Discovery of neuronal cell types by pairing whole cell reconstructions with RNA expression profiles

Tiago A. Ferreira[✉], Mark Eddison, Monique Copeland, Mary Lay, Emily Tenshaw, Monet Weldon, David Schauder, Donald J. Olbris, Konrad Rokicki, Nelson Spruston, Paul W. Tillberg, Wyatt Korff[✉], Joshua T. Dudman[✉], and The MouseLight Project Team

Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA

[✉] Correspondence: dudmanj@janelia.hhmi.org; korffw@janelia.hhmi.org; ferreirat@janelia.hhmi.org

Abstract

Effective classification of neuronal cell types requires both molecular and morphological descriptors to be collected *in situ* at single cell resolution. However, current spatial transcriptomics techniques are not compatible with imaging workflows that successfully reconstruct the morphology of complete axonal projections. Here, we introduce a new methodology that combines tissue clearing, submicron whole-brain two photon imaging, and Expansion-Assisted Iterative Fluorescence *In Situ* Hybridization (EASI-FISH) to assign molecular identities to fully reconstructed neurons in the mouse brain, which we call morphoFISH. We used morphoFISH to molecularly identify a previously unknown population of cingulate neurons projecting ipsilaterally to the dorsal striatum and contralaterally to higher-order thalamus. By pairing whole-brain morphometry, improved techniques for nucleic acid preservation and spatial gene expression, morphoFISH offers a quantitative solution for discovery of multimodal cell types and complements existing techniques for characterization of increasingly fine-grained cellular heterogeneity in brain circuits.

Introduction

There is broad agreement that a critical step to understanding the remarkable complexity of the mammalian brain is to develop an increasingly detailed and comprehensive understanding of both its molecular constituent cell types as well as their complex projection patterns[1]. The ability to ultimately take such parts lists of cell types and assemble them into increasingly sophisticated models of brain function will depend both on understanding how molecular diversity contributes to the development and physiology of neurons as well as understanding the detailed connectivity of the diverse neuronal types. Connectivity is determined in substantial part by the detailed morphology of

long-range axons that can span the entire mammalian brain [2–4]. For example, the elaboration of cortex over mammalian development is substantially driven by the elaboration of subtle morphological and molecular variants of long range projection neurons that are thought to be critical to empower new functions [5].

There has been remarkable progress over the last several years on complementary fronts: complete reconstructions of the extended morphology of neurons throughout the entire brain[3,4,6–8] and increasingly deep analysis of the molecular constituents of individual cells *in situ*[6,9,10]. However, for a host of technical reasons, these advances have also made these two major classes of measurement incommensurate. Thus, to date, it has not been possible to obtain complete anatomical reconstructions and molecular cell-type characterization from individual cells in the same tissue. Rather, it has generally been necessary to compare across datasets to identify average projection patterns of molecularly defined cell types [2] or infer the likely molecular profile of individual reconstructed neurons [11,12]. Pioneering approaches to spatial transcriptomics and single-cell barcoding have led to great insight into the intersection of whole brain projection targets and molecular profiling[4]. However, a critical unmet need is methods that allow one to combine complete and detailed morphological reconstruction at submicron resolution with molecular characterization in individual neurons.

To accomplish this goal, we sought to develop a pipeline that combines molecular identification and complete morphological reconstruction of the same neurons. This required refining software to obtain complete and accurate single neuron morphological reconstruction, as well as developing a new method that preserves ribonucleic acids (RNA) and allows for multi-round fluorescent *in situ* hybridization (FISH) after weeks of processing time required for whole brain imaging. To obtain the highest topological fidelity, we quantified the most common pitfalls to neuronal tracing and used that data to refine workflows and software that obtain complete, single neuron reconstructions in a high-fidelity regime. To permit molecular profiling of these neuronal types *in situ*, we build upon the EASI-FISH approach to develop a protocol with vastly improved methods for preservation of RNA many weeks post-fixation, tissue clearing, and imaging.

Here we describe a substantial addition of hundreds of highly accurate, complete neuronal reconstructions distributed across multiple cell types in the mouse brain, spanning a range of reconstruction problems from massive branching in midbrain dopamine neurons to long axons with many sparse branches of neurons in the pontine brainstem. Moreover, we describe and share HortaCloud, the software that greatly facilitates multi-user tracing in the cloud and validation of reconstruction quality. As a proof of principle, we describe the use of morphoFISH for the discovery of a rare morphomolecular subtype of long range projection neuron specific to the cingulate cortex in mice.

Results

Release of new morphological reconstructions

The complete reconstruction of single axons in mammalian neurons is both essential and profoundly challenging[13]. A significant constraint on the function of a neuron are the other brain areas and cell types to which it connects via its axonal projections. In the previous data releases for the MouseLight Project Team, we have observed a broad distribution of complete axonal length and number of brain areas targeted from a sampling of ~1200 reconstructed neurons across multiple, but necessarily limited cell type and brain area sampling. Here, we reconstructed an additional ~500 neurons. These include a subset of neurons from frontal cortical areas with particularly long axonal arbors relative to other cortical neurons. This allowed direct, quantitative comparison to recent large datasets[7]. In addition, we now include reconstructions from underrepresented subcortical areas including a large population of medulla projection neurons and sparse reconstructions from cell types that define the upper and lower bounds of the overall distribution observed in the previous ~1200 reconstructions (**Fig. 1a** and **Data Availability**).

Cable lengths of individual neuron axonal arbors in the mouse brain can range from a few millimeters to total up to ~50 cm^3. These large cable lengths can be observed either in neurons that connect very densely to just a single brain region by elaborating complex and plentiful short terminal axonal branches. This is typified by midbrain dopamine neurons of the substantia nigra (SN; **Fig. 1a–c**). At the other extreme, a neuron may elaborate a similar cable length of axon and yet innervate several tens of different brain regions with relatively short and sparse terminal axonal branches. This is typified by projection neurons of the nucleus incertus (NI; **Fig. 1a,d**). Cell types from the cortex to the hindbrain fill in an approximately continuous distribution between these extremes.

This broad range of axonal complexity presents a number of critical challenges for reconstruction. In particular, it is often the case that a neuron may have a single or few axonal branch points located somewhere along centimeters of axon that determines its connectivity to one or several brain areas via exceptionally variable length axonal branches (**Fig. 1f–h**). Accurate identification of all brain areas targeted by single neurons is a critical functional feature and essential for identifying morphotypes of projection neurons and distinguishing amongst related

projection neurons from a given area **(Fig. 1i)**. This can lead to important quantitative differences in the inferred properties of single cells depending upon the reconstruction method used and the relative certainty about completeness and accuracy (**Fig. 1f–h**).

Development of morphoFISH

Complete axonal tracing in mouse brain tissue requires extensive chemical modification of the brain tissue to accommodate whole brain imaging (**Fig. 2a**). This has been previously achieved at the expense of RNA integrity. Indeed, RNA is barely detected in MouseLight imaged tissue[3] (**Fig. 2b**). Taking inspiration from EASI-FISH[14]—a highly sensitive technique that stabilizes RNA by capturing it in an expandable hydrogel—we developed a protocol that prevents RNA degradation during histological handling and imaging of the brain by embedding in a non-swelling polyacrylamide hydrogel. After whole brain imaging, regions containing cell somas are later targeted for expansion and EASI-FISH processing. The method both increases the initial RNA retention and slows down its degradation. We took a number of steps to preserve RNA in whole brain preparations. First, we implemented a 3-step transcardiac perfusion in which the animal is first perfused with paraformaldehyde (PFA) alone; then with methacrylic acid-NHS ester, which reacts with proteins and will later anchor them to the gel; then with a solution including acrylamide monomer, bis(acryloyl)cystamine crosslinker, and the APS/TEMED initiator system, which will later polymerize to form the gel. A polymerization inhibitor (4-hydroxy-TEMPO) is added to this mixture to ensure gelation solution has time to diffuse throughout the brain before the polymerization reaction takes place. Once the brain is dissected and gel allowed to polymerize, it becomes embedded in an acrylamide gel where proteins are covalently anchored to the gel polymer (**Fig. 2a**). PFA post-fixation is complemented with glutaraldehyde (GA) and melphalan, an alkylating agent that reacts with nucleic acids and bears a primary amine that is fixable with aldehydes[14,15], creating extensive crosslinking amongst proteins, RNA, and the acrylamide gel. Indeed, with this protocol, we effectively doubled the capture of total RNA in the brain when compared to prior perfusion approaches (**Fig. 2b**).

We next focused on mitigating RNA loss during the weeks-long histological processing of the tissue, which includes delipidation and refractive index (RI) matching. While the former is aimed at removal of light-scattering lipids from the tissue, the latter is aimed at minimizing the spatial variation of refractive index due to the heterogeneity of tissue components. Combined, these are the longest processing steps and proved to be problematic for RNA preservation in tissue. To monitor detection sensitivity after removal of lipids, we compared tissue with and without delipidation for abundant RNA transcripts. We found similar expression levels across individual cells regardless of whether

delipidation had been performed (**Fig. 2e,f**). We then tested refractive index matching strategies [16–19] for RNA retention and settled on the DMSO/iodixanol cocktail already adopted by our platform [3,20]. To further improve RNA preservation, we supplemented the imaging bath with polyvinylsulfonic acid (PVSA), a negatively charged polymer with side structures mimicking the conformation of mRNA molecules that has been shown to be an effective competitive inhibitor for ribonucleases[21].

In total, the complete morphoFISH protocol was successful at yielding approximately the same preservation of RNA after 30 days post-perfusion, as achieved with standard fixation procedures in prior EASI-FISH protocols (day 0 standard fixation, D0SF) (**Fig. 2b,c**). Consistent with improved RNA retention, cleared tissue left unattended for 5 days in the serial-sectioning two-photon microscope used by MouseLight retained comparable D0SF levels (**Fig. 2d**). Typically, reconstruction of complete axonal arbors requires signal amplification of fluorophores along dim axon collaterals using whole brain immunohistochemistry (ref). This procedure would take several weeks on top of the 30 days to process the brain and is therefore prone to compounded mRNA loss. Rather than incorporating this extra step into morphoFISH, we focused on optimizing a pipeline for reconstructing single neurons labeled with endogenous, unamplified fluorescent reporters, which enables effective, albeit somewhat lower yield labeling[22].

High-fidelity neuronal reconstructions

We obtained sparsely labeled neuronal populations by injecting a mixture of low-titer AAV Cre and a high-titer Cre-dependent reporter as previously described[3]. In order to further improve reconstruction accuracy, we first quantified the most common errors afflicting reconstructions of long-range axonal projections. Second, we used insights from the most problematic errors to optimize the development of our custom reconstruction software, HortaCloud, in preparation for making it available to the broader scientific community. Third, we released a standardized protocol and guidelines for tracing long-range projections using the open HortaCloud tool.

For error quantification, expert annotators independently reconstructed the same cell, and a ground-truth reconstruction was generated by comparing their work (**Fig. 3, and Methods**). After repeated tracing of a mediodorsal (MD) thalamic neuron with more than 100 mm of cable length, we found that: 1) tracing errors could be grouped into four categories; 2) errors were spread across the ramified sections of the axonal arbor; 3) errors were common even amongst expert tracers; and 4) errors occurred at frequencies that were low but nonetheless sufficient to distort the topology of the cell and complete annotation of its target areas (**Fig. 3a-d**). We found that ‘crossovers’ between neurites (either from the same cell or from neighboring cells) have the highest potential to alter

projection patterns while being the most time-consuming to resolve. For many neuronal morphotypes, this can often require expert perusal of as much as extra 12 cm of cable length per cell, because it may involve tracing neighboring neurites from multiple cells (**Fig. S1**). Building on previous work[3] we found that assembling consensus reconstructions from two independent annotators provides an acceptable balance between accuracy and throughput. However, tracings obtained from a large group of annotators that did not follow expert guidelines were found to be of low quality, as assessed by DIADEM score comparisons[23] (**Fig. 3e**). Collectively, these results are at odds with previous studies proposing that spot-checks after automated segmentation are sufficient to obtain acceptable accuracy[7,24]. Indeed, we noticed that while such spot-check approaches are biased towards bright neurites, we found that fluorophore availability is not strongly correlated with the ability to fully reconstruct long-range projections (**Fig. S2**), which may contribute to the observed quantitative differences in the distribution of cable lengths between MouseLight datasets and reconstructions from frontal cortical neurons by other groups[7] (**Fig. 1f-h**).

Identification of a novel morphomolecular neuron subtype

The successful preservation of RNA in tissue allows multi-round *in situ* hybridization-based molecular profiling of individual neurons within tissue suitable for complete tracing of sparsely labeled neurons. While this approach is necessarily limited in throughput, it can nonetheless, in principle, empower the molecular profiling of neurons with unique axonal morphologies.

Areas of the frontal cortex, in particular cingulate areas, appear to harbor molecular subtypes of deep layer, putative long-range projection neurons that are unique and not found in other cortical areas [12]. Despite the detailed molecular characterization, it is currently not possible to map these rare or unique molecular subtypes onto unique morphological projection patterns, which will likely be critical for complete cell typing [1]. Building upon our extensive tracing of individual neurons in frontal cortical areas we looked for individual neurons with rare axonal projection patterns. We identified one particular neuron with a highly unique projection pattern based upon previous examinations of bulk and large scale single neuron reconstructions. Specifically, we identified a neuron in the cingulate cortex that projects ipsilaterally to the dorsal striatum (STRd), or the caudoputamen (CP), and contralaterally to the higher-order parafascicular thalamic nucleus (PF) (**Fig. 4a–c**). Given the absence of any such combination of projections in a single neuron in other cortical areas, this suggested that this neuron could represent a unique morphomolecular subtype.

Following complete tracing of this neuron, we were able to profile the molecular expression

pattern through several rounds of non-destructive *in situ* hybridization using methods modified from the EASI-FISH protocol previously described[14,25] (**Fig. 4d**). Taking advantage of critical resources of gene expression profiling using deep sequencing of cells in the cingulate cortex, we identified and validated a family of marker genes sufficient to identify unique molecular types of deep layer projection neurons (**Sup. Information**). Using multiple rounds, we probed the target neuron with 2 probes across 6 rounds of *in situ* hybridization and imaging. Quantification of expression level of each gene across the population of neurons within the imaged tissue section identified the target neuron as positive for Galnt14 and Kccn2 and negative for a host of other genes—consistent with a unique subclass of deep layer projection neuron found in cingulate cortex. This connects a unique morphotype of projection neuron to relatively rare molecular subtypes of projection neurons in the cingulate/retrosplenial cortices identified from extensive sequencing of cortical neurons in the mouse brain [26].

We next sought to confirm the molecular identity of projection neurons with this unique innervation pattern and discover the relative prevalence of this morphotype in a higher throughput manner. The identification of the unique projection properties allowed us to design an intersectional viral labeling scheme. We injected retrograde viruses into the contralateral PF (expressing Cre-recombinase) and ipsilateral dorsal striatum (expressing GFP) in 3 tdTomato cre-dependent reporter mice (**Fig. 4e**). The unique morphology of the target neuron should be reflected in a sparse population of dual-labeled (tdTomato+ and GFP+) ipsilateral to the dorsal striatal injection site and located in the cingulate cortex. This is expected to be contrasted with a much larger population of tdTomato+ deep layer projection neurons ipsilateral to the Pf injection site, as thalamocortical projections are primarily ipsilateral not derived from the bilateral intratelencephalic projection subtypes that innervate striatum[3]. To provide a comprehensive estimate of the relative prevalence of the STR^{IPSI}/PF^{CONTRA} morphotypes, we performed mesoscale imaging of the majority of the rostral brain and confirmed that the intersectional labeling strategy did identify a small but clear population of STR^{IPSI}/PF^{CONTRA} projection neurons localized to the rostral cingulate cortex (**Fig. 4f**). Finally, to confirm that this population was consistent with the morphomolecular type identified from morphoFISH, we performed FISH on the labeled tissue (**Fig. 4g**). We found that the dual-labeled subset of STR^{IPSI}/PF^{CONTRA} projection neurons were indeed excitatory neurons ($Slc17a7^{+}$) from the subset of $Galnt14^{+}$ deep layer projection neurons in the cingulate.

Discussion

Here we present several innovations, including refinements to techniques for tissue preparation and software for efficient, complete, and accurate reconstructions that combine to empower simultaneous multi-round *in situ* hybridization-based molecular

profiling of individual neurons with complete axonal reconstructions in the mouse brain—a resource we refer to as morphoFISH. We provide one particular example of a unique morphomolecular cell type identified in the cingulate cortex. Discovery of this unique cell type could be rapidly translated into customized, higher-throughput techniques suitable for future functional experiments. The cingulate cortex is thought to contain a number of unique molecular and presumed anatomical types of projection neurons that may contribute to the unique and critical functions of the cingulate cortex. The sharing of output mediated by these unique projection neurons to ipsilateral high-order thalamic areas and contralateral dorsal striatum may play critical roles in coordinating interhemispheric coordination of motor skill learning coordinated with internal state.

While morphoFISH presents a unique resource that can be a critical component for a broad effort to identify morphomolecular cell types in the mouse brain, it also presents some fundamental challenges to throughput. Molecular profiling via repeated rounds of *in situ* hybridization is powerful and has been used to discover novel functional correlates of molecularly defined cell types [27]. The approaches for dramatically improved RNA retention described here are general and offer the promise of enhancing many approaches to molecular profiling, including the improved characterization of low-expressing genes that may be critical for some questions. However, while this approach offers potentially very high sensitivity for individual transcripts, it also requires substantial time due to modest (≤4) multiplexing as compared to higher throughput approaches for spatial transcriptomics. Autofluorescent lipofuscin granules start to accumulate in the mouse brain after 8 weeks [28], and so we set our quantification routines to ignore lipofuscin foci. As a consequence, we are likely undercounting HCR spots across probing rounds. We predict that the same type of experiments performed in younger animals would allow for more straightforward quantification.

The recent advent of approaches allowing the full reconstruction of sparsely labeled neurons in the mouse brain has already yielded large databases that include thousands of reconstructed neurons. At the same time, this is a small sampling relative to the entire mouse brain and is heavily biased towards select neuron types in specific brain areas. Automatic processing via modern methods and software continues to rapidly develop and has dramatically improved the throughput of reconstruction. However, there remains a critical role of human intervention, as even exceptionally rare errors, e.g., a missing branch, can produce large differences in reconstruction accuracy. There is no ground truth to date, and thus differences in average properties across different reconstruction methods, a small subset of which we describe here, will be critical to evaluate carefully as reconstructed neurons proliferate. In *Drosophila*, light-level reconstructions have been indispensable for interpreting EM connectomes and have guided retrieval of molecular, functional, and mesoscale data used to catalog neuron types in detail [29]. Given the expert

curation of our morphologies, and the multi-modal characterization enabled by morphoFISH, we predict such data will be critical to facilitate similar advances in the mouse.

The detailed reconstructions and multi-user annotations that constitute the database of reconstructed neurons described here were all performed in HortaCloud – a new cloud-based tool for collaborative and accurate neuron reconstruction. Unlike traditional desktop applications, HortaCloud enables collaborative tracing across remote locations using a web browser. HortaCloud builds on the proven track record of MouseLight's Horta desktop application[3] with a cloud system that any user or institution can easily deploy to avoid the complex and costly set up previously required for handling terabyte-sized anatomical datasets with on-premise infrastructure. HortaCloud presents a freely available and open source application that is already 'future proofed' for new imaging modalities and file types (e.g. OME-ZARR; ExaSPIM) as they become available and includes a number of critical analysis tools to lower the burden for early adopters (e.g. EASI-Tower, SNT, Python notebooks). Finally, to facilitate adoption of the methods described here, we have shared detailed, step-by-step protocols[20] that democratize access to this technology.

Methods

Animals

Wild-type C57BL/6 animals were obtained from Charles River Laboratories and cre lines from the Jackson Laboratory (Sup. Table 1). Adult females (12-20 weeks) were used for all experiments and were group housed with sex-matched littermates. Mice had access to ad libitum food and water and were housed in an enriched environment. All experimental protocols were conducted according to the National Institutes of Health guidelines for animal research and were approved by the Institutional Animal Care and Use Committee at Howard Hughes Medical Institute, Janelia Research Campus (Protocol #14–115).)

Viral Labeling

Dual injection of rAAV2-retro helpers[30] (Addgene plasmid #81070; RRID:Addgene_81070)

Imaging

Dual Whole brain: Whole brain imaging was performed as previously described[3]. To ensure RNAse activity would not accumulate in the imaging bath over time, a peristaltic pump was used to replenish imaging solution at a constant speed of 1ml/min. An Ambion™ RNaseAlert™ lab test kit (ThermoFisher) was used to monitor RNAse activity and solution replaced in case of detection. Imaging progressed as described previously[3,22] but brain slices (180-200um) were collected every 2 cuts using a sterile brush and placed in 24-well plates. Each slice was washed (once in RNAse-free water and twice with 70% ethanol) and placed in 70% ethanol at 4°C until probing[25]. Serial-sectioning: Following standard perfusion and protocols, dissected brains were embedded in 5% agarora, and coronal sections (100μm thick) were cut on a vibratome (Leica). Sections mounted in Vectashield® hardset media with DAPI (Vector Laboratories) and imaged on a confocal slide scanner using the TissueFAXS platform (TissueGnostics).

Image Processing

Stitching, segmentation of neurites, and brain registration against the Allen CCF were performed as previously described[3].

Morphometry

Connectivity analyses (Fig. 1, 3, and 4): Only CCF compartments of ontology depth 7 with publicly available meshes were considered (141 compartments). Morphometry comparisons, volume analysis, and reconstruction renderings were performed in SNTv4.9.9[31].

Tissue preparation

Detailed morphoFISH protocols are published at protocols.io[20]. In short: Transfected mice were anesthetized with an overdose of isoflurane and then transcardially perfused with the following solutions sequentially: 1) perfusion fixative (4% PFA); 2) protein anchoring solution (4% PFA, 20mM MA-NHS); 3) gel solution (4% PFA, 20% AcAm, 0.1% BAC, 0.2% TEMED, 0.01%4-Hydroxy-TEMPO, 0.2%2 APS); and 4) post-fixation solution (4% PFA, 0.05% glutaraldehyde, 10mM MOPs (pH 7.0), 1mg/ml melphalan). Brains were extracted and gel allowed to polymerize for 1 hour in a nitrogen-saturated enclosure. Brains were then post-fixed in solution 4) at 4°C overnight (40 hours) and washed in PBS to remove all traces of excess fixative. Delipidation and index matching proceeded as detailed in the companion protocols[20,32], brain was embedded in 4% low melting agarose, and solutions supplemented with 10mg/ml PVSA.

Quantification of total RNA

200 µm-thick brain sections were sliced on a vibratome and collected into individual wells of a 24-well plate in 1x RNAse-free PBS (Invitrogen). 2 slices were collected at the reported time points in Fig. 2f, and processed for cytoDAPI detection as previously described[25]. Tissue was imaged on a Light-sheet microscope (Zeiss), cells segmented with Cellpose3[33], and intensities of segmented ROIs measured in Fiji[34]. Intensities were normalized to the overall mean obtained from D0SF slices.

EASI-FISH

Probing was performed as previously[14] and detailed in the companion protocol[25] with two genes probed per round. Quantification of HCR spots was performed as follows: Cellpose3[33] was used for 3D segmentation of cytoplasmic DAPI signal, applying its default “cyto3” model on both XY and YZ planes. Spot-to-cell assignment was performed by iterating through each mask and running a 3D-maxima detection routine[35] on the masked voxels for each color channel. Maxima with the same centroid coordinates on both channels were assumed to report lipofuscin and were discarded. Histograms of spot counts (Fig.4) were obtained from all the segmented cells in the field of view. The amplitude of Gaussian distributions overlaid on frequency data was adjusted so that the area under the curve (AUC) of both distributions matched. Summary expression was obtained by normalizing the spot counts associated with the labeled cell to the 10th and 90th percentile of the distribution obtained from all cells (≈1100) in the field of view (447x447x480µm). Other EASI-FISH quantifications were performed using the EASI-FISH pipeline[14] via its Nextflow Tower interface (https://janeliascicomp.github.io/multifish/). Finalized charts were rendered in GraphPad Prism 10.3.1 (GraphPad Software).

HortaCloud Implementation

The reconstruction software used by the MouseLight Project Team—Horta (How Outstanding Researchers Trace Axons)—was migrated to the cloud by adoption of Amazon's AppStream Virtualized Desktop Infrastructure (VDI) framework, which allows the desktop software to be run in the cloud from a web browser by multiple concurrent users connected to the internet. This effort required several changes to the original codebase and is described in detail elsewhere.

Reconstructions

The detailed protocol is described in protocols.io[36]. In summary, cells with axonal arbors fully resolvable are identified and traced by an experienced annotator. Upon completion, a second annotator peruses the traced centerlines, annotating any mismatched branches. Differences are then extracted between the two annotations and a 'best-guess' consensus is obtained.

Author Contributions

ME, PWT, and TAF established the morphoFISH protocol. MC prepared samples for imaging. TAF performed whole-brain imaging, image processing, and data analyses. ME performed EASI-FISH experiments. CA, ET, ML, MW provided reconstructions and brain registrations. DS, DJO, and KR developed the HortaCloud client. TAF and JTD defined the probing library. JTD designed the AAV2-retro experiment. JC, JTD, and TAF conceptualized the project. TAF oversaw all experiments. JTD and TAF wrote the manuscript. JTD and WK oversaw the project. Remaining members of the MouseLight Project Team contributed to the general development of the project.

The MouseLight Project Team during this time consisted of: Andrew Recknagel, Cameron Arshadi, Emily Tenshaw, Jayaram Chandrashekar, Mary Lay, Monet Weldon, Patrick Edson (Leap Scientific), and Tiago A. Ferreira. Joshua T. Dudman, Karel Svoboda, Nelson Spruston, and Wyatt Korff were part of the MouseLight steering committee.

Acknowledgments

The authors would like to thank Janelia Experimental Technologies, especially Alex Sohn, Daniel Flickinger, Jon Arnold, Mioara Gugiu, Steven Sawtelle, and Vasily Goncharov for hardware design, fabrication, and maintenance; Hyun Ah Yi and Janelia Virus Tools for viral reagents; Michael DeSantis for imaging assistance; Judith Baka for revision of neuronal reconstructions; Maya Tondravi and Gudrun Ihrke for coordinating experimental work; Janelia Scientific Computing, especially Adam Taylor, Ben Arthur, and David Ackerman for

help with data processing; Ken Carlile and Scientific Computing Systems for help with data handling; and Janelia Vivarium, including Rae Demars and Sarah Lindo, for animal care and surgical assistance. We thank Luke Lavis for advice on morphoFISH chemistry. We are also grateful to the developers of the open source libraries used in this study and several laboratories/institutions for releasing data publicly, including the Allen Institute for Brain Science (Allen Mouse Common Coordinate Framework and Cell Types Database: RNA-Seq Data) and CEBSIT (ION) (fMOST reconstructions). This work was supported by the Howard Hughes Medical Institute.

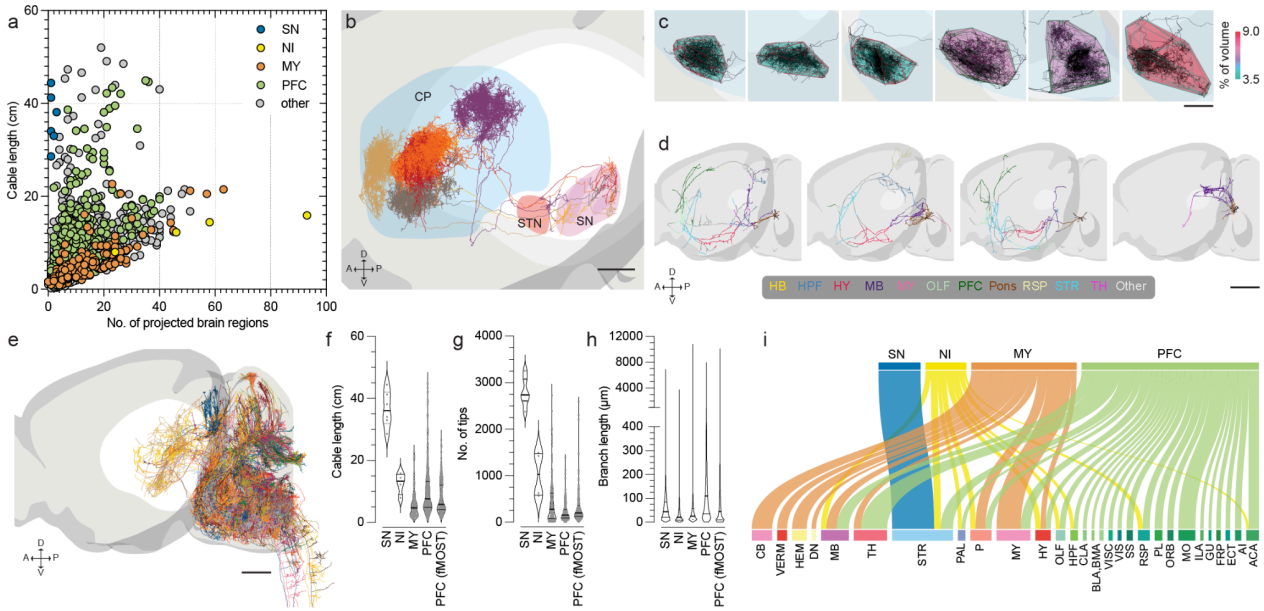
This article is subject to HHMI's Open Access to Publications policy. HHMI lab heads and Project Team Leaders have previously granted a nonexclusive CC BY 4.0 license to the public and a sublicensable license to HHMI in their research articles. Pursuant to those licenses, the author-accepted manuscript of this article can be made freely available under a CC BY 4.0 license immediately upon publication.

Data availability

Neuron reconstructions were deposited at https://ml-neuronbrowser.janelia.org/. Imagery is hosted on AWS Open Data (https://registry.opendata.aws/janelia-mouselight/) and accessible in all HortaCloud instances https://hortacloud.org/. Analysis scripts used in this manuscript are available at https://github.com/MouseLightPipeline/morphoFISH.

Figures

Figure 1. Whole-brain sampling reveals broader morphological diversity of neuronal projection types.



a. The mouse brain (Allen Mouse Brain Common Coordinate Framework (CCF)) was divided into 141 mid-ontology regions and cable length of individual axons plotted against the total number of unique regions innervated by their axonal endings (N= 1638 cells). Cells from Substantia Nigra (SN, N=6), Nucleus Incertus (NI, N=4), Medulla (MY, N=142), and prefrontal cortex (PFC, N=420) are indicated.

b–e. Detailed analysis of populations highlighted in a. dorsoventral (D-V) and anterior-posterior (A-P) axes are indicated.

b–c. SN neurons projecting exclusively to the ipsilateral caudoputamen (CP). Single axonal tufts can cover as much as 9% of the ipsilateral CP lobe, as measured by the 3D convex hull formed by the axon end points. Subthalamic nucleus (STN) is also highlighted. Scale bar: XX mm

d. NI neurons in the pontine brainstem projecting to many areas in the forebrain (including hippocampal formation (HPF), hypothalamus (HY), olfactory areas (OLF), PFC, striatum (STR), and thalamus (TH)), retrosplenial area (RSP), midbrain (MB), and hindbrain (HB, including Pons (P) and MY). Scale Bar XX mm.

e. Medulla neurons with projections to the vermal regions (VERM) of the cerebellar cortex, TH, HY, palladium (PAL), MB, cerebellum (CB), brain stem (BS), and spinal cord. Scale bar: XX mm

f–h. Morphometric properties of the 4 groups of cells. PFC reconstructions from Gao L. et al, 2022 (fMOST, N=426) are included for comparison.

i. Sankey diagram summarizing axonal projection patterns. “Flow weights” reflect normalized innervation to target areas. Abbreviations and leaf colors reflect CCF nomenclature.

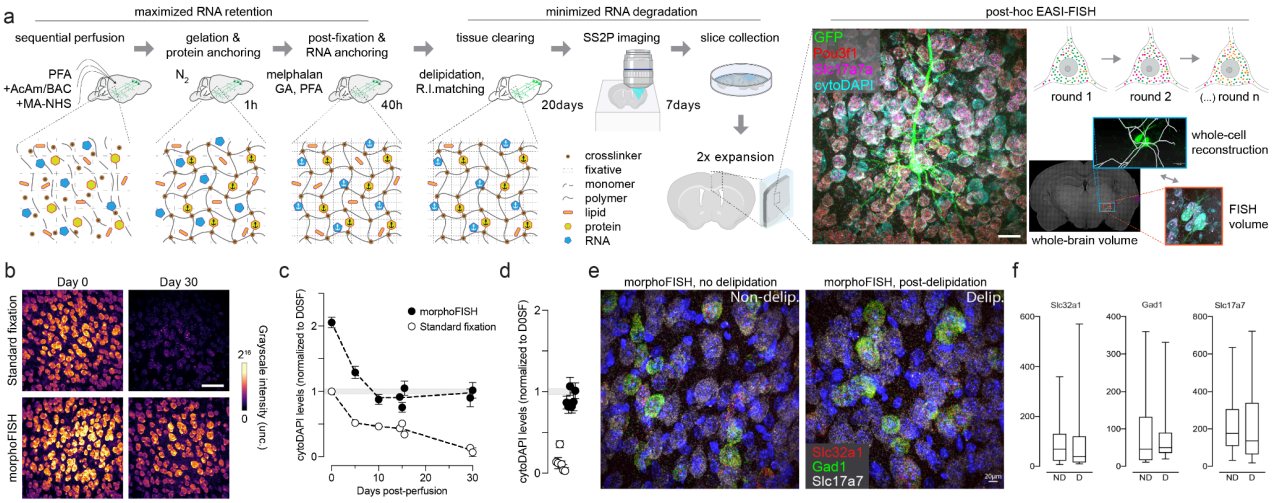


Figure. 2. morphoFISH allows for sensitive detection of transcripts in cleared, imaged tissue.

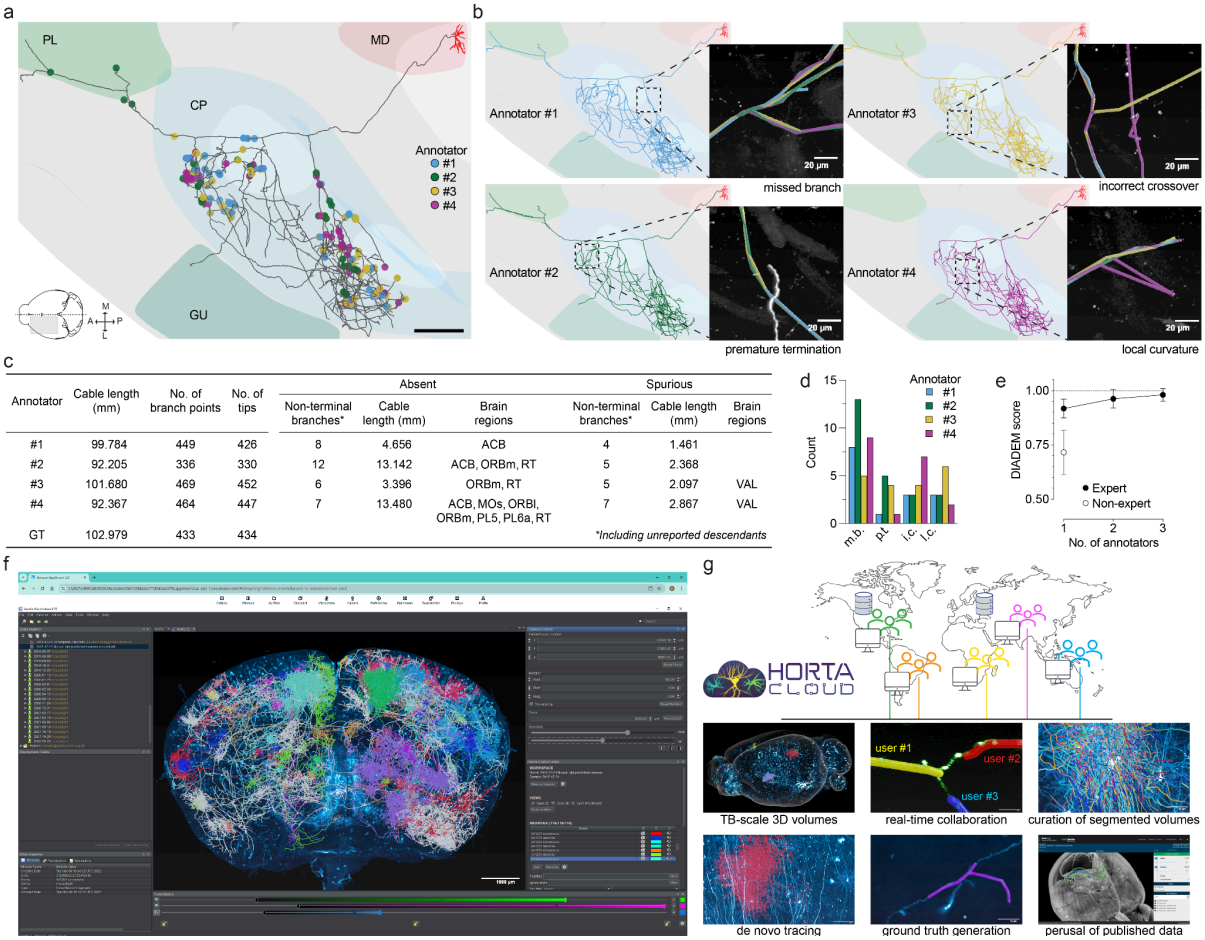
a. Diagram of the morphoFISH protocol. Paraformaldehyde (PFA), acrylamide (AcAm) monomers, AcAm cross linker (BAC), and amine-reactive MA-NHS are sequentially perfused into a brain with sparsely labeled neurons expressing a fluorescent protein (FP). After brain dissection gel is polymerized and proteins are covalently anchored to the gel. During the post-fixation step, PFA fixation is complemented with glutaraldehyde (GA), and RNA is cross-linked to the scaffold of gel/proteins by the alkylating agent melphalan. Next, the whole brain is cleared by removal of lipids and index-matched to a high refractive index solution, and imaged on a serial-sectioning two-photon (SS2P) microscope. Duration of individual steps is indicated. After imaging, sections are harvested and multiple rounds of EASI-FISH performed on expanded sections with FP positive somata and imaged by light-sheet microscopy. Preservation of FP signal during EASI-FISH enables labeled neurons to be mapped back to their original location in the SS2P whole brain volume, and assigned to its morphological reconstruction.

b. Conceptual diagram of RNA anchoring to the hydrogel scaffold through the synergistic action of melphalan and GA.

c,d. Cytosolic DAPI-stained RNA (cytoDAPI) monitored in 200µm brain slices left at room temperature for 30 days. **c.** Representative images (scale bar: XXum) **d.** Time-course quantification**.** morphoFISH allows for a two-fold detection of RNA relatively to Day 0 standard fixation protocols (D0SF). While loss of RNA occurs in both conditions, the high-retention provided by morphoFISH yields comparable D0SF levels at day 30, while RNA is virtually undetected under standard fixation. Individual data points reflect measurements from 300-500 cells per slice (mean±SD). Dashed line connects means between sampled days.

e. Quantification of cytoDAPI levels after a mock imaging experiment in which cleared tissue (delipidated and R.I matched) was left on the imaging tray of the SS2P rig for 5 days.

f,g. Impact of delipidation on morphoFISH transcript detection for Slc32a1, Gad1, and Slc17a7 (counts / cell). **f**. Representative images (scale bar: XXµm) **g**. Spot quantification for segmented cells in **f**.



c

Annotator	Cable length (mm)	No. of branch points	No. of tips	Absent			Spurious		
				Non-terminal branches*	Cable length (mm)	Brain regions	Non-terminal branches*	Cable length (mm)	Brain regions
#1	99.784	449	426	8	4.656	ACB	4	1.461	
#2	92.205	336	330	12	13.142	ACB, ORBm, RT	5	2.368	
#3	101.680	469	452	6	3.396	ORBm, RT	5	2.097	VAL
#4	92.367	464	447	7	13.480	ACB, MOs, ORBl, ORBm, PL5, PL6a, RT	7	2.867	VAL
GT	102.979	433	434						

*Including unreported descendants

Figure 3. Axonal reconstructions guided by quantitative indicators of rigor.

a-d. The axonal arbor of an ipsilateral mediodorsal thalamic neuron was independently reconstructed by multiple expert annotators and tracing errors quantified.
a. Ground-truth (GT) obtained from the four individual reconstructions. Markers highlight the locations at which major errors occurred. Note that when subsidiary branches are involved, only the first error in the parental branch is marked. Prelimbic area (PL), Caudoputamen (CP), Mediodorsal nucleus of thalamus (MD), and gustatory areas (GU) are highlighted. Dendrites are displayed in red and were excluded from analysis. Bottom-left diagram highlights the location of the neuron in the mouse brain. Midline (dashed line) and medio-lateral/anterior-posterior axes (M-L/A-P) are indicated.
b. Result of individual annotators. Inserts highlight the four most common types of detected errors (see **Methods**).
c. Morphometric properties of axonal reconstructions in **a,b**. Table includes sections that were not traced despite being present in GT ("Absent"), and sections from neighboring cells that were traced but do not exist in GT ("Spurious"). Brain regions (as defined in Fig. 1a) associated with at least one axonal termination are listed. Note that non-terminal branches typically ramify into several subsidiaries containing downstream errors not included in this analysis.
d. Absolute frequencies of categorized errors grouped by annotator: missed branch (m.b.), premature termination (p.t.), incorrect crossover (i.c.), and local curvature (l.c.).

e. Increase in accuracy as a function of the number of expert annotators converging into GT, as assessed by pairwise DIADEM scores (mean±SD) against the GT consensus. For context, scores from 12 non-expert annotators are also included. Scores were computed using default settings without adjustment of parameters.

f. Screenshot of the HortaCloud user interface running in a web browser. Quantitative error analysis guided its development to favor reconstruction accuracy.

g. Conceptual illustration of HortaCloud's architecture. Centralized servers ("instances") host shared workspaces (directories that aggregate 3D volumes, metadata, and annotations) that can be accessed concurrently by multiple users ("clients") from internet connected locations. The ability of users to access and modify workspaces is managed by a permissions framework. Vignettes highlight key system capabilities.

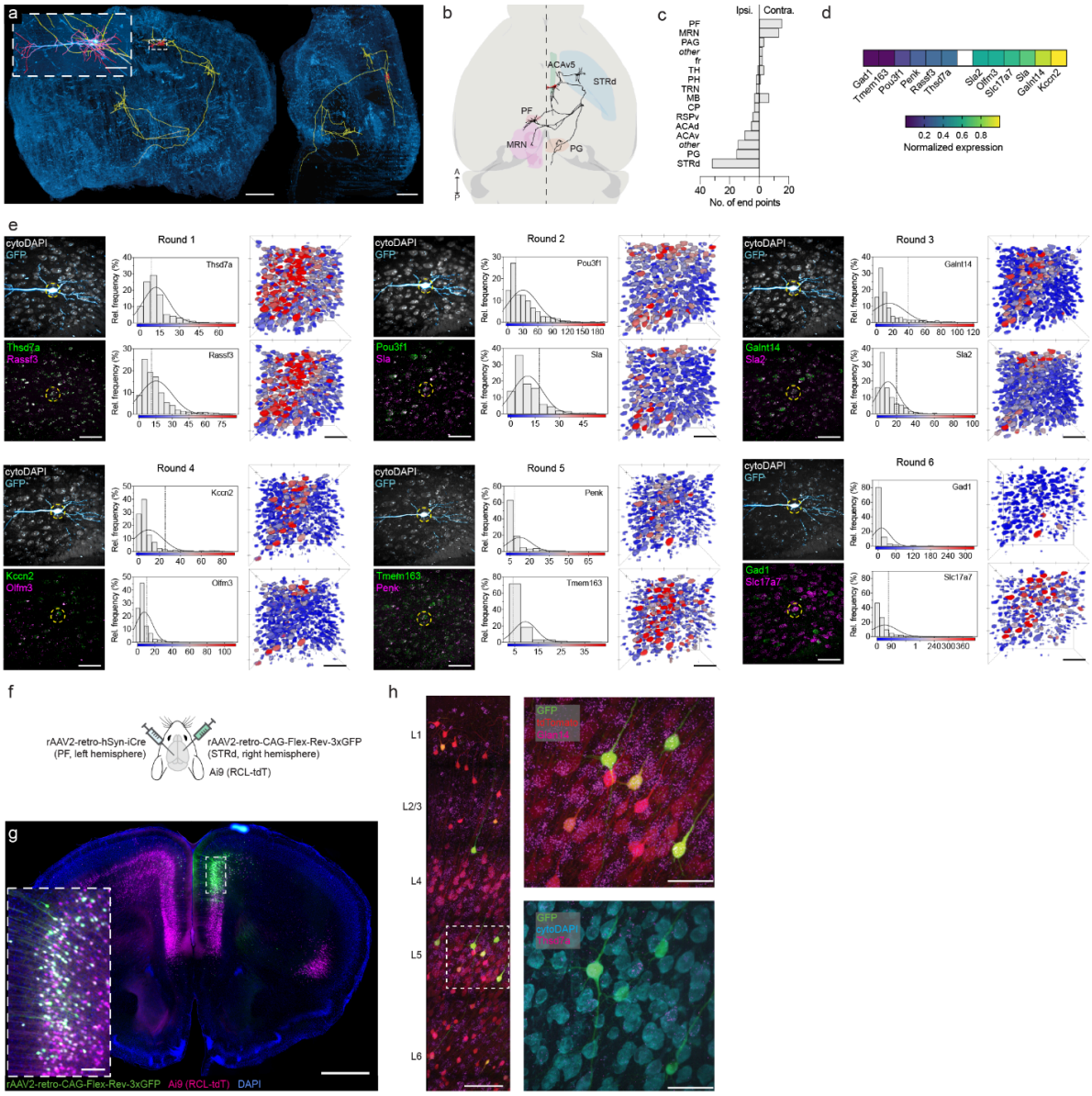


Figure 4 Molecular characterization of a novel class of cingulate neurons.

a. Coronal (left) and sagittal (right) views of a ventral anterior cingulate area layer 5 (ACAv5) neuron (AA1624) traced in HortaCloud. Dendrites are magnified in insert and highlighted in red, with axon in yellow. Scale bars: 1mm (whole brain), 100µm (magnified insert).

b. Dorsal view of the same cell in CCF coordinates, depicting ipsilateral projections to the dorsal striatum (STRd) and pontine gray (PG), and contralateral projections to the parafascicular nucleus (PF) and the midbrain reticular nucleus (MRN). Dashed line highlights midline. Anterior-posterior (A–P) axis is indicated.

c. Frequency histogram of end points of the same neuron grouped by brain areas and split by ipsilateral/contralateral hemisphere. “Other” aggregates brain areas associated with a single end point or non-cellular regions. Abbreviations according to CCF nomenclature: ACAd: Anterior cingulate area dorsal part; fr: fasciculus retroflexus; MB: Midbrain; MRN: Midbrain reticular nucleus; PAG: Periaqueductal gray; PH: Posterior hypothalamic nucleus; TH: Thalamus; TRN: Tegmental reticular nucleus.

d. Summary of the gene-expression profile of AA1624 (detailed in e.). Reported values are normalized to the 10th and 90th percentiles of all cells probed.

e. morphoFISH characterization of AA1624 (12 transcripts probed across 6 rounds). For each round, a field of view centered

on the cell's soma was imaged for cyto-DAPI, GFP, and two HCR-probed transcripts. A 100μm maximum intensity Z-projection is shown. Histograms depict counts of HCR spots per cytoDAPI segmented somata. Vertical line highlights the position of the cell. Overlaid gaussian fit reports mean and standard deviation. Heatmaps depict cytoDAPI segmented somata and color-coded by the number of HCR spot counts, scaled to the lowest and highest count detected in the volume.

f. Diagram of dual-stereotaxic injection, in which a Ai9 (RCL-tdT) reporter mouse was injected with rAAv2-retro-CAG-Flex-Rev-3xGFP in STRd (right hemisphere), and AAv2-retro-hSyn-iCre in PF (left hemisphere).

g. Expression pattern of dual-stereotaxic injection in the PFC (see Fig. S3 for complete pattern). Magnified insets depict somata of STRdPSI/PFCONTRA pGFP+/tdTom+ cells in the ACA. Scale bars: 1mm and 100μm (inset).

h. Molecular characterization of GFP-positive cells in ACAv5 described in f,g. Magnified insert highlights high levels of Galnt14 expression in GFP-positive cells relative to Thsd7a expression

Supplementary Information

Attached

References

1. Zeng, H. What is a cell type and how to define it? *Cell* **185**, 2739–2755 (2022).
2. Oh, S. W. *et al.* A mesoscale connectome of the mouse brain. *Nature* **508**, 207–214 (2014).
3. Winnubst, J. *et al.* Reconstruction of 1,000 projection neurons reveals new cell types and organization of long-range connectivity in the mouse brain. *Cell* **179**, 268–281.e13 (2019).
4. Chen, X. *et al.* High-Throughput Mapping of Long-Range Neuronal Projection Using In Situ Sequencing. *Cell* **179**, 772–786.e19 (2019).
5. Greig, L. C., Woodworth, M. B., Galazo, M. J., Padmanabhan, H. & Macklis, J. D. Molecular logic of neocortical projection neuron specification, development and diversity. *Nat. Rev. Neurosci.* **14**, 755–769 (2013).
6. Chen, X. *et al.* Whole-cortex in situ sequencing reveals input-dependent area identity. *Nature* (2024) doi:10.1038/s41586-024-07221-6.
7. Gao, L. *et al.* Single-neuron projectome of mouse prefrontal cortex. *Nat. Neurosci.* **25**, 515–529 (2022).
8. Qiu, S. *et al.* Whole-brain spatial organization of hippocampal single-neuron projectomes. *Science* **383**, eadj9198 (2024).
9. Fang, R. *et al.* Conservation and divergence of cortical cell organization in human and mouse revealed by MERFISH. *Science* **377**, 56–62 (2022).
10. BRAIN Initiative Cell Census Network (BICCN). A multimodal cell census and atlas of

the mammalian primary motor cortex. *Nature* **598**, 86–102 (2021).

11. Muñoz-Castañeda, R. *et al.* Cellular anatomy of the mouse primary motor cortex. *Nature* **598**, 159–166 (2021).
12. Tasic, B. *et al.* Shared and distinct transcriptomic cell types across neocortical areas. *Nature* **563**, 72–78 (2018).
13. Economo, M. N., Winnubst, J., Bas, E., Ferreira, T. A. & Chandrashekar, J. Single-neuron axonal reconstruction: The search for a wiring diagram of the brain. *J. Comp. Neurol.* **527**, 2190–2199 (2019).
14. Wang, Y. *et al.* EASI-FISH for thick tissue defines lateral hypothalamus spatio-molecular organization. *Cell* **184**, 6361–6377.e24 (2021).
15. Brox, L. W., Gowans, B. & Belch, A. L-phenylalanine mustard (melphalan) uptake and cross-linking in the RPMI 6410 human lymphoblastoid cell line. *Cancer Res.* **40**, 1169–1172 (1980).
16. Susaki, E. A. *et al.* Advanced CUBIC protocols for whole-brain and whole-body clearing and imaging. *Nat. Protoc.* **10**, 1709–1727 (2015).
17. Klingberg, A. *et al.* Fully automated evaluation of total glomerular number and capillary tuft size in nephritic kidneys using lightsheet microscopy. *J. Am. Soc. Nephrol.* **28**, 452–459 (2017).
18. Huang, J. *et al.* A cationic near infrared fluorescent agent and ethyl-cinnamate tissue clearing protocol for vascular staining and imaging. *Sci. Rep.* **9**, 521 (2019).
19. Staudt, T., Lang, M. C., Medda, R., Engelhardt, J. & Hell, S. W. 2,2'-thiodiethanol: a new water soluble mounting medium for high resolution optical microscopy. *Microsc. Res. Tech.* **70**, 1–9 (2007).

20. MouseLight Project Team & Ferreira, T. A. morphoFISH Protocols. *protocols.io* (2024) doi:10.17504/protocols.io.q26g7pej3gwz/v1.

21. Earl, C. C., Smith, M. T., Lease, R. A. & Bundy, B. C. Polyvinylsulfonic acid: A Low-cost RNase inhibitor for enhanced RNA preservation and cell-free protein translation. *Bioengineered* **9**, 90–97 (2018).

22. Economo, M. N. *et al.* A platform for brain-wide imaging and reconstruction of individual neurons. *Elife* **5**, e10566 (2016).

23. Gillette, T. A., Brown, K. M. & Ascoli, G. A. The DIADEM metric: comparing multiple reconstructions of the same neuron. *Neuroinformatics* **9**, 233–245 (2011).

24. Zhang, L. *et al.* Collaborative augmented reconstruction of 3D neuron morphology in mouse and human brains. *Nat. Methods* 1–11 (2024).

25. Eddison, M., Copeland, M. & Ferreira, T. A. Expansion-assisted iterative fluorescence in situ hybridization (EASI-FISH) for morphoFISH. *protocols. io* (2024) doi:10.17504/protocols.io.5jyl824n6l2w/v1.

26. Yao, Z. *et al.* A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell* **184**, 3222–3241.e26 (2021).

27. Xu, S. *et al.* Behavioral state coding by molecularly defined paraventricular hypothalamic cell type ensembles. *Science* **370**, eabb2494 (2020).

28. Nakano, M., Oenzil, F., Mizuno, T. & Gotoh, S. Age-related changes in the lipofuscin accumulation of brain and heart. *Gerontology* **41 Suppl 2**, 69–79 (1995).

29. Nern, A. *et al.* Connectome-driven neural inventory of a complete visual system. *Neuroscience* (2024).

30. Tervo, D. G. R. *et al.* A designer AAV variant permits efficient retrograde access to

projection neurons. *Neuron* **92**, 372–382 (2016).

31. Arshadi, C., Günther, U., Eddison, M., Harrington, K. I. S. & Ferreira, T. A. SNT: a unifying toolbox for quantification of neuronal anatomy. *Nat. Methods* **18**, 374–377 (2021).
32. Copeland, M., Chandrashekar, J. & Ferreira, T. A. Aqueous (SBiP) Delipidation for Whole Mouse Brain after morphoFISH Perfusion. *protocols. io* (2024) doi:10.17504/protocols.io.rm7vzjz54lx1/v1.
33. Stringer, C. & Pachitariu, M. Cellpose3: one-click image restoration for improved cellular segmentation. *bioRxiv* 2024.02.10.579780 (2024) doi:10.1101/2024.02.10.579780.
34. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
35. Ollion, J., Cochennec, J., Loll, F., Escudé, C. & Boudier, T. TANGO: a generic tool for high-throughput 3D image analysis for studying nuclear organization. *Bioinformatics* **29**, 1840–1841 (2013).
36. Tenshaw, E. *et al.* Reconstructing Complete Neurons in Whole Brains Using HortaCloud. *protocols.io* (2024) doi:10.17504/protocols.io.yxmvm38wnl3p/v1.