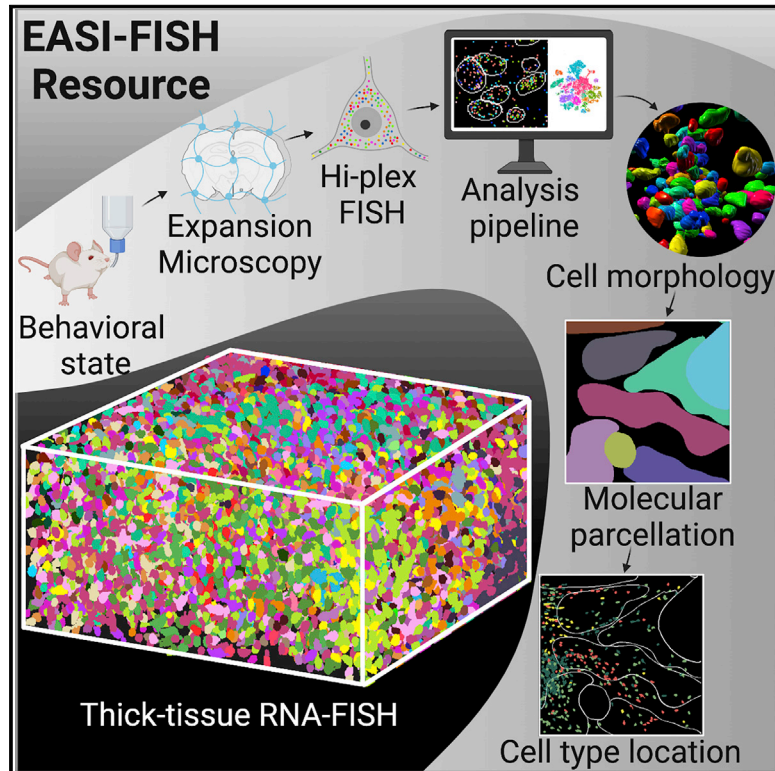# EASI-FISH for thick tissue defines lateral hypothalamus spatio-molecular organization

Graphical abstract

Authors

Yuhan Wang, Mark Eddison,
Greg Fleishman, ..., Wyatt Korff,
Scott M. Sternson, Paul W. Tillberg

Correspondence

ssternson@health.ucsd.edu (S.M.S.),
tillbergp@janelia.hhmi.org (P.W.T.)

In brief

Multiplexed FISH in thick tissue sections with deep-learning based 3D segmentation enables facile characterization of spatial organization for different cell types in defined brain regions.

Highlights

- EASI-FISH enables robust gene expression profiling in thick brain slices

- A turnkey analysis pipeline for facile analysis of large EASI-FISH image datasets

- EASI-FISH reveals molecularly defined subregions of the lateral hypothalamus

- Cell type subdivisions corresponding to morphology, functional tuning, and location

CellPress

Cell

Resource

EASI-FISH for thick tissue defines lateral hypothalamus spatio-molecular organization

Yuhan Wang,[1] Mark Eddison,[1] Greg Fleishman,[1] Martin Weigert,[2,3] Shengjin Xu,[1,8] Tim Wang,[1] Konrad Rokicki,[1] Cristian Goina,[1] Fredrick E. Henry,[1,9] Andrew L. Lemire,[1] Uwe Schmidt,[2] Hui Yang,[1,4] Karel Svoboda,[1] Eugene W. Myers,[2] Stephan Saalfeld,[1] Wyatt Korff,[1] Scott M. Sternson,[1,5,6,7,10,*] and Paul W. Tillberg[1,6,7,*]

[1]Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA 20147, USA
[2]Max Planck Institute of Molecular Cell Biology and Genetics, 01307 Dresden, Germany
[3]Institute of Bioengineering, School of Life Sciences, EPFL, 1015 Lausanne, Switzerland
[4]Dominick P. Purpura Department of Neuroscience, Albert Einstein College of Medicine, Bronx, New York 10461, USA
[5]Department of Neurosciences, Howard Hughes Medical Institute, University of California, San Diego, La Jolla, CA 92093, USA
[6]These authors contributed equally
[7]Senior authors
[8]Present address: Institute of Neuroscience, State Key Laboratory of Neuroscience, CAS Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai 200031, China
[9]Present address: Adaptive Biotechnologies, 1551 Eastlake Ave E, Ste 200, Seattle, WA 98102, USA
[10]Lead contact
*Correspondence: ssternson@health.ucsd.edu (S.M.S.), tillbergp@janelia.hhmi.org (P.W.T.)
https://doi.org/10.1016/j.cell.2021.11.024

SUMMARY

Determining the spatial organization and morphological characteristics of molecularly defined cell types is a major bottleneck for characterizing the architecture underpinning brain function. We developed Expansion-Assisted Iterative Fluorescence *In Situ* Hybridization (EASI-FISH) to survey gene expression in brain tissue, as well as a turnkey computational pipeline to rapidly process large EASI-FISH image datasets. EASI-FISH was optimized for thick brain sections (300 μm) to facilitate reconstruction of spatio-molecular domains that generalize across brains. Using the EASI-FISH pipeline, we investigated the spatial distribution of dozens of molecularly defined cell types in the lateral hypothalamic area (LHA), a brain region with poorly defined anatomical organization. Mapping cell types in the LHA revealed nine spatially and molecularly defined subregions. EASI-FISH also facilitates iterative reanalysis of scRNA-seq datasets to determine marker-genes that further dissociated spatial and morphological heterogeneity. The EASI-FISH pipeline democratizes mapping molecularly defined cell types, enabling discoveries about brain organization.

INTRODUCTION

Molecular definition of cell types provides a way to classify neurons, and marker-genes can be leveraged to interrogate the function of cell types in neural circuits (Luo et al., 2018). Single-cell RNA sequencing (scRNA-seq) has greatly accelerated discovery of molecularly defined cell types (Saunders et al., 2018; Tasic et al., 2018; Zeisel et al., 2018). Fluorescence *in situ* hybridization (FISH) is well suited to establish the spatial organization of cell types identified from scRNA-seq, but this requires mapping the co-expression patterns of dozens of genes in the same cells in three-dimensional (3D) tissue volumes. Although several multiplex FISH methods have been reported (Chen et al., 2015b; Codeluppi et al., 2018; Moffitt et al., 2018; Nicovich et al., 2019; Qian et al., 2020; Shah et al., 2016; Wang et al., 2018), it remains a bottleneck for many research labs due to complex procedures and challenging computational analyses. Additionally, most methods are restricted to thin tissue

sections (10-20 μm), obscuring three-dimensional (3D) relationships of cell types within brain structures.

To overcome these limitations, we developed Expansion-Assisted Iterative-FISH (EASI-FISH) with multi-round multiplexed RNA-FISH in 300-μm-thick brain sections. Expansion microscopy (ExM) (Chen et al., 2015a; Tillberg et al., 2016) is advantageous for high-resolution imaging in thick tissue, allowing 3D reconstruction of tissue volumes. We also developed an analysis pipeline for rapid and automated data processing.

We applied EASI-FISH to the mouse lateral hypothalamic area (LHA), an important motivational center regulating ingestive, social, arousal, and autonomic functions (Bernardis and Bellinger, 1993; Petrovich, 2018; Rossi et al., 2021; Stuber and Wise, 2016). Despite extensive functional investigation, an understanding of the LHA is limited by poor anatomical definition. We mapped molecularly defined cell type markers identified from LHA scRNA-seq datasets and uncovered an unexpected molecular parcellation of the LHA not previously identified from
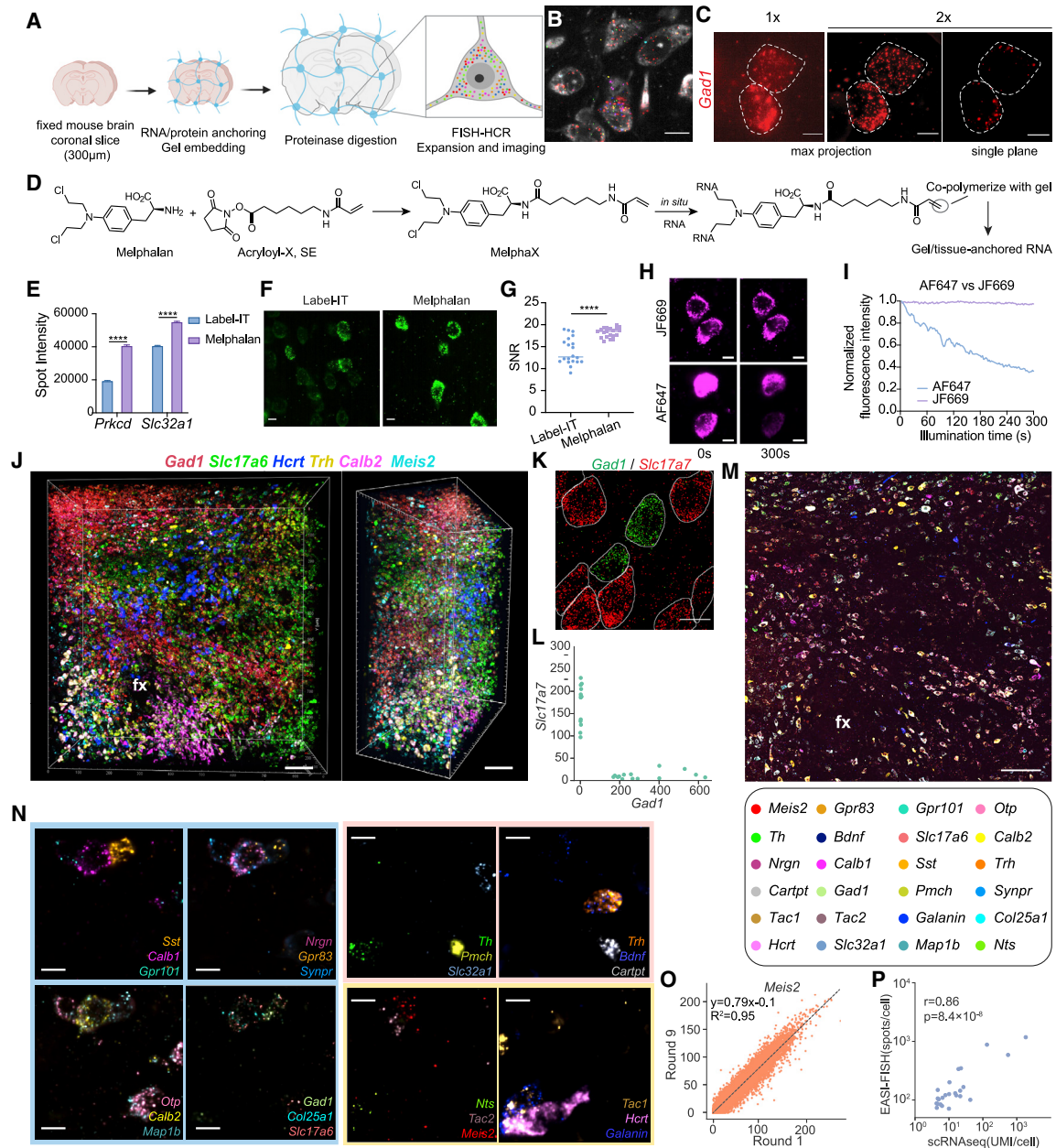
Figure 1. EASI-FISH method

(A) Schematic of the EASI-FISH process.

(B) Example EASI-FISH image. Scale bar: 10 μm.

(C) Axial projection and single-plane images of *Gad1* expression in two cortical neurons before (1×) and after (2×) expansion. Scale bar: 5 μm.

(D) Chemical reaction of Melphalan with Acryloyl-X, SE to produce MelphaX, which reacts with RNA and can be incorporated into a hydrogel matrix.

(E) Spot fluorescence intensities with RNA anchored by Label-IT (0.1 mg/ml) or Melphalan (0.1 mg/ml). ****p < 0.0001. Error bars: SEM. Statistics: Table S1.

(F and G) (F) Representative FISH images (*Gad1* in cortex) and (G) signal-to-noise ratio (SNR) from Label-IT or Melphalan-anchored tissue samples. Scale bar: 5 μm. SNR = Avg (pixel values) / SD (pixel values). ****p < 0.0001. Statistics: Table S1.

(H and I) (H) Representative images and (I) quantification comparing photostability of hairpins conjugated with Alexa Fluor 647 (AF647) and Janelia Fluor 669 (JF669). Scale bar: 5 μm.

(J) 3D rendering of a thick tissue volume generated by EASI-FISH with 2 rounds of 3-plex FISH. Scale bar:100 μm. fx: fornix.

(K and L) *Slc17a7* and *Gad1* expression in the cortex. Scale bar: 10 μm.

(M) LHA tissue volume with 24 marker-genes (single optical plane shown). Genes and colormap listed below. Tissue dimension: 0.8 mm × 0.8 mm × 0.3 mm. Scale bar: 100 μm. fx: fornix.

(legend continued on next page)

cell density measurements. Our results demonstrate the capability of the EASI-FISH data acquisition and analysis pipeline to reveal the organization of cell types in the LHA that underlie the multifaceted functions of this important brain area.

RESULTS

EASI-FISH protocol development

We designed and implemented EASI-FISH (Figure 1A) in thick tissue sections from cortex, central amygdala (CeA), and LHA by building on expansion microscopy enhanced smFISH (exFISH) (Chen et al., 2016), where tissue is physically expanded by embedding in a swellable hydrogel (Chen et al., 2015a; Tillberg et al., 2016). As in exFISH, uniform expansion is achieved by proteolytic digestion of the embedded tissue, while preserving RNA via covalent attachment to the hydrogel mesh for detection by FISH methods, including amplification with the hybridization chain reaction (HCR) (Figures 1A and 1B). Expansion reduces tissue autofluorescence and light scattering. It also increases the effective imaging resolution, increasing the number of individual RNA molecules that may be resolved per cell (Figure 1C).

We optimized the procedure for improved detection accuracy and robust sample processing across multiple rounds. First, to covalently anchor RNA molecules to the hydrogel, we used the bis-nitrogen mustard, Melphalan, instead of the exFISH linker, Label-IT. Melphalan is an alkylating agent with a primary amine available for conjugation to NHS esters, but it is widely available through major chemical vendors, 50 times less expensive than Label-IT, has two alkylating moieties per molecule, and could be incorporated into the gel matrix (Figure 1D). Importantly, RNA retention with Melphalan was comparable to Label-IT (Figure S1A), with significantly increased spot brightness (Figure 1E) and reduced background compared with Label-IT (Figure 1F), improving the signal-to-noise ratio by 30% (Figure 1G). We also improved tissue clearing and reagent penetration through 300-μm-thick tissue volumes when compared to the original exFISH protocol (Chen et al., 2016) (STAR Methods; Figure S1B).

For amplification of FISH signal, we used HCR v3.0 (Choi et al., 2018) because the probe and amplification oligos are short (50–100 nt) and can therefore rapidly penetrate thick tissue, while also reducing nonspecific spots. In contrast, another FISH signal amplification method, RNAscope (Wang et al., 2012), did not show sufficient reagent penetration in thick tissues (Figures S1C and S1D). We optimized hybridization conditions to improve detection specificity (STAR Methods; Figure S1E). Image acquisition was performed using selective plane illumination microscopy (SPIM, Zeiss Z.1 microscope), readily achieving single transcript sensitivity following HCR amplification.

To minimize mobile spots due to light-induced fragmentation of HCR products, we used low laser intensity or anti-fade compounds (STAR Methods; Video S1; Table S2; Figures S1F and S1G). To improve signal photostability, we custom-labeled the

corresponding HCR amplification hairpins with the photostable far-red dye Janelia Fluor 669 (JF669) (Grimm et al., 2017) (Figures 1H and 1I)

Three transcript species were detected in each imaging round, compatible with the spectral capabilities of standard fluorescence microscopes. For multiple rounds of FISH, we adapted a stripping and reprobing strategy that uses DNase I to remove probes and HCR amplification product from each previous round (Figure S1H) (Lubeck et al., 2014). For multi-round registration and cell segmentation, we used cytosolic RNA stained with DAPI (4′,6-diamidino-2-phenylindole) (Xu et al., 2020), which we refer to as cytoDAPI. Although DAPI is primarily used as a DNA stain, after DNase treatment, it provides a good near-UV cytosolic stain that is abolished by RNase treatment (Figure S1I). With these improvements, EASI-FISH enables robust, high quality and multiplexed FISH imaging of thick tissue volumes (∼1 mm × 1 mm × 0.3 mm in pre-expansion dimensions) (Figure 1J).

EASI-FISH data processing

Multi-round, high resolution imaging of thick tissue specimens for EASI-FISH produces multi-terabyte (TB) images. Therefore, we built computational image processing tools to handle these large datasets in a consistent and efficient manner (Figures 2A and 2B).

Stitching

For EASI-FISH in expanded thick tissue samples, multiple sub-volumes (tiles) were sequentially acquired, followed by computational stitching into a single large image. We used an Apache Spark–based high-performance stitching pipeline (Gao et al., 2019), which includes a flat-field correction step, followed by globally optimal translation for each tile (Figure S2A).

Round-to-round registration

Next, image volumes from each round of FISH were aligned using cytosolic contours of cytoDAPI using an automated nonrigid 3D registration pipeline (Figure S2B). The pipeline is highly accurate, with 99% ± 0.8% structural similarity between fixed and nine moving image volumes (STAR Methods; Figure S2C), and it is more than 10-times faster compared to other deformable registration methods (e.g., ANTs) (Yushkevich et al., 2016).

Cell segmentation

CytoDAPI provided a cytosolic signal for generating cell segmentation masks. The high accuracy of the registration pipeline allowed us to apply cell segmentation masks from a single round of imaging to all other rounds, which simplified analysis and reduced computation time.

Accurate segmentation of *in situ*-stained volumetric (3D) fluorescence image data has been a long-standing challenge that can degrade the accuracy of multiplexed FISH analysis. We developed a deep learning-based automated 3D segmentation algorithm, called *Starfinity* (Figure 2C). *Starfinity* is an extension of *StarDist*, a nuclear detection method (Schmidt et al., 2018; Weigert et al., 2020). In contrast to *StarDist*, *Starfinity* aggregates the dense prediction of cell border distances into pixel affinities

(N) 24 marker-gene expression in cells from 3 ROIs in (M). Scale bar: 10 μm.
(O) Spot counts for *Meis2* from EASI-FISH round 1 and round 9.
(P) Correlation between EASI-FISH spot counts and scRNA-seq UMIs for 24 marker-genes from the LHA.
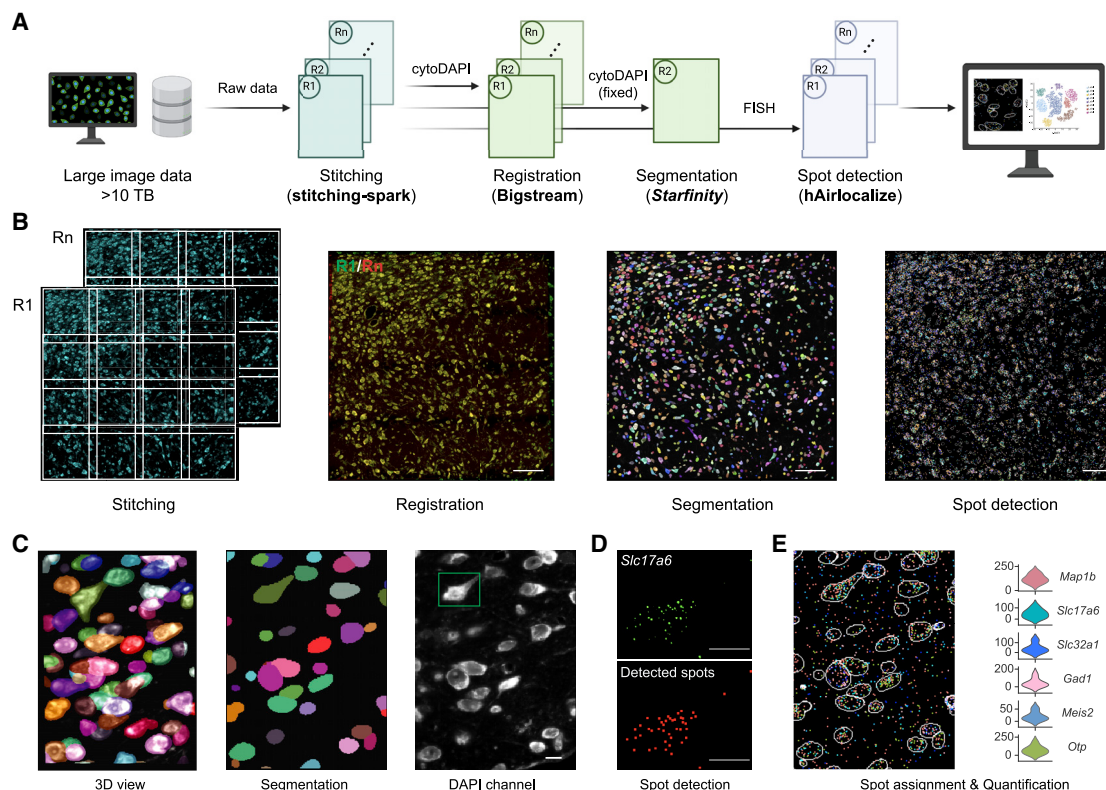
Figure 2. EASI-FISH analysis pipeline
(A) EASI-FISH data processing workflow.
(B) Representative images showing stitching, registration, segmentation, and spot detection in large image volumes. Scale bar: 100 μm.
(C) Example of *Starfinity* segmentation from cytoplasmic DAPI.
(D) Representative hAirlocalize-enabled spot detection in cell highlighted with green square in (C). Scale bar: 10 μm.
(E) Spot counts per neuron.

without imposing convex cell boundary constraints (STAR Methods). After generating appropriate training data, a *Starfinity* model was trained to predict cell body shapes from cytoDAPI images, outperforming several other segmentation methods (Figure S2D; Table S3). We manually inspected ~5% of automatically segmented cells from 4 samples (~4,000 out of 80,000 cells) and found that 93% of cells were properly segmented, 4% of cells over-segmented, 1% of cells undersegmented, and 2% of cells were contaminated by neighboring cells (Figure S2E). Because most oversegmented cells (62%) can be identified and rapidly corrected post hoc by quantitative and semi-automated criteria (STAR Methods), the final estimated segmentation accuracy was 95.5%.

Spot detection

We adapted Airlocalize (Lionnet et al., 2011) for spot detection (Figure 2D). To rapidly process more than 10 TB of image data, we developed hAirlocalize (high throughput spot detection based on Airlocalize) to parallelize the spot detection process. For cells with very high gene expression, where single spots cannot be resolved even with 2× expansion, we measured the total intensity per cell and converted the intensity to spot counts based on measured well-isolated single spot intensities for these genes (STAR Methods; Figures S2F and S2G).

We combined these computational modules into a self-contained, platform-agnostic computational pipeline for end-to-end EASI-FISH data analysis. The EASI-FISH pipeline can rapidly process large datasets greater than 10 TB in size with minimal manual intervention (STAR Methods).

EASI-FISH is sensitive and stable

Using this analysis pipeline, we evaluated performance of the EASI-FISH procedure. RNA detection efficiency with 20-probe sets was 92% ± 2% for *Gad1* based on targeting single transcripts with two colors of interleaved 10-probe sets (STAR Methods; Figure S1J). To determine the sensitivity of EASI-FISH, we analyzed genes from LHA with low expression levels according to scRNA-seq. Low-expressed genes, *Klhl13* (RNA-seq UMI_{mean} = 48) and *Igf1* (UMI_{mean} = 15), were co-expressed in all melanin-concentrating hormone (*Pmch*)-expressing neurons in the LHA, so we used the fraction of $Pmch^+$ neurons in which we cannot detect these genes as an estimate of false negative rate at the cell level. Among the $Pmch^+$ neurons that were analyzed, 34/34 (100%) expressed *Klhl13* and 38/41 (93%) expressed *Igf1* with an average background-subtracted spot count per cell of 195 for *Klhl13* and 41 for *Igf1* (Figure S1K), indicating a low dropout rate with EASI-FISH.
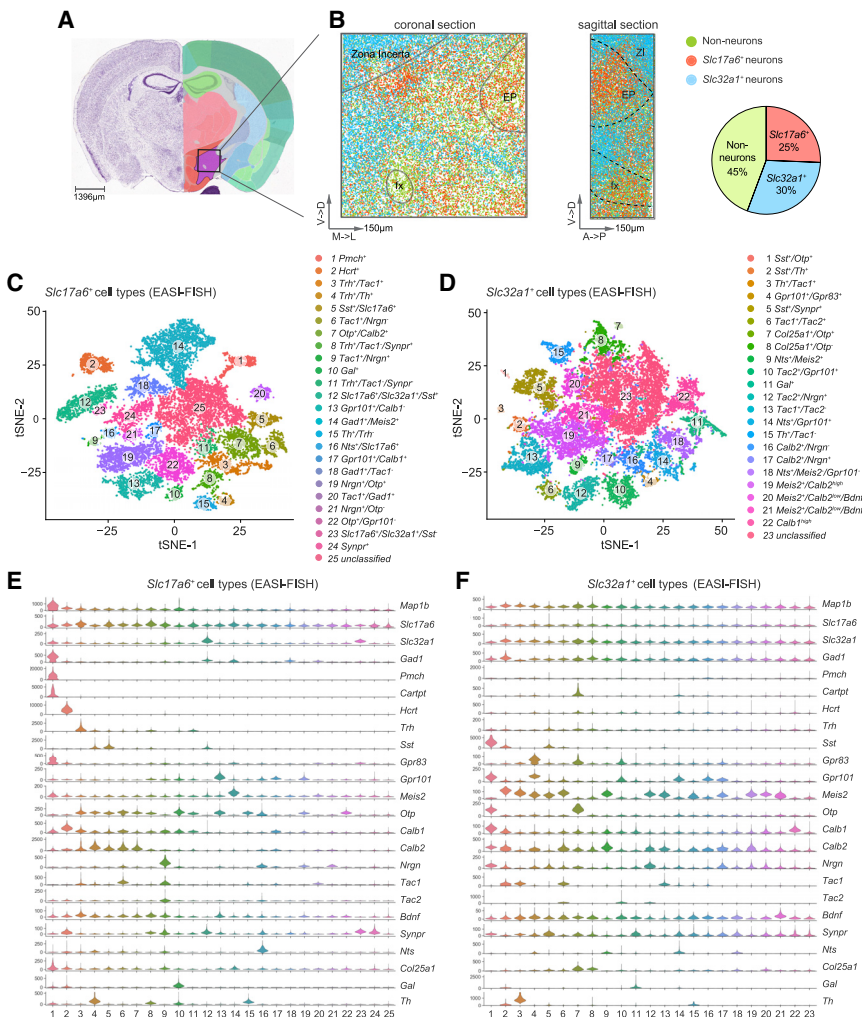
Figure 3. Profiling LHA molecular markers with EASI-FISH

(A) Location of imaged region from LHA mapped to the coronal mouse brain atlas. Image credit: Allen Institute.

(B) Spatial organization and proportion of excitatory and inhibitory neurons and non-neurons in the imaged LHA region (total: 66,488 cells from 3 mice). Scale arrows: 150 μm. EP: entopeduncular nucleus, fx: fornix, ZI: zona incerta.

(C and D) t-distributed stochastic neighbor embedding (tSNE) plot for (C) excitatory and (D) inhibitory neurons in the LHA, with cell types color-coded by EASI-FISH cluster.

(E and F) Expression (spot counts) of 24 FISH marker-genes in the (E) excitatory and (F) inhibitory EASI-FISH clusters, shown by violin plots.

well correlated with scRNA-seq data (r = 0.96, p = 0.0081, based on measurement of *Klhl13*, *Igf1*, *Pdyn* and *Tacr3*). High RNA retention was observed after re-probing the same targets after at least 7 rounds (Figure S1O), even with an elapsed time of more than 40 days (93.5%, n = 4 genes). Taken together, EASI-FISH showed high reproducibility between samples, minimal RNA loss across hybridization rounds, good correlation with scRNA-seq data, high sensitivity, and a low dropout rate.

Application of EASI-FISH for profiling LHA marker-genes

We applied the EASI-FISH sample processing and data analysis pipeline for in-depth examination of the molecularly defined cell types in the LHA. We performed scRNA-seq on manually picked LHA neurons and combined this data with published LHA scRNA-seq data collected with droplet-based methods (Mickelsen et al., 2019; Rossi et al., 2019) to determine consensus cell clusters across three datasets (Figure S3A). Clustering analysis of the combined data identified 17 glutamatergic *Slc17a6* (encoding *Vglut2*)-expressing clusters (labeled e1-e17) and 17 GABAergic *Slc32a1* (encoding *Vgat*)-expressing clusters (labeled i1-i17) (Figures S3B–S3D). Each cluster included cells from at least 2 of the 3 datasets (Figure S3E). A combinatorial set of 24 marker-genes was selected for subsequent EASI-FISH experiment to map major cell types (STAR Methods; Figures S3F and S3G; Table S5; and Data S1).

We used EASI-FISH to profile these marker-genes in tissue volumes (~1 mm × 1 mm × 0.3 mm) taken from the tuberal LHA (Figure 3A). The tuberal LHA is associated with eating, drinking, and arousal; and it corresponds to the area used for the scRNA-seq tissue samples. Nine rounds of 3-plex FISH (24 unique genes) were performed on three samples, including repeat-rounds for validation at the end of the data collection sequence to assess sample stability (Figures 1M and 1N), which

We examined two sources of false-positive spot assignments: nonspecific probe binding and amplification and true-positive transcript detections that may not originate within the segmented cell body. Because we used HCR v3.0, nonspecific probe binding and amplification was low (1 per 3000 μm³), which we determined by applying FISH probes to samples in the absence of the target gene (*GFP* probes in wild-type mouse tissue; Figure S1E). We estimated background (false-positive) spot detection rate using two genes with mutually exclusive expression patterns based on scRNA-seq (*Pdyn* and *Tacr3*) (1 per 50 μm³, ~30 spots/cell) (STAR Methods; Figure S1L). We observed a similarly low false-positive background detection rate for mutually exclusive genes for *Slc17a7* (*Vglut1*) and *Gad1* in cortical neurons (Figures 1K and 1L). These background spots, detected in cells in which we confirmed correct cell boundary segmentation, are likely true RNA transcripts detected *en passant* neuronal processes that are adjacent to the cell body or a result of light-induced mobile spots fragmented from positive cells.

Spot count measurements were highly reproducible across multiple rounds (Figure S1M), between replicates of EASI-FISH experiments from different animals (Figure S1N) and were also
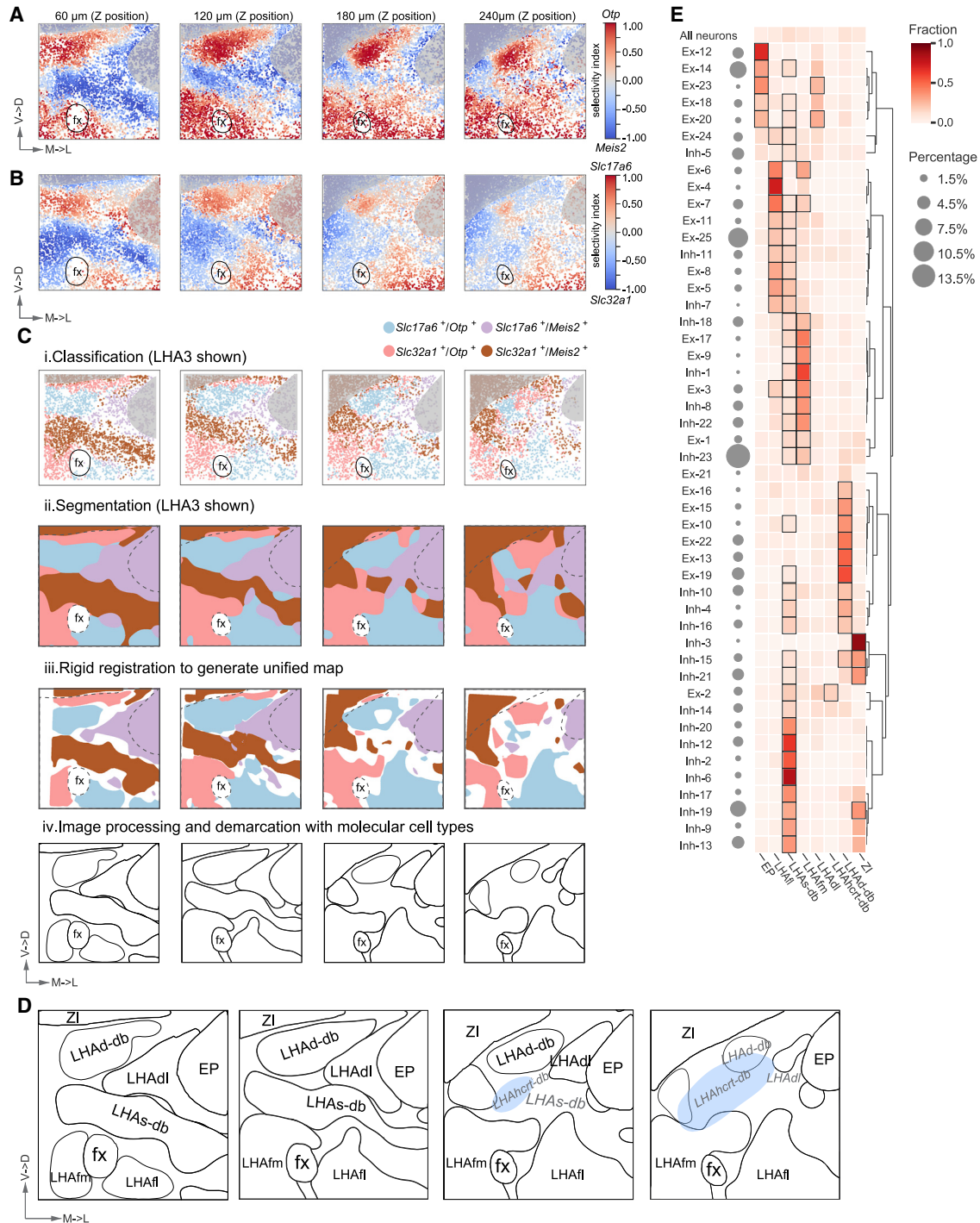
Figure 4. Spatial reconstruction of LHA reveals spatio-molecularly defined subregions

(A and B) Selective neighborhood enrichment of (A) *Otp/Meis2* and (B) *Slc17a6/Slc32a1*. Each dot indicates the centroid position of a neuron. Panels are 60 μm axial projections of sub-volumes. From left to right: anterior to posterior.

(C) Automated workflow for 3D molecular parcellation. C-i: Classify regions based on their relative enrichment for *Otp*, *Meis2*, *Slc17a6* and *Slc32a1*. C-ii: 3D segmentation with Gaussian Mixture Models. C-iii: Consensus parcellation map after rigid registration across animals. C-iv: Post-processing of parcellation map. Neighboring brain regions (ZI and EP) are shaded in gray in Figures 4A and 4B and C-i and highlighted with dotted line in C-ii and C-iii.

(D) Spatio-molecular parcellation of the LHA, including an additional *Hcrt* subregion (light blue). From left to right: anterior to posterior. ZI: zona incerta; EP: entopeduncular nucleus; fx: fornix; LHAd-db: LHA dorsal diagonal band; LHAdl: LHA dorsal lateral region; LHAs-db: LHA suprafornical diagonal band; LHAfm:

(legend continued on next page)

showed excellent RNA retention (90%, n = 2 genes, 3 mice) between round 1 and round 9 (Figures 1O and S1P). There was high correlation with scRNA-seq UMI counts (r = 0.86, p = 8.4 × 10^{-8}), and the EASI-FISH pipeline detected an average of 13-fold (±1.6) more molecules per cell compared to scRNA-seq UMI counts (Figure 1P). In these large LHA samples, false-positive detection rate was low, as measured by spot counts from genes that showed orthogonal expression patterns (Figure S1Q).

A total of ~86,000 ROIs were identified from three specimens (Table S6), with 36,423 intact neurons (*Map1b*$^+$) detected in all imaging rounds (Figure 3B). We used a *de novo* approach to identify cell types in the LHA based on 24-plex marker-gene expression. We grouped neurons into *Slc17a6*-expressing (45%, 16,394 cells) and *Slc32a1*-expressing (55%, 20,029 cells) (Figures 3C and 3D). Among these neurons, 79% could be classified based on differential expression of marker-genes, while 7% of *Slc17a6*$^+$ neurons (n = 2787) and 14% of *Slc32a1*$^+$ neurons (n = 5034) showed low or no expression of marker-genes other than *Map1b*, *Slc17a6*, *Slc32a1*, or *Gad1*, and were grouped as unclassified clusters from each type (Ex-25 and Inh-23). Clustering *Slc17a6*$^+$ neurons by marker-genes separated them into 24 molecularly defined clusters (Figures 3E, S4A, and S4C; Data S1) and the *Slc32a1*-expressing population was clustered into 22 molecularly defined neuronal subtypes (Figures 3F, S4B, and S4D; Data S1). These molecularly defined clusters were detected in all three animals (Figures S4A and S4B), except Inh-1 (STAR Methods). Among the molecularly defined clusters, Inh-3, Inh-15, and Inh-21 were dominated by cells from the zona incerta (ZI); Ex-12 and Ex-23 were enriched for cells from the entopeduncular nucleus (EP) (Data S1).

Strong correlations were found between many scRNA-seq and EASI-FISH clusters (Figures S4F and S4G). scRNaseq clusters that did not map well are either rare or not present in the region of the LHA that we profiled with EASI-FISH, based on marker-gene distribution from Allen Brain ISH Data (STAR Methods).

Spatial reconstruction of LHA reveals molecularly defined subregions

The LHA is one of the largest and most intensively studied regions in the hypothalamus. However, previous studies do not demarcate subregions of the lateral hypothalamus in mouse (Franklin and Paxinos, 1997; Lein et al., 2007), and, in the rat, limited parcellation has been proposed by combining cytoarchitectural and connectivity information (Geeraedts et al., 1990; Hahn et al., 2019; Hahn and Swanson, 2010, 2015; Veening et al., 1987).

Computational LHA parcellation

Because the EASI-FISH pipeline provides detailed molecular information along with location of every cell, we pursued a machine learning approach to identify LHA subregion boundaries by combining molecular, spatial, and cell density information in

these datasets. We prioritized reproducibility by defining consensus regions that can be automatically aligned between samples from multiple mice.

To examine the spatio-molecular organization of the LHA with a limited number of marker-genes, we leveraged the hierarchy of the cell type gene expression profiles, selecting neurotransmitter transporters for excitatory and inhibitory neurons (*Slc17a6* and *Slc32a1*) and transcription factors (*Otp* and *Meis2*), which have important developmental and cell specification functions in the LHA (Romanov et al., 2020). Principal component analysis of the spatial distribution of all 24 marker-genes also identified these four genes as having the largest magnitude weights and revealed distinct 3D spatial patterns in the LHA (Figures S5A–S5D). The expression of these genes showed a central *Meis2*-expressing wedge within the LHA that bisected *Otp*-enriched areas near the fornix and the ZI (Figuress 4A and S5E), which were further subdivided by *Slc17a6* and *Slc32a1* (Figures 4B and S5F).

To execute an unbiased parcellation of the LHA, we developed a computational approach with the following steps: (1) automated volumetric segmentation based on spatial distribution of cells co-expressing combinations of *Otp* or *Meis2*, and *Slc17a6*, or *Slc32a1*; (2) rigid registration of the segmented volumes to align samples from biological replicates; and (3) generation of a consensus parcellation across multiple samples (Figure 4C). All steps were performed using the entire image volume, which we present using axial projections of 4 subvolumes (Figures 4A–4C).

First, for tissue volumes from each animal, we plotted the spatial distribution of cells selectively expressing *Otp* or *Meis2*, and *Slc17a6* or *Slc32a1*; and applied spatial density-based smoothing (example images in Figure 4A and 4B). We then classified each neuron into broad cell classes (*Otp/Slc17a6, Otp/Slc32a1, Meis2/Slc17a6, Meis2/Slc32a1*) based on the relative spatial enrichment of the two pairs of genes (example images in Figure 4C-i) and used Gaussian mixture models for 3D segmentation of the imaged tissue volumes (example images in Figure 4C-ii). To evaluate the generality of this strategy, we also applied our analysis to a spatial transcriptomic dataset from the visual cortex (Wang et al., 2018) and recovered the expected cortical lamination (Figures S5G–S5L).

We assessed generalization of the LHA parcellation across animals by performing rigid alignment on the segmented tissue volumes. The degree of alignment was quantified on fiducial landmarks (ZI, EP and the fornix) (average intersection over union between LHA1 and LHA3: 0.71; between LHA2 and LHA3: 0.76) as well as spatial distribution of marker-genes (Figure S5M). To estimate the consensus anatomical parcellation, we performed Simultaneous Truth and Performance Level Estimation (STAPLE) (Warfield et al., 2004), which eliminated discordance at boundaries of the segmented subregions (Figure 4C-iii).

LHA medial fornical region; LHAfl: LHA lateral fornical region. Solid lines: parcellation based on *Otp*, *Meis2*, *Slc17a6*, and *Slc32a1* expression. Gray/italic: transition zones. Scale arrows in Figures 4A–4D: 150 μm.

(E) Excitatory and inhibitory EASI-FISH cluster enrichment in parcellated subregions, as compared to distribution from all neurons (top, STAR Methods). Gray outlines: statistically significant (permutation test, p < 0.05, effect size: Cohen's h > 0.2) enrichment of cell types. Gray circles: percent of neurons in each molecularly defined cell type. Color bar: fraction of neurons from selected cell type in each subregion. Statistics: Table S1.

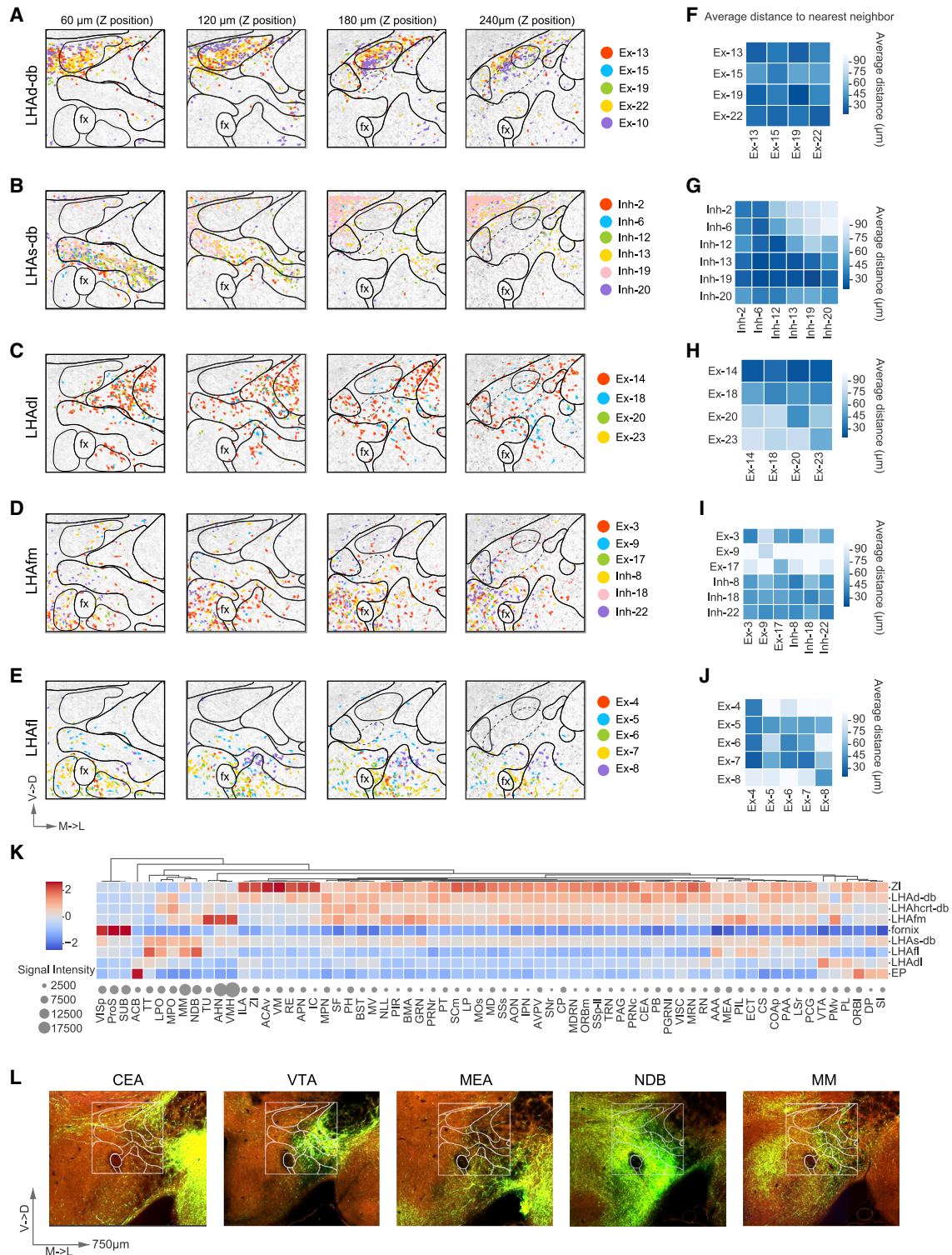**Figure 5. Molecularly defined cell types and afferent axons are enriched in LHA subregions**

(A–E) Molecularly defined clusters enriched in (A) LHAd-db, (B) LHAs-db, (C) LHAdl, (D) LHAfm, (E) LHAfl subregions. Panels are 60 μm axial projections of sub-volumes. From left to right: anterior to posterior. Cytosolic DAPI stain with all cells in light gray. Scale arrows: 150 μm.

(legend continued on next page)

Computational parcellation based on *Otp*, *Meis2*, *Slc17a6*, and *Slc32a1* defined 5 zones in the LHA (Figures 4C-iv and 4D), most of which have not been described previously. Two prominent bands run diagonally at an approximately 60-degree angle from each other. The dorsal diagonal band (LHAd-db) is below the ZI and is enriched for excitatory neurons. A suprafornical diagonal band (LHAs-db) is dorsal and lateral to the fornix and is enriched in inhibitory neurons. The diagonal bands surround the wedge-like *Meis2*-enriched excitatory subregion in the dorsal lateral region (LHAdl), which is flanked laterally by the EP. The medial fornical area (LHAfm) is modestly enriched in inhibitory neurons. The subregion lateral to the fornix (LHAfl) is enriched for excitatory neurons. In the posterior portion, the LHAd-db and LHAdl become less defined, leaving this portion of the volume not well demarcated by the above 4 genes. Instead, this region is enriched for hypocretin (*Hcrt*⁺) neurons (Data S1, Ex-2), an important neuropeptide secreting population, which is enriched in a diagonal band running in the same direction as the LHAd-db. We defined this sixth subregion as the hypocretin neuron-enriched diagonal band (LHAhcrt-db).

The subregions in the LHA that we identified by *Otp*, *Meis2*, *Slc17a6*, and *Slc32a1* co-expression coarsely tracked with neuronal density in this region (Figure S5N). For example, a dense group of neurons running from the fornix to the ZI was noted previously in the rat brain and named the suprafornical LHA (Hahn et al., 2019; Hahn and Swanson, 2010, 2015). However, we discovered that consideration of molecular identity revealed a more intricate structural organization that was subdivided between the LHAfm, LHAs-db, and LHAd-db.

Neuronal cell types in LHA subregions

Next, we considered the relationship of molecularly defined cell types to these LHA subregions. Although molecularly defined cell types were intermingled (mean neighborhood composition in a 50-μm-radius: 16 cell types), they were not randomly distributed (Complete Spatial Randomness testing, p value < 0.05, Table S1). Most cell types were spatially enriched in one or more of the LHA subregions (Figures 4E, 5A, 5B, 5C, 5D, 5E and S6A), with good correlation between animals (Figure S6B). Additional differential spatial enrichment of molecularly defined cell types was observed within LHAfl and LHAs-db, subdividing the LHA into nine subregions (dotted lines in Figure S6C and quantification in Figures S6D and S6E).

After clustering the molecularly defined cell types based on their spatial overlap (STAR Methods), we found groups of molecularly defined cell types enriched in regions corresponding to LHAs-db, LHAd-db, LHAfl and LHAfm (Figures S6F and S6G). Taken together, the 3D-molecular organization of LHA subregions can be determined with either a limited set of marker-genes for broad cell classes or a larger set of marker-genes for individual cell types.

LHAd-db

The LHAd-db contains a mixture of mainly excitatory and a few inhibitory cell types. Excitatory cell types Ex-13 (*Gpr101/Calb1*⁻),

Ex-15 (*Th/Trh*⁻), Ex-19 (*Nrgn/Otp*), and Ex-22 (*Otp/Gpr101*⁻) are primarily localized to LHAd-db (Figure 5A). Ex-10 (*Gal/Slc17a6*) is also enriched in this subregion, but unlike the other cell types, Ex-10 forms a band around the LHAd-db (Figure 5A). In addition, a variety of inhibitory cell types are present in LHAd-db that are also distributed in the adjacent ZI and LHAs-db (Figure 4E). Some broadly distributed inhibitory cell types are also observed in LHAd-db, such as Inh-4 (*Gpr101/Gpr83*), Inh-10 (*Tac2/Gpr101*), and Inh-16 (*Calb2/Nrgn*⁻) (Figure 4E).

LHAs-db

Most LHAs-db cell types are inhibitory, with small clusters Inh-2 (*Sst/Th*), Inh-6 (*Tac2/Tac1*), Inh-12 (*Tac2/Nrgn*), and Inh-20 (*Meis2/Calb2^{low}/Bdnf*⁻) almost exclusively localized to this subregion (Figure 5B). Several LHAs-db clusters are also found in the ZI, such as Inh-9 (*Nts/Meis2*), Inh-13 (*Tac1/Tac2*⁻), Inh-17 (*Calb2/Nrgn*), and Inh-19 (*Meis2/Calb2^{high}*) (Figures 4E and 5B). There were few excitatory cell types in LHAs-db. Spatial positioning of cell types within the LHAs-db revealed additional spatial segregation. Inh-2, 6, 9, 17, and 19 are enriched in the medial part of the LHAs-db; and Inh-12 is primarily concentrated in the lateral part of the LHAs-db (Figures S6C–S6D), while Inh-13 and Inh-20 are more evenly distributed.

LHAdl

LHAdl is a relatively cell-sparse zone (Figure S5N), which is largely comprised of excitatory cell types similar to those in the adjacent EP: Ex-14 (*Gad1/Meis2*), Ex-18 (*Gad1/Tac1*⁻), Ex-20 (*Tac1/Gad1*), and Ex-23 (*Slc17a6/Slc32a1/Sst*⁻) (Figures 4E and 5C).

LHAfm

A mix of excitatory (Ex-3 (*Trh/Tac1*), Ex-9 (*Tac1/Nrgn*), and Ex-17 (*Gpr101/Calb1*)); and inhibitory (Inh-1 (*Sst/Otp*), Inh-8 (*Col25a1/Otp*⁻), Inh-18 (*Nts/Meis2*⁻/*Gpr101*⁻), and Inh-22 (*Calb1^{high}*)) cell types are localized in the LHAfm, with a higher fraction of inhibitory neurons (Figures 4E and 5D). Because the LHAfm is at the border of the dissection region for scRNA-seq analyses, there appears to be a larger number of cells lacking specific marker-genes (i.e., Ex-25 and Inh-23). Some of the LHAfm cell types are also shared with the adjacent medial-ventral LHAfl, such as Ex-6 (*Tac1/Nrgn*) and 7 (*Otp/Calb2*) (Figure 5D and 5E).

LHAfl

LHAfl is primarily comprised of excitatory cell types with Ex-4, Ex-8, Ex-11 (*Trh*-expressing cell types) and Ex-5 (*Sst/Slc17a6*), Ex-6 (*Tac1/Nrgn*⁻), Ex-7 (*Otp/Calb2*) (Figure 4E and 5E). Spatial positioning of cell types within LHAfl revealed additional subdomains (Figures S6C and S6E). Ex-4, Ex-6, Ex-7 are located medial-ventrally (LHAfl-mv) and Ex-11 dorsal-laterally (LHAfl-dl), with additional poorly classified excitatory neurons (Ex-25) in the ventral lateral subdomain (LHAfl-vl) (Figure S6E). In addition, there is a sparse distribution of inhibitory cell types, such as Inh-11 (*Gal*). Another two inhibitory cell types (Inh-8 and Inh-22) that were primarily localized in the LHAfm were also in LHAfl.

Molecularly defined cell types enriched in the same subregions are typically spatially intermixed, based on average

---

(F–J) Average distance to nearest neighbor (ANN) among molecularly defined cell types enriched in LHAd-db, LHAs-db, LHAdl, LHAfm and LHAfl, as shown in (A–E). Rows: cell type used to find nearest neighbor; columns: corresponding cell type used to compute ANN (STAR Methods).

(K) Quantification of axons projecting to or passing through the LHA subregions based on data from the Allen Brain Atlas Connectivity database.

(L) Representative images showing differential projection patterns in the LHA from the central amygdala (CEA), ventral tegmental area (VTA), medial amygdala (MEA), diagonal band nucleus (NDB) and medial mammillary nucleus (MM).
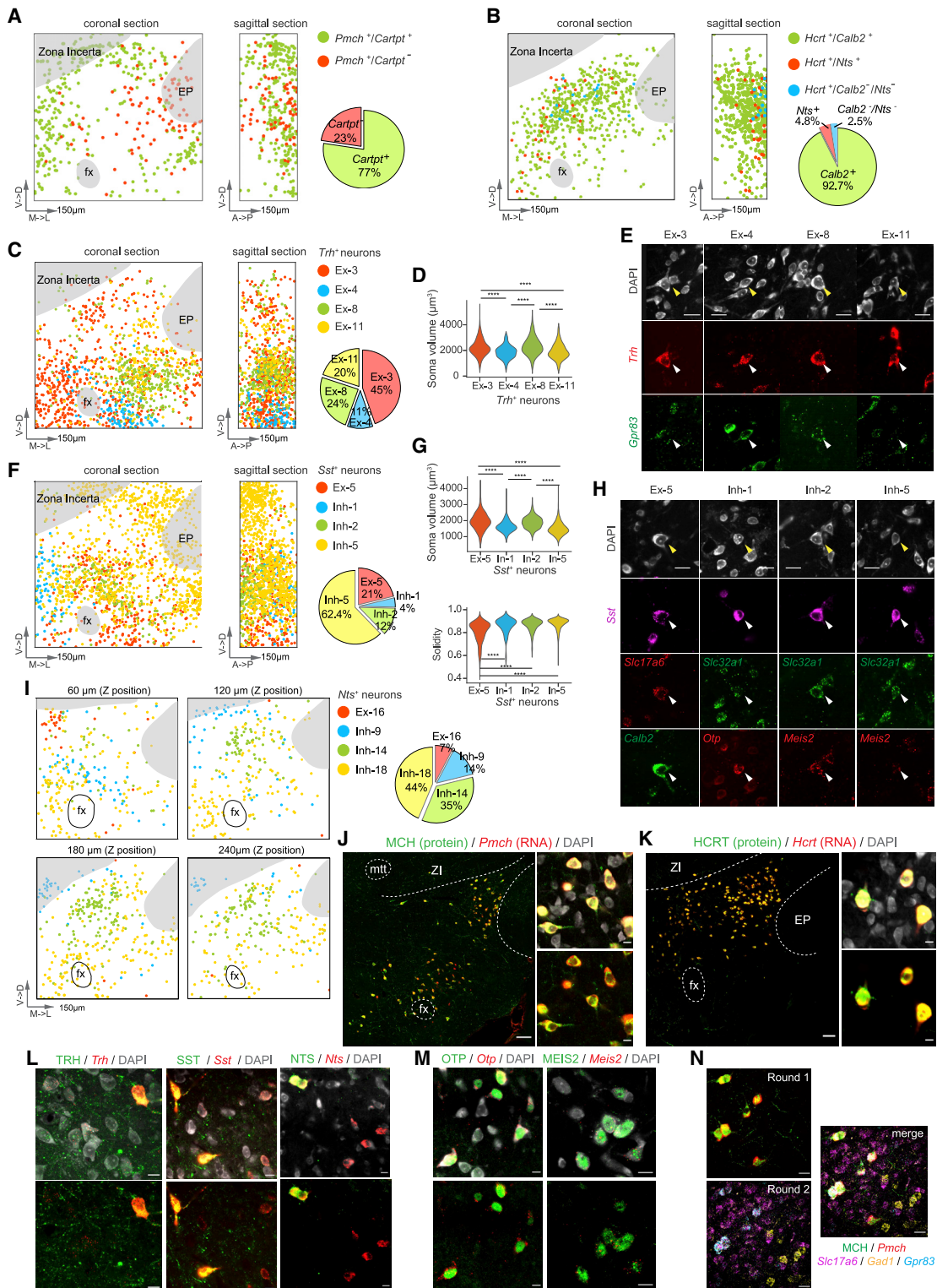
Figure 6. Molecular and spatial organization of major neuropeptide neurons in the LHA

(A) $Pmch^+$ neuron subpopulations. Max-projected coronal and sagittal sections with $Cartpt^+$ in green and $Cartpt^-$ in red.

(B) $Hcrt^+$ neuron subpopulations are spatially intermixed.

(C) Trh^+ neuron cell types. Left: Spatial distribution of Trh^+ clusters. Right: Percent of Trh-expressing neurons in each cluster.

(D) Trh^+ clusters soma volumes. ****$p < 0.0001$. Statistics: Table S1.

(legend continued on next page)

distance to nearest neighbor (ANN) analysis (Figures 5F–5J). But segregation of some cell types within these subregions is also evident, which supports the additional subdivisions of LHAs-db and LHAfl noted above (Figures S6C, 5G, and 5J).

Transcriptional relatedness of spatially clustered cell types

For molecularly defined cell types enriched in the same subregion, some corresponded to the same scRNA-seq clusters. For example, in the LHAd-db, Ex-13, and Ex-19 both have high correlation with seq-e9 (*Otp/Gpr101*) from scRNA-seq (Figures S4F and S4G), revealing additional heterogeneity within scRNA-seq clusters. However, instances of intermingled excitatory and inhibitory cell types were common in all regions, such as Ex-3 and Inh-8 in the LHAfm and Ex-13, Ex-19, and Inh-4 in the LHAd-db (Figure 4E).

LHA subregions robust to marker-gene selection

Many cell types showed restricted distribution in the LHA, but most individual marker-genes (except *Hcrt*) were not restricted to a single subregion (Data S1; Figure S6H). Thus, co-expression relationships between multiple marker-genes are essential for revealing the underlying spatial organization of molecularly defined cell types. To determine how well combinatorial gene expression could predict neuronal spatial positions in the LHA, we trained a random-forest regression model. Combinatorial expression of 24 marker-genes explained $60 \pm 2\%$ of spatial variation. Excluding any single marker-gene or *Otp, Meis2, Slc17a6*, and *Slc32a1* in combination did not substantially decrease the prediction accuracy, indicating that the model was not dominated by individual genes (Figures S6I and S6J).

To evaluate the platform independence of these subregions, we analyzed a brain slice from the open-source (Vizgen, 2021) MERFISH mouse brain receptor map dataset, consisting of 483 genes with neuron and glia markers. Principal component analysis of spatial gene expression patterns was consistent with the subregions described above (Figures S6K and S6L).

Axonal projections in LHA subregions

We analyzed the spatial distribution of afferent axons projecting to or passing through the LHA by mapping this detailed LHA parcellation onto the Common Coordinate Framework from the Allen Mouse Connectivity Atlas (Oh et al., 2014). Most axonal projections were not broadly distributed across the LHA. Selectivity was greatest between the LHAdl and LHAfl regions, which had largely mutually exclusive input patterns. Some axonal projections mapped onto specific LHA subregions (Figure 5K). For example, CEA projections were enriched in the LHAd-db, VTA projections were enriched in LHAdl, and MM projections were enriched in LHAfl and LHAs-db (Figure 5L).

Neuropeptide cell types in the LHA

We identified subgroupings of several neuropeptide-expressing cell types that were spatially distinct and were associated with morphological differences.

Pmch⁺ neurons and Hcrt⁺ neurons

Pmch⁺ and *Hcrt*⁺ neurons are involved in feeding and arousal. Within the tissue volume that we analyzed, 83% of *Pmch*⁺ neurons were in the LHA and 17% in the ZI. *Pmch*⁺ neurons could be subdivided into two populations based on *Cartpt* expression that encodes a neuropeptide that regulates energy homeostasis (*Cartpt*⁺: 77%, 388/501 neurons; *Cartpt*⁻: 22%,113/501), largely consistent with scRNA-seq data (Mickelsen et al., 2019). Nearly all *Pmch*⁺ neurons in the ZI were *Cartpt*⁺ (99%). Within the LHA, we observed distinct spatial distributions of the two *Pmch*⁺ subpopulations, with the *Pmch/Cartpt*⁻ neurons enriched in the LHAdl, while the *Pmch/Cartpt*⁺ population segregated into a medial population and a ventral lateral population (Figure 6A). More than 92% of *Pmch*⁺ neurons analyzed co-expressed *Gad1* and *Slc17a6*, consistent with scRNA-seq data and previous report (van den Pol et al., 2004). Most *Pmch*⁺ neurons (77%) co-expressed the obesity-related GPCR, *Gpr83*, and there was a 2-fold greater frequency of *Gpr83* co-expressing cells in the *Cartpt*⁺ population (*Pmch/Gpr83* co-expression: *Cartpt*⁺: 87%, *Cartpt*⁻: 43%).

Hcrt⁺ neurons were spatially restricted in the LHA samples we examined and were enriched in a dorsal diagonal band that was caudal to the LHAd-db. Our manually picked scRNA-seq dataset contained many *Hcrt*⁺ neurons, revealing two main subdivisions based on the expression of *Calb2* and *Nts*. The majority of *Hcrt*⁺ neurons (93%, 593/640) expressed *Calb2*, with a small population expressing *Nts* (5%, 31/640). The remainder of *Hcrt*⁺ neurons were negative for both markers (2%, 16/640), which were enriched in a more caudal position. Although there were some spatial differences in the distribution of *Hcrt*⁺ subtypes, they were largely intermingled (Figure 6B).

Trh⁺ neurons

The thyrotropin-releasing hormone (*Trh*)-expressing neurons in the LHA have been implicated in promoting arousal behaviors (Horjales-Araujo et al., 2014). EASI-FISH identified four prominent *Trh*⁺ cell types (Ex-3, Ex-4, Ex-8, and Ex-11), which were spatially and molecularly distinct (Figure 6C). Consistent with scRNA-seq data, 97.6% (1483/1519) of *Trh*⁺ neurons co-expressed *Otp* and were found in *Otp* zones (LHAfm and LHAfl). Compared to scRNA-seq data, EASI-FISH revealed additional molecular heterogeneity within *Trh*⁺ neurons that were spatially segregated. Ex-3 was spatially enriched in the LHAfm, had high expression of *Trh*, and co-expressed *Calb1* and *Tac1*. In contrast, Ex-4, Ex-8, and Ex-11 were positioned lateral to the

---

(E) Cell body morphology from *Trh*⁺ subtypes, highlighted by arrowheads.

(F) *Sst*⁺ neuron clusters. Left: Spatial distribution of *Sst*⁺ neuronal clusters. Right: Percent of *Sst*⁺ neurons in each cluster.

(G) Cell body volume (top) and solidity (bottom) in *Sst*⁺ clusters. ****p < 0.0001. Statistics: Table S1.

(H) Morphology of *Sst*⁺ subtypes, highlighted by arrowheads.

(I) *Nts*⁺ neuron clusters. Left: Spatial distribution of *Nts*⁺ neuronal clusters. From left to right: anterior to posterior. Right: Percent of *Nts*⁺ neurons in each cluster. Scale arrows: 150 μm.

(J–L) IF and FISH co-detection of five neuropeptides, MCH (J), HCRT (K), TRH, SST and NTS (L). Scale bars: 100μm (J and K) left panel, 10μm (J and K) right panels and (L).

(M) IF and FISH co-detection of OTP (left) and MEIS2 (right). Scale bar: 10μm.

(N) IF (MCH) and FISH (*Pmch*) co-detection in round 1, followed by IF signal removal and 3-plex FISH (round 2). Scale bar: 20 μm.
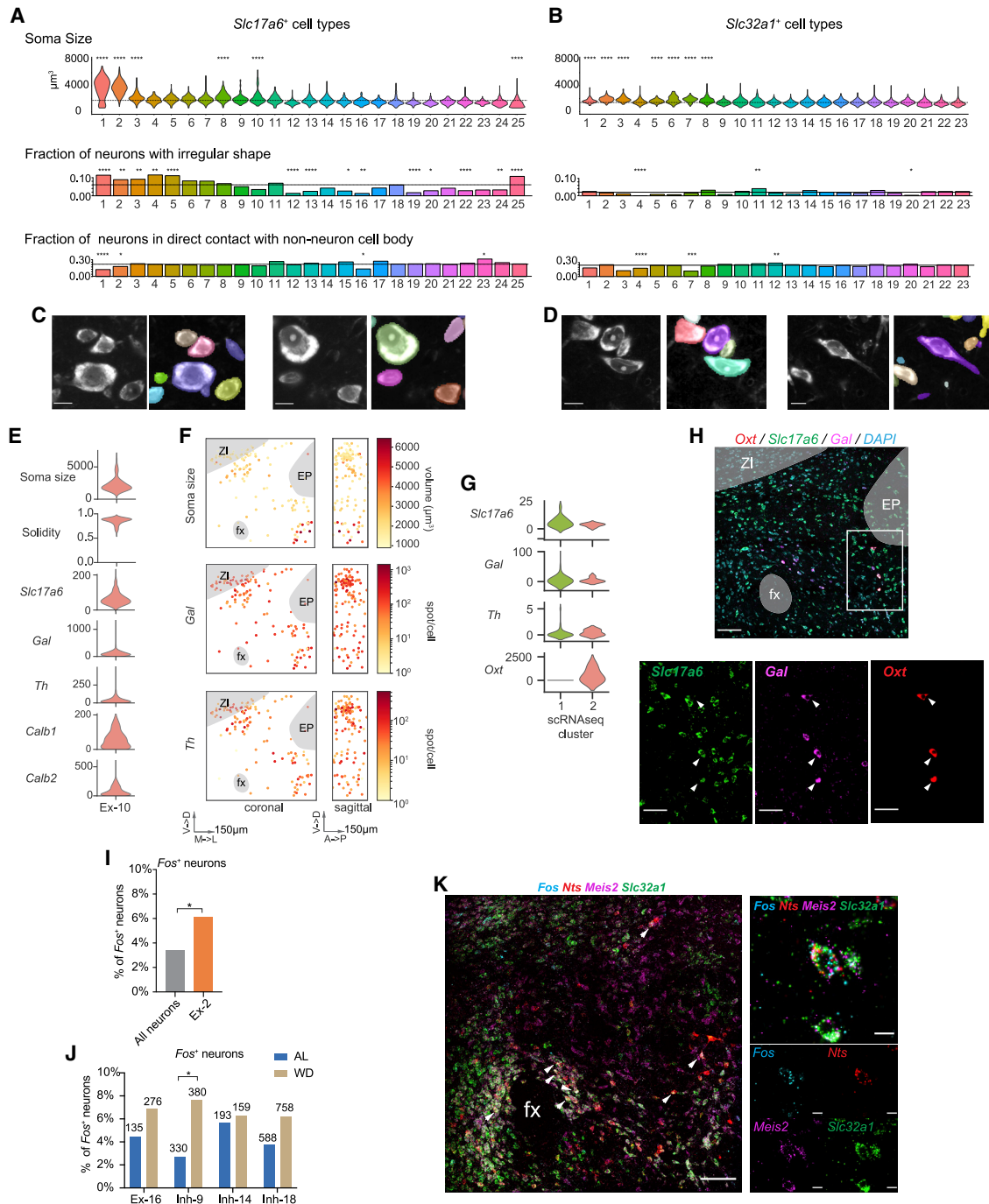
Figure 7. Morphological diversity and enrichment of *IEGs* in the LHA clusters

(A and B) Morphological characteristics in (A) excitatory and (B) inhibitory clusters as measured by soma volume (top), fraction of neurons with low solidity and eccentricity (middle) and fraction of neuron somata contacting a non-neuron cell body (bottom). Dotted lines: average measurements. *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001. Statistics: Table S1.

(C) Representative images showing soma size diversity and accurate segmentation.

(D) Representative images showing soma shape diversity. Scale bar in (C) and (D): 10 μm.

(E) Soma size, solidity, and marker-gene expression in Ex-10 cluster.

(F) Spatial distribution of soma size (top panel), *Galanin (Gal,* middle panel), and *Th* (bottom panel) in Ex-10 cluster. Ex-10 subpopulation enriched in the LHAfl subregion with large somata.

(legend continued on next page)

fornix in the LHAfl. Ex-4 was enriched in the LHAfl-mv subdomain and co-expressed *Th* and *Calb2*. Ex-8 and Ex-11 were intermingled in a separate position in the LHAfl-dl subdomain and were discriminated by expression of *Gpr83* and *Synpr*, respectively (Figure 3E). In addition to spatial differences, the four *Trh*+ cell types also showed different cell volumes, with larger cell bodies in Ex-3 and Ex-8 relative to Ex-4 and Ex-11 (Figures 6D and 6E).

Sst+ neurons

For *Sst*+ populations, we identified one excitatory (Ex-5) and three inhibitory (Inh-1, Inh-2, and Inh-5) cell types in the imaged LHA volume. These populations were spatially separated, with Ex-5 enriched in the LHAs-db and LHAfl, Inh-1 enriched in the LHAfm, and Inh-2 enriched in the LHAs-db. Inh-5 was dispersed across multiple LHA subregions (Figure 6F). Inh-1 had the highest *Sst* expression level (Figure 3F), and co-expressed *Gpr101* and *Nrgn*. The excitatory *Sst*+ cluster Ex-5 and inhibitory cluster Inh-2 had larger cell bodies than Inh-1 and Inh-5 (Figures 6G and 6H). In addition, many neurons in the Ex-5 cluster were less convex compared to the inhibitory *Sst*+ neuron clusters (Figures 6G and 6H). We also observed co-expression of *Sst* in a subset of the *Trh*+ cluster, Ex-4 and *Trh* in the *Sst*+ cluster, Ex-5. Ex-4 and Ex-5 differed in their *Th* expression, with Ex-5 negative for *Th*.

Nts+ neurons

Nts+ neurons in the LHA have been examined in multiple behaviors, including thirst (Brown et al., 2015; Kempadoo et al., 2013; Patterson et al., 2015). *Nts* was expressed in several transcriptionally distinct excitatory (Ex-16) and inhibitory (Inh-9, Inh-14, Inh-18) cell types. Ex-16, the single *Slc17a6*-expressing *Nts*+ population, was enriched in the dorsomedial region and was more anterior than the inhibitory populations (Figure 6I). Inh-9 was spatially enriched in the ZI and LHAs-db, and co-expressed *Meis2*. Inh-14 was enriched in a dorsal diagonal band that spatially overlapped with *Hcrt*+ neurons (33% overlap) and co-expressed *Gpr101* and *Galanin.* Inh-18 was spatially dispersed in the LHA.

Slc17a6/Slc32a1 co-expressing populations

Our image volume included glutamate and GABA co-releasing populations (Ex-12) in the EP (Wallace et al., 2017). 62% (742/1188) of *Slc17a6/Slc32a1* EP neurons co-expressed *Sst* (Figure S7A). We also identified a cluster of this dual neurotransmitter cell type in the anterior part of the LHAd-db (Figures S7B–S7C).

Relationship of somatic RNA to protein expression

We examined the relationship of somatic RNA and protein detection. For this, we developed a method to combine immunofluorescence (IF) with EASI-FISH (STAR Methods; Figures S7D and S7E). We found that all neuropeptide expressing neurons detected by IF also showed positive RNA signal with FISH (Figures 6J–6L). For neuropeptides that are well-localized to the soma (MCH and HCRT) (Figure 6J and 6K), nearly all FISH-positive neurons were also detected with IF (Figure S7F). For neuropep-

tides that localize primarily to distal processes (TRH, SST and NTS) (Figure 6L), we detected IF in the soma of some FISH-positive neurons (Figure S7F). We also found good mRNA-to-protein correspondence for the transcription factors (OTP, MEIS2) used for LHA parcellation (Figures 6M and S7F). Thus, EASI-FISH can be readily extended to evaluate the protein and mRNA content of the same cells. This method is also compatible with multi-round FISH following IF (STAR Methods and Figure 6N).

Somatic morphology in the LHA

We characterized the somatic morphological diversity in the LHA (Figures 7A and 7B), using the accurate, automated 3D segmentation masks generated with *Starfinity*. *Pmch*+ ($3412 \pm 76.1\ \mu m^3$) and *Hcrt*+ ($3690 \pm 40.6\ \mu m^3$) neurons were the primary large neuronal populations in this region. These neurons were ~2.5-fold larger in cell body volume than the average LHA neuron volume (average excluding *Pmch*+ and *Hcrt*+: $1533 \pm 3.5\ \mu m^3$) (Figure 7C). Excluding the *Pmch*+ and *Hcrt*+ neurons, we found that the *Slc17a6*+ neurons ($1624 \pm 6\ \mu m^3$) on average are nearly $250\ \mu m^3$ larger in volume than *Slc32a1*+ neurons ($1389 \pm 3\ \mu m^3$) ($p < 0.0001$). The somatic volume was positively correlated with total RNA content, as indicated by cytosolic DAPI staining ($r = 0.93$, $p = 0$) and *Map1b* expression ($r = 0.82$, $p = 0$) (Figures S7G and S7H).

We also evaluated 3D somatic shape (Figures S7I and S7J) and identified cells with the most extreme somatic shapes based on their solidity and eccentricity (Figure 7D). This revealed a higher fraction of irregular somatic shapes in *Slc17a6*+ neurons (6%) compared to the *Slc32a1*+ neurons (2%) (Figures 7A and 7B). Neurons with irregular somatic shapes were enriched in the LHAfl (Figure S7K).

Iterative refinement of cell type marker-genes

We found that the distribution of cell sizes in Ex-10 (*Slc17a6/Gal*) had a long tail of large neurons (Figure 7E). Neurons in this cluster were enriched in two subregions (LHAd-db and LHAfl-vl) (Figure 7F), with the large neurons primarily in the LHAfl-vl subregion. This raised the possibility that this is a distinct but rare cell type.

To test this hypothesis, we subdivided the most similar scRNA-seq cluster (seq-e9) and identified *Oxytocin* (*Oxt*) as the top differentially expressed gene between the subdivided scRNA-seq clusters (Figure 7G). To examine whether *Oxt* can be used to separate this population from the mediodorsal *Slc17a6*+/*Gal*+ subpopulation, we probed for *Slc17a6*, *Gal* and *Oxt* with EASI-FISH in a new sample to independently validate the existence of this population. Indeed, *Oxt* expression was detected in a small group of neurons in the ventral lateral part of the LHA. Consistent with our predictions, these neurons had large cell bodies ($3089 \pm 140\ \mu m^3$) and co-expressed *Slc17a6* and *Gal* (100%, 13/13) (Figure 7H). This demonstrated how spatial and morphological measurements can facilitate the discovery

---

(G) scRNA-seq cluster corresponding to Ex-10 was subdivided into subclusters differentiated by *Oxt* expression.

(H) *Oxt* marks an Ex-10 subpopulation. Top: Spatial distribution of Ex-10 subpopulation marked by *Oxt* expression. Scale bars: 100 μm. Bottom: inset showing co-expression of *Slc17a6* and *Galanin (Gal)* in the *Oxt* subpopulation. Scale bar: 25 μm.

(I) *Fos* enrichment in *Hcrt* neurons (Ex-2). *p < 0.05. Statistics: Table S1.

(J) *Fos* in *Nts* subpopulations in *ad libitum* food and water (AL) and water-deprived (WD) animals. Numbers indicate neurons from each cell type in each group. *p < 0.05. Statistics: Table S1.

(K) Image showing *Fos* expression in Inh-9 (*Nts/Meis2*) neurons (arrowheads on left). Scale bar: 100 μm. Right: representative Inh-9 neuron. Scale bar: 10 μm.

of rare cell types with unusually large cell bodies, using a limited number of marker-genes.

Cell-type-specific expression of *Fos*

Expression of the immediate early gene, *Fos*, is associated with plasticity and changes in neuronal activity. We noticed basal expression of *Fos* in *Hcrt* neurons in scRNA-seq datasets (Figure S7L) which was confirmed with EASI-FISH (Figure 7I). This may be associated with the broadly tuned arousal functions of these neurons (Mileykovskiy et al., 2005) in the context of experimental handling. *Hcrt⁻* clusters showed low *Fos* expression, and we went on to use EASI-FISH to examine enrichment of *Fos* in the LHA following behavioral perturbation.

The LHA is involved in multiple motivated behaviors, including thirst, which has been localized to a broad set of *Nts⁺* neurons that express *Fos* in response to dehydration (Brown et al., 2015; Brown et al., 2019) and selectively elicit water-drinking when activated (Kurt et al., 2019). Because we identified 4 *Nts* populations by EASI-FISH (1 excitatory and 3 inhibitory), we performed a separate, targeted analysis comparing mice with *ad libitum* (AL) food and water to water-deprived (WD) mice to determine which of these *Nts* cell types was involved in the response to dehydration. *Fos* was significantly elevated in WD mice in 1 out of the 4 *Nts⁺* cell types, Inh-9 (*Nts/Meis2*) (Figure 7J), which is located selectively in the LHAs-db (Figure 6I). In contrast, fourteen additional clusters that could be assigned from the marker-genes used for this experiment did not show statistically significant *Fos* enrichment (Figure S7M). The identification of a single *Nts⁺* subpopulation located in a specific LHA compartment shows that the molecularly defined cell types and subregions predicted by our analysis have distinct response profiles to physiological perturbations.

DISCUSSION

We report a resource that includes methods and a turnkey computational analysis pipeline for multi-round FISH datasets in thick sections of brain tissue. EASI-FISH enables quantitative *in situ* measurements of gene expression with cellular resolution using commercial laboratory equipment and provides single RNA puncta counts, detailed spatial information, and morphological characteristics of the underlying cells.

EASI-FISH is optimized for multi-round, multiplex FISH analysis of all cells within continuous tissue volumes that are thicker (300 μm) than what has been achieved with previous methods. This thickness is suitable for aligning samples from different subjects. We found that volumetric mapping of neuronal gene co-expression in brain tissue was critical for identifying the structural subdivisions that we discovered in the LHA across multiple animals. In addition, use of thick samples facilitated LHA parcellation that showed inter-subject consensus. A 300-μm axial depth is also consistent with tissue dimensions for other neuroscience data collection modalities that could potentially be combined with EASI-FISH, for example brain slice recordings or *in vivo* two-photon calcium imaging (Xu et al., 2020). Alternative methods that are constrained to thin tissue samples for FISH probe penetration or imaging and analysis face the challenge of sub-sectioning and aligning many adjacent sections to obtain

a volumetric perspective. EASI-FISH is likely to be valuable for mapping the molecular characteristics of continuous brain regions as well as in experimental pipelines that integrate measurements in thick tissue samples (including from living animals) with post hoc FISH.

We used EASI-FISH to characterize the molecular, spatial, morphological, and functional diversity of LHA neurons. We discovered a molecularly defined subregional LHA parcellation into which distinct cell types were selectively localized. The EASI-FISH pipeline generated high-quality quantitative spatial gene expression data, which enabled a machine learning approach to identify LHA subregions based on combinations of four marker-genes. This approach also recapitulated cortical lamination based on only 4 genes. Past efforts at parcellation of the rat LHA have not elucidated most of the regions that we examined (Geeraedts et al., 1990; Hahn, 2010; Hahn and Swanson, 2010); for example, as with prior work in the rat brain, we found evidence for a suprafornical cell dense zone (Hahn and Swanson, 2010) in the mouse. However, consideration of molecularly defined cell type information subdivided this suprafornical cell density into distinct, molecularly defined laminar structures that extended laterally. More generally, molecularly defined cell types are increasingly appreciated as a fundamental unit of brain organization (Zeng and Sanes, 2017). Thus, brain parcellation should incorporate organizational principles that are reliant on cell type distributions, which requires detailed spatial analysis of cellular multi-gene co-expression relationships.

High-quality cell segmentation with the EASI-FISH pipeline allowed us to examine the relationship of marker-gene expression to cell size and somatic morphology in the LHA. With morphological analysis of LHA cell types, we demonstrated that Ex-10 (*Th/Gal*) could be subdivided to reveal an *Oxt*-expressing subpopulation that defined a spatially segregated set of large magnocellular neurons. We also used EASI-FISH to refine the diverse *Nts⁺* neuron population in the LHA, which we showed comprises four subtypes, and only a single population, Inh-9, was modulated by water deprivation. Thus, molecularly defined neuron classifications in the LHA that are derived from gene co-expression relationships can specify groups of neurons with distinct cell size or functional tuning.

Our analysis of the LHA highlights the extent of spatial diversity in cell types and axonal inputs that should be considered when investigating LHA function. Thus, detailed examination of LHA function will be dependent on convenient and robust multi-gene co-expression analysis. EASI-FISH fills an important gap in neural circuit research by facilitating a seamless extension of cell-type analysis generated in dissociated neurons to experimental modalities that also require information about *in situ* anatomical localization. This will lower the barrier to make progress toward extending the molecular revolution in neuroscience to functional and systems analysis, which is essential for understanding the interplay between neural coding and molecular properties in behavior and disease.

Limitations of study

To prioritize ease of use, we chose to use non-barcoded sequential gene probing, which simplifies experimental design but limits gene number. Spot-to-spot alignment (Moffitt et al., 2018; Shah

et al., 2017) is likely possible using this method, which would facilitate barcoding methods and increase gene number, although this would also increase imaging complexity and analysis requirements.

The molecularly defined subregions as well as the distribution of cell types described in this study only covered a limited portion of the LHA, which spans ~3 mm anterior to posterior; thus, future work will extend these methods to the entire LHA. Moreover, additional studies are needed to establish the causal contributions of the LHA subregions and molecularly defined cell types to behavior and physiology.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Mice
- METHOD DETAILS
  - Single-cell RNA sequencing
  - EASI-FISH protocol summary
  - Reagents and chemicals
  - Tissue fixation and preparation
  - RNA anchoring, gelation, and Proteinase K digestion
  - DNase I digestion
  - *In situ* Hybridization and HCR
  - Image acquisition, sample handling, and multiplexing
  - Notes on Expansion factor, tissue thickness, and imaging conditions
  - Combining immunofluorescence (IF) detection with EASI-FISH
  - *Fos* activity during water deprivation
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - scRNA-Seq analysis and marker-gene selection for EASI-FISH
  - Integration of scRNA-Seq datasets
  - EASI-FISH data processing summary
  - Image Stitching
  - Image Registration
  - FISH Spot Detection
  - 3D Soma Segmentation
  - End-to-end Analysis Pipeline
  - Visualization
  - Post-processing
  - Evaluation of EASI-FISH detection efficiency, selectivity, and sensitivity
  - FISH cluster analysis
  - FISH to scRNA-Seq data mapping
  - LHA boundaries and neighboring brain regions
  - Marker-gene spatial expression analysis with PCA
  - Spatial enrichment of selected marker-genes
  - Anatomical segmentation of the LHA

- Regional parcellation of the visual cortex with STAR-Map data
- Neighborhood cell type composition
- Average distance to nearest neighbor (ANN) analysis
- Spatial distribution of molecularly defined cell types
- Spatial overlap of molecularly defined cell types
- Prediction of spatial position with gene expression
- Cross-correlation analysis between biological replicates
- Vizgen MERFISH data analysis
- Connectivity analysis
- Fos activity in molecularly defined cell types
- Statistics and figure scales

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2021.11.024.

AUTHOR CONTRIBUTIONS

Conceptualization and supervision: S.M.S., P.W.T., W.K., K.S., E.W.M.; Methodology: P.W.T., Y.W., M.E.; Investigation: Y.W., M.E., F.E.H., H.Y.; Software: G.F., T.W., S.X. (registration), M.W., U.S., S.S. (segmentation); K.R., C.G. (Nextflow pipeline), A.L.L. (sequencing). Formal analysis: Y.W.; Writing: Y.W. and S.M.S. with input from all co-authors.

DECLARATION OF INTERESTS

REFERENCES

Alon, S., Goodwin, D.R., Sinha, A., Wassie, A.T., Chen, F., Daugharthy, E.R., Bando, Y., Kajita, A., Xue, A.G., Marrett, K., et al.; IMAXT Consortium (2021). Expansion sequencing: Spatially precise in situ transcriptomics in intact biological systems. Science *371*. https://doi.org/10.1126/science.aax2656.

Bernardis, L.L., and Bellinger, L.L. (1993). The lateral hypothalamic area revisited: neuroanatomy, body weight regulation, neuroendocrinology and metabolism. Neurosci. Biobehav. Rev. *17*, 141–193.

Betley, J.N., Cao, Z.F., Ritola, K.D., and Sternson, S.M. (2013). Parallel, redundant circuit organization for homeostatic control of feeding behavior. Cell *155*, 1337–1350.

Brown, J.A., Woodworth, H.L., and Leininger, G.M. (2015). To ingest or rest? Specialized roles of lateral hypothalamic area neurons in coordinating energy balance. Front. Syst. Neurosci. *9*, 9.

Brown, J.A., Wright, A., Bugescu, R., Christensen, L., Olson, D.P., and Leininger, G.M. (2019). Distinct subsets of lateral hypothalamic neurotensin neurons are activated by leptin or dehydration. Sci. Rep. 9, 1873.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. 36, 411–420.

Cembrowski, M.S., Phillips, M.G., DiLisio, S.F., Shields, B.C., Winnubst, J., Chandrashekar, J., Bas, E., and Spruston, N. (2018). Dissociable structural and functional hippocampal outputs via distinct subiculum cell classes. Cell 173, 1280–1292.e18.

Chen, F., Tillberg, P.W., and Boyden, E.S. (2015a). Optical imaging. Expansion microscopy. Science 347, 543–548.

Chen, K.H., Boettiger, A.N., Moffitt, J.R., Wang, S., and Zhuang, X. (2015b). RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. Science 348, aaa6090.

Chen, F., Wassie, A.T., Cote, A.J., Sinha, A., Alon, S., Asano, S., Daugharthy, E.R., Chang, J.B., Marblestone, A., Church, G.M., et al. (2016). Nanoscale imaging of RNA with expansion microscopy. Nat. Methods 13, 679–684.

Choi, H.M.T., Schwarzkopf, M., Fornace, M.E., Acharya, A., Artavanis, G., Stegmaier, J., Cunha, A., and Pierce, N.A. (2018). Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. Development 145, e1202. https://doi.org/10.1242/dev.165753.

Codeluppi, S., Borm, L.E., Zeisel, A., La Manno, G., van Lunteren, J.A., Svensson, C.I., and Linnarsson, S. (2018). Spatial organization of the somatosensory cortex revealed by osmFISH. Nat. Methods 15, 932–935.

Cressie, N.A.C. (2015). Statistics for Spatial Data (Wiley).

Di Tommaso, P., Chatzou, M., Floden, E.W., Barja, P.P., Palumbo, E., and Notredame, C. (2017). Nextflow enables reproducible computational workflows. Nat. Biotechnol. 35, 316–319.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21.

Dubé, D., and Pelletier, G. (1979). Effect of colchicine on the immunohistochemical localization of somatostatin in the rat brain: light and electron microscopic studies. J. Histochem. Cytochem. 27, 1577–1581.

Fischler, M.A., and Bolles, R.C. (1981). Random sample consensus - A paradigm for model-fitting with applications to image-analysis and automated cartography. Commun. ACM 24, 381–395.

Franklin, K.B.J., and Paxinos, G. (1997). The Mouse Brain in Stereotaxic Coordinates (Academic Press).

Gao, R., Asano, S.M., Upadhyayula, S., Pisarev, I., Milkie, D.E., Liu, T.L., Singh, V., Graves, A., Huynh, G.H., Zhao, Y., et al. (2019). Cortical column and whole-brain imaging with molecular contrast and nanoscale resolution. Science 363. https://doi.org/10.1126/science.aau8302.

Geeraedts, L.M., Nieuwenhuys, R., and Veening, J.G. (1990). Medial forebrain bundle of the rat: IV. Cytoarchitecture of the caudal (lateral hypothalamic) part of the medial forebrain bundle bed nucleus. J. Comp. Neurol. 294, 537–568.

Grimm, J.B., Brown, T.A., Tkachuk, A.N., and Lavis, L.D. (2017). General Synthetic Method for Si-Fluoresceins and Si-Rhodamines. ACS Cent. Sci. 3, 975–985.

Gur, D., Bain, E.J., Johnson, K.R., Aman, A.J., Pasoili, H.A., Flynn, J.D., Allen, M.C., Deheyn, D.D., Lee, J.C., Lippincott-Schwartz, J., and Parichy, D.M. (2020). In situ differentiation of iridophore crystallotypes underlies zebrafish stripe patterning. Nat. Commun. 11, 6391.

Hafemeister, C., and Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol. 20, 296.

Hahn, J.D. (2010). Comparison of melanin-concentrating hormone and hypocretin/orexin peptide expression patterns in a current parceling scheme of the lateral hypothalamic zone. Neurosci. Lett. 468, 12–17.

Hahn, J.D., and Swanson, L.W. (2010). Distinct patterns of neuronal inputs and outputs of the juxtaparaventricular and suprafornical regions of the lateral hypothalamic area in the male rat. Brain Res. Brain Res. Rev. 64, 14–103.

Hahn, J.D., and Swanson, L.W. (2015). Connections of the juxtaventromedial region of the lateral hypothalamic area in the male rat. Front. Syst. Neurosci. 9, 66.

Hahn, J.D., Sporns, O., Watts, A.G., and Swanson, L.W. (2019). Macroscale intrinsic network architecture of the hypothalamus. Proc. Natl. Acad. Sci. USA 116, 8018–8027.

Hempel, C.M., Sugino, K., and Nelson, S.B. (2007). A manual method for the purification of fluorescently labeled neurons from the mammalian brain. Nat. Protoc. 2, 2924–2929.

Horjales-Araujo, E., Hellysaz, A., and Broberger, C. (2014). Lateral hypothalamic thyrotropin-releasing hormone neurons: distribution and relationship to histochemically defined cell populations in the rat. Neuroscience 277, 87–102.

Kahn, D., Abrams, G.M., Zimmerman, E.A., Carraway, R., and Leeman, S.E. (1980). Neurotensin neurons in the rat hypothalamus: an immunocytochemical study. Endocrinology 107, 47–54.

Kawana, E., and Watanabe, K. (1981). A cytoarchitectonic study of zona incerta in the rat. J. Hirnforsch. 22, 535–541.

Kempadoo, K.A., Tourino, C., Cho, S.L., Magnani, F., Leininger, G.M., Stuber, G.D., Zhang, F., Myers, M.G., Deisseroth, K., de Lecea, L., and Bonci, A. (2013). Hypothalamic neurotensin projections promote reward by enhancing glutamate transmission in the VTA. J. Neurosci. 33, 7618–7626.

Kolmac, C., and Mitrofanis, J. (1999). Distribution of various neurochemicals within the zona incerta: an immunocytochemical and histochemical study. Anat. Embryol. (Berl.) 199, 265–280.

Kurt, G., Woodworth, H.L., Fowler, S., Bugescu, R., and Leininger, G.M. (2019). Activation of lateral hypothalamic area neurotensin-expressing neurons promotes drinking. Neuropharmacology 154, 13–21.

Lein, E.S., Hawrylycz, M.J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A.F., Boguski, M.S., Brockway, K.S., Byrnes, E.J., et al. (2007). Genome-wide atlas of gene expression in the adult mouse brain. Nature 445, 168–176.

Lionnet, T., Czaplinski, K., Darzacq, X., Shav-Tal, Y., Wells, A.L., Chao, J.A., Park, H.Y., de Turris, V., Lopez-Jones, M., and Singer, R.H. (2011). A transgenic mouse for in vivo detection of endogenous labeled mRNA. Nat. Methods 8, 165–170.

Lubeck, E., Coskun, A.F., Zhiyentayev, T., Ahmad, M., and Cai, L. (2014). Single-cell in situ RNA profiling by sequential hybridization. Nat. Methods 11, 360–361.

Luo, L., Callaway, E.M., and Svoboda, K. (2018). Genetic Dissection of Neural Circuits: A Decade of Progress. Neuron 98, 256–281.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell 161, 1202–1214.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. J. 17, 10.

Mickelsen, L.E., Bolisetty, M., Chimileski, B.R., Fujita, A., Beltrami, E.J., Costanzo, J.T., Naparstek, J.R., Robson, P., and Jackson, A.C. (2019). Single-cell transcriptomic analysis of the lateral hypothalamic area reveals molecularly distinct populations of inhibitory and excitatory neurons. Nat. Neurosci. 22, 642–656.

Mileykovskiy, B.Y., Kiyashchenko, L.I., and Siegel, J.M. (2005). Behavioral correlates of activity in identified hypocretin/orexin neurons. Neuron 46, 787–798.

Moffitt, J.R., Bambah-Mukku, D., Eichhorn, S.W., Vaughn, E., Shekhar, K., Perez, J.D., Rubinstein, N.D., Hao, J., Regev, A., Dulac, C., and Zhuang, X. (2018). Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. Science 362. Published online November 1, 2018. https://doi.org/10.1126/science.aau5324.

Napari contributors (2019). napari: a multi-dimensional image viewer for python (Zenodo). https://doi:10.5281/zenodo.3555620.

Nicovich, P.R., Taormina, M.J., Baker, C.A., Nguyen, T.N., Thomsen, E.R., Garren, E., Long, B., Gorham, M., Miller, J.A., Hage, T., et al. (2019). Multimodal cell type correspondence by intersectional mFISH in intact tissues. bioRxiv, 525451.

Oh, S.W., Harris, J.A., Ng, L., Winslow, B., Cain, N., Mihalas, S., Wang, Q., Lau, C., Kuan, L., Henry, A.M., et al. (2014). A mesoscale connectome of the mouse brain. Nature 508, 207–214.

Patterson, C.M., Wong, J.M., Leininger, G.M., Allison, M.B., Mabrouk, O.S., Kasper, C.L., Gonzalez, I.E., Mackenzie, A., Jones, J.C., Kennedy, R.T., and Myers, M.G., Jr. (2015). Ventral tegmental area neurotensin signaling links the lateral hypothalamus to locomotor activity and striatal dopamine efflux in male mice. Endocrinology 156, 1692–1700.

Petrovich, G.D. (2018). Lateral Hypothalamus as a Motivation-Cognition Interface in the Control of Feeding Behavior. Front. Syst. Neurosci. 12, 14.

Pietzsch, T., Saalfeld, S., Preibisch, S., and Tomancak, P. (2015). BigDataViewer: visualization and processing for large image data sets. Nat. Methods 12, 481–483.

Preibisch, S., Saalfeld, S., and Tomancak, P. (2009). Globally optimal stitching of tiled 3D microscopic image acquisitions. Bioinformatics 25, 1463–1465.

Qian, X., Harris, K.D., Hauling, T., Nicoloutsopoulos, D., Muñoz-Manchado, A.B., Skene, N., Hjerling-Leffler, J., and Nilsson, M. (2020). Probabilistic cell typing enables fine mapping of closely related cell types in situ. Nat. Methods 17, 101–106.

Romanov, R.A., Tretiakov, E.O., Kastriti, M.E., Zupancic, M., Häring, M., Korchynska, S., Popadin, K., Benevento, M., Rebernik, P., Lallemend, F., et al. (2020). Molecular design of hypothalamus development. Nature 582, 246–252.

Rossi, M.A., Basiri, M.L., McHenry, J.A., Kosyk, O., Otis, J.M., van den Munkhof, H.E., Bryois, J., Hübel, C., Breen, G., Guo, W., et al. (2019). Obesity remodels activity and transcriptional state of a lateral hypothalamic brake on feeding. Science 364, 1271–1274.

Rossi, M.A., Basiri, M.L., Liu, Y., Hashikawa, Y., Hashikawa, K., Fenno, L.E., Kim, Y.S., Ramakrishnan, C., Deisseroth, K., and Stuber, G.D. (2021). Transcriptional and functional divergence in lateral hypothalamic glutamate neurons projecting to the lateral habenula and ventral tegmental area. Neuron. S0896-6273(21)00693-0. https://doi.org/10.1016/j.neuron.2021.09.020.

Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol. 33, 495–502.

Saunders, A., Macosko, E.Z., Wysoker, A., Goldman, M., Krienen, F.M., de Rivera, H., Bien, E., Baum, M., Bortolin, L., Wang, S., et al. (2018). Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain. Cell 174, 1015–1030.e16.

Schmidt, U., Weigert, M., Broaddus, C., and Myers, G. (2018). Cell Detection with Star-Convex Polygons. Lecture Notes in Computer Science . 11071, 265–273.

Shah, S., Lubeck, E., Zhou, W., and Cai, L. (2016). In Situ Transcription Profiling of Single Cells Reveals Spatial Organization of Cells in the Mouse Hippocampus. Neuron 92, 342–357.

Shah, S., Lubeck, E., Zhou, W., and Cai, L. (2017). seqFISH Accurately Detects Transcripts in Single Cells and Reveals Robust Spatial Organization in the Hippocampus. Neuron 94, 752–758.e1.

Smith, K., Li, Y., Piccinini, F., Csucs, G., Balazs, C., Bevilacqua, A., and Horvath, P. (2015). CIDRE: an illumination-correction method for optical microscopy. Nat. Methods 12, 404–406.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell 177, 1888–1902.e21.

Stuber, G.D., and Wise, R.A. (2016). Lateral hypothalamic circuits for feeding and reward. Nat. Neurosci. 19, 198–205.

Tang, M., Kaymaz, Y., Logeman, B.L., Eichhorn, S., Liang, Z.S., Dulac, C., and Sackton, T.B. (2020). Evaluating Single-Cell Cluster Stability Using The Jaccard Similarity Index. Bioinformatics 37, 2212–2214.

Tasic, B., Yao, Z., Graybuck, L.T., Smith, K.A., Nguyen, T.N., Bertagnolli, D., Goldy, J., Garren, E., Economo, M.N., Viswanathan, S., et al. (2018). Shared and distinct transcriptomic cell types across neocortical areas. Nature 563, 72–78.

Tillberg, P.W., Chen, F., Piatkevich, K.D., Zhao, Y., Yu, C.C., English, B.P., Gao, L., Martorell, A., Suk, H.J., Yoshida, F., et al. (2016). Protein-retention expansion microscopy of cells and tissues labeled using standard fluorescent proteins and antibodies. Nat. Biotechnol. 34, 987–992.

van den Pol, A.N., Acuna-Goycolea, C., Clark, K.R., and Ghosh, P.K. (2004). Physiological properties of hypothalamic MCH neurons identified with selective expression of reporter gene after recombinant virus infection. Neuron 42, 635–652.

Veening, J.G., Te Lie, S., Posthuma, P., Geeraedts, L.M.G., and Nieuwenhuys, R. (1987). A topographical analysis of the origin of some efferent projections from the lateral hypothalamic area in the rat. Neuroscience 22, 537–551.

Vizgen (2021). Vizgen Data Release V1.0. https://info.vizgen.com/mouse-brain-map?submissionGuid=f1a3404d-97eb-4d37-89e3-2e0ff6cce452.

Wallace, M.L., Saunders, A., Huang, K.W., Philson, A.C., Goldman, M., Macosko, E.Z., McCarroll, S.A., and Sabatini, B.L. (2017). Genetically Distinct Parallel Pathways in the Entopeduncular Nucleus for Limbic and Sensorimotor Output of the Basal Ganglia. Neuron 94, 138–152.e5.

Waller, L.A., and Crawford, C.A.G. (2004). Applied Spatial Statistics for Public Health Data (Wiley-Interscience).

Wang, Z., Bovik, A.C., Sheikh, H.R., and Simoncelli, E.P. (2004). Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. 13, 600–612.

Wang, F., Flanagan, J., Su, N., Wang, L.C., Bui, S., Nielson, A., Wu, X., Vo, H.T., Ma, X.J., and Luo, Y. (2012). RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. J. Mol. Diagn. 14, 22–29.

Wang, X., Allen, W.E., Wright, M.A., Sylwestrak, E.L., Samusik, N., Vesuna, S., Evans, K., Liu, C., Ramakrishnan, C., Liu, J., et al. (2018). Three-dimensional intact-tissue sequencing of single-cell transcriptional states. Science 361, eaat5691.

Warfield, S.K., Zou, K.H., and Wells, W.M. (2004). Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans. Med. Imaging 23, 903–921.

Weigert, M., Schmidt, U., Haase, R., Sugawara, K., and Myers, G. (2020). Star-convex Polyhedra for 3D Object Detection and Segmentation in Microscopy. Ieee Wint Conf Appl, 3655–3662.

Xu, S., Yang, H., Menon, V., Lemire, A.L., Wang, L., Henry, F.E., Turaga, S.C., and Sternson, S.M. (2020). Behavioral state coding by molecularly defined paraventricular hypothalamic cell type ensembles. Science 370. https://doi.org/10.1126/science.abb2494.

Yushkevich, P.A., Pluta, J., Wang, H., Wisse, L.E., Das, S., and Wolk, D. (2016). Fast Automatic Segmentation of Hippocampal Subfields and Medial Temporal Lobe Subregions in 3 Tesla and 7 Tesla MRI. Alzheimers Dement. 12, 126–P127.

Zappia, L., and Oshlack, A. (2018). Clustering trees: a visualization for evaluating clusterings at multiple resolutions. Gigascience 7. https://doi:10.1093/gigascience/giy083.

Zeisel, A., Hochgerner, H., Lönnerberg, P., Johnsson, A., Memic, F., van der Zwan, J., Häring, M., Braun, E., Borm, L.E., La Manno, G., et al. (2018). Molecular Architecture of the Mouse Nervous System. Cell 174, 999–1014.e22.

Zeng, H., and Sanes, J.R. (2017). Neuronal cell-type classification: challenges, opportunities and the path forward. Nat. Rev. Neurosci. 18, 530–546.

STAR★METHODS

KEY RESOURCES TABLE

Reagent or resource	Source	Identifier
Antibodies		
Anti-MCH	Atlas Antibodies	Cat. # HPA046055; RRID:AB_2679517
Anti-Orexin A (KK09)	Santa Cruz	Cat. # sc-80263; RRID:AB_1126868
Anti-OTP	Thermo Fisher	Cat. # PA531513; RRID:AB_2548986
Anti-MEIS2	Atlas Antibodies	Cat. # HPA003256; RRID:AB_1079356
Anti-pro-TRH	Millipore Sigma	Cat. # ABN1658; RRID:AB_2884885
Anti-SST	Atlas Antibodies	Cat. # HPA019472; RRID:AB_1857360
Anti-NTS	Phoenix Pharmaceuticals Inc	Cat. # H-048-03
Anti-VGLUT2	Synaptic Systems	Cat. # 135 416; RRID:AB_2619824
Anti-VGAT	Synaptic Systems	Cat. # 131 011; RRID:AB_887872
Secondary antibody HCR probe (B4), Donkey Anti-Rabbit	Molecular Instruments	N/A
Chemicals, peptides, and recombinant proteins		
Melphalan	Cayman Chemicals	Cat. # 16665
Acryloyl-X, SE	Thermo Fisher	Cat. # A20770
Label-IT Amine	Mirus Bio	Cat. # MIR 3900
Ribonucleoside Vanadyl Complex	NEB	Cat. # S1402S
RNase-Free DNase Set (50)	QIAGEN	Cat. # 79254
Proteinase K	NEB	Cat. # P8107S
DAPI	Sigma	Cat. # D9542
Janelia Fluor 669, SE	Tocris	Cat. # 6420
N,N,N′,N′-Tetramethyl ethylenediamine	Sigma	Cat. # T22500
Ammonium persulfate	Sigma	Cat. # A3678
Acrylamide solution	Sigma	Cat. # A4058
4-Hydroxy-TEMPO	Sigma	Cat. # 176141
N, N'-Methylenebisacrylamide	Sigma	Cat. # M7279
Acrylamide	Sigma	Cat. # A9099
Acrylic Acid	Sigma	Cat. # 147230
DMSO	Sigma	Cat. # 570672
MOPS buffer	Sigma	Cat. # M1254
20x SSC	Thermo Fisher	Cat. # AM9763
Nuclease-free water	Thermo Fisher	Cat. # AM9932
NaOH	Fisher scientific	Cat. # SS267
Poly-L-Lysine	Pelco	Cat. # 18026
DEXTRAN SULFATE 50%, 100ML	Sigma	Cat. # S4030
Formamide	Fisher scientific	Cat. # BP227-100
PBS	Fisher scientific	Cat. # BP24384
RNase away/DNase away	Fisher scientific	Cat. # 10328011
Photo-Flo 200	EMS	Cat. # 74257
RNase-free BSA	Thermo Fisher	Cat. # AM2616
Critical commercial assays		
QIAquick Nucleotide Removal Kit (50)	QIAGEN	Cat. # 28304
Deposited data		
LHA scRNA-seq (10x Genomics)	Mickelsen et al., 2019	GEO: GSE125065
LHA scRNA-seq (DropSeq)	Rossi et al., 2019	GEO: GSE130597

(Continued on next page)

Continued

Reagent or resource	Source	Identifier
LHA scRNA-seq (Smart-SCRB)	this study	GEO: GSE167293
EASI-FISH pipeline example image data	this study	https://janelia.figshare.com/collections/EASI-FISH_analysis_pipeline/5276708
LHA EASI-FISH data	this study	https://janelia.figshare.com/articles/dataset/EASI-FISH_enabled_spatial_analysis_of_molecular_cell_types_in_the_lateral_hypothalamus/13749154
Starfinity training data and trained model for 3D segmentation	this study	https://janelia.figshare.com/articles/dataset/Ground-truth_cell_body_segmentation_used_for_Starfinity_training/13624268
Experimental models: Organisms/strains		
Mouse: C57BL/6	Jackson Laboratory	JAX stock #000664
Mouse: *Agrp*-IRES-Cre	Jackson Laboratory	JAX stock #012899
Mouse: Ai6	Jackson Laboratory	JAX stock #007906
Oligonucleotides		
HCR probes	Molecular Instrument	Table S5
HCR Amplifier B1	Molecular Instrument	N/A
HCR Amplifier B2	Molecular Instrument	N/A
HCR Amplifier B3	Molecular Instrument	N/A
HCR Amplifier B4	Molecular Instrument	N/A
HCR Amplifier B5	Molecular Instrument	N/A
Software and algorithms		
EASI-FISH pipeline	this study	https://zenodo.org/record/5652239; https://github.com/JaneliaSciComp/multifish
Starfinity 3D segmentation	this study	https://github.com/stardist/stardist/tree/refinement
BigStream registration pipeline	this study	https://github.com/GFleishman/bigstream
Stitching-Spark stitching pipeline	Gao et al., 2019	https://github.com/saalfeldlab/stitching-spark
Airlocalize	Lionnet et al., 2011	https://github.com/timotheelionnet/AIRLOCALIZE/
Custom code	this study	https://zenodo.org/record/5658557; https://github.com/multiFISH/EASI-FISH
Seurat 4.0.1	Stuart et al., 2019	Satijalab.org; RRID: SCR_016341
Ilastik	Ilastik	https://www.ilastik.org/
MATLAB	Mathworks	https://www.mathworks.com/products/matlab.html
Fiji	ImageJ	https://imagej.net/software/fiji/
Python v3.7		Python.org; RRID: SCR 008394
Nextflow	Di Tommaso et al., 2017	https://www.nextflow.io/
Greedy registration	(Yushkevich et al., 2016)	https://github.com/pyushkevich/greedy
n5-viewer	Saalfeld lab	https://github.com/saalfeldlab/n5-viewer
Napari	(Napari contributors, 2019)	https://napari.org/
Other		
Zeiss Lightsheet Z.1 microscope	Zeiss	https://www.zeiss.com/microscopy/us/products/imaging-systems/light-sheet-microscope-for-lsfm-imaging-of-live-and-cleared-samples-lightsheet-7.html
Press-to-Seal Silicone Isolator with Adhesive	Thermo Fisher	Cat. # P24743
8mm glass coverslip	Harvard Apparatus	Cat. # BS4 64-0701
Zeiss Lightsheet Z.1 imaging holder	Svoboda Lab and Janelia Experimental Technology	https://www.janelia.org/open-science/zeiss-lightsheet-z1-sample-holder

RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources and reagents should be directed to Scott Sternson (ssternson@health.ucsd.edu).

Materials availability
This study did not generate new unique reagents.

Data and code availability
All EASI-FISH data generated in this study have been deposited at figshare (https://janelia.figshare.com/articles/dataset/EASI-FISH_enabled_spatial_analysis_of_molecular_cell_types_in_the_lateral_hypothalamus/13749154). The single cell RNA-seq dataset generated in this study has been deposited to GEO (Gene Expression Omnibus, https://www.ncbi.nlm.nih.gov/geo/) with accession number GSE167293.

Software, pipeline, and custom code generated for EASI-FISH data processing are available on GitHub (https://zenodo.org/record/5652239, https://zenodo.org/record/5658557; for active updates: https://github.com/JaneliaSciComp/multifish and https://github.com/multiFISH/EASI-FISH/) with detailed instructions. Example dataset and *Starfinity* training data and model used in this manuscript have been deposited at figshare (https://janelia.figshare.com/collections/EASI-FISH_analysis_pipeline/5276708).

Any additional information required to reanalyze the data reported in this paper is available from the lead contacts upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mice
Adult mice were used for all experiments. Specific breakdown of age/sex are given for each experiment in the corresponding methods description. For single-cell RNA sequencing, male *Agrp-IRES-Cre × Ai9* mice (age 6-8 weeks) were used. For all EASI-FISH experiments, C57BL/6J male mice (8 weeks old) were used. All methods for animal care and use were conducted according to National Institutes of Health guidelines for animal research and approved by the Institutional Animal Care and Use Committee (IACUC) at Janelia Research Campus. In all cases mice were housed in a 12 h light/12 h dark cycle and had *ad libitum* access to water and chow diet unless otherwise noted.

METHOD DETAILS

Single-cell RNA sequencing
The single-cell RNA sequencing performed by us was focused on the suprafornical, tuberal LHA and its surrounding areas (additional datasets that we combined with ours included a larger volume of LHA). For single cell dissociation and collection, we used transgenic animal *Agrp-IRES-Cre × Ai9* crosses, where the tdTomato-labeled AGRP neurons extend a prominent set of axonal projections to the suprafornical region of the LHA (Betley et al., 2013) and provide a useful signal for visually guided dissection of the targeting brain region. The manual sorting procedure to isolate non-fluorescent cell bodies from micro-dissected brain slices was similar to that previously described in (Hempel et al., 2007). Briefly, we sectioned 300 μm coronal slices from male *Agrp-IRES-Cre × Ai9* mice (age 6-8 weeks) and used the tdTomato fluorescence signal from the AGRP neuron axon bundle terminal in the LHA to identify the boundaries of the LHA and then manually dissected with spring scissors. The dissected tissue sections were then subject to protease digestion (0.5 mg/mL Protease (Sigma), 1hr at room temperature), after which cells were dissociated through gentle pipetting until a single-cell suspension is obtained based on visual inspection using a microscope. Dissociated neurons from 7 animals were pooled and intact neurons were manually transferred into individual wells based on size. Each sorted single cell was lysed with 3 μl lysis buffer (0.2% Triton X-100 (Sigma) and 0.1 U/μl RNase inhibitor (Lucigen)) and cDNA libraries were prepared using the Smart-SCRB chemistry as described previously (Cembrowski et al., 2018; Xu et al., 2020). Barcoded cDNA libraries were then pooled and sequenced on a NextSeq 550 high-output flowcell with 27 bp in read 1 to obtain the barcode and UMI, and 125 bp in read 2 for cDNA. PhiX control library (Illumina) was spiked in at a final concentration of 15% to improve color balance in read 1. Libraries were sequenced to an average depth of 135,025 ± 38,401 (mean ± SD) reads per cell.

Sequencing alignment was performed as previously described (Gur et al., 2020). Sequencing adapters were trimmed from the sequencing reads with Cutadapt v2.10 (Martin, 2011) prior to alignment with STAR v2.7.5c (Dobin et al., 2013) to the M. musculus GRCm38.90. genome assembly from Ensembl (ensembl.org). Gene counts were generated using the STARsolo algorithm (https://github.com/alexdobin/STAR/blob/master/docs/STARsolo.md). Gene counts for the subset of barcodes used in each library were extracted using custom R scripts (https://zenodo.org/record/5662268).

EASI-FISH protocol summary
A brief description of EASI-FISH procedure is provided, followed by the detailed protocol below. First, RNA in tissue was preserved via covalent attachment to the hydrogel mesh. To do this, the alkylating agent, melphalan was first reacted with the succinimidyl ester of 6-((Acryloyl)amino) hexanoic Acid (Acryloyl-X, SE) to produce MelphaX. Next, MelphaX was applied to tissue to react with nucle-

otides including RNA, functionalizing them for incorporation into a tissue-gel network by polymerization of the acryloyl "tails" (Figure 1D). Tissue was then embedded into the polyacrylate gel, followed by proteolytic digestion. We found that inclusion of the ionic detergent, sodium dodecylsulfate (SDS), and increased tissue expansion through reduced salt concentration in the protease digestion step improved optical clearing and reagent penetration through thick tissue volumes. Proteolytic digestion and expansion (2 × in each dimension) reduces tissue autofluorescence (95% reduction) and light scattering, yielding a composite material that is refractive index-matched for water immersion objective lenses. For detection of mRNA transcripts, tissue-gel samples were hybridized with FISH probes bearing HCR initiators, signal amplified with HCR hairpins and imaged. DNase I digestion was used for round-to-round stripping followed by cytosolic DAPI staining to generate the cytosolic staining pattern for stitching, registration, and segmentation. With 2 × expansion, acquiring a native sample volume of 1 mm × 1 mm × 0.3 mm requires collecting an image 8-times this volume. Selective plane illumination fluorescence microscopy (SPIM or 'light sheet microscopy') avoids photobleaching of out-of-focus fluorophores and accelerates image acquisition ∼100-fold compared with confocal microscopy and thus provides a decisive advantage for rapid imaging of such large tissue volumes at high resolution.

Reagents and chemicals

MelphaX
Melphalan (Cayman Chemicals, 16665) was dissolved to 2.5 mg/ml in anhydrous DMSO (Sigma). Acryloyl-X, SE (Thermo Fisher, 20770) was dissolved to 10mg/mL in anhydrous DMSO. 4 parts of Melphalan stock were combined with 1 part of Acryloyl-X, SE stock and reacted overnight with shaking at room temperature to make MelphaX (2 mg/ml). Aliquots were stored at −20°C in a desiccated environment and used at 1 mg/ml by diluting in MOPS buffer (20 mM, pH 7.7).

Sodium Acrylate (4M)
We made sodium acrylate by reacting acrylic acid (14.6 M; Sigma, 147230) with NaOH. Briefly, in a fume hood, acrylic acid (5.5 ml) was mixed with nuclease free water (4.5 ml). 10 M NaOH (7.2 ml) (Fisher Scientific) was added gradually to prevent excessive heating. Then 1M NaOH (Fisher Scientific) was added dropwise until the pH reached 7.6-7.8. Water was added to reach a final volume of 20 ml.

Stock-X
4 M Sodium Acrylate (4.6 ml), 50% Acrylamide (w/v in water, 1 ml) (Sigma, A9099), 2% N, N'-Methylenebisacrylamide (1.5 ml) (Sigma, M7279), 5 M NaCl (8 ml), 10 × PBS (2 ml), Nuclease Free Water (1.7 ml). Aliquots were stored at −20°C.

ExM gel solution
Before gelation, Stock-X was mixed with 0.5% 4-Hydroxy-TEMPO (Sigma, 176141), 10% TEMED (Sigma, T22500) and 10% APS (Sigma, A3678) at a ratio of 94:2:2:2.

Proteinase K digestion buffer
50 mM Tris-HCl (pH 8), 50 mM NaCl, 1 mM EDTA, 0.5% Triton X-100 and 0.3% SDS.

DNase I Buffer
10 mM Tris-HCl (pH 8), 2.5 mM $MgCl_2$, 0.5 mM $CaCl_2$.

Poly-L-Lysine coating solution
Photo-Flo 200 (3.2 μl, EMS 74257) was added to Poly-L-Lysine solution (1.6 ml. Pelco, 18026) to make the Poly-L-Lysine coating solution.

HCR hairpin conjugation with custom fluorophores
We conjugated the photostable fluorophore JF669, NHS (Tocris, 6420) to amine-modified hairpin oligos from Molecular Instruments. Briefly, 100 μM amine-modified hairpin oligos (h1 and h2) (5 μl each) were air-dried with SpeedVac for 30 min. Dried oligos were dissolved in 0.1 M sodium bicarbonate (3 μl) (pH 8-9). JF669, NHS (0.1 mg) was dissolved in DMSO (2 μl) and reacted with amine-modified oligos overnight in the dark at room temperature. Excess JF669, NHS was removed with QIAquick Nucleotide removal kit (QIAGEN, 28304) and the conjugated JF669 hairpins were reconstituted with nuclease free water to a final concentration of 60 ng/ml (∼3 μM).

Tissue fixation and preparation

C57BL/6J male mice (8 weeks old) were used for all FISH experiments described in this study. Animals were anesthetized with isoflurane and perfused with RNase-free PBS (15ml) followed by ice-cold 4% paraformaldehyde (PFA) (50 ml). Brain tissue was dissected and fixed in 4% PFA overnight before sectioning on a vibratome. Brain coronal slices (300 μm) were sectioned and stored in 70% ethanol at 4°C for up to 6 months. For the lateral hypothalamus experiment, LHA region (∼2.5 × 4 mm) around Bregma −1.5 to −1.1 was cut out using anatomical landmarks as boundaries, including the mammillothalamic tract (mtt), zona incerta (ZI), fornix and optic tract. For ease of orientation and optimal imaging, the tissue was cut as a rectangle. An RNase-free paintbrush was used for tissue handling.

RNA anchoring, gelation, and Proteinase K digestion

The tissue slice was rehydrated in PBS at room temperature (RT) (2 × 15 min) and incubated in MOPS buffer (20 mM, pH 7.7, 30 min). Tissue was incubated overnight (37°C) in MOPS buffer (50 μl) with 1 mg/ml MelphaX and 0.1mg/ml Acryloyl-X, SE. The next day, tissue was rinsed in PBS (2 × 5 min) and placed in a 9 mm wide × 0.5 mm deep gasket (Invitrogen) on a glass slide that was previously

coated with Poly-L-Lysine (1 μl) and allowed to dry. Gel solution was freshly made (see recipe above) and kept on ice. Tissue was equilibrated with gel solution (40 μl, 3 × 10 min) at 4°C. A coverslip was used to seal the gasket and gel was allowed to form at 37°C for 2 hours. The coverslip and gasket were then removed to recover the tissue-gel. The tissue-gel was trimmed to a rectangle shape, with a corner cut to help with orientation. Tissue-gel samples were then transferred into a 2 mL Eppendorf tube and digested overnight (37°C) in Proteinase K buffer (750 μl) with 7.5 μl of 800 U/ml Proteinase K (NEB, P8107S). After digestion, samples were trimmed again and washed in PBS (4 × 15 min). Hands-on time for this step is around 30 min.

DNase I digestion

Tissue-gel samples were equilibrated in DNase I buffer (750 μl) for 30 min and then incubated with RNase-free DNase I (2.7 Kunitz units/μl, 50 μl) (QIAGEN) in DNase I buffer (450 μl) at 37°C for 2 h. After DNase I digestion, samples were washed in PBS (4 × 15 min) to remove DNase I. A stronger DNase I treatment (2 × 2 h) was performed before the first round of FISH to completely digest nuclear DNA.

In situ Hybridization and HCR

For hybridization, tissue-gel samples were first equilibrated in hybridization buffer (500 μl) for 30 min at 37°C. Samples were then hybridized with probe sets (1 μM each probe, 3μl) in hybridization buffer (300 μl) overnight at 37°C. The next morning, samples were washed in probe wash buffer (2 × 15 min, then 3 × 30 min), followed by PBS (6 × 30 min) at 37°C. At this stage, hybridization chain reaction (HCR) can be carried out immediately. Alternatively, samples can be left in PBS overnight at room temperature for HCR the next day or stored in 4°C for longer storage. For HCR, samples were first equilibrated in amplification buffer (500 μl) for 30 min. Hairpins (3 μM) (conjugated to AF488, AF546, JF669) were heated and snap-cooled in a PCR thermocycler (95°C for 90 s then cooled at room temperature for 30 min). To initiate HCR, hairpins h1 and h2 were mixed and diluted to 30 nM in 300 μl fresh amplification buffer. Samples were then left in the dark for HCR reaction at room temperature for 3 h. After HCR, samples were first washed in 5 × SSCT (5 × SSC + 0.1% Tween) (2 × 30 min), then washed in 0.5 × SSCT (0.5 × SSC + 0.1% Tween) (2 × 30 min) and stored in 0.5 × SSC at 4°C before imaging. All HCR v3.0 probe and hairpin oligos were purchased from Molecular Instruments. *Gad1* HCR v3.0 probes used for assessing the detection efficiency of EASI-FISH were designed based on (Choi et al., 2018) and purchased from IDT.

Image acquisition, sample handling, and multiplexing

Although EASI-FISH samples can be imaged with traditional wide-field and confocal microscopes, bleaching and poor signal-to-noise ratio (SNR) would be problematic for thick tissue specimens. Thus, selective plane illumination microscopy is preferred. In this study, all samples were imaged on a Zeiss Lightsheet Z.1 microscope. A 20 × water-immersion objective (20x/1.0 W Plan-Apochromat Corr DIC M27 75mm, RI = 1.33) was used for imaging with 1 × zoom. Single-side illumination was used to reduce light exposure and imaging time. Images were collected at 0.23 × 0.23 μm pixel resolution (post-expansion) and 0.42 μm z step size with dual camera detection of two channels in each of two tracks: the 488 nm and 669 nm channels were acquired together, followed by the 405 nm and 546 nm channels. For large volume imaging, each image tile was 1920 × 1920 pixels (438.5 μm × 438.5 μm, post-expansion) in size with around 1500 z-slices, with overlap between tiles set to 8%. For imaging of the lateral hypothalamus samples, zona incerta, the fornix and optic tract were used to guide selection of field of view (FOV). 4 × 4 tiles were taken from the LHA tissue-gel sample (usually 2-3mm in x and y dimension). Imaging of sample volumes described above (0.8 × 0.8 × 0.3 mm pre-expansion, 2 × expanded) with 4 × 4 tiles (438.5 μm × 438.5 μm lateral dimension, ~1500 z-slices each tile) and 4 channels (in 2 tracks) took 2.5 hours.

Before imaging, samples were stained in PBS with 5 μg/ml DAPI (2 × 30 min). At this concentration and in the absence of DNA, DAPI stains the RNA in the cytoplasm. Samples were mounted to a Poly-L-Lysine coated 8mm glass coverslip (Harvard Apparatus) that was glued to a custom-made plastic holder (https://www.janelia.org/open-science/zeiss-lightsheet-z1-sample-holder) and imaged in PBS (Expansion factor: 2 ×). After image acquisition the probes and HCR products were removed using DNase I (see above) with samples attached to the holder. To remove samples from the holder for re-hybridization, they were incubated at room temperature in 10% dextran sulfate (500 μl) (Fisher Scientific) in PBS for 30 min. A paintbrush can be used to assist with gel removal. Each round of EASI-FISH includes DNase I digestion, *in situ* hybridization, HCR amplification and imaging, with 1-2 hours of hands-on time. Note that all units discussed in this paragraph regarding pixel size and image size are in post-expansion units (in contrast to our convention in the rest of the paper where scales are usually in pre-expansion units, unless otherwise noted).

Notes on Expansion factor, tissue thickness, and imaging conditions
Expansion factor

For the current application, EASI-FISH imaging was performed on a 2 × expanded specimen, which offered suitable tissue clearing and acceptable data size. By reducing salt concentration with 0.05 × SSC, EASI-FISH specimens can be expanded up to 3 × , which further improves optical resolution but also increases physical sample size, imaging time, and data size.

Tissue thickness

We have successfully used EASI-FISH in tissue slices up to 400 μm thick (pre-expansion). Further modification of the procedure (e.g., longer incubation times) may enable application of EASI-FISH in thicker tissue sections. However, substantially thicker samples would increase data size after expansion and may require the use of longer working distance, lower NA objective lenses, which would

reduce imaging resolution for spot detection. Larger continuous volumes can also be achieved by serial thick tissue sections, which may be aligned based on continuity of spatio-molecular domains.

Imaging condition

We found that HCR spots were susceptible to light-induced fragmentation, producing mobile spots of reduced brightness (wigglers). Wigglers were difficult to detect without time-resolved imaging and led to false positive spots outside of cell bodies. Carefully controlling the light dose (reduced laser power and exposure time) alleviated light-induced fragmentation of HCR spots. Anti-fade compounds, such as p-phenylenediamine (PPD) and DABCO also reduced spot fragmentation (Table S2 and Video S1). However, we noted that antifade compounds dramatically increased the bleaching rate of one of the commercially available HCR hairpin fluorophores (AF546) (Figure S1H).

Combining immunofluorescence (IF) detection with EASI-FISH

We extended the EASI-FISH method to allow for co-detection with immunofluorescence (IF). 100 μm tissue sections were used due to limitations of antibody penetration. As described above, C57BL/6J male mice (8 weeks old) were perfused with RNase-free PBS (15 ml) followed by ice-cold 4% paraformaldehyde (PFA) (50 ml). Brain coronal slices (100 μm) were sectioned and stored in 70% ethanol at 4°C. The LHA region was identified as described above and rehydrated in PBS. Immunofluorescence was performed by first permeabilizing with 0.5% Triton X-100 in PBS (15 min) at room temperature, followed by incubation in blocking buffer containing 1% RNase-free BSA (Thermo Fisher, AM2616) and RVC (ribonuclease inhibitor, NEB, S1402S) in PBS with 0.1% Triton X-100 (1 h) at room temperature. Tissue slices were then incubated with primary antibody (1:200 in blocking buffer, O/N) on a shaker at room temperature. Tissue slices were washed with PBST (0.1% Tween-20 in PBS) (4 × 30 min). Secondary antibody (1:500 in blocking buffer, O/N) was applied on a shaker at room temperature, followed by washes in PBST (0.1% Tween-20 in PBS) (4 × 30 min). After IF staining, RNA and protein (including antibodies) were anchored with MelphaX (1 mg/ml) and AcX (0.2 mg/ml) overnight at 37°C, embedded in gel, and proteinase K treated as described above, followed by probe hybridization and HCR. For IF combined with multi-round FISH, an HCR initiator-labeled-secondary antibody (Molecular Instruments) was used (antibody signal was amplified together with FISH). After the first-round imaging of IF and FISH, both FISH signal and amplified IF signal were removed with DNase I, followed by another round of FISH as described above. This method works well to preserve both IF and FISH signals as indicated by IF signals before and after EASI-FISH (Fig.S7D) and mRNA transcript counts with or without IF (Figure S7E).

For immunofluorescence detection of neuropeptides that localize primarily to distal processes (TRH, SST and NTS) (Dubé and Pelletier, 1979; Horjales-Araujo et al., 2014; Kahn et al., 1980), intracerebral colchicine injection (10 μg in 500 nl) (Brown et al., 2019) was performed on C57BL/6J male mouse (8 weeks old). Tissue was collected 48 h later and processed as described above.

Fos activity during water deprivation

C57BL/6J male mice (8 weeks old) were acclimated for three days before the start of the experiment. On the day of the experiment, animals either had *ad libitum* (AL) access to food and water (home-cage, n = 2) or the water bottle was removed for 24 h starting at 10 am (WD, n = 3). Precautions were taken to eliminate perturbation and noise in the animal room where this experiment was performed. Before and after the experiment, animal body weight and food intake were monitored. Mice were perfused the next morning, followed by brain dissection and sectioning as described above. Four rounds of EASI-FISH were performed on anatomically matched brain slices from these animals to allow for detection of 10 cell type markers relevant for identifying *Nts* cell types as well as anatomical parcellation markers and *Fos* (*Fos*, *Otp*, *Meis2*, *Nts*, *Slc17a6*, *Slc32a1*, *Tac1*, *Tac2*, *Gpr101*, *Hcrt*, *Sst*).

QUANTIFICATION AND STATISTICAL ANALYSIS

scRNA-Seq analysis and marker-gene selection for EASI-FISH

Analysis of the single-cell RNA sequencing data, including filtering, variable gene selection, dimensionality reduction and clustering was performed with Seurat (v2.3.4) (Butler et al., 2018; Satija et al., 2015) in R (v3.4.3). First, cell doublets/multiplets and low-quality cells were filtered based on the total number of detected genes (1,500-7,500), relative abundance of mitochondrial transcripts (percent.mito < 0.055) and number of unique molecular identifiers (nUMI) per cell (< 2×10^5) respectively. Genes expressed in less than 3 cells were also removed. The resulting dataset consisted of 1,507 cells and 17,535 genes. The filtered dataset was then log-normalized and scaled, while regressing out the effects of latent variables including nUMI, and percent.mito. Next, we performed principal component analysis (PCA) and used the first 31 principal components for downstream analysis. For clustering, we used the graph-based clustering approach implemented in Seurat, with the original Louvain algorithm and 10 iterations. Non-neuronal cells were identified by expression of non-neuronal markers (*Aqp4*, *Olig2*, *Opalin*, *Pdgfra*, *Ctss*, *Flt1*, etc.) and absence of neuronal markers (*Snap25*, *Syp*, *Tubb3*, *Map1b*, etc.) and then removed from the dataset before further analysis.

The remaining 1,425 cells were processed similarly to what was described above and 4 neuronal cell types were identified, whose preliminary identities were assigned based on unique expression of enriched genes: Group 1 (70%, 998 cells) consisted primarily of cells with high levels of the neuropeptide *Hcrt*, Group 2 (15%, 214 cells) was a heterogeneous population best characterized by common expression of *Sparcl1*, Group 3 (10%, 143 cells) contained a high percentage of cells strongly expressing *Nts*, and Group 4 (5%, 67) was defined by very high levels of *Pmch*. Assessment of differential gene expression between neuronal cell types was performed

using the FindAllMarkers function in Seurat (Wilcoxon rank sum test, logfc.threshold = 0.55, min.pct = 0.25), with p values adjusted based on the Bonferroni correction. The full list of enriched genes for each major neuronal subclass is provided in Table S4.

To identify marker-genes for EASI-FISH, we started with the list of differentially expressed genes as outlined in Table S4. We applied a series of selection criteria designed to allow classification of a maximum number of unique cell types using the fewest number of genes possible. As such, in addition to limiting our search to genes with an adjusted p value cutoff of at least 0.05 and an average log-fold change of 0.55 or over, we also specifically selected markers with as close to binary "on/off" expression patterns in the cell type of interest as possible, based on high percentage of marker-positive cells in the target population compared to low percentage of marker-positive cells outside the target population (displayed as pct.1 and pct.2 in Table S4, respectively). While not explicitly used as a limiting factor for the selection of marker-genes in this experiment, we found that a value of 0.4 for the ratio of minimum difference in the fraction of detection between the two groups to be an informative rubric for aiding in selection of marker-genes with close to binary characteristics. We also looked for genes that could further split the *Hcrt* and *Pmch* neurons as they are well known neuropeptide secreting neurons with important functions. Using these parameters, alongside manual inspection of the Allen brain ISH Data (Lein et al., 2007) for cross-validation, we settled on the following genes to represent the neuronal cell types of greatest value and highest confidence given the number of assessed cells in the scRNaseq dataset: *Hcrt*, *Calb2*, *Nts*, *Gpr83*, *Pmch*, *Cartpt*, *Tac2*. Additionally, *Slc17a6* and *Slc32a1* were included to specify excitatory and inhibitory neurons respectively and *Map1b* was used as a pan-LHA neuronal marker. Due to the over-representation of specific neuronal cell types (e.g *Hcrt* neurons) and under-representation of some neuronal cell types (e.g inhibitory neurons) in this dataset (likely due to bias during hand sorting), we chose to supplement the marker list with those identified in a recently published dataset (Mickelsen et al., 2019). Additional genes were selected for inclusion in an effort to represent a broader diversity of cell types in the LHA. The collected final list of marker-genes is in Table S5. We note that this is not the only combination of genes that could feasibly serve to represent these molecularly defined cell types.

Integration of scRNA-Seq datasets

To obtain a broader diversity of molecularly defined cell types in the LHA, we integrated our scRNA-Seq data with published LHA scRNA-Seq datasets (Mickelsen et al., 2019; Rossi et al., 2019). Processed gene count expression matrices for the two published datasets were directly downloaded from Gene Expression Omnibus (GEO). The Mickelsen et al. dataset (GSE125065) was obtained via the 10 × Genomics platform, data from both female and male mice were used. The Rossi et al. dataset (GSE130597) was obtained via the Drop-seq method (Macosko et al., 2015) and only control groups (male animals) were included. Before datasets integration, the three scRNA-Seq datasets were pre-processed independently to filter out non-neurons as well as cells of low quality. First, similar to described above, genes that were expressed in less than 5 cells and cells with fewer than 200 detected features were first removed. Then cells with unique gene counts over 7,500 or under 500 were removed to eliminate multiplets as well as cells with low quality. Next, gene expression in remaining cells were normalized to total expression and log-transformed. Then the top 2,000 highly variable features were selected from each dataset after variance-stabilizing transformation (vst) (Hafemeister and Satija, 2019). Z-score transformed data was used for dimension reduction and clustering analysis. Non-neuronal clusters were identified by expression of non-neuronal markers (e.g., *Aqp4*, *Olig2*, *Opalin*, *Pdgfra*, *Ctss*, *Flt1*, etc.) and absence of neuronal markers (*Snap25*, *Syp*, *Tubb3*, *Map1b*, etc.) and removed.

After pre-processing, 4,418 cells from the Mickelsen et al. dataset and 2,087 cells from Rossi et al. dataset were included for integration with our dataset (1,425 cells). Consensus features (genes = 13,613) across all three datasets were used for downstream analysis. Integration of multiple scRNA-Seq datasets was performed using Seurat (v3.2.0) (Stuart et al., 2019). Briefly, the top 2,000 highly variable features across three datasets were selected independently with the FindVariableFeatures function in Seurat and used to identify cells with common molecular features across datasets (aka. integration anchors) after canonical correlation analysis (CCA) based dimensionality reduction. These integration anchors were filtered (k.filter = 200) and ranked (k.score = 30) based on their shared overlap of mutual neighborhoods. This step is executed with the FindIntegrationAnchors function in Seurat. These selected anchor sets were used to compute a weighted integration vector (weighted by anchor pairs' similarity and anchor correspondence, k.weight = 100), which is used to transform the three datasets into a common space (merging our dataset and the Rossi et al. dataset into the Mickelsen et al. dataset) using Seurat's IntegrateData function.

The integrated dataset was then used as input for dimensionality reduction and clustering, as described above. Neurons were first separated into excitatory and inhibitory populations and then clustered. The Silhouette score, SC3 stability index and the Jaccard index distribution after bootstrapping were used to determine the optimal resolution and neighborhood size for clustering. The Silhouette score is a measure of how similar a cell is to its own cluster compared to other clusters and the silhouette score for each cell in a specified cluster was calculated as the Euclidean distance in PCA space using the CalculateSilhouette function in Seurat. The SC3 stability index is used to evaluate the optimal resolution parameter. It looks at how cells cluster over different resolutions. Clusters that share the same cells across resolutions get higher index. The SC3 stability index was calculated using the *clustree* package in R (Zappia and Oshlack, 2018). For bootstrap analysis, we randomly selected 80% of cells from the integrated dataset and performed dimensionality reduction and clustering. We then calculated the Jaccard index between the most similar new cluster and the original cluster. This procedure was repeated 100 times and the distribution of Jaccard index across clusters were plotted and used to evaluate cluster stability. Clusters with high stability show consistently high Jaccard similarity index with bootstrapping. This bootstrap analysis was performed using *sccusteval* package in R (Tang et al., 2020) with modifications. Based

on these evaluations, the following parameters were chosen for SNN based clustering of the *Slc17a6* and *Slc32a1* neurons: *Slc17a6* clusters, k.param = 20, resolution = 0.6; *Slc32a1* clusters, k.param = 60, resolution = 1.2.

To determine how well selected marker-genes could recover major cell types, we correlated the expression of selected marker-genes from individual neurons in scRNA-Seq to cluster average and assigned the cell type identity with the highest correlation. Comparing the true and predicted identity of each cell, we found that the proportion of neurons that can be correctly assigned in this way scaled linearly up to the selected 24 marker-genes, and further increase in the number of marker-genes did not substantially improve this (Figure S3F). Based on this analysis, we found that most scRNaseq clusters (29/34) can be identified by selected marker-genes (Figure S3G), with the remaining scRNaseq clusters either lacking highly selective markers (seq-e16, seq-i2, seq-i3, seq-i4) or excluded due to contamination with oligodendrocytes (seq-e4, based on *Opalin* expression).

EASI-FISH data processing summary

EASI-FISH analysis pipeline is critical for accurate and quantitative characterization of thick tissue, multiplex FISH datasets and includes image stitching, registration, FISH spot detection and 3D cell segmentation. Large volumetric EASI-FISH datasets collected on a Zeiss Lightsheet Z.1 microscope were saved as single multidimensional (multi-tile, multi-channel, z stack) CZI files. We also exported a metadata file (MVL format) that includes tile configurations for the stitching step, which were later converted into a JSON format.

Image Stitching

We used a previously developed flat-field correction and stitching package (Gao et al., 2019) that enabled rapid processing of 3D image tiles with Apache Spark-based high-performance computing environments (https://github.com/saalfeldlab/stitching-spark). The pipeline first automatically performs a flat-field correction for each tile to account for intensity variations across the lightsheet. The flat-field correction was based on corrected intensity distributions using regularized energy minimization (CIDRE) (Smith et al., 2015), where the flat-fields were calculated for each tile and channel independently and applied to each tile stack prior to stitching. It then derives the globally optimal translation for each tile that minimizes the sum of square distances to competing optimal pairwise translations estimated by phase-correlation (Preibisch et al., 2009) with the DAPI channel. The same transformation was applied to the other three image channels. Stitched image volumes were then exported and saved into N5 format. This flat-field correction and stitching pipeline enabled rapid and automated data processing. All flat-field correction, stitching, and data export were executed on HHMI Janelia's LSF computing cluster.

Image Registration

Because sample handling could cause small deformations and 3D shifts in field-of-view (FOV) during image acquisition, to correct for this, image volumes from each round were registered after stitching based on the DAPI channels. To enable rapid and robust data processing, we developed a registration package that combined random sample consensus (RANSAC) (Fischler and Bolles, 1981) based feature matching with non-symmetric, diffeomorphic image registration, called Bigstream (https://github.com/GFleishman/bigstream). First, a Difference of Gaussian (DoG) filter was applied to 8 × 8 × 4 downsampled (1.84 × 1.84 × 1.68 μm, post-expansion) image pairs and local maxima above a selected threshold were selected as features, matched with RANSAC and affine transformed. After applying this global affine transformation to 4 × 4 × 2 downsampled (0.92 × 0.92 × 0.84 μm, post-expansion) image volumes, the transformed image volumes were split into 256 × 256 × 256 pixel-chunks with 12.5% spatial overlaps along each boundary for further processing. Another round of feature selection and affine transformation was performed on each image chunk at 0.92 × 0.92 × 0.84 μm (post-expansion) scale, followed by deformable registration. For better integration with the rest of the image processing pipeline, we created an implementation in Python of the non-symmetric diffeomorphic registration algorithm from the Greedy software package (Yushkevich et al., 2016). This enables in-memory data sharing of objects produced by different steps of the pipeline and thus avoids data saving in intermediate steps. It also guarantees compatibility of object formats such as transforms. All registration steps on image chunks can be executed in parallel. The global affine, piecewise affine, and piecewise deformable transforms were composed to a single displacement vector field stored in N5 format. Both forward transform and inverse transforms were computed. The forward transform maps on-grid-positions from the fixed coordinate system to corresponding (potentially off-grid) locations in the moving coordinate system and is required for resampling moving image data to match fixed image data. The inverse transform maps on-grid-positions in the moving coordinate system to (potentially off-grid) locations in the fixed coordinate system and is required for moving explicit positional data (such as detected FISH spot coordinates) from the moving coordinate system to the fixed coordinate system. For assessing the registration accuracy, mean structural similarity index (SSIM) (Wang et al., 2004) between the moving and fixed image was computed.

FISH Spot Detection

Airlocalize (Lionnet et al., 2011) was adapted and optimized for spot detection in large image volumes. It implements a local background correction step that subtracts background fluorescence surrounding each spot that is associated with out-of-focus spots due to light-sheet excitation beam thickness. Briefly, for Airlocalize, images were first passed through a Difference of Gaussian (DoG) filter and global background subtracted. Next, a pre-detection step was performed that identifies local maxima above threshold within a 5-pixel radius, which was used to estimate the rough spot position. Then a local background correction was per-

formed based on background estimation around each spot and a 3D Gaussian fit was used to estimate the spot location as well as the spot fluorescence intensity. Airlocalize was originally written in MATLAB and works best for epifluorescence and confocal image data ~1 GB in size.

For large image data associated with the EASI-FISH data collection pipeline, we developed a high-throughput version to allow for distributed data access and processing. In order to integrate it with other parts of the EASI-FISH pipeline, we compiled Airlocalize (with minor modifications) into a Python package for distribution. Large image volumes were split into overlapping chunks (~1GB in size, 10% overlap in z and 5% overlap in x and y). Spot detection on each chunk was performed simultaneously in parallel. The detected spots from individual image chunks were combined, with repetitive detections in the overlap regions removed. Detected spot positions and intensities were exported as a single CSV file. All spot detection procedure was performed on full resolution images. Crosstalk correction was applied to images acquired in the red channel (AF546) to correct for bleed-through from the co-acquired DAPI channel before spot detection.

We called this Airlocalize-based high-throughput spot detection approach hAirlocalize. After spot detection from each EASI-FISH round, the inverse transformation matrix acquired from the registration step (see above) was applied to the spot point cloud to transform spots to fixed image coordinate. Spot detection and warping were confirmed by visual inspections.

3D Soma Segmentation

Most approaches for cell segmentation use thin tissue sections and an assumption of a single cell layer, which is often invalid, and prior approaches in thicker tissue lack accurate segmentation (Alon et al., 2021; Wang et al., 2018). To achieve automatic and accurate detection and segmentation of cells in 3D image volumes, we developed a deep learning-based segmentation method based on the previously published 3D *StarDist* approach (Weigert et al., 2020). In a first step, *StarDist* predicts for each pixel its cell center probability and its radial distances to the nearest cell borders. In a second step, it selects the center point of each cell and uses the radial distances at this point to create a star-convex polyhedra that approximates the cell outlines. While this approach provides high detection accuracy in situations with many crowded and dense objects, the resulting segmentation masks are imprecise for non-convex cell shapes which could lead to unwanted misallocation of FISH spots. To address this problem, we retained the center probability and distance prediction step from *StarDist*, but afterward aggregated pixel affinity maps from the densely predicted distances, an approach that we called *Starfinity*. The final segmentation masks were obtained by applying a watershed segmentation on the affinity maps while using the thresholded center probability as seeds. We trained such a *Starfinity* model with annotated cells from 3 different brain areas (LHA, CEA and cortex) using an iterative approach. Manual annotation was performed using Paintera (https://github.com/saalfeldlab/paintera). First, image stacks with a total of 248 intact cells were manually annotated at full resolution (0.23 × 0.23 × 0.42 μm) to train a preliminary *Starfinity* model. Then from predictions made from this preliminary model, we selected images where cells were mistakenly segmented and corrected the errors in these images manually. We then fed the corrected annotations (827 cells) to train the model. For both training and predictions, images were downsampled (4 × 4 × 2-fold) to increase the receptive field of the model, make the pixel resolution approximately isotropic, and reduce computational demands. The training and prediction were based on the cytosolic DAPI channel. The trained model gave a segmentation accuracy of 93%, with 4% over-segmentation errors, 1% under-segmentation errors and another 2% contaminated by neighboring cells.

The *Starfinity* segmentation result then underwent a post hoc, semi-automated correction for over-segmentation errors. First, we identified the over-segmented ROI pairs based on high gene expression correlation between neighboring ROIs (1. Pearson Correlation Coefficient greater than 0.998, this cutoff was determined by maximum Youden's index calculated from the ROC analysis of manually inspected data; 2. the centroid positions of the selected ROI pair were less than 23 μm (pre-expansion) apart and the two ROIs were touching/connected). Over-segmentation errors usually occur with the primary dendrite (small fragments) separated from its soma (usually average sized) (Figure S2E), therefore we used ROI size as another filter. To eliminate false merges between neuronal processes and between neuronal process and non-neurons (which are usually quite small), we further filtered out ROI pairs that are smaller than 600 μm^3 in size. To avoid merging two neighboring neurons with highly correlated gene expression, for ROI pairs greater than 1500 μm^3 in size (average soma size in this region) and ROIs that had more than one matches, we first ranked the corresponding ROI pairs based on their correlation coefficient and then inspected them manually (these comprises < 15% of all flagged ROI pairs). Finally, identified ROI pairs were computationally merged into one single ROI. This method was estimated to eliminate 62% over-segmentation errors, bringing down the total over-segmentation error to less than 2%.

End-to-end Analysis Pipeline

To enable easy adoption of EASI-FISH, we built a self-contained, highly distributed, and platform agnostic computational pipeline (https://doi.org/10.5281/zenodo.5652239, for active updates: https://github.com/JaneliaSciComp/multifish) for analyzing EASI-FISH data. The pipeline was built using Nextflow (Di Tommaso et al., 2017), with all software containerized, which makes it portable and reproducible. The pipeline has been tested and proven to work on multiple platforms, including stand-alone workstations, the IBM Load Sharing Facility (LSF) computing cluster and the cloud platform, Amazon Web Services (AWS). It could also be adapted to execute on other batch schedulers (such as SLURM) and cloud platforms.

The pipeline is also highly flexible. We provide an end-to-end pipeline for EASI-FISH users to analyze their imaging data and directly translate over 10 terabytes of image data into spot counts per cell. The EASI-FISH steps (stitching, registration, segmentation, spot detection, etc.) in the pipeline are modularized, allowing flexibility for users to select only the necessary modules for their specific

	image scale	Stand-alone workstation (8 h total)	LSF cluster (3 h total)
Stitching	full resolution	2.5 h	1.5 h
Registration	4 × 4 × 2 or 8 × 8 × 4 downsampled	0.5 h	1.5 h
Spot detection	full resolution	7.5 h (2.5 h/channel)	1.5 h
Segmentation	4 × 4 × 2 downsampled	0.2 h	0.2 h
Post-processing	–	0.05 h	0.05 h

application. Furthermore, each module is containerized using Docker, and can therefore be easily substituted with a different implementation or algorithm.

The pipeline is openly available and includes extensive documentation and automatic download of example datasets for push-button replication. Monitoring of task execution and resource utilization is available using the Nextflow Tower web UI.

For analysis of an example dataset (35GB, 4 channels, 600 z-slices for each EASI-FISH round), we provide the processing time below for each step on a stand-alone workstation (128GB RAM, 40 cores) and the IBM LSF computing cluster. For data that is larger than 100GB, we recommend usage of cluster or cloud platforms for improved data parallelization and speed.

Although this breakdown shows the time for each step, it is important to note that, once the stitching step is finished, the subsequent steps (registration, segmentation, and spot detection) can be performed simultaneously in parallel. The pipeline can be used to analyze a complete multi-round EASI-FISH experiment all at once. Users can also choose to interleave the data collection and analysis with minimal modification. This way, the time lag between data acquisition and analysis will be negligible and users can start running through the pipeline once two rounds of EASI-FISH data have been acquired.

Visualization
The stitched and registered volumetric image data was visualized with Fiji plugin N5 Viewer (https://github.com/saalfeldlab/n5-viewer) based on BigDataViewer (Pietzsch et al., 2015) for interactive browsing of multichannel multiscale N5 datasets. More flexible visualization options, including overlaying with segmentation masks, inspection of detected spots, ROIs and spots queries were performed using Napari (Napari contributors, 2019) with custom scripting in Python (https://zenodo.org/record/5658557).

Post-processing
With the segmentation masks from the DAPI channel and spots extracted from FISH channels, extracted spots from each FISH channel were then assigned to individual ROIs to obtain a spot count for each FISH channel (gene). We also performed the following steps for improved data quality, all code available at https://zenodo.org/record/5658557.

Filter ROIs
Cells on the edge of the field of view that were only partially captured in 3D or failed to be captured in one or more EASI-FISH image rounds were removed from the analysis. Additionally, we observed autofluorescence in the red channel on the surface of the specimen in later rounds (> 7) and excluded part of the specimen (~30-50μm pre-expansion, from the top surface along the z axis) from the downstream analysis. After this procedure, 77% of detected ROIs were used for the following analysis (Table S6).

Lipofuscin spots
We observed autofluorescent punctate signals in the tissue-gel samples that are likely due to lipofuscin. Lipofuscin puncta are lysosomal storage bodies, and their presence could lead to false positive spot detection. We took advantage of the fact that lipofuscin has a broad excitation/emission spectrum and can therefore be identified by signal colocalization in more than one channels. In the current application, the 488 nm and 669 nm channels were acquired at the same time and were used to identify lipofuscin spots by signal colocalization. We identified spots whose centroid positions were within a radius of 3 pixels (345 nm, pre-expansion) between the 488 nm and 647 nm channels and subtracted these spots for all FISH channels. For genes with high expression levels (spot count exceeding 200 per cell, detected in round 4 to round 10), we chose to use the median lipofuscin spot counts for that cell across all rounds to avoid subtraction of real spots due to high spot density in one or both channels.

Dense spots
For cells with highly expressed genes, the integrated fluorescence intensity was used instead to estimate spot counts. First, we analyzed the distribution of spot fluorescence intensity from the hAirlocalize output for each gene. The spot intensity followed a right-skewed distribution. Outliers with high intensity values were likely detection of multiple spots in cells with high expression. The mode of the distribution was used as an estimate for single spot intensity for any given gene. The spot count in any given cell was calculated by dividing the sum of fluorescence intensity in that cell with the estimated single spot intensity. We compared the estimated spot counts with hAirlocalize spot counts for genes that showed low or medium expression levels (less than 200 spots/cell) in the majority of cells and observed that the two measurements were highly correlated and comparable in the low to medium spot density range. *Slc32a1*, for example (Figure S2F), shows a least-squares linear regression fitting to the data indicating a slope of 1.03 and R^2 of 0.93. Because intensity-estimated spot counts can be sensitive to background fluorescence in cells with low gene expression, we chose to use hAirlocalize measurement in these cells. Based on scRNA-Seq, we found that gene expression

variability between cells goes up with mean expression level (Figure S2G), such that spots for a given gene may be resolvable in some cells but not others. Therefore, we chose a spot density threshold to determine whether to use hAirlocalize spot counts or intensity-estimated spot counts for a given gene in each cell. For cells with spot density higher than 0.01spot/voxel (corresponding to spot separation on average ~1.3 μm apart), we applied the intensity-estimated spot counts.

Neuronal morphological analysis

Taking advantage of the accurate 3D segmentation, we measured the morphological properties of neuronal cell body (soma) segmentation masks using the scikit-image implementation of *regionprops*. The spatial position of each neuron was defined as the centroid of the segmentation mask. Solidity was a measure of the overall concavity and was defined as the neuronal cell body volume divided by that of its convex hull. The eccentricity was defined as the length ratio between the minor axis and the major axis (Figure S7I). To calculate fraction of neurons with the most extreme somatic shapes in each cluster, we selected for cells with low eccentricity (elongated, < 0.7) and solidity (less convex, < 0.7) (Figure 7D). Other measurements that were not discussed here, but have been performed include major axis length, minor axis length, orientation, extent and whether a neuron cell body is in direct contact with a non-neuron cell body.

All measurements reported here are in pre-expansion units in μm, unless otherwise noted.

Evaluation of EASI-FISH detection efficiency, selectivity, and sensitivity

To assess transcript detection efficiency, *Gad1* mRNA transcripts were labeled with interleaved probe sets conjugated to two different fluorophores (10 probes in each set, conjugated to AF488 or AF546). Assuming independence of probe binding between the two probe-sets, the probability of AF488 channel detection within the whole population of *Gad1* transcripts should be equivalent to that within the AF546-positive subpopulation. We measured this detection probability for the AF488 channel as 66% ± 2%. Similarly, the detection probability for the AF546 channel (within the subpopulation of AF488-positive spots) was 75% ± 6%. Thus, the false negative rates for the two probe sets are 34% ± 2% and 25% ± 6%, respectively. Again, assuming probe-set binding independence, the false negative rate for both probe-sets together is the product of both probe-sets' individual false negative rates, or 8% ± 2%, corresponding to a detection rate of 92% ± 2% for the full set of 20 probes.

We examined two sources of false-positive spot assignments: non-specific probe binding and amplification and true-positive transcript detections that may not originate within the segmented cell body. Because we used HCR v3.0, non-specific probe binding and amplification was low (1 per 3000 μm³, 1011 spots detected in tissue volume 219.38 × 219.38 × 61.47 μm³ in size), which we determined by applying FISH probes to samples in the absence of the target gene (GFP probes in wild-type mice). For analysis of false positive background spots, we selected genes that were known to be mutually exclusive. For example, in the LHA, cells expressing *Tacr3* have undetectable levels of *Pdyn* and vice versa based on scRNA-Seq (Figure S1L) and *Slc17a7* and *Gad1* expression in the cortex.

To test cell-level detection sensitivity of the EASI-FISH method, we focused on a single neuronal cell type from the scRNA-Seq dataset, the *Pmch* neurons. We identified genes that are selectively expressed in this population at various levels (Table S4). *Klhl13* (UMI$_{mean}$ = 48) and *Igf1* (UMI$_{mean}$ = 15) were selected specifically due to their unique expression in most of *Pmch* neurons (high pct.1) and relative absence in other cell types (low pct.2), as well as their low absolute levels of gene expression as indicated by UMI counts.

FISH cluster analysis

A total of ~86,000 cells were identified from three specimens. Incomplete cells on the tissue surface and cells that failed to be included in all imaging rounds were removed from downstream analysis, leaving 66,488 (77%) cells (Table S6). Neurons were then identified using a two-components Gaussian Mixture Model (GMM) based on spot count of the LHA pan-neuronal marker, *Map1b* with probability greater than 0.7. 36,423 (55%) cells were identified as neurons in this way. Clustering analysis was performed on z-score normalized spot counts of 24 marker-genes. Unlike the scRNA-Seq analysis, no logarithmic transformation was applied to minimize the weight of false positive spot detection. As described above, principal component analysis (PCA) was performed, and graph-based SNN clustering analysis implemented in Seurat was used for an initial clustering using all marker-genes. Like the scRNA-Seq analysis, this separated neurons into *Slc17a6* and *Slc32a1* populations. Subsequent clustering was performed on *Slc17a6* and *Slc32a1* populations separately. For *Slc17a6* population, *Map1b* and *Slc17a6* were excluded from the clustering analysis as they were not considered variables. For *Slc32a1* population, non-variable genes (widely expressed: *Map1b*, *Slc32a1*; not expressed: *Bdnf*, *Cartpt*, *Trh*, *Pmch*, *Hcrt*) were excluded from the clustering analysis. The clusters were reordered based on the cluster-average of total spot counts from high to low.

As above, the Silhouette score, SC3 stability index and Jaccard index distribution as calculated by bootstrapping 80% of the data 100 times were used to choose clustering parameters: the nearest neighborhood size k and resolution to maximize the number of stable clusters (k parameter chosen for *Slc17a6* cells: 20, k parameter chosen for *Slc32a1* cells: 10). Cell types with average total spot counts from all genes below the 20th percentile of spot counts in all neurons were aggregated and assigned as poorly classified clusters due to lack of marker-gene detection. Using this cutoff, 80% of neurons from the *Slc17a6* population (Ex-1 to Ex-24) and 74.5% of neurons from the *Slc32a1* population (Inh-1 to Inh-22) were classified. tSNE was used for visualization (Figures 3C and 3D) with perplexity of 50 for *Slc17a6*⁺ population and 40 for *Slc32a1*⁺ population. Most molecularly defined clusters were detected in all three animals (Figures S4A and S4B), except Inh-1, which was located medially and was only captured in two out of three samples because of slight spatial differences at the edge of the field of view (Data S1). Gene expression within identified molecularly

defined clusters was highly correlated across biological replicates (Figure S4E), indicating no batch effects for EASI-FISH measurements and analysis

FISH to scRNA-Seq data mapping

To map FISH data to scRNA-Seq clusters, we first log-transformed the scRNA-Seq data and calculated the average marker-gene expression in each scRNA-Seq cluster. Marker-gene expression measured via EASI-FISH was also log-transformed, thresholded (> 30 spot counts), scaled and then correlated to scRNA-Seq clusters. Neurons were assigned to the scRNA-Seq cluster with the highest correlation. Neurons with low correlation coefficients (< 0.5) to all scRNA-Seq clusters were not assigned. This analysis was performed using *mfishtools* package in R with modifications. We then calculated the fraction of neurons from each EASI-FISH cluster that mapped to each scRNA-Seq cluster and assigned correspondence based on the fraction of mapping. We were able to find correspondence in 19 out of the 29 scRNaseq clusters that we expected to identify in the LHA. Clusters that did not show strong mapping are either rare clusters (seq-e1, seq-e3, seq-e10, seq-e13, seq-e17) or not present in the region of the LHA that we profiled with EASI-FISH, based on marker-gene distribution from Allen Brain ISH Data (seq-e11, seq-e12, seq-i12 and seq-i14).

LHA boundaries and neighboring brain regions

The zona incerta (ZI) is characterized by a high density of inhibitory neurons with small and regular cell bodies (Kawana and Watanabe, 1981; Kolmac and Mitrofanis, 1999). Therefore, we drew boundaries between the ZI and the LHA based on neuronal density and morphology, as well as *Slc17a6* and *Slc32a1* expression. The entopeduncular nucleus (EP) is enriched for glutamate/GABA co-releasing somatostatin neurons (Wallace et al., 2017), which were used to identify boundaries for EP. The fornix was identified based on its location, circular profile, and lack of neuron somas.

Marker-gene spatial expression analysis with PCA

To identify highly correlated spatial patterns in the LHA and marker-genes with greatest spatial variation, we decomposed the 24 marker-gene expression patterns into principal components (PC). Specifically, the gene expression data (spot counts) were first z-score normalized and mapped in 3D to reconstruct the expression patterns. 24 such images with each one representing the expression pattern of one marker-gene were then smoothed and decomposed into principal components (PCs) that corresponded to spatial variations in gene expression patterns. The first 4 PCs accounted for 66% of the data variation. Genes that were selected to parcellate the LHA (*Meis2*, *Slc32a1*, *Slc17a6*, *Otp* and *Hcrt*) were among the ones with the most positive or negative weight (account for greatest variations) on the first 4 PCs.

Spatial enrichment of selected marker-genes

To compute the selective spatial enrichment for *Otp/Meis2*, we first counted the number of Otp^+ and $Meis2^+$ neurons within a 50 μm radius neighborhood of any given neuron. The selectivity index for *Otp/Meis2* is calculated as $(Otp^+_{num} - Meis2^+_{num}) / (Otp^+_{num} + Meis2^+_{num})$. For *Slc17a6/Slc32a1*, the same procedure was performed based on the number of $Slc17a6^+$ and $Slc32a1^+$ neurons in the neighborhood, except for cells that co-expressed *Slc17a6* and *Slc32a1*, which were excluded from the analysis. This procedure was advantageous compared to an image smoothing filter as it preserved neuronal density information, which was useful for segmentation later.

Anatomical segmentation of the LHA

The initial segmentation of the LHA was based on expression of two pairs of genes *Otp/Meis2* and *Slc17a6/Slc32a1*. We first classified neurons into 4 classes based on their spatial enrichment for these four genes (Class 1: *Slc17a6/Otp*; Class 2: *Slc32a1/Otp*; Class 3: *Slc17a6/Meis2*; Class 4: *Slc32a1/Meis2*) and then used Gaussian Mixture Models (n components = 50) implemented in scikit-learn to generate the 3D segmentation (1 μm isotropic voxel resolution). This segmentation was performed separately on each LHA volume. Then, the segmentation masks from LHA1 and LHA2 were aligned to LHA3 using the rigid registration implemented in Greedy (Yushkevich et al., 2016). The fiducial landmarks (ZI, EP and fornix) and spatial distribution of marker-genes from three animals were used to cross validate the registration. Simultaneous truth and performance level estimation (STAPLE) (Warfield et al., 2004) implemented in SimpleITK was used to generate a unified atlas with a probability threshold of more than 0.9. Disconnected regions that were from the same molecularly defined class (1-4) were further split and assigned unique ROIs. Small ROIs (< 0.5×10^6 μm^3) were removed and the segmentation boundaries were smoothed with a Gaussian filter (sigma = 30), with ZI and EP identified as distinct regions.

Regional parcellation of the visual cortex with STARMap data

To determine whether this computational parcellation method could be generalized to other brain regions, we applied this method to a part of the mouse primary visual cortex from the STARmap thick tissue (100μm) dataset, a known brain structure (Wang et al., 2018). As described above, marker-genes were first z-score normalized and mapped in 3D to reconstruct the expression patterns. Images containing expression patterns of 28 marker-genes were decomposed into principal components (PCs). The eigen-images from the top 4 PCs captured the laminar structures in the visual cortex (Figure S5G) and explained 87% of the variance in the dataset (Figure S5H). We

selected genes that have the largest magnitude of weight on each principal component as candidates for structural parcellation. Based on this criterion, *Pcp4, Mbp*, *Cux2* and *Plcxd2* were chosen for parcellation. As with the LHA parcellation, these four genes were used to parcellate the imaged tissue volume into 6 cortical layers (L1, L2/3, L4, L5 and L6), the white matter and the subcortical area.

Neighborhood cell type composition

The spatial distribution of molecularly defined cell types was analyzed with neighborhood complexity and purity analysis, similar to previously described (Moffitt et al., 2018). Briefly, the neighborhood complexity computes the number of distinct molecularly defined cell types within a 50-μm-radius neighborhood. The neighborhood purity was defined as the fraction of neurons within the 50-μm-radius neighborhood that were part of the most abundant molecularly defined cell type. We found that in the LHA a 50-μm-radius neighborhood consists of 16 molecularly defined cell types on average (neighborhood complexity), with on average 27% of neurons in this radius corresponding to the most abundant cluster (neighborhood purity).

Average distance to nearest neighbor (ANN) analysis

The nearest neighbor of each neuron was queried using kdtree (scipy) and the average distance to the nearest neighbor in each molecularly defined cell type was defined as the mean distance of neurons in a cell type to its nearest neighbor. Note that neuron number in each molecularly defined cell type is different, therefore it is not a one-to-one relationship (e.g., multiple neurons can have the same nearest neighbor), and the ANN between cell type A and cell type B can be different depending on which cell type was queried. For all ANN graphs generated in this study, rows represent molecularly defined cell types that were used to look for nearest neighbor and columns represent the corresponding molecularly defined cell types from which the nearest neighbor was identified.

Spatial distribution of molecularly defined cell types

First, to determine the spatial distribution of molecularly defined cell types (whether they were clustered, dispersed or uniformly distributed), we compared the distribution of each molecularly defined cell type with a CSR (complete spatial randomness) process and performed a Monte Carlo test of CSR (Cressie, 2015; Waller and Crawford, 2004). For a molecularly defined cell type with N neurons, we simulated the CSR process by randomly sampling N neurons from the entire dataset (regardless of their identity) and computed their average distance to nearest neighbors (ANN_{CSR}). This process was repeated 1,000 times to generate a distribution of the average distance to nearest neighbor under CSR (ANN_{CSR}). The number of randomly sampled cells was matched to that in each molecularly defined cell type. The ANN from each molecularly defined cell types (ANN_{Mol}) was calculated and compared to the CSR distribution to calculate the p value. We found that all (48/48) of the molecularly defined cell types were spatially clustered ($ANN_{Mol} < ANN_{CSR}$, $p < 0.05$) compared to a CSR process (Table S1).

To determine the statistical significance of subregion enrichment of molecularly defined cell types, we performed a permutation test. We shuffled the positions of all neurons 1,000 times to generate random spatial distributions. For each molecularly defined cell type, the observed fraction of neurons enriched in each subregion was compared to the distribution of such fractions generated by shuffling to calculate the p values and effect size (Figure 4E).

Spatial overlap of molecularly defined cell types

To determine the spatial overlap of cells from molecularly defined clusters, we first calculated the local neuronal density of each molecularly defined cell type using kernel density estimation (kde) (scipy). To eliminate outliers, we threshold the density distribution above 95th percentiles and generated a binarized segmentation mask for each cell type to represent its regional enrichment. Binarized segmentation masks from all molecularly defined cell types were combined to generate a pixelated image with 'intensity values' at each voxel (10μm × 10μm × 10μm) representing the number of overlapping cell types occupying that voxel. This overlaid image was used to further subdivide the LHAfl and LHAs-db subregions. The segmentation masks generated for each molecularly defined neuronal type was used to calculate the spatial overlap between pairs of molecularly defined clusters. The fraction of spatial overlap between molecularly defined cell types were calculated as the number of overlapping voxels that selected cell types occupy divided by the total number of voxels occupied by either cell type.

Prediction of spatial position with gene expression

To determine whether neuronal spatial position could be predicted based on combinatorial expression of marker-genes, we trained a multi-output Random Forest regressor to predict the spatial positions of neurons. First, gene expression data were normalized (z-score) to a normal distribution with similar scale ranges. We trained the random forest regressor using the z-score normalized expression of 24 marker-genes as input and the x, y, z positions (pre-expansion, in μm) as output. The prediction performance was evaluated using 10-fold cross-validation and random permutation cross-validation (Shuffle & Split). Both coefficient of determination (R^2) and the explained variance score were calculated, and they were identical when rounded to 2 decimal places. To evaluate the importance of individual genes and their combinations in predicting neuronal position, models were trained and tested by removing one or combinations of features (genes) and as described above, cross validation was used to evaluate the predictive power of the model based on R^2 score. To evaluate statistical significance, we shuffled the relationship between the gene expression and neuronal position 1,000 times and compared the prediction accuracy of our model to the distribution of prediction accuracy from shuffled data to calculate the p value.

Cross-correlation analysis between biological replicates

To determine whether findings presented here were reproducible across animals, we performed cross-correlation analysis on (1) gene expression of molecularly defined cell types (Figure S4E), (2) marker-gene spatial distribution (Figure S5M), and (3) molecularly defined cell type spatial distribution (Figure S6B) between biological replicates. For molecularly defined cell type gene expression, average marker-gene expressions in any given molecularly defined cell type were used to calculate the correlation coefficient. For the spatial distribution of marker-genes and molecularly defined cell types, correlation analyses were performed on tissue volumes after rigid alignment. For gene expression spatial distribution, gene expression was binarized and cells with spot count greater than 30 were defined as positive for any given gene. The image volumes from different animals were discretized into 100μm × 100μm × 100μm bins and the number of positive cells for any given gene was used to generate the correlation coefficient. For molecularly defined cell type spatial distribution, instead of the number of positive cells for any given gene, the number of cells from any given molecularly defined cluster in each bin was used for cross-correlation. All correlation analyses were performed by comparing LHA1 and LHA2 to LHA3 and p value < 0.05 were reported here.

Vizgen MERFISH data analysis

One brain slice (S2R2) from the (Vizgen, 2021) MERFISH mouse brain receptor map dataset (Vizgen Data Release V1.0. May 2021) was downloaded and the LHA was identified and cropped out based on boundaries from neighboring brain regions (ZI, EP and the optic tract) and the fornix. This sample was a single thin section (10 μm) corresponding to a posterior portion of our reconstructed volume. For analysis, neurons were separated from non-neurons and divided into excitatory and inhibitory neurons based on relative expression of *Slc17a6* and *Slc32a1*. Similar to the EASI-FISH analysis that was described above, 483 marker-genes were z-score normalized and mapped to reconstruct spatial expression patterns. Images containing expression patterns of 483 marker-genes were decomposed into principal components (PCs). The eigen-images from the top 3 PCs of excitatory and inhibitory neurons were displayed, which explained ~70% of variance in the data. LHA from both hemispheres were analyzed and the right hemisphere was shown in Figure S6K.

Connectivity analysis

To determine whether afferent axons projecting to or passing through the LHA follow the regional parcellation, we looked at dataset collected from the Allen Mouse Brain Connectivity database. We utilized the spatial search function from the Allen Brain Atlas API and selected for experiments meeting the following criteria: 1) rAAV-EGFP tracing in wild-type animals for broad neuronal cell type tracing; 2) density of projection signal (ratio of thresholded fluorescence pixel over all pixels in the structure) greater than 0.1 in the LHA (6350, 5850, 6850 in the reference space was chosen as the LHA target location). 148 experiments met these criteria. Affine alignment was applied to transform the LHA parcellation map to the reference atlas based on fiducial landmarks (ZI, EP and fornix). Mean signal intensities in each LHA subregion were extracted from the 148 experiments. For multiple experiments from the same injection site, cross-correlation analysis was performed to remove outlier experiments, which are most likely due to injection site variability. Experiments with the injection site in the LHA or its neighboring brain regions (AHN, VMH) were removed because there could be injection site contamination in the LHA. This allowed us to examine a total of 64 different brain regions that project to or pass through the selected LHA target location and determine their relative signal intensities in LHA subregions.

### Fos activity in molecularly defined cell types

To identify molecularly defined cell types based on marker-gene expression, we performed cross-correlation analysis based on selected marker-gene expression and assigned the cell type identity with the highest correlation to each neuron. In addition to the *Nts* cell types that we initially set out to identify, the selected marker-genes can also recover additional clusters that were defined from the full 24-marker EASI-FISH LHA dataset: excitatory cell types (Ex-2, Ex-6, Ex-14, Ex-17, Ex-20, Ex-23) and inhibitory cell types (Inh-1, Inh-3, Inh-4, Inh-6, Inh-10, Inh-12, Inh-16, Inh-17). For cross-correlation, marker-gene spot counts were first z-score normalized before computing the Pearson's correlation coefficient between each neuron and the molecular cluster averages. Neurons were assigned to the cell type identity with the highest correlation coefficient and only neurons with correlation greater than 0.8 were included in downstream analysis to eliminate ambiguous assignment. To identify *Fos* positive neurons, *Fos* spot counts were z-score normalized and neurons with z-score higher than 2.0 (p < 0.05) were identified as *Fos* positive. Neurons were combined from the *ad libitum* food and water group (AL) and water-deprived group (WD) and Chi-square test for independence was performed to compare the statistical significance of the enrichment of *Fos*-positive neurons in molecular cell types from the two groups (Figure S7M). P values were adjusted for multiple comparison with the false discovery rate method.

Statistics and figure scales

Statistical analyses were performed in Python, R or Prism GraphPad and described in methods, figure legends and Table S1.

All scales (scale bar and scale arrows) in the text, figures, and figure legends were converted to pre-expansion units in μm, unless noted otherwise.

Supplemental figures
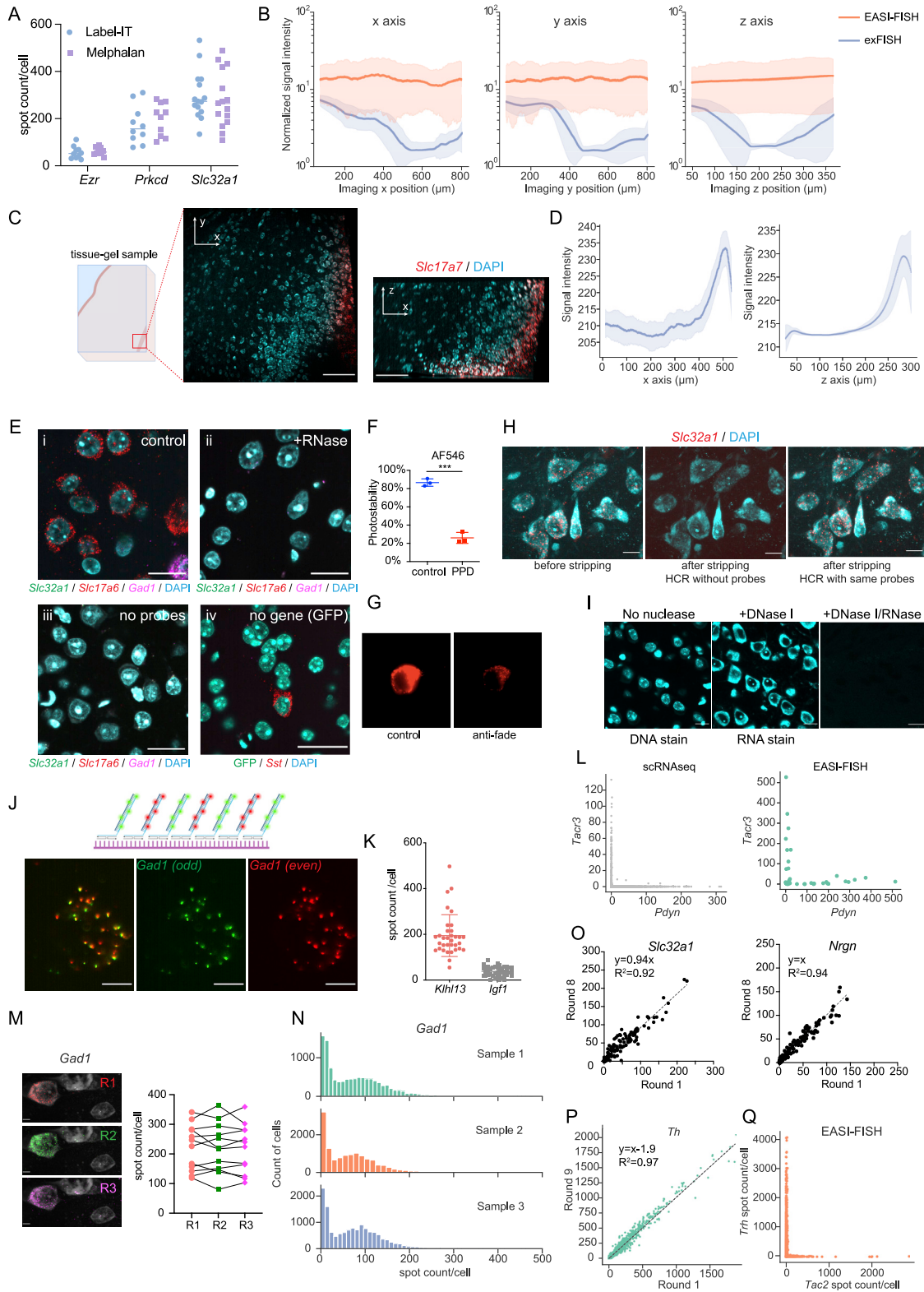
Figure S1. Optimization of EASI-FISH protocol, related to Figure 1

(A) *Ezr*, *Prkcd* and *Slc32a1* spot counts detected with different RNA anchoring chemicals (Label-IT or Melphalan). Statistics: Table S1.

(B) EASI-FISH was optimized to reduce light scattering in large tissue specimens (0.8 mm × 0.8 mm × 0.3 mm), as demonstrated by stable signal intensity across the x, y, and z axes when compared to the original exFISH protocol (Chen et al., 2016), which shows a substantial decay across all dimensions. Single-side illumination during imaging. Shading denotes standard deviation (SD) around mean.

(C) Representative images showing RNAscope detection of *Slc17a7* in 300μm EASI-FISH cortical sample. A small field of view on the right edge of the tissue-gel sample is shown here. Left: schematic showing where the representative image volume (middle and right) was taken, middle: single optical plane from the image volume, right: side-view of the image volume. Scale bar: 100μm.

(D) Signal intensity quantification of the image volume shown in (C). Shading denotes standard deviation (SD) around mean.

(E) EASI-FISH is optimized for high specificity, as indicated by minimal spot detection in the absence of RNA (ii), low non-specific binding/HCR initialization of hairpins in the absence of probes (iii), and low spot detection in the absence of the target gene GFP (iv, *Sst* detection as positive control) as compared to control (i). Scale bar: 25 μm.

(F) Quantification and (G) representative image showing rapid photobleaching of Alexa Fluor 546 (AF546) in the presence of anti-fade (PPD). ***p < 0.001. Statistics: Table S1.

(H) DNase I can be used to remove FISH signals and allows for re-probing without signal loss. Top: before DNase I stripping, middle: after DNase I stripping, bottom: re-probed after DNase I stripping, shown as axial projection. Scale bar: 10 μm.

(I) DAPI staining of DNA (left) and RNA (middle) in EASI-FISH samples. Note that cells with low RNA content can have very weak DAPI RNA staining. Right panel: no DAPI staining in tissue in the absence of oligonucleotides (DNase and RNase treatment). Scale bar: 25 μm.

(J) EASI-FISH detection efficiency assessed by co-localization of interleaved HCR probes targeting the same gene, *Gad1*. Scale bar: 5 μm.

(K) EASI-FISH false negative detection as indicated by co-detection of low expressors *Klhl13* and *Igf1* in *Pmch* neurons from the LHA.

(L) EASI-FISH false positive detection in mutually exclusive genes, *Pdyn* and *Tacr3*. Left: UMIs of *Pdyn* and *Tacr3* in LHA neurons detected with scRNA-Seq, Right: Spot counts of *Pdyn* and *Tacr3* in neurons detected with EASI-FISH.

(M) Representative images (left) and quantification (right) showing reliable detection of the same gene (*Gad1*) across 3 rounds of FISH in cortex. Scale bar: 5 μm.

(N) Reliable spot detection of the same gene (*Gad1*) in biological replicates.

(O) Spot counts of *Slc32a1* and *Nrgn* in the same neurons from EASI-FISH round 1 and round 8. Statistics: Table S1.

(P) RNA stability and (Q) false positive detection in the LHA samples used for EASI-FISH demonstration. Dotted lines in (O) and (P): linear regression fit (equation and R^2). Statistics: Table S1.
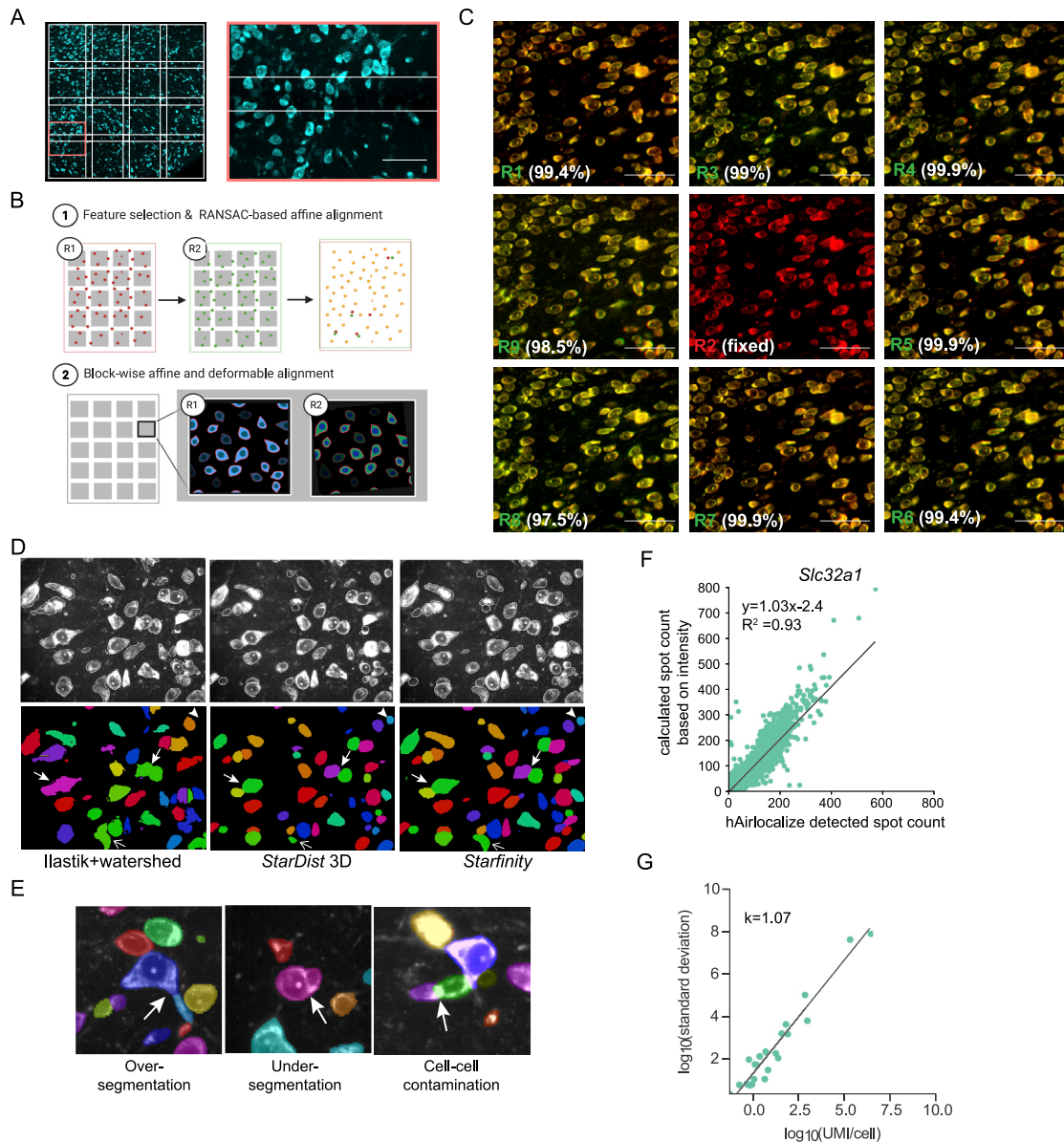
Figure S2. EASI-FISH data processing pipeline, related to Figure 2

(A) Representative multi-tile images after stitching. Left: Representative single image plane after tile-tile stitching. White grids indicate the dimensions of tile overlaps. Right: zoom-in of region highlighted in red box (left). Scale bar: 50 μm.

(B) Schematic illustration (cartoon images) of the round-to-round registration pipeline.

(C) Representative images showing registration of 9 rounds of EASI-FISH images based on RNA-staining (cytoDAPI). Round 2 was used as the fixed image round (shown in red, all other rounds in green). Structural similarity to fixed round is shown in parentheses. Scale bar: 50 μm.

(D) Comparison of segmentation methods: Ilastik in combination with watershed (left), *StarDist* (middle) and *Starfinity* (right). Representative segmentation errors are highlighted by white arrows. Thick arrows indicate under-segmentation errors (cell-cell merges); Arrowheads indicate under-detection errors; Thin arrows indicate *StarDist*-specific errors due to star-convexity constraints.

(E) Segmentation errors in *Starfinity*, over-segmentation (one cell was split into two labels), under-segmentation (two cells assigned the same label), contamination (segmentation boundaries not properly drawn), highlighted with white arrows.

(F) Spot count comparison between hAirlocalize and integrated intensity-based estimation in low-expressor-cells (*Slc32a1*). Black line: linear regression fit (equation and R^2). Statistics: Table S1.

(G) 24 marker-gene expression level variation, shown as $\log_{10}$ (standard deviation), as a function of average gene expression $\log_{10}$ (average UMI/cell) in scRNA-Seq population. Black line: power-law fit (slope: k). Statistics: Table S1.
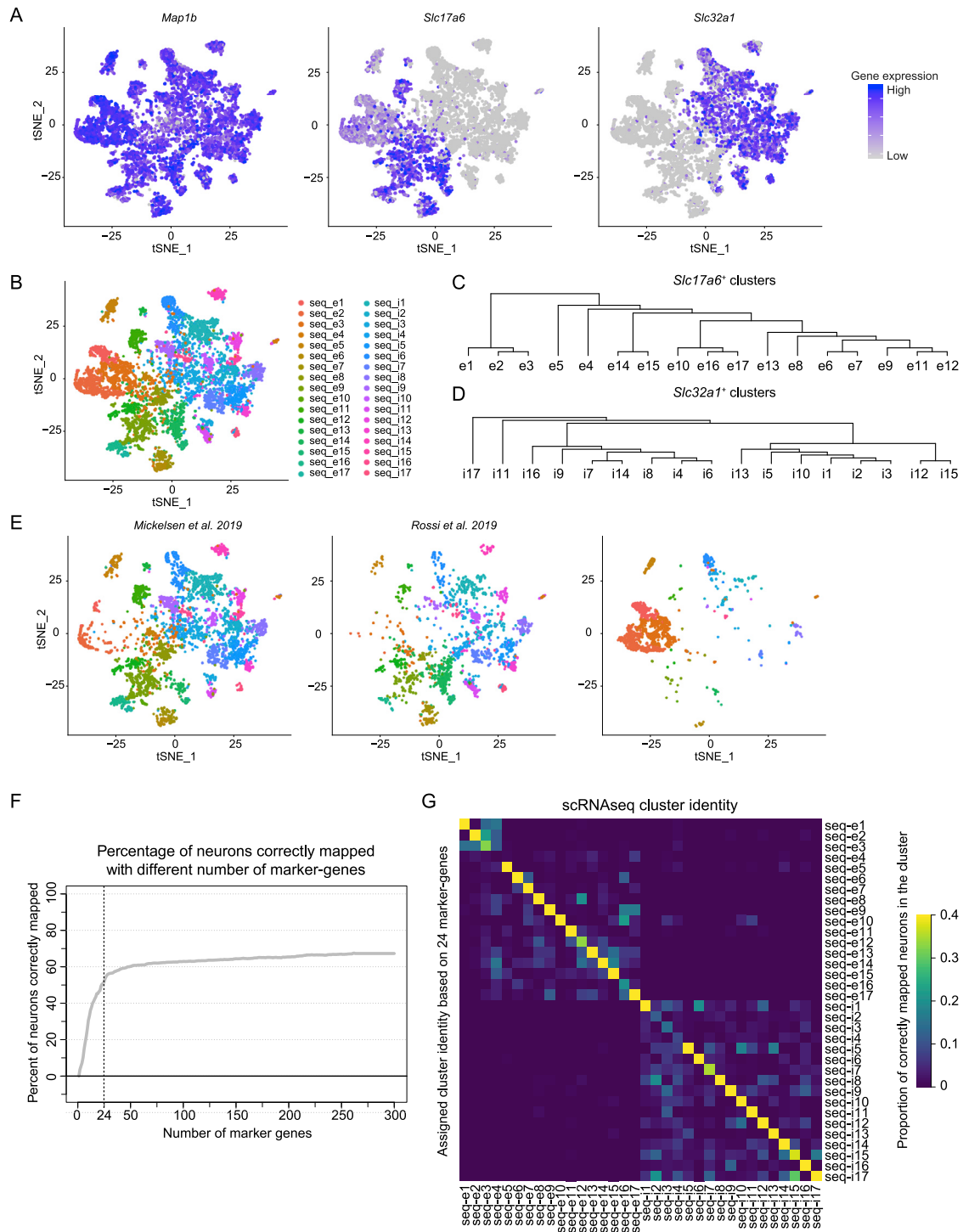
Figure S3. scRNA-seq analysis of molecularly defined cell types in the LHA, related to STAR Methods

(A) tSNE plot of scRNA-Seq data showing *Map1b, Slc17a6*, and *Slc32a1* expression in the LHA neurons.

(B) tSNE plot for molecularly defined clusters in the LHA, with cells color-coded by cluster.

(C and D) Hierarchical analysis of the *Slc17a6*+ clusters (C) and *Slc32a1*+ clusters (D).

(E) Cell cluster distribution across three datasets.

(F) Percent of correctly mapped neurons with increased number of marker-genes (dotted vertical line indicates percent of neurons correctly mapped with the selected 24 marker-genes). The top 300 most differentially expressed marker-genes were chosen for this analysis.

(G) Proportion of neurons correctly mapped to each scRNA-Seq cluster based on selected 24 EASI-FISH marker-genes.
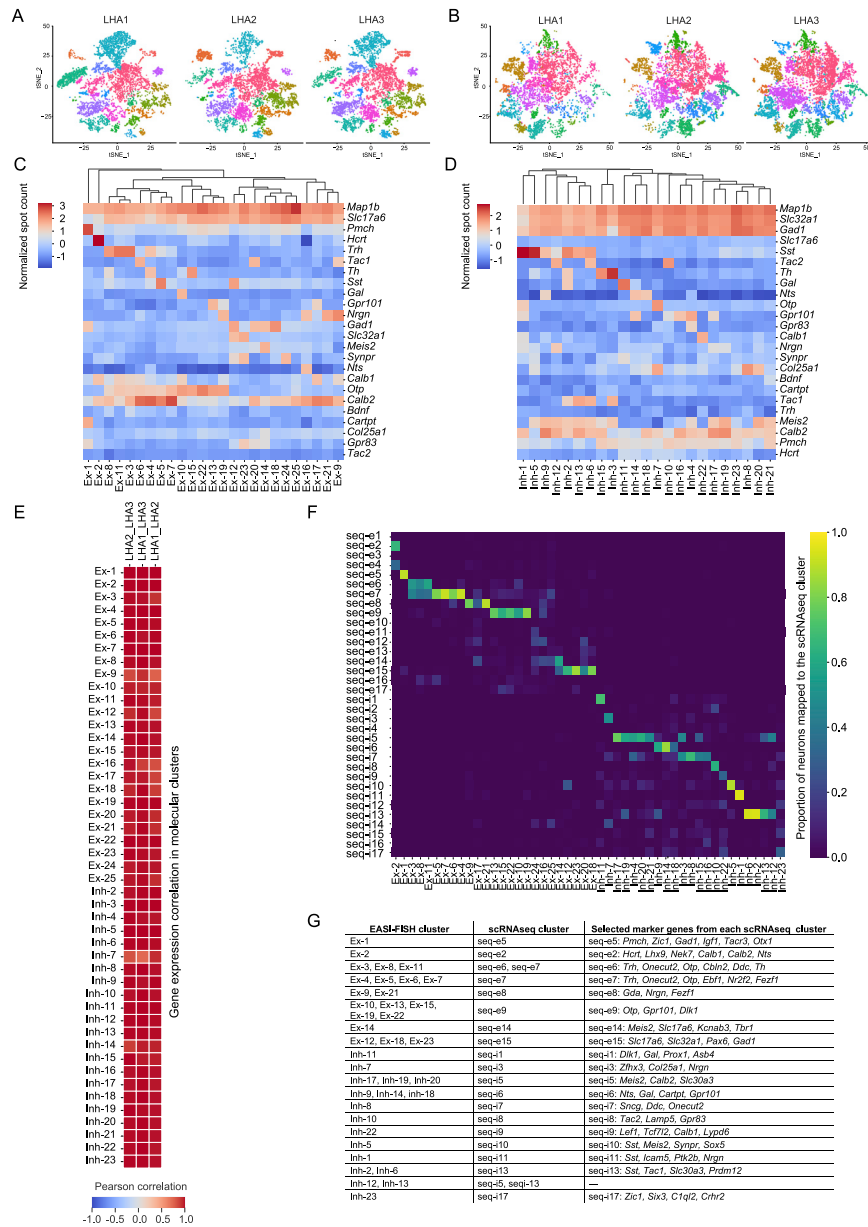
Figure S4. EASI-FISH for profiling LHA molecular markers, related to Figure 3

(A and B) tSNE plot showing *Slc17a6*⁺ and *Slc32a1*⁺ cluster distribution in different animals (from left to right: LHA1, LHA2 and LHA3), see Figure 3C and 3D) for cell type correspondence.

(C and D) Heatmap and hierarchical analysis of marker-gene expression for (C) *Slc17a6*⁺ and (D) *Slc32a1*⁺ molecularly defined cell types. Color bar: z-score normalized spot count.

(E) Correlation analysis of marker-gene expression between samples within molecularly defined cell types.

(F) Proportion of neurons from EASI-FISH clusters mapped to scRNA-Seq clusters based on cross-correlation of marker-gene expression.

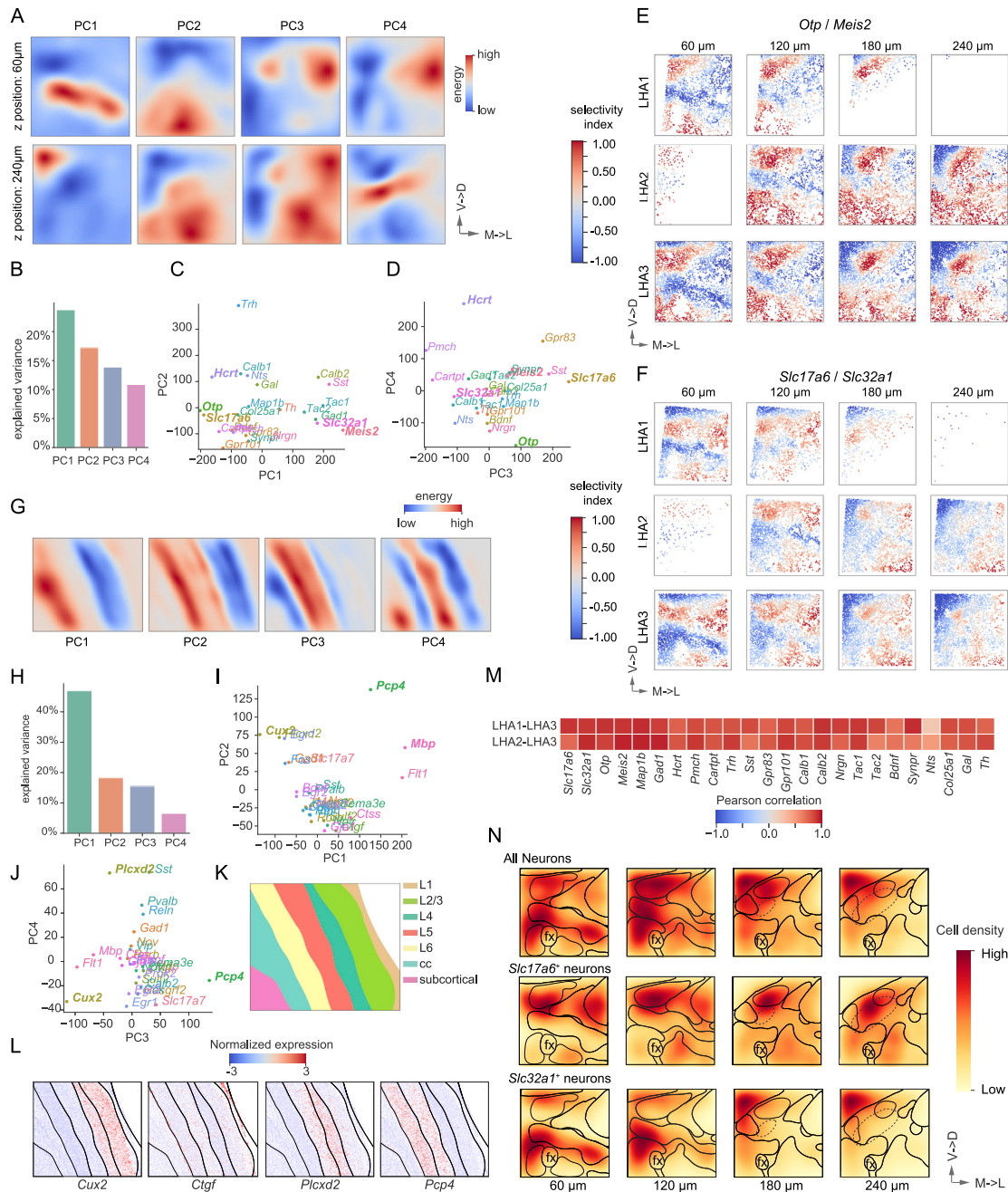(G) Summary table showing correspondence between EASI-FISH cluster and scRNAseq clusters.

Figure S5. Spatio-molecular parcellation of the LHA, related to Figure 4

(A–D) Principal component analysis (PCA) of the marker-genes spatial distribution. (A) Eigen-images from the top 4 principal components (PCs), two different axial depths (60 μm and 240 μm) shown here. Scale arrows: 150 μm. (B) Explained variance from the top 4 PCs. (C and D) PCA plot showing the spatial expression variations of marker-genes on the top 4 PC axes. Scale arrows: 150 μm.

(E) Relative *Otp/Meis2* enrichment across three samples. Scale arrows: 150 μm.

(F) Relative *Slc17a6/Slc32a1* enrichment across three samples. LHA1 and LHA2 were aligned to LHA3 with rigid registration. Scale arrows: 150 μm.

(G–L) Computational parcellation of a region of the visual cortex using published STARmap dataset. (G) PCA of the marker-genes spatial distribution and eigen-images from the top 4 PCs. (H) Explained variance from the top 4 PCs. (I and J) PCA plot showing spatial expression variations of marker-genes on the top 4 PC axes. (K) Computational parcellation of the visual cortex with 4 genes (*Mbp*, *Pcp4*, *Cux2*, *Plcxd2*) selected from the PCA analysis. (L) Expression patterns of known layer-specific marker-genes overlaid with computational parcellation.

(M) Correlation analysis of marker-gene spatial distributions across samples. Image volumes were first aligned and then binned to 10 × 10 × 4 bins. The number of marker-gene positive cells in each bin was calculated and used for the correlation analysis.

(N) Spatial neuronal density (top), *Slc17a6*⁺ neuronal density (middle) and *Slc32a1*⁺ neuronal density (bottom) in the LHA, computed using kernel density estimation (kde). Scale arrows: 150 μm.
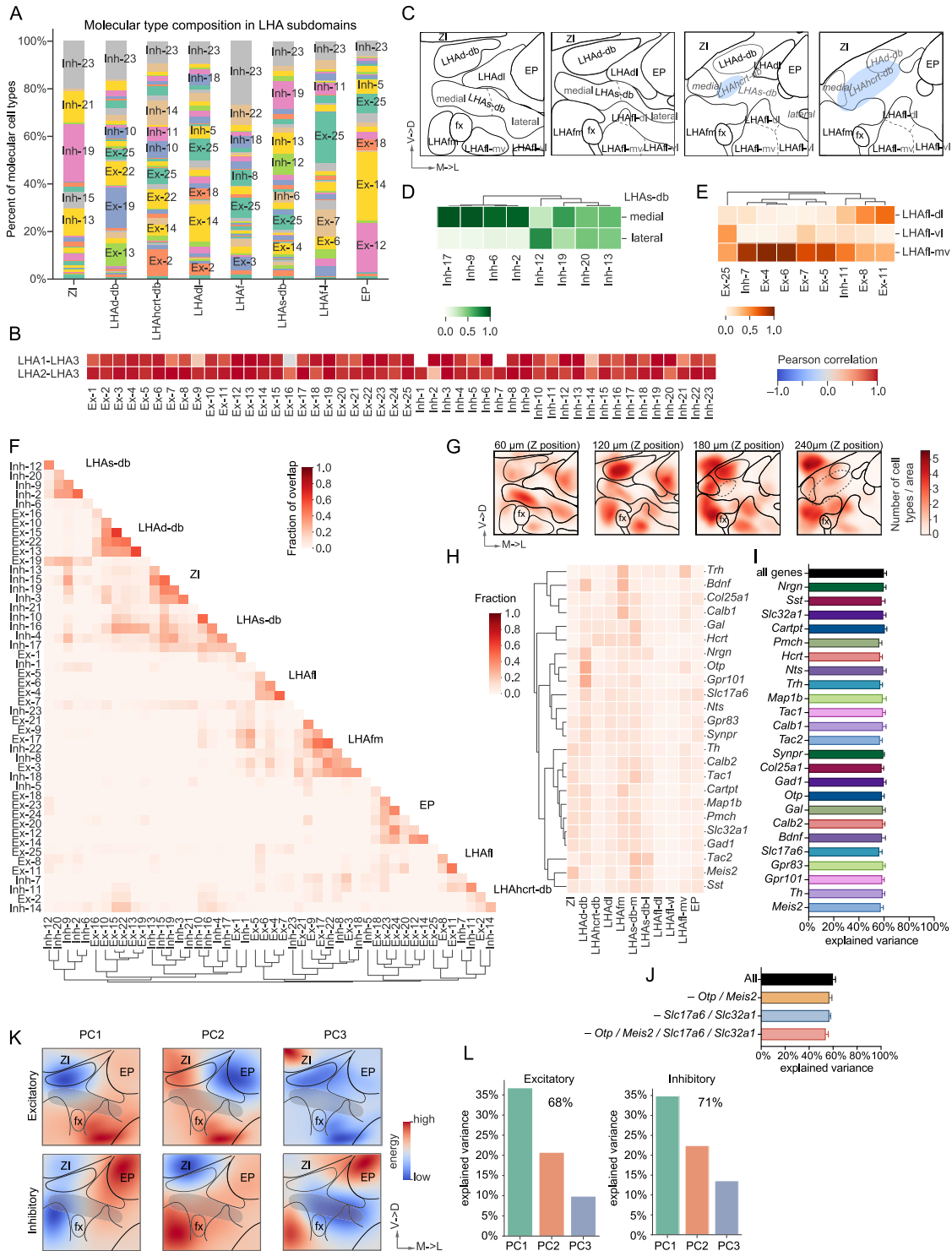
Figure S6. Molecularly defined cell types are enriched in LHA subregions, related to Figure 5

(A) Molecularly defined cell type compositions in LHA subregions.

(B) Correlation analysis of the spatial distribution of molecularly defined clusters across samples. Fraction of molecularly defined neuron clusters in each sub-region was used for the correlation analysis.

(C) Spatio-molecular parcellation of the LHA (as in Fig. 4D) with additional subdivisions based on segregation of molecularly defined cell types. Annotations in gray and dotted line boundaries indicate subdomains within LHAs-db and LHAfl based on spatial segregation of molecularly defined cell types. Scale arrows: 150 μm.

(legend continued on next page)

(D) Fraction of neurons from selected molecularly defined clusters in the medial and lateral part of the LHAs-db.

(E) Fraction of neurons from selected molecularly defined clusters in the LHAfl subdomains.

(F) Molecularly defined cell types grouped by their spatial overlap.

(G) Molecularly defined-cluster-occupancies in the LHA. Colormap indicates the number of clusters occupying each location. Scale arrows: 150 μm.

(H) Spatial distribution of marker-gene expression in parcellated LHA subregions.

(I) Random Forest regression models were used to predict the spatial position of neurons based on marker-gene expression. The prediction accuracy with all 24 marker-genes were calculated and compared to that with 23 marker-genes, with the selected marker-gene removed.

(J) Prediction accuracy of neuronal spatial position in the presence or absence of *Otp*, *Meis2*, *Slc17a6*, *Slc32a1*. Error bars in (I and J) represent SD of model prediction accuracy with cross validation.

(K and L) PCA on the spatial distribution of genes from Vizgen MERFISH dataset, separated by excitatory (top) and inhibitory (bottom) populations. (K) Eigen-images from the top 3 PCs. Parcellation map drawn based on *Slc17a6/Slc32a1* expression. LHAs-db is shown in gray shading because it is transitioning away as this sample corresponds to posterior portion of the volume that we analyzed. Although gene selection for the MERFISH dataset was not ideally suited to describe LHA cell types, PCA analysis of the gene expression patterns showed spatial variability in gene expression that corresponded to five of the LHA subregions described above (LHAd-db, LHAdl, LHAfl, LHAfm and LHAs-db) in eigen-images from the top PCs. For excitatory neurons, LHAd-db and LHAfl are apparent in PC1 and PC3, with EP and LHAdl in PC2. For inhibitory neurons, LHAfm is observed in all 3 PCs and LHAs-db is a distinctive region in PC3. ZI can be identified from both excitatory PC3 and inhibitory PC2. Scale arrows: 150 μm. (L) Explained variance from the top 3 PCs, which explain ~70% variance from the data.
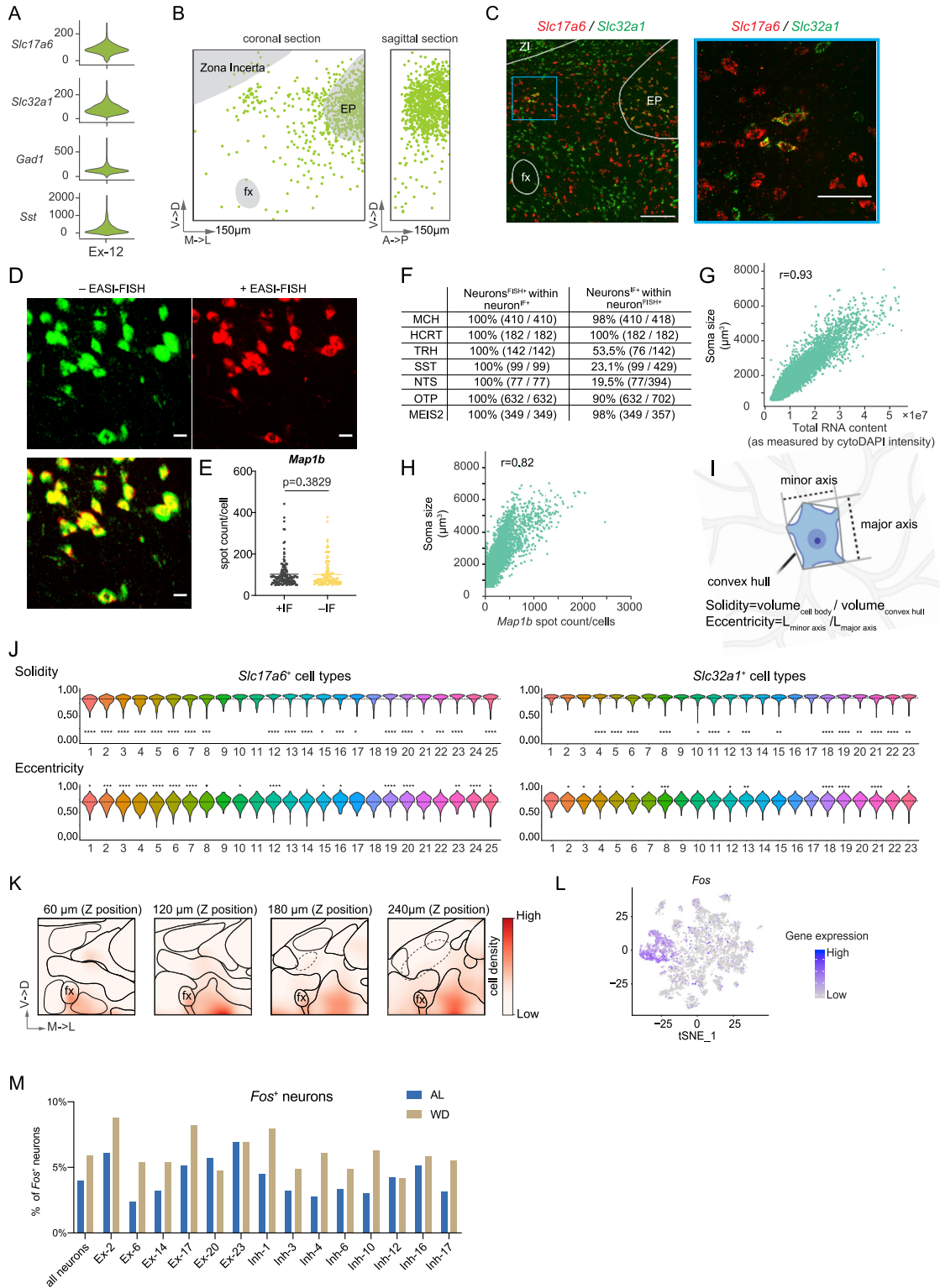
Figure S7. EASI-FISH identified molecularly defined clusters and their morphology and enrichment in *IEGs*, related to Figures 6 and 7

(A) Expression of selected genes in *Slc17a6* and *Slc32a1* co-expressing neurons.

(B) Spatial distribution of *Slc17a6* and *Slc32a1* co-expressing cluster, Ex-12. Majority of these neurons are in the entopeduncular nucleus (EP), with a small cluster in the LHAd-db. Scale arrows: 150 μm.

(legend continued on next page)

(C) Representative image highlighting the *Slc17a6* and *Slc32a1* co-expressing neurons in the EP and LHA. Left: Representative single optical plane image showing *Slc17a6* and *Slc32a1* expression in the LHA. Scale bar: 100μm. Right: zoom-in of region highlighted in blue box (left) to show the *Slc17a6/Slc32a1* neurons in the LHA, Scale bar: 50μm.

(D) Immunofluorescence (IF) signals before (−) and after (+) EASI-FISH procedure, suggesting that IF signals were well preserved during the EASI-FISH procedure. Scale bar: 10μm.

(E) Minimal RNA loss was observed during IF procedure.

(F) Summary table showing number of positive neurons detected with IF and FISH for neuropeptides (MCH, Orexin, TRH, SST, NTS) and transcription factors (OTP, MEIS2).

(G) Correlation between total RNA content and soma size. Statistics in Table S1.

(H) Correlation between pan-neuronal marker, *Map1b* expression and soma size. Statistics in Table S1.

(I) Schematic showing 3D morphological measurement, solidity and eccentricity performed in (J).

(J) Solidity and eccentricity measurement in EASI-FISH excitatory (left) and inhibitory (right) clusters. Statistics in Table S1.

(K) Density analysis of neurons with irregular shape. Four different axial depth shown as representative images. From left to right: anterior to posterior (rostral→caudal). Scale arrow: 150 μm.

(L) tSNE plot showing baseline *Fos* expression in the LHA based on scRNA-Seq. see Figures S3B and S4G for corresponding molecular cell types.

(M) Incidence of *Fos* expression in additional molecularly defined neurons from *ad libitum* (AL) and water-deprived (WD) groups. Number of neurons analyzed from each molecularly defined cell type in each group. Differences in *Fos* expression for these cell types were not statistically significant after correction for multiple comparisons. Statistics in Table S1.