Epigenome profiling and editing of neocortical progenitor cells during development

Mareike Albert, Nereo Kalebic, Marta Florio, Naharajan Lakshmanaperumal, Christiane Haffner, Holger Brandl, Ian Henry and Wieland B Huttner

APPENDIX TABLE OF CONTENTS

APPENDIX FIGURE LEGENDS

Appendix Figure S1. Marker gene expression in the neural cell populations.

mRNA expression of cell type-specific genes was analyzed by RT-qPCR in the defined neural cell populations. Two independent biological replicates are shown. Error bars represent SD of three qPCR amplifications.

Appendix Figure S2. Distribution of histone methylation at highly and lowly expressed genes.

The average read count for H3K4me3, H3K27me3 and Input in the defined neural cell populations was plotted across gene bodies for the 5,000 most highly (green) and lowly (orange) expressed genes using the ngs.plot package in R. Note that H3K4me3 marks are enriched around TSS of highly expressed genes, whereas H3K27me3 marks are broadly enriched across lowly expressed genes. Dark green and orange lines represent mean values; light green and orange shading represents SEM.

Appendix Figure S3. ChIP-seq tracks of replicate samples for H3K4me3 and H3K27me3.

Example genome browser tracks of ChIP-seq replicates for H3K4me3 and H3K27me3. For comparison, whole E14.5 brain (Brain) histone methylation from ENCODE/LICR is displayed. The replicates represent biological replicates of batches of 50,000 cells sorted from different pools of embryos for all samples, except for H3K27me3 of aRG-N where technical replicates are shown.

Appendix Figure S4. Validation of histone methylation signals by ChIP-qPCR for selected loci.

A H3K4me3 (green) and H3K27me3 (red) ChIP-seq tracks at selected genes in the defined neural cell populations and in whole E14.5 brain (Brain, from ENCODE/LICR). The ChIP-seq tracks for *Neurog2* and *Tubb3* (Figure 1B) are replotted here to facilitate comparison with ChIP-qPCR data.

B ChIP-qPCR for H3K4me3 (green) and H3K27me3 (red) in the defined neural cell populations for genes shown in (A). The data for the IgG control ChIP is identical for H3K4me3 and H3K27me3, and was obtained simultaneously. The location of the primers used for ChIP-qPCR is indicated in (A). Error bars represent SD of three qPCR amplifications. Similar results were obtained in two independent ChIP experiments.

Appendix Figure S5. Bivalent genes in neocortical development.

A, B GO term enrichment analysis of the genes that are H3K27me3-positive (A; n=2740) or bivalent (B; n=1166) in all the defined neural cell populations. The eight GO term clusters with the lowest p-value are shown (category: biological process). The number of genes present in the tested list over the number of genes in the reference gene set is indicated in the bars. The p-value is given next to each bar.

C Percentage of H3K27me3-posititve genes in the defined neural cell populations that are H3K27me3-positive only (red) or bivalent (yellow). Note that the percentage of bivalent genes is highest in NEC at E9.5.

D, E mRNA expression of genes that are bivalent in ESCs and remain bivalent in neurons at E14.5, analyzed at E15 (D) and P1 (E), presented as dot plots. Plots were generated using The Developing Cortical Neuron Transcriptome Resource (DeCoN: http://decon.fas.harvard.edu/ (Molyneaux et al., 2015)). The diameter of the dots refers to expression levels. Color represents relative neuron subtype-specific expression (CPN, callosal

projection neurons; ScPN, subcerebral projection neurons; CThPN, corticothalamic projection neurons). Genes with a >5-fold increase in gene expression from E15 to P1 in at least one of the neuronal subtypes are shown.

Appendix Figure S6. Histone methylation patterns at oligodendrocyte and astrocyte genes.

A H3K4me3 (green) and H3K27me3 (red) ChIP-seq signals around TSS in the defined neural cell populations. Genes expressed in oligodendrocyte progenitors (OPC), oligodendrocytes and astrocytes are shown. For comparison, E14.5 whole brain (Brain) ChIP-seq profiles from ENCODE/LICR are shown.

B mRNA expression of genes shown in (A), analyzed by RNA-seq. Error bars represent SD of four or five biological replicates.

C mRNA expression of genes shown in (A) in oligodendrocyte precursor cells (OPC), newly formed oligodendrocytes (new oligodendrocytes), myelinating oligodendrocytes (myel. oligodendrocytes), astrocytes and microglia isolated from postnatal mice at P7, using RNA-seq data from Zhang *et al.* (*J Neurosci* 2014; available at: http://web.stanford.edu/group/barres_lab/brain_rnaseq.html).

Appendix Figure S7. Genes with broad H3K4me3 domains are co-expressed and include induced neural stem cell reprogramming factors.

A Networks of genes with cell type-specific broad H3K4me3 domains (see Table EV5) were analyzed using the GeneMANIA Cytospace plugin. Lines connecting nodes (circles) represent a gene co-expression, co-variation and/or physical interaction relationship, as inferred by previous studies.

Genes connected by first-order interactions are shown. Genes in the GO term cluster ‘forebrain development’ are highlighted by yellow circles.

B Many induced neural stem cell (iNSC) reprogramming factors (7/11; see Table EV6) are marked by broad H3K4me3 domains in the developing mouse neocortex. The ChIP-seq tracks for *Hes1*, *Neurog2* and *Sox2* (Figure 1B) are replotted here for completeness of the list of known iNSC factors.

C mRNA expression of genes shown in (B), analyzed by RNA-seq. Error bars represent SD of four or five biological replicates.

Appendix Figure S8. Principle component analysis of ChIP-seq replicates.

A, B Principle component analysis (PCA) of H3K4me3 (A) and H3K27me3 (B) ChIP-seq replicate samples. The percentage of variance covered by the first two components is indicated.

Appendix Figure S9. Differential gene expression.

A Number of genes that are differentially expressed between the defined neural cell populations, determined by RNA-sequencing. RNA-seq data for RG and neurons are from (Florio et al., 2015), those for mouse NECs were determined here. Red, up-regulated genes; blue, down-regulated genes.

B Number of genes that are enriched within the defined neural cell populations.

C Gene ontology term enrichment analyses of genes enriched in the defined neural cell populations (see B). The three most enriched GO term clusters are shown (category: biological process). GO term enrichment

analysis was performed using DAVID (http://david.abcc.ncifcrf.gov), with default settings.

D Heat map depicting the mRNA expression levels (FPKM values) of selected cell type-specific genes, determined by RNA-seq.

E mRNA expression (in FPKM) of genes that gain (left) or loose (right) H3K27me3 from aRG-P to aRG-N, shown as box plots.

F mRNA expression (in FPKM) of genes that gain (left) or loose (right) H3K27me3 from bRG to N, shown as box plots.

Data information: Boxes represent first quartile (bottom), median and third (top) quartile; whiskers refer to 10th and 90th percentiles. Log scale was used to facilitate viewing of FPKM values in the lower range. Significance relative to aRG-P (E, orange) or relative to bRG (F, orange) was calculated using a Kruskal-Wallis test; no significant change was observed.

Appendix Figure S10. *In vitro* testing of *Eomes* gRNAs.

A Schematic overview of the localization of PCR templates and gRNAs at the *Eomes* locus. gRNAs were chosen to target the CpG islands upstream (gRNA 1-3) and downstream (gRNA 4-6) of the *Eomes* TSS.

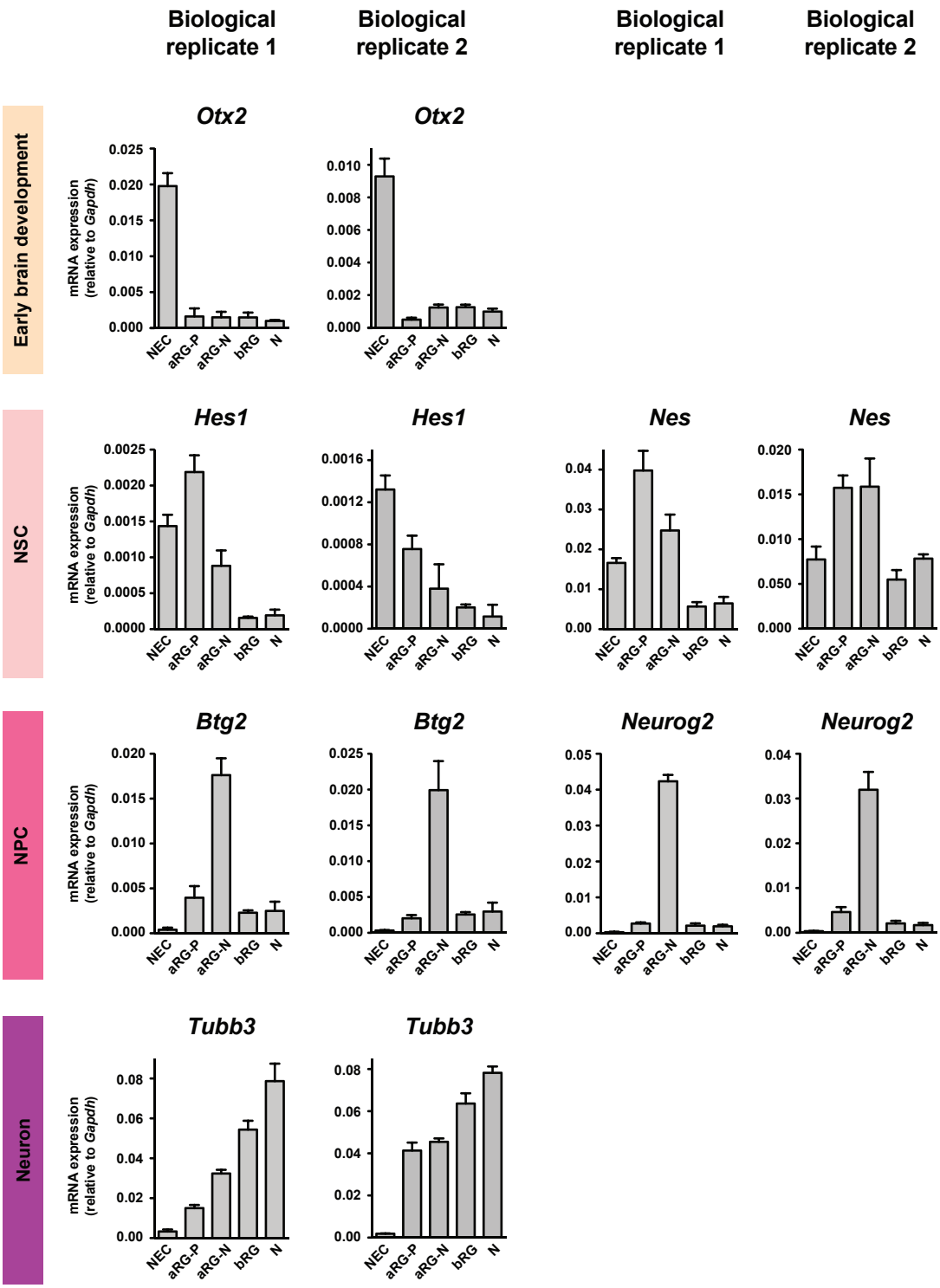
B Scheme of PCR templates used for *in vitro* testing of guide RNA efficiencies. The sizes of PCR templates, guide RNA binding sites and expected sizes of cut fragments are indicated.

C Agarose gel showing the effects of the gRNAs 1–6 to direct Cas9-mediated cutting of PCR templates (see A and B). Uncut templates were loaded in lane 5 (template 1-3) and lane 9 (template 4-6). gRNAs 2 and 4 (red arrows) were selected for *in vivo* epigenome editing.

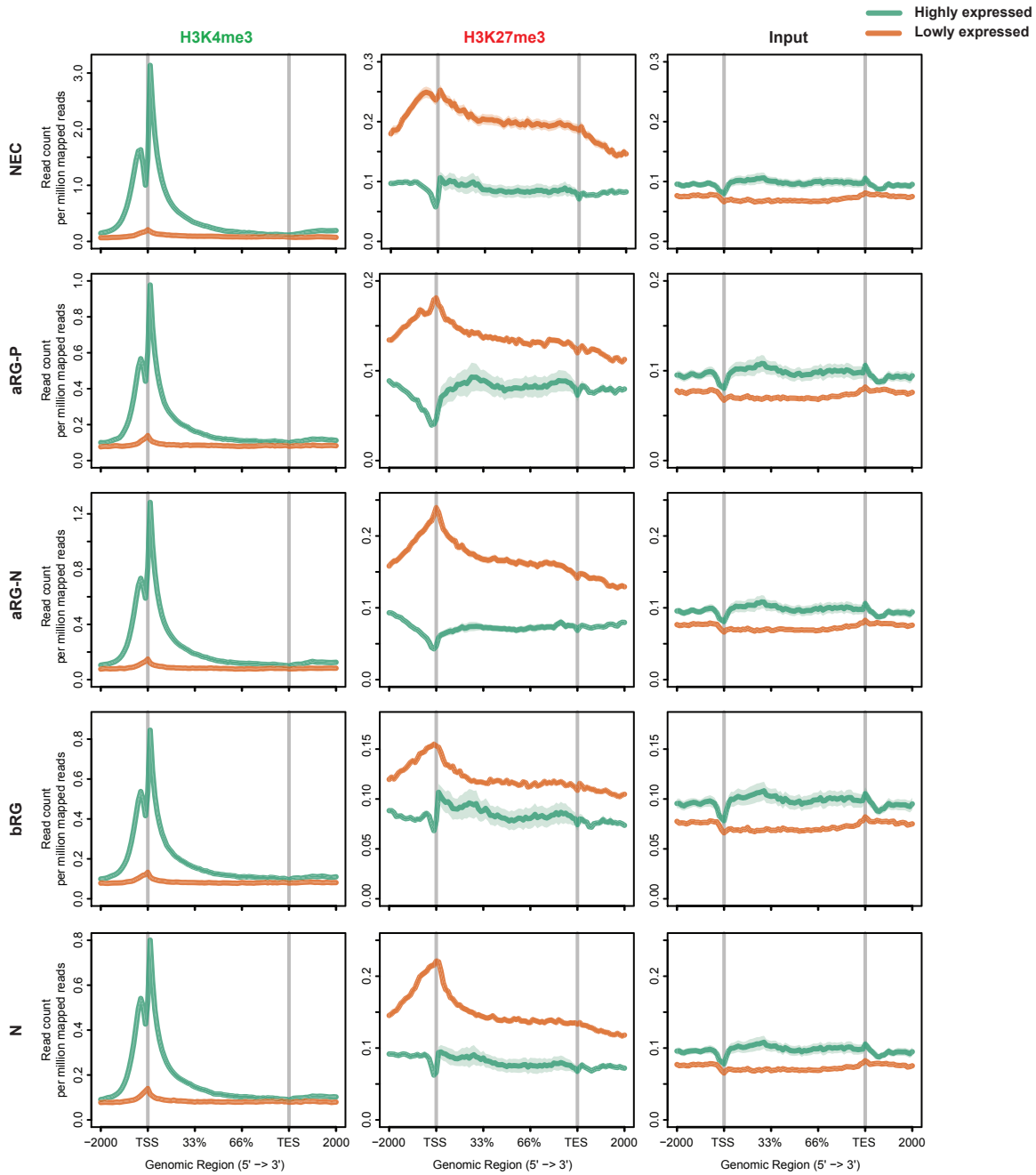
D Quantification of the percentage of dCas9-positive cells that are Tbr2-positive 48 hours after g*LacZ* dCas9-Ezh2 (white), g*Eomes* dCas9-Ezh2 (black) or g*Eomes* dCas9-Ezh2* mutant (grey) IUE. Data for individual IUE experiments (biological replicates) is shown.

E Quantification of the number of dCas9-positive cells (based on RFP) per microscopic field 48 hours after g*LacZ* dCas9-Ezh2 (white), g*Eomes* dCas9-Ezh2 (black) or g*Eomes* dCas9-Ezh2* mutant (grey) IUE. Data represent mean ± SD. **; ns, not significant.

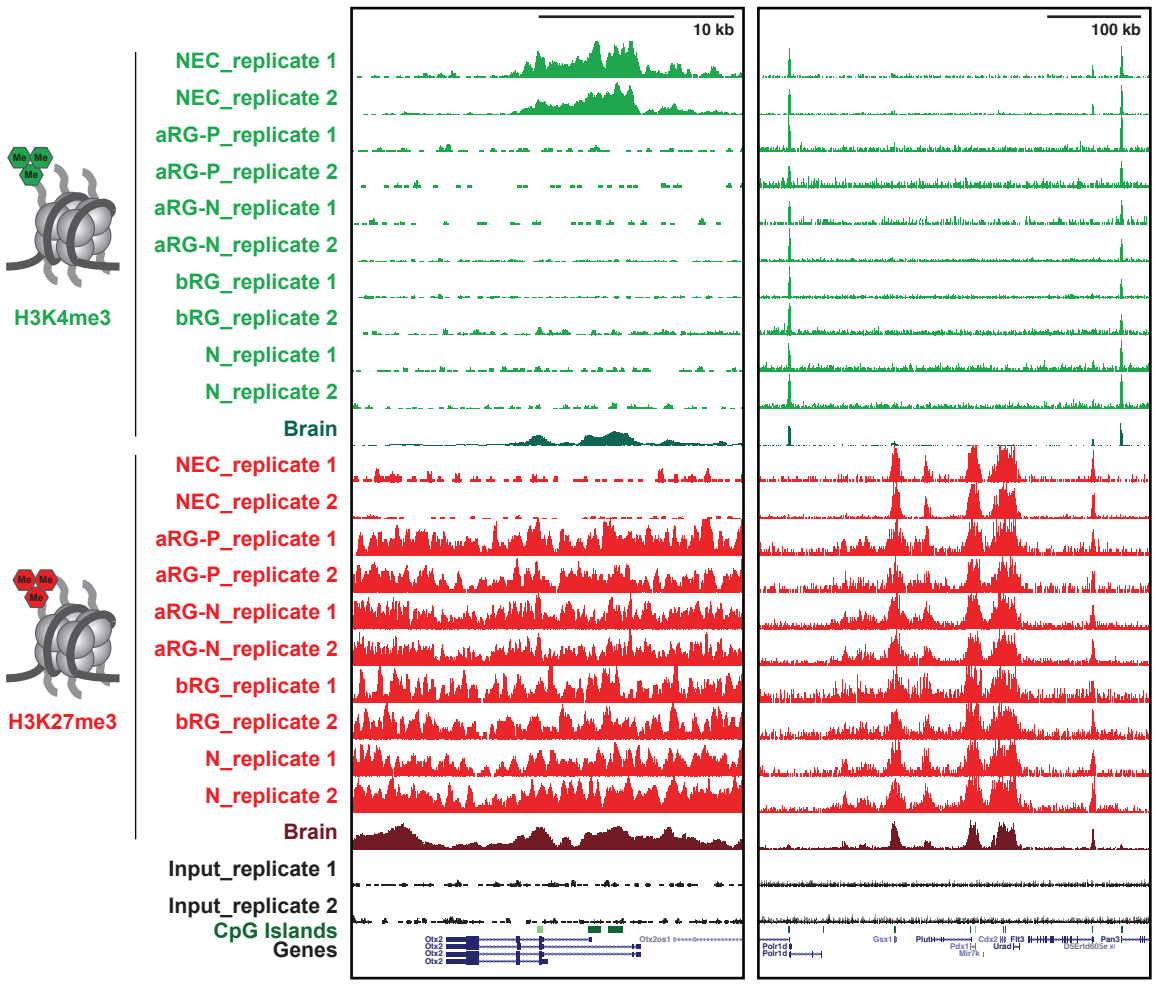
F Fold changes, expressed as log (logFC), of H3K27me3 levels around the TSS of *Eomes* and an unrelated gene (*Actb*) after IUE (related to Fig 7G). Error bars represent SD of biological replicates (n=4 IUE experiments). Fold changes are relative to each respective control (g*Eomes* dCas0-Ezh2). Significance was calculated using a One sample *t*-test. *, $p<0.05$; **, $p<0.01$; ***, $p<0.001$; ns, not significant.



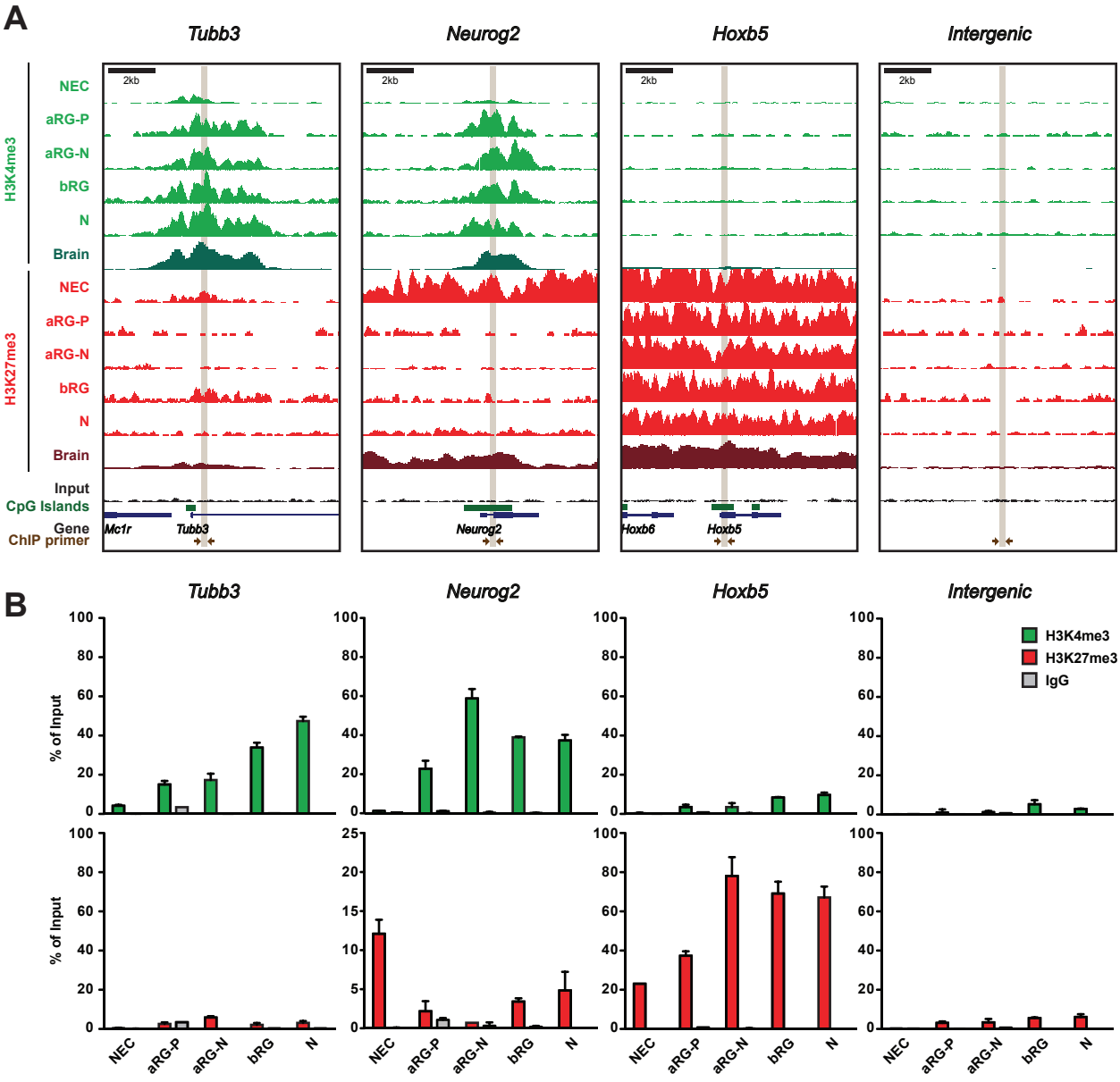
Appendix Figure S1



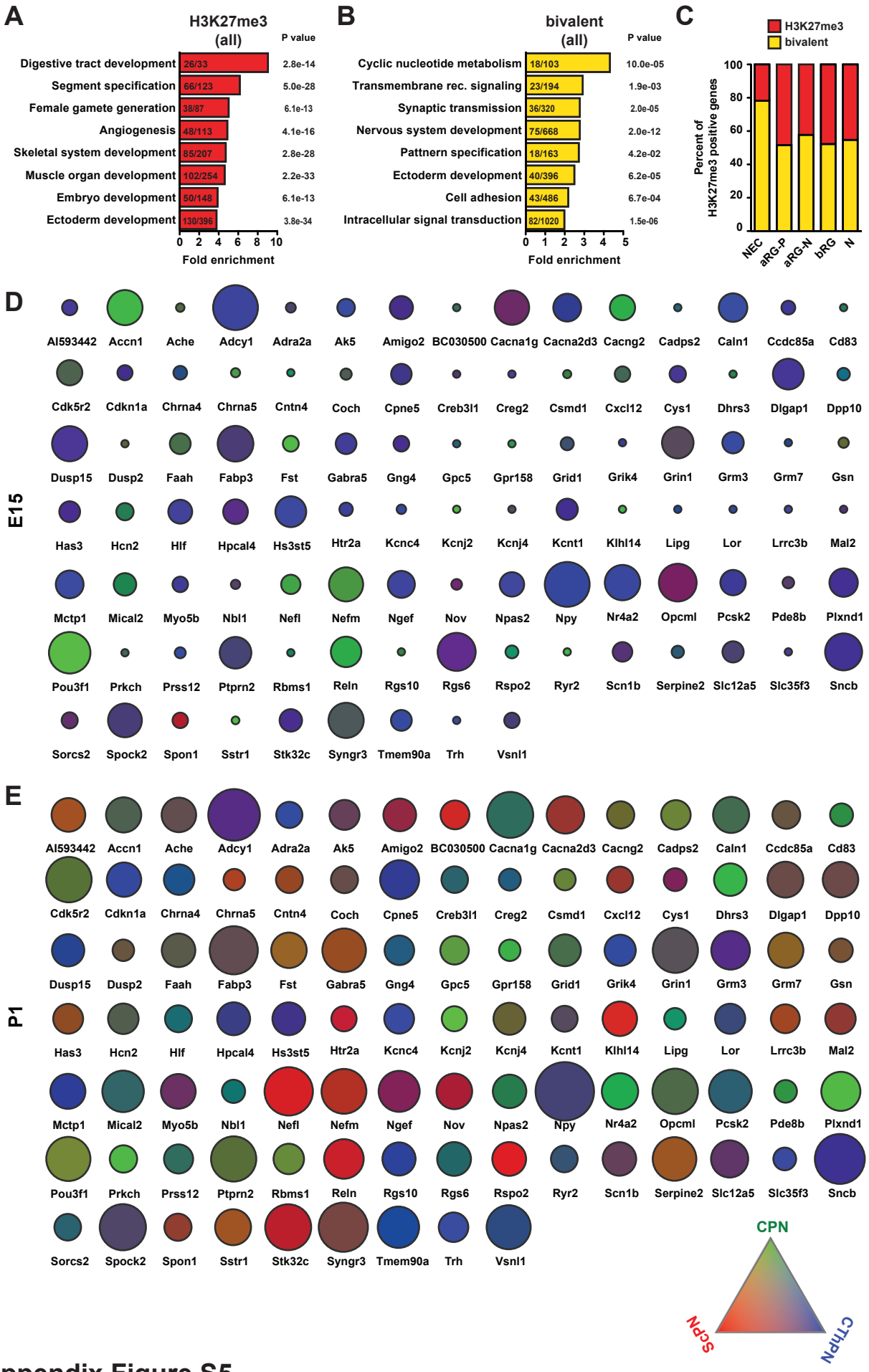
Appendix Figure S2



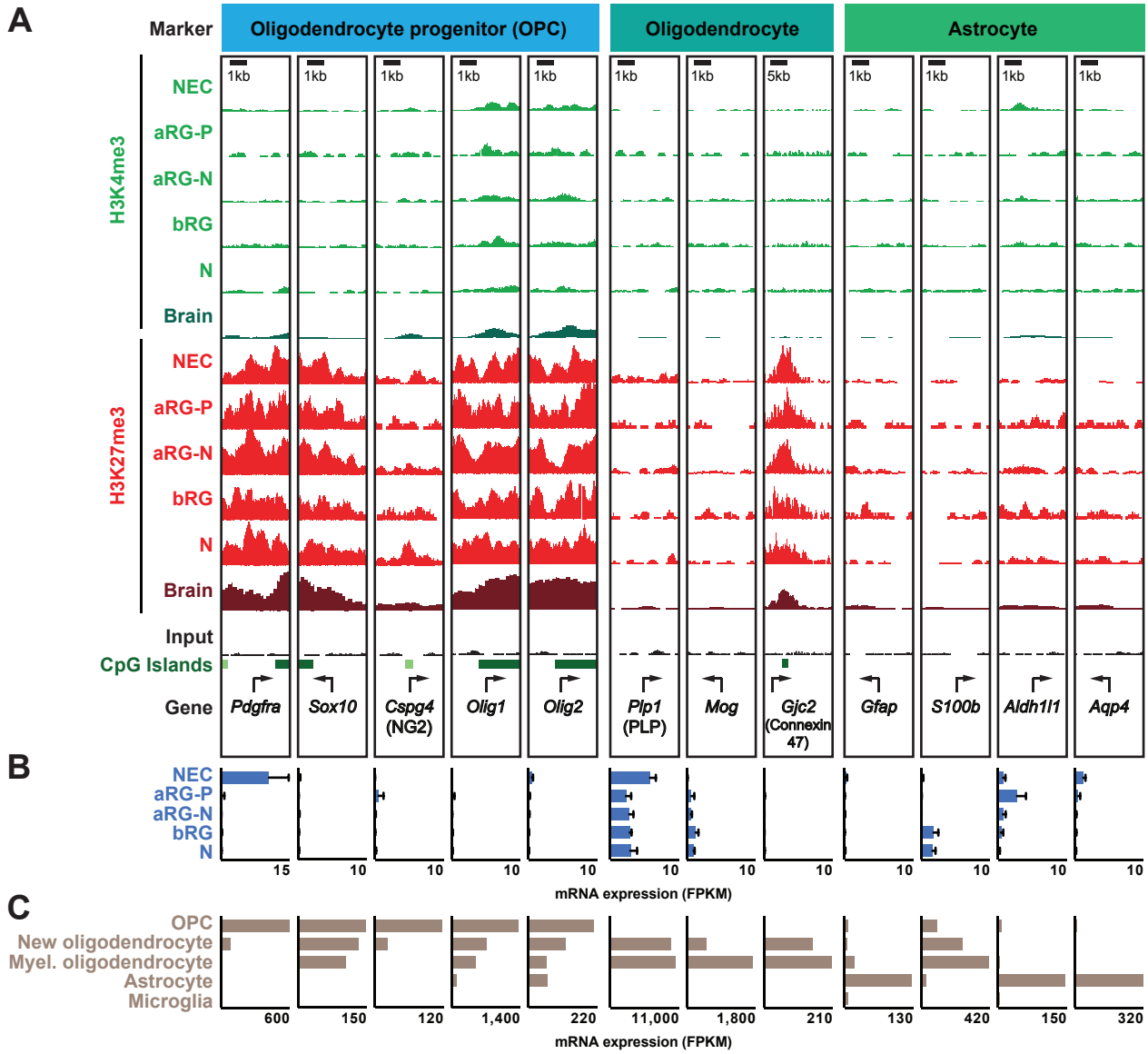
Appendix Figure S3



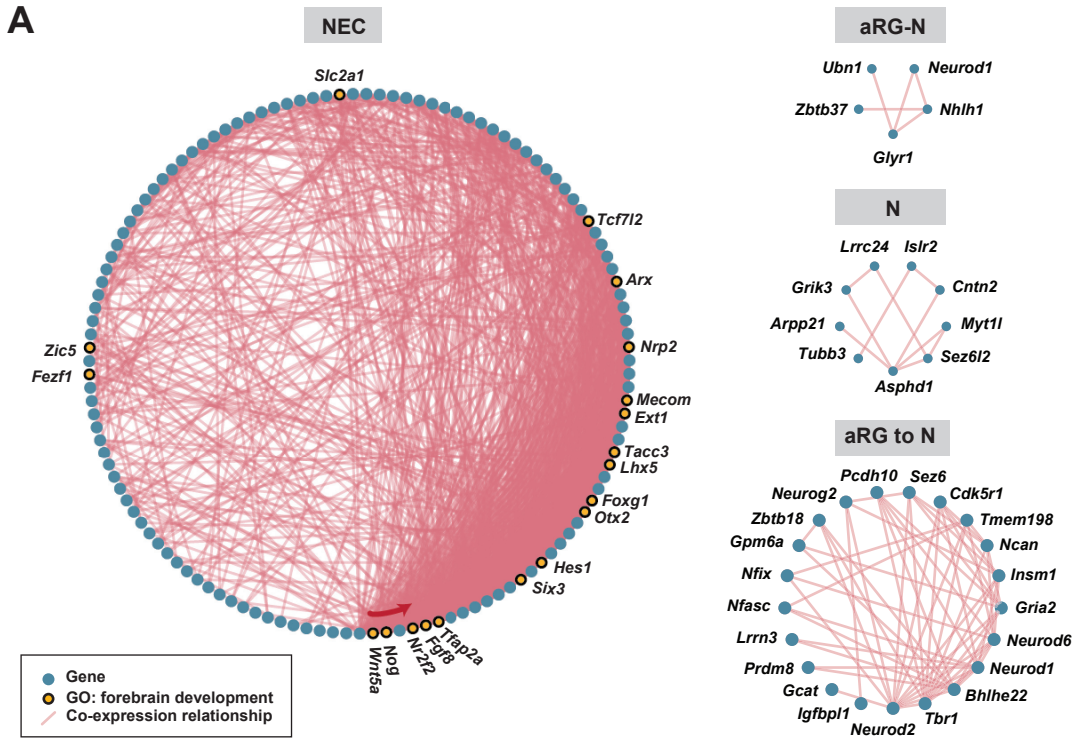
Appendix Figure S4

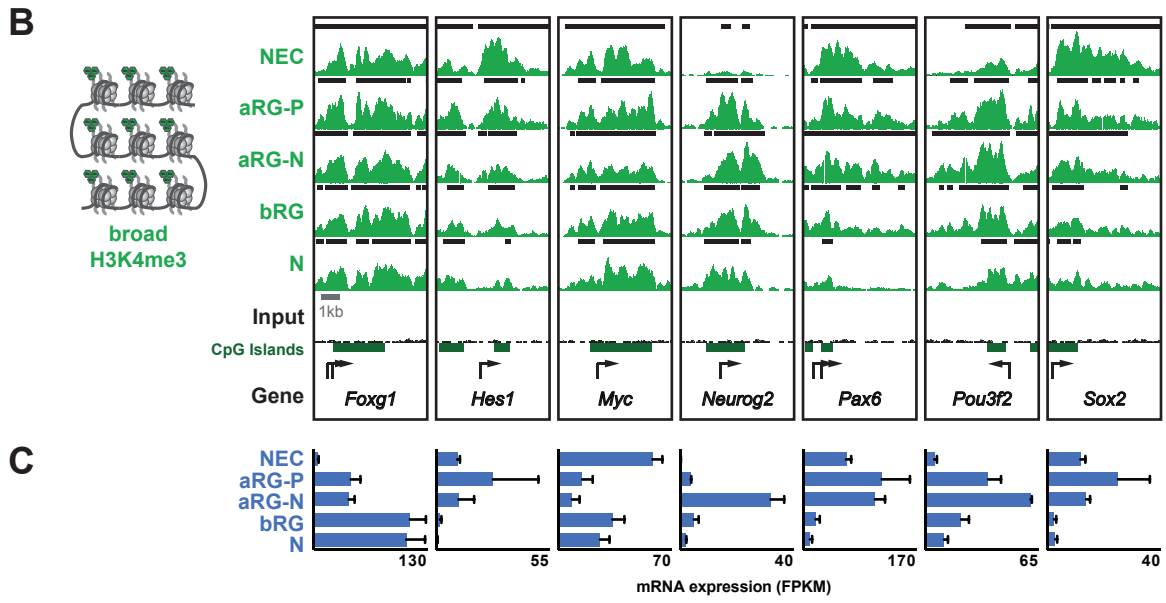


Appendix Figure S5

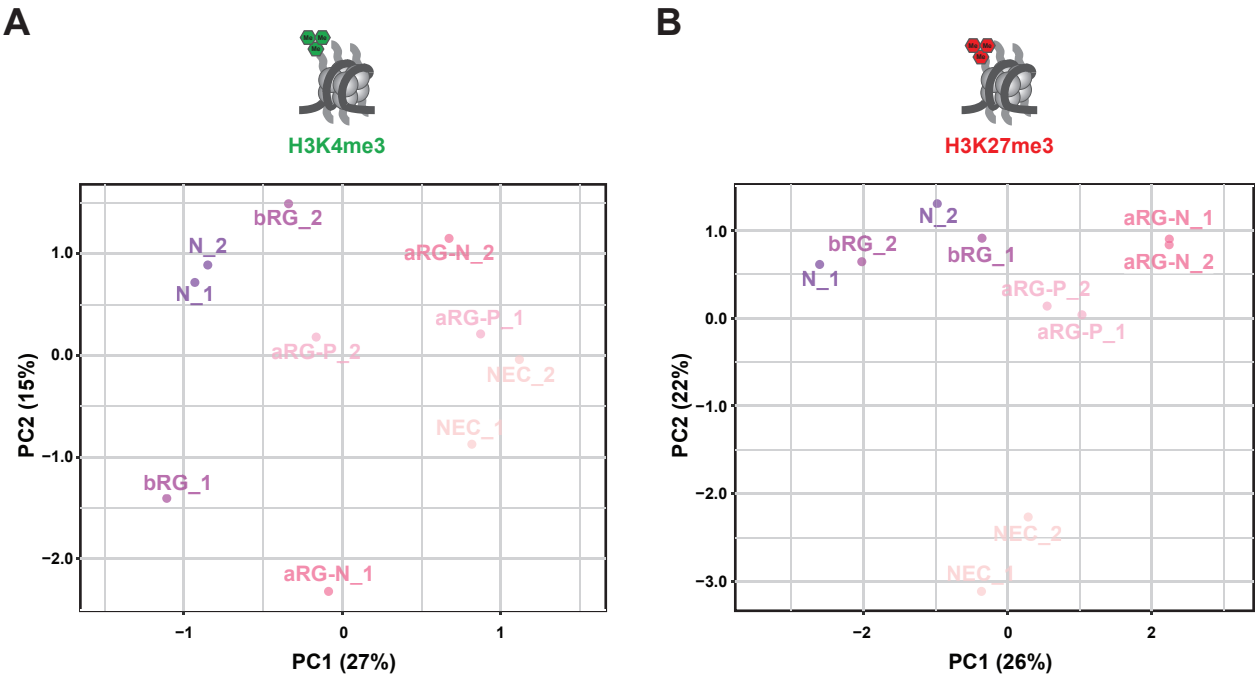


Appendix Figure S6

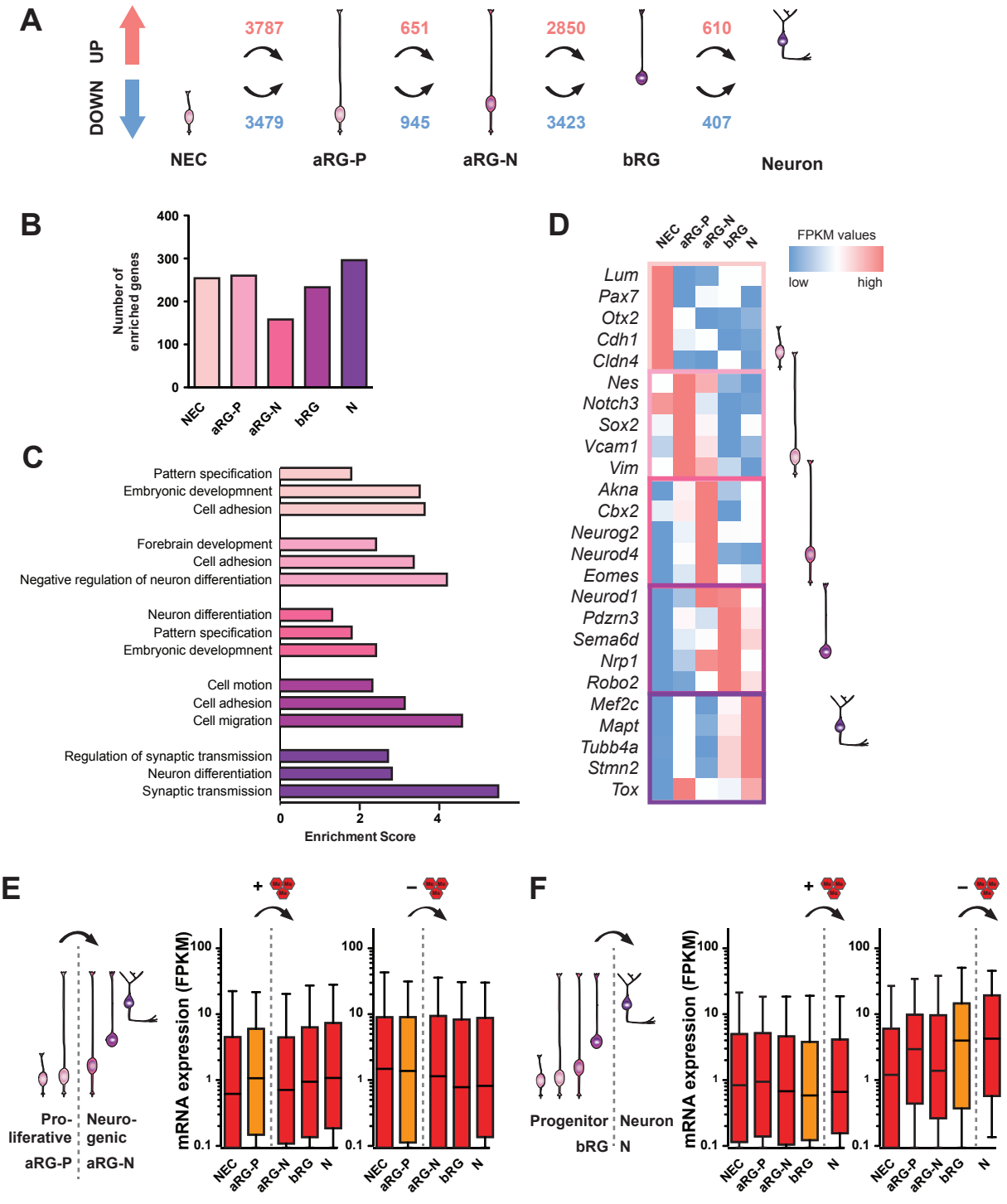




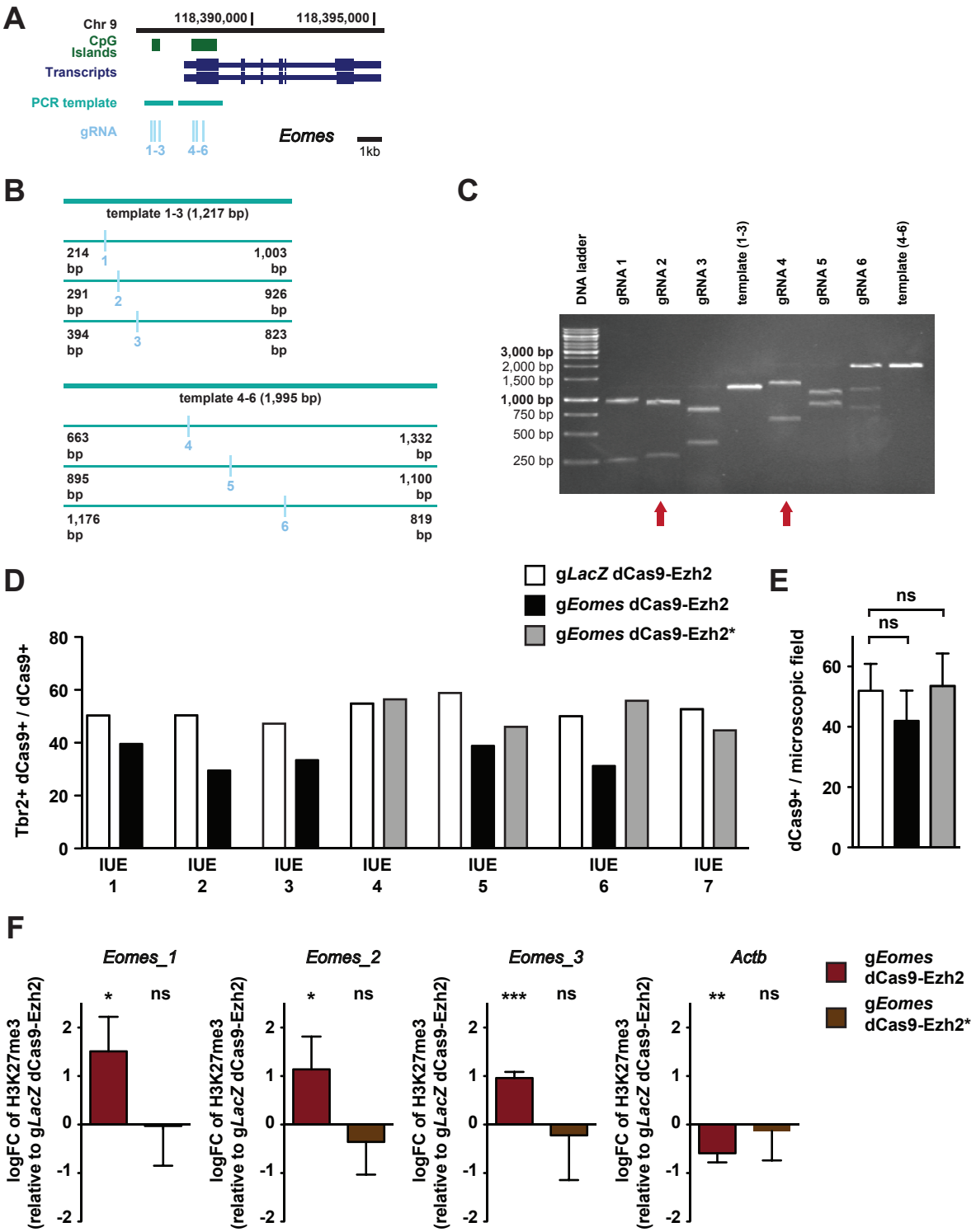
Appendix Figure S7



Appendix Figure S8



Appendix Figure S9



Appendix Figure S10