residues accounting together for almost half of the residues. Using Prosite and ELM software, we found that this domain features three putative phosphorylation sites conserved in all SoxC proteins: a casein kinase-2 phosphorylation site at position S363 in Sox11, a glycogen synthase kinase-3 and a Polo-like kinase site at position T371; and a Polo-like kinase site at position T376 (unpublished results). Although mutations mimicking loss-of-function or gain-of-function of these sites did not affect the activity of Sox11 in our transient reporter assays (unpublished results), it remains possible that phosphorylation of these conserved sites contributes to modulating the activity of the SoxC proteins in response to various signaling pathways in their natural context. Our observations that the C33 domain is sufficient for transactivation raise the question of the role of the serine-rich region preceding the TAD in Sox4 and Sox11. This region was predicted to participate in transactivation, based on the fact that TADs are often serine-rich (8,9). Although this region was not needed for transactivation of GAL4 and Sox reporters, we cannot exclude the possibility that it may modulate the activity of the proteins on endogenous target genes *in vivo*. It is unlikely, however, that its role is highly specific and essential, given that this region greatly varies in length and sequence across species.

We confirmed that Sox11 is a more potent transactivator than Sox4 (9), and we also demonstrated that Sox12 is a transactivator, but weaker than Sox4 and Sox11. This conclusion was reached using somewhat artificial reporter genes that contained different Sox sites, flanking sequences and Sox site copy numbers, but it was also reached using the natural *Tubb3* promoter, the only alleged target gene currently known for Sox4 and Sox11 (15). The reproducibility of our results using very different reporters thus strongly suggests that the three proteins also transactivate their direct target genes *in vivo* with different efficiencies. We then showed that even though they are highly conserved, the TADs of the SoxC proteins are responsible at least for a large part for the differences in transactivation efficiencies between the three proteins. The fusion of GAL4 with the Sox11 TAD was indeed more potent than the fusion with the Sox4 or Sox12 TAD. Moreover, Sox4 and Sox12 were also more active with a swapped Sox11 TAD than with either of their domains, and Sox11 was less active with a Sox4 or Sox12 TAD than with its own domain. Secondary structure models predicted that 20 residues of the Sox11 TAD form a continuous  $\alpha$ -helix, but that this helix is shorter and interrupted in both Sox4 and Sox12. The length and stability of this helix is thus likely to determine the strength of the SoxC TADs. Interestingly, the secondary structure of each SoxC TAD is highly conserved between orthologues, strongly suggesting that each protein has conserved or acquired unique residues that determine its

specific activity, and thus that co-expression of the three genes does not necessarily imply redundancy.

We next demonstrated that the three SoxC proteins differ not only in their transactivation efficiency, but also in their DNA-binding efficiency. We showed that full-length Sox4 avidly binds DNA in EMSA, as do Sox proteins from other groups, namely the SoxD Sox6 and SoxE Sox9, whereas full-length Sox12 binds DNA less efficiently, and full-length Sox11 hardly binds DNA. Our data on full-length Sox11 are thus consistent with those published by Wiebe and collaborators (12). This group reported that they tried to improve binding of Sox11 to DNA by varying EMSA conditions, but did not succeed. Similarly, we tested many parameters, including salt and detergent type and concentration, pH, Sox and other protein concentration, and different nonspecific DNA competitors, but also failed (unpublished results). We thus conclude that a potent mechanism may exist that blocks binding of Sox11 to DNA. Given that their HMG box domains are virtually identical, we postulated that the differences in DNA-binding efficiency between the three SoxC proteins were due to sequence differences lying outside the HMG box. We verified the accuracy of this prediction by showing that the three SoxC proteins avidly and similarly bind DNA when their C-terminal half or third is deleted. We confirmed the report by Wiebe and collaborators (12) that deletion of the main acidic stretch in the middle of Sox11 increases the ability of the protein to bind DNA. We, however, provided additional data that significantly extend the conclusion reached by this group. We showed that deleting the other acidic region of the protein or the C52 domain had a similar effect. Moreover, deleting the acidic regions of Sox12 also increased the efficiency of DNA binding of the protein. Our data thus indicated that several regions in Sox11 and Sox12 have the ability to interfere with binding of the HMG box to DNA *in vitro*. They prompted us to test whether the SoxC proteins also differ in their DNA-binding efficiency in intact cells.

Evidence from our study and previous studies that Sox11 robustly transactivates reporters harboring Sox recognition sites in transfected cells provides irrefutable demonstration that Sox11 is capable of binding DNA in cells. The question then arises of the efficiency with which Sox11 binds DNA compared to Sox4 and Sox12. Our observation that a plateau of maximum transactivation of reporter genes is reproducibly obtained with less Sox4 or Sox12 protein than Sox11 protein is consistent with the notion that Sox11 does not bind DNA as efficiently as Sox4 and Sox12 in cells. The confirmation of the previous finding (12) that Sox11 is a better transactivator in cells upon deletion of its acidic regions is also consistent with the notion that the acidic regions reduce the DNA-binding efficiency of Sox11. We obtained a similar result for Sox12, though, indicating that the two proteins may be similarly affected. To test the role of the C-terminal region, which contains the TAD, we tested whether SoxC proteins lacking the C-terminus were able to interfere with the activity of SoxC full-length proteins. We found that Sox11 lacking the C33 domain was capable of interfering in a dominant-negative manner with the activity of



any SoxC full-length protein. Its efficiency was similar to that of Sox4 and Sox12 lacking the same domain. This result thus suggests that the C33 domain may also reduce the ability of Sox11 to bind DNA in cells. Interestingly, truncations of Sox11 in the C-terminal half had similar effects whether they removed the C33 domain only or also removed the acidic regions. Among several scenarios that can be proposed to account for these DNA-binding inhibitory effects of distinct protein domains, we like to favor a scenario, whereby these regions would act in concert with one another. We propose that the acidic domains might act as hinges to allow the C33 domain to block access of the HMG box domain to DNA. In this model, the C33 domain would have the dual function of mediating transactivation and indirectly restricting its own activity by limiting binding of the protein to DNA. The control of SoxC activity *in vivo* could thus involve factors capable of preventing negative autoregulation. These factors could include transcription factors that have been shown to physically interact and synergize with the Sox HMG box in binding DNA, such as Brn2 that we used in this study, but also other POU domain proteins, homeodomain proteins, basic helix-loop-helix proteins, and other proteins (46,47). These factors may synergize with SoxC proteins at least in part by blocking access of other regions of the SoxC proteins to the HMG box domain. Other and yet unknown factors involved in preventing negative autoregulation could include the transcriptional co-activators that are likely to interact with the C33 domain, or proteins that may specifically interact with the acidic regions of the SoxC proteins. Identifying SoxC-interacting proteins in future work using for instance a yeast two-hybrid approach should help increase understanding of the mode of action and mode of regulation of these proteins.

Based on the fact that all three SoxC proteins are transcriptional activators, they can be predicted to function similarly where co-expressed *in vivo*. However, our results suggest that Sox12 may act to temper the activity of Sox4 and Sox11, by successfully competing with its group members for DNA binding and thereby preventing the other two proteins, which have a stronger TAD, to exert their full activity. The consequences of inactivating Sox12 *in vivo* are not known yet. One can now speculate that inactivation of *Sox12* may lead to effects opposite to those observed upon inactivating *Sox4* and *Sox11* in sites of co-expression. On the other hand, the consequences of inactivating *Sox4* and/or *Sox11* may be partially or completely compensated by *Sox12*. It will thus be important to compare the consequences of inactivating the genes individually and simultaneously. Generating double and triple knockouts for these genes is predicted to be a laborious and difficult task, as it is for any set of genes, but most particularly considering that the *Sox4* and *Sox11* single knockout mice die early. Another approach could be to express a SoxC mutant protein capable of blocking the activity of wild-type SoxC proteins in a dominant-negative manner.

In this context, our data demonstrating that any of the three SoxC proteins lacking its C33 domain is capable of inhibiting transactivation of reporter constructs by any of

the wild-type SoxC proteins is particularly interesting. These truncated SoxC proteins are likely to act by competing with the wild-type proteins for DNA binding. Notably, robust inhibition was seen when equal amounts of wild-type and mutant protein were expressed in transfected cells. This result strongly suggests that spontaneous or targeted mutations of the TAD in only one SoxC allele could cause a disease more severe than a complete gene knockout. Although no human disease or spontaneous mouse mutation has yet been reported that is due to a *SoxC* mutation, these considerations, together with the expression pattern of the three genes and the already known consequences of inactivating *Sox4* and *Sox11* in the mouse, strongly suggest that diseases due to mutations in a *SoxC* gene could be very severe. Not only gene mutations could interfere with the activity of SoxC proteins, but also physiological or pathological events that could induce protein posttranslational modifications, such as proteolysis, phosphorylation or SUMOylation.

In conclusion, this study significantly increases our knowledge of the relative expression pattern and molecular properties of the three SoxC proteins. Sox4 and Sox11 were previously shown to be essential in several major developmental, physiological and pathological processes. Our data strongly suggest that the two genes are likely to functionally interact with each other as well as with their close relative Sox12 in many processes, both positively and negatively, and therefore that the extent to which they have been thought so far to contribute to important processes *in vivo* has probably been largely underestimated. Their roles in these and other processes will be fully uncovered only by studying the three genes simultaneously.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank Pallavi Bhattaram for critical advice on the article. This work was supported by National Institutes of Health/National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIH/NIAMS) (AR54153 to V.L.), Arthritis Foundation (Postdoctoral Fellowship to A.P.) and National Multiple Sclerosis Society (Postdoctoral Fellowship to C.E.P.). Funding to pay the Open Access publication charged for this article was provided by the NIH/NIAMS.

*Conflict of interest statement.* None declared.

## REFERENCES

- Schepers, G.E., Teasdale, R.D. and Koopman, P. (2002) Twenty pairs of Sox: extent, homology, and nomenclature of the mouse and human Sox transcription factor gene families. *Dev. Cell*, **3**, 167–170.
- Lefebvre, V., Dumitriu, B., Penzo-Méndez, A., Han, Y. and Pallavi, B. (2007) Control of cell fate and differentiation by Sry-related high-mobility-group box (Sox) transcription factors. *Int. J. Biochem. Cell. Biol.*, **39**, 2195–2214.

3. Koopman,P., Gubbay,J., Vivian,N., Goodfellow,P. and Lovell-Badge,R. (1991) Male development of chromosomally female mice transgenic for Sry. *Nature*, **351**, 117–121.
4. Laudet,V., Stehelin,D. and Clevers,H. (1993) Ancestry and diversity of the HMG box superfamily. *Nucleic Acids Res.*, **21**, 2493–2501.
5. Bianchi,M.E. and Agresti,A. (2005) HMG proteins: dynamic players in gene regulation and differentiation. *Curr. Opin. Genet. Dev.*, **15**, 496–506.
6. Bowles,J., Schepers,G. and Koopman,P. (2000) Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators. *Dev. Biol.*, **227**, 239–255.
7. Crémazy,F., Berta,P. and Girard,F. (2001) Genome-wide analysis of Sox genes in *Drosophila melanogaster*. *Mech. Dev.*, **109**, 371–375.
8. van de Wetering,M., Oosterwegel,M., van Norren,K. and Clevers,H. (1993) Sox-4, an Sry-like HMG box protein, is a transcriptional activator in lymphocytes. *EMBO J.*, **12**, 3847–3854.
9. Kuhlbrodt,K., Herbarth,B., Sock,E., Enderich,J., Hermans-Borgmeyer,I. and Wegner,M. (1998) Cooperative function of POU proteins and SOX proteins in glial cells. *J. Biol. Chem.*, **273**, 16050–16057.
10. Jay,P., Goze,C., Marsollier,C., Taviaux,S., Hardelin,J.P., Koopman,P. and Berta,P. (1995) The human SOX11 gene: cloning, chromosomal assignment and tissue expression. *Genomics*, **29**, 541–545.
11. Maschhoff,K.L., Anziano,P.Q., Ward,P. and Baldwin,H.S. (2003) Conservation of Sox4 gene structure and expression during chicken embryogenesis. *Gene*, **320**, 23–30.
12. Wiebe,M.S., Nowling,T.K. and Rizzino,A. (2003) Identification of novel domains within Sox-2 and Sox-11 involved in autoinhibition of DNA binding and partnership specificity. *J. Biol. Chem.*, **278**, 17901–17911.
13. Jay,P., Sahly,I., Gozé,C., Taviaux,S., Poulat,F., Couly,G., Abitbol,M. and Berta,P. (1997) Sox22 is a new member of the Sox gene family, mainly expressed in human nervous tissue. *Hum. Mol. Genet.*, **6**, 1069–1077.
14. Cheung,M., Abu-Elmagd,M., Clevers,H. and Scotting,P.J. (2000) Roles of Sox4 in central nervous system development. *Brain. Res. Mol. Brain. Res.*, **79**, 180–191.
15. Bergsland,M., Werme,M., Malewicz,P.M., Perlmann,T. and Muhr,J. (2006) The establishment of neuronal properties is controlled by Sox4 and Sox11. *Genes Dev.*, **20**, 3475–3486.
16. Lioubinski,O., Muller,M., Wegner,M. and Sander,M. (2003) Expression of Sox transcription factors in the developing mouse pancreas. *Dev. Dyn.*, **227**, 402–408.
17. 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.* (1996) Defects in cardiac outflow tract formation and pro-lymphocyte expansion in mice lacking Sox-4. *Nature*, **380**, 711–714.
18. Reppe,S., Rian,E., Jemtland,R., Olstad,O.K., Gautvik,V.T. and Gautvik,K.M. (2000) Sox-4 messenger RNA is expressed in the embryonic growth plate and regulated via the parathyroid hormone/parathyroid hormone-related protein receptor in osteoblast-like cells. *J. Bone Miner. Res.*, **15**, 2402–2412.
19. Hargrave,M., Wright,E., Kun,J., Emery,J., Cooper,L. and Koopman,P. (1997) Expression of the Sox11 gene in mouse embryos suggests roles in neuronal maturation and epithelio-mesenchymal induction. *Dev. Dyn.*, **210**, 79–86.
20. Sock,E., Rettig,S.D., Enderich,J., Bosl,M.R., Tamm,E.R. and Wegner,M. (2004) Gene targeting reveals a widespread role for the high-mobility-group transcription factor Sox11 in tissue remodeling. *Mol. Cell. Biol.*, **24**, 6635–6644.
21. Weigle,B., Ebner,R., Temme,A., Schwind,S., Schmitz,M., Kiessling,A., Rieger,M.A., Schackert,G., Schackert,H.K. and Rieber,E.P. (2005) Highly specific overexpression of the transcription factor SOX11 in human malignant gliomas. *Oncol. Rep.*, **13**, 139–144.
22. Liu,P., Ramachandran,S., Ali Seyed,M., Scharer,C.D., Laycock,N., Dalton,W.B., Williams,H., Karanam,S., Datta,M.W., Jaye,D.L. *et al.* (2006) Sex-determining region Y box 4 is a transforming oncogene in human prostate cancer cells. *Cancer Res.*, **66**, 4011–4009.
23. Pramoongjago,P., Baras,A.S. and Moskaluk,C.A. (2006) Knockdown of Sox4 expression by RNAi induces apoptosis in ACC3 cells. *Oncogene*, **25**, 5626–5639.
24. Schilham,M.W., Moerer,P., Cumano,A. and Clevers,H.C. (1997) Sox-4 facilitates thymocyte differentiation. *Eur. J. Immunol.*, **27**, 1292–1295.
25. Wilson,M.E., Yang,K.Y., Kalousova,A., Lau,J., Kosaka,Y., Lynn,F.C., Wang,J., Mrejen,C., Episkopou,V., Clevers,H.C. *et al.* (2005) The HMG box transcription factor Sox4 contributes to the development of the endocrine pancreas. *Diabetes*, **54**, 3402–3409.
26. Nissen-Meyer,L.S., Jemtland,R., Gautvik,V.T., Pedersen,M.E., Paro,R., Fortunati,D., Pierroz,D.D., Stadelmann,V.A., Reppe,S., Reinholt,F.P. *et al.* (2007) Osteopenia, decreased bone formation and impaired osteoblast development in Sox4 heterozygous mice. *J. Cell Sci.*, **120**, 2785–2795.
27. Hoser,M., Baader,S.L., Bösl,M.R., Ihmer,A., Wegner,M. and Sock,E. (2007) Prolonged glial expression of Sox4 in the CNS leads to architectural cerebellar defects and ataxia. *J. Neurosci.*, **27**, 5495–5505.
28. Potzner,M.R., Griffel,C., Lütjen-Drecoll,E., Bösl,M.R., Wegner,M. and Sock,E. (2007) Prolonged Sox4 expression in oligodendrocytes interferes with normal myelination in the central nervous system. *Mol. Cell. Biol.*, **27**, 5316–5326.
29. Kornblum,H.I., Yanni,D.S., Easterday,M.C. and Seroogy,K.B. (2000) Expression of the EGF receptor family members ErbB2, ErbB3, and ErbB4 in germinal zones of the developing brain and in neurosphere cultures containing CNS stem cells. *Dev. Neurosci.*, **22**, 16–24.
30. Zheng,X.S., Yang,X.F., Liu,W.G., Shen,G., Pan,D.S. and Luo,M. (2007) A novel method for culturing neural stem cells. *In Vitro Cell Dev. Biol. Anim.*, **43**, 155–158.
31. Penzo-Méndez,A., Dy,P. and Lefebvre,V. (2007) Generation of mice harboring a Sox4 conditional null allele. *Genesis*, **45**, 776–780.
32. Smits,P., Li,P., Mandel,J., Zhang,Z., Deng,J.M., Behringer,R.R., de Crombrughe,B. and Lefebvre,V. (2001) The transcription factors L-Sox5 and Sox6 are essential for cartilage formation. *Dev. Cell*, **1**, 277–290.
33. Thompson,J.D., Gibson,T.J., Plewniak,F., Jeanmougin,F. and Higgins,D.G. (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.*, **25**, 4876–4882.
34. Guermeur,Y., Geourjon,C., Gallinari,P. and Deleage,G. (1999) Improved performance in protein secondary structure prediction by inhomogeneous score combination. *Bioinformatics*, **15**, 413–421.
35. Lefebvre,V., Huang,W., Harley,V.R., Goodfellow,P.N. and de Crombrughe,B. (1997) SOX9 is a potent activator of the chondrocyte-specific enhancer of the pro- $\alpha$ 1(II) collagen gene. *Mol. Cell. Biol.*, **17**, 2336–2346.
36. Lefebvre,V., Li,P. and de Crombrughe,B. (1998) A new long form of Sox5 (L-Sox5), Sox6 and Sox9 are coexpressed in chondrogenesis and cooperatively activate the type II collagen gene. *EMBO J.*, **17**, 5718–5733.
37. Yuan,H., Corbi,N., Basilico,C. and Dailey,L. (1995) Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. *Genes Dev.*, **9**, 2635–2645.
38. Lefebvre,V., Zhou,G., Mukhopadhyay,K., Smith,C.N., Zhang,Z., Eberspaecher,H., Zhou,X., Sinha,S., Maity,S.N. and de Crombrughe,B. (1996) An 18-base-pair sequence in the mouse pro- $\alpha$ 1(II) collagen gene is sufficient for expression in cartilage and binds nuclear proteins that are selectively expressed in chondrocytes. *Mol. Cell. Biol.*, **16**, 4512–4523.
39. MacGregor,G.R. and Caskey,C.T. (1989) Construction of plasmids that express *E. coli* beta-galactosidase in mammalian cells. *Nucleic Acids Res.*, **17**, 2365.
40. Bazan,E., Alonso,F.J., Redondo,C., Lopez-Toledano,M.A., Alfaro,J.M., Reimers,D., Herranz,A.S., Paino,C.L., Serrano,A.B., Cobacho,N. *et al.* (2004) In vitro and in vivo characterization of neural stem cells. *Histol. Histopathol.*, **19**, 1261–1275.
41. Campos,L.S. (2004) Neurospheres: insights into neural stem cell biology. *J. Neurosci. Res.*, **78**, 761–769.
42. Sudo,H., Kodama,H.A., Amagai,Y., Yamamoto,S. and Kasai,S. (1983) In vitro differentiation and calcification in a new clonal osteogenic cell line derived from newborn mouse calvaria. *J. Cell. Biol.*, **96**, 191–198.

43. Harley, V.R., Jackson, D.I., Hextall, P.J., Hawkins, J.R., Berkovitz, G.D., Sockanathan, S., Lovell-Badge, R. and Goodfellow, P.N. (1992) DNA binding activity of recombinant SRY from normal males and XY females. *Science*, **255**, 453–456.
44. van Beest, M., Dooijes, D., van De Wetering, M., Kjaerulf, S., Bonvin, A., Nielsen, O. and Clevers, H. (2000) Sequence-specific high mobility group box factors recognize 10-12-base pair minor groove motifs. *J. Biol. Chem.*, **275**, 27266–27273.
45. Sandberg, M., Källström, M. and Muhr, J. (2005) Sox21 promotes the progression of vertebrate neurogenesis. *Nat. Neurosci.*, **8**, 995–1001.
46. Kamachi, Y., Uchikawa, M. and Kondoh, H. (2000) Pairing SOX off: with partners in the regulation of embryonic development. *Trends Genet.*, **16**, 182–187.
47. Wissmüller, S., Kosian, T., Wolf, M., Finzsch, M. and Wegner, M. (2006) The high-mobility-group domain of Sox proteins interacts with DNA-binding domains of many transcription factors. *Nucleic Acids Res.*, **34**, 1735–1744.