SUPPLEMENTARY FIGURE LEGENDS

Figure S1. Expression of LHX2 in hESC-derived neural cells.

(A-B) Immunocytochemical analysis of differentiating hESCs at IVD day 20 showed that LHX2 is co-expressed with neural progenitor markers Nestin and OTX2. (C-D) Immunocytochemical analysis of differentiating hESCs at IVD day 20 revealed that LHX2 is co-expressed with the neuronal markers Tuj1 and NeuN. (E) Percentage of LHX2 and PAX6 immunopositive cells at IVD day 10, 20 and 30. Error bars represent the mean ± s.e.m. (n = 3).

Figure S2. Characterization of i*LHX2* and sh*LHX2*-expressing hESCs.

(A) RT-PCR analysis of *OCT4*, *SOX2*, and *REX1* gene expression using mRNA isolated from undifferentiated i*GFP* and i*LHX2* hESCs, and sh*Luc* and sh*LHX2*-expressing hESCs. (B) AP staining and OCT4 and SSEA4 immunofluorescence analysis of undifferentiated i*GFP*, i*LHX2*, sh*Luc* and sh*LHX2* cells show that the transgenic hESCs expressed alkaline phosphatase, OCT4 and SSEA4.

Figure S3. Ectopic expression of *LHX2* in differentiating NTU1 hESCs and iCFB46 iPSCs.

qRT-PCR analysis of the indicated genes using mRNA isolated from NTU1 hESCs and CFB46 iPSCs infected with lentivirus encoding *LHX2* or *GFP*. Error bars represent the mean ± s.e.m. (n = 3).

Figure S4. LHX2 does not regulate PAX6 expression through its promoters.

(A-B) Sequence analysis of *PAX6* promoters P0 (bp -818 to -411) and P1 (bp +7671 to +8083). Incomplete LHX2 binding motifs (TAATTA) are indicated in boldface. The PCR product “0” containing incomplete LHX2 binding sites is underlined. The PCR product “1” (acting as a negative control) is double underlined. (C) Combined ChIP and qPCR analyses of DNA isolated from hESC-NPs using an antibody against LHX2 revealed that LHX2 does not bind to the *PAX6* promoters P0 and P1. DNA from undifferentiated hESCs served as a negative control. Error bars represent the mean ± s.e.m. (n = 3).

Figure S5. Expression of *CER1* during neural differentiation of hESCs.

(A) RT-PCR and qRT-PCR analysis of *LHX2* and *CER1* mRNA isolated from differentiating hESCs on the indicated days. Error bars represent the mean ± s.e.m. (n = 3). (B) Immunocytochemical analysis of differentiating hESCs at IVD days 10 and 20 with antibodies against LHX2 and CER1, demonstrating that CER1 is co-expressed with LHX2 in neural progenitor cells. (C) qRT-PCR analysis of *CER1* using mRNA isolated from i*LHX2* hESCs with transiently expressing sh*Luc* or sh*PAX6* through lentiviral infection. dox^{-} i*LHX2* hESCs were used as a control. *ACTIN* was used as the internal control for normalization (n = 3).Error bars represent the mean ± s.e.m. Significance: *$P < 0.05$.

Figure S6. The effects of CER1 on neural and mesodermal differentiation.

(A) qRT-PCR analysis of indicated genes using mRNA isolated from differentiating sh*CER1*-expressing hESCs at IVD day 12. *ACTIN* was used as the internal control for normalization (n = 3).Error bars represent the mean ± s.e.m. Significance: *$P < 0.05$. (B) Western blot analysis of protein isolated from sh*Luc*- or sh*LHX2*-expressing hESCs transiently expressing *CER1* through lentiviral infection, using antibodies against ACTB and proteins downstream of the BMP4 (phosphor-SMAD1 and SMAD1) and WNT (phosphor-β-catenin and β-catenin) signaling pathways. Proteins isolated from hESCs expressing GFP were used as a control. Signals were determined by Image J software with consistent index and all values were normalized to sh*Luc* cells with GFP expression.

Figure S7. Identification of LHX2 binding target on CER1.

(A) Combined ChIP and qPCR analysis of DNA isolated from hESC-NPs using antibodies against LHX2. DNA from undifferentiated hESCs served as a negative control. LHX2 does not bind to the *CER1* promoter. (B) VISTA Browser display of a 47 kb fragment of the *CER1* region on Chr. 9 of the human genome (chr9: 14,681,866-14,728,766). VISTA plots are shown in the six-way (Human-Rhesus-Dog-Horse-Mouse-Rat-Chicken) alignment. Based on the annotation, conserved regions above 70%/ 100bp in VISTA are above the cutoff, and are colored under the curve (red: conserved noncoding regions; blue: conserved exons; turquoise: untranslated regions). Lines under the VISTA plot show screened conserved regions among species (black: regions I and III without putative LHX2 binding sites; gray: regions II and VI with putative LHX2 binding sites but without LHX2 binding signals; red: regions IV, V, VII, VIII and IX with LHX2 binding signals; blue: predicted promoter region). Error bars represent the mean ± s.e.m. (n = 3).

Figure S8. Forebrain gene expression in hESC-derived neural progenitors and the effect of LHX2 in forebrain lineage differentiation.

(A) qRT-PCR analysis of indicated genes using mRNA isolated from undifferentiated and differentiating H9 hESCs at IVD day 12. *ACTIN* was used as the internal control for normalization. Error bars represent the mean ± s.e.m. (n = 3). Significance: $^{*}P < 0.05$. (B) Transcriptome profiling of hESCs with altered *LHX2* expression, with regard to regional patterning. The bar chart shows the number of region-specific genes with greater than 1.3 fold expression changes, as determined using Genespring software.

Figure S1

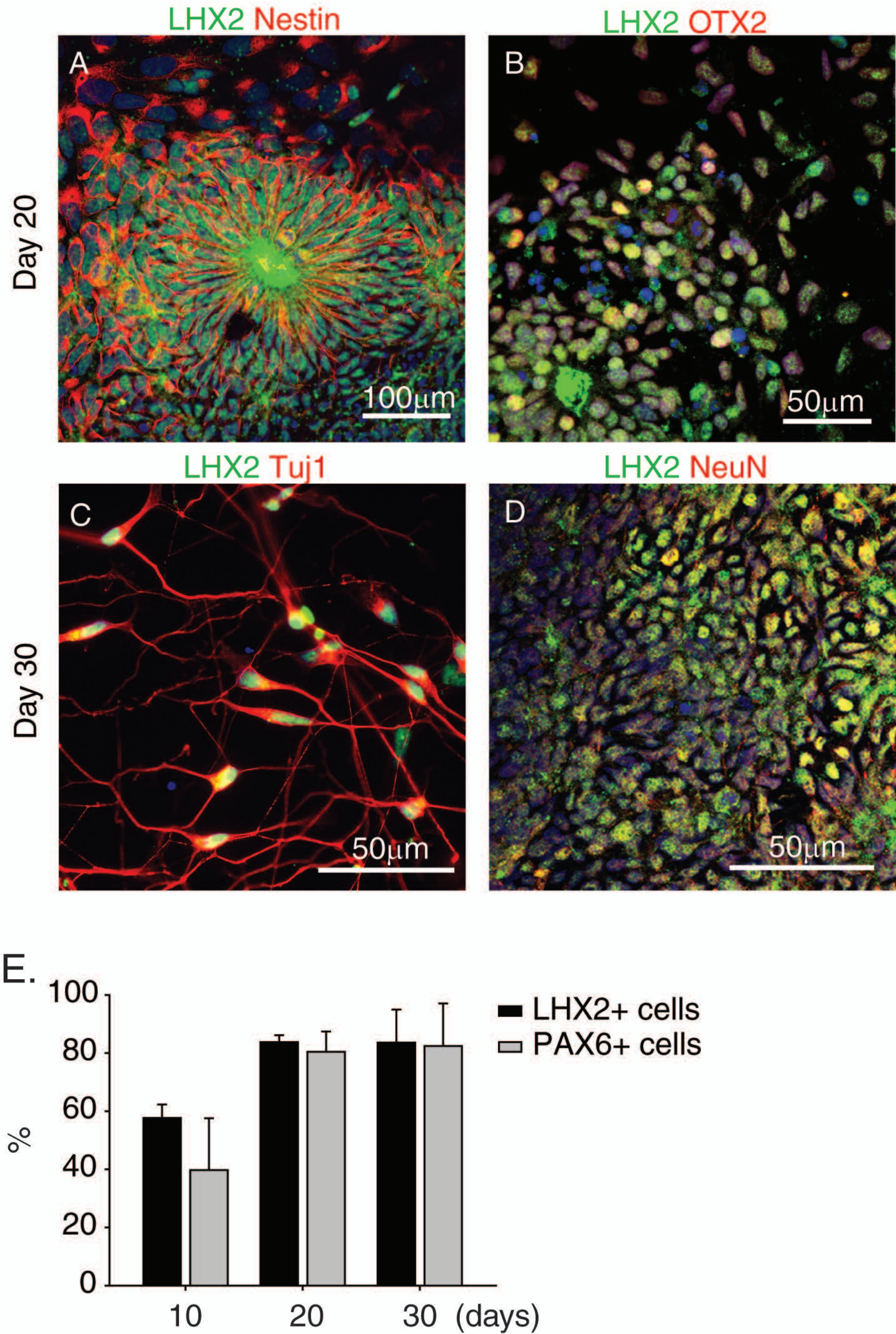

LHX2 Nestin
LHX2 OTX2
A
B
Day 20
100μm
50μm
LHX2 Tuj1
LHX2 NeuN
C
D
Day 30
50μm
50μm
E.
100
80
60
40
20
0
%
10
20
30 (days)
LHX2+ cells
PAX6+ cells

Figure S2

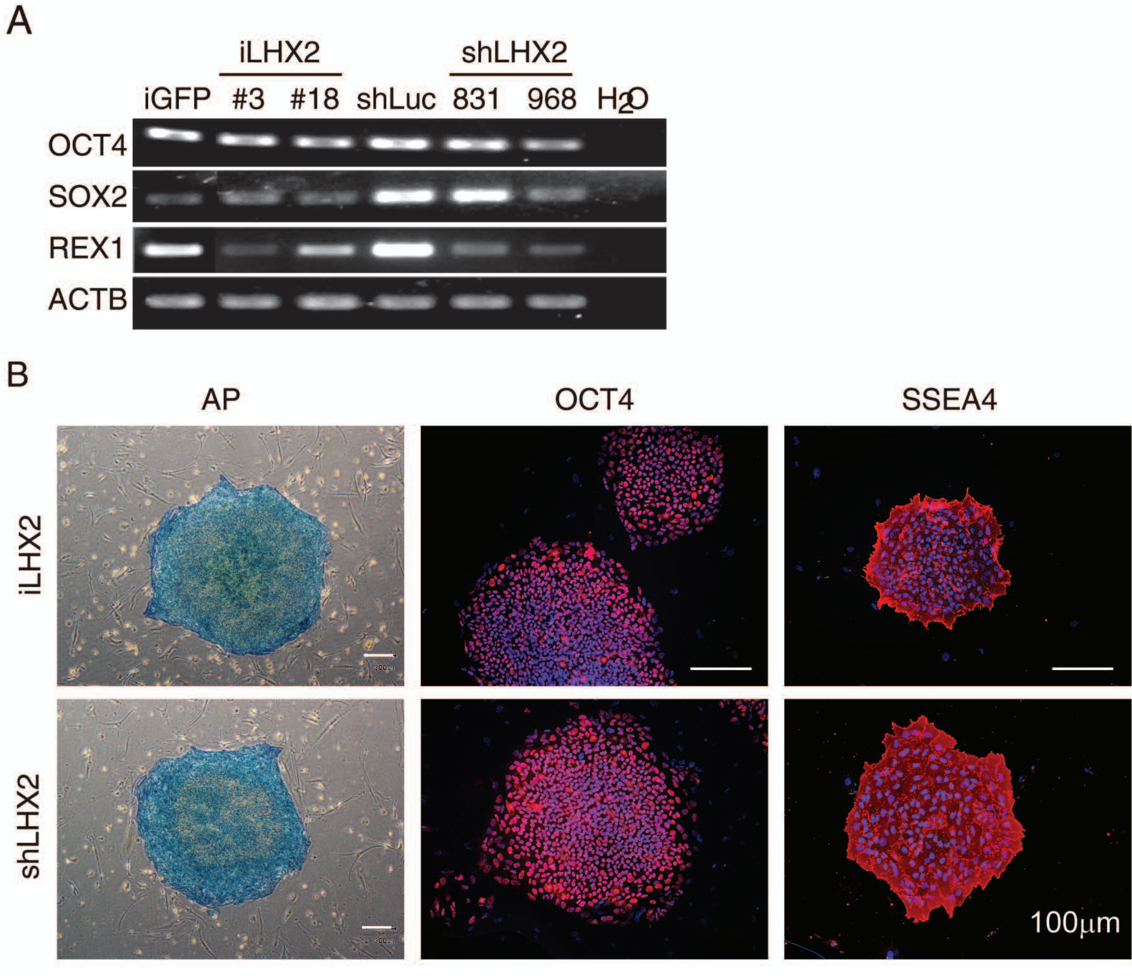
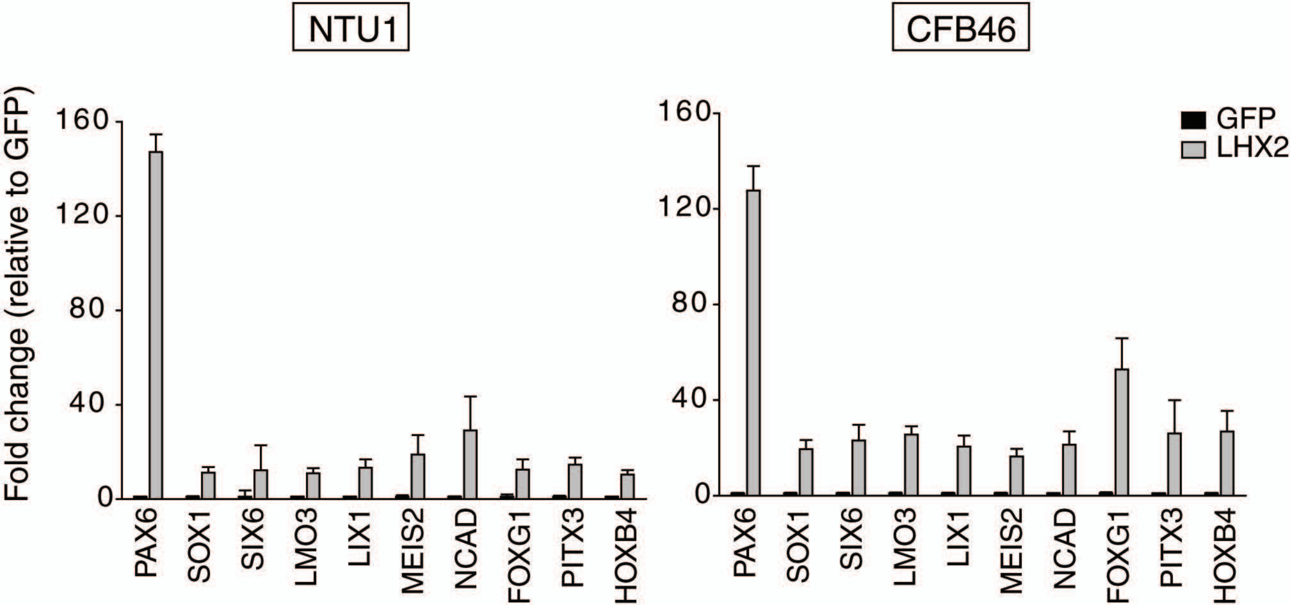

A
iLHX2
shLHX2
iGFP
#3
#18
shLuc
831
968
H_2O
OCT4
SOX2
REX1
ACTB
B
AP
OCT4
SSEA4
iLHX2
shLHX2
100μm

Figure S3



NTU1
CFB46
Fold change (relative to GFP)
160
120
80
40
0
■ GFP
▨ LHX2
PAX6
SOX1
SIX6
LMO3
LIX1
MEIS2
NCAD
FOXG1
PITX3
HOXB4

Figure S4

A. P0 promoter

-818 AGACCAGCAA GTCAGTGTCC TGCAGACCCC GCCCCAATTT CTATGAGTAT TGACTTCAGA ATCTGGGATT
TTCCTGTTTT CCTCCTCTAA GTCACAAAGT CAACAGT**TAA TT**CAAAGTCA AAGATAAATA CAATCACCTA
CATTTTCTAA TGTGGTTGGA GCCTTTCAGC CAGAGGGCGA GGGAAGCCCG GGTAGGCCCC CTTTAGGGCT
TCCCTCTTGA GAACCCAGCA GGCCTGGAGA GACCTTTGGC CTAGGCCCTG AAAAAGGGGT CGCATGTCCT
CTTCCCGGAG CCCCCGTCTG TGCCCAGCTA GTGACTTGCG GGCTCGAGGG CCAGGTTGAG GGTACTCATC
GAGCCTCGAA CTCCTCCTAA AAATGATTCC TGCCAAAAGC GCCTCTCCAT CCCGGCG

B. P1 promoter

+7671 CCGCGCCGAA TCTCGGCGTG CGCGGAGCGG GGAGATGCAG GCGAGCGCCA GAGCCCGGGC TCGGGGGCCC
TGCGCCGGGG AGAGGAGCCG GGACCCACCG GCGGAGCCGA AAACAAGTGT ATTCATATTC AAACAAACGG
ACC**AATT**GCA CCAGGCGGGG AGAGGGAGCA TCCAATCGGC TGGCGCGAGG CCCCGGCGCT GCTTTGCATA
AAGCAATATT TTGTGTGAGA GCGAGCGGTG CATTTGCATG TTGCGGAGTG ATTAGTGGGT TTGAAAAGGG
AACCGTGGCT CGGCCTCATT TCCCGCTCTG GTTCAGGCGC AGGAGGAAGT GTTTTGCTGG AGGATGATGA
CAGAGGTCAG GCTTCGCTAA TGGGCCAGTG AGGAGCGGTG GAGGCGAGGC CGGGCGCCGG CA

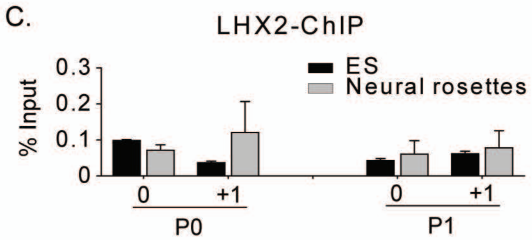


Figure S5

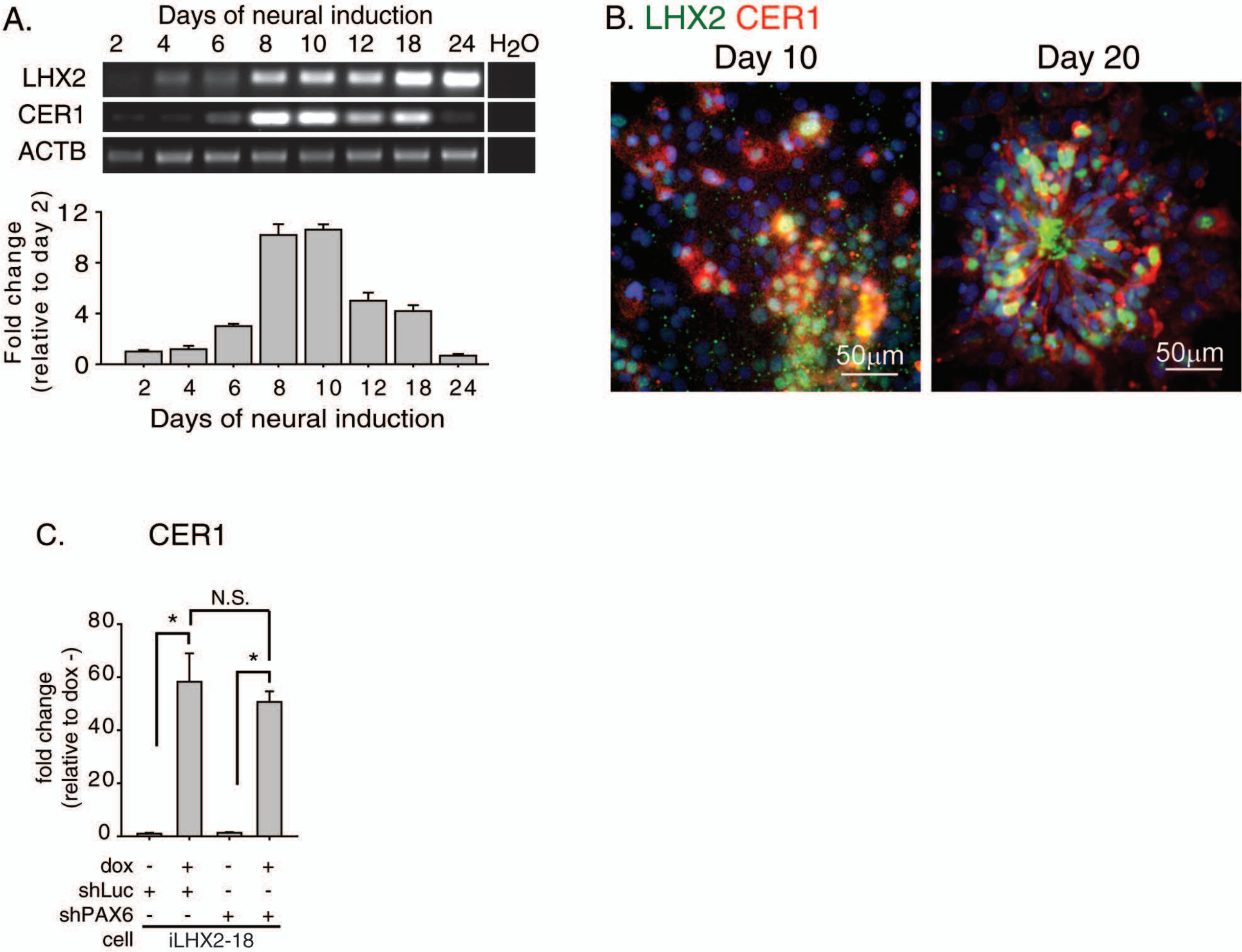

A.
Days of neural induction
2 4 6 8 10 12 18 24 H2O
LHX2
CER1
ACTB
Fold change (relative to day 2)
0 4 8 12
2 4 6 8 10 12 18 24
Days of neural induction
B. LHX2 CER1
Day 10
Day 20
50μm
50μm
C. CER1
fold change (relative to dox -)
0 20 40 60 80
*
N.S.
*
dox - + - +
shLuc + + - -
shPAX6 - - + +
cell iLHX2-18

Figure S6

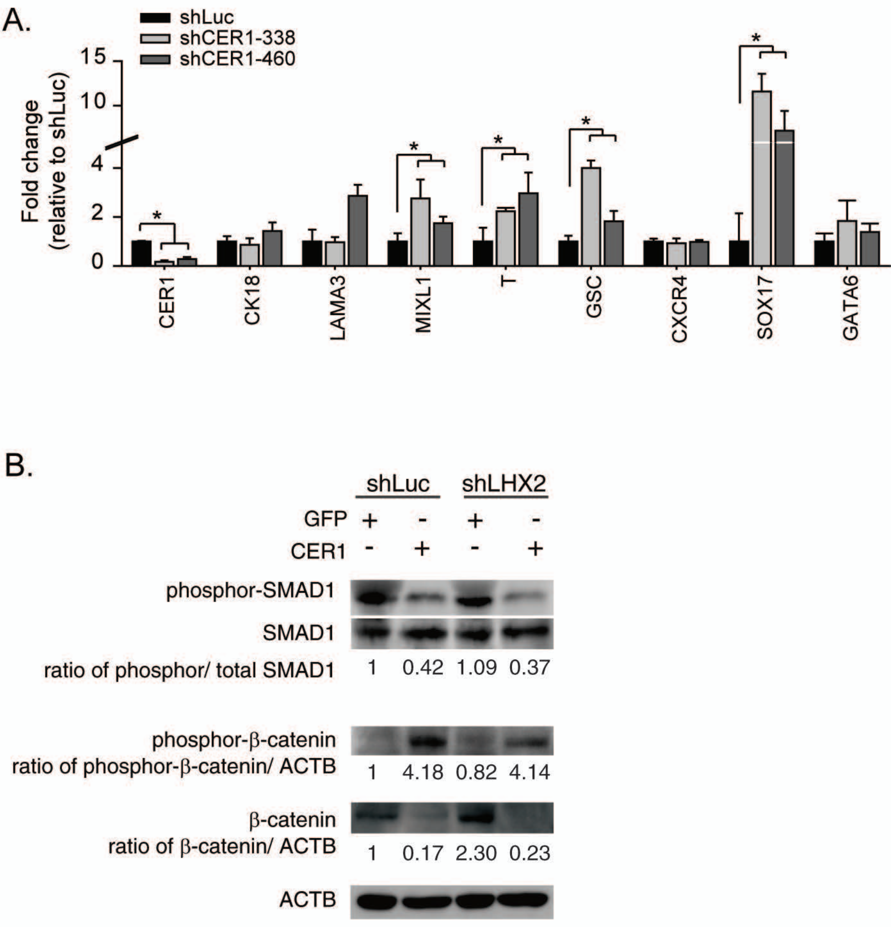

A.
shLuc
shCER1-338
shCER1-460
Fold change
(relative to shLuc)
15
10
4
2
0
*
CER1
CK18
LAMA3
MIXL1
T
GSC
CXCR4
SOX17
GATA6
B.
shLuc
shLHX2
GFP + - + -
CER1 - + - +
phosphor-SMAD1
SMAD1
ratio of phosphor/ total SMAD1 1 0.42 1.09 0.37
phosphor-β-catenin
ratio of phosphor-β-catenin/ ACTB 1 4.18 0.82 4.14
β-catenin
ratio of β-catenin/ ACTB 1 0.17 2.30 0.23
ACTB

Figure S7

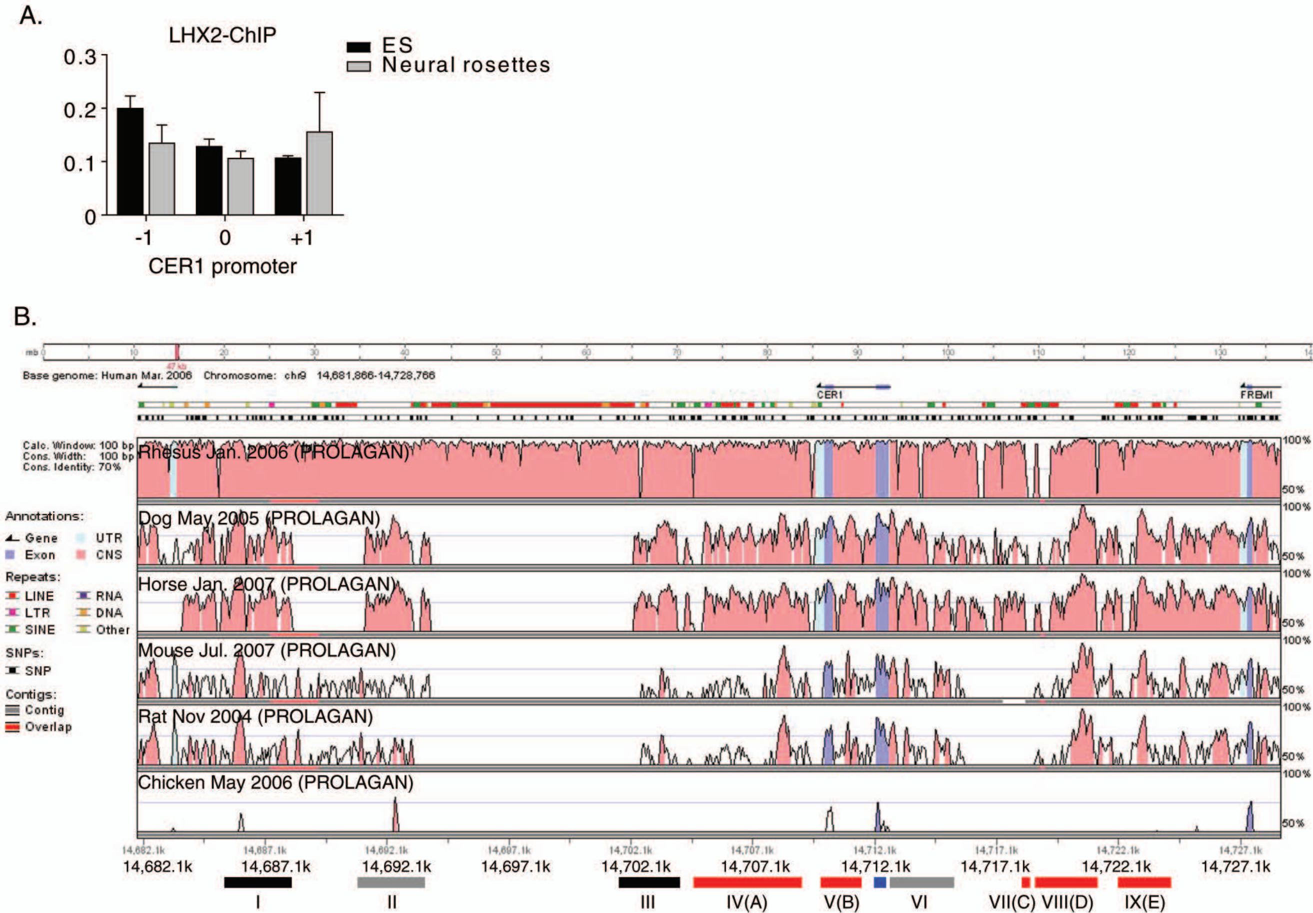
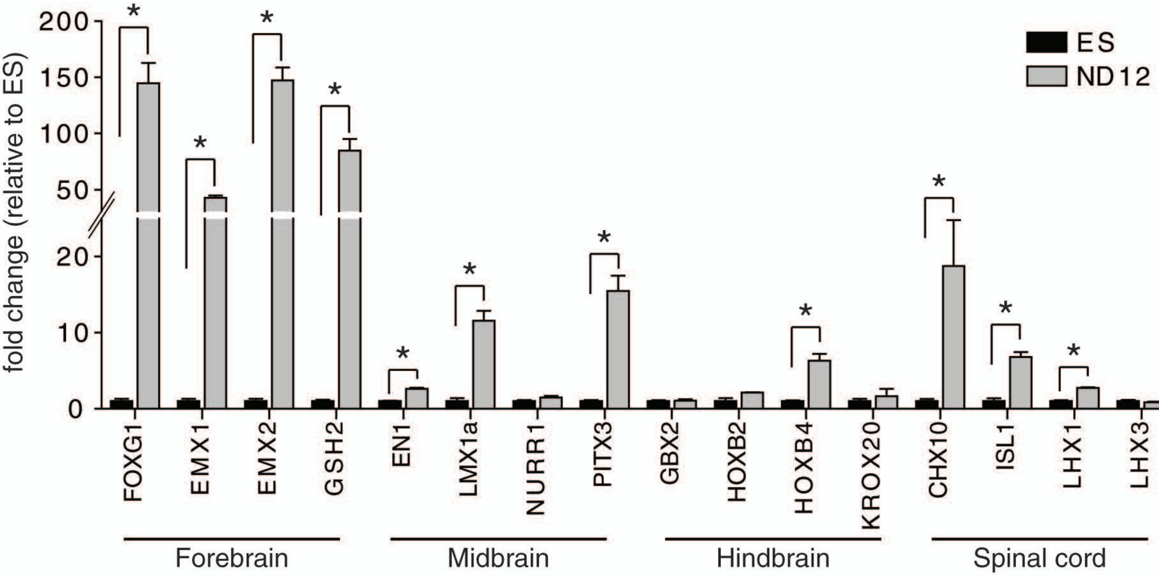

A.
LHX2-ChIP
ES
Neural rosettes
0.3
0.2
0.1
0
-1
0
+1
CER1 promoter
B.
mb
47 kb
Base genome: Human Mar. 2006 Chromosome: chr9 14,681,866-14,728,766
CER1
FRMD1
Calc. Window: 100 bp
Cons. Width: 100 bp
Cons. Identity: 70%
Rhesus Jan. 2006 (PROLAGAN)
Dog May 2005 (PROLAGAN)
Horse Jan. 2007 (PROLAGAN)
Mouse Jul. 2007 (PROLAGAN)
Rat Nov 2004 (PROLAGAN)
Chicken May 2006 (PROLAGAN)
100%
50%
Annotations:
Gene
UTR
Exon
CNS
Repeats:
LINE
RNA
LTR
DNA
SINE
Other
SNPs:
SNP
Contigs:
Contig
Overlap
14,682.1k
14,687.1k
14,692.1k
14,697.1k
14,702.1k
14,707.1k
14,712.1k
14,717.1k
14,722.1k
14,727.1k
I
II
III
IV(A)
V(B)
VI
VII(C)
VIII(D)
IX(E)

Figure S8

A.



B. shLuc v.s. shLHX2

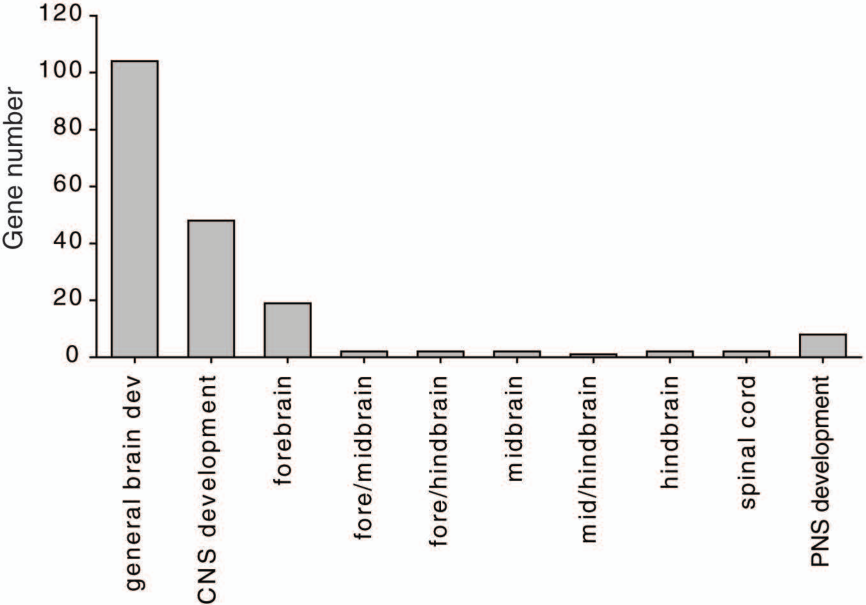


Table SI. Primers for RT-PCR, qRT-PCR and q-ChIP

RT-PCR	NAME	Forward	Reverse
	OCT4	CTTGCTGCAGAAGTGGGTGGAGGAA	CTGCAGTGTGGGTTTCGGGCA
	NESTIN	GCCACCTCACTGCAGTAGTG	AGTCAGCACATGGGAGTGC
	LHX2	GAAGGGGCGGCCGAGGAAAC	GCTGGTCACGGTCCAGGTGC
	PAX6	AACAGACACAGCCCTCACAAACA	CGGGAACTTGAACTGGAACTGAC
	SOX1	CAATGCGGGGAGGAGAAGTC	CTCTGGACCAAACTGTGGCG
	LMO3	TGTGCCTGCTGTGACTGTCGC	CAGGACGCGCAGTTTCCCGT
	LIX1	GGAATTTTGGGAAAGCAAGC	CAGCACTGAAAGTTGCCAAA
	RAX	AGCGCGCGTTCGAGAAGTCC	GTTGACCTTGCCGGCCAGCT
	OTX2	CGCCTTACGCAGTCAATGGG	CGGGAAGCTGGTGATGCATAG
	NCAD	TGGCTCTACTGCTGGGTCCT	TGGCCCCCAGTCGTTCAGGT
	MEIS2	CCAGGGGACTACGTTTCTCA	TAACATTGTGGGGCTCTGTG
	MAP2	GCATGAGCTCTTGGCAGG	CCAATTGAACCCATGTAAAGCC
	GFAP	AGGGCTGACACGTCCAC3	GCCTTAGAGGGGAGAGGAG
	GAPDH	ACCACAGTCCATGCCATCAC	TCCACCACCCTGTTGCTGTA
qRT-PCR	OCT4	CCGAAAGAGAAAGCGAACCAG	AGAACCACACTCGGACCACATC
	NANOG	CAGCAGATGCAAGAACTCTCCA	CATTGCTATTCTTCGGCCAGT
	PAX6	GATAACATACCAAGCGTGTCATCAATA	TGCGCCCATCTGTTGCT
	SOX1	TCCTGGAGTATGGACTGTCCG	GAATGCAGGCTGAATTCGG
	SIX6	AACAAGAATGAGTCGGTGCT	CAGCGGGAACTTCTTCCTTA
	LMO3	ACGGAGCGCCCGAGATGATG	GGCACAGGATGAGGTTGGCC
	LIX1	GGAATTTTGGGAAAGCAAGC	CAGCACTGAAAGTTGCCAAA
	MEIS2	CCAGGGGACTACGTTTCTCA	TAACATTGTGGGGCTCTGTG
	NCAD	TCCTGATATATGCCCAAGACAA	TGACCCAGTCTCTCTTCTGC
	FOXG1	AGAAGAACGGCAAGTACGAGA	TGTTGAGGGACAGATTGTGGC
	EMX1	GAGACGCAGGTGAAGGTGT	GTTGATGTGATGGGAGCCCT
	EMX2	CACAGAAACGGACAACATGG	CTTTAGACGAGGGTCGCTTG
	GSH2	CTCGCTCATCATCAAGGACA	AGTGCAGGTGCGAAGTGAC
	PITX3	TGGACTAGGCCCTACACACAGA	GCGCACGTTTATTTCATTTATCTTT
	EN1	ACCGCACCACCAACTTTTTC	TCCTTTTTGCAGCCGAAGTC
	NURR1	TTCTCCTTTAAGCAATCGCCC	AAGCCTTTGCAGCCCTCACAG
	LMX1a	ACATGAGCAGGGCTGAGACT	TGGTGTCTCCTTCAGCACAG
	HOXB4	AAAGAGCCCGTCGTCTACC	GTGTAGGCGGTCCGAGAG
	GBX2	CTCGCTGCTCGCCTTCTC	GCCAGTCAGATTGTCATCCG
	HOXB2	TTTAGCCGTTCGCTTAGAGG	CGGATAGCTGGAGACAGGAG
	KROX20	TTGACCAGATGAACGGAGTG	CAGAGACGGGAGCAAAGC
	ISL1	ATGACAAAACTAATATCCAGGGG	ATTGACCAGTTGCTGAAAAGC
	CHX10	CCCATCAGTGGAGTCCAGAT	TGTGAGGCATAGGACATGGA
	LHX1	TAAATGTAAAGCCCTCCGTATCA	CCATCTTACCGAGTGTGTC
	LHX3	GCTGGGCCCGGGAAAGTTCG	GTGCTAGCAGCAGGTCGCCTC
	βACTIN	GAGCACAGAGCCTCGCCTTT	ACATGCCGGAGCCGTTGTC
	CK18	TGGCAGACATCCGGGCCCAA	TGTGGTGCTCTCCTCAATCTGCT
	LAMA3	CCTGGGGCAGTGTCTGGGCTA	GCTACAACCCTGGCTGGGTCTAA
	MIXL1	AACGAAATGTCTGAAGCCCCA	TCCTCCCATGAGTCCAGCTTT
	T	GGCATCGGAGGCGCGAGAAC	CGGACCAAGACTGTCCCCGCT
	GSC	CGGAGAAGTGGAACAAGACGT	GCGTGTGCAAGAAAGTAGCATC
	CXCR4	GGATCAGCATCGACTCCTTCA	TTCTCAAACTCACACCCTTGCTT
	SOX17	GCCAAGGGCGAGTCCCGTATC	CGACTTGCCCAGCATCTTGCTCA
	GATA6	GGCGCCTCCTCTCTCCTTTTTATTC	TCTAGCTCCTCGGGCGGGC
q-ChIP	P0-0	CTAAGTCACAAAGTCAACAGT	GCTGAAAGGCTCCAACCA

	P0-1	CCTGGAGAGACCTTTGG	GAGCCCGCAAGTCACTA
	P1-0	TGTATTCATATTCAAACAAACGGAC	CGATTGGATGCTCCCTC
	P1-1	ATGTTGCGGAGTGATTAGT	ACCAGAGCGGGAAATGA
	E6-1	AGATCCTGCTACGTCTGTC	GATGGTTCAAGGCACCG
	E6-0	GGTGCCTTGAACCATCTT	CGCAACTACCGCCTCTA
	E6+1	GTCCCTCCCTCACCCCA	TCGGTGGAGCCTCAGGG
	ETel-A-2	AGTTACTGAGGGAGGGATAG	CAGAAGAGCACACTCGG
	ETel-A-1	ATCGCAGTTCACGACAT	GCGGCAGGGTGTCTTTA
	ETel-A-0	CTCCCAGCGTGTGATTT	TCAGTGAGAATCGCTAATTATTAGG
	ETel-A+1	GAAGTGTCCCCTTAGTGAG	GCTGAACCACAAGTTAGTCT
	ETel-A+2	TCATAAATCAAACCCAGCCAT	ACGATCTCCAATAAACATCTCTAAT
	ETel-B-0	GGTGACATAATTACCTCTGACCA	GAGAAGGCGACAGAGGA
	ETel-B+1	CCGTGATAGGCTTGTCTTT	ATCTGAATAGTAATTCCAGTCATCC
	ETel-B+2	TGGGTGGTGTCATTTGTTG	TGCCTCTGATTCCTCACG
	+156kA-1	ATAACCTTCCTTCCCAAAGTC	GAAGATCTGAATTTAATCCTGGC
	+156kA0	TAACATGTTCACAAACTCTGTTAATTA	AATGCCTTTAAGCACTTGCAG
	+156kA1	ACTGGCAGAAATGACACT	GTTAGGAGCACAGGCTTT
	+156kA2	AACTATCTGCCATGCCC	CTTCAGGACACAGGCTTT
	+156kB-2	GGAGGATTTAATAGTAAGGGTAGTT	TTTGTCTGTTGCTCACTTGG
	+156kB-1	CCTTTGGAATTAGGAAATATCAACA	CAGGGCATTAGTAAGTCTTTTAATC
	+156kB0	AAGGTTCCTTATGTAGACTTAATTA	TTTTCTTAGGTAGCCACTTTTC
	+156kB1	TACTTAGGAGGCTGAGGCA	GCAATCTCGGCTCACTG
	+156kB2	CCTGTAATCCCAGCACTTT	CGTCTCGATCTCCTGAC
	+156kC-2	AAATTGCTCCCATTTTTCCC	GCAGTGTTATATGAATTCTTCTGAT
	+156kC-1	AATAGAGATTATCCAAAACAGTAGT	AAGTTTAAAACAAATTCTGCTTTCC
	+156kC0	TGATGTTTTGTATAATATTTAAATAATTATGC	CATGCTTAAGCATATACGTATATTA
	+156kC1	CACCATCACATGAGGCAAA	TGTCACAGGAGCAGTGTAAA
	+156kD0	GGAGGATTTAATAGTAAGGGTAGTT	CCCTTGTCTCTTGGGGAA
	+156kD1	GCCTCCTGAGTAGCTGAA	ATCTAAAGCAAAGAAGAGAATCACA
	+156kD2	GTGTGTGTAAGAACCCTGTAGA	AAACAGCTTCATCCTAGGTC
	CER1-pro-1	CTGGGCTTTTTCTAAATGTATGC	GCCTTGTGCTGGAATAGAAT
	CER1-pro-0	GTTGTCTTTGCTAATACTGCTC	CAGTGAAAGCCTGGGAC
	CER1-pro+1	CAAAGTGACGGCAGGAG	GTCAGTAGCCAGGCAGA
	CER1-A-2	AATGATGTATAGAAAGTGTTTGGC	CAATAGGTATCATCTACTGAGGTTA
	CER1-A-1	TACAATCCTGAAAGAAACCTCC	GCCATCATTAATTAAATTATAAGGTCCA
	CER1-A0	ATAAAGATATAATTACAAGCTGATATGAGT	GACAAGTTTGGGTTCACG
	CER1-A1	ATCCTCTACTCATTTATTCATATTCTTTG	GTCTATGATGAGCCTGGTC
	CER1-A2	CACTTCTAGGAGATAACAAATATAAAGC	GACTACTAATTATTGAGTTCTTACTATGTG
	CER1-B-2	AGCAAACACATTTCCATTACG	TCCTCAGATTTACAGTCCATAGC
	CER1-B-1	TGTCTAATTGGATGTGAGCG	TATGTATTGGGAGCCTACTATGC
	CER1-B0	ACCTTTACCTCTCATAACTCCA	TTTGAAATTCTTCTTTTGAATTCTATTGC
	CER1-B1	TATCTTCATGTTTAAGGACCTAATGC	CATCCTACTGTCTACTTCCTCT
	CER1-B2	CTATTAAAGAAGTCCCTCATAAACCTA	CCAAATACTCTCATATCCATTTACGA
	CER1-C0	GCCTTTCACATAGTAAATGCC	GAATTAAAGTGGGATAGGGAAGAA
	CER1-C1	GCACTTTGGGAGACCGA	TAGAGACAGGGTTTCACCAT
	CER1-D0	TGTCTCAACGGAACCCT	AGGCAGAGCAGTCAAAG
	CER1-D1	AGAACCCATCAGATAAGTGAG	AGTGTGGAAAGGAGAAACGAA
	CER1-D2	CTCCATGTCACTGAGATGTATTT	CGAGGAATCACCGAATAGC
	CER1-E-2	AAGTTATATGTGTAAGTTATGAGGATAGTA	GCCTTTAGAACAGGATCGT
	CER1-E-1	TATAACTCCAGAATATTTCCATCATCTC	TGGGACTAACTGCCAATAGGTA
	CER1-E0	CTGCTCAGAATGCTAATAGATATG	GAGAAAGTTACTGTGTGAAGAAAT

	CER1-E1	CTCATCTCTCTCTCCCGTA	CCTCTGCCTTGATTGATAGTT

Table SII. Region position

Name	Position (UCSC genome browser: Human Feb. 2009 GRCh37/hg19)
PAX6 related	
P0	chr11:31,839,282-31,840,731
P1	chr11:31,832,704-31,834,156
-6k	chr11:31,847,523-31,848,122
Etel	chr11:31,837,445-31,839,211
+156k	chr11:31,679,369-31,690,371
CER1 related	
promoter	chr9:14,722,713-14,723,559
A	chr9:14,715,743-14,716,426
B	chr9:14,720,391-14,721,701
C	chr9:14,728,766-14,729,088
D	chr9:14,729,390-14,729,961
E	chr9:14,733,116-14,735,245