Developmental Cell, Vol. 4, 509–519, April, 2003, Copyright ©2003 by Cell Press

Functional Analysis of Chicken *Sox2* Enhancers Highlights an Array of Diverse Regulatory Elements that Are Conserved in Mammals

Masanori Uchikawa,[1] Yoshiko Ishida,[1,2] Tatsuya Takemoto,[1] Yusuke Kamachi,[1] and Hisato Kondoh[1,*]
[1]Graduate School of Frontier Biosciences
Osaka University
1-3 Yamadaoka
Suita, Osaka 565-0871
Japan

Summary

***Sox2* expression marks neural and sensory primordia at various stages of development. A 50 kb genomic region of chicken *Sox2* was isolated and scanned for enhancer activity utilizing embryo electroporation, resulting in identification of a battery of enhancers. Although *Sox2* expression in the early embryonic CNS appears uniform, it is actually pieced together by five separate enhancers with distinct spatio-temporal specificities, including the one activated by the neural induction signals emanating from Hensen's node. Enhancers for *Sox2* expression in the lens and nasal/otic placodes and in the neural crest were also determined. These functionally identified *Sox2* enhancers exactly correspond to the extragenic sequence blocks conspicuously conserved between chicken and mammals, which are not discernible by sequence comparison among mammals.**

Introduction

SOX proteins are a group of transcriptional regulators having analogous HMG domains for their DNA binding, and are highly involved in regulation of cell differentiation (Kamachi et al., 2000; Pevny and Lovell-Badge, 1997; Wegner, 1999). There are roughly 20 *Sox* genes in each vertebrate species (Schepers et al., 2002), which are classified into nine groups, according to the organization of the encoding proteins and the HMG domain sequences. Cumulative evidence indicates that SOX proteins form a complex with their partner transcription factors unique to each SOX group, thereby selecting group-specific and partner-dependent regulatory target genes (Kamachi et al., 2000, 2001; Wilson and Koopman, 2002).

Among the *Sox* genes, those belonging to group B1 (*Sox1*, *Sox2*, and *Sox3*) are expressed widely in the nervous system, and are implicated in neural development (Collignon et al., 1996; Pevny et al., 1998; Rex et al., 1997; Uchikawa et al., 1999; Uwanogho et al., 1995; Wood and Episkopou, 1999). In particular, *Sox2* expression consistently marks neural primordial cells at various stages of development (Cai et al., 2002; Kishi et al., 2000; Mizuseki et al., 1998; Rex et al., 1997; Streit et al., 1997; Uchikawa et al., 1999; Uwanogho et al., 1995). Furthermore, a recent report investigating the neural primordia of embryos indicates that *Sox2* expression is highly correlated with the multipotent neural stem cell state (Cai et al., 2002). Because *Sox2* is expressed more or less uniformly in the early neural tube, it is regarded as a "pan-neural" marker in early embryonic stages (Darnell et al., 1999; Streit et al., 1997). In addition, *Sox2* is commonly expressed in sensory placodes (Uchikawa et al., 1999).

Another important aspect of *Sox2* regulation is that its expression in the CNS is first activated upon neural induction elicited by signals from the organizer (Fernandez-Garre et al., 2002; Streit et al., 1997). Therefore, initiation of *Sox2* expression must be an essential part of the mechanism of neural induction. These aspects of *Sox2* expression in the early embryo are common among vertebrates (Kishi et al., 2000; Rex et al., 1997; Streit et al., 1997; Wood and Episkopou, 1999).

To elucidate the molecular mechanisms underlying *Sox2* regulation in the early neural tissues, we analyzed a wide genomic region encompassing the *Sox2* gene. Here we report the following findings: (1) *Sox2* in the neural tissues is regulated by multiple extragenic enhancers having distinct spatio-temporal specificities, in spite of its apparently uniform "pan-neural" expression; (2) one of the enhancers responds to the neural induction signals emanated from Hensen's node, opening a straightforward access to the mechanism of neural induction; and (3) consistent with the conserved *Sox2* expression pattern, the functionally identified *Sox2* enhancers exactly correspond to extragenic sequence blocks conspicuously conserved between the chicken and mouse/human, but these blocks are obscured when mouse and human sequences are compared because of the existence of wider spans of sequence similarity. Thus, the enhancers associated with a genetic locus are accurately predictable as sequence blocks conserved between chicken and mammals.

Results and Discussion

Detailed *Sox2* Expression Patterns in Early Chicken Embryos

Sox2 expression consistently marks neural primordial cells at various stages of development. *Sox2* expression in the CNS is first activated by neural induction elicited by signals from the organizer. In chicken embryos, *Sox2* expression is initiated at stage 4 (st. 4) in the presumptive cephalic neural plate, the posteriorly notched horseshoe-shaped area surrounding Hensen's node (Figure 1A, compare st. 3 and st. 4). As Hensen's node moves posteriorly, a newer *Sox2*-expressing area is continuously generated and added to the region of previously activated *Sox2* (Figure 1A, st. 5–8). At neurulation stages when the sheet of *Sox2*-expressing cells converges (Figure 1A, st. 8) and forms a neural tube, posterior addition of new *Sox2*-expressing cells still continues (Figure 1A, st. 8–14).

*Correspondence: kondohh@fbs.osaka-u.ac.jp
[2]Present address: Hitachi, Ltd., Life Science Group, 1-3-1 Minamidai, Kawagoe, Saitama, 350-1165 Japan.

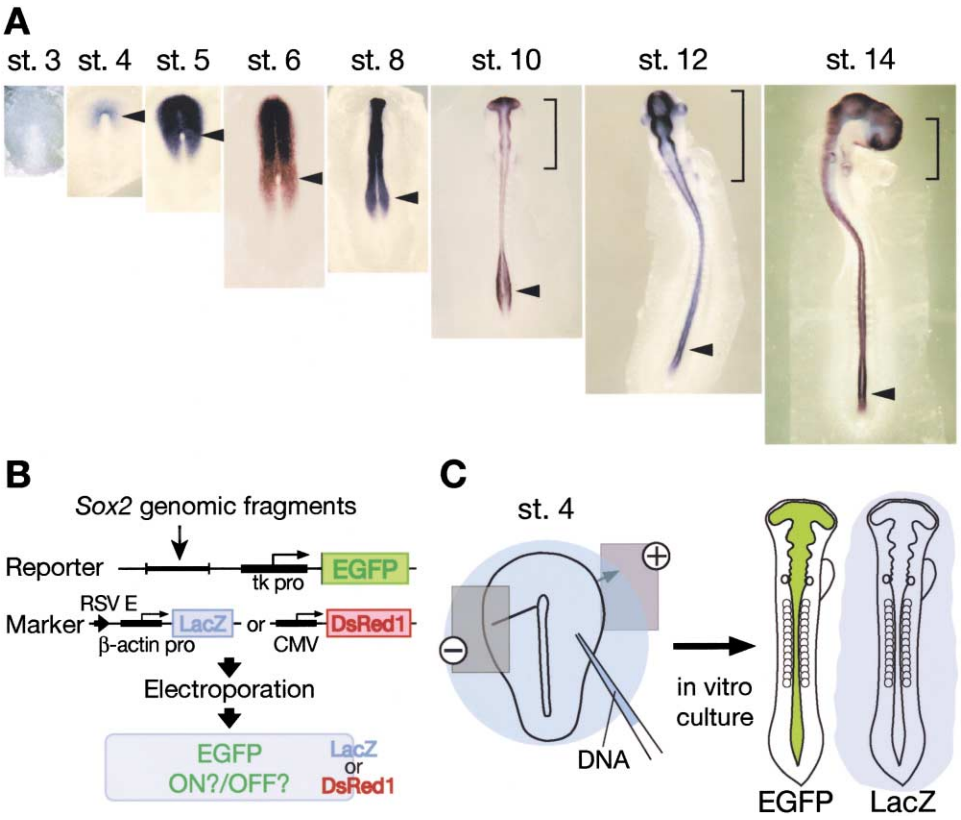


Figure 1. Development and Electroporation of Chicken Embryo

(A) Expression of *Sox2* in the early chicken embryo. *Sox2* expression is initiated in the presumptive cephalic neural plate at st. 4 in the area surrounding Hensen's node (arrowhead). The *Sox2*-expressing area extends posteriorly with posterior regression of the node. After st. 10, *Sox2* expression is induced in the cephalic placodes, and their derivatives are indicated by a bracket.

(B) To identify the regulatory elements responsible for expression of *Sox2* in the early CNS and placodes, we cloned an ~50 kb genomic region encompassing the *Sox2* gene of the chicken, prepared a number of small fragments scanning the whole region, and placed the fragments in a reporter expressing EGFP.

(C) The reporter constructs were electroporated into chicken embryos under New culture from st. 4 to st. 7, and expression of EGFP in tissues of the embryo was examined after 6–24 hr. *LacZ* or *DsRed1* genes driven by nonspecific enhancer/promoters were coelectroporated to monitor electroporation.

Around st. 10, *Sox2* expression is induced in the second set of ectodermal cell populations, the sensory placodes (Figure 1A, st. 10–14). *Sox2* is also activated in the ventral to lateral parts of the head ectoderm. The expression in the lateral head ectoderm is extinguished by stage 11, except for the region contacted by the optic vesicle at st. 10, which is marked by induction of strong *Sox2* expression and develops into the lens placode (Kamachi et al., 1998). The lens placode then invaginates and forms a lens vesicle, the primitive lens (st. 14–16). Strong *Sox2* expression continues in later lens development in the chicken (Kamachi et al., 1998; Uchikawa et al., 1999).

Identification of *Sox2* Enhancers

Definition of regulatory sequences of the *Sox2* gene accounting for its expression in the early CNS and placodes is essential to elucidate the molecular mechanisms underlying the development of these tissues. A roughly 50 kb region of the chicken genome covering the *Sox2* locus (16.7 kb upstream and 32.5 kb downstream; Figure 2) was isolated, the whole nucleotide sequence was determined, and various portions were analyzed for enhancer activity.

Chicken embryos at st. 4 under New culture were electroporated in the epiblast and ectoderm with EGFP reporter constructs carrying various genomic fragments of the *Sox2* locus (Figure 1B), and expression of the EGFP gene in the developing embryos (Figure 1C) was examined. Successful electroporation of wide tissue areas was confirmed by coelectroporation of LacZ or DsRed1 vectors carrying a nonspecific enhancer. At least six independent electroporations were done for each genomic fragment, which gave essentially the same results.

Initial screening using DNA fragments ~5 kb long indicated association of a certain enhancer activity with most of the fragments examined, whereas the specificity of each enhancer was distinct. We narrowed down the DNA fragments bearing the enhancer activity step by step to a few hundred base pairs as shown in Figure 2. We also included in the analysis the DNA fragments corresponding to the conserved sequence blocks identified by comparison of chicken and mammalian sequences as described below. Through this analysis, we

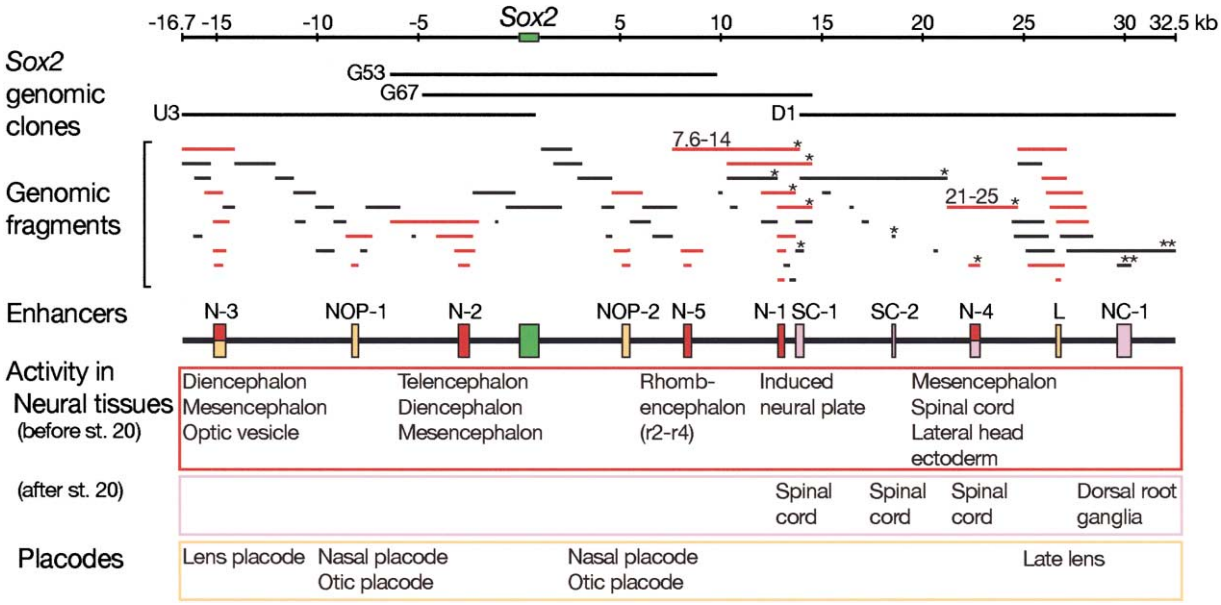


Figure 2. Identification of Enhancers Responsible for *Sox2* Expression in Various Domains of the CNS and Sensory Placodes

The genomic fragments that cover the ~50 kb *Sox2* locus of the chicken were tested for enhancer activity by electroporation of embryos. Position +1 indicates the translational start site. The *Sox2* coding sequence shown in green has no intron. Fragments shown in red exhibited an enhancer activity by electroporation of st. 4 embryos and observation at st. 5–20. An active fragment and its subfragments generally showed the same enhancer activity. Fragments electroporated in the st. 11 neural tube and showing enhancer activity are indicated by * (activity in the spinal cord) or ** (activity in the dorsal root ganglia). Fragments of 7.6–14 and 21–25 are used in Figures 4E and 4F. In the lower map, neural enhancers detectable in stages before st. 20 are indicated by red boxes, those detected in later stages are indicated by purple boxes, and the placodal enhancers are indicated by ocher boxes. Their domain specificities in the CNS and placodes are indicated.

identified eight distinct enhancers active in different domains of the early CNS and sensory placodes, as summarized in Figure 2 (red and ocher boxes). These enhancers were active in both dorsal and ventral compartments of respective domains of the CNS.

Enhancer N-1 Responding to Neural Induction Signals

Activity of the enhancer N-1 accounts for the Hensen's node-induced *Sox2* expression. A few hours after electroporation (st. 5), EGFP fluorescence reflecting the enhancer activity appeared in the area surrounding Hensen's node. Continuous observations of the same developing embryos indicated that the tissue area supporting the enhancer activity moved caudally, always surrounding the node undergoing posterior regression (Figure 3A). Use of a destabilized form of EGFP (d2EGFP) did not significantly alter the tissue area of the EGFP fluorescence, indicating that the EGFP reporter expression follows the temporal change of the enhancer activity.

We explanted Hensen's node of quail to nonneural regions of st. 4 chicken embryos, in the distal parts of area pellucida (embryogenic region) or in area opaca (extraembryonic region), which had been electroporated with the N-1/EGFP construct beforehand. The Hensen's node explant induced the activity of the enhancer at these ectopic sites, in parallel with ectopic induction of the endogenous *Sox2* as previously reported (Streit et al., 1997; Figure 3B). These observations indicate that the enhancer N-1 activates *Sox2* expression responding to neural induction signals from Hensen's node.

In later phases, the activity of enhancer N-1 is always confined to the region surrounding the node, and is not retained in more rostral regions where the enhancer had the activity in earlier periods; this argues that enhancer N-1 is primarily involved in the activation of *Sox2* expression in the process to generate the neural plate.

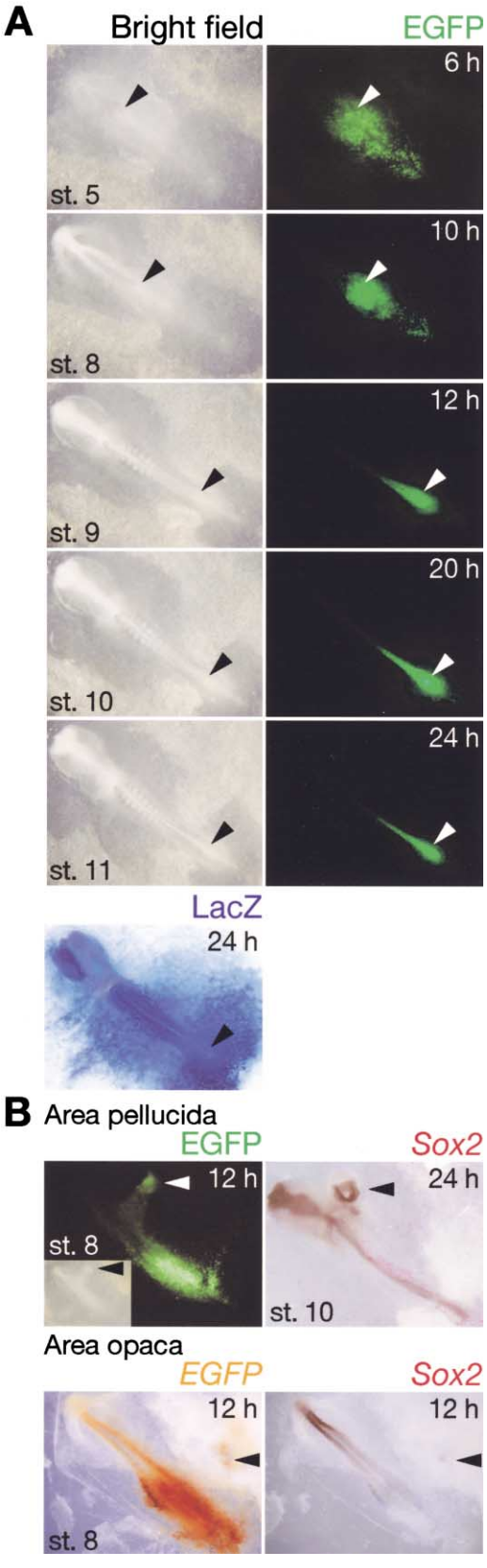
It is indicated that in early neurogenesis (e.g., st. 6), the epiblast cells in the proximity of the node are newly induced to take part in the neural plate, whereas later in neurogenesis, neural precursors surrounding the node proliferate and migrate caudally along with the node (Brown and Storey, 2000; Darnell et al., 1999; Mathis et al., 2001; Storey et al., 1992). In both phases, enhancer N-1 is active in the cell population that surrounds the node and is in the nascent phase of neurogenesis.

Enhancers for Specific Domains of the CNS

In other regions of the neural plate, enhancers N-2 to N-5 regulate *Sox2* expression.

Enhancer N-2 is activated around st. 5 analogous to enhancer N-1, and covers the most rostral part of the CNS. The activity of N-2 is initially broad, covering the region rostral to Hensen's node before st. 7 (Figure 4A, 6 hr), and at st. 9–11 the activity covers prosencephalon, mesencephalon, and rostral rhombencephalon (Figure 4A, st. 11). The activity is gradually restricted more rostrally, and eventually to the dorsal telencephalon (Figure 4A and data not shown).

Enhancer N-3 has the activity detectable after st. 8 in the neural folds roughly corresponding to the level of the mesencephalon. Later the activity of enhancer N-3 covers the retina, diencephalon, and the rostral mesencephalon (Figure 4B).



Enhancer N-4 has an activity detectable after st. 10 in the mesencephalon, rhombencephalon, and the spinal cord, in which the caudal margin of the expression lies around the level of somite formation. The enhancer activity in the rhombomeres r2–r4 is significantly lower than other regions (Figure 4C). Enhancer N-4 is also active in the lateral and ventral head ectoderm, accounting for the ectodermal *Sox2* expression (Figure 4C, arrow).

Enhancer N-5 is active after st. 12 exclusively in the rhombencephalon at the levels r2–r4 in a way complementing the activity of enhancer N-4 (Figure 4D).

Thus, enhancers N-1 to N-5 individually regulate *Sox2* in distinct domains of the CNS, and combination of the expression domains covers the entire region of the CNS. This indicates that "pan-neural" expression of *Sox2* is actually brought about by a patchwork of region-specific enhancer activities.

The nucleotide sequences of these early neural enhancers are shown in Figure 5. Inspection of these sequences indicates that potential binding sites for SOX/LEF/TCF (HMG domain) transcription factors (Schilham and Clevers, 1998) are distributed among the enhancers. The sequence ATTA, characteristic of homeodomain protein binding sites, is also common. In particular, potential sites for POU family proteins, a subclass of the homeodomain proteins, are carried by enhancers N-2 and N-3. The occurrence of SOX/LEF/TCF factor binding sites raises the possibility that these enhancers are activated by SOX2 itself, which is first turned on by the action of enhancer N-1, and suggests an autoregulatory loop in maintaining enhancer activity. Alternatively, they could be under regulation of the Wnt signaling system, having LEF-1/TCF transcription factors as their downstream target (Clevers and van de Wetering, 1997).

It has been indicated that HMG domain proteins and

Figure 3. Activity of Enhancer N-1 Responding to Neural Induction Signals

(A) The same embryo electroporated with EGFP reporter vector carrying enhancer N-1 at st. 4 and observed at different stages of development under New culture. Arrowheads indicate the position of Hensen's node. EGFP fluorescence representing enhancer activity appears after a few hr in the region surrounding Hensen's node, and the region with the activity changes location as the node moves caudally. The LacZ panel shows expression of coelectroporated nonspecific *LacZ* gene, confirming gene transfer in a wide area of the embryo.

(B) Effect of Hensen's node explant on the activity of enhancer N-1. A piece of node from the quail embryo (arrowhead) was grafted in st. 4 embryos, which had been electroporated with N-1-bearing EGFP reporter plasmid.

Upper: graft at the edge of area pellucida. Left: an embryo 12 hr after the graft. The enhancer is activated (arrowhead) in the chicken tissues surrounding the graft, in addition to those surrounding the authentic node. The inset shows a bright field view of the same embryo. Right: the same embryo fixed after 24 hr and processed for *Sox2* in situ hybridization, showing induction of endogenous *Sox2* expression in response to the node graft.

Lower: graft in area opaca. Left: an embryo 12 hr after the graft, processed for *EGFP* in situ hybridization and color development with INT/BCIP. Right: the same specimen after removal of the pigment deposit of INT/BCIP processed for *Sox2* in situ hybridization and color development with NBT/BCIP.

st., developmental stages; h, hours after electroporation.

homeodomain proteins often interact in regulation of transcription of the target genes (Dailey and Basilico, 2001; Kamachi et al., 2000). The regulation of *nestin* gene, which is also expressed in the neural primordial cells, requires binding of a POU protein (e.g., Brn1) to a site in its neural enhancer (Josephson et al., 1998). We have recently shown that the *nestin* enhancer is also dependent on the binding of a SOX protein (e.g., SOX2; S. Tanaka, Y.K., N. Jin, H. Hamada, and H.K., unpublished results). It is tempting to speculate that at least in the rostral domain of the early CNS where *Sox2* expression is activated by enhancers N-2 and N-3, the neural primordial cells are maintained by a combinatorial action of POU and SOX transcription factors.

Enhancers N-1, N-4, and N-5 have the binding sites for Zfhx (zinc finger-homeodomain) family proteins δEF1 (Sekido et al., 1994, 1997)/SIP1 (Verschueren et al., 1999) in their nucleotide sequences, although its significance remains to be examined.

Enhancers for Later Neural Development

The enhancers active in later developmental stages of the CNS were also investigated by electroporating st. 11 embryos in ovo in the ventricular zone of the spinal cord, and assessing enhancer activities at st. 23 (48 hr later). A few regions located 3′ of the *Sox2* gene have analogous enhancer activity in the ventricular zone of the spinal cord (Figure 2, fragments marked by *; Figures 4E and 4F). Besides enhancer N-4 active from the early spinal cord, a few other regions, such as conserved blocks 18 and 22 (see below) designated as enhancers SC-1 and SC-2, respectively, were active in the spinal cord.

The same analysis also indicated an enhancer activity in the dorsal root ganglia, reflecting the *Sox2* expression in the neural population of neural crest cells (Figure 4G; Uchikawa et al., 1999; Uwanogho et al., 1995). The activity was associated with the fragment located the most 3′ of the region of the analysis (Figure 2, fragments marked by **), which was narrowed down to the conserved sequence block 25 designated as enhancer NC-1 (neural crest-1; see below).

Enhancers for Lens Development

In previous work, we identified three distinct stages of *Sox2* expression in the early lens development starting from the lateral head ectoderm (Kamachi et al., 1998): (1) low *Sox2* expression in the ventral to lateral head ectoderm from before the stages of contact by the optic vesicle; (2) high *Sox2* expression induced by the contact of the optic vesicle (st. 11) and leading to lens placode formation (st. 12) and initiation of lens development in concert with more widely expressed *Pax6* (Kamachi et al., 2001); and (3) high *Sox2* expression for lens development after the vesicle stage (st. 14) in the chicken, which is replaced by *Sox1* expression in the mouse, rendering *Sox1* knockout mouse defective in lens development (Nishiguchi et al., 1998).

Enhancers exactly corresponding to all these three stages of lens development were identified. Enhancer N-4 is initially active in the lateral to ventral head ectoderm besides mesencephalon and spinal cord (Figure 4C, arrow). The enhancer N-3 is then activated by the lens induction signal of the optic vesicle (Figure 4B, arrow). In embryos where one of the optic vesicles failed to make contact with the lateral ectoderm under the culture condition, this enhancer was not activated in that side (data not shown). Enhancer N-3 remains active in the lens-forming cells up to early lens vesicle stage (st. 14). Then late lens enhancer, L, becomes active and persists to later lens development, which was identified by electroporation of st. 10 embryo head ectoderm (Figure 4J, arrow).

It is interesting to note that the enhancer involved in lens induction (N-3) is distinct from that in neural induction (N-1), implying that different extracellular signals induce these tissues.

Nasal and Otic Placode Enhancers

The analysis employing electroporation of st. 4 embryos also identified placodal enhancers. Two enhancers, NOP-1 and NOP-2, are found which are individually active in early nasal and otic placodes (Figures 4H and 4I). This observation implies that development of these placodes is under analogous regulation at least in the initial phases.

Matching of the Enhancers to the Array of Sequence Blocks Conserved between Chicken and Mammals

As summarized in Figure 2, we identified an array of enhancers with distinct spatio-temporal coverages that account for expression of *Sox2* in the embryonic CNS and sensory placodes. Because *Sox2* expression pattern in these tissues is well conserved between the chicken and the mouse (Kamachi et al., 1998; Rex et al., 1997; Uwanogho et al., 1995; Wood and Episkopou, 1999), we compared the nucleotide sequences of the *Sox2*-flanking regions between chicken and mouse, and between chicken and human, using a stringent criterion to assess similarity (>60% identity in a stretch longer than 100 bp). The analysis revealed 25 blocks of sequence highly conserved between chicken and mammals and distributed in the region of analysis. Most remarkably, all ten identified neural and placodal enhancers (except for enhancer L; Figure 2) match these blocks (Figure 6A; Table 1). The sequence alignment of the neural enhancers N-1 to N-5 derived from the three animal species (Figure 5) confirms the high degree of conservation, including the putative transcription factor binding sites. Previously, Zappone et al. (2000) described an enhancer located 4 kb 5′ of mouse *Sox2* gene, and active in the dorsal telencephalon of midgestation mouse embryos. This enhancer corresponds to the block 8/enhancer N-2 (Figure 6A; Table 1), confirming the conserved enhancer activity.

DNA fragments representing the conserved sequence blocks were also analyzed by electroporation of st. 4 embryos, but indicated no enhancer other than those shown in Figure 2 (Table 1). Electroporation of st. 11 neural tubes, however, demonstrated the presence of additional enhancers active in the spinal cord, blocks 18 (SC-1) and 22 (SC-2). In addition, block 25 (enhancer NC-1) was found to bear enhancer activity in the dorsal root ganglia. These enhancer activities had been indicated for larger DNA fragments (Figure 2, those marked by * or **).

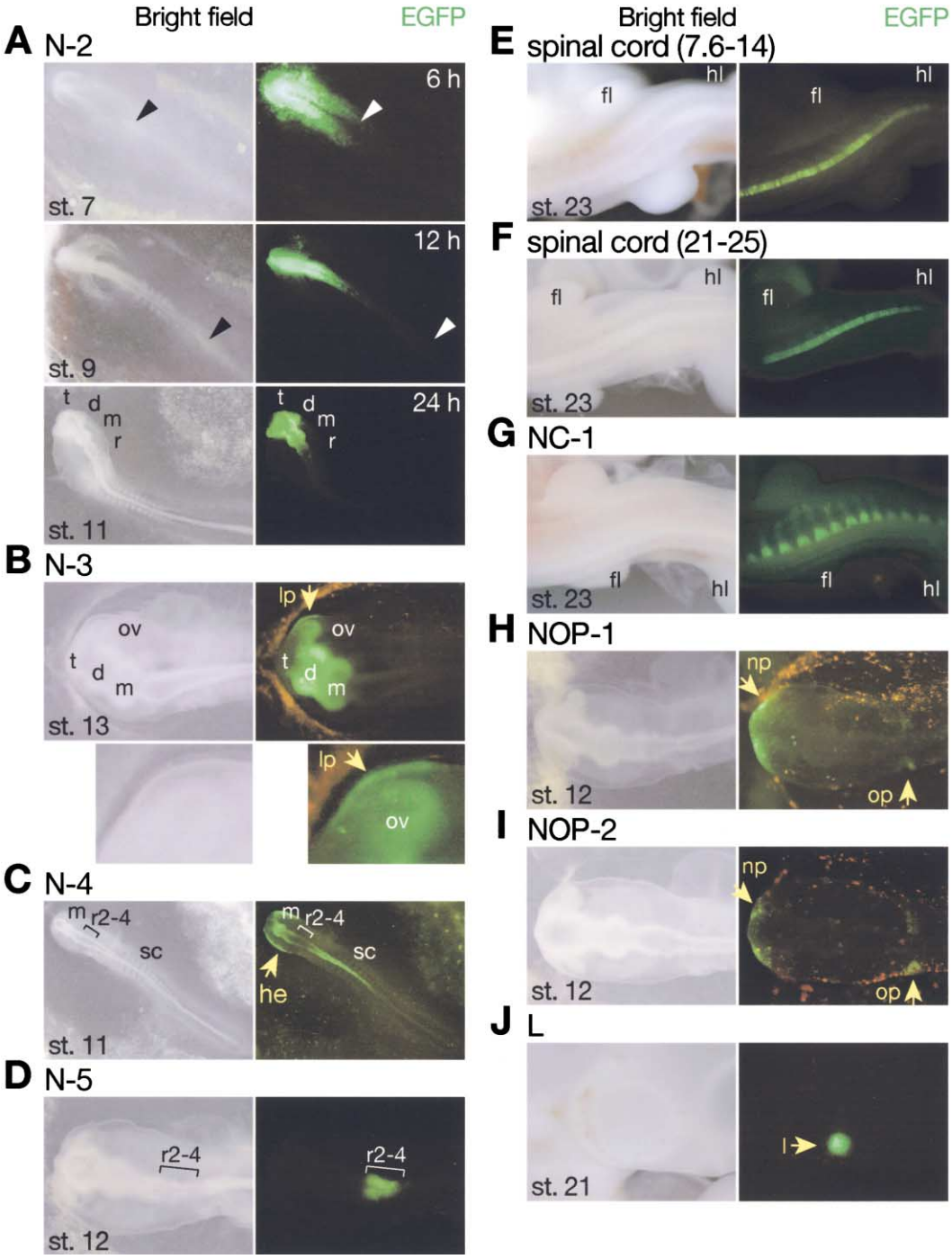


Figure 4. Activity of Individual *Sox2* Enhancers in Electroporated Chicken Embryo

(A–D) Enhancers N-2 to N-5 individually regulating *Sox2* in distinct domains of the CNS.

(A) Enhancer N-2 at three embryonic stages showing the activity in the telencephalon (t), diencephalon (d), mesencephalon (m), and the anterior rhombencephalon (r).

(B) Enhancer N-3, showing activity in the diencephalon, rostral mesencephalon, and retina (optic vesicle, ov). N-3 also shows activity in the lens placode (lp; arrow), which is dependent on apposition by the optic vesicle. The lower panels are enlargements showing the lens placode-optic vesicle region.

(C) Enhancer N-4 showing activity in the mesencephalon (m), rhombencephalon (activity lower in rhombomeres r2–r4), and spinal cord (sc). This enhancer also shows activity in the lateral head ectoderm (he; arrow). This activity in the lateral head ectoderm is attenuated by st. 11.

(D) Enhancer N-5, active in the rhombencephalon from r2 to r4.

(E–G) Enhancers detected by electroporation of the st. 11 neural tube.

(E) Spinal cord enhancer activity of the DNA fragment 7.6–14 (Figure 2) containing the conserved sequence blocks 14–18.

(F) Spinal cord enhancer activity of fragment 21–25 (Figure 2) containing enhancer N-4 (conserved sequence block 24).

(G) Enhancer activity in the dorsal root ganglia of enhancer NC-1 (block 25).

Other blocks that have not been assigned to an enhancer may be active in other tissues, such as esophagus (Ishii et al., 1998), or in later development. Indeed, block 11 corresponds to the recently described enhancer active in mouse ES cells (Tomioka et al., 2002). On the other hand, we have not identified enhancers for expression in epibranchial placodes. These and some other enhancers may be present outside the region of this current analysis.

Thus, the data demonstrate that the *Sox2* enhancers are highly conserved between chicken and mammals, and the corresponding enhancers are arranged colinearly in the genomes (Figure 6A). Comparison of the chicken sequence with fugu (puffer fish) *Sox2* locus (scaffold 2806) indicated that blocks 4 and 10 are conserved in the fish, but the other blocks are not (data not shown). Sequence comparison between the mouse and human genomes, however, showed that the enhancer regions are buried within longer stretches of sequence similarity (Figures 6A and 6D). Presumably the phylogenic distance between birds and mammals is optimal for identifying the conserved regulatory sequences.

The recent advance of genome sequencing has indicated that conserved sequences of noncoding regions are candidates for genetic regulatory regions, but strategies to sort out functionally significant conserved regions have been sought (see Hardison, 2000 for review). Our analysis shown in Figure 6 demonstrates that, at least in the extragenic sequences of the *Sox2* locus, sequence conservations between chicken and mammals are a reliable indication of functional enhancers. This observation may be generalized; that is, if an array of sequence blocks is conserved between chicken and mammals and is associated with a genetic locus, it likely represents a set of important regulatory sequences. Indeed, there is a precedence that a missing neural enhancer of *SCL* gene was identified as a sequence block conserved between chicken and mammals and located distantly from the gene (Gottgens et al., 2000).

Comparison of chicken and mouse sequences is sensitive enough to identify subtle differences of *Sox2* regulation between the species. The late lens enhancer (L) of the chicken has no counterpart in the mouse genome, accounting for the absence of *Sox2* expression in the mouse lens after the lens vesicle stage (Kamachi et al., 1998; Nishiguchi et al., 1998). It is interesting to note that some of the blocks conserved between chicken and human (blocks 3, 21, and 25) are not found in the mouse sequence (Figure 6A, boxes labeled "h"; Table 1).

Conclusions

We identified distinct enhancers for *Sox2* expression in the embryonic CNS and placodes, including those responding to signals for neural and lens induction. Analysis of these enhancers will allow the unveiling of molecular mechanisms that initiate development of these tissues. Expression of *Sox2* in the neural tissues depends on a patchwork of activities of separate enhancers that have distinct spatial and temporal specificities. Perhaps the expression of *Sox2* is an absolute requirement for the neuroepithelial cells, but the extracellular and intracellular conditions of these cells presumably vary from one tissue level to another and from one developmental stage to another, and these variations necessitate the contribution of numerous enhancers of different specificities to establish a uniform and continuous *Sox2* expression. These enhancers correspond to an array of phylogenically conserved extragenic sequences, which are clearly definable by comparing chicken and mammalian sequences. Thus, comparison of chicken and mammalian genome sequences provides definitive information to identify and characterize the regulatory sequences.

Experimental Procedures

Genomic Clones of the Chicken *Sox2* Locus and Sequence Analysis

Sox2 genomic clones G53 and G67 were isolated from a λFIXII genomic library of the chicken by screening with *Sox2* cDNA as probe (Kamachi et al., 1995), and additional clones U3 and D1 by walking (Figure 2). Comparison of the sequences was done using Pustell matrix analysis (Pustell and Kafatos, 1982) in MacVector software. The hash size, window size, and minimal percent scores were chosen as six, 30, and 60, respectively. When a valid match was found, we examined to see whether the conserved sequence with no lower than 60% identity continues over 100 bp in length. The sequence alignment was also confirmed by the VISTA (Mayor et al., 2000) and Gene Works programs.

Plasmid Construction

The plasmid ptkEGFP was constructed by insertion of the Herpes simplex virus thymidine kinase promoter in the polylinker HindIII site of pCAT3-basic vector (Promega), and by replacing the CAT gene with the EGFP gene (Clontech). To identify enhancers associated with the *Sox2* locus, various DNA fragments of *Sox2* and flanking sequences were inserted in the polylinker SmaI site of ptkEGFP. In some cases, d2EGFP (Clontech), a destabilized form EGFP, was employed for the coding region, but the expression pattern in the embryo was indistinguishable between EGFP and d2EGFP, whereas fluorescence was brighter with the former, causing us to compile the full data set with EGFP-encoding vectors.

Electroporation

We adopted electroporation to transfer plasmid DNA into the epiblast/ectoderm layer of early chicken embryos. Stage 4 chicken embryos (Hamburger and Hamilton, 1951) were cultured by a modified New technique (Kamachi et al., 1998). An egg containing a staged embryo was delivered from the shell in a 10 cm dish. Thick albumen was removed from the area surrounding the blastoderm, a ring of filter paper was placed on the vitelline membrane encircling the blastoderm, and the membrane was cut at the circumference of the filter ring. The embryo anchored to the filter ring through the vitelline membrane was transferred into warmed Hank's solution, and gently washed to remove the adhering yolk. The embryo was placed upside down (vitelline membrane side down) onto the electrode (cathode), which was made of platinum foil 2 × 2 mm square

(H–J) Enhancers are active only in the placodal lineages.
(H) Enhancer NOP-1, having activity in nasal (np) and otic (op) placodes.
(I) Enhancer NOP-2, having activity in nasal (np) and otic (op) placodes, similar to NOP-1.
(J) Enhancer L, having activity in the lens (l) after the lens vesicle stage.
The orange fluorescence in (B), (H), and (I) is derived from strong DsRed1 expression in extraembryonic tissues.
st., developmental stages; h, hours after electroporation.

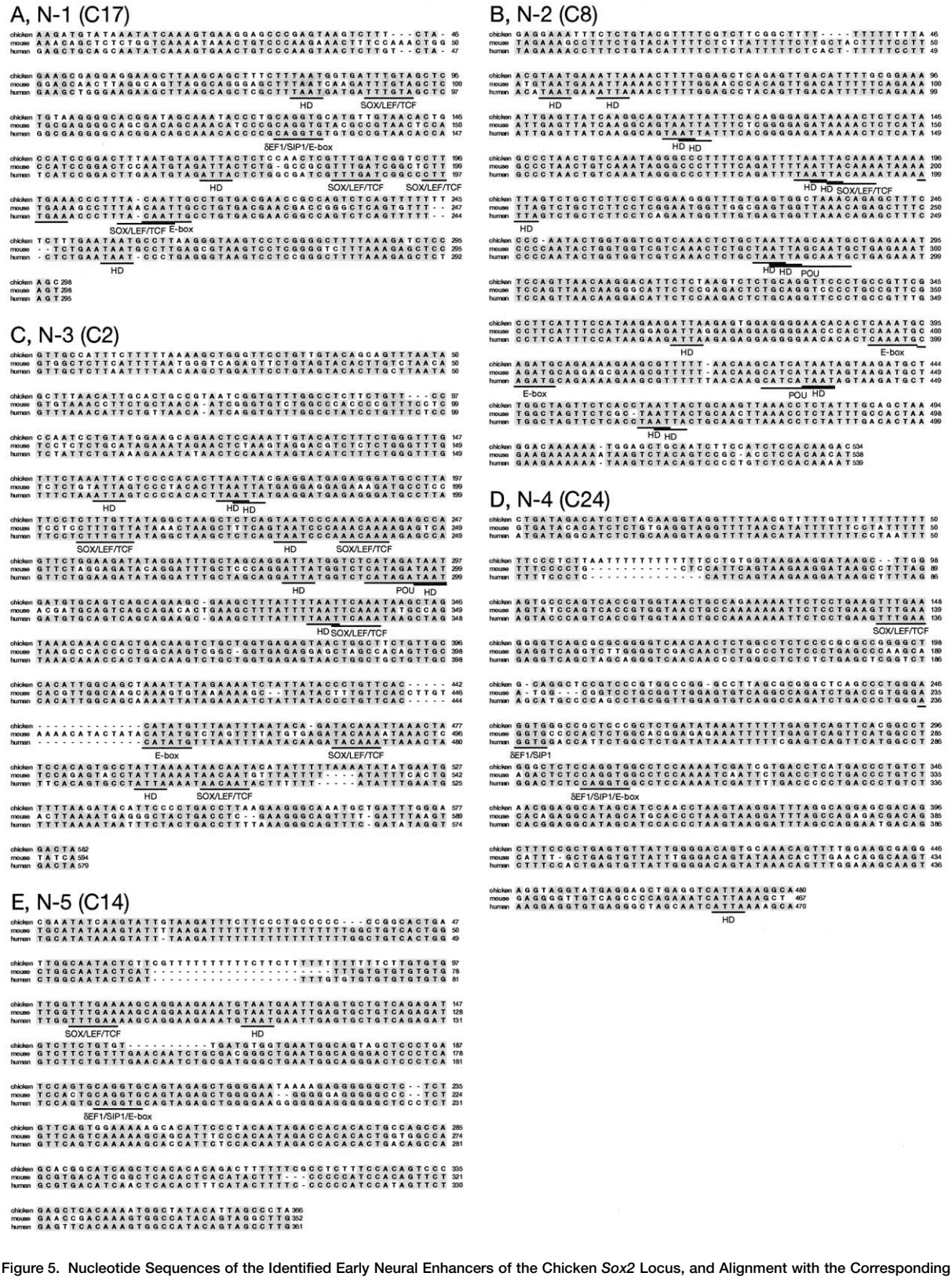


Figure 5. Nucleotide Sequences of the Identified Early Neural Enhancers of the Chicken *Sox2* Locus, and Alignment with the Corresponding Mouse and Human Sequences

Enhancers N-1 (A), N-2 (B), N-3 (C), N-4 (D), and N-5 (E). Sequences are shaded when the nucleotide residue is conserved in two of the species. Putative binding sequences of various transcription factors conserved among the animal species are underlined; putative binding sequences of SOX/LEF/TCF proteins ([A/T][A/T]CAA[A/T](G)), POU proteins ($CATN_{0\text{-}3}$[T/A]AAT), and δEF1/SIP1 proteins (CACCT(G)) (Sekido et al., 1994, 1997; Verschueren et al., 1999) overlapping with E2 box (CACCTG), homeodomain (HD) proteins (ATTA), and E box (CANNTG).

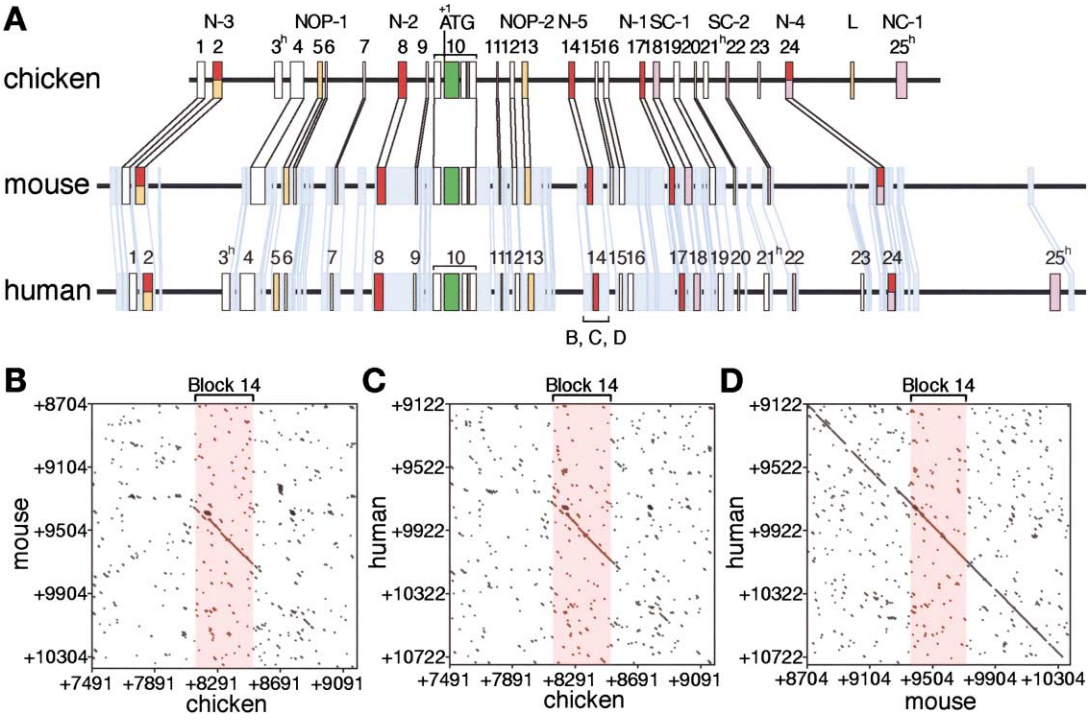


Figure 6. Conservation of the Identified Enhancers across Species in Amniotes

(A) Summary of the analysis of conserved sequence blocks of the *Sox2* locus. The blocks of sequences (>60% identity over >100 bp) conserved between chicken and mouse and between chicken and human are indicated as boxes on the maps of respective animal species. The *Sox2* coding sequence is shown in green. The *Sox2* enhancers shown in Figure 2 are indicated using the same color codes. These match the conserved sequence blocks except for enhancer L, which is unique to the chicken. The blocks labeled (h) are conserved between the chicken and human but not in the mouse. The regions of sequence conservation between mouse and human determined by the same criteria are indicated in blue.

(B–D) Dot matrix analysis comparing the nucleotide sequences encompassing the conserved sequence block 14 corresponding to enhancer N-5. A dot was produced with the window size of ten nucleotides and the minimal score of 60% identity.

(B) Comparison of chicken and mouse sequences.

(C) Comparison of chicken and human sequences.

(D) Comparison of mouse and human sequences. Note that the block 14 (enhancer N-5) sequence is embedded in a longer stretch of sequence similarity when the sequence comparison is made between mouse and human.

fixed to the bottom of the chamber. A DNA solution containing the reporter plasmid (~1 μl volume with a final concentration of 2 μg/μl) and the marker plasmid (pMiwZ or pDsRed1-N1 [Clontech], 1 μg/μl) was injected between the blastoderm and the vitelline membrane using a glass capillary. An anodal electrode was quickly placed on the hypoblast side of the embryo with the interelectrode distance of 4 mm. Electroporation was performed using a CUY21 electroporator (BEX) with five pulses of 10 volts for a duration of 50 msec and with intervals of 100 msec. The embryo was rinsed with warmed Hank's solution and incubated with the hypoblast side up on the agar culture medium made of liquid albumen, 0.15% glucose, 61.5 mM NaCl, and 0.3% agar noble (DIFCO). The embryo was overlaid with 5% yolk supernatant in Hank's solution and incubated at 38°C with 100% humidity overnight. To examine the effect of an ectopic node, a chicken embryo was electroporated as described above, and Hensen's node of stage 4 quail embryo was excised using a tungsten needle and grafted onto the host chick embryo at the position of area pellucida-area opaca boundary or at the position of area opaca with the same anteroposterior level as the host node.

Whole-Mount In Situ Hybridization

Embryos were fixed overnight in 4% paraformaldehyde (PFA) in phosphate-buffered saline (PBS) at 4°C, washed twice in PBT (PBS containing 0.1% Tween-20), and dehydrated with 25%, 50%, and 75% methanol in PBT, and then with 100% methanol and stored. The embryos were rehydrated through the methanol/PBT series in reverse and washed twice in PBT; they were then bleached with 6% hydrogen peroxide in PBT for 1 hr, followed by two PBT washes. Embryos were treated with 10 μg/ml proteinase K for appropriate periods depending on their stages, washed briefly in PBT, and re-fixed in 0.2% glutaraldehyde/4% PFA in PBT for 20 min. They were then washed twice in PBT for 10 min, and preincubated in hybridization buffer (50% formamide, 5× SSC [pH 5], 1% CHAPS, 50 μg/ml heparin, 50 μg/ml yeast tRNA) for 1 hr at 70°C. Hybridization with 0.5–1 μg/ml digoxigenin-labeled *Sox2* RNA probe (Kamachi et al., 1998) was carried out overnight at 70°C. The embryos were washed twice in solution 1 (50% formamide, 5× SSC [pH 5], 1% SDS) for 30 min at 70°C. After three quick washes with solution 2 (0.5 M NaCl, 10 mM Tris-HCl [pH 7.5], 0.1% Tween-20) at 37°C, the embryos were treated with 100 μg/ml RNase A in solution 2 for 30 min at 37°C. They were rinsed in solution 2, then in solution 3 (50% formamide, 2× SSC [pH 5]), washed twice in solution 3 at 65°C for 30 min, and three times in TBST (150 mM NaCl, 100 mM Tris-HCl [pH 7.5], 0.1% Tween-20) for 5 min each time. After treatment with 1.5% blocking reagent (Roche) in TBST for 1 hr, the embryos were incubated overnight at 4°C with alkaline phosphatase-conjugated anti-digoxigenin antibody diluted to 1/2500 and preabsorbed with chick embryo powder. The embryos were then washed several times in TBST for 1 hr each at 4°C, followed by an overnight wash. After three washings in NTMT (100 mM NaCl, 100 mM Tris-HCl [pH 9.5], 50 mM $MgCl_2$, 0.1% Tween-20, 2 mM levamisole) for 10 min each time, the embryos were incubated in NTMT containing 4.5 μl NBT and 3.5 μl BCIP per ml or in BM purple (Roche) until the color developed. After the staining reaction was terminated, the embryos were fixed further

Table 1. Identified Enhancers of *Sox2* Gene and Their Conservation across Amniote Species

Conserved Sequence Block	Enhancer	Assigned Activity before st. 20 (the Earliest Stage of Enhancer Activity)	Enhancer Activity at st. 23	Length in Chicken Sequence (bp)	Position in Chicken Sequence[e]	Similarity in Human Sequence (%)	Similarity in Mouse Sequence (%)
1		None	None[d]	496	−16175 to −15680	77	68
2	N-3	Di/mesencepalon early lens (st. 8)		582	−15133 to −14552	89	71
3		None		487	−11102 to −10616	62	Absent[f]
4		None		875	−10068 to −9194	71	68
5	NOP-1	Nasal and otic placodes (st. 11)	None[d]	323	−8311 to −7989	83	81
6		None		145	−7803 to −7659	75	68
7		None		143	−5310 to −5168	75	71
8	N-2	Tel/di/mesencephalon (st. 5[c])	None[d]	534	−3012 to −2479	90	87
9		None		147	−1198 to −1052	84	81
10[a]		None[b]	None	2768	−662 to 2106	69	69
11		None	None[d]	103	3427 to 3529	63	61
12		None		287	4314 to 4600	83	83
13	NOP-2	Nasal and otic placodes (st. 11)		374	5105 to 5478	85	77
14	N-5	Rhombencephalon (r2–r4) (st. 12)	None	366	8151 to 8516	75	72
15		None	None	188	9872 to 10059	71	71
16		None	None	344	10445 to 10788	87	86
17	N-1	Neural induction (st. 5[c])	None	298	12825 to 13122	83	72
18	SC-1	None	Spinal cord	413	13686 to 14098	81	65
19		None	None	362	15041 to 15402	84	80
20		None	None	127	16389 to 16515	74	61
21		None	None	292	16988 to 17279	61	Absent[f]
22	SC-2	None	Spinal cord	171	18450 to 18620	70	69
23		None	None	158	20551 to 20708	67	(55)[g]
24	N-4	Head ectoderm Mesencephalon Spinal cord (st. 10)	Spinal cord	480	22339 to 22818	78	73
	L	Late lens (st. 17)	None	224	26592 to 26815	Absent[f]	Absent[f]
25	NC-1	None	Dorsal root ganglia	676	29610 to 30285	85	Absent[f]

[a] The block including the *Sox2* coding region.
[b] The *Sox2* promoter (Wiebe et al., 2000) did not show any significant specificity in the assay.
[c] The earliest stage of detectable EGFP expression after electroporation at st. 4. The enhancer activity must have been present before st. 5.
[d] Large fragments including two to four conserved blocks were analyzed.
[e] The position of the *Sox2* translational initiation site is taken as +1.
[f] "Absent" indicates no corresponding blocks identified even under less stringent conditions for detecting a sequence similarity.
[g] A block of sequence with similarity score of 55% is found in the corresponding region, although the score does not satisfy the condition for registering as a conserved sequence block (similarity higher than 60%).

with 4% PFA in PBT overnight at 4°C. When in situ hybridization was done in sequence using two different probes, the color reaction was done with INT/BCIP for the first hybridization reaction. After photographic recording, the developed pigments were removed by washing in methanol and then the second probe was applied for hybridization reaction, and the color of the signal was developed using NBT/BCIP.

Acknowledgments

We thank members of the Kondoh laboratory for stimulating discussions. This work was supported by grants from the Ministry of Education, Science, Sports, and Culture of Japan, and from the Human Frontier Science Program.

Received: December 3, 2002
Revised: February 4, 2003
Published: April 7, 2003

References

Brown, J.M., and Storey, K.G. (2000). A region of the vertebrate neural plate in which neighbouring cells can adopt neural or epidermal fates. Curr. Biol. *10*, 869–872.

Cai, J., Wu, Y., Mirua, T., Pierce, J.L., Lucero, M.T., Albertine, K.H., Spangrude, G.J., and Rao, M.S. (2002). Properties of a fetal multipotent neural stem cell (NEP cell). Dev. Biol. *251*, 221–240.

Clevers, H., and van de Wetering, M. (1997). TCF/LEF factor earn their wings. Trends Genet. *13*, 485–489.

Collignon, J., Sockanathan, S., Hacker, A., Cohen-Tannoudji, M., Norris, D., Rastan, S., Stevanovic, M., Goodfellow, P.N., and Lovell-Badge, R. (1996). A comparison of the properties of Sox-3 with Sry and two related genes, Sox-1 and Sox-2. Development *122*, 509–520.

Dailey, L., and Basilico, C. (2001). Coevolution of HMG domains and homeodomains and the generation of transcriptional regulation by Sox/POU complexes. J. Cell Physiol. *186*, 315–328.

Darnell, D.K., Stark, M.R., and Schoenwolf, G.C. (1999). Timing and cell interactions underlying neural induction in the chick embryo. Development *126*, 2505–2514.

Fernandez-Garre, P., Rodriguez-Gallardo, L., Gallego-Diaz, V., Alvarez, I.S., and Puelles, L. (2002). Fate map of the chicken neural plate at stage 4. Development *129*, 2807–2822.

Gottgens, B., Barton, L.M., Gilbert, J.G., Bench, A.J., Sanchez, M.J., Bahn, S., Mistry, S., Grafham, D., McMurray, A., Vaudin, M., et al. (2000). Analysis of vertebrate SCL loci identifies conserved enhancers. Nat. Biotechnol. *18*, 181–186.

Hamburger, V., and Hamilton, H.L. (1951). A series of normal stages in the development of the chick embryo. J Morphol. *88*, 49–92.

Hardison, R.C. (2000). Conserved noncoding sequences are reliable guides to regulatory elements. Trends Genet. *16*, 369–372.

Ishii, Y., Rex, M., Scotting, P.J., and Yasugi, S. (1998). Region-specific expression of chicken Sox2 in the developing gut and lung epithelium: regulation by epithelial-mesenchymal interactions. Dev. Dyn. *213*, 464–475.

Josephson, R., Muller, T., Pickel, J., Okabe, S., Reynolds, K., Turner, P.A., Zimmer, A., and McKay, R.D. (1998). POU transcription factors control expression of CNS stem cell-specific genes. Development *125*, 3087–3100.

Kamachi, Y., Sockanathan, S., Liu, Q., Breitman, M., Lovell-Badge, R., and Kondoh, H. (1995). Involvement of SOX proteins in lens-specific activation of crystallin genes. EMBO J. *14*, 3510–3519.

Kamachi, Y., Uchikawa, M., Collignon, J., Lovell-Badge, R., and Kondoh, H. (1998). Involvement of Sox1, 2 and 3 in the early and subsequent molecular events of lens induction. Development *125*, 2521–2532.

Kamachi, Y., Uchikawa, M., and Kondoh, H. (2000). Pairing SOX off: with partners in the regulation of embryonic development. Trends Genet. *16*, 182–187.

Kamachi, Y., Uchikawa, M., Tanouchi, A., Sekido, R., and Kondoh, H. (2001). Pax6 and SOX2 form a co-DNA-binding partner complex that regulates initiation of lens development. Genes Dev. *15*, 1272–1286.

Kishi, M., Mizuseki, K., Sasai, N., Yamazaki, H., Shiota, K., Nakanishi, S., and Sasai, Y. (2000). Requirement of Sox2-mediated signaling for differentiation of early *Xenopus* neuroectoderm. Development *127*, 791–800.

Li, H.S., Yang, J.M., Jacobson, R.D., Pasko, D., and Sundin, O. (1994). Pax-6 is first expressed in a region of ectoderm anterior to the early neural plate: implications for stepwise determination of the lens. Dev. Biol. *162*, 181–194.

Mathis, L., Kulesa, P.M., and Fraser, S.E. (2001). FGF receptor signalling is required to maintain neural progenitors during Hensen's node progression. Nat. Cell Biol. *3*, 559–566.

Mayor, C., Brudno, M., Schwartz, J.R., Poliakov, A., Rubin, E.M., Frazer, K.A., Pachter, L.S., and Dubchak, I. (2000). VISTA: visualizing global DNA sequence alignments of arbitrary length. Bioinformatics *16*, 1046–1047.

Mizuseki, K., Kishi, M., Matsui, M., Nakanishi, S., and Sasai, Y. (1998). *Xenopus* Zic-related-1 and Sox-2, two factors induced by chordin, have distinct activities in the initiation of neural induction. Development *125*, 579–587.

Nishiguchi, S., Wood, H., Kondoh, H., Lovell-Badge, R., and Episkopou, V. (1998). Sox1 directly regulates the γ-crystallin genes and is essential for lens development in mice. Genes Dev. *12*, 776–781.

Pevny, L.H., and Lovell-Badge, R. (1997). Sox genes find their feet. Curr. Opin. Genet. Dev. *7*, 338–344.

Pevny, L.H., Sockanathan, S., Placzek, M., and Lovell-Badge, R. (1998). A role for SOX1 in neural determination. Development *125*, 1967–1978.

Pustell, J., and Kafatos, F.C. (1982). A high speed, high capacity homology matrix: zooming through SV40 and polyoma. Nucleic Acids Res. *10*, 4765–4782.

Rex, M., Orme, A., Uwanogho, D., Tointon, K., Wigmore, P.M., Sharpe, P.T., and Scotting, P.J. (1997). Dynamic expression of chicken Sox2 and Sox3 genes in ectoderm induced to form neural tissue. Dev. Dyn. *209*, 323–332.

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.

Schilham, M.W., and Clevers, H. (1998). HMG box containing transcription factors in lymphocyte differentiation. Semin. Immunol. *10*, 127–132.

Sekido, R., Murai, K., Funahashi, J., Kamachi, Y., Fujisawa-Sehara, A., Nabeshima, Y., and Kondoh, H. (1994). The δ-crystallin enhancer-binding protein δEF1 is a repressor of E2-box-mediated gene activation. Mol. Cell. Biol. *14*, 5692–5700.

Sekido, R., Murai, K., Kamachi, Y., and Kondoh, H. (1997). Two mechanisms in the action of repressor δEF1: binding site competition with an activator and active repression. Genes Cells *2*, 771–783.

Storey, K.G., Crossley, J.M., De Robertis, E.M., Norris, W.E., and Stern, C.D. (1992). Neural induction and regionalisation in the chick embryo. Development *114*, 729–741.

Streit, A., Sockanathan, S., Perez, L., Rex, M., Scotting, P.J., Sharpe, P.T., Lovell-Badge, R., and Stern, C.D. (1997). Preventing the loss of competence for neural induction: HGF/SF, L5 and Sox-2. Development *124*, 1191–1202.

Tomioka, M., Nishimoto, M., Miyagi, S., Katayanagi, T., Fukui, N., Niwa, H., Muramatsu, M., and Okuda, A. (2002). Identification of Sox-2 regulatory region which is under the control of Oct-3/4-Sox-2 complex. Nucleic Acids Res. *30*, 3202–3213.

Uchikawa, M., Kamachi, Y., and Kondoh, H. (1999). Two distinct subgroups of Group B Sox genes for transcriptional activators and repressors: their expression during embryonic organogenesis of the chicken. Mech. Dev. *84*, 103–120.

Uwanogho, D., Rex, M., Cartwright, E.J., Pearl, G., Healy, C., Scotting, P.J., and Sharpe, P.T. (1995). Embryonic expression of the chicken Sox2, Sox3 and Sox11 genes suggests an interactive role in neuronal development. Mech. Dev. *49*, 23–36.

Verschueren, K., Remacle, J.E., Collart, C., Kraft, H., Baker, B.S., Tylzanowski, P., Nelles, L., Wuytens, G., Su, M.T., Bodmer, R., et al. (1999). SIP1, a novel zinc finger/homeodomain repressor, interacts with Smad proteins and binds to 5′-CACCT sequences in candidate target genes. J. Biol. Chem. *274*, 20489–20498.

Wegner, M. (1999). From head to toes: the multiple facets of Sox proteins. Nucleic Acids Res. *27*, 1409–1420.

Wiebe, M.S., Wilder, P.J., Kelly, D., and Rizzino, A. (2000). Isolation, characterization, and differential expression of the murine Sox-2 promoter. Gene *246*, 383–393.

Wilson, M., and Koopman, P. (2002). Matching SOX: partner proteins and co-factors of the SOX family of transcriptional regulators. Curr. Opin. Genet. Dev. *12*, 441–446.

Wood, H.B., and Episkopou, V. (1999). Comparative expression of the mouse Sox1, Sox2 and Sox3 genes from pre-gastrulation to early somite stages. Mech. Dev. *86*, 197–201.

Zappone, M.V., Galli, R., Catena, R., Meani, N., De Biasi, S., Mattei, E., Tiveron, C., Vescovi, A.L., Lovell-Badge, R., Ottolenghi, S., and Nicolis, S.K. (2000). Sox2 regulatory sequences direct expression of a (β)-geo transgene to telencephalic neural stem cells and precursors of the mouse embryo, revealing regionalization of gene expression in CNS stem cells. Development *127*, 2367–2382.

Accession Numbers

The compiled 49.2 kb sequence of the chicken *Sox2* locus is in the GenBank/DDBJ databases (AB092842). The *Sox2* locus sequences of the mouse, human, and fugu are AL606746 (GenBank), AC117415 (GenBank), and CAAB01002806 (The Fugu Genome Sequencing Consortium), respectively.