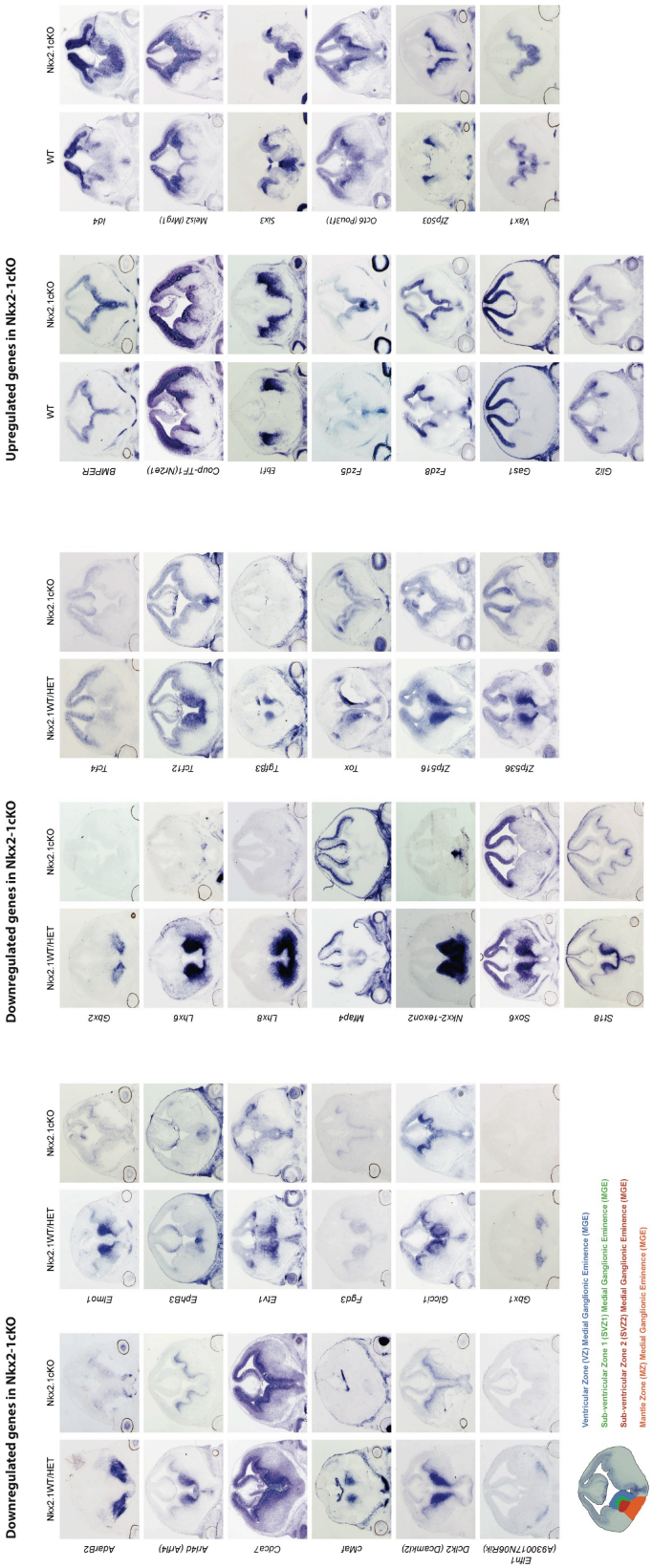
Neuron, Volume *91*

Supplemental Information

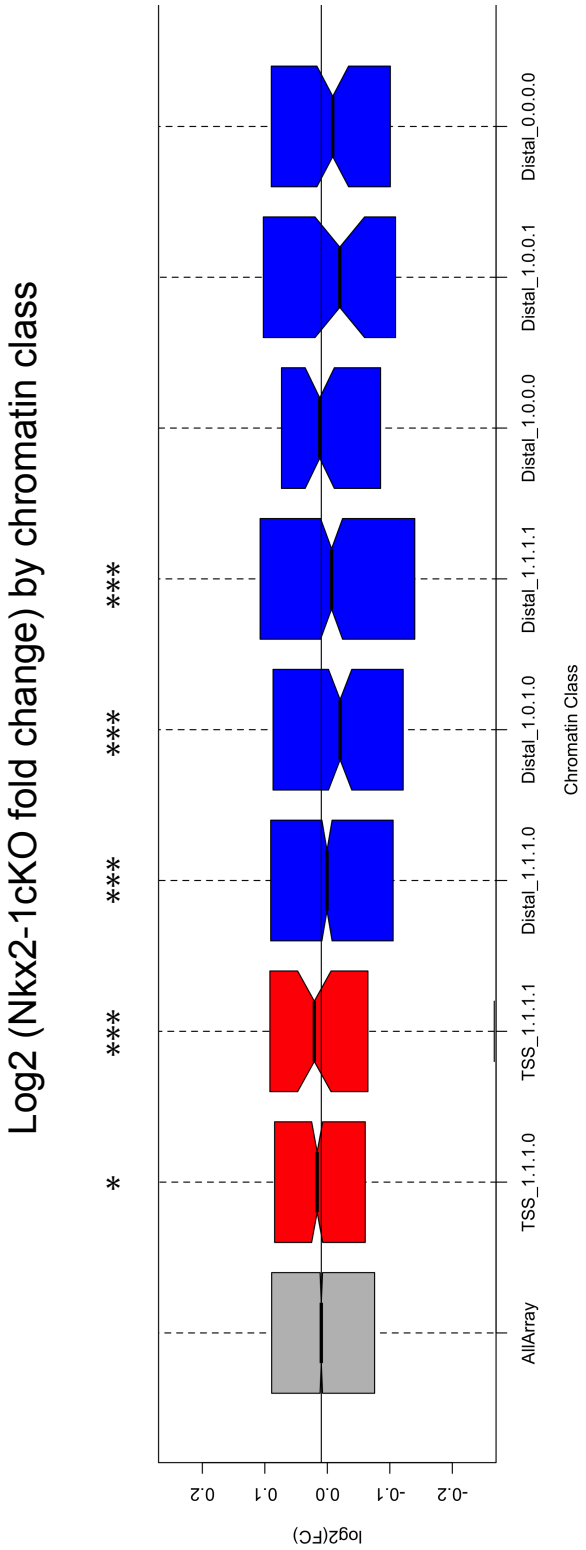
Transcriptional Networks Controlled by NKX2-1 in the Development of Forebrain GABAergic Neurons

Magnus Sandberg, Pierre Flandin, Shanni Silberberg, Linda Su-Feher, James D. Price, Jia Sheng Hu, Carol Kim, Axel Visel, Alex S. Nord, and John L.R. Rubenstein

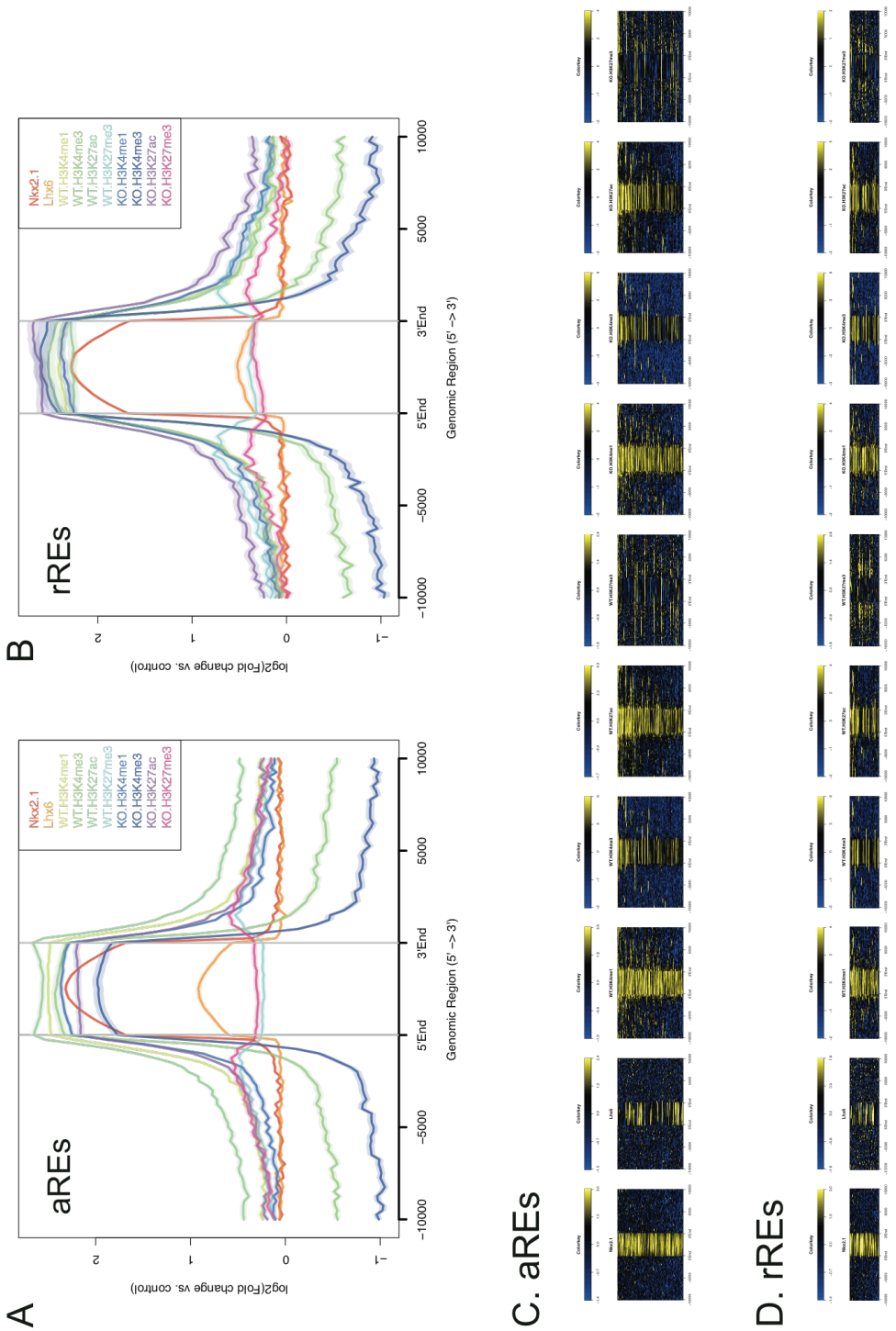
Sandberg *et al.* Figure S1



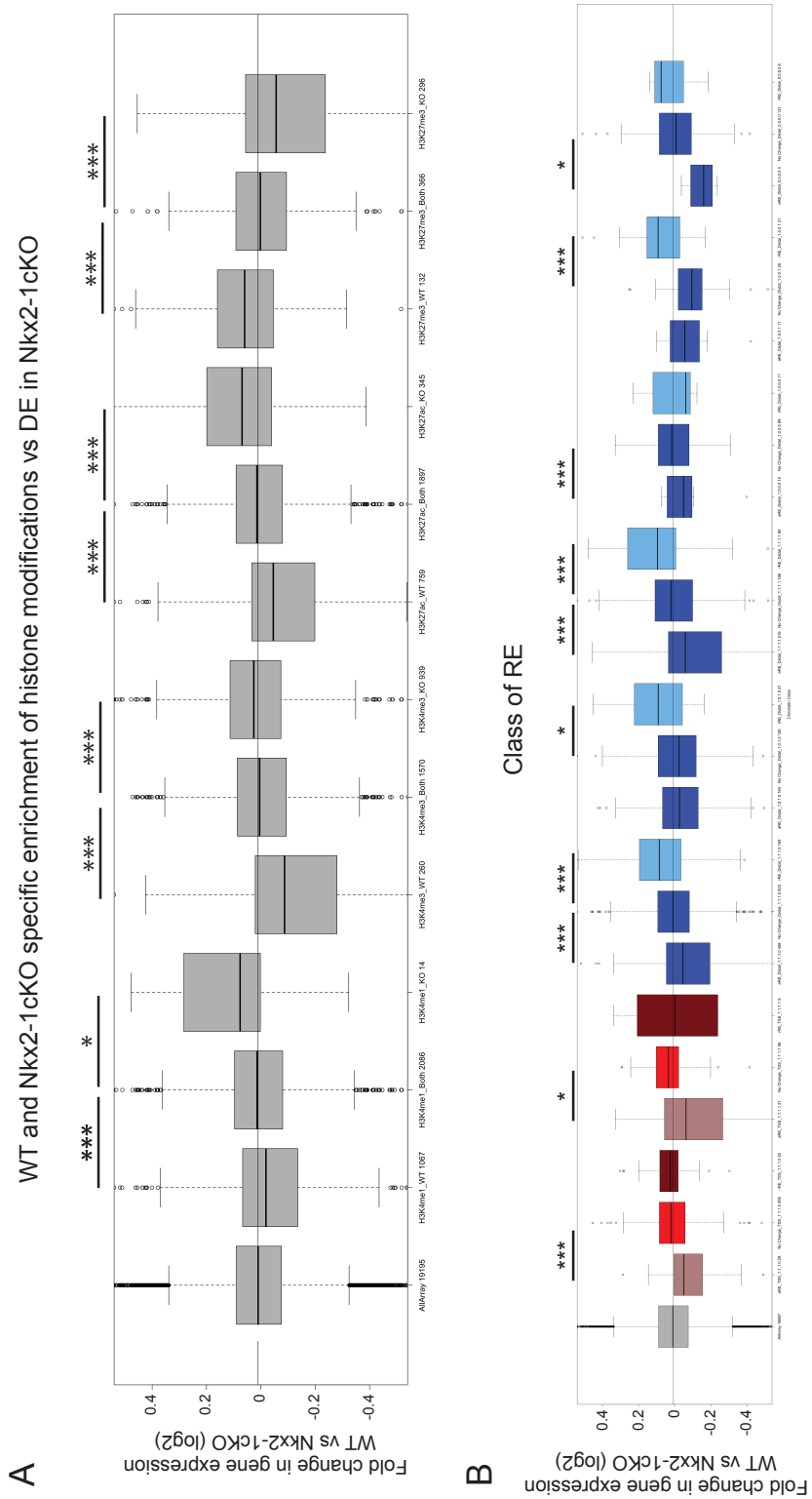
Sandberg *et al*. Figure S2



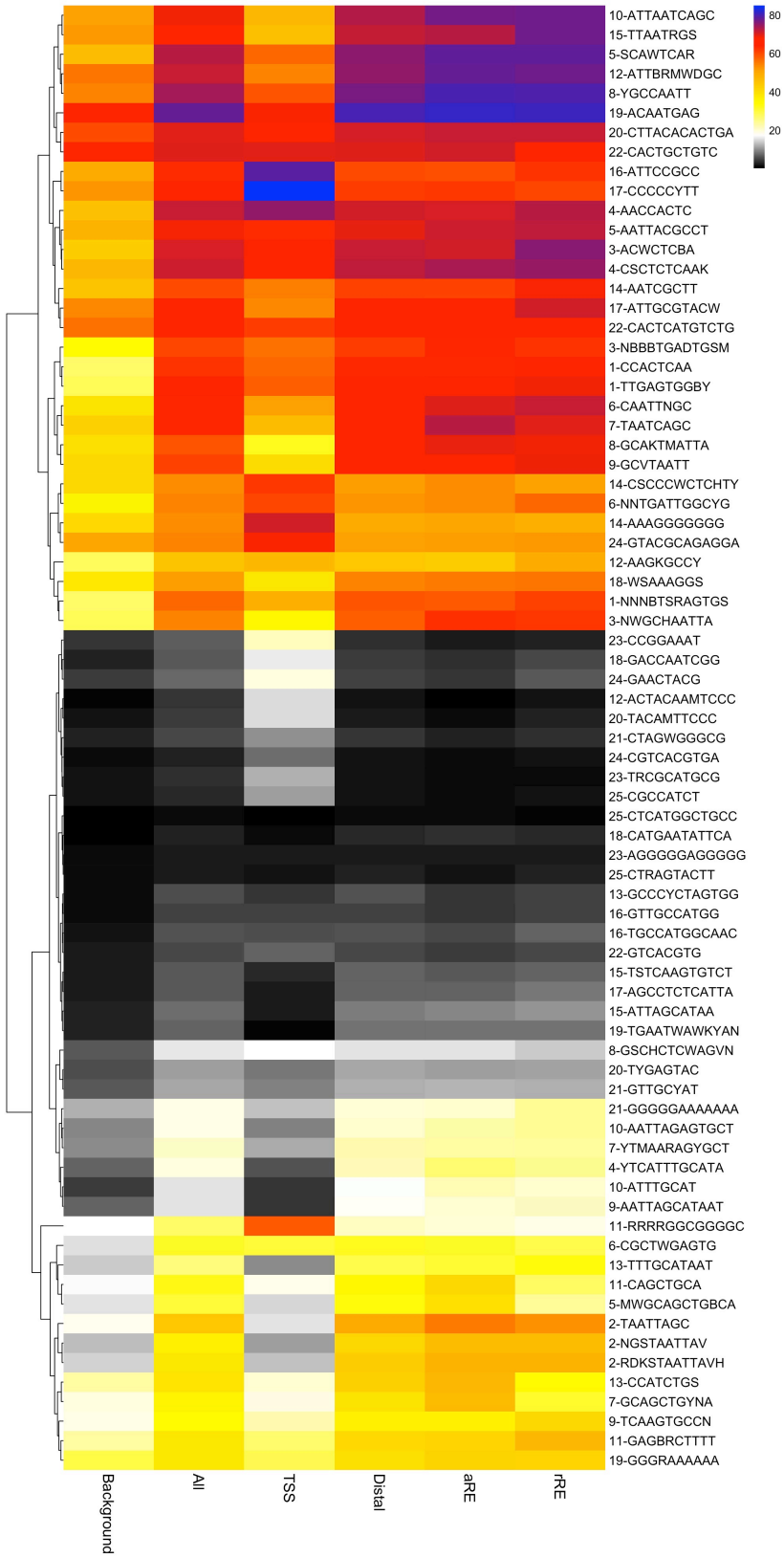
Sandberg *et al*. Figure S3



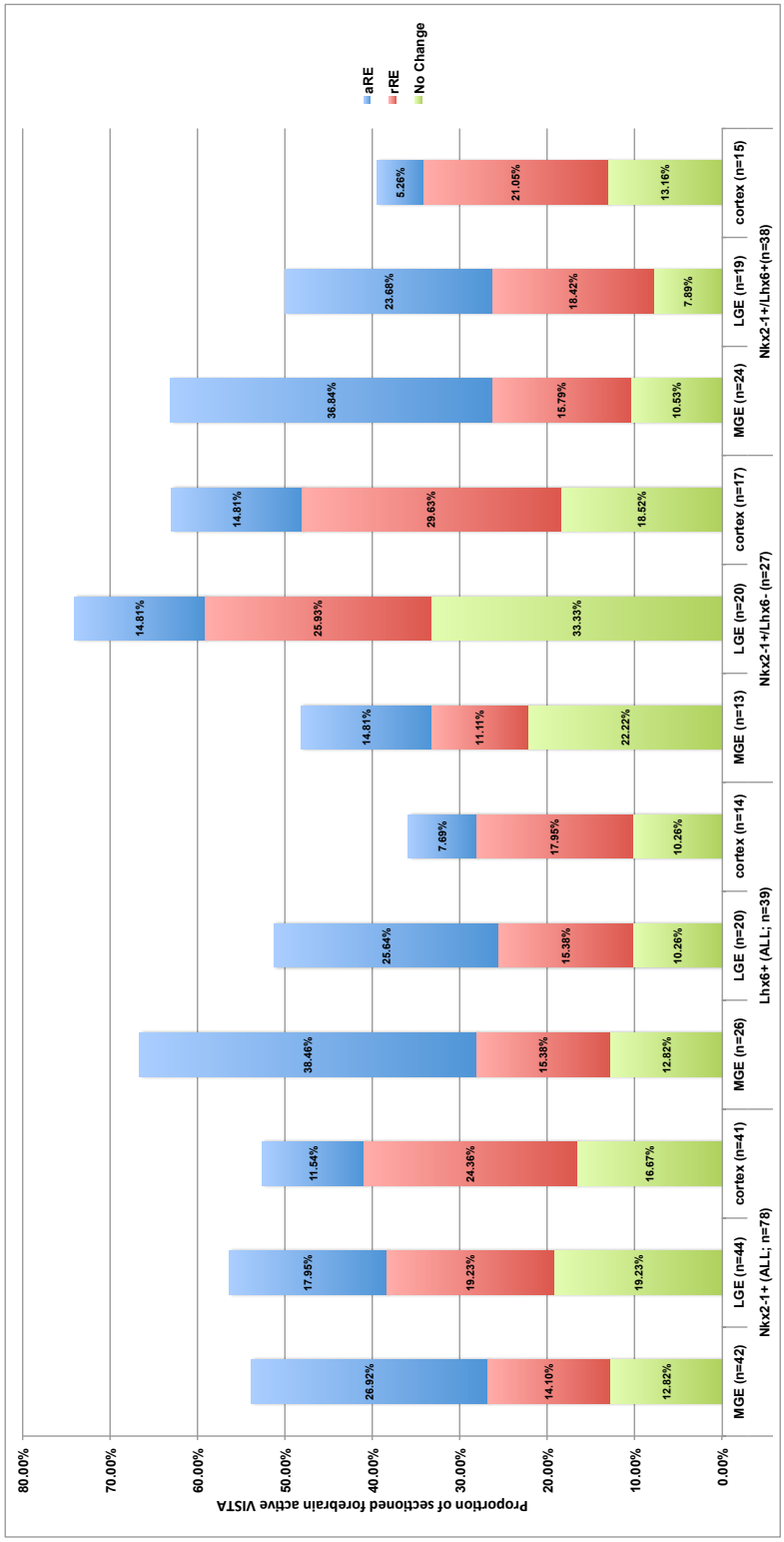
Sandberg *et al.* Figure S4



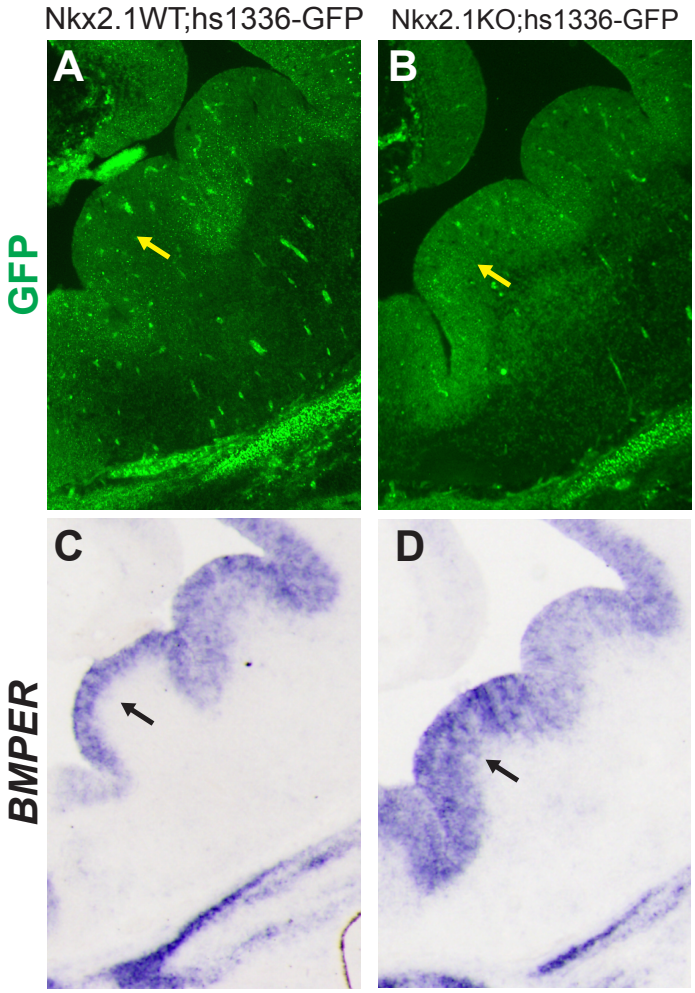
Sandberg *et al.* Figure S5

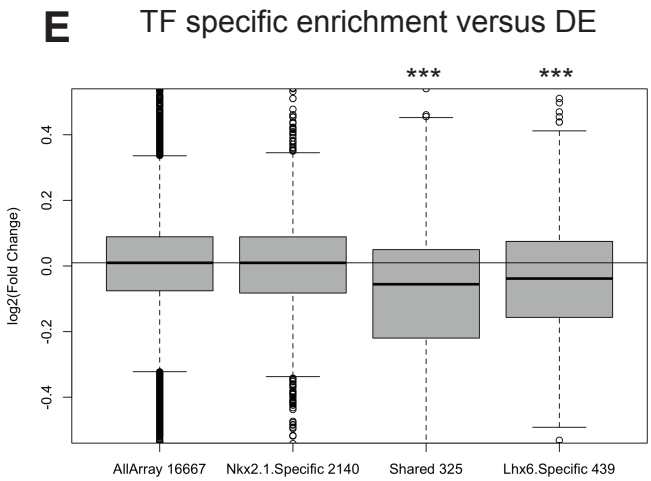


Sandberg *et al.* Figure S6

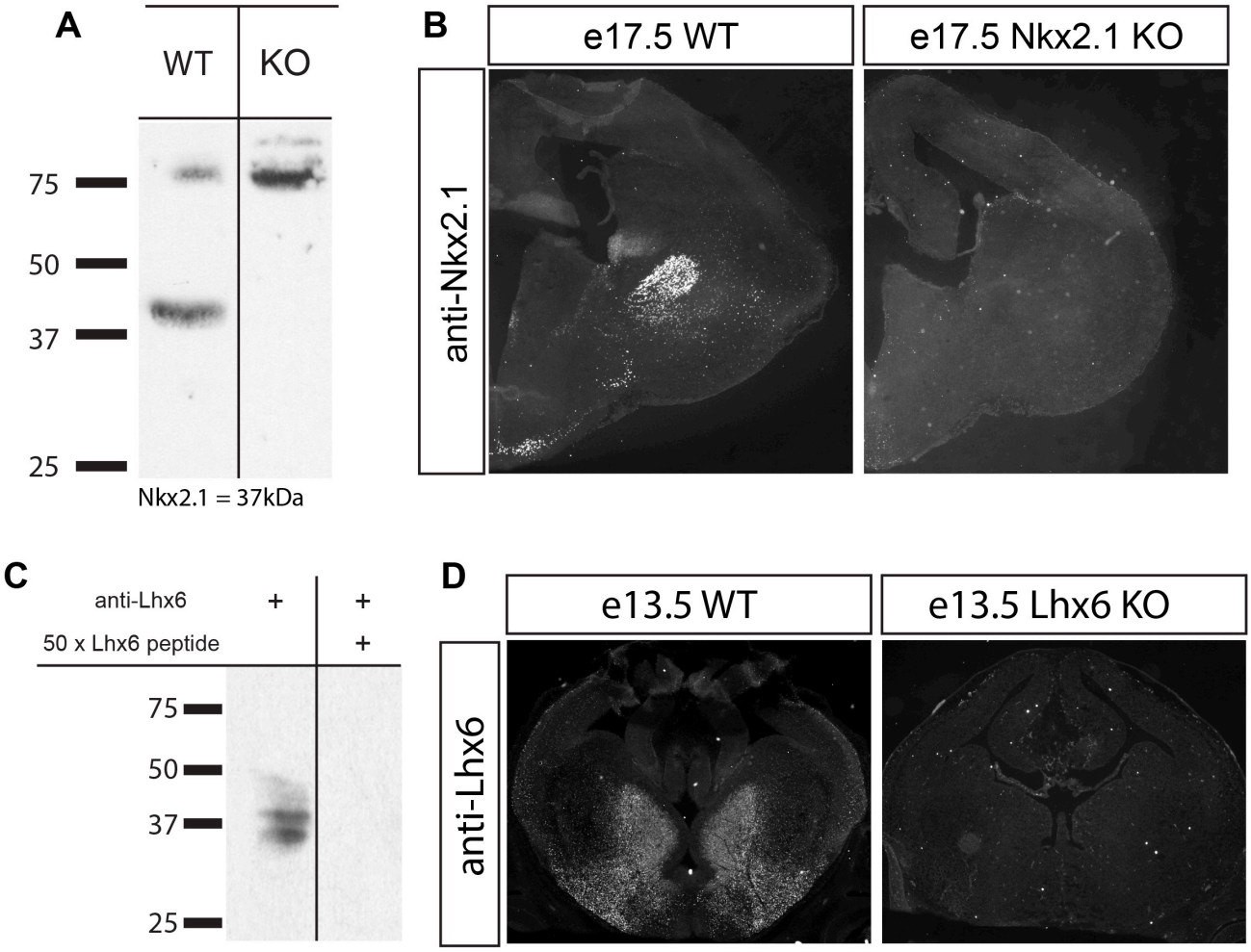


Sandberg *et al.* Figure S7





Sandberg *et al.* Figure S8



Supplemental Figure Legends

Figure S1. *ISH* on differentially expressed genes in the *Nkx2-1*cKO MGE at e13.5. Related to Figures 1, 7 and 8.
Downregulated genes; *AdarB2*, *Arl4d* (*Arfl4*), *Cdca7*, *cMAF*, Dclk2 (*Dcamkl2*), *Elfn1* (*A930017N06Rik*), *Elmo1*, *EphB3*, *Etv1*, *Fgd3*, *Glcci1*, *Gbx1*, *Gbx2*, *Lhx6*, *Lhx8*, *Mfap4*, *Nkx2-1exon2*, *St18*, *Tcf4*, *Tcf12*, *Tgfβ3*, *Tox*, *Zfp516* and *Zfp536*. Cutoff, -1.25 > log2FC. Upregulated genes; *BMPER*, *Ebf1*, *Fzd5*, *Fzd8*, *Gas1*, *Gli2*, *Id4*, *Meis2* (*Mrg1*), *Six3*, *Oct6* (*Pou3f1*), *Zfp503*, *Vax1*. Cutoff, log2FC < 1.25.

Figure S2. Average differential gene expression of closes TSS to NKX2-1 bound REs. Related to Figures 3 and 5.
NKX2-1 bound REs divided into 8 groups based on their location and histone profile in the e13.5 WT MGE; 1 - Active TSS (1.1.1.0 = H3K4me1+, H3K4me3+, H3K27ac+ and H3K27me3-), 2 - Bivalent TSS (1.1.1.1), 3 - Active (strong) Distal (1.1.1.0), 4 - Active (weak) distal (1.0.1.0), 5 - Bivalent distal (1.1.1.1), 6 - Latent distal (1.0.0.0), 7 - Repressed distal (1.0.0.1), 8 - No Histone (0.0.0.0). Average log2 fold differential gene expression (WT vs *Nkx2-1*cKO) for each NKX2-1 RE group. Two sample T-test was used to test significance between the groups: ***p < 0.001, *p < 0.05.

Figure S3. Identification of NKX2-1 bound REs mediating transcriptional activation and repression. Related to Figure 5.
(A-B) Average profile of NKX2-1, LHX6, H3K4me1, H3K4me3, H3K27ac and H3K27me3 at aREs and rREs in WT and *Nkx2-1c*KO MGE at e13.5. (C) Heatmap showing NKX2-1, LHX6, H3K4me1, H3K4me3, H3K27ac and H3K27me3 in WT and *Nkx2-1*cKO MGE (e13.5) at aREs. (D) Heatmap showing NKX2-1, LHX6, H3K4me1, H3K4me3, H3K27ac and H3K27me3 in WT and *Nkx2-1*cKO MGE (e13.5) at rREs.

Figure S4. Comparison of differential gene expression and changes in histone profiles at NKX2-1 REs. Related to Figure 5.
(A) NKX2-1 bound REs with significant H3K4me1, H3K4me3, H3K27ac and H3K27me3 in WT, *Nkx2-1*cKO or both WT and *Nkx2-1c*KO MGE at e13.5. This plot show the average log2 fold differential gene expression (WT vs *Nkx2-1*cKO) for each NKX2-1 RE group. (B) NKX2-1 bound REs divided into 24 groups based on their WT histone profile, combined with changes in histone profile in the e13.5 *Nkx2-1*cKO MGE. WT RE groups are defined as follow; Active_TSS (1.1.1.0 = H3K4me1+, H3K4me3+, H3K27ac+ and H3K27me3-), Bivalent_TSS (1.1.1.1), Active (strong) distal (1.1.1.0), Active (weak) distal (1.0.1.0), Bivalent distal (1.1.1.1.), Repressed distal (1.0.0.1). aREs have reduced H3K4me3 and/or H3K27ac, and/or increased H3K27me3. rREs have increased H3K4me3 and/or H3K27ac, and/or reduced H3K27me3. This plot show the average log2 fold differential gene expression (WT vs *Nkx2-1*cKO) for each NKX2-1 RE group. Wilcoxon test was used to test significance between aREs and rREs compared to REs with no histone change: ***p < 0.001, *p < 0.05.

Figure S5. *De novo* motif analysis of NKX2-1 bound REs. Related to Figure 5.
Frequency of discovered *de novo* motifs at; background, all REs, TSS, distal REs, aREs and rREs. All motif sequences indicated.

Figure S6. Proportion of aREs and rREs of NKX2-1 and/or LHX6 bound VISTA. Related to Figure 7.
Proportion of MGE, LGE and cortex positive VISTA elements. The VISTA elements were divided into the following groups based on NKX2-1 and LHX6 binding; NKX2-1 (all REs), LHX6 (all REs), +/-REs and +/+REs. Each group has been scored for the proportions of aREs, rREs and REs with no change in the histone configuration.

Figure S7. Activity of BMPER enhancer, hs1336, in WT and *Nkx2-1c*KO MGE at e13.5. Related to Figure 8.
Activity of hs1336-GFP in the basal ganglia of control (A) and *Nkx2-1*cKO (B) at e13.5. *In situ* analysis of *BMPER* transcription in WT (C) and *Nkx2-1*cKO (D) forebrain at e13.5. Yellow arrows (A and B) indicate increased hs1336 activity in the mutant SVZ of the MGE. Black arrows (C and D) indicate a consistent increase in *BMPER* expression in the VZ and SVZ1 of the mutant MGE. (E) Average differential gene

expression of closes TSS to +/-REs, +/+REs and -/+REs. Wilcoxon test was used to test differences between control genes (whole array) and the different groups: ***$p < 0.001$.

Figure S8. Validation of polyclonal NKX2-1 and LHX6 antibodies. Related to Figure 2, 6 and Experimental Procedure.
(A) Western blot analysis on WT and *Nkx2-1*KO MGE tissue (e13.5), using NKX2-1 polyclonal for the blot. (B) NKX2-1 IHC on WT and *Nkx2-1*KO telencephalon (e17.5). (C) Western blot analysis on WT and *Lhx*6KO MGE tissue (e13.5), using LHX6 polyclonal for the blot. (D) LHX6 IHC on WT and *Lhx6*KO telencephalon (e13.5).

Data S1. Summary of *Nkx2-1*cKO RNA expression micro-array and ChIP-seq analysis. Related to Figures 1-3 and Figures 5-8.

Supplemental Experimental Procedures

Control primers ChIP-qPCR

Nkx2-1_positive: 5' TCTCTCCCCCTCTCCCTGCAC, 3'GCTTGGATTGCAAAGGAACTAAGCG
Lhx6_positive: 5'GCCATTAATGAACCTATGACAGCC, 3'AGCAGTGTTCAGCTCCTCTG
H3K4me1_positive1: 5'ACGTCCAGAAGCATCTGCATTTCCA,
3'ACTGTTCAAGGGAAGACACTGACCT
H3K4me3_H3K27me3_positive: 5' AACTCACTCAAGCACACGGT,
3'GAGTTGGGCGCCAATCCTAT
H3K27Ac_positive: 5'GCAGTGGCTCATTCCCTCCCC, 3'GCAGCCTCGCCTTCCTCCAG
Negative_control_1: 5'AGCCAGACCCCTGCAGCCATT, 3'AGTGGCTTGCTTAGGCAGGGGAG
Negative_control_2: 5'TGAAAGTGCGTGGCAGCCCAA, 3' GCTGTCCTCTGCATTCCCCCA
Negative_control_3: 5'TGCAGCAGCCTCTCCAGGTG, 3'TGCCAAATGTGCTCAGTACCCTGC

Primers used to generate in situ probes

Adarb2: 5'TCTCCGAGACAGCCGAGAAG, 3'CTCACGCCACTAAGGAGAGG
Arl4d: 5'CTGCATGTCGTCGTCATTGG, 3'CGTCTCTTCTTGCTCGACCG
BMPER: 5'GGAGGGTGTGGTGACAGAGT, 3'GTCCAGGTCGATGTGGAAGT
Cdca7: 5'ATCGTCCTCCTCCGATGACA, 3'AGTAAGCATGGACGTTCCCG
Dclk2: 5'TTGAAGAACGGGACAAAAGG, 3'GACTTCGGAGTGGAAAGCTG
Dlgap1: 5'AGGCTTCCCGGAGTAACAAT, 3'CACTTCAACGGCAGAACTCA
Elfn1: 5'TACAGCGTGGGTACTTGTGG, 3'GTAGTCAGGATCACCAGCGG
Elmo1: 5'AGCCAGTTAGCCCCAAGC, 3' TCCACAGCATGTTTAAGGGA
EphB3: 5'CGCTAACGCTGTGGAGGT, 3'CATGAAGACAAACCCGGC
Fgd3: 5'GAACTTCCCCTGTGAGGAGAG, 3'GGTACACAGTAGAGCATCACAT
Fzd5: 5'CACTCAAGACTCCGGAGAGG, 3'GTAGCGGAATCGTTCCATGT
Fzd8: 5'CTGCCACAACCCCTTCTTTA, 3'CGGTTGTGCTGCTCATAGAA
Gas1: 5'CACTGCATCTCGGCGCTTA, 3'AAGTGTGACCCGAGCAGC
Glcci1: 5'ACGGGACCCTCATGTTCA, 3'AGTGTTGCCAGAGCCGAG
Gli2: 5'GGAGGGAAGGTACCATTATGACC, 3'GAATGGTGATGGGGTTCACGG
Id4: 5'GGCCAGAGCAGAAATTAAGAGA, 3'GAAACTGGATACTGGGCAAAAC
Meis2: 5'TCTGGAGTTAGAAAAGGTCCACG, 3'GTCCATAACCTGTCCGCCAA
Mfap4: 5'TACAAGCTGGGCTTTGGC, 3'GGTGTGGCAGTGTGCAAG
Nkx2-1-exon2: 5'CTGTGCGTTTGTCGCTTACA, 3'CTCAAGCGCATCTCACGTCT
Oct6: 5'CCCTCTACGGTAACGTGTTCTC, 3'AGTGTGTGGTGGTGGTGGT
St18: 5'TTGGTAATGTGAGCTGTGAAGG, 3'TGTGTGTGTATGTGCATGTGTG
Tcf12: 5'TCTCGAATGGAAGACCGC, 3'CTCCCTCCTGCCAGGTTT

Tgfβ3: 5'GGACTTCGGCCACATCAAGA, 3'CCAGGTTGCGGAAGCAGTAA
TOX: 5'TGCCTGGACCCCTACTATTG, 3'CATTGATGCCACAATCTTCG
Vax1: 5'AGAAGAAGGACCAGGGCAAG, 3'CTCAGGTAACGGGCTGAGAG
ZFP503: 5'CAAGAAAGATCCGGAAGCTG, 3'CCCATTAGCAGACACCCAGT
ZFP516: 5'GCAGGAAGCAAGAACAGACC, 3'CAGCCTCTTCACCACTGTCA
ZFP536: 5'GCTCATCAGTCACGTGGAGA, 3'CTGGTGGTAGGTTCGGAAAA

Additional probes were described in Long et al., 2009.

Generation of mm1429 transgenic

mm1429 was amplified from mouse genomic DNA by polymerase chain reaction (PCR), sequence validated and inserted into pCR-Blunt II-TOPO (Zero Blunt TOPO PCR Kloning Kit, Invitrogen). mm1429 was further sub-cloned into an Hsp68-LacZ reporter vector (5'-XhoI:mm1429:XhoI-3') (Kothary et al., 1989). Generation of transgenic mice and embryo staining was done as previously described (Poulin et al., 2005). Transgenic embryos were sectioned and β-galactosidase was detected using an anti- β-galactosidase antibody (MP Biomedicals, Cat.# 0856028).

Site directed mutagenesis of mm1429
NKX2-1 and LHX6 motifs were mutated in pCR-Blunt II-TOPO, sequence verified and sub-cloned into a pGL4.23-Luciferase reporter with a minimal β-globin-promoter using SacI and XhoI. The luciferase reporter was generated from pGL4.23- β-globin-promoter-GFP and pGL4.23-Luciferase as templates.

Primers used to generate mm1429 reporters

mm1429_WT_F, CGGAGTCTGCTCCTAGACCCAGATTCCATAC
mm1429_WT_R, CGCTCGAGCCACACACAGGCACAAAAGTAC
mm1429_mut_site#1-F, GCGGCTGGAATTTTTTCATAAGTACCCTTTGGGGGTGG
mm1429_mut_site#1-R, CCACCCCCAAAGGGTACTTATGAAAAAATTCCAGCCGC
mm1429_mut_site#2-F, CCTGGGCTTCAGGAAAGGATTTTTTGGTGTCTCTGGAAATTGCCATC
mm1429_mut_site#2-R, GATGGCAATTTCCAGAGACACCAAAAAATCCTTTCCTGAAGCCCAGG
mm1429_mut_site#3-F, CTTGAAGGCACAAATACTGCCAAAAAAACTCTAAGCCGATCCAAATAG
mm1429_mut_site#3-R, CTATTTGGATCGGCTTAGAGTTTTTTTGGCAGTATTTGTGCCTTCAAG
mm1429_mut_site#4-F, CTTTCTTTCAAATTTAAAAAATGTGTACTTTTGTGCC
mm1429_mut_site#4-F, GGCACAAAAGTACACATTTTTTAAATTTGAAAGAAAG

Sequence pGL4.23-β-globin-promoter-Luciferase

GGGTTACTAGTACGTCTCTCAAGGATAAGTAAGTAATATTAAGGTACGGGAGGTATTGGACA
GGCCGCAATAAAATATCTTTATTTTCATTACATCTGTGTGTTGGTTTTTTGTGTGAATCGATAG
TACTAACATACGCTCTCCATCAAAACAAAACGAAACAAAACAAACTAGCAAAATAGGCTGTC
CCCAGTGCAAGTGCAGGTGCCAGAACATTTCTCTGGCCTAACTGGCCGGTACCTGAGCTCGCT
AGCCTCGAGGATATCAAGATCTGGCCTCGGCGGCCAAGCTTAGCCCGGGCTGGGCATAAAAG
TCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTAGCCTGCAGGTCGAGGAGCGCAGCCTT
CCAGAAGCAGAGCGCGGCGACCGGTCCACCATGGAAGATGCCAAAAACATTAAGAAGGGCC
CAGCGCCATTCTACCCACTCGAAGACGGGACCGCCGGCGAGCAGCTGCACAAAGCCATGAAG
CGCTACGCCCTGGTGCCCGGCACCATCGCCTTTACCGACGCACATATCGAGGTGGACATTACC
TACGCCGAGTACTTCGAGATGAGCGTTCGGCTGGCAGAAGCTATGAAGCGCTATGGGCTGAA
TACAAACCATCGGATCGTGGTGTGCAGCGAGAATAGCTTGCAGTTCTTCATGCCCGTGTTGGG
TGCCCTGTTCATCGGTGTGGCTGTGGCCCCAGCTAACGACATCTACAACGAGCGCGAGCTGCT
GAACAGCATGGGCATCAGCCAGCCCACCGTCGTATTCGTGAGCAAGAAAGGGCTGCAAAAGA
TCCTCAACGTGCAAAAGAAGCTACCGATCATACAAAAGATCATCATCATGGATAGCAAGACC

GACTACCAGGGCTTCCAAAGCATGTACACCTTCGTGACTTCCCATTTGCCACCCGGCTTCAAC
GAGTACGACTTCGTGCCCGAGAGCTTCGACCGGGACAAAACCATCGCCCTGATCATGAACAG
TAGTGGCAGTACCGGATTGCCCAAGGGCGTAGCCCTACCGCACCGCACCGCTTGTGTCCGATT
CAGTCATGCCCGCGACCCCATCTTCGGCAACCAGATCATCCCCGACACCGCTATCCTCAGCGT
GGTGCCATTTCACCACGGCTTCGGCATGTTCACCACGCTGGGCTACTTGATCTGCGGCTTTCGG
GTCGTGCTCATGTACCGCTTCGAGGAGGAGCTATTCTTGCGCAGCTTGCAAGACTATAAGATT
CAATCTGCCCTGCTGGTGCCCACACTATTTAGCTTCTTCGCTAAGAGCACTCTCATCGACAAGT
ACGACCTAAGCAACTTGCACGAGATCGCCAGCGGCGGGGCGCCGCTCAGCAAGGAGGTAGGT
GAGGCCGTGGCCAAACGCTTCCACCTACCAGGCATCCGCCAGGGCTACGGCCTGACAGAAAC
AACCAGCGCCATTCTGATCACCCCCGAAGGGGACGACAAGCCTGGCGCAGTAGGCAAGGTGG
TGCCCTTCTTCGAGGCTAAGGTGGTGGACTTGGACACCGGTAAGACACTGGGTGTGAACCAGC
GCGGCGAGCTGTGCGTCCGTGGCCCCATGATCATGAGCGGCTACGTTAACAACCCCGAGGCTA
CAAACGCTCTCATCGACAAGGACGGCTGGCTGCACAGCGGCGACATCGCCTACTGGGACGAG
GACGAGCACTTCTTCATCGTGGACCGGCTGAAGAGCCTGATCAAATACAAGGGCTACCAGGT
AGCCCCAGCCGAACTGGAGAGCATCCTGCTGCAACACCCCAACATCTTCGACGCCGGGGTCG
CCGGCCTGCCCGACGACGATGCCGGCGAGCTGCCCGCCGCAGTCGTCGTGCTGGAACACGGT
AAAACCATGACCGAGAAGGAGATCGTGGACTATGTGGCCAGCCAGGTTACAACCGCCAAGAA
GCTGCGCGGTGGTGTTGTGTTCGTGGACGAGGTGCCTAAAGGACTGACCGGCAAGTTGGACG
CCCGCAAGATCCGCGAGATTCTCATTAAGGCCAAGAAGGGCGGCAAGATCGCCGTGTAATAA
TTCTAGAGTCGGGGCGGCCGGCCGCTTCGAGCAGACATGATAAGATACATTGATGAGTTTGGA
CAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCT
TTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGT
TTCAGGTTCAGGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTA
AAATCGATAAGGATCCGTCGACCGATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCCTTCCGG
TGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGTCTTCTTTATCATGCAACTCGTAG
GACAGGTGCCGGCAGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGC
TGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGAT
AACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCG
CGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAA
GTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCC
CTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGG
GAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTC
CAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTA
TCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAG
GATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACG
GCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAA
GAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCA
AGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGT
CTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGG
ATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGT
AAACTTGGTCTGACAGCGGCCGCAAATGCTAAACCACTGCAGTGGTTACCAGTGCTTGATCAG
TGAGGCACCGATCTCAGCGATCTGCCTATTTCGTTCGTCCATAGTGGCCTGACTCCCCGTCGTG
TAGATCACTACGATTCGTGAGGGCTTACCATCAGGCCCCAGCGCAGCAATGATGCCGCGAGA
GCCGCGTTCACCGGCCCCCGATTTGTCAGCAATGAACCAGCCAGCAGGGAGGGCCGAGCGAA
GAAGTGGTCCTGCTACTTTGTCCGCCTCCATCCAGTCTATGAGCTGCTGTCGTGATGCTAGAGT
AAGAAGTTCGCCAGTGAGTAGTTTCCGAAGAGTTGTGGCCATTGCTACTGGCATCGTGGTATC
ACGCTCGTCGTTCGGTATGGCTTCGTTCAACTCTGGTTCCCAGCGGTCAAGCCGGGTCACATG
ATCACCCATATTATGAAGAAATGCAGTCAGCTCCTTAGGGCCTCCGATCGTTGTCAGAAGTAA
GTTGGCCGCGGTGTTGTCGCTCATGGTAATGGCAGCACTACACAATTCTCTTACCGTCATGCC
ATCCGTAAGATGCTTTTCCGTGACCGGCGAGTACTCAACCAAGTCGTTTTGTGAGTAGTGTAT
ACGGCGACCAAGCTGCTCTTGCCCGGCGTCTATACGGGACAACACCGCGCCACATAGCAGTA
CTTTGAAAGTGCTCATCATCGGGAATCGTTCTTCGGGGCGGAAAGACTCAAGGATCTTGCCGC
TATTGAGATCCAGTTCGATATAGCCCACTCTTGCACCCAGTTGATCTTCAGCATCTTTTACTTT
CACCAGCGTTTCGGGGTGTGCAAAAACAGGCAAGCAAAATGCCGCAAAGAAGGGAATGAGT
GCGACACGAAAATGTTGGATGCTCATACTCGTCCTTTTTCAATATTATTGAAGCATTTATCA

Luciferase reporter assay in primary MGE cultures

MGE tissue was dissected from e13.5 embryos, triturated and plated onto 24-well plates (1 embryo/2wells). Primary cultures were transfected with a total of 500ng DNA using Lipofectamin 2000 (Thermo Fischer) and cultured in Neurobasal Medium (Thermo Fischer) supplemented with 0.5% Glucose, GlutaMAX (Thermo Fischer Scientific) and B27 (Thermo Fischer Scientific). Luciferase assays were performed 48h after transfection using Dual Luciferase Reporter Assay System (Promega). Statistical differences between experimental groups were determined with Chi-square.

Western blot analysis

MGE tissue from one litter e13.5 and e17.5 MGE was harvested and homogenized in RIPA buffer (150 mM NaCl, 1% NP-40, 0.1% SDS, 0.5% sodium deoxycholate in 50 mM Tris) with protease inhibitor (Complete EDTA-free, Roch) and the total protein was separated in SDS-PAGE gels and transferred to nitrocellulose membranes. Western blotting was performed by standard procedures. Antibodies used: NKX2-1 polyclonal (Santa Cruz, sc-13040) and LHX6 polyclonal antibody (Genscript)

Differential Expression

Sample preparation, labeling, and array hybridizations were performed according to standard protocols from the UCSF Shared Microarray Core Facilities and Agilent Technologies (http://www.arrays.ucsf.edu and http://www.agilent.com). Total RNA quality was assessed using a Pico Chip on an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA). RNA was amplified and labeled with Cy3-CTP using the Agilent low RNA input fluorescent linear amplification kits following the manufacturers protocol (Agilent). Labeled cRNA was assessed using the Nanodrop ND-100 (Nanodrop Technologies, Inc., Wilmington DE), and equal amounts of Cy3 labeled target were hybridized to Agilent whole mouse genome 4x44K Ink-jet arrays (Agilent). Hybridizations were performed for 14 hrs, according to the manufacturers protocol (Agilent). Arrays were scanned using the Agilent microarray scanner (Agilent) and raw signal intensities were extracted with Feature Extraction v9.1 software (Agilent). Datasets were normalized using the *quantile* normalization method as earlier described(Bolstad et al., 2003). Hybridization performance was assessed by *Agilent's Feature Extraction 10.1*. For each array, a diagnostic pdf summarizing several diagnostic plots and some summary statistics was produced, and all arrays passed QC analysis. No background subtraction was performed, and the median feature pixel intensity was used as the raw signal before normalization. For NKX2-1, four replicates of e13.5 *Nkx2-1*cKO or WT MGE were used. After normalization, Log2 fold change was calculated from the average values across replicates where available. The probe with the highest intensity value was selected for each gene. Fold changes of 1.2 or greater were used to define differentially expressed genes for gene set analysis. Fold changes of 1.22 or greater were used in genomic binding analysis. Heatmaps show log2 fold change and normalized intensity, with genes clustered based on Gaussian expression distributions as estimated using Mclust (http://www.stat.washington.edu/mclust/). Gene set enrichment performed using DAVID(Huang et al., 2009) with all mouse genes as background, terms shown selected based on Benjamini-Hochberg corrected significance $P<0.05$ and non-redundancy.

ChIP-seq Computational Analysis

For ChIP-seq datasets, clustering, base calling and quality metrics were performed using standard Illumina software. FASTQ files for ChIP, Negative control, and Input libraries were analyzed for overall quality and then reads were filtered for adaptor sequence and by quality and sequence reads were mapped to the mouse genome (mm9) using BWA(Li and Durbin, 2009), with the run command:

```
bwa aln -t 5 -l 25 mm9 lib.fastq
```

Peak calling on aligned bam files was performed using Macs2 using settings with model building omitted (Feng et al., 2011). Comparison of peak calling with model versus without suggested that the no model settings performed better for our datasets, including for NKX2-1 and LHX6. All peak calling was performed using both Negative binding control and input DNA control. All datasets included in this work showed no enrichment in the Negative binding control or in the Input DNA control. For peaks analyzed further, we used the ChIP versus Input control analysis. Example Macs2 commands:

Peak Calling

```
macs2 callpeak -t Nkx2.1.merged.ChIP.bam -c Nkx2.1.merged.Input.bam -f BAM -g mm -n
Nkx2.1_ChIPvsInput_p0.00001 --outdir Nkx2.1_ChIPvsInput_p0.00001 -p 0.00001 --nolambda --
nomodel --extsize=250 --call-summits –bdg
```

Differential enrichment

NKX2-1 vs. LHX6

```
macs2 bdgdiff --t1 Nkx2.1_ChIPvsInput_p0.00001_treat_pileup.bdg --c1
Nkx2.1_ChIPvsInput_p0.00001_control_lambda.bdg --t2
Lhx6_ChIPvsInput_p0.00001_treat_pileup.bdg --c2
Lhx6_ChIPvsInput_p0.00001_control_lambda.bdg --d1 89956032 --d2 50979153 -l 150 --outdir
Nkx2.1_vs_Lhx6_differential --o-prefix Nkx2.1_vs_Lhx6_differential
```

WT vs cKO Histone

```
macs2 bdgdiff --t1 H3K27ac_WT_ChIPvsInput_p0.00001_treat_pileup.bdg --c1
H3K27ac_WT_ChIPvsInput_p0.00001_control_lambda.bdg --t2
H3K27ac_KO_ChIPvsInput_p0.00001_treat_pileup.bdg --c2
H3K27ac_KO_ChIPvsInput_p0.00001_control_lambda.bdg --d1 59772579 --d2 60128750 -l 150 --
outdir H3K27ac_differential --o-prefix H3K27ac_differential
```

Genome wide coverage generated from aligned bam files using custom scripts, with reads extended to 300bp and coverage estimates were generated in 25bp windows across the genome. Coverage and heatmap diagrams were produced using the ngs.plot.r package (Shen et al., 2014), with coverage normalized against input and for overall library size. Peaks where a significant proportion of reads originated in repetitive sequence were omitted from further analysis. For NKX2-1 and LHX6, merged peak sets were generated across the three replicates. Peaks were annotated to genomic features, including RefSeq and UCSC gene builds and evolutionarily conserved elements (Phastcons, GERP), as well as to nearest gene represented in the *Nkx2-1*cKO microarray dataset. We filtered out peaks that overlapped with alignment gaps, highly repetitive regions, and ENCODE blacklist peaks. For binding to expression analysis, microarray expression estimates were used for all TSS peaks that were also present in the microarray, as well as for peaks with a represented microarray gene within 100kb. Functional enrichment was performed using GREAT with genomic background and default settings. Images of genomic coverage and peaks calls were generated via uploading our datasets to the UCSC genome browser and using UCSC internal tools for image capture and data extraction(Dreszer et al., 2012). For plotting, ChIP-seq replicates bam files were merged to produce combined coverage values. Motif analysis performed using the HOMER package (http://homer.salk.edu/homer/)(Heinz et al., 2010), with default settings and genomic background.

Example HOMER command:

```
findMotifsGenome.pl Nkx2.1.all.bed mm9 Nkx2.1.All -size -250,250 -mask
```

De novo motifs discovered using the full NKX2-1 and LHX6 ChIP-seq peak sets were then used as the motif set for analysis of peak subsets to identify motifs enriched at certain features (e.g. TSSs) or in specific peak sets (e.g. peaks that show chromatin change in the *Nkx2-1cKO*). De novo motifs were annotated against the internal HOMER database to identify potential transcription factors. All peaks were annotated with motif occurrences using HOMER.

Classification of chromatin state for NKX2-1 and LHX6 peaks performed based on presence of overlapping or flanking histone modification peak (H3K4me1, H3K4me3, H3K27ac, H3K27me3) in datasets from wild-type stage and region matched MGE. Histone ChIP-seq datasets processed as described above. A flanking/overlapping peak was classified as present if merged replicate analysis of histone ChIP-seq datasets had a peak called that reached a significance level of $P < 0.00001$ in the Macs2 analysis. To

identify NKX2-1-bound REs where chromatin state was sensitive to change in the *Nkx2-1cKO*, we performed ChIP-seq on cKO e13.5 MGE, processing the datasets as described above. After screening to ensure no bias in Negative control or Input datasets, we compared differential ChIP enrichment by running the WT versus the cKO using Macs2 as described above with a cutoff of $P < 0.00001$. We performed both WT vs. cKO and cKO vs. WT for each histone mark. Custom R scripts were used for computational analysis and are available at request.

Supplemental References

Bolstad, B.M., Irizarry, R.A., Astrand, M., Speed, T.P., 2003. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 19, 185–193.

Dreszer, T.R., Karolchik, D., Zweig, A.S., Hinrichs, A.S., Raney, B.J., Kuhn, R.M., Meyer, L.R., Wong, M., Sloan, C.A., Rosenbloom, K.R., Roe, G., Rhead, B., Pohl, A., Malladi, V.S., Li, C.H., Learned, K., Kirkup, V., Hsu, F., Harte, R.A., Guruvadoo, L., Goldman, M., Giardine, B.M., Fujita, P.A., Diekhans, M., Cline, M.S., Clawson, H., Barber, G.P., Haussler, D., James Kent, W., 2012. The UCSC Genome Browser database: extensions and updates 2011. Nucleic Acids Res. 40, D918–23. doi:10.1093/nar/gkr1055

Feng, J., Liu, T., Zhang, Y., 2011. Using MACS to identify peaks from ChIP-Seq data. Curr Protoc Bioinformatics Chapter 2, Unit 2.14–2.14.14. doi:10.1002/0471250953.bi0214s34

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., Glass, C.K., 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Molecular Cell 38, 576–589. doi:10.1016/j.molcel.2010.05.004

Huang, D.W., Sherman, B.T., Lempicki, R.A., 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4, 44–57. doi:10.1038/nprot.2008.211

Kothary, R., Clapoff, S., Darling, S., Perry, M.D., Moran, L.A., Rossant, J., 1989. Inducible expression of an hsp68-lacZ hybrid gene in transgenic mice. Development 105, 707–714.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760. doi:10.1093/bioinformatics/btp324

Poulin, F., Nobrega, M.A., Plajzer-Frick, I., Holt, A., Afzal, V., Rubin, E.M., Pennacchio, L.A., 2005. In vivo characterization of a vertebrate ultraconserved enhancer. Genomics 85, 774–781. doi:10.1016/j.ygeno.2005.03.003

Shen, L., Shao, N., Liu, X., Nestler, E., 2014. ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. BMC Genomics 15, 284. doi:10.1186/1471-2164-15-284