

Figure S1

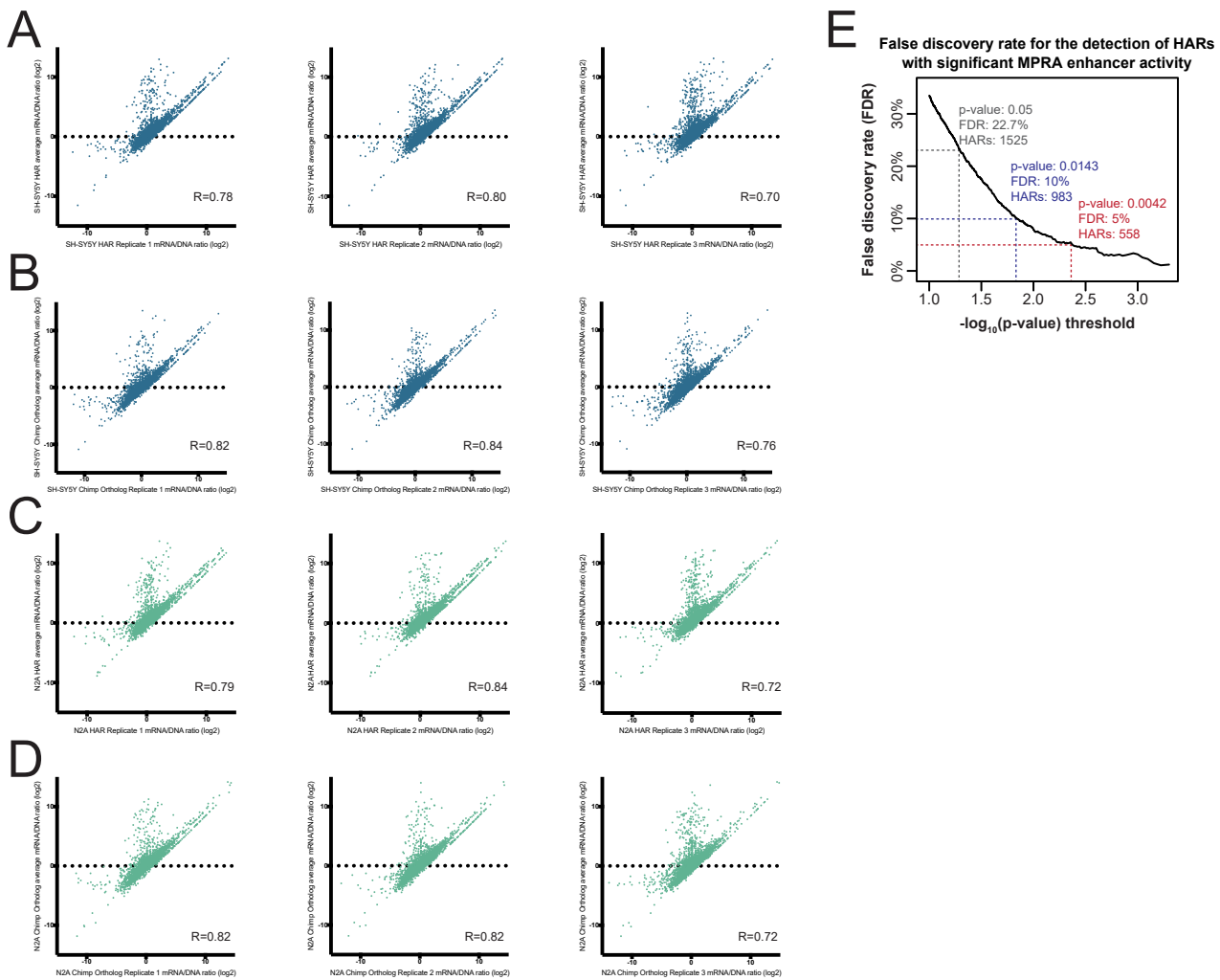














Figure S1 | caMPRA measurements are highly reproducible across replicates, Related to Figure 1.

mRNA/DNA ratio for all replicates (x-axis) compared the condition's average ratio (y-axis) for all HARs and chimpanzee orthologous sequences demonstrates high reproducibility with correlations above 0.8 for most comparisons. For each cell line and species sequence tested, three replicates are shown. **(A)** Human HAR sequence relative activity reproducibility in human neuroblastoma cells (SH-SY5Y). **(B)** Chimpanzee ortholog sequence reproducibility in human neuroblastoma cells (SH-SY5Y). **(C)** Human HAR sequence relative activity reproducibility in mouse neuroblastoma cells (N2A). **(D)** Chimpanzee ortholog sequence reproducibility in mouse neuroblastoma cells (N2A). **(E)** False discovery rate (FDR) for the detection of HARs with significant enhancer activity at either the chimpanzee or human ortholog in either SY-SY5Y or N2A cells. a p-value cutoff of 0.05 was used for most analyses in this paper, which has an FDR of 22%.

Figure S2

A

Known Motif	TF	p-value	% with motif
	RBPJ	1.00E-03	18.91%
	TEAD	1.00E-02	12.38%
	TBR2	1.00E-02	37.23%
	PAX7	1.00E-02	4.26%
	ATOH1	1.00E-02	17.33%
	PAX6	1.00E-02	2.57%

Known Motif	TF	p-value	% with motif
	HSF1	1.00E-03	1.78%
	LRXE	1.00E-03	1.19%
	BACH1	1.00E-02	0.89%
	STAT5	1.00E-02	6.14%
	STAT4	1.00E-02	16.73%
	HIF2A	1.00E-02	4.06%

B

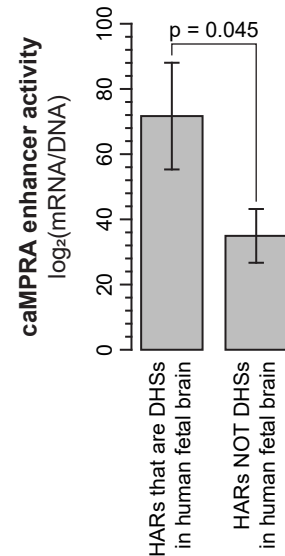


Figure S2 | TF binding elements enriched in HARs that gain activity in humans, Related to Figure 1.

(A) Motif enrichment for transcription factor binding motifs within HARs that gained activity in comparison to their chimpanzee ortholog in captureMPRA. The p-value and the percentage of HARs containing the motif are located to the right of each transcription factor. Motifs are grouped into transcriptional regulators known to play key roles in neurogenesis (top) and cell homeostasis (bottom). Only TFs that were significant are shown.

(B) Mean caMPRA signal at HARs that do or do not show features of in vivo enhancer activity in fetal brain tissue, as indicated by DNaseI hypersensitivity. Error bars represent standard error of the mean. p-value calculated using two-sided t-test

Figure S3

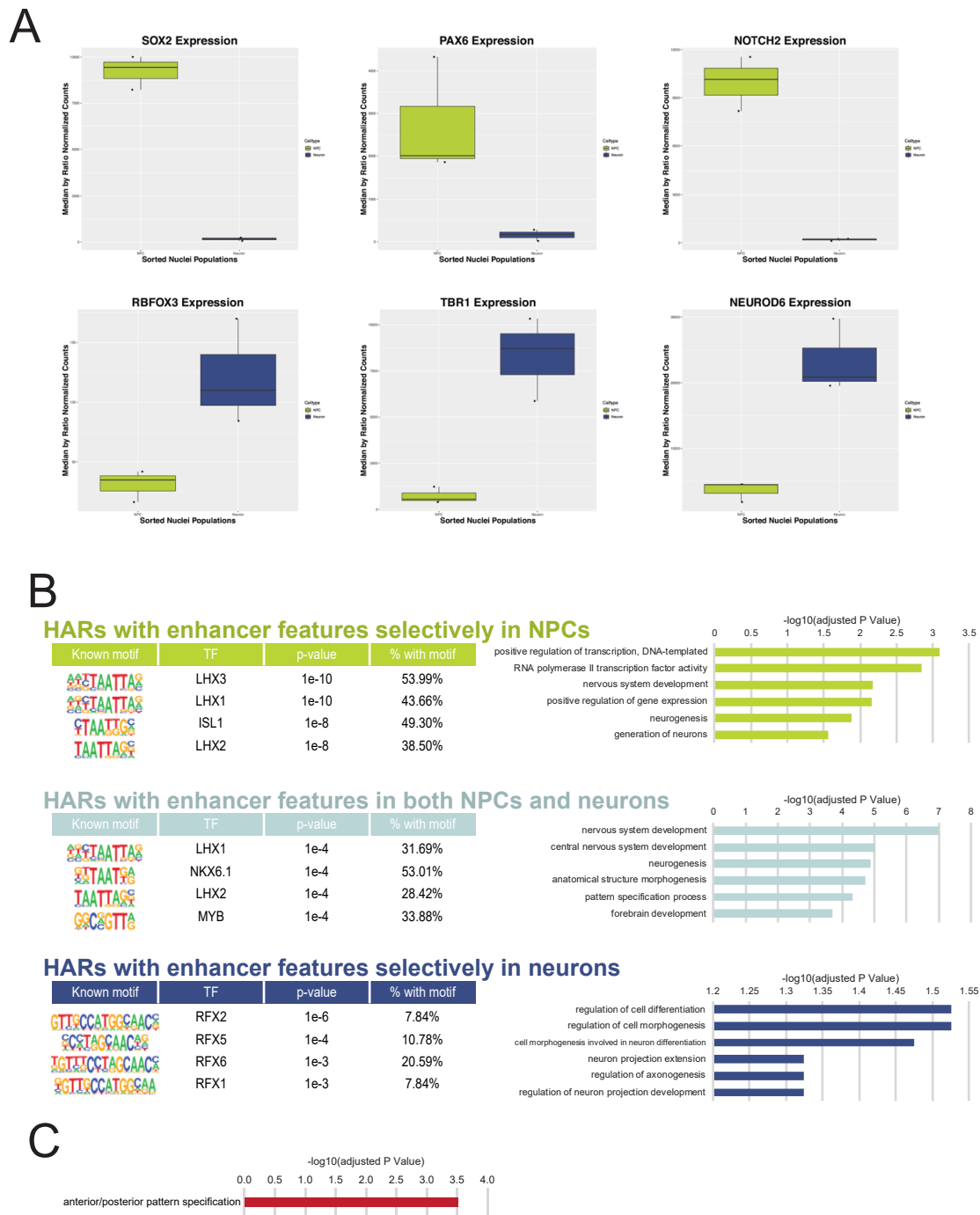


Figure S3 | HARs contain cell-type-specific neurodevelopmental regulatory elements, Related to Figure 2.

(A) Boxplots of normalized RNA transcript counts for both sorted NPC (green) and neurons (blue) for (top) three canonical NPC genes (SOX2, PAX6, NOTCH2) and (bottom) three canonical neuronal genes (RBFOX3/NeuN, TBR1, NEUROD6).

(B) (left) Motif enrichment for transcription factor binding motifs within HARs that are selectively active in NPCs (top), neurons (bottom) or both (middle). (right) Gene-Ontology sequencing (GO-seq) analysis of genes most proximal to these HARs.

(C) For HARs with enhancer features in NPCs and/or neurons, presented is the GOzilla GO-enrichment of genes with the strongest Hi-C interaction in human fetal cortical plate

Figure S4

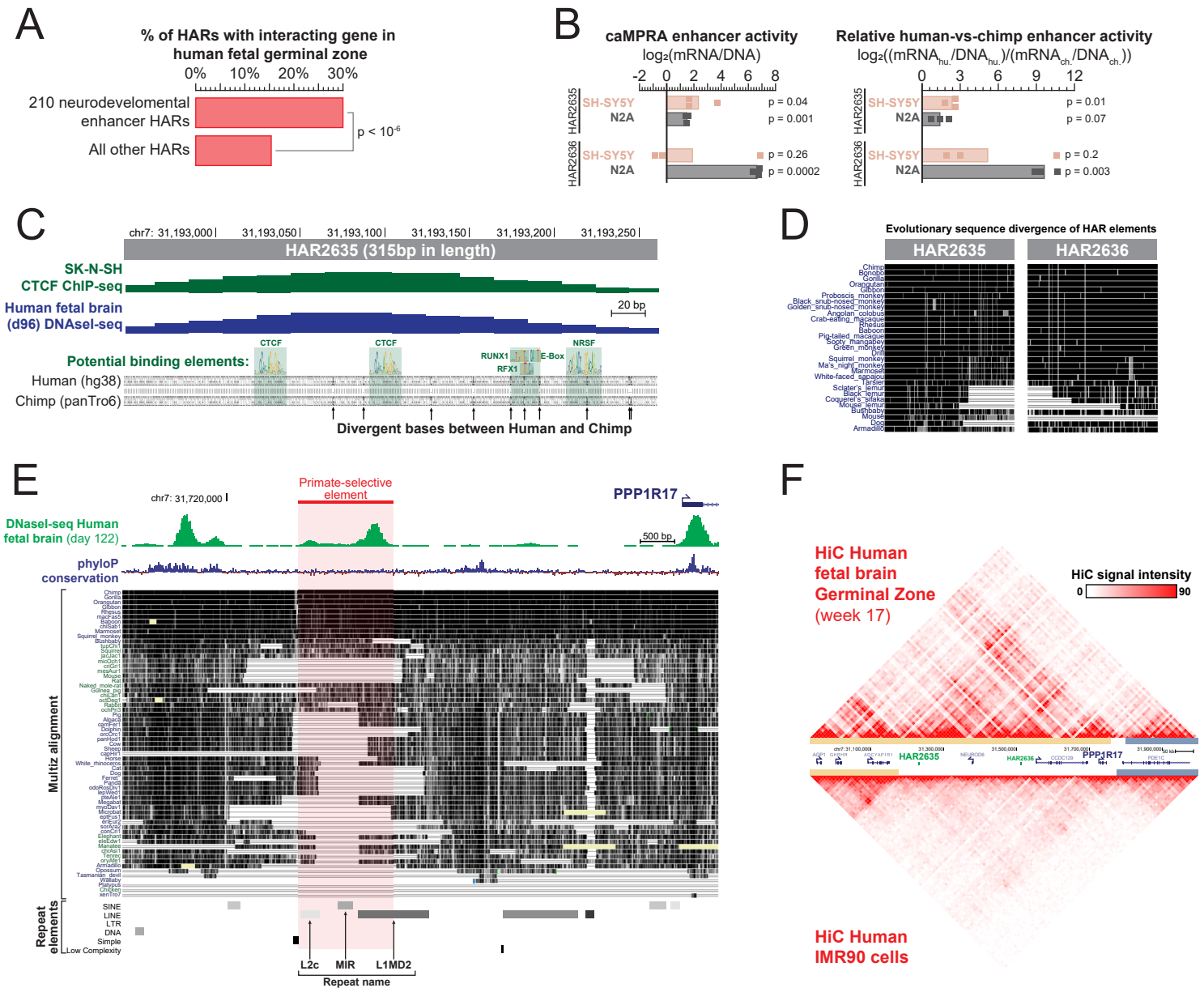


Figure S4 | HAR2635 is a neurodevelopmental enhancer, Related to Figure 3.

(A) Percentage of HARs with an interacting gene in human fetal brain germinal zone based on HiC data, as a function of whether the HAR demonstrates both *in vitro* and *in vivo* enhancer activity as defined in Figure 3A. p-value calculated using Z-test.

(B) CaptureMPRA activity of both HAR2635 and HAR2636 in SH-SY5Y and N2A cells, as well as relative human-vs-chimp activity. Data from each replicate is indicated by points. p-values calculated using a 1-sided t-test to detect deviation from 0.

(C) HAR2635 shares 96.8% genomic similarity to its ortholog in chimp, with 10 base pair changes over its 315 bp length. Comparison of transcription factor binding motifs between these species shows a gain of several TF binding elements in the human lineage, as well as a preservation of CTCF binding elements between both lineages.

(D) Sequence divergence of HAR2625 and HAR2636 between primate and non-primate species. Black segments represent bases conserved in that species.

(E) Genomic locus of *PPP1R17* showing human fetal brain DNase-seq signal in addition to evolutionary conservation and repeat elements. Red box indicates primate-selective regulatory element.

(F) HiC Interaction frequencies of the *PPP1R17* locus in both (top) human fetal brain germinal zone as well as (bottom) human fetal lung fibroblast (IMR90) cells. Scale identical for both graphs.

Figure S5

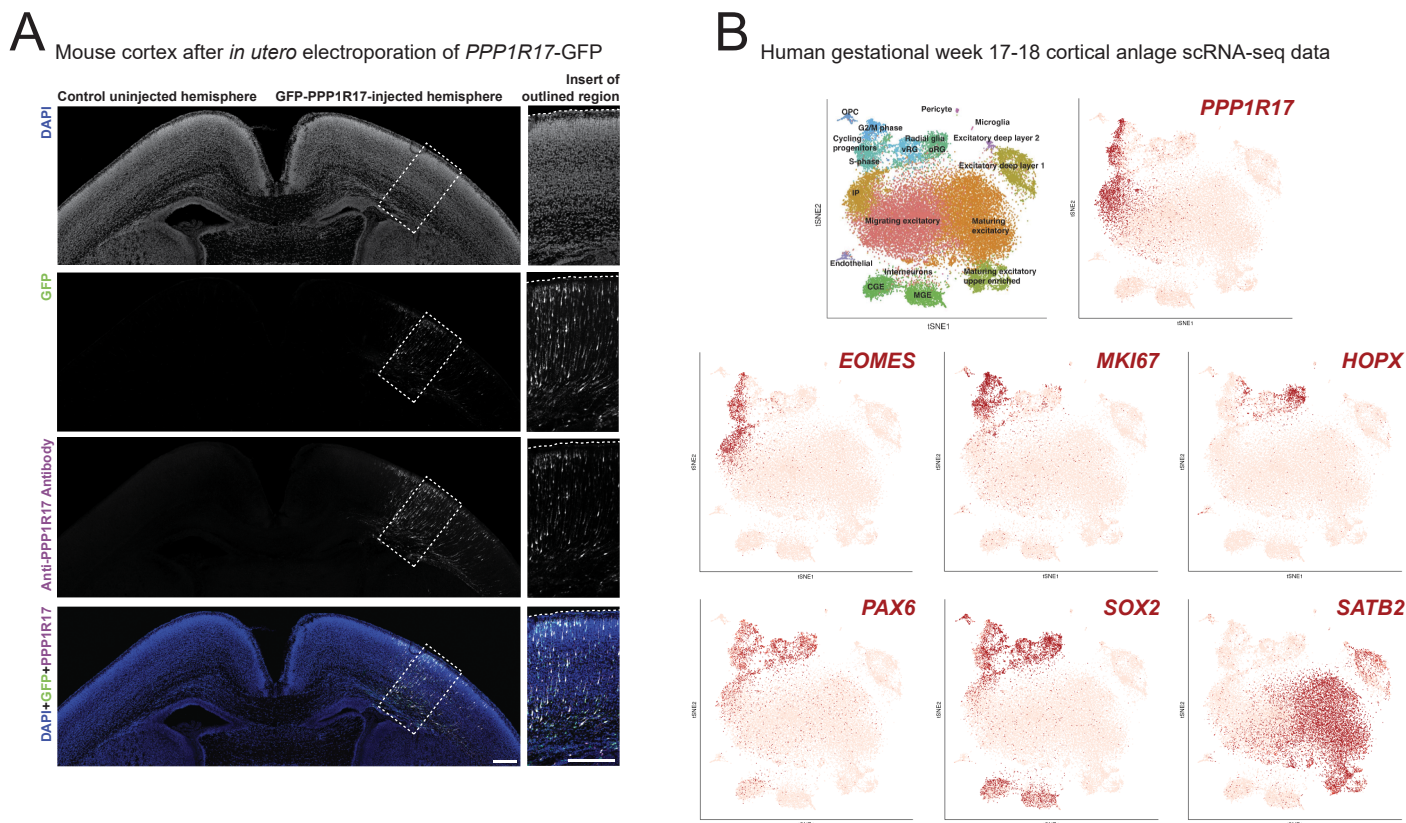


Figure S5 | Divergent expression of *PPP1R17* cortical expression, Related to Figure 4.

(A) Lack of *PPP1R17* expression in the mouse cortex as demonstrated by *in utero* electroporation of *GFP-PPP1R17* construct in only a single mouse cortex. Staining for GFP, *PPP1R17*, or the DNA marker DAPI demonstrates *PPP1R17* staining selectively within the region injected with the *GFP-PPP1R17* construct.

(B) tSNE plot from single cell RNA-seq data obtained using human gestational week 17-18 cortical anlage tissue (Polioudakis et al., Neuron 2019). Individual cell types are labeled according to Polioudakis et al. Overlap of *PPP1R17*, *EOMES* (*TBR2*), *MKI67* (*KI67*), *SOX2*, *PAX6*, *HOPX* and *SATB2* expression demonstrating that *PPP1R17* expression is largely restricted to proliferating intermediate progenitor cells and not mature neurons or radial glia.

Figure S6

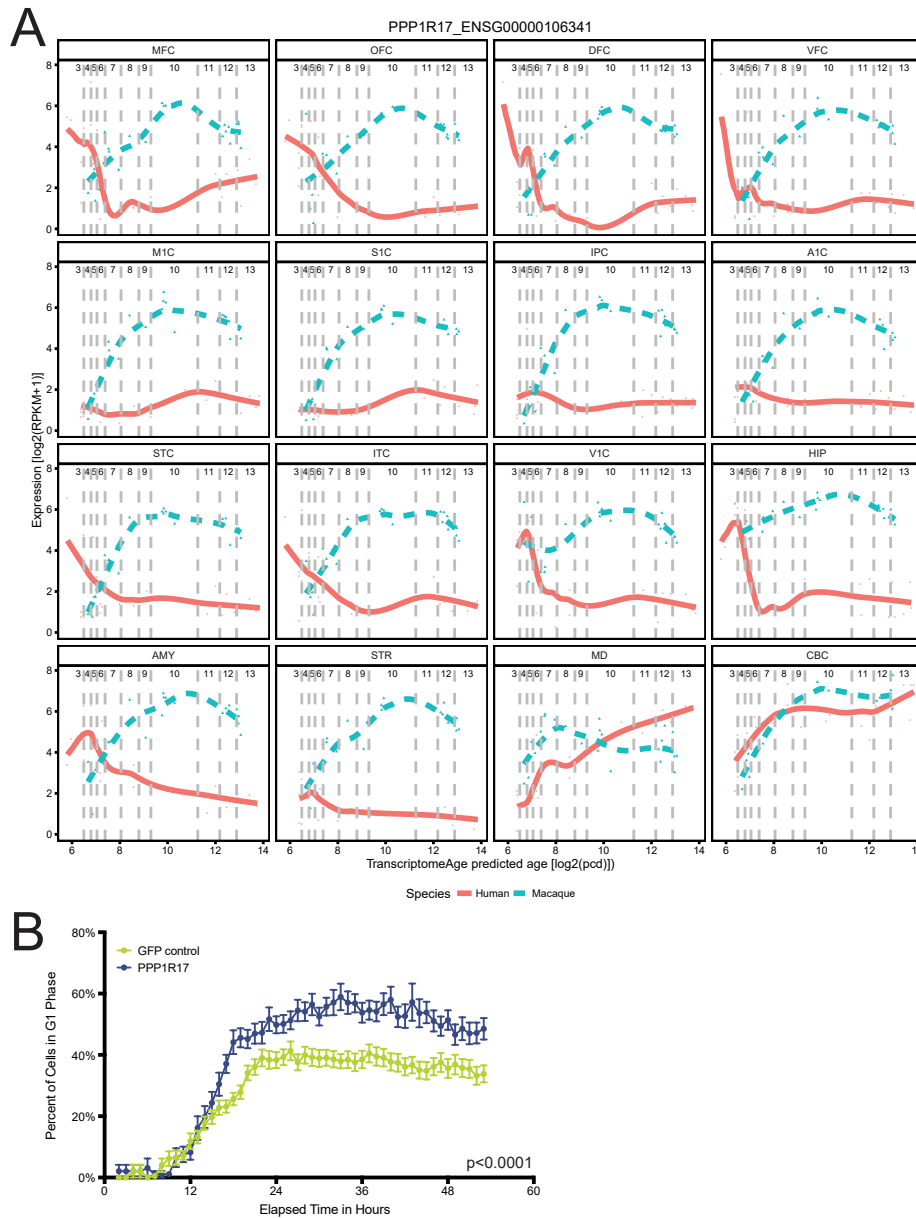


Figure S6 | PPP1R17 expression across the developing brain in humans and macaque, Related to Figures 5 and 6.

(A) PPP1R17 expression using single-cell RNA-sequencing at various developmental time stages in human and macaque brain development from fetal to adult. Brain regions profiled are demonstrated above, and the RNA expression patterns as a function of developmental time (calculated using TranscriptomeAge predicted age as defined in Zhu et al., 2018) are shown below. Brain region abbreviations: Medial prefrontal cortex, MFC; orbital prefrontal cortex, OFC; dorsolateral prefrontal cortex, DFC; ventrolateral prefrontal cortex, VFC; primary motor cortex, M1C; primary somatosensory cortex, S1C; inferior posterior parietal cortex, IPC; primary auditory cortex, A1C; superior temporal cortex ITC; primary visual cortex, V1C; hippocampus, HIP; amygdala, AMY; striatum, STR; mediodorsal nucleus of the thalamus, MD; cerebellar cortex, CBC.

(B) Labeling with cell cycle indicator demonstrates increase of G1 phase in *PPP1R17*⁺ cells ($p < 0.0001$, 2-way ANOVA). Percent of cells expressing both the G1 marker and either GFP or PPP1R17-GFP were quantitated from 4 images per well, for each of $n = 13$ GFP-transfected wells, $n = 13$ PPP1R17-transfected wells. Error bars = SEM.

Figure S7

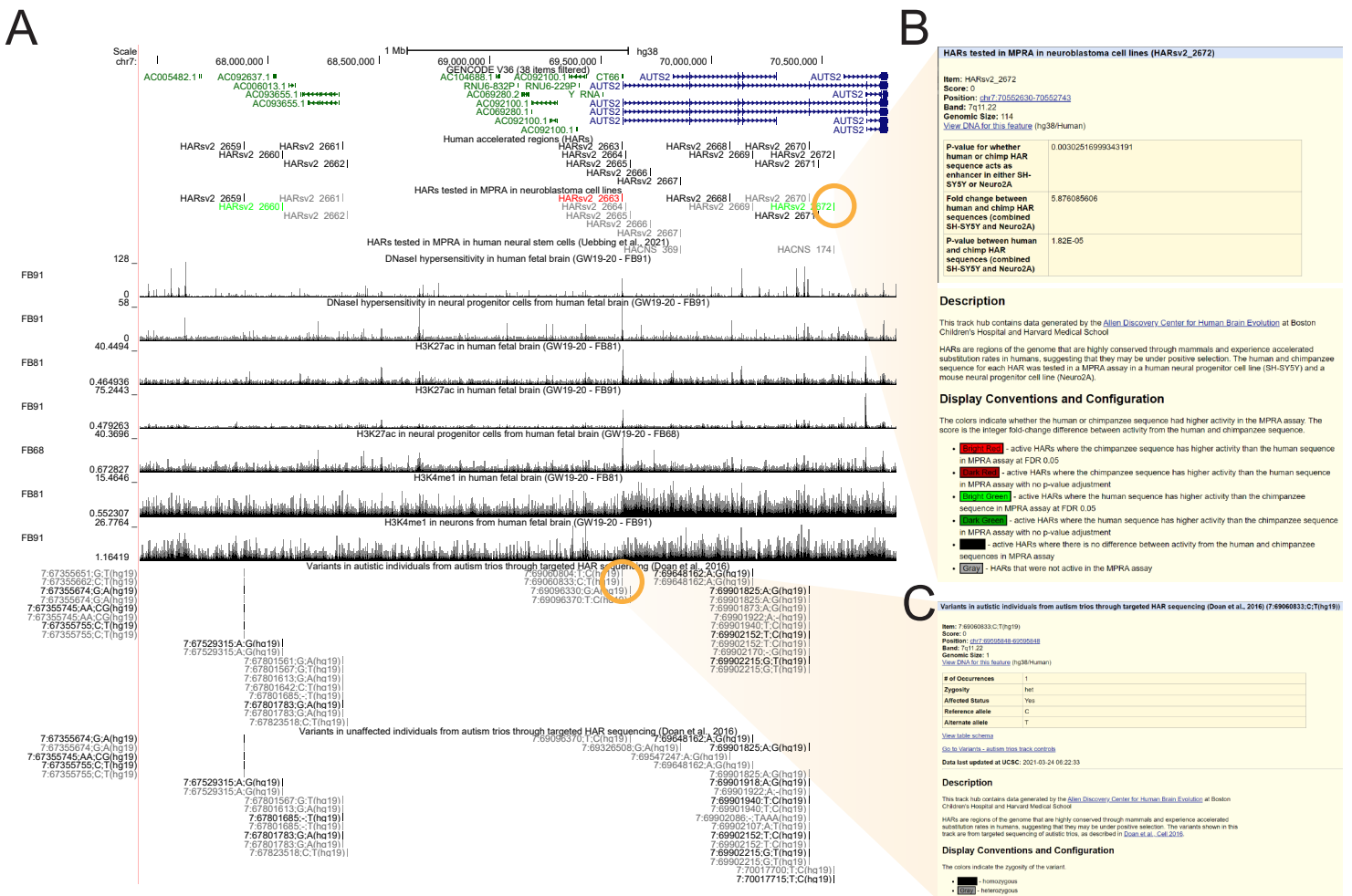


Figure S7 | HARhub UCSC genome track hub, Related to Figures 1 and 2.

(A) UCSC genome browser shot showing the *AUTS2* locus, which has been previously associated with autism. Shown are representative tracks of HARs tested in this study versus the Uebbing et al., 2021 study. Also shown are chromatin accessibility and histone mark patterns, in addition to genetic variation at these HAR elements.

(B) Upon clicking on one of the HARs, information regarding that HAR will be displayed, including the relative enhancer activity of the human versus chimpanzee paralog.

(C) Upon clicking on one of the HAR SNPs, the genetic details of that SNP will be displayed.