Supplementary Material

Evolutionary loss of genomic proximity to conserved noncoding elements impacted the gene expression dynamics during mammalian brain development

Meenakshi Bagadia[§], Keerthivasan Raanin Chandradoss[§], Yachna Jain, Harpreet Singh, Mohan Lal, Kuljeet Singh Sandhu*

Department of Biological Sciences
Indian Institute of Science Education and Research (IISER) - Mohali
Knowledge City, Sector - 81, SAS Nagar 140306, India

[§]*Equal contribution*

**To whom correspondence should be addressed*
Kuljeet Singh Sandhu
Assistant Professor
Department of Biological Sciences
Indian Institute of Science Education and Research (IISER) - Mohali
E. mail: sandhuks@iisermohali.ac.in

Supplementary Figure Legends

Figure S1. A flowchart illustrating overall strategy to obtain syntenic and LOP CNE-gene pairs.

Figure S2. Distribution of CNE density per gene (within 1Mb distance) in LOP and syntenic sets in human. P-value was calculated using two-tailed Mann Whitney U test.

Figure S3. Enrichment of (a) Mammalian Phenotype Ontology (MPO) terms, and (b) Bgee anatomical terms in syntenic (left panel) and rat-LOP (right panel) gene-sets. P-values were corrected using the Benjamini-Hochberg method.

Figure S4. Extension of epigenomic analyses of syntenic and rat-LOP CNEs. (a-b) H3K4me1 ChIP enrichment (over input DNA) around syntenic (black) and rat-LOP (red) CNEs in (a) fetal, and (b) post-natal tissues. (c) The strategy used to assess the CNE-TSS interactions using HiC datasets of human fetal and adult brains. All-to-all HiC interactions were filtered for TSS-to-all interactions for the genes in syntenic and rat-LOP sets. The resultant data was analogous to 4C and was analyzed using method atypical for 4C analysis. Loess regression line was fit to 4C counts as a function of genomic distance from the reference point (TSS in this case). The distance of 3-standard deviation from this regression line was taken as significance cut-off for the interactions impinging onto TSS. (d) Example of CNE-TSS interactions identified in the rat-LOP set. Upper and lower panels represent fetal and adult brain data. Red-line: 4C signal; grey line: H3K4me1; Black line: Loess fit; Dotted line: 3-standard deviation cut-off for significance.

Figure S5. Motif enrichment analysis of rat-LOP CNEs. 'Peak-motifs' from RSAT was used to identify over-represented sequence motifs in the rat-LOP CNEs while taking syntenic CNEs as background control. (a) Sequence motifs, their e-values and the matching transcription factors (TFs) from JASPAR. (b) Tissue specificity analyses of TFs. Red bars represent the tissues, wherein the TFs exhibit significant specificity ($P<0.05$). (c) Time course gene expression of TFs during human brain development. The red curve represents the average expression of TFs and grey color denotes 90% confidence interval.

Figure S6. Analysis of distance dependent categories of rat-LOP genes**. (a)** Enrichment of gene ontology terms (as revealed by 'toppgene' web-server) among genes in three different non-overlapping gene-sets prepared by calculating tertiles of CNE-gene distance distribution in rat-LOP set. (b) Corresponding developmental gene expression trajectories of human and rat brains. Red curve signifies rat-LOP genes, and the black line with 95% confidence intervals (grey) denotes null distribution obtained from rest of the genome. P-values were calculated using one-tailed t-tests for the fetus-specific upregulation of rat-LOP genes as compared to postnatal.

Figure S7. Analysis of gain of proximity instances in human. The gain of proximity instances were identified using the strategy shown. No enrichment was found in the GO process terms. The most enriched GO term is shown.

Figure S8. Analysis of 'loss of proximity' events that were common to rodents. (a) Enrichment of gene ontology terms among genes in rodent-LOP-2 set (syntenic in human and chimp, but lost proximity in rat and mouse, n=832). Rodent-LOP-1 set (intersection of rat-LOP and mouse-LOP, n=62) did not exhibit signifcant enrichment of any GO term at 10% FDR. (b) Developmental gene expression patterns of human, rat and mouse brains. Red curve signifies rodent-LOP genes, and black line denotes the null distribution. P-values were calculated using one tailed t-tests for the fetus-specific upregulation of rodent-LOP genes as compared to post-natal.

Figure S9. Enrichment of H3K4me1 mark on syntenic and rodent-LOP CNEs. Shown are the fetal-to-adult ratio of normalized H3K4me1 enrichment in +/- 200 bp from the center of the CNEs in rodent-LOP and syntenic sets. Rodent-LOP-1 set is obtained by inducting mouse data onto original rat-LOP data. Rodent-LOP-2 set was obtained by independently comparing CNE-gene pairs in primates (human and chimp) and rodents (rat and mouse). P-values were calculated using two-tailed t-tests.

Figure S10. Analysis of multually exclusive 'loss of proximity' events in mouse and rat. Shown are the developmental gene expression trajectories for the genes that were syntenic in mouse and human but lost proximity in rat (n=125, left panel), and for the genes that were syntenic in rat and human but lost proximity in mouse (n=23, right panel). The correlation between loss of proximity and the fetal brain specific upregulation of genes can be seen. The gene-sets in this analysis were obtained by mapping mouse data onto the main datasets of CNE-gene pairs. P-values were calculated using one tailed t-tests for the fetal brain specific upregulation of LOP genes as compared to postnatal.

Figure S11. Comparison of rat-LOP CNE-gene pairs in rn5 and rn6 genome assemblies. (a) Shown is the scatter plot of CNE-gene distances (log10 scale) of the rat-LOP set in rn5 and rn6 assemblies. (b) Gene Ontology enrichment analysis of genes that had lost synteny consistently in both assemblies (83.5% of total). P-values are adjusted using the Benjamini-Hochberg method.

Supplementary Data legend

Spreadsheet tables having raw datasets used in the analyses.

Figure S1

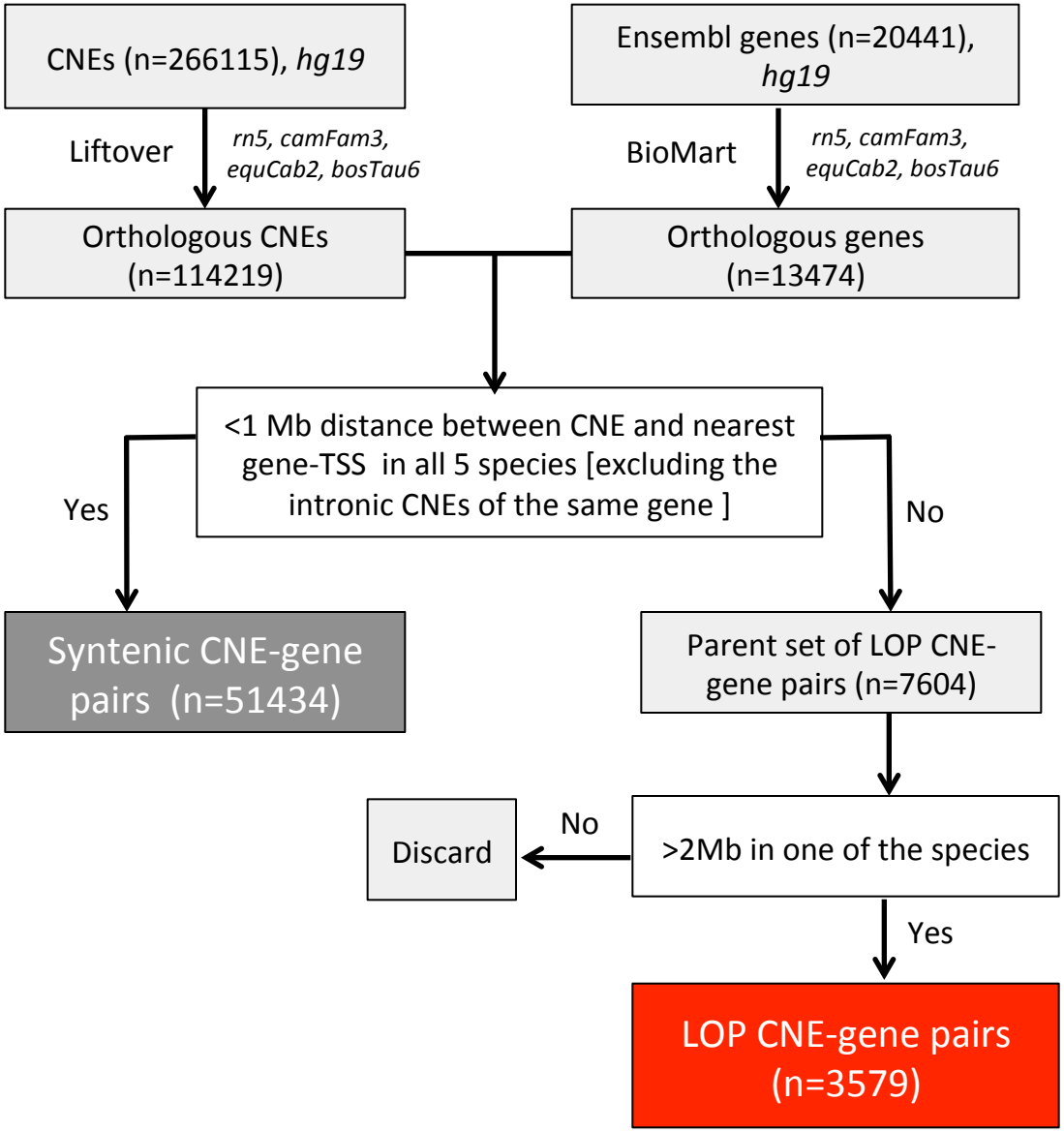

CNEs (n=266115), hg19
Ensembl genes (n=20441), hg19
Liftover
rn5, camFam3, equCab2, bosTau6
BioMart
rn5, camFam3, equCab2, bosTau6
Orthologous CNEs (n=114219)
Orthologous genes (n=13474)
<1 Mb distance between CNE and nearest gene-TSS in all 5 species [excluding the intronic CNEs of the same gene]
Yes
No
Syntenic CNE-gene pairs (n=51434)
Parent set of LOP CNE-gene pairs (n=7604)
Discard
No
>2Mb in one of the species
Yes
LOP CNE-gene pairs (n=3579)

Figure S2

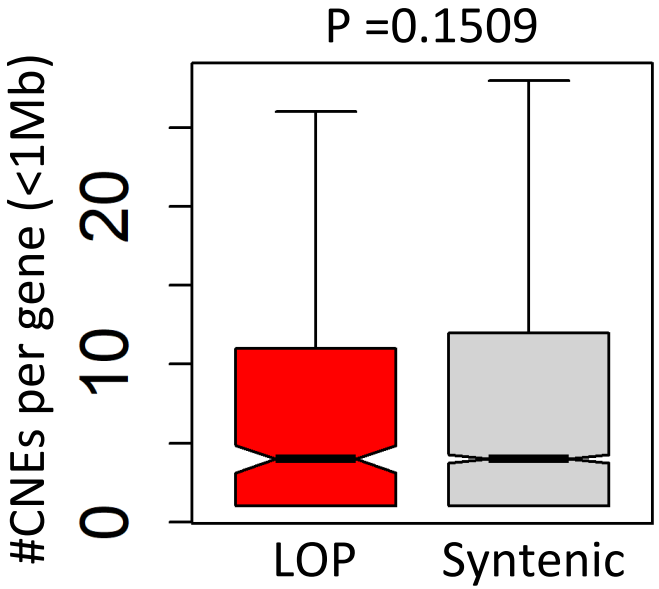


Figure S3

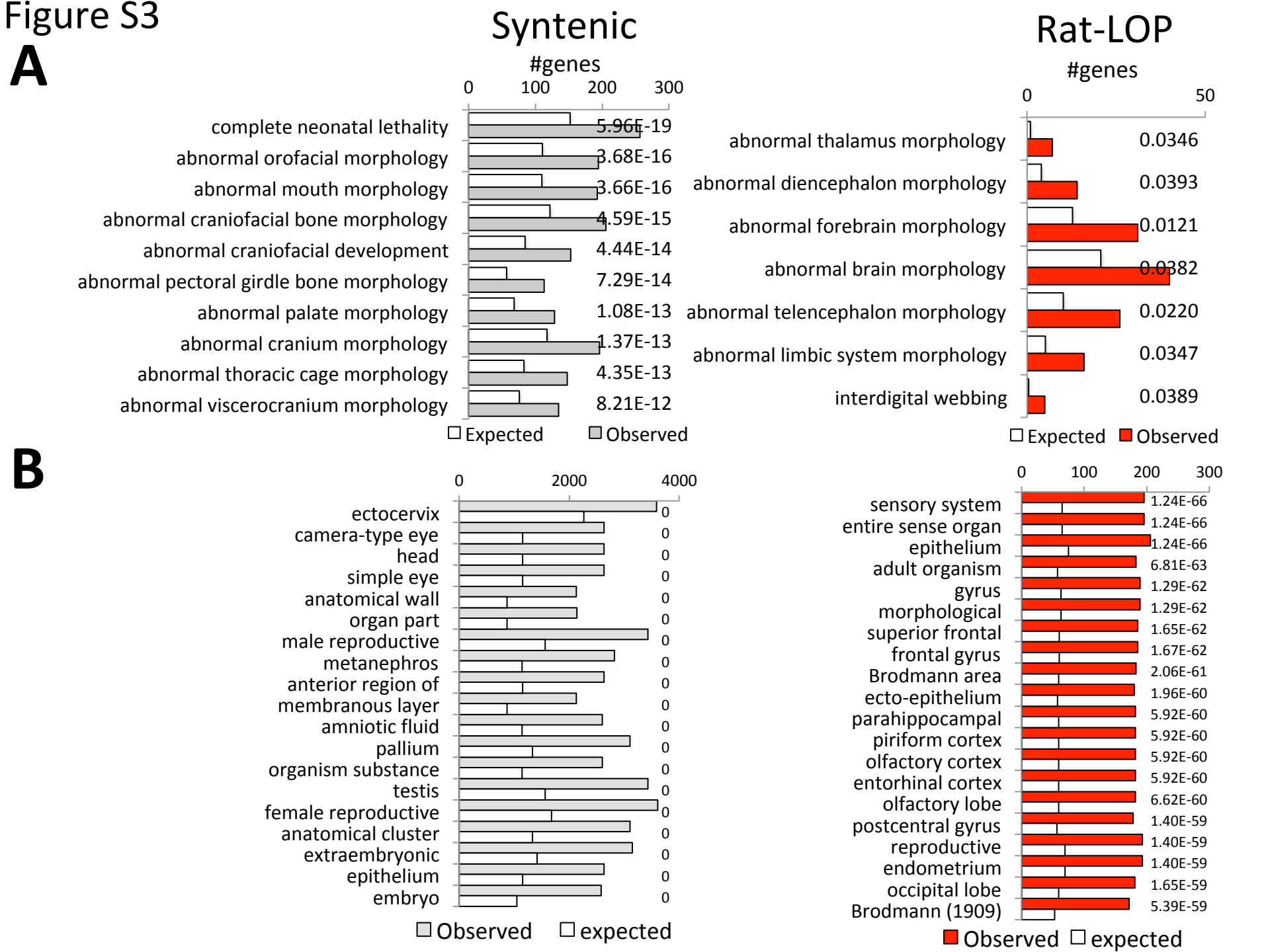

A
Syntenic
#genes
0
100
200
300
complete neonatal lethality
abnormal orofacial morphology
abnormal mouth morphology
abnormal craniofacial bone morphology
abnormal craniofacial development
abnormal pectoral girdle bone morphology
abnormal palate morphology
abnormal cranium morphology
abnormal thoracic cage morphology
abnormal viscerocranium morphology
5.96E-19
3.68E-16
3.66E-16
4.59E-15
4.44E-14
7.29E-14
1.08E-13
1.37E-13
4.35E-13
8.21E-12
Expected
Observed
Rat-LOP
#genes
0
50
abnormal thalamus morphology
abnormal diencephalon morphology
abnormal forebrain morphology
abnormal brain morphology
abnormal telencephalon morphology
abnormal limbic system morphology
interdigital webbing
0.0346
0.0393
0.0121
0.0382
0.0220
0.0347
0.0389
Expected
Observed
B
0
2000
4000
ectocervix
camera-type eye
head
simple eye
anatomical wall
organ part
male reproductive
metanephros
anterior region of
membranous layer
amniotic fluid
pallium
organism substance
testis
female reproductive
anatomical cluster
extraembryonic
epithelium
embryo
0
Observed
expected
0
100
200
300
sensory system
entire sense organ
epithelium
adult organism
gyrus
morphological
superior frontal
frontal gyrus
Brodmann area
ecto-epithelium
parahippocampal
piriform cortex
olfactory cortex
entorhinal cortex
olfactory lobe
postcentral gyrus
reproductive
endometrium
occipital lobe
Brodmann (1909)
1.24E-66
1.24E-66
1.24E-66
6.81E-63
1.29E-62
1.29E-62
1.65E-62
1.67E-62
2.06E-61
1.96E-60
5.92E-60
5.92E-60
5.92E-60
5.92E-60
6.62E-60
1.40E-59
1.40E-59
1.40E-59
1.65E-59
5.39E-59
Observed
expected

Figure S4

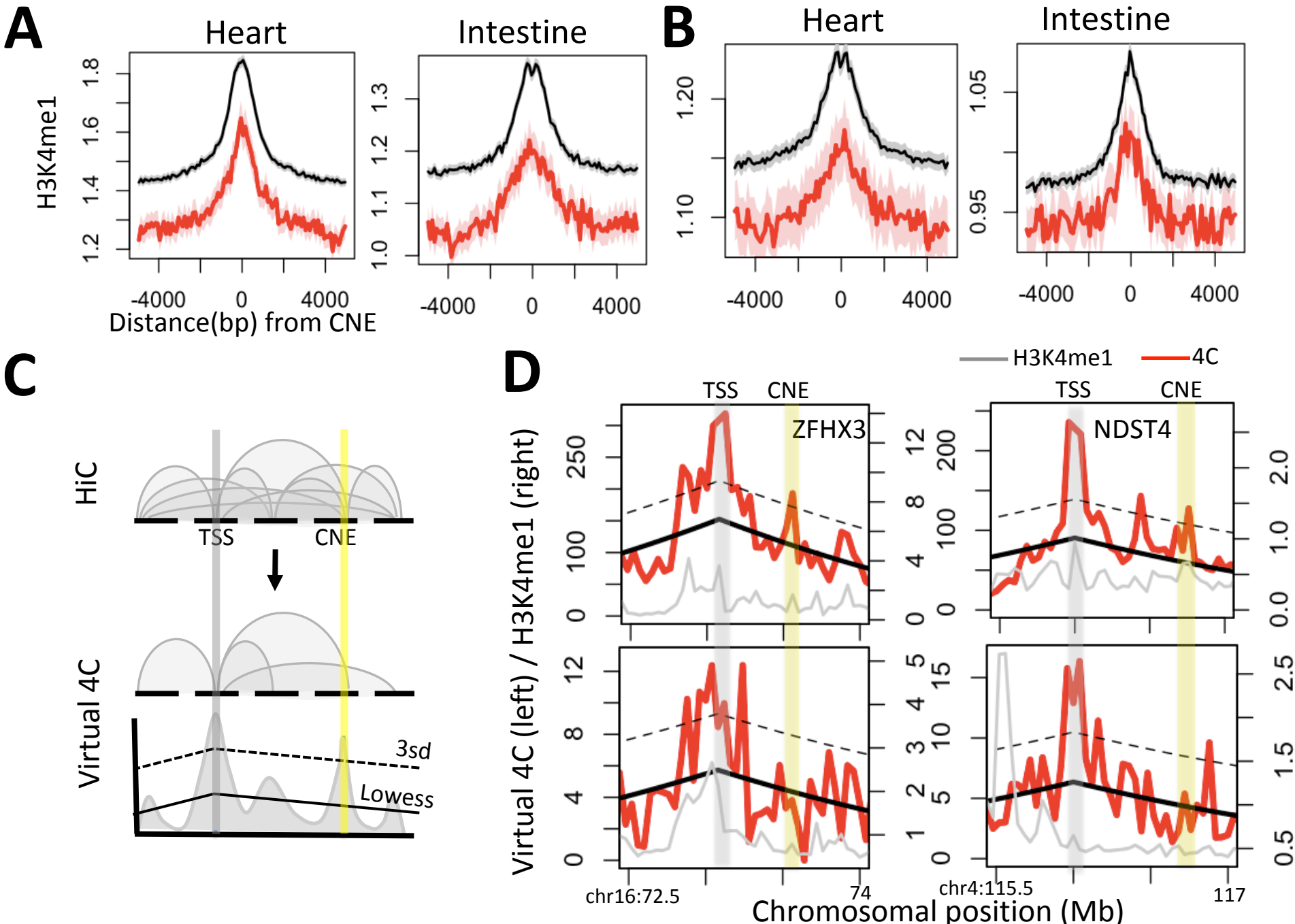
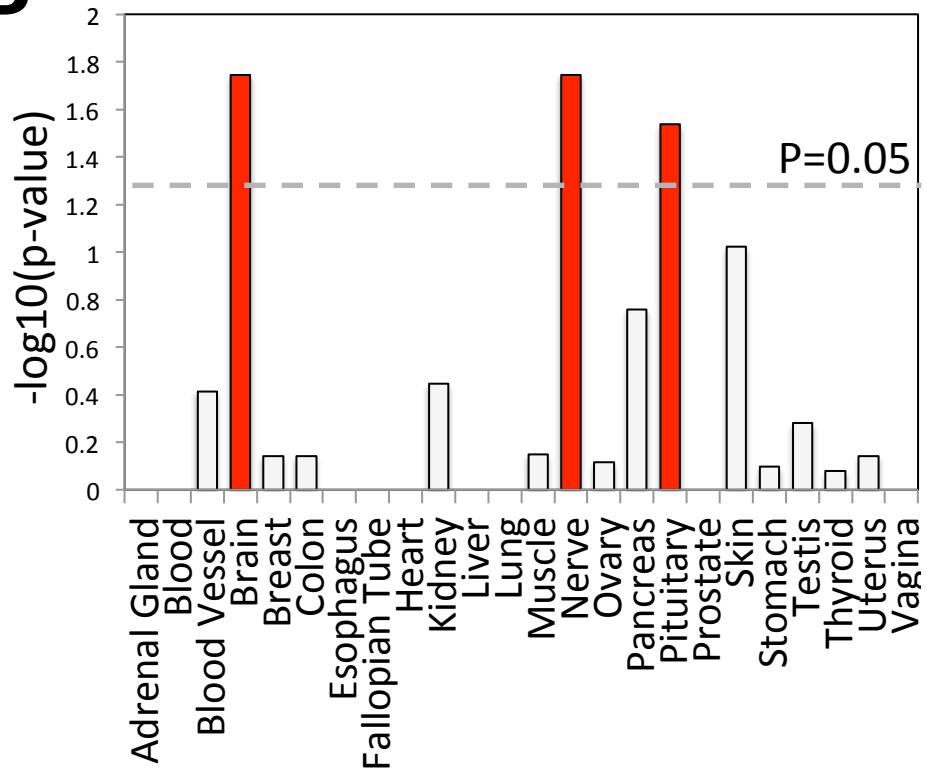

A
Heart
Intestine
H3K4me1
Distance(bp) from CNE
B
Heart
Intestine
C
HiC
TSS
CNE
Virtual 4C
3sd
Lowess
D
H3K4me1
4C
TSS
CNE
ZFHX3
NDST4
Virtual 4C (left) / H3K4me1 (right)
chr16:72.5
74
chr4:115.5
117
Chromosomal position (Mb)

Figure S5

A

Motif	E-value	TF match
mdGCAAATTAaww	0.00025	mix_a,RAX,NOTO,HESX1,GSX2,LHX2,ESX1,LBX2,Dlx1,MIXL1,MNX1,Arid3b,POU2F2,Pou2f3,POU1F1,POU3F2,POU3F1,PRRX1,SHOX,UNCX,Shox2,RAX2,Prrx2,LMX1B,ISX,LHX9,MSX2,PAX4,POU3F3,LBX1,Msx3,Nobox,DLX6,Dlx3,VSX2,NKX6_2,VSX1,MSX1,LMX1A,EN1,Dlx4,Lhx4,POU3F4,Dlx2,POU5F1B,NKX6_1,BSX,PDX1,POU5F1,Lhx3,POU2F1,Arid3a,Nkx2_5
datGaTATAATGwa	0.00038	NKX6_2
wwCATTATGwh	0.0019	NKX6_2,NR3C2,POU6F2,POU2F1,POU3F3
wrCAtAATaAaw	0.0043	ZNF410,LMX1B,POU2F1,POU3F3
hrmATGCAAat	0.0032	POU5F1,POU2F1,POU5F1B,POU3F4,POU3F3,Ddit3_Cebpa,POU2F2,POU1F1,POU3F1,POU3F2,Pou5f1_Sox2,Pou2f3

B



C

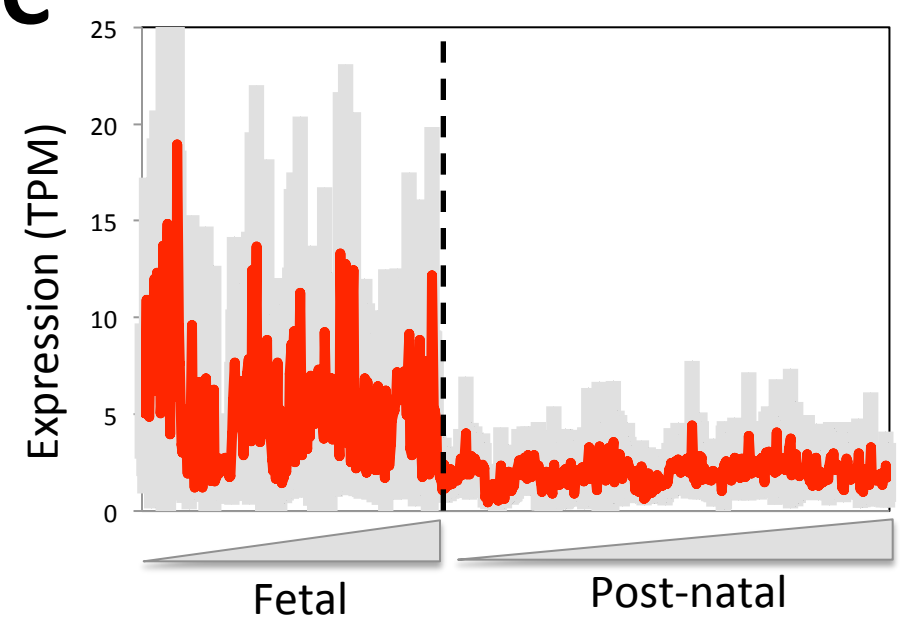


Figure S6

A

1st tertile (0-62Kb), #81

GO term	P-value
regulation of neuron differentiation	7.59E-05
phosphatidylinositol dephosphorylation	1.46E-04
regulation of glial cell differentiation	2.13E-04
regulation of nervous system dev	2.51E-04
regulation of neurogenesis	3.63E-04
negative reg of neuron projection dev	4.18E-04
phospholipid dephosphorylation	4.40E-04
long-chain fatty acid metabolic process	5.06E-04
neuron differentiation	5.28E-04
negative regulation of axon regeneration	5.96E-04

2nd tertile (62-200Kb), #81

GO term	P-value
cardiac cell fate commitment	1.96E-05
muscle structure development	8.79E-05
central nervous system neuron diff	1.92E-04
muscle cell differentiation	3.25E-04
regulation of neurogenesis	4.07E-04
specification of animal organ identity	4.57E-04
myoblast differentiation	4.87E-04
cell fate specification	5.31E-04
cranial suture morphogenesis	7.63E-04
central nervous system development	8.95E-04

3rd tertile (200-1000Kb), #82

GO term	P-value
positive regulation of synapse assembly	1.24E-05
regulation of synapse assembly	4.06E-05
synapse assembly	4.52E-05
regulation of grooming behavior	1.85E-04
regulation of synapse organization	2.22E-04
tricarboxylic acid cycle	2.67E-04
peptidyl-tyrosine phosphorylation	2.68E-04
peptidyl-tyrosine modification	2.77E-04
type II pneumocyte differentiation	3.87E-04
citrate metabolic process	3.94E-04

B

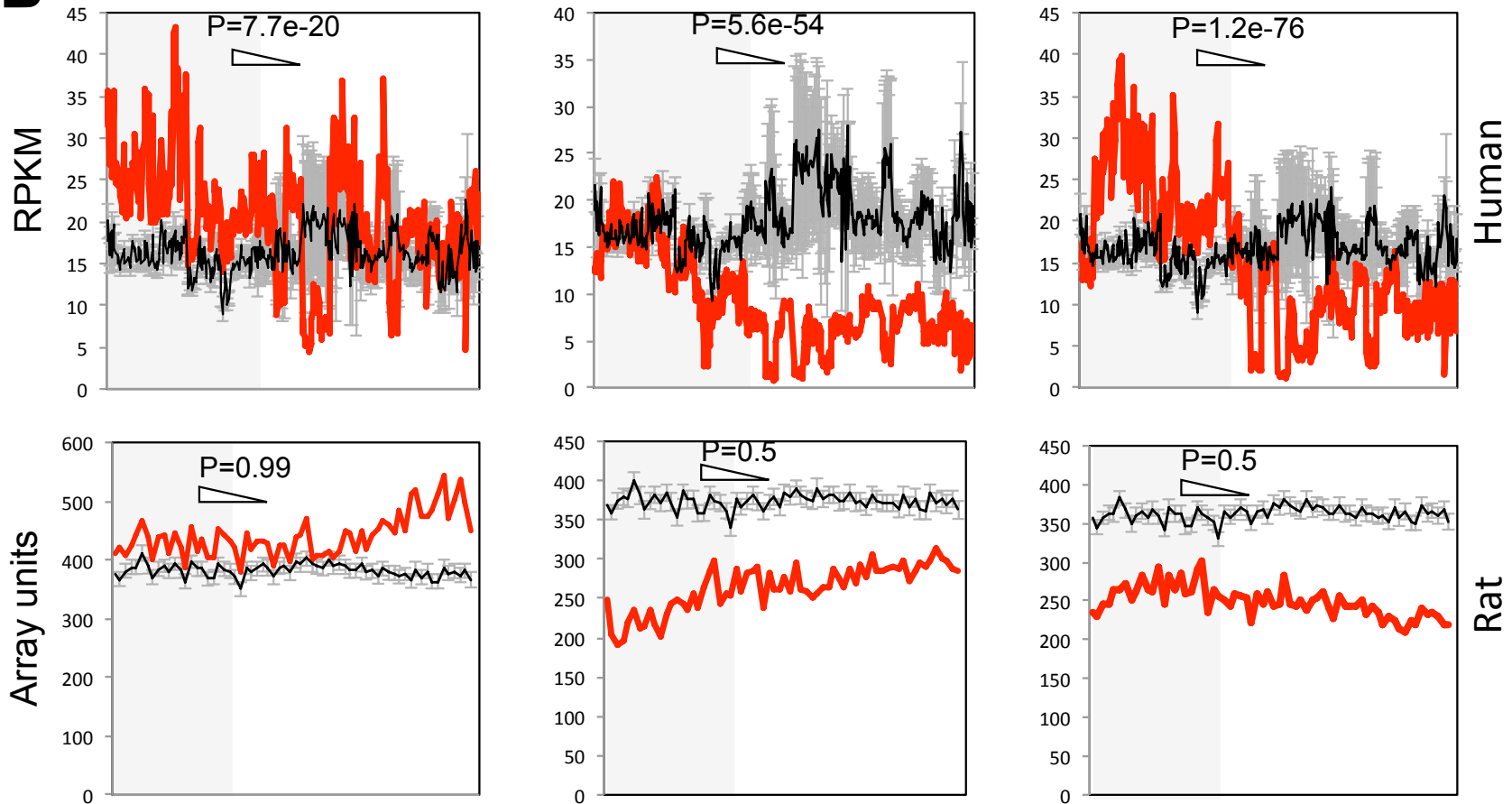
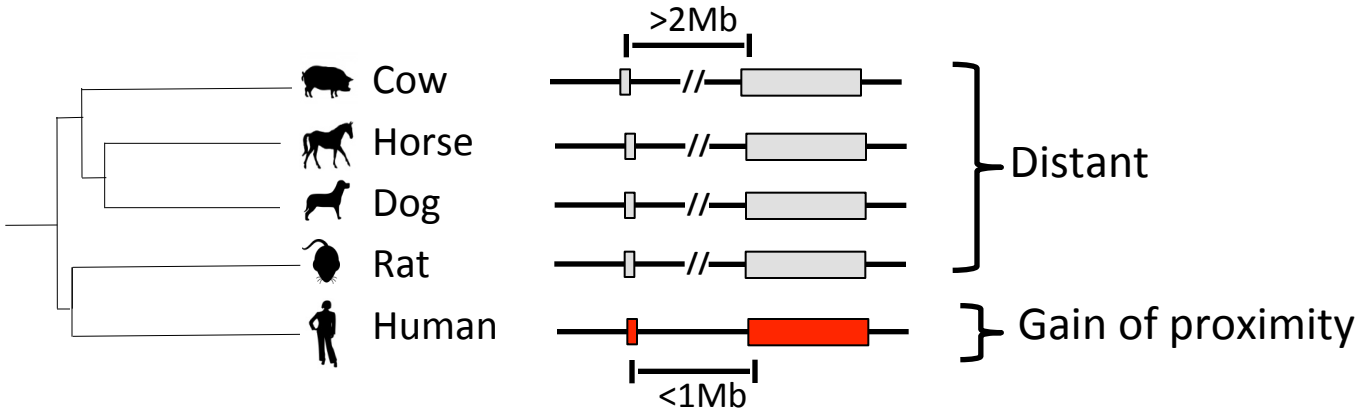


Figure S7



Genes	GO Term Name	Expected	Observed	Obs %	Fold change	FDR
388	G-protein coupled receptor signaling pathway	22.86409	41	10.5%	1.793205	1

Figure S8

A

Rodent-LOP-2
[vs. syn in primates]
(#832)

GO term	P-value	FDR
negative regulation of cell migration	1.74E-05	6.92E-02
negative regulation of cell motility	3.92E-05	6.92E-02
regulation of cell motility	4.85E-05	6.92E-02
regulation of cell migration	5.13E-05	6.92E-02
response to hormone	6.06E-05	6.92E-02
regulation of stress fiber assembly	8.21E-05	6.92E-02
Reg. of actin filament polymerization	8.45E-05	6.92E-02
Reg. of actin filament bundle assembly	8.88E-05	6.92E-02
actin filament polymerization	1.05E-04	7.28E-02
Reg. of cellular component movement	1.39E-04	7.85E-02

B

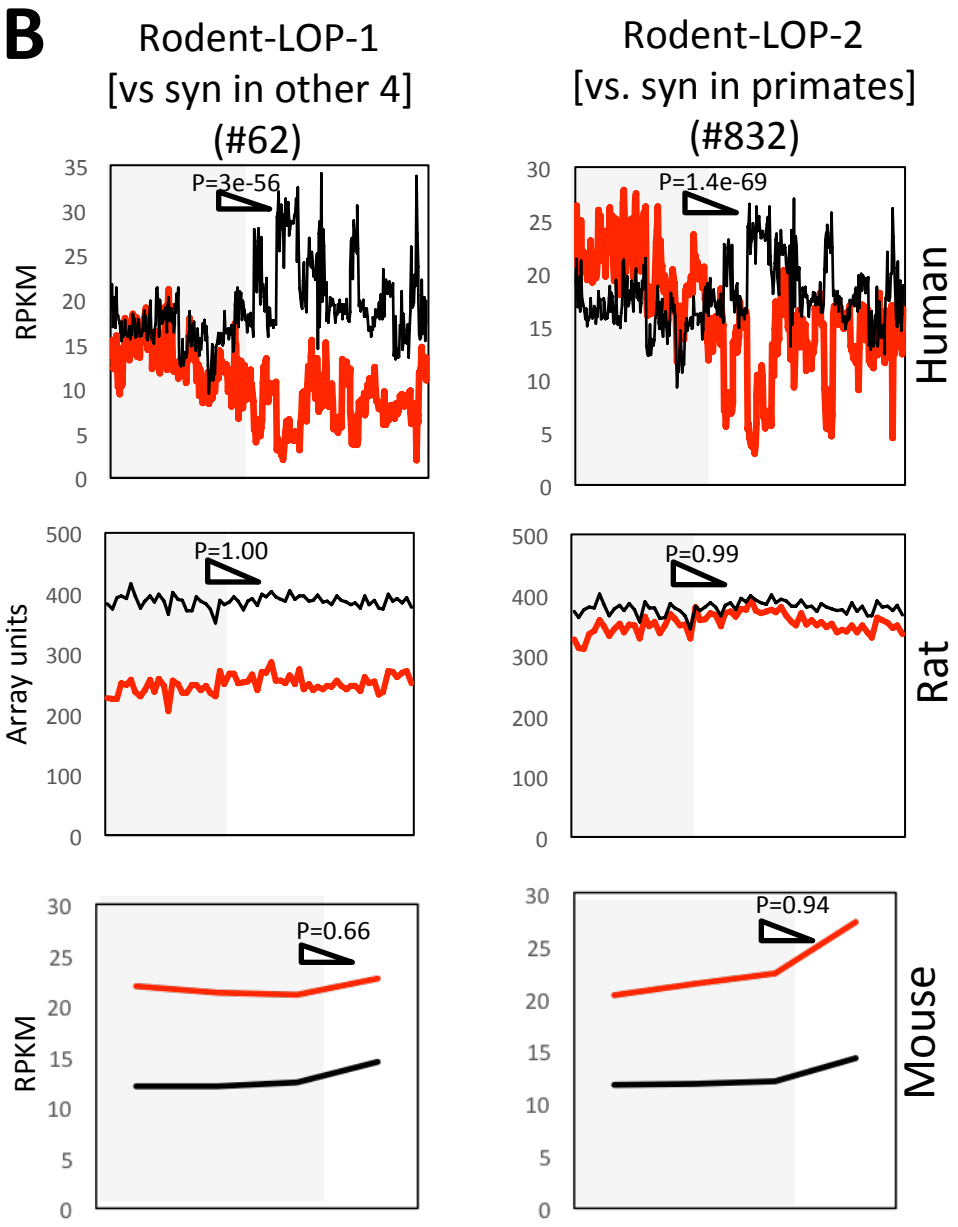


Figure S9

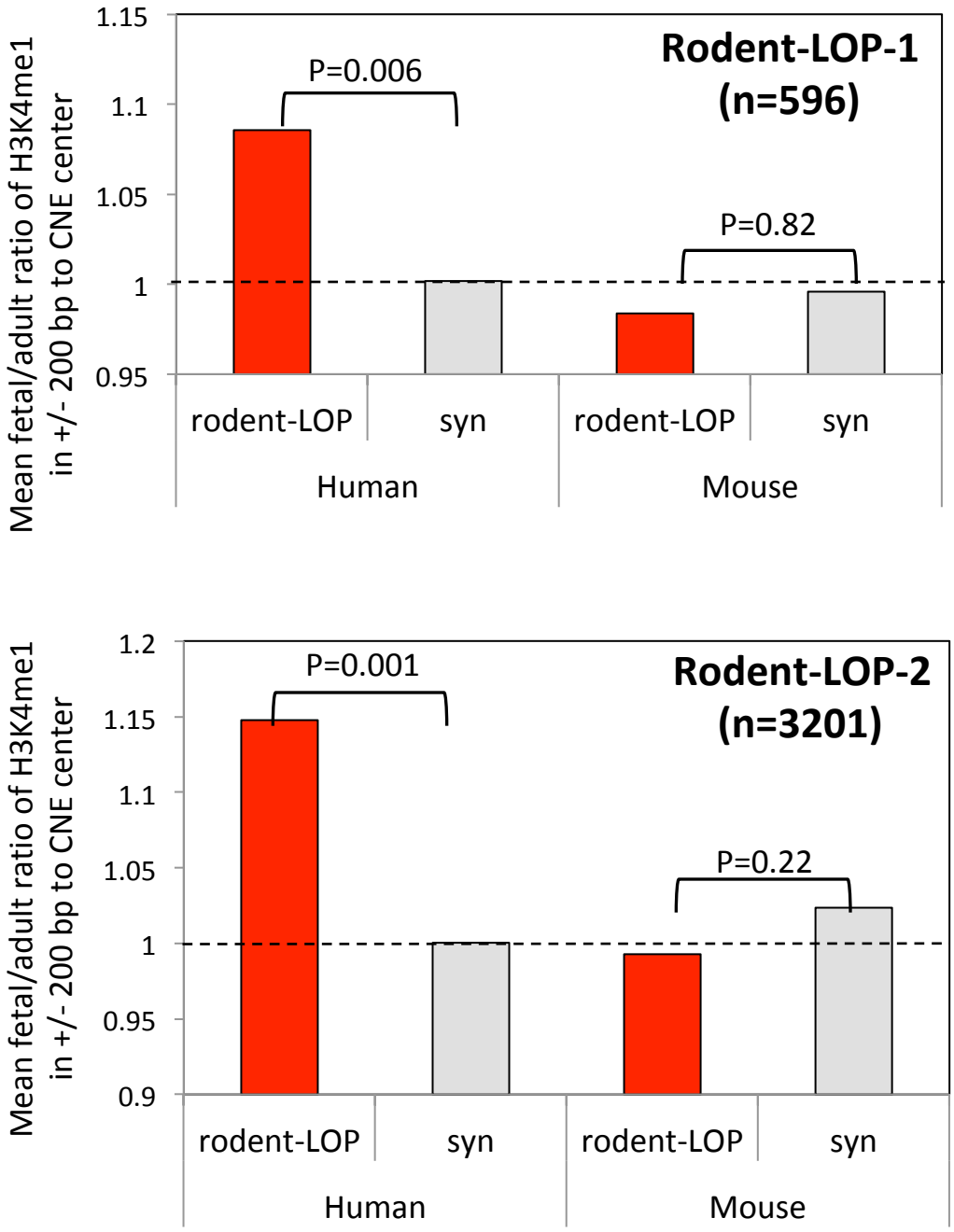

Rodent-LOP-1
(n=596)
P=0.006
P=0.82
Mean fetal/adult ratio of H3K4me1
in +/- 200 bp to CNE center
1.15
1.1
1.05
1
0.95
rodent-LOP
syn
rodent-LOP
syn
Human
Mouse
Rodent-LOP-2
(n=3201)
P=0.001
P=0.22
Mean fetal/adult ratio of H3K4me1
in +/- 200 bp to CNE center
1.2
1.15
1.1
1.05
1
0.95
0.9
rodent-LOP
syn
rodent-LOP
syn
Human
Mouse

Figure S10

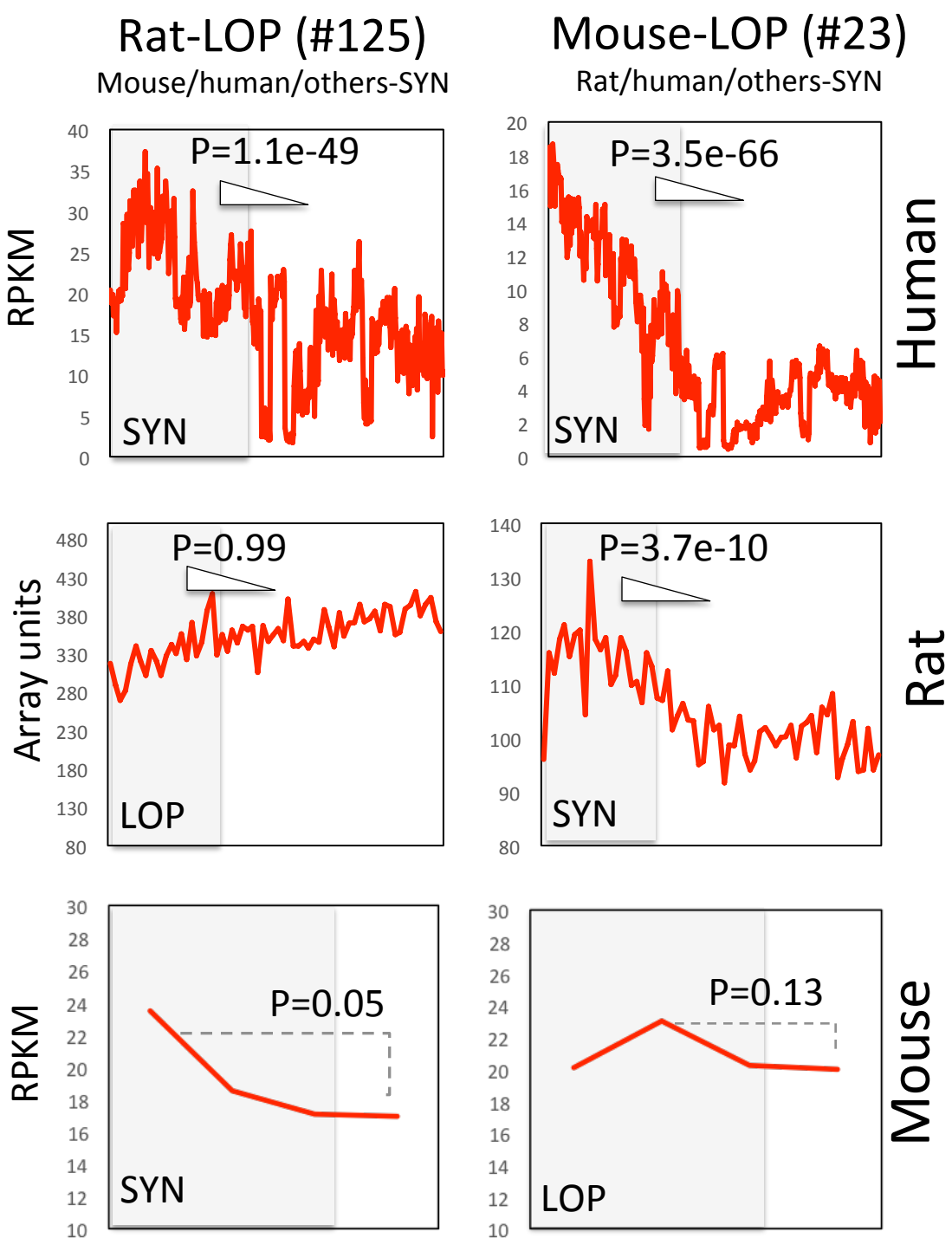
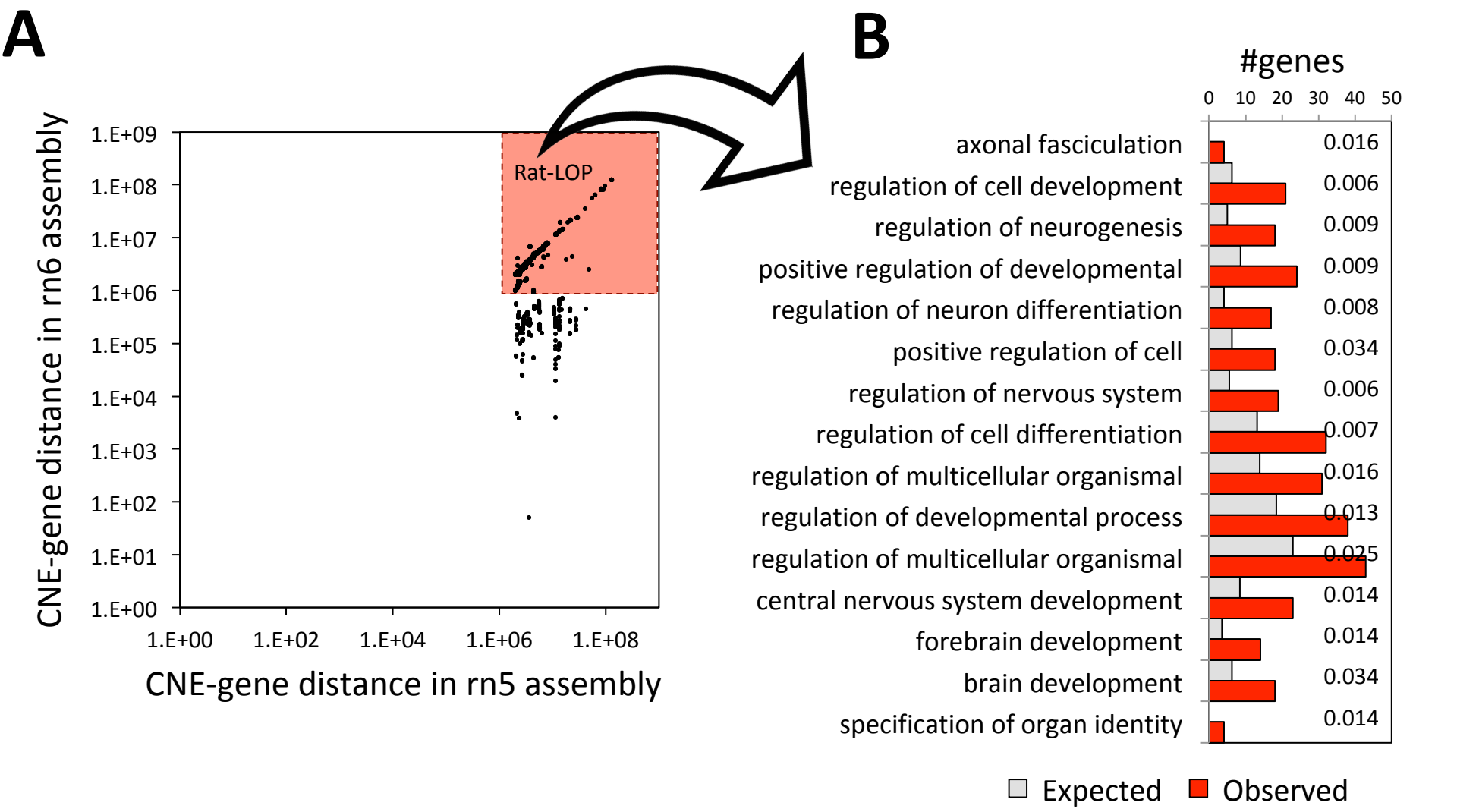

Rat-LOP (#125)
Mouse/human/others-SYN
Mouse-LOP (#23)
Rat/human/others-SYN
RPKM
P=1.1e-49
SYN
P=3.5e-66
SYN
Human
Array units
P=0.99
LOP
P=3.7e-10
SYN
Rat
RPKM
P=0.05
SYN
P=0.13
LOP
Mouse

Figure S11


A
B
Rat-LOP
CNE-gene distance in rn6 assembly
CNE-gene distance in rn5 assembly
1.E+09
1.E+08
1.E+07
1.E+06
1.E+05
1.E+04
1.E+03
1.E+02
1.E+01
1.E+00
1.E+00
1.E+02
1.E+04
1.E+06
1.E+08
#genes
0
10
20
30
40
50
axonal fasciculation
regulation of cell development
regulation of neurogenesis
positive regulation of developmental
regulation of neuron differentiation
positive regulation of cell
regulation of nervous system
regulation of cell differentiation
regulation of multicellular organismal
regulation of developmental process
regulation of multicellular organismal
central nervous system development
forebrain development
brain development
specification of organ identity
0.016
0.006
0.009
0.009
0.008
0.034
0.006
0.007
0.016
0.013
0.025
0.014
0.014
0.034
0.014
Expected
Observed