Sox proteins: regulators of cell fate specification and differentiation

Yusuke Kamachi* and Hisato Kondoh*

Summary
Sox transcription factors play widespread roles during development; however, their versatile functions have a relatively simple basis: the binding of a Sox protein alone to DNA does not elicit transcriptional activation or repression, but requires binding of a partner transcription factor to an adjacent site on the DNA. Thus, the activity of a Sox protein is dependent upon the identity of its partner factor and the context of the DNA sequence to which it binds. In this Primer, we provide an mechanistic overview of how Sox family proteins function, as a paradigm for transcriptional regulation of development involving multi-transcription factor complexes, and we discuss how Sox factors can thus regulate diverse processes during development.

Key words: HMG domain, Partner factors, Transcriptional regulation, Stem cells, Cell lineage specification

Introduction
An emerging view of transcriptional regulation in developmental processes is that transcription factor complexes, rather than single transcription factors, play a major role (Reményi et al., 2004; Verger and Duterque-Coquillaud, 2002). This idea has been pioneered, most rigorously tested and studied in a variety of developmental contexts in the case of Sox (SRY-related HMG-box) family proteins. Sox family proteins are a conserved group of transcriptional regulators (see Box 1) defined by the presence of a highly conserved high-mobility group (HMG) domain that mediates DNA binding. This domain was first identified in Sry, a crucial factor involved in mammalian male sex determination (Gubbay et al., 1990; Sinclair et al., 1990). Multiple other Sox proteins have subsequently been identified and analysed. Vertebrate genomes contain ~20 Sox family members with highly divergent developmental functions, as was indicated by the initial analyses of Sox expression in embryos (Collignon et al., 1996; Uwanogho et al., 1995). Although Sox proteins are also conserved in invertebrate lineages and exert regulatory functions (Phochanukul and Russell, 2010), we here confine our discussion to the vertebrate Sox family members. In this Primer, we first outline the structure and molecular activities of Sox proteins, before discussing their roles during vertebrate development. In particular, the action and function of Sox proteins will be overviewed as a paradigm for transcriptional regulation mediated by a family of transcription factors.

Sox protein structure
Sox proteins can bind to ATTGTT or related sequence motifs through their HMG domain, which consists of three α helices (Badis et al., 2009; Kondoh and Kamachi, 2010). This binding is established by the interaction of the HMG domain with the minor groove of the DNA, which widens the minor groove and causes DNA to bend towards the major groove (Reményi et al., 2003). It is speculated that DNA bending itself may contribute to the regulatory functions of Sox proteins (Pevny and Lovell-Badge, 1997), but this has not been proved experimentally. Sox proteins are classified into groups A-H, depending on the amino acid sequence of the HMG domain (Fig. 1A). Outside the HMG domain, strong homology, with regards to amino acid sequence and the overall organization of protein domains, is found only within a group (Bowles et al., 2000; Schepers et al., 2002) (Fig. 1B) (see also Box 1).

The SoxA group contains only Sry, which is encoded in mammalian Y chromosomes. Although the HMG box of Sry is highly conserved between species, sequences outside this domain are highly divergent among mammalian species (Sekido, 2010). The SoxB group is split into two sub-groups. SoxB1 comprises Sox1, Sox2 and Sox3; fish species also possess Sox19. These family members have short N-terminal sequences followed by the HMG domain was first identified in Sry, a crucial factor involved in mammalian male sex determination (Gubbay et al., 1990; Sinclair et al., 1990). Multiple other Sox proteins have subsequently been identified and analysed. Vertebrate genomes contain ~20 Sox family members with highly divergent developmental functions, as was indicated by the initial analyses of Sox expression in embryos (Collignon et al., 1996; Uwanogho et al., 1995). Although Sox proteins are also conserved in invertebrate lineages and exert regulatory functions (Phochanukul and Russell, 2010), we here confine our discussion to the vertebrate Sox family members. In this Primer, we first outline the structure and molecular activities of Sox proteins, before discussing their roles during vertebrate development. In particular, the action and function of Sox proteins will be overviewed as a paradigm for transcriptional regulation mediated by a family of transcription factors.

Sox protein structure
Sox proteins can bind to ATTGTT or related sequence motifs through their HMG domain, which consists of three α helices (Badis et al., 2009; Kondoh and Kamachi, 2010). This binding is established by the interaction of the HMG domain with the minor groove of the DNA, which widens the minor groove and causes DNA to bend towards the major groove (Reményi et al., 2003). It is speculated that DNA bending itself may contribute to the regulatory functions of Sox proteins (Pevny and Lovell-Badge, 1997), but this has not been proved experimentally. Sox proteins are classified into groups A-H, depending on the amino acid sequence of the HMG domain (Fig. 1A). Outside the HMG domain, strong homology, with regards to amino acid sequence and the overall organization of protein domains, is found only within a group (Bowles et al., 2000; Schepers et al., 2002) (Fig. 1B) (see also Box 1).

The SoxA group contains only Sry, which is encoded in mammalian Y chromosomes. Although the HMG box of Sry is highly conserved between species, sequences outside this domain are highly divergent among mammalian species (Sekido, 2010). The SoxB group is split into two sub-groups. SoxB1 comprises Sox1, Sox2 and Sox3; fish species also possess Sox19. These family members have short N-terminal sequences followed by the HMG domain was first identified in Sry, a crucial factor involved in mammalian male sex determination (Gubbay et al., 1990; Sinclair et al., 1990). Multiple other Sox proteins have subsequently been identified and analysed. Vertebrate genomes contain ~20 Sox family members with highly divergent developmental functions, as was indicated by the initial analyses of Sox expression in embryos (Collignon et al., 1996; Uwanogho et al., 1995). Although Sox proteins are also conserved in invertebrate lineages and exert regulatory functions (Phochanukul and Russell, 2010), we here confine our discussion to the vertebrate Sox family members. In this Primer, we first outline the structure and molecular activities of Sox proteins, before discussing their roles during vertebrate development. In particular, the action and function of Sox proteins will be overviewed as a paradigm for transcriptional regulation mediated by a family of transcription factors.

Box 1. The evolution of Sox genes
Sox (SRY-related HMG-box) genes encoding the Sox proteins are conserved throughout the animal kingdom. The genes for Sox/Tcf type high-mobility group (HMG) domain transcription factors are already present in unicellular choanoflagellates, the animals most closely related to multicellular metazoans, although Hox and other segmentation-related transcription factor genes are absent, indicating an ancient origin of Sox factors (King et al., 2008). True Sox factors became distinguished in the most primitive metazoan Trichoplax (Srivastava et al., 2008). Even in invertebrates, Sox genes can be classified into B, C, D, E and F groups (Phochanukul and Russell, 2010). Sry (group A) is unique to mammalian Y chromosomes and is assumed to have been derived from Sox3 (group B1) (Foster and Graves, 1994), which is encoded on the X chromosome. The presence of multiple Sox genes in most Sox groups in vertebrates is a consequence of multiple rounds of genome duplication, which began in primitive chordates. Fish genomes have experienced three rounds of duplications, as opposed to two rounds in other higher vertebrates. This has resulted in frequent duplicated Sox genes, which are exemplified by Sox1a and Sox1b (Kamachi et al., 2009; Okuda et al., 2006).

The functional redundancy within a group of Sox genes allows individual protein members of the group to change their amino acid sequence and hence their function over evolutionary time. In mammals, Sox15 is a singlet group G Sox, which was derived from an ancestor of Sox19 (group B1) during vertebrate evolution, as indicated by its conserved synteny (Kamachi et al., 2009; Okuda et al., 2006). Other taxon-restricted Sox genes appear to have arisen analogously, e.g. SoxH (Sox30) from SoxD, as evidenced by the presence of remnant of coiled-coil domain sequence (Fig. 1C), or by tandem duplication of a genomic segment, e.g. Sox22 in fish from duplicated Sox17 (Kobayashi et al., 2006).
domain and long C-terminal sequences after the HMG domain. The C-terminal sequence includes a domain assigned as a transcriptional activation domain, based on an assay using fusion to heterologous DNA-binding domains, as detailed below (Kamachi et al., 1998). SoxB2 proteins, which comprise Sox14 and Sox21, not only have an HMG domain similar to SoxB1 proteins, but also share a short basic amino acid sequence, known as ‘B homology’, with them. Unlike SoxB1 proteins, however, SoxB2 proteins have a repression domain in the C-terminal sequence (Uchikawa et al., 1999).

SoxC, SoxE and SoxF groups have protein organizations that are analogous to SoxB1 proteins, i.e. a total length of 300-500 amino acids with HMG domains located close to the N terminus and an activation domain in the C-terminal region. However, the amino acid sequences outside the HMG domain are unique to individual groups. SoxC comprises Sox4, Sox11 and Sox12 (Dy et al., 2008); SoxE comprises Sox8, Sox9 and Sox10 (Stolt and Wegner, 2010); and SoxF comprises Sox7, Sox17 and Sox18 (Francois et al., 2010). SoxE proteins also contain a self-dimerization domain on

Fig. 1. Grouping and structures of Sox proteins. (A) An unrooted phylogenetic neighbour-joining tree of the high mobility group (HMG) domains of the mouse in comparison with outgroup HMG domains of Lef1/Tcf7 (lymphoid enhancer binding factor 1/transcription factor 7, T cell specific) and Hmgb1. (B) Alignment and classification of the Sox HMG domains. Sequences (79 amino acids) of the HMG domains from 20 mouse Sox proteins are aligned according to their relationships. The conservation levels of the amino acid sequences are shown below the alignment with sequence logos. (C) The domain structures of mouse Sox proteins, showing representatives of the groups. The basic structures are highly conserved among members of each group.
the N-terminal side of the HMG domain (Bernard et al., 2003; Sock et al., 2003; Stolt and Wegner, 2010).

The SoxD group, which comprises Sox5, Sox6 and Sox13, is unique because SoxD proteins have a long N-terminal sequence containing a coiled-coil domain that allows dimerization with other SoxD proteins (Lefebvre, 2010).

Finally, mammal-specific SoxG (Sox15) (Meeson et al., 2007) and SoxH (Sox30) (Osaki et al., 1999) proteins are structurally related to SoxB1 and SoxD proteins, respectively (see Box 1).

**The regulation of Sox protein expression and activity**

Sox activity can be regulated at multiple levels. As discussed in detail below, Sox proteins function together with partner factors to elicit their action. However, the expression of Sox genes themselves is frequently subject to auto-regulation or control by other Sox proteins. Sox expression is also regulated post-transcriptionally by miRNAs (Peng et al., 2012; Xu et al., 2009b) (see Box 2). Sox protein function is known to be dose-dependent (Pevny and Nicolis, 2010), so modulating Sox protein levels is an important mode of regulation. In addition, Sox protein activity in a cell is modulated by covalent modification or by interaction with various proteins. Notably, Sox-dependent regulations intersect with signalling systems such as the sonic hedgehog (Shh) and Wnt signalling pathways, in which Sox-Gli and Sox-β-catenin interactions, respectively, are implicated (Bernard and Harley, 2010; Leung et al., 2011; Malki et al., 2010; Oosterveen et al., 2012; Peterson et al., 2012). Various post-translational modifications have also been reported to modulate the activity, stability and intracellular localization of Sox2, Sox9 and other Sox proteins. For example, studies on cultured cells have shown that Sox2 is subject to various covalent modifications such as phosphorylation (Jeong et al., 2010; Van Hoof et al., 2009), sumoylation (Tsuruzoe et al., 2006), acetylation (Baltus et al., 2009), methylation (Zhao et al., 2011) and glycosylation (Jang et al., 2012) (Fig. 2), although the biological significance of these modifications during embryogenesis remains to be confirmed. Interestingly, the regions in which Sox2 post-translational modifications can occur are located within two short amino acid stretches: one in the HMG domain and the other in the C-terminal domain, which suggests that these modifications may compete or cooperate with each other. For example, Sox2 is phosphorylated in two regions. Phosphorylation of a threonine in the HMG domain by Akt kinases enhances protein stability by inhibiting ubiquitin-mediated proteolysis (Jeong et al., 2010), whereas phosphorylation of serine residues in the C-terminal domain leads to phosphorylation-dependent sumoylation of the adjacent lysine residue (Van Hoof et al., 2009). Acetylation can also regulate Sox protein activity. Acetylation by p300 of a lysine residue close to the nuclear localization signal of Sry, and of one in the nuclear export signal region of Sox2, promotes nuclear import and nuclear export, respectively (Baltus et al., 2009; Thevenet et al., 2004). Because these two sites are highly conserved, these acetylation reactions may finely tune the subcellular localization of a broad range of Sox proteins.

In a few cases, the impact of Sox protein modifications has been demonstrated in developing embryos. Sry and Sox9 are phosphorylated by protein kinase A (PKA), which enhances their DNA-binding activity (Malki et al., 2010). This phosphorylation of Sox9 also leads to its translocation into the nucleus in cells of the testis (Malki et al., 2010), and is essential in neural crest (NC) cells for the Sox9-Snail interaction that promotes NC delamination (Liu et al., 2013). The functional relevance of sumoylation has been best demonstrated for SoxE proteins in NC development. SoxE sumoylation inhibits NC development, causing loss of pigmentation (Taylor and Labonne, 2005). Sumoylated SoxE

---

**Box 2. Modulation of Sox protein expression levels by microRNAs**

Sox (Sry-related HMG-box) protein expression levels are modulated post-transcriptionally by microRNAs (miRNAs; miRs), which repress the translation and/or promote the degradation of their target mRNAs. Several examples of such regulation are detailed below.

- **miR-145** expression is highly upregulated during the differentiation of human embryonic stem cells (ESCs). miR-145 directly targets 3' UTRs (untranslated regions) of the miRNAs of Sox2 and other ‘core pluripotency factors’, and promotes differentiation into mesoderm and ectoderm lineages (Xu et al., 2009b).

- **miR-200 family members** express at high levels in mouse ESCs and neural stem/progenitor cells. Sox2 is directly targeted by miR-200c, whereas Sox2 trans-activates the promoter of the miR-200c/141 gene, thereby forming a negative-feedback loop (Peng et al., 2012). This miR-200-mediated negative regulation may lead to gradual reduction in Sox2 levels during neural differentiation.

- **Dykyrprrtkhtlmk** Enhances self-association

- **Gsvkseassspvvtssshs** Enhances nuclear import (Sry)

- **Dykyprrktkhtlmk** Enhances stability

- **Gsvkseassspvvtssshs** Inhibits transcriptional activation

**Fig. 2. Post-translational modification of Sox2.** Post-translational modification sites of Sox2 (Sry-related HMG-box 2) are shown with their functional effects. Data obtained from studies of human and mouse Sox2 are combined on the mouse amino acid sequence. The nuclear export signal (NES, blue), nuclear localization signal (NLS, purple) and Group B homology region (striped red) are highlighted.

---
Mechanism of action: Sox proteins complex with partner factors

An important characteristic of Sox proteins is that they generally exhibit their gene regulatory functions only by forming complexes with partner transcription factors (Kamachi et al., 2000; Kondoh and Kamachi, 2010). Thus, a functional Sox-binding site is accompanied by a binding site for a second partner protein, which is required for Sox-dependent transcriptional regulation, and the binding of a single Sox protein alone to a DNA site does not lead to transcriptional activation or repression (e.g. Kamachi et al., 2001; Yuan et al., 1995) (Fig. 3A).

As Sox proteins in many cases interact with the partner factors in the absence of DNA, it is presumed that the Sox-partner complexes can form first and then recognize target DNA sites as a complex (Fig. 3A). For SoxB1/C/F proteins, their partner factors are heterologous transcription factors belonging to protein families such as the Pou and Pax families. SoxE proteins, by contrast, exhibit two modes of partner interactions: one that employs heterologous partners and another that requires homologous dimerization. SoxD proteins also dimerize via their coiled-coil dimerization motif. Partner factors for SoxH have not yet been determined.

Classic examples of the regulatory targets of Sox-partner complexes are the Fgf4 (fibroblast growth factor 4) enhancer, which is activated by Sox2 and Pou5f1/Oct4 (POU domain, class 5, transcription factor 1) in embryonic stem cells (ESCs) (Ambrosetti et al., 1997; Yuan et al., 1995), and the δ-crystallin DC5 enhancer, which is activated by Sox2 and Pax6 (paired box gene 6) during lens development (Kamachi et al., 1995; Kamachi et al., 2001). In both cases, the enhancers are activated by the co-binding of Sox2 and a partner factor, and mutations in the binding site for either Sox2 or the co-factor inactivate the enhancer.

The binding site sequences of the Sox2-partner complexes are not simply the sum of the consensus binding sequences that are determined using single factors (Fig. 3B,C). For example, the replacement of the Pax6-binding site within the DC5 sequence bound by the Sox2-Pax6 complex with the Pax6 consensus binding sequence inactivates the enhancer (Kamachi et al., 2001). (D) Sox9 homodimer-binding sequences found in the enhancers of collagen and the related genes indicated. In contrast to the heterodimers, the Sox9 dimer-binding sequences deviate from the Sox consensus in both sites (Bernard et al., 2003; Bridgewater et al., 2003; Dy et al., 2012; Han and Lefebvre, 2008, Sock et al., 2003). Acan, aggrecan; Col9a1, collagen, type IX, α1; Col10a1, collagen, type X, α1; Col11a2, collagen, type XI, α2; Dppa4, developmental pluripotency associated 4; Fbxo15, F-box protein 15; Fgf4, fibroblast growth factor 4; Pou5f1, POU domain, class 5, transcription factor 1; Utf1, undifferentiated embryonic cell transcription factor 1.
sequence (Bridgewater et al., 2003; Dy et al., 2012; Han and Lefebvre, 2008) (Fig. 3D). Evidence indicates that a complex of a Sox and its partner factor assumes a specific conformation to elicit transcriptional regulation upon binding to its unique target sequences (Kamachi et al., 2001).

Whether a Sox-partner complex elicits transcriptional activation or repression depends on whether the complex recruits co-activator or repressor. Indeed, recent proteomics analysis identified Trtrap co-activator (transformation/transcription domain-associated protein), Ncor/SMRT co-repressors (Ncor1/2, nuclear receptor co-repressors 1/2) and NuRD co-repressor complexes as strongly interacting co-factors of Sox2 (Engelen et al., 2011), suggesting that Sox2 can participate in transcriptional activation or repression depending on the co-factor recruited by the Sox2-partner factor complex.

A popular way of characterizing the transactivation/repression potential of a transcription factor is to excise regions of its non-DNA-binding domains and fuse them to the DNA-binding domain of another protein, such as that of the GAL4 protein, which binds to DNA as a dimer. If the resulting fusion protein can recruit co-activators in this context, it will activate a reporter gene, and the protein region is registered as an ‘activation domain’ that can recruit a co-activator under a certain molecular context. A ‘repression domain’ that can recruit a co-repressor is determined analogously. The various domains of Sox proteins indicated in Fig. 1 have been determined in this fashion. For example, SoxB1 proteins harbouring an activation domain participate in the transactivation process, e.g. those occurring in the Fgf4 and δ-crystallin enhancers (Ambrosetti et al., 1997; Kamachi et al., 2001; Yuan et al., 1995), whereas SoxB2 proteins with a repression domain participate in transrepression processes (Bylund et al., 2003; Uchikawa et al., 1999).

However, it is important to emphasize that these DNA-binding domain fusion assays do not represent the full repertoire of the regulatory potential of a transcription factor or a Sox-partner complex. For example, the interaction of Sox2 with co-repressor (Engelen et al., 2011) was not indicated by using the fusion assay (Kamachi et al., 1999; Kamachi et al., 1998). Moreover, although SoxD proteins do not possess an activation/repression domain based on the DNA-binding domain fusion assay, Sox6 is known to interact with the co-repressor CtBP (C-terminal binding protein) and to repress Fgf3 expression during inner ear development (Murakami et al., 2001). It has also been demonstrated that Sox9 has both activation and repression modes, depending on the target site (and hence the partner factor), and in the repression mode Sox9 recruits Gli proteins as the partner factor (Leung et al., 2011). Furthermore, Sox9 interacts with Runx2 (runt related transcription factor 2) and inhibits Runx2-dependent transactivation of osteoblast-specific genes (Zhou et al., 2006). Therefore, the transcriptional regulatory activity of a Sox-partner complex needs to be assessed in individual contexts of the cells and target DNAs.

**Sox-partner exchange allows stepwise progression of developmental processes**

Some Sox-partner complexes can activate the expression of their own respective genes, which stabilizes the cell state (Kondoh and Kamachi, 2010; Kondoh et al., 2004), as seen in the case of Sox2 and Pou5f1/Oct4 (Fig. 4A). However, the replacement of either component of the complex can result in large-scale changes in target genes, which in turn can contribute to developmental progression. Two representative cases of Sox-dependent developmental regulation are discussed below.

**Fig. 4. Paradigms of developmental regulation by Sox-partner complexes.** (A) Sox-partner complexes can activate cell-specific genes and their own genes, thereby stabilizing the cell state. Sox2-Pou5f1/Oct4 in ESCs is shown. Sox2- and Pax6-dependent lens cell development (not shown) is also reliant on Sox2-Pax6-co-regulatory loops. (B) Sox-partner complexes can also activate the gene for the partner protein required during the next developmental step, thereby promoting a domino-like progression of the developmental process. Sox10-Pax3 interactions, which then activate Mitf expression and Sox10-Mitf complex formation, during melanocyte development from the neural crest are shown. Schwann cell development (not shown) is also a representative of this paradigm, as discussed in the main text. (C) Sox-partner complexes can activate a second Sox gene that acts in the following step employing the same partner factor, thereby promoting the developmental process. The case of Sox9 induction by Sry-Sf1 during male gonad development is shown. (D) Sox trios in cases. Although Sox9 often employs itself as its own partner and dimerizes, Sox9 dimers further require SoxD (Sox5/6) dimers that bind to nearby DNA sites for their activity in many occasions of chondrocyte development. An example of Acan enhancer regulation is shown. The nature of Sox9-Sox5/6 interaction has not been clarified. Amt, anti-Mullerian hormone; Dct, dopachrome tautomerase; Mitf, microphthalmia-associated transcription factor; Pax3, paired box gene 3; Pgd, prostaglandin D synthase; Pou5f1, Pou domain, class 5, transcription factor 1; Sf1, steroidogenic factor 1; Sox, SRY-related HMG-box.

In the first case, a Sox-partner pair activates the gene for another transcription factor, which then functions as a new Sox partner (Fig. 4B). For example, during melanocyte development from the NC, a Sox10-Pax3 pair activates expression of the transcription factor Mitf (microphthalmia-associated transcription factor), which then acts as a Sox10 partner to activate the Dct (dopachrome tautomerase) and Tyr (tyrosinase) genes involved in melanocyte differentiation (Bondurand et al., 2000; Ludwig et al., 2004; Murisier et al., 2007). In the second case, a Sox factor changes while the same partner factor is shared, as exemplified by the Sry-Sox9 switch during gonadal development (Fig. 4C).

Another example of such a developmental switch is seen during neural tube development (Fig. 5A). In embryonic neural tubes, neural progenitor cells in the ventricular zone express SoxB1 proteins, while the postmitotic neuronal precursors in subventricular zone express SoxC proteins (Bergsland et al., 2011; Tanaka et al., 2004; Uwanogho et al., 1995). The mechanism
underlying this SoxB1-SoxC switch has not been characterized. SoxB1 and SoxC proteins both form complexes with Pou3f factors (e.g. Pou3f2/Bm2), which are expressed and nuclear localized in the ventricular and subventricular zones. The switching of Sox expression changes the gene expression profile of the cells from the SoxB1-Pou3f target group to the SoxC-Pou3f target group. These sets of target genes show some overlap, e.g. Nes (nestin) (Tanaka et al., 2004), resulting in graded, rather than abrupt, changes between cells.

**Developmental functions of Sox group proteins**

The aforementioned modes of Sox protein-dependent regulation are found in various developmental processes, and there are strong associations between individual Sox groups and specific cell lineages. It is often observed that the members of the same Sox group, sharing equivalent functions, are expressed in the same developing tissues (e.g. SoxB1 proteins in the CNS, SoxE proteins in the NC and SoxP proteins in the vascular system), although their precise spatial and temporal tissue expression patterns differ slightly. This situation creates functional redundancy within the group members that co-regulate the same targets, which safeguard the developmental processes against genetic variations between individual animals. However, the contribution of individual members of a Sox group to a developmental process is not always equivalent, with regard to the timing of expression, expression levels and the activity of a protein, as exemplified by the more minor contribution by Sox8 among SoxE proteins in the development of NC-derived tissues (Kellerer et al., 2006; Stolt et al., 2005). Documented cases of Sox-dependent gene regulation in various developmental contexts are listed in Table 1, highlighting Sox-partner interactions. It is interesting that the members of the same Sox group tend to interact with the same partner factor. even in different lineages, e.g. Sox10-Mef2c in NC versus Sox9-Mef2c in chondrogenesis. Mutant and tissue-specific knockout mouse phenotypes, as well as the congenital disorders in humans associated with Sox protein function, are also indicated.

Representative cases of cell lineage-specific regulation by particular Sox groups are summarized below.

**The Sox2-Pou5f1/Oct4 complex in embryonic stem cells and the epiblast**

Many ESC-expressed genes, such as *Fgf4*, are regulated by the Sox2-Pou5f1/Oct4 complex (Boyer et al., 2005; Chen et al., 2008; Kondoh and Kamachi, 2010). As these genes include Sox2 and Pou5f1 themselves, it appears that ESC states are stabilized by an autoregulatory loop. In addition, Sox2 and Pou5f1/Oct4 are two of the essential four transcription factors initially identified for induced pluripotent stem cell (iPSC) production (Takahashi and Yamanaka, 2006), indicating their pivotal contribution to the ESC state. Sox2-Pou5f1/Oct4 complex function is also essential for the formation of the inner cell mass of pre-implantation embryos, from which ESCs are derived (Avilion et al., 2003), and of the epiblast of the post-implantation embryo. However, this complex is downregulated in embryos after somatic differentiation is initiated (Iwafuchi-Doi et al., 2012). In zebrafish embryos, the SoxB1-Pou5f1 complex also regulates early developmental processes, including embryo patterning (Okuda et al., 2010; Onichtchouk et al., 2010; Shih et al., 2010).

**SoxB1 proteins in neural development**

The neural plate (the embryonic neural progenitor tissue) expresses Sox2 (Kishi et al., 2000; Takemoto et al., 2011; Uwanogho et al., 1995), and this expression persists in all neural progenitor/stem cells throughout life (Pevny and Placzek, 2005; Pevny and Nicolis, 2010; Suh et al., 2007). Presumably, the same Sox2 protein recruits different partner proteins during development that fulfill context-dependent regulatory functions, in addition to Pou3f proteins, which are common Sox2 partners during neural development (Tanaka et al., 2004). SoxB1 proteins maintain the neural stem/progenitor state of cells by preventing the transition to a more advanced developmental stage that is marked by SoxC expression (Graham et al., 2003). In addition, it is possible that SoxB1 proteins...
Table 1. Examples of Sox factor participation in the developmental processes

<table>
<thead>
<tr>
<th>Sox group</th>
<th>Tissues</th>
<th>Participating Sox</th>
<th>Known partner factors*</th>
<th>Known target genes*</th>
<th>Phenotypes of Sox KO/tissue-specific KO/hypomorphs</th>
<th>Associated congenital human disease</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Gonad</td>
<td>Sry</td>
<td>Sox9</td>
<td></td>
<td>Male gonad defect</td>
<td>XY gonad dysgenesis</td>
<td>(Sekido and Lovell-Badge, 2008)</td>
</tr>
<tr>
<td>B1</td>
<td>ICM, ESCs, iPSCs, epiblast</td>
<td>Sox2</td>
<td>Pou5f1 (Oct4)</td>
<td>Sox2, Pou5f1, Nanog, Fgf4, Uf1, Fbxo15, Lefty1, Mycn, Tcf3, Dppa4, Trp53, Rest, Zfp42</td>
<td>Failure in implantation (KO)</td>
<td>Septo-optic dysplasia</td>
<td>(Boyer et al., 2005; Chen et al., 2008; Kondoh and Kamachi, 2010)</td>
</tr>
<tr>
<td>B1</td>
<td>Embryonic neural tube</td>
<td>Sox2/3/1</td>
<td>Pou3fs</td>
<td>Nes, Shh, Sox2</td>
<td>Neural hypoplasia, abnormal thalamus, retinal defects</td>
<td>Septo-optic dysplasia</td>
<td>(Favaro et al., 2009; Miyagi et al., 2006; Tanaka et al., 2004; Zhao et al., 2012)</td>
</tr>
<tr>
<td>B1</td>
<td>Ventral neural tube</td>
<td>Sox2/3/1</td>
<td>Activating form Gli proteins</td>
<td>Nkx2.2, Olig2, Nkx6.1/6.2</td>
<td>Septo-optic dysplasia</td>
<td>(Oosterveen et al., 2012; Peterson et al., 2012)</td>
<td></td>
</tr>
<tr>
<td>B1</td>
<td>Neural progenitors in adult brain</td>
<td>Sox2</td>
<td>Pou3fs</td>
<td>Shh</td>
<td>Loss of neural progenitor/stem cells</td>
<td></td>
<td>(Favaro et al., 2009; Zhao et al., 2012)</td>
</tr>
<tr>
<td>B1</td>
<td>Neural stem cells derived from ESCs</td>
<td>Sox2</td>
<td>Chd7</td>
<td>Jag1, Tulp3, Hes5, Mycn</td>
<td>AEG (anophthalmia-esophageal-genital) syndrome</td>
<td></td>
<td>(Engelen et al., 2011; Williamson et al., 2006)</td>
</tr>
<tr>
<td>B1</td>
<td>Pituitary</td>
<td>Sox2/3</td>
<td></td>
<td></td>
<td>Hypogonadism caused by pituitary defect (Sox3)</td>
<td>Hypogonadism caused by pituitary defect (SOX2)</td>
<td>(Kelberman et al., 2006; Rizzoti et al., 2004; Sato et al., 2007)</td>
</tr>
<tr>
<td>B1</td>
<td>Lens</td>
<td>Sox2/1</td>
<td>Pax6, Pou2f1 (previously Oct1), Maf1</td>
<td>Sox2, Pax6, crystallins, N-cadherin</td>
<td>Lens defects</td>
<td>Anophthalmia, cataractus lens</td>
<td>(Aota et al., 2003; Donner et al., 2007; Inoue et al., 2007; Kamachi et al., 2001; Matsumata et al., 2005; Rajaram and Kerppola, 2004; Smith et al., 2009)</td>
</tr>
<tr>
<td>C</td>
<td>Early embryonic cells</td>
<td>Sox4/11</td>
<td>Tead2</td>
<td></td>
<td>Organ hypoplasia</td>
<td></td>
<td>(Bhattaram et al., 2010; Schilham et al., 1996; Sock et al., 2004)</td>
</tr>
<tr>
<td>C</td>
<td>Developing neurons</td>
<td>Sox4/11</td>
<td>Tubb3</td>
<td></td>
<td></td>
<td></td>
<td>(Bergsland et al., 2006)</td>
</tr>
<tr>
<td>C</td>
<td>Kidney</td>
<td>Sox11</td>
<td>Wt1</td>
<td>Wnt4</td>
<td>Aberrant histogenesis of neocortex</td>
<td></td>
<td>(Murugan et al., 2012; Azim et al., 2009; Kwan et al., 2008; Lai et al., 2008)</td>
</tr>
<tr>
<td>D</td>
<td>Neocortex neurons</td>
<td>Sox5/6</td>
<td>Sox5/6</td>
<td></td>
<td>Precocious oligodendrocyte differentiation</td>
<td></td>
<td>(Stolt et al., 2002; Stolt et al., 2006)</td>
</tr>
<tr>
<td>D</td>
<td>Oligodendrocytes</td>
<td>Sox5/6</td>
<td>Sox5/6</td>
<td>Myelin basic protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>Chondrocytes</td>
<td>Sox5/6</td>
<td>Sox5/6, Sox9</td>
<td>Collagens, Acan</td>
<td>Chondrogenesis defects</td>
<td></td>
<td>(Lefebvre et al., 1998; Smits et al., 2001)</td>
</tr>
<tr>
<td>E</td>
<td>Hair bulge</td>
<td>Sox9</td>
<td></td>
<td></td>
<td>Alopecia</td>
<td>Hypertrichosis (Sox9 dysregulation)</td>
<td>(Fantauzzo et al., 2012; Vidal et al., 2005)</td>
</tr>
<tr>
<td>E</td>
<td>Otic placode</td>
<td>Sox9</td>
<td></td>
<td></td>
<td>Otic placode invagination defect</td>
<td></td>
<td>(Barrionuevo et al., 2008)</td>
</tr>
<tr>
<td>E</td>
<td>Neural stem cells</td>
<td>Sox9</td>
<td></td>
<td></td>
<td>Failure of neural stem maintenance</td>
<td></td>
<td>(Cheng et al., 2009; Scott et al., 2010; Stolt et al., 2003)</td>
</tr>
<tr>
<td>E</td>
<td>Glial progenitors</td>
<td>Sox9</td>
<td>Nf1a</td>
<td></td>
<td>Impaired gliogenesis</td>
<td></td>
<td>(Kang et al., 2012; Stolt et al., 2003)</td>
</tr>
</tbody>
</table>

Table 1. Continued on next page
<table>
<thead>
<tr>
<th>Sox group</th>
<th>Tissues</th>
<th>Participating Sox</th>
<th>Known partner factors*</th>
<th>Known target genes*</th>
<th>Phenotypes of Sox K/O/tissue-specific KO/hypomorphs</th>
<th>Associated congenital human disease</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>Oligodendrocyte progenitors/ oligodendrocytes</td>
<td>Sox9/10</td>
<td>Suppressor of fused, myelin basic protein</td>
<td>Loss of oligodendrocytes</td>
<td>(Pozniak et al., 2010; Stolt et al., 2003; Stolt et al., 2002; Stolt et al., 2006)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>Neural crest</td>
<td>Sox9</td>
<td>Snail2</td>
<td>Snail2</td>
<td>Loss of neural crest delamination</td>
<td>(Liu et al., 2013; Sakai et al., 2006)</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>Neural crest</td>
<td>Sox9</td>
<td>Ets1, Myb</td>
<td>Sox10</td>
<td></td>
<td>(Betancur et al., 2010; Cheung and Briscoe, 2003)</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>Cranial neural crest</td>
<td>Sox10</td>
<td>Mef2c</td>
<td></td>
<td>Waardenburg-Hirschprungs disease</td>
<td>(Agarwal et al., 2011)</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>Gonad</td>
<td>Sox9</td>
<td>SF1, Sox9</td>
<td>Amh, Pgds</td>
<td>Male gonad defect</td>
<td>XY gonad dysgenesis</td>
<td>(De Santa Barbara et al., 1998; Sekido and Lovell-Badge, 2008; Wilhelm et al., 2007)</td>
</tr>
<tr>
<td>E</td>
<td>Chondrocytes</td>
<td>Sox9</td>
<td>Sox9, Sox5/6</td>
<td>Collagen genes, Acan</td>
<td>Chondrogenesis defects</td>
<td>Campomelic dysplasia</td>
<td>(Bell et al., 1997; Bi et al., 2001; Han and Lefebvre, 2008; Lefebvre et al., 1997; Leung et al., 2011; Zhao et al., 1997)</td>
</tr>
<tr>
<td>E</td>
<td>Chondrocytes</td>
<td>Sox9</td>
<td>Mef2c</td>
<td>Col10a1 (activation)</td>
<td>Chondrogenesis defects</td>
<td>Campomelic dysplasia</td>
<td>(Dy et al., 2012)</td>
</tr>
<tr>
<td>E</td>
<td>Chondrocytes</td>
<td>Sox9</td>
<td>Gli2/3</td>
<td>Col10a1 (repression)</td>
<td>Chondrogenesis defects</td>
<td>Campomelic dysplasia</td>
<td>(Leung et al., 2011)</td>
</tr>
<tr>
<td>F</td>
<td>Endoderm</td>
<td>Sox17/7</td>
<td>Pou5f1 (previously Brn2), Egr2</td>
<td></td>
<td>Gut defects</td>
<td></td>
<td>(Aksoy et al., 2013; Kanai-Azuma et al., 2002)</td>
</tr>
<tr>
<td>F</td>
<td>Vascular endothelium</td>
<td>Sox18/17</td>
<td>Mef2c</td>
<td>Vcam1, N-cadherin, Mmp7</td>
<td>Vascular defects</td>
<td>Hypotrichosis-lymphoedema-telangiectasia syndrome</td>
<td>(Downes et al., 2009; Hoeth et al., 2012; Hosking et al., 2001; Hosking et al., 2004)</td>
</tr>
<tr>
<td>F</td>
<td>Lymphatic endothelium</td>
<td>Sox18</td>
<td>Prox1</td>
<td></td>
<td>Lymphatic defects</td>
<td>Hypotrichosis-lymphoedema-telangiectasia syndrome</td>
<td>(Francois et al., 2008)</td>
</tr>
<tr>
<td>F</td>
<td>Dermal papilla of hair follicle</td>
<td>Sox18</td>
<td></td>
<td></td>
<td>Reduction of zigzag hair (null), alopecia (dominant negative)</td>
<td>Hypotrichosis-lymphoedema-telangiectasia syndrome</td>
<td>(Pennisi et al., 2000a; Pennisi et al., 2000b)</td>
</tr>
<tr>
<td>G</td>
<td>Muscle satellite cells</td>
<td>Sox15</td>
<td>Fhl3</td>
<td>Foxk1</td>
<td>Muscle regeneration defect</td>
<td></td>
<td>(Meeson et al., 2007)</td>
</tr>
<tr>
<td>H</td>
<td>Male germ cells</td>
<td>Sox20</td>
<td></td>
<td></td>
<td></td>
<td>(Osaki et al., 1999)</td>
<td></td>
</tr>
</tbody>
</table>

*The partner factors and the target genes do not necessarily correspond. In some cases, interacting partners are known without defined target genes, whereas in other cases target genes have been identified without the knowledge of partner factors.

Amh, anti-Mullerian hormone; Atoh1, atonal homologue 1; Chd7, chromodomain helicase DNA binding protein 7; Col10a1, collagen, type X, α1; Dct, dopachrome tautomerase; Dppa4, developmental pluripotency associated 4; Egr2, early growth response 2; Ets1, embryonic stem cells; Ets1, E26 avian leukemia oncogene 1, 5 domain; Eya1, eyes absent 1 homolog; Fbxo15, F-box protein 15; Fgf4, fibroblast growth factor 4; Fhl3, four and a half LIM domain 3; Foxk1, forkhead box K1; Hes5, hairy and enhancer of split 5; Icm, inner cell mass; IpsiCs, induced pluripotent stem cells; Jag1, jagged 1; KO, knockout; Lefty1, left right determination factor 1; Maf1, MAF1 homologue; Mef2c, myocyte enhancer factor 2C; Mef2d, myocyte enhancer factor 2D; Mmp7, matrix metallopeptidase 7; Mysn, myeloblastosis oncogene; Nkx2, nk2 transcription factor; POU3F1, POU domain, class 5, transcription factor 1; Pou5f1, POU domain, class 5 transcription factor (which comprise Pou5f1 (Oct6), Pou5f2 (Brn2), Pou5f3 (Brn3) and Pou5f4 (Brn4)); Pou5f1, POU domain, class 5, transcription factor 1; Prx1, prospero-related homeobox 1; Rett, Rett-silencing transcription factor; Ret, Ret proto-oncogene; Sf1, steroidogenic factor 1; Shh, sonic hedgehog; Six1, sine oculis-related homeobox 1; Sox9, SRY-related HMG-box; Tcf3, transcription factor 3; Tubb3, tubulin β3 class II; Tulp3, tubby-like protein 3; Tyr, tyrosinase; Uf1, undifferentiated embryonic cell transcription factor 1; Vcam1, vascular cell-adhesion molecule 1; Wnt4, wingless-related MMTV integration site 4; Wt1, Wilms' tumour 1; Zfp42, zinc-finger protein 42.
also support the survival of neural stem/progenitor cells, as a large decrease in Sox8B1 activity results in the massive neural tissue degeneration (Ferri et al., 2004). An important regulatory target of Sox2 in the brain is the Shh gene, the activity of which is essential for stem cell maintenance in the hippocampus and for hypothalamus development (Favaro et al., 2009; Zhao et al., 2012). The downstream effectors of the hedgehog (Hh) pathway, the Gli proteins, in turn act as Sox2 partners during the Hh signal-dependent activation of transcription factor genes, such as Nkx2.2 (NK2 homeobox 2), Olig2 (oligodendrocyte transcription factor 2), Nkx6.1 and Nkx6.2, which are characteristic of ventral spinal cord neural progenitors (Oosterwegel et al., 2012; Peterson et al., 2012). This represents a paradigm for signal-dependent activation of Sox-partner complexes (Fig. 5B).

The expression of Sox3, another Sox8B1 protein, largely overlaps with that of Sox2 in neural and sensory tissues (Dee et al., 2008; Nishimura et al., 2012; Okuda et al., 2006; Uchikawa et al., 2011; Wood and Episkopou, 1999). Sox3 itself is dispensable in mice (Rizzoti et al., 2004; Weiss et al., 2003), but it safeguards Sox2 functions by functional redundancy (Iwafuchi-Doi et al., 2011; Rizzoti and Lovell-Badge, 2007). Sox1, the third Sox8B1, is first activated weakly in a part of the developing anterior neural plate, then strongly in the developing spinal cord, and finally throughout the neural tube in mouse and chicken embryos (Cajal et al., 2012; Uchikawa et al., 2011; Wood and Episkopou, 1999). The functions of Sox1 are largely shared with those of Sox2 and Sox3, although it has a unique function during the migration of neuronal precursors in the ganglionic eminence (Ekonomou et al., 2005) and in the lens (Nishiguchi et al., 1998).

Other Sox proteins in the nervous system
Sox21, a Sox8B2 protein with transcriptional repression activity, is often co-expressed with Sox2 in the neural, sensory and endodermal tissues (Uchikawa et al., 1999). The repressive activity of Sox21 modifies the balance between progenitor cell maintenance and the progression to postmitotic neural development, during which time Sox8C activity is involved (Graham et al., 2003; Sandberg et al., 2005). In the adult hippocampus, Sox21 represses the expression of the transcriptional repressor Hes5 to promote neurogenesis (Matsuda et al., 2012). Sox9 is also expressed in the developing CNS, and is essential for the establishment and maintenance of multipotent neural stem cells (Cheng et al., 2009; Scott et al., 2010). During the late stages of neural development, Sox9 regulates gliogenesis in neural stem cells (Stolt et al., 2003) using NFIA as its partner protein (Kang et al., 2012). Within the glial cells, oligodendrocyte development is redundantly supported by co-expressed Sox9 and Sox10 (Stolt et al., 2003), which repress suppressor of fused (Sufu) expression (Pozniak et al., 2010).

SoxE proteins during neural crest development
Neural crest cells delaminate from the dorsal border of closing neural tube and give rise to the peripheral nervous system and melanocytes, and (in the cephalic region) to muscle and bone precursors as well. The SoxE family member Sox9 is expressed initially in the NC-competent cell population present in the dorsal neural tube (Cheung and Briscoe, 2003). As these cells migrate from the neural tube, the expression of Sox10, the major NC regulator, is activated in a Sox9-dependent manner (Betancur et al., 2010). Sox9 cooperates with Ets1 and cMyb to activate Sox10 in the cranial NC (Betancur et al., 2010). Sox10 first interacts with the following partner factors: Metf2e for the cranial NC (Agarwal et al., 2011); Pax3 for melanocytes (Bondurand et al., 2000); and Pou3f1/2 for Schwann cell development (Kuhlbrodt et al., 1998). These Sox10-partner complexes then activate a second set of Sox10 partner genes, such as Mitf for melanocytes (Bondurand et al., 2000) and Egr2 for Schwann cells (Ghislain and Charnay, 2006; Reiprich et al., 2010), thereby activating the genes characteristic of their respective terminally differentiated cells, e.g. connexin 32 (Gjb1) and myelin protein zero during Schwann cell development (Bondurand et al., 2001; LeBlanc et al., 2007).

Sox2 in the sensory placodes
Sensory placodes (i.e. the nasal, lens, otic and taste bud placodes, as well as the hypophyseal placode), which are the precursors of sensory tissues, develop from the cephalic ectoderm and express Sox2 as a key regulator (Okubo et al., 2006; Schlosser et al., 2008; Uchikawa et al., 2011; Wood and Episkopou, 1999). All of these placodes express N-cadherin (cadherin 2, Cdh2) in a Sox2-dependent fashion during their morphogenesis (Matsumata et al., 2005). Pax6 and Pou2f1 (Oct1) act as Sox2 partners during nasal and lens placode development (Donner et al., 2007; Kamachi et al., 1998; Kamachi et al., 2001). The hypophyseal placode also expresses Sox3, and Sox3 deficiency causes pituitary defects in mice, consistent with an important role for Sox3 in this tissue (Rizzoti et al., 2004).

The otic placodes express Sox9 and Sox2 (Mak et al., 2009), both of which play specific roles during inner ear development. Sox9 is involved during early otocyst development (Saint-Germain et al., 2004), and its mutational disruption in mouse leads to the failure of otic placode invagination (Barriouneau et al., 2008). Sox2 is involved in the production and maintenance of the neural progenitors in the otocysts, i.e. the sensory patches. Inner ear-specific loss of Sox2 expression in mutant mice results in the loss of neural components: hair cells and supporting cells (Kiernan et al., 2005). During hair cell development, Sox2 cooperates with the Six1-Eya1 complex to activate the bHLH protein gene Atoh1 (Ahmed et al., 2012; Neves et al., 2011; Neves et al., 2012). Sox2 and Atoh1 then cooperate to prime cells for hair cell development, which is initiated upon Sox2 downregulation (Dabdoub et al., 2008). An analogous scenario occurs during the development of vestibulo-cochlear ganglion neurons, where the bHLH proteins are neurogenin 1 and Neurod1 (Evans et al., 2013; Puligilla et al., 2010).

SoxG (Sox15) in skeletal muscle regeneration
Sox15-null mice are grossly normal, but show defects in skeletal muscle regeneration from satellite cells, which are muscle stem cells for regeneration (Lee et al., 2004; Maruyama et al., 2005). It has been shown that Sox15 recruits Fh3 (four and a half LIM domains 3) as a partner factor to regulate the Fok1 gene (forkhead box protein K1), which is essential for the cell cycle progression of myogenic progenitor cells (Meeson et al., 2007).

The Sox trio (Sox9 and Sox5/6) in chondrogenesis during skeletal development
The crucial role of Sox9 during skeletal development was first indicated when heterozygous human Sox9 mutations were identified as the cause of campomelic dysplasia, a skeletal dysmorphology syndrome associated with male-to-female sex reversal (Foster et al., 1994; Wagner et al., 1994). Sox9-heterozygous mouse mutants die perinatally because of skeletal abnormalities (Bi et al., 2001). Sox9 is highly expressed throughout the chondrocyte lineage, from mesenchymal chondroprogenitors to
hypertrophic chondrocytes, exhibiting an expression profile similar to that of the type II collagen gene Col2a1 (Ng et al., 1997; Zhao et al., 1997), and it is essential throughout these developmental stages (Dy et al., 2012). Sox9 binding to the Col2a1 intronic enhancer is required for the activation of the Col2a1 gene (Bell et al., 1997; Lefebvre et al., 1997), but strong activation of the 48 bp Col2a1 minimal enhancer occurs only when Sox9 and the SoxD proteins Sox5 and Sox6 (Sox5/6) together bind to several sites in the enhancer (Lefebvre et al., 1998). The same combination of Sox9 and Sox5/6, known as the ‘Sox trio’ is involved in the activation of many extracellular matrix genes, e.g. aggrecan (Acan) and Col11a2, which are secreted by the chondrocytes (Han and Lefebvre, 2008). This ‘Sox trio’ primarily functions as combination of a Sox9 homodimer and multiple Sox5/6 dimers that bind to adjacent DNA sites (Fig. 4D). Indeed, individuals with mutations in the SOX9 dimerization domain develop campomelic dysplasia without causing sex reversal (Bernard et al., 2003; Sock et al., 2003). In line with this, single knockouts of Sox5 or Sox6 in mice lead to mild skeletal abnormalities, whereas a Sox5/6 double knockout leads to a lack of chondrogenesis despite normal Sox9 expression (Smits et al., 2001), thereby confirming the functional cooperation of SoxD and SoxE proteins during chondrogenesis. The molecular mechanism underlying the SoxD-SoxE cooperation has not been clarified.

Sry and Sox9 in sex determination and testis development
Following the discovery of SRY/Sry (Gubbay et al., 1990; Sinclair et al., 1990), the involvement of Sox9 in male sex determination was recognized on the basis of genetic analysis of human campomelic dysplasia, as mentioned above, in which the majority of XY individuals with heterozygous SOX9 mutations exhibit male-to-female sex reversal (Foster et al., 1994; Wagner et al., 1994). In mouse embryos, Sry is expressed transiently between E10.5 and E12.5 in the somatic supporting cells of the undetermined gonads, which triggers the differentiation of these cells into Sertoli cells during subsequent development. Sry directly activates Sox9 synergistically with steroidogenic factor 1 (Sf1, Nr5a1) via the testis-specific enhancer of Sox9 (Tes) (Sekido and Lovell-Badge, 2008). After Sox9 expression has been activated beyond a crucial level, its expression is maintained in Sertoli cells via positive-feedback loops, including autoregulation, whereby Sox9 activates the Tes enhancer cooperatively with Sf1. Sox9 then activates other genes that play crucial roles during testis development, such as Amh (anti-Müllerian hormone) and Pdgds (prostaglandin D synthase) (Fig. 4C); Amh is activated by the cooperative action of Sox9 and Sf1 (De Santa Barbara et al., 1998), while Pdgds is activated by the Sox9 dimer (Wilhelm et al., 2007). Another SoxE gene, Sox8, is also activated in the Sertoli cells following Sox9 expression, and plays essential roles in the maintenance of testicular functions at later stages (Barriónuevo et al., 2009).

SoxF proteins in the initiation of endoderm development
The involvement of SoxF proteins in endoderm development was first indicated in Xenopus (Hudson et al., 1997). In early mouse embryos, the SoxF protein Sox7 is expressed in the primitive endoderm, whereas another SoxF protein, Sox17, is expressed primarily in the definitive endoderm. During the derivation of the endoderm from the epiblast, a Sox2-Pou5f1-Oct4 complex switches to a Sox17-Pou5f1/Oct4 pair (Aksoy et al., 2013), analogous to the case shown in Fig. 5A. Sox7/17-expressing cells intercalate with each other when the definitive endoderm is formed (Kwon et al., 2008). Sox17 knockout mouse embryos develop severe defects in gut tube formation (Kanai-Azuma et al., 2002). In fish species, endodermal Sox17 is under the regulation of Sox32 (also known as ‘casanova’), a fish-specific SoxF (Dickmeis et al., 2001; Kikuchi et al., 2001). During later endodermal development, Sox2 also participates in the regulation of foregut and trachea in mice (Que et al., 2009; Que et al., 2007).

SoxF proteins in vascular/lymphatic development
The SoxF proteins Sox7/17/18 are all expressed in endothelial cells during mouse vascular development. The first insight into SoxF functions during vascular development was provided by the spontaneous mouse mutation Ragged, which results in the synthesis of Sox18 with a truncated C-terminal domain (Pennisi et al., 2000b). Homozygous Ragged embryos exhibit severe cardiovascular defects, as well as hair follicle defects (discussed below), presumably owing to dominant-negative effects over other SoxF and related Sox proteins (Pennisi et al., 2000b). By contrast, Sox18-null knockout mice are viable and exhibit only minor coat defects (Pennisi et al., 2000a). Sox17 knockout mouse embryos exhibit cardiovascular defects, whereas Sox17/18 double knockout embryos have more severe defects than Sox17 single knockout in regions where Sox7 is only weakly expressed. Sox7-null embryos die around E14.5 due to cardiovascular defects (Wat et al., 2012). These results support the overlapping functions of Sox7/17/18 (Sakamoto et al., 2007). It is thought that these SoxF proteins play important roles during the establishment and maintenance of the integrity of blood vessels (Downes et al., 2009), employing Mef2c (myocyte enhancer factor 2C) as a partner factor (Hosking et al., 2001) and by regulating Vcam1 (vascular cell-adhesion molecule 1) (Hosking et al., 2004), N-cadherin and Mmp7 (metalloprotease 7) expression (Hoeth et al., 2012).

Sox proteins and hair follicle development
The hair follicle, the site of hair growth, comprises two tissue components, the dermal papilla of condensed mesenchyme and the epithelial hair bulb cells that serve as the stem cells of hair shaft, which interact with each other to control overall hair growth. Sox2 and Sox18 play major regulatory roles in the dermal papilla and Sox9 in the hair bulb (Clavel et al., 2012; Driskell et al., 2009; Pennisi et al., 2000b). In the skin on the back of the mouse, four different hair types exist, which are produced in three waves during embryogenesis: guard hairs during the first wave (~E14.5), awl and auchen hairs during the second wave (~E16.5), and zigzag hairs during the third wave (E18.5). Sox2 is expressed in the developing papillae in the first and second waves, and in the papillae of guard/awl/auchen hairs throughout life, but not in the papillae of zigzag hairs (Driskell et al., 2009). Hair follicle-specific knockout of Sox2 results in growth impairment of the non-zigzag hairs (Clavel et al., 2012). This effect of Sox2 inactivation is partly attributed to loss of the expression of a BMP antagonist sclerostin domain-containing protein 1 (Sostdc1), a direct regulatory target of Sox2 (Clavel et al., 2012). All embryonic dermal papillae express Sox18, but Sox18-null mutant mice show only a reduction of the
zigzag hair type (Pennisi et al., 2000a). This difference in Sox2 dependence and Sox18 dependence distinguishes non-zigzag and zigzag hairs. However, expression of the dominant-negative form of Sox18 (Ragged) in the heterozygous condition causes the loss of achaene hairs in addition to zigzag hairs, and in the homozygous condition leads to the entire loss of hairs (Pennisi et al., 2000b). This observation suggests inhibition of the function of related proteins, possibly Sox2, by dominant-negative Sox18. In the hair bulb compartment, Sox9 plays a major role in hair development, and skin-specific knockout of Sox9 results in the loss of hair shaft stem cells, causing alopecia (Vidal et al., 2005), whereas loss of the Sox9 repressor Trps1 (a GATA-type transcription factor) causes hypertrichosis in humans (Fantauzzo et al., 2012). Additionally, Sox21 expressed in the hair bulb plays a role in regulating the hair cycle (Kiso et al., 2009).

**SoxC proteins have versatile functions**

The SoxC proteins Sox4/11/12 are expressed widely in embryonic tissues, with the highest expression levels found in neural and mesenchymal progenitor cells. Sox4 and Sox11 knockout mice develop multiple organ defects (Schilham et al., 1996; Sock et al., 2004), whereas Sox12 knockout mice are grossly normal (Hoser et al., 2008). However, various combinations of homo- and heterozygous SoxC mutations indicate that continued and cumulative activities of these three genes are required for widespread organ development; increasingly severe organ hypoplasia phenotypes were found with decreasing wild-type SoxC alleles (Bhattaram et al., 2010). Partially SoxC-deficient embryos (Sox11–/−, Sox4+/−Sox11−/−, etc.) also display malformation of various organs. For example, Sox4−/−Sox11−/− embryos are minute at E9.5 and show highly severe organ defects and extensive cell death. Nevertheless, these embryos appropriately express genes involved in cell lineages and embryo patterning, such as Pax1/7, Otx2 (orthodenticle homolog 2), Shh and Bmp4 (bone morphogenetic protein 4) (Bhattaram et al., 2010). This suggests that SoxC proteins play important roles during early embryogenesis in sustaining neural and mesenchymal progenitor cells, rather than in cell lineage specification. In support of this, a reported target of SoxC is Tead2 (TEA domain family member 2), which is a transcriptional mediator of the Hippo signalling pathway that is crucial for organ size regulation (Bhattaram et al., 2010).

However, during more advanced developmental stages, SoxC proteins regulate cell differentiation and tissue patterning. For example, following the initiation of neurogenesis in the CNS, SoxC protein expression becomes confined to post-mitotic differentiating neurons, which complements SoxB1 protein expression in neural progenitor cells (Tanaka et al., 2004). During these stages, a SoxC target is the neuron-specific tubulin gene Tubb3 (tubulin β3, class 3) (Bergsland et al., 2006). During *Xenopus* pronephros development, Sox11 and Wt1 (Wilms tumor 1 homolog) combine to regulate Wnt4, which is essential for nephric tissue development (Murugan et al., 2012).

**Genome-wide approaches to studying Sox proteins: new interacting partners, interacting sites and an emerging pioneer factor function**

The recent development of high-throughput proteomic/genomic approaches has enabled the systematic identification of Sox2-interacting proteins and Sox2-bound genomic sites. For example, Engelen et al. (Engelen et al., 2011), using neural stem cell lines, identified a series of Sox2-interacting proteins, including 19 transcription factors as candidate Sox2 partners, in addition to the Sox2-interacting co-activator and co-repressors discussed above. Among these transcription factors, chromodomain helicase DNA-binding protein 7 (Chd7) was investigated and was shown to bind close to genomic positions occupied by Sox2 in the vicinity of many genes. Moreover, knockdown of Sox2 reduced the occupancy of both Sox2 and Chd7, and reduced expression of the potential target genes in a way analogous to Chd7 knockdown, indicating that Chd7 functions as an important partner of Sox2 in neural stem cells. In support of this notion, Chd7 often colocalizes with Sox2 in ESC genomes (Schnetz et al., 2010). Furthermore, using a proteomics approach, Sox21 was found to be associated with Sox2 during ESC differentiation (Mallanna et al., 2010).

Analyses of the genomic regions bound by Sox2 and other factors (Boyer et al., 2005; Chen et al., 2008) not only confirmed that Sox2 and Pou5f1 are functional partners in ESC regulation, but also indicated that the same genomic sites are frequently co-bound by other transcription factors, such as Nanog. These observations suggest that the partnering between a Sox protein and its co-DNA-binding protein serves as a core of higher-order transcription factor complex.

It must be noted that, although ESCs and other cells cultured in vitro offer an abundant source of material for such high-throughput studies, these in vitro cultured cells may differ from real embryonic cells. For example, oestrogen receptor-related b (Esrrb) has been identified as a Sox2 partner in ESCs (Hutchins et al., 2013) and was shown to be essential for ESC maintenance in serum-free culture media (Festuccia et al., 2012; Martello et al., 2012), but in normal embryogenesis Esrrb is not expressed in the embryo proper but is required only for placental development (Mitsunaga et al., 2004; Pettersson et al., 1996). In the future, the analysis of cells isolated from embryonic tissue will be needed to validate the in vivo relevance of these interactions.

The recently identified function of transcription factors as pioneer factors (Liber et al., 2010; Xu et al., 2009a) may also apply to Sox proteins. The binding of a pioneer factor to a regulatory region does not elicit gene activation, but places the gene in a state poised for transcription. The subsequent recruitment of additional transcriptional factors to the site, or replacement of the pioneer factor with an acting factor, then activates the genes that are characteristic of differentiated cells (Bergsland et al., 2011). Sox2-bound genomic sites in ESCs are often associated with such genes, which are not expressed in ESCs but are activated in, for example, neural progenitor cells. Moreover, many SoxB1-bound genomic sites present in neural progenitor cells are also involved in the regulation of neuron-specific genes that will be activated later when SoxB1 is replaced with SoxC (Bergsland et al., 2011), suggesting that Sox proteins exhibit general pioneer factor functions. If these SoxB1 factors do indeed function as pioneers, their knockdown will affect occupancy by the second factor, thereby inhibiting developmental progression. However, many interesting issues remain unresolved, such as whether the pioneer factor function of Sox proteins still need partners.

**Conclusions and perspective**

As exemplified by the cases documented in this Primer, Sox proteins do not exert their activating or repressive regulatory functions when they bind to DNA on their own, but instead bind specific DNA sequences in combination with partner proteins to exert their effects. The combinatorial code of target specificity produced by Sox proteins and their partner proteins (the ‘Sox-partner code’) (Kamachi et al., 2000; Kondoh and Kamachi, 2010) ensures the stringent selection of target genes in the genome. In
addition, changes in one component of the Sox-partner complex facilitate the stepwise progression of developmental processes, whereas the formation of co-activating regulatory loops between Sox and its partner proteins stabilizes the state of cell differentiation (Kondoh and Kamachi, 2010; Kondoh et al., 2004). These basic principles will broadly apply to other transcription factor families that function as transcription factor complexes.

The recent advancement of genome-wide, high-throughput approaches holds promise to gaining an overall view of the gene regulatory networks that operate in the spatiotemporal dynamics of embryogenesis. Availability of more microscale analysis will be of great help in the analysis of embryonic tissues. To inquire into and to experimentally validate the regulatory networks in developing embryos, the refinement of gene knockout/knock-in technologies, such as inducible systems and CRISPR/Cas technology-based tailor-made gene manipulations (e.g. Cong et al., 2013) will make invaluable contributions. Such approaches will allow highly time-resolved and tissue-restricted analysis of the impact of changes in the activity of Sox or other transcription factors. Such molecular manipulations could include mutations in the Sox-partner molecular interface. In this respect, the information that lags behind the above-mentioned advances is the structural characterization of Sox-partner complexes. The structural data available to date are only those of DNA-bound Sox HMG domains (e.g. Reményi et al., 2003), rather than real complexes of full-length Sox and partner factors. Solving the three-dimensional structures of representative Sox-partner complexes on target DNA will contribute greatly to the elucidation of Sox partner-dependent regulations at the molecular level.

Acknowledgements
The authors appreciate discussions with the members of Kondoh laboratory and Sox research community, which made this article possible, and thank anonymous reviewers for helpful suggestions. Because of the space constraints and the focus of this article, many original studies were not referred to or cited, with apology.

Funding
Studies by the authors described in this article were funded in part by grants cited, with apology.

Competing interests statement
The authors declare no competing financial interests.

References


comparison of the properties of Sox-3 with Sry and two related genes, Sox-1 and Sox-2. Development 122, 509-520.


