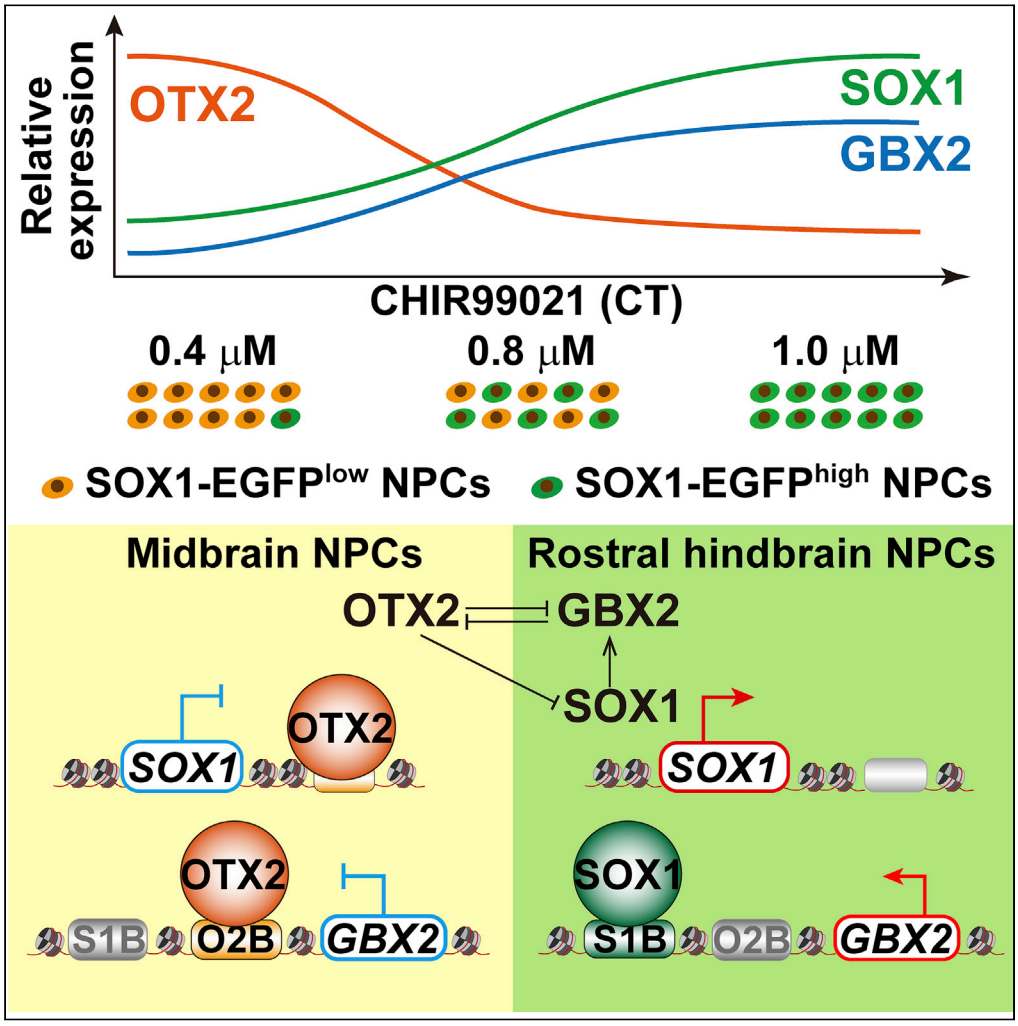
iScience

CellPress
OPEN ACCESS

Article

SOX1 Is Required for the Specification of Rostral Hindbrain Neural Progenitor Cells from Human Embryonic Stem Cells



Xinyuan Liu, Zhuoqing Fang, Jing Wen, ..., Naihe Jing, Dongmei Lai, Ying Jin

laidongmei@hotmail.com (D.L.)
yjin@sibs.ac.cn (Y.J.)

HIGHLIGHTS

SOX1 is highly expressed in rostral hindbrain NPCs derived from hESCs

OTX2 inhibits SOX1 expression in addition to its inhibition on GBX2 expression

SOX1 contributes to the specification of rostral hindbrain NPCs from hESCs

GBX2 is a key factor for SOX1 to function in the rostral hindbrain NPC specification

Liu et al., iScience 23, 101475
September 25, 2020
https://doi.org/10.1016/j.isci.2020.101475

iScience

CellPress
OPEN ACCESS

Article

SOX1 Is Required for the Specification of Rostral Hindbrain Neural Progenitor Cells from Human Embryonic Stem Cells

Xinyuan Liu,[1,7] Zhuoqing Fang,[1,7] Jing Wen,[1] Fan Tang,[2] Bing Liao,[2] Naihe Jing,[4,5] Dongmei Lai,[6,*] and Ying Jin[1,2,3,5,8,*]

SUMMARY

Region-specific neural progenitor cells (NPCs) can be generated from human embryonic stem cells (hESCs) by modulating signaling pathways. However, how intrinsic transcriptional factors contribute to the neural regionalization is not well characterized. Here, we generate region-specific NPCs from hESCs and find that SOX1 is highly expressed in NPCs with the rostral hindbrain identity. Moreover, we find that OTX2 inhibits SOX1 expression, displaying exclusive expression between the two factors. Furthermore, *SOX1* knockout (KO) leads to the upregulation of midbrain genes and downregulation of rostral hindbrain genes, indicating that SOX1 is required for specification of rostral hindbrain NPCs. Our SOX1 chromatin immunoprecipitation sequencing analysis reveals that SOX1 binds to the distal region of *GBX2* to activate its expression. Overexpression of *GBX2* largely abrogates *SOX1*-KO-induced aberrant gene expression. Taken together, this study uncovers previously unappreciated role of SOX1 in early neural regionalization and provides new information for the precise control of the OTX2/GBX2 interface.

INTRODUCTION

During embryonic development, the neuroectoderm develops into the forebrain, midbrain, hindbrain, and spinal cord along a rostral-caudal (R-C) axis. How neural regionalization is precisely controlled remains a critical unsolved issue. Insights from model animals show that the R-C identity of neural progenitor cells (NPCs) can be controlled by modulating WNT and retinoic acid (RA) signaling, and that the dorsal-ventral (D-V) identity can be controlled by modulating sonic hedgehog (Shh) signaling (Kiecker and Niehrs, 2001a, 2001b; Maden, 2007). These neural patterning principles have been applied to derive region-specific NPCs from human embryonic stem cells (hESCs). Recent advances permit highly efficient generation of NPCs from hESCs via dual inhibition of SMAD signaling (dSMADi) (Chambers et al., 2009; Fasano et al., 2010). Based on the dSMADi and aggregate formation strategy, several protocols were subsequently developed to generate regionally specified NPCs and neurons from hESCs (Imaizumi et al., 2015; Kirkeby et al., 2012; Kriks et al., 2011). The treatment with patterning factors, including the porcupine inhibitor IWP-2 (a WNT antagonist, hereafter referred to as IWP2), the glycogen synthase kinase 3 (GSK3) inhibitor CHIR99021 (a WNT agonist, hereafter referred to as CT), and RA, enables the generation of NPCs with distinct R-C identities from hESCs efficiently. A ventral identity can also be induced by the combined treatment with Shh proteins and purmorphamine (an SHH agonist).

It is obvious that extrinsic signals play important roles in the process of neural regionalization. Meanwhile, intrinsic transcriptional factors also serve as key determinants during this process. Otx2 and Gbx2 were shown to be required for specification of the forebrain-midbrain and anterior hindbrain, respectively (Joyner et al., 2000). *Otx2* and *Gbx2* are among the earliest genes expressed in the nervous system. *Otx2* is expressed within the forebrain and midbrain with a caudal limit at the midbrain-hindbrain boundary (MHB), whereas *Gbx2* is expressed in the hindbrain (Bally-Cuif et al., 1995a, 1995b; Bouillet et al., 1995; Shamim and Mason, 1998; Simeone et al., 1992). Some genetic analyses of *Otx2* and *Gbx2* in mice demonstrated that the generation of the *Otx2-Gbx2* border in the right place is important to position the MHB and favors the normal development of the midbrain and hindbrain (Acampora et al., 1995; Ang et al., 1996; Broccoli et al., 1999; Katahira et al., 2000; Matsuo et al., 1995; Millet et al., 1996, 1999; Simeone, 1998). However, how

[1]CAS Key Laboratory of Tissue Microenvironment and Tumor, Shanghai Institute of Nutrition and Health, CAS Center for Excellence in Molecular Cell Science, University of Chinese Academy of Sciences, Chinese Academy of Sciences, 320 Yueyang Road, Shanghai 200031, China

[2]Department of Histoembryology, Genetics and Developmental Biology, Shanghai Key Laboratory of Reproductive Medicine, Shanghai JiaoTong University School of Medicine, 225 South Chongqing Road, Shanghai 200025, China

[3]Basic Clinical Research Center, Renji Hospital, Shanghai JiaoTong University School of Medicine, 160 Pujian Road, Shanghai 200127, China

[4]State Key Laboratory of Cell Biology, CAS Center for Excellence in Molecular Cell Science, Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, University of Chinese Academy of Sciences, 320 Yue Yang Road, Shanghai 200031, China

[5]School of Life Science and Technology, ShanghaiTech University, 100 Haike Road, Shanghai 201210, China

[6]International Peace Maternity and Child Health Hospital, Shanghai JiaoTong University School of Medicine, Shanghai 200030, China

[7]These authors contributed equally

[8]Lead Contact

*Correspondence: laidongmei@hotmail.com (D.L.), yjin@sibs.ac.cn (Y.J.)

https://doi.org/10.1016/j.isci.2020.101475

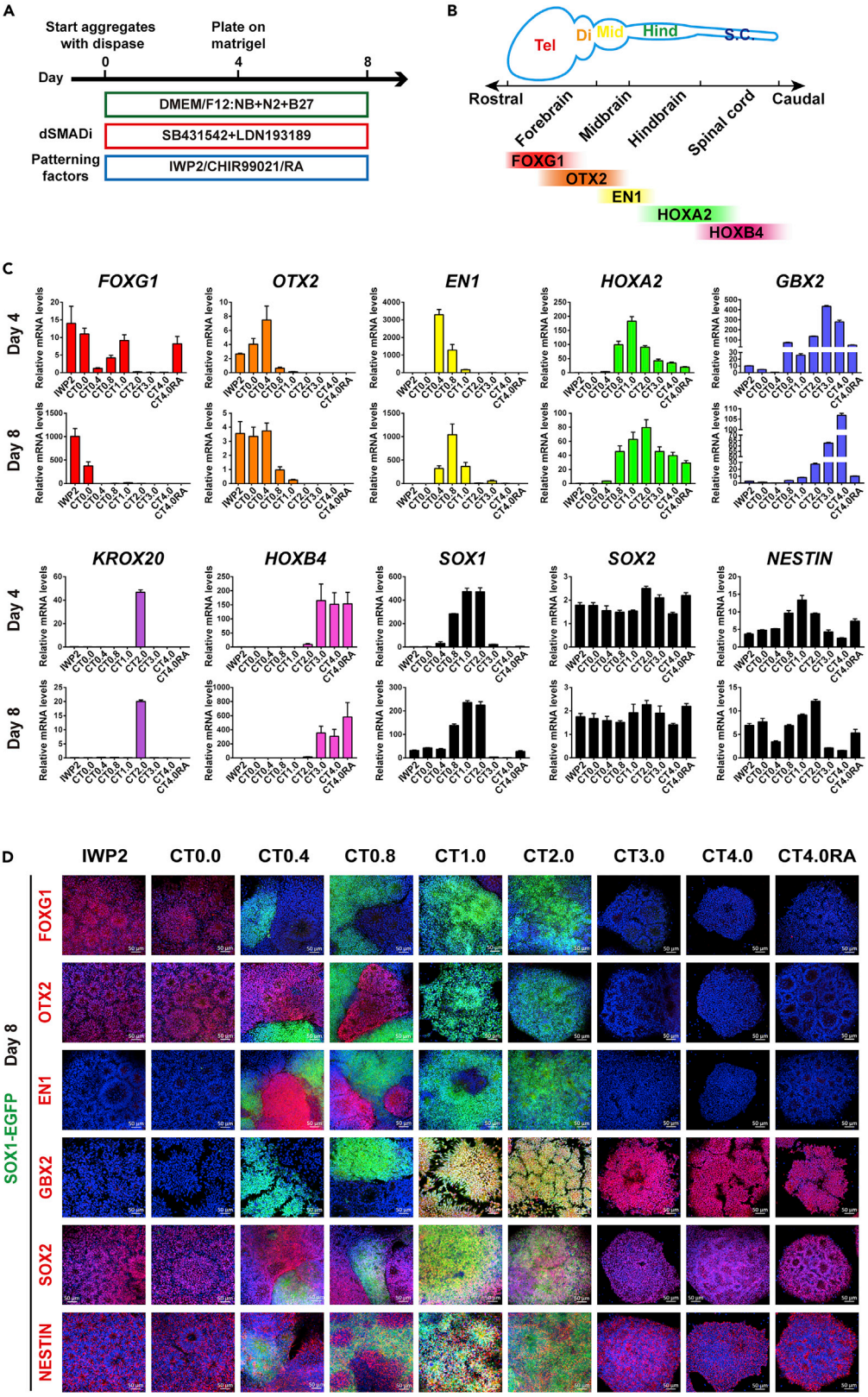

A
Start aggregates with dispase
Plate on matrigel
0
4
8
Day
DMEM/F12:NB+N2+B27
dSMADi
SB431542+LDN193189
Patterning factors
IWP2/CHIR99021/RA
B
Tel
Di
Mid
Hind
S.C.
Rostral
Caudal
Forebrain
Midbrain
Hindbrain
Spinal cord
FOXG1
OTX2
EN1
HOXA2
HOXB4
C
FOXG1
OTX2
EN1
HOXA2
GBX2
KROX20
HOXB4
SOX1
SOX2
NESTIN
Day 4
Day 8
Relative mRNA levels
IWP2
CT0.0
CT0.4
CT0.8
CT1.0
CT2.0
CT3.0
CT4.0
CT4.0RA
D
IWP2
CT0.0
CT0.4
CT0.8
CT1.0
CT2.0
CT3.0
CT4.0
CT4.0RA
SOX1-EGFP
Day 8
FOXG1
OTX2
EN1
GBX2
SOX2
NESTIN

Figure 1. SOX1 Was Highly Expressed in Rostral Hindbrain NPCs

(A) The illustration of our differentiation protocol. dSMADi, dual inhibition of SMAD signaling; NB, neurobasal; RA, retinoid acid. The following patterning factors were used to derive NPCs with R-C identities from hESCs: 2 μM -IWP2, 0–4.0 μM CHIR99021 (CT0.0-CT4.0), and 1 μM RA with 4.0 μM CT (CT4.0RA).

(B) A diagram showing representative R-C markers in the neural tube.

(C) qRT-PCR analysis of mRNA levels of R-C markers as well as *SOX1*, *SOX2*, and *NESTIN* in NPCs at days 4 and 8, relative to undifferentiated WT hESCs (day 0). n = 3 independent experiments. Data are shown as mean ± SEM.

(D) Representative results of immunofluorescence staining of NPCs derived from SOX1-EGFP reporter hESCs (clone #6) using antibodies against FOXG1, OTX2, EN1, GBX2, SOX2, and NESTIN, respectively, at day 8. The nucleus was stained by DAPI. Scale bars, 50 μm.

See also Figures S1 and S2.

OTX2 and GBX2 function and what upstream factors regulate their expression during human neural regionalization are poorly understood.

SOX1 is a member of the B1 group of SOX family transcription factors, harboring a high-mobility group DNA-binding domain (Malas et al., 1997). The embryonic expression pattern of Sox1 in the mouse has been well characterized (Aubert et al., 2003; Pevny et al., 1998; Uchikawa et al., 2011). The study by Pevny et al. reported that Sox1 is detected throughout the neural plate and early neural tube, marking the dividing neural epithelial cells within the embryonic neural tube (Pevny et al., 1998). Moreover, using *Sox1-gfp* knockin mice, Aubert et al. showed that, at E9.5, $Sox1^{GFP}$ is expressed along the entire neural tube but in no other tissue, suggesting a potential role of Sox1 for early neural development (Aubert et al., 2003). Moreover, it was reported that, when somitogenesis begins after E8, strong *Sox1* expression initiates in the closed posterior neural tube first, and then the anterior neural plate in mouse embryos (Uchikawa et al., 2011; Wood and Episkopou, 1999). This expression feature raises a question of whether Sox1 is involved with early neural regionalization. Homozygous mutant mice ($Sox1^{-/-}$) are viable but exhibit defects such as spontaneous seizures and microphthalmia (Malas et al., 2003; Nishiguchi et al., 1998). Compared with Sox1 studies in the mouse, little is known about the expression pattern of SOX1 during early human development and the function of SOX1 in the process of human neural regionalization. hESCs provide an attractive *in vitro* tool to address these issues. In the current study, with a goal to explore the function of SOX1 in the neural regionalization process of hESCs, we established SOX1-EGFP hESC reporter cell lines and generated NPCs with different regional identities from both wild-type (WT) and *SOX1* knockout (KO) hESCs. Our study uncovers that SOX1 is required for the specification of NPCs with the rostral hindbrain identity from hESCs primarily by directly promoting *GBX2* expression.

RESULTS

SOX1 Is Highly Expressed in Rostral Hindbrain NPCs

To investigate the molecular regulation during specification of region-specific NPCs from hESCs, we first generated NPCs with different regional identities from hESCs of the H9 line (Thomson et al., 1998) based on published protocols with minor modifications (Imaizumi et al., 2015; Kirkeby et al., 2012). According to the dSMADi strategy (Chambers et al., 2009), WNT and RA signaling pathways were modulated to generate the R-C identities of hESC-derived NPCs (Figures 1A and 1B). We then examined the expression of R-C markers in these NPCs at days 4 and 8 of differentiation by quantitative real-time PCR (qRT-PCR) (Figure 1C). On day 8, the telencephalic marker *FOXG1* (Xuan et al., 1995) was highly expressed in NPCs treated with IWP2 or 0.0 μM CT, whereas the expression of *OTX2* was high in NPCs treated with CT concentrations below 1.0 μM. *EN1*, which was reported to be detected in the midbrain and rostral hindbrain (Hanks et al., 1995), was detected in NPCs treated with 0.4–1.0 μM CT. In contrast, *GBX2*, known to be broadly expressed in the hindbrain and spinal cord (Bouillet et al., 1995; Luu et al., 2011), was detected in NPCs of CT1.0-CT4.0RA groups, and its level was especially high in NPCs treated with 3.0–4.0 μM CT. Moreover, *KROX20*, known to be expressed in the rhombomeres (r) 3 and 5 of the developing hindbrain (Schneider-Maunoury et al., 1993; Wilkinson et al., 1989), was only detected in NPCs treated with 2.0 μM CT. The expression of rostral hindbrain marker *HOXA2* (Prince and Lumsden, 1994) peaked in NPCs treated with 1.0–2.0 μM CT, whereas the caudal hindbrain marker *HOXB4* (Hunt et al., 1991) was highly expressed in NPCs treated with 3.0–4.0 μM CT. The combination of RA and 4.0 μM CT further elevated the expression level of *HOXB4*. Thus, we successfully established the *in vitro* model of hESC-derived NPCs with different R-C regional identities.

Interestingly, the expression of *SOX1* peaked in NPCs treated with a narrow CT concentration range (1.0–2.0 μM), a pattern similar to that of *HOXA2*. However, the expression of *SOX2* and *NESTIN* showed no

preference among NPCs of different groups (Figure 1C). The result could be reproduced in another independently derived hESC line, SHhES2 (Li et al., 2010), under the same differentiation condition (Figure S1A).

To verify the expression pattern of SOX1 along the R-C axis, we generated a SOX1-EGFP reporter hESC line by knocking in a P2A-EGFP-Neo cassette following the endogenous *SOX1* coding sequence (removing the *SOX1* stop codon) in H9 hESCs (Figure S1B), using the CRISPR/Cas9 nickase system with a pair of small guide RNAs (sgRNAs) targeting the *SOX1* sequence around its stop codon. The 2A self-cleaving sequences were used to generate a multicistronic reporter cassette that enables expression of EGFP and neomycin resistance gene flanked by the *loxP* sites (NeoR), allowing for the positive selection of correctly targeted clones. The excision of the NeoR cassette was catalyzed by Cre-induced recombination afterward. The correct integration was identified by genomic DNA PCR. Clone #6 hESCs were correctly targeted in one allele, whereas the non-targeted allele was intact. Heterozygous insertion of the reporter cassette was confirmed by genomic DNA PCR and Sanger sequencing (Figures S1C and S1D). The karyotype of clone #6 hESCs was normal (46, XX) (Figure S1E). We then performed flow cytometric analysis for NPCs derived from the SOX1-EGFP reporter hESCs and found that about 97% NPCs in the CT1.0 group exhibited high levels of both SOX1 and EGFP (Figure S1F), indicating the consistency between the expression of SOX1 and EGFP. In contrast, most NPCs in the CT0.0 group had low SOX1 and EGFP levels. As a control, SOX2 was highly expressed in both $EGFP^{low}$ NPCs of the CT0.0 group and $EGFP^{high}$ NPCs of the CT1.0 group (Figure S1F). These results support the notion that the SOX1-EGFP reporter hESCs could faithfully indicate the endogenous SOX1 expression.

Moreover, immunofluorescence staining analysis of expression patterns of some neural markers in NPCs derived from the SOX1-EGFP reporter hESCs (clone #6) further validated the SOX1 expression pattern and the robustness of our *in vitro* differentiation model at a protein level (Figure 1D). Intriguingly, we observed that SOX1-$EGFP^{low}$ and SOX1-$EGFP^{high}$ NPCs were present in the same dish when the CT concentration was relatively low (0.4 and 0.8 μM) and that EGFP-expressing NPCs and OTX2-expressing NPCs were mutually exclusive (Figure 1D). When the CT concentration increased to relatively high levels (1.0 and 2.0 μM), most NPCs became SOX1-$EGFP^{high}$. In contrast, expression of SOX2 and NESTIN was similar in NPCs of all groups, in line with their transcript levels (Figure 1C). As NPCs treated with 1.0 and 2.0 μM CT highly expressed rostral hindbrain genes, we considered them as rostral hindbrain NPCs for simplicity. NPCs treated with 0.4 μM CT were considered as midbrain NPCs due to their high expression of midbrain genes. Therefore, SOX1 was highly expressed in the rostral hindbrain NPCs.

In addition to the R-C identity, we also examined the expression of SOX1 in the D-V axis by modulating SHH signaling. When combined with SHH protein and purmorphamine (an SHH agonist) (termed the ''+SHH'' group) (Maroof et al., 2013), we could efficiently pattern NPCs toward a ventral fate (Figure S1G). In previous studies, PAX6 and PAX7 were used as the dorsal NPC markers, whereas NKX2.1 and NKX2.2 served as ventral NPC markers (Imaizumi et al., 2015; Kirkeby et al., 2012). Consistently, our qRT-PCR result indicated that the high expression of *PAX6* and *PAX7* was mainly found in NPCs of the "-SHH" group, although *PAX6* was also highly expressed in CT4.0RA NPCs of the "+SHH" group. *NKX2.1* and *NKX2.2* were detected in NPCs of the "+SHH'' group. Notably, SOX1 was only detected in NPCs of the ''-SHH'' group (Figure S1H). Thus, in the following experiments, we only studied the role of SOX1 in the R-C patterning process of NPCs with the dorsal identity.

To avoid the cell line bias, we generated additional two SOX1-EGFP reporter hESC lines using the same strategy. By genomic DNA PCR and Sanger sequencing, we identified hESCs of clones #4 and #8 having heterozygous insertion of the reporter cassette (Figures S2A and S2B). The flow cytometric analysis for NPCs derived from these two hESC clones showed that the NPCs in the CT1.0 group exhibited high levels of both SOX1 and EGFP, and the NPCs in the CT0.0 group had low levels of both SOX1 and EGFP. In contrast, SOX2 was highly expressed in both $EGFP^{low}$ NPCs of the CT0.0 group and $EGFP^{high}$ NPCs of CT1.0 group (Figure S2C). These results clearly show that these two hESC clones could faithfully reflect the expression of the endogenous SOX1. We also repeated the immunofluorescence staining analysis of OTX2 and GBX2 for NPCs derived from these two SOX1-EGFP reporter hESC clones, and results similar to those of NPCs derived from SOX1-EGFP reporter hESCs of clone #6 were obtained (Figure S2D). Thus, we reveal the unique expression pattern of SOX1 during early neural regionalization of hESCs and generate the SOX1-EGFP reporter hESC lines.

OTX2 Inhibits SOX1 Expression

As mentioned earlier, both SOX1-$EGFP^{low}$ and SOX1-$EGFP^{high}$ NPCs could be present in the same dish or even in the same aggregate when the CT concentration was 0.8 μM (Figure 2A). To determine the

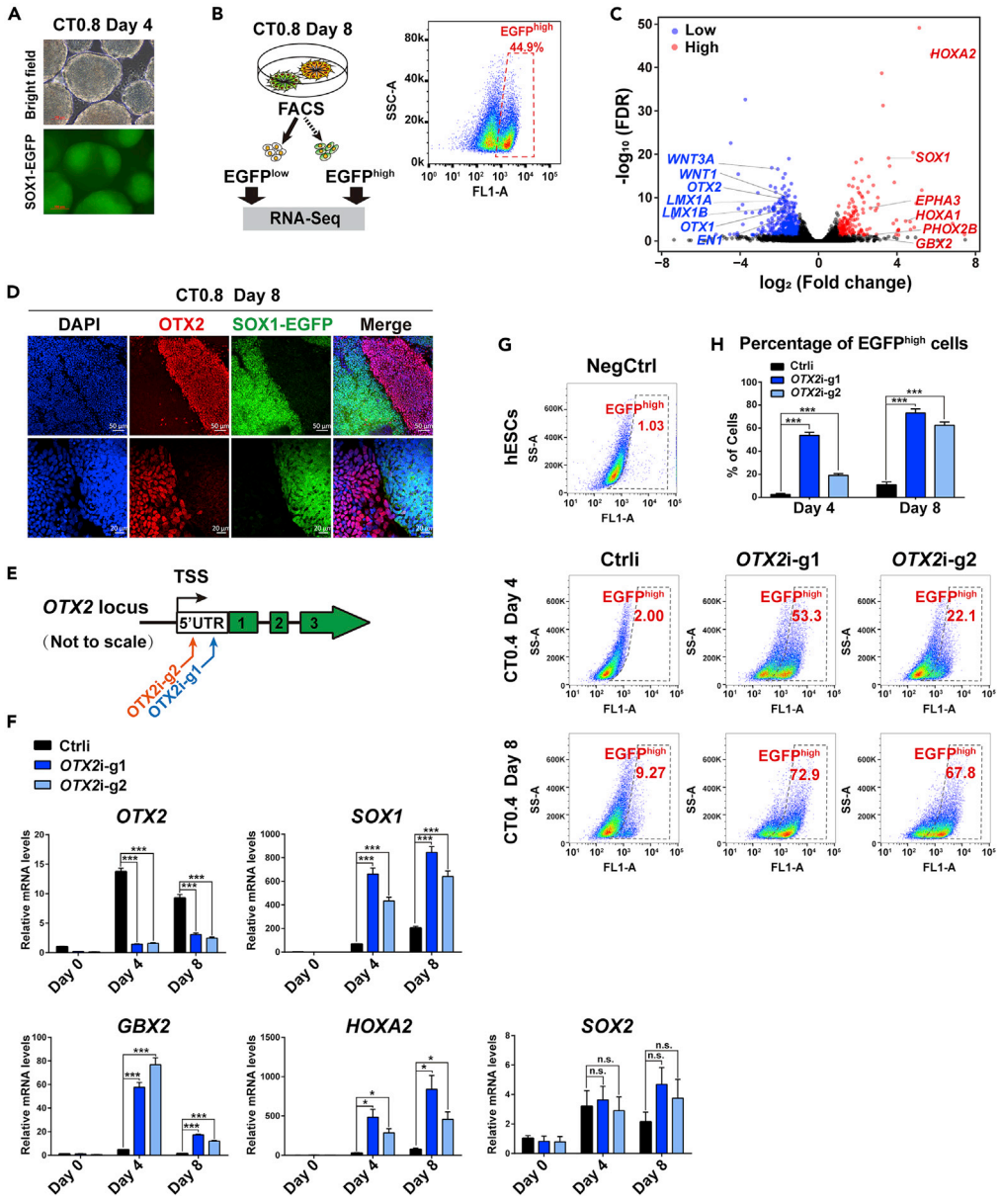


Figure 2. OTX2 Negatively Regulates SOX1 Expression

(A) Representative images of aggregates at day 4 treated with 0.8 μM CT. Scale bars, 100 μm.

(B) The design of sorting NPCs treated with 0.8 μM CT at day 8 into EGFP^{high} and EGFP^{low} subpopulations by fluorescence-activated cell sorting (FACS) for RNA-seq.

(C) The RNA-seq volcano plot is shown. The blue and red dots indicate the DEGs significantly enriched (FDR <0.05) in the EGFP^{low} subpopulation (Low) and in the EGFP^{high} subpopulation (High), respectively. Black dots indicate the genes with no significant changes between the two subpopulations.

(D) Representative results for OTX2 immunofluorescence staining of SOX1-EGFP reporter hESC (clone #6)-derived NPCs treated with 0.8 μM CT at day 8. Scale bars: 50 μm in the upper row; 20 μm in the lower row from a different view using an oil immersion lens.

(E) A diagram showing the targeting sites of 2 sgRNAs (*OTX2*i-g1 and *OTX2*i-g2) used for *OTX2* knockdown in SOX1-EGFP reporter hESCs (clone #6).

(F) qRT-PCR analysis of the *OTX2* knockdown efficiency as well as mRNA levels of *SOX1* and other indicated neural markers in NPCs (CT0.4, day 4) derived from *OTX2* knockdown (*OTX2*i-g1 or *OTX2*i-g2) and control knockdown (Ctrli)

Figure 2. *Continued*
SOX1-EGFP reporter hESCs, relative to Ctrli hESCs (day 0). n = 3 independent experiments. Data are shown as mean ± SEM. *p < 0.05, ***p < 0.001; n.s., not significant.
(G) Flow cytometric analysis for the percentage of $EGFP^{high}$ cells in NPCs (CT0.4, day 4) derived from *OTX2* knockdown (*OTX2*i-g1 or *OTX2*i-g2) and control knockdown (Ctrli) SOX1-EGFP reporter hESCs, respectively. NegCtrl: undifferentiated WT hESCs.
(H) Quantitative analysis of the percentage of $EGFP^{high}$ cells for results of the experiment shown in (G). n = 3 independent experiments, Data are shown as mean ± SEM. ***p < 0.001.
See also Figure S3.

differences between these two populations, we sorted NPCs (CT0.8, day 8) into $EGFP^{low}$ and $EGFP^{high}$ subpopulations with three biological replications for RNA sequencing (RNA-seq) (Figure 2B). Analysis of differentially expressed genes (DEGs) identified 170 genes enriched in $EGFP^{high}$ cells and 336 genes enriched in $EGFP^{low}$ cells (fold changes >1.5 and false discovery rate [FDR] < 0.05). Based on our previously published dataset obtained from positionally patterned NPCs derived from WT hESCs (Fang et al., 2019), clustering analysis of DEGs revealed that the $EGFP^{high}$ cells were grouped with NPCs treated with 0.8–2.0 μM CT, whereas $EGFP^{low}$ cells were grouped with NPCs treated with 0.4 μM CT (Figure S3A). Moreover, as indicated by the volcano plot, *OTX1*, *OTX2*, *LMX1A*, *LMX1B*, *EN1*, *WNT1*, and *WNT3A* were enriched in the $EGFP^{low}$ cells, whereas *GBX2*, *HOXA1*, *HOXA2*, and *SOX1* were enriched in the $EGFP^{high}$ cells (Figure 2C). These analyses indicate the association of SOX1 with rostral hindbrain NPCs.

As *OTX2* was highly expressed in $EGFP^{low}$ cells and our immunofluorescence staining analysis showed the mutually exclusive expression of OTX2 and SOX1 (Figure 2D), we suspected that OTX2 may repress the expression of *SOX1* in the $EGFP^{low}$ NPCs. By analyzing published chromatin immunoprecipitation sequencing (ChIP-seq) data of OTX2 in hESCs, endoderm cells, and ectoderm cells (Tsankov et al., 2015), we found an OTX2-binding site located downstream of the *SOX1* locus (Figure S3B), suggesting that OTX2 may bind to the regulatory region of *SOX1* to inhibit its expression. In addition, we found an OTX2-binding site downstream of the *GBX2* locus (Figure S3B). Our ChIP-qPCR analysis validated binding of OTX2 to these two sites (Figure S3C) in NPCs of the CT0.4 group. To test whether OTX2 could regulate *SOX1* expression, we knocked down *OTX2* in SOX1-EGFP reporter hESCs using the CRISPRi system (Mandegar et al., 2016) with two different sgRNAs (OTX2i-g1 and OTX2i-g2) targeting the 5′ UTR sequence of *OTX2* gene and differentiated them into NPCs with 0.4 μM CT (Figure 2E). Results of qRT-PCR analysis indicated that *SOX1* mRNA levels in NPCs were substantially increased by *OTX2* knockdown. Expression of *GBX2* and *HOXA2* was also upregulated, whereas *SOX2* levels were not significantly altered by *OTX2* knockdown (Figure 2F). Furthermore, flow cytometric analysis showed that *OTX2* knockdown resulted in a significant increase in the proportion of $EGFP^{high}$ NPCs (Figures 2G and 2H). Thus, we uncover a new role of OTX2 for inhibiting *SOX1* expression and provide experimental evidence for the repressive effect of OTX2 on *GBX2* expression in NPCs derived from hESCs.

SOX1 Is Required for Specification of Rostral Hindbrain NPCs from hESCs

To understand the role of SOX1 in specifying rostral hindbrain NPCs from hESCs, we generated *SOX1* knockout (*SOX1*-KO) H9 hESC lines by the CRISPR/Cas9 nickase system. The PGK-PuroR cassette was used to replace the whole coding sequence (CDS) of *SOX1* gene (Figure 3A). Our western blot result identified five hESC colonies lacking SOX1 expression (Figure 3B). Two colonies, #26 (S1KO-#26) and #48 (S1KO-#48), were selected for Sanger sequencing, which verified the precise insertion of the PGK-PuroR cassette (Figure S4A). hESCs with *SOX1* homozygous deletion from these two clones were differentiated into NPCs with different regional identities along the R-C axis, including the rostral forebrain (IWP2), midbrain (CT0.4), and rostral hindbrain (CT1.0). RNA samples of undifferentiated hESCs (day 0) and NPCs from these three groups at days 4 and 8 were collected for RNA-seq (Figure 3C). Compared with our previously published RNA-seq data of WT samples (Fang et al., 2019), DEGs were identified (Figure 3D). Among the three groups of NPCs, *SOX1*-KO gave rise to the most DEGs in the CT1.0 group and the least DEGs in the IWP2 group (Figure 3D), suggesting that *SOX1*-KO affected NPCs of the CT1.0 group most. Therefore, Gene Ontology analyses were performed for DEGs between *SOX1*-KO (CT1.0) and WT NPCs (CT1.0) on day 8 (Figure 3E). Notably, upregulated DEGs in *SOX1*-KO NPCs enriched terms of pattern specification process and midbrain development. Moreover, both up- and downregulated DEGs enriched the term of WNT signaling. In addition, based on our previously published dataset (Fang et al., 2019), clustering analysis of DEGs revealed that the *SOX1*-KO NPCs (CT1.0) were grouped with WT NPCs of CT0.8 and CT0.4 groups (Figure 3F), suggesting that the regional identity of *SOX1*- KO NPCs of the CT1.0 group was closer to that of WT NPCs of CT0.8 and CT0.4 groups than to that of WT NPCs of the CT1.0 group. In contrast,

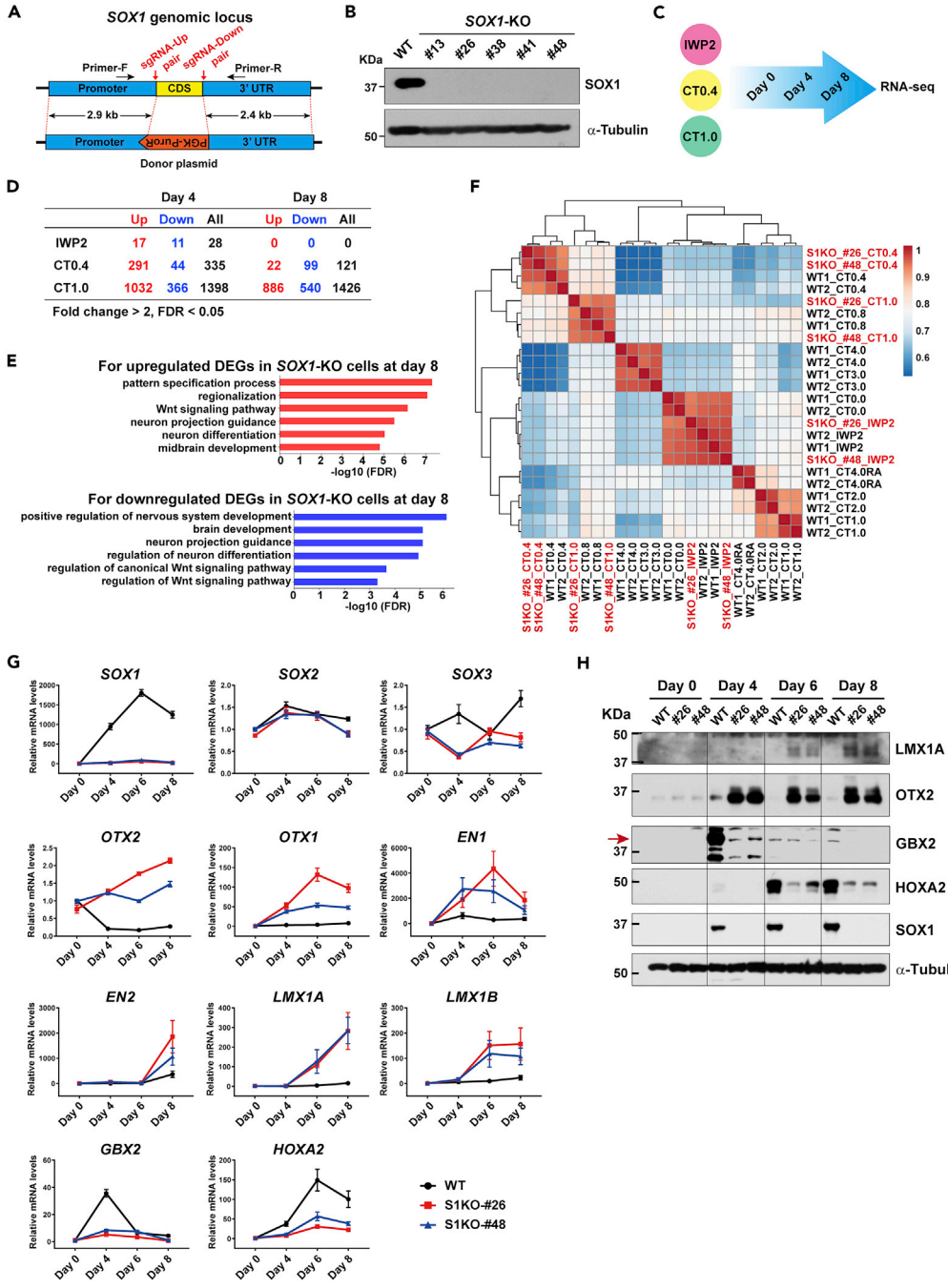


Figure 3. SOX1 Knockout Impairs the Specification of Rostral Hindbrain NPCs from hESCs

(A) The design for the construction of *SOX1*-KO hESC lines with two pairs of sgRNAs. The cleavage sites are indicated by red arrows. The sgRNA-Up pair: a pair of sgRNAs targeting the *SOX1* genomic sequence around the 5′-end of its CDS region; the sgRNA-Down pair: a pair of sgRNAs targeting the *SOX1* genomic sequence around the 3′-end of its CDS region.

(B) Representative results of western blot analysis for the identification of five hESC clones lacking SOX1 expression.

(C) A diagram of sample collection for RNA-seq analysis.

(D) The summary of the number of upregulated, downregulated, and total DEGs in *SOX1*-KO NPCs of each group. Fold change > 2.0, FDR < 0.05.

(E) Gene Ontology (GO) analyses for DEGs between WT and *SOX1*-KO NPCs (CT1.0). Only the top six GO terms with FDR < 0.05 and combined score > 15 are shown.

Figure 3. *Continued*
(F) Unsupervised hierarchical clustering and Pearson correlation between WT and *SOX1*-KO NPCs at day 8.
(G) qRT-PCR analysis of mRNA levels of neural markers in WT and *SOX1*-KO NPCs (CT1.0) at the indicated time points, relative to undifferentiated WT hESCs (day 0). n = 3 independent experiments. Data are shown as mean ± SEM.
(H) Representative results for western blot analysis of neural markers in WT and *SOX1*-KO cells at days 0, 4, 6, and 8 of neural differentiation (CT1.0). A red arrow indicates the specific band of GBX2 proteins.
See also Figure S4.

the *SOX1*-KO NPCs treated with IWP2 or 0.4 μM CT were grouped with WT NPCs under the same treatment. These results indicate that *SOX1*-KO impairs the specification of rostral hindbrain NPCs.

In addition, the aberrant expression of neural markers (Figure 3G) and WNT signaling-related genes (Figure S4B) were verified by qRT-PCR analysis. *OTX1*, *OTX2*, *EN1*, *EN2*, *LMX1A*, and *LMX1B* were upregulated, whereas *HOXA2* was downregulated, in NPCs of both *SOX1*-KO clones, when compared with WT NPCs. Of note, *SOX1*-KO caused most evident downregulation of *GBX2* at day 4 of differentiation in NPCs (CT1.0). However, *SOX1*-KO did not alter the expression of *SOX2* (Figure 3G). Western blot analyses also showed the elevation of OTX2 and LMX1A protein levels, and the reduction in GBX2 and HOXA2 protein levels in *SOX1*-KO NPCs (Figure 3H). Consistent with our qRT-PCR result (Figure S4B), luciferase assay results showed that *SOX1*-KO significantly enhanced the β-catenin/TCF reporter (TOP/FOPFlash) activity (Figure S4C), suggesting a repressive role of SOX1 for WNT signaling in rostral hindbrain NPCs.

As Sox1 has been reported to mark proliferating cells within the neural tube of mouse embryos (Pevny et al., 1998), we determined whether *SOX1*-KO would affect the cell cycle and proliferation of NPCs treated with 1.0 μM CT. Flow cytometric analyses indicated that there were not obvious differences in the distribution of cell cycle phases (Figures S4D and S4E) and the percentage of EdU^+ cells between WT and *SOX1*-KO NPCs (Figures S4F and S4G).

SOX1 Activates *GBX2* Expression Contributing to Rostral Hindbrain NPC Specification

To understand how SOX1 contributed to the specification of rostral hindbrain NPCs from hESCs, we carried out ChIP-seq of SOX1 in rostral forebrain NPCs at day 8 (IWP2 D8), rostral hindbrain NPCs at day 4 (CT1.0 D4), and rostral hindbrain NPCs at day 8 (CT1.0 D8) (Figure 4A). A total of 4,223 highly reliable SOX1-binding regions were identified using the MACS2 (FDR q-value < 0.01), with the most binding regions in the NPCs of CT1.0 D8 group (Figure S5A). Most SOX1-binding regions located in distal intergenic regions (Figure S5B). The DNA motif analysis revealed that 73.48% of SOX1-binding regions contained a [CT][AT]TTGT-enriched sequence (p value < 1×10^{-1234}), consistent with the common consensus motif of SOX proteins (Figure 4B). To identify downstream target genes of SOX1, we overlaid the DEGs (between *SOX1*-KO and WT NPCs) and SOX1-binding genes obtained from our SOX1 ChIP-seq data of NPCs (CT 1.0) on day 4. Ninety-two putative downstream targets of SOX1 in rostral hindbrain NPCs were identified (Figure 4C). Among these 92 genes, we focused on *GBX2* and *HOXA2* due to their close relationships with hindbrain development. SOX1-binding regions were found about 80 kb downstream of the *GBX2* locus (Figure 4D) and in the second exon of the *HOXA2* locus (Figure S5C).

To test the role of GBX2 and HOXA2 in the SOX1-mediated function, we established doxycycline (Dox)-inducible hESC lines overexpressing *EGFP*, *GBX2*, and *HOXA2*, respectively, in WT and *SOX1*-KO hESCs (clones #26 and #48). During neural differentiation under the CT 1.0 μM condition, Dox was added from day 0 to day 8, and NPCs were collected at days 4 and 8 for analyses. Our result of qRT-PCR analysis showed that *GBX2* overexpression in *SOX1*-KO NPCs could largely abolish *SOX1*-KO-induced upregulation of midbrain markers and downregulation of *HOXA2* (Figure 4E). Of note, overexpression of *GBX2* dramatically reduced *OTX2* expression levels in both WT and *SOX1*-KO NPCs, in agreement with the previous reports that Gbx2 represses *Otx2* expression in embryos (Katahira et al., 2000; Millet et al., 1999). Furthermore, overexpression of *GBX2* evidently blocked *SOX1*-KO-caused upregulation of *WNT1*, *AXIN2*, and *WLS* as well as downregulation of *SFRP2* and *SFRP4*, to various degrees (Figure 4E). Western blot analysis further revealed that overexpression of *GBX2* abrogated alterations in protein levels of OTX2 and HOXA2 induced by *SOX1*-KO (Figure 4F). In contrast, *HOXA2* overexpression in *SOX1*-KO NPCs could not efficiently abolish *SOX1*-KO-caused alterations in the expression of the most tested genes (Figure S5D). Collectively, SOX1 may contribute to the specification of rostral hindbrain NPCs from hESCs primarily by controlling *GBX2* expression.

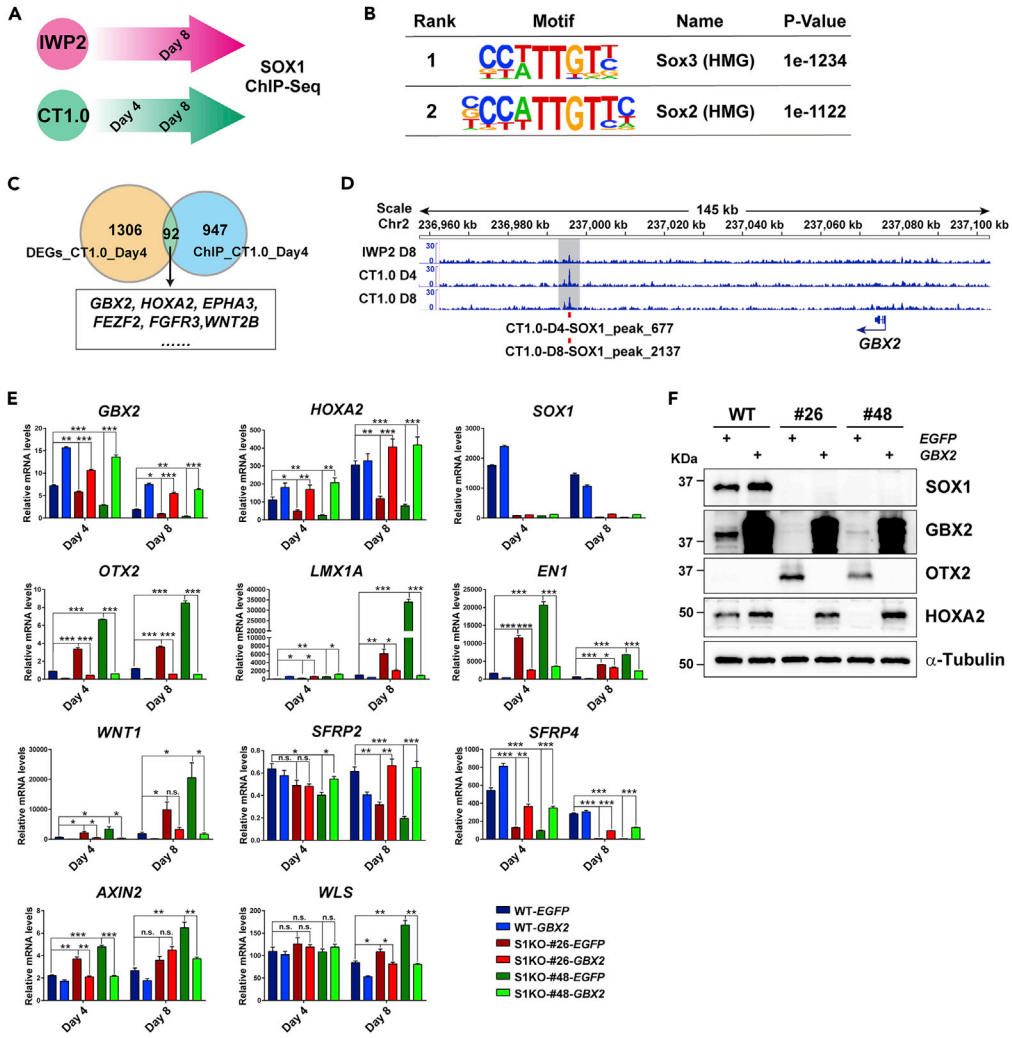


Figure 4. SOX1 Regulates *GBX2* Expression Contributing to Rostral Hindbrain NPC Specification

(A) A diagram of sample collection for SOX1 ChIP-seq analysis.

(B) The DNA motif analysis of SOX1-binding regions.

(C) The Venn diagram showing the overlap between DEGs (WT versus *SOX1*-KO NPCs treated with 1.0 μM CT at day 4) and SOX1-binding genes identified by SOX1 ChIP-seq analysis (CT1.0, Day 4).

(D) The SOX1-binding region downstream of the *GBX2* locus in NPCs of the three groups indicated in (A).

(E) qRT-PCR analysis of mRNA levels of neural markers and WNT signaling-related genes at the indicated time points in WT and *SOX1*-KO NPCs (CT1.0) overexpressing *EGFP* or *GBX2*, relative to undifferentiated WT hESCs (day 0). n = 3 independent experiments. Data are shown as mean ± SEM. *p < 0.05, **p < 0.01, ***p < 0.001; n.s., not significant. Relative expression levels of *SOX1* were not analyzed statistically.

(F) Representative results for western blot analysis of neural marker protein levels at day 4 in WT and *SOX1*-KO NPCs (CT1.0) overexpressing *EGFP* or *GBX2*.

See also Figure S5.

Identification of the Regulatory Region for *GBX2* Expression

As mentioned earlier, there was a SOX1-binding region downstream of the *GBX2* locus in rostral hindbrain NPCs (Figure 4D), and analysis of published datasets (Rada-Iglesias et al., 2011) revealed that p300 and mono-methylation of histone H3 at lysine 4 (H3K4me1) were highly enriched around this region in hESCs (Figure 5A). However, whether this region would be functionally important for *GBX2* expression has remained unclear. To address this question, we deleted the core SOX1-binding region (named S1B region) using the CRISPR/Cas9 system with a pair of sgRNAs (S1B-sgRNA1 and S1B-sgRNA2) (Figures 5A and 5B). Two clones (S1BKO-#5 and S1BKO-#27)

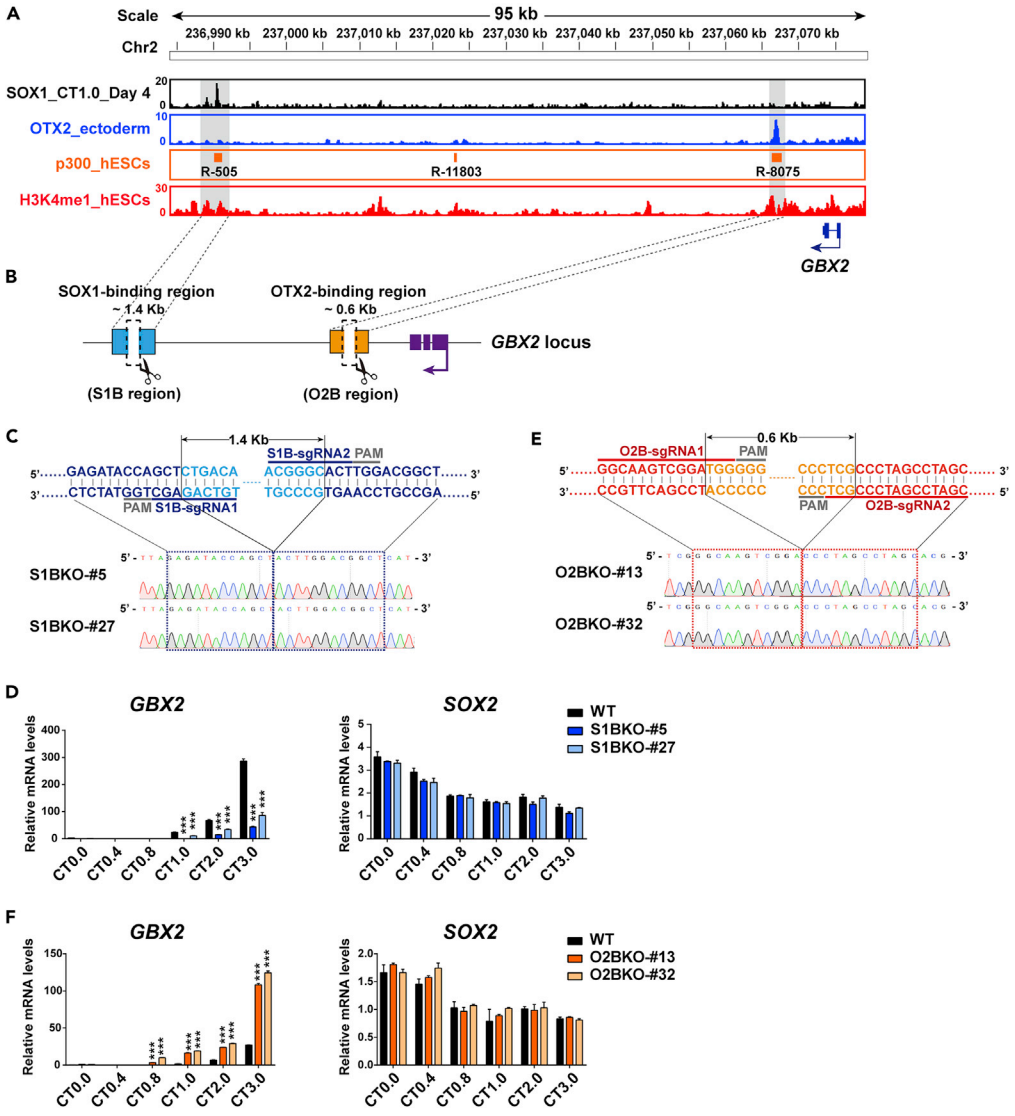


Figure 5. Identification of Two Regulatory Regions for *GBX2* Expression

(A) Genomic binding sites of SOX1 downstream of the *GBX2* locus in NPCs of the CT1.0 group at day 4. ChIP-seq signal profiles of OTX2 in ectoderm cells (OTX2_ectoderm) and p300 and H3K4me1 in hESCs (p300_hESCs and H3K4me1_hESCs) were obtained from published datasets.

(B) The strategy to delete the core SOX1-binding region (S1B region) and OTX2-binding region (O2B region) in hESCs.

(C) Representative Sanger sequencing peak maps verified a deletion of 1.4 kb in hESCs of S1BKO-#5 and S1BKO-#27 clones.

(D) qRT-PCR analysis of mRNA levels of *GBX2* and *SOX2* in WT and S1BKO NPCs under the indicated CT concentrations at day 4, relative to undifferentiated WT hESCs (day 0). n = 3 independent experiments. Data are shown as mean ± SEM. ***$p < 0.001$.

(E) Representative Sanger sequencing peak maps verified a deletion of 0.6 kb in hESCs of O2BKO-#13 and O2BKO-#32 clones.

(F) qRT-PCR analysis of mRNA levels of *GBX2* and *SOX2* in WT and O2BKO NPCs under the indicated CT concentrations at day 4, relative to undifferentiated WT hESCs (day 0). n = 3 independent experiments. Data are shown as mean ± SEM. ***$p < 0.001$.

possessing biallelic deletion of the S1B region were identified and confirmed by Sanger sequencing (Figure 5C) and genomic DNA PCR analysis (data not shown). We then differentiated hESCs of WT and S1B KO clones into NPCs with different CT concentrations ranging from 0.0 to 3.0 μM. We found that deletion of the S1B region

downregulated *GBX2* dramatically in the NPCs of CT1.0, CT2.0, and CT3.0 groups, validating an important activating role of the S1B region for *GBX2* expression (Figure 5D).

It has been shown that Otx2 could repress *Gbx2* expression in animal models (Acampora et al., 1995, 1997; Ang et al., 1996; Matsuo et al., 1995; Simeone, 1998). However, the underlying mechanisms have not been fully elucidated, particularly in hESC-derived NPCs. Based on published OTX2 ChIP-seq data (Tsankov et al., 2015), we found an OTX2-binding region about 20 kb downstream of the *GBX2* locus (Figure S3B). We deleted the core OTX2-binding region (named O2B region) using the CRISPR/Cas9 system with a pair of sgRNAs (O2B-sgRNA1 and O2B-sgRNA2) to determine the role of the O2B region for *GBX2* expression (Figures 5A and 5B). Two clones (O2BKO-#13 and O2BKO-#32) with a homozygous deletion of the O2B region were established and verified by Sanger sequencing (Figure 5E) and genomic DNA PCR analysis (data not shown). Deletion of the O2B region led to obvious upregulation of *GBX2* in the NPCs of CT0.8, CT1.0, CT2.0, and CT3.0 groups (Figure 5F), indicating a repressive role of this O2B region for *GBX2* expression in NPCs. In contrast, deletion of either S1B region or O2B region had no clear effects on the expression of *SOX2* in NPCs, favoring the notion that these two regulatory regions are specifically responsible for the expression of *GBX2*. The question of how these regions participate in the control of *GBX2* expression remains unclear.

DISCUSSION

In this study, we report the expression pattern and function of transcriptional factor SOX1 in hESC-derived NPCs. Our results show that *SOX1* was highly expressed in NPCs with the rostral hindbrain identity at the early stage of hESC neural differentiation. Using the SOX1-EGFP reporter hESC line established in the current study, we found that SOX1-EGFP^{low} and SOX1-EGFP^{high} NPCs were present in the same culture dish when the CT concentration was at 0.8 μM, and we were able to selectively purify these two subpopulations for transcriptomic analysis. As expected, EGFP^{high} cells highly expressed rostral hindbrain markers, whereas EGFP^{low} cells highly expressed midbrain markers, arguing for the association of SOX1 with rostral hindbrain NPCs. Furthermore, our immunofluorescence staining results showed that SOX1-EGFP signal and OTX2 expression were mutually exclusive when they were detected in the same culture dish. These observations point to the close relationship of SOX1 expression with rostral hindbrain NPCs. The following lines of evidence indicate a role of SOX1 in the specification of rostral hindbrain NPCs: (1) analysis of DEGs between WT and *SOX1*-KO NPCs showed that *SOX1*-KO affected gene expression in NPCs of the CT1.0 group most, (2) *SOX1*-KO downregulated hindbrain markers and upregulated midbrain markers, (3) SOX1-binding regions were mostly detected in NPCs of the CT1.0 group, and (4) SOX1 regulated expression of *GBX2* and *HOXA2*. Therefore, this study unraveled a new role of SOX1 for the specification of rostral hindbrain NPCs from hESCs. These findings will help to elucidate how human neural regionalization is regulated at a molecular level.

To dissect the molecular mechanism underlying the function of SOX1, we searched for its downstream targets and identified GBX2 as a key factor responsible for the function of SOX1 in the specification of rostral hindbrain NPCs from hESCs, based on the following evidence: (1) our SOX1 ChIP-seq analysis identified a SOX1-binding region downstream of the *GBX2* locus, (2) *SOX1*-KO in NPCs of the CT1.0 group resulted in significant downregulation of *GBX2*, and (3) overexpression of *GBX2* largely abrogates *SOX1*-KO-induced alterations in gene expression. It is significant to uncover a new regulator of *GBX2* expression, as Gbx2 is an important factor for the normal development of the rostral hindbrain and the correct positioning of the MHB (Nakayama et al., 2013; Wassarman et al., 1997; Waters and Lewandoski, 2006).

One of the important contributions of this study is to propose a model for the interaction between OTX2, GBX2, and SOX1 in hESC-derived NPCs (Figure 6). It is known that a set of genes act in distinct domains around the Otx/Gbx boundary to specify cells of different fates. It remains an open question how these genes interact to pattern the cell fates around the boundary. We identified SOX1 as a downstream target of OTX2, a previously unreported regulatory relationship. Thus, OTX2 functions by repressing *SOX1* expression in addition to its suppression of *GBX2*. Our discovery of mutual exclusion between SOX1-expressing NPCs and OTX2-expressing NPCs led to the identification of SOX1-EGFP^{high} rostral hindbrain NPCs and an important role of SOX1 in neural regionalization. Moreover, this study not only uncovered an activating role of SOX1 for *GBX2* expression but also identified a SOX1-binding site-containing core regulatory region required for *GBX2* expression. Furthermore, we provided experimental evidence for the role of an OTX2-binding site-containing core regulatory region in repressing *GBX2* expression. These findings would greatly enhance our understanding of how the OTX2

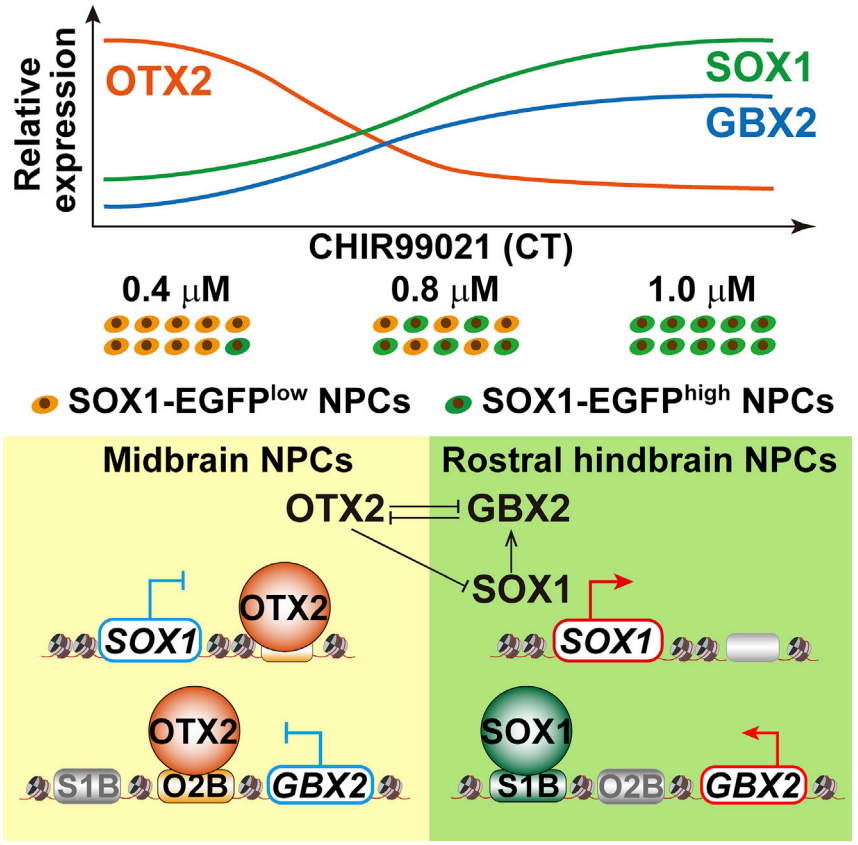


Figure 6. A Working Model

A proposed model illustrating the interaction between OTX2, SOX1, and GBX2 in hESC-derived NPCs. SOX1 and GBX2 are highly expressed in rostral hindbrain NPCs (CT1.0), whereas OTX2 is highly expressed in midbrain NPCs (CT0.4). OTX2 can bind to the regulatory region downstream of *SOX1* and *GBX2* to inhibit the expression of these two factors. SOX1 can bind to the distal region downstream of *GBX2* to activate its expression. S1B, the core SOX1-binding region; O2B, the core OTX2-binding region; NPC, neural progenitor cells.

and GBX2 boundary is regulated at a transcriptional level. However, our finding also raises more questions, such as how the SOX1-binding site-containing core regulatory region contributes to the spatiotemporal regulation of *GBX2* expression during neural regionalization? What is the role of the local chromatin state for SOX1-mediated regulation of *GBX2* expression? On the other hand, SOX proteins usually collaborate with their partner factors to control gene expression (Kondoh and Kamachi, 2010). Thus, it will be helpful to find factors associated with SOX1 in rostral hindbrain NPCs. In addition, comparison of the genome-wide chromatin accessibility and configuration between WT and *SOX1*-KO NPCs would be useful for comprehensively elucidating the molecular mechanisms by which SOX1 controls cell identity during early neural regionalization. The SOX1-EGFP reporter hESC lines generated in this study, allowing reliable visualization and purification of NPCs with the rostral hindbrain identity during hESC neural regionalization, will assist addressing the aforementioned issues.

Limitations of the Study

Our findings were obtained from an *in vitro* differentiation system, which may recapitulate the early stage of human neural regionalization. It remains to be tested whether our findings are consistent with the early human neural development process *in vivo*.

Resource Availability

Lead Contact

Further information and requests for resources should be directed to and fulfilled by the Lead Contact, Ying Jin (yjin@sibs.ac.cn).

Materials Availability

All plasmids generated in this study and their information are available from the corresponding author upon reasonable request.

Data and Code Availability

The accession number for the RNA-seq and ChIP-seq data reported in this paper is GEO: GSE138218.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.101475.

ACKNOWLEDGMENTS

This work was supported by grants from the Ministry of Science and Technology of China (2016YFA0100100), the National Natural Science Foundation of China (31730055, 31801224 and 31871373), Strategic Priority Research Program of the Chinese Academy of Sciences (XDB19020100), and the Innovative Research Team of High-level Local Universities in Shanghai (SSMU-ZLCX20180401).

AUTHOR CONTRIBUTIONS

X.L., Z.F., Y.J., and D.L. designed the project. X.L. and Z.F. performed major experiments, analyzed data, and wrote the manuscript. J.W and F.T helped with immunostaining. B.L. and N.J. discussed the project and edited the manuscript. Y.J. directed the project and wrote the manuscript. X.L. and Z.F. contributed equally to this study.

DECLARATION OF INTERESTS

The authors declare that they have no conflicts of interest with the contents of this article.

Received: December 10, 2019
Revised: July 11, 2020
Accepted: August 17, 2020
Published: September 25, 2020

REFERENCES

Acampora, D., Mazan, S., Lallemand, Y., Avantaggiato, V., Maury, M., Simeone, A., and Brulet, P. (1995). Forebrain and midbrain regions are deleted in Otx2-/- mutants due to a defective anterior neuroectoderm specification during gastrulation. Development *121*, 3279–3290.

Acampora, D., Avantaggiato, V., Tuorto, F., and Simeone, A. (1997). Genetic control of brain morphogenesis through Otx gene dosage requirement. Development *124*, 3639–3650.

Ang, S.L., Jin, O., Rhinn, M., Daigle, N., Stevenson, L., and Rossant, J. (1996). A targeted mouse Otx2 mutation leads to severe defects in gastrulation and formation of axial mesoderm and to deletion of rostral brain. Development *122*, 243–252.

Aubert, J., Stavridis, M.P., Tweedie, S., O'Reilly, M., Vierlinger, K., Li, M., Ghazal, P., Pratt, T., Mason, J.O., Roy, D., et al. (2003). Screening for mammalian neural genes via fluorescence-activated cell sorter purification of neural precursors from Sox1-gfp knock-in mice. Proc. Natl. Acad. Sci. U S A *100*, 11836–11841.

Bally-Cuif, L., Cholley, B., and Wassef, M. (1995a). Involvement of Wnt-1 in the formation of the mes/metencephalic boundary. Mech. Dev. *53*, 23–34.

Bally-Cuif, L., Gulisano, M., Broccoli, V., and Boncinelli, E. (1995b). c-otx2 is expressed in two different phases of gastrulation and is sensitive to retinoic acid treatment in chick embryo. Mech. Dev. *49*, 49–63.

Bouillet, P., Chazaud, C., Oulad-Abdelghani, M., Dolle, P., and Chambon, P. (1995). Sequence and expression pattern of the Stra7 (Gbx-2) homeobox-containing gene induced by retinoic acid in P19 embryonal carcinoma cells. Dev. Dyn. *204*, 372–382.

Broccoli, V., Boncinelli, E., and Wurst, W. (1999). The caudal limit of Otx2 expression positions the isthmic organizer. Nature *401*, 164–168.

Chambers, S.M., Fasano, C.A., Papapetrou, E.P., Tomishima, M., Sadelain, M., and Studer, L. (2009). Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. Nat. Biotechnol. *27*, 275–280.

Fang, Z., Liu, X., Wen, J., Tang, F., Zhou, Y., Jing, N., and Jin, Y. (2019). SOX21 ensures rostral forebrain identity by suppression of WNT8B during neural regionalization of human embryonic stem cells. Stem Cell Reports *13*, 1038–1052.

Fasano, C.A., Chambers, S.M., Lee, G., Tomishima, M.J., and Studer, L. (2010). Efficient derivation of functional floor plate tissue from human embryonic stem cells. Cell Stem Cell *6*, 336–347.

Hanks, M., Wurst, W., Anson-Cartwright, L., Auerbach, A.B., and Joyner, A.L. (1995). Rescue of the en-1 mutant phenotype by replacement of en-1 with en-2. Science *269*, 679–682.

Hunt, P., Gulisano, M., Cook, M., Sham, M.H., Faiella, A., Wilkinson, D., Boncinelli, E., and Krumlauf, R. (1991). A distinct Hox code for the branchial region of the vertebrate head. Nature *353*, 861–864.

Imaizumi, K., Sone, T., Ibata, K., Fujimori, K., Yuzaki, M., Akamatsu, W., and Okano, H. (2015). Controlling the regional identity of hPSC-derived

neurons to uncover neuronal subtype specificity of neurological disease phenotypes. Stem Cell Reports *5*, 1010–1022.

Joyner, A.L., Liu, A., and Millet, S. (2000). Otx2, Gbx2 and Fgf8 interact to position and maintain a mid-hindbrain organizer. Curr. Opin. Cell Biol. *12*, 736–741.

Katahira, T., Sato, T., Sugiyama, S., Okafuji, T., Araki, I., Funahashi, J., and Nakamura, H. (2000). Interaction between Otx2 and Gbx2 defines the organizing center for the optic tectum. Mech. Dev. *91*, 43–52.

Kiecker, C., and Niehrs, C. (2001a). A morphogen gradient of Wnt/beta-catenin signalling regulates anteroposterior neural patterning in Xenopus. Development *128*, 4189–4201.

Kiecker, C., and Niehrs, C. (2001b). The role of prechordal mesendoderm in neural patterning. Curr. Opin. Neurobiol. *11*, 27–33.

Kirkeby, A., Grealish, S., Wolf, D.A., Nelander, J., Wood, J., Lundblad, M., Lindvall, O., and Parmar, M. (2012). Generation of regionally specified neural progenitors and functional neurons from human embryonic stem cells under defined conditions. Cell Rep. *1*, 703–714.

Kondoh, H., and Kamachi, Y. (2010). SOX-partner code for cell specification: regulatory target selection and underlying molecular mechanisms. Int. J. Biochem. Cell Biol. *42*, 391–399.

Kriks, S., Shim, J.W., Piao, J., Ganat, Y.M., Wakeman, D.R., Xie, Z., Carrillo-Reid, L., Auyeung, G., Antonacci, C., Buch, A., et al. (2011). Dopamine neurons derived from human ES cells efficiently engraft in animal models of Parkinson's disease. Nature *480*, 547–551.

Li, C., Yang, Y., Lu, X., Sun, Y., Gu, J., Feng, Y., and Jin, Y. (2010). Efficient derivation of Chinese human embryonic stem cell lines from frozen embryos. In Vitro Cell. Dev. Biol. Anim. *46*, 186–191.

Luu, B., Ellisor, D., and Zervas, M. (2011). The lineage contribution and role of Gbx2 in spinal cord development. PLoS One *6*, e20940.

Maden, M. (2007). Retinoic acid in the development, regeneration and maintenance of the nervous system. Nat. Rev. Neurosci. *8*, 755–765.

Malas, S., Duthie, S.M., Mohri, F., Lovell-Badge, R., and Episkopou, V. (1997). Cloning and mapping of the human SOX1: a highly conserved gene expressed in the developing brain. Mamm. Genome *8*, 866–868.

Malas, S., Postlethwaite, M., Ekonomou, A., Whalley, B., Nishiguchi, S., Wood, H., Meldrum, B., Constanti, A., and Episkopou, V. (2003). Sox1-deficient mice suffer from epilepsy associated with abnormal ventral forebrain development and olfactory cortex hyperexcitability. Neuroscience *119*, 421–432.

Mandegar, M.A., Huebsch, N., Frolov, E.B., Shin, E., Truong, A., Olvera, M.P., Chan, A.H., Miyaoka, Y., Holmes, K., Spencer, C.I., et al. (2016). CRISPR interference efficiently induces specific and reversible gene silencing in human iPSCs. Cell Stem Cell *18*, 541–543.

Maroof, A.M., Keros, S., Tyson, J.A., Ying, S.W., Ganat, Y.M., Merkle, F.T., Liu, B., Goulburn, A., Stanley, E.G., Elefanty, A.G., et al. (2013). Directed differentiation and functional maturation of cortical interneurons from human embryonic stem cells. Cell Stem Cell *12*, 559–572.

Matsuo, I., Kuratani, S., Kimura, C., Takeda, N., and Aizawa, S. (1995). Mouse Otx2 functions in the formation and patterning of rostral head. Genes Dev. *9*, 2646–2658.

Millet, S., Bloch-Gallego, E., Simeone, A., and Alvarado-Mallart, R.M. (1996). The caudal limit of Otx2 gene expression as a marker of the midbrain/hindbrain boundary: a study using in situ hybridisation and chick/quail homotopic grafts. Development *122*, 3785–3797.

Millet, S., Campbell, K., Epstein, D.J., Losos, K., Harris, E., and Joyner, A.L. (1999). A role for Gbx2 in repression of Otx2 and positioning the mid/hindbrain organizer. Nature *401*, 161–164.

Nakayama, Y., Kikuta, H., Kanai, M., Yoshikawa, K., Kawamura, A., Kobayashi, K., Wang, Z., Khan, A., Kawakami, K., and Yamasu, K. (2013). Gbx2 functions as a transcriptional repressor to regulate the specification and morphogenesis of the mid-hindbrain junction in a dosage- and stage-dependent manner. Mech. Dev. *130*, 532–552.

Nishiguchi, S., Wood, H., Kondoh, H., Lovell-Badge, R., and Episkopou, V. (1998). Sox1 directly regulates the gamma-crystallin genes and is essential for lens development in mice. Genes Dev. *12*, 776–781.

Pevny, L.H., Sockanathan, S., Placzek, M., and Lovell-Badge, R. (1998). A role for SOX1 in neural determination. Development *125*, 1967–1978.

Prince, V., and Lumsden, A. (1994). Hoxa-2 expression in normal and transposed rhombomeres: independent regulation in the neural tube and neural crest. Development *120*, 911–923.

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

Schneider-Maunoury, S., Topilko, P., Seitandou, T., Levi, G., Cohen-Tannoudji, M., Pournin, S., Babinet, C., and Charnay, P. (1993). Disruption of Krox-20 results in alteration of rhombomeres 3 and 5 in the developing hindbrain. Cell *75*, 1199–1214.

Shamim, H., and Mason, I. (1998). Expression of Gbx-2 during early development of the chick embryo. Mech. Dev. *76*, 157–159.

Simeone, A. (1998). Otx1 and Otx2 in the development and evolution of the mammalian brain. EMBO J. *17*, 6790–6798.

Simeone, A., Acampora, D., Gulisano, M., Stornaiuolo, A., and Boncinelli, E. (1992). Nested expression domains of four homeobox genes in developing rostral brain. Nature *358*, 687–690.

Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998). Embryonic stem cell lines derived from human blastocysts. Science *282*, 1145–1147.

Tsankov, A.M., Gu, H., Akopian, V., Ziller, M.J., Donaghey, J., Amit, I., Gnirke, A., and Meissner, A. (2015). Transcription factor binding dynamics during human ES cell differentiation. Nature *518*, 344–349.

Uchikawa, M., Yoshida, M., Iwafuchi-Doi, M., Matsuda, K., Ishida, Y., Takemoto, T., and Kondoh, H. (2011). B1 and B2 Sox gene expression during neural plate development in chicken and mouse embryos: universal versus species-dependent features. Dev. Growth Differ. *53*, 761–771.

Wassarman, K.M., Lewandoski, M., Campbell, K., Joyner, A.L., Rubenstein, J.L., Martinez, S., and Martin, G.R. (1997). Specification of the anterior hindbrain and establishment of a normal mid/hindbrain organizer is dependent on Gbx2 gene function. Development *124*, 2923–2934.

Waters, S.T., and Lewandoski, M. (2006). A threshold requirement for Gbx2 levels in hindbrain development. Development *133*, 1991–2000.

Wilkinson, D.G., Bhatt, S., Chavrier, P., Bravo, R., and Charnay, P. (1989). Segment-specific expression of a zinc-finger gene in the developing nervous system of the mouse. Nature *337*, 461–464.

Wood, H.B., and Episkopou, V. (1999). Comparative expression of the mouse Sox1, Sox2 and Sox3 genes from pre-gastrulation to early somite stages. Mech. Dev. *86*, 197–201.

Xuan, S., Baptista, C.A., Balas, G., Tao, W., Soares, V.C., and Lai, E. (1995). Winged helix transcription factor BF-1 is essential for the development of the cerebral hemispheres. Neuron *14*, 1141–1152.

iScience, Volume *23*

Supplemental Information

SOX1 Is Required for the Specification of Rostral Hindbrain Neural Progenitor Cells from Human Embryonic Stem Cells

Xinyuan Liu, Zhuoqing Fang, Jing Wen, Fan Tang, Bing Liao, Naihe Jing, Dongmei Lai, and Ying Jin

Supplemental figures and legends

Figure S1

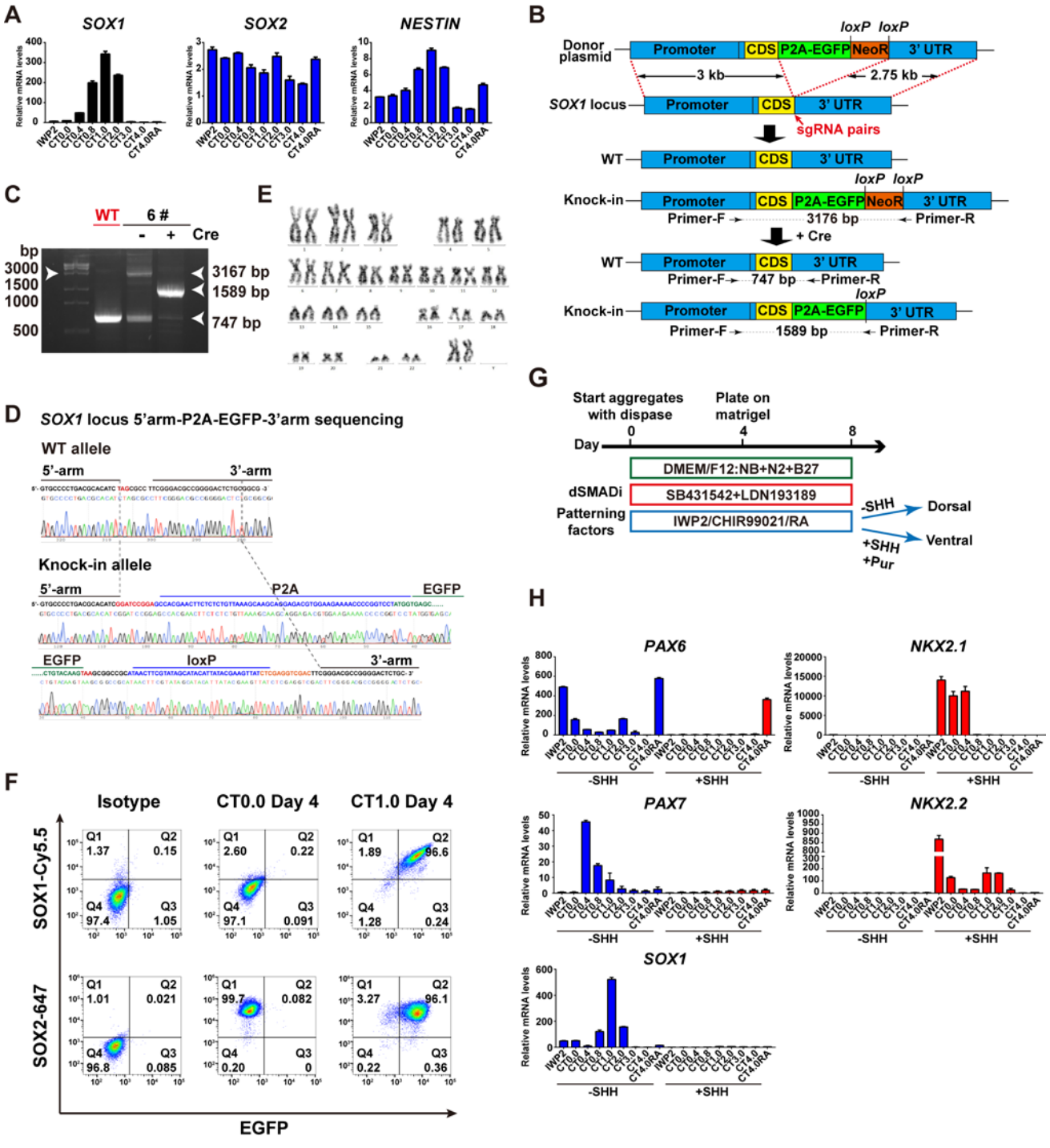


Fig S1. Related to Fig 1.

(**A**) qRT-PCR analysis of mRNA levels of *SOX1*, *SOX2* and *NESTIN* in NPCs derived from another hESC line (SHhES2) at day 4, relative to undifferentiated WT hESCs (day 0). n=3 independent experiments. Data are shown as mean ± SEM.

(**B**) A diagram showing the design of the targeting vector and the screening strategy to identify correctly targeted SOX1-EGFP reporter hESCs. The cleavage sites are indicated by a red arrow.

(**C**) The representative genomic DNA PCR result showing the correct targeting and the excision of the PGK-NeoR cassette after Cre-mediated recombination in hESCs of the clone #6.

(**D**) Representative results for Sanger sequencing peak maps verified precise joining of 5' homologous arm with the P2A-EGFP cassette, and a *loxP* site remained with 3' homologous arm after Cre-mediated recombination in SOX1-EGFP reporter hESCs of clone #6.

(**E**) The representative result from karyotype analysis of hESCs of the clone #6: normal 46, XX.

(**F**) Flow cytometric profiles of SOX1 and EGFP as well as SOX2 and EGFP in NPCs derived from SOX1-EGFP reporter hESCs of clone #6 under indicated conditions. Similar results were obtained in 3 independent experiments.

(**G**) The illustration of our differentiation protocol to generate NPCs with both D-V and R-C identities from hESCs. SHH: Sonic Hedgehog C24II; Pur: purmorphamine.

(**H**) qRT-PCR analysis of mRNA level of *SOX1* and D-V markers in NPCs of "-SHH" and "+SHH" groups at day 4, relative to undifferentiated WT hESCs (day 0). n=3 independent experiments. Data are shown as mean ± SEM.

Figure S2

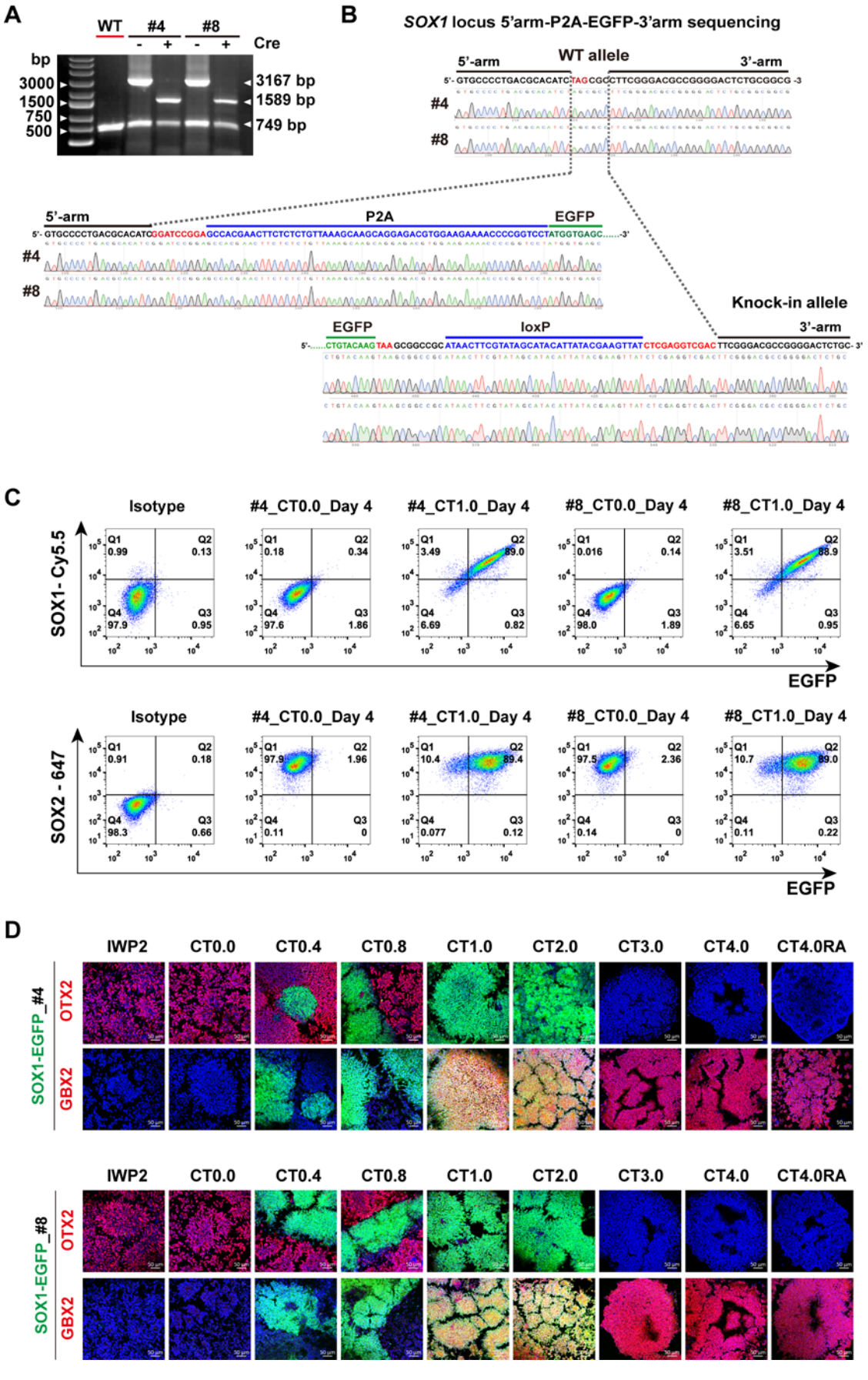

A
WT
#4
#8
Cre
bp
3000
1500
750
500
3167 bp
1589 bp
749 bp
B
SOX1 locus 5'arm-P2A-EGFP-3'arm sequencing
5'-arm
WT allele
3'-arm
#4
#8
P2A
EGFP
Knock-in allele
loxP
C
Isotype
#4_CT0.0_Day 4
#4_CT1.0_Day 4
#8_CT0.0_Day 4
#8_CT1.0_Day 4
SOX1- Cy5.5
SOX2 - 647
EGFP
D
IWP2
CT0.0
CT0.4
CT0.8
CT1.0
CT2.0
CT3.0
CT4.0
CT4.0RA
SOX1-EGFP_#4
SOX1-EGFP_#8
OTX2
GBX2

Fig S2. Related to Fig 1.

(**A**) The representative genomic DNA PCR result showing the correct targeting and the excision of the PGK-NeoR cassette after Cre-mediated recombination in SOX1-EGFP reporter hESCs of the clones #4 and #8.

(**B**) Representative Sanger sequencing peak maps verified precise joining of 5' homologous arm with the P2A-EGFP cassette, and a *loxP* site remained with 3' homologous arm after Cre-mediated recombination in SOX1-EGFP reporter hESCs of clones #4 and #8.

(**C**) Flow cytometric profiles of SOX1 and EGFP as well as SOX2 and EGFP in NPCs derived from SOX1-EGFP reporter hESCs of clones #4 and #8 under indicated conditions. Similar results were obtained in 3 independent experiments.

(**D**) Representative results of immunofluorescence staining of NPCs derived from SOX1-EGFP reporter hESCs of clones #4 and #8 using antibodies against OTX2 and GBX2, respectively, at day 8. The nucleus was stained by DAPI. Scale bars, 50 μM.

Figure S3

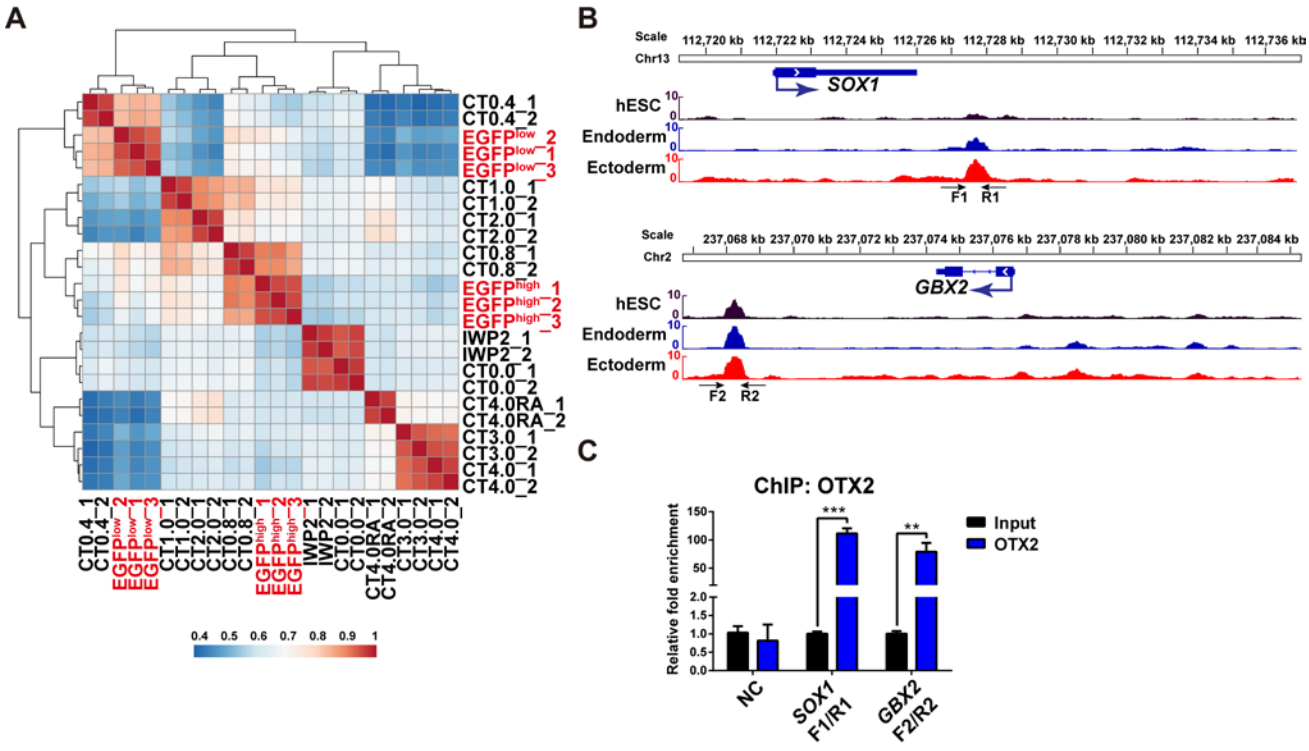


Fig S3. Related to Fig 2.

(**A**) Unsupervised hierarchical clustering and pearson correlation between EGFP^{high}, EGFP^{low} and WT hESC-derived NPCs with different regional identities at day 8.

(**B**) ChIP-seq signal profiles of OTX2 at the *SOX1* and *GBX2* loci in hESCs, endoderm and ectoderm cells were obtained from published data.

(**C**) ChIP-qPCR results of OTX2 at the *SOX1* and *GBX2* loci as well as a negative control region in NPCs of the CT0.4 group at day 4. n = 3 independent experiments. Data are shown as mean ± SEM. ** p < 0.01, *** p < 0.001. NC: negative control, a gene desert region (chr21:25509072+25509220, GRCh37/hg19).

Figure S4

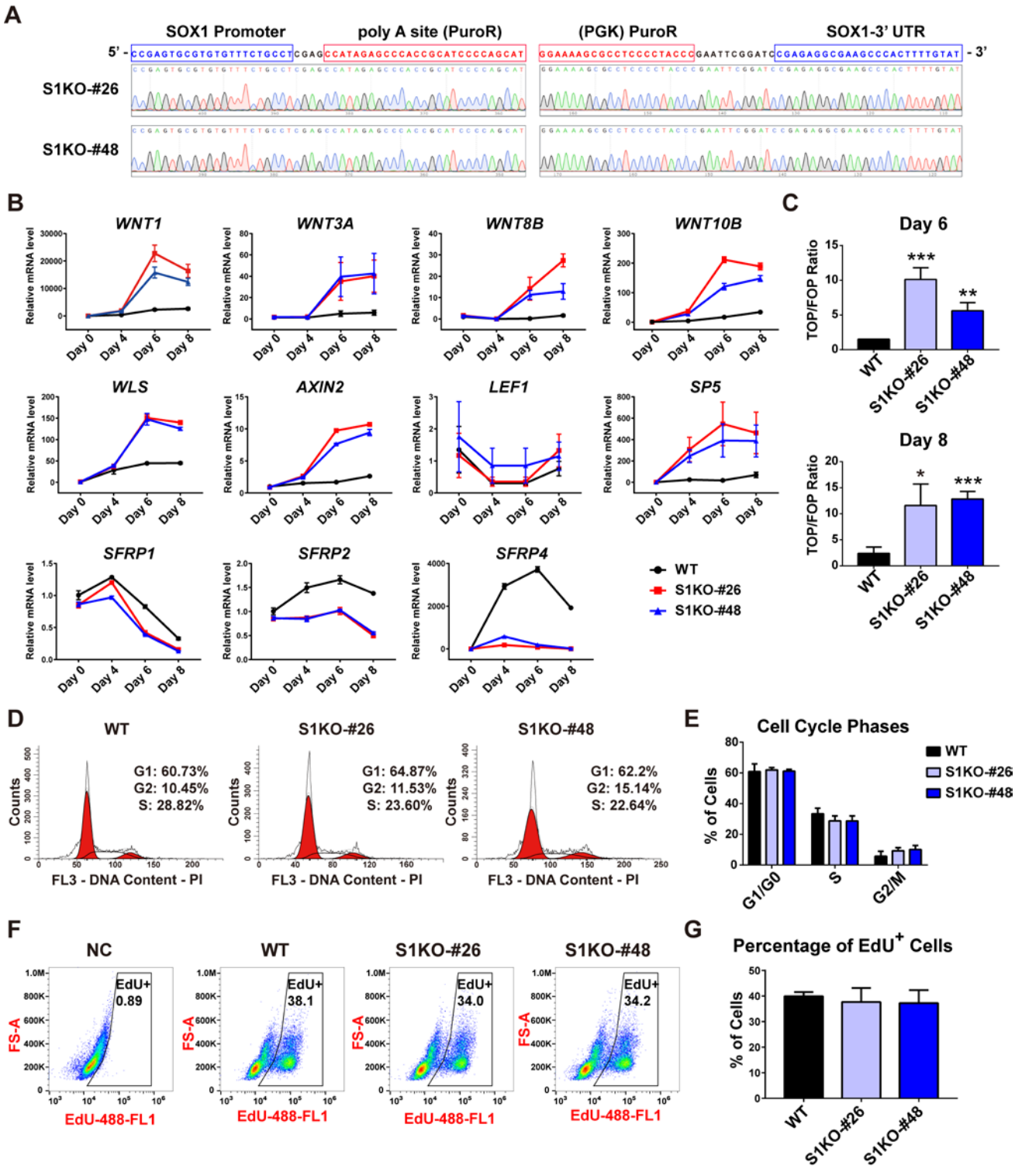


Fig S4. Related to Fig 3.

(**A**) Representative Sanger sequencing peak maps verified the replacement of the CDS region of *SOX1* by the PGK-Puro cassette in hESCs of S1KO-#26 and S1KO-#48 clones.

(**B**) qRT-PCR analysis of mRNA levels of WNT signaling associated genes in WT and *SOX1*-KO NPCs (CT1.0) at the indicated time points, relative to undifferentiated WT hESCs (day 0). n = 3 independent experiments. Data are shown as mean ± SEM.

(**C**) Examination of WNT signaling activity by TOP/FOPFlash reporter assays in WT and *SOX1*-KO NPCs (CT1.0) at day 6 and day 8. n = 3 independent experiments. Data are shown as mean ± SEM. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

(**D**) A representative flow cytometric analysis of the cell cycle using propidium iodide (PI) DNA staining assays in WT and *SOX1*-KO NPCs (CT1.0) at day 8.

(**E**) Quantitative analysis of results from the experiment shown in (D). n = 3 independent experiments. Data are shown as mean ± SEM.

(**F)** A representative flow cytometric analysis of the cell proliferation rate by EdU labeling assays in WT and *SOX1*-KO NPCs (CT1.0) at day 8. NC: WT NPCs (CT1.0) at day 8 without EdU treatment.

(**G**) Quantitative analysis of results shown in (F) n = 3 independent experiments. Data are shown as mean ± SEM.

Figure S5

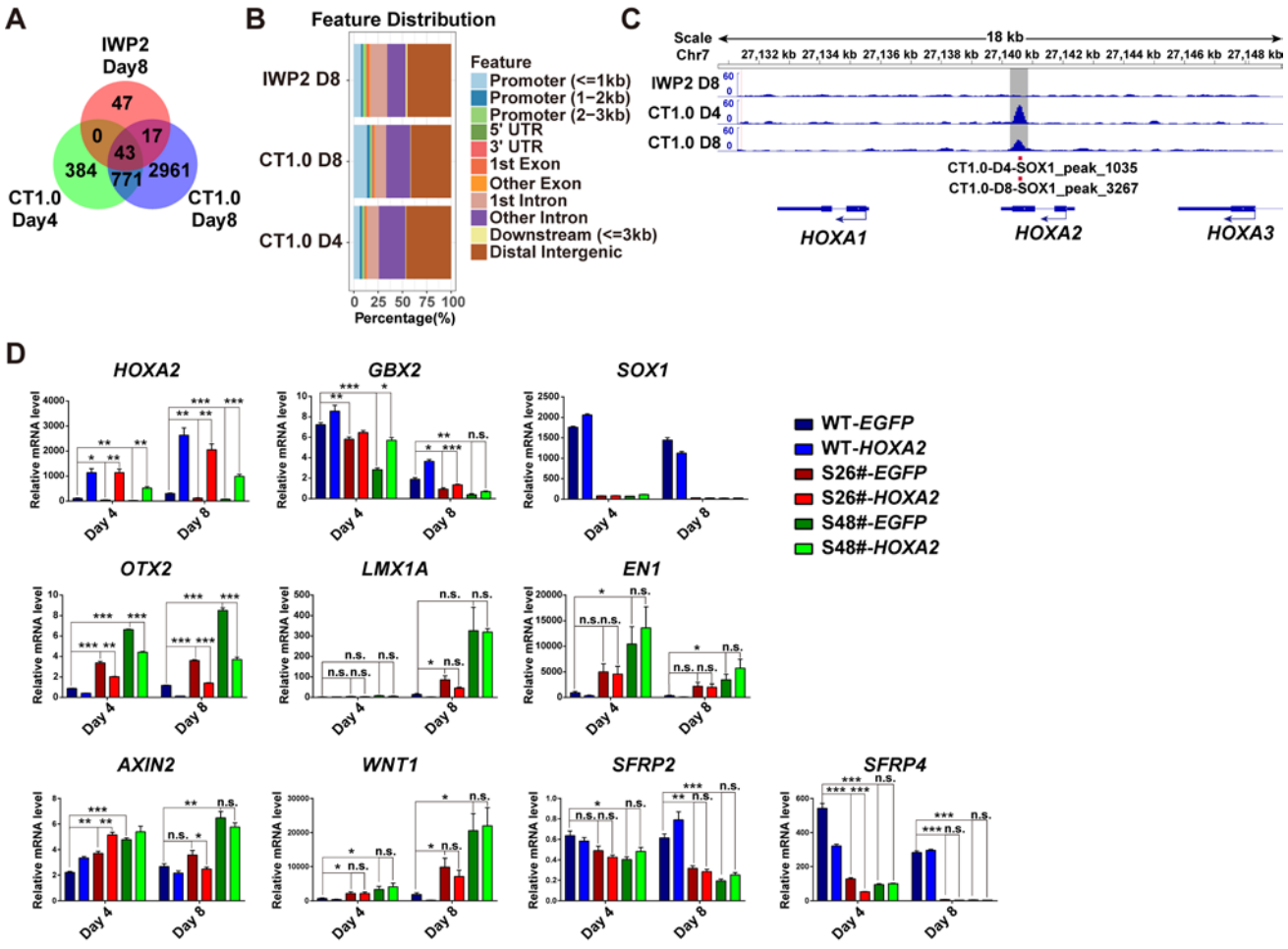


Fig S5. Related to Fig 4.

(**A**) The Venn diagram showing the number of SOX1 binding regions in NPCs of IWP2 day 8, CT1.0 day 4 and CT1.0 day 8 groups.

(**B**) Genomic distribution of SOX1-binding regions relative to their nearest RefSeq genes using the cis-regulatory element annotation system.

(**C**) The SOX1 binding region at the *HOXA2* locus in NPCs of the three groups indicated in Fig. 4A.

(**D**) qRT-PCR analysis of mRNA levels of neural markers and WNT signaling related genes at indicated time points in WT and *SOX1*-KO NPCs (CT1.0) overexpressing *EGFP* or *HOXA2*, relative to undifferentiated WT hESCs (day 0). n = 3 independent experiments. Data are shown as mean ± SEM. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, n.s.: not significant.

Supplemental tables

Table S1. The list of qRT-PCR primers used in this study, related to Figs 1-5, S1, S3, S4 and S5.

qPCR Primers		
Gene	**Forward 5' to 3'**	**Reverse 5' to 3'**
GAPDH	GCACCGTCAAGGCTGAGAAC	AGGGATCTCGCTCCTGGAA
SOX1	GGAATGGGAGGACAGGATTT	ACTTTTATTTCTCGGCCCGT
SOX2	TGCTGCCTCTTTAAGACTAGGAC	CCTGGGGCTCAAACTTCTCT
SOX3	TGGAGAACTGCAACGCCTACGC	GATCACGGCAGAAATCACCAACTC
NESTIN	TTCCCTCAGCTTTCAGGACCCCAA	AAGGCTGGCACAGGTGTCTCAA
FOXG1	GCAGCACTTTGAGTTACAACGGCA	AGTTCTGAGTCAACACGGAGCTGT
OTX2	ACAAGTGGCCAATTCACTCC	GAGGTGGACAAGGGATCTGA
OTX1	CACTAACTGGCGTGTTTCTGC	GGCGTGGAGCAAAATCG
LMX1A	AGGCCATCGCTCCCAATGAG	TGGTTCTCGGACGTTTGGGG
LMX1B	TTCCTGATGCGAGTCAACGAG	GCAGTACAGTTTCCGATCCCG
EN1	CCGCGCACCAGGAAGCTGAA	CAGCGCCAGGCCGTTCTTGA
EN2	CATGGCACAGGGCTTGTA	TACTCGCTGTCCGACTTGC
GBX2	AAAGAGGGCTCGCTGCTC	ATCGCTCTCCAGCGAGAA
KROX20	TTGACCAGATGAACGGAGTG	TGGTTTCTAGGTGCAGAGACG
HOXA2	CGTCGCTCGCTGAGTGCCTG	TGTCGAGTGTGAAAGCGTCGAGG
PAX6	TCTTTGCTTGGGAAATCCG	CTGCCCGTTCAACATCCTTAG
PAX7	ACCCCTGCCTAACCACATC	GCGGCAAAGAATCTTGGAGAC
NKX2.1	CGCATCCAATCTCAAGGAAT	CAGAGTGTGCCCAGAGTGAA
NKX2.2	AAACCATGTCACGCGCTCA	GGCGTTGTACTGCATGTGCT
WNT1	GCCATTGAACAGCTGTGAGC	CGTGGCTCTGTATCCACGTT
WNT3A	GTGTTCCACTGGTGCTGCTA	CCCTGCCTTCAGGTAGGAGT
WNT8B	TTGTCGATGCCCTGGAAACA	TTGAGTGCTGCGTGGTACTT
WN10B	TGAGCTCGGTGAGAGCAAAG	TTAAACCGTGGGGAGACTGC
WLS	CACAACGGCAGTGTCCTACA	ACTCATCTCCATGTGGGGGA
AXIN2	ACTTCTGGTTTGCCTGCAATGGA	GTGGCAGGCTTCAGCTGCTT
LEF1	CAGATCACCCCACCTCTTGG	GTGAGGATGGGTAGGGTTGC

SP5	GCACGTCAAGACTCACCAGA	CATTTTGGGAGGCAGGCAAC
SFRP1	ACCACCGTCTGTCTCAGAGT	TCTGTGCCTACAGAGAGCCT
SFRP2	AAGGAAAAGCCCACCCGAAT	ACAACAACCAACCAGACCCA
SFRP4	GCGGAGAACAGTTCAGGACA	AGTCGGAAGTCTCCGCTTTG
OTX2 ChIP-qPCR primers		
NC	GGGGGATCAGATGACAGTAAA	AATGCCAGCATGGGAAATA
***SOX1*-F1/R1**	GGGCCGAGGGTTTTGTGTAG	AAGGGTTTGACTTGCGGCT
***GBX2*-F2/R2**	TATTTCACAGCCACTGGGTCC	GGGCCCTTTAGGACGAAGG

Table S2. The list of genomic DNA PCR primers used in this study, related to Figs 3, 5, S1 and S2.

Primers for construction of pBSK-SOX1-P2A-EGFP donor plasmid	
LeftArm-F	AAGCTTCCTGAACGTTGAGCACTAGTGT
LeftArm-R	TCCGGATCCGATGTGCGTCAGGGGCACCG
RightArm-F	GTCGACGCCTTCGGGACGCCGGGGACTCT
RightArm-R	ATCGATTGAATTGGAGCGACACTGCT
Primers for construction of pBSK-SOX1-Puro-Donor plasmid	
PuroF	TGGGTACCGTCGACGGGTAGGGGAGGCGCTTTT
PuroR	CTGAATTCCCATAGAGCCCACCGCATCC
S1KO-LA-F	GTCGACGAAAAGCGTGCGCCATATCA
S1KO-LA-R	GGATCCGAATTCCTCGAGGCAGAAACACACGCACTCG
S1KO-RA-F	GAATTCGGATCCGAGAGGCGAAGCCCACTTTT
S1KO-RA-R	GCGGCCGCCATATGGGCTCACTTTTGGACGGACA
Primers for verification of S1B region and O2B region knockout	
S1BS-F	TCCTCCCCATCTCCAGACAG
S1BS-R	GATTTCAGCCCCGTGTGTCT
O2BS-F	GGGACGAGGTCAGGATTTGG
O2BS-R	CCCTCATCCCCTCTCTGGAA

Table S3. The list of guide sequences of individual sgRNAs, related to Figs 2, 3 and 5.

pX335-SOX1-StopA	GAAGGCGCTAGATGTGCGTC
pX335-SOX1-StopB	GGACGCCGGGGACTCTGCGG
pX462-SOX1-UpA	ATGCTGTACATCGGGGCGGC
pX462-SOX1-UpB	ACCGACCTGCACTCGCCCGG
Ctrli	GGAGACGGACGTCTCC
OTX2i-g1	TGCTCCAAACCCACCCACCA
OTX2i-g2	GGGCTGGTTTACTGCTTCGG
S1B-sgRNA1	TACATTCGGACTGTCAGAGC
S1B-sgRNA2	ACTGCCCATGACGGGCTACT
O2B-sgRNA1	GCATCGGGCAAGTCGGATGG
O2B-sgRNA2	CACGTGCTAGGCTAGGGCGA

Table S4. The list of antibodies used in this study, related to Figs 1-4 and S2.

Antibody	**Company**
SOX1	R&D, AF3369
OTX2	R&D, AF1979
FOXG1	Abcam, ab18259
EN1	DSHB, 4G11
SOX2	Custom
NESTIN	Millipore, MAB5326
GBX2	Sigma, HPA067809
GBX2	EPIGENTEK, A69507-050
HOXA2	Sigma, HPA029774
α-Tubulin	Sigma, T9026
NESTIN-Alexa647	BD Biosciences, 51-9007230
SOX1-Cy5.5	BD Biosciences, 561549
SOX2-Alexa647	BD Biosciences, 51-9006407
Isotype-Alexa647	BD Biosciences, 557783

Isotype-Cy5.5	BD Biosciences, 550795
Cy3-Donkey Anti-Mouse	Jackson ImmunoResearch 715-165-150
Cy3-Donkey Anti-Rat	Jackson ImmunoResearch 712-165-150
Cy3-Donkey Anti-Rabbit	Jackson ImmunoResearch 711-165-152
Cy3-Donkey Anti-Rabbit	Jackson ImmunoResearch 705-165-147
Alexa Fluor 555-Donkey Anti-Rabbit	Invitrogen, A-31572

Transparent methods

Construction of donor plasmids and sgRNA plasmids

For the pBSK-SOX1-P2A-EGFP donor plasmid used to generate SOX1-EGFP reporter hESC lines, the P2A-EGFP-loxP-NeoR-loxP cassette was synthesized and inserted into pBlueScript-II-SK (+). The left homologous arm fragment of about 3 kb in length was amplified by PCR from the genomic DNA of H9 hESCs with 2 primers (LeftArm-F and LeftArm-R). The right homologous arm fragment of about 2.75 kb in length was amplified by PCR from the genomic DNA with 2 primers (RightArm-F and RightArm-R). The primer sequences are provided in Table S2. For the pCAGGS-CRE-T2A-puro plasmid used to remove the NeoR cassette, the T2A-Puro cassette was synthesized and inserted into the pCAGGS-CRE plasmid (gift from Niwa's lab). For the pBSK-SOX1-Puro-Donor plasmid used to generate *SOX1*-KO hESC lines, the PGK-PuroR cassette was amplified by PCR from the pBigT (gift from Frank Costantini's lab) with 2 primers (PGK-Puro-F and PGK-Puro-R) and inserted into pBlueScript-II-SK (+). The left homologous arm fragment of about 2.9 kb in length was amplified by PCR from the genomic DNA of H9 hESCs with 2 primers (S1KO-LA-F and S1KO-LA-R). The right homologous arm fragment of about 2.4 kb in length was amplified by PCR from the genomic DNA of H9 hESCs with 2 primers (S1KO-RA-F and S1KO-RA-R). For sgRNA plasmids, Cas9 sgRNA vectors (pX335, or pX462 or pX459) were digested with BbsI, and gel extraction was performed to obtain the fragments of interest. A pair of oligos including targeting sequences was annealed and cloned into the BbsI-digested Cas9 sgRNA vector. The guide sequences of individual sgRNAs are provided in Table S3. All plasmids constructed in this study were sequenced to ensure their exactness.

Cell culture and differentiation of hESCs

The H9 (Karyotype, 46, XX) and SHhES2 (Karyotype, 46, XX) hESC lines were cultured on hESC-qualified Matrigel (BD, 354277) supplemented with mTeSR1 (Stemcell, #85850), and dissociated with Gentle Cell Dissociation Reagent (Stemcell, #07174) every 5 to 7 days for routine passaging. For neural differentiation, hESC colonies were detached with 1 mg/mL Dispase II (Gibco, 17105041) and suspended in neural differentiation medium, which was prepared as follows: DMEM/F12: Neurobasal (1:1, Gibco, 11320033), N2 (Gibco, 17502048), B27 without vitamin A (Gibco, 12587010), GlutaMax-I (Gibco, 35050061), MEM Non-Essential Amino Acids Solution (Gibco, 11140050), 5 μM SB431542 (Stemgent, 04-0010),

and 50 nM LDN193189 (Stemgent, 04-0074). Rho Kinase (ROCK)-inhibitor (Y-27632, 10 μM, Selleck, S1049) was present from day 0 to day 2. On day 4, aggregates were plated onto Matrigel-coated 6-well plates. From day 0 to day 8, IWP-2 (2 μM, Millipore, 686770), or CHIR99021 (0-4.0 μM, Stemgent,04-0004-10), or RA (1 μM, Sigma, R2625) was included in the differentiation medium for region-specific neural differentiation.

Electroporation

hESCs were cultured in mTeSR1 (Stemcell, #85850) with the Rho Kinase (ROCK)-inhibitor (Y-27632, 10 μM, Selleck, S1049) for 2 hrs prior to electroporation. Cells were dissociated by Accutase (Stemcell, #07920) for 8 minutes at 37 °C. Cells were dispersed into single cells, and 1×10^7 cells were electroporated with appropriate combinations of plasmids in 500 μL of the electroporation buffer (5 mM KCl, 5 mM $MgCl_2$, 15 mM HEPES, 102.94 mM Na_2HPO_4, 47.06 mM NaH_2PO_4, pH 7.2) using the Gene Pulser Xcell System (Bio-Rad, 165-2661) at 250 V, 500 μF in 0.4 cm cuvettes (Bio-Rad, #1652088) (Chen et al., 2015).

Generation of the SOX1-EGFP reporter hESC line

H9 hESCs were electroporated with a pair of sgRNA plasmids (pX335-SOX1-StopA and pX335-SOX1-StopB) as well as the donor plasmid pBSK-SOX1-P2A-EGFP, and screened with G418 (125 μg/mL) for 5 days. The clones were characterized by genomic DNA PCR with two primers: Primer-F (5'-CTGCCGCAGCACTACCAG-3'); Primer-R (5'-GTGCTTGGACCTGCCTTACT-3'). To excise NeoR cassette, the hESCs from #4, #6 and #8 clone were electroporated with the plasmid pCAGGS-CRE-T2A-puro, and screened with 0.5 μg/mL puromycin for 1 day. The colonies were individually picked and expanded, followed by genomic DNA PCR and sequencing analyses for successful recombination using Primer-F and Primer-R as mentioned above.

Generation of SOX1 knockout hESC lines

H9 hESCs were electroporated with two pairs of sgRNA plasmids（sgRNA-Up：pX462-SOX1-UpA and pX462-SOX1-UpB; sgRNA-Down: pX335-SOX1-StopA and pX335-SOX1-StopB） as well as donor plasmid pBSK-SOX1-Puro-Donor, and screened with 0.5 μg/mL puromycin for 5 days. The colonies were individually picked and expanded, followed by genomic DNA PCR with Primer-F and Primer-R as mentioned above. The guide sequences of individual

sgRNAs are provided in Table S3. The passages of *SOX1* knockout hESC lines used for neural differentiation ranged from 5 to 23 after *SOX1* deletion.

Generation of S1B region or O2B region knockout hESC lines

For S1B region knockout, H9 hESCs were electroporated with 2 sgRNA plasmids (pX459-S1B-sgRNA1 and pX459-S1B-sgRNA2). For O2B region knockout, H9 hESCs were electroporated with 2 sgRNA plasmids (pX459-O2B-sgRNA1 and pX459-O2B-sgRNA2). The cells were screened with 0.2 μg/mL puromycin for 2 days, and individual colonies were picked and expanded, followed by genomic DNA PCR and sequencing analyses. The primers for DNA PCR and sequencing analyses are provided in Table S2. The guide sequence of individual sgRNAs are provided in Table S3.

Generation of the OTX2 knockdown hESC line

The pAAVS1-CRISPRi plasmid (gift from Mohammad A. Mandegar) was used to generate the parental dCas9-KRAB SOX1-EGFP reporter hESCs (clone #6) with the dCas9-KRAB cassette integrated at the AAVS1 locus. The plentiGuide-puro vector was used to generate the sgRNA expressing plasmid (Addgene, #117986). The detailed method was described previously (Mandegar et al., 2016). Two sgRNAs targeting the 5' UTR of *OTX2* (OTX2i-g1 and OTX2i-g2) and control sgRNA (Ctrli) were synthesized and annealed into the BsmBI-digested plentiGuide-puro vector, respectively. These plasmids were used to generate *OTX2* knockdown and control knockdown SOX1-EGFP reporter hESCs using the parental dCas9-KRAB SOX1-EGFP reporter hESCs (clone #6) mentioned above. The guide sequences of individual sgRNAs are provided in Table S3. Details of lentiviral production and infection were previously described (Rubin et al., 2019).

Karyotyping

The well-grown SOX1-EGFP hESCs were treated with 100 ng/mL colchicine for 16 hrs and karyotyped as previously described (Ma et al., 2012).

Flow cytometric analysis and fluorescence-activated cell sorting

For flow cytometric analysis, cells were dissociated into single cells and counted. Then they were fixed with 4% PFA for 15 min at room temperature, washed 3 times with 1×PBS, and

incubated in the blocking buffer (10% Bovine Serum Albumin (BSA) and 0.2% Triton X-100 in PBS) for 30 min at room temperature, followed by incubation with antibodies for 1 hr at room temperature. The data of stained samples were acquired on the Beckman Gallios cytometer.

For fluorescence-activated cell sorting, cells were dissociated into single cells and counted. Then the cell suspension was analyzed by the Beckman Moflo Astrios cytometer. About 1×10^6 to 1×10^7 of $EGFP^{low}$ or $EGFP^{high}$ NPCs were collected. Subsequently, total RNA was extracted by the TRIzol reagent (Life Technologies, #15596026).

Luciferase reporter assays

Cells were dissociated into single cells and transfected with 500 ng of the 8×TOPFlash (Addgene, #12456) or 8×FOPFlash (Addgene, #12457) plasmid together with 50 ng of the pRL-TK internal control plasmid (Promega). After 48 hrs, luciferase activity was examined with the Dual Glow Luciferase Assay System (Promega, #E1960).

Cell cycle and cell proliferation analysis

For cell cycle analysis, cells were dissociated into single cells and counted. For each sample, 500 μL Cell Cycle Rapid Detection Solution (Dakewe, DKW41-DKK) was added into 1×10^6 cells and analyzed by the Beckman Gallios cytometer. For cell proliferation analysis, cells were incubated with EdU (10 μM; BD Biosciences) for 1 hr and labeled using a Cell-Light EdU Apollo488 *In Vitro* Flow Cytometry Kit (RIBO). Cells were then dissociated into single cells, fixed with 4% paraformaldehyde (PFA) solution, and permeabilized with 0.2% Triton X-100/3% BSA in PBS. Finally, the samples were analyzed by the Beckman Gallios cytometer.

ChIP-seq and ChIP-qPCR assays

ChIP-seq and ChIP-qPCR assays were performed as previously described (Zhu et al., 2017). All primers used in ChIP-qPCR assays are listed in Table S1.

For ChIP-seq, DNA libraries were constructed using the NEB Next Ultra DNA Library Prep Kit for Illumina, and sequenced using Illumina HiSeq X-ten at a target sequencing depth of 20 million reads. Published ChIP-Seq datasets (GEO:GSM1521760 for OTX2; GEO:GSM602291 for p300; GEO: GSM537679 for H3K4me1) were downloaded from the Gene Expression Omnibus for integrated analysis. For SOX1 ChIP-seq data, reads were

aligned to the hg19 reference assembly using bowtie2 with default parameter settings. Uniquely mapped reads were kept and extended by 200 bp for further analysis. We used MACS2 with q < 0.01 to identify significant binding events for SOX1. Peaks were then annotated according to their proximity to transcription start sites using Homer. ChIP-seq signal profiles at the specific locus were visualized with the Integrative Genomics Viewer (IGV).

RNA-seq

RNA was extracted from cell samples using the TRIzol reagent. Sequencing libraries for RNA with Poly(A) tails were prepared and sequenced on the Illumina Hiseq X-ten platform according to Illumina manufacturer instructions. Gene expression was quantified by the Salmon (Patro et al., 2017), and DEGs were analyzed by DESeq2 (Love et al., 2014) with following settings: fold changes > 2 and FDR < 0.05. GO analysis for DEGs was conducted using the GSEApy (https://github.com/zqfang/GSEApy) (Fang, 2020).

RNA extraction, cDNA synthesis and quantitative real time PCR (qRT-PCR)

Total RNA was extracted using the TRIzol reagent (Life Technologies, #15596026). Total RNA of 4 μg was reversely transcribed into cDNAs using a ReverTra Ace reverse transcriptase (Toyobo, #FSK-101) according to the manual. QRT-PCR was performed on the ABI ViiA7 Real-Time PCR system, using the SYBR Premix Ex Taq II (Takara, #RR820L) following manufacturer's instructions. GAPDH was used as an internal control, and a list of primers for qRT-PCR is provided in Table S1.

Genomic DNA extraction and PCR

Genomic DNA was extracted with the Pure Link™ Genomic DNA Mini Kit (Invitrogen, K1820-01). Genomic DNA PCR was carried out using the Q5® High-Fidelity DNA Polymerase (New England Biolabs, M0492L). A list of primers for genomic DNA PCR is provided in Table S2.

Western blot analysis

Cells were rinsed with PBS twice and lysed with the Co-IP buffer (50 mM Tris-HCl, 150 mM NaCl, 5 mM $MgCl_2$, 0.2 mM EDTA, 20% Glycerol, 0.1% NP-40, 3 mM β-Mercaptoethanol; pH 7.5) supplemented with a protease inhibitor cocktail (Selleck, B14001). Total protein

concentration was measured using the Pierce BCA Protein Assay Kit (Thermo Scientific, #23227), following the manufacturer's instructions. Total proteins of 20 μg (per lane) were separated by the SDS-PAGE and transferred to 0.45 μM nitrocellulose blotting membranes (GE Healthcare, #10600002). Membranes were blocked with 5% BSA diluted in TBST (19 mM Tris, 2.7 mM KCl, 137 mM NaCl, 0.1% Tween-20; pH 7.4) for 1 hr and incubated with specific primary antibodies overnight at 4 ^{o}C, followed by incubation with HRP-conjugated secondary antibodies for 1 hr. The signals were detected from blotted membranes by exposing to the SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific, #34580). All western blot analyses were conducted in at least three independent experiments. Antibodies used are listed in Table S4.

Immunofluorescence staining

Aggregates formed during neural differentiation were attached to Matrigel-coated coverslips (Fisherbrand, #12-545-82). They were fixed with 4% PFA for 15 min at room temperature, rinsed 3 times with 1×PBS, and incubated in the blocking buffer (10% donkey serum and 0.2% Triton X-100 in PBS) for 30 min at room temperature followed by incubation with primary antibodies overnight at 4 ^{o}C. Next day, coverslip cultures were rinsed 3 times with PBS, and incubated with secondary antibodies (Cy3, 1:200) for 1 hr at room temperature. The nuclei were stained with DAPI (Sigma, #D9542). Images were captured using a Zeiss Cell Observer microscope. Antibodies used are listed in Table S4.

Statistical analysis

The unpaired Student's *t* test was used for statistical tests unless stated otherwise. Data are shown as mean$\pm$SEM. Numbers of biological replicates relevant for individual experiments are stated in figure legends.

Supplemental references

Chen, Y., Cao, J., Xiong, M., Petersen, A.J., Dong, Y., Tao, Y., Huang, C.T., Du, Z., and Zhang, S.C. (2015). Engineering Human Stem Cell Lines with Inducible Gene Knockout using CRISPR/Cas9. Cell stem cell 17, 233-244.

Fang, Z. (2020). GSEApy: Gene Set Enrichment Analysis in Python. Zenodo.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 550.

Ma, Y., Li, C., Gu, J., Tang, F., Li, C., Li, P., Ping, P., Yang, S., Li, Z., and Jin, Y. (2012). Aberrant gene expression profiles in pluripotent stem cells induced from fibroblasts of a Klinefelter syndrome patient. The Journal of biological chemistry 287, 38970-38979.

Mandegar, Mohammad A., Huebsch, N., Frolov, Ekaterina B., Shin, E., Truong, A., Olvera, Michael P., Chan, Amanda H., Miyaoka, Y., Holmes, K., Spencer, C.I., et al. (2016). CRISPR Interference Efficiently Induces Specific and Reversible Gene Silencing in Human iPSCs. Cell stem cell.

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nature methods 14, 417-419.

Rubin, A.J., Parker, K.R., Satpathy, A.T., Qi, Y., Wu, B., Ong, A.J., Mumbach, M.R., Ji, A.L., Kim, D.S., Cho, S.W., et al. (2019). Coupled Single-Cell CRISPR Screening and Epigenomic Profiling Reveals Causal Gene Regulatory Networks. Cell 176, 361-376 e317.

Zhu, Z., Li, C., Zeng, Y., Ding, J., Qu, Z., Gu, J., Ge, L., Tang, F., Huang, X., Zhou, C., et al. (2017). PHB Associates with the HIRA Complex to Control an Epigenetic-Metabolic Circuit in Human ESCs. Cell stem cell 20, 274-289 e277.