PhD Program in
Translational and Molecular Medicine
DIMET

XXVII Cycle, Academic Year 2013-2014

University of Milano – Bicocca
School of Medicine and Faculty of Science

Functional characterization of regulatory sequences targeted by the transcription factor SOX2, identified by studies of long-range chromatin interactions in brain-derived neural stem/precursor cells

Coordinator: Prof. Andrea Biondi
Tutor: Prof. Silvia K. Nicolis

Dr. Jessica Armida Bertolini
Matr. No. 077100

TABLE OF CONTENTS

CHAPTER 1:

GENERAL INTRODUCTION

1. SOX FAMILY GENES

The *Sox* family genes encode for transcriptional factors relevant during embryonic development and cell fate determination. The common characteristic of these factors is the presence of a binding domain, called HMG-box (High-Mobility Group box), in their protein sequence. This domain, composed by 79 amino acids, recognizes the DNA in a sequence-specific way, binding the DNA minor groove. The consensus sequence is: 5'-(A/T)(A/T)CAA(A/T)G-3' [Wegner, 1999].

The *Sox* genes have been identified by their homology with the HMG-box of *Sry* ("Sex determining region Y"), the gene important for the male determining sex in mammalian and located on the chromosome Y. In fact, *Sox* means *Sry*-related HMG-box.

The *Sox* genes have been identified in mammalians, reptiles, amphibians, fishes, insects and nematodes [Bowles *et al*., 2000]. In vertebrates, the SOX proteins are divided in ten groups, called A-J, and they have a crucial role in the development of nervous system, eye, cartilage, blood vessels, sex determining and development of testis and heart [Bowles *et al*., 2000]. The tissue-specific protein-protein interactions with other transcription factors, the spatial-temporal expression pattern and the specific sequence of the HMG-box domain allow SOX factors to be specific for their targets (Zhong *et al.*, 2011).

All SOX factors are able to recognize and bind to the same 7-bases consensus sequence but every single factor is able to regulate selectively the expression of specific target genes. Some SOX factors are expressed in more cell types and are able to regulate the expression of different genes, activating or repressing them. Moreover, a specific cell type can express more than one SOX factor, each one with its specific target gene group.

There are also a lot of number of partner factors. For the SOX2 transcription model, SOX2 binds the DNA in a weak way and the presence of a specific partner factor (that recognizes a nearby consensus sequence) can stabilize the binding and activate the transcription of the target gene. The distribution of partner factors is cell type specific and the choice of the partner depends on their availability in the different tissues (Kamachi *et al.*, 2000).

The B group of SOX factors play a crucial role in neurogenesis, morphogenesis and gonadogenesis. They are subdivided in two other subgroups (B1 and B2), based on the differences of the protein sequence and their functional role. SoxB1 proteins act as transcriptional activators, instead SoxB2 act as transcriptional repressors (Zhong *et al.*, 2011).

The genes included in the SoxB1 group are *Sox1*, *Sox2* and *Sox3*. They present a high level of similarity. *Sox3* in expressed during early embryonic developmental stages and in the central nervous system. Also *Sox1* is involved in the neurogenesis and in the development of the crystalline (Uchikawa *et al.*, 1999). Looking at the expression of these genes during the embryonic development, we observe an overlap of their expression patterns, suggesting a functional interaction of these genes during the organogenesis and a regulation of the target genes expression (Uchikawa *et al.*, 1999). SoxB1 genes have redundant functions and the loss of one of them can often be complemented by the expression of another gene of the same group (Graham *et al.*, 2003). After neurogenesis, *Sox1*, *Sox2* and *Sox3* are co-expressed in neural precursor cells in active proliferation along the anterior-posterior axis of the developing embryo. These genes remain active in neural progenitors and stem cells in the neurogenic regions of the adult central nervous system, suggesting a their possible role in the maintenance of the neural precursors and neural stem cells identity and in the inhibition of neural differentiation (Zappone et al., 2000; Ferri et al., 2004; Graham *et al.*, 2003). On the other hand, the SoxB2 genes seem to promote the exit from the cell-cycle and to induce the neural differentiation (Jager *et al.*, 2011).

2. THE SOX2 TRANSCRIPTION FACTOR AND ITS IMPLICATION IN BRAIN DEVELOPMENT

The SOX2 transcription factor belongs to SoxB1 group of *Sox* family genes. It is transiently expressed in the inner cell mass and the epiblast of the blastocyst and, later, throughout the developing neuroephitelium.

This gene is localized on the chromosome 3, both in mouse and human. It is highly conserved and it is composed by a single exon, encoding for 2,4 kilobases transcript. The protein is composed by three regions: an hydrophobic region on the N-terminal portion; a central part with the HMG-box domain; a trans-activation domain on the C-terminal portion.

During the mouse embryonic development, *Sox2* expression is still traceable in the oocyte stage and in the morula at the second day of embryonic development

(E2,5). Its expression remains in the blastocyst inner cell mass till E3,5 in the epiblast till E6 and extra-embryonic ectoderm till E6,5. Later, the expression is reduced to the anterior part of the ectoderm (neuroectoderm), to the future neural plate (E7-7,5) and neural tube (E8,5). Later, the *Sox2* expression becomes pan-neural (Fig. 1). It seems to be present also in the brachial arches and in germ cells (Avilion *et al.*, 2003).

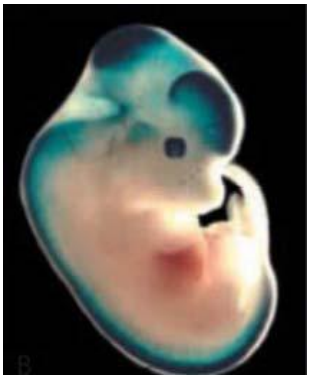


Figure 1. *Sox2*$^{\beta\text{-}geo}$ expression (X-gal staining) in a E12,5 mouse embryo; *Sox2* has a pan-neural expression [Ferri *et al.*, 2004].

During later stages of embryonic development, *Sox2* expression remains at high level in the ventricular zone, in active proliferation, while it decreases in the marginal zone where the differentiation begins (Ferri *et al.*, 2004).

In the adult brain, *Sox2* remains expressed in sporadic cells in the differentiated cerebral regions, like cortex, thalamus, striatum. Instead, *Sox2* remains highly expressed in ependyma and in neurogenic regions: the lateral ventricle (where the rostral migratory stream starts and through which the neural precursors reach the olfactory bulbs) and the hippocampal dentate gyrus (Ferri *et al.*, 2004).

Sox2 expression is crucial during the early stages of embryonic development. Homozygous *Sox2*-KO (knock-out) mice die for the loss of the stem cells of the blastocyst inner cell mass (ICM) (Avilion *et al.*, 2003; Penvy *et al.*, 1998).

The Sox2$^{-/-}$ ICM stem cells stop their proliferation and some of them start an inappropriate differentiation, expressing trophoblast markers. The loss of stem cells cause the early lethality of mutants.

Hence, to study SOX2 later functions in neural development, our laboratory generated, through gene targeting, a "*Sox2*^{flox}" mutation, in which the *Sox2* gene is flanked by lox sites; these are the substrates for Cre-recombinase, which, expressed by suitable transgenes, allows the spatially and temporally controlled ablation of *Sox2.* We compared the defects in mice obtained after *Sox2*-ablation operated by two different Cre-transgenes: *Nestin-Cre* and *Bf1-Cre*.

<u>*Nestin-Cre* transgene:</u>

Cre activity driven by the *Nestin-Cre* transgene starts at embryonic day 10,5 (E10,5) and induces the loss of *Sox2* in all the central nervous system (CNS) till E12,5. The *Sox2*-deleted mutant mice are born but most of them died by 4 weeks of age. At birth (postnatal day 0, P0), the brain defects in mutant mice were quite limited. Instead, subsequently (at P7), the development of hippocampus was compromised: its size was markedly reduced, in comparison with wild-type (wt), in particular in the caudal zone, resulting in an underdeveloped dentate gyrus. The defects in hippocampus was also observed in the adult *Sox2*-mutant brain (Favaro *et al.*, 2009).

Moreover, the defects of these *Sox2*-mutant brains mimic the effect observed after the *Shh* loss. SHH is a cytokine important for various aspects of CNS development, including hippocampus development Analyzing *Shh* mRNA expression in the *Sox2*-deleted brains, we noticed that it was strongly reduced at E14,5 in telencephalon and diencephalon of *Sox2*-mutant mice, but not in midbrain and spinal cord. At birth, *Shh* mRNA was absent in the hippocampal hilus of *Sox2*-mutant mice, where it was clearly detectable in wt mice (Favaro *et al.*, 2009).

We tried to rescue the pathological phenotype by the administration of a SHH-pharmacological agonist (SHH-Ag) to pregnant mice, starting at E12,5. The defects in hippocampus development of *Sox2*-deleted mice were partially rescued, confirming that a stimulation of the SHH pathway was able to reduce the defects due to *Sox2* loss (Favaro *et al.*, 2009).

In vitro, the deletion of *Sox2* caused loss of self-renewal of neurosphere cultures, obtained from the dissection of *Sox2*-deleted mouse brains. Moreover, analyzing the *Shh* mRNA expression in these cells, we noticed that, also in this case, its expression was completely lost (Favaro *et al.*, 2009).

Interestingly, by infecting the *Sox2*-deleted neuroshpere cultures with a lentivirus encoding *Sox2*, we observed a partial rescue of the defects previously described. The infected neurospheres showed rescue in self-renewal and in *Shh* mRNA expression (Favaro *et al.*, 2009).

Bf1-Cre transgene (this work is the object of Chapter 3 of this thesis):

As shown before, the defects observed in *Nestin-Cre* mutant mice are quite limited. Thus, in our laboratory we used a different transgene able to induce *Sox2* loss at earlier stages. The Cre activity, driven by this transgene, starts at E9,5 and it is specific for the developing telencephalon. Following *Sox2* ablation, the mutant mice die just after birth, suggesting the presence of more severe defects in development than the ones observed using the *Nestin-Cre* transgene.

At E12,5, the telencephalic vesicles were reduced and the eyes were abnormal. Interestingly, although the whole telencephalon was affected, the ventral part was much more severely compromised than the dorsal one, with major tissue loss. At E18,5, mutant embryos showed important brain defects: the head and the telencephalon were smaller than in wild-type, while the midbrain was almost unaffected; moreover, the olfactory bulbs and the midline ventral structure were absent; the eyes were abnormal and extremely reduced in size; finally, the hippocampus was severely underdeveloped [Ferri *et al.*, 2013].

This phenotype resembles that of mutants in the gene encoding the Sonic hedgehog (SHH) cytokine, as well as that of mutants in the gene encoding transcription factor NKX2.1, a known transcriptional activator of *Shh*; interestingly, *Nkx2.1* expression itself is also stimulated by SHH signaling by a feedback mechanism. *In situ* hybridization studies showed that *Shh* expression is severely reduced in *Sox2*-mutants in the midline region at E12,5 and already severely down-regulated in the medial ventral telencephalon at E11,5. Also *Nkx2.1* expression is severely down-regulated in *Sox2*-mutants at E11,5, and this reduction in mRNA expression is already detectable also at earlier stages (E9,5 and E10,5), preceding phenotypic abnormalities [Ferri *et al.*, 2013].

As already seen for the *Nestin-Cre* transgene, by administering a SHH-agonist (SHH-ag) to the pregnant mice at E8,5 (just before the *Sox2* ablation) and E10,5, the mutant mice analyzed showed a partial rescue of the defects: the expression of some ventral determinants remained at E14,5, while it was lost in untreated mutants; the morphology of the ventral brain is almost recovered. Again, it means

that *Shh* is crucial and the lack of its signaling is an important cause of the defects observed in *Sox2*-mutant mice [Ferri *et al.*, 2013].

Moreover, in this work we proved that *Nkx2.1* gene, the other ventral determinant down-regulated in *Sox2*-mutant mice, is a direct SOX2 target gene and that its regulation is mediated by SOX2 in a dose-dependent way. It is known that NKX2.1 is an activator of *Shh*, so we could suppose that, in *Sox2*-mutant mice, the lack of SOX2 causes the defective activation of *Shh*, through the loss of its activator NKX2.1 [Ferri *et al.*, 2013] (see Chapter 3).

3. PROTEIN COMPLEXES INVOLVING SOX2 AND ITS CO-FACTORS

It is known that often SOX2 regulates its target genes through the interaction with other transcription factors that act as co-factors, forming a protein complex. One SOX2 target gene regulated by the SOX2 protein complex is *δ-crystalline*. In this case, SOX2 binds the co-factor PAX6, interacting on an enhancer element of the target gene. This protein-complex formation is crucial for the beginning of the crystalline development [Kamachi *et al.*, 2001].

There are many other examples of protein interaction between SOX2 and POU proteins (other transcription factors important during development), such as the SOX2-OCT4 protein complex that binds the *Fgf4* (fibroblast growth factor 4) gene [Ambrosetti *et al.*, 1997]. These same factors are also involved in the regulation of the *Utf1* (undifferentiated transcription factor 1) gene expression [Nishimoto et al., 1999]. It is also known that *Sox2* and *Oct4* are both expressed in pluripotent cells of the blastocyst inner cell mass, where they have crucial roles [Avilion *et al.*, 2003]. Their combined activity in embryonic stem cells (ESCs) is the basis for their crucial role in the maintenance of pluripotent state of ICM stem cells [Kamachi *et al.*, 2000; Boyer *et al.*, 2005]. A further mechanism is that the SOX2-OCT4 protein complex is also implicated in the regulation of *Nanog,* another gene essential for pluripotency [Rodda *et al.*, 2005]

Instead, the protein complex SOX2-BRN2 is active in more differentiated cells among neural lineage (neural stem/progenitor cells) [Lodato *et al.*, 2013].

4. SOX2 AND HUMAN DISEASES

Heterozygous *Sox2* mutations in humans cause neurological defects: in particular, mutations (including missense, frameshift and nonsense mutations) identified in the *Sox2* locus cause defects in the development of eyes (anophthalmia, microphthalmia) [Fantes *et al.*, 2003; Schneider *et al.*, 2009] and defects in hippocampus, with neurological pathology including epilepsy, motor control problems and learning disabilities [Ragge *et al.*, 2005; Sisodiya *et al.*, 2006; Kelberman *et al.*, 2006]. Other pathological characteristics of patients with heterozygous *Sox2* mutations are mild facial dysmorphism, developmental delay, esophageal atresia [Kelberman *et al.*, 2006], psychomotor retardation and hypothalamo-pituitary disorders [Tziaferi *et al.*, 2008].

5. LONG-RANGE INTERACTIONS IN CHROMATIN

5.1. LONG-RANGE INTERACTIONS AND CHIA-PET TECHNIQUE

Recently, it was found that transcriptional regulatory elements of genes are not always localized in the proximity of the gene they control, but often they lie very far from it on the linear chromosome map. It means that the gene regulatory networks are organized by spatially connectivity between distal regulatory elements (DREs) and their corresponding promoters [Zhang *et al.*, 2013; Li *et al.*, 2012; Fullwood *et al.*, 2009; Cheutin and Cavalli, 2014].

It has been developed a new approach for the genome-wide mapping of long-range interactions: the Chromatin Interaction Analysis by Paired-End Tag sequencing (ChIA-PET). This technique is performed by cross-linking of the chromatin to block the DNA fragments that are brought together by long-range interactions, followed by chromatin immunoprecipitation with specific antibodies (in Zhang *et al.* [2013], the antibody was against the hypophosphorylated form of RNA polymerase II, present in the pre-initiation complexes), ligation of “junction fragments” and high-throughput sequencing of the interacting regions (Fig. 2) [Zhang *et al.*, 2013; Fullwood *et al.*, 2009].

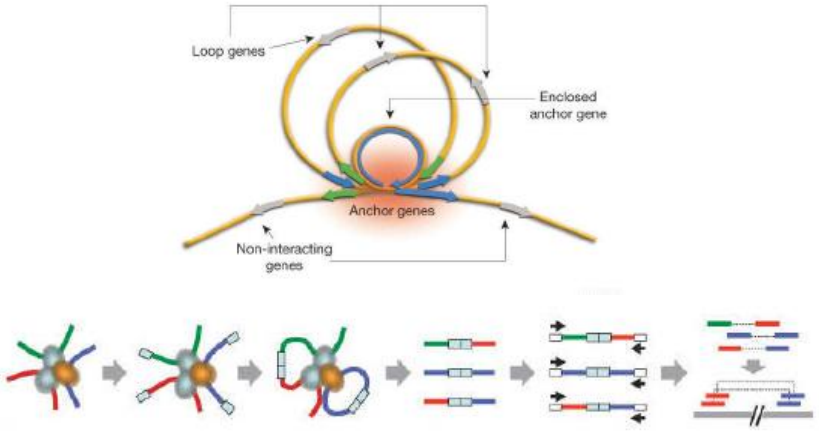


Figure 2. Genome-wide detection of long-range DNA interactions in chromatin (ChIA-PET) [Fullwood *et al.*, 2009]

Zhang *et al.* (2013) performed this technique on different type of cells: embryonic stem cells (ESCs), neural stem cells (NSCs) and neurosphere stem/progenitors cells (NPCs). NPCs are neural progenitor cells derived *ex vivo* from mice forebrain telencephalic region [Zappone *et al.*, 2000]. Using the ChIA-PET analysis, they found the majority of the interactions surrounding promoter regions, with three possible conformations: two interacting promoters, promoters connecting to intergenic regions or to intragenic regions. Thus, this connections showed a large numbers of putative enhancers located in these inter- and intragenic regions. In many of them it was been possible identified also other enhancer characteristics, such as an enrichment in the presence of monomethylated histone H3 lysine 4 (H3K4me1), sequence conservation and presence of binding sites for co-activator p300 transcription factor. Interestingly, these data suggest that a consistent proportion of the identified putative enhancers do not rule their nearest gene, as previously assumed, but they are connected by long-range interactions to gene also very far from them [Zhang *et al.*, 2013].

Moreover, among all the putative enhancers indentified, a portion were defined "poised enhancers" [Zhang *et al.*, 2013]. In ESCs, a poised enhancer is proposed to prime the associated gene for a subsequent transcription, such as a cell-type specific transcription during development [Rada-Iglesias *et al.*, 2011]. In their work, Zhang *et al.* (2013) found that a high number of poised enhancers were associated to genes with "bivalent promoters", consisting in large regions of H3 lysine 27 tri-methylation (H3K27me3) harboring smaller regions of H3 lysine 4

mono-methylation (H3K4me1) [Bernstein *et al.*, 2006; Zhang *et al.*, 2013]. The H3K27me3 represses transcription by promoting compact chromatin structure, while the H3K4me1 regulates positively the transcription by the recruitment of nucleosome remodelers and histone acetylases that open the chromatin structure [Bernstein *et al.*, 2006]. In ESCs, bivalent promoters are often localized upstream of genes that encode transcription factors necessary for development [Bernstein *et al.*, 2006].

Instead, genes with enhancer-promoter interactions in single-gene complexes were more likely to be tissue-specific or developmentally regulated [Li *et al.*, 2012].

5.2. DISEASE-ASSOCIATED NON-CODING ELEMENTS

Given the importance of distal regulatory elements in transcriptional regulation, one may expect that mutations in these elements can cause pathology, due to the deregulation of the associated genes. Indeed for example, a single nucleotide mutation, found in the regulatory sequence located 460 kilobases (kb) upstream of the *Shh* gene, was discovered in an individual with holoprosencephaly; the mutation reduced the activity of the distant enhancer in transgenic assays [Jeong *et al.*, 2008].

A further example involved another one *Shh* enhancer, located 1 megabases away from the *Shh* gene and embedded in the intronic region of *LMBR1*; a point mutation in this enhancer site causes preaxial polydactyly, a common congenital limb malformation in mammals [Lettice *et al.*, 2002].

Another example involves the *PAX6* gene: it is known that the correct expression of *PAX6* is dependent on regulatory elements inside the last intron of the neighboring gene *ELP4*. Breakpoints within *ELP4*, which leave the coding sequence of *PAX6* intact, have also been shown to cause aniridia. The phenotype is not caused by loss of *ELP4* function but rather by loss of *PAX6* expression, thus suggesting that essential regulatory elements driving *PAX6* reside inside *ELP4* [Navratilova *et al.*, 2009; Kleinjan *et al.*, 2001].

These are just few examples of the importance of distal regulatory elements to rule the associated genes *via* long-range interactions.

5.3. *Sox2* IS INVOLVED IN LONG-RANGE INTERACTIONS

It is possible to use the ChIA-PET approach to evaluate if chromatin organization is able to reflect the cell-specific transcription regulatory circuitry. In particular, Zhang *et al.* (2013) analyzed the spatial connectivity of reprogramming genes, such as *Pou5f1*, *Nanog*, *Lin28a*, *Klf4*, *Myc* and *Sox2* [Yu *et al.*, 2007], in embryonic stem cells (ESCs), through RNApolII-mediated interaction maps. These genes are known to govern pluripotency in ESCs though coordinated autoregulatory loops [Jaenisch and Young, 2008]. The authors found that three *Klf* genes (*Klf1*, *Klf2* and *Klf4*) were directly connected to *Sox2*. Moreover, extending the analysis from one to two connectivity hubs, all of the reprogramming genes were found to be connected within one major hub, except for *Myc* and *Lin28a*. It means that all of these genes could co-localize in the nucleus within the same "transcription factory". Among them, *Nanog* and *Pou5f1* have limited connections whereas *Sox2* has extensive connectivity [Zhang *et al.*, 2013]. Moreover, in ESCs, they found that the *Sox2* promoter is connected to clusters of ESC-specific enhancers to other pluripotency related genes; instead, in neural stem cells (NSCs) it could be observed a very different *Sox2* connectivity profile and different enhancers mediated cell-specific connectivity [Zhang *et al.*, 2013].

6. GENOME-WIDE ANALYSIS OF REGULATORY ELEMENTS

Evolutionary constraint of non-coding sequences can predict the location of enhancers in the genome [Woolfe *et al.*, 2005], but does not reveal when and where these enhancers are active *in vivo*. The acetyltransferase and transcriptional coactivator p300 is a near-ubiquitously expressed component of enhancer-associated protein assemblies and it is critically required for embryonic development [Merika *et al.*, 1998].

Visel *et al.* (2009) determined the genome-wide occupancy of p300 in forebrain, midbrain and limb tissue isolated directly from developing mouse embryos at embryonic day 11,5 (E11,5). Using a transgenic mouse reporter assay, they show that p300 binding in these embryonic tissues predicts, with high accuracy, not only where enhancers are located in the genome, but also in what tissues they are active *in vivo*. They cloned the human genomic sequences orthologous to these enhancer candidate regions into an enhancer reporter vector and generated transgenic mice. A high number of the orthologous human sequences, tested in mouse by transgenesis, gave rise to an activation of the reporter gene, proving

the high evolutionary conservation of these regulatory sequences [Visel *et al.*, 2009; Visel *et al.*, 2013].

These data provide a primary resource for investigating gene regulatory mechanisms of telencephalon development and enable studies of the role of distant-acting enhancers in neurodevelopmental disorders [Visel *et al.*, 2013].

Many of these p300-binding detected sequences, and some of the ones validated by transgenesis in mouse, are located within the putative regulatory sequences studied in our experiments (see Chapter 2).

7. *IN VIVO* EXPERIMENTS IN ZEBRAFISH

7.1. TRANSGENESIS TECHNIQUE

A general strategy for testing whether non-coding regulatory sequences are functionally relevant involves testing their ability to up-regulate reporter gene expression *in vivo*. "Enhancer assays" using mouse transgenic are slow and laborious. An alternative approach highly used is transgenesis in zebrafish (*Danio rerio*) embryos [Woolfe *et al.*, 2005; Bessa *et al.*, 2009]. This technique is particularly suitable for the availability of large numbers of fertilized eggs, ease of micro-injection and transparency of the developing embryos. Thus, hundreds of individuals may be screened at the same time, increasing the throughput of this functional assay [Woolfe *et al.*, 2005].

The commonly used transgenesis vectors consist of a shuttle vector, a minimal promoter and an *in vivo* reporter gene. These characteristics face some practical problems: first, the random integration of the shuttle vector in the genome often causes it to be exposed to the enhancer activity present in the surrounding genomic regions, resulting in reporter gene expression that does not result from the DNA sequenced cloned in the vector; second, the lack of a positive control of transgenesis makes it difficult to determine the efficiency of the integration events [Bessa *et al.*, 2009].

Bessa *et al.* (2009) described a novel vector to test the enhancer activity of putative regulatory elements in zebrafish. This Zebrafish Enhancer Detector (ZED) vector is based on the Tol2 transposon and it presents a series of improvements:

- a *gata2a* minimal promoter [Ellingsen *et al.*, 2005], selected from several promoters, to drive the enhanced green fluorescent protein (EGFP) expression;

- the presence of a Gateway entry site, at the 5' of *gata2a* minimal promoter, to facilitate the insertion of the sequence in study;

- the presence of insulator sequences around the enhancer reporter cassette;

- the presence of a further reporter gene, encoding for a red fluorescent protein (DsRed), guided by a cardiac actin promoter, used as control for transgenesis efficiency.

We used this ZED vector for our *in vivo* experiments in zebrafish (see Chapter 2).

7.2. THE INVOLVMENT OF SOX2 IN ZEBRAFISH EMBRYONIC DEVELOPMENT

In zebrafish, B1 *Sox* family gene comprises *sox1a/1b/2/3/19a/19b*. In zebrafish, *sox3/19a/19b* are expressed in the blastula [Okuda *et al.*, 2006], whereas the corresponding early expression in mice is covered by *Sox2* [Avilion *et al.*, 2003]. Following this stage, the B1 *sox* genes are important for specification of the embryonic ectoderm into the neuroectoderm lineage. Among the B1 *sox* genes of zebrafish, *sox2/3/19a/19b* are expressed at high levels during early development with extensive regional overlaps [Okuda *et al.*, 2006; Okuda *et al.*, 2010]. *sox19b* mRNA is maternally supplied. *sox3* and *sox19a* are activated around the 1000-cell stage and *sox2* around the 30% epiboly stage [Okuda *et al.*, 2006; Okuda *et al.*, 2010] (Fig. 3). The expression of *sox3/19a/19b* initially covers the entire blastoderm, but gradually disappears at the embryonic margin after 30% epiboly stage. At the shield stage, the expression of *sox2/3/19a/19b* covers the future ectoderm, but then becomes confined to the presumptive neuroectoderm [Okuda *et al.*, 2006; Okuda *et al.*, 2010]. Expression of *sox1a/1b* is initiated only during late gastrulation stages. Later, the expression of the B1 *sox* genes continues in neural precursors, where they are able to maintain the neural progenitor states [Graham *et al.*, 2003; Tanaka *et al.*, 2004; Okuda *et al.*, 2010].

The similarities in the characteristics of the B1 SOX proteins as transcriptional regulators [Okuda *et al.*, 2006; Tanaka *et al.*, 2004] suggest redundant functions in tissues where they are co-expressed [Graham *et al.*, 2003]. In fact, single *Sox1*, or *Sox3*, knock-out mice present only mild abnormalities in the central nervous system (CNS), presumably due to this redundancy effect [Nishiguchi *et al.*, 1998; Rizzoti et al., 2004; Weiss *et al.*, 2003], whereas *Sox2* knock-out mouse embryos die around implantation [Avilion *et al.*, 2003]. Consistently, a single *sox2* or *sox3* knock-down (KD) in zebrafish causes only mild developmental abnormalities [Dee *et al.*, 2008; Kamachi *et al.*, 2008; Okuda *et al.*, 2010].

Morpholino antisense oligonucleotides (MO) have been widely used to inhibit gene function in zebrafish embryos and are usually used as sequence-specific translation-blocking agents [Kamachi *et al.*, 2008].

Kamachi et al. (2008) designed a MO specifically directed against the *Sox2* mRNA, through the binding on the complementary sequence in the *Sox2* 5'-UTR (untranslated region). They demonstrated that this MO was the most effective to down-regulate *Sox2* levels, compared to other two tested MOs. In fact, they measured that it was possible reach the 69% and the 85% inhibition by injection of 0,9 and 1,8 ng of this molecule, respectively [Kamachi *et al.*, 2008]. We used this same MO in our *in vivo* experiments in zebrafish (see Chapter 2).

To investigate the function of B1 *sox* in early stage embryos, Okuda *et al.* (2010) knocked down *sox2/3/19a/19b* either individually or in combination, using morpholino antisense oligonucleotides (MO). No gross abnormalities were observed in the embryo morphology when any one of *sox2/3/19a/19b* was knocked-down, although the development of the CNS may be slightly perturbed and 75% of the *sox2* morphants showed an upturned tail phenotype. When any three of *sox2/3/19a/19b* were simultaneously knocked-down (triple KD), a range of morphological abnormalities was observed depending on the combination of KD targets. Triple KDs of *sox2/19a/19b* and *sox2/3/19b* caused only mild morphological defects, presumably because the remaining *sox3* and *sox19a* genes, respectively, mostly cover the B1 *sox* expression domains. *sox3/19a/19b* morphants often showed stronger yet variable defects in their posterior structures, presumably reflecting the weak *sox2* expression in the posterior neuroectoderm. *sox2/3/19a* morphants appeared normal during gastrulation, but later developed morphological abnormalities, likely because *sox19b* expression decreases in later stages. In contrast to the triple KDs, the quadruple knockdown of *sox2/3/19a/19b* resulted in very severe developmental abnormalities, suggesting essential functions of B1 *sox* in early embryogenesis [Okuda *et al.*, 2010].

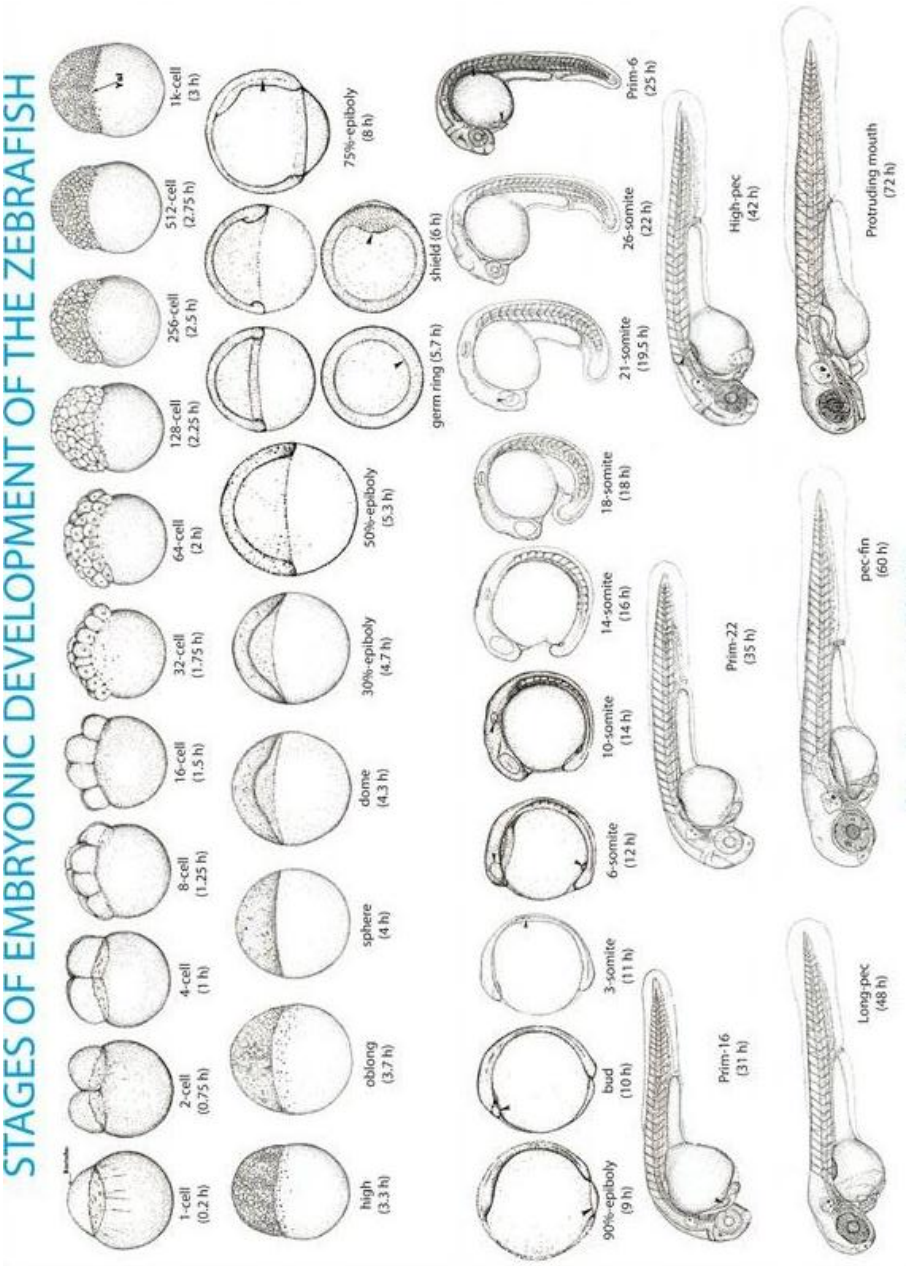


Figure 3. Stages of zebrafish embryonic development.

8. REFERENCES

- Ambrosetti DC, Basilico C, Dailey L. 1997. Synergistic activation of the fibroblast growth factor 4 enhancer by Sox2 and Oct-3 depends on protein-protein interactions facilitated by a specific spatial arrangement of factor binding sites. *Mol Cell Biol.* **17**: 6321-9.

- 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**: 126-40.

- Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, Jaenisch R, Wagschal A, Feil R, Schreiber SL, Lander ES. 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**: 315-26.

- Bessa J, Tena JJ, de la Calle-Mustienes E, Fernández-Miñán A, Naranjo S, Fernández A, Montoliu L, Akalin A, Lenhard B, Casares F, Gómez-Skarmeta JL. 2009. Zebrafish enhancer detection (ZED) vector: a new tool to facilitate transgenesis and the functional analysis of cis-regulatory regions in zebrafish. *Dev Dyn.* **238**: 2409-17.

- Bowles J, Schepers G, Koopman P. 2000. Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators. *Dev Biol.* **227**: 239-55.

- Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, Gifford DK, Melton DA, Jaenisch R, Young RA. 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**: 947-56.

- Cheutin T, Cavalli G. 2014. Polycomb silencing: from linear chromatin domains to 3D chromosome folding. *Curr Opin Genet Dev.* **25**: 30-7.

- Dee CT, Hirst CS, Shih YH, Tripathi VB, Patient RK, Scotting PJ. 2008. Sox3 regulates both neural fate and differentiation in the zebrafish ectoderm. *Dev Biol.* **320**: 289-301.

- Ellingsen S, Laplante MA, König M, Kikuta H, Furmanek T, Hoivik EA, Becker TS. 2005. Large-scale enhancer detection in the zebrafish genome. *Development* **132**: 3799-811.

- Fantes J, Ragge NK, Lynch SA, McGill NI, Collin JR, Howard-Peebles PN, Hayward C, Vivian AJ, Williamson K, van Heyningen V, FitzPatrick DR. 2003. Mutations in SOX2 cause anophthalmia. *Nat Genet.* **33**: 461–3.

- Favaro R, Valotta M, Ferri AL, Latorre E, Mariani J, Giachino C, Lancini C, Tosetti V, Ottolenghi S, Taylor V, Nicolis SK. 2009. Hippocampal development and neural stem cell maintenance require Sox2-dependent regulation of Shh. *Nat Neurosci.* **12**: 1248-56.

- Ferri A, Favaro R, Beccari L, Bertolini J, Mercurio S, Nieto-Lopez F, Verzeroli C, La Regina F, De Pietri Tonelli D, Ottolenghi S, Bovolenta P, Nicolis SK. 2013. Sox2 is required for embryonic development of the ventral telencephalon through the activation of the ventral determinants Nkx2.1 and Shh. *Development* **140**: 1250-61.

- Ferri AL, Cavallaro M, Braida D, Di Cristofano A, Canta A, Vezzani A, Ottolenghi S, Pandolfi PP, Sala M, De Biasi S, Nicolis SK. 2004. Sox2 deficiency causes neurodegeneration and impaired neurogenesis in the adult mouse brain. *Development* **131**: 3805-19.

- Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, Mohamed YB, Orlov YL, Velkov S, Ho A, Mei PH, Chew EG, Huang PY, Welboren WJ, Han Y, Ooi HS, Ariyaratne PN, Vega VB, Luo Y, Tan PY, Choy PY, Wansa KD, Zhao B, Lim KS, Leow SC, Yow JS, Joseph R, Li H, Desai KV, Thomsen JS, Lee YK, Karuturi RK, Herve T, Bourque G, Stunnenberg HG, Ruan X, Cacheux-Rataboul V, Sung WK, Liu ET, Wei CL, Cheung E, Ruan Y. 2009. An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* **462**: 58-64.

- Graham V, Khudyakov J, Ellis P, Pevny L. 2003. SOX2 functions to maintain neural progenitor identity. *Neuron.* **39**: 749-65.

- Jaenisch R, Young R. 2008. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell* **132**: 567-82.

- Jager M, Quéinnec E, Le Guyader H, Manuel M. 2011. Multiple Sox genes are expressed in stem cells or in differentiating neuro-sensory cells in the hydrozoan Clytia hemisphaerica. *Evodevo* **2**: 12.

- Jeong Y, Leskow FC, El-Jaick K, Roessler E, Muenke M, Yocum A, Dubourg C, Li X, Geng X, Oliver G, Epstein DJ. 2008. Regulation of a remote Shh forebrain enhancer by the Six3 homeoprotein. *Nat Genet.* **40**: 1348-53.

- Kamachi Y, Okuda Y, Kondoh H. 2008. Quantitative assessment of the knockdown efficiency of morpholino antisense oligonucleotides in zebrafish embryos using a luciferase assay. *Genesis* **46**: 1-7.

- Kamachi Y, Uchikawa M, Tanouchi A, Sekido R, Kondoh H. 2001. Pax6 and SOX2 form a co-DNAbinding partner complex that regulates initiation of lens development. *Genes Dev.* **15**: 1272-86.

- Kamachi Y, Uchikawa M, Kondoh H. 2000. Pairing SOX off: with partners in the regulation of embryonic development. *Trends Genet.* **16**: 182-7.

- Kelberman D, Rizzoti K, Avilion A, Bitner-Glindzicz M, Cianfarani S, Collins J, Chong WK, Kirk JM, Achermann JC, Ross R, Carmignac D, Lovell-Badge R, Robinson IC, Dattani MT. 2006. Mutations within Sox2/SOX2 are associated with abnormalities in the hypothalamo-pituitary-gonadal axis in mice and humans. *J Clin Invest.* **116**: 2442–55.

- Kleinjan DA, Seawright A, Schedl A, Quinlan RA, Danes S, van Heyningen V. 2001. Aniridia-associated translocations, DNase hypersensitivity, sequence comparison and transgenic analysis redefine the functional domain of PAX6. *Hum Mol Genet.* **10**: 2049-59.

- Lettice LA, Horikoshi T, Heaney SJ, van Baren MJ, van der Linde HC, Breedveld GJ, Joosse M, Akarsu N, Oostra BA, Endo N, Shibata M, Suzuki M, Takahashi E, Shinka T, Nakahori Y, Ayusawa D, Nakabayashi K, Scherer SW, Heutink P, Hill RE, Noji S. 2002. Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. *Proc Natl Acad Sci U S A* **99**: 7548-53.

- Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, Poh HM, Goh Y, Lim J, Zhang J, Sim HS, Peh SQ, Mulawadi FH, Ong CT, Orlov YL, Hong S, Zhang Z, Landt S, Raha D, Euskirchen G, Wei CL, Ge W, Wang H, Davis C, Fisher-Aylor KI, Mortazavi A, Gerstein M, Gingeras T, Wold B, Sun Y, Fullwood MJ, Cheung E, Liu E, Sung WK, Snyder M, Ruan Y. 2012. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148**: 84-98.

- Lodato MA, Ng CW, Wamstad JA, Cheng AW, Thai KK, Fraenkel E, Jaenisch R, Boyer LA. 2013. SOX2 co-occupies distal enhancer elements with distinct POU factors in ESCs and NPCs to specify cell state. *PLoS Genet.* **9**: e1003288.

- Merika M, Williams AJ, Chen G, Collins T, Thanos D. 1998. Recruitment of CBP/p300 by the IFN beta enhanceosome is required for synergistic activation of transcription. *Mol Cell.* **1**: 277-87.

- Navratilova P, Fredman D, Hawkins TA, Turner K, Lenhard B, Becker TS. 2009. Systematic human/zebrafish comparative identification of cis-regulatory activity around vertebrate developmental transcription factor genes. *Dev Biol.* **327**: 526-40.

- Nishiguchi S, Wood H, Kondoh H, Lovell-Badge R, Episkopou V. 1998. Sox1 directly regulates the gamma-crystallin genes and is essential for lens development in mice. *Genes Dev.* **12**: 776-81.

- Nishimoto M, Fukushima A, Okuda A, Muramatsu M. 1999. The gene for the embryonic stem cell coactivator UTF1 carries a regulatory element which selectively interacts with a complex composed of Oct-3/4 and Sox-2. *Mol Cell Biol.* **19**: 5453-65.

- Okuda Y, Ogura E, Kondoh H, Kamachi Y. 2010. B1 SOX coordinate cell specification with patterning and morphogenesis in the early zebrafish embryo. *PLoS Genet.* **6**: e1000936.

- 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**: 811-25.

- Pevny LH, Sockanathan S, Placzek M, Lovell-Badge R. 1998. A role for SOX1 in neural determination. *Development* **125**: 1967-78.

- Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. 2011. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**: 279-83.

- Ragge NK, Lorenz B, Schneider A, Bushby K, de Sanctis L, de Sanctis U, Salt A, Collin JR, Vivian AJ, Free SL, Thompson P, Williamson KA, Sisodiya SM, van Heyningen V, Fitzpatrick DR. 2005. SOX2 anophthalmia syndrome. *Am J Med Genet A.* **135**: 1–7.

- Rizzoti K, Brunelli S, Carmignac D, Thomas PQ, Robinson IC, Lovell-Badge R. 2004. SOX3 is required during the formation of the hypothalamo-pituitary axis. *Nat Genet.* **36**: 247-55.

- Rodda DJ, Chew JL, Lim LH, Loh YH, Wang B, Ng HH, Robson P. 2005. Transcriptional regulation of nanog by OCT4 and SOX2. *J Biol Chem.* **280**: 24731-7.

- Schneider A, Bardakjian T, Reis LM, Tyler RC, Semina EV. 2009. Novel SOX2 mutations and genotype-phenotype correlation in anophthalmia and microphthalmia. *Am J Med Genet A.* **149A**: 2706–15.

- Sisodiya SM, Ragge NK, Cavalleri GL, Hever A, Lorenz B, Schneider A, Williamson KA, Stevens JM, Free SL, Thompson PJ, van Heyningen V, Fitzpatrick DR. 2006. Role of SOX2 mutations in human hippocampal malformations and epilepsy. *Epilepsia* **47**: 534–42.

- Tanaka S, Kamachi Y, Tanouchi A, Hamada H, Jing N, Kondoh H. 2004. Interplay of SOX and POU factors in regulation of the Nestin gene in neural primordial cells. *Mol Cell Biol.* **24**: 8834-46.

- Tziaferi V, Kelberman D, Dattani MT. 2008. The role of SOX2 in hypogonadotropic hypogonadism. *Sex Dev.* **2**: 194–9.

- Uchikawa M, Kamachi Y, 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**: 103-20.

- Visel A, Taher L, Girgis H, May D, Golonzhka O, Hoch RV, McKinsey GL, Pattabiraman K, Silberberg SN, Blow MJ, Hansen DV, Nord AS, Akiyama JA, Holt A, Hosseini R, Phouanenavong S, Plajzer-Frick I, Shoukry M, Afzal V, Kaplan T, Kriegstein AR, Rubin EM, Ovcharenko I, Pennacchio LA, Rubenstein JL. 2013. A high-resolution enhancer atlas of the developing telencephalon. *Cell* **152**: 895-908.

- Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, Afzal V, Ren B, Rubin EM, Pennacchio LA. 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**: 854-8.

- Wegner M. 1999. From head to toes: the multiple facets of Sox proteins. *Nucleic Acids Res.* **27**: 1409-20.

- Weiss J, Meeks JJ, Hurley L, Raverot G, Frassetto A, Jameson JL. 2003. Sox3 is required for gonadal function, but not sex determination, in males and females. *Mol Cell Biol.* **23**: 8084-91.

- Woolfe A, Goodson M, Goode DK, Snell P, McEwen GK, Vavouri T, Smith SF, North P, Callaway H, Kelly K, Walter K, Abnizova I, Gilks W, Edwards YJ, Cooke JE, Elgar G. 2005. Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol.* **3**: e7.

- Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, Nie J, Jonsdottir GA, Ruotti V, Stewart R, Slukvin II, Thomson JA. 2007. Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**: 1917-20.

- Zappone MV, Galli R, Catena R, Meani N, De Biasi S, Mattei E, Tiveron C, Vescovi AL, Lovell-Badge R, Ottolenghi S, Nicolis SK. 2000. Sox2 regulatory sequences direct expression of a (beta)-geo transgene to telencephalic neural stem cells and precursors of the mouse embryo, revealing regionalization of gene expression in CNS stem cells. *Development* **127**: 2367-82.

- Zhang Y, Wong CH, Birnbaum RY, Li G, Favaro R, Ngan CY, Lim J, Tai E, Poh HM, Wong E, Mulawadi FH, Sung WK, Nicolis S, Ahituv N, Ruan Y, Wei CL. 2013. Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. *Nature* **504**: 306-10.

- Zhong L, Wang D, Gan X, Yang T, He S. 2011. Parallel Expansions of Sox Transcription Factor Group B Predating the Diversifications of the Arthropods and Jawed Vertebrates. *PLoS One* **6**: e16570.

AIM OF THE THESIS

The aim of my PhD project was to identify and functional characterize novel transcriptional regulatory elements of genes implicated in neural development, candidate to be putative targets of the SOX2 transcription factor, thus possibly mediating its function in brain development and disease.

The main part of my PhD work is presented in Chapter 2. Our laboratory used the new ChIA-PET (Chromatin Interaction Analysis by Paired-End Tag sequencing) approach to obtain a genome-wide map of long-range chromatin interactions, comparing neurosphere cultures of wild-type (wt) neural stem/precursor cells (NPCs) and *Sox2*-deleted NPCs cultured from the mouse forebrain. Wt and *Sox2*-deleted NPCs were expanded in parallel for a few passages, chromatin was cross-linked and immune-precipitated with anti-RNApolII antibodies directed against the non-phosphorilated form of RNApolII (which is found in the pre-initiation complex) and analyzed by ChIA-PET. We chose this method based on recent discoveries regarding the importance of distal regulatory elements in gene expression. It is already known that these elements can lie very far from the gene they control on the linear chromosome map, but mutations in their sequences can cause important effects on the expression of the "connected" gene. Moreover, we analyzed the wt NPCs by ChIP-seq with anti-SOX2 antibodies, to define a genome-wide map of SOX2 binding sites. We noticed that a high number of putative distal regulatory elements presented a SOX2 ChIP-seq binding site and were associated to neural genes by long-range interactions.
In the context of this wider project, the aim of my work was to verify if some putative distal regulatory sequences, identified by ChIA-PET and presenting SOX2 binding sites (validated by ChIP-seq), were really able to work as regulatory elements in an *in vivo* situation, guiding the expression of a reporter gene in transgenic experiments in zebrafish. This would point to the value of this approach for identifying novel transcriptional regulatory elements spread in the genome. It would also allow us to identify novel molecular targets of SOX2, potentially involved in its important function in brain development and disease. Second, a further scope of my work was to verify if the activity of these distal regulatory sequences, tested *in vivo*, was responsive to SOX2 levels. To verify this hypothesis, I used two approaches: loss of function experiments in *in vivo* models of transgenic zebrafish lines and transfection assays in *in vitro* cultured cells (co-transfection with *Sox2* expression vectors).

The second part of my work is presented in Chapter 3. Our laboratory had found that mice deleted for *Sox2* (by a *Bf1*-Cre transgene at E9,5) presented a strong defective phenotype with major tissue loss in the ventral telencephalon. We noticed that also some tissue-specific determinants were down-regulated after *Sox2* deletion, such as NKX2.1 and SHH, very important effectors in the development of the ventral telencephalon. These observations suggested the hypothesis that SOX2 was able to operate a direct regulation on the expression of these determinants.
The aim of my work within this project was to address if *Nkx2.1* was a direct SOX2 target gene, by testing the responsivity of its promoter, which carries putative SOX2 binding sites, to SOX2. I cloned the *Nkx2.1* promoter region in a luciferase vector and tested it in transfection experiments in cultured cells, increasing amounts of SOX2 by co-transfection of a *Sox2* expression vector. Moreover, to confirm the requirement for the SOX2 consensus binding sites in the promoter, I performed site-specific mutagenesis of the putative SOX2 binding sites identified on the *Nkx2.1* promoter region. Finally, I tested if the activation I observed was SOX2-specific or if it could be replaced by other SOX transcription factors.

The third part of my work is presented in Chapter 4. In this paper, our laboratory observed that the transcription factor EMX2 was able to work as transcriptional repressor, inhibiting the BRN2 binding to a specific enhancer of *Sox2* gene. It is already known that often SOX2 and BRN2 act as co-factors, binding together their target genes. We asked if the inhibitory mechanism operated by EMX2, sequestering BRN2, could be generalized for other SOX2 target genes or if it was restricted to the *Sox2* locus.
In the experiments I did in the context of this work, I used a luciferase vector carrying an enhancer sequence of the *Nestin* gene, known to be bound by SOX2 and BRN2. I evaluated the capability of EMX2 to antagonize the luciferase activity, co-transfecting different amounts of SOX2 and BRN2 (alone or together) in presence of EMX2.

CHAPTER 2:

(paper in preparation)

SOX2 is required in brain-derived neural stem/precursor cells to maintain a genome-wide pattern of long-range chromatin interactions involving enhancers active in the brain

Jessica A. Bertolini[1,#], Rebecca Favaro[1,#], Yubo Zhang[2,#], Ben Martynoga[3], Cristiana Barone[1], Simona Delvecchio[1], Sara Mercurio[1], Jessica Mariani[1], Sergio Ottolenghi[1], Paul Robson[4], François Guillemot[3], Giulio Pavesi[5], Marcos J. Cardozo[6], Chia-Lin Wei[2], Paola Bovolenta[6], Silvia K. Nicolis[1]

[1] Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza 2, 20126 Milano, Italy.
[2] DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Walnut Creek, CA 94598, USA.
[3] MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK.
[4] Genome Institute of Singapore, 60 Biopolis St, Singapore 138672.
[5] University of Milano, Via Festa del Perdono 7, 20122 Milano, Italy.
[6] Centro de Biología Molecular Severo Ochoa, CSIC-UAM, c/Nicolás Cabrera 1, 28049 Cantoblanco, Madrid, Spain.

[#]Joint first authors

Abstract

The identification and characterization of regulatory sequences is crucial for understanding the spatial and temporal control on gene expression. It is already known that many transcriptional regulatory elements are localized very far from the genes they control on the linear chromosome map and they are able to reach the proximity of these genes through the formation of chromatin loops, called long-range interactions. Moreover, a lot of these distal regulatory elements (DREs) are localized in non-coding regions of the genome, in “gene deserts” or in introns of not-related genes. Comparing neurosphere cultures of wild-type (wt) neural stem/precursor cells (NPCs) and *Sox2*-deleted NPCs, we obtained a genome-wide map of long-range interactions of wt and, in parallel, *Sox2*-deleted NPCs, through a Chromatin Interaction Analysis by Paired-End Tag sequencing (ChIA-PET) method using an antibody against RNA polymerase II. We also obtained a genome-wide map of SOX2 binding DNA sites in wt NPCs, by ChIP-seq. We noticed that around half of the ChIA-PET long-range interactions lost in *Sox2*-

deleted NPCs included a SOX2 ChIP-seq peak. Then, we sought to identify and validate some putative SOX2-dependent DREs, by an *in vivo* (transgenesis in zebrafish) and *in vitro* (transfection in cultured cells) approach. In this way, we confirmed that some of the putative DREs, involved in wt-specific long-range interactions, were able to work as enhancers and guide the expression of a reporter gene in *in vivo* experiments to embryonic regions always including the forebrain. The activity of a subgroup of them responded to experimental changes in SOX2 levels, *in vivo* (by anti-*Sox2* morpholino oligonucleotides) or in transfected cells (by co-transfection of *Sox2*). We demonstrated that this genome-wide approach is a good method for identifying DREs spread among the genome map, active in the brain. Some of the SOX2-dependent DREs are connected to genes that are important for aspects of brain development, that are defective in *Sox2*-mutated patients (e.g. hippocampus development). Others are connected to genes associated with other brain diseases, that we observe in our mouse mutants (e.g. microcephaly), suggesting that *Sox2* deficiency may contribute also to these diseases in humans. Thus, this genome-wide approach could be useful to identify other SOX2-dependent DREs, including those associated to genes involved in genetic disease, to better investigate the regulation of the transcriptional mechanism and the implication of SOX2 in the onset of pathologies.

Introduction

Sox2 belongs to the *Sox* (*S*ry-related HMG b*ox*) family of transcription factors, playing important roles in development and differentiation. *Sox2* is expressed from early developmental stages in the morula and blastocyst inner cell mass (ICM) [Avilion *et al.*, 2003]; later, its expression is confined to the developing neural plate and subsequently to the neural tube. In the developing neural tube, *Sox2* expression remains high in the ventricular zone in active proliferation, while it decreases in the marginal zone where differentiation begins [Ferri *et al.*, 2004].

Sox2 expression is crucial during the early stages of embryonic development. Homozygous *Sox2*-KO (knock-out) mice die following loss of the stem cells of the blastocyst inner cell mass [Avilion *et al.*, 2003; Penvy *et al.*, 1998]. Using two *Sox2* conditional knock-outs in mouse (*Nestin-Cre* and *Bf1-Cre* transgene, activated at two different time-points during embryonic development), our laboratory discovered that *Sox2* is important for the development of the brain (hippocampus and basal ganglia) and for the maintenance of neural stem cells both *in vivo* (in

the hippocampus) and *in vitro* (for long-term self renewal) [Favaro *et al*., 2009; Ferri *et al.*, 2013].

Heterozygous *Sox2* mutations in humans cause defects in the development of eyes (anophthalmia, microphthalmia) [Fantes *et al.*, 2003; Schneider *et al.*, 2009] and hippocampus, with neurological pathology including epilepsy, motor control problems and learning disabilities [Ragge *et al.*, 2005; Sisodiya *et al.*, 2006; Kelberman *et al.*, 2006].

Recently, it was found that transcriptional regulatory elements are not always localized in the proximity of the genes they control but often lie very far from them on the linear chromosome map and they are able to reach the proximity of these genes, and rules their expression, through the formation of chromatin loops, called long-range interactions [Zhang *et al.*, 2013; Li *et al.*, 2012; Cheutin and Cavalli, 2014]. Mutations in these elements can cause pathology, due to the deregulation of the associated gene. For example, a single nucleotide mutation, found in the regulatory sequence located 460 kb upstream of the *Shh* gene, was discovered in an individual with holoprosencephaly; the mutation reduced the activity of the distant enhancer [Jeong *et al.*, 2008].

For this reason, the identification and functional characterization of regulatory sequences is crucial for understanding the spatial and temporal control on gene expressions.

In this work, we compared long-range DNA interactions in chromatin of wild-type mouse neural stem/precursor cells (NPCs) and *Sox2*-deleted cells (by *Nestin-Cre* transgene, at embryonic day E12,5; Favaro *et al.*, 2009), using the Chromatin Interaction Analysis by Paired-End Tag sequencing (ChIA-PET) technique [Fullwood *et al.*, 2009; Zhang *et al.*, 2013]. ChIA-PET mapping strategy is an unbiased whole-genome approach for the *de novo* analysis of chromatin interactions and for studying higher-order organization of chromosomal structures and functions [Fullwood *et al.*, 2009]. Using an antibody against RNA polymerase II, we obtained a genome-wide map of long-range chromatin interactions in wild-type and *Sox2*-deleted NPCs [Fullwood *et al.*, 2009; Zhang *et al.*, 2013].

We functionally validated a selected sample of distal elements, connected to genes in a SOX2-dependent way, characterizing their ability to work as transcriptional regulatory elements by *in vivo* (transgenesis in zebrafish) and *in vitro* (transfection in cultured cells) approaches.

Materials and methods

ZED constructs

The putative regulatory elements chosen were PCR amplified from genomic DNA obtained from telencephalic tissue of a CD-1 mouse. The PCR products were cloned in TOPO vector (Invitrogen, pCR®8/GW/TOPO® TA cloning KIT) and then recombined to each vector in a Zebrafish Enhancer Detection vector [Bessa *et al.*, 2009], by the Gateway recombination system (Invitrogen, gateway LR clonase II Enzyme mix). The primers used to amplified each putative regulatory element are listed in Table 2. The mouse genomic coordinates (mm9) for each putative regulatory element are listed in Table 1.

Luciferase constructs

We used a TK-LUC vector (provided by A. Okuda, Saitama Medical School, Saitama, Japan).
The putative regulatory elements *Sox4*DA and *Coup-TF1*DA1 were extracted from the TOPO vector (previously described), using EcoRI restriction enzyme, and cloned in pBluescript SK plasmid in EcoRI restriction site; then cut from pBluescript, using SacI and XhoI restriction enzymes, and clone in TK-LUC in SacI-XhoI restriction sites.
*Sox3*DA1 was extracted from the TOPO vector, using EcoRI enzyme, and cloned in pBluescript SK plasmid in EcoRI restriction site; then cut from pBluescript, using SmaI and XhoI restriction enzymes, and clone in TK-LUC in SmaI-XhoI restriction sites
*Sox3*DA2 and *Cxcr4*DA were extracted from the TOPO vector, using EcoRI enzyme, and cloned in TK-LUC in EcoRI restriction site.
The element Sox4PA were amplified from genomic DNA from telencephalic tissue of a CD-1 mouse and cloned in luciferase vector in BglII-SacI restriction sites.
hAKT3int element (provided by A. Visel, Berkley, California) were cloned in TK-LUC vector in KpnI restriction site.

In vivo experiments

Transgenesis in zebrafish
For each construct to test, one nanoliter containing 40 ng/µl of ZED plasmid was injected into one-cell-stage embryos of wild-type zebrafish, along with 50 ng/µl of *Tol2* mRNA to facilitate genomic integration. Green fluorescence was monitored

at 24 hours post fertilization (hpf) and the red fluorescence, as injection control, at 48 hpf.

Generation of stable transgenic lines

F0 transient transgenic zebrafishes were crossed with wild-type zebrafishes and their GFP+ F1 progeny grown to maturity. The F1 fishes were mated to obtain the F2 progeny. At the same time, the F1 fishes were mated to obtain eggs to use for loss-of-function experiments.

Morpholino injection

We used the morpholino antisense oligonucleotides (MO), specifically directed against the *Sox2* mRNA, previously validated by Kamachi *et al.* (2008); *Sox2*-MO sequence: 5'-GAAAGTCTACCCCACCAGCCGTAAA-3' (Gene Tools LLC). As control, we used a MO sequence not related to any mRNA transcripts; ctrl-MO sequence: 5'-CCTCTTACCTCAGTTACAATTTATA-3' (Gene Tools LLC). For each stable lines, one nanoliter containing 0,5 µM of *Sox2*-MO (or ctrl-MO) and 75 ng/µl of RFP mRNA, as control of injection, was injected into one-cell-stage embryos of zebrafish F1 stable lines. Green fluorescence and red fluorescence (for the injection control) were monitored during the early developmental stages, starting from the 5-somites stage, till at least the prim-6 stage (around 26 hpf).

Transfection experiments and luciferase assayes

P19 (a mouse teratocarcinoma cell line) cells were plated $4x10^4$/well in 12-well plates in 1 ml of α-MEM (Euroclone) for well and transfected after 24 hours, with calcium phosphate transfection method. For each transfection, we used 600 ng of luciferase vector and increasing amounts of *Sox2* and/or *Mash1* expressing vectors, calculated as molar ratios (1:0,075, 1:0,125, 1:0,250, 1:0,500; luciferase-vector:expressing-factor-vector). As control, we used equimolar amounts of *Sox2* and/or *Mash1* "empty" vectors. As control, 5 ng of Renilla luciferase was added to each transfection. pBluescript plasmid was added to each transfection to equalize total DNA to 2 µg. Luciferase activity was measured after 24 hours, using Dual-Luciferase® Reporter Assay System (Promega). The luciferase lectures were normalized on Renilla luciferase lectures.

Results

Sox2 loss causes profound changes in the genome-wide pattern of long-range interactions mediated by RNApolII in neural stem cell chromatin

In previous work, we had established that neural stem/precursor cell (NPC) cultures from the forebrain of mice, in which *Sox2* had been deleted by a *Nestin-Cre* transgene, and from their control non-deleted littermates [Favaro *et al.*, 2009]. While *Sox2*-mutant and control NPCs initially expand in culture with similar kinetics *Sox2*-deleted NPCs later fail to self-renew in long-term culture, pointing to a requirement for *Sox2* in NPCs maintenance that we also observed in *in vivo* NPCs of the brain (hippocampus) [Favaro *et al.*, 2009].

In the present work, we sought to determine the effect of *Sox2* loss on the genome-wide pattern of RNA-polII-mediated long-range interactions. We had previously determined the long-range interaction pattern of the wild-type cells by ChIA-PET analysis with anti-RNApolII antibodies [Zhang *et al.,* 2013]. We now compared, in the same way, ex-vivo cultures of NPCs from several normal and *Sox2*-deleted forebrains at P0.

It is significant that ex-vivo-brain-derived NPCs present gene expression and long-range-connectivity patterns that make them different (more "brain-related") than others, widely used NSC models, such as NS-5 cells (a clonal NSC line obtained by *in vitro* differentiation of ES cells) [Zhang *et al.*, 2013]. Indeed, various key regulators of forebrain development are expressed, and connected *via* long-range interactions, in forebrain-derived NPCs, but not in NS-5 cells, whereas other neural genes expressed more posteriorly along the neuraxis, such as *Hoxa* genes, and/or at earlier neural development stages, are preferentially active and connected in NS-5 cells [Zhang *et al.*, 2013]. Brain-derived NPCs thus appear to be particularly suitable for the study of the mechanisms of gene regulation related to forebrain identity.

Wild-type and *Sox2*-deleted NPCs were expanded in parallel for a few passages, chromatin was cross-linked and immune-precipitated with anti-RNApolII antibodies directed against the non-phosphorilated form of RNApolII (which is found in the pre-initiation complex) and analyzed by ChIA-PET [Zhang *et al.*, 2013] (Fig. 1a). Long-range interactions were comparatively classified as "common" (found in both wild-type and mutant), "specific" (found in wild-type, but not in mutant, or in mutant, but not in wild-type), and "alternative" (in which one of the two "interacting anchors" was the same in wild-type and mutant, but the other one changed) (Fig. 1b). Out of 7066 long-range interactions defined in wild-type (wt) NPCs, 2734 were lost in *Sox2*-deleted cells ("wt-specific" interactions), 3364 interactions were of the "alternative usage" type and 968 interactions were

"common" (unchanged) (Fig. 1c); moreover, 650 new interactions were detected specifically in *Sox2*-deleted NPCs ("mut-specific" interactions).
Among these SOX2-regulated interactions, most (2378 out of 2734 "wt-specific" interactions; 85% of wt-specific interactions) involve gene promoters and are localized at ±2,5 kilobases (kb) from the transcription start site of UCSC known genes, in agreement with previous RNApolII-ChIA-PET analyses of wild-type NSCs and ES cells [Zhang *et al.*, 2013]. Specifically, among these "wt-specific" interactions, about half (40,3%) connected two promoter-containing anchors, whereas the others connected a promoter-containing anchor to a distal intergenic (26%) or intragenic (18,6%) DNA region (Fig. 1d), in agreement with similar ratios previously found in normal NSCs, NS5 and ES cells [Zhang et al., 2013]. The remaining 15,1% of "wt-specific" interactions connected non-promoter regions.
It was previously found that the most cell-type specific interactions were those connecting promoter-containing anchors to distal anchors, whereas promoter-to-promoter-containing anchors interactions were more conserved between different cell types [Zhang *et al.*, 2013].

***Sox2*-dependent long-range interaction anchors are enriched in SOX2 binding**

Changes in long-range interactions following *Sox2* loss may be caused by a direct positive role of *Sox2* in maintaining interactions (the "wt-specific" interactions, lost following *Sox2* ablation), or by a direct negative role of *Sox2* in preventing the formation of interactions (the "mut-specific" interactions, appearing following *Sox2* ablation), or also by indirect *Sox2* functions.
To begin to relate direct SOX2 binding to SOX2-regulated long-range interactions, and more widely to obtain a reliable dataset of SOX2 binding data within our brain-derived NPCs as a basis for analyses of *Sox2* function, we performed a SOX2 ChIP-seq analysis of our wild-type brain-derived NPCs. The ca. 15000 SOX2 peaks detected (on duplicate samples) in our analysis showed significant overlap with previous SOX2 ChIP-seq datasets obtained with NS-5 cells [Lodato *et al.*, 2013; Engelen *et al.*, 2011] and other ES-derived NSCs [Bergsland *et al.*, 2011], but also specificities, as expected on the basis of the different origin of the cells. We first asked whether SOX2 peaks were enriched within interacting anchors, as compared with a random distribution within the genome. SOX2 peaks were indeed highly enriched within anchors, in agreement with a functional link between long-range interactions and direct SOX2 binding. We then evaluated the enrichment in SOX2 peaks of the different categories of interactions (Fig. 2), with special attention to the SOX2-dependent interactions. In "wt-specific" interactions (lost following *Sox2* ablation), about 45% of the interactions carried a SOX2 peak,

whereas by contrast, in "mut-specific" interactions, only 28% of the interactions carry a SOX2 peak (this is the interaction category that has by far the lowest frequency of SOX2 peaks). This suggests that *Sox2* is, on average, more frequently involved in maintaining interactions, rather than in preventing their formation. On the other hand, also "common" interactions are highly enriched in SOX2 peaks (about 71%) (Fig. 2), indicating that SOX2 is not strictly required in the maintenance of this subset of interactions.

Sox2-dependent long-range interaction anchors are enriched in enhancers active in the forebrain of transgenic mice

Do long-range interaction anchors identify DNA elements that have active functional roles for gene expression in the brain? To begin to address this question, we first looked at genes that we knew to be directly regulated by SOX2 and to play important functions in brain defects caused by *Sox2* loss. *Nkx2.1* encodes a transcription factor, required for the development of the ventral forebrain, whose expression is drastically down-regulated following *Sox2* early deletion in the developing telencephalon [Ferri *et al.*, 2013]. *Sox2* telencephalic deletion causes a dramatic loss in ventral telencephalic tissue, pointing to *Nkx2.1* as an important mediator of SOX2 function in the developing brain; SOX2 directly binds and regulates the *Nkx2.1* promoter [Ferri *et al.*, 2013]. In our ChIA-PET analysis, *Nkx2.1* is connected to a DNA region downstream to the gene in wild-type, but not in *Sox2*-mutant cells (Fig. 3a). Within this region, an enhancer is found, that was previously shown by ChIP-seq (Fig. 3c) [Visel *et al.*, 2009] to be bound by the transcriptional coactivator p300 in the forebrain of the E11,5 mouse embryos, but not in the mesencephalon or in the limb. These experiments had found that p300 tissue-specific binding was highly predictive of enhancer activity within the examined tissues (forebrain, or limb) in transgenic assays [Visel *et al.*, 2009]; the *Nkx2.1*-connected region directed expression of a reporter LacZ transgene to the embryonic forebrain (diencephalon), in a region which is part of the endogenous *Nkx2.1* expression domain (Fig. 3d). Interestingly, the *Nkx2.1*-connected region, cloned in LacZ transgene and used to obtain transgenic mice is the human genomic sequence orthologous to the enhancer candidate region identified by p300 ChIP-seq [Visel *et al.*, 2013]. It reveals a high conservation in the transcription mechanism during evolution. Thus, our ChIA-PET analysis detects a SOX2-dependent interaction between *Nkx2.1*, a gene regulated by SOX2, and a forebrain enhancer.

We next looked at other SOX2-dependent interactions, asking whether they might involve other enhancers previously identified by the p300 ChIP-seq assay and

validated by transgenesis (VISTA Enhancer Browser, http://enhancer.lbl.gov) [Visel *et al.*, 2013]. For all of the enhancers already validated by Visel *et al.*, the enhancer candidate region cloned in the LacZ transgene is the corresponding human genomic sequence, orthologous to the one identified by p300 ChIP-seq assay in mouse.

The *Sox4* transcription factor gene is important to maintain neuronal cells throughout the developing central nervous system [Bergsland *et al.*, 2006; Cheung *et al.*, 2000]; a SOX2-dependent interaction connects this gene to a VISTA enhancer, active in the telencephalon and neural tube, containing a SOX2 ChIP-seq peak (Fig. 4). This enhancer is located 650 kilobases (kb) away from *Sox4*, within an intron of a gene active in liver and pancreas (Fig. 4).

The *Sox3* transcription factor gene, coexpressed with *Sox2* in the developing nervous system, is important for neural development, and its mutation in humans leads to defects of the hypothalamic-pituitary axis [Alatzoglou *et al.*, 2009]. Two SOX2-dependent long-range interactions involve *Sox3*; one of them (*Sox3*DA1) connects the *Sox3* gene to a telencephalic enhancer, located 350 kb away; the second one (*Sox3*DA2) contains a very high SOX2 ChIP-seq peak (Fig. 5).

Another VISTA forebrain enhancer is located within an intron of the *Akt3* gene, associated to microcephaly and intellectual disability [Boland *et al.*, 2007], and it is connected to the promoter of the *Zfp238* gene, encoding a transcription factor, by another SOX2-dependent interaction. *AKT3* gene is also associated to a rare case of megalencephaly [Riviere *et al.*, 2012].

Given these interesting individual examples, we asked whether VISTA forebrain enhancers would be more represented within interacting anchors of the different categories, than within the total DNAaseI-hypersensitive sites, representing potential regulatory elements. As the p300 ChIP-seq experiment [Visel *et al.*, 2009] had been conducted, in parallel, on the E11,5 forebrain and limb, detecting a set of limb-specific enhancers, we also asked, comparatively, about enrichment of our anchors in these limb enhancers. This analysis detected a strong selective enrichment, within SOX2-dependent interaction anchors, of forebrain enhancers, as compared to limb enhancers. This result indicates that DNA regions, involved in SOX2-dependent long range interactions, are enriched in enhancers active within the developing forebrain.

An inspection of long-range interactions affecting other neural genes identified the involvement of other enhancers, previously identified by transgenic assays. An interesting case is *Sox10*, a transcription factor involved in gliogenesis and myelination and mutated in several human genetic diseases affecting these functions [Inoue *et al.*, 2004]. Two enhancers active in the developing CNS are connected in normal, but not mutant cells; in mutant cells, a novel interaction develops, that connects the gene promoter to one of such enhancers.

SOX2-dependent long-range interactions predict novel forebrain enhancers active in transgenic fishes

The detection of previously validated forebrain enhancers within SOX2-dependent interacting anchors was an intriguing finding. However, only a fraction of the anchors (about 4%) contained enhancers that had been previously validated in the VISTA enhancer atlas. Thus, we wished to address, in a more general way, the functional regulatory properties of DNA regions, involved in SOX2-dependent long-range interactions. We chose a transgenic assay in zebrafish (*Danio rerio*), allowing to test the ability of a mouse DNA region to drive expression of a GFP transgene (directed by a minimal promoter that is inactive by itself) throughout embryogenesis in zebrafish, using a Zebrafish Enhancer Detection (ZED) vector. The ZED vector [Bessa *et al.*, 2009] carries, together with the GFP reporter gene, also a red fluorescence reporter gene active in the developing muscle, allowing to count transgenic embryos independently from the GFP expression and useful as control for the transgenesis efficiency. For this functional test, we chose anchors involved in SOX2-dependent long-range interactions ("wt-specific" interactions) and carrying SOX2 ChIP-seq peaks. Among these, we focused on distal anchors (DAs) located outside genes, or within gene introns, as opposed to promoter-containing anchors (PAs); these two categories together represent almost the 50% of the total "wt-specific" interactions. The reason for this choice lies in our previous finding [Zhang *et al.*, 2013] that promoter-to-non promoter interactions are the more cell-type-specific category of interactions.
The selected distal anchors, to test as putative distal regulatory elements (DREs), are associated in a SOX2-dependent way to genes important for neural development, candidate to be putative SOX2 target genes, in mouse wt NPCs.
Out of 13 constructs generated with each selected SOX2-dependent distal anchors, 12 directed GFP expression to the developing forebrain (Table 1); some were further active in the more posterior brain region and neural tube. The GFP expression, guided by each tested DRE, was compared to the expression of the endogenous zebrafish gene, homologous to the mouse gene connected to the tested sequence (Fig. 6-7). Remarkably, GFP expression closely matched the endogenous expression in zebrafish forebrain, or part of it (Fig. 6-7). For example, GFP expression, driven by an anchor embedded within the intron of a pancreatic/hepatic gene and connected in mouse to the *Sox4* promoter, was detected within the telencephalic area of endogenous *Sox4* expression (Fig. 6). Similar data were obtained with anchors connected to important regulators of forebrain development: *Sp8*DA, *Cxcr4*DA, *Sox3*DA1, *Nkx2.1*DA, *Irx1*DA (Fig. 7; supplementary fig. 1), *COUP-TF1* DA1-2 (not shown).

The only construct, in which GFP expression was detected also outside the endogenous expression pattern of the associated gene, was the *Irx1*-connected region. In addition to the comparable expression in brain, this construct gave rise to a very strong GFP expression in the neural tube, not matching the *Irx1* pattern (Fig. 7).
Overall, these data indicate that RNApolII-mediated SOX2-dependent long-range interactions between genes and distal non-promoter regions identify novel forebrain enhancers, contained within anchors, with high confidence.
Interestingly, we noted that 12, out of 13, mouse DREs are able to guide reporter gene expression in developing zebrafish forebrain, indicating that the regulatory mechanisms underlying their function are highly conserved in evolution.

Experimental manipulation of SOX2 levels affects GFP-expression mediated by some SOX2-dependent distal elements

We further wished to ask whether these forebrain enhancers were also responsive to SOX2 levels in their transcription-activating function.
We obtained 8 stable zebrafish transgenic lines carrying enhancer-GFP constructs, from the initial 13 transient transgenic lines. To test if the enhancer activity of DREs is regulated by SOX2, we decided to use a loss of function approach: we injected transgenic eggs from stable transgenic zebrafish lines with a morpholino antisense oligonucleotides (MO) previously optimized to specifically down-regulate endogenous *Sox2* levels [Kamachi *et al.*, 2008], as compared to MO control. When embryos carrying transgenes driven by the *Sp8*-, or the *Sox3*DA1-connected distal anchors, were injected with *Sox2*-MO, a selective down-regulation of transgene expression in the forebrain was observed (Fig. 8, *Sp8*DA-GFP*;* Fig. 9, *Sox3*DA1-GFP). Other transgenic constructs (5) did not respond this way (e.g. *Sox4*DA-GFP), suggesting heterogeneity in the degree and/or stage-specificity of SOX2 responsiveness among enhancers, or compensation of SOX2 function by other SoxB proteins.
These results point to a responsiveness to SOX2 levels of the activity of at least some of the newly identified forebrain enhancers *in vivo*.

Cotransfection of SOX2 alone does not significantly activate the distal regulatory elements tested in luciferase assays in cultured cells

We further addressed SOX2 responsiveness in transfections; we cloned distal anchors upstream to a tk minimal promoter and luciferase and co-transfected

these constructs, with increasing amounts of a *Sox2* expression vector, in P19 cultured cells (a mouse teratocarcinoma cell line with neural characteristics). When the *Sox4* distal anchor constructs were tested in co-transfection with *Sox2* expression vector only, no significant increase in luciferase expression was observed (Fig. 12a); similar results were obtained with the *Sox3*DA1; *Coup-TF1* DA1 showed a moderate increase (about 3-fold maximum stimulation by SOX2 observed, to be compared with 1,5x stimulation of the "empty" tk-luciferase control vector) (Fig. 10). Furthermore, they did not seem to work as strong enhancers, because their luciferase basal level was comparable with the control one (tk-luc vector), in contrast with their ability to work as enhancers already seen in transgenesis in zebrafish. Curiously, some DREs-luciferase constructs (*Sox3*DA2- and *Cxcr4*-connected region; Fig. 10) seemed to work as a silencer in *in vitro* experiments and the co-transfection with increasing amounts of *Sox2* expression vector did not significantly increase the luciferase activity.

As control and in parallel to the experiments regarding the distal anchors (DAs) tested as distal regulatory elements (DREs), we cloned also two promoter-containing regions (PAs) in luciferase vector. The PAs are the DNA regions identified to be associated to their corresponding DAs *via* long-range interactions in mouse NPCs, that are localized in proximity of gene promoter regions. We decide to clone a small portion (highly enriched in SOX2 ChIP-seq peaks) of the *Zfp335*-promoter-containing region (*Zfp335*PA), the element connected with the *Zfp335*-connected region (*Zfp335*DA) (Fig. 11), upstream of the tk minimal promoter, to test it as enhancer. Then, we cloned the *Sox4*-promoter containing region (*Sox4*PA), connected to the *Sox4*DA, upstream of luciferase gene, to test it as promoter (Table 1). Interestingly, both the PAs sequences, transfected alone in P19 cultured cells, showed a significant increase in luciferase activity (Fig. 12b), particularly strong for the *Sox4*PA, in comparison with the control vector without the PA sequence, proofing the capability of the *Zfp335*-promoter-containing region and of the *Sox4*-promoter-containing region to work respectively as an enhancer and a promoter. Moreover, the *Zfp335*-promoter-containing region luciferase construct showed a SOX2 dose-dependent activation, increasing amounts of the co-transfected *Sox2* expression vector, and about a 3-times fold-increase in luciferase activity in presence of the maximum quantity of *Sox2* expression vector used, in comparison with its own basal level (Fig. 11). Interestingly, the *Zfp335*-promoter-containing region, as the Z*fp335*-connected region too, were the only sequences tested in transgenesis in zebrafish that didn't present any enhancer characteristic, resulting to be not able to guide GFP expression. The *Sox4*-promoter-containing region showed a 2-times fold increase in luciferase activity in presence of the maximum quantity of *Sox2* expression vector.

Curiously, the DREs sequences, already validated as forebrain enhancers in transgenesis in zebrafish, are not able to increase luciferase activity in transfection experiments in P19 cultured cells. Increasing amounts of SOX2 in cells don't produce any positive results in transcriptional activation. On the contrary, the two promoter-containing regions tested are actively regulated by SOX2.

SOX2, in presence of its co-factor MASH1, is able to strongly increase transcriptional activity in luciferase assays

Looking at the previous data, the DREs tested in *in vitro* experiments didn't work as enhancers in cultured cells and *Sox2* alone didn't seem to have a significant role to increase the luciferase activity. However, we noticed that some enhancers, whose activity in transgenic brain had been demonstrated, drove expression preferentially to the differentiating cells of the marginal zone, rather than to the stem cell-containing ventricular zone, and were active within a specific subregion of the brain, the ganglionic eminences (primordia of the basal ganglia) (VISTA enhancer atlas) (e.g. *Sox4*DA; Fig. 4). This raised the hypothesis that SOX2 may not be sufficient by itself, but may require combination with other transcription factors, active in specific regions of the differentiating brain neuroepithelium (e.g. ganglionic eminences). *Mash1/Ascl1* encodes a transcription factor expressed in, and important for, the differentiating ganglionic eminences [Castro *et al.*, 2011]; intriguingly, a MASH1 binding regions had been detected by ChIP-seq in neural stem (NS-5) cells [Castro *et al.*, 2011], that precisely overlaps the SOX2-binding peak detected in the distal anchor connected to *Sox4* in our cells. We thus co-transfected the luciferase construct carrying *Sox4*-connected region with *Sox2* and *Mash1* expression vectors. Whereas MASH1 alone, as SOX2 alone, was unable to transactivate the construct, a strong synergy was observed by co-transfecting with *Sox2* and *Mash1* together, leading to a 10-20-fold increase in luciferase activity (Fig. 12a). A similar result (with a lower level of transactivation) was seen with the human sequence of *AKT3* intronic enhancer (not tested in transgenesis in zebrafish by us but confirmed to be a forebrain enhancer in transgenic mice by Visel *et al.*, 2013) (Fig. 13), that also carries a MASH1 ChIP-seq peak.
Interestingly, the promoter of the *Sox4* gene is strongly active in transfection experiments in the absence of added SOX2 (as seen before) and MASH1; when the distal SOX2/MASH1-binding enhancer is further included in the construct, the activity is additively increased, and synergistically stimulated by *Sox2* and *Mash1*, as seen with the enhancer alone (Fig. 12b)
Instead, other DREs (*Cxcr4-*, *Sox3*DA2-connected region), that present a SOX2 peak close to a MASH1 peak, do not show any increase in transcription activity if

co-transfected with *Sox2* and *Mash1* expression vectors together (data not shown).
This results prompted us to examine the degree of overlap between SOX2 and MASH1/ASCL1 binding peaks at the genome-wide level. About 25% of the SOX2 peaks overlapped with MASH1 peaks (4798); these represented about 25% of the total MASH1 peaks (about 19000; Castro *et al.*, 2011). This result indicates that a significant proportion of SOX2 peaks within anchors may be jointly regulated by SOX2 and MASH1.

Discussion

The identification and characterization of regulatory sequences is crucial for understanding the spatial and temporal control on gene expression. It is already known that many transcriptional regulatory elements are localized very far from the genes they control on the linear chromosome map and they are able to reach the proximity of these genes through the formation of chromatin loops, called long-range interactions.
In this work we compared neurosphere cultures of wild-type (wt) neural stem/precursor cells (NPCs) and *Sox2*-deleted NPCs, and we obtained a genome-wide map of long-range interactions of wt and, in parallel, *Sox2*-deleted NPCs, through a Chromatin Interaction Analysis by Paired-End Tag sequencing (ChIA-PET) method using an antibody against RNA polymerase II.
We observed that, following *Sox2*-loss, a high proportion of RNApolII-mediated long-range interactions, identified in wt NPCs chromatin by ChIA-PET, were lost in *Sox2*-mutated NPCs, pointing to the importance of the presence of SOX2 in their maintenance. Moreover, we observed that 85% of the wt-specific long-range interactions, lost in mutated cells, involved the promoter region of UCSC known genes, many of which important for neural development. Interestingly, we noticed the appearance of about 650 new long-range interactions after the deletion of *Sox2*. In this case, SOX2 could have also a negative role in preventing the formation of interactions.
In parallel, we determined the genome-wide map of SOX2 binding sites in chromatin of wild-type NPCs, by ChIP-seq. We noticed that SOX2 peaks were highly enriched within long-range interaction anchors: at least half of the SOX2-dependent long-range interactions contain a SOX2 ChIP-seq peak, suggesting that SOX2 has a direct role in their maintenance. A significantly lower proportion of SOX2 peaks have been identified in the new mutant-specific long-range interactions, formed after *Sox2* loss. This suggests that that SOX2 is more

frequently involved in maintaining interactions, rather than in preventing their formation.
Moreover, we observed that a portion of the wt-specific long-range interactions, identified by ChIA-PET in wt NPCs and lost in *Sox2*-mutated NPCs, were enriched in enhancer sequences identified by p300 ChIP-seq on mouse forebrain tissue and already validated as forebrain enhancers in transgenic mice [Visel *et al.*, 2009; Visel *et al.*, 2013]. One very interesting thing of these experiments is that the sequences tested in transgenic mice were the human genomic sequences orthologous to the enhancer candidate regions identified by p300 ChIP-seq in mouse tissues. The human sequences were able to guide reporter gene expression in mouse and work as enhancers [Visel *et al.*, 2009; Visel *et al.*, 2013]. Three of the interaction anchors tested in our work in *in vivo* experiments, as putative distal regulatory elements (DREs), included a p300 validated forebrain enhancer in their sequence: the *Nkx2.1*-, *Sox4*- and *Sox3*DA1-connected regions. An additional human sequence, located in a intron of *AKT3* gene, was also a forebrain enhancer in transgenic mice and we tested it in *in vitro* assay.
The DRE connected to *Sox4* (Fig. 4) is also particularly interesting because it is located at 650 kb from the *Sox4* gene, forming a very long chromatin loop. In fact, this particular enhancer is localized within an intron of a gene active in liver and pancreas but, through transgenesis in zebrafish, we noticed its capability to work as forebrain enhancer. This supports the idea that the regulatory elements could be localized also a many kb from the gene they control, an important aspect in the overall function of long-range interactions [Zhang *et al.*, 2013].

Using a transgenic assay in zebrafish, we tested 13 DREs connected to genes important for brain development in wt mouse NPCs, but whose connection to the gene was lost in *Sox2*-mutant NPCs.
Out of 13 tested DREs, 12 of them were able to drive the expression of the reporter GFP gene in a regulated way, proving to work as forebrain enhancers during embryonic development in zebrafish. Significantly, the GFP expression pattern observed matched the endogenous expression pattern of the DRE-associated neural gene, or at least a part of it, in zebrafish at the same developmental stage analyzed. This confirmed the idea that the putative DREs, identified by the ChIA-PET assay in brain-derived mouse NPCs, were able to work as enhancers in the brain in *in vivo* experiments.
A further consideration is that the sequences tested in zebrafish were cloned from mouse DNA, but they were able to guide the GFP expression pattern in a different organism. It focuses the attention on the evolutionary conservation of the gene regulation mechanism.

Moreover, looking at the tested DREs, that include a sequence already validated as forebrain enhancer in transgenic mice by Visel *et al.* (2013), we can notice a partial overlap in the GFP expression pattern, that we obtained in zebrafish using the mouse sequence, and the LacZ expression pattern observed in mouse with the orthologous human sequence. This focuses the attention on the importance of these enhancer sequences during the embryonic development of vertebrates and, again, the conservation of their mechanism of action during evolution.
Only one of the 13 tested DREs (the *Irx1*-connected region; Fig. 7) appeared to guide GFP expression in a wider domain, compared to the endogenous expression of the associated gene in zebrafish. We could speculate that this specific sequence alone may not be sufficient to regulate GFP expression *in vivo* but it may also need the presence of other regulatory sequences absent in the genomic context in study.
One distal anchor (*Zfp335*DA) did not work as enhancer in transgenic zebrafish. Curiously, also the *Zfp335*-promoter containing anchor (*Zfp335*PA: the DNA element associated to *Zfp335*DA *via* long-range interaction) did not present any enhancer activity in transgenesis experiments, neither alone nor in combination with the distal anchor (data not shown). Also in this case we could speculate that these sequences may need the presence of other regulatory sequences absent in the genomic context in study. Indeed, one of these sequences (a SOX2 peaks containing region within the *Zfp335*PA) worked as enhancer in a cellular context, where it responded to SOX2 levels, suggesting that it possesses some characteristics of transcriptional regulatory elements. Alternatively, perhaps these regulatory elements evolved their proprieties during evolution; in fact, these sequences are conserved between mouse and human but we could not detect any conservation with fish (even though the fish presents a gene homologous to *Zfp335*).

We also wished to ask whether these forebrain enhancers were responsive to SOX2 levels in their transcription-activating function. Out of 8 stable transgenic zebrafish lines, obtained from the 13 transient transgenic lines, two responded to a SOX2 decrease. We injected a morpholino antisense oligonucleotides (MO), specifically directed again *Sox2* mRNA, in zebrafish embryos of stable transgenic lines. The lines transgenic for the *Sp8*-connected region and one of the two *Sox3*-connected region showed a relevant decrease in GFP expression, specifically in forebrain, at early developmental stages (Fig. 8-9). In comparison, the GFP expression in somites (for *Sp8*DA-GFP line) or in more posterior neural regions (for *Sox3*DA1-GFP line) was respectively unaffected and relatively unaffected. This shows that a subgroup of the identified DREs are really responsive to SOX2 levels. For the other lines that do not present relevant changes in GFP expression after

the MO injection, we can speculate that the presence of other transcription factors, perhaps belonging to the SoxB1 family genes, could be able to compensate for the absence of *Sox2*, while for the transgenic lines of *Sp8*- and *Sox3*-connected regions this compensation did not occur.
We further cloned some of the DREs, previously tested in transgenic experiments in zebrafish, in luciferase vector, upstream of a tk minimal promoter. By transfecting P19 cultured cells, we observed that none of 5 tested DREs was able to work as powerful enhancer alone and to increase significantly its luciferase activity following SOX2 co-transfection (Fig. 10; Fig. 12a). Only the two promoter-containing regions tested demonstrated an ability to activate luciferase transcription and to be responsive to SOX2 levels by a dose-dependent increase in luciferase activity (Fig. 11; Fig. 12b). We could speculate that the DREs didn't work as enhancers in cultured cells, even if their ability to guide a reporter gene expression had been already validated in transgenesis in zebrafish, because of differences in the context: the lack of some other sequences able to interact with them in an *in vivo* situation (such as the connected promoter region identified by ChIA-PET), or the lack of other transcription factors, not expressed in P19 cells, that might help SOX2 binding to the tested sequences.
In relation to this hypothesis, we noticed that our SOX2 ChIP-seq peaks are often localized in proximity of ChIP-seq peaks of another transcription factor, MASH1/ASCL1 [Castro *et al.*, 2011], active in more differentiated neural regions. Considering the analyzed DREs, we observed that the *Sox4*-connected region presented these characteristics: the presence of a SOX2 binding site close to a MASH1 binding site. Co-transfecting the *Sox4*DA-luiferase vector with both the *Sox2* and *Mash1* expression vectors, we observed a significant increase (10-20 fold) in luciferase activity. Only very low concentrations of the *Mash1* expressing vector were sufficient to induce a SOX2 dose-dependent response in transcriptional activation. While the co-transfection of the luciferase vector with the *Sox2* or *Mash1* expression vectors individually did not give rise to an increase in luciferase activity (Fig. 12a).
As already seen, the *Sox4*-promoter containing region (alone and co-transfected with *Sox2* expression vector) is strongly active in P19 cell. We wondered what could happen if, close to the *Sox4* promoter, we also cloned the connected DRE region, identified by long-range interaction, to create a situation more similar to that obtained after the formation of the chromatin loop between the two interacting regions in NPCs. For this specific anchors combination, we noticed an additive increase in luciferase activity, due to the presence of both sequences (*Sox4*DA and PA) responsive to *Sox2* and *Mash1* combined amounts (Fig. 12b). We could suppose that, by adding an enhancer element to a promoter sequence (even if it is already strongly active), the responsiveness to a specific combination

of transcription factors could be increased. This could be a possible explanation of the utility of the chromatin loop formation.
Another interesting example is the element localized within a long intron of the human *AKT3* gene and already validated as forebrain enhancer by Visel *et al.* (2013). This is the human sequence orthologous to the one identified in mouse tissue by p300 ChIP-seq [Visel *et al.*, 2009]. Comparing this sequence with the ChIA-PET analysis on our NPCs from mouse forebrains, we found that this region (in the *Akt3* mouse gene) interacted with the *Zfp238* promoter region, another gene involved in neural development, and presented a SOX2 binding site close to a MASH1 binding site. The *AKT3* gene is also particularly interesting because it has been found associated to human pathologies, such as microcephaly and intellectual disability [Boland *et al.*, 2007]. Considering the high evolutionary conservation of the element between human and mouse, we cloned in a luciferase vector the same human sequence tested in transgenic mice by Visel *et al.* In transfection experiments, the human-*AKT3*-intron-containing (h*AKT*3in) region alone appeared to work as a silencer, reducing the luciferase activity to less than 1/3, compared to the control vector without the tested sequence. Co-transfecting the luciferase vector with the *Sox2* and *Mash1* expression vectors, we can observe a dose-dependent activation increasing transcription factors amounts, up to a 3-fold increase in luciferase activity. Instead, the co-transfection of the luciferase vector with *Sox2* or *Mash1* expression vectors individually, did not produce an increase in the luciferase activity (Fig. 13). We can speculate that, even if the h*AKT*3in region worked as a silencer element in a cellular context (we didn't check it in fish), it is transactivated by SOX2 and MASH1, if they are present together.
As some regulatory elements need a combination of transcription factors to be activated, we could speculate that this could happen also for other transcription factors, known to be SOX2 co-factors. In fact, it is known that SOX2 often binds co-factors to regulate some of its target genes. Many examples document an interaction between SOX2 and POU proteins, such as OCT4 and BRN2 (transcription factors important during development). The SOX2-OCT4 protein complex regulates the expression of different genes in embryonic stem (ES) cells, such as *Fgf4* [Ambrosetti *et al.*, 1997], *Utf1* [Nishimoto *et al.*, 1999] and *Nanog* [Rodda *et al.*, 2005]. Instead, the SOX2-BRN2 protein complex is active in more differentiated cells of the neural lineage [Lodato *et al.*, 2013]. However, P19 cells do not express the *Brn2* gene, although they express the *Oct4* gene [Jin *et al.*, 2009]. So, we can speculate that the majority of our tested DREs did not show enhancer characteristics in *in vitro* assays because of the lack of co-factor BRN2. Instead, in the context of a whole organism, this protein is present and the sequences are able to work as enhancers through the formation of the SOX2-

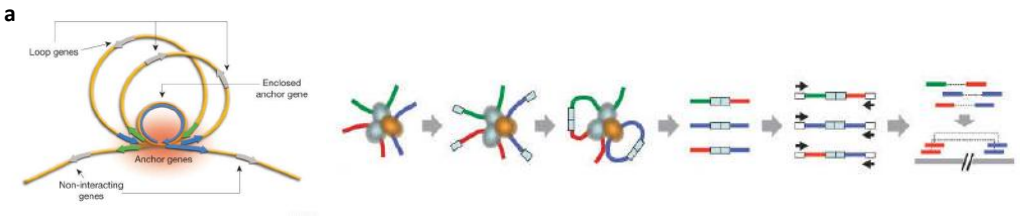
BRN2 protein complex. Thus, to test this hypothesis, we could try to transactivate the DRE-luciferase vectors with *Brn2* and *Sox2* expression vectors combined in P19 cells.

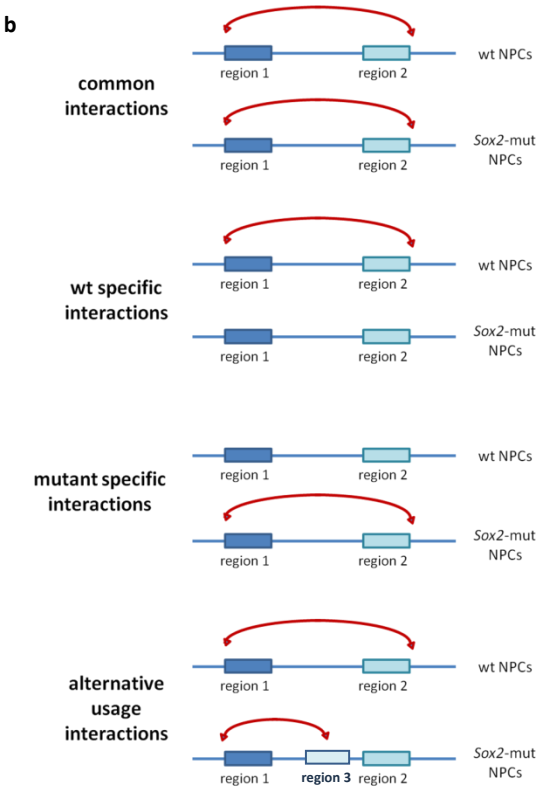
Analyzing the data obtained from the ChIA-PET assay, it is possible to identify many interesting situations in which the loss of *Sox2* induces loss of long-range interactions involving genes connected to different diseases. Among the genes I have studied, several are involved in human genetic diseases: mutations in *Sox3* are associated to X-linked hypopituitarism [Alatzoglou *et al.*, 2009; Rizzoti *et al.*, 2004], while mutations in *Coup-TF1* are associated to optic atrophy with intellectual disabilities [Bosch *et al.*, 2014].
Moreover, long-range interactions are localized in regions affected by heterozygous deletions found in patients affected by CNS genetic diseases, in which a single causative gene has not yet been identified. One interesting case involves a region found deleted in patients with Wolf-Hirschhorn syndrome (Fig. 14), a disease due to deletions in the human chromosome 4 (in the region orthologous to the mouse region in study) and characterized by mental retardation, microcephaly and cranial malformations [Battaglia *et al.*, 1999]; of note, microcephaly and cranial malformations are also detected in *Sox2*-mutant mice [Ferri *et al.*, 2013]. This region appears to be a hub of SOX2-dependent long-range interactions, lost in *Sox2*-mutant cells. This information might suggest a link between lack of *Sox2* and Wolf-Hirschhorn syndrome, by a mechanism that might involve linking genes that have to work coordinately, perhaps an example of how long-range interactions could be important and predictive in the identification of regulatory elements involved in genetic diseases.

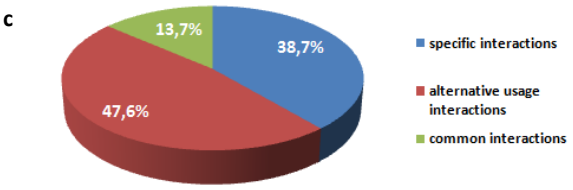
Conclusion

In this work we demonstrated that genome-wide analysis by ChIA-PET approach is a good method for identifying distal regulatory elements (DREs) spread within the genome, specifically regulated by SOX2. Moreover, considering the importance of SOX2 during development, this genome-wide approach could be useful to identify other SOX2-dependent DREs associated to genes involved in genetic disease, to better investigate the mechanism behind the regulation of these genes mediated by SOX2 and the implication of SOX2 in the onset of pathologies.

Figures


a
Loop genes
Enclosed anchor gene
Anchor genes
Non-interacting genes


b
common interactions
wt specific interactions
mutant specific interactions
alternative usage interactions
region 1
region 2
region 3
wt NPCs
Sox2-mut NPCs


c
13,7%
38,7%
47,6%
specific interactions
alternative usage interactions
common interactions

d

	Prom-Prom	Prom-Intragenic	Prom-Intergenic	Non Promoter
Specific in wt-NPCs	40,3%	18,6%	26,0%	15,1%
Common in wt-NPCs	53,2%	23,2%	17,6%	6,0%
Alternative in wt-NPCs	47,5%	22,9%	22,7%	7,0%
Specific in mut-NPCs	24,4%	15,8%	25,6%	34,2%
Common in mut-NPCs	52,9%	23,3%	17,4%	6,4%
Alternative in mut-NPCs	41,5%	22,7%	21,5%	14,3%

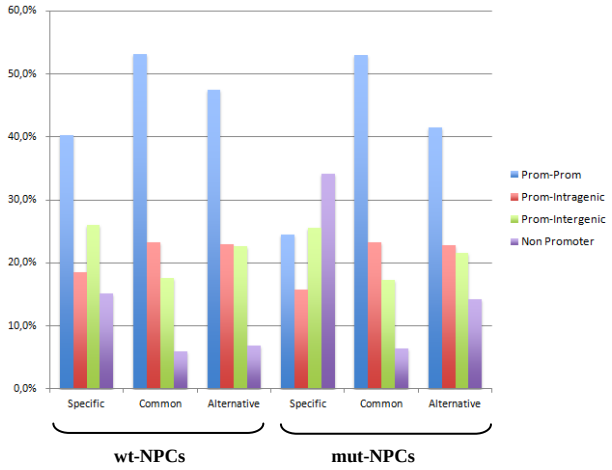


Figure 1. a. Scheme of genome-wide detection of long-range DNA interactions in chromatin (ChIA-PET) [image source: Fullwood *et al.*, 2009]. **b.** Scheme of the interaction classification. **c.** Percentage of the three interaction types in wt-NPCs. **d.** Percentage of the "promoter-centric" (classified as "promoter-promoter", "promoter-intragenic region" and "promoter-intergenic region") interactions and "non-promoter" interactions, identified in wt-NPCs and *Sox2* mutant-NPCs.

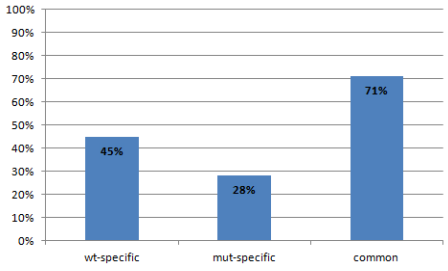


Figure 2. Enrichment in SOX2 ChIP-seq peaks within the different categories of ChIA-PET interactions.

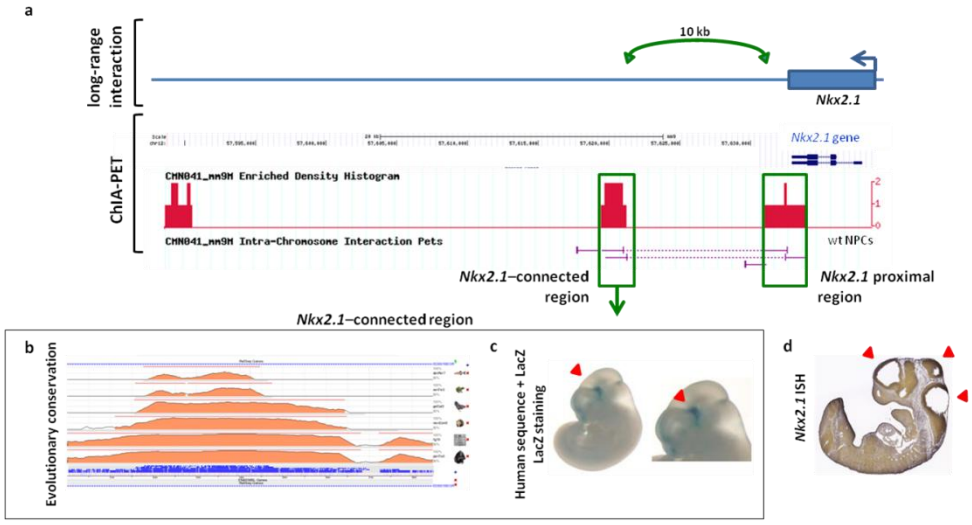


Figure 3. Analysis of the *Nkx2.1* long-range interaction. a. Representation of SOX2-dependent long-range interaction (lost in *Sox2*-mutant NPCs) identified in wt NPCs by ChIA-PET assay and its localization on genome map. **b.** Multispecies vertebrate conservation plot (ECR browser program) of the *Nkx2.1*-connected region identified by p300 ChIP-seq assay made by Visel *et al.* (2013). **c.** LacZ staining of transgenic mouse for human genomic sequence orthologous to *Nkx2.1*-connected region (coronal section, E11,5 mouse embryo); LacZ-stained in transgenic mouse forebrain (red arrowheads) [Visel *et al*., 2013]. **d.** *in situ* hybridization (ISH) for *Nkx2.1* in E11,5 mouse embryo (coronal section); *Nkx2.1* is expressed in forebrain, midbrain, hindbrain (red arrowheads) [Allen Brain Atlas].

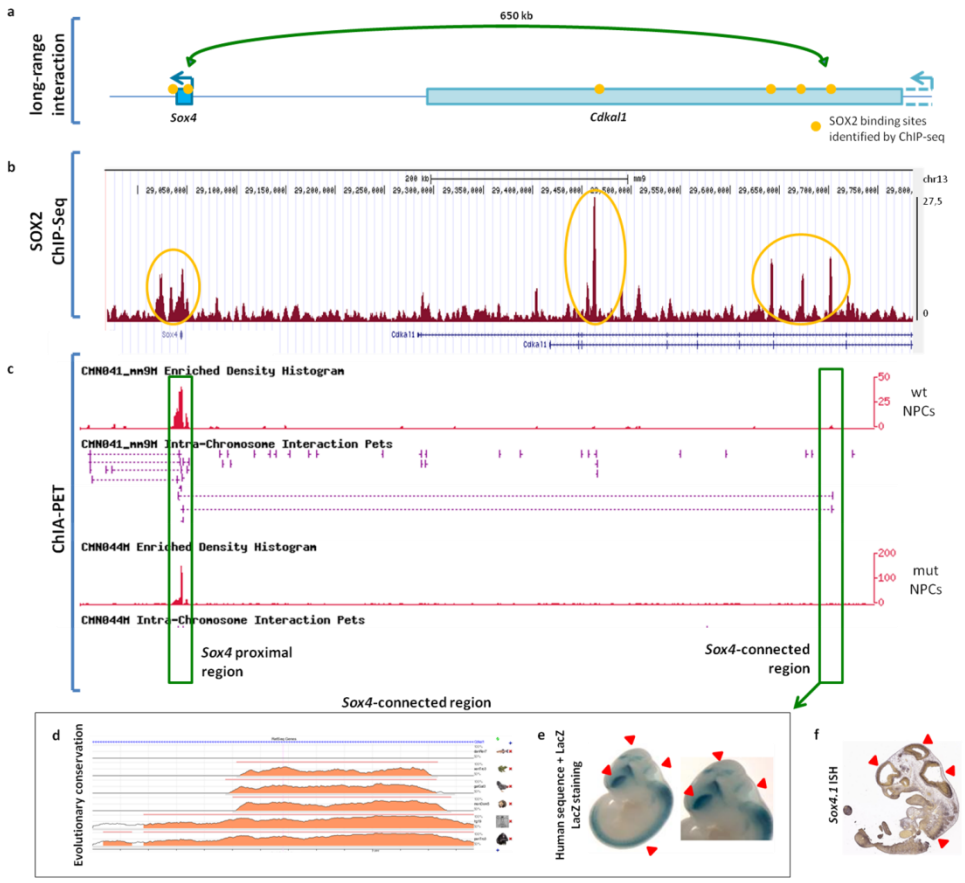


Figure 4. Analysis of the *Sox4* long-range interaction. a. Representation of SOX2-dependent long-range interaction (lost in *Sox2*-mutant NPCs) identified in wt NPCs by ChIA-PET assay and the position of the SOX2 ChIP-seq peaks on the scheme. **b.** Plot of the SOX2 ChIP-seq peaks (in orange circles) identified in the analyzed region. **c.** Plot of the long-range interactions identified in wt NPCs *vs* *Sox2*-mutant NPCs by ChIA-PET assay; in green rectangles the long-range interaction in analysis. **d.** Multispecies vertebrate conservation plot (ECR browser program) of the *Sox4*-connected region identified by p300 ChIP-seq assay made by Visel *et al.* (2013). **e.** LacZ staining of transgenic mouse for human genomic sequence orthologous to *Sox4*-connected region (coronal section, E11,5 mouse embryo); LacZ-stained in transgenic mouse forebrain, midbrain, hindbrain and neural tube (red arrowheads) [Visel *et al.*, 2013]. **f.** *in situ* hybridization (ISH) for *Sox4* in E11,5 mouse embryo (coronal section); *Sox4* is expressed in forebrain, midbrain, hindbrain and neural tube (red arrowheads) [Allen Brain Atlas].

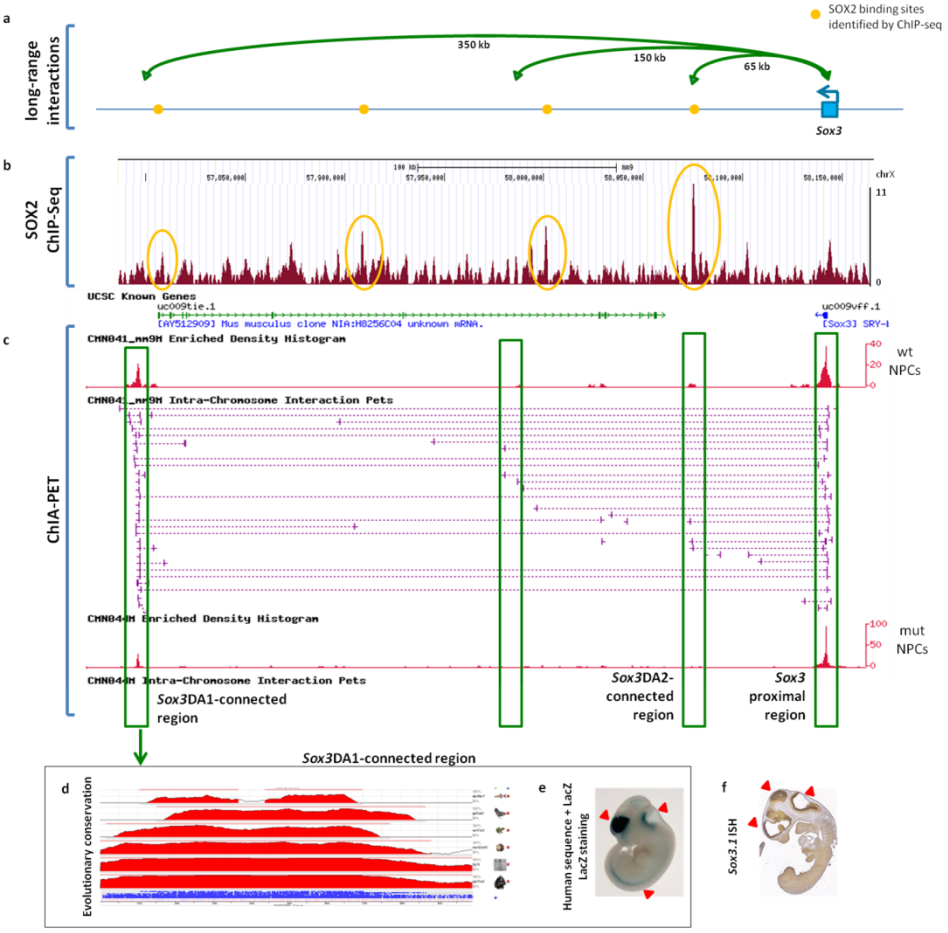


Figure 5. Analysis of the *Sox3* long-range interactions. a. Representation of SOX2-dependent long-range interaction (lost in *Sox2*-mutant NPCs) identified in wt NPCs by ChIA-PET assay and the position of the SOX2 ChIP-seq peaks on the scheme. **b.** Plot of the SOX2 ChIP-seq peaks (in orange circles) identified in the analyzed region. **c.** Plot of the long-range interactions identified in wt NPCs *vs* *Sox2*-mutant NPCs by ChIA-PET assay; in green rectangles the long-range interactions in analysis. **d.** Multispecies vertebrate conservation plot (ECR browser program) of the *Sox3*DA1-connected region identified by p300 ChIP-seq assay made by Visel *et al.* (2013). **e.** LacZ staining of transgenic mouse for human genomic sequence orthologous to *Sox3*DA1-connected region (coronal section, E11,5 mouse embryo); LacZ-stained in transgenic mouse forebrain, hindbrain and neural tube (red arrowheads) [Visel *et al.*, 2013]. **f.** *in situ* hybridization (ISH) for *Sox3* in E11,5 mouse embryo (coronal section); *Sox3* is expressed in forebrain, midbrain and hindbrain (red arrowheads) [Allen Brain Atlas].

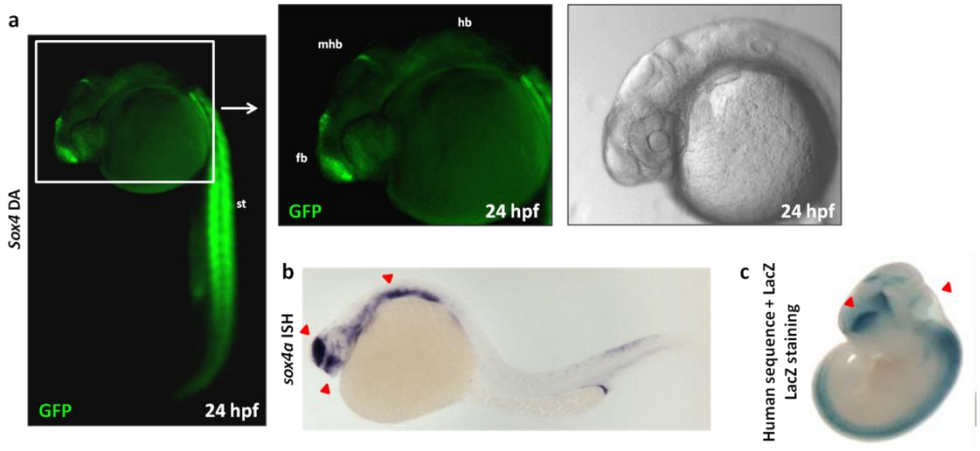


Figure 6. Expression pattern of *Sox4* DA. **a.** GFP expression pattern guided by the mouse *Sox4*-connected distal anchor (DA) in zebrafish at 24 hours post fertilization (hpf). Fb: forebrain; mhb: midbrain-hindbrain boundary; hb: hindbrain; st: somites. **b.** *in situ* hybridization (ISH) of the endogenous *sox4a* in zebrafish at 24 hpf [ZFIN, http://zfin.org/]; red arrowheads: region where GFP expression pattern in zebrafish matches the endogenous expression pattern of the gene. **c.** LacZ staining of transgenic mouse for human genomic sequence orthologous to *Sox4*-connected region (coronal section, E11,5 mouse embryo) [Visel *et al*., 2013]; red arrowheads: region where GFP expression pattern in zebrafish matches the corresponding orthologous human sequence in transgenic mouse.

Figure 7 (following page). Expression pattern of some tested DAs at 24 hpf. GFP expression pattern guided by the mouse DAs in zebrafish at 24 hours post fertilization (hpf) (first column); the corresponding images of fish with visible light (central column); *in situ* hybridization (ISH) of the zebrafish endogenous gene, associated to the mouse DA in NPCs, at 24 hpf [ZFIN, http://zfin.org/] (last column). Fb: forebrain; mhb: midbrain-hindbrain boundary; hb: hindbrain: sp: spinal cord. Red arrowheads: region where GFP expression pattern in zebrafish matches the expression pattern of the endogenous gene.

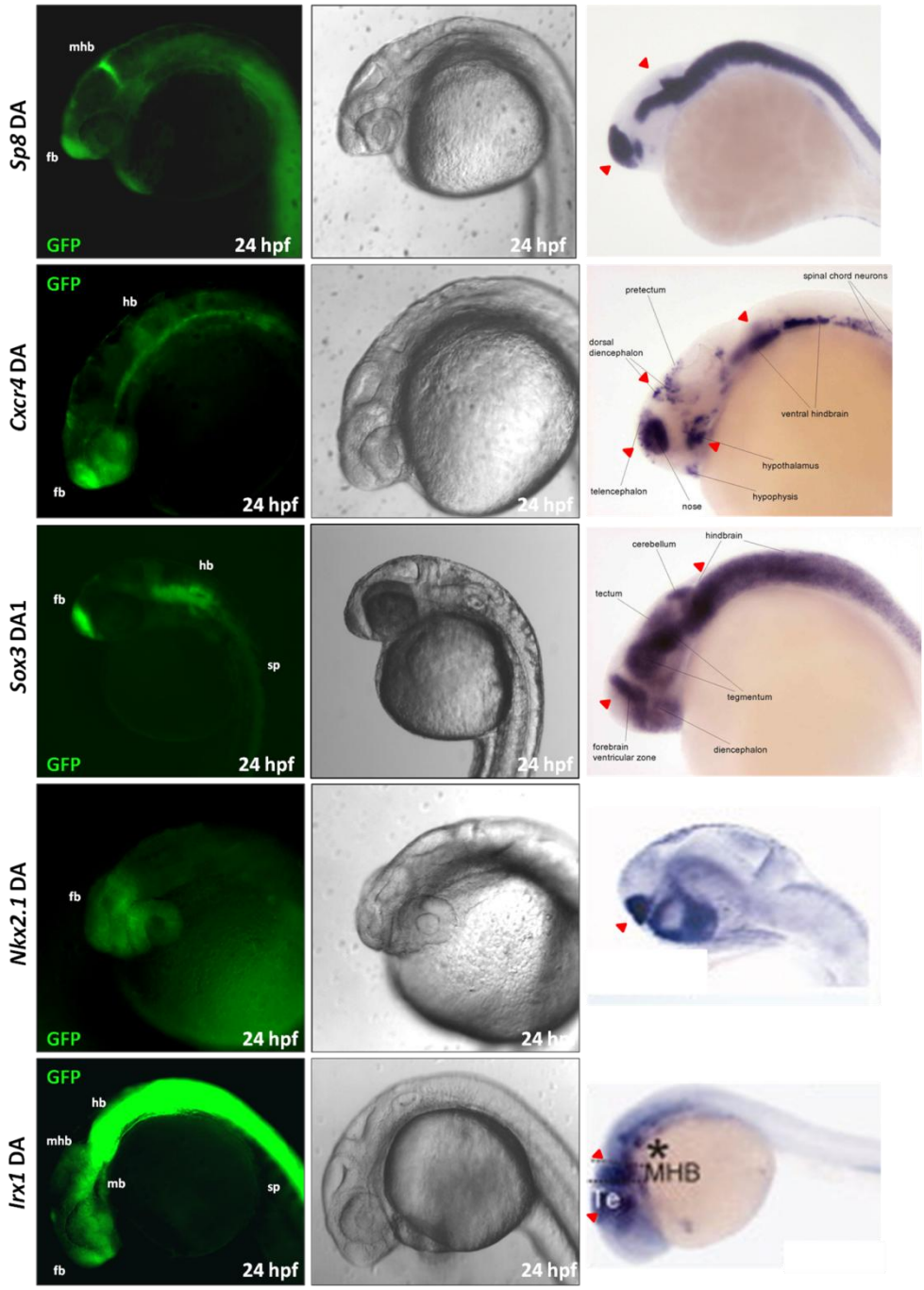
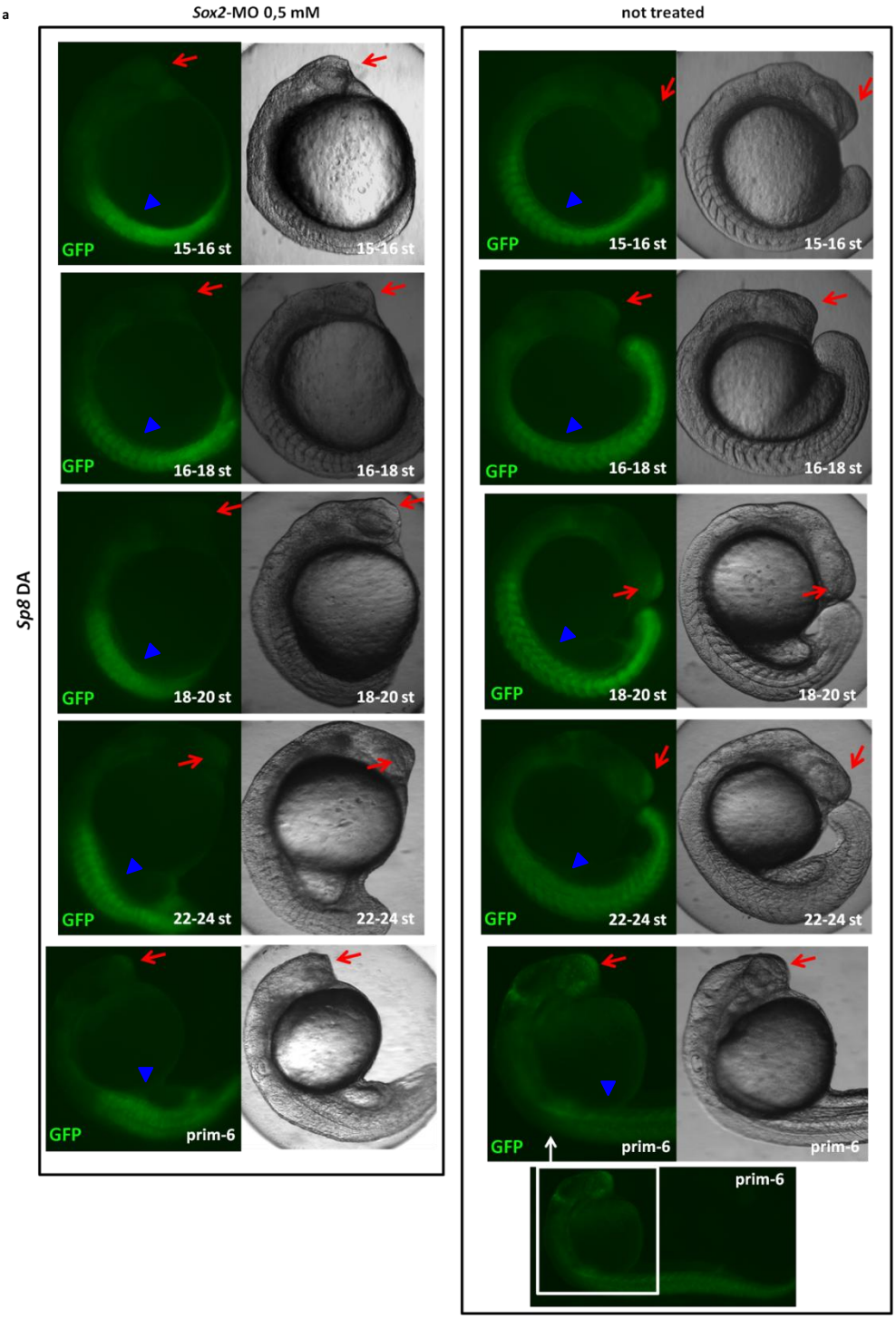
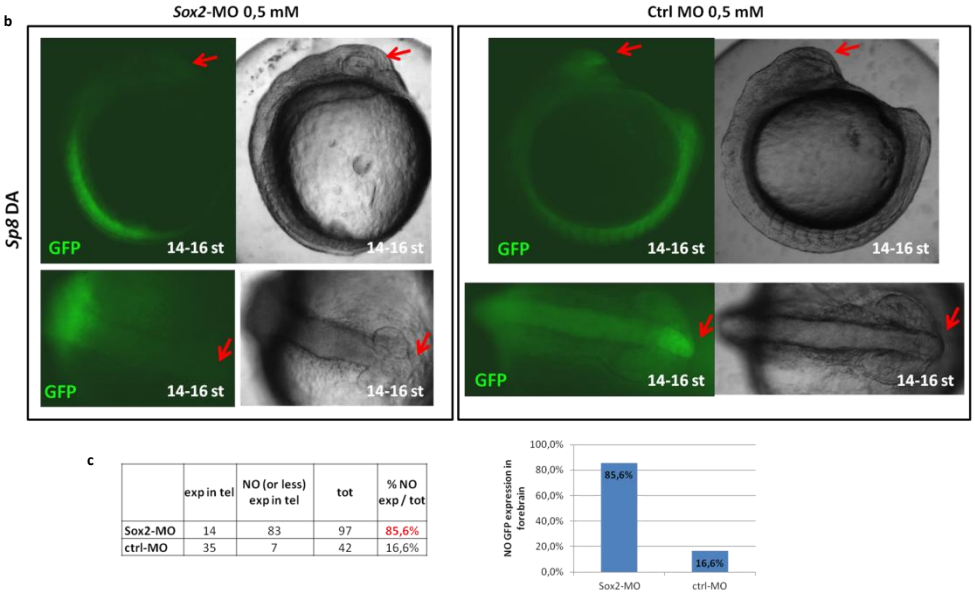


Fig. 7. (description in previous page)

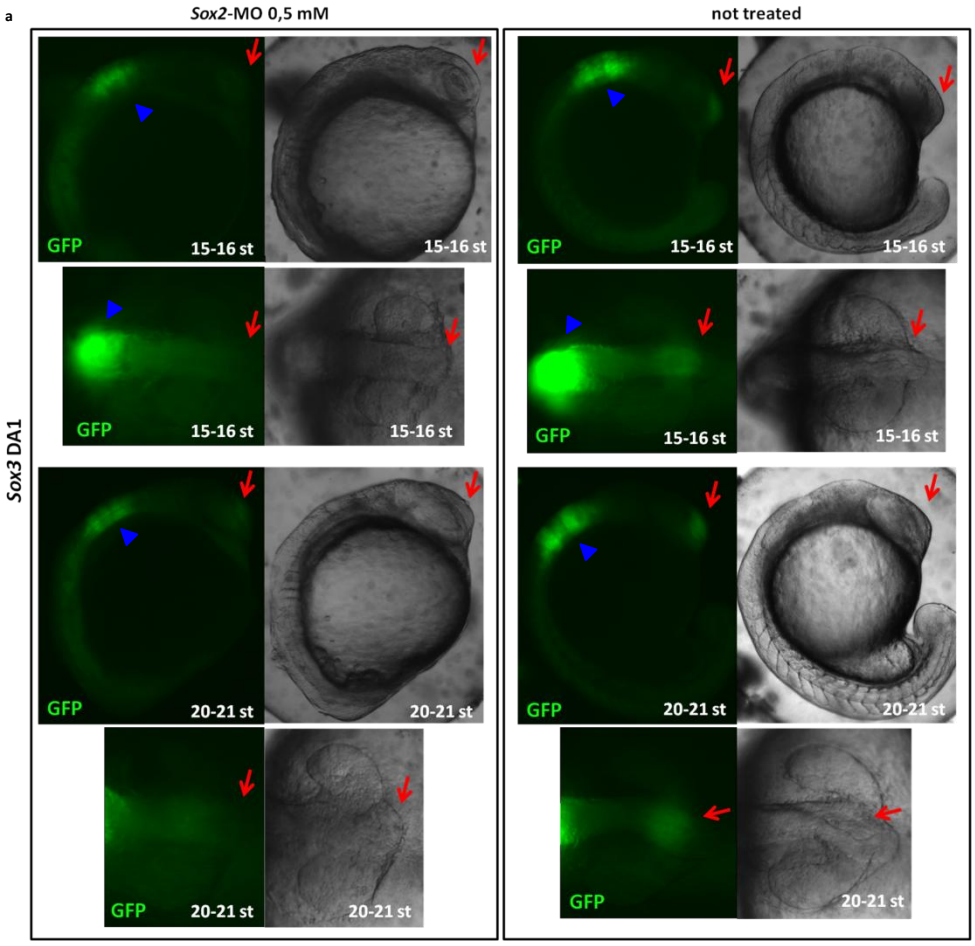

a
Sox2-MO 0,5 mM
not treated
Sp8 DA
GFP
15-16 st
16-18 st
18-20 st
22-24 st
prim-6

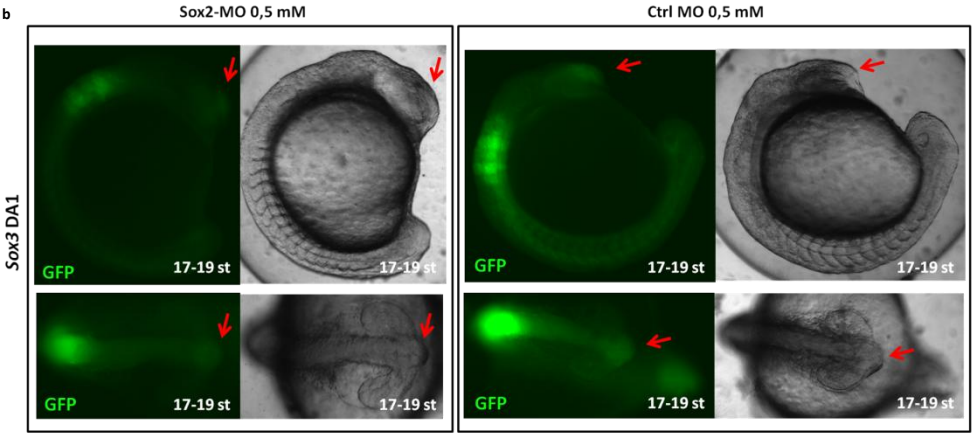


c

	exp in tel	NO (or less) exp in tel	tot	% NO exp / tot
Sox2-MO	14	83	97	85,6%
ctrl-MO	35	7	42	16,6%

Figure 8. *Sox2*-MO injection in *Sp8*DA-GFP stable line. a. *Sox2*-MO injection in *Sp8*DA-GFP stable line and comparison between injected fishes and not treated fishes during progressive stages in embryonic development. GFP expression is visible in not treated (nt) embryos (right column) from 15-16-somites (st) stage (top) to prim-6 stage (bottom); at all the developmental stages the GFP expression in forebrain is down-regulated in *Sox2*-MO treated embryos (left column). Note that GFP expression in somites are comparatively not affected (blue arrowheads); red arrows: forebrain. **b**. Comparison between injection of *Sox2*-MO and ctrl-MO at 0,5 mM, observed at 14-16-somites (st) developmental stage, in lateral and dorsal views; red arrows: forebrain. **c.** Percentage of fishes in which GFP expression is lost (or reduced) in forebrain, represented in table and histogram; results are the mean of two independent experiments.


a
Sox2-MO 0,5 mM
not treated
Sox3 DA1
GFP
15-16 st
20-21 st



c

	exp in tel	NO (or less) exp in tel	tot	% NO exp / tot
Sox2- MO	1	8	9	88,9%
ctrl MO	5	0	5	0,0%

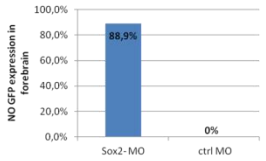


Figure 9. *Sox2*-MO injection in *Sox3*DA1-GFP stable line. a. *Sox2*-MO injection in *Sox3*DA1-GFP stable line and comparison between injected fishes and not treated fishes, in lateral and dorsal views. GFP expression is visible in not treated (nt) embryos (right column) in 15-16-somites (st) stage (top) and in 20-21 somites stage (bottom); in both stages the GFP expression in forebrain is down-regulated in *Sox2*-MO treated embryos (left column). Note that GFP expression in hindbrain are comparatively less affected (blue arrowheads); red arrows: forebrain. **b**. Comparison between injection of *Sox2*-MO and ctrl-MO at 0,5 mM, observed at 17-19-somites (st) developmental stage, in lateral and dorsal views; red arrows: forebrain. **c.** Percentage of fishes in which GFP expression is lost (or reduced) in forebrain, represented in table and histogram; results are the mean of two independent experiments.

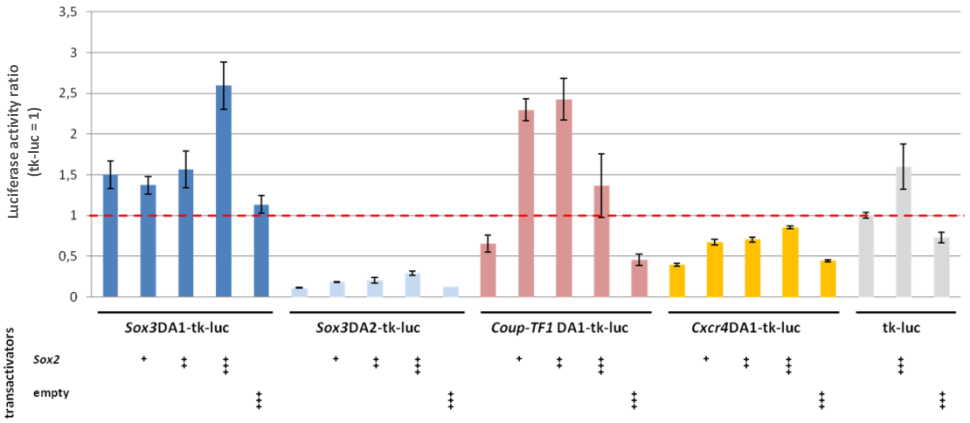


Figure 10. Luciferase assays for Sox3DA1-2, Coup-TF1 DA1 and Cxcr4DA.

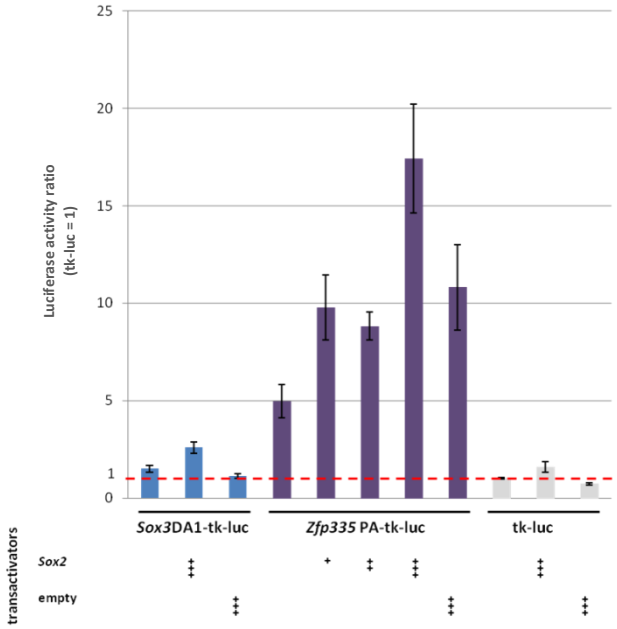


Figure 11. Luciferase assay for *Zfp335*DA.

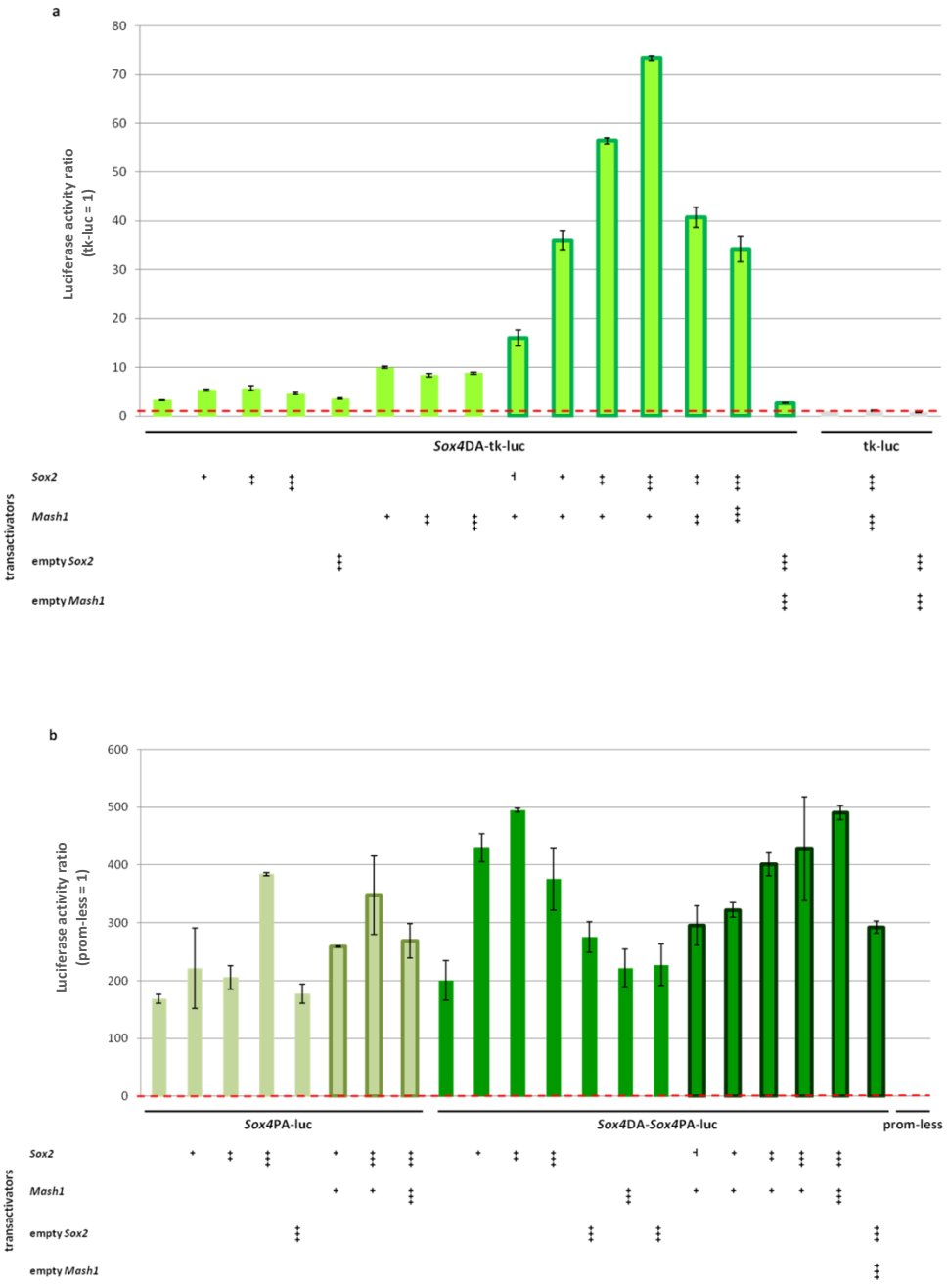


Figure 12. a. Luciferase assay for *Sox4*DA. b. Luciferase assay for *Sox4*PA and combination Sox4PA-DA.

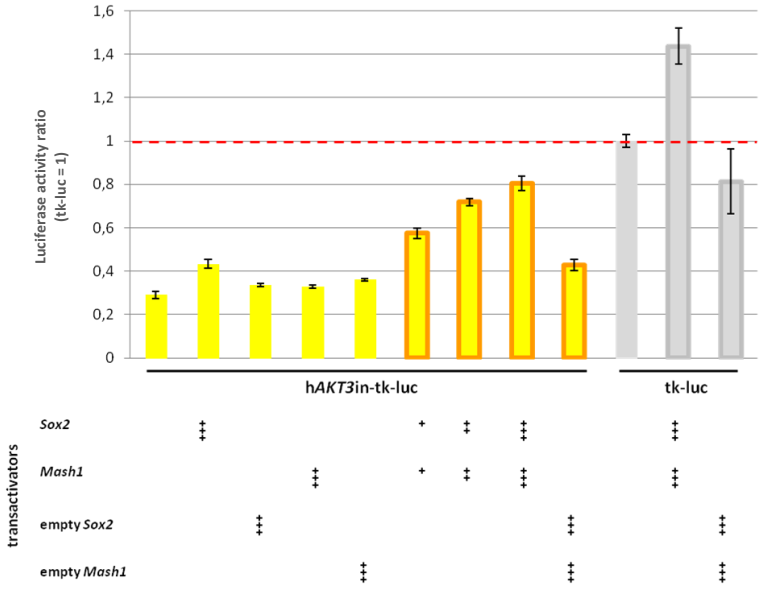


Figure 13. Luciferase assay for h*AKT3*in.

Figures 10-13. Luciferase assays for different regulatory elements. Co-transfection in P19 cells of DAs/PAs-tk-luciferase vector with increasing amounts of *Sox2* and/or *Mash1* expression vectors, or the corresponding "empty" vectors (in molar ratio compared to the luciferase vector at 1; +, 1:0,075; ++, 1:0,125; +++, 1:0,25; ++++, 1:0,5). Results are the mean of one transfection, in triplicate, for *Sox3*DA1-2, *Coup-TF1* DA1 and *Cxcr4*DA luciferase assays; at least two independent transfections, in triplicate, for *Zfp335*DA, h*AKT3*in, *Sox4*DA, *Sox4*PA and *Sox4*PA-DA. Luciferase values are normalized on Renilla luciferase values. Luciferase ratio data: fold-increase compared to the corresponding "empty" luciferase vector (tk-promoter only or promoter-less, as on the y axis of each histogram) set at 1(red line). Vertical bars on each histogram represent the standard error.

Figure 14. Analysis of long-range interactions in a region involved in deletions found in Wolf-Hirschhorn syndrome. a. Representation of SOX2-dependent long-range interactions (lost in *Sox2*-mutant NPCs) identified in wt NPCs by ChIA-PET assay and the position of the SOX2 ChIP-seq peaks on the scheme. **b.** Plot of the SOX2 ChIP-seq peaks (in orange circles) identified in the analyzed region. **c.** Plot of the long-range interactions identified in wt NPCs *vs Sox2*-mutant NPCs by ChIA-PET assay; in green rectangles the long-range interactions identified in the region.

Tables

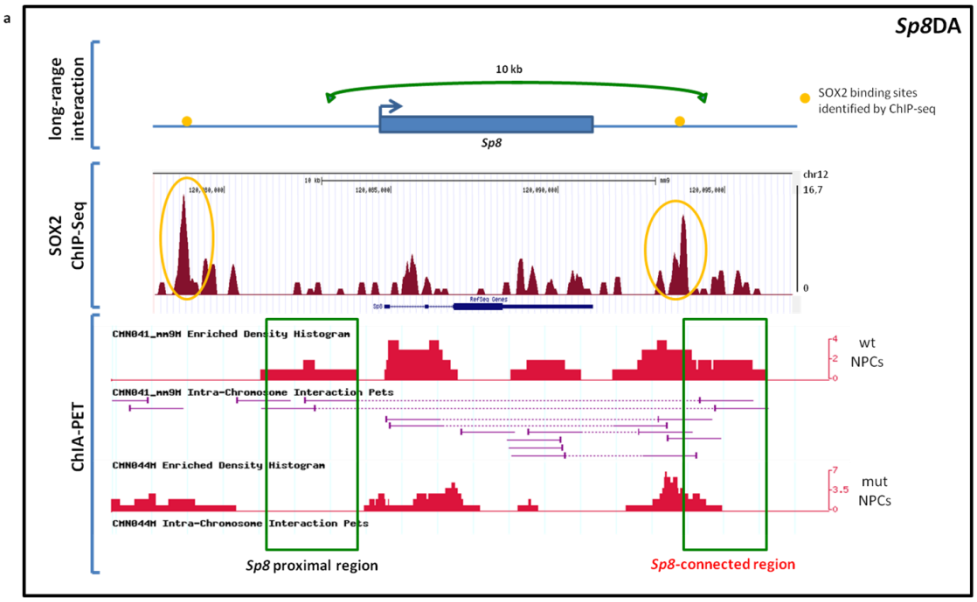
Associated gene	Cloned anchor [DA: distal anchor; PA: proximal anchor] mouse sequence (mm9)	Distance PA-DA [bp]	Presence of already defined forebrain enhancer [Visel *et al.*, 2009, 2013]		Presence of MASH1 ChIP-seq peak [Castro *et al.*, 2011]	Tested in zebrafish
			p300 binding site	validated enhancer		
Nkx2.1	DA - chr12:57621170-57622007	10.000	✓	✓	-	✓
Sp8	DA - chr12:120093525-120095900	10.000	-	-	-	✓
Coup-TF1	DA1 - chr13:78323285-78325305	12.000	-	-	-	✓
	DA2 - chr13:78301244-78303397	27.000	-	-	-	✓
Ntng1	DA - chr3:110053025-110054127	100.000	-	-	-	✓
Irx1	DA - chr13:71644069-71646540	450.000	✓	-	✓	✓
Socs3	DA - chr11:117800525-117801836	32.000	✓	-	✓	✓
Chd7	DA - chr4:8737687-8738538	40.000	-	-	-	✓
Sox3	DA1 - chrX:57797208-57798198	350.000	✓	✓	-	✓
	DA2 - chrX:58077513-58079137	65.000	-	-	✓	✓
Sox4	DA - chr13:29701468-29702829	650.000	✓	✓	✓	✓
	PA - chr13:29045461-29046297		-	-	✓	nt
Cxcr4	DA - chr1:130300526-130303029	180.000	✓	-	✓	✓
Zfp335	DA - chr2:164629282-164632033	125.000	-	-	-	-
	PA - chr2:164736357-164737128		-	-	-	-

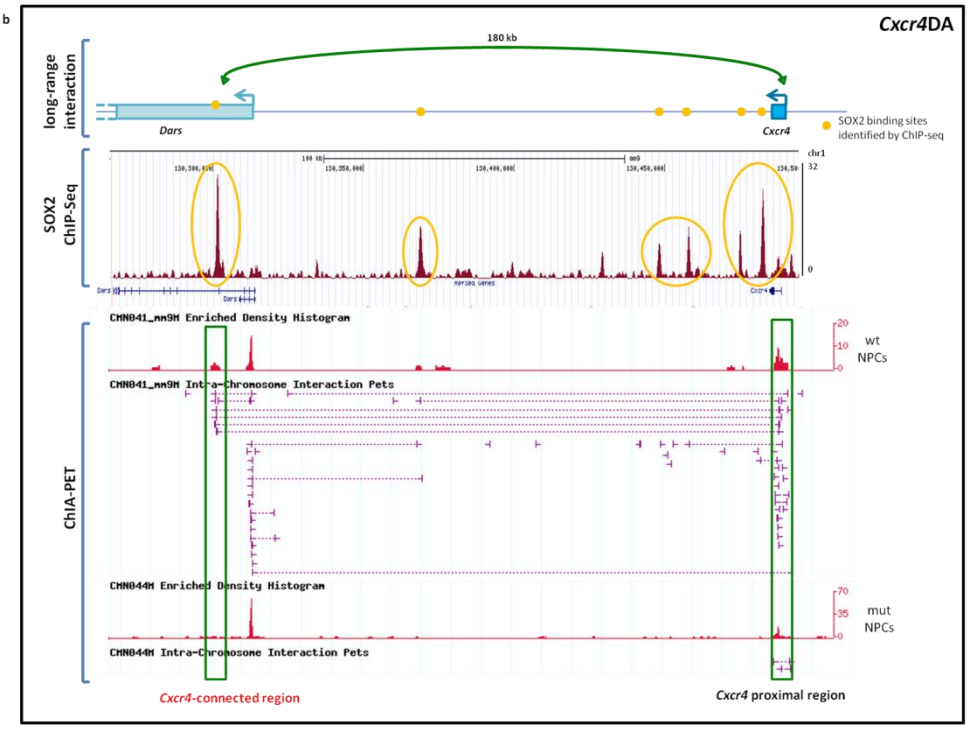
Table 1. List of the cloned anchors. Characteristic of each cloned regulatory element; nt: not tested.

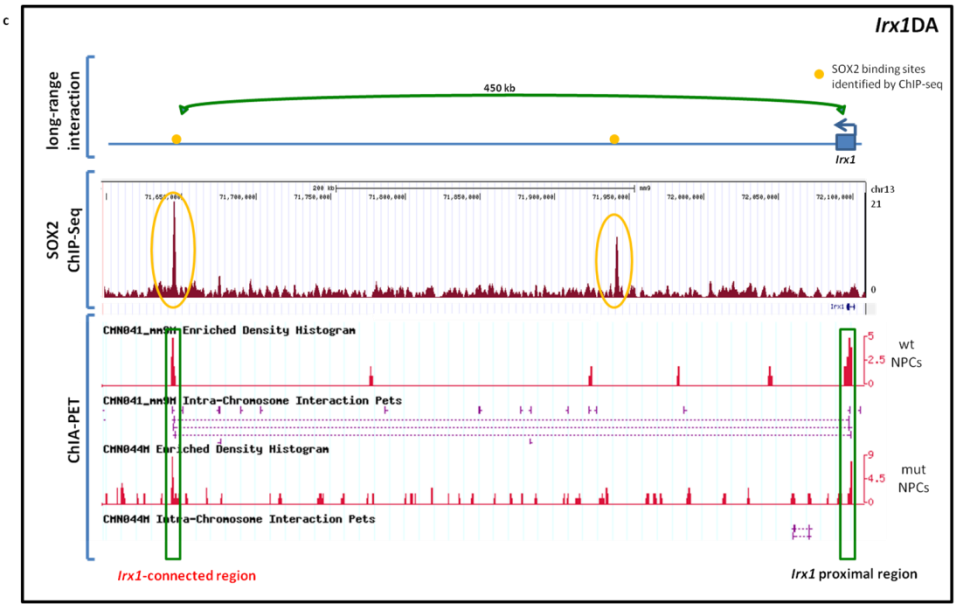
cloned anchor	primer forward 5' → 3'	primer reverse 5' → 3'
Nkx2.1 DA	TCCCGTTCCTAGTCTTTGATACTT	CGAGCAACAGGAGAGGAATAATTT
Sp8 DA	GGGGAAGAGTTCCTAGCCATT	GTGGGAGCTCAATTCATCTAA
Coup-TF1 DA1	GCTCCAGCGTCTACTGAGAAAT	AGCAGAATCCCTGAGACTTCAC
Coup-TF1 DA2	CCAGTGAAACACCTACTCACCA	AAGTTGGCATTTTTAGGACTCG
Ntng1 DA	GTAGAGGCGCGGAACCATAG	GGGGTAAAAGGAAAGGGCAAA
Irx1 DA	CAGCAAAGCATTGTAACTGTGA	TGGGGCTTTAACACAAGCAT
Socs3 DA	GCTCACACTGACCCATAGGTTT	TTGCCTCTCAGAGTGAACCA
Chd7 DA	AGGCAAGCTCACCAGCTCT	GATTTCAAAGGCAGCCACAT
Sox3 DA1	GGAGGCACATGAAAGCAATAA	GGGTAAGGTTAAAATGGCTTTTG
Sox3 DA2	ACTGTCCATTTAGTTTTCATAAATCA	GTGGGCAGGGATACCTTAGTCT
Sox4 DA	GTCCTTCAGCAAGCTCTAAACA	AATGGTGGTGAAATCTGCAAGT
Sox4 PA	ATATAGATCTAGTTCCCCCACTGCAGACT with BglII restriction site	ATATGAGCTCCTGCCTTTCACAAACAGCAC with SacI restriction site
Cxcr4 DA	GGACCCCTCAGTGAATATTAAGG	TTTGCACTGTGGTACACATTTT
Zfp335 DA	CCTAGCACTCAACCCTGAGATT	GACTTCAGAATGGAGCCAGAAC
Zfp335 PA (in ZED vector)	GGAAGTAGTTCCGGTTCGAG	GTTTGAGGCTTTCACTCTGCTG
Zfp335 PA (in luc vector)	ATATGGTACCGGAAGTAGTTCCGGTTCGAG with KpnI restriction site	ATATCTCGAGGTTTGAGGCTTTCACTCTGCTG with XhoI restriction site

Table 2. List of primers. List of the primers used for amplifying the chosen sequences from mouse DNA by PCR and cloning them in ZED or luciferase vectors.

Supplementary figures







Supplementary figure 1. Analysis of long-range interactions involving *Sp8, Cxcr4* and *Irx1* genes. Representation of SOX2-dependent long-range interactions (lost in *Sox2*-mutant NPCs) identified in wt NPCs by ChIA-PET assay and their localization on genome map; plot of the SOX2 ChIP-seq peaks (in orange circles) identified in the analyzed region; plot of the long-range interactions identified in wt NPCs *vs Sox2*-mutant NPCs by ChIA-PET assay (in green rectangles the long-range interactions in analysis), involving *Sp8* (**a**), *Cxcr4* (**b**) and *Irx1* (**c**) genes.

References

- Alatzoglou KS, Kelberman D, Dattani MT. 2009. The role of SOX proteins in normal pituitary development. *J Endocrinol.* **200**: 245-58.

- Allen Brain Atlas - data portal: http://www.brain-map.org/

- Ambrosetti DC, Basilico C, Dailey L. 1997. Synergistic activation of the fibroblast growth factor 4 enhancer by Sox2 and Oct-3 depends on protein-protein interactions facilitated by a specific spatial arrangement of factor binding sites. *Mol Cell Biol.* **17**: 6321-9.

- 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**: 126-40.

- Battaglia, A., Carey, J. C., Cederholm, P., Viskochil, D. H., Brothman, A. R., Galasso, C. 1999. Natural history of Wolf-Hirschhorn syndrome: experience with 15 cases. *Pediatrics* **103**: 830-6.

- Bergsland M, Ramsköld D, Zaouter C, Klum S, Sandberg R, Muhr J. 2011. Sequentially acting Sox transcription factors in neural lineage development. *Genes Dev.* **25**: 2453-64.

- Bergsland M, Werme M, Malewicz M, Perlmann T, Muhr J. 2006. The establishment of neuronal properties is controlled by Sox4 and Sox11. *Genes Dev.* **20**: 3475-86.

- Bessa J, Tena JJ, de la Calle-Mustienes E, Fernández-Miñán A, Naranjo S, Fernández A, Montoliu L, Akalin A, Lenhard B, Casares F, Gómez-Skarmeta JL. 2009. Zebrafish enhancer detection (ZED) vector: a new tool to facilitate transgenesis and the functional analysis of cis-regulatory regions in zebrafish. *Dev Dyn.* **238**: 2409-17.

- Boland E, Clayton-Smith J, Woo VG, McKee S, Manson FD, Medne L, Zackai E, Swanson EA, Fitzpatrick D, Millen KJ, Sherr EH, Dobyns WB, Black GC. 2007. Mapping of deletion and translocation breakpoints in 1q44 implicates the serine/threonine kinase AKT3 in postnatal microcephaly and agenesis of the corpus callosum. *Am J Hum Genet.* **81**: 292-303.

- Bosch DG, Boonstra FN, Gonzaga-Jauregui C, Xu M, de Ligt J, Jhangiani S, Wiszniewski W, Muzny DM, Yntema HG, Pfundt R, Vissers LE, Spruijt L, Blokland

EA, Chen CA; Baylor-Hopkins Center for Mendelian Genomics, Lewis RA, Tsai SY, Gibbs RA, Tsai MJ, Lupski JR, Zoghbi HY, Cremers FP, de Vries BB, Schaaf CP. 2014. NR2F1 mutations cause optic atrophy with intellectual disability. *Am J Hum Genet.* **94**: 303-9.

- Castro DS, Martynoga B, Parras C, Ramesh V, Pacary E, Johnston C, Drechsel D, Lebel-Potter M, Garcia LG, Hunt C, Dolle D, Bithell A, Ettwiller L, Buckley N, Guillemot F. 2011. A novel function of the proneural factor Ascl1 in progenitor proliferation identified by genome-wide characterization of its targets. *Genes Dev.* **25**: 930-45.

- Cheung M, Abu-Elmagd M, Clevers H, Scotting PJ. 2000. Roles of Sox4 in central nervous system development. *Brain Res Mol Brain Res.* **79**: 180-91.

- Engelen E, Akinci U, Bryne JC, Hou J, Gontan C, Moen M, Szumska D, Kockx C, van Ijcken W, Dekkers DH, Demmers J, Rijkers EJ, Bhattacharya S, Philipsen S, Pevny LH, Grosveld FG, Rottier RJ, Lenhard B, Poot RA. 2011. Sox2 cooperates with Chd7 to regulate genes that are mutated in human syndromes. *Nat Genet.* **43**: 607-11.

- Fantes J, Ragge NK, Lynch SA, McGill NI, Collin JR, Howard-Peebles PN, Hayward C, Vivian AJ, Williamson K, van Heyningen V, FitzPatrick DR. 2003. Mutations in SOX2 cause anophthalmia. *Nat Genet.* **33**: 461–3.

- Favaro R, Valotta M, Ferri AL, Latorre E, Mariani J, Giachino C, Lancini C, Tosetti V, Ottolenghi S, Taylor V, Nicolis SK. 2009. Hippocampal development and neural stem cell maintenance require Sox2-dependent regulation of Shh. *Nat Neurosci.* **12**: 1248-56.

- Ferri A, Favaro R, Beccari L, Bertolini J, Mercurio S, Nieto-Lopez F, Verzeroli C, La Regina F, De Pietri Tonelli D, Ottolenghi S, Bovolenta P, Nicolis SK. 2013. Sox2 is required for embryonic development of the ventral telencephalon through the activation of the ventral determinants Nkx2.1 and Shh. *Development* **140**: 1250-61.

- Ferri AL, Cavallaro M, Braida D, Di Cristofano A, Canta A, Vezzani A, Ottolenghi S, Pandolfi PP, Sala M, De Biasi S, Nicolis SK. 2004. Sox2 deficiency causes neurodegeneration and impaired neurogenesis in the adult mouse brain. *Development* **131**: 3805-19.

- Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, Mohamed YB, Orlov YL, Velkov S, Ho A, Mei PH, Chew EG, Huang PY, Welboren WJ, Han Y, Ooi HS, Ariyaratne PN, Vega VB, Luo Y, Tan PY, Choy PY, Wansa KD, Zhao B, Lim KS, Leow SC, Yow JS, Joseph R,

Li H, Desai KV, Thomsen JS, Lee YK, Karuturi RK, Herve T, Bourque G, Stunnenberg HG, Ruan X, Cacheux-Rataboul V, Sung WK, Liu ET, Wei CL, Cheung E, Ruan Y. 2009. An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* **462**: 58-64.

- Inoue K, Khajavi M, Ohyama T, Hirabayashi S, Wilson J, Reggin JD, Mancias P, Butler IJ, Wilkinson MF, Wegner M, Lupski JR. 2004. Molecular mechanism for distinct neurological phenotypes conveyed by allelic truncating mutations. *Nat Genet.* **36**: 361-9.

- Jeong Y, Leskow FC, El-Jaick K, Roessler E, Muenke M, Yocum A, Dubourg C, Li X, Geng X, Oliver G, Epstein DJ. 2008. Regulation of a remote Shh forebrain enhancer by the Six3 homeoprotein. *Nat Genet.* **40**: 1348-53.

- Jin Z, Liu L, Bian W, Chen Y, Xu G, Cheng L, Jing N. 2009. Different transcription factors regulate nestin gene expression during P19 cell neural differentiation and central nervous system development. *J Biol Chem.* **284**: 8160-73.

- Kamachi Y, Okuda Y, Kondoh H. 2008. Quantitative assessment of the knockdown efficiency of morpholino antisense oligonucleotides in zebrafish embryos using a luciferase assay. *Genesis* **46**: 1-7.

- Kelberman D, Rizzoti K, Avilion A, Bitner-Glindzicz M, Cianfarani S, Collins J, Chong WK, Kirk JM, Achermann JC, Ross R, Carmignac D, Lovell-Badge R, Robinson IC, Dattani MT. 2006. Mutations within Sox2/SOX2 are associated with abnormalities in the hypothalamo-pituitary-gonadal axis in mice and humans. *J Clin Invest.* **116**: 2442–55.

- Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, Poh HM, Goh Y, Lim J, Zhang J, Sim HS, Peh SQ, Mulawadi FH, Ong CT, Orlov YL, Hong S, Zhang Z, Landt S, Raha D, Euskirchen G, Wei CL, Ge W, Wang H, Davis C, Fisher-Aylor KI, Mortazavi A, Gerstein M, Gingeras T, Wold B, Sun Y, Fullwood MJ, Cheung E, Liu E, Sung WK, Snyder M, Ruan Y. 2012. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148**: 84-98.

- Li G, Fullwood MJ, Xu H, Mulawadi FH, Velkov S, Vega V, Ariyaratne PN, Mohamed YB, Ooi HS, Tennakoon C, Wei CL, Ruan Y, Sung WK. 2010. ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. *Genome Biol.* **11**: R22.

- Lodato MA, Ng CW, Wamstad JA, Cheng AW, Thai KK, Fraenkel E, Jaenisch R, Boyer LA. 2013. SOX2 co-occupies distal enhancer elements with distinct POU factors in ESCs and NPCs to specify cell state. *PLoS Genet.* **9**: e1003288.

- Nishimoto M, Fukushima A, Okuda A, Muramatsu M. 1999. The gene for the embryonic stem cell coactivator UTF1 carries a regulatory element which selectively interacts with a complex composed of Oct-3/4 and Sox-2. *Mol Cell Biol.* **19**: 5453-65.

- Pevny LH, Sockanathan S, Placzek M, Lovell-Badge R. 1998. A role for SOX1 in neural determination. D*evelopment* **125**: 1967-78.

- Ragge NK, Lorenz B, Schneider A, Bushby K, de Sanctis L, de Sanctis U, Salt A, Collin JR, Vivian AJ, Free SL, Thompson P, Williamson KA, Sisodiya SM, van Heyningen V, Fitzpatrick DR. 2005. SOX2 anophthalmia syndrome. *Am J Med Genet A.* **135**: 1–7.

- Riviere JB, Mirzaa GM, O'Roak BJ, Beddaoui M, Alcantara D, Conway RL, St-Onge J, Schwartzentruber JA, Gripp KW, Nikkel SM, Worthylake T, Sullivan CT, and 29 others. 2012. De novo germline and postzygotic mutations in AKT3, PIK3R2 and PIK3CA cause a spectrum of related megalencephaly syndromes. *Nature Genet.* **44**: 934-40.

- Rizzoti K, Brunelli S, Carmignac D, Thomas PQ, Robinson IC, Lovell-Badge R. 2004. SOX3 is required during the formation of the hypothalamo-pituitary axis. *Nat Genet.* **36**: 247-55.

- Rodda DJ, Chew JL, Lim LH, Loh YH, Wang B, Ng HH, Robson P. 2005. Transcriptional regulation of nanog by OCT4 and SOX2. *J Biol Chem.* **280**: 24731-7.

- Schneider A, Bardakjian T, Reis LM, Tyler RC, Semina EV. 2009. Novel SOX2 mutations and genotype-phenotype correlation in anophthalmia and microphthalmia. *Am J Med Genet A.* **149A**: 2706–15.

- Sisodiya SM, Ragge NK, Cavalleri GL, Hever A, Lorenz B, Schneider A, Williamson KA, Stevens JM, Free SL, Thompson PJ, van Heyningen V, Fitzpatrick DR. 2006. Role of SOX2 mutations in human hippocampal malformations and epilepsy. *Epilepsia* **47**: 534–42.

- Visel A, Taher L, Girgis H, May D, Golonzhka O, Hoch RV, McKinsey GL, Pattabiraman K, Silberberg SN, Blow MJ, Hansen DV, Nord AS, Akiyama JA, Holt A, Hosseini R, Phouanenavong S, Plajzer-Frick I, Shoukry M, Afzal V, Kaplan T,

Kriegstein AR, Rubin EM, Ovcharenko I, Pennacchio LA, Rubenstein JL. 2013. A high-resolution enhancer atlas of the developing telencephalon. *Cell* **152**: 895-908.

- Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, Afzal V, Ren B, Rubin EM, Pennacchio LA. 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**: 854-8.

- ZFIN – the Zebrafish Model Organism Database: http://zfin.org/

- Zhang Y, Wong CH, Birnbaum RY, Li G, Favaro R, Ngan CY, Lim J, Tai E, Poh HM, Wong E, Mulawadi FH, Sung WK, Nicolis S, Ahituv N, Ruan Y, Wei CL. 2013. Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. *Nature* **504**: 306-10.

CHAPTER 3:

Development 140, 1250-1261 (2013)

Sox2 is required for embryonic development of the ventral telencephalon through the activation of the ventral determinants *Nkx2.1* and *Shh*

Anna Ferri[1], Rebecca Favaro*[1], Leonardo Beccari*[2,3], Jessica Bertolini[1], Sara Mercurio[1], Francisco Nieto-Lopez[2,3], Cristina Verzeroli[1], Federico La Regina[4], Davide De Pietri Tonelli[5], Sergio Ottolenghi[1], Paola Bovolenta[2,3], Silvia K. Nicolis[1#]

[1]Dipartimento di Biotecnologie e Bioscienze, Università di Milano-Bicocca, p.zza della Scienza 2, 20126 Milan, Italy
[2]Centro de Biología Molecular Severo Ochoa. CSIC-UAM and [3] CIBER de Enfermedades Raras (CIBERER), c/ Nicolás Cabrera, 1 28049 Cantoblanco, Madrid, Spain
[4]European Brain Research Institute (EBRI Rita Levi-Montalcini), via del Fosso Fiorano, 64 Rome, Italy
[5]Italian Institute of Technology, via Morego 30, 16163 Genova, Italy

*equally contributing authors
[#] corresponding author

Development 140, 1250-1261 (2013) doi:10.1242/dev.073411

Sox2 is required for embryonic development of the ventral telencephalon through the activation of the ventral determinants Nkx2.1 and Shh

Anna Ferri[1], Rebecca Favaro[1,*], Leonardo Beccari[2,3,*], Jessica Bertolini[1], Sara Mercurio[1], Francisco Nieto-Lopez[2,3], Cristina Verzeroli[1], Federico La Regina[4], Davide De Pietri Tonelli[5], Sergio Ottolenghi[1], Paola Bovolenta[2,3] and Silvia K. Nicolis[1,‡]

SUMMARY

The Sox2 transcription factor is active in stem/progenitor cells throughout the developing vertebrate central nervous system. However, its conditional deletion at E12.5 in mouse causes few brain developmental problems, with the exception of the postnatal loss of the hippocampal radial glia stem cells and the dentate gyrus. We deleted Sox2 at E9.5 in the telencephalon, using a Bf1-Cre transgene. We observed embryonic brain defects that were particularly severe in the ventral, as opposed to the dorsal, telencephalon. Important tissue loss, including the medial ganglionic eminence (MGE), was detected at E12.5, causing the subsequent impairment of MGE-derived neurons. The defect was preceded by loss of expression of the essential ventral determinants Nkx2.1 and Shh, and accompanied by ventral spread of dorsal markers. This phenotype is reminiscent of that of mice mutant for the transcription factor Nkx2.1 or for the Shh receptor Smo. Nkx2.1 is known to mediate the initial activation of ventral telencephalic Shh expression. A partial rescue of the normal phenotype at E14.5 was obtained by administration of a Shh agonist. Experiments in Medaka fish indicate that expression of Nkx2.1 is regulated by Sox2 in this species also. We propose that Sox2 contributes to Nkx2.1 expression in early mouse development, thus participating in the region-specific activation of Shh, thereby mediating ventral telencephalic patterning induction.

KEY WORDS: Brain development, Sox2, Ventral telencephalon, Mouse, Neurogenesis, Sonic hedgehog, Nkx2.1

INTRODUCTION

The transcription factor Sox2 is necessary for the maintenance of pluripotency in epiblast and embryonic stem cells; its knockout is early embryonic lethal (Avilion et al., 2003; Masui et al., 2007). Later in development, Sox2 is required in various tissue stem cells and early progenitors, in particular in the nervous system (Que et al., 2009; Basu-Roy et al., 2010; Pevny and Nicolis, 2010). Throughout vertebrate evolution, Sox2 is expressed in the developing neuroectoderm from its earliest stages (Wegner and Stolt, 2005). In the embryonic nervous system, Sox2 marks undifferentiated neural precursor cells, including neural stem cells (NSCs). Postnatally, Sox2 is expressed in NSCs within the neurogenic niches of the subventricular zone (SVZ) and hippocampus dentate gyrus (DG) (Zappone et al., 2000; Ellis et al., 2004; Ferri et al., 2004; Suh et al., 2007). Sox2 is also expressed in some differentiating neural cells and neurons (Ferri et al., 2004; Taranova et al., 2006; Cavallaro et al., 2008).

Interestingly, heterozygous Sox2 mutations in humans cause a characteristic spectrum of CNS abnormalities, including eye, hippocampus, hypothalamus and basal ganglia defects, with neurological pathology including epilepsy and motor control problems (Fantes et al., 2003; Kelberman et al., 2008; Sisodiya et al., 2006).

Sox2 gain-of-function and dominant-negative experiments established roles for Sox2 in the maintenance of NSC/progenitor cells in chicken and frog (Kishi et al., 2000; Bylund et al., 2003; Graham et al., 2003). Moreover, neonatal and embryonic NSCs grown *in vitro* from mice with a nestin-Cre-driven conditional ablation of Sox2 in the neural tube at embryonic day of development (E) 12.5 became prematurely exhausted in long-term culture experiments (Favaro et al., 2009).

Despite the severe *in vitro* defects of NSC maintenance, *in vivo* embryonic brain abnormalities in Sox2-nestin-Cre mutants are rather limited (Miyagi et al., 2008; Favaro et al., 2009); the only prominent defect is early postnatal failure to maintain hippocampal NSCs (radial glia) and neurogenesis, followed by loss of the hippocampal dentate gyrus. These defects were preceded by embryonic-perinatal loss of sonic hedgehog (Shh) expression in the telencephalon (but not in midbrain and in spinal cord), and could be rescued by a chemical Shh agonist (Favaro et al., 2009).

The reasons for the limited effects of Sox2 deletion on brain development remain unclear. Other Sox proteins, such as Sox1 and Sox3, which play roles similar to those of Sox2 (Bylund et al., 2003; Graham et al., 2003), might compensate *in vivo* for Sox2 absence. Alternatively, the timing of embryonic Sox2 deletion in previous experiments (Favaro et al., 2009) might have been too late, thus failing to uncover essential earlier functions of Sox2.

Here, we have used an early-acting Bf1 (Foxg1)-Cre transgene, which completely ablated Sox2 by E9.5 in the developing telencephalon, two days earlier than the deletion with nestin-Cre (Miyagi et al., 2008; Favaro et al., 2009). This caused defects much more severe than those observed with nestin-Cre (Miyagi et al.,

[1]Dipartimento di Biotecnologie e Bioscienze, Università di Milano-Bicocca, piazza della Scienza 2, 20126 Milan, Italy. [2]Centro de Biologia Molecular Severo Ochoa, CSIC-UAM and [3]CIBER de Enfermedades Raras (CIBERER), c/Nicolás Cabrera, 1 28049 Cantoblanco, Madrid, Spain. [4]European Brain Research Institute (EBRI Rita Levi-Montalcini), via del Fosso Fiorano, 64 Rome, Italy. [5]Italian Institute of Technology, via Morego 30, 16163 Genova, Italy.

*These authors contributed equally to this work
‡Author for correspondence (silvia.nicolis@unimib.it)

Accepted 3 January 2013

2008; Favaro et al., 2009). Unexpectedly, these defects were markedly region specific, with much more pronounced ventral than dorsal telencephalic alterations. The medial ganglionic eminence (MGE) was completely lost at E12.5, preceded by an earlier failure to express the ventral determinants Nkx2.1 (Nkx2-1) and Shh. Treatment with a Shh agonist (Shh-ag) *in vivo* was sufficient to rescue the ventral (MGE) phenotype to a significant, but not complete, extent. Furthermore, we show that Sox2 regulates Nkx2.1, a known direct activator of Shh (Jeong et al., 2006).

MATERIALS AND METHODS

Mouse strains

Sox2$^{flox/+}$ mice (Favaro et al., 2009) were bred to *Bf1-Cre* mice (Hébert and McConnell, 2000) to obtain compound *Sox2*$^{flox/+}$ *Bf1-Cre* heterozygotes, which were bred to *Sox2*$^{flox/flox}$ mice to generate Sox2-deleted embryos. *Bf1-Cre* mice were maintained by brother-sister mating, and subsequently on a 129 background (Hébert and McConnell, 2000).

Histology, *in situ* hybridisation (ISH), immunohistochemistry and Shh agonist treatment

Histology, ISH and immunohistochemistry were carried out as previously described (Ferri et al., 2004; Favaro et al., 2009). Antibodies used were: anti-SOX2, anti-SOX1, anti-SOX3, anti-SOX9 mouse monoclonals (R&D Systems); anti-Nkx2.1 rabbit polyclonal (BIOPAT); anti-SHH rabbit polyclonal (Santa Cruz); and anti-SHH mouse monoclonal [Developmental Studies Hybridoma Bank (DSHB)]. BrdU (Sigma B5002, 15 mg/ml in PBS) was administered to pregnant females at 6 µl/g body weight; females were sacrificed after 30 minutes. BrdU immunofluorescence and TUNEL analysis were carried out as described by Favaro et al. (Favaro et al., 2009) and Ferri et al. (Ferri et al., 2004), respectively.

Shh agonist #1.2 (Frank-Kamenetsky et al., 2002) was administered to pregnant females at E8.5 and E10.5, by oral gavage of a 1.5 mg/ml solution in 0.5% methylcellulose/0.2% Tween 80 at 100 µl/g body weight.

Mosaic deletion of *Sox2* by Sox2CreERT2 was by tamoxifen administration at E8.5 by oral gavage of a 20 mg/ml solution in 1:10 ethanol/corn oil, 0.1 mg/g body weight (Favaro et al., 2009).

Nkx2.1 regulation studies

Transgenic constructs

The genomic sequence spanning nucleotides −495 to +1842 relative to the mouse upstream *Nkx2.1* transcription start site was PCR amplified (primers: forward: 5′-GAGTAGAGAGCACTCTTCAAGGAG-3′; reverse: 5′-GGCGTCGGCTGGAGGAGGAAGGAAG-3′) and cloned into the vector IsceI-EGFP (Conte and Bovolenta, 2007) generating mNkx2.1 wt long:EGFP. The Sox2 consensus sites were mutated using the Multisite Quickchange Lightening Kit (Strataclone).

Luciferase constructs

Appropriate fragments were amplified by PCR (with primers: forward: 5′-ATCTCGAGCCGACCAAATTGGACCGCGG-3′, added *Xho*I site underlined; reverse: 5′-GCGAGATCTTGCCAAATATTCTGGTGTT-ACCTTAACG-3′, added *Bgl*II site underlined) and cloned upstream to the luciferase gene into the TK-LUC vector (provided by A. Okuda, Saitama Medical School, Saitama, Japan) previously deleted of the TK minimal promoter.

Chromatin immunoprecipitation (ChIP)

ChIP was performed using stage 16-18 Medaka fish (*Oryzia latipes*) embryos. Chromatin was immunoprecipitated with 2 µg of anti-Sox2 (R&D Systems) or a non-related IgG (Sigma). DNA was analysed by Q-PCR (Roche). Fold-enrichment was expressed as the ratio of Sox2 to IgG signal. Q-PCR of the 18S rRNA region and the 3′ UTR of the *Nkx2.1* gene, lacking *Sox2*-binding consensuses (negative controls), and of the *Nkx2.1* promoter/enhancer, were performed using the following specific primers: 18S Forward: 5′-GGTAACCCGCTGAACCCCAC-3′; 18S Reverse: 5′-CCATCCAATCGGTAGTAGCG-3′; Nkx2.1-3′UTR Forward: 5′GCCC-TACAGGTTCAGTCCAG-3′; Nkx2.1-3′UTR Reverse: 5′ACTGGG-ACTGGGGTTCTTTT-3′; Nkx2.1enhancer Forward: 5′-CAATTAAG-GCGGACTTGAGG-3′; Nkx2.1enhancer Reverse: 5′-AGAAGGCA-AGGCAATCTCTC-3′.

Transfection experiments

P19 cells (2×10^5/well) were plated in 6-well plates and transfected after 24 hours in 1 ml of Opti-MEM (Invitrogen) with Lipofectamine 2000 (Invitrogen) with 1 µg luciferase plasmid (Nkx2.1-luciferase, or 'empty'-luciferase), and increasing amounts of *Sox2* expression vector (Favaro et al., 2009). In control experiments, equimolar amounts of *Sox2* 'empty' vector were used. pBluescript was added to each transfection to equalise total DNA to 2 µg. Luciferase activity was measured after 24 hours. For transgenesis experiments in Medaka, plasmids purified using the Genopure Plasmid Midi Kit (Roche) were injected at the one-cell stage into Medaka oocytes CAB strain, at 15 ng/µl (Conte and Bovolenta, 2007). Embryos were analysed for EGFP expression (by fluorescence and confocal microscopy) in the hypothalamus at stage 19. To determine whether Sox2 regulates reporter expression, Nkx2.1 *wt-long*-EGFP was co-injected with *Sox2* mRNA or a Sox2-specific, already validated morpholino (MO) (Beccari et al., 2012). ISH was as described (Conte and Bovolenta, 2007) using probes against Medaka *Nkx2.1*, *Arx* and *Dmbx* (*Arx* and *Dmbx* representing diencephalic and mesencephalic markers, respectively). Subsequently, three independent stable transgenic lines were selected.

In utero electroporation

E13.5 C57/Bl6 pregnant mice were anesthetised and DNA introduced by electroporation *in utero* as described (Sanchez-Camacho and Bovolenta, 2008) using a solution containing a 1:1 mixture of Nkx2.1 *wt-long*::EGFP and pCAG-Cherry (2µg/µl). Embryos were collected and analysed after 48 hours (E15.5) by sectioning the brains in 50-µm-thick frontal sections. GFP expression was enhanced by immunostaining with rabbit anti-GFP (1:1000, Molecular Probes).

RESULTS

Sox2 early deletion severely impairs embryonic brain development

To ablate Sox2 in the early embryonic brain, we bred mice carrying a *Sox2*^{flox} conditional mutation (Favaro et al., 2009) to mice expressing the Cre-recombinase gene under the control of the *Bf1* regulatory regions, specifically active in the developing telencephalon from embryonic day (E) 9.5 of development (Bf1cre 'knock-in') (Hébert and McConnell, 2000). In *Sox2*$^{flox/flox}$*;Bf1cre* embryos, Sox2 protein was completely ablated by E9.5 in the telencephalon, though not in more posterior neural tube regions, as expected (Fig. 1A). This caused early morphological defects: at E12.5, telencephalic vesicles were reduced and the eyes were abnormal (Fig. 1B,C). Interestingly, although the whole telencephalon was affected, the ventral part was much more severely compromised than the dorsal one (Fig. 1C,F); histological sections (Fig. 1F) showed that the ventral primordia of the medial ganglionic eminence (MGE), involved in the generation of the basal ganglia (Sur and Rubenstein, 2005; Hébert and Fishell, 2008), were severely reduced (Fig. 1F, arrowhead). These initial defects developed into profoundly abnormal development, leading to death just after birth. At E18.5, mutant pups had a smaller head (Fig. 1E) and the telencephalon was smaller than in wild type (Fig. 1D,G: compare with the almost unaffected midbrain); also, the olfactory bulbs and the midline ventral structures were absent (Fig. 1D, black arrowhead pointing to ventral 'hole'). In tissue sections, the ventral midline and the immediately adjacent territories were missing (Fig. 1G, arrowheads).

In agreement with the early MGE abnormalities, GABAergic cortical interneurons, which originate in the MGE and then migrate to more dorsal locations (Sur and Rubenstein, 2005; Hébert and Fishell, 2008; Elias et al., 2008), were strongly decreased in mutants, as indicated by the almost complete loss of somatostatin (SS)-positive and the strong reduction of the neuropeptide Y (NPY)-

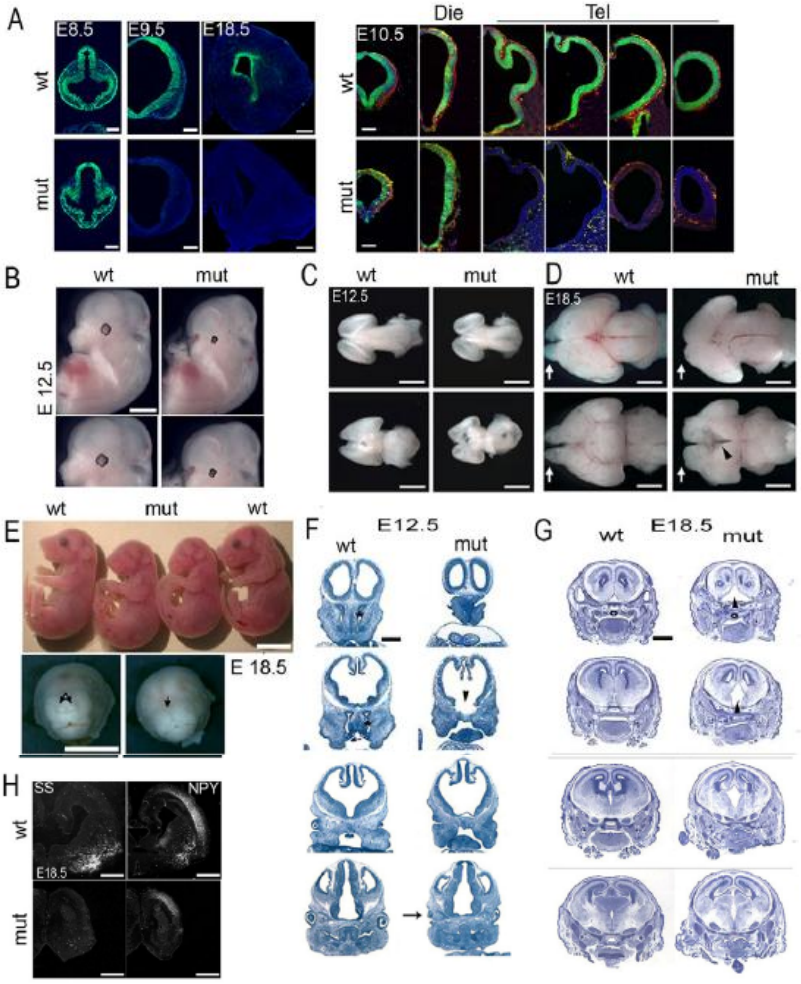


Fig. 1. Early telencephalic ablation of Sox2 with Bf1Cre causes impairment of embryonic brain development. (**A**) Sox2 immunofluorescence (green) on telencephalic sections of normal (*Sox2flox/flox*) and mutant (*Sox2flox/flox;Bf1cre*) mouse embryos. Left: E8.5, E9.5 and E18.5 sections. Sox2 ablation is complete by E9.5. Right: E10.5 sections (posterior left to anterior right). Sox2 ablation is seen in the telencephalon (Tel) but not in the diencephalon (Die). (**B-D**) Brain abnormalities. (B) E12.5 whole embryos. Note the reduced telencephalon, the comparatively unaffected midbrain and the undeveloped eye. (C) Dissected E12.5 brains, viewed dorsally (top) and ventrally (bottom). Note the smaller telencephalic vesicles and the initial ventral tissue loss. (D) Dissected E18.5 brains viewed dorsally (top) show, in mutant, smaller telencephalon (compare to unaffected midbrain) and absence of olfactory bulbs (arrows). Ventral view (bottom) reveals extensive tissue loss (arrowhead) in mutant. (**E**) Mutant E18.5 embryos show smaller head and eyes compared with wild type (wt; top), and facial abnormalities including fusion of the anterior nasal plate (bottom; double arrow in wt, single arrow in mutant) and slightly increased eye proximity. (**F**) E12.5 coronal sections, thyonine stained, anterior (top) to posterior. Arrowhead indicates ventral tissue loss (MGE) in mutant; arrow indicates defective mutant eye. Note olfactory epithelium (asterisk in wt) is missing in the mutant. Note the comparatively unaffected diencephalon in the last section. (**G**) E18.5 coronal sections (thyonine stain) reveal major loss of ventral territories, including striatum region (arrowheads). Circle indicates defective maxillary region (palate). (**H**) ISH for somatostatin (SS) and neuropeptide Y (NPY) shows strong downregulation in the mutant, particularly for SS. Scale bars: 150 μm.

positive subsets of neurons (Markram et al., 2004; Toledo-Rodriguez et al., 2005; Elias et al., 2008; Hébert and Fishell, 2008) (Fig. 1H). SS-positive interneurons originate from the (dorsal) MGE progenitors and require the Nkx2.1 transcription factor for their development (see below) (Hébert and Fishell, 2008; Butt et al., 2008; Flandin et al., 2011). NPY-positive neurons originate from the progenitor domain of the adjacent preoptic area (Gelman et al., 2009), which may be somewhat less severely affected.

Additional abnormalities included absence of the olfactory epithelium [Fig. 1F, asterisk in wild type (wt)] and face abnormalities: the nasal plate, normally developing a characteristic bilateral symmetry, was consistently centrally fused (Fig. 1E, arrows) and underdeveloped. Furthermore, the eyes were abnormal and extremely reduced in size (Fig. 1B,E,F) (see also Taranova et al., 2006); maxillary structures, e.g. the palate, were also abnormal (Fig. 1G); the cortex (Fig. 1B,D,G) was reduced; and the

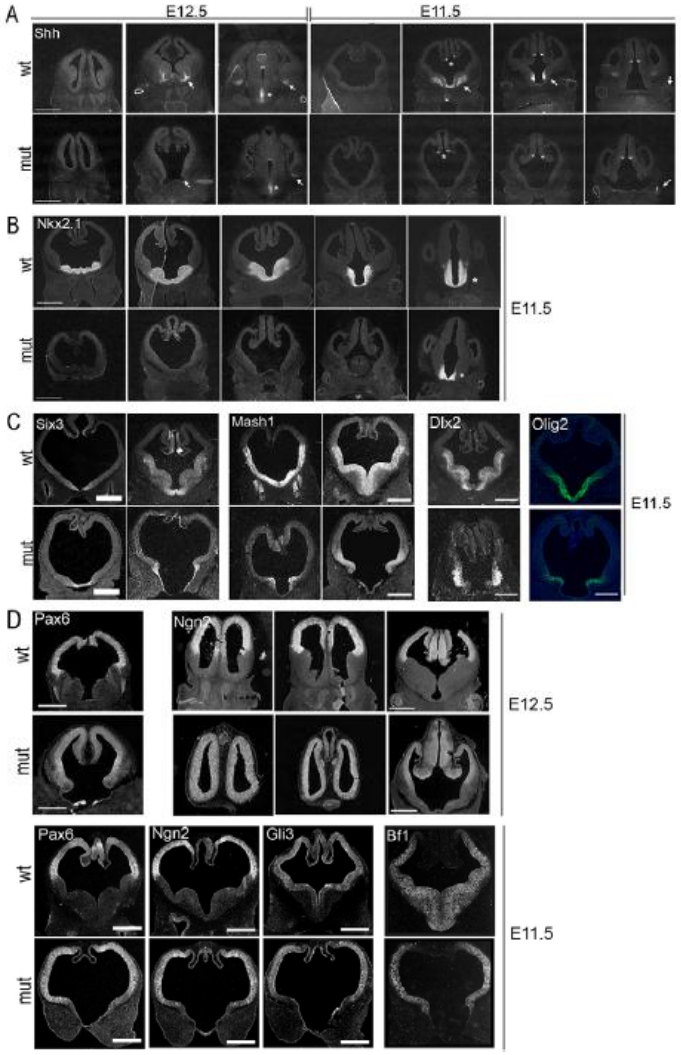


Fig. 2. Expression of ventral determinants is impaired in *Sox2* mutants. (**A**) ISH with *Shh* probe on E12.5 (left) and E11.5 (right) normal (top) and mutant (bottom) mouse embryos (left anterior to right posterior). Arrows indicate the *Shh* signal in wild type, and its absence (midline region) or important reduction (amygdala region) in mutants. Asterisks indicate the signal in diencephalon, a non-*Sox2*-deleted region, as an internal control, showing similar intensity. Arrows in the bottom far-right panel indicate the impaired mutant eyes. (**B**) ISH with *Nkx2.1* probe on E11.5 embryos (left anterior to right posterior). The signal is detected in all telencephalic sections in wild type, but not in mutant. Asterisks indicate the signal in non-*Sox2*-deleted diencephalon, as internal control. (**C**) ISH with probes for ventrally expressed genes at E11.5. Probes are indicated on each panel. Ventral gene expression shows loss or strong downregulation in mutants. Note *Mash1* and *Six3* hybridisation to the olfactory epithelium of wt, but not mutants. (**D**) Expression of some dorsally, or dorsally/ventrally, expressed genes in E12.5 and E11.5 mutants, compared with wild type. Expression of *Pax6* and *Ngn2* is maintained but clearly shifted ventrally in E12.5 mutants. Expression of *Bf1* at E11.5 is retained in the mutant (though lost ventrally where tissue loss is observed). Scale bars: 150 µm.

hippocampus (at E18.5) was severely underdeveloped (not shown). None of the defects described above was seen in control mice (*Sox2*$^{flox/+}$*Bf1-Cre*; *Sox2*$^{flox/+}$; *Sox2*$^{flox/flox}$) (not shown).

Early expression of ventral forebrain determinants is impaired in Sox2 mutants

We focused on the most severely affected region, the ventral telencephalon, to study genes known to be involved in its specification and development. We first analysed embryos by ISH at E12.5, when the morphological defect becomes overt, and at E11.5, when the defective morphology can first be appreciated. The *Shh* gene is expressed in the developing ventral telencephalon, and is crucial at early stages for the development of this region (Fuccillo et al., 2004; Sousa and Fishell, 2010). Furthermore, we had previously found that *Shh* is a Sox2 target, acting as its functional effector in postnatal hippocampal development (Favaro et al., 2009). By E12.5, *Shh* mRNA is completely absent in the midline region following the loss of the tissue expressing it, and is strongly downregulated in the amygdala region (Fig. 2A); in E11.5 mutant embryos, *Shh* is already severely downregulated in the medial ventral telencephalon (Fig. 2A). Indeed, deletion of the *Shh* gene, or that of its receptor Smo, from the early ventral telencephalon using the same *Bf1-Cre* transgene (Fuccillo et al., 2004) produces abnormalities very similar to those of our mutants. Importantly, these abnormalities are less severe than those seen in the complete *Shh* knockout, in which *Shh* expression in the prechordal plate mesoderm is also lost (Chiang et al., 1996).

The transcription factor Nkx2.1, a direct regulator of Shh (Sussel et al., 1999; Jeong et al., 2006), is specifically expressed in the MGE within the developing brain, and is absolutely required for its development (Sussel et al., 1999; Butt et al., 2008; Nóbrega-Pereira et al., 2008). In *Sox2* mutants, *Nkx2.1* expression was already undetectable at E11.5 in the telencephalon (Fig. 2B), but still observed in the non-*Sox2*-deleted diencephalon (Fig. 2B).

Six3, another transcription factor essential for ventral telencephalic development (Lagutin et al., 2003; Geng et al., 2008),

is also a direct activator of Shh (Jeong et al., 2008); expression of *Six3* was only slightly reduced at E11.5, in coincidence with the initial tissue loss (Fig. 2C). Expression of the gene encoding Mash1 (Ascl1 – Mouse Genome Informatics), a transcription factor expressed in the MGE and lateral ganglionic eminence (LGE) and important for GABAergic interneuron development (Guillemot 2007), was essentially lost in regions close to the midline, and reduced more laterally (Fig. 2C). The genes encoding Dlx2 and Olig2, two transcription factors expressed in the MGE and LGE, downstream of Shh activity (Fuccillo et al., 2004), and required for ventral telencephalic development (Sur and Rubenstein, 2005; Hébert and Fishell, 2008), were similarly reduced (Fig. 2C). The Ebf1 transcription factor is expressed within the developing LGE, but not the MGE (Fuccillo et al., 2006; Geng et al., 2008); expression of *Ebf1* was maintained, to some extent, in mutants (supplementary material Fig. S2). These data are consistent with a severe loss of MGE, but some degree of maintenance of LGE primordia.

In contrast to the strong reduction of the 'ventral' effectors described above, expression of transcription factor genes marking the dorsal brain and required for its development, such as *Pax6*, *Ngn2* (*Neurog2* – Mouse Genome Informatics) and *Gli3*, was maintained at E11.5-12.5 in mutants, with a clear tendency for dorsal-specific expression to spread ventrally (Fig. 2D), particularly at E12.5.

Expression of the gene encoding Bf1, a transcription factor expressed both dorsally and ventrally, but required mainly in ventral regions (Gutin et al., 2006; Hébert and Fishell, 2008), was maintained in lateral and dorsal regions, though it was severely reduced in the area affected by initial tissue loss (Fig. 2D, lower-right panel).

Early downregulation of Nkx2.1 precedes ventral tissue loss

As morphological abnormalities are already evident at E11.5, we investigated whether any gene expression defects precede their development. At E10.5 and E9.5, *Nkx2.1* expression was clearly detectable in the ventral telencephalon of the wild type, but was strongly downregulated or absent in the mutant (Fig. 3A). Consistent with a relationship between *Sox2* and *Nkx2.1* expression, the latter was clearly present in diencephalon (Fig. 3A), where *Sox2* was normally expressed (Fig. 1A). Similarly, *Shh* expression, which largely overlaps with that of *Nkx2.1*, was absent or weak in a few of the mutant embryos at E10.5 (not shown). *Six3* expression was only slightly decreased in mutants at E10.5 and E9.5 (Fig. 3B,C). By contrast, the gene encoding Bf1, which acts in parallel with Shh (Hébert and Fishell, 2008), was normally expressed in *Sox2* mutants, compared with controls (Fig. 3C). *Sox1* and *Sox3*, members of the same Sox transcription subfamily as *Sox2*, are widely co-expressed with *Sox2* in the telencephalon; they do not show major variations in mutant embryos at these stages (Fig. 3C). Sox9, which stimulates NSC growth after E10.5-11.5 (Scott et al., 2010), was normally expressed at these early stages (supplementary material Fig. S1).

We conclude that *Sox2* deletion affects the expression of early, important determinants of brain development, in a region-specific manner: several ventral fate genes are severely affected, whereas activity of dorsal genes is maintained. Notably, one essential effector of ventral telencephalon and MGE development, and activator of Shh, Nkx2.1, is downregulated at early stages.

Increased apoptotic cell death in early *Sox2*-mutant ventral telencephalon

We investigated whether ventral tissue loss in *Sox2* mutants was due to impaired cell proliferation and/or increased cell death. Cell

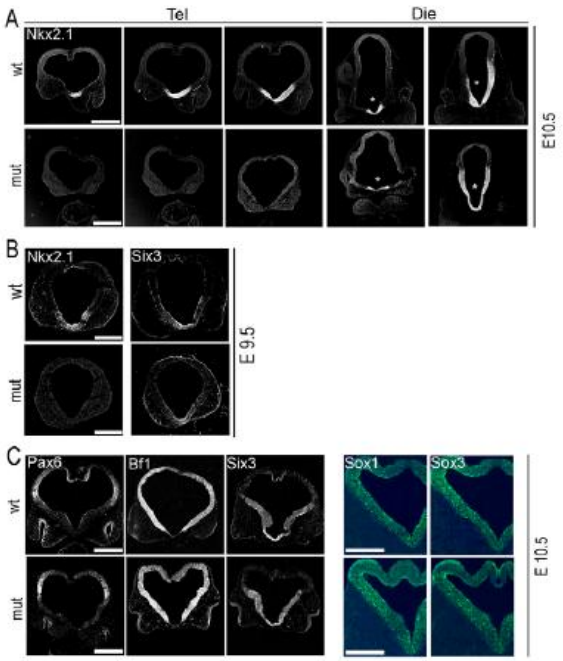


Fig. 3. Gene expression abnormalities are detected by *in situ* hybridisation at early stages of development, preceding morphologic impairment in mutants. (**A**,**B**) *Nkx2.1* expression is not established in the telencephalon (Tel) of mouse mutants at E10.5 (A) or E9.5 (B), but is preserved in the adjacent non-*Sox2*-deleted diencephalon (Die). *Six3* expression is only slightly reduced at E9.5. Asterisks indicate the Nkx2.1 signal in non-Sox2-deleted diencephalon. (**C**) *Pax6*, *Bf1*, *Six3* and (by immunofluorescence) Sox1 and Sox3 do not show major changes in mutants at E10.5. Scale bars: 200 µm.

proliferation, assessed by BrdU labelling at E9.5 and E10.5 just prior to the appearance of morphological defects, was not decreased overall in mutant telencephalon or specifically in the ventral region (Fig. 4A). Apoptotic cell death, assayed by TUNEL, was comparable between normal and mutant embryos at E9.5, but a threefold increase in TUNEL-positive cells was observed in the ventral telencephalon of E10.5 mutants (Fig. 4B).

Thus, increased cell death could directly cause ventral tissue loss in the mutants. Apoptotic death is a possible consequence of impaired ventral gene expression (e.g. loss of Shh, which has anti-apoptotic activities) (Cayuso et al., 2006), which precedes by at least one day the increase in cell death.

Defective expression of ventral genes and morphological abnormalities of *Sox2* mutants are rescued by a Shh agonist

The ventral defects observed in Bf1-cre-deleted *Sox2* mutants are very similar to those observed in mutants of the sonic hedgehog pathway [in which the Shh receptor smoothened (Smo) is conditionally ablated with the same deleter, Bf1cre] (Fuccillo et al., 2004), as well as to that of *Nkx2.1* mutants (Sussel et al., 1999). Indeed, *Sox2* mutants show (Figs 2, 3) severely impaired expression of both Shh and Nkx2.1, a direct activator of Shh (Jeong et al., 2006).

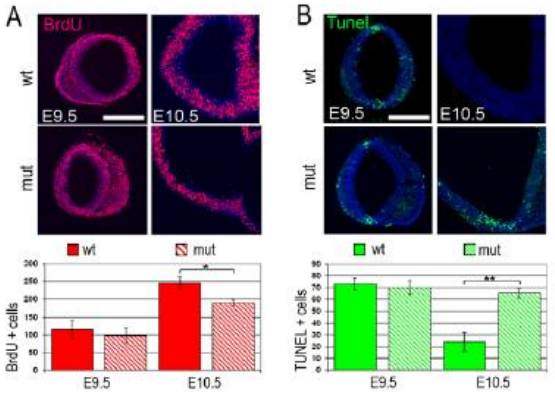


Fig. 4. Cell death is ventrally increased in *Sox2* mutant telencephalon. (**A**) Immunofluorescence for BrdU in normal (wt) and mutant (mut) mouse telencephalon; histogram shows quantification of BrdU-positive cells in the ventral half of the telencephalon. (**B**) TUNEL assay of normal and mutant telencephali. Sections (top) show increased TUNEL signal in mutant, concentrated ventrally. Histogram shows quantification; significantly higher numbers of TUNEL-positive cells are found in mutants compared with wild type at E10.5 (n=5 wild-type and mutant embryos analysed, for both assays). Values on the y-axis represent the mean±s.d. of the total number of cells counted, on every fifth 20-µm section throughout the telencephalon (four or five total sections counted for E9.5 or 10.5 brains, respectively). *P<0.01; **P<0.001 (Student's t-test). Scale bars: 150 µm.

Hence, we tested whether Shh signalling was involved in the *Sox2* mutant phenotype, by administering mice an agonist (Shh-ag) that activates the Shh co-receptor smoothened (Frank-Kamenetsky et al., 2002). Shh-ag was administered at E8.5 (just prior to *Sox2* ablation) and E10.5. Already at E14.5, expression of ventral determinants *Mash1* and *Dlx2*, which is impaired in the untreated mutants, recovered to a significant, albeit not complete, extent (Fig. 5); morphologically, the ventral brain also recovered a somewhat more normal shape, with ventral bulges reminiscent of wild-type ganglionic eminences (Fig. 5). By contrast, no major effect was observed on brain morphology or gene expression of treated wild-type littermates (Fig. 5).

We conclude that failure to activate Shh signalling is an important cause of the defects observed in *Sox2*-mutant embryonic telencephalon.

Sox2 activates Nkx2.1 cell-autonomously

Nkx2.1 is a direct activator of the *Shh* gene and is required for its expression *in vivo* (Sussel et al., 1999; Jeong et al., 2006; Sousa and Fishell, 2010); however, *Nkx2.1* expression is also stimulated in response to Shh signalling (Fuccillo et al., 2004; Xu et al., 2005; Gulacsi and Anderson, 2006; Sousa and Fishell, 2010). *Nkx2.1* expression failed to be established early in *Sox2* mutants, and remained absent at later stages (Figs 2, 3); we thus investigated whether loss of *Nkx2.1* expression depends on Sox2 cell-autonomously or is secondary to the loss of Shh expression. To this end, we used a *Sox2CreERT2* transgene, encoding the tamoxifen-activatable Cre guided by the *Sox2* telencephalic enhancer/promoter (Favaro et al., 2009). Tamoxifen treatment at E8.5 and E10.5 caused a 'salt-and-pepper' deletion of Sox2, as seen by immunofluorescence at E14.5 (Fig. 6); some Shh expression,

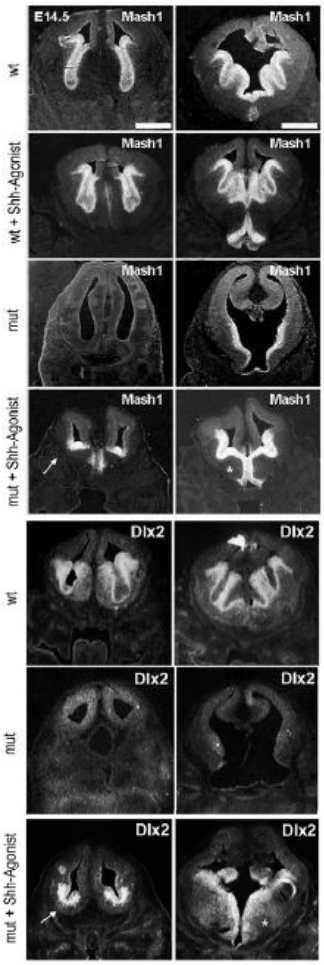


Fig. 5. A pharmacological Shh agonist significantly rescues ventral gene expression and morphology in *Sox2* mutants. ISH for ventral markers *Mash1* (top) and *Dlx2* (bottom) on normal (wt) and *Sox2*-deleted mouse embryos (mut), treated with Shh agonist or untreated. Telencephalic sections at two levels, anterior (left) and posterior (right), are shown. Expression of *Dlx2* and *Mash1*, strongly impaired (particularly anteriorly) in mutants by E14.5, is significantly rescued in treated mutants, together with an improved ventral morphology. No major effect is seen in the same region on normal embryos from the same litter. A representative experiment is shown out of n=4 mutant embryos analysed. Scale bars: 200 µm.

presumably arising from non-deleted cells, was maintained (Fig. 6), and no major abnormality was noticed in these mosaic-deleted embryos. We analysed Nkx2.1 and Sox2 expression by immunofluorescence in the ventral telencephalic ventricular zone. In control embryos, most cells co-expressed Sox2 and Nkx2.1 (Fig. 6). In tamoxifen-treated embryos, Sox2-expressing cells were strongly reduced; Nkx2.1 expression was retained in cells in which Sox2 was still expressed, but was rarely, if ever, seen in cells that did not express Sox2 (Fig. 6). We conclude that expression of Nkx2.1 requires Sox2 cell-autonomously.

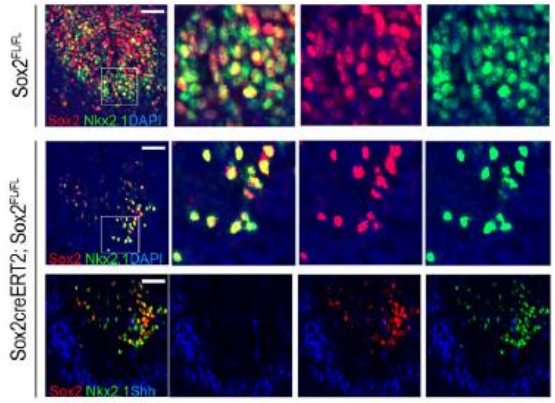


Fig. 6. Mosaic Sox2 ablation via a *Sox2CreERT2* transgene leads to cell-autonomous loss of Nkx2.1. Immunofluorescence for Nkx2.1 (green) and Sox2 (red) in normal mouse embryos (*Sox2flox/flox*; top row) at E14.5 in the ventral telencephalon. Left-hand panel: general view at low magnification; right-hand panels (top and intermediate rows) show a magnification of the boxed area, with merged and separated colour channels. Sox2 and Nkx2.1 are co-expressed in most cells in the wild type. In *Sox2CreERT2; Sox2flox/flox* embryos treated with tamoxifen at E8.5 and E10.5, only a fraction of ventral telencephalic cells retains Sox2 expression (second row, compare stained cells with total DAPI-labelled nuclei); Nkx2.1 is detected in those cells that show Sox2 expression. Shh is detectable by immunofluorescence (blue in lower row) in this region. One representative experiment is shown out of *n*=3 embryos analysed. Scale bars: 500 µm.

Regulation of *Nkx2.1* by Sox2

The early loss of *Nkx2.1* following Sox2 telencephalic ablation (Figs 2, 3) raised the possibility that *Nkx2.1* expression is directly controlled by Sox2, within a specific subregion of the Sox2 pan-neural expression domain.

In a survey for evolutionarily conserved regions in the *Nkx2.1* genomic locus, we detected a small conservation peak just upstream to the second *Nkx2.1* exon (Fig. 7A,B). Evolutionary conservation within this region was present across vertebrate evolution (Fig. 7A). This region included a single and a twin potential Sox2-binding sites; both sites are conserved in mammals, and at least one site is conserved in vertebrates, including teleostean fishes (Fig. 7A). The *Nkx2.1* gene has two promoters, one upstream to the first exon ('distal' promoter), the other in the intron between exon 1 and 2 ('proximal' promoter), both of which are functional *in vitro* and *in vivo* (including E10.5 and E14.5 telencephalon; supplementary material Fig. S3), though the latter might be the stronger (Pan et al., 2004; Hamdan et al., 1998). The Sox2-binding sites (Fig. 7B) lie in the region between the 'distal' and 'proximal' transcriptional start sites. ChIP from E14.5 embryos gave a moderate (2.5-fold) enrichment for this region (not shown). To develop a functional reporter assay for promoter sequences, we cloned a fragment including the conserved *Sox2* sequences from the *Nkx2.1* region upstream to a green fluorescent protein (*GFP*) gene (Fig. 7B), and tested it in Medaka embryos. These sequences drove GFP activity in forebrain regions superimposable with those showing endogenous *Nkx2.1* expression (Fig. 8A,B). In line with these observations, ChIP from stage 16-18 Medaka embryos with anti-Sox2 antibody revealed a 30-fold enrichment (relative to ChIP with a non-related IgG) of the *Nkx2.1* intronic conserved element, which was not observed in negative control regions (a different region of the *Nkx2.1* gene devoid of Sox2 consensus sites and the 18S RNA-encoding gene) (Fig. 8C), indicating that Sox2 binds to the *Nkx2.1* promoter/enhancer *in vivo*.

Co-injection of *Sox2* mRNA enhanced expression of the *Nkx2.1-GFP* transgene (Fig. 8F,J, compare with 8D,H), as well as of the endogenous *Nkx2.1* gene, which is both increased and expanded anteriorly, as detected by ISH (Fig. 8N, compare with 8L). Conversely, co-injection of anti-*Sox2* morpholino (Sox2 MO) (Beccari et al., 2012) (Fig. 8H,I) antagonised the activity of the co-injected *Nkx2.1-GFP* transgene (Fig. 8E,I, compare with 8D,H), as well as endogenous *Nkx2.1* expression, the domain of which was also reduced (Fig. 8M, compare with 8L).

To evaluate the importance of a direct action of Sox2 on transgene regulation, we mutated the conserved Sox2 sites within the *Nkx2.1-GFP* transgene. GFP expression required the integrity of the Sox2-consensus sites, as their mutation caused a substantial loss of transgene activity (Fig. 8G,K, compare with 8D,H). This result is consistent with experiments showing that mutation of the same Sox2 sites in a luciferase-reporter gene driven by the 'proximal' promoter abolishes the response to co-transfected Sox2 in P19 teratocarcinoma cells (Fig. 7C).

These experiments show that Sox2 is an important regulator of *Nkx2.1* expression in Medaka fish. In Medaka fish, the telencephalon is substantially reduced in size and lacks detectable endogenous *Nkx2.1* expression (Fig. 8A,L). This raises the question of whether the *Nkx2.1* regulatory elements studied here are sufficient to drive expression in the ventral telencephalon of the mouse. We thus tested the GFP construct described above in E13.5 mouse telencephalon by transient electroporation. Supplementary material Fig. S4 shows that two days after electroporation the transgene is expressed in the ventral telencephalon. At E13.5-15.5, mutation of the Sox2 sites had little effect on telencephalic expression, indicating that, at this developmental stage, other transcription factor-binding sites play a role in the regulation of this construct in the telencephalon (see Discussion).

DISCUSSION

The Sox2 transcription factor is crucial for the maintenance of several types of stem cells, including pluripotent, neural and osteogenic stem cells (Masui et al., 2007; Favaro et al., 2009; Basu Roy et al., 2010). Despite the importance of Sox2 in NSCs *in vitro*, major abnormalities in brain development were not detected by conditional ablation of Sox2 at midgestation (E12.5) in mouse, with the exception of defects in postnatal development of the hippocampus dentate gyrus and of the retina (Taranova et al., 2006; Miyagi et al., 2008; Favaro et al., 2009). Here, we examined the hypothesis that Sox2 is required in the developing telencephalon at early developmental stages. By conditionally deleting *Sox2* by E9.5 in all the developing telencephalon, we discovered that *Sox2* deletion strongly affects embryonic development of the ventral telencephalon. Patterning of the ventral telencephalon is crucially dependent on the induction of the diffusible factor Shh, which is mediated by the transcription factor Nkx2.1 (Sousa and Fishell, 2010). The crucial role of Shh is highlighted by the severe abnormalities in patients affected with holoprosencephaly, a developmental defect of the brain ventral midline, caused by *SHH* mutations (Dubourg et al., 2004; Roessler and Muenke, 2010). Here, we show that Sox2 is required for the early expression of *Nkx2.1*, thus controlling downstream ventral patterning genes such as *Shh*.

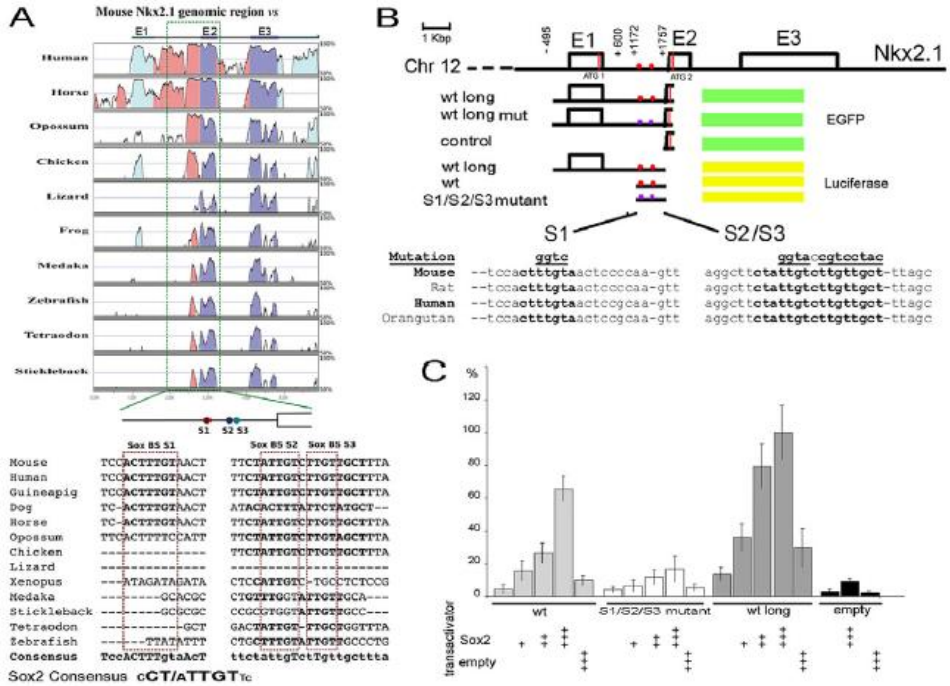


Fig. 7. Nkx2.1 regulation by Sox2. (**A**) The *Nkx2.1* intronic promoter/enhancer is evolutionary conserved in vertebrates. The genomic *Nkx2.1* sequences from the different vertebrate species were retrieved from the UCSC genome browser and aligned using the Shuffle-LAGAN of Vista. Pink, pale-blue and violet peaks represent conserved non-coding elements (75% conservation over 100 bp), mRNA untranslated sequence and coding sequence, respectively. Putative Sox2-binding sites, indicated as S1, S2 and S3, localise to a conserved element in the first intron. The S1 binding site is conserved among mammals but not in other vertebrates. S2 and S3 binding sites were conserved among most vertebrates. The indicated Consensus within the alignment was derived with Simple Consensus Maker (http://www.hiv.lanl.gov/content/sequence/CONSENSUS/SimpCon.html). The general Sox2 consensus is from Engelen et al. (Engelen et al., 2011). (**B**) Schematic of the *Nkx2.1* gene (Sox2 sites indicated by red dots) with constructs used for transgenic (EGFP) and transfection experiments (luciferase); sequence shows the Sox sites and the mutations introduced. EGFP constructs: *wt-long*: a region from nucleotide –495 (5′ to exon 1) to nucleotide +1942 in the second exon was cloned in frame with the EGFP reporter (green box); this fragment comprises the conserved elements in the first intron. *wt-long-mut*: carries mutations in the Sox2 sites, shown below (underlined). *control*: promoter-less EGFP construct. Luciferase constructs: *wt-long*: same region as in the EGFP construct, here linked to a luciferase reporter. *wt*: a shorter region encompassing the two Sox2 sites, from +1172 to +1757. *S1/S2/S3 mutant*: same as wt, with the same Sox2 mutations as in the *wt-long-mut*-EGFP reporter. (**C**) Co-transfection in P19 cells of *Nkx2.1* promoter with luciferase vectors (1 µg) and their mutant versions (shown in B), with increasing amounts (+, 0.125 µg; ++, 0.25 µg; +++, 0.5 µg) of *Sox2*-expressing vector (Sox2), or with the corresponding 'empty' vector. Results are the mean of at least three independent transfections, in triplicate.

Early Sox2 loss affects *Nkx2.1* and *Shh* expression

In *Bf1-cre Sox2*-deleted embryos, extensive ventral tissue loss occurs starting at ~E11.5, developing into major abnormalities of the ganglionic eminences (particularly the MGE) and of MGE-derived GABAergic neurons at later stages (Fig. 1); the expression of the dorsal markers *Pax6* and *Ngn2* (Fig. 2) also tends to spread ventrally, pointing to abnormalities of the ventral versus dorsal specification of the telencephalon. These defects strongly resemble those observed in *Nkx2.1* germ-line deletion and in the conditional ablation (via Bf1-cre) of the Shh receptor smoothened (Sussel et al., 1999; Fuccillo et al., 2004; Sousa and Fishell, 2010). We confirmed the connection to Shh by showing that *Shh* expression is strongly diminished in the ventral region of the mutant telencephalon concomitantly and prior to the onset of tissue loss (Figs 2, 3). Moreover, treatment of the embryos with a Shh agonist substantially rescued ventral development in the mutant brain (Fig. 5), though prenatal lethality still occurred. This rescue is reminiscent of that of the hippocampal dentate gyrus stem cells and postnatal growth by the same drug, in *nestin-cre Sox2*-deleted mice (Favaro et al., 2009). Local cell death in the ventral telencephalon is detected just prior to the onset of tissue loss (Fig. 4); this might also relate to loss of Shh signalling, which activates the anti-apoptotic gene *Bcl2* (Cayuso et al., 2006). These data, together with those of Favaro et al. (Favaro et al., 2009), highlight an unexpected role of Sox2 in mediating the development of specific brain regions at defined stages via Shh-dependent regulation.

How does Sox2 regulate *Shh* expression in the ventral telencephalon? Sox2 might regulate genes involved in local specification of ventral regions. A prime candidate target is *Nkx2.1*, essential for ventral brain development and correct dorsoventral patterning (Sussel et al., 1999; Sousa and Fishell, 2010). Nkx2.1 is thought to mediate the early 'homogenetic' induction of Shh in the ventral telencephalon, in response to the gradient of mesendoderm-derived Shh (Sousa and Fishell, 2010). Indeed, *Nkx2.1* mutant mice fail to express Shh in the ventral region (Sussel et al., 1999), and their phenotype resembles both that of *Shh* or smoothened mutants

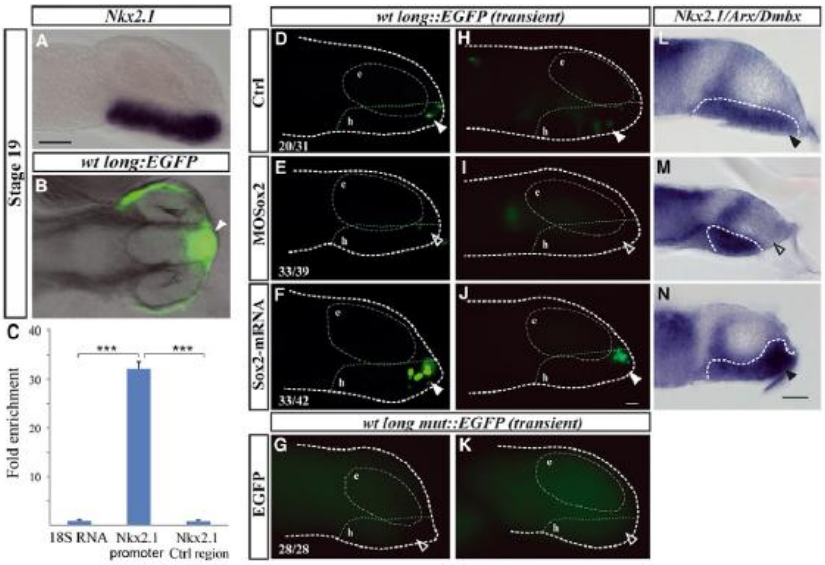


Fig. 8. *Nkx2.1* transgene and endogenous gene regulation by Sox2 in Medaka. (**A**) Lateral view of stage 19 Medaka embryo hybridised with an *Nkx2.1*-specific probe. (**B**) Dorsal view of a stable transgenic Medaka fish embryo carrying the EGFP reporter driven by the mouse *Nkx2.1* promoter/enhancer (*wt long*). Note the expression of the EGFP reporter corresponding to anterior hypothalamus (white arrowhead). (**C**) ChIP performed with anti-Sox2 on chromatin from stage 19 Medaka embryos. The histograms show the mean value+s.e.m. of a representative experiment performed in triplicate. Fold enrichment for the tested regions (*Nkx2.1* promoter, and a control region (Nkx2.1-Ctrl region) located 2783 bp downstream of the predicted *Nkx2.1* start codon) was normalised to control IgG and compared with 18s RNA. ***$P<0.0001$. (**D-N**) Medaka fish embryos were co-injected with control (D,H,L,G,K), Sox2 MO (E,I,M) or *Sox2* mRNA (F,J,N). Lateral views of transient transgenic Medaka fish embryos carrying the EGFP reporter driven by the *wt long* (D-F,H-J) or mutated (*wt long mut*) (G,K) mouse promoter/enhancer (confocal microscopy, G-N) and of embryos hybridised *in toto* with probe against diencephalic (*Arx*), mesencephalic (*Dmbx1*) and hypothalamic (*Nkx2.1*) markers (L-N). As expected for transient transgenic embryos, injections of the *wt long::EGFP* construct activates reporter expression with a variable (compare D with H, white arrowheads) mosaic pattern in discrete cells always in the anterior domain of the hypothalamus. Reporter expression is no longer observed when embryos are injected with the mutated version (unfilled arrowheads in G,K) or co-injected with Sox2 MO (empty arrowheads in E,I), whereas expression is increased in intensity upon *Sox2* mRNA injection (white arrowheads in F,J). Note also the parallel reduction (unfilled arrowhead in M) or expansion (black arrowhead in N) of the anterior hypothalamic mRNA distribution of *Nkx2.1* (black arrowhead in L), whereas the distribution of diencephalic and mesencephalic markers did not change. The shape of the embryos and of the eye (e) are outlined by dashed white lines. The *Nkx2.1*-positive hypothalamic domain (h) is outlined by green dotted lines. Frequency of the observed phenotype for each one of the experimental conditions is indicated in the bottom-left corner of panels (D-G). Two prototypic embryos are shown for each experimental condition. Scale bars: 50 μm in A-K; 40 μm in L-N.

(Fuccillo et al., 2004), and that of the present *Sox2* mutant. Furthermore, mutations destroying a consensus *Nkx2.1*-binding site in a distant *Shh* enhancer, active in telencephalon, impair the transcription of reporter constructs in transgenic mice (Jeong et al., 2006). Finally, *Nkx2.1* is required for expression of transcription factors Lhx6 and Lhx8 (also known as Lhx7) (Sussel et al., 1999), which coordinately activate Shh in neurons in the developing MGE (Flandin et al., 2011).

The early severe impairment of *Nkx2.1* expression in *Sox2* mutants already by E10.5 (Fig. 3), and the absence of *Shh* at least from E10.5/11.5 onwards (Fig. 2), are consistent with the hypothesis that a large part of the phenotypic effects of *Sox2* ablation is initially mediated by *Nkx2.1* deficiency.

Do other transcription factors mediate the effects of *Sox2* deficiency?

Presently, we can neither rule out nor implicate other genes besides *Nkx2.1* (and *Shh*) in the early effects of *Sox2* ablation. *SIX3* mutations are found in some human patients affected with holoprosencephaly (Jeong et al., 2008), and *Six3* haploinsufficiency caused by the 'knock-in' of a human mutant *SIX3* gene impairs Shh expression and MGE development in mouse, recapitulating features of the human phenotype (Geng et al., 2008). Moreover, a mutation in a *SIX3*-binding site, within a *SHH* long-range acting enhancer, has been detected in a human holoprosencephalic patient (Jeong et al., 2008). In our mutants, *Six3* expression was only slightly diminished in the ventral region at early stages, when *Nkx2.1* expression was already substantially affected (Figs 2, 3), making it unlikely that the effects of *Sox2* ablation are mediated by *Six3* deficiency. Interestingly, in the Medaka telencephalon, *Sox2* activates *Six3*, but the two genes seem to have antagonistic function in the hypothalamus (Beccari et al., 2012). The expression of *Bf1*, another candidate gene (Gutin et al., 2006; Hébert and Fishell, 2008), was also not significantly affected at these early stages, despite the *Bf1* hemizygosity due to cre 'knock-in' (Fig. 3).

After *Nkx2.1* (and thus *Shh*) expression is initially activated by mesendoderm-derived Shh, its activity is normally maintained, at later stages, by Shh itself (Xu et al., 2005; Xu et al., 2010; Sousa and Fishell, 2010). Following mosaic *Sox2* deletion at E8.5 (Fig. 6), Nkx2.1 is poorly expressed later on in *Sox2*-deleted cells, even in the presence of Shh. These data do not contradict the notion that Shh maintains later expression of Nkx2.1 (Xu et al., 2005; Xu et

al., 2010), but simply highlight an early requirement for Sox2 in establishing this process.

The ventral telencephalic defects due to *Sox2* early ablation point to a marked regional specificity of Sox2 requirement in development (Fig. 1). As an example, Sox2 is required for *Nkx2.1* (and *Shh*) (see also Favaro et al., 2009) expression only within a specific subregion of the Sox2 pan-neural expression domain (Figs 2, 3). This might depend on local Sox2 concentrations, and/or on the presence of additional co-regulators. The transcription factors Sox1 and Sox3 are closely related to Sox2 (Wegner and Stolt, 2005), and recognise similar DNA sequences *in vitro* (Kondoh and Kamachi, 2010; Wegner, 2010). Thus, in regions in which Sox2 ablation causes few or no defects, Sox1 and/or Sox3 might compensate for Sox2 deficiency. Indeed, embryos doubly mutant for *Sox2* and *Sox3* ($Sox2^{+/-}$; $Sox3^{-/Y}$) develop diencephalic defects, mirroring those observed following early diencephalic deletion of the *Shh* gene (Zhao et al., 2012); by contrast, no such defects were observed in single mutants. Thus, Sox3 does compensate for some Sox2 functions in the diencephalon. By contrast, in the ventral telencephalon (present paper), Sox2 has some region-specific functions that cannot be complemented by Sox3 and Sox1, in spite of their normal expression levels (as also observed in Medaka) (Beccari et al., 2012). Most sequence diversity between Sox1, Sox2 and Sox3 occurs outside the DNA-binding domain; unique partnerships between Sox2 and co-factors (Kondoh and Kamachi, 2010; Bernard and Harley, 2010; Wegner, 2010) might mediate its specific functions in ventral (and hippocampal) (Favaro et al., 2009) brain development.

Among Sox2-specific interactors/DNA-binding proteins, CHD7 is known to activate a set of common targets relevant for anophtalmia (caused by *SOX2* mutations in humans and mice) and CHARGE syndrome (caused by *CHD7* mutations) (Engelen et al., 2011). The specific expression of different Sox2 interacting/cooperating factors in various tissues might impart regional specificity to the defects caused by the absence of Sox2; indeed, an important (antagonistic) relationship of Sox2 with Pax6 was reported in a study of the development of neural competence in the optic cup (Matsushima et al., 2011).

How does Sox2 regulate *Nkx2.1*?

In Medaka, modulation of Sox2 levels correlates with changes in endogenous *Nkx2.1* expression intensity and spatial distribution (Fig. 8L-N). Moreover, the *Nkx2.1-GFP* construct faithfully recapitulates endogenous *Nkx2.1* expression, and requires intact Sox2-binding sites for activity (Fig. 8B,D-K), consistent with transfection results in P19 cells (Fig. 7C). These data identify *Nkx2.1* as a Sox2 target in Medaka. In Medaka, unlike in mouse, Nkx2.1 is not active in the telencephalon, and we thus cannot directly extrapolate from the Medaka results to infer direct regulation of Nkx2.1 by Sox2 in the telencephalon. We tested the same *Nkx2.1* construct in mouse by electroporation in the telencephalon at E13.5-15.5, showing that it is active in the telencephalon, preferentially in the MGE and other ventral regions (supplementary material Fig. S4); however, the mutation of the Sox2 sites did not substantially affect the activity of the construct (not shown). These results formally rule out the possibility that, at the E13.5-15.5 stage, the Sox2-binding sites, per se, are required for activity of the *Nkx2.1* promoters in the telencephalon. As a consequence, it remains unclear whether, in *Sox2* mutant mouse, the observed loss of *Nkx2.1* expression depends on the loss of a direct activity of Sox2 on the *Nkx2.1* promoter (so far unproven), on additional effects on other regulatory elements, or on indirect effects mediated by other Sox2-dependent factors.

How do we reconcile the data obtained by electroporation in mouse telencephalon with the low activity of the Sox2-mutated reporter in Medaka (and in *in vitro* transfected P19 cells)? We speculate that the requirement for Sox2 binding to the *Nkx2.1* promoter (if any) might be limited to the early stages of development. We know that late *Sox2* ablation (E12.5) has little effect on ventral telencephalic development and gene expression (Favaro et al., 2009; unpublished data), whereas early ablation (E9.5) causes important defects. We therefore hypothesise that Sox2-binding sites in the *Nkx2.1* promoter might be required for *Nkx2.1* regulation in mouse telencephalon at early (~E9.5), but not late, stages of development. Once established, *Nkx2.1* expression might be maintained, at E14.5, by transcription factors other than Sox2, and additional regulatory regions might be involved in controlling *Nkx2.1* expression. Unfortunately, the present constructs show very low activity following electroporation at E9.5, and similar transgenic constructs were not expressed in embryonic ventral telencephalon (Pan et al., 2004), making it difficult to test this hypothesis. Several regions adjacent to *Nkx2.1* bind Sox2 in ChIP experiments, and distal intergenic regions exhibit long-range interactions with the *Nkx2.1* gene (not shown), and might potentiate the promoter. These sequences will be investigated in the future in order to determine to their regulatory potential.

Conclusions

Sox2, despite its ubiquitous expression in neural stem/progenitor cells at all levels of the developing central nervous system, is absolutely required, in a stage- and region-specific way, in a limited set of locations, here exemplified by the early ventral telencephalon and by the hippocampus (Favaro et al., 2009). In the ventral telencephalon, Nkx2.1 is likely to be the main (although not necessarily the exclusive) mediator of Sox2 effects; other factors might mediate Sox2 activities in different regions. Expression defects of Shh (a target of Nkx2.1 in the ventral telencephalon), are common to both territories affected by Sox2 loss (ventral telencephalon and hippocampus); it will be interesting to examine other embryonic brain sites expressing Shh for defects caused by Sox2 ablation at different developmental stages.

Acknowledgements

We thank Carol Yan for help with *Sox2CreERT2* transgenic experiments, Sara Lazzarin for *in situ* hybridisation; Barbara Costa for advice on pharmacological treatments; and Laura Croci and Giacomo Consalez for the *Ebf1a* probe. S.N. dedicates this work to the dear memory of Cecilia Tiveron and Larysa Pevny, good and generous scientists and friends, who researched development with insight and with love.

Funding

S.K.N. is supported by Cariplo Foundation [grant 20100673]; ASTIL Regione Lombardia [SAL-19, N Prot FL 16874]; Telethon [GGP12152]; and Associazione Italiana Ricerca sul Cancro [AIRC IG-5801]. P.B. is supported by the Spanish Ministerio de Economia y Competividad [BFU2010-16031]; and the Comunidad Autonoma de Madrid [BDM-2315].

Competing interests statement

The authors declare no competing financial interests.

Supplementary material

Supplementary material available online at
http://dev.biologists.org/lookup/suppl/doi:10.1242/dev.073411/-/DC1

References

Avilion, A. A., Nicolis, S. K., Pevny, L. H., Perez, L., Vivian, N. and Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev.* **17**, 126-140.

Basu-Roy, U., Ambrosetti, D., Favaro, R., Nicolis, S. K., Mansukhani, A. and Basilico, C. (2010). The transcription factor Sox2 is required for osteoblast self-renewal. *Cell Death Differ.* **17**, 1345-1353.

Beccari, L., Conte, I., Cisneros, E. and Bovolenta, P. (2012). Sox2-mediated differential activation of Six3.2 contributes to forebrain patterning. *Development* **139**, 151-164.
Bernard, P. and Harley, V. R. (2010). Acquisition of SOX transcription factor specificity through protein-protein interaction, modulation of Wnt signalling and post-translational modification. *Int. J. Biochem. Cell Biol.* **42**, 400-410.
Butt, S. J., Sousa, V. H., Fuccillo, M. V., Hjerling-Leffler, J., Miyoshi, G., Kimura, S. and Fishell, G. (2008). The requirement of Nkx2-1 in the temporal specification of cortical interneuron subtypes. *Neuron* **59**, 722-732.
Bylund, M., Andersson, E., Novitch, B. G. and Muhr, J. (2003). Vertebrate neurogenesis is counteracted by Sox1-3 activity. *Nat. Neurosci.* **6**, 1162-1168.
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.
Cayuso, J., Ulloa, F., Cox, B., Briscoe, J. and Martí, E. (2006). The Sonic hedgehog pathway independently controls the patterning, proliferation and survival of neuroepithelial cells by regulating Gli activity. *Development* **133**, 517-528.
Chiang, C., Litingtung, Y., Lee, E., Young, K. E., Corden, J. L., Westphal, H. and Beachy, P. A. (1996). Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function. *Nature* **383**, 407-413.
Conte, I. and Bovolenta, P. (2007). Comprehensive characterization of the cis-regulatory code responsible for the spatio-temporal expression of olSix3.2 in the developing medaka forebrain. *Genome Biol.* **8**, R137.
Dubourg, C., Lazaro, L., Pasquier, L., Bendavid, C., Blayau, M., Le Duff, F., Durou, M. R., Odent, S. and David, V. (2004). Molecular screening of SHH, ZIC2, SIX3, and TGIF genes in patients with features of holoprosencephaly spectrum: Mutation review and genotype-phenotype correlations. *Hum. Mutat.* **24**, 43-51.
Elias, L. A., Potter, G. B. and Kriegstein, A. R. (2008). A time and a place for nkx2-1 in interneuron specification and migration. *Neuron* **59**, 679-682.
Ellis, P., Fagan, B. M., Magness, S. T., Hutton, S., Taranova, O., Hayashi, S., McMahon, A., Rao, M. and Pevny, L. (2004). SOX2, a persistent marker for multipotential neural stem cells derived from embryonic stem cells, the embryo or the adult. *Dev. Neurosci.* **26**, 148-165.
Engelen, E., Akinci, U., Bryne, J. C., Hou, J., Gontan, C., Moen, M., Szumska, D., Kockx, C., van Ijcken, W., Dekkers, D. H. et al. (2011). Sox2 cooperates with Chd7 to regulate genes that are mutated in human syndromes. *Nat. Genet.* **43**, 607-611.
Fantes, J., Ragge, N. K., Lynch, S. A., McGill, N. I., Collin, J. R., Howard-Peebles, P. N., Hayward, C., Vivian, A. J., Williamson, K., van Heyningen, V. et al. (2003). Mutations in SOX2 cause anophthalmia. *Nat. Genet.* **33**, 461-463.
Favaro, R., Valotta, M., Ferri, A. L., Latorre, E., Mariani, J., Giachino, C., Lancini, C., Tosetti, V., Ottolenghi, S., Taylor, V. et al. (2009). Hippocampal development and neural stem cell maintenance require Sox2-dependent regulation of Shh. *Nat. Neurosci.* **12**, 1248-1256.
Ferri, A. L., Cavallaro, M., Braida, D., Di Cristofano, A., Canta, A., Vezzani, A., Ottolenghi, S., Pandolfi, P. P., Sala, M., DeBiasi, S. et al. (2004). Sox2 deficiency causes neurodegeneration and impaired neurogenesis in the adult mouse brain. *Development* **131**, 3805-3819.
Flandin, P., Zhao, Y., Vogt, D., Jeong, J., Long, J., Potter, G., Westphal, H. and Rubenstein, J. L. R. (2011). Lhx6 and Lhx8 coordinately induce neuronal expression of Shh that controls the generation of interneuron progenitors. *Neuron* **70**, 939-950.
Frank-Kamenetsky, M., Zhang, X. M., Bottega, S., Guicherit, O., Wichterle, H., Dudek, H., Bumcrot, D., Wang, F. Y., Jones, S., Shulok, J. et al. (2002). Small-molecule modulators of Hedgehog signaling: identification and characterization of Smoothened agonists and antagonists. *J. Biol.* **1**, 10.
Fuccillo, M., Rallu, M., McMahon, A. P. and Fishell, G. (2004). Temporal requirement for hedgehog signaling in ventral telencephalic patterning. *Development* **131**, 5031-5040.
Fuccillo, M., Rutlin, M. and Fishell, G. (2006). Removal of Pax6 partially rescues the loss of ventral structures in Shh null mice. *Cereb. Cortex* **16 Suppl. 1**, i96-i102.
Gelman, D. M., Martini, F. J., Nóbrega-Pereira, S., Pierani, A., Kessaris, N. and Marín, O. (2009). The embryonic preoptic area is a novel source of cortical GABAergic interneurons. *J. Neurosci.* **29**, 9380-9389.
Geng, X., Speirs, C., Lagutin, O., Inbal, A., Liu, W., Solnica-Krezel, L., Jeong, Y., Epstein, D. J. and Oliver, G. (2008). Haploinsufficiency of Six3 fails to activate Sonic hedgehog expression in the ventral forebrain and causes holoprosencephaly. *Dev. Cell* **15**, 236-247.
Graham, V., Khudyakov, J., Ellis, P. and Pevny, L. (2003). SOX2 functions to maintain neural progenitor identity. *Neuron* **39**, 749-765.
Guillemot, F. (2007). Spatial and temporal specification of neural fates by transcription factor codes. *Development* **134**, 3771-3780.
Gulacsi, A. and Anderson, S. A. (2006). Shh maintains Nkx2.1 in the MGE by a Gli3-independent mechanism. *Cereb. Cortex* **16 Suppl. 1**, i89-i95.
Gutin, G., Fernandes, M., Palazzolo, L., Paek, H., Yu, K., Ornitz, D. M., McConnell, S. K. and Hébert, J. M. (2006). FGF signalling generates ventral telencephalic cells independently of SHH. *Development* **133**, 2937-2946.
Hamdan, H., Liu, H., Li, C., Jones, C., Lee, M., deLemos, R. and Minoo, P. (1998). Structure of the human Nkx2.1 gene. *BBA* **1396**, 336-348.
Hébert, J. M. and Fishell, G. (2008). The genetics of early telencephalon patterning: some assembly required. *Nat. Rev. Neurosci.* **9**, 678-685.
Hébert, J. M. and McConnell, S. K. (2000). Targeting of cre to the Foxg1 (BF-1) locus mediates loxP recombination in the telencephalon and other developing head structures. *Dev. Biol.* **222**, 296-306.
Jeong, Y., El-Jaick, K., Roessler, E., Muenke, M. and Epstein, D. J. (2006). A functional screen for sonic hedgehog regulatory elements across a 1 Mb interval identifies long-range ventral forebrain enhancers. *Development* **133**, 761-772.
Jeong, Y., Leskow, F. C., El-Jaick, K., Roessler, E., Muenke, M., Yocum, A., Dubourg, C., Li, X., Geng, X., Oliver, G. et al. (2008). Regulation of a remote Shh forebrain enhancer by the Six3 homeoprotein. *Nat. Genet.* **40**, 1348-1353.
Kelberman, D., de Castro, S. C., Huang, S., Crolla, J. A., Palmer, R., Gregory, J. W., Taylor, D., Cavallo, L., Faienza, M. F., 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.
Kishi, M., Mizuseki, K., Sasai, N., Yamazaki, H., Shiota, K., Nakanishi, S. and Sasai, Y. (2000). Requirement of Sox2-mediated signalling for differentiation of early Xenopus neuroectoderm. *Development* **127**, 791-800.
Kondoh, H. and Kamachi, Y. (2010). SOX-partner code for cell specification: Regulatory target selection and underlying molecular mechanisms. *Int. J. Biochem. Cell Biol.* **42**, 391-399.
Lagutin, O. V., Zhu, C. C., Kobayashi, D., Topczewski, J., Shimamura, K., Puelles, L., Russell, H. R., McKinnon, P. J., Solnica-Krezel, L. and Oliver, G. (2003). Six3 repression of Wnt signaling in the anterior neuroectoderm is essential for vertebrate forebrain development. *Genes Dev.* **17**, 368-379.
Markram, H., Toledo-Rodriguez, M., Wang, Y., Gupta, A., Silberberg, G. and Wu, C. (2004). Interneurons of the neocortical inhibitory system. *Nat. Rev. Neurosci.* **5**, 793-807.
Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A. A. et al. (2007). Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat. Cell Biol.* **9**, 625-635.
Matsushima, D., Heavner, W. and Pevny, L. H. (2011). Combinatorial regulation of optic cup progenitor cell fate by SOX2 and PAX6. *Development* **138**, 443-454.
Miyagi, S., Masui, S., Niwa, H., Saito, T., Shimazaki, T., Okano, H., Nishimoto, M., Muramatsu, M., Iwama, A. and Okuda, A. (2008). Consequence of the loss of Sox2 in the developing brain of the mouse. *FEBS Lett.* **582**, 2811-2815.
Nóbrega-Pereira, S., Kessaris, N., Du, T., Kimura, S., Anderson, S. A. and Marín, O. (2008). Postmitotic Nkx2-1 controls the migration of telencephalic interneurons by direct repression of guidance receptors. *Neuron* **59**, 733-745.
Pan, Q., Li, C., Xiao, J., Kimura, S., Rubenstein, J., Puelles, L. and Minoo, P. (2004). In vivo characterization of the Nkx2.1 promoter/enhancer elements in transgenic mice. *Gene* **331**, 73-82.
Pevny, L. H. and Nicolis, S. K. (2010). Sox2 roles in neural stem cells. *Int. J. Biochem. Cell Biol.* **42**, 421-424.
Que, J., Luo, X., Schwartz, R. J. and Hogan, B. L. (2009). Multiple roles for Sox2 in the developing and adult mouse trachea. *Development* **136**, 1899-1907.
Roessler, E. and Muenke, M. (2010). The Molecular Genetics of Holoprosencephaly. *Am. J. Med. Genet.* **154C**, 52-61.
Sánchez-Camacho, C. and Bovolenta, P. (2008). Autonomous and non-autonomous Shh signalling mediate the in vivo growth and guidance of mouse retinal ganglion cell axons. *Development* **135**, 3531-3541.
Scott, C. E., Wynn, S. L., Sesay, A., Cruz, C., Cheung, M., Gomez Gaviro, M. V., Booth, S., Gao, B., Cheah, K. S., Lovell-Badge, R. et al. (2010). SOX9 induces and maintains neural stem cells. *Nat. Neurosci.* **13**, 1181-1189.
Sisodiya, S. M., Ragge, N. K., Cavalleri, G. L., Hever, A., Lorenz, B., Schneider, A., Williamson, K. A., Stevens, J. M., Free, S. L., Thompson, P. J. et al. (2006). Role of SOX2 mutations in human hippocampal malformations and epilepsy. *Epilepsia* **47**, 534-542.
Sousa, V. H. and Fishell, G. (2010). Sonic hedgehog functions through dynamic changes in temporal competence in the developing forebrain. *Curr. Opin. Genet. Dev.* **20**, 391-399.
Suh, H., Consiglio, A., Ray, J., Sawai, T., D'Amour, K. A. and Gage, F. H. (2007). In vivo fate analysis reveals the multipotent and self-renewal capacities of Sox2+ neural stem cells in the adult hippocampus. *Cell Stem Cell* **1**, 515-528.
Sur, M. and Rubenstein, J. L. (2005). Patterning and plasticity of the cerebral cortex. *Science* **310**, 805-810.
Sussel, L., Marin, O., Kimura, S. and Rubenstein, J. L. (1999). Loss of Nkx2.1 homeobox gene function results in a ventral to dorsal molecular respecification within the basal telencephalon: evidence for a transformation of the pallidum into the striatum. *Development* **126**, 3359-3370.
Taranova, O. V., Magness, S. T., Fagan, B. M., Wu, Y., Surzenko, N., Hutton, S. R. and Pevny, L. H. (2006). SOX2 is a dose-dependent regulator of retinal neural progenitor competence. *Genes Dev.* **20**, 1187-1202.

Toledo-Rodriguez, M., Goodman, P., Illic, M., Wu, C. and Markram, H. (2005). Neuropeptide and calcium-binding protein gene expression profiles predict neuronal anatomical type in the juvenile rat. *J. Physiol.* **567**, 401-413.

Wegner, M. (2010). All purpose Sox: The many roles of Sox proteins in gene expression. *Int. J. Biochem. Cell Biol.* **42**, 381-390.

Wegner, M. and Stolt, C. C. (2005). From stem cells to neurons and glia: a Soxist's view of neural development. *Trends Neurosci.* **28**, 583-588.

Xu, Q., Wonders, C. P. and Anderson, S. A. (2005). Sonic hedgehog maintains the identity of cortical interneuron progenitors in the ventral telencephalon. *Development* **132**, 4987-4998.

Xu, Q., Guo, L., Moore, H., Waclaw, R. R., Campbell, K. and Anderson, S. A. (2010). Sonic hedgehog signaling confers ventral telencephalic progenitors with distinct cortical interneuron fates. *Neuron* **65**, 328-340.

Zappone, M. V., Galli, R., Catena, R., Meani, N., De Biasi, S., Mattei, E., Tiveron, C., Vescovi, A. L., Lovell-Badge, R., Ottolenghi, S. et al. (2000). Sox2 regulatory sequences direct expression of a (beta)-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.

Zhao, L., Zevallos, S. E., Rizzoti, K., Jeong, Y., Lovell-Badge, R. and Epstein, D. J. (2012). Disruption of SoxB1-dependent Sonic hedgehog expression in the hypothalamus causes septo-optic dysplasia. *Dev. Cell* **22**, 585-596.

CHAPTER 4:

Nucleic Acids Research 40, 6461-6476 (2012)

Emx2 is a dose-dependent negative regulator of *Sox2* telencephalic enhancers

Jessica Mariani[1], Rebecca Favaro[1], Cesare Lancini[1], Giulia Vaccari[2], Anna L. Ferri[1], Jessica Bertolini[1], Diletta Tonoli[1], Elisa Latorre[1], Roberta Caccia[1], Antonella Ronchi[1], Sergio Ottolenghi[1], Satoru Miyagi[3], Akihiko Okuda[3], Vincenzo Zappavigna[2], Silvia K. Nicolis[1]

[1]Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza 2, 20126 Milano, Italy.
[2]Department of Animal Biology, University of Modena and Reggio Emilia, Via G. Campi 213/d, Modena 41100, Italy.
[3]Division of Developmental Biology, Research Center for Genomic Medicine, Saitama Medical School, Saitama 350-1241, Japan.

Emx2 is a dose-dependent negative regulator of Sox2 telencephalic enhancers

J. Mariani[1], R. Favaro[1], C. Lancini[1], G. Vaccari[2], A. L. Ferri[1], J. Bertolini[1], D. Tonoli[1], E. Latorre[1], R. Caccia[1], A. Ronchi[1], S. Ottolenghi[1], S. Miyagi[3], A. Okuda[3], V. Zappavigna[2] and S. K. Nicolis[1,*]

[1]Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza 2, 20126 Milano, [2]Department of Animal Biology, University of Modena and Reggio Emilia, Via G. Campi 213/d, Modena 41100, Italy and [3]Division of Developmental Biology, Research Center for Genomic Medicine, Saitama Medical School, Saitama 350-1241, Japan

Received June 9, 2011; Revised March 16, 2012; Accepted March 19, 2012

ABSTRACT

The transcription factor Sox2 is essential for neural stem cells (NSC) maintenance in the hippocampus and *in vitro*. The transcription factor Emx2 is also critical for hippocampal development and NSC self-renewal. Searching for 'modifier' genes affecting the Sox2 deficiency phenotype in mouse, we observed that loss of one Emx2 allele substantially increased the telencephalic β-geo (LacZ) expression of a transgene driven by the 5′ or 3′ Sox2 enhancer. Reciprocally, Emx2 overexpression in NSC cultures inhibited the activity of the same transgene. *In vivo*, loss of one Emx2 allele increased Sox2 levels in the medial telencephalic wall, including the hippocampal primordium. In hypomorphic Sox2 mutants, retaining a single 'weak' Sox2 allele, Emx2 deficiency substantially rescued hippocampal radial glia stem cells and neurogenesis, indicating that Emx2 functionally interacts with Sox2 at the stem cell level. Electrophoresis mobility shift assays and transfection indicated that Emx2 represses the activities of both Sox2 enhancers. Emx2 bound to overlapping Emx2/POU-binding sites, preventing binding of the POU transcriptional activator Brn2. Additionally, Emx2 directly interacted with Brn2 without binding to DNA. These data imply that Emx2 may perform part of its functions by negatively modulating Sox2 in specific brain areas, thus controlling important aspects of NSC function in development.

INTRODUCTION

The transcription factor Sox2, essential in pluripotent stem cells of the blastocyst (1), is also highly expressed in neural stem cells (NSC) and their early progeny (2–6). Decreased expression of Sox2 in a mouse hypomorphic Sox2 mutant causes important brain and neurologic defects (3,7), which mimic significant aspects of the pathology of Sox2-deficient patients (8,9). In this hypomorphic mutant, we combined the deletion of one Sox2 allele ($Sox2^{\beta\text{-geo}}$ knock-in) with the deletion, on the other allele, of an upstream enhancer of Sox2 ($Sox2^{\Delta Enh}$), important for its expression in telencephalic NSC (3,6,10–13). The hypomorphic mutant, expressing Sox2 at a level about 30% that of the wild-type, shows hippocampal stem cells loss, corpus callosum interruption, parenchymal loss in striatum and thalamus, decreased numbers of GABAergic neurons and neurological defects, including epilepsy (3,7). Recently (14), we showed that Sox2 embryonic deletion leads to complete perinatal loss of hippocampal stem cells. NSC from the forebrain of such mutants become rapidly exhausted in *in vitro* neurosphere culture.

The Emx2 transcription factor is expressed in the developing dorsal telencephalon, including prospective hippocampus and cerebral cortex, from early embryogenesis (15,16). Its expression is maintained postnatally in brain neurogenic regions, the subventricular zone (SVZ) and hippocampus dentate gyrus (DG) (17,18).

Emx2 inactivation in mouse causes delayed hippocampal development, with reduced cerebral cortex and abnormal specification of cortical areas (15,19–21). *In vitro*, mutant $Emx2^{-/-}$ NSC show increased proliferation in long-term neurosphere cultures (17).

*To whom correspondence should be addressed. Tel: +39 02 6448 3339 (3315); Fax: +39 02 6448 3565; Email: silvia.nicolis@unimib.it

A common aspect of the defects in Sox2 and Emx2 mutants is the abnormal hippocampal development, together with important NSC abnormalities in *in vitro* long-term culture (see above, and 3,14,15,17,21). In this work, we explored potential functional interactions between Sox2 and Emx2 at the molecular level and, *in vivo*, in mouse. We report that Emx2 negatively regulates two Sox2 telencephalic-specific enhancers *in vivo* and in transfection assays, by interfering with binding of positive regulators to their cognate sites within the enhancers. *In vivo*, Emx2 deficiency leads to some increase of Sox2 in the medial wall of the telencephalon, and partially counteracts hippocampal neurogenesis defects observed in Sox2 deficient (hypomorphic) mouse mutants.

MATERIALS AND METHODS

Mouse lines, X-gal staining and immunohistochemistry

The 5′ and 3′ enhancer- β-geo transgenic mice lines were described in (6,11,22,23). The Sox2-hypomorphic (Sox2 $^{\Delta Enh}$) and null (Sox2 $^{\beta\text{-geo}}$) mutant alleles were as in (3). The Emx2 null mutant mice (kindly provided by A. Mallamaci) were described in (15).

X-gal staining, immunohistochemistry (IHC) and histology were as reported (6).

Glial fibrillary acidic protein (GFAP)/nestin and BrdU IHC on hippocampus and all histological analyses were carried out as previously reported (3). IHC with anti-Emx2 antibodies was as described (18). IHC with anti-Brn2 antibody, a SantaCruz goat antibody (22) was used (1:100).

Experimental procedures involving animals were approved by the Italian Ministry of Health.

Transgenic neurosphere culture and lentiviral transduction

Neurosphere cultures were derived from E15.5 dorsal telencephalon of transgenic brains as described (6,14), expanded (in the presence of 400 µg/ml G418) and transduced (in the absence of G418) with an internal ribosome entry site-green fluorescent protein (Emx2-IRES-GFP) (or GFP-only control)-encoding lentivirus at a multiplicity of infection of 5. To generate the Emx2-transducing virus, the Emx2 coding sequence was cloned, upstream to IRES-GFP, in place of the Sox2 coding sequence, in the lentiviral vector described in refs. 7 and 14.

Two passages after transduction RNA was extracted, reverse transcribed and analysed by real-time PCR (MESA GREEN qPCR Master Mix Eurogentec) for the expression of β-geo (lacZ), Emx2 and Sox2 with the following primers: LacZ-f CTGGATCAAATCTGTCGATCC, LacZ-r CGTATTCGCAAAGGATCAGC, Emx2-f GTC CCAGCTTTTAAGGCTAGA, Emx2-r CTTTTGCCTT TTGAATTTCGTTC, Sox2-f GGCAGCTACAGCATG ATGCAGGAGC, Sox2-r CTGGTCATGGAGTTGTAC TGCAGG; HPRT-f TCCTCCTCAGACCGCTTT, HPRT-r CCTGGTTCATCATCGCTAATC; the dataset are analysed with a 7500 System Software v1.4 (Applied Biosystem). Neurospheres cultured as above from Emx2$^{-/-}$ or wild-type embryonic brains were expanded for one passage, total RNA was extracted and analysed by real-time RT–PCR as above with the same Sox2 and Emx2 primers. Expression levels normalized versus HPRT expression.

Luciferase reporter constructs

The Sox2 5′ telencephalic enhancer core region of 400 bp was PCR amplified from the 0.4a-Sox2 promoter-β-geo vector (22) using the following primers:

Fw: 5′ CGAGGTACCGTCAAATAGGGCCCTTTT CAG 3′; Rv: 5′ TATCTCGAGAAGCCAACTGACA ATGTTGTGG 3′ containing a KpnI and XhoI restriction sites (underlined), for further cloning into the pGL3-based plasmid containing a 215 bp minimal tk promoter (a gift of Hitoshi Niwa) (5′enh-tk-luc).

The reporter plasmid carrying mutations in the ATTA-3 site (ATTA-3 site mut) was obtained as above, starting from 0.4a-MUT Sox2 promoter-β-geo vector (22). All other ATTA-site mut reporter constructs were similarly obtained by PCR-based site-directed mutagenesis.

Primers for the ATTA-1/2 site mut plasmid:

Fw: 5′CGAGGTACCGTCAAATAGGGCCCTTTT CAGATTTTA**AGGA**CAAAATAAA**AGGA**GTCTGC TC3′

Rv: 5′TATCTCGAGAAGCCAACTGACAATGTT GTGG 3′ containing the desidered mutations (in bold) and KpnI/XhoI restriction sites (underlined).

The ATTA-4 site mut plasmid was generated by replacing a PstI cassette of the 5′enh-tk-luc with a corresponding cassette, containing the desired mutation (in bold), obtained after amplification with the following primers:

Fw: 5′ ACTCTGCAGGTCCCCTGCCGTTCGCCTTC ATTTCCATA**AGGA**GAGGAGGAGAGGAGG 3′

Rv: 5′ CGGGTCGCTGCAGGGTCGCTCGGTGTTC G 3′

PstI restriction site (underlined) in both primers.

The ATTA-5/6 sites mut plasmid was generated using two overlapping primers containing the desidered mutations (in bold) to separately amplify the 5′- and 3′-portions of the 5′ enhancer, in conjunction with external primers flanking the KpnI and XhoI sites of the 5′enh-tk-luc vector. The full mutated enhancer was obtained by reamplification of the obtained fragments with the same external 5′enh-tk-luc primers. The sequences of the primers used are:

Fw1: 5′ GCATCAACCTAGTAAGATGCTTGGCTAG TTCTCGCTA**AGGT**CTGCAAC 3′

Rv1 (XhoI-external primer): 5′ TATCTCGAGAAGCCA ACTGACAATGTTGTGG 3′

Fw2 (KpnI-external primer): 5′ CGAGGTACCGTCAA ATAGGGCCCTTTTCAG 3′

Rv2: 5′ GTTGCAGACCTTAGCGAGAACTAGCCAA GCATCTTACTA**GGT**TGATGC 3′

The reporter plasmid carrying mutations in five ATTA-sites was obtained by combining the mutations via PCR.

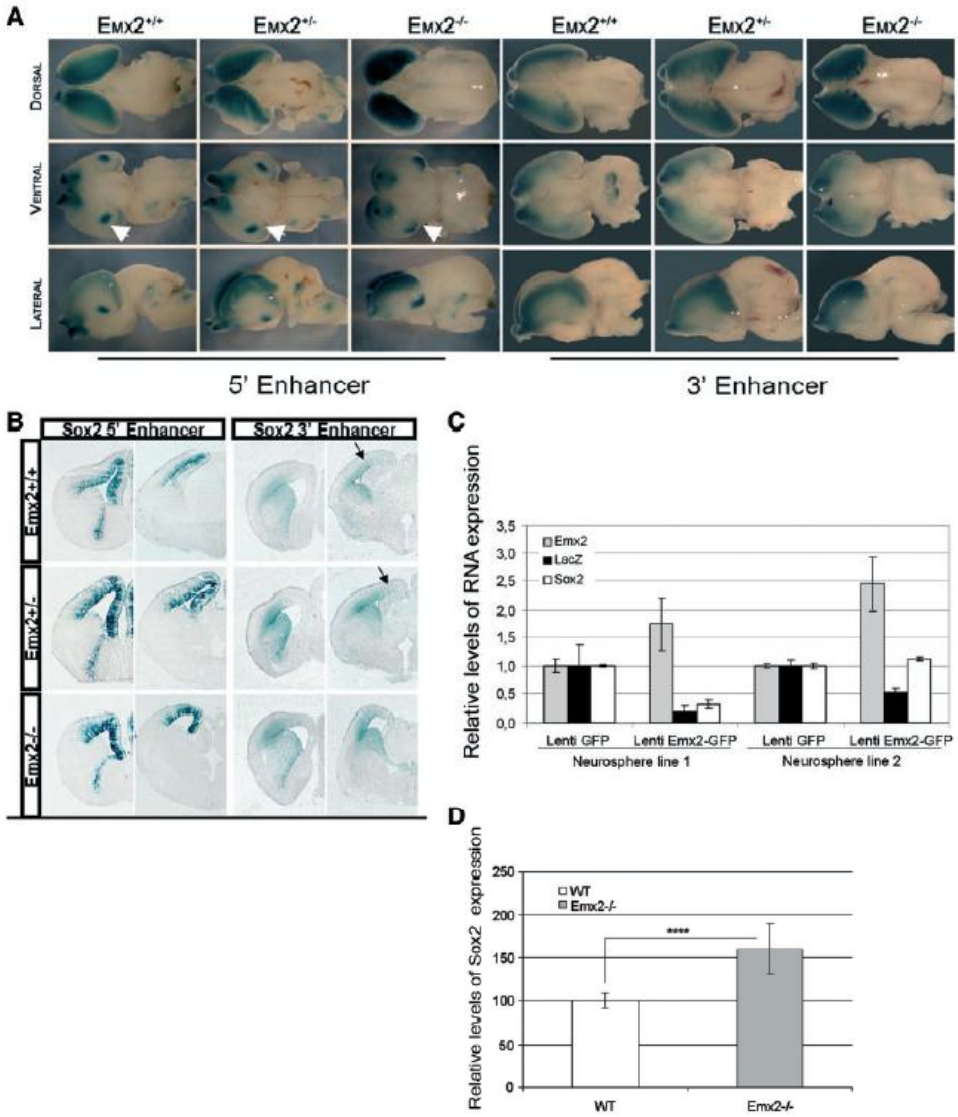


Figure 1. Emx2 deficiency increases activity of Sox2 telencephalic enhancers-driven lacZ transgenes. (**A**) X-gal stained E15.5 brains carrying β-geo transgenes driven by the 5′ Sox2 telencephalic enhancer (left) or by the 3′ enhancer (right), of $Emx2^{+/+}$, $Emx2^{+/-}$, or $Emx2^{-/-}$ genotype, as indicated. Dorsal (top row), ventral (middle row) and lateral (bottom row) views are shown. Increased X-gal staining is seen, most clearly in dorsal views, in $Emx2^{+/-}$ as compared to $Emx2^{+/+}$ brains, and in $Emx2^{-/-}$ as compared to $Emx2^{+/-}$ brains. In the 5′enhancer-transgenic brains, an X-gal-positive spot on the ventral telencephalic vesicles, visibile in the ventral (arrow) and lateral views, has comparable intensity in $Emx2^{+/+}$ and $Emx2^{+/-}$ brains, acting as an internal control for staining. Overall, 7/7 $Emx2^{+/-}$ transgenic embryos (5′ construct, E15.5) showed increased lacZ expression relative to $Emx2^{+/+}$ from the same litter (4 embryos). Similarly, 7/8 $Emx2^{+/-}$ embryos carrying the 3′ transgene showed increased lacZ activity relative to $Emx2^{+/+}$ controls (4 embryos). Homozygous $Emx2^{-/-}$ 5′ transgenic embryos were always (7/7) more intensely stained than their control heterozygotes ($Emx2^{+/-}$) littermates (11 embryos); 7/7 of the $Emx2^{-/-}$ 3′ transgenics were more stained than their $Emx2^{+/-}$ heterozygous controls (10 embryos). (**B**) X-gal stained brain coronal sections of 5′ or 3′ enhancer-lacZ transgenic forebrains of $Emx2^{+/+}$ (top row), $Emx2^{+/-}$ (middle) and $Emx2^{-/-}$ (bottom) genotype. Arrow in B (3′ enhancer) points to some dorsal expansion of X-gal staining signal in $Emx2^{+/-}$, as compared to $Emx2^{+/+}$ brain. (**C**) Relative RNA levels (real-time RT–PCR) of Emx2, β-geo (lacZ), and endogenous Sox2 in cultured NSC (neurospheres) from Sox2 3′ enhancer-β-geo transgenic brains, transduced with Emx2-GFP or GFP (control)—encoding lentiviruses, as indicated. RNA levels in control (GFP-lenti-transduced) cells are set = 1 (for non-normalized data, see Supplementary Figure S1). The values represent the mean ± SD of $n = 2$ independent RT–PCR experiments on each line performed in triplicate (all RNA levels—LacZ, Emx2, Sox2—significantly differ between Lenti-Emx2 and Lenti-GFP transductions ($P < 0.003$ by Student's t-test), except for endogenous Sox2 levels in line 2, which are comparable). (**D**) Relative RNA levels (real-time RT–PCR) of endogenous Sox2 RNA in cultured NSC (neurospheres) from $Emx2^{-/-}$ versus wild type embryonic brains. The values represent the mean ± SD of $n = 3$ independent RT–PCR experiments on two wild-type and three Emx2-mutant independent cultures tested, each performed in triplicate. (****$P = 6.3\,e^{-11}$ by Wilcoxon's one-tailed test).

For constructing the 3X POU/ATTA site 3 plasmid, the combined POU/ATTA site (in bold type characters) was multimerized to three copies, and subcloned into the KpnI/XhoI site of the pGL3-tk luciferase vector, using the following primers:

Fw: 5′CACTGCTAATTAGCAATGCTAGGGTGCTAA**TTAGCAATGCTAGGGTGCTAATTAGCAATGCT**AGC 3′

Rv: 5′TCGAGCTAGCATTGCTAATTAGCACCCTAG**CATTGCTAATTAGCACCCTAGCATTGCTAAT**TAGCAGTGGTAC 3′

For constructing the 2X ATTA site 1,2 plasmid, the ATTA site 1,2 core sequence,

5′ TTAATTACAAAATAAAATTAGTCTGCTCTTC 3′, was dimerized (as a synthetic oligonucleotide) and subcloned into the KpnI/XhoI site of the pGL3-tk luciferase vector.

The Luciferase reporter vectors bearing BamHI/SalI genomic DNA fragments of the 3′enhancer were described (11,23); their core sequence was essentially as in (24): 5′GGATCCCTAATTAATGCAGAGACTCTA AAAGAATTTCCCGGGCTCGGGCAGCCATTGTGA TGCATATAGGATTATTCACGTGGTAATGAGCACA GTCGAC 3′

These fragments were subcloned into the BamHI/SalI site located 3′ to the Luciferase gene.

The Nestin258-luciferase construct (a gift from H. Kondoh) was previously described (25).

P19 transfection assays

For transfection experiments, P19 cells were grown in MEM-ALPHA medium supplemented with PenStrep, L-glutamine and 10% fetal bovine serum. 2×10^5/well exponentially growing P19 cells were plated in 6-well-plates, and transfected the following day with Lipofectamine 2000 (Invitrogen) according to supplier's instructions. Briefly, medium in each well was replaced with 1 ml of Opti-MEM medium (Invitrogen) with 10 µl of Lipofectamine 2000 (Invitrogen), with DNA. For transfection experiments, we used 1 µg of the luciferase reporter plasmid, and 500 ng of the Emx2 expression vector (pCAGGS-Emx2, a gift from V. Broccoli) per well if not otherwise indicated. In control experiments, equimolar amounts of the Emx2 'empty' vector (pCAGGs) or the Otx2 expression vector (pCAGGS-Otx2, a gift from V. Broccoli), were used. The pBluescript vector was added to each transfection to equalize the total amount of transfected DNA to 2 µg total in each reaction. After 24 h, total cellular extracts were prepared and Luciferase activity was measured according to the Promega Luciferase reporter system protocol.

For cotransfection experiments with Brn2 and Emx2 expression vectors, Brn2 expression vector (a gift from D. Mejiers) (or the 'empty' control) was transfected at the fixed amount of 500 ng/transfection or at increasing amounts (+, ++, +++: 125, 500, 1000 ng) where indicated; in these experiments, Emx2 expression vector was added (100 to 1000 ng) as indicated in the relevant Figures. The 'empty' vector was added to each transfection at the proper concentration to equalize the total amount of DNA transfected in each reaction to 2 µg. Sox2 expression vector (activating the Nestin258-luciferase construct in conjunction with Brn2, ref. 25) was added at the fixed amount of 500 ng/transfection.

In vitro protein expression and purification

Emx2 (in pSG5), Brn2, GATA1 and GATA2 (in pBluescript) were produced using *in vitro* transcription—translation reticulocyte lysate system (TNT, Promega), according to the manufacturer's indication, in a total volume of 50 µl for 1.5 h at 30°C, using 2 µg plasmid template with 25 µl of reticulocyte lysate, and then frozen at −80°C.

The amounts in µl of the TNT reactions used in different experiments are indicated in Figure legends. To use equivalent amounts of *in vitro*-synthethized proteins (Brn2, Emx2, GATA1 or GATA2), TNT reactions were performed in the presence of 35S methionine, the amounts of protein produced were estimated by autoradiography of western blot, normalized for the numbers of methionines in each protein, and equivalent amounts of each recombinant protein were used.

The Emx2 (or CP2 control, ref. 26) cDNAs were cloned in frame into the pGEX2T vector. The *Escherichia coli* BL21 strain cells were transformed with the above plasmid and cultures were grown at midlogarithmic phase (0.6 A_{600}). Protein expression was induced with 0.1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) for 3 h at 37°C. The GST-EMX2 protein present in the soluble fraction was bound to GST-Sepharose 4B (Amersham Bioscience) and purified according to the manufacturer's instructions.

Protein was eluted from sepharose, quantitated by Coomassie blue staining in comparison to BSA standards, and 1 µg of total protein (for GST-Emx2, GST-CP2 and GST-only resins) was used for GST-pulldown of ^{35}S Brn2-containing TNT reaction as in (26,27).

Electrophoretic mobility shift assay and Chromatin Immunoprecipitation

Electrophoretic mobility shift assay (EMSA) was performed (28,29) by preincubating TNT-produced proteins or nuclear extract (from the hippocampal stem cell line AHP or from neurosphere cultures) for 30 min on ice in 20 µl of binding buffer (75 mM NaCl, 20% Ficoll, 10 mM Tris–HCl, pH 7.5, 0.5 mM EDTA, 10 mM DTT, 1 µg of poly(dI-dC), together with 2 µl (5×10^4 cpm) of ^{32}P-end-labelled oligonucleotide probes. The incubation mixture was resolved by electrophoresis on a 5 or 6% polyacrylamide gel (29:1, acrylamide/bisacrylamide ratio) in 50 mM Tris borate, 1 mM EDTA, pH 8.2 (0.5XTBE) buffer run at 4°C at 150 V for 3 h. Gel were dried and exposed to a Kodak X-AR film at −80°C. For 'supershift' reactions, 1 µl of the 1:10 diluted mouse α-Emx2 antibody (mouse ascites, kindly provided by F. Mavilio) or 8 µl of the goat α-Brn2 antibody (undiluted) (Santa Cruz Biotechnology) were added to the complete binding reaction just prior to the addition of the labelled probe.

The following double-stranded oligonucleotides were used as probes for EMSA (only the top strand is shown)

(underlined sequences correspond to mutated nucleotides):

ATTA-site 3: 5′-TCGTCAAACTCTGCTAATTAGCAA TGCTGAGAAA-3′;
ATTA-site 3 mut1: 5′-TCGTCAAACTCTGC ATCCTTGCAGAGCTGAGAAA-3′;
ATTA-site 3 mut2: 5′-TCGTCAAACTCTGCTACGGCG CAATGCTGAGAAA-3′;
3′-Enh: 5′-GGCAGGTTCCCCTCTAATTAATGCAGA GACTC-3′;

ATTA-1/2 sites:

5′-GGGCCCTTTTCAGATTTTAATTACAAAATAAA ATTAGTCTGCTCTTCCTCGG-3′;

ATTA-1/2 sites mut:

5′-GGGCCCTTTTCAGATTTTAAGGACAAAATAA AAGGAGTCTGCTCTTCCTCGG-3′;

Delta1-Enh: 5′-AGAGAGCAGGTGCTGTCTGCATT ACCATACAGCTGAGCGC-3′;
Nestin-Enh: 5′-GTGTGGACAAAAGGCAATAATT AGCATGAGAATCGGCCTC-3′.

Chromatin immunoprecipitation (ChIP) was as described (13).

RESULTS

Emx2 negatively regulates transgenic Sox2-β-geo reporters

We initially bred mice carrying transgenic β-geo reporters driven by Sox2 regulatory elements to Emx2 mutant ($Emx2^{+/-}$) mice. The Sox2-β-geo transgene (6) is driven by 5.7 kb of the Sox2 promoter/enhancer, and its neural expression is progressively confined to the telencephalon, after E11.5. The SRR2 transgene (11,23) is driven by the tk-promoter linked to an enhancer normally located immediately 3′ to the Sox2 coding region (these mouse lines are denominated 5′ and 3′ enhancer lines, respectively; alternative names in the literature for the 5′ and 3′ enhancers include SRR1 and N2, and SRR2, respectively, 11,24,30). Breeding with Emx2-mutant mice, we obtained E15.5 progeny consisting of embryos carrying the transgene in the heterozygous state, together with the three possible Emx2 genotypes (wild-type, +/+; heterozygote, +/−; homozygote, −/−).

For both constructs, loss of one Emx2 allele is associated to significantly increased β-geo expression (evaluated by classical X-gal staining) (Figure 1A); a further strong increase is observed in $Emx2^{-/-}$ mice (note, however, that the $Emx2^{-/-}$ brain is abnormal, as expected (15).

We confirmed these results by X-gal staining of brain sections (Figure 1B). The 5′enhancer construct is expressed in dorsal and medial areas of the telencephalic ventricular zone and, to lower levels, ventrally, along the ganglionic eminence, whereas the 3′ enhancer construct is more active in ventrolateral areas. In $Emx2^{+/-}$ heterozygotes, the respective domains of expression were more intensely stained, both anteriorly and posteriorly; additionally, the extension of the X-gal-positive region was somewhat increased towards the midline, in mice carrying the 3′enhancer construct (arrows). As expected, homozygous $Emx2^{-/-}$ mutants showed increased β-geo expression, although matching the different areas is problematic due to morphological abnormalities. We conclude from these Emx2 loss-of-expression experiments that Emx2 negatively modulates two different telencephalic enhancers of Sox2 in the developing mouse brain.

We further asked whether abnormally increased levels of Emx2 could inhibit Sox2 telencephalic enhancers in neural cells. To this end, we derived independent NSC cultures from the dorsal telencephalon of two E15.5 mice carrying the 3′enhancer-β-geo construct, and we transduced them with an Emx2-GFP-expressing lentivirus (or with control GFP-expressing virus). In both cell populations, the β-geo reporter expression was strongly inhibited by the Emx2-expressing virus, as compared to the control virus (Figure 1C). In one of the two lines, which expressed moderate levels of Sox2, also the endogenous Sox2 level was significantly decreased; in the second line, which showed a much higher expression of endogenous Sox2, no significant inhibition could be observed (Figure 1C and Supplementary Figure S1); it is possible that the 'isolated' enhancer guiding β-geo more readily responds to acute, exogenous variations of Emx2 levels than the 'full' Sox2 locus, which is controlled by various different regulatory regions. Taken together, the above results indicate that Emx2 negatively regulates, *in vivo* and in *ex vivo* derived neural progenitors, the activities of Sox2' telencephalic enhancers.

We also wished to address whether Emx2 deficiency would affect endogenous Sox2 levels in NSC cultures, as it does *in vivo* in the developing brain. To this end, we measured endogenous Sox2 mRNA levels in NSC cultures derived from Emx2-mutant telencephalon, as compared to wild-type littermates cultures, by real-time RT–PCR (Figure 1D). Emx2 ablation led to a significant (average 50%) increase in endogenous Sox2 levels in three independent mutant cultures tested (Figure 1D and Supplementary Figure S2), confirming that expression levels of the resident Sox2 gene are modulated by Emx2 in cultured neural stem/progenitor cells.

Loss of a single Emx2 allele significantly rescues the hippocampal NSC deficiency of hypomorphic Sox2 mutant mice

To begin to address whether the Emx2-dependent inhibition of Sox2 telencephalic regulatory elements has any *in vivo* effects on Sox2-dependent brain phenotypes, we selected for further studies the hippocampus neural stem/progenitor cells of the hypomorphic $Sox2^{\beta\text{-geo}/\Delta Enh}$ mutant (3,7), that expresses Sox2 (from the single residual knock-down allele) at low levels (30% relative to normal, whereas heterozygotes express 65%) (3,7). In these mice, postnatal neurogenesis is strongly diminished, particularly in the hippocampus. In particular, the number of nestin/GFAP double-positive radial glia cells (a stem/

progenitor cell expressing Sox2 (3,5) is drastically decreased (3).

In Sox2 hypomorphic mutants, heterozygosis for a mutated Emx2 allele was sufficient to substantially rescue the number of GFAP/nestin stem/progenitor cells from about 20% to 60% of wild-type levels (Figure 2A and B); additionally, the radial glia was converted from a thin, poorly-developed appearance typical of cells of the hypomorphic mutant, to quasi-normal morphology (Figure 2A). In agreement, BrdU incorporation (Figure 2B) was increased to about 45% of wild-type levels in $Sox2^{\beta\text{-geo}/\Delta Enh}$; $Emx2^{+/-}$, versus about 30% in $Sox2^{\beta\text{-geo}/\Delta Enh}$; $Emx2^{+/+}$ controls (even if loss of a single Emx2 allele, per se, causes some decrease of BrdU incorporation (Figure 2B, ref. 31, see section 'Discussion').

To interpret this result, we examined Sox2 expression in wild-type mice in the prospective hippocampal area during development. In this area, Sox2 and Emx2 are coexpressed in a large proportion of cells (Figure 2C). At E 15.5, both the medial and lateral walls of the telencephalon expressed Sox2; however the medial wall of the lateral ventricle, from which the hippocampus will originate, expressed Sox2 at comparatively lower levels than the lateral wall in the wild-type (Figure 2C, filled versus empty arrowheads). On the other hand, the Emx2 level was higher in the medial as compared to the lateral wall (Figure 2C, arrowheads see also 15,20), pointing to an inverse relation between Sox2 and Emx2 expression. Within the medial telencephalic wall (prospective hippocampus region), an inverse Sox2/Emx2 relation is also seen with an Emx2-high, Sox2-low region developing adjacent to a comparatively Emx2-low, Sox2-high region (Figure 2C, thin arrowheads; Supplementary Figure S3).

In $Emx2^{+/-}$ heterozygotes we noted a significant upregulation of Sox2 expression in the medial telencephalic, relative to the lateral wall, when compared to wild-type mice (Figure 2C, arrowheads). This inverse correlation suggests that, within the area from which the hippocampus will arise, Emx2 may negatively modulate Sox2 levels. This result is consistent with the possibility that the loss of a single Emx2 allele in Sox2 hypomorphic / $Emx2^{+/-}$ double mutants contributes, by upregulating the deficient Sox2 expression, to the observed radial glia rescue.

Emx2 transfection in Sox2-positive P19 teratocarcinoma cells inhibits the activity of reporter genes driven by the 5′ or 3′ Sox2 enhancer

The previous *in vivo* results, indicating that Emx2 somehow negatively modulates the Sox2 enhancers, raise the question whether Emx2 effects on Sox2 are direct or mediated by other factors. The 5′- and 3′-enhancers 'core' elements were previously defined *in vivo* by transgenic assays and, *in vitro*, by transfection in Embryonic Stem (ES) Cells (11,22,23). Both elements contain POU sites, known to be functionally important in ES and brain cells, which bind specific transcription factors (Oct4 in ES, Brn1 and Brn2 in neural cells) (11,22,23). In transgenic mice, ~400 nt of the 5′ enhancer recapitulate full expression, but as little as 120 nt are sufficient for some specific activity (22). The 400 nt enhancer contains, in addition to the two POU sites, several ATTA sites (referred to as ATTA-1 to ATTA-6, Figure 3A), which represent the core of potential homeobox transcription factor-binding motifs (22), including Emx2. The more 5′ POU site is combined with ATTA-3 site within a single overlapping sequence. The 3′ enhancer similarly contains several ATTA sites, together with a previously characterized POU-binding element (23) (Figure 3A).

To address the possibility that Emx2 directly affects Sox2 enhancer function, and to investigate its molecular mechanisms of action, we developed a simplified *in vitro* transfection system. We performed transfection experiments in P19 teratocarcinoma cells, using the 5′ and 3′ enhancer 'core' regions linked to a luciferase reporter. P19 cells express Sox2 at high levels, but are negative for both Emx2 and the putative neural Sox2 activators Brn1 and Brn2 (11,22), although they express the related POU factor Oct3/4, an activator of Sox2 in ES cells (22,23); this allows us to test for the effects of adding these exogenous factors in appropriate combinations and dosage, and to evaluate the effects on enhancer functions of different, specific point mutations within transcription factors recognition sites.

We first transfected into P19 cells a luciferase reporter gene, driven by the minimal tk promoter linked to the core 5′Sox2 enhancer, in the absence or presence of an Emx2-expression vector.

Emx2 cotransfection strongly repressed the activity of the enhancer, to a level just above that of the control enhancer-less tk-luciferase vector (Figure 3B). Cotransfection with a vector expressing Otx2, a related homeobox gene, or with empty vector gave no significant repression. Similarly, Emx2 repressed the activity of the 3′Sox2 telencephalic enhancer (11,23), when assayed with both a full size and a 'core' enhancer (22) construct (Figure 3C), though the observed repression was less pronounced than that observed with the 5′ enhancer. The repression caused by Emx2 was dose-dependent for both the 5′ and 3′ enhancers (Figure 3D).

To identify the site where Emx2 binds to repress transcription, we mutated, in different combinations, each of six sites characterized by the ATTA sequence in the 5′enhancer. Unexpectedly, all the mutations strongly decreased the activity (in the absence of cotransfected Emx2) (Figure 3E); the simultaneous mutation of five out of six sites (1/2/4/5/6, leaving only ATTA-3), essentially abolished the activity of the core enhancer (Figure 3E). In these experiments, Emx2 cotransfection further reduced the residual activity of the mutants to the background level corresponding to the activity of the tk-promoter-luciferase construct.

These experiments suggest that the mutation of the ATTA sites destroys the binding of some (yet unidentified) activator protein. In contrast, as the repressive Emx2 activity is not abolished by any of the mutations, Emx2 either binds to other unidentified sites, or somehow antagonizes the activator at each of the defined sites.

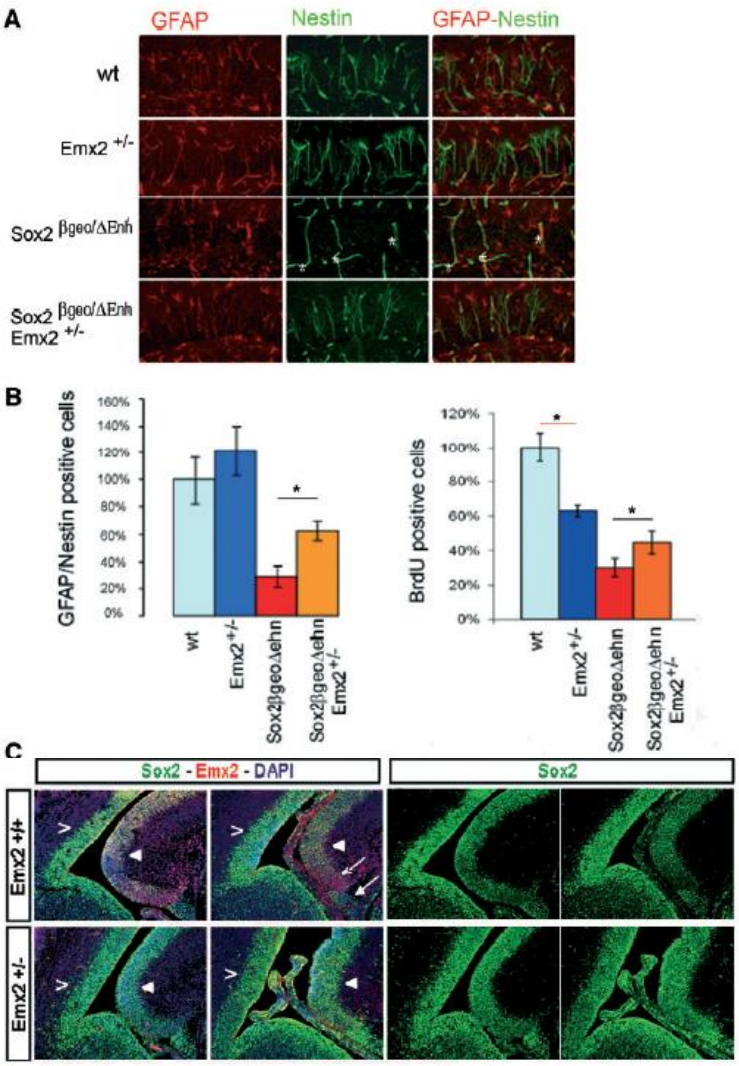


Figure 2. Emx2 deficiency ($Emx2^{+/-}$) rescues GFAP/nestin stem cells impairment in the hippocampus of Sox2-deficient ($Sox2^{\beta\text{-geo}/\Delta Enh}$) mutant mice. (**A**) GFAP/nestin double immunofluorescence of hippocampus DG in the indicated genotypes. GFAP/nestin-positive cells, strongly depleted in Sox2-hypomorphic ($Sox2^{\beta\text{-geo}/\Delta Enh}$) mutants, recover to a significant extent in $Sox2^{\beta\text{-geo}/\Delta Enh}$; $Emx2^{+/-}$ double mutants (asterisks mark vessels, showing non-specific fluorescence). (**B**) GFAP/nestin-positive cells and BrdU-positive cells. Wild-type is set = 100%. $n = 5$ mice per genotype; $^*P \leq 0.002$ by Student's t-test. (**C**) Double immunofluorescence with anti Emx2 (red) and anti Sox2 (green) antibodies on E15.5 telencephalic sections (confocal microscopy), in wild-type ($Emx2^{+/+}$, top) and $Emx2^{+/-}$ heterozygotes (two different representative mice per genotype, out of $n = 5$ mice analysed). The Sox2 (green) channel is also separately shown on the right panels. In $Emx2^{+/-}$ brains, compared to $Emx2^{+/+}$ controls, a comparative increase in the intensity of Sox2 staining is seen in the medial telencephalic wall (filled arrowhead) (comprising the hippocampal primordium region, arrows), as compared with the outer/lateral wall (empty arrowhead) within the same section. In the lower medial telencephalic wall, in the region of the prospective hippocampus, a boundary can be appreciated between two regions showing Sox2-high/Emx2 low (filled arrow) and comparatively Sox2-low/Emx2 high levels (non-filled arrow) (see also magnification and channel separation in Supplementary Figure S3).

Emx2 binds a composite POU/Emx2 binding-site (ATTA-3), and inhibits the binding of Brn2 to the same site

We characterized by EMSA the binding of recombinant Emx2 to all of the ATTA sites in the core 5′ enhancer. ATTA-3 resembles (Figure 4A) one of the few characterized Emx2-binding sites, that of the Wnt1 gene (32,33); furthermore, a similar site is located in the 3′ enhancer (ATTA-4) just upstream to the already studied (11,23), functionally important, POU site. In EMSA,

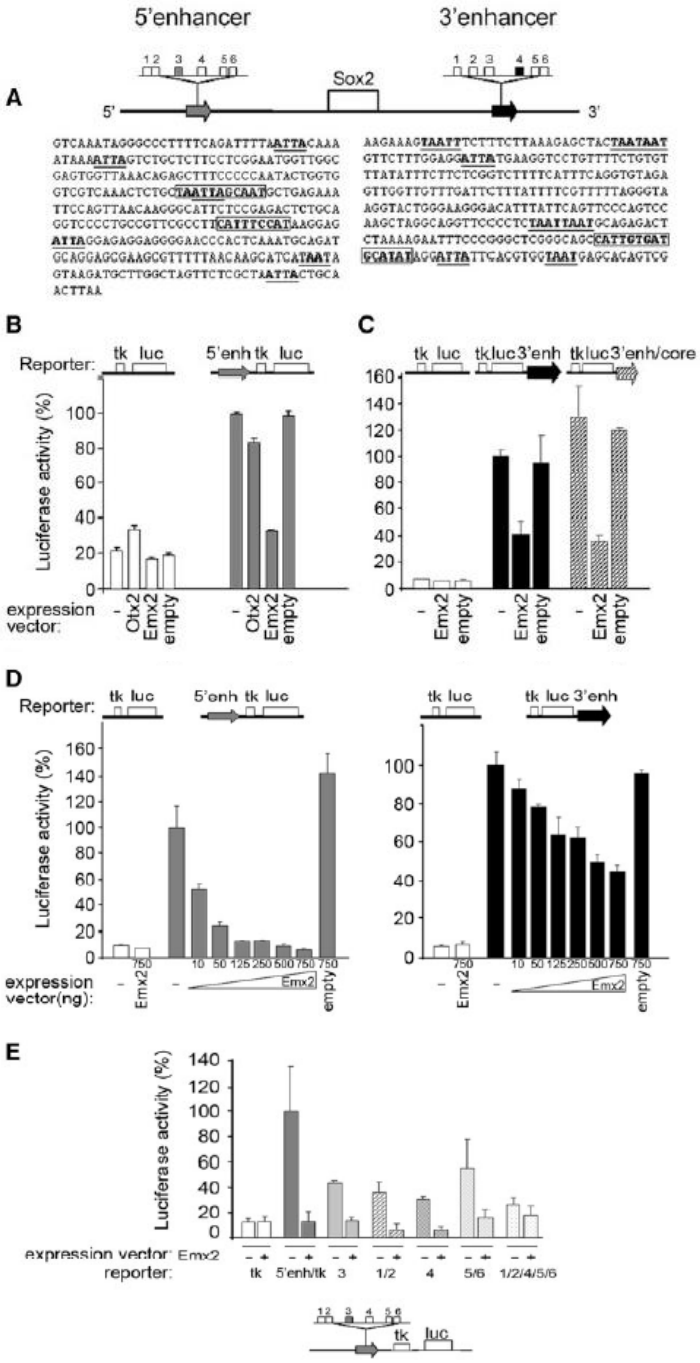


Figure 3. Emx2 represses the activity of the 5′ and 3′ Sox2 telencephalic enhancers in transfection assays. (**A**) 5′ and 3′ Sox2 telencephalic enhancers. Numbered squares: ATTA sites, underlined and bold in the sequences below. Boxed bold sequences: POU sites (11,22,23) in 5′ and 3′ enhancers (**B** and **C**). Cotransfection of 5′ or 3′ enhancer-driven (black bars, full enhancer; striped bars, 'core' enhancer) tk-luciferase vectors, or 'empty' tk-luciferase vector (white bars), with Emx2 or Otx2 expression vectors, or with 'empty' vector. The mean activity of the enhancer-driven constructs (with no cotransfected expression vector) is set = 100% luciferase activity. (**D**) Cotransfection of 5′ and 3′-enh. luciferase constructs with increasing amounts of Emx2-expression vector. (**E**) Luciferase activity of 5′ enhancer constructs carrying mutations in the indicated ATTA sites, and their response to cotransfection of the Emx2 expression vector (500 ng). Values represent the mean ± SD of $n \geq 3$ independent transfection experiments, with each transfection done in triplicate.

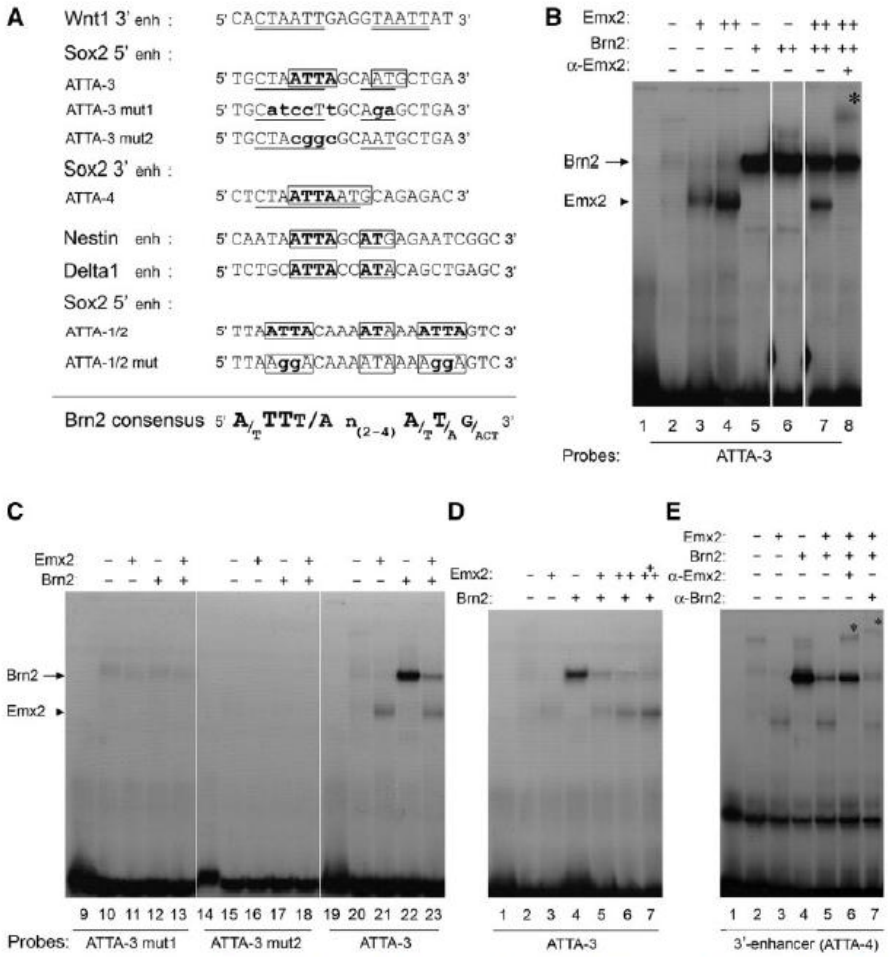


Figure 4. Emx2 binds to ATTA sites within the Sox2 5′ and 3′ enhancers, and antagonizes binding of the activator Brn2. (**A**) ATTA sequences binding Emx2 and/or Brn2. Lowermost line: Brn2/POU consensus based on TFBS cluster (http://hscl.cimr.cam.ac.uk/TFBScluster) and our data. Letter size is proportional to nucleotide frequency. The spacer (n) is 2–3 nt in previously validated sites (34,35). For the interaction of a POU factor with its binding site, and spacer length, see (27). Boxed sequences are homologies to the Brn2 consensus. Underlined sequences correspond to the previously reported Emx2 binding sequence (footprint) in the Wnt1 enhancer (32,33), and to homologous sequences within the 5′ and 3′ Sox2 enhancers. (**B**) EMSA with an ATTA-3 site probe (5′ enhancer) and recombinant Emx2 and Brn2 proteins (as indicated above the lanes). Anti-Emx2 antibody was added in lane 8. Asterisk: supershifted band. Amounts of TNT product used (see section 'Materials and Methods'): Emx2: + = 3 µl; ++ = 6 µl. Brn2: + = 3 µl; ++ = 6 µl. α-Emx2 antibody: + = 1 µl (of a 1:10 dilution of ascites fluid). (**C**) EMSA with wild-type (lanes 19–23) and two different mutated (lanes 9–13; 14–18) ATTA-3 site probes (5′ enhancer). Amounts of TNT product used: Emx2: + = 6 µl; Brn2: + = 1.5 µl. (**D**) Addition of increasing amounts of Emx2 (lanes 5–7) to ATTA-3 site probe (5′ enhancer) together with a fixed amount of Brn2 (as in lane 4). An Emx2 retarded band appears, while the Brn2 band progressively disappears. Amounts of TNT product used: Emx2: + = 4 µl; ++ = 8 µl; +++ = 12 µl. Brn2: + = 1 µl. (**E**) EMSA with a probe from the 3′ enhancer ATTA-4 site, showing ability to bind Emx2 or Brn2. Addition of Emx2 together with Brn2 (lane 5) antagonizes Brn2 binding. Asterisks indicate bands supershifted by antibodies (lanes 6 and 7). Amounts of TNT product used: Emx2: + = 5 µl. Brn2: + = 2 µl. α-Emx2 antibody: + = 1 µl (of a 1:10 dilution of ascites fluid). α-Brn2 antibody (Santa Cruz): + = 8 µl.

recombinant Emx2 (Supplementary Figure S4, panel A) bound to the Wnt-1 oligonucleotide (originally characterized only by foot-printing) generating a complex, that was super shifted by an anti-Emx2 antibody (Supplementary Figure S4, panel B). Similarly, ATTA-3 was efficiently bound by Emx2 (Figure 4B, lanes 3–4; Figure 4C, lane 21); two different mutations of ATTA-3 abolished Emx2 binding (Figure 4C, lanes 11 and 16, versus lane 21). Further, the ATTA-3/Emx2 binding was efficiently competed by excess unlabelled Wnt-1 or ATTA-3 oligonucleotides, with similar kinetics (Supplementary Figure S4B). We conclude that ATTA-3 can be bound, *in vitro*, by Emx2.

An oligonucleotide including the combined ATTA/POU site (ATTA-3) binds (21) the ES cell factor OCT4 and its brain homologues Brn1 and Brn2. As Emx2

negatively modulates the activity of Sox2 telencephalic enhancers in brain (Figure 1), we asked if Emx2 binding to the POU sites in brain cells might interfere with the binding of Brn factors. Brn2 bound, as expected, the composite POU/ATTA-site 3 (ATTA-3) of the 5′enhancer, that was shown to bind Emx2 (Figure 4B, lanes 5 and 6). When Brn2 and Emx2 were added together, no ternary Emx2-Brn2-probe complex was detected, suggesting that the binding was mutually exclusive. Addition of anti-Emx2 antibody caused the loss of the Emx2 band and its supershift, but did not affect the Brn2 band (Figure 4B, lanes 7 and 8). Importantly, Brn2 binding was abolished (Figure 4C, lanes 12 and 17 as compared to lane 22) by the same mutations that caused loss of Emx2 binding.

Adding increasing amounts of Emx2, in the presence of a fixed amount of Brn2, proportionally increased Emx2 binding, whereas Brn2 binding was strongly decreased (Figure 4D, lanes 5–7). The repression of Brn2 binding was observed already at relatively low levels of added Emx2 (and Emx2 binding), and under conditions of a large excess of labelled oligonucleotide probe; this suggests that the repression of Brn2 binding is not simply the result of a direct competition on the same DNA molecule, but rather entails other indirect mechanisms (see below).

In the 3′enhancer, a motif (ATTA-4) similar to the Emx2-binding ATTA-3 site is located just upstream to an already studied (11,22), functionally important, POU site. We performed EMSA experiments with Emx2 and Brn2 using the 3′ enhancer ATTA-4-site. Again, ATTA-4 (Figure 4A) bound both Brn2 and Emx2 (Figure 4E), and addition of Emx2 greatly decreased the binding of Brn2 (Figure 4E, lanes 4 and 5). Similarly, to the 5′ site, mutation of this site abolished the binding of both Emx2 and Brn2 (data not shown).

Emx2 inhibits Brn2 binding to ATTA sites 1,2 without directly binding to DNA

The ATTA motif is part of a large number of core sequences of distinct transcription factor-binding motifs, which are difficult to identify purely on the basis of the DNA sequence. As the POU/ATTA sequence (ATTA-3) binds both Oct3/4 and Brn1/Brn2 (21), and other sequences containing an ATTA motif bind Brn1 and Brn2 (33,34; Figure 4A), we tested all ATTA sites in the 5′ enhancer for binding to these factors. Brn2 bound (Figure 5A) an oligonucleotide containing both sites 1 and 2 (ATTA-1/2), whereas Emx2 did not bind (the weak band migrating slightly faster than Brn2 in lane 3, arrowhead, is due to a protein contained in the TNT extract used for Brn2 synthesis, see lane 2). We could not detect any binding of Emx2 to the ATTA-1/2 probe even when adding Emx2 in the absence of Brn2, in amounts equal or greater than those able to generate a strong shifted band with the ATTA-3 probe in a control binding run in parallel (Supplementary Figure S5). Mutation of the conserved TT doublet in the ATTA motif abolished Brn2 binding, leaving only the fast TNT-derived band (lanes 10 and 11). The Brn2 band was almost completely ablated by addition of anti-Brn2 antibody (lane 4). Finally, excess unlabelled ATTA-1/2 oligonucleotide competed the binding of the previously validated Brn2-binding site, ATTA-3 in the 5′ enhancer (22 and present article) as efficiently as unlabelled ATTA-3 site oligonucleotide did (Figure 5B, lanes 4 and 5, versus lane 3). In contrast, a mutated ATTA-1/2 site oligonucleotide failed to compete (lane 6). We conclude that Brn2 can bind to the ATTA-1/2 site in a sequence-specific way.

As shown in Figure 4D, Emx2 might inhibit the binding of Brn2 to the POU/ATTA site (ATTA-3) oligonucleotide both by direct DNA binding and by other indirect mechanisms. We tested the effects of Emx2 addition to the ATTA-1/2 site oligonucleotide, in the presence of Brn2. Emx2 addition (Figure 5A, lane 5) almost completely abolished Brn2 binding, already at low Emx2 concentrations. Similar or higher amounts of the haematopoietic transcription factors GATA-1 and GATA-2 did not interfere with Brn2 binding (Figure 5A, lanes 6 and 7), nor did addition of a TNT lysate prepared by transcription/translation of an 'empty' vector in control experiments (data not shown).

In additional experiments (Figure 5C) Emx2 prevented Brn2 binding, in a dose-dependent fashion, to two independently characterized Brn2-binding sites (Figure 4A), those in the Delta and Nestin genes neural enhancers (34,35).

ATTA-1/2 and ATTA 3 probes were also tested with nuclear extracts from the neural (adult hippocampal) AHP cell line (Figure 5D and E); endogenous Brn2 bound to both probes (Figure 5E, lanes 1 and 2; Figure 5E, lane 1) generating bands that were supershifted by anti-Brn2, but not anti-GATA-1 (Figure 5D, lanes 3 and 4); and were properly competed by the same unlabelled oligonucleotide, but not by its mutated version (Figure 5D, competitors: lane 5:ATTA-1/2; lanes 6 and 7: mutated ATTA-1/2; lane 8: ATTA-3). Also in this neural cell context, the addition of increasing amounts of Emx2 (but not of GATA-1) caused a sharp decrease of Brn2 binding, already at low Emx2 concentrations (Figure 5E, lanes 2–4, compare to lanes 1 and 5).

Overall, the experiments reported above (Figures 4 and 5) demonstrate that Emx2 prevents the binding of transcription factors (in this case Brn2) to their cognate motifs via mechanisms independent of its binding to DNA; one possible mechanism might be protein–protein interaction between Emx2 and Brn2. In a GST-pull down assay, a GST–Emx2 fusion protein retained *in vitro* synthesized Brn2 (Figure 5F), indicating that Emx2 and Brn2 proteins are able to physically interact.

Emx2 functionally antagonizes Brn2

POU factors, including Oct4 and neural transcription factors Brn1 and Brn2, were characterized as activators of the Sox2 3′ enhancer in co-transfection experiments, and the mutation of the POU/ATTA site (ATTA-3 site) in the 5′enhancer (22) or of the POU site in the 3′enhancer (11,22) substantially decreased the activity of Sox2 transgenic constructs, suggesting that Brn1 and Brn2 factors

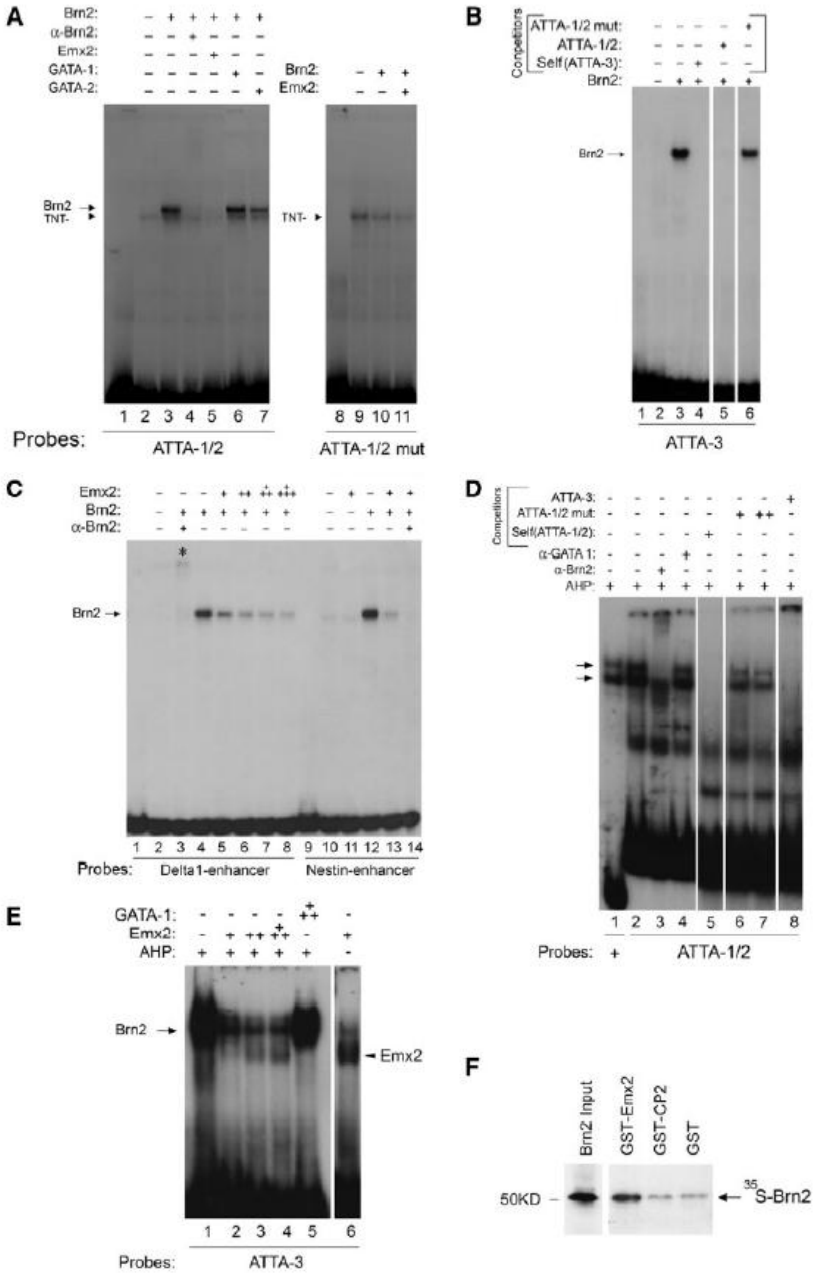


Figure 5. Emx2 antagonizes the binding of Brn2 to ATTA-1/2 sites in the 5′ enhancer, and to previously characterized Brn2 binding sites in other neural enhancers. (**A**) EMSA with a probe containing ATTA sites 1 and 2 (5′ enhancer); added recombinant proteins, and Brn2 antibody, are indicated above the lanes. The probe binds recombinant Brn2 (arrow), but not Emx2 (TNT- arrowhead indicates a non-specific band seen also with TNT extract only). Addition of Emx2 antagonizes Brn2 binding (lane 5). No antagonism is seen upon addition of GATA1 or GATA2 (lanes 6 and 7). Amounts of TNT product used (see section 'Materials and Methods'): Brn2: + = 1.5 μl; Emx2: + = 4 μl; GATA1/2: + = 4 μl. α-Brn2 antibody (Santa Cruz): + = 8 μl. (**B**) EMSA with an ATTA-3 site probe (a previously validated Brn2 binding site in the 5′ enhancer (11,22,23); binding of Brn2 is efficiently competed by wild-type non-labelled ATTA-1/2 sites oligonucleotide (lane 5), but not by its mutated version (lane 6). Competition is as efficient as with the 'self' oligonucleotide (lane 4). Amounts of TNT product used: Brn2: + = 1.5 μl. Competitor amounts: Self (ATTA-3): + = 25X (molar excess of cold competitor oligonucleotide); ATTA-1/2: + = 25X; ATTA-1/2 mut: + = 25X. (**C**) EMSA with probes containing previously validated Brn2 binding sites in the nestin and Delta-1 enhancers. Brn2 binding (arrow) is antagonized by simultaneous Emx2 addition in a dose-dependent way. Asterisk: Brn2 antibody-supershifted band. Amounts of TNT product used for Delta 1-enhancer : Brn2: + = 2 μl; Emx2: + = 1 μl; ++ = 2 μl; +++ = 3 μl; ++++ = 4 μl. α-Brn2 antibody: + = 8 μl (Santa Cruz). Amounts of TNT product used for

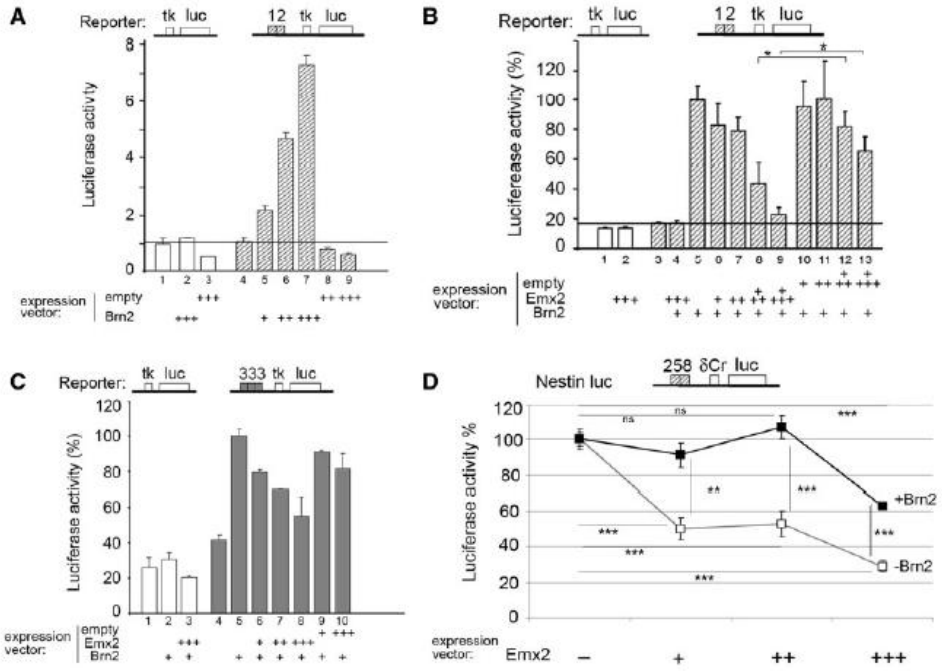


Figure 6. Emx2 represses Brn2-transactivated ATTA-1/2 and ATTA-3 sites—tk luciferase reporter constructs in a dose-dependent way. (**A**) Brn2 dose-dependent transactivation of ATTA-1/2 sites (5′ enhancer) (Brn2: +, ++, +++, 125, 500, 1000 ng) (B and C) Emx2 dose-dependent repression of Brn2-dependent transactivation of ATTA-1/2 sites construct (**B**) and of ATTA site 3 construct (**C**) (Brn2: +, 500 ng; Emx2: +, ++, +++, ++++; 100, 200, 500, 1000 ng). In **A**, luciferase activity is expressed in arbitrary units, where 1 is the activity of the tk luc reporter; in B and C, 100% luciferase activity is set to the maximum observed activity. The horizontal line in A and B represents the background activity of the ATTA-1/2 site construct in the absence of cotransfected Brn2. (*$P < 0.007$ by Student's *t*-test, non-parametric, Welch correction) (**D**) Nestin-enhancer (258 wt)-driven luciferase reporter, transactivated by cotransfected Sox2 expression vector 500 ng (as in 25), is cotransfected with increasing amounts of Emx2 expression vector, in the presence (+Brn2) or absence (−Brn2) of Brn2 expression vector (Brn2: 500 ng; Emx2: +, ++, +++; 125, 250, 500 ng). Brn2 addition antagonizes the repressive effect of Emx2 (**$P < 0.001$; ***$P < 0.0002$ by Student's *t*-test, non-parametric, Welch correction). 100% luciferase activity is set to that observed without Emx2 cotransfection for both +Brn2 and −Brn2 samples. In the absence of cotransfected Brn2, the activity of the construct was about 20% lower (marginally significant, data not shown) than in the presence of Brn2. Note that a Sox2-expression vector was cotransfected in all the experiments, as Sox2 had been reported to increase the activity of the present reporter in other cell types (25). Values represent the mean ± SD of $n \geq 3$ independent transfection experiments, with each transfection in triplicate.

may be positive regulators of Sox2 transcription in the brain (11,22,23).

To evaluate the respective roles of Brn2 and Emx2 in transfection experiments we linked to the minimal tk-promoter the ATTA-1/2 or the POU/ATTA (ATTA-3) site (the latter as a trimer) from the 5′enhancer. We transfected the construct into P19 teratocarcinoma cells (which express Sox2) in the presence of different amounts of Brn2-and/or Emx2 expression vectors (Figure 6). In the absence of Emx2, Brn2 strongly stimulated the activity of the ATTA-1/2 construct in a dose-dependent way and, to a lesser extent, that of the ATTA-3 construct (Figure 6A and C). The Brn2-dependent stimulation of the ATTA-1/2 construct was repressed to basal levels (just above the level of the tk-luc reporter, lane 9 versus lanes 1 and 2), by cotransfection of progressively increasing amounts of the Emx2-expression vector (Figure 6B). Cotransfection of

Figure 5. Continued
Nestin-enhancer: Brn2: + = 2 µl; Emx2: + = 4 µl. α-Brn2 antibody: + = 8 µl (Santa Cruz). (**D**) EMSA with ATTA-1/2 site probe and nuclear extracts from AHP neural cells. Two complexes are generated (arrows) with both ATTA-3 (lane 1, '+' as in ref. 21) and ATTA-1/2 (lane 2), which are supershifted by anti-Brn2 (lane 3), but not anti-GATA1 antibodies (lane 4). Binding of Brn2 to ATTA-1/2 is efficiently competed by unlabelled ATTA-3 (lane 8), by 'self' ATTA-1/2 (lane 5), but not by mutated ATTA-1/2 (lanes 6 and 7) oligonucletides. Unlabeled competitor oligonucleotides were added in 25-fold molar excess (+) or 50-fold molar excess (++). (**E**) EMSA with ATTA-3 probe and nuclear extracts from AHP cells. Added recombinant proteins (Emx2, GATA-1) are indicated above the lanes. The Brn2 retarded complex (lane 1, arrow) (see also ref. 21 and panel D) is sharply decreased following addition of Emx2 (lanes 2–4), but not of control GATA-1 (lane 5). The lower, Emx2-containing complex is progressively increased in parallel with the addition of Emx2. This complex has the same mobility of that generated by direct binding of recombinant Emx2 to the ATTA-3 probe (lane 6). (**F**) Emx2 and Brn2 directly interact in a GST pulldown assay. Brn2 is retained by GST-Emx2, but not by GST-CP2 control resin (which gives a weak signal equivalent to that seen with the 'empty' resin GST).

control 'empty' vector, instead of Emx2^{-} expression vector, yielded a slight inhibition only at the highest tested levels, ensuring specificity of the Emx2 repression observed (Figure 6B, lanes 10–13). Similarly, on the ATTA-3 construct, Brn2-dependent stimulation was inhibited by Emx2 (Figure 6C), though the observed repression is weaker. Thus, Brn2 is an activator at both the ATTA-3 [as previously shown *in vivo* and *in vitro* (22)] and the ATTA-1/2 sites, and Emx2 inhibits the transcriptional activity at the same sites, antagonizing Brn2-dependent stimulation. As Emx2 does not bind to ATTA-1/2 site sequences (Figure 5A), this inhibition may be caused by mechanisms that do not strictly require Emx2 binding to the DNA. The somewhat lower effect of Emx2 in the Brn2-dependent system, as compared to the drastic effect observed with the full 'core' element (in the absence of cotransfected Brn2) (Figure 3), probably reflects the modest enhancer activity of the individual ATTA sites in isolation, as compared with the cooperative activity of the multiple sites active in the full enhancer (Figure 3).

Is the Emx2 inhibitory effect limited to the Sox2 5′ and 3′ enhancers? To evaluate this point, we performed experiments using the nestin enhancer, that is positively regulated by transfection of Brn factors, in conjunction with Sox2 (25). As shown above (Figure 5C), Emx2 addition antagonized Brn2 binding (in EMSA) to this enhancer. In transfection experiments in P19 cells, in the absence of transfected Brn2, Emx2 strongly inhibited the activity of the enhancer already at low concentrations (Figure 6D, '−Brn2', open squares). In this condition, enhancer activity likely depends on the related POU factor Oct3/4, expressed in P19 cells. In contrast, the addition of Brn2 ('+Brn2', filled squares) completely prevented the Emx2 repression at low/intermediate Emx2 levels, and substantially attenuated it at the highest Emx2 level (Figure 6D). Note that, in P19 cells, the addition of Brn2 *per se* stimulates the activity of the enhancer only minimally (~20%, data not shown), indicating that the observed Brn2-dependent attenuation of Emx2-mediated repression is not the result of independent stimulation of gene activity by Brn2; rather, the excess Brn2 may directly 'titrate' Emx2 activity. These results, together with those reported above, suggest that Emx2 and Brn2 antagonize each other's activities.

Emx2 binds to the 5′enhancer *in vivo*

To ascertain if Emx2 interacts in brain cells with the Sox2 regulatory elements, we performed *in vitro* ChIP with anti-Emx2 antibodies, using chromatin from embryonic telencephalon (E14.5), from wild-type and Emx2-null (negative control) embryos. A fragment comprising the ATTA-3 and the adjacent ATTA-1/2 sites was bound by Emx2 in wild-type chromatin, but not in Emx2-null chromatin (Figure 7). No binding was detected in an adjacent region B, comprising ATTA-5 and 6 sites, and lying 3′ to the bound DNA region. We conclude that Emx2 likely functionally interacts with the Sox2 regulatory region *in vivo*.

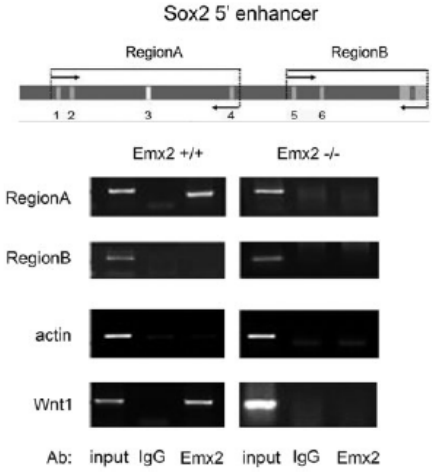


Figure 7. Emx2 is bound to the Sox2 enhancer *in vivo*. ChIP with anti-Emx2 antibodies of E14.5 embryonic brain chromatin from wild-type and Emx2$^{-/-}$ control embryos. Region A, containing ATTA-3 site is immunoprecipitated from wild-type, but not Emx2-null chromatin. The previously described Wnt1 enhancer containing an Emx2 binding site (33) is used as a control (Wnt1), and is similarly precipitated from wild-type, but not mutant, chromatin. Antibodies used are indicated below the lanes. Input: input chromatin. IgG: anti-IgG control antibodies. Emx2: anti-Emx2 antibodies.

DISCUSSION

Emx2 is a transcription factor involved in hippocampal growth and in cortex patterning (19,36). With the exception of the Wnt1 and FGF8 genes (32,37,38), there are few identified neural Emx2 targets. Here we show, by *in vivo* and *in vitro* experiments, that Emx2 negatively regulates two characterized Sox2 enhancers. Loss of a single Emx2 allele increases Sox2 expression in the E15.5 medial telencephalic wall and partially rescues an hippocampal phenotype dependent on Sox2 deficiency. Our results, together with data of the literature, suggest that Emx2 may control aspects of Sox2 expression and brain development by antagonizing the activities of transcriptional activators, such as Brn2.

Emx2 negatively modulates Sox2 telencephalic enhancers *in vivo*

Sox2 neural expression in chick (30) and mouse (6,11,22,23,39) is regulated by multiple enhancers. Among the best characterized mouse enhancers are the 5′ and 3′ Sox2 enhancers studied here, which direct transgenic reporter gene expression to the telencephalon, the 5′ enhancer being more active in dorso-medial regions, and the 3′ enhancer in ventro-lateral regions. Emx2 is expressed in the dorsal telencephalon in a posterior medial to anterior lateral concentration gradient, that intercepts the wide Sox2 expression domain; at the cellular level, the two expression domains substantially overlap within the ventricular zone, allowing for potential cross-regulation (3,19,36). In our experiments (Figure 1), the loss of one or both Emx2 copies substantially

increases the expression of transgenes driven by the 5′ or the 3′ Sox2 enhancers, indicating that the normal levels of Emx2 may inhibit to some extent the activities of enhancers of Sox2. The inhibitory activity of Emx2 is further reflected in the decreased activity of the same enhancers brought about by Emx2 overexpression in NSC cultures (Figure 1C).

Does the altered regulation of Sox2 enhancers by Emx2 modify the levels of endogenous Sox2 *in vivo*? Overall, in developing brain, Sox2 levels are not highly changed in Emx2 +/− embryos, but significant modulation can be appreciated at specific locations. In the late embryo, Sox2 and Emx2 are coexpressed in the prospective hippocampal domain; at this stage, in the lateral ventricle, regions of high Sox2 expression show relatively lower Emx2, and regions of high Emx2 expression have lower Sox2 levels (Figure 2C). This inverse correlation is particularly evident in the medial telencephalic wall, where the hippocampus primordia develop (Figure 2C and Supplementary Figure S3). Loss of a single Emx2 allele results, in this region, in increased Sox2 expression (Figure 2C and Supplementary Figure S3). Thus, effects of Emx2 deficiency on Sox2 may be more evident in specific regions/developmental stages, possibly depending on expression levels of Emx2 itself, or interactions with other factors. This conclusion is in agreement with the rather subtle phenotipic effects of changes in Emx2 levels observed in the cortex as compared to the hippocampus (36).

Heterozygous Emx2 deficiency antagonizes the hippocampal NSC loss of Sox2 hypomorphic mutants

An important question arising from our present results, is whether heterozygous Emx2-deficiency (that, by itself, has little effect on brain development, 15,20,36) has any phenotypic consequences on Sox2-dependent functions.

Both Emx2 homozygous mice and Sox2 mutants (Sox2 hypomorphic and Sox2 conditional-null mice) show severe hippocampal defects, indicating that both genes are essential for hippocampal development (3,14,15). In the hippocampus, Sox2 is required for postnatal NSC survival; complete Sox2 ablation by E12.5 results in the loss of hippocampal neurogenesis and DG severe hypoplasia, between P2 and P7 (14). Moreover, in adult Sox2 hypomorphic ($Sox2^{\beta\text{-geo}/\Delta Enh}$) mutants, expressing 30% of the normal Sox2 RNA, nestin/GFAP radial glia cells (a Sox2-expressing neural stem/progenitor cell) (3,5,40) in the hippocampus are importantly decreased (Figure 2). As Sox2 and Emx2 are coexpressed in the hippocampal primordium (Figure 2) and in the adult hippocampus (as well as in the hippocampal AHP cell line (Supplementary Figures S6 and S7), a reduction in Emx2 dosage may be expected to affect Sox2-dependent functions in this region.

In adult hypomorphic Sox2 mutants, heterozygous Emx2-deficiency strongly increases the number of nestin/GFAP radial glia cells and, to a lesser extent, BrdU incorporation (Figure 2A and B); note that, in wild-type mice, heterozygous Emx2 deficiency, *per se*, only slightly raises the number of nestin/GFAP radial glia cells and decreases, rather than increases, BrdU incorporation (Figure 2A and B; ref. 31). These results demonstrate that Emx2 deficiency rescues, in part, at least one well characterized Sox2-dependent NSC phenotype. These data, taken together with the increased Sox2 expression in the medial lateral ventricle wall (that includes the prospective hippocampus) of heterozygous Emx2-deficient mice (Figure 2C, Supplementary Figure S3), are thus consistent with the hypothesis that Emx2 deficiency may contribute to phenotypic rescue by raising Sox2 expression. Of course, additional mechanisms might also contribute to the observed phenotype.

Emx2 represses telencephalic enhancers in transfection assays, directly binds to enhancer sequences *in vitro* and antagonizes binding of POU transcriptional activators

To begin to address at the molecular level of our *in vivo* data, we performed EMSA and transfection experiments with P19 cells, that express Sox2, and can be manipulated by transfection to express Brn2 and/or Emx2 (absent in the basal state). We propose two mechanisms whereby Emx2 might downregulate Sox2 enhancer activity (Figures 4, 5 and 6; Supplementary Figure S8).

Firstly, Emx2, by directly binding to its cognate sites in DNA, might prevent the activity of other transcription factors which bind to sites overlapping the Emx2 motif; secondly, Emx2 might repress transcription by antagonizing the binding to DNA of transcription factors, without direct DNA binding, through protein–protein interactions. As to the first mechanism, Emx2 directly binds to 5′ (ATTA-3) and 3′enhancer (ATTA-4) sites in the Sox2 gene (Figure 4), which are also bound by the POU factors Brn1 and Brn2; importantly, these factors were previously implicated in Sox2 activity on the basis of *in vivo* experiments (transfection, transgenic mice and ChIP (11,22,23). As mutations at the ATTA-3 site abolish the binding of both Emx2 and Brn2, their binding sites are likely functionally overlapping, and their binding might be mutually exclusive; indeed, we did not detect by EMSA (even at high protein concentrations, data not shown) any band of mobility slower than that of Brn2, that might suggest the formation of a ternary complex of DNA with both factors. Moreover, in EMSA, the addition of increasing amounts of Emx2 resulted in increased binding of Emx2, together with progressive disappearance of the Brn2 band (Figure 4).

The second mechanism is suggested by the following observations: the binding of Brn2 to ATTA-sites in Sox2 enhancers and to other previously validated Brn2 sites (nestin, delta; 22,34,35) is prevented by Emx2 addition, in the absence of any binding of Emx2 itself to the same sequences (Figure 5). Emx2 and Brn2 might be reciprocally antagonistic (Figure 6) through direct protein–protein interaction, which would prevent Brn2 binding to regulatory sequences, and transcriptional activation (Figure 6). Indeed, GST pull-down experiments show that Brn2 and Emx2 may physically interact (Figure 5F). Our present interpretation is in agreement with data reported by other authors in a different experimental system; Sahara *et al.* (37) reported that Emx2

represses SP8 trancription factor-dependent activity of the FGF8 promoter without binding to the promoter itself; moreover, other authors reported that Emx2 and SP8 physically interact (38).

The DNA motif recognized by Brn2 in our experiments is a rather degenerate one, centred on an ATTA motif potentially recognized by many transcription factors (41). It can be hypothesized that, in addition to Brn2, other transcription factors, particularly, among neural factors, the Brn1 homolog, or Oct6, might bind to this sequence, and could thus be antagonized by Emx2. The 5′ enhancer is also bound by Oct3/4 at early developmental stages, and Oct3/4 is required for its activity in ES cells (22) and in the early neural plate (13), leading to the suggestion that an exchange between early Oct3/4 and later Brn/Oct factors binding is important in the early function of this element (13,22). Of note, this enhancer is active in P19 cells, which do not express neural POU factors, but do express Oct3/4; Emx2 addition by cotransfection can antagonize both basal 5′ enhancer activity (presumably dependent on Oct3/4 binding, 13,22), as well as Brn-stimulated activity (Figures 3 and 6). This may, again, reflect a wider ability by Emx2 to antagonize different POU factors, which may differently act on Sox2 (and other genes) regulation at different developmental stages and in different telencephalic regions.

Additional data suggest that these mechanisms do operate *in vivo*. In fact, Emx2 binds to a fragment comprising the 5′ ATTA-3 site in nuclei from normal telencephalon, in ChIP experiments (Figure 7). This fragment lies within a 120 bp DNA region that mediates POU site-dependent reporter gene expression in the telencephalon of transgenic embryos (22).

In conclusion, we propose that Emx2 contributes to the modulation of Sox2 expression by antagonizing Brn2 and possibly other activators able to bind the ATTA core sequence. The mechanism may not be restricted to the Sox2 enhancers, as at least the nestin enhancer may be similarly regulated (Figures 5C and 6D). It provides a wide scope for regulation, depending on the affinities of Emx2 for its DNA target and/or protein interactors, and on the relative ratios between Emx2 and brain transcription factors.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figures 1–8.

ACKNOWLEDGEMENTS

We thank Pierfrancesco Pagella and Florian Ebner for the generation of the Emx2 lentiviral genome, Vania Broccoli for the pCAGGS-Emx2 construct, Hisato Kondoh for the Nes258-luciferase construct, Dieter Chichung Lie, Esra Karaca and Rusty Gage for the AHP cell line and advice on its culture, and Barbara Costa and Ivan Scotti for help with statistical analysis.

FUNDING

Fondazione Cariplo (20100673); Associazione Italiana Ricerca sul Cancro (AIRC – IG5801); Fondazione Banca del Monte di Lombardia, Ministero dell'Istruzione, Università e Ricerca (MIUR, Cofin 2008); and ASTIL Regione Lombardia (SAL-19, N Prot FL-16874, to S.K.N.). Funding for open access charge: Fondo Ateneo di Ricerca (FAR 2009–2011).

Conflict of interest statement. None declared.

REFERENCES

1. Avilion,A.A., Nicolis,S.K., Pevny,L.H., Perez,L., Vivian,N. and Lovell-Badge,R. (2003) Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev.*, **17**, 126–140.
2. Ellis,P., Fagan,B.M., Magness,S.T., Hutton,S., Taranova,O., Hayashi,S., McMahon,A., Rao,M. and Pevny,L. (2004) SOX2, a persistent marker for multipotential neural stem cells derived from embryonic stem cells, the embryo or the adult. *Dev. Neurosci.*, **26**, 148–165.
3. Ferri,A.L., Cavallaro,M., Braida,D., Di Cristofano,A., Canta,A., Vezzani,A., Ottolenghi,S., Pandolfi,P.P., Sala,M., DeBiasi,S. *et al.* (2004) Sox2 deficiency causes neurodegeneration and impaired neurogenesis in the adult mouse brain. *Development*, **131**, 3805–3819.
4. Graham,V., Khudyakov,J., Ellis,P. and Pevny,L. (2003) SOX2 functions to maintain neural progenitor identity. *Neuron*, **39**, 749–765.
5. Suh,H., Consiglio,A., Ray,J., Sawai,T., D'Amour,K.A. and Gage,F.H. (2007) In vivo fate analysis reveals the multipotent and self-renewal capacities of Sox2+ neural stem cells in the adult hippocampus. *Cell Stem Cell*, **1**, 515–528.
6. Zappone,M.V., Galli,R., Catena,R., Meani,N., De Biasi,S., Mattei,E., Tiveron,C., Vescovi,A.L., Lovell-Badge,R., Ottolenghi,S. *et al.* (2000) Sox2 regulatory sequences direct expression of a (beta)-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.
7. 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.
8. Kelberman,D., de Castro,S.C., Huang,S., Crolla,J.A., Palmer,R., Gregory,J.W., Taylor,D., Cavallo,L., Faienza,M.F., 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.
9. Sisodiya,S.M., Ragge,N.K., Cavalleri,G.L., Hever,A., Lorenz,B., Schneider,A., Williamson,K.A., Stevens,J.M., Free,S.L., Thompson,P.J. *et al.* (2006) Role of SOX2 mutations in human hippocampal malformations and epilepsy. *Epilepsia*, **47**, 534–542.
10. Foshay,K.M. and Gallicano,G.I. (2008) Regulation of Sox2 by STAT3 initiates commitment to the neural precursor cell fate. *Stem Cells Dev.*, **17**, 269–278.
11. Miyagi,S., Nishimoto,M., Saito,T., Ninomiya,M., Sawamoto,K., Okano,H., Muramatsu,M., Oguro,H., Iwama,A. and 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.
12. Papanayotou,C., Mey,A., Birot,A.M., Saka,Y., Boast,S., Smith,J.C., Samarut,J. and Stern,C.D. (2008) A mechanism regulating the onset of Sox2 expression in the embryonic neural plate. *PLoS Biol.*, **6**, e2.
13. Iwafuchi-doi,M., Yoshida,Y., Onichtchouk,D., Leichsenring,M., Driever,W., Takemoto,T., Uchikawa,M., Kamachi,Y. and Kondoh,H. (2011) The Pou5f1/Pou3f-dependent but Sox3B-independent regulation of conserved enhancer N2 initiates

Sox2 expression during epiblast to neural plate stages in vertebrates. *Dev. Biol.*, **352**, 354–366.
14. Favaro,R., Valotta,M., Ferri,A.L., Latorre,E., Mariani,J., Giachino,C., Lancini,C., Tosetti,V., Ottolenghi,S., Taylor,V. *et al.* (2009) Hippocampal development and neural stem cell maintenance require Sox2-dependent regulation of Shh. *Nat. Neurosci.*, **12**, 1248–1256.
15. Pellegrini,M., Mansouri,A., Simeone,A., Boncinelli,E. and Gruss,P. (1996) Dentate gyrus formation requires Emx2. *Development*, **122**, 3893–3898.
16. Simeone,A., Gulisano,M., Acampora,D., Stornaiuolo,A., Rambaldi,M. and Boncinelli,E. (1992) Two vertebrate homeobox genes related to the Drosophila empty spiracles gene are expressed in the embryonic cerebral cortex. *EMBO J*, **11**, 2541–2550.
17. Galli,R., Fiocco,R., De Filippis,L., Muzio,L., Gritti,A., Mercurio,S., Broccoli,V., Pellegrini,M., Mallamaci,A. and Vescovi,A.L. (2002) Emx2 regulates the proliferation of stem cells of the adult mammalian central nervous system. *Development*, **129**, 1633–1644.
18. Gangemi,R.M., Daga,A., Marubbi,D., Rosatto,N., Capra,M.C. and Corte,G. (2001) Emx2 in adult neural precursor cells. *Mech. Dev.*, **109**, 323–329.
19. O'Leary,D.D., Chou,S.J. and Sahara,S. (2007) Area patterning of the mammalian cortex. *Neuron*, **56**, 252–269.
20. Rash,B.G. and Grove,E.A. (2006) Area and layer patterning in the developing cerebral cortex. *Curr. Opin. Neurobiol.*, **16**, 25–34.
21. Roelink,H. (2000) Hippocampus formation: an intriguing collaboration. *Curr. Biol.*, **10**, R279–R281.
22. Catena,R., Tiveron,C., Ronchi,A., Porta,S., Ferri,A.L.M., 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.
23. 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.
24. 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.
25. Tanaka,S., Kamachi,Y., Tanouchi,A., Hamada,H., Jing,N. and Kondoh,H. (2004) Interplay of SOX and POU factors in regulation of the Nestin gene in neural primordial cells. *Mol. Cell. Biol.*, **24**, 8834–8846.
26. Bosé,F., Fugazza,C., Casalgrandi,M., Capelli,A., Cunningham,J.M., Zhao,Q., Jane,S.M., Ottolenghi,S. and Ronchi,A. (2006) Functional interaction of CP2 with GATA-1 in the regulation of erythroid promoters. *Mol. Cell. Biol.*, **26**, 3942–3954.
27. Scully,K.M., Jacobson,E.M., Jepsen,K., Lunyak,V., Viadiu,H., Carriere,C., Rose,D.W., Hooshmand,F., Aggarwal,A.K. and Rosenfeld,M.G. (2000) Allosteric effects of Pit-1 DNA sites on long-term repression in cell type specification. *Science*, **290**, 1127–1131.
28. Di Rocco,G., Gavalas,A., Popperl,H., Krumlauf,R., Mavilio,F. and Zappavigna,V. (2001) The recruitment of SOX/OCT complexes and the differential activity of HOXA1 and HOXB1 modulate the Hoxb1 auto-regulatory enhancer function. *J. Biol. Chem.*, **276**, 20506–20515.
29. Capellini,T.D., Vaccari,G., Ferretti,E., Fantini,S., He,M., Pellegrini,M., Quintana,L., Di Giacomo,G., Sharpe,J., Selleri,L. *et al.* (2010) Scapula development is governed by genetic interactions of Pbx1 with its family members and with Emx2 via their cooperative control of Alx1. *Development*, **137**, 2559–2569.
30. Uchikawa,M., Ishida,Y., Takemoto,T., Kamachi,Y. and 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.
31. Muzio,L., Soria,J.M., Pannese,M., Piccolo,S. and Mallamaci,A. (2005) A mutually stimulating loop involving emx2 and canonical wnt signalling specifically promotes expansion of occipital cortex and hippocampus. *Cereb. Cortex*, **15**, 2021–2028.
32. Iler,N., Rowitch,D.H., Echelard,Y., McMahon,A.P. and Abate-Shen,C. (1995) A single homeodomain binding site restricts spatial expression of Wnt-1 in the developing brain. *Mech. Dev.*, **53**, 87–96.
33. Ligon,K.L., Echelard,Y., Assimacopoulos.,Danielian,P.S., Kaing,S., Grove,E.A., McMahon,A.P. and Rowitch,D.H. (2003) Loss of Emx2 function leads to ectopic expression of Wnt1 in the developing telencephalon and cortical dysplasia. *Development*, **130**, 2275–2287.
34. Castro,D.S., Skowronska-Krawczyk,D., Armant,O., Donaldson,I.J., Parras,C., Hunt,C., Critchley,J.A., Nguyen,L., Gossler,A., Gottgens,B. *et al.* (2006) Proneural bHLH and Brn proteins coregulate a neurogenic program through cooperative binding to a conserved DNA motif. *Dev. Cell*, **11**, 831–844.
35. 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.
36. O'Leary,D.D. and Sahara,S. (2008) Genetic regulation of arealization of the neocortex. *Curr. Opin. Neurobiol.*, **18**, 90–100.
37. Sahara,S., Kawakami,Y., Izpisua Belmonte,J.C. and O'Leary,D.D. (2007) Sp8 exhibits reciprocal induction with Fgf8 but has an opposing effect on anterior-posterior cortical area patterning. *Neural Dev.*, **2**, 10.
38. Zembrzycki,A., Griesel,G., Stoykova,A. and Mansouri,A. (2007) Genetic interplay between the transcription factors Sp8 and Emx2 in the patterning of the forebrain. *Neural Dev.*, **2**, 8.
39. Takanaga,H., Tsuchida-Straeten,N., Nishide,K., Watanabe,A., Aburatani,H. and Kondo,T. (2009) Gli2 is a novel regulator of sox2 expression in telencephalic neuroepithelial cells. *Stem Cells*, **27**, 165–174.
40. Doetsch,F. (2003) The glial identity of neural stem cells. *Nat. Neurosci.*, **6**, 1127–1134.
41. Noyes,M.B., Christensen,R.G., Wakabayashi,A., Stormo,G.D., Brodsky,M.H. and Wolfe,S.A. (2008) Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. *Cell*, **133**, 1277–1289.

CHAPTER 5:
CONCLUSIONS AND FUTURE PERSPECTIVES

1. THE IMPORTANCE OF STUDYING SOX2

1.1 SOX2 IS REQUIRED DURING EMBRYONIC DEVELOPMENT

Sox2 belongs to the *Sox* (*S*ry-related HMG b*ox*) family of transcription factors, playing important roles in development and differentiation. SOX2 binds a specific DNA sequence (contacting the minor grove of DNA) through the HMG box, that also acts as a protein-to-protein interaction domain.

Sox2 is expressed from early developmental stages in the morula and blastocyst inner cell mass (ICM); later its expression is confined to the developing neural plate and subsequently to the neural tube. During the later stages of embryonic development, its neural expression remains high in the ventricular zone in active proliferation, while it decreases in the marginal zone where differentiation begins [Ferri *et al.*, 2004].

Sox2 expression is crucial during the early stages of embryonic development. Homozygous *Sox2*-KO (knock-out) mice die following loss of the stem cells of the blastocyst inner cell mass (ICM) [Avilion *et al.*, 2003; Penvy *et al.*, 1998]. Hence, to study SOX2 later functions in neural development, our laboratory generated, through gene targeting, a "*Sox2*^{flox}" mutation, in which the *Sox2* gene is flanked by lox sites; these are the substrates for Cre-recombinase, which, expressed by suitable transgenes, allows the spatially and temporally controlled ablation of *Sox2.*

Using two *Sox2* conditional knock-outs in mouse (*Nestin-Cre* and *Bf1-Cre* transgene, activated at two different time-points during embryonic development), our laboratory discovered that *Sox2* is important for the development of the brain (hippocampus and basal ganglia) and for the maintenance of neural stem cells both *in vivo* (in the hippocampus) and *in vitro* (for long-term self renewal) [Favaro *et al*., 2009; Ferri *et al.*, 2013]. Moreover, the brain defects in mice were associated to the down-regulation of ventral determinant markers (such as, *Nkx2.1* and *Shh,* both SOX2 target genes). We observed that, administrating a SHH-agonist to the pregnant mice, the defects in mutant developing mice were partially rescued. Thus, the defects in brain development, caused by failure to activate a critical SOX2 target (*Shh*), could be partially compensated by supplying a

drug mimicking the action of the target gene product (SHH agonist)[Favaro *et al.*, 2009; Ferri *et al.*, 2013].

1.2. SOX2 AND HUMAN DISEASES

In mouse, the viability doesn't seem to be affected if only one *Sox2* allele is lost. Moreover, *Sox2* heterozygous mice do not show any overt pathology, with the exception of some mild ventricle enlargement [Ferri *et al.*, 2004]. Instead, heterozygous *SOX2* mutations in humans (including microdeletions, missense, frameshift and nonsense mutations) cause neurological defects, such as defects in development of eyes (anophthalmia, microphthalmia) [Fantes *et al.*, 2003; Schneider *et al.*, 2009] and defects in hippocampus, with neurological pathology including epilepsy, motor control problems and learning disabilities [Ragge *et al.*, 2005; Sisodiya *et al.*, 2006; Kelberman *et al.*, 2006]. Other pathological characteristics of patients with heterozygous *SOX2* mutations are mild facial dysmorphism, developmental delay, esophageal atresia [Kelberman *et al.*, 2006], psychomotor retardation and hypothalamo-pituitary disorders [Tziaferi *et al.*, 2008].

Central nervous system abnormalities, similar to the severe ones observed in human patients, had been identified in another mouse model generated in our lab: a *Sox2*$^{\beta geo/\Delta enh}$ mouse [Ferri *et al.*, 2004]. This model is heterozygous for *Sox2* gene and, in the other allele, it carries the deletion of a neural cell-specific enhancer of *Sox2* [Zappone *et al.*, 2000]. These mice are born in lower numbers, if compared to the expected frequency, with severe brain malformations and defects in neural stem cells proliferation. It was probably due to the reduced SOX2 level (25-30%, compared to the wild-type) produced by these heterozygous mice. Moreover, 40% of these mice presented epileptic-like spikes in cerebral cortex and hippocampus. Furthermore, brain abnormalities of these mice could be associated to those observed in other mouse models for neurological disease: for example, loss of thalamo-striatal parenchyma with ventricle enlargement are associated to primary neurodegeneration as in Huntington and Alzheimer's diseases [Ferri *et al.*, 2004; Capsoni *et al.*, 2000; Yamamoto *et al.*, 2000]; of note, intracellular inclusions in neurons are comparable to protein inclusions in neurodegenerative diseases [Ferri *et al.*, 2004].

Thus, *Sox2* mutation results in pathological disorders, in humans as well as mouse models. Trying to better understand the mechanism behind its involvement in disease would be useful to learn more about the pathogenesis.

2. SOX2-DEPENDENT LONG-RANGE INTERACTIONS AND DISEASES

Often regulatory elements are localized very far from the genes they control on the linear chromosome map and they are able to reach the proximity of these genes, and regulate their expression, through the formation of chromatin loops, called long-range interactions. Many of these distal regulatory elements are localized in non-coding regions of the genome, in gene desert regions, or within introns of not-related genes. Mutations in their sequences could cause dramatic effects on the expression of the regulated gene. For example, a single nucleotide mutation, found in a regulatory sequence located 460 kb upstream of the *Shh* gene, was discovered in an individual with holoprosencephaly; the mutation reduced the activity of the distant enhancer [Jeong *et al.*, 2008].

For this reason, the identification and functional characterization of regulatory sequences is crucial for understanding the spatial and temporal control on gene expression.

We have observed that SOX2 is involved in chromatin loops formation and its presence is crucial for the maintenance of a high number of long-range interactions, which are lost in *Sox2*-deleted cells. On the other hand, the absence of *Sox2* induces the formation of new interactions. Thus, SOX2 is strongly implicated in the regulation and in the expression of a lot of genes but, in some circumstances, it may also be implicated in preventing loops formation. With our *in vivo* experiments in transgenesis in zebrafish, we demonstrated that this genome-wide approach (Chromatin Interaction Analysis by Paired-End Tag sequencing – ChIA-PET; Zhang *et al.*, 2013) is a good method for identifying distal regulatory elements scattered within the genome. A small subgroup of them resulted also directly responsive to experimental modulation of SOX2 levels.

Some SOX2-dependent long-range interactions, found analyzing the ChIA-PET data, involve regions (not yet tested in *in vivo* or *in vitro* experiments) highly connected to the Wolf-Hirschhorn syndrome, a disease characterized by mental retardation, microcephaly (also observed in *Sox2*-mutant mice, Ferri et al., 2013) and cranial malformations [Battaglia *et al.*, 1999]. This region appeared to be a hub of SOX2-dependent long-range interactions, that resulted to be lost in *Sox2*-mutant cells. This could be an example of an interesting link between SOX2 and the insurgence of a genetic pathology associated to heterozygous deletions. Perhaps, the deletion removes enhancer(s) of genes, or perturbs their pattern of long-range interactions, causing them to be hypofunctional, even if they are not deleted. The enhancer within the intron of the *Akt3* gene, connected to the promoter of a different gene, may be one example. This situation might be

compared to heterozygous mutations of important enhancers of the globin genes (LCR) in delta-beta thalassemia, or of the *Pax6* gene in aniridia [Bhatia *et al.*, 2013].

There are many other intriguing examples, that can be considered for future studies, in which SOX-dependent long-range interactions involve pathological genes, such as Fam161a (associated to retinitis pigmentosa; Langmann *et al.*, 2010). In other SOX2-dependent interactions, the genes involved are related to axon guidance or neural development.

In sum, considering that mutations in regulatory sequences can cause important effects on the expression of their associated genes, this genome-wide approach could be useful to identify other SOX2-dependent DREs, eventually associated to genes involved in genetic disease, to better investigate the regulation of the transcriptional mechanism and the implication of SOX2 in the onset of pathologies.

3. *Nkx2.1* AS A DEFINED SOX2 TARGET GENE

Sox2 deletion by *Bf1-Cre* transgene causes loss of an extended portion of the ventral telencephalon, similar to that observed in the conditional knock-out of the *Shh* [Fuccillo *et al.*, 2004] and *Nkx2.1* [Sussel *et al.*, 1999] genes, two important determinants of brain development. Moreover, the expression of these same determinants (*Nkx2.1* and *Shh*) were lost in our *Sox2*-deleted mice [Ferri *et al.*, 2013]. Furthermore, it is known that NKX2.1 is a direct activator for *Shh*. Giving a pharmacological agonist for SHH to pregnant mice, we noticed that the developing mice presented a partial rescue of brain defects observed after the *Sox2* loss. It proved that the defective activation of *Shh*, due to the loss of its activator NKX2.1, was an important cause of the brain defects seen in *Sox2*-mutant mice.

We proved that *Nkx2.1* is a direct SOX2 target gene and, knowing that NKX2.1 has a crucial role for the ventral patterning of telencephalon [Sussel *et al.*, 1999], our results provide evidence that misregulation of *Nkx2.1* could provide one mechanism by which SOX2 could regulate ventral telencephalic development.

Understanding how SOX2 is able to regulate its target genes and their own implication in embryonic development, could give important information on the regulatory mechanisms behind brain development and cell differentiation.

4. EMX2 AS A SOX2 REPRESSOR IN CORTICAL PATTERNING?

Emx2 is an important homeotic gene, expressed in dorsal telencephalon. Homozygous *Emx2* knock-out in mice present severe abnormalities in brain development, including small cerebral hemispheres and olfactory bulbs [Yoshida *et al.*, 1997].

We found that the transcription factor EMX2 was able to inhibit the binding of BRN2 on its own binding site within an enhancer of *Sox2* [Mariani *et al.*, 2013; see Chapter 4]; we thus wondered if this inhibitory mechanism mediated by EMX2 could be generalized or if it was specific for *Sox2* regulation. Using an enhancer of the *Nestin* gene, in which binding sites for SOX2 and BRN2 were both present, we discovered that EMX2 was able to antagonize the transactivation operated by the complex SOX2-BRN2 and/or also the transactivation operated by SOX2 only, also in this situation [Mariani *et al.*, 2013]. The degree of inhibition was rather significant, already at low doses of EMX2. This confirmed that EMX2 and BRN2 antagonize each other, through a mechanism not specific for *Sox2* locus only.

Moreover, EMX2 appeared to directly antagonize SOX2 function. This is an important consideration, if we consider that EMX2 is a homeotic factor, expressed in the dorsal telencephalon and important for brain patterning. For example, it could antagonize specific ventral genes (such as *Nkx2.1*), limiting their expression domain to the ventral telencephalon. *In vivo* observations are consistent with this hypothesis: in homozygous mice knock-out for *Emx2* and *Pax6*, an expansion of *Nkx2.1* expression to more dorsal telencephalic regions had been observed [Muzio *et al.*, 2002].

Thus, the identification of EMX2 as a direct transcriptional repressor of *Sox2*, acting through the inhibition of BRN2 binding on *Sox2* telencephalic enhancers, and a SOX2 antagonist for the activation of its own target genes, suggests that EMX2 gradients might affect SOX2 (and its targets) levels in different cortical regions. By this means, it could take part in the control of the balance between self-renewal and commitment to neural stem cell differentiation.

5. REFERENCES

- 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**: 126-40.

- Battaglia, A., Carey, J. C., Cederholm, P., Viskochil, D. H., Brothman, A. R., Galasso, C. 1999. Natural history of Wolf-Hirschhorn syndrome: experience with 15 cases. *Pediatrics* **103**: 830-6.

- Bhatia S, Bengani H, Fish M, Brown A, Divizia MT, de Marco R, Damante G, Grainger R, van Heyningen V, Kleinjan DA. 2013. Disruption of autoregulatory feedback by a mutation in a remote, ultraconserved PAX6 enhancer causes aniridia. *Am J Hum Genet.* **93**: 1126-34.

- Capsoni S, Ugolini G, Comparini A, Ruberti F, Berardi N, Cattaneo A. 2000. Alzheimer-like neurodegeneration in aged antinerve growth factor transgenic mice. *Proc Natl Acad Sci U S A*. **97**: 6826-31.

- Fantes J, Ragge NK, Lynch SA, McGill NI, Collin JR, Howard-Peebles PN, Hayward C, Vivian AJ, Williamson K, van Heyningen V, FitzPatrick DR. 2003. Mutations in SOX2 cause anophthalmia. *Nat Genet.* **33**: 461–3.

- Ferri AL, Cavallaro M, Braida D, Di Cristofano A, Canta A, Vezzani A, Ottolenghi S, Pandolfi PP, Sala M, De Biasi S, Nicolis SK. 2004. Sox2 deficiency causes neurodegeneration and impaired neurogenesis in the adult mouse brain. *Development* **131**: 3805-19.

- Jeong Y, Leskow FC, El-Jaick K, Roessler E, Muenke M, Yocum A, Dubourg C, Li X, Geng X, Oliver G, Epstein DJ. 2008. Regulation of a remote Shh forebrain enhancer by the Six3 homeoprotein. *Nat Genet.* **40**: 1348-53.

- Kelberman D, Rizzoti K, Avilion A, Bitner-Glindzicz M, Cianfarani S, Collins J, Chong WK, Kirk JM, Achermann JC, Ross R, Carmignac D, Lovell-Badge R, Robinson IC, Dattani MT. 2006. Mutations within Sox2/SOX2 are associated with abnormalities in the hypothalamo-pituitary-gonadal axis in mice and humans. *J Clin Invest.* **116**: 2442–55.

- Langmann T, Di Gioia SA, Rau I, Stohr H, Maksimovic NS, Corbo JC, Renner AB, Zrenner E, Kumaramanickavel G, Karlstetter M, Arsenijevic Y, Weber BHF, Gal A, Rivolta C. 2010. Nonsense mutations in FAM161A cause RP28-associated recessive retinitis pigmentosa. *Am. J. Hum. Genet.* **87**: 376-81.

- Mariani J, Favaro R, Lancini C, Vaccari G, Ferri AL, Bertolini J, Tonoli D, Latorre E, Caccia R, Ronchi A, Ottolenghi S, Miyagi S, Okuda A, Zappavigna V, Nicolis SK. 2012. Emx2 is a dose-dependent negative regulator of Sox2 telencephalic enhancers. *Nucleic Acids Res.* **40**: 6461-76.

- Muzio L, DiBenedetto B, Stoykova A, Boncinelli E, Gruss P, Mallamaci A. 2002. Conversion of cerebral cortex into basal ganglia in Emx2(-/-) Pax6(Sey/Sey) double-mutant mice. *Nat Neurosci.* **5**: 737-45.

- Pevny LH, Sockanathan S, Placzek M, Lovell-Badge R. 1998. A role for SOX1 in neural determination. D*evelopment* **125**: 1967-78.

- Ragge NK, Lorenz B, Schneider A, Bushby K, de Sanctis L, de Sanctis U, Salt A, Collin JR, Vivian AJ, Free SL, Thompson P, Williamson KA, Sisodiya SM, van Heyningen V, Fitzpatrick DR. 2005. SOX2 anophthalmia syndrome. *Am J Med Genet A.* **135**: 1–7.

- Riviere JB, Mirzaa GM, O'Roak BJ, Beddaoui M, Alcantara D, Conway RL, St-Onge J, Schwartzentruber JA, Gripp KW, Nikkel SM, Worthylake T, Sullivan CT, and 29 others. 2012. De novo germline and postzygotic mutations in AKT3, PIK3R2 and PIK3CA cause a spectrum of related megalencephaly syndromes. *Nature Genet.* **44**: 934-40.

- Schneider A, Bardakjian T, Reis LM, Tyler RC, Semina EV. 2009. Novel SOX2 mutations and genotype-phenotype correlation in anophthalmia and microphthalmia. *Am J Med Genet A.* **149A**: 2706–15.

- Sisodiya SM, Ragge NK, Cavalleri GL, Hever A, Lorenz B, Schneider A, Williamson KA, Stevens JM, Free SL, Thompson PJ, van Heyningen V, Fitzpatrick DR. 2006. Role of SOX2 mutations in human hippocampal malformations and epilepsy. *Epilepsia* **47**: 534–42.

- Tziaferi V, Kelberman D, Dattani MT. 2008. The role of SOX2 in hypogonadotropic hypogonadism. *Sex Dev.* **2**: 194–9.

- Yamamoto A, Lucas JJ, Hen R. 2000. Reversal of neuropathology and motor dysfunction in a conditional model of Huntington's disease. *Cell* **101**: 57-66.

- Yoshida M, Suda Y, Matsuo I, Miyamoto N, Takeda N, Kuratani S, Aizawa S. 1997. Emx1 and Emx2 functions in development of dorsal telencephalon. *Development* **124**: 101-11.

- Zappone MV, Galli R, Catena R, Meani N, De Biasi S, Mattei E, Tiveron C, Vescovi AL, Lovell-Badge R, Ottolenghi S, Nicolis SK. 2000. Sox2 regulatory sequences direct expression of a (beta)-geo transgene to telencephalic neural stem cells and precursors of the mouse embryo, revealing regionalization of gene expression in CNS stem cells. *Development* **127**: 2367-82.

- Visel A, Taher L, Girgis H, May D, Golonzhka O, Hoch RV, McKinsey GL, Pattabiraman K, Silberberg SN, Blow MJ, Hansen DV, Nord AS, Akiyama JA, Holt A, Hosseini R, Phouanenavong S, Plajzer-Frick I, Shoukry M, Afzal V, Kaplan T, Kriegstein AR, Rubin EM, Ovcharenko I, Pennacchio LA, Rubenstein JL. 2013. A high-resolution enhancer atlas of the developing telencephalon. *Cell* **152**: 895-908.

- Zhang Y, Wong CH, Birnbaum RY, Li G, Favaro R, Ngan CY, Lim J, Tai E, Poh HM, Wong E, Mulawadi FH, Sung WK, Nicolis S, Ahituv N, Ruan Y, Wei CL. 2013. Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. *Nature* **504**: 306-10.