

Supplementary Materials for
**Evolution of pallium, hippocampus, and cortical cell types revealed by
single-cell transcriptomics in reptiles**

Maria Antonietta Tosches,*† Tracy M. Yamawaki,* Robert K. Naumann, Ariel A. Jacobi,
Georgi Tushev, Gilles Laurent†

*These authors contributed equally to this work.

†Corresponding author. Email: gilles.laurent@brain.mpg.de (G.L.); maria.tosches@brain.mpg.de (M.A.T.)

Published 3 May 2018 on *Science* First Release
DOI: 10.1126/science.aar4237

This PDF file includes:

- Materials and Methods
- Figs. S1 to S17
- Tables S1 and S2
- References

Materials and Methods

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Drop-seq experiments were performed using 6 female turtles from the species *Trachemys scripta elegans* weighing 450-600 grams and 8 male lizards from the species *Pogona vitticeps* weighing 200-400 grams. An additional animal of each species was used for 3' end sequencing of bulk tissue samples. *In situ* hybridizations (ISH) were performed using turtles and lizards of both genders in a similar size range. Animals were obtained from external breeders, selected for sex, size, weight, and health. All experiments were performed in accordance with Hessian, German and EU laws on animal experimentation. Animals were sacrificed according to § 4 (3) Tierschutzgesetz (TierSchG, German animal welfare law) and § 2 Tierschutz-Versuchstierverordnung (TierSchVersV).

METHOD DETAILS

Sequencing Experiments

Single-cell isolation

Animals were deeply anesthetized prior to decapitation (turtle: Ketamine 10-15 mg/kg and Demethomidine 0.05-0.1 mg/kg, administered by intramuscular injection; lizard: Tiletamine 20-40 mg/kg and zolazepam 20-40 mg/kg (Zoletil 100) administered by intramuscular injection before placement in a whole body chamber with Isoflurane 5% vol.). To remove blood, brains were perfused with Ringer solution through the carotid arteries. (Turtle Ringer solution: 96mM NaCl, 2.6mM KCl, 31mM NaHCO₃, 2mM MgCl₂, 4mM CaCl₂ and 20mM D-glucose; Lizard Ringer solution: 126 mM NaCl, 3 mM KCl, 24mM NaHCO₃, 0.72mM NaH₂PO₄, 1.2mM MgSO₄, 1.8 mM CaCl₂ and 10mM D-glucose). While immersed in cold Ringer solution oxygenated with carbogen (5% CO₂, 95% O₂), brains were dissected out of the skull, and the dura mater and arachnoid were removed. Next, brains were sectioned coronally (turtle: 400µm, lizard: 300µm) in oxygenated sucrose Ringer solution (2.6mM KCl, 31mM NaHCO₃, 10mM MgCl₂, 0.5mM CaCl₂, 20mM D-glucose, 200mM sucrose) using a vibratome (Leica VT 1200 S). Brain regions (cortex, DVR, etc.) were dissected from slices under a stereo microscope (see fig. S7 for dissections).

Brain tissue was dissociated using a modified protocol from the Worthington Papain Dissociation System. Dissected turtle slices were cut into small chunks that were then digested with 40 units/mL Papain, 200 units/mL DNaseI, 2mM L-cysteine, 1mM EDTA in EBSS (Worthington Biochemical Corporation) at room temperature. After 20 minutes tissue pieces were passed through a fire-polished glass pipette ten times. The process was repeated three times in five-minute intervals using pipettes with progressively narrower tip diameters. Dissected lizard slices were cut into small chunks that were digested with 20 units/mL Papain, 250 µg/ml Liberase (Roche), 100 units/mL DNaseI, 1mM L-

cysteine, 1mM EDTA in EBSS at room temperature. After 40 minutes, tissue pieces were dissociated with a series of fire-polished pipettes of decreasing tip diameter until a uniform suspension was produced. Cell suspensions were filtered through a 100 μ m cell strainer and pelleted in a benchtop centrifuge at 300 x g for 5 minutes. Pellets were resuspended in 1mg/mL ovomucoid inhibitor and 1mg/mL albumin in EBSS and strained through a 40 μ m filter.

To remove cell debris, 1mL of cell suspension was layered on a gradient composed of 2mL layers of 11, 14, 17 and 20% Percoll Plus (GE Healthcare) pH 7.4 in EBSS and centrifuged for 4 min at 430xg with no brake. The bottom 2 mL of the gradient were collected, mixed with 2 mL EBSS and cells pelleted at 550xg. Cells were resuspended in 100-200 μ L of EBSS with 0.2% BSA and 0.3% glucose. After counting, cells were diluted with 1xPBS to a final concentration of 100 cells/ μ L for Drop-seq.

Drop-seq

Drop-seq was performed according to the Drop-seq Laboratory Protocol v3.1 (<http://mccarrolllab.com/dropseq/>), an optimized version of the Drop-seq technique described in (8) and developed by Evan Macosko, Melissa Goldman and Steve McCarroll. Animals used, dissections from these animals and Drop-seq libraries are summarized in Tables S1 and S2. Cells and beads were run through the microfluidics devices to achieve a final droplet occupancy of 2.3% for cells and 4.6% for barcoded beads (ChemGenes). Sample 15.1-3 was run at a concentration so that 1.7% of droplets were occupied by cells. Beads collected from droplets were reverse-transcribed in bulk, then separated into batches of 4,300-6,000 beads each, corresponding to approximately 100-115 cells, for PCR amplification for 15 cycles (turtle samples: 16, 17, lizard samples: 19-25), 16 cycles (turtle samples: 8, 11, 12, 15, lizard samples: 8, 9, 22) or 17 cycles (lizard sample 22). Batches of 3 to 10 PCR reactions from one individual were pooled for fragmentation and indexed with a Nextera_N7xx primer (Table S2). Fragmentation products were pooled for a final library that included an estimated 5500-7000 cells. Libraries were sequenced on an Illumina NextSeq500 using the 150 cycle high output kit with Read1 set at 20 bases and Read2 at 138 bases. Additional reads from samples 8.1, 8.2 and 9.1 were generated on two lanes of a HiSeq2000 in rapid mode with Read1 set at 25 bases.

Histology Experiments

Tissue Preparation

Animals were anesthetized and decapitated as described above. Brains for *in situ* hybridizations were prepared using RNase-free solutions prepared with DEPC-treated H₂O. Brains were fixed by perfusion through the carotid arteries with 50-100mL 1x PBS followed by 50-100mL cold 4% PFA in PBS (pH 7.4). The entire brain was removed from the skull and dura mater and arachnoid were removed. The tissue was then post-fixed overnight in 4% PFA in PBS at 4°C. For cryoprotection, post-fixed brains were

then transferred to a 30% sucrose 1xPBS solution for one to two days at 4°C before sectioning. 50 µm frontal or sagittal sections were obtained using a cryotome (Hyrax S50, Zeiss). All stainings were performed on free-floating sections.

Immunohistochemistry and in situ hybridizations (ISHs)

Antibody stainings were performed on free-floating sections. Sections were first incubated overnight in blocking solution (PBS with 2% BSA and 0,1% Triton) and later incubated for 24-72 h in primary antibodies: mouse anti-Satb2 (Abcam, ab51502) 1:200, rat anti-SST (Millipore, MAB354) 1:1000, mouse anti-NeuN (Millipore, MAB377) 1:200, rabbit anti-Gfap (Z0334, Dako) 1:200, rabbit anti-Sox9 (Merck Millipore AB5535) 1:500, rabbit anti-HuB (ELAVL2, Sigma H1538) 1:500, rabbit anti-calbindin D-28k (Swant, CB38a) 1:500 in blocking solution. Alexa Fluor conjugated secondary antibodies (Life Technologies) were used at a 1:500 dilution in blocking solution.

To prepare ISH probes, genes of interest (0.5-1.5kb fragments from coding regions) were PCR-amplified from *T. scripta* or *P. vitticeps* brain cDNA libraries and cloned into the PCRII vector (TA cloning kit, Invitrogen). All clones were verified by Sanger sequencing. DIG- or FITC-labeled RNA probes were *in vitro* transcribed from linearized plasmids and purified with the RNeasy kit (Qiagen).

The protocol for chromogenic ISHs was modified from (38). Sections were postfixed overnight in 4% PFA/PBS at 4°C; after the proteinase K permeabilization step, sections were acetylated (5 min in 1% TEA/PBT, 5 min in 3µl/ml acetic anhydride/PBT) and postfixed (4%PFA/PBT, 20min) before pre-hybridization and hybridization (hybridization buffer: 50% formamide, 5x SSC, 50µg/ml heparin, 250µg/ml yeast tRNA, 5x Denhardt's solution, 0.2% Tween-20, 500µg/ml salmon sperm DNA, 5% dextran sulfate). Hybridization and post-hybridization washes (2x 40min washes in 2x SSC, 0.1% Tween-20, 50% formamide, followed by 2x 40min washes in 0.2x SSC, 0.1% Tween-20) were carried at hybridization temperature (55°C for turtle sections, 60°C for lizard sections). Sections were incubated overnight with an AP-conjugated anti-DIG antibody (Roche, 1:5000 dilution), then washed in TNT (0.1M Tris-HCl pH7.5; 0.15M NaCl; 0.05% Tween-20), and finally incubated at RT with NBT 4.5 µl/ml, BCIP 3.5 µl/ml (Roche) in alkaline phosphatase buffer (with 5% polyvinylalkohol) for signal development.

For dual colorimetric ISHs, a DIG- and a FITC-labeled probe were used for hybridization, and signals were detected after sequential incubations with anti-DIG and anti-FITC AP-conjugated antibodies and stainings (38). Signals were developed with NBT/BCIP and INT/BCIP (Roche), following manufacturer's instructions.

Fluorescent ISHs were done as above, except the following steps. To quench endogenous alkaline phosphatases, sections were incubated for 10 min in PBS + 1% H₂O₂ before the permeabilization step, and for 10 min with Dako Dual Endogenous Enzyme Block (Agilent) in PBT after the post-hybridization washes. Sections were then incubated with an anti-DIG POD antibody (Roche, 1:500), and signal was developed with

the TSA Plus Cyanine 5 and Fluorescein System kit (Pelkin Elmer) following manufacturer's instructions. In double fluorescent ISHs, signals from DIG- and FITC-labeled probes were detected sequentially using anti-DIG and anti-FITC POD-conjugated antibodies, and the same TSA Plus Cyanine 5 and Fluorescein System kit. Fluorescent ISHs were combined with immunohistochemistry by adding primary antibodies at the anti-DIG POD step, followed by incubations with secondary antibodies after the detection of mRNA signal.

After mounting, chromogenic ISHs were imaged at a 20x magnification using an automatic digital slide scanner (Pannoramic MIDI II, 3DHISTECH). Images were adjusted in Adobe Photoshop as follows. Whenever necessary, background surrounding the section of interest was removed (e.g. to remove a neighbouring section, or bubbles forming after mounting). Color balance, white balance, brightness and contrast were adjusted evenly across the images.

Fluorescent signals from ISHs and immunohistochemistry were imaged with a Zeiss LSM 710 confocal microscope. Images were adjusted in Adobe Photoshop or Fiji. Brightness, contrast and gamma corrections were adjusted evenly across the images.

QUANTIFICATION AND STATISTICAL ANALYSIS

Refinement of *C. picta* and *P. vitticeps* 3' gene annotations

For the analysis of our Drop-seq data from *Trachemys scripta elegans*, we used the genome of the closely related *Chrysemys picta bellii* to align reads. Previously, over 96% of a draft *Trachemys* genome was found to align to the *C. p. bellii* genome (39). In line with this, more than 85% of reads from our in-house RNA sequencing experiments of bulk *T. scripta* neuronal tissue aligned to the *C. p. bellii* genome. Reads from the lizard *P. vitticeps* were aligned to the *P. vitticeps* genome, pvi1.1 assembly (40).

Since Drop-seq libraries are enriched for mRNA 3'-ends, we used targeted sequencing of mRNA 3'-ends from bulk cortical samples to improve the annotation of 3' UTRs of the turtle and lizard reference genomes (fig. S2). The MACE (Massive Analysis of cDNA Ends) technique (41) (www.genxpro.net, GenXPro, Frankfurt am Main, Germany) was chosen to prepare 3'-end libraries from RNA extracted from *T. scripta* or *P. vitticeps* microdissected cortices. MACE libraries were sequenced on a NextSeq500 platform with 68 nt long reads. Reads were filtered for PCR duplicates by unique molecule identifiers (UMIs), and aligned to the genome of interest (*C. picta* or *P. vitticeps*) using STAR Aligner (42). Parameters "outFilterScoreMinOverLread" and "outFilterMatchNminOverLread" were used to allow poly(A) tail soft clip at reads end. Only uniquely aligned reads were considered for downstream analysis. Internal priming events were avoided by filtering alignments hitting poly(A) rich genomic regions. Such regions were identified by aligning 10 As to the genome of interest (43).

Poly(A) supported sites (PASS) were identified using a HTSLib driven tool called PASSFinder (<https://github.molgen.mpg.de/MPIBR-Bioinformatics/PASSFinder>). This

tool clusters reads based on Poly(A) tail using an algorithm developed by Jim Kent at UCSC (<https://github.com/jstjohn/KentLib/blob/master/lib/dnautil.c>). Alignments were piled up and 3'-most non-A base was maximised to precisely pinpoint poly(A) site position. PAS sites were then associated with upstream genes up to 20kB and the species annotations were extended accordingly (*C. picta* genome, *Chrysemys_picta_bellii-3.0.3* assembly; *P. vitticeps* genome, *pvi1.1* assembly).

After a first search, we found that there was no *SST* gene annotated in *Pogona*. A tblastn search for somatostatin (*SST*) in the *P. vitticeps* genome revealed the presence of a somatostatin sequence with high sequence homology to the *Anolis* *SST* gene. This region was updated in our annotation and correspondingly extended according to the MACE analysis described above.

Alignment and Demultiplexing Reads

Preprocessing of raw reads and generation of digital expression tables (DGEs) were performed using Drop-seq tools (version 1.12) described in (44). In short, cell barcodes (CIDs) and unique molecular identifiers (UMIs) were extracted for each read. Reads with at least one base in either the CID or UMI that had a quality score below 10 were discarded. CID and UMI assignments were also corrected for truncated CID errors particular to the split and pool synthesis method. The polyA sequence and any 5' adapter sequence were then trimmed from the mRNA read. Trimmed reads were aligned to either the *C. picta* genome (*Chrysemys_picta_bellii-3.0.3* assembly) or *P. vitticeps* genome (*pvi1.1* assembly) with STAR Aligner (version 2.5.0b). Finally, molecule counts per gene per cell were assigned based on our updated genome annotations (see above) to generate a DGE matrix.

Quality Control and cell filtering

Our analysis included CIDs with at least 800 genes detected for turtle and at least 500 genes for lizard. The initial analysis revealed several clusters that contained cells likely to be damaged or dying, indicated by their high fraction of mitochondrial gene transcripts detected, high fraction of intergenic reads and correspondingly lower percentage of reads mapping to coding or UTR regions (figure S3A-E,J). To identify similar poor quality cells and remove them from analysis, we selected a random subset of 100 turtle CIDs or 50 lizard CIDs from these clusters and an additional 100 turtle CIDs or 50 lizard CIDs from the remaining clusters to use as a training set for a support vector machine (SVM) classifier (R package e1071). The SVM classifier was trained on seven statistics: number of UMIs detected, number of genes detected, fraction of reads aligning to intergenic, coding, UTR, or intronic regions, and the fraction of transcripts from mitochondrial genes (Supplemental Figure S3F,K). CIDs identified as poor quality from our classifier were removed from analysis along with any remaining CIDs that met at least one of the following criteria: at least 7% mitochondrial transcripts, at least 60%

intergenic reads, at least 30% intronic reads, less than 20% coding reads, and less than 20% UTR reads.

Our secondary analysis of neuronal clusters was performed using a higher gene detection threshold since in general neurons have a larger transcriptome than other cell types in the brain (6). We included neuronal CIDs with at least 2,000 genes detected in turtles and 1,500 genes detected in lizards.

Sequencing Statistics

Turtle cells (CIDs >800 genes detected) were sequenced to a median read depth of 15,336.5 reads/CID, corresponding to a median 2,658 UMIs/CID and median 1,536 genes/CID. Across sequencing libraries, median read depth ranged from 9,001 reads/CID in library d4 to 18,614 reads/CID in library d3. This corresponds to a range of median transcripts detected from 2,099 UMIs/CID in library d2d4 to 2,960 UMIs/CID in library d5 and a range of median genes detected from 1,309 genes/CID in library d2d4 to 1,719.5 genes/CID in library d5 (figure S2E). The narrower range of transcripts and genes detected indicates we are approaching sequencing saturation, and adding additional reads will not significantly increase the amount of transcripts or genes detected. Library d2d4 contains reads from libraries d2 and d4 while library d4 was sequenced solely in library d4 (see Table S2). Lizard cells (CIDs >500 genes detected) were sequenced to a median read depth of 19,257 reads/CID, corresponding to a median 1,594 UMIs/CID and median 994 genes/CID. Although for the lizard libraries, the mean read depth ranged from 7,389 reads/CID in library d4 to 21,540 reads/CID in library d8, the range of median transcripts detected was again much narrower from 1,401.5 UMIs/CID in library d4 to 1,638 UMIs/CID in library d7, corresponding to a median gene detection ranging from 966 genes in library d4 to 1,023.5 genes/CID in library d3 (figure S2F). Note that statistics for lizard cells from library d4 include reads from library d2.

After quality control (QC) filtering of poor quality cells, turtle cells had a median of 12,745 reads/CID, 2,731.5 UMIs/CID and 1,584 genes/CID. Turtle neurons analyzed had a median of 33,751 reads/CID, 8,545 UMIs/CID and 3,621 genes/CID. Lizard cells had a median of 21,937.5 reads/CID, 1,918 UMIs/CID and 1,179.5 genes/CID. Lizard neurons analyzed had a median of 49,696 reads/CID, 4,284 UMIs/CID and 2,204 genes/CID. (Figure S5A-C).

Normalization and Batch Correction

Gene expression was normalized by total number of transcripts detected, then multiplied by a scaling factor of 10,000. The data was then log transformed (ln) after the addition of a pseudo-count of 1.

In the first round of analysis, we noticed that minor batch effects could be attributed to the animal of origin. In tSNE space, every point cloud contained intermixed cells from

different animals, with the exception of the ependymoglia cells cluster. Within this cluster, cells were segregating by animal of origin (tsEG, fig. S3G). We then adjusted for batch effects by animal of origin using ComBat (sva R package) as described in (44,45). ComBat was run on the normalized data with the parameter “par.prior=T” (default parametric adjustment mode). ComBat-corrected data were then scaled and centered by row, and used for PCA. Clustering with ComBat-corrected data produced clusters with cells from different animals intermixed in tSNE plots (cfr. tsEG cells in fig. S3G and fig. S3H; see also fig. S3L and fig. S8E). Note that different brain regions were dissected from each animal (fig. S7A); for this reason some neuronal clusters do not contain cells from all animals (for example, pallial thickening clusters; cfr. also Table S1 and S2).

Principal Component Analysis (PCA)

PCA, tSNE and SNN cluster analysis were performed in R (46) using the Seurat package (v1.4.0). Highly expressed and highly variable genes were identified using MeanVarPlot() function as described in (47). Briefly, genes were grouped by mean expression into twenty bins and the dispersion (variance/mean) was then z-normalized in each bin. The variable genes were identified using the following thresholds: turtle cells: xlow-0.25, xhigh-3, ylow-0.5; turtle neurons: xlow-0.1 xhigh-3 ylow-0.3; lizard cells: xlow-0.25, xhigh-3, ylow-0.5; lizard neurons: xlow-0.25, xhigh-3, ylow-0.5, with the x and y axes corresponding to mean expression and normalized dispersion respectively.

PCA was performed using the PCAFast() function in the Seurat package using high variable genes (turtle cells: 1101 genes, turtle neurons: 1979 genes, lizard cells: 971 genes, lizard neurons: 892 genes). Statistically significant principal components were determined using a permutation method described in (44). Briefly, expression values for variable genes were shuffled across cells 1,000 times, and PCA (big.PCA() function, bigpca R package) was performed on these shuffled matrices to determine the maximum eigenvalue across replicates. Principal components from non-shuffled data were considered significant if their eigenvalue determined using big.PCA() was greater than this shuffled maximum.

tSNE embeddings (48) of significant PCs (turtle cells: 32 PCs, turtle neurons: 38 PCs, lizard cells: 22 PCs, lizard neurons: 18 PCs) were calculated using the runTSNE() function in Seurat with the default perplexity setting of 30.

Shared Nearest Neighbor (SNN) Cluster Analysis

Clusters of cells were identified with the FindClusters() function of the Seurat R package, which applies a graph-clustering algorithm on significant PCs (as described in (49)). The following parameters were used for clustering: turtle cells: resolution 0.3, k parameter 30; lizard cells: resolution 0.6, k parameter 30; lizard neurons: resolution 3, k parameter 10; method: original Louvain algorithm.

For the turtle cells dataset, we observed that this method was unable to separate the tsMur cluster from a neighboring cluster made of cell doublets (grey in Fig. 1A). Doublets were recognized from the coexpression of mural and microglial markers, and from the absence of cluster-specific marker genes. We found that the Infomap algorithm (50), when the k -NN graph was built with $k=30$, was able to separate tsMur and doublets. Hence in the tSNE in Fig. 1A and in analysis in fig. S4D-F, the tsMur cluster and the neighboring cluster of doublets have been separated on the basis of cluster assignments found with Infomap.

For the turtle neurons dataset, we initially observed that the Louvain algorithm was unable to identify small and large clusters equally well with a global k parameter. For this reason, glutamatergic and GABAergic cells were clustered separately, using the following settings: turtle glutamatergic neurons: resolution 10, k parameter 25; turtle GABAergic neurons: resolution 5, k parameter 15. For the turtle and lizard neurons datasets, data were initially over-clustered, and final clusters were determined by merging similar clusters on the basis of differential expression, similar to the approach described in (44). For merging, we iteratively calculated clusters similarities based on average expression of variable genes and built a dendrogram from these similarities (BuildClusterTree() function in Seurat). Clusters were only merged if they shared a terminal node on this dendrogram and had fewer than 40 differentially expressed genes (turtle neurons) or 16 differentially expressed genes (lizard neurons). For the turtle neurons dataset, a gene was considered differentially expressed between two clusters if it was detected in more than 20% of cells in at least one of the two clusters, if the average expression in the two clusters differed by at least $\ln(1.5)$ and if the differential expression was significant, *i.e.* if it had a p -value <0.05 as determined using a zero-inflated regression model implemented in the MAST (v1.2.2) R package (see below). For the lizard dataset, with fewer cells and fewer genes detected on average, we used more stringent criteria: a gene was considered differentially expressed between two clusters if it was detected in more than 40% of neurons in at least one of the two clusters, and if it had a minimal average expression difference of $\ln(2)$ between the two clusters compared.

Identification of Differentially Expressed Genes

For the calculation of differentially expressed genes, we used the zlm.SingleCellAssay() function from the MAST (v1.2.1) R package (51). The package uses raw UMIs count data and a hurdle model to account for zero-inflated data (such as single-cell RNA sequencing datasets). The zlm.SingleCellAssay() function models expression on the basis of cluster membership and number of detected genes per cell, therefore accounting for technical variations in transcript capture rate. To find differentially expressed genes between two clusters, we considered only genes detected above a minimum fraction of cells in at least one of the two clusters (typically 0.2).

Identification of One-to-one Orthologs

One-to-one orthologs were identified using the EggNOG mapper (beta version) web interface and the EggNOG (v.4.5) orthology data (52). EggNOG functional terms were determined for the turtle, lizard, mouse (GRCm38 assembly), and human (GRCh38 assembly) predicted proteomes, limited to one-to-one orthologs within the taxonomic scope of vertebrates. Within a species, we pruned both functional groups and genes (i) if any functional group that had multiple mappings to proteins derived from different genes, or (ii) if different protein isoforms of a gene mapped to different functional groups. Then the EggNOG functional group assignment for a gene was used as an intermediate to pair gene orthologs between two species.

Cross-species Pairwise Cluster Correlations

Gene expression in a cluster of single cells, i.e. a cell type, was calculated using the AverageExpression() function in Seurat. Clusters of cells, or groups of them (*e.g.* “superclusters”), were compared across species from these average gene expression data. The datasets used in these comparisons were acquired with Drop-seq (turtle and lizard, this paper), Fluidigm C1 single-cell RNAseq (mouse cortex (6)), SMARTer single-cell RNAseq (mouse cortex (7)), SMART-seq v4 single-cell RNAseq (mouse cortex (53)), single-nucleus RNAseq (mouse hippocampus (19)) and microarrays on microdissected brain regions (human pallium (13)). The single cell methods differ in number of genes detected, number of cells sequenced, quality of clusters, and dynamic range. Our approach was designed to mitigate potential problems and biases arising from the comparison of data acquired with these different methods, in different laboratories and from different species.

For the published mouse single-cell datasets (from (6), (7) and (19)), we used cluster assignments defined in the original publications. For the Allen Brain Institute ALM dataset (5,992 cells, <http://celltypes.brain-map.org/rnaseq>, cfr. fig. S14), we used Seurat as described above (with 32 principal components, clustering with resolution=7) and identified 33 clusters of glutamatergic neurons. These clusters were annotated on the basis of the expression of layer-specific markers (see also (53)).

In short, cell type correlation analysis between two sets of cell types in two species consists in: (1) the selection of a common gene set for the comparison, i.e. the intersection of differentially expressed one-to-one orthologs; (2) the transformation of the two “gene set × cell type” matrices in gene specificity matrices; (3) the calculation of Spearman rank correlations (and their *p*-values) of gene specificities among all pairs of cell types in the two species. Our criteria for gene selection, data transformation and correlation analysis are detailed below.

Gene set selection

Non-differentially expressed genes are not informative for comparative purposes, and correlations calculated on entire transcriptomes are driven by confounding factors such as the number of expressed genes (9). For this reason, we used differentially expressed one-to-one orthologs for the calculation of pairwise cluster correlations. In this way, we sought to limit our analyses to potentially conserved cell type markers. One-to-many orthologs cannot be compared directly: they cannot be considered functionally equivalent, because duplicated genes usually undergo neofunctionalization. We first determined the genes differentially expressed within each of the two datasets to compare, by calculating differentially expressed genes among all possible pairs of clusters in the dataset. Differentially expressed genes were computed with the MAST package as described above, and we typically selected genes differentially expressed between any pair of clusters with a minimum average difference of $\ln(1.5)$ (or $\ln(1.25)$) for the lizard data in fig. S11B). For the human microarray dataset (13), we used the R limma package (v3.28.21) to find differentially expressed genes among the pallial regions included in the dataset (with a fold-change $> \log_2(2)$ and an Holm-adjusted p -value $< 1e-05$). Of these genes, we considered only genes with high differential stability (DS) scores (as defined in (13)) for downstream analysis.

After identifying differentially expressed genes independently for each dataset and species of interest, we defined the gene set for comparison as the intersection of the differentially expressed one-to-one orthologs. With this choice, only genes detected and differentially expressed in both species/datasets compared were included in the analysis. This way, we also excluded genes absent in at least one of the two datasets compared, with the reasoning that, besides biology, differences in sequencing technique and depth might also account for non-detected genes. Notably, we did see similar trends in correlations when we expanded our gene set to include one-to-one orthologs that varied across cell classes in only one of the two species considered (*i.e.* union of the differentially expressed one-to-one orthologs in the two species).

Our criteria led to the inclusion of genes expressed in multiple clusters, that is, genes selected for comparison were not necessarily just those expressed in only one cell type and nowhere else. Consequently, this analysis also detects similarities between cell types within a species, and can explain the blocks of similarities we see in our correlation matrices. In other words, with this analysis a cell type will show similarities to multiple cell types both within and across species. Note also that the size of the gene set depends on the diversity of cell types included in the analysis. Heterogeneous sets of cells (for example, neurons plus non-neuronal types) had more differentially expressed genes, hence larger gene sets. In the comparisons of similar cell types (for example, glutamatergic neurons), gene sets were smaller, because there were fewer differentially expressed genes.

For the analyses restricted to transcription factors, a reference list of transcription factors was downloaded from Ensembl using the gene ontology terms GO:0003700 (transcription factor activity), GO:0003702 (RNA polymerase II transcription factor

activity), GO:0003709 (RNA polymerase III transcription factor activity), GO:0016563 (transcriptional activator activity) and GO:0016564 (transcriptional repressor activity).

The pairwise cluster correlations reported in this paper were calculated from the following one-to-one orthologs gene sets:

- turtle glutamatergic superclusters and human pallial regions: 1017 genes (Fig. 2E) and, of these, 73 transcription factors (Fig. 2F);
- turtle and lizard hippocampal clusters (Fig. 3B, top): 242 genes;
- turtle and mouse hippocampal clusters (Fig. 3B, bottom): 222 genes;
- turtle aDC glutamatergic clusters and mouse V1 neocortical layer types (Fig. 4C): 260 genes;
- turtle and mouse GABAergic clusters (Fig. 5B): 406 genes;
- turtle and lizard neuronal and non-neuronal clusters (fig. S4D): 1199 genes;
- turtle and mouse neuronal and non-neuronal clusters (fig. S4E): 1418 genes;
- lizard and mouse neuronal and non-neuronal clusters (fig. S4F): 1080 genes;
- turtle and lizard glutamatergic pallial areas (fig. S9E): 350 genes;
- lizard and mouse hippocampal clusters (fig. S11B): 548 genes.
- turtle aDC and mouse ALM glutamatergic clusters (fig. S14B): 250 genes.
- turtle aDC and mouse S1 glutamatergic clusters (fig. S14C): 248 genes.
- turtle aDVR and mouse V1 neocortical layer types (fig. S14D): 260 genes.
- turtle aDVR and mouse ALM glutamatergic clusters (fig. S14E): 210 genes.
- turtle aDVR and mouse S1 glutamatergic clusters (fig. S14F): 203 genes.

Gene specificity index

With the criteria described above, we identified a common gene set for the two datasets and species to compare, and we reduced the data to the two matrices “gene set \times cell type set species 1” and “gene set \times cell type set species 2”. Before calculating pairwise cluster correlations, we transformed each of these matrices into gene specificity matrices. Given a cell type set C with N cell types, the specificity index $s_{g,c}$ of the gene g for the cell type c within C is defined as the ratio between g_c (the expression level of g in c) and the mean expression of g in C :

$$s_{g,c} = \frac{g_c}{\frac{1}{N} \sum_{i \in C} g_i}$$

where $s_{g,c} \in [0, N]$. If g is not expressed in c , but is expressed in other cells in C , then $s_{g,c} = 0$ (genes expressed in none of the cell types are not included in the analysis). $s_{g,c} = 1$ if g_c is equal to the mean expression of g in the cell type set C , and $s_{g,c} = N$ if g is expressed only in cell type c and nowhere else in C . Note that cell types here are defined as clusters in transcriptomics space. Practically, gene expression g_c is the expression of g averaged across all the single cells in cluster c . The mean expression of g

in C (denominator of the equation above) is calculated as the mean of all g_i ($i \in C$), i.e. of all averaged cell type expression values.

Such a specificity score helps to resolve cross-platform differences in expression quantitation, allowing us to compare our Drop-seq data to single-cell and microarray data from other publications. In contrast to methods that binarize data into expressing or non-expressing cell classes, this index preserves relative expression of a gene between high and low-expressing cell classes. This specificity index, similar to those previously used to compare tissue types (4), is also useful to sidestep the confound of inherent similarities in gene expression within a class of genes (e.g. cytoskeleton genes have higher expression levels than transcription factors) that lead to high correlation values when comparing raw expression data. In other words, after this transformation each gene is scaled within the same $[0, N]$ range and can contribute to the correlation regardless of its absolute expression level. Finally, using a specificity index instead of raw expression data mitigates potential biases arising from the comparison of data acquired with different technologies (i.e. from differences in depth, dynamic range etc.), because all data are within $[0, N]$.

Correlations

From the gene specificity matrices determined as above, pairwise Spearman rank order correlations were calculated. Spearman rank correlation is well suited to compare different cell type sets (each with its own number N of cell types), because it compares the rank orders of gene specificities, and not their absolute values.

Significance for correlation coefficients was determined with a permutation test. Gene expression values were shuffled 1,000 times across cell types and the resulting Spearman correlation coefficient rho was calculated. The p -value was calculated as the fraction of the absolute value of rho values that were greater than or equal to the absolute value of rho for the actual (i.e. non-shuffled) data.

Limitations

With this pairwise cluster correlation analysis, data acquired with different technologies become comparable after the selection of an appropriate gene set and the gene specificity transformation. Our pipeline is conservative, and this comes with a few limitations. First, an analysis restricted to differentially expressed one-to-one orthologs leads to a reduced representation of cell type molecular diversity. Genes excluded for biological and technical reasons (one-to-many orthologs, genes detected in only one of the two datasets to compare, see above) might be important to define cellular identity and reconstruct the evolutionary diversification of cell types. Second, the discriminatory power of this analysis depends on the cell type set of choice, because gene sets and gene specificity scores depend on the cell type set. For example, the comparisons of neuronal and non-neuronal cells in fig. S4D-F do not reveal fine correspondences between excitatory subtypes, because the gene set used is dominated by markers that distinguish neuronal vs non-neuronal cells. For similar reasons, this type of comparison is

meaningful only among cells within the same class or anatomical region (GABAergic neurons, dorsal pallium glutamatergic neurons etc).

Given the limitations described above, we cannot completely exclude the existence of finer similarities and homologies between subtypes of cells (for example, between turtle and mouse subtypes of GABAergic interneurons). With the data currently available, restricting the analysis to a smaller number of cell types (e.g. only to turtle and mouse PV-like interneurons) results in sets of differentially expressed one-to-one orthologs that are too small for reliable statistics.

In addition, our analysis was limited by the availability and quality of mammalian data; this applies, in particular, to the comparisons of pallial regions (Fig. 2E-F) and of GABAergic interneurons (Fig. 5B). A mammalian pallium single-cell RNAseq dataset would have been the most suitable for a comparison with our reptilian data, but such a dataset is not available yet. The microarray data from (13) are the only ones, to date, encompassing different pallial regions, including neocortex, hippocampus, claustrum, piriform cortex, and amygdala. The tissue resolution of the microarray data prompted us to reconstruct turtle tissue data *in silico* by pooling cells from the same anatomically-defined pallial region (“superclusters”). Even if the human microarray data do not have single-cell resolution, the selection of differentially expressed one-to-one orthologs excludes automatically most non-neuronal and neuronal GABAergic markers, and correlations are driven by similarities of neuronal (glutamatergic) types.

For GABAergic interneurons (Fig. 5B), we compared turtle MGE and CGE-derived cells, sampled from the entire pallium, with mouse cells, sampled from the visual cortex (7). A comparison of GABAergic cells sampled from the same brain regions in the two species would have been more appropriate. However, single-cell data for non-cortical mammalian GABAergic neurons are not available yet. As of our turtle dataset, all MGE- and CGE-derived GABAergic clusters contain cells sampled from dorsal cortex dissections (fig. S7C-D). This suggests that interneurons from all these clusters should be present in the dorsal cortex, and that the comparison with mammalian cortical interneurons (from (7)) is legitimate.

Despite these limitations, we observe that all our pairwise correlation analyses lead to conclusions consistent with the analysis of conserved cell type markers known from the literature (e.g. Fig. 1B,D, Fig. 2D, Fig. 3C, Fig. 5A, Supplementary Figures) and with weighted gene co-expression network analysis (see below).

Weighted Gene Co-expression Network Analysis

Weighted gene co-expression network analysis (WGCNA) was performed with functions in the WGCNA R package (54). To attenuate the effects of noise and outliers, the analyses were performed on pseudocells, calculated as averages of 4-10 cells randomly chosen within each cluster.

For the analysis of turtle glutamatergic data (Fig. 2C), the 4,854 turtle glutamatergic cells were converted to 490 pseudocells (44 averages of 9 cells and 445 averages of 10 cells). High variable genes among the cells of interest were calculated using the distance to the median (DM) metric (55). The top 1500 variable genes determined in this way were used for analysis. An adjacency matrix, representing a “signed” gene network, was built setting the soft power parameter to 12 (calculated from the pickSoftThreshold() function). From this adjacency matrix, a topological overlap matrix (TOM) was calculated (TOMsimilarityFromExpr() function); the TOM dissimilarity measure (1-TOM) was then used to cluster genes (hierarchical clustering, “average” method). Modules were identified with the dynamic tree cutting algorithm using the function cutreeDynamic(), with a minimum module size of 50 and the PAM stage parameter set to TRUE. Module eigengenes, defined as the first principal component of the expression matrix of a module, were calculated using the moduleEigengene() function.

For the analysis and comparison of modules across species, gene modules were calculated as above from one-to-one orthologs expressed with high variance in the cell types of interest in either one or the other species. Pseudocells calculated as averages of 4 or 5 cells randomly selected from each cluster were used for analysis. Genes expressed in less than 20% of cells in all clusters of either one or the other species were excluded from analysis. High variable genes among the cells of interest were calculated using the distance to the median (DM) metric (55). Genes (one-to-one orthologs) were then ranked by DM value in each of the two species, and the union of the top n genes in the two species was taken for further analysis. In parallel analyses, we verified that gene modules obtained from these gene sets (constrained by the need to compare two species) were similar to those calculated from high variable genes selected independently for each dataset.

For the comparison of hippocampal glutamatergic neurons, the top 600 turtle and mouse genes were selected, for a total of 1094 genes. Adjacency matrices were calculated independently for the two species, setting the soft threshold parameter to 4, and modules were calculated with minimum module size of 200. For the comparison of turtle aDC glutamatergic clusters and mouse neocortical glutamatergic neurons, the settings were: 1000 top turtle and mouse genes, for a total of 1657 genes; soft threshold parameters: 5; minimum module size: 20. For the comparison of turtle and mouse cortical GABAergic interneurons, the settings were: 750 top turtle and mouse genes, for a total of 1208 genes; soft threshold parameters: 6 (turtle) and 7 (mouse); minimum module size: 30.

Overlaps of gene modules (*i.e.* number of genes in common between any of two modules) across species were computed, and p -values of overlaps were calculated using the hypergeometric distribution (Fisher’s exact test). To plot gene networks, for each module network edges were ranked by weight, and the top 750 edges (hippocampus) or 500 edges (GABAergic interneurons) were selected, together with the nodes (*i.e.* genes) connected by them. Networks were plotted in Cytoscape (v3.5.1) using the “Perfuse Force Directed Layout”; then turtle and mouse networks were merged with the Union function.

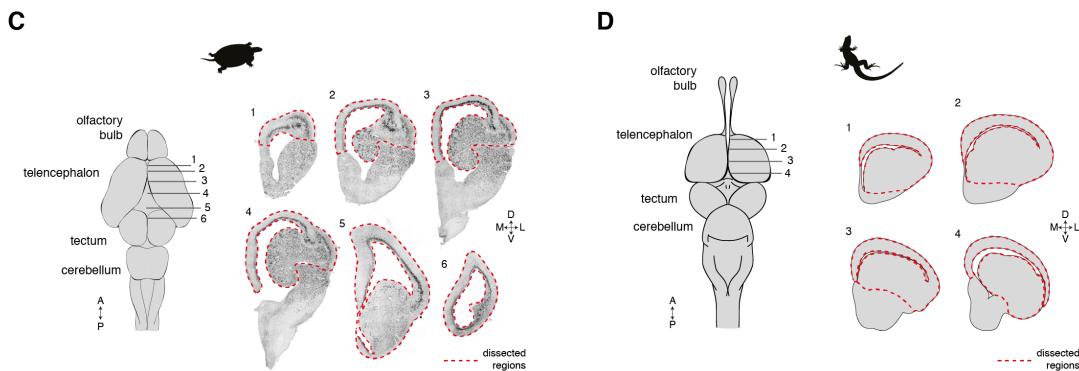
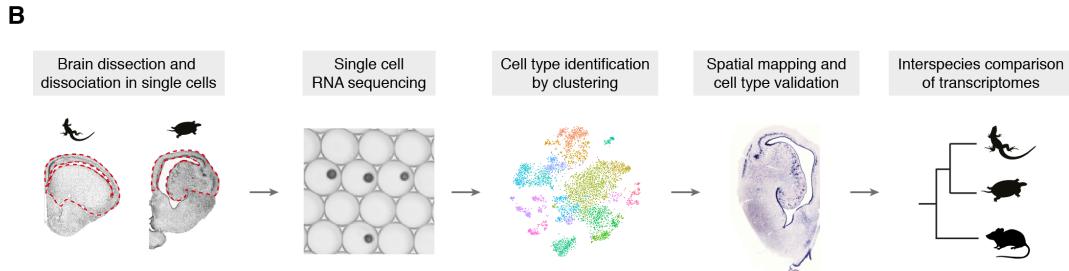
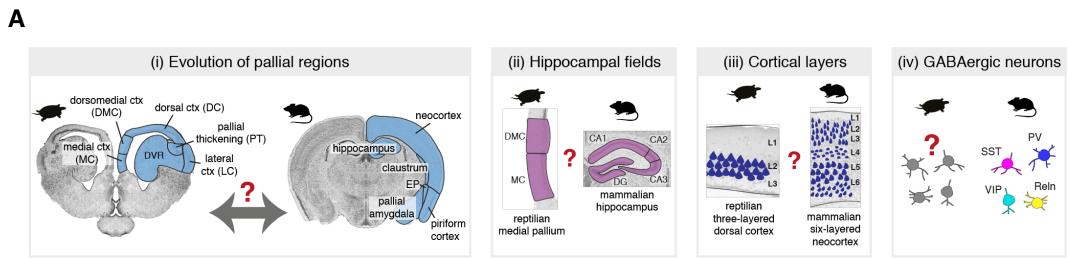


Fig. S1. Main questions, methods overview and brain dissections.

(A) Main questions addressed in this study. (i) The evolutionary relationships between anatomically defined pallial regions in amniotes are heavily disputed. In particular, it is not clear whether the reptilian anterior DVR is the homolog of the mammalian neocortex (1,2,10-12). (ii) The evolutionary relationships between subdivisions of the reptilian and mammalian hippocampi are also unclear. The presence of a dentate gyrus (DG) homolog in reptiles is disputed (3). (iii) The stem ancestor of all amniotes (reptiles, birds and mammals) presumably had a three-layered cortex. The reptilian cortex is three-layered along the entire mediolateral extent (birds lost a layered cortical architecture). In mammals, the dorsal part of the cortex acquired a six-layered organization (“neocortex”), whereas the hippocampus (medial) and the piriform cortex (lateral) retained a three-layered organization. The cell type relationships between the reptilian three-layered dorsal cortex and the mammalian six-layered neocortex are still poorly understood (1,4,5). (iv) Cortical GABAergic neurons arise from conserved embryonic territories (ganglionic eminences) in the subpallium (31,32). It is unclear, however, whether the same types of GABAergic neurons exist in reptiles and mammals.

(B) Methods overview.

| (C) and (D) Schematics of turtle and lizard brains (dorsal views) and frontal sections through their forebrains. The regions dissected for single-cell RNA sequencing are highlighted. See also fig. S7A for a more detailed illustration of turtle dissections.

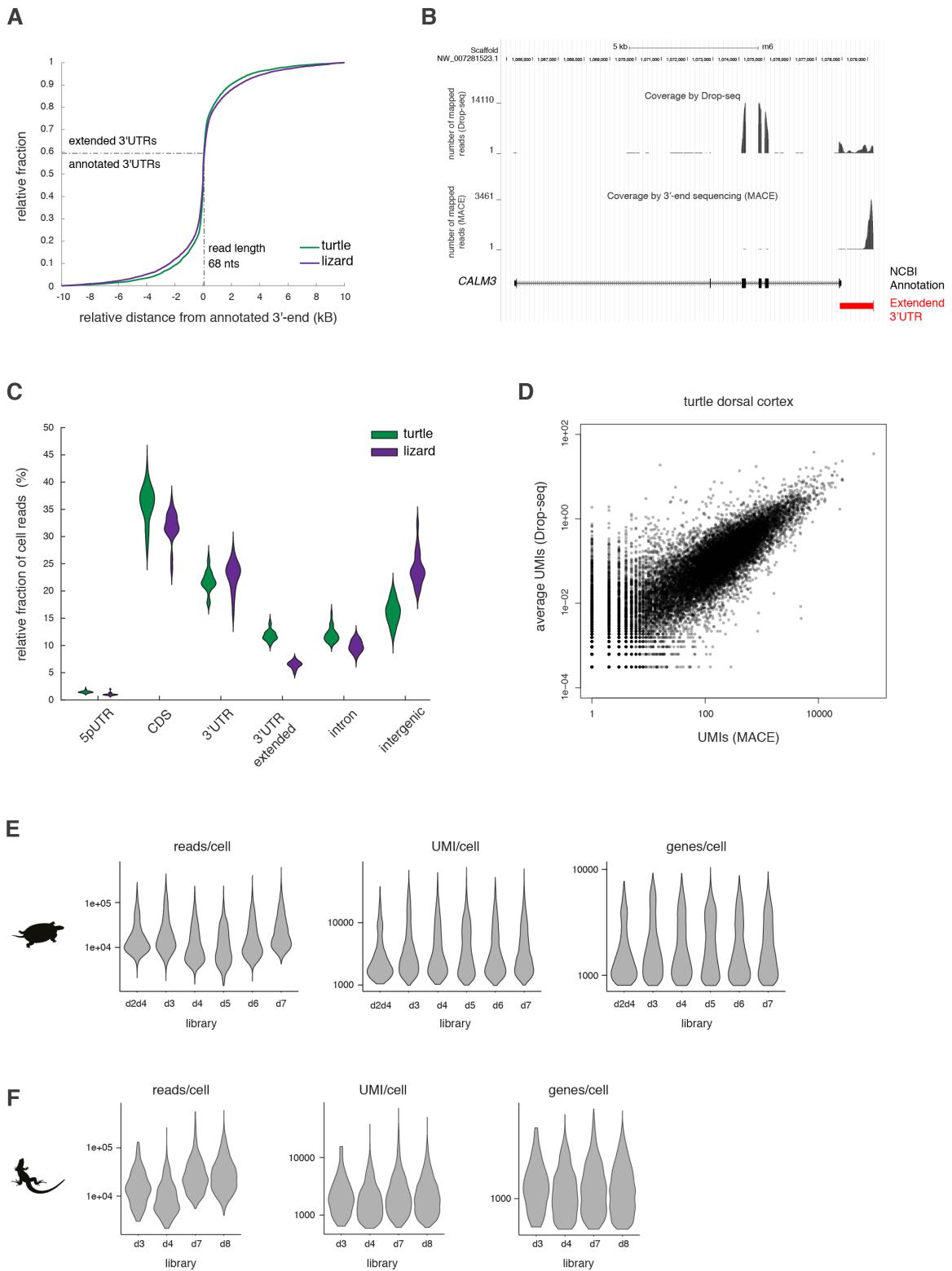


Fig. S2. Turtle and lizard 3'UTR annotation and Drop-seq statistics.

- (A) Cumulative plots showing the relative distances (in kB) between the previously annotated turtle and lizard 3'-ends and our extended annotations (see *Methods*).
- (B) Representative example of an extended 3'-end annotation. Top: reads from Drop-seq experiments mapping on the turtle *CALM3* gene and its neighborhood. A significant fraction of reads maps on the genomic region downstream the NCBI-annotated *CALM3* 3'-end. Bottom: in the same genomic region, there is a sharp peak of 3'-MACE reads downstream the annotated *CALM3* 3'-end. The extended 3'-end annotation (red), based on the MACE data, includes the genomic region up to this peak.
- (C) Violin plots showing mapping of Drop-seq reads on UTRs, coding regions (CDS), introns and intergenic regions. A substantial percentage of reads (>5%) maps to extended 3'UTRs.
- (D) Correlation between the average number of transcripts (UMIs) detected by Drop-seq in turtle cortical samples and the number of transcripts measured with 3'-MACE from turtle cortical samples.
- (E) and (F) Violin plots showing the number of reads/cell, UMIs/cell and genes/cell in turtle and lizard datasets, before filtering of low-quality cells (see *Methods*).

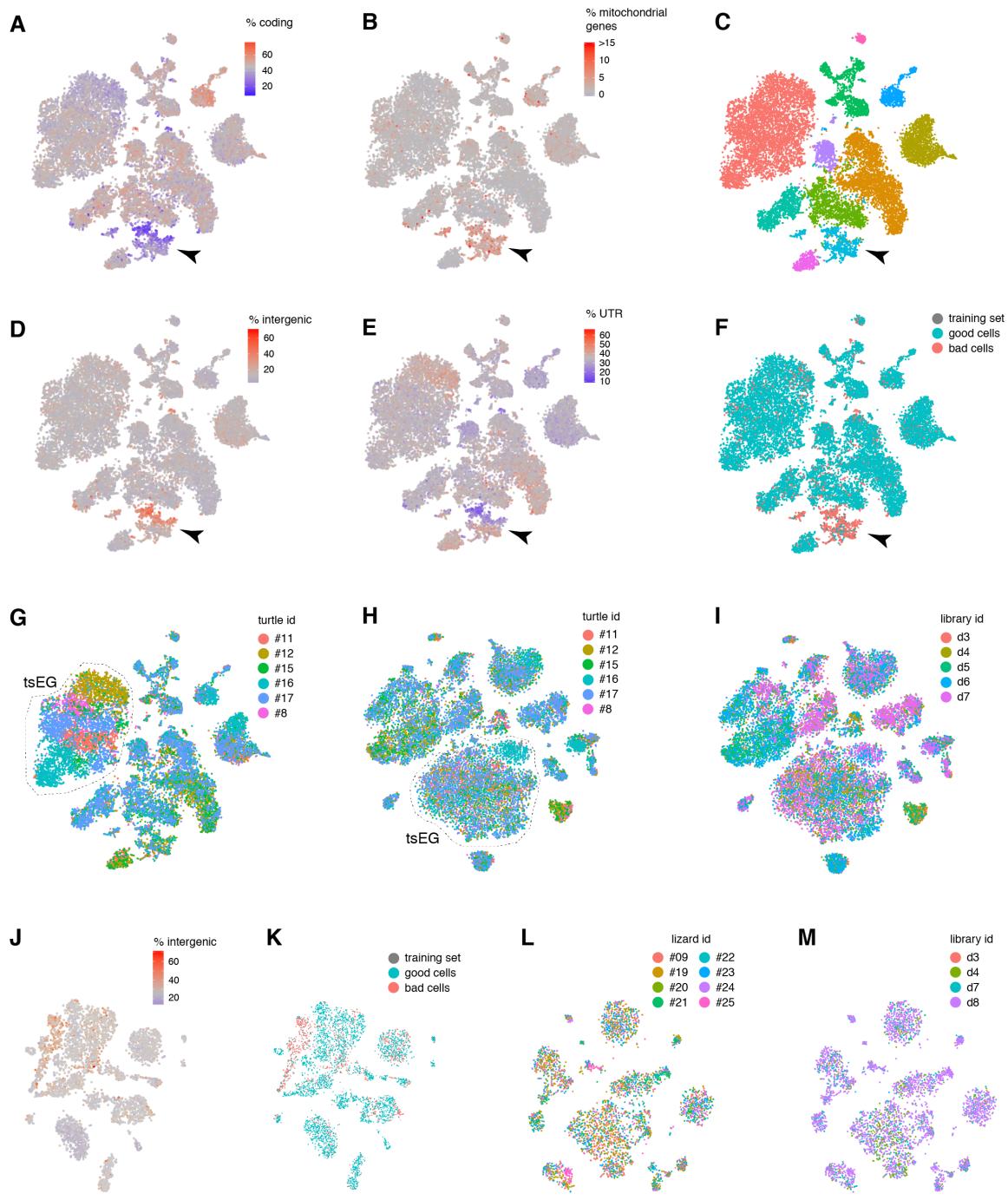


Fig. S3. Drop-seq data filtering and batch correction.

(A), (B), (D) and (E) tSNE plots of turtle single-cell transcriptomes before filtering. Cells are color-coded according to read mapping statistics.

(C) tSNE plot showing clusters of turtle cells before filtering. Bad quality cells cluster together (cyan cluster, arrowhead).

(F) Classification of good and bad-quality turtle cells with a support vector machine (SVM) classifier trained on cell quality metrics (see *Methods*).

(G) tSNE plot of turtle cells before SVM-based filtering and batch correction, cells color-coded by animal of origin. Cells in the ependymoglia cells cluster (tsEG) segregate by animal of origin.

(H) and (I) tSNE plots of turtle cells (18,828 cells) after SVM-based filtering and batch correction, cells color-coded by animal (H) or library id (I).

(J) tSNE plots of lizard cells before SVM-based filtering, cells color-coded by percentage of reads mapping to intergenic regions.

(K) Classification of good and bad-quality lizard cells with SVM.

(L) and (M), tSNE plots of lizard cells (4,187 cells) after SVM-based filtering and batch correction, cells color-coded by animal (L) or library (M).

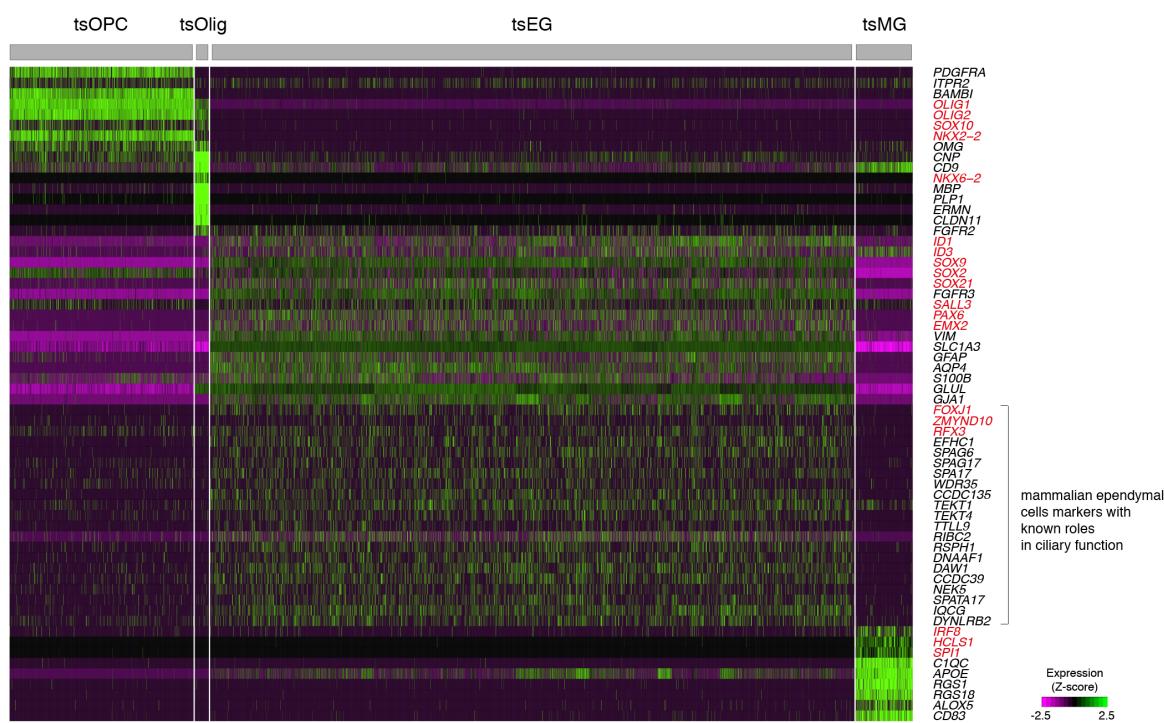
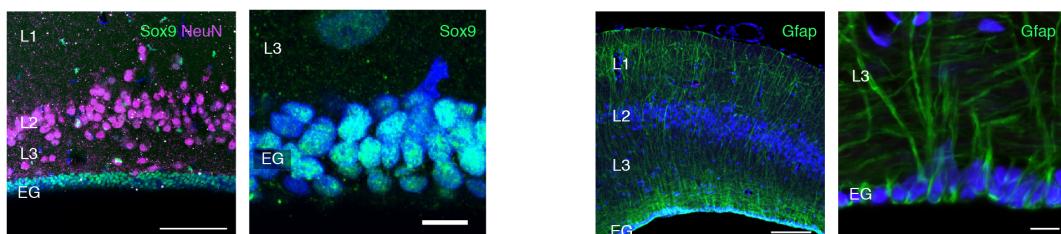
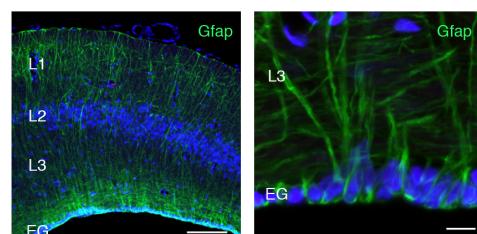
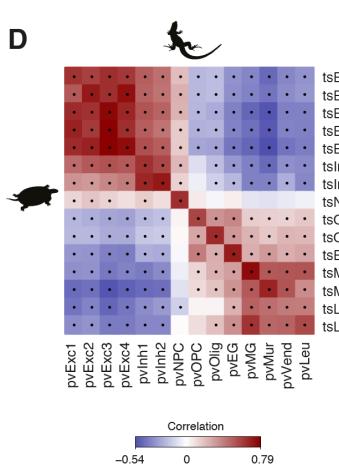
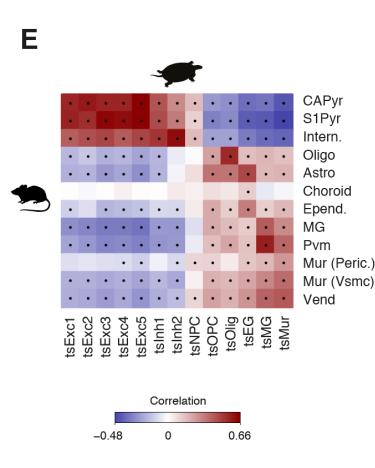
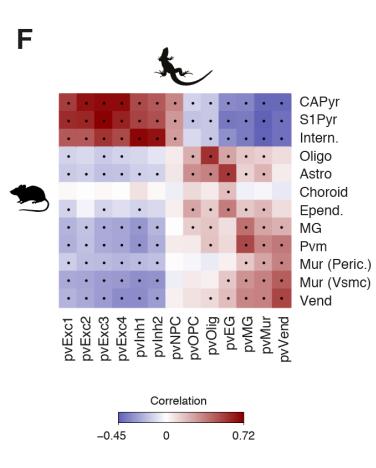
A**B****C****D****E****F**

Fig. S4. Glial cells in the turtle and lizard pallium.

- (A) Heatmap of gene expression in turtle glial cells, showing expression of mammalian markers of immature and mature oligodendrocytes, of astrocytes and ependymal cells, and microglia (6). Note that mammalian ependymal cells markers, including genes with known roles in ciliary function, are expressed in turtle ependymoglia cells. Names of transcription factors in red. tsOPC: oligodendrocyte precursor cells; tsOlig: mature oligodendrocytes; tsEG: ependymoglia cells; tsMG: microglia.
- (B) Left: immunohistochemistry on a turtle cortical section, showing the ependymoglia cells marker Sox9 (green), the pan-neuronal marker NeuN (magenta) and DAPI (blue). Right: close-up. Scale bars: 100 μ m (left) and 10 μ m (right).
- (C) Left: immunohistochemistry on a lizard cortical section, showing the ependymoglia cells marker Gfap (green) and DAPI (blue). Right: close-up. Scale bars: 100 μ m (left) and 10 μ m (right).
- (D), (E) and (F): Pairwise Spearman correlations of high-level neuronal and non-neuronal clusters. Correlations calculated from differentially expressed genes (see *Methods*).
- (D) Comparison of lizard (columns) and turtle (rows) neuronal and non-neuronal clusters.
- (E) Comparison of turtle (columns) and mouse (rows) neuronal and non-neuronal clusters.
- (F) Comparison of lizard (columns) and mouse (rows) neuronal and non-neuronal clusters. In (E) and (F), mouse data are from (6).
- CAPyr: CA Pyramidal cells; S1Pyr: S1 pyramidal cells; Intern.: GABAergic interneurons; Astro.: astrocytes; Epend.: ependymal cells; MG: microglia; Pvm: perivascular macrophages; Mur: mural cells, subdivided in pericytes (Peric.) and Vascular smooth muscle cells (Vsmc); Vend.: vascular endothelial cells.

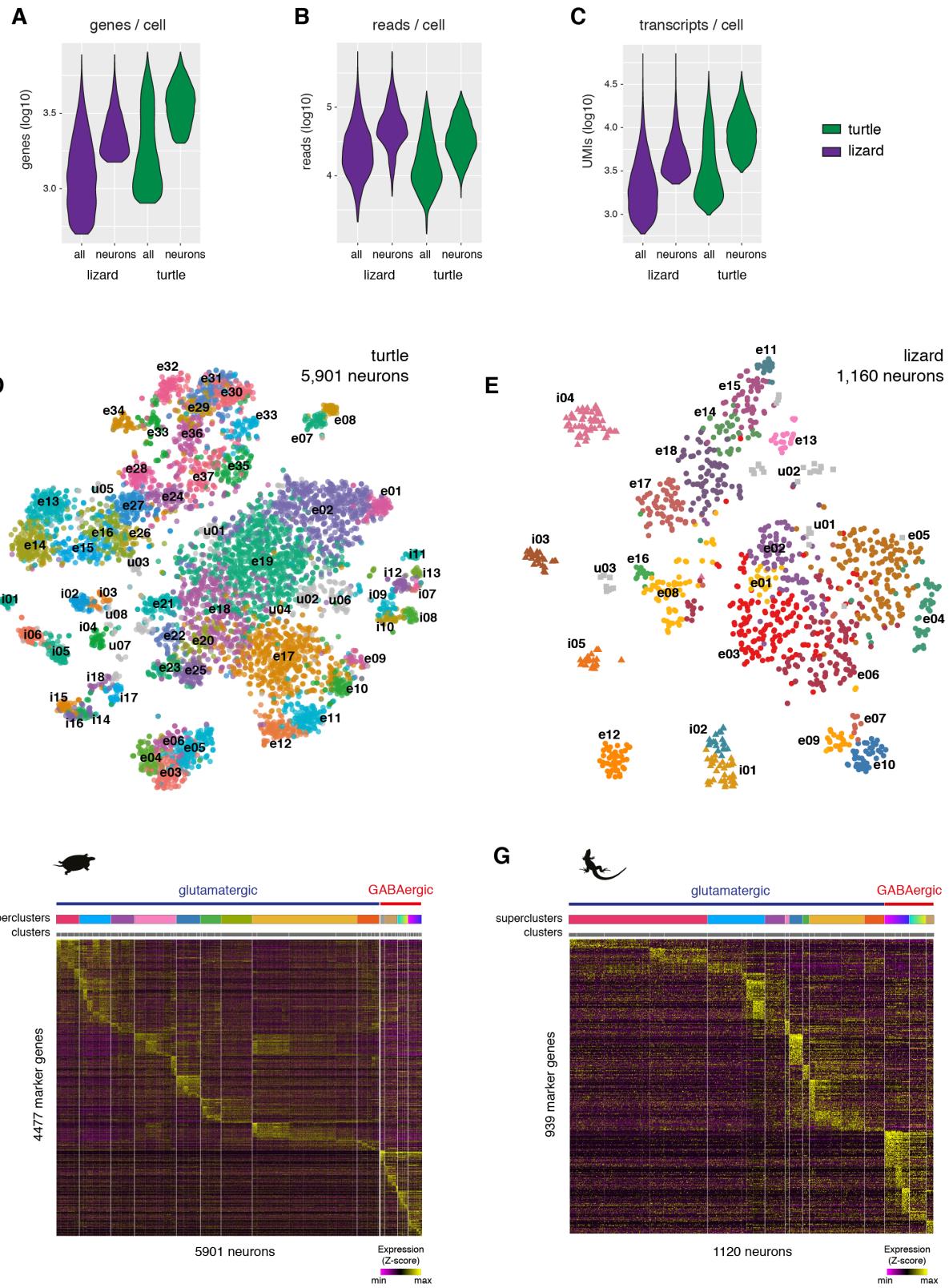


Fig. S5. Subclustering of turtle and lizard neurons.

- (A) to (C) Statistics (reads, transcripts and genes/cell) of the turtle and lizard final neuronal datasets (“neurons”) compared to the whole datasets (“all”, neurons and non-neuronal cells), after quality filtering. Cutoffs of 2000 genes/cell or 1500 genes/cell were used for the turtle and lizard neuron datasets, respectively.
- (D) tSNE plot and clusters of 5,901 high-quality excitatory (e) and inhibitory (i) turtle neurons. Same as in Fig. 1E, with cluster names included. u: unassigned clusters (grey).
- (E) tSNE plot and clusters of 1,160 high-quality excitatory (e) and inhibitory (i) lizard neurons. u: unassigned cells (grey). In (D) and (E), clusters are not matched by name or color.
- (F) Heatmap showing expression of cluster-specific genes (rows) in turtle single neurons (columns), single-cells arranged by cluster (gray bars) and high-level classification (glutamatergic vs. GABAergic and superclusters, see Fig. 2).
- (G) As in (F) for the lizard neuronal dataset.

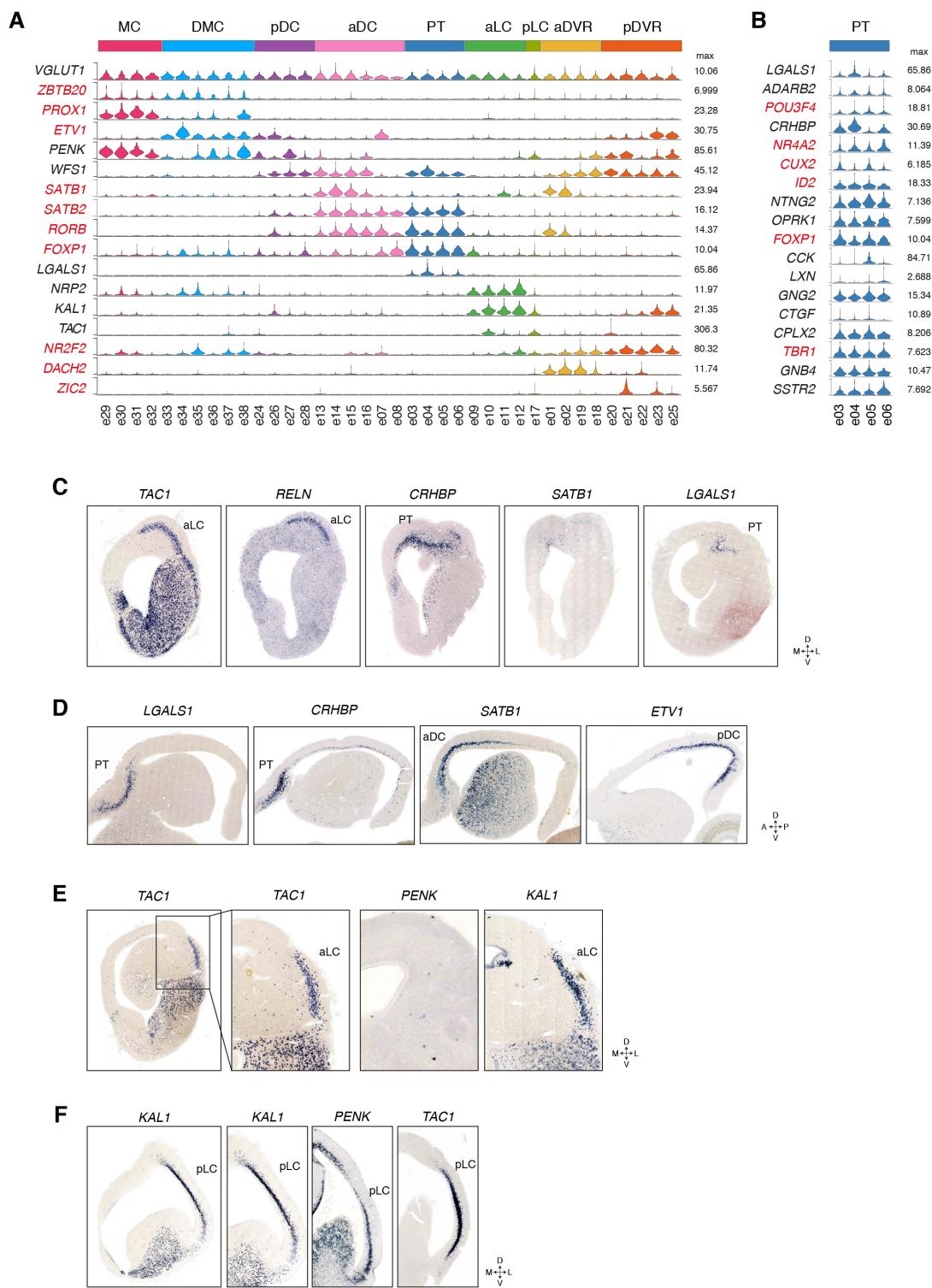


Fig. S6. Expression of region-specific glutamatergic markers in turtle.

- (A) Violin plots showing expression of genes in turtle pallial glutamatergic types. The combinatorial and spatially-restricted expression of these genes was used to map glutamatergic types on anatomy.
- (B) Violin plots showing expression of turtle PT (“pallial thickening”) markers (*LGALS1*, *ADARB2*, *POU3F4*, *CRHBP*) and of mammalian claustrum markers (all other genes) in cells mapping to the turtle PT.
- In (A) and (B), for each gene, violin plots are scaled to the maximum number of transcripts (UMIs) detected for that gene (max). Names of transcription factors in red.
- (C) Frontal sections through the anterior turtle forebrain (~F1 in fig. S8, *LGALS1* section more posterior). *RELN* and *TAC1* are expressed in the anterior lateral cortex (aLC), the reptilian homolog of the mammalian piriform cortex. *RELN*, a marker of semilunar cells in the piriform cortex, is expressed in a subset of cells of the turtle and lizard aLC (cfr also fig. S9F). *CRHBP* and *LGALS1* label the pallial thickening (PT). The anterior dorsal cortex marker *SATB1* is not expressed in PT (see also Fig. 2D).
- (D) Sagittal sections showing expression of the PT markers *LGALS1* and *CRHBP*, of the aDC marker *SATB1* and the posterior dorsal cortex (pDC) marker *ETV1*.
- (E) Frontal sections (~F3 in fig. S8) showing expression of *TAC1* and *KAL1* but not *PENK* in the anterior lateral cortex (aLC).
- (F) Frontal sections (~F5 in fig. S8) showing expression of *KAL1*, *PENK* and *TAC1* in posterior lateral cortex (pLC). See also Fig. 2.



Fig. S7. Dissections from the turtle pallium.

- (A) Schematics showing different pallial dissections, from which separate Drop-seq experiments were performed. “Cortex” indicates all cortical regions (note that the “cortex” dissections include also part of the posterior DVR), “posterior cortex” indicates caudal cortical regions, including the cortical amygdala. DC: dorsal cortex; MC: medial cortex; DMC: dorsomedial cortex; DVR: dorsoventricular ridge.
- (B) tSNE plot of turtle neurons, cells color-coded by dissection (same color code as in A).
- (C) Cumulative barplot showing, for each cluster, the relative contribution of cells from different dissections (color code as in A). The pallial origin attributed to each cluster, as reconstructed from the *in situ* hybridizations with cluster-specific markers (cfr Fig. 2 and fig. S6), is consistent with the contributions of different dissections to the clusters. For example, the majority of cells assigned to clusters e03-e06 (which map to pallial thickening) came from those samples containing the pallial thickening (“cortex” and “cortex wo DC”, cfr fig. S7A). Imprecise dissections can explain the presence of cells from neighboring regions (in this example, from “DVR” dissections).
- (D) Barplot showing, for each cluster, the number of neurons coming from the “DC” (dorsal cortex) dissections. All clusters of MGE- and CGE-derived GABAergic interneurons (i07-i18, cfr. Fig. 5) contain cells dissected from the dorsal cortex.
- (E) Cumulative barplot showing, for each cluster, the total number of neurons sampled from the “DVR” dissections and from all the remaining dissections (grey).

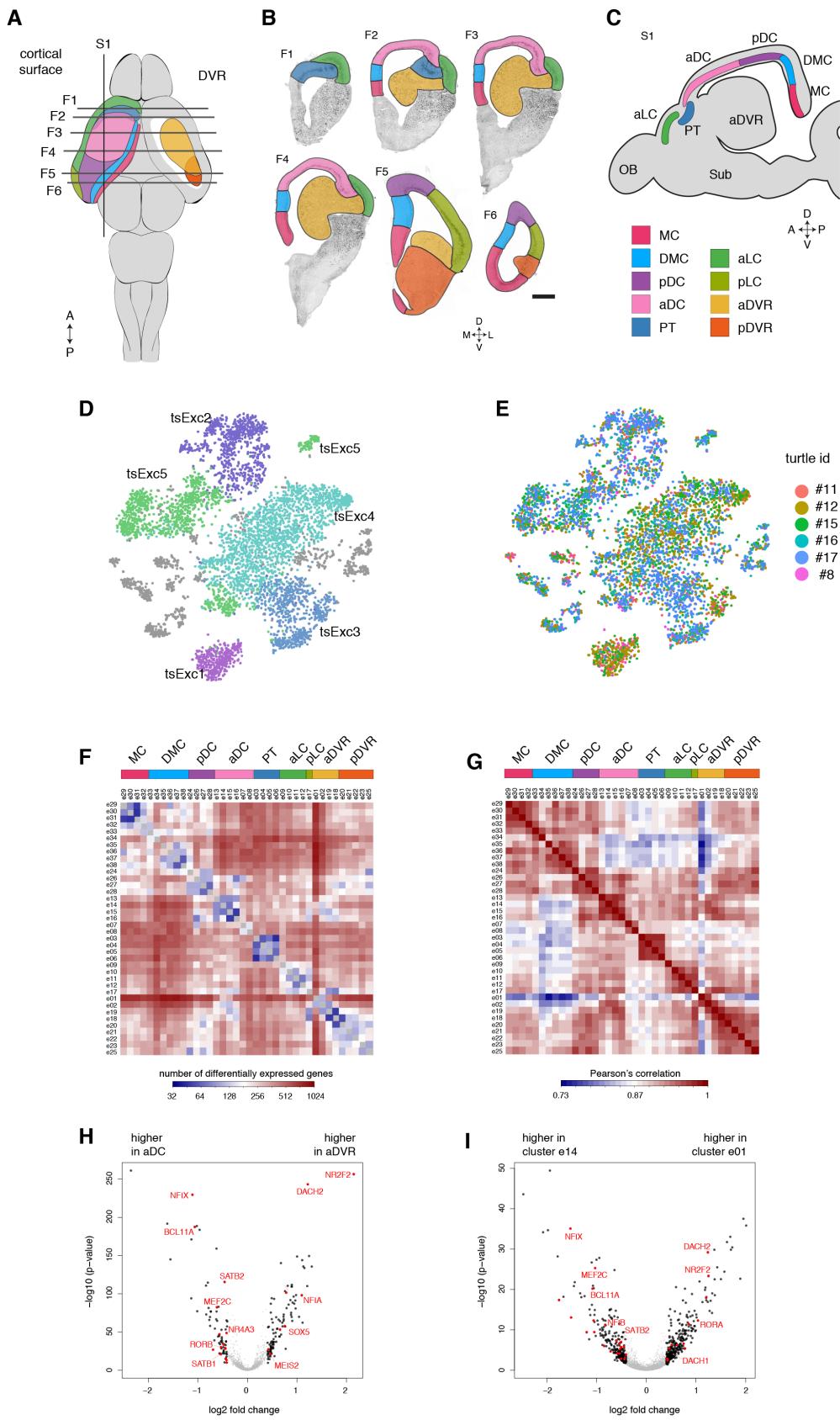


Fig. S8. Pallial regions in turtle.

- (A) Schematic of the turtle brain, dorsal view, illustrating pallial regions.
(B) Annotated frontal sections (levels indicated in (A)) stained with the pan-neuronal marker NeuN. Scale bar: 1mm.
(C) Schematic of a sagittal section through the turtle forebrain (level indicated in (A)), showing the relative positions of the regions depicted in (A). OB: olfactory bulb; Sub: subpallium. MC: medial cortex; DMC: dorsomedial cortex; pDC and aDC: posterior and anterior dorsal cortex; PT: pallial thickening; aLC and pLC: anterior and posterior lateral cortex; aDVR and pDVR: anterior and posterior dorsal ventricular ridge. Same color scheme in (A), (B) and (C).
(D) tSNE plot showing the turtle neurons dataset with glutamatergic cells color-coded by cluster assignment in Fig. 1A (high-level clustering of turtle cells). Non-glutamatergic cells in grey.
(E) tSNE plot showing the turtle neurons dataset with cells color-coded by animal of origin. Within each cluster, cells coming from different animals are evenly distributed, indicating that batch effects do not contribute to clustering. Not every pallial area was sampled from each animal (see also Table S1). This explains why some clusters (for example, PT) are made of cells from some, but not all, animals.
(F) Number of differentially expressed genes among glutamatergic clusters (genes detected in more than 40% of cells in at least one cluster, average expression difference $> \ln(1.5)$ in at least one of the pairwise comparisons). Clusters are ordered by pallial region.
(G) Pearson's correlations among turtle glutamatergic clusters, computed using the same differentially expressed genes used for (F). Clusters are ordered by pallial region.
(H) Volcano plot showing the 181 genes differentially expressed (black dots) between turtle aDC and aDVR. Transcription factors in red.
(I) Volcano plot showing the 579 genes differentially expressed (black dots) between turtle glutamatergic clusters e14 (in aDC) and e01 (in aDVR). Among all aDC and aDVR clusters, these two clusters have the lowest number of differentially expressed genes (cfr. (F) and (G)); these cells correspond most likely to neurons receiving visual inputs in cortex and aDVR. Transcription factors in red.

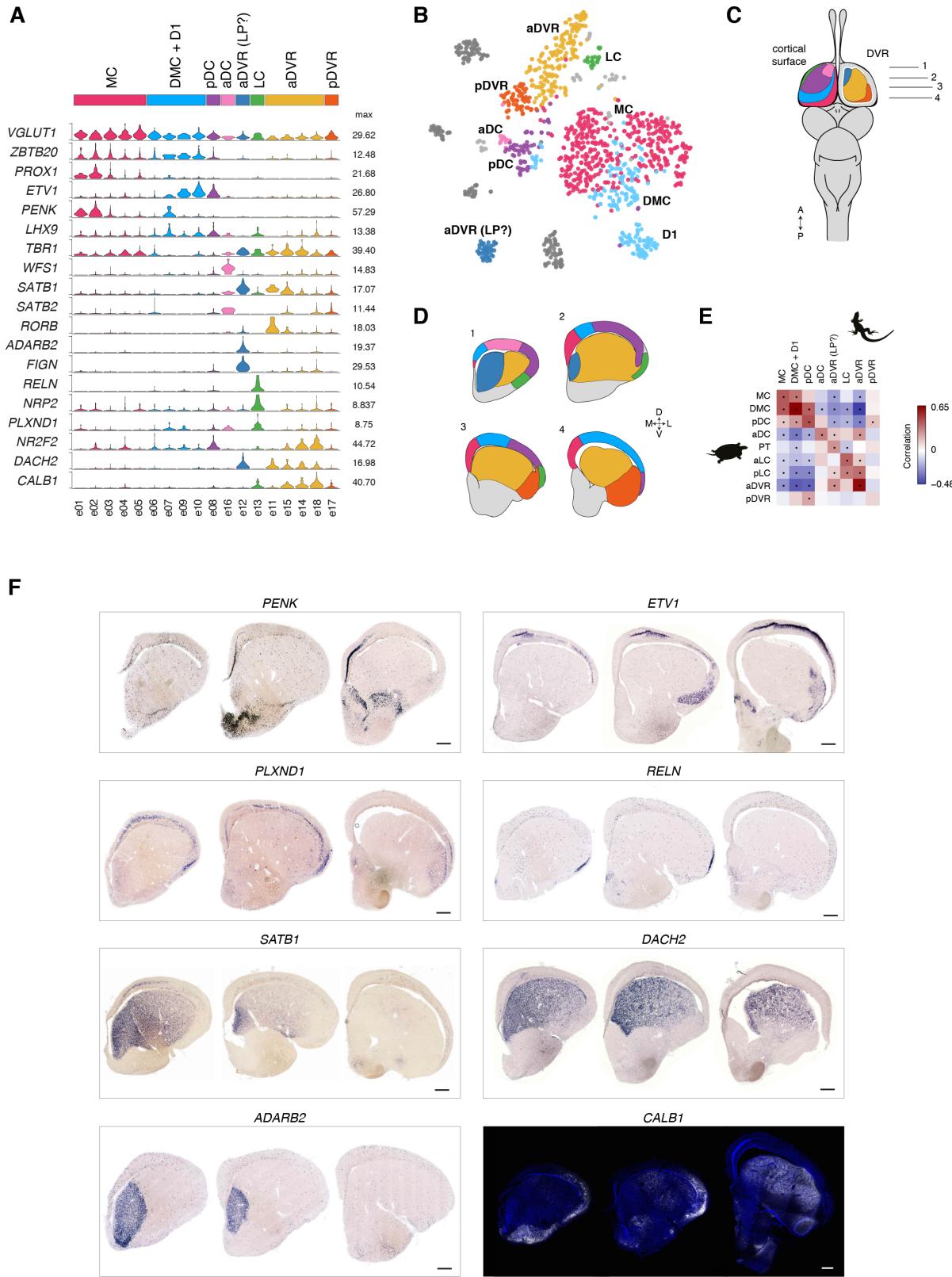


Fig. S9. Pallial regions in lizard.

- (A) Violin plots showing expression of genes in lizard pallial glutamatergic types. The combinatorial and spatially-restricted expression of these genes was used to map glutamatergic types on anatomy.
- (B) tSNE plot of lizard neurons, showing glutamatergic cells color-coded by area assignment. GABAergic neurons in grey. aDC: anterior dorsal cortex; pDC: posterior dorsal cortex; LC: lateral cortex; MC: medial cortex; DMC: dorsomedial cortex; D1: D1 area of dorsal cortex; aDVR: anterior DVR; pDVR: posterior DVR. aDVR (LP?) indicates a molecularly distinct region of the aDVR which has been recently proposed as a Lateral Pallium (LP) derivative in lizards, and as the homolog of the turtle pallial thickening and the bird mesopallium (1).
- (C) Schematic of the lizard brain from a dorsal view, showing the regions defined in (B) on the cortical surface (left) and in the DVR (right).
- (D) Frontal sections (levels 1-4 in (C)) of the lizard forebrain, with the regions defined in (B) labeled. Color scheme from (A) and (B).
- (E) Pairwise correlations of turtle and lizard superclusters, based on differentially expressed genes (see *Methods*).
- (F) Expression of some of genes shown in (A) at different rostro-caudal levels (frontal sections). These expression patterns were used to map lizard clusters to pallial regions. Note that some of these specific regional markers do not have conserved expression in turtle and lizard (for example, *CALB1*, expressed in lizard but not turtle aDVR). Conserved expression patterns in turtle and lizard include: *ETV1* in DMC, pDC and parts of pDVR; *RELN* in a subset of aLC cells; *SATB1* in the aDC and parts of the aDVR; *DACH2* in the aDVR; *ADARB2* in the turtle PT and the lizard aDVR/LP region (cfr. also Fig. 2). Scale bars: 500 μ m

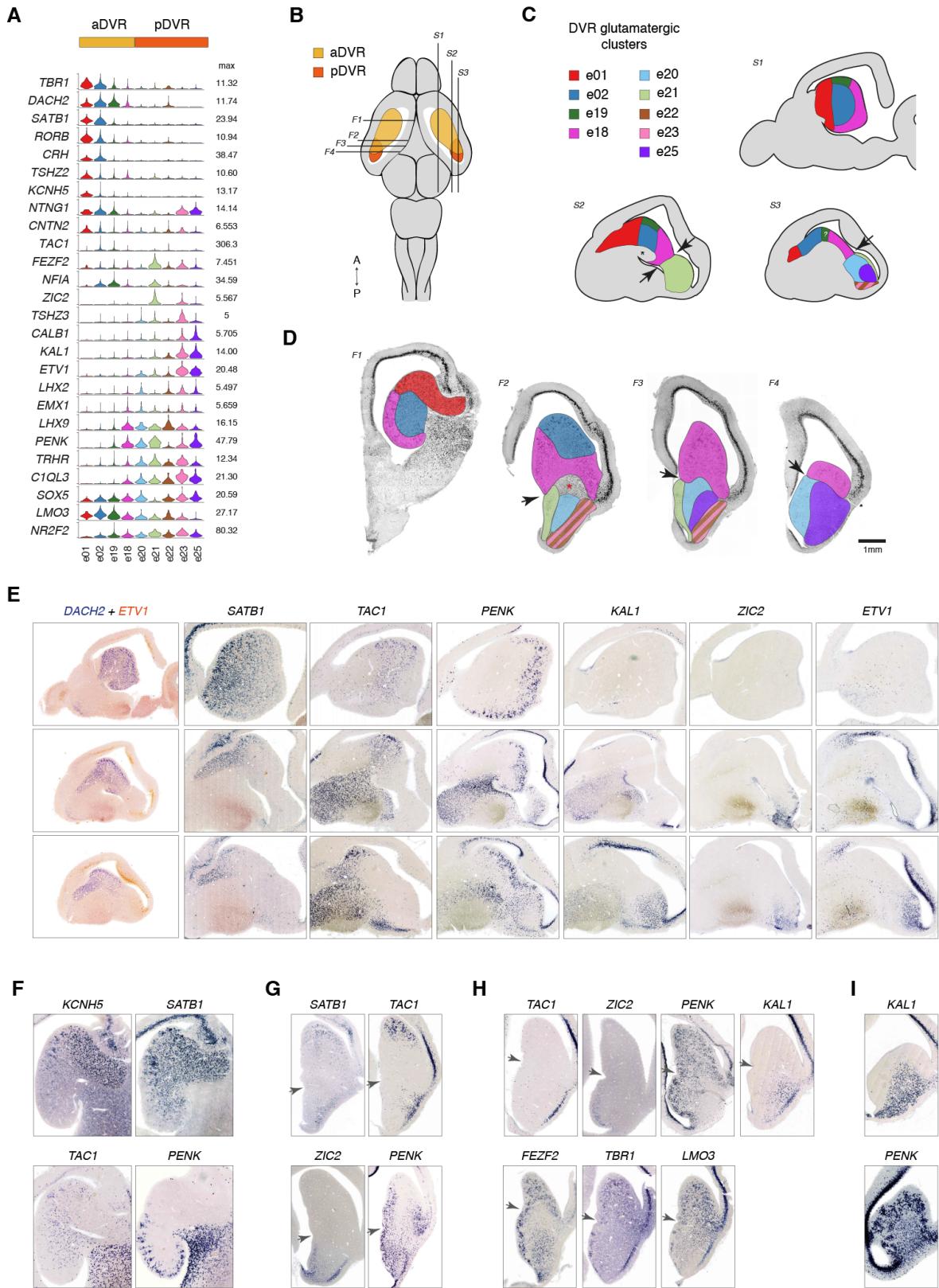


Fig. S10. The turtle Dorsal Ventricular Ridge (DVR).

- (A) Violin plot showing genes differentially expressed among DVR clusters, and used for mapping these clusters on anatomy.
- (B) Schematic representation of the turtle brain, with anterior and posterior DVR and frontal and sagittal sections.
- (C) Mapping of turtle DVR clusters (in (A)) on sagittal sections.
- (D) Mapping of turtle DVR clusters (in (A)) on frontal sections. The arrows in (C) and (D) indicate the sulcus separating aDVR and pDVR. The asterisk (*) labels an area occupied by *FOXP2*+ GABAergic neurons, see also fig. S16D.
- (E) Expression of markers of DVR clusters in sagittal sections. Sections are arranged from medial (top row) to lateral (bottom row), and correspond approximately to the schematics S1, S2 and S3 in (C). The first column shows double colorimetric ISHs for the aDVR marker *DACH2* (blue) and for *ETV1*, expressed in a subset of pDVR cells (orange).
- (F) Expression of markers of aDVR clusters in frontal sections, ~ level F1 in (D). *KCNH5* labels the visuorecipient area of the turtle aDVR.
- (G) Expression of markers of aDVR and pDVR clusters in frontal sections, ~ level F2 in (D). Arrows indicate the boundary between aDVR and pDVR.
- (H) Expression of markers of aDVR and pDVR clusters in frontal sections, ~ level F3 in (D).
- (I) Expression of markers of aDVR and pDVR clusters in frontal sections, ~level F4 in (D).

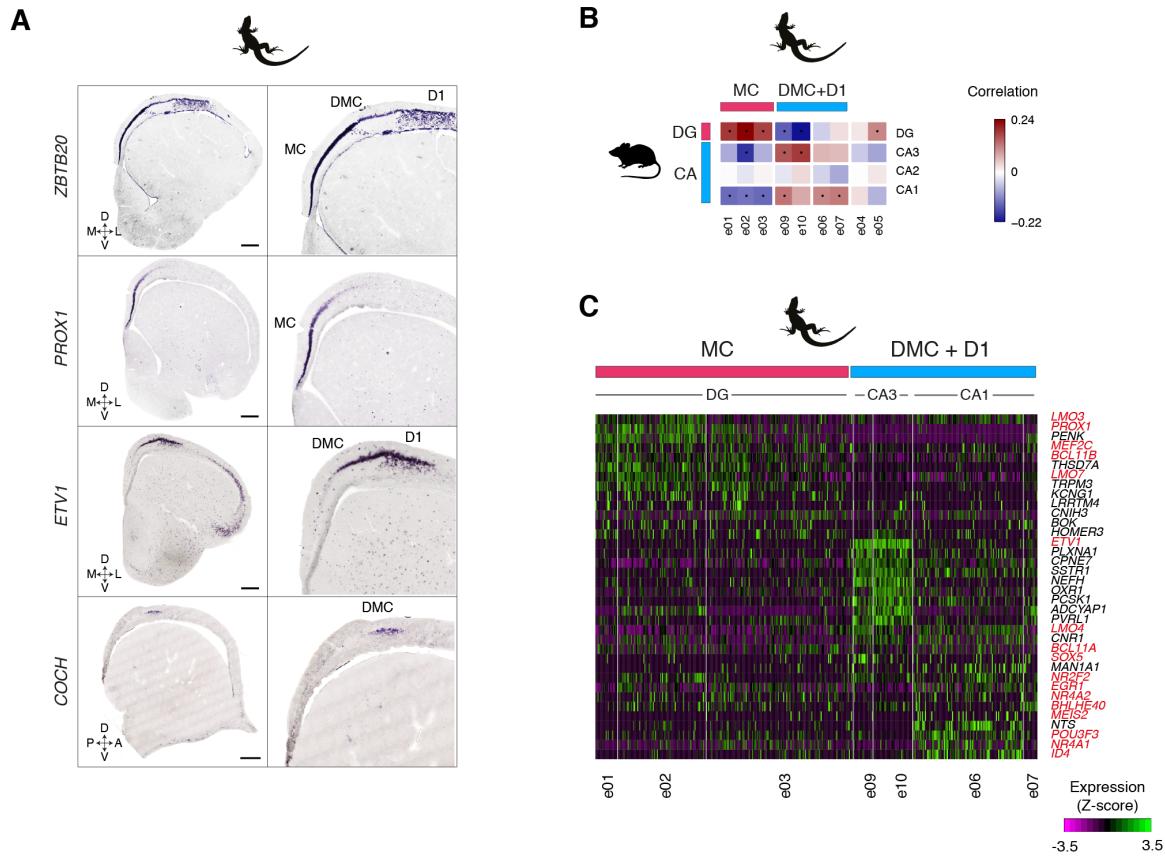


Fig. S11. Lizard hippocampus.

(A) *In situ* hybridizations showing expression of the pan-hippocampal marker *ZBTB20*, the DG marker *PROX1*, the CA marker *ETV1* and the CA3 marker *COCH* in lizard. Frontal sections for *ZBTB20*, *PROX1* and *ETV1*; sagittal section for *COCH*. Scale bar: 500μm.

(B) Pairwise correlations between glutamatergic clusters mapping to the lizard hippocampus and mouse hippocampal glutamatergic neurons. Mouse data from (19).

(C) Heatmap showing expression of markers differentially expressed in lizard hippocampal clusters. Names of transcription factors in red.

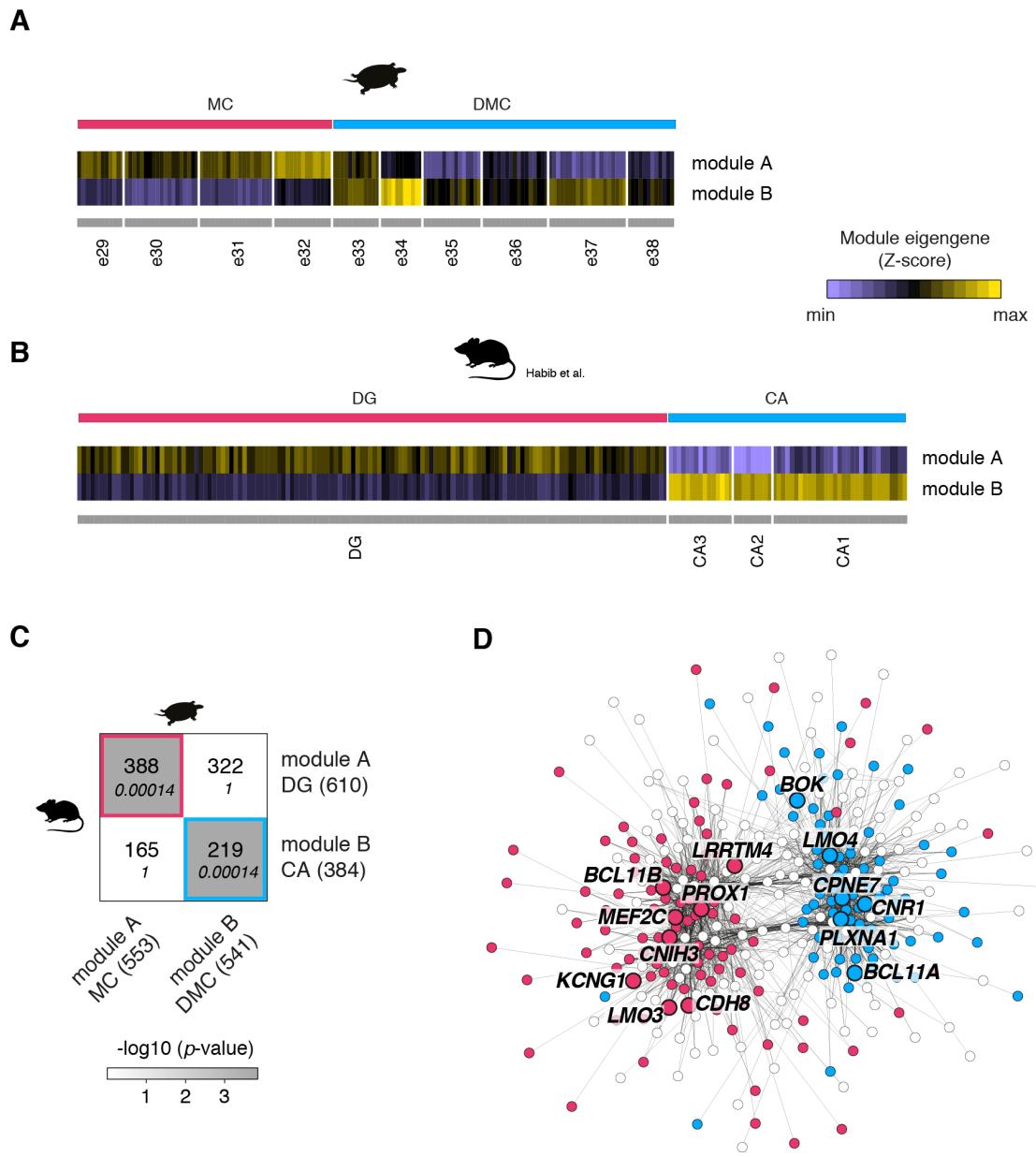


Fig. S12. Hippocampus gene network analysis.

- (A) Expression of module eigengenes (first principal component of genes modules, rows) calculated from turtle MC and DMC cells with WGCNA (see *Methods*).
- (B) Expression of module eigengenes of mouse hippocampal modules (rows) calculated from mouse DG and CA glutamatergic cells with WGCNA. Mouse expression data and clusters from (19).
- (C) Overlaps of turtle (columns) and mouse (rows) hippocampal gene modules. The table indicates the number of overlapping genes and *p*-values of overlaps (italics). Number of genes in each module in parentheses.
- (D) Illustration of the union of turtle and mouse gene networks. Nodes: genes, color-coded according to (C); edges: high-ranking gene-gene correlations.

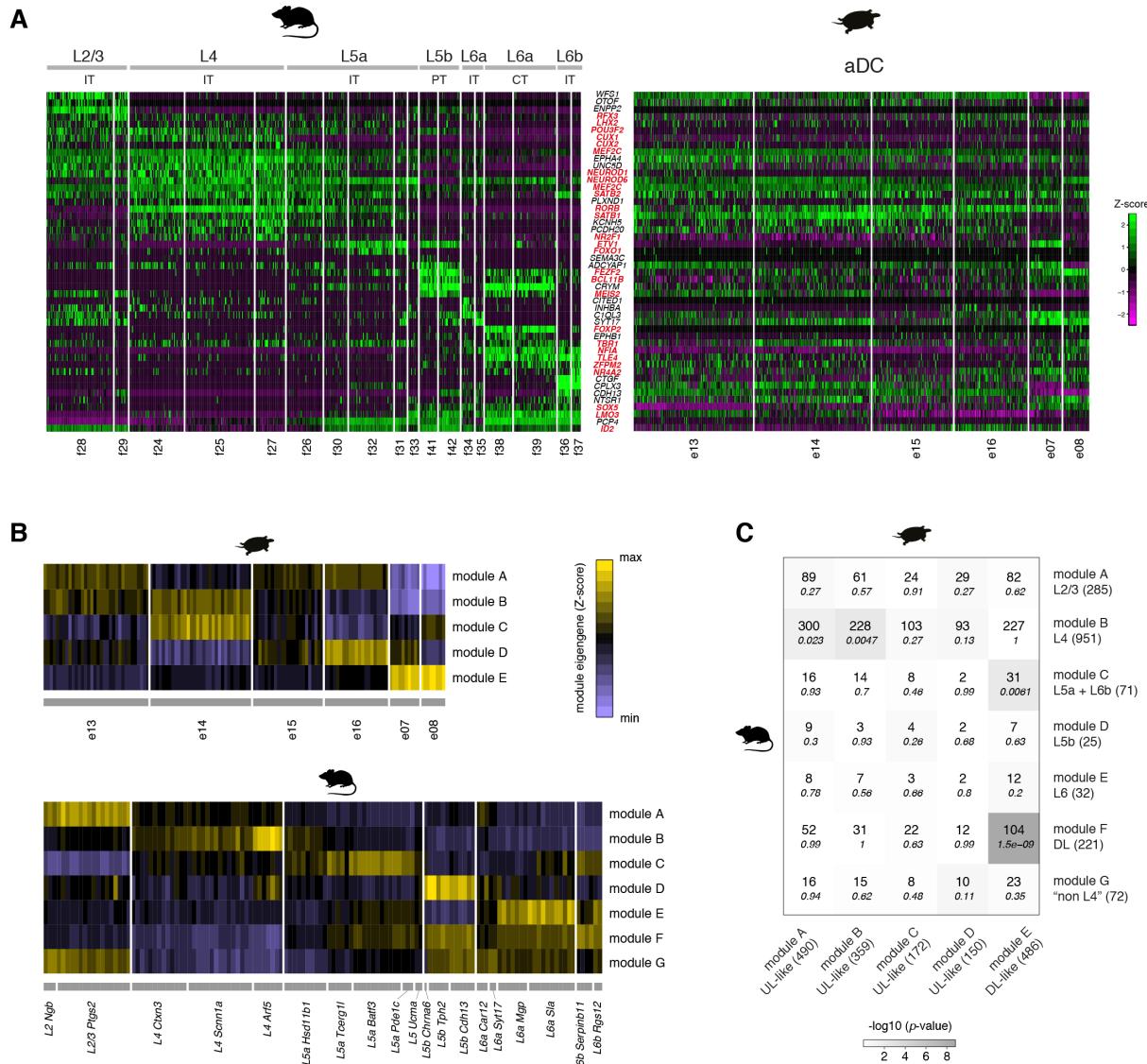


Fig. S13. Comparison of gene expression in turtle anterior dorsal cortex and mouse neocortex.

(A) Heatmap showing expression of markers (rows) of neocortical layers selected from the literature (6, 7, 25, 26, 35), in mouse visual cortex single cells and turtle anterior dorsal cortex. Mouse expression data and clusters from (7). Cells (columns) are grouped by cluster. Names of transcription factors in red. IT: intratelencephalic neurons, PT: pyramidal tract neurons, CT: corticothalamic neurons

(B) Top: expression of module eigengenes of gene modules (rows) obtained from turtle aDC cells with WGCNA. Bottom: expression of module eigengenes of gene modules (rows) calculated from mouse neocortical cells with WGCNA. Mouse data from (7).

(C) Gene overlaps between modules identified in mouse neocortex single cells and turtle aDC, and *p*-values indicating significance of overlaps (Fisher's exact test). Number of genes in each module in parentheses. DL: deep layers.

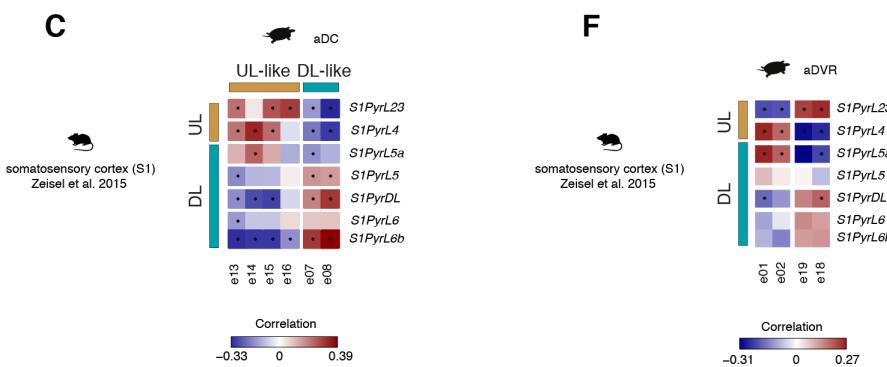
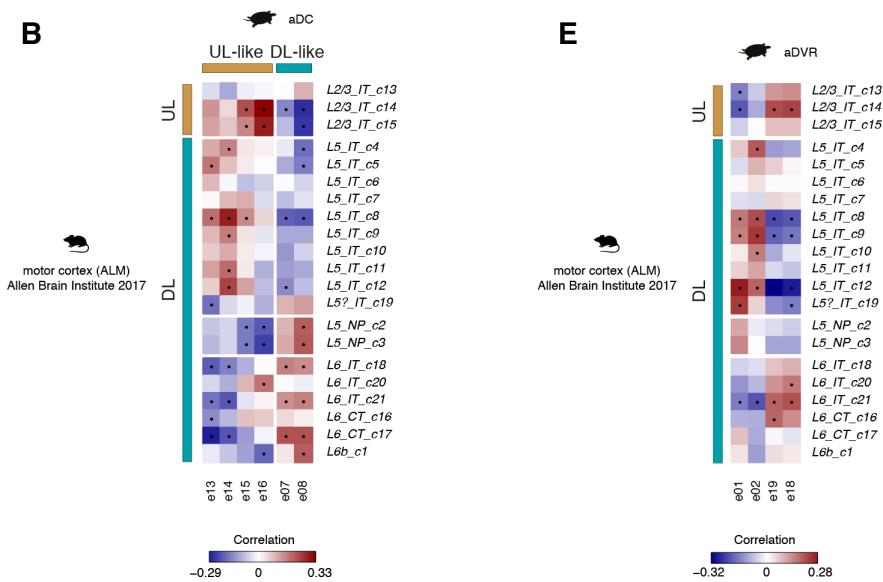
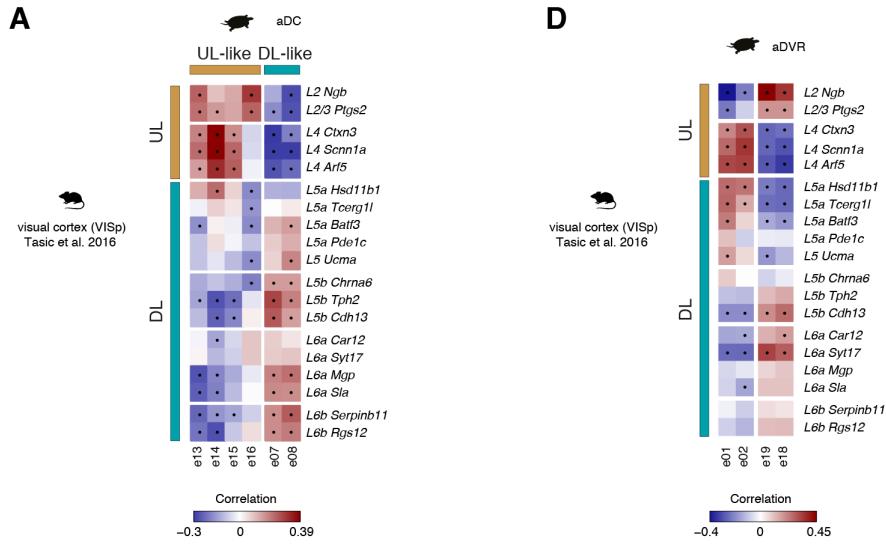


Fig. S14. Comparison of turtle anterior dorsal cortex (aDC) and anterior DVR clusters with mammalian single-cell transcriptomes from different cortical areas.

(A)-(C) Pairwise correlations of turtle anterior dorsal cortex (aDC, columns) and mouse neocortex (rows) glutamatergic types. Dots: statistically significant correlations. ((A) same as in Fig. 4C).

(D)-(F) Pairwise correlations of turtle anterior DVR (columns) and mouse neocortex (rows) glutamatergic types. Dots: statistically significant correlations.

Different mammalian datasets were used for these comparison: in (A) and (D), the Tasic et al 2016 visual cortex data (7); in (B) and (E), the Allen Brain Institute Anterior Lateral Motor cortex (ALM) dataset; in (C) and (F), the Zeisel et al 2015 somatosensory cortex dataset (6). See also *Methods*.

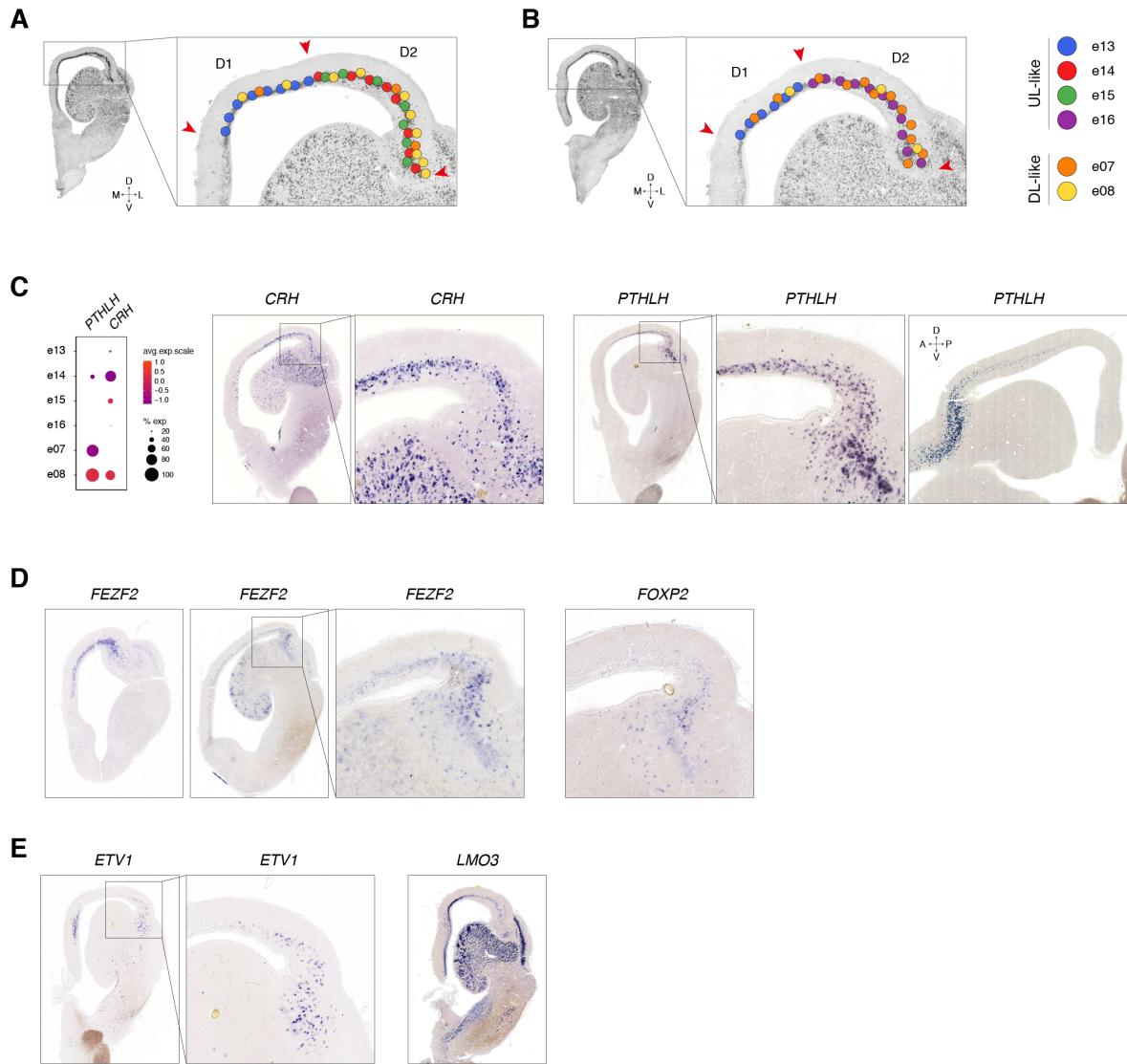


Fig. S15. Spatial distribution of glutamatergic neurons in turtle anterior dorsal cortex (aDC).

(A) and (B) Schematic representation of the spatial distribution of the six turtle aDC clusters. From the transcriptomics data, we selected markers specific of each of these clusters, or of groups of them (cfr. Fig. 4A and fig. S15C), performed *in situ* hybridizations on sections sampled at different rostro-caudal levels, and analyzed them according to known anatomical landmarks (see below). In these schematics, the medio-lateral subdivisions D1 and D2 are indicated; they can be distinguished by the thickness of layer 3. The *SOX5*-negative cluster e13 maps to D1 (cfr. Fig. 4B). The *CRH*⁺ cluster e14 labels a subset of cells in anterior D2 (see below). The *LMO3*-negative cluster e16 corresponds to cells in the posterior D2 (see Fig.4 and below). The DL-like cells e07 and e08, expressing the marker *PTHLH*, are more abundant in D2; *FOXP2*⁺ cells from cluster

e08 are preferentially found in anterior D2, whereas *ETV1*⁺ cells from cluster e07 are abundant in posterior D2. The color-code chosen here for the representation of aDC clusters is not related to the color-code used in Fig.4.

(C) *PTHLH* and *CRH* are expressed in scattered cells in aDC. Note that *PTHLH* is also expressed in the pallial thickening (PT) region.

(D) Expression of the DL-cells markers *FEZF2* and *FOXP2* in lateral anterior D2.

(E) Expression of the DL-cells marker *ETV1* in lateral posterior D2, and lack of *LMO3* expression in the same area.

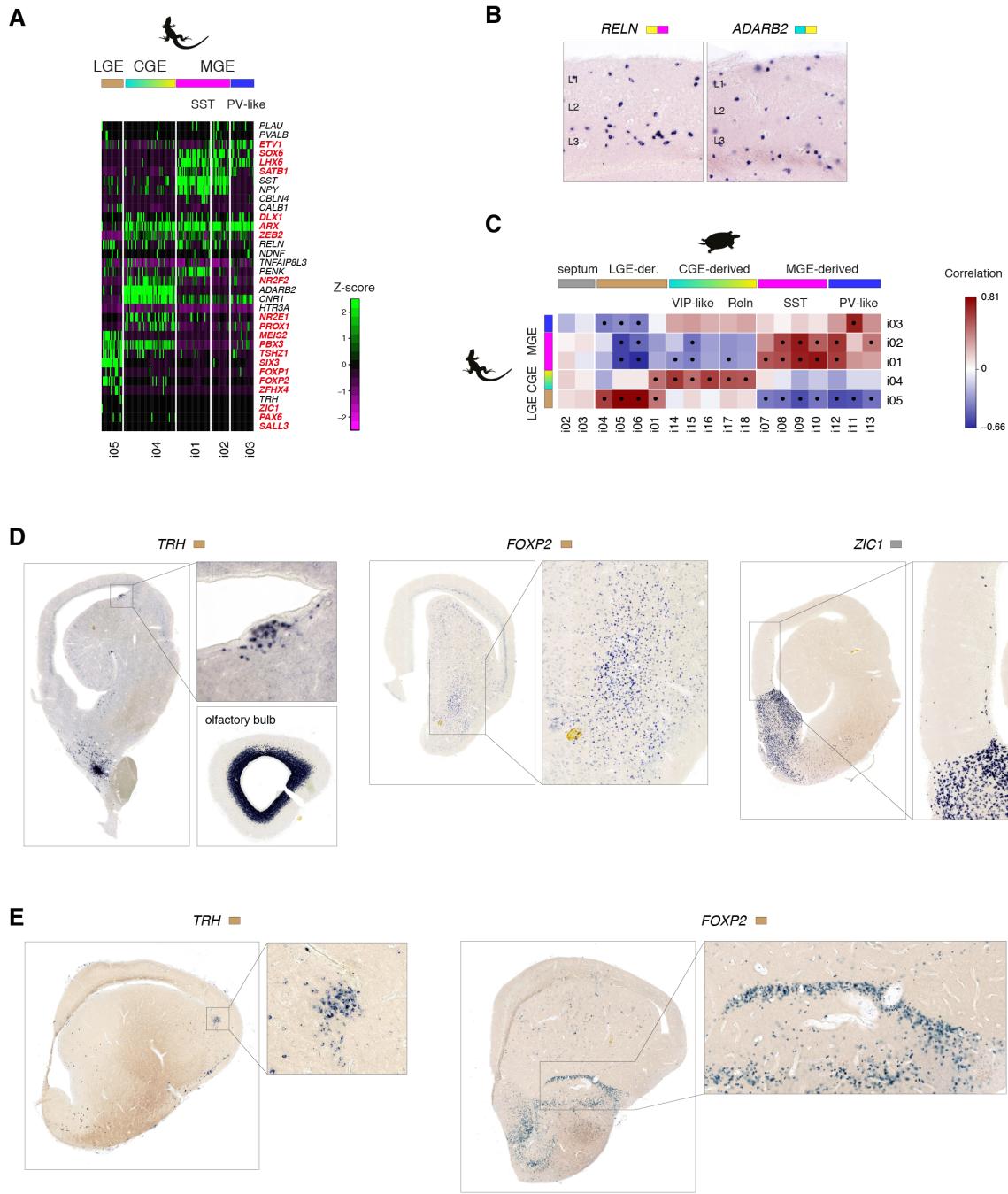


Fig. S16. Lizard GABAergic interneurons and reptilian LGE-derived GABAergic neurons

- (A) Heatmap showing expression of canonical mammalian GABAergic interneuron markers in lizard GABAergic clusters.
- (B) *In situ* hybridizations showing expression of *RELN* and *ADARB2* in the lizard dorsal cortex.
- (C) Pairwise correlations of turtle and lizard GABAergic interneuron types, calculated from differentially expressed genes.

(D) and (E): expression of markers of additional types of GABAergic neurons: *TRH*, *FOXP2* and *ZIC1* in turtle (D) and *TRH* and *FOXP2* in lizard (E). Besides the MGE- and CGE-derived GABAergic clusters described in Fig. 5, we found also GABAergic clusters made of cells accidentally dissected from the septum (i03) and striatum (i04: *MEIS2*+, *FOXP1*+, *CALB1*+, *TAC1*+, *PENK*+, *FOXP2*-). In addition, there are clusters resembling mammalian olfactory bulb (OB) interneurons by molecular identity (i02: septum-derived dopaminergic neurons *TH*+, *PAX6*+, *ZIC1*+, *SALL3*%; i01: LGE-derived *TRH*+ interneurons), even though the OB was not included in our dissections. ISH with cluster-specific genes revealed that i01 and i02 were sampled from the turtle rostral migratory stream (RMS), the path of newborn GABAergic neurons migrating to the OB (cfr. *TRH* in (D) and (E)). With a similar approach, we identified a putative reptilian counterpart of the mammalian intercalated cells of the amygdala (i04, i05: *MEIS2*+, *FOXP2*+, *TSHZ1*+, *SP8*+, *PBX3*+, *TAC1*-, *PENK*-) (cfr. *FOXP2* expression in (D) and (E), and Fig. 5A). Markers of LGE- and septal-derived neurons were never expressed in cortical layers (except for a few *ZIC1*+ cells in MC, (D) right). Taken together, this indicates that LGE- and septal-derived GABAergic neurons have similar molecular identities and migratory behaviors in reptiles and mammals.

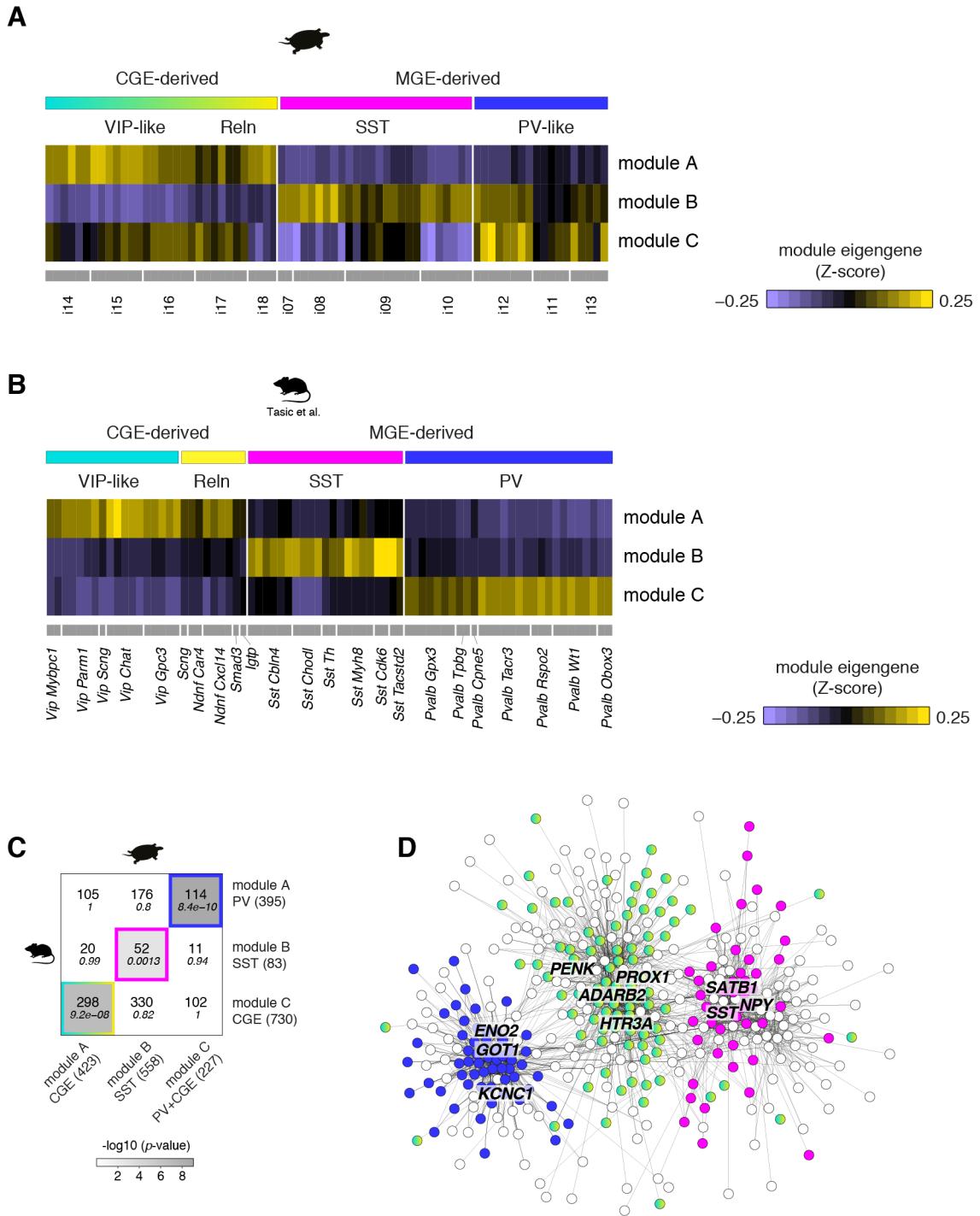


Fig. S17. GABAergic interneurons gene network analysis

- (A) Module eigengenes of gene modules (rows) calculated from turtle GABAergic cells with WGCNA.
- (B) Module eigengenes of gene modules (rows) calculated from mouse cortical GABAergic cells with WGCNA. Mouse data and clusters from (7).

(C) Overlaps of turtle (columns) and mouse (rows) GABAergic gene modules. The table indicates the number of overlapping genes and *p*-values of overlaps (italics). Number of genes in each module in parentheses.

(D) Illustration of the union of turtle and mouse gene networks. Nodes: genes, color-coded according to (C); edges: high-ranking gene-gene correlations.

Table S1. Brain dissections.

Table indicating the dissections from turtles (*Trachemys scripta elegans*) and lizards (*Pogona vitticeps*), including animal number, sample number, and number of cells in the final analysis (after filtering). See also fig. S7A.

Species	animal id	sample id	brain region	number of cells
<i>T. scripta</i>	8	1	cortex	842
<i>T. scripta</i>	8	2	cortex	746
<i>T. scripta</i>	11	1	cortex	983
<i>T. scripta</i>	11	2	cortex	1435
<i>T. scripta</i>	11	3	DVR	490
<i>T. scripta</i>	12	1	cortex	1306
<i>T. scripta</i>	12	2	cortex	1016
<i>T. scripta</i>	12	3	DVR	1892
<i>T. scripta</i>	15	1	DC	566
<i>T. scripta</i>	15	2	Cortex w/o DC	905
<i>T. scripta</i>	15	3	DVR	813
<i>T. scripta</i>	16	1	DC	751
<i>T. scripta</i>	16	2	Posterior Cortex	1713
<i>T. scripta</i>	16	3	MC + DMC	423
<i>T. scripta</i>	16	4	DVR	868
<i>T. scripta</i>	17	1	DC + MC + DMC	936
<i>T. scripta</i>	17	2	Posterior Cortex	1998
<i>T. scripta</i>	17	3	DVR	1145
<i>P. vitticeps</i>	9	1	cortex	654
<i>P. vitticeps</i>	19	1	Cortex	1177
<i>P. vitticeps</i>	20	1	Cortex + DVR	202
<i>P. vitticeps</i>	21	1	Cortex + DVR	675
<i>P. vitticeps</i>	22	1	Cortex + DVR	193
<i>P. vitticeps</i>	23	1	Cortex + DVR	598
<i>P. vitticeps</i>	24	1	Cortex + DVR	357
<i>P. vitticeps</i>	25	1	Cortex + DVR	331

Table S2. Drop-seq libraries.

The table reports the correspondences between animals, samples, sequencing libraries, and number of cells used for analysis after filtering.

Drop-seq library	index	species	animal id	sample id	brain region	Number of cells	Sequencing Platform
d2	N701	<i>T. scripta</i>	8	1	cortex	resequenced in Library4 N704	HiSeq
d2	N702	<i>T. scripta</i>	8	1	cortex	resequenced in Library4 N701	HiSeq
d2	N703	<i>T. scripta</i>	8	1	cortex	resequenced in Library4 N702	HiSeq
d2	N704	<i>T. scripta</i>	8	1	cortex	resequenced in Library4 N703	HiSeq
d2	N705	<i>T. scripta</i>	8	1	cortex	resequenced in Library4 N704	HiSeq
d2	N706	<i>T. scripta</i>	8	2	cortex	resequenced in Library4 N705	HiSeq
d2	N707	<i>T. scripta</i>	8	2	cortex	resequenced in Library4 N705	HiSeq
d2	N710	<i>P. vitticeps</i>	9	1	cortex	resequenced in Library4 N706	HiSeq
d2	N711	<i>P. vitticeps</i>	9	1	cortex	resequenced in Library4 N706	HiSeq
d3	N701	<i>T. scripta</i>	8	1	cortex	123	NextSeq
d3	N702	<i>T. scripta</i>	8	2	cortex	401	NextSeq
d3	N703	<i>T. scripta</i>	8	2	cortex	333	NextSeq
d3	N704	<i>T. scripta</i>	11	1	cortex	385	NextSeq
d3	N705	<i>T. scripta</i>	11	1	cortex	335	NextSeq
d3	N706	<i>T. scripta</i>	11	1	cortex	346	NextSeq
d3	N707	<i>T. scripta</i>	11	2	cortex	483	NextSeq
d3	N710	<i>T. scripta</i>	11	2	cortex	433	NextSeq
d3	N711	<i>T. scripta</i>	11	2	cortex	683	NextSeq
d3	N712	<i>T. scripta</i>	11	3	DVR	386	NextSeq
d3	N714	<i>T. scripta</i>	11	3	DVR	266	NextSeq
d3	N715	<i>P. vitticeps</i>	9	1	cortex	164	NextSeq
d4	N701	<i>T. scripta</i>	8	1	cortex	202	NextSeq
d4	N702	<i>T. scripta</i>	8	1	cortex	190	NextSeq
d4	N703	<i>T. scripta</i>	8	1	cortex	199	NextSeq
d4	N704	<i>T. scripta</i>	8	1	cortex	236	NextSeq
d4	N705	<i>T. scripta</i>	8	2	cortex	84	NextSeq
d4	N706	<i>P. vitticeps</i>	9	1	cortex	632	NextSeq
d4	N707	<i>T. scripta</i>	12	1	cortex	738	NextSeq
d4	N710	<i>T. scripta</i>	12	1	cortex	724	NextSeq
d4	N711	<i>T. scripta</i>	12	2	cortex	578	NextSeq
d4	N712	<i>T. scripta</i>	12	2	cortex	557	NextSeq
d4	N714	<i>T. scripta</i>	12	3	DVR	490	NextSeq
d4	N715	HEK/BRL	n/a	n/a	n/a	100	NextSeq
d5	N701	<i>T. scripta</i>	12	3	DVR	397	NextSeq
d5	N702	<i>T. scripta</i>	12	3	DVR	410	NextSeq
d5	N703	<i>T. scripta</i>	12	3	DVR	392	NextSeq

d5	N704	<i>T. scripta</i>	12	3	DVR	458	NextSeq
d5	N705	<i>T. scripta</i>	12	3	DVR	330	NextSeq
d5	N706	<i>T. scripta</i>	15	1	DC	331	NextSeq
d5	N707	<i>T. scripta</i>	15	1	DC	259	NextSeq
d5	N710	<i>T. scripta</i>	15	2	Cortex w/o DC	257	NextSeq
d5	N711	<i>T. scripta</i>	15	2	Cortex w/o DC	347	NextSeq
d5	N712	<i>T. scripta</i>	15	2	Cortex w/o DC	372	NextSeq
d5	N714	<i>T. scripta</i>	15	3	DVR	476	NextSeq
d5	N715	<i>T. scripta</i>	15	3	DVR	439	NextSeq
d6	N701	<i>T. scripta</i>	16	1	DC	301	NextSeq
d6	N702	<i>T. scripta</i>	16	1	DC	495	NextSeq
d6	N703	<i>T. scripta</i>	16	2	Posterior Cortex	275	NextSeq
d6	N704	<i>T. scripta</i>	16	2	Posterior Cortex	192	NextSeq
d6	N705	<i>T. scripta</i>	16	3	MC + DMC	224	NextSeq
d6	N706	<i>T. scripta</i>	16	3	MC + DMC	233	NextSeq
d6	N707	<i>T. scripta</i>	16	4	DVR	465	NextSeq
d6	N710	<i>T. scripta</i>	16	4	DVR	525	NextSeq
d6	N711	<i>T. scripta</i>	17	1	DC + MC + DMC	433	NextSeq
d6	N712	<i>T. scripta</i>	17	1	DC + MC + DMC	579	NextSeq
d6	N714	<i>T. scripta</i>	17	3	DVR	617	NextSeq
d6	N715	<i>T. scripta</i>	17	3	DVR	595	NextSeq
d7	N701	<i>T. scripta</i>	16	2	Posterior Cortex	479	NextSeq
d7	N702	<i>T. scripta</i>	16	2	Posterior Cortex	537	NextSeq
d7	N703	<i>T. scripta</i>	16	2	Posterior Cortex	584	NextSeq
d7	N704	<i>T. scripta</i>	17	2	Posterior Cortex	516	NextSeq
d7	N705	<i>T. scripta</i>	17	2	Posterior Cortex	495	NextSeq
d7	N706	<i>T. scripta</i>	17	2	Posterior Cortex	481	NextSeq
d7	N707	<i>T. scripta</i>	17	2	Posterior Cortex	422	NextSeq
d7	N710	<i>T. scripta</i>	17	2	Posterior Cortex	335	NextSeq
d7	N711	<i>P. vitticeps</i>	19	1	Cortex	209	NextSeq
d7	N712	<i>P. vitticeps</i>	19	1	Cortex	282	NextSeq
d7	N714	<i>P. vitticeps</i>	19	1	Cortex	237	NextSeq
d7	N715	<i>P. vitticeps</i>	19	1	Cortex	310	NextSeq
d8	N701	<i>P. vitticeps</i>	19	1	Cortex	539	NextSeq
d8	N702	<i>P. vitticeps</i>	20	1	Cortex + DVR	194	NextSeq
d8	N703	<i>P. vitticeps</i>	20	1	Cortex + DVR	213	NextSeq
d8	N704	<i>P. vitticeps</i>	21	1	Cortex + DVR	264	NextSeq
d8	N705	<i>P. vitticeps</i>	21	1	Cortex + DVR	219	NextSeq
d8	N706	<i>P. vitticeps</i>	21	1	Cortex + DVR	423	NextSeq
d8	N707	<i>P. vitticeps</i>	22	1	Cortex + DVR	216	NextSeq
d8	N710	<i>P. vitticeps</i>	23	1	Cortex + DVR	371	NextSeq

d8	N711	<i>P. vitticeps</i>	23	1	Cortex + DVR	443	NextSeq
d8	N712	<i>P. vitticeps</i>	24	1	Cortex + DVR	162	NextSeq
d8	N714	<i>P. vitticeps</i>	24	1	Cortex + DVR	259	NextSeq
d8	N715	<i>P. vitticeps</i>	25	1	Cortex + DVR	422	NextSeq

References

1. L. Puelles *et al.*, in *Evolution of Nervous Systems*, J. H. Kaas, Ed. (Elsevier, ed. 2, 2017), vol. 1, pp. 519–555.
2. G. F. Striedter, The telencephalon of tetrapods in evolution. *Brain Behav. Evol.* **49**, 179–194 (1997). [doi:10.1159/000112991](https://doi.org/10.1159/000112991) [Medline](#)
3. G. F. Striedter, Evolution of the hippocampus in reptiles and birds. *J. Comp. Neurol.* **524**, 496–517 (2016). [doi:10.1002/cne.23803](https://doi.org/10.1002/cne.23803) [Medline](#)
4. T. G. Belgard, J. F. Montiel, W. Z. Wang, F. García-Moreno, E. H. Margulies, C. P. Ponting, Z. Molnár, Adult pallium transcriptomes surprise in not reflecting predicted homologies across diverse chicken and mouse pallial sectors. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 13150–13155 (2013). [doi:10.1073/pnas.1307444110](https://doi.org/10.1073/pnas.1307444110) [Medline](#)
5. J. Dugas-Ford, J. J. Rowell, C. W. Ragsdale, Cell-type homologies and the origins of the neocortex. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 16974–16979 (2012). [doi:10.1073/pnas.1204773109](https://doi.org/10.1073/pnas.1204773109) [Medline](#)
6. A. Zeisel, A. B. Muñoz-Manchado, S. Codeluppi, P. Lönnberg, G. La Manno, A. Juréus, S. Marques, H. Munguba, L. He, C. Betsholtz, C. Rolny, G. Castelo-Branco, J. Hjerling-Leffler, S. Linnarsson, Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138–1142 (2015). [doi:10.1126/science.aaa1934](https://doi.org/10.1126/science.aaa1934) [Medline](#)
7. B. Tasic, V. Menon, T. N. Nguyen, T. K. Kim, T. Jarsky, Z. Yao, B. Levi, L. T. Gray, S. A. Sorensen, T. Dolbeare, D. Bertagnolli, J. Goldy, N. Shapovalova, S. Parry, C. Lee, K. Smith, A. Bernard, L. Madisen, S. M. Sunkin, M. Hawrylycz, C. Koch, H. Zeng, Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat. Neurosci.* **19**, 335–346 (2016). [doi:10.1038/nn.4216](https://doi.org/10.1038/nn.4216) [Medline](#)
8. E. Z. Macosko, A. Basu, R. Satija, J. Nemesh, K. Shekhar, M. Goldman, I. Tirosh, A. R. Bialas, N. Kamitaki, E. M. Martersteck, J. J. Trombetta, D. A. Weitz, J. R. Sanes, A. K. Shalek, A. Regev, S. A. McCarroll, Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015). [doi:10.1016/j.cell.2015.05.002](https://doi.org/10.1016/j.cell.2015.05.002) [Medline](#)
9. G. La Manno, D. Gyllborg, S. Codeluppi, K. Nishimura, C. Salto, A. Zeisel, L. E. Borm, S. R. W. Stott, E. M. Toledo, J. C. Villaescusa, P. Lönnberg, J. Ryge, R. A. Barker, E. Arenas, S. Linnarsson, Molecular diversity of midbrain development in mouse, human, and stem cells. *Cell* **167**, 566–580.e19 (2016). [doi:10.1016/j.cell.2016.09.027](https://doi.org/10.1016/j.cell.2016.09.027) [Medline](#)
10. H. J. Karten, Neocortical evolution: Neuronal circuits arise independently of lamination. *Curr. Biol.* **23**, R12–R15 (2013). [doi:10.1016/j.cub.2012.11.013](https://doi.org/10.1016/j.cub.2012.11.013) [Medline](#)
11. A. B. Butler, Z. Molnár, Development and evolution of the collopallium in amniotes: A new hypothesis of field homology. *Brain Res. Bull.* **57**, 475–479 (2002). [doi:10.1016/S0361-9230\(01\)00679-7](https://doi.org/10.1016/S0361-9230(01)00679-7) [Medline](#)
12. L. L. Bruce, T. J. Neary, The limbic system of tetrapods: A comparative analysis of cortical and amygdalar populations. *Brain Behav. Evol.* **46**, 224–234 (1995). [doi:10.1159/000113276](https://doi.org/10.1159/000113276) [Medline](#)

13. M. Hawrylycz, J. A. Miller, V. Menon, D. Feng, T. Dolbeare, A. L. Guillozet-Bongaarts, A. G. Jegga, B. J. Aronow, C.-K. Lee, A. Bernard, M. F. Glasser, D. L. Dierker, J. Menche, A. Szafer, F. Collman, P. Grange, K. A. Berman, S. Mihalas, Z. Yao, L. Stewart, A.-L. Barabási, J. Schulkin, J. Phillips, L. Ng, C. Dang, D. R. Haynor, A. Jones, D. C. Van Essen, C. Koch, E. Lein, Canonical genetic signatures of the adult human brain. *Nat. Neurosci.* **18**, 1832–1844 (2015). [doi:10.1038/nn.4171](https://doi.org/10.1038/nn.4171) [Medline](#)
14. P. V. Hoogland, E. Vermeulen-Vanderzee, Efferent connections of the dorsal cortex of the lizard *Gekko gecko* studied with *Phaseolus vulgaris*-leucoagglutinin. *J. Comp. Neurol.* **285**, 289–303 (1989). [doi:10.1002/cne.902850302](https://doi.org/10.1002/cne.902850302) [Medline](#)
15. H. Yoneshima, S. Yamasaki, C. C. J. Voelker, Z. Molnár, E. Christophe, E. Audinat, M. Takemoto, M. Nishiwaki, S. Tsuji, I. Fujita, N. Yamamoto, Er81 is expressed in a subpopulation of layer 5 neurons in rodent and primate neocortices. *Neuroscience* **137**, 401–412 (2006). [doi:10.1016/j.neuroscience.2005.08.075](https://doi.org/10.1016/j.neuroscience.2005.08.075) [Medline](#)
16. N. Moreno, A. González, Evolution of the amygdaloid complex in vertebrates, with special reference to the anamniote-amniotic transition. *J. Anat.* **211**, 151–163 (2007). [doi:10.1111/j.1469-7580.2007.00780.x](https://doi.org/10.1111/j.1469-7580.2007.00780.x) [Medline](#)
17. P. R. Manger, D. A. Slutsky, Z. Molnár, Visual subdivisions of the dorsal ventricular ridge of the iguana (*Iguana iguana*) as determined by electrophysiologic mapping. *J. Comp. Neurol.* **453**, 226–246 (2002). [doi:10.1002/cne.10373](https://doi.org/10.1002/cne.10373) [Medline](#)
18. L. Medina, A. Abellán, E. Desfilis, Contribution of genoarchitecture to understanding hippocampal evolution and development. *Brain. Behav. Evol.* **90**, 25–40 (2017). [doi:10.1159/000477558](https://doi.org/10.1159/000477558) [Medline](#)
19. N. Habib, Y. Li, M. Heidenreich, L. Swiech, I. Avraham-David, J. J. Trombetta, C. Hession, F. Zhang, A. Regev, Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons. *Science* **353**, 925–928 (2016). [doi:10.1126/science.aad7038](https://doi.org/10.1126/science.aad7038) [Medline](#)
20. J. M. Shen, A. R. Kriegstein, Turtle hippocampal cortex contains distinct cell types, burst-firing neurons, and an epileptogenic subfield. *J. Neurophysiol.* **56**, 1626–1649 (1986). [doi:10.1152/jn.1986.56.6.1626](https://doi.org/10.1152/jn.1986.56.6.1626) [Medline](#)
21. M. S. Cembrowski, L. Wang, K. Sugino, B. C. Shields, N. Spruston, Hipposeq: A comprehensive RNA-seq database of gene expression in hippocampal principal neurons. *eLife* **5**, e14997 (2016). [doi:10.7554/eLife.14997](https://doi.org/10.7554/eLife.14997) [Medline](#)
22. S. Reiter, H.-P. Liaw, T. M. Yamawaki, R. K. Naumann, G. Laurent, On the value of reptilian brains to map the evolution of the hippocampal formation. *Brain. Behav. Evol.* **90**, 41–52 (2017). [doi:10.1159/000478693](https://doi.org/10.1159/000478693) [Medline](#)
23. F. F. Ebner, in *Evolution of Brain and Behavior in Vertebrates*, R. B. Masterton, M. E. Bitterman, C. B. G. Campbell, N. Hotton, Eds. (Lawrence Erlbaum Associates, Hillsdale, NJ, 1976), pp. 115–167.
24. A. Reiner, Neurotransmitter organization and connections of turtle cortex: Implications for the evolution of mammalian isocortex. *Comp. Biochem. Physiol. Comp. Physiol.* **104**, 735–748 (1993). [doi:10.1016/0300-9629\(93\)90149-X](https://doi.org/10.1016/0300-9629(93)90149-X) [Medline](#)
25. B. J. Molyneaux, P. Arlotta, J. R. L. Menezes, J. D. Macklis, Neuronal subtype specification

- in the cerebral cortex. *Nat. Rev. Neurosci.* **8**, 427–437 (2007). [doi:10.1038/nrn2151](https://doi.org/10.1038/nrn2151) [Medline](#)
26. L. T. Gray, Z. Yao, T. N. Nguyen, T. K. Kim, H. Zeng, B. Tasic, Layer-specific chromatin accessibility landscapes reveal regulatory networks in adult mouse visual cortex. *eLife* **6**, 1–30 (2017). [doi:10.7554/eLife.21883](https://doi.org/10.7554/eLife.21883) [Medline](#)
 27. P. S. Ulinski, Organization of corticogeniculate projections in the turtle, *Pseudemys scripta*. *J. Comp. Neurol.* **254**, 529–542 (1986). [doi:10.1002/cne.902540406](https://doi.org/10.1002/cne.902540406) [Medline](#)
 28. M. G. Blanton, A. R. Kriegstein, Morphological differentiation of distinct neuronal classes in embryonic turtle cerebral cortex. *J. Comp. Neurol.* **310**, 558–570 (1991). [doi:10.1002/cne.903100405](https://doi.org/10.1002/cne.903100405) [Medline](#)
 29. D. Arendt, J. M. Musser, C. V. H. Baker, A. Bergman, C. Cepko, D. H. Erwin, M. Pavlicev, G. Schlosser, S. Widder, M. D. Laubichler, G. P. Wagner, The origin and evolution of cell types. *Nat. Rev. Genet.* **17**, 744–757 (2016). [doi:10.1038/nrg.2016.127](https://doi.org/10.1038/nrg.2016.127) [Medline](#)
 30. R. Tremblay, S. Lee, B. Rudy, GABAergic interneurons in the neocortex: From cellular properties to circuits. *Neuron* **91**, 260–292 (2016). [doi:10.1016/j.neuron.2016.06.033](https://doi.org/10.1016/j.neuron.2016.06.033) [Medline](#)
 31. C. Métin, C. Alvarez, D. Moudoux, T. Vitalis, C. Pieau, Z. Molnár, Conserved pattern of tangential neuronal migration during forebrain development. *Development* **134**, 2815–2827 (2007). [doi:10.1242/dev.02869](https://doi.org/10.1242/dev.02869) [Medline](#)
 32. O. Marín, J. L. R. Rubenstein, A long, remarkable journey: Tangential migration in the telencephalon. *Nat. Rev. Neurosci.* **2**, 780–790 (2001). [doi:10.1038/35097509](https://doi.org/10.1038/35097509) [Medline](#)
 33. T. Nomura, H. Gotoh, K. Ono, Changes in the regulation of cortical neurogenesis contribute to encephalization during amniote brain evolution. *Nat. Commun.* **4**, 2206 (2013). [doi:10.1038/ncomms3206](https://doi.org/10.1038/ncomms3206) [Medline](#)
 34. A. M. Goffinet, C. Daumerie, B. Langerwerf, C. Pieau, Neurogenesis in reptilian cortical structures: ^3H -thymidine autoradiographic analysis. *J. Comp. Neurol.* **243**, 106–116 (1986). [doi:10.1002/cne.902430109](https://doi.org/10.1002/cne.902430109) [Medline](#)
 35. S. Lodato, P. Arlotta, Generating neuronal diversity in the mammalian cerebral cortex. *Annu. Rev. Cell Dev. Biol.* **31**, 699–720 (2015). [doi:10.1146/annurev-cellbio-100814-125353](https://doi.org/10.1146/annurev-cellbio-100814-125353) [Medline](#)
 36. G. M. Shepherd, The microcircuit concept applied to cortical evolution: From three-layer to six-layer cortex. *Front. Neuroanat.* **5**, 30 (2011). [doi:10.3389/fnana.2011.00030](https://doi.org/10.3389/fnana.2011.00030) [Medline](#)
 37. S. M. Suryanarayana, B. Robertson, P. Wallén, S. Grillner, The lamprey pallium provides a blueprint of the mammalian layered cortex. *Curr. Biol.* **27**, 3264–3277.e5 (2017). [doi:10.1016/j.cub.2017.09.034](https://doi.org/10.1016/j.cub.2017.09.034) [Medline](#)
 38. J. L. Ferran *et al.*, in *In Situ Hybridization Methods*, G. Hauptmann, Ed. (Springer, 2015), vol. 99, pp. 83–107.
 39. H. B. Shaffer, P. Minx, D. E. Warren, A. M. Shedlock, R. C. Thomson, N. Valenzuela, J. Abramyan, C. T. Amemiya, D. Badenhorst, K. K. Biggar, G. M. Borchert, C. W. Botka, R. M. Bowden, E. L. Braun, A. M. Bronikowski, B. G. Bruneau, L. T. Buck, B. Capel, T.

- A. Castoe, M. Czerwinski, K. D. Delehaunty, S. V. Edwards, C. C. Fronick, M. K. Fujita, L. Fulton, T. A. Graves, R. E. Green, W. Haerty, R. Hariharan, O. Hernandez, L. W. Hillier, A. K. Holloway, D. Janes, F. J. Janzen, C. Kandoth, L. Kong, A. P. de Koning, Y. Li, R. Literman, S. E. McGaugh, L. Mork, M. O’Laughlin, R. T. Paitz, D. D. Pollock, C. P. Ponting, S. Radhakrishnan, B. J. Raney, J. M. Richman, J. St John, T. Schwartz, A. Sethuraman, P. Q. Spinks, K. B. Storey, N. Thane, T. Vinar, L. M. Zimmerman, W. C. Warren, E. R. Mardis, R. K. Wilson, The western painted turtle genome, a model for the evolution of extreme physiological adaptations in a slowly evolving lineage. *Genome Biol.* **14**, R28 (2013). [doi:10.1186/gb-2013-14-3-r28](https://doi.org/10.1186/gb-2013-14-3-r28) [Medline](#)
40. A. Georges, Q. Li, J. Lian, D. O’Meally, J. Deakin, Z. Wang, P. Zhang, M. Fujita, H. R. Patel, C. E. Holleley, Y. Zhou, X. Zhang, K. Matsubara, P. Waters, J. A. M. Graves, S. D. Sarre, G. Zhang, High-coverage sequencing and annotated assembly of the genome of the Australian dragon lizard *Pogona vitticeps*. *Gigascience* **4**, 45 (2015). [doi:10.1186/s13742-015-0085-2](https://doi.org/10.1186/s13742-015-0085-2) [Medline](#)
41. B. K. Zajac, J. Amendt, R. Horres, M. A. Verhoff, R. Zehner, De novo transcriptome analysis and highly sensitive digital gene expression profiling of *Calliphora vicina* (Diptera: Calliphoridae) pupae using MACE (Massive Analysis of cDNA Ends). *Forensic Sci. Int. Genet.* **15**, 137–146 (2015). [doi:10.1016/j.fsigen.2014.11.013](https://doi.org/10.1016/j.fsigen.2014.11.013) [Medline](#)
42. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013). [doi:10.1093/bioinformatics/bts635](https://doi.org/10.1093/bioinformatics/bts635) [Medline](#)
43. S. Müller, L. Rycak, F. Afonso-Grunz, P. Winter, A. M. Zawada, E. Damrath, J. Scheider, J. Schmäh, I. Koch, G. Kahl, B. Rotter, APADB: A database for alternative polyadenylation and microRNA regulation events. *Database (Oxford)* **2014**, bau076 (2014). [doi:10.1093/database/bau076](https://doi.org/10.1093/database/bau076) [Medline](#)
44. K. Shekhar, S. W. Lapan, I. E. Whitney, N. M. Tran, E. Z. Macosko, M. Kowalczyk, X. Adiconis, J. Z. Levin, J. Nemesh, M. Goldman, S. A. McCarroll, C. L. Cepko, A. Regev, J. R. Sanes, Comprehensive classification of retinal bipolar neurons by single-cell transcriptomics. *Cell* **166**, 1308–1323.e30 (2016). [doi:10.1016/j.cell.2016.07.054](https://doi.org/10.1016/j.cell.2016.07.054) [Medline](#)
45. W. E. Johnson, C. Li, A. Rabinovic, Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007). [doi:10.1093/biostatistics/kxj037](https://doi.org/10.1093/biostatistics/kxj037) [Medline](#)
46. R Core Team, R: A language and environment for statistical computing (R Foundation, Vienna, 2012); www.R-project.org/
47. R. Satija, J. A. Farrell, D. Gennert, A. F. Schier, A. Regev, Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015). [doi:10.1038/nbt.3192](https://doi.org/10.1038/nbt.3192) [Medline](#)
48. L. J. P. Van Der Maaten, G. E. Hinton, Visualizing high-dimensional data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
49. L. Waltman, N. J. Van Eck, A smart local moving algorithm for large-scale modularity-based community detection. *Eur. Phys. J. B* **86**, 471 (2013). [doi:10.1140/epjb/e2013-40829-0](https://doi.org/10.1140/epjb/e2013-40829-0)

50. M. Rosvall, C. T. Bergstrom, Maps of random walks on complex networks reveal community structure. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 1118–1123 (2008). [doi:10.1073/pnas.0706851105](https://doi.org/10.1073/pnas.0706851105) [Medline](#)
51. G. Finak, A. McDavid, M. Yajima, J. Deng, V. Gersuk, A. K. Shalek, C. K. Slichter, H. W. Miller, M. J. McElrath, M. Prlic, P. S. Linsley, R. Gottardo, MAST: A flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 278 (2015). [doi:10.1186/s13059-015-0844-5](https://doi.org/10.1186/s13059-015-0844-5) [Medline](#)
52. J. Huerta-Cepas, K. Forslund, L. P. Coelho, D. Szklarczyk, L. J. Jensen, C. von Mering, P. Bork, Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* **34**, 2115–2122 (2017). [doi:10.1093/molbev/msx148](https://doi.org/10.1093/molbev/msx148) [Medline](#)
53. B. Tasic *et al.*, Shared and distinct transcriptomic cell types across neocortical areas. BioRxiv 229542 [Preprint]. 6 December 2017. <https://doi.org/10.1101/229542>
54. P. Langfelder, S. Horvath, WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008). [doi:10.1186/1471-2105-9-559](https://doi.org/10.1186/1471-2105-9-559) [Medline](#)
55. A. A. Kolodziejczyk, J. K. Kim, J. C. H. Tsang, T. Ilicic, J. Henriksson, K. N. Natarajan, A. C. Tuck, X. Gao, M. Bühler, P. Liu, J. C. Marioni, S. A. Teichmann, Single cell RNA-sequencing of pluripotent states unlocks modular transcriptional variation. *Cell Stem Cell* **17**, 471–485 (2015). [doi:10.1016/j.stem.2015.09.011](https://doi.org/10.1016/j.stem.2015.09.011) [Medline](#)