LISTS OF SUPPLEMENTAL FIGURES AND TABLES

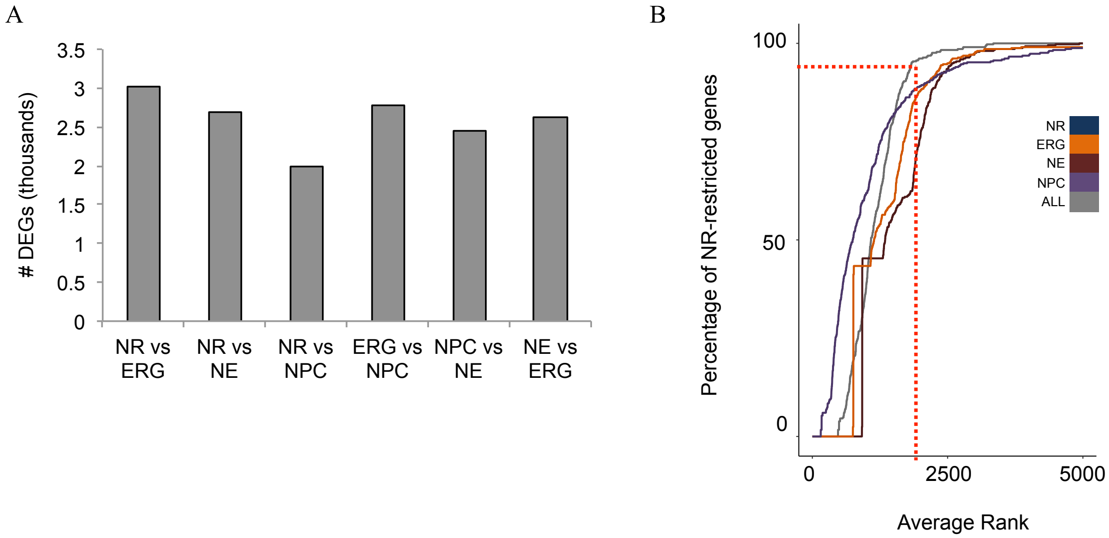

Figure S1. Distinct neural rosette transcriptional profiles compared to 2D neural stem cells. Related to Figure 1. A, Bar plot showing the number of differentially expressed genes (DEGs) in the pairwise comparisons among all cell types. Number of DEGs is expressed as thousands. See also Table S1B. B, Comparison between NR-restricted genes defined using entropy-based method and DEGs as measured by Rank Product. Red dashed line indicates 96% NR-restricted genes recovered in the top 10% for Rank Product for NR vs all 2D NSCs. See also Table S1C.

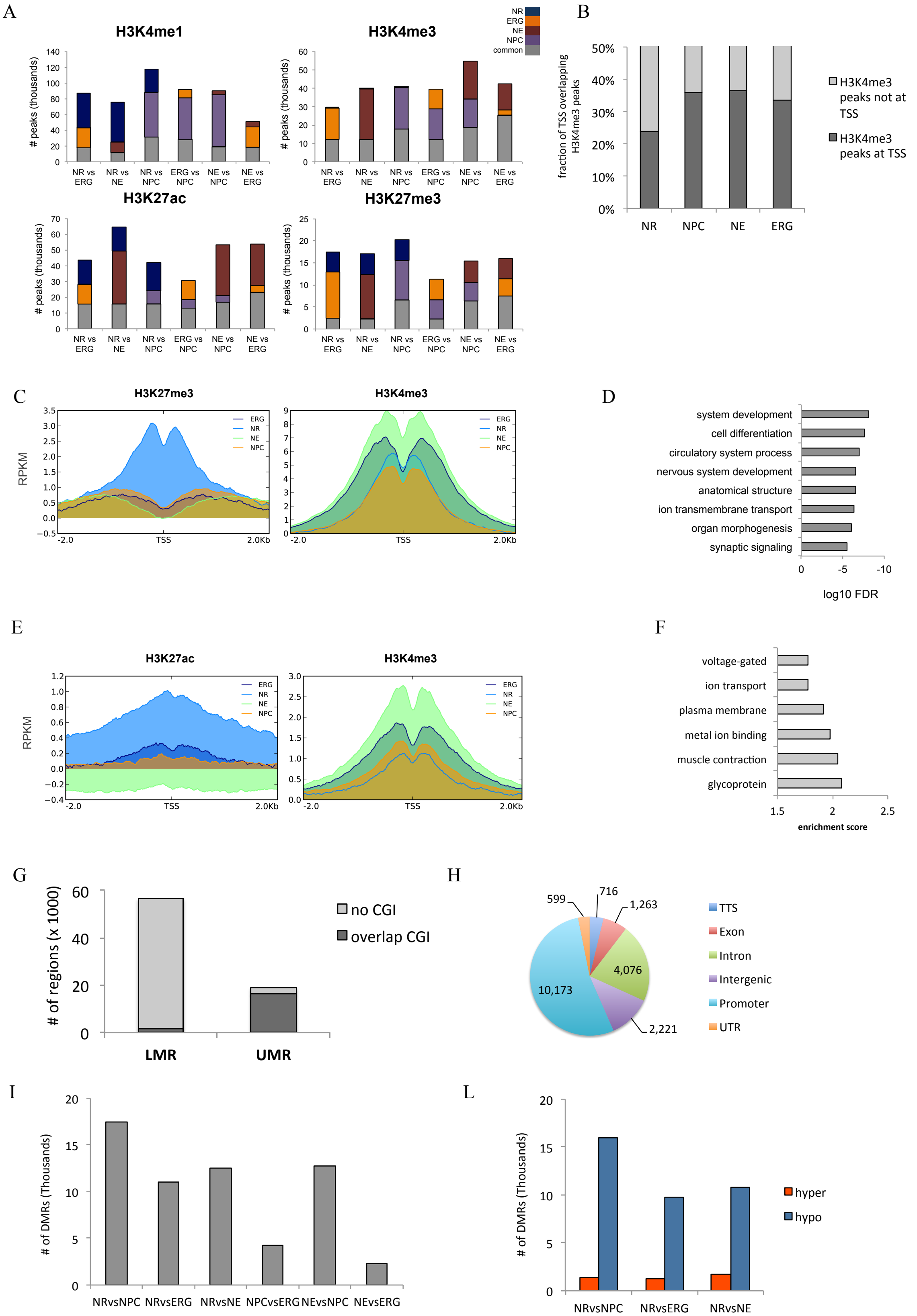
A
H3K4me1
# peaks (thousands)
140
120
100
80
60
40
20
0
NR vs ERG
NR vs NE
NR vs NPC
ERG vs NPC
NE vs NPC
NE vs ERG
H3K4me3
# peaks (thousands)
60
50
40
30
20
10
0
NR
ERG
NE
NPC
common
H3K27ac
# peaks (thousands)
70
60
50
40
30
20
10
0
H3K27me3
# peaks (thousands)
25
20
15
10
5
0
B
fraction of TSS overlapping H3K4me3 peaks
50%
40%
30%
20%
10%
0%
NR
NPC
NE
ERG
H3K4me3 peaks not at TSS
H3K4me3 peaks at TSS
C
H3K27me3
H3K4me3
RPKM
ERG
NR
NE
NPC
-2.0
TSS
2.0Kb
D
system development
cell differentiation
circulatory system process
nervous system development
anatomical structure
ion transmembrane transport
organ morphogenesis
synaptic signaling
0
-5
-10
log10 FDR
E
H3K27ac
H3K4me3
RPKM
F
voltage-gated
ion transport
plasma membrane
metal ion binding
muscle contraction
glycoprotein
1.5
2
2.5
enrichment score
G
# of regions (x 1000)
60
40
20
0
LMR
UMR
no CGI
overlap CGI
H
599
716
1,263
4,076
10,173
2,221
TTS
Exon
Intron
Intergenic
Promoter
UTR
I
# of DMRs (Thousands)
20
15
10
5
0
NRvsNPC
NRvsERG
NRvsNE
NPCvsERG
NEvsNPC
NEvsERG
L
# of DMRs (Thousands)
20
15
10
5
0
NRvsNPC
NRvsERG
NRvsNE
hyper
hypo

Figure S2. Epigenomic landscapes at neural rosette promoters. Related to Figure 2. A, Bar plots show the overlap between peaks for each pairwise comparisons. Number of peaks is expressed as thousands. B, Bar chart shows the fraction (expressed as percentage of the total) of TSS marked by H3K4me3 per each cell types. C, Profile plots of the averaged ChIP-seq signal intensity (RPKM input normalized) over all NR-unique H3K37ac peaks that overlapped promoter regions (defined as +/- 1kb around TSS - 2,780 peaks). D, Bar plot shows the most significantly enriched GO terms associated with genes whose promoters are shown in C. E, Profile plots of the averaged ChIP-seq signal intensity (RPKM input normalized) over all NR-unique bivalent promoters (H3K4me3/H3K27me3 regions). F, Bar plot shows the most significantly enriched cluster of GO terms associated with genes having bivalent promoters - shown in E. G, Bar plot showing the number of UMRs and LMRs identified in NRs and their overlap with CG islands (CGIs), as defined by UCSC. H, Pie chart shows distribution of UMRs among genomic features. I, Bar plot showing the number of DMRs or each pairwise comparisons. L, Bar plot showing the number of hypomethylated and hypermethylated DMRs in pairwise comparisons between NRs and each of the 2D NSCs. Number of peaks is expressed as thousands.

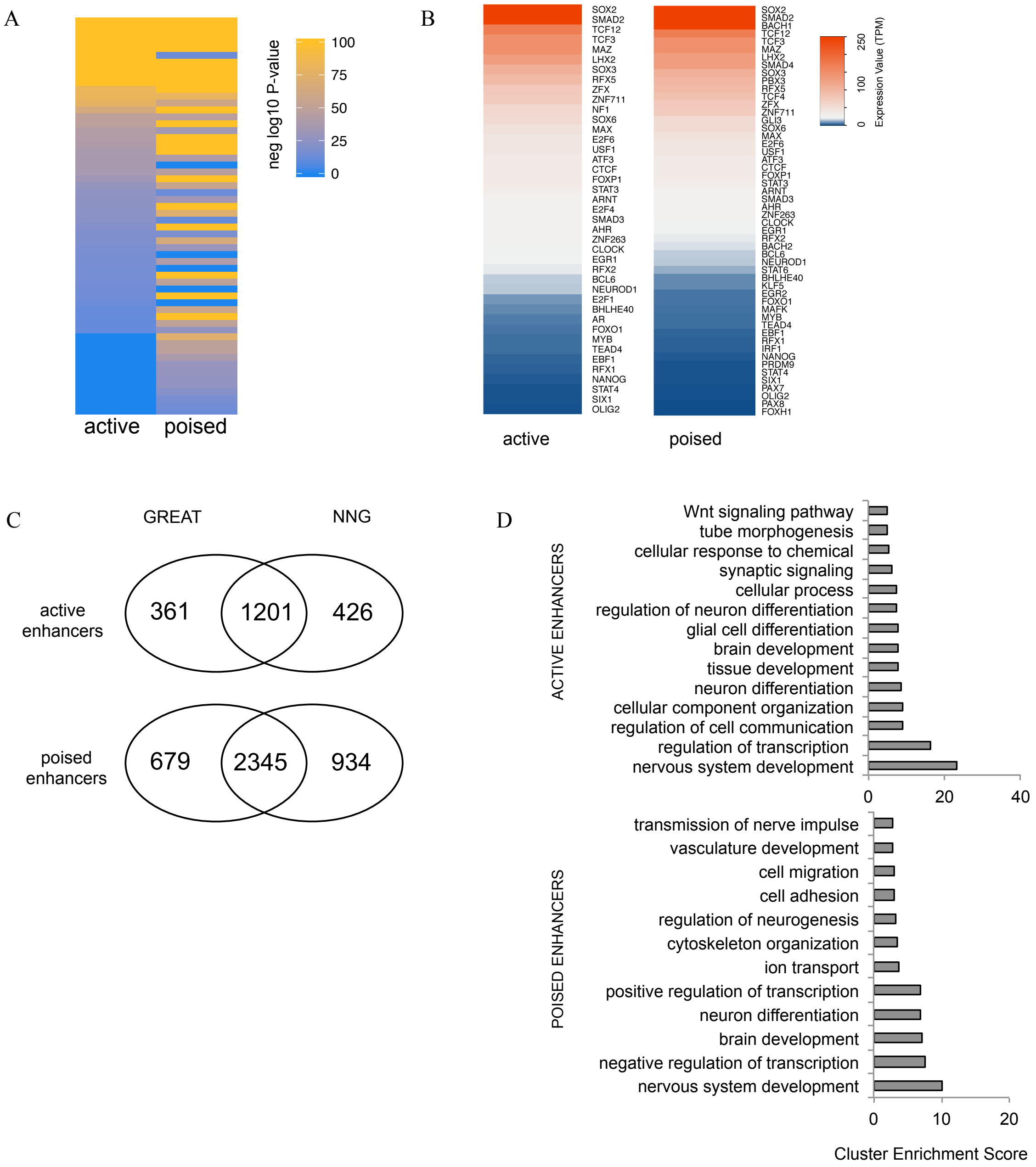

Figure S3. NR regulatory enhancer networks and dynamics during nervous system development. Related to Figure 3. A, Heatmap representation of negative log10 p-value for motif enrichment at enhancer regions. Only motifs for TF expressed (TPM>1) in NR are shown. See Table S3B for complete list. B, Heatmap representations of expression levels (TPM) of genes coding for TFs whose motif were found significantly enriched at active (left) and poised (right) enhancers. Only TFs whose gene was detected by RNA-seq are shown. C, Venn diagram shows overlap between putative target genes assigned using NNG and GREAT method. D, Bar plots show the most significantly enriched clusters of GO terms associated with putative target genes assigned by GREAT for enhancers shown in Cluster 1 of Figure 3D.

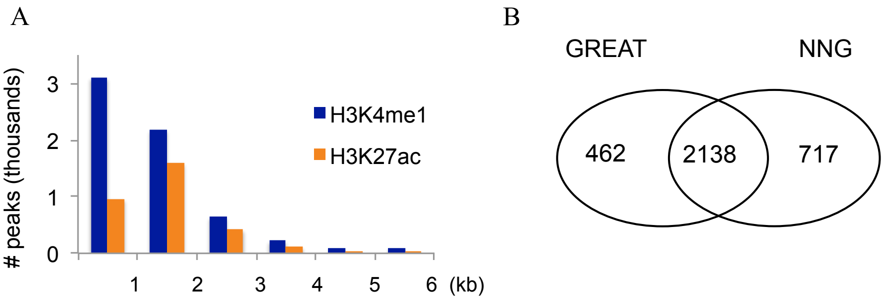

Figure S4. Broad distal regulatory domains reveal master regulators of neural tube formation. Related to Figure 4. A, Histogram plot shows the distribution of peak length for both H3K4me1 and H3K27ac. Number of peaks per length range is expressed as thousands. B, Venn diagram shows overlap between putative target genes assigned using NNG and GREAT method. See Table S4 for target genes.

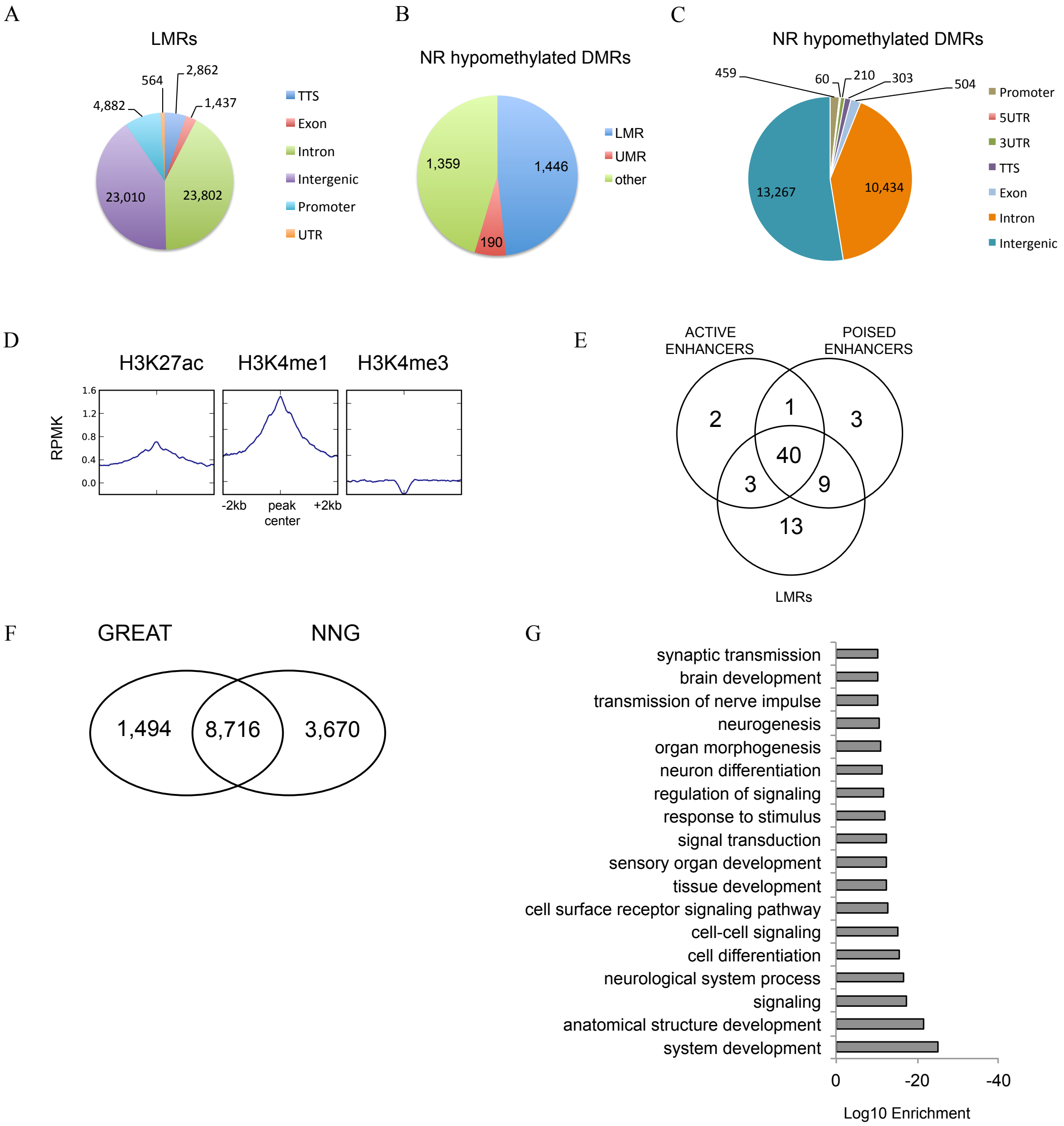

Figure S5. LMR predictions of distal regulatory domains and their developmental dynamics. Related to Figure 5. A, Pie chart shows distribution of LMRs among genomic features. B, Pie chart shows overlap between NR hypomethylated DMRs (2,995 regions common across all three pairwise comparisons against 2D NSCs) and UMRs and LMRs. C, Pie chart shows distribution among genomic features of hypomethylated DMRs in any of three pairwise comparisons between NRs and 2D NSCs (25,237 regions). D, Profile plots of the averaged ChIP-seq signal intensity (RPKM input normalized) over hypomethylated DMRs in any of three pairwise comparisons between NRs and 2D NSCs (25,237 regions). E, Venn diagram of overlap among significantly (q-value < 0.01) TF binding site motifs at active enhancers, poised enhancers and unmarked LMRs. Only TFs detected by RNA-seq are shown. F, Venn diagram shows overlap between putative target genes of unmarked LMRs assigned using NNG and GREAT method. See Table S5 for target genes. G, Bar plots show the most significantly GO terms associated with putative target genes assigned by GREAT for unmarked LMRs.

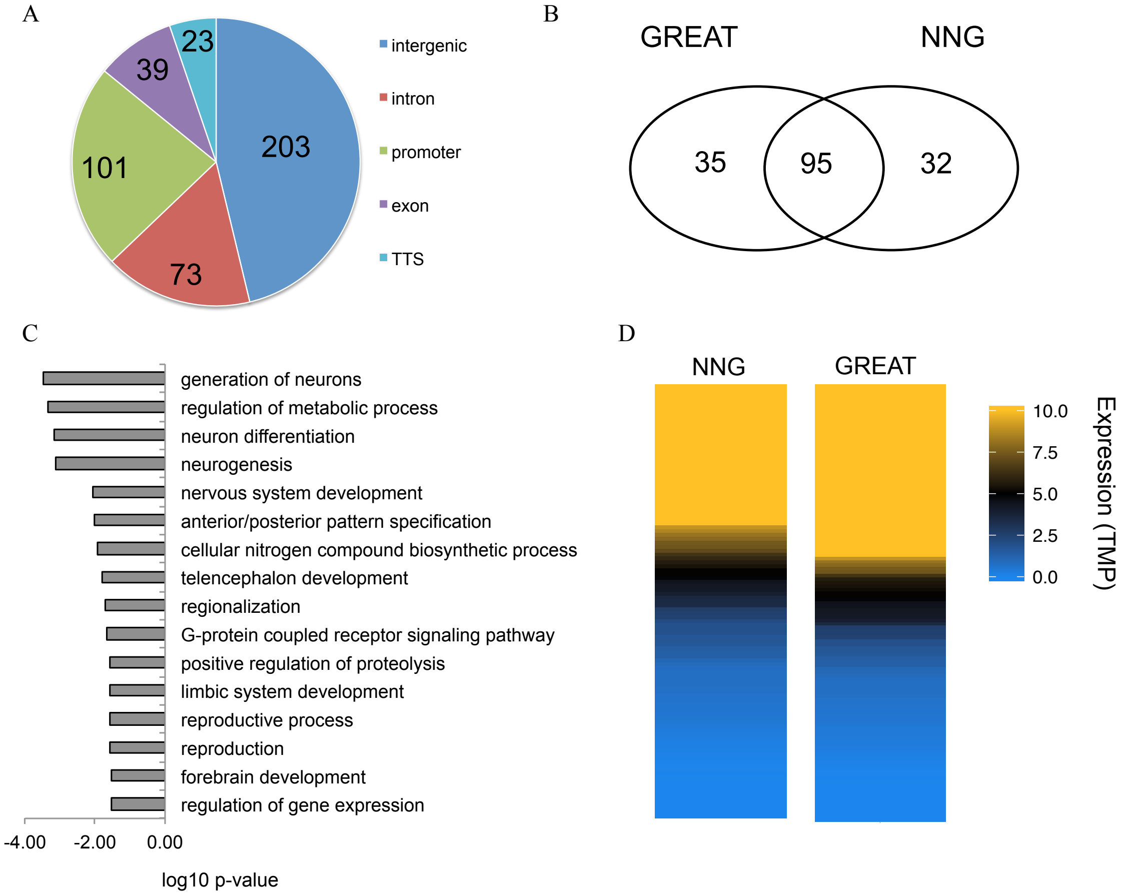

Figure S6. Folate-associated CpGs overlap NR regulatory elements. Related to Figure 6. A, Pie chart shows distribution of the 443 folate-associated CpGs identified by Joubert et al., 2014 among genomic features. B, Venn diagram shows unique and shared putative target genes assigned by NNG and GREAT methods. See Table S6 for target genes. C, Bar chart shows most significantly enriched GO terms for GREAT putative targets. D, Heatmap representation of the gene expression values (TPM) of the putative target genes of folate-associated CpGs at NR regulatory elements.

Table S1. Cell-type restricted and differentially expressed genes. Related to Figure 1.
Table S2. ChIP-seq datasets comparisons. Related to Figure 2.
Table S3. TFBS motifs and NNGs of NR enhancers. Related to Figure 3.
Table S4. TFBS motifs and NNGs of NR broad enhancers. Related to Figure 4.
Table S5. TFBS motifs for unmarked LMRs. Related to Figure 5.
Table S6. NNGs of the Folate-associated CpGs at NR regulatory elements. Related to Figure 6.
Table S7. Traits and conditions associated SNPs enriched at NR enhancers. Related to Figure 7.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

hESC culture and neural differentiation

hES cells (WA09, WiCell) were cultured on Matrigel (BD) in mTERS1 (Stemcell Technologies) according to manufacturer's protocol. NRs were differentiated in STEMdiff™ Neural Induction Medium (Stemcell Technologies) according to manufacturer's manual. Briefly, hESCs were dissociated with Accutase and embryonic bodies (EB) were formed using AggreWell™ 800 plates (Stemcell Technologies). 10,000 hES cells were aggregated per EB. These plates have V-bottom microwells (approximately 300 per well) that allow for control the number of cells that aggregate per EB and force cells to aggregate to a homogenous size and spherical shape. This approach allowed us to have highly uniform EB formation. EBs were cultured for 4 days in AggreWell plates and then transferred to Poly-Ornhitine (Sigma) and Laminin (Sigma) coated plates, where EBs where culture in adhesion for additional 7 days. Within 48 hours, 3D tubular-structures (neural rosettes, NRs) covered the surface of EBs. NRs were harvested at day 12 of differentiation preserving the 3D structure by incubating in STEMdiff™ Neural Rosette Selection Reagent (Stemcell Technologies) for 30-50 minutes, then manually dislodging the tubular structures.

Immunocytochemistry

NRs at day 12 of differentiation were fixed in 4% paraformaldehyde for 30 minutes at room temperature and then blocked in 0.5% BSA for 1 hour at room temperature. Immunostaining was performed in blocking buffer at 4ºC overnight. Primary antibodies used: chicken anti-PAX6 (DSHB), rabbit anti-PAX6 (Covance), mouse anti-ZO-1 (BD), mouse anti-NESTIN (R&D), rabbit anti-NESTIN (Abcam), mouse anti-SOX2 (R&D), rabbit anti-SOX2 (Cell Signaling), goat anti-SOX1 (Santa Cruz), rabbit anti-SOX1 (Cell Signaling). Appropriate Alexa Fluor antibodies were used as secondary. Fluorescent images were collected using a Leica TCS or Zeiss LSM confocal microscopes.

RNA-sequencing (RNA-seq) library construction, sequencing and data analysis

RNA was purified using AllPrep DNA/RNA Mini Kit (Qiagen), treated with RNease-free DNase 1 (NEB) and rRNA removed using Ribo-Zero Gold Kit (Epicentre). RNA-seq was performed using ScriptSeq™ v2 RNA-Seq Library Preparation Kit (Epicentre) according to manufacturer's protocol to generate whole transcriptome strand-specific libraries. Libraries were sequenced in pair-end 100 cycles run on HiSeq2500.

Duplicates from two independent experiments were performed. After testing they were highly correlated, duplicates were merged for subsequent analyses.

Abundances of RNA-Seq reads were quantified by Kallisto (Bray et al., 2016) using Gencode reference transcriptome v19 and expression levels were quantified as transcripts per million (TPM). Cell-specific expression was determined using an approach in which Shannon entropy is used to rank genes from having equal expression across all groups to having significant expression in one group only, as described in (Schug et al., 2005). Briefly, in this method the relative expression of a gene g in a tissue t from N tissue groups is defined as $p_{t|g} = w_{g,t}/\sum 1 \leq t \leq N w_{g,t}$ where $w_{g,t}$ is the expression level of the gene in the tissue. The entropy of a gene's expression distribution is defined as $H_g = \sum 1 \leq t \leq N - p_{t|g} \log2(p_{t|g})$. H_g is zero for genes expressed only in one tissue and log2(N) for genes equally in all tissue groups. For the expression of each gene in each tissue a specificity score is defined by $Q_{g|t} = Hg - \log2(p_{t|g})$. Low values of $Q_{g|t}$ indicate tissue specific expression. We used a cutoff of $Q_{g|t} < 1.5$ to define cell specific expression.

Alternative pairwise differential expression analysis of RNA-seq was performed with RankProd R package (Carratore FD, AJFH, Wittner B, Breitling R and Battke F - RankProd: Rank Product method for identifying differentially expressed genes with application in meta-analysis. R package version 3.2.0.).

Chromatin Immunoprecipitation Sequencing (ChIP-Seq) library construction, sequencing and data analysis

ChIP-seq was performed as previously described (Hawkins et al., 2010) with minor modifications. Briefly, cells were cross-linked for 10 min a room temperature in 1% formaldehyde. After quenching and washing, cells were lysed and nuclei sonicated for 10-12 minutes using Covaris S2 (Covaris). 20 µg of chromatin were incubated O/N at 4C with 11 µl of Dynabeads (Invitrogen) coated with 3 µg of the following antibodies: anti-H3K4me1 (Diagenode, C15410037), anti-H3K4me3, anti-H3K27ac and anti-H3K27me3 (Active Motif, 39159, 39133 and 39155, respectively). Bead-antibody-chromatin complexes were washed in RIPA buffer with 500 mM LiCl for 8 times, chromatin eluted in TE supplemented with 1% SDS and 300 mM NaCl for 30 minutes at 65C and then reverse cross-linked O/N 65C. Proteinase K and RNase treatments were performed afterward and DNA purified using phenol:chloroform:isoamyl alcohol followed by absolute ethanol precipitation. DNA was end repaired using End Repair Kit (NEB), dA was ligated at 3' using Klenow (NEB) and Illumina-compatible barcoded adapters ligated using T4 ligase (NEB). AMPure Beads (Beckman Culture) were used for reaction clean-up and to size-select ligated DNA so that fragments between 300bp and 700bp were selected for be amplified in PCR. 10-14 cycles of PCR were performed using KAPA HiFi HotStart ReadyMix PCR Kit (Kapa Biosystems). Clean-up was performed using AMPure Beads. Final libraries were quantified using dsDNA HS Assay (Invitrogen) and average size determined using High Sensitivity DNA Kit (Agilent). Libraries were sequenced on single-end 50 or 75 cycles on HiSeq2500 or NextSeq550 (Illumina). As control, a library was constructed for each input chromatin. For each histone modifications, experiments were performed in duplicates.

Sequencing raw reads were trimmed for low quality (q score < 30) and adapters, then mapped to human genome (hg19) using Bowtie2 (Ben Langmead and Salzberg, 2012). After testing they were highly correlated, duplicates were merged for subsequent analyses. Peaks were called on merged duplicates using MACSv1.4 using the --nomodel mode (Zhang et al., 2008). Peaks overlap was performed using Homer suite (mergePeaks -d given) (Heinz et al., 2010). ChIP-seq signals were expressed as per kilobase per million reads (RPKM) and normalized by subtraction of input using deepToolsv v2.3 suite (bamCompare, --normalizeUsingRPKM --ratio subtract --ignoreDuplicates). k-means clustering analysis was performed using deepTools v2.3 suite (plotHeatmap --kmeans). Peak annotation to human genome (hg19) and GO term enrichment analysis of the associated nearest genes was performed using Homer suite (annotatePeaks.pl hg19 -go). Transcription Factor Binding Site (TFBS) motif enrichment analysis was performed using Homer suite (findMotifsGenome.pl -size given -nomotif). For identifying motifs enriched at broad enhancers with respect to standard enhancers, all enhancer regions smaller than 3kb were used as background. TFs network construction, first we collected the protein-protein interactions from BioGRID (Chatr-Aryamontri et al., 2015), HPRD, STRING (Szklarczyk et al., 2015) and irefR (Mora and Donaldson, 2011) databases and built a unique network after removing redundant interactions and self-loops. Then, we retrieved the interactions between enriched TFs from this unique network. The network visualizations were created using cytoscape software (Shannon et al., 2003).

Whole Genome Bisulfite Sequencing (WGBS) library construction, sequencing and data analysis

WGBS was performed as previously described (Lister et al., 2008) with minor modifications and using 1 μg - 500 ng of genomic DNA per library. Briefly, genomic DNA was purified using QIAMP DNA Mini Kit (Qiagen) and fragmented using Covaris S2 (Covaris). Lambda DNA (unmethylated DNA) was used as control for determine bisulfite conversion efficiency and was spiked in at 0.5%. DNA was end repaired using End Repair Kit (NEB), dA was ligated at 3' using Klenow (NEB) and Illumina compatible, methylated barcoded adapters ligated using T4 ligase (NEB). AMPure Beads (Beckman Culture) were used for reaction clean-up and to size-select ligated DNA. Bisulfite treatment was performed using MethylCode Bisulfite Conversion Kit and DNA was amplified for 4-8 cycles using Kapa HiFi Uracil (Kapa Biosystems). Clean-up was performed using AMPure Beads. Final libraries were quantified using dsDNA HS Assay (Invitrogen) and average size determined using High Sensitivity DNA Kit (Agilent). Libraries were sequenced on HiSeq2500 in paired-end 100 cycles runs. All experiments were performed in duplicates.

Raw reads were adapter trimmed using the bbduk algorithm in BBMap (https://sourceforge.net/projects/bbmap/) and aligned to human genome hg19 and methylation quantified using BSeeker2 (Guo et al., 2013) and Bowtie2 (Ben Langmead and Salzberg, 2012) using the end-to-end mode. To identify differentially methylated regions (DMRs), CG dinucleotide data for each strand were merged with differential regions identified using DSS (Wu et al., 2013) with detection p-values <0.01, delta-mC > 0.2, number of CG's >= 3, window size > 50 and merging of DMRs with < 100 bases between them. Samples with replicates were merged before calling DMRs. DMRs were calculated for all pairwise comparisons with the union of all DMRs being used for downstream comparative analyses. To visualize differences in mC between cell types, mCG/CG was calculated for the union set of all DMRs that had coverage in all samples. Unmethylated and low methylated regions (UMRs and LMRs) were identified using MethylSeekR (Bulger and Groudine, 2011) with FDR < 5% for regions with < 10% (UMRs) or < 50% (LMRs) mCG/CG and > 5 CGs per region. CpG islands (CGIs) coordinates were obtained for human genome hg19 form UCSC (http://hgdownload.soe.ucsc.edu/goldenPath/hg19/database/cpgIslandExt.txt.gz). CGIs (28,691 regions) were defined as regions > 200bp with fraction of CpG > 11.7%. Region annotation to human genome (hg19) and GO term enrichment analysis of the associated nearest genes was performed using Homer suite (annotatePeaks.pl hg19 -go). Comparing of UMR/LMRs with promoter and enhancer maps was performed using bedtools/2.24.0 suite (Quinlan & Hall, 2010).

Genetic variants analysis

Enhancers overlapping SNPs and DHS were identified using bedtools intersect. Probability of SNP/enhancer overlaps was found using Fisher's exact test. Bar plot showing overlaps of NR enhancers with neural related SNPs and DHS was made with the upsetR R package (Lex AR, et. al, 2014). Alluvial plots for overlapping SNPs were made with the R package alluvial (https://github.com/mbojan/alluvial).

Publically available data used in this study

DHS data for human fetal brain (18 weeks gestation) and fetal spinal cord (15 week gestation) were obtained from the Roadmap Epigenome Project, accession number GSM595913 and GSM878661, respectively. Raw sequencing reads were obtained and data were processed using same pipeline used for ChIP-seq data analysis.

Folate-associated CpGs coordinates were obtained from Joubert et al., 2016. Annotation to genomic features (hg19) was performed using Homer suite. GO term enrichment analysis was performed using DAVID (Da Wei Huang et al., 2009).

Raw data of RNA-seq, ChIP-seq and WGBS for NPC was obtained from Xie et al., 2013, and for NE and ERG from Ziller et al., 2014.

SNPs for all available traits were downloaded from the NHGRI-EGI GWAS Catalog (https://www.ebi.ac.uk/gwas/). SNAP (https://www.broadinstitute.org/mpg/snap/ldsearchpw.php) was used to collect SNPs in linkage disequilibrium (r2 threshold of 0.8, distance limit 500). For each condition, SNPs inside of the genomic coordinates of NR enhancers were identified. For each condition, p-values for the likelihood of the number of SNPs found overlapping enhancers were calculated using the hypergeometric test and adjusted for multiple test with the Bonferroni correction.

SUPPLEMENTAL REFERENCES

Ben Langmead, Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat Meth 9, 357–359. doi:10.1038/nmeth.1923

Bray, N.L., Pimentel, H., Melsted, P., Pachter, L., 2016. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol 34, 525–527. doi:10.1038/nbt.3519

Bulger, M., Groudine, M., 2011. Functional and mechanistic diversity of distal transcription enhancers. Cell 144, 327–339. doi:10.1016/j.cell.2011.01.024

Chatr-Aryamontri, A., Breitkreutz, B.-J., Oughtred, R., Boucher, L., Heinicke, S., Chen, D., Stark, C., Breitkreutz, A., Kolas, N., O'Donnell, L., Reguly, T., Nixon, J., Ramage, L., Winter, A., Sellam, A., Chang, C., Hirschman, J., Theesfeld, C., Rust, J., Livstone, M.S., Dolinski, K., Tyers, M., 2015. The BioGRID interaction database: 2015 update. Nucleic Acids Res 43, D470–D478. doi:10.1093/nar/gku1204

Da Wei Huang, Sherman, B.T., Lempicki, R.A., 2009. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 37, 1–13. doi:10.1093/nar/gkn923

Guo, W., Fiziev, P., Yan, W., Cokus, S., Sun, X., Zhang, M.Q., Chen, P.-Y., Pellegrini, M., 2013. BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. BMC Genomics 14, 774–774. doi:10.1186/1471-2164-14-774

Hawkins, R.D., Hon, G.C., Lee, L.K., Ngo, Q., Lister, R., Pelizzola, M., Edsall, L.E., Kuan, S., Luu, Y., Klugman, S., Antosiewicz-Bourget, J., Ye, Z., Espinoza, C., Agarwahl, S., Shen, L., Ruotti, V., Wang, W., Stewart, R., Thomson, J.A., Ecker, J.R., Ren, B., 2010. Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. Cell Stem Cell 6, 479–491. doi:10.1016/j.stem.2010.03.018

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. Mol. Cell 38, 576–589. doi:10.1016/j.molcel.2010.05.004

Joubert, B.R., Dekker, den, H.T., Felix, J.F., Bohlin, J., Ligthart, S., Beckett, E., Tiemeier, H., van Meurs, J.B., Uitterlinden, A.G., Hofman, A., Håberg, S.E., Reese, S.E., Peters, M.J., Andreassen, B.K., Steegers, E.A.P., Nilsen, R.M., Vollset, S.E., Midttun, Ø., Ueland, P.M., Franco, O.H., Dehghan, A., de Jongste, J.C., Wu, M.C., Wang, T., Peddada, S.D., Jaddoe, V.W.V., Nystad, W., Duijts, L., London, S.J., 2016. Maternal plasma folate impacts differential DNA methylation in an epigenome-wide meta-analysis of newborns. Nature Communications 7, 10577–10577. doi:10.1038/ncomms10577

Mora, A., Donaldson, I.M., 2011. iRefR: an R package to manipulate the iRefIndex consolidated protein interaction database. BMC Bioinformatics 12, 455–455. doi:10.1186/1471-2105-12-455

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics, 26(6), 841–842. http://doi.org/10.1093/bioinformatics/btq033

Schug, J., Schuller, W.-P., Kappen, C., Salbaum, J.M., Bucan, M., Stoeckert, C.J., 2005. Promoter features related to tissue specificity as measured by Shannon entropy. Genome Biology 6, R33–R33. doi:10.1186/gb-2005-6-4-r33

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Research 13, 2498–2504. doi:10.1101/gr.1239303

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., Kuhn, M., Bork, P., Jensen, L.J., Mering, von, C., 2015. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 43, D447–D452. doi:10.1093/nar/gku1003

Wu, L., Wang, L., Shangguan, S., Chang, S., Wang, Z., Lu, X., Zhang, Q., Wang, J., Zhao, H., Wang, F., Guo, J., Niu, B., Guo, J., Zhang, T., 2013. Altered methylation of IGF2 DMR0 is associated with neural tube defects. Mol Cell Biochem 380, 33–42. doi:10.1007/s11010-013-1655-1

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W., Liu, X.S., 2008. Model-based Analysis of ChIP-Seq (MACS). Genome Biology 9, R137. doi:10.1186/gb-2008-9-9-r137