www.sciencemag.org/cgi/content/full/science.aau5324/DC1

Supplementary Material for

Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region

Jeffrey R. Moffitt, Dhananjay Bambah-Mukku, Stephen W. Eichhorn, Eric Vaughn, Karthik Shekhar, Julio D. Perez, Nimrod D. Rubinstein, Junjie Hao, Aviv Regev, Catherine Dulac*, Xiaowei Zhuang*

*Corresponding author. Email: dulac@fas.harvard.edu (C.D.); zhuang@chemistry.harvard.edu (X.Z.)

Published 1 November 2018 as *Science* First Release
DOI: 10.1126/science.aau5324

This PDF file includes:

Materials and Methods
Figs. S1 to S22
Tables S4, S5, and S9
References

Other Supplementary Material for this manuscript includes the following:
(available at www.sciencemag.org/content/science.aau5324/DC1)

Tables S1 to S3, S6 to S8, S10, and S11 as separate Excel files

Materials and Methods

Animals

Adult C57Bl6/J male and female mice aged 50-90 days were used in this study. Animals were maintained on a 12 hour:12 hour light/dark cycle (2pm-2am dark period) with *ad libitum* access to food and water. Animal care and experiments were carried out in accordance with NIH guidelines and were approved by the Harvard University Institutional Animal Care and Use Committee (IACUC).

Single-cell dissociation and library preparation

Imaging and digestion of perineuronal nets. Single-cell analysis of adult mouse brain tissue has been challenging due to the high rates of cell death during dissociation, and several protocols have been developed to partially address this issue (*70*). To further improve cell survival following dissociation of adult brain tissue, we examined the possible role of age-dependent increases in extracellular matrix glycoproteins such as chondroitin sulfate proteoglycans and hyaluronans (perineuronal nets) in increased cell death during dissociation of adult brain tissue (*76*). To visualize extracellular matrix in the preoptic region, mice aged post-natal day (PND) 10 and PND 65 were perfused transcardially with 1× phosphate buffered saline (PBS) followed by a solution of 4% v/v paraformaldehyde (PFA) in 1×PBS. Brains were harvested and post fixed overnight in 4% v/v PFA in 1×PBS and sectioned at 40-μm thickness on a vibratome (Leica, VT1000s). To reveal perineuronal nets, sections harboring the hypothalamus were blocked for 1 hour in 5% normal donkey serum in 1×PBS, followed by overnight incubation in biotinylated Wisteria Floribunda lectin (Vector Laboratories, B-1355). Sections were washed with 1% v/v Triton X-100 in 1×PBS, incubated in Alexa568-conjugated streptavidin (Invitrogen, S11226), washed again, placed on coverslips, and imaged on a Zeiss Axioscan Z1 microscope. A substantial increase in perineuronal net signal was seen in the hypothalamus of adult compared to PND 10 mice (**fig. S1B**). To address the challenge such nets might place in live cell dissociation, we modified published dissociation protocols (*70*) to include enzymes that digest extracellular matrix glycoproteins, as detailed below. As shown in **fig. S1C**, these modifications substantially improved cell survival during dissociation.

Preparation of single-cell suspensions. Adult virgin male and female mouse brains (aged ~7-8 weeks) were harvested using standard lab protocols (*5*). Tissue was pooled from 4-5 mice per sex per replicate. We generated single-cell gene-expression profiles from 3 replicates each for adult male and female mice. Each male and female replicate was processed in parallel to reduce batch effects. Briefly, mice were anesthetized with isoflurane, and brains were rapidly dissected in ice cold 1×PBS on an adult mouse brain matrix (Zivic instruments, BSMAS001-1). A ~2.5 mm × 2.5 mm × 1.1 mm (Bregma +0.5 to -0.6) tissue block spanning the preoptic region was dissected, and this tissue was placed into papain dissociation buffer comprising 8 U/ml Papain (Worthington, LK003178), 100 U/ml DNase1 (Worthington, LK003172), 50 U/ml chondroitinase ABC (Sigma, C3667), 0.07% hyaluronidase (MP BIOMEDICALS, 0210074080) 0.8 mM kynurenic acid (Sigma, K3375), 1X Glutamax (Life Technologies, 35050061), 0.05 mM (2R)-amino-5-phosphonovaleric acid (APV; Thermo Fisher Scientific, 010510), 0.01 mM Y27632 dihydrochloride (Sigma, Y0503), 0.2X B27 supplement (Thermo Fisher Biosciences, 17504044), and 1% w/v D(+)trehalose (Sigma, T9531) in hibernate A media (Life Technologies, A1247501)). The tissue was incubated in this buffer for 2 hours with gentle rotation at 37 °C. The tissue was then washed three times in Hibernate A buffer (0.8 mM kynurenic acid; 1X Glutamax; 0.05 mM APV; 1% w/v D(+)trehalose; 0.2X B27 supplement; 0.01 mM Y27632

dihydrochloride in 1X Hibernate A media) containing 0.1 mg/ml trypsin inhibitor (Sigma, T2011) at room temperature and then gently triturated in Hibernate A buffer using fire polished Pasteur pipettes of three different diameters (600 μm, 300 μm, 200 μm) to further aid the dissociation of cells. The resulting suspension was washed three times by centrifugation (400g, 10 minutes) and resuspended in Hibernate A buffer. The cell suspension was then passed through a 20-μm nylon filter (Thermo Fisher Scientific, 130-101-812) to remove debris and clumped tissue. For the comparison shown in **fig. S1C**, the dissociation cocktail labeled as the "published protocol" was the same as described above with the exception that chondroitinase ABC, hyaluronidase, and Y27632 dihydrochloride were not included.

Fluorescence activated cell sorting (FACS) and single-cell RNA-seq. To quantify the fraction of live cells as well as to sort live from dead cells, the resulting suspension was incubated for 30 minutes at 37 °C in Hibernate A buffer containing 0.1 μM calcein green AM (Life Technologies, C3099) and 7.5 μM propidium iodide (Life Technologies, P3566) to label live and dead cells, respectively. To remove excess dye, the suspension was washed and resuspended in Hibernate A buffer. Live cells were sorted by selecting calcein positive and propidium iodide negative cells on a Biorad S3 Cell sorter. The resulting suspension was examined for viability using the trypan blue exclusion method and loaded into a 10X Genomics Chromium single-cell chip at concentrations of ~300 live cells/μL according to the manufacturer's instructions. We aimed to obtain ~5,000-7,000 cells per run. Downstream preparation of sequencing libraries was carried out using the 10X Genomics Chromium Single Cell Kit V2. The libraries were sequenced on an Illumina NextSeq500 instrument using the instructions provided by 10X Genomics. Paired-end sequencing with read lengths of 150 nt was performed for all samples.

Analysis of single-cell RNA-seq

Sequence alignment. Illumina sequencing reads were aligned to the mouse genome (mm10-1.2.0) using the 10X Genomics CellRanger pipeline (version 2.0.0) with the default parameters. The data were normalized to the lowest saturated sample (78.1%), leading to 101,771 reads (mean), 2,461 genes (median), and 5,513 unique molecular identifier (UMI) counts (median) per cell. Raw reads and the output of the CellRanger pipeline are available on GEO (GSE113576).

Variable gene identification. To identify transcriptionally distinct groups of cells, we utilized approaches that have been described elsewhere (*28*, *31*, *73*). Specifically, we utilized a Poisson-Gamma-distribution model to determine the predicted variability for individual genes given their observed expression profile and identified genes with a measured variability higher than this predicted value, as described previously (*73*). In this model, the coefficient of variation, *CV*, for each gene is described by

$$CV = \sqrt{{}^{1}/_{\mu} + {}^{1}/_{\alpha}}\,, \tag{1}$$

where μ is the mean of the number of transcripts per cell for that gene and the parameter α is identical for all genes and captures technical sources of variation, such as variation in the efficiency with which transcripts within individual cells are captured. α is determined by fitting the total number of transcripts per cell, N, relative to the mean value across all cells, $\langle N \rangle$, to a gamma distribution

$$P\left({}^{N}/_{\langle N \rangle}, \alpha\right) = \frac{\alpha^{\alpha} N^{(\alpha-1)}}{\Gamma(\alpha)} e^{-\alpha \frac{N}{\langle N \rangle}}. \tag{2}$$

The set of variable genes were selected as those that had a CV at least 1.5 times greater than that predicted from the null model in Eq. (1). Small variations in this threshold did not substantially affect subsequent clustering results.

To reduce the impact of potential batch effects as well as of perturbations due to cell preparation, we first removed overall variations in the number of transcripts observed for each cell by dividing the observed counts within each cell by the total number of transcripts within that cell. These normalized counts were then scaled by the average number of transcripts observed within all cells across all replicates, creating a matrix, E_{ij}, of normalized counts for each gene, *i*, in each cell, *j*. Unless otherwise noted, we supplemented this expression matrix with a pseudo count of 1 and log transformed it, i.e. $\log_{10}(E_{ij}+1)$, for the subsequent cell clustering analysis, as has been described previously (*28*, *31*, *73*).

The cell dissociation and capture process can modify the expression of some mRNAs. In particular, it has been documented to increase the expression of ribosomal-protein RNAs, mitochondrial RNAs, and immediate-early genes as well as to lead to changes in the total number of transcripts recovered per cell (*77*, *78*). We used these known proxies for cell-dissociation-induced expression changes to mitigate expression variations in all genes that might be due to cell dissociation. Specifically, we first computed a score for each cell associated with the ribosomal, mitochondrial, and immediate-early gene expression level within each cell. Each of these three scores was calculated from the average of the $\log_{10}(E_{ij}+1)$ values within each cell for each of the following gene sets. The ribosomal score was calculated from the average of all ribosomal-protein genes, i.e. all genes with names starting with 'Rpl' or 'Rps'; the mitochondrial score was calculated from the average of all mitochondrial transcripts, *i.e.* all genes with names starting with 'mt-', and the immediate-early gene score was calculated from the average of Fos, Egr1, and Npas4. We then constructed a linear regression model for the $\log_{10}(E_{ij}+1)$ for each gene across all cells as a function of the ribosomal, mitochondrial, and immediate early gene scores as well as the total number of transcripts per cell, to subtract the effect of cell-dissociation-induced expression changes, as previously described (*28*, *79*), and the resulting $\log_{10}(E_{ij}+1)$ after this treatment was used for clustering. However, all reported expression values were calculated from the log transformed data prior to this regression.

To further remove any potential effects of cell-dissociation-induced expression changes on clustering, we removed 'variable' genes with an expression level that correlated across all cells with any of the above scores (ribosomal, mitochondrial, and immediate-early gene), as described previously (*28*, *79*) . Genes with a Pearson correlation coefficient with any of these scores larger than 0.2 or smaller than -0.2 were excluded. We also excluded genes that were expressed in 20 or fewer cells from the 'variable' gene list. 859 genes were contained in the final set of variable genes.

Principal component analysis (PCA) and community-based cell clustering. To identify transcriptionally distinct clusters of cells, we first reduced the dimensionality of these data with a PCA on the z-score of the regressed $\log_{10}(E_{ij}+1)$ values for the final set of variable genes. The z-score was calculated for individual genes across all cells. To determine the number of principal components to keep from this analysis, we estimated the largest eigenvalue one would expect to obtain in PCA computed on random permutations of the data. Specifically, we conducted 50 iterations, where, in each case, we randomly permuted the expression values of each gene independently across cells. This permutation maintains the univariate distribution for every gene, but deflates gene-gene correlations. For every iteration, we calculated the maximum eigenvalue

from a PCA of the randomized matrix. We kept all principal components that had an eigenvalue larger than the mean of the distribution of maximum eigenvalues across the 50 iterations.

To then identify transcriptionally distinct clusters of cells, we utilized the Jaccard-Louvain community detection approach on a *k*-nearest neighbor graph as implemented by the Phenograph package (*30*) with the following modifications. First, we performed cell clustering based on the principle components identified above using the Jaccard-Louvain community detection approach on a *k*-nearest neighbor graph (*30*). We then identified stable clusters by running a bootstrap analysis to identify cells that were consistently clustered together. Specifically, we selected a random 50% of cells and repeated the PCA, graph construction, and community detection for those cells. We then compared the clustering results derived from this subset of cells to that derived from the original clustering of all cells and computed the maximum fraction of cells from each of the original clusters found in any of the clusters generated in the bootstrap. We defined an original cluster as stable if the average of this value across 20 replicate bootstraps was greater than or equal to 0.5. We found that the vast majority of clusters were not sensitive to small changes in this threshold. Cells that were members of unstable clusters or which clustered into small clusters (<10 cells) were marked as unstable and discarded from subsequent analysis.

In addition, we found that the number of stable clusters and the number of cells assigned to stable clusters was somewhat sensitive to the choice of the neighborhood size, *k*, used to create the nearest neighbor graph. Thus, we developed a method to guide the selection of this number. We performed clustering and the cluster stability analysis for a range of *k* values and then examined the dependence of the number stable clusters and the fraction of cells assigned to stable clusters on *k* (**fig. S2A**). Initially, as we increased the value of *k*, the number of cells that were assigned to stable clusters increased substantially with a concomitant but slow decrease in the number of stable clusters. However, after *k* was increased to pass a certain value (the "elbow" point), further increase in *k* caused a precipitous decrease in the number of stable clusters with little to no change in the number of cells assigned to stable clusters. This behavior allowed us to find a value of *k* that balanced the desire to better resolve similar groups of cells (i.e. increase the number of stable clusters) with the desire to explain the majority of the data (i.e. maximize the number of cells within stable clusters). The specific *k* value selected ($k = 10$) is depicted in **fig. S2A**.

After clustering, we identified several clusters that expressed mixtures of markers for multiple cell types, e.g. neurons and oligodendrocytes, and that did not contain any genes that were uniquely enriched in these clusters (as judged by the largest z-score of individual genes for the average expression profile of individual clusters). These clusters were marked as ambiguous, possibly representing doublets, and were excluded from subsequent analysis.

This clustering analysis produced major cell classes, including inhibitory neurons, excitatory neurons, microglia, astrocytes, mature and immature oligodendrocytes, ependymal cells, endothelial cells, and mural cells, some of which contain further sub-divisions. The major cell classes and sub-divisions of the non-neuronal cell classes derived from this analysis are summarized in **Table S1**, but the sub-division results of neurons from this round of analysis are not depicted in **Table S1** because the clustering of inhibitory and excitatory neurons were further refined, as described below.

To further refine our detection of transcriptionally distinct populations of inhibitory and excitatory neurons, we utilized the expression of Gad1 and Vglut2 (Slc17a6) to assign individual clusters as inhibitory or excitatory neurons, respectively. We then pooled all inhibitory neurons together and all excitatory neurons together, and repeated the procedure of variable gene

identification, PCA, and community-based clustering as described above for cells in these two major classes separately. We identified 246 and 172 variable genes for inhibitory and excitatory neurons, respectively. We identified the optimal value of *k* for each of these sets of neurons using the same bootstrap analysis described above (*k* = 20 and 10 for excitatory and inhibitory neurons, respectively; **fig. S2B, C**). We initially named these clusters as 'i' for inhibitory clusters or 'e' for excitatory clusters based on the relative strength of inhibitory (Gad1) or excitatory (Slc17a6) markers and then numbered them in order of cell abundance with e1 and i1 as the highest abundance excitatory and inhibitory clusters, respectively. One excitatory cluster (e18) was identified as unstable from our bootstrap analysis and removed. No inhibitory clusters were unstable. Two inhibitory clusters (i33 and i34) were identified as putative doublets based on the strong expression of non-neuronal markers and were removed from subsequent analysis. We did not relabel the remaining clusters hence the cluster IDs e18, i33, and i34 are missing from our final data set. In addition, we identified several clusters which expressed both inhibitory and excitatory markers (i19, i26, and e14) and renamed these clusters as (h1, h2, and h3). Hence, cluster IDs i19, i26, and e14 are also missing from our final data set. To confirm that these hybrid clusters were not composite clusters or doublets, we first examined individual cells within this cluster and determined that many cells co-expressed both the inhibitory and excitatory markers. In addition, we used a doublet detection software (https://github.com/JonathanShor/DoubletDetection) to confirm that these clusters and all other clusters were not enriched in doublets. The sub-divisions of inhibitory neurons and excitatory neurons derived from this round of analysis are also summarized in **Table S1**.

tSNE calculation. The t-distributed stochastic neighbor embedding (tSNE) for all cells, inhibitory neurons, and excitatory neurons was performed based on the same principal components derived from the variable genes described above, namely 859, 246, and 172 genes for all cells, inhibitory neurons, and excitatory neurons, respectively. tSNE values were calculated using the Barnes-Hut implementation (*80*), which is publicly available (https://github.com/lvdmaaten/bhtsne).

Marker gene identification. For each neuronal cluster, we utilized the model-based analysis of single-cell transcriptomics (MAST) package (*75*) to identify the set of all genes that were enriched within that cluster relative to all other neurons. If the enrichment of a given gene within a given cluster had a false-discovery-rate (FDR) adjusted p-value of 0.01 or less, its enrichment was judged as statistically significant.

Figure 2A-C report all clusters that are enriched in expression of Gal, Th, or Adcyap1, respectively. The marker genes listed for each cluster in **Fig. 2A-C** represent statistically significant genes (not exhaustive) from each cluster, in addition to Gal, Th, or Adcyap1/Bdnf. The violin plots in **Fig. 2A-C** contain both marker genes selected in this fashion (with gene name colored in black) as well as genes of particular interest that are expressed but not necessarily statistically significantly enriched within some of the displayed clusters (with gene name colored in blue and green).

Hierarchical tree construction. To build a hierarchical tree that explains the relationship between neuronal clusters, we used the same significant principal components (PCs) used for clustering (described above). Using all the cells for a given cluster, the mean PC loadings were calculated. A distance matrix was then calculated in Euclidean space and this was fed to the r function 'hclust' using the 'complete' method for constructing the tree. Genes marking branches within the trees in **Fig. 1C, D** were identified by finding differentially expressed genes between

all cells within the clusters on one side of the branch relative to all cells within the clusters on the other side of the branch. The listed genes are not necessarily the most differentially expressed.

Gene set enrichment analysis (GSEA). To explore the potential enrichment of different gene categories within the genes that mark different neuronal clusters, we utilized a gene set enrichment analysis (*81*). Specifically, we first constructed manually curated lists for three categories of genes (1) neuropeptides and genes involved in neuromodulator production and transport, (2) transcription factors, and (3) neuromodulator (neuropeptide and hormone) receptors, from public resources. The genes included within each of these categories are listed in **Table S3**. Next, to determine the set of differentially expressed genes, we estimated the z-score of $\log_{10}(E_{ij}+1)$ for each gene for each cluster via a Wilcoxon rank-sum test for neurons within each cluster versus all other neurons. We selected putative markers as any gene that had a z-score greater than 5 within any cluster, a relatively permissive threshold that produced a set of 3610 putative markers. To then rank these genes based on the strength with which they mark individual clusters, we calculated, for each gene, the fold change for the average $\log_{10}(E_{ij}+1)$ value within each cluster relative to the average $\log_{10}(E_{ij}+1)$ value for cells of all other clusters. We then sorted all genes based on their maximum fold change value in any cluster. To determine if any of these three gene categories were enriched in the stronger markers, we utilized a publicly available, light-weight implementation of the GSEA algorithm in R (https://github.com/JEFworks/liger). The results of this analysis are provided in **fig. S6** and the p-values are also provided in **Fig. 1E**. These results did not depend critically on our z-score threshold for defining putative markers.

Nomenclature of scRNAseq clusters. In addition to the cluster ID (e1, e2, …, i1, i2, …, and h1, h2, h3) as described above, we also assigned a descriptive name to each neuronal cluster. On the basis of our finding that neuropeptides and molecules involved in neuromodulator synthesis and release were enriched in marker genes of neuronal clusters in the scRNA-seq dataset, we chose a naming scheme for preoptic neuronal clusters that highlights this fact and also identifies markers for potential future functional manipulation of specific neuronal populations. A two-marker nomenclature was devised in which the first marker gene mostly relies on neuromodulators enriched in a cluster. In addition, if needed and possible for full identification, a second marker gene was provided that either reflects a second peptide, or another biologically important gene co-expressed in a cluster. In the absence of identifiable peptide or neuromodulator marker, we used specifically enriched/highly expressed genes as markers. In some cases, two clusters expressed the same two cardinal markers and differed only in the expression levels of these markers. In the case where no genes satisfactorily distinguished clusters for manipulation purposes, we choose genes that varied in expression levels between clusters. In these cases, the secondary or primary marker gene of these clusters were chosen based on their higher expression in one cluster over the other. Finally, for the inhibitory and excitatory clusters that we could not readily identify two notable marker genes, we designated them as Gaba and Glut, respectively, together with the name of an additional gene to help differentiate among these less defined clusters when possible.

Bulk RNA-seq of the preoptic region

Estimates of RNA expression in the preoptic region were derived from previously published bulk RNA-seq data of this brain region [GSE22131] (*43*). Specifically, we reanalyzed and averaged the results from two replicates (F1i, F1r) using a previously published pipeline (*82*). In some cases, the sequencing data are not sufficient to distinguish the expression level of two or

more similar isoforms of the same gene. In this pipeline, this situation is identified, and the combined expression of these isoforms is reported. These bulk RNA-seq data were used to determine the specificity of potential MERFISH probes (described below). We also used these bulk RNA-seq data for comparison with the average abundance per cell measured with MERFISH (**fig. S11B**).

Selection of genes for MERFISH

In order to discriminate transcriptionally distinct cell populations with MERFISH, we designed a MERFISH gene panel comprised of 155 genes. To discriminate major cell classes, we included in this panel a set of 33 established markers for inhibitory neurons, excitatory neurons, mature and immature oligodendrocytes, astrocytes, microglia, endothelial cells, and mural cells, which we confirmed were expressed within the expected cell classes in our scRNA-seq data. To discriminate different neuronal types, we took two approaches to select genes. In the first approach, we selected a panel of genes likely to be relevant to neuronal function in the hypothalamus (*33*, *37*). Specifically, we generated a panel of 16 neuropeptide genes and 28 neuropeptide and hormone receptor genes. In addition, we also included a panel of 11 genes reported to be expressed differentially between male and female mice in some portions of the preoptic region of the hypothalamus (*41*, *42*). This strategy for gene selection serves the dual purpose of revealing important functional properties of neurons as well as including genes that likely discriminate many neuronal populations. Some genes appeared in multiple of these categories; thus, this set contained a total of 52 unique genes. When combined with the 33 markers for cell classes, this pre-selected gene set contained a total of 85 unique genes.

In the second approach to select genes that discriminate different neuronal types, we utilized our scRNA-seq data to generate a set of putative marker genes for the identified neuronal clusters. This marker set started with all 'variable' genes identified for the inhibitory and excitatory neurons, as described above. To include genes that capture the transcriptional heterogeneity observed in our data, we identified putative markers by clustering the inhibitory and excitatory neurons using the approach described above but with two modifications designed to include more cell clusters and, thus, allow for the identification of genes that also display expression heterogeneity within some of the reported scRNA-seq clusters. First, we purposefully selected a value of the nearest neighborhood size k that was smaller than what we selected using our guidelines above so as to sub-divide clusters that contained some transcriptional heterogeneity (k = 7 for both excitatory and inhibitory neurons). Second, we included all clusters even if they were not identified as stable via the bootstrap method described above. We computed the average value of $\log_{10}(E_{ij} + 1)$ for each gene across all clusters and then computed the z-score of these values for individual genes across all clusters. We selected putative markers as any gene that had a z-score greater than 2 in any cluster and was expressed in at least 25% of the cells in that cluster.

We then trimmed this list of putative markers for neuronal clusters by identifying genes that are potentially challenging for highly multiplexed FISH imaging experiments. These include (1) genes that were relatively short and hence would only allow binding of relatively few hybridization probes (the number of probes we used for each transcript is described in more detail below) and (2) genes that were expressed at an average of 10 or more UMI counts in any cell cluster as determined from our scRNA-seq data. We chose this threshold of 10 UMI counts to mitigate the concern that some RNAs may be expressed at abundance levels too high to measure with MERFISH. Given the reported capture efficiencies for droplet-based methods of

~10% (*27*, *28*), we estimate that an average count of 10 UMIs corresponds to ~100 RNA copies per cell. This 10-UMI threshold is a conservative threshold to use because even at a RNA density substantially greater than ~100 copies per cell, individual RNA molecules can still be resolved by diffraction-limited imaging. We note that some of genes that did not pass the above criteria on RNA length and expression but are important for discriminating cell populations or describing cell function were still included in our MERFISH measurements, but not as a part of the combinatorial single-molecule FISH (smFISH) imaging run as described in our previous MERFISH work (*20*, *26*, *40*); instead, these genes were imaged in a non-combinatorial manner in a set of sequential, multi-color FISH imaging rounds with one single gene measured in each color channel per round, following the combinatorial smFISH imaging run, as described below.

This trimmed list of putative markers for neuronal clusters contained 199 genes, of which 7 were in the list of 85 pre-selected gene set described above. However, we found that many of these markers were expressed in the same sets of neuronal clusters and, thus, marked the same neuronal types. To reduce this redundancy, we utilized an iterative approach to identify pairs of genes that most strongly correlated in their expression across different neuronal clusters. We first computed the Pearson correlation coefficients between the average expression of individual pairs of genes across all clusters. We then identified the pair of genes with the largest correlation coefficient and removed one of these genes at random. Using this approach, we reduced this set of marker genes for neuronal clusters to a final number of 70 genes. Combining this set with the first set of 85 pre-selected genes as described above, we included a total of 155 genes in our MERFISH measurements.

Among these 155 genes, 135 genes were imaged in the combinatorial smFISH imaging run. However, all of the neuropeptides as well as several genes that we considered to be important in discriminating cell populations, i.e. Th, Mbp, Sln, Nnat, were either expressed at high levels or are short, and hence are potentially challenging for the combinatorial smFISH imaging run. These amounted to a total of 20 genes. We included these 20 genes in our overall MERFISH measurements but detected them in the non-combinatorial, sequential rounds of multi-color FISH following the combinatorial smFISH imaging run. For behavioral measurements, cFos was added to the set of genes measured with sequential rounds of FISH. The set of genes measured with MERFISH are described in **Table S6**.

Design of MERFISH encoding probes

MERFISH encoding probes for the 135 genes that were included in the combinatorial smFISH imaging run were designed using a pipeline that we have described previously (*40*). Each MERFISH encoding probe contains a 30-mer targeting region that is complementary to the RNA of interest and, thus, targets that probe to that RNA, as well as a series of 20-mer readout sequences that encode the specific barcode assigned to each RNA.

We first identified all possible 30-mer targeting regions within each desired transcript as we have previously described (*40*). From the set of all possible 30-mers for each gene, we selected 92, 30-mers at random. Also as we have previously described (*83*), only a fraction of possible MERFISH probes bind to each RNA, likely due the exclusion of binding sites by proteins or RNA secondary structure. Thus, we allowed these 30-mers to overlap by as much as 20 nt to increase the number of probes utilized for shorter RNAs. This allowed us to include genes as short as ~1000 nt (*83*).

We next assigned to each of the 135 RNAs a unique binary barcode drawn from a previously published, 16-bit, Hamming-Distance-4, Hamming-Weight-4 encoding scheme (*20*).

The encoding scheme contains 140 possible barcodes; we used only 135 of them to code RNAs and left 5 of these barcodes unassigned so that measurement of these 5 'blank' barcodes would provide a direct measure of the false-positive rate in MERFISH (**figs. S9F-H, S11C**).

We then utilized these barcodes to assign readout sequences to the encoding probes associated with each gene with each encoding probe containing two readout sequences, as previously described (*40*). For the 16-bit encoding scheme, a total of 16 readout sequences were used, and the collection of encoding probes for each gene together contain 4 of the 16 readout sequences according to the barcode assigned to the gene. In addition, we concatenated to each encoding probe sequence two PCR primers, the first comprising the T7 promoter, and the second being a random 20-mer designed to have no region of homology greater than 15 nt with any of the encoding probe sequences designed above, as we previously described (*40*).

Encoding probes for the 20 genes imaged in the non-combinatorial sequential multi-color FISH rounds were produced in the same fashion, except that each of these RNAs was associated with a single unique readout sequence and that a single copy of this readout sequence was placed at the 5' end of the target region for each probe. Each of these RNAs contained 36 or more encoding probes. The readout sequences used here were different from the 16 readout sequences used for the combinatorial smFISH run. The sequences of all template molecules for encoding probes are included in **Table S10**.

Construction of MERFISH encoding probes

MERFISH encoding probes were created from the template molecules above, following a procedure as we have described previously (*26*). Briefly, template molecules were synthesized as a complex oligopool (CustomArray). This pool was then amplified via PCR to produce a set of templates for an *in vitro* transcription reaction, which in turn generated a large quantity of RNA molecules complementary to these templates. The RNA molecules were purified via size-exclusion chromatography (Zeba desalting columns; ThermoFisher, 89894) and then converted back to DNA using reverse transcription (Maxima H- reverse transcriptase; ThermoFisher, EP0751). We utilized the USER enzyme system (New England Biolabs, M5505S) to excise the reverse transcription primer, which contained a penultimate 'U' at the 3' position, as described previously (*83*).

Encoding probes to the 135 genes measured in the combinatorial smFISH run or the 20 genes measured in sequential multicolor FISH rounds were produced separately and then mixed during staining. Probes for cFos and Th were synthesized via direct solid phase synthesis (Integrated DNA Technologies [IDT]).

Tissue section preparation for MERFISH experiments

Mice (aged 7-8 weeks) were individually housed for four to five days and then euthanized directly from the home cage or exposed to specific behavioral stimuli (as described below) prior to harvest. Mice were euthanized with isoflurane, the brain quickly harvested, frozen immediately in optimal cutting temperature compound (Tissue-Tek O.C.T.; VWR, 25608-930), and stored at -80 °C for at least 6 hours before cutting. In parallel, coverslips for mounting the tissue sections for MERFISH imaging experiments were prepared as described previously (*26*).

Frozen brains were manually dissected to a ~ 3-mm × 3-mm × 3-mm block containing the entire preoptic region and surrounding nuclei and were sectioned at -18 °C on a cryostat (MICROM, HM550). Slices were removed and discarded until the preoptic region was reached. For the animals that were not exposed to behavioral stimuli (naïve animals), a contiguous set of

sixty 10-μm-thick slices were collected with every fifth slice (12 slices total) placed onto coverslips for imaging. The other slices were either placed on additional coverslips for additional use or discarded. Each coverslip contained either 6 slices from the anterior portion or 6 slices from the posterior portion of the preoptic region. These slices are labeled in order from anterior to posterior as 1 to 12, with 1 representing the most anterior slice and 12 the most posterior slice.

This sampling frequency corresponds to a 50-μm resolution along the anterior-posterior axis, and we adopted this sampling frequency because 50 μm is smaller than the sizes of all established nuclei in the preoptic region, and previous measured gene expression patterns do not change substantially over a 50-μm anterior-posterior distance in this region. Thus, we anticipated that this sampling frequency would be sufficient to capture the cell diversity in this region. Indeed, we find that all but the ~5 very low abundance neuronal clusters identified via MERFISH are present in multiple subsequent slices (**fig. S17**), confirming that this anterior-posterior resolution was sufficient to sample the spatial distributions of the neuronal clusters in the preoptic region.

For the animals that were exposed to behavioral treatments, we collected the same contiguous set of 10-μm-thick slices and kept every 15th slice. These slices were labeled in order from anterior to posterior as 1 to 4. Given the spatial distribution observed for neuronal clusters identified in naïve animals with the higher sampling frequency described above (**fig. S17**), we anticipated that this reduction in sampling frequency would not result in significant loss in neuronal clusters identified. Indeed, we found that all neuronal clusters identified with the higher sampling frequency, naïve animal data were observed in the lower sampling frequency data measured for animals exposed to behavioral treatments.

Tissue slices were allowed to briefly thaw on coverslips at room temperature, and then were fixed by treating with 4% PFA in 1×PBS for 10 minutes. We then washed coverslips three times with 1×PBS and stored them in 70% ethanol at 4 °C for at least 18 hours to permeabilize cell membranes.

Staining of MERFISH samples was conducted as we have described previously (*26*). Briefly, we removed the 70% ethanol from samples and washed them once with encoding-probe wash buffer (30% formamide in 2× saline sodium citrate [SSC]) for five minutes at room temperature. The wash buffer was then aspirated from a coverslip, and the coverslip was inverted onto a 50 μL droplet of probe mixture on a parafilm coated petri dish. The probe mixture comprised ~1 nM of each encoding probe for the combinatorial smFISH run, ~10 nM of each encoding probe for the sequential multi-color FISH rounds, and 1 μM of a polyA-anchor probe (IDT) in 2×SSC with 30% v/v formamide, 0.1% wt/v yeast tRNA (Life Technologies, 15401-011) and 10% v/v dextran sulfate (Sigma, D8906). We then incubated the sample at 37 °C for 36 hours. The sequence of the polyA-anchor probe was /5Acryd/ TTGAGTGGATGGAGTGTAAT T+TT+TT+TT+TT+TT+TT+TT+TT+TT+T, were T+ indicates locked nucleic acid and /5Acryd/ represents a 5' acrydite modification. This polyA-anchor probe hybridizes to the polyA sequence on the polyadenylated mRNAs and allows these RNAs to be anchored to a polyacrylamide gel as described below. To wash away excess encoding probes and polyA-anchor probes, each coverslip was washed in encoding-probe wash buffer for 30 minutes at 47 °C for a total of two times. We optimized the concentrations of the encoding probes each time these probes were created.

To clear our samples and remove fluorescence background, we embedded tissue slices in a thin polyacrylamide gel and then treated these samples with protease and detergent as we described previously (*26*) but with a modified digestion protocol that we found to improve

clearing in the mouse brain. Specifically, we utilized a digestion buffer comprising 2×SSC, 2% v/v sodium dodecyl sulfate (SDS; ThermoFisher, AM9823), 0.5% v/v Triton X-100 (Sigma, X100), and 1:100 proteinase K (New England Biolabs, P8107S) and digested samples for 48 hours at 37 °C. The digestion buffer was exchanged once during this incubation. After digestion, coverslips were washed in 2×SSC for 30 minutes for a total of four washes and then stored at 4 °C for up to a week in 2×SSC supplemented with 1:100 Murine RNase inhibitor (New England Biolabs, M0314S) prior to imaging. We observed no degradation in the quality of the sample over this time.

MERFISH imaging

To prepare for a MERFISH measurement, we first filled a home-built reagent cartridge with fresh aliquots of buffers utilized in each round of staining and imaging. This cartridge contained 40 mL of a readout-probe-wash buffer comprised of 2×SSC, 10% v/v ethylene carbonate (Sigma, E26258), and 0.1 % v/v Triton X-100; 40 mL of a cleavage buffer comprised of 50 mM Tris(2-carboxyethyl)phosphine (TCEP; Sigma, 646547) in 2×SSC; 40 mL of a TCEP-wash buffer comprised of 2×SSC; and 40 mL of an imaging buffer comprising 5 mM 3,4-dihydroxybenzoic acid (Sigma, P5630), 2 mM trolox (Sigma, 238813), 50 μM trolox quinone, 1:500 recombinant protocatechuate 3,4-dioxygenase (rPCO; OYC Americas), 1:500 Murine RNase inhibitor, and 5 mM NaOH (to adjust pH to 7.0) in 2×SSC. We used the O_2 scavenging enzyme rPCO and its substrate 3,4-dihydrozybenzoic acid to remove O_2 and the combination of Trolox and Trolox quinone to decrease photobleaching during imaging (*84*, *85*). All buffers were degassed for 5 minutes under vacuum prior to use. The cleavage buffer and the imaging buffers were stored under a layer of mineral oil (Sigma, 330779) to prevent oxygenation during the course of the measurement.

In parallel, we prepared a separate reagent cartridge containing the readout probe mixtures appropriate for each round of hybridization. The readout probe mixture was comprised of the readout-probe-wash buffer supplemented with 3 nM each of the appropriate readout probes. Two probes were included in each mixture, one labeled with Cy5 and the other with Alexa750. These dyes were conjugated to the readout probe via a disulfide bond and were synthesized by Biosynthesis, Inc. A total of 8 readout probe mixtures (16 readout probes) were used for the combinatorial smFISH run and a total of 11 readout probe mixtures (22 readout probes) were used for the sequential multi-color FISH rounds. We utilized 5 mL of each readout probe mixture.

Both reagent cartridges were loaded into a home-built, automated fluidics system as we have described previously (*20*). The sequences of all readout probes as well as the associated dyes are described in **Table S11**.

To prepare the sample for imaging, we first stained it with a readout hybridization mixture containing the readout probes associated with the first round of imaging in the combinatorial smFISH run, as well as a probe complementary to the polyA-anchor probe and conjugated via a disulfide bond to the dye Alexa488 (**Table S11**), also at a concentration of 3 nM. The sample was incubated in this mixture for 15 minutes at room temperature, and then washed in the readout-probe-wash buffer supplemented with 1 μg/mL DAPI for 10 minutes to both wash away excess readout probe and stain nuclei within the sample. The sample was then washed briefly in 2×SSC and imaged immediately.

We imaged samples largely as we have described previously (*26*). Briefly, the sample was loaded into a commercial flow chamber (Bioptechs, FCS2) with a 0.75-mm-thick flow gasket

(Bioptechs, 1907-100; DIE# F18524). Imaging buffer was introduced into the chamber and the sample was imaged with a low magnification objective (Olympus, PlanN 10×/0.25 NA) with 405-nm illumination to produce a low-resolution mosaic of all slices in the DAPI channel. We then used this mosaic image to identify the center position of each slice and generated a 9×9 grid of field-of-view (FOV) positions covering 1.8-mm ×1.8-mm around each slice center.

We then switched the microscope to a high magnification, high-numerical aperture objective (Olympus, PlanApo 60×/1.3 NA, silicon oil) and imaged each of the FOV positions generated above. In the first round of imaging, we collected images in the 750-nm, 650-nm, 488-nm, and 405-nm channels to image the first two readout probes (conjugated to Alexa750 and Cy5, respectively), the total polyA-mRNA signal as revealed by the probe complementary to the polyA-anchor probe and conjugated to Alexa488, and the DAPI signal (405-nm channel). We utilized the latter two channels for cell segmentation as described below. To image the entire volume of each slice, we collected seven 1.5-μm-thick z-stacks for each of these color channels. In addition, we also collected a single image of the orange fiducial beads on the surface of the coverslip using the 560-nm illumination channel. This image served as a spatial reference, and all subsequent images of each FOV were referenced to this image to correct for slight misalignments in the stage position that naturally occur when the stage is moved.

After the first round of imaging, the dyes were cleaved from the readout probes by flowing 1.65 mL of cleavage buffer (50 mM TCEP in 2x SSC) through the flow chamber and then incubating the sample in this buffer for 15 minutes. Excess cleavage buffer was washed away by flowing 1.5 mL of TCEP-wash buffer over the span of three minutes. To perform the second round of readout probe hybridization, we flowed 3 mL of the readout probe mixture containing the appropriate readout probes across the chamber and incubated the sample in this mixture for a total of 15 minutes. To wash off excess readout probe, we flowed 1.25 mL of readout wash buffer across the system and incubated the sample in this buffer for 5 minutes. Finally, we flowed 1.5 mL of imaging buffer across the sample over the span of three minutes. We then imaged the sample at all FOV locations in the 750-nm, 650-nm, and 560-nm channels as described above. We repeated this process for all rounds of hybridization and imaging to complete the combinatorial smFISH run (8 rounds of hybridization and imaging) and the sequential multi-color FISH rounds (10 or 11 rounds of hybridization and imaging). The RNAs measured within the sequential multi-color FISH rounds were often so abundant when expressed that individual RNA molecules were not observed. For this reason, we did not need measurements of the entire cell volume to produce a statistically robust measure of expression; thus, to reduce imaging time, we only imaged the central z-plane for both the 650-nm and 750-nm channels in these rounds. These samples were imaged on a MERFISH microscope that we have described previously (*26*, *40*).

MERFISH image analysis and cell segmentation

All MERFISH image analysis was performed with algorithms largely similar to what we have described previously (*26*, *40*) but in a revised pipeline optimized for running in massive parallel on the Harvard Odyssey high-performance computer cluster. All software is available at github.com/ZhuangLab/MERFISH-analysis.

As described previously (*20*, *26*, *40*), we identified the location of the fiducial beads in each FOV in each round of imaging and used these locations to create an affine transformation that reregisters all z-stacks of images for each FOV in each imaging round to the position of that FOV in the first imaging round. We high-pass filtered these image stacks to remove background,

deconvolved them using 20 rounds of Lucy-Richardson deconvolution to tighten RNA spots, and low-pass filtered them to account for small movements in the centroid of RNAs between imaging rounds as we have observed previously (*20*). Individual RNA molecules imaged in the combinatorial smFISH run were identified as describe in our previous MERFISH studies (*26, 40*). Briefly, we assigned individual pixels to specific barcodes by finding the barcode with the smallest Euclidean distance between the normalized intensity profile across all 16 images in the combinatorial smFISH imaging run for that pixel and the predicted intensity profile across all 16 bits for that barcode. To remove spurious assignments, we discarded any pixels for which this distance was larger than the distance between a given barcode and any of the error-containing barcodes generated via a single bit flip. Contiguous sets of pixels in each image that matched the same barcode where then combined to form a putative RNA.

We removed a background of spurious barcodes generated by random fluorescent fluctuations by removing putative RNAs that contained a small number of contiguous pixels (< 4) and putative RNAs that were dim. To identify dim RNAs, we utilized the observation that two peaks were present in a histogram of the average brightness of identified RNAs and set a brightness threshold based on the local minimum in probability between these two populations, as previously described (*26*, *40*).

The identity of the RNAs imaged in the non-combinatorial, sequential multicolor FISH rounds were naturally provided by the round and color channel in which each RNA was imaged.

We identified cell segmentation boundaries in each FOV using a seeded watershed approach. To identify seeds, i.e. regions of the image that must contain a cell, we first low-pass filtered the DAPI image stack to remove noise and then applied an adaptive intensity threshold to identify putative nuclei. To resolve closely spaced nuclei, we eroded this image stack using a disk of 2-μm diameter, less than half the observed average diameter for nuclei. We identified the center of these putative nuclei from regional maximum of this eroded image. To account for slight offsets in the center of the same nuclei in different z-planes, we used image dilation to connect putative nuclei center in different z-planes that were within 1-μm of one another. To identify regions of high polyA signal that correspond to individual cells, we similarly applied a low pass filter and an adaptive intensity threshold to the polyA image stack. To create catch basins for the watershed algorithm, we then inverted this image and set the regions associated with the nuclei seeds to zero. To remove local variations in the polyA intensity and force all watershed basins to contain a seed, we used morphological reconstruction to remove all local image minima other than the regions associated with seeds. Finally, we applied a watershed algorithm on this processed polyA stack to identify segmentation boundaries. Because each of these image operations was applied to the full z-stack of images for each FOV, as opposed to individual z-sections, the generated cell segmentation boundaries were naturally 3D. This algorithm was able to largely address the situation in which two adjacent cells partially overlapped along the axial direction. Nonetheless, given the lower optical resolution in the axial direction relative to the lateral directions, we expect that some cells that overlapped in the z-direction were not properly segmented. In this case, we anticipate that our ability to detect doublets during clustering should further allow some of this small fraction of remaining doublet cells to be identified and rejected.

This segmentation approach generated a small population of spurious segmentation artifacts, which resulted in putative cells with very small total volumes. We removed these features from subsequent analysis by requiring that all cells have a volume of at least 100 μm^3.

We assigned individual RNAs to segmentation boundaries based on whether or not they fell within those boundaries. We note that given the thickness of our slices, some cells could not be imaged completely. Thus, to remove differences in RNA counts due to imaged cell volumes, we normalized the RNA counts per cell by the imaged volume of each cell. In addition, we observed a modest batch effect between MERFISH runs. Specifically, we found that the mean total number of RNAs identified per cell varied by ~20% from run-to-run. This variation could arise from variation in the RNA integrity between samples or in the signal-to-noise ratio for individual RNA spots. We sought to remove the influence of these batch effects by normalizing the mean total RNA density per cell for each MERFISH data set so that this mean value was the same across all datasets.

We observed that the polyA stain dropped substantially in brightness outside of the soma of cells such that, in the vast majority of cases, we could not identify neuronal or glial processes from these stains, and thus the boundaries that we segmented corresponded to the boundaries of the cell soma. Some RNA molecules that we identified fell outside of these boundaries and are, thus, candidates for RNAs found within neuronal or glial processes.

For the RNAs imaged in the non-combinatorial sequential FISH rounds, since many of them were so abundant that individual molecules could not be resolved, we quantified the signal from these sequential FISH rounds by summing the fluorescence intensity of all pixels that fell within the segmentation boundary associated with the central z-plane (the only plane imaged for these sequential FISH rounds). We then normalized this signal by the area of the segmentation boundary in the central z-slice. We observed that the fluorescence background observed in these sequential FISH rounds was different in both the 650-nm and 750-nm channels. To remove this differential background, we exploited the fact that no cells expressed a majority of the genes measured in the sequential FISH rounds; thus, we estimated the background in the 650-nm and 750-nm channels for each cell by taking the median of the signal for each sequential FISH round in each of these channels. We then subtracted these background estimates from the measured fluorescence intensities of each sequential FISH round. A small population of cells did not contain a segmentation boundary within the central z-slice, presumably because these represent small cell fragments present only in the top or bottom portions of the slice, and these cells were removed from subsequent analysis.

Clustering of MERFISH data

MERFISH data were clustered using the same approach as described for the clustering of scRNA-seq data with several notable exceptions. First, given that the total number of genes in the MERFISH panel is only 155, a reduction in the number of included genes was not computationally necessary. Thus, we did not select ‘variable’ genes but rather used all genes for clustering. Second, MERFISH has a small false positive rate, as revealed by the small but finite frequency at which we observe the blank barcodes (**figs. S9F-H, S11C**). A logarithmic transform would tend to increase the weight of such false counts. To avoid this effect, we performed all clustering on the z-scores of the normalized RNA densities per cell, with the z-score calculated for individual genes across all cells, without applying a logarithmic transform. Following the scRNA-seq approach described above, we then reduced the dimensionality of these data with PCA and identified clusters using the Jaccard-Louvain community-based detection with a bootstrap analysis to both identify stable clusters and select the optimal value of the nearest neighborhood size *k* as described earlier for the scRNA-seq clustering analysis. We utilized a *k*

value of 10 for all cells, a value of 12 for inhibitory neurons, and a value of 10 for excitatory neurons.

As with the scRNA-seq measurements, we first performed clustering of all cells and identified major cell classes, including inhibitory neurons, excitatory neurons, microglia, astrocytes, mature and immature oligodendrocytes, ependymal cells, endothelial cells, and mural cells, some of which contain further sub-partitions. Again, we observed a few cell clusters (~3% of all cells) that co-expressed markers for more than one cell class and thus were candidates for putative doublets, i.e. adjacent cells that our cell segmentation algorithm failed to separate. We utilized the same criteria for doublet selection as we did with the scRNA-seq data to identify and remove these doublets. However, one advantage of MERFISH is that we have an image of every member of each of these clusters. Thus, we visually inspected a random subset of cells within each of these putative doublet clusters and confirmed that indeed these clusters were enriched in connected cells that were not separated by our automated segmentation algorithm.

We also identified a few clusters of cells (~1% of all cells) that had very low average intensities of DAPI and polyA as well as low total RNA counts, suggesting that these segmented features did not represent real cells or, perhaps, represented very small cell fragments. Indeed, visual inspection of members of these groups revealed that they appeared to be small cell fragments or spurious fluorescence signals in the DAPI and polyA channels. These groups were also removed from subsequent analysis.

In this clustering analysis, we observed sub-divisions of several of the cell classes, as we did with the scRNA-seq data. We manually aggregated the sub-divisions of non-neuronal cell classes for all subsequent analysis. To further distinguish transcriptionally distinct clusters of inhibitory and excitatory neurons, we performed a second round of clustering on the inhibitory neurons and excitatory neurons separately. We also identified clusters of neurons that were enriched in markers associated with non-neuronal cells (~10% of all neurons), and we assigned these groups as putative doubles of neuronal and non-neuronal cells. Again, visual inspection of members of these groups supported this assignment. To formalize this visual assessment of putative doublets, we created a polarization metric, defined as the distance between the average center of mass of all observed RNAs for the neuronal markers, i.e. Gad1, Syt4, Slc17a6, and the center of mass of all observed RNAs for a panel of non-neuronal markers (Ndrg1, Lpar1, Ermn, Gjc3, Opalin, Sgk1, Ttyh2, Sox6, Traf4, Pdgfra, Cspg5, Aqp4, Aldh1l1, Lmod1, Myh11, Slco1a4, Fn1, Selplg, Slc15a3). Indeed, we found that this polarization metric was larger for these putative doublet clusters than for all other clusters. We removed these doublet clusters from subsequent analysis.

We then assigned a basic ID to each neuronal cluster comprised of an ‘I’ or ‘E’ to mark inhibitory or excitatory clusters, determined based upon the relative expression of the inhibitory marker Gad1 or the excitatory marker Slc17a6, and a number indicating the relative cell abundance of that cluster with respect to all other inhibitory or excitatory clusters, e.g. E-1 and I-1 are the most abundant excitatory and inhibitory clusters, respectively. Of these clusters, we identified one (I-28) as a putative hybrid cluster based on the co-expression of inhibitory and excitatory markers and renamed it as H-1, which showed correspondence to the scRNA-seq cluster h2. Again, we did not relabel the other clusters, and hence the cluster ID I-28 is missing from our final dataset. In total, this analysis identified ~40 inhibitory and ~30 excitatory neuronal clusters after removal of putative doublets.

MERFISH data from naïve male and female animals, as well as animals subjected to defined behavioral stimuli (described below), were clustered together. The expression level of

cFos, which was measured only for the animals exposed to behavioral stimuli, was not included in clustering. With the exception of **Fig. 8** and **fig. S20**, all reported analyses on MERFISH data were conducted with the data derived from naïve animals.

tSNE visualization of MERFISH data were performed as described for the scRNA-seq data, using the same principle components calculated from all 155 genes that were used to cluster cells. Marker genes that were statistically enriched within specific MERFISH clusters were identified by comparing the expression level of a given gene (in normalized RNA density, as described above) within cells of the given cluster relative to the expression level of this gene in cells of all other clusters using a two-sided t-test. An enrichment was judged to be significant if it had a p-value less than 0.05 after correction for a false discovery rate of 5%. Genes included within the violin plots for **Fig. 6-8** and within the dot plots in **fig. S15**, were drawn from the set of statistically enriched genes for each of the listed MERFISH cluster. However, these figures also include genes of interest that were expressed within these clusters but not necessarily statistically significantly enriched.

Nomenclature of MERFISH neuronal clusters. In addition to the cluster ID (E-1, E-2, …, I-1, I-2, …, and H-1) as described above, we also assigned a descriptive name to each neuronal cluster on the basis of the spatial measurements of these neuronal clusters. We determined the nuclei in which these clusters were primarily enriched (**Fig. 5C**). We then assigned names to individual neuronal clusters based on this nuclei assignment. For the small fraction of dispersed clusters that were not enriched in specific nucleic, we provided a gene-based name on the basis to two marker genes. We note that I-17 was found at the very edge of our imaged region and falls outside of the boundaries of the nearest imaged nuclei, the VLPO. Given its location, it is likely located within the HDB, a nuclei which is immediately next to but largely outside of the imaged region (**fig. S17, S18**). For this reason, we gave this cluster a gene-based name with HDB in the parenthesis (I-17:Tac1 (HDB).

Correspondence between MERFISH and scRNA-seq

To identify neuronal clusters identified with scRNA-seq and MERFISH that have similar expression profiles and, thus, might represent corresponding clusters of cells, we first computed the pairwise Pearson correlation coefficient between the average expression profiles of each MERFISH neuronal cluster and each scRNA-seq neuronal cluster (**Fig. 4C**; **fig. S14A**). The expression profiles for both groups of clusters were measured in z-scores. We then utilized this correlation analysis to identify the scRNA-seq cluster(s) that correlated the most strongly with each MERFISH cluster (**Fig. 4D**; **fig. S14C**). Specifically, we first identified the scRNA-seq cluster with the largest correlation coefficient for each MERFISH cluster. To then identify correlation coefficients that were statistically indistinguishable from these maximum values, we performed a bootstrap analysis. Specifically, we computed the average expression profiles for all MERFISH clusters for a random ~75% of the replicates of the anterior and posterior portions of the preoptic region. Similarly, we computed the average expression profiles for all scRNA-seq clusters from a random ~75% of the scRNA-seq replicates. We then computed the pairwise Pearson correlation coefficients between the average expression profiles for these bootstrap replicates, as described above. For each MERFISH cluster, we identified any scRNA-seq cluster(s) that had a correlation coefficient larger than that of the scRNA-seq cluster with the largest correlation coefficient in the original analysis. If a scRNA-seq cluster had a larger correlation coefficient in at least 10% of the bootstrap replicates, it was considered to be

statistically indistinguishable from the largest value, and this cluster was included as an additional, putative corresponding cluster to that MERFISH cluster.

To remove potentially spurious correspondences, we determined the null probability distribution of maximum Pearson correlation coefficients when no correspondence exists. We performed an analysis in which we randomized the gene labels associated with the average expression profiles for each scRNA-seq cluster and recomputed the pairwise Pearson correlation coefficients between scRNA-seq (after randomization) and MERFISH clusters. From this analysis, we found that there was a 10% probability of generating a maximum correlation coefficient greater than 0.34 when no correspondence exists, and we removed any putative corresponding clusters with a correlation coefficient less than or equal to 0.34.

In parallel, we trained a neural network classifier (single-layer perceptron, 100 nodes) on the MERFISH data and predicted the MERFISH neuronal cluster identity for each neuron measured with scRNA-seq (**fig. S14D**). For the vast majority of clusters, the scRNA-seq cluster that best correlated in its expression profile with a given MERFISH cluster was also the scRNA-seq cluster that was most frequently assigned the label associated with that MERFISH cluster via this classifier (**fig. S14A, D**), providing further support for the putative correspondence constructed above.

We utilized the results from the classifier approach in two additional ways. First, we removed any putative correspondence between a MERFISH cluster and a scRNA-seq cluster (determined using the correlation approach) for which this classifier did not provide significant support for this correspondence, i.e. <10% of cells assigned a given MERFISH label were found within the putative corresponding scRNA-seq cluster. Reflecting the agreement between the classifier results and the correspondence determined via correlation, this cut removed only ~10% of putative correspondences identified via correlation. The final set of putative correspondence between MERFISH and scRNA-seq clusters with their Pearson correlation coefficients is included in **Fig. 4D**, **fig. S14C** and **Table S9**.

Second, to provide further support of the correspondence that we drew between MERFISH and scRNA-seq clusters based on correlation, we identified the scRNA-seq cluster most closely associated with a given MERFISH cluster based on these classifier results. Specifically, for each MERFISH cluster, we identified the scRNA-seq cluster to which this MERFISH cluster label had been assigned most frequently. To also include assignments that were statistically indistinguishable from this maximum frequency, we adopted a similar bootstrap approach as described above for the correlation approach but applied it to the measured frequency with which this MERFISH cluster label was applied to each scRNA-seq cluster. We also required that a given scRNA-seq cluster comprise at least 10% of the cells that were given this MERFISH cluster label. From this analysis we found a small number of putative correspondences between MERFISH clusters and scRNA-seq clusters for which the Pearson correlation coefficient between their average expression profiles was lower than our false-positive threshold of 0.34, described above. We, thus, removed these elements, which reduced the number of putative correspondences by only ~25%. The resulting set of correspondences between MERFISH and scRNA-seq clusters produced from this classier analysis (**fig. S14E**) agreed well with that produced by the correlation analysis above (**fig. S14C**).

Because our scRNAseq measurements were performed over a larger tissue volume than our MERFISH measurements, several scRNA-seq clusters can be readily determined to be located in structures outside of the region imaged by MERFISH. Specifically, we identified these clusters as being outside of the MERFISH imaged region by using the most specific markers for each

cluster and examining coronal images from the Allen Brain Atlas (*35*) and our own in situ hybridization data for these marker genes. Marker genes of clusters i6, i14, i28, i30, and i31 place these cells in the SCN, marker genes of clusters e20, e21, h3, and i27 place them in the PVN, and marker genes for cluster i3 place it to the olfactory tubercle (OT), which were all outside of the region imaged via MERFISH. Marker genes of clusters i35 and i44 place these cells to the LPO and MnPO, respectively. Although MnPO and LPO were partially covered in our imaged region, our in situ hybridization data revealed that neurons co-expressing markers of these clusters were found in portions of these nuclei outside our imaged region. Marker genes for clusters e9, i42 and i43 placed these cells in the HDB; only a small portion of this nuclei was included in our imaged region. Thus, we did not seek a correspondence between MERFISH clusters and these scRNA-seq clusters (i3, i6, i14, i27, i28, i30, i31, i35, i42, i43, i44, h3, e9, e20, and e21). We note that it is possible that some of the remaining scRNA-seq clusters were still outside the MERFISH imaged region because the some of these clusters had no markers that were present in the Allen Brain Atlas and, thus, we were unable to estimate their location.

To select galanin (Gal)-enriched MERFISH clusters for the purpose of comparison to galanin-enriched scRNA-seq clusters, we selected MERFISH clusters with an above average and statistically significant enrichment in Gal relative to all cells (as determined via a two-sided t-test with a p-value cut-off of 0.05). In examining the corresponding scRNA-seq clusters to the Gal-enriched MERFISH clusters, we determined that one MERFISH cluster (E-28), which was not identified as Gal-enriched but had the next highest level of Gal expression, corresponded to a Gal-enriched scRNA-seq cluster. Similarly, we found that the scRNA-seq cluster e3, which corresponded to a Gal-enriched MERFISH cluster, also had a high but not statistically enriched level of Gal expression. Because these two clusters expressed appreciable levels of Gal and each showed putative correspondence to a Gal-enriched cluster measured by the other approach, they were included in the analysis shown in **Fig. 4E**.

For comparison purpose, we also partitioned our scRNA-seq neurons into two random sets each comprised of half of the measured cells (bootstrapped replicates). We then re-clustered these bootstrapped replicates of neurons using the same protocols described above, computed average expression profiles for each of the newly formed clusters in each data subset, and calculated the pairwise Pearson correlation coefficients and determined the most similar clusters as described above. We found that the number of corresponding clusters as well as the range of correlation coefficients observed for these most similar clusters (**fig. S14F, G**) is comparable to or slightly higher than what we observed between the most similar MERFISH and scRNA-seq clusters (**Fig. 4C, D**; **fig. S14A, C**).

Spatial mixing and spatial heterogeneity analysis of neuronal clusters

Spatial mixing analysis of neuronal clusters. To compute the neighborhood composition for neurons, we determined the number of neurons and the number of unique neuronal clusters within a given radius around each neuron as a function of the radius size. We then computed the fraction of cells within this radius that correspond to the most abundant cluster.

Spatial heterogeneity analysis within individual clusters. To explore the spatial heterogeneity of gene expression within individual clusters, we performed a principal component analysis of the expression of all genes within each cluster. Expression was measured in z-scores with z-score calculated for individual genes across all cells within that cluster. We then plotted the spatial distribution of these cells colored by the amplitude of one of the first two principal components (PC). **Figure S16** contains several examples of the clusters in which there was a clear spatial

gradient in one PC. While we chose to display only a single PC in these examples, we note that spatial correlations were observed in the amplitude of multiple different PCs in some clusters.

Overlap fraction analysis of neuronal clusters

To determine the fractional overlap between different neuronal clusters, we first computed the local density of each neuronal cluster by convolving the location of each cell for each neuronal cluster with a gaussian distribution of 25-μm width. We then discretized this density into 25-μm × 25-μm bins, effectively creating a pixelated image of this density. We eliminated the effect of outlier cells by thresholding this discretized density distribution on a minimum density of 1 cell per 25-μm × 25-μm pixel. The overlap fraction was calculated between all pairs of clusters as the total number of pixels above threshold for both clusters divided by the total number of pixels above threshold for either cluster. The measured overlap fraction was not strongly sensitive to the bin width or the density threshold. Only the overlap fractions between the aromatase- and Esr1-enriched clusters are reported in **Fig. 6F**.

Comparison of spatial measurements of neuronal clusters between biological replicates

We first calculated the density of individual neuronal clusters in all slices of all replicates, as described above. Then for each neuronal cluster in each tissue section, we computed the area covered by the cluster as the number of pixels above threshold, as described above, and determined the variation of this area across all replicates. The area of clusters varied across replicate animals by only ~10% on average (averaged over all clusters). In addition, we also calculated the relative variation in the overlap fraction between all pairs of neuronal clusters. For each pair of neuronal clusters in each tissue section, we computed the overlap fraction as described above and then determined the relative variation in this value across all replicates. Again the relative variation in the overlap fraction across replicate animals was only ~10% on average (averaged across all clusters). Finally, we computed the distance between the centroid positions of each pair of neuronal clusters in each tissue section, determined the variation in these values across replicate animals, and also found that centroid distances varied across replicate animals by ~10% (median value across all cluster pairs). Thus, in general, we find that the area of each neuronal cluster, the overlap fraction between clusters, and the distance between cluster centroids is well reproduced between replicate animals. Only clusters that contained at least 25 cells in a given slice were used for these calculations.

Behavioral assays for cFos measurements

Animals were individually housed for ~4-5 days before behavioral assays, which were performed as described previously (*5*, *72*, *74*), beginning >1 hour after the onset of the dark cycle under dim red lighting. Mice were habituated to the testing room for 1 hour before the start of each behavioral assay. Only mice that performed the desired behaviors (described below) after the defined behavioral stimuli were selected for tissue collection.

Parental behavior. Virgin females: A single mouse pup aged 2-3 days was introduced into one corner of the home cage of an adult virgin female mouse (age ~8 weeks), away from the nesting material. A parental response was defined as retrieval of the pup to the nest combined with nesting, crouching, grooming and licking behaviors. Mice were sacrificed for tissue harvesting 30 minutes following the retrieval of the pup to the nest. Only mice displaying all the above behavioral subroutines were selected for further processing.

Mothers and Fathers: Virgin male and female mice (age ~7 weeks) were co-housed in pairs until the birth of pups (~21 days). Mice were allowed to remain with pups for 3 days before mothers were moved to a fresh cage and pups were removed from the home cage of the father. Two days later, parental behavior was tested as described above with virgin females.

Pup-directed aggression. A single mouse pup aged 2-3 days was introduced into one corner of the home cage of an adult virgin male mouse (age ~8 weeks), away from the nesting material. Mice were sacrificed for tissue harvesting 30 minutes following aggressive attacking of the pup, characterized by biting behavior leading to audible pup vocalizations and occasionally preceded by tail rattling. Pups were immediately removed from the cage following an attack.

Inter-male aggression. Castrated C57Bl6/J males (aged ~6-8 weeks) swabbed with intact adult male urine (~100 μL) were introduced into the home cage of an adult virgin male mouse (age ~8 weeks). Mice were sacrificed for tissue harvesting 30 minutes following aggressive attacking of the intruder, characterized by rapid bouts of biting leading to defensive reactions by the intruder (rearing and escaping). Intruder males were removed from the cage 10 minutes after the commencement of the first attack.

Male mating. Ovariectomized female mice (aged ~6-8 weeks) were hormonally primed to be in estrous as previously described (*41*) (injected subcutaneously with 10 μg of 17 β-estradiol benzoate in 100 μL sesame oil on 2 consecutive days preceding the test day). On the test day, females received a subcutaneous injection of 50 μg of progesterone in 50 μL sesame oil at least 6 hours before testing and were introduced into the home cage of an adult virgin male mouse (age ~8 weeks). Mice were sacrificed for tissue harvesting 30 minutes following intromission of the female. Intromission usually occurred within 5 minutes of the first mount. Female mice were removed from the cage 10 minutes after the first intromission. None of the male mice ejaculated during this short testing period.

Female mating. An intact adult sexually experienced male mouse was introduced into the home cage of an adult virgin female mouse (age ~8 weeks) in estrous (determined by a vaginal smear). Females were sacrificed for tissue harvesting 30 minutes following intromission by the male. Males were removed from the cage 10 minutes after the first intromission. None of the male mice ejaculated during this short testing period.

Calculation of cFos-positive-cell enrichment

To determine the enrichment of cFos-positive cells within each cluster, we first defined neurons as cFos-positive if the cFos signal within those cells was greater than two standard deviations above the mean cFos value for all neurons. We then calculated the enrichment of cFos-positive cells within each cluster for each behavior and each replicate as the fraction of cells within that cluster that were cFos-positive over the fraction of all neurons in all replicates that were cFos-positive. We reported the average of this enrichment as well as the standard deviation over all replicates for each behavior in **Fig. 8**. To determine enrichments that were statistically significant, we used the binomial distribution to calculate the probability of observing a given fraction of cFos-positive cells within a cluster given the probability of observing a cFos-positive cell across all neurons. We selected a p-value cut-off using a false-discovery rate of 5%. Clusters highlighted in red in **Fig. 8** and **fig. S20** satisfied this p-value cut-off and had an average enrichment greater than 1. In addition, we excluded clusters from this analysis that were not present in sufficient numbers to allow reliable measurements of the presence of a small fraction of cFos-positive cells. Specifically, we only analyzed clusters in which at least 10 cells were present in at least two of the 3-4 replicate animals.

Two- and three-color FISH measurements by RNAscope

Double- and triple-label FISH were performed using the RNAscope assay V2 kit (Advanced Cell Diagnostics, 32110). Brains from naïve animals, or brains from animals following the relevant behavioral stimulus, were harvested and frozen in OCT as described above. 16-µm sections encompassing the preoptic region were cut as described above and placed on microscope slides (VWR, 48311-703) chilled to -18 °C. These slides were then stored at -80 °C prior to staining. RNAscope staining was performed according to the manufacturer's instructions.

Supplementary Figures

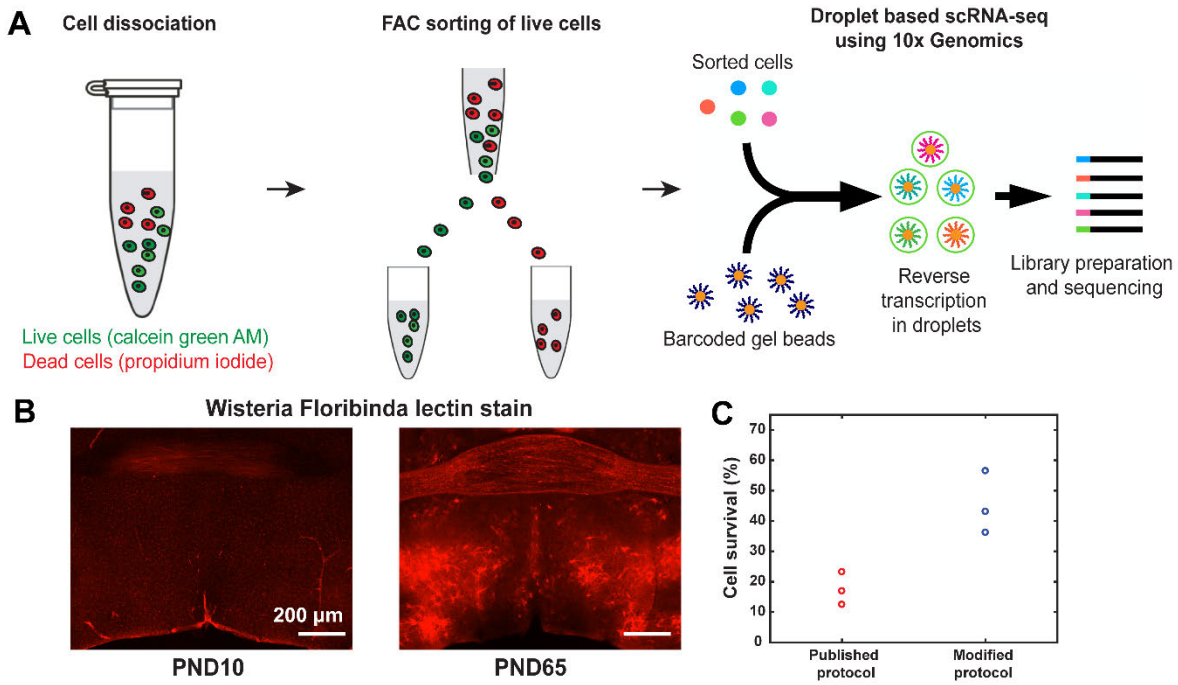


Fig. S1. Sample preparation for single-cell RNA-seq. (A) Schematic depiction of the process by which cells are dissociated and stained with markers of live (Calcein green AM) and dead (Propidium iodide) cells, FACS sorted to enrich for live cells, and encapsulated in droplets with barcoded beads using the 10x Genomics platform in order to produce single-cell cDNA libraries for next-generation sequencing. **(B)** Wisteria floribunda lectin staining from the preoptic region shows an increase in perineuronal nets in adult (post-natal day 65; PND65), compared to young mice (PND10). Because of the accumulation of perineuronal nets during maturation, we included enzymes that digest extracellular matrices to increase the cell survival rate during dissociation. **(C)** The measured fraction of cells that survived dissociation for three replicates prepared with previous (*70*) (red) or our improved protocol (blue).

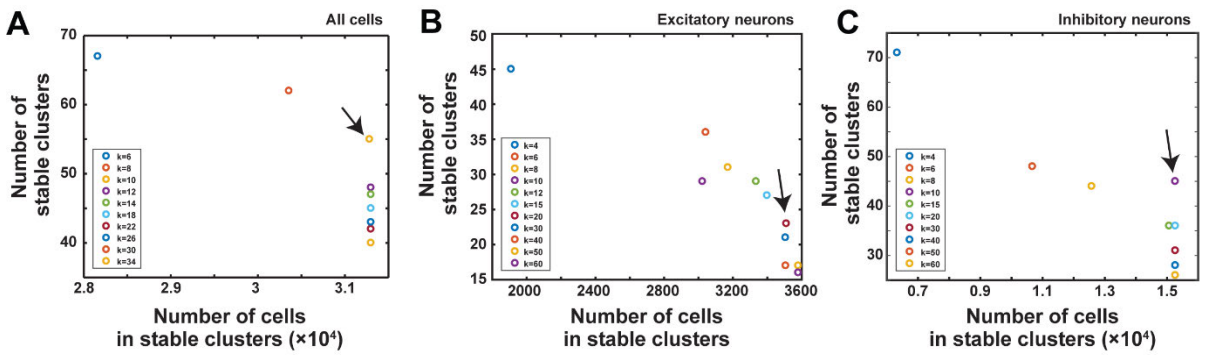


Fig. S2. Determination of the nearest neighbor number (*k*) used in the Jaccard-Louvain community detection approach for cell clustering. The number of clusters identified by the Jaccard-Louvain community detection algorithm depends on the number of nearest neighbors, *k*, used to construct the nearest neighbor graph upon which this algorithm operates. Similarly, the stability of these clusters—the fraction of cells that are co-clustered together when random subsets of cells are re-clustered (bootstrap analysis)—also depends on this value. We select a value of *k* that balances the desire to have more clusters with the desire to have most measured cells assigned to stable clusters. (**A-C**) The number of stable clusters versus the number of cells assigned to stable clusters (determined using the bootstrap analysis) as a function of the number of nearest neighbors (*k*) used for determining distinct cell clusters for all cells (A), for excitatory neurons only (B), and for inhibitory neurons only (C). Arrows indicate the *k* values selected, below which the number of cells included in the stable clusters decreases substantially with decreasing *k*.

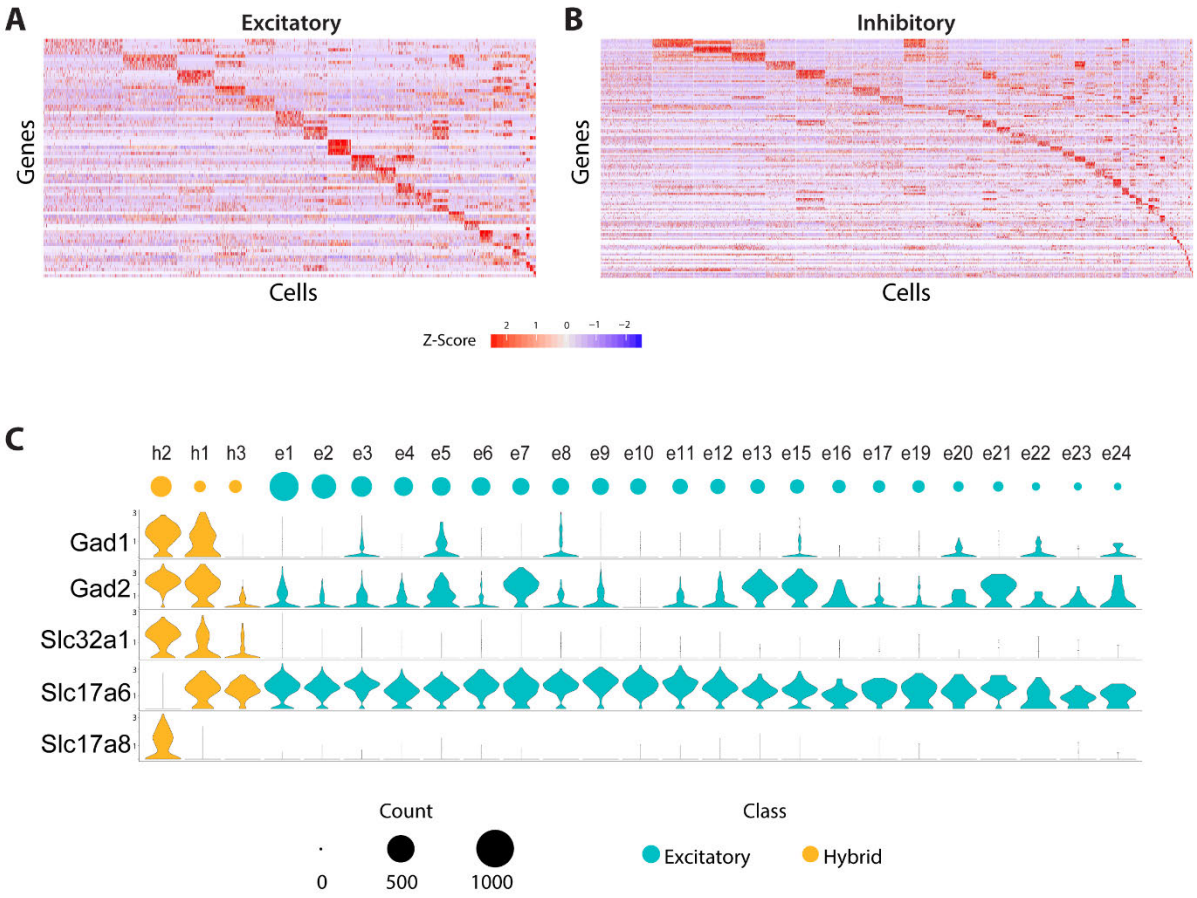


Fig. S3. Expression of the top 5 marker genes in all neuronal clusters identified via scRNA-seq and the distribution of inhibitory and excitatory markers in excitatory and hybrid neuronal clusters. (**A, B**) Expression profiles for all excitatory (A) or inhibitory neurons (B) with individual cells displayed in columns and individual genes in rows. The depicted genes represent the set of the top five most enriched genes within each cluster (on the basis of averaged normalized expression fold-change). Expression profiles are measured in z-scores of the normalized expression values for individual genes across all cells. The normalized expression value for a gene is computed by normalizing to each cell's total UMI count, multiplied by 10,000 (for scale), and log-transformed. Z-scores are computed by subtracting the gene's mean expression (across all cells) and dividing by the gene's standard deviation in expression. (**C**) Normalized expression distributions of inhibitory (Gad1, Gad2, and Slc32a1) and excitatory (Slc17a6, Slc17a8) markers in the identified hybrid and excitatory clusters. Expression distributions are calculated on logarithmically scaled counts, as described above, and the scale of each individual row is set by the 95% quantile of the cluster that expresses the largest level of that gene. Expression values for these genes in these clusters are included in **Table S2**.

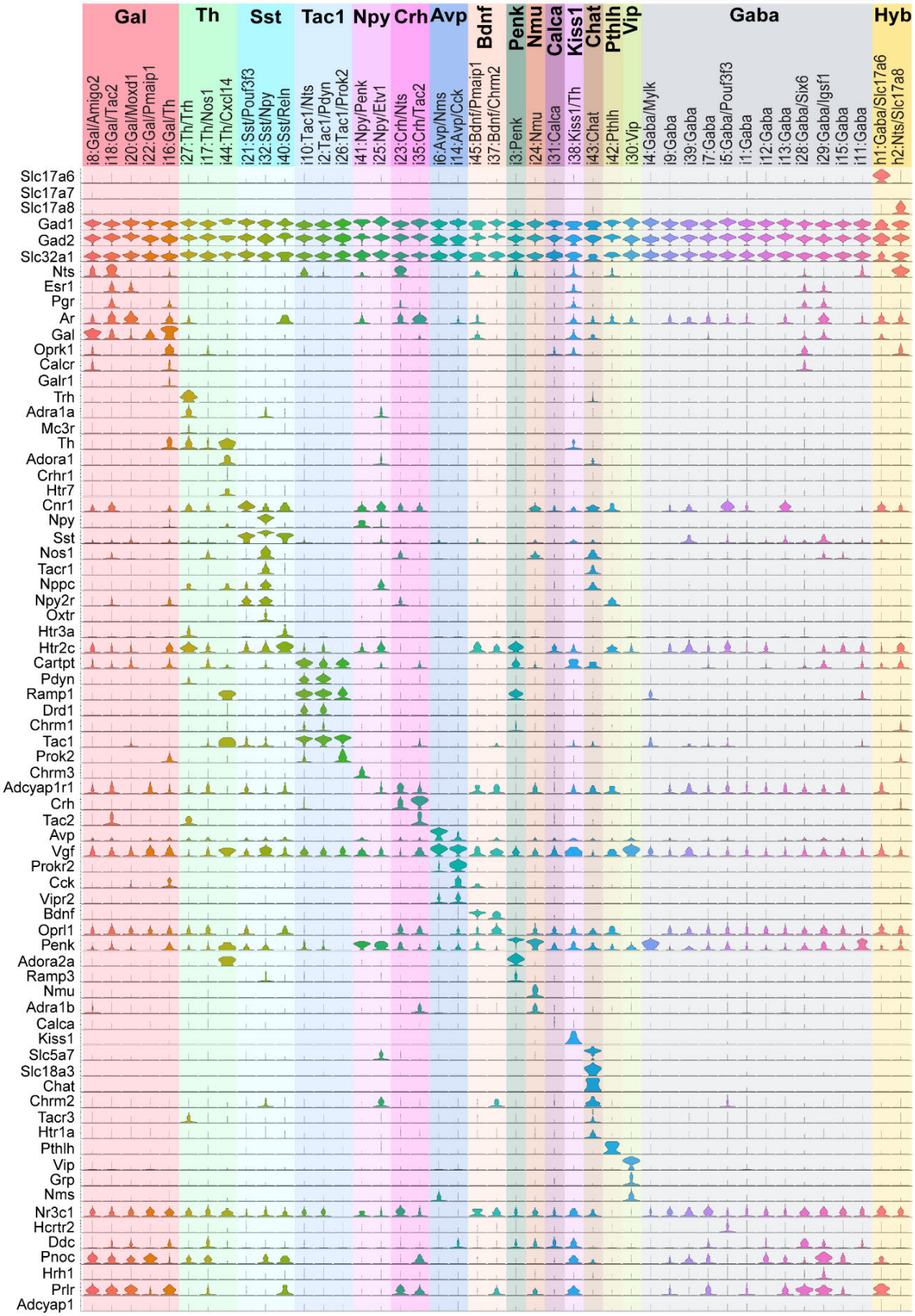


Fig. S4. **Expression distributions of notable genes in inhibitory neurons identified via scRNA-seq.** Expression distributions for notable neuropeptides, neuromodulators, and

peptide/hormone receptors in all inhibitory neuronal clusters. All displayed genes were enriched in at least one inhibitory cluster (as determined via MAST (*75*) with a false-discovery rate of less than 0.01). Neuronal clusters are grouped based on the expression of neuropeptides and neuromodulators designated by the same colors as in **Fig. 1C**. Expression distributions were calculated as in **Fig. 2**. Expression values for these genes in these clusters are included in **Table S2**. The hybrid neuronal clusters, h1 and h2, are listed here since they were identified after the second round of clustering of inhibitory neurons.

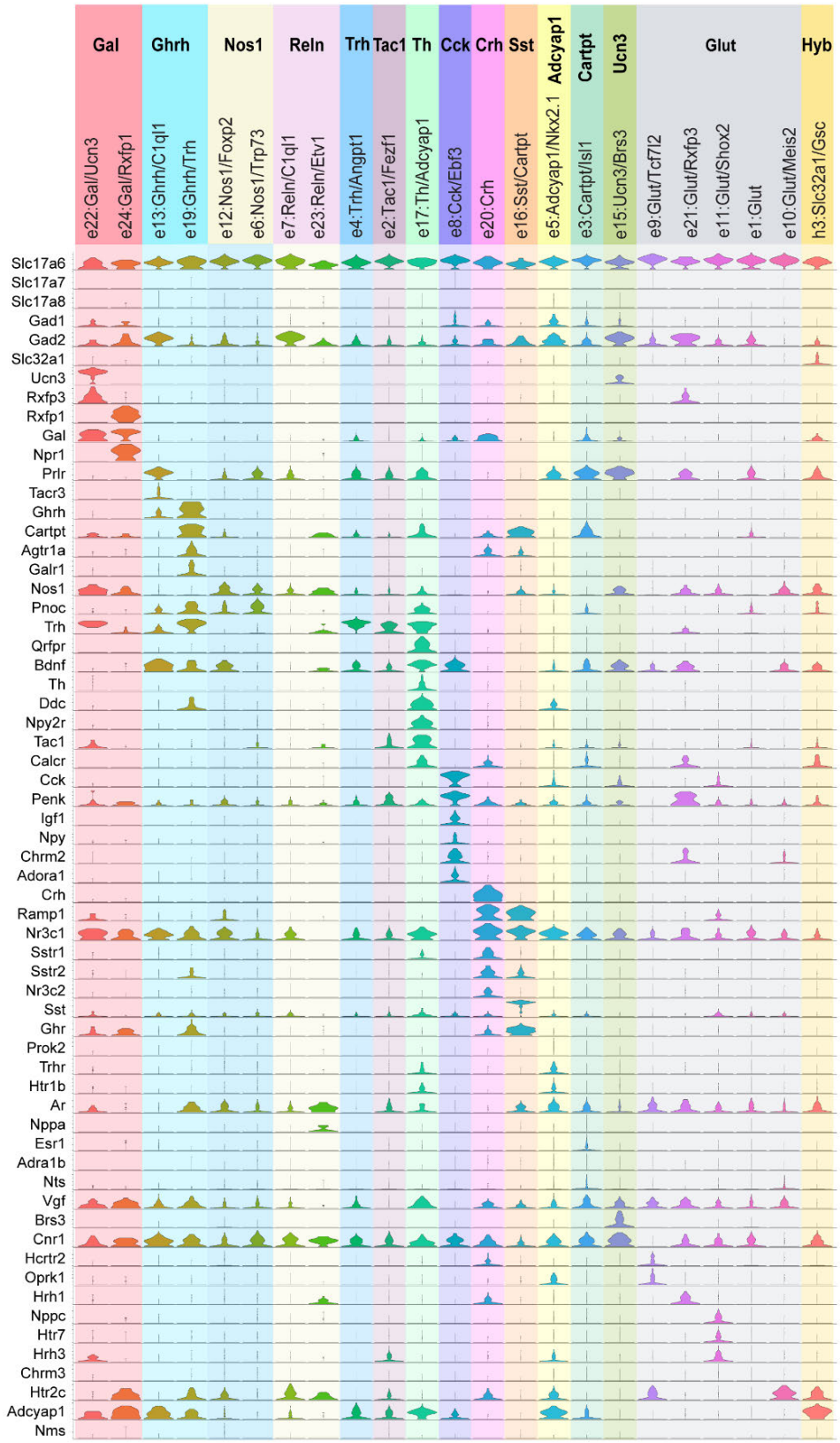


Fig. S5. **Expression distributions of notable genes in excitatory neurons identified via scRNA-seq.** Expression distributions for notable neuropeptides, neuromodulators, and

peptide/hormone receptors in all excitatory neuronal clusters. All displayed genes were enriched in at least one excitatory cluster (as determined via MAST (*75*) with a false-discovery rate of less than 0.01). Expression distributions were calculated as in **Fig. 2**. Neuronal clusters are grouped based on the expression of neuropeptides and neuromodulators designated by the same color as in **Fig. 1D**. Expression values for these genes in these clusters are included in **Table S2**. The hybrid neuronal cluster, h3, is listed here since it was identified after the second round of clustering of excitatory neurons.

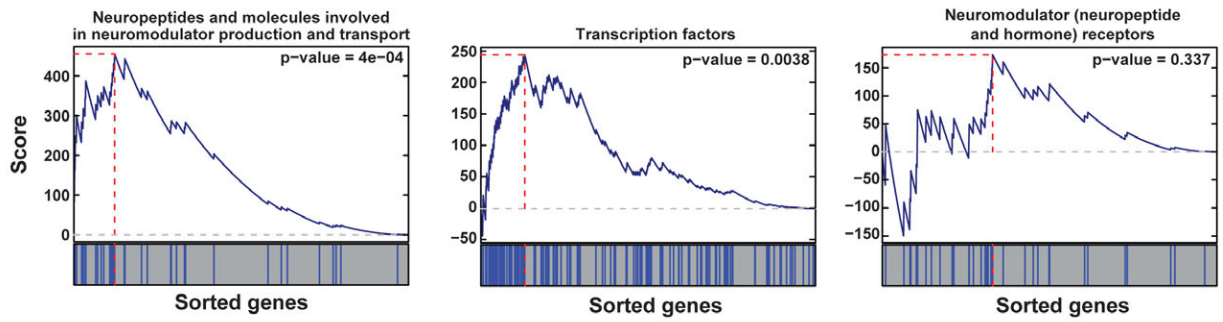
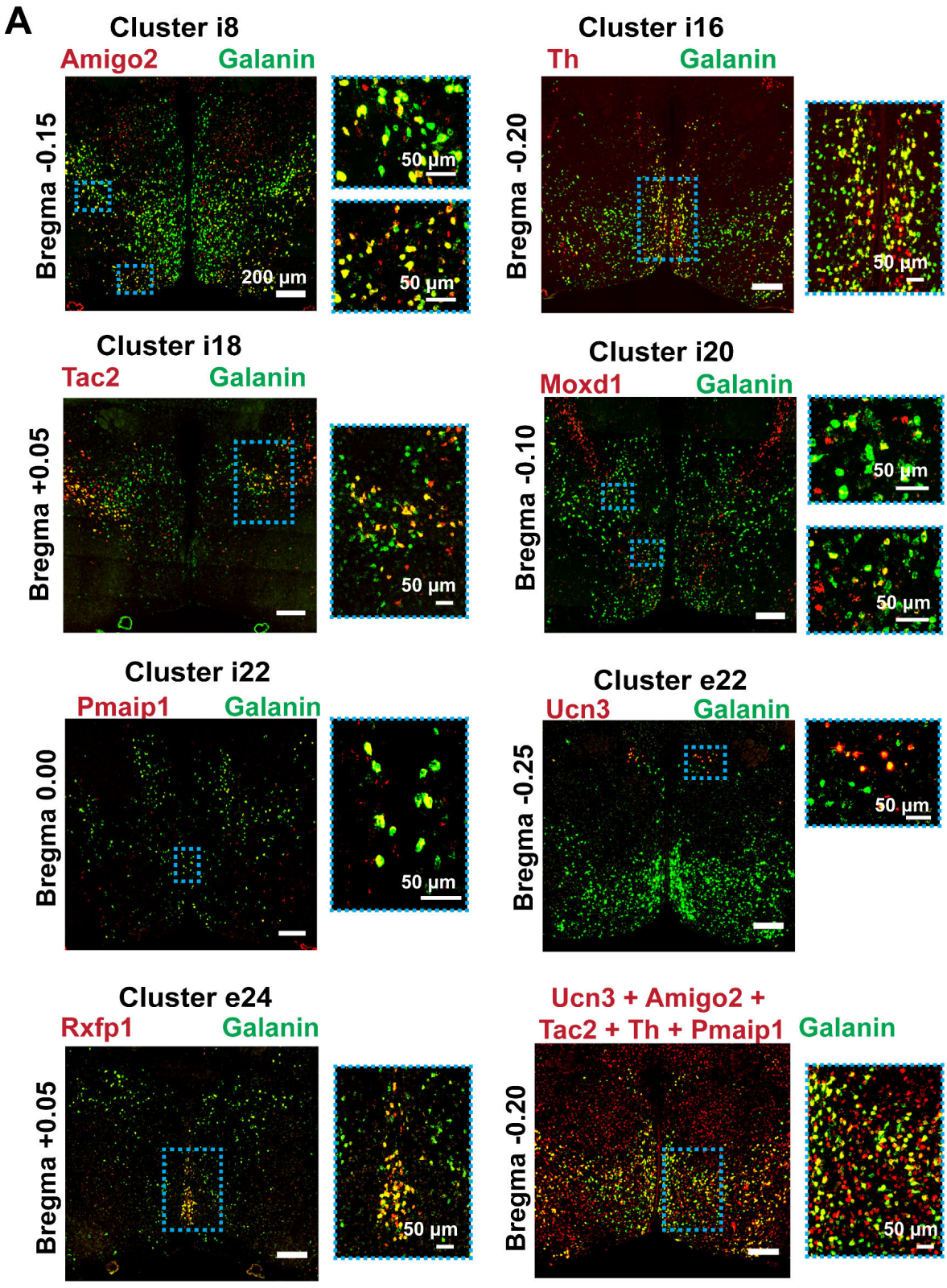


Fig. S6. **Gene-set enrichment analysis of three gene categories in the differentially expressed genes among the neuronal clusters determined by scRNA-seq.** We observed that different categories of genes appeared enriched in the top markers of neuronal clusters identified via scRNA-seq. To determine the significance of this enrichment, we performed a Gene-Set Enrichment Analysis (GSEA). The three categories of genes considered in this analysis are (1) neuropeptides and molecules involved in neuromodulator production and transport, (2) transcription factors, and (3) neuromodulator (neuropeptide and hormone) receptors. In this analysis, genes are sorted based on the strength with which they mark individual neuronal clusters identified in **Fig. 1** (*29*) and arrayed from strongest to weakest markers from left to right with genes that belong to the specified category marked by blue lines and other genes marked in grey (bottom). The score on the vertical axis is an accumulating weight that either increases or decreases depending on whether each gene does or does not belong to the specified category under consideration and progresses from the strongest to the weakest marker gene. The p-value represents the probability of observing the maximum value of the score (red dashed line) if the genes belong to the specified category are distributed randomly along the horizontal axis. The p-values for these three categories derived from this GSEA test are also reported in **Fig. 1E**. The genes included within each of these categories are provided in **Table S3**. On the basis of this calculation, the differentially expressed genes for distinct neuronal clusters are enriched for neuropeptides and molecules involved in neuromodulator production and transport, as well as for transcription factors, but not for neuromodulator receptors. However, we note that neuromodulator receptors are often expressed at relative low levels, which could potentially make the enrichment of these genes more difficult to detect.


A
Cluster i8
Amigo2
Galanin
Bregma -0.15
200 µm
50 µm
50 µm
Cluster i16
Th
Galanin
Bregma -0.20
50 µm
Cluster i18
Tac2
Galanin
Bregma +0.05
50 µm
Cluster i20
Moxd1
Galanin
Bregma -0.10
50 µm
50 µm
Cluster i22
Pmaip1
Galanin
Bregma 0.00
50 µm
Cluster e22
Ucn3
Galanin
Bregma -0.25
50 µm
Cluster e24
Rxfp1
Galanin
Bregma +0.05
50 µm
Ucn3 + Amigo2 +
Tac2 + Th + Pmaip1
Galanin
Bregma -0.20
50 µm

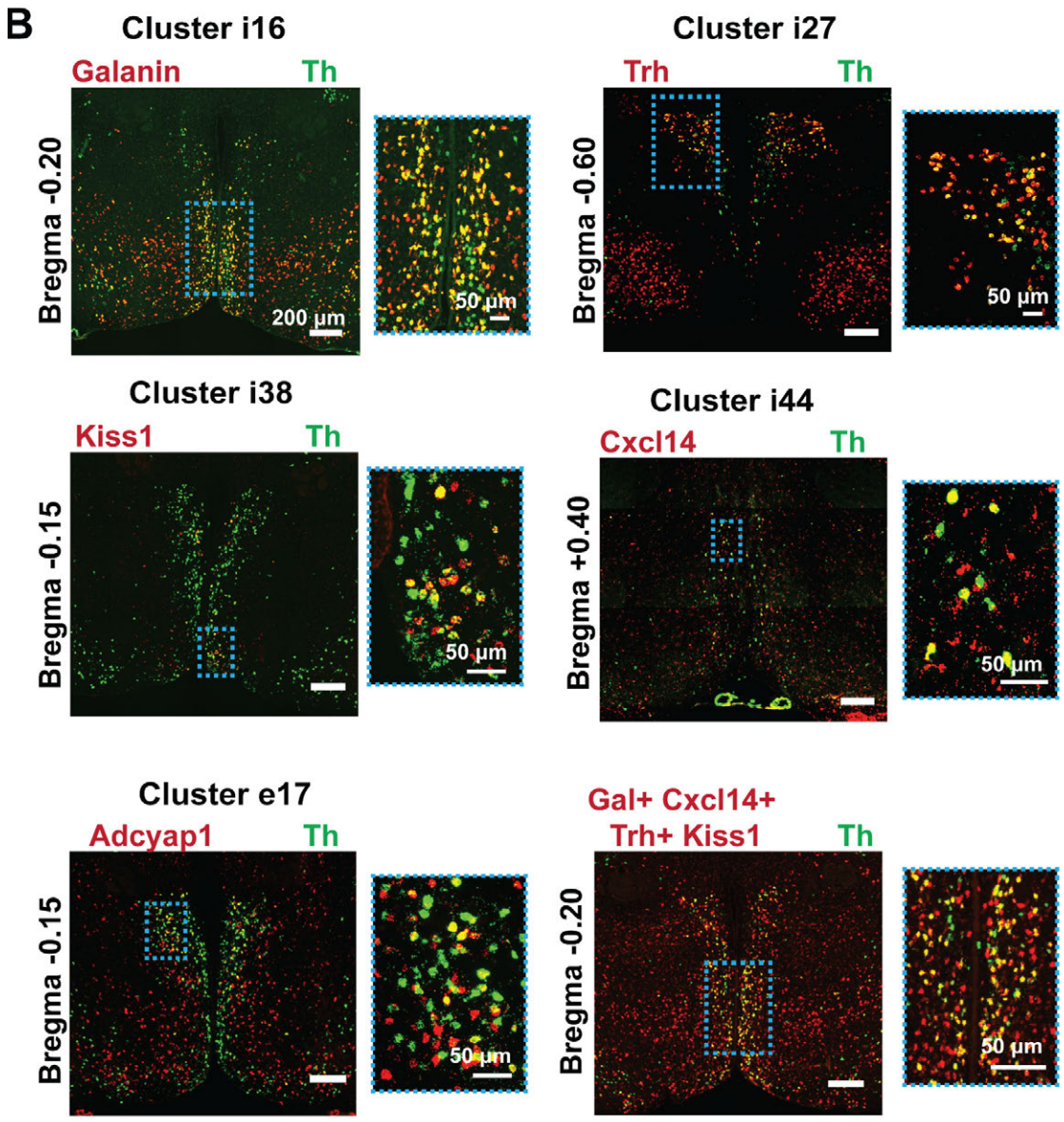


Fig. S7. In situ hybridization experiments to validate the co-expression of marker genes for galanin-enriched clusters and Th-enriched clusters identified by scRNA-seq. (**A**) Two-color FISH images in 16-µm-thick sections of the preoptic region show co-localization between Galanin (green) and another marker (red) for each Galanin-enriched cluster in **Fig. 2A**. Yellow represents co-expression of the markers. The blue dashed rectangles indicate the region of the zoom-in depicted to the right of each image. (**B**) As in (A) but for the Th-enriched clusters depicted in **Fig. 2B**. Markers are selected from the set of most enriched genes within each cluster. Only clusters that can be distinguished by a unique marker gene are tested by these experiments, hence the Th-enriched cluster i17 was not tested. Images of sections at different positions along the anterior-posterior axis are shown for different clusters to optimize the overlap between the two imaged marker genes in each case, and location in Bregma coordinates is provided. The images for the galanin- and Th-enriched cluster i16 in (B) were reproduced from (A) for clarity with the green and red colors exchanged to match the color scheme of (B). Scale bars: 200 µm. Scale bars for zoom-in: 50 µm.

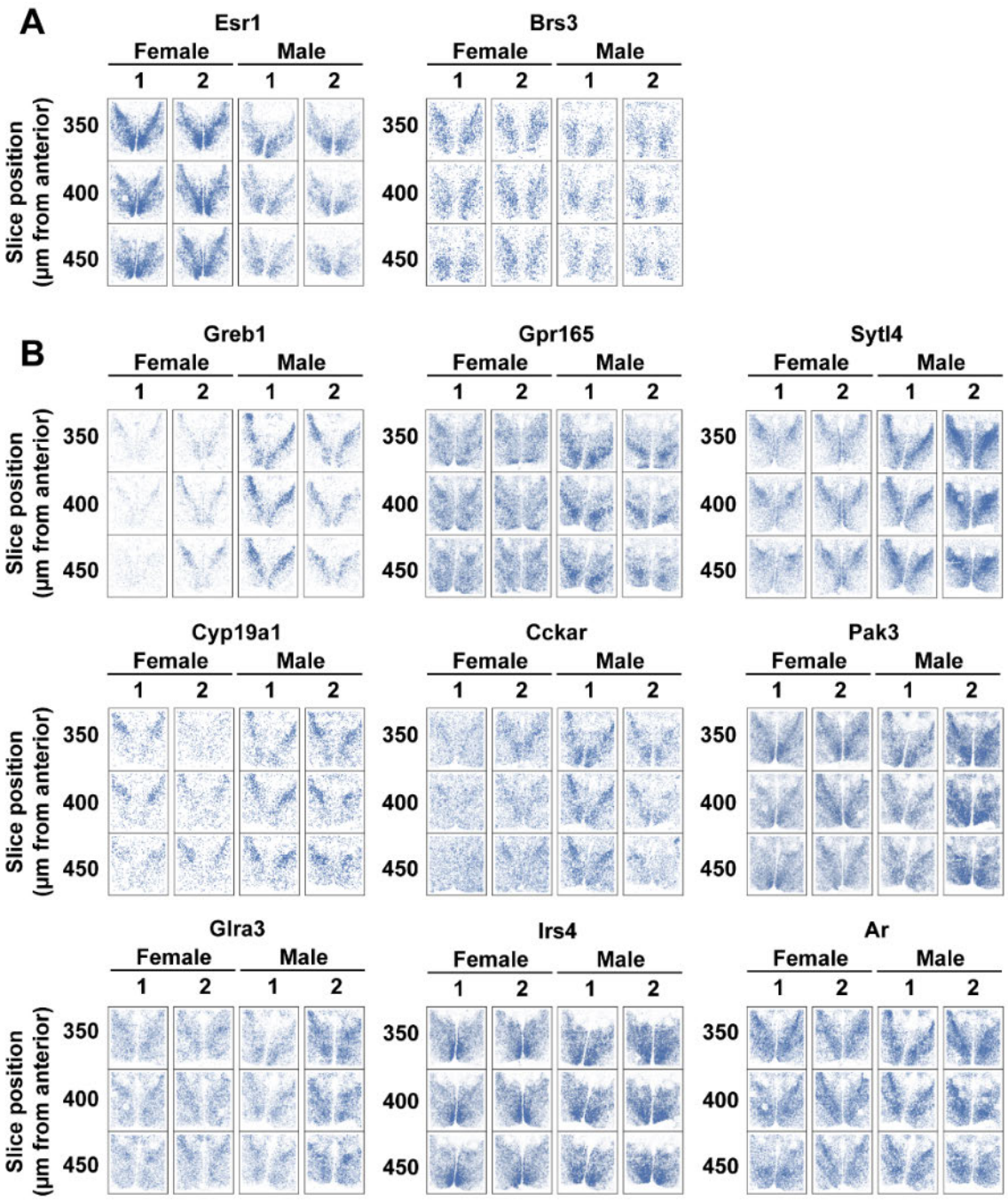
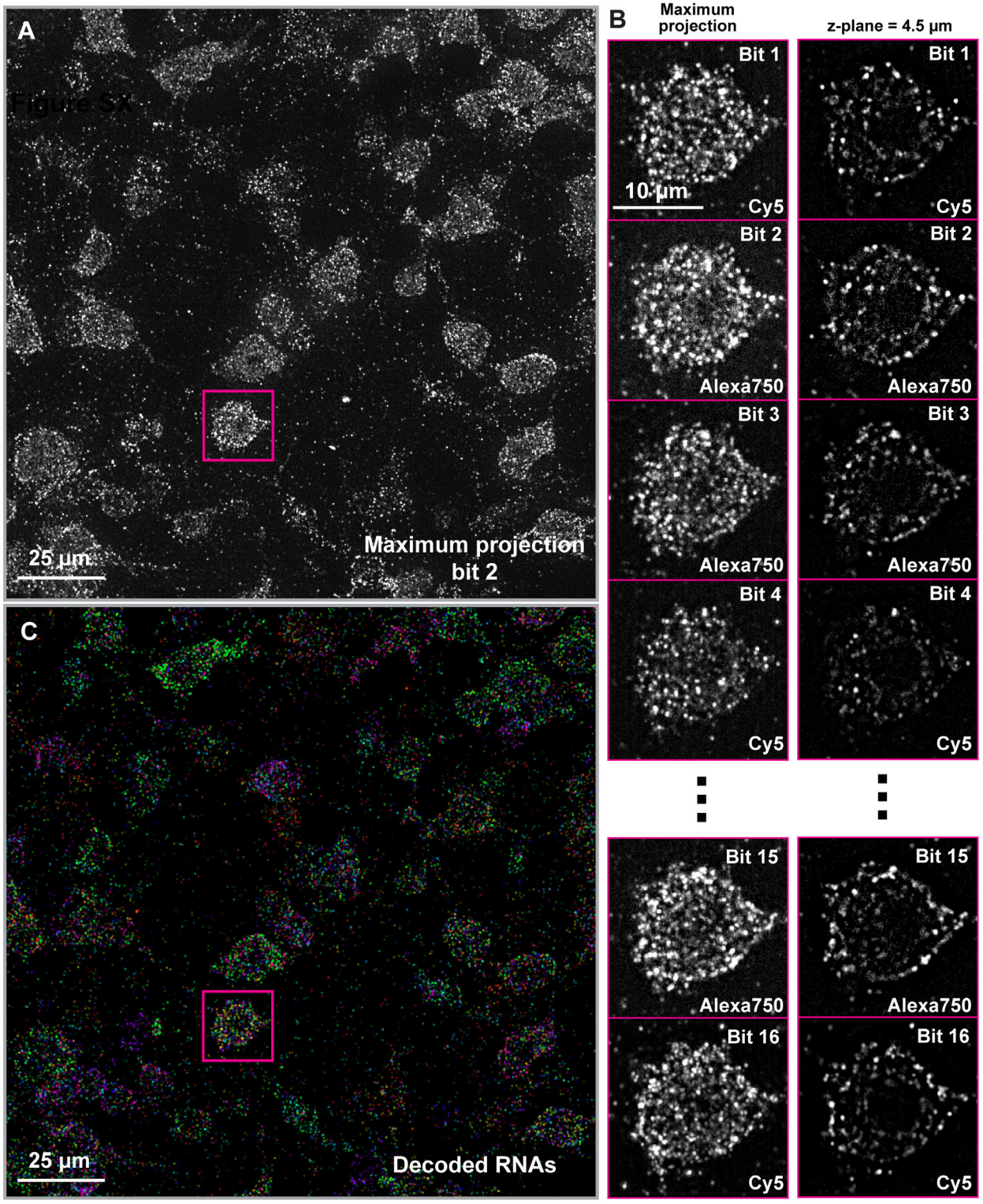


Fig. S8. Expression determined by MERFISH for 11 genes previously reported to display sexually dimorphic expression. (**A**) MERFISH images of RNAs for 2 genes that have been previously reported (*42*) to express at a higher level in females than in males in the preoptic region. (**B**) Same as (A) for 9 genes that have been previously reported (*41*, *42*) to express at a higher level in males than in females in the preoptic region. Three of the twelve 1.8-mm × 1.8-mm slices imaged (at indicted anterior-posterior positions) for two female and two male replicate animals are shown.


A
Figure SX
25 µm
Maximum projection
bit 2
B
Maximum projection
z-plane = 4.5 µm
Bit 1
10 µm
Cy5
Bit 2
Alexa750
Bit 3
Alexa750
Bit 4
Cy5
Bit 15
Alexa750
Bit 16
Cy5
C
25 µm
Decoded RNAs

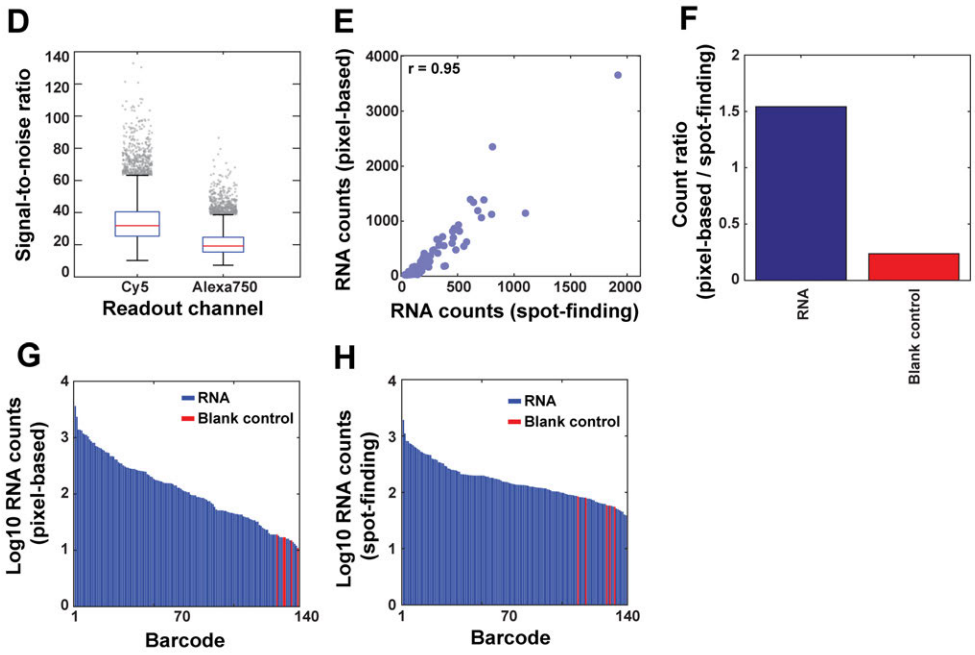


Fig. S9. Identification of RNA molecules from the combinatorial smFISH images. (**A**) A maximum projection image generated from the seven z-planes imaged for one representative field-of-view (FOV). Depicted is the image of one of the 16 imaged bits (two color channels (Cy5 or Alexa750) and eight hybridization rounds). (**B**) The maximum projection (left) or a single z-plane (right) images for the cell marked in the red box in (A) in 6 of the 16 bits that comprise the combinatorial smFISH measurements. The bit number and color channel associated with each image are indicated. Background has been subtracted in (A) and (B) with a high-pass filter. (**C**) Decoded identity of the RNA molecules identified in the FOV depicted in (A). Each colored dot represents the position of a single RNA molecule identified, the color indicates the identity of the gene, and the opacity is proportional to the brightness of the RNA signal. (**D**) Box plots of the measured signal-to-noise ratio for all RNA molecules across all rounds of hybridizations in combinatorial smFISH imaging in each color channel. The red line is the median value, the box indicates the 25th to 75th quantile range, and the whiskers cover the 99% quantile range. Individual dots represent measurements that fall beyond this range. (**E-H**) Identifying hundreds of millions of RNAs in 1.1 million cells requires computationally efficient algorithms. To this end, we used a pixel-based algorithm (*29*, *40*) that associates each pixel with a specific RNA barcode based on the intensity profile of that pixel across all 16 bits and then combines adjacent pixels assigned to the same RNA species into a single detected molecule. This pixel-based algorithm recently developed by us (*29*, *40*) is far more computationally efficient than our previous spot-finding based approach in which individual RNA molecules are localized with a spot-finding algorithm and then connected across bits to form barcodes (*20*). In addition to being computationally much faster, the pixel-based algorithm also has higher performance in terms of RNA identification efficiency and accuracy, as shown in (E-H). **(E)** The total counts observed for each RNA species in the FOV shown in (A) when RNAs were identified with the pixel-based approach versus the spot-finding algorithm. The results from the two algorithms are highly correlated with a Pearson correlation coefficient of 0.95. **(F)** The ratio of total RNA counts (blue) or total blank-control barcode counts (red) when RNAs were identified with the pixel-based algorithm relative to when RNAs were identified with the spot-finding algorithm.

Five barcodes were not assigned to RNAs (blank controls) and the detection of these barcodes provide an estimate of the false-positive detection rate. The pixel-based algorithm detects more RNA spots per cell (blue) and also has a substantially lower false-positive detection rate (red). **(G)** The total counts observed for each RNA species (barcode) for the FOV in (A) when RNAs were identified with a pixel-based algorithm. Barcodes are sorted from highest to lowest abundance. Red bars represent the observed counts for the five barcodes not assigned to RNAs (blank controls). **(H)** As in (G) but for RNAs identified with the spot-finding algorithm in the same FOV. For both algorithms, the blank-control barcodes are detected at lower rate than most RNA-encoding barcodes, except for the barcodes encoding the small fraction of RNA species with very low abundance, but this fraction is smaller for pixel-based algorithm than for the spot-finding algorithm. Comparison of our decoded results with the total number of spots found by the spot-finding algorithm (correct + incorrect spots) indicate that our error-detection scheme rejects ~60-70% of the smFISH spots, which likely represent stray probes and fluorescent background.

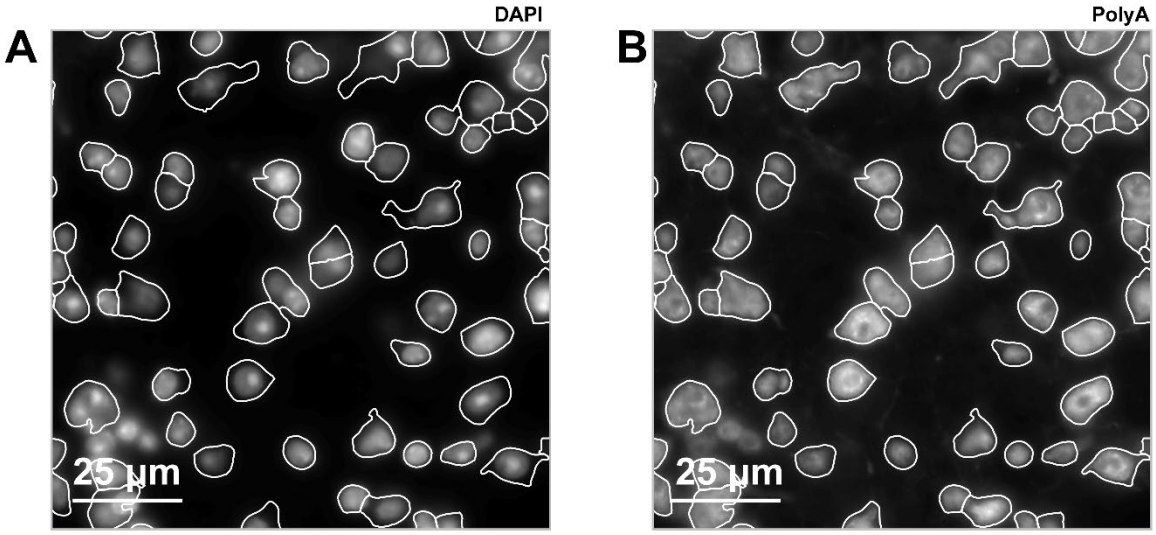


Fig. S10. Segmentation of cells with DAPI and polyA co-stains. (**A, B**) DAPI (A) and total polyA RNA (B) images of a single field-of-view (FOV) (same FOV as in **fig. S9**) shown together with the identified segmentation boundaries for each cell marked in white. Shown is the image from the central slice in the z-stack collected at this FOV, and the depicted boundaries were identified in this z-plane. To identify cell segmentation boundaries from the DAPI and polyA z-stacks, adaptive image thresholding was used to identify regions of the DAPI image that are bright and, thus, contain nuclei. Image erosion was then used to separate closely spaced nuclei and identify nuclei centers. To determine cell segmentation boundaries, we used a seeded watershed approach in which the intensity of the polyA image acts as watershed catch-basin, determining the boundaries of cells, while the location of nuclei act as 'seeds', i.e. locations in the image that must contain a cell. We utilize a three-dimensional watershed algorithm on our z-stacks, producing three-dimensional cell segmentation boundaries.

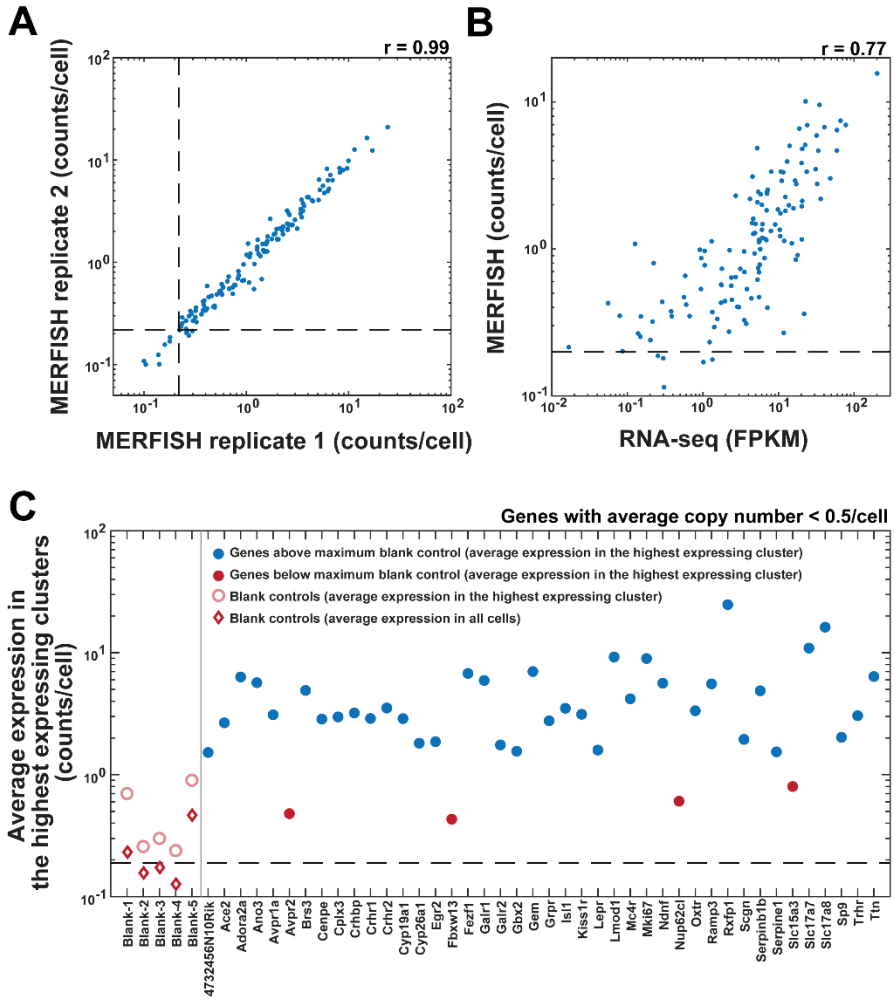


Fig. S11. Correlation between MERFISH and bulk RNA-seq, and reproducibility and false-positive detection rates in MERFISH. (**A**) Scatter plot of average copy number per cell for two MERFISH measurements conducted on two replicate animals for all genes measured in the combinatorial smFISH run. The dashed lines represent the average copy number per cell observed for the five "blank" control MERFISH barcodes (barcodes not assigned to any gene) and is a measure of the false-positive rate in MERFISH. We note that good correlation between replicates are observed even for low-expression genes whose average count per cell is below the average blank count. This is because not all cells express these genes and in individual cells that express these genes, their expression levels are typically higher than the average blank count (see panel (C)). (**B**) Scatter plot of average copy number per cell determined by MERFISH versus expression level determined by bulk RNA-seq for all genes measured in the combinatorial smFISH run. The dashed line represents the measured false-positive rate as described in (A). The Pearson correlation coefficients (r) between the logarithmic expression values depicted in (A, B) are provided. (**C**) The average expression of a gene across all cells (as in A, B) may not be representative of the average expression within individual clusters. For example, an RNA expressed at low levels when averaged across all cells may, nonetheless, be highly expressed in the cells of a low abundance cluster. To evaluate the fraction of genes that are lowly expressed when averaged across all cells but are still expressed above background within individual

clusters, we computed the average expression of each gene within every cluster and determined, for each gene, the expression level observed in its highest expressing cluster. This maximum expression level is plotted (blue and red solid circles) for the 42 low-expression genes with an average copy number across all cells of less than 0.5 copy per cell. Also depicted are the average copy number in the highest expressing clusters (red open circles) and the average copy number across all cells (red diamonds) for the five blank controls. Of these 42 low-expression genes, only four genes (red circles) have their average expression level within their respective highest expressing clusters below the maximum value observed for any of the blank controls. Thus, the vast majority of these low-expression genes are expressed above background in at least one cluster.

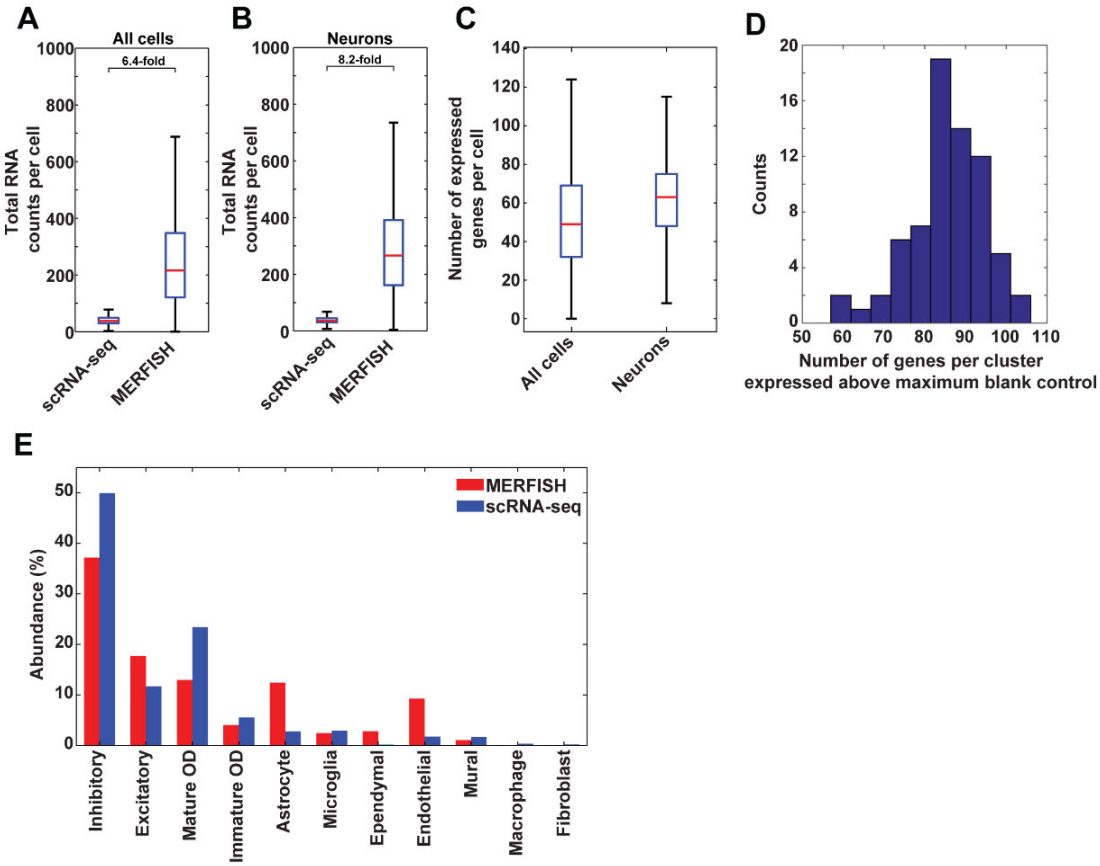


Fig. S12. Comparison of RNA copy numbers and cell abundance detected in MERFISH and scRNA-seq measurements. (**A**) Box plots of the total RNA copy number detected per cell for the 135 genes measured in the combinatorial smFISH run (right) in comparison with the total UMI number measured per cell with scRNA-seq for the same 135 genes (left). Depicted are values for all measured cells. The fold difference in the average counts observed with MERFISH relative to that observed with scRNA-seq is indicated. (**B**) As in (A) but for the total RNA copy number per cell for the 135 genes measured in all neurons. We note that many of the genes in the MERFISH panel that we chose for their biological significance happened to have low expression levels, such as neuromodulator receptors; thus, this panel of genes would be expected to produce relatively few counts per cell. (**C**) Box plots of the number of genes expressed per cell as determined by MERFISH. Depicted are values for all cells or all neurons measured with MERFISH. A gene is defined as expressed if a single copy is measured within a cell. In the boxplots, the red line is the median value, the box indicates the 25th to 75th quantile range, and the whiskers cover the 99% quantile range. (**D**) The blank-control barcodes measured in MERFISH allow a direct measure of the background false detection rate. To determine the number of genes expressed above this background level for each cluster, we computed the average expression of each gene within each cluster and compared those values to the average count of the highest-count blank-control barcode in that cluster. Depicted is the histogram of the number of genes in each cluster that are expressed above this background level, histogrammed across all clusters. (**E**) Fractional abundance of cells observed in MERFISH (red) and scRNA-seq (blue) for the major cell classes.

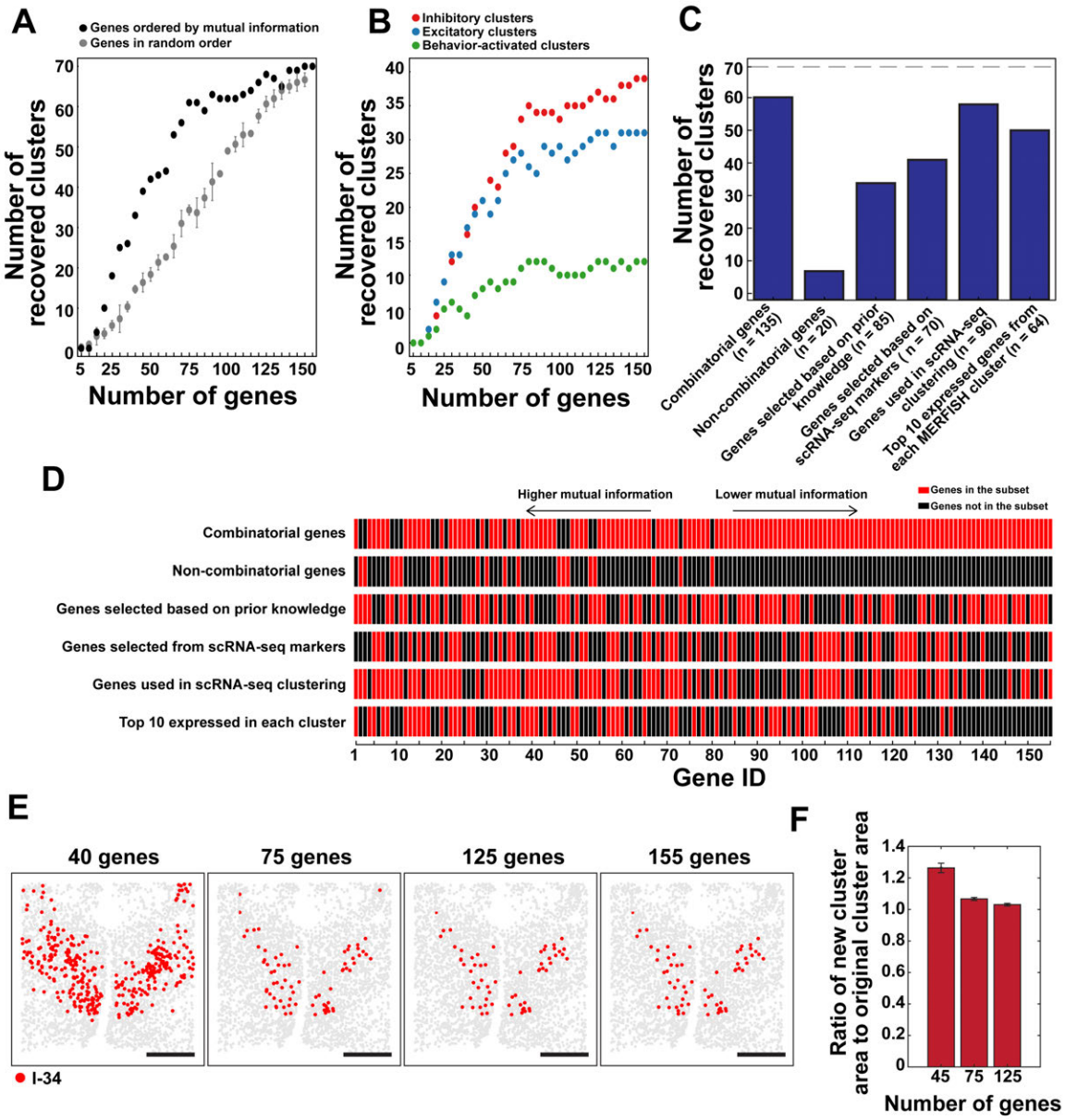


Fig. S13. The dependence of the neuronal clusters recovered on the number of genes used for clustering. To determine the relative importance of the number of genes as well as different gene categories in clustering, we examined the effect of different gene subsets on re-clustering the neurons identified with MERFISH. In parallel, we also used an information theory quantity known as mutual information to determine the relative importance of each gene to the definition of our clusters, as described previously (*67*). In addition, because different genes can share similar expression patterns across clusters and, thus, carry partially redundant information, we calculated the mutual information for each gene conditional on the expression of all genes with higher levels of mutual information; thus, partially removing this redundancy and generating a sorted list of most to least informative genes. (**A**) The number of the originally identified neuronal clusters (using all 155 genes) that were recovered when the imaged neurons were re-

clustered using different numbers of genes. A cluster is judged to be recovered if more than half of the cells in the original cluster are grouped together in a cluster in the new clustering scheme, a commonly used criterion (*31*). Genes were added in order of most to least mutual information (black markers) or were added in random order (gray markers; average over three random gene orders with error bars representing standard deviation (n = 3)). (**B**) The values in (A) displayed separately for the inhibitory clusters (red), excitatory clusters (blue), and the clusters that were activated during behavioral stimulus (green; clusters indicated by red bars with * symbol in **Fig. 8**). Re-clustering results in (A) and (B) show that increasing the number of genes rapidly increased the number of clusters recovered up to ~75 genes, which allowed ~90% of the neuronal clusters to be recovered. Beyond 75 genes, the number of clusters recovered increased more slowly with the number of genes included. (**C**) The number of clusters recovered when the imaged neurons were re-clustered with the following 6 specific gene subsets: 1) the 135 genes measured in the combinatorial smFISH rounds, 2) the 20 genes measured with non-combinatorial FISH, 3) the 85 pre-selected genes that are known markers of major cell class or are known to be of functional importance to hypothalamus, 4) the 70 additional genes added, which are marker genes for specific neuronal clusters identified by scRNA-seq but not already included in the 85 preselected genes, 5) the 96 genes in the MERFISH panel that were among the genes used for clustering the scRNA-seq neurons, and 6) the 64 genes that represent the top 10 most expressed genes in any MERFISH neuronal cluster. The dashed line marks the total number of clusters identified when all 155 genes were used for clustering. (**D**) The mutual information for genes contained within the subsets depicted in (C). Individual bars represent genes, with genes sorted from highest to lowest mutual information from left to right. For each gene subset, red represents a gene that is within that subset while black indicates a gene that is not within the subset but is among the 155 measured genes. Results in (C) show that both the 85 genes selected based on prior biological knowledge and the 70 additional marker genes identified by scRNA-seq were important for identifying neuronal clusters, as only about half of the clusters were recovered if either subset was used alone for clustering. Nearly 90% of the clusters were recovered with the 135 genes measured in the combinatorial smFISH; however, the 20 genes measured in the sequential FISH rounds were also highly informative, as shown in (D). In addition, because these genes contained many functionally relevant genes such as neuropeptides, they were important for inferring the functional properties of the neuronal clusters. (**E, F**) In addition to recovering fewer clusters, the accuracy of the spatial distributions of the identified clusters is also reduced when clustering with fewer genes. (**E**) An example cluster (I-34) that was originally detected with the 155 genes and recovered by clustering with 40, 75, and 125 genes. As the number of genes is decreased, fewer clusters were recovered and each cluster tends to include more cells. (**F**) To quantify this effect, we compared the spatial areas covered by the recovered clusters when re-clustering with different numbers of genes to the areas of the corresponding original clusters. Plotted are the ratios of the area of the recovered cluster to that of the original cluster averaged over all clusters (error bars represent S.E.M. over the number of recovered clusters, n = 39, 61, 68, respectively). When 40 genes were used to re-cluster the data, the recovered clusters displayed a ~30% increase in the area relative to the original cluster. By contrast, once the majority of clusters were recovered with 75 or 125 genes, this error dropped substantially, to <10% and <5%, respectively. Together these results illustrate both the importance of imaging a large gene panel and the adequacy of imaging 155 genes for the cell populations identified here.

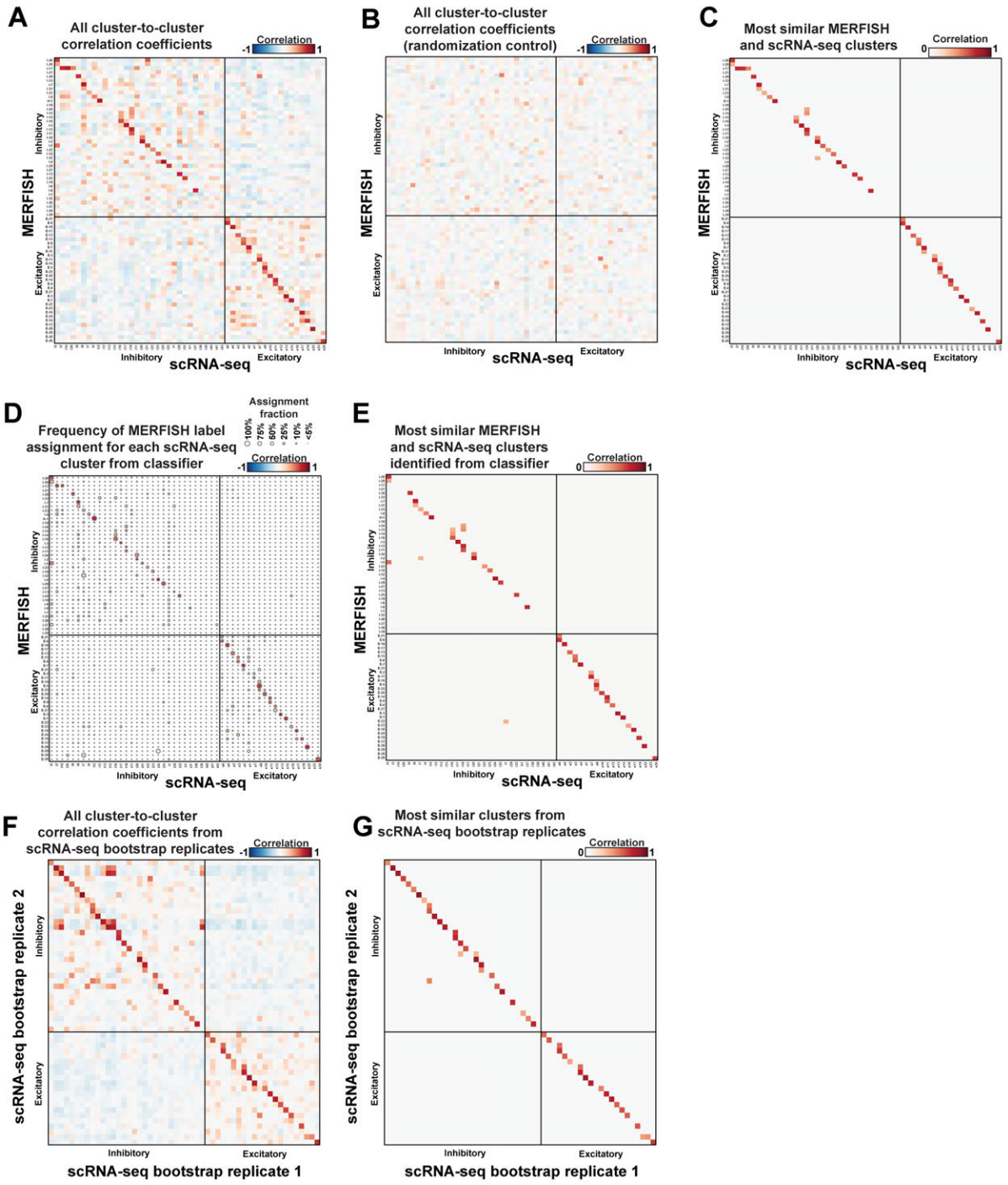


Fig. S14. Correlation between MERFISH neuronal clusters and scRNA-seq neuronal clusters. (**A**) Pairwise Pearson correlation coefficients between the average expression profiles for individual neuronal cluster identified with MERFISH and those identified with scRNA-seq. The average expression profiles are expressed in z-scores calculated for individual genes across all clusters. (**B**) A randomization control showing the Pearson correlation coefficient as in (A)

but calculated when the order of the genes in the expression profiles is randomized. (**C**) The scRNA-seq cluster(s) that correspond to or are most similar to a given MERFISH cluster, identified as the scRNA-seq cluster with the highest Pearson correlation coefficient or the set of scRNA-seq clusters with highest Pearson correlation coefficients that were approximately indistinguishable from one another (*29*). The Pearson correlation coefficients are indicated by color and given in **Table S9**. Panels (A) and (C) are reproduced from **Fig. 4C, D** for reference. (**D**) To further validate our assignment of the most similar scRNA-seq cluster(s) for each MERFISH cluster, we trained a neural network on the expression profiles (measured in z-scores) and cluster labels for the MERFISH data and used this network to assign a MERFISH cluster label to each neuron measured with scRNA-seq based on its expression profile (measured in z-scores (*29*)). We then identified all scRNA-seq cells that are assigned the label of each MERFISH cluster and plotted the fraction of these labeled cells that are within each of the scRNA-seq neuronal clusters. The size of the circle represents this fraction and the color represents the Pearson correlation coefficient between the average expression profiles of the scRNA-seq and MERFISH clusters as described in (A). (**E**) scRNA-seq cluster(s) that correspond to or are most similar to a given MERFISH cluster, identified as the scRNA-seq cluster or the set of scRNA-seq clusters that contained the largest and approximately indistinguishable fraction of the MERFISH label assigned via the neural network classifier approach described in (D). As a support for our association between scRNA-seq and MERFISH clusters by the correlation analysis shown in (A) and (C), we find that this classifier analysis identifies the vast majority of the corresponding clusters that were identified by the correlation analysis, and vice versa. (**F, G**) For comparison purpose, we partitioned our scRNA-seq neurons into two equally sized groups, i.e. bootstrapped replicates, clustered each of these groups independently, and then compared the expression profiles of these clusters using the approaches in (A) and (C). Depicted in (F) are the pairwise Pearson correlation coefficients between all clusters formed from these two bootstrapped replicates, as in (A), and depicted in (G) are these values for the most similar pairs of clusters, as in (C). We find that 81% of the clusters identified in each of bootstrapped replicates had a corresponding cluster(s) in the other replicate, and the correlation coefficients for these most similar clusters ranged from ~0.4-0.8, with an average of 0.6. In comparison, 73% of MERFISH clusters had a corresponding scRNA-seq cluster(s), 75% of scRNA-seq clusters had a corresponding MERFISH cluster(s), and the correlation coefficients also ranged from ~0.4-0.8, with an average of 0.6.

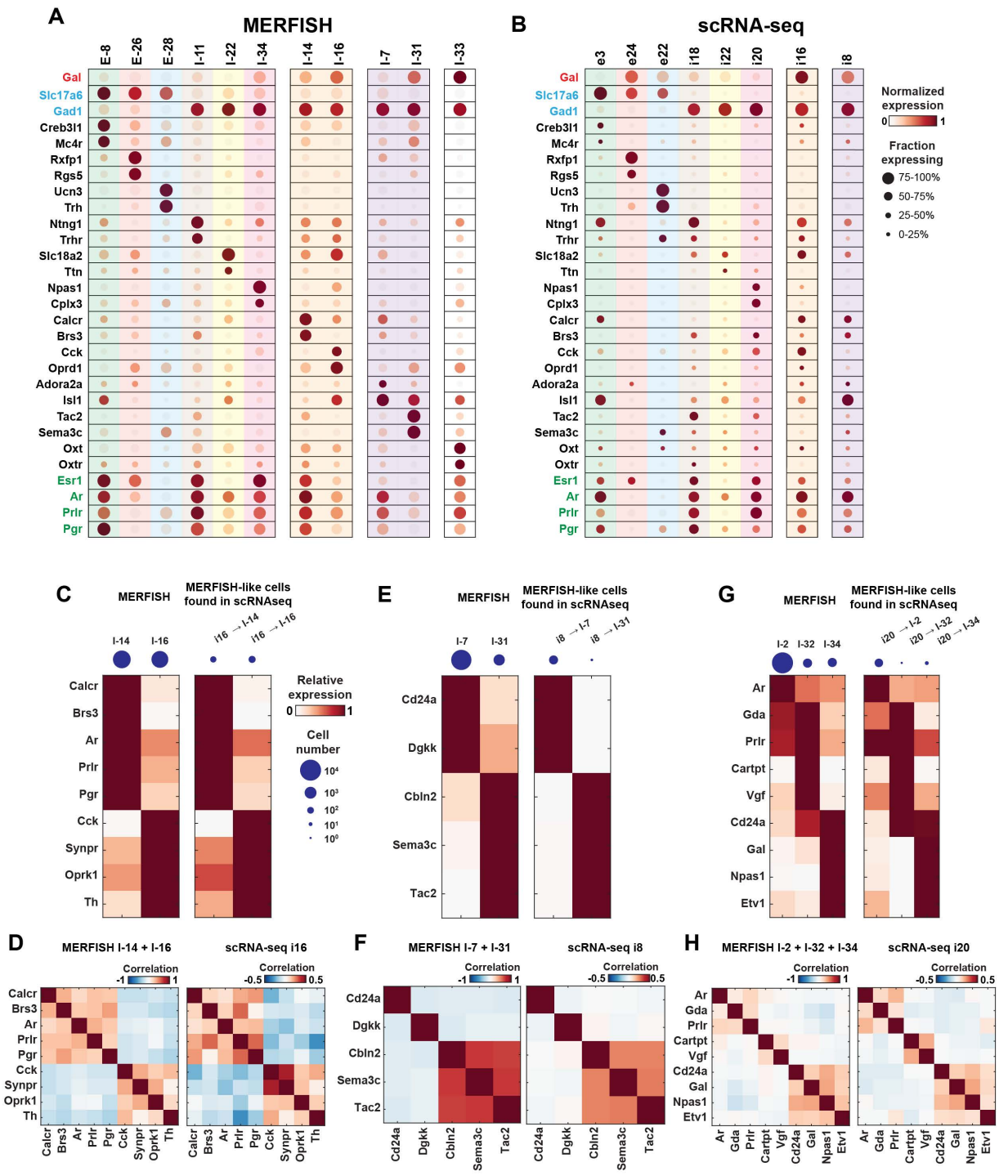


Fig. S15. Correspondence between Gal-enriched neuronal clusters identified in MERFISH and scRNA-seq. (**A, B**) Expression of selected marker genes and genes of interest in Gal-enriched clusters identified by MERFISH (A) or scRNA-seq (B). The size of circles indicates the fraction of cells in each cluster expressing a given gene, and the shade of circles indicates the average expression level of a given gene within each cluster normalized to the maximum expression level of that gene in the depicted clusters. The genes depicted here are Gal (red), the inhibitory and excitatory markers Gad1 and Slc17a6 (blue), the top two markers of each Gal-

enriched neuronal cluster identified by MERFISH (black), and the sex hormone receptors, Esr1, Prlr, Ar, Pgr (green). The putative corresponding MERFISH and scRNA-seq clusters are highlighted by vertical bars of the same color, and situations in which two adjacent MERFISH neuronal clusters share the same color of vertical bar represent situations in which two MERFISH neuronal clusters corresponded to a single scRNA-seq neuronal cluster. (**C**) Identification of sub-populations of cells within the Gal-enriched scRNA-seq cluster i16 that express markers for the two Gal-enriched MERFISH clusters, I-14 and I-16. The average expression profile for selected genes in these two MERFISH clusters (left) agrees well with the average expression profile (right) observed for sub-populations of scRNA-seq cells in i16 that were classified as most similar to either I-14 or I-16 using a neural network classifier. The listed genes are selected from the most differentially expressed genes between I-14 and I-16. Circles (top) depict the number of cells in the original MERFISH clusters and the number of cells in i16 classified as each of these MERFISH clusters. (**D**) (Left) Pearson correlation coefficients between the expression levels of pairs of genes across all cells in MERFISH clusters I-14 and I-16 combined, for genes depicted in (C). (Right) Pearson correlation coefficients between expression levels of pairs of genes across all cells in scRNA-seq cluster i16. Gene expression within individual cells was expressed in z-scores. (**E,F**) As in (C,D) but for the two Gal-enriched MERFISH clusters, I-7 and I-31, that showed putative correspondence to the Gal-enriched scRNA-seq cluster, i8. (**G,H**) As in (C,D) but for the Gal-enriched cluster I-34 and the two non-Gal-enriched clusters I-2 and I-32, all of which showed putative correspondence to the Gal-enriched cluster i20.

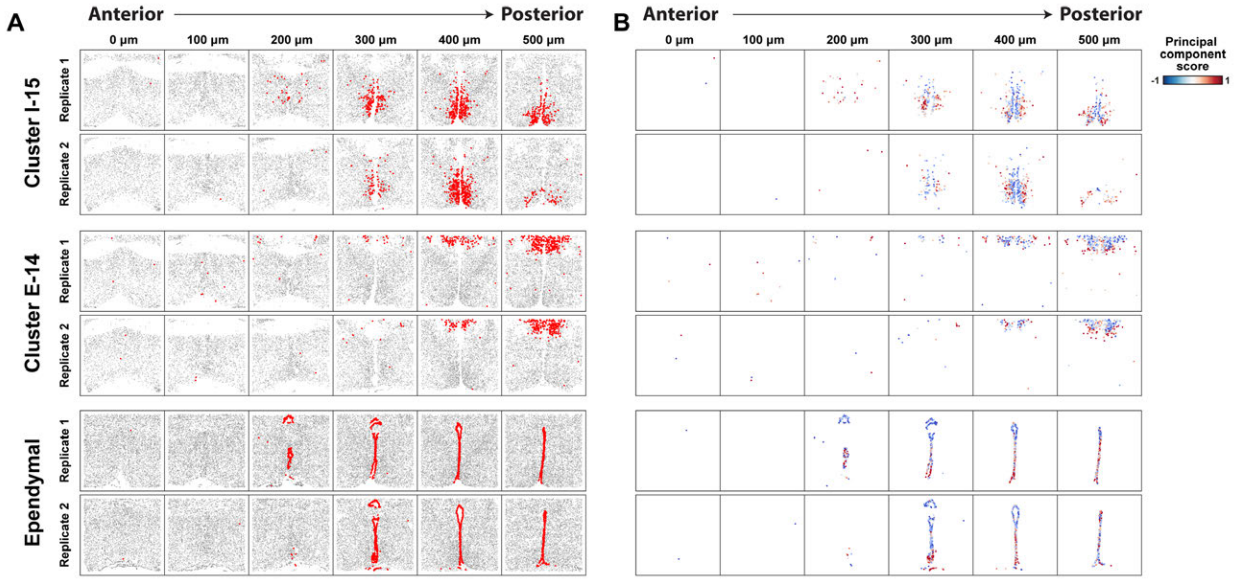


Fig. S16. Heterogeneity in gene expression as a function of position within a few selected clusters. (**A**) Spatial distribution of cells in neuronal cluster I-15 (top), E-14 (middle), and ependymal cell class (bottom) for six of the twelve 1.8-mm × 1.8-mm imaged slices taken from two replicate female mice. For clusters I-15 and E-14, red markers indicate cells of the specified neuronal clusters and gray markers indicate all other neurons. For the ependymal cell example, red markers indicate all ependymal cells and gray markers indicate all other cells. (**B**) Spatial variation of gene expression within the same clusters shown in (A). Only cells within the specified cluster are depicted, and the color of each cell is determined by its amplitude along the first principal component as determined by principal component analysis (PCA) of the variation in gene expression among cells of each specified cluster. The loadings associated with this principle component are dominated by a handful of genes, suggesting that gradients in the expression of a small number of genes drive these variations. Notable genes that contribute to the loadings of the first principal component are the following: Irs4, Pak3, and Slc18a2 for I-15; Necab1, Cbln1, and Cbln2 for E-14; and Cd24a, Aqp4, and Mlc1 for ependymal cells.

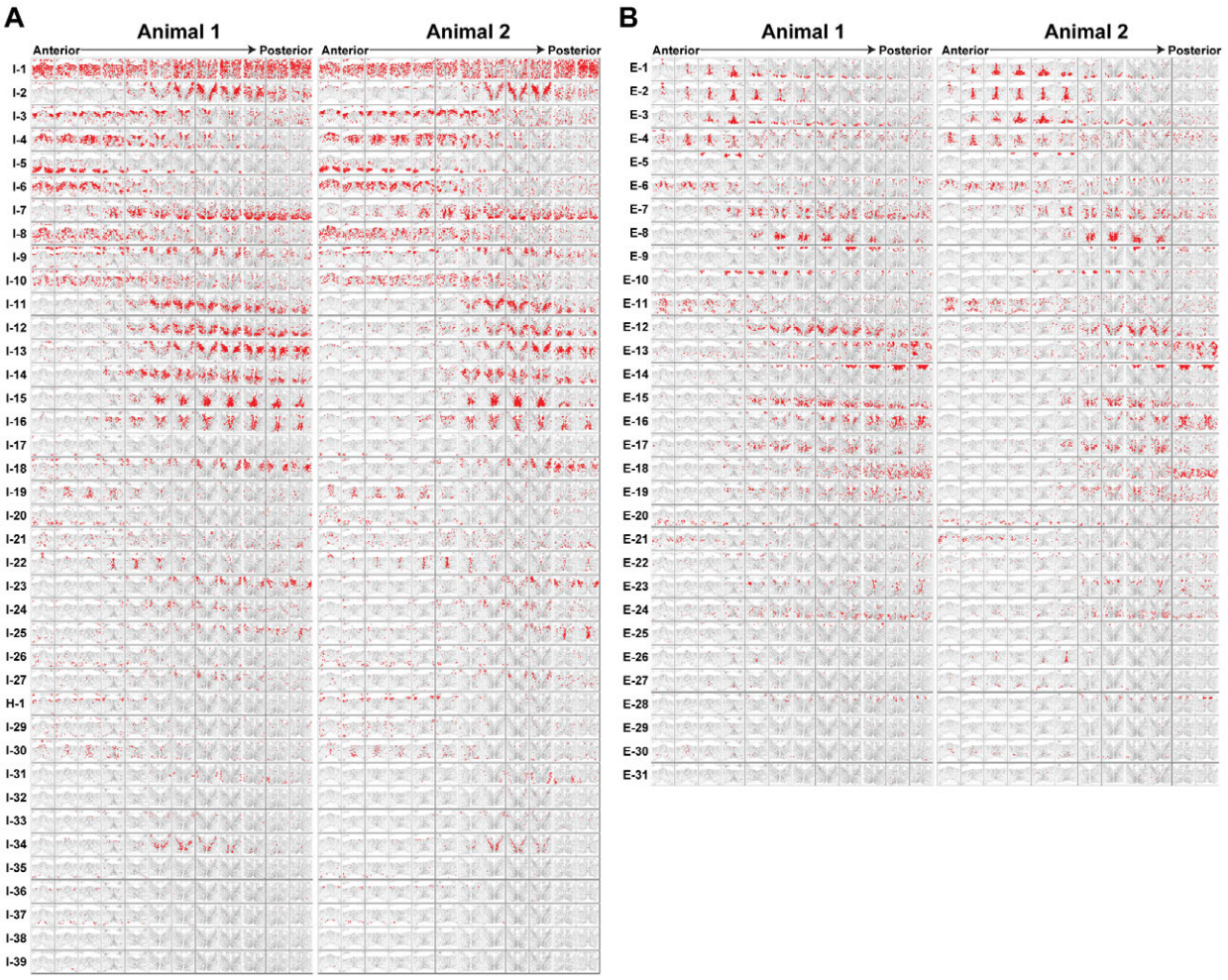
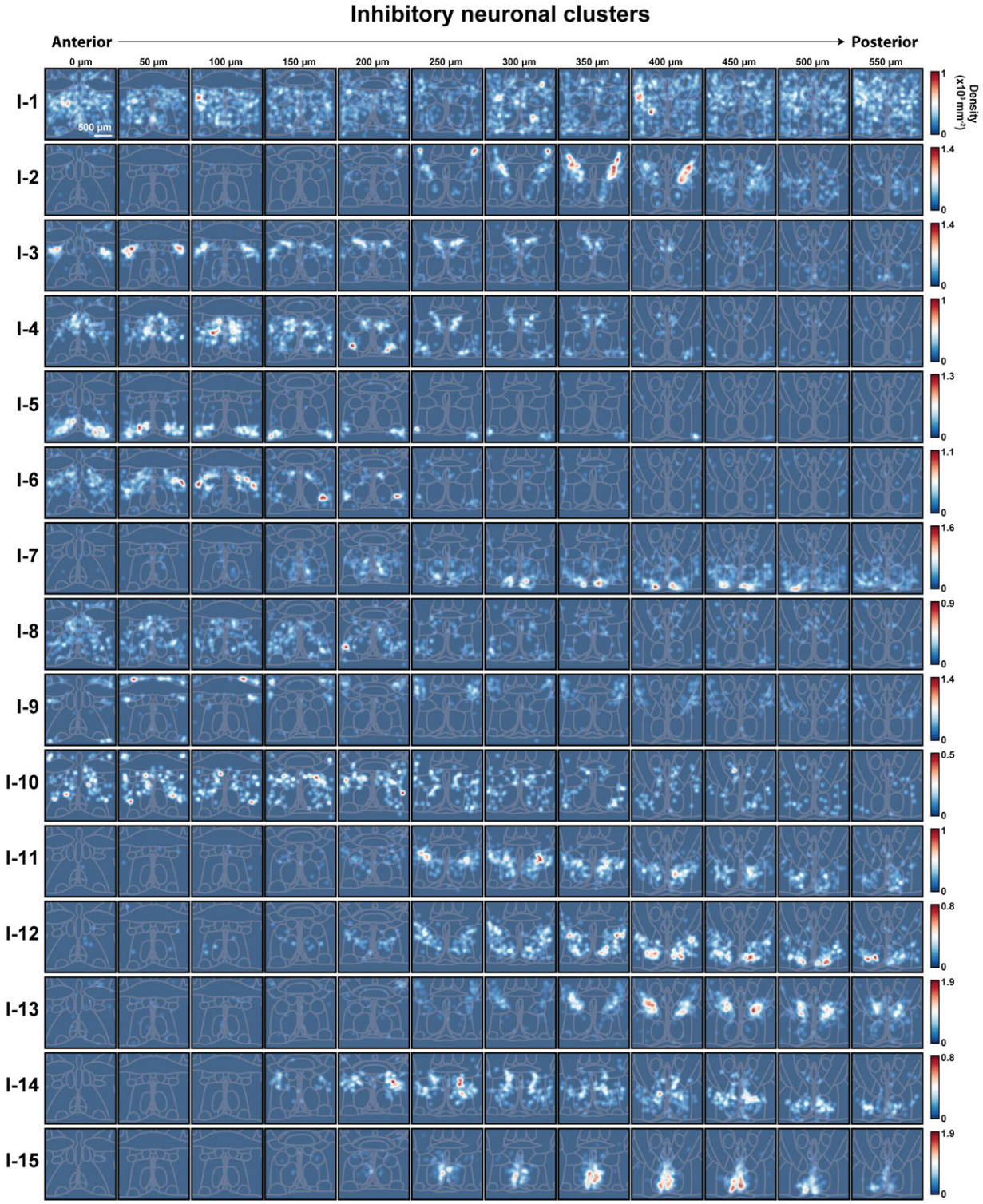
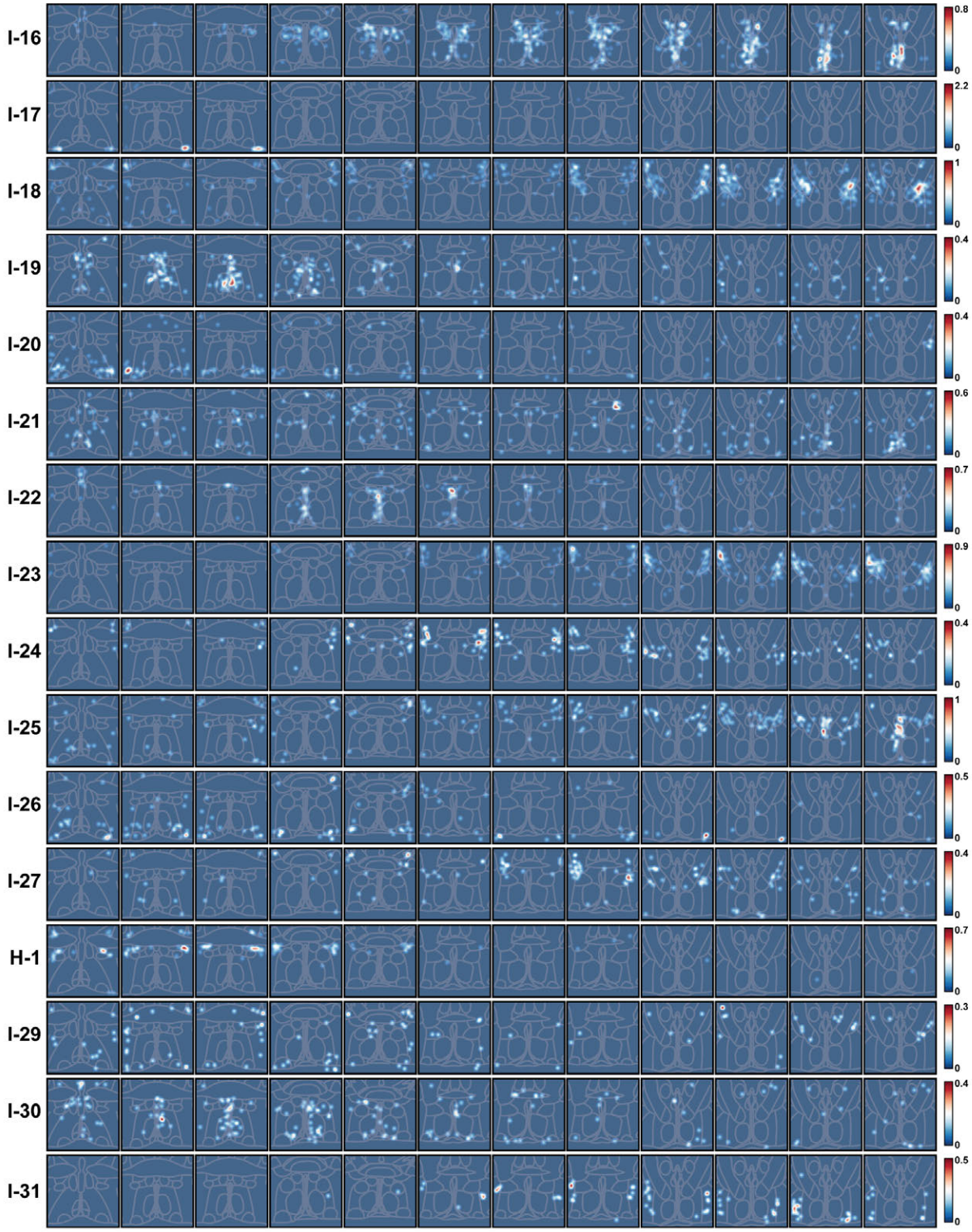
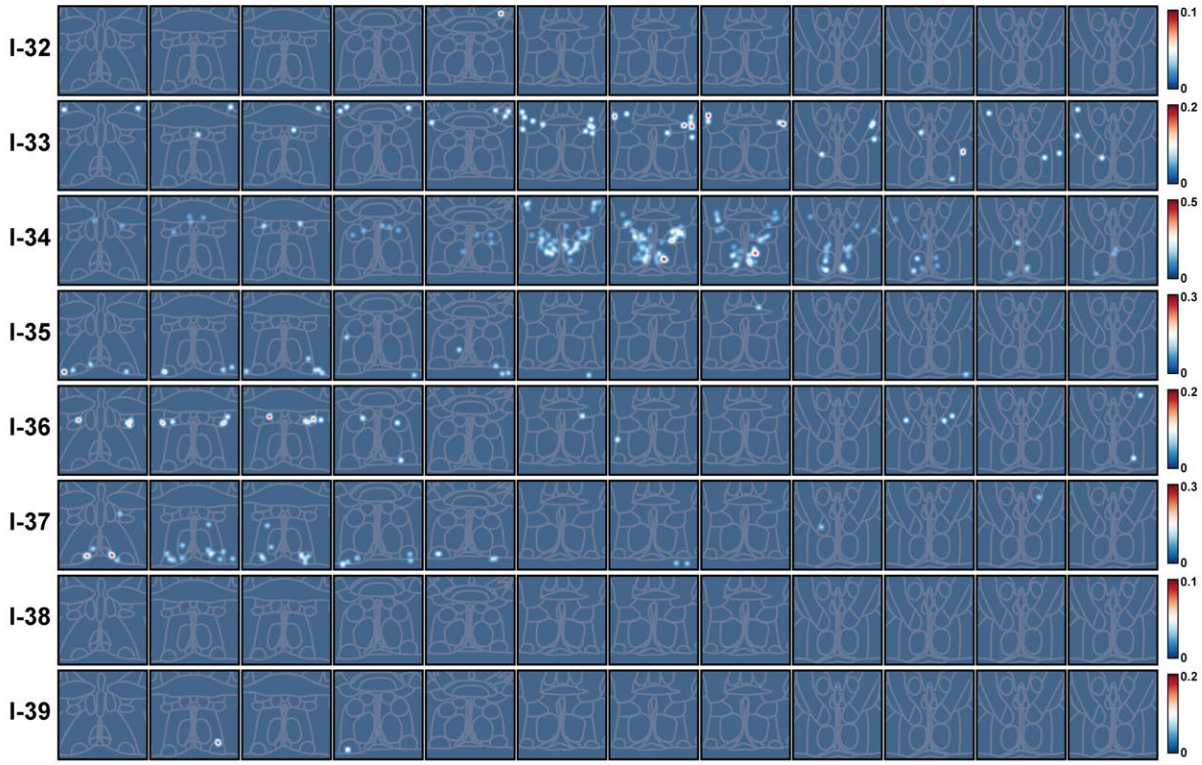
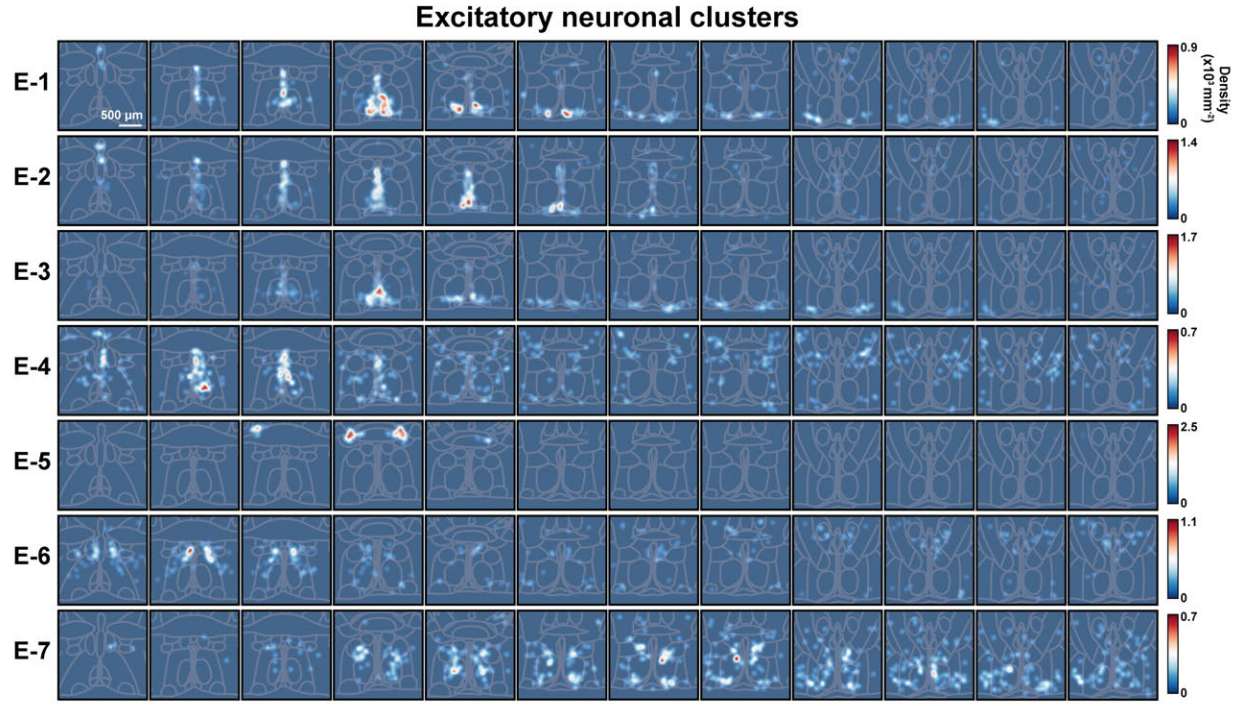


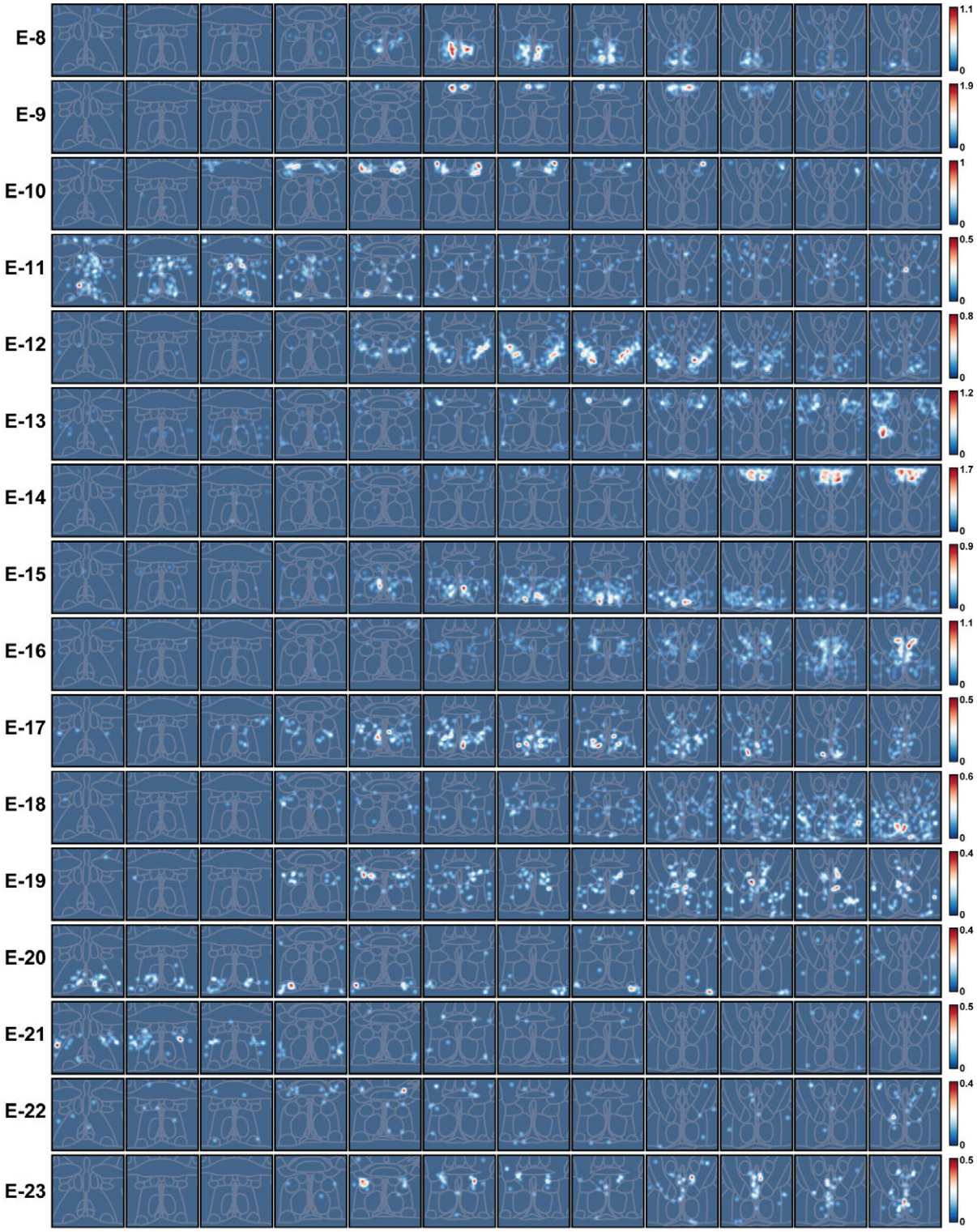
Fig. S17. Spatial distribution of all neuronal clusters in two replicate animals. (**A, B**) Spatial distribution of all inhibitory (A) and excitatory (B) neuronal clusters for the twelve 1.8-mm × 1.8-mm slices from the different anterior-posterior positions of the preoptic region for two replicate female animals. Red markers indicate cells of the specified neuronal clusters and gray markers indicate all other neurons. Slices progress from the anterior portion to the posterior portion of the preoptic region from left to right. The hybrid cluster, H-1 (originally I-28), is listed with the inhibitory neurons, as it was identified in the second round of clustering with inhibitory neurons. Results from 2 of the 9 measured animals are shown here. For some of the 9 animals, only a subset of the twelve slices were measured. We note that, despite our effort to measure ~1 million cells and keep only stable clusters, a few of the identified clusters had extremely low abundance (e.g. I-38, I-39, E-31) and may not represent specific cell populations. In addition, we note that cluster I-1 is spread relatively uniformly throughout the imaged region and is devoid of marker genes other than the expression of Gad1, and thus may not represent a specific neuronal population.


Inhibitory neuronal clusters
Anterior
Posterior
0 μm
50 μm
100 μm
150 μm
200 μm
250 μm
300 μm
350 μm
400 μm
450 μm
500 μm
550 μm
500 μm
I-1
I-2
I-3
I-4
I-5
I-6
I-7
I-8
I-9
I-10
I-11
I-12
I-13
I-14
I-15
Density (x10³ mm⁻²)
1
0
1.4
0
1.4
0
1
0
1.3
0
1.1
0
1.6
0
0.9
0
1.4
0
0.5
0
1
0
0.8
0
1.9
0
0.8
0
1.9
0


I-16
0.8
0
I-17
2.2
0
I-18
1
0
I-19
0.4
0
I-20
0.4
0
I-21
0.6
0
I-22
0.7
0
I-23
0.9
0
I-24
0.4
0
I-25
1
0
I-26
0.5
0
I-27
0.4
0
H-1
0.7
0
I-29
0.3
0
I-30
0.4
0
I-31
0.5
0


I-32
I-33
I-34
I-35
I-36
I-37
I-38
I-39
0.1
0.2
0.5
0.3
0.2
0.3
0.1
0.2


Excitatory neuronal clusters
E-1
E-2
E-3
E-4
E-5
E-6
E-7
500 μm
0.9
1.4
1.7
0.7
2.5
1.1
0.7
Density
(x10³ mm⁻²)


E-8
1.1
0
E-9
1.9
0
E-10
1
0
E-11
0.5
0
E-12
0.8
0
E-13
1.2
0
E-14
1.7
0
E-15
0.9
0
E-16
1.1
0
E-17
0.5
0
E-18
0.6
0
E-19
0.4
0
E-20
0.4
0
E-21
0.5
0
E-22
0.4
0
E-23
0.5
0

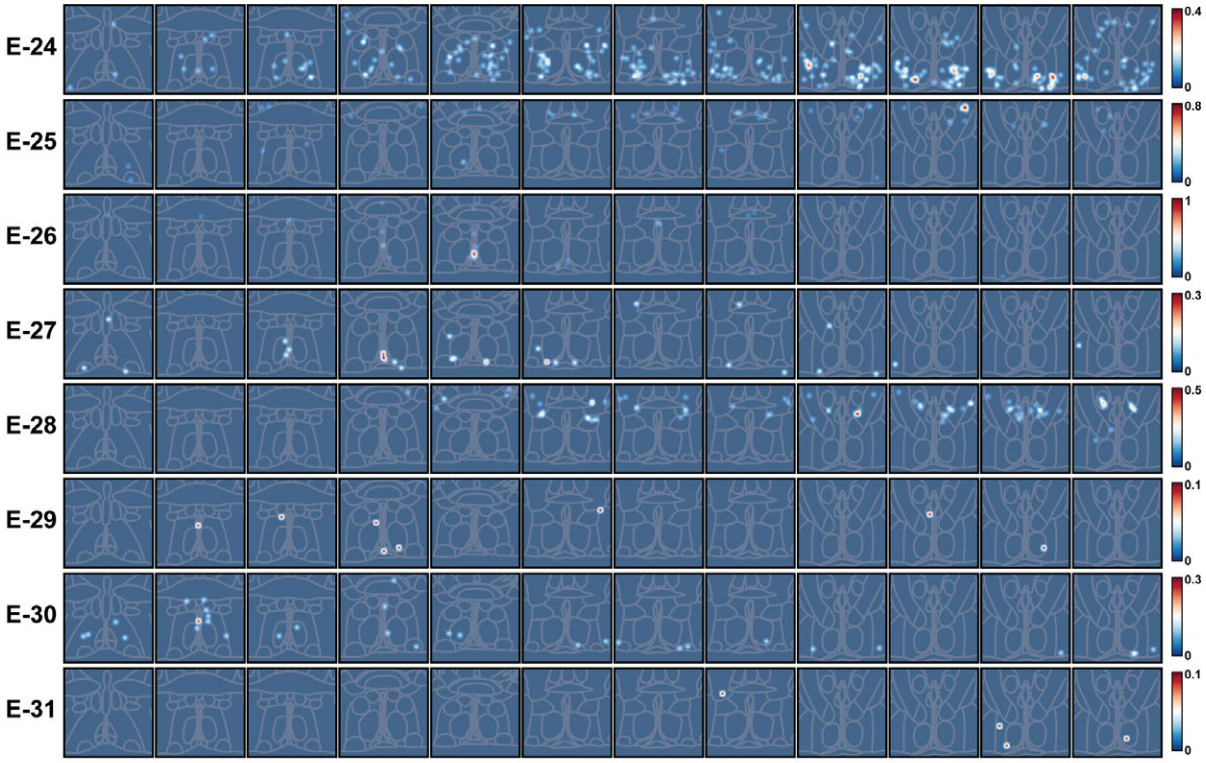


Fig. S18. Density plots of all neuronal clusters. The density of each MERFISH neuronal cluster in each of the twelve 1.8-mm × 1.8-mm slices from the different anterior-posterior positions of the preoptic region for one female replicate (Animal 1 depicted in **fig. 17**). Slices are arrayed from the anterior on the left to the posterior on the right. Regions of low neuronal density are depicted in blue, and regions of high density are depicted in red. The density is normalized for each neuronal cluster to the maximum value observed across all slices, and the density scale is listed for each neuronal cluster. The density is estimated by convolving a gaussian distribution of 25-μm width with the location of each cell within a given neuronal cluster. The schematic lines indicating the boundaries of the nuclei for each slice are depicted in white. The nucleus boundaries are drawn based on Paxinos (*45*) and approximately aligned to the tissue section imaged based on the locations of landmarks, such as the anterior commissure, ventricle, and fornix. The nucleus identities are depicted in **Fig. 5B**. The resolution of the nuclei map along the anterior-posterior axis is not as fine as the distance between our slices; thus, individual nuclei maps have been repeated for different slices.

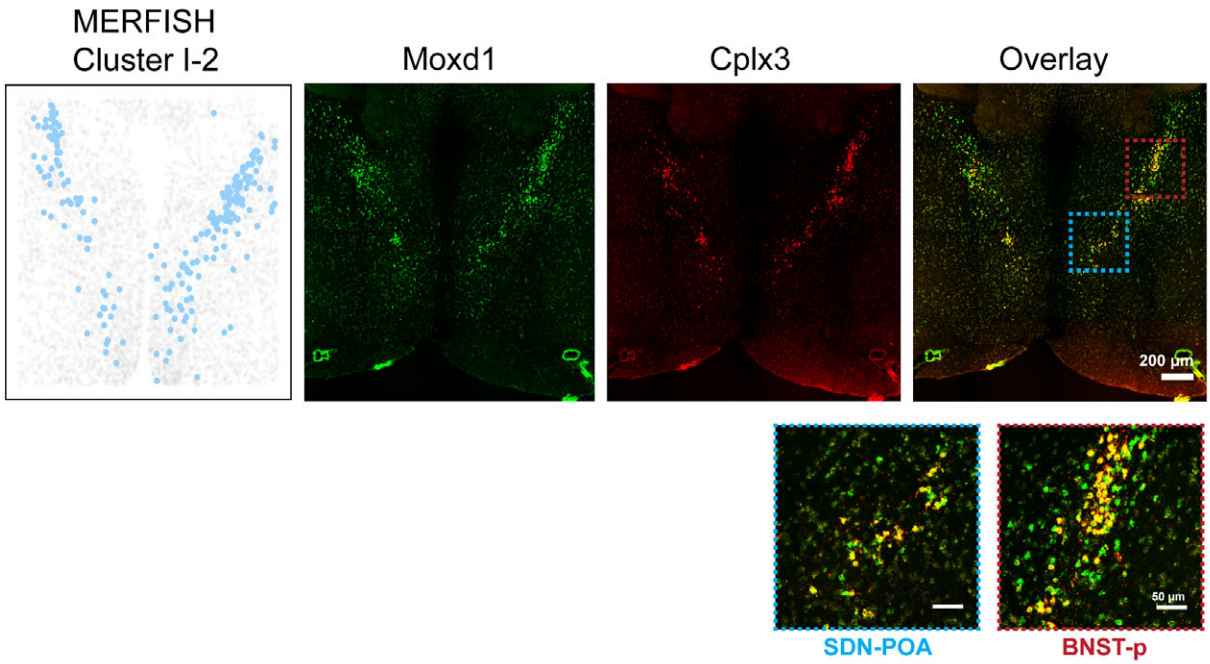


Fig. S19. Cluster I-2 partially overlaps in space with the sexually dimorphic nucleus of the pre-optic area (SDN-POA). Left panel: Spatial distribution of the neuronal cluster I-2 identified by MERFISH in a single slice, with neurons in I-2 depicted in blue and all other neurons in grey. Right panels: In situ hybridization with Moxd1 (red) and Cplx3 (green), and the overlay of these stains (yellow marks co-expression of these markers). Bottom: Zoom-in on the BNST-p region (red boxed region) and the SDN-POA region (blue boxed region). Moxd1 is a known marker of the SDN-POA. Cplx3 marks the MERFISH cluster I-2 and Moxd1, though not part of the MERFISH gene panel, marks the scRNA-seq cluster i-20 that corresponds to the MERFISH cluster I-2 (**Table S9**). Moxd1 and Cplx3 are co-expressed in cells in the SDN-POA, as well as in cells in the BNST-p region, and both the SDN-POA and BNST-p spatially overlap with I-2. The position of the depicted slice along the anterior-posterior axis is Bregma -0.15.

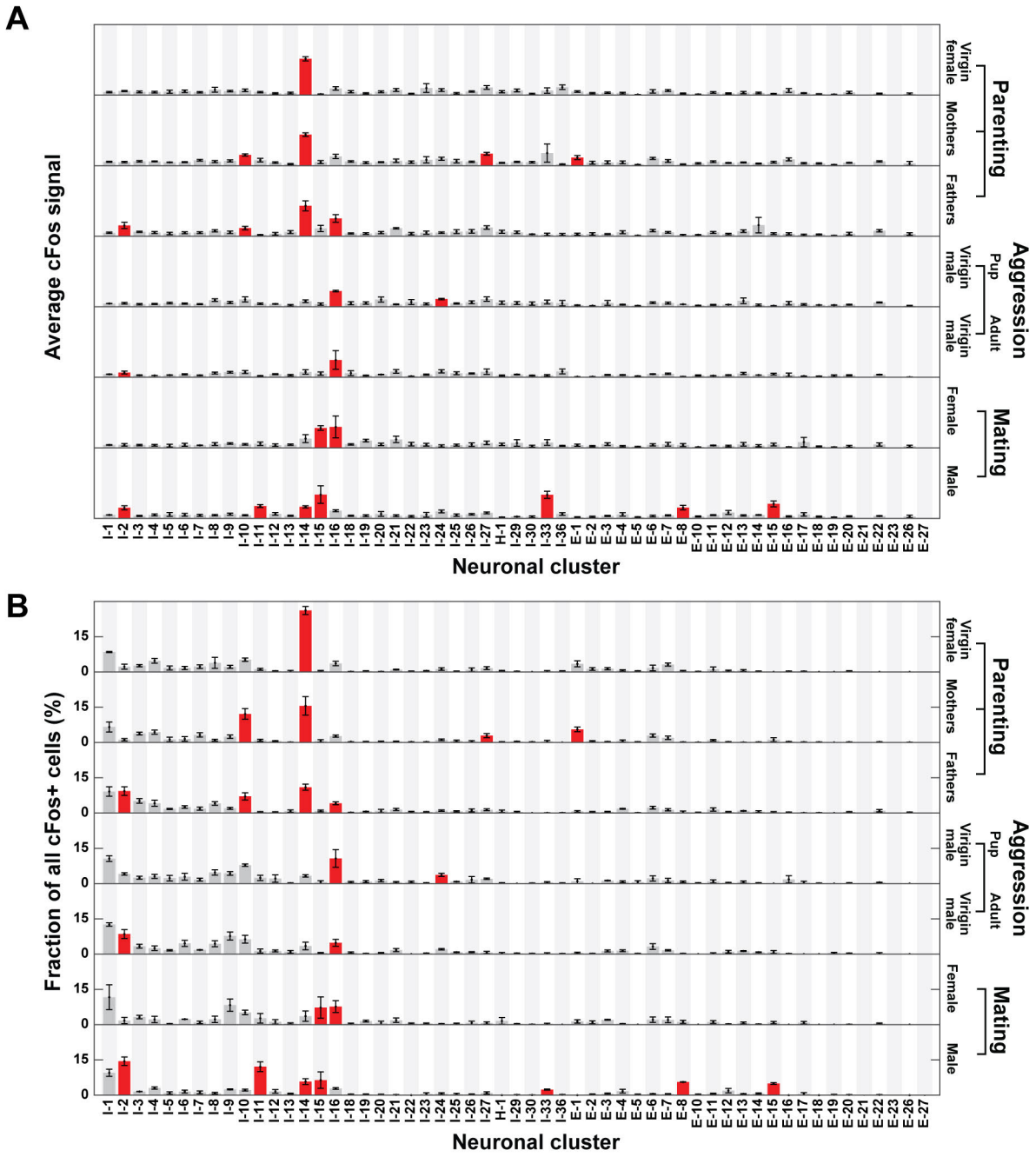


Fig. S20. cFos expression in neuronal clusters after specific social behaviors. (**A**) Average cFos expression level observed within each neuronal cluster in males and females after parenting, mating, and aggressive behavior. The cFos level is in arbitrary fluorescence intensity units and the scale is the same for all panels. (**B**) Fraction of all cFos-positive cells distributed among individual clusters. Error bars represent the standard error of the mean across replicate animals (N = 3 – 5 replicate animals). Red bars represent clusters that displayed a statistically significant enrichment in cFos-positive cells (binomial test, false discovery rate < 5%) as determined in **Fig. 8A**. Only clusters in which at least 10 cells are present in two or more replicates of each sex and behavioral stimulus are depicted.

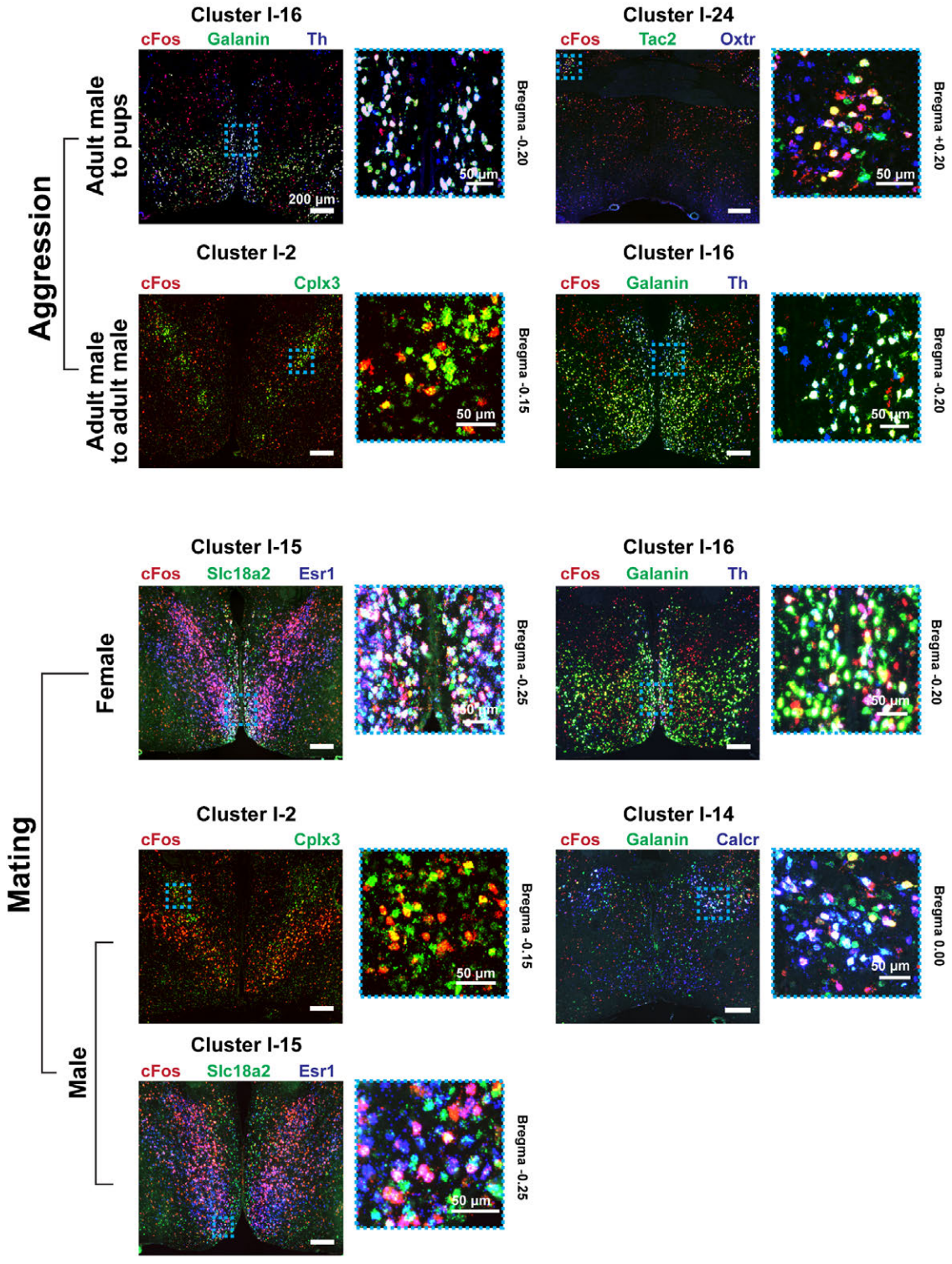


Fig. S21. In situ hybridization images showing co-expression of cFos and markers of neuronal clusters activated by various behaviors. Representative two- or three-color in situ

hybridization images of 16-μm-thick sections from the preoptic region show co-localization of markers of the indicated neuronal clusters with cFos after various behaviors, including pup-directed aggression in virgin males, adult-directed aggression in virgin males, mating in females, and mating in males. Regions in blue dashed boxes are magnified and shown on the right. Red, green, and blue mark the listed genes and white indicates co-expression in the overlay images. For I-2, only one marker gene Cplx3 (green) was imaged together with cFos (red) and yellow indicated co-expression in the overlay image. Among clusters that displayed an enrichment in cFos-positive cells after these behaviors (highlighted in red in **Fig. 8A**), only those clusters that can be distinguished by the combination of one to two marker genes and their spatial location are tested here. Images of sections are shown at different Bregma positions for different clusters to match the anterior-posterior positions of the clusters determined by MERFISH. The number of cFos-positive, Esr- and Slc18a2-expressing cells are relatively small in male-mating samples compare to female-mating samples, consistent with our observations that I-15 cells are less abundant in males than in females (**Fig. 6H**), that enrichment of cFos-positive cells in I-15 is weaker in males than in females (**Fig. 8A**), and that Slc18a2 shows a gradient of expression in I-15 (**fig. S16**) and is expressed in only a fraction of I-15 cells (smaller fraction in males than in females). Scale bars: 200 μm. Scale bar for zoom-in: 50 μm.

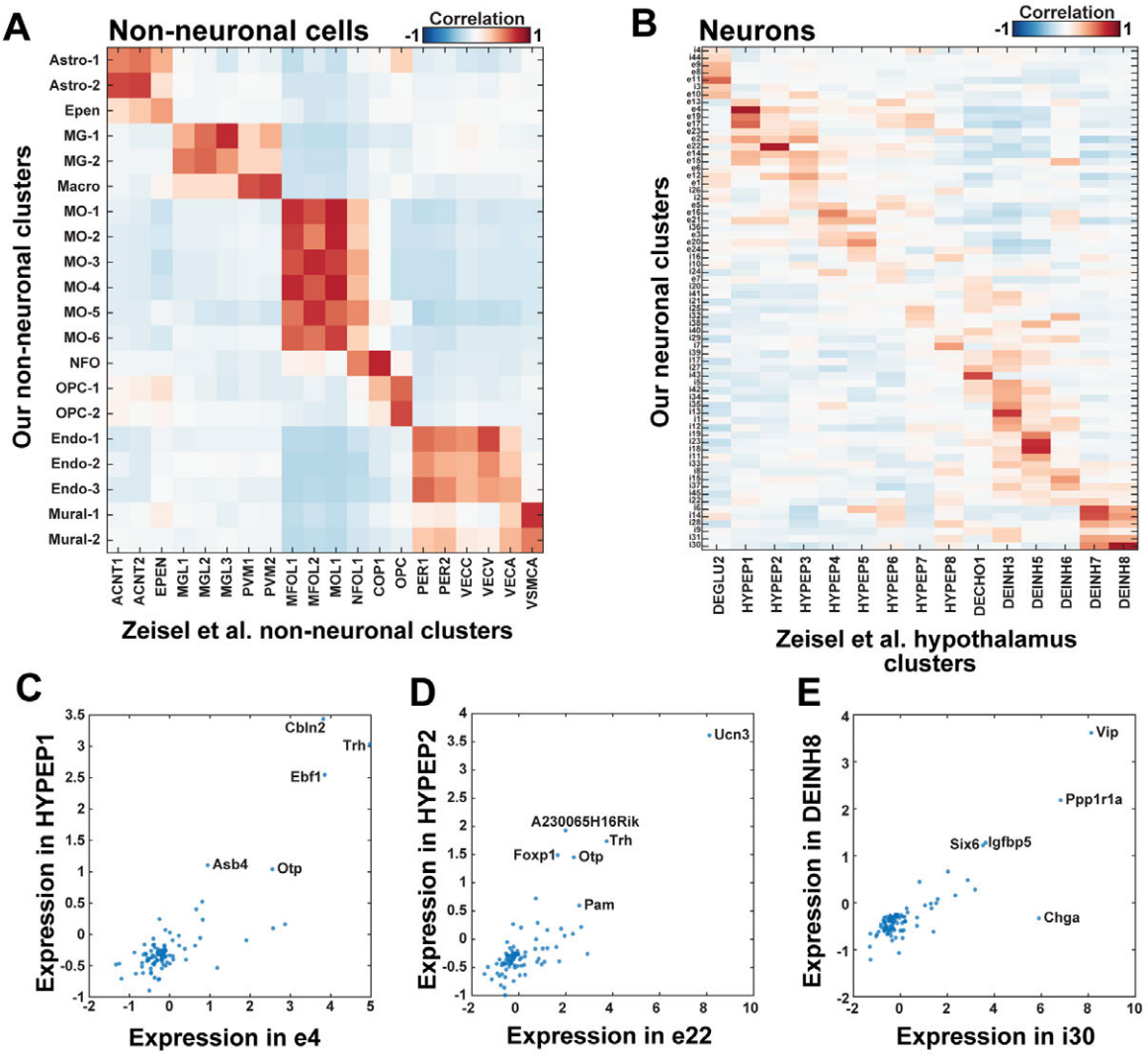


Fig. S22. Comparison of the clusters identified here with published clusters from the mouse hypothalamus [Zeisel et al. (*69*)**].** We compared the cell clusters that we identified for the preoptic region using scRNA-seq with published clusters found in the hypothalamus in a recent large-scale scRNA-seq study of the mouse brain (*69*). (**A**) Pairwise Pearson correlation coefficients between the average expression profiles for the non-neuronal clusters identified in our work and those for the comparable non-neuronal clusters from Zeisel et al (*69*) that were found in the entire central nervous system, including the hypothalamus. The average expression profiles are expressed in z-scores calculated for individual genes across all clusters. For the clusters identified in this work: Astro: astrocytes; Epen: ependymal; MG: microglia; Macro: macrophage; MO: mature oligodendrocytes; NFO: newly formed oligodendrocytes; OPC: oligodendrocyte precursor cells; Endo: endothelial cells. For the clusters published by Zeisel et al: ACTN: astrocytes; EPEN: ependymal; MGL: microglia; PVM: perivascular macrophages; MFOL: myelin forming oligodendrocytes; MOL: mature oligodendrocytes; NFOL: newly formed oligodendrocytes; COP: committed oligodendrocytes; OPC: oligodendrocyte precursor cells; PER: pericytes; VECC/VECV/VECA: vascular endothelial cells, VSMCA: vascular smooth muscle cells. (**B**) As in (A) but for all neuronal clusters identified from our scRNA-seq

data and the 15 published hypothalamic neuronal clusters identified from the ~2000 hypothalamic neurons measured by Zeisel et al. (*69*). The cluster labels in Zeisel et al: DEGLU: diencephalon glutamatergic neuronal cluster; HYPEP: hypothalamus peptidergic neuronal cluster; DECHO: diencephalon cholinergic neuronal cluster; DEINH: diencephalon inhibitory neuronal cluster. (**C-E**) Scatter plots of the expression of genes within three example hypothalamic neuronal clusters identified by Zeisel et al. versus the corresponding gene expression observed in the most similar neuronal clusters described in our work. Expression is measured in z-score. The listed genes are among the top markers for each cluster. In general, our scRNA-seq data and those reported in Zeisel et al (*69*) are in good correlation. However, because we used ~10 times more neuronal scRNA-seq profiles to characterize about 1/5 of the whole hypothalamus (the preoptic region), and sequenced each cell to greater depth (we detected 1.6x more UMIs and genes per cell), we were able to analyze the preoptic region with a greater depth and thus gain substantially finer delineation of cell populations. Similar but moderately weaker correlations were observed between the expression profiles of our neuronal clusters and those of another scRNA-seq survey of the whole hypothalamus (*68*), perhaps because this older study contained only ~1,000 neurons from naïve animals.

Supplementary Tables

Table S1. Clustering of scRNA-seq data (provided as a separate xls file). "Cell name" contains a unique ID for each measured cell; "Sex" indicates the sex of the animal from which each cell was derived; "Replicate number" indicates the replicate measurement from which each cell was derived; "Cell class" indicates the major cell class assigned to each cell. An entry of 'ambiguous' indicates that the cell was marked as a putative doublet and was not further analyzed. An entry of 'unstable' indicates that the cell was not part of a stable cluster and was not further analyzed. "Non-neuronal cluster" indicates the distinct cluster within individual non-neuronal cell class to which each cell was assigned. The assignment of cells to major cell classes and sub-divisions of non-neuronal cell classes was derived from clustering analysis of all cells. "Neuronal cluster" indicates the neuronal cluster to which each cell was assigned, and this assignment was derived from clustering analysis of inhibitory neurons and excitatory neurons separately. The gene expression profile of each cell is provided in GEO, GSE113576.

Table S2. Gene expression in the neuronal clusters identified in scRNA-seq (provided as a separate xls file). Gene names are listed in the first column, and individual clusters are listed in subsequent columns. For each gene, the mean expression ("Mean expression"), the expression value corresponding to the 95% quantile ("95% quantile"), and the fraction of cells expressing at least one copy of that gene ("Fraction of cells expressing gene") are provided for each cluster. In addition to the genes listed in **Figs. 1**, **2**, **S3**, **S4**, S**5**, the provided genes are those used to cluster either the inhibitory or excitatory neurons as well as all genes measured with MERFISH.

Table S3. Genes categories used for gene-set enrichment analysis (provided as a separate xls file). Individual columns contain gene names associated with the specified gene categories in **Fig. 1E** and **fig. S6**.

Region	Abbreviation
Anteroventral periventricular nucleus	AvPe
Bed nucleus of the anterior commissure	BAC
Bed nucleus of the stria terminalis	BNST
Bed nucleus of the stria terminalis, medial division, anterolateral part	BNST-mal
Bed nucleus of the stria terminalis, medial division, ventral part	BNST-mv
Bed nucleus of the stria terminalis, posterior part	BNST-p
Lateral preoptic area	LPO
Medial preoptic area	MPA
Medial preoptic nucleus	MPN
Median preoptic nucleus	MnPO
Nucleus of the horizontal limb of the diagonal band	HDB
Olfactory tubercle	OT
Parastrial nucleus	PS
Paraventricular hypothalamic nucleus, anterior parvicellular	PaAP
Paraventricular thalamic nucleus	PVA
Periventricular hypothalamic nucleus	Pe
Septohypothalamic nucleus	SHy
Sexually dimpophic nucleus of the preoptic area	SDN-POA
Striohypothalamic nucleus	StHy
Suprachiasmatic Nucleus	SCN
Ventrolateral preoptic nucleus	VLPO
Ventromedial preoptic nucleus	VMPO

Table S4. Abbreviations of hypothalamic nuclei. “Region” contains the full name of different hypothalamic nuclei mentioned in this work. “Abbreviation” lists the common abbreviation for each of these nuclei.

Category	Function	Region	Marker genes	Ref.
REPRODUCTION	Puberty	Preoptic region	Gnrh1	(*86*)
	Puberty	MPO	Kisspeptin	(*87*)
	Sex differentiation	MPOA+BNST and other regions	Aromatase	(*41*, *88*)
	Sex differentiation	MPOA+BNST and other regions	Esr1	(*41*, *89*)
	Sex differentiation	MPOA+BNST and other regions	AR	(*90*)
SLEEP	Sleep	VLPO	Galanin	(*91*)
	Sleep	VLPO and vicinity	Cck, Crh, Tac1	(*11*)
SOCIAL BEHAVIORS	Male and female parenting	MPOA	Galanin	(*5*)
	Female parenting	AvPe	Th	(*6*)
	Male aggression	AvPe	Th	
	Female sexual motivation	MPOA	Nts	(*92*)
	Male-typical mating in males and females	MPOA	Esr1	(*64*)
	Parenting in males and females (virgin and mothers)	MPOA	Esr1	(*63*, *64*)
	Female sexual behavior	Rostral periventricular area	Kiss1	(*93*)
	Several male and female behaviors	MPOA/BNST and rest of hypothalamus	Sexually dimorphic genes	(*42*)
HUNTING	Hunting in males	MPOA	CamkIIalpha	(*94*)
FEEDING	Feeding	MPOA	Galanin	(*38*)
	Feeding	BNST	VGAT	(*95*)
HOMEOSTASIS	Warm sensing	VMPO	Adcyap1/Bdnf	(*8*)
	Thirst	MnPO	Agtr1a, Nxph4, Nos1	(*9*, *10*, *97*)
	Hypothermia/drinking	POA	Vglut2	(*96*)

Table S5. Summary of previous associations between marker genes and different hypothalamic nuclei and functions. "Category" summarizes the major functional role played by each neuronal population previously described. "Function" summarizes the biological function associated with each neuronal population. "Region" summarizes the hypothalamic

nuclei in which each neuronal population has been observed. “Marker genes” summarizes previously established marker genes for each population. “Ref.” summarizes the references associated with these previous observations.

Table S6. Summary of genes included in the MERFISH measurements (provided as a separate xls file). “Gene name” is the common name associated with each gene. Blank 1-5 are five MERFISH barcodes that are not used for any genes, providing a measure of the false-positive detection rate in MERFISH. “Barcode” represents the 16-bit binary barcode assigned to that gene or blank control for MERFISH measurements. An entry of ‘Sequential stain’ indicates that the given gene was measured in the non-combinatorial, sequential FISH imaging rounds after the combinatorial smFISH imaging run. “Readout probes” lists the names of the specific readout probes associated with each gene. The sequences of these probes are provided **in Table S10**. “Gene category” indicates whether the gene was selected based on prior biological knowledge or as derived from our scRNA-seq measurements.

Table S7. Example MERFISH data. Properties of all measured cells in one example slice from an animal are included in this table. “Cell ID” is a unique ID associated with each cell. “Animal ID” is a unique ID associated with animal. “Animal sex” is the gender of the animal in which the cell was imaged. “Behavior” describes the behavioral treatment of the animal. ‘Naïve’ indicates that no treatment was performed. “Bregma” indicates the approximate location of the slice in bregma coordinates. “Centroid X” is the x coordinate of the centroid position for the cell in μm. “Centroid Y” is the y coordinate of the centroid position for the cell in μm. “Cell class” lists the major cell class to which a cell was assigned. A value of “Ambiguous” represents cells that were identified as putative doublets and were not further analyzed. “Neuron cluster_ID” represents the neuronal cluster to which a cell was assigned. This field is empty if the cell was not a neuron. Columns with a gene name, e.g. ‘Ace2’, contain the expression values for that gene in that cell. Expression values for the 135 genes measured in the combinatorial smFISH run were determined as the total counts per cell divided by the cell volume and scaled by 1000. Expression values for the 20 genes measured in the non-combinatorial, sequential FISH rounds were arbitrary fluorescence units per μm^3, but the same scale is used for all cells. The 135 genes measured via combinatorial smFISH are listed first and the 20 genes measured in the non-combinatorial, sequential FISH rounds are listed last. Expression values have been batched corrected (*29*). A table containing these values for all cells measured via MERFISH for all replicates under all conditions is available on Dryad (doi:10.5061/dryad.8t8s248).

Table S8. Gene expression in the neuronal clusters identified in MERFISH (provided as a separate xls file). Gene names are listed in the first column, and individual clusters are listed in subsequent columns. For each gene, the mean expression (“Mean expression”), the expression value corresponding to the 95% quantile (“95% quantile”), and the fraction of cells expressing at least one copy of that gene (“Fraction of cells expressing gene”) are provided for each cluster. All 155 genes measured with MERFISH are included. Note that genes measured during the combinatorial smFISH rounds (see **Table S6**) are measured in RNA counts per μm^3 (scaled by 1000) while RNAs measured in the non-combinatorial, sequential FISH rounds (see **Table S6**) are measured in arbitrary fluorescence units per μm^3, but the same scale is used for all cells.

	scRNA-seq clusters	Notable genes observed in scRNA-seq		MERFISH clusters	Notable genes observed in MERFISH	Pearson correlation coefficients	Notes	Behavioral activation	Possible correspondence with previous functional studies	Ref
		Not included in the MERFISH library	Included in MERFISH library							
Excitatory clusters with putative correspondence between scRNA-seq and MERFISH	e2:Tac1/Fezf1	Fezf1	Trh, Tac1	E-13:BAC	Trh, Tac1	0.58				
	e3:Cartpt/Isl1		Cartpt, Isl1, Nts, Creb3l1	E-8:MPN/Pe/VMPO	Cartpt, Isl1, Nts, Creb3l1	0.58	Clusters E-8 and E-15 are differentiated in part by Creb3l1 expression	Male mating		
				E-15:MPN/VMPO	Cartpt, Isl1, Nts	0.53		Male mating		
	e4:Trh/Angpt1	Angpt1	Trh, Adcyap1	E-18:MPA/MPN	Trh, Adcyap1	0.75				
	e5:Adcyap1/Nkx2.1	Nkx2.1	Adcyap1, Crh low	E-7:MPA/MPN/StHy	Adcyap1	0.76	Clusters E-17 and E-31 are differentiated in part by Crh expression			
				E-31:Crh	Adcyap1, Crh	0.43				
	e6:Nos1/Trp73	Trp73	Nos1, Sema3c	E-6:MPA/PaAP/SHy	Nos1, Sema3c	0.69	Clusters E-6 and E-21 are differentiated in part by Nos1 expression			
				E-21:LPO/PS	Sema3c	0.52				
	e7:Reln/C1ql1	Reln, C1ql1, Agtr1a, Nxph4	Onecut2, Bmp7, Etv1, Nos1	E-2:AvPe/Pe/MnPO	Onecut2, Etv1, Nos1	0.39	Clusters E-2 and E-11 are differentiated in part by Etv1 and Bmp7 expression. ~25% of Cluster e7 cells express Agtr1a.		Agtr1a, Nxph4 and Nos1 cells involved in thirst	(*9*, *10*, *97*)
				E-11:Glut/Bmp7	Onecut2, Bmp7	0.7				
	e8:Cck/Ebf3		Cck, Ebf3, Penk, low Slc17a7	E-5:BNST	Cck, Ebf3, Penk	0.72	Clusters E-5 and E-10 are differentiated in part by Cck expression, Cluster E-25 expresses Slc17a7 but not Cck, Ebf3 or Penk.			
				E-10:BAC	Ebf3, Penk	0.38				
				E-25:BNST	Slc17a7	0.5				
	e10:Glut/Meis2	Meis2	Fst	E-20:VLPO	Fst	0.66				

	scRNA-seq clusters	Notable genes observed in scRNA-seq		MERFISH clusters	Notable genes observed in MERFISH	Pearson correlation coefficients	Notes	Behavioral activation	Possible correspondence with previous functional studies	Ref
		Not included in the MERFISH library	Included in MERFISH library							
Excitatory clusters with putative correspondence between scRNA-seq and MERFISH	e11:Glut/Shox2	Shox2	Ntng1, low Cck	E-9:PVA	Ntng1	0.6	Clusters E-9 and E-14 are differentiated in part by Cck expression			
				E-14:PVA	Ntng1, Cck	0.71				
	e12:Nos1/Foxp2	Foxp2	Nos1	E-4:AvPe/Pe/MnPO/MPA	Nos1	0.55				
	e13:Ghrh/C1ql1	Ghrh, C1ql1	Nup62cl, Onecut2	E-3:AvPe/Pe/VMPO/VLPO	Nup62cl, Onecut2	0.83			Adcyap1 cells involved in warm sensing, Tac3R/Kiss1R cells involved in hot flashes	(*8*, *58*)
	e15:Ucn3/Brs3		Brs3, Ucn3 low	E-1:AvPe/Pe/MnPO/VMPO	Brs3, Ucn3 low	0.77		Maternal		
	e16:Sst/Cartpt		Sst, Cartpt	E-23:PaAP/Pe	Sst, Cartpt	0.39				
	e17:Th/Adcyap1		Trh, Th, Omp, Adcyap1	E-16:MPN/PaAP/StHy	Trh, Th, Omp, Adcyap1	0.77				
	e19:Ghrh/Trh	Ghrh	Cartpt, Trh, Onecut2	E-19:Cartpt/Trh	Cartpt, Trh, Onecut2	0.68				
	e22:Gal/Ucn3		Gal, Ucn3, Trh, Etv1	E-28:PaAP	Ucn3, Gal, Trh, Etv1	0.77				
	e24:Gal/Rxfp1		Gal, Rxfp1, Etv1	E-26:Pe	Gal, Rxfp1, Etv1	0.67				
Excitatory clusters without correspondence	e1:Glut	Nkx2.1								
	e9:Glut/Tcf7l2	Tcf7l2								
	e23:Reln/Etv1	Reln	Etv1							
				E-12:MPA/MPN/StHy	Tac2, Nts, Mc4r					
				E-17:MPN	Ntsr1, Gda, Crhr1					
				E-22:Ndnf/Cplx3	Ndnf, Cplx3					
				E-24:MPA/VLPO	Col25a1, Crhr2					
				E-27:Glut/Necab1	Necab1, Tacr1, Crhr2					
				E-29:Sst	Sst, Gem, Tacr3					
				E-30:Gnrh1	Gnrh1, Isl1				Gnrh cells of the preoptic area	

	scRNA-seq clusters	Notable genes observed in scRNA-seq		MERFISH clusters	Notable genes observed in MERFISH	Pearson correlation coefficients	Notes	Behavioral activation	Possible correspondence with previous functional studies	Ref
		Not included in the MERFISH library	Included in MERFISH library							
Excitatory clusters outside of the region imaged by MERFISH	e20:Crh		Crh							
	e21:Glut/Rxfp3	Rxfp3								
Inhibitory clusters with putative correspondence between scRNA-seq and MERFISH	i2:Tac1/Pdyn	Pdyn	Tac1, Ano3	I-17:Tac1 (HDB)	Tac1, Ano3	0.74(i2)	Cluster I-17 does not split perhaps due to the lack of Pdyn, Drd1, and Prok2 in the MERFISH library.		This cluster may correspond to neurons manipulated by Chung et al, as sleep promoting.	(*11*)
	i10:Tac1/Nts	Drd1, Nts, Pdyn	Tac1, Ano3			0.72(i10)				
	i26:Tac1/Prok2	Prok2	Tac1, Ano3			0.51(i26)				
	i4:Gaba/Mylk		Penk, Sp9	I-9:BNST	Penk, Sp9	0.49				
	i5:Gaba/Pou3f3	Satb1, Pou3f3	Sox6, Necab1, Tac1	I-20:VLPO	Sox6	0.37	Clusters I-20 and I-26 are differentiated in part by Necab1 and Tac1 expression			
				I-26:VLPO	Sox6, Necab1, Tac1	0.7			This cluster may correspond to neurons manipulated by Chung et al, as sleep promoting.	(*11*)
	i7:Gaba		Scgn	I-23:BNST	Scgn	0.38				
	i8:Gal/Amigo2		Gal, Amigo2	I-7:MPN/VMPO	Gal, Amigo2, Dgkk	0.76	Clusters I-7 and I-31 are differentiated in part by Tac2 and Dgkk expression		Galanin cells involved in sleep	(*91*)
				I-31:LPO	Gal, Amigo2, Tac2	0.36				
	i9:Gaba		Ttn, Onecut2	I-30:AvPe/Pe/MnPO	Ttn, Onecut2	0.71				

	scRNA-seq clusters	Notable genes observed in scRNA-seq		MERFISH clusters	Notable genes observed in MERFISH	Pearson correlation coefficients	Notes	Behavioral activation	Possible correspondence with previous functional studies	Ref
		Not included in the MERFISH library	Included in MERFISH library							
Inhibitory clusters with putative correspondence between scRNA-seq and MERFISH	i16:Gal/Th		Gal, Th, Calcr, Brs3	I-14:MPA/MPN/StHy	Gal, Calcr, Brs3	0.4	Clusters I-14 and I-16 are differentiated in part by Th and Calcr/Brs3 expression	Virgin female maternal, maternal, paternal, male mating	Galanin cells involved in parenting	(*5*)
				I-16:AvPe/Pe/SHy	Gal, Th	0.61		Paternal, inter-male aggression, pup-directed aggression, female mating	Th cells involved in maternal behavior /Male aggression	(*6*)
	i17:Th/Nos1		Th, Nos1	I-3:BNST/SHy/PaAP/PS	Th, Nos1	0.79				
	i18:Gal/Tac2		Gal, Tac2, Nts	I-11:MPN/StHy	Gal	0.63	Clusters I-11, I-12, I-15, and I-24 are differentiated in part by expression of Gal, Nts, Egr2, and Tac2	Male mating		
				I-12:MPA/MPN/StHy	Nts	0.74				
				I-15:AvPe/Pe/VMPO	Egr2	0.46		Male mating, Female mating		
				I-24:BNST	Tac2	0.4		Pup-directed aggression		
	i20:Gal/Moxd1	Moxd1	Cplx3, Aromatase, low Cck, low Gal, low Cartpt	I-2:BNST/StHy/MPN	Cplx3, Aromatase	0.77	Clusters I-32, I-2 and I-34 are differentiated in part by Gal and Cartpt expression	Paternal, inter-male aggression, Male mating	Aromatase neurons of the sexually dimorphic nucleus	(*39*, *41*, *46*)
				I-32:BNST	Cplx3, Cck, Cartpt	0.38				
				I-34:MPN/StHy	Cplx3, Gal	0.51				
	i21:Sst/Pou3f3	Pou3f3	Sst low, Npy2,r	I-5:VLPO	Sst low, Npy2r	0.6				
	i22:Gal/Pmaip1	Pmaip1	Gal, Slc18a2	I-22:MnPO/Pe	Gal, Slc18a2	0.46				
	i23:Crh/Nts		Crh, Nts	I-10:MPA/PaAP/SHy	Crh, Nts	0.56		Maternal, paternal		
	i24:Nmu	Nmu	Penk, Serpinb1b	I-4:MPA/PaAP/SHy/VLPO	Penk, Serpinb1b	0.84				
	i25:Npy/Etv1	Npy	Etv1, Crhbp	I-29:Gaba/Etv1	Etv1, Crhbp	0.66				
	i29:Pnoc/Igsf1	Igsf1	Pnoc, Sytl4	I-13:BNST/StHy	Pnoc, Sytl4	0.72				
	i32:Sst/Npy	Npy	Sst, Nos1	I-25:BNST/Pe	Sst, Nos1	0.68				
	i37:Bdnf/Chrm2		Bdnf, Pchd11x	I-8:MPA/SHy	Bdnf, Pchd11x	0.73				

	scRNA-seq clusters	Notable genes observed in scRNA-seq		MERFISH clusters	Notable genes observed in MERFISH	Pearson correlation coefficients	Notes	Behavioral activation	Possible correspondence with previous functional studies	Ref
		Not included in the MERFISH library	Included in MERFISH library							
Inhibitory clusters without correspondence	i1:Gaba						This cluster is defined by a lack of markers and might represent a population of cells that have not been properly resolved			
	i11:Gaba	Nrgn	Nts, Penk							
	i12:Pnoc	Serpinb1b	Pnoc							
	i13:Gaba	Lhx8, Lypd6								
	i15:Gaba	Nr4a2, Fbxw13	Pnoc, Amigo2							
	i38:Kiss1/Th	Kiss1	Th						Kiss1 Neurons involved in puberty and female sexual behavior	(*87*, *93*)
	i39:Gaba	Lypd1, Pax6								
	i40:Sst/Reln	Reln	Sst							
	i41:Npy/Penk	Npy	Penk							
	i45:Bdnf/Pmaip1	Pmaip1	Bdnf							
				I-1:Gaba			This cluster is defined by a lack of markers and might represent a population of cells that have not been properly resolved			
				I-6:LPO/PS/SHy	Pnoc, Vgf, Tacr3					
				I-18:BNST	Cxcl14, Col25a, Sp91					
				I-19:AvPe/Pe/MnPO	Crhr2, Gpr165					
				I-21:Trh/Nts	Trh, Nts, Bmp7					
				I-27:BNST	Crh, Vgf, Cck, Creb3l1			Maternal		
				I-33:Gal/Cartpt	Gal, Cartpt, Galr1, Npy1r			Male mating		

	scRNA-seq clusters	Notable genes observed in scRNA-seq		MERFISH clusters	Notable genes observed in MERFISH	Pearson correlation coefficients	Notes	Behavioral activation	Possible correspondence with previous functional studies	Ref
		Not included in the MERFISH library	Included in MERFISH library							
Inhibitory clusters without correspondence				I-35:Gaba/Rxfp1	Rxfp1, Oxtr					
				I-36:Crh/Tac1	Crh, Tac1, Pnoc, Cckar, Cckbr					
				I-37:Gaba/Chat	Chat , Htr2c, Ntsr1					
				I-38:Gaba/Fbxw13	Fbxw13, Fezf1					
				I-39:Cartpt/Penk	Cartpt, Penk, Igf1r, Igf2r					
Inhibitory clusters outside of the region imaged by MERFISH	i3:Penk/Nts		Nts, Penk							
	i6:Avp/Nms	Avp, Nms								
	i14:Avp/Cck	Avp	Cck, Vgf							
	i27:Th/Trh		Trh Th							
	i28:Gaba/Six6	Six6								
	i30:Vip	Vip								
	i31:Calca	Calca								
	i35:Crh/Tac2	Pde11a	Crh, Tac2							
	i42:Pthlh	Pthlh								
	i43:Chat	Ntrk1, Slc18a3	Chat							
	i44:Th/Cxcl14	Cxcl14	Th							
Hybrid clusters	h1:Gaba/Slc17a6	Nr2f2	Slc17a6, Gad1							
	h2:Nts/Slc17a8		Nts, Slc17a8, Gad1	H-1:BNST	Nts, Slc17a8, Gad1	0.78				
Hybrid clusters outside of the region imaged by MERFISH	h3:Slc32a1/Gsc	Gsc, Slc32a1								

Table S9. Summary of properties of all neuronal clusters identified with scRNA-seq and MERFISH. "scRNA-seq clusters" represents the name of the scRNA-seq cluster. "Notable genes observed in scRNA-seq" lists notable genes expressed within each scRNA-seq cluster with genes divided into genes that were or were not included in the MERFISH panel. "MERFISH clusters" represents the name of the MERFISH cluster. "Notable genes observed in MERFISH" lists notable genes expressed within each MERFISH cluster. The most similar scRNA-seq and MERFISH clusters have been

grouped in the same row. Situations in which a single scRNA-seq cluster corresponds to multiple MERFISH clusters, or vice versa, are denoted by grouped rows that are colored identically (gray or white). “Pearson correlation coefficients” lists the Pearson correlation coefficients between the expression profiles of these most similar clusters. “Notes” provides a description of notable genes that differentiate multiple MERFISH clusters that are most similar to a single scRNA-seq cluster, or vice versa. The listed genes are only a subset of the genes that are differentially expressed and, thus, distinguish these clusters. “Behavioral activation” summarizes the behavioral condition in which each neuron was activated as identified in this work. “Possible correspondence with functional studies” lists possible associations with previously annotated neuronal populations. “Ref” provides the appropriate references for these previous populations. Excitatory, inhibitory, and hybrid clusters that do or do not have a correspondence between scRNA-seq and MERFISH as well as scRNA-seq clusters that are outside of the region imaged with MERFISH are grouped. We note that although we name the clusters by no more than two marker genes, these two-gene combinations do not always uniquely distinguish the cluster. For example, both e7:Reln/C1ql1 and e13:Ghrh/C1ql1express both Reln and C1ql1, and e7 is distinguished from e13, by the absence of Ghrh. See Materials and Methods (*29*) for detailed description of our naming strategy. In addition, we note that some of the MERFISH clusters still contain substantial representations in the most anterior or most posterior slice imaged, suggesting that their spatial distributions likely extend to nuclei outside the region imaged in this work, which are not represented in the spatial names provided for MERFISH clusters here. MERFISH cluster I-17:Tac1 (HDB) is localized at the edges of the imaged region and falls outside of the boundaries of the nearest imaged nuclei, the VLPO. Given its location, it is likely located within the HDB, a nuclei which is immediately next to but largely outside of the imaged region.

Table S10. Sequences of encoding probes used for MERFISH imaging (provided as a separate xls file). "Gene" contains the common gene name. "Isoform" contains the specific isoform ID used. Entries that list different isoform IDs for the same gene represent isoforms that could not be distinguished via bulk RNA-sequencing (*29*). "Probe" is the sequence of each encoding probe designed to the given gene.

Table S11. Sequences of readout probes used in the MERFISH measurements (provided as a separate xls file). "Readout probe name" is the name of each readout probe. "Sequence" is the sequence of each probe. "Fluorophore" is the name of the fluorophore attached to each probe. "Purpose" is a summary of what each readout probe is used to measure, either a specific bit associated with MERFISH barcodes used in the combinatorial smFISH run or an individual gene measured during the non-combinatorial, sequential FISH rounds.

References and Notes

1. L. W. Swanson, Cerebral hemisphere regulation of motivated behavior. *Brain Res.* **886**, 113–164 (2000). doi:10.1016/S0006-8993(00)02905-X Medline

2. R. B. Simerly, in *The Rat Nervous System*, G. Paxinos, Ed. (Elsevier, ed. 3, 2004), pp. 335–368.

3. S. M. Sternson, Hypothalamic survival circuits: Blueprints for purposive behaviors. *Neuron* **77**, 810–824 (2013). doi:10.1016/j.neuron.2013.02.018 Medline

4. J. Lu, A. A. Bjorkum, M. Xu, S. E. Gaus, P. J. Shiromani, C. B. Saper, Selective activation of the extended ventrolateral preoptic nucleus during rapid eye movement sleep. *J. Neurosci.* **22**, 4568–4576 (2002). doi:10.1523/JNEUROSCI.22-11-04568.2002 Medline

5. Z. Wu, A. E. Autry, J. F. Bergan, M. Watabe-Uchida, C. G. Dulac, Galanin neurons in the medial preoptic area govern parental behaviour. *Nature* **509**, 325–330 (2014). doi:10.1038/nature13307 Medline

6. N. Scott, M. Prigge, O. Yizhar, T. Kimchi, A sexually dimorphic hypothalamic circuit controls maternal care and oxytocin secretion. *Nature* **525**, 519–522 (2015). doi:10.1038/nature15378 Medline

7. K. Sokolowski, S. Esumi, T. Hirata, Y. Kamal, T. Tran, A. Lam, L. Oboti, S.-C. Brighthaupt, M. Zaghlula, J. Martinez, S. Ghimbovschi, S. Knoblach, A. Pierani, N. Tamamaki, N. M. Shah, K. S. Jones, J. G. Corbin, Specification of select hypothalamic circuits and innate behaviors by the embryonic patterning gene dbx1. *Neuron* **86**, 403–416 (2015). doi:10.1016/j.neuron.2015.03.022 Medline

8. C. L. Tan, E. K. Cooke, D. E. Leib, Y.-C. Lin, G. E. Daly, C. A. Zimmerman, Z. A. Knight, Warm-sensitive neurons that control body temperature. *Cell* **167**, 47–59.e15 (2016). doi:10.1016/j.cell.2016.08.028 Medline

9. D. E. Leib, C. A. Zimmerman, A. Poormoghaddam, E. L. Huey, J. S. Ahn, Y.-C. Lin, C. L. Tan, Y. Chen, Z. A. Knight, The forebrain thirst circuit drives drinking through negative reinforcement. *Neuron* **96**, 1272–1281.e4 (2017). doi:10.1016/j.neuron.2017.11.041 Medline

10. W. E. Allen, L. A. DeNardo, M. Z. Chen, C. D. Liu, K. M. Loh, L. E. Fenno, C. Ramakrishnan, K. Deisseroth, L. Luo, Thirst-associated preoptic neurons encode an aversive motivational drive. *Science* **357**, 1149–1155 (2017). doi:10.1126/science.aan6747 Medline

11. S. Chung, F. Weber, P. Zhong, C. L. Tan, T. N. Nguyen, K. T. Beier, N. Hörmann, W.-C. Chang, Z. Zhang, J. P. Do, S. Yao, M. J. Krashes, B. Tasic, A. Cetin, H. Zeng, Z. A. Knight, L. Luo, Y. Dan, Identification of preoptic sleep neurons using retrograde labelling and gene profiling. *Nature* **545**, 477–481 (2017). doi:10.1038/nature22350 Medline

12. J. F. Poulin, B. Tasic, J. Hjerling-Leffler, J. M. Trimarchi, R. Awatramani, Disentangling neural cell diversity using single-cell transcriptomics. *Nat. Neurosci.* **19**, 1131–1141 (2016). doi:10.1038/nn.4366 Medline

13. A. Tanay, A. Regev, Scaling single-cell genomics from phenomenology to mechanism. *Nature* **541**, 331–338 (2017). doi:10.1038/nature21350 Medline

14. H. Zeng, J. R. Sanes, Neuronal cell-type classification: Challenges, opportunities and the path forward. *Nat. Rev. Neurosci.* **18**, 530–546 (2017). doi:10.1038/nrn.2017.85 Medline

15. E. Lein, L. E. Borm, S. Linnarsson, The promise of spatial transcriptomics for neuroscience in the era of molecular cell typing. *Science* **358**, 64–69 (2017). doi:10.1126/science.aan6827 Medline

16. N. Crosetto, M. Bienko, A. van Oudenaarden, Spatially resolved transcriptomics and beyond. *Nat. Rev. Genet.* **16**, 57–66 (2015). doi:10.1038/nrg3832 Medline

17. J. M. Levsky, S. M. Shenoy, R. C. Pezo, R. H. Singer, Single-cell gene expression profiling. *Science* **297**, 836–840 (2002). doi:10.1126/science.1072241 Medline

18. R. Ke, M. Mignardi, A. Pacureanu, J. Svedlund, J. Botling, C. Wählby, M. Nilsson, In situ sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods* **10**, 857–860 (2013). doi:10.1038/nmeth.2563 Medline

19. J. H. Lee, E. R. Daugharthy, J. Scheiman, R. Kalhor, J. L. Yang, T. C. Ferrante, R. Terry, S. S. F. Jeanty, C. Li, R. Amamoto, D. T. Peters, B. M. Turczyk, A. H. Marblestone, S. A. Inverso, A. Bernard, P. Mali, X. Rios, J. Aach, G. M. Church, Highly multiplexed subcellular RNA sequencing in situ. *Science* **343**, 1360–1363 (2014). doi:10.1126/science.1250212 Medline

20. K. H. Chen, A. N. Boettiger, J. R. Moffitt, S. Wang, X. Zhuang, RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015). doi:10.1126/science.aaa6090 Medline

21. S. Shah, E. Lubeck, W. Zhou, L. Cai, In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus. *Neuron* **92**, 342–357 (2016). doi:10.1016/j.neuron.2016.10.001 Medline

22. X. Wang *et al.*, *Science* **361**, eaat5691 (2018).

23. M. E. Greenberg, E. B. Ziff, L. A. Greene, Stimulation of neuronal acetylcholine receptors induces rapid gene transcription. *Science* **234**, 80–83 (1986). doi:10.1126/science.3749894 Medline

24. A. M. Femino, F. S. Fay, K. Fogarty, R. H. Singer, Visualization of single RNA transcripts in situ. *Science* **280**, 585–590 (1998). doi:10.1126/science.280.5363.585 Medline

25. A. Raj, P. van den Bogaard, S. A. Rifkin, A. van Oudenaarden, S. Tyagi, Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5**, 877–879 (2008). doi:10.1038/nmeth.1253 Medline

26. J. R. Moffitt, J. Hao, D. Bambah-Mukku, T. Lu, C. Dulac, X. Zhuang, High-performance multiplexed fluorescence in situ hybridization in culture and tissue with matrix imprinting and clearing. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 14456–14461 (2016). doi:10.1073/pnas.1617699113 Medline

27. A. M. Klein, L. Mazutis, I. Akartuna, N. Tallapragada, A. Veres, V. Li, L. Peshkin, D. A. Weitz, M. W. Kirschner, Droplet barcoding for single-cell transcriptomics applied to

embryonic stem cells. *Cell* **161**, 1187–1201 (2015). doi:10.1016/j.cell.2015.04.044 Medline

28. 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 Medline

29. Materials and methods are available as supplementary materials.

30. J. H. Levine, E. F. Simonds, S. C. Bendall, K. L. Davis, A. D. Amir, M. D. Tadmor, O. Litvin, H. G. Fienberg, A. Jager, E. R. Zunder, R. Finck, A. L. Gedman, I. Radtke, J. R. Downing, D. Pe'er, G. P. Nolan, Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. *Cell* **162**, 184–197 (2015). doi:10.1016/j.cell.2015.05.047 Medline

31. 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 Medline

32. S. M. Wojcik, S. Katsurabayashi, I. Guillemin, E. Friauf, C. Rosenmund, N. Brose, J.-S. Rhee, A shared vesicular carrier allows synaptic corelease of GABA and glycine. *Neuron* **50**, 575–587 (2006). doi:10.1016/j.neuron.2006.04.016 Medline

33. R. A. Romanov, A. Zeisel, J. Bakker, F. Girach, A. Hellysaz, R. Tomer, A. Alpár, J. Mulder, F. Clotman, E. Keimpema, B. Hsueh, A. K. Crow, H. Martens, C. Schwindling, D. Calvigioni, J. S. Bains, Z. Máté, G. Szabó, Y. Yanagawa, M.-D. Zhang, A. Rendeiro, M. Farlik, M. Uhlén, P. Wulff, C. Bock, C. Broberger, K. Deisseroth, T. Hökfelt, S. Linnarsson, T. L. Horvath, T. Harkany, Molecular interrogation of hypothalamic organization reveals distinct dopamine neuronal subtypes. *Nat. Neurosci.* **20**, 176–188 (2017). doi:10.1038/nn.4462 Medline

34. S. J. Shabel, C. D. Proulx, J. Piriz, R. Malinow, Mood regulation. GABA/glutamate co-release controls habenula output and is modified by antidepressant treatment. *Science* **345**, 1494–1498 (2014). doi:10.1126/science.1250469 Medline

35. E. S. Lein, M. J. Hawrylycz, N. Ao, M. Ayres, A. Bensinger, A. Bernard, A. F. Boe, M. S. Boguski, K. S. Brockway, E. J. Byrnes, L. Chen, L. Chen, T.-M. Chen, M. C. Chin, J. Chong, B. E. Crook, A. Czaplinska, C. N. Dang, S. Datta, N. R. Dee, A. L. Desaki, T. Desta, E. Diep, T. A. Dolbeare, M. J. Donelan, H.-W. Dong, J. G. Dougherty, B. J. Duncan, A. J. Ebbert, G. Eichele, L. K. Estin, C. Faber, B. A. Facer, R. Fields, S. R. Fischer, T. P. Fliss, C. Frensley, S. N. Gates, K. J. Glattfelder, K. R. Halverson, M. R. Hart, J. G. Hohmann, M. P. Howell, D. P. Jeung, R. A. Johnson, P. T. Karr, R. Kawal, J. M. Kidney, R. H. Knapik, C. L. Kuan, J. H. Lake, A. R. Laramee, K. D. Larsen, C. Lau, T. A. Lemon, A. J. Liang, Y. Liu, L. T. Luong, J. Michaels, J. J. Morgan, R. J. Morgan, M. T. Mortrud, N. F. Mosqueda, L. L. Ng, R. Ng, G. J. Orta, C. C. Overly, T. H. Pak, S. E. Parry, S. D. Pathak, O. C. Pearson, R. B. Puchalski, Z. L. Riley, H. R. Rockett, S. A. Rowland, J. J. Royall, M. J. Ruiz, N. R. Sarno, K. Schaffnit, N. V. Shapovalova, T.

Sivisay, C. R. Slaughterbeck, S. C. Smith, K. A. Smith, B. I. Smith, A. J. Sodt, N. N. Stewart, K.-R. Stumpf, S. M. Sunkin, M. Sutram, A. Tam, C. D. Teemer, C. Thaller, C. L. Thompson, L. R. Varnam, A. Visel, R. M. Whitlock, P. E. Wohnoutka, C. K. Wolkey, V. Y. Wong, M. Wood, M. B. Yaylaoglu, R. C. Young, B. L. Youngstrom, X. F. Yuan, B. Zhang, T. A. Zwingman, A. R. Jones, Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2007). doi:10.1038/nature05453 Medline

36. 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 Medline

37. T. Shimogori, D. A. Lee, A. Miranda-Angulo, Y. Yang, H. Wang, L. Jiang, A. C. Yoshida, A. Kataoka, H. Mashiko, M. Avetisyan, L. Qi, J. Qian, S. Blackshaw, A genomic atlas of mouse hypothalamic development. *Nat. Neurosci.* **13**, 767–775 (2010). doi:10.1038/nn.2545 Medline

38. Z. A. Knight, K. Tan, K. Birsoy, S. Schmidt, J. L. Garrison, R. W. Wysocki, A. Emiliano, M. I. Ekstrand, J. M. Friedman, Molecular profiling of activated neurons by phosphorylated ribosome capture. *Cell* **151**, 1126–1137 (2012). doi:10.1016/j.cell.2012.10.039 Medline

39. R. B. Simerly, Wired for reproduction: Organization and development of sexually dimorphic circuits in the mammalian forebrain. *Annu. Rev. Neurosci.* **25**, 507–536 (2002). doi:10.1146/annurev.neuro.25.112701.142745 Medline

40. J. R. Moffitt, J. Hao, G. Wang, K. H. Chen, H. P. Babcock, X. Zhuang, High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 11046–11051 (2016). doi:10.1073/pnas.1612826113 Medline

41. M. V. Wu, D. S. Manoli, E. J. Fraser, J. K. Coats, J. Tollkuhn, S. Honda, N. Harada, N. M. Shah, Estrogen masculinizes neural pathways and sex-specific behaviors. *Cell* **139**, 61–72 (2009). doi:10.1016/j.cell.2009.07.036 Medline

42. X. Xu, J. K. Coats, C. F. Yang, A. Wang, O. M. Ahmed, M. Alvarado, T. Izumi, N. M. Shah, Modular genetic control of sexually dimorphic behaviors. *Cell* **148**, 596–607 (2012). doi:10.1016/j.cell.2011.12.018 Medline

43. C. Gregg, J. Zhang, B. Weissbourd, S. Luo, G. P. Schroth, D. Haig, C. Dulac, High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science* **329**, 643–648 (2010). doi:10.1126/science.1190830 Medline

44. T. Kudo, M. Uchigashima, T. Miyazaki, K. Konno, M. Yamasaki, Y. Yanagawa, M. Minami, M. Watanabe, Three types of neurochemical projection from the bed nucleus of the stria terminalis to the ventral tegmental area in adult mice. *J. Neurosci.* **32**, 18035–18046 (2012). doi:10.1523/JNEUROSCI.4057-12.2012 Medline

45. G. Paxinos, K. Franklin, *The Mouse Brain in Stereotaxic Coordinates* (Academic Press, ed. 3, 2007).

46. Y. Tsuneoka, S. Tsukahara, S. Yoshida, K. Takase, S. Oda, M. Kuroda, H. Funato, Moxd1 is a marker for sexual dimorphism in the medial preoptic area, bed nucleus of the stria

terminalis and medial amygdala. *Front. Neuroanat.* **11**, 26 (2017). doi:10.3389/fnana.2017.00026 Medline

47. B. J. Marlin, R. C. Froemke, Oxytocin modulation of neural circuits for social behavior. *Dev. Neurobiol.* **77**, 169–189 (2017). doi:10.1002/dneu.22452 Medline

48. M. Zelikowsky, M. Hui, T. Karigo, A. Choe, B. Yang, M. R. Blanco, K. Beadle, V. Gradinaru, B. E. Deverman, D. J. Anderson, The neuropeptide Tac2 controls a distributed brain state induced by chronic social isolation stress. *Cell* **173**, 1265–1279.e19 (2018). doi:10.1016/j.cell.2018.03.037 Medline

49. R. Andero, B. G. Dias, K. J. Ressler, A role for Tac2, NkB, and Nk3 receptor in normal and dysregulated fear memory consolidation. *Neuron* **83**, 444–454 (2014). doi:10.1016/j.neuron.2014.05.028 Medline

50. Y. Tsuneoka, T. Maruyama, S. Yoshida, K. Nishimori, T. Kato, M. Numan, K. O. Kuroda, Functional, anatomical, and neurochemical differentiation of medial preoptic area subregions in relation to maternal behavior in the mouse. *J. Comp. Neurol.* **521**, 1633–1663 (2013). doi:10.1002/cne.23251 Medline

51. T. Gallopin, P. Fort, E. Eggermann, B. Cauli, P.-H. Luppi, J. Rossier, E. Audinat, M. Mühlethaler, M. Serafin, Identification of sleep-promoting neurons in vitro. *Nature* **404**, 992–995 (2000). doi:10.1038/35010109 Medline

52. J. Lu, M. A. Greco, P. Shiromani, C. B. Saper, Effect of lesions of the ventrolateral preoptic nucleus on NREM and REM sleep. *J. Neurosci.* **20**, 3830–3842 (2000). doi:10.1523/JNEUROSCI.20-10-03830.2000 Medline

53. Z. D. Zhao, W. Z. Yang, C. Gao, X. Fu, W. Zhang, Q. Zhou, W. Chen, X. Ni, J.-K. Lin, J. Yang, X.-H. Xu, W. L. Shen, A hypothalamic circuit that controls body temperature. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 2042–2047 (2017). doi:10.1073/pnas.1616255114 Medline

54. C. L. Sisk, D. L. Foster, The neural basis of puberty and adolescence. *Nat. Neurosci.* **7**, 1040–1047 (2004). doi:10.1038/nn1326 Medline

55. H. Yoon, L. W. Enquist, C. Dulac, Olfactory inputs to hypothalamic neurons controlling reproduction and fertility. *Cell* **123**, 669–682 (2005). doi:10.1016/j.cell.2005.08.039 Medline

56. U. Boehm, Z. Zou, L. B. Buck, Feedback loops link odor and pheromone signaling with reproduction. *Cell* **123**, 683–695 (2005). doi:10.1016/j.cell.2005.09.027 Medline

57. J. T. George, S. B. Seminara, Kisspeptin and the hypothalamic control of reproduction: Lessons from the human. *Endocrinology* **153**, 5130–5136 (2012). doi:10.1210/en.2012-1429 Medline

58. S. L. Padilla, C. W. Johnson, F. D. Barker, M. A. Patterson, R. D. Palmiter, A neural circuit underlying the generation of hot flushes. *Cell Reports* **24**, 271–277 (2018). doi:10.1016/j.celrep.2018.06.037 Medline

59. F. S. vom Saal, Time-contingent change in infanticide and parental behavior induced by ejaculation in male mice. *Physiol. Behav.* **34**, 7–15 (1985). doi:10.1016/0031-9384(85)90069-1 Medline

60. J. Kohl, B. M. Babayan, N. D. Rubinstein, A. E. Autry, B. Marin-Rodriguez, V. Kapoor, K. Miyamishi, L. S. Zweifel, L. Luo, N. Uchida, C. Dulac, Functional circuit architecture underlying parental behaviour. *Nature* **556**, 326–331 (2018). doi:10.1038/s41586-018-0027-0 Medline

61. C. Dulac, L. A. O'Connell, Z. Wu, Neural control of maternal and paternal behaviors. *Science* **345**, 765–770 (2014). doi:10.1126/science.1253291 Medline

62. G. C. Teskey, M. Kavaliers, Effects of opiate agonists and antagonists on aggressive encounters and subsequent opioid-induced analgesia, activity and feeding responses in male mice. *Pharmacol. Biochem. Behav.* **31**, 43–52 (1988). doi:10.1016/0091-3057(88)90309-7 Medline

63. Y.-Y. Fang, T. Yamaguchi, S. C. Song, N. X. Tritsch, D. Lin, A hypothalamic midbrain pathway essential for driving maternal behaviors. *Neuron* **98**, 192–207.e10 (2018). doi:10.1016/j.neuron.2018.02.019 Medline

64. Y.-C. Wei, S.-R. Wang, Z.-L. Jiao, W. Zhang, J.-K. Lin, X.-Y. Li, S.-S. Li, X. Zhang, X.-H. Xu, Medial preoptic area in mice is capable of mediating sexually dimorphic behaviors regardless of gender. *Nat. Commun.* **9**, 279 (2018). doi:10.1038/s41467-017-02648-0 Medline

65. S. Arber, Motor circuits in action: Specification, connectivity, and function. *Neuron* **74**, 975–989 (2012). doi:10.1016/j.neuron.2012.05.011 Medline

66. A. Paul, M. Crow, R. Raudales, M. He, J. Gillis, Z. J. Huang, Transcriptional architecture of synaptic communication delineates GABAergic neuron identity. *Cell* **171**, 522–539.e20 (2017). doi:10.1016/j.cell.2017.08.032 Medline

67. H. Li, F. Horns, B. Wu, Q. Xie, J. Li, T. Li, D. J. Luginbuhl, S. R. Quake, L. Luo, Classifying *Drosophila* olfactory projection neuron subtypes by single-cell RNA sequencing. *Cell* **171**, 1206–1220.e22 (2017). doi:10.1016/j.cell.2017.10.019 Medline

68. R. Chen, X. Wu, L. Jiang, Y. Zhang, Single-cell RNA-seq reveals hypothalamic cell diversity. *Cell Reports* **18**, 3227–3241 (2017). doi:10.1016/j.celrep.2017.03.004 Medline

69. A. Zeisel, H. Hochgerner, P. Lönnerberg, A. Johnsson, F. Memic, J. van der Zwan, M. Häring, E. Braun, L. E. Borm, G. La Manno, S. Codeluppi, A. Furlan, K. Lee, N. Skene, K. D. Harris, J. Hjerling-Leffler, E. Arenas, P. Ernfors, U. Marklund, S. Linnarsson, Molecular architecture of the mouse nervous system. *Cell* **174**, 999–1014.e22 (2018). doi:10.1016/j.cell.2018.06.021 Medline

70. J. N. Campbell, E. Z. Macosko, H. Fenselau, T. H. Pers, A. Lyubetskaya, D. Tenen, M. Goldman, A. M. J. Verstegen, J. M. Resch, S. A. McCarroll, E. D. Rosen, B. B. Lowell, L. T. Tsai, A molecular census of arcuate hypothalamus and median eminence cell types. *Nat. Neurosci.* **20**, 484–496 (2017). doi:10.1038/nn.4495 Medline

71. S. Blackshaw, S. Scholpp, M. Placzek, H. Ingraham, R. Simerly, T. Shimogori, Molecular pathways controlling development of thalamus and hypothalamus: From neural specification to circuit formation. *J. Neurosci.* **30**, 14925–14930 (2010). doi:10.1523/JNEUROSCI.4499-10.2010 Medline

72. T. Kimchi, J. Xu, C. Dulac, A functional circuit underlying male sexual behaviour in the female mouse brain. *Nature* **448**, 1009–1014 (2007). doi:10.1038/nature06089 Medline

73. S. Pandey, K. Shekhar, A. Regev, A. F. Schier, Comprehensive identification and spatial mapping of habenular neuronal types using single-cell RNA-seq. *Curr. Biol.* **28**, 1052–1065.e7 (2018). doi:10.1016/j.cub.2018.02.040 Medline

74. L. Stowers, T. E. Holy, M. Meister, C. Dulac, G. Koentges, Loss of sex discrimination and male-male aggression in mice deficient for TRP2. *Science* **295**, 1493–1500 (2002). doi:10.1126/science.1069259 Medline

75. 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 Medline

76. N. Horii-Hayashi, T. Sasagawa, W. Matsunaga, M. Nishi, Development and structural variety of the chondroitin sulfate proteoglycans-contained extracellular matrix in the mouse brain. *Neural Plast.* **2015**, 256389 (2015). doi:10.1155/2015/256389 Medline

77. T. Ilicic, J. K. Kim, A. A. Kolodziejczyk, F. O. Bagger, D. J. McCarthy, J. C. Marioni, S. A. Teichmann, Classification of low quality cells from single-cell RNA-seq data. *Genome Biol.* **17**, 29 (2016). doi:10.1186/s13059-016-0888-1 Medline

78. S. C. van den Brink, F. Sage, Á. Vértesy, B. Spanjaard, J. Peterson-Maduro, C. S. Baron, C. Robin, A. van Oudenaarden, Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nat. Methods* **14**, 935–936 (2017). doi:10.1038/nmeth.4437 Medline

79. F. Buettner, K. N. Natarajan, F. P. Casale, V. Proserpio, A. Scialdone, F. J. Theis, S. A. Teichmann, J. C. Marioni, O. Stegle, Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat. Biotechnol.* **33**, 155–160 (2015). doi:10.1038/nbt.3102 Medline

80. L. van der Maaten, *J. Mach. Learn. Res.* **15**, 3221–3245 (2014).

81. A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, J. P. Mesirov, Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550 (2005). doi:10.1073/pnas.0506580102 Medline

82. J. D. Perez, N. D. Rubinstein, D. E. Fernandez, S. W. Santoro, L. A. Needleman, O. Ho-Shing, J. J. Choi, M. Zirlinger, S.-K. Chen, J. S. Liu, C. Dulac, Quantitative and functional interrogation of parent-of-origin allelic expression biases in the brain. *eLife* **4**, e07860 (2015). doi:10.7554/eLife.07860 Medline

83. G. Wang, J. R. Moffitt, X. Zhuang, Multiplexed imaging of high-density libraries of RNAs with MERFISH and expansion microscopy. *Sci. Rep.* **8**, 4847 (2018). doi:10.1038/s41598-018-22297-7 Medline

84. I. Rasnik, S. A. McKinney, T. Ha, Nonblinking and long-lasting single-molecule fluorescence imaging. *Nat. Methods* **3**, 891–893 (2006). doi:10.1038/nmeth934 Medline

85. C. E. Aitken, R. A. Marshall, J. D. Puglisi, An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophys. J.* **94**, 1826–1835 (2008). doi:10.1529/biophysj.107.117689 Medline

86. T. J. Wu, M. J. Gibson, M. C. Rogers, A. J. Silverman, New observations on the development of the gonadotropin-releasing hormone system in the mouse. *J. Neurobiol.* **33**, 983–998 (1997). doi:10.1002/(SICI)1097-4695(199712)33:7<983:AID-NEU9>3.0.CO;2-4 Medline

87. M. L. Gottsch, M. J. Cunningham, J. T. Smith, S. M. Popa, B. V. Acohido, W. F. Crowley, S. Seminara, D. K. Clifton, R. A. Steiner, A role for kisspeptins in the regulation of gonadotropin secretion in the mouse. *Endocrinology* **145**, 4073–4077 (2004). doi:10.1210/en.2004-0431 Medline

88. N. J. MacLusky, F. Naftolin, Sexual differentiation of the central nervous system. *Science* **211**, 1294–1302 (1981). doi:10.1126/science.6163211 Medline

89. K. M. Lenz, B. M. Nugent, R. Haliyur, M. M. McCarthy, Microglia are essential to masculinization of brain and behavior. *J. Neurosci.* **33**, 2761–2772 (2013). doi:10.1523/JNEUROSCI.1268-12.2013 Medline

90. S. A. Juntti, J. Tollkuhn, M. V. Wu, E. J. Fraser, T. Soderborg, S. Tan, S. Honda, N. Harada, N. M. Shah, The androgen receptor governs the execution, but not programming, of male sexual and territorial behaviors. *Neuron* **66**, 260–272 (2010). doi:10.1016/j.neuron.2010.03.024 Medline

91. J. E. Sherin, J. K. Elmquist, F. Torrealba, C. B. Saper, Innervation of histaminergic tuberomammillary neurons by GABAergic and galaninergic neurons in the ventrolateral preoptic nucleus of the rat. *J. Neurosci.* **18**, 4705–4721 (1998). doi:10.1523/JNEUROSCI.18-12-04705.1998 Medline

92. J. A. McHenry, J. M. Otis, M. A. Rossi, J. E. Robinson, O. Kosyk, N. W. Miller, Z. A. McElligott, E. A. Budygin, D. R. Rubinow, G. D. Stuber, Hormonal gain control of a medial preoptic area social reward circuit. *Nat. Neurosci.* **20**, 449–458 (2017). doi:10.1038/nn.4487 Medline

93. V. Hellier, O. Brock, M. Candlish, E. Desroziers, M. Aoki, C. Mayer, R. Piet, A. Herbison, W. H. Colledge, V. Prévot, U. Boehm, J. Bakker, Female sexual behavior in mice is controlled by kisspeptin neurons. *Nat. Commun.* **9**, 400 (2018). doi:10.1038/s41467-017-02797-2 Medline

94. S.-G. Park, Y.-C. Jeong, D.-G. Kim, M.-H. Lee, A. Shin, G. Park, J. Ryoo, J. Hong, S. Bae, C.-H. Kim, P.-S. Lee, D. Kim, Medial preoptic circuit induces hunting-like actions to target objects and prey. *Nat. Neurosci.* **21**, 364–372 (2018). doi:10.1038/s41593-018-0072-x Medline

95. J. H. Jennings, D. R. Sparta, A. M. Stamatakis, R. L. Ung, K. E. Pleil, T. L. Kash, G. D. Stuber, Distinct extended amygdala circuits for divergent motivational states. *Nature* **496**, 224–228 (2013). doi:10.1038/nature12041 Medline

96. S. B. G. Abbott, C. B. Saper, Median preoptic glutamatergic neurons promote thermoregulatory heat loss and water consumption in mice. *J. Physiol.* **595**, 6569–6583 (2017). doi:10.1113/JP274667 Medline

97. V. Augustine, S. K. Gokce, S. Lee, B. Wang, T. J. Davidson, F. Reimann, F. Gribble, K. Deisseroth, C. Lois, Y. Oka, Hierarchical neural architecture underlying thirst regulation. *Nature* **555**, 204–209 (2018). doi:10.1038/nature25488 Medline