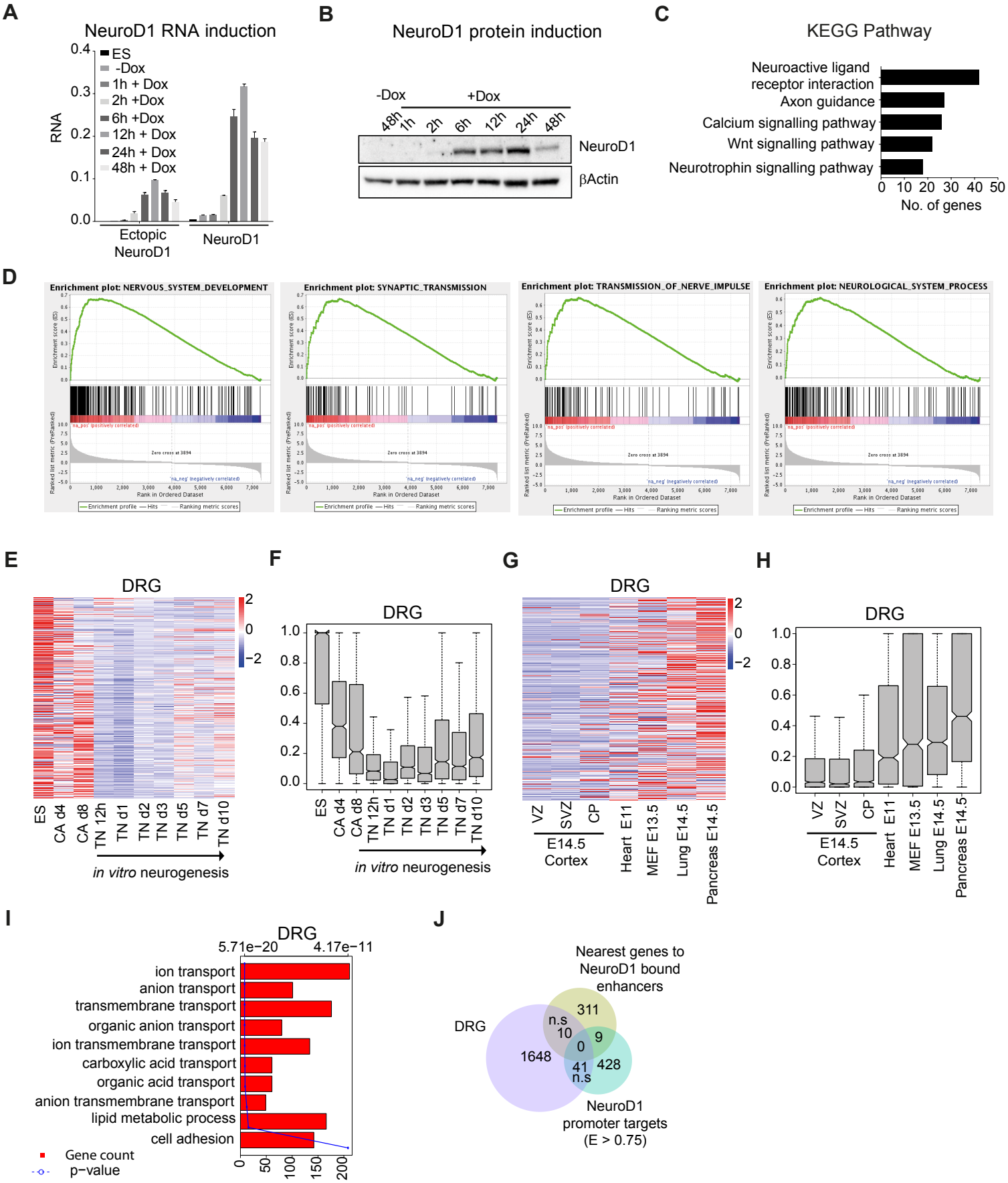
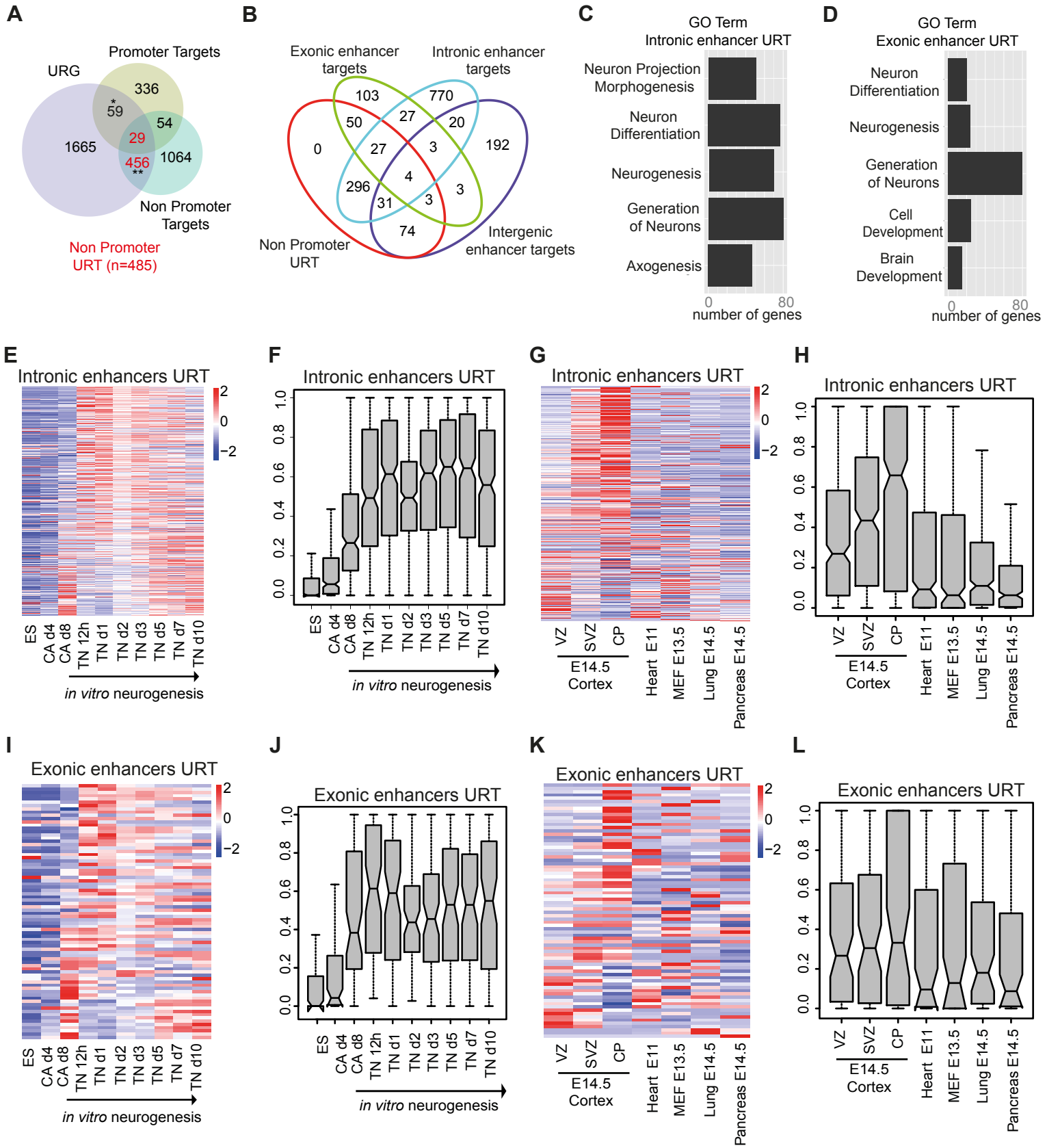
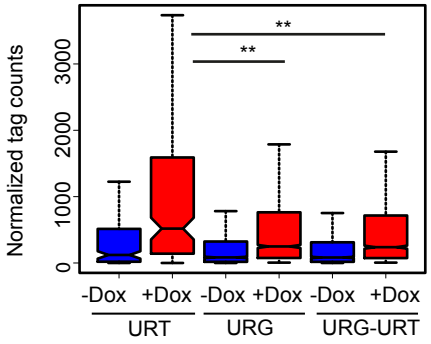
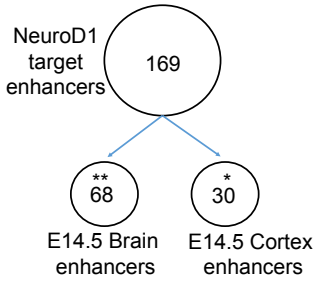

A
NeuroD1 RNA induction
RNA
ES
-Dox
1h + Dox
2h +Dox
6h +Dox
12h + Dox
24h + Dox
48h + Dox
Ectopic NeuroD1
NeuroD1
B
NeuroD1 protein induction
-Dox
+Dox
48h
1h
2h
6h
12h
24h
48h
NeuroD1
βActin
C
KEGG Pathway
Neuroactive ligand receptor interaction
Axon guidance
Calcium signalling pathway
Wnt signalling pathway
Neurotrophin signalling pathway
No. of genes
D
Enrichment plot: NERVOUS_SYSTEM_DEVELOPMENT
Enrichment plot: SYNAPTIC_TRANSMISSION
Enrichment plot: TRANSMISSION_OF_NERVE_IMPULSE
Enrichment plot: NEUROLOGICAL_SYSTEM_PROCESS
E
DRG
ES
CA d4
CA d8
TN 12h
TN d1
TN d2
TN d3
TN d5
TN d7
TN d10
in vitro neurogenesis
F
DRG
in vitro neurogenesis
G
DRG
VZ
SVZ
CP
E14.5 Cortex
Heart E11
MEF E13.5
Lung E14.5
Pancreas E14.5
H
DRG
I
DRG
5.71e−20
4.17e−11
ion transport
anion transport
transmembrane transport
organic anion transport
ion transmembrane transport
carboxylic acid transport
organic acid transport
anion transmembrane transport
lipid metabolic process
cell adhesion
Gene count
p−value
J
Nearest genes to NeuroD1 bound enhancers
DRG
311
n.s
10
9
0
1648
41
428
n.s
NeuroD1 promoter targets (E > 0.75)


A
URG
Promoter Targets
336
59
54
29
1665
456
1064
Non Promoter Targets
Non Promoter URT (n=485)
B
Exonic enhancer targets
Intronic enhancer targets
103
770
50
27
20
0
27
3
192
4
296
31
3
74
Non Promoter URT
Intergenic enhancer targets
C
GO Term
Intronic enhancer URT
Neuron Projection Morphogenesis
Neuron Differentiation
Neurogenesis
Generation of Neurons
Axogenesis
0
80
number of genes
D
GO Term
Exonic enhancer URT
Neuron Differentiation
Neurogenesis
Generation of Neurons
Cell Development
Brain Development
0
80
number of genes
E
Intronic enhancers URT
2
0
−2
ES
CA d4
CA d8
TN 12h
TN d1
TN d2
TN d3
TN d5
TN d7
TN d10
in vitro neurogenesis
F
Intronic enhancers URT
0.0
0.2
0.4
0.6
0.8
1.0
ES
CA d4
CA d8
TN 12h
TN d1
TN d2
TN d3
TN d5
TN d7
TN d10
in vitro neurogenesis
G
Intronic enhancers URT
2
0
−2
VZ
SVZ
CP
E14.5 Cortex
Heart E11
MEF E13.5
Lung E14.5
Pancreas E14.5
H
Intronic enhancers URT
0.0
0.2
0.4
0.6
0.8
1.0
VZ
SVZ
CP
E14.5 Cortex
Heart E11
MEF E13.5
Lung E14.5
Pancreas E14.5
I
Exonic enhancers URT
2
0
−2
ES
CA d4
CA d8
TN 12h
TN d1
TN d2
TN d3
TN d5
TN d7
TN d10
in vitro neurogenesis
J
Exonic enhancers URT
0.0
0.2
0.4
0.6
0.8
1.0
ES
CA d4
CA d8
TN 12h
TN d1
TN d2
TN d3
TN d5
TN d7
TN d10
in vitro neurogenesis
K
Exonic enhancers URT
2
0
−2
VZ
SVZ
CP
E14.5 Cortex
Heart E11
MEF E13.5
Lung E14.5
Pancreas E14.5
L
Exonic enhancers URT
0.0
0.2
0.4
0.6
0.8
1.0
VZ
SVZ
CP
E14.5 Cortex
Heart E11
MEF E13.5
Lung E14.5
Pancreas E14.5

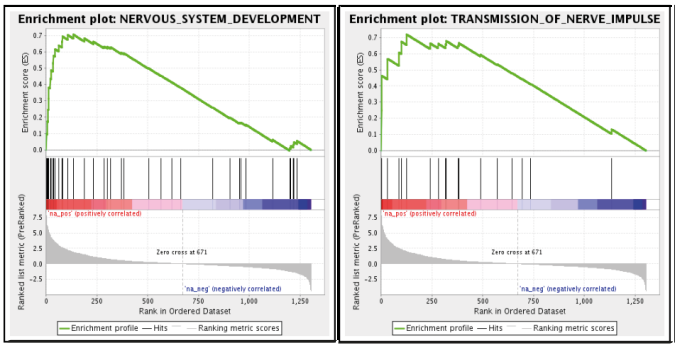
**A**



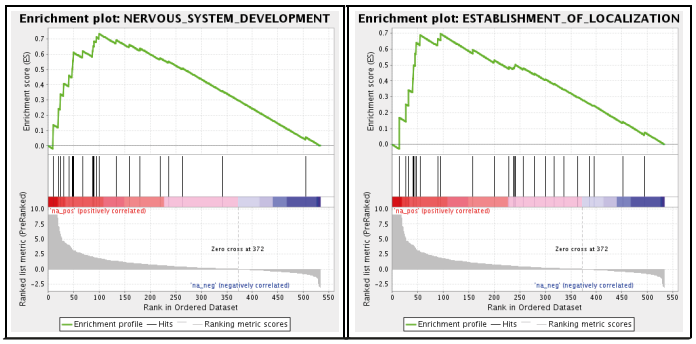

**B**




**C**




**D**




**E**

Promoter motifs

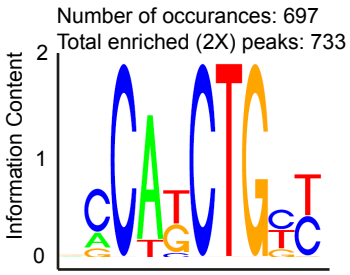
| Rank | Motif | Name | p-val |
|---|---|---|---|
| 1 |  | Atoh1(bHLH)/Cerebellum-Atoh1-ChIP-Seq(GSE22111)/Homer | 1e-4 |
| 2 |  | NeuroD1(bHLH)/Islet-NeuroD1-ChIP-Seq(GSE30298)/Homer | 1e-3 |
| 3 |  | Olig2(bHLH)/Neuron-Olig2-ChIP-Seq(GSE30882)/Homer | 1e-3 |

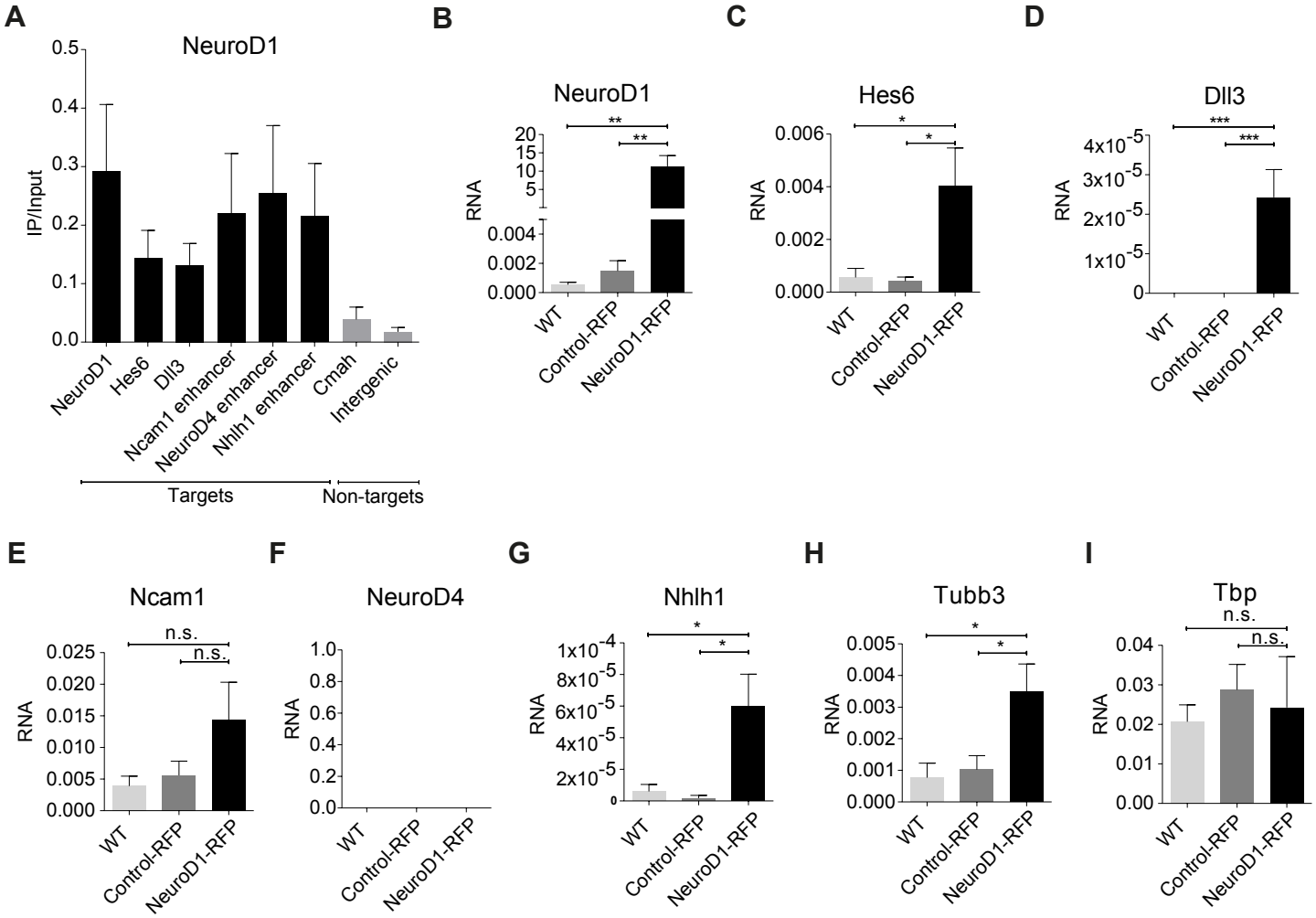
**F**

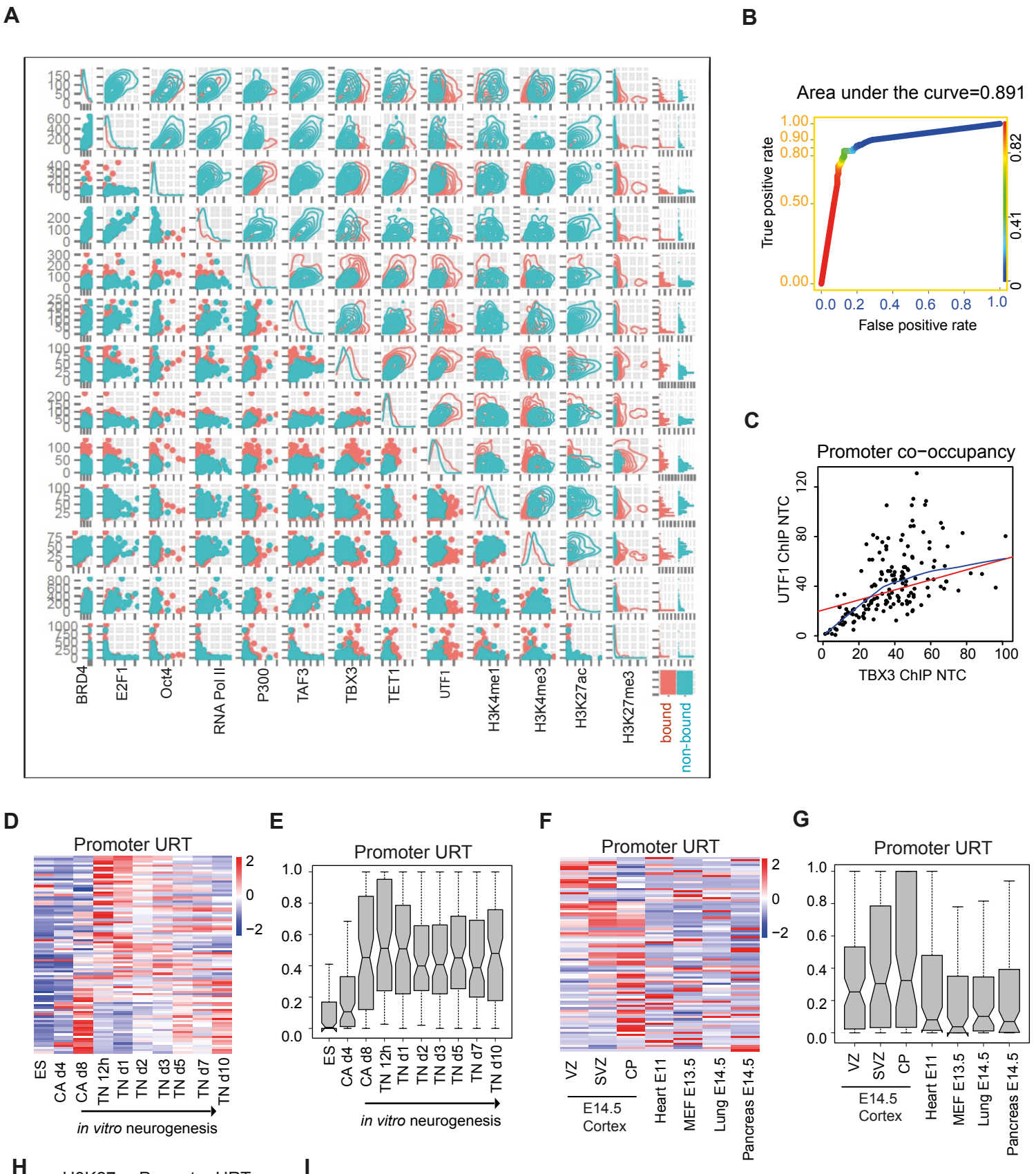
Enhancer motifs

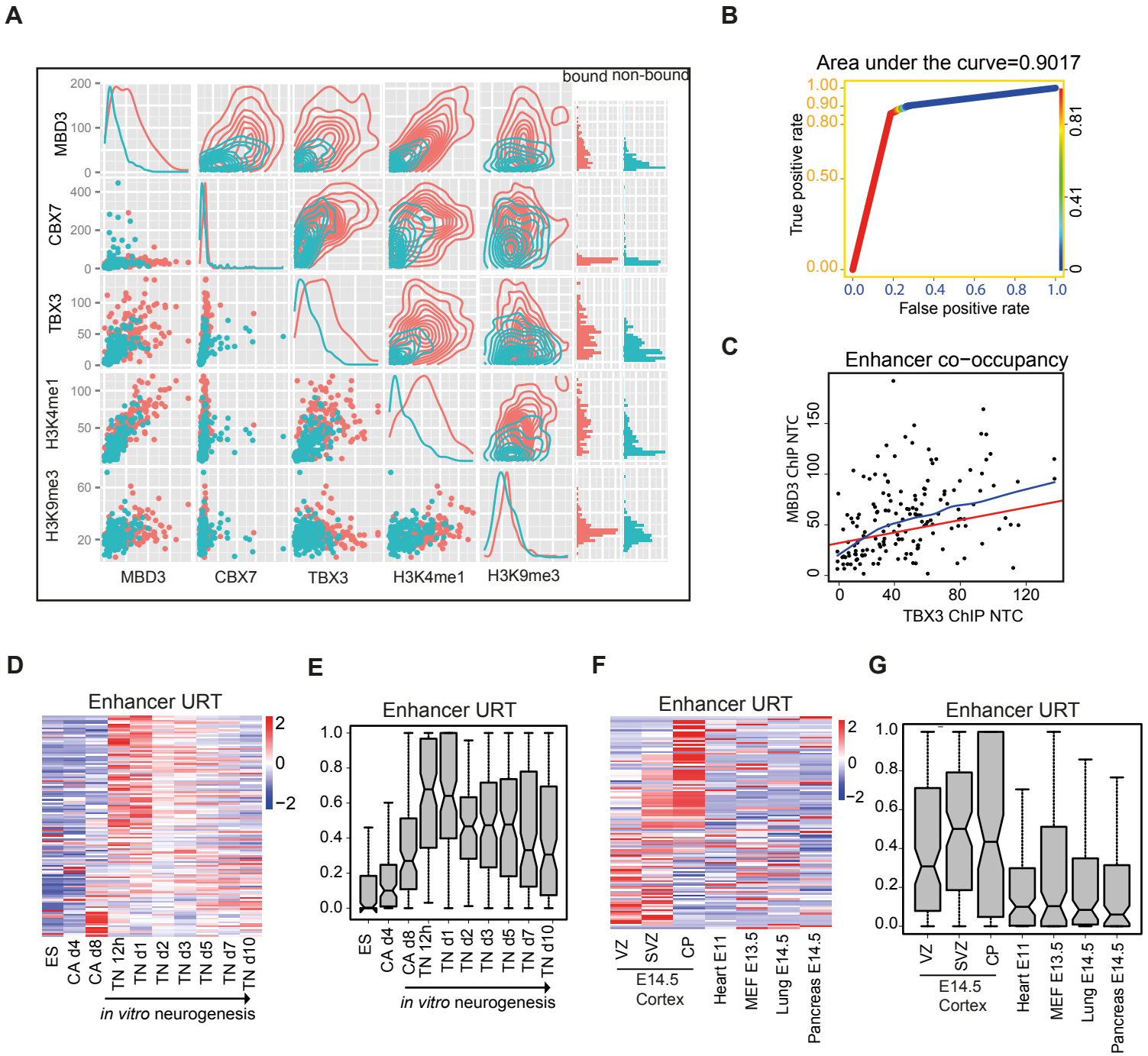
| Rank | Motif | Name | p-val |
|---|---|---|---|
| 1 |  | Atoh1(bHLH)/Cerebellum-Atoh1-ChIP-Seq(GSE22111)/Homer | 1e-8 |
| 2 |  | Olig2(bHLH)/Neuron-Olig2-ChIP-Seq(GSE30882)/Homer | 1e-8 |
| 3 |  | NeuroD1(bHLH)/Islet-NeuroD1-ChIP-Seq(GSE30298)/Homer | 1e-7 |
| 4 |  | Sox3(HMG)/NPC-Sox3-ChIP-Seq(GSE33059)/Homer | 1e-5 |
| 5 |  | Ptf1a(HLH)/Panc1-Ptf1a-ChIP-Seq(GSE47459)/Homer | 1e-4 |
| 6 |  | Sox2(HMG)/mES-Sox2-ChIP-Seq(GSE11431)/Homer | 1e-4 |
| 7 |  | Unknown-ESC-element/mES-Nanog-ChIP-Seq(GSE11724)/Homer | 1e-3 |
| 8 |  | Lhx3(Homeobox)/Neuron-Lhx3-ChIP-Seq(GSE31456)/Homer | 1e-3 |
| 9 |  | Tcf12(HLH)/GM12878-Tcf12-ChIP-Seq(GSE32465)/Homer | 1e-3 |
| 10 |  | SCL(HLH)/HPC7-Scl-ChIP-Seq(GSE13511)/Homer | 1e-3 |
| 11 |  | Nanog(Homeobox)/mES-Nanog-ChIP-Seq(GSE11724)/Homer | 1e-2 |
| 12 |  | E2A(HLH)/proBcell-E2A-ChIP-Seq(GSE21978)/Homer | 1e-2 |
| 13 |  | Nkx6.1(Homeobox)/Islet-Nkx6.1-ChIP-Seq(GSE40975)/Homer | 1e-2 |
| 14 |  | Sox6(HMG)/Myotubes-Sox6-ChIP-Seq(GSE32627)/Homer | 1e-2 |
| 15 |  | Isl1(Homeobox)/Neuron-Isl1-ChIP-Seq(GSE31456)/Homer | 1e-2 |
| 16 |  | MyoD(HLH)/Myotube-MyoD-ChIP-Seq(GSE21614)/Homer | 1e-2 |
| 17 |  | Smad3(MAD)/NPC-Smad3-ChIP-Seq(GSE36673)/Homer | 1e-2 |
| 18 |  | EBF(EBF)/proBcell-EBF-ChIP-Seq(GSE21978)/Homer | 1e-2 |

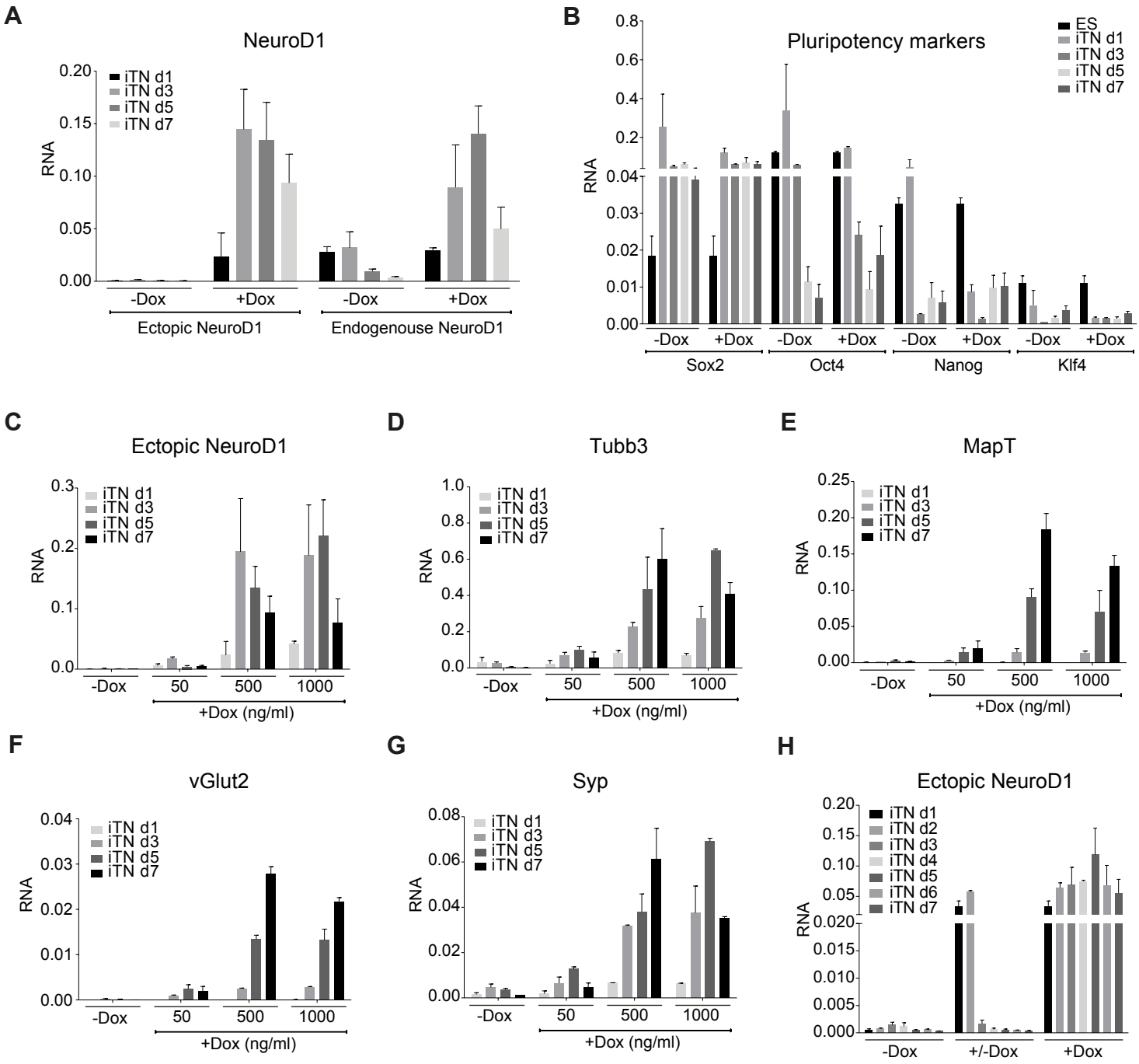
**G**




A
NeuroD1
IP/Input
0.5
0.4
0.3
0.2
0.1
0.0
NeuroD1
Hes6
Dll3
Ncam1 enhancer
NeuroD4 enhancer
Nhlh1 enhancer
Cmah
Intergenic
Targets
Non-targets
B
NeuroD1
RNA
20
15
10
5
0.004
0.002
0.000
**
**
WT
Control-RFP
NeuroD1-RFP
C
Hes6
RNA
0.006
0.004
0.002
0.000
*
*
WT
Control-RFP
NeuroD1-RFP
D
Dll3
RNA
4x10-5
3x10-5
2x10-5
1x10-5
0
***
***
WT
Control-RFP
NeuroD1-RFP
E
Ncam1
RNA
0.025
0.020
0.015
0.010
0.005
0.000
n.s.
n.s.
WT
Control-RFP
NeuroD1-RFP
F
NeuroD4
RNA
1.0
0.8
0.6
0.4
0.2
0.0
WT
Control-RFP
NeuroD1-RFP
G
Nhlh1
RNA
1x10-4
8x10-5
6x10-5
4x10-5
2x10-5
0
*
*
WT
Control-RFP
NeuroD1-RFP
H
Tubb3
RNA
0.005
0.004
0.003
0.002
0.001
0.000
*
*
WT
Control-RFP
NeuroD1-RFP
I
Tbp
RNA
0.04
0.03
0.02
0.01
0.00
n.s.
n.s.
WT
Control-RFP
NeuroD1-RFP


A
B
Area under the curve=0.891
True positive rate
False positive rate
C
Promoter co−occupancy
UTF1 ChIP NTC
TBX3 ChIP NTC
BRD4
E2F1
Oct4
RNA Pol II
P300
TAF3
TBX3
TET1
UTF1
H3K4me1
H3K4me3
H3K27ac
H3K27me3
bound
non-bound
D
Promoter URT
ES
CA d4
CA d8
TN 12h
TN d1
TN d2
TN d3
TN d5
TN d7
TN d10
in vitro neurogenesis
E
Promoter URT
in vitro neurogenesis
F
Promoter URT
VZ
SVZ
CP
E14.5 Cortex
Heart E11
MEF E13.5
Lung E14.5
Pancreas E14.5
G
Promoter URT
H
H3K27ac Promoter URT
ES
NP
TND1
TND10
CD4
Liver
MEF
NIH3T3
I
H3K27ac Promoter URT


A
bound non-bound
MBD3
CBX7
TBX3
H3K4me1
H3K9me3
B
Area under the curve=0.9017
True positive rate
False positive rate
C
Enhancer co−occupancy
MBD3 ChIP NTC
TBX3 ChIP NTC
D
Enhancer URT
ES
CA d4
CA d8
TN 12h
TN d1
TN d2
TN d3
TN d5
TN d7
TN d10
in vitro neurogenesis
E
Enhancer URT
in vitro neurogenesis
F
Enhancer URT
VZ
SVZ
CP
E14.5 Cortex
Heart E11
MEF E13.5
Lung E14.5
Pancreas E14.5
G
Enhancer URT
H
H3K27ac Enhancer URT
ES
NP
TND1
TND10
CD4
Liver
MEF
NIH3T3
I
H3K27ac Enhancer URT
J
H3K27ac Intronic URT
K
H3K27ac Exonic URT


A
NeuroD1
RNA
iTN d1
iTN d3
iTN d5
iTN d7
-Dox
+Dox
Ectopic NeuroD1
Endogenouse NeuroD1
B
Pluripotency markers
ES
Sox2
Oct4
Nanog
Klf4
C
Ectopic NeuroD1
+Dox (ng/ml)
D
Tubb3
E
MapT
F
vGlut2
G
Syp
H
Ectopic NeuroD1
iTN d2
iTN d4
iTN d6
+/-Dox

**LEGENDS FOR APPENDIX FIGURES**

**Appendix Figure S1**. **Doxycycline-induced expression of NeuroD1 in ES cells causes repression of non-neuronal genes.** (A) Time course of the doxycycline-induced expression dynamics of ectopic and total NeuroD1. RT-qPCRs were performed in biological replicates, RNA reflects the relative gene expression normalized to a housekeeping gene (Rpl19), and the error bars indicate the S.E.M. of two biological replicates. (B) Western blot time-course assay for HA-NeuroD1 protein as well as beta-actin after Dox induction in ES cells. (C) Bar plot showing the KEGG pathways that are enriched in URG (n=2209). Bar lengths represents number of genes contributing to a pathway enrichment (D) Gene set enrichment plot for the top 4 enriched GO terms for URG, as determined by GSEA depicting density for level of upregulation that is exhibited by genes that contributes for a particular GO term. (E) Heat map depicting the expression of DRG (Fig. 1F) in an *in vitro* neuronal RNA-seq time course using biological triplicates. Each row represents one promoter where gene expression is scaled from red (high expression) to blue (low expression). (F) Same as (E) but represented as boxplot. (G) Heatmap depicting the RNA-seq expression of DRG in various embryonic tissues. (H) Same as (G) but represented as a boxplot. (I) Bar plot depicting top GO terms enriched in DRG. Bar length represents gene-count in the dataset contributing to a GO term. Blue line represents respective p-values. (J) Venn diagram depicting overlap of URT (promoter and enhancer) with DRG.

Data information: Error bars reflect standard error of the mean. Significance was determined by Fisher's test as n.s, non-significant. Boxplots depicting RNA-seq data contain expression values scaled between 0 and 1 on y-axis. GEO IDs for all sequencing data used are provided in Table EV1.

**Appendix Figure S2. NeuroD1 target enhancers are also located in introns and exons and they function to induce expression of neuronal genes during neurogenesis.** (A) Venn diagram depicting overlap of URG with NeuroD1 promoter and non-promoter target

(intergenic, exonic and intronic enhancer) associated genes. (B) Venn diagram depicting overlap of total non-promoter NeuroD1 enhancer target associated genes with intronic, exonic and intergenic NeuroD1 enhancer associated genes. GO terms plot depicting top GO terms enriched for NeuroD1 intronic (C) and exonic (D) target enhancer associated genes. Bar length is determined by number of genes contributing to respective GO term. (E) Heat map depicting the expression upregulated NeuroD1 bound genes associated with intronic target enhancers in an *in vitro* neuronal RNA-seq time course using biological triplicates. Each row represents one promoter where gene expression is scaled from red (high expression) to blue (low expression). (F) Same as (E) but represented as boxplot. (G) Same as in (E) but for RNA-seq expression in various embryonic tissues. (H) Same as (G) but represented as a boxplot. (I) Heat map depicting the expression upregulated NeuroD1 bound genes associated with exonic target enhancers in an *in vitro* neuronal RNA-seq time course using biological triplicates. Each row represents one promoter where gene expression is scaled from red (high expression) to blue (low expression). (J) Same as (I) but represented as boxplot. (K) Same as in (I) but for RNA-seq expression in various embryonic tissues. (L) Same as (K) but represented as a boxplot.

Data information: Significance shown in (A) was determined by fischer's test with * $P < 0.05$, ** $P < 0.01$. Boxplots depicting RNA-seq data contain expression values scaled between 0 and 1 on y-axis. GEO IDs for all sequencing data used are provided in Table EV1.

**Appendix Figure S3. NeuroD1 binds its target sites in a highly sequence-specific manner to induce neuronal gene expression program**. (A) Boxplot depicting RNA-seq normalized tag counts of URT (set 1) and URG (set 2) as well as URG-URT (set 3) in –Dox and +Dox conditions after 48 h of NeuroD1 induction in ES cells. (B) Overlap of NeuroD1 bound enhancer sites regulating induced genes with publicly available tissue specific enhancer sites in E14.5 cortex and brain. (C-D) GO term enrichment plot from GSEA for the top two GO terms of promoter URT (C) and enhancer URT (D). (E-F) Motif enriched at NeuroD1-bound promoters (E) and NeuroD1-bound enhancer sites (F). (G) *De novo* motif

built on enriched NeuroD1 peaks (E > 1) represented by information content of each nucleotide in a positional manner. 95% of peaks showed the depicted motif at peak center.
Data information: Significance shown in (A) was determined by *t*-test with * $P < 0.05$, ** $P < 0.01$. Significance shown in (B) was determined by fischer's test with * $P < 0.05$, ** $P < 0.01$.

**Appendix Figure S4. NeuroD1 is able to bind and induce its targets in murine fibroblasts**. (A) ChIP qPCR results for the enrichment of NeuroD1 at target sites (black) and control regions (grey) in NIH3T3 fibroblasts after transient overexpression for 48 h. DNA normalized to the respective input DNA is plotted on the y-axis (IP/Input). The error bars indicate S.E.M. of four biological replicates. (B-I) RT-qPCR results for the expression of NeuroD1 targets (B-G), the neuronal marker Tubb3 (H) and a housekeeping control (Tbp, I) in FAC-sorted NIH3T3 cells upon transfection with NeuroD1-IRES-RFP or Control-IRES-RFP plasmids for 48 h or in non-transfected control cells (wild type, WT). RNA reflects the relative gene expression normalized to a housekeeping gene (Rpl19), error bars indicate S.E.M. of four biological replicates.
Data information: Significance was determined by *t*-test with * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

**Appendix Figure S5. Identification of features that discriminate between NeuroD1-bound and non-NeuroD1-bound promoters using Bayesian modeling.** (A) Distribution plot shows how ChIP-seq signal strength of all factors predicted by bayesian model correlates at NeuroD1 promoter URT to classify between control (non-target promoter of URG) and targets (promoter URT) with cutoff for information gain coefficient of 0.1. NeuroD1 promoter URT are depicted in red while control promoters in cyan. Diagonal and top-right charts group show one and two dimensional density plots respectively. Bottom left charts depict scatter plots for ChIP-seq enrichment. (B) Receiver operating characteristic (ROC) plot for a Bayesian model developed to distinguish between targets and control promoters from (A) (C) Scatter plot for UTF1 and TBX3 co-occupancy at NeuroD1 promoter URT. The

linear regression line is shown in red, and the trend line is shown in blue. (D) Heat map depicting the expression of promoter URT in an *in vitro* neuronal RNA-seq time course using biological triplicates. Each row represents one promoter where gene expression is scaled from red (high expression) to blue (low expression). (E) Same as (D) but represented as boxplot. (F) Same as in (D) but for RNA-seq expression in various embryonic tissues. (G) Same as (F) but represented as a boxplot. (H) Heatmap depicting relative H3K27ac enrichments at promoter URT in a neuronal timecourse and different embryonic (E14.5) tissues / cell-types (CD4 T-lymphocytes, liver, MEFs, NIH3T3 cells). (I) Same as in (H) but depicted as boxplot.

Data information: Boxplots depicting RNA-seq as well as ChIP-seq data contain expression values scaled between 0 and 1 on y-axis. GEO IDs for all sequencing data used are provided in Table EV1.

**Appendix Figure S6. Identification of features that discriminate between NeuroD1-bound and non-NeuroD1-bound enhancers using Bayesian modeling.** (A) Distribution plot shows how ChIP-seq signal strength of all factors predicted by Bayesian model to discriminate targets and control (cutoff for information gain 0.1) correlates at enhancers associated to URT. Enhancers of URT are depicted in red while non-target enhancer associated to URG are depicted in cyan. Diagonal and top-right charts group show one and two dimensional density plots respectively. Bottom left charts depict scatter plot of ChIP-seq enrichments. (B) Receiver operating characteristic (ROC) plot for the Bayesian model developed to distinguish targets and control. (C) Scatter plot for the co-occupancy of MBD3 and TBX3 at enhancers associated to URT. (D) Heat map depicting the expression of enhancer URT in an *in vitro* neuronal RNA-seq time course using biological triplicates. Each row represents one promoter where gene expression is scaled from red (high expression) to blue (low expression). (E) Same as (D) but represented as boxplot. (F) Same as in (D) but for RNA-seq expression in various embryonic tissues. (G) Same as (F) but represented as a boxplot. (H) Heatmap depicting relative H3K27ac enrichments at enhancers associated to

URT in a neuronal timecourse and different embryonic (E14.5) tissues / cell-types (CD4 T-lymphocytes, liver, MEFs, NIH3T3 cells). (I) Same as in (H) but depicted as boxplot. (J) Same as in (I) but for intronic enhancers associated to URT. (K) Same as in (I) but for exonic enhancers associated to URT.

Data information: Boxplots depicting RNA-seq as well as ChIP-seq data contain expression values scaled between 0 and 1 on y-axis. GEO IDs for all sequencing data used are provided in Table EV1

**Appendix Figure S7. Expression dynamics of neuronal and pluripotency markers during iTN formation.** (A) RT-qPCR results for the expression dynamics of ectopic and endogenous NeuroD1 during iTN differentiation in continuously induced (+Dox) and control condition (-Dox). (B) RT-qPCR results for the expression dynamics of pluripotency markers during iTN differentiation in continuously induced (+Dox) and control conditions (-Dox) as well as wild type embryonic stem cells (ES). (C-G) RT-qPCR results for the dose-dependent induction (50, 500 and 1000 ng/ml) of ectopic NeuroD1 (C) and neuronal markers (D-G) during iTN formation. (H) RT-qPCR results for the expression dynamics of ectopic NeuroD1 during iTN differentiation in continuously induced condition (+Dox) and upon Dox removal after 48 h (+/- Dox) in comparison to non-induced control (-Dox).

Data information: RNA reflects the relative gene expression normalized to a housekeeping gene (Rpl19). The data are from two biological replicates and the error bars indicate the S.E.M.