LINE-1 retrotransposon activation intrinsic to interneuron development

Gabriela O. Bodea[1,2,*], Maria E. Ferreiro[1], Francisco J. Sanchez-Luque[3,4], Juan M. Botto[1], Jay Rasmussen[1], Muhammed A. Rahman[1], Laura R. Fenlon[1], Carolina Gubert[5], Patricia Gerdes[2], Liviu-Gabriel Bodea[6], Prabha Ajjikuttira[1], Peter Kozulin[1], Victor Billon[1,7], Santiago Morell[2,†], Marie-Jeanne H.C. Kempen[4], Chloe J. Love[5], Lucy M. Palmer[5], Adam D. Ewing[2], Dhanisha J. Jhaveri[1,2], Sandra R. Richardson[2], Anthony J. Hannan[5], Geoffrey J. Faulkner[1,2,*]

[1]Queensland Brain Institute, University of Queensland, Brisbane QLD 4072, Australia.
[2]Mater Research Institute - University of Queensland, TRI Building, Woolloongabba QLD 4102, Australia.
[3]GENYO. Pfizer-University of Granada-Andalusian Government Centre for Genomics and Oncological Research, PTS Granada 18016, Spain.
[4]MRC Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, United Kingdom.
[5]Florey Institute of Neuroscience and Mental Health, University of Melbourne, Parkville VIC 3052, Australia.
[6]Clem Jones Centre for Ageing Dementia Research, Queensland Brain Institute, University of Queensland, Brisbane QLD 4072, Australia.
[7]Biology Department, École Normale Supérieure Paris-Saclay, 91190 Gif-sur-Yvette, France.

†Current address: Department of Genetics, University of Cambridge, Cambridge, CB2 3EH, United Kingdom.
*Corresponding authors: gabriela.bodea@gmail.com (G.O.B.); faulknergj@gmail.com (G.J.F)

Transposable elements (TEs) are a reservoir of new transcription factor binding sites for protein-coding genes[1–3]. Developmental programs that activate TE-derived regulatory elements could, in principle, manifest in lineage-specific TE mobility. While somatic LINE-1 (L1) retrotransposon insertions have been detected in human neurons[4–6], the impact of L1 insertions on neurodevelopmental gene regulation, and whether L1 mobility is restricted to certain neuronal lineages, is unknown. Here, we reveal programmed L1 activation by SOX6, a transcription factor critical for parvalbumin (PV^+) interneuron development[7–9]. PV^+ neurons harbor unmethylated and euchromatic L1 promoters, express L1 mRNA, and permit L1 transgene mobilization *in vivo*. Elevated L1 expression in adult dentate gyrus PV^+ neurons is however attenuated by environmental enrichment. Nanopore sequencing of PV^+ neurons identifies unmethylated L1 loci providing alternative promoters to core PV^+ neuron genes, such as CAPS2. These data depict SOX6-mediated L1 activation as an ingrained component of the mammalian PV^+ neuron developmental program.

The retrotransposon L1 comprises ~18% of the human and mouse genomes, and remains a source of gene and regulatory sequence variation[10–12]. To mobilize, L1 initiates transcription of a full-length (>6kbp) mRNA from its internal 5′UTR promoter. The mRNA encodes two proteins, denoted ORF1p and ORF2p, that mediate L1 retrotransposition[13–15]. While the promoters of the youngest human (L1HS) and mouse (T_F) L1 subfamilies differ in composition, they are each regulated by CpG methylation and contain binding sites for YY1 and SOX transcription factors[6,16–23] (**Fig. 1a**). SOX proteins bind two L1HS 5′UTR sites[20] previously linked to SOX2-mediated L1 transcriptional repression[18] (**Fig. 1b**), and SOX2 downregulation is thought to enable L1 retrotransposition in the neuronal lineage[4–6,17,18,24,25]. To functionally assess the L1HS 5′UTR SOX binding sites, we used a quantitative L1-EGFP retrotransposition reporter assay[13,26,27] in cultured PA-1 embryonal carcinoma cells. Inversion of either SOX site, or scrambling of the second site (+570 to +577), had no effect upon the mobility of L1.3, a highly mobile human L1[28,29] (**Fig. 1b**). Scrambling the first SOX site (+470 to +477) consistently reduced L1.3 retrotransposition efficiency by more than 40% (**Fig. 1b**). Of several SOX proteins known to be expressed in the brain[8,30], including SOX2, SOX5, SOX6 and SOX11, the first SOX site most closely matched SOX6 (**Fig. 1b**), and this motif coincided with the center of an ENCODE SOX6 ChIP-seq peak (**Fig. 1a**) obtained from K562 human myelogenous leukemia cells[31]. Based on these *in vitro* data, we hypothesized SOX6 is an L1 activator.

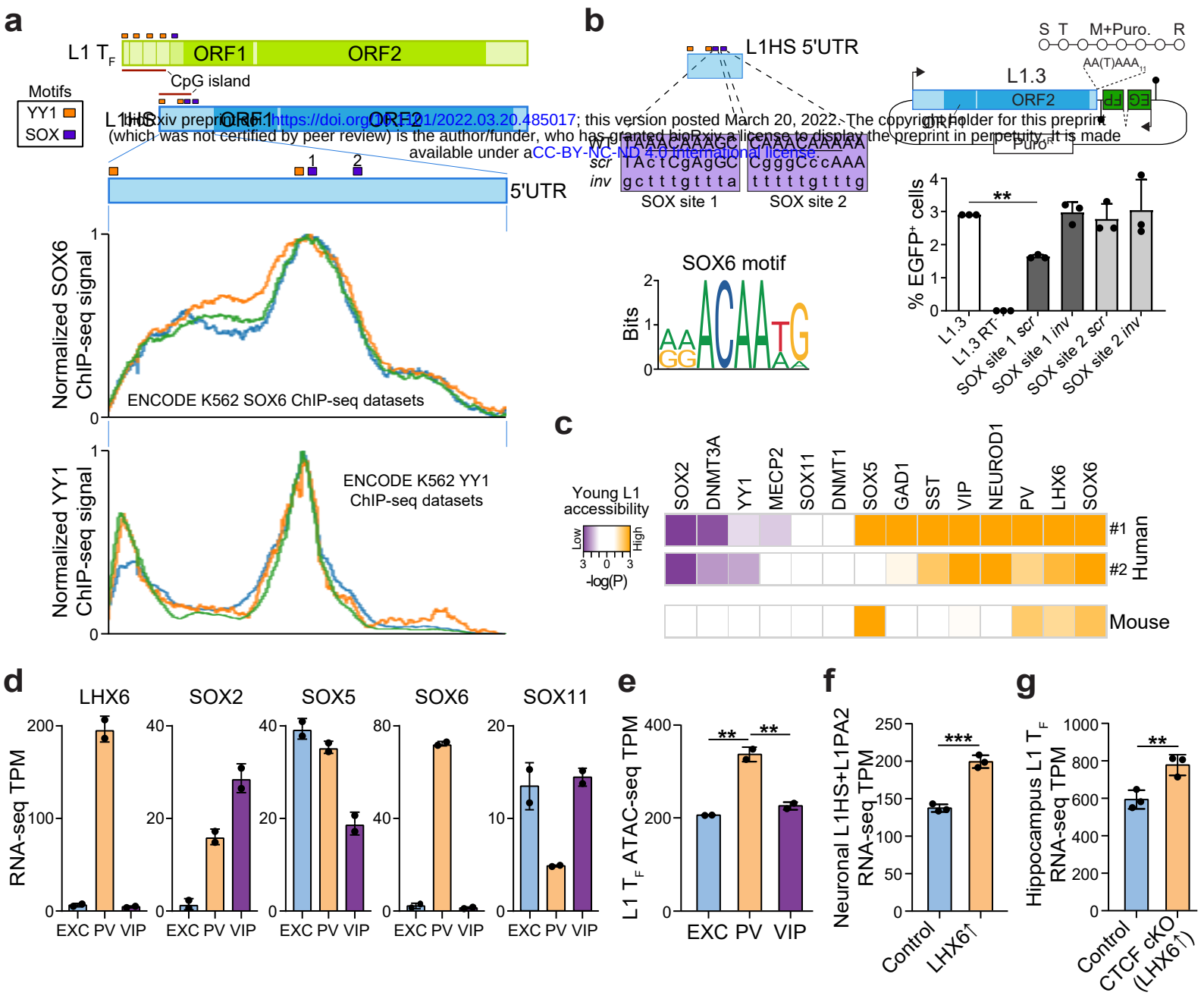


Fig. 1: L1 activation by the LHX6/SOX6 transcriptional program. a, Mouse (L1 T_F) and human (L1HS) mobile L1s. Each cartoon indicates two ORFs, as well as 5′UTR embedded YY1- (orange) and SOX-binding (purple) sites, with the latter numbered 1 and 2 and corresponding to L1HS positions +470 to +477 and +570 to +577, respectively. Mouse L1 T_F 5′UTR sequences are composed of multiple monomers, with 3.5 shown here, in addition to a non-monomeric sequence. Displayed underneath are MapR-RCon[31] profiles of ENCODE K562 SOX6 and YY1 ChIP-seq profiles of the L1HS 5′UTR. **b,** Left: annotated L1HS SOX-binding sites 1 and 2[20], highlighted in bold, were scrambled (*scr*) or inverted (*inv*). Site 1 most closely matched the JASPAR[91] SOX6 binding site motif (matrix ID: MA0515.1). Right: L1 retrotransposition efficiency measured in cultured PA-1 cells using an enhanced green fluorescent protein (EGFP) L1 reporter assay[13,26,27]. The assay design (top) shows L1.3, a highly mobile L1HS element[28,29], expressed from its native promoter (black arrow) and tagged with an EGFP cassette activated only upon retrotransposition and driven by a cytomegalovirus promoter (CMVp) (S, seeding; T, transfection; M, change of media; R, result analysis; filled lollipop, polyadenylation signal). AA(T)AAA indicates where a thymine base was removed to ablate the natural L1.3 polyadenylation signal. Cells are selected for puromycin resistance ($Puro^R$) and retrotransposition efficiency is measured as the percentage of $EGFP^+$ sorted cells. Tested elements (bottom) included, in order, positive (L1.3) and negative (L1.3 RT^-, D702A mutant) controls[13,28], followed by L1.3 vectors where the SOX-binding sites were scrambled or inverted. $^{**}P<0.01$, $n=3$ replicates. **c,** L1 chromatin accessibility in hippocampal tissue as measured by human[33] and mouse[32] scATAC-seq. Cells were grouped based on selected accessible genes known to regulate L1 activity or define neuronal populations. In each group, the average number of reads aligned to a young L1 (mouse: L1 T_F, human: L1HS or L1PA2) was calculated, with statistical significance compared to the remaining cells determined via permutation test ($n=1000$). Human data were available for two individuals (#1 and #2). **d,** LHX6, SOX2, SOX5, SOX6 and SOX11 expression in excitatory (EXC) pyramidal neuron, PV^+ interneuron and vasoactive intestinal peptide (VIP) interneuron cortex populations defined by Mo *et al.*[35], measured by RNA-seq tags per million (TPM). $^{***}P<0.001$, $N=2$. **e,** Proportion of ATAC-seq reads aligned to peaks associated with full-length L1 T_F copies in neuronal populations defined by Mo *et al.*[35]. $^{**}P<0.01$. **f,** Young human L1 subfamily expression measured by RNA-seq TPM in neurons derived via *in vitro* differentiation of induced pluripotent stem cells[36], with (LHX↑) and without (control) LHX overexpression. $^{***}P=0.0004$, $N=3$. **g,** L1 T_F subfamily expression measured by RNA-seq TPM in bulk hippocampus[37] of animals with (CTCF cKO) and without (control) conditional knockout of CTCF and associated induction of LHX6 expression. $^{**}P=0.01$, two-tailed t test, $N=3$. Note: histogram data in (b, d, e, f and g) are represented as mean ± SD. Significance testing for (b and e) was via one-way ANOVA with Tukey's multiple comparison test and for (f and g) via two-tailed t test.

SOX6 coordinates a major transcriptional program of the embryonic and adult brain, functioning downstream of LHX6 to ensure PV^+ neuron development[7–9]. To evaluate association of LHX6/SOX6 and L1 activity *in vivo*, we analyzed human and mouse hippocampus single-cell assay for transposase-accessible chromatin sequencing (scATAC-seq) datasets[32,33]. The known[6,17,18,24,34] L1 repressors SOX2, YY1 and DNMT3A were negatively correlated with human L1 accessibility (**Fig. 1c**). By contrast, LHX6, SOX6 and L1 accessibility were positively correlated in both species (**Fig. 1c**). We then examined mouse bulk excitatory and inhibitory neuron RNA-seq and ATAC-seq datasets[35] and found L1 accessibility, as well as LHX6 and SOX6 expression, were highest in PV^+ neurons, while SOX2, SOX5 and SOX11 expression were not specific to PV^+ neurons (**Fig. 1d,e**). Overexpression of LHX6 during human *in vitro* interneuron differentiation[36] (**Fig. 1f**) or via murine conditional *in vivo* knockout of the LHX6 inhibitor CTCF[37] (**Fig. 1g**) significantly increased L1 transcript abundance. These analyses suggested SOX6 could activate L1 in the PV^+ neuron lineage.

As an orthogonal *in vivo* approach to examine L1 activity in PV^+ neurons, we generated a transgenic mouse line carrying an L1-EGFP retrotransposition reporter system (**Fig. 2a** and **Extended Data Fig. 1a**). Here, L1.3 incorporated T7 and 3×FLAG epitope tags on L1 ORF1p and ORF2p, respectively (**Fig. 2a**). Immunofluorescence revealed $EGFP^+$ neurons *in vivo* (**Fig. 2b**). In agreement with prior transgenic L1 experiments[18], nearly all $EGFP^+$ cells were found in the brain, apart from occasional $EGFP^+$ ovarian interstitial cells (**Extended Data Fig. 1b-e**). Tagged ORF1p and ORF2p expression was observed in $EGFP^+$ neurons, indicating the L1 protein machinery coincided with retrotransposition (**Fig. 2c,d**). Crucially, 85.4% of $EGFP^+$ hippocampal cells were PV^+ neurons, on average (**Fig. 2e,f**). $PV^+/EGFP^+$ neurons were found throughout the hippocampal dentate gyrus (DG) and cornu ammonis regions 1-3 (CA1-3, referred to here as CA) (**Fig. 2g**) but were infrequent in the cortex (**Fig. 2h**). $EGFP^+$ cells also expressed GAD1 (**Extended Data Fig. 1f**), another inhibitory interneuron marker[38]. Seeking a complementary method, we used *in utero* electroporation to deliver to the embryonic hippocampus a codon-optimized synthetic mouse L1 T_F (L1SM)[39] bearing an EGFP reporter (**Extended Data Fig. 2a**). We observed occasional hippocampal $EGFP^+/PV^+$ neurons in electroporated neonates (**Extended Data Fig. 2b**). No $EGFP^+$ cells were present when the reporter, with disabled L1 ORF2p endonuclease and reverse transcriptase activities[39], was electroporated into the contralateral hemisphere (**Extended Data Fig. 2c**). As mouse and human

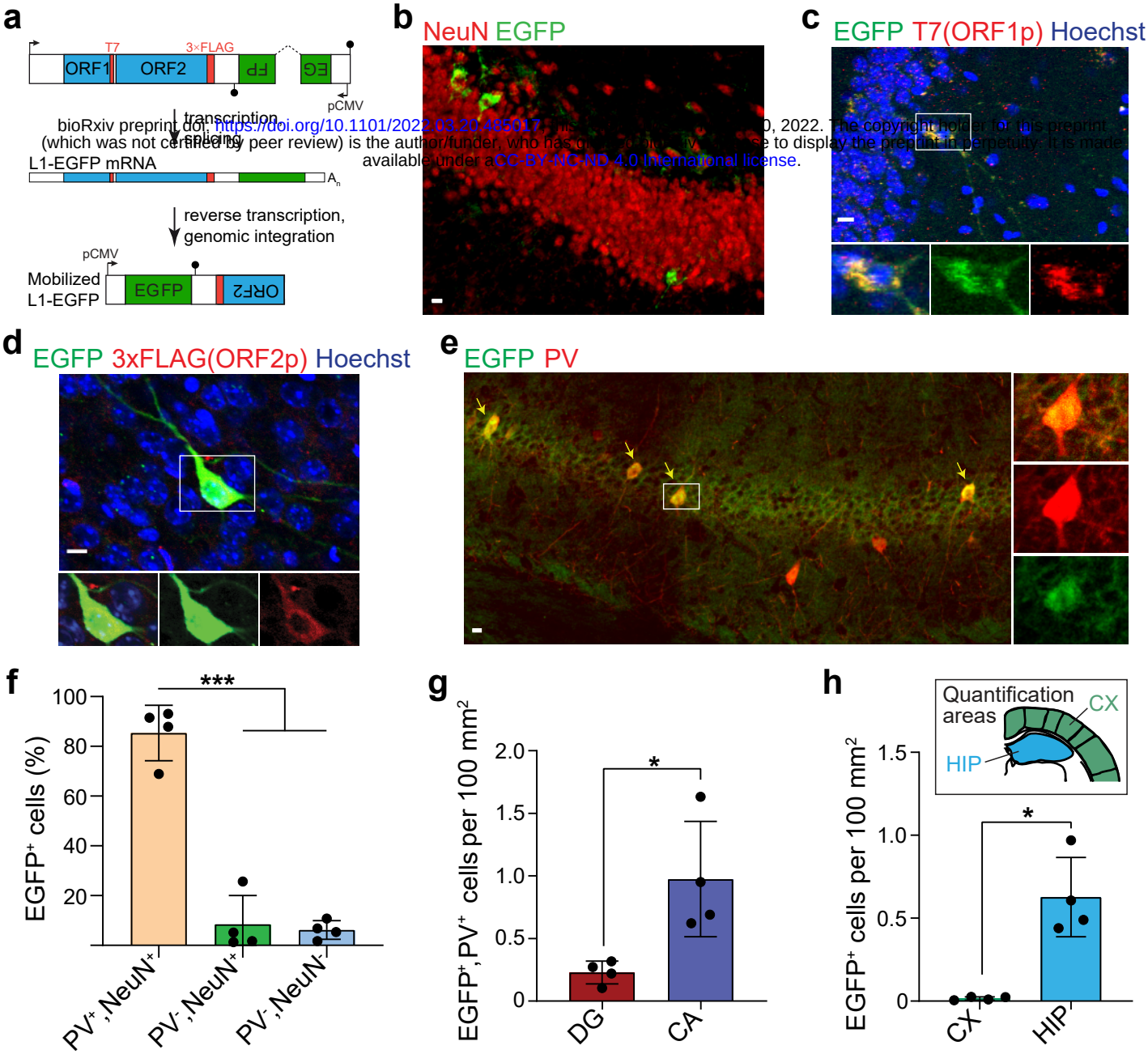
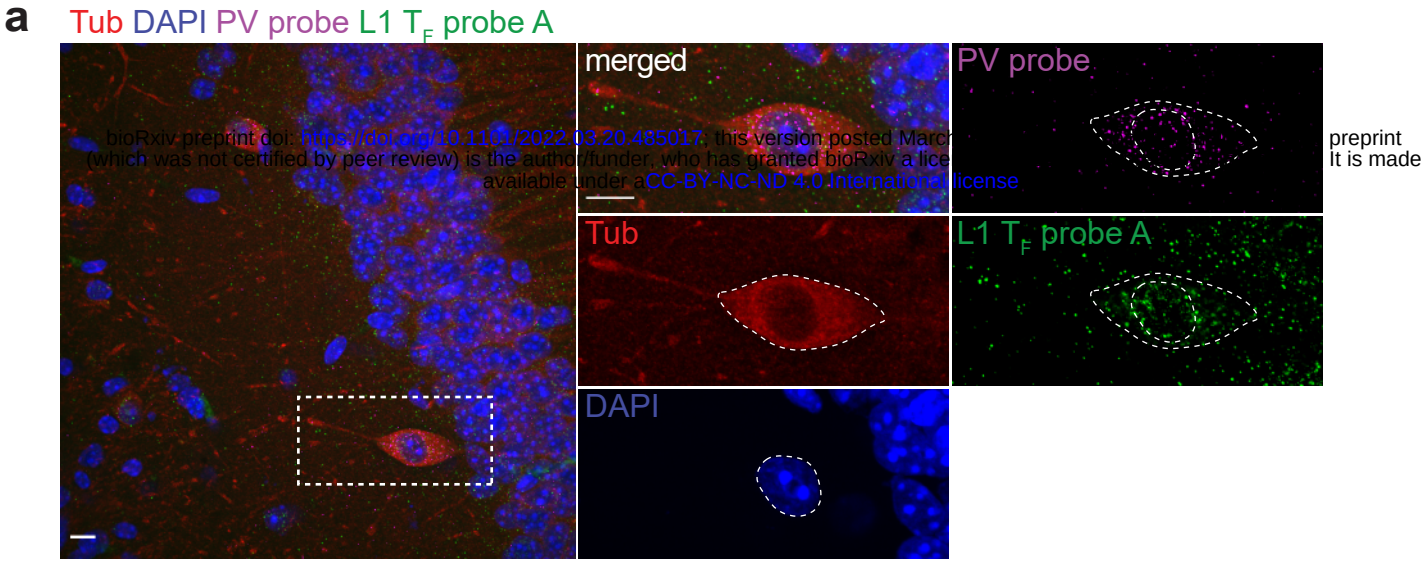


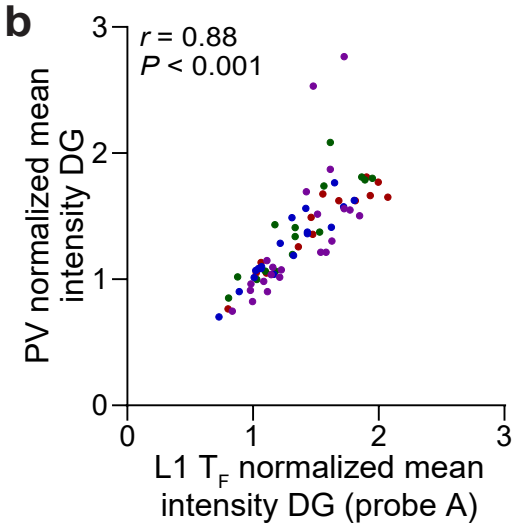
Fig. 2: Engineered L1 retrotransposition in PV⁺ neurons. a, L1-EGFP reporter schematic. A retrotransposition-competent human L1[26,28,29] is expressed from its native promoter, harbors epitope tagged ORF1 (T7) and ORF2 (3×FLAG) sequences, and carries an EGFP indicator cassette. The EGFP is antisense to the L1, incorporates a ɤ-globin intron in the same orientation as the L1, expressed from a cytomegalovirus promoter (CMVp), and terminated by a polyadenylation signal (filled black lollipop). L1-EGFP transcription and mRNA splicing removes the ɤ-globin EGFP intron and, if this mRNA is retrotransposed, allows EGFP expression. **b,** Example EGFP⁺ cells detected in the hippocampus. **c,** Representative confocal image of ORF1p immunostaining (T7 tag) of L1-EGFP adult mouse brain. Image insets show a selected cell in merged (top) and single channels for EGFP (green) and ORF1p (red). **d,** As for (c), except for 3×FLAG-tagged ORF2p. **e,** EGFP and PV immunostaining of an L1-EGFP animal coronal hippocampus section. Yellow arrows indicate EGFP⁺ neurons. **f,** Percentages of hippocampal EGFP⁺ cells colocalized with NeuN and PV. ***P=0.0002, Welch's ANOVA, N(mice)=4. **g,** Distribution of counted EGFP⁺/PV⁺ cells in hippocampal substructures. *P=0.0453, two-tailed t test. **h,** EGFP⁺ cell counts in cortex (CX) and hippocampus (HIP). *P=0.014, two-tailed t test. Note: panels (f-h) represent data as mean ± SD. Scale bars in (b-e) indicate 100µm.

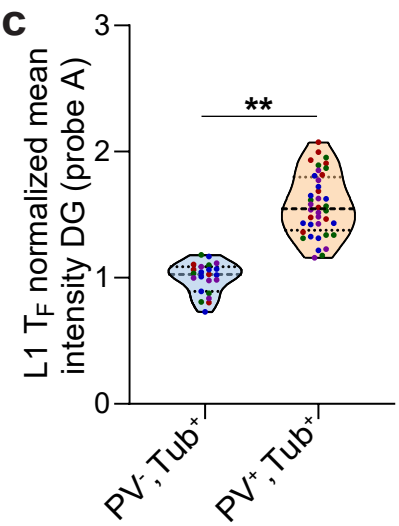
L1s delivered by distinct methods retrotransposed almost exclusively in PV^+ neurons *in vivo*, we inferred this lineage would likely satisfy the molecular prerequisites for endogenous L1 mobility.

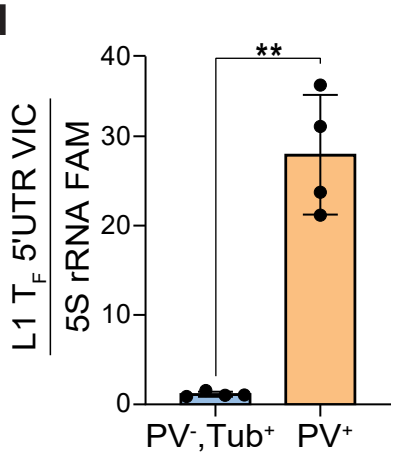
To quantify single L1 mRNA molecules in neuronal subpopulations, we designed a custom RNA fluorescence *in situ* hybridization (FISH) probe against the monomeric 5′UTR of the mouse L1 T_F subfamily[40–42] (**Extended Data Fig. 3**). With multiplexed RNA FISH, we measured cytoplasmic L1 and PV mRNA abundance in adult β-tubulin (Tub) immunostained neurons (**Fig. 3a** and **Extended Data Fig. 3**). L1 T_F and PV expression were strongly correlated in DG (Spearman r=0.88, P<0.001) (**Fig. 3b**) and CA (r=0.82, P<0.001) (**Extended Data Fig. 4a**) neurons. L1 T_F transcription was significantly higher in PV^+ neurons, compared to PV^- neurons, in DG (P=0.0016) (**Fig. 3c**) and CA (P=0.0251) (**Extended Data Fig. 4b**). A second L1 T_F 5′UTR RNA FISH probe (**Extended Data Fig. 3**) showed L1 mRNA enrichment in hippocampal and cortical PV^+ neurons (**Extended Data Fig. 4c-g**). We then used TaqMan qPCR to measure L1 expression in hippocampal PV^+ and PV^- neurons sorted from pooled neonate littermates (**Extended Data Fig. 5**). Three qPCR primer/probe combinations (**Extended Data Fig. 3**) detecting the L1 T_F 5′UTR each indicated significantly higher (P<0.03) expression in PV^+ neurons (**Fig. 3d** and **Extended Data Fig. 6a,b**). By contrast, qPCR targeting the L1 T_F ORF2 region, expected to mainly detect immobile 5′ truncated L1s incorporated in other cellular mRNAs, showed no difference between PV^+ and PV^- neurons (**Extended Data Fig. 6c,d**). We concluded PV^+ neurons were enriched for L1 T_F mRNA.

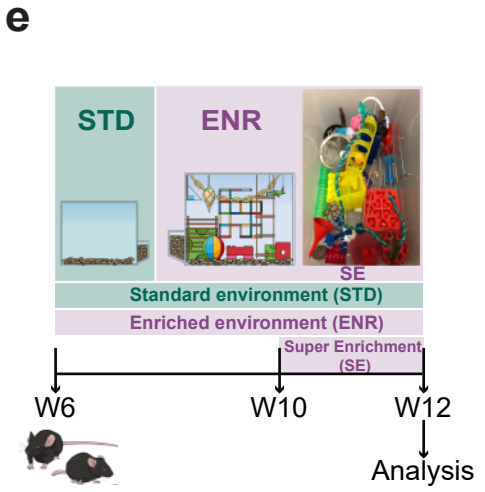
Given environmental stimuli can evoke molecular phenotypes in circuits involving PV^+ neurons[38], we analyzed L1 activity in adult mice housed in standard (STD) and enriched (ENR) environments. ENR cages were larger and incorporated ladders, tunneling objects, and toys of various textures, sizes and shapes (**Fig. 3e**). RNA FISH revealed a moderate (10.7%, P=0.049) reduction in DG PV^+ neuron L1 T_F transcript abundance in the ENR group compared to the STD group (**Fig. 3f**) and a smaller, non-significant reduction (7.5%, P=0.189) for CA PV^+ neurons (**Extended Data Fig. 7a**). No significant differences were seen for DG, CA or cortex PV^- neurons or cortical PV^+ neurons (**Fig. 3g** and **Extended Data Fig. 7b-e**). Access to voluntary running wheel exercise in lieu of an enriched environment did not significantly alter L1 T_F mRNA levels in DG PV^+ neurons (**Extended Data Fig. 7f**). Stereological analysis confirmed similar PV^+ neuron counts in ENR and STD animals (**Extended Data Fig. 7g,h**). Consistent, though non-significant, decreases in L1 T_F 5′UTR mRNA abundance were observed by qPCR in

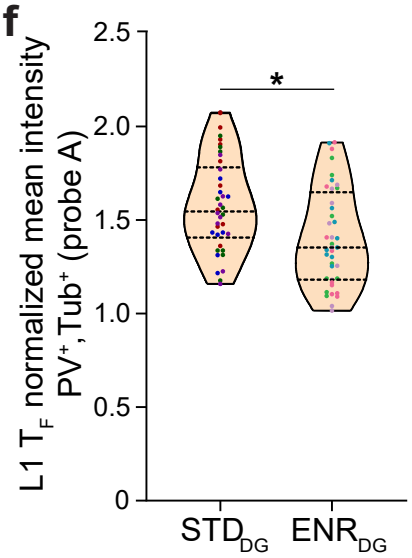












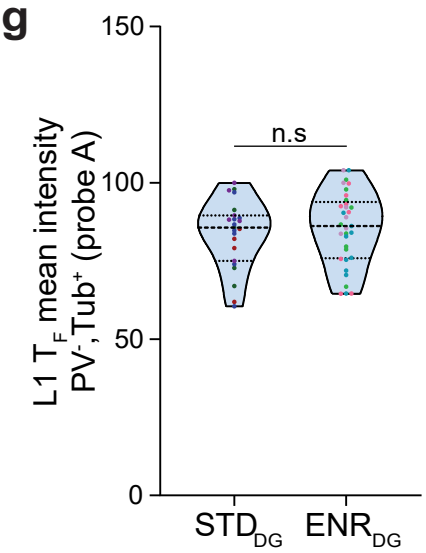


Fig. 3: L1 mRNA is abundant in PV$^+$ neurons. a, Representative maximum intensity projection confocal image of a coronal hippocampus section showing L1 T_F (green) and PV (magenta) transcripts detected by RNA FISH, β-tubulin (Tub, red) immunohistochemistry and DAPI staining (blue). Image insets show higher magnification of a selected PV$^+$ neuron (dashed rectangle). Dashed lines in image insets demark nuclear and cellular boundaries defined for PV and L1 mRNA quantification. Scale bar: 10μm. **b,** Mean PV RNA FISH intensity in dentate gyrus (DG) Tub$^+$ neurons, as a function of L1 T_F (probe A) signal. Spearman r=0.88, P<0.001 at α=0.05, n(cells)=69, N(mice)=4. Cells from different mice are color coded. **c,** Mean L1 T_F RNA FISH intensity in DG PV$^-$/Tub$^+$ and PV$^+$/Tub$^+$ neurons. **P=0.0016, PV$^-$/Tub$^+$ n=23, PV$^+$/Tub$^+$ n=40, N=4. **d,** Multiplexed TaqMan qPCR[34] measuring mRNA abundance of the L1 T_F monomeric 5′UTR (VIC channel) relative to 5S rRNA (FAM channel) in PV$^-$/Tub$^+$ and PV$^+$ neurons. Cells were sorted from pooled neonate (P0) litter hippocampi. **P=0.004, N=4 litters. Data are represented as mean ± SD. **e,** Standard (STD) and enriched (ENR) environment housing schematics. Mice (aged 6 weeks) were placed in either STD or ENR housing for 6 weeks. ENR consisted of: large cage for spatial stimuli; ladders, tunneling objects and toys of various textures, sizes, and shapes for sensory, cognitive and motor stimulation. Between week 10 and 12, ENR mice were exposed three times a week for one hour to 'super-enriched' condition in a larger playground arena with novel toys. **f,** Mean L1 T_F RNA FISH intensity in PV$^+$/Tub$^+$ neurons from STD and ENR animal DG tissue. *P=0.049, STD n=38, ENR n=40, N=4. **g,** As for (f), except comparing DG PV$^-$/Tub$^+$ neurons. P=0.444, STD n=23, ENR n=33. Note: PV$^+$ values in (c, d, f) were normalized to PV$^-$/Tub$^+$ mean values. Significance testing was via two-tailed t test, comparing animal or litter mean values.

ENR bulk hippocampus, compared to STD samples, while ORF2 was not lower (**Extended Data Fig. 7i-k**). Considered alongside these results, our RNA FISH data suggested adult DG PV^+ neuron L1 transcription was specifically attenuated *in vivo* by environmental enrichment.

DNA methylation mediates L1 promoter silencing[6,22,23]. Therefore, to explain the apparent specificity of L1 transcription to PV^+ neurons, we performed L1 T_F 5′UTR monomer bisulfite sequencing[43] on neonate hippocampal cell populations. L1 T_F was significantly (P=0.03) less methylated on average in PV^+ neurons (83.9%) than in PV^- neurons (91.8%) (**Fig. 4a,b**). Unmethylated L1 T_F monomers were only observed in PV^+ neurons (**Fig. 4a,c**). DNMT1, DNMT3A and MeCP2 effect methylation-associated transcriptional repression in PV^+ neurons[44–47]. These genes all expressed significantly (P<0.05) less mRNA in neonate PV^+ neurons than in PV^- neurons (**Fig. 4d,e** and **Extended Data Fig. 8a**). MeCP2 protein expression was on average 10.5% lower in adult PV^+ neurons, compared to PV^- neurons (P=0.0007) (**Extended Data Fig. 8b,c**). L1 repression thus appeared broadly relaxed in PV^+ neurons.

Long-read Oxford Nanopore Technologies (ONT) sequencing allows genome-wide analysis of TE family methylation, as well as that of individual TE loci[48]. We therefore ONT sequenced PV^+ and PV^- cells from pooled neonate hippocampus samples to ~25× and ~15× combined genome-wide depths, respectively. Amongst the potentially mobile TE families surveyed, only the youngest mouse L1s (T_F, G_F and A-type) were significantly (P<0.05) less methylated in PV^+ cells than in PV^- cells (**Fig. 4f**). L1 loci supplied the vast majority (82%) of differentially methylated TEs (**Fig. 5a**). Of 545 differentially methylated (P<0.01) full-length L1 T_F loci, 543 were less methylated in PV^+ cells (**Supplementary Table 1**). Notably, the T_F subfamily can be further divided into three additional groups distinguishable by ONT sequencing, denoted T_{FI}, T_{FII}, and T_{FIII}, where T_{FIII} is the oldest and diverges in its 5′UTR when compared to $T_{FI/II}$[49]. We found by far the highest fraction (72%) of strongly demethylated L1s corresponded to the T_{FIII} subfamily (**Fig. 5b,c**). We noted the non-monomeric T_{FIII} 5′UTR contained two predicted SOX protein binding sites, whereas other young mouse L1 5′UTRs contained only one of these motifs (**Fig. 5d**). Remarkably, significantly hypomethylated L1 T_{FIII} copies with intact ORFs were observed in the introns of protein-coding genes critical to PV^+ neuron development and function, such as CAPS2[50,51], CHL1[52], and ERBB4[53] (**Fig. 5e,f** and **Extended Data Fig. 9** and **Supplementary Table 1**). In CAPS2, for example, the L1 5′UTR was completely unmethylated in numerous PV^+ neurons (**Fig. 5f**). Analysis of ENCODE PacBio

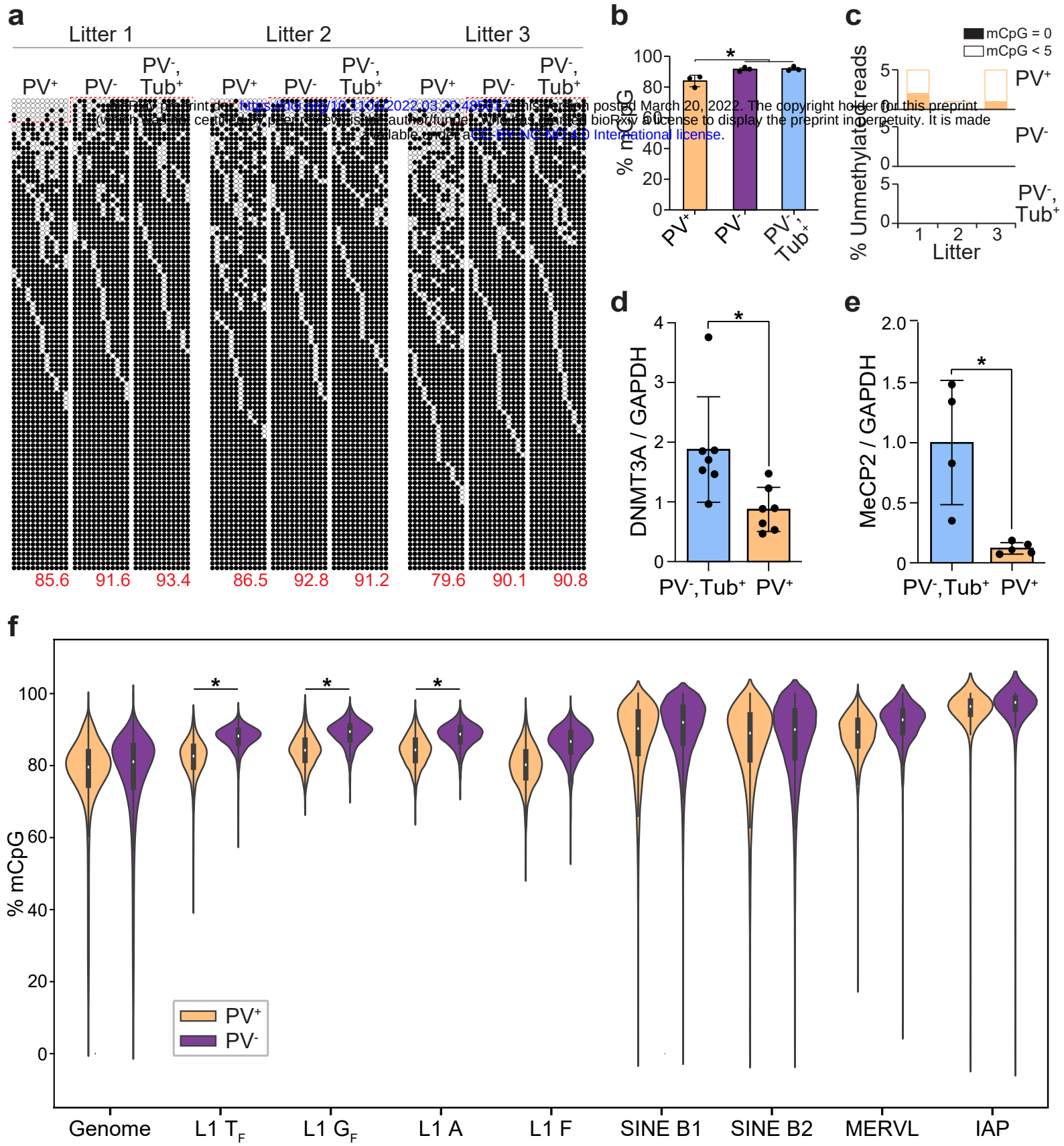


Fig. 4: Global L1 T_F promoter hypomethylation in PV$^+$ neurons. a, Targeted bisulfite sequencing of L1 T_F promoter monomer CpG islands[43] was performed on PV$^+$, PV$^-$ and PV$^-$/Tub$^+$ cells sorted from pooled hippocampal tissue from each of three neonate (P0) litters. Each cartoon displays 100 non-identical randomly selected sequences, where methylated CpGs (mCpGs) and unmethylated CpGs are represented by black and white circles, respectively, as well as the overall mCpG percentage (red numbers). Amplicons above the dotted red line contain <5 mCpGs. **b,** L1 T_F monomer methylation was significantly lower (*P<0.05, two-tailed t test) in PV$^+$ neurons than in either PV$^-$ or PV$^-$/Tub$^+$ cells. **c,** Fully (mCpG=0) and nearly (mCpG<5) unmethylated L1 T_F monomers were only found in PV$^+$ neurons. **d,** DNMT3A mRNA abundance measured by qPCR in PV$^-$/Tub$^+$ and PV$^+$ neurons, relative to GAPDH. *P=0.024, two-tailed t test, N=7 litters. **e,** As for (d), except for MeCP2. *P=0.041, N=4. **f,** CpG methylation ascertained by barcoded ONT sequencing upon matched hippocampal PV$^+$ and PV$^-$ cells from ten separate neonate litter pools. Results are shown for the whole genome (10kbp windows), T_F, G_F, A-type and F-type L1s >6kbp, B1 (>140bp) and B2 (>185bp) SINEs, and MERVL MT2 (>470bp) and IAP (>320bp) long terminal repeats. Included elements accrued at least 20 methylation calls and 4 reads in each of the combined PV$^+$ and PV$^-$ datasets. *P<0.05, Friedman test followed by Dunn's multiple comparison test, N=10 litter means. Note: histogram data are represented as mean ± SD.

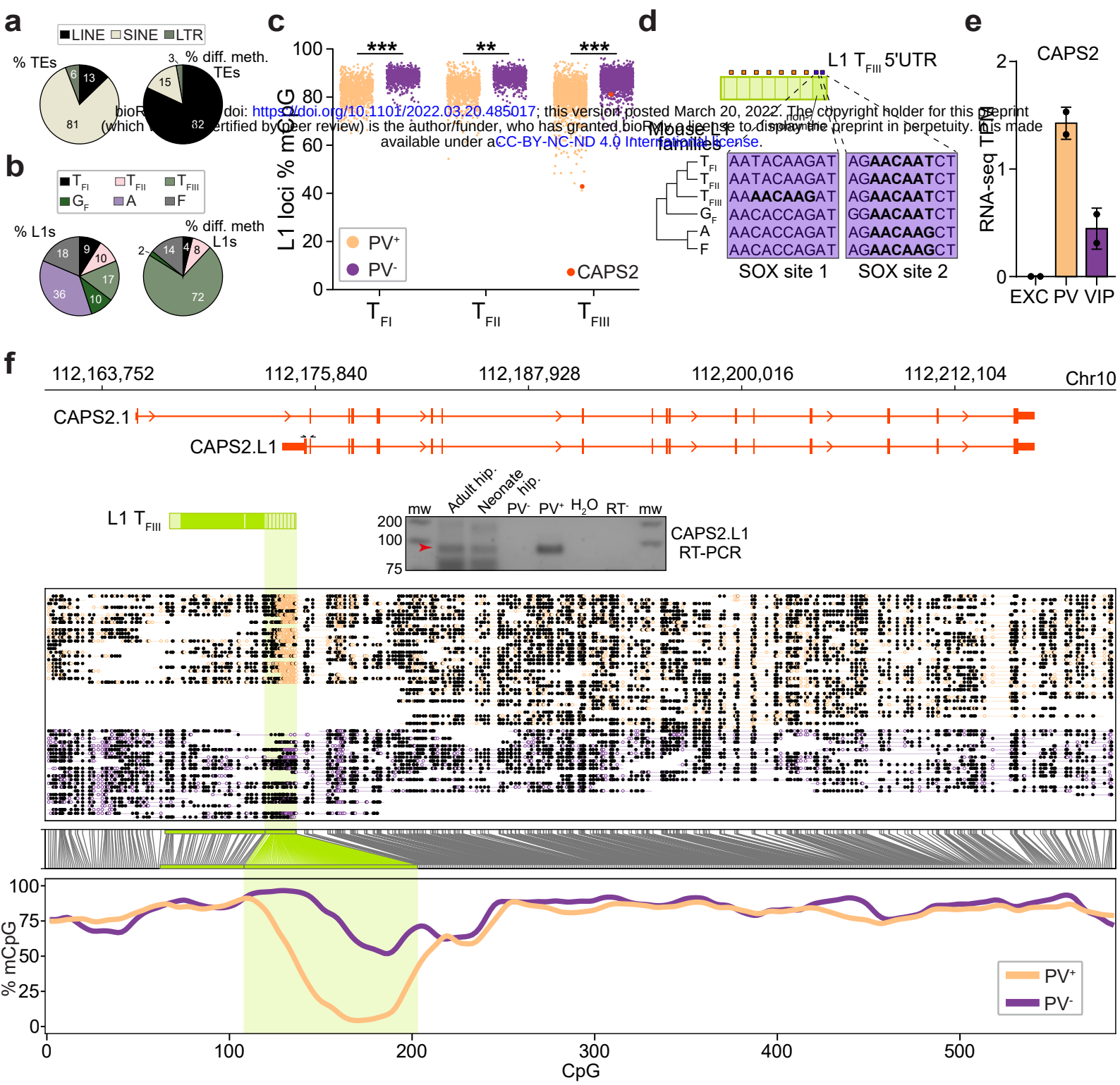


Fig. 5: PV⁺ neuron genes harbor hypomethylated L1 T_F promoters. a, Composition of all young (left) and differentially methylated (P<0.01, Supplementary Table 1) young (right) TEs, by superfamily. Note: MERVL and IAP are LTR retrotransposons. **b,** As per (a), except showing the breakdown of young L1 subfamilies (left) and their contribution to the 50 differentially methylated (P<0.01) loci showing the largest absolute change in methylation percentage. **c,** L1 T_{FI}, T_{FII} and T_{FIII} subfamily CpG methylation strip plots for PV⁺ and PV⁻ cells, as represented collectively by the L1 T_F violin plot in Fig. 4f. Each point represents an L1 locus, with an example intronic to CAPS2 highlighted by an orange dot. **P<0.01 and ***P<0.001, Friedman test followed by Dunn's multiple comparison test, N=10 litter means. **d,** The mouse L1 T_{FIII} subfamily 5′UTR is composed of multiple monomers, with 7.5 shown here, in addition to a non-monomeric sequence. YY1- (orange) and SOX-binding (purple) sites are shown. The L1 T_{FIII} consensus non-monomeric region contains two predicted SOX-binding sites, highlighted in bold, whereas the other young L1s have only one. The depicted cladogram is based on a multiple sequence alignment of non-monomeric sequences. **e,** CAPS2 expression in excitatory (EXC) pyramidal neuron, PV⁺ interneuron and vasoactive intestinal peptide (VIP) interneuron cortex populations defined by Mo *et al.*[35], measured by RNA-seq tags per million (TPM). N=2. **f,** Methylation profile of the CAPS2 locus obtained from ONT sequencing. The first panel shows a full-length L1 T_{FIII} with intact ORFs, as highlighted in (c), orientated antisense to the first intron of the canonical CAPS2.1 isoform. ENCODE long-read transcriptome sequencing of hippocampus tissue (ENCLB505CBY) indicated a chimeric transcript, labeled here CAPS2.L1, spliced into CAPS2.1 exon 2 and encoding an ORF in frame with the CAPS2.1 ORF. The gel image displays PCR products generated using primers specific to CAPS2.L1 (marked by opposing black arrows), with input template cDNA from bulk adult hippocampus (5′RACE) and neonate hippocampus (reverse transcribed total RNA from bulk and sorted PV⁺ and PV⁻ cells). A red arrow indicates on-target products confirmed by capillary sequencing. The second panel displays aligned ONT reads, with unmethylated CpGs colored in orange (PV⁺) and purple (PV⁻), and methylated CpGs colored black. The third panel indicates the relationship between CpG positions in genome space and CpG space, including those corresponding to the L1 T_{FIII} 5′UTR (shaded light green). The fourth panel indicates the fraction of methylated CpGs for each cell type across CpG space.

long-read hippocampus transcriptome sequencing[54] indicated the L1 5′UTR initiated an antisense novel transcript, which we termed CAPS2.L1, spliced into downstream CAPS2 exons and in frame with the canonical CAPS2 ORF (**Fig. 5f**). By 5′RACE and RT-PCR, we reliably detected CAPS2.L1 in adult and neonate hippocampus tissue, and in PV^+ cells, but not PV^- cells (**Fig. 5f**). Aside from the CAPS2 example, we identified 43 young mouse L1s whose 5′UTR promoted expression of a spliced transcript annotated by GenBank or detected by the abovementioned ENCODE PacBio datasets (**Supplementary Table 1**). These results suggested unmethylated L1s can provide alternative promoters otherwise repressed during neurogenesis[12,55,56] to key genes required for proper mouse PV^+ neuron development and function.

In sum, this study reveals L1 activity in the PV^+ neuron lineage governed by SOX6 (**Extended Data Fig. 10**). PV^+ neurons are "node" cells that connect neural circuits associated with memory consolidation and other core cognitive processes[57,58]. The potential for neurodevelopmental L1 mobility as a consequence of PV^+ neuron genes incorporating unmethylated retrotransposition-competent L1s is notable given the proposed roles for stochastic L1-mediated genome mosaicism in the brain[18,59–61]. Our results do not however preclude other neuronal subtypes or brain regions from expressing L1 mRNA. Engineered L1 reporter experiments have thus far generated data congruent with endogenous L1 mobility in the early embryo[42,62], neurons[4–6,18,24,63] and in cancer[13,16,64]. While we and others have mapped endogenous L1 retrotransposition events in individual human neurons[4–6], the composition of the L1 T_F 3′UTR appears to severely impede such analyses in mouse[42,43,65]. Prior pan-neuronal studies reported elevated L1-EGFP mobility associated with exercise-induced adult neurogenesis[66] and, as a result of early-life stress, increased L1 DNA copy number[34]. That elevated PV^+ neuron L1 activity in adult DG was here moderately attenuated, rather than increased, by ENR housing, and not affected by voluntary exercise, was therefore unexpected. Environmental enrichment may counter physiological stress upon PV^+ neurons[67–69], and could enhance L1 repression[70]. As shown in other biological contexts[1,3,10,21], L1 activation in PV^+ neurons illustrates how retrotransposons may be incorporated into transcription factor programs guiding cell fate and mature function, potentially in an experience-dependent manner.

Acknowledgements

We thank J.D. Boeke, J.L. Garcia-Perez and J.V. Moran for sharing L1SM and L1.3 plasmids,

and P. Sah, R. Lister, R. Sullivan, S. van de Wakker, A. Gaudin, S.W. Cheetham, N. Jansz and members of the Faulkner laboratory for helpful discussions. We acknowledge the QBI and TRI Flow Cytometry suites for technical advice and the QBI Advanced Microscopy Facility for technical assistance and equipment, supported by ARC LIEF grant LE130100078. This work was supported by NHMRC-ARC Dementia Research Development Fellowship GNT1108258 and DFG fellowship BO4460/1-1 (G.O.B.), NHMRC Investigator Grants GNT1173476 (S.R.R.) and GNT1173711 (G.J.F.), ARC Discovery Project DP200102919 (S.R.R. and G.J.F.), NHMRC Project Grants GNT1106206 (G.J.F., A.J.H., L.M.P.) and GNT1126393 (G.J.F.), a CSL Centenary Fellowship (G.J.F), and the Mater Foundation.

Author contributions

G.O.B. and G.J.F. designed the project and wrote the manuscript. G.O.B., M.E.F., F.J.S-L., J.M.B., J.R., M.A.R., L.R.F., C.G., P.G., L-G.B, P.A., V.B., S.M., M-J.H.C.K. and C.J.L. performed experiments. G.O.B., M.E.F., P.K., A.D.E., D.J.J., S.R.R., A.J.H. and G.J.F. analyzed the data. G.O.B., L.M.P., S.R.R., A.J.H. and G.J.F. provided resources.

Competing interests

The authors declare no competing interests.

References

1. Chuong, E. B., Elde, N. C. & Feschotte, C. Regulatory activities of transposable elements: from conflicts to benefits. *Nat. Rev. Genet.* **18**, 71–86 (2017).
2. Britten, R. J. & Davidson, E. H. Gene regulation for higher cells: a theory. *Science* **165**, 349–357 (1969).
3. Senft, A. D. & Macfarlan, T. S. Transposable elements shape the evolution of mammalian development. *Nat. Rev. Genet.* **22**, 691–711 (2021).
4. Erwin, J. A. *et al.* L1-associated genomic regions are deleted in somatic cells of the healthy human brain. *Nat. Neurosci.* **19**, 1583–1591 (2016).
5. Evrony, G. D. *et al.* Cell lineage analysis in human brain using endogenous retroelements. *Neuron* **85**, 49–59 (2015).
6. Sanchez-Luque, F. J. *et al.* LINE-1 Evasion of Epigenetic Repression in Humans. *Mol. Cell*

75, 590–604 (2019).

7. Batista-Brito, R. *et al.* The cell-intrinsic requirement of Sox6 for cortical interneuron development. *Neuron* **63**, 466–481 (2009).
8. Connor, F., Wright, E., Denny, P., Koopman, P. & Ashworth, A. The Sry-related HMG box-containing gene Sox6 is expressed in the adult testis and developing nervous system of the mouse. *Nucleic Acids Res.* **23**, 3365–3372 (1995).
9. Munguba, H. *et al.* Postnatal Sox6 regulates synaptic function of cortical parvalbumin-expressing neurons. *J. Neurosci.* (2021).
10. Kazazian, H. H., Jr & Moran, J. V. Mobile DNA in Health and Disease. *N. Engl. J. Med.* **377**, 361–370 (2017).
11. Moran, J. V., DeBerardinis, R. J. & Kazazian, H. H., Jr. Exon shuffling by L1 retrotransposition. *Science* **283**, 1530–1534 (1999).
12. Denli, A. M. *et al.* Primate-specific ORF0 contributes to retrotransposon-mediated diversity. *Cell* **163**, 583–593 (2015).
13. Moran, J. V. *et al.* High frequency retrotransposition in cultured mammalian cells. *Cell* **87**, 917–927 (1996).
14. Wei, W. *et al.* Human L1 retrotransposition: cispreference versus trans complementation. *Mol. Cell. Biol.* **21**, 1429–1439 (2001).
15. Feng, Q., Moran, J. V., Kazazian, H. H., Jr & Boeke, J. D. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* **87**, 905–916 (1996).
16. Scott, E. C. *et al.* A hot L1 retrotransposon evades somatic repression and initiates human colorectal cancer. *Genome Res.* **26**, 745–755 (2016).
17. Kuwabara, T. *et al.* Wnt-mediated activation of NeuroD1 and retro-elements during adult neurogenesis. *Nat. Neurosci.* **12**, 1097–1105 (2009).
18. Muotri, A. R. *et al.* Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* **435**, 903–910 (2005).
19. Athanikar, J. N., Badge, R. M. & Moran, J. V. A YY1-binding site is required for accurate human LINE-1 transcription initiation. *Nucleic Acids Res.* **32**, 3846–3855 (2004).
20. Tchénio, T., Casella, J. F. & Heidmann, T. Members of the SRY family regulate the human LINE retrotransposons. *Nucleic Acids Res.* **28**, 411–415 (2000).
21. Deniz, Ö., Frost, J. M. & Branco, M. R. Regulation of transposable elements by DNA

modifications. *Nat. Rev. Genet.* **20**, 417–431 (2019).

22. de la Rica, L. *et al.* TET-dependent regulation of retrotransposable elements in mouse embryonic stem cells. *Genome Biol.* **17**, 234 (2016).
23. Castro-Diaz, N. *et al.* Evolutionally dynamic L1 regulation in embryonic stem cells. *Genes Dev.* **28**, 1397–1409 (2014).
24. Coufal, N. G. *et al.* L1 retrotransposition in human neural progenitor cells. *Nature* **460**, 1127–1131 (2009).
25. Macia, A. *et al.* Engineered LINE-1 retrotransposition in nondividing human neurons. *Genome Res.* **27**, 335–348 (2017).
26. Ostertag, E. M., Prak, E. T., DeBerardinis, R. J., Moran, J. V. & Kazazian, H. H., Jr. Determination of L1 retrotransposition kinetics in cultured cells. *Nucleic Acids Res.* **28**, 1418–1423 (2000).
27. Kopera, H. C. *et al.* LINE-1 cultured cell retrotransposition assay. *Methods Mol. Biol.* **1400**, 139–156 (2016).
28. Sassaman, D. M. *et al.* Many human L1 elements are capable of retrotransposition. *Nat. Genet.* **16**, 37–43 (1997).
29. Dombroski, B. A., Scott, A. F. & Kazazian, H. H., Jr. Two additional potential retrotransposons isolated from a human L1 subfamily that contains an active retrotransposable element. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 6513–6517 (1993).
30. Bowles, J., Schepers, G. & Koopman, P. Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators. *Dev. Biol.* **227**, 239–255 (2000).
31. Sun, X. *et al.* Transcription factor profiling reveals molecular choreography and key regulators of human retrotransposon expression. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E5526–E5535 (2018).
32. Sinnamon, J. R. *et al.* The accessible chromatin landscape of the murine hippocampus at single-cell resolution. *Genome Res.* **29**, 857–869 (2019).
33. Corces, M. R. *et al.* Single-cell epigenomic analyses implicate candidate causal variants at inherited risk loci for Alzheimer's and Parkinson's diseases. *Nat. Genet.* **52**, 1158–1168 (2020).
34. Bedrosian, T. A., Quayle, C., Novaresi, N. & Gage, F. H. Early life experience drives

structural variation of neural genomes in mice. *Science* **359**, 1395–1399 (2018).

35. Mo, A. *et al.* Epigenomic Signatures of Neuronal Diversity in the Mammalian Brain. *Neuron* **86**, 1369–1384 (2015).
36. Yuan, F. *et al.* Induction of human somatostatin and parvalbumin neurons by expressing a single transcription factor LIM homeobox 6. *Elife* **7**, e37382 (2018).
37. Sams, D. S. *et al.* Neuronal CTCF Is Necessary for Basal and Experience-Dependent Gene Regulation, Memory Formation, and Genomic Structure of BDNF and Arc. *Cell Rep.* **17**, 2418–2430 (2016).
38. Donato, F., Rompani, S. B. & Caroni, P. Parvalbumin-expressing basket-cell network plasticity induced by experience regulates adult learning. *Nature* **504**, 272–276 (2013).
39. Han, J. S. & Boeke, J. D. A highly active synthetic mammalian retrotransposon. *Nature* **429**, 314–318 (2004).
40. Naas, T. P. *et al.* An actively retrotransposing, novel subfamily of mouse L1 elements. *EMBO J.* **17**, 590–597 (1998).
41. DeBerardinis, R. J., Goodier, J. L., Ostertag, E. M. & Kazazian, H. H. Rapid amplification of a retrotransposon subfamily is evolving the mouse genome. *Nat. Genet.* **20**, 288–290 (1998).
42. Richardson, S. R. *et al.* Heritable L1 retrotransposition in the mouse primordial germline and early embryo. *Genome Res.* **27**, 1395–1405 (2017).
43. Schauer, S. N. *et al.* L1 retrotransposition is a common feature of mammalian hepatocarcinogenesis. *Genome Res.* **28**, 639–653 (2018).
44. Lavery, L. A. *et al.* Losing Dnmt3a dependent methylation in inhibitory neurons impairs neural function by a mechanism impacting Rett syndrome. *Elife* **9**, e52981 (2020).
45. Wu, H. *et al.* Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. *Science* **329**, 444–448 (2010).
46. Feng, J. *et al.* Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons. *Nat. Neurosci.* **13**, 423–430 (2010).
47. Muotri, A. R. *et al.* L1 retrotransposition in neurons is modulated by MeCP2. *Nature* **468**, 443–446 (2010).
48. Ewing, A. D. *et al.* Nanopore Sequencing Enables Comprehensive Transposable Element Epigenomic Profiling. *Mol. Cell* **80**, 915–928.e5 (2020).

49. Sookdeo, A., Hepp, C. M., McClure, M. A. & Boissinot, S. Revisiting the evolution of mouse LINE-1 in the genomic era. *Mob. DNA* **4**, 3 (2013).
50. Shinoda, Y. *et al.* Calcium-dependent activator protein for secretion 2 (CAPS2) promotes BDNF secretion and is critical for the development of GABAergic interneuron network. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 373–378 (2011).
51. Ratai, O. *et al.* An Alternative Exon of CAPS2 Influences Catecholamine Loading into LDCVs of Chromaffin Cells. *J. Neurosci.* **39**, 18–27 (2019).
52. Schmalbach, B. *et al.* Age-dependent loss of parvalbumin-expressing hippocampal interneurons in mice deficient in CHL1, a mental retardation and schizophrenia susceptibility gene. *J. Neurochem.* **135**, 830–844 (2015).
53. Chen, Y.-J. *et al.* ErbB4 in parvalbumin-positive interneurons is critical for neuregulin 1 regulation of long-term potentiation. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 21818–21823 (2010).
54. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
55. Faulkner, G. J. *et al.* The regulated retrotransposon transcriptome of mammalian cells. *Nat. Genet.* **41**, 563–571 (2009).
56. Jönsson, M. E. *et al.* Activation of neuronal genes via LINE-1 elements upon global DNA demethylation in human neural progenitors. *Nat. Commun.* **10**, 3182 (2019).
57. Fuchs, E. C. *et al.* Recruitment of parvalbumin-positive interneurons determines hippocampal function and associated behavior. *Neuron* **53**, 591–604 (2007).
58. Ognjanovski, N. *et al.* Parvalbumin-expressing interneurons coordinate hippocampal network dynamics required for memory consolidation. *Nat. Commun.* **8**, 15039 (2017).
59. Erwin, J. A., Marchetto, M. C. & Gage, F. H. Mobile DNA elements in the generation of diversity and complexity in the brain. *Nat. Rev. Neurosci.* **15**, 497–506 (2014).
60. McConnell, M. J. *et al.* Intersection of diverse neuronal genomes and neuropsychiatric disease: The Brain Somatic Mosaicism Network. *Science* **356**, eaal1641 (2017).
61. Muotri, A. R. & Gage, F. H. Generation of neuronal variability and complexity. *Nature* **441**, 1087–1093 (2006).
62. Kano, H. *et al.* L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. *Genes Dev.* **23**, 1303–1312 (2009).

63. Baillie, J. K. *et al.* Somatic retrotransposition alters the genetic landscape of the human brain. *Nature* **479**, 534–537 (2011).
64. Miki, Y. *et al.* Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer Res.* **52**, 643–645 (1992).
65. Howell, R. & Usdin, K. The ability to form intrastrand tetraplexes is an evolutionarily conserved feature of the 3'end of L1 retrotransposons. *Mol. Biol. Evol.* **14**, 144–155 (1997).
66. Muotri, A. R., Zhao, C., Marchetto, M. C. N. & Gage, F. H. Environmental influence on L1 retrotransposons in the adult hippocampus. *Hippocampus* **19**, 1002–1007 (2009).
67. Francis, D. D., Diorio, J., Plotsky, P. M. & Meaney, M. J. Environmental enrichment reverses the effects of maternal separation on stress reactivity. *J. Neurosci.* **22**, 7840–7843 (2002).
68. van Dellen, A., Blakemore, C., Deacon, R., York, D. & Hannan, A. J. Delaying the onset of Huntington's in mice. *Nature* **404**, 721–722 (2000).
69. Chen, C.-C., Lu, J., Yang, R., Ding, J. B. & Zuo, Y. Selective activation of parvalbumin interneurons prevents stress-induced synapse loss and perceptual defects. *Mol. Psychiatry* **23**, 1614–1625 (2018).
70. Zocher, S., Overall, R. W., Lesche, M., Dahl, A. & Kempermann, G. Epigenetic rejuvenation of the hippocampus by environmental enrichment. *bioRxiv* (2019) doi:10.1101/776310.
71. Garcia-Perez, J. L. *et al.* LINE-1 retrotransposition in human embryonic stem cells. *Hum. Mol. Genet.* **16**, 1569–1577 (2007).
72. Alisch, R. S., Garcia-Perez, J. L., Muotri, A. R., Gage, F. H. & Moran, J. V. Unconventional translation of mammalian LINE-1 retrotransposons. *Genes Dev.* **20**, 210–224 (2006).
73. Paolino, A. *et al.* Differential timing of a conserved transcriptional network underlies divergent cortical projection routes across mammalian brain evolution. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 10554–10564 (2020).
74. Blaess, S. *et al.* Temporal-spatial changes in Sonic Hedgehog expression and signaling reveal different potentials of ventral mesencephalic progenitors to populate distinct ventral midbrain nuclei. *Neural Dev.* **6**, 29 (2011).
75. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in

342 eukaryotic genomes. *Mob. DNA* **6**, 11 (2015).
343 76. Kumaki, Y., Oda, M. & Okano, M. QUMA: quantification tool for methylation analysis.
344 *Nucleic Acids Res.* **36**, W170–5 (2008).
345 77. Faulkner, G. J. *et al.* A rescue strategy for multimapping short sequence tags refines surveys
346 of transcriptional activity by CAGE. *Genomics* **91**, 281–288 (2008).
347 78. Lanciano, S. & Cristofari, G. Measuring and interpreting transposable element expression.
348 *Nat. Rev. Genet.* **21**, 721–736 (2020).
349 79. Hashimoto, T. *et al.* Probabilistic resolution of multi-mapping reads in massively parallel
350 sequencing data using MuMRescueLite. *Bioinformatics* **25**, 2613–2614 (2009).
351 80. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21
352 (2013).
353 81. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
354 *EMBnet.journal* **17**, 10–12 (2011).
355 82. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
356 *arXiv [q-bio.GN]* arXiv:1303.3997 (2013).
357 83. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–
358 2079 (2009).
359 84. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
360 85. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–
361 3100 (2018).
362 86. Simpson, J. T. *et al.* Detecting DNA cytosine methylation using nanopore sequencing. *Nat.*
363 *Methods* **14**, 407–410 (2017).
364 87. Li, H. Tabix: fast retrieval of sequence features from generic TAB-delimited files.
365 *Bioinformatics* **27**, 718–719 (2011).
366 88. Cheetham, S. W., Kindlova, M. & Ewing, A. D. Methylartist: Tools for Visualising
367 Modified Bases from Nanopore Sequence Data. *bioRxiv* (2021).
368 89. Gubert, C. & Hannan, A. J. Environmental enrichment as an experience-dependent
369 modulator of social plasticity and cognition. *Brain Res.* **1717**, 1–14 (2019).
370 90. Mazarakis, N. K. *et al.* 'Super-Enrichment' Reveals Dose-Dependent Therapeutic Effects of
371 Environmental Stimulation in a Transgenic Mouse Model of Huntington's Disease. *J.*
372 *Huntingtons Dis.* **3**, 299–309 (2014).

91. Khan, A. *et al.* JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* **46**, D260–D266 (2018).
92. Wang, F. *et al.* RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. *J. Mol. Diagn.* **14**, 22–29 (2012).
93. Bodea, L.-G. *et al.* Neurodegeneration by activation of the microglial complement-phagosome pathway. *J. Neurosci.* **34**, 8546–8556 (2014).
94. Filice, F., Vörckel, K. J., Sungur, A. Ö., Wöhr, M. & Schwaller, B. Reduction in parvalbumin expression not loss of the parvalbumin-expressing GABA interneuron subpopulation in genetic parvalbumin and shank mouse models of autism. *Mol. Brain* **9**, (2016).
95. O'Driscoll, C., Kaufmann, W. E. & Bressler, J. Relationship between Mecp2 and NFκb signaling during neural differentiation of P19 cells. *Brain Res.* **1490**, 35–42 (2013).

Data availability

ONT sequencing data (.fastq and .fast5) generated from hippocampal PV^+ and PV^- cell pools are available from the European Nucleotide Archive (ENA) under accession number PRJEB47835.

Materials availability

L1.3 retrotransposition assay constructs carrying mutant SOX6 binding sites are available from Geoffrey J. Faulkner and require a material transfer agreement.

Code availability

Nanopore methylation analyses were performed with MethylArtist (https://github.com/adamewing/methylartist). Bisulfite sequencing results were visualized with QUMA (http://quma.cdb.riken.jp/). RNA-seq and ATAC-seq datasets were analyzed by pipelines joining together in serial published bioinformatic tools (see **Methods**).

Methods

Cultured PA-1 cell L1-EGFP assay

Retrotransposition efficiency was measured for L1.3, a highly mobile human L1HS[28,29], carrying an enhanced green fluorescent protein (EGFP) reporter cassette driven by a cytomegalovirus

promoter (CMVp), with the L1 expressed from its native promoter and delivered by a plasmid backbone also incorporating a puromycin resistance gene for selecting transfected cells[13,26,27]. In this system, the entire L1 3′UTR, with the thymine base deleted from within its native polyadenylation signal, precedes the EGFP cassette[6]. Vectors carrying wild-type and reverse transcriptase mutant[13] (D702A) L1.3 sequences, as well as L1.3 sequences with either of their 5′UTR SOX binding sites scrambled or inverted[20], were tested in cultured PA-1 cells, in normal media only and treated with trichostatin A prior to flow cytometry, as described previously[6]. L1 constructs with altered SOX6 sites were built by PCR fusion using overlapping primers that included the desired mutations. The results shown in **Fig. 1b** are one representative experiment of three biological replicates showing a similar trend per assayed construct. As a quality check, plasmid transfection efficiencies were calculated by co-transfecting with pCEP-EGFP into each cell line[71,72]. No untransfected PA-1 cells survived treatment with puromycin, ensuring untransfected cells did not contribute to $EGFP^-$ cells on the day of analysis. Untransfected cells not treated with puromycin were used to set the $EGFP^-$ signal level in flow cytometry.

L1-EGFP transgenic mice

To trace retrotransposition of an engineered L1 reporter *in vivo*, we generated a new transgenic L1-EGFP mouse line harboring L1.3, with epitope tags on ORF1p and ORF2p and an EGFP indicator cassette[13,26] embedded in its 3′UTR. To assemble the L1 transgene, we cloned the NotI-BstZ17I fragment from pJM101/L1.3-ORF1-T7-ORF2-3×FLAG (containing T7 gene 10 epitope tag on the C-terminus of ORF1 and a 3×FLAG tag on the C-terminus of ORF2) into p99-GFP-LRE3, yielding p99-GFP-L1.3-ORF1-T7-ORF2-3×FLAG. Both pJM101/L1.3-ORF1-T7-ORF2-3×FLAG and p99-GFP-LRE3 were kind gifts from Jose Garcia-Perez (University of Edinburgh). In p99-GFP-L1.3-ORF1-T7-ORF2-3×FLAG, transgene transcription was driven by the native L1.3 promoter, with an SV40 polyadenylation signal (pA) located downstream of the EGFP retrotransposition indicator cassette. The EGFP cassette was equipped with a cytomegalovirus (CMV) promoter and a herpes simplex virus type 1 (HSV) thymidine kinase (TK) polyadenylation signal, facilitating EGFP expression upon genomic integration via retrotransposition. In preparation for pronuclear injection, EGFP-L1.3-ORF1-T7-ORF2-3×FLAG was released by digestion with Not1 and MluI restriction enzymes, separated from the vector backbone on a 0.7% agarose gel, purified by phenol-chloroform extraction, and eluted in microinjection buffer (7.5mM

Tris-HCl, 0.15mM EDTA pH7.4). Transgenic L1-EGFP mice were produced by the Transgenic Animal Service of Queensland (TASQ), University of Queensland, using a standard pronuclear injection protocol. Briefly, zygotes were collected from superovulated C57BL/6 females. The microinjection buffer containing EGFP-L1.3-ORF1-T7-ORF2-3×FLAG was then transferred to the zygote pronuclei. Successfully injected zygotes were transplanted into the oviducts of pseudopregnant females. Primers flanking the EGFP cassette were used to screen potential founders by PCR (**Supplementary Table 2**). Identified founder L1-EGFP animals were bred on a C57BL/6 background. All procedures were followed as approved by the University of Queensland Animal Ethics Committee (TRI/UQ-MRI/381/14/NHMRC/DFG and MRI-UQ/QBI/415/17).

In utero *electroporation*

Embryonic *in utero* electroporation was employed to simultaneously deliver control (pmCherry) and experimental (L1) plasmids. Here, pmCherry *w*as a 4.7kb plasmid that expressed mCherry fluorescent protein under the control of a CMV promoter (Addgene 632524). L1 plasmids consisted of pUBC-L1SM-UBC-EGFP and pMut2-UBC-L1SM-UBC-EGFP. pUBC-L1SM-UBC-EGFP was a derivative of cep99-GFP-L1SM, which contained a full-length codon-optimized synthetic mouse L1 T_F element (L1SM, kindly shared by Jef Boeke, NYU Langone)[39], where mouse ubiquitin C (UBC) promoters were substituted for the CMV promoters used to drive L1SM and EGFP expression in cep99-GFP-L1SM. pMut2-UBC-L1SM-UBC-EGFP was identical to pUBC-L1SM-UBC-EGFP, apart from two non-synonymous mutations in the L1SM ORF2 sequence known to disable ORF2p reverse transcriptase and endonuclease activities. *In utero* electroporation was performed as described previously[73], with the day of mating defined as embryonic day 0 (E0). Briefly, time-mated pregnant CD1 mice were anesthetized at E14.5 via an intraperitoneal injection of ketamine/xylazine (120mg/kg ketamine and 10mg/kg xylazine). Embryos were exposed via a laparotomy and 0.5-1.0μL of plasmid DNA combined with 0.0025% Fast Green dye, to aid visualization, was injected into the lateral ventricle of each embryo using a glass-pulled pipette connected to a Picospritzer II (Parker Hannifin). Injections involved either combinations of pUBC-L1SM-UBC-EGFP and pmCherry (1μg/μL each) or pMut2-UBC-L1SM-UBC-EGFP and pmCherry (1μg/μL each). Half of the pups from each litter were co-injected with pUBC-L1SM-UBC-EGFP and pmCherry into the left hemisphere and the other half with pMut2-UBC-L1SM-UBC-EGFP and pmCherry into the right hemisphere. Plasmids were directed into the

forebrain by placement of 3mm diameter microelectrodes across the head, which delivered 5 (100ms, 1Hz) approximately 36V square wave pulses via an ECM 830 electroporator (BTX). Once embryos were electroporated, uterine horns were replaced inside the abdominal cavity and the incision sutured closed. Dams received 1mL of Ringer's solution subcutaneously and an edible buprenorphine gel pack for pain relief. Dams were monitored daily until giving birth to live pups, which were collected for analysis at P10.

Histology

Adult transgenic L1-EGFP mice (12-16 weeks) were anesthetized using isoflurane, and perfused intracardially with PBS and 4% PFA. CD1 pups, having been electroporated *in utero* with mouse L1-EGFP plasmids, were euthanized at postnatal day 10 by cervical dislocation. 12-week old CBA×C57BL/6 mice, intended for RNA FISH, were injected intraperitoneally with sodium pentobarbital (50mg/kg), followed by cervical dislocation to ensure euthanasia. All brains were dissected and fixed in PFA for 24h. For cryopreservation, fixed brains were immersed first in 15% sucrose and then 30% sucrose to submersion, and embedded in optimal cutting temperature (OCT) compound and stored at -80^{o}C. Transgenic L1-EGFP animal brains were sectioned on a cryostat (Leica, settings OT=-20^{o}C, CT=-20^{o}C) at 40µm thickness. Free-floating sections were collected in PBS and stored at 4^{o}C. CBA×C57BL/6 brains were sectioned on a cryostat (Leica, settings OT=-22^{o}C, CT=-22^{o}C) at 30µm thickness. Free-floating sections were collected in cryoprotectant (25% glycerol, 35% ethylene glycol, in PBS) and immediately stored at -20^{o}C.

Tissue processing and immunofluorescent staining with primary and secondary antibodies were carried out as described previously[74]. Primary antibodies and dilutions were as follows: rabbit anti-GFP, 1:500 (Thermo Fisher A11122); chicken anti-GFP, 1:500 (Millipore AB16901); mouse anti-T7, 1:500 (Millipore 69522); rabbit anti-T7, 1:500 (Millipore AB3790); goat anti-tdTomato, 1:1000 (Sicgen T2200); mouse anti-NeuN, 1:250 (Millipore MAB377); guinea pig anti-NeuN, 1:250 (Millipore ABN90), rabbit anti-Gad65/67 (GAD1), 1:500 (Sigma G5163); mouse anti-parvalbumin (PV), 1:2000 (Sigma P3088); rabbit anti-β tubulin III (Tub), 1:500 (Sigma T2200); rabbit anti-MeCP2, 1:500 (Abcam ab2828). Secondary antibodies and dilutions were as follows: donkey anti-guinea pig Dylight 405, 1:200 (Jackson Immunoresearch 706475148); donkey anti-mouse Dylight 405, 1:200 (Jackson Immunoresearch 715475150); donkey anti-chicken Alexa Fluor 488, 1:500 (Jackson Immunoresearch 703546155); donkey anti-rat Alexa Fluor 488, 1:500

(Jackson Immunoresearch 712546150); donkey anti-rabbit Alexa Fluor 488, 1:500 (Thermo Fisher A21206); donkey anti-goat Alexa Fluor 594, 1:500 (Jackson Immunoresearch 705586147); donkey anti-rabbit Cy3, 1:200 (Jackson Immunoresearch 711165152); donkey anti-mouse Cy3, 1:500 (Jackson Immunoresearch 715165150); donkey anti-guinea pig Alexa Fluor 647, 1:500 (Millipore AP193SA6); donkey anti-mouse Alexa Fluor, 1:500 (Jackson Immunoresearch 715606150). For nuclei labelling: BisBenzimide H33258 (Sigma B2883). Blocking serum: normal donkey serum (Jackson Immunoresearch 017000121).

Imaging

$EGFP^+$ cells were imaged on a Zeiss LSM510 confocal microscope. Acquisition of high magnification, Z-stack images was performed with Zen 2009 software. Images of EGFP, NeuN and PV immunostaining for quantification were taken from hippocampal and adjacent cortical areas using a Zeiss AxioObserver Z1 microscope and Zen 2009 software, equipped with an ApoTome system and a 10× objective. Visualization and imaging of EGFP, NeuN and PV in *in utero* electroporated mice was performed using a Zeiss Plan-Apochromat 20x/0.8 NA air objective and a Plan-Apochromat 40×/1.4 NA oil-immersion objective on a confocal/two-photon laser-scanning microscope (LSM 710, Carl Zeiss Australia) built around an Axio Observer Z1 body and equipped with two internal gallium arsenide phosphide (GaAsP) photomultiplier tubes (PMTs) and three normal PMTs for epi- (descanned) detection and two external GaAsP PMTs for non-descanned detection in two-photon imaging, and controlled by Zeiss Zen Black software. RNA FISH for sections of hippocampus and adjacent cortical areas, as well as MeCP2, NeuN and PV immunostainings were imaged on a spinning-disk confocal system (Marianas; 3I, Inc.) consisting of a Axio Observer Z1 (Carl Zeiss) equipped with a CSU-W1 spinning-disk head (Yokogawa Corporation of America), ORCA-Flash4.0 v2 sCMOS camera (Hamamatsu Photonics), using a 63×/1.4 NA C-Apo objective and a 20×/0.8 NA Plan-Apochromat objective, respectively. All Z-stack spinning-disk confocal image acquisition was performed using SlideBook 6.0 (3I, Inc). PV stereology was performed on an upright Axio Imager Z2 fluorescent microscope (Carl Zeiss) equipped with a motorized stage and Stereo Investigator software (MBF Bioscience). Contours were drawn based on DAPI staining using a 5×/0.16 NA objective. Counting was performed on a 10×/0.3 NA objective. All image processing and analysis post acquisition were performed using Fiji for Windows (ImageJ 1.52d).

Single molecule RNA fluorescence in situ *hybridization (FISH)*

Two custom RNAscope probes were designed against the RepBase[75] L1 T_{FI} subfamily consensus sequence (**Extended Data Fig. 3**). L1 probe A (design #NPR-0003768, Advanced Cell Diagnostics, Cat. #ADV827911C3) targeted the L1 T_{FI} 5′UTR monomeric and non-monomeric region (consensus positions 827 to 1688). L1 probe B (design #NPR-000412, Advanced Cell Diagnostics, Cat. #ADV831481C3) targeted the L1 T_{FI} 5′UTR monomeric region (consensus positions 142 to 1423). Weak possible off-target loci for probe A and B comprised the pseudogene Gm-17177, two non-coding RNAs (LOC115486508 for probe A and LOC115490394 for probe B) and a minor isoform of the PPCDC gene (only for probe A), none of which were expressed beyond very low levels or with specificity to PV^+ neurons. Using the L1 T_F RNAscope probes, we performed fluorescence *in situ* hybridization (FISH) on fixed, frozen brain tissue according to the manufacturer's specifications (RNAscope Fluorescent Multiplex Reagent Kit part 2, Advanced Cell Diagnostics, Cat. #320850) and with the following modifications: 30µm coronal sections instead of 15µm, and boiling in target retrieval solution for 10min instead of 5min. To identify neurons, we performed immunohistochemistry using a rabbit anti-β-tubulin antibody (Sigma Cat. #T2200) and donkey anti-rabbit Cy3 secondary antibody (Jackson Immunoresearch, Cat. #711165152) following a previously described protocol[74]. To identify PV^+ neurons we employed a validated mouse PV RNAscope probe (Mm-Pvalb-C2, Advanced Cell Diagnostics, Cat. #ADV421931C2). Probes for the ubiquitously expressed mouse peptidylprolyl isomerase B (PPIB) gene and *Escherichia coli* gene *dapB* were used as positive and negative controls, respectively, for each FISH experiment.

Cell quantifications

L1 T_F and *PV RNA FISH:* We analyzed four hippocampal sections per animal for each L1 T_F 5′UTR probe (**Extended Data Fig. 3**) using Imaris 9.5.1 (Bitplane, Oxford Instruments). To render 3D visualizations for a given neuron, we used Tub and DAPI staining to outline its soma and nucleus along Z-stack planes where the cell was detected. We set voxels outside the cell, and inside the nucleus, to a channel intensity value of zero to only retain cytoplasmic L1 mRNA signal and avoid nuclear L1 DNA. We then calculated the mean intensity of the L1 and PV channels within the cytoplasm. To quantify relative L1 T_F mRNA expression in PV^+/Tub^+ versus PV^-/Tub^+ neurons

we normalized values to the mean value of PV^-/Tub^+ neurons from each mouse. As a result, data from PV^-/Tub^+ neurons are presented as mean intensity raw values. *MeCP2*: To quantify MeCP2 protein expression we analyzed two hippocampal sections per animal. For each cell, we drew the contours of NeuN immunostaining along the relevant Z-stack planes and rendered a cell 3D visualization. We then calculated the mean MeCP2 channel intensity in $PV^+/NeuN^+$ and $PV^-/NeuN^+$ neurons. *PV stereology*: We stained and analyzed every 12th hippocampal section per animal. Cell density was calculated using the total number of PV^+ cells and the total subregion area from ~6 sections per animal. *L1-EGFP*: To quantify $EGFP^+$ cells we stained and analyzed every 12th hippocampal section (again, ~6 sections per animal). To visualize colocalization, we used Adobe Photoshop CC 2017. We counted $EGFP^+$, $EGFP^+/NeuN^+$ and $EGFP^+/PV^+$ cells across the entire hippocampus and adjacent cortex. The average number of double-labeled cells per $100mm^2$ was determined for each animal. All statistical analyses were performed using Prism (v8.3.1)

Cell sorting and nucleic acid isolation

Neonate litters were obtained from time-mated C57BL/6 mice bred in-house at the QBI animal facility. The day of birth was defined as postnatal day 0 (P0). From each P0 litter of ~6 pups we dissected and pooled hippocampus tissue. Tissues were dissociated in a papain solution, containing approximately 20U papain (Worthington) and 0.025mg DNase I (Worthington). Prior to use, papain was dissolved in HBSS (Gibco) with 1.1mM EDTA (Invitrogen), 0.067mM mercaptoethanol (Sigma) and 5mM cysteine-HCL (Sigma), and diluted in Hibernate E medium (Gibco). Tissue was incubated for 10min at 37°C with 0.5mL papain solution per embryo. Following digestion, the cell suspension was passed through a 70µm mesh cell strainer, washed into Hibernate E supplemented with B27 (Gibco) and then centrifuged at 300g for 5min. From this point in the protocol onwards, reagents were pre-chilled and the remaining procedures performed on ice. The cell pellet was resuspended in a blocking buffer (HBSS with 5% BSA). A rabbit anti-PV conjugated Alexa Fluor 647 antibody (Bioss bs-1299R-A647, dilution 1:2000) was directly added to the blocking buffer cell suspension and incubated for 1h at 4°C, then passed through a 40µm mesh cell strainer and subjected to flow cytometry. The cell suspension was run through a 100µm nozzle at low pressure (28psi) on a BD FACSAria II flow cytometer (Becton Dickinson). This first sort isolated PV^+ and PV^- cells. To further isolate PV^- neurons, PV^- cells from the first

sort were collected in tubes containing 40U RNAseOUT ribonuclease inhibitor (Invitrogen), then fixed in ice cold 50% ethanol for 5min and centrifuged at 300g for 7min. Following centrifugation, cells were immunostained in blocking buffer containing mouse anti-beta III Tubulin (Tub) conjugated Alexa Fluor 488 antibody (Abcam ab195879, dilution 1:1000) and DAPI (Sigma D9542, 1µg/mL) for 15min at 4ºC. Tub^+ immunostained cells were subjected to a second sort on the same FACS machine and specification as above. Four populations of cells were collected: PV^+ and PV^- (**Extended Data Fig. 5a,** sort 1) and PV^-/Tub^+ and PV^-/Tub^- (**Extended Data Fig. 5a,** sort 2). DNA and RNA were then extracted from each cell population. For RNA extractions, cells were sorted directly into the lysis buffer provided in the NucleoSpin RNA XS kit (Macherey Nagel), with RNA extraction performed following the manufacturer's specifications, except DNAse treatment was performed on a column twice for 20min, instead of once for 15min. For DNA extraction, purified cells were collected into a DNA lysis buffer containing TE buffer (10mM Tris-HCl pH 8 and 0.1mM EDTA), 2% SDS and 100µg/mL proteinase K, and DNA was extracted following a standard phenol-chloroform protocol.

Quantitative PCR on sorted cells and bulk hippocampus

Total RNA extracted from purified PV^+, PV^-, PV^-/Tub^+ and PV^-/Tub^- (**Extended Data Fig. 5a,** sorts 1 and 2) populations was used as input for SYBR Green and TaqMan qPCR assays. qPCR reactions were carried out using 300pg RNA/µL from purified PV^+ and PV^- cells and 100pg RNA/µL from purified PV^-/Tub^+ and PV^-/Tub^- cells. An RNA integrity number (RIN) above 6, as measured on an Agilent Bioanalyzer (Agilent Technologies, RNA 6000 Pico Kit, Cat. #5067-1513), was set as the minimum cutoff for RNA quality. All qPCRs were carried out on a LightCycler 480 Real-Time PCR system (Roche Life Science). Oligonucleotide PCR primers, as listed in **Supplementary Table 2**, were purchased from Integrated DNA Technologies. *SYBR Green assay*: PCR reactions were prepared using the Power SYBR Green RNA-to-CT 1 step kit (Applied Biosystems, Cat. #4391112). Reactions contained a 2× Power SYBR Green RT-PCR Mix, 10pmol of each primer, 1µL RNA input template and 1× reverse transcriptase enzyme mix in a 10µL final volume. Cycling conditions were as follows: 48ºC for 30min, 95ºC for 10min, followed by 40 cycles of 95ºC, 15sec; 60ºC, 1min. To assess potential DNA contamination, an L1 T_F qPCR using primers L1Md_5UTR_F and L1Md_5UTR_R was performed with and without reverse transcriptase. A three or more cycle difference between experiments run with and without

reverse transcriptase, and detection after cycle 30 in the latter, was considered as non-DNA contaminated RNA. *TaqMan assay*: Applied Biosystems custom L1, URR1 and 5S rRNA TaqMan MGB probes, as listed in **Supplementary Table 2**, were purchased from Thermo Fisher (Cat. #4316032), as was a proprietary mouse GAPDH combination (Cat. #4352339E). TaqMan qPCR reactions contained: 4× TaqPath 1-Step RT-qPCR multiplex reaction master mix (ThermoFisher, Cat. #A28521), 4pmol of each primer, 1pmol probe (with the exception of the ORF2/URR1 TaqMan reaction, for which we used 1pmol ORF2 primers) and 1µL RNA input template in a 10uL final volume. Cycling conditions were as follows: 37ºC for 2min; 50ºC for 15min; 95ºC for 2min, followed by 40 cycles of 95ºC, 3sec; 60ºC, 30sec. TaqMan assays for L1 were multiplexed with assays for either 5S rRNA, GAPDH or URR1 controls. L1 probes were conjugated to VIC or 6FAM fluorophores. Controls were conjugated to HEX, VIC or 6FAM fluorophores. Primer/probe sequences and the associated detection channels are listed in **Supplementary Table 2**. For each assay, the relative mRNA expression in a particular sample was calculated by the delta delta-CT method, using the negative population in the respective sort as control, i.e. PV^+ was compared to PV^- (**Extended Data Fig. 5a**, sort 1) and PV^-/Tub^+ compared to PV^-/Tub^- (**Extended Data Fig. 5a**, sort 2). As the PV^-/Tub^+ and PV^-/Tub^- populations were isolated as a result of two sortings in serial, for some assays sufficient RNA was only available to perform qPCR on PV^+ and PV^- populations. For qPCR on bulk hippocampus, tissue was isolated from 12-week old animals housed in standard (STD, *N*=12) and enriched (ENR, *N*=14) environments. RNA extraction was performed by Trizol following the manufacturer's specifications (Trizol reagent, Invitrogen Cat. #15596026). Quantitative TaqMan PCR assays were performed as described above, using 40ng of RNA as input.

CAPS2.L1 5'RACE and RT-PCR

For 5′RACE, hippocampus tissue from three adult C57BL/6 mice was pooled and RNA extracted (Trizol reagent, Invitrogen Cat. #15596026). RNA was used as input for a FirstChoice RLM-RACE Kit (Invitrogen, Cat. #AM1700) to generate cDNA from capped, full-length mRNAs, following the manufacturer's specifications. Total RNA extracted from purified PV^+, PV^- and pooled neonate hippocampi was reverse transcribed using a High-Capacity cDNA Reverse Transcription Kit (Invitrogen, Cat. #4368814). PCR amplification was then performed using 1U MyTaq HS DNA Polymerase (BioLine) in 1× MyTaq buffer, 10pmol primer CAPS2.L1_F,

10pmol primer CAPS2.L1_R, 1µL cDNA in a 20µL final volume reaction. PCR cycling conditions were as follows: 95ºC for 1min, (95ºC for 15sec; 55ºC for 15sec; 72ºC for 10sec)×38, 72ºC for 5min. Reaction products were run on a 1.5% agarose gel in 1×TAE, stained with SYBR Safe DNA gel stain.

L1 T_F promoter bisulfite sequencing

Targeted bisulfite sequencing was performed as described previously[43] to assess L1 T_F 5′UTR monomer CpG methylation genome-wide. Briefly, this involved extraction of genomic DNA from PV^+, PV^- and PV^-/Tub^+ populations purified from hippocampus tissue pooled from neonate littermates (**Extended Data Fig. 5**). Approximately 4×10^4 events per population were obtained from each of 3 litters (experimental triplicates). DNA was extracted via a conventional phenol-chloroform method and ethanol precipitation aided by glycogen (Ambion). DNA concentration was assessed with a Qubit dsDNA HS assay kit. Next, 20ng of genomic DNA was bisulfite converted using the EZ-DNA Methylation Lightning kit (Zymo Research, Cat #D5030) following the manufacturer's specifications. Bisulfite PCR reactions used MyTaq HS DNA polymerase (Bioline), and contained 1× reaction buffer, 12.5pmol of each primer, 2µL bisulfite treated DNA input template and 1U of enzyme in a 25µL final volume. PCR cycling conditions were as follows: 95ºC for 2min, followed by 40 cycles of 95ºC, 30sec; 54ºC, 30sec; 72ºC, 30sec and 1 cycle of 72ºC, 5min. Primer sequences (BS_L1_TF_F and BS_L1_TF_R) were as provided in **Supplementary Table 2**. PCR products were visualized by electrophoresis on a 2% agarose gel, followed by the excision of fragments of expected size and DNA extracted using a MinElute gel extraction kit (Qiagen, Cat #28604) following the manufacturer's specifications. DNA concentration was assessed with a Qubit dsDNA HS assay kit and 30ng converted DNA was used as input for library preparation. Libraries were prepared using a NEBNext Ultra II DNA library prep kit (NEB, E7645S) and NEBNext Multiplex Oligos for Illumina (NEB, Cat# E6609S). Libraries were eluted in 15µL H_20 and concentrations measured with an Agilent 2100 Bioanalyzer using an Agilent HS DNA kit (Agilent Technologies, Cat. 5067-4627). Barcoded libraries of PV^+ and PV^-/Tub^+ populations from each of the 3 litters were mixed in equimolar quantities, diluted to 8nM, and combined with 50% PhiX spike-in control (Illumina, Cat #FC-110-3001). Single-end 300mer sequencing was then performed on a MiSeq platform (Illumina) using a MiSeq Reagent v3 kit (Illumina, Cat #MS-102-3003). Data were then analyzed as described elsewhere[6]. To

summarize, reads with the L1 T_F bisulfite PCR primers at their termini were retained and aligned to the mock converted T_F monomer target amplicon sequence with blastn. Reads where non-CpG cytosine bisulfite conversion was <95%, or ≥5% of CpG dinucleotides were mutated, or ≥5% of adenine and guanine nucleotides were mutated, were removed. 100 reads per triplicate cell population, excluding identical bisulfite sequences, were randomly selected and analyzed using QUMA[76] with default parameters, with strict CpG recognition.

RNA-seq analysis

The mappability of individual TE copies generally varies as a function of sequencing read length, as well as TE subfamily age and copy number[77,78]. We therefore adapted a prior approach to quantify young mouse (L1 T_F) and human (L1HS and L1PA2) subfamily-level transcript abundance with RNA-seq[55,77,79]. Analyzed datasets included Sams *et al.*[37], bulk hippocampus single-end (1×61mer) RNA-seq obtained from wild-type and conditional CTCF knockout animals (SRA: SRP078142, *N*=3 pools of 2 animals per group), and Yuan *et al.*[36] bulk single-end (1×49mer) RNA-seq of neurons differentiated *in vitro* from human induced pluripotent stem cells, with and without LHX6 overexpression (SRA: SRP147748, *N*=3 per group). For each RNA-seq library, we aligned reads to the reference genome (mouse: mm10, human: hg38) genome assembly with STAR[80] version 2.6 (parameters --twopassMode Basic --outSAMprimaryFlag AllBestScore --winAnchorMultimapNmax 1000 --outFilterMultimapNmax 1000) and marked duplicate reads with Picard MarkDuplicates (http://broadinstitute.github.io/picard). We expected the high copy number and limited divergence of young L1 subfamilies to cause most of the corresponding RNA-seq reads to "multi-map" to multiple genomic loci[77,78]. As conceived previously, we assigned multi-map reads a weighting at each of their aligned positions based on the abundance of uniquely mapping reads aligned within 100bp in the same library[55,77,79]. Each position was then assigned a weighting proportionate to the fraction of uniquely mapped reads found there, out of the total number of uniquely mapped reads within 100bp of any mapping position for the multi-mapping read. If no uniquely mapped reads were found near any of the aligned positions for a multi-mapped read, all positions were given an equal weighting. We then intersected the unique and weighted multi-map alignments with RepeatMasker coordinates and produced a total read count for L1 T_F (RepeatMasker: "L1Md_T"), L1HS and L1PA2 genome-wide, normalized by dividing by the total mapped read count for that RNA-seq library (tags-per-million).

Bulk ATAC-seq analysis

Mouse cortex ATAC-seq data were previously generated by Mo *et al*.[35] for excitatory pyramidal neurons (marked by CAM2KA), PV interneurons and VIP interneurons, via the isolation of nuclei tagged in specific cell types (INTACT) method. Paired-end fastq files were obtained from the Sequence Read Archive (SRA identifiers SRR1647880-SRR1647885). Trim Galore (parameters --max_n 2 --length 50 --trim-n) was used to apply CutAdapt[81] to read pairs to trim adapters and low quality bases. Processed reads were aligned to the reference genome (mm10) using bwa mem[82] with parameters (-a) to output all multimapping alignments. Alignments were filtered to keep only those with an alignment score equal to the maximum achieved for that read. The resultant bam files were sorted using samtools[83]. Peaks for each combined pair of duplicate experiments were called using MACS2[84] with default parameters, intersected with young L1 genomic coordinates, and then used to calculate the fraction of reads in each replicate aligned to at least one L1-associated peak.

scATAC-seq analyses

Mouse hippocampus scATAC-seq data reported by Sinnamon *et al*.[32] were obtained from the SRA (identifiers SRR7749424 and SRR7749425). Only reads corresponding to the 2,346 cell identifiers reported by Sinnamon *et al.* were retained. Human hippocampus scATAC-seq data were reported by Corces *et al*.[33] (SRA identifiers SRR11442501 and SRR11442502). Human read pairs were retained if the corresponding barcode was present in the 10x Genomics scATAC-seq Unique Molecular Identifier (UMI) whitelist (737K version 1). Human and mouse read pairs were processed and aligned as per the bulk ATAC-seq above, using the hg38 and mm10 reference genome assemblies, respectively. Cells (UMIs) with fewer than 10,000 uniquely aligned read pairs were discarded. For each cell, we then determined the TPM fraction of reads overlapping -1000bp to +500bp of the annotated genomic start position of at least one young L1. For the mouse analysis, 2,629 young full-length (>6kbp) T_F L1s were identified "L1Md_T" as listed by the UCSC Genome Browser RepeatMasker track. For the human analysis, a cohort of 840 full-length (>5.9kbp) L1HS and L1PA2 elements defined previously[6] were employed. Cells were grouped based on having at least one read aligned within the genomic coordinates of a given gene, with these coordinates defined as the first 50kbp of genes longer than 50kbp. The average fraction of young L1-associated

reads was then calculated for each cell group, compared to all other cells. To assess the statistical significance of the observed L1-associated read fractions, permutation tests were performed to determine this fraction for random resamplings of the same number of cell identifiers, with 10^3 permutations.

Nanopore methylation analysis

High molecular weight DNA was extracted using a Nanobind CBB Big DNA Kit (Circulomics, NB-900-001-01) from PV^+ and PV^- (**Extended Data Fig. 5**) populations purified from 10 neonate (P0) littermate hippocampus sample pools. DNA samples were sheared to ~10kb average size, prepared as barcoded libraries using a Ligation Sequencing Kit (Oxford Nanopore Technologies, SQK-LSK109), and sequenced on two flow cells of an ONT PromethION platform (Kinghorn Centre for Clinical Genomics, Australia). Bases were called with Guppy 4.0.11 (Oxford Nanopore Technologies) and reads aligned to the mm10 reference genome using minimap2 version 2.20[85] and samtools version 1.12[83]. Reads were indexed and per-CpG methylation calls generated using nanopolish version 0.13.2[86]. Methylation likelihood data were sorted by position and indexed using tabix version 1.12[87]. Methylation statistics for the genome divided into 10kbp bins, and reference TEs defined by RepeatMasker coordinates (http://www.repeatmasker.org/), were generated using MethylArtist version 1.0.4[88], using commands db-nanopolish, segmeth and segplot with default parameters. Only full-length (>6kbp) L1s were included. Methylation profiles for individual loci were generated using the MethylArtist command locus, where parameters specified a 30bp sliding window with a 2bp step, and smoothed with a window size of 8 for the Hann function. To identify individual differentially methylated TEs (**Supplementary Table 1**), we required elements to have at least 4 reads and 20 methylation calls in each sample. Comparisons were carried out via Fisher's Exact Test using methylated and non-methylated call counts, with significance defined as a Bonferroni corrected P value of less than 0.01.

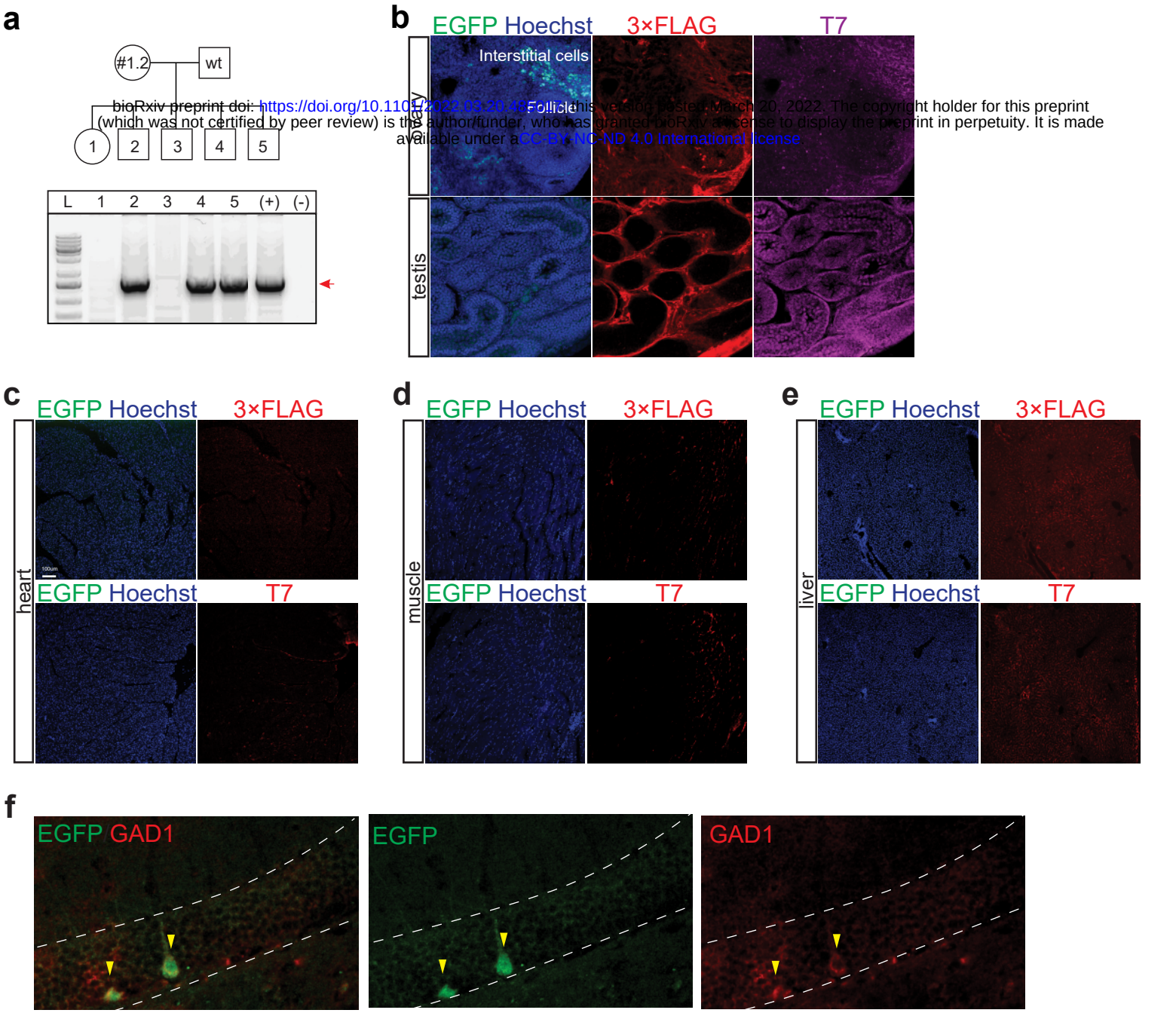
Environmental enrichment and exercise experimental design

At six weeks of age, CBA×C57BL/6 mice were randomly assigned to either a standard (STD), enriched (ENR) environment or exercise (EXE) group, as described previously[68]. All mice were exposed to their assigned housing condition for 6 weeks. Briefly, STD housing consisted of an open-top standard mouse cage (34 × 16 × 16cm; 4 mice/box) with basic bedding and nesting

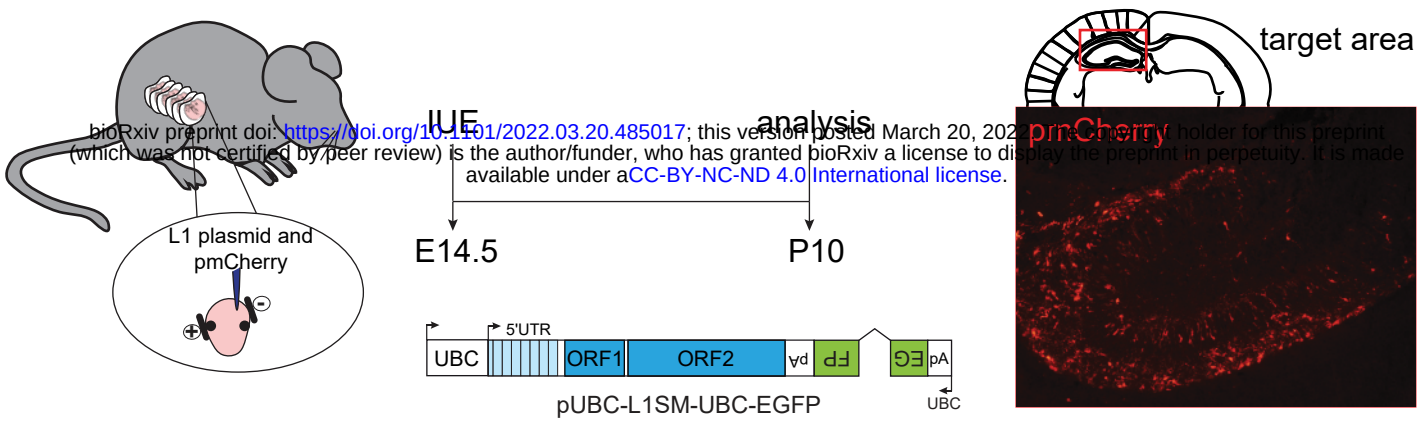
materials. ENR and EXE mice were housed in larger cages (40 × 28 × 18cm; 4 mice/box) containing the same basic bedding and nesting materials as the STD plus specific features. ENR cages contained climbing and tunneling objects together with inanimate objects of various textures, sizes, and shapes, which altogether confer the enhancement of sensory, cognitive and motor stimulation[89]. These cages were changed weekly to ensure novelty for ENR mice. In addition, from week ten, ENR mice were exposed three times a week for one hour to an extra 'super-enriched' condition in a larger playground arena (diameter: 57cm, height: 90cm) as previously described[90]. Each EXE cage contained two running wheels (12cm in diameter) to ensure mice had access to voluntary wheel running. Running wheels were excluded from the ENR housing to ensure the effects of physical activity were exclusive to the EXE mice. All mice had *ad libitum* access to food and water and were housed in a controlled room at 22°C and 45% humidity on a 12:12 hour light/dark cycle. All procedures were approved by The Florey Institute of Neuroscience and Mental Health Animal Ethics Committee (19-012-FINMH) and were performed in accordance with the relevant guidelines and regulations of the Australian National Health and Medical Research Council Code of Practice for the Use of Animals for Scientific Purposes.

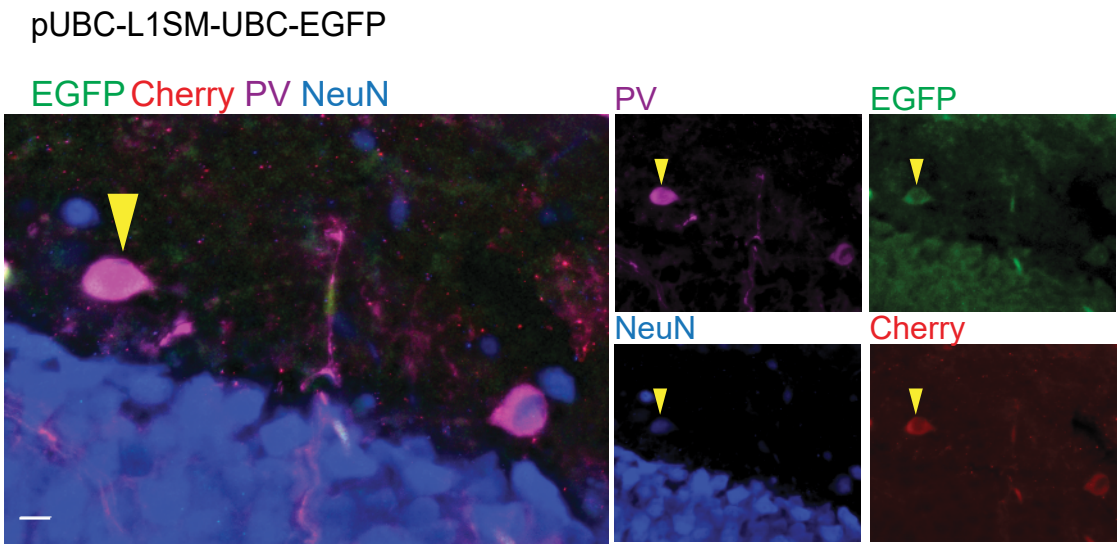
Supplementary Table 2. Primer and probe information.

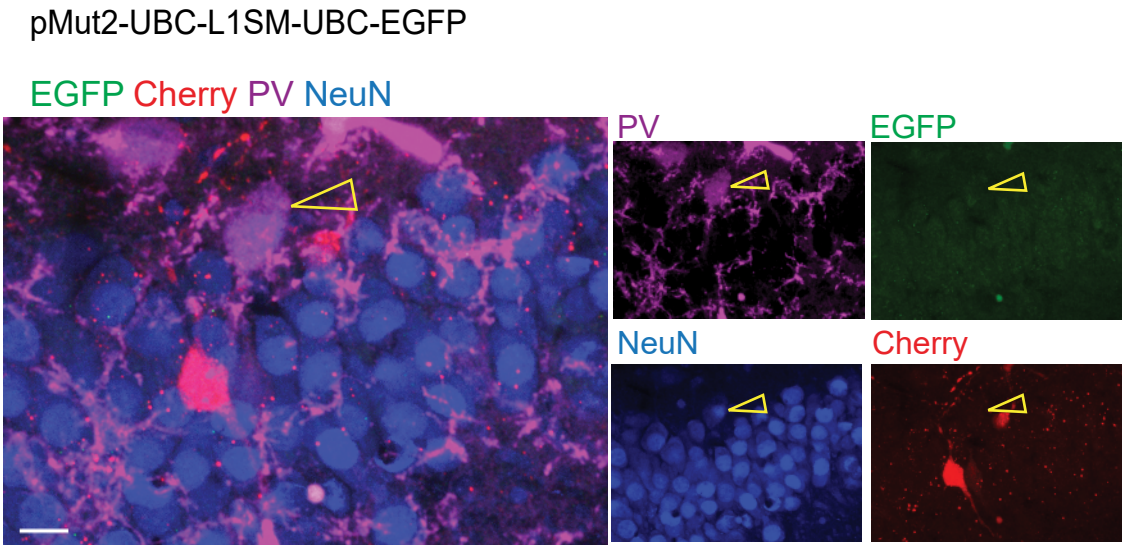
Name	Sequence	Target	Assay	Source
BS_L1TFI/II_F	GTTGAGGTAGTATTTTGTGTGGGT	L1 $T_{FI/II}$ 5′UTR	Bisulfite PCR	[43]
BS_L1TFI/II_R	TTCCAAAAACTATCAAATTCTCTAAC	L1 $T_{FI/II}$ 5′UTR	Bisulfite PCR	[43]
URR1_F	GTAGCAAAATTATAGTTATGAAGTA GC	URR1	TaqMan qPCR	This study.
URR1_R	CTCAACCTTCCTAATGCT	URR1	TaqMan qPCR	This study.
URR1 probe	HEX-TCACCACAACATGAGGAACTGTATTA -MGBNFQ	URR1	TaqMan qPCR	This study.
L1TFnonmon_F	AGGTCCAAATACAAGATATCTGC	L1 T_F 5′UTR	TaqMan qPCR	This study.
L1TFnonmon_R	ATCAGCAGACCTGGGAGACA	L1 T_F 5′UTR	TaqMan qPCR	This study.
L1TFnonmon probe	FAM-TGCCAGCAGAGAGTGCTCTGAGC -MGBNFQ	L1 T_F 5′UTR	TaqMan qPCR	This study.
TaqORF2 Fw	CATCAATGTAATCCATTATATAAAC	L1 T_F ORF2	TaqMan qPCR	This study.
TaqORF2 Rv	TTTATCATGAATGGGTGTTG	L1 T_F ORF2	TaqMan qPCR	This study.
TaqORF2 probe	FAM-CACATGATCATCTCGTTAGATGCAGA -MGBNFQ	L1 T_F ORF2	TaqMan qPCR	This study.
Mouse GAPDH assay	Proprietary VIC®/MGB probe and primers	GAPDH	TaqMan qPCR	ThermoFisher
L1Md_5UTR_F	CAGCCGGCCACCTTCC	L1 $T_{FI/II}$ 5′UTR	TaqMan qPCR	[34]
L1Md_5UTR_R	GGTCCCGGACCAAGATGG	L1 $T_{FI/II}$ 5′UTR	TaqMan qPCR	[34]
L1Md_5UTR probe	VIC-CGGAGGACAGGTGC-MGBNFQ	L1 $T_{FI/II}$ 5′UTR	TaqMan qPCR	[34]
5S_rRNA_F	ACGGCCATACCACCCTGAA	5S rRNA	TaqMan qPCR	[34]
5S_rRNA_R	GGTCTCCCATCCAAGTACTAACCA	5S rRNA	TaqMan qPCR	[34]
5S_rRNA_probe	FAM-CCGAGATCAGACGAGAT-MGBNFQ	5S rRNA	TaqMan qPCR	[34]
GAPDH_F	AACTTTGGCATTGTGGAAGG	GAPDH	SYBR qPCR	[93]
GAPDH_R	GGATGCAGGGATGATGTTCT	GAPDH	SYBR qPCR	[93]
PV_F	TGTCGATGACAGACGTGCTC	PV	SYBR qPCR	[94]
PV_R	TTCTTCAACCCCAATCTTGC	PV	SYBR qPCR	[94]
MeCP2_F	AGGAGAGACTGGAGGAAAAGT	MeCP2	SYBR qPCR	[95]
MeCP2_R	CTTAAACTTCAGTGGCTTGTCT	MeCP2	SYBR qPCR	[95]
L1eGFP_6154	AATATCACGGGTAGCCAACG	EGFP	Genotyping	This study.
L1eGFP_7850	TAGCGCTACCGGACTCAGAT	EGFP	Genotyping	This study.
CAPS2.L1_F	GTCACTCTTCCTGCCTGCTC	CAPS2.L1	RACE/RT-PCR	This study.
CAPS2.L1_R	AGTCAACAAGGCTTCCCAGA	CAPS2.L1	RACE/RT-PCR	This study.
Dnmt1_F	CCTAGTTCCGTGGCTACGAGGAGAA	DNMT1	SYBR qPCR	[46]
Dnmt1_R	TCTCTCTCCTCTGCAGCCGACTCA	DNMT1	SYBR qPCR	[46]
Dnmt3a_F	GCCGAATTGTGTCTTGGTGGATGACA	DNMT3A	SYBR qPCR	[46]
Dnmt3a_R	CCTGGTGGAATGCACTGCAGAAGGA	DNMT3A	SYBR qPCR	[46]



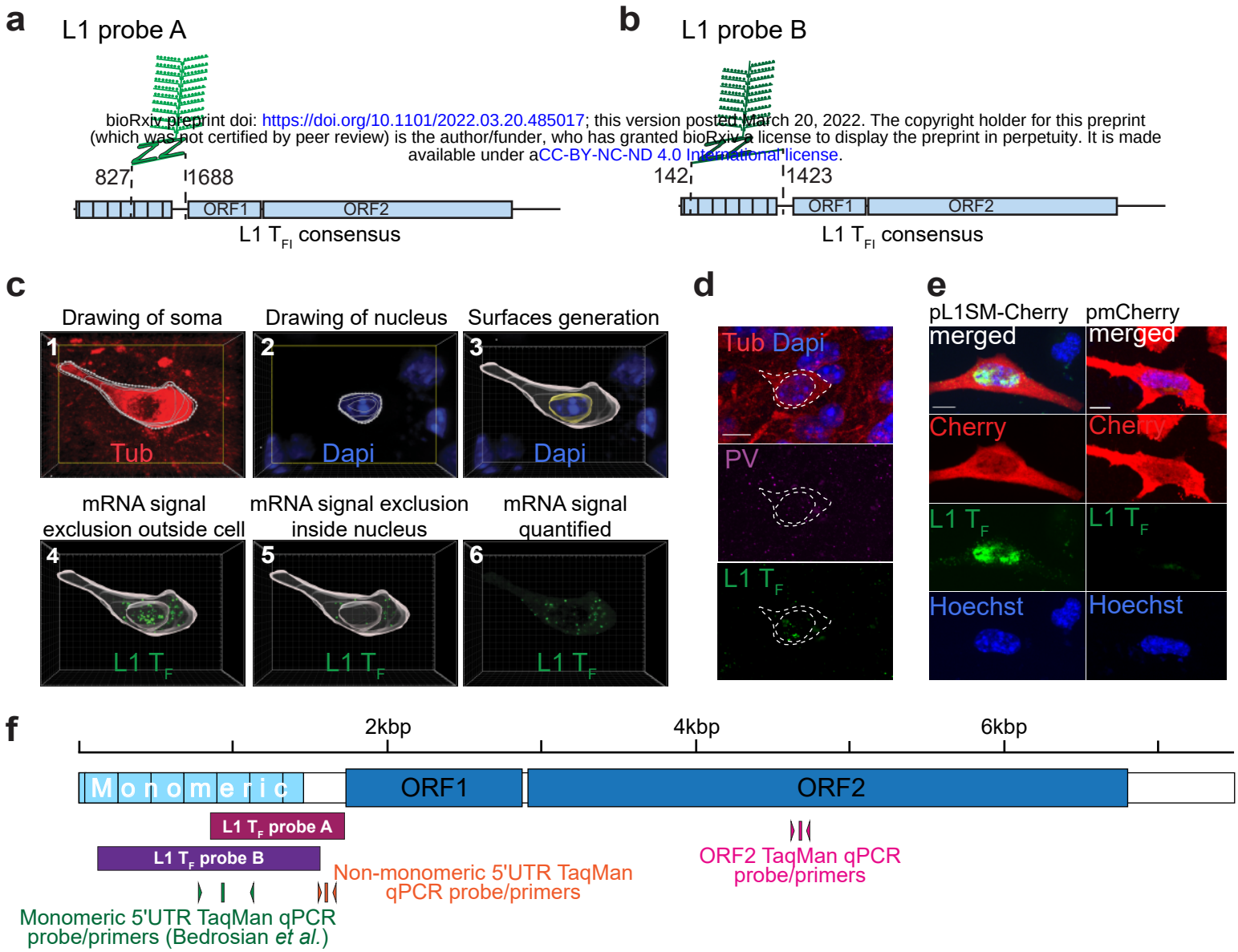
Extended Data Fig. 1: Assessment of L1 retrotransposition in brain and non-brain tissues of L1-EGFP animals. a, EGFP cassette genotyping PCR results for the offspring of founder animal #1.2. Circles and squares represent female and male mice, respectively. The indicated 1.7kbp PCR product (red arrow) corresponds to the integrated intron-containing EGFP indicator cassette. Gel labels are as follows: L, ladder; 1-5, transgenic offspring littermates; +, EGFP positive control plasmid DNA; -, H_20. Offspring 2, 4 and 5 carried the L1-EGFP transgene. **b,** Ovary and testis **c,** heart **d,** muscle and **e,** liver tissues of adult L1-EGFP mice were immunostained for EGFP and L1 proteins (T7-tagged ORF1p and 3×FLAG-tagged ORF2p). $EGFP^+$ cells were observed in the interstitial cells of the ovaries but not in other tissues. DNA was stained with Hoechst dye (blue). **f,** Representative maximum projection confocal image of a coronal hippocampus section from a transgenic L1-EGFP animal showing immunostaining for EGFP (green) and the interneuron marker GAD1 (red). Yellow arrowheads indicate $EGFP^+/GAD1^+$ neurons in DG. The image is presented as merged and single channels for EGFP and GAD1. Scale bar: 100µm.



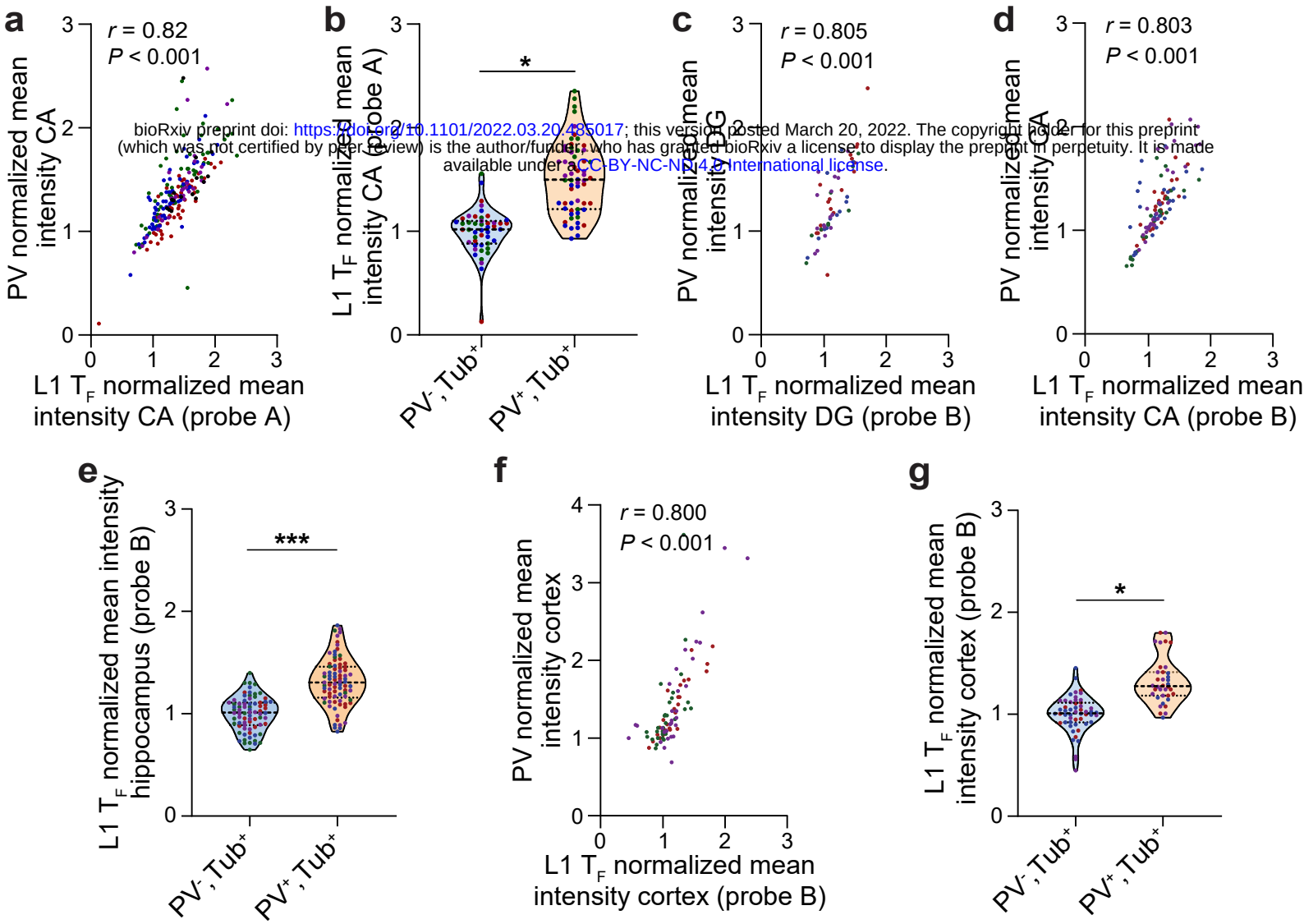




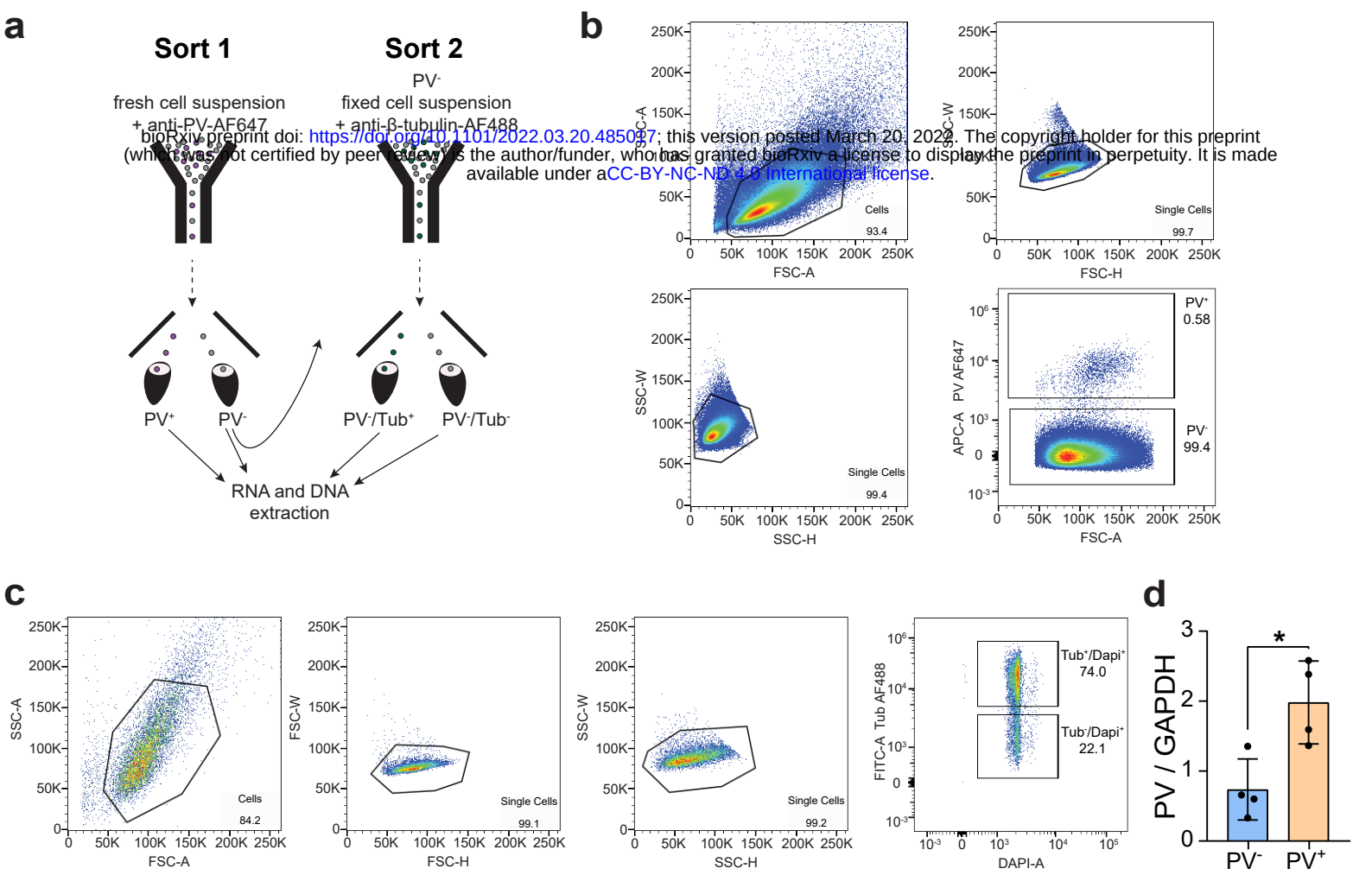
Extended Data Fig. 2: Retrotransposition of an engineered mouse L1 delivered *in utero*. a, Schematic representation of L1-EGFP reporter *in utero* electroporation (IUE). A coronal view of electrode positioning is shown at left. Embryos were co-injected with pmCherry (where a CMV promoter controls mCherry expression) and a second plasmid, carrying a mouse L1[39] tagged with an EGFP indicator cassette (pUBC-L1SM-UBC-EGFP), into the left hemisphere. As a negative control, embryos were co-injected with pmCherry and a disabled L1 reporter plasmid (pMut2-UBC-L1SM-UBC-EGFP) into the right hemisphere. The red inset, shown at right, displays a coronal section of an electroporated mouse brain with pmCherry visible in the targeted hippocampal region. IUE was performed on embryonic day (E)14.5. Embryos were born and then sacrificed at postnatal day (P)10. Note: UBC-L1SM-UBC-EGFP consists of a heterologous UBC promoter driving expression of a synthetic mouse L1 T_F (L1SM) containing a native L1 monomeric 5′UTR promoter, codon-optimized ORF1 and ORF2 sequences, the 5′ part of the L1 3′UTR, and an intron-interrupted EGFP indicator cassette with its own UBC promoter and polyadenylation signal (pA). In this system, a cell becomes EGFP positive only when the L1-EGFP mRNA is transcribed, spliced, reverse transcribed and integrated into the genome, allowing a functional EGFP to be expressed from its UBC promoter. Mut2-UBC-L1SM-UBC-EGFP is the same as UBC-L1SM-UBC-EGFP, except it carries mutations known to disable ORF2p reverse transcriptase and endonuclease activities[39]. **b,** Example maximum intensity projection confocal image of a hippocampus section from an embryo electroporated with UBC-L1SM-UBC-EGFP. A PV^+ (magenta), $Cherry^+$ (red), $NeuN^+$ (blue), $EGFP^+$ cell is indicated with a yellow arrowhead. **c,** No $EGFP^+$ cells were found for the retrotransposition-incompetent Mut2-UBC-L1SM-UBC-EGFP plasmid. An empty arrowhead points to a PV^+, $NeuN^+$, $Cherry^-$, $EGFP^-$ cell.



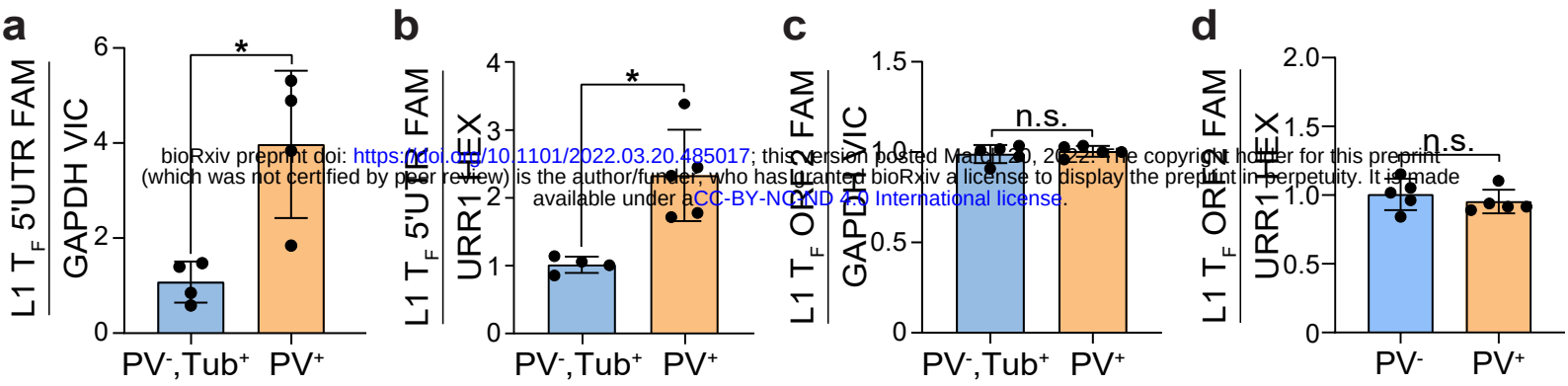
Extended Data Fig. 3: Single-molecule L1 T_F RNA FISH experimental design. a, L1 T_F RNAscope probe A consisted of 20 "ZZ" oligo pairs[92] targeting the L1 T_{FI} 5′UTR monomeric and non-monomeric region (consensus positions 827 to 1688). **b,** L1 T_F probe B was composed of 17 ZZ oligo pairs and targeted the L1 T_{FI} 5′UTR monomeric region (positions 142 to 1423). **c,** Imaris analysis performed on Z-stack images of L1 T_F and PV RNA FISH coronal hippocampus sections immunostained for Tub. Imaris workflow: 1: neuron identification, based on cytoplasmic Tub (red) fluorescence, and cell volume drawing, 2: nucleus definition by DAPI (blue) staining followed by drawing of nuclear volume, 3: delineation of cell and nucleus surfaces, 4: cell surface masking to eliminate voxels outside the cell, 5: nucleus surface masking to exclude voxels inside the nucleus, and 6: L1 T_F mRNA (green) fluorescence signal within the defined cytoplasmic volume. **d,** Maximum intensity projection confocal image of a hippocampal section showing L1 T_F probe A (green) and PV (magenta) RNA FISH, and Tub (red) immunohistochemistry, in a selected PV^- neuron. Dashed lines demark nuclear and cellular boundaries defined for PV and L1 mRNA quantification. **e,** Confocal images of N2A cells transfected with either mouse L1 construct (pL1SM-Cherry) or control (pmCherry) showing specificity of L1 T_F RNA FISH signal in cells expressing the L1 construct. Scale bar: 10µm. **f,** Schematic showing L1 T_F sequences assayed by RNA FISH and TaqMan qPCR.



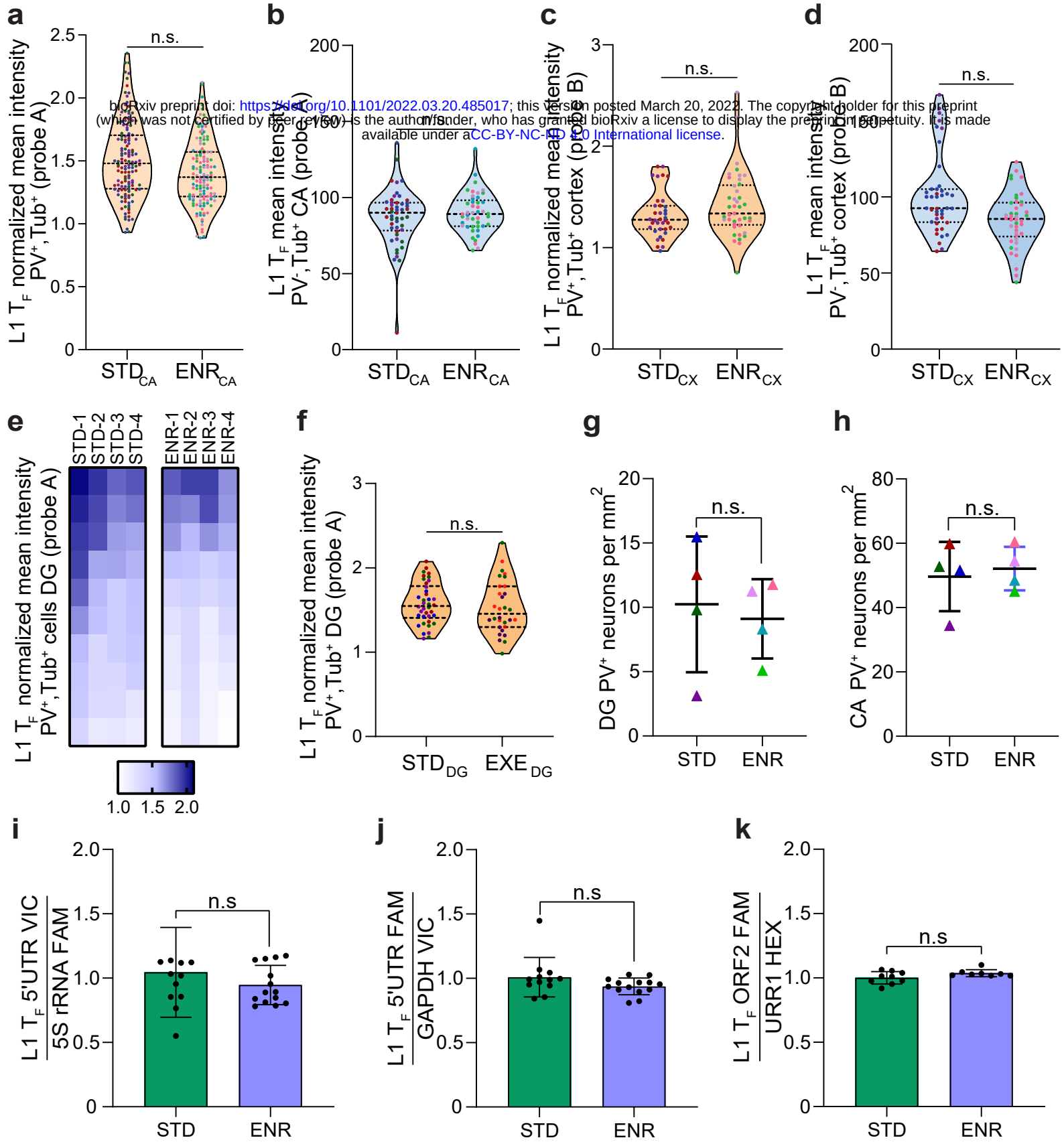
Extended Data Fig. 4: Additional L1 T_F RNA FISH data for hippocampus and cortex. a, Mean PV RNA FISH intensity in CA Tub^+ neurons, as a function of L1 T_F (probe A) signal. Spearman r=0.82, P<0.001 at α=0.05, n(cells)=210, N(mice)=4. Cells from different mice are color coded. **b,** Mean L1 T_F RNA FISH (probe A) intensity in CA PV^-/Tub^+ (blue plot) and PV^+/Tub^+ neurons (orange plot). *P=0.0251, two-tailed t test comparing animal means, PV^-/Tub^+ n=47, PV^+/Tub^+ n=60. **c,** Mean RNA FISH intensity of L1 T_F probe B and PV in DG Tub^+ neurons. n=55, N=4. **d,** As per (c), except showing results for CA. n=102, N=4. **e,** Mean L1 T_F RNA FISH (probe B) intensity in hippocampal PV^-/Tub^+ neurons and PV^+/Tub^+ neurons. ***P=0.0008, PV^-/Tub^+ n=76, PV^+/Tub^+ n=81, N=4. **f,** As per (c), except for cortex. n=87, N=4. **g,** As per (e), except for cortex. *P=0.0193, PV^-/Tub^+ n=52, PV^+/Tub^+ n=36, N=4. Note: PV^+/Tub^+ values were normalized to PV^-/Tub^+ mean values.



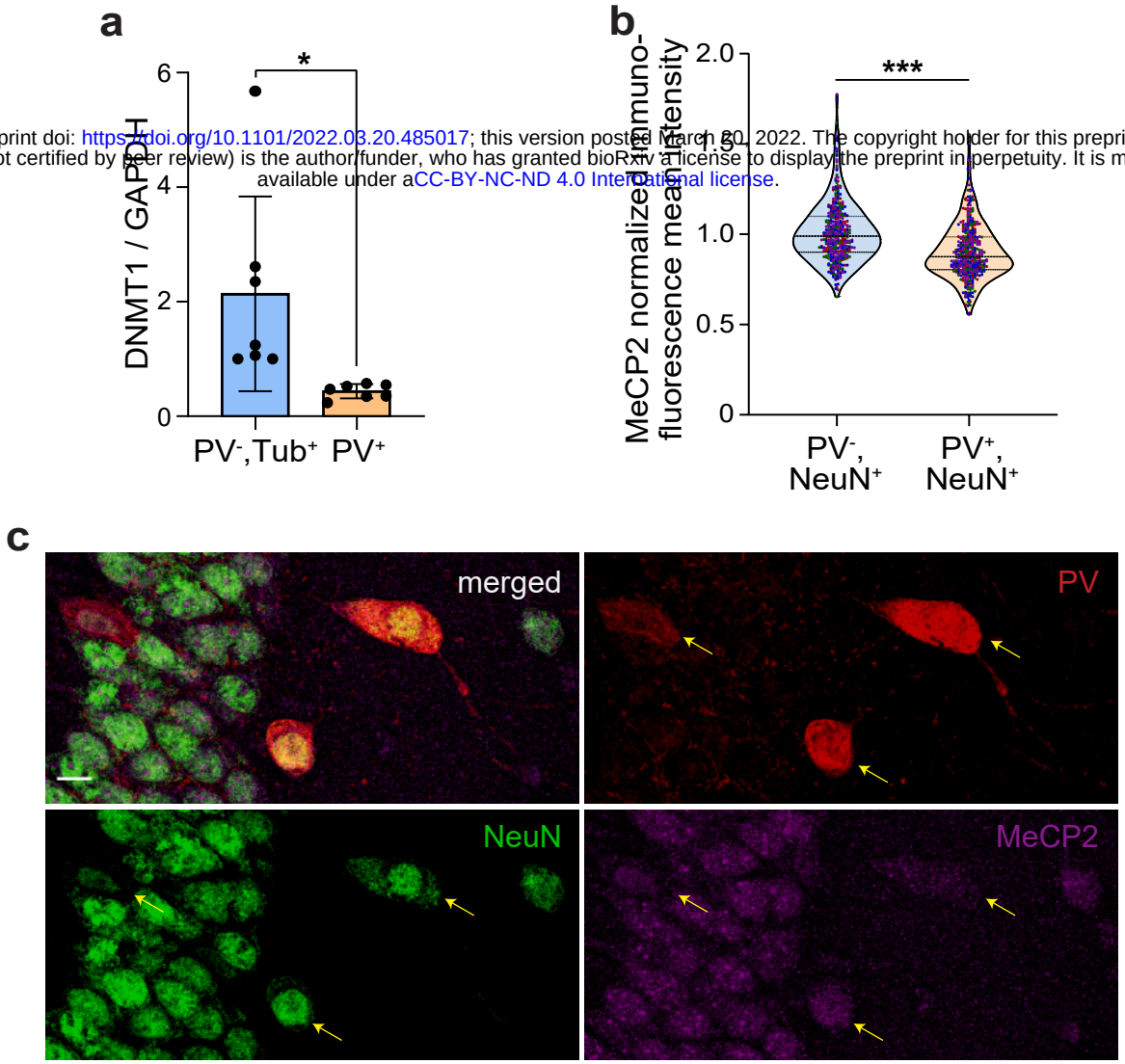
Extended Data Fig. 5: PV$^+$ neuron isolation. a, Schematic of PV$^+$ and PV$^-$/Tub$^+$ neuron isolation from pooled neonate (P0) litter hippocampal tissue. An anti-PV conjugated antibody (AF647) was used to label and isolate PV$^+$ cells. Freshly sorted PV$^-$ cells were subsequently labeled with an anti-Tub conjugated antibody (AF488) and sorted again to isolate PV$^-$/Tub$^+$ neurons. **b,** Gating strategy and purity of fluorescence activated cell sorting (FACS) for PV$^+$ cells. **c,** As for (b), except showing PV$^-$/Tub$^+$ neurons. **d,** Quality control qPCR of relative PV mRNA expression in sorted cells. PV mRNA enrichment in the PV$^+$ population was observed, as expected. Data are relative to GAPDH and presented as mean ± SD. *P=0.017, two-tailed t test, N(litters)=3.



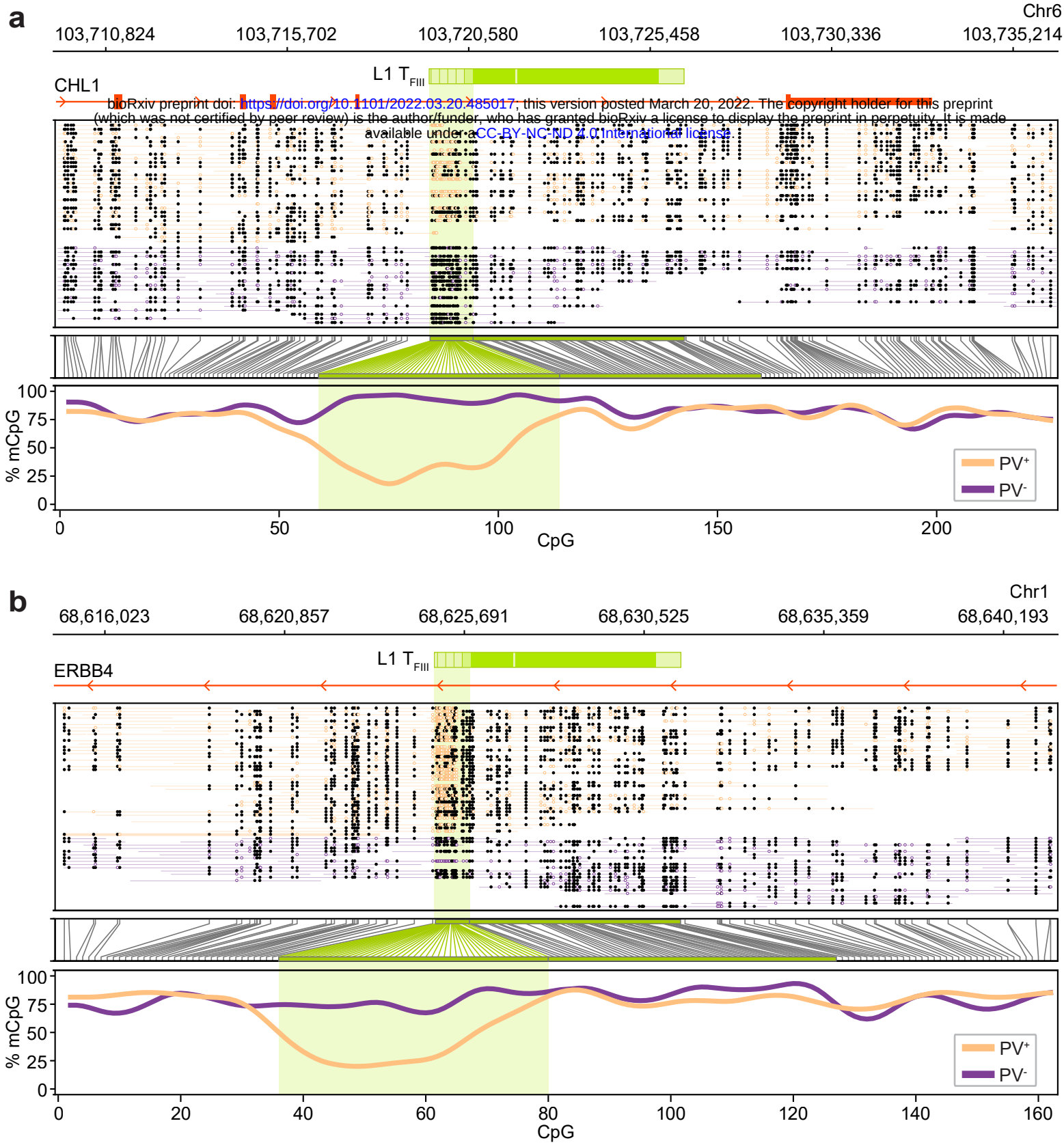
Extended Data Fig. 6: Elevated L1 transcription in PV$^+$ neurons. a, Multiplexed TaqMan qPCR measuring mRNA abundance of the L1 T_F non-monomeric 5′UTR sequence (FAM channel) relative to GAPDH (VIC channel) in sorted PV$^-$/Tub$^+$ and PV$^+$ neurons. Cells were sorted from pooled neonate (P0) litter hippocampi. *P=0.029, N=4 litters. **b,** As for (a), except relative to URR1 repetitive DNA (HEX channel). **P=0.011, N=4-5 litters. **c,** As for (a), except measuring L1 T_F ORF2 (FAM channel) relative to GAPDH (VIC channel). P=0.582, N=5 litters. **d,** As for (a), except for L1 T_F ORF2 (FAM channel) relative to URR1 repetitive DNA (HEX channel) in PV$^-$ cells and PV$^+$ neurons. P=0.444, N=5 litters. Note: Data are represented as mean ± SD with significance values determined via two tailed t tests.



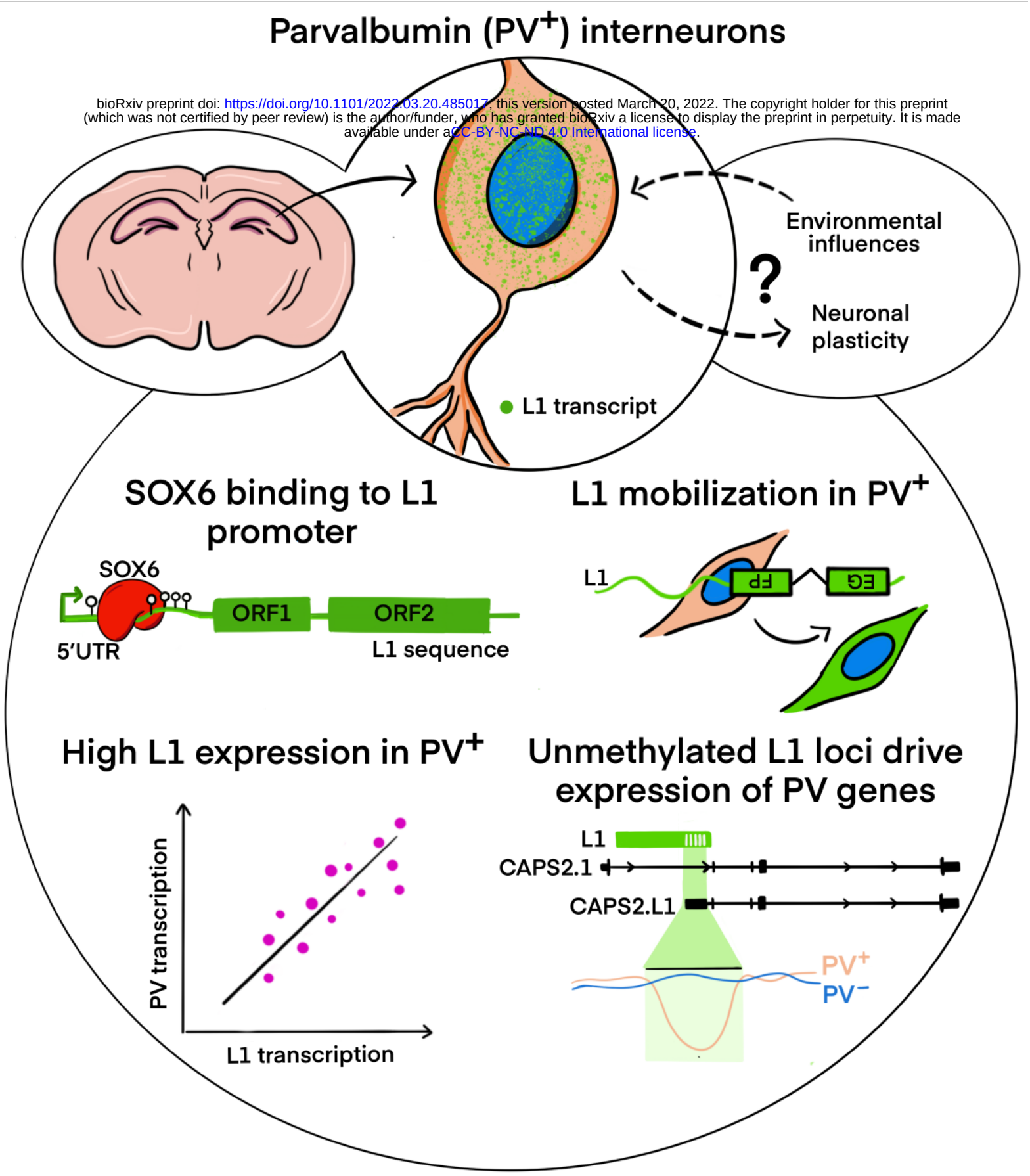
Extended Data Fig. 7: Environmental enrichment does not impact PV⁺ neuron count, or L1 mRNA abundance in PV⁻ neurons. a, Mean L1 T_F RNA FISH (probe A) intensity in PV⁺/Tub⁺ neurons from STD and ENR animal CA tissue. *P*=0.189, STD *n*(cells)=140, ENR *n*=138, *N*(mice)=4. Values were normalized to PV⁻/Tub⁺ mean values. Cells from each mouse are color coded. **b,** Mean L1 T_F RNA FISH intensity (probe A) in PV⁻/Tub⁺ CA neurons. *P*=0.294, STD *n*=49, ENR *n*=45, *N*=4. **c,** As per (a), except showing cortex data obtained with L1 T_F probe B. *P*=0.648, STD *n*=36, ENR *n*=41, *N*=3. **d,** As for (b), except displaying cortex data obtained with L1 T_F probe B. *P*=0.276, STD *n*=42, ENR *n*=42. **e,** Heatmaps comparing normalized mean L1 T_F RNA FISH intensity (probe A) in DG neurons from STD and ENR mice. Each column represents an individual animal, while each cell represents a neuron and is colored based on L1 T_F mRNA abundance relative to the median value of the STD group. *n*=10, *N*=4. **f,** Mean L1 T_F RNA FISH intensity (probe A) in PV⁺/Tub⁺ neurons of DG from mice housed in STD conditions or with access to exercise (EXE). *P*=0.71, STD *n*=38, EXE *n*=30, *N*=4. **g,** Stereological estimation of PV⁺ neuron number in DG of STD and ENR mice. *P*=0.726, *N*=4. **h,** As for (g), except displaying CA data. *P*=0.714, *N*=4. **i,** TaqMan qPCR measuring abundance of the L1 T_F mRNA monomeric 5′UTR (VIC channel) relative to 5S rRNA (FAM channel) in bulk hippocampus samples from STD and ENR mice. *P*=0.38, STD *N*=12, ENR *N*=14. **j,** As for (i), except targeting the L1 T_F non-monomeric 5′UTR (FAM channel) relative to GAPDH (VIC channel). *P*=0.15. **k,** As for (i), except measuring L1 T_F ORF2 (FAM channel) relative to URR1 (HEX channel). *P*=0.07, STD *N*=9, ENR *N*=8. Note: values in (g-k) are represented as mean ± SD. Significance values for all but (e) were obtained via two-tailed t test comparing means of animals or groups, where appropriate.



Extended Data Fig. 8: Relaxation of epigenetic repression in PV$^+$ neurons. a, DNMT1 mRNA abundance measured by qPCR in PV$^-$/Tub$^+$ and PV$^+$ neurons, relative to GAPDH. *P=0.038, two-tailed t test, N=7 litters. **b,** MeCP2 protein expression in PV$^-$/NeuN$^+$ (blue plot) and PV$^+$/NeuN$^+$ (orange plot) neurons. MeCP2 immunofluorescence mean intensities were obtained from coronal hippocampus sections stained for MeCP2, PV and NeuN, and normalized to the PV$^-$/NeuN$^+$ population mean. ***P=0.0007, PV$^-$/NeuN$^+$ n(cells)=414, PV$^+$/NeuN$^+$ n=414, N(mice)=4. Cells from each mouse are color coded. **c,** Representative immunostaining image of a coronal hippocampus section showing colocalization of MeCP2 (magenta) with PV (red) and the pan-neuronal marker NeuN (green). Yellow arrows indicate PV$^+$ neurons on single channel images. Scale bar: 10µm.



Extended Data Fig. 9: Demethylated L1s in PV$^+$ neuron genes, as identified by ONT sequencing. a, Methylation profile of a full-length L1 T_{FIII} element with intact ORFs, intronic to CHL1. The first panel shows the L1 orientated in sense to the last intron of CHL1. The second panel displays aligned ONT reads, with unmethylated CpGs colored in orange (PV$^+$) and purple (PV$^-$), and methylated CpGs colored black. The third panel indicates the relationship between CpG positions in genome space and CpG space, including those corresponding to the L1 T_{FIII} 5′UTR (shaded light green). The fourth panel indicates the fraction of methylated CpGs for each cell type across CpG space. **b,** As for (a), except displaying an L1 T_{FIII} antisense and intronic to ERBB4.



Extended Data Fig. 10: Model of L1 activation in PV⁺ neurons by SOX6.