SCUOLA DI DOTTORATO

UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

School of Medicine and Faculty of Science

University of Milano – Bicocca

PhD program in Translational and Molecular Medicine DIMET

Cycle XXIX

Identification and functional characterization of Sox2-target genes involved in brain disease and abnormal brain development

Dr. Miriam Pagin

Registration number 787791

Tutor: **Prof. Silvia K. Nicolis**

Coordinator: **Prof. Andrea Biondi**

ACADEMIC YEAR 2015-2016

TABLE OF CONTENTS

CHAPTER 1:

GENERAL INTRODUCTION

1. SOX FAMILY GENES

Sox genes encode a group of transcriptional factors that carry a DNA binding HMG (High-Mobility Group) domain and additional domains implicated in transcriptional regulation. *Sox* genes are expressed in various phases of embryonic development and cell differentiation in a manner linked to cell specification [Kamachi et al., 2000]. The HMG 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 male sex determination in mammalian and located on the chromosome Y. In fact, Sox means Sry-related HMG-box.

HMG domains of SOX proteins share highly similar DNA-binding properties, recognizing only 6–7 base pair DNA sequences. Nevertheless, an individual SOX protein appears to interact selectively with and to regulate a unique repertoire of target genes. The same SOX protein appears to regulate different sets of target genes depending on the cell type in which the protein is expressed [Kamachi et al., 2000].

Sox genes have been identified in mammalians, reptiles, amphibians, fishes, insects and nematodes [Bowles et al., 2000]. In vertebrates, SOX proteins are divided in ten groups, called A-J, and they have a crucial role in the

development of the nervous system, eye, cartilage, blood vessels, sex determination 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].

The B group of SOX factors plays 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 proteins 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 embryonic development, we observe an overlap of their expression patterns, suggesting a functional interaction of these genes during organogenesis and potential common target genes [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 they have a role in the maintenance of 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 neural differentiation [Jager et al., 2011].

2. THE SOX2 TRANSCRIPTION FACTOR AND ITS IMPLICATION IN BRAIN DEVELOPMENT AND NEURAL STEM CELLS AND DIFFERENTIATION

Sox2 is a member of the Sox (SRY-related HMG box) gene family that encodes transcription factors with a single HMG DNA-binding domain. SOX2 belongs to the SOX B1 subgroup based on homology within and outside the HMG box. It is transiently expressed in the inner cell mass and the epiblast of the blastocyst and, later, throughout the developing neuroephitelium.

The *Sox2* gene is localized on chromosome 3, both in mouse and human. It is highly conserved and it is composed of a single exon, encoding for 2,4 kilobases transcript. The protein is composed of 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 mouse embryonic development, *Sox2* is first detected in some cells at morula stages (E2.5) and in the blastocysts (E3.5) it is expressed specifically within the ICM (Inner Cell Mass). *Sox2* expression persists throughout the epiblast, but by mid-late-streak stages (E7-E7.5), it becomes restricted to presumptive neuroectoderm anteriorly, while it is excluded from the posterior, including cells ingressing through the primitive streak. A second site of expression initiates in the ectoplacental cone, and this is maintained in ExE at E6.5. By E9.5, *Sox2* RNA is seen throughout the brain, neural tube, sensory placodes, branchial arches, and gut endoderm (Figure

1). It is also present in both male and female germ cells [Avilion et al., 2003].

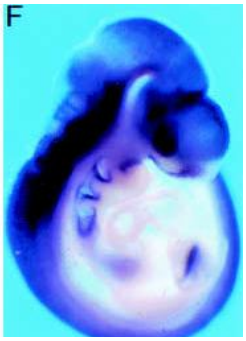


Figure 1

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].

Figure 2: *Sox2*β-geo expression (X-gal staining) in a E12,5 mouse embryo; *Sox2* has a pan-neural expression [Ferri et al., 2004].

In the adult brain, SOX2 is expressed both in sparse cells within differentiated regions (cortex, striatum, thalamus) and within neurogenic regions (periventricular ependyma and subependyma, hippocampus). In particular, within differentiated regions, SOX2 is expressed in at least some characteristic pyramidal cells in the cerebral cortex, in differentiated neurones in the striatum and, abundantly, in the thalamus (particularly in the periventricular nuclei). In the ventricular zone, Sox2 is strongly expressed in ependymal cells and choroid plexus. Furthermore, Sox2 remains highly expressed in neurogenic regions: the lateral ventricle, from where expression extends along the entire rostral migratory stream (RMS), along which dividing precursors migrate to the olfactory bulb, and in the germinative layer of the hippocampus 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; Pevny 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 *Sox2* mutants.

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 studied the defects derived after *Sox2*-ablation by *Nestin-Cre* transgene. Sox2 conditional knock-out experiments have shown that Sox2 is necessary for neural stem cell maintenance *in vitro* (in neurosphere cultures) and *in vivo* (in the dentate

gyrus of the hippocampus), and its action is partially mediated by the direct regulation of the Shh-pathway [Favaro et al., 2009]

Nestin-Cre transgene:

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) starting at 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 the hippocampus was compromised: its size was markedly reduced, particularly caudally, in comparison with wild-type (wt) pups, resulting in an underdeveloped dentate gyrus. The hippocampal defect was also observed in the adult *Sox2*-mutant brain [Favaro et al., 2009].

In vitro, deletion of *Sox2* caused loss of self-renewal of neurosphere cultures, obtained from the dissection of *Sox2*-deleted mouse forebrains; this observation supports the important role of *Sox2* in NSC maintenance observed *in vivo* [Favaro et al., 2009].

Interestingly, by infecting *Sox2*-deleted neuroshpere cultures with a lentivirus encoding *Sox2*, it was observed a partial rescue of the defects previously described. The infected neurospheres showed rescue in self-renewal [Favaro et al., 2009].

In the nervous system, *Sox2* is expressed, and is functionally important, at the earliest developmental stages of the vertebrates. In the mouse nervous system, *Sox2* is expressed in stem cells and early precursors, and in few mature neurons [Zappone et al., 2000; Ferri et al., 2004]. Adult *Sox2*-deficient mice, in which *Sox2* expression is decreased by about 70%, exhibit

neural stem/precursor cell proliferative defects in the hippocampus and periventricular zone [Ferri et al., 2004]. Moreover, neurons containing neurofilament/ubiquitin-positive aggregates are observed, together with dead neurons, in thalamic and striatal parenchyma, which are already substantially reduced in size at early developmental stages. These observations point to a possible role for *Sox2* in the maturation and/or survival of embryonic and adult neurons. *In vitro* differentiation studies on neurosphere-derived neural cells were performed: *in vitro*, *Sox2*-deficient cells exhibit a striking differentiation defect, characterized by abnormal morphology and decreased expression of mature differentiation markers. This is confirmed by the in vitro rescue experiment with a *Sox2*-expressing lentivirus. *Sox2* overexpression in mutant cells at the onset of differentiation is necessary to rescue the well-arborized β-tubulin-positive, MAP2-positive phenotype observed in normal, but not in mutant cells [Cavallaro et al., 2008]. However the mutant cells generate normal glia; in neuronal committed cells, *Sox2* might act to repress part of a gliogenic transcription program.

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 DC5, a minimal enhancer element of the target gene. This protein-complex formation is crucial for the beginning of the lens differentiation and crystalline development [Kamachi et al., 2001].

There are also many examples of protein interaction between SOX2 and POU proteins (other transcription factors important during development), such as the SOX2-OCT3 protein complex that binds the *Fgf4* (fibroblast growth factor 4) gene. OCT-3 POU domain and the SOX2 HMG domain can mediate both direct protein-protein interaction and cooperative binding to the FGF-4 enhancer DNA. Direct protein-protein contact between SOX2 and OCT-3 is a prerequisite for the formation of a functional OCT-3 complex and that this interaction is facilitated by the specific arrangement of the factor binding sites within the FGF-4 enhancer. This observation underscores the fundamental role played by enhancer DNA in stereospecific assembly of active transcription complexes and illustrates one regulatory mechanism by which common DNA elements, which are widely dispersed throughout the genome and can bind multiple proteins, can nonetheless direct tissue-specific, selective transcriptional activation [Ambrosetti et al., 1997]. These same factors are also involved in the regulation of the *Utf1* (undifferentiated transcription factor 1) gene expression: SOX2 is involved in stimulation of transcription of the *Utf1* gene by cooperating with OCT3-4 [Nishimoto et al., 1999]. It is also known that *Sox2* and *Oct3-4* 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-OCT3-4 protein complex is also implicated in the regulation of *Nanog,* another gene essential for pluripotency. Nanog represents the first known transcription factor appearing after compaction, and specific to the inner cells of the morula, both *Oct4* and *Sox2* are expressed prior to compaction in all blastomeres [Rodda et al., 2005].

Instead, the protein complex SOX2-BRN2 is active in more differentiated cells among neural lineage (neural stem/progenitor cells). Consistent with a functional role, forced expression of BRN2 in differentiating ESCs led to recruitment of SOX2 to a subset of NPC distal enhancers. This recruitment was associated with changes in chromatin structure, activation of neighboring genes, and ultimately precocious differentiation toward a neural-like state [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, optic nerve hypoplasia, ocular coloboma, retinal and chorioretinal dystrophy) [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 Many of these DREs, including cell specific enhancers, were characterized for their vital function in development and differentiation. Increasing evidence has shown that DREs can function over long distances, even on a different chromosome from their target genes. [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 formaldehyde 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 (Figure 3) [Zhang et al., 2013; Fullwood et al., 2009].

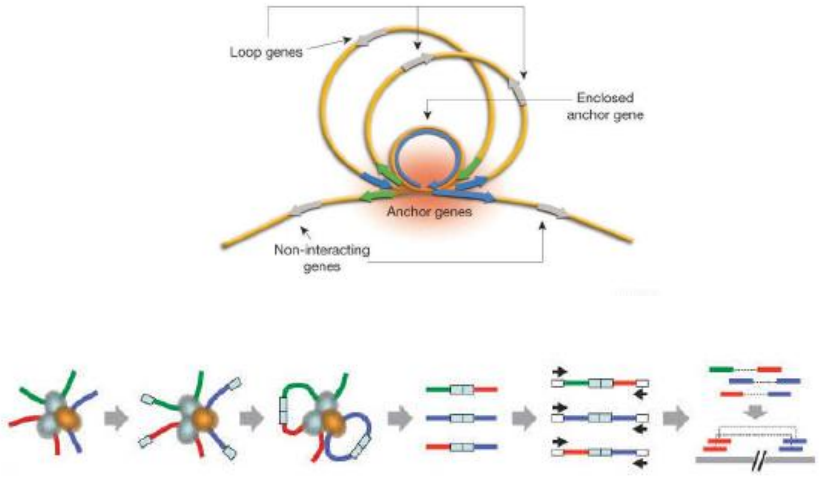


Figure 3. Genome-wide detection of long-range DNA interactions in chromatin (ChIA-PET) [Fullwood et al., 2009]

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. The expression levels of the genes involved in the RNAPII interactions are significantly higher than those with no detected interaction, suggesting that their promoters are

transcriptionally more active. Interestingly, these data suggest that a consistent proportion of the identified putative enhancers do not regulate 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 ES cells, bivalent domains frequently overlay developmental transcription factor genes expressed at very low levels. Bivalent domains tend to resolve during ES cell differentiation and, in differentiated cells, developmental genes are typically marked by broad regions selectively enriched for either Lys27 or Lys4 methylation. This suggests that bivalent domains silence developmental genes in ES cells while keeping them poised for activation. [Bernstein et al., 2006].

In addition, 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 CHROMOSOME CONFORMATION CAPTURE TECHNOLOGIES

3C technology: toward three-dimensional (3D) genomics.

3C is another genomics strategy to analyze contact frequencies between selective genomic sites in cell populations. The initial step in 3C and 3C-derived methods is to establish a representation of the 3D organization of the DNA. To this end, the chromatin is fixed using a fixative agent, most often formaldehyde [Dekker et al. 2002]. Next, the fixed chromatin is cut with a restriction enzyme recognizing 6 base pairs (bp)—such as HindIII, BglII, SacI, BamHI. In the subsequent step, the sticky ends of the cross-linked DNA fragments are religated under diluted conditions to promote intramolecular ligations (i.e., between cross-linked fragments). DNA fragments that are far away on the linear template, but colocalize in space, can, in this way, be ligated to each other. A template is thereby created that is, in effect, a onedimensional (1D) cast of the 3D nuclear structure. The way to establish the 3D conformation of a locus or chromosome is to measure the number of ligation events between nonneighboring sites. In 3C, this is done by semiquantitative or quantitative PCR [De Wit et al., 2012].

Chromosome conformation capture-on-chip (4C) technology.

4C originally combined 3C technology with microarrays to analyze the contacts of a selected genomic site with all of the genomic fragments that are represented on the array [Simonis et al., 2006]. 4C-seq refers to the same strategy, but uses next-generation sequencing (NGS) instead of microarrays to analyze contacting sequences [Splinter et al. 2011]. in 4C technology, the ligated 3C template is processed with a second round of DNA digestion and ligation to create small DNA circles. Using viewpoint-

specific primers, inverse PCR specifically amplifies all sequences contacting this chromosomal site. They can then be analyzed by microarrays or, nowadays, by NGS methods [De Wit et al, 2012].

Chromosome conformation capture carbon copy (5C) technology.

5C can be described as a ‘‘many versus many’’ technology. It allows concurrent determination of interactions between multiple sequences [Dostie et al. 2006]. In 5C, the 3C template is hybridized to a mix of oligonucleotides, each of which partially overlaps a different restriction site in the genomic region of interest. Pairs of oligonucleotides that correspond to interacting fragments are juxtaposed on the 3C template and can be ligated together. Since all 5C oligos carry one of two universal sequencesat their 5’ ends, all ligation products can subsequently be amplified simultaneously in a multiplex PCR reaction. Readout of these junctions occurs either on a microarray or by high-throughput sequencing. The resolution of the technique is determined by the spacing between neighboring oligonucleotides on the linear chromosome template [De Wit et al., 2012].

5.3. DISEASE-ASSOCIATED NON-CODING ELEMENTS

Given the importance of distal regulatory elements in transcriptional regulation, 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 *Sonic hedgehog (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 regulate the associated genes *via* long-range interactions.

5.4. 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. The expression of these key reprogramming genes is known to govern pluripotency in ESCs through coordinated

autoregulatory loops [Jaenisch and Young, 2008]. The authors found that three *Klf* genes (*Klf1*, *Klf2* and *Klf4*) were directly connected to *Sox2*. When the network analysis was extended from one to two hops of connectivity, all of the reprogramming genes were found to be connected within one major hub, except for *Myc* and *Lin28a*, suggesting 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, like *Sall1, Asf1b, Dusp6* and *Jund*; instead, in neural stem cells (NSCs) it could be observed a very different *Sox2* connectivity profile and different enhancers mediated cell-specific connectivity. Such cell-specific connectivity is mediated through differential enhancer usage. Similarly, distinct connectivity maps constructed by oligodendrocyte transcription factors *Olig1* and *Olig2*, genes important for neural cell fate determination are found in NSCs and NPCs, respectively. *Olig1* and *Olig2* are directly connected to many genes critical for neuronal development in NSCs, including neuropilin (*Neto2*), *Fabp7* and *Bsg*. [Zhang et al., 2013].

6. RNA-seq

RNA sequencing (RNA-seq) is a recently developed approach to transcriptome profiling that uses next generation sequencing techniques (also known as deep sequencing because of their potential for high coverage) to study RNA. Understanding the transcriptome is essential for interpreting the functional elements of the genome and revealing the

molecular constituents of cells and tissues, and also for understanding development and disease [Wang et al., 2009; Yongjun Chu and David R. Corey, 2012].

Various technologies have been developed to deduce and quantify the transcriptome, including hybridization-or sequence-based approaches. Specialized microarrays have also been designed; these methods have several limitations, which include: reliance upon existing knowledge about genome sequence; high background levels owing to cross-hybridization; and a limited dynamic range of detection owing to both background and saturation of signals. Moreover, comparing expression levels across different experiments is often difficult and can require complicated normalization methods. the development of novel high-throughput DNA sequencing methods has provided a new method for both mapping and quantifying transcriptomes, the RNA-seq, that has clear advantages over existing approaches RNA-seq library preparation usually includes reverse transcription [Wang et al., 2009].

In general, a population of RNA (total or fractionated, such as poly(A)+) is converted to a library of cDNA fragments with adaptors attached to one or both ends (Figure 4). Each molecule, with or without amplification, is then sequenced in a high-throughput manner to obtain short sequences from one end (single-end sequencing) or both ends (pair-end sequencing). The reads are typically 30–400 bp depending on the DNA-sequencing technology used. Following sequencing, the resulting reads are either aligned to a reference genome or reference transcripts, or assembled *de novo* without the genomic sequence to produce a genome-scale transcription map that consists of both the transcriptional structure and/or level of expression for each gene [Wang et al., 2009].

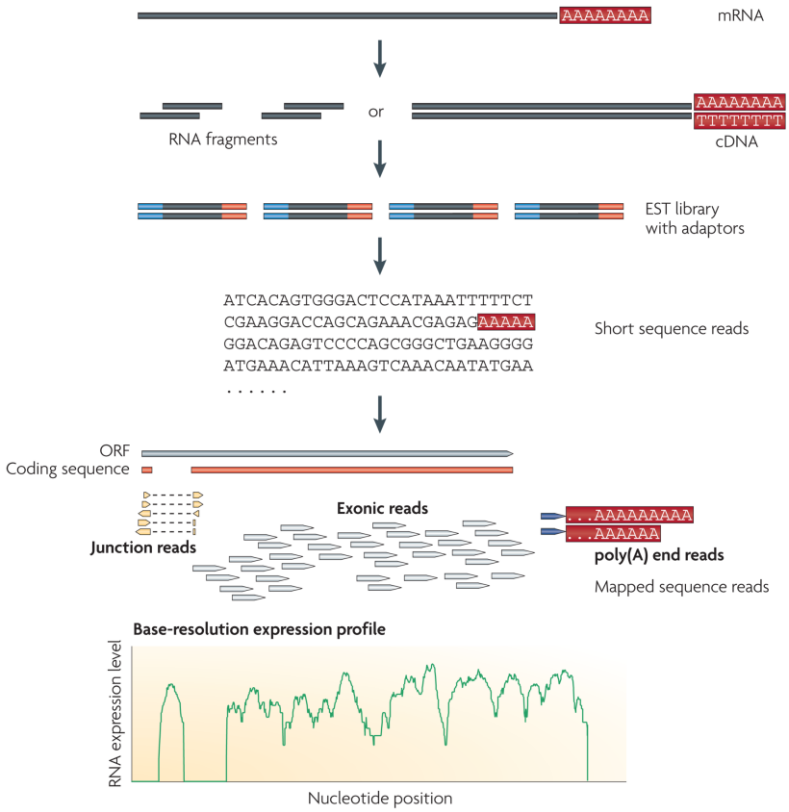


Figure 4. A typical RNA-seq experiment [Wang et al., 2009]. Long RNAs are first converted into a library of cDNA fragments through either RNA fragmentation or DNA fragmentation. Sequencing adaptors (blue) are subsequently added to each cDNA fragment and a short sequence is obtained from each cDNA using high-throughput sequencing technology. The resulting sequence reads are aligned with the reference genome or transcriptome, and classified as three types: exonic reads, junction reads and poly(A) end-reads.

RNA-seq offers several advantages over existing technologies: unlike hybridization-based approaches, RNA-seq is not limited to detecting transcripts that correspond to existing genomic sequence. This makes RNA-seq particularly attractive for non-model organisms with genomic sequences that are yet to be determined. RNA-seq can reveal the precise

location of transcription boundaries, to a single-base resolution Furthermore, 30-bp short reads from RNA-seq give information about how two exons are connected, whereas longer reads or pairend short reads should reveal connectivity between multiple exons. These factors make RNA-seq useful for studying complex transcriptomes. In addition, RNA-seq can also reveal sequence variations (for example, SNPs) in the transcribed regions. A second advantage of RNA-seq relative to DNA microarrays is that RNA-seq has very low background signal. RNA-seq has also been shown to be highly accurate for quantifying expression levels and the results have high levels of reproducibility. Taking together all these advantages, RNA-seq is the first sequencing method that allows the entire transcriptome to be surveyed in a very high-throughput and quantitative manner.

Unlike small RNAs (microRNAs (miRNAs), Piwi-interacting RNAs (piRNAs), short interfering RNAs (siRNAs) and many others), which can be directly sequenced after adaptor ligation, larger RNA molecules must be fragmented into smaller pieces (200–500 bp) to be compatible with most deep-sequencing technologies. Common fragmentation methods include RNA fragmentation (RNA hydrolysis or nebulization) and cDNA fragmentation (DNase I treatment or sonication) [Wang et al., 2009].

In some cases the direction of the transcription is lost during strand amplification, but methods are available for defining transcript direction. Data analysis of RNA-seq may include transcript assembly, alternatively spliced transcript, or novel transcript discovery and transcript quantitation. [Chu and Corey, 2012]

Mapping the RNA-seq reads through informatics instruments leads to several challenges including the development of efficient methods to store,

retrieve and process large amounts of data and also remove low-quality reads. Once high-quality reads have been obtained, the first task of data analysis is to map the short reads from RNA-seq to the reference genome. There are several programs for mapping reads to the genome, including ELAND, SOAP, MAQ and RMAP. For large transcriptomes, alignment is also complicated by the fact that a significant portion of sequence reads match multiple locations in the genome.

A major complication in quantification is the fact that RNA-seq reads do not always map uniquely to a single gene or isoform and there are a few informatics tools that can perform accurate trascriptome quantification in the absence of a reference genome: one of these is RSEM, which has an important role for the quantification of gene expression for the reconstruction of *the novo* transcriptome [Li and Dewey, 2011].

The RNA-seq analysis is useful to compare gene expression for example between biological different conditions to determine which genes are differentially expressed by using tools that count the number of fragments for each genes and compare them between the different analysed conditions. The most common tools for this type of analysis are DESeq [Anders and Huber, 2010] and EdgeR [Robinson et al., 2010].

7. *Socs3*: A *Sox2*-TARGET GENE

One of the putative Sox2 target genes derived from ChIA-PET and RNA-seq experiments is Socs3. Socs3 (Suppressor Of Cytokine Signaling 3) is an inducible negative feedback inhibitor of cytokine signalling.

Cytokines are essential regulators of diverse biological processes such as embryogenesis, hemopoiesis, and innate and adaptive immunity. Many cytokines, including interleukins, interferons, and hemopoietic growth factors, exert their biological effects via the JAK-STAT (Janus kinase-signal transducer and activator of transcription) pathway. Cytokines binding to specific cell surface receptors leads to the activation of receptor-associated JAK kinases and JAK-mediated phosphorylation of the receptor cytoplasmic domains. Phosphorylated receptors then recruit STAT proteins, which are in turn phosphorylated, translocate to the nucleus and exert transcriptional control via binding to target gene regulatory sequences. The suppressor of cytokine signaling (Socs) proteins are key regulators of the JAK-STAT pathway, ensuring that activation of the pathway for essential cellular processes is spatially and temporally controlled to prevent pathology. The Socs family has eight members—Socs1, Socs2, Socs3, Socs4, Socs5, Socs6, Socs7, and CIS (cytokine-inducible SH2 protein) [White and Nicola, 2013].

The expression of Socs3 is tightly regulated at both the mRNA and protein levels. *Socs3* transcription is rapidly induced by a range of type I and type II cytokines, signaling via STAT1 and/or STAT3. The best characterized of these are members of the gp130 family, including IL-6, IL-11, and LIF. Induction of *Socs3* by IL-6 also requires specificity protein 3 (Sp3), while TNFα induces *Socs3* expression via activation of the MKK6/MAPK cascade, and LPS induction of *Socs3* involves the MAPK-ERK1/2 and JNK pathways downstream of Toll-like receptor 4. In addition, growth hormone induces *Socs3* expression via CREB/c-Fos/c-Jun and FOXO3a and cAMP induces *Socs3* expression via Epac-1 (exchange protein directly activated by cAMP). Other known inducers of *Socs3* transcription include leptin, IFN-λ, IL-1, IL-9, IL-10 and IL-21. Socs3 protein levels are tightly controlled via proteolysis,

using both proteasomal and non-proteasomal pathways [White and Nicola, 2013].

Homozygous null mutation of *Socs3* is lethal between embryonic day 11 (E11) and E13, due to defects in placental development [White and Nicola, 2013].

Moreover, several evidences suggested that Socs3 plays a negative regulatory role in *Stat3*. Socs3 has been shown to interact with gp130 and JAK, and this interaction results in relatively specific inhibition of gp130 signaling [Nicholson et al. 2000; Schmitz et al. 2000]. Transient expression of Socs3 inhibits LIF-induced *Stat3* reporter gene activation [Rakesh and Agrawal 2005]. Overexpression of Socs3 promoted neurogenesis and inhibited astrogliogenesis in neural stem cells. Socs3 also promoted maintenance of neural stem cells, probably through up-regulation of Notch1 [Cao et al., 2006; Schmidt-Edelkraut et al., 2013].

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.

- Anders S, Huber W. 2010. Differential expression analysis for sequence count data. Genome biology, 11(10):R106, 2010.

- 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.

- 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.

- Cao F, Hata R, Zhu P, Ma YJ, Tanaka J, Hanakawa Y, Hashimoto K, Niinobe M, Yoshikawa K, Sakanaka M. 2006.

Overexpression of SOCS3 inhibits astrogliogenesis and promotes m aintenance of neural stem cells. J. Neurochem 98: 459-70.

- Cavallaro M1, Mariani J, Lancini C, Latorre E, Caccia R, Gullo F, Valotta M, DeBiasi S, Spinardi L, Ronchi A, Wanke E, Brunelli S, Favaro R, Ottolenghi S, Nicolis SK. 2008. Impaired generation of mature neurons by neural stem cells from hypomorphic Sox2 mutants. Development. 135: 541-57.

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

- Chu Y, Corey DR. 2012. RNA sequencing: platform selection, experimental design, and data interpretation. Nucleic Acid Ther. 22: 271-4.

- Dekker J, RippeK, Dekker M, Kleckner N. 2002. Capturing chromosome conformation. Science. 295: 1306-11.

- De Wit E, de Laat W. 2012. A decade of 3C technologies: insights into nuclear organization. Genes Dev. 26: 11-24.

- Dostie J, Richmond TA, Arnaout RA, Selzer RR, Lee WL, Honan TA, Rubio ED, Krumm A, Lamb J, Nusbaum C, Green RD, Dekker J. 2006. Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. Genome Res. 16: 1299-309.

- 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.

- Li B, Dewey CN. 2011. Rsem: accurate transcript quantification from rna-seq data with or without a reference genome. BMC bioinformatics, 12(1):323.

- 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.

- 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.

- Nicholson SE, De Souza D, Fabri LJ, Corbin J, Willson TA, Zhang JG, Silva A, Asimakis M, Farley A, Nash AD, Metcalf D, Hilton DJ, Nicola NA, Baca M. 2000. Suppressor ofm cytokine signaling-3 preferentially binds to the SHP-2-binding site on the shared cytokine receptor subunit gp130. Proc. Natl Acad. Sci. U S A 97, 6493–6498.

- 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. 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.

- Rakesh K. and Agrawal D. K. (2005) Controlling cytokine signaling by constitutive inhibitors. Biochem. Pharmacol. 70, 649–657.

- 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.

- Robinson MD, McCarthy DJ, Smyth GK. 2010. Edger: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics, 26: 139–140.

- 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.

- Schmidt-Edelkraut U, Hoffmann A, Daniel G, Spengler D. 2013. Zac1 regulates astroglial differentiation of neural stem cells through Socs3. Stem Cells. 31: 1621-32.

- Schmitz J, Weissenbach M, Haan S, Heinrich PC, Schaper F. 2000. SOCS3 exerts its inhibitory function on interleukin-6 signal transduction through the SHP2 recruitment site of gp130. J. Biol. Chem. 275, 12 848–12 856

- 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.

- Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E, van Steensel B, de Laat W. 2006. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat Genet. 38: 1348-54.

- 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.

- Splinter E, de Wit E, Nora EP, Klous P, van de Werken HJ, Zhu Y, Kaaij LJ, van Ijcken W, Gribnau J, Heard E, de Laat W. 2011. The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA. Genes Dev. 25: 1371-83.

- Tziaferi V, Kelberman D, Dattani MT. 2008. The role of SOX2 in hypogonadotropic hypogonadism. Sexual Development 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. Mechanisms of Development 84: 103-20.

- Wegner M. 1999. From head to toes: the multiple facets of Sox proteins. Nucleic Acids Research. 27: 1409-1420

- White CA, Nicola NA. 2013. An essential physiological inhibitor of signalling by interleukin-6 and G-CSF family cytokines. JAK-STAT 2:4, e25045.

- 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 functionally characterize novel SOX2-target genes involved in brain disease and abnormal brain development.

It is well known that some SOX2-target genes are involved in different central nervous system 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, optic nerve hypoplasia, ocular coloboma, retinal and chorioretinal dystrophy) [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].

In order to understand the role of *Sox2* in neural development, our laboratory generated *Sox2* conditional knockout (KO) mutations in mouse [Favaro et al., 2009; Ferri et al., 2013]. *Sox2* ablation at different developmental time points produced important brain defects, more serious when the ablation was early. We observed loss of hippocampal neural stem cells (NSC), leading to severe hippocampal reduction [Favaro et al., 2009], and loss of neural stem/precursor cells that constitute the developing ventral telencephalon, leading to sever defects in its embryogenesis [Ferri

et al, 2013]. *Sox2* conditional deletion allowed to observe an important function for *Sox2* also in the maintenance of NSC self-renewal in long-term in vitro NSC cultures. *Sox2*-ablated NSC, cultured as neurospheres, self-renewed for several passages in culture, as the wild-type (wt) ones, but then underwent a decrease in growth, with progressive culture exhaustion. Sphere formation could be rescued by lentiviral *Sox2* [Favaro et al., 2009]. This reveled an essential role for *Sox2* in the development of multiple CNS regions and in the maintenance of NSC.

To understand the mechanisms of *Sox2* function, a central question is which genes SOX2 regulates as a transcription factor, by what mechanisms SOX2 acts in regulating them, and which SOX2-regulated genes are critical mediators of its function. A new way in which SOX2 regulates its targets has been recently observed in our laboratory: SOX2 maintains a high number of long-range interactions between genes and distal enhancers, that regulate gene expression. The ChIA-PET (Chromatin Interaction Analysis by Paired-End Tag sequencing) technique was used (in collaboration with Dr. C.L. Wei) to obtain a genome-wide map of these long-range chromatin interactions, comparing neurosphere cultures of wild-type NPC and *Sox2*-deleted NPC. Wt and *Sox2*-deleted NPC 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. 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 (in collaboration with F. Guillemot, London) the wt NPC 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 project, the aims of my work were:

- to study the differentiation of *Sox2*-ablated cells, as compared to controls; in previous work, it was shown that *Sox2*-deleted cells have an important self-renewal defect and their growth in culture becomes exhausted after few passages [Favaro et al., 2009]. I studied differentiation into neurons and glia: very few beta-tubulinIII (neuronal marker) stained cells were observed in *Sox2*-ablated cells differentiated, indicating that *Sox2* has an important role in neuronal differentiation.

- to identify which genes were deregulated after *Sox2* ablation during self-renewal and differentiation of NSC, using RNA-seq analysis, which identified hundreds of genes deregulated after *Sox2* ablation. The most down-regulated gene was *Socs3*.

- To test selected genes, deregulated following *Sox2* ablation, as possible mediators of *Sox2* function in neural stem cells, by rescuing experiments. *Socs3* is a gene severely downregulated in mutant cells following *Sox2* loss. I reintroduced *Socs3* in mutant cells via a lentiviral vector; this significantly rescued the self-renewal defect of mutant cells. I also tested if the reintroduction of *Socs3* could rescue the neuronal differentiation defect of mutant cells and my initial experiments suggest that this might be the case. Then I aimed to test the role of some of the other most deregulated genes

as mediators of *Sox2* function in self-renewal and differentiation, by rescuing experiments of mutant cells. I generated lentiviral vectors to express a small group of genes, deregulated following *Sox2* loss and known to play important roles in cell proliferation (*Fos, Jun, JunB, Egr1, Egr2*), and I will use them to transduce mutant neural stem cells (in combination, or individually), as I did for *Socs3*. I will evaluate if any specific gene, among those, is able to rescue, upon overexpression, the ability to self-renew and the neuronal differentiation in mutant cells.

- Finally, I aim to test if *Sox2* reintroduction in mutant cells could rescue the long-range interactions of a small number of identified target genes, lost in *Sox2*-deleted cells, by 3C experiments.

REFERENCES

- 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.

- 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.

- 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.

- 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. Sexual Development 2: 194-9.

CHAPTER 2:

(Manuscript submitted to Cell Stem Cell)

Sox2 is required for global functional chromatin connectivity in brain-derived neural stem cells

Jessica Bertolini[1]*, Rebecca Favaro[1]*, Chee-Hong Wong[2]*, Miriam Pagin[1], Chew Yee Ngan[2], Marit Vermunt[3], Ben Martynoga[4], Cristiana Barone[1], Jessica Mariani[1], Marcos Julián Cardozo[5], Noemi Tabanera[5], Federico Zambelli[6], Sara Mercurio[1], Sergio Ottolenghi[1], Paul Robson[2,7], Menno Creyghton[3], Paola Bovolenta[5], Giulio Pavesi[6], Francois Guillemot[4], Silvia K. Nicolis[1#§], Chia-Lin Wei[2,8#]

1 Department of Biotechnology and Biosciences, University Milano-Bicocca, 20126 Milano, Italy
2 The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA
3 Hubrecht Institute-KNAW and University Medical Center Utrecht 3584CT, Utrecht, The Netherlands
4 The Francis Crick Institute, Midland Road, London NW 1AT, UK
5 Centro de Biología Molecular Severo Ochoa, Consejo Superior de Investigaciones Científicas - Universidad Autónoma de Madrid and Ciber de Enfermedades Raras (CIBERER), ISCIII Madrid Spain
6 Department of Biosciences, University of Milano, 20100 Italy
7 Stem Cell & Regenerative Biology, Genome Institute of Singapore, Singapore
8 Institute for Systems Genomics, University of Connecticut, CT, USA

* co-first authors
# corresponding authors
§ Lead contact

Summary

Lineage-specific transcription factors establish cell type-specific gene expression by binding to promoters and distal enhancers, which are brought into contact via long-range chromatin interactions. The SOX2 transcription factor is critical for neural stem cells (NSC) maintenance and brain development. We determined, by genome-wide chromatin interaction analysis (RNApolII ChIA-PET), the global pattern of long-range chromatin interactions in normal and *Sox2*-deleted mouse NSC. In normal

NSC, distal regions interacting with promoters were highly enriched in SOX-2 bound enhancers. *Sox2* deletion caused extensive loss of long-range interactions and reduced expression of a subset of genes associated with *Sox2*-dependent interactions. Expression of one of these genes, *Suppressor of Cytokine Signaling 3* (*Socs3*), rescued the self-renewal defect of SOX2-ablated NSC. Overall, our work identifies SOX2 as a major regulator of functional chromatin connectivity in NSC, and demonstrates the role of genes associated with *Sox2*-dependent interactions in NSC maintenance and, potentially, in neurodevelopmental disorders.

Keywords: SOX2, Chia-PET, Neural stem cells, transcription factors, chromatin connectivity

Introduction

Neural stem cells are critical for brain development and for postnatal maintenance of neurogenesis in specific brain areas. SOX2, a transcription factor (TF) essential for pluripotency (Avilion et al., 2003; Takahashi and Yamanaka, 2006), is also required for correct brain development. In humans, *SOX2* mutations cause genetically dominant nervous system disease, involving hippocampus and eye defects, epilepsy and learning disabilities (OMIM 206900). In mice, *Sox2* ablation causes similar defects, such as hippocampal hypoplasia, microcephaly, ventral forebrain depletion, and anophtalmia, some of which may result from a defect in NSC self-renewal (Favaro et al., 2009; Ferri et al., 2013). These in vivo defects are reflected in the inability of Sox2-deleted neural stem cells to self-renew in long-term cultures (Favaro et al., 2009). The molecular mechanisms by which *Sox2* regulates NSC remain mostly unknown.

Gene promoters are regulated by transcription factors via looping to enhancers (Deng et al., 2012; Noordermeer et al., 2011), often located at large distances and skipping intervening genes (de Laat and Duboule, 2013; Rivera and Ren, 2013; Sanyal et al., 2012; Zhang et al., 2013). Genome-wide analyses of long range interactions in chromatin (Sanyal et al., 2012; Zhang et al., 2013) define complex three-dimensional networks (the connectome), whereby a promoter may interact not only with enhancers but also with additional promoters, which are in turn connected to further promoter(s) and/or enhancer(s). The genome-wide connectome is cell-type specific (Gorkin et al., 2014; Zhang et al., 2013), presumably reflecting cell type-specific transcription factor representation. So far, it is unknown to what extent a given transcription factor is relevant for the establishment/maintenance of a specific genome-wide interaction network and for the function of the interaction network in controlling gene activity.

We previously used ChIA-PET to identify RNA-polII-mediated long-range interactions in ES cells (ESC) and in brain-derived neural stem/progenitor cells (NSC) cultures (Zhang et al., 2013). Most enhancer-promoter interactions were highly cell-type-specific, differing between ESC and NSC. In this work, we sought to identify molecular mechanisms underlying Sox2-dependent gene regulation in NSC, as well as genes involved in Sox2-dependent maintenance of long-term NSC self-renewal. We thus deleted Sox2 in NSC in mouse embryo, and studied the effects of loss of embryonic Sox2 onto the RNApolII-mediated chromatin long-range interaction network in neonatal NSC grown in vitro (as in (Favaro et al., 2009). We unexpectedly observed a major loss of long-range interactions in Sox2-deleted (mut NSC) versus undeleted NSC (wt NSC). We identified thousands

of genes connected via long-range interactions to distal Sox2-bound, epigenetically defined enhancers; many of these genes, including important neurodevelopmental genes, were downregulated upon Sox2 ablation. We validated one of these as a critical downstream Sox2 target whose re-expression in mut NSC is sufficient to rescue their self-renewal defect.

Results

Sox2 loss profoundly alters the genome-wide RNApolII-mediated pattern of long-range chromatin interactions in NSC

We established (Favaro et al., 2009) NSC cultures from the forebrain of conditionally *Sox2*-ablated mice and their control non-deleted littermates. While *Sox2*-deleted (mut) and control (wt) NSC initially expand in culture with similar kinetics, mut NSC later fail to self-renew, pointing to a requirement for Sox2 in NSC maintenance that we also observed *in vivo* in the hippocampus (Favaro et al., 2009).

To determine the effect of Sox2 loss on the genome-wide pattern of RNA-polII-mediated long-range interactions, we performed ChIA-PET analysis with anti-polII antibodies, specific for the preinitiation complex (Zhang et al., 2013), comparing, at P0, ex-vivo NSC cultures derived from normal and Sox2-deleted (at E11.5) forebrains (Fig. 1A). These cultures contain forebrain–specific, but not more posterior, transcripts, e.g. from Hoxa genes, indicating that the NSC maintain a forebrain identity (Zappone et al., 2000; Zhang et al., 2013).

ChIA-PET identifies inter-molecular ligation reads, derived from distant DNA regions brought together in association with immunoprecipitated RNApolII, defining long-range interactions (Zhang et al., 2013). In total, we identified

7046 and 2984 interactions in wt and mut NSC, respectively (see Table S1 for summary of sequencing data processing). The lower number of nodes detected in mut NSC was not due to reduced efficiency in the ChIA-PET procedure in these cells. In fact, the sequencing depths were equivalent between these two cell types. The numbers of intrachromosomal-ligated paired end tags (PETs) to determine both the RNApolII binding sites and the interactions in mut NSC (1.3 million) was more than two-fold higher that for wt NSC (691K). While similar numbers of RNApolII binding sites were detected (Fig. S1), less than half of the intra-chromosomal interactions were observed in mut NSC (Table S1). DNA regions participating in such interactions are termed "anchors", and overlapping anchors are termed "nodes" (Fig. 1B). Nodes or anchors residing within 2.5kb from an annotated transcription start site (TSS) are listed as "promoter node/anchor" (proximal), whereas the remaining ones are termed "non-promoter node/anchor" (distal) (Fig. 1B and Table S1). In mut NSC, the number of unique nodes was strongly decreased relative to wt NSC (4653 versus 8732); both promoter (P) and non-promoter (non-P) nodes were decreased by almost 50% (Fig. 1B, Table S1). Only a small proportion of the nodes (1864 promoter nodes, 561 non-promoter nodes) were common to wt and mut NSC, and the majority of the non-common nodes were found in wt NSC (3219 P nodes, 3088 non-P nodes; Fig. 1B). These data indicate that *Sox2* is required for the maintenance of a large fraction of the nodes of wt NSC.

Interactions were classified on the base of characteristics of <u>both</u> interacting anchors; they were subdivided according to i) type of interacting element, i.e.promoter (P) or non-promoter (non-P) (Fig. 1C), and ii) detection of the interaction in wt, or mut, or both cell types (WT, MUT, COMMON interaction)(Fig. 1D). We also defined "alternative usage"

interactions, in which only one of the two connected anchors is conserved between wt and mut NSC, and the other one(s) vary (Fig. 1D, bottom). About 90% of the interactions found in wt cells were mediated through promoters (promoter-promoter (P-P) or promoter-non promoter (P-nonP) interactions)(Fig. 1C); numbers of both categories were strongly decreased in mut NSC, without changes in the relative proportion of P-P and P-nonP interactions (Fig. 1C,E). Similarly, the number of alternative usage interactions was greatly reduced in mut NSC (Fig. 1D, bottom row; two interactions with a shared left anchor in wt NSC can be replaced by one in mut). "COMMON" interactions were rare; note that many of these interactions, although detected in mut NSC, are represented by lower numbers of PET counts than in wt NSC, indicating a loss of these interactions in a proportion of the mut NSC (not shown). Interestingly, "MUT-specific" interactions show a much lower percentage of P-P, and a correspondingly higher percentage of non-P interactions, than observed in all other types of interactions (Fig.1E). This suggests that many interactions that are lost may re-engage in other, non-functional interactions. Examples of interactions maps across multiple unique nodes (Fig. 1F), and around single genes (Fig. 1G), in both wt and mut NSC, are shown. To make sure that the loss of (WT-specific) interactions observed in *Sox2*-mutant cells does not reflect a lack of RNA polymerase detection, we analyzed polymerase occupancy in both cell types, as detected by RNApolII ChIA-PET. Most (92%) of the regions marked by RNA polII are retained also in mut NSC, irrespective of whether or not they were connected (Fig. S1); further, the frequent appearance of "alternative" interactions between wt and mut NSC implies a real change of anchor connectivity, independent of RNA-polII detection. In conclusion, *Sox2* ablation causes the loss of a substantial subset of canonical interactions, in particular promoter-centered

interactions (both P-P and P-nonP), and leads to the generation of new interactions with high frequency of non-canonical types (non P-non P).

Long-range interaction anchors are enriched in SOX2 binding and in chromatin epigenetic marks of active/poised enhancers

Changes in long-range interactions following SOX2 loss may reflect a direct positive role of SOX2 in maintaining interactions, or a negative role in preventing the formation of interactions, or also indirect effects of SOX2 loss on other factors involved in creating/maintaining the interactions. To explore the relation of direct SOX2 binding to SOX2-regulated long-range interactions, we identified SOX2-bound sites through genome-wide ChIP-Seq of wt brain-derived NSC in culture. We also performed ChIP-Seq, in both wt and mut NSC, for histone modifications H3K27ac and H3K4me1, allowing the identification of active (H3K27ac+ and H3K4me1+), as well as "poised" (H3K4me1+) enhancers, with the potential to be activated (Cantone and Fisher, 2011; Creyghton et al., 2010; Rivera and Ren, 2013). For this analysis, we used both a "peak-calling" and a "segmentation" approach (chromHMM, see Methods), which led to qualitatively consistent conclusions between the two methods (Fig.2, Fig. S2). Finally, we linked SOX2-binding sites to both epigenetic marks and interacting anchors (Fig. 2). For a summary of data, see Table S2.

SOX2-bound sites were rarely located at promoters (i.e a region linked to a refseq annotated TSS), and more frequently in intronic (ca. 8000) and intergenic (ca. 8000) distal regions (Fig. 2A and data not shown). Over 90% of Sox2-bound sites were associated (in wt NSC) with nucleosomes characterized by the presence of either or both histone modifications (Fig.

2A). Virtually all Sox2-bound promoters (ca. 1600) were also H3K27ac+; as to SOX2-bound distal regions (ca.16000), most (ca. 85%) carried epigenetic enhancer marks (EM) (Fig. 2A, right). About half of these regions showed both modifications (H3K27ac and H3K4me1), whereas the other half showed H3K4me1 only (Fig. 2A). Thus, distal SOX2-bound regions were overwhelmingly associated with epigenetic enhancer marks (Fig. 2A). Of note, within distal H3K27ac+ peaks in general, a large proportion were SOX2-bound, whereas fewer H3K27ac+ promoters were SOX2-bound (Fig. 2B, Fig. S2).

SOX2 peaks were highly enriched within anchors; a high proportion (ca. 42 %) of interactions in wt NSC carried a SOX2-bound site within at least one of the two interacting anchors (Table S3). In turn, distal anchors were highly enriched for epigenetic EM. In particular, active enhancer signatures (H3K27ac+, H3K4me1+) were present (in wt NSC) on 90% of common distal non-P anchors, and 60-70% of WT-specific and WT-alternative anchors (Fig. 2C); over 90% of distal SOX2-bound anchors were associated with active enhancer marks within all different categories (Fig. 2C), pointing to potential functional role(s) of SOX2. MUT-specific distal anchors are much less frequently labeled by epigenetic EM, suggesting that many of these anchors may be devoid of transcriptional function (Fig. 2C).

Finally, distal H3K27ac+ regions that are SOX2-bound were more often engaged in interactions (i.e. overlapping with anchors) than SOX2-negative H3K27ac+ regions (Fig. 2D top); similar results were observed when considering epigenetically marked regions in general (Fig. 2D middle, bottom), further implying a role for SOX2 in long-range interactions.

We then asked whether the loss of interactions in mut NSC might lead to changes in patterns of histone modification: enrichment in H3K27ac,

H3K4me1, both, or none (see Methods). Overall, we observed only little loss or gain of the modification between wt and mut NSC (Fig. S2).

In conclusion, SOX2 is frequently bound to distal anchors, where it is highly associated with epigenetic EMs; presence of SOX2 in epigenetically marked regions (distal, and/or proximal) is highly significantly associated with the involvement of that region in interactions (as compared to Sox2-negative regions). This suggests that SOX2 binding may be relevant for the establishment, and/or maintenance, of the interactions. On the other hand, SOX2 ablation causes a widespread loss of long-range interactions, without greatly affecting the chromatin signatures surrounding the SOX2 binding sites.

SOX2-dependent long-range interactions predict novel forebrain enhancers

The strong enrichment of anchors in epigenetic EMs suggested that long-range interaction anchors might be used to identify regulatory elements driving gene expression in the developing brain. For example, the Nkx2.1 gene, a critical SOX2 target mediating its function in the developing ventral forebrain (Ferri et al., 2013), is connected by a WT-specific interaction to a characterized VISTA forebrain enhancer (Fig. 3A); VISTA forebrain enhancers are identified by a p300 ChIP-seq peak and by transgenesis (http://enhancer.lbl.gov)(Visel et al., 2009).

Overall, VISTA enhancers were enriched within interaction anchors, particularly SOX2-dependent WT-specific and WT-alternative anchors (Fig. S3). For example, a Sox2-dependent interaction connects the Sox4 gene to a SOX2 peak within a neural VISTA enhancer, located 650kb away, within an

intron of a liver gene (Fig. 3B). The Sox3 gene, whose mutation causes mental retardation and hypothalamic-pituitary defects (Laumonnier et al., 2002) interacts with a SOX2 peak in a telencephalic enhancer 350kb away (Fig. 3C). A SOX2-dependent interaction connects a SOX2-bound neural enhancer, within the Akt3 gene, to the Zbtb18 (Zfp238 in man) TF gene (Fig. 3D), whose mutation causes microcephaly in man and mouse (de Munnik et al., 2014); in man, deletions including Akt3, or balanced translocations separating Akt3 from Zfp238 (Boland et al., 2007)(Fig. 3D) also cause microcephaly, raising the possibility that disruption of the enhancer and its connections might contribute to the pathology.

Next, we tested distal regions involved in SOX2-dependent long-range interactions, using a transgenic enhancer assay in zebrafish. Eleven out of 13 reporter constructs containing SOX2-dependent and SOX2-binding distal anchors, directed GFP expression to the developing forebrain, (Fig. 4A; Table S4). GFP expression closely matched part of, or the whole, forebrain expression pattern of the endogenous zebrafish ortholog of the mouse gene connected to the analyzed enhancer (Fig. 4A-C). Similar data were obtained with anchors connected to important regulators of forebrain development, including *Sp8, Cxcr4, Sox3, COUP-TF1, Irx1* and *Socs3* (Fig. 4; Table S4).

We asked whether the activity of the transgenic enhancers in zebrafish was dependent on SOX2. We silenced endogenous Sox2 with morpholinos in stable transgenic lines for 7 enhancer-GFP constructs; GFP expression driven by 2 out of 7 constructs (including the Sp8- and Sox3-connected distal anchors) was downregulated (Fig. 4D and data not shown). Finally, SOX2 strongly activated, by transfection, constructs containing the anchors

connected to the Sox4 promoter and to the Akt1 gene, in synergy with the TF Ascl1/Mash1 (Fig. S4 and data not shown).

These data indicate that RNA-polII-mediated, SOX2-dependent interactions between genes and distal regions identify with high confidence novel forebrain enhancers.

Reduced gene expression in *Sox2*-mutant cells correlates with loss of long-range interactions

To correlate the expression levels of genes to the observed pattern of their long-range interactions, we analyzed by RNA-seq gene expression in wt and mut NSC. First, we partitioned genes in classes, ranked according to their expression levels (transcripts per million, TPM) in wt NSC and determined, for each class, the fraction of genes with an interaction (Fig. 5A). The higher the expression level of genes, the higher the percentage of genes whose promoter is involved in interactions (from 2-3% for non-expressed genes to about 60% for highly expressed genes (Fig. 5A). Interactions were both between promoters (P-P) and between promoters and distal non-promoter anchors (P-nonP; Fig. 5A). Higher expression levels were also associated with an increase in number of interactions per gene, particularly with P-nonP interactions, which show an almost two-fold increase of expression for every added “enhancer” (Fig. 5B and Fig. S5). Thus, more highly expressed genes are in general more connected than less expressed genes.

To determine how the presence of interactions is reflected into gene expression levels, we subdivided genes according to their connectivity (non-connected, connected with promoters or enhancers, or connected specifically with enhancers) and determined the distribution of gene

expression, in both wt and mut NSC (Fig. 5C). In both cell types, genes associated to nodes (i.e. involved in interactions, Fig. 1B) showed a distribution of expression levels shifted towards higher values than the overall population of expressed genes; the highest expression levels were seen with nodes involved specifically in interactions with enhancers (P-E) (Fig. 5C) (p-value < $2.2x10^{-16}$)

We next compared mut to wt NSC. In mut NSC, the distributions of expression levels of all genes, and of each category of genes, had lower median and mean values than in wt NSC, with an overall distribution significantly shifted towards lower values (Fig. 5C). Indeed, a Wilcoxon paired signed-rank test (see Methods) showed that the differences between wt and mut distributions of gene expression were highly significant (p-value < $2.2x10^{-16}$ for every pair considered). To further assess the significance of the observed expression decrease in mut NSC, we also plotted the distribution of the variation of the expression of each gene between wt and mut cells, defined as log-fold ratios (log2 [TPM_wild type/TPM_mutant])(Fig. 5D). To avoid bias from genes with low transcript levels, we considered only genes with TPM > 1. The plot was clearly shifted towards positive values, indicating that the majority of genes were more highly expressed in wt than in mut NSC (Fig. 5D). To ask whether the decrease of gene expression in mut NSC was related to any specific type of interaction, we partitioned the log2 fold ratio distribution into different gene categories, according to their type of connections (Fig. 5E, Fig. S5B). The bias towards positive fold ratios (log2 [TPM_wild type/TPM_mutant]) remained visible for all classes of genes, with or without interactions, but shifted more visibly away from zero, with increasing median values, when considering genes whose promoter node is connected to a distal non-P sequence in wt NSC (Fig. 5E, lanes 8-10 and Fig. S5B). To assess the

significance of these differences, we compared the distributions of the fold-ratios associated with all the interaction classes to the distribution of the "no interaction" class with a Mann-Whitney-Wilcoxon test (Fig. 5E). All three classes of genes whose promoter is associated, in wt NSC, with regions with enhancer marks ("enhancer" in general, WT-specific "enhancer", SOX2-bound "enhancer") (Fig. 5E, lanes 8-10) showed highly significant differences relative to the no interaction class (lane 1)($P<10^{-4}$), while classes of genes whose promoter is associated with any type of interaction (P-P or P-nonP; lane 6) were much less, although still, significant ($P<0.01$). Genes whose promoter is part of WT-specific nodes (lane 7) also yielded significant P-values ($P<0.01$), which contrast with the non-significant P-values of promoters linked to nodes common to both wt and mut NSC (lane 2), or linked to nodes present in mut NSC (lane 3).

These results indicate that loss of SOX2 is associated with an overall gene expression decrease, that is more relevant for genes involved in interactions with distal non-P sequences (in wt NSC) that are lost in mut NSC, or that contain a SOX2-bound site (in wt NSC)(Fig. 5E, lanes 8-10; see also Fig. S5B). To further validate these results, we identified three groups of genes (Fig. 5F): group 1: genes (TPM>5) showing a significant decrease of expression in mut versus wt NSC (677 genes); group 2: genes showing a moderate but not statistically significant decrease of expression in mutant (2194 genes); group 3: genes not showing any relevant expression change (7263 genes)(See Methods). Then we considered the different types of interactions associated with genes, to determine whether genes in each expression variation group could be significantly associated with any type of interaction (i.e., if their number, within a given interaction class, was higher, or lower, than the number expected by chance). We summarize the results by defining a "coassociation score" (see Methods) as the $-\log_{10} p$ of

the probability of observing a given number of genes associated with a given type of interaction in each of the three groups. We denoted a number lower than the expected value by multiplying the result by -1 (Fig. 5F, and Fig. S6; Table S5). The results confirmed a highly significant overlap between genes with significantly decreased expression in mut NSC and those characterized, in wt NSC, by a promoter to "enhancer" interaction that is lost in mut NSC (Fig. 5F, lanes 2,3 and 5), or by interactions involving SOX2-bound sites (Fig. 5F); note that the correlation holds true also for "COMMON" SOX2+ (but not Sox2-) interactions)(Fig. S6B, lanes 11-13). P-P interactions, instead, showed moderately significant overlap only with genes with moderate or no expression change (Fig. S6B, lanes 16,18). It has previously been proposed that some interacting promoters could influence each others' activities (Li et al., 2012). Our data, while not ruling out this model, suggest that overall, the numerous P-P interactions detected in NSC play a comparatively minor role relative to P-nonP interactions (Fig. 5F). This is consistent with our observation (Zhang et al., 2013) that P-non P interactions are more cell-type specific than P-P interactions, and with the cell-type specificity of Sox2 functions in neural cells.

Overexpression of Socs3, a multi-connected Sox2 target, rescues long-term self-renewal of mut NSC

Sox2-mutant NSC have a severe self-renewal defect, and their growth in culture becomes exhausted after 7-10 passages (approximately 30 days)(Favaro et al., 2009). To evaluate if any specific gene, among those whose connectivity and expression are affected by *Sox2* loss, was able to rescue long-term self-renewal in mut NSC we expressed in mut NSC the *Socs3* gene, an inhibitor of Jak/Stat signaling, which antagonizes precocious

differentiation of NSC into astroglia (Cao et al., 2006). *Socs3* is strongly down-regulated in mut NSC (Table S6), and shows both a SOX2 peak on the promoter, and multiple distal *Sox2*-dependent interactions (Fig. 6A), including one with a SOX2-bound anchor, that already tested positive in transgenic Zebrafish (Fig. 4). We transduced both wt and *Sox2*-mut NSC with a lentiviral *Socs3*-vector coexpressing GFP, performing three experiments, at virus to cell ratios transducing 20, 30 or 50% of the NSC. *Socs3*-transduced wt NSC grew at a similar rate as untransduced wt NSC, and continued to grow long-term, whereas untransduced mut NSC stopped growing between passages 8-12, as expected (Fig. 6B,C). In contrast, *Socs-3* transduced mut NSC continued to grow long term, even after the untransduced mut NSC had stopped growing. At the time of initial divergence of the growth curves of transduced and untransduced NSC, most or all neurospheres (from transduced mut NSC) contained GFP+ cells, and over 70% of the cells were positive by FACS, indicating a strong enrichment of the transduced cells; eventually, all cells became GFP+ (Fig. 6D). Note that *Socs3*- transduced wt NSC were not positively selected relative to untransduced wt NSC, indicating that, in wt NSC, the endogenous *Socs3* level was not limiting for optimal growth (not shown). This result identifies a SOX2-regulated gene, involved in SOX2-dependent interactions, whose abnormal regulation in mut NSC may be responsible for their defective long-term maintenance.

Discussion

SOX2 is necessary to maintain long-range interactions between regulatory elements in NSC

Here, we show that neural loss of Sox2, a TF necessary in NSC and in forebrain development (Favaro et al., 2009; Ferri et al., 2013; Ferri et al., 2004), causes a major decrease of the long-range interactions typical of wt NSC (Fig. 1), together with expression changes in a substantial proportion of genes. Decreased gene expression in mut NSC was significantly correlated with the loss of promoter-enhancer (P-nonP) interactions (but not to loss of P-P interactions), and/or to loss of SOX2 binding to interaction anchors in wt NSC (Fig. 5E,F). This demonstrates the relevance of a specific TF, SOX2, in shaping the pattern of chromatin connectivity genome-wide, and its reflection on gene activity. The striking impact on genome-wide interactions, of the loss of just a single TF was not expected.

The only precedent, to our knowledge, for a similar effect is the loss of *Oct4* or *Nanog* in pluripotent ES cells, which was shown to affect long-range interactions (de Wit et al., 2013). However, loss of these TFs by itself leads to downregulation of the core pluripotency TF network and initiation of abnormal differentiation, hence to drastic changes in the identity of the TF-deleted cells. Further, ablation of non-tissue-specific factors, such as CTCF or Cohesin, in ES or non-pluripotent cells causes some transcriptional dysregulation together with changes in long-range interactions (Dixon et al., 2016; Merkenschlager and Odom, 2013; Sofueva et al., 2013). However, the latter factors are thought to act primarily as “architectural proteins” (Phillips-Cremins et al., 2013) in regulating chromatin interactions.

What is the role of Sox2 in the observed changes of connectivity and gene expression? Distal interaction anchors are highly enriched in epigenetic EMs (Fig. 2), as well as in forebrain VISTA enhancers (Visel et al., 2009)(see Fig. 3, Fig. S3), or in enhancers identified in transgenic zebrafish (Fig. 4). In

addition, SOX2-bound sites are highly enriched in regions marked by epigenetic enhancer signatures, as well as in anchors (Fig. 2); moreover, SOX2-bound epigenetically defined enhancer regions are much more represented than non-SOX2-bound regions in anchors (Fig. 2). We hypothesize that SOX2, in addition to playing a conventional enhancer activator function (Fig. 4; Fig. S4), might act as a "bridging" factor, helping to bring together distant regulatory elements carrying other bound TFs, which may cooperate in transcriptional activation. SOX2 is an interactor of several TFs, as well as of proteins involved in determining chromatin structure, such as NurD complex, SWI/SNF, CHD7 and SMRT/NCOR (Engelen et al., 2011), which are thus possible candidates for mediating such a function. In addition, structural factors, such as Mediator, Cohesin and CTCF (Kagey et al., 2010; Merkenschlager and Odom, 2013; Phillips-Cremins et al., 2013; Sofueva et al., 2013), which are important determinants of DNA looping and gene expression, might be candidates for roles in establishing SOX2-dependent interactions. However, not all the interactions that are lost carry, within the involved anchors, SOX2-bound sites; it is therefore possible that, in mut NSC, secondary changes in the expression of TFs or chromatin factors, resulting from the primary SOX2 loss, cause the loss of these interactions.

While SOX2 plays an important role in the regulation of long-range interactions and gene activity, epigenetic EM marking enhancers are not lost genome wide in *Sox2*-mut NSC. *Sox2*, being importantly involved in forebrain development, might be expected to be relevant to the establishment of tissue-specific patterns of chromatin EM; it may be surprising, at first sight, that SOX2 ablation, and the ensuing loss of interactions, have little impact on chromatin EM (Fig. S2). A possible explanation is that SOX2 was ablated at a stage (E11.5) when specific EM

were already established within NSC. SOX2 might be initially important in determining the transition from an ectodermic cell to a NSC, and thus the establishment of the proper chromatin EM. However, our data show that SOX2 is not strictly required for the maintenance of the established chromatin EM. These observations dissociate the presence of an epigenetic EM from the actual engagement of the enhancer in an interaction with a promoter.

Not all genes that lose interactions in mut NSC are importantly down-regulated. There are several possible explanations: first, maintenance of expression of some genes may not require the actual maintenance of a long-range interaction, once the appropriate active chromatin organization has been initially established at the promoter via the long-range interaction with a distal enhancer. Second, many genes interact with multiple distal regulatory elements (Fig. 3, Fig. 6A, Fig. S7), which may act redundantly. Third, a given interaction may be relevant for gene expression only at specific developmental time points or in particular cell type contexts (see (de Laat and Duboule, 2013; Nord et al., 2015).

SOX2 loss ablates long-range interactions and downregulates key genes relevant to cell proliferation control

Sox2 is expressed in NSC throughout life (Zappone et al., 2000), and is essential for NSC maintenance in culture and, in vivo, in the hippocampus (Favaro et al., 2009). It is so far unknown which genes downstream to *Sox2* mediate its function in long-term NSC maintenance. Among the genes most significantly downregulated in *Sox2*-mutant cells is *Socs3*, a highly connected gene (Fig. 6; Table S6). By overexpressing *Socs3*, we rescued

long-term growth of *Sox2*-mutant NSC, providing evidence for a crucial role of *Socs3* in *Sox2*-dependent NSC maintenance (Fig. 6). Interestingly, several additional genes (*Fos, Jun, JunB, Btg2, Egr1 and Egr2*), encoding well-known regulators of cell proliferation, are expressed at high levels in wt NSC and are substantially downregulated in *Sox2*-mutant cells (Table S6); these genes show multiple promoter-enhancer interactions in wt NSC, that disappear in mut NSC (Fig. 6A, Table S6; Fig. S7). These findings may open the way to the identification, by similar studies, of a network of interacting genes, mediating *Sox2* functions in NSC. Hippocampal defects observed in *Sox2* mutant mice (as in humans) have been related to defects in NSC (Favaro et al 2009). The discovery of mediators of *Sox2* function in these cells may be relevant to the understanding of in vivo defects.

Many mouse homologs of genes affected in neurodevelopmental brain diseases are involved in long-range interactions with distant regions carrying enhancer marks

Our results indicate the importance of long-range interactions between promoters and enhancers for the activity in NSC of many genes, which are mostly not closely linked to their interacting anchor (and would thus not be detected by the conventional approach of ascribing a given enhancer to the nearest adjacent gene). Overall, in NSC, about 2000 expressed genes present such connections (Fig. 5B). Our study therefore identifies hundreds to thousands of potential regulatory elements, each connected to a specific gene active in neural cells. Thousands of polymorphisms in non-coding elements in man may be linked to brain disease or neurodevelopmental disorders (Doan et al., 2016; Nord et al., 2015); changes in TF- binding sequences may affect chromatin modifications locally and at distant sites,

affecting gene activity over great distances (Denker and de Laat, 2015). The comparison of the regulatory elements that we identified in mouse with conserved orthologous sequences in man may allow identification of genes regulated by such enhancers, and which might be dysfunctional in individuals carrying mutations at these elements.

Interestingly (Table 1), several genes, affected in important human neurodevelopmental disorders (Alcantara and O'Driscoll, 2014; Kim et al., 2012; Lee and Young, 2013; Peters et al., 2008; Saez et al., 2016; Tan et al., 2014; Williamson and FitzPatrick, 2014) are involved in long-range interactions, many of which are lost in *Sox2*-mutant cells; in particular, a large proportion of genes mutated in primary recessive microcephaly, severe intellectual disability and eye disease have mouse homologs involved in long-range interactions and/or showing a SOX2 peak in their promoter. Significantly, the pathology of *SOX2* mutant patients includes brain (mainly hippocampal) abnormalities, some degree of mental retardation, severe eye defects and seizures (Ragge et al., 2005; Sisodiya et al., 2006); microcephaly and some of the pathology observed in humans are also prominent in mouse Sox2 mutants (Favaro, 2009; Ferri 2004, 2013). Our data suggest that some of the genes showing connections in Table 1 might play a role in the human and mouse Sox2-dependent pathology.

Acknowledgements

Supporting grants: AIRC, Telethon, Regione Lombardia ASTIL, Fondazione Cariplo, ERANET to SKN; MINECO (BFU2013-43213-P), CIBERER, ISCIII, ERANET and an Institutional Grant from Fundación Ramon Areces to PB; the Francis Crick Institute, which receives its core funding from Cancer

Research UK (FC0010089), the UK Medical Research Council (FC0010089), and the Wellcome Trust (FC0010089), to FG and BM; the UK Biotechnology and Biological Sciences Research Council to F.G. (BB/K005316/1). JM was supported by short-term fellowships from EMBO and Fondazione Cariplo for two stays in Madrid.

Author contributions

C-L.W and S.K.N. designed the study. S.K.N., C-L.W., G.P., S.O., F.G., M.C., P.B. discussed the results and wrote the manuscript with input from all other authors. R.F. and M.P. prepared NSC for all chromatin studies. C-H. W, C.Y.N. performed and analyzed ChIA-PET, and did RNA-seq; B.M., M.V., did Sox2 and histone marks ChIP-Seq; P.R. and J.M did initial RNAseq work. G.P., F.Z. did bioinformatics/biostatitistics analyses. R.F., C.B., J.B., S.M. made DNA constructs and transfections. J.B., M.J.C., N.T. performed zebrafish studies, under the guidance of P.B. M.P. did the mut NSC rescue experiments.

FIGURES

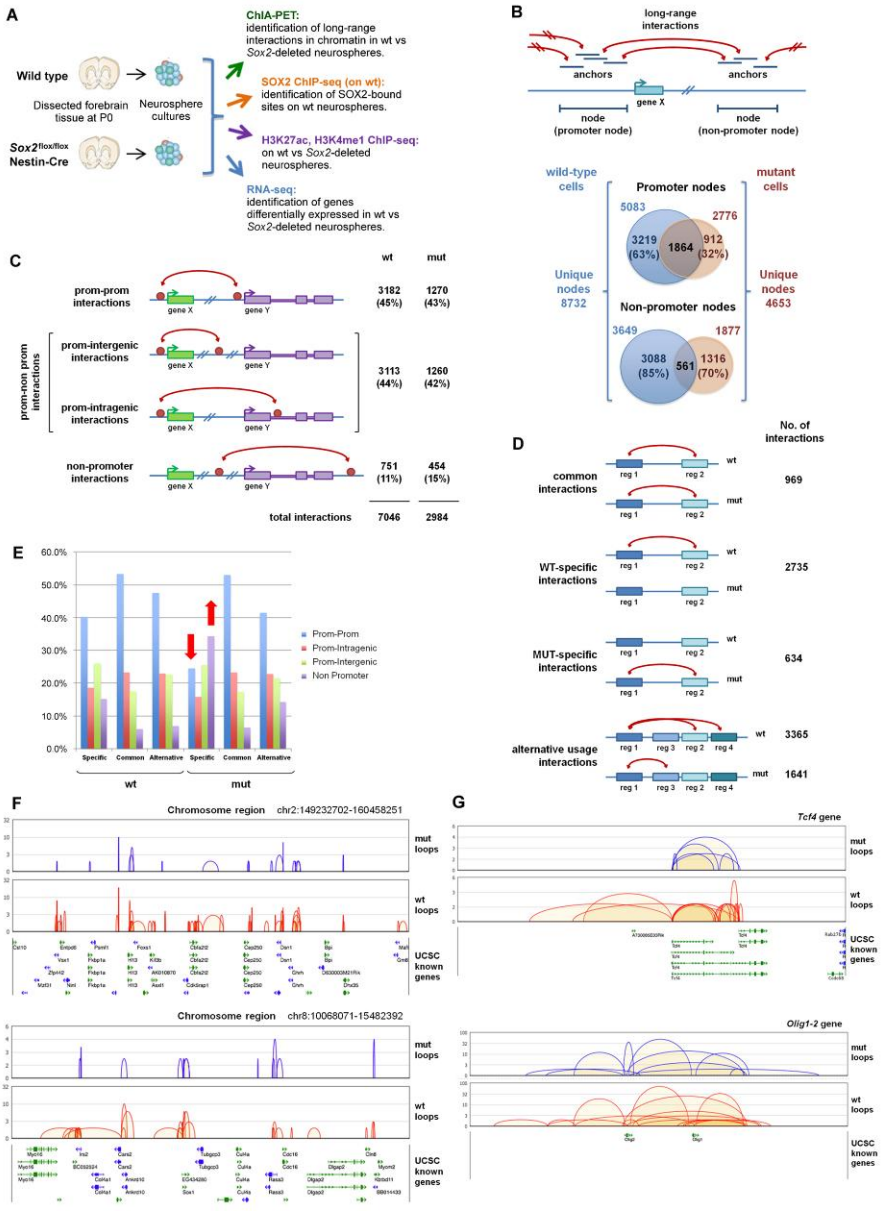


Figure 1: Sox2 ablation causes major loss of long-range interactions in brain-derived NSC.

(A) Functional genomics analyses. (B) Top: "anchors" and "nodes" connected by long-range interactions; bottom: numbers of promoter/non-promoter nodes in wt and mut NSC. (C) Interaction types, according to the nature of the connected regions. "Promoter": annotated TSS-containing region. (D) Interaction types, according to their presence, or not, in wt and mut NSC. (E) Frequency of the interaction types shown in panel C among specific, common, alternative interactions, in wt and mut NSC.(F,G) Connectivity diagrams in wt (orange) and mut (purple) cells, across chromosome regions (F) or specific loci (G). PET counts (Y axes); note different scales in some panels. In mut NSC, an overall decrease of "looping" is seen, but some interactions are lost, others are maintained.

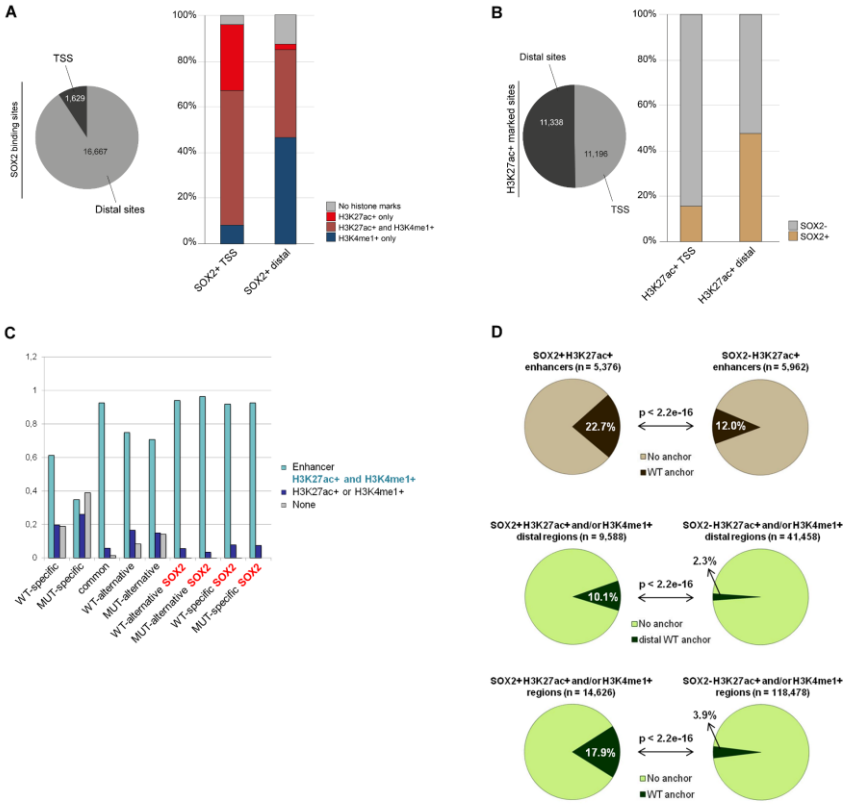


Figure 2: Sox2-bound regions carrying epigenetic enhancer marks (EM) show significantly higher overlap with anchors, than Sox2-negative EM-positive regions.

(A) Number of Sox2-bound sites in regions linked to annotated TSS (+/-1000nt) and to distal, non-P sites (left); fractions of different EM-positive regions within

Sox2-bound TSS-linked or distal regions (right). (B) Fraction of Sox2-bound sites within EM-positive regions (H3K27Ac+) on TSS-linked or distal anchors (peak calling). (C) Fraction of regions carrying EM on different categories of distal anchors (chromHMM). (D) Fraction of Sox2+ vs. Sox2- EM-positive regions which overlap with anchors. Top: distal (1000nt+/- from TSS) Ac+ regions, peak-calling. Middle: distal epigenetically marked regions: overlap with distal anchors. Bottom: all epigenetically marked regions, overlap with all anchor types, chromHMM.

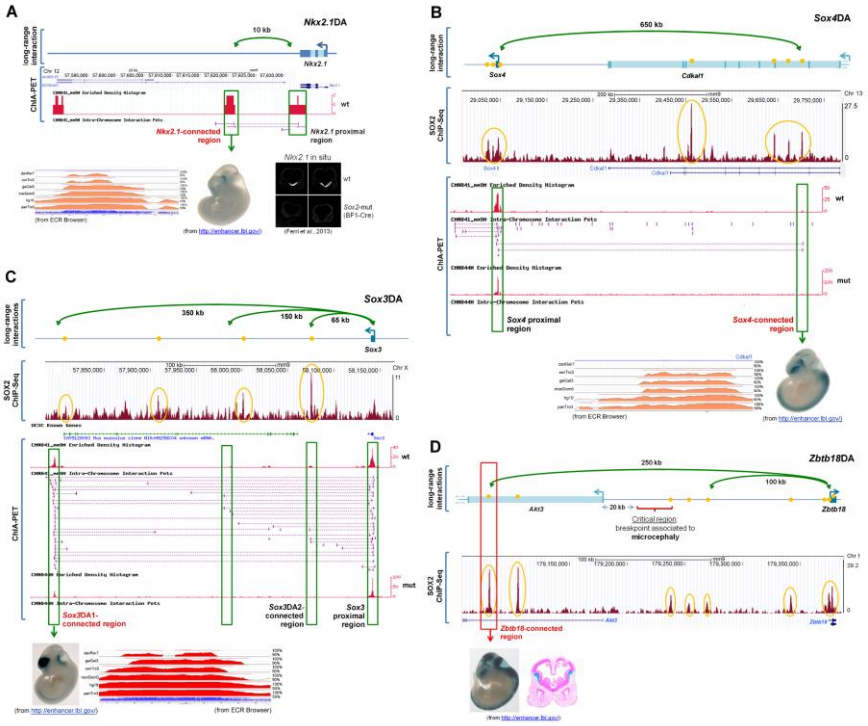


Figure 3: Distal anchors in Sox2-dependent interactions correspond to enhancers known to be active in the brain of transgenic mice.
(A,B,C,D) Sox2-dependent ChIA-PET interactions (green arrows) between four different genes (Nkx2.1, Sox4, Sox3, Zbtb18) and distal regions overlapping previously characterized enhancers [Visel et al, 2009]; SOX2 ChIPseq peaks (present paper), lacZ-stained transgenic embryos (from http://enhancer.lbl.gov), and evolutionary conservation (ECR browser). In A, loss of ventral Nkx2.1 expression in Sox2-mutant embryonic forebrain [Ferri et al, 2013].

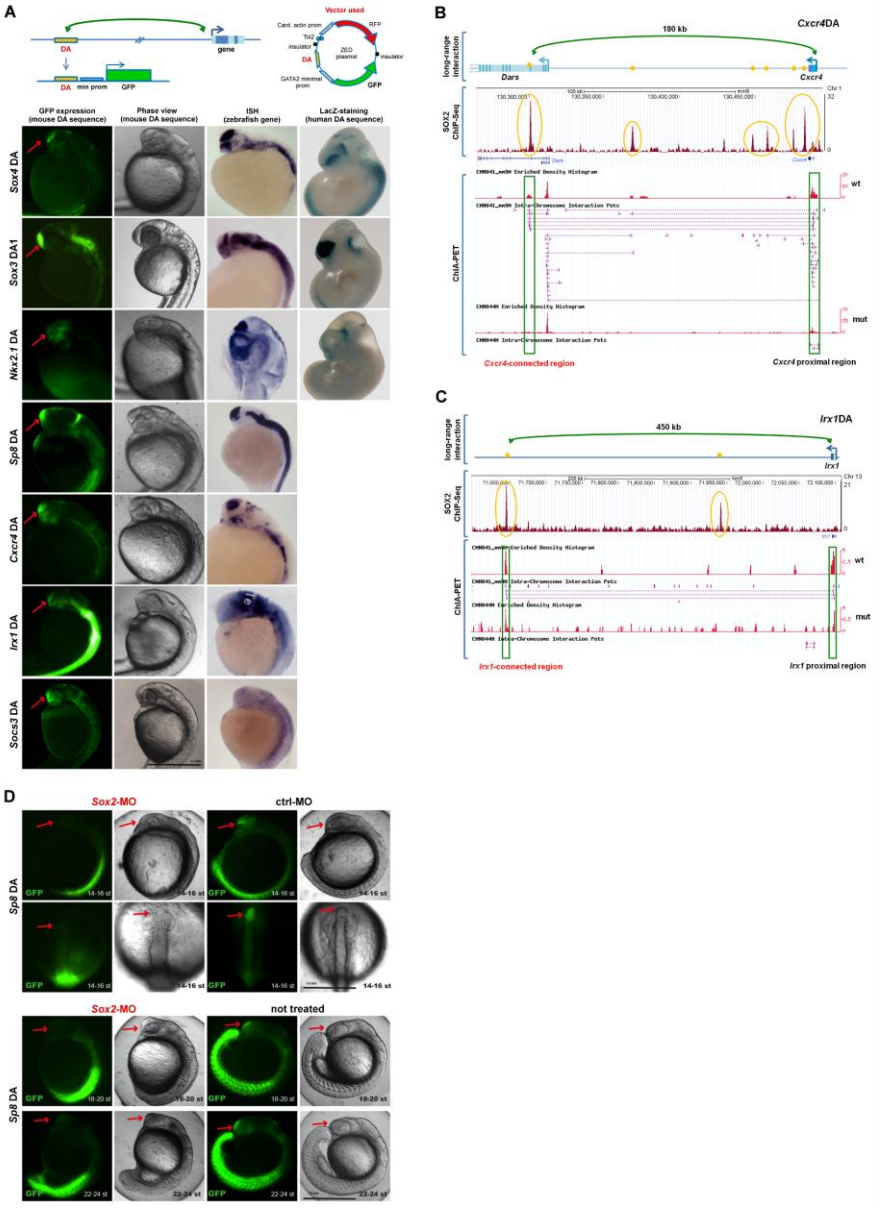


Figure 4: distal anchors in Sox2-dependent interactions drive GFP transgene activity to zebrafish brain.

(A) Top, enhancer-dependent GFP- reporter (ZED): distal anchors (DA) from Sox2-dependent interactions are cloned upstream to a minimal promoter and GFP. Bottom; first and second left columns: GFP expression in transgenic embryos and bright-field images. Third column: expression (in situ hybridization, from http://zfin.org) of the endogenous zebrafish gene corresponding to the gene

connected, in mouse, to the tested anchor. Fourth column: forebrain lacZ staining driven by transgenes carrying the human enhancers corresponding to the anchor (from http://enhancer.lbl.gov; Visel et al, 2009). (B,C) interactions, SOX2 peaks and ChIA-PET data for the Cxcr4 and Irx1 anchors. (D) treatment of embryos transgenic for Sp8 DA-ZED with anti-Sox2 morpholino (Sox2-MO, left) or control (ctrl-MO, right) at (two different stages). Reduced GFP signal is seen in Sox2-MO, but not ctrl-MO embryos, in forebrain (red arrow) but not in more posterior regions (internal control).

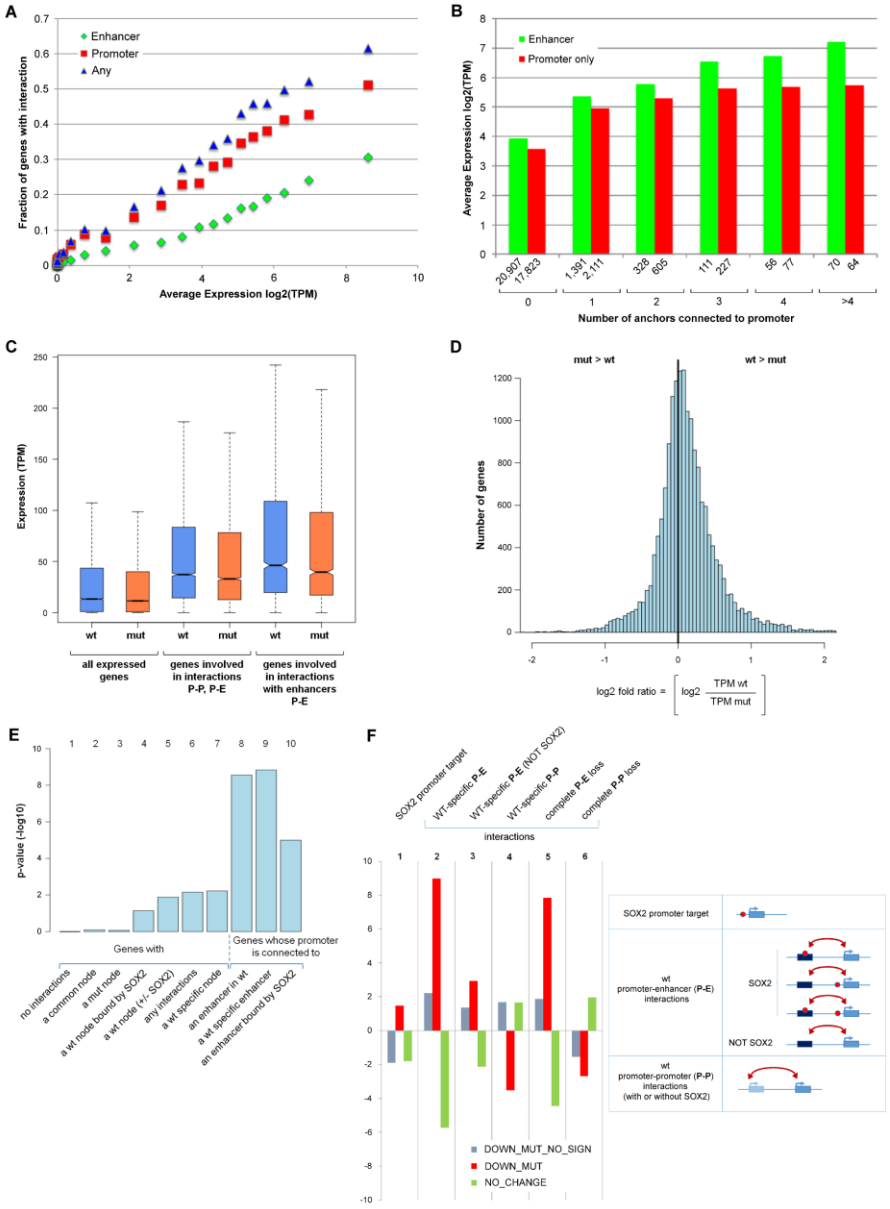


Figure 5: Reduced gene expression in Sox2-mutant NSC correlates with loss of long-range interactions.

(A) Fraction of genes whose promoter is involved in interactions (Y axis) within gene groups with increasing expression levels (X axis). Enhancer: interaction involving the gene promoter (TSS), and a distal non-P region. Promoter: interaction involving the gene promoter and another promoter. (B) Average gene expression (Y axis) of genes according to the number of interactions in which the

gene promoter is involved (X axis). Numbers of genes involved in interactions with enhancers (green), or distal promoters only (red), are shown (X axis). (C) Distribution of expression values (TPM) of genes with TPM > 0. wt NSC: orange; mut NSC: blue. From left to right: all genes; genes whose promoter is a node (P-P, P-E interactions); genes whose promoter is connected to an enhancer (P-E interactions). (D) Distribution of the fold ratio values for all genes with TPM > 0 defined as log2(TPM_WT/TPM_MUT). It confirms results shown in (C): the fold ratio is shifted towards positive values, i.e. a majority of genes have expression in wt higher than in mut NSC. (E): Comparison of fold ratio distributions (from panel D, see also Fig. S5) for different categories of interactions, with the "no category" fold ratio distribution. p-values (as –log10) by the Mann–Whitney U test comparing each interaction category to the "no category" one. Asterisks denote significant comparisons. (F) Coassociation scores (histograms) between genes (TPM>5) showing either reduced gene expression in mut NSC, or no change, and the indicated categories of interactions (see Methods). DOWN_MUT, DOWN_MUT NO_SIGN: statistically significant, or moderate, reduction, respectively; NO_CHANGE: no relevant change.

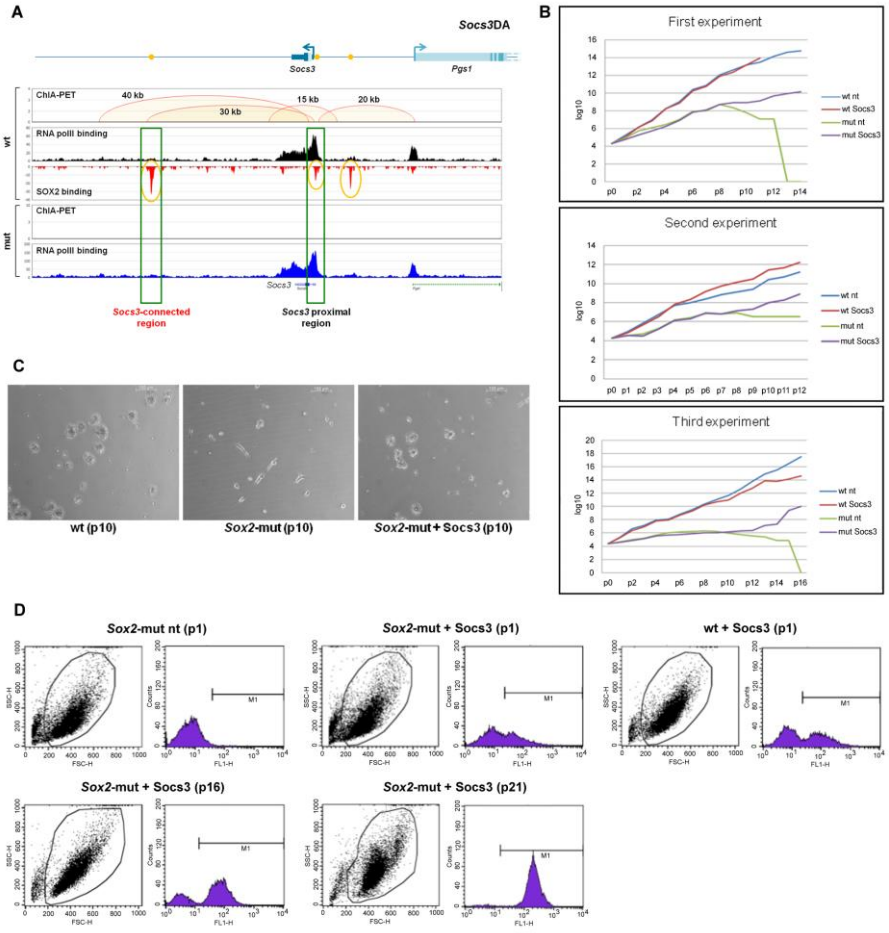


Figure 6: SOCS3 re-expression in mut NSC prevents NSC exhaustion and restores self-renewal.

(A) Top, Socs3 gene. ChIA-PET interactions, SOX2 peaks, and ChIA-PET reads in wt NSC; bottom, loss of interactions in Sox2-mut NSC. (B) growth curves of mut NSC, transduced with a Socs3-GFP-expressing lentivirus (mut socs3) or not (mut), and of wt controls (wt, wt socs3). (C) Images of mut NSC, or Socs3-transduced mut NSC, three days after passage 12; neurospheres develop only from Socs3-transduced NSC. (D) FACS analysis (GFP) of mut, wt and mut Socs3 cells, at the indicated passage number. With passaging, the fraction of GFP+ NSC progressively increases in mut Socs3 NSC, eventually reaching 100%.

Category of disease and gene name in mouse	Disease gene promoter				
	interaction with				SOX2-bound
	distal enhancer			SOX2-bound enh.	
	wt-sp	wt-alt	com		
Microcephaly associated to defects in					
centrosome and spindle microtubule (1)					
positive: 10/22; p-value: 0.006					
Cdk5rap2 (Mcph3)					●
Casc5					●
Cenpj					●
Stil					●
Cep63		●		●	
Kif2a		●			
Kif11	●				
Tubb2b		●	●	●	●
Tuba1a		●	●	●	●
Poc1a					●
origin recognition complex core (1)					
positive: 2/5					
Orc4					●
Cdt1					●
DNA damage response and repair (1)					
positive: 5/19					
Lig4					●
Phc1 (Mcph11)		●		●	●
Xrcc2					●
Xrcc4					●
Blm (Recql3)					●
Other microcephalies					
Gpr56		●	●	●	●
Cdk19					●
Arx			●		
Zbtb18					●
Angelman and Angelman-like syndromes (2)					
(intellectual disability and absent speech)					
positive: 10/12; p-value: 0.0000065					
Ube3a					●
Tcf4	●	●	●	●	●
Ehmt1					●
Herc2		●		●	
Adsl					●
Cdkl5					●
MeCP2					●
Foxg1		●		●	
Atrx					●
Zeb2	●	●		●	●

	Disease gene promoter				
Category of disease and gene name in mouse	**interaction with**				**SOX2-bound**
	distal enhancer			**SOX2-bound enh.**	
	wt-sp	**wt-alt**	**com**		
Histone modification, chromatin remodelling					
and mediator mutations (3,4,5)					
(intellectual disability)					
positive: 8/12; p-value: 0.0007					
Med17					●
Med23					●
Med25					●
Smarca2					●
Arid1a		●	●	●	
Arid1b		●	●		
Jmjd1c		●	●	●	
Phf21a					●
Cohesin subunit mutations (6)					
(psicomotor delay, intellectual disability)					
positive: 4/14					
Smc3					●
Rad21 (Scc1)					●
Stag1		●		●	
Stag2		●			
Microphthalmia / Anopththalmia /					
Coloboma and other eye pathologies (7)					
Otx2		●			●
Pax6					●
Six3					●
Bmp7	●			●	
Grcc10 (C12orf57)		●			●
Sall2					●
RarB	●				●
Smoc1	●			●	
Wdr19					●
COUP-TF1 (Nr2f1)	●			●	
Abhd12	●			●	

References:

(1) Alcantara and O'Driscoll, 2014

(2) Tan et al., 2014

(3) Lee and Young, 2013

(4) Saez et al., 2016

(5) Kim et al., 2012

(6) Peters et al., 2008

(7) Williamson and Fitzpatrick, 2014

Table 1: Interactions and SOX2 peaks in mouse homologs of genes involved in inherited neural disease in man.
We list genes related to the main categories of neurodevelopmental disorders in man that are either present in SOX2-mutant patients (hippocampal, eye defects; intellectual disability, seizures) and/or in Sox2-mutant mice (microcephaly, hippocampal defects, seizures, eye defects). Disease genes are analysed for interactions of the respective mouse gene promoter with distal enhancers, or presence of SOX2. p-values are given, when significant, for the association between the above characteristics and type of disease

Methods

Neural stem/progenitor cell (NSC) cultures

P0 brain-derived NSC cultures were obtained from dissected telencephalon of wild type and Sox2-deleted mice, and grown, as described in (Favaro et al., 2009) and (Zhang et al., 2013). Cells were grown in parallel for two passages in 10 ng/ml Epidermal Growth Factor (EGF) and 10 ng/ml basic Fibroblast Growth Factor (bFGF), followed by cell dissociation and expansion in the presence of EGF, but not bFGF, for 3-4 more passages, as described in Zhang et al. (2013), for optimal maintenance of mutant NSC (Favaro et al., 2009). For ChIA-PET analysis, pooled NSC from four wild type and six mutant brains were used; for Sox2 and histone modifications studies, four wild type and four mutant brains were used for each analysis. Preparation of NSC for ChIA-PET analysis was as in Zhang et al., 2013.

Socs3 transduction in NSC

At day 0 wild type and Sox2-Mut neurospheres were dissociated to single cells and seeded at a density of 25000 cells/ 1ml/ well (9 wells for each cell type) in 24 well plates, in DMEM-F12 with Glutamax (Gibco) containing EGF only as mitogen. After 4 hours wt and mut NSC were transduced with a

TWEEN lentiviral vector expressing SOCS3 from the CMV promoter and GFP driven by the human PGK promoter (Francipane et al., 2009), at a multiplicity of infection (MOI) of 3.5-5.5 and incubated overnight at 37°C. Then 1ml per well of fresh medium was added both to transduced and non-transduced (control) cells. After 4 days, cells were dissociated to single cells, counted, and seeded at a density of 20000 cell/well in the same EGF medium described above. In addition, 500000 cells for each sample (from pooled wells) were fixed using PFA 4% and washed in PBS in order to analyze the GFP fluorescence by flow cytometry (BD FACSCalibur™): 10,000 events were analyzed for each sample. The samples were excited at 488 nm (blue laser) and the resulting fluorescence measured at wavelengths >530 nm. The results were analyzed using CellQuest Pro software (BD Biosciences).

ChIA-PET analyses

The ChIA-PET experiments and data processing were as in Zhang et al., 2013.

ChIPseq analyses

H3K27Ac and H3K4me1 ChIPseq

NSC from wild type and Sox2-deleted brains (Favaro et al., 2009) (two populations each) were used for ChIP-sequencing on H3K27ac and H3K4me1 as described previously (Vermunt et al., 2016). For each line, neurospheres were dissociated and 4 million single cells were crosslinked with 1% formaldehyde for 10 minutes at room temperature. Reaction was

quenched with 0.125 M Glycine, cells were washed with cold PBS and lysed according to Vermunt et al., 2016. Nuclei pellets were resuspended in 160µl sonication buffer and divided over two microtubes for shearing in the Covaris S series with the following settings for 12 cycles of 60 seconds: intensity 3, duty cycle 20%, 200 cycles/bursts. Chromatin immunoprecipitation steps after sonication were performed as described previously (Vermunt et al., 2016) using 50 µl DynaI protein G beads that were preincubated with 5 µg Ab4729 (Abcam) for H3K27ac or 5 µg Ab8894 (Abcam) for H3K4me1. Whole cell extract of 4 million cells was split onto both antibodies, resulting in the use of 2 million cells per ChIP. Libraries were made using the Illumina Truseq DNA library protocol and sequencing was done at the MIT BioMicro Center (http://openwetware.org/wiki/BioMicroCenter). Obtained sequences were aligned onto the mm9 mouse genome assembly using Bowtie 1.1.0 (bowtie-bio.sourceforge.net) excluding reads that had more than 1 mismatch or that could map to multiple genomic locations. MACS2 was used for peak calling (p-value threshold = 10-5, extsize = 400, local lambda = 100,000) and narrowpeaks were extended to a minimum of 2000 basepairs (bps) to match peak resolution. Overlapping enriched regions were merged and were considered promoters when located within 1000 bps from annotated mm9 transcriptional start sites (TSSs) and considered putative distal enhancers when located more than 1000 bps away from TSSs.

Sox2 ChIPseq

NSC from wild type and Sox2-deleted brains (Favaro et al., 2009) (two populations each) were fixed sequentially with di(N-succimidyl) glutarate and 1% formaldehyde in phosphate-buffered saline and then lysed,

sonicated and immunoprecipitated as described previously (Mateo et al., 2015 and references therein). SOX2 was immunoprecipitated with 3mg of goat anti-SOX2 (Santa Cruz sc-17320).

DNA libraries were prepared from 10ng of immunoprecipitated DNA and 10ng of input DNA control, according to the standard Illumina ChIP-seq protocol. Libraries were sequenced with the Genome Analyzer IIx (Illumina). The raw reads were mapped to the mouse genome (mm9 including random chromosomes) with Bowtie (Langmead, 2010) version 0.12.5. We used MACS (Zhang et al., 2008) version 2.0.9 to define Sox2 bound regions (peaks). As this tool is very sensitive to the unbalanced number of reads in the real and the input set, we decided to reduce the larger input dataset to match the number of mapped reads in the smaller IP dataset by randomly downsampling reads, as described previously (Mateo et al., 2015).

Analysis of histone modifications colocalization

Co-localization of histone modifications was performed with ChromHMM version 1.4. (Ernst and Kellis, 2012). Briefly, the software partitions the genome into non overlapping segments of 200 bps. Then, given a set of histone modification ChIP-Seq experiments, associates to each segment each of the histone modifications if the number of reads mapping in the segment can be considered to be enriched according to a random background Poisson distribution. Then, given a number of states as input, it evaluates the co-occurrence of histone modifications in the genome segments, building a model in which each of the states is characterized by a given combination of modifications.

The program was run setting a different number of states, and by processing either wild type (WT) samples alone and mutant (MUT) samples alone, and on both WT and MUT samples combined. In every setting, the model recovered consistently four main states, corresponding to the joint presence of H3K27ac and H3K4me1, either modification alone, and neither modification. More importantly, all the analyses run on the combined WT and MUT samples failed to identify "differential" states in which one of the two modifications was present only in WT or MUT samples. That is, the model built, regardless of the number of states given as input, consistently contained four more states corresponding to 1) the presence of both H3K27ac and H3K4me1 in both WT and MUT samples; 2) to H3K27ac in both WT and MUT samples; 3) to H3K4me1 in both WT and MUT samples; 4) to neither modification in WT or MUT samples.

Differentially enriched 200bp samples were identified with an approach similar to the one of ChromHMM, by comparing for each modification the number of mapped reads in the segments in the two WT samples to the number of reads of the two MUT samples. Given a region with *n* reads in WT and *m* reads for MUT, the we compute the probability of finding *n* and *m* reads by chance, given N mapped reads for WT and M for MUT, with a Chi-Square test. We considered "differentially enriched" all regions with the resulting p-value lower than 10^{-4}.

Accession numbers

The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE90561.

Zebrafish transgenesis

Sequences from 15 of the identified anchors (13 distal, and 2 proximal) were amplified from the mouse genome using specific primers (listed below) and cloned into a pBluescript vector. Individual fragments were then transferred, using Gateway technology (Invitrogen), to the ZED (Zebrafish Enhancer Detection) vector (Bessa et al., 2009). ZED contains a cardiac actin promoter-driven RFP gene, used as an internal transgenesis control, and a minimal promoter linked to the putative enhancer being tested, and driving GFP expression. Plasmid DNA was purified using the Genopure plasmid Midi kit (Roche) following manufacturer instructions. Zebrafish embryos were microinjected at one-cell stage with 3–5 nl of a solution containing 25 nM of each of the construct to be tested and 25 ng/ml of Tol2 RNA. Putative transgenic embryos, as determined by the expression of cardiac-actin:RFP, were screened for tissue-specific enhancer activity by looking for EGFP expression in the brain at 15, 18, 24 and 48 hpf stages. Fluorescent images were acquired with a black-and white highly-sensitive camera (Leica DC350FX) and converted into colour images with the associated Leica acquisition program. EGFP distribution was compared with the expression pattern of the putative regulated genes, as determined by *in situ* hybridization analysis (from http://zfin.org). EGFP-positive (in forebrain) embryos were collected and propagated to generate three independent F1 transgenic lines each by crossing with wild type animals. To confirm a link between Sox2 and the identified elements, embryos derived from the F1 lines were microinjected with a *sox2* specific morpholino (GeneTools; 50-500 nM), previously reported to efficiently interfere with Sox2 expression in zebrafish (Okuda et al., 2010). Embryos were then analysed for increased or reduced reporter expression.

Gene associated with proximal anchor	Cloned anchor	Forward primer (5'->3')	Reverse primer (5'->3')
Nkx2.1	DA504	TCCCGTTCCTAGTCTTTGATACTT	CGAGCAACAGGAGAGGAATAATTT
Sp8	PA545	GACATCCAGACCTCTGTTTTCC	TCTACCAGAGGTGGGATTCAAG
Sp8	DA545	GGGGAAGAGTTCCTAGCCATT	GTGGGAGCTCAATTCATCTAA
Coup-TFI	DA2467	GCTCCAGCGTCTACTGAGAAAT	AGCAGAATCCCTGAGACTTCAC
Coup-TFI	DA231	CCAGTGAAACACCTACTCACCA	AAGTTGGCATTTTTAGGACTCG
Ntng1	DA1414	GTAGAGGCGCGGAACCATAG	GGGGTAAAAGGAAAGGGCAAA
Irx1	DA597	CAGCAAAGCATTGTAACTGTGA	TGGGGCTTTAACACAAGCAT
Socs3	DA463	GCTCACACTGACCCATAGGTTT	TTGCCTCTCAGAGTGAACCA
Chd7	DA1439	AGGCAAGCTCACCAGCTCT	GATTTCAAAGGCAGCCACAT
Sox3	DA2733	GGAGGCACATGAAAGCAATAA	GGGTAAGGTTAAAATGGCTTTTG
Sox3	DA2702	ACTGTCCATTTAGTTTTCATAAATCA	GTGGGCAGGGATACCTTAGTCT
Sox4	DA2458	GTCCTTCAGCAAGCTCTAAACA	AATGGTGGTGAAATCTGCAAGT
Cxcr4	DA62	GGACCCCTCAGTGAATATTAAGG	TTTGCACTGTGGTACACATTTT
Zfp355	DA1303	CTAGCACTCAACCCTGAGATT	GACTTCAGAATGGAGCCAGAAC
Zfp355	PA1303	AAGTAGTTCCGGTTCGAG	TTTGAGGCTTTCACTCTGCTG

RNA-seq analysis

RNA extraction was performed on three independent NSC populations for both wild type and mutant cells, using Trizol and RNeasy Kit (Qiagen). 1/5 volume of chloroform was added to one volume of Trizol. Aqueuos phase was transferred into a new tube. 1.5x volume of ethanol was added and mixed well. Mixture was was filtered through Rneasy (Qiagen) column. Column was washed with Buffer RW1. On-column DNase treatment was performed as described by the RNase-Free DNase Kit (Qiagen). Post treatment column was clean up with Buffer RW1 and two washes of RPE buffer. Column was then dried and total RNA was eluted with RNase-free water. PolyA Stranded Truseq Libraries were generated using the Truseq Stranded mRNA Sample Preparation Kit (Illumina). First, mRNA was purified from 1µg of total RNA using magnetic beads containing poly-T oligos. mRNA was then fragmented and reversed transcribed using Superscript II (Invitrogen), followed by second strand synthesis. Double stranded cDNA was treated with end-pair, A-tailing, adapter ligation and 8 cycles of PCR amplification.

RNA-Seq was performed on triplicates for the two genotypes studied, yielding 51 bp single-end reads. The number of sequences obtained in each sample ranged from 7.5 to 12.5 millions.

Read counts and transcript levels for each sample were computed with the RSEM software package version 1.17 (Li and Dewey, 2011), on the RefSeq gene annotation available at the UCSC Genome Browser for mouse genome assembly mm9 (24,148 genes). For downstream analyses, expression levels measured as Transcripts per Million (TPM) were employed.

Expression box plots and subsequent tests were generated using the R functions “boxplot” and “Wilcox.test”. Differential expression analysis was performed on the TPM values with the Noiseq package (Tarazona et al., 2015)(and refs. therein) using the NoiseqBIO method that handles replicate experiments. In this analysis, we considered to be having a significant variation of expression (Group 1) those genes with both 1) a fold ratio of the average transcript level in the two conditions greater than 1.5 and 2) an associated false discovery rate lower than 0.05 (corresponding to a Noiseq q-value greater than 0.95). As control for the analysis on co-associations between interacting anchors and differential expression, we also defined as having a “moderate” change of expression (Group 2) genes that did not satisfy either of the two previous conditions but had a FDR < 0.2 (q-value > 0.8), and “not changing” (Group 3) all the remaining genes with FDR > 0.2.

Co-association scores

Co-association scores are based on the significance of the overlap of two sets of genes, in this case one set showing a significant change of expression in the RNA-Seq data and a second one of genes whose promoter associated with a given type of interaction. Given two sets of genes of size *n* and *m* out of *N* annotated genes, and *k* genes in common to the two sets, the probability of having *k* genes in common by chance can be estimated by a Fisher’s exact test with parameters *k* (number of successes), *n* (size of the sample), *m* (number of successes on the population), *N* (size of the population). The test was applied by computing the number *k* of genes involved in any type of interaction in each of the three expression variation groups. Co-association scores were computed starting from the p-value *p* resulting from the test, defined as $-\log_{10} p$ if *k* was greater than the

expected value (hence showing a co-association between the two gene classes greater than what expected by chance), $\log_{10} p$ otherwise (hence showing a negative co-association). Thus, the higher positive co-association scores are, the more significant is the overlap between the two categories considered, with co-association scores greater than 2 showing a statistically significant overlap. Vice versa for negative scores, showing a significantly low overlap between two sets if lower than 2.

REFERENCES

- Alcantara, D., and O'Driscoll, M. (2014). Congenital microcephaly. Am J Med Genet C Semin Med Genet 166C, 124-139.

- 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.

- Bessa, J., Tena, J.J., de la Calle-Mustienes, E., Fernandez-Minan, A., Naranjo, S., Fernandez, A., Montoliu, L., Akalin, A., Lenhard, B., Casares, F., et al. (2009). Zebrafish enhancer detection (ZED) vector: a new tool to facilitate transgenesis and the functional analysis of cis-regulatory regions in zebrafish. Developmental dynamics : an official publication of the American Association of Anatomists 238, 2409-2417.

- Boland, E., Clayton-Smith, J., Woo, V.G., McKee, S., Manson, F.D., Medne, L., Zackai, E., Swanson, E.A., Fitzpatrick, D., Millen, K.J., et al. (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.

- Cantone, I., and Fisher, A.G. (2011). Unraveling epigenetic landscapes: the enigma of enhancers. Cell Stem Cell 8, 128-129.

- Cao, F., Hata, R., Zhu, P., Ma, Y.J., Tanaka, J., Hanakawa, Y., Hashimoto, K., Niinobe, M., Yoshikawa, K., and Sakanaka, M. (2006). Overexpression of SOCS3 inhibits astrogliogenesis and promotes maintenance of neural stem cells. Journal of neurochemistry 98, 459-470.

- Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A 107, 21931-21936.

- de Laat, W., and Duboule, D. (2013). Topology of mammalian developmental enhancers and their regulatory landscapes. Nature 502, 499-506.

- de Munnik, S.A., Garcia-Minaur, S., Hoischen, A., van Bon, B.W., Boycott, K.M., Schoots, J., Hoefsloot, L.H., Knoers, N.V., Bongers, E.M., and Brunner, H.G. (2014). A de novo non-sense mutation in ZBTB18 in a patient with features of the 1q43q44 microdeletion syndrome. European journal of human genetics : EJHG 22, 844-846.

- de Wit, E., Bouwman, B.A., Zhu, Y., Klous, P., Splinter, E., Verstegen, M.J., Krijger, P.H., Festuccia, N., Nora, E.P., Welling, M., et al.

(2013). The pluripotent genome in three dimensions is shaped around pluripotency factors. Nature 501, 227-231.

- Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P.D., Dean, A., and Blobel, G.A. (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. Cell 149, 1233-1244.

- Denker, A., and de Laat, W. (2015). A Long-Distance Chromatin Affair. Cell 162, 942-943.

- Dixon, J.R., Gorkin, D.U., and Ren, B. (2016). Chromatin Domains: The Unit of Chromosome Organization. Mol Cell 62, 668-680.

- Doan, R.N., Bae, B.I., Cubelos, B., Chang, C., Hossain, A.A., Al-Saad, S., Mukaddes, N.M., Oner, O., Al-Saffar, M., Balkhy, S., et al. (2016). Mutations in Human Accelerated Regions Disrupt Cognition and Social Behavior. Cell 167, 341-354 e312.

- 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.

- Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. Nature methods 9, 215-216.

- 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., Favaro, R., Beccari, L., Bertolini, J., Mercurio, S., Nieto-Lopez, F., Verzeroli, C., La Regina, F., De Pietri Tonelli, D., Ottolenghi, S., et al. (2013). Sox2 is required for embryonic development of the ventral telencephalon through the activation of the ventral determinants Nkx2.1 and Shh. Development 140, 1250-1261.

- 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.

- Francipane, M.G., Eterno, V., Spina, V., Bini, M., Scerrino, G., Buscemi, G., Gulotta, G., Todaro, M., Dieli, F., De Maria, R., et al. (2009). Suppressor of cytokine signaling 3 sensitizes anaplastic thyroid cancer to standard chemotherapy. Cancer Res 69, 6141-6148.

- Gorkin, D.U., Leung, D., and Ren, B. (2014). The 3D genome in transcriptional regulation and pluripotency. Cell Stem Cell 14, 762-775.

- Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature 467, 430-435.

- Kim, H.G., Kim, H.T., Leach, N.T., Lan, F., Ullmann, R., Silahtaroglu, A., Kurth, I., Nowka, A., Seong, I.S., Shen, Y., et al. (2012). Translocations disrupting PHF21A in the Potocki-Shaffer-syndrome region are associated with intellectual disability and craniofacial anomalies. Am J Hum Genet 91, 56-72.

- Langmead, B. (2010). Aligning short sequencing reads with Bowtie. Current protocols in bioinformatics Chapter 11, Unit 11 17.

- Laumonnier, F., Ronce, N., Hamel, B.C., Thomas, P., Lespinasse, J., Raynaud, M., Paringaux, C., Van Bokhoven, H., Kalscheuer, V., Fryns, J.P., et al. (2002). Transcription factor SOX3 is involved in X-linked mental retardation with growth hormone deficiency. Am J Hum Genet 71, 1450-1455.

- Lee, T.I., and Young, R.A. (2013). Transcriptional regulation and its misregulation in disease. Cell 152, 1237-1251.

- Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC bioinformatics 12, 323.

- Li, G., Ruan, X., Auerbach, R.K., Sandhu, K.S., Zheng, M., Wang, P., Poh, H.M., Goh, Y., Lim, J., Zhang, J., et al. (2012). Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell 148, 84-98.

- Mateo, J.L., van den Berg, D.L., Haeussler, M., Drechsel, D., Gaber, Z.B., Castro, D.S., Robson, P., Crawford, G.E., Flicek, P., Ettwiller, L., et al. (2015). Characterization of the neural stem cell gene

regulatory network identifies OLIG2 as a multifunctional regulator of self-renewal. Genome Res 25, 41-56.

- Merkenschlager, M., and Odom, D.T. (2013). CTCF and cohesin: linking gene regulatory elements with their targets. Cell 152, 1285-1297.

- Noordermeer, D., de Wit, E., Klous, P., van de Werken, H., Simonis, M., Lopez-Jones, M., Eussen, B., de Klein, A., Singer, R.H., and de Laat, W. (2011). Variegated gene expression caused by cell-specific long-range DNA interactions. Nature cell biology 13, 944-951.

- Nord, A.S., Pattabiraman, K., Visel, A., and Rubenstein, J.L. (2015). Genomic perspectives of transcriptional regulation in forebrain development. Neuron 85, 27-47.

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

- Peters, J.M., Tedeschi, A., and Schmitz, J. (2008). The cohesin complex and its roles in chromosome biology. Genes Dev 22, 3089-3114.

- Phillips-Cremins, J.E., Sauria, M.E., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S., Ong, C.T., Hookway, T.A., Guo, C., Sun, Y., et al. (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell 153, 1281-1295.

- Ragge, N.K., Lorenz, B., Schneider, A., Bushby, K., de Sanctis, L., de Sanctis, U., Salt, A., Collin, J.R., Vivian, A.J., Free, S.L., et al. (2005).

SOX2 anophthalmia syndrome. Am J Med Genet A 135, 1-7; discussion 8.

- Rivera, C.M., and Ren, B. (2013). Mapping human epigenomes. Cell 155, 39-55.

- Saez, M.A., Fernandez-Rodriguez, J., Moutinho, C., Sanchez-Mut, J.V., Gomez, A., Vidal, E., Petazzi, P., Szczesna, K., Lopez-Serra, P., Lucariello, M., et al. (2016). Mutations in JMJD1C are involved in Rett syndrome and intellectual disability. Genetics in medicine : official journal of the American College of Medical Genetics 18, 378-385.

- Sanyal, A., Lajoie, B.R., Jain, G., and Dekker, J. (2012). The long-range interaction landscape of gene promoters. Nature 489, 109-113.

- 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.

- Sofueva, S., Yaffe, E., Chan, W.C., Georgopoulou, D., Vietri Rudan, M., Mira-Bontenbal, H., Pollard, S.M., Schroth, G.P., Tanay, A., and Hadjur, S. (2013). Cohesin-mediated interactions organize chromosomal domain architecture. Embo J 32, 3119-3129.

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

- Tan, W.H., Bird, L.M., Thibert, R.L., and Williams, C.A. (2014). If not Angelman, what is it? A review of Angelman-like syndromes. Am J Med Genet A 164A, 975-992.

- Tarazona, S., Furio-Tari, P., Turra, D., Pietro, A.D., Nueda, M.J., Ferrer, A., and Conesa, A. (2015). Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. Nucleic Acids Res 43, e140.

- Vermunt, M.W., Tan, S.C., Castelijns, B., Geeven, G., Reinink, P., de Bruijn, E., Kondova, I., Persengiev, S., Netherlands Brain, B., Bontrop, R., et al. (2016). Epigenomic annotation of gene regulatory alterations during evolution of the primate brain. Nat Neurosci 19, 494-503.

- Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., et al. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature 457, 854-858.

- Williamson, K.A., and FitzPatrick, D.R. (2014). The genetic architecture of microphthalmia, anophthalmia and coloboma. European journal of medical genetics 57, 369-380.

- 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.

- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol 9, R137.

- Zhang, Y., Wong, C.H., Birnbaum, R.Y., Li, G., Favaro, R., Ngan, C.Y., Lim, J., Tai, E., Poh, H.M., Wong, E., et al. (2013). Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. Nature 504, 306-310.

Supplemental Information

Supplemental Figures

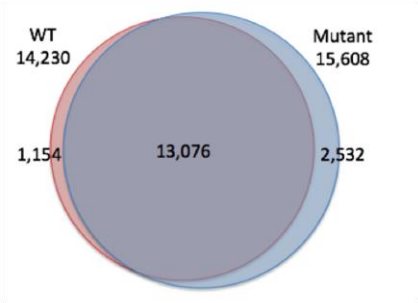


Figure S1, related to Figure 1. RNApolII peaks in wild type and mutant cells.

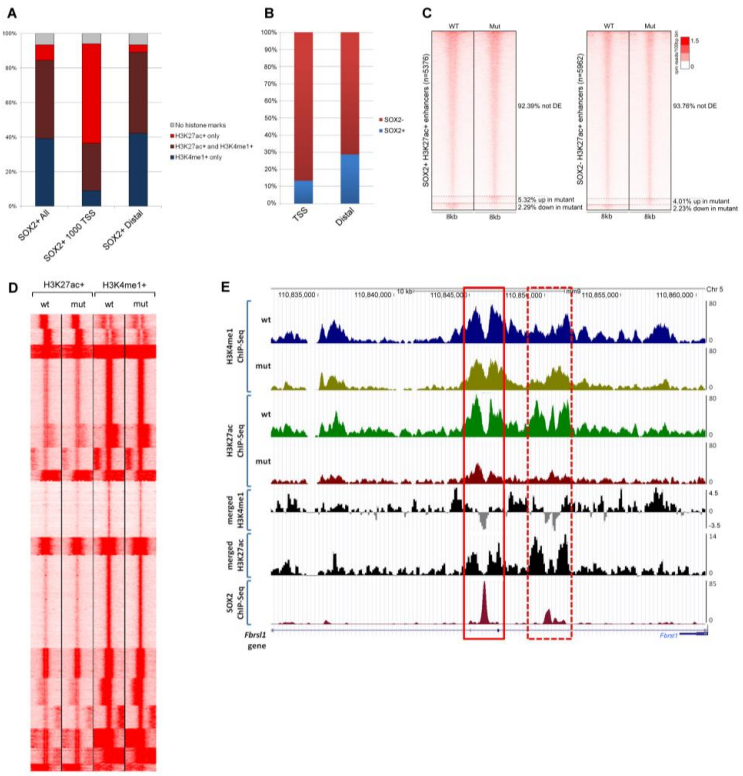
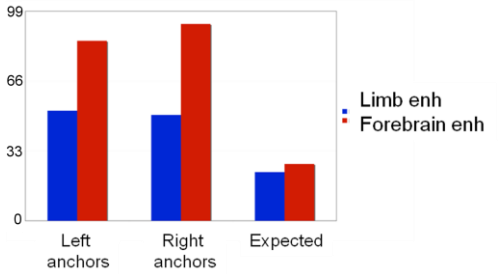


Figure S2, related to Figure 2. Long-range interaction anchors are enriched in SOX2 binding and in chromatin epigenetic marks of active/poised enhancers. (A) Fractions of different EM-positive regions within SOX2-bound TSS-linked or distal regions (chromHMM analysis). (B) Fraction of SOX2+ and SOX2- regions within epigenetically marked regions (i.e. H3K27ac+ and/or H3K4me1+) (ChromHMM). (C) Sox2 loss does not result in H3K27ac-enrichment changes at enhancers. Heatmap depicts H3K27ac enrichment (for one wt and one mutant line) for SOX2-positive and SOX2-negative enhancers. Reads were counted within 40 bins of 100 bps up- and downstream of the enhancer centre. The fraction of (not) significantly differentially enriched enhancers, as defined using DESeq2, is indicated. (D) Sox2 loss does not result in H3K27ac- or H3K4me1-enrichment changes at regions bound by SOX2 in wt cells. (E) A representative example of quantitative differences in epigenetic enhancer marks between wt and mut cells (SOX2 peaks within the Fbrsl1/AUTS2l gene, boxed in red). Merged profiles represent the variation of enrichment, positive values correspond to greater enrichment in wt, negative in mut, respectively. Typically, H3K27ac has two peaks flanking the SOX2 binding site: the height of these peaks is decreased in mutant cells, whereas H3K4me1 has a moderate increase in mut closer to SOX2 binding sites, as also shown for the second SOX2 peak on the right, where similar changes occur.

A



B

Forebrain p300 enhancers	Total number		2,454				
OVERLAP WITH ANCHOR TYPE	**Enhancers in all anchors**	**Enhancers in DISTAL anchors**	**% overlap**	**% distal in overlap**	**No. of anchors**	**No. of distal**	**% distal**
common	119	59	4.8	**49.6**	1,938	512	26.4
WT-specific	167	115	6.8	**68.9**	5,470	2,047	37.4
WT-alternative	206	132	8.4	**64.1**	6,730	2,080	30.9
MUT-specific	23	19	0.9	**82.6**	1,268	697	55.0
MUT-alternative	82	39	3.3	**47.6**	3,282	1,195	36.4

Figure S3, related to Figure 3. Enrichment of VISTA enhancers within interaction anchors. (A) Enrichment of VISTA enhancers from limb or forebrain in interaction anchors. (B) Overlap of forebrain VISTA enhancers with anchor types.

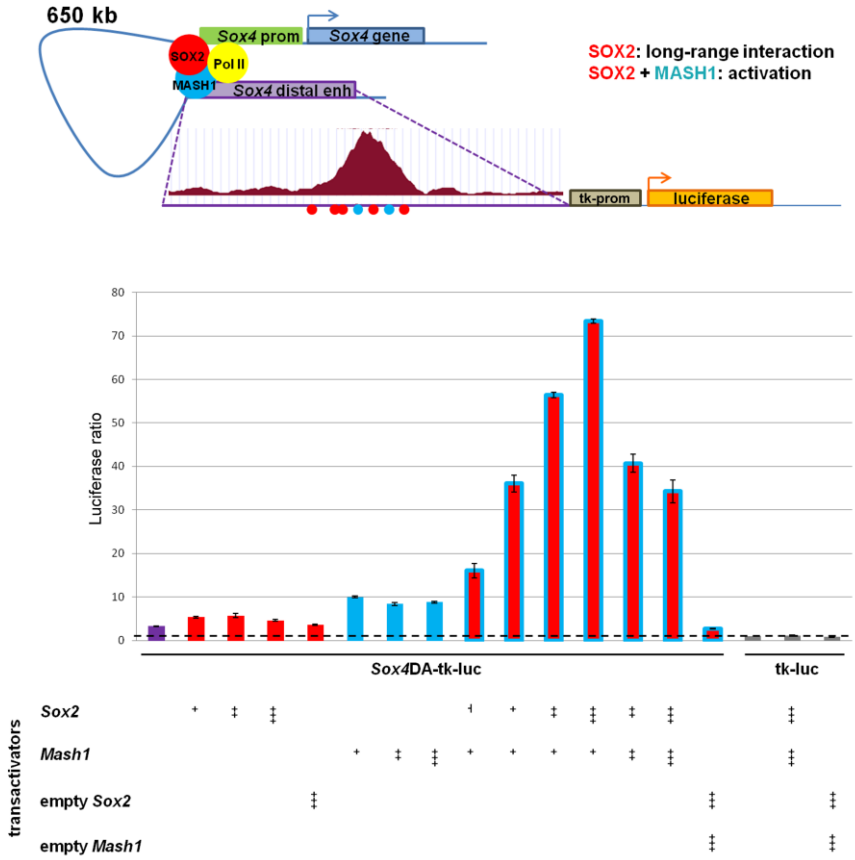


Figure S4, related to Figures 3 and 4. MASH1 is necessary in association with SOX2 to transactivate the distal Sox4 enhancer in transient cotransfection studies. Diagram showing the SOX2 peak in the Sox4 distal anchor; red and blue dots indicate SOX2 and ASCL1/MASH1 putative binding sites. Cotransfection studies in P19 and (not shown) Neuro2A cells indicate that SOX2 alone poorly transactivates the distal Sox4-connected enhancer/tk promoter, but is very active in association with ASCL1/MASH1.

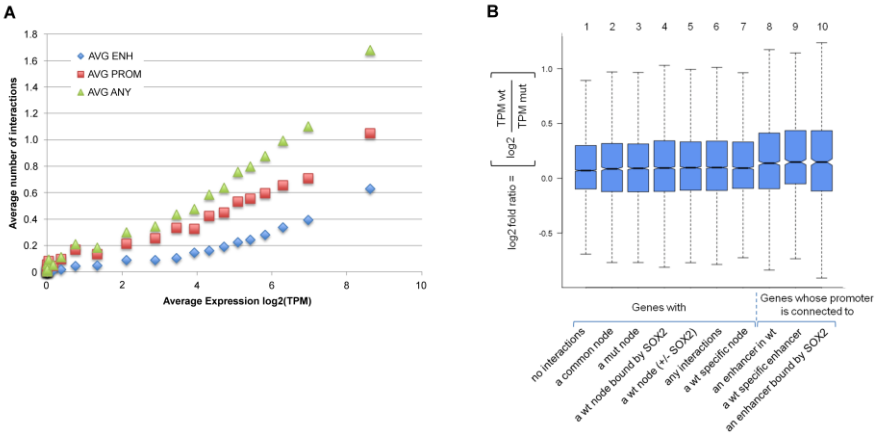


Figure S5, related to Figure 5. (A) Relation between expression levels and numbers of interactions. (B) Distribution of gene expression variation in different interaction categories.

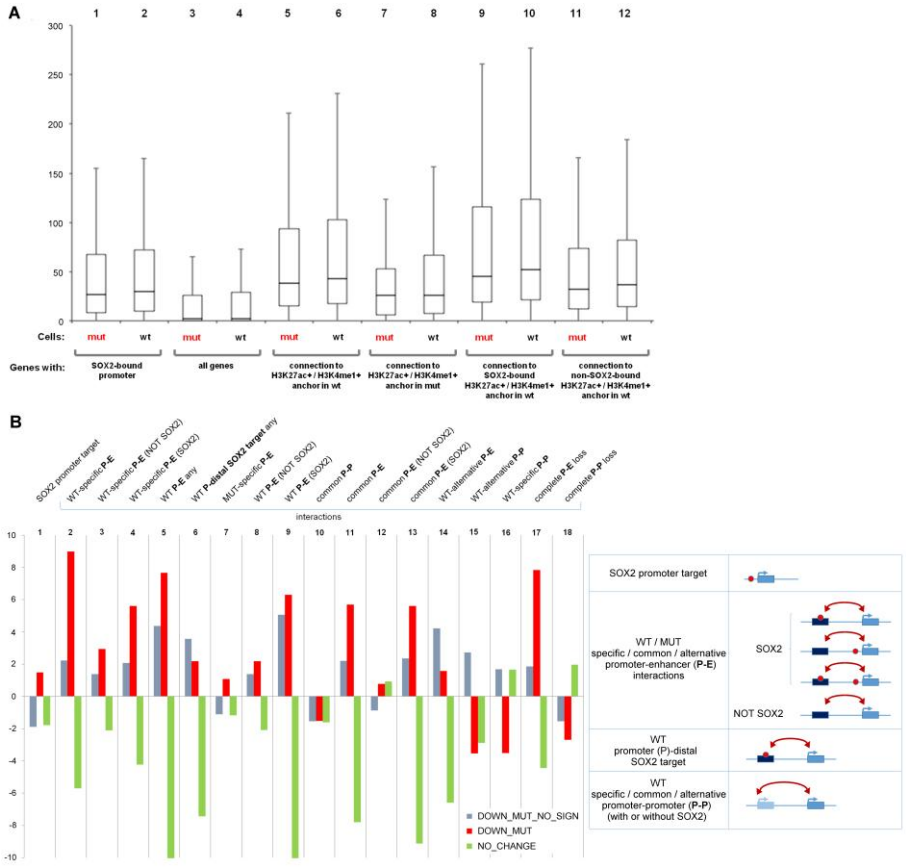


Figure S6, related to Figure 5. Coassociation scores: relationship between decreased gene expression in mut cells and different interaction categories. (A) Expression levels (RNAseq, reads), in wt and mut NSC, of genes classified according to presence/absence of SOX2-bound sites, and/or interactions (B) Coassociation scores, related to Fig. 5F. The definitions of the interactions are given below. Anchors were categorized as in Fig. 1, i.e. as "promoter" (containing a TSS) (P) or "enhancer" (E) carrying enhancer marks.

SOX2 promoter target: the gene has a SOX2-bound promoter (involved or not in an interaction).

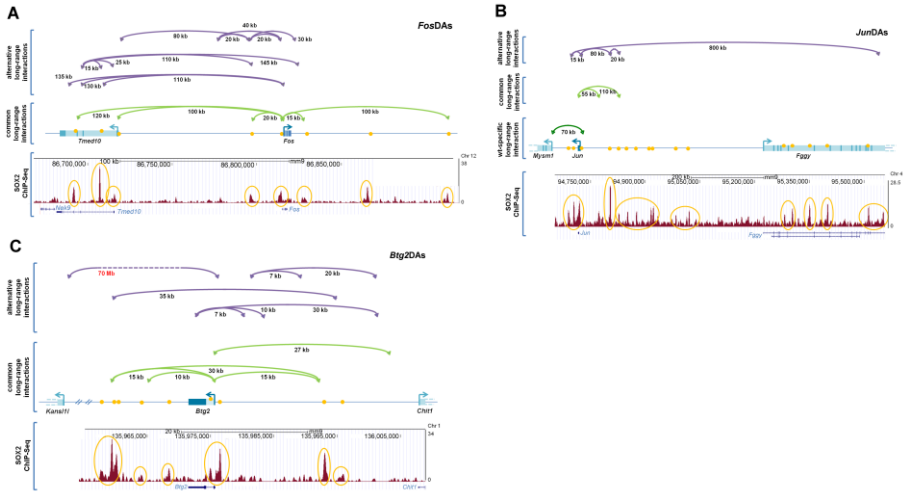


Figure S7, related to Figure 6. Interactions, in wt and mut NSC, of genes downregulated in mut NSC, and known to be related to cell growth.

Supplementary Tables

	wt brain		*Sox2* ablated brain	
Total PETs	166.7 M		170.8M	
# of PETs w/full linker	161.2 M		170.5M	
	chimeric	non-chimeric	chimeric	non-chimeric
# of PETs	17.9 M	143.3 M	12.1M	151.9 M
non-chimeric ratio		88.90%		93.19%

Mapped PETs	17.8 M	143.1 M	12.1M	151.7M
non-redundant PETs	0.9 M	6.8 M	1.9 M	22.6 M
% of non-redundancy	5.30%	4.70%		
self-ligated PETs		684 K		2.7M
# of bindng sites		11819		12068
intra-chr ligated PETs		691 K		1.3 M
# of intra-chromosomal interactions		**7046 (2+)**		**2984 (2+)**
inter-chr ligated PETs		6.3 M		18.4 M
# of inter-chromosomal interactions		19 (3+)		273 (3+)
Promoter-promoter		3182 (45%)		1270 (43%)
Promoter-enhancer		3113 (44%)		1260 (42%)
Intergenic-intergenic		283 (4%)		162 (5%)
Total # of unique nodes		8732		4653
promoter nodes		5083 (58%)		2776 (60%)
common		1864		1864
non-promoter nodes		3649 (42%)		1877 (40%)
intergenic nodes		1697		846
intragenic nodes		1952		1031

common	561	561

	WT vs MUT		WT vs MUT	
Common interactions	965	14%	981	33%
Interactions shared anchors	3376	48%	1389	47%
Specific interactions	2705	38%	614	21%
Total interactions	7046		2984	

Table S1. ChIA-PET analysis: general features.

	Peaks
SOX2 total peaks	**18361**
of which:	
3’-UTR	122
miscRNA	37
miRNA	0
TTS	203
pseudogene	2
Exon	140
Intron	8564
Intergenic	8044
Promoter	1194
5’-UTR	53

snoRNA	0
rRNA	0

Table S2, related to Figure 2. Location of SOX2 ChIP-seq peaks.

Interactions	Total SOX2-bound anchors	Interactions involving ONE SOX2-positive anchor	Interactions involving TWO SOX2-positive anchors	Total SOX2+ interactions	Total number of interactions	% of interactions with SOX2	% of anchor regions (single) with SOX2
common	684	381	151	532	969	**54.9**	35.3
WT-alternative	1738	1196	271	1467	3365	**43.6**	25.8
WT-specific	1140	820	160	980	2735	**35.8**	20.8
MUT-alternative	887	671	108	779	1641	**47.5**	27.0
MUT-specific	167	113	27	140	634	**22.1**	13.2

Table S3, related to Figure 2. SOX2-ChIP-seq. Numbers and frequency of SOX2-bound anchors within the different interaction categories

Gene associated with the proximal anchor	Cloned anchor [DA: distal anchor; PA: proximal anchor]	Distance PA-DA [bp]	Presence of SOX2 peaks (ChIP-seq)				Presence of already defined forebrain enhancer (VISTA enhancer)		Tested in fish (GFP+ at 24 hpf)				
			cloned anchor		associated anchor		p300 binding site	validated enhancer	reproducible GFP+ expression in forebrain	GFP+ in forebrain / tot transgenics		GFP+ in forebrain and elsewhere / tot transgenics	intensity of GFP+ expression in forebrain
			in our NSCs	in other ChIP-seq	in our NSCs	in other ChIP-seq				%	n°	n°	
Nkx2.1	DA504	10,000	-	✓	-	-	✓	✓	+	94.7%	36/38	37/38	++++
Sp8	PA545	10,000	-	-	✓	✓	-	-	+	75.0%	15/20	15/20	+
	DA545		✓	✓	-	-	-	-	+	94.9%	37/39 (1)	39/39	++++
Coup-TF1	DA2467	12,000	✓	-	✓	✓	-	-	+	35.0%	7/20	7/20	+
	DA602	27,000	✓	-	-	-	-	-	+	40.0%	8/20	8/20	+
Ntng1	DA1414	100,000	✓(2)	✓	✓(2)	-	-	-	+/-	2.6%	(1-3)/39 (3)	(5-7)/39 (4)	+
Irx1	DA597	450,000	✓	✓	-	✓	✓	(6)	+	49.1%	29/59	53/59 (5)	++
Socs3	DA463	32,000	✓	✓	✓	✓	✓	(6)	+	23.5%	8/34 (7)	16/34	+
Chd7	DA1439	40,000	✓	✓	-	-	-	-	+	13.6%	3/22	9/22	++
Sox3	DA2733	350,000	-	- (8)	-	-	✓	✓	+	100.0%	130/130	130/130	+++++
	DA2702	65,000	✓	✓	-	-	-	-	+	81.2%	13/16	15/16	+++
Sox4	DA2458	650,000	✓	✓	✓	✓	✓	✓	+	93.4%	114/122	119/122	+++
Cxcr4	DA62	180,000	✓	✓	-	-	✓	(6)	+	31.2%	30/96	56/96	++
Zfp335	DA1303	125,000	✓	-	-	✓	-	-	-	0%	0/56	4/56	-
	PA1303		-	✓	✓	-	-	-	-	0%	0/25	5/25	-

(1): 10/37 GFP+ also in midbrain-hindbrain boundary;
(2): the peak is not completely included;
(3): one embryo with clearly positive signal, two with weak signal;
(4): five embryos with clearly positive signal, two with weak signal;
(5): 18/53 in more posterior neural tube regions;
(6): not present in VISTA enhancer;
(7): 2 embryos with clearly positive signal, 6 with weak signal;
(8): presence of a SOX3 peak.

Table S4, related to Figure 4. Summary of transgenic experiments in zebrafish.

	Gene_id	Mean mut expression	Mean wt expression	T test	log10P	Reliable	Reliable FR	FR	wt P-E any	wt P-E NOT SOX2
1	Gfap	66.35	1133.94	0.094602	-100	1	1	-4.092982		
2	Socs3	29.32	254.43	0.052057	-100	1	1	-3.112967	✓	
3	Clu	34.10	227.71	0.172497	-100	1	1	-2.735763		
4	Cyr61	105.61	482.53	0.059767	-100	1	1	-2.190841	✓	
5	Mt2	316.08	1427.69	0.009051	-100	1	1	-2.174954		
6	Fos	818.55	2868.27	0.064026	-100	1	1	-1.808920	✓	
7	Junb	105.83	359.06	0.104522	-100	1	1	-1.761470	✓	
8	Egr1	493.33	1595.90	0.069378	-100	1	1	-1.693543	✓	
9	Aldoc	211.03	666.20	0.007900	-100	1	1	-1.658048		
10	Pygb	146.33	417.48	0.052679	-100	1	1	-1.511874	✓	
11	Mt1	692.17	1963.79	0.010874	-100	1	1	-1.504307	✓	
12	Gja1	318.71	903.56	0.014802	-100	1	1	-1.503073	✓	
13	Ttyh1	276.23	697.78	0.072625	-100	1	1	-1.336611	✓	
14	Mt3	531.49	1334.62	0.014240	-100	1	1	-1.328157	✓	
15	Jun	248.82	621.48	0.066673	-100	1	1	-1.320266	✓	
16	Glud1	436.58	965.63	0.008381	-100	1	1	-1.145048	✓	
17	Cpe	796.49	1539.72	0.109824	-100	1	1	-0.950858	✓	
18	3110035E14Rik	30.92	194.73	0.317008	-100	1	1	-2.651125		
19	Zfp36	39.91	233.94	0.084719	-100	1	1	-2.548223		
20	Epb4.1l2	79.03	213.23	0.363927	-100	1	1	-1.430752		
21	Btg2	133.53	359.17	0.141180	-100	1	1	-1.426775	✓	
22	S100a16	111.17	292.21	0.283060	-100	1	1	-1.393408	✓	
23	Dynll2	136.12	316.42	0.306526	-100	1	1	-1.216393	✓	✓
24	Fads2	608.34	1047.26	0.087281	-97.414231	1	1	-0.783574	✓	
25	C4b	4.50	92.93	0.241502	-97.386090	1	1	-4.336997		
26	Sparc	88.31	319.68	0.037103	-95.733576	1	1	-1.854744		
27	Ier2	146.45	345.89	0.229963	-94.992777	1	1	-1.239305	✓	
28	Sox10	17.18	95.25	0.370489	-94.410367	1	1	-2.464458		
29	C1ql1	19.29	97.67	0.353574	-89.845053	1	1	-2.333831	✓	✓
30	Glul	301.65	644.60	0.007536	-82.455305	1	1	-1.095267	✓	
31	Mir6935	37.64	178.08	0.064755	-81.295952	1	1	-2.239270		
32	Abhd4	193.44	451.24	0.006222	-73.746762	1	1	-1.221618	✓	
33	Fads1	291.81	531.83	0.205661	-72.815326	1	1	-0.865738	✓	✓
34	Pfn2	277.00	534.70	0.113147	-71.463314	1	1	-0.948593	✓	
35	Aqp4	107.36	280.54	0.120083	-70.139739	1	1	-1.384893	✓	
36	Ppap2b	314.38	589.26	0.018052	-67.157923	1	1	-0.906166	✓	
37	Sfxn5	111.48	314.63	0.004630	-66.849952	1	1	-1.496064		
38	Parvb	20.61	99.84	0.283996	-66.515995	1	1	-2.270919		
39	Scrg1	161.72	342.54	0.128059	-65.271455	1	1	-1.082260		
40	Ntrk2	221.67	449.67	0.074916	-62.503610	1	1	-1.020123	✓	
41	Prnp	191.81	409.04	0.009562	-61.715415	1	1	-1.092153		
42	Tmeff2	22.42	83.36	0.359233	-59.191752	1	1	-1.889665	✓	✓
43	Cd81	836.25	1237.76	0.037905	-58.015305	1	1	-0.565675	✓	
44	March8	66.70	183.78	0.205097	-57.134272	1	1	-1.460818		
45	Abat	104.27	247.17	0.001370	-56.636506	1	1	-1.244380	✓	
46	Oat	140.49	336.16	0.010311	-56.172702	1	1	-1.258069		
47	Gadd45g	66.11	210.72	0.010497	-55.149065	1	1	-1.670963	✓	
48	Dpysl3	129.83	303.46	0.057578	-52.067122	1	1	-1.224300		
49	Tspan7	244.48	474.79	0.032942	-51.541017	1	1	-0.957307	✓	✓
50	Cadm4	305.11	509.91	0.193040	-50.769312	1	1	-0.740730		
51	Id4	22.17	96.89	0.220038	-50.359209	1	1	-2.122685	✓	
52	Gltp	52.39	129.57	0.326475	-49.781293	1	1	-1.304822		
53	Nrbp2	60.01	161.84	0.191962	-49.180618	1	1	-1.429782		
54	Abhd3	24.77	118.62	0.000182	-47.938622	1	1	-2.254851		

55	Gabbr1	38.09	154.84	0.002757	-47.439965	1	1	-2.020569	✓	✓
56	Igfbp5	31.54	115.79	0.170279	-46.172042	1	1	-1.872935		
57	Appl2	63.65	180.94	0.027868	-45.177386	1	1	-1.505785	✓	
58	S1pr1	49.84	162.84	0.017027	-44.619242	1	1	-1.706073	✓	
59	Lrig1	125.21	251.39	0.023648	-44.338702	1	1	-1.005019	✓	
60	Paqr8	33.81	120.94	0.073512	-43.675695	1	1	-1.835661	✓	
61	Mgst1	113.86	234.81	0.004540	-43.468626	1	1	-1.043560		
62	Rsrp1	153.63	267.58	0.220788	-43.233825	1	1	-0.800111		
63	S100a1	30.96	105.91	0.198400	-41.565266	1	1	-1.770870		
64	Prdx6	358.47	567.50	0.013855	-41.375650	1	1	-0.662628		
65	Tmem163	1.26	24.72	0.384390	-37.696114	1	1	-4.193171		
66	Atp1b2	207.55	366.10	0.067644	-37.409991	1	1	-0.818514	✓	
67	Npdc1	78.73	166.90	0.239866	-36.414433	1	1	-1.083064		
68	Ppp2r2b	94.55	203.78	0.113873	-35.204928	1	1	-1.107019		
69	Entpd2	1.34	31.34	0.262111	-34.707594	1	1	-4.451651		
70	Clip3	121.29	211.19	0.338827	-33.555965	1	1	-0.799535		
71	Csdc2	10.30	69.86	0.053313	-33.042657	1	1	-2.749416		
72	Trim9	60.53	147.81	0.107453	-33.005931	1	1	-1.286695	✓	✓
73	Fbrsl1	17.55	79.15	0.116694	-32.694751	1	1	-2.166682	✓	
74	Map1b	10.26	45.52	0.341062	-32.383792	1	1	-2.138537		
75	Dhrs1	104.00	227.87	0.006881	-32.097655	1	1	-1.130941		
76	Fhl1	129.60	264.75	0.055687	-32.031263	1	1	-1.029978		
77	Cnn3	474.08	688.19	0.089943	-31.495588	1	1	-0.537565	✓	
78	Gstm1	349.88	532.78	0.003577	-31.246915	1	1	-0.606538		
79	Gsn	13.86	68.51	0.131781	-30.693785	1	1	-2.297535		
80	Mgll	32.00	104.36	0.067502	-30.382314	1	1	-1.702455		
81	Pdlim4	9.87	65.97	0.036992	-30.081581	1	1	-2.728257		
82	Rtn1	123.71	232.74	0.048703	-29.671616	1	1	-0.911271		
83	Ctsl	140.17	251.57	0.119835	-29.457245	1	1	-0.843327		
84	Lamp1	510.90	738.01	0.017125	-29.214635	1	1	-0.530515		
85	Egr2	82.58	164.05	0.124663	-28.971762	1	1	-0.989376	✓	
86	Pink1	70.73	152.67	0.141882	-28.064392	1	1	-1.108965		
87	Gipc1	68.98	131.74	0.292247	-27.760602	1	1	-0.932412		
88	Phyhd1	12.94	65.65	0.035118	-27.760086	1	1	-2.333678		
89	Psd2	7.09	46.88	0.117958	-27.515662	1	1	-2.708550		
90	Asrgl1	350.93	524.76	0.004405	-27.211109	1	1	-0.580361	✓	
91	Gdpd2	47.44	117.06	0.137619	-27.179936	1	1	-1.301307		
92	Rgma	55.68	139.55	0.051674	-26.918892	1	1	-1.323910	✓	
93	Bfsp2	6.97	39.85	0.241633	-26.916025	1	1	-2.498293		
94	Naaa	10.19	47.12	0.234864	-26.241817	1	1	-2.197688		
95	Tst	38.37	112.28	0.002412	-26.057154	1	1	-1.546622	✓	
96	Gdi1	216.22	331.74	0.167877	-25.979768	1	1	-0.617299		
97	Syt11	410.02	603.82	0.020930	-25.879013	1	1	-0.558270		
98	Ddt	65.57	146.81	0.055617	-24.984276	1	1	-1.161553		
99	Nkain4	16.43	67.25	0.086095	-24.956802	1	1	-2.026663		
100	Ctsb	144.98	259.84	0.002068	-24.940977	1	1	-0.841363		

Table S6, related to Figures 5, 6. List of the 100 more down-regulated genes in mut NSC.

Gene	Promoter interaction with distal enhancer	SOX2-bound distal anchor	SOX2-bound promoter
Apoe	●		●
Cst3	●	●	●
Fos	●	●	
Mt1	●	●	
Egr1	●	●	
Cpe	●	●	
Mt2			
Mt3	●		●
Ednrb	●		
Ndrg2	●	●	
Fads2	●	●	●
Sirt2			●
Glud1	●	●	
Fth1	●		
Gja1	●		●
Mmd2	●	●	
Ttyh1	●	●	
Atp1a2	●		●
Aldoc			●
Glul	●	●	●
Wsb1	●		●
Jun	●	●	●
Ppap2b	●	●	●
Prdx6			
Pfn2	●	●	●

Fads1	•		
Rnd2	•	•	
Bcas1			
Cadm4			
Gpr17	•		•
Slc6a11	•		
Cyr61	•	•	•
Tspan7	•		
Abhd4	•		•
Ntrk2	•	•	•
Pygb	•	•	•
Prnp		•	•
Atp1b2	•	•	
Sepp1			•
Btg2	•	•	•
Junb	•	•	
Ier2	•	•	•
Scrg1			•
Oat		•	
Pisd-ps1			
Sparc			
Dynll2	•		
Sfxn5			
Dpysl3			•
Map1lc3b			
S100a16	•	•	•
Tubb4a			
Aqp4	•	•	•
Rsrp1			
Fhl1			

Mrps6	•	•	
Ctsb			
Socs3	•		•
Gfap			

Table S7, related to Figure 5. Interactions and/or SOX2 peaks within the most expressed protein coding genes among genes down-regulated in mutant cells (DOWN-MUT).

CHAPTER 3:

SUPPLEMENTARY RESULTS

1. DIFFERENTIATION OF SOX2-MUTANT NSC IS LINKED TO PROFOUND EXPRESSION CHANGES IN GENE SUBSETS DEFINED BY SOX2 BINDING AND LONG-RANGE INTERACTIONS.

To evaluate the effects of *Sox2* loss and of chromatin interactions changes in NSC upon subsequent differentiation, we performed RNA-seq analysis on cells induced to differentiation towards neurons and astroglia, according to established protocols [Gritti A et. al, 1996; Gritti et al., 2001](Figure 2, 3). We evaluated early stages (day 4) and more advanced stages (day 11), in which cells with differentiated characteristics (β-tubulin positive neurons, GFAP-positive astrocytes) are clearly present. At this stage, neurons (approximately 20%) are mainly of the Gaba-ergic type, and show a subset of markers of terminally differentiated neurons (such as β-tubulinIII); the remaining cells are mostly astroglia, with few (1-2%) oligodendrocytes [Gritti et al., 1996; Gritti et al., 2001].

We first evaluated differentiation of neurosphere cultures derived from wt and *Sox2*-deleted P0 mouse forebrain, by immunofluorescence with neuronal and glial markers (Figure 3). Undifferentiated neurospheres, dissociated to single cells, were made to adhere to slides, in the presence of bFGF. After 4 days, bFGF was removed, and 1% FBS was added, leading to differentiation within 11 days from initial plating (Figure 2; [Gritti et al., 1996; Gritti et al., 2001; Cavallaro et al., 2008]). We stained cells at day 4 and 11 with antibodies against β-tubulin III (neuronal marker) and GFAP (astrocytic marker)(Figure 3). At day 4, WT cells begin to show β-tubulin-

positive neurons, exhibiting initial branching morphology, together with a majority of weakly GFAP-positive cells with glial-like morphology; no, or very few, doubly stained cells were visible. In strong contrast, mutant cells showed weakly β-tubulin stained cells, with a less differentiated morphology than wild-type neuronal cells, and the majority, if not all, of these cells showed GFAP-positivity; we could not identify signicant numbers of cells exhibiting neuronal- or glial-type morphology comparable to wild type cells. At day 11, whereas wild-type cells showed clear phenotypic differences between neuronal and glial cells, mutant cells showed strongly decreased numbers of β-tubulin-posititve cells, with poor differentiated morphology, and the majority, if not all, of these cells showed GFAP-positivity.

These data indicate that mutant cells tend to differentiate poorly into neurons, and have an impaired ability to resolve their characteristics into well defined neuronal and astroglial populations. This result is suggestive of a "promiscuous" pattern of gene activity, leading to excess glial-type gene expression at day 4, as described by RNA-seq.

We then analysed gene expression, on parallel cultures differentiated as above (Figure 2), and analysed by RNAseq, at day 4 (D4) and day 11 (D11) of differentiation (Figure 4).

In wild-type cells, at day 4, (D4 WT) over 5000 genes showed clearly decreased levels of expression relative to undifferentiated (day 0, D0 WT) cells, whereas only about 1000 genes were up-regulated. In mutant cells, however, similar numbers of genes were up- or down-regulated relative to D0 (1723 vs 2271 genes). When comparing gene expression between D4 MUT versus D4 WT cells, 2618 genes were more expressed in MUT than in

WT cells, whereas only 291 genes were more expressed in WT versus MUT cells. Strikingly, almost all (2439/2618 genes) of the genes more expressed in MUT versus WT cells corresponded to the class of genes down-regulated at day 4 versus day 0 in WT cells. This trend holded true at Day 11, when approximately 50% of the genes more expressed in MUT versus WT cells at D4 remained more expressed than their WT counterpart at Day 11 (Figure 4). These data hint at a possible negative role of Sox2 on the expression of a subclass of genes during differentiation.

We then asked whether the deregulated genes in mutant cells were predominantly cell type-specific (i.e. neuronal or glial-specific), or not (i.e. common to both lineages). To this end, we used as a comparison a data base of gene expression [Zhang et al., 2014] for mature neurons and glial cells; although full maturation of neurons is not obtained in our in vitro differentiation protocol, a large proportion of neuronal genes from the Zhang et al. data base is expressed in our in vitro culture. Genes deregulated at day 4 in mutant cells versus WT (i.e. up- or down-regulated) were highly significantly enriched in astrocytic-, but not neuronal-type, genes.

Similarly, at day 11, genes overexpressed in MUT cells (versus WT cells) were highly significantly enriched in astrocytic-type genes, but not in neuronal-type genes; however, importantly, genes overexpressed in WT cells versus MUT cells were enriched in both gene categories, neuronal- and astrocytic-type genes. Thus, whereas glial-type genes are overexpressed in MUT, versus WT cells, both at the initial (day 4) and advanced (day 11) differentiation stages, neuronal-type genes are significantly downregulated in MUT, versus WT cells, at the late stages of differentiation. The implications of this finding is that either less neurons are generated in

mutant cells at day 11 (Figure 3), or the neurons show delayed/diminished expression of some genes.

2. OVEREXPRESSION OF *Socs3* PARTIALLY RESCUES THE DIFFERENTIATION DEFECT OF SOX2-MUTANT NEURAL STEM CELLS.

I had found that overexpression of *Socs3* in *Sox2*-mutant cells was able to partially rescue the proliferation defect of *Sox2*-mutant NSC (see submitted manuscript in chapter 2, Figure 5). I subsequently evaluated if overexpression of *Socs3* could be able to rescue also the differentiation defect of *Sox2*-mutant cells.

Neurosphere cultures derived from wt, *Socs3*-transduced wt cells and Socs3-transduced mutant cells (see submitted manuscript in chapter 2 and Figure 5) were dissociated to single cells and were made to adhere to slides, in the presence of bFGF; after 4 days, bFGF was removed, and 1% FBS was added, leading to differentiation within 11 days from initial plating, as described above.

We stained cells at differentiation day 4 and 11 with antibodies against beta-tubulin III and GFAP: at day 4, untransduced (control) wild-type cells (wt nt) began to show some β-tubulin-positive neurons, exhibiting initial branching morphology, together with a majority of weakly GFAP-positive cells with glial-like morphology, as expected and shown before (Figure 4, left panels; see also Figure 2). In contrast, Socs3-transduced cells were all GFAP-negative, and they all appeared β-tubulin-positive (Figure 4, central panels). At day 11, wt control cells (wt nt) showed β-tubulin-positive cells with mature morphology and extensive arborization, as expected (Figure 5,

left panels); *Socs3*-transduced wt cells showed fewer mature neurons compared to the control sample, but the majority of cells were GFAP-positive, similarly to untransduced cells (Figure 5, central panels). In contrast, *Socs3*-transduced mutant cells showed no GFAP-positivity at day 11; instead, all cells were β-tubulin-positive, though they seemed to have a suffering morphology (Figure 4, right panels).

It should be noted, however, that in these (initial) experiments, I analyzed the *Socs3* transduced mutant cells that were all *Socs3* positive (derived from the experiments in the submitted manuscript, Figure 6), whereas only a fraction of the wild type cells were *Socs3*-transduced (about 30% by FACS analysis, not shown).

This result suggested that *Socs3* antagonizes precocious differentiation of NSC into astroglia (as previously demostrated, Cao et al., 2006; Schmidt-Edelkraut et al., 2013) and it might drive the differentiation towards neurons. In this experiment, *Socs3*-transduced wt cells were only a relatively small proportion (30% by FACS analysis, not shown) of the entire sample; this could be the reason why these cells were still able to differentiate into glial cells at advanced differentiation stages (day 11). Excessive levels of ectopic *Socs3* could have toxic effects on neuronal differentiation, which might explain the reduction in β-tubulin-positive cells with respect to the control cells (wt nt). Moreover, *Socs3*-transduced mutant cells, which are all *Socs3* positive cells (Figure 6), differentiated only into β-tubulin-positive cells: at day 4 these cells showed underdeveloped neuronal morphology, as expected, but later they appeared to be suffering without a mature neuronal phenotype, probably caused by the absence of glial cells, whose differentiation was inhibited by *Socs3*.

It's important to remark that the *Socs3*-transduced mutant cells I used in these first experiments had been expanded for many passages in the undifferentiated state, and it is possible that, together with *Socs3* expression, they have been selected for some additional characteristic conferring them a growth advantage in these conditions.
I plan to repeat these experiments by freshly transducing, in parallel, mutant and control (wild type) cells with the *Socs3*-GFP-expressing lentivirus. I plan to use conditions (multiplicity of infection, MOI 5) in which only a fraction of the cells is transduced (about 20%-30%). I will then ask whether transduced mutant cells recover the ability to fully differentiate into neurons. An advantage of having only a fraction of the mutant cells expressing *Socs3* is the expected persistence of glial cells (differentiating from non-transduced cells), which are lost following *Socs3* expression in all cells, and which might be required for full neuronal differentiation of the rescued mutant cells. These experiments will allow me to make more final conclusions as to the ability of *Socs3* to rescue neuronal differentiation of mutant cells.

3. 3C EXPERIMENTS.

I performed 3C (Chromosome Conformation Capture) experiments to ask if the long-range interaction involving Socs3 and a distant enhancer detected by ChIA-PET (see the manuscript at chapter 2, figure 7A) could also be observed in vivo in the developing brain. The chromatin was prepared (as described by Dekker et al., 2002) from E 12.5 wild-type mouse telencephalon. The cross-linked chromatin was cut using the HindIII restriction enzyme; I scanned 88 Kb spanning the Socs3 locus using primers

flanking HindIII restriction sites (see Figure 1 as a scheme of 3C method and Figure 7). One of these long-range interactions was confirmed ("LIG" lane from cross-linked chromatin, digested and ligated, Figure 8). A number of controls is essential for correct interpretation of 3C data. I confirmed that the 3C primers amplified artificial 3C products in vitro (BAC1 and BAC2, Bacterial Artificial Chromosome), and did not amplify digested but not ligated (DIG) chromatin (Figure 8). Furthermore, the 3C product derived from crosslinked template (LIG) was sequenced and corresponded to the expected sequence (not shown). Thus, it might be possible to use 3C to validate larger numbers of ChIA-PET interactions in the early embryo. This result will probably allow me to study the same interaction in *Sox2*-transduced cells, about the ability to obtain the rescue of this interaction by Sox2.

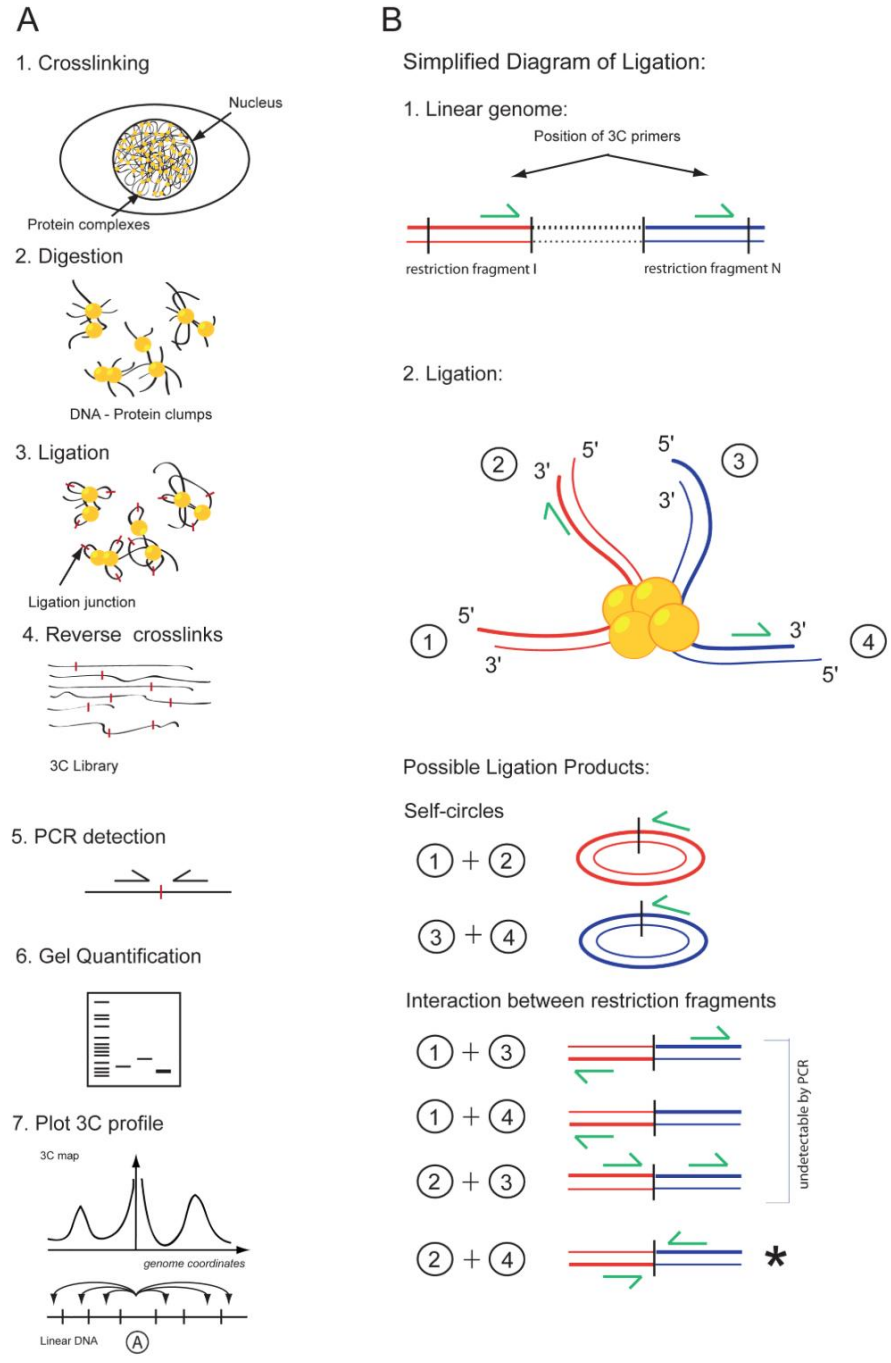


Figure 1: Illustration of 3C method [Naumova et al., 2012]

4. FIGURES.

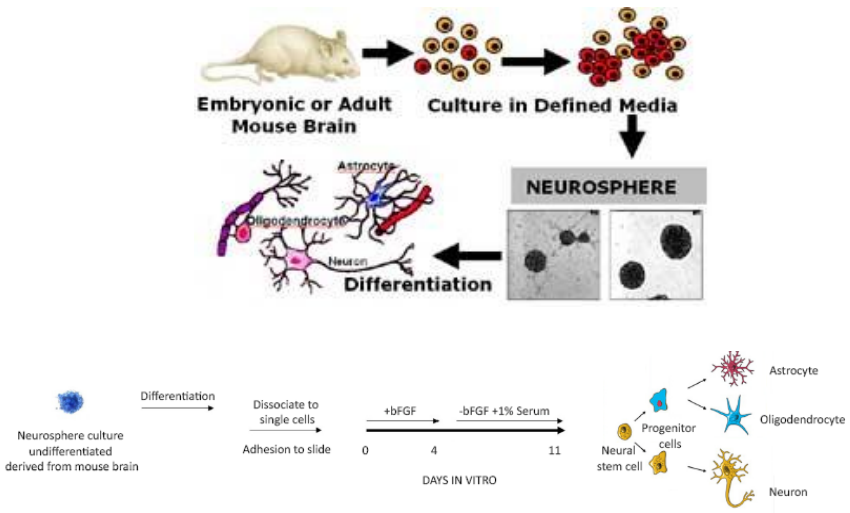
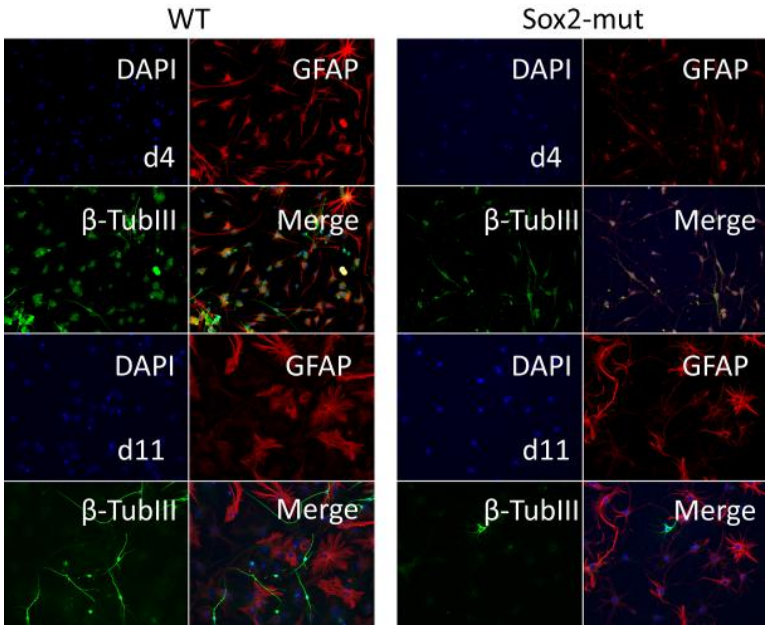


Figure 2: in vitro neural stem cell differentiation scheme (adapted from Cavallaro et al., 2008).



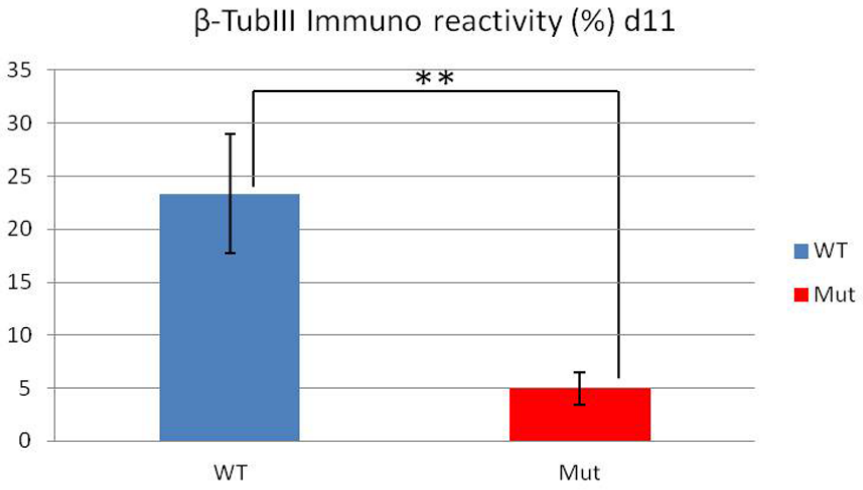


Figure 3: β-tubulinIII and GFAP immunofluorescence on differentiation day 4 and day 11: at day 4, wild-type cells begin to show beta-tubulin-positive neurons, exhibiting initial branching morphology, together with a majority of weakly GFAP-positive cells with glial-like morphology. Mutant cells showed weakly beta-tubulin stained cells, with a less differentiated morphology than wild-type

neuronal cells, and the majority, if not all, of these cells showed GFAP-positivity. At day 11, most neuronal cells in wt cultures showed mature morphology with extensive arborisation, while neurons produced by Sox2-deleted cells were reduced in numbers and showed rare arborizations. GFAP-positive cells did not vary significantly (ca. 70% in wt sample and ca. 80% in Sox2-ablated cells, compared to the total number of cells, not shown). Below: histogram showing quantification of the % of β-tubulinIII-positive cells at day 11 in wild type and mutant cells. Histogram bars represent mean ± sd of the three wild type and three mutant differentiated cell populations from which the RNA-seq data were generated. β-tubulin-positive cells are more than 20%, compared to the total number of cells, in wt sample; in contrast, β-tubulin-positive cells in Sox2-ablated cells are ca. 5%. ** P value < 0.001 (Chi-Square test). n=5 independent experiments.

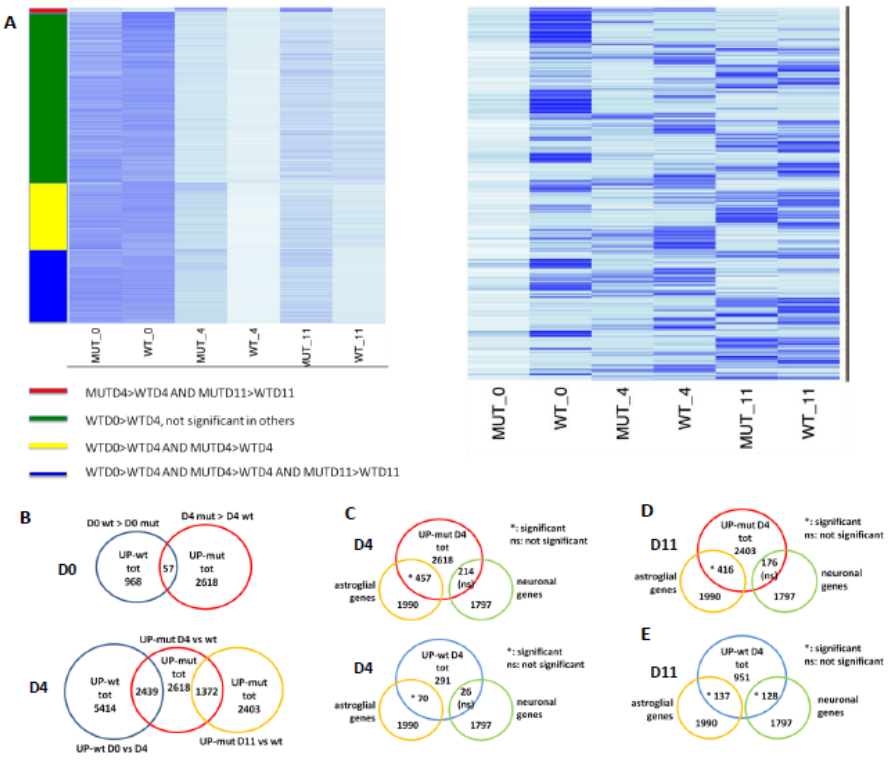


Figure 4: Sox2 loss causes gene expression changes in differentiating NSC.

(A) Left: Heat map showing expression levels of genes downregulated between D0 (wt) and D4 (wt). Sub-categories are shown in different colours. Right: Heat map showing expression levels of genes that have reduced expression in mutant cells at D0, with their expression levels at D4, and D11. (B, C, D, E) Venn diagrams showing numbers and overlaps between different categories of genes changing their expression between wild type and mutant cells, and/or during differentiation.

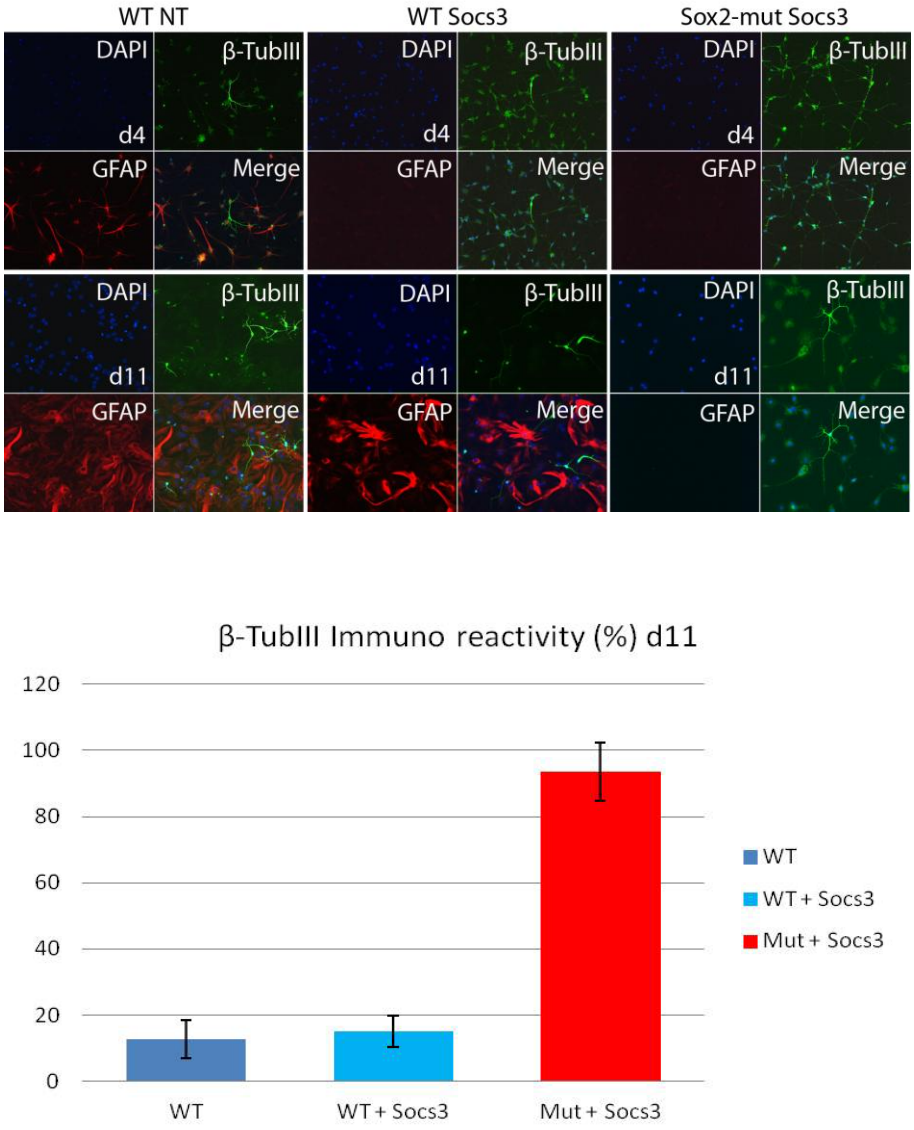


Figure 5: β-tubulinIII and GFAP immunofluorescence on differentiation day 4 (d4) and day 11 (d11) in wild type untransduced (WT NT, control) or Socs-3 transduced wild type (WT Socs3), or Sox2-mutant (Sox2-mut Socs3), cells. At d4, Socs3-transduced cells were all β-tubulinIII-positive. At d11, whereas transduced WT cells showed similar number of mature neurons with respect to the control WT NT, Sox2-mut Socs3 cells showed only β-tubulinIII positive cells and no GFAP-positive glia (only 1% compared to the total number of cells, not shown), though cells have a somewhat suffering morphology and do not appear morphologically

completely differentiated. GFAP positive cells are ca. 80%, compared to the total number of cells, both in WT NT sample and WT Socs3 sample. Below: histogram showing quantification of the percentage of β-tubulinIII-positive cells at day 11: β-tubulin-positive cells are ca 15%, compared to the total number of cells, in wt sample; β-tubulin-positive cells are more than 18% in wt+Socs3 sample; in contrast, β-tubulin-positive cells in rescued cells are ca. 95%. Histogram bar represent mean ± sd. n=4 independent experiments.

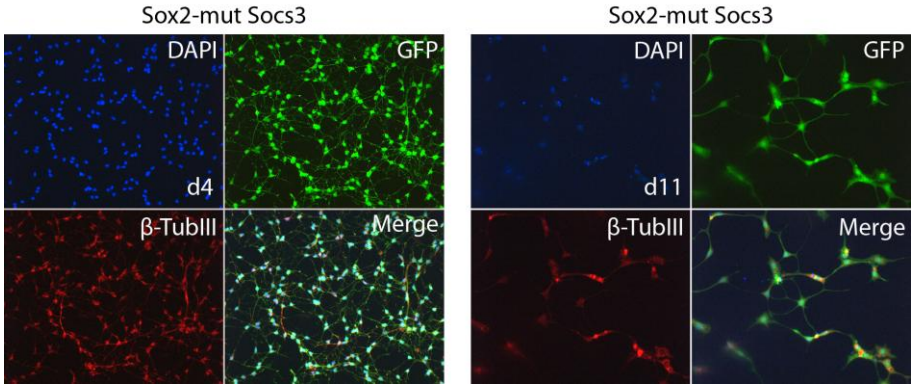


Figure 6: β-tubulinIII and GFP immunofluorescence on Socs3-transduced Sox2-mutant cells. All cells are GFP-positive, indicating that all cells carry the Socs3-lentiviral vector, coexpressing GFP.

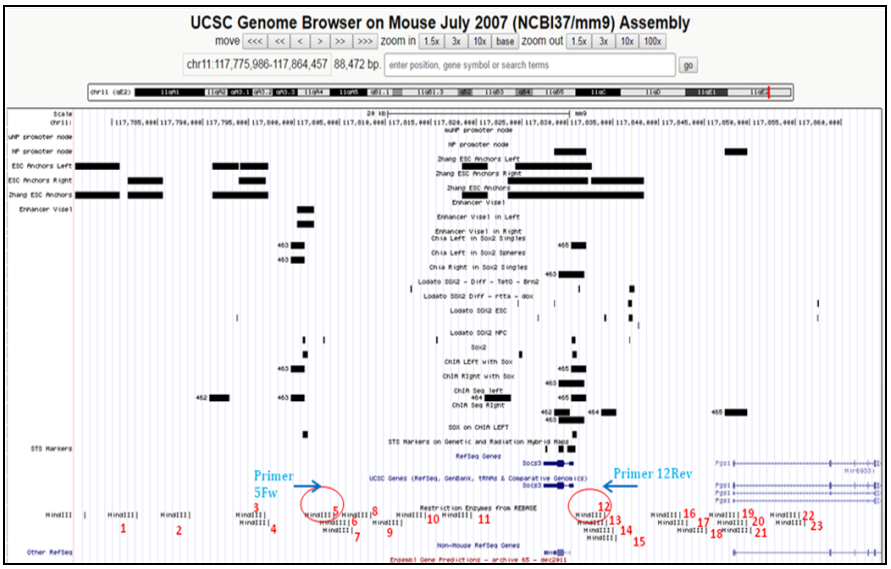


Figure 7: Socs3 locus and primers (Primer 5Fw; Primer 12Rev) used for 3C PCR shown below (Figure 8).

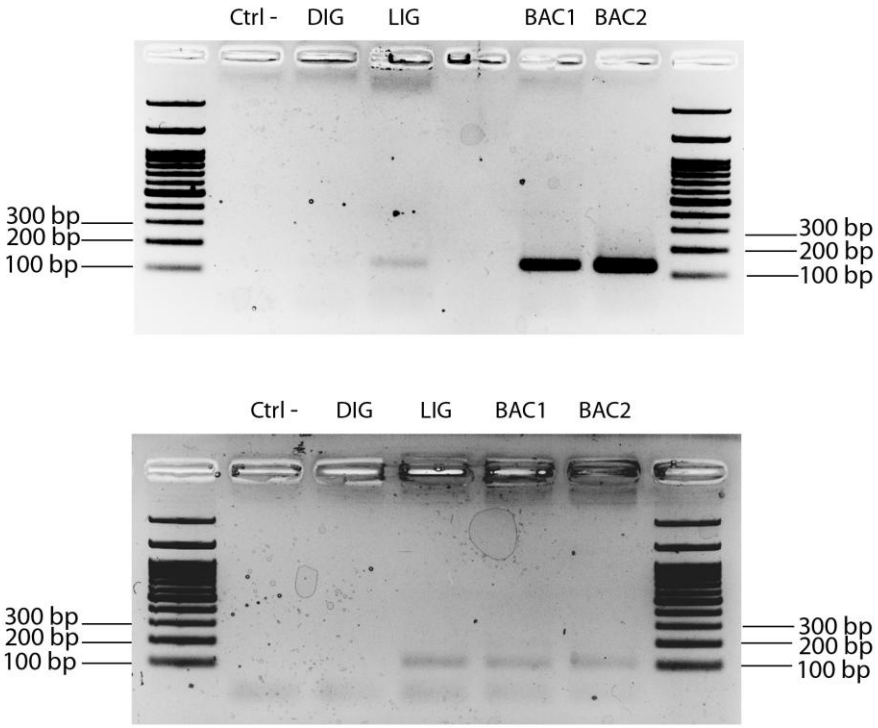


Figure 8: PCR amplification obtained combining specific primers of one of the long-range interactions involving *Socs3*. The PCR products of LIG, BAC1 and BAC2 lanes (agarose gel above) are fragments of 150 bp, as expected. To confirm this result, I designed nested internal primers for the same region: the PCR products of LIG, BAC1 and BAC2 lanes are fragments of 100 bp, as expected (agarose gel below). All controls (Ctrl- and DIG) are negative, as expected. Ctrl- contains all PCR reagents without DNA.

5. REFERENCES

- Cao F, Hata R, Zhu P, Ma YJ, Tanaka J, Hanakawa Y, Hashimoto K, Niinobe M, Yoshikawa K, Sakanaka M. 2006. Overexpression of SOCS3 inhibits astrogliogenesis and promotes m aintenance of neural stem cells. J. Neurochem 98: 459-70.

- Cavallaro M1, Mariani J, Lancini C, Latorre E, Caccia R, Gullo F, Valotta M, DeBiasi S, Spinardi L, Ronchi A, Wanke E, Brunelli S, Favaro R, Ottolenghi S, Nicolis SK. 2008. Impaired generation of mature neurons by neural stem cells from hypomorphic Sox2 mutants. Development. 135: 541-57.

- Dekker J, RippeK, Dekker M, Kleckner N. 2002. Capturing chromosome conformation. Science. 295: 1306-11.

- Gritti A, Parati EA, Cova L, Frolichsthal P, Galli R, Wanke E, Faravelli L, Morassutti DJ, Roisen F, Nickel DD, Vescovi AL. 1996. Multipotential stem cells from the adult mouse brain proliferate and self-renew in response to basic fibroblast growth factor. J. Neurosci. 16: 1091-1100.

- Gritti A, Galli R, Vescovi AL. 2001. Cultures of stem cells of the central nervous system. In Protocols for Neural Cell culture. 3rd edn (ed. S. Fedoroff and A. Richardson), pp. 173-198. Totowa, NJ: Humana Press.

- Schmidt-Edelkraut U, Hoffmann A, Daniel G, Spengler D. 2013. Zac1 regulates astroglial differentiation of neural stem cells through Socs3. Stem Cells. 31: 1621-32.

- Zhang Y, Chen K, Sloan SA, Bennett ML, Scholze AR, O'Keeffe S, Phatnani HP, Guarnieri P, Caneda C, Ruderisch N, Deng S, Liddelow SA, Zhang C, Daneman R, Maniatis T, Barres BA, Wu JQ. 2014.

  An RNAsequencing transcriptome and splicing database of glia, neu rons, and vascular cells of the cerebral cortex. J Neurosci. 34: 11929-4.

CHAPTER 4:

SUMMARY

The *Sox2* gene encodes a transcription factor active in stem/progenitor cells during the development of central nervous system in vertebrates. Lineage-specific transcription factors, as SOX2, establish cell type-specific gene expression by binding to promoters and distal enhancers, which are brought into contact via long-range chromatin interactions. The SOX2 transcription factor is essential for the specification and maintenance of pluripotent stem cells (ES and iPS cells). Heterozygous *Sox2* mutations in humans cause a characteristic spectrum of CNS abnormalities, involving the hippocampus and the eye, and causing epilepsy, learning disabilities and defective motor control [Fantes et al., 2003; Schneider et al., 2009; Ragge et al., 2005; Sisodiya et al., 2006; Kelberman et al., 2006]. Some of its target genes are involved in different central nervous system diseases. In order to understand the role of *Sox2* in neural development, our laboratory generated *Sox2* conditional KO mutations in mouse Favaro et al., 2009; Ferri et al., 2013]. The consequences of *Sox2* ablation at different developmental time points produced important brain defects, more serious when the ablation was early. *Sox2* conditional deletion allowed to observe an important function for *Sox2* also in the maintenance of Neural Stem Cell (NSC) self-renewal in long-term in vitro NSC cultures [Favaro et al., 2009]. Sox2-ablated NSC, cultured as neurospheres from P0 mouse forbrain, self-renewed for several passages in culture, as the wild-type (wt) ones, but then underwent a decrease in growth, with progressive culture exhaustion. Sphere formation could be rescued by lentiviral *Sox2* [Favaro et al., 2009].

This reveled an essential role for *Sox2* in the development of multiple CNS regions and in the maintenance of NSC.

To understand the mechanisms of SOX2 function, a central question is which genes SOX2 regulates as a transcription factor, by what mechanisms SOX2 acts in regulating them, and which SOX2-regulated genes are critical mediators of its function. A new way in which SOX2 regulates its targets has been recently observed in our laboratory: SOX2 maintains a high number of long-range interactions between genes and distal enhancers, that regulate gene expression. We determined, by genome-wide chromatin interaction analysis (RNApolII ChIA-PET, Chromatin Interaction Analysis by Paired-End Tag sequencing, in collaboration with Dr. CL Wei), the global pattern of long-range chromatin interactions in normal and *Sox2*-deleted mouse NSC. Moreover, we analyzed (in collaboration with F. Guillemot, London) the wt NPC by ChIP-seq with anti-SOX2 antibodies, to define a genome-wide map of SOX2 binding sites. In normal NSC, distal regions interacting with promoters were highly enriched in SOX2 bound enhancers.

Sox2 deletion caused extensive loss of long-range interactions and reduced expression of a subset of genes associated with *Sox2*-dependent interactions, which expression was analyzed using RNA-seq technique. Expression of one of these genes, *Suppressor of Cytokine Signaling 3* (*Socs3*), rescued the self-renewal defect of *Sox2*-ablated NSC. Overall, our work identifies SOX2 as a major regulator of functional chromatin connectivity in NSC, and demonstrates the role of genes associated with *Sox2*-dependent interactions in NSC maintenance and, potentially, in neurodevelopmental disorders.

In this work, I studied the differentiation of *Sox2*-ablated cells into neurons and glia, as compared to controls: the neurosphere cultures, derived from

P0 mouse forebrain, were dissociated into single cells and plated in differentiation conditions: at day 11 (advanced differentiation stage), very few β-tubulinIII (neuronal marker) positive cells were observed in *Sox2*-deleted cells differentiated, with poor differentiated morphology, while the majority of these cells showed GFAP (glial marker) positivity. This result showed the importance of *Sox2* in the development into mature neurons.

We also analyzed the changes in gene expression resulting from *Sox2* deletion by RNA-seq analysis of three samples for both wt and *Sox2*-mutant cells in undifferentiated cells, and two differentiation conditions (early stage, day 4, and advanced stage, day 11). Hundreds of genes were deregulated in mutant cells. The most down-regulated gene was *Socs3*, so we transduced *Sox2*-mutant cells with a lentiviral *Socs3*–vector, coexpressing GFP. The untransduced mutant cells grew at a slower rate than wt cells and progressively decreased, as expected, and finally exhausted. *Socs3*-transduced wt cells grew at a similar rate as untransduced cells and continued to grow long-term; *Socs3*-transduced mutant cells initially grew as the untransduced cells (only a proportion of the cells had been tranduced), but continued to grow even after the untransduced mutant cells were completely exhausted, and transduced cells were positively selected. These results suggested that *Socs3* partially rescued the proliferation defect of mutant cells. I also tested if the reintroduction of *Socs3* could rescue the neuronal differentiation defect of mutant cells and my initial experiments suggest that this might be the case: at day 4, untransduced (control) wild-type cells (wt nt) began to show some β-tubulin-positive neurons, exhibiting initial branching morphology, together with a majority of weakly GFAP-positive cells with glial-like morphology, as expected; in contrast, Socs3-transduced cells were all GFAP-negative, and they all appeared β-tubulin-positive. At day 11, wt control cells (wt nt)

showed β-tubulin-positive cells with mature morphology and extensive arborization, as expected. *Socs3*-transduced wt cells showed little more β-tubulin-positive cells compared to the control sample, but the majority of cells were GFAP-positive, similarly to untransduced cells. In contrast, *Socs3*-transduced mutant cells showed no GFAP-positivity at day 4; instead, all cells were β-tubulin-positive, though they seemed to have a suffering morphology.

Then I aimed to test the role of some of the other most deregulated genes as mediators of *Sox2* function in self-renewal and differentiation, by rescuing experiments of mutant cells. I generated lentiviral vectors to express a small group of genes, deregulated following Sox2 loss and known to play important roles in cell proliferation (*Fos, Jun, JunB, Egr1, Egr2*), and I will use them to transduce mutant neural stem cells (in combination, or individually), as I did for *Socs3*. I will evaluate if any specific gene, among those, is able to rescue, upon overexpression, the ability to self-renew and the neuronal differentiation in mutant cells.

Finally, I aim to test if *Sox2* reintroduction in mutant cells could rescue the long-range interactions of a small number of identified target genes, lost in Sox2-deleted cells, by 3C experiments.

CONCLUSIONS AND FUTURE PERSPECTIVES

1. THE IMPORTANCE OF STUDYING SOX2

1.1 SOX2 IS REQUIRED DURING EMBRYONIC DEVELOPMENT

SOX2 belongs to the SOX (Sry-related HMG box) 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 Pevny and Nicolis, 2010]. 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. 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; Pevny et al., 1998]. For this reason, 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. Lox sites 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]. *In vivo* evidence for a role of *Sox2* in NSC maintenance was further confirmed by a progressive and complete loss of *in vitro* NSC renewal in *Sox2*-deleted neurosphere cultures at P0 and P7 and by the rescue of neurosphere formation obtained using *Sox2* lentiviral transduction of mutant neurospheres [Favaro et al., 2009].

1.2. SOX2 AND HUMAN DISEASE

In mouse, the viability doesn't seem to be affected if only one *Sox2* allele is lost. Moreover, heterozygous Sox2 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]. This points to a differential dosage sensitivity to SOX*2* reduction in mice and humans. Indeed, central nervous system abnormalities, similar to the severe ones observed in human patients, had been already identified in the first 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, on the other allele, it carries the deletion of a telencephsalicl-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 levels (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. In addition, 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]; 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 DISEASE

Regulatory elements are often 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]. Thousands of polymorphisms in non-coding elements in man may be linked to brain disease or neurodevelopmental disorders [Doan et al., 2016; Nord et al., 2015]; changes in transcription factor-binding sequences may affect chromatin modifications locally and at distant sites, affecting gene activity over great distances [Denker and de Laat, 2015]. The comparison of the regulatory elements that we identified in mouse with conserved orthologous sequences in man may allow identification of genes regulated by such enhancers, and which might be dysfunctional in individuals carrying mutations at these elements. For this reason, the identification and functional characterization of regulatory sequences are crucial for understanding the spatial and temporal control on gene expression.

We observed (see submitted manuscript, in chapter 2) that neural loss of *Sox2* causes a major decrease of the long-range interactions typical of wt NSC, together with expression changes in a substantial proportion of genes. Decreased gene expression in mut NSC was significantly correlated with the loss of promoter-enhancer (P-nonP) interactions (but not to loss of P-P interactions), and/or to loss of *Sox2* binding to interaction anchors in wt NSC. This demonstrates the relevance of a specific transcription factor (TF), SOX2, in shaping the pattern of chromatin connectivity genome-wide, and its reflection on gene activity. The striking impact on genome-wide interactions, of the loss of just a single TF was not expected.

The only precedent, to our knowledge, for a similar effect is the loss of *Oct4* or *Nanog* in pluripotent ES cells, which was shown to affect long-range interactions [de Wit et al., 2013]. However, loss of these TFs by itself leads to downregulation of the core pluripotency TF network and initiation of abnormal differentiation, hence to drastic changes in the identity of the TF-deleted cells.

Some *Sox2*-dependent long-range interactions, found analyzing the ChIA-PET data, involve regions 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.

Several genes, affected in important human neurodevelopmental disorders [Alcantara and O'Driscoll, 2014; Kim et al., 2012; Lee and Young, 2013;

Peters et al., 2008; Saez et al., 2016; Tan et al., 2014; Williamson and FitzPatrick, 2014] are involved in long-range interactions, many of which are lost in *Sox2*-mutant cells; in particular, a large proportion of genes mutated in primary recessive microcephaly, severe intellectual disability and eye disease have mouse homologs involved in long-range interactions and/or showing a SOX2 peak in their promoter.

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.

2.1. *Socs3* AS SOX2 TARGET GENE

It so far unknown which genes downstream to *Sox2* mediate its function in long-term NSC maintenance. Among the genes most significantly downregulated in *Sox2*-mutant cells is Socs3, a highly connected gene. By overexpressing *Socs3*, we rescued long-term growth of *Sox2*-mutant NSC, providing evidence for a crucial role of *Socs3* in *Sox2*-dependent NSC maintenance. Interestingly, several additional genes (*Fos, Jun, JunB, Egr1* and *Egr2*), encoding well-known regulators of cell proliferation, are expressed at high levels in wt NSC and are substantially downregulated in *Sox2*-mutant cells (Table 1); these genes show multiple promoter-enhancer interactions in wt NSC, that disappear in mutant NSC. These findings may open the way to the identification, by similar studies, of a network of interacting genes, mediating *Sox2* functions in NSC. Hippocampal defects

observed in *Sox2* mutant mice (as in humans) have been related to defects in NSC [Favaro et al., 2009]. The discovery of mediators of *Sox2* function in these cells may be relevant to the understanding of *in vivo* defects.

3. FUTURE PERSPECTIVES AND ONGOING EXPERIMENTS

To test the role of some of the most deregulated genes as mediators of *Sox2* function in self-renewal and differentiation, I'm currently doing rescuing experiments of mutant cells. I generated lentiviral vectors to express a small group of genes (Figure 1), downregulated following *Sox2* loss and known to play important roles in cell proliferation (*Fos, Jun, JunB, Egr1, Egr2*), as described above, and I will use them to transduce mutant neural stem cells (in combination, or individually), as I did for *Socs3*. I will evaluate if any specific gene, among those, is able to rescue, upon overexpression, the ability to self-renew and the neuronal differentiation in mutant cells.

I cloned *Fos, Jun, JunB, Egr1* and *Egr2* cDNAs in the CSIΔNGFR lentiviral vector, using the BamHI restriction site (Figure1).

Finally, I will test if *Sox2* reintroduction in mutant cells could rescue the long-range interactions of a small number of identified target genes, lost in *Sox2*-deleted cells, by 3C experiments.

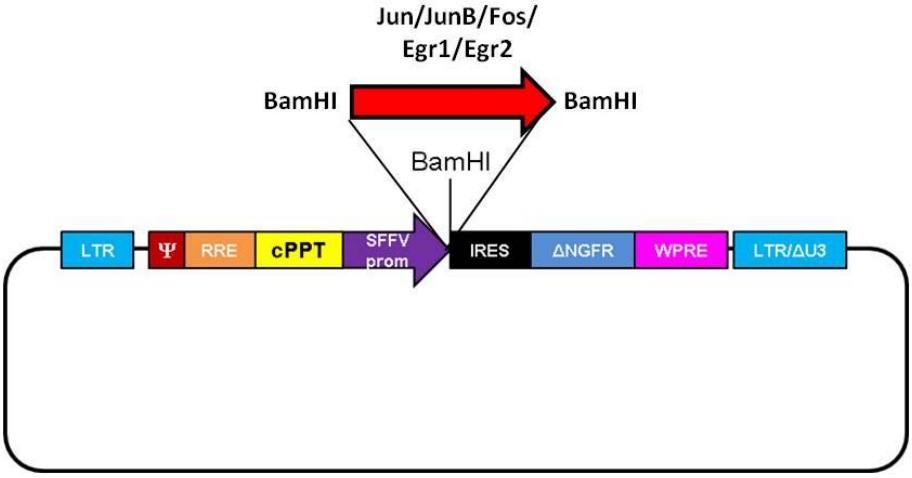


Figure 1: CSINΔNGFR lentiviral vector with cDNA insert (*Jun, JunB, Fos, Egr1* and *Egr2*) cloned using BamHI restriction sites on the CSIΔNGFR vector and at the ends of cDNAs.

Gene	MEAN_WT	MEAN_MUT	FOLD RATIO =log2(WT/MUT)
Socs3	254,43	29,32	3,117312
Fos	2.868,27	818,55	1,809046
Jun	621,48	248,82	1,320613
JunB	359,06	105,83	1,762431
Egr1	1.595,90	493,33	1,693745
Egr2	164,05	82,58	0,9902424

Table 1: the top downregulated genes in Sox2-deleted NSC obtained by RNA-seq analysis. RNA-seq was performed on triplicates for the two genotypes studied, yielding 51 bp single-end reads. The number of sequences obtained in each sample ranged from 7.5 to 12.5 millions. Read counts and transcript levels for each sample were computed with the RSEM software package version 1.17 (Li and Dewey, 2011), on the RefSeq gene annotation available at the UCSC Genome Browser for mouse genome assembly mm9 (24,148 genes). For downstream analyses, expression levels were measured as Transcripts per Million (TPM).

4. REFERENCES

- Alcantara D, O'Driscoll M. 2014. Congenital microcephaly. Am J Med Genet C Semin Med Genet. 166C: 124-39.

- 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.

- 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.

- de Wit E, Bouwman BA, Zhu Y, Klous P, Splinter E, Verstegen MJ, Krijger PH, Festuccia N, Nora EP, Welling M, Heard E, Geijsen N, Poot RA, Chambers I, de Laat W. 2013. The pluripotent genome in three dimensions is shaped around pluripotency factors. Nature. 501: 227-31.

- Denker A1, de Laat W2. 2015. A Long-Distance Chromatin Affair. Cell. 162: 942-3.

- Doan RN, Bae BI, Cubelos B, Chang C, Hossain AA, Al-Saad S, Mukaddes NM, Oner O, Al-Saffar M, Balkhy S, Gascon GG; Homozygosity Mapping Consortium for Autism, Nieto M, Walsh CA. 2016. Mutations in Human Accelerated Regions Disrupt Cognition and Social Behavior. Cell. 167: 341-354

- 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 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.

- 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.

- Li B, Dewey CN. 2011. Rsem: accurate transcript quantification from rna-seq data with or without a reference genome. BMC bioinformatics, 12(1):323.

- Lee TI, Young RA. 2013. Transcriptional regulation and its misregulation in disease. Cell. 152: 1237-51.

- Naumova N, Smith EM, Zhan Y, Dekker J. 2012. Analysis of long-range chromatin interactions using Chromosome Conformation Capture. Methods. 58:192-203.

- Nord AS, Pattabiraman K, Visel A, Rubenstein JL. 2015. Genomic perspectives of transcriptional regulation in forebrain dev elopment. Nueron. 85: 27-47

- Pevny LH, Nicolis SK. 2010. Sox2 roles in neural stem cells. Int J Biochem Cell Biol. 42: 421-4.

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

- Peters JM, Tedeschi A, Schmitz J. 2008. The cohesin complex and its roles in chromosome biology. Genes Dev. 22: 3089-114.

- 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.

- Sáez MA, Fernández-Rodríguez J, Moutinho C, Sanchez-Mut JV, Gomez A, Vidal E, Petazzi P, Szczesna K, Lopez-Serra P, Lucariello M, Lorden P, Delgado-Morales R, de la Caridad OJ, Huertas D, Gelpí JL, Orozco M, López-Doriga A, Milà M, Perez-Jurado LA, Pineda M, Armstrong J, Lázaro C, Esteller M. 2016. Mutations in JMJD1C are involved in Rett syndrome and intellectual disability. Genet Med. 18: 378-85.

- 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.

- Tan WH, Bird LM, Thibert RL, Williams CA. 2014. If not Angelman, what is it? A review of Angelman-like syndromes. Am J Med Genet A. 164A: 975-92.

- Tziaferi V, Kelberman D, Dattani MT. 2008. The role of SOX2 in hypogonadotropic hypogonadism. Sexual Development 2: 194-9.

- Williamson KA, FitzPatrick DR. 2014. The genetic architecture of microphthalmia, anophthalmia and colo boma. Eur J Med Genet. 57: 369-80.

- 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.

- 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.