A Proximal *Sox2* Enhancer Cluster is Required for the Anterior Regional Identity of Neural Progenitors

Ian C Tobias[1&], Sakthi D Moorthy[1,2&], Virlana M Shchuka[1], Lida Langroudi[1], Zoe E Gillespie[1], Andrew G Duncan[1], Ruxiao Tian[1], Mariia Cherednychenko[1], Natalia A Gajewska [1], Raphaël B Di Roberto[1], and Jennifer A Mitchell[1*]

1) Department of Cell and Systems Biology, University of Toronto, Toronto, Ontario, M5S3G5, Canada

2) Current Address: Center for Commercialization of Regenerative Medicine, Toronto, Ontario, M5G1M1, Canada

*Corresponding Author

E-mail: ja.mitchell@utoronto.ca (J.A.M.)

[&]These authors made equal contributions to this work

Abstract

Embryonic development depends on spatially and temporally orchestrated gene regulatory networks. Expressed in neural stem and progenitor cells (NSPCs), the transcription factor sex-determining region Y box 2 (Sox2) is critical for embryogenesis and stem cell maintenance in neural development. Whereas the enhancers of *Sox2* are well defined in early embryogenesis, little is known about *Sox2* gene regulation in the neural lineage. Using functional genome data, we identify an enhancer cluster region that regulates *Sox2* transcription in NSPCs derived from mouse embryonic stem cells (ESCs). By generating allelic mutants using CRISPR-Cas9 mediated deletions, we show that this proximal enhancer cluster, termed *Sox2* regulatory regions 2-18 (SRR2-18), controls *Sox2* allelic dosage during neural differentiation. Transcriptome analyses demonstrate that SRR2-18 is necessary for maintaining the anterior regional identity of ESC-generated NSPCs. Biallelic deletion of the *Sox2* neural enhancer cluster perturbs the regulation of neurodevelopmental and regionalization genes. Furthermore, homozygous NSPC deletants exhibit cell-type autonomous self-renewal defects and altered developmental potency. Altogether, our data define a *cis*-regulatory enhancer cluster controlling *Sox2* in the neural lineage and highlight how the allelic *Sox2* dosage is critical for the anterior-posterior regionalization of NSPCs.

Introduction

Specialized cell types within the nervous system are generated during embryonic development from populations of self-renewing stem and progenitor cells that emerge from defined neurogenic zones (Johe et al. 1996; Molofsky et al. 2003). These neural stem and progenitor cells (NSPCs) differentiate into interconnected networks of neuronal as well as glial cell types, which include astrocytes and oligodendrocytes (Götz and Barde 2005; Shen et al. 2006). More than 80 different neuronal and glial cell types corresponding to developmental stage, regional specificity, and cellular function have been identified by single cell gene expression patterns in the mouse postnatal brain and spinal cord (Rosenberg et al. 2018; Zeisel et al. 2018). To execute the complex behaviours of the vertebrate nervous system, regionally distinct gene expression profiles instruct NSPC populations to organize specialized cranial and spinal cell types along the rostral-caudal and dorsal-ventral axes (Mathis and François Nicolas 2000; Hagey et al. 2016).

The proliferative activity, developmental potency, and region-specific identity of NSPCs are controlled by genetic programs defined by differential regulatory element usage (Ensini et al. 1998; Ziller et al. 2015). Cell fate decisions involve extensive transcriptional changes that are governed by gene regulatory networks — systems of DNA sequences (such as enhancers) responsive to transcription factors (TFs) (Brandenberger et al. 2004; Bailey et al. 2006; Loh et al. 2006). Enhancers are orientation-independent regulatory elements that influence when and where genes become transcriptionally active (Banerji et al. 1981). Enhancers can be positioned upstream or downstream and at variable genomic distances from the genes they regulate (Tuan et al. 1989; Sanyal et al. 2012). Depending on their genomic location, enhancers have been shown to contribute to transcriptional regulation through the formation of chromatin-chromatin interactions (Carter et al. 2002; Tolhuis et al. 2002; Zhang et al. 2013). Distal enhancer elements can be guided to their target promoters by architectural elements, either through the combined action of the zinc finger protein CCCTC-binding factor (CTCF) and the ring-shaped Cohesin complex, or by CTCF-independent mechanisms (Tang et al. 2015; Weintraub et al. 2017; Taylor et al. 2022).

Enhancers exhibit a variety of context-dependent chromatin states, which are often inferred from DNA methylation and post-translational histone modifications (ENCODE Project Consortium 2012). Additionally, enhancers appear to be more dynamic than promoter elements across various cell types, tissues, and taxonomic groups (Heintzman et al. 2009; Roadmap Epigenomics Consortium et al. 2015; Villar et al. 2015). By leveraging our knowledge of epigenetically-informed chromatin states (e.g., Histone H3 lysine 27 acetylation [H3K27Ac] in the context of enhancers), TF occupancy and DNA accessibility, and co-activator recruitment (e.g., Mediator complex, EP300) (Creyghton et al. 2010; Kagey et al. 2010; Rada-Iglesias et al. 2011), we can make predictions about enhancer activity. However, a single gene can be regulated by multiple enhancers or redundant enhancers in different cellular contexts (Hay et al. 2016; Moorthy et al. 2017; Osterwalder et al. 2018; Taylor et al. 2022). As a result, identifying the locations of enhancers and determining their precise roles within a specific cellular context has been challenging (reviewed in (Tobias et al. 2021)). In models that explore cell differentiation processes, coordinated changes in TF binding occur at enhancer regions, which directly impact gene expression and restrict cellular developmental outcomes (Wamstad et al. 2012; Kieffer-Kwon et al. 2013; Huang et al. 2016).

Sex-determining region Y box 2 (*Sox2*) encodes a TF that is initially expressed in the inner cell mass and epiblast of preimplantation embryos (Gubbay et al. 1990; Avilion et al. 2003). *Sox2* expression is downregulated upon pluripotent cell commitment towards endodermal or mesodermal lineages; however, its transcription is maintained within the neuroectodermal lineage, where SOX2 inhibits mesendoderm fate determinants (Ivanova et al. 2006; Kopp et al. 2008; Thomson et al. 2011). *Sox2* represents one of the SOXB1 family transcription factors (*Sox1-3*) with partially overlapping functions, which are broadly co-expressed in the neural ectoderm and the anterior gut endoderm (Taranova et al. 2006; Que et al. 2007; Cavallaro et al. 2008). SOX2 collaborates with other TFs, including *Pou5f1* and *Pax6*, to ensure proper regulation of transcriptional processes that are essential for maintaining tissue-specific identities (Masui et al. 2007; Chen et al. 2008; Lodato et al. 2013; Quevedo et al. 2019).

The deletion of *Sox2* in the E10.5-12.5 forebrain produces viable mice with mild neurodevelopmental impairments at birth (Miyagi et al. 2008). Conversely, *Sox2* is essential for neurogenesis in the adult mouse hippocampal niche, which exhibits a Sonic hedgehog (Shh) signaling-dependent depletion of NSPCs in the absence of *Sox2* (Favaro et al. 2009). *In vivo*, NSPCs display regional and temporal heterogeneity, with both quiescent and amplifying progenitor cell populations that retain self-renewal and differentiation capabilities when cultured *in vitro* (Reynolds and Weiss 1992; Garcia et al. 2004; Suh et al. 2007). Diffusible factors can induce NSPCs to shift positional neural cell identities along the anterior-posterior or ventral-dorsal axes; however, the primary allocation of embryonic cells to cranial and spinal progenitors occurs in the vertebrate epiblast prior to neural induction (Bertacchi et al. 2013; Hallmann et al. 2016; Metzis et al. 2018). These forms of regionalized gene regulation in neural progenitor populations are essential to prevent the specification of inappropriate or mixed neural cell identities.

Sox2 and its genic flanking sequences fall within a synteny block on chromosome 3 (q26.33) in placental mammals, which may suggest evolutionary selection for functional relationships (Zhang et al. 2022). Two enhancers located proximally to the *Sox2* transcriptional start site (TSS), known as *Sox2* regulatory regions (SRR) 1 and 2, have been shown to independently drive reporter gene expression in both ESCs and NSPCs (Zappone et al. 2000; Tomioka et al. 2002; Catena et al. 2004; Miyagi et al. 2004). Biallelic deletion of SRR1 in mouse embryos has been shown to induce a transient *Sox2* expression deficit in the anterior neural plate (Iwafuchi-Doi et al. 2011). *Sox2* haplodeficient mice with a monoallelic deletion of SRR1 on the intact *Sox2* allele display cerebral malformations and a decrease in Nestin (*Nes*)-positive hippocampal NSPCs (Ferri et al. 2004). Conversely, SRR2 is thought to recruit transcriptional repressors and decrease *Sox2* expression during *in vitro* differentiation of embryonic and neural stem cells (Li et al. 2012; Marqués-Torrejón et al. 2013). Another *Sox2* regulatory element homologous to the N1 enhancer in chickens is first activated in the caudal lateral epiblast of mouse embryos and later controls the fate of *Sox2*-expressing neuromesodermal progenitor cells (Takemoto et al. 2006, 2011).

Multiple regulatory elements with regionally delineated activities are posited to coordinate *Sox2* transcription in diverse neurogenic contexts where SOX2 is required (Uchikawa et al. 2003; Uchikawa and Kondoh 2016). However, there still exists a significant gap in understanding between the identification of *Sox2* regulatory elements and the mechanisms governing their tissue-specific actions. The ESC-to-NSPC differentiation model provides a genetically tractable system for studying how epigenetic remodeling and chromatin organization influence neural gene regulatory networks (Phillips-Cremins et al. 2013). We have shown that the *Sox2* control region (SCR), a distal cluster of enhancers located at position 104-111 kb downstream of the *Sox2* TSS, regulates *Sox2* expression in mouse ESCs (Chen et al. 2012; Zhou et al. 2014). Interestingly, proximal enhancers of *Sox2* are not necessary for the maintenance of *Sox2* expression in mouse ESCs (Zhou et al. 2014). As pluripotent cells differentiate, TF binding is redistributed genome-wide, and the chromatin state is altered at the SCR. This involves a loss of H3K27ac enrichment and long-range interactions with the *Sox2* promoter (Beagan et al. 2016; Thompson et al. 2022; Chakraborty et al. 2023). Reducing the dosage of *Sox2* in $SCR^{-/-}$ ESCs delays but does not abolish pan-neural progenitor gene activation during spontaneous differentiation *in vitro* (Zhou et al. 2014). The *Sox2* locus has been extensively used in various studies that investigated long-range gene regulation in pluripotent stem cells. However, a more comprehensive understanding of how regulatory elements contribute to *Sox2* gene regulation and resultant early neural lineage commitment is still required (Brosh et al. 2023; Chakraborty et al. 2023; Platania et al. 2023). To gain more insight into the *Sox2* regulatory mechanisms in neural cells, we investigated the contribution of candidate neural lineage *Sox2* enhancers using mouse ESC-derived NSPCs and CRISPR/Cas9-mediated genome editing models.

Results

Sox2 flanking regions are responsible for enhancement of *Sox2* transcription during neural stem/progenitor cell differentiation.

SRR1 and SRR2 exhibit regulatory activity in the mouse epiblast and in ESCs, suggesting their potential role in *Sox2* gene regulation early in development (Iwafuchi-Doi et al. 2011). However, neither of these enhancers produce an obvious neurodevelopmental phenotype when both allelic copies are deleted in genetically engineered mouse models (Ferri et al. 2004; Moncho-Amor et al. 2021). Examining the coactivator-associated modification H3K27ac in mouse ESCs and E14.5 forebrain tissue, we reanalyzed chromatin immunoprecipitation with sequencing (ChIP-seq) data from the ENCODE consortium to identify additional enhancer candidates active during early neural development (**Fig. 1A**). To refine our search to regions contributing to RNA polymerase II (RNAPII)-mediated transcription, we also reanalyzed Mediator subunit MED1 ChIP-seq data from *in vitro* generated NSPCs (Quevedo et al. 2019).

Examination of these H3K27ac ChIP-seq profiles showed that both ESCs and NSPCs retain distinct regions of H3K27ac enrichment near the *Sox2* gene, extending from the upstream flanking region, through SRR1, and into a 16.7 kb region downstream of the *Sox2* TSS containing SRR2 (SRR2-18) (**Fig. 1A**). MED1 chromatin recruitment signals were largely restricted to a 10 kb region immediately downstream of *Sox2* in cultured neural progenitors (**Fig. 1A**). We also inspected publicly available H3K27ac ChIP-seq profiles generated in E12.5 ganglionic eminences (Rhodes et al. 2022), neural progenitor-rich tissues that produce neurons in the cerebral cortex. Our analysis verified that these *Sox2*-flanking regions display high levels of H3K27ac specifically in murine neurogenic tissues (**Supplemental Fig. S1A**).

Since H3K27ac- and MED1-enriched chromatin profiles are often predictive of active enhancers (Chen et al. 2012), we asked whether SRR1 and SRR2-18 play a role in coordinating *Sox2* expression during neural lineage commitment. To investigate the temporal dynamics of gene expression during neural differentiation, we modeled neural induction and NSPC generation using mouse hybrid F1 (*M. musculus*/129 × *M. castaneus*) ESC cultures (**Fig. 1B**). We used RT-qPCR to measure the levels of *Sox2*, *Pou5f1*, *Pax6* and *Nes* RNA over the differentiation time course. We observed a gradual decrease in *Pou5f1* transcript levels as mouse ESCs transitioned from states of naïve pluripotency to

the neural lineage in neural stem expansion (NSE) medium (P = 1.53 x 10^{-2}, Wilcoxon Test; **Supplemental Fig. S1B**). Conversely, *Pax6* expression was significantly increased by day 5 of neural differentiation (P = 3.42 x 10^{-5}, Welch's T Test; **Supplemental Fig. S1C**), consistent with its role as a neural fate determinant (Suter et al. 2009). Expression of NSPC marker *Nes* steadily increased with culturing in NSE medium (P = 1.21 x 10^{-5}, Welch's T Test; **Supplemental Fig. S1D**), indicating successful maintenance of stem and progenitor cell-rich populations (Lendahl et al. 1990). In general, *Sox2* RNA levels showed more developmental stage-dependent fluctuations and peaked at 3 days of neural differentiation (P = 2.47 x 10^{-2}, Welch's T Test; **Supplemental Fig. S1E**). This observation aligns with the notion that ESCs undergo sequential differentiation processes to acquire NSPC identity in similar differentiation protocols (Hitoshi et al. 2004; Iwafuchi-Doi et al. 2012). Immunofluorescence imaging confirmed that ESC-derived NSPCs co-express both SOX2 and NES (**Fig. 1C**).

We previously used CRISPR/Cas9-mediated deletion analyses in hybrid mouse ESCs to establish two different enhancer-deleted cell lines: one lacking both allelic copies of the SCR ($\Delta SCR^{129/Cast}$), and the other harboring monoallelic compound deletions of candidate enhancers located within 18 kb of the *Sox2* TSS ($\Delta SRR1,2\text{-}18^{Cast}$) (Zhou et al. 2014). The latter deletion encompasses most of the neural H3K27ac domain flanking the *Sox2* gene. To monitor *Sox2* expression from each allele and across a range of *Sox2* transcription levels, we used allele-specific RT-qPCR to determine the proportion of the sum *Sox2* transcript quantity detected from the 129 or Cast allele, as we have done previously (Moorthy et al. 2017). Allele-specific *Sox2* RT-qPCR showed the expected depletion of *Sox2* RNA from both the 129 and Cast alleles in $\Delta SCR^{129/Cast}$ ESCs prior to neural differentiation (P = 2.51 x 10^{-2}, Wilcoxon Test; **Fig. 1D**). However, $\Delta SCR^{129/Cast}$ cells showed a significant increase in *Sox2* expression during neural differentiation (P = 3.73 x 10^{-2}, Wilcoxon Test; **Fig. 1D**), which was not significantly different from the parent line at differentiation day 7 or in early-passage NSPC cultures (P =0.46, Wilcoxon Test; **Fig. 1D**). We did not observe any change in Cast *Sox2* transcript levels in $\Delta SRR1,2\text{-}18^{Cast}$ cells versus in undifferentiated mouse ESCs or during the first two days of neural differentiation (P = 0.46 and P = 0.88, respectively, Wilcoxon Tests; **Fig. 1D**). This is consistent with our previous

finding that *Sox2* proximal elements are dispensable for *Sox2* expression in pluripotent mouse ESCs (Zhou et al. 2014). Interestingly, ΔSRR1,2-18^{Cast} cells showed significantly decreased Cast-derived *Sox2* RNA in early-passage NSPCs compared to that of parent NSPCs (P = 2.86 x 10^{-2}, Wilcoxon Test; **Fig. 1D**). These findings indicate that the distal SCR is dispensable for *Sox2* expression in NSPCs established from ESCs *in vitro*. Instead, gene proximal elements within the SRR1 and/or SRR2-18 regions contribute to ESC-derived NSPC *Sox2* expression.

We next asked whether SRR2 or additional elements within the downstream flanking region are required for *Sox2* transcription in NSPCs. To this end, we generated clonal populations of ESCs harboring a large deletion of the ~100 kb region spanning SRR2 and SRR104 (ΔSRR2-104^{Cast}), a mutant generated to remove all the possible proximal downstream enhancers (**Fig. 1E**). We also established NSPCs from mutant ESCs with a monoallelic deletion of SRR2 alone (ΔSRR2^{Cast}) and assessed allele-specific *Sox2* transcript levels. This allowed us to distinguish between the contribution of SRR2, an element with reported activity in epiblast stem cell-derived neural progenitors (Iwafuchi-Doi et al. 2011), and other elements within the SRR2-104 region. Interestingly, NSPC clones with either ΔSRR2^{Cast} or ΔSRR2-104^{Cast} genotypes showed a significant decrease in *Sox2* transcription from the targeted allele (P = 1.39 x 10^{-2} and P = 4.46 x 10^{-4}, respectively, Tukey's Test; **Fig. 1E**). Additionally, the proportion of Cast-derived *Sox2* transcript was further decreased in ΔSRR2-104^{Cast} cells compared to that in ΔSRR2^{Cast} NSPCs (P = 6.31 x 10^{-3}, Tukey's Test). Together, these data demonstrate that one or more regions downstream of SRR2 contribute additively to *Sox2* expression in ESC-derived NSPCs.

***Sox2* transcription is maintained in NSPCs despite deletion of SRR1 and flanking CTCF-bound regions**

The architectural domain harboring *Sox2* in pluripotent and neural cells is delimited on the centromeric end by three CTCF-bound regions directly upstream of the *Sox2* promoter. These sites bookend SRR1, the anterior neural plate enhancer predicted to contribute to *Sox2* expression in the neuroectoderm

(Iwafuchi-Doi et al. 2011). To scrutinize the protein-binding elements with potential architectural and regulatory roles upstream of *Sox2*, we used ENCODE ChIP-seq data that profile CTCF binding in E14.5 mouse forebrain tissue. We also reanalyzed publicly available ChIP-seq data for Cohesin complex subunit structural maintenance of chromosomes 1 (SMC1) retention in isolated NSPCs. Each CTCF binding site showed SMC1 accumulation in NSPCs (**Fig. 2A**), raising the possibility that local chromatin topology contributes to *Sox2* expression in neural progenitors.

To determine whether the upstream SRR1 and the promoter proximal CTCF sites are required for *Sox2* expression in our ESC-derived NSPC model, we deleted upstream regions encompassing SRR1 in hybrid ESCs and differentiated genetically modified clones into NSPCs. Although these elements are found near alternative promoters and exons in the *Sox2* overlapping transcript (*Sox2-ot*) long non-coding RNA, previous *in vivo* deletions targeting *Sox2-ot* features that are conserved in vertebrates resulted in no discernible phenotype in mouse development (Li et al. 2021). As an external control, we confirmed the *cis*-linked decrease in *Sox2* allelic ratio in NSPCs established from ΔSRR1,2-18^{Cast} ESCs (P = 2.86 x 10^{-2}, Welch's T Test; **Fig. 2B**). Deletion of SRR1 on the Cast allele (ΔSRR1^{Cast}) did not significantly change the allelic ratio of *Sox2* transcripts compared to WT hybrid NSPCs (P = 0.77, Welch's T Test; **Fig. 2B**). Moreover, we expanded the boundary of the SRR1-deleted region to include all three upstream CTCF binding sites in mouse neural cells and brain tissue (ΔSRR1-3xCTCF^{Cast}). We did not observe any significant difference in *Sox2* allelic ratio in ΔSRR1-3xCTCF^{Cast} NSPCs (P = 0.90, Welch's T Test; **Fig. 2B**), which is consistent with our recent finding that disruption of *Sox2* promoter proximal CTCF sites in the E9.5-10.5 embryo does not affect SOX2 levels in the central nervous system (CNS) (Chakraborty et al. 2023). Notably, the SRR1-3xCTCF^{Cast} deletion (-11.5 kb to -2.3 kb of TSS) contained a conserved zinc finger protein motif upstream of the *Sox2* promoter (pSox2). This motif has been reported to mediate *Sox2* transcriptional activation (-2.7 kb of TSS) in primary NSPCs through arsenite-resistance protein 2 (ARS2) binding (Andreu-Agullo et al. 2012). We have found that genes transactivated by dispersed regulatory architectures can have redundant enhancers that are not required for maintenance of gene transcription (Moorthy et al. 2017). It remained a possibility that the

extended CpG island promoter region, alongside formally undefined regulatory elements, contributes to *Sox2* expression in neural progenitors. Yet after establishing NSPC clones with a larger deletion (-15.6 kb to -1.1 kb of TSS) that spanned all proximal CTCF binding sites and greater than half of the CpG island promoter region on the Cast allele (ΔSRR1 CTCF-pSox2^{Cast}), the fraction of *Sox2* transcript from the Cast allele did not significantly differ from that of control NSPCs (P = 0.79, Welch's T Test; **Fig. 2B**). In summary, our data suggest neither SRR1, the CTCF binding sites, the ARS2 bound region, nor the full-length CpG island promoter are required for *Sox2* transcription in neural lineage commitment.

SRR2-18 regulates *Sox2* allelic dosage and monoallelic deletion disrupts transcriptional regulatory programs in NSPCs

The ability of genomic sites or synthetic constructs to bind multiple TFs correlates positively with enhancer regulatory activity by gain-of-function (e.g. reporter-based assays) and loss-of-function (e.g. knock-out of regulatory elements) experiments in both cell line and animal models (Zinzen et al. 2009; Moorthy et al. 2017; Singh et al. 2021). To this end, we examined open access transcription factor ChIP-seq data in cultured neural progenitors or embryonic cerebral cortex tissue to locate core modules of the predicted enhancer cluster. Aside from the *Sox2* CpG island promoter, regions that were approximately +3.7 kb (SRR2), +9.8 kb (chick N-5 homologue; hereafter SRR10), +14 kb (chick N-1 homologue; hereafter SRR14), and +17.1 kb (hereafter SRR17) downstream of the *Sox2* TSS showed binding of TFs associated with neuroepithelial/radial glia identity (ASCL1, BRN2, PAX6, SOX2), stem cell homeostasis (FOXO3, JUN, MAX, SOX9, TCF3), and neurogenesis (FEZF2, NFI, OLIG2, SMAD3, SOX4, SOX21, TBR1) (**Fig. 3A; Supplemental Table S1**). Notably, we observed only weak evidence of multiple TF binding at SRR1 in neural cell and tissue ChIP-seq data (- 4.2 kb; chick N-2 homologue). We also discovered a neural multiple TF-bound region upstream (-7.8 kb; hereafter SRR-8). SRR-8 was deleted alongside SRR1 in ΔSRR1-3xCTCF^{Cast} and ΔSRR1 CTCF-pSox2^{Cast} cells (**Supplemental**

Fig. S2A), with neither deletion showing disrupted *Sox2* expression (**Fig. 2B**); hence, SRR-8 appears not to contribute to *Sox2* transcription in this context.

We performed luciferase reporter assays in ESCs or NSPCs to assess the sufficiency of each individual downstream region to induce transcriptional activation of a modified plasmid vector with the mouse *Sox2* promoter (-2209 to -347 bp of TSS; **Supplemental Fig. S2B**). Control experiments were carried out in ESCs or NSPCs with a set of previously validated enhancers, namely SRR107 (a component of the pluripotency-associated SCR) or the *Nes* intron 2 enhancer (Nes-I2). This allowed us to verify that cell type-selective enhancer activity is reproduced with an ectopic reporter approach (Josephson et al. 1998; Zhou et al. 2014). Indeed, enhancer activity from *Sox2* promoter vectors carrying SRR107 was significantly lower in NSPCs compared to that in ESCs ($P = 8.88 \times 10^{-4}$, Wilcoxon Test; **Supplemental Fig. S2C**); however, Nes-I2 reporter vectors showed significantly greater transcriptional output in NSPCs relative to ESCs ($P = 1.01 \times 10^{-2}$, Wilcoxon Test; **Supplemental Fig. S2C**). SRR2-driven luciferase reporter activity was not significantly different between ESC and NSPC transfectants ($P = 0.17$, Wilcoxon Test; **Supplemental Fig. S2B**). We observed a wide range of transcriptional outputs from the SRR14 reporter vector in both ESC and NSPC transfectants (**Fig. 3B**). The SRR14 region notably overlaps a CpG island, a CTCF binding site and an alternative promoter for the *Sox2-ot* long non-coding RNA (**Supplemental Fig. S2D**). Neither SRR10 nor SRR17 showed transcriptional enhancer activity in ESCs and instead exhibited a significant decrease in luciferase activity ($P = 1.14 \times 10^{-3}$ and $P = 1.56 \times 10^{-2}$, respectively, Wilcoxon Tests; **Fig. 3B**). NSPCs transfected with reporter plasmids carrying SRR10 or SRR17 had significantly increased reporter activity compared to cells with the *Sox2* promoter plasmid alone ($P = 1.64 \times 10^{-2}$ and $P = 3.14 \times 10^{-2}$, respectively, Wilcoxon Tests; **Fig. 3B**). Collectively, these observations suggest SRR2-18 is comprised of multiple discrete enhancers that can increase transcriptional output in NSPCs.

In vivo reporter assays have shown that different *Sox2* proximal enhancers can transactivate reporter genes with distinct spatial domains in the developing vertebrate CNS (Uchikawa et al. 2003). However, enhancer loss-of-function approaches in a native chromatin context have yet to be completed to define

the enhancers required to support *Sox2* transcription in NSPC populations. To test the role of this 16.7 kb region in neural lineage specification from pluripotent ESCs, we decoupled *Sox2* expression from SRR2-18 *cis*-regulation by CRISPR-Cas9 mediated deletion. To facilitate the isolation of enhancer-deleted NSPCs for extensive clonal deletion analyses, we generated a reporter line with an mCherry insertion following the *Sox2* coding sequence and the cleavable P2A peptide on the *Mus musculus*^{129} allele. Independent derivations of NSPC cultures from the parent *Sox2*-mCherry^{129} ESC line displayed adherent cell morphologies consistent with undifferentiated neuroectodermal and progenitor cells (Okabe et al. 1996) (**Supplemental Fig. S3A**). We routinely confirmed homogeneous mCherry expression in ESC-derived *Sox2*-mCherry^{129} NSPCs and all sub-clones by flow cytometry (**Supplemental Fig. S3B**).

From the *Sox2*-mCherry^{129} line, we produced monoallelic deletions of SRR2-18 to assess the significance of this proximal enhancer cluster for *Sox2* expression in NSPCs. Allele-specific RT-qPCR analysis showed ΔSRR2-18^{129} NSPCs had a significantly increased Cast *Sox2* allelic ratio compared to that of non-deletion counterparts ($P = 5.50 \times 10^{-3}$, Welch's T Test; **Fig. 4A**). Reciprocally, ΔSRR2-18^{Cast} NSPCs exhibited a significant decrease in the fraction of *Sox2* transcript from the Cast allele compared to that in *Sox2*-mCherry^{129} NSPCs ($P = 2.11 \times 10^{-4}$, Welch's T Test; **Fig. 4A**). These data suggest that the SRR2-18 region functions as a *cis*-regulator of *Sox2* in the neural lineage. Accordingly, we performed ribosomal RNA-depleted total RNA-sequencing (RNA-seq) for transcriptome analysis. Comparison of normalized RNA-seq data indicated that *Sox2*-mCherry^{129} NSPCs at different passage numbers uncovered high dataset concordance (Pearson Correlation > 0.8 on raw counts and > 0.98 on log-transformed counts, **Fig. 4B**). Furthermore, reanalysis of bulk or transcription factor expression-sorted neural progenitors followed by hierarchical clustering analysis (HCA) revealed that *Sox2*-mCherry^{129} NSPCs group with *in vitro* derived *Sox1*$^{+}$ neural progenitor cells (SOX1+ NPC) and *Hes5*$^{+}$ telencephalic radial glia in an active Notch signaling state (HES5+ RG) (Bonev et al. 2017) (**Fig. 4B**). Importantly, NSPCs generated from hybrid ES cells were dissimilar and agglomerated separately from

bulk hybrid ESCs in the undifferentiated (ES F1) and ES-derived neural rosette states (ES-NEcto) (**Fig. 4B**).

We next sought to confirm whether the monoallelic deletion of SRR2-18 specifically disturbs the *Sox2* allelic dosage and no other *cis*-regulated gene. To account for mapping bias to reference mouse alleles, we assembled a SNP-substituted mouse genome using the 129S1/SvImJ sub-strain as a reference and N-masked discriminatory CAST/EiJ sub-strain SNPs (**Supplemental Fig. S3C**). We resolved the relative allelic expression at each locus independently of total gene expression to infer *cis* or *trans* modes of gene regulation. Profiles of allele-sorted RNA-seq reads showed that monoallelic deletion of SRR2-18 decreased steady state *Sox2* transcription from the targeted allele (**Fig. 4C**). Differential allele-resolved expression analysis in monoallelic deletants identified *Sox2* as the only chromosome 3 gene regulated by the SRR2-18 cluster (Cast/Total ratio log2FC > |0.5|; $Q < 0.05$) (**Supplemental Fig. S3D**). This analysis also highlighted that imbalances in allelic dosage between the 129 and Cast haplotypes are pervasive across a wide range of gene expression levels in hybrid NSPCs.

Germline *Sox2* hypomorphism results in viable mice that present with posteromedial cerebral malformations and disrupted patterning of the anterior diencephalon (hypothalamus, pituitary, optic cup). These models suggest that decreases in *Sox2* dosage below a certain threshold can negatively impact neural development (Ferri et al. 2004; Kelberman et al. 2006; Langer et al. 2012). To provide insight into the role of SRR2-18 as a *cis*-regulator of *Sox2* expression, in global gene regulation and in the establishment of neural progenitor identity, we analyzed total RNA-seq data in a manner agnostic to the hybrid cell haplotypes in SRR2-18 heterozygous NSPCs. Differential expression analysis revealed moderate alterations to the SRR2-18$^{+/-}$ NSPC transcriptome, which included 192 upregulated and 37 downregulated RNAs (log2FC > |0.5|; $Q < 0.01$) (**Fig. 4D**). Monoallelic SRR2-18 deletion induced the expression of the *Hoxb* locus genes that are associated with caudal patterning of the neural tube (*Hoxb6-9*), ventral patterning of the neural tube (*Sim2*, *Evc2*), and transmembrane proteins involved in cell migration and morphogenesis (*Icam1*, *Dnah5*, *Tmem179*) (**Fig. 4D-E; Supplemental Table S2**).

We detected greater steady-state *Shh* levels in SRR2-18 heterozygous NSPCs, suggesting that enhancer cluster deletion can promote a tissue organizer response similar to that of the ventral progenitors and floor plate cells of the neural tube (Wijgerde et al. 2002). Gene ontology (GO) analysis revealed terms with top enrichment scores (ES) among the upregulated gene set in heterozygous ΔSRR2-18 NSPCs, including, but not limited to, anterior/posterior pattern specification (*ES* = 0.69, *Q* = 3.1 x 10^{-3}) and regionalization (*ES* = 0.61, *Q* = 7.5 x 10^{-3}; **Supplemental Fig. S4A**; **Supplemental Table S3**). In summary, these data show that SRR2-18 regulates *Sox2* expression in a *cis*-acting mechanism. Moreover, hybrid NSPCs are sensitive to allelic differences in transcriptional *Sox2* dosage, whereby SRR2-18 monoallelic deletants activate neural tube patterning pathways under steady-state culture conditions.

Biallelic deletion of SRR2-18 in NSPCs disrupts global gene regulation

To investigate the contribution of transcriptional dosage compensation events attributable to the SRR2-18 region remaining on the intact allele, we derived clonal lines lacking the SRR2-18 cluster from both alleles (ΔSRR2-$18^{129/Cast}$). Established ΔSRR2-$18^{129/Cast}$ NSPCs do not exhibit imbalanced *Sox2* allelic dosage and maintain nuclear SOX2 immuno-reactivity (**Supplemental Fig. S4B-C**). However, ΔSRR2-$18^{129/Cast}$ NSPCs have distinctive colony and individual cell morphotypes (**Fig. 5A**). In maintenance culture conditions, ΔSRR2-$18^{129/Cast}$ NSPCs initially show no difference in exponential-phase growth rate but adopt lower rates of cell division with prolonged time in culture (**Fig. 5B**). Prior to the onset of proliferative defects, ΔSRR2-$18^{129/Cast}$ NSPCs had greater transcriptomic deviation from parent NSPCs than did the monoallelic series of SRR2-18 deletions (**Fig. 5C**). NSPCs established from homozygous SRR2-18 deleted-ESCs group further away from the parent control line in principal component space. We also compared our NSPCs decoupled from SRR2-18 mediated regulation to aggregates of self-renewing primary NSPCs (commonly referred to as neurospheres) established from wild-type and *Sox2*-ablated mice from reanalyzed open access RNA-seq data (Bertolini et al. 2019). The

transcriptome of enhancer-deleted NSPCs markedly deviated from *Sox2*-knockout (KO) neurospheres that were cultured for fewer than 24 days after a conditional *Sox2* deletion (**Fig. 5C**). These findings are suggestive of longer-term cascading regulatory adaptations to the absence of SRR2-18 during neural differentiation.

To identify transcript-level features that distinguish ΔSRR2-18$^{129/Cast}$ NSPCs from the parent line, we performed differential gene expression analysis. A total of 104 transcripts showed a significant decrease in overall abundance and 277 transcripts showed a significant increase in biallelic SRR2-18 deletants (log2FC > |0.5|; $Q < 0.01$, Wald test under negative binomial model) (**Fig. 5D; Supplemental Table S4**). Reanalysis of RNA-seq data from primary neurospheres recovered following *Sox2* ablation revealed that the acute transcriptional response to *Sox2* loss-of-function resulted in only 46 differentially expressed genes (35 decreased expression; 11 increased expression; log2FC > |0.5|; $Q < 0.01$, Wald test under negative binomial model) (**Supplemental Fig. S4E; Supplemental Table S5**). We found no overlap in the set of genes activated in SRR2-18-deleted NSPCs and *Sox2*-KO neuropheres, suggesting that neural enhancer-mediated *Sox2* dosage influences a broader set of biological processes than has been previously appreciated. Gene ontology (GO) analysis revealed terms enriched among the downregulated gene set in ΔSRR2-18$^{129/Cast}$ NSPCs, including, but not limited to, nucleobase biosynthesis ($ES = -0.41$, $Q = 2.8 \times 10^{-2}$) and CNS neuron differentiation ($ES = -0.52$, $Q = 3.0 \times 10^{-2}$, one-sided Fisher's Tests) (**Supplemental Fig. S4F; Supplemental Table S6**). Reciprocally, analysis of the induced gene set suggested that tube development ($ES = 0.50$, $Q = 2.8 \times 10^{-3}$) and positive regulation of MAPK cascade ($ES = 0.57$, $Q = 1.3 \times 10^{-2}$, one-sided Fisher's Tests) were among the top enriched biological processes in ΔSRR2-18$^{129/Cast}$ NSPCs (**Fig. 5E**). These data indicated that biallelic SRR2-18-deleted cells showed dysregulated nucleobase metabolism and MAPK pathways as well as altered neurodevelopmental signatures.

To assess whether biallelic deletion of SRR2-18 perturbs further differentiation into post-mitotic CNS cell types, we grew NSPCs in medium without exogenous epidermal growth factor (EGF) to facilitate exit from a self-renewing state. We introduced a defined serum-containing supplement for ten days to

promote mixed lineage differentiation into neurons and glial cells. Total RNA-sequencing was performed on bulk progenies differentiated from ΔSRR2-18$^{129/Cast}$ or parent NSPCs (**Fig. 5F**). As anticipated, the progenies differentiated from Sox2-mCherry^{129} NSPC progeny showed decreased expression of cell cycle process genes (*Mki67*, *E2f2*), and induced expression of signals and determinants of astrocyte (*Cntf*, *Sox9*), oligodendrocyte (*Daam2*, *Cxcr4*), and neuronal (*Neurog2*, *Brinp1*) maturation (**Supplemental Fig. S5A**; **Supplemental Table S7**). However, undirected differentiation of Sox2-mCherry^{129} NSPCs or ΔSRR2-18$^{129/Cast}$ NSPCs produced post-mitotic derivatives with distinctive morphologies (**Supplemental Fig. S5B**). The mixed progenies of biallelic SRR2-18 deleted cells exhibited pronounced differences in the number of differentially expressed genes (1235 increased expression; 1133 decreased expression; log2FC > |0.5|; $Q < 0.01$, Wald test under negative binomial model) compared to bulk derivatives of parent NSPCs (**Supplemental Fig. S5C, Supplemental Table S8**). We observed a marked depletion of transcripts associated with biological processes involved in brain development, including, but not limited to, forebrain development (ES = -0.62, $Q = 5.6 \times 10^{-5}$, one-sided Fisher's Test), telencephalon development (ES = -0.60 , $Q = 2.8 \times 10^{-3}$, one-sided Fisher's Test), and generation of neurons (ES = -0.31, $Q = 7.4 \times 10^{-3}$, one-sided Fisher's Test) in differentiated cells from a ΔSRR2-18$^{129/Cast}$ background compared to those in differentiated cells with intact SRR2-18 (**Fig. 5F; Supplemental Table S9**). These results indicate that SRR2-18-mediated *Sox2* regulation plays an important role in NSPC differentiation, particularly in establishing an anterior neural progenitor identity and facilitating neuronal differentiation.

To determine whether the nuclear regulatory functions of SOX2 protein were disrupted in our allelic series of deletions, we monitored the incorporation of SOX2 into nuclear protein complexes with RNA polymerase II (RNAPII) using proximity ligation assays with confocal microscopy (Dhaliwal et al. 2018, 2019). We specifically used an antibody targeting the phosphorylated serine 5 residue on the carboxy-terminal domain of RNAPII. This allowed us to quantitate the protein interaction frequencies of SOX2 with RNAPII molecules engaged in the elongation of primary transcripts that are spatially resolved within 3D image renderings of single nuclei (Komarnitsky et al. 2000). In undifferentiated NSPCs, SOX2

interacts with RNAPII at a median of ~ 21 visible nucleoplasmic foci. However, we observed a range of SOX2:RNAPII interaction foci that likely reflect asynchronous post-translational SOX2 regulation by cell cycle effectors that contribute to heterogeneity in bulk ESC-derived NSPC cultures (**Fig. 6A**) (Julian et al. 2013). SOX2:RNAPII interactions in ΔSRR2-18^{Cast} NSPCs were significantly decreased compared to those in the parent line (median = 7, $P = 6.55 \times 10^{-6}$, Wilcoxon Test; **Fig. 6B**). The frequency of SOX2:RNAPII association in NSPC nuclei were further decreased in the ΔSRR2-18$^{129/Cast}$ line compared to that of monoallelic deletants (median = 4, $P = 7.25 \times 10^{-3}$, Wilcoxon Test; **Fig. 6B**). Together, these data suggest that deletion of SRR2-18 disrupts global gene regulation in ESC-derived NSPCs through reduced *Sox2* transcription, which affects the participation of SOX2 protein in transcriptionally productive complexes.

Decoupling Sox2 transcriptional regulation from SRR2-18 alters the accessible chromatin landscape of NSPCs

SOX2 has a well-recognized role in supporting a permissive chromatin environment for the proper induction of neurogenic genes in NSPCs (Amador-Arjona et al. 2015; Bertolini et al. 2019). We set out to obtain a comprehensive understanding of the gene regulatory network elicited downstream from *Sox2* decoupled from SRR2-18 regulation. We performed genome-scale identification and analysis of protein-DNA interaction sites inferred from accessible chromatin regions using ATAC-seq. Reproducible ATAC-seq peaks were largely distributed at intergenic regions, annotated TSSs, and gene bodies (**Supplemental Fig. 6A-B**). Consistent with their transcriptional dysregulation of *Sox2*, ΔSRR2-18$^{129/Cast}$ NSPCs showed a decrease in chromatin accessibility at the *Sox2* promoter compared to that in parental cells (**Fig. 6C**). Count-based analysis of ATAC-seq reads within a merged peak set identified 88,910 regions with increased and 65,152 regions with decreased chromatin accessibility (log2FC > |1.5|; $Q < 10e^{-5}$, Wald test under negative binomial model) (**Fig. 6D**; **Supplemental Table**

S10). These data suggest that *Sox2* enhancer-perturbed NSPCs adopt a distinct regulome that reflects a unique spatial and functional cellular identity.

We considered that those candidate regulatory elements represented in the set of decreased chromatin accessibility regions in ΔSRR2-18$^{129/Cast}$ NSPCs were potential SOX2 targets in multipotent neural progenitors. To this end, we intersected our union ATAC-seq peak set shared between parent and deletant NSPCs with publicly available H3K27ac ChIP-seq peaks from Sox1-GFP^{+} *in vitro* differentiated NSPCs (Bonev et al. 2017). We reanalyzed ChIP-seq data obtained from GEO to establish a non-redundant set of SOX2-occupied regions in NSPCs (74,233 unique regions; **Supplemental Table S11**). This allowed us to distinguish between regions of accessible chromatin associated with SOX2 regulation and loci with a similar chromatin profile but that are not direct SOX2 targets (**Fig. 6E**). Average profiles of ATAC-seq signal in the Sox2-mCherry^{129} NSPCs across qualifying genomic intervals showed enrichment of ATAC-seq signal at SOX2-bound regions (mean summit coverage = 6.7 RPM at 7,462 retained peaks) compared to signal at non-SOX2 target H3K27ac-modified regions (mean summit coverage = 3.2 RPM at 1,011 retained peaks) (**Fig. 6F**). As expected for an accessible and acetylated chromatin state, motif enrichment analysis of SOX2-bound regions predominantly recovered SOX, E26 transformation specific (ETS), and activator protein-1 (AP-1) motifs corresponding to the transcription factor families previously recognized to contribute to NSPC self-renewal (**Supplemental Table S12**) (Pagin et al. 2021). Furthermore, SOX2-bound and H3K27ac-modified regions display a lower degree of ATAC-seq signal enrichment in ΔSRR2-18$^{129/Cast}$ cells compared to that in parent NSPCs (**Fig. 6F**). This observation reflects a change in cumulative protein-chromatin interactions at shared SOX2-occupied genomic sites with regulatory activity in neural progenitors.

To better understand the biological processes likely to be affected by altered protein-chromatin interactions genome-wide, GO analysis was performed on the differentially accessible regions annotated by the nearest TSS. Terms overrepresented in the set of ATAC-seq peaks showing decreased accessibility in ΔSRR2-18$^{129/Cast}$ NSPCs were related, but not limited to, forebrain development (*ES* = -0.34, *Q* = 2.9x10^{-10}, one-sided Fisher's Test) and neurogenesis (*ES* = -0.33, *Q* =

$1.7x10^{-35}$, one-sided Fisher's Test) (**Fig. 7A; Supplemental Table S13**). Integrative analysis of the regulome and transcriptome in ΔSRR2-$18^{129/Cast}$ NSPCs revealed that multiple proneural factors known to prime and regulate neuronal specification show a loss of transcriptional activation, including Achaete-scute family TF 1 (*Ascl1*; also known as *Mash1*) (**Fig. 7B**). Reciprocally, genes near ATAC-seq peak regions with increased chromatin accessibility in ΔSRR2-$18^{129/Cast}$ NSPCs included, but were not limited to, terms associated with regulation of anatomical structure morphogenesis (*ES* = 0.37, *Q* = $8.4x10^{-38}$, one-sided Fisher's Test), anterior/posterior pattern specification (*ES* = 0.31, *Q* = $3.0x10^{-8}$, one-sided Fisher's Test), and neural tube development (*ES* = 0.25, *Q* = $4.2x10^{-4}$, one-sided Fisher's Test) (**Fig. 7C; Supplemental Table S14**). We observed a specific increase in chromatin accessibility and transcription in *Hoxb* locus genes (*Hoxb5-9*) related to caudal neural tube patterning (**Supplemental Fig. S6C**), consistent with the regional identity bias of SRR2-18-deleted NSPCs. In the developing vertebrate trunk, caudal type homeobox (CDX) factors function upstream of the *Hox* loci and are required for spinal cord specialization and patterning (Chawengsaksophak et al. 2004; Mazzoni et al. 2013). ΔSRR2-$18^{129/Cast}$ NSPCs exhibit a switch-like gain of active transcriptional features at the *Cdx2* locus compared to the parent Sox2-mCherry^{129} line (**Fig. 7D**). These findings indicate that the induced expression of caudal positional factors and downregulation of anterior proneural factors in biallelic SRR2-18 deletants are supported by genome-wide alterations in the accessible chromatin landscape.

We explored the possibility that biallelic SRR2-18 deletion activates gene regulatory programs associated with posterior/trunk cell types. To assess the similarity of ΔSRR2-$18^{129/Cast}$ NSPCs with tissue-specific chromatin accessibility signatures, we intersected our NSPC ATAC-seq peaks with a collection of E14.5 excised mouse tissue ATAC-seq peaks from the ENCODE Consortium. *Sox2*-mCherry^{129} ESC-derived NSPCs show the greatest overlap in accessible chromatin intervals with forebrain ATAC-seq peaks (non-promoter overlaps = 41,737, odds ratio = 114.8; **Fig. 7E**). Overlap with other anterior neural plate-derived tissues – the midbrain and hindbrain – is consistent with control NSPCs exhibiting a neural chromatin landscape configuration without a well-defined ventral-dorsal identity (non-promoter overlaps = 16,630 & 22,093, odds ratio = 98.5 & 127.1; **Fig. 7E**). Parent NSPCs

exhibited fewer accessible chromatin interval intersections with the caudal portion of the neural tube (non-promoter overlaps = 21,118, odds ratio = 28.4; **Fig. 7E**). Indeed, reproducible ATAC-seq peaks from parent NSPCs were significantly enriched in the forebrain peak set (odds ratio = 15.0, $P = 1.28 \times 10^{-35}$, hypergeometric test) but not in the neural tube (odds ratio = 0.3, $P = 1$, hypergeometric test). The ΔSRR2-18$^{129/Cast}$ NSPC were observed to have fewer shared accessible chromatin regions with forebrain tissue (non-promoter overlaps = 26,672, odds ratio = 18.1) and a relatively greater degree of overlap with neural tube tissue (non-promoter overlaps = 27,146, odds ratio = 92.9; **Fig. 7E**). Accessible chromatin regions from ΔSRR2-18$^{129/Cast}$ NSPC were not significantly enriched in forebrain peaks (odds ratio = 1.7, $P = 1$, hypergeometric test) but did significantly overlap with the neural tube peak set (odds ratio = 2.9, $P = 7.21 \times 10^{-11}$, hypergeometric test).

We performed hierarchical clustering on normalized ATAC-seq reads we reanalyzed from open access data on samples isolated from cranial and trunk neural tissues. Cell type-specific patterns in chromatin accessibility with appropriate biological process enrichment could be identified (**Supplemental Fig. S7A-B**); however, the majority of ΔSRR2-18$^{129/Cast}$ NSPC open chromatin regions form a partially overlapping cluster with moderately accessible genomic regions in ventral neural tube progenitors (Cluster 7). GO terms from the nearest gene are associated with chordate development and cell differentiation in the spinal cord cell. Predictably, these bulk cell and tissue analyses fell short of pinpointing a specific developmental cell state associated with the accessible chromatin profile of ΔSRR2-18$^{129/Cast}$ NSPCs. This contrasted with the NSPCs containing intact SRR2-18, which grouped closer to forebrain NSPCs, astrocytes, and excitatory neurons in unsupervised HCA and PCA analyses (**Supplemental Fig. S7A-C**).

To assess how genome-wide protein-DNA interactions are altered in ΔSRR2-18$^{129/Cast}$ NSPCs, we next predicted the repertoire of TF motifs found at accessible chromatin regions. For this, we carried out digital footprinting, an approach that infers TF occupancy from ATAC-seq data based on the transposition frequency at base-pair resolution, after estimation and correction of positional bias (Bentsen et al. 2020). We found several dozen sequence-specific TFs with a differential occupancy

score outside of statistical thresholds in NSPCs between the two genotypes (score > |0.2|, $Q < 0.01$; **Fig. 7F**, Two-sided one-sample T test under background distribution). Notably, SOX2 footprints were not differentially enriched between parental and ΔSRR2-18$^{129/Cast}$ NSPCs. However, a relative lack of motif content differences across SoxB1 family members could not be ruled out. Moreover, functional compensation between co-expressed SoxB family transcription factors in neural tissues is well documented (Graham et al. 2003; Miyagi et al. 2008). Several transcription factors that regulate NSPC proliferation and neuronal differentiation in various parts of the brain were markedly depleted in ΔSRR2-18$^{129/Cast}$ NSPCs, including KLF15, NFIX, ASCL1, and the Notch downstream effectors HES1 and HEY2 (**Fig. 7F**). Footprinting analyses highlighted several TF motifs that participate in the AP-1 complex and are enriched within NSPCs derived without SRR2-18 mediated *Sox2* regulation (**Fig. 7F**). Additionally, spinal cord specification and patterning factors such as CDX2/4, HOXB-D, and PBX1/2 also contribute to the altered chromatin accessibility profile in ΔSRR2-18$^{129/Cast}$ NSPCs (**Fig. 7F**). These results indicate that decoupling Sox2 expression from the SRR2-18 enhancer cluster induces a loss of anterior neural tissue identity. This loss is characteristic of cultured NSPCs driven by posterior regionalization factors in a complex regulatory network that is sensitive to *Sox2* transcript abundance and protein function.

Discussion

The mechanisms of enhancer-mediated lineage- and cell type-specific gene activation at different developmental stages or spatial contexts are unclear. Previous studies have shown that enhancer switching is common during cell differentiation by comparing lineage-specific TF binding sites in stem/progenitor and mature cells (Soleimani et al. 2012; Alder et al. 2014; Chen et al. 2018a). We found that *Sox2*, a key factor in pluripotency and NSPC fate (Masui et al. 2007; Foshay and Gallicano 2008), exhibited a dynamic temporal pattern of *Sox2* RNA levels during mouse ESC to neural differentiation. Leveraging the genetic background of *M. musculus*/129 × *M. castaneus* ESCs and their derivatives, we demonstrated that these dynamics were coincident with a switch in regulatory control of *Sox2* from the pluripotency associated SCR to a proximal downstream region that determines *Sox2* transcription during the ESC-to-NSPC transition. Additionally, we found that the SRR2-18 region is necessary to maintain the level of *Sox2* transcriptional output in ESC-derived NSPCs under self-renewing conditions. Using an allelic deletion series, we showed that SRR2-18-mediated *Sox2* regulation is critical for sustaining the proliferative rate of NSPCs, establishing an anterior neural progenitor identity, and activating neuronal differentiation genes. We identified patterns in differential chromatin accessibility and TF motif representation between parental and ΔSRR2-18$^{129/Cast}$ NSPCs that propose how neural *Sox2* regulation safeguards a neurogenic program and a regionally biased neural progenitor identity through gene regulatory network activation.

Other investigations have described the regional selectivity of enhancer-mediated gene expression for candidate *Sox2* proximal enhancers in neural tissues using transgenic mouse reporter assays (Uchikawa and Kondoh 2016). The proximal enhancers SRR1 and SRR2 show reporter gene activity in neurospheres and spatially restrict expression to the telencephalon in murine neural development (Zappone et al. 2000; Miyagi et al. 2004, 2006), however, neither SRR1 nor candidate distal enhancers (Beagan et al. 2017; Chakraborty et al. 2023) are sufficient to functionally compensate for the loss of SRR2-18-mediated regulation of *Sox2* in NSPCs. To this point, we observed no difference in the level of *Sox2* knockdown in compound monoallelic ΔSRR1,SRR2-18 (31-42%) and monoallelic ΔSRR2-18

NSPCs (32-38%). A similar level of *Sox2* transcriptional deficiency in NSPCs of the developing forebrain has been described using a *Sox2* haplodeficient mouse model harboring a heterozygous deletion of SRR1 on the intact copy of *Sox2* (Ferri et al. 2004). However, in this case, reduced *Sox2* level in NSPCs *in vivo* was dependent on engineering a null *Sox2* allele. These findings indicate that, despite its functional sufficiency to activate gene transcription in primitive neural cell populations, SRR1 is dispensable for maintained *Sox2* expression in mouse NSPCs. This is consistent with the finding that SRR1 homozygous deletion *in vivo* showed no neurodevelopmental pathology (Ferri et al. 2004).

Conditional biallelic deletion of the *Sox2* coding sequence resulted in impaired hippocampal NSPC survival and differentiation competence, as well as self-renewal defects in primary NSPCs upon extended maintenance in culture (Favaro et al. 2009; Bertolini et al. 2019). These studies highlighted the roles of immediate early genes and cytokine signaling in the transcriptional amplification cascades linked to NSPC self-renewal. Because clustered enhancers may function in a partially redundant manner within the genomic context (Moorthy et al. 2017), our approach was to delete the 16.7 kb region constituting the clustered *Sox2* proximal elements. Using allele-specific transcriptomics, we confirmed that, aside from *Sox2*, no other coding or non-coding RNAs were controlled in a direct *cis*-acting mechanism attributable to SRR2-18. Decoupling *Sox2* regulation from the SRR2-18 enhancer cluster was permissive to neural induction; however, NSPCs harboring a SRR2-18 deletion on even one allele showed increased chromatin accessibility and expression at multiple positional factor genes that specify the caudal and ventral identity of neural cells. The organizing stimulus of gene expression programs characteristic of caudal and ventral neural cell fates in adherent NSPCs is unclear. A reduction in *Sox2* dosage during axial regionalization *in vivo* has been reported to preconfigure the regulatory landscape of the caudal epiblast for WNT signaling input to specify posterior neural fates (Metzis et al. 2018; Blassberg et al. 2022). Additionally, fibroblast growth factor (FGF) signaling is involved in multiple stages of neural tube development, including specification of spinal cord identity, caudal extension and patterning, and directing of the onset of ventral neural tube identity and patterning (Reviewed in (Diez Del Corral and Morales 2017)). However, given that all NSPC culturing media

(containing exogenous FGF2 and EGF) are chemically defined and devoid of exogenous retinoids and Smoothened receptor ligands, the patterning responses observed in ΔSRR2-18$^{129/Cast}$ NSPCs are consistent with cell-derived instructive factors. It will be interesting to examine if SRR2-18-mediated SOX2 regulation can acutely modify embryonic cell responses to FGF or WNT signaling during neural induction and axial regionalization.

Regulating chromatin accessibility by the cumulative action of protein binding and chromatin modification is an important mechanism for controlling gene expression patterns (Thurman et al. 2012). Changes in DNA accessibility or topology can alter gene regulation programs during cell differentiation, leading to the establishment of unique spatial and functional cell identities (Iwafuchi-Doi et al. 2016). SOX2 regulates the transcriptional permissiveness of its target genomic sites in multiple cellular contexts and facilitates regulatory chromatin-chromatin interactions in NSPCs that safeguard against depletion of the progenitor cell pool during brain development (Bertolini et al. 2019; Dodonova et al. 2020). In cranial neural progenitors, SOX2 transcriptionally primes a set of conserved basic helix-loop-helix TFs known as proneural genes, which induce neuron differentiation during brain development (Amador-Arjona et al. 2015). ESC-derived NSPCs exhibited transcriptome and accessible genome features similar to forebrain progenitors and showed primarily anterior neural cell chromatin accessibility profiles. We observed specific patterns in genome-wide chromatin accessibility in ΔSRR2-18$^{129/Cast}$ cells, which are associated with footprints for TFs that have previously established roles in rostral-caudal regionalization of the nervous system. However, the highest scoring differential motif occupancy scores were classical immediate-early genes of the JUN, FOS, and ATF protein families. Components of the AP-1 complex, these multifunctional TFs play roles in cell proliferation, cell migration, neuronal maturation and activity, and the cell stress response (Reviewed in (Bejjani et al. 2019)). Different genetically engineered mouse models deficient for stress- or mitogen-activated phosphorylases upstream of the AP-1 complex show defects in neural tube closure (Sabapathy et al. 1999; Chi et al. 2005); however, there are currently no high-confidence AP-1 TF family targets implicated in neural tube morphogenesis. SOX2 cooperates with distinct TF repertoires to target

diverse genomic sites across different *Sox2*-expressing developmental fates, including in the spinal cord (Hagey et al. 2018). Given that ESC-derived NSPCs were propagated in adherent culture and FGF2-containing medium prior to transcriptome and regulome profiling, the temporal progression in regulatory mechanisms that direct this shift in target DNA recognition remains to be determined.

In this study, we examined the *Sox2* locus and identified a proximal cluster of known and predicted *Sox2* enhancers in NSPCs. Using ectopic reporter assays and allele-specific deletion analyses, we determined that the SRR2-18 enhancer cluster functions as a *cis*-regulator of *Sox2* in NSPCs. We show that SRR2-18-driven *Sox2* allelic dosage is necessary for maintaining the rostro-caudal identity of ESC-generated NSPCs. Elucidating the regulatory mechanisms of *Sox2* is essential for understanding the transcriptional programs controlling neurodevelopment and the spatial diversity of NSPC fates.

Materials and Methods

Embryonic Stem Cell Culture

F1 (*M. musculus*/129 × *M. castaneus*) ESCs were obtained from Barbara Panning, University of California San Francisco (Mlynarczyk-Evans et al. 2006). Cells were maintained on 0.1% (mass/vol) porcine gelatin coated plates and Dulbecco's Modified Eagle Medium (Thermo Fisher 11960044) supplemented with 15% (vol/vol) ESC-qualified fetal bovine serum (Wisent Bioproducts 920040), 0.1 mM nonessential amino acids (Thermo Fisher 11140050), 1 mM sodium pyruvate (Thermo Fisher 11360070), 2 mM GlutaMAX (Thermo Fisher 35050061), 0.1 mM 2-mercaptoethanol (Thermo Fisher 21985023), 1X Penicillin-Streptomycin (Thermo Fisher 15140122), 1000 U/mL recombinant leukemia inhibitory factor (LIF), 3 μM GSK3β inhibitor CHIR99021 (Millipore Sigma 361559-5MG), and 1 μM MEK inhibitor PD0325901 (Millipore Sigma 444966-5MG), referred to here as "LIF/2i medium". Incubators were humidified and maintained at 37°C, 5% CO_2 under normoxic conditions. All cultures were confirmed to be free of mycoplasma contamination during routine testing with the Mycoplasma PCR Detection Kit (BioVision Inc. K1476-100).

In Vitro Neural Differentiation and Neural Stem/Progenitor Cell Culture

F1 mouse ESCs were differentiated toward NSPC *in vitro* (Ying et al. 2003). ESCs were seeded into 10 cm gelatinized culture plates in LIF/2i medium (day 0) at a density of 1 x 10^4 cells per cm^2. On day 1, culture medium was changed to a 1:1 mixture of Neurobasal medium (Thermo Fisher 21103049) and DMEM/F12 medium (Thermo Fisher 10565018) supplemented with 1X N2 (Thermo Fisher 17502048), 1X B27 without Vitamin A (Thermo Fisher 12587010), 2 mM GlutaMAX, 10 mM HEPES (Thermo Fisher 15630080), 0.015% Fraction V BSA (mass/vol), 0.1 mM 2-mercaptoethanol and 1X Penicillin-Streptomycin, referred to here as "N2B27 medium". On day 5, adherent differentiating cells were detached with Accutase (Millipore Sigma A6964-500ML) and transferred to low adherence 10 cm plates in N2B27 medium supplemented with 10 ng/mL murine epidermal growth factor (Peprotech Inc. 315-09-100UG) and 10 ng/mL fibroblast growth factor 2 (Cedarlane Labs CLCYT386), referred to here as "NSE medium". On days 10-12, the cell aggregates formed in suspension culture were dissociated with Accutase and seeded onto T25 flasks coated with 100 µg/mL of poly-D-Lysine (PDL) (Millipore Sigma 27964-99-4) and 10 µg/mL laminin (Millipore Sigma L2020) in NSE medium at a density of 2 x 10^4 cells per cm^2. This is considered passage 0 (p0) of neural stem/progenitor cell cultures. NSPCs were seeded into poly-D-lysine- and laminin-coated T25 flasks in NSE medium at a density of 2 x 10^4 cells per cm^2 and passaged every 4-5 days for a maximum of 25 passages. NSPC multipotency was assessed by differentiation into neuronal, astrocyte, and oligodendrocyte lineages following removal of EGF from NSE medium for one day and later by culture in NeuroCult™ Differentiation Medium (Stem Cell Technologies 05704) for an additional 9 days.

CRISPR-Cas9 Mediated Deletion and Fluorescent Reporter Line Generation

Cas9-mediated deletions were carried out as described (Moorthy and Mitchell, 2016). Briefly, Cas9 targeting guide RNAs (gRNAs) flanking predicted enhancer regions are provided in **Supplemental**

Table 15. gRNA plasmids were assembled in the gRNA Cloning Vector (a gift from George Church; Addgene plasmid # 41824), using the protocol described by Mali and colleagues (Mali et al. 2013). The sequence of the resulting guide plasmid was confirmed by Sanger sequencing. For deletions, F1 ESC were transfected with 5 µg each of 5′ gRNA, 3′ gRNA, and pCas9_GFP plasmid (a gift from Kiran Musunuru; Addgene plasmid # 44719) (Ding et al. 2013). To tag *Sox2* at the C-terminal, we constructed a HDR donor vector using Gibson assembly. The following sequences were assembled using the Gibson Assembly cloning kit (New England Biolabs E5510S): mCherry2-N1 amplicon (a gift from Michael Davidson; Addgene plasmid # 54517), P2A sequence through oligos, and homology arms amplified from F1 genomic DNA excluding the gRNA target site in the 3' arm (**Supplemental Table 17**). For homology-directed repair knock-in, F1 ESC were transfected with 5 µg each of *Sox2* 3' untranslated region (UTR) gRNA and pCas9_GFP plasmid, as well as 10 µg donor fragment purified from EcoRV restriction digest. Transfections were performed using the Neon™ Transfection System (Thermo Fisher Scientific). Forty-eight hours after transfection, cells were sorted on a BD FACSAria (Temerty Faculty of Medicine Flow Cytometry Facility, Toronto) for GFP-positive cells (in knock-out experiments) or mCherry-positive cells (for tag knock-in). The collected cells were seeded at 1 x $10^4/cm^2$ on 10 cm gelatinized culture dishes and grown for 5–6 days until large well-defined colonies formed. Isolated colonies were picked and propagated for genotyping and gene expression analysis. For DNA extraction from 96-well plates, cells were washed once with PBS and subsequently lysed in a humidified chamber with 50 µL 1X cell lysis buffer using the prepGEM Universal Kit (PUN0100, MicroGEM), according to manufacturer's recommendations. The 96-well plates were screened for deletions using allelic-specific qPCR-based assays. All deletions were confirmed by Sanger sequencing and analysis using primers 5′ and 3′ from the gRNA target sites. SNPs within the amplified product confirmed the genotype of the deleted allele.

Immunofluorescence Microscopy and Proximity Ligation Assays

Cells grown on coated glass coverslips were fixed with 10% neutral formalin (Millipore Sigma HT501128) for 15 min. Cells were permeabilized and blocked with 10% fetal bovine serum (FBS) and 0.1% Triton-X100 in 1X phosphate buffer saline (PBS) for 60 min at room temperature. Primary antibodies were applied for overnight incubation at 4 °C at a dilution of 1:40. Cells were washed twice with 0.1% Tween-20 in PBS and secondary antibodies were applied at a dilution of 1:1000 for 2 hours. Primary and secondary antibodies were diluted in 0.2% FBS, and 0.15% Tween-20 in PBS. Antibody product identifiers and dilutions used for immunofluorescence are listed in **Supplemental Table S16**. Nuclei were counterstained with 5 ng/mL 4',6-diamidino-2-phenylindole (DAPI) for 30 seconds. Following five washes in PBS, coverslips were mounted onto slides with Vectashield (Vector Laboratories H100010) and sealed using transparent nail polish. Epifluorescence imaging was performed on a Zeiss Axio Imager Z2 Upright Microscope (ZEISS), equipped with a X-cite Exacte light source (Excelitas Technologies) and a CoolCube 1 camera (MetaSystems). Raw images were processed in FIJI (Schindelin et al., 2012).

For *in situ* protein detection by proximity-dependent DNA ligation assays (Fredriksson et al. 2002), NSPCs were fixed and processed on coverslips as above. Primary antibodies against phospho-serine 5 carboxy terminal domain (S5P CTD) RNA polymerase II (Abcam ab5131; RRID: AB_449369) and SOX2 (R and D Systems MAB2018; RRID: AB_358009) (MAB2018, R&D Systems), were diluted 1:200 and 1:1000 (respectively) into 0.2% FBS/0.1% Triton X-100 in PBS and incubated on coverslips overnight at 4 °C. Following this, coverslips with NSPCs were washed three times in 0.1% Tween-20 in PBS for 5 minutes each. Duolink probes (mouse plus and rabbit minus) were prepared and applied as described by manufacturer's instructions for the Duolink® PLA kit Cy5 (Millipore-Sigma DUO92013). Coverslips with NSPCs were mounted in VECTASHIELD Antifade Mounting Medium with DAPI. Images were collected using a Leica TCS SP8 confocal microscope and 100X magnification objective lens with oil immersion at constant exposure. Individual nuclei were identified using Imaris 7.1 (Oxford Instruments; RRID:SCR_007370). 3D models rendered and PLA focal signals were automatically counted per nucleus. All PLA experiments were carried out on at least three independent cultures of

NSPCs and a minimum of 100 individual nuclei were imaged per genotype. Single antibody and PLA probe hybridization negative controls were also conducted.

RNA Extraction and Reverse-Transcription Quantitative PCR

Total RNA from cultured cells was isolated using RNeasy RNA kit (Qiagen) and processed with TURBO DNA-free kit (Thermo Fisher AM1907). DNase-treated RNA was reverse transcribed with random primers using the high-capacity cDNA synthesis kit (Thermo Fisher 4368814). Target gene expression was monitored by quantitative PCR and normalized to *Gapdh* and *Eef2a* levels with gene-specific primers (oligonucleotide sequences provided in **Supplemental Table S17**). We monitored expression of *Sox2* by allele-specific reverse transcription quantitative real-time PCR (RT-qPCR) due to the presence of discriminatory SNPs within the *Sox2* transcript, as previously described (Zhou et al. 2014). For each biological replicate, qPCR reactions were performed in technical duplicate using SYBR Select Master mix (Thermo Fisher 4472908) and the CFX 384 Real-Time Detection system (BIO-RAD; RRID:SCR_018057). Expression levels were interpolated from a standard curve dilution series of F1 ESC genomic DNA using CFX Maestro Software. All RNA samples were confirmed not to have DNA contamination because no amplification was observed in reverse transcriptase negative samples.

Reporter Plasmid Construction

SRR2 and SRR107 were amplified from a Sox2 BAC (RP23-274P9) by PCR using Phusion polymerase (New England Biolabs E0553S) and cloned into the pJET1.2/blunt cloning vector (Thermo Fisher K1232) by Zhou et al. (Zhou et al. 2014). The remaining *Sox2* neural enhancer candidates were amplified from genomic DNA isolated from Bl6 mouse tissue (Jackson Labs) using the DNeasy Blood and Tissue Kit (Qiagen 69504). The 1.8-kb *Sox2* promoter (pSox2) was amplified from mouse genomic DNA by PCR using DreamTaq polymerase (Thermo Fisher EP0702) and ligated into the BanII site to

replace the TATA-box minimal promoter of the pGL4.23 vector. All primers used in plasmid construction are provided in **Supplemental Table S18**. Enhancer sequences were amplified using primers containing 15 bp 5'overhangs homologous to the sequences flanking the NotI restriction site in the pGL4.23 luciferase reporter vector. Amplicons were cloned into the NotI site of the pGL4.23 vector with pSox2 by In-Fusion® cloning (Takara Bio 639650). All plasmids were purified from bacterial culture using Presto™ Midi Plasmid Kit (Geneaid Biotech Inc. PI025).

Dual Luciferase Reporter Assay

To evaluate enhancer activity in a luciferase reporter assay, F1 ESCs were seeded on gelatin-coated 96-well plates at a density of 10,000 cells per well. Alternatively, F1 NSPCs were seeded onto PDL/laminin-coated 96-well plates at a density of 50,000 cells per well. After 24 hours, the cells were co-transfected using Lipofectamine 3000 (Thermo Fisher L3000008) with pSox2-pGL4.23 reporter vectors and pGL4.75 encoding Renilla luciferase (Promega E8411 & E6931) at a 50:1 molar ratio. The Renilla plasmid served as an internal control for transfection efficiency in each well. After 24 hours, spent growth medium was replaced with fresh growth medium, according to the cell type transfected. Luciferase activity in cell lysates was assayed 48 hours post-transfection using the Dual Luciferase Reporter Assay kit (Promega E1910), and the Fluoroskan Ascent FL microplate reader (Thermo Fisher Scientific). After background signal correction, the ratio of firefly/Renilla luciferase activity was calculated for each tested enhancer candidate and was normalized to that of the empty vector.

RNA-sequencing Library Preparation and Data Processing

Total RNA from cultured NSPCs was isolated and processed in the same way as samples prepared for reverse-transcription quantitative PCR assays. Total RNA was sent to The Centre for Applied Genomics (The Hospital for Sick Children, Toronto) for paired-end rRNA-depleted total RNA-seq

(Illumina 2500, 125 bp). Raw RNA-seq reads from the public domain were retrieved from the GEO, which include data from the following studies (Bertolini et al. 2019; Lattke et al. 2021). Raw reads were assessed for quality, adapter content and duplication rates with FastQC and trimmed with Fastp v0.20.0 (Chen et al. 2018b). Trimmed reads were aligned to mm10 (GENCODE M7) with with STAR v2.9.0a (Dobin et al. 2013), with the following: parameters --alignEndsType Local -- --outFilterMultimapNmax 20. Exon-mapped reads were quantified using featureCounts (Liao et al. 2014) and imported into R v4.4.1 for differential expression analysis with DESeq2 (Love et al. 2014). Genes with an absolute log2 fold-change > 1 and false discovery rate-adjusted P < 0.05 in total exon-mapped read counts were considered differentially expressed.

For allele-specific RNA-seq, a hybrid version of mm10 (GENCODE M7) was created with SNPsplit v0.3.4 (Krueger and Andrews 2016) using CAST/EiJ SNPs to N-mask discriminatory variants on a 129S1/SvImJ SNP-substituted reference based on GRCh38 from the Mouse Genome Project (dbSNP142). Trimmed reads were aligned to the hybrid N-masked assembly with STAR v2.9.0a, with the following parameters: --alignEndsType EndToEnd --outSAMattributes NH HI NM MD --outMultimapperOrder Random --outSAMmultNmax 1. Allele-specific sorting of alignment files was performed using SNPsplit with default parameters. Genes with an absolute log2 fold-change > 0.5 and false discovery rate-adjusted P < 0.05 in allele-sorted read counts using DESeq2 were considered to have significant allelic imbalances.

The alignment files of all replicates per sample group were combined using Samtools merge and used to create bedgraphs that were normalized to counts per million mapped reads (CPM) using Deeptools bamCoverage (Ramírez et al. 2014).

Assay for Transposase Accessible Chromatin with Sequencing (ATAC-seq) Library Preparation and Data Processing

NSPC ATAC-seq multiplex library preparation and next-generation sequencing were performed at the Princess Margaret Genomics Centre, Toronto, Canada (www.pmgenomics.ca). 30,000 – 50,000 Sox2-mCherry^{129} and ΔSRR2-18$^{129/Cast}$ NSPCs were dissociated with Accutase and washed once with 0.04% BSA in PBS. Intact nuclei were isolated and processed using standard ATAC-seq procedures developed by Buenrostro et al. (Buenrostro et al. 2013) with modifications to the lysis and transposition steps based on the OMNI-ATAC protocol by Corces et al. (Corces et al. 2017). Demultiplexed raw ATAC-seq reads were processed using the ENCODE ATAC-seq pipeline (v.1.10.0) established by the Kundaje Lab (RRID:SCR_023100). In this pipeline, MACS2 was used for peak-calling and conserved peaks in biological replicates were filtered using Irreproducible Discovery Rate (IDR; RRID:SCR_017237) (Li et al. 2011). Accessible chromatin regions identified by conservative thresholded IDR peaks were retained for downstream analyses. The alignment files of all replicates per sample group were combined using Samtools merge and used to create bedGraphs that were normalized to counts per million mapped reads (CPM) using Deeptools bamCoverage. IDR peaks were imported into R and annotated to the nearest transcriptional start site of genes using the ChIPpeakanno package (Zhu et al. 2010). Normalization of read counts and analysis of differential accessibility for the union peak regions were performed using DiffBind package (Stark and Brown 2011). Peaks with padj ≤ 0.01 and log2FC ≥ 2 were considered differentially accessible.

Transcription Factor Motif Enrichment

Transcription factor digital footprint analysis was performed using TOBIAS with standard settings (Bentsen et al. 2020) and a reference compendium of non-redundant mammalian TF motifs from the Jaspar 2022 release (Castro-Mondragon et al. 2022). Non-redundant motifs with an absolute log2 fold-change > 0.2 and *Q* value < 0.01 were considered significantly enriched in each condition. Independent cell culture replicates (n = 3) were merged into a single BAM file for each treatment for Tn5 bias

correction and footprint scoring. A union peak set was exported from DiffBind. ENCODE mm10 blacklist regions (ENCFF543DDX) were excluded (Amemiya et al. 2019).

Statistics and Reproducibility

Data were plotted and analyzed with R version 4.4.1 (R Core Team 2020) and the Tidyverse package (Wickham et al. 2019). Statistical methods, *P* values or adjusted *P* values for each comparison are listed in the figure legend and/or in the corresponding results section. Assumptions of normality and homogeneity of variance were assessed with Shapiro-Wilk and Fligner-Killeen tests. For all experiments, sample size was determined empirically. Investigators were not blinded to allocation during experiments or outcome assessments. Sequencing data and processed bedGraph or narrowPeak files from this work were submitted to the GEO repository under the reference series GSE237778.

Competing Interests Statement

The authors declare no competing interests.

Acknowledgments

The authors thank members of the Mitchell group for helpful discussions. This work was supported by the Canadian Institutes of Health Research (FRN PJT153186 and PJT180312), the Canada Foundation for Innovation, and the Ontario Ministry of Research and Innovation (operating and infrastructure grants held by J.A.M.). Salary support for S. D. M. was provided by Canadian Institutes of Health Research Project Grant (FRN 153186). Salary support for I.C.T was provided by an institutional fellowship. Salary support for Z. E. G. was provided by a Canadian Institutes of Health Research fellowship. Studentship

funding was provided by the Natural Science and Engineering Research Council of Canada (CGS Doctoral held by V. M. S.) This research was enabled in part by support provided by Compute Ontario (https://www.computeontario.ca/) and the Digital Research Alliance of Canada (https://www.alliancecan.ca). We thank the ENCODE Consortium for data sets queried here. We also thank Alan M. Moses for allowing A.G.D. to use computational resources while in his lab.

Author Contributions

I.C.T., S.D.M., and J.A.M. conceived the project and designed the experiments. I.C.T, S.D.M., V.M.S., L.L., Z.E.G., N.G., M.C., and R.B.D. made experimental contributions to the gene expression analysis, generation of deleted ESC lines, NSPC derivation and maintenance, plasmid construction and reporter assays, sample preparation for fluorescence microscopy, confocal imaging, RNA-seq and ATAC-seq data processing. I.C.T and A.G.D curated and formally analyzed all experimental data. I.C.T. generated data visualizations and wrote the original draft. J.A.M. acquired funding and provided resources, administration, and supervision to this project. Review and editing of the manuscript were carried out by all co-authors.

References

Alder O, Cullum R, Lee S, Kan AC, Wei W, Yi Y, Garside VC, Bilenky M, Griffith M, Morrissy AS. 2014. Hippo signaling influences HNF4A and FOXA2 enhancer switching during hepatocyte differentiation. *Cell Rep* **9**: 261–271.

Amador-Arjona A, Cimadamore F, Huang C-T, Wright R, Lewis S, Gage FH, Terskikh AV. 2015. SOX2 primes the epigenetic landscape in neural precursors enabling proper gene activation during hippocampal neurogenesis. *Proc Natl Acad Sci* **112**. https://pnas.org/doi/full/10.1073/pnas.1421480112 (Accessed March 12, 2023).

Amemiya HM, Kundaje A, Boyle AP. 2019. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci Rep* **9**: 9354.

Andreu-Agullo C, Maurin T, Thompson CB, Lai EC. 2012. Ars2 maintains neural stem-cell identity through direct transcriptional activation of Sox2. *Nature* **481**: 195–198.

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

Bailey PJ, Klos JM, Andersson E, Karlén M, Källström M, Ponjavic J, Muhr J, Lenhard B, Sandelin A, Ericson J. 2006. A global genomic transcriptional code associated with CNS-expressed genes. *Exp Cell Res* **312**: 3108–3119.

Banerji J, Rusconi S, Schaffner W. 1981. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**: 299–308.

Beagan JA, Duong MT, Titus KR, Zhou L, Cao Z, Ma J, Lachanski CV, Gillis DR, Phillips-Cremins JE. 2017. YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. *Genome Res* **27**: 1139–1152.

Beagan JA, Gilgenast TG, Kim J, Plona Z, Norton HK, Hu G, Hsu SC, Shields EJ, Lyu X, Apostolou E, et al. 2016. Local Genome Topology Can Exhibit an Incompletely Rewired 3D-Folding State during Somatic Cell Reprogramming. *Cell Stem Cell* **18**: 611–624.

Bejjani F, Evanno E, Zibara K, Piechaczyk M, Jariel-Encontre I. 2019. The AP-1 transcriptional complex: Local switch or remote command? *Biochim Biophys Acta BBA - Rev Cancer* **1872**: 11–23.

Bentsen M, Goymann P, Schultheis H, Klee K, Petrova A, Wiegandt R, Fust A, Preussner J, Kuenne C, Braun T. 2020. ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. *Nat Commun* **11**: 4267.

Bertacchi M, Pandolfini L, Murenu E, Viegi A, Capsoni S, Cellerino A, Messina A, Casarosa S, Cremisi F. 2013. The positional identity of mouse ES cell-generated neurons is affected by BMP signaling. *Cell Mol Life Sci CMLS* **70**: 1095–1111.

Bertolini JA, Favaro R, Zhu Y, Pagin M, Ngan CY, Wong CH, Tjong H, Vermunt MW, Martynoga B, Barone C, et al. 2019. Mapping the Global Chromatin Connectivity Network for Sox2 Function in Neural Stem Cell Maintenance. *Cell Stem Cell* **24**: 462-476.e6.

Blassberg R, Patel H, Watson T, Gouti M, Metzis V, Delás MJ, Briscoe J. 2022. Sox2 levels regulate the chromatin occupancy of WNT mediators in epiblast progenitors responsible for vertebrate body formation. *Nat Cell Biol* **24**: 633–644.

Bonev B, Mendelson Cohen N, Szabo Q, Fritsch L, Papadopoulos GL, Lubling Y, Xu X, Lv X, Hugnot J-P, Tanay A, et al. 2017. Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* **171**: 557-572.e24.

Brandenberger R, Wei H, Zhang S, Lei S, Murage J, Fisk GJ, Li Y, Xu C, Fang R, Guegler K, et al. 2004. Transcriptome characterization elucidates signaling networks that control human ES cell growth and differentiation. *Nat Biotechnol* **22**: 707–716.

Brosh R, Coelho C, Ribeiro-dos-Santos AM, Ellis G, Hogan MS, Ashe HJ, Somogyi N, Ordoñez R, Luther RD, Huang E. 2023. Synthetic regulatory genomics uncovers enhancer context dependence at the Sox2 locus. *Mol Cell* **83**: 1140–1152.

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**: 1213–1218.

Carter D, Chakalova L, Osborne CS, Dai Y, Fraser P. 2002. Long-range chromatin regulatory interactions in vivo. *Nat Genet* **32**: 623–626.

Castro-Mondragon JA, Riudavets-Puig R, Rauluseviciute I, Berhanu Lemma R, Turchi L, Blanc-Mathieu R, Lucas J, Boddie P, Khan A, Manosalva Pérez N. 2022. JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **50**: D165–D173.

Catena R, Tiveron C, Ronchi A, Porta S, Ferri A, Tatangelo L, Cavallaro M, Favaro R, Ottolenghi S, Reinbold R, et al. 2004. Conserved POU Binding DNA Sites in the Sox2 Upstream Enhancer Regulate Gene Expression in Embryonic and Neural Stem Cells. *J Biol Chem* **279**: 41846–41857.

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

Chakraborty S, Kopitchinski N, Zuo Z, Eraso A, Awasthi P, Chari R, Mitra A, Tobias IC, Moorthy SD, Dale RK, et al. 2023. Enhancer–promoter interactions can bypass CTCF-mediated boundaries and contribute to phenotypic robustness. *Nat Genet* **55**: 280–290.

Chawengsaksophak K, de Graaff W, Rossant J, Deschamps J, Beck F. 2004. Cdx2 is essential for axial elongation in mouse development. *Proc Natl Acad Sci U S A* **101**: 7641–7645.

Chen AF, Liu AJ, Krishnakumar R, Freimer JW, DeVeale B, Blelloch R. 2018a. GRHL2-Dependent Enhancer Switching Maintains a Pluripotent Stem Cell Transcriptional Subnetwork after Exit from Naive Pluripotency. *Cell Stem Cell* **23**: 226-238.e4.

Chen C, Morris Q, Mitchell JA. 2012. Enhancer identification in mouse embryonic stem cells using integrative modeling of chromatin and genomic features. *BMC Genomics* **13**: 152.

Chen S, Zhou Y, Chen Y, Gu J. 2018b. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**: i884–i890.

Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orlov YL, Zhang W, Jiang J, et al. 2008. Integration of External Signaling Pathways with the Core Transcriptional Network in Embryonic Stem Cells. *Cell* **133**: 1106–1117.

Chi H, Sarkisian MR, Rakic P, Flavell RA. 2005. Loss of mitogen-activated protein kinase kinase kinase 4 (MEKK4) results in enhanced apoptosis and defective neural tube development. *Proc Natl Acad Sci U S A* **102**: 3846–3851.

Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, Satpathy AT, Rubin AJ, Montine KS, Wu B. 2017. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods* **14**: 959–962.

Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, et al. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci* **107**: 21931–21936.

Dhaliwal NK, Abatti LE, Mitchell JA. 2019. KLF4 protein stability regulated by interaction with pluripotency transcription factors overrides transcriptional control. *Genes Dev* **33**: 1069–1082.

Dhaliwal NK, Miri K, Davidson S, Tamim El Jarkass H, Mitchell JA. 2018. KLF4 Nuclear Export Requires ERK Activation and Initiates Exit from Naive Pluripotency. *Stem Cell Rep* **10**: 1308–1323.

Diez Del Corral R, Morales AV. 2017. The Multiple Roles of FGF Signaling in the Developing Spinal Cord. *Front Cell Dev Biol* **5**: 58.

Ding Q, Lee Y-K, Schaefer EA, Peters DT, Veres A, Kim K, Kuperwasser N, Motola DL, Meissner TB, Hendriks WT. 2013. A TALEN genome-editing system for generating human stem cell-based disease models. *Cell Stem Cell* **12**: 238–251.

Dodonova SO, Zhu F, Dienemann C, Taipale J, Cramer P. 2020. Nucleosome-bound SOX2 and SOX11 structures elucidate pioneer factor function. *Nature* **580**: 669–672.

ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57.

Ensini M, Tsuchida TN, Belting HG, Jessell TM. 1998. The control of rostrocaudal pattern in the developing spinal cord: specification of motor neuron subtype identity is initiated by signals from paraxial mesoderm. *Dev Camb Engl* **125**: 969–982.

Favaro R, Valotta M, Ferri ALM, Latorre E, Mariani J, Giachino C, Lancini C, Tosetti V, Ottolenghi S, Taylor V, et al. 2009. Hippocampal development and neural stem cell maintenance require Sox2-dependent regulation of Shh. *Nat Neurosci* **12**: 1248–1256.

Ferri ALM, Cavallaro M, Braida D, Di Cristofano A, Canta A, Vezzani A, Ottolenghi S, Pandolfi PP, Sala M, DeBiasi S, et al. 2004. Sox2 deficiency causes neurodegeneration and impaired neurogenesis in the adult mouse brain. *Development* **131**: 3805–3819.

Foshay KM, Gallicano GI. 2008. Regulation of Sox2 by STAT3 initiates commitment to the neural precursor cell fate. *Stem Cells Dev* **17**: 269–278.

Fredriksson S, Gullberg M, Jarvius J, Olsson C, Pietras K, Gústafsdóttir SM, Östman A, Landegren U. 2002. Protein detection using proximity-dependent DNA ligation assays. *Nat Biotechnol* **20**: 473–477.

Garcia ADR, Doan NB, Imura T, Bush TG, Sofroniew MV. 2004. GFAP-expressing progenitors are the principal source of constitutive neurogenesis in adult mouse forebrain. *Nat Neurosci* **7**: 1233–1241.

Götz M, Barde Y-A. 2005. Radial glial cells defined and major intermediates between embryonic stem cells and CNS neurons. *Neuron* **46**: 369–372.

Graham V, Khudyakov J, Ellis P, Pevny L. 2003. SOX2 functions to maintain neural progenitor identity. *Neuron* **39**: 749–765.

Gubbay J, Collignon J, Koopman P, Capel B, Economou A, Münsterberg A, Vivian N, Goodfellow P, Lovell-Badge R. 1990. A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature* **346**: 245–250.

Hagey DW, Klum S, Kurtsdotter I, Zaouter C, Topcic D, Andersson O, Bergsland M, Muhr J. 2018. SOX2 regulates common and specific stem cell features in the CNS and endoderm derived organs. *PLoS Genet* **14**: e1007224.

Hagey DW, Zaouter C, Combeau G, Lendahl MA, Andersson O, Huss M, 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.

Hallmann A-L, Araúzo-Bravo MJ, Zerfass C, Senner V, Ehrlich M, Psathaki OE, Han DW, Tapia N, Zaehres H, Schöler HR, et al. 2016. Comparative transcriptome analysis in induced neural stem cells reveals defined neural cell identities in vitro and after transplantation into the adult rodent brain. *Stem Cell Res* **16**: 776–781.

Hay D, Hughes JR, Babbs C, Davies JOJ, Graham BJ, Hanssen LLP, Kassouf MT, Oudelaar AM, Sharpe JA, Suciu MC, et al. 2016. Genetic dissection of the α-globin super-enhancer in vivo. *Nat Genet* **48**: 895–903.

Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, et al. 2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**: 108–112.

Hitoshi S, Seaberg RM, Koscik C, Alexson T, Kusunoki S, Kanazawa I, Tsuji S, van der Kooy D. 2004. Primitive neural stem cells from the mammalian epiblast differentiate to definitive neural stem cells under the control of Notch signaling. *Genes Dev* **18**: 1806–1811.

Huang J, Liu X, Li D, Shao Z, Cao H, Zhang Y, Trompouki E, Bowman TV, Zon LI, Yuan G-C, et al. 2016. Dynamic Control of Enhancer Repertoires Drives Lineage and Stage-Specific Transcription during Hematopoiesis. *Dev Cell* **36**: 9–23.

Ivanova N, Dobrin R, Lu R, Kotenko I, Levorse J, DeCoste C, Schafer X, Lun Y, Lemischka IR. 2006. Dissecting self-renewal in stem cells with RNA interference. *Nature* **442**: 533–538.

Iwafuchi-Doi M, Donahue G, Kakumanu A, Watts JA, Mahony S, Pugh BF, Lee D, Kaestner KH, Zaret KS. 2016. The Pioneer Transcription Factor FoxA Maintains an Accessible Nucleosome Configuration at Enhancers for Tissue-Specific Gene Activation. *Mol Cell* **62**: 79–91.

Iwafuchi-Doi M, Matsuda K, Murakami K, Niwa H, Tesar PJ, Aruga J, Matsuo I, Kondoh H. 2012. Transcriptional regulatory networks in epiblast cells and during anterior neural plate development as modeled in epiblast stem cells. *Development* **139**: 3926–3937.

Iwafuchi-Doi M, Yoshida Y, Onichtchouk D, Leichsenring M, Driever W, Takemoto T, Uchikawa M, Kamachi Y, Kondoh H. 2011. The Pou5f1/Pou3f-dependent but SoxB-independent regulation of conserved enhancer N2 initiates Sox2 expression during epiblast to neural plate stages in vertebrates. *Dev Biol* **352**: 354–366.

Johe KK, Hazel TG, Muller T, Dugich-Djordjevic MM, McKay RD. 1996. Single factors direct the differentiation of stem cells from the fetal and adult central nervous system. *Genes Dev* **10**: 3129–3140.

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

Julian LM, Vandenbosch R, Pakenham CA, Andrusiak MG, Nguyen AP, McClellan KA, Svoboda DS, Lagace DC, Park DS, Leone G. 2013. Opposing regulation of Sox2 by cell-cycle effectors E2f3a and E2f3b in neural stem cells. *Cell Stem Cell* **12**: 440–452.

Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS, et al. 2010. Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**: 430–435.

Kelberman D, Rizzoti K, Avilion A, Bitner-Glindzicz M, Cianfarani S, Collins J, Chong WK, Kirk JMW, Achermann JC, Ross R, et al. 2006. Mutations within Sox2/SOX2 are associated with abnormalities in the hypothalamo-pituitary-gonadal axis in mice and humans. *J Clin Invest* **116**: 2442–2455.

Kieffer-Kwon K-R, Tang Z, Mathe E, Qian J, Sung M-H, Li G, Resch W, Baek S, Pruett N, Grøntved L, et al. 2013. Interactome Maps of Mouse Gene Regulatory Domains Reveal Basic Principles of Transcriptional Regulation. *Cell* **155**: 1507–1520.

Komarnitsky P, Cho E-J, Buratowski S. 2000. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev* **14**: 2452–2460.

Kopp JL, Ormsbee BD, Desler M, Rizzino A. 2008. Small increases in the level of Sox2 trigger the differentiation of mouse embryonic stem cells. *Stem Cells* **26**: 903–911.

Krueger F, Andrews SR. 2016. SNPsplit: Allele-specific splitting of alignments between genomes with known SNP genotypes. F1000Res 5, 1479.

Langer L, Taranova O, Sulik K, Pevny L. 2012. SOX2 hypomorphism disrupts development of the prechordal floor and optic cup. *Mech Dev* **129**: 1–12.

Lattke M, Goldstone R, Ellis JK, Boeing S, Jurado-Arjona J, Marichal N, MacRae JI, Berninger B, Guillemot F. 2021. Extensive transcriptional and chromatin changes underlie astrocyte maturation in vivo and in culture. *Nat Commun* **12**: 4335.

Lendahl U, Zimmerman LB, McKay RD. 1990. CNS stem cells express a new class of intermediate filament protein. *Cell* **60**: 585–595.

Li H, Collado M, Villasante A, Matheu A, Lynch CJ, Cañamero M, Rizzoti K, Carneiro C, Martínez G, Vidal A, et al. 2012. p27Kip1 Directly Represses Sox2 during Embryonic Stem Cell Differentiation. *Cell Stem Cell* **11**: 845–852.

Li P-Y, Li S-Q, Gao S-G, Dong D-Y. 2021. CRISPR/Cas9-mediated gene editing on Sox2ot promoter leads to its truncated expression and does not influence neural tube closure and embryonic development in mice. *Biochem Biophys Res Commun* **573**: 107–111.

Li Q, Brown JB, Huang H, Bickel PJ. 2011. Measuring reproducibility of high-throughput experiments.

Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**: 923–930.

Lodato MA, Ng CW, Wamstad JA, Cheng AW, Thai KK, Fraenkel E, Jaenisch R, Boyer LA. 2013. SOX2 Co-Occupies Distal Enhancer Elements with Distinct POU Factors in ESCs and NPCs to Specify Cell State ed. G.S. Barsh. *PLoS Genet* **9**: e1003288.

Loh Y-H, Wu Q, Chew J-L, Vega VB, Zhang W, Chen X, Bourque G, George J, Leong B, Liu J, et al. 2006. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet* **38**: 431–440.

Love M, Anders S, Huber W. 2014. Differential analysis of count data–the DESeq2 package. *Genome Biol* **15**: 10–1186.

Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. 2013. RNA-guided human genome engineering via Cas9. *Science* **339**: 823–826.

Marqués-Torrejón MÁ, Porlan E, Banito A, Gómez-Ibarlucea E, Lopez-Contreras AJ, Fernández-Capetillo Ó, Vidal A, Gil J, Torres J, Fariñas I. 2013. Cyclin-Dependent Kinase Inhibitor p21 Controls Adult Neural Stem Cell Expansion by Regulating Sox2 Gene Expression. *Cell Stem Cell* **12**: 88–100.

Masui S, Nakatake Y, Toyooka Y, Shimosato D, Yagi R, Takahashi K, Okochi H, Okuda A, Matoba R, Sharov AA, et al. 2007. Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat Cell Biol* **9**: 625–635.

Mathis L, François Nicolas J. 2000. Different clonal dispersion in the rostral and caudal mouse central nervous system. *Development* **127**: 1277–1290.

Mazzoni EO, Mahony S, Peljto M, Patel T, Thornton SR, McCuine S, Reeder C, Boyer LA, Young RA, Gifford DK, et al. 2013. Saltatory remodeling of Hox chromatin in response to rostrocaudal patterning signals. *Nat Neurosci* **16**: 1191–1198.

Metzis V, Steinhauser S, Pakanavicius E, Gouti M, Stamataki D, Ivanovitch K, Watson T, Rayon T, Mousavy Gharavy SN, Lovell-Badge R, et al. 2018. Nervous System Regionalization Entails Axial Allocation before Neural Differentiation. *Cell* **175**: 1105-1118.e17.

Miyagi S, Masui S, Niwa H, Saito T, Shimazaki T, Okano H, Nishimoto M, Muramatsu M, Iwama A, Okuda A. 2008. Consequence of the loss of Sox2 in the developing brain of the mouse. *FEBS Lett* **582**: 2811–2815.

Miyagi S, Nishimoto M, Saito T, Ninomiya M, Sawamoto K, Okano H, Muramatsu M, Oguro H, Iwama A, Okuda A. 2006. The Sox2 Regulatory Region 2 Functions as a Neural Stem Cell-specific Enhancer in the Telencephalon. *J Biol Chem* **281**: 13374–13381.

Miyagi S, Saito T, Mizutani K, Masuyama N, Gotoh Y, Iwama A, Nakauchi H, Masui S, Niwa H, Nishimoto M, et al. 2004. The Sox-2 Regulatory Regions Display Their Activities in Two Distinct Types of Multipotent Stem Cells. *Mol Cell Biol* **24**: 4207–4220.

Mlynarczyk-Evans S, Royce-Tolland M, Alexander MK, Andersen AA, Kalantry S, Gribnau J, Panning B. 2006. X chromosomes alternate between two states prior to random X-inactivation. *PLoS Biol* **4**: e159.

Molofsky AV, Pardal R, Iwashita T, Park I-K, Clarke MF, Morrison SJ. 2003. Bmi-1 dependence distinguishes neural stem cell self-renewal from progenitor proliferation. *Nature* **425**: 962–967.

Moncho-Amor V, Chakravarty P, Galichet C, Matheu A, Lovell-Badge R, Rizzoti K. 2021. SOX2 is required independently in both stem and differentiated cells for pituitary tumorigenesis in p27-null mice. *Proc Natl Acad Sci* **118**: e2017115118.

Moorthy SD, Davidson S, Shchuka VM, Singh G, Malek-Gilani N, Langroudi L, Martchenko A, So V, Macpherson NN, Mitchell JA. 2017. Enhancers and super-enhancers have an equivalent regulatory role in embryonic stem cells through regulation of single or multiple genes. *Genome Res* **27**: 246–258.

Okabe S, Forsberg-Nilsson K, Spiro AC, Segal M, McKay RD. 1996. Development of neuronal precursor cells and functional postmitotic neurons from embryonic stem cells in vitro. *Mech Dev* **59**: 89–102.

Osterwalder M, Barozzi I, Tissières V, Fukuda-Yuzawa Y, Mannion BJ, Afzal SY, Lee EA, Zhu Y, Plajzer-Frick I, Pickle CS, et al. 2018. Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* **554**: 239–243.

Pagin M, Pernebrink M, Giubbolini S, Barone C, Sambruni G, Zhu Y, Chiara M, Ottolenghi S, Pavesi G, Wei C-L. 2021. Sox2 controls neural stem cell self-renewal through a Fos-centered gene regulatory network. *Stem Cells* **39**: 1107–1119.

Phillips-Cremins JE, Sauria MEG, Sanyal A, Gerasimova TI, Lajoie BR, Bell JSK, Ong C-T, Hookway TA, Guo C, Sun Y, et al. 2013. Architectural Protein Subclasses Shape 3D Organization of Genomes during Lineage Commitment. *Cell* **153**: 1281–1295.

Platania A, Erb C, Barbieri M, Molcrette B, Grandgirard E, de Kort MA, Meaburn K, Taylor T, Shchuka VM, Kocanova S. 2023. Competition between transcription and loop extrusion modulates promoter and enhancer dynamics. *BioRxiv* 2023–04.

Que J, Okubo T, Goldenring JR, Nam K-T, Kurotani R, Morrisey EE, Taranova O, Pevny LH, Hogan BLM. 2007. Multiple dose-dependent roles for Sox2 in the patterning and differentiation of anterior foregut endoderm. *Development* **134**: 2521–2531.

Quevedo M, Meert L, Dekker MR, Dekkers DHW, Brandsma JH, van den Berg DLC, Ozgür Z, van IJcken WFJ, Demmers J, Fornerod M, et al. 2019. Mediator complex interaction partners organize the transcriptional network that defines neural stem cells. *Nat Commun* **10**: 2669.

R Core Team. 2020. R Core Team R: a language and environment for statistical computing. *Found Stat Comput*.

Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. 2011. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**: 279–283.

Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. 2014. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**: W187–W191.

Reynolds BA, Weiss S. 1992. Generation of Neurons and Astrocytes from Isolated Cells of the Adult Mammalian Central Nervous System. *Science* **255**: 1707–1710.

Rhodes CT, Thompson JJ, Mitra A, Asokumar D, Lee DR, Lee DJ, Zhang Y, Jason E, Dale RK, Rocha PP, et al. 2022. An epigenome atlas of neural progenitors within the embryonic mouse forebrain. *Nat Commun* **13**: 4196.

Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, et al. 2015. Integrative analysis of 111 reference human epigenomes. *Nature* **518**: 317–330.

Rosenberg AB, Roco CM, Muscat RA, Kuchina A, Sample P, Yao Z, Graybuck LT, Peeler DJ, Mukherjee S, Chen W, et al. 2018. Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* **360**: 176–182.

Sabapathy K, Jochum W, Hochedlinger K, Chang L, Karin M, Wagner EF. 1999. Defective neural tube morphogenesis and altered apoptosis in the absence of both JNK1 and JNK2. *Mech Dev* **89**: 115–124.

Sanyal A, Lajoie BR, Jain G, Dekker J. 2012. The long-range interaction landscape of gene promoters. *Nature* **489**: 109–113.

Schoenfelder S, Furlan-Magaril M, Mifsud B, Tavares-Cadete F, Sugar R, Javierre B-M, Nagano T, Katsman Y, Sakthidevi M, Wingett SW, et al. 2015. The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Res* **25**: 582–597.

Shen Q, Wang Y, Dimos JT, Fasano CA, Phoenix TN, Lemischka IR, Ivanova NB, Stifani S, Morrisey EE, Temple S. 2006. The timing of cortical neurogenesis is encoded within lineages of individual progenitor cells. *Nat Neurosci* **9**: 743–751.

Singh G, Mullany S, Moorthy SD, Zhang R, Mehdi T, Tian R, Duncan AG, Moses AM, Mitchell JA. 2021. A flexible repertoire of transcription factor binding sites and a diversity threshold determines enhancer activity in embryonic stem cells. *Genome Res* **31**: 564–575.

Soleimani VD, Yin H, Jahani-Asl A, Ming H, Kockx CE, van Ijcken WF, Grosveld F, Rudnicki MA. 2012. Snail regulates MyoD binding-site occupancy to direct enhancer switching and differentiation-specific transcription in myogenesis. *Mol Cell* **47**: 457–468.

Stark R, Brown G. 2011. DiffBind: differential binding analysis of ChIP-Seq peak data. *R Package Version* **100**.

Suh H, Consiglio A, Ray J, Sawai T, D'Amour KA, Gage FH. 2007. In Vivo Fate Analysis Reveals the Multipotent and Self-Renewal Capacities of Sox2+ Neural Stem Cells in the Adult Hippocampus. *Cell Stem Cell* **1**: 515–528.

Suter DM, Tirefort D, Julien S, Krause K-H. 2009. A Sox1 to Pax6 switch drives neuroectoderm to radial glia progression during differentiation of mouse embryonic stem cells. *Stem Cells Dayt Ohio* **27**: 49–58.

Takemoto T, Uchikawa M, Kamachi Y, Kondoh H. 2006. Convergence of Wnt and FGF signals in the genesis of posterior neural plate through activation of the Sox2 enhancer N-1. *Development* **133**: 297–306.

Takemoto T, Uchikawa M, Yoshida M, Bell DM, Lovell-Badge R, Papaioannou VE, Kondoh H. 2011. Tbx6-dependent Sox2 regulation determines neural or mesodermal fate in axial stem cells. *Nature* **470**: 394–398.

Tang Z, Luo OJ, Li X, Zheng M, Zhu JJ, Szalaj P, Trzaskoma P, Magalska A, Wlodarczyk J, Ruszczycki B, et al. 2015. CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell* **163**: 1611–1627.

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

Taylor T, Sikorska N, Shchuka VM, Chahar S, Ji C, Macpherson NN, Moorthy SD, Kort MAC de, Mullany S, Khader N, et al. 2022. Transcriptional regulation and chromatin architecture maintenance are decoupled functions at the Sox2 locus. *Genes Dev* **36**: 699–717.

Thompson JJ, Lee DJ, Mitra A, Frail S, Dale RK, Rocha PP. 2022. Extensive co-binding and rapid redistribution of NANOG and GATA6 during emergence of divergent lineages. *Nat Commun* **13**: 4257.

Thomson M, Liu SJ, Zou L-N, Smith Z, Meissner A, Ramanathan S. 2011. Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. *Cell* **145**: 875–889.

Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B. 2012. The accessible chromatin landscape of the human genome. *Nature* **489**: 75–82.

Tobias IC, Abatti LE, Moorthy SD, Mullany S, Taylor T, Khader N, Filice MA, Mitchell JA. 2021. Transcriptional enhancers: from prediction to functional assessment on a genome-wide scale. *Genome* **64**: 426–448.

Tolhuis B, Palstra R-J, Splinter E, Grosveld F, de Laat W. 2002. Looping and Interaction between Hypersensitive Sites in the Active β-globin Locus. *Mol Cell* **10**: 1453–1465.

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

Tuan DY, Solomon WB, London IM, Lee DP. 1989. An erythroid-specific, developmental-stage-independent enhancer far upstream of the human "beta-like globin" genes. *Proc Natl Acad Sci* **86**: 2554–2558.

Uchikawa M, Ishida Y, Takemoto T, Kamachi Y, Kondoh H. 2003. Functional Analysis of Chicken Sox2 Enhancers Highlights an Array of Diverse Regulatory Elements that Are Conserved in Mammals. *Dev Cell* **4**: 509–519.

Uchikawa M, Kondoh H. 2016. Regulation of Sox2 via Many Enhancers of Distinct Specificities. In *Sox2*, pp. 107–129, Elsevier https://linkinghub.elsevier.com/retrieve/pii/B9780128003527000074 (Accessed March 12, 2023).

Villar D, Berthelot C, Aldridge S, Rayner TF, Lukk M, Pignatelli M, Park TJ, Deaville R, Erichsen JT, Jasinska AJ. 2015. Enhancer evolution across 20 mammalian species. *Cell* **160**: 554–566.

Wamstad JA, Alexander JM, Truty RM, Shrikumar A, Li F, Eilertson KE, Ding H, Wylie JN, Pico AR, Capra JA, et al. 2012. Dynamic and Coordinated Epigenetic Regulation of Developmental Transitions in the Cardiac Lineage. *Cell* **151**: 206–220.

Weintraub AS, Li CH, Zamudio AV, Sigova AA, Hannett NM, Day DS, Abraham BJ, Cohen MA, Nabet B, Buckley DL, et al. 2017. YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell* **171**: 1573-1588.e28.

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J. 2019. Welcome to the Tidyverse. *J Open Source Softw* **4**: 1686.

Wijgerde M, McMahon JA, Rule M, McMahon AP. 2002. A direct requirement for Hedgehog signaling for normal specification of all ventral progenitor domains in the presumptive mammalian spinal cord. *Genes Dev* **16**: 2849–2864.

Ying Q-L, Stavridis M, Griffiths D, Li M, Smith A. 2003. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat Biotechnol* **21**: 183–186.

Zappone MV, Galli R, Catena R, Meani N, Biasi SD, Mattei E, Tiveron C, Vescovi AL, Lovell-Badge R, Ottolenghi S, et al. 2000. Sox2 regulatory sequences direct expression of a β-geo transgene to telencephalic neural stem cells and precursors of the mouse embryo, revealing regionalization of gene expression in CNS stem cells. *Development* **127**: 2367–2382.

Zeisel A, Hochgerner H, Lönnerberg P, Johnsson A, Memic F, van der Zwan J, Häring M, Braun E, Borm LE, La Manno G, et al. 2018. Molecular Architecture of the Mouse Nervous System. *Cell* **174**: 999-1014.e22.

Zhang J, Zhou Y, Yue W, Zhu Z, Wu X, Yu S, Shen Q, Pan Q, Xu W, Zhang R. 2022. Super-enhancers conserved within placental mammals maintain stem cell pluripotency. *Proc Natl Acad Sci* **119**: e2204716119.

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

Zhou HY, Katsman Y, Dhaliwal NK, Davidson S, Macpherson NN, Sakthidevi M, Collura F, Mitchell JA. 2014. A Sox2 distal enhancer cluster regulates embryonic stem cell differentiation potential. *Genes Dev* **28**: 2699–2711.

Zhu LJ, Gazin C, Lawson ND, Pagès H, Lin SM, Lapointe DS, Green MR. 2010. ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics* **11**: 1–10.

Ziller MJ, Edri R, Yaffe Y, Donaghey J, Pop R, Mallard W, Issner R, Gifford CA, Goren A, Xing J. 2015. Dissecting neural differentiation regulatory networks through epigenetic footprinting. *Nature* **518**: 355–359.

Zinzen RP, Girardot C, Gagneur J, Braun M, Furlong EEM. 2009. Combinatorial binding predicts spatio-temporal cis-regulatory activity. *Nature* **462**: 65–70.

Figures

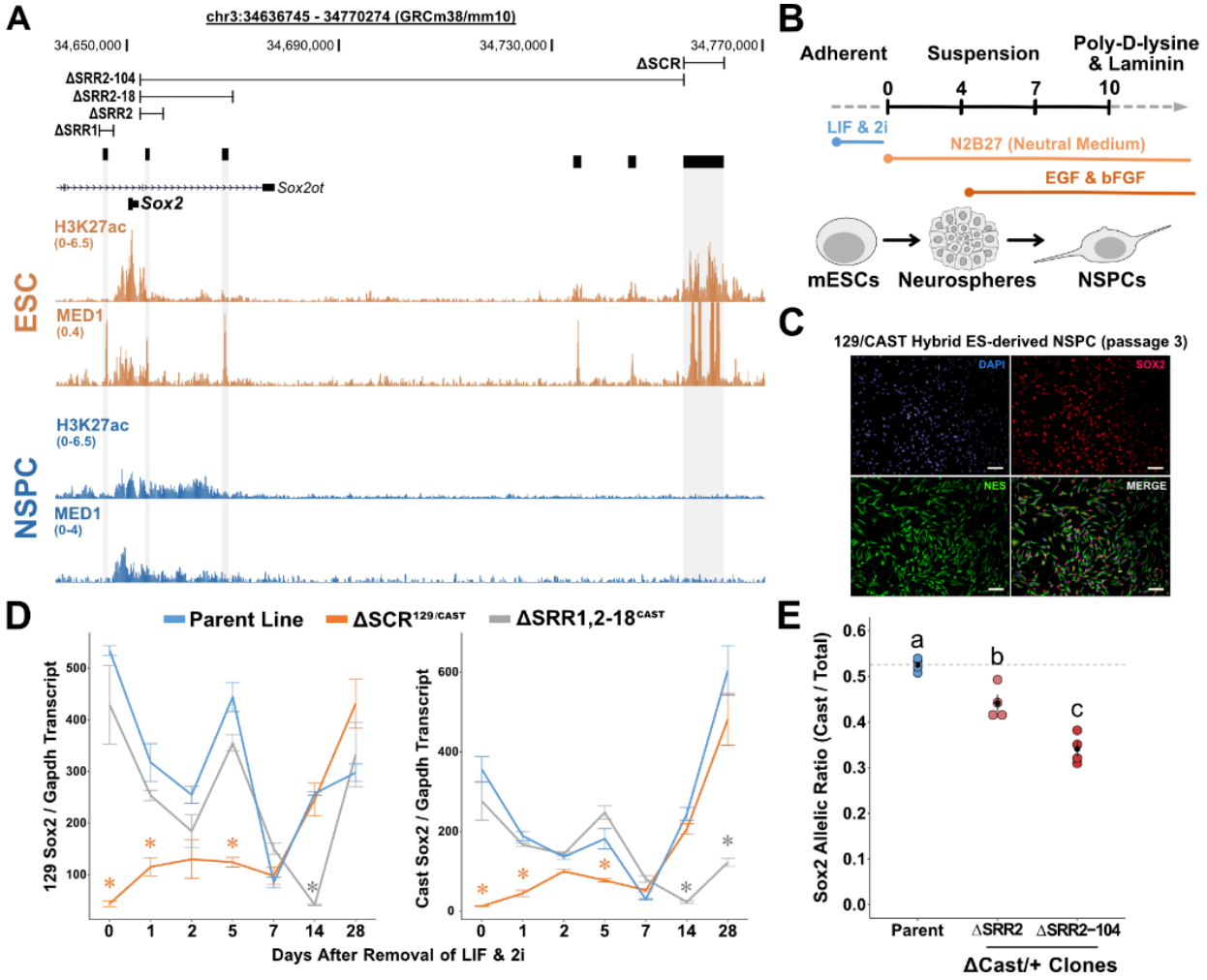

Fig 1. | Enhancers regulate *Sox2* dynamics during neural differentiation of mouse ESCs. (A) H3K27ac and MED1 ChIP-seq tracks in mouse embryonic stem cells (mESCs; orange) and neural stem/progenitor cells (NSPCs; blue) visualized by UCSC Genome Browser (mm10). Shaded bars mark *Sox2* regulatory regions (SRRs) active in mESCs. H3K27ac and MED1 co-enrichment at *cis*-acting enhancers are located upstream and downstream of *Sox2*. Ranges show the margins of CRISPR-Cas9-mediated deletions of different SRRs or the *Sox2* control region (SCR). (B) Schematic of ESC differentiation following 2i/LIF withdrawal, aggregate suspension culture in neutral N2B27 medium, and directed differentiation into NSPCs. (C) Immunofluorescence of NES overlayed with SOX2 in passage 3 adherent NSPCs. Scale bar distance is 50 µm. In D-E, *Sox2* expression in differentiating ESCs or stably derived NSPCs by allele-specific RT-qPCR comparing wildtype lines to lines harboring monoallelic deletions of enhancers with the indicated genotypes. Expression levels for each allele are shown relative to the total transcript levels. Error bars represent SEM. n ≥ 3. (D) Significant differences from the day 0 timepoint determined from Wilcoxon tests are indicated (∗) $P < 0.05$, (∗∗) $P < 0.01$, (∗∗∗) $P < 0.001$, (∗∗∗∗)

$P < 0.0001$. (E) Groups determined by Tukey's Honestly Squared Difference tests to be significantly different ($P < 0.05$) from one another are labeled with different letters; to indicate $P > 0.05$, groups would be labeled with the same letter.

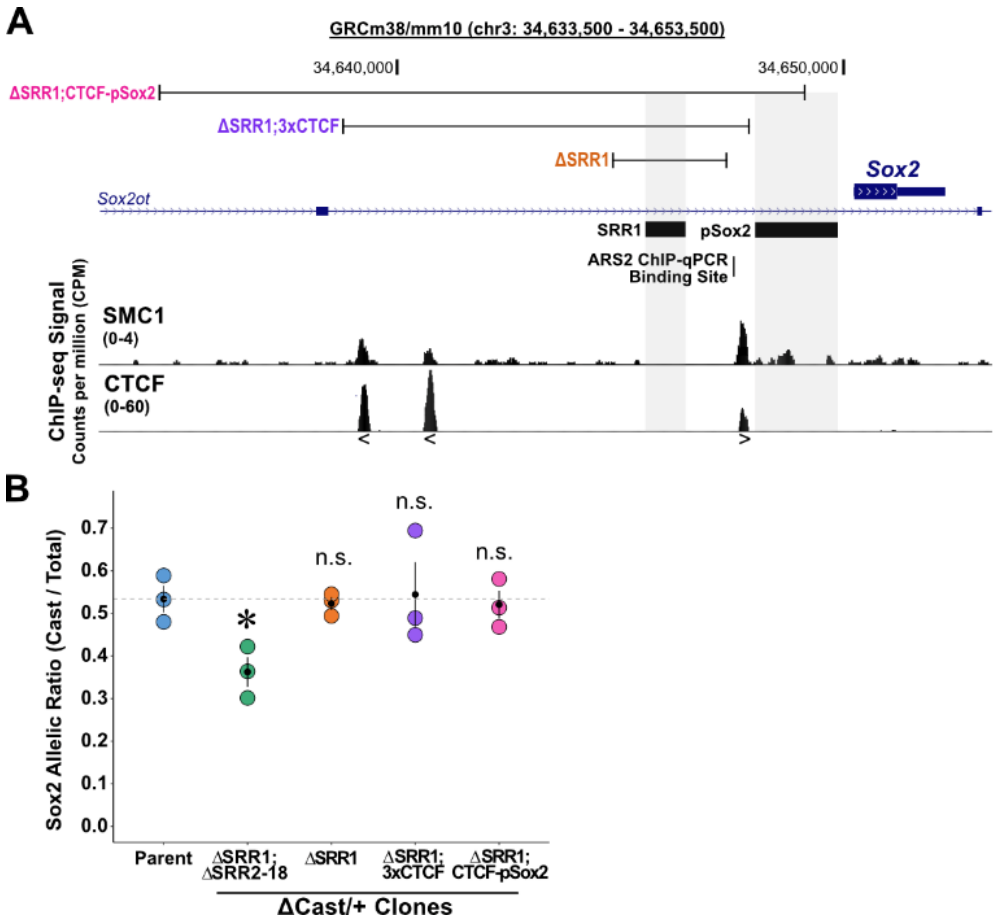

Fig 2. | SRR1 is dispensable for *Sox2* expression in mouse NSPCs. (A) SMC1 and CTCF ChIP-seq tracks visualized by the UCSC Genome Browser (mm10). SMC1/CTCF enrichment at elements that control chromatin topology are located upstream of *Sox2*. Ranges show the margins of CRISPR-Cas9 -mediated deletions of SRR1, clustered CTCF binding sites (3x CTCF), zinc finger protein motifs (ZNF) or the *Sox2* promoter (pSox2). Angled brackets indicate CTCF motif orientation. (B) *Sox2* expression from the Cast allele in ESC-derived NSPCs by allele-specific RT-qPCR comparing wildtype lines to lines harboring monoallelic deletions of enhancers with the indicated genotypes. Expression levels for each allele are shown relative to the total transcript level. Error bars represent SEM. n ≥ 3. Significant differences from the parent line values by Wilcoxon tests are denoted (∗) P < 0.05, (∗∗) P < 0.01, (∗∗∗) P < 0.001, (∗∗∗∗) P < 0.0001, (n.s.) not significant.

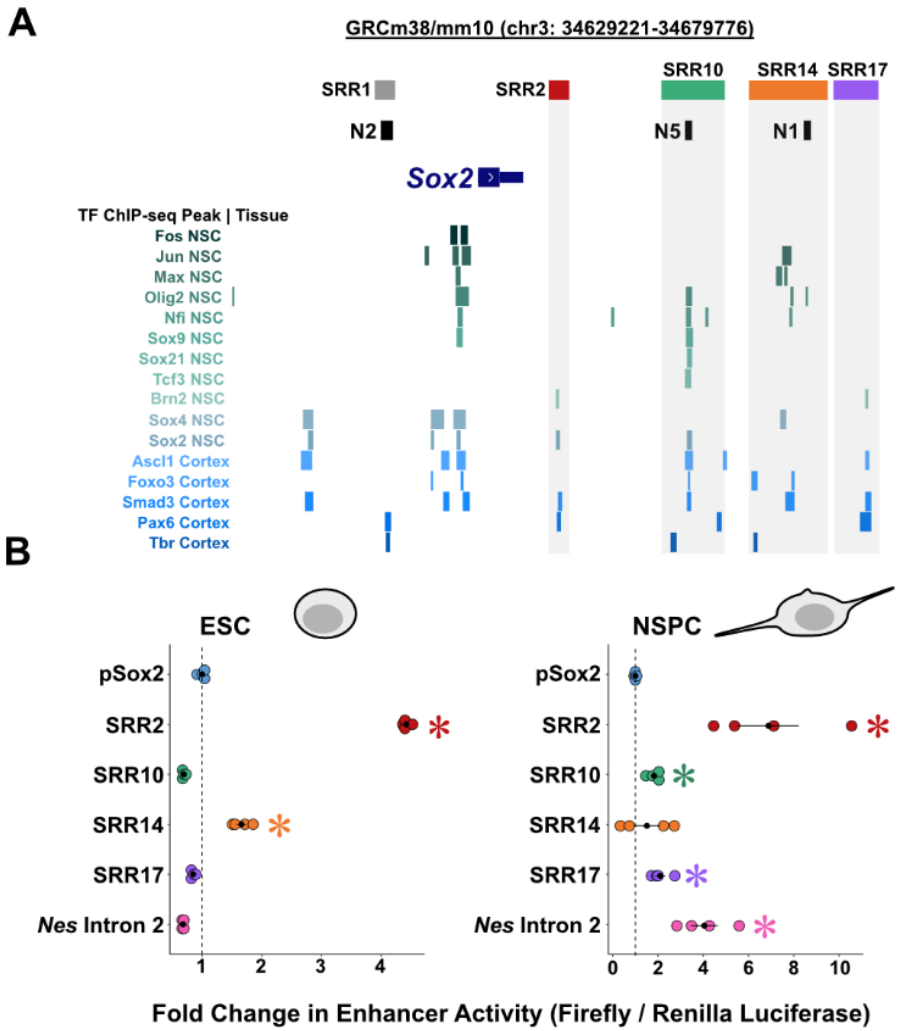

Fig 3. | SRR2-18 contains multiple NSPC enhancers. (A) The SRR2-18 genomic region is displayed on the UCSC Genome Browser (mm10). *Sox2* regulatory regions (SRRs, top) correspond to transcription factor-bound regions derived from NSC and embryonic cerebral cortex ChIP-seq data sets compiled from public databases (bottom). (B) Clusters of transcription factor binding sites in neural cells identifying distinct elements capable of activating transcription in luciferase reporter assays. Dashed line indicates mean enhancer activity of the *Sox2* promoter (pSox2) alone. The *Nes* enhancer located in the second intron of the full-length *Nes* gene served as a positive control for neural progenitor regulatory activity. Error bars represent SEM. n ≥ 3. Significant differences from the pSox2 values by Wilcoxon tests are indicated. (∗) $P < 0.05$, (∗∗) $P < 0.01$, (∗∗∗) $P < 0.001$, (∗∗∗∗) $P < 0.0001$, (ns) not significant.

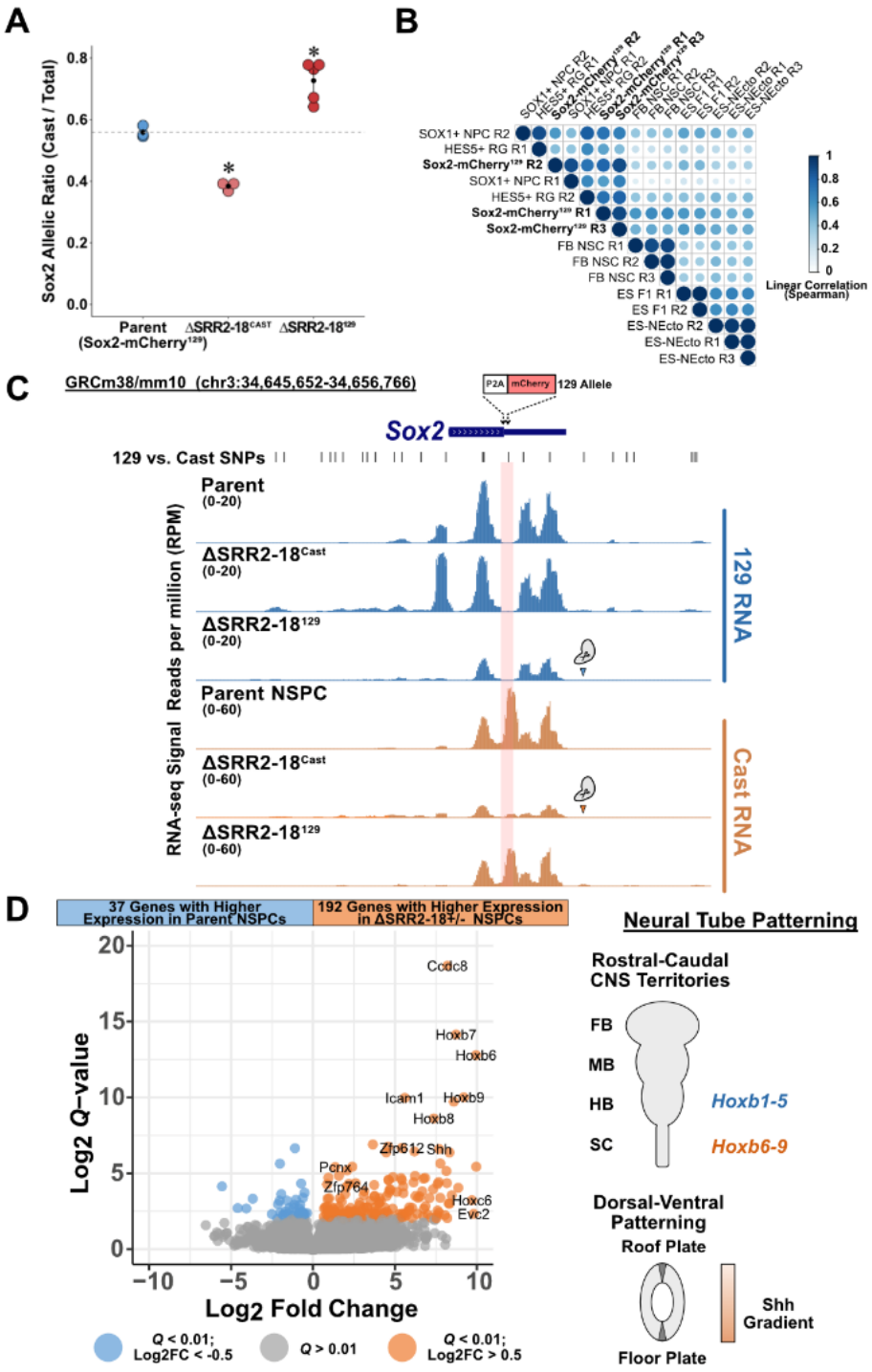

Fig 4. | Monoallelic SRR2-18 deletion affects *Sox2* allelic dosage and positional identity in NSPCs. (A) *Sox2* expression from the Cast allele in ESC-derived NSPCs by allele-specific RT-qPCR comparing the parent *Sox2*-mCherry^{129} lines to lines harboring monoallelic deletions of the SRR2-18 genomic region with the indicated genotypes. Expression levels for each allele are shown relative to the total transcript levels. Error bars represent SD. n ≥ 3. Significant differences from the parent line values by Welch's unequal variance t-test are indicated. (∗) P < 0.05, (∗∗) P < 0.01, (∗∗∗) P < 0.001, (∗∗∗∗) P < 0.0001, (ns)

not significant. (B) Correlogram showing Pearson correlations (represented by circle size and color) of normalized transcript counts in bulk and TF-expressing subsets of neuroepithelial and neural progenitor cells from RNA-seq. *Sox2*-tagged hybrid NSPCs compared to F1 hybrid ESCs (F1 ES), wildtype ESC-derived neurectoderm (ES-NEcto), forebrain-derived primary NSC (FB NSC), *Sox1*-expressing neural progenitors (SOX1 NPCs), and *Hes5*-expressing radial glia-like cells (HES5+ RG) compiled from the GEO repository. (C) Allele-sorted RNA-seq read pileup over the *Sox2* gene body, displayed on the UCSC Genome Browser (mm10). Red shaded bar indicates discriminatory 129/Cast single nucleotide polymorphism (SNP) on the 129 allele that is destroyed by the P2A-mCherry tag. Globular scissor protein icon denotes allele harboring a deletion of the SRR2-18 genomic region. (D) Volcano plot with the displayed number of differentially expressed genes between the parent and SRR2-18$^{+/-}$ NSPCs. Within table, number of genes that passed the cutoffs of absolute log2 fold-change > 1 and $Q < 0.01$. Q values represent the adjusted *P* values computed with the Benjamini & Hochberg method for controlling false discovery rate (FDR). *P* values were extracted from a negative binomial regression model by Wald Chi-Squared tests. (E) Schematic of the dorsoventral sonic hedgehog (Shh) signaling gradient and HoxB cluster expression domains along the rostrocaudal axis in the vertebrate neural tube. CNS regions annotated: forebrain (FB), midbrain (MB), hindbrain (HB), and spinal cord (SC).

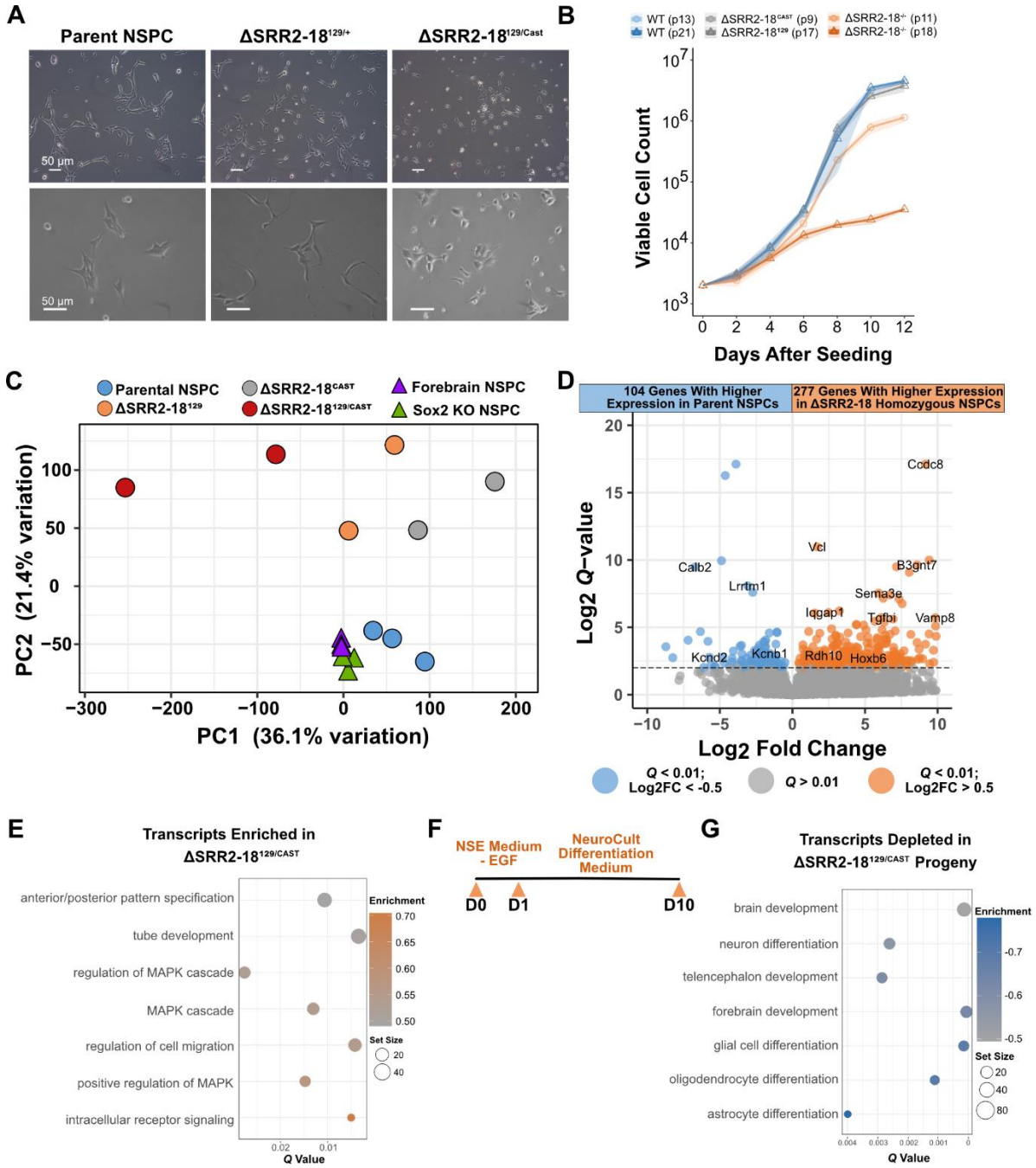

Fig 5. | SRR2-18 is required for anterior NSPC fates. (A) Phase contrast micrographs displaying representative NSPC morphologies with Sox2-mCherry^{129}, SRR2-18$^{+/-}$, and SRR2-18$^{-/-}$ genotypes. Scale bar distance is 50 µm. (B) Cell proliferation curves from different clonal NSPC lines and passage numbers showing differences in acute growth observed in SRR2-18$^{-/-}$ NSPCs. (C) Biplot of principle component analysis comparing RNA-seq data of parent and enhancer-deleted NSPCs to wild-type and *Sox2* knockout (KO) neurosphere RNA-seq data compiled from the GEO repository. (D) Volcano plot with the displayed number of differentially expressed genes between the parent and SRR2-18$^{-/-}$ NSPCs that

passed the cutoffs of absolute log2 fold-change > 1 and $Q < 0.01$. *P* values were extracted from a negative binomial regression model by Wald Chi-Squared tests. (E) Bubble plot of the significantly enriched gene sets of upregulated genes in SRR2-18$^{-/-}$ vs. control NSPCs. (F) Schematic of NSPC differentiation following sequential EGF and FGF2 withdrawal and successive undirected differentiation in a serum-containing supplement until differentiation day 10. (G) Bubble plot of the depleted gene sets of upregulated genes in derivatives of SRR2-18$^{-/-}$ vs. control NSPCs generated by 10 days of *in vitro* differentiation. *Q* values represent the adjusted *P* values computed with the Benjamini & Hochberg method for controlling false discovery rate (FDR)

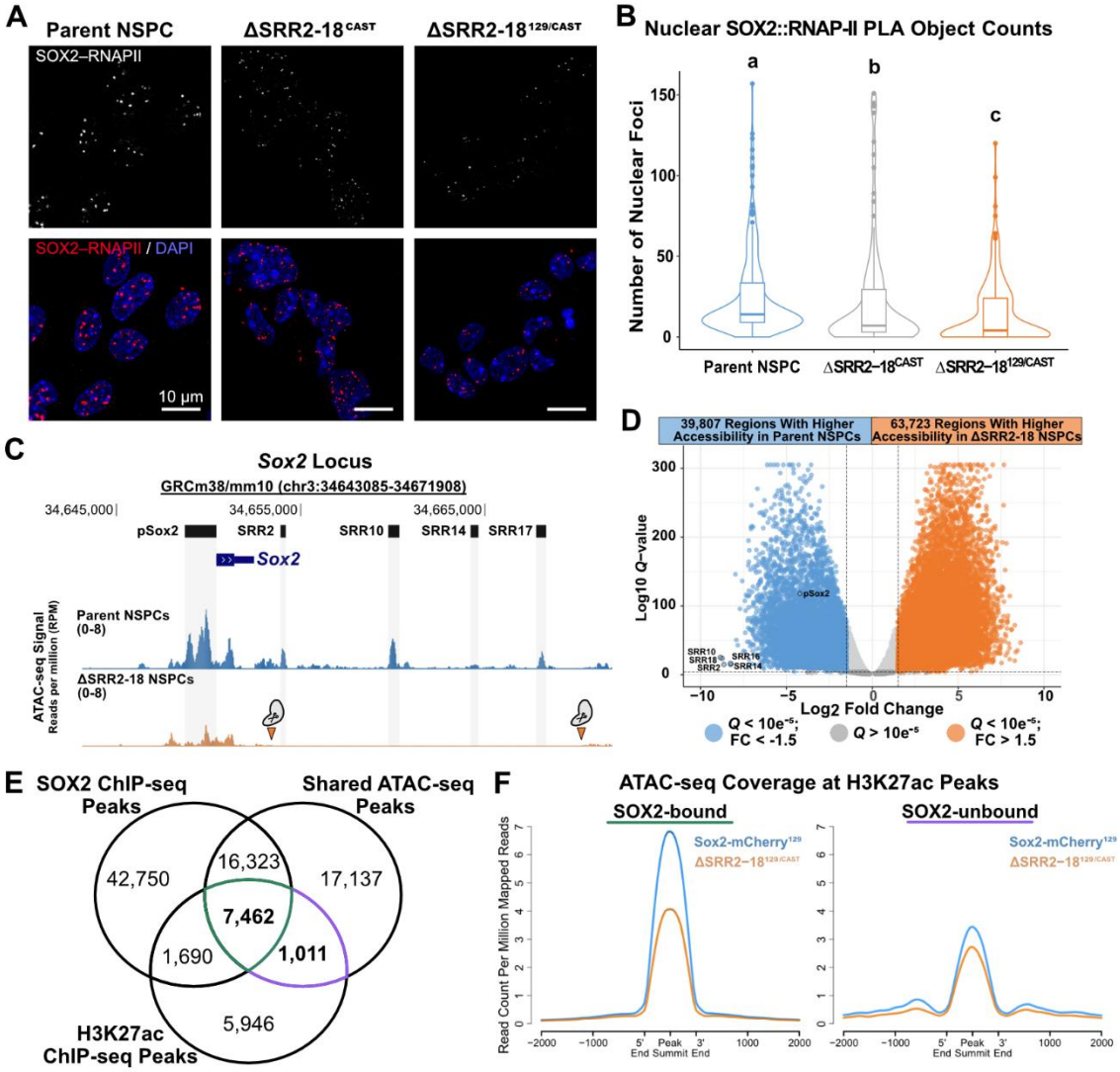

Fig 6. | Loss of SRR2-18 alters SOX2 transcriptional regulation. (A) Proximity ligation amplification (PLA) indicating the interaction between SOX2/RNAPII-S5P in parent and SRR2-18 deletion NSPCs. Images show median-intensity Z stack projections. Scale bar distance is 50 µm. (B) Box-and-whisker plots mark the number of PLA foci per nucleus. Boxes indicate interquartile range of intensity values and whiskers indicate the 10th and 90th percentiles. Outliers are represented by dots. Images were collected from at least three biological replicate samples and over 100 nuclei were quantified for each denoted genotype. Groups determined by Wilcoxon rank sum tests to be significantly different ($P < 0.05$) from one another are labeled with different letters; to indicate $P > 0.05$, groups would be labeled with the same letter. (C) The *Sox2* gene and SRR2-18 genomic regions displayed on the UCSC Genome Browser (mm10). *Sox2* promoter and regulatory regions (SRRs, top) in NPSCs and ATAC-seq tracks from parent

and SRR2-18$^{-/-}$ NSPCs (bottom). Globular scissor protein icon denotes the margins of a biallelic SRR2-18 deletion. (D) Volcano plot, with number of differentially accessible genomic regions between the parent and SRR2-18$^{-/-}$ NSPCs. Within table, number of genes that passed the cutoffs of absolute log2 fold-change > 1.5 and Q < 0.0001. *Q* values represent the adjusted *P* values computed with the Benjamini & Hochberg method for controlling false discovery rate (FDR). *P* values extracted from a negative binomial regression model by Wald Chi-Squared tests in DESeq2. (E) Shared accessible genomic regions between parent and SRR2-18$^{-/-}$ NSPCs divided into SOX2-bound, SOX2-unbound, H3K27ac-modified and non-H3K27ac modified subsets based on NSPC ChIP-seq data available in the GEO data depository. (F) Normalized ATAC-seq read coverage at SOX2-bound/H3K27ac and SOX2-unbound/H3K27ac accessible genomic regions between parent and SRR2-18$^{-/-}$ NSPCs.

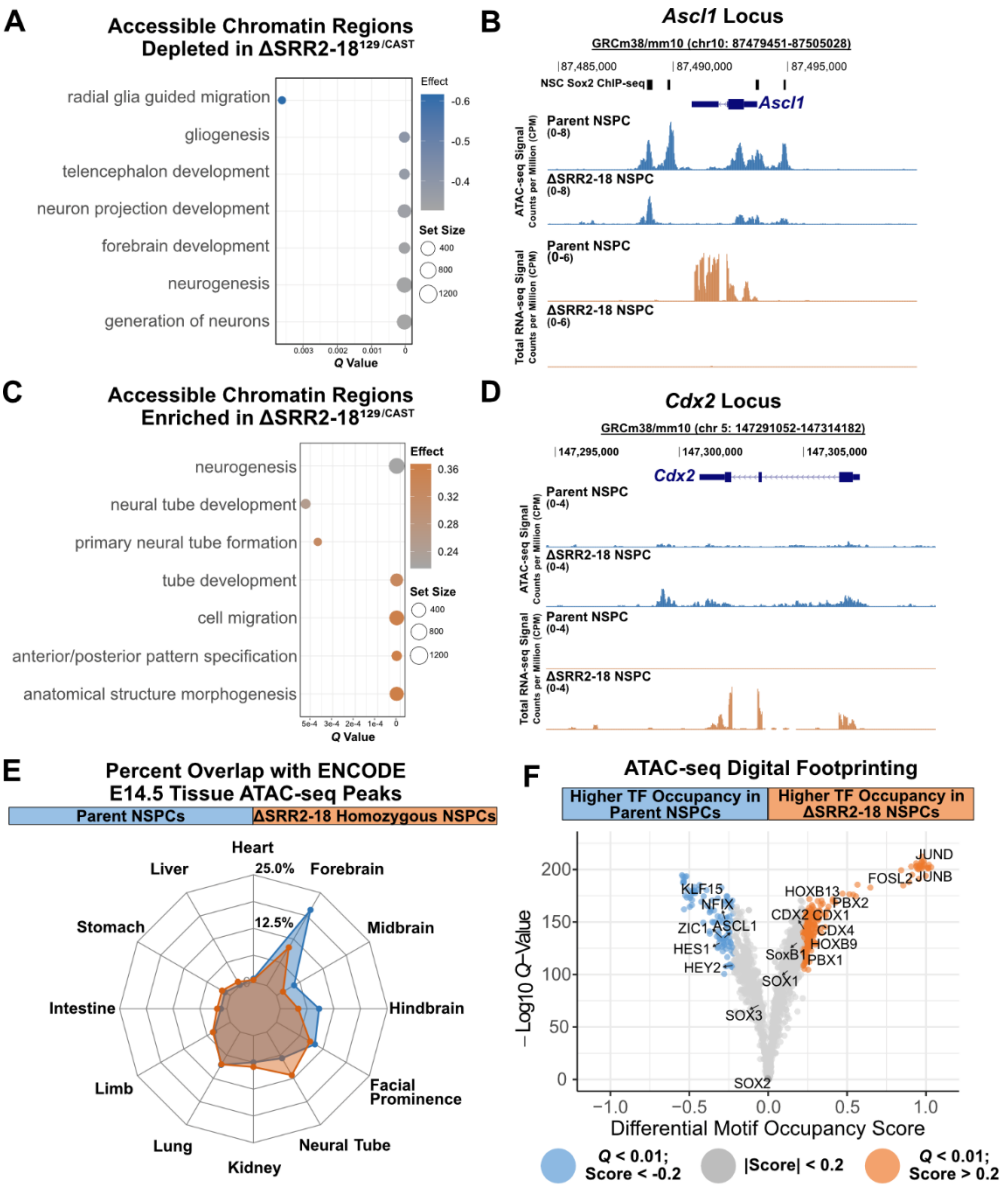

Fig 7. | Induction of caudal patterning factors and chromatin remodeling in NSPCs decoupled from SRR2-18 mediated regulation. (A) Bubble plot of the underrepresented genomic regions in the set of decreased accessibility regions in SRR2-18$^{-/-}$ vs. control NSPCs. (B) Normalized ATAC-seq and RNA-seq tracks over the *Ascl1/Mash1* gene and flanking genomic regions displayed on the UCSC Genome Browser (mm10). Signal tracks show reduced accessibility at the *Ascl1* promoter and decreased total *Ascl1* RNA in SRR2-18$^{-/-}$ NSPCs. (C) Bubble plot displaying enriched genomic regions in the set of increased accessibility regions in SRR2-18$^{-/-}$ vs. control NSPCs. (D) Normalized ATAC-seq and RNA-seq tracks over the *Cdx2* gene and flanking genomic regions displayed on the UCSC Genome Browser (mm10). Signal tracks show increased accessibility at the *Cdx2* locus and increased total *Cdx2* RNA in SRR2-18$^{-/-}$ NSPCs. (E) Radar diagram of accessible genomic regions called from ATAC-seq data in parent NSPCs (blue border) and SRR2-18$^{-/-}$ NSPCs (orange border) overlapping ENCODE consensus

ATAC-seq peaks from E14.5 tissue samples. Height of the borders at each tissue intersection corresponds to the percentage of non-promoter accessible genomic regions that overlap. (F) Volcano plot, with number of differentially accessible genomic regions between the parent and SRR2-18$^{-/-}$ NSPCs. Within table, number of genes that passed the cutoffs of absolute log2 fold-change > 1.5 and Q < 0.0001. *Q* values represent the adjusted *P* values computed with the Benjamini & Hochberg method for controlling false discovery rate (FDR).