Stem Cell Reports
Article

ISSCR

OPEN ACCESS

SOX21 Ensures Rostral Forebrain Identity by Suppression of WNT8B during Neural Regionalization of Human Embryonic Stem Cells

Zhuoqing Fang,[1,6] Xinyuan Liu,[1,6] Jing Wen,[1] Fan Tang,[2] Yang Zhou,[1,5] Naihe Jing,[3,4] and Ying Jin[1,2,4,*]

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

[2]Basic Clinical Research Center, Renji Hospital, 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]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

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

[5]Present address: Department of Biomedical Engineering, School of Medicine and School of Engineering, The University of Alabama at Birmingham, Birmingham, AL 35294-0019, USA

[6]Co-first author

*Correspondence: yjin@sibs.ac.cn

https://doi.org/10.1016/j.stemcr.2019.10.013

SUMMARY

The generation of brain region-specific progenitors from human embryonic stem cells (hESCs) is critical for their application. However, transcriptional regulation of neural regionalization in humans is poorly understood. Here, we applied a rostrocaudal patterning system from hESCs to dissect global transcriptional networks controlling early neural regionalization. We found that SOX21 is required for rostral forebrain fate specification. *SOX21* knockout led to activation of Wnt signaling, resulting in caudalization of regional identity of rostral forebrain neural progenitor cells. Moreover, we identified *WNT8B* as a SOX21 direct target. Deletion of *WNT8B* or inhibition of Wnt signaling in *SOX21* knockout neural progenitor cells restored rostral forebrain identity. Furthermore, SOX21 interacted with β-catenin, interfering with the binding of TCF4/β-catenin complex to the *WNT8B* enhancer. Collectively, these results unveil the unknown role of SOX21 and shed light on how a transcriptional factor modulates early neural regionalization through crosstalk with a key component of Wnt signaling.

INTRODUCTION

During early mammalian embryogenesis, the central nervous system is patterned to acquire regional identities and divides into dorsoventral and rostrocaudal compartments (Rubenstein and Rakic, 2013). Along the rostral-caudal (R-C) axis, the neural tube developes into the forebrain, midbrain, hindbrain, and the spinal cord at the early phase of R-C specification (Itasaki, 2015). Each region will give rise to specific subtypes of neural progenitor cells (NPCs) and neurons (Kiecker and Lumsden, 2012; Wurst and Bally-Cuif, 2001). Among these regions, the forebrain is the most complex, containing the telencephalon and diencephalon. The precise patterning of the forebrain is critically important for the proper wiring of the cerebrum.

Investigations of how neural regionalization is controlled during the early stages of development have been conducted primarily in model animals. Although many signaling molecules, including fibroblast growth factor, Wnt, and retinoic acid (RA), are involved in rostrocaudal patterning (Irioka et al., 2005; Tuazon and Mullins, 2015), Wnt is perhaps the most crucial (Andoniadou and Martinez-Barbera, 2013; Twyman, 2009). A morphogen gradient of Wnt signaling is important for rostrocaudal patterning, and low levels of Wnt/β-catenin activities are required to specify the telencephalic fate (Fossat et al., 2011, 2012; Kiecker and Niehrs, 2001; Nordstrom et al., 2002). Furthermore, RA plays a central role in inducing the posterior hindbrain and spinal cord (Glover et al., 2006). Despite these advances, the question of how neural cells acquire regional identities by the Wnt gradient is far from completely understood. Several Wnt ligands are expressed in the neural tube, such as Wnt1, Wnt3a, and Wnt8b (Ciani and Salinas, 2005). Among Wnt ligands, Wnt8b has been found to localize in the dorsal midline of the caudal forebrain at the eight-somite stage (Liu et al., 2010). The wedge-shaped expression domain of *Wnt8b* in the diencephalon has been suggested to indicate the location of the zona limitans intrathalamica (Puelles and Martinez, 2013). Moreover, Wnt8b has been reported to regulate the diencephalic versus telencephalic fate in zebrafish (Houart et al., 2002). However, how WNT8B expression is spatiotemporally regulated in the forebrain remains largely unknown.

Difficulties in obtaining and manipulating human embryonic neural tissues have severely hindered the investigation of human neural development. Highly robust methods for neural differentiation from human pluripotent stem cells (hPSCs) have been developed (Chambers

et al., 2009). However, the protocol to generate regionally specified NPCs from human PSCs was not available until recent years (Imaizumi et al., 2015; Kirkeby et al., 2012; Kriks et al., 2011). Similar to neural development *in vivo*, the rostrocaudal identity (ranging from the telencephalon to the spinal cord) can be induced by the dose-dependent control of Wnt and RA signaling, and the dorsoventral identity can be modulated by Shh signaling. Rostrocaudally patterned NPCs have been obtained by application of the porcupine inhibitor IWP-2 (IWP2, a Wnt signaling inhibitor), different dosages of the glycogen synthetase kinase 3 (GSK3) inhibitor CT99021 (CT, a Wnt signaling activator), or CT combined with RA (Imaizumi et al., 2015; Kirkeby et al., 2012). The system provides a useful tool to address the question of how Wnt signaling is precisely modulated for the establishment and maintenance of the rostrocaudal identity in early human neural development.

In this study, we adopted a robust protocol to control the regional identities of human embryonic stem cell (hESC)-derived NPCs and focused on the role of transcriptional regulation for Wnt signaling and cell identity in positional patterning. Through conducting genome-wide transcriptional analyses for rostrocaudally specified NPCs ranging from the telencephalic forebrain to the spinal cord, we revealed that the transcriptional factor SOX21 plays an important role in forebrain regionalization. SOX21 is a member of the B2 group of SOX family proteins, harboring a high-mobility group DNA binding domain (Kamachi and Kondoh, 2013; Uchikawa et al., 1999). The expression of SOX21 initiates in the anterior neural plate of mouse embryos at the embryonic day 8.0 (E8.0) (Uchikawa et al., 2011). Previous studies have suggested roles of Sox21 in the progression of neurogenesis in the mouse, chicken, and *Xenopus* (Matsuda et al., 2012; Sandberg et al., 2005; Whittington et al., 2015). However, the role of SOX21 in neural patterning, particularly in human neural development, remains unclear. Here, we reveal that SOX21 represses WNT8B expression, participating in the control of forebrain regionalization. This study unveils a previously unappreciated role of SOX21 and sheds light on the transcriptional control of rostrocaudal patterning during early neural differentiation of hESCs.

RESULTS

The Rostrocaudal Patterning of hESC-Derived NPCs

To dissect the transcriptional network controlling the rostrocaudal specification of early neural differentiation, we generated NPCs from hESCs of the H9 line (Thomson et al., 1998) with defined regional identities ranging from the rostral forebrain to the spinal cord based on published protocols (Imaizumi et al., 2015; Kirkeby et al., 2012; Kriks et al., 2011). Initially, hESCs were suspended in N2B27 medium containing dual SMAD inhibitors (LSB: 50 nM LDN193189 and 5 μM SB431542) and patterning factors, including IWP2, CT, and RA, for 4 days. At day 4, aggregates were attached to culture dishes and cultured for a further 4 days to yield NPCs with more specific regional identities along the R-C axis (Figure 1A). Generally, rostrocaudal patterning was established by the dose-dependent control of WNT signaling. Inhibition of Wnt signaling by IWP2 (2 μM) induced the rostral forebrain fate, while activation of Wnt signaling by CT induced the diencephalic, mesencephalic, and rhombencephalic fate, depending on CT concentrations ranging from 0.4 to 4.0 μM. RA (1.0 μM) coupled with 4.0 μM CT (CT4.0RA) was used to induce the spinal cord fate. The ventralization factor was not added in the current differentiation system because we mainly focused on the transcriptional regulation of rostrocaudal patterning.

We first examined the morphology of aggregates at days 4 and 8 (Figure 1B). Consistent with a report by Imaizumi et al. (2015), the size of the aggregates increased with higher CT concentrations. The proportion of $SOX2^{+}NANOG^{-}$ NPCs was very high (>97%) under all differentiation conditions used (Figure 1C), suggesting the successful neural induction. To validate the regional features of these NPCs, we then carried out RNA sequencing (RNA-seq) for cell samples collected from 19 groups (each with 2 biological replicates), covering 3 time points (days 0, 4, 8) and 9 differentiation conditions (IWP2, CT0.0, CT0.4, CT0.8, CT1.0, CT2.0, CT3.0, CT4.0, CT4.0RA) (Figures 1A and 1B). To analyze transcriptomes of the positionally patterned NPCs, we applied unsupervised hierarchical clustering and found six major regions along the R-C axis at day 8, designated as region 1 (IWP2 and CT0.0), region 2 (CT0.4), region 3 (CT0.8), region 4 (CT1.0 and CT2.0), region 5 (CT3.0 and CT4.0), and region 6 (CT4.0RA) (Figure 1D). The six regions could already be classified at day 4 (Figure S1A). To better characterize gene expression patterns along the R-C axis, we applied principal-component analysis (Figures 1E and S1B). The second principal component accounted for the trajectory of rostrocaudal patterning, while the first principal component separated region 5 (NPCs treated with CT 3.0–4.0 μM at day 8) from the rest, indicating that these NPCs might have unique characteristics (Figure 1E). In addition, we compared our RNA-seq data with the CORTECON system (van de Leemput et al., 2014). Spearman correlation analysis revealed that all of our positionally patterned NPCs highly correlated with the neural differentiation stage (day 7), suggesting the progenitor nature of our NPCs (Figure 1F).

To systematically investigate the coexpression relationships among genes in positionally patterned NPCs, we performed weighted gene coexpression network analysis

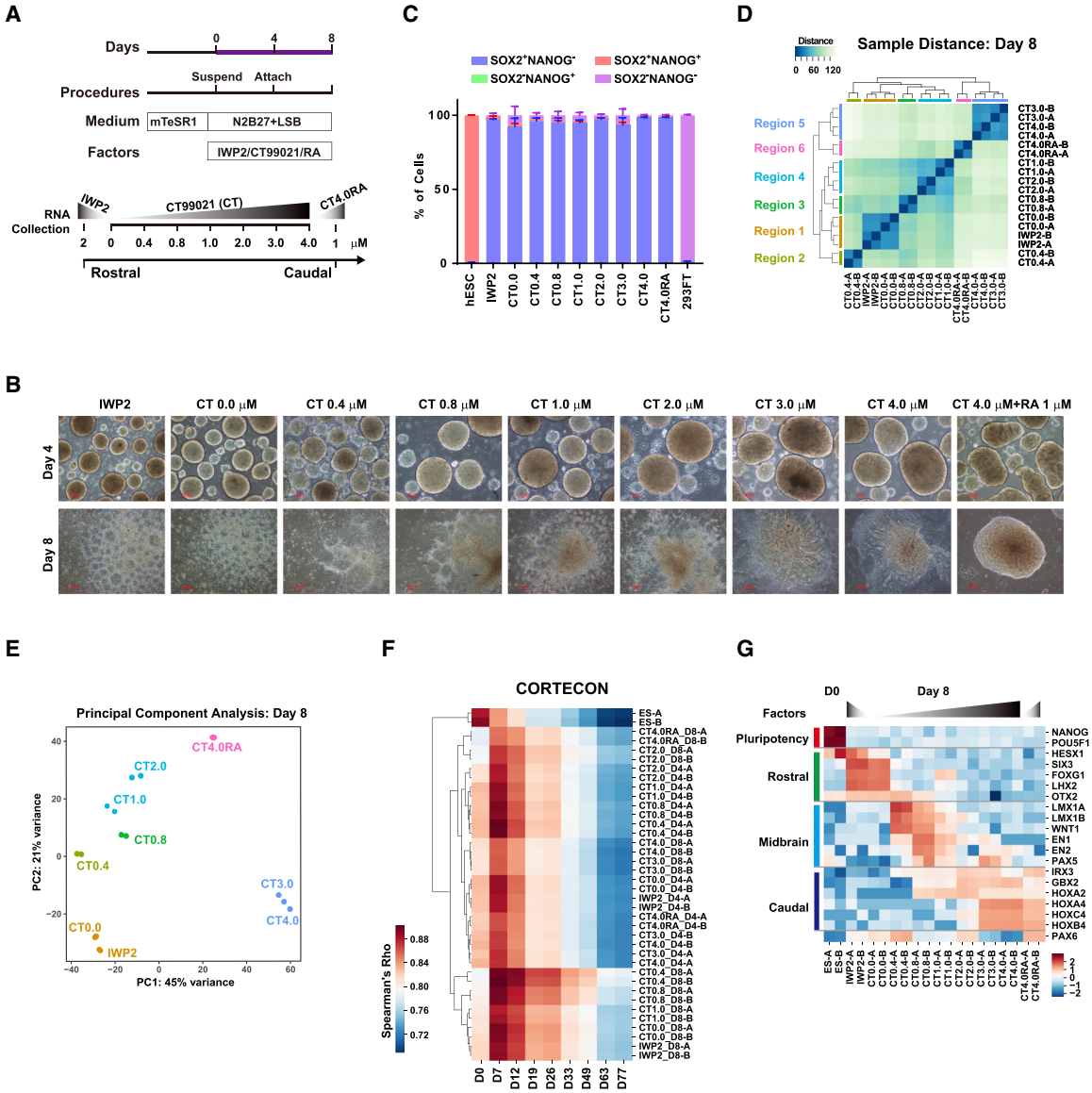


Figure 1. Transcriptome Analysis of the Rostrocaudal Patterning System

(A) A schematic illustration of an aggregation-based rostrocaudal neural differentiation protocol. Differentiation was induced by LSB (50 nM LDN193189 and 5 μM SB431542). The following patterning factors were used for the generation of region-specific neural progenitor cells along the rostrocaudal axis: 2 μM IWP-2 (IWP2), 0–4.0 μM CT99021 (CT), and 1 μM RA with 4.0 μM CT (CT4.0RA) from days 0 to 8.

(B) Bright-field images of neural aggregates at day 4 and rosettes at day 8. The RNA samples for RNA-seq were collected at days 4 and 8 from all groups. Scale bars, 100 μm.

(C) The percentage of $SOX2^{+}NANOG^{-}$ cells quantified by flow cytometric analysis at day 8. 293FT cells and undifferentiated WT hESCs were used as controls. Data are shown as means ± SEM; n = 3.

(D) Unsupervised hierarchical clustering of 18 samples at day 8.

(E) Principal-component analysis of 18 samples at day 8.

(F) Comparative analysis of our RNA-seq data with the CORTECON neural differentiation system. Spearman correlations with different days of CORTECON samples are shown.

(G) A heatmap of RNA-seq results for region-specific marker genes at day 8. Suffixes A and B represent biological replicates.

See also Figure S1.

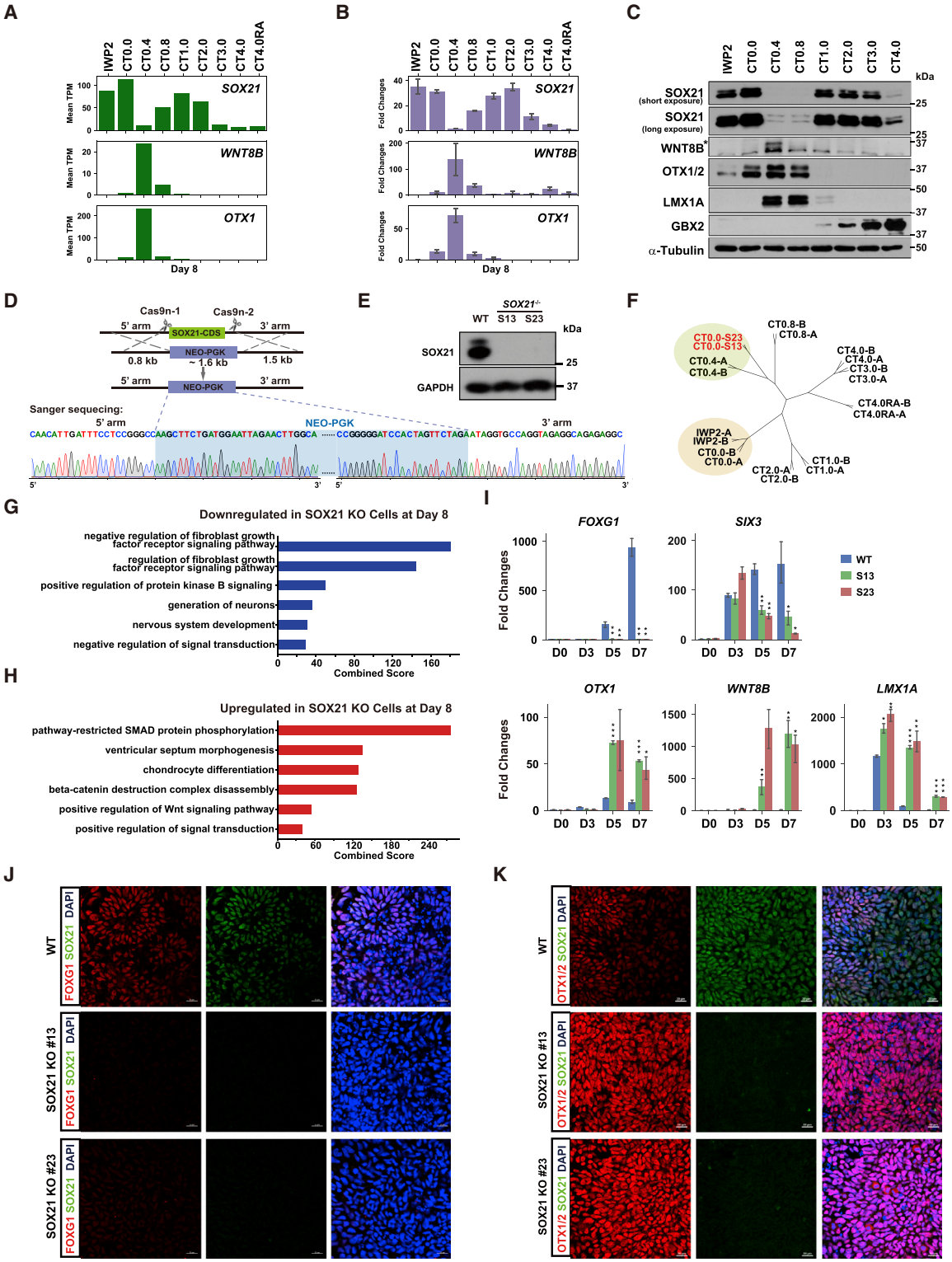


Figure 2. SOX21 Participates in Forebrain Regionalization

(A) Expression levels of *SOX21*, *WNT8B*, and *OTX1* at day 8 of rostrocaudal differentiation obtained from our RNA-seq data.

(B) Real-time qPCR validation of *SOX21*, *WNT8B*, and *OTX1* expression levels at day 8. Data are represented as fold changes relative to undifferentiated WT hESCs (day 0) and shown as means ± SEM; n = 3.

(legend continued on next page)

(WGCNA) and identified 26 distinct coexpression modules corresponding to clusters of correlated genes (Figure S1C; Table S1). Notably, seven modules showed region-specific expression patterns and might represent core gene networks along the R-C axis (Figure S1D). Functional enrichment analysis was conducted to characterize these modules (Figure S1E). The Darkgrey module (region 1) was enriched for terms such as anteroposterior axis specification and embryonic axis specification, including genes such as *SIX3*, *LHX2*, *HESX1*, and *FEZF1*, which started to be expressed as early as day 4. Terms enriched in the Darkgreen module (region 1) were extracellular organization and metabolic process, including genes such as *FOXG1* and *DLX5*, which were induced at day 8. Although the Pink module (region 2) was enriched for terms highly correlated with midbrain development, such as *LMX1A*, *LMX1B*, and *WNT1*, some genes that are known to also be expressed in the caudal forebrain were present, such as *WNT8B*, *EMX2*, *OTX1*, and *DLX2*. Caudal region modules, including the Lightgreen, Lightcyan, Black, and Grey60, were enriched for terms such as nervous system development, regionalization, and morphogenesis. In addition, the Turquoise module (day 0, undifferentiated hESCs) was enriched for pluripotency-related terms (Figure S1E). Taken together, these analyses provide rich information for unveiling gene coexpression networks controlling positional patterning.

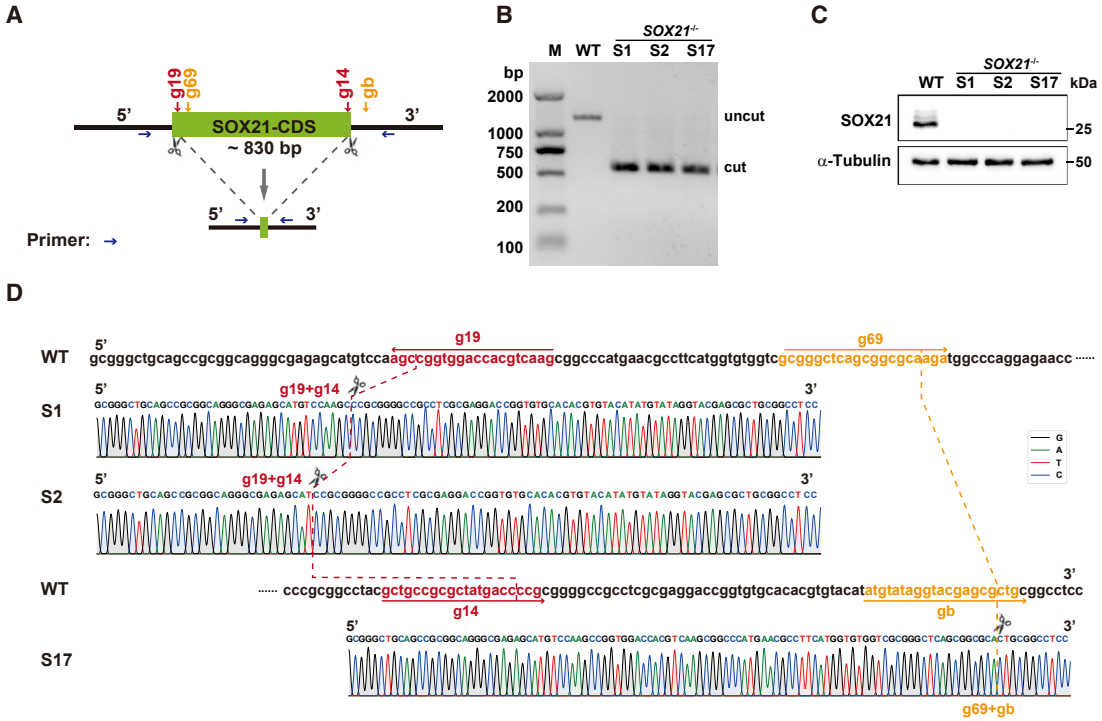
With these RNA-seq data, we examined the relative gene expression levels of known regional markers (Figure 1G). The simultaneously high expression of *HESX1*, *SIX3*, *FOXG1*, and *LHX2* in NPCs from the IWP2 or 0.0 μM CT group (region 1) indicated a rostral forebrain identity. Some midbrain markers (*LMX1A*, *LMX1B*, and *WNT1*) were detected with CT concentrations at 0.4 to 1.0 μM, whereas other midbrain markers (*EN1*, *EN2*, *PAX5*) were highly expressed in NPCs treated with 0.8 μM CT (region 3). NPCs from groups treated with CT concentrations at and above 2.0 μM highly expressed markers of the hindbrain, including *GBX2*, *HOXA2*, *HOXA4*, *HOXC4*, and *HOXB4*. Notably, NPCs from the CT4.0RA group (region 6) highly expressed *PAX6*, *HOXB4*, and *HOXC4*, representing the spinal cord identity. Altogether, these results verified that patterning factors (IWP2/CT/RA) efficiently induced the rostrocaudal identities of NPCs, recapitulating gene expression patterns *in vivo* to some extent.

SOX21 Is Required for Rostral Forebrain Specification from hESCs

The aforementioned WGCNA analysis revealed that *WNT8B* was specifically expressed in NPCs of the 0.4 μM CT group (region 2) (Figure S1D). Interestingly, our RNA-seq results showed that the mRNA of *SOX21* was very low in NPCs of the 0.4 μM CT group (Figure 2A). The expression patterns of SOX21, WNT8B, and OTX1 at both the mRNA and protein levels were verified by real-time qPCR and western blot analysis (Figures 2B and 2C). A reverse relationship appeared between the expression levels of SOX21 and WNT8B in NPCs of the 0.4 μM CT group. Previously, *Sox21* expression has been reported to be very low in the dorsal diencephalon at the 11- to 12-somite stage of mouse embryos (Uchikawa et al., 2011), whereas *Wnt8b* has been detected at the diencephalic-telencephalic boundary of mouse embryos at E10.5 (Fotaki et al., 2010) and at that of human embryos at Carnegie stages 15 and 16 (Lako et al., 1998). Given the unique expression pattern of *Sox21* and *Wnt8b* as well as the implications of Wnt8b in forebrain patterning (Houart et al., 2002), we anticipated that SOX21 might have a role in the control of *WNT8B* expression and forebrain patterning . To test this hypothesis, we knocked out *SOX21* in hESCs of the H9 line (Thomson et al., 1998) using the CRISPR/Cas9 nickase system via homology-directed repair (Figure 2D). Two clones (S13 and S23) harboring homozygous *SOX21* deletions were established, and the absence of SOX21 in NPCs was verified by western blot and Sanger sequencing analyses (Figures 2D and 2E).

To obtain evidence for a role of SOX21 in early forebrain patterning, we generated rostral forebrain NPCs from hESCs

(C) Representative western blot analysis of SOX21, WNT8B, OTX1/2, LMX1A, and GBX2 at day 8. Antibodies of OTX1/2 recognize both OTX1 and OTX2. A star indicates the specific band of WNT8B proteins.
(D) The strategy to establish *S0X21* KO clones using the CRISPR/Cas9 nickase with four gRNAs. A map of Sanger sequencing results shows the homology recombination site in *S0X21* KO hESCs. The PGK-NEO cassette is highlighted with light blue. CDS, coding sequence.
(E) Representative western blot analysis of SOX21 protein level in hESC-derived NPCs. WT, wild-type cells; S13 and S23 are two *S0X21* KO clones.
(F) Hierarchical clustering between *S0X21* KO samples (S13 and S23) and WT rostrocaudal differentiation samples at day 8.
(G and H) GO terms enriched for downregulated (G) and upregulated (H) genes in *S0X21* KO cells. Only the top 6 GO terms with FDR <0.05 and combined score >15 are listed.
(I) Real-time qPCR validation of DEGs at days 0, 3, 5, and 7 (CT = 0.0 μM). Data are represented as fold changes relative to undifferentiated WT hESCs (day 0) and shown as means ± SEM. *p < 0.05, **p < 0.01, ***p < 0.001 (n = 3).
(J and K) Typical immunofluorescence images of FOXG1, OTX1/OTX2, and SOX21 at day 7. Antibodies against OTX1/OTX2 recognize both OTX1 and OTX2. Scale bars, 20 μm.
See also Figure S2.



(legend on next page)

based on the protocol described in Figure 1A (CT = 0.0 μM) (Kirkeby et al., 2012). This protocol was used hereafter in all neural differentiation experiments of this study. Between wild-type (WT) and *SOX21* knockout (KO) NPCs, the percentage of $SOX2^{+}NANOG^{-}$ cells was similar (Figure S2A), and there was no detectable difference in the NPC proliferation rate as quantified by EdU^{+} cells (Figure S2B). Next, to uncover genes regulated by SOX21 at a genome-wide scale, RNA was collected from NPCs at differentiation days 0, 4, and 8, with two biological replicates for RNA-seq. The Pearson's correlation coefficient was greater than 0.98 between the S13 and S23 clones, providing evidence for the specific effect of *SOX21* KO generated by the CRSPR/Cas9 approach (Figure S2C). Analysis of differentially expressed genes (DEGs) identified 363 upregulated and 367 downregulated genes (fold changes >2 and false discovery rate [FDR] <0.05) in *SOX21* KO NPCs at day 8 (Figure S2D). Notably, clustering analysis of DEGs revealed that the regional identity of *SOX21* KO NPCs was closer to that of NPCs treated with 0.4 μM CT at day 8 (Figure 2F), indicating that the regional identity of rostral forebrain NPCs was caudalized in the absence of *SOX21*. Furthermore, gene ontology (GO) analysis of DEGs showed that downregulated genes were enriched for the term of nervous system development, including rostral forebrain markers, such as *FOXG1* and *SIX3*, whereas upregulated genes were enriched for the term of positive regulation of Wnt signaling pathway, including *WNT8B* (Figures 2G, 2H, and S2D). Real-time qPCR analysis verified significant downregulation of *FOXG1* and *SIX3* as well as upregulation of *OTX1*, *LMX1A*, and *WNT8B* in *SOX21* KO cells compared with WT cells (Figure 2I). Immunofluorescence staining results also revealed a substantial reduction in FOXG1 and SIX3 expression (Figures 2J and S2E) but an increase in OTX1/OTX2 expression in *SOX21* KO NPCs (Figure 2K). Western blot analysis also showed higher protein levels of WNT8B and OTX1/OTX2 in *SOX21* KO NPCs than in WT NPCs (see results in Figure 5B). To confirm the function of SOX21 in early neural differentiation and avoid the cell line bias, we generated *SOX21* KO hESCs (clone ES8-S11) in another hESC line, SHhES8 (Zhu et al., 2017) using the same gene editing strategy described above. Alterations in gene expression similar to those in NPCs derived from *SOX21* KO H9 hESCs were observed (Figure S2F).

In addition, to further rule out potential off-target effects in the CRISPR/Cas9 nickase-mediated *SOX21* KO cells, we generated three additional *SOX21* KO clones (S1, S2, S17) using a donor-free paired gRNA-guided CRISPR/Cas9 KO strategy (paired-KO) (Liu et al., 2016) and two new sets of paired gRNAs (g19 + g14 and g69 + gb) (Figure 3A). PCR analysis suggested that the three *SOX21* KO clones had biallelic deletion of *SOX21* genomic DNA (Figure 3B). Western blot analysis and Sanger sequencing results validated the lack of SOX21 (Figures 3C and 3D). Consistently, *FOXG1*, *SIX3*, and *LHX2* were all repressed, while *WNT8B*, *OTX1*, *LMX1A*, and *IRX3* were significantly upregulated in *SOX21* KO NPCs (Figure 3E). Furthermore, we knocked down *SOX21* using two shRNAs, and similar alterations in gene expression were found in *SOX21* knockdown NPCs (Figure 3F). These results clearly show that the alterations observed in *SOX21* KO NPCs were specifically caused by the absence of *SOX21* and that SOX21 is required for the specification of the rostral forebrain NPCs from hESCs.

Identification of SOX21 Putative Target Genes

To understand how SOX21 participates in the control of forebrain regional identity, we carried out chromatin immunoprecipitation assays coupled to high-throughput DNA sequencing (ChIP-seq) in NPCs at day 5 and identified 3,765 confident (FDR q <0.01) binding peaks of SOX21 with a high signal-to-noise ratio (Figure S3A). Associations of SOX21 with some of these genes were verified by ChIP-qPCR (Figure S3B). Approximately 22% of SOX21 peaks were in promoters, and 40% were in distal intergenic regions (Figure 4A). *De novo* DNA motif analysis revealed that 66.16% of the peaks contained a C[AT]TTGT enriched sequence ($p < 1 \times 10^{-1032}$), consistent with the common consensus motif of SOX proteins (Figure 4B) (Harley

Figure 3. Generation of Additional SOX21 Knockout hESC Clones

(A) Schematic representation of the paired-KO strategy for *SOX21* KO in H9 hESCs. Four different gRNAs targeting different loci were designed (g19 is paired with g14, and g69 is paired with gb). Primers for amplification of genomic DNA are indicated by blue arrows. CDS, coding sequence.

(B) Representative genomic DNA PCR results show the biallelic deletion of *SOX21* in hESCs.

(C) Representative western blot analysis of SOX21 protein level in hESC-derived NPCs. WT, wild-type cells; S1, S2 and S17 are three *SOX21* KO clones.

(D) Sanger sequencing results show ligation sites (dashed line) in two gRNA pairs (g19 + g14 and g69 + gb).

(E) Results from real-time qPCR analysis of regional marker expression in WT and $SOX21^{-/-}$ NPCs at days 0, 4, and 8. Data are represented as fold changes relative to undifferentiated WT hESCs (day 0) and shown as means ± SEM. *p < 0.05, **p < 0.01, ***p < 0.001 (n = 3).

(F) Results from real-time qPCR analysis of regional marker expression in NT NPCs and *SOX21* knockdown NPCs at day 8. For inducible knockdown of *SOX21*, cells were treated with doxycycline (100 ng/mL) from days 0 to 8. Data are represented as fold changes relative to undifferentiated WT ESCs (day 0) and shown as means ± SEM. *p < 0.05, **p < 0.01, ***p < 0.001 (n = 3). NT, nontarget control.

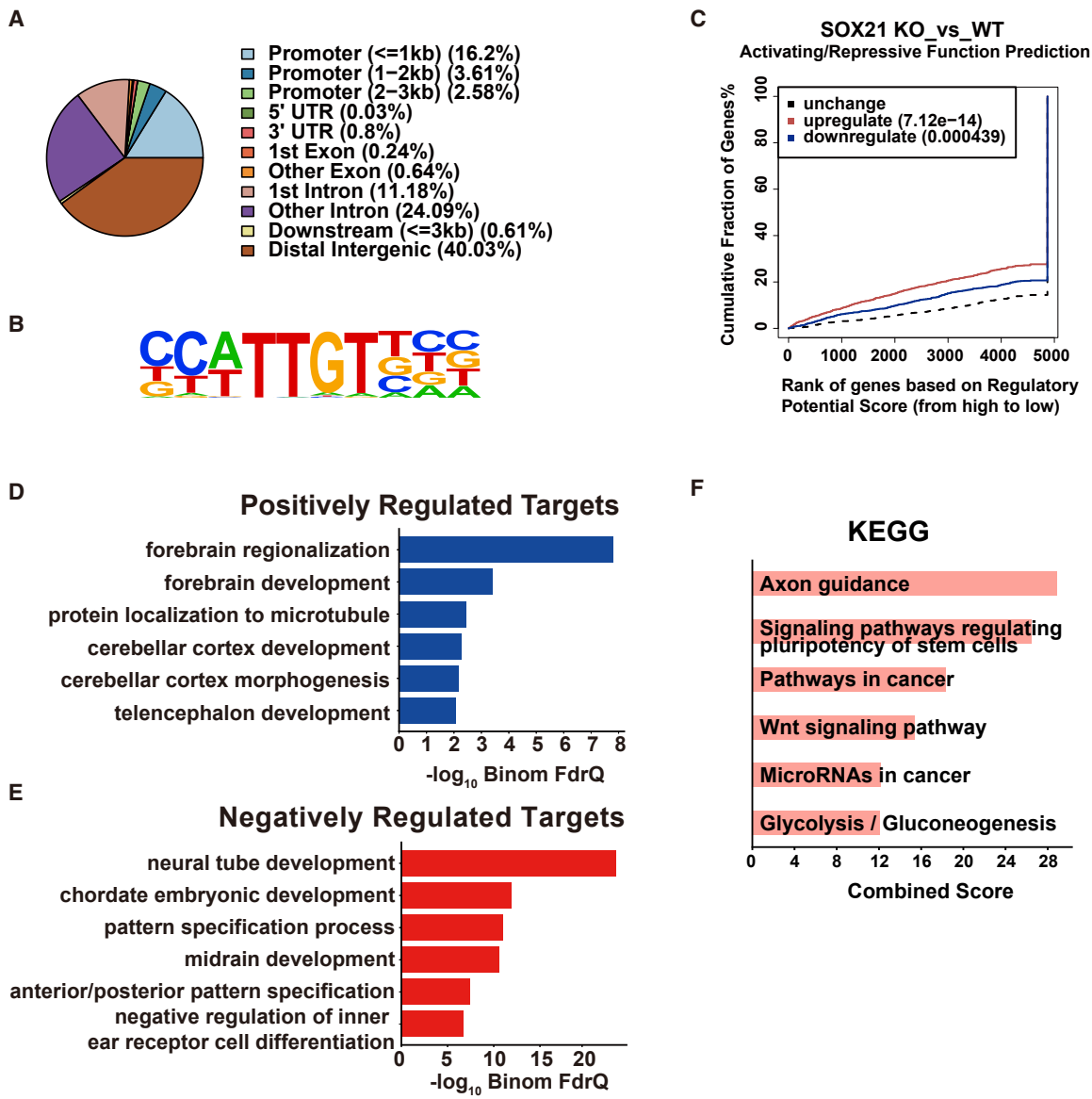


Figure 4. Genome-wide Analyses of SOX21 Binding Peaks

(A) A pie chart to show the distribution of SOX21-bound regions.

(B) The top-ranked DNA motif conserved in SOX21-bound regions revealed by the *de novo* motif discovery analysis.

(C) The BETA analysis of SOX21 functions by the integration of the ChIP-seq and RNA-seq datasets from *SOX21* KO versus WT NPCs. Cumulative proportions of genes either downregulated (positively, blue), upregulated (negatively, red), or unchanged (black) increased over different regulatory potential score cutoffs. p values that represent the significance of the up- or down-group distributions were compared with those of the unchanged group by the Kolmogorov-Smirnov test.

(D and E) GO analysis of SOX21 positively (D) and negatively (E) regulated targets enriched by GREAT software. BinomFdrQ, the binomial FDR q value.

(F) KEGG signaling pathway analysis of SOX21 negatively regulated targets. The top six terms are shown.

See also Figure S3.

et al., 1994). To identify putative target genes (PTG) of SOX21, we overlaid our transcriptomic dataset with that of SOX21 ChIP-seq and found that 358 and 490 PTGs were significantly downregulated ($p < 4.39 \times 10^{-4}$, Table S2) and upregulated ($p < 7.12 \times 10^{-14}$, Table S3), respectively (Figure 4C), in *SOX21* KO cells. GO analysis of biological processes showed that terms of forebrain regionalization and telencephalon development were enriched by

SOX21-positively regulated PTGs (Figure 4D), whereas PTGs negatively regulated by SOX21 were enriched for terms of pattern specification and midbrain development (Figure 4E). Moreover, disease ontology of mouse phenotype analysis showed that the positively regulated PTGs were associated with abnormal telencephalon development (Figure S3C), while the negatively regulated PTGs were associated with severe disease, including abnormal white matter morphology, abnormal neural tube, and exencephaly (Figure S3D). Moreover, our SOX21 function prediction results suggested that upregulated PTGs had much higher regulatory scores than the downregulated and unchanged ones in *SOX21* KO NPCs (Figure 4C), suggesting that SOX21 tends to have more repressive than activating activities. Therefore, we examined the signaling pathways related to upregulated PTGs and found that the Wnt signaling pathway was enriched (Figure 4F), implying that SOX21 might have a repressive role in Wnt signaling. Taken together, these results further confirm that SOX21 contributes to rostral forebrain regional specification during early neural differentiation of hESCs.

SOX21 Represses Wnt Signaling to Ensure Rostral Forebrain Identity

To validate the role of SOX21 in repressing Wnt signaling, we compared Wnt signaling activities between WT and *SOX21* KO NPCs using the β-catenin/TCF reporter (TOP/FOPFlash) luciferase assay. Consistent with our RNA-seq data, *SOX21* KO significantly elevated the reporter activity by approximately 30- to 50-fold (Figure 5A), supporting a repressive effect of SOX21 on the transcriptional activity of β-catenin. Moreover, *SOX21* KO increased the protein levels of OTX1/OTX2 and WNT8B as well as active β-catenin, phosphorylated LRP6, DVL2, and DVL3 (Figure 5B). Importantly, overexpression of *DKK1* (an antagonist of Wnt signaling) abrogated the increase in those protein levels caused by *SOX21* KO (Figure 5B). Similar results were obtained when cells were treated with IWP-2 (2 μM) (Figure 5C). Furthermore, real-time qPCR results showed that *DKK1* overexpression restored the expression of *FOXG1* and *SIX3* and suppressed the expression of *OTX1*, *LMX1A*, and Wnt signaling genes (*AXIN2*, *WNT8B*, *WLS*), while *SOX21* expression was not altered by *DKK1* overexpression (Figure S4A). In addition, three Wnt signaling inhibitors, IWP-2 (2 μM), IWR1-e (10 μM), and DKK1 (100 ng/mL), all efficiently corrected alterations in the expression of *FOXG1* and *SIX3* as well as *OTX1*, *OTX2*, and *WNT8B* (Figure 5D). As positive controls, Wnt signaling genes, such as *AXIN2*, *WLS*, and *SP5*, were all downregulated by the inhibitors (Figure 5D). These results clearly show that aberrant activation of Wnt signaling could account for most, if not all, *SOX21* KO-caused changes in gene expression. Therefore, we propose that SOX21 represses Wnt signaling to ensure the rostral forebrain identity of hESC-derived NPCs.

WNT8B Is a Major Downstream Target of SOX21

To dissect how SOX21 represses canonical Wnt signaling, we focused on *WNT8B*, which ranked first among significantly upregulated SOX21 PTGs in *SOX21* KO cells (Figure S5A). To functionally test the hypothesis that *WNT8B* is a major target of SOX21 in the control of rostral forebrain identity, we knocked out *WNT8B* by targeting the second exon of *WNT8B* using the CRISPR/Cas9 system in WT and *SOX21* KO cells (Figure S5B). Notably, double KO of *WNT8B* and *SOX21* (named S13W5, S13W13, and S23W1) largely blocked the aberrant upregulation of WNT8B and OTX1/OTX2, as well as pLRP6 and active β-catenin, at the protein level (Figure 6A). Real-time qPCR results revealed that alterations in the expression of regional markers and Wnt signaling genes in *SOX21* KO NPCs were corrected in $WNT8B^{-/-}SOX21^{-/-}$ NPCs to different extents (Figure S5C). Moreover, TOP/FOPFlash luciferase reporter assays showed that the activation of Wnt signaling caused by *SOX21* KO was abolished when *SOX21* and *WNT8B* were both absent in NPCs (Figure 6B). Taken together, these results suggest that SOX21 ensures the rostral forebrain identity primarily by repressing *WNT8B* expression.

To understand how SOX21 repressed *WNT8B* expression, we first examined whether SOX21 could be recruited to the *WNT8B* enhancer. ChIP-qPCR results validated the association of SOX21 with the enhancer of *WNT8B* in NPCs (Figure 6C), suggesting a potentially direct control of *WNT8B* expression by SOX21. Next, we searched for SOX21 protein partners by coimmunoprecipitation coupled with a mass spectrometric analysis using specific antibodies against SOX21 in both WT and *SOX21* KO NPCs. Among the protein partners identified (Table S4), we were most interested in β-catenin, as it usually forms complexes with TCF transcription factors to activate Wnt target genes (Nusse and Clevers, 2017). The specific interaction between SOX21 and β-catenin was verified by reciprocal coimmunoprecipitation assays (Figure 6D). Moreover, we analyzed published genome-wide DNA binding profiles of TCF4 in hESCs and anterior neural ectoderm cells (ANECs) (Tsankov et al., 2015) and found that TCF4 was associated with the enhancer of *WNT8B* in hESCs, but not in the ANECs (Figure 6E), suggesting that TCF4 might be excluded from the enhancer of *WNT8B* in NPCs. Furthermore, our ChIP-qPCR results showed that the recruitment of β-catenin and TCF4 to the *WNT8B* enhancer was greater in *SOX21* KO NPCs than in WT NPCs (Figure 6F), further supporting the notion that SOX21 has an inhibitory effect on TCF4/β-catenin recruitment to the *WNT8B* locus. In

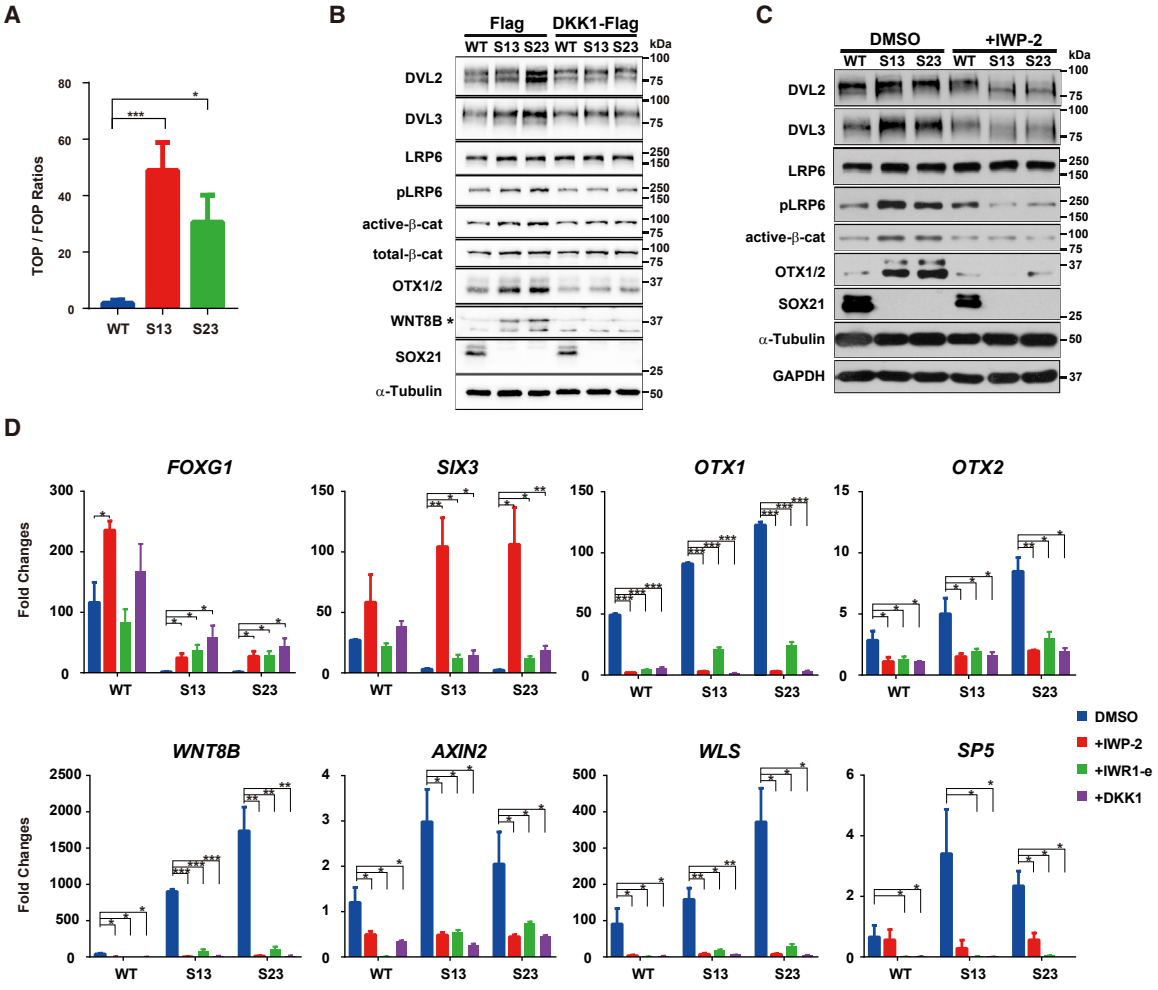


Figure 5. SOX21 Represses Wnt Signaling to Ensure Rostral Forebrain Specification

(A) Examination of Wnt signaling activities by TOP/FOPFlash reporter assays in WT and *SOX21* KO NPCs at day 5. Data are shown as means ± SEM. *p < 0.05, **p < 0.01, ***p < 0.001 (n = 3).

(B) Representative western blot analysis of Wnt signaling components and regional markers in WT and *SOX21* KO NPCs when *DKK1-Flag* expression was induced by doxycycline (100 ng/mL) from days 0 to 6. The Flag-tagged vector was used as a control. Antibodies against OTX1/OTX2 recognize both OTX1 and OTX2. A star indicates the specific band of WNT8B proteins.

(C) Representative western blot analysis of levels of various proteins in WT and *SOX21* KO NPCs in the absence or presence of the Wnt signaling inhibitor IWP-2 (2 μM) for 6 days.

(D) Real-time qPCR analysis of relative expression levels for regional markers as well as Wnt signaling-associated genes in WT and *SOX21* KO NPCs treated with or without Wnt signaling inhibitors, including DKK1 (100 ng/mL), IWP-2 (2 μM), and IWR1-e (10 μM), for 5 days. Data are represented as fold changes relative to undifferentiated WT hESCs (day 0) and shown as means ± SEM. *p < 0.05, **p < 0.01, ***p < 0.001 (n = 3). DMSO is the vehicle used to resolve the inhibitors.

See also Figure S4.

addition, *SOX21* KO enhanced the interaction intensity between TCF4 and β-catenin (Figure 6G). These results suggest that SOX21 may inhibit *WNT8B* expression via interference with the formation of TCF4/β-catenin complexes and their recruitment to the *WNT8B* locus.

Finally, to specifically show the relationship between TCF4/β-catenin and SOX21 in the regulation of *WNT8B* expression, we conducted reporter assays using a reporter vector containing either the WT *WNT8B* enhancer sequence or the enhancer sequence with mutated TCF4 binding motifs (Figure 6H). An increase in WT *WNT8B* reporter activity by CT (0.4 μM) treatment verified the activation of *WNT8B* expression by Wnt signaling (Figure 6H). Mutation of the TCF4 binding sequence in the *WNT8B* enhancer dramatically reduced reporter activity even under CT treatment (Figure 6H), suggesting that TCF4

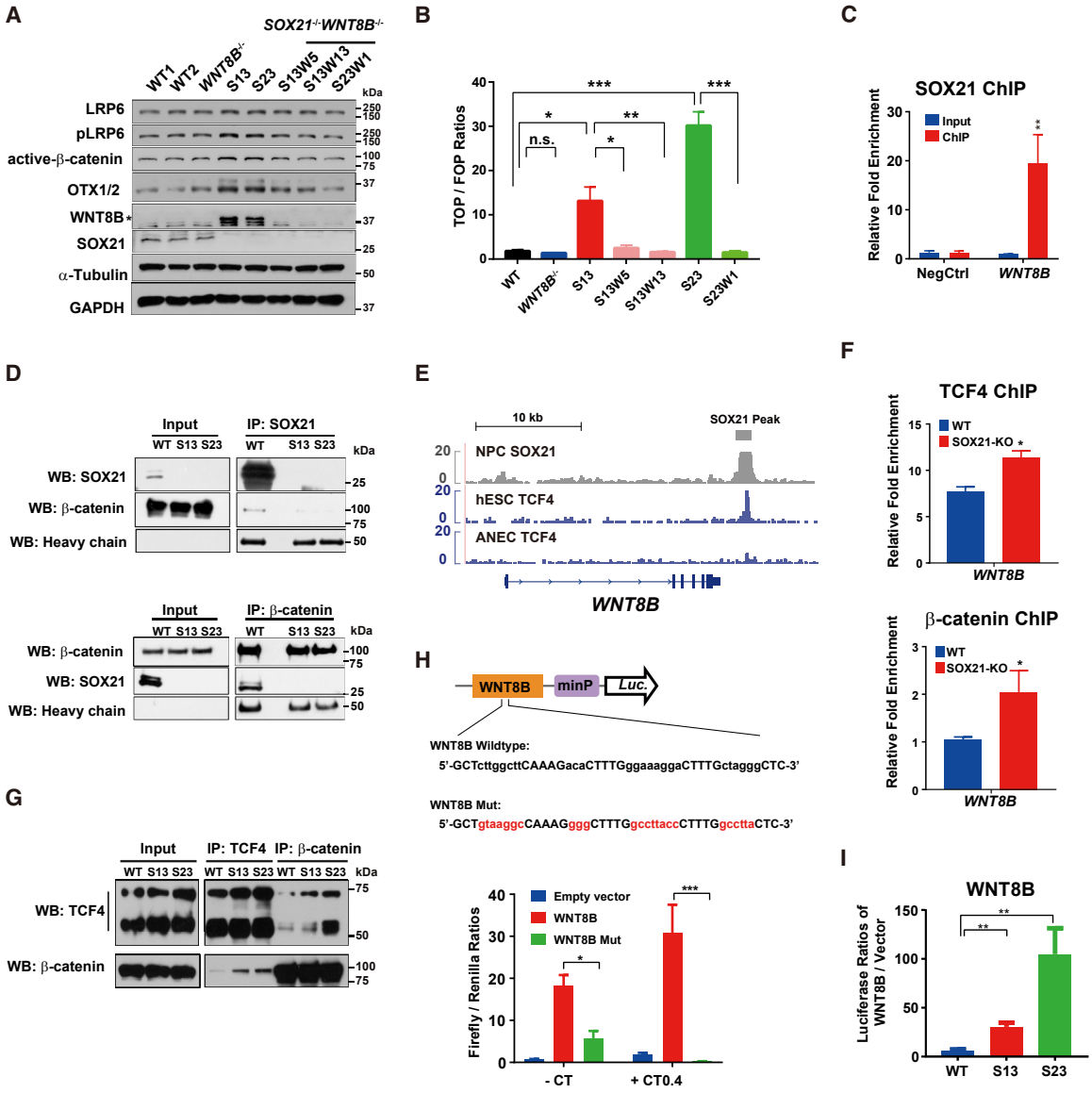


Figure 6. WNT8B Is a Major Downstream Target of SOX21

(A) Representative western blot analysis of regional markers and Wnt signaling proteins in WT, *SOX21*$^{-/-}$, *WNT8B*$^{-/-}$, and *SOX21*$^{-/-}$*WNT8B*$^{-/-}$ NPCs at day 6. WT, wild type; *SOX21* and *WNT8B* double KO clones: S13W5 (*SOX21*$^{-/-}$ no. 13, *WNT8B*$^{-/-}$ no. 5), S13W13 (*SOX21*$^{-/-}$ no. 13, *WNT8B*$^{-/-}$ no. 13), S23W1 (*SOX21*$^{-/-}$ no. 23, *WNT8B*$^{-/-}$ no. 1).

(B) Examination of Wnt signaling activity by TOP/FOPFlash reporter assays in WT, *SOX21*$^{-/-}$, *WNT8B*$^{-/-}$, and *SOX21*$^{-/-}$*WNT8B*$^{-/-}$ NPCs at day 5. Data are shown as means ± SEM. *p < 0.05, **p < 0.01, ***p < 0.001 (n = 4).

(C) ChIP-qPCR results of SOX21 at the *WNT8B* locus in NPCs at day 5. Data are shown as means ± SEM. *p < 0.05, **p < 0.01 (n = 3). NegCtrl stands for the negative control, a gene desert region (chr21:25509072 + 25509220, GRCh37/hg19).

(D) Representative coimmunoprecipitation results for interactions between endogenous SOX21 and β-catenin in WT and *SOX21* KO NPCs at day 7. IP: immunoprecipitation; WB: western blot.

(E) ChIP-seq signal profiles of SOX21 at the *WNT8B* locus in WT NPCs at day 5. ChIP-seq signal profiles of TCF4 in hESCs and anterior neural ectoderm cells (ANECs) were obtained from published data. The gray box above the signal tracks indicates the SOX21 binding site at the *WNT8B* locus.

(F) ChIP-qPCR analysis of TCF4 and β-catenin binding at the enhancer of *WNT8B* in WT and *SOX21* KO NPCs at day 5. Data are shown as means ± SEM. *p < 0.05, **p < 0.01 (n = 3).

(legend continued on next page)

recruitment plays a key role in *WNT8B* expression. Notably, *WNT8B* reporter activity was significantly higher in *SOX21* KO NPCs than in WT NPCs (Figure 6I), in line with the increased *WNT8B* mRNA levels in *SOX21* KO NPCs (Figures 2I, 3E, and S2F). Collectively, these results further support our conclusion that SOX21 interferes with the formation of TCF4/β-catenin complexes and their recruitment to the *WNT8B* enhancer to repress Wnt signaling in NPCs during early neural differentiation.

DISCUSSION

In this study, we applied a rostrocaudal patterning system from hPSCs to mimic early human neural regionalization. The expression patterns of regional markers indicate that the Wnt gradient could induce rostrocaudally specified NPCs from hESCs *in vitro*. Transcriptomic analysis divided regional identities along the R-C axis into six regions as follows: region 1 (IWP2 and CT0.0), region 2 (CT0.4), region 3 (CT0.8), region 4 (CT1.0 and CT2.0), region 5 (CT3.0 and CT4.0), and region 6 (CT4.0RA). Gene coexpression network analysis further revealed that seven region-specific modules might represent the core regulatory network along the R-C axis. The global gene regulatory networks in regionally defined hESC-derived NPCs gained in this study will help future efforts aimed at the efficient generation of region-specific NPCs.

We revealed that SOX21 is required for rostral forebrain specification through repressing Wnt signaling in NPCs. In *SOX21* KO NPCs, rostral forebrain genes were significantly downregulated, while caudal forebrain genes were upregulated. Importantly, our clustering analysis revealed that the *SOX21* KO NPCs were most closely associated with NPCs of region 2 (CT0.4 group) with diencephalic-mesencephalic features. Therefore, the regional identity of rostral forebrain NPCs was caudalized in the absence of *SOX21*. Forebrain patterning requires the inhibition of posteriorizing Wnt signaling. Consistently, *SOX21* KO substantially elevated Wnt signaling activity. Functionally, inhibition of Wnt signaling in *SOX21* KO NPCs restored the rostral forebrain identity, providing direct evidence that SOX21 suppresses Wnt signaling to ensure rostral forebrain identity and prevent the caudal fate. Nevertheless, Sox proteins function dependently on developmental stages and cellular contexts, and the question of how SOX21 activates genes in the rostral forebrain needs further investigation. In addition to the rostral forebrain, SOX21 proteins were also detected in NPCs treated with 1.0–4.0 μM CT, suggesting that SOX21 has other functions in neural differentiation of hESCs and might act through different mechanisms. *Sox21* KO mice are born normally, although they display cyclic alopecia and postnatal growth deficiency (Cheung et al., 2017; Kiso et al., 2009). This relatively minor phenotype could be explained by the functional redundancy among members of the Sox family, as they display overlaps of expression domains in the developing neural tissue (Uchikawa et al., 2011).

A few transcription factors have been known to repress Wnt signaling for forebrain specification in vertebrate development. For example, Six3 has been reported to inhibit the expression of *Wnt1* and *Wnt8b* directly in the mouse anterior neural ectoderm (Lagutin et al., 2003; Liu et al., 2010), and Hesx1 has been shown to antagonize Wnt/β-catenin signaling in the mouse and zebrafish forebrain (Andoniadou et al., 2011). However, how these factors inhibit the transcription of Wnt/β-catenin targets and how canonical Wnt signaling is repressed in human forebrain development remain unclear. Many Sox proteins are known to suppress Wnt signaling via different mechanisms (Kormish et al., 2010; Wegner, 2010). Here, we found that *SOX21* KO enhanced the recruitment of TCF4/β-catenin to the *WNT8B* enhancer. Thus, we propose that SOX21 might interfere with the binding of TCF4/β-catenin complexes to their target genes, repressing the activation of Wnt target genes. In addition, SOX21 formed protein complexes with β-catenin, and *SOX21* KO increased the interaction between TCF4 and β-catenin. Because *SOX21* KO also resulted in higher protein levels of TCF4, the increased interaction between TCF4 and β-catenin in *SOX21* KO NPCs could be a result of elevated TCF4 protein level or the absence of a competitor.

Among SOX21 PTGs, we focused on *WNT8B*. *WNT8B* KO largely abrogated the defect of rostral forebrain

(G) Representative coimmunoprecipitation results of interactions between endogenous TCF4 and β-catenin in WT and *S0X21* KO NPCs at day 7. IP: immunoprecipitation; WB: western blot.

(H) Results of luciferase reporter assays using an empty vector containing only a minimal promoter (minP) and vectors containing either the WT *WNT8B* enhancer sequence or the enhancer sequence with mutated TCF4 binding sites (WNT8B Mut), respectively, in response to Wnt signaling activation by CT99021 (0.4 μM) at day 6. Mutations at the TCF4 binding sites are indicated by red letters. Data are shown as means ± SEM. *p < 0.05, **p < 0.01, ***p < 0.001 (n = 4).

(I) Examination of *WNT8B* enhancer luciferase reporter activity in WT and *S0X21* KO NPCs at day 6. Data are represented as ratios of luciferase activity relative to that with an empty vector containing a minP only. Data are shown as means ± SEM. *p < 0.05, **p < 0.01, ***p < 0.001 (n = 4).

See also Figure S5.

specification caused by *SOX21* KO, supporting the hypothesis that *WNT8B* is a key target of SOX21. Taken together, these results suggest that SOX21 prevents TCF4/β-catenin complexes from binding to the *WNT8B* enhancer in hESC-derived rostral forebrain NPCs. In the absence of SOX21, more TCF4/β-catenin complexes bind to and activate *WNT8B*, leading to the caudalization of regional identity of rostral forebrain NPCs. The transcriptional repression of *WNT8B* by SOX21 is required to generate rostral forebrain NPCs. The discovery of the new function of SOX21 will facilitate our understanding of the transcriptional control of early human forebrain patterning.

EXPERIMENTAL PROCEDURES

Experimental methods are briefly summarized. See Supplemental Information for details.

Culture and Differentiation of hESCs

hESCs of the H9 (karyotype, XX) and SHhES8 (karyotype, XX) lines were maintained in mTeSR1 medium. For neural differentiation, hESC colonies were detached with dispase (Gibco) and resuspended in neural differentiation medium. The differentiation medium was prepared as follows: DMEM/F12:Neurobasal (1:1, Gibco), N2 (Gibco), B27 without vitamin A (Gibco), 2 mM glutamine, 0.1 mM nonessential amino acid, 5 μM SB431542 (Stemgent), and 50 nM LDN193189 (Stemgent). Rock inhibitor (Y-27632, 10 μM, Selleck) was present from days 0 to 2. On day 4, aggregates were plated onto Matrigel-coated six-well plates. From days 0 to 8, IWP-2 (2 μM, Millipore), or CT99021 (0–4.0 μM, STEMCELL), or RA (1 μM, Sigma) were included in the differentiation medium for region-specific neural differentiation.

Generation of KO hESC Lines

Homology arms were amplified from genomic DNA and inserted into the PGK-NEO-DTA targeting vector. The CRISPR/Cas9 plasmid was digested with Bbs I and ligated to the designed gRNAs. hESCs were digested into single cells and electroporated with appropriate combinations of CRISPR/Cas9 plasmids, with or without a donor plasmid. After a week of drug selection, colonies were collected for genotyping analyses.

RNA Extraction, cDNA Synthesis, and Real-Time qPCR

Total RNA was extracted using TRIzol reagent, and 2 μg total RNA was reverse transcribed into cDNA. Real-time qPCR was performed on the ABI ViiA7 Real-Time PCR system using SYBR Premix Ex Taq II (Takara). A list of primers used is provided in Table S5.

Western Blotting

Cells were lysed in the coimmunoprecipitation buffer, and protein concentrations were measured using the Pierce BCA Protein Assay Kit (Thermo Scientific). Twenty micrograms of total protein was separated by 10% SDS-PAGE, transferred to nitrocellulose blotting membranes, and blotted using primary and secondary antibodies. The antibodies used are listed in Table S6. All western blot analyses were conducted in at least three independent experiments, unless otherwise indicated.

Luciferase Reporter Assays

The WT or mutated *WNT8B* enhancer was cloned and inserted into the pGL4.22 vector (Promega). Then, 500 ng of the enhancer reporter plasmid or an empty vector and 10 ng of a pRL-TK internal control plasmid (Promega) were used for transfection. For Wnt signaling reporter assays, 500 ng of the 8xTOPFlash (Addgene, no. 12456) or 8xFOPFlash (Addgene, no. 12457) plasmid together with 50 ng of the pRL-TK internal control plasmid (Promega) were used for transfection. Luciferase activities were examined with the Dual-Glo Luciferase Assay System (Promega).

Flow Cytometric Analysis and Cell Proliferation Assays

Cells were fixed, permeabilized (Cytofix/Cytoperm Kit, BD Biosciences) and stained using antibodies against SOX2 and NANOG. For cell proliferation analysis, cells were incubated with 10 μM EdU (BD Biosciences) for 1 h and labeled using a Cell-Light EdU Apollo488 *In Vitro* Flow Cytometry Kit (RIBO). The data for stained samples were acquired on an Accuri C6 flow cytometer.

ChIP Assays

ChIP assays were performed as described previously with modifications (Zhu et al., 2017). See detailed methods in Supplemental Information. All primers used in ChIP-qPCR assays are listed in Table S7.

GO Analysis

GO analysis for DEGs and region-specific expressed genes was conducted using Enrichr (Kuleshov et al., 2016). Combined score (c) is the multiplication of the log of p value (p) with *Z* score (z); $c = \log(p) \times z$.

Statistical Analysis

The unpaired Student's t test was used for statistical tests unless stated otherwise, and data are presented as the mean ± SEM. Numbers of biological replicates relevant for individual experiments are stated in figure legends.

ACCESSION NUMBERS

The RNA-seq and ChIP-seq data presented in this manuscript are accessible through GEO series accession number GEO: GSE110506.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.stemcr.2019.10.013.

AUTHOR CONTRIBUTIONS

Z.F. and X.L. designed the project, performed major experiments, analyzed data, and wrote the manuscript. J.W. and F.T. helped with immunostaining. Y.Z. participated in the initiation of the

project. N.J. discussed the project and edited the manuscript. Y.J. directed the project and wrote the manuscript. Z.F. and X.L. contributed equally to this study.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China grants 31801224, 31730055, and 31871373; Ministry of Science and Technology of China grant 2016YFA0100100; Strategic Priority Research Program of the Chinese Academy of Sciences grant XDB19020100; and the Innovative Research Team of High-level Local Universities in Shanghai. The authors declare that they have no conflicts of interest with the contents of this article.

Received: January 31, 2019
Revised: October 23, 2019
Accepted: October 28, 2019
Published: November 21, 2019

REFERENCES

Andoniadou, C.L., and Martinez-Barbera, J.P. (2013). Developmental mechanisms directing early anterior forebrain specification in vertebrates. Cell Mol. Life Sci. *70*, 3739–3752.

Andoniadou, C.L., Signore, M., Young, R.M., Gaston-Massuet, C., Wilson, S.W., Fuchs, E., and Martinez-Barbera, J.P. (2011). HESX1- and TCF3-mediated repression of Wnt/beta-catenin targets is required for normal development of the anterior forebrain. Development *138*, 4931–4942.

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.

Cheung, L.Y., Okano, H., and Camper, S.A. (2017). Sox21 deletion in mice causes postnatal growth deficiency without physiological disruption of hypothalamic-pituitary endocrine axes. Mol. Cell Endocrinol. *439*, 213–223.

Ciani, L., and Salinas, P.C. (2005). WNTs in the vertebrate nervous system: from patterning to neuronal connectivity. Nat. Rev. Neurosci. *6*, 351–362.

Fossat, N., Jones, V., Khoo, P.L., Bogani, D., Hardy, A., Steiner, K., Mukhopadhyay, M., Westphal, H., Nolan, P.M., Arkell, R., et al. (2011). Stringent requirement of a proper level of canonical WNT signalling activity for head formation in mouse embryo. Development *138*, 667–676.

Fossat, N., Jones, V., Garcia-Garcia, M.J., and Tam, P.P.L. (2012). Modulation of WNT signaling activity is key to the formation of the embryonic head. Cell Cycle *11*, 26–32.

Fotaki, V., Larralde, O., Zeng, S., McLaughlin, D., Nichols, J., Price, D.J., Theil, T., and Mason, J.O. (2010). Loss of Wnt8b has no overt effect on hippocampus development but leads to altered Wnt gene expression levels in dorsomedial telencephalon. Dev. Dyn. *239*, 284–296.

Glover, J.C., Renaud, J.S., Lampe, X., and Rijli, F.M. (2006). Hindbrain development and retinoids. Adv. Dev. Biol. *16*, 145–180.

Harley, V.R., Lovell-Badge, R., and Goodfellow, P.N. (1994). Definition of a consensus DNA binding site for SRY. Nucleic Acids Res. *22*, 1500–1501.

Houart, C., Caneparo, L., Heisenberg, C., Barth, K., Take-Uchi, M., and Wilson, S. (2002). Establishment of the telencephalon during gastrulation by local antagonism of Wnt signaling. Neuron *35*, 255–265.

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.

Irioka, T., Watanabe, K., Mizusawa, H., Mizuseki, K., and Sasai, Y. (2005). Distinct effects of caudalizing factors on regional specification of embryonic stem cell-derived neural precursors. Brain Res. Dev. Brain Res. *154*, 63–70.

Itasaki, N. (2015). Vertebrate embryo: neural patterning (eLS. John Wiley) https://doi.org/10.1002/9780470015902.a0000737.pub3.

Kamachi, Y., and Kondoh, H. (2013). Sox proteins: regulators of cell fate specification and differentiation. Development *140*, 4129–4144.

Kiecker, C., and Lumsden, A. (2012). The role of organizers in patterning the nervous system. Annu. Rev. Neurosci. *35*, 347–367.

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

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.

Kiso, M., Tanaka, S., Saba, R., Matsuda, S., Shimizu, A., Ohyama, M., Okano, H.J., Shiroishi, T., Okano, H., and Saga, Y. (2009). The disruption of Sox21-mediated hair shaft cuticle differentiation causes cyclic alopecia in mice. Proc. Natl. Acad. Sci. U S A *106*, 9292–9297.

Kormish, J.D., Sinner, D., and Zorn, A.M. (2010). Interactions between SOX factors and Wnt/beta-catenin signaling in development and disease. Dev. Dyn. *239*, 56–68.

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.

Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q.N., Wang, Z.C., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. *44*, W90–W97.

Lagutin, O.V., Zhu, C.C., Kobayashi, D., Topczewski, J., Shimamura, K., Puelles, L., Russell, H.R., McKinnon, P.J., Solnica-Krezel, L., and Oliver, G. (2003). Six3 repression of Wnt signaling in the anterior neuroectoderm is essential for vertebrate forebrain development. Genes Dev. *17*, 368–379.

Lako, M., Lindsay, S., Bullen, P., Wilson, D.I., Robson, S.C., and Strachan, T. (1998). A novel mammalian wnt gene, WNT8B, shows brain-restricted expression in early development, with sharply delimited expression boundaries in the developing forebrain. Hum. Mol. Genet. *7*, 813–822.

Liu, W., Lagutin, O., Swindell, E., Jamrich, M., and Oliver, G. (2010). Neuroretina specification in mouse embryos requires Six3-mediated suppression of Wnt8b in the anterior neural plate. J. Clin. Invest. *120*, 3568–3577.

Liu, Z., Hui, Y., Shi, L., Chen, Z., Xu, X., Chi, L., Fan, B., Fang, Y., Liu, Y., Ma, L., et al. (2016). Efficient CRISPR/Cas9-mediated versatile, predictable, and donor-free gene knockout in human pluripotent stem cells. Stem Cell Reports *7*, 496–507.

Matsuda, S., Kuwako, K., Okano, H.J., Tsutsumi, S., Aburatani, H., Saga, Y., Matsuzaki, Y., Akaike, A., Sugimoto, H., and Okano, H. (2012). Sox21 promotes hippocampal adult neurogenesis via the transcriptional repression of the Hes5 gene. J. Neurosci. *32*, 12543–12557.

Nordstrom, U., Jessell, T.M., and Edlund, T. (2002). Progressive induction of caudal neural character by graded Wnt signaling. Nat. Neurosci. *5*, 525–532.

Nusse, R., and Clevers, H. (2017). Wnt/beta-catenin signaling, disease, and emerging therapeutic modalities. Cell *169*, 985–999.

Puelles, L., and Martinez, S. (2013). Chapter 8-patterning of the diencephalon. In Patterning and Cell Type Specification in the Developing CNS and PNS, J.L.R. Rubenstein and P. Rakic, eds. (Academic Press), pp. 151–172.

Rubenstein, J.L.R., and Rakic, P. (2013). Comprehensive developmental neuroscience: patterning and cell type specification in the developing CNS and PNS (Elsevier/Academic Press).

Sandberg, M., Kallstrom, M., and Muhr, J. (2005). Sox21 promotes the progression of vertebrate neurogenesis. Nat. Neurosci. *8*, 995–1001.

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.

Tuazon, F.B., and Mullins, M.C. (2015). Temporally coordinated signals progressively pattern the anteroposterior and dorsoventral body axes. Semin. Cell Dev. Biol. *42*, 118–133.

Twyman, R.M. (2009). Wnt pathway and neural patterning A2-squire. In Encyclopedia of Neuroscience, R. Larry, ed. (Academic Press), pp. 497–502.

Uchikawa, M., Kamachi, Y., and Kondoh, H. (1999). Two distinct subgroups of Group B Sox genes for transcriptional activators and repressors: their expression during embryonic organogenesis of the chicken. Mech. Dev. *84*, 103–120.

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.

van de Leemput, J., Boles, N.C., Kiehl, T.R., Corneo, B., Lederman, P., Menon, V., Lee, C., Martinez, R.A., Levi, B.P., Thompson, C.L., et al. (2014). CORTECON: a temporal transcriptome analysis of in vitro human cerebral cortex development from human embryonic stem cells. Neuron *83*, 51–68.

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

Whittington, N., Cunningham, D., Le, T.K., De Maria, D., and Silva, E.M. (2015). Sox21 regulates the progression of neuronal differentiation in a dose-dependent manner. Dev. Biol. *397*, 237–247.

Wurst, W., and Bally-Cuif, L. (2001). Neural plate patterning: upstream and downstream of the isthmic organizer. Nat. Rev. Neurosci. *2*, 99–108.

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.e7.

Stem Cell Reports, Volume *13*

Supplemental Information

SOX21 Ensures Rostral Forebrain Identity by Suppression of WNT8B during Neural Regionalization of Human Embryonic Stem Cells

Zhuoqing Fang, Xinyuan Liu, Jing Wen, Fan Tang, Yang Zhou, Naihe Jing, and Ying Jin

Inventory of Supplemental Information

1. Supplemental Figures and Legends

1) Figure S1. Related to Figure 1. Region- specific transcriptome analysis of hESC-derived neural progenitor cells.

2) Figure S2. Related to Figure 2. SOX21 participates in forebrain regionalization.

3) Figure S3. Related to Figure 4. Genome-wide analysis of SOX21 binding peaks.

4) Figure S4. Relative to Figure 5. Restoration of the rostral forebrain fate in SOX21 knockout cells by DKK1 overexpression.

5) Figure S5. Related to Figure 6. WNT8B knockout largely rescues defects caused by SOX21 deletion.

2. Supplemental Tables

1) Table S1. Related to Figure S1. Modules classified by WGCNA.

2) Table S2. Related to Figure 4. The list of SOX21 positively regulated putative target genes.

3) Table S3. Related to Figure 4. The list of SOX21 negatively regulated putative target genes.

4) Table S4. Related to Figure 6. The list of potential SOX21 interacting proteins identified by mass spectrometric analysis.

5) Table S5. The list of real-time qPCR primers used in this study.

6) Table S6. The list of antibodies used in this study.

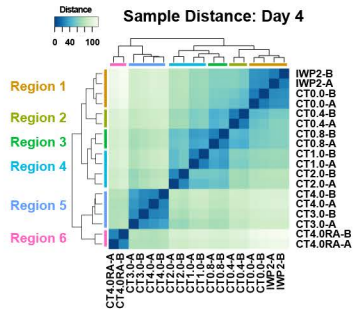
7) Table S7. The list of ChIP-qPCR primers used in this study.

3. Supplemental Experimental Procedures

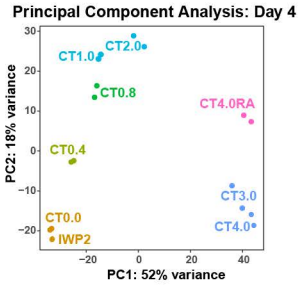
4. Supplemental References

Figure S1

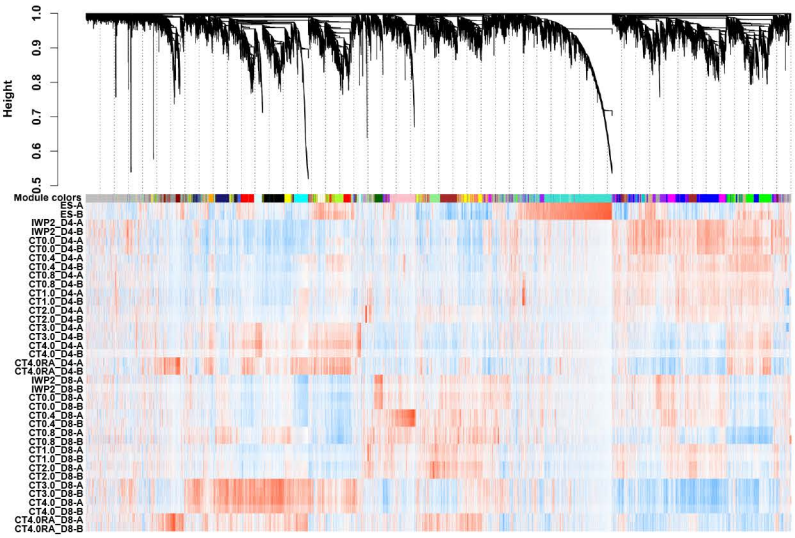
A



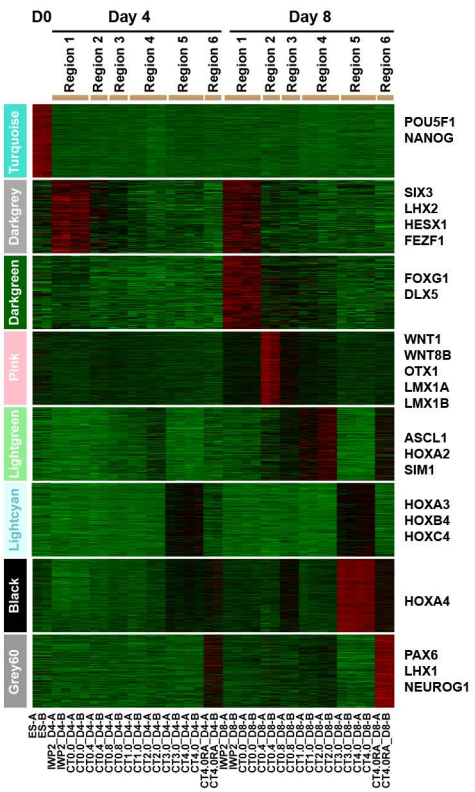
B



C



D



E

	Biological Process	Combined Score
Turquoise	response to interferon-gamma (GO:0034341)	34.3
	cellular response to cytokine stimulus (GO:0071345)	28.6
	cytokine-mediated signaling pathway (GO:0019221)	28.5
	cofactor metabolic process (GO:0051186)	27.3
	cellular response to interferon-gamma (GO:0071346)	26.2
Darkgrey	axis specification (GO:0009798)	20.5
	embryonic axis specification (GO:0000578)	19.5
	pattern specification process (GO:0007389)	19.4
	mitotic sister chromatid segregation (GO:0000070)	19.1
	anterior/posterior axis specification (GO:0009948)	19.1
Darkgreen	extracellular matrix organization (GO:0030198)	52.5
	extracellular structure organization (GO:0043062)	52.3
	glycosaminoglycan metabolic process (GO:0030203)	34.7
	aminoglycan metabolic process (GO:0006022)	33.5
	keratan sulfate catabolic process (GO:0042340)	33.0
Pink	axon guidance (GO:0007411)	49.9
	neuron projection guidance (GO:0097485)	49.8
	regulation of osteoblast differentiation (GO:0045667)	48.4
	midbrain development (GO:0030901)	47.6
	extracellular matrix organization (GO:0030198)	47.6
Lightgreen	neuron differentiation (GO:0030182)	42.6
	negative regulation of nervous system development (GO:0051961)	36.1
	regulation of neuron differentiation (GO:0045664)	34.0
	central nervous system neuron differentiation (GO:0021953)	33.7
	neuroblast proliferation (GO:0007405)	32.8
Lightcyan	anterior/posterior pattern specification (GO:0009952)	88.4
	regionalization (GO:0003002)	81.0
	pattern specification process (GO:0007389)	60.0
	organ morphogenesis (GO:0009887)	49.4
	embryonic morphogenesis (GO:0048598)	40.9
Black	regulation of translation (GO:0006417)	27.3
	posttranscriptional regulation of gene expression (GO:0010608)	27.3
	alcohol biosynthetic process (GO:0046165)	26.9
	positive regulation of translation (GO:0045727)	22.9
	cholesterol biosynthetic process (GO:0006695)	22.4
Grey60	pattern specification process (GO:0007389)	28.3
	regulation of neuron differentiation (GO:0045664)	26.6
	embryonic morphogenesis (GO:0048598)	24.3
	cellular response to growth hormone stimulus (GO:0071378)	19.4
	nervous system development (GO:0007399)	19.3

Figure S2

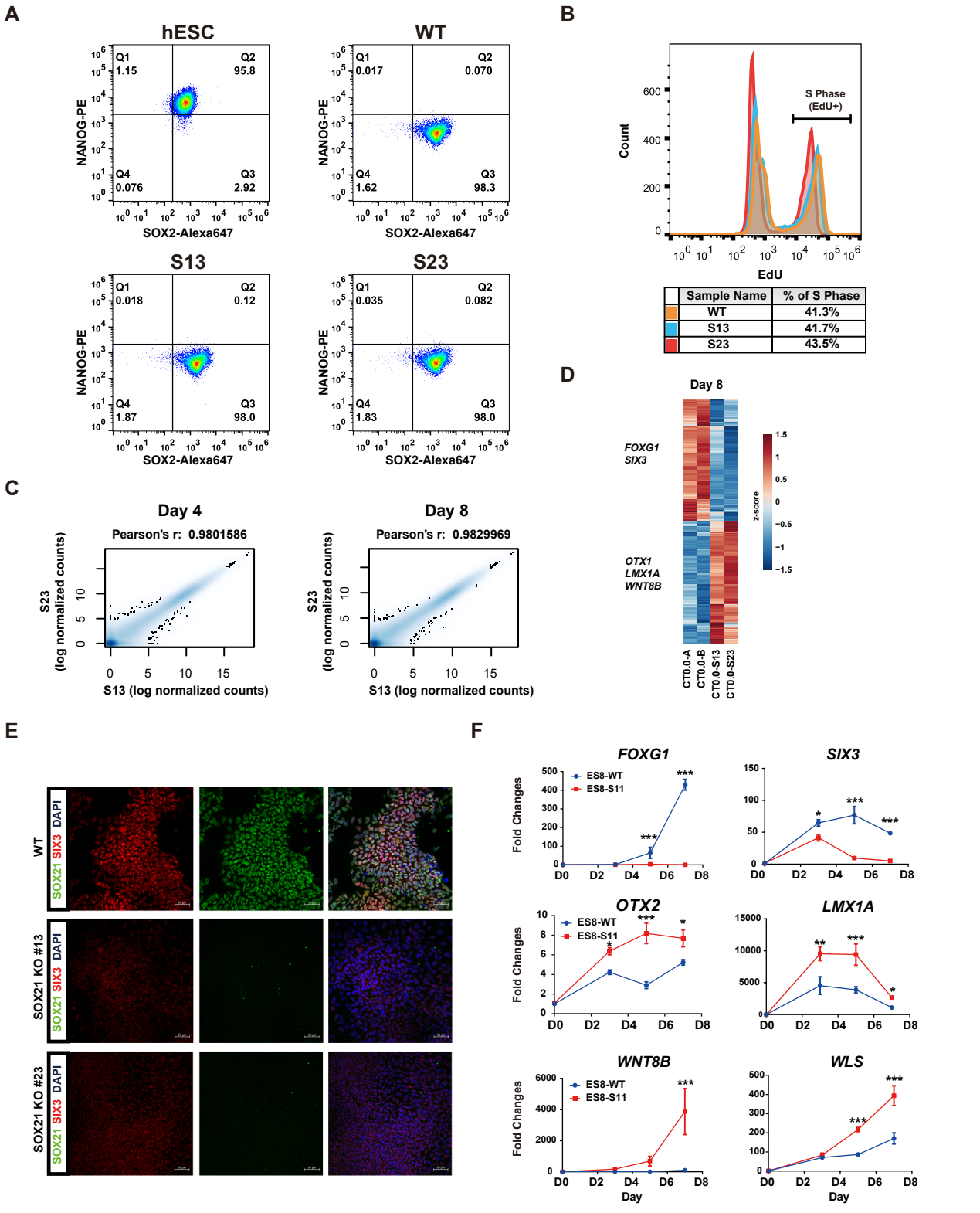

A
hESC
Q1 1.15
Q2 95.8
Q4 0.076
Q3 2.92
WT
Q1 0.017
Q2 0.070
Q4 1.62
Q3 98.3
S13
Q1 0.018
Q2 0.12
Q4 1.87
Q3 98.0
S23
Q1 0.035
Q2 0.082
Q4 1.83
Q3 98.0
NANOG-PE
SOX2-Alexa647
B
S Phase (EdU+)
Count
EdU
Sample Name | % of S Phase
WT | 41.3%
S13 | 41.7%
S23 | 43.5%
C
Day 4
Pearson's r: 0.9801586
Day 8
Pearson's r: 0.9829969
S23 (log normalized counts)
S13 (log normalized counts)
D
Day 8
FOXG1
SIX3
OTX1
LMX1A
WNT8B
z-score
CT0.0-A
CT0.0-B
CT0.0-S13
CT0.0-S23
E
WT
SOX21 KO #13
SOX21 KO #23
SOX21 SIX3 DAPI
F
FOXG1
SIX3
OTX2
LMX1A
WNT8B
WLS
ES8-WT
ES8-S11
Fold Changes
Day
D0 D2 D4 D6 D8

Figure S3

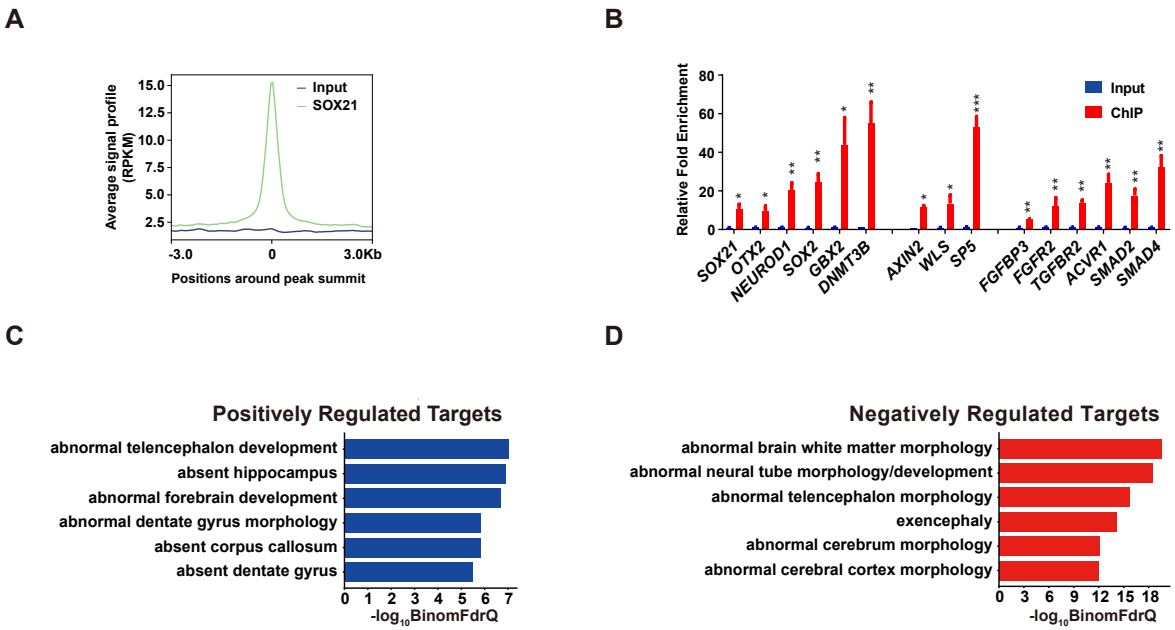

A
Average signal profile (RPKM)
15.0
12.5
10.0
7.5
5.0
2.5
Input
SOX21
-3.0
0
3.0Kb
Positions around peak summit
B
Relative Fold Enrichment
80
60
40
20
0
Input
ChIP
SOX21
OTX2
NEUROD1
SOX2
GBX2
DNMT3B
AXIN2
WLS
SP5
FGFBP3
FGFR2
TGFBR2
ACVR1
SMAD2
SMAD4
C
Positively Regulated Targets
abnormal telencephalon development
absent hippocampus
abnormal forebrain development
abnormal dentate gyrus morphology
absent corpus callosum
absent dentate gyrus
0 1 2 3 4 5 6 7
-log10BinomFdrQ
D
Negatively Regulated Targets
abnormal brain white matter morphology
abnormal neural tube morphology/development
abnormal telencephalon morphology
exencephaly
abnormal cerebrum morphology
abnormal cerebral cortex morphology
0 3 6 9 12 15 18
-log10BinomFdrQ

Figure S4

A

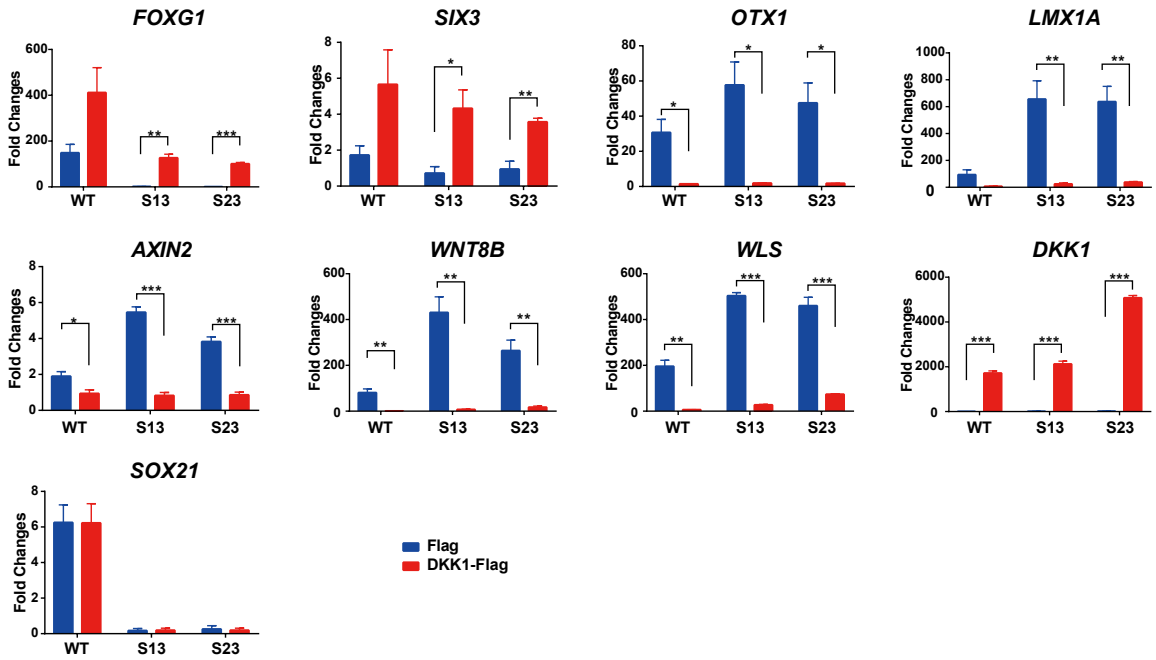

FOXG1
SIX3
OTX1
LMX1A
AXIN2
WNT8B
WLS
DKK1
SOX21
Fold Changes
WT
S13
S23
Flag
DKK1-Flag

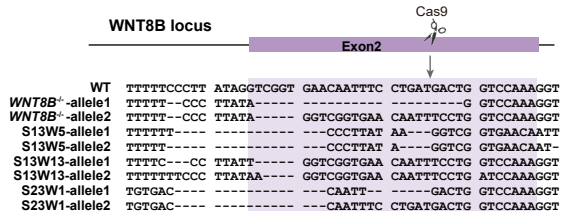
Figure S5

A

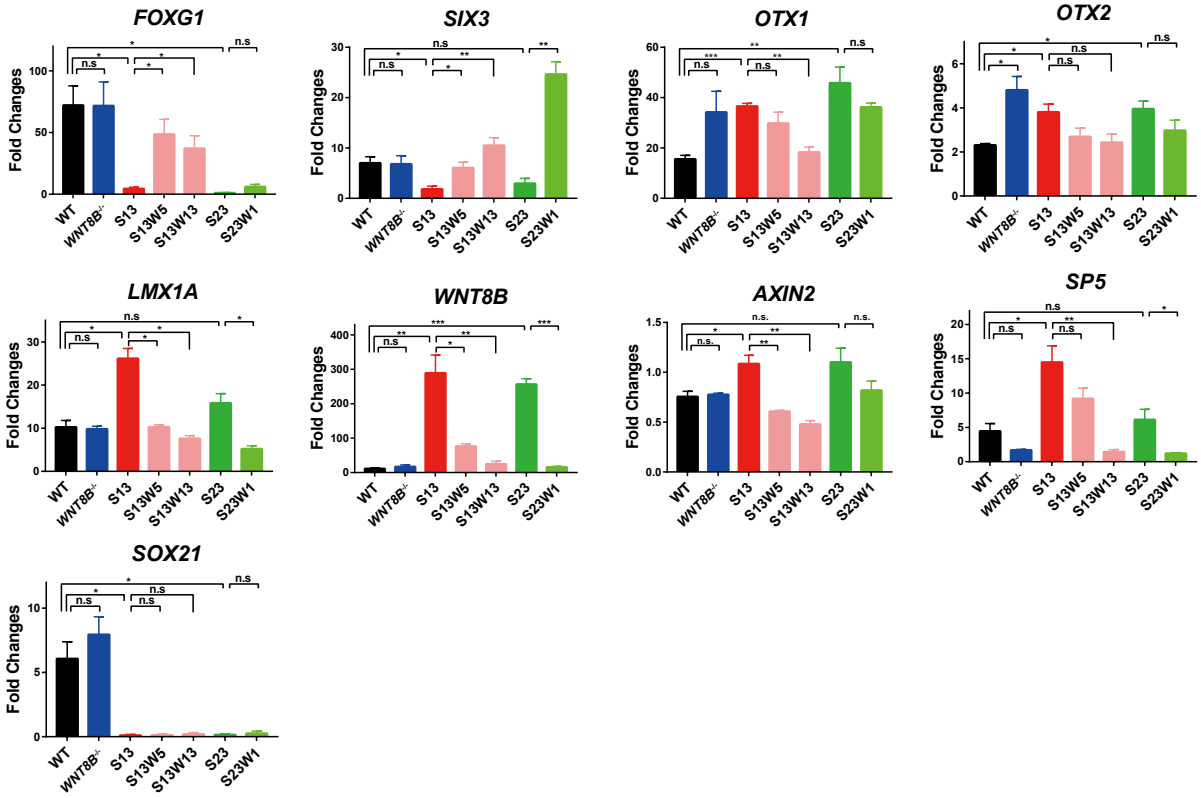
Top 10 Upregulated Putative Target Genes at Day 8

Gene Name	log2FoldChange	FDR
WNT8B	8.35	6.88E-60
HES3	5.68	3.05E-19
FOXB1	5.00	9.96E-21
WLS	4.87	2.46E-61
MTUS1	4.24	4.41E-08
NR5A1	4.09	6.64E-04
RSPO2	4.01	1.64E-07
DMRT3	3.89	1.02E-03
FGFR2	3.79	1.33E-81
TMEM125	3.50	2.21E-01

B



C



Supplemental Information

Supplemental Figure legends

Figure S1. Related to Figure 1. Region- specific transcriptome analysis of hESC-derived neural progenitor cells (NPCs).

(A) Unsupervised hierarchical clustering of 18 samples at day 4.

(B) Principal component analysis of 18 samples at day 4.

(C) The gene dendrogram and co-expression modules were identified by the weighted gene co-expression network analysis (WGCNA). The first color band underneath the dendrogram is modules determined by dynamic tree cut. Remaining color bands reveal highly correlated genes for the particular sample.

(D) The heatmap shows the relative gene expression in 7 representative region- specific modules and 1 module of undifferentiated WT hESCs. Representative genes belonging to each module are shown on the right side.

(E) Top 5 gene ontology terms for each module are shown.

See also Table S1.

Figure S2. Related to Figure 2. SOX21 participates in forebrain regionalization

(A) The representative result from flow cytometric analysis of $NANOG^{-}SOX2^{+}$ cells in WT and *SOX21* KO cells at day 7 of neural differentiation. $NANOG^{+}SOX2^{+}$ undifferentiated WT hESCs were used as a control.

(B) The representative result from cell proliferation analysis by EdU labeling assays in WT and *SOX21* KO cells at day 7. The percentage of EdU positive cells is indicated by the horizontal bar and summarized in the table at the bottom.

(C) The correlation analysis of RNA-seq data between two *SOX21* KO clones at day 4 and 8, respectively.

(D) A heatmap of 730 differentially expressed genes (DEGs) between WT (CT0.0-A and CT0.0-B) and *SOX21* KO cells (S13 and S23) (fold changes > 2, FDR < 0.05) at day 8 (CT = 0.0 μM). Some known markers involved in neural development are listed on the left. Each row of the heatmap represents the Z-score transformed and normalized read count values of DEGs across all samples.

(E) Representative result of immunofluorescence staining using antibodies against SIX3 and SOX21 at day 7. Scale bars: 50 μm.

(F) Results from real-time qPCR analysis of DEGs in another *SOX21* KO hESC line (SHhES8). Data are represented as fold changes relative to undifferentiated WT hESCs (day 0) and shown as mean ± SEM. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ (n=3).

Figure S3. Related to Figure 4. Genome-wide analysis of SOX21 binding peaks.

(A) The average signal profile of SOX21 relative to the summit of SOX21 ChIP-seq peaks.

(B) Results of ChIP-qPCR assays to validate the recruitment of SOX21 to the selected loci. Data are shown as mean ± SEM. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ (n=3).

(C-D) Disease ontology analysis of positively and negatively regulated targets by SOX21, respectively. BinomFdrQ: the Binomial FDR q-value.

Figure S4. Relative to Figure 5. Restoration of the rostral forebrain fate in SOX21 knockout cells by DKK1 overexpression

(A) The gene expression of regional markers and Wnt signaling genes was quantified by real-time qPCR at day 5 of WT and *SOX21* KO NPCs with *DKK1* or Flag control overexpression. Data are represented as fold changes relative to undifferentiated WT hESCs (day 0) and shown as mean ± SEM. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ (n=3).

Figure S5. Related to Figure 6. WNT8B knockout largely rescues defects caused by SOX21 deletion

(A) The top 10 upregulated targets in *SOX21* KO cells at day 8. SOX21 targets were identified by integrated analysis of ChIP-seq and RNA-seq data and ranked by the log2FoldChange.

(B) The strategy to knockout *WNT8B* using the wild type CRISPR/Cas9. A guide RNA was targeted to the *WNT8B* exon 2, which is showed in the purple. Sanger sequencing for 2 mutated alleles (with suffix -1 and -2, respectively) of *WNT8B* gene in WT and *SOX21* KO hESCs, as well as for a normal allele in WT hESCs. WT, wild-type; double KO clones: S13W5 (*SOX21*$^{-/-}$#13 *WNT8B*$^{-/-}$#5), S13W13 (*SOX21*$^{-/-}$#13 *WNT8B*$^{-/-}$#13), S23W1 (*SOX21*$^{-/-}$#23 *WNT8B*$^{-/-}$#1).

(C) Results from real-time qPCR analysis of regional markers and Wnt signaling genes in WT, *SOX21*$^{-/-}$, *WNT8B*$^{-/-}$, *SOX21*$^{-/-}$*WNT8B*$^{-/-}$ NPCs at day 5. Data are represented as fold changes relative to undifferentiated WT hESCs (day 0) and shown as mean ± SEM. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ (n=3); n.s, not significant.

Table S1. Related to Figure S1. Modules classified by WGCNA

Table S2. Related to Figure 4. The list of SOX21 positively regulated putative target genes

Table S3. Related to Figure 4. The list of SOX21 negatively regulated putative target genes

Table S4. Related to Figure 6. The list of potential SOX21 interacting proteins identified by mass spectrometric analysis

Table S5. The list of real-time qPCR primers used in this study

Human genes		
Gene	**Forward 5' to 3'**	**Reverse 5' to 3'**
ACTB	TGAAGTGTGACGTGGACATC	GGAGGAGCAATGATCTTGAT
AXIN2	ACTTCTGGTTTGCCTGCAATGGA	GTGGCAGGCTTCAGCTGCTT
FOXG1	TGCAATGTGGGGAGAATACA	CAGGTTTGAATGAAATGGCA
LHX2	GAAGGGGCGGCCGAGGAAAC	GCTGGTCACGGTCCAGGTGC
LMX1A	GCTCAGATCCCTTCCGACAG	GCCAGTGACCCCTCAAAGAA
OTX1	CACTAACTGGCGTGTTTCTGC	GGCGTGGAGCAAAATCG
DKK1	ATAGCACCTTGGATGGGTATTCC	CTGATGACCGGAGACAAACAG
OTX2	ACAAGTGGCCAATTCACTCC	GAGGTGGACAAGGGATCTGA
IRX3	GCGGAACAGATCGCTGTAGT	GAGAGCCGATAAGACCAGGG
SIX3	ACCGGCCTCACTCCCACACA	CGCTCGGTCCAATGGCCTGG
SOX21	CCGAGTGGAAACTGCTCACA	TTCTTGAGGAGCGTCTTGGG
SP5	CTTCGGGTGTCCATGCCTC	GTGCGGTCCTGGAGAAAGG
WLS	TGCCATGAAGACCTTCCTTACG	CCAGTCAAACCCGATGGAAAAC
WNT8B	TTGTCGATGCCCTGGAAACA	TTGAGTGCTGCGTGGTACTT

Table S6. The list of antibodies used in this study

Antibody	Company
SOX21	R&D, AF3538
GAPDH	Bioworld, 20301707-1
FOXG1	Abcam, ab18259
OTX1/2	Abcam, ab25985
SIX3	Abcam, ab221750
LMX1A	Abcam, ab139726
p-LRP6	CST, #2568P
LRP6	CST, #3395P
DVL2	CST, #3216
DVL3	CST, #3218
TCF4	CST, #2569
Active-β-catenin	CST, #8814
β-catenin	BD Biosciences, 610154
α-Tubulin	Sigma, T9026
GBX2	Sigma, HPA067809
WNT8B	Lifespan, LS-C117181-50
SOX2-Alexa647	BD Biosciences, 51-9006407
NANOG-PE	BD Biosciences, 560483
Isotype-Alexa647	BD Biosciences, 557783
Isotype-PE	BD Biosciences, 555749

Table S7. The list of ChIP-qPCR primers used in this study

Gene	Forward 5' to 3'	Reverse 5' to 3'
NegCtrl	GGGGGATCAGATGACAGTAAA	AATGCCAGCATGGGAAATA
ACVR1	TTTGAACGCTGCTTGCATGG	AGGCTCTTGGTCACATCTGC
AXIN2	TGACCAAGCAGACGACGAAG	TTGCGTTTGGGCAAGGTACT
DNMT3B	TTGACTTGGTGATTGGCGGA	CCAGGAACCGTGAGATGTCC
FGFBP3	AATAAGGCTTTGGCGCCTCT	CAGCTTCGGAGGAGTCATGC
FGFR2	CCTGCGGAGACAGGTAACAG	GGTGTCTGCCGTTGAAGAGA
GBX2	GCGGTCAGGCTTAATAGGATCA	CCGCCTTCGTCCTAAAGGG
NEUROD1	CGCTTTGCAAGGGCTTATCC	AGGCGACTGGTAGGAGTAGG
OTX2	TGCATTCCTATCCCTACATTTGC	CTGCAAATGGCCCCAATCAA
SMAD2	AGCCAATGGCAAGTGAAGGA	CACCAAGGATGCAGCCACTA
SMAD4	GTGACACCACCCTCCTAAGTG	AGAGCTCTGAATGATCCAGCC
SOX2	CGAGGCTTTGTTTGACTCCG	ATAGGTAGGCGCTCAATGCG
SOX21	TTCTTGGCCGGTAAACCTATTCA	TGACAAAACGGTGAAAGGGGAA
SP5	GGCCCCCTTTGATCAGGAAA	AGTTTGCCGCTACCCAATCA
TGFBR2	GCTCTGGTGCTCTGGGAAAT	CCAGCACTCAGTCAACGTCT
WLS	ACTCCTGGGGAATAGGGACCA	TCCACAAAAGATTATGGGGCACCT
WNT8B	CCGGCCTTTCTTCCCTTCAA	CGGGCTAGATGTGTGTGTGT

Supplemental Experimental Procedures

Culture and differentiation of human ESCs

Human ESCs of H9 (Karyotype, XX) and SHhES8 (Karyotype, XX) lines were cultured on Matrigel (BD, #354277) coated dishes with a daily change of the mTeSR1 medium (STEMCELL). For neural differentiation, human ESC colonies were detached with dispase (1 mg/ml, Gibco) and suspended for aggregate formation in the DMEM/F12:Neurobasal (1:1, Gibco), N2 supplement (1:100, Gibco), and B27 supplement without vitamin A (1:50, Gibco). Rock inhibitor (Y-27632, 10 μM, Selleck) was present from day 0 to day 2. On day 4, aggregates were plated onto 6-well plates coated with Matrigel, and grew in the same medium. From day 0 to day 8, SB431542 (5 μM, Stemgent) and LDN193189 (50 nM, Stemgent) were present in the medium, with or without IWP-2 (2 μM, Millipore), CHIR99021 (0.0 - 4.0 μM, STEMCELL), and RA (1μM, Sigma, #R2625) for region- specific neural differentiation.

Targeting plasmid construction

For donor-based *SOX21* knockout (KO), the single exon of *SOX21* was replaced with a PGK-neo cassette. A 0.8 kb short homology arm upstream the translation initiation site and a 1.5 kb long homology arm downstream the stop codon of *SOX21* were generated by PCR from H9 ESC genomic DNA using primers that had restriction sites on the ends of each amplicon and cloned into the PGK-NEO-DTA targeting vector. SalI/Hind III were used for the short arm and XbaI/NotI for the long arm: SOX21-short_fw 5'-acgcgtcgacagccgggagaacttcctct-3', SOX21-short_rev 5'-cccaagcttggcccggaggaaatcaatgt-3'; SOX21-long_fw 5'-gctctagaataggtgccaggtagaggca-3', SOX21-long_rev 5'-

ataagaatgcggccgcaggcacgtaagggcaattca-3'. Two paired gRNAs flanking the coding sequencing of *SOX21* were designed using the online tool from Zhang-lab (http://tools.genome-engineering.org). The gRNAs were cloned into px335 (Addgene plasmid, #48873) using the following gRNA pairs: for Cas9n-1, using g1A 5'-agccggtggaccacgtcaag-3', and g1B 5'-gacatgctctcgccctgccg-3'; for Cas9n-2, using g2A 5'-atgtataggtacgagcgctg-3', and g2B 5'-gcacaccggtcctcgcgagg-3'. For donor-free gene targeting, we designed 2 different sets of gRNAs, which were cloned into px459 (Addgene plasmid, #48139). The sequence of gRNAs for human *SOX21*: g19 5'-cttgacgtggtccaccggct-3', g69 5'- gcgggctcagcggcgcaaga-3', g14 5' -gctgccgcgctatgaccccg-3', gb 5'- atgtataggtacgagcgctg-3'. For *SOX21* knockdown experiments, 2 oligonucleotide sequences (3# 5'-tcgcaattttatcgaagatta-3'; 5# 5'-tccctgtttgtactatttgaa-3') targeting *SOX21* were designed to construct the *SOX21* shRNA inserts, respectively, according to a published protocol (Chang et al., 2013), and the AAVS1-TRE3G-EGFP plasmid (Qian et al., 2014) was used as a vector to generate *SOX21* shRNA plasmids. We also designed a nontarget (NT) control oligonucleotide sequence (5'-aggaattataatgcttatcta-3') to make the NT control shRNA plasmid. The sequence of gRNA for *WNT8B* KO is 5'-ttacctttggaccagtcatc-3'.

Generation of genetically modified hESC lines

Human ESCs were treated with Rho kinase inhibitor (Y27632, 10 μM) for 1 hr before electroporation. Cells were digested into single cells by Accutase (Innovative Cell Techonologies) for 7 min, and 5 x 10^6 cells were electroporated with plasmids in 200 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) (Chen et al., 2015), using the Gene Pulser Xcell System (Bio-Rad) at 250 V, 500 μF in 0.4 cm cuvettes (Bio-Rad). For *SOX21* KO experiments, 3.125 μg of each gRNA plasmid and 12.5 μg donor plasmid was used for electroporation. For *WNT8B* KO experiments, 25 μg of the gRNA plasmid was used for electroporation. For *DKK1* overexpression experiments, the *DKK1* cDNA with the Flag sequence at its C-terminus was cloned into the AAVS1-TRE3G-EGFP (Qian et al., 2014) to replace the EGFP sequence and generate the AAVS1-TRE3G-DKK1-Flag plasmid. Fifteen μg of the AAVS1-TRE3G-DKK1-Flag plasmids or the *SOX21* shRNA plasmids, plus 6.25 μg of AAVS1 gRNA plasmids for nickase (gA 5'-gtccctagtggccccactgt-3', gB 5'-gacagaaaagccccatcctt-3'), were used for electroporation. Cells were subsequently plated onto Matrigel coated 6-well plates in the mTeSR1 containing 10 μM Y27632. Twenty-four hrs later, the medium was changed to the mTeSR1 without the Rho kinase inhibitor. Three days after electroporation, puromycin (0.2 μg/ml) or G418 (125 μg/ml) was added into the culture medium for 5 - 10 days until colonies appeared. Individual colony were then picked up and transferred into 12-well plates in the mTeSR1 medium. Positive colonies were identified by genomic PCR analysis and replated onto 6-well plates. Primers for *SOX21* KO identification are S21ko-fw 5'-ctcggccggagacactaagg-3' and S21ko-rev 5'-gtgggtcaaaacgcaacagg-3'.

RNA extraction, cDNA synthesis and quantitative real time PCR

Total RNA was extracted using TRIzol reagent (Life Technologies, #15596026) and 2 μg total RNA were reversely transcribed into cDNA using a ReverTra Ace reverse transcriptase (Toyobo, #FSK-101), according to the manuals. Real-time qPCR was performed on the ABI ViiA7 Real-Time PCR system, using the SYBR Premix Ex Taq II

(Takara, #RR820L) according to manufacture instructions. *ACTB* was used as an internal control, a list of primers is provided in Table S5.

Western blot, coimmunoprecipitation and mass spectrometric analysis

Cells were washed 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 7.9) combined with a protease inhibitor cocktail (Selleck, B14001). Total protein concentration was measured using the Pierce BCA Protein Assay Kit (Thermo Scientific, #23227), according to the manufacture instructions. Twenty μg total proteins (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 °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). For SOX21 coimmunoprecipitation experiments, NPCs at differentiation day 7 were lysed in the Co-IP Buffer with 1 x protease inhibitor cocktail. Then, 10 mg NPC lysates were incubated with 10 μg of SOX21 antibodies or goat IgG (Sigma) at 4 °C overnight. Then, 100 μl protein G Magnetic Beads (Millipore, #LSKMAGG10) were used to pull down SOX21 antibody- or IgG- bound proteins for 2 hrs at 4 °C. Next, protein complexes were washed twice with the Co-IP buffer, once with the high salt buffer (50 mM Tris-HCl, 250 mM NaCl, 5 mM $MgCl_2$, 0.2 mM EDTA, 20% Glycerol, 0.1% NP-40, 3 mM β-Mercaptoethanol; pH 7.9), once with the Co-IP buffer, and finally eluted in the 2x SDS

loading buffer. The eluted mixture was separated by precast-GLgel (4-15%, Sangon Biotech, C621104-001). Proteins in the gel were stained with a fast silver stain kit (Beyotime, P0017S), and sequenced by mass spectrometry (Jiyun Biotech, Shanghai). All WB analyses were conducted in at least three independent experiments, unless otherwise indicated.

Immunofluorescence staining

For neural differentiation, aggregates were attached to Matrigel-coated coverslips (Fisherbrand, #12-545-82). On neural differentiation day 7, cells were fixed with 4% paraformaldehyde for 15 min at room temperature, washed 3 times with 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 washed 3 times with PBS, and incubated with secondary antibodies (FITC, 1:100; 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 S6.

Luciferase reporter assays

Aggregates of NPCs formed for 4 days were dissociated with Accutase, and 3 x 10^5 cells were transfected with luciferase plasmids using Lipo2000 (Invitrogen, #11668019). For enhancer reporter assays, the *WNT8B* enhancer was cloned by genomic DNA PCR using primers having restriction sites and the reverse primer had a minimal promoter sequence (Bgl II and Hind III): WNT8B-en-F 5'-ctggcctgtgagcctaaaca-3'; WNT8B-en-R 5'-

gccaagcttctggaagtcgagcttccattatataccctctaaccctagtaaggcaggggag-3'. The wild type and mutant *WNT8B* enhancers containing a minimal promoter were inserted into pGL4.22 (Promega). Then, 500 ng of the enhancer reporter plasmid and 10 ng of the pRL-TK internal control plasmid (Promega) were used for transfection. For Wnt signaling reporter experiments, 500 ng of the 8XTOPFlash (Addgene plasmid, #12456) or 8XFOPFlash (Addgene plasmid, #12457) plasmid together with 50 ng of the pRL-TK internal control plasmid (Promega) were used for transfection. Cells were plated onto 24-well plates in the N2B27 medium supplemented with 50 nM LDN193189, 5 μM SB431542 and 10 μM Y27632. The medium was changed 24 hrs post transfection. Cells were collected at 48 hrs post transfection and luciferase activities were examined with the Dual Glow Luciferase Assay System (Promega, #E1960) according to the manufacture instructions. The luciferase intensity ratio was determined by (TOP firefly/renilla) / (FOP firefly/renilla).

Flow cytometric analysis and cell proliferation assays

Cells were fixed, permeabilized (Cytofix/Cytoperm Kit, BD Biosciences) and stained using antibodies against SOX2 and NANOG. For cell proliferation analysis, cells were incubated with 10 μM EdU (BD Biosciences) for 1 hr, and labeled using a Cell-Light EdU Apollo488 *In Vitro* Flow Cytometry Kit (RIBO) according to manufacture instructions. Stained samples were acquired on an Accuri C6 flow cytometer (BD Accuri Cytometers), and data were analyzed using a FlowJo software.

ChIP-qPCR

Cells (1×10^7) were cross-linked with 1% formaldehyde (Thermo Scientific, #28906) for 10 min and quenched by 0.125 M glycine for 5 min at room temperature. Cell cultures were

suspended in the 1% SDS FA lysis buffer (50 mM HEPES-KOH, pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, and 1% SDS) supplemented with 1x protease inhibitors (Roche, #04693132001) and rotated for 15 min at 4 °C. Cell pellets were collected by centrifugation at 15,000 g, at 4 °C for 30 min, and resuspended in 1 ml of the 0.1% SDS FA lysis buffer (50 mM HEPES-KOH, pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, and 0 .1% SDS) with 1x protease inhibitors. Cell pellets were sheared by sonication (50% amplitude, 30s ON, 30s OFF, incubated in ice water) for 20 min (UibraCELL), and followed by centrifugation at 12,000 g for 10 min at 4 °C. An aliquot of 10 μl supernatant was used as an input, and the remaining supernatant was precleared and incubated with 20 μg of specific antibodies overnight at 4 °C. On the next day, the reaction mixture was incubated with 100 μl Protein G Magnetic Beads (Millipore, #LSKMAGG10) for 2 hrs at 4 °C, and followed by gentle wash with the 0.1% SDS FA lysis buffer (3 times), the high salt buffer (50 mM HEPES-KOH, pH 7.5, 350 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate and 0 .1% SDS), the ChIP wash buffer (250 mM LiCl, 0.5% NP40, 0.5% sodium deoxycholate, 1 mM EDTA, 10 mM Tris-HCl; pH8.0), and the TE buffer (10 mM Tris-HCl, 1 mM EDTA; pH8.0) sequentially. The chromatin was eluted from beads with 200 μl ChIP elution buffer (50 mM Tris-HCl, 10 mM EDTA, 1% SDS; pH7.5) for 30 min at 37 °C. The Input and eluted chromatin were decross-linked by adding 8 μl of 5 M NaCl at 65 °C for less than 16 hrs, followed by adding 2 μl of 20 mg/ml protease K (Sigma) for 2 hrs at 56 °C. The resulting DNA was purified using a QIAquick PCR Purification Kit (QIAGEN, #28106) and quantified using a Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen, P11496). The purified DNA was used for

examination of site specific enrichments. To evaluate site specific enrichments, 0.5 ng of input or enriched DNA was used for qPCR. Results were presented as the relative fold enrichment to the input. All primers used in ChIP-qPCR assays are listed in Table S7.

RNA-seq

Total RNA was isolated from cells using TRIzol Reagent (Invitrogen). Isolated RNA was enriched by Poly(A) tails and fragmented. Sequencing libraries were prepared according to Illumina manufacturer instructions. Paired-end RNA-seq of 2 x 150 base pair reads were sequenced on the Illumina Hiseq X-ten by Annoroad Gene Technology (Beijing, PR China) Co., Ltd. The Salmon (Patro et al., 2017) was used to calculate samples' TPM (Transcript Per Million) and raw counts. DESeq2 (Love et al., 2014) were then used to identify differentially expressed genes with following settings: fold changes > 2 and FDR < 0.05.

ChIP-seq

ChIP assays were performed as described in the ChIP-qPCR section, and ChIP-seq DNA libraries were constructed using the NEB Next Ultra DNA Library Prep Kit for Illumina, according to the manufacture instructions. Purified libraries were sequencing on Illumina Hiseq X-ten platforms by Annoroad Gene Technology (Beijing, PR China) Co., Ltd.

ChIP-seq data for TCF4 (GSE61475) were downloaded from the Gene Expression Omnibus (GEO), including peak files and bigwig files for visualization. For SOX21 ChIP-seq data, sequencing reads were aligned to the human genome (hg19) and uniquely mapped reads were kept for further analysis. Peaks were called using MACS2 with q < 0.01, and annotated by the GREAT(McLean et al., 2010). ChIP-seq signal profiles at the specific locus were visualized with the Integrative Genomics Viewer (IGV).

Weighted gene co-expression network analysis (WGCNA)

WGCNA was performed by following the tutorial written by Langfelder and Horvath (Langfelder and Horvath, 2008). RNA-seq data sets of positionally patterned NPCs were used to construct a signed weighted correlation network using the WGCNA. First, we created a signed Pearson correlation matrix between 19861 genes (genes with the maximum raw read count < 30 across all samples were filtered out). In order to construct adjacency matrix, we adjusted the soft power β to 18, which is a soft-threshold of the correlation matrix. We then calculated the topological overlap matrix and performed hierarchical clustering to group genes with highly similar co-expression relationships. Modules were defined by the Dynamic Hybrid Tree Cut algorithm from the hierarchical clustering tree. The expression profile of each module can be summarized by module eigengene, which represented the first principal component of module expression profile.

Supplemental Reference

Chang, K., Marran, K., Valentine, A., and Hannon, G.J. (2013). Creating an miR30-based shRNA vector. Cold Spring Harbor protocols *2013*, 631-635.

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.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC bioinformatics *9*, 559.

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

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nature biotechnology *28*, 495-501.

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

Qian, K., Huang, C.L., Chen, H., Blackbourn, L.W.t., Chen, Y., Cao, J., Yao, L., Sauvey, C., Du, Z., and Zhang, S.C. (2014). A simple and efficient system for regulating gene expression in human pluripotent stem cells and derivatives. Stem Cells *32*, 1230-1238.