SCUOLA DI DOTTORATO
UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

Department of

Medicine and Surgery

PhD program Molecolar and Translational Medicine Ciclo / Cycle XXXIII

THESIS TITLE

Sox2-dependent molecular functions in the transcriptional control of glioma and normal neural stem cells

Barone Cristiana

Registration number 758789

Tutor: Prof. Silvia K Nicolis

Coordinator: Prof. Andrea Biondi

ACADEMIC YEAR 2019/2020

A mia sorella.

Table of contents

Chapter 1	Introduction	6
	1. Stem cells and cancer	6
	2. Sox2	8
	2.1 Roles of Sox2 in normal development and differentiation	8
	2.2 Sox2 in cancer	10
	3. Neural tumors and stem cells	12
	4. Gliomas	12
	4.1 PDGF and oligodendroglioma	14
	4.2 Sox2 role in PDGF B-induced murine oligodendroglioma model	14
	4.3 Targeting SOX2 in anticancer therapy	21
	5. Sox2 loss-of-function in human developmental diseases	23
	5.1 Study Sox2 downstream target genes in CNS development	24
	5.2 Long-Range interactions and ChIA-PET technique	24
	6. Scope of the thesis	27
Chapter 2	Sox2-dependent maintenance of mouse oligodendroglioma involves the Sox2-mediated downregulation of Cdkn2b, Ebf1, Zfp423 and Hey2	38
Chapter 3	Sox2 Functions in Neural Cancer Stem Cells: The Importance of the Context	90
Chapter 4	Mapping the Global Chromatin Connectivity Network for Sox2 Function in Neural Stem Cell Maintenance	116
Chapter 5	Sox2 controls neural stem cell self-renewal through a Fos-centered gene regulatory network	192

Chapter 6 1. Summary 235
1.1 Sox2 and its downstream target genes relevance in CSCs maintenance. 235
1.2 Sox2 and its downstream target genes relevance in NSCs maintenance 238
2. Conclusions 241
3. Future perspectives 243
3.1 Sox2 interactome in NSCs and CSCs 243
3.2 ChIA-PET on pHGG cells 244
3.3 miRNA deregulation following Sox2 loss 246
3.4 Towards the definition of therapeutic targets 250

Chapter 1

Introduction

1. Stem cells and cancer

Stem cells are defined as "cells that have the ability to perpetuate themselves through self-renewal and to generate mature cells of a particular tissue through differentiation" (Reya, Morrison, Clarke, & Weissman, 2001). The stem cell niche is the surrounding molecular and cellular environment, that strictly regulates stem cells proliferation and differentiation. Stem cells constitute the basis of development and, in adult life, are responsible for tissue homeostasis and repair. Tissues are constituted of differentiated cells, which permit the functioning of the tissue and the organs. This population is sustained by stem cells. Under precise stimuli, stem cells, usually quiescent, re-enter the cell cycle and undergo an asymmetric division: each stem cell gives rise to a stem cell (self-renewal) and a more differentiated cell (differentiation), called progenitor. The progenitor is still an undifferentiated cell; however, it has limited self-renewal ability and it has lost a certain degree of potency. The progenitors duplicate several times and give rise to terminally differentiated cells. In this way, only the progenitor compartment undergoes massive proliferation, avoiding replication-associated accumulation of mutations in the stem cell compartment. Under different conditions and in response to stress, stem cells can also undergo symmetric divisions; in this case, a stem cell can give rise to two stem cells or two progenitors.

In 1997 it was proposed for the first time by Bonnet et al (Bonnet & Dick, 1997) that tumors could have the same hierarchical organization we find in normal tissues. In fact tumors can be described as aberrant organs originated from a Tumor-Initiating Cell (TIC) also called Cancer Stem Cell (CSC) (Reya et al., 2001). In

tumors we can find heterogeneous combinations of cells with different phenotypic characteristics and different proliferative potentials (Reya et al., 2001)(Kreso et al, 2014). Within the populations of tumor cells there is a sub-population of cells called CSCs, that share similar characteristics with normal stem cells. CSCs have long term ability to self-renew and are able to generate a more differentiated, short-term proliferating and non-tumorigenic progeny that will compose the bulk of the tumor (Reya et al., 2001; Shackleton, Quintana, Fearon, & Morrison, 2009). Most of the tumor bulk cells are typically proliferating while CSC can be slowly proliferating and even "quiescent". Because of this characteristic, CSC are typically resistant to chemotherapy drugs (Barone et al., 2018). This observation leads to the hypothesis that tumor cells may utilize the machinery physiologically used by stem cells to self-renew. However there are some differences between CSCs and normal stem cells such as increased proliferative kinetics (Reya et al., 2001), independency from environmental signals (L. Li & Neaves, 2006) and chromosomal aberrations.

The Cancer Stem Cells (CSCs) hypothesis is thus redefining the way we look at tumors and, as a consequence, it is leading to new therapeutic approaches in the war on cancer. Since CSC are resistant to classical therapeutic approaches, one of the focuses of oncological research is finding which are the underlying genes that give cancer cells "stemness" properties.

Targeting cancer stem cells could represent a major advancement in cancer treatment. The ablation of CSCs would make the tumor incapable of generating new cells. After the complete loss of the stem cell compartment, the tumor would probably go on growing for a short period, sustained by the last divisions of non-stem neoplastic cells; however, these cells would start entering replicative senescence or, in any case, they would exhaust their replicative potential (Reya et al., 2001). This would lead to the spontaneous degeneration of the tumor.

One way to approach the understanding of tumorigenesis from this new perspective is to focus on factors essential for normal stem cells, that could have a key role also in maintaining CSCs. Many studies are defining Sox2 as a stem factor also involved to tumorigenesis and survival of CSCs.

2. Sox2

Sox2 belongs to the Sox (SRY-related HMG box) gene family that encodes transcription factors characterized by a single HMG DNA binding domain. Sox2, with Sox1 and Sox3 forms the B1 subgroup (Kamachi, Uchikawa, & Kondoh, 2000). Sox2 is required for stem cell maintenance, functionality and differentiation during both development and adult life in the nervous system; moreover, it has been recently demonstrated that Sox2 is expressed in a wide variety of tumors and can correlates with high malignancy and poor prognosis.

2.1 Roles of Sox2 in normal development and differentiation

Sox2 is expressed, in mouse, both during development and during adult life, in different types of highly undifferentiated cells (Ferri et al., 2004)In earlier development, Sox2 is expressed by pluripotent embryonic stem cells of the inner cell mass (Ferri et al., 2004). Later in development, Sox2 expression is restricted to the developing nervous system, mainly in the most undifferentiated precursors (Ferri et al., 2004), but also in some differentiated nervous cells (Mercurio, Serra, & Nicolis, 2019) present in restricted stem niches i.e. the subventricular zone (SVZ) of the forebrain and the sub-granular zone (SGZ) in the dentate gyrus of hippocampus. Sox2 positive cells, isolated from both the developing nervous system and the adult neurogenic regions and cultured under appropriate conditions, are able to grow as neurospheres (Ferri et al., 2004). These cells are able to self-

renew indefinitely in culture, and they are able to differentiate to all neural lineages (neurons, astrocytes and oligodendrocytes) (Ferri et al., 2004).

The study of Sox2 role in the formation of Central Nervous System (CSC) has been addressed mainly by knock-out experiments. The generation of Sox2-null mutant mice showed that Sox2 is strictly required for the maintenance of the pluripotent stem cells of the epiblast (Avilion et al., 2003); Sox2 knock-out impairs the formation of the blastocyst, inducing cells from the inner cell mass to differentiate into trophoblastic cells. The Sox2 knock-out is therefore embryonic lethal (Avilion et al., 2003). On these basis it was then demonstrated that Sox2 acts together with Oct4, Nanog and N-myc to maintain pluripotency in embryonic stem cells; these four genes are able to induce pluripotency in terminally differentiated cells (Takahashi & Yamanaka, 2006) reprogramming them to induced pluripotent stem cells (iPSCs). Given the embryonic lethality of the Sox2 complete knock-out, Sox2 roles in later development and in stem cell physiology have been investigated using a conditional knock-out mouse model.

Sox2 conditional knock-out at different embryonal stages and restricted to different CSC regions shows that Sox2 is necessary for neural stem cell maintenance in vitro (in neurosphere cultures) and, *in vivo,* for the correct development of different brain structures such as in the dentate gyrus of the hippocampus (Favaro et al., 2009), the ventral telencephalon (Ferri et al., 2004), retina-talamo-cortical connections involved in the visual system (Mercurio, Serra, Motta, et al., 2019). Sox2 knock-out on neurosphere cultures induces complete exhaustion of the culture after a few passages (Favaro et al., 2009). These data indicate Sox2 as a fundamental gene for NSCs maintenance and physiology; its key role in reprogramming and dedifferentiation has attracted attention on its potential role in tumorigenesis, notably in CSCs.

2.2 Sox2 in cancer

The discovery of CSCs in tumors focused attention on Sox2 also from the point of view of this pathological stem cell type. Indeed, Sox2 is expressed in different tumors of neural origin, such as gliomas (the most common primary brain tumors, whose most malignant and lethal subtype is glioblastoma multiforme), medulloblastomas (the most common brain tumor in childhood), and melanoma (a tumor arising from neural crest type cells); in all of these tumors, CSC have been identified, and found to express Sox2 (Garros-Regulez et al., 2016; Nicolis, 2007; Vanner et al., 2014).

Many groups are studying the functional role of Sox2 in different types of tumors. Interestingly, although it is expressed in many types of tumors, Sox2 is essential only for some of them. Regarding neural tumors, our laboratory and others proved that Sox2 is essential in maintenance of glioma stem cells both in mouse models and in human patient samples (Favaro et al., 2014; Gangemi et al., 2009). Importantly, conditional knockout and RNA interference experiments showed that SOX2 controls CSS functions also in some non-neural tumors, such as in skin, lung, and esophagus squamous cell carcinomas (SCC), osteosarcomas, Ewing sarcoma, and small cell lung carcinomas (SCLC)(Bass et al., 2009; Basu-Roy et al., 2012; Boumahdi et al., 2014; Riggi et al., 2010; Rudin et al., 2012). Surprisingly Sox2 is dispensable, although expressed, in two models of tumor with a neural origin: mouse melanoma and SMOM2-induced medulloblastoma(Schaefer et al., 2017; Schuller et al., 2008).

Sox2 is essential for the maintenance of CSC and tumorigenesis in some neural tumor types, while being dispensable in others. It is possible that different stem cell programs control the maintenance of, for example, glioma versus melanoma stem

cells, and that only the first one requires Sox2. Alternatively, it is possible that, at least in some cases, Sox2 acts redundantly with other Sox factors, co-expressed with it in some tumor types (i.e. Sox3 in MB (Xie et al., 1998). On the other hand, we noted that a requirement for Sox2 is found not only in neural CSC within gliomas, but also in very different, non-neural tumors, such as skin and esophagus SCC, lung SCLC, and osteosarcomas. Perhaps, although they differ by histology and by cell of origin, these tumors share a "core", Sox2-controlled gene regulatory network, active in their CSC. Thus, it will be important to comparatively characterize the gene regulatory networks controlled by Sox2 in these CSCs.

These findings have implications for therapy approaches. On one hand, they suggest it might be advantageous to "classify" tumors according to the gene regulatory networks that function in the maintenance of their CSC, that in turn might involve shared efficacy of CSC-targeting drugs. On the other hand, they emphasize the need for functional experiments, to address the importance of specific gene products (here, Sox2, and its downstream targets), to distinguish driver from bystander roles, in order to appropriately target future CSC-directed therapy approaches.

In some case Sox2 dysregulation can be a lesion causative of the tumor. Numerous studies have shown an amplification of the SOX2 gene locus and an increased SOX2 expression in different cancer types such as in esophagus and lung squamous cells carcinoma (Bass et al., 2009), breast cancer (Chen et al., 2008), colorectal cancer (Long & Hornick, 2009), gliomas (Annovazzi, Mellai, Caldera, Valente, & Schiffer, 2011) and others (Novak et al., 2019).

3. Neural tumors and stem cells

Normal NSCs are found in different areas of the brain, mainly the subventricular zone and the dentate gyrus of the hippocampus (Hadjipanayis & Van Meir, 2009). Neural stem cells are defined functionally: they are able to self-renew and to give rise, by differentiation, to neurons, astroglia and oligodendroglia within the clonal progeny of a single cell (Pevny & Nicolis, 2010). When grown in non-differentiating conditions, they proliferate extensively.

CSCs are able to self-renew, to proliferate indefinitely and to differentiate. Multipotency is not necessary, given that some tumors show just a single differentiation lineage (Hadjipanayis & Van Meir, 2009). Although CSCs show many similarities with NSCs, there are also significant differences between them, and, notably, between CSCs isolated from different tumors. NSCs are extremely rare in the nervous system (< 1% of total cell population), while CSCs can represent 1-25% of the overall tumoral mass (Piccirillo et al., 2006), and they are extremely dependent on the context. Moreover, cancer stem cells can resemble neural stem cells in different proportions, depending on the tumor and, notably, on the cell of origin of that class of cancer (Hadjipanayis & Van Meir, 2009).

4. Gliomas

Brain tumours, such as gliomas, can arise from stem, progenitor and/or more mature cells (Azzarelli, Simons, & Philpott, 2018). Gliomas are classified according to:

- histology and cellular markers: astrocytomas, oligodendrogliomas, oligoastrocytomas and ependymomas.

- aggressiveness: Low-grade and High-grade.

Genetic and transcriptomic expression studies have allowed a more detailed molecular classification identifying four GBM subtypes: (i) classical, with EGFR amplification and overexpression, CDKN2A and PTEN deletion, NES overexpression and activation of NOTCH and SHH signaling pathways; (ii) mesenchymal, with loss of NF1, TP53, and PTEN, overexpression of MET, CHI3L1, CD44, and MERTK, and activation of the TNF and NF-kB pathways; (iii) proneural, with PDGFR amplification, loss or mutation of IDH1, PIK3K, TP53, CDKN2A, and PTEN, and activation of HIF, PI3K, and PDGFR pathways; and (iv) neural, with EGFR amplification and overexpression, and expression of neuronal markers, such as NEFL, SYT1, and/or GABRA1(Azzarelli, Simons, & Philpott, 2018).

Among High-Grade astrocytomas we can find Glioblastomas, which are the most common and malignant human brain tumor (Louis, Holland, & Cairncross, 2001). CSCs have been isolated and characterized in human glioblastomas (Barone, Pagin et al. 2018) and it was demonstrated by Gangemi et al. (Gangemi et al., 2009) that SOX2 has a central role in controlling proliferation of human glioblastoma CSCs both in vivo and in vitro, as it happens in normal murine neural stem cells. SOX2 silencing in human glioblastoma CSC leads to loss of tumorigenicity. These results suggest that targeting SOX2 or its downstream genes could be effective in therapeutical approach. To test this hypothesis is useful to dispose of murine models of this disease.

4.1. PDGF and oligodendroglioma

The platelet-derived growth factor receptor (PDGFR) is expressed by neuroepithelial cells at E8.5 of mouse embryo development (Andrae, Hansson, Afink, & Nister, 2001). Later in development, PDGF is known to function as oligodendrocyte precursor cells mitogen. PDGF is important for the regulation of oligodendrocyte precursor number and for oligodendrocyte production in vivo. While most oligodendrocyte precursors go into mature oligodendrocyte differentiation early in postnatal life, in adult brain remains a slowly dividing population of oligodendrocyte precursors (Ffrench-Constant, Miller, Kruse, Schachner, & Raff, 1986; Wolswijk & Noble, 1989).

Apart from its developmental roles, PDGF signaling has been associated to the formation of brain tumors. In most cases of oligodendrogliomas and astrocytomas an activation of the pathway occurs (Guha, Dashner, Black, Wagner, & Stiles, 1995; Varela et al., 2004). This activation has been observed both in low- and high-grade tumors, with similar frequency, suggesting that the overexpression of PDGF/PDGFR pathway may be important for tumor initiation.

Alteration of PDGF-B signaling is commonly observed in human gliomas of different histopathological grades PDGF-B signaling is usually altered, and previous studies demonstrate the ability of PDGF-B to induce gliomas in mouse perinatal and adult neural stem cells and progenitors (Di Rocco, Carroll, Zhang, & Black, 1998; Hermanson et al., 1992; Nister et al., 1988; Varela et al., 2004).

4.2 Sox2 role in PDGF B-induced murine oligodendroglioma model

Appolloni *et al*. generated a model of murine oligodendroglioma by overexpression of PDGF-B in neural precursors of mouse embryos (E14.5), by retroviral

transduction of embryonic brains in utero (Appolloni et al., 2009); the tumors that have developed (PDGF-induced high-grade gliomas, pHGG) express Sox2 (Favaro et al., 2014). pHGGs contain tumor stem cells that will reform the same tumor type after in vivo transplantation of dissociated tumor tissue or pHGG cells maintained in culture (Calzolari et al., 2008). It has been shown that Sox2 is required for the maintenance of oligodendroglioma stem cells by combining a conditional Sox2 flox/flox mutation in the mouse glioma model (Favaro et al., 2014). If Sox2 is deleted from the cells obtained from this glioma model, they lose the ability to form tumors after transplantation into the brain of mice, and, in culture, the deletion of Sox2 causes a decrease in cell growth, an increase in cell death and a "Differentiation" involving the activation of the expression of markers of oligodendrocytes and differentiated astrocytes (see Figure 1; Figure 2) (Favaro et al., 2014).

The authors investigated genes differentially expressed in Sox2-deleted pHGG cells, as compared to Sox2 Wild-Type (WT) pHGG cells, performing a microarray analysis. These genes include both up-regulated (134) and down-regulated (12) genes following the loss of Sox2. Of the genes upregulated following the deletion of Sox2, many are already known as tumor suppressors in different types of tumors. On the contrary, among the the genes downregulated following the loss of Sox2, some are known to have pro-tumorigenic characteristics (Favaro et al., 2014) (Table 1).

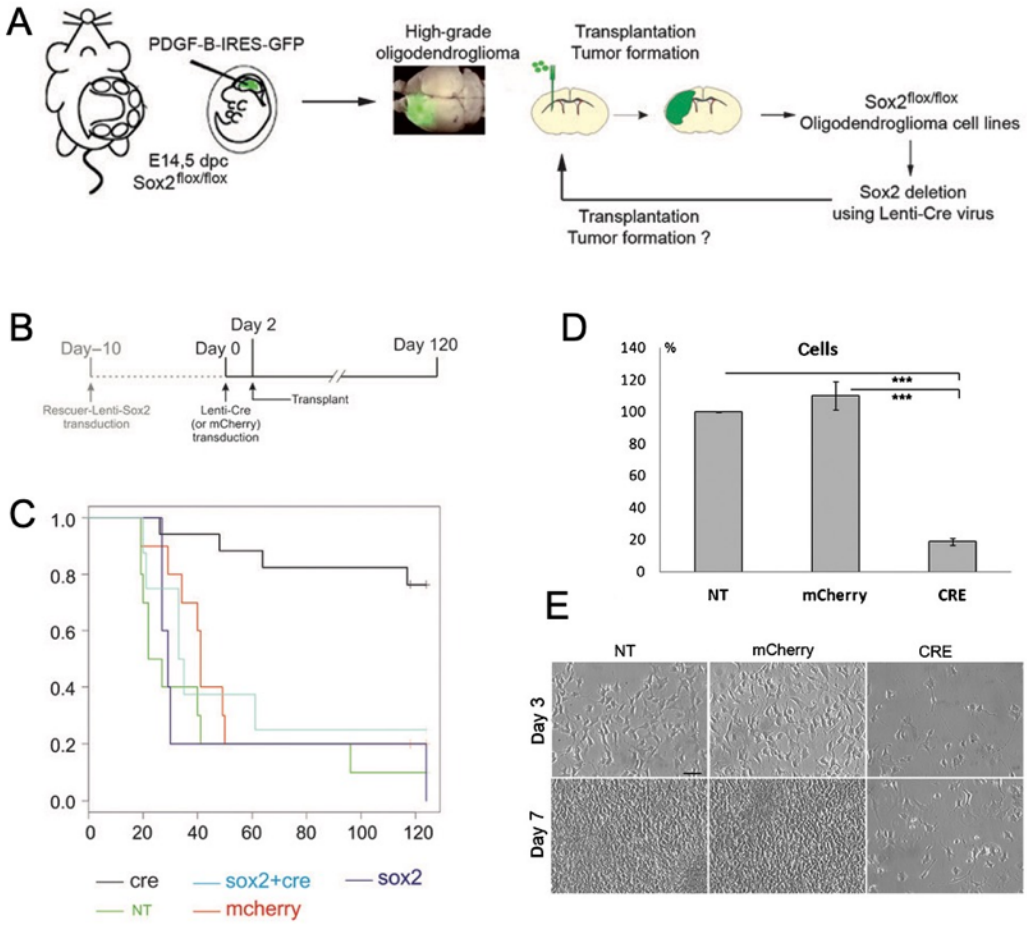


Figure1. Sox2 is required for the maintenance of cancer stem cells in a High Grade Glioma (pHHG) mouse model. A, Schematic representation of the experimental procedure to obtain and study the Sox2 flox/flox oligodendroglioma. B, Diagram of viral transduction and transplantation experiments in the brain of mice; C, Survival curves (Kaplan – Meier) of mice transplanted with untreated pHGG cells (NT, green line), Lenti-Cre transduced pHGG cells (Sox2 delete, black line) transduced with a control virus (mCherry, red line), transduced with a virus that allows the 'rescuing' of Sox2 + LentiCre (blue line) and finally pHGG cells transduced only with the virus for the 'rescuing' of Sox2 (blue line). D, pHGG cell

numbers, obtained with non-transduced cells (NT) or following transduction with control (mCherry) or Cre virus (CRE), after 7 days in serum-containing medium. The cell number obtained with NT cells (>1,000 cells counted for each experiment) is set at 100%. E, Representative images showing cell density at day 3 and 7 are shown. (Favaro et al., 2014)

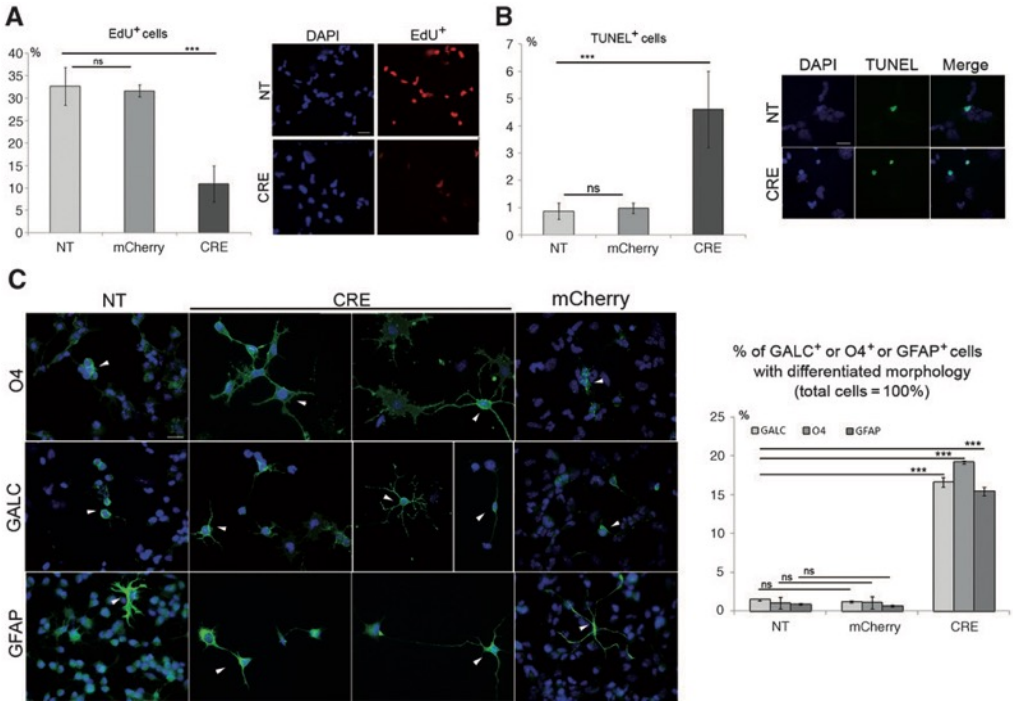


Figure2. In vitro effects of Sox2 flox/flox deletion on pHGG cells. A, EdU incorporation obtained following transduction with control (mCherry) or Cre virus after 2 days in serum-containing medium. Histograms report the percentage of EdU-positive nuclei over the total number of (DAPI-positive) nuclei. Representative images showing EdU-positive cells are shown besides the histograms (scale bar, 20 mm). B, TUNEL analysis of cells 2 days after transduction

with Cre or control mCherry virus. Representative images are shown, with TUNEL-positive nuclei in green (scale bar, 20 mm). C, immunofluorescence with antibodies against O4, GalC, or GFAP (green) of cells transduced with control mCherry or Cre virus, or untransduced (NT), after 7 days in culture in serum- containing medium (scale bar, 20 mm). Arrowheads, some examples of cells with "differentiated" morphology. (Favaro et al., 2014)

Table1

Gene (up/down-regulated after Sox2 loss)	Gene product function	Associated tumor type	Known function in cancer	Sox2 peaks in NPC	Ref.
Hopx (10x up)	TF	Lung tumor	Tumor suppressor in squamous lung cancer	Yes	Cancer Cell 2013, 23:725
Sdc4 (7.5x up)	Cell adhesion (Cell surface proteoglycan, co-receptor for integrins)	Expressed in gliomas and other tumors Induced by hypoxia	Inhibits K-ras-dependent cell invasion into collagen; Modulates cell adhesion, binding to fibronectin via FNIII repeat (competed byTenascin)	Yes	Matrix Biol. 2011, 30:207
Rgs2 (7x up)	Ras signalling Quiescence-induced regulator of Wnt signal	Astrocytoma cells	downregulated in many solid tumors	Yes	BBA 2001, 1541:201
Ebf1 (6.8X up)	TF; interactor of TET2 (DNA demethylase) in glioma and leukemia	Glioma; Leukemia (ALL), pancreas tumors	Tumor suppressor Regulates ATF5 promoter	Yes	Nat. Commun. 2013, 4:2166
Wif1 (6.75x up)	Signalling molecule, secreted Wnt antagonist (antagonizes beta-cat/TCF/LEF activation)	Lost/downreg. in most (75%) glioblastomas; hypermethyl. in astrocytomas; colorectal cancer non small cell lung cancer et al	Tumor suppressor GBM in vitro: inhibits cell prolif., anchorage-indep. growth in soft agar; abolishes tumorigenicity in vivo. Overexpr. induces senescence. Repressed by c-myc and methylation. Note: Wnt/bcat signaling is downstream of met	Yes	J Clin Neurosci 2012, 19:1428 Neurooncol 2011, 13:736
Irf6 (6.7x up)	TF interacts with tumor suppressor maspin	Breast cancer; Squamous cell carcinoma	Tumor suppressor in breast cancer; in keratinocytes Mediator of quiescence/differentiation?	Yes	Mol Cell Biol 2008, 28:2235
Pcp4 (PEP-19) (6x up)	Promotes neuronal differentiation in normal NSC. Signalling modulator	None found		Yes	J Comp Neurol 2011, 519:2279
Cntnap2 (CASPR2) (5x up)	Contactin-associated protein-like 2; cell adhesion molecule	Glioma, oligodendroglioma	Tumor suppressor in glioma: Mutated in gliomas/oligodendrogliomas; overexpression causes decreased proliferation, apoptosis	Yes	Oncogene 2010, 29:6138
Tnc (4.2x up)	Extracell. glycoprotein, adhesion-modulatory ECM	Highly expressed in many tumors, incl. gliomas	Invasion Antigen for immunotherapy.	Yes	Can Res 2001, 61:8586

	molecule				
Cryab (4.8x up)	Crystallyn alphaB Cell adhesion	Nasopharyngeal carc.	Tumor suppressor in nasopharyngeal carcinoma; Associates w Betacatenin/E-cadherin Deleted in breast cancer methastases	Yes	Oncogene 2012, 31:3709
Hey2 (4x up)	TF; effector in Notch signalling	Glioma (incl. OG), GBM (downregulated); Gastric cancer		Yes	Oncogene 2011, 30:3454
Trp53inp1 (3.8x up)	Nuclear protein; transcr. activated by p53 and p73		Overexpression induces cell cycle arrest and cell death in pancreas and lung cells	Yes	Oncogene 2005, 24:8093
Tgfb2 (3.8x up)	Transforming GF Acts via SMAD TFs	High-grade glioma	Pro-tumorigenic in glioma; Antisense oligonucleotide, blocking TGFb2, antagonizes glioma. Antagonizes T cell infiltration	Yes	Curr Pharm Biotechnol 2011, 12:2150
Cdkn2b (3.66x up)	Cyclin-dependent kynase inhibitor Downstream to PDGFR-ras;upstream to Rb	CDKN2B deleted in 50-70% of gliomas (adjacent to CDKN2A/p16Ink4a and p14Arf)	Tumor suppressor in GBM. In GBM cells, Cdkn2b overexpression in Cdkn2a-deficient cells inhibits cell growth and telomerase, induces senescence.	No	J Neurooncol 2010, 96:169
Myo5B (4x down)	dimeric nonfilamentous myosin. Membrane recycling regulator, with Rab GTPases	Downregulated in gastric and bladder cancer:	Inactivation can promote proliferation, invasion, motility of gastric cancer cells. Functional interactor of c-met (inhibits HGF-stimulated met degradation)	No	Dig Dis Sci 2013, 58:2038
Prss23 (3x down)	Serin protease	Breast cancer	Knockdown may suppress estrogen-driven cell proliferation of breast tumor cells in vitro	Yes	PloS One 2012, 7:e30397
Pkib (2.9x down)	Protein kinase inhibitor	Prostate cancer, glioma	Knockdown causes growth suppression in prostate cancer; poorly expr.in normal tissue	Yes	Oncogene 2009, 28:2849
Ptger3 (EP3) (2.43x down)	Prostaglandin E receptor, subt. 3; Positive regulator of angiogenesis	Hematopoietic and breast cancer, lung adenocarcinoma lines	KO in BM cells suppresses tumor growth and angiogenesis of wt cells (transplant of BMC)	No	BBRC 2009, 382:720
Tfrc (2.4x down)	Transferrin receptor	Upregulated in many tumors,	Enhances cellular proliferation and tumorigenesis; RNA interference causes growth arrest	Yes	J Membr Biol 2014, Feb27
Mmp17 (2.12x down)	Membrane-anchored matrix Metalloproteinase	Expressed in glioma		Yes	Canc. met. Rev 2008, 27:289
Rbl1(p107) (2x down)	TF	melanoma	Tumor suppressor; corepressor of Sox2, with E2f3a (feedback?)	Yes	JBC 2013, 288:24809; Cell StemCell 2013,12:440

4.3 Targeting SOX2 in anticancer therapy

Because of its more and more clear role in many types of tumors, many studies propose Sox2 as a pharmacological target in cancer treatment. Here I list several studies that tried to target Sox2 directly or through its regulative pathways.

In glioma, Ikushima et al. (Ikushima et al., 2009) demonstrated that TGF- beta signaling is essential for the stemness of glioma-initiating cells. The authors also showed that TGF-beta exerts its effect via its direct target SOX4. SOX4, in turn, induces the expression of SOX2 resulting in highly tumorigenic and dedifferentiated glioma-initiating cells. Moreover, it was shown that targeting the TGF-beta/SOX4/SOX2 pathway by using TGF-beta inhibitors leads to a decreased lethal potency in intracranial transplantation assay. Hence, the authors suggested that disrupting the TGF-beta/SOX4/SOX2 pathway could be a therapeutic option for the treatment of glioma.

SOX2 was proposed as a target for the treatment of prostate cancer. The simultaneous inhibition of SOX2 and HH signaling pathways leads to reduced proliferation and migration and a drastic increase in the percentage of apoptotic cells, suggesting that a combinatorial therapy might be the most effective strategy (Kar, Sengupta, Deb, Pradhan, & Patra, 2017).

Moreover, since it was demonstrated that SOX2 is involved in the regulation of cancer features like migration and invasion in breast cancer by the SOX2/miR-181a-5p, miR-30e-5p/TUSC3 axis, it was suggested to target this axis (Liu et al., 2017). Zhao et al. (Zhao, Li, Zhang, Yang, & Chang, 2015) showed that expression of miR-126 is low in hepatocellular carcinoma and that a restoration of miR-126 expression results in apoptosis and inhibition of proliferation and cell cycle

progression. They also demonstrated that SOX2 was a target of miR-126 and that SOX2 overexpression could partially reverse the effects of miR-126. It was concluded that miR-126 might serve as a therapeutic target since it acts as tumor suppressor partially by targeting SOX2. Active hexose-correlated compound (AHCC) was found to target specifically SOX2 in pancreatic cancer. It was proposed that AHCC might be a candidate for combinatorial anticancer therapy (Nawata et al., 2014).

A study on lung and esophageal squamous cell carcinoma used an interesting technique to target SOX2. The authors established a zinc finger-based artificial transcription factor, which allowed them to selectively suppress SOX2 in cancer cells while the viability of normal human cells was not influenced. Hence, the authors suggested using this artificial transcription factor for anticancer targeted therapy in lung and esophageal squamous cell carcinoma (Yokota et al., 2017). The same technique was also shown in breast cancer to target SOX2. The use of zinc finger-based artificial transcription factor targeting SOX2 results in decreased cancer cell proliferation and colony formation in vitro and inhibited breast cancer cell growth in vivo. The authors suggested using this technique in cancer therapy to obtain a long-lasting downregulation of oncogenic transcription factors (Stolzenburg et al., 2012). A novel approach to target SOX2 was presented by Favaro et al. (Favaro et al., 2014). The authors tested SOX2 as a therapeutic target by using SOX2 peptides for mouse immunization. They found that the immunized mice display a delayed tumor onset and prolonged survival (Favaro et al., 2014). In an earlier study, Schmitz et al. (Schmitz et al., 2007) suggested to use SOX2 as a glioma-specific antigen for T cell immunotherapy of glioma patients since SOX2 is overexpressed in tumor tissue samples and in glioma cells. They showed that

SOX2-derived peptides are able to activate cytotoxic T lymphocytes which could lyse glioma cells (Schmitz et al., 2007).
Many publications suggested not to target SOX2 directly but instead to target pathways, which regulate SOX2 or which are regulated by SOX2. My main PhD project fits in this context. In fact I identified a tumor-suppressive gene network downstream of Sox2 in the PDGF-induced murine model of oligodendroglioma (described in Paragraph 4.2).

5. Sox2 loss-of-function in human developmental diseases

As discussed in Paragraph 2.1 Sox2 has a key role in normal CNS development. The functional role of Sox2 in CNS development was addressed by Nicolis group, characterizing several mouse conditional KO mutants, in which the Sox2 deletion affects CNS in a time- and region-specific way. Sox2 conditional KO results in developmental defect of different CNS structures (as discussed in Paragraph 2.1). What we saw in mouse mirrors SOX2-dipendent defects observed in human patients. 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, Bardakjian, Reis, Tyler, & Semina, 2009) and defects in hippocampus, with neurological pathology including epilepsy, motor control problems and learning disabilities (Kelberman et al., 2006; Ragge et al., 2005; Sisodiya 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 hypothalamic-pituitary disorders (Tziaferi, Kelberman, & Dattani, 2008).

5.1 Study Sox2 downstream target genes in CNS development

As previously discussed for Sox2 function in tumorigenesis, also in the case of normal CNS development it is important to define downstream Sox2 target genes, that could be key effectors of Sox2 role.
Our laboratory focused on this aim using three different genome-wide methods on mouse Neural Stem Cells (NSCs), grown in vitro from postnatal day 0 (P0) mouse brains:

- RNA sequencing (RNAseq) of Sox2-deleted vs Sox2-WT Neural Stem Cells (NSCs).
- Chromatin Immuno Precipitation of SOX2 in Sox2-WT NSCs (ChIPseq).
- Chromatin Interaction Analysis by Paired-End Tag Sequencing (ChIA-PET, see Paragraph 5.2) of Sox2-deleted vs Sox2-WT Neural Stem Cells (NSCs).

5.2 Long-Range interactions and the 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 spatial 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 (Cheutin & Cavalli, 2014; Fullwood et al., 2009; G. Li et al., 2012; Zhang et al., 2013).

A new approach has been developed 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. 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, these connections showed a large number of putative enhancers located in these inter- and intragenic regions. In many of them it has been possible to identify 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 identified, a subset 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 positively regulates 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 (G. Li et al., 2012).

In the Chapter 4 I will discuss how Sox2 is involved in the formation of chromatinic loop, and how, in this way, it regulates the transcription of its downstream target genes.

6. Scope of the thesis

During my PhD thesis work I studied the molecular functions of the transcription factor Sox2 molecular functions in the transcriptional control of glioma and normal neural stem cells.

My main project (Chapter 2) has been the study of the role of Sox2 in glioma formation. I found that Sox2 plays a key role as "stem factor" in glioma stem cells by acting as a transcriptional repressor of a tumor suppressive gene network (Barone *et al*., 2020).

The role of Sox2 in different kinds of tumors will be reviewed in Chapter 3 (Barone *et al*., 2018).

In addition, I collaborated with my colleagues in other two projects aimed at dissecting the molecular functions of Sox2 in the maintenance of normal neural stem cells (NSCs), and therefore in embryonal development of the CNS (Chapters 4 and 5).

In particular, in Chapter 4, I describe how Sox2 acts as transcriptional regulator of its target genes in neural stem cells, through mediating the formation of chromatinic long-range interactions between enhancers and promoters of its downstream target genes (Bertolini *et al*. 2019).

In Chapter 5, I will go more in depth in the study of Sox2 downstream target genes in NSCs. Indeed, through functional experiments of gain-of-function and loss-of-function, we (Pagin *et al*.) found that Sox2 regulates a gene network involved in NSC proliferation and long-term maintenance.

References

Andrae, J., Hansson, I., Afink, G. B., & Nister, M. (2001). Platelet-derived growth factor receptor-alpha in ventricular zone cells and in developing neurons. *Mol Cell Neurosci, 17*(6), 1001-1013. doi:10.1006/mcne.2001.0989

Annovazzi, L., Mellai, M., Caldera, V., Valente, G., & Schiffer, D. (2011). SOX2 expression and amplification in gliomas and glioma cell lines. *Cancer Genomics Proteomics, 8*(3), 139-147. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/21518820

Appolloni, I., Calzolari, F., Tutucci, E., Caviglia, S., Terrile, M., Corte, G., & Malatesta, P. (2009). PDGF-B induces a homogeneous class of oligodendrogliomas from embryonic neural progenitors. *International journal of cancer. Journal international du cancer, 124*(10), 2251-2259. doi:10.1002/ijc.24206

Avilion, A. A., Nicolis, S. K., Pevny, L. H., Perez, L., Vivian, N., & Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes & development, 17*(1), 126-140. doi:10.1101/gad.224503

Azzarelli, R., Simons, B. D., & Philpott, A. (2018). The developmental origin of brain tumours: a cellular and molecular framework. *Development, 145*(10). doi:10.1242/dev.162693

Bass, A. J., Watanabe, H., Mermel, C. H., Yu, S., Perner, S., Verhaak, R. G., . . . Meyerson, M. (2009). SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. *Nature genetics, 41*(11), 1238-1242. doi:10.1038/ng.465

Basu-Roy, U., Seo, E., Ramanathapuram, L., Rapp, T. B., Perry, J. A., Orkin, S. H., . . . Basilico, C. (2012). Sox2 maintains self renewal of tumor-initiating cells in osteosarcomas. *Oncogene, 31*(18), 2270-2282. doi:10.1038/onc.2011.405

Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., . . . Lander, E. S. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell, 125*(2), 315-326. doi:10.1016/j.cell.2006.02.041

Bonnet, D., & Dick, J. E. (1997). Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat Med, 3*(7), 730-737. doi:10.1038/nm0797-730

Boumahdi, S., Driessens, G., Lapouge, G., Rorive, S., Nassar, D., Le Mercier, M., . . . Blanpain, C. (2014). SOX2 controls tumour initiation and cancer stem-cell functions in squamous-cell carcinoma. *Nature, 511*(7508), 246-250. doi:10.1038/nature13305

Calzolari, F., Appolloni, I., Tutucci, E., Caviglia, S., Terrile, M., Corte, G., & Malatesta, P. (2008). Tumor progression and oncogene addiction in a PDGF-B-induced model of gliomagenesis. *Neoplasia, 10*(12), 1373-1382, following 1382. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19048116

Chen, Y., Shi, L., Zhang, L., Li, R., Liang, J., Yu, W., . . . Shang, Y. (2008). The molecular mechanism governing the oncogenic potential of SOX2 in breast cancer. *The Journal of biological chemistry, 283*(26), 17969-17978. doi:10.1074/jbc.M802917200

Cheutin, T., & Cavalli, G. (2014). Polycomb silencing: from linear chromatin domains to 3D chromosome folding. *Curr Opin Genet Dev, 25*, 30-37. doi:10.1016/j.gde.2013.11.016

Di Rocco, F., Carroll, R. S., Zhang, J., & Black, P. M. (1998). Platelet-derived growth factor and its receptor expression in human oligodendrogliomas. *Neurosurgery, 42*(2), 341-346. doi:10.1097/00006123-199802000-00080

Fantes, J., Ragge, N. K., Lynch, S. A., McGill, N. I., Collin, J. R., Howard-Peebles, P. N., . . . FitzPatrick, D. R. (2003). Mutations in SOX2 cause anophthalmia. *Nature genetics, 33*(4), 461-463. doi:10.1038/ng1120

Favaro, R., Appolloni, I., Pellegatta, S., Sanga, A. B., Pagella, P., Gambini, E., . . . Nicolis, S. K. (2014). Sox2 is required to maintain cancer stem cells in a mouse model of high-grade oligodendroglioma. *Cancer research, 74*(6), 1833-1844. doi:10.1158/0008-5472.CAN-13-1942

Favaro, R., Valotta, M., Ferri, A. L., Latorre, E., Mariani, J., Giachino, C., . . . Nicolis, S. K. (2009). Hippocampal development and neural stem cell maintenance require Sox2-dependent regulation of Shh. *Nature neuroscience, 12*(10), 1248-1256. doi:10.1038/nn.2397

Ferri, A. L., Cavallaro, M., Braida, D., Di Cristofano, A., Canta, A., Vezzani, A., . . . Nicolis, S. K. (2004). Sox2 deficiency causes neurodegeneration and impaired neurogenesis in the adult mouse brain. *Development, 131*(15), 3805-3819. doi:10.1242/dev.01204

Ffrench-Constant, C., Miller, R. H., Kruse, J., Schachner, M., & Raff, M. C. (1986). Molecular specialization of astrocyte processes at nodes of Ranvier in rat optic nerve. *J Cell Biol, 102*(3), 844-852. doi:10.1083/jcb.102.3.844

Fullwood, M. J., Liu, M. H., Pan, Y. F., Liu, J., Xu, H., Mohamed, Y. B., . . . Ruan, Y. (2009). An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature, 462*(7269), 58-64. doi:10.1038/nature08497

Gangemi, R. M., Griffero, F., Marubbi, D., Perera, M., Capra, M. C., Malatesta, P., . . . Corte, G. (2009). SOX2 silencing in glioblastoma tumor-initiating

cells causes stop of proliferation and loss of tumorigenicity. *Stem cells, 27*(1), 40-48. doi:10.1634/stemcells.2008-0493

Garros-Regulez, L., Garcia, I., Carrasco-Garcia, E., Lantero, A., Aldaz, P., Moreno-Cugnon, L., . . . Matheu, A. (2016). Targeting SOX2 as a Therapeutic Strategy in Glioblastoma. *Front Oncol, 6*, 222. doi:10.3389/fonc.2016.00222

Guha, A., Dashner, K., Black, P. M., Wagner, J. A., & Stiles, C. D. (1995). Expression of PDGF and PDGF receptors in human astrocytoma operation specimens supports the existence of an autocrine loop. *International journal of cancer. Journal international du cancer, 60*(2), 168-173. doi:10.1002/ijc.2910600206

Hadjipanayis, C. G., & Van Meir, E. G. (2009). Tumor initiating cells in malignant gliomas: biology and implications for therapy. *J Mol Med (Berl), 87*(4), 363-374. doi:10.1007/s00109-009-0440-9

Hermanson, M., Funa, K., Hartman, M., Claesson-Welsh, L., Heldin, C. H., Westermark, B., & Nister, M. (1992). Platelet-derived growth factor and its receptors in human glioma tissue: expression of messenger RNA and protein suggests the presence of autocrine and paracrine loops. *Cancer research, 52*(11), 3213-3219. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/1317261

Ikushima, H., Todo, T., Ino, Y., Takahashi, M., Miyazawa, K., & Miyazono, K. (2009). Autocrine TGF-beta signaling maintains tumorigenicity of glioma-initiating cells through Sry-related HMG-box factors. *Cell stem cell, 5*(5), 504-514. doi:10.1016/j.stem.2009.08.018

Kamachi, Y., Uchikawa, M., & Kondoh, H. (2000). Pairing SOX off: with partners in the regulation of embryonic development. *Trends Genet, 16*(4), 182-187. doi:10.1016/s0168-9525(99)01955-1

Kar, S., Sengupta, D., Deb, M., Pradhan, N., & Patra, S. K. (2017). SOX2 function and Hedgehog signaling pathway are co-conspirators in promoting androgen independent prostate cancer. *Biochim Biophys Acta Mol Basis Dis, 1863*(1), 253-265. doi:10.1016/j.bbadis.2016.11.001

Kelberman, D., Rizzoti, K., Avilion, A., Bitner-Glindzicz, M., Cianfarani, S., Collins, J., . . . Dattani, M. T. (2006). Mutations within Sox2/SOX2 are associated with abnormalities in the hypothalamo-pituitary-gonadal axis in mice and humans. *The Journal of clinical investigation, 116*(9), 2442-2455. doi:10.1172/JCI28658

Li, G., Ruan, X., Auerbach, R. K., Sandhu, K. S., Zheng, M., Wang, P., . . . Ruan, Y. (2012). Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell, 148*(1-2), 84-98. doi:10.1016/j.cell.2011.12.014

Li, L., & Neaves, W. B. (2006). Normal stem cells and cancer stem cells: the niche matters. *Cancer research, 66*(9), 4553-4557. doi:10.1158/0008-5472.CAN-05-3986

Liu, K., Xie, F., Gao, A., Zhang, R., Zhang, L., Xiao, Z., . . . Lan, X. (2017). SOX2 regulates multiple malignant processes of breast cancer development through the SOX2/miR-181a-5p, miR-30e-5p/TUSC3 axis. *Mol Cancer, 16*(1), 62. doi:10.1186/s12943-017-0632-9

Long, K. B., & Hornick, J. L. (2009). SOX2 is highly expressed in squamous cell carcinomas of the gastrointestinal tract. *Hum Pathol, 40*(12), 1768-1773. doi:10.1016/j.humpath.2009.06.006

Louis, D. N., Holland, E. C., & Cairncross, J. G. (2001). Glioma classification: a molecular reappraisal. *Am J Pathol, 159*(3), 779-786. doi:10.1016/S0002-9440(10)61750-6

Mercurio, S., Serra, L., Motta, A., Gesuita, L., Sanchez-Arrones, L., Inverardi, F., . . . Nicolis, S. K. (2019). Sox2 Acts in Thalamic Neurons to Control the Development of Retina-Thalamus-Cortex Connectivity. *iScience, 15*, 257-273. doi:10.1016/j.isci.2019.04.030

Mercurio, S., Serra, L., & Nicolis, S. K. (2019). More than just Stem Cells: Functional Roles of the Transcription Factor Sox2 in Differentiated Glia and Neurons. *Int J Mol Sci, 20*(18). doi:10.3390/ijms20184540

Nawata, J., Kuramitsu, Y., Wang, Y., Kitagawa, T., Tokuda, K., Baron, B., . . . Nakamura, K. (2014). Active hexose-correlated compound down-regulates sex-determining region Y-box 2 of pancreatic cancer cells. *Anticancer Res, 34*(9), 4807-4811. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/25202061

Nicolis, S. K. (2007). Cancer stem cells and "stemness" genes in neuro-oncology. *Neurobiol Dis, 25*(2), 217-229. doi:10.1016/j.nbd.2006.08.022

Nister, M., Libermann, T. A., Betsholtz, C., Pettersson, M., Claesson-Welsh, L., Heldin, C. H., . . . Westermark, B. (1988). Expression of messenger RNAs for platelet-derived growth factor and transforming growth factor-alpha and their receptors in human malignant glioma cell lines. *Cancer research, 48*(14), 3910-3918. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/2454731

Novak, D., Huser, L., Elton, J. J., Umansky, V., Altevogt, P., & Utikal, J. (2019). SOX2 in development and cancer biology. *Semin Cancer Biol*. doi:10.1016/j.semcancer.2019.08.007

Pevny, L. H., & Nicolis, S. K. (2010). Sox2 roles in neural stem cells. *The international journal of biochemistry & cell biology, 42*(3), 421-424. doi:10.1016/j.biocel.2009.08.018

Piccirillo, S. G., Reynolds, B. A., Zanetti, N., Lamorte, G., Binda, E., Broggi, G., . . . Vescovi, A. L. (2006). Bone morphogenetic proteins inhibit the tumorigenic potential of human brain tumour-initiating cells. *Nature, 444*(7120), 761-765. doi:10.1038/nature05349

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. A., Flynn, R. A., & Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature, 470*(7333), 279-283. doi:10.1038/nature09692

Ragge, N. K., Lorenz, B., Schneider, A., Bushby, K., de Sanctis, L., de Sanctis, U., . . . Fitzpatrick, D. R. (2005). SOX2 anophthalmia syndrome. *Am J Med Genet A, 135*(1), 1-7. doi:10.1002/ajmg.a.30642

Reya, T., Morrison, S. J., Clarke, M. F., & Weissman, I. L. (2001). Stem cells, cancer, and cancer stem cells. *Nature, 414*(6859), 105-111. doi:10.1038/35102167

Riggi, N., Suva, M. L., De Vito, C., Provero, P., Stehle, J. C., Baumer, K., . . . Stamenkovic, I. (2010). EWS-FLI-1 modulates miRNA145 and SOX2 expression to initiate mesenchymal stem cell reprogramming toward Ewing sarcoma cancer stem cells. *Genes & development, 24*(9), 916-932. doi:10.1101/gad.1899710

Rudin, C. M., Durinck, S., Stawiski, E. W., Poirier, J. T., Modrusan, Z., Shames, D. S., . . . Seshagiri, S. (2012). Comprehensive genomic analysis identifies SOX2 as a frequently amplified gene in small-cell lung cancer. *Nature genetics, 44*(10), 1111-1116. doi:10.1038/ng.2405

Schaefer, S. M., Segalada, C., Cheng, P. F., Bonalli, M., Parfejevs, V., Levesque, M. P., . . . Sommer, L. (2017). Sox2 is dispensable for primary melanoma and metastasis formation. *Oncogene, 36*(31), 4516-4524. doi:10.1038/onc.2017.55

Schmitz, M., Temme, A., Senner, V., Ebner, R., Schwind, S., Stevanovic, S., . . . Weigle, B. (2007). Identification of SOX2 as a novel glioma-associated antigen and potential target for T cell-based immunotherapy. *British journal of cancer, 96*(8), 1293-1301. doi:10.1038/sj.bjc.6603696

Schneider, A., Bardakjian, T., Reis, L. M., Tyler, R. C., & Semina, E. V. (2009). Novel SOX2 mutations and genotype-phenotype correlation in anophthalmia and microphthalmia. *Am J Med Genet A, 149A*(12), 2706-2715. doi:10.1002/ajmg.a.33098

Schuller, U., Heine, V. M., Mao, J., Kho, A. T., Dillon, A. K., Han, Y. G., . . . Ligon, K. L. (2008). Acquisition of granule neuron precursor identity is a critical determinant of progenitor cell competence to form Shh-induced medulloblastoma. *Cancer Cell, 14*(2), 123-134. doi:10.1016/j.ccr.2008.07.005

Shackleton, M., Quintana, E., Fearon, E. R., & Morrison, S. J. (2009). Heterogeneity in cancer: cancer stem cells versus clonal evolution. *Cell, 138*(5), 822-829. doi:10.1016/j.cell.2009.08.017

Sisodiya, S. M., Ragge, N. K., Cavalleri, G. L., Hever, A., Lorenz, B., Schneider, A., . . . Fitzpatrick, D. R. (2006). Role of SOX2 mutations in human hippocampal malformations and epilepsy. *Epilepsia, 47*(3), 534-542. doi:10.1111/j.1528-1167.2006.00464.x

Stolzenburg, S., Rots, M. G., Beltran, A. S., Rivenbark, A. G., Yuan, X., Qian, H., . . . Blancafort, P. (2012). Targeted silencing of the oncogenic transcription factor SOX2 in breast cancer. *Nucleic acids research, 40*(14), 6725-6740. doi:10.1093/nar/gks360

Takahashi, K., & Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell, 126*(4), 663-676. doi:10.1016/j.cell.2006.07.024

Tziaferi, V., Kelberman, D., & Dattani, M. T. (2008). The role of SOX2 in hypogonadotropic hypogonadism. *Sex Dev, 2*(4-5), 194-199. doi:10.1159/000152035

Vanner, R. J., Remke, M., Gallo, M., Selvadurai, H. J., Coutinho, F., Lee, L., . . . Dirks, P. B. (2014). Quiescent sox2(+) cells drive hierarchical growth and relapse in sonic hedgehog subgroup medulloblastoma. *Cancer Cell, 26*(1), 33-47. doi:10.1016/j.ccr.2014.05.005

Varela, M., Ranuncolo, S. M., Morand, A., Lastiri, J., De Kier Joffe, E. B., Puricelli, L. I., & Pallotta, M. G. (2004). EGF-R and PDGF-R, but not bcl-2, overexpression predict overall survival in patients with low-grade astrocytomas. *J Surg Oncol, 86*(1), 34-40. doi:10.1002/jso.20036

Wolswijk, G., & Noble, M. (1989). Identification of an adult-specific glial progenitor cell. *Development, 105*(2), 387-400. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/2680425

Xie, J., Murone, M., Luoh, S. M., Ryan, A., Gu, Q., Zhang, C., . . . de Sauvage, F. J. (1998). Activating Smoothened mutations in sporadic basal-cell carcinoma. *Nature, 391*(6662), 90-92. doi:10.1038/34201

Yokota, E., Yamatsuji, T., Takaoka, M., Haisa, M., Takigawa, N., Miyake, N., . . . Naomoto, Y. (2017). Targeted silencing of SOX2 by an artificial transcription factor showed antitumor effect in lung and esophageal squamous cell carcinoma. *Oncotarget, 8*(61), 103063-103076. doi:10.18632/oncotarget.21523

Zappone, M. V., Galli, R., Catena, R., Meani, N., De Biasi, S., Mattei, E., . . . Nicolis, S. K. (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*(11), 2367-2382. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10804179

Zhang, Y., Wong, C. H., Birnbaum, R. Y., Li, G., Favaro, R., Ngan, C. Y., . . . Wei, C. L. (2013). Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. *Nature, 504*(7479), 306-310. doi:10.1038/nature12716

Zhao, C., Li, Y., Zhang, M., Yang, Y., & Chang, L. (2015). miR-126 inhibits cell proliferation and induces cell apoptosis of hepatocellular carcinoma cells partially by targeting Sox2. *Hum Cell, 28*(2), 91-99. doi:10.1007/s13577-014-0105-z

Chapter 2

Paper published, ***Glia***, 25 Sept 2020, **https://doi.org/10.1002/glia.23914**

Scope of the work

In this work, that represents the main part of my Ph.D. project, I investigate how the transcription factor Sox2, key stem factor during normal development, explicates an oncogenic role in glioma initiating cells. In this work I and my colleagues show as Sox2 is able to repress several oncosuppressor genes. Through the ectopic re-activation of these Sox2-repressed oncosuppressor genes we were able to obtain an anti-proliferative effect *in vitro*, and an anti-tumorigenic effect *in vivo* in a model of mouse oligodendroglioma. The in vitro anti-proliferative potential of these oncosuppressor factors is confirmed also in human glioblastoma primary cells. Many studies propose Sox2 as pharmacological target in therapies against cerebral tumors. We think that the understanding of transcriptional network regulated by Sox2 in a pathological context could open the way to novel therapeutic targets.

Sox2-dependent maintenance of mouse oligodendroglioma involves the Sox2-mediated downregulation of Cdkn2b, Ebf1, Zfp423 and Hey2

Running title:

Sox2 inhibits antioncogenes in glioma

Cristiana Barone[1], Mariachiara Buccarelli[2], Francesco Alessandrini[3§], Miriam Pagin[1], Laura Rigoldi[1], Irene Sambruni[1], Rebecca Favaro[1#], Sergio Ottolenghi[1], Roberto Pallini[4], Lucia Ricci-Vitiani[2], Paolo Malatesta[3] and Silvia K. Nicolis[1*]

[1]Department of Biotechnology and Biosciences, University of Milano-Bicocca, piazza della Scienza 2, 20126 Milano, Italy
[2]Department of Oncology and Molecular Medicine, Istituto Superiore di Sanità, Viale Regina Elena 299, 00161 Roma, Italy
[3]Dipartimento di Medicina Sperimentale, Università di Genova, and Ospedale Policlinico San Martino, IRCCS, largo Rosanna Benzi 10, 16132 Genova, Italy
[4]Institute of Neurosurgery, Fondazione Policlinico Universitario A. Gemelli IRCCS, Università Cattolica del Sacro Cuore, 00168 Rome, Italy
[§]Present address: Northwestern University, Feinberg School of Medicine, Chicago (IL), USA
[#]Present address: Humanitas University, Pieve Emanuele (MI), Dermatology Unit, Humanitas Research Hospital, IRCCS, Rozzano (MI), Italy
*Correspondence, lead contact: silvia.nicolis@unimib.it
The authors declare no potential conflicts of interest.

Acknowledgements

This research was supported by Associazione Italiana per la Ricerca sul Cancro (AIRC) grant IG 2014 – 16016 to S.K.N., and IG 2019 Id.23154 to R.P. C.B. is the recipient of a DIMET (Doctorate in Molecular and Translational Medicine) PhD fellowship. M.P. is the recipient of a Dipartimenti di Eccellenza fellowship.
The authors wish to thank Alessandra Boe for highly qualified technical assistance in flow cytometry of human GBM cells.

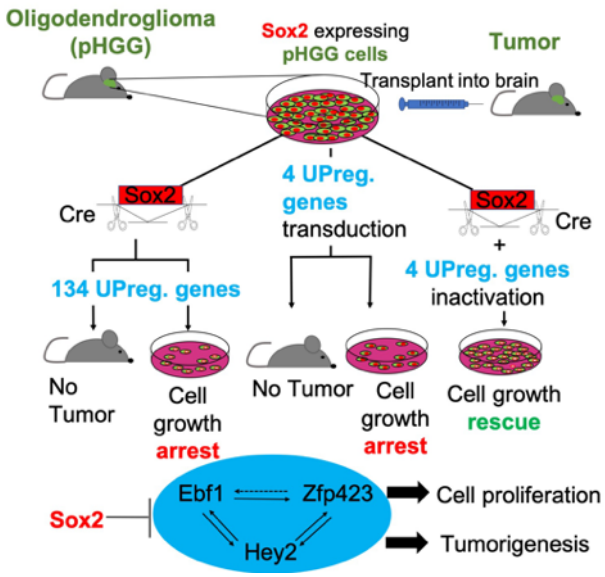
Abstract

Cancer stem cells (CSC) are essential for tumorigenesis. The transcription factor Sox2 is overexpressed in brain gliomas, and is essential to maintain CSC. In mouse

high-grade glioma pHGG cells in culture, Sox2 deletion causes cell proliferation arrest and inability to reform tumors after transplantation *in vivo;* in Sox2-deleted cells, 134 genes are derepressed. To identify genes mediating Sox2 deletion effects, we overexpressed into pHGG cells nine among the most derepressed genes, and identified four genes, Ebf1, Hey2, Zfp423 and Cdkn2b, that strongly reduced cell proliferation *in vitro* and brain tumorigenesis *in vivo*. CRISPR/Cas9 mutagenesis of each gene, individually or in combination (Ebf1+Cdkn2b), significantly antagonized the proliferation arrest caused by Sox2 deletion. The same genes also repressed clonogenicity in primary human glioblastoma-derived CSC-like lines. These experiments identify a network of critical tumor suppressive Sox2-targets whose inhibition by Sox2 is involved in glioma CSC maintenance, defining new potential therapeutic targets.

Keywords: Sox2, mouse oligodendroglioma, cancer stem cells, transcription factors, gene regulatory networks, tumorigenesis, human glioblastoma

Table of Content Image



Main Points

Sox2 maintains glioma tumorigenicity by repressing the antioncogenic activity of a regulatory network involving the Ebf1, Hey2, Zfp423 and Cdkn2b genes.
Mutation of these genes prevents the cell proliferation arrest of Sox2-deleted glioma cells.

1. Introduction

The Sox2 transcription factor has become widely known for its prominent roles in stem cells. It is required to maintain embryonic (ES), as well as various types of tissue-specific stem cells (Arnold et al., 2011; Avilion et al., 2003; Kondoh H, 2016). In the normal nervous system, Sox2 is important to maintain neural stem cells (NSC), *in vivo* in the hippocampus, as well as in long-term *in vitro* culture (Bertolini et al., 2019; Favaro et al., 2009; Gómez-López et al., 2011).
The discovery of a subset of tumor cells, capable to reform the tumor of origin following conventional chemotherapy (to which they are resistant), pointed to roles of a minority component of tumors in maintaining their growth; these cells are usually referred to, in general, as Cancer Stem Cells (CSC) (Chen et al., 2012; Nicolis, 2007; S. K. Singh et al., 2004); in the context of glioblastoma studies, we will refer to them as Glioma Stem Cells (GSC).
Sox2 is overexpressed in a variety of tumors (Barone, 2018; Cesarini et al., 2017; S. M. Schaefer et al., 2017), by unknown mechanisms, possibly via dysregulation of miR:Sox2 axes (Sathyan et al., 2015).
The ability of Sox2 to maintain ES and NSC led to the hypothesis that Sox2 may retain, in CSC, and specifically in GSC, an essential role (Barone, 2018; Nicolis, 2007; Wuebben & Rizzino, 2017).

In neural tumors, Sox2 is often active at high levels, although it is typically not mutated. This raised the question whether Sox2 is required by GSC for tumorigenesis, or whether Sox2 expression just parallels the oncogenic transformation, but has no role in maintaining transformed cells. In previous work (Favaro et al., 2014), we found that Sox2 Cre-mediated homozygous deletion in a mouse model of high-grade oligodendroglioma, obtained by PDGF-B viral overexpression (mouse PDGF-induced high-grade glioma, pHGG), completely prevents tumor reinitiation following transplantation into a mouse host brain (the assay that identifies GSC as tumor-reinitiating cells). *In vitro*, Sox2-deletion in pHGG cells significantly decreases their proliferation, and activates oligodendrocytic-like differentiation (Favaro et al., 2014). These findings paralleled those obtained in human patient-derived glioblastoma multiforme (GBM) cells, where Sox2 downregulation incapacitated cell proliferation and tumor reinitiation following transplantation (Bulstrode et al., 2017; Gangemi et al., 2009). Several mechanisms maintaining human GSC through SOX2 expression have been proposed (Ikushima et al., 2009; T. Schaefer et al., 2019; D. K. Singh et al., 2017; Suva et al., 2014); at variance, in an experimental model resembling the genesis of low-grade glioma, Sox2 inhibition appears to mediate a block of NSC "differentiation" leading to self-renewing and invasiveness (Modrek et al., 2017). Additionally, the importance of SOX2 in tumorigenesis and CSC was widened to various different tumor types: for example, SOX2 amplification was identified as an oncogenic driver of esophageal and lung squamous cell carcinoma, a common type of lung cancer (Bass et al., 2009), and SOX2 was found to be required within skin tumors stem cells (Boumahdi et al., 2014). An exception to these findings is represented by the Sox2-expressing neural crest-derived tumor melanoma; indeed, Sox2 conditional deletion in two different mouse models showed that Sox2 is entirely dispensable for tumorigenesis in both cases (Barone, 2018; Cesarini et al.,

2017; S. M. Schaefer et al., 2017). Overall, these findings extend the significance of the study of the "dark side of Sox2" (Wuebben & Rizzino, 2017), i.e. its ability to sustain tumorigenesis, to a broad sample of tumor types.

A central remaining open question is what are the mechanisms whereby Sox2 controls CSC maintenance, i.e. which downstream target genes of Sox2 mediate its function in CSC maintenance and tumorigenesis. Previously (Favaro et al., 2014), we identified genes that are differentially expressed following Sox2 deletion in pHGG oligodendroglioma cells, and found that most of these genes are upregulated following Sox2 loss. In the present work, we individually overexpressed, within cells carrying intact Sox2, nine among the genes most upregulated following Sox2 deletion, and asked if this would reproduce, at least in part, the effect of Sox2 deletion. We identified a subset of four Sox2 downstream target genes, whose experimental overexpression (in Sox2 non-deleted cells) leads to decreased cell proliferation and to increased cell differentiation *in vitro*, as well as to loss of tumor-initiating ability *in vivo*, pointing to these genes as important mediators of Sox2 function.

2. Results

Following Sox2 Cre-mediated deletion in pHGG cells, the first significant changes in gene expression are detected at 96 hours after Cre transduction and consist mainly in gene upregulation, involving 134 genes (putative tumor suppressor genes) (Favaro et al., 2014). In this work, we focused on these genes asking whether we could identify, among them, genes whose upregulation causally contributes to the anti- tumorigenic effect of Sox2 loss.

2.1 Lentiviral expression in pHGG cells of a specific subset of genes upregulated following Sox2 deletion causes drastic reduction of cell proliferation, reproducing the effect of Sox2 deletion.

To test the effect of upregulating, in pHGG cells, individual genes found overexpressed in Sox2-deleted cells, we cloned the cDNAs of nine among the most significantly overexpressed genes (Table 1) into lentiviral expression vectors, and asked if the overexpression of any of them into Sox2-positive pHGG cells would reproduce the effects of Sox2 deletion (Figure 1a). These genes included the top five upregulated genes, and four more upregulated genes among the top 25 (listed in Table S1) of potential interest, comprising the cell cycle regulator Cdkn2b and transcription factors Hey2 and Zfp423; the latter was recently reported to be a possible Sox2 activator in human gliomas (Signaroldi et al., 2016). The cDNAs (Table 1) were cloned upstream to an IRES and a delta-NGF-receptor (dNGFR) marker gene (lacking the intracellular domain), which is thus coexpressed with the cDNA; this allows to identify transduced cells by FACS analysis with antibodies recognizing dNGFR. We point out that, based on the results obtained with the control empty dNGFR vector (EV) (see below, Figure 1 and subsequent experiments using EV, Figures 2 and 3 and Figure S3), the expression of dNGFR on the cell membranes is devoid of any biological activity, as expected in view of the absence of the intracellular signaling domain. pHGG cells were transduced with each of the cDNA-expressing vectors, or with a control empty vector; the percentage of transduced cells was close to 100% by FACS analysis, and all cDNAs were significantly overexpressed in comparison to empty vector-transduced control cells (not shown). At 96 hr after transduction, cells were counted; empty vector-transduced cells had proliferated and reached high numbers, comparable to those of non- transduced cell controls; however, cells transduced with the cDNAs encoding

Ebf1, Hey2, Zfp423, and Cdkn2b showed substantially lower numbers, ranging from 20% (Hey2) to 40% (Zfp423) of cell numbers found in controls (Figure 1b). In a separate experiment, Ebf1- and Cdkn2b-encoding viruses were co-transduced in pHGG cells; the combination of the two viruses showed an additive effect on the reduction of cell numbers, as compared to cells in which each of the two genes was transduced individually (Figure S1). Further, 48 hr after transduction, we analyzed cells by FACS for dNGFR positivity to evaluate the percentage of transduced cells expressing the lentiviral constructs and its evolution through time (at 96 hr, 10 days and 17 days post-transduction) (Figure 1c). Cells transduced with the empty vector maintained similar high percentages from early to late stages (still representing 80–95% of total cells at day 17); in contrast, cells transduced with Cdkn2b, Ebf1, Zfp423, and Hey2 progressively reduced their relative abundance, representing only about 10% of total cells at day 17, indicative of a disadvantage caused by the over- expressed cDNAs (Figure 1c). In one experiment, we also evaluated the percentage of cells positive for phospho-histone H3, marking cells undergoing mitosis; phospho-histone H3-positive cells ranged between 16 and 20% in controls, but were strongly reduced (to 1–6%) in cells transduced with the cDNAs (Figure 1d). Apoptotic cell death was significantly increased by each of the transduced genes, by about two-fold, although less than observed after Sox2 deletion (Favaro et al., 2014) (Figure 1e).

We also asked if overexpression of these cDNAs caused cell differentiation, as previously observed for Sox2-deleted cells (Favaro et al., 2014). We analyzed cells for oligodendrocyte (GALC, O4) and astrocyte (GFAP) markers expression 7 days after transduction, by immunofluorescence (IF). Cdkn2b, Ebf1, and Zfp423 overexpression caused a significant increase (1–2% to 7%) in the numbers of cells positive for oligodendrocytic (O4, GalC) and, more variably, astrocytic (GFAP) markers (Figure 2a), with some cells exhibiting a “differentiated” morphology

(Figure 2b). This may actually be an underestimate of the number of differentiated cells induced by cDNA overexpression, because differentiation is first overt (by IF) at day 7, when the percentage of dNGFR-positive (transduced) cells is already reduced (Figure 1c); indeed, for Zfp423, where we could evaluate the percentage of differentiated cells specifically among the transduced cells (thanks to a FLAG tag carried by the Zfp423 cDNA), the percentage of GFAP-positive cells was higher among the Flag+ than among the Flag- cells (Figure 2a-IV). With Hey2, no increase of differentiation relative to the background in controls was observed (not shown). Overall, individual overexpression of the four Sox2-inhibited genes (Cdkn2b, Ebf1, Hey2, and Zfp423) is sufficient to reproduce the effect on cell proliferation of Sox2 deletion. Cell differentiation, on the other hand, is less affected by individual gene overexpression than it is by Sox2 deletion; it is possible that effects on cell differentiation would require the overactivity of all four Sox2-inhibited genes, possibly together with additional unidentified targets.

Finally, we did not observe cell numbers reduction, nor differentiation increase, in cells overexpressing Sdc4, Cryab, Rgs2, Wif1, Hopx (not shown). We thus continued our subsequent analyses focusing on the first group of four genes, hereafter termed “Sox2-inhibited tumor suppressive targets”.

We further asked if the Sox2-inhibited tumor suppressive targets may regulate each other. Viral upregulation of Ebf1 led to significant upregulation (by qRT-PCR) of endogenous Zfp423 and Hey2, and in turn, viral upregulation of Hey2 led to upregulation of Ebf1 and Zfp423 (Figure 3a). Zfp423 overexpression led to a small, but significant increase, of Ebf1 only; Cdkn2b overexpression had no effect on any of the other three genes (Figure S2). In all cases, control Wild Type (untransduced) and control EV-transduced pHGG cells showed no significant difference between them in the expression levels of the four genes (Figure S3). Of note, Sox2 levels

were unaltered in cells transduced with the four tested genes (Figure 3b), indicating that these genes do not act on cell proliferation by reducing Sox2 expression; hence, the changes in gene expression observed in Ebf1 and Hey2-transfected cells were not indirect effects of changes in Sox2 levels. These results point to a connection of these genes within a Sox2-dependent gene regulatory network (Figure 3c).

2.2 Sox2-inhibited tumor suppressive targets act downstream to Sox2 in maintaining glioma cell proliferation

The experiments described above demonstrate anti-oncogenic/anti-proliferative activity of the four identified Sox2-inhibited tumor suppressive targets (Cdkn2b, Ebf1, Hey2, Zfp423) upon overexpression in pHGG cells; however, is the upregulation of these genes, following Sox2 deletion, responsible for the proliferation arrest demonstrated by (Favaro et al., 2014)? To test this point, we individually mutated Ebf1, Cdkn2b or Zfp423 in pHGG cells (by CRISPR/Cas9-mediated technology), followed by Sox2 deletion by lentiviral CRE transduction (Favaro et al., 2014); we then evaluated cell numbers at 96 hr post transduction (Figure 4a). Cells where Sox2 had been deleted, and that had also been mutated in Ebf1, Cdkn2b, Zfp423, or Hey2 (Sox2−, Target−, Figure 4b), showed significantly higher cell numbers, compared to cells carrying Sox2 deletion, but no mutation in the Ebf1, Cdkn2b, Zfp423 target gene (Sox2−, Target+, Figure 4b). Importantly, ablation of each of the four target genes in cells carrying intact Sox2 genes had no significant effect, indicating that these genes importantly counteract cell proliferation only in the absence of Sox2 (when the genes are upregulated); in the presence of Sox2 (wild type cells), the expression of these genes is likely too low to significantly antagonize cell proliferation (contrary to what happens after Sox2

deletion), hence no effect is observed when the genes are mutated (Figure 4b). These experiments, taken together with the upregulation of these genes shown in Sox2-deleted cells, are thus consistent with the proposal that cell proliferation arrest upon Sox2 deletion requires upregulation of the Ebf1, Cdkn2b, Hey2, and Zfp423 genes.

We also evaluated the effect of the simultaneous mutation of two Sox2-inhibited tumor suppressive targets (Zfp423 and Cdkn2b or Ebf1 and Cdkn2b), followed by Sox2 deletion; mutating both genes had an additive effect on cell proliferation, when compared to the individual mutation of each of the genes (Figure 4c). Interestingly, the double Ebf1 and Cdkn2b mutation of Sox2-deleted cells increased their number from about 10% (relative to intact pHGG cells) to almost 50%, pointing to a significant rescue of their proliferative ability.

Finally, we evaluated the role of the Notch pathway, of which Hey2 is an effector, in a similar type of experiment. We inhibited the Notch pathway by the DAPT inhibitor of the gamma-secretase enzyme, acting on the initial step of the Notch signaling pathway. The inhibitor had no effect on the proliferation of cells carrying intact copies of Sox2; however, when added in combination with Cre-mediated Sox2 deletion, it significantly increased cell numbers, as compared to cells treated only with Cre (Figure 4d). In this experiment, Hey2 was strongly decreased by DAPT (Figure S4); this result is in agreement with our observations that Hey2 overexpression inhibits cell proliferation (Figure 1), and that Hey2 mutation in Sox2-deleted cells stimulates cell proliferation (Figure 4b).

2.3 Upregulation of Sox2-inhibited tumor suppressive targets identified in vitro antagonizes tumorigenesis in vivo

pHGG cells carrying wild-type Sox2 quickly and efficiently give rise to tumors of the same type of the tumor of origin, following orthotopic transplantation into the mouse brain; however, Cre-mediated Sox2 deletion completely prevents tumor reinitiation (Favaro et al., 2014). To test whether the experimental upregulation of Cdkn2b, Zfp423, Ebf1, Hey2 would also be effective in antagonizing tumorigenesis in vivo, we upregulated these genes by individual transduction of the corresponding lentiviral vectors into pHGG cells, and transplanted them into host brains, to assess their tumorigenic efficiency; Empty Vector-transduced cells served as controls. We set up conditions to obtain a multiplicity of infection giving rise to a ratio of about 60% transduced/40% non-transduced cells (dark cells and white cells in Figure 5a). This made it possible to retrospectively analyze, in tumors, the ratio of transduced versus non-transduced cells, as a measure of the relative tumorigenic efficiency of the two cell types (Figure 5). As shown in Figure 5b, Empty Vector-transduced cells developed tumors in 5/5 transplanted brains, whereas cells transduced with the vectors upregulating the Sox2-inhibited tumor suppressive targets gave rise to only 1–2 tumors out of the same number of transplanted brains. The (few) tumors developing with these vectors, and those obtained with the empty vector, were then dissected, dissociated to single cells, and analyzed by FACS for their percentage of dNGFR-positive (i.e., transduced) cells. In all tumors obtained with Empty Vector control cells, the percent- age of dNGFR-positive cells was similar to the percentage at the time of transplantation (i.e., about 60%), if not higher; however, for cells transduced with vectors overexpressing the Sox2-inhibited tumor suppressive targets, the percentage of dNGFR-positive cells was importantly reduced in comparison to the percentage prior to transplantation (from about 60% to about 20–25%), indicating a disadvantage of the transduced cells in tumor formation (Figure 5c). These findings indicate that upregulation of the Sox2-inhibited tumor suppressive targets, identified in vitro, antagonizes tumorigenesis in vivo.

2.4 Upregulation of Sox2-inhibited tumor suppressive targets identified in mouse pHGG cells antagonizes cell growth in different patient-derived, Sox2-expressing glioblastoma stem-like cell (GSC) lines

SOX2 expression has been previously documented in human GBM, the most aggressive and lethal human neural tumor, and experimental downregulation of SOX2 expression in two patient-derived cell lines was shown to severely reduce tumor-reinitiating ability of the cells following transplantation into the mouse brain (Gangemi et al., 2009). We thus wished to ask if upregulation of the Sox2-inhibited tumor suppressive targets, identified in pHGG cells, would also antagonize tumor cell growth of human patient-derived GSCs. We took advantage of a collection of patient-derived GSC lines (D'Alessandris et al., 2017; Marziali et al., 2016; Ricci-Vitiani et al., 2010), which all express SOX2, but at different levels (Figure S5a). These gliomas express Ebf1 at very variable levels, between different tumors, whereas Hey2, Cdkn2b, and Zfp423 expression is in general quite low; the heterogeneity of the genetic origin of these tumors, together with their limited number, preclude, at the present time, an assessment of potential correlations with Sox2 expression, although there is a trend toward higher Ebf1 levels in Sox2-low tumors (Figure S5b, c). Cells from three different lines, expressing different levels of SOX2 (see Figure S5a), were transduced with the same vectors previously used for mouse cells (encoding Cdkn2b, Zfp423, Ebf1, Hey2), and the clonogenic efficiency was tested (Figure 6a, b). The cells transduced with vectors expressing the Sox2-inhibited tumor suppressive targets showed significantly reduced clonogenic efficiency, in comparison to Empty Vector-transduced cells, although the extent of reduction varied between different cell lines (Figure 6b). The reduction was more pronounced for those cell lines expressing the highest SOX2 levels

(GSC#1, GSC#163), as compared to the line expressing lower SOX2 (GSC#83) (Figure 6b). These observations indicate that Sox2 targets, identified as mediators of Sox2 function in pHGG cells, also antagonize cell growth of various SOX2-expressing GBM-derived cell lines.

Further, only Hey2 transduction (Figure 6c), and not that of the others targets, also significantly reduced migration ability of GSC#1 and GSC#163 cells (although not of GSC#83, not shown), tested by a transwell-migration assay. Migration ability is an in vitro correlate of invasiveness, contributing to metastatic tumor development; our data suggest that Hey2 may also contribute to the regulation of this important feature of tumorigenic cells.

3. Discussion

Previously, we demonstrated that Sox2 deletion arrests cell proliferation in pHGG oligodendroglioma cells and prevents in vivo tumorigenesis by such cells (a test of CSC function, Singh et al., 2008), when transplanted into mouse brain (Favaro et al., 2014). We now show that overexpression (in non-Sox2-deleted pHGG cells) of genes upregulated following Sox2 deletion inhibits their in vitro proliferation, and prevents in vivo tumorigenesis, thus mimicking the effects of Sox2 deletion.

3.1 Sox2 maintains tumor cell properties by inhibiting four genes able to antagonize cell growth in vitro and in vivo

In this work, we identify four genes, that are importantly upregulated following Sox2 deletion in pHGG cells, as mediators of the proliferation arrest and inhibition of tumorigenesis due to Sox2 deletion. In fact, each of these genes (Cdkn2b, Ebf1, Hey2, Zfp423) significantly reduced the proliferation of non-Sox2-deleted pHGG cells, upon viral transduction (Figure 1), and inhibited tumorigenesis after

transplantation of the transduced cells in mouse brain. Not only are these genes able to reproduce the effects of Sox2 deletion in pHGG cells; their activity is also essential, as shown by CRISPR/ Cas9 mutagenesis or pharmacological inhibition (Figure 4), for repressing cell proliferation in Sox2-deleted pHGG cells. These results indicate that a critical contribution of Sox2 to the maintenance of tumorigenesis is represented by its ability to inhibit the expression of at least four genes acting as tumor suppressors. Interestingly, three out of the four genes, all encoding transcription factors (Ebf1, Hey2, and Zfp423), appear to be connected in a functional interaction network (Figure 3), in which both Ebf1 and Hey2 activate the other two transcription factors (Figure 3c), whereas the fourth gene (the cell cycle regulator Cdkn2b) does not affect the activity of the three transcription factor genes, yet it strongly cooperates with them to repress proliferation. This points to coordinated mechanisms for inhibiting cell proliferation stemming from Sox2 inhibition; importantly, none of these genes affects Sox2 activity (Figure 3).

It is further interesting that, among the genes that we tested by overexpression, there are some, such as Hopx, Wif1, and Cryab, that are known to act as tumor suppressors in other types of tumors, although they did not affect proliferation in our experiments. It is possible that the anti-tumorigenic effect of these genes is context- dependent (i.e., specific for a given tumor type), but we cannot rule out that other types of assay might reveal a role of these genes also in pHGG cells.

Finally, initial studies of three primary human glioblastoma- derived cell lines essentially confirm a repressive ability of the Sox2-inhibited tumor suppressive targets identified in pHGG cells, also in spontaneously arising human tumors, suggesting novel targets of possible therapeutic relevance.

3.2 How do the identified putative tumor suppressors affect tumor cell proliferation and tumorigenesis?

Of the four identified Sox2-inhibited tumor suppressive targets, three have previously been proposed to be somehow involved in spontaneous gliomas in man. Cdkn2b is commonly deleted in glioblastoma, often together with the adjacent Cdkn2a gene (Liu et al., 2014; Melin, 2011). Ebf1 encodes a transcription factor, that is an interaction partner for TET2, an enzyme mediating DNA demethylation; TET2 is inhibited by 2-hydroxyglutarate, an oncometabolite generated by mutated isocitrate dehydrogenase (IDH) 1 and 2, which are frequently mutated in low-grade gliomas, and other tumors. As these tumors share a hypermethylated DNA phenotype, it is possible that Ebf1, as part of the TET2-Ebf1 complex, abnormally binds to the hypermethylated genes (Guilhamon et al., 2013; Liao, 2009). Furthermore, both Ebf1 and Ebf3 (a member of the family) are found to be mutated and possibly inactivated in a variety of tumors (Liao, 2009), and Ebf1is mutated in Grade II and Grade III gliomas (Suzuki et al., 2015), implying a repressive role for these genes. In particular, EBF3 is mutated in gliomas, and activates genes involved in cell cycle arrest and apoptosis while repressing genes involved in cell survival and proliferation (Liao, 2009). Hey 2, a transcription factor, is downstream to the Notch receptor signaling pathway. Notch1 and Notch2 genes are frequently mutated by inactivating mutations in gliomas, particularly in Low Grade Gliomas associated to IDH mutations and 1p/19q losses (Bai et al., 2016) (Cancer Genome Atlas Research et al., 2015) (Suzuki et al., 2015), suggesting an antagonistic role to gliomagenesis of Notch itself or of downstream genes in the signaling pathway. Based on genomic data, a correlation of loss of the Notch pathway activity and particularly of Hey2 levels with oligodendroglioma was also proposed by (Halani et al., 2018). The experimental inactivation of Notch signaling had a stimulatory

effect on mouse glioblastoma cells proliferation (Giachino et al., 2015), although Ying et al. (2011) reported opposite conclusions regarding the effect of Notch repression in Retinoic Acid induced glioblastoma differentiation (Ying et al., 2011). Our present data are consistent with a tumor suppressive effect of Notch and Hey2 activation, as proposed by Giachino et al.; however, following Sox2 deletion in pHGG cells, only the expression of Hey2, but not of Notch, was increased, pointing to a direct effect of Sox2 on Hey2 activity. Finally, Zfp342 was also reported to have antigliomagenic activity, upon transfection, in one of three tested mouse astroglioma lines derived from Ink4a/Arf –/–; EGFR-mutant mice, acting, however, via SOX2 inhibition (Signaroldi et al., 2016).

From the analysis of spontaneous mutations leading to gliomas, specific pathways involving antioncogenic proteins, such as Retinoblastoma (Rb) and p53, appear to be involved, with different mutations often acting in conjunction. Thus, CDKN2A and CDKN2B are commonly deleted; they inhibit CDK4 and CDK6 kinases, which in turn activate Rb, the net effect being loss of Rb activity. Similarly, p53 may be inactivated by amplification of their inhibitor MDM proteins (Rao, Edwards, Joshi, Siu, & Riggins, 2010). In mouse, Notch signaling inactivation, combined with p53 loss, leads to the generation of aggressive brain tumors (Giachino et al., 2015); in agreement, in man Notch1 inactivating mutations are detected in gliomas, and Notch pathway effectors Hey2 and Hes5 expression levels are inversely correlated with tumor severity (Giachino et al., 2015). Our results indicate that at least two of the genes upregulated following Sox2 deletion in pHGG cells fit well in this general scheme: Cdkn2b and Hey2 increases implicate an involvement of the signaling upstream to Rb and of the Notch pathway, respectively. It is to be noted that the promoters of Ebf1, Cdkn2b, Zfp423 and Hey2 are directly bound by Sox2 in a human GBM- derived cell line (Fang et al., 2011). This implies, in particular for

Hey2, that Sox2 might directly impact on Hey2 activity, rather than indirectly through Notch pathway modulation. Other effects of Sox2 on levels of various mRNAs might also be regulated through changes in microRNA activities (Fang et al., 2011; Lopez-Bertoni et al., 2020), or through interactions with the transcriptional repressor Groucho (Liu et al., 2014) and with ribonucleoproteins (Fang et al., 2011).

3.3 Perspectives

Sox2 is overexpressed in several different tumors in man, and particularly in brain tumors. While a functional role has been demonstrated for Sox2 in the maintenance of neural tumors (Favaro et al., 2014; Gangemi et al., 2009), Sox2 mutations have not been demonstrated. It also remains unclear how Sox2 acts in the maintenance of tumors. We know from studies of wild type NSCs that, in the absence of Sox2, these cells progressively lose their replicative ability and are finally lost (Favaro et al., 2009). So far, an important mediator of Sox2 activity in sustaining long-term proliferation in vitro of NSC has been identified as Socs3, a signaling-controlling molecule (Bertolini et al., 2019). At variance with NSC, in oligodendroglioma we have now identified genes which are repressed, directly or indirectly, by Sox2, to maintain the tumor proliferation. As discussed above, these genes participate in regulatory circuits which eventually affect known tumor suppressors. So, a possible mode of action of Sox2 might be to somehow repress, in gliomas, genes involved in the suppression of tumorigenesis. In the case of Cdkn2b, a gene very often deleted in GBM, Sox2 action mimics the effect of the deletion by repressing the activity of Cdkn2b.

This finding, together with observations by several authors, points to the possibility of identifying proteins which are “druggable” and thus allowing to slow down

tumor progression. In this regard, a recent paper (Rubin & Sage, 2019; Walter et al., 2019) showed that mutating the Cdk2 gene in lung adenocarcinoma tumor cells sensitized these cells to the action of palbociclib, an inhibitor of Cdk4/6 kinases (themselves repressed by CDKN2b), which are frequently activated in these tumors. Our observation that Sox2 inhibits several genes which act as tumor suppressors points to the exciting possibility of developing drugs able to prevent the ability of Sox2 (or interacting proteins, such as Groucho) to repress these genes. It is also possible to propose that the products of the identified tumor suppressors (e.g., the mRNA) could be delivered to tumor cells by carriers (e.g., nanoparticles) targeted to the tumor-reinitiating cells, via the recognition of specific cell surface antigens carried by them (Haas et al., 2017).

Although Sox2 itself may be envisaged as a target for therapy approaches, and immunotherapy against SOX2 significantly increased survival time in mouse models (Favaro et al., 2014), the nuclear localization of SOX2 makes it a difficult target for efficient pharmacological recognition.

Overall, the identification of multiple downstream Sox2 targets, impacting on various signaling pathways, representing important mediators of Sox2 function, may hopefully contribute to the design of specific, multi-hit therapy approaches.

4. Materials and methods

4.1 Lentiviral constructs and infections

The Ebf1, Cdkn2b, Hey2 cDNAs were generated by RT-PCR from E14.5 mouse telencephalon RNA (primers: Table S3), the FLAG- Zfp432 cDNA was cut with XhoI/AgeI from a pMSCV-puro vector (gift from G. Testa) (Signaroldi et al., 2016); all were cloned into a unique BamHI site upstream to the IRES-dNGFR

cassette of the pHR SIN BX IR/EMW (Barbarani, Fugazza, Barabino, & Ronchi, 2019) (a gift from A.Ronchi, Milano).
Lentiviral vectors for CRISPR-Cas9 experiments were from Addgene: lentiCRISPRV2puro (Addgene #98290, RRID: Addgene_104990) and lentiCRISPRv2neo (Addgene #98292, RRID: Addgene_104992). The RNA guide sequences were designed and cloned according to: https://media.addgene.org/cms/filer_public/53/09/53091cde-b1ee-47ee-97cf9b3b05d290f2/lenticrisprv2-and-lentiguide-oligo-cloning-protocol.pdf, and are: Ebf1 5' -CGACAGACAGGGCCAGCCCG-3' , Cdkn2b 5' -CAGGGCGTTGGGATCTGCGC-3', Zfp423 5'-TCACAACATCCGGCCGGCC-3'.
For Cre-encoding virus see (Favaro et al., 2014). Lentiviral particles were produced by the calcium phosphate transfection protocol in the packaging human embryonic kidney cell line 293T and infection performed as previously described (Ricci-Vitiani et al., 2004).

4.2 Cell cultures

GSCs were isolated from surgical samples of adult patients who underwent craniotomy at the Institute of Neurosurgery, Catholic University of Rome, upon patient informed consent and approval by the local ethical committee (Pallini et al., 2008). GSC cultures were established from surgical specimens through mechanical dissociation and culturing in a serum-free medium supplemented with 20 ng/ml epidermal growth factor (EGF) and 10 ng/ml basic fibroblast growth factor (bFGF) (Peprotech, 100-15 and 100-18B) as previously described (D'Alessandris et al., 2017; Pallini et al., 2008).

4.3 In vitro overexpression assay of Sox2-inhibited tumor suppressive targets

For in vitro overexpression experiments (Figures 1 and 2) cells were plated at 300,000 cells/well in Matrigel-coated 6-well plates, in DMEM-F12 (Life Technologies, cod. 31331028) Complete Medium (CM), that is, supplemented with 1 ml/50 ml B27 (Thermo Fisher, cod.17504044), 400 µl/50 ml Penicillin/Streptomicin (Euroclone, cod. ECB3001D) and 2% of Fetal Bovine Serum (FBS) (Euroclone, cod. ECS0180L). Cells were transduced 4 hr after plating, with cDNA overexpressing lentivirus or Empty Vector control lentivirus at MOI 10. The medium was changed 15 hr after transduction. After 96 hr cells were collected, counted (Figure 1b), re-plated at 300.000 cells/well in Matrigel-coated 6-well plates, and collected again at 10 and 17 days post-infection. At every passage, an aliquot of cells (min 50,000-max 500,000) was fixed with 4% paraformaldehyde (PFA) and stained as in (Barbarani et al., 2019) with an anti-human CD271 (dNGFR) antibody, conjugated with Phycoerythrin (PE) (BioLegend, Cat. No. 345106, RRID: AB_2152647) (dilution 1:200) and analyzed by CytoFLEX (Beckman- Coulter) to determine the percentage of infected cells (Figure 1c). The % of dNGFR-positive cells always approached 100%. With the double overexpression of Ebf1 and Cdkn2b, and control Ebf1and EV, or Cdkn2b and EV (Figure S1) we used MOI 8 for each vector, that still results in almost 100% efficiency of transduction upon single virus transduction. To evaluate the differentiation rate, cells collected at 96 hr were plated at a density of 15,000 cells/well in Matrigel-coated 24-well plates. After 3 days, cells were fixed with 4% paraformaldehyde. To estimate proliferation transduced cells were collected 48 hr after transduction and plated at a density of 30,000 cells per well in Matrigel-coated 24-well chambered slides. After 24 hr cells were fixed with 4% PFA. Antibodies against FLAG (1:800; Sigma-Ald. F7425 RRID: AB_439687), GFAP (1:100; Millipore MAB3402 RRID: AB_94844) and phosphohistone H3 (1:1,000; Millipore 06-570 RRID: AB_310177) were used for IF performed as in (Cavallaro

et al., 2008). O4 and GalC IF used anti-O4 and anti-GalC hybridomas (undiluted supernatant) as in (Favaro et al., 2014), a gift from C. Taveggia (Figure 2a, b).

For in vitro overexpression of Ebf1, Hey2, Zfp423, and Cdkn2b in human GSCs (Figure 6), cells were transduced with cDNA-expressing lentivirus or empty vector as previously described (Ricci-Vitiani et al., 2004). After 48 hr, the efficiency of infection was evaluated by FACSCanto (Becton Dickinson) analysis with anti-human CD271 (dNGFR) PE-conjugated antibody, as above.

Colony-forming ability was evaluated by plating a single cell/well in 96-well plates. After 3–4 weeks, each well was examined and the number of spheres/aggregates were counted.

Migration ability was evaluated by plating the cells in Corning FluoroBlock Multiwell Inserts System (Corning Life Sciences, REF351164), according to manufacturer's instructions. Briefly, 3 × 103 GSC were added to the upper chambers in stem cell medium. Stem cell medium supplemented with growth factors (EGF and b-FGF) was used as chemoattractant in the lower chambers. The plates were incubated for 48 hr at 37°C, then the fluorescent dye calcein acetoxymethylester (calcein-AM, Life Technologies Corporation, cod C1430) was added to the lower chamber for 30 min. The cell viability indicator calcein-AM is a non-fluorescent, cell permeant compound that is hydrolyzed by intra- cellular esterases into the fluorescent anion calcein and can be used to fluorescently label viable cells before microscope observation. The number of migrated cells was evaluated by counting the cells after imaging acquisition using a fluorescence microscope.

4.4 Quantitative reverse transcriptase PCR

Primers for the quantification of mRNAs of Sox2-inhibited antioncogenic targets are reported in Table S4. RNA isolation and Real Time PCR were performed as described in (Barbarani et al., 2019).

4.5 CRISPR-Cas9 assays

Cells were plated at density of 1 × 106 cells/Matrigel-coated 100 mm plate, in DMEM-F12 CM, without FBS but with 10 ng/ml EGF and bFGF (Peprotech 100-15 and 100-18). Cells were transduced, 4 hr after plating, with 15 μl of lentiCRISPRV2puro or lentiCRISPRV2neo lentivirus for each plate. The medium was changed 15 hr after transduction. 48 hr after transduction the cells were selected with 5 μg/ml Puromycin (Sigma-Aldrich, cod. P8833) or 500 μg/ml G418 (Sigma-Aldrich, cod. A1720) for 3 days. After selection an aliquot of cells (100,000–300,000 cells) was collected to evaluate the efficiency of mutagenesis by cloning and sequencing the target gene. For experiments in which cells had to be mutated with two different sgRNA (Figure 4c), they were first treated with lentiCRISPRV2puro and selected in Puromycin for 3 days. After selection cells were collected and re-plated at a density of 1 × 106 cells in Matrigel-coated 100 mm plate, subsequently treated with lentiCRISPRV2neo, then selected with G418 for 3 days. After selection the cells were collected and plated at a density of 300,000 cells in Matrigel-coated 6-well. After 4 hr from plating cells were transduced with Cre-encoding virus (MOI 7). The medium was changed after 15 hr. After 96 hr cells were collected and counted (Figure 4b, c). To evaluate the percentage of residual Sox2 positive cells, cells were plated at a density of 30,000 cells/well in Matrigel-coated 24-well plates, and analyzed by Sox2 IF as in (Cavallaro et al., 2008). The per- centage of residual SOX2-positive cells 4 days after CRE transduction was comparable among different conditions, and was routinely <10%, and never more than 30%. To evaluate the efficiency of CRISPR- Cas9-mediated mutagenesis DNA

was extracted and the genomic region surrounding the sgRNA-targeted site was amplified by PCR, with primers listed in Table S5. The amplified DNA was cloned in pGEM-T Easy (Promega, A1360) by transformation of TOP10 E. coli, and inserts from individual colonies were sequenced. The sequences from CRISPR-Cas9-treated cells were compared to wild-type sequences, by using BLAST NCBI. An indel mutation was found in 100% of cases.

Notch pathway inhibition by DAPT (Sigma, cod. D5942): cells were plated at a density of 300,000 cells/well in Matrigel-coated 6-well plates and, 4 hr after plating, were transduced (MOI 7) with the Cre-encoding virus (Favaro et al., 2014). After 15 hr, the medium was replaced with medium supplemented with DAPT 10 μM or 2,5 μM (++ and +, respectively, in Figure 4d). After 96 hr, cells were collected and counted (Figure 4d); an aliquot was used to evaluate the Hey2 and Sox2 mRNA relative abundance (Figure S1) by RT-PCR using primers reported in Table S4. Sox2 residual mRNA in Cre-treated cells was <10% compared to control (not shown).

4.6 Transplantation of virally transduced cells into mouse brain

All in vivo experiments were approved by the Ethical Committee for Animal Experimentation (CSEA) of the IRCCS AOU San Martino IST, Genova, and by the Italian Ministry of Health. Animals were handled in agreement with Italian current regulations about animal use for scientific purposes (D.lvo January 27, 1992, n. 116). Cell injections in the brain parenchyma of adult mice were as described (Favaro et al., 2014; Gambini et al., 2012). Cells to be transplanted were treated with viruses at MOI 5, to obtain about 60% transduced cells. We transplanted 30,000 cells per mouse of five C57Bl/6J mice per condition. Animals were monitored daily and, 3 days after the first mouse started to show signs of

neurological symptoms, we sacrificed all the transplanted mice (38 days after pHGG cells transplantation, 2 days before to the overall median survival of mice transplanted with pHGG cells (Favaro et al., 2014). The presence of the green fluorescent protein (GFP) fluorescent reporter expressed by the tumor cells was evaluated under a Leica fluorescence stereomicroscope on the whole brain and, in the instances where clear masses were not visible, the brain was coronally sectioned to check for the possible presence of small masses. Mice showing pHGG cells outside the injection site were evaluated as tumor-bearing. The GFP–positive tumor area was dissected under the fluorescence microscope, and trypsinized for 20 min to obtain a single-cell suspension, then analysed by FACS as described previously (Barbarani, Fugazza, Barabino, & Ronchi, 2019) to assess the percentage of infected (dNGFR-positive) cells (Figure 5c).

Acknowledgments

This research was supported by Associazione Italiana per la Ricerca sul Cancro (AIRC) grant IG 2014-16016 to S.K.N., and IG 2019 Id. 23154 to R.P. C.B. is the recipient of a DIMET (Doctorate in Molecular and Translational Medicine) PhD fellowship. M.P. is the recipient of a Dipartimenti di Eccellenza fellowship. The authors wish to thank Alessandra Boe for highly qualified technical assistance in flow cytometry of human GBM cells, Giuseppe Testa for the the FLAG-Zfp432 cDNA, and Carla Taveggia for the anti-O4 and anti-GalC hybridomas supernatants.

Conflict of interest

The authors declare no potential conflicts of interest.

Data availability statement

Data availability: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Arnold, K., Sarkar, A., Yram, M. A., Polo, J. M., Bronson, R., Sengupta, S., . . . Hochedlinger, K. (2011). Sox2(+) adult stem and progenitor cells are important for tissue regeneration and survival of mice. *Cell stem cell, 9*(4), 317-329. doi:10.1016/j.stem.2011.09.001

Avilion, A. A., Nicolis, S. K., Pevny, L. H., Perez, L., Vivian, N., & Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes & development, 17*(1), 126-140. doi:10.1101/gad.224503

Bai, H., Harmanci, A. S., Erson-Omay, E. Z., Li, J., Coskun, S., Simon, M., . . . Gunel, M. (2016). Integrated genomic characterization of IDH1-mutant glioma malignant progression. *Nature genetics, 48*(1), 59-66. doi:10.1038/ng.3457

Barbarani, G., Fugazza, C., Barabino, S. M. L., & Ronchi, A. E. (2019). SOX6 blocks the proliferation of BCR-ABL1(+) and JAK2V617F(+) leukemic cells. *Sci Rep, 9*(1), 3388. doi:10.1038/s41598-019-39926-4

Barone, C., Pagin, M., Serra, L., Motta, L., Rigoldi, L., Giubbolini, S., Badiola-Sanga, A., Mercurio, S. and Nicolis, S.K. (2018). Sox2 functions in neural cancer stem cells: the importance of the context. *Insights of Neuro-Oncology, 2*(1), 18-26.

Bass, A. J., Watanabe, H., Mermel, C. H., Yu, S., Perner, S., Verhaak, R. G., . . . Meyerson, M. (2009). SOX2 is an amplified lineage-survival oncogene in

lung and esophageal squamous cell carcinomas. *Nature genetics, 41*(11), 1238-1242. doi:10.1038/ng.465

Bertolini, J. A., Favaro, R., Zhu, Y., Pagin, M., Ngan, C. Y., Wong, C. H., . . . Wei, C. L. (2019). Mapping the Global Chromatin Connectivity Network for Sox2 Function in Neural Stem Cell Maintenance. *Cell stem cell, 24*(3), 462-476 e466. doi:10.1016/j.stem.2019.02.004

Boumahdi, S., Driessens, G., Lapouge, G., Rorive, S., Nassar, D., Le Mercier, M., . . . Blanpain, C. (2014). SOX2 controls tumour initiation and cancer stem-cell functions in squamous-cell carcinoma. *Nature, 511*(7508), 246-250. doi:10.1038/nature13305

Bulstrode, H., Johnstone, E., Marques-Torrejon, M. A., Ferguson, K. M., Bressan, R. B., Blin, C., . . . Pollard, S. M. (2017). Elevated FOXG1 and SOX2 in glioblastoma enforces neural stem cell identity through transcriptional control of cell cycle and epigenetic regulators. *Genes & development, 31*(8), 757-773. doi:10.1101/gad.293027.116

Cesarini, V., Guida, E., Todaro, F., Di Agostino, S., Tassinari, V., Nicolis, S., . . . Dolci, S. (2017). Sox2 is not required for melanomagenesis, melanoma growth and melanoma metastasis in vivo. *Oncogene, 36*(31), 4508-4515. doi:10.1038/onc.2017.53

Chen, J., Li, Y., Yu, T. S., McKay, R. M., Burns, D. K., Kernie, S. G., & Parada, L. F. (2012). A restricted cell population propagates glioblastoma growth after chemotherapy. *Nature, 488*(7412), 522-526. doi:10.1038/nature11287

D'Alessandris, Q. G., Biffoni, M., Martini, M., Runci, D., Buccarelli, M., Cenci, T., . . . Pallini, R. (2017). The clinical value of patient-derived glioblastoma tumorspheres in predicting treatment response. *Neuro Oncol, 19*(8), 1097-1108. doi:10.1093/neuonc/now304

Fang, X., Yoon, J. G., Li, L., Yu, W., Shao, J., Hua, D., . . . Lin, B. (2011). The SOX2 response program in glioblastoma multiforme: an integrated ChIP-seq, expression microarray, and microRNA analysis. *BMC Genomics, 12*, 11. doi:10.1186/1471-2164-12-11

Favaro, R., Appolloni, I., Pellegatta, S., Sanga, A. B., Pagella, P., Gambini, E., . . . Nicolis, S. K. (2014). Sox2 is required to maintain cancer stem cells in a mouse model of high-grade oligodendroglioma. *Cancer research, 74*(6), 1833-1844. doi:10.1158/0008-5472.CAN-13-1942

Favaro, R., Valotta, M., Ferri, A. L., Latorre, E., Mariani, J., Giachino, C., . . . Nicolis, S. K. (2009). Hippocampal development and neural stem cell maintenance require Sox2-dependent regulation of Shh. *Nature neuroscience, 12*(10), 1248-1256. doi:10.1038/nn.2397

Gambini, E., Reisoli, E., Appolloni, I., Gatta, V., Campadelli-Fiume, G., Menotti, L., & Malatesta, P. (2012). Replication-competent herpes simplex virus retargeted to HER2 as therapy for high-grade glioma. *Mol Ther, 20*(5), 994-1001. doi:10.1038/mt.2012.22

Gangemi, R. M., Griffero, F., Marubbi, D., Perera, M., Capra, M. C., Malatesta, P., . . . Corte, G. (2009). SOX2 silencing in glioblastoma tumor-initiating cells causes stop of proliferation and loss of tumorigenicity. *Stem cells, 27*(1), 40-48. doi:10.1634/stemcells.2008-0493

Gómez-López, S., Wiskow, O., Favaro, R., Nicolis, S. K., Price, D. J., Pollard, S. M., & Smith, A. (2011). Sox2 and Pax6 maintain the proliferative and developmental potential of gliogenic neural stem cells In vitro. *Glia, 59*(11), 1588-1599. doi:10.1002/glia.21201

Guilhamon, P., Eskandarpour, M., Halai, D., Wilson, G. A., Feber, A., Teschendorff, A. E., . . . Beck, S. (2013). Meta-analysis of IDH-mutant

cancers identifies EBF1 as an interaction partner for TET2. *Nat Commun, 4*, 2166. doi:10.1038/ncomms3166

Haas, T. L., Sciuto, M. R., Brunetto, L., Valvo, C., Signore, M., Fiori, M. E., . . . De Maria, R. (2017). Integrin alpha7 Is a Functional Marker and Potential Therapeutic Target in Glioblastoma. *Cell stem cell, 21*(1), 35-50 e39. doi:10.1016/j.stem.2017.04.009

Halani, S. H., Yousefi, S., Velazquez Vega, J., Rossi, M. R., Zhao, Z., Amrollahi, F., . . . Brat, D. J. (2018). Multi-faceted computational assessment of risk and progression in oligodendroglioma implicates NOTCH and PI3K pathways. *NPJ Precis Oncol, 2*, 24. doi:10.1038/s41698-018-0067-9

Ikushima, H., Todo, T., Ino, Y., Takahashi, M., Miyazawa, K., & Miyazono, K. (2009). Autocrine TGF-beta signaling maintains tumorigenicity of glioma-initiating cells through Sry-related HMG-box factors. *Cell stem cell, 5*(5), 504-514. doi:10.1016/j.stem.2009.08.018

Kondoh H, L.-B. R. e. b. (2016). *Sox2, Biology and Role in Development and Disease* (Vol. ISBN: 978-0-12-800352-7): Elsevier, Associated Press.

Liao, D. (2009). Emerging roles of the EBF family of transcription factors in tumor suppression. *Mol Cancer Res, 7*(12), 1893-1901. doi:10.1158/1541-7786.MCR-09-0229

Liu, Y. R., Laghari, Z. A., Novoa, C. A., Hughes, J., Webster, J. R., Goodwin, P. E., . . . Scotting, P. J. (2014). Sox2 acts as a transcriptional repressor in neural stem cells. *BMC Neurosci, 15*, 95. doi:10.1186/1471-2202-15-95

Lopez-Bertoni, H., Kotchetkov, I. S., Mihelson, N., Lal, B., Rui, Y., Ames, H., . . . Laterra, J. (2020). A Sox2:miR-486-5p Axis Regulates Survival of GBM Cells by Inhibiting Tumor Suppressor Networks. *Cancer research, 80*(8), 1644-1655. doi:10.1158/0008-5472.CAN-19-1624

Marziali, G., Signore, M., Buccarelli, M., Grande, S., Palma, A., Biffoni, M., . . . Ricci-Vitiani, L. (2016). Metabolic/Proteomic Signature Defines Two Glioblastoma Subtypes With Different Clinical Outcome. *Sci Rep, 6*, 21557. doi:10.1038/srep21557

Melin, B. (2011). Genetic causes of glioma: new leads in the labyrinth. *Curr Opin Oncol, 23*(6), 643-647. doi:10.1097/CCO.0b013e32834a6f61

Modrek, A. S., Golub, D., Khan, T., Bready, D., Prado, J., Bowman, C., . . . Placantonakis, D. G. (2017). Low-Grade Astrocytoma Mutations in IDH1, P53, and ATRX Cooperate to Block Differentiation of Human Neural Stem Cells via Repression of SOX2. *Cell Rep, 21*(5), 1267-1280. doi:10.1016/j.celrep.2017.10.009

Nicolis, S. K. (2007). Cancer stem cells and "stemness" genes in neuro-oncology. *Neurobiol Dis, 25*(2), 217-229. doi:10.1016/j.nbd.2006.08.022

Rao, S. K., Edwards, J., Joshi, A. D., Siu, I. M., & Riggins, G. J. (2010). A survey of glioblastoma genomic amplifications and deletions. *J Neurooncol, 96*(2), 169-179. doi:10.1007/s11060-009-9959-4

Ricci-Vitiani, L., Pallini, R., Biffoni, M., Todaro, M., Invernici, G., Cenci, T., . . . De Maria, R. (2010). Tumour vascularization via endothelial differentiation of glioblastoma stem-like cells. *Nature, 468*(7325), 824-828. doi:10.1038/nature09557

Rubin, S. M., & Sage, J. (2019). Manipulating the tumour-suppressor protein Rb in lung cancer reveals possible drug targets. *Nature, 569*(7756), 343-344. doi:10.1038/d41586-019-01319-y

Sathyan, P., Zinn, P. O., Marisetty, A. L., Liu, B., Kamal, M. M., Singh, S. K., . . . Majumder, S. (2015). Mir-21-Sox2 Axis Delineates Glioblastoma Subtypes with Prognostic Impact. *The Journal of neuroscience : the*

official journal of the Society for Neuroscience, 35(45), 15097-15112. doi:10.1523/JNEUROSCI.1265-15.2015

Schaefer, S. M., Segalada, C., Cheng, P. F., Bonalli, M., Parfejevs, V., Levesque, M. P., . . . Sommer, L. (2017). Sox2 is dispensable for primary melanoma and metastasis formation. *Oncogene, 36*(31), 4516-4524. doi:10.1038/onc.2017.55

Schaefer, T., Ramadoss, A., Leu, S., Tintignac, L., Tostado, C., Bink, A., . . . Boulay, J. L. (2019). Regulation of glioma cell invasion by 3q26 gene products PIK3CA, SOX2 and OPA1. *Brain Pathol, 29*(3), 336-350. doi:10.1111/bpa.12670

Signaroldi, E., Laise, P., Cristofanon, S., Brancaccio, A., Reisoli, E., Atashpaz, S., . . . Testa, G. (2016). Polycomb dysregulation in gliomagenesis targets a Zfp423-dependent differentiation network. *Nat Commun, 7*, 10753. doi:10.1038/ncomms10753

Singh, D. K., Kollipara, R. K., Vemireddy, V., Yang, X. L., Sun, Y., Regmi, N., . . . Bachoo, R. M. (2017). Oncogenes Activate an Autonomous Transcriptional Regulatory Circuit That Drives Glioblastoma. *Cell Rep, 18*(4), 961-976. doi:10.1016/j.celrep.2016.12.064

Singh, S. K., Hawkins, C., Clarke, I. D., Squire, J. A., Bayani, J., Hide, T., . . . Dirks, P. B. (2004). Identification of human brain tumour initiating cells. *Nature, 432*(7015), 396-401. doi:10.1038/nature03128

Suva, M. L., Rheinbay, E., Gillespie, S. M., Patel, A. P., Wakimoto, H., Rabkin, S. D., . . . Bernstein, B. E. (2014). Reconstructing and reprogramming the tumor-propagating potential of glioblastoma stem-like cells. *Cell, 157*(3), 580-594. doi:10.1016/j.cell.2014.02.030

Suzuki, H., Aoki, K., Chiba, K., Sato, Y., Shiozawa, Y., Shiraishi, Y., . . . Ogawa, S. (2015). Mutational landscape and clonal architecture in grade II and III gliomas. *Nature genetics, 47*(5), 458-468. doi:10.1038/ng.3273

Walter, D. M., Yates, T. J., Ruiz-Torres, M., Kim-Kiselak, C., Gudiel, A. A., Deshpande, C., . . . Feldser, D. M. (2019). RB constrains lineage fidelity and multiple stages of tumour progression and metastasis. *Nature, 569*(7756), 423-427. doi:10.1038/s41586-019-1172-9

Wuebben, E. L., & Rizzino, A. (2017). The dark side of SOX2: cancer - a comprehensive overview. *Oncotarget, 8*(27), 44917-44943. doi:10.18632/oncotarget.16570

Ying, M., Wang, S., Sang, Y., Sun, P., Lal, B., Goodwin, C. R., . . . Xia, S. (2011). Regulation of glioblastoma stem cells by retinoic acid: role for Notch pathway inhibition. *Oncogene, 30*(31), 3454-3467. doi:10.1038/onc.2011.58

Tables and Figures

Table 1

Sox2 target	Ranking	Fold Increase (Sox2-del cells 96hpd/WTcells 96hpd)	Mean expression values in wt	Mean expression values in mCherry Control	Mean expression values in Sox2-deleted cells 48hpd	Mean expression values in Sox2-deleted cells 96hpd
Hopx	1	12.4x UP	213.3	229.7	226.4	2640.0
Sdc4	2	8.2x UP	42.3	50.2	86.4	346.5
Wif1	3	7.1x UP	107.2	133.6	157.5	763.6
Rgs2	4	7.1x UP	732.0	787.4	1154.3	5204.4
Ebf1	5	7x UP	40.0	48.9	91.6	281.1
Cryab	11	5.3x UP	17.1	16.1	21.4	91.6
Hey2	19	4.1x UP	29.6	26.7	32.8	121.2
Cdkn2b	25	3.8x UP	135.0	144.4	172.2	517.5
Zfp423	147	2.0x UP	78.8	83.4	89.6	211.4

Table 1

Expression of Sox2-inhibited genes in Cre-transduced Sox2-deleted pHGG cells and in control (mCherry-transduced) cells, at 48 and 96 hours post Cre transduction

hpd: hours post deletion (Expression data from Favaro et al. 2014).

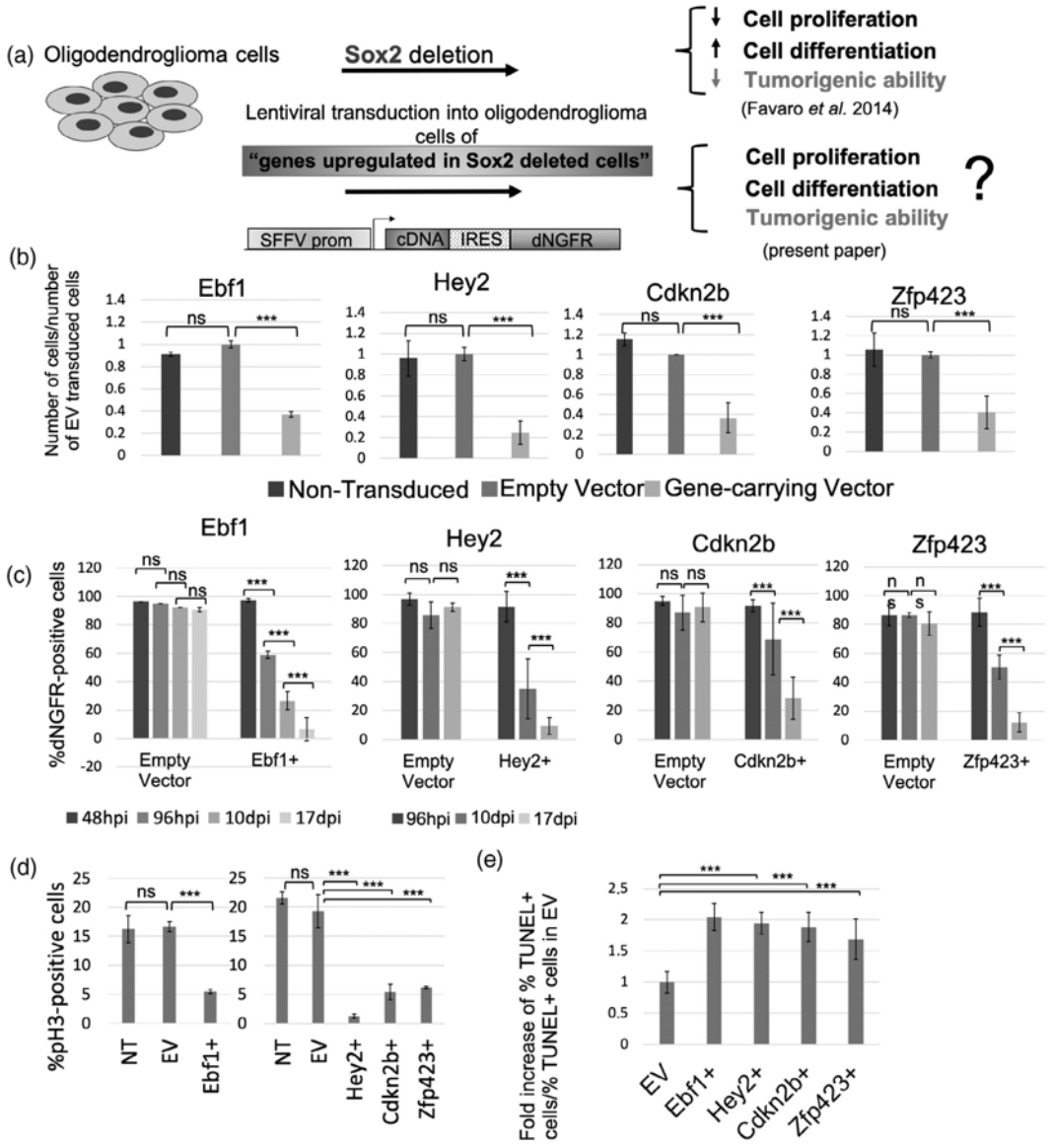


Figure 1 Transduction into pHGG oligodendroglioma cells of genes upregulated following Sox2 deletion reduces cellproliferation. (a) Scheme of the rationale of the experiment, and the lentiviral vector used for gene overexpression. (b) Cell numbers obtained following transduction of the cDNAs indicated above the histograms ("Gene-carrying Vector"), normalized to numbers obtained with Empty Vector (EV)- transduced cells, 96 hr after transduction. Error bars represent

mean ± SD from at least two independent experiments, each performed in triplicate (***p < .001; Paired T test). ns: non-significant. (c) Frequency (%) of dNGFR-positive (i.e., transduced) oligodendroglioma cells (out of total cells = 100%) transduced with lentiviruses expressing the indicated cDNAs, or with Empty Vector, at different time points from the transduction, indicated below the histograms. Hpi, hours post infection; dpi, days post infection. Error bars represent mean ± SD from at least two independent experiments, performed in triplicate (***p < .001, **p < .01, *p < .05 Paired T test). (d) Left, histograms: frequency (%) of cells undergoing mitosis (phospho-histone H3, pH 3-positive) within cDNA (Ebf1, or Hey2, Cdkn2b, Zfp423, on X axis)- or Empty Vector (EV)- transduced control cells. NT: non-transduced cells. Error bars represent mean ± SD from an experiment performed in triplicate. (***p < .001, **p < .01, *p < .05 Paired T test). (e) Fold increase of % of TUNEL-positive oligodendroglioma cells, after lentiviral transduction of the indicated cDNAs, relative to Empty Vector (EV, set = 1, corresponding to an average of 7.5% TUNEL-positive cells). Error bars represent mean ± SD from two experiments performed in triplicate. (***p < .001; Chi-square test).

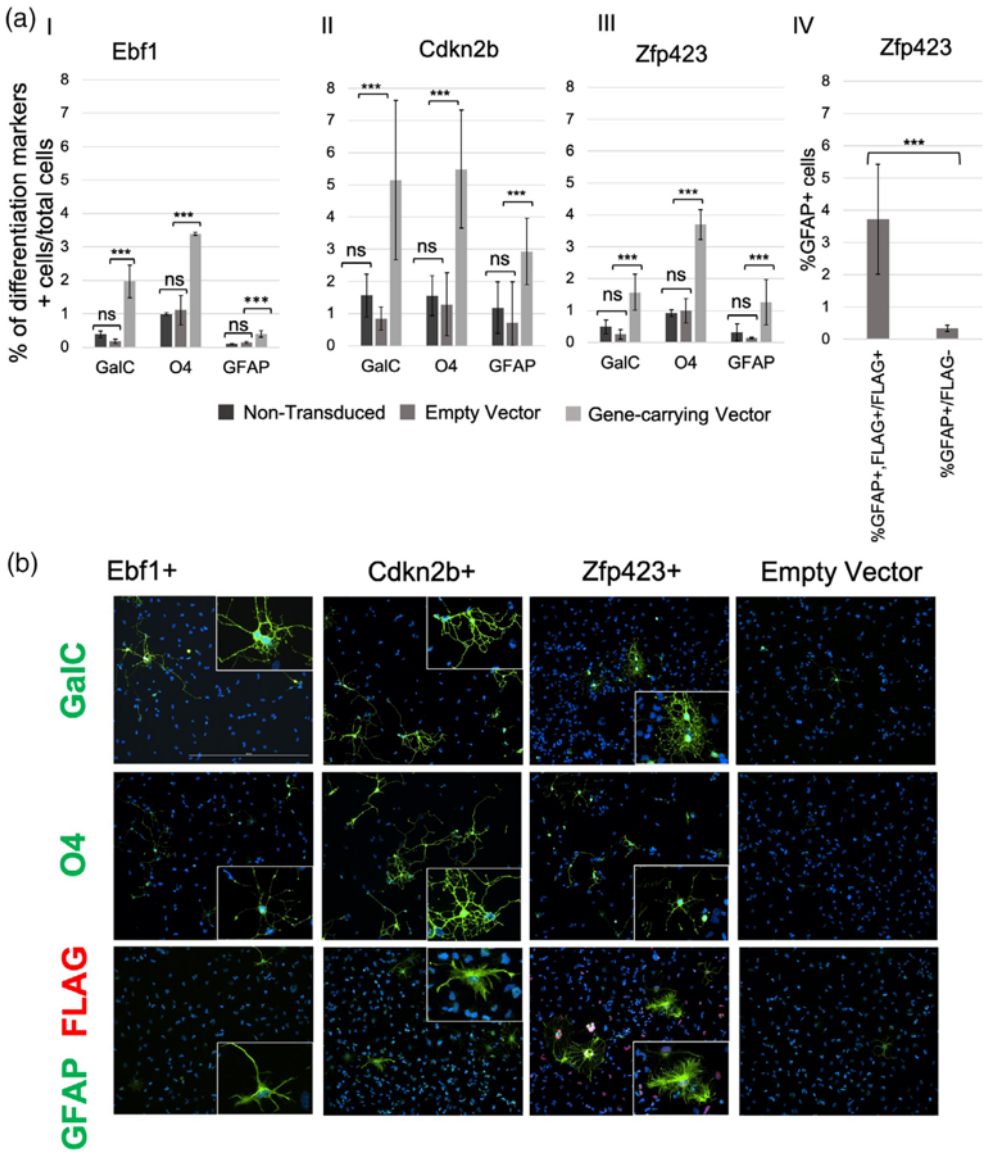


Figure 2 Transduction into pHGG cells of genes upregulated following Sox2 deletion induces glial differentiation markers and morphology in some cells. (a) I-II- III: Frequency (%) of cells immunopositive for the indicated differentiation markers (GalC, O4, GFAP) within cells transduced with the cDNAs above the

histograms, or with Empty Vector- transduced control cells. IV: Frequency of GFAP-positive, FLAG-positive (i.e., Zfp423-transduced) cells within total FLAG-positive cells, compared to the frequency of GFAP-positive cells within FLAG-negative cells. Error bars represent mean ± SD from at least two independent experiments, of which at least one performed in triplicate (***$p < .001$, Chi-square test). (b) Immunofluorescence for glial differentiation markers GalC, O4, GFAP (green) on cells transduced with the indicated cDNAs, or with Empty Vector; magnifications of individual cells are in the boxed areas. DAPI (blue) visualizes cell nuclei, after 7 days in culture after transduction, in serum-containing medium (scale bar, 100 um). Zfp423-overexpressing cells were double stained with antiGFAP (green) and antiFLAG (Red) antibodies.

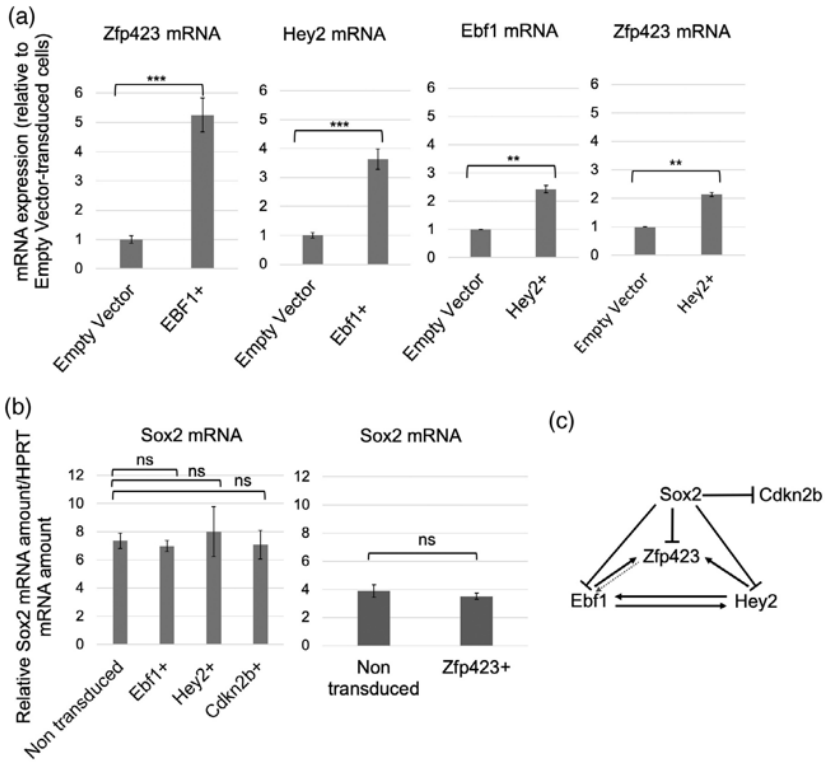


Figure 3 Transduction into pHGG cells of Ebf1 or Hey2 demonstrates cross-regulatory interactions between genes downstream to Sox2. (a) Zfp423, Hey2, Ebf1, and Zfp423 mRNA (qRT-PCR) in cells transduced with Empty Vector (control, set = 1) or Ebf1 (Ebf1+) or Hey2 (Hey2+) and collected 96 hr after transduction. Error bars represent mean ± SD from at least two independent experiments, performed in triplicate (***p < .001, **p < .01, Paired T test). Expression levels of the genes: the basal amount (in Empty Vector-transduced cells) of Sox2-targets mRNA relative to HPRT mRNA is: Ebf1, 0.61 ± 0.057; Cdkn2b, 0.35 ± 0.011; Hey2, 0.38 ± 0.029; Zfp423, 0.37 ± 0.0015. The values are the mean ± SD of two independent experiments, performed in triplicate. (b) Sox2 mRNA (qRT-PCR) (normalized to HPRT) in non-transduced cells or cells individually transduced with Ebf1 (Ebf1+), Hey2 (Hey2+), Cdkn2b (Cdkn2b+) (left), or Zfp423

(right). Error bars represent mean ± SD from at least two independent experiments, performed in triplicate. ns: non-significant (paired T test). C, Model for cross-regulatory interactions between Sox2, Zfp423, Ebf1, and Hey2.

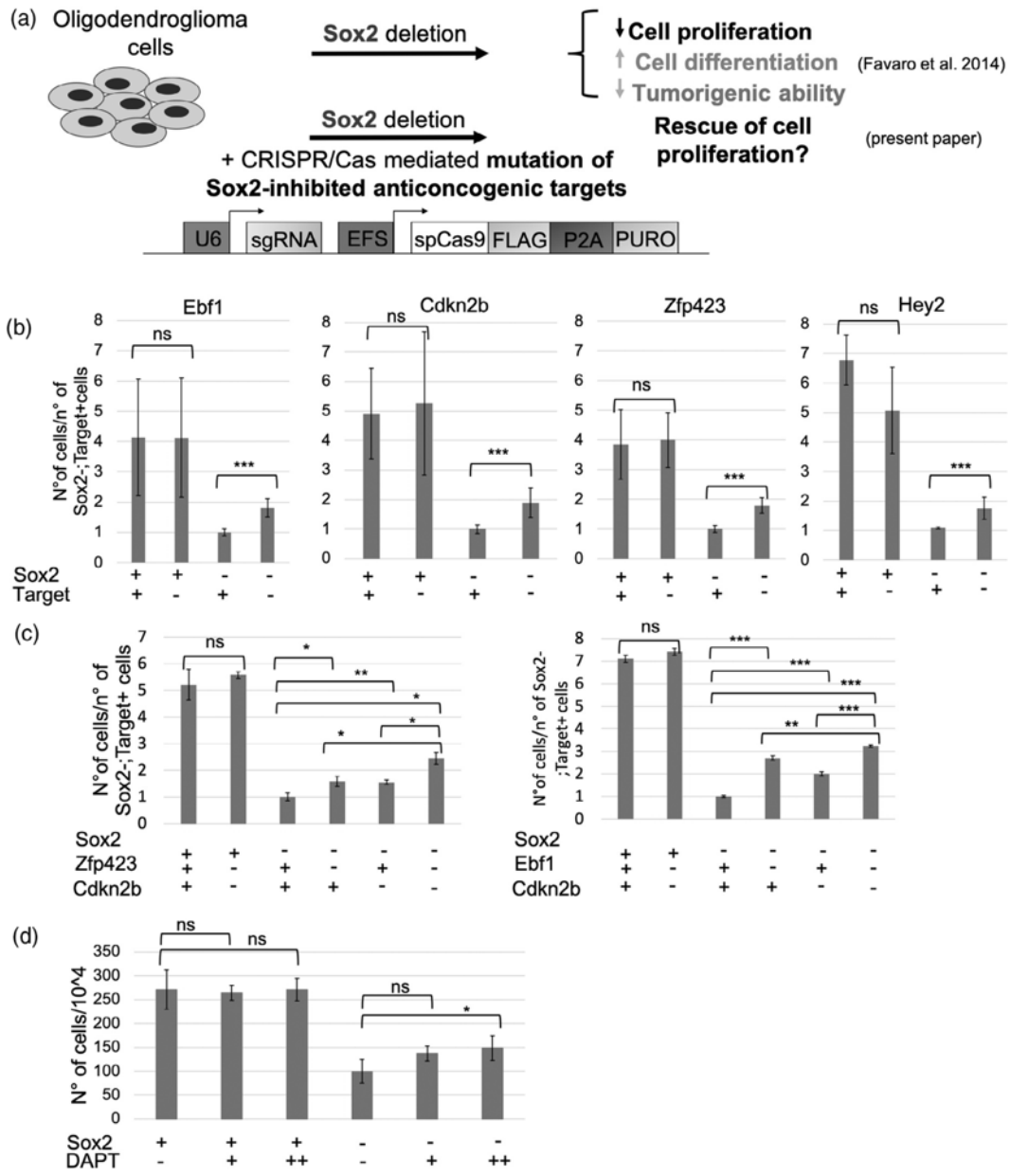
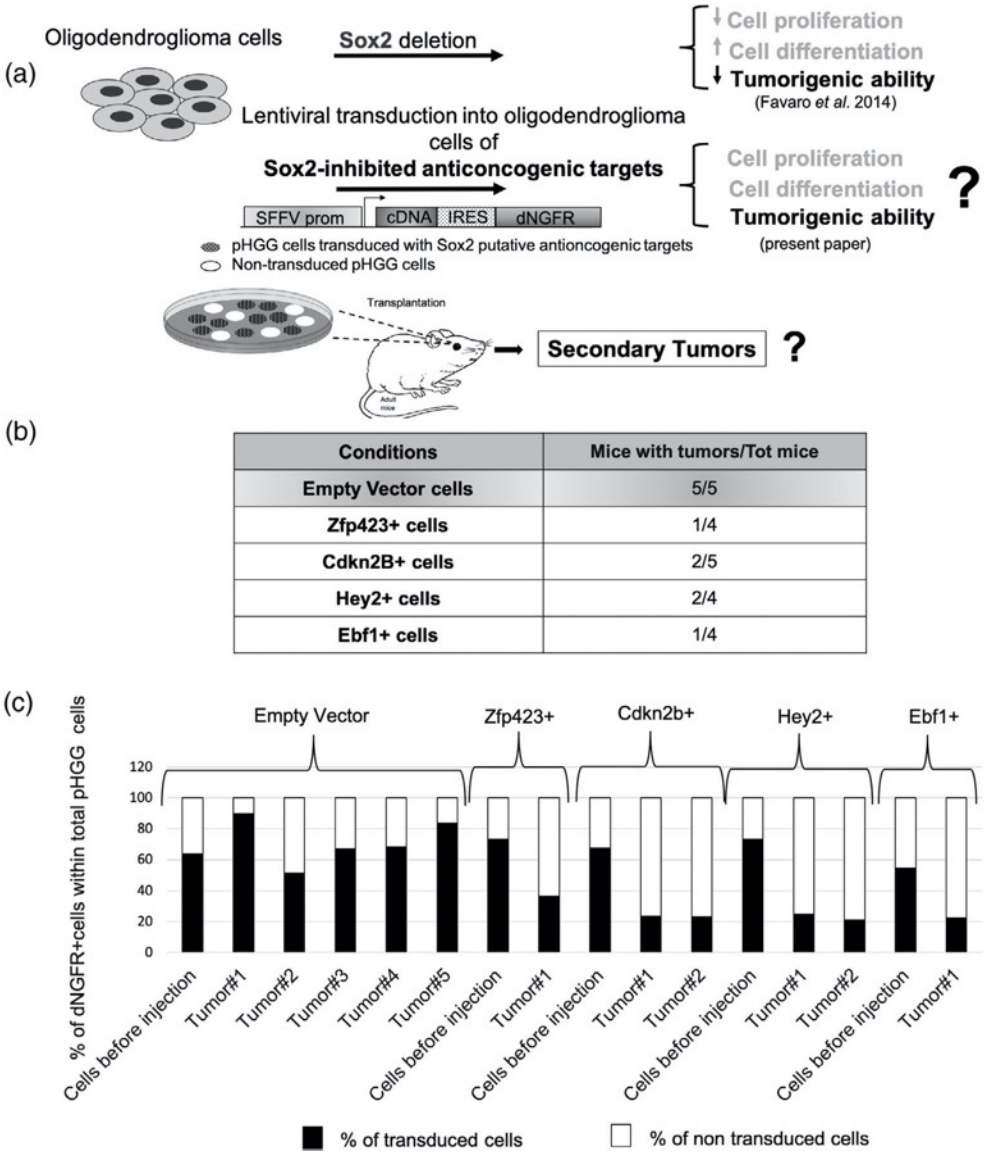


Figure 4 Mutation by CRISPR/Cas9, or pharmacological inhibition, of individual Sox2-inhibited tumor suppressive targets partially rescues cell proliferation of Sox2-deleted oligodendroglioma cells. (a) Scheme of the experiment. (b) Numbers of pHGG cells, mutated by CRISPR/Cas9 as indicated above the histograms, and CRE-treated (or not) to delete Sox2 (counting was 96 hr after Cre transduction). Presence of intact (+) or mutated (–) Sox2, and Cas9-targeted gene (Target +, intact; Target –, mutated), is indicated below the histograms. Cell numbers were normalized over the numbers of cells obtained in Sox2-, Target+ cells (set = 1). Error bars represent mean ± SD from three independent experiments each performed in triplicate (***p < .001, **p < .01, Paired T test). C, pHGG cell numbers, following mutagenesis of both Zfp423 and Cdkn2b (left) or both Ebf1 and Cdkn2b (right) Presence of intact (+) or deleted (–) Sox2 and indicated targets is reported below the histograms. The number of Sox2-, Zfp423+, Cdkn2b + cells (left) or Sox2-, Ebf1+, Cdkn2b + cells (right) is set = 1. Error bars represent mean ± SD from three independent experiments, each performed in triplicate (**p < .01, *p < .05, Paired T test). (d) pHGG cell numbers, following treatment with the Notch pathway inhibitor (DAPT, 2,5 μM [+] or 10 μM [++]) of cells carrying intact (+) or Cre-deleted (–) Sox2. Error bars represent mean ± SD from four experiments each performed in triplicate (*p < .05, Paired T test).



Conditions	Mice with tumors/Tot mice
Empty Vector cells	5/5
Zfp423+ cells	1/4
Cdkn2B+ cells	2/5
Hey2+ cells	2/4
Ebf1+ cells	1/4

Figure 5 Lentiviral transduction of Sox2-inhibited tumor suppressive targets antagonizes tumorigenesis in vivo. (a) Scheme of the experiment. (b) Fraction of mice developing secondary tumors following injection in the brain of cells transduced with empty vector, or the indicated cDNAs. (c) Frequency (%) of

dNGFR-positive (i.e., transduced) pHGG cells (black) before injection, and in different tumors obtained following injection.

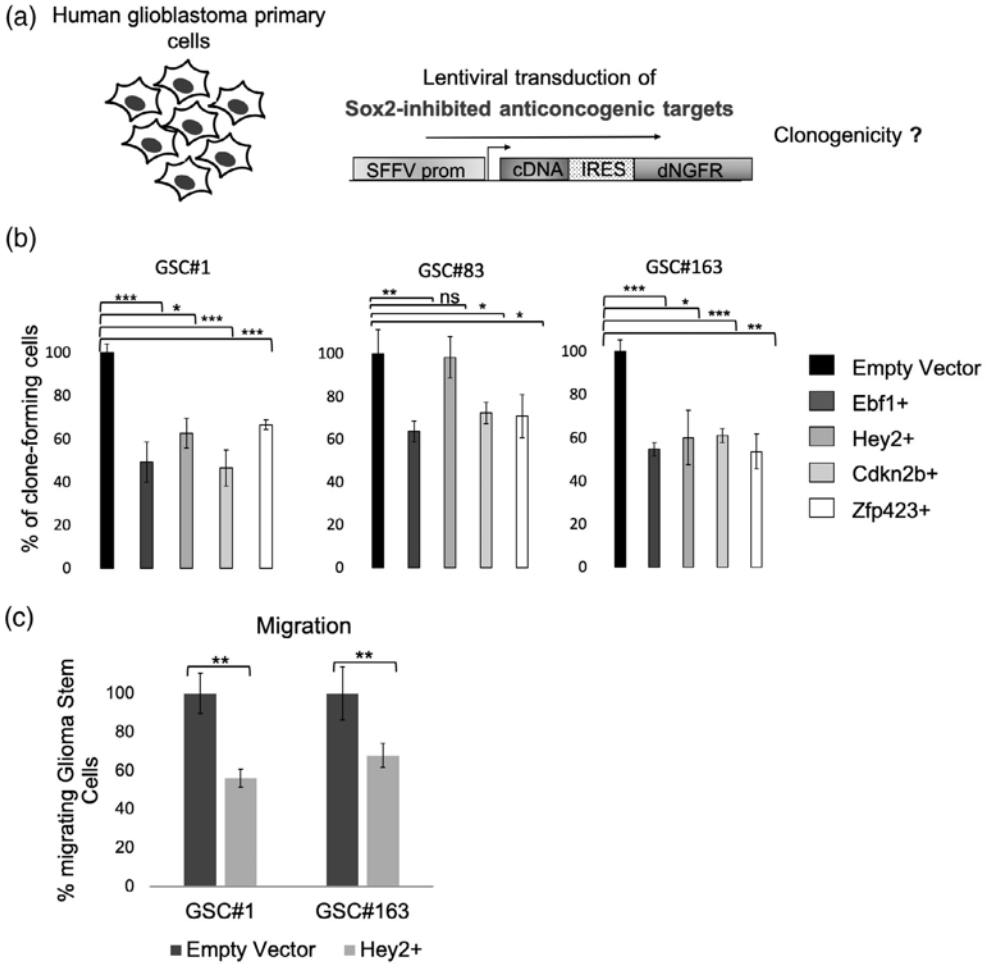
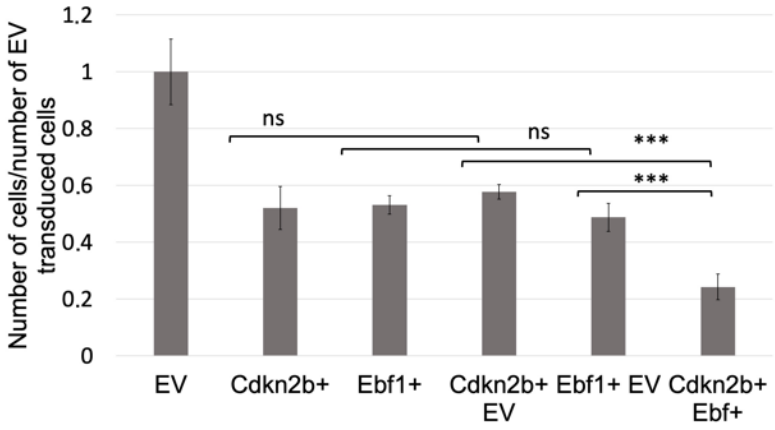


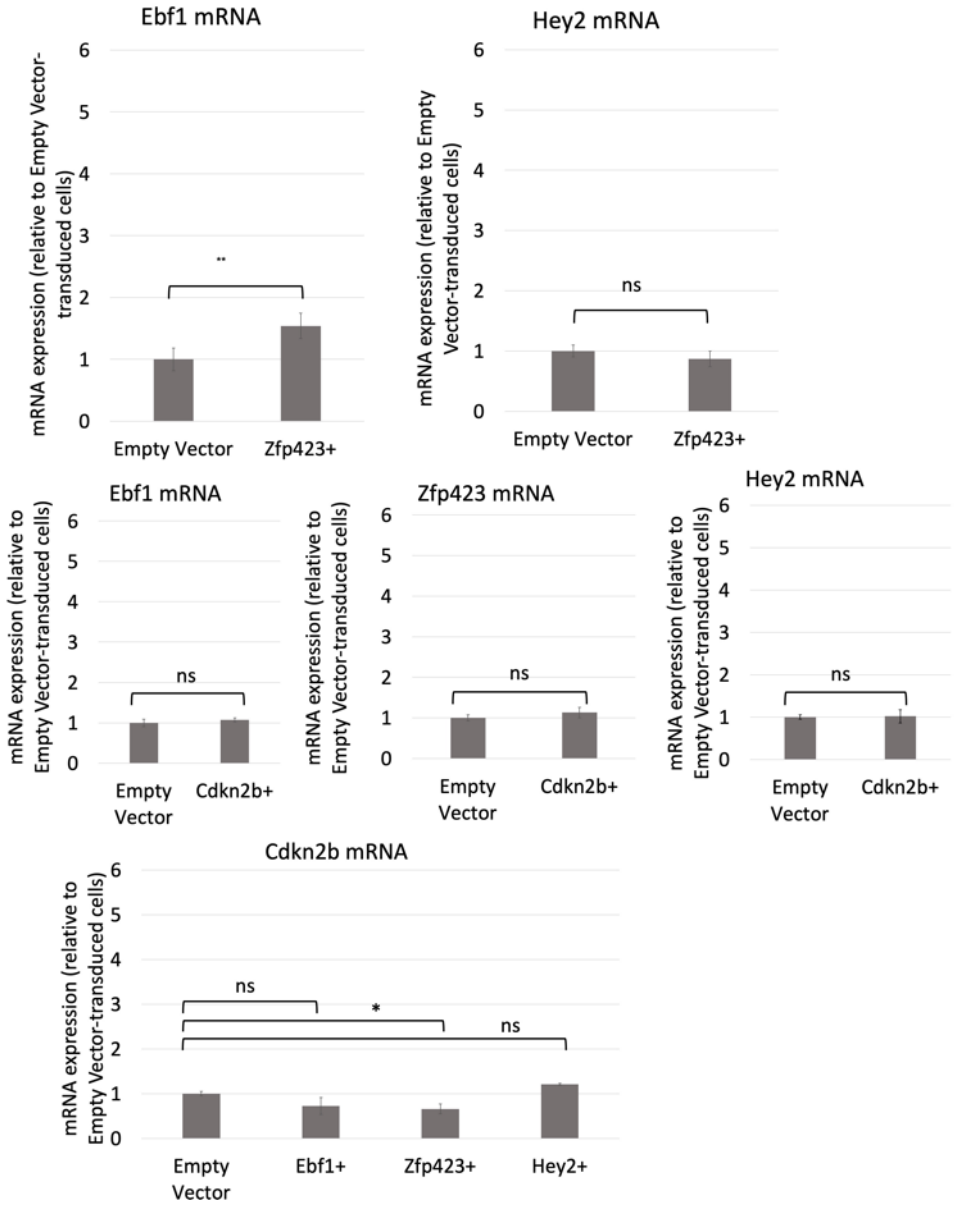
Figure 6 Lentiviral transduction of Sox2-inhibited tumor suppressive targets reduces clonogenicity and migration in different human patient-derived glioblastoma stem-like cell lines. (a) Scheme of the experiment. (b) Clonogenic ability of different cell lines transduced with the cDNAs indicated on the right, or Empty Vector. GSC, glioma stem-like cells. (c) Migration ability (% of migrated cells relative to Empty Vector) of cells transduced with Hey2, or Empty Vector. (b,

c) *p < .05; **p < .01; ***p < .001, Chi-square test. A representative experiment is shown, out of two independent experiments each performed in triplicate.

Supporting information

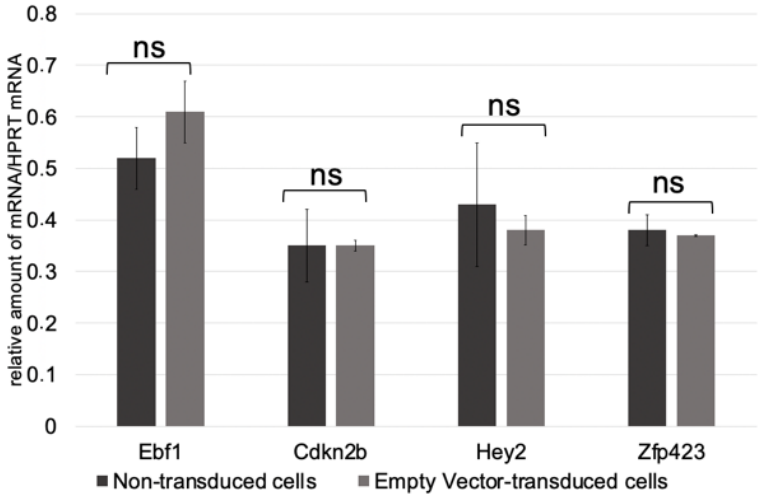


Supplementary Fig. 1 Cell numbers obtained following transduction of the cDNAs indicated below the bars ("Gene-carrying Vector"), normalized to numbers obtained with Empty Vector (EV)-transduced cells, 96 hours after transduction. Error bars represent mean ± SD from two independent experiments each performed in triplicate (***. P<0.001; Paired T test). ns: non-significant.



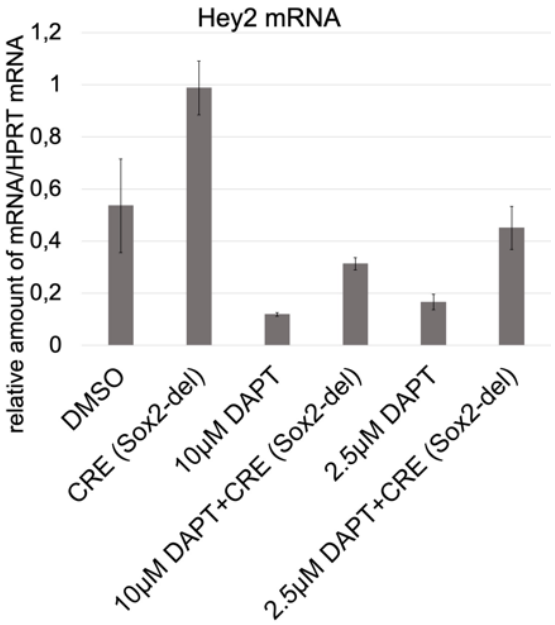
Supplementary Fig. 2 mRNA expression levels (relative to HPRT) of the genes indicated above the panels, in pHGG cells overexpressing the genes indicated under the histograms. The levels are given as fold-change relative to EV-transduced cells; samples were collected 96 hrs after transduction. Error bars represent

mean ± SD from two independent experiments performed in triplicate (**.P<0.01.*.P<0.05 Paired T test). ns: non-significant.



Supplementary Fig. 3

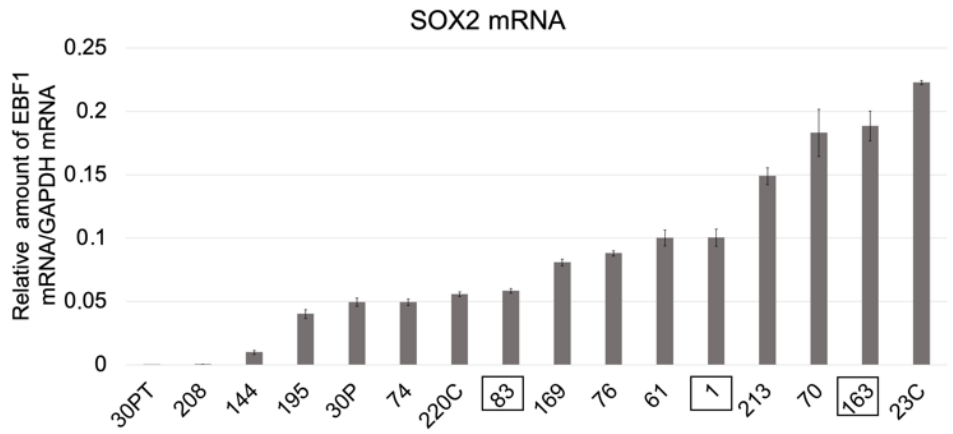
mRNA expression levels (relative to HPRT) of the genes indicated under the histograms, in pHGG cells non-transduced (dark grey) or tranduced with empty vector (light gray). ns: non significant.



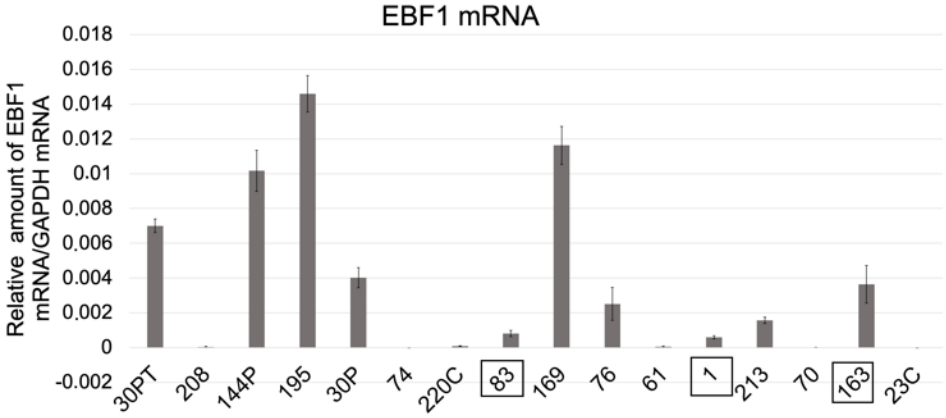
Supplementary Fig. 4

Hey2 mRNA levels in DAPT-treated and untreated control cells.

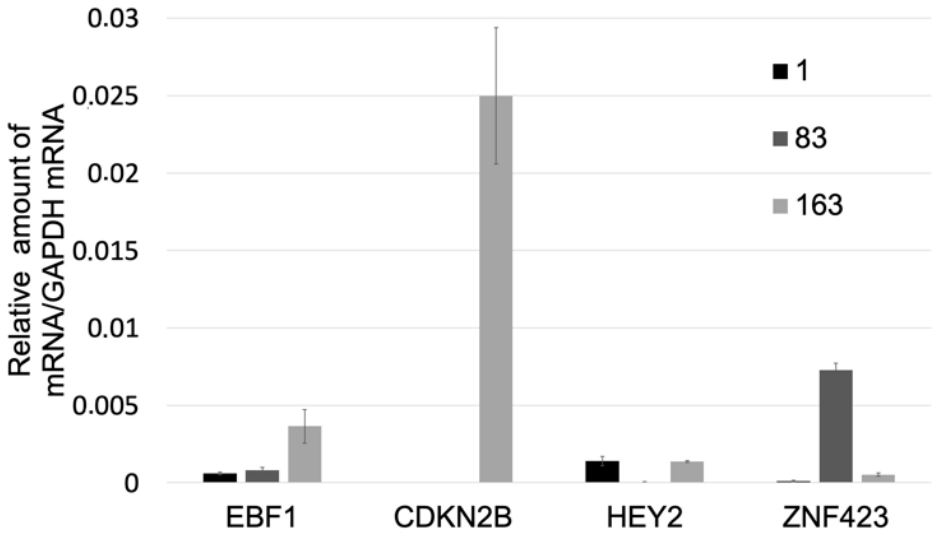
A


SOX2 mRNA
Relative amount of EBF1 mRNA/GAPDH mRNA
0.25
0.2
0.15
0.1
0.05
0
30PT
208
144
195
30P
74
220C
83
169
76
61
1
213
70
163
23C

B


EBF1 mRNA
Relative amount of EBF1 mRNA/GAPDH mRNA
0.018
0.016
0.014
0.012
0.01
0.008
0.006
0.004
0.002
0
-0.002
30PT
208
144P
195
30P
74
220C
83
169
76
61
1
213
70
163
23C

C



Supplementary Fig. 5

A. SOX2 mRNA levels (relative to GAPDH) in different human primary glioblastoma cells. Lines boxed in black are studied in Fig. 6.

B. EBF1 mRNA levels (relative to GAPDH) in different human primary glioblastoma cells. Lines are ordered by increasing amounts of SOX2 (See also Suppl. Table 2).

C. Endogenous EBF1, CDKN2B, HEY2 and ZNF423 mRNA levels (relative to GAPDH) in the human primary glioblastoma cell lines used in Fig. 6.

Ranking	Gene	Description	Fold Increase Cre96h/NT	p-values*
1	Hopx	HOP homeobox	12.37654	0.005821101
2	Sdc4	syndecan 4	8.197359	0.005821101
3	Wif1	Wnt inhibitory factor 1	7.12466	0.003497308
4	Rgs2	regulator of G-protein signaling 2	7.109481	0.003497308
5	Ebf1	early B-cell factor 1	7.023428	0.003497308
6	Irf6	interferon regulatory factor 6	6.775218	0.000775201
7	Pcp4	Purkinje cell protein 4	6.071423	0.003497308
8	Ttr	transthyretin	5.957583	0.005821101
9	Slc14a1	solute carrier family 14 (urea transporter). member 1	5.561991	0.003497308
10	Tnfaip6	tumor necrosis factor alpha induced protein 6	5.521515	0.017074448
11	Cryab	crystallin. alpha B	5.353396	0.022243778
12	Cntnap2	contactin associated protein-like 2	5.028883	0.003497308
13	Rnasel	ribonuclease L (2'. 5'-oligoisoadenylate synthetase-dependent)	4.673196	0.019143448
14	Calml4	calmodulin-like 4	4.387243	0.01728057
15	Ttr	transthyretin	4.386028	0.006584671
16	Nfe2	nuclear factor. erythroid derived 2	4.254731	0.014111669
17	Shisa2	shisa homolog 2 (Xenopus laevis)	4.197811	0.006584671
18	Tnc	tenascin C	4.171262	0.024512824
19	Hey2	hairy/enhancer-of-split related with YRPW motif 2	4.09025	0.005821101
20	Gstt1	glutathione S-transferase. theta 1	4.04022	0.006584671
21	Trp53inp1	transformation related protein 53 inducible nuclear protein 1	4.039775	0.016600061
22	Vasn	vasorin	3.993707	0.043593647
23	Crlf1	cytokine receptor-like factor 1	3.954931	0.022533083
24	Tgfb2	transforming growth factor. beta 2	3.879516	0.020027034
25	Cdkn2b	cyclin-dependent kinase inhibitor 2B (p15. inhibits CDK4)	3.832073	0.023211677

26	Slc26a2	solute carrier family 26 (sulfate transporter). member 2	3.818559	0.031908678
27	Limch1	LIM and calponin homology domains 1	3.744043	0.034944017
28	Cdc42ep3	CDC42 effector protein (Rho GTPase binding) 3	3.666592	0.037979355
29	Adhfe1	alcohol dehydrogenase. iron containing. 1	3.634825	0.041014694
30	Arc	activity regulated cytoskeletal-associated protein	3.519879	0.044050032

Supplementary table 1 Top 30 upregulated genes in Sox2-deleted pHGG cells, compared to non-deleted cells (data from Favaro et al., 2014). *false-discovery rate correction of p-values were applied; differentially expressed genes with p-values<0.05 were selected (see Favaro et al., 2014).

Human glioblastoma cell lines	SOX2 mRNA relative to GAPDH mRNA (by qRT-PCR)	% SOX2(protein)-positive cells (by Cytofluorimetric assay)	Mean of SOX2-fluorescence (by Cytofluorimetric assay)
30PT	0.000017291	47.4	4.9
208	0.000451238	2	1.25
144P	0.009809254	45.7	2.6
195	0.040028347	20.7	1.82
30P	0.049329329	47.4	4.9
74	0.049360763	30.95	2.68
220C	0.055629892	79.6	8.4
83P	0.058343361	77.7	3.9
169	0.080738768	47.4	7.04
76	0.08788486	88.4	14.9
61	0.100006601	81.6	4.95
1	0.100328146	95.2	18.4
213	0.148886851	92	37.2
70	0.183084501	38	5.7
163	0.188432125	89.2	17.3
23C	0.222796757	97.3	40.1

Supplementary Table 2 SOX2 expression (mRNA and protein) in human GSC lines (see also Supplementary Fig. 4); the lines studied in Fig. 6 are in bold type. Data from D'Alessandris et al., 2017.

Gene	Forward/Reverse	Sequence
Ebf1	Forward	ACCATGGACTACAAGGACGA
Ebf1	Reverse	TCACATGGGAGGGACAATCAT
Hey2	Forward	ATATGGATCCCAGTAGCTGCTCCTCCTTCG
Hey2	Reverse	ATATGGATCCATTGCTGCTGTGTGGAACTG
Cdkn2b	Forward	ATATGGATCCCGAAGGACCATTTCTGCCAC
Cdkn2b	Reverse	ATATGGATCCTCGTGCTTGCAGTCTTCCTA

Supplementary Table 3 PCR Primers (see Materials and Methods)

Gene	Forward/Reverse	Sequence
hEBF1	Forward	CCTGGTGTTGTGGAAGTCACA
hEBF1	Reverse	GATGGTGGGTTCGTTGAGC
mEbf1	Forward	ACCCTGAAATGTGCCGAGTATT
mEbf1	Reverse	GGGTTTCCTGCATTCTTTAGGC
hHEY2	Forward	TGGGGAGCGAGAACAATTAC
hHEY2	Reverse	TTTTCAAAAGCAGTTGGCACA
mHey2	Forward	TGGGGAGCGAGAACAATTACC
mHey2	Reverse	CCCTCTCCTTTTCTTTCTTGCC
hCDKN2B	Forward	CACCCCCACCCACCTAATTC
hCDKN2B	Reverse	TGAGTGTCGAGGGCCAGATA
mCdkn2b	Forward	GCCCAATCCAGGTCATGATGAT
mCdkn2b	Reverse	ATACCTCGCAATGTCACGGTG
hZNF423	Forward	CTTCTCGCTGGCCTGGGATT
hZNF423	Reverse	GGTCTGCCAGAGACTCGAAGT
mZfp423	Forward	ATGTCCAGGCGGAAGCAG
mZfp423	Reverse	TTTCCGATCACACTCTGGCT
hSOX2	Forward	AACCCCAAGATGCACAACTC
hSOX2	Reverse	CGGGGCCGGTATTTATAATC
mSox2	Forward	GGCAGCTACAGCATGATGCAGGAGC
mSox2	Reverse	CTGGTCATGGAGTTGTACTGCAGG
hGAPDH	Forward	ACGGATTTGGTCGTATTGGG
hGAPDH	Reverse	TGATTTTGGAGGGATCTCGC
mHprt	Forward	TCCTCCTCAGACCGGTTT
mHprt	Reverse	CCTGGTTCATTCATCGCTAATC

Supplementary table 4 RT-PCR Primers (see Materials and Methods). h: human; m: mouse.

Gene	Forward/Reverse	Sequence
Ebf1	Forward	GAGTGGCATTTGTCCGGTTC
Ebf1	Reverse	TTCTGAGCCCGGGACTACTA
Cdkn2b	Forward	CCAATCTAGTGCCGAGGGAT
Cdkn2b	Reverse	CTCACCGAAGCTACTGGGTC
Zfp423	Forward	TGAAGCCTAATTGCCCCTGA
Zfp423	Reverse	CCCTTGGGAAGTGGCCTATG
Hey2	Forward	CTTCTACGCCGGATCAGAGT
Hey2	Reverse	GCCAACTGCCTTTACTTGCT

Supplementary table 5 PCR Primers (see Materials and Methods)

Chapter 3

Review published**,** ***Insight of Neuro-Oncology,*** 13 oct 2018,
DOI: 10.36959/828/332

Sox2 Functions in Neural Cancer Stem Cells: The Importance of the Context

Cristiana Barone§, Miriam Pagin§, Linda Serra, Alessia Motta, Laura Rigoldi, Simone Giubbolini, Alexandra Badiola-Sanga, Sara Mercurio and Silvia K Nicolis*
Department of Biotechnology and Biosciences, University of Milano-Bicocca, Italy §Joint first authors

Abstract
The Sox2 transcription factor is expressed in different neural tumors. In particular, it is active within the "cancer stem cell" (CSC) subpopulation of tumor cells, able to reinitiate tumorigenesis after conventional chemotherapy (to which it is usually resistant). This led to hypothesize that Sox2 (and its downstream regulated genes) may qualify as promising targets for therapeutic strategies directed against CSC. However, the potential relevance of Sox2 in this regard depends on whether it is functionally important to maintain CSC. Here, we comparatively examine the effects of Sox2 genetic ablation within mouse models of different neural tumor types. Sox2 ablation in mouse glioma (and in human glioblastoma- derived CSC) demonstrated a critical function for Sox2 in the maintenance of CSC. Surprisingly, however, Sox2 ablation in two different mouse models of melanoma (a neural crest-related tumor), and in a mouse model of medulloblastoma of the Sonic Hedgehog subgroup, showed that, in these contexts, Sox2 is dispensable for tumorigenesis.

This heterogeneous situation has a parallel in the normal development of the nervous system, where generalized Sox2 ablation in neural stem/ progenitor cells selectively affects the development of some neural regions, but not other ones. Molecular mechanisms underlying these specificities may involve the regulation, by Sox2, of different sets of target genes in different tumors, but also a redundant regulation of the same targets by different Sox transcription factors, differentially coexpressed with Sox2 in different tumors. Collectively, these findings point to the need to experimentally address the requirement for Sox2, and its downstream targets, within different tumor types, as a prerequisite to fully exploit its potential as a target for novel therapeutic approaches.

Keywords

Sox2, Cancer stem cells, Cancer, Glioblastoma, Oligodendroglioma, Medulloblastoma, Melanoma, Mouse genetic models

Abbreviations

CSC: Cancer Stem Cells; ESFT: Ewing Sarcoma Family Tumors; MB: Medulloblastoma; NSC: Neural Stem Cells; SCC: Squamous Cell Carcinoma; SCLC: Small Cell Lung Carcinoma

1. Introduction

Stem cells are “cells that have the ability to perpetuate themselves through self-renewal and to generate mature cells of a particular tissue through differentiation” Reya et al.,2001). “Cancer stem cells” (CSC) were identified in human brain tumors (Reya et al., 2001; Nicolis, 2007; Singh et al., 2004D), as a minority subpopulation able to reinitiate tumor development following conventional chemotherapy (to which they are usually resist- ant), or, experimentally, following transplantation in a host (mouse) brain. Much evidence has accumulated supporting the importance

of CSC for tumor relapse and propagation (Nicolis, 2007; Beck and Blanpain, 2013; Pattabiraman and Weinberg, 2014) and CSC are now considered a central target for therapeutic approaches aimed at eradicating tumor development. CSC self-renew, and also produce "pseudo-differentiated" cells, constituting the tumor bulk (Chen et al., 2012; Suva et al., 2014; Vanner et al., 2014). While most of the tumor bulk cells are typically proliferating (though they are called "differentiated" from the CSC perspective), CSC can be slowly proliferating and even "quiescent" (Nicolis, 2007; Vanner et al., 2014), making the use of drugs hitting proliferating cells uneffective in killing them.

Sox2 is a member of the *S*ry-related HMG-b*ox* (SOX) family of transcription factors; the Sox2 gene is located on chromosome 3, in both humans and mice, and in both species it is constituted by a single coding exon (Kondoh and Robin, 2016). Sox2 was first studied in the context of normal embryonic development, where knock-out experiments in mouse demonstrated its essential role to maintain the pluripotent stem cells of the early embryo (Avilion et al., 2003), as well as its function in several tissue-specific stem cells, including neural stem cells (NSC); NSC cultured *in vitro* from the mutant neonatal mouse brain fail to self-renew in long- term culture, and, *in vivo*, postnatal hippocampal NSC are impaired (Arnold et al.,2011; Bertolini et al., 2016; Favaro et al., 2009; Pevny and Nicolis, 2010).

The discovery of CSC in tumors focused attention on Sox2 also from the point of view of this pathological stem cell type. Indeed, Sox2 is expressed in different tumors of neural origin, such as gliomas (the most common primary brain tumors, whose most malignant and lethal subtype is glioblastoma multiforme), medulloblastomas (the most common brain tumor in childhood), and melanoma (a tumor arising from neural crest type cells); In all of these tumors, CSC have been identified, and found to express Sox2 (Nicolis, 2007; Vanner et al., 2014; Garros-Regulez et al., 2016).

In this paper, we focus on the comparative review of recent experiments, that made use of conditional mutation of Sox2 in different mouse models of neural tumors (and of genetic Sox2 ablation in human CSC-enriched cultures of the same tumor types), to investigate the functional relevance of Sox2 in tumorigenesis, and, in particular, in the maintenance of CSC. These experiments demonstrated an absolute requirement for Sox2 of CSC able to reinitiate tumorigenesis of gliomas (glioblastoma and oligodendroglioma) in mouse; unexpectedly, how- ever, they also showed that Sox2 is dispensable for tumorigenesis in a model of Sonic Hedgehog (SHH)-sub-group medulloblastoma, and in two different models of melanoma. We discuss a parallel of this situation with the identification of region-specific Sox2 functions in the context of the development of the normal neuroepithelium, and possible molecular mechanisms underlying the context-specificity of Sox2 functions. Finally, we discuss the need and approaches to identify functional downstream effectors of Sox2 in Sox2-dependent CSC, that could complement Sox2 as targets of therapeutic strategies directed against CSC.

2. Sox2-dependent neural cancer stem cells in gliomas

Gliomas are the most common cerebral neoplasias (86%) (Lau et al., 2017; Rasmussen et al., 2017; Reifenberger et al., 2017). Glioblastoma multiforme (GBM), the most aggressive and deadly among gliomas (average patients' survival is about 15 months), was one of the first tumor types in which CSC were originally described (reviewed in [Nicolis, 2007]) The development of serum-free *in vitro* cultures of tumor-derived cells has allowed to expand CSC (in equilibrium with more differentiated progeny) from GBM tissue of many different patients; these cells retain the ability to re-form a tumor with the same characteristics of the tumor of origin following transplantation into a host mouse brain (xenograft), and thus

represent important *in vitro* models of CSC. GBM, and CSC-enriched cultures derived from it, consistently ex- press Sox2 (Lee et al., 2006).

Gangemi, et al. (2009) first addressed the consequences of lowering Sox2 levels in some patient-derived CSC-enriched cell cultures, by expressing anti-Sox2 shRNAs causing strong reduction of Sox2 mRNA levels. This resulted in reduced cell proliferation and reduced clonogenicity *in vitro*, and to loss of tumorigenicity *in vivo*, in mouse xenografts (Gangemi et al., 2009). A role for Sox2 in maintaining other patient-specific GBM CSC was further supported in additional important work by other laboratories (Garros-Regulez et al., 2016).

In mouse cells, complete Sox2 ablation was obtained *in vitro* by conditional knockout in a genetically defined model of glioma: a high-grade oligodendroglioma induced by overexpression of Platelet-derived growth fac- tor B (PDGF-B) (Favaro et al., 2014). Oligodendroglioma is the second most common tumor in adults; patients with high-grade oligodendroglioma have a median survival of 3-4 years (Lau et al., 2017; Reis-Filho et al., 2000). Alterations of PDGF-B signaling are common molecular lesions in human gliomas, including oligodendrogliomas: PDGF and PDGF receptor have both been found overexpressed in glial tumor-derived cells and glioma surgical samples, and amplification of the gene encoding the PDGF receptor-A occurs in high-grade oligodendrogliomas (Appolloni et al., 2009; Jackson et al., 2006; Shih and Holland, 2006). In addition, PDGF can initiate the "reprogramming" of normal, committed O2A oligodendrocyte progenitors to neural stem-like cells, in a process that requires Sox2 (Kondo and Raff, 2004). PDGF-B was used to induce tumor development in embryonic brain; cells cultured from such tumors would re-initiate tumor development following transplantation, thus behaving as CSC. Sox2 ablation was achieved by conditional knock- out, in which Cre recombinases delete the endogenous Sox2 gene, that has been previously flanked by loxP sites, the Cre substrate ("Sox2 flox" allele) (Favaro et al., 2009).

Sox2 Cre-mediated ablation *in vitro* in such glioma cells completely prevented tumor reinitiation following *in vivo* trans- plantation; over the time window in which non-deleted cells developed deadly tumors, mice transplanted with Sox2-deleted cells remained tumor free. The only two tumors developing from Cre-treated cells were SOX2-positive, demonstrating they were derived from non-deleted cells. *In vitro*, Sox2 ablation caused proliferation reduction, apoptosis activation, and aberrant
differentiation into cells expressing oligodendrocyte and astrocyte markers (Favaro et al., 2014). Counterintuitively, Sox2, or pep- tide fragments of it, though being a nuclear protein, can be found exposed on the cell surface of tumor cells, in association with the Major Histocompatibility Complex (MHC), presumably as a result of tumor cell lysis and processing of its proteins; based on this, immunotherapy against SOX2 protein was attempted, which resulted in a doubling of survival time of mice transplanted with oligodendroglioma cells (Favaro et al., 2014).

3. Sox2 cooperating genes

Recent experiments investigated transcription fac- tors cooperating with Sox2 in cell "reprogramming" to tumor-propagating cells (cancer stem cells). In recent work (Suva et al., 2014), SOX2 was able to reprogram "differentiated" GBM cells from human tumors (DCGs) to stem-like tumor-propagating cells (TPC), when transduced into the DCGs together with transcription factors POU3F2 (BRN2), SALL2, and OLIG2. In GBM-derived CSC-enriched cultures grown in serum-free conditions, genomic mapping (by ChIPseq) of H3K27Ac, a histone modification carried by active enhancers and promoters, had initially revealed regulatory elements specific of TPC (versus DCGs), and these were enriched in the DNA

sequence recognized by SOX proteins. RNA seq showed that expression of SOX2 (and of SOX1, SOX5, SOX8, SOX21) was higher in TPC than in DCGs ("differentiated" by serum or BMP4 addition). Similarly, transcription factors SALL2, POU3F, and OLIG2 were selected based on two considerations: enrichment, in TPC versus DCGs, of their expression, and the presence of H3K27Ac (an epigenetic mark of transcriptionally active state) on their binding sites on gene promoters and enhancers. The combination of SOX2 (but not SOX1), SALL2, POU3F and OLIG2 (or Rest Corepressor 2, RCOR2, a transcriptional corepressor) could reprogram DCGs to TPCs, that carried a genome-wide pattern of H3K27Ac sites superimposable to that of the TPCs that had been directly grown from the tumor. Further, the 4 factors co-bind to a large number of distal regulatory elements specifically active in TPCs. Interestingly, a minority of cells co-expressed the 4 transcription factors in the tumor, as demonstrated by immunofluorescence. Looking for therapeutic implications, two mediators acting downstream to the 4 transcription factors were also identified: RCOR2 (that can replace OLIG2 in the reprogramming cocktail) and histone demethylase LSD1, whose repression caused cell death specifically in TPCs. Interestingly, it had previously been found that LSD1-specific inhibitors impaired the growth of Sox2-expressing, but not that of Sox2-negative, lung squamous cell carcinomas (SCC), and Sox2 expression was associated with sensitivity to LSD1 inhibition in lung, breast, ovarian, and other carcinoma cells (Zhang et al., 2013). This indicates that LSD1 is a mediator of tumorigenic effects downstream to Sox2, but not other tumorigenic factors.

In a related study, ChIPseq in GBM cells showed that DNAase hypersensitive sites in patient-derived GBM cells, mapped by ATAC-seq (a technique that allows to map DNA regions that are more accessible in chromatin, usually in correlation with the binding of transcription factors), are enriched in DNA recognition sequences for SOX2 and FOXG1, a transcription factor active in embryonic brain

development; ectopic expression of SOX2 and FOXG1 in postmitotic astrocytes reactivated proliferation and stem cell properties. The coexpression of FOXG1 with SOX2 in GBM led to hypothesize that they could, together, contribute to cell "reprogramming" to stemness in gliomagenesis (Bulstrode et al., 2017).

4. Sox2 relevance for non-neural tumors. Common downstream Sox2 effectors?

Importantly, conditional knockout and RNA interference experiments showed that SOX2 controls CSS functions also in some non-neural tumours, such as in skin, lung, and aesophagus squamous cell carcinomas (SCC), osteosarcomas, Ewing sarcoma, and small cell lung carcinomas (SCLC) (Bass et al., 2009; Basu-Roy et al., 2012; Boumahdi et al., 2014; Riggi et al., 2010; Rudin et al., 2012). This suggests that some downstream effectors of Sox2 function may be conserved between different tumors. In this regard, in osteosarcoma-initiating cells, Sox2 directly represses the genes encoding two activators of the "hippo" signaling pathway, Nf2 (Neurofibromin 2, also called Merlin, encoding a protein involved in connecting the cytoskeleton with proteins of the cell membrane) and WWC1 (also called Kibra, encoding a cytoplasmic phosphoprotein), which in turn negatively regulate the transcriptional co-activator YAP, important for promoting tumor growth; interestingly, Sox2 depletion led to upregulation of Nf2 and WWC1, downregulation of YAP, and to reduced cell clonogenicity, in both osteosarcoma and GBM cells, indicating that these effectors may be shared between these tumor types (Basu-Roy et al.,2012). In osteosarcomas, as well as in laryngeal cancer cells, Sox2 was also proposed to control migration and invasion via the Wnt/beta-catenin signalling pathway (Tang et al., 2018; Yang et al., 2014), though enforced Sox2 expression in lung adenocarcinoma was reported to promote cell migration and

invasion, but to inhibit Wnt/beta-catenin signaling activity (He et al., 2017). Thus, Sox2 might regulate the Wnt/beta-catenin pathway in different tumor types, though not necessarily with the same functional outcome. In hematopoietic tumors, Sox2 expression was detected in cultured cells isolated from ALK-positive an- aplastic large cell lymphoma, and Sox2 downregulation impaired their clonogenicity and tumorigenic ability; oxidative stress increased Sox2 expression and cancer stem cell properties in a subpopulation of cells, and Sox2 was reported to bind DNA more efficiently (Wu et al., 2018). It is possible that specific partners had become available in oxy- gen-stressed cells, or that increased levels of Sox2 were important for binding; alternatively, posttranslational modifications of Sox2 were proposed to be involved in this "activation" following oxidative stress (Gupta et al., 2018).

Interestingly, some papers reported Sox2 expression and some functional effects of Sox2 downregulation also in human mammary cancer cell lines (Leis et al., 2012; Wu et al., 2012). However, in a widely studied mouse model of breast cancer, produced by expression of a transgene encoding a mutated ErbB2/Neu oncogene driven to mammary tissue by the MMTV promoter (Guy et al., 1992; Muller et al.,1988), and activated by a mammary-specific Cre recombinase (Wagner et al., 1997), Sox2 ablation had no effect on tumorigenesis (A.B.S. and S.K.N., unpublished observations). It is possible that breast tumors are heterogeneous regarding functional requirements for Sox2 in tumorigenesis. Indeed, Sox2 expression was observed in some breast tumors (mostly belonging to early stages of tumor progression), but not in others, and some, but not all breast tumor-derived stem cell-enriched cultured cell lines are reported to express Sox2 (Leis et al., 2012). A future characterization of the gene regulatory networks acting in breast cancer CSC might allow to better categorize them with respect to Sox2 function.

It is interesting to note that Sox2 is important also in the normal cell counterparts of CSC within several of the non-neural tissues (see above) known to develop Sox2-

dependent tumors, as previously seen with NSC (Favaro et al., 2009): indeed, normal osteoblasts, dermal papilla cells, and cells of the developing foregut (giving rise to lung, esophagus and trachaea) require Sox2 function (Basu-Roy et al., 2010; Clavel et al., 2012; Domyan et al., 2011). This indicates that Sox2 function, already present in tissue-specific stem/progenitor cells, is retained by CSC of (at least some) tumors of the same tissue type. A different situation was documented in Ewing Sarcoma Family Tumors (ESFT), mesenchimal tumors thought to arise from primary mesenchymal stem cells. Here, SOX2 expression is strongly activated *de novo* by the oncogenic transcription factor that characterizes a high proportion of ESFT, encoded by the EWS-FLI-1 fusion gene. EWS-FLI-1 activates SOX2 (together with OCT4 and NANOG), and SOX2 is a key factor in the emergence of a ESFT CSC population; its downregulation in ESFT cells antagonizes cell proliferation and tumorigenesis (Riggi et al., 2010).

Collectively, these findings show that SOX2 is re- quired by CSC in various gliomas, in mouse and human (as well as by CSC of several non-neural cell types). This raises the possibility that SOX2 may qualify as a target for CSC-directed therapy strategies; note that, though Sox2 is highly expressed in gliomas and CSC, its expression is very limited in normal brain tissue surrounding

the tumor. They further indicate that the identification of new downstream mediators of Sox2 function in gliomas may be of relevance for therapy approaches, not only in glioma, but also in other Sox2-dependent tumor types, where some relevant Sox2-controlled gene regulatory networks may be conserved (see also the recent reviews on the subject by [Garros-Regulez et al., 2016; Wuebben et al., 2017]).

5. Medulloblastoma Development in a Sox2- ablated Mouse Model

Medulloblastoma (MB), the most common brain tumor in childhood, was shown early on to harbour CSC (Reya et al., 2001; Nicolis, 2007). Quite some studies were devoted to the Sonic Hedgehog (SHH) subgroup of MB, representing about 30% of total MB, presenting aberrant SHH signaling because of loss of function of negative regulators (including PTCH1, SUFU), activating mutations of positive transducers (SMO), or amplification of transcriptional effectors, like GLI2 (Northcott et al., 2012). SHH subgroup MB were shown to originate from cerebellar granule neuron precursors, that proliferate during normal development under the physiological stimulus of SHH (Schuller et al., 2008). In these tumors, SHH pathway inhibitors entered clinical trials, but reports of resistance and relapse indicate the possibility that an insensitive CSC might be spared (Vanner et al., 2014).

In mouse, a model of the SHH subgroup MB is the irradiated Ptch +/- mouse, where postnatal irradiation increases tumor frequency from 20% to 80%. Recent work in this model has shown that rare, quiescent cells, expressing Sox2, behave as tumor-propagating cells following transplantation, and in primary tumors *in situ* (Vanner et al., 2014). Sox2-expressing cells, and their progeny, were labelled *in vivo* through activation of a GFP transgene by an inducible Cre (CreERT2) driven by the Sox2 locus, and characterized through tumor development, by immunofluorescence. This revealed that rare Sox2-positive cells (less than 5% of total) produce rapidly proliferating progenitors (marked by doublecortin, DCX), that, together with their non-dividing progeny (positive for NeuN) constitute the tumor bulk. The fraction of Sox2-positive cells increased following anti-mitotic, or anti-SMO, therapy, which kills dividing cells; this suggests that Sox2-positive cells were spared by these therapies and could be responsible for the observed relapse. These findings indicate Sox2-positive cells as a promising target for an- ti-CSC therapy in SHH

subgroup MB but leave open the question of whether Sox2 itself is functionally relevant for them.

Sox2 function in mouse MB tumorigenesis has been tested by conditional Sox2 knockout within another model of SHH subgroup MB, the lox-stop-lox-SmoM2-YFP mouse (Ahlfeld et al., 2013). SmoM2 encodes a mutated version of the Smo gene, encoding the SHH co-receptor; SmoM2 was originally discovered as an activating mutation of Smo (and thus of the SHH pathway) in basal cell carcinomas (Xie et al., 1998). In the mouse model, the SmoM2 transgene activates SHH signaling following Cre recombinase-mediated excision of the stop sequence (Schuller et al., 2008). In hGFAPCre:SmoM2 flox/+ mice, SmoM2 is activated throughout the neuroepithelium, and MB develops (Schuller et al., 2008); these MB expresses Sox2, with sparse cells showing particularly high expression levels (Xie et al., 1998). In hGFAPCre:Sox2flox/flox:SmoM2flox/+ mice, tumors developed, where SOX2 was not detected. However, the development and morphology of the tumors were not significantly altered by Sox2 loss; further, the survival time was not significantly different between Sox2-positive and Sox2-negative genotypes.

It should be noted that, in this model, SHH signaling is activated uniformly within the many cells in which SmoM2 is activated by Cre, which does not mirror the pathological situation in which an oncogenic mutation is first present in just one cell. However, in this model, Sox2 ablation did not preclude MB tumorigenesis. It is possible that, in MB, Sox2 is upstream to genes regulating SHH signaling, as seen in normal NSC (Favaro et al., 2009); in the SmoM2 model, Sox2 function might be bypassed by autonomous, constitutive activation of the SHH pathway.

In this model, expression of Sox3, a transcription factor belonging to a subgroup of Sox genes coexpressed with Sox2 in the developing neuroepithelium (Bertolini et al., 2016), was detected in the tumor, and found upregulated in mutant cerebella following Sox2 ablation (Xie et al., 1998). Sox transcription factors co-bind to

many target genes (Bergsland et al., 2011), so these findings raise the possibility that Sox3 acts redundantly with Sox2 to maintain MB tumorigenesis. Vice versa, pro-differentiative Sox factors (Sox5, 6, 21) were found to be downregulated during malignant glioma progression, their genetic ablation increased the ability of cells to form glioma-like tumors, and expression of high levels of Sox5/6/21 in primary human GBM cells antagonized their tumorigenic capacity (Kurtsdotter et al., 2017). A similar, antagonistic function of Sox9 and Sox10 was also described in melanoma (Shakhova et al., 2015) (see below Figure 1). It is possible that the Sox genes coexpressed with Sox2 in tumors and CSC may represent a generally important factor conditioning Sox2 function in CSC. Their importance may be tested by genetic ablation experiments, in combination with Sox2.

6. Sox2-independent Neural Cancer Stem Cells in Melanoma

Melanoma is a skin tumor originating from malignant transformation of melanocytes, cells derived from the embryonic neural crest (Shakhova, 2014). About 50% of human melanomas express Sox2 (Laga et al., 2011); further, Sox2 has functional roles in normal melanocyte progenitors of the developing neural crest, as Sox2 gradual downregulation permits the differentiation of neural crest (and Schwann cells)-derived melanocyte progenitors into melanocytes (Adameyko et al., 2012; Schaefer et al 2017). Sox2 has thus been considered a candidate to play functional roles also in melanoma. A different Sox factor, Sox 10, had been previously shown by conditional knock-out to play an essential role in melanoma development within a mouse genetic model of NrasQ61K-driven melanomagenesis (Shakhova et al., 2012), prompting related experiments for Sox2.

The study of Sox2 function in human melanoma-derived cells by RNAi-mediated knockdown approaches had given controversial results. Whereas some Sox2

knockdown experiments indicated a contribution of Sox2 to the maintenance of patient-derived melanoma spheres self-renewal and xenograft tumor development (Santini et al., 2014), and Sox2 overexpression increased melanoma cell invasion (Girouard et al., 2012), in other reports Sox2 knockdown failed to elicit effects (Laga et al., 2011). In more recent experiments, Sox2 was fully ablated via CRISPR/Cas9-mediated mutagenesis in human patient-derived melanoma cells carrying a N-Ras Q61L oncogenic mutation and expressing high Sox2 levels (Schaefer et al., 2017). Following xenotransplantation into immunocompromised mice, Sox2-deleted cells were equally capable of generating tumors than non-deleted cells (Schaefer et al., 2017), pointing to a dispensable role for Sox2.

In mouse, Sox2 function in *in vivo* melanomagenesis was recently studied in two different genetic models of melanoma, by conditional Sox2 knockout. The first model is the Tyr:CreERT2 :: NRasQ61K Ink4a-/- mouse, in which a transgene (Tyr:CreERT2), expressing the inducible CreERT2 in melanocytes, is coupled to a Cre-activatable transgene, carrying the oncogenic NRasQ61K mutation, together with a homozygous mutation of Ink4a (Schaefer et al., 2017). This model recapitulates all phases of melanomagenesis from benign nevi formation from melanocytes, to malignant transformation, metastases, and tumor dissemination at distant sites; in this model, the requirement for Sox10 in melanomagenesis had previously been demonstrated (Shakhova et al., 2012). In the second model, a BrafV600E mutation cooperates with Pten loss to induce melanoma (Cesarini et al., 2017). Both tumors express Sox2, at least in a fraction of cells. In both cases, Sox2 ablation was obtained by combining these oncogenic mutations with a homozygous Sox2flox allele; this caused Sox2 deletion, by the same Cre recombinases that activated the oncogenic transgenes. In both cases, Sox2 ablation did not prevent tumor formation, and the kinetics and characteristics of tumor growth did not differ detectably in Sox2-negative and Sox2-positive (control) tumors, nor did the

survival; further, to test for Sox2 requirement in different phases of tumorigenesis, deletion was induced before the appearance of primary melanomas, or when metastases formed, and again no difference was observed with controls, despite efficient Sox2 deletion (Schaefer et al., 2017; Cesarini et al., 2017). This indicates that, at least in these models, Sox2 is dispensable for tumor development from melanocytes. It is possible that the different findings in some human cell lines and in the two *in vivo* mouse models (see above) reflect a previously unappreciated heterogeneity in melanomas regarding Sox2 requirement. Functional experiments will be important in discriminating, among melanomas, those requiring Sox2 function for tumorigenesis.

7. A Parallel: Different Requirements for Sox2 by Stem/Progenitor Cells of Different Regions of the Normal Developing Nervous System

It may be interesting to note that Sox2 function is highly context-dependent also in the development of the normal nervous system. Following Sox2 deletion throughout the developing neural tube at embryonic day 11.5 (E11.5) (via a nestin-Cre transgene), development of the hippocampus is severely perturbed, whereas neural development in general is comparatively spared (Favaro et al., 2009). Following even earlier Sox2 ablation throughout the developing telencephalon, at E9.5 (via a FoxG1Cre transgene), the ventral telencephalon (medial ganglionic eminences, the primordium of the basal ganglia) is essentially lost, as is the olfactory neuroepithelium, whereas the dorsal telencephalon (cortex primordium) is comparatively less affected (Panaliappan et al., 2018). Sox2 ablation in the developing midbrain/hindbrain led to impaired development of the cerebellar vermis, and of postnatal cerebellar Bergmann glia, but not (or much less) of other cerebellar regions and cell types (Cerrato et al., 2018). These findings indicate that

Sox2 normally functions in a stage-, region-, and cell type-specific way in neural cells, driving context-specific gene regulatory networks.

8. Conclusions and Perspective

We have summarized evidence showing that Sox2 is essential for the maintenance of CSC and tumorigenesis in some neural tumor types, while being dispensable in others. What molecular mechanisms could be involved in this context-specificity of functions in CSC? It is possible that different stem cell programs control the maintenance of, for example, glioma versus melanoma stem cells, and that only the first one requires Sox2. Alternatively, it is possible that, at least in some cases, Sox2 acts redundantly with other Sox factors, coexpressed with it in some tumor types (e.g. Sox3 in MB, see above). The combined ablation, in model systems, of the Sox genes potentially involved, by conditional knockout or CRISPR-Cas-mediated mutagenesis, should clarify these points.

On the other hand, we noted that a requirement for Sox2 is found not only in neural CSC within gliomas, but also in very different, non-neural tumors, such as skin and esophagus SCC, lung SCLC, and osteosarcomas. Perhaps, although they differ by histology and by cell of origin, these tumors share a "core", Sox2-controlled gene regulatory network, active in their CSC. Thus, it will be important to comparatively characterize the gene regulatory networks controlled by Sox2 in these CSCs.

These findings have implications for therapy approaches. On one hand, they suggest it might be advantageous to "classify" tumours according to the gene regulatory networks that function in the maintenance of their CSC, that in turn might involve shared efficacy of CSC-targeting drugs. On the other hand, they emphasize the need for functional experiments, to address the importance of specific gene products

(here, Sox2, and its downstream targets), to distinguish driver from bystander roles, in order to appropriately target future CSC-directed therapy approaches.

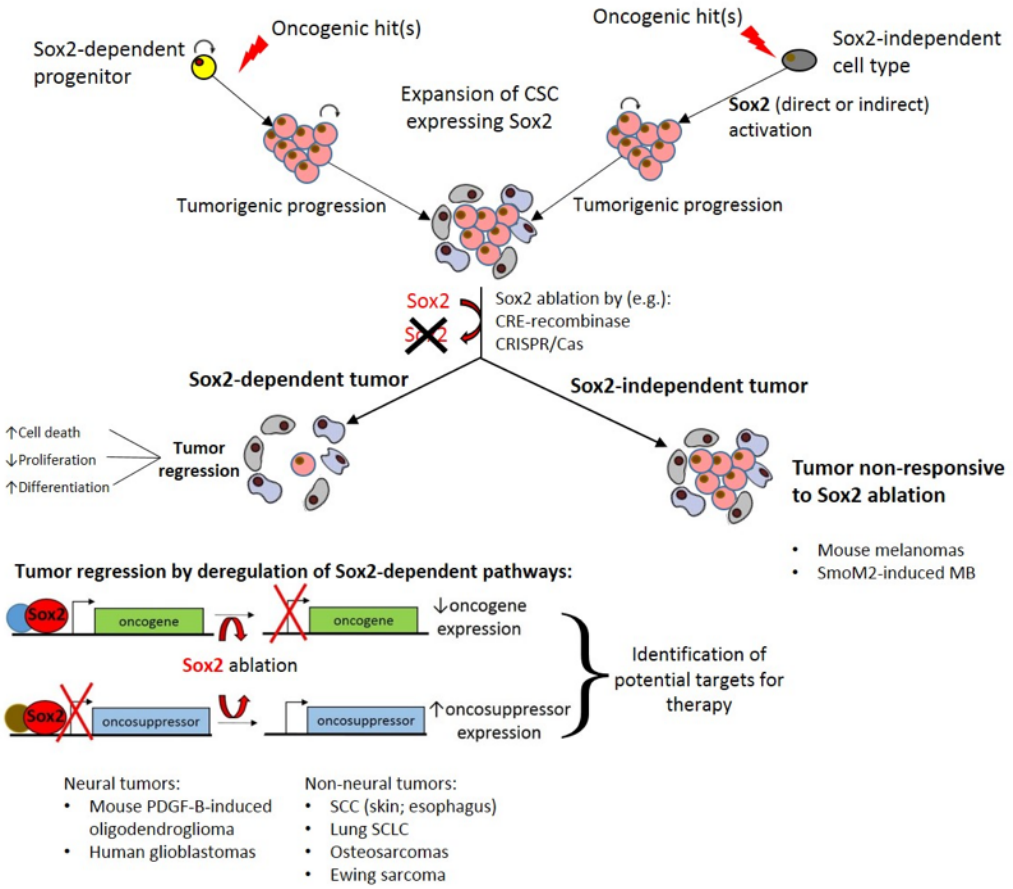


Figure 1: Sox2 experimental ablation distinguishes Sox2-dependent from Sox2-independent CSC.

Oncogenic hits can convert normal cells to CSC. Sox2 ablation can antagonize CSC and tumor progression (left). Here, Sox2 downstream genes can be identified, which vary their expression following Sox2 loss; downregulated as well as upregulated genes are identified. Experimental manipulation of the activity of these

targets can identify important mediators of Sox2 effects on cancer cells, providing potential targets for therapy.

Acknowledgements

Work in the Nicolis laboratory has been supported by Associazione Italiana per la Ricerca sul Cancro (AIRC- IG), the European Community (ERANET-NEURON ImprovVision), Telethon, Fondazione Cariplo. We wish to thank Sergio Ottolenghi for insightful discussion.

References

Adameyko I, Lallemend F, Furlan A, et al. (2012) Sox2 and Mitf cross-regulatory interactions consolidate progenitor and melanocyte lineages in the cranial neural crest. Devel- opment 139: 397-410.

Ahlfeld J, Favaro R, Pagella P, et al. (2013) Sox2 require- ment in Sonic hedgehog-associated medulloblastoma. Cancer Res 73: 3796-3807.

Appolloni I, Calzolari F, Tutucci E, et al. (2009) PDGF-B induces a homogeneous class of oligodendrogliomas from embryonic neural progenitors. Int J Cancer 124: 2251-2259.

Arnold K, Sarkar A, Yram MA, et al. (2011) Sox2(+) adult stem and progenitor cells are important for tissue regenera- tion and survival of mice. Cell Stem Cell 9: 317-329.

Avilion AA, Nicolis SK, Pevny LH, et al. (2003) Multipotent cell lineages in early mouse development depend on SOX2 function. Genes Dev 17: 126-140.

Bass AJ, Watanabe H, Mermel CH, et al. (2009) SOX2 is an amplified lineage-survival oncogene in lung and esopha- geal squamous cell carcinomas. Nat Genet 41: 1238-1242.

Basu-Roy U, Ambrosetti D, Favaro R, et al. (2010) The transcription factor Sox2 is required for osteoblast self-re- newal. Cell death and differentiation 17: 1345-1353.

Basu-Roy U, Seo E, Ramanathapuram L, et al. (2012) Sox2 maintains self renewal of tumor-initiating cells in oste- osarcomas. Oncogene 31: 2270-2282.

Beck B, Blanpain C (2013) Unravelling cancer stem cell po- tential. Nat Rev Cancer 13: 727-738.

Bergsland M, Ramsköld D, Zaouter C, et al. (2011) Se- quentially acting Sox transcription factors in neural lineage development. Genes Dev 25: 2453-2464.

Bertolini J, Mercurio S, Favaro R, et al. (2016) Sox2-depen- dent regulation of neural stem cells and CNS development. In: Kondoh H, Lovell-Badge R, Sox2, Biology and role in development and disease. Elsevier.

Boumahdi S, Driessens G, Lapouge G, et al. (2014) SOX2 controls tumour initiation and cancer stem-cell functions in squamous-cell carcinoma. Nature 511: 246-250.

Bulstrode H, Johnstone E, Marques-Torrejon MA, et al. (2017) Elevated FOXG1 and SOX2 in glioblastoma enforces neural stem cell identity through transcriptional control of cell cycle and epigenetic regulators. Genes Dev 31: 757-773.

Cerrato V, Mercurio S, Leto K, et al. (2018) Sox2 condition- al mutation in mouse causes ataxic symptoms, cerebellar vermis hypoplasia, and postnatal defects of Bergmann glia. Glia.

Cesarini V, Guida E, Todaro F, et al. (2017) Sox2 is not required for melanomagenesis, melanoma growth and melanoma metastasis *in vivo*. Oncogene 36: 4508-4515.

Chen J, Li Y, Yu TS, et al. (2012) A restricted cell population propagates glioblastoma growth after chemotherapy. Nature 488: 522-526.

Clavel C, Grisanti L, Zemla R, et al. (2012) Sox2 in the der- mal papilla niche controls hair growth by fine-tuning BMP signaling in differentiating hair shaft progenitors. Dev cell 23: 981-994.

Domyan ET, Ferretti E, Throckmorton K, et al. (2011) Sign- aling through BMP receptors promotes respiratory identity in the foregut via repression of Sox2. Development 138: 971-981.

Favaro R, Appolloni I, Pellegatta S, et al. (2014) Sox2 is required to maintain cancer stem cells in a mouse model of high-grade oligodendroglioma. Cancer Res 74: 1833-1844.

Favaro R, Valotta M, Ferri AL, 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, 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.

Gangemi RM, Griffero F, Marubbi D, et al. (2009) SOX2 silencing in glioblastoma tumor-initiating cells causes stop of proliferation and loss of tumorigenicity. Stem Cells 27: 40-48.

Garros-Regulez L, Garcia I, Carrasco-Garcia E, et al. (2016) Targeting Sox2 as a therapeutic strategy in glioblastoma. Front Oncol 6: 222.

Girouard SD, Laga AC, Mihm MC, et al. (2012) SOX2 con- tributes to melanoma cell invasion. Laboratory investigation 92: 362-370.

Gupta N, Gopal K, Wu C, et al. (2018) Phosphorylation of Sox2 at threonine 116 is a potential marker to identify a subset of breast cancer cells with high tumorigenecity and stem-like features. Cancers 10.

Guy CT, Webster MA, Schaller M, et al. (1992) Expression of the neu protooncogene in the mammary epithelium of transgenic mice induces metastatic disease. Proc Natl Acad Sci U S A 89: 10578-10582.

He J, Shi J, Zhang K, et al. (2017) Sox2 inhibits Wnt-beta-catenin signaling and metastatic potency of cisplatin-resistant lung adenocarcinoma cells. Mol med rep 15: 1693-1701.

Kondoh H, Robin LB (2016) Sox2, biology and role in development and disease. Elsevier, Associated Press.

Kondo T, Raff M (2004) Chromatin remodeling and histone modification in the conversion of oligodendrocyte precur- sors to neural stem cells. Genes Dev 18: 2963-2972.

Kurtsdotter I, Topcic D, Karlen A, et al. (2017) SOX5/6/21 Prevent oncogene-driven transformation of brain stem cells. Cancer Res 77: 4985-4997.

Jackson EL, Garcia-Verdugo JM, Gil-Perotin S, et al. (2006) PDGFR alpha-positive B cells are neural stem cells in the adult SVZ that form glioma-like growths in response to in- creased PDGF signaling. Neuron 51: 187-199.

Laga AC, Zhan Q, Weishaupt C, et al. (2011) SOX2 and nestin expression in human melanoma: An immunohistochemical and experimental study. Exp dermatol 20: 339- 345.

Lau CS, Mahendraraj K, Chamberlain RS (2017) Oligodendrogliomas in pediatric and adult patients: an out- come-based study from the surveillance, epidemiology, and end result database. Cancer Manag Res 9: 159-166.

Lee J, Kotliarova S, Kotliarov Y, et al. (2006) Tumor stem cells derived from glioblastomas cultured in bFGF and EGF more closely mirror the phenotype and

genotype of primary tumors than do serum-cultured cell lines. Cancer cell 9: 391-403.

Leis O, Eguiara A, Lopez-Arribillaga E, et al. (2012) Sox2 expression in breast tumours and activation in breast cancer stem cells. Oncogene 31: 1354-1365.

Muller WJ, Sinn E, Pattengale PK, et al. (1988) Single-step induction of mammary adenocarcinoma in transgenic mice bearing the activated c-neu oncogene. Cell 54: 105-115.

Nicolis SK (2007) Cancer stem cells and "stemness" genes in neuro-oncology. Neurobiol Dis 25: 217-229

Northcott PA, Jones DT, Kool M, et al. (2012) Medulloblastomics: The end of the beginning. Nat Rev Cancer 12: 818- 834.

Panaliappan TK, Wittmann W, Jidigam VK, et al. (2018) Sox2 is required for olfactory pit formation and olfactory neurogenesis through BMP restriction and Hes5 upregulation. Development 145.

Pattabiraman DR, Weinberg RA (2014) Tackling the cancer stem cells - what challenges do they pose? Nat Rev Drug Discov 13: 497-512.

Pevny LH, Nicolis SK (2010) Sox2 roles in neural stem cells. Int J Biochem Cell Biol 42: 421-424.

Rasmussen BK, Hansen S, Laursen RJ, et al. (2017) Epidemiology of glioma: clinical characteristics, symptoms, and predictors of glioma patients grade I-IV in the danish neuro-oncology registry. J Neurooncol 135: 571-579.

Reifenberger G, Wirsching HG, Knobbe-Thomsen CB, et al. (2017) Advances in the molecular genetics of gliomas - implications for classification and therapy. Nat Rev Clin Oncol 14: 434-452.

Reis-Filho JS, Faoro LN, Carrilho C, et al. (2000) Evaluation of cell proliferation, epidermal growth factor receptor, and bcl-2 immunoexpression as prognostic factors for patients with World Health Organization grade 2 oligodendroglioma. Cancer 88: 862-869.

Reya T, Morrison SJ, Clarke MF, et al. (2001) Stem cells, cancer, and cancer stem cells. Nature 414: 105-111.

Riggi N, Suva ML, De Vito C, et al. (2010) EWS-FLI-1 modulates miRNA145 and SOX2 expression to initiate mesenchymal stem cell reprogramming toward Ewing sarcoma cancer stem cells. Genes Dev 24: 916-932.

Rudin CM, Durinck S, Stawiski EW, et al. (2012) Comprehensive genomic analysis identifies SOX2 as a frequent- ly amplified gene in small-cell lung cancer. Nat Genet 44: 1111-1116.

Santini R, Pietrobono S, Pandolfi S, et al. (2014) SOX2 regulates self-renewal and tumorigenicity of human melanoma-initiating cells. Oncogene 33: 4697-4708.

Schaefer SM, Segalada C, Cheng PF, et al. (2017) Sox2 is dispensable for primary melanoma and metastasis formation. Oncogene 36: 4516-4524.

Schuller U, Heine VM, Mao J, et al. (2008) Acquisition of granule neuron precursor identity is a critical determinant of progenitor cell competence to form Shh-induced medulloblastoma. Cancer cell 14: 123-134.

Shakhova O, Cheng P, Mishra PJ, et al. (2015) Antagonistic cross-regulation between Sox9 and Sox10 controls an anti-tumorigenic program in melanoma. PLoS Genet 11: e1004877.

Shakhova O (2014) Neural crest stem cells in melanoma development. Curr Opin Oncol 26: 215-221.

Shakhova O, Zingg D, Schaefer SM, et al. (2012) Sox10 promotes the formation and maintenance of giant congenital naevi and melanoma. Nat cell biol 14: 882-890.

Shih AH, Holland EC (2006) Platelet-derived growth factor (PDGF) and glial tumorigenesis. Cancer lett 232: 139-147.

Singh SK, Hawkins C, Clarke ID, et al. (2004) Identification of human brain tumour initiating cells. Nature 432: 396-401.

Suva ML, Rheinbay E, Gillespie SM, et al. (2014) Recon- structing and reprogramming the tumor-propagating potential of glioblastoma stem-like cells. Cell 157: 580-594.

Tang L, Wang D, Gu D (2018) Knockdown of Sox2 inhibits OS cells invasion and migration via modulating wnt/beta-catenin signaling pathway. Patholo oncol res 24: 907-913.

Vanner RJ, Remke M, Gallo M, et al. (2014) Quiescent sox2(+) cells drive hierarchical growth and relapse in sonic hedgehog subgroup medulloblastoma. Cancer cell 26: 33- 47.

Wagner KU, Wall RJ, St-Onge L, et al. (1997) Cre-mediated gene deletion in the mammary gland. Nucleic Acids Res 25: 4323-4330.

Wu C, Gupta N, Huang YH, et al. (2018) Oxidative stress enhances tumorigenicity and stem-like features via the activation of the Wnt/beta-catenin/MYC/Sox2 axis in ALK-positive anaplastic large-cell lymphoma. BMC cancer 18: 361.

Wuebben EL, Rizzino A (2017) The dark side of SOX2: Cancer - a comprehensive overview. Oncotarget 8: 44917- 44943.

Wu F, Zhang J, Wang P, et al. (2012) Identification of two novel phenotypically distinct breast cancer cell subsets based on Sox2 transcription activity. Cellular signal 24: 1989-1998.

Xie J, Murone M, Luoh SM, et al. (1998) Activating smoothened mutations in sporadic basal-cell carcinoma. Nature 391: 90-92.

Yang N, Hui L, Wang Y, et al. (2014) Overexpression of SOX2 promotes migration, invasion, and epithelial-mesenchymal transition through the Wnt/beta-catenin pathway in laryngeal cancer Hep-2 cells. Tumour Biol 35: 7965-7973.

Zhang X, Lu F, Wang J, et al. (2013) Pluripotent stem cell protein Sox2 confers sensitivity to LSD1 inhibition in cancer cells. Cell Rep 5: 445-457.

Chapter 4

Paper published, ***Cell stem cell***, 7 Mar 2019,
https://doi.org/10.1016/j.stem.2019.02.004

Scope of the work

During my PhD studies I gave my contribute to a big collaborative work that have disclosed mechanisms by which Sox2 acts in gene regulation in normal neural stem cells at the genome-wide level, by maintaining and regulating a network of long-range interactions in chromatin, connecting gene promoters to distant enhancers. These long-range interaction maps have open to the dentification of novel significant downstream mediators of Sox2 function important for neural stem cell self-renewal and relevant for Sox2-dependent neurodevelopmental diseases.

Mapping the Global Chromatin Connectivity Network for Sox2 Function in Neural Stem Cell Maintenance

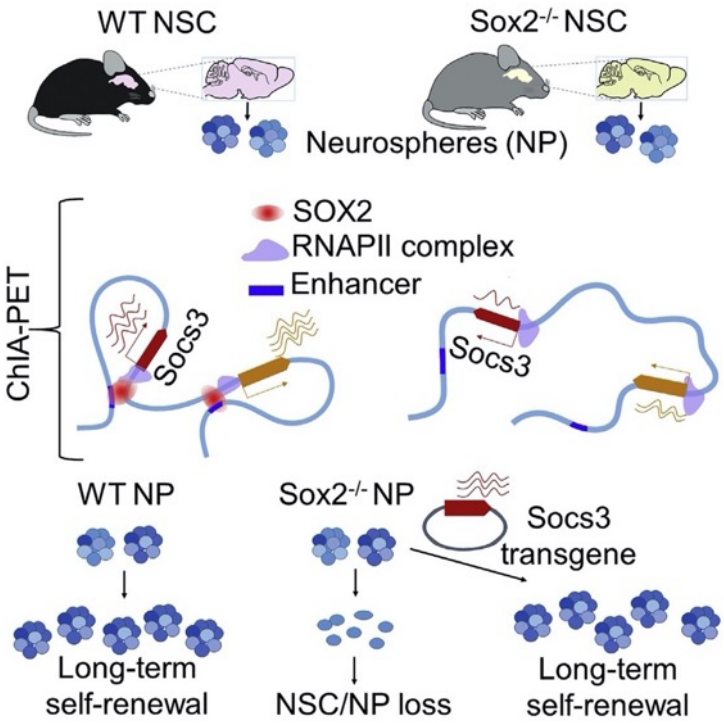
Jessica A. Bertolini,[1,9] Rebecca Favaro,[1,9] Yanfen Zhu,[2,9] Miriam Pagin,[1] Chew Yee Ngan,[2] Chee Hong Wong,[2] Harianto Tjong,[2] Marit W. Vermunt,[3,8] Ben Martynoga,[4] Cristiana Barone,[1] Jessica Mariani,[1] Marcos Julian Cardozo,[5] Noemi Tabanera,[5] Federico Zambelli,[6] Sara Mercurio,[1] Sergio Ottolenghi,[1] Paul Robson,[2,7] Menno P. Creyghton,[3] Paola Bovolenta,[5] Giulio Pavesi,[6] Francois Guillemot,[4] Silvia K. Nicolis,[1,10,*] and Chia-Lin Wei[2,*]

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 Biologia Molecular Severo Ochoa, Consejo Superior de Investigaciones Cientificas-Universidad Autonoma de Madrid and Ciber de Enfermedades Raras (CIBERER), ISCIII Madrid, Spain
6 Department of Biosciences, University of Milano, 20133 Milano, Italy
7 Stem Cell and Regenerative Biology, Genome Institute of Singapore, Singapore
8 Present address: Division of Hematology, The Children's Hospital of Philadelphia, Philadelphia, PA, USA
9 These authors contributed equally
10 Lead Contact
*Correspondence: silvia.nicolis@unimib.it (S.K.N.), chia-lin.wei@jax.org (C.-L.W.)

Table of Content Image



Abstract

The SOX2 transcription factor is critical for neural stem cell (NSC) maintenance and brain development. Through chromatin immunoprecipitation (ChIP) and chromatin interaction analysis (ChIA-PET), we determined genome-wide SOX2-bound regions and Pol II-mediated long-range chromatin interactions in brain-derived NSCs. SOX2-bound DNA was highly enriched in distal chromatin regions interacting with promoters and carrying epigenetic enhancer marks. Sox2 deletion caused widespread reduction of Pol II-mediated long-range interactions and decreased gene expression. Genes showing reduced expression in Sox2-deleted

cells were significantly enriched in interactions between promoters and SOX2-bound distal enhancers. Expression of one such gene, Sup- pressor of Cytokine Signaling 3 (Socs3), rescued the self-renewal defect of Sox2-ablated NSCs. Our work identifies SOX2 as a major regulator of gene expression through connections to the enhancer network in NSCs. Through the definition of such a connectivity network, our study shows the way to the identification of genes and enhancers involved in NSC maintenance and neurodevelopmental disorders.

1. Introduction

Neural stem cells (NSCs) 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 anophthalmia, 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 NSCs to self-renew in long-term cultures (Favaro et al., 2009). SOX2 functions and targets are the subject of intense investigation (Engelen et al., 2011; Gonc ̧ alves et al., 2016; Hagey et al., 2016; Lodato et al., 2013; Zhu et al., 2013).

Transcriptional regulation is mediated via DNA looping between gene promoters and their corresponding distal enhancers, 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 net- works (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 single transcription factor influences the function of genome-wide interaction networks in controlling cell-specific transcriptional activity.

We previously used chromatin interaction analysis with paired-end tag sequencing (ChIA-PET) to identify RNA polymerase II (Pol II)-mediated long-range interactions in embryonic stem cells (ESCs) and in brain-derived NSC or progenitor cell cultures (Zhang et al., 2013). In the present work, we sought to identify molecular mechanisms underlying *Sox2*-dependent gene regulation in NSCs, as well as genes involved in *Sox2*- dependent maintenance of long-term NSC self-renewal. We thus deleted *Sox2* in NSCs in mouse embryonic brain and studied the effects of embryonic loss of *Sox2* on RNA expression in neonatal NSCs grown *in vitro* (see Favaro et al. 2009) and its relationship to the Pol II-mediated chromatin long-range interaction network. 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 *Sox2* mutant NSCs is sufficient to rescue their self-renewal defect.

2. Results

2.1 Comparison of Genome-wide Pol II-Mediated Long-Range Chromatin Interactions in Wild-Type and Sox2-Deleted NSC.

We established NSC cultures from the neonatal forebrain of conditionally (at E11.5) *Sox2*-ablated mice and their control non-deleted littermates (Favaro et al., 2009). Freshly isolated *Sox2*-deleted (mutant; MUT) and control (wild-type; WT) NSCs efficiently expand in culture at early passages (Favaro et al., 2009), however, MUT NSCs later fail to self-renew long-term, pointing to a requirement for *Sox2* in NSC maintenance that matches a defect observed also *in vivo* after P0 in the hippocampus (Favaro et al., 2009). Sox2-deleted NSCs retain the ability to differentiate into glia and neurons upon differentiation induction, however, under self-renewal culture conditions, they do not spontaneously differentiate, as indicated by morphological and immunochemical criteria and by comparison of RNA sequencing (RNA-seq) data (by Pearson correlation, hierarchical clustering, and principal component analysis) of un-induced WT and MUT cells (day 0) with WT cells induced to initial differentiation (day 4) (data not shown).

To determine the effect of *Sox2* loss on the genome-wide pattern of Pol II-mediated long-range chromatin interactions, we first performed ChIA-PET analysis with anti-polII antibodies, specific for the preinitiation complex (Zhang et al., 2013), comparing *ex vivo* NSC cultures derived at P0. These cultures express forebrain-specific transcripts, indicating that the NSCs

maintain a forebrain identity (Zappone et al., 2000; Zhang et al., 2013). ChIA-PET identifies protein-mediated genome- wide long-range chromatin contacts through proximity ligation and chromatin immunoprecipitation (ChIP). We generated ChIA-PET data from both normal (WT) and *Sox2*-ablated (MUT) brain cells to determine Pol II-mediated long-range connectivity (distant enhancer-promoter and promoter-promoter connectivity) (Figure S1). We began to analyze the chromatin

connectivity by following the original version of the protocol (Zhang et al., 2013) that performed nuclei lysis followed by proximity ligation and Pol II ChIP from hundreds of million pooled NSCs from WT and Sox2-deleted (MUT) neonatal (P0) forebrains (4 WT and 6 MUT, deleted at E11.5) (Favaro et al., 2009) neonates. These datasets (Table S1) are referred to as wTR1 and mTR1, respectively. We observed a substantial reduction of normalized inter- actions (numbers of significant interactions per million intra-molecular ligation PETs) in *Sox2*-deleted (2,295 in mTR1) versus WT NSCs (10,197 in wTR1) (Table 1 and Figure 1, discussed below). We further verified that such reduction of overall Pol II connectivity observed in MUT NSCs did not result from different Pol II immunoprecipitation efficiencies between WT and MUT cells; indeed, we observed highly similar normalized density profiles between WT and MUT genomic regions (Figure S2). Specifically, the vast majority (92%) of the ChIA-PET-defined Pol II binding regions in WT NSCs were retained in MUT NSCs, irrespective of whether or not they were connected (Figures 1 and S2). To confirm that the reduction of Pol II-mediated interactions in MUT NSCs was not influenced by the pooling of heterogeneous samples or correlated with the experimental procedure, we subsequently generated additional replicate datasets using an improved ChIA-PET method (named *in situ* ChIA-PET) (Figure S1). In the *in situ* Pol II ChIA-PET protocol, instead of per- forming proximity ligation in chromatin mixtures from hundreds of million cells that is intrinsically noisy (as evident by the high level of inter-chromosomal PETs), proximity ligation was per- formed in permeabilized intact nuclei, followed by ChIP and a transposase-mediated ‘‘tagmentation’’ to generate PETs for sequencing analysis (Figure S1; STAR Methods). As such, the *in situ* ChIA-PET method results in higher efficiency in capturing intra-molecular ligation PETs, thus requires significantly lower numbers of cells to yield highly sensitive detection of protein- mediated chromatin interactions (Mumbach et al.,

2016; Table 1). Therefore, this approach allowed us to analyze cultures from individual neonatal forebrains (two WT, wTR2 and wTR3, and two MUTs, mTR2 and mTR3). The improvement can be shown by the ratio of the intra-molecular interaction PETs; in the TR1 (original version), only 5%–10% of the uniquely aligned paired reads (unique PETs, Table 1, line 3) were intra-molecular chromatin ligated PETs (defined by *cis*-interaction PETs, Table 1, line 6), whereas in TR2 and TR3, 50%–60% of the uniquely aligned paired reads were intra-molecular chromatin ligated PETs (Table 1). Despite two versions of the protocols being applied, the three ChIA-PET experiments (TR1, TR2, and TR3) showed overall highly similar interaction patterns when the same genotypes (WT or MUT) were considered, at different resolution (Figure S2); 74% and 80% of the interactions detected in wTR1were also detected in the combined (i.e., the sum of) wTR2 and wTR3 interactions (within a window ± 5 kb and ± 10 kb, respectively). Further, the average reproducibility score between wTR2 and wTR3 based on SCC (stratum-adjusted correlation coefficient) (Yang et al., 2017) over all 20 chromosomes was 0.935 and 0.839 for wTR2-wTR3 and mTR2-mTR3, respectively (Table S2).

The *in situ* proximity ligation adopted in the improved ChIA- PET method effectively captures specific Pol II-mediated long- range interactions between regulatory elements through the Pol II ChIP enrichment (Figure S1). From a total of 6–13 million intra-molecular PETs in WT and MUT TR2 and TR3, we defined between 15,000 to 96,000 ‘‘significant interactions (loops)’’ (false discovery rate [FDR] <0.05, $p < 0.05$), enriched in Pol II-mediated interactions) (Table 1, line 7; Figure S1; STAR Methods). Among these significant interactions from TR2 and TR3, 2,878–18,022 interactions were mediated by Pol II (defined by the interactions with Pol II binding at *both* DNA regions connected by the interactions) (Table 1, line 8, and STAR Methods). We defined DNA regions connected by an interaction as ‘‘anchors’’ and overlapping anchors as ‘‘nodes’’ (Figure 1B). We

further annotated nodes and anchors as promoter nodes or anchors if they resided within 2.5 kb from annotated transcription start sites (TSS) and the remaining ones as non-promoter nodes/anchors. Based on PhastCons Score Threshold analysis, these non-promoter nodes were significantly more conserved among vertebrate genomes relative to random intergenic regions, suggesting their potential function in chromatin organization (not shown).

Similarly to mTR1 NSCs, mTR2 and mTR3 exhibited a global reduction in the numbers of chromatin interactions from both the general chromatin contacts and the specific Pol II-mediated interactions if compared to WT NSCs (wTR2 and wTR3, respectively) (Figure 1B). The number of normalized significant interactions (''number of loops detected per million of intra-molecular ligated reads'') ranged between 6–10 K in WT but were reduced to only 2–3 K in MUT cell samples (Table 1, line 9), while the normalized significant Pol II-mediated interactions were 719 and 1,374 in WT, but only 307 and 525 in MUT cells samples, respectively (Table 1, line 10). Consistent with the reduction of the Pol II-mediated interactions in MUT cells, the number of no- des was lower in MUT cells (Figure 1B). Indeed, from 8,295 promoter nodes and 6,549 non-promoter nodes in combined wTR2 and wTR3, the corresponding numbers in MUT cells were reduced to 3,356 and 1,726, respectively. On the other hand, the majority of the nodes observed in mTR2 and mTR3 samples were detected in WT cells as well. These changes of interactions can be found in many specific loci. The changes were not uniform across the genome, but rather highly variable, exhibiting loci showing drastic reduction interspersed with loci showing little or no reduction (Figure 1; see also Figures S3 and S5, Table S3, and screenshots throughout the paper). In some cases, while the data from mTR1 showed apparent loss of interactions (as compared to wTR1), the data obtained by the more sensitive *in situ* ChIA-PET, providing a higher depth in the intra-molecular ligated PETs, showed a reduction in frequency

instead of a complete loss of the interactions (Figures 1D and 1G, compare TR2 with TR1; see also Figure S3); the above discussed dependence of the detection of some interactions on obtaining high numbers of interacting PETs makes it difficult to prove the complete loss of any specific interaction. In some regions, we actually observed new sets of loops emerging in MUT cells (Figure 1).

In conclusion, the data indicate that in the absence of *Sox2*, chromatin connectivity was substantially altered genome-wide, with an overall decreased interaction frequency, in particular at selected loci.

2.2 SOX2-Bound Distal Regions Carrying Enhancer Marks Are Highly Enriched Within Interactions.

The changes in connectivity observed following *Sox2* ablation point to a role for SOX2 DNA binding within chromatin in the generation and/or maintenance of long-range interactions. We thus identified SOX2-bound sites through genome-wide ChIP-seq of WT brain-derived NSCs in culture (Figure S4; STAR Methods). We also performed ChIP-seq, in both WT and MUT NSCs, for histone modifications H3K27ac and H3K4me1 (Figure S4, two replicates), allowing for the identification of active (H3K27ac+ and H3K4me1+), as well as ‘‘poised’’ (H3K4me1+ only) enhancers, with the potential to be activated (Cantone and Fisher, 2011; Creyghton et al., 2010; Rivera and Ren, 2013). For the latter analysis, we used both a ‘‘peak-calling’’ and a ‘‘segmentation’’ approach (chromHMM), which both led to qualitatively consistent results (see STAR Methods).

Finally, we linked SOX2-binding sites to both epigenetic marks and interacting anchors, as defined in WT NSCs by Zhang et al. (2013) (corresponding to wTR1) and in *in situ* ChIA-PET experiments (wTR2, wTR3) (Figure 2). For a summary of data, see Tables S4 and S5.

SOX2-bound sites were rarely located at promoters (±1 kb TSS of RefSeq genes) and more frequently in intronic and inter- genic distal regions (Figure 2A and data not shown). Over 90% of SOX2-bound sites were associated with nucleosomes characterized by the presence of either or both the histone modifications investigated in WT NSCs (Figure 2A, histograms: peak calling and chromHMM); in particular, both SOX2-bound promoters and distal regions were mostly (90%) H3K27ac. On the other hand, within H3K27ac+ regions, a much smaller proportion of promoters than of distal regions were SOX2-bound (Figures 2B, right, and S4); indeed, within distal H3K27ac+ regions, SOX2-bound regions were highly enriched in comparison to 1,000 sets of random genomic loci ($p < 0.001$, random sam- pling, see STAR Methods).

To identify the binding of SOX2, if any, within long-range inter- actions, we first classified interactions in WT cells according to the type of interacting element (i.e., promoter [P] or non-promoter [non-P]) (Figure 2C). Approximately 85%–90% of the interactions were mediated through promoters (promoter-promoter [P-P] or promoter-non-promoter [P-non-P] interactions) in both WT and MUT cells, equally subdivided between P-P and P-non-P classes (Figure 2C); within P-non-P interactions, a promoter is connected to either an intergenic or an intragenic region. Only a small number (10%–15%) of interactions connect two non-promoter regions (Figure 2C). An almost identical distribution was observed using interactions defined from TR2 and TR3 (Figure 2C).

Within interactions, SOX2 peaks were highly abundant; ca. 35%–46% of all interactions in WT NSCs (TR1, TR2, and TR3) carried a SOX2-bound site within at least one of the two interacting anchors (Figure 2D; Tables S4 and S5). Specifically, approximately half (44%–53%) of P-nonP interactions (also called P-E, see below) were SOX2-positive with 34%–43% of distal elements (putative enhancers) being SOX2-bound (Figure 2D). Approximately 95% of SOX2-bound

distal anchors were in regions carrying both active enhancer marks (H3K27ac+, H3K4me1+) and the remainder with either one or the other mark; among non-SOX2-bound distal anchors, ~70% were associated with both marks and ~15% with either (not shown). From now on, we refer to these interactions as ''promoter-enhancer'' (P-E) interactions. Importantly, SOX2-positive epigenetically marked (EM) distal regions (whether H3K27Ac+ and/or H3K4Me1+) were significantly more involved in interactions than SOX2-negative EM regions: this was observed in the original ChIA-PET (TR1), as well as in *in situ* ChIA-PET data (TR2 and TR3) (Figures 2E and S4 and data not shown). Thus, the high frequency of SOX2-bound distal anchors is the result of both the enrichment of SOX2 binding within EM distal regions (Figure 2B) and the preferential engagement in long-range interactions of SOX2-positive EM regions versus non-SOX2-bound regions (Figure 2E).

In conclusion, the high enrichment of SOX2-positive EM regions within distal anchors in P-E interactions points to a functional role of SOX2 binding at the level of these interactions.

Interestingly, the loss of SOX2 in MUT NSCs does not lead to important changes (loss or gain) in the patterns of histone modifications: enrichment in H3K27ac, H3K4me1, both, or none (Figure S4).

2.3 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 (Figures 3 and S5; Table S6). We identified several genes playing important developmental roles in the forebrain, some of

which are homologs of human genes involved, when mutated, in inherited brain diseases (microcephaly, intellectual disability, etc.) (Table S6), which were connected by multiple P-E (and P-P) interactions to SOX2-bound elements (possible enhancers). Some of the distal anchors connected to these genes by P-E interactions overlapped with previously identified p300-bound enhancers, already shown to be active in forebrain as transgenic constructs (VISTA enhancers, https://enhancer.lbl.gov) (Visel et al., 2009; Figures 3 and S5). Interestingly, VISTA enhancers were enriched within TR1, TR2, and TR3 interaction anchors (with 13% to 23% of VISTA enhancers overlapping with anchors), and particularly so within distal non-promoter interaction anchors; in *Sox2* MUT cells the proportion of VISTA enhancers overlapping with anchors dropped to ca. 4%, and their representation in distal anchors was drastically reduced (Figure S5). These data point to potential functional roles of distal non-promoter anchors identified by ChIA-PET in gene regulation *in vivo* in the developing brain.

Next, we tested distal regions involved in SOX2-bound long- range interactions using a transgenic enhancer assay in zebra- fish. Fifteen out of seventeen reporter constructs containing SOX2-bound distal anchors directed GFP expression to the developing forebrain (Figure 4; Table S7). GFP expression matched part of, or the whole, forebrain expression pattern of the endogenous zebrafish ortholog of the mouse gene con- nected to the analyzed enhancer (Figure 4). Similar data were obtained with anchors connected to important regulators of fore- brain development, including *Sp8*, *Cxcr4*, *Sox3*, *Nr2f1*, *Irx1*, *Socs3* and c-*fos* (Figure 4; Table S6). Further, the expression of 3 out of 7 enhancers tested was affected by anti-*Sox2* morpholinos or by injected *Sox2* mRNA (Figure S6 and data not shown). Thus, RNA-polII-mediated, SOX2-bound interactions identify with high confidence novel forebrain enhancers.

SOX2-Bound Promoter-to-Enhancer Long-Range Interactions Are the Main Determinant of SOX2- Dependent Transcription.

To correlate the expression of genes to the observed pattern of long-range interactions, we analyzed by RNA-seq the transcriptomes of WT and MUT NSCs (Tables S4 and S5).

To determine how the presence of Pol II-mediated interactions is reflected into gene expression levels, we subdivided genes according to the Pol II ChIA-PET connectivity of their promoters (Zhang et al., 2013) (non-connected promoter, promoter connected with promoters or enhancers, or promoter connected specifically with enhancers, see Figure 2D). Then, we determined the distribution of transcript levels in both WT and MUT NSCs for the different interaction categories (Figure 5A). Considering only transcribed genes, those involved in interactions (P-P and P-E) showed a distribution of expression levels shifted toward higher values than the overall population of expressed genes (data for TR1, TR2, and TR3, Figure 5A); the highest expression levels were seen with genes involved specifically in interactions with enhancers (P-E) (Figure 5A) (p value < 2.2 3 1016). Interestingly, higher expression levels were also associated with an increase in number of interactions per gene, particularly with P-E interactions, which show a clear trend toward higher expression values for every added ‘‘enhancer’’ (TR1, TR2, and TR3, Figure 5B). Moreover, genes whose promoter interacted with SOX2-bound enhancers were more expressed than those connected to enhancers not bound by SOX2 (TR1, TR2, and TR3, Figure 5A). Very similar results were obtained by both the analysis of data obtained by the original ChIA-PET (Zhang et al., 2013) (TR1) and by *in situ* ChIA-PET (TR2 and TR3).

We next compared gene expression between MUT and WT NSCs. In MUT NSCs, the distributions of expression levels of all genes, and of each category of genes, had lower median and mean values than in WT NSCs, with an overall distribution

significantly shifted toward lower values (Figure 5A). Indeed, a Wilcoxon paired signed-rank test (see STAR Methods) showed that the differences between WT and MUT distributions of gene expression were highly significant (p value < 2.2 3 1016 for every pair considered), for TR1, TR2, and TR3 (Fig- ure 5A). 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 be- tween WT and MUT cells, defined as log-fold ratios (log2 [TPM_wild-type/TPM_mutant]) (TPM, transcripts per million) (Figure 5C). To avoid bias from genes with low transcript levels, we considered only genes with TPM >1. The plot was clearly shifted toward positive values, indicating that the majority of genes were more highly expressed in WT than in MUT NSCs (Figure 5C).

Taken together, the above results indicate that loss of *Sox2* is associated with an overall gene expression decrease that is more relevant for genes involved in Pol II-mediated P-E interactions in WT NSCs; moreover, genes with SOX2-positive P-E interactions were more expressed than genes with P-E interactions in general. To further validate these results, we considered genes with TPM >5 in either WT or MUT cells and divided them into three groups (Figure 5D): group 1, genes showing a significant decrease of expression in MUT versus WT NSCs (677 genes); group 2, genes showing a visible but not statistically significant decrease of expression in MUT (2194 genes); group 3, all the other genes with TPM >5 (see STAR Methods). We then 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 summarized the results by defining a ‘‘co-association score’’ (see STAR Methods) as the -log10 *p* of the probability of observing, in each of the three expression groups, a given number of genes associated with a given type of interaction. We

denoted a number lower than the expected value by multiplying the result by 1 (Figure 5D; Table S8). In all three experiments (wTR1, wTR2, and wTR3), the results confirmed a highly significant overlap be- tween genes showing significantly decreased expression in MUT NSCs (group 1 genes) and those characterized, in WT NSCs, by a promoter to enhancer interaction (Figure 5D, lanes 1, 5, and 9). Moreover, the influence of SOX2 binding was evident by comparing group 1 genes connected to enhancers bound by SOX2, which yielded highly significant co-association scores (lanes 2, 6, and 10), with those connected to enhancers not bound by SOX2 (lanes 3, 7, and 11), which yielded only margin- ally significant scores in wTR1 and nonsignificant scores in wTR2 and wTR3. On the other hand, SOX2 binding to promoters was only marginally associated to genes showing significantly decreased expression levels (p value >0.01, not shown), pointing to the fact that binding of SOX2 to a connected distal enhancer is much more functionally effective than its binding to a promoter. Finally, the P-P interactions category had no significant overlap with genes showing significant expression changes (group 1). However, group 2 genes showing only mildly decreased expression in MUT cells (not reaching the threshold requested for statistical significance), were significantly associated with the P-P interaction category (lanes 4, 8, and 12); we speculate that P-P interactions might be responsible for moderate positive effects on transcription, and their loss in MUT cells might predominantly result into a minor decrease of expression (as observed for group 2 genes), rather than into the stronger decrease observed with group 1 genes.

In conclusion, the identification of thousands of long-range interaction enhancers in NSCs, many of which are bound by SOX2, demonstrates an important role of SOX2 in controlling gene expression at the connected promoter.

Overexpression of Socs3, a Multi-Connected SOX2 Target, Rescues Long-Term Self-Renewal of MUT NSC *Sox2* MUT NSCs have a severe self-renewal defect, and their growth in culture becomes exhausted after 7–10 passages ($30 days) (Favaro et al., 2009). To evaluate if any specific gene (from among those whose expression is affected by *Sox2* loss) was able to rescue long-term self-renewal in MUT NSC, we expressed in MUT NSCs the *Socs3* gene, an inhibitor of Jak/Stat signaling, which antagonizes precocious differentiation of NSCs into astroglia (Cao et al., 2006). *Socs3* is strongly down- regulated (down to 10%–15% of WT values) in MUT NSCs (Tables S4 and S5) and shows both a SOX2 peak on the promoter and multiple interactions (Figure 6A; see also Figure 1E), including one with a SOX2-bound anchor that already tested active in transgenic Zebrafish assays (Figure 4). We transduced both WT and *Sox2* MUT NSCs with a lentiviral *Socs3*-vector co- expressing GFP, performing three experiments at virus-to-cell ratios transducing 20%, 30%, or 50% of the NSCs. *Socs3*-trans- duced WT NSCs grew at a similar rate as untransduced WT NSCs and continued to grow long-term, whereas untransduced MUT NSCs stopped growing between passages 8–12 (Figures 6B and 6C). In contrast, *Socs-3* transduced MUT NSCs continued to grow long term, even after the untransduced MUT NSCs had stopped growing and could eventually be grown in bulk to generate large cell populations. At the time of initial divergence of the growth curves of transduced and untransduced MUT NSCs (experiment 3), most or all neurospheres (from transduced MUT NSCs) contained GFP+ cells, and over 70% of the cells were positive by fluorescence-activated cell sorting (FACS), indicating a strong enrichment of the transduced cells; eventually, all cells became GFP+ (Figure 6D). Note that *Socs3*- transduced WT NSCs were not positively selected relative to untransduced WT NSCs, indicating that in WT NSCs, 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 NSCs may be responsible for their defective long-term maintenance.

3. Discussion

We show that SOX2 is critically involved in long-range chromatin interactions in NSCs; its ablation during early mouse development leads to a predominant decrease in long-range chromatin connectivity, particularly at some loci, when tested in neonatal forebrain-derived NSC cultures. SOX2 binding is greatly enriched on DNA regions connected by interactions (anchors) at either promoters or enhancers. The loss of *Sox2* decreases the expression of ~1,000 genes. The identification of thousands of epigenetically defined enhancers involved in long-range interactions allowed us to demonstrate that SOX2-bound long- range interactions represent the most relevant functional category associated with the observed gene downregulation in MUT NSCs (Figure 5D). mRNAs encoding important transcription factors, and signal transduction molecules, are significantly reduced in MUT cells; among these factors, SOCS3 is able to rescue long-term self-renewal in MUT NSC, when overexpressed in these cells.

3.1 SOX2 Enrichment in Pol II-Mediated Long-Range Interactions between Promoters and Epigenetic Enhancers in NSC.

A relevant role for SOX2 in long-range interactions can be hypothesized on the basis of the following observations: distal interaction anchors are highly enriched in epigenetic enhancer marks (Figure 2) and are highly represented in enhancers

active in transgenic zebrafish (Figure 4) and in forebrain VISTA enhancers (Visel et al., 2009) (see Figures 3, 4, and S5). In addition, SOX2-bound sites are highly enriched in regions marked by epigenetic enhancer signatures (Figure 2B); in particular, SOX2-bound epigenetically defined enhancer regions are much more represented than non-SOX2-bound regions in anchors (Figure 2E). Finally, upon ablation of *Sox2*, there is a reduction in interactions frequency, which is detected by ChIA-PET (Figures 1, 3, S3, and S5; Table 1) at large numbers of loci. It is thus possible to hypothesize that SOX2, perhaps in complex with additional factors, may contribute to generate, or maintain, the network of interactions characteristic of NSCs.

A subset of P-P and P-E interactions is decreased in frequency in MUT cells. This might result from the loss of SOX2 from the interacting anchors (an appealing hypothesis) and the ensuing global chromatin conformation changes; an additional contribution to interaction loss might be represented by the transcriptional deregulation of many SOX2-controlled transcription factors (Tables S4 and S5) potentially contributing to interactions. Further, SOX2 interacts with several TFs, as well as with 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 interactions.

3.2 Decreased Gene Expression in Sox2 MUT Cells Is Significantly Correlated with Genes Whose Promoter Is Connected with an Enhancer Bound by SOX2 in WT

Cells The decreased transcription of ~1,000 genes (Tables S4 and S5) in *Sox2* MUT cells could in principle be ascribed to either loss of an effect of SOX2 on the gene promoter, or to loss of an effect on a connected enhancer. We demonstrate a

predominant role of SOX2-bound enhancers versus non-SOX2-bound enhancers on the regulation of their connected genes (Figure 5D); in contrast, SOX2 binding to promoters is much less functionally relevant. Some interacting promoters may influence each other's activities (Li et al., 2012); our data, while not ruling out this model, suggest that, overall, the numerous P-P interactions detected in NSCs play a comparatively minor role relative to P-E interactions in SOX2-dependent regulation (Figures 5B and 5D). This is consistent with our observation (Zhang et al., 2013) that P-E interactions are more cell-type specific than P-P interactions, and is in agreement with the known cell-type specificity of SOX2 functions in neural cells. The predominant transcriptional effect of SOX2 at distal enhancer regions might be related either to an activating effect of SOX2 onto the connected promoter or to a stabilizing effect of SOX2 onto the interaction itself (as suggested by the decreased frequency of the interaction upon *Sox2* ablation), or both.

The role of interactions in controlling gene activity has been addressed by the knockout of genes encoding CTCF or Cohesin components in regulating chromatin interactions in cell lines. Only moderate transcriptional deregulation was observed in connection with widespread and deep changes in long-range interactions (Merkenschlager and Odom, 2013; Rao et al., 2017; Schwarzer et al., 2017); these proteins, how- ever, are thought to act primarily as architectural proteins (Phil- lips-Cremins et al., 2013), in contrast to the well-established role of SOX2 as a transcription factor. During completion of this manuscript, the transcription factor YY1 was identified as a mediator of promoter-enhancer interactions in embryonic stem cells (Weintraub et al., 2017). YY1, contrary to SOX2, is ubiquitously expressed and acquires cell-type specificity of binding to DNA thanks to RNA and other undefined factors (Weintraub et al., 2017). Intriguingly, in neural progenitor cells (NPCs), YY1 is bound to a large subset of NPC-specific pro- moter-enhancer

interactions (Beagan et al., 2017). It will be interesting to ask whether SOX2 and YY1 may functionally interact at this level.

3.3 Sox2 Loss Does Not Substantially Alter Epigenetic Enhancer Marks

While SOX2 is bound to a very large proportion of epigenetic enhancers in NSC and plays an important role in the regulation of a subset of genes, epigenetic marks on enhancers are not lost in *Sox2* MUT NSCs (Figure S4). This might be explained by the fact that no gene is completely silenced in the absence of SOX2, and interactions might be decreased, but not completely lost. Additionally, *Sox2* was ablated at a stage (E11.5) when specific EM might already have been established within NSCs; SOX2 might be initially important in determining the transition from an ectodermic cell to a NSC, and thus the establishment of proper chromatin EM, but might not be required afterward for their maintenance. These observations dissociate the presence of an epigenetic EM from the actual presence of a critical transcription factor on interacting promoter-enhancer complexes.

3.4 SOX2 Loss Affects the Activity of Key Genes Relevant to Cell Proliferation Control.

Sox2 is expressed in NSCs 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 unknown which genes downstream to *Sox2* mediate its function in long-term NSC maintenance. *Socs3*, a highly connected gene (Figures 1 and 6) is strongly downregulated in *Sox2* MUT NSCs. *Socs3* transduction of a proportion of

Sox2 MUT cells leads to a slow but progressive increase of the growth rate of the culture, with eventual recovery of a population of actively growing cells, providing evidence for a crucial role of *Socs3* in *Sox2*-dependent NSC maintenance (Figure 6). Interestingly, several additional genes *(*c-*fos*, *Jun*, *JunB*, *Btg2*, *Egr1*, and *Egr2*), encoding well- known regulators of cell proliferation, are expressed at high levels in WT NSCs and are substantially downregulated in *Sox2* MUT cells (Tables S4 and S5). These genes show multiple promoter-enhancer interactions in WT NSCs (not shown and Table S5) and might be part of a network of interacting genes required, together with *Socs3*, for optimal *Sox2*-dependent maintenance of NSCs. Hippocampal defects observed in *Sox2* MUT mice have been related to defects in NSCs (Favaro et al., 2009). The discovery of mediators of *Sox2* function in NSCs may be relevant to the understanding of *in vivo* defects.

3.5 Many Mouse Homologs of Genes Affected in Neurodevelopmental Brain Diseases Are Involved in Long-Range Interactions with Distant Regions Carrying Enhancer Marks.

Thousands of polymorphisms in non-coding elements in man may be linked to brain disease or neurodevelopmental disorders (Nord et al., 2015). In NSCs, between ca. 1,750 and 3,500 expressed genes (Figure 5B, wTR1, wTR2, and wTR3) present long-range promoter-enhancer connections; 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, which might be dysfunctional in individuals carrying mutations at these elements.

Interestingly, the mouse homologs of several genes known to be involved in human neural disease show SOX2-bound interactions (Figures 3B–3D and S5). A SOX2-bound neural enhancer within the *Akt3* gene is connected to the *Zbtb18* (*ZFP238* in

man) TF gene (Figure 3B), whose mutation causes microcephaly in man and mouse (de Munnik et al., 2014); in man, deletions including *AKT3*, or translocations separating *AKT3* from *ZFP238* (Boland et al., 2007) (Figure 3B) also cause microcephaly. Both *Gpr56*, a gene whose promoter is bound by SOX2, and *Arid1a* (Figure 3) are connected to distant SOX2- bound epigenetic enhancers; mutation of the *GPR56* promoter causes structural neocortical abnormalities and *ARID1A* mutation is responsible for intellectual disability (Bae et al., 2014; Lee and Young, 2013, and references therein). The mutation of *SOX3* (see its enhancer in Figure 4) causes mental retardation and hypothalamic-pituitary defects (Laumonnier et al., 2002). Table S6 (see also Figure S5) lists additional genes involved in long-range interactions (many of which are SOX2-bound), whose human homologs are affected in neurodevelopmental disorders. In particular, Table S6 includes a large proportion of genes mutated in primary recessive microcephaly, severe intellectual disability, and eye disease. Significantly, the pathology of *SOX2* mutant patients includes brain (mainly hippocampal) abnormalities, some degree of mental retardation, and eye defects (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 et al., 2009; Ferri et al., 2004, 2013). Our data suggest that some of the genes showing connections in Table S5 might play a role in human and mouse *Sox2*-dependent pathology; additionally, it might be interesting to search for mutations of the connected enhancers in human diseases such as intellectual disability and microcephaly.

Acknowledgment

We thank Daofeng Li and Ting Wang (Washington University, http://wang.wustl.edu/) for helping us host and visualize our genomic data on the WashU browser. We thank Dr. Tae-Kyung Kim for the *c-fos* enhancers. Research re- ported in this publication was supported by Telethon, AIRC, Regione Lombardia ASTIL, Fondazione Cariplo, ERANET ImprovVision (NEURON8-Full-815- 091 to S.K.N.), MINECO (BFU2013-43213-P and BFU2016-75412-R with FEDER support), CIBERER, ISCIII, ERANET/PCIN-2015-176-C02-01, an Institutional Grant from Fundacion Ramon Areces and Banco Santander (to P.B.), the 4DN (U54 DK107967), ENCODE Consortia (UM1 HG009409 to C.-L.W.), NCI (P30CA034196 to C.-L.W. and C.Y.N.), the Francis Crick Institute (which receives its core funding from Cancer Research UK; FC0010089), the UK Med- ical Research Council (FC0010089), the Wellcome Trust (FC0010089 to F.G. and B.M.), and the UK Biotechnology and Biological Sciences Research Council (BB/K005316/1 to F.G.). J.B. 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., and P.B. discussed the results and wrote the manuscript with input from all other authors. R.F. and M.P. prepared NSCs for all chromatin studies. Y.Z. and H.T. performed and analyzed the TR2 and TR3 ChIA-PET data. C.-H.W. analyzed TR1 ChIA-PET data. C.Y.N. performed RNA-seq. B.M. and M.V. did Sox2 and histone marks ChIP-seq. P.R. and J.M. did initial RNA-seq work. G.P. and F.Z. planned and performed bioinformatics and biostatistics analyses. R.F., C.B., J.B., and S.M. made DNA constructs and transfections. J.B., M.J.C., and N.T. performed zebrafish studies under the guidance of P.B. M.P. did the MUT NSC rescue experiments.

Experimental model and subject detail

Animals

Sox2 conditional mutant mice

Mutant and wild-type mice were sacrificed at P0, to obtain forebrains for NSC cultures (sex was indifferent). *Sox2* deletion was obtained by breeding *Sox2*flox mutant mice to *nestinCre* transgenic mice (as in Favaro et al., 2009). The experiments were approved by the Italian Ministery of Health as conforming to the relevant regulatory standards.

Zebrafish

AB and *tupl* wild-type zebrafish strains were maintained and bred according to standard procedures (Fishman et al, 1997). All experiments conform to the guidelines from the European Community Directive and the Spanish legislation for the experimental use of animals.

Cell lines, primary cultures, microbe strains

Primary ex-vivo neural stem/progenitor cell 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), see Method Details section below.

Microbe strains.

Standard cloning procedures were carried out in *E. coli* TOP ten and DH5alpha.

Lentivirus packaging cell lines

Packaging of the *Socs3*-expressing lentivirus was performed in 293T cells.

Method details

Cultures of neural stem/progenitor cells from wild-type and Sox2-deleted mouse P0 forebrain

After forebrain dissection and cell dissociation, we plated cells in 25 mL flasks and cultured them to expand their number in complete medium (2% (vol/vol) B27 in DMEM F12 with Glutamax), supplemented with 10 ng/ml EGF, 10 ng/ml of basic fibroblast growth factor (FGF) with 0.2% (vol/vol) heparin. We cultured wild-type and mutant cells in parallel for two initial passages (until sphere formation was detected, 3-7 days, normally 4 days) (Favaro et al., 2009), followed by sphere dissociation and further expansion in the presence of EGF, but not bFGF, for 3-5 more passages (4 days), as described in Zhang et al. (2013), for optimal maintenance of mutant NSC (Favaro et al., 2009). Cell passaging involved 0.25% trypsin treatment for 5 min, followed by block with 1 mg/ml trypsin inhibitor (Sigma T6522) for 5 min; neurospheres were then mechanically dissociated by pipetting and cells were seeded at 80,000 per ml in T150 flasks.

After obtaining appropriate numbers of wild-type and mutant cells, we started collecting part of the neurospheres at each passage, continuing the culture of the remaining cells as long as the growth of mutant cells was comparable to that of wild-type cells. In order to proceed to ChIA-PET and ChIP-seq experiments, we mechanically dissociated and crosslinked the neurospheres as described below.

ChIA-PET experiments

ChIA-PET experiments for TR1were performed as described (Zhang et al., 2013), using pooled NSC from four wild-type and six mutant brains from littermates. Wild-type and mutant cultures were processed in parallel. wTR1 data are from Zhang et al. (2013); mTR1 data, present paper. Neurospheres were cross-linked by standard formaldehyde treatment and the pellets were then snap-frozen in nitrogen.

The crosslinked cells were lysed to release the chromatin–DNA complexes followed by fragmentation to an average size of 300 base pairs (bp). The sonicated chromatin–DNA complexes were incubated with the Pol II monoclonal antibody-coated magnetic Protein G beads (8WG16, Covance). To determine the ChIP quality, a small portion of ChIP DNA was eluted for quantitative PCR (qPCR) analysis. For ChIA-PET library construction, ChIP-enriched chromatin complexes were divided into two aliquots. To distinguish the intramolecular proximity ligation products from the chimeras resulting from non-specific inter- molecular ligations, two different barcoded biotinylated half-linkers (linker A and linker B) were ligated to the ends of polished bead-bound-DNA fragments and used to join the juxtaposed chromatin regions. The half-linker-ligated chromatin–DNA fragments were pooled for phosphorylation and proximity-based circularization. MmeI was subsequently used to release the paired-end tags (PETs). The full-length linkers AA/BB resulting from intra-molecular circularization were considered to be non-chimeric PETs. Conversely, the chimeric full linkers AB/BA resulting from intermolecular ligation were considered to be ligation noises. The biotin-labeled PET constructs were amplified and subjected to sequencing analysis.

For the *in situ* Pol II ChIA-PET 10 million formaldehyde crosslinked cells for each analysis (wTR2, wTR3 and mTR2, mTR3), (cells grown from individual brains, as above) were suspended with 100 mL 0.55% SDS and incubated at RT for 10 min, 62C for 10 min and 37C for 10 min to permeabilize nuclei, which was followed by addition of 25 mL 20% Triton X-100, incubation of 30 min at 37C to quench SDS, followed by addition of 50 mL of blunt-end four-cutter AluI (cat# R0137L, NEB), 50 mL 10 3 CutSmart buffer and 275 mL H2O and incubation at 37 C overnight for *in situ* digestion. After pelleting and washing once with 1 mL 1 3 CutSmart buffer, the nuclei were suspended in 500 mL A tailing solution composed of 50 mL 10 3 CutSmart buffer, 10 mL BSA (cat#B9000S, NEB), 10 mL 10 mM dATP

(cat#N0440S, NEB), 10 mL Klenow (30-50 exo-) (cat#M0202L, NEB), and 420 mL H2O and incubated at RT for 1 hr. The *in situ* proximity ligation with biotinylated bridge linker was performed by adding 200 mL 5 3 ligation buffer (cat#B6058S, NEB), 6 mL Bridge linker (200 ng/ul), 10 mL T4 DNA ligase (cat#M0202L, NEB) and 284 mL H2O and incubating at 16C overnight. The nuclei with *in situ* proximity ligation were then subjected to sonication and chromatin immunoprecipitation with anti-Pol II antibody (8WG16, cat# 920102, Biolegend, San Diego, CA), Tn5 tagmentation, and biotin selection; PETs were amplified by PCR and sequenced.

ChIA-PET data analyses

The TR1 sequence data generated from the original ChIA-PET protocol were analyzed using ChIA-PET tool (Li et al., 2010). In brief, non-redundant PET sequence reads were first analyzed for linker barcode composition and non-chimeric PETs were used for further analysis. Next, the linker sequences were trimmed, and the PET sequences were mapped to the mouse reference genome (mm9) with 1 mismatch allowed. The PETs with genomic locations from both head and tail tags within 2 bp were merged to further filter the redundancy arising from clonal PCR amplification. Based on the mapping coordinates, the specific-ligation PETs were used for further classification as inter-chromosomal, intra-chromosomal and self-ligation PETs. Inter-chromosomal PETs were defined as both the head and tail of the PETs uniquely mapped onto different chromosomes. To define highly reliable interaction clusters, we adopted the false discovery rate (FDR) of the hyper-geometric model. Such model takes into consideration the tag counts from both anchor regions and the sequencing depth to determine reliable, i.e., significant, interactions. A FDR cutoff 0.05 was used. Finally, we performed a random shuffling simulation to evaluate the correlation between noise level and PET cluster

counts. The simulation broke down the pairing relationship of different PET clusters and the tags were randomly paired to generate simulated PETs. We further compared the interaction numbers between simulations versus our experimental data and determined the noise level (number of simulated clusters/number of real clusters) for the PET-2+ (PET cluster with 2 counts and above) to be 1.2%. Therefore, we chose the PET cluster > = 2 to keep the noise level low and singletons (PET = 1) were considered as noise. For TR2 and TR3 generated by the *in situ* ChIA-PET protocol, ChIA-PET data were processed with ChIA-PET Utilities, a scalable re-implementation of ChIA-PET tool. Briefly, sequencing adaptors incorporated during the tagmentation reaction in the library construction process were removed from the paired reads. Tags identified (> = 18bp) were mapped to mouse genome (mm9) using BWA alignment (Li and Dur- bin, 2009) and memarXiv:1303.3997 [q-bio.GN], https://arxiv.org/abs/1303.3997 according to their tag length. The duplicated pair- end tags arising from clonal PCR amplification were filtered and the uniquely mapped, non-redundant PETs were classified as inter- chromosomal (L tags and R tags mapped onto different chromosomes), intra-chromosomal (L tags and R tags mapped onto the same chromosome with genomic distance > 8Kb) and self-ligation PETs (L tags and R tags mapped onto the genome % 8Kb). Multiple intra-chromosomal PETs whose respective ends were found within 1 Kb were then clustered as iPET-2, 3 . We further per- formed statistical assessment of the PET clusters interaction significance using ChiaSigScaled, a scalable re-implementation of ChiaSig (Paulsen et al., 2014). Interaction clusters with member size 3 and above (iPET 3+) and FDR < 0.05 were classified as significant interactions (Table 1, line 7). Among these, the interactions with Pol II binding at both anchors were further defined as Pol II-mediated interactions (Table 1, line 8). Next, the interactions were classified based on their anchors overlapped with gene models. Each anchor was annotated with the gene that overlapped at 1bp

overlap. To classify each anchor, priority was given to promoter (P) region (defined as ± 2.5kb of TSS) followed by gene region (G). Anchors that do not overlap with any gene or promoter region were classified as intergenic (I). The interaction classification is just the combination of its anchors classification. To determine the reproducibility between TR2 and TR3, we adopted the method used in the Hi-C data (Yang et al., 2017) to determine the SCC, Stratum-adjusted Correlation Coefficient. To compute the SCC, ChIA-PET loops for each library were aggregated into 10-kb bin matrix. SCC score is computed with HiCrep method (hicrep library in R).

For each chromosome, the smoothing parameter was used as recommended (h = 3), with maximum distance 1 Mb. Because the methods used to fragment chromatins and generate tags for sequencing were different in TR1, TR1 was not included in the SCC analysis. As TR1 used sonication shearing and MmeI digestion while TR2 and TR3 used AluI digestion and Tn5 transposon-based tagmentation, the exact anchor locations defined by these two methods cannot be directly compared.

ChIP-seq ANALYSES

SOX2 ChIPseq

NSC from 6 wild-type forebrains, at P0 (Favaro et al., 2009), were independently grown, and analyzed by ChIPseq in duplicate. Individual cultures were pooled together, at a stage when neurospheres were still relatively small, and then divided into two aliquots: neurospheres were directly fixed for the first ChIP (‘‘spheres’’ ChIPseq), whereas single cells (resulting from dissociation of the same neurospheres) were used for the second ChIP (‘‘singles’’ ChIPseq) (see GEO).

Cells 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).

By using default significance thresholds, this resulted in 18,359 SOX2-bound peak regions for the first (''spheres'') ChIPseq. Since fixation of whole neurospheres might not, in theory, be equally efficient for internal relative to external cells, we also performed another ChIP experiment (''singles'' ChIPseq). Read mapping and peak calling were performed with the same parameters used in the first experiment, producing 43,070 bound regions at the same significance thresholds. Of the first dataset (''spheres'' ChIPseq), the vast majority (15,985 peaks, 87%) were contained in the latter (''singles'' ChIPseq) peak list. Since also the 13% non-overlapping peaks for the first experiment showed enrichment in the corresponding loci in the second ChIPseq, even if below the ''peak detection threshold,'' we kept for all the subsequent analyses presented in the paper all the peaks returned by the first experiment (''spheres'').

The difference between ''spheres'' and ''singles'' ChIPseq appears to be due mainly to the presence in ''singles'' of numerous additional small peaks that have no (significant) corresponding peaks in the ''spheres'' sample, and may represent more marginal binding sites. Indeed, comparison of profiles between ''spheres''

and ‘‘singles’’ showed little differences between them, indicating that the main binding sites are very similar (data not shown).

Finally, over 50% of the ca. 18400 peaks observed in our forebrain NSC ‘‘spheres’’ are found also within the ca. 24800 peaks detected in a ChIPseq analysis of the NS-5 cell line (Mateo et al., 2015), an ES-derived NSC line that shows a general (nonfore- brain-specific) neural phenotype. For these reasons, we used for subsequent analyses the data from ‘‘spheres.’’

H3K27Ac and H3K4me1 ChIPseq

NSC were derived from six forebrains of wild-type and 6 forebrains of *Sox2*-deleted mice, at P0 (Favaro et al., 2009). NSC were initially cultured individually, then pooled according to wild-type or mutant genotype to generate two independent pools. Each independent pool was divided in two parts, and used for ChIP-sequencing on H3K27ac or H3K4me1, as described previously (Vermunt et al., 2016). 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 160 mL sonication buffer and divided over two microtubes for shearing in the Covaris S series with the following settings for 12 cycles of 60 s: intensity 3, duty cycle 20%, 200 cycles/bursts. Chromatin immunoprecipitation steps after sonication were performed as described previously (Vermunt et al., 2016) using 50 mL DynaI protein G beads that were preincubated with 5 mg Ab4729 (Abcam) for H3K27ac or 5 mg 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 (https://openwetware.org/wiki/BioMicroCenter).

Obtained se- quences were aligned onto the mm9 mouse genome assembly using Bowtie 1.1.0 (http://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 = 105, 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.

Analysis of histone modifications colocalization by ChromHMM

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-occur- rence of histone modifications in the genome segments, building a model in which each of the states is characterized by a given com- bination 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 sam- ples; 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, then 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

Enrichment of SOX2-bound sites within H3K27Ac-enriched regions

Significance of overlap between H3K27ac-enriched regions and SOX2 binding sites was analyzed in comparison to 1,000 sets of random genomic DNA (random sampling). These sets comprised the same number of elements of similar size selected randomly from the mm9 genome excluding GAP regions (UCSC genome browser), blacklisted regions (cite PMID: 22955616) and unmappable regions. To define unmappable regions, bam-files from all H3K27ac and H3K4me1 ChIP-seq datasets (n = 8) were merged and map- ped onto the mm9 genome that was binned in sliding windows of 3000 bp with an overlap of 500 bp. Bins with zero reads were defined as unmappable and thus excluded.

Zebrafish transgenesis

Sequences from 17 of the identified anchors (15 distal, and 2 proximal) were amplified from the mouse genome using specific primers (Table S1) and cloned into a pBluescript vector by a pCR8/GW/TOPO TA Cloning® Kit (Life Technologies). Individual fragments were then transferred, using recombination-mediated, ClonaseTM-assisted, 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 trans- genesis 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 color 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 trans- genic lines each by crossing with wild-type animals. In the case of the en1 and en2 c-*fos* enhancer embryos were analyzed only at F0. To confirm a link between *Sox2* and the identified elements, embryos derived from the F1 lines were microinjected in one blastomere at 1-2 cell stage with a *sox2* specific morpholino (GeneTools; 50-500 nM), previously reported to efficiently interfere with *sox2* expres- sion in zebrafish, without however causing major morphological defects (Okuda et al., 2010). To complement this study, the pCS2 plasmid containing the *sox2* coding

sequence was linearized and *in vitro* transcribed using the SP6 MessagemMachine kit (Ambion). The synthesized mRNA (Esteve et al., 2004) was purified using Quiaquick RNeasy columns (Quiagen), precipitated, quantified and injected at the concentration of 100ng/ul into embryos derived from the F1 lines as above. Embryos were grown and scored for increased or reduced reporter expression at 15, 18, 24 and 48 hpf stages against the levels presented in untreated embryos or embryos injected with either a Genetools standard control MO or an unrelated (mCherry) mRNA, used as controls.

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 1mg 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, adaptor 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.

Cell Stem Cell *24*, 462–476.e1–e6, March 7, 2019 e5

Expression boxplots 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 (wTR1, wTR2, wTR3) 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 –log10 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), log10 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. The coassociation scores have been calculated using genes expressed at levels of at least 5 TPM in wt NSC; we also did the same analysis using all expressed genes: the results are qualitatively similar, and are not shown.

Socs3 transduction in NSC

The TWEEN lentiviral vector expressing SOCS3 from the CMV promoter and GFP driven by the human PGK promoter (Francipane et al., 2009) was transfected into low-passage 293T cells by calcium phosphate precipitation (overnight) together with the VSV-G plasmid (encoding ENV), CMV R8.74 (packaging) and pRSV-REV (encoding reverse transcriptase). Following replacement with fresh medium, the cell supernatants were collected at 24-48 hours from transfection. For NSC transduction, wild-type and *Sox2* mutant neurospheres (obtained as in Favaro et al., 2009) were grown for 2-3 passages in bFGF and EGF, and for one more passage in EGF only. They were then dissociated to single cells and seeded at a density of 25,000 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 trans- duced with the *Socs3*-expressing vector at a multiplicity of

infection (MOI) of 3.5-5.5 and incubated overnight at 37C. 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 20,000 cell/well in the same EGF medium described above. In addition, 500,000 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).

DATA AND SOFTWARE AVAILABILITY

Accession numbers

The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (GEO). The accession number for the ChIA-PET, ChIPseq and RNAseq data reported in this paper is GEO: GSE90561

The genomic data can be visualized through the WashU browser:

http://epigenomegateway.wustl.edu/legacy/?genome=mm9&datahub=https://wangftp.wustl.edu/~dli/7131149234337a58201ae3da174ecc51/hub&coordinate=chr8:87120161-87587163

References

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

Bae, B.I., Tietjen, I., Atabay, K.D., Evrony, G.D., Johnson, M.B., Asare, E., Wang, P.P., Murayama, A.Y., Im, K., Lisgo, S.N., et al. (2014). Evolutionarily dynamic alternative splicing of GPR56 regulates regional cerebral cortical patterning. Science *343*, 764–768.

Beagan, J.A., Duong, M.T., Titus, K.R., Zhou, L., Cao, Z., Ma, J., Lachanski, C.V., Gillis, D.R., and Phillips-Cremins, J.E. (2017). YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. Genome Res. *27*, 1139–1152.

Bessa, J., Tena, J.J., de la Calle-Mustienes, E., Ferna´ ndez-Min˜ a´ n, A., Naranjo, S., Ferna´ndez, A., Montoliu, L., Akalin, A., Lenhard, B., Casares, F., and Go´ mez-Skarmeta, J.L. (2009). Zebrafish enhancer detection (ZED) vector: a new tool to facilitate transgenesis and the functional analysis of cis-regulatory regions in zebrafish. Dev. Dyn. *238*, 2409–2417.

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. J. Neurochem. *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 pre- dicts developmental state. Proc. Natl. Acad. Sci. USA *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., Garcı´a-Min˜au´r, 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. Eur. J. Hum. Genet. *22*, 844–846.

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 dis- covery and characterization. Nat. Methods *9*, 215–216.

Esteve, P., Lopez-Rios, J., and Bovolenta, P. (2004). SFRP1 is required for the proper establishment of the eye field in the medaka fish. Mech. Dev. *121*, 687–701.

Favaro, R., Valotta, M., Ferri, A.L., Latorre, E., Mariani, J., Giachino, C., Lancini, C., Tosetti, V., Ottolenghi, S., Taylor, V., and Nicolis, S.K. (2009). Hippocampal development and neural stem cell maintenance require Sox2-dependent regu- lation of Shh. Nat. Neurosci. *12*, 1248–1256.

Ferri, A.L., Cavallaro, M., Braida, D., Di Cristofano, A., Canta, A., Vezzani, A., Ottolenghi, S., Pandolfi, P.P., Sala, M., DeBiasi, S., et al. (2004). Sox2 defi- ciency causes neurodegeneration and impaired neurogenesis in the adult mouse brain. Development *131*, 3805–3819.

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.

Fishman, M.C., Stainier, D.Y., Breitbart, R.E., and Westerfield, M. (1997). Zebrafish: genetic and embryological methods in a transparent vertebrate em- bryo. Methods Cell Biol. *52*, 67–82.

Francipane, M.G., Eterno, V., Spina, V., Bini, M., Scerrino, G., Buscemi, G., Gulotta, G., Todaro, M., Dieli, F., De Maria, R., and Stassi, G. (2009). Suppressor

of cytokine signaling 3 sensitizes anaplastic thyroid cancer to standard chemotherapy. Cancer Res. *69*, 6141–6148.

Gonc ¸ alves, J.T., Schafer, S.T., and Gage, F.H. (2016). Adult Neurogenesis in the Hippocampus: From Stem Cells to Behavior. Cell *167*, 897–914.

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

Hagey, D.W., Zaouter, C., Combeau, G., Lendahl, M.A., Andersson, O., Huss, M., and Muhr, J. (2016). Distinct transcription factor complexes act on a permissive chromatin landscape to establish regionalized gene expression in CNS stem cells. Genome Res. *26*, 908–917.

Langmead, B. (2010). Aligning short sequencing reads with Bowtie. Curr. Protoc. 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, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Li, G., Fullwood, M.J., Xu, H., Mulawadi, F.H., Velkov, S., Vega, V., Ariyaratne, P.N., Mohamed, Y.B., Ooi, H.S., Tennakoon, C., et al. (2010). ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. Genome Biol. *11*, R22.

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.

Lodato, M.A., Ng, C.W., Wamstad, J.A., Cheng, A.W., Thai, K.K., Fraenkel, E., Jaenisch, R., and Boyer, L.A. (2013). SOX2 co-occupies distal enhancer ele- ments with distinct POU factors in ESCs and NPCs to specify cell state. PLoS Genet. *9*, e1003288.

Mateo, J.L., van den Berg, D.L., Haeussler, M., Drechsel, D., Gaber, Z.B., Castro, D.S., Robson, P., Lu, Q.R., Crawford, G.E., Flicek, P., 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.

Mumbach, M.R., Rubin, A.J., Flynn, R.A., Dai, C., Khavari, P.A., Greenleaf, W.J., and Chang, H.Y. (2016). HiChIP: efficient and sensitive analysis of pro- tein-directed genome architecture. Nat. Methods *13*, 919–922.

Nord, A.S., Pattabiraman, K., Visel, A., and Rubenstein, J.L.R. (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.

Paulsen, J., Rodland, E.A., Holden, L., Holden, M., and Hovig, E. (2014).). A statistical model of ChIA-PET data for accurate detection of chromatin 3D interactions. Nucleic Acids Res. 42, e143.

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 anophthal- mia syndrome. Am. J. Med. Genet. A. 135, 1–7, discussion 8.

Rao, S.S.P., Huang, S.C., Glenn St Hilaire, B., Engreitz, J.M., Perez, E.M., Kieffer-Kwon, K.R., Sanborn, A.L., Johnstone, S.E., Bascom, G.D., Bochkov, I.D., et al. (2017). Cohesin Loss Eliminates All Loop Domains. Cell 171, 305–320.

Rivera, C.M., and Ren, B. (2013). Mapping human epigenomes. Cell *155*, 39–55.
Sanyal, A., Lajoie, B.R., Jain, G., and Dekker, J. (2012). The long-range interaction landscape of gene promoters. Nature *489*, 109–113.

Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., Fonseca, N.A., Huber, W., H Haering, C., Mirny, L., and Spitz, F. (2017). Two independent modes of chromatin organization revealed by cohe- sin removal. Nature *551*, 51–56.

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.

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.

Tarazona, S., Furio ´ -Tarı ´, 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.
Cell Stem Cell 24, 462–476, March 7, 2019 475

Vermunt, M.W., Tan, S.C., Castelijns, B., Geeven, G., Reinink, P., de Bruijn, E., Kondova, I., Persengiev, S., Bontrop, R., Cuppen, E., et al.; Netherlands Brain Bank (2016). Epigenomic annotation of gene regulatory alterations during evo- lution 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.

Weintraub, A.S., Li, C.H., Zamudio, A.V., Sigova, A.A., Hannett, N.M., Day, D.S., Abraham, B.J., Cohen, M.A., Nabet, B., Buckley, D.L., et al. (2017). YY1 Is a Structural Regulator of Enhancer-Promoter Loops. Cell 171, 1573–1588.

Yang, T., Zhang, F., Yardımcı, G.G., Song, F., Hardison, R.C., Noble, W.S., Yue, F., and Li, Q. (2017). HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. Genome Res. 27, 1939–1949.

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

Zhu, J., Adli, M., Zou, J.Y., Verstappen, G., Coyne, M., Zhang, X., Durham, T., Miri, M., Deshpande, V., De Jager, P.L., et al. (2013). Genome-wide chromatin state transitions associated with developmental and environmental cues. Cell 152, 642–654.

Tables and Figures

		wTR1	wTR2	wTR3	mTR1	mTR2	mTR3
1	Total PETs with linker	143.3 M	35,295,370	33,767,613	151.9 M	24,976,862	21,514,824
2	Mapped PETs	143.1 M	23,672,163	20,624,486	151.7 M	16,512,951	11,050,816
3	Unique PETs	6.8 M	22,208,305	16,873,008	22.6 M	15,756,748	10,160,601
4	Self-ligated PETs	684 K	4,296,406	2,869,465	2.7 M	2,185,670	1,729,424
5	Number of Pol II binding sites ($p < 1e{-}5$)	11,819[a]	41,187[a]	36,641[a]	12,068[a]	21,020[a]	36,139[a]
6	Intra-molecular chromatin ligated PETs	691 K	13,105,813	10,207,155	1.3 M	9,360,576	6,119,151
7	Significant interactions (loops) (FDR <0.05, $p < 0.05$)	7,046[a]	96,295	63,458	2,984[a]	29,713	15,561
8	Significant loops with Pol II peaks on both anchors		18,022[a]	7,346[a]		2,878[a]	3,202[a]
9	Number of significant loops per million intra-chr PETs	10,197	7,348	6,217	2,295	3,174	2,543
10	Number of significant Pol II-bound loops per million intra-chr PETs		1,375	720		307	523

Table 1. Summary of the ChIA-PET Sequencing and Interaction Analysis

ChIA-PET data in triplicates were processed to define the binding peaks and significant interactions (see STAR Methods). For full list of significant interactions, see Table S1. PET, paired end tag; unique PETs, PETs for which the sequence of either side of the linker (e.g., the biotinylated linker in Figure S1B) can be uniquely mapped to one specific point in the reference genome. [a]Numbers of the significant interactions used in further analyses are indicated.

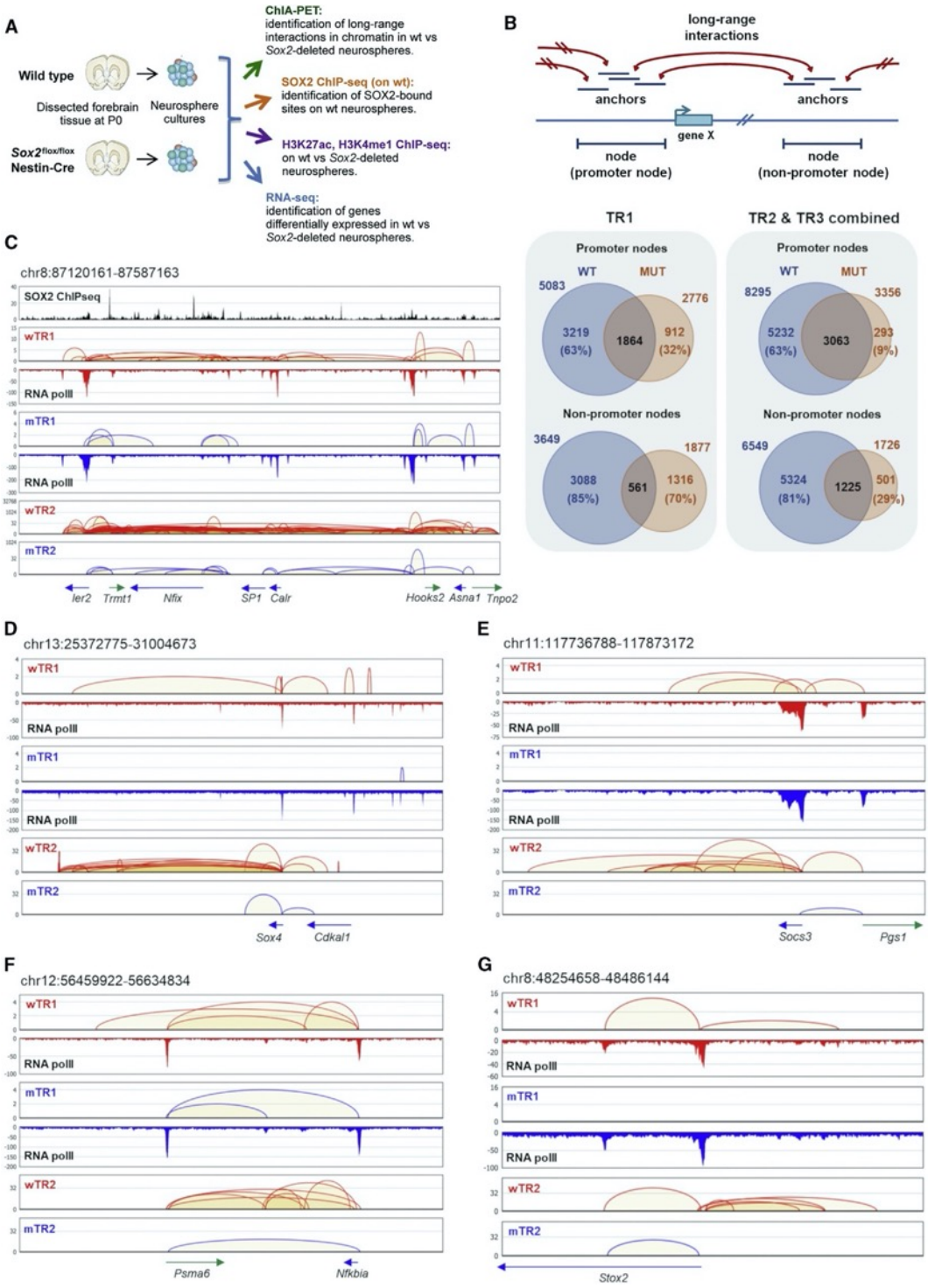

A
Wild type
Dissected forebrain tissue at P0
Neurosphere cultures
Sox2flox/flox Nestin-Cre
ChIA-PET:
identification of long-range interactions in chromatin in wt vs Sox2-deleted neurospheres.
SOX2 ChIP-seq (on wt):
identification of SOX2-bound sites on wt neurospheres.
H3K27ac, H3K4me1 ChIP-seq:
on wt vs Sox2-deleted neurospheres.
RNA-seq:
identification of genes differentially expressed in wt vs Sox2-deleted neurospheres.
B
long-range interactions
anchors
anchors
gene X
node (promoter node)
node (non-promoter node)
TR1
TR2 & TR3 combined
Promoter nodes
WT
MUT
5083
2776
3219 (63%)
1864
912 (32%)
Non-promoter nodes
3649
1877
3088 (85%)
561
1316 (70%)
Promoter nodes
WT
MUT
8295
3356
5232 (63%)
3063
293 (9%)
Non-promoter nodes
6549
1726
5324 (81%)
1225
501 (29%)
C
chr8:87120161-87587163
SOX2 ChIPseq
wTR1
RNA polII
mTR1
RNA polII
wTR2
mTR2
Ier2
Trmt1
Nfix
SP1
Calr
Hooks2
Asna1
Tnpo2
D
chr13:25372775-31004673
wTR1
RNA polII
mTR1
RNA polII
wTR2
mTR2
Sox4
Cdkal1
E
chr11:117736788-117873172
wTR1
RNA polII
mTR1
RNA polII
wTR2
mTR2
Socs3
Pgs1
F
chr12:56459922-56634834
wTR1
RNA polII
mTR1
RNA polII
wTR2
mTR2
Psma6
Nfkbia
G
chr8:48254658-48486144
wTR1
RNA polII
mTR1
RNA polII
wTR2
mTR2
Stox2

Figure 1. Sox2 Ablation Causes Major Loss of Long-Range Interactions in Brain-Derived NSCs.

(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 NSCs, left: TR1 and right: TR2 and TR3 combined. (C–G) Connectivity diagrams in WT NSCs (WT interactions; red) and MUT NSCs (MUT interactions; blue), across 5 different chromosome regions, in the wTR1, wTR2, and mTR1, mTR2 analysis; regions coordinates are: chr8:87120161-87587163. (C), chr13:25372775-31004673. (D), chr11:117736788-117873172. (E), chr12:56459922-56634834. (F) and chr8:48254658-48486144 (G). Their genomic coordinates are indicated above each panel, and genes within each region shown below the panels. Pol II- and SOX2-binding peaks are shown. PET counts (Y axes); note different (log10) scales in some panels. In MUT NSCs, an overall decrease of ‘‘looping’’ is seen, but some interactions are lost, others are maintained. Note the persistence of Pol II binding in MUT NSCs and the frequent coincidence (in C) of SOX2 peaks with interaction anchors. See also Figures S1, S2, S3 and Table 1.

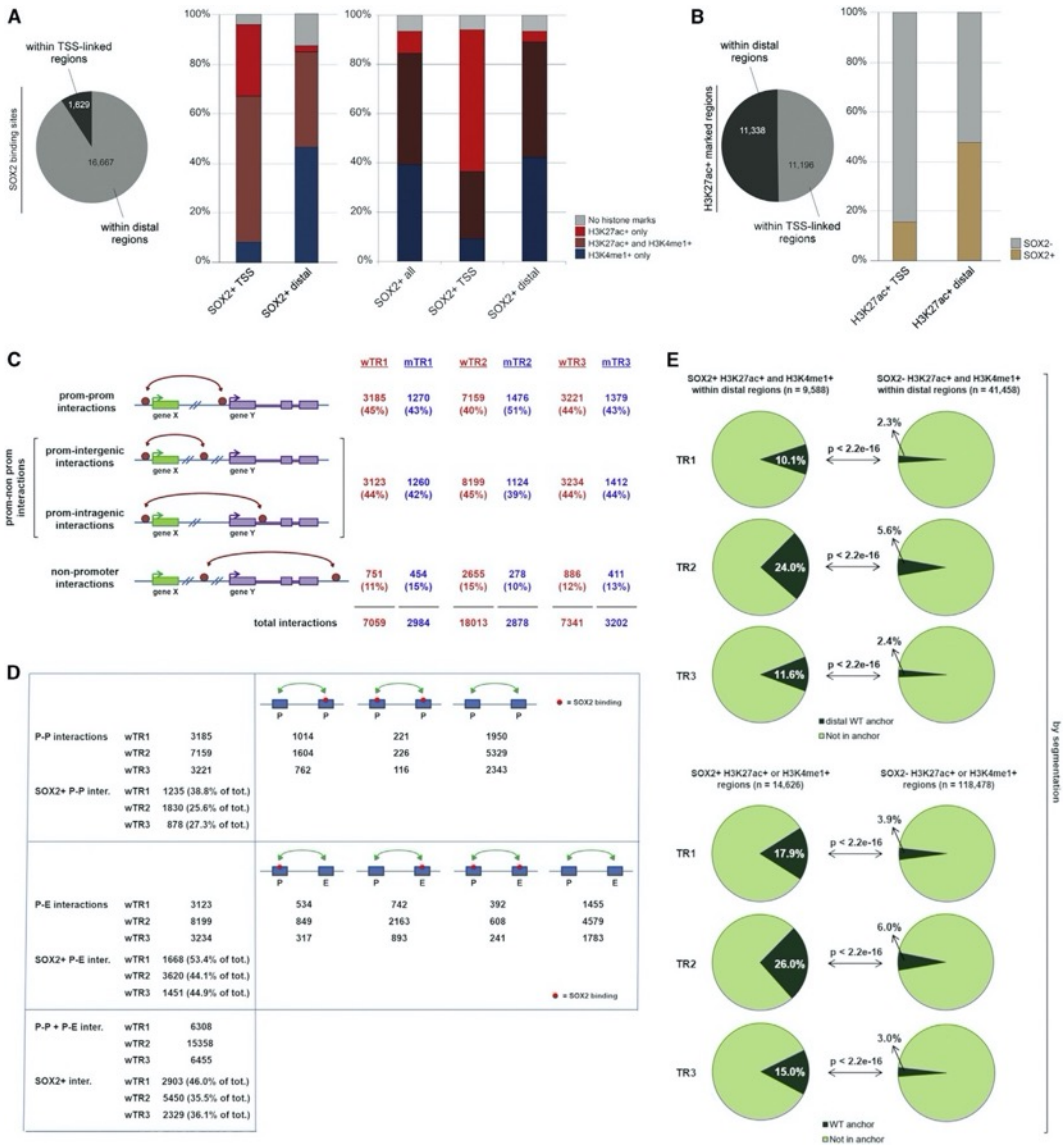


Figure 2. SOX2-Bound Regions Carrying Epigenetic Enhancer Marks (EM) Show Significantly Higher Overlap with Anchors Than SOX2- Negative EM-Positive Regions. (A) Left: number of SOX2-bound sites in regions linked to annotated TSS (±1,000 nt) and in distal, non-P regions. Right (histograms): percentage of different enhancer marks (EMs)-positive regions within SOX2-bound TSS-linked (SOX2+ TSS) or distal (SOX2+ distal) regions (left histograms, peak

calling; right histo- grams, chromHMM). (B) Fraction of SOX2-bound sites within EM-positive regions (H3K27Ac+) on TSS-linked or distal regions (peak calling). (C) Interaction types according to the nature of the connected regions, for wTR1, wTR2, wTR3, mTR1, mTR2, and mTR3. ''Prom,'' annotated TSS-containing region (i.e., promoter). (D) Numbers of P-P and P-nonP (P-E) SOX2-positive interactions in WT cells in wTR1, wTR2, and wTR3. See also Table S4. (E) Fraction of SOX2+ (left) versus SOX2 (right) EM-positive regions that overlap with anchors in wTR1, wTR2, and wTR3. Top: distal epigenetically marked regions (H3K27Ac+ and H3K4me1+) overlap with distal anchors. Bottom: all epigenetically marked regions (H3K27Ac+ or H3K4me1+) overlap with all anchor types, chromHMM. See also Figure S4 and Tables S4 and S5.

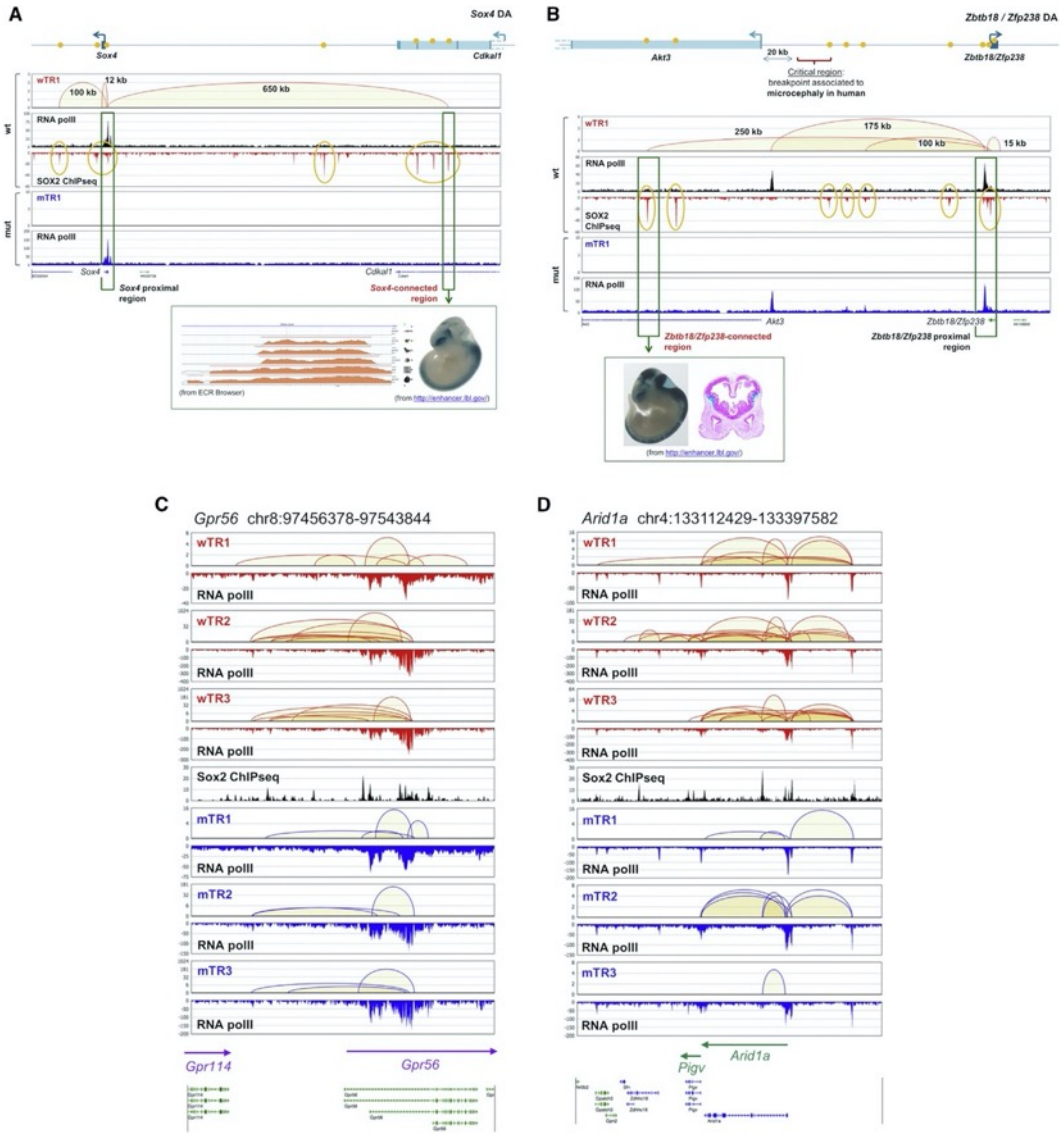


Figure 3. Distal Anchors Connected by Sox2-Dependent Interactions to Genes Important in Neural Development and Disease. (A and B) *Sox2*-dependent ChIA-PET interactions (TR1) between two different genes (*Sox4*, A; *Zbtb18*, B) and distal regions overlapping previously characterized ‘‘VISTA’’ enhancers (Visel et al., 2009); SOX2 ChIP-seq peaks (present paper), lacZ-stained transgenic embryos (from https://enhancer.lbl.gov), and evolutionary conservation (ECR

browser). (C and D) *Sox2*-dependent interactions (wTR1, wTR2, wTR3, mTR1, mTR2, and mTR3) involving *Gpr56* (C) and *Arid1a* (D), two genes whose human homologs are involved in neurodevelopmental brain disease. See also Figure S5 and Tables S5 and S6.

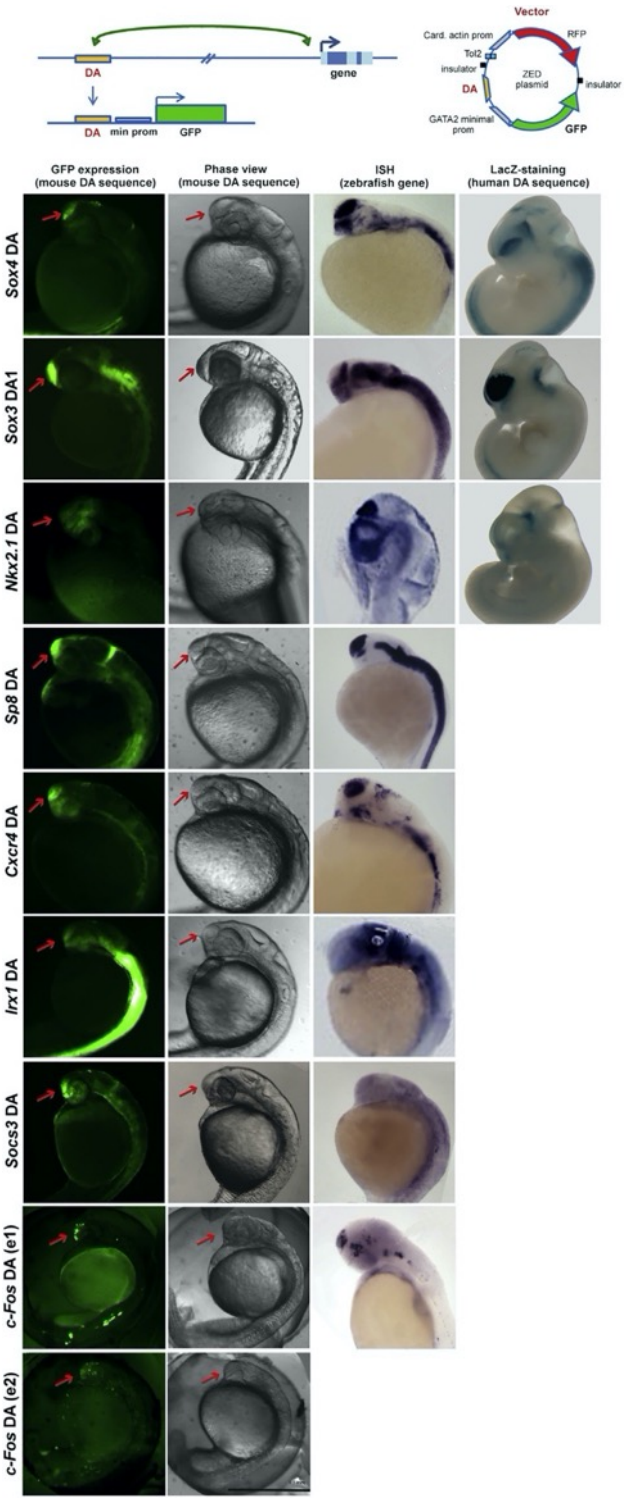

DA
gene
DA
min prom
GFP
Vector
Card. actin prom
RFP
Tol2
insulator
ZED
plasmid
insulator
DA
GATA2 minimal
prom
GFP
GFP expression
(mouse DA sequence)
Phase view
(mouse DA sequence)
ISH
(zebrafish gene)
LacZ-staining
(human DA sequence)
Sox4 DA
Sox3 DA1
Nkx2.1 DA
Sp8 DA
Cxcr4 DA
Irx1 DA
Socs3 DA
c-Fos DA (e1)
c-Fos DA (e2)

Figure 4. Distal Anchors in Sox2-Dependent Interactions Drive GFP Transgene Activity to Zebrafish Brain. 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 (F1 of stable lines, except for c-*fos*, transient trans- genics). 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 https://enhancer.lbl.gov) (Visel et al., 2009). See also Figure S6 and Table S6.

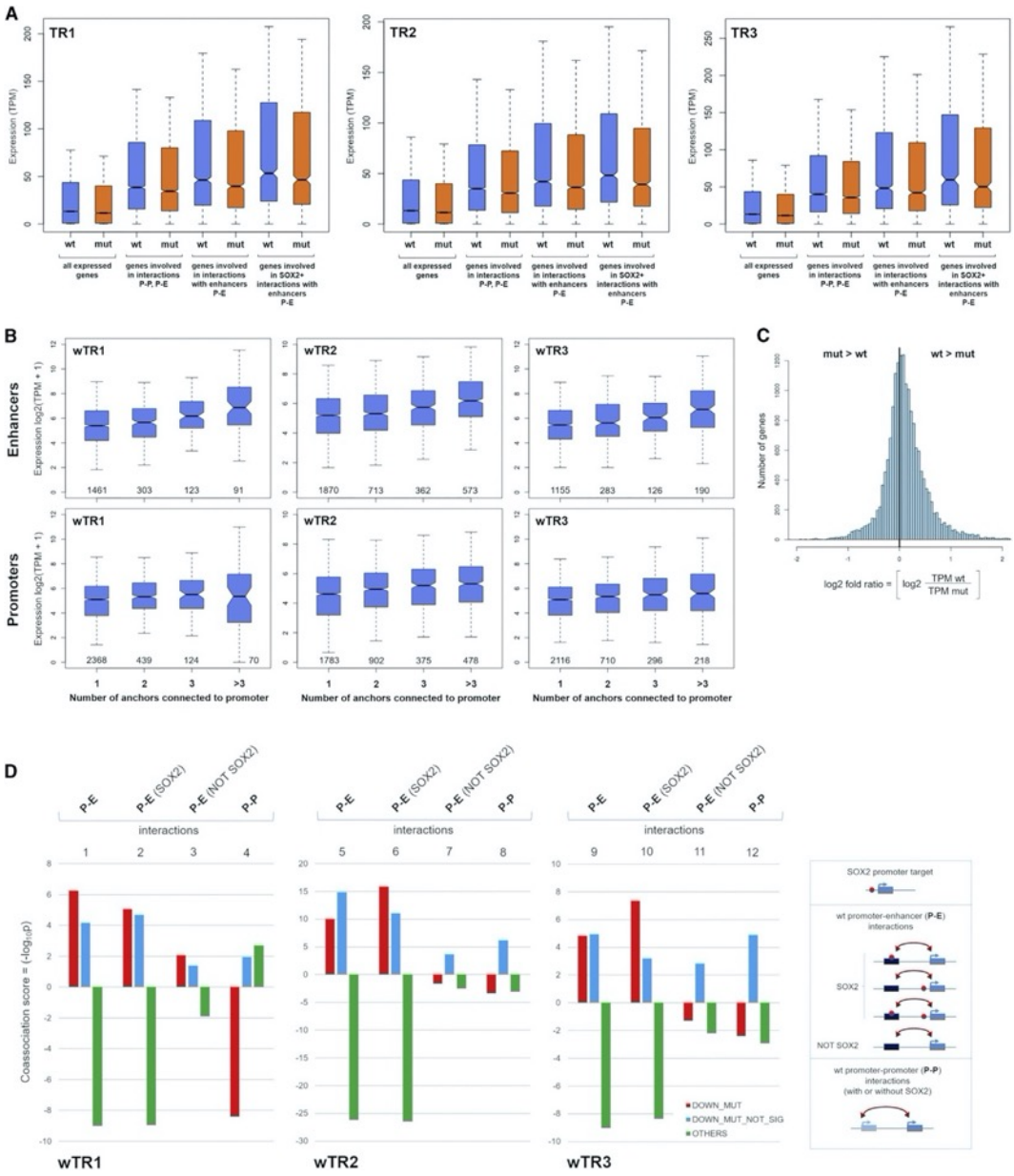


Figure 5. Reduced Gene Expression in Sox2 MUT NSCs Correlates with Loss of Long-Range Interactions. (A) Distribution of expression values (TPM) of genes with TPM >0. Blue, WT NSCs; orange, MUT NSCs. From left to right: all genes; genes whose promoter is a node (P-P, P-E interactions); genes whose

promoter is connected to an enhancer (P-E interactions); genes with SOX2-positive P-E interactions. (B) Distribution of expression values (y axis) of genes according to the number and type of element (enhancer or promoter anchors) interacting with the gene promoter (x axis) in wTR1, wTR2, and wTR3. Top: interactions with enhancers. Bottom: interactions only with promoters. The number of genes involved is shown in each diagram inside the box along the x axis. (C) Distribution of the fold ratio values for all genes with TPM >0 defined as log2 (TPM_WT/TPM_MUT). It confirms results shown in (B): the fold ratio is shifted toward positive values (i.e., a majority of genes have expression in WT higher than in MUT NSCs).

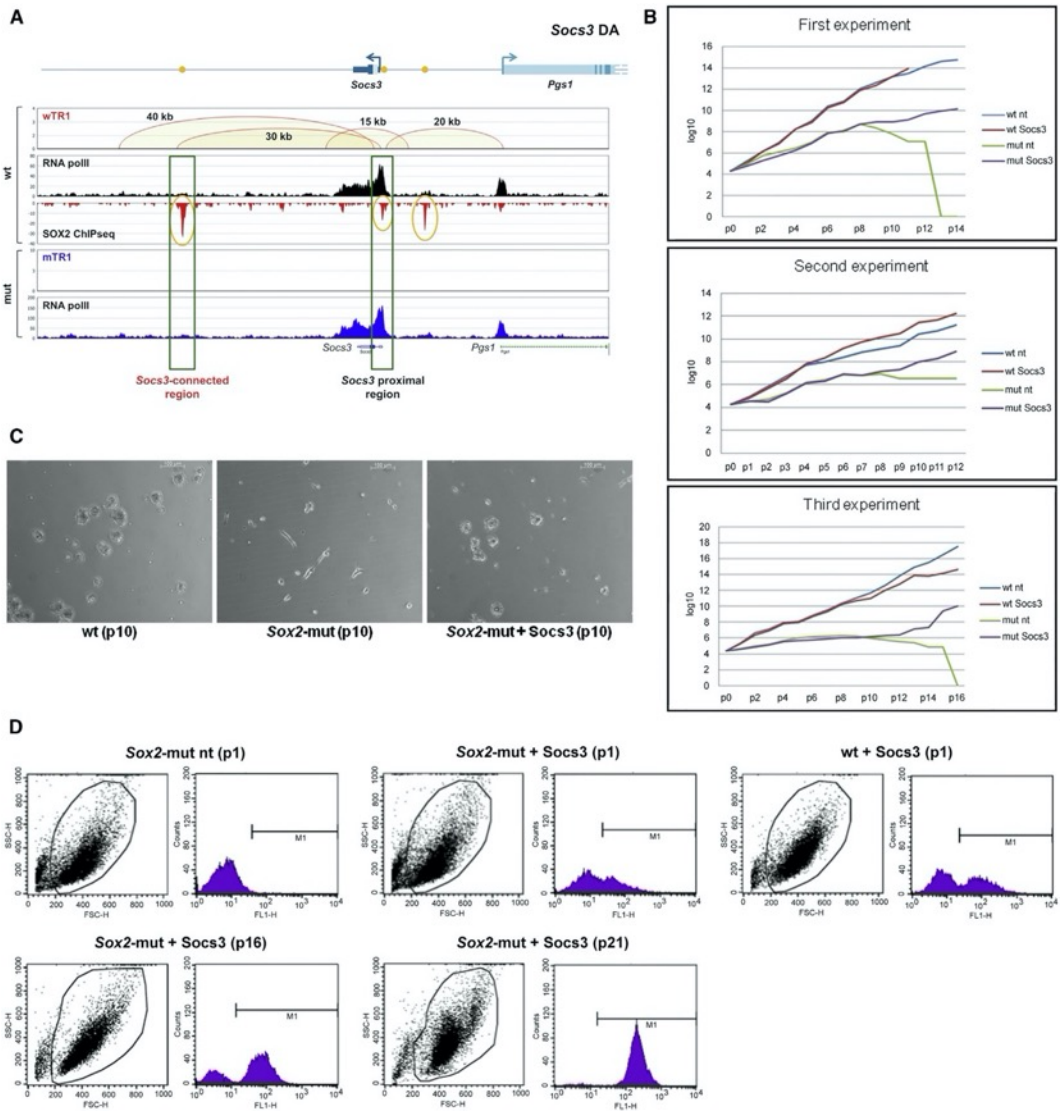


Figure 6. SOCS3 Re-expression in MUT NSCs Prevents NSC Exhaustion and Restores Self-Renewal. (A) Top: *Socs3* gene. ChIA-PET interactions, SOX2 peaks, and ChIA-PET reads in WT NSCs. Bottom: loss of interactions in *Sox2*-MUT NSCs. (B) Growth curves of MUT NSCs, not transduced (MUT) or transduced (MUT *Socs3*) with a *Socs3*-GFP-expressing lentivirus, and of WT controls (WT or WT *Socs3*). (C) Images (phase-contrast) of MUT or *Socs3*-

transduced MUT NSCs 3 days after passage 12; neurospheres develop only from *Socs3*-transduced NSCs. For comparison, WT NSCs. (D) FACS analysis (GFP) of MUT, WT, and MUT *Socs3* cells at the indicated passage number. With passaging, the fraction of GFP+ NSCs progressively increases in MUT *Socs3* NSCs, eventually reaching 100%.

Supplemental information

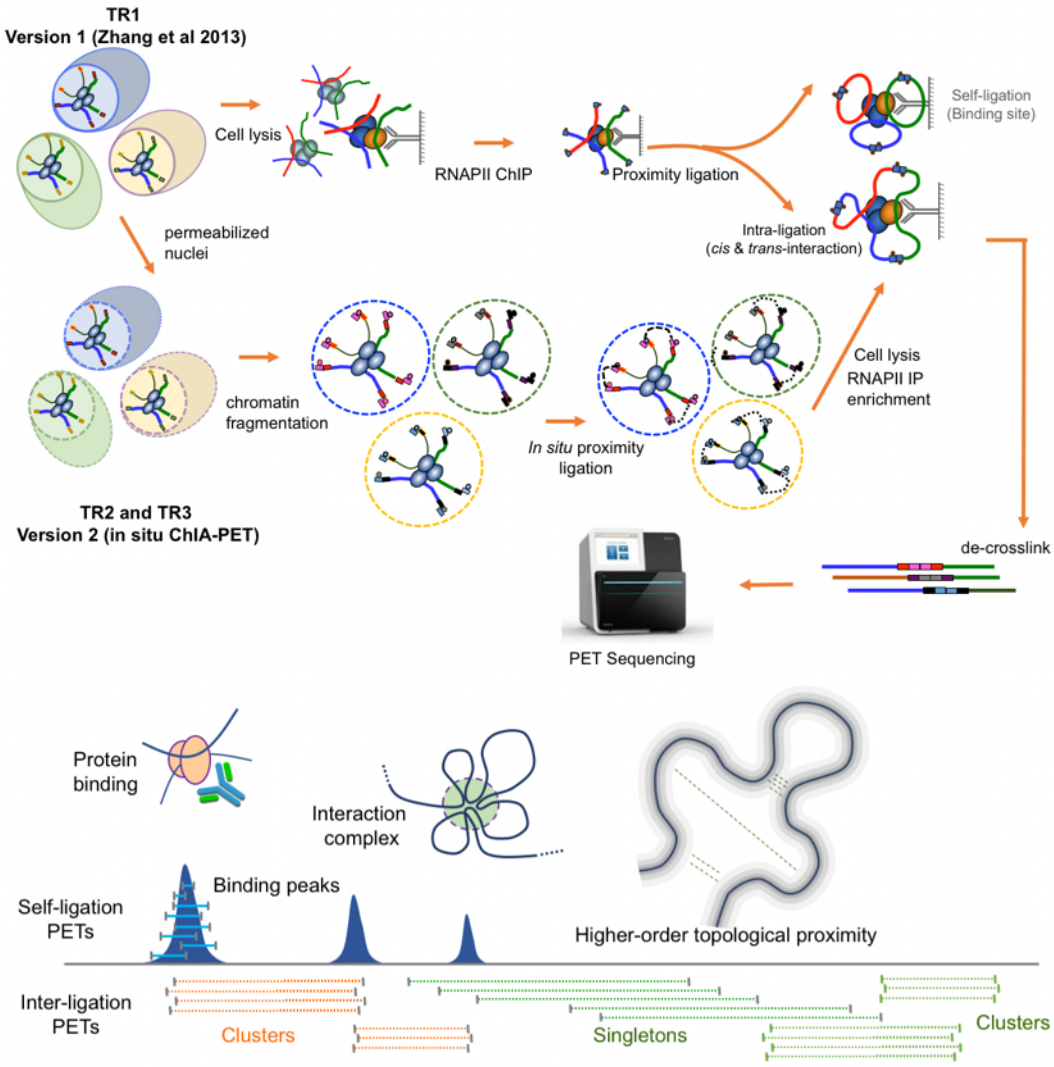


Figure S1, related to Figure 1. Overview of Chia-PET and in situ ChIA-PET workflow ChIA-PET was as previously described by Zhang et al., 2013; after crosslinking, all additional procedures were performed after cell lysis. For in situ ChIA-PET, the nuclei of the crosslinked cells were permeabilized when still intact, allowing to process chromatin for proximity ligation prior to cell lysis (see

Methods). In the bottom panel, different types of inter-ligation PETs are shown. In addition to revealing RNAPolII-mediated interactions, in situ ChIA-PET may reveal, when considering the entire set of intra-molecular ligated PETs (bottom panel) also general chromatin contacts that represent broad spatial topological associated domains, similar to the global contact maps generated by Hi-C based approaches (data not shown).

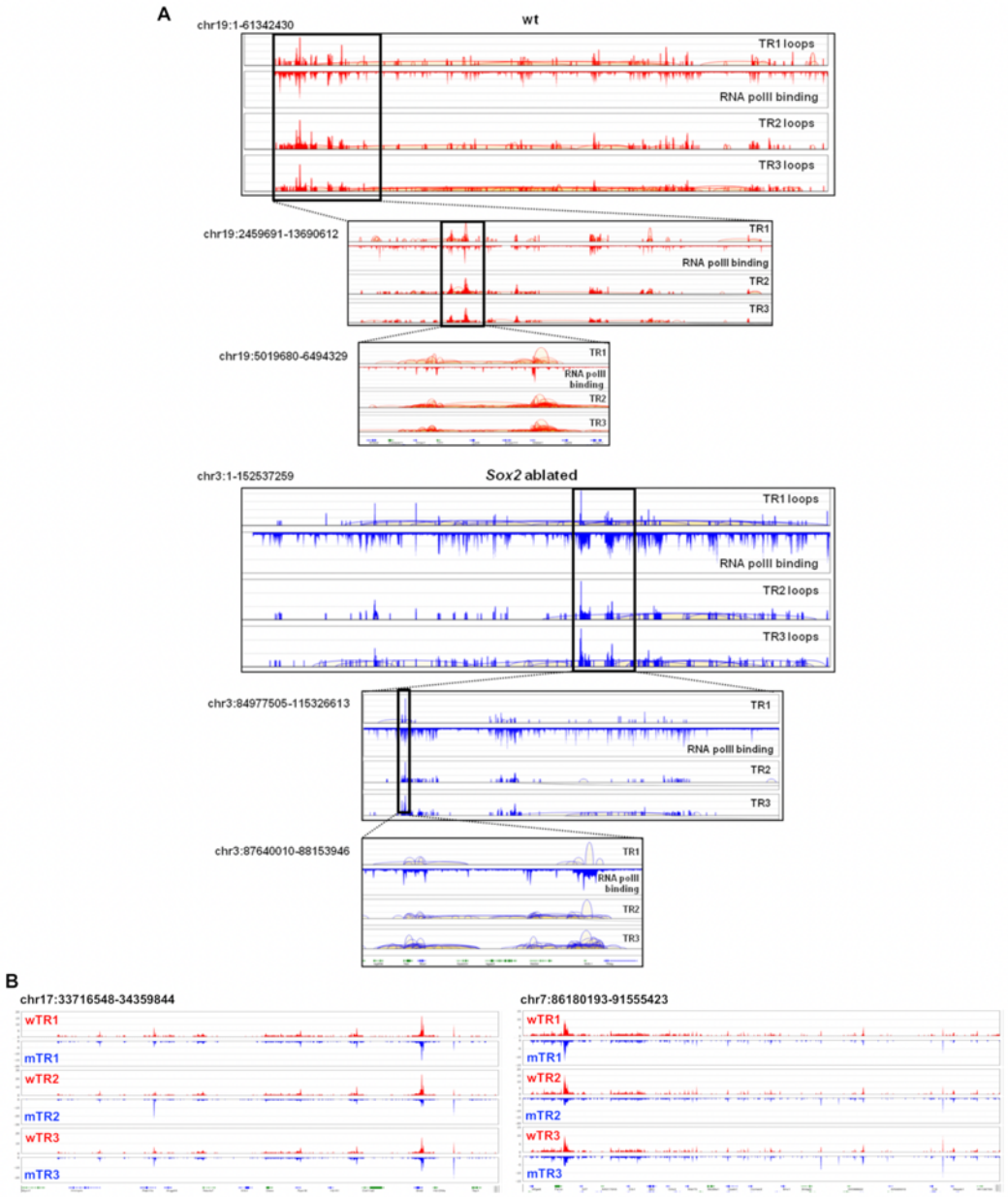


Figure S2, related to Figure 1. Examples of reproducibility in different biological replicates (A) Examples of interactions (loops) profiles in TR1, TR3, TR3, in wild-type cells (red, top) and mutant cells (blue, bottom), in the indicated chromosome regions. The boxed regions are zoomed-in to allow observation at

higher resolution. (B) Examples of RNApolII binding profiles in wild type and mutant TR1, TR2 and TR3 samples.

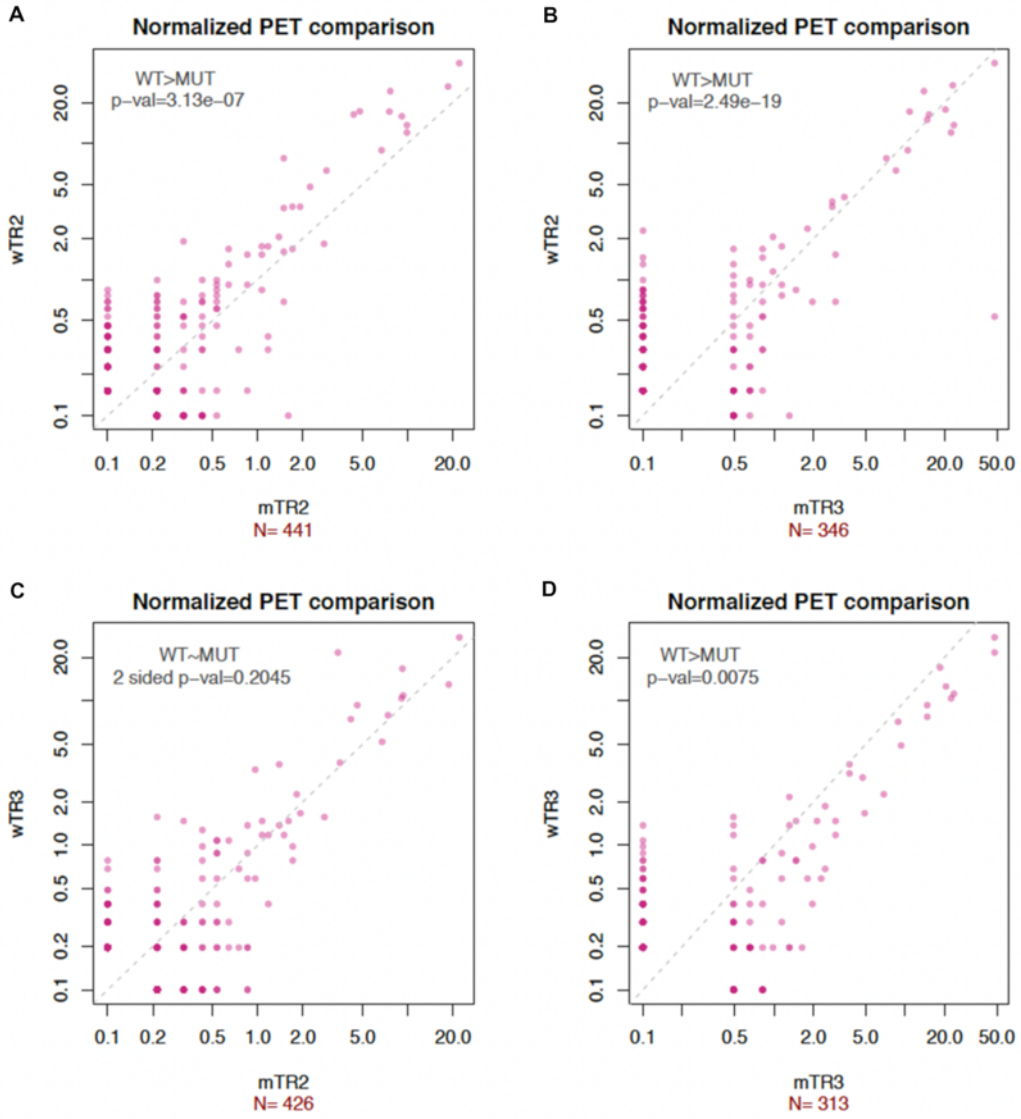


Figure S3, related to Figure 1. Comparison of PET counts between wild-type and Sox2-mutant libraries (TR2 and TR3) across 9 view points. Comparisons

are:A) between wtTR2 and mtTR2; B) between wtTR2 and mtTR3; C) between wtTR3 and mtTR2; D) between wtTR3 and mtTR3. All cis-interactions along the chromosome from compared libraries that originated from all view points listed in the Table below (+/- 5kb) were collected. The interaction anchors were then evaluated to assess whether they overlap. As a result, a new set of non-overlapping interactions was then collected (see the Table for number of interactions or dots on each panel). When an interaction was absent in one of the two libraries, its PET count was assigned as 0. The compared data set is compiled from the output of a significance-calling algorithm, using down-sampled data (Table 1) to minimize sequencing depth bias. The PET counts for each interaction (normalized by dividing by the total number of intrachromosomal PETs -Table 1, line 6- , then multiplying by one million) are plotted as dots, with WT PET count on the Y axis, and MUT PET count on the X axis; the dashed line represents equal interaction frequency, so points above this line indicate reduction of interaction frequency in MUT. Using log-log scale plotting, the zero PET counts are all augmented by 0.1. Because the dots are transparent, the color intensity of the dots indicates multiple dots having the same WT/MUT PET counts. We statistically evaluated differences between wild type and mutant; differences for MUT < WT are shown as p-values on the plot area (2-sided Wilcoxon rank sum test p-values). mTR3 was highly significantly reduced relative to wTR2 and wTR3; mTR2 was less significantly reduced. The comparison of WT and MUT replicates between themselves yielded non-significant p-values (0.029 and 0.09039, respectively). For this analysis, the threshold for interaction calling was set to PET count = 2, FDR < 0,05 across all 9 loci, yielding a number of interactions adequate for statistical analysis. N: the total number of interactions (dots on the plots) evaluated for each comparison. The numbers of PET counts are listed in Table S3.

For each viewpoint, the genomic coordinates are given. Number of interactions (dots) are shown on the right for each viewpoint, and each panel.

	Viewpoint	CHR	Mid_Coordinate	panel A	panel B	panel C	panel D
1	c-Fos-enh5	chr12	86,797,329	43	36	39	32
2	c-Fos-prom	chr12	86,814,992	63	53	70	58
3	c-Fos_enh3	chr12	86,829,043	54	47	38	29
4	Sox3_prom	chrX	58,145,388	10	7	11	8
5	Olig1_prom	chr16	91,270,745	66	54	74	53
6	Olig2_prom	chr16	91,224,410	102	80	93	70
7	Kat2b_prom	chr17	53,707,872	24	20	20	13
8	Tcf4_prom	chr18	69,505,556	29	10	34	17
9	Stox2_prom	chr8	48,372,611	50	39	47	33
Total				**441**	**346**	**426**	**313**

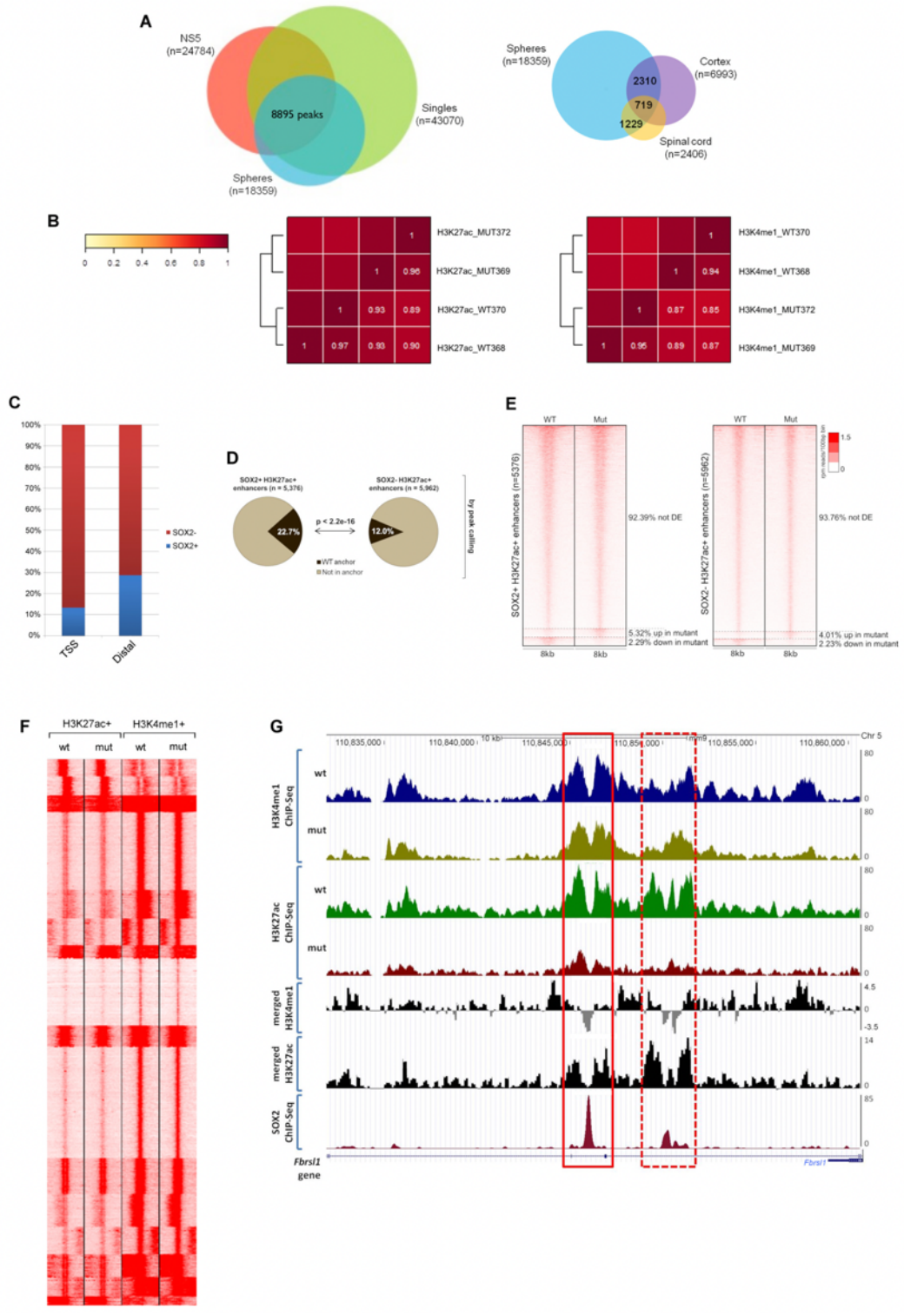
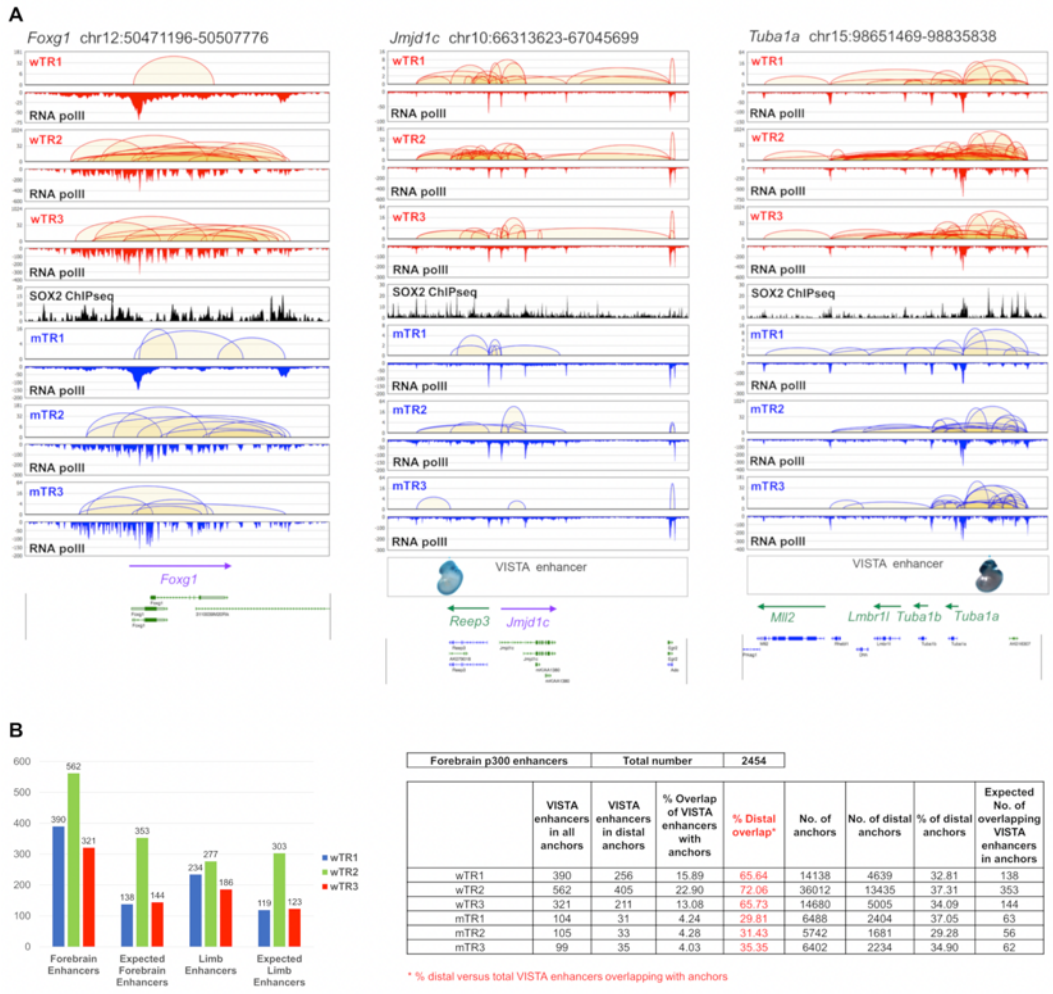

A
NS5
(n=24784)
8895 peaks
Singles
(n=43070)
Spheres
(n=18359)
Spheres
(n=18359)
2310
719
1229
Cortex
(n=6993)
Spinal cord
(n=2406)
B
0
0.2
0.4
0.6
0.8
1
H3K27ac_MUT372
H3K27ac_MUT369
H3K27ac_WT370
H3K27ac_WT368
H3K4me1_WT370
H3K4me1_WT368
H3K4me1_MUT372
H3K4me1_MUT369
C
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
SOX2-
SOX2+
TSS
Distal
D
SOX2+ H3K27ac+
enhancers (n = 5,376)
SOX2- H3K27ac+
enhancers (n = 5,962)
22.7%
p < 2.2e-16
12.9%
WT anchor
Not in anchor
Key peak switch
E
WT
Mut
SOX2+ H3K27ac+ enhancers (n=5376)
92.39% not DE
5.32% up in mutant
2.29% down in mutant
SOX2- H3K27ac+ enhancers (n=5962)
93.76% not DE
4.01% up in mutant
2.23% down in mutant
8kb
F
H3K27ac+
H3K4me1+
wt
mut
wt
mut
G
Chr 5
H3K4me1 ChIP-Seq
H3K27ac ChIP-Seq
merged H3K4me1
merged H3K27ac
SOX2 ChIP-Seq
wt
mut
Fbrsl1
gene

Figure S4, related to Figure 2. Long-range interaction anchors are enriched in SOX2 binding and in chromatin epigenetic marks of active/poised enhancers. (A) SOX2 ChIPseq was performed on both intact ("spheres") and dissociated ("singles") neurospheres (see Methods). Ca. 87% of peaks identified in spheres are also present in singles; ca. 50% of peaks in spheres overlap with peaks in the ES cell-derived NS-5 neural stem cell line (Mateo et al., 2015). Ca. 33% and 51% of peaks detected in neural progenitors of cortical and spinal cord origin, and 68% of their intersection (Hagey et al., 2016), are also detected in spheres. (B) Pearson correlations between the samples for each antibody used in duplicate samples. (C) Fraction of SOX2+ and SOX2- regions within epigenetically marked regions (i.e. H3K27ac+ and/or H3K4me1+) (ChromHMM). (D) Fraction of Sox2+ vs. Sox2-EM-positive regions which overlap with anchors. Distal (>1000nt from TSS) Ac+ regions, peak-calling. (E) *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. (F) *Sox2* loss does not result in H3K27ac- or H3K4me1-enrichment changes at regions bound by SOX2 in wt cells. (G) 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.



Forebrain p300 enhancers	Total number	2454

	VISTA enhancers in all anchors	VISTA enhancers in distal anchors	% Overlap of VISTA enhancers with anchors	% Distal overlap*	No. of anchors	No. of distal anchors	% of distal anchors	Expected No. of overlapping VISTA enhancers in anchors
wTR1	390	256	15.89	65.64	14138	4639	32.81	138
wTR2	562	405	22.90	72.06	36012	13435	37.31	353
wTR3	321	211	13.08	65.73	14680	5005	34.09	144
mTR1	104	31	4.24	29.81	6488	2404	37.05	63
mTR2	105	33	4.28	31.43	5742	1681	29.28	56
mTR3	99	35	4.03	35.35	6402	2234	34.90	62

* % distal versus total VISTA enhancers overlapping with anchors

Figure S5, related to Fig. 3. Enrichment of VISTA enhancers within interaction anchors. (A) Examples of replicates of ChIA-PET analysis of regions surrounding genes encoding proteins important for inherited brain developmental defects (See also Table S7). Images (from https://enhancer.lbl.gov/) of embryos carrying lacZ transgenes driven by VISTA enhancers located within the considered genomic regions are shown. (B) Left: Numbers of VISTA enhancers from forebrain or limb (forebrain enhancers, limb enhancers; Visel et al., 2009) in interaction anchors, as compared to the numbers expected based on a random distribution.

Limb enhancers are presented as a term of comparison between regulatory elements active in forebrain neural tissue versus non-neural tissue. Right: Overlap of forebrain VISTA enhancers with anchor types.

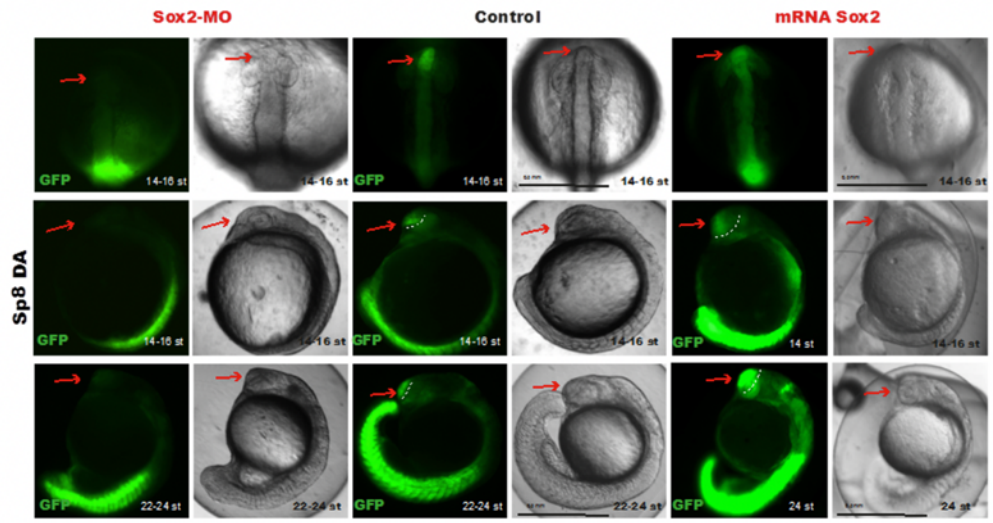


Figure S6, related to Figures 3 and 4. Regulation of the Sp8 enhancer-dependent GFP-reporter construct by decreased or increased Sox2 expression. Transgenic embryos injected with anti-Sox2 morpholino (*Sox2*-MO, left), control morpholino (center), Sox2mRNA and analyzed at two different stages. Reduced GFP signal is seen in *Sox2*-MO, but not ctrl-MO embryos, in forebrain (red arrow) but not in more posterior regions (internal control). Sox2 mRNA extends the forebrain expression of the transgene (in 19/54 embryos). Top, dorsal view; middle and bottom, lateral view.

Tables S2, S6, S7

(Excel Supplemental Tables are provided separately)

Table S2, related to Figure 1
Average reproducibility score (SCC, Stratum-adjusted Correlation Coefficient*) between TR2 & TR3 over all 20 chromosomes

WT: 0.935 (wTR2 vs. wTR3). MUT: 0.839 (mutTR2 vs. mutTR3).

SCC break down by chromosome between replicas:

Chrom	WT	MUT
chr1	0.938	0.846
chr2	0.918	0.813
chr3	0.935	0.819
chr4	0.946	0.859
chr5	0.920	0.826
chr6	0.920	0.815
chr7	0.944	0.857
chr8	0.936	0.854
chr9	0.930	0.827
chr10	0.925	0.813
chr11	0.924	0.823
chr12	0.946	0.866
chr13	0.964	0.877
chr14	0.951	0.867
chr15	0.933	0.849
chr16	0.941	0.850
chr17	0.939	0.862
chr18	0.944	0.855
chr19	0.942	0.855
chrX	0.905	0.755

Method:

To compute the SCC, ChiA-PET loops for each library were aggregated into 10-kb bin matrix.

SCC score is computed with HiCrep tool (hicrep library in R; Yang et al. (2017) Genome Research 27:1939). For each chromosome, the smoothing parameter used as recommended (h=5) and maximum distance 1 Mb.

*reference: https://genome.cshlp.org/content/early/2017/08/30/gr.220640.117

Pearson's correlation coefficient (PCC) is computed for loops within the same genomic distance. From all distances considered, the PCCs are then weight averaged.

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	•	•		•	•

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		
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, Am J Med Genet C Semin Med Genet 166C, 124

(2) Tan et al., 2014, Am J Med Genet A 164°, 975-992

(3) Lee and Young, 2013, Cell 152, 1237-1251

(4) Saez et al., 2016, Genetics in Medicine 18, 378-385

(5) Kim et al., 2012, Am J Hum Genet 91, 56-72

(6) Peters et al., 2008, Genes Dev 22, 3089-3114

(7) Williamson and Fitzpatrick, 2014, Eur J Med Genet 57, 369-380

Table S6, related to Figure 3, Fig. S5 and Table S5. 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 analyzed for interactions of the respective mouse gene promoter with distal enhancers in wTR1, or presence of SOX2. For replicates (wTR2 and wTR3) of connectivity analyses for some of these genes see Fig. 3 and Fig. S5. p-values are given, when significant, for the association between the above characteristics and type of disease.

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 (10)	in our NSCs	in other ChIP-seq (10)				%	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	-
c-Fos	DA852 (e1)	40,000	✓	-	✓	✓	✓	✓	+	22.2%	12/54	54/54 (9)	+++
	DA854 (e2)	20,000	✓	✓	✓	-	✓	✓	+	9.6%	7/73	73/73 (9)	+

(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;
(9): all transgenics showed some expression in the trunk;
(10): Lodato *et al.*, 2013; Engelen *et al.*, 2011.

Chapter 5

Submitted to ***Stem Cell, available on biorxv***
doi: https://doi.org/10.1101/2020.03.17.995621

Sox2 controls neural stem cell self-renewal through a Fos-centered gene regulatory network

Running head:
A Sox2-Fos axis controlling neural stem cells

Miriam Pagin[1], Simone Giubbolini[1], Cristiana Barone[1], Gaia Sambruni[1], Yanfen Zhu[2], Sergio Ottolenghi[1], Chia-Lin Wei[2], Silvia K. Nicolis[1]*

[1]Department of Biotechnology and Biosciences, University of Milano-Bicocca, piazza della Scienza 2, 20126 Milano, Italy
[2]The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA
Authors' contributions:
Miriam Pagin: Conception and design, Collection and assembly of data, data analysis and interpretation, manuscript writing, final manuscript approval
Simone Giubbolini: Collection and assembly of data, data analysis and interpretation, final manuscript approval
Cristiana Barone: Collection and assembly of data, data analysis and interpretation, manuscript writing, final manuscript approval
Gaia Sambruni: Collection and assembly of dada, data analysis and interpretation
Yanfen Zhu: Collection and assembly of data, final manuscript approval

Sergio Ottolenghi: Conception and design, data analysis and interpretation, manuscript writing, final manuscript approval
Chia-Lin Wei: Data analysis and interpretation
Silvia K. Nicolis: Conception and design, data analysis and interpretation, manuscript writing, financial support, overall coordination and supervision

*Correspondence: Silvia K. Nicolis, PhD Department of Biotechnology and Biosciences University of Milano-Bicocca piazza della Scienza 2, 20126 Milano, Italy, AIRG IG- +39 02 6448 3339(office)/3315(lab)
silvia.nicolis@unimib.it

This work was supported by ERANET-ImprovVision (NEURON8-full-815-091) and Associazione Italiana per la Ricerca sul Cancro (AIRC) IG 2014-16016 grants to SKN. MP is the recipient of a fellowship from the Italian Ministery of University and Research (MIUR) through the grant “Dipartimenti di Eccellenza 2017” to the University of Milano-Bicocca, Department of Biotechnology and Biosciences.

Keywords:
Neural stem cells (NSCs), Sox2, Fos, transcription factors, self-renewal, gene expression, lentiviral vector, CRISPR, conditional knockout, cell biology

Abstract

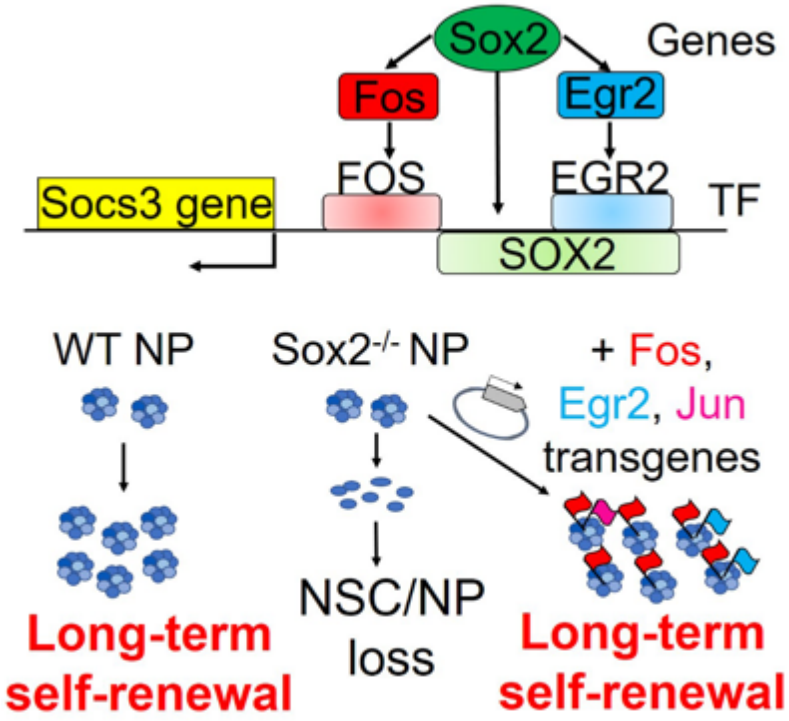
The Sox2 transcription factor is necessary for the long-term self-renewal of neural stem cells (NSC). Its mechanism of action is still poorly defined. To identify molecules regulated by Sox2, and acting in mouse NSC maintenance, we transduced, individually or in combination, into Sox2-deleted NSC, genes whose expression is strongly downregulated following Sox2 loss (Fos, Jun, Egr2). Fos alone rescued long-term proliferation, as shown by in vitro cell growth and clonal analysis. Further, Fos requirement for efficient long-term proliferation was demonstrated by the strong reduction of NSC clones capable of long-term expansion following CRISPR/Cas9-mediated Fos inactivation. Previous work showed that the Suppressor of cytokine signaling 3 (Socs3) gene is strongly downregulated following Sox2 deletion, and its reexpression by lentiviral transduction rescues long-term NSC proliferation. Fos appears to be an upstream regulator of Socs3, possibly together with Jun and Egr2; indeed, Sox2 reexpression in Sox2-deleted NSC progressively activates both Fos and Socs3 expression; in turn, Fos transduction activates Socs3 expression. Based on available SOX2 ChIPseq and ChIA-PET data, as well as results from the literature, we propose a model whereby Sox2 is a direct activator of both Socs3 and Fos, as well as possibly Jun and Egr2; in turn, Fos, Jun and Egr2 may activate Socs3. These results provide the basis for developing a model of a network of interactions, regulating critical effectors of NSC proliferation and long-term maintenance.

Significance statement

Proliferation and maintenance of NSC are essential during normal brain development, and, postnatally, for the maintenance of hippocampal function and memory until advanced age. Little is known about the molecular mechanisms that

maintain the critical aspects of NSC biology (quiescence and proliferation) in postnatal age. Our work provides a methodology, transduction of genes deregulated following Sox2 deletion, that allows to test many candidate genes for their ability to sustain NSC proliferation. In principle, this may have interesting implications for identifying targets for pharmacological manipulations.

Table of Content Image



1. Introduction

The transcription factor Sox2 is critically important in the development of the brain. In humans, Sox2 mutations cause brain abnormalities (hippocampal dysplasia, learning disabilities, epilepsy, motor control problems, eye and vision defects) (Bertolini et al.,2016; Fantes et al., 2003; Pevny and Nicolis; 2010; Sisodiya et al., 2006). Sox2 is expressed in neural stem and progenitor cells; its deletion during mouse embryogenesis causes early postnatal loss of NSC in the hippocampus, and its postnatal loss causes decreased hippocampal neural stem/progenitor cell proliferation (Favaro et al. 2009). In vitro experiments on NSC showed that Sox2 loss causes progressive exhaustion of NSC, in contrast to the long-term proliferation of control wild type cells (Favaro et al 2009); differentiation defects of progenitor cells have also been reported in Sox2-mutant cells by Cavallaro et al., 2008 and Cimadamore et al., 2013. The loss of self-renewal in Sox2-mutant cells grown in vitro is reversible; in fact, re-expression, in Sox2-deleted cells, of Socs3, the gene most downregulated following Sox2 loss, rescued long-term self-renewal of mutant cells (Bertolini et al. 2019). These findings point to the possibility of identifying the network of regulatory processes acting downstream to Sox2 in allowing NSC proliferation and preserving their functional integrity. To this end, we expressed, in Sox2-deleted NSC, some of the most downregulated genes identified by RNAseq comparison of Sox2-deleted and wild type NSC.

We show that Fos, a transcription factor, efficiently rescues long-term NSC proliferation of Sox2-deleted cells, and increases Socs3 levels; also, its mutation in wild type NSC greatly reduces their ability to generate long-term proliferating clones. Mining of our previous NSC genomic analyses of Sox2 ChIPseq and RNApolII ChIA-PET, which detect long-range promoter-enhancer chromatin interactions [8], shows that Socs3, Fos and its partner Jun, and Egr2 (a regulator of

Socs3), are all bound by Sox2, defining a network of regulatory interactions sustaining Sox2-dependent long-term NSC self-renewal.

2. Materials and Methods

2.1 Primary ex-vivo neural stem/progenitor cell cultures

Brain-derived NSC cultures were obtained from dissected telencephalon of wild-type and Sox2-deleted P0 mice, and grown in fresh medium (FM): DMEM-F12 with Glutamax (GIBCO), supplemented with 1/50 vol/vol B27 (Life Technologies), 1% of Penicillin-Streptomycin (Euroclone) supplemented with EGF (10 ng/ml, Tebu-bio) as mitogen, as described in [5,8,9].

2.2 Lentiviral constructs

Fos, Jun and Egr2 cDNAs were derived from Addgene plasmids: Fos: "pcDNA3FLAGFos WT", from John Blenis (#8966); Jun: "FlagJunWTMyc", from Axel Behrens (#47443); Egr2:"mEgr2/LZRS", from David Wiest (#27784). cDNAs were amplified by PCR and cut with BamHI, and cloned into a unique BamHI site upstream to the IRES-dNGFR cassette of the pHR SIN BX IR/EMW [10](a gift from A. Ronchi, Milano).

The lentiviral vector for CRISPR-Cas9 experiments was lentiCRISPRV2puro (Addgene #98290, RRID: Addgene_104990). The guide RNA (sgRNA) for Fos mutagenesis was designed and cloned according to:

https://media.addgene.org/data/plasmids/52/52961/52961attachment_B3xTwla0bkYD.pdf and is:
5'-CCGCTGCAGTAGCGCCTCCC-3'
As a control, we designed also a scramble non-targeting guide RNA (Mock):
5'- ATGTTGCAGTTCGGCTCGAT-3'

Lentiviral vectors were produced by calcium phosphate transfection into the packaging human embryonic kidney cell line 293T, of the VSV-G plasmid (encoding ENV), CMV R8.74 (packaging) and pRSV-REV (encoding reverse transcriptase)[8,10]. Briefly, after transfection, following replacement with DMEM high glucose (Euroclone), containing 10% Fetal Bovine Serum (Sigma), 1% penicillin-streptomycin (Euroclone), 1% of L-Glutamine (Euroclone), the cell supernatants were collected 24-48 hours after transfection. Lentiviral vectors were titrated on HEK-293T cells by measuring the percentage of eGFP (for the Sox2-transducing vector) or of dNGFR-positive cells (for the Fos, Jun, Egr2-transducing vectors) by Flow Cytometry [10].

2.3 NSC transduction with lentiviral constructs encoding Fos, Jun, Egr2, or Empty Vector.

NSC transduction [8]: wild-type and Sox2-deleted neurospheres were grown for 2-3 passages (3-4 days each) in FM supplemented with bFGF (10ng/ml, Tebu-bio) and EGF (10 ng/ml, Tebu-bio), and for one more passage in FM supplemented with EGF only. For passaging, neurospheres were first incubated in 0,25% trypsin (GIBCO) for 5 minutes at 37°C and, subsequently, ovomucoid (Leibovitz's L15 medium (GIBCO) containing trypsin inhibitor (Sigma), BSA (Sigma) and 40 μg/ml DNaseI (Sigma) for 5 minutes at 37°C). Spheres were then carefully dissociated

mechanically by gently pipetting up and down, centrifuged at 1200 rpm for 4 minutes and cells were resuspended in 1ml of FM, counted and replated at a density of 25,000 cells/1ml/well (9 wells for each cell type) in 24-well plates, in FM. After 4 hours cells (Sox2-deleted or wild-type) were transduced with appropriate lentiviral vectors (single or in combination), at a multiplicity of infection (MOI) 5.5 for each vector. Cells were incubated overnight at 37°C, then 1ml per well of FM with EGF was added both to transduced and non-transduced (control) cells. After 4 days, cells were dissociated to single cells as above, counted, and seeded at a density of 20,000 cells/well in FM with EGF. In addition, 500,000 cells for each sample (from pooled wells) were fixed using PFA 4% and stained as in [10] with an anti-human CD271 (dNGFR) antibody, conjugated with Phycoerythrin (PE) (BioLegend, Cat. No. 345106, RRID: AB_2152647) (dilution 1:200) and analyzed by CytoFLEX (Beckman-Coulter) to determine the percentage of infected cells: 10,000 events were analyzed for each sample. Cells were dissociated, counted and seeded at the same density every 3-4 days, to obtain a cumulative growth curve (Fig 2A, 2B).

2.4 Identification of Fos, Jun, Egr2 lentivirus integrated in clones of lentivirally transduced NSC

Cell populations of NSC transduced with lentiviruses encoding Fos, Jun, Egr2, obtained after 6 or more passages from the transduction, were plated at limiting dilution into 96-well plates (down to a dilution of 7.5 cells/well, which gave either one or no clone in every well). Individual clones (neurospheres) were analyzed by PCR following a short expansion (2-3 passages), DNA was extracted and analyzed by PCR using primers designed to detect the virally transduced, but not the

endogenous, Fos, Jun and Egr2 (forward primer is in the viral SFFV promoter driving the cDNA expression; reverse primer is in the cDNA).

Forward primer SFFV: 5'-CTCACAACCCCTCACTCG-3'

Fos Reverse primer: 5'- AGGTCATTGGGGATCTTGCA-3' (and 5'-GGCTGGGGAATGGTAGTAGG-3')

Jun Reverse primer: 5'- GGTTCAAGGTCATGCTCTGT-3'

Egr2 Reverse primer: 5'- TGCCCGCACTCACAATATTG-3'

2. 5 Transduction of Sox2-deleted NSC with Sox2-encoding lentiviral vector

Sox2-deleted NSCs were dissociated to single cells and seeded at a density of $5x10^5$ cells/T25cm^2 flask/5ml, in FM. After 4 hours NSCs were transduced with a GFP-Sox2- expressing lentivirus [5], at a multiplicity of infection (MOI) 8 and 12 (Suppl. Fig. 3). Cells were incubated overnight at 37°C. Virus was removed by medium change at 24 h: cells were centrifuged at 1000 rpm for 4 minutes and resuspended in FM with EGF. Non-transduced controls were treated equivalently (without virus). At every passage, every 3-4 days, cells were dissociated to single cells as above, counted, and replated in FM with EGF at a density of 20,000 cells/well/1ml, to generate a cumulative growth curve (Suppl. Fig 3). At successive passages, aliquots of Sox2-transduced cells (50,000-500,000 cells, from pooled wells) were fixed using 4% paraformaldehyde (PFA) and analyzed for GFP fluorescence by CytoFLEX (Beckman-Coulter) to determine the percentage of infected cells: 10,000 events were analyzed for each sample.

2.6 Measure of Fos and Socs3 expression levels in Sox2-deleted cells transduced with Sox2- or Fos-encoding lentiviral vectors

Sox2-deleted NSCs were dissociated to single cells and seeded at a density of $5x10^5$ cells/T25cm^2 flask/5ml, in FM. After 4 hours NSCs were transduced with lentiviruses expressing Sox2 or Fos at a multiplicity of infection (MOI) 10. Cells were incubated overnight at 37°C. Virus was removed by medium change at 24 h: cells were centrifuged at 1000 rpm for 4 minutes and resuspended in FM. Non-transduced controls were treated equivalently (without virus).

At every passage, every 3-4 days, cells were dissociated to single cells as above, counted, and replated in FM with EGF at a density of 800,000 cells/T75cm^2/12ml.; 50,000-500,000 cells were analyzed by FACS for GFP or dNGFR expression, as described above, to determine the percentage of infected cells. Additionally, aliquots of cells (100,000-500,000 cells) were lysed in TRIzol (Life Technologies) for total RNA extraction. RNA was purified using Direct-zol RNA Miniprep (Zymo Research), treated with DNase (Zymo Research) for 30 minutes at 37° C and 400ng of RNA were retrotranscribed (High Capacity cDNA Reverse Transcription Kit, Applied Biosystem). About 5µl of a 1:25 dilution (adjusted following normalization by *Hprt* expression) were used for the real time PCR. Negative control reactions (without Reverse Transcriptase) gave no PCR amplification. Real time analysis was performed using ABI Prism 7500 (Applied Biosystems). Samples from each experiment were analyzed in duplicate. Specific PCR product accumulation was monitored by using SsoAdvanced™ Universal SYBR® Green Supermix (Bio-Rad) fluorescent dye in a 12-µl reaction volume. Dissociation curves confirmed the homogeneity of PCR products.

Primers for the quantification of Fos and Socs3 mRNAs are:

Fos Forward primer: 5'-CTGTCCGTCTCTAGTGCCAA-3'

Fos Reverse primer: 5'-TGCTCTACTTTGCCCCTTCT-3'
Socs3 Forward primer: 5'-ACCTTTGACAAGCGGACTCT-3'
Socs3 Reverse primer: 5'-AGGTGCCTGCTCTTGATCAA-3'
HPRT Forward primer: 5'-TCCTCCTCAGACCGGTTT-3'
HPRT Reverse primer: 5'-CCTGGTTCATTCATCGCTAATC-3'
All qRT-PCRs for Fos and Socs3 were performed in parallel with HPRT qRT-PCR, for normalization.

2.7 CRISPR-Cas9 assays

Wild-type neurospheres were dissociated into single cells, as described above, and seeded at a density of $5x10^5$ cells/T25cm^2 flask/5ml, in FM with EGF. Cells were transduced, after 4 hours from plating, with a previously defined amount of lentiCRISPRV2puro lentivirus expressing the anti-Fos guide RNA or the scramble non-targeting guide RNA. Medium was changed after 3 days and cells were selected in FM with 5μg/ml Puromycin (Sigma-Aldrich, cod. P8833) for 3 days. To remove puromycin, cells were then centrifuged at 1000 rpm for 4 minutes and resuspended in FM without puromycin and grown for 15 days to allow recovery from the stress of transduction and selection. Neurospheres were then dissociated to single cells; similar numbers of cells treated with scramble guide RNA (Mock) or with anti-Fos guide RNA (Mut) were plated at clonal dilution in individual wells of a 96-well plate in FM with EGF. With 15 cells/well, essentially all wells showed, after 10 days, at least one or more neurospheres; at a nominal limiting dilution of 7.5 cells/well, single clones (neurospheres) appeared only in a proportion of the wells (the "primary clones"). These clones were then further used to determine the proportion of long-term growing clones (see below). Two experiments were performed: 258 and 207 primary clones were obtained from cells treated with

scramble guide RNA; 73 and 198 primary clones were detected from cells treated with anti-Fos guide RNA.

Individual primary clones from wells showing a single neurosphere were picked up, disaggregated and replated ("replated clones", Fig. 4C, 4D) in a single well of a 48-well plate, for further expansion in a 24-well plate and, finally, in a 6-well plate and in flasks. At every passage, the number of clones which did not proliferate was annotated. Clones that continued to grow in flasks or 6-well plates were considered to be "long-term growing clones".

To verify the efficiency of Fos mutagenesis, at the beginning of the experiment (see above and Fig. 4B) an aliquot of the puromycin-resistant cells was collected, DNA was extracted and amplified by PCR with primers surrounding the site of hybridization of the sgRNA. The amplified DNA was cloned in pGEM®-T Easy (Promega, A1360) by transformation of TOP10 E.coli, and inserts from individual colonies were sequenced. The sequences from CRISPR-Cas9-treated cells were compared to wild-type sequences, by using BLAST NCBI.

PCR primers:

Fos Forward: 5'-TCACAGCGCTTCTATAAAGGC-3' and 5'-CTACTACTCCAACCGCGACT-3'

Fos Reverse: 5'-CTGCGAGTCACACCCCAG-3' and 5'-CGCCAGTCTCCCTCCAGA-3'

2.8 Immunocytochemistry

Transduced Sox2-deleted, or wild type control NSCs, were dissociated to single cells and seeded on Matrigel™-coated glass coverslips. After 4h cells were fixed (20 minutes) with 4% PFA in phosphate-buffered saline (PBS; pH 7.4) and rinsed 3 times with PBS. Coverslips were then incubated for 90 minutes at 37°C in PBS

containing 10% normal goat serum (NGS), 0.2% Triton X-100. Coverslips were incubated with the primary anti-SOX2 antibody (mouse monoclonal IgG2a, 1:100, R&D Systems), overnight at 4°C. Following thorough washing with PBS, cells were incubated for 45 minutes (room temperature) with secondary rhodamine (TRITC)-conjugated goat anti-mouse IgG2a antibodies (1:1000, AlexaFluor Life Technologies). Coverslips were rinsed 3 times in PBS and mounted on glass slides with Fluoromount (Sigma) with DAPI (4′,6-diamidino-2-phenylindole).

3. Results

3.1 Reactivation of Fos expression in Sox2-deleted neural stem cells restores long-term self-renewal

We identified, by RNAseq, about 900 genes downregulated in NSC grown in vitro from neonatal brain, after Cre-mediated Sox2 deletion during embryogenesis (Bertolini et al., 2019). To identify the network of genes mediating Sox2 function in NSC self-renewal, we reasoned that genes downregulated following Sox2 deletion likely include those required for self-renewal; indeed, we previously showed that the most downregulated gene, Socs3, encoding a signaling protein acting at the cell membrane, rescued long-term self-renewal when expressed via a lentivirus in Sox2-deleted cell (Bertolini et al., 2019). We therefore overexpressed, in Sox2-mutant cells, three among the most downregulated genes (Fig. 1A), Fos, Jun, and Egr2, encoding transcription factors which might possibly be involved in the regulation of Socs3 (or other critical factors).

To test for the functional role of the downregulated genes, we individually cloned their cDNAs into a lentiviral expression vector, and transduced them into Sox2-mutant cells, cultured from P0 mouse brains from which Sox2 had been deleted in

vivo, at embryonic day (E) 11.5, via a nestin-Cre transgene (Favaro et al., 2009; Bertolini et al., 2019) (Fig. 1B). We first tested a combination of vectors expressing Fos and Jun, since FOS and JUN are known to act together, forming the AP1 complex, in transcriptional regulation (Efferl et al., 2003). The vector contains a delta-NGF receptor (dNGFR) marker gene that allows to identify transduced cells by anti-dNGFR immunofluorescence (Fig. 1B). In preliminary experiments, we transduced NSC using single vectors, including each of the genes to be tested, to evaluate the proportion of the cells that were transduced by each type of vector: the transduction efficiency was 23,3% for Fos; for Jun and for Egr2, it was 18,5% and 17,1%, respectively.

We then transduced mutant cells at early stages of their culture, using combinations of different vectors at the same multiplicity of infection (MOI) as that used for single-vector transduction (Fig. 2). After initial passages, non-transduced mutant cells progressively slowed down their proliferation and stopped to grow, the culture eventually becoming exhausted, as expected (Favaro et al., 2009; Bertolini et al., 2019); by contrast, cells transduced with Fos and Jun continued to grow exponentially, with a kinetics matching that of control wild-type cells (Fig. 2A). Taking advantage of the delta-NGF receptor (dNGFR) marker, that is coexpressed with the cDNAs from the lentiviral vectors, we followed the percentage of dNGFR-positive (transduced) cells though cell passaging. While, at the beginning (passage 2 after transduction), dNGFR-positive cells represented about 50% of total cells, their percentage progressively rose to 100% (passage 15) through passaging (Fig. 2C; Suppl. Fig. 1), indicating a selective advantage of transduced cells.

We also verified that cells that had recovered self-renewal were all SOX2-negative, by immunofluorescence (IF) for SOX2 (Suppl. Fig. 2A); this ruled out the possibility that a positive selection of rare, SOX2-positive cells persisting in Sox2-deleted brains could contribute to the recovery of mutant cells.

We also tested a second combination of vectors, expressing Fos and Egr2 (Fig. 2B). In parallel with this vector combination, we also transduced mutant cells with the Fos-expressing vector only (Fig. 2B). The viral upregulation of Fos and Egr2 in mutant cells prevented cell exhaustion, and restored long-term self-renewal (Fig. 2B), similarly to what had been found with Fos+Jun (Fig. 2A). A similar result was obtained also with the Fos vector alone; self-renewal was restored, and the growth kinetics of the cells was similar to that of wild type cells (Fig. 2B). We then measured the percentage of transduced cells by FACS analysis with an anti-dNGFR antibody, as previously done for the Fos+Jun-transduced cells; this was about 20% at the beginning (passage 2), and progressively rose to about 100% through passaging (for both Fos+Egr2 and Fos only), again pointing to a selective advantage of the transduced cells (Fig. 2C; Suppl. Fig. 1).

As a control experiment, also wild type cells were transduced with the Fos-expressing vector; this did not lead to any change in the growth kinetics of the cells, that remained the same as that of non-transduced cells (Suppl. Fig. 2B). As a further control, we also verified that mutant cells, transduced with an "empty" dNGFR lentiviral vector, did not change their growth kinetics in comparison to that of non-transduced cells (Suppl. Fig. 2B).

Overall, these results point to Fos as the major player in the rescue of NSC self-renewal in transduced cells.

Do any of the other co-transduced cDNAs play any additional role in the effect shown? To this end, we plated the transduced cells (Fos + Jun transduction) at limiting dilution, in 96 well plates, to obtain clones. These were then individually expanded, and tested for the presence of the two lentiviral inserts (Fos, Jun). Out of 14 clones tested by PCR analysis, only two contained Jun, but all contained the Fos-expressing vector (Fig. 3A). This indicated that Fos is required for the observed

recovery of self-renewal, as suggested. Importantly, no clones exclusively transduced with Jun (but not with Fos) were detected, implying that Jun, by itself, is not sufficient to rescue Sox2-deleted stem cells. The proportion of doubly transduced clones (Fos+, Jun+) is roughly that expected on the basis of the probability of chance transduction of a Fos-transduced cell also with a Jun vector, based on the previous assessment of the transduction efficiency of individual vectors (see above).

Similarly, in the Fos+Egr2-transduction, out of 14 clones expanded and analyzed, 8 contained the Egr2 vector, but all contained the Fos vector; none contained the Egr2 vector alone (Fig. 3B). Again, this indicates that Egr2 alone cannot rescue NSC long-term proliferation.

3.2 Fos mutation antagonizes NSC growth

Having found that Fos, a putative target of Sox2, is able to replace Sox2 to maintain NSC self-renewal, when expressed in Sox2-deleted cells, we asked if Fos itself is required to maintain NSC, in the presence of wild type Sox2. We thus mutated the endogenous Fos gene, using the CRISPR/Cas9 system. A short guide RNAs (sgRNA) was identified, which efficiently mutagenized Fos (sgRNA-fos) (Fig. 4A; Methods). Wild type NSC were transduced with a Cas9+sgRNA-expressing lentivirus, carrying a puromycin-resistance gene, or a control virus expressing "scramble" sgRNA that is not expected to mutate Fos; transduced cells were then selected by puromycin for 3 days (Fig. 4B). After puromycin withdrawal, the cells were grown for 15 days to allow recovery from the stress of transduction and selection; at this initial stage, control and mutated cells proliferated in bulk culture with comparable kinetics (not shown). In parallel, at the time of puromycin withdrawal, an aliquot of the puromycin-resistant cells was collected, DNA was

extracted and amplified by PCR with primers surrounding the site of hybridization of the sgRNA, and the PCR product cloned in a plasmid, and sequenced (Fig. 4B). 9/10 individual clones were mutated with the anti-Fos sgRNA, whereas no mutated Fos copy was found in scramble-treated cells (Suppl. Table 1).

We then tested the ability of NSC to self-renew generating stem cells capable of giving rise to clonal progeny (Gritti et al., 1996; Cavallaro et al. 2000) (Fig. 4C). In this type of experiment, cells were plated at clonal dilution in individual wells, and the appearance of single neurospheres in a well was evaluated at the microscope. When the growth of the neurosphere had generated a relatively large number of cells (a "primary clone"), this was picked up, disaggregated and replated ("replated clones", Fig. 4C, D) in a single well, for further expansion. Primary clones could be obtained from both scramble-treated controls (mock) and sgRNA-fos treated cells. We then asked whether each of these clones represents the bona fide progeny of a long-term self-renewing stem cell, by testing for its ability to expand through serial passaging, to give long-term growth (Fig. 4C, D, Suppl. Table 2, "long-term proliferating clones"). We performed two experiments (Fig. 4D, Suppl. Table 2): in both experiments, the number of primary clones (from sgRNA-fos treated cells), capable to reform secondary clones demonstrating long-term growth, was greatly, and significantly, diminished in comparison to controls from mock-treated cells (Fig. 4D, Suppl. Table 2, line 2). These data indicate that Fos-mutant NSC progressively deteriorate during culture, failing to self-renew, a phenomenon reminiscent of the previous observations obtained with Sox2-deleted NSC (Bertolini et al., 2019).

3.3 Sox2 upregulation in Sox2-mutant cells restores expression of endogenous Fos and Socs3

These observations made us wonder if re-expression of Sox2 into Sox2-deleted cells, leading to recovery of self-renewal (Favaro et al., 2009) (Suppl. Fig. 3), would also re-increase levels of Fos and Socs3 expression. We thus analyzed by qRT-PCR the expression of Fos, and of Socs3, in mutant cells (three different mutants) transduced with Sox2, which re-acquire the ability to self-renew long term. Transduction of Sox2 initially (until passages 5-6) had little or no effect on the expression of Fos and Socs3 (Fig. 5 A,B); however, at later stages in culture, after 5-8 passages, there was a substantial increase of both Fos and Socs3 mRNAs that lasted at least until late passages (passages 20-22, when the experiment was interrupted)(Fig. 5A,B). In some experiments, we also checked the expression of Socs3 in Sox2-mutant cells transduced with Fos (Fig. 5C) (alone or together with Jun or Egr2); again, an increase of Socs3 expression was observed. A selective advantage of Sox2- transduced cells was again observed by FACS analysis through passaging, until almost 100% of the cells were represented by transduced cells (Suppl. Fig. 3).

4. Discussion

Socs3 overexpression is able to rescue the self-renewal defect of Sox2-mutant NSC (Bertolini et al., 2019). Do any additional genes, upstream to Socs3, or acting in parallel with it, contribute to mediate Sox2 effects on self-renewal? In the present work, we tested some candidate genes which are strongly downregulated in Sox2-mutant cells. We found that Fos alone is able to fully rescue Sox2-mutant cells (Fig. 2). The analysis of the rescued clones, obtained by Fos+Jun or Fos+Egr2

transduction, shows that neither Jun, nor Egr2 play critical roles in the rescue. In fact, we do not find any rescued clones that have been transduced exclusively with one of these genes, and, when these transduced genes are present in clones, they are always accompanied by Fos. Further, the proportion of Fos-transduced clones, that are also transduced by Jun (Fos + Jun), is not higher than that expected on the basis of chance cotransduction of a Fos-transduced cells also by a Jun vector, indicating that Jun does not significantly cooperate with Fos in the rescue (Fig. 3). However, the proportion of Fos/Egr2 cotransduced rescued clones (8 out of 14 total rescued clones) is higher than expected based on the assumption that Fos transduced cells should be transduced with Egr2, by chance, with a probability of 0.17, the frequency of Egr2 transduction; the result obtained might suggest that Fos-transduced cells, also transduced with Egr2, have some proliferative advantage over Fos-only-transduced NSC, indicating that Egr2 could cooperate with Fos. Finally, the progressive decrease of clonogenicity that occurs in Sox2-positive NSC mutated in Fos by CRISPR/Cas9 (Fig. 4; Suppl. Tables 1, 2) confirms that Fos is a critical component of the genetic program controlling in vitro self-renewal of NSC.

Fos, a component, with Jun or related factors, of the AP-1 protein complex, is well known to control cell life and death by regulating the expression and function of cell cycle regulators such as Cyclin D1, p53, p21cip1/waf1, p19ARF and p16 (Shaulian, Karin, 2001). In NSC grown in vitro, factors such as FGF2 and bEGF control cell proliferation through a cascade involving ERK signaling, Fos, and cyclins (Adepoju et al., 2014) . The strong decrease of Fos in Sox2-deleted cells (Fig. 1A), although less profound than that observed with Socs3, is in line with the known role of Fos in cell proliferation control. Thus, three lines of evidence support the conclusion that Fos is an important mediator of Sox2 function in the maintenance of NSC self-renewal: i) the rescue of self-renewal of Sox2-deleted cells by Fos overexpression (which, by itself, has no effect in wild type NSC); ii)

the decreased long-term self-renewal in Fos-mutated NSC; iii) the known proliferation-related activity of Fos in several different cell types.

The effects on Fos and Socs3 levels of Sox2 deletion, and of Sox2 re-expression in Sox2-deleted cells, suggest that Sox2 might act directly on these two genes. In favor of this hypothesis is the observation that Fos is connected by long-range chromatin interactions to two different functionally validated enhancers (Bertolini et al., 2019; Joo et al., 2016), shown by SOX2 ChIPseq to be bound by SOX2 (see Table S7 in [8]), and that Socs3 is bound by SOX2 on its promoter (Fig. 6B) as well as on a connected functional enhancer (Fig. 6A) (Bertolini et al., 2019). Additionally, also Jun and Egr2 regions show SOX2 binding (Fig. 7). Sox2 loss disrupts a large proportion of promoter-enhancer interactions in NSC, including those in the Fos, Socs3, Jun and Egr2 regions (Fig. 7), providing a plausible explanation for the decreased expression of all these genes.

Finally, additional studies in non-neural cells showed that both the AP1 (FOS+JUN) complex and Egr1/2 interact with the Socs3 promoter to activate it in transfection experiments (Fig. 7) (Barclay et al., 2007; Li et al., 2012). Preliminary experiments (not shown) do in fact suggest that Egr2 transduction in Sox2-deleted cells transiently increases Socs3 levels during the initial cell culture passages.

4.1 Conclusions and Summary

Overall, these data indicate that a Sox2-dependent Fos/Jun/Egr-Socs3 network is involved in the regulation of in vitro NSC self-renewal (Fig. 6, 7). Within this network, Fos and its putative target Socs3 appear to be essential for NSC renewal. While our transduction results do not document a critical role for the Jun and Egr genes in NSC maintenance, their proposed involvement in Socs3 regulation (Fig. 6), dependence on Sox2 for expression, and binding by Sox2 (Fig. 7) point to some

role of these genes in Sox2-dependent proliferation. Indeed, it remains possible that the overall residual levels of Jun and Egr1/2 in Sox2-deleted cells are sufficient (and required) to sustain their regulatory functions in NSC, once Fos is provided to the cells by transduction, as in Fig. 2. Under these conditions, further adding Jun and Egr expression, as in Fig. 2, is not expected to reveal their necessity.
The present data open a new window on factors necessary for NSC maintenance.

Acknowledgements

We thank A. Ronchi for providing the pHR SIN BX IR/EMW vector. This research was supported by ERANET – NEURON ImprovVision and Associazione Italiana per la Ricerca sul Cancro (AIRC) grant IG 2014 – 16016 to S.K.N. M.P. is the recipient of a Dipartimenti di Eccellenza fellowship. C.B. is the recipient of a DIMET (Doctorate in Molecular and Translational Medicine) PhD fellowship.

Disclosure of potential conflicts of interest

The authors declare no potential conflicts of interest.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Adepoju A, Micali N, Ogawa K et al. FGF2 and insulin signaling converge to regulate cyclin D expression in multipotent neural stem cells. Stem Cells 2014;32(3):770-778.

Barbarani G, Fugazza C, Barabino SML et al. SOX6 blocks the proliferation of BCR-ABL1(+) and JAK2V617F(+) leukemic cells. Scientific reports 2019;9(1):3388.

Barclay JL, Anderson ST, Waters MJ et al. Regulation of suppressor of cytokine signaling 3 (SOC3) by growth hormone in pro-B cells. Molecular endocrinology 2007;21(10):2503-2515.

Bertolini JA, Favaro R, Zhu Y et al. Mapping the Global Chromatin Connectivity Network for Sox2 Function in Neural Stem Cell Maintenance. Cell Stem Cell 2019;24(3):462-476 e466.

Bertolini JA, Mercurio, S., Favaro, R., Mariani, J., Ottolenghi, S., Nicolis, S.K. . Sox2-dependent regulation of neural stem cells and CNS development. In: Kondoh H, Lovell-Badge, R., ed. Sox2, Biology and role in development and disease. Elsevier, 2016.

Cavallaro M, Mariani J, Lancini C et al. Impaired generation of mature neurons by neural stem cells from hypomorphic Sox2 mutants [Research Support, Non-U.S. Gov't] [in eng]. Development 2008;135(3):541-557.

Cimadamore F, Amador-Arjona A, Chen C et al. SOX2-LIN28/let-7 pathway regulates proliferation and neurogenesis in neural precursors. Proc Natl Acad Sci U S A 2013;110(32):E3017-3026.

Eferl R, Wagner EF. AP-1: a double-edged sword in tumorigenesis. Nature reviews Cancer 2003;3(11):859-868.

Fantes J, Ragge NK, Lynch SA et al. Mutations in SOX2 cause anophthalmia [in eng]. Nat Genet 2003;33(4):461-463.

Favaro R, Valotta M, Ferri AL et al. Hippocampal development and neural stem cell maintenance require Sox2-dependent regulation of Shh [Research Support, Non-U.S. Gov't] [in eng]. Nat Neurosci 2009;12(10):1248-1256.

Gritti A, Parati EA, Cova L et al. Multipotential stem cells from the adult mouse brain proliferate and self-renew in response to basic fibroblast growth factor. J Neurosci 1996;16(3):1091-1100.

Joo JY, Schaukowitch K, Farbiak L et al. Stimulus-specific combinatorial functionality of neuronal c-fos enhancers. Nat Neurosci 2016;19(1):75-83.

Li S, Miao T, Sebastian M et al. The transcription factors Egr2 and Egr3 are essential for the control of inflammation and antigen-induced proliferation of B and T cells. Immunity 2012;37(4):685-696.

Pevny LH, Nicolis SK. Sox2 roles in neural stem cells. Int J Biochem Cell Biol 2010;42(3):421-424.

Shaulian E, Karin M. AP-1 in cell proliferation and survival. Oncogene 2001;20(19):2390-2400.

Sisodiya SM, Ragge NK, Cavalleri GL et al. Role of SOX2 mutations in human hippocampal malformations and epilepsy [Comparative Study Research Support, Non-U.S. Gov't] [in eng]. Epilepsia 2006;47(3):534-542.

Zappone MV, Galli R, Catena R et al. 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 [Research Support, Non-U.S. Gov't] [in eng]. Development 2000;127(11):2367-2382.

Zhang Y, Wong CH, Birnbaum RY et al. Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. Nature 2013;504(7479):306-310.

Figures and tables

A

Gene	wt	Sox2-mut	Fold ratio wt/Sox2-mut
Socs3	254.43	29.32	8.7
Fos	2868.27	818.55	3.5
Jun	621.48	248.82	2.5
Egr2	164.05	82.58	2.0

B

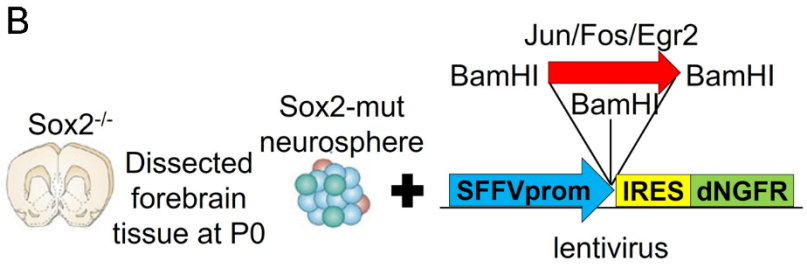
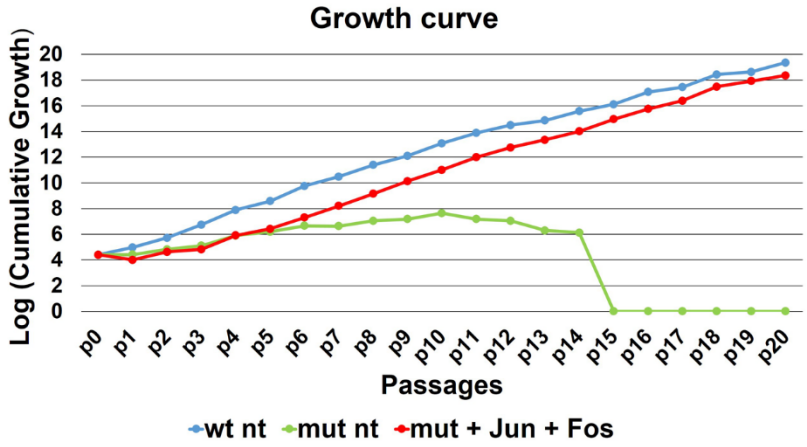


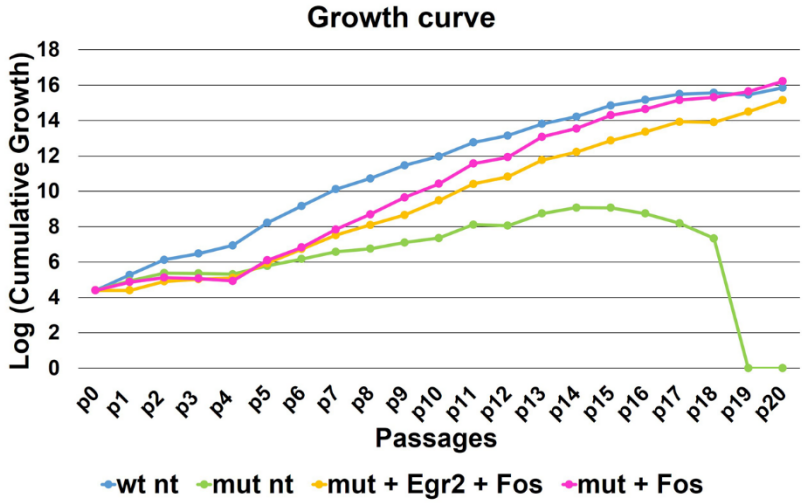
Figure 1 A, Expression levels of selected transcription factor genes among the most downregulated genes in Sox2-deleted NSC (transcripts per million; data from Bertolini et al, 2019).

B, Scheme of the Sox2 target transduction experiment and of the lentiviral vector

A


Growth curve
Log (Cumulative Growth)
0
2
4
6
8
10
12
14
16
18
20
p0 p1 p2 p3 p4 p5 p6 p7 p8 p9 p10 p11 p12 p13 p14 p15 p16 p17 p18 p19 p20
Passages
wt nt
mut nt
mut + Jun + Fos

B


Growth curve
Log (Cumulative Growth)
0
2
4
6
8
10
12
14
16
18
p0 p1 p2 p3 p4 p5 p6 p7 p8 p9 p10 p11 p12 p13 p14 p15 p16 p17 p18 p19 p20
Passages
wt nt
mut nt
mut + Egr2 + Fos
mut + Fos

C

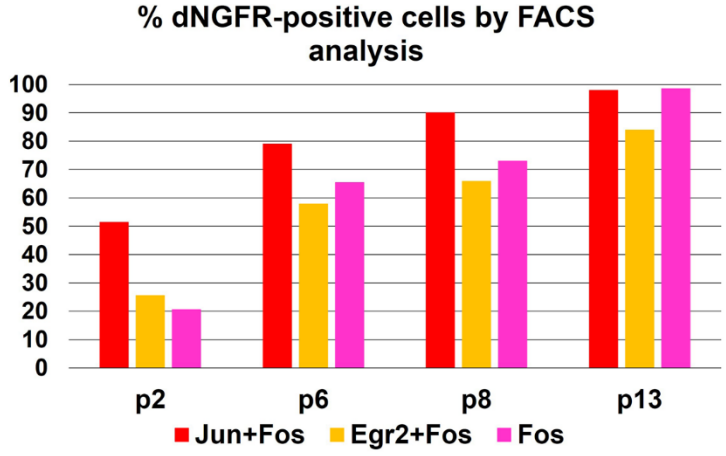

% dNGFR-positive cells by FACS analysis
0
10
20
30
40
50
60
70
80
90
100
p2
p6
p8
p13
Jun+Fos
Egr2+Fos
Fos

Figure 2 Sox2-mutant cells transduced with Fos+Jun, Fos+Egr2, or Fos alone recover the ability to self-renew long term

A,B, growth curves of Sox2-deleted cells (green), and of the same cells transduced with the indicated vectors

C, Percentage of dNGFR-positive cells by FACS analysis

A

Fos+Jun transduction	
Fos+, Jun-	12
Fos+, Jun+	2
Fos-, Jun+	0
Fos-, Jun-	0

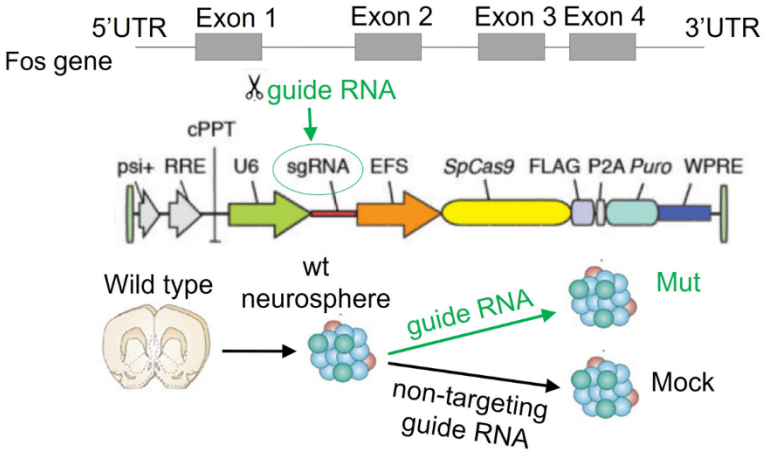
B

Fos+Egr2 transduction	
Fos+, Egr2-	6
Fos+, Egr2+	8
Fos-, Egr2+	0
Fos-, Egr2-	0

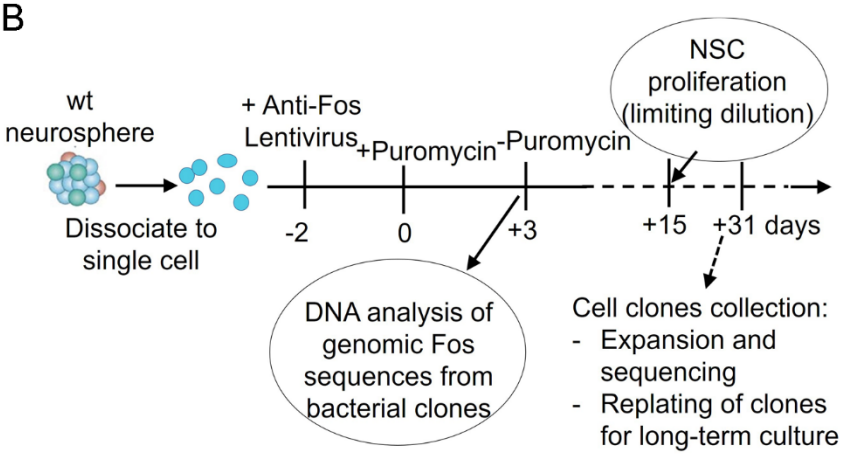
Figure 3 Clones cultured from Sox2-mutant cells rescued with Fos+Jun (A), or Fos+Egr (B), always contain the Fos-expressing lentivirus

Numbers of clones are indicated. The transduction efficiency of the cells by Jun alone, in preliminary experiments, was 18.5%; hence, we would expect that 18.5% of the 14 Fos-positive clones, i.e 2-3 clones, should also be Jun-positive, by chance. For Egr2, the expected frequency of Fos/Jun positive clones is 17.1% of the 14 Fos-positive clones, i.e. 2-3 clones.

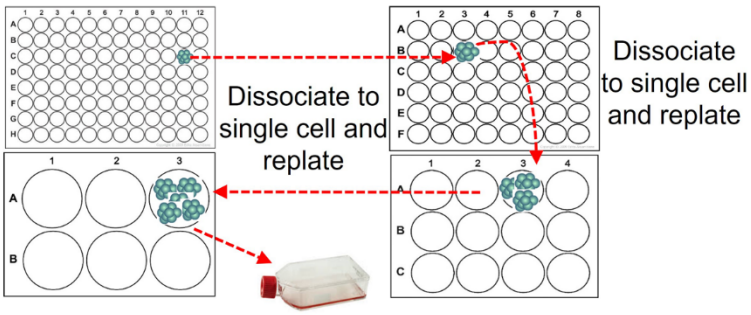
A



B



C



D

Long-term proliferating clones				
	Experiment 1		Experiment 2	
	Mut	Mock	Mut	Mock
Total replated clones	23	11	25	20
Long-term proliferating clones	7 (30,4%)	8 (72,7%)	8 (32%)	16 (80%)

Figure 4 Fos mutagenesis impairs the clonogenic ability of NSC

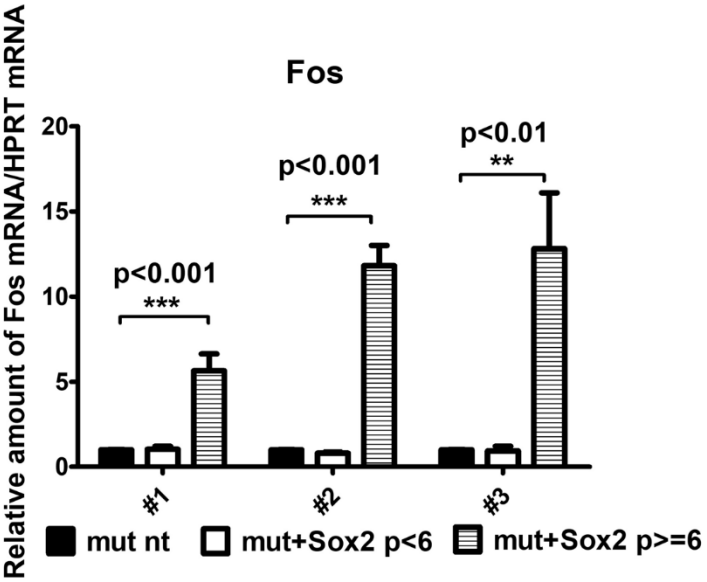
A, Position of the guide RNA-recognized sequence in the Fos gene; lentiviral vector for mutagenesis expressing the guide RNA and Cas9

B, Fos mutagenesis experiment and obtainment of transduced, puromycin-resistant NSC as single clones at limiting dilution (primary clones)

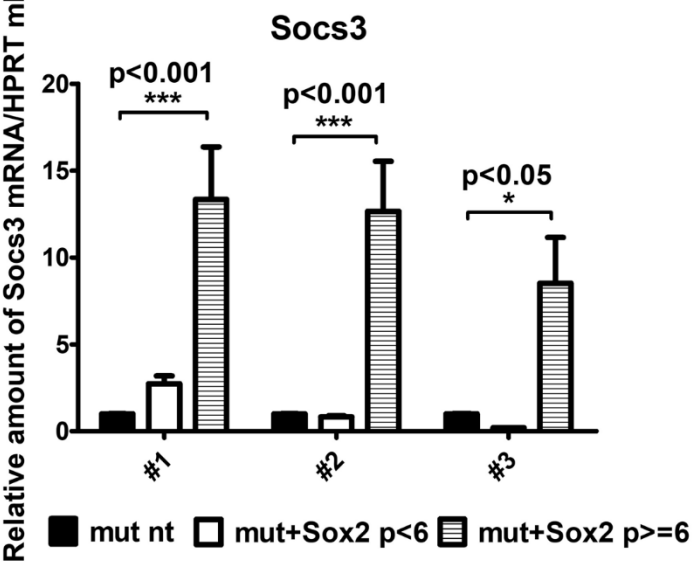
C, Replating of primary single clones for progressive expansion into long-term growing NSC populations (long-term proliferating clones).

D, Fos mutagenesis in wild type NSC reduces the percentage of long-term proliferating clones. The significance of the difference between number of clones obtained with anti-Fos (Mut) and scramble (Mock) sgRNA was validated by paired one-tailed t-test ($p=0.03127$)

A


Fos
Relative amount of Fos mRNA/HPRT mRNA
20
15
10
5
0
p<0.001
***
p<0.001
***
p<0.01
**
#1
#2
#3
mut nt
mut+Sox2 p<6
mut+Sox2 p>=6

B


Socs3
Relative amount of Socs3 mRNA/HPRT mRNA
20
15
10
5
0
p<0.001
***
p<0.001
***
p<0.05
*
#1
#2
#3
mut nt
mut+Sox2 p<6
mut+Sox2 p>=6

C

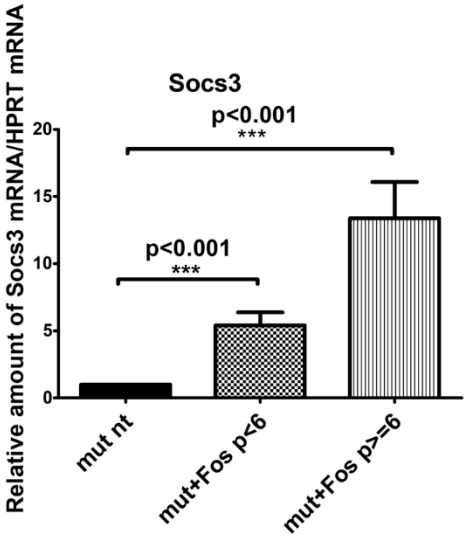

Socs3
Relative amount of Socs3 mRNA/HPRT mRNA
20
15
10
5
0
p<0.001
***
p<0.001
***
mut nt
mut+Fos p<6
mut+Fos p>=6

Figure 5 Sox2-rescued Sox2-deleted cells recover expression of endogenous Fos and Socs3

A,B, Expression of Fos (A) and Socs3 (B) in Sox2-mutant cells (three different mutants, #1, #2, #3) transduced with Sox2-expressing lentivirus. P<6 and p>6 are the average expression values for passages 2-5 and 6-22, respectively.

C, Expression of Socs3 in Sox2-deleted cells transduced with Fos. mut: Sox2-deleted. nt: non transduced.

In A,B,C, expression values are calculated as ratios of Fos or Socs3 mRNA/HPRT mRNA, as determined by qRT-PCR. The values obtained for untransduced mutant (mut nt) are set = 1, and the values obtained for Sox2- or Fos-transduced mutants (mut + Sox2; mut + Fos) are normalized to the corresponding mut nt.

Histograms in A,B,C represent the mean of results obtained with the following numbers of qRT-PCRs at different passages: mut 1 and 2: 20 qRT-PCRs performed in duplicate for mut nt and mut+Sox2 p>6, and 13 qRT-PCRs performed in duplicate for mut nt and mut+Sox2 p<6; mut 3: 10 qRT-PCRs for nt and mut+Sox2 p>6; 3 qRT-PCRs for mut nt and mut+Sox2 p<6. ***p<0.001; *p<0.05, paired t-test. ns: not significant.

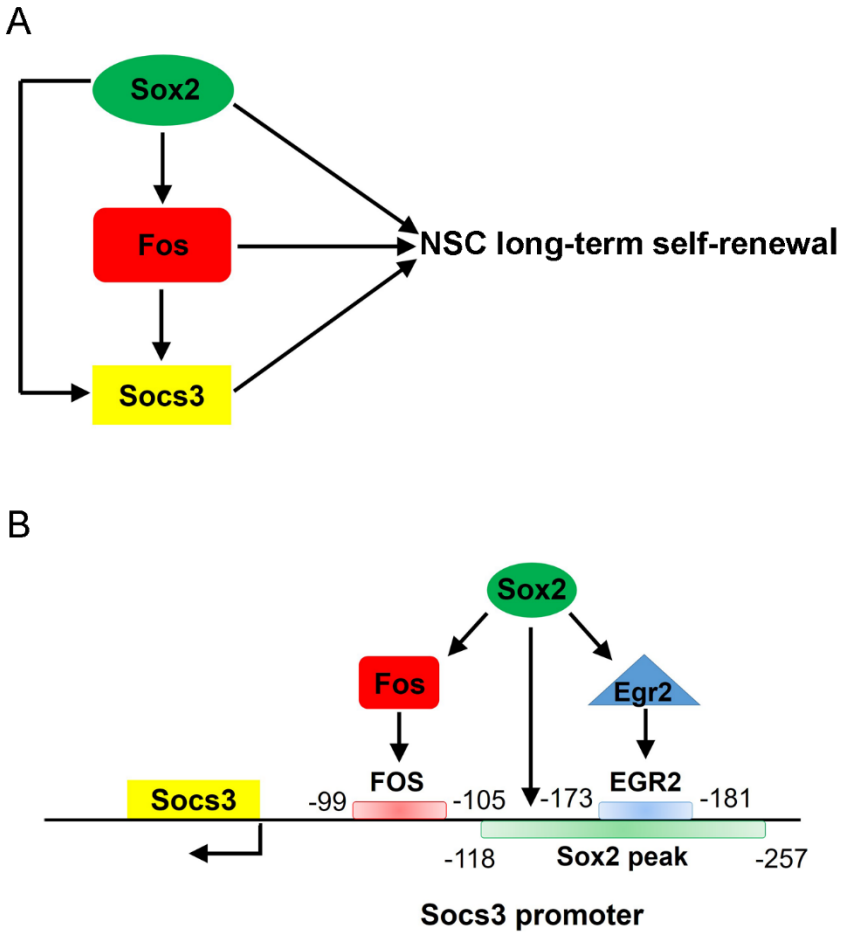
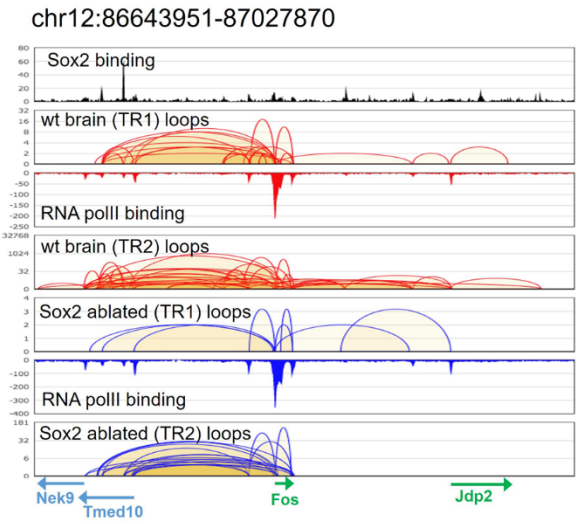


Figure 6 A model for a Sox2-controlled gene regulatory network involving Fos, Egr2 and Socs3

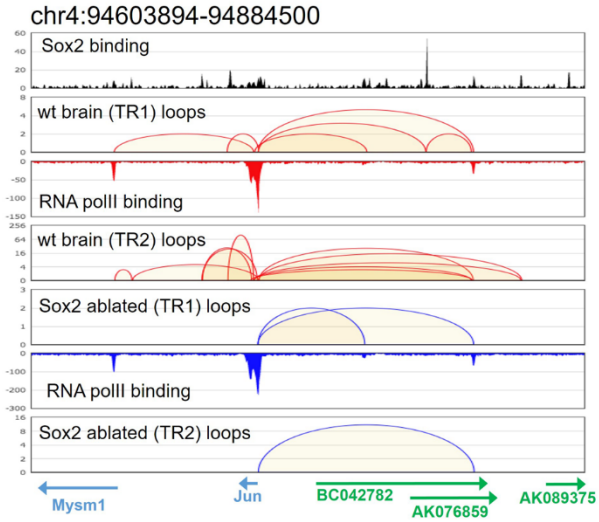
A, Regulatory relations between Sox2, Fos, and Socs3, in NSC long-term self-renewal control

B, Sox2, Fos and Egr2 functional interactions on the Socs3 promoter. For Fos and Egr2 binding data, see [17,18]. For Sox2 binding, see [8].

A


chr12:86643951-87027870
Sox2 binding
wt brain (TR1) loops
RNA polII binding
wt brain (TR2) loops
Sox2 ablated (TR1) loops
RNA polII binding
Sox2 ablated (TR2) loops
Nek9
Tmed10
Fos
Jdp2

B


chr4:94603894-94884500
Sox2 binding
wt brain (TR1) loops
RNA polII binding
wt brain (TR2) loops
Sox2 ablated (TR1) loops
RNA polII binding
Sox2 ablated (TR2) loops
Mysm1
Jun
BC042782
AK076859
AK089375

C

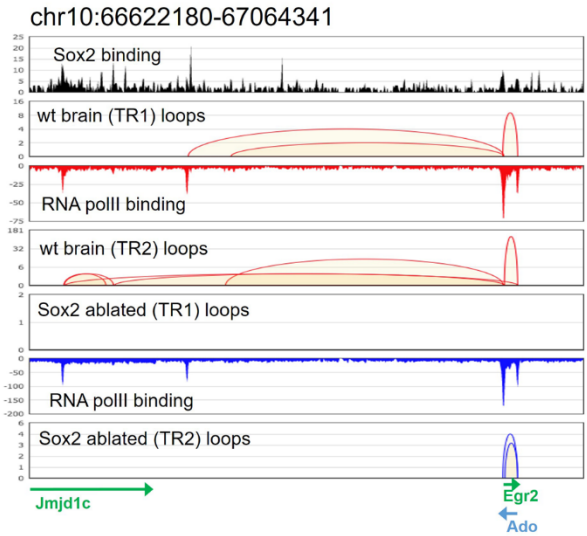
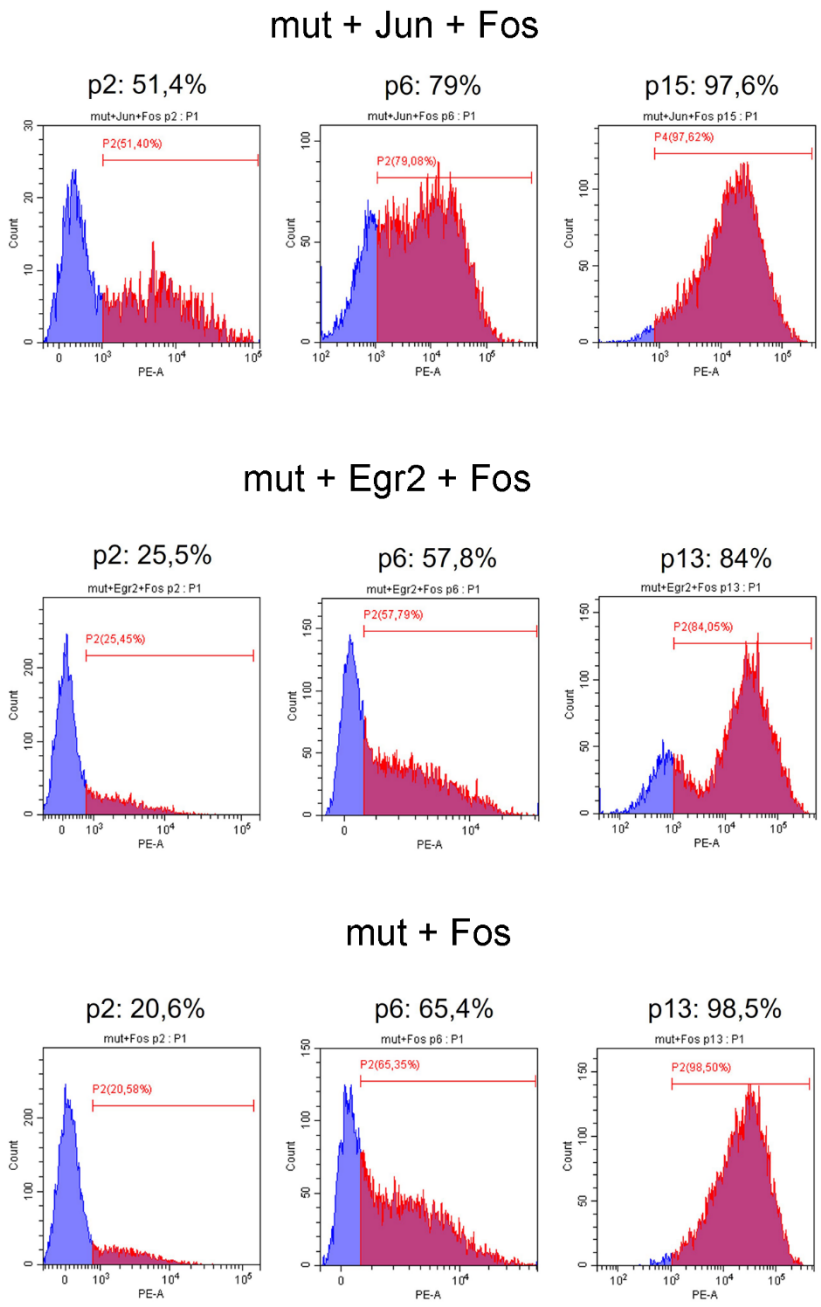

chr10:66622180-67064341
Sox2 binding
wt brain (TR1) loops
RNA polII binding
wt brain (TR2) loops
Sox2 ablated (TR1) loops
RNA polII binding
Sox2 ablated (TR2) loops
Jmjd1c
Egr2
Ado

Figure 7 Long-range chromatin interactions detected by ChIA-PET on the Fos (A), Jun (B), and Egr2 (C) regions in wild type and Sox2-deleted NSC

The Fos (A), Jun (B) and Egr2 (C) genes are indicated below each panel. Long-range interactions between the genes and distal regions are indicated by loops (wild type cells, red; Sox2-deleted cells, blue). Tracks with SOX2 and RNApolII binding peaks are shown. TR1 and TR2 refer to different ChIA-PET experiments. Note the occurrence of SOX2 binding peaks in correspondence of distant regions connected with each of the three genes. Data from (Bertolini et al, 2019).

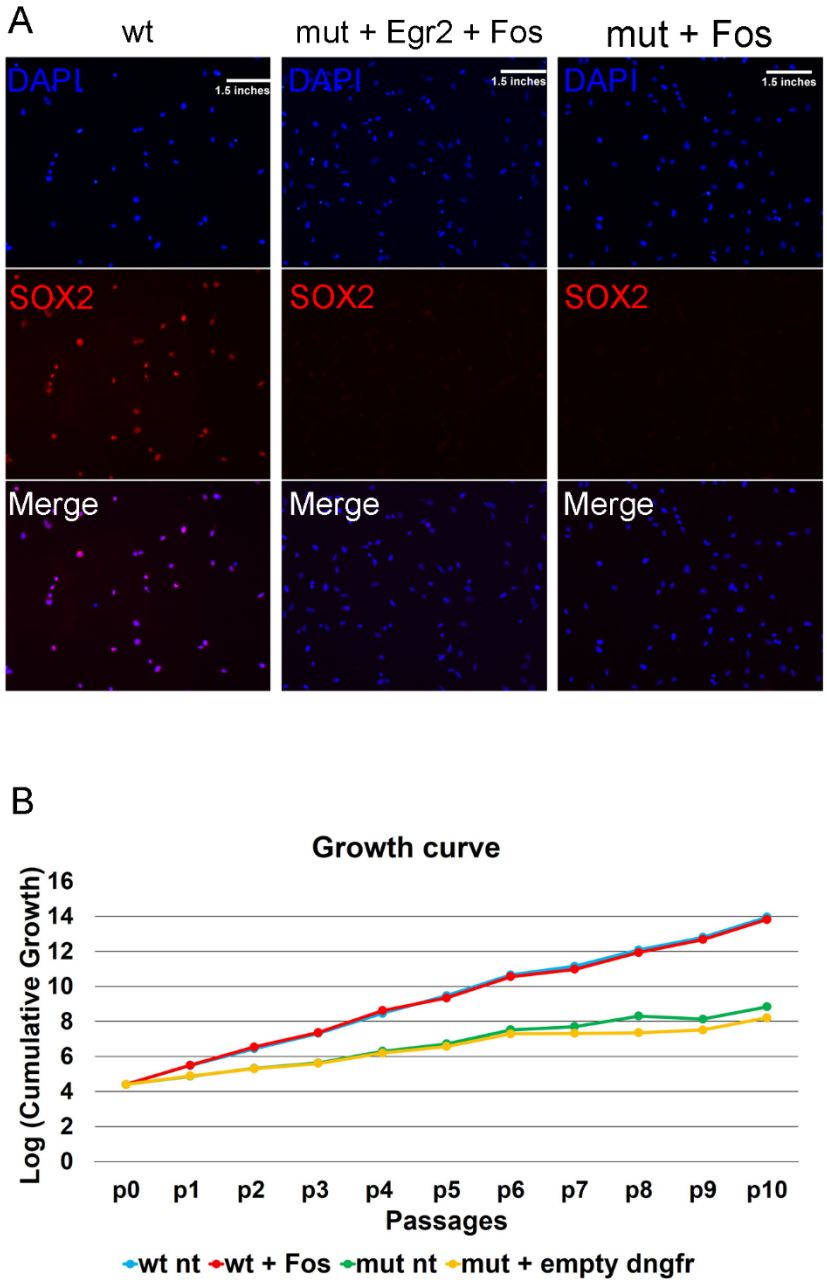
Supplemental information



Supplementary Fig. 1

FACS analysis of dNGFR-positive Sox2-deleted cells transduced with Jun and Fos, Egr2 and Fos, or Fos only

dNGFR-positive cells are indicated by the red color, and the corresponding percentage is indicated. Their proportion increases progressively at different passages.

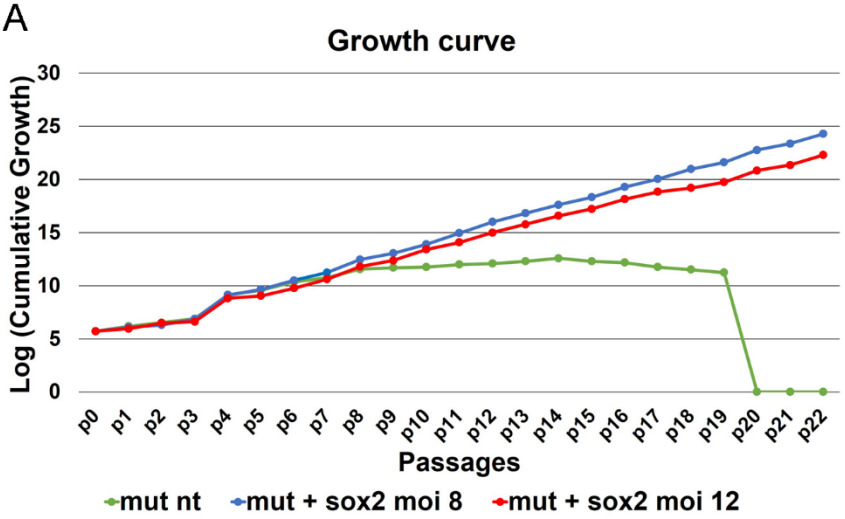


Supplementary Fig. 2

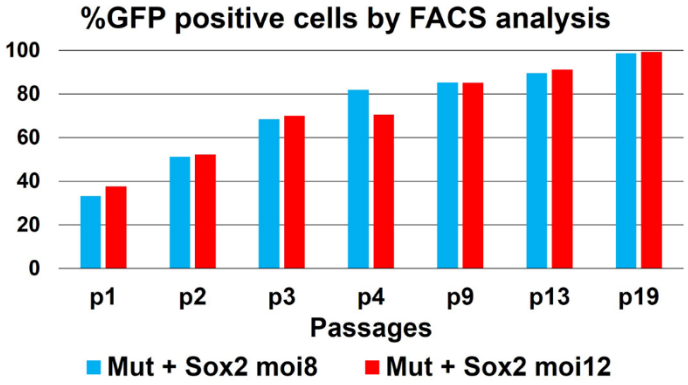
A, Absence of residual SOX2-positive cells in long-term cultures of Sox2-deleted NSC transduced with Fos+Egr2, or Fos only

B, Growth of wild type NSC transduced with Fos or not transduced (nt). Growth of Sox2-deleted (mut) cells transduced with empty vector (empty dngfr) or not transduced (nt).

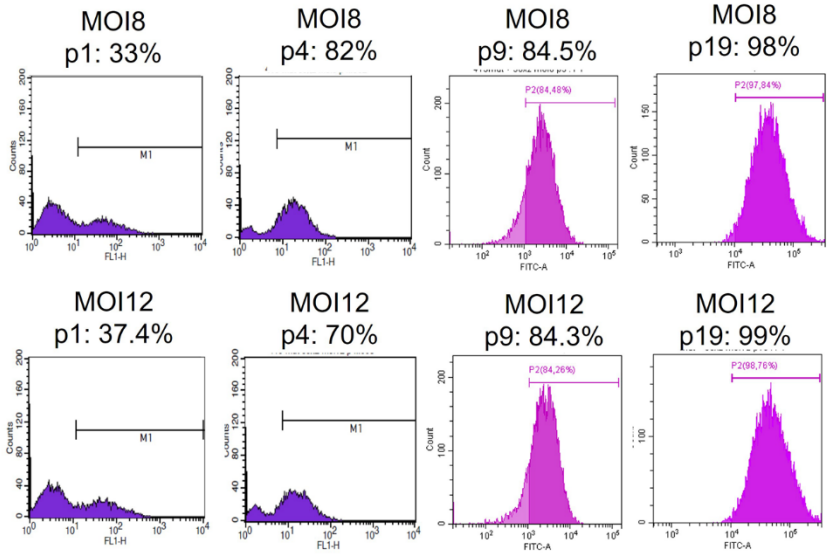
A


Growth curve
Log (Cumulative Growth)
30
25
20
15
10
5
0
p0 p1 p2 p3 p4 p5 p6 p7 p8 p9 p10 p11 p12 p13 p14 p15 p16 p17 p18 p19 p20 p21 p22
Passages
mut nt
mut + sox2 moi 8
mut + sox2 moi 12

B


%GFP positive cells by FACS analysis
100
80
60
40
20
0
p1 p2 p3 p4 p9 p13 p19
Passages
Mut + Sox2 moi8
Mut + Sox2 moi12

C


MOI8
p1: 33%
MOI8
p4: 82%
MOI8
p9: 84.5%
MOI8
p19: 98%
MOI12
p1: 37.4%
MOI12
p4: 70%
MOI12
p9: 84.3%
MOI12
p19: 99%
M1
FL1-H
FITC-A
Count
Counts

Supplementary Fig. 3

Rescue of long-term proliferation of Sox2-deleted NSC by lentiviral Sox2 transduction

A, Growth curve of Sox2-deleted (mut) cells, untransduced (green) or transduced with Sox2- and GFP-expressing lentivirus at MOI 8 (blue) or 12 (red).

B, Percentage of GFP-positive cells at different passages

C, Examples of FACS analyses at different passages

	Mutant bacterial clones			Mutant alleles in cellular clones
	36h after puromycin selection	8 days after puromycin selection	20 days after puromycin selection	
Mut	9/10 tot	10/10	16/16 tot	14/18 tot *
Mock	0/3 tot			0/8

* One of the Mut clones was wild type homozygous

Supplementary Table 1

Number of mutant/total alleles of Fos following CRISPR/Cas9 mutagenesis

DNA was extracted from cells treated with lentiviruses at 36 hours, 8 days, and 20 days after puromycin selection; the Fos gene region containing the sgRNA-targeted sequence was amplified by PCR, and cloned into a pGEM®-T Easy plasmid. Bacterial clones were individually grown, plasmid DNA was extracted, and the plasmid insert was sequenced. The number of clones carrying a mutated Fos region is reported, over the total number of sequenced clones. The number of mutant alleles was also determined in individual cellular clones obtained after long-term propagation. Mutant clones were not found in Mock-transduced controls.

	Experiment 1		Experiment 2	
	Mut	Mock	Mut	Mock
Clones collected from 96-well plate	**23**	**11**	**25**	**20**
Clones which died in 48-well plate	6	0	9	0
Clones that do not proliferate after seeding in 12-well plate	10	3	8	4
Clones that continue to proliferate (in 6-well plate or in flask)	**7**	**8**	**8**	**16**

Supplementary Table 2

Comparison between antiFos Cas9-treated and Mock-treated NSC, as to their ability to generate clones capable of long-term expansion

Numbers of primary clones in two experiments (clones from 96-well plates) obtained for further replating for long-term growth analysis (first row). After replating, some of the clones continued to efficiently proliferate long-term in 6-well plates or in flasks, for at least two months (fourth row). Only these clones are scored as derived from long-term self-renewing NSC.

Chapter 6

1. Summary

1.1 *The relevance of Sox2 and its downstream target genes in CSCs maintenance*

If we think about brain development as a hierarchically organized process, the neural stem cells occupy the top of the hierarchy, as they are capable of self-renewal (by symmetric division) and to give rise to a progeny of more differentiated cells that will form a functional organ (by asymmetric division). It is well known that Sox2 plays a key role in the self-renewal of stem cells. It is required from the early stages of mouse embryo development in the maintenance of Embryonic Stem cells of the Inner Cell Mass (Avilion et al., 2003). Later in development Sox2 expression and functional role become restricted to the neural tube and here it is essential for the maintenance of neural stem cells and the development of several central nervous system structures (Favaro et al., 2009; A. L. Ferri et al., 2004; Mercurio et al., 2019; Pevny & Nicolis, 2010).

In 1997, it was proposed for the first time by Bonnet et al that tumors could have the same hierarchical organization that we find in normal tissues (Bonnet & Dick, 1997). In fact, tumors can be described as aberrant organs originated from a Tumor-Initiating Cell (TIC) also called Cancer Stem Cell (CSC) (Reya, Morrison, Clarke, & Weissman, 2001).

In the field of cancer therapies, there is a growing interest in understanding the molecular mechanisms involved in survival and self-renewal of CSCs, in order to target these cells responsible of tumor relapse, metastasis and drug resistance (Reya et al., 2001; Wang, Ma, & Cooper, 2013).

My thesis work is placed in this context. Starting from the two points discussed above, we hypothesized that Sox2 could have a key function in the maintenance of

CSCs, as previously found for normal NSCs. Several groups showed that Sox2 has a functional role in many types of tumor, including brain tumors (Barone et al., 2018; Hüser et al., 2018).

In previous work, Favaro *et al. (*Favaro *et al., 2014)* demonstrated an essential role of transcription factor Sox2 for the maintenance of neural CSCs, using a mouse model of PDGF-induced high-grade glioma (pHGG), in which Sox2 can be deleted by a CRE-ricombinase/loxP system. Transplanting wild-type pHGG cells into the mouse brain generated lethal tumors, but mice transplanted with Sox2-deleted cells remained tumor-free. Cultured Sox2-deleted pHGG cells show decreased growth-rate, activation of glial differentiation, and increased cell death compared to pHGG cells that express Sox2. The authors tested SOX2 as a therapeutic target by using SOX2 peptides for mouse immunization. They found that the immunized mice displayed a delayed tumor onset and prolonged survival. As Favaro *et al.*, many other studies are testing Sox2 as direct pharmacological target (Huser, Novak, Umansky, Altevogt, & Utikal, 2018). It is our opinion that an alternative and cooperative approach could be to find "druggable" molecular pathways regulated by Sox2, and acting as effectors of its function in CSCs. In order to address this, we performed a differential expression analysis by microarray, comparing Sox2-WT and Sox2-deleted pHGG cells. After Sox2 Cre-mediated deletion, 134 genes are up-regulated and 12 are down-regulated. My hypothesis was that, since the final effect of Sox2 loss was the loss of tumorigenicity, among the genes upregulated after Sox2 loss we could find tumor suppressor genes, repressed by Sox2 in pHGG. To test this hypothesis, I performed gain-of-function and loss-of-function experiments in pHGG cells. I tested only a subset of the genes upregulated after Sox2 loss (9 genes), choosing the most upregulated ones, and the ones already known to have a role in cancer.

In this way, I identified four genes, that are importantly upregulated following Sox2 deletion in pHGG cells, as mediators of the proliferation arrest and inhibition of tumorigenesis due to Sox2 deletion. In fact, each of these genes (Cdkn2b, Ebf1, Hey2, Zfp423), when experimentally overexpressed in Sox2-wt pHGG cells by viral transduction, significantly reduced cell proliferation, and inhibited tumorigenesis after transplantation of the transduced cells into the mouse brain. Not only are these genes able to reproduce the effects of Sox2 deletion in pHGG cells; their activity is also essential for repressing cell proliferation in Sox2-deleted pHGG cells, as shown by CRISPR/CAS9 mutagenesis or pharmacological inhibition. These results indicate that a critical contribution of Sox2 to the maintenance of tumorigenesis is represented by its ability to inhibit the expression of at least four genes acting as tumor suppressors. Interestingly, three out of the four genes, all encoding transcription factors (Ebf1, Hey2 and Zfp423), appear to be connected in a functional interaction network, thanks to which the overexpression of one of them cause the upregulation of the others.

The tumor suppressive function of Cdkn2b, Ebf1, Hey2, Zfp423 found in this murine model of oligodendroglioma is confirmed in human glioblastoma samples from patients (thanks to a collaboration with Mariachiara Buccarelli, Roberto Pallini and Lucia Ricci-Vitiani of the “Istituto Superiore di Sanità” in Rome). In fact, we found that gene overexpression in three human glioblastoma primary cell lines was able to reduce their clonogenic efficiency, index of stemness grade. Moreover Hey2 overexpression reduced migration in two out of three of these lines. With this work, we discovered a gene regulatory network downstream to Sox2, that suggests novel targets of possible therapeutic relevance.

1.2 Sox2 and its downstream target genes relevance in NSCs maintenance.

During my PhD work, I had the possibility to collaborate with my colleagues in projects about Sox2 function in normal neural stem cells (NSCs). These collaborations opened the possibility to understand possible parallelism or differences in Sox2 functions between NSCs and CSCs.

The study of Sox2 role in normal brain development has a translational implication because 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; Kelberman et al., 2008; Ragge et al., 2005; Schneider, Bardakjian, Reis, Tyler, & Semina, 2009; Sisodiya et al., 2006). In order to understand the role of Sox2 in neural development, our laboratory generated Sox2 conditional KO mutations in mouse (Favaro et al., 2009; A. Ferri et al., 2013). The consequences of Sox2 ablation at different developmental time points produced important brain defects, more serious when the ablation was earlier. 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 forebrain, 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).

These studies highlighted 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, key questions are:

which genes are SOX2 targets? How does SOX2 regulate their expression?, Which SOX2-regulated genes are critical mediators of its function?

A new way in which SOX2 regulates its target genes has been recently shown in a study I participated in (see Chapter 4): 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 chromatin. Moreover, we defined a genome-wide map of SOX2 binding sites in wt NSC chromatin by ChIP-seq with anti-SOX2 antibodies (in collaboration with F. Guillemot, London).

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, whose expression was analyzed by RNA-seq technique.

Confirming Sox2 role in NSCs self-renewal, we observed that genes encoding well-known regulators of cell proliferation *(Socs3,* c-*fos*, *Jun*, *JunB*, *Btg2*, *Egr1*, and *Egr2*), are expressed at high levels in WT NSCs and are substantially downregulated in Sox2 mut cells. These genes show multiple promoter-enhancer interactions in wt NSCs. In Bertolini *et al*. (see Chapter 4), we found that Socs3 overexpression rescued the self-renewal defect of Sox2-ablated NSC indicating that this gene is a key mediator of Sox2 function.

My contribution to this collaborative paper focused on the selection of distal interaction "anchors", connected to gene promoters via long-range interactions identified by RNApolII ChIA-PET, for their functional validation as transcriptional enhancers. Within each anchor (2-3 kb of average length), I defined a region most likely to represent the putative enhancer "core", based on the presence of transcription factor binding sites, and evolutionary conservation. I generated many

constructs carrying these sequences upstream to reporter genes (luciferase for transfection; GFP for transgenesis in zebrafish). These constructs allowed to prove that 14 out of 15 such distal sequences, when tested in a transgenesis assay in the zebrafish embryo, acted as enhancers driving GFP expression to the developing forebrain. Some of the enhancers were also active in transfection and cotransfection assays. Collectively, these results showed that most distal sequences identified via RNApolII ChIA-PET represent "bona fide" enhancers, active in the forebrain, the original source of the NSC onto which we performed the original ChIA-PET analysis.

I collaborated also with Pagin *et al*. (see Chapter 5), to better investigate the regulatory gene network downstream of Sox2, responsible of NSCs self-renewal, taking advantage from genome-wide analysis done in the previous work (Bertolini et al., 2019).

To identify molecules regulated by Sox2, and acting in mouse NSC maintenance, we transduced, individually or in combination, into Sox2-deleted NSC, genes whose expression is strongly downregulated following Sox2 loss (Fos, Jun, Egr2). Fos alone rescued long-term proliferation. Further, Fos requirement for efficient long-term proliferation was demonstrated by the strong reduction of NSC clones capable of long-term expansion following CRISPR-Cas9-mediated Fos inactivation. Previous work showed that the Socs3 gene is strongly downregulated following Sox2 deletion, and its re-expression by lentiviral transduction rescues long-term NSC proliferation. Fos appears to be an upstream regulator of Socs3, possibly together with Jun and Egr2. These results provide the basis for developing a model of a network of interactions, regulating critical effectors of NSC proliferation and long-term maintenance.

2. Conclusions

The translational importance of my research is due to the need to understand the molecular mechanisms that sustain the functionality of CSCs, in order to interfere with them and to hit this subpopulation of cells responsible of tumor relapse, metastasis and drug resistance. What we observed in the murine model of oligodendroglioma used in my work is likely to be conserved in human, because of our result on human glioblastoma samples, but also because our observations are supported by previous published works.

Of the four identified putative tumor suppressor genes (Ebf1, Hey2, Cdkn2b, Zfp423), three have previously been proposed to be somehow involved in spontaneous gliomas in humans. From the analysis of spontaneous mutations leading to gliomas, specific pathways involving antioncogenic proteins, such as Retinoblastoma (Rb), appear to be involved, with different mutations often acting in conjunction. For example, mutations/deletions in CDKN2A and CDKN2B are are often found in gliomas, these genes are inhibitors of CDK4 and CDK6 kinases which are activators of Rb, therefore these mutations result in loss of Rb activity (Rao, Edwards, Joshi, Siu, & Riggins, 2010).

Ebf1 is a transcription factor known to have an important role in B-cell differentiation (Nechanitzky et al., 2013) and in neuronal differentiation during embryogenesis (Garel et al., 1997). In addition, Ebf1 was found among commonly mutated genes in two large cohorts of human grade II and III gliomas (Suzuki et al., 2015) identified among commonly mutated genes also EBF1. Furthermore, both Ebf1 and Ebf3 (a member of the family) have been found to be mutated and possibly inactivated in a variety of tumors (Liao, 2009), implying a tumor suppressive role for these genes. In particular, EBF3 is mutated in gliomas, and it activates genes

involved in cell cycle arrest and apoptosis while repressing genes involved in cell survival and proliferation (Liao, 2009); whether Ebf1, when increased following Sox2 deletion in pHGG cells, plays a similar role, is an interesting possibility.

Hey 2, a transcription factor, is an effector downstream to the Notch signaling pathway; there are contrasting results reported in glioblastoma, regarding the effects of the activation of Notch signaling on glioblastoma cells proliferation (Giachino et al., 2015; Ying et al., 2011). Our present results are consistent with an antioncogenic effect of Hey2 activation, as proposed by Giachino et al. In mouse, Notch signaling inactivation, combined with p53 loss, leads to the generation of aggressive brain tumors (Giachino et al., 2015); in agreement, in man Notch1 inactivating mutations are detected in gliomas, and Notch pathway effectors Hey2 and Hes5 expression levels are inversely correlated with tumor severity (Giachino et al., 2015). Our results indicate that at least two of the genes upregulated following Sox2 deletion in pHGG cells fit well in this general scheme: Cdkn2b and Hey2 increases implicate an involvement of the signaling upstream to Rb, and of at least one member of the Notch pathway.

Finally, Zfp342 was also reported to have tumor suppressive activity in glioma (Signaroldi et al., 2016).

It is to be noted that the promoters of Ebf1, Cdkn2b, Zfp423 and Hey2 are directly bound by Sox2 in a human GBM-derived cell line (Fang et al., 2011), suggesting that they are direct Sox2 target in both murine and human gliomas.

The importance of studying the downstream targets of Sox2 in normal CNS development and maintenance of NSCs has the same rationale discussed above regarding Sox2 function in CSCs.

One possible translational meaning of this study is to find "druggable" pathways downstream to Sox2 to rescue neurological defects due to heterozygous SOX2 mutations in humans.

Moreover, the identification of long-range interactions in chromatin from neural stem cells that require Sox2 opened novel possibilities in terms of translational applications. Mutations in regulatory sequences could cause dramatic effects on the expression of the regulated gene, and then lead to genetic disease. 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, Pattabiraman, Visel, & Rubenstein, 2015) changes in transcription factor-binding sequences may affect chromatin modifications locally and at distant sites, affecting gene activity over great distances (Denker & 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.

3. Future perspectives

3.1 Sox2 interactome in NSCs and CSCs

An interesting observation is that Sox2 acts mainly as transcriptional activator in NSCs (Bertolini et al., 2019; see Chapter 4 and 5), whereas its function in pHGG is

mainly to repress downstream target genes (Favaro et al., 2014; Barone et al., 2020; see Chapter 2). It is already known that Sox2 could act as a transcriptional activator or repressor depending on its interactors (Cox, Mallanna, Luo, & Rizzino, 2010; Liu et al., 2014). A direct transcriptional repressor activity of Sox2 in neural stem cells, mediated by its interaction with the transcriptional repressor Groucho, has been reported (Liu et al., 2014). In future it would be interesting to compare the interactome of Sox2 in NSCs and in pHGG cells to understand if tumor-specific interactors could be found, and if we can target them to block tumor formation and/or progression.

3.2 ChIA-PET on pHGG cells

In Bertolini et al. (2019), we used RNA PolII Chromatin Interaction Analysis by Paired End Tagging (Chia-PET) (see Chapter 1; see Chapter 4) to analyze long-range interactions in transcriptional active chromatin of NSCs Sox2-wt and Sox2-mut. In Sox2-mut NSCs, this network is altered, with major reduction of RNA PolII-mediated interactions. As already discussed above, following Sox2 loss, about 900 genes were down-regulated (with relatively few being upregulated), and SOX2-bound genes involved in long-range interactions were the most represented category within downregulated genes, pointing to an activatory function for SOX2, acting via enhancer-promoter interactions (Bertolini et al., 2019) (see Chapter 3). Notably, in glioma pHGG stem cells, most of the genes deregulated following Sox2 deletion show increased expression (Favaro et al.,2014; Barone et al., 2020; see Chapter 2), rather than downregulation, as in normal NSCs. This suggests that, in pHGG cells, Sox2 may act, directly or indirectly, as a repressor. This finding points to potentially important differences between the Sox2-dependent chromatin organization in wild-type NSCs and tumorigenic stem cells. Hence, in collaboration

with Dr. Chia-Lin Wei, we proposed to study by RNA PolII ChIA-PET the chromatin of pHGG cells before and after Sox2 deletion, and to compare these tumorigenic neural cells with wild type NSCs.

The recent development (in Dr. Wei's laboratory) of PRC2-ChIA-PET, identifying long-range interactions mediated by the Polycomb Repressor Complex 2 (PRC2), demonstrated that these long-range interactions associate with gene repression, and involve silencers (Ngan et al., 2020). In ES cells, early genome-wide ChIP-on Chip studies (Boyer et al., 2005) had shown that SOX2 is bound to the promoters of active genes (in association with RNA PolII), as well as silent genes (in association with PRC2), due to be activated later in development. It is possible that tumorigenesis implies a shift in SOX2 distribution from activating (polII-mediated) to repressive (PRC2-mediated) interaction loops. PRC2 Chip-Seq may be used to compare the distribution on PRC2 on enhancers and promoters bound by Sox2 in Sox2-wt pHGG cells with the distribution in Sox2-deleted cells.

The experiments done in human glioblastoma (GBM) primary cells from patients show a decrease in self-renewing capability of cells overexpressing the four tumor suppressing genes. To confirm the *in vitro* results we are planning, in collaboration with Prof. Roberto Pallini, to xenotransplant these cells in immunocompromised mice, in order to see if the overexpression of the four tumor suppressing genes is sufficient to reduce the tumorigenicity of human GBM cells *in vivo*.

To better validate our molecular model obtained in mouse, we are also planning to test the SOX2-dipendance of the human GBM primary cells used for our experiments, using CRISPR-Cas9 technology to mutate the resident SOX2 gene.

Furthermore, this approach could allow us to investigate, by RNA-seq or microarrays, deregulated gene expression, following Sox2 loss, in human GBM

primary cells, in order to compare these data with data obtained from mouse pHGG cells.

3.3 miRNA deregulation following Sox2 loss

It is of growing interest among researchers to investigate the role of microRNA (miRNA) in normal cellular as well as in disease processes. miRNAs are a family of small non-coding RNAs which were reported to regulate the expression of various oncogenes or tumor suppressor genes (Reddy, 2015).

In order to identify potential miRNAs downstream of Sox2, I performed a microarray experiment on normal and Sox2-deleted pHGG cells. I found 63 up-regulated miRNAs and only 3 down-regulated miRNAs following Sox2 deletion, suggesting that Sox2 acts as a repressor also of miRNAs. Many of the de-regulated miRNAs, are already known for their role in cancer (see Table 1): there are known tumor suppressor among the upregulated ones, and oncogenic factors among the downregulated ones. This observation fits with the tumorigenic role of Sox2 in our murine model of oligodendroglioma.

I was interested in finding, among the miRNAs potentially downstream of Sox2, miRNAs that target genes that are upregulated in GBM samples compared to normal tissues. To do this I took advantage of GEPIA (Gene Expression Profiling Interactive Analysis)

the web-based tool based on TCGA and GTEx data, I could have access to databases of genes differentially expressed in GBM samples comparing to normal brain tissue. I crossed this list of genes with a list of genes potentially targets (according to miRbase, http://www.mirbase.org) of upregulated miRNAs following Sox2 loss. My hypothesis is that it could be interesting to functionally test that miRNAs that target (and so downregulate) genes knew to be upregulated in GBM

samples comparing to normal tissues (see Figure 1a). I found two miRNAs (miR-133c and miR362-3p) that have a great overlap between their target genes and genes upregulated in GBM (see Figure 1b).

We are planning to overexpress the two miRNAs mentioned above, cloning them in a lentiviral vector, optimized for short hairpin RNA (pTRIPZ, Thermo Fisher Scientific, Waltham, MA, USA) in Sox2-wt pHGG cells, in order to see if this is sufficient to interfere with the tumorigenicity of these cells.

Table 1

miRNA (Fold-change Sox2-del vs control)	**Role in tumor**	**DOI of relevant papers**
miR-329 (3,7x UP)	Tumor suppressor in NSCLC, gastric cancer, pituitary tumor, breast cancer, pancreatic cancer	DOI: 10.18632/oncotarget.7517; DOI: 10.18632/oncotarget.2755;DOI: 10.18632/oncotarget.5003; https://doi.org/10.1038/cdd.2015.116; DOI: 10.18632/oncotarget.7375
miR-34c-3p (3,3x UP)	Tumor suppressor in glioma; Low er circulating doses in Brest Cancer patients	DOI: 10.3892/ol.2013.1579 ; https://doi.org/10.3349/ymj.2017.58.4.697
miR-376b-3p (3,3x UP)	Lower circulating doses in glioma patients	DOI: 10.3233/CBM-160146
miR-299b (3,2x UP)	Downregulated in Cervical Cancer	DOI: 10.1007/s11033-013-2998-0
miR-135b-3p (3,2x UP)	Tumor suppressor in GBM	DOI: 10.18632/oncotarget.5925

miR-202-3p (2,9x UP)	Tumor suppressor in gastric cancer, in Cervical cancer, in osteosarcoma	DOI: 10.1371/journal.pone.0069756; DOI: 10.18632/oncotarget.12499; https://doi.org/10.1007/s11010-014-2195-z
miR-205-3p (2,6x UP)	Tumor suppressor in Brest Cancer	DOI: 10.1172/JCI73351
miR-451b (2,6x UP)	Tumor suppressor in T-ALL	https://doi.org/10.1084/jem.20102384
miR-6946-5p (2,5x UP)	Tumor suppressor in colon-rectal cancer	DOI: 10.1038/onc.2017.4; DOI: 10.4238/gmr.15027730
miR-452 (2,3x UP)	Tumor suppressor in lung cancer	https://doi.org/10.1159/000430362
miR-297b (2,3x UP)	Tumor suppressor in lymphoma	https://doi.org/10.3109/10428194.2012.678005
miR-92a-3p (2,6x DOWN)	Oncogenic in lung cancer and osteosarcoma	DOI: 10.3892/ijo.2017.3999; https://doi.org/10.3892/or.2017.5484

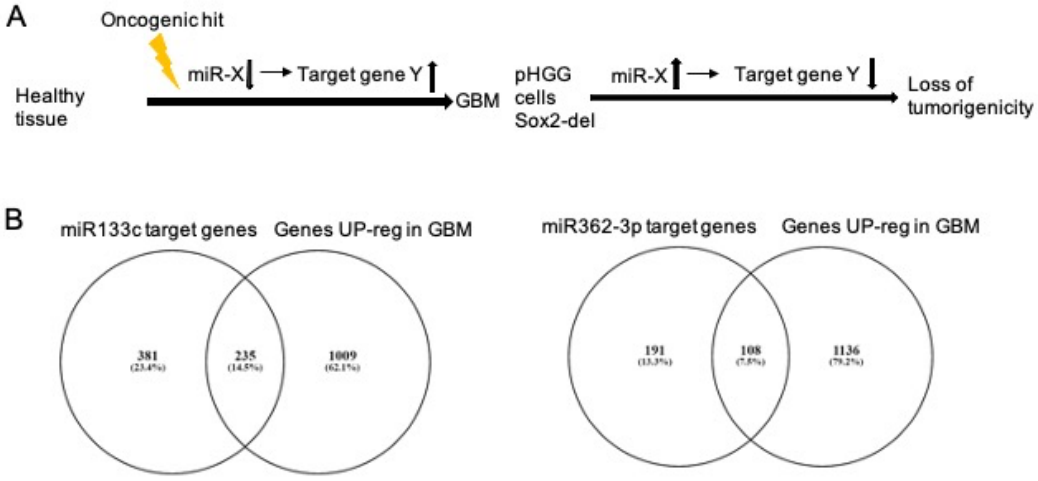


Figure1 Potential tumor suppressor role of up-regulated miRNA following Sox2 loss in pHGG cells. A. Rationale scheme. B. Overlap between the number of upregulated miRNA target genes and genes upregulated in GBM samples comparing healthy tissue.

3.4 Towards the definition of therapeutic targets

As discussed above, we found four genes (Ebf, Hey2, Cdkn2b, Zfp423) that participate in regulatory circuits with tumor suppressive capabilities. Therefore, a possible mode of action of Sox2 might be to repress, in gliomas, genes involved in the suppression of tumorigenesis. In the case of Cdkn2b, a gene very often deleted in GBM, Sox2 action mimics the effect of Cdkn2b deletion since it represses the activity of Cdkn2b.

Our observation that Sox2 inhibits several genes which act as tumor suppressors points to the exciting possibility of developing drugs able to prevent the ability of Sox2 (or interacting proteins, such as Groucho) to repress these genes.

Recent clinical trials ranging from phase 1 to phase 3 (clinicaltrials.gov) are testing the drug Palbociclib to see if it blocks cancer progression. Palbociclib is an inhibitor of CDK4/6, it thus mimics Cdkn2b activity. This inhibitor could be used for those type of tumors that have deletions in the Cdkn2b locus. In our model Sox2 represses Cdkn2b (also called Ink4b) and is important for the self-renewal of CSCs. It could be interesting to see if this inhibitor, Palbociclib, could suppress CSCs ability to self-renew. Another possible therapeutic approach could be to deliver the products of the identified anti-oncogenes (e.g. the mRNA) to tumor cells by carriers (e.g. nanoparticles) that reach the tumor-reinitiating cells(Haas et al., 2017).

Although many studies consider Sox2 itself as a target for therapy approaches (Hüser et al., 2018), the nuclear localization of SOX2 makes it a difficult target for efficient pharmacological recognition. Overall, the identification of multiple downstream Sox2 targets, impacting on various signaling pathways, representing important mediators of Sox2 function, may hopefully contribute to the design of specific, multi-hit therapy approaches.

References

Avilion, A. A., Nicolis, S. K., Pevny, L. H., Perez, L., Vivian, N., & Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes & development, 17*(1), 126-140. doi:10.1101/gad.224503

Bertolini, J. A., Favaro, R., Zhu, Y., Pagin, M., Ngan, C. Y., Wong, C. H., . . . Wei, C. L. (2019). Mapping the Global Chromatin Connectivity Network for Sox2 Function in Neural Stem Cell Maintenance. *Cell stem cell, 24*(3), 462-476 e466. doi:10.1016/j.stem.2019.02.004

Bonnet, D., & Dick, J. E. (1997). Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat Med, 3*(7), 730-737. doi:10.1038/nm0797-730

Boyer, L. A., Lee, T. I., Cole, M. F., Johnstone, S. E., Levine, S. S., Zucker, J. P., . . . Young, R. A. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell, 122*(6), 947-956. doi:10.1016/j.cell.2005.08.020

Cox, J. L., Mallanna, S. K., Luo, X., & Rizzino, A. (2010). Sox2 uses multiple domains to associate with proteins present in Sox2-protein complexes. *PLoS One, 5*(11), e15486. doi:10.1371/journal.pone.0015486

Denker, A., & de Laat, W. (2015). A Long-Distance Chromatin Affair. *Cell, 162*(5), 942-943. doi:10.1016/j.cell.2015.08.022

Doan, R. N., Bae, B. I., Cubelos, B., Chang, C., Hossain, A. A., Al-Saad, S., . . . Walsh, C. A. (2016). Mutations in Human Accelerated Regions Disrupt Cognition and Social Behavior. *Cell, 167*(2), 341-354 e312. doi:10.1016/j.cell.2016.08.071

Fang, X., Yoon, J. G., Li, L., Yu, W., Shao, J., Hua, D., . . . Lin, B. (2011). The SOX2 response program in glioblastoma multiforme: an integrated ChIP-

seq, expression microarray, and microRNA analysis. *BMC Genomics, 12*, 11. doi:10.1186/1471-2164-12-11

Fantes, J., Ragge, N. K., Lynch, S. A., McGill, N. I., Collin, J. R., Howard-Peebles, P. N., . . . FitzPatrick, D. R. (2003). Mutations in SOX2 cause anophthalmia. *Nature genetics, 33*(4), 461-463. doi:10.1038/ng1120

Favaro, R., Appolloni, I., Pellegatta, S., Sanga, A. B., Pagella, P., Gambini, E., . . . Nicolis, S. K. (2014). Sox2 is required to maintain cancer stem cells in a mouse model of high-grade oligodendroglioma. *Cancer research, 74*(6), 1833-1844. doi:10.1158/0008-5472.CAN-13-1942

Favaro, R., Valotta, M., Ferri, A. L., Latorre, E., Mariani, J., Giachino, C., . . . Nicolis, S. K. (2009). Hippocampal development and neural stem cell maintenance require Sox2-dependent regulation of Shh. *Nature neuroscience, 12*(10), 1248-1256. doi:10.1038/nn.2397

Ferri, A., Favaro, R., Beccari, L., Bertolini, J., Mercurio, S., Nieto-Lopez, F., . . . Nicolis, S. K. (2013). Sox2 is required for embryonic development of the ventral telencephalon through the activation of the ventral determinants Nkx2.1 and Shh. *Development, 140*(6), 1250-1261. doi:10.1242/dev.073411

Ferri, A. L., Cavallaro, M., Braida, D., Di Cristofano, A., Canta, A., Vezzani, A., . . . Nicolis, S. K. (2004). Sox2 deficiency causes neurodegeneration and impaired neurogenesis in the adult mouse brain. *Development, 131*(15), 3805-3819. doi:10.1242/dev.01204

Garel, S., Marin, F., Mattei, M. G., Vesque, C., Vincent, A., & Charnay, P. (1997). Family of Ebf/Olf-1-related genes potentially involved in neuronal differentiation and regional specification in the central nervous system. *Dev Dyn, 210*(3), 191-205. doi:10.1002/(SICI)1097-0177(199711)210:3<191::AID-AJA1>3.0.CO;2-B

Giachino, C., Boulay, J. L., Ivanek, R., Alvarado, A., Tostado, C., Lugert, S., . . . Taylor, V. (2015). A Tumor Suppressor Function for Notch Signaling in Forebrain Tumor Subtypes. *Cancer Cell, 28*(6), 730-742. doi:10.1016/j.ccell.2015.10.008

Haas, T. L., Sciuto, M. R., Brunetto, L., Valvo, C., Signore, M., Fiori, M. E., . . . De Maria, R. (2017). Integrin alpha7 Is a Functional Marker and Potential Therapeutic Target in Glioblastoma. *Cell stem cell, 21*(1), 35-50 e39. doi:10.1016/j.stem.2017.04.009

Huser, L., Novak, D., Umansky, V., Altevogt, P., & Utikal, J. (2018). Targeting SOX2 in anticancer therapy. *Expert Opin Ther Targets, 22*(12), 983-991. doi:10.1080/14728222.2018.1538359

Jeong, Y., Leskow, F. C., El-Jaick, K., Roessler, E., Muenke, M., Yocum, A., . . . Epstein, D. J. (2008). Regulation of a remote Shh forebrain enhancer by the Six3 homeoprotein. *Nature genetics, 40*(11), 1348-1353. doi:ng.230 [pii]

10.1038/ng.230

Kelberman, D., de Castro, S. C., Huang, S., Crolla, J. A., Palmer, R., Gregory, J. W., . . . Dattani, M. T. (2008). SOX2 plays a critical role in the pituitary, forebrain, and eye during human embryonic development. J *Clin Endocrinol Metab, 93(*5), 1865-1873. doi:jc.2007-2337 [pii]

10.1210/jc.2007-2337

Liao, D. (2009). Emerging roles of the EBF family of transcription factors in tumor suppression. Mo*l Cancer Res, 7(12*), 1893-1901. doi:10.1158/1541-7786.MCR-09-0229

Liu, Y. R., Laghari, Z. A., Novoa, C. A., Hughes, J., Webster, J. R., Goodwin, P. E., . . . Scotting, P. J. (2014). Sox2 acts as a transcriptional repressor in neural stem cells. BM*C Neurosci, 15,* 95. doi:10.1186/1471-2202-15-95

Mercurio, S., Serra, L., Motta, A., Gesuita, L., Sanchez-Arrones, L., Inverardi, F., . . . Nicolis, S. K. (2019). Sox2 Acts in Thalamic Neurons to Control the Development of Retina-Thalamus-Cortex Connectivity. iS*cience, 15,* 257-273. doi:10.1016/j.isci.2019.04.030

Nechanitzky, R., Akbas, D., Scherer, S., Gyory, I., Hoyler, T., Ramamoorthy, S., . . . Grosschedl, R. (2013). Transcription factor EBF1 is essential for the maintenance of B cell identity and prevention of alternative fates in committed cells. Na*t Immunol, 14(8*), 867-875. doi:10.1038/ni.2641

Ngan, C. Y., Wong, C. H., Tjong, H., Wang, W., Goldfeder, R. L., Choi, C., . . . Wei, C. L. (2020). Chromatin interaction analyses elucidate the roles of PRC2-bound silencers in mouse development. Na*ture genetics, 52(3*), 264-272. doi:10.1038/s41588-020-0581-x

Nord, A. S., Pattabiraman, K., Visel, A., & Rubenstein, J. L. R. (2015). Genomic perspectives of transcriptional regulation in forebrain development. Ne*uron, 85(1*), 27-47. doi:10.1016/j.neuron.2014.11.011

Pevny, L. H., & Nicolis, S. K. (2010). Sox2 roles in neural stem cells. Th*e international journal of biochemistry & cell biology, 42(3*), 421-424. doi:10.1016/j.biocel.2009.08.018

Ragge, N. K., Lorenz, B., Schneider, A., Bushby, K., de Sanctis, L., de Sanctis, U., . . . Fitzpatrick, D. R. (2005). SOX2 anophthalmia syndrome. Am *J Med Genet A, 135(1*), 1-7. doi:10.1002/ajmg.a.30642

Rao, S. K., Edwards, J., Joshi, A. D., Siu, I. M., & Riggins, G. J. (2010). A survey of glioblastoma genomic amplifications and deletions. J *Neurooncol, 96(2*), 169-179. doi:10.1007/s11060-009-9959-4

Reddy, K. B. (2015). MicroRNA (miRNA) in cancer. Ca*ncer Cell Int, 15,* 38. doi:10.1186/s12935-015-0185-1

Reya, T., Morrison, S. J., Clarke, M. F., & Weissman, I. L. (2001). Stem cells, cancer, and cancer stem cells. Na*ture, 414(6*859), 105-111. doi:10.1038/35102167

Schneider, A., Bardakjian, T., Reis, L. M., Tyler, R. C., & Semina, E. V. (2009). Novel SOX2 mutations and genotype-phenotype correlation in anophthalmia and microphthalmia. Am *J Med Genet A, 149A(12*), 2706-2715. doi:10.1002/ajmg.a.33098

Signaroldi, E., Laise, P., Cristofanon, S., Brancaccio, A., Reisoli, E., Atashpaz, S., . . . Testa, G. (2016). Polycomb dysregulation in gliomagenesis targets a Zfp423-dependent differentiation network. Na*t Commun, 7,* 10753. doi:10.1038/ncomms10753

Sisodiya, S. M., Ragge, N. K., Cavalleri, G. L., Hever, A., Lorenz, B., Schneider, A., . . . Fitzpatrick, D. R. (2006). Role of SOX2 mutations in human hippocampal malformations and epilepsy. Ep*ilepsia, 47(3*), 534-542. doi:10.1111/j.1528-1167.2006.00464.x

Suzuki, H., Aoki, K., Chiba, K., Sato, Y., Shiozawa, Y., Shiraishi, Y., . . . Ogawa, S. (2015). Mutational landscape and clonal architecture in grade II and III gliomas. Na*ture genetics, 47(5*), 458-468. doi:10.1038/ng.3273

Wang, J., Ma, Y., & Cooper, M. K. (2013). Cancer stem cells in glioma: challenges and opportunities. Tr*ansl Cancer Res, 2(5*), 429-441. doi:10.3978/j.issn.2218-676X.2013.08.01

Ying, M., Wang, S., Sang, Y., Sun, P., Lal, B., Goodwin, C. R., . . . Xia, S. (2011). Regulation of glioblastoma stem cells by retinoic acid: role for Notch pathway inhibition. On*cogene, 30(3*1), 3454-3467. doi:10.1038/onc.2011.58