Complex architecture and regulated expression of the *Sox2ot* locus during vertebrate development

PAULO P. AMARAL,[1] CHRISTINE NEYT,[1] SIMON J. WILKINS,[1] MARJAN E. ASKARIAN-AMIRI,[1] SUSAN M. SUNKIN,[2] ANDREW C. PERKINS,[1] and JOHN S. MATTICK[1]

[1]ARC Special Research Centre for Functional and Applied Genomics, Institute for Molecular Bioscience, The University of Queensland, St Lucia, QLD 4072, Australia
[2]Allen Institute for Brain Science, Seattle, Washington 98103, USA

ABSTRACT

The *Sox2* gene is a key regulator of pluripotency embedded within an intron of a long noncoding RNA (ncRNA), termed *Sox2 overlapping transcript* (*Sox2ot*), which is transcribed in the same orientation. However, this ncRNA remains uncharacterized. Here we show that *Sox2ot* has multiple transcription start sites associated with genomic features that indicate regulated expression, including highly conserved elements (HCEs) and chromatin marks characteristic of gene promoters. To identify biological processes in which *Sox2ot* may be involved, we analyzed its expression in several developmental systems, compared to expression of *Sox2*. We show that *Sox2ot* is a stable transcript expressed in mouse embryonic stem cells, which, like *Sox2*, is down-regulated upon induction of embryoid body (EB) differentiation. However, in contrast to *Sox2*, *Sox2ot* is up-regulated during EB mesoderm-lineage differentiation. In adult mouse, *Sox2ot* isoforms were detected in tissues where *Sox2* is expressed, as well as in different tissues, supporting independent regulation of expression of the ncRNA. *Sox2dot*, an isoform of *Sox2ot* transcribed from a distal HCE located >500 kb upstream of *Sox2*, was detected exclusively in the mouse brain, with enrichment in regions of adult neurogenesis. In addition, *Sox2ot* isoforms are transcribed from HCEs upstream of *Sox2* in other vertebrates, including in several regions of the human brain. We also show that *Sox2ot* is dynamically regulated during chicken and zebrafish embryogenesis, consistently associated with central nervous system structures. These observations provide insight into the structure and regulation of the *Sox2ot* gene, and suggest conserved roles for *Sox2ot* orthologs during vertebrate development.

Keywords: noncoding RNA; *Sox2*; embryonic stem cells; zebrafish; central nervous system

INTRODUCTION

Over the last decade, transcriptomic studies have revealed that most of the genome of higher eukaryotes is transcribed in developmentally regulated patterns, to produce a complex set of overlapping and interlacing coding and noncoding transcripts from both strands (Mattick and Makunin 2006; Willingham and Gingeras 2006). Over 30,000 putative full-length long (>200 nucleotides [nt]) non-protein-coding RNAs (ncRNAs) have been catalogued in mouse (Carninci et al. 2005), many of which are subjected to splicing, but the biological roles of only a small number have been investigated. Nevertheless, their specific expression patterns, conservation of promoters, splice site sequences and predicted structures, association with particular chromatin signatures, regulation by morphogens and transcription factors, alternative splicing and altered expression in disease states, as well as evidence of evolutionary selection, all suggest that many of these transcripts are functional (Cawley et al. 2004; Carninci et al. 2005; Washietl et al. 2005; Pang et al. 2006, 2007; Pollard et al. 2006; Ravasi et al. 2006; Torarinsson et al. 2006; Ponjavic et al. 2007; Rinn et al. 2007; Dinger et al. 2008a; Mercer et al. 2008; Guttman et al. 2009; Zhang et al. 2009). Indeed, ncRNAs have been shown to regulate gene expression through a large variety of mechanisms, including the control of chromatin structure, transcription, mRNA processing, and translation (Amaral et al. 2008) and are emerging as important players in gene regulation during eukaryotic development (Amaral and Mattick 2008).

Given the lack of currently known intrinsic features that might provide insight into specific long ncRNA functions,

Reprint requests to: John S. Mattick, ARC Special Research Centre for Functional and Applied Genomics, Institute for Molecular Bioscience, The University of Queensland, St Lucia, QLD 4072, Australia; e-mail: j.mattick@imb.uq.edu.au; fax: 61-7-3346-2111.

Article published online ahead of print. Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.1705309.

a common strategy to study ncRNAs has been the analysis of their expression and genomic relationship with adjacent protein-coding loci. This approach has successfully revealed *cis*-regulatory roles for several ncRNAs (Sleutels et al. 2002; Feng et al. 2006; Martianov et al. 2007; Beltran et al. 2008; Dinger et al. 2008a; Ebralidze et al. 2008; Pandey et al. 2008; Wang et al. 2008; Yu et al. 2008). Recently, we identified 174 long, spliced ncRNAs that are dynamically regulated during mouse embryonic stem (ES) cell differentiation, many of which are associated and coordinately expressed with genes that have important roles in the differentiation process, including several homeobox-containing genes (Dinger et al. 2008a). In fact, long ncRNA genes are commonly associated with transcription factor loci that control development and are proposed to be a general component in their regulation (Engstrom et al. 2006; Amaral and Mattick 2008). Moreover, a number of ncRNAs exhibit expression profiles similar to genes that have crucial roles in ES pluripotency, such as *Sox2*, *Nanog*, and *Pou5f1*, indicating that long ncRNAs may also participate in the regulatory networks that define the pluripotent state (Dinger et al. 2008a).

Sox2 (SRY-box containing gene 2) is a key regulatory factor involved in establishing and maintaining pluripotency in vertebrate stem cells. It is a highly conserved single-exon gene that encodes a high-mobility group transcription factor expressed very early in development, including in totipotent cells from the morula and inner cell mass of blastocysts, and its ablation causes early embryonic lethality in mice (Avilion et al. 2003; Masui et al. 2007). Conditional knockouts of *Sox2* have further demonstrated important roles in conserved ontogenetic processes, especially in neural development (Mizuseki et al. 1998; Bylund et al. 2003; Okuda et al. 2006). *Sox2* is required for both neural induction of the ectoderm and maintenance of progenitor and neural stem cells during embryonic and adult neurogenesis (Cai et al. 2002; Wegner and Stolt 2005; Cavallaro et al. 2008; Kim et al. 2008). *Sox2* is also expressed in glial precursors and proliferating astrocytes, but it is down-regulated in neurons and in quiescent astrocytes (Ferri et al. 2004; Komitova and Eriksson 2004; Miyagi et al. 2004, 2006; Bani-Yaghoub et al. 2006). However, the mechanisms determining its differential regulation in these neural cell lineages are unclear. In addition, *Sox2* is required for mouse retina and lens development (Kamachi et al. 2001; Taranova et al. 2006), and heterozygous deficiency in humans is associated with eye defects, such as microphthalmia and anophtalmia, as well as other abnormalities, including forebrain and hippocampal malformation, and epilepsy (Fantes et al. 2003; Kelberman et al. 2006, 2008; Sisodiya et al. 2006).

Human *SOX2* is required for ES self-renewal and pluripotency (Fong et al. 2008). It has recently generated great therapeutic interest as a reprogramming factor (along with Oct4/Klf4/c-Myc or Oct4/Nanog/Lin28) capable of inducing pluripotency when ectopically expressed in a variety of differentiated murine and human cells, the so-called induced pluripotent stem (iPS) cells (Takahashi and Yamanaka 2006; Okita et al. 2008; Park et al. 2008). Neural progenitor cells, which express *Sox2*, can be reprogrammed into iPS cells when transduced with only Oct4/Klf4, in which case endogenous Sox2 is thought to facilitate reprogramming (Eminli et al. 2008). Therefore, elucidating the mechanisms that regulate *Sox2* expression is important to better understand and manipulate its roles in the control of pluripotency.

Multiple mechanisms are implicated in the developmental and tissue-specific regulation of *Sox2*, notably transcriptional control via an array of highly conserved enhancers located upstream and downstream from the gene, which have stage- and tissue-specific activities (Zappone et al. 2000; Uchikawa et al. 2003, 2004; Catena et al. 2004; Miyagi et al. 2004).

Recently, a study reporting the involvement of mutations in *SOX2* in anophtalmia made the intriguing observation that the *SOX2* gene lies in an intron of a long multiexon ncRNA that is transcribed in the same orientation (Fantes et al. 2003). The described transcript in humans, named *SOX2OT* or "*Sox2 overlapping transcript*" (approved by the HUGO Gene Nomenclature Committee; also "non-protein-coding RNA 43," HGNC ID:20209), is 3.4 kb long, and shares 88% identity with orthologous sequences in mouse (Fantes et al. 2003). *SOX2OT* has been postulated to participate in *SOX2* regulation (Fantes et al. 2003), but its expression patterns and potential functions have not been studied.

Our group previously observed that at least one isoform of *Sox2ot* is expressed in the mouse brain in regions of adult neurogenesis (Mercer et al. 2008), indicating that this ncRNA may have roles in processes related to the *Sox2* gene. In this work, to obtain insights into its function and regulation, we examined the expression of *Sox2ot* transcripts in different developmental systems and its relationship to *Sox2* expression. We found that *Sox2ot* is dynamically expressed in different developmental processes in which *Sox2* plays important roles, including mouse ES cell and neural cell differentiation, but it is also present in tissues where *Sox2* is not detected. *Sox2ot* orthologs are expressed in different vertebrates, and are regulated during development of mouse, chicken, and zebrafish embryos, with a conserved association with the central nervous system. The expression profiles of *Sox2ot*, as well as its evolutionarily preserved genomic association with *Sox2*, suggest that it has conserved functions in vertebrate development, and that it may participate in the regulation of *Sox2* or related processes.

RESULTS

Sox2ot transcripts and locus structure

SOX2OT was originally reported as a spliced ncRNA mapping to human chromosome 3q26.3-q27, with an

intron overlapping the *SOX2* gene in the same transcriptional orientation (Fantes et al. 2003). We analyzed the orthologous region in mouse and observed expressed sequence tags (ESTs) derived from several alternative isoforms of *Sox2ot*, with multiple alternative promoters, splicing and polyadenylation sites (Fig. 1; Supplemental Fig. 1). AceView annotation (Thierry-Mieg and Thierry-Mieg 2006) supports 16 different canonical GT-AG introns in at least 13 splicing variants and six alternative transcription start sites (TSSs). Four *Sox2ot* full-length sequences cloned from mouse tissues (Strausberg et al. 2002; Numata et al. 2003; Carninci et al. 2005) range in size from 638 nt (GenBank accession no. BY721402) to ~3.5 kb (accession no. AK031919).

Analysis of protein-coding potential of mouse and human *Sox2ot/SOX2OT* full-length cloned sequences with the software CRITICA, which uses a combination of statistical and comparative parameters, such as open reading frame (ORF) length, synonymous versus nonsynonymous base substitution rates, and similarity to known proteins (Badger and Olsen 1999; Frith et al. 2006), indicates no significant coding potential, although it remains possible that small proteins or peptides are encoded within the transcript (Dinger et al. 2008b).

The existence of poly(A) and 5′ cap structures, inferred from full-length cloning strategies, suggests that *Sox2ot* is transcribed by RNA polymerase II (RNAP II), and therefore represents a putative "mRNA-like" ncRNA (Numata et al. 2003). Mapping of TSSs of *Sox2ot* determined from EST clones that include the 5′end of transcripts (Carninci et al. 2005; Kimura et al. 2006) showed several novel isoforms with alternative TSSs (Fig. 1A; Supplemental Fig. 1).

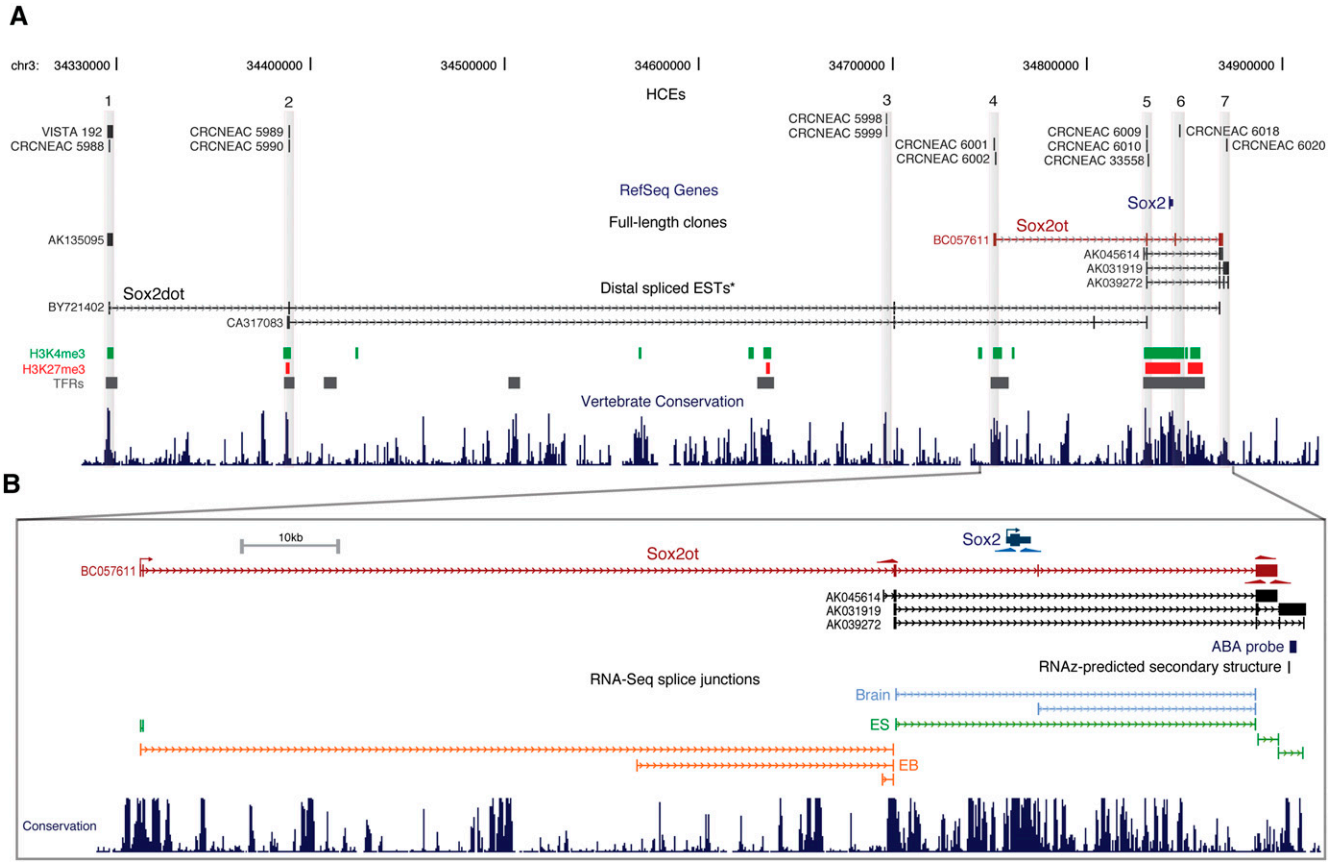

FIGURE 1. Genomic organization of *Sox2ot* locus in mouse. (*A*) Genome Browser (Hinrichs et al. 2006) view of *Sox2 overlapping transcripts* showing "Full-length clones" (mouse mRNAs from GenBank) and distal "spliced ESTs" tracks (accession numbers on the *left*). *The depicted ESTs represent some of *Sox2ot* splicing isoforms (for complete list of spliced ESTs see Supplemental Table 1), including "Sox2 distal overlapping transcript" (*Sox2dot*). Mouse highly conserved elements (HCEs 1–7) associated with *Sox2ot* start sites and exons in mouse or human (guide gray boxes, see Supplemental Fig. 1) are depicted by a Condor identifier (http://condor.fugu.biology.qmul.ac.uk/) (Woolfe et al. 2007). Also indicated are the Multiz 17-way vertebrate conservation (phastCons score histogram), TFRs, H3K4me3 (green), and H3K27me3 (red) chromatin marks from mouse brain obtained from UCSC Genome Browser tracks. (*B*) Detailed view of *Sox2ot* region proximal to *Sox2*, indicating positions of primer pairs used in RT-PCR analysis of *Sox2* (blue arrows) and *Sox2ot* expression (opposite arrows *below* and *above* transcript representing "Sox2ot a" and "Sox2ot b" primer pairs, respectively); splicing isoforms identified by available RNA-Seq data in mouse brain (blue), undifferentiated ES cells (green), and differentiating EBs (orange). The positions of an RNAz predicted secondary structure ($P > 0.9$) (Supplemental Fig. 3) and the Allen brain probe (ABA probe) used to determine *Sox2ot* expression (Mercer et al. 2008) in mouse brain sections are shown.

We also analyzed published chromatin modification maps of active (tri-methylated histone 3 lysine 4, H3K4me3) or repressed (trimethylated histone 3 lysine 27, H3K4me27) promoters obtained by chromatin immunoprecipitation-sequencing (ChIP-Seq) (Mikkelsen et al. 2007), and found support for multiple TSSs of *Sox2ot* (Fig. 1A). These promoter regions are associated with noncoding sequence blocks of high conservation in vertebrates (Woolfe et al. 2004, 2007), here termed highly conserved elements (HCEs) 1–7 encompassed in the *Sox2ot* locus. These blocks also largely correspond to "transposon free regions" (TFRs) (sequences over 5 kb long, which have remained resistant to transposon fixation throughout vertebrate evolution and appear to comprise extended regulatory sequences controlling the expression of genes that are central to early development) (Fig. 1A; Supplemental Fig. 1; Simons et al. 2006). The novel transcripts include a distal isoform associated with HCE 1, whose TSS is located more than 0.5 Mb upstream of *Sox2* in mouse, and is hereafter referred to as *Sox2dot* ("Sox2 distal overlapping transcript"). HCE 1 has been shown to have enhancer activity specific to the forebrain of developing mouse embryos (Supplemental Fig. 2; Pennacchio et al. 2006).

In addition, we identified a putative promoter between HCEs 2 and 3 that is marked by H3K4me3 and H3K27me3 modifications, as well as a TFR, but which is not associated with any mapped alternative TSS of *Sox2ot* or any EST sequence (Fig. 1A). However, we found a novel ncRNA associated with this element from a published collection of recently identified long intergenic ncRNAs (lincRNAs) (Guttman et al. 2009; data not shown). This, and additional ESTs located in the introns of *Sox2ot*, further highlight the transcriptional complexity associated with the *Sox2ot* locus.

Expression of *Sox2ot* in mouse tissues and ES cell differentiation

Inspection of the distribution of spliced ESTs in mouse tissues indicated a predominance of *Sox2ot* transcription in the central nervous system (CNS) (Supplemental Table 1). Moreover, analysis of RNA deep sequencing (RNA-Seq) data (Cloonan et al. 2008; Mortazavi et al. 2008) showed the presence of alternative splicing isoforms of *Sox2ot* in mouse brain, as well as in pluripotent and differentiating embryonic stem cells (Fig. 1B).

Because of the importance of *Sox2* in these systems, we first examined the regulation of *Sox2ot*, along with *Sox2*, in undifferentiated mouse ES cells and during differentiation as embryoid bodies (EBs) over a 16-d time course (see Materials and Methods) using quantitative real-time PCR (qRT-PCR). Similarly to *Sox2*, *Sox2ot* is expressed in pluripotent ES cells and is initially down-regulated upon EB differentiation (Fig. 2A). However, unlike *Sox2*, the ncRNA is reactivated after 6 d of differentiation (Fig. 2A), which coincides with induction of mesoderm differentiation programs (Bruce et al. 2007; Dinger et al. 2008a). We did not detect expression of the *Sox2dot* isoform in these samples at any time point (data not shown).

In addition, to confirm the transcription of *Sox2ot* by RNAPII and to assess transcript stability, we treated mouse ES cells with RNAPII inhibitor α-amanitin for 12, 24, and 48 h, and quantified the levels of *Sox2* and *Sox2ot* in treated and nontreated cells during EB differentiation. While *Sox2* mRNA levels are dramatically reduced after 12 h of treatment, *Sox2ot* levels were not significantly affected in the same period and significant reduction was observed only after 24 h of treatment (Fig. 2B). Similar results were observed using nondifferentiating ES cells treated with α-amanitin (data not shown), indicating that *Sox2ot* is more stable than *Sox2* mRNA in ES cells.

By examination of *Sox2ot* expression in mouse embryos (E10–E12 days post-fertilization [dpf]) and in adult brain using Northern blot, we detected a major band of ~3 kb (Fig. 2C). The expression was stronger in the brain, where multiple additional bands of ~1, 4, 6, and >10 kb, presumably consisting of alternative isoforms, can be detected. To further evaluate the regulation of the expression of *Sox2ot* RNAs and the relationship with *Sox2* in mouse, we analyzed their expression in mouse embryo, adult tissues, and cell lines by RT-PCR. By using a primer pair that detected *Sox2ot* last exon ("Sox2ot a") (Fig. 1B), the ncRNA was detected in mouse embryo (E10–E12 dpf) and in different adult mouse tissues, both in tissues where *Sox2* was also present, such as brain, lung, and heart (Fig. 2D, right panels), and in tissues where the mRNA was not detected, such as testis (Fig. 2D, left panels). Using primers spanning an exon–exon junction ("Sox2ot b") (Fig. 1B), a cassette exon was detected, which is differentially regulated in mouse tissues (Fig. 2D, cf. lower band of "Sox2ot b" in brain and ES cells). The bands corresponding to the different isoforms were sequenced to confirm their identity, and the regulated expression of isoforms was also evidenced by detection of different splicing junctions in the RNA-Seq data from different sources (Fig. 1B).

Sox2dot isoform was present in whole embryo and newborn mouse, but was detected only in the brain among the adult tissues tested (Fig. 2D). The expression of *Sox2dot* and isoforms of *Sox2ot* with alternative TSSs in the mouse brain is supported by the presence of H3K4me3 histone modifications associated with their promoter regions (Fig. 1A). In addition, the observation that H3K27me3 modifications can also be associated with alternative TSSs indicates that initiation of *Sox2ot* transcription from different promoters can be regulated at the chromatin level (Fig. 1A).

Sox2ot and *Sox2dot* are dynamically regulated during mouse neurogenesis

Sox2ot expression was previously observed in the mouse brain in zones of adult neurogenesis (Mercer et al. 2008).

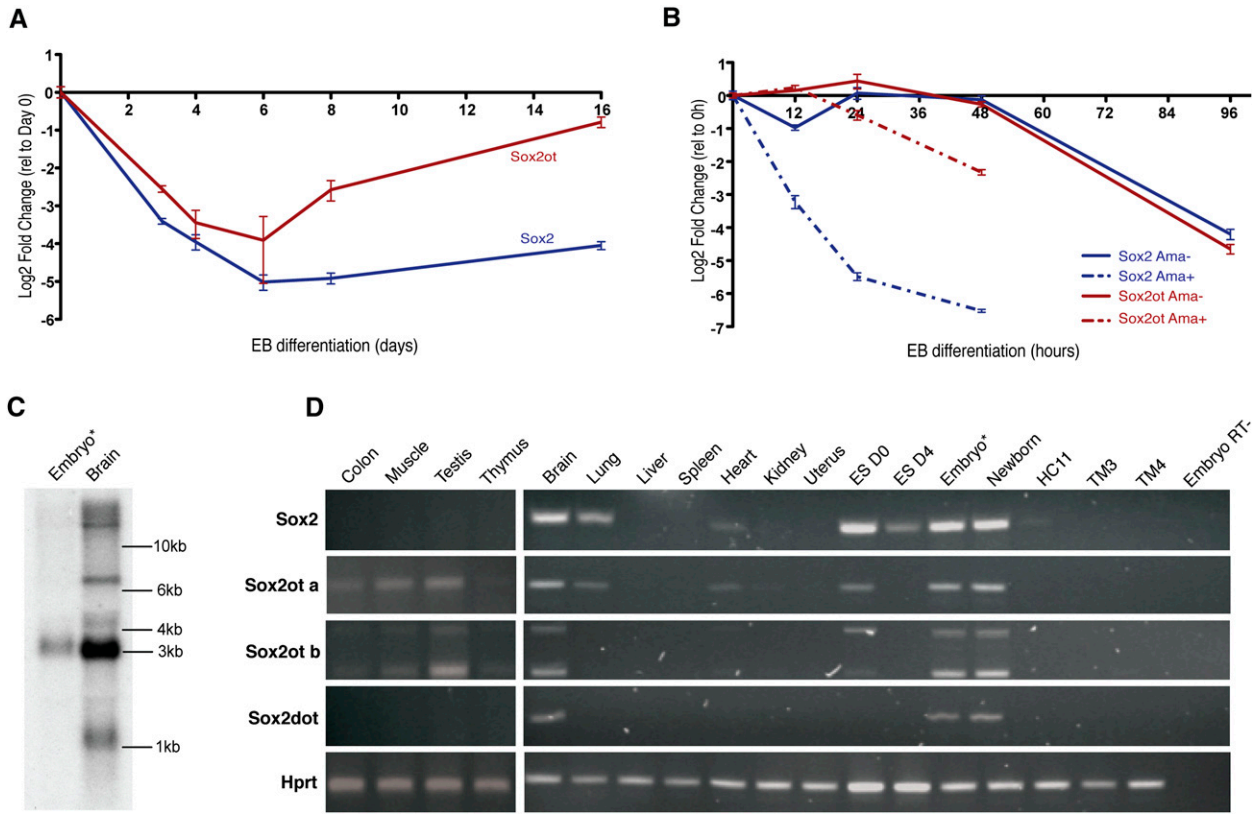

FIGURE 2. Expression profile of *Sox2* and *Sox2ot* in mouse ES cells and tissues. (*A*) qRT-PCR analysis of *Sox2* and *Sox2ot* levels in EB over a 16-d differentiation time course (expressed relative to day 0), and (*B*) in differentiating EBs treated (solid lines) or nontreated (dashed lines) with α-amanitin (expressed relative to 0 h of treatment). Error bars show standard deviation (SD) for at least three replicates. (*C*) Northern blot detection of *Sox2ot* expression in mouse embryos and adult brain. (*D*) RT-PCR performed with different primer pairs (see Materials and Methods for primer sequences) detecting different regions and isoforms of *Sox2ot* ("*Sox2ot* a," "*Sox2ot* b," and *Sox2dot*). *Embryo: RNA pool from E10–E12 dpf embryos. RT-: No reverse transcriptase control with embryonic RNA.

This pattern was detected by in situ hybridization (ISH) with a probe targeting a long (~2.8 kb) last exon (Fig. 1B, GenBank accession no. AK031919). In addition, *Sox2ot* ESTs are found in mouse differentiating neural stem cells libraries (Supplemental Table 1), and its expression can be detected in a variety of mouse primary neuronal cell cultures (data not shown).

To further explore the possible involvement of *Sox2ot* in neural differentiation processes, we next examined the expression of the RNAs using the "neurosphere assay," an in vitro system of neurogenesis with cultures of neurospheres originated from neural stem cells and undifferentiated precursors present in the adult mouse subventricular zone (Reynolds and Rietze 2005, 2006). Neurospheres were cultured for up to 7 d in differentiation media, which progressively enriches for neurons and glial cells (Rietze and Reynolds 2006). Analysis of Beta-III tubulin (*Tubb3*) and Glial fibrillary acidic protein (*Gfap*) mRNA expression by qRT-PCR indicates an increasing enrichment for neurons and astrocytes, respectively, after 2 and 7 d of culture (Fig. 3A). *Sox2* and *Sox2ot* were present in the cultured neurospheres (passage 2) (Fig. 3A), observed at similar levels after eight passages (data not shown). In addition, both transcripts were concordantly up-regulated in neurospheres after 2 d of culture in differentiation medium, and slightly down-regulated after 7 d, a pattern similar to that observed for *Nestin* (Fig. 3A). *Sox2ot* absolute expression levels were significantly higher in these neural cell types than in ES cells (data not shown).

Sox2dot was expressed at low levels in neurospheres, but showed a marked increase after 7 d of culture, indicating specific expression in the more differentiated population of neurons and glial cells (Fig. 3B). We then performed ISH to examine the expression of *Sox2dot* in adult mouse brain sections. Like *Sox2ot*, *Sox2dot* was also detected in zones of adult neurogenesis (Zhao et al. 2008) in the subventricular zone, rostral migratory stream and olfactory bulb, with a higher intensity in the latter (Fig. 3C), which contains a higher number of differentiated neurons, consistent with the expression observed in the neurospheres enriched in differentiated cells.

Although the in vitro neurosphere differentiation system originates a heterogeneous population of cells, including progenitors and differentiating neuronal and glial cells

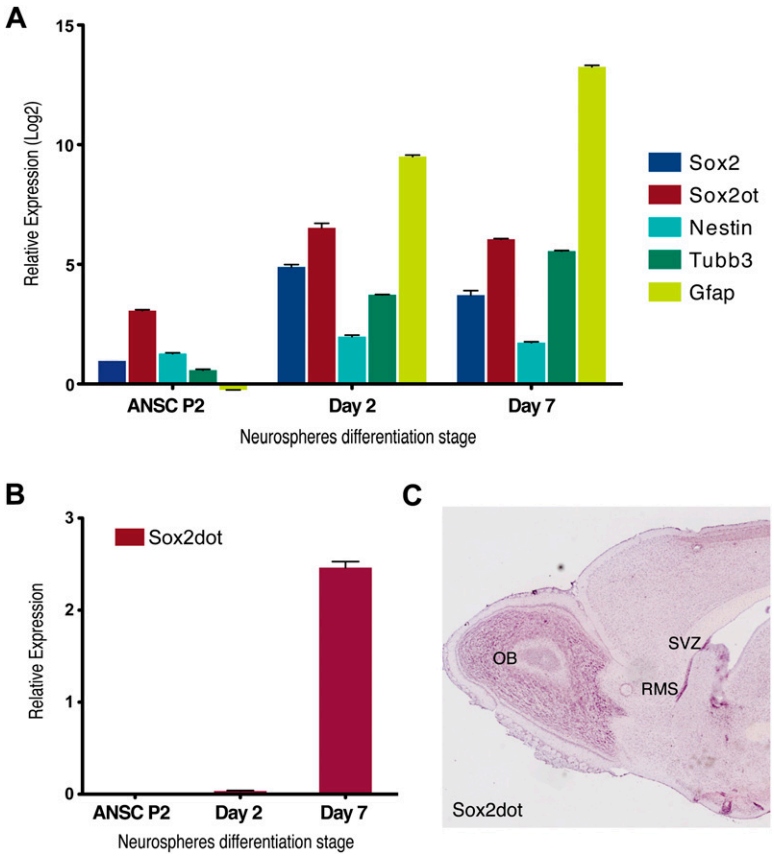

FIGURE 3. *Sox2ot* dynamic expression in mouse adult neurogenesis. Detection of (*A*) *Sox2* and *Sox2ot*, and (*B*) *Sox2dot*, in mouse differentiating neurospheres by qRT-PCR (expressed relative to *Hprt* levels). Levels of neural markers *Nestin*, Beta-III tubulin (*Tubb3*), and Glial fibrillary acidic protein (*Gfap*) were also measured, and error bars showing SD were determined from at least three replicates. (*C*) In situ hybridization of sagittal adult mouse brain section showed expression of *Sox2dot* in zones of adult neurogenesis. Detailed anterior view indicating the olfactory bulb (OB), rostral migratory stream (RMS), and subventricular zone (SVZ).

(Jensen and Parmar 2006), these results suggest that *Sox2ot* isoforms are present in neural cells and are dynamically regulated during differentiation. Interestingly, highly enriched binding of neuronal transcriptional repressor NRSF/REST has been previously validated in conserved sites upstream of *Sox2ot* (Mortazavi et al. 2006). NRSF/REST has been proposed to target several long ncRNAs in mouse and human neural cell differentiation models, and may represent a general mechanism of regulation of ncRNA expression in neuronal development (Johnson et al. 2009).

Sox2ot is transcribed from highly conserved elements in vertebrates

We observed that multiple *SOX2OT* TSSs in humans, determined by 5′-end sequencing of "oligo-cap" cDNA libraries (Kimura et al. 2006), are associated with orthologous HCEs (Supplemental Fig. 1; Supplemental Table 1). Analysis of ultraconserved elements (UCEs), blocks with at least 100 base pairs (bp) that have 100% conservation in placental mammals (Stephen et al. 2008), showed that the ~850 kb locus spanned by *SOX2OT/DOT* transcripts in humans contains 46 UCEs that occupy 1.09% of this region (average block size of 204 bp) (Supplemental Fig. 1A). This represents 15.6-fold enrichment relative to the entire genome (UCE coverage of 0.07%, average block size of 155 bp), underlying a high density of functional elements spanned by *SOX2OT* transcription. The longest UCE in the region in humans (646 bp long) corresponds to HCE 1, which in mouse comprises a sequence of 554 bp that presents 95% identity with human and 83% with zebrafish genomic sequences.

As observed in mouse, these HCEs associated with *SOX2OT* TSSs closely correspond to TFRs in humans; which further suggests that these sequences have been subjected to functional constraints across mammals (Supplemental Fig. 1). Moreover, HCEs upstream of *Sox2* are also identified in chicken (Uchikawa et al. 2004) and zebrafish genomes (Supplemental Fig. 1), and associated *Sox2ot* ESTs are not only found in these species, but also in other vertebrates with low EST coverage, such as *Xenopus*, cow, macaque, and orangutan (data not shown). Not all HCEs associated with *Sox2ot* TSSs in human and mouse are associated with TSSs in zebrafish (Supplemental Fig. 1), but this may be due to the small EST coverage and lack of 5′ end-cloned EST information in the latter.

At the primary sequence level, although Fantes et al. (2003) reported an identity of 88% of human *SOX2OT* transcribed from HCE 4 and 5 with mouse ESTs, the sequence similarity is restricted to specific regions of the transcript, and, because different exons have different conservation levels, the overall sequence conservation varies among alternative isoforms (Supplemental Fig. 1A). For example, the sequence of 2,971 bases [excluding the poly(A) tail] from *Sox2ot* isoform transcribed from HCE 4 in mouse (GenBank accession no. BC057611) is highly conserved in the rat genome, presenting ~90% identity in a continuous stretch of 2967 bases and all GT-AG splice sites conserved. However, the same sequence restricts to 384 nt with 91.2% identity in the human genome (concentrating in the exons immediately flanking *Sox2*, whose

splice sites are also conserved). These stretches can overlap highly conserved elements, such as HCE 7 that falls inside the last exons of mouse and human *Sox2ot/SOX2OT* (Supplemental Fig. 4). This element is part of an evolutionarily constrained region that encompasses a sequence of 106 nt predicted to form a conserved secondary structure (RNAz, $P > 0.9$) (Fig. 1B; Supplemental Fig. 3), which may represent a conserved functional element within *Sox2ot*. Finally, some exons of *SOX2OT* may be species- or clade-specific, as evidenced by the exon between HCEs 1 and 2, which is present in all *SOX2DOT* ESTs in humans and has splice site conserved only in primates, but is not detected in mouse (not present in ESTs nor detected by RT-PCR) (Supplemental Fig. 1A,B, data not shown).

Developmental regulation of *Sox2ot* in vertebrates

In light of the association of *Sox2ot* with highly conserved sequences and the identification of ESTs in different vertebrates, we evaluated whether *Sox2ot* transcripts are present and regulated during vertebrate embryonic development by examining the expression of *Sox2* and *Sox2ot/dot* orthologs in chicken and zebrafish embryos (Fig. 4). Using qRT-PCR, *Sox2ot* was detected during chicken development, with increasing levels observed after 2 d of embryonic development, similar to the pattern observed for *Sox2* (Fig. 4A), although with lower absolute expression levels. Expression of both *Sox2* and *Sox2ot* was enriched to the head of 6.5-d chicken embryos, compared to the limbs and whole body (Fig. 4A).

In zebrafish embryos, Northern blot analysis of *sox2ot* expression at 10, 30, and 78 h post-fertilization (hpf) stages showed a major band of ~2.5 kb at 30 hpf, as well as less intense bands of ~1.5 and ~6 kb, which are also present in the other stages (Supplemental Fig. 4A), indicating developmental regulation of alternative isoforms. Moreover, PCR using cDNA from zebrafish embryos and primers positioned in the first and last exons of *sox2ot* EST (accession no. EB947627), followed by sequencing, demonstrated the presence of different splicing isoforms (Supplemental Fig. 1C), including an isoform that skips the exon associated with HCE 5 and an isoform that contains a novel exon of 207 bases positioned between HCE 4 and 5 (deposited as GenBank accession nos. GO599877, GO599878, and GO599879).

In qRT-PCR analysis of a zebrafish developmental series from two-cell stage embryos to 5 dpf, *sox2* and *sox2ot* presented concordant up-regulation after sphere stage (Fig. 4B). This corresponds to the activation of zygotic transcription at midblastula transition, suggesting that *sox2* and *sox2ot* are both early zygotic genes in zebrafish. However, there is an apparent divergence in their expression patterns after tailbud stage when *sox2ot* total levels are reduced, although later in development at 5 dpf, both *sox2* and *sox2ot* are highly expressed (Fig. 4B). Interestingly, after zygotic activation, *sox2dot* is expressed at very low levels, but is dramatically induced at 5 dpf (Fig. 4D). By this developmental stage, the zebrafish has completed most aspects of organogenesis and already contains well-developed neuroanatomical features (Kimmel et al. 1995), a result that may parallel the expression of *Sox2dot* in mouse brain and in differentiated neurosphere cells.

To explore the temporospatial expression pattern of the transcripts in zebrafish embryos, we used whole-mount ISH with riboprobes for *sox2* and *sox2ot* at tailbud and 28 hpf stages. We detected presumptive neurectodermal expression of *sox2* at tailbud stage, as previously described (Okuda et al. 2006). At the same stage, *sox2ot* was also expressed in neurectodermal tissue, but with a more diffused pattern (Fig. 4C). Later at 28 hpf, both transcripts were highly expressed throughout the developing brain and eyes, and also at lower levels posteriorly in the neural tube (Fig. 4C). Specific expression of *sox2ot* in the retina and CNS was also observed by section ISH using 48 hpf embryos (Supplemental Fig. 4), and this was maintained in the brain throughout embryonic development until at least 6 dpf (data not shown). However, *sox2ot* expression at 28 hpf was broader, observed at high levels in the pharyngeal arches, but lower in some regions of *sox2* expression, such as the otic placodes. Moreover, at this stage, unlike *sox2*, *sox2ot* was also detected at posterior mesoderm, with distinct expression in the gut, lateral mesoderm, and posterior blood island (Fig. 4C), a pattern that was supported by section ISH with 48 hpf embryos (Supplemental Fig. 4). These results point to a complex expression pattern for *sox2ot,* which overlaps with *sox2* particularly in the CNS, but presents a broader distribution, especially in mesoderm tissues during development.

Finally, to gain insight into the expression of *SOX2OT* in humans, we used EST and published expression data information to examine transcript distribution. Available microarray expression data for several normal human tissues show the highest expression of *SOX2OT* in the brain and spinal cord (GEO Data Set Records GDS423 and GDS1085) (Shyamsundar et al. 2005; Yanai et al. 2005). qRT-PCR analysis of several human tissues confirmed that the levels of *SOX2OT* are orders of magnitude higher in the brain, where *SOX2* is also predominantly expressed, but is also detected in other tissues such as testis (Fig. 4E). In addition, we analyzed the presence of 202 spliced ESTs originating from the *SOX2OT*-associated HCEs in different EST libraries (Supplemental Table 1). The majority of *SOX2OT* spliced ESTs in humans (85% or 172/202 ESTs) originate from HCE 5, which is also the HCE with the highest number of associated *Sox2ot* ESTs in mouse (Supplemental Table 1). Consistent with the microarray and qRT-PCR data, the relative EST frequency of *SOX2OT* in human tissues (see Materials and Methods) indicates a predominance in CNS libraries, including fetal brain and eyes, with only 16 out of 202 spliced ESTs not assigned to

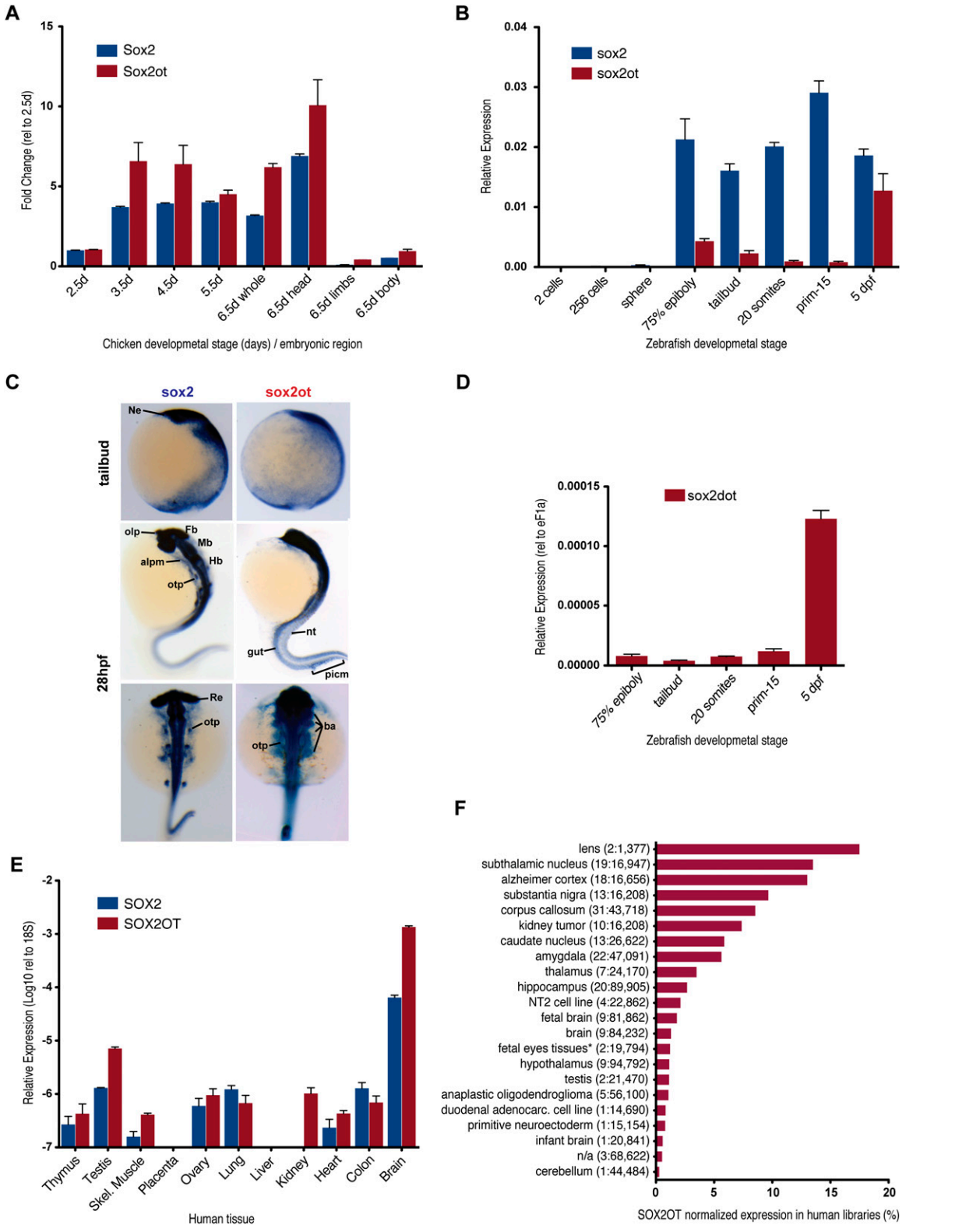

FIGURE 4. Expression of *Sox2ot* orthologs in different vertebrate species. (*A*) Relative expression of *Sox2* and *Sox2ot* in chicken embryos as determined by qRT-PCR (relative to 2.5-d embryos). (*B–D*) Expression of *sox2* and *sox2ot/dot* in zebrafish embryos determined by qRT-PCR (*B*,*D*) (expressed relative to *eF1α* levels), and whole mount in situ hybridization with zebrafish embryos at tailbud stage and 28 h post-fertilization (hpf) embryos (*C*; anterior to the *top* with lateral view of embryos in *upper* and *middle* panels, and dorsal view in *lower* panel). alpm: Anterior lateral plate mesoderm, ba: branchial arches, Fb: forebrain, Hb: hindbrain, Mb: midbrain, Ne: neurectoderm, nt: neural tube, olp: olfactory placode, otp: otic placode, picm: posterior intermediate cell mass, and Re: retina. (*E*) Expression of *SOX2OT* and *SOX2* in human tissues determined by qRT-PCR (relative to 18S rRNA levels). (*F*) Normalized EST distribution of *SOX2OT* ESTs in human tissue libraries (%), where (*x*:*n*) indicates the number of *SOX2OT* spliced ESTs (*x*) over the total number of ESTs (*n*) in each tissue (see Materials and Methods). n/a: Not available. Primer sequences are provided in Supplemental Table 2.

brain or eye tissues (of which two were obtained from testis and 10 from kidney tumor) (Fig. 4F), with similar results in mouse (Supplemental Table 1). The tissue with greatest enrichment in human *SOX2OT* ESTs is "lens," with a proportion of 17.5% of *SOX2OT* expression relative to all libraries (although this library has a smaller coverage, with ~1300 ESTs), followed by subthalamic nucleus (13.5%) and Alzheimer cortex (13%), as well as an enrichment in kidney tumor libraries (7.4%), while it is found in lowest relative abundance in cerebellum (0.3%). It is important to notice that we only considered spliced ESTs, for which there is information about the transcript orientation and connectivity to *SOX2OT*. Nonetheless, this distribution shows that *SOX2OT* is highly enriched in the human brain and suggests that its expression is spatially regulated.

DISCUSSION

Most RNAs transcribed from eukaryotic genomes have little or no protein-coding potential, and the vast majority is currently defined as "transcripts of unknown function" (TUF) (Willingham and Gingeras 2006). A first necessary step toward their characterization is to determine the tissues and physiological processes in which such RNAs are expressed and may play specific roles. Several studies have shown specific expression of large numbers of ncRNAs (Ravasi et al. 2006; Louro et al. 2007; Nakaya et al. 2007; Dinger et al. 2008a; Mercer et al. 2008; Sunwoo et al. 2008). Moreover, long ncRNAs are dynamically regulated during animal development, supporting the idea that they are involved in differentiation and developmental processes (Amaral and Mattick 2008).

An emerging common feature of the set of conserved genes for regulatory proteins that direct multicellular development is their genomic association with long ncRNAs, with accumulating examples of ncRNAs that regulate their associated genes at different levels (Amaral and Mattick 2008). This set includes the transcription factor *SOX2*, embedded within the intron of a putative long ncRNA, *SOX2OT*, which was proposed to play a role in the transcriptional regulation of *SOX2* expression (Fantes et al. 2003). In mouse, *Sox2* regulates a large number of genes (Boyer et al. 2005; Chen et al. 2008), and precise regulation of its dosage is critical for the temporal and spatial regulation of differentiation processes, with small changes dramatically altering the self-renewal and pluripotency capacity of ES cells (Taranova et al. 2006; Kopp et al. 2008). Indeed, *Sox2* is thought to function as a molecular "rheostat" controlling key transcriptional regulatory networks in stem cells (Boer et al. 2007; Kopp et al. 2008). However, the mechanisms regulating *Sox2* expression are still being elucidated, and include regulation at the chromatin structure level (Sikorska et al. 2008), multiple transcription factors (Boyer et al. 2005), and post-transcriptional regulation by microRNAs (Tay et al. 2008).

In this work, we examined the expression patterns of *Sox2ot* and *Sox2* in developmental systems, especially those where *Sox2* has well-characterized roles. We found that *Sox2ot* is a stable RNA transcribed in mouse ES cells, and its expression profile was similar to *Sox2* in mouse early differentiating EBs and neurospheres, as well as in some adult tissues. However, *Sox2ot* was also differentially regulated and present in tissues where *Sox2* was not detected, indicating that expression of *Sox2ot* does not simply reflect the transcriptional status of *Sox2*. Indeed, *sox2ot* expression in zebrafish embryos was broader than for *sox2*, including in mesodermal tissues, a result that may parallel the upregulation of *Sox2ot* in mouse ES differentiation into mesoderm lineages.

These expression patterns indicate that *Sox2* and *Sox2ot* may have complex functional relationships. Because *Sox2ot* expression overlapped with *Sox2* expression in all of the different systems examined, *Sox2ot* could be involved in the activation of *Sox2* in some tissues. If this is the case, the fact that *Sox2ot* is also present in tissues where *Sox2* is not detected indicates that it is not sufficient for the activation, but must act in conjunction with the different regulatory factors that control the expression of *Sox2* (see above). Nevertheless, although the conserved association between *Sox2ot* and *Sox2* in vertebrates supports a functional relationship, *Sox2ot* may have additional roles that are unrelated or tangential to the regulation of *Sox2*. The evolutionarily preserved association and the overlap in the expression patterns could be a consequence of the sharing of regulatory elements that control the expression of both *Sox2ot* and *Sox2*. Finally, regulation of *Sox2ot* by *Sox2* is also possible, but we did not find evidence of Sox2 binding in *Sox2ot* proximal promoter regions in published ChIP-Seq data for Sox2 in mouse ES cells (Chen et al. 2008; data not shown).

Supporting a role for *Sox2ot* in vertebrate development, we observed that the expression of the ortholog transcripts is differentially regulated in human tissues and dynamically modulated in zebrafish and chicken embryos. The observed expression was consistently associated with CNS tissues, which is a pattern observed for a large number of ncRNAs that may play roles in animal brain development and function (Amaral and Mattick 2008). Nevertheless, while the inferred promoter regions of *Sox2ot* are highly conserved, the primary sequence of the transcript has conservation in limited regions and contains repeat elements, features that are also present in other characterized ncRNAs (Amaral and Mattick 2008). In addition, these observations are consistent with previous findings that ncRNA sequences have more plasticity to evolve (Pang et al. 2006), but have also, on average, higher promoter conservation compared to protein-coding genes (Carninci et al. 2005; Ponjavic et al. 2007).

We observed that several *Sox2ot* isoforms originating from alternative TSSs are also associated with chromatin

modifications characteristic of bona fide promoters and with HCEs (Fig. 1; Supplemental Fig. 1). Indeed, it is known that when HCEs (which often have enhancer activity) are localized in introns, they are generally associated with alternative downstream TSSs from the host gene (Vavouri et al. 2007). Moreover, the upstream TSSs observed for *Sox2ot* isoforms are common in the organization of the metazoan transcriptome, in which many transcripts have distal 5′ exons, sometimes located several hundreds of kilobases upstream, and are expressed from alternate promoters that are tissue- or cell-line-specific (Kimura et al. 2006; Manak et al. 2006; Denoeud et al. 2007). We found evidence that *Sox2ot* alternative promoters can be differentially regulated, as indicated by *Sox2dot*, a novel isoform of *Sox2ot* that has a unique tissue expression pattern (detection in adult mouse restricted to brain) and different profiles in the ES and neurosphere differentiation assays. The differential regulation of *Sox2ot* isoforms, including the extensive occurrence of alternative splicing, may indicate specific roles for these different transcripts.

Similar to *Sox2ot*, the long spliced ncRNAs *Evf-1* and *Evf-2* are transcribed from alternative TSSs that correspond to ultraconserved enhancers associated with the *Dlx5/6* genes, which have important roles in development (Kohtz and Fishell 2004; Feng et al. 2006). Interestingly, in mouse neural cells, the *Evf-2* isoform was shown to mediate the transcriptional activity of the associated ultraconserved element by interacting and recruiting the Dlx2 transcriptional activator to the enhancer (Feng et al. 2006). Most recently, the function of *Evf-2* was analyzed in mouse embryos by insertion of premature termination sites that truncate *Evf-2,* which caused a specific neural phenotype and demonstrated regulation of the enhancer by the ncRNA in vivo (Bond et al. 2009). A similar mechanism has been suggested for ncRNAs transcribed from highly conserved elements associated with other *Dlx* genes (Dinger et al. 2008a). Given the high density of conserved elements in the locus and the association of *Sox2ot* transcripts with different HCEs, including *Sox2dot* with the HCE 1 neural enhancer, one possibility is that *Sox2ot* transcripts or transcription are also involved in regulation of enhancer activity.

Sikorska et al. (2008) showed that epigenetic modifications of the *SOX2* enhancers SRR1 and SRR2A correlate with high level of *SOX2* expression during cell differentiation (Sikorska et al. 2008), and here we observed that *Sox2ot*-associated HCEs are marked by chromatin activation and/or repression marks (Fig. 1A). In addition, the region surrounding *Sox2* flanked by *Sox2ot* exons corresponds to a very large TFR (>30 kb in mouse and human), and comprises a broad region (>20 kb) marked by histone modifications characteristic of transcriptionally active promoters (H3K4me3) in the mouse brain (Fig. 1A; Supplemental Fig. 1). This unusual chromatin structure has been reported in mouse ES cells, in which transcriptional elongation marks (H3K36me3) also extend far downstream (>15 kb) from the *Sox2* annotated 3′ end (Mikkelsen et al. 2007). Considering these features associated with the chromatin at *Sox2* enhancers and flanking region, it is also possible that *Sox2ot* plays a role in regulating chromatin structure, a rapidly emerging theme for regulatory RNAs (Rinn et al. 2007; Ebisuya et al. 2008; Pandey et al. 2008; Yu et al. 2008; for review, see Mattick et al. 2009). Indeed, recent evidence indicate that a large proportion of lincRNAs, which are also conserved long, spliced ncRNAs, directly associate with different chromatin modifying complexes and can affect the expression of genes that are targeted by these complexes (Khalil et al. 2009). Finally, a regulatory role for *Sox2ot* may depend on the act of transcription itself, as observed for specific ncRNA loci involved in chromatin and transcriptional remodeling in yeast and mammalian cells (Ebisuya et al. 2008; Hirota et al. 2008). However, the stability, regulated splicing patterns, predicted secondary structure, and conservation of *Sox2ot* transcripts (Fig. 1) suggest that the transcript has intrinsic functions.

Gain- and loss-of-function strategies to determine *Sox2ot* mechanism of action and relationship with *Sox2* are warranted, for example, using morpholinos to block *sox2ot* splicing in zebrafish embryos and siRNA knockdown in ES cells, an approach that has proved successful in identifying functional roles for an increasing number of ncRNAs (Mattick 2009). The elucidation of these roles may help to understand the different developmental processes, and potentially the related disorders, in which *Sox2ot* is expressed. Linkage disequilibrium analysis has already implicated the upstream regions of *SOX2OT* as a myopia common susceptibility locus, and the ncRNA was suggested to be a plausible candidate (Andrew et al. 2008), with perturbations in its expression patterns possibly involved, as has also been proposed for ncRNAs present in other disease susceptibility loci (Shirasawa et al. 2004; He et al. 2009).

In conclusion, this study provides insight into the range of structural isoforms, epigenetic modifications, and expression patterns of the *Sox2ot* gene and its transcripts. It also highlights the hidden complexity of such loci and of the biological processes in which they are likely to participate, consistent with the emerging view of the eukaryotic genome as an RNA machine (Amaral et al. 2008). Indeed, another ncRNA located ~100 kb downstream from *Sox2* was recently shown to be expressed specifically in mouse ES cells and to be a direct target of Sox2 and Oct4, as part of a collection of over a thousand novel spliced ncRNAs identified (Guttman et al. 2009). Similar to the *Sox2ot* locus, identification and characterization of unnanotated transcripts with orthologs in amenable model organisms may represent a general approach to explore the roles of ncRNAs in developmental processes.

MATERIALS AND METHODS

Mouse tissue samples and cell culture

Mouse W9.5 ES cells at low passage number were maintained in 15% fetal calf serum (FCS) on mitotically inactive MEFs with 1000 U/mL LIF, as previously described (Bruce et al. 2007; Dinger et al. 2008a). Embryoid bodies were differentiated in Dulbecco's modified Eagle's medium (DMEM) containing 10% FCS in 1% methylcellulose (GIBCO 10912-012), which efficiently induces mesoderm-differentiation programs (Bruce et al. 2007; Dinger et al. 2008a). For RNA Polymerase II inhibition experiments, ES cells were seeded at 1×10^5 cells/mL in 6 cm plates in 10% FCS DMEM with or without 20 μg/mL α-amanitin (Sigma A2263) (and with or without LIF) and harvested after 0, 12, 24, and 48 h of culture. Whole neurospheres were generated from adult neural stem cells and proliferating precursor cells harvested from the subventricular zone of 12 mo old mice, passaged and differentiated as described by Rietze and Reynolds (2006), and were kindly provided by Beatrice Large and Rod Rietze. Mouse mammary epithelial HC11 cells (Ball et al. 1988) were cultured in RPMI 1640 medium supplemented with 10% FCS, 5 μg/mL insulin, and 10 ng/mL EGF. Mouse testis-derived cell lines TM3 (ATCC no. CRL-1714) and TM4 (ATCC no. CRL-1715) were cultured in high-glucose DMEM medium containing 10% FCS, 2.5 mM L-glutamine and 0.5 mM sodium pyruvate.

Mouse tissues dissected from adult CD1 or C57BL mice, and embryos dissected from pregnant C57BL mice at 10, 11, and 12 dpc (E10–E12) were frozen in liquid nitrogen and stored at −80°C until later used for RNA extraction.

RNA isolation and RT-PCR expression analysis

RNA from zebrafish embryos, mouse tissues, and cell cultures was purified using Trizol (Invitrogen) or RNeasy Mini Kit (Qiagen) and treated with DNase I (Invitrogen), according to instructions of the manufacturers. RNA samples from chicken embryos were kindly provided by Evgeny Glazov, and total RNA from human tissues was purchased from Ambion (AM6000). cDNA preparation and quantitative real-time PCR (qRT-PCR) analysis were performed as described (Bruce et al. 2007). Primers were designed based on *Sox2ot* EST sequences available for different vertebrates, spanning splice sites in most cases (Supplemental Table 2), and PCR products were sequenced to confirm the identity of the fragments. For normalization of transcript expression levels, *Hprt* and *eF1α* were used as internal controls in mouse and zebrafish samples, respectively. For qRT-PCR analysis of α-amanitin inhibition experiments, as well as for human and chicken expression analysis, cDNA was produced using random primers and expression values normalized relative to 18S RNA expression. For mouse tissue expression analysis, cDNA was used in PCR for 35 cycles and amplification products were visualized after electrophoresis in 2%–3% agarose gels. For in situ hybridization (ISH) probe preparation, cDNA from ES cells, and mouse or zebrafish embryos was amplified (see primers in Supplemental Table 2) and PCR products were cloned into pGEM-T Easy Vectors (Promega) (sequenced to confirm the identity of the insert), which were then used in PCR with T7 and SP6 primers to generate DNA templates for in vitro transcription reactions. ISH probe sequences are listed in Supplemental Table 2.

Northern blot analysis

For Northern blot analysis, 10 μg of total RNA were electrophoresed at 90 V for 3 h and 20 min through 1.2% agarose gels containing 1.1% formaldehyde in buffer 1X MOPS, 0.2 M sodium acetate (pH 7.0), and 0.01 M Na_2EDTA (1 M stock solution to pH 8.0). After electrophoresis, RNA gels were rinsed twice in distilled water and then washed in 10X SSC (3 M NaCl, 0.3 M sodium citrate at pH 7.0) at room temperature for 45 min. RNA was transferred overnight to Hybond-N+ membranes (Amersham Biosciences) by capillary blotting and fixed to the membranes by baking at 80°C for 2 h. The blots were probed with ^{32}P-labeled probes (Supplemental Table 2), which were prepared by PCR and purified amplification products labeled using Megaprime DNA Labeling Systems (Amersham Biosciences), following the manufacturer's instructions. Band size was estimated using 0.5–10 kb RNA Ladder (Invitrogen).

Zebrafish material and in situ hybridization

AB strain zebrafish stocks were reared and embryos obtained as described (Westerfield 1994). Embryos were incubated at 28.5°C and staged as previously described (Kimmel et al. 1995). All experiments were performed subject to animal ethics committee approval (AEC no. IMB/167/06/ARC).

For whole-mount in situ hybridization with zebrafish embryos, riboprobe templates were generated by PCR using primers that included T7/SP6 promoter sequences as described above. Riboprobe synthesis and in situ hybridizations were performed as previously described (Wilkins et al. 2008), and section ISH performed with 7 μm sections, as described by Wilhelm et al. (2007).

Electronic expression and secondary structure analysis

Available microarray expression data for *SOX2OT* in several normal human tissues were obtained from the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/), deposited under GEO Data set Record GDS423, which contains expression data for 12 normal human tissues interrogated with Affymetrix GeneChip HG-U95B (ProbeSet 55950) (Yanai et al. 2005), and Data set Record GDS1085, which analyzed the expression profile of 35 different types of normal tissues (ProbeSet 34804) (Shyamsundar et al. 2005).

For electronic expression analysis, *SOX2OT* spliced ESTs in human tissue libraries present in GenBank were catalogued, and the number of *SOX2OT* EST clones was computed for each tissue library (Supplemental Table 1). *SOX2OT* EST frequency (f) in each tissue was calculated by dividing the number of *SOX2OT* EST clones (x) by the total number of clones per tissue (n). Finally, the "relative *SOX2OT* expression" in human tissues was obtained by normalizing the frequency in each library (f) by the total number of *SOX2OT* ESTs in all libraries (expressed as a percentile) (Fig. 4F; Supplemental Table 1).

In the secondary structure analysis, a 106 nt region associated to the conserved sequence of *Sox2ot* was identified by RNAz (Washietl et al. 2005) with $P > 0.9$, as described by Dinger et al. (2008a). Additional secondary structure analysis was performed using the most recent RNAalifold algorithm with RIBOSUM scoring

(http://rna.tbi.univie.ac.at/cgi-bin/RNAalifold.cgi). RNAalifold predicts the consensus structure from a set of aligned sequences taking into account both thermodynamic stability and sequence covariation (Bernhart et al. 2008). The 17-Way species multiple genome alignment from the UCSC genome browser was used as input.

SUPPLEMENTAL MATERIAL

Supplemental material can be found at http://www.rnajournal.org.

ACKNOWLEDGMENTS

We thank Ke-lin Ru for laboratory assistance, Martin Smith for assistance with secondary structure analysis, Beatrice Large and Rod Rietze for providing cultured neurospheres, Evgeny Glazov for providing chicken embryonic RNA, and Marcel Dinger for critical reading of the manuscript. This work was supported by the Australian Research Council (Grant nos. FF0561986 and S00001543) and The University of Queensland, and the experiments with zebrafish were supported by the Australian Zebrafish Phenomics Facility (NHMRC Grant no. 455871).

Received April 24, 2009; accepted August 18, 2009.

REFERENCES

Amaral PP, Mattick JS. 2008. Noncoding RNA in development. *Mamm Genome* **19:** 454–492.

Amaral PP, Dinger ME, Mercer TR, Mattick JS. 2008. The eukaryotic genome as an RNA machine. *Science* **319:** 1787–1789.

Andrew T, Maniatis N, Carbonaro F, Liew SH, Lau W, Spector TD, Hammond CJ. 2008. Identification and replication of three novel myopia common susceptibility gene loci on chromosome 3q26 using linkage and linkage disequilibrium mapping. *PLoS Genet* **4:** e1000220. doi: 10.1371/journal.pgen.1000220.

Avilion AA, Nicolis SK, Pevny LH, Perez L, Vivian N, Lovell-Badge R. 2003. Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes & Dev* **17:** 126–140.

Badger JH, Olsen GJ. 1999. CRITICA: Coding region identification tool invoking comparative analysis. *Mol Biol Evol* **16:** 512–524.

Ball RK, Friis RR, Schoenenberger CA, Doppler W, Groner B. 1988. Prolactin regulation of β-casein gene expression and of a cytosolic 120-kd protein in a cloned mouse mammary epithelial cell line. *EMBO J* **7:** 2089–2095.

Bani-Yaghoub M, Tremblay RG, Lei JX, Zhang D, Zurakowski B, Sandhu JK, Smith B, Ribecco-Lutkiewicz M, Kennedy J, Walker PR, et al. 2006. Role of Sox2 in the development of the mouse neocortex. *Dev Biol* **295:** 52–66.

Beltran M, Puig I, Pena C, Garcia JM, Alvarez AB, Pena R, Bonilla F, de Herreros AG. 2008. A natural antisense transcript regulates *Zeb2/Sip1* gene expression during Snail1-induced epithelial-mesenchymal transition. *Genes & Dev* **22:** 756–769.

Bernhart SH, Hofacker IL, Will S, Gruber AR, Stadler PF. 2008. RNAalifold: Improved consensus structure prediction for RNA alignments. *BMC Bioinformatics* **9:** 474. doi: 10.1186/1471-2105-9-474.

Boer B, Kopp J, Mallanna S, Desler M, Chakravarthy H, Wilder PJ, Bernadt C, Rizzino A. 2007. Elevating the levels of Sox2 in embryonal carcinoma cells and embryonic stem cells inhibits the expression of *Sox2:Oct-3/4* target genes. *Nucleic Acids Res* **35:** 1773–1786.

Bond AM, Vangompel MJ, Sametsky EA, Clark MF, Savage JC, Disterhoft JF, Kohtz JD. 2009. Balanced gene regulation by an embryonic brain ncRNA is critical for adult hippocampal GABA circuitry. *Nat Neurosci* **12:** 1020–1027.

Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, et al. 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122:** 947–956.

Bruce SJ, Gardiner BB, Burke LJ, Gongora MM, Grimmond SM, Perkins AC. 2007. Dynamic transcription programs during ES cell differentiation toward mesoderm in serum versus serum-freeBMP4 culture. *BMC Genomics* **8:** 365. doi: 10.1186/1471-2164-8-365.

Bylund M, Andersson E, Novitch BG, Muhr J. 2003. Vertebrate neurogenesis is counteracted by Sox1-3 activity. *Nat Neurosci* **6:** 1162–1168.

Cai J, Wu Y, Mirua T, Pierce JL, Lucero MT, Albertine KH, Spangrude GJ, Rao MS. 2002. Properties of a fetal multipotent neural stem cell (NEP cell). *Dev Biol* **251:** 221–240.

Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, Oyama R, Ravasi T, Lenhard B, Wells C, et al. 2005. The transcriptional landscape of the mammalian genome. *Science* **309:** 1559–1563.

Catena R, Tiveron C, Ronchi A, Porta S, Ferri A, Tatangelo L, Cavallaro M, Favaro R, Ottolenghi S, Reinbold R, et al. 2004. Conserved POU binding DNA sites in the Sox2 upstream enhancer regulate gene expression in embryonic and neural stem cells. *J Biol Chem* **279:** 41846–41857.

Cavallaro M, Mariani J, Lancini C, Latorre E, Caccia R, Gullo F, Valotta M, DeBiasi S, Spinardi L, Ronchi A, et al. 2008. Impaired generation of mature neurons by neural stem cells from hypomorphic *Sox2* mutants. *Development* **135:** 541–557.

Cawley S, Bekiranov S, Ng HH, Kapranov P, Sekinger EA, Kampa D, Piccolboni A, Sementchenko V, Cheng J, Williams AJ, et al. 2004. Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell* **116:** 499–509.

Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orlov YL, Zhang W, Jiang J, et al. 2008. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133:** 1106–1117.

Cloonan N, Forrest AR, Kolle G, Gardiner BB, Faulkner GJ, Brown MK, Taylor DF, Steptoe AL, Wani S, Bethel G, et al. 2008. Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat Methods* **5:** 613–619.

Denoeud F, Kapranov P, Ucla C, Frankish A, Castelo R, Drenkow J, Lagarde J, Alioto T, Manzano C, Chrast J, et al. 2007. Prominent use of distal 5′ transcription start sites and discovery of a large number of additional exons in ENCODE regions. *Genome Res* **17:** 746–759.

Dinger ME, Amaral PP, Mercer TR, Pang KC, Bruce SJ, Gardiner BB, Askarian-Amiri ME, Ru K, Solda G, Simons C, et al. 2008a. Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res* **18:** 1433–1445.

Dinger ME, Pang KC, Mercer TR, Mattick JS. 2008b. Differentiating protein-coding and noncoding RNA: Challenges and ambiguities. *PLoS Comput Biol* **4:** e1000176. doi: 10.1371/journal.pcbi.1000176.

Ebisuya M, Yamamoto T, Nakajima M, Nishida E. 2008. Ripples from neighbouring transcription. *Nat Cell Biol* **10:** 1106–1113.

Ebralidze AK, Guibal FC, Steidl U, Zhang P, Lee S, Bartholdy B, Jorda MA, Petkova V, Rosenbauer F, Huang G, et al. 2008. PU.1 expression is modulated by the balance of functional sense and antisense RNAs regulated by a shared *cis*-regulatory element. *Genes & Dev* **22:** 2085–2092.

Eminli S, Utikal J, Arnold K, Jaenisch R, Hochedlinger K. 2008. Reprogramming of neural progenitor cells into induced pluripotent stem cells in the absence of exogenous Sox2 expression. *Stem Cells* **26:** 2467–2474.

Engstrom PG, Suzuki H, Ninomiya N, Akalin A, Sessa L, Lavorgna G, Brozzi A, Luzi L, Tan SL, Yang L, et al. 2006. Complex loci in

human and mouse genomes. *PLoS Genet* **2:** e47. doi: 10.1371/journal.pgen.0020047.
Fantes J, Ragge NK, Lynch SA, McGill NI, Collin JR, Howard-Peebles PN, Hayward C, Vivian AJ, Williamson K, van Heyningen V, et al. 2003. Mutations in SOX2 cause anophthalmia. *Nat Genet* **33:** 461–463.
Feng J, Bi C, Clark BS, Mady R, Shah P, Kohtz JD. 2006. The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator. *Genes & Dev* **20:** 1470–1484.
Ferri AL, Cavallaro M, Braida D, Di Cristofano A, Canta A, Vezzani A, Ottolenghi S, Pandolfi PP, Sala M, DeBiasi S, et al. 2004. Sox2 deficiency causes neurodegeneration and impaired neurogenesis in the adult mouse brain. *Development* **131:** 3805–3819.
Fong H, Hohenstein KA, Donovan PJ. 2008. Regulation of self-renewal and pluripotency by Sox2 in human embryonic stem cells. *Stem Cells* **26:** 1931–1938.
Frith MC, Bailey TL, Kasukawa T, Mignone F, Kummerfeld SK, Madera M, Sunkara S, Furuno M, Bult CJ, Quackenbush J, et al. 2006. Discrimination of non-protein-coding transcripts from protein-coding mRNA. *RNA Biol* **3:** 40–48.
Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, et al. 2009. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458:** 223–227.
He H, Nagy R, Liyanarachchi S, Jiao H, Li W, Suster S, Kere J, de la Chapelle A. 2009. A susceptibility locus for papillary thyroid carcinoma on chromosome 8q24. *Cancer Res* **69:** 625–631.
Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, Diekhans M, Furey TS, Harte RA, Hsu F, et al. 2006. The UCSC Genome Browser Database: Update 2006. *Nucleic Acids Res* **34:** D590–D598.
Hirota K, Miyoshi T, Kugou K, Hoffman CS, Shibata T, Ohta K. 2008. Stepwise chromatin remodelling by a cascade of transcription initiation of noncoding RNAs. *Nature* **456:** 130–134.
Jensen JB, Parmar M. 2006. Strengths and limitations of the neurosphere culture system. *Mol Neurobiol* **34:** 153–161.
Johnson R, Teh CH, Jia H, Vanisri RR, Pandey T, Lu ZH, Buckley NJ, Stanton LW, Lipovich L. 2009. Regulation of neural macroRNAs by the transcriptional repressor REST. *RNA* **15:** 85–96.
Kamachi Y, Uchikawa M, Tanouchi A, Sekido R, Kondoh H. 2001. Pax6 and SOX2 form a co-DNA-binding partner complex that regulates initiation of lens development. *Genes & Dev* **15:** 1272–1286.
Kelberman D, Rizzoti K, Avilion A, Bitner-Glindzicz M, Cianfarani S, Collins J, Chong WK, Kirk JM, Achermann JC, Ross R, et al. 2006. Mutations within Sox2/SOX2 are associated with abnormalities in the hypothalamo-pituitary-gonadal axis in mice and humans. *J Clin Invest* **116:** 2442–2455.
Kelberman D, de Castro SC, Huang S, Crolla JA, Palmer R, Gregory JW, Taylor D, Cavallo L, Faienza MF, Fischetto R, et al. 2008. SOX2 plays a critical role in the pituitary, forebrain, and eye during human embryonic development. *J Clin Endocrinol Metab* **93:** 1865–1873.
Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, Thomas K, Presser A, Bernstein BE, van Oudenaarden A, et al. 2009. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci* **106:** 11667–11672.
Kim J, Chu J, Shen X, Wang J, Orkin SH. 2008. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* **132:** 1049–1061.
Kimmel CB, Ballard WW, Kimmel SR, Ullmann B, Schilling TF. 1995. Stages of embryonic development of the zebrafish. *Dev Dyn* **203:** 253–310.
Kimura K, Wakamatsu A, Suzuki Y, Ota T, Nishikawa T, Yamashita R, Yamamoto J, Sekine M, Tsuritani K, Wakaguri H, et al. 2006. Diversification of transcriptional modulation: large-scale identification and characterization of putative alternative promoters of human genes. *Genome Res* **16:** 55–65.
Kohtz JD, Fishell G. 2004. Developmental regulation of EVF-1, a novel noncoding RNA transcribed upstream of the mouse Dlx6 gene. *Gene Expr Patterns* **4:** 407–412.
Komitova M, Eriksson PS. 2004. Sox-2 is expressed by neural progenitors and astroglia in the adult rat brain. *Neurosci Lett* **369:** 24–27.
Kopp JL, Ormsbee BD, Desler M, Rizzino A. 2008. Small increases in the level of Sox2 trigger the differentiation of mouse embryonic stem cells. *Stem Cells* **26:** 903–911.
Louro R, Nakaya HI, Amaral PP, Festa F, Sogayar MC, da Silva AM, Verjovski-Almeida S, Reis EM. 2007. Androgen responsive intronic noncoding RNAs. *BMC Biol* **5:** 4. doi: 10.1186/1741-7007-5-4.
Manak JR, Dike S, Sementchenko V, Kapranov P, Biemar F, Long J, Cheng J, Bell I, Ghosh S, Piccolboni A, et al. 2006. Biological function of unannotated transcription during the early development of *Drosophila melanogaster*. *Nat Genet* **38:** 1151–1158.
Martianov I, Ramadass A, Serra Barros A, Chow N, Akoulitchev A. 2007. Repression of the human dihydrofolate reductase gene by a noncoding interfering transcript. *Nature* **445:** 666–670.
Masui S, Nakatake Y, Toyooka Y, Shimosato D, Yagi R, Takahashi K, Okochi H, Okuda A, Matoba R, Sharov AA, et al. 2007. Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat Cell Biol* **9:** 625–635.
Mattick JS. 2009. The genetic signatures of noncoding RNAs. *PLoS Genet* **5:** e1000459.
Mattick JS, Makunin IV. 2006. Noncoding RNA. *Hum Mol Genet* **15:** R17–R29.
Mattick JS, Amaral PP, Dinger ME, Mercer TR, Mehler MF. 2009. RNA regulation of epigenetic processes. *Bioessays* **31:** 51–59.
Mercer TR, Dinger ME, Sunkin SM, Mehler MF, Mattick JS. 2008. Specific expression of long noncoding RNAs in the mouse brain. *Proc Natl Acad Sci* **105:** 716–721.
Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, et al. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448:** 553–560.
Miyagi S, Saito T, Mizutani K, Masuyama N, Gotoh Y, Iwama A, Nakauchi H, Masui S, Niwa H, Nishimoto M, et al. 2004. The Sox-2 regulatory regions display their activities in two distinct types of multipotent stem cells. *Mol Cell Biol* **24:** 4207–4220.
Miyagi S, Nishimoto M, Saito T, Ninomiya M, Sawamoto K, Okano H, Muramatsu M, Oguro H, Iwama A, Okuda A. 2006. The Sox2 regulatory region 2 functions as a neural stem cell-specific enhancer in the telencephalon. *J Biol Chem* **281:** 13374–13381.
Mizuseki K, Kishi M, Matsui M, Nakanishi S, 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.
Mortazavi A, Leeper Thompson EC, Garcia ST, Myers RM, Wold B. 2006. Comparative genomics modeling of the NRSF/REST repressor network: From single conserved sites to genome-wide repertoire. *Genome Res* **16:** 1208–1221.
Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5:** 621–628.
Nakaya HI, Amaral PP, Louro R, Lopes A, Fachel AA, Moreira YB, El-Jundi TA, da Silva AM, Reis EM, Verjovski-Almeida S. 2007. Genome mapping and expression analyses of human intronic noncoding RNAs reveal tissue-specific patterns and enrichment in genes related to regulation of transcription. *Genome Biol* **8:** R43. doi: 10.1186/gb-2007-8-3-r43.
Numata K, Kanai A, Saito R, Kondo S, Adachi J, Wilming LG, Hume DA, Hayashizaki Y, Tomita M. 2003. Identification of putative noncoding RNAs among the RIKEN mouse full-length cDNA collection. *Genome Res* **13:** 1301–1306.

Okita K, Nakagawa M, Hyenjong H, Ichisaka T, Yamanaka S. 2008. Generation of mouse induced pluripotent stem cells without viral vectors. *Science* **322:** 949–953.

Okuda Y, Yoda H, Uchikawa M, Furutani-Seiki M, Takeda H, Kondoh H, Kamachi Y. 2006. Comparative genomic and expression analysis of group B1 *sox* genes in zebrafish indicates their diversification during vertebrate evolution. *Dev Dyn* **235:** 811–825.

Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, Komorowski J, Nagano T, Mancini-Dinardo D, Kanduri C. 2008. *Kcnq1ot1* antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol Cell* **32:** 232–246.

Pang KC, Frith MC, Mattick JS. 2006. Rapid evolution of noncoding RNAs: Lack of conservation does not mean lack of function. *Trends Genet* **22:** 1–5.

Pang KC, Stephen S, Dinger ME, Engstrom PG, Lenhard B, Mattick JS. 2007. RNAdb 2.0—an expanded database of mammalian noncoding RNAs. *Nucleic Acids Res* **35:** D178–D182.

Park IH, Lerou PH, Zhao R, Huo H, Daley GQ. 2008. Generation of human-induced pluripotent stem cells. *Nat Protoc* **3:** 1180–1186.

Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, Minovitsky S, Dubchak I, Holt A, Lewis KD, et al. 2006. In vivo enhancer analysis of human conserved noncoding sequences. *Nature* **444:** 499–502.

Pollard KS, Salama SR, Lambert N, Lambot MA, Coppens S, Pedersen JS, Katzman S, King B, Onodera C, Siepel A, et al. 2006. An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* **443:** 167–172.

Ponjavic J, Ponting CP, Lunter G. 2007. Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome Res* **17:** 556–565.

Ravasi T, Suzuki H, Pang KC, Katayama S, Furuno M, Okunishi R, Fukuda S, Ru K, Frith MC, Gongora MM, et al. 2006. Experimental validation of the regulated expression of large numbers of noncoding RNAs from the mouse genome. *Genome Res* **16:** 11–19.

Reynolds BA, Rietze RL. 2005. Neural stem cells and neurospheres—re-evaluating the relationship. *Nat Methods* **2:** 333–336.

Rietze RL, Reynolds BA. 2006. Neural stem cell isolation and characterization. *Methods Enzymol* **419:** 3–23.

Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, et al. 2007. Functional demarcation of active and silent chromatin domains in human *HOX* loci by noncoding RNAs. *Cell* **129:** 1311–1323.

Shirasawa S, Harada H, Furugaki K, Akamizu T, Ishikawa N, Ito K, Ito K, Tamai H, Kuma K, Kubota S, et al. 2004. SNPs in the promoter of a B cell-specific antisense transcript, *SAS-ZFAT*, determine susceptibility to autoimmune thyroid disease. *Hum Mol Genet* **13:** 2221–2231.

Shyamsundar R, Kim YH, Higgins JP, Montgomery K, Jorden M, Sethuraman A, van de Rijn M, Botstein D, Brown PO, Pollack JR. 2005. A DNA microarray survey of gene expression in normal human tissues. *Genome Biol* **6:** R22. doi: 10.1186/gb-2005-6-3-r22.

Sikorska M, Sandhu JK, Deb-Rinker P, Jezierski A, Leblanc J, Charlebois C, Ribecco-Lutkiewicz M, Bani-Yaghoub M, Walker PR. 2008. Epigenetic modifications of SOX2 enhancers, SRR1 and SRR2, correlate with in vitro neural differentiation. *J Neurosci Res* **86:** 1680–1693.

Simons C, Pheasant M, Makunin IV, Mattick JS. 2006. Transposon-free regions in mammalian genomes. *Genome Res* **16:** 164–172.

Sisodiya SM, Ragge NK, Cavalleri GL, Hever A, Lorenz B, Schneider A, Williamson KA, Stevens JM, Free SL, Thompson PJ, et al. 2006. Role of *SOX2* mutations in human hippocampal malformations and epilepsy. *Epilepsia* **47:** 534–542.

Sleutels F, Zwart R, Barlow DP. 2002. The noncoding *Air* RNA is required for silencing autosomal imprinted genes. *Nature* **415:** 810–813.

Stephen S, Pheasant M, Makunin IV, Mattick JS. 2008. Large-scale appearance of ultraconserved elements in tetrapod genomes and slowdown of the molecular clock. *Mol Biol Evol* **25:** 402–408.

Strausberg RL, Feingold EA, Grouse LH, Derge JG, Klausner RD, Collins FS, Wagner L, Shenmen CM, Schuler GD, Altschul SF, et al. 2002. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. *Proc Natl Acad Sci* **99:** 16899–16903.

Sunwoo H, Dinger ME, Wilusz JE, Amaral PP, Mattick JS, Spector DL. 2008. *MEN ε/β* nuclear retained noncoding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. *Genome Res* **19:** 347–359.

Takahashi K, Yamanaka S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126:** 663–676.

Taranova OV, Magness ST, Fagan BM, Wu Y, Surzenko N, Hutton SR, Pevny LH. 2006. SOX2 is a dose-dependent regulator of retinal neural progenitor competence. *Genes & Dev* **20:** 1187–1202.

Tay Y, Zhang J, Thomson AM, Lim B, Rigoutsos I. 2008. MicroRNAs to *Nanog, Oct4* and *Sox2* coding regions modulate embryonic stem cell differentiation. *Nature* **455:** 1124–1128.

Thierry-Mieg D, Thierry-Mieg J. 2006. AceView: A comprehensive cDNA-supported gene and transcripts annotation. *Genome Biol* (Suppl.1) **7:** S12. doi: 10.1186/gb-2006-7-s1-s12.

Torarinsson E, Sawera M, Havgaard JH, Fredholm M, Gorodkin J. 2006. Thousands of corresponding human and mouse genomic regions unalignable in primary sequence contain common RNA structure. *Genome Res* **16:** 885–889.

Uchikawa M, Ishida Y, Takemoto T, Kamachi Y, Kondoh H. 2003. Functional analysis of chicken Sox2 enhancers highlights an array of diverse regulatory elements that are conserved in mammals. *Dev Cell* **4:** 509–519.

Uchikawa M, Takemoto T, Kamachi Y, Kondoh H. 2004. Efficient identification of regulatory sequences in the chicken genome by a powerful combination of embryo electroporation and genome comparison. *Mech Dev* **121:** 1145–1158.

Vavouri T, Walter K, Gilks WR, Lehner B, Elgar G. 2007. Parallel evolution of conserved noncoding elements that target a common set of developmental regulatory genes from worms to humans. *Genome Biol* **8:** R15. doi: 10.1186/gb-2007-8-2-r15.

Wang X, Arai S, Song X, Reichart D, Du K, Pascual G, Tempst P, Rosenfeld MG, Glass CK, Kurokawa R. 2008. Induced ncRNAs allosterically modify RNA-binding proteins in *cis* to inhibit transcription. *Nature* **454:** 126–130.

Washietl S, Hofacker IL, Lukasser M, Huttenhofer A, Stadler PF. 2005. Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome. *Nat Biotechnol* **23:** 1383–1390.

Wegner M, Stolt CC. 2005. From stem cells to neurons and glia: A Soxist's view of neural development. *Trends Neurosci* **28:** 583–588.

Westerfield M. 1994. The zebrafish book. In University of Oregon, Eugene, OR.

Wilhelm D, Hiramatsu R, Mizusaki H, Widjaja L, Combes AN, Kanai Y, Koopman P. 2007. SOX9 regulates prostaglandin D synthase gene transcription in vivo to ensure testis development. *J Biol Chem* **282:** 10553–10560.

Wilkins SJ, Yoong S, Verkade H, Mizoguchi T, Plowman SJ, Hancock JF, Kikuchi Y, Heath JK, Perkins AC. 2008. Mtx2 directs zebrafish morphogenetic movements during epiboly by regulating microfilament formation. *Dev Biol* **314:** 12–22.

Willingham AT, Gingeras TR. 2006. TUF love for 'junk' DNA. *Cell* **125:** 1215–1220.

Woolfe A, Goodson M, Goode DK, Snell P, McEwen GK, Vavouri T, Smith SF, North P, Callaway H, Kelly K, et al. 2004. Highly conserved noncoding sequences are associated with vertebrate development. *PLoS Biol* **3:** e7. doi: 10.1371/journal.pbio.0030007.

Woolfe A, Goode DK, Cooke J, Callaway H, Smith S, Snell P, McEwen GK, Elgar G. 2007. CONDOR: A database resource of developmentally associated conserved noncoding elements. *BMC Dev Biol* **7:** 100. doi: 10.1186/1471-213X-7-100.

Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, Ophir R, Bar-Even A, Horn-Saban S, Safran M, Domany E, et al. 2005. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* **21:** 650–659.

Yu W, Gius D, Onyango P, Muldoon-Jacobs K, Karp J, Feinberg AP, Cui H. 2008. Epigenetic silencing of tumor suppressor gene *p15* by its antisense RNA. *Nature* **451:** 202–206.

Zappone MV, Galli R, Catena R, Meani N, De Biasi S, Mattei E, Tiveron C, Vescovi AL, Lovell-Badge R, Ottolenghi S, et al. 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.

Zhang X, Lian Z, Padden C, Gerstein MB, Rozowsky J, Snyder M, Gingeras TR, Kapranov P, Weissman SM, Newburger PE. 2009. A myelopoiesis-associated regulatory intergenic noncoding RNA transcript within the human HOXA cluster. *Blood* **113:** 2526–2534.

Zhao C, Deng W, Gage FH. 2008. Mechanisms and functional implications of adult neurogenesis. *Cell* **132:** 645–660.

Complex architecture and regulated expression of the *Sox2ot* locus during vertebrate development

Paulo P. Amaral, Christine Neyt, Simon J. Wilkins, et al.

RNA 2009 15: 2013-2027 originally published online September 18, 2009
Access the most recent version at doi:10.1261/rna.1705309