An image resource of subdivided *Drosophila* GAL4-driver expression patterns for neuron-level searches

Geoffrey W. Meissner*, Zachary Dorman, Aljoscha Nern, Kaitlyn Forster, Theresa Gibney, Jennifer Jeter, Lauren Johnson, Yisheng He, Kelley Lee, Brian Melton, Brianna Yarbrough, Jody Clements, Cristian Goina, Hideo Otsuna, Konrad Rokicki, Robert R. Svirskas, Yoshinori Aso*, Gwyneth M. Card*, Barry J. Dickson*, Erica Ehrhardt[1], Jens Goldammer[1], Masayoshi Ito, Wyatt Korff*, Ryo Minegishi, Shigehiro Namiki[2], Gerald M. Rubin*, Gabriella Sterne[3], Tanya Wolff, Oz Malkesman; for FlyLight Project Team

Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, United States

*For correspondence: meissnerg@janelia.hhmi.org (GWM); asoy@janelia.hhmi.org (YA); cardg@janelia.hhmi.org (GMC); dicksonb@janelia.hhmi.org (BJD); korffw@janelia.hhmi.org (WK); rubing@janelia.hhmi.org (GMR)

Present address: [1]Institute of Zoology, University of Cologne, Cologne, Germany; [2]Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Japan; [3]University of California, Berkeley, Berkeley, CA, United States

Abstract

Precise, repeatable genetic access to specific neurons via the GAL4/UAS system and related methods is a key advantage of *Drosophila* neuroscience. Neuronal targeting is typically documented using light microscopy of full GAL4 expression patterns, which mostly lack the single-cell resolution required for reliable cell type identification. Here we use stochastic GAL4 labeling with the MultiColor FlpOut approach to generate cellular resolution confocal images at large scale. We are releasing aligned images of 27,000 such adult central nervous systems.

An anticipated use of this resource is to bridge the gap between electron microscopy-identified neurons and light microscopy-based intersectional genetic approaches such as the split-GAL4 system. Identifying the individual neurons that make up each GAL4 expression pattern improves the prediction of which GAL4 enhancer fragments best combine via split-GAL4 to target neurons of interest. To this end we have developed the NeuronBridge search tool, which matches these light microscope neuronal images to neurons in the recently published FlyEM hemibrain. This work thus provides a resource and search tool that will significantly enhance both the efficiency and efficacy of split-GAL4 targeting of EM-identified neurons and further advance *Drosophila* neuroscience.

Introduction

Many experimental approaches to understanding the nervous system require the ability to repeatedly target specific neurons in order to efficiently explore their anatomy, physiology, gene expression or function. In *Drosophila melanogaster* the dominant approaches to targeting cells have been GAL4/UAS and related binary systems (Brand & Perrimon, 1993; Lai & Lee, 2006; Pfeiffer, et al., 2010; Potter, et al., 2010). The GAL4 protein, expressed from one transgene, binds upstream activation sequence (UAS) elements inserted in a separate transgene and activates the expression and translation of an adjacent functional protein. An extensive toolkit of UAS transgenes has been developed (reviewed in Guo, et al., 2019). Large collections of GAL4 driver lines have been created, including collections (referred to here as "Generation 1" or "Gen1" GAL4 lines) in which GAL4 expression is typically controlled by 2 to 4 kilobase fragments of enhancer and promoter regions (Pfeiffer, et al., 2008; Jenett, et al., 2012; Tirian & Dickson, 2017). Published image libraries of the expression patterns of these GAL4 lines are available and provide a basis for visual or computational searches for driver lines expressed in cell populations of interest.

Despite these extensive resources, obtaining precise experimental access to individual neuronal cell types remains challenging. A GAL4 driver line from one of the above collections is typically expressed in tens or more neuronal cell types and even more individual neurons, which is not sufficiently specific for many experiments. Several intersectional approaches have been designed to improve targeting specificity (reviewed in Guo, et al., 2019), the most widely used of which is the split-GAL4 system (Luan, et al., 2006; Pfeiffer, et al., 2010). In brief, to create a split-GAL4 driver the activation domain (AD) and DNA binding domain (DBD) of GAL4 are individually placed under control of separate enhancer fragments. The AD and DBD are attached to leucine zipper motifs that further stabilize binding. Only in those neurons where both enhancer fragments are active is a functional GAL4 reassembled to activate UAS, resulting in a positive intersection between enhancer expression patterns. The split-GAL4 system provides the required targeting specificity and has been used at an increasingly large scale (e.g. Gao, et al., 2008; Tuthill, et al., 2013; Aso, et al., 2014; Wu, et al., 2016; Namiki, et al., 2018; Wolff & Rubin, 2018; Dolan, et al., 2019; Davis et al., 2020), but good split combinations remain challenging to predict.

Split-GAL4 construction typically begins with the identification of GAL4 driver lines with expression in the cell type of interest. While the stereotyped shape of fly neurons can sometimes be distinguished by anatomy, the specific features of a neuron are often obscured by other cells in a GAL4 expression pattern. Several stochastic labeling methods that reveal single cells present in broader expression patterns have been developed (reviewed in Germani, et al., 2017). While large libraries of single cell images exist (Chiang, et al., 2011), these were mainly generated using a few widely expressed GAL4 lines. MultiColor FlpOut (MCFO; Nern, et al., 2015) enables the labeling of stochastic subsets of neurons within a GAL4 or split-GAL4 pattern in multiple colors. Labeling a GAL4 pattern using MCFO allows for the efficient determination of a significant fraction of the neurons present within it.

The need for resources to map single cell morphologies to genetic tools (GAL4 lines) has become more urgent due to recent advances in connectomics. Comprehensive electron microscopy (EM) mapping of specific brain regions or whole nervous systems is transforming neuroscience (e.g. Zheng, et al., 2018; Maniates-Selvin, et al., 2020; Scheffer, et al., 2020) by providing anatomy at unparalleled resolution, near complete cell type coverage, and connectivity information. However, leveraging these new datasets to understand more than pure anatomy will be greatly facilitated by the ability to genetically

target specific neurons and circuits. Light microscopy (LM) data also complement EM datasets by revealing features outside a reconstructed EM volume or by providing independent validation of cell shapes with a greater sample size. To integrate these requires datasets and methods for matching EM neurons with LM-derived GAL4/split-GAL4 data.

Recently developed techniques allow searching for neuron shapes (including neuron fragments, whole neurons, or overlapping groups of neurons) in coregistered LM and EM data. Two leading approaches are NBLAST (Costa, et al., 2016), which performs comparisons between segmented neurons, and color depth MIP search (Otsuna, et al., 2018), which efficiently compares bitmap images using color to represent depth within the samples. Advanced anatomical templates such as JRC2018 improve point-to-point mapping between samples and modalities (Bogovic, et al., 2019). These search tools and templates bridge the EM/LM gap but require single-cell-level image collections that cover many neurons present within Gen1 GAL4 patterns to reach their maximum utility. In particular, to identify multiple Gen1 GAL4s that can be combined to make a split-GAL4 driver, the morphologies of individual neurons within many GAL4 lines must be available.

Here we used MCFO to dissect Gen1 GAL4 line patterns at scale to create a resource for linking EM-reconstructed neurons to GAL4 lines, and to improve the process of making split-GAL4 reporters to target neurons, whether they were first identified in EM or LM. We therefore focused on 4562 Gen1 GAL4 lines that have already been converted into split-GAL4 hemidrivers, performing two rounds of MCFO labeling to improve coverage of neurons. In the first, completed, phase we employed Flp-recombinase drivers with weak pan-neuronal expression (R57C10-Flp MCFO reporters) to induce MCFO labeling. In the second phase we are expanding labeling of a subset of lines using temperature-induced Flp-expression (hs-Flp). While the second phase remains in progress, we are presently releasing the first phase of the data, along with the NeuronBridge tool to search between the FlyEM hemibrain, Gen1 MCFO data, and published split-GAL4 data.

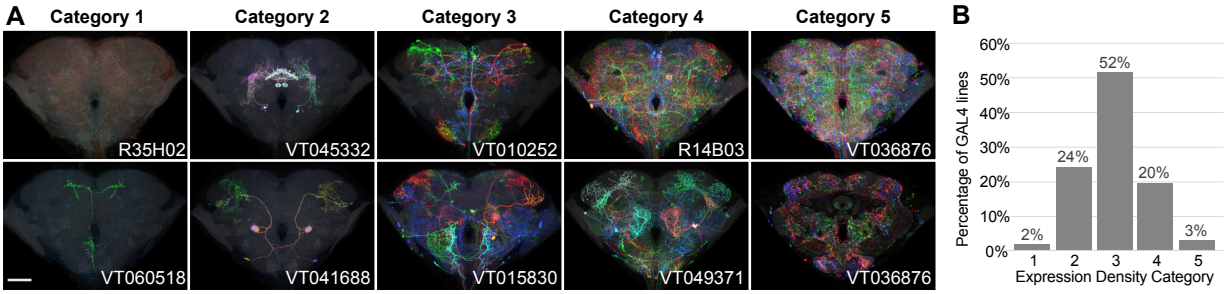

Figure 1. Generation 1 MCFO expression density categories.
(A) Two example brain maximum intensity projections (MIPs) are shown for each expression density category, except Category 5, where a single brain is shown both as a MIP and a single confocal slice through its center. Qualitative categorization was manually performed on a line level based on the full CNS expression pattern. Category 1 lines contained no visible neurons or only commonly repeated ones. Categories 2 to 4 labeled identifiable neurons with increasing density. Category 5 lines had such dense expression that the immunohistochemical labeling approach failed to fully label the center of the brain. Category 1 and 5 lines were generally excluded from imaging and the collection as a whole. Scale bar, 50 μm.
(B) The frequency distribution of lines within the different expression density categories are shown. Sample size is all 4919 lines considered for inclusion in either phase. 95% of lines were within the desired range.

Results

Phase 1

MCFO labeling of *Drosophila* neurons was performed with a pan-neuronal Flp recombinase (R57C10-Flp) on 4562 Generation 1 GAL4 lines in Phase 1. We generated images of 27,226 central brains and 26,512 ventral nerve cords (VNCs) from 27,729 flies. The central nervous system was typically dissected from six flies per line. A medium-strength Flp transgene (*R57C10-Flp2::PEST in attP18*; Nern, et al., 2015) was used for almost all lines, regardless of GAL4 expression density, yielding a wide range of neuronal labeling in each MCFO sample. 238 of the sparser lines were crossed to an MCFO reporter with a stronger Flp transgene (*R57C10-FlpL in su(Hw)attP8*), and 71 lines were crossed to both reporters. A "hybrid" labeling protocol was used, in which a chemical tag (Brp-SNAP and SNAP-tag ligand) labels the neuropil reference, and immunohistochemistry of MCFO markers labels specific GAL4 neurons (Kohl, et al., 2014; Nern, et al., 2015; Meissner, et al., 2018). Chemical tag labeling of the Brp reference is not as bright as Brp antibody staining with nc82, but is more consistent and has lower background.

Imaging was optimized in several ways to maximize throughput. We focused on the central brain and VNC due to production constraints, choosing to exclude the optic lobes. Although the optic lobe contains more neurons than the central brain, it has a repetitive structure and many of its cell types have been anatomically described, often at both the light and EM level (e.g. Fischbach and Dittrich, 1989; Morante and Desplan, 2008; Takemura, et al., 2013; Nern, et al., 2015; Takemura, et al., 2015). Collections of split-GAL4 driver lines for many optic lobe cell types are also available (Tuthill, et al., 2013; Wu, et al., 2016; Davis, et al., 2020). We also note that many neurons that connect the optic lobe with the central brain can still be identified in our dataset based on their central brain arborizations. The optic lobe anatomy of such cells could be further characterized in follow-up experiments with the identified GAL4 lines.

All imaging was performed using Zeiss LSM 710 and 780 laser scanning confocal microscopes. We used 40x oil objectives due to their good axial resolution and their field of view covering the central brain. The VNC was imaged in two tiles and stitched together (Yu & Peng, 2011). The isotropic 0.44x0.44x0.44 micron voxel size was based on the approximate maximum axial resolution of the objective in our setup.

GAL4 lines were qualitatively categorized by density of expression within the central brain and VNC, ranging from Category 1 yielding no unique neurons per sample, to Category 5 being so dense that it overwhelmed our immunohistochemical approach, leaving a shell of partially labeled neurons around the outside of each sample (Figure 1A). Category 2 lines were characterized by sparse, easily-separable neurons, whereas Category 3 yielded denser but identifiable neurons. Category 4 displayed densely-labeled neurons that were challenging to distinguish. Most lines ranged between Categories 2 and 4 (Figure 1B).

Phase 2

In order to further increase the number of identifiable neurons labeled across GAL4 lines, a subset of lines was examined again with altered parameters. Phase 2 of the project is expected to generate images of an additional 18,000 to 26,000 central brains and 6,000 to 9,000 VNCs from the combination of 6,000 Category 2 samples and 12,000 to 19,000 Category 3 samples (Figure 2A). For Phase 2, GAL4 expression density was optimized by (1) selecting lines with expression most likely useful for split halves, (2) adjusting MCFO parameters to maximize separable neurons obtained per sample, and (3) limiting brains and VNCs processed per line to minimize the diminishing returns associated with oversampling.

A uniform labeling strategy was used across all lines in Phase 1, resulting in a wide range of neuron densities. For Phase 2, we considered the optimal GAL4 line expression density both for yield of neurons in MCFO and utility as split halves. Category 1 and 5 lines appear to be of low value for targeting specific neurons in the adult central brain and VNC and were therefore excluded from further work. High neuron density within Category 4 means that although the theoretical neuron yield from each sample is high, our ability to distinguish individual neurons within the pattern with current technology is low (although future improvements to neuron segmentation approaches are expected to improve yields).

A

Phase	Flp	Density Category	Brains per line	VNCs per line	Lines	Total brains	Total VNCs
Phase 1	R57C10-Flp	Cat 2	6	6	1126	6498	6276
		Cat 3	6	6	2461	14992	14637
		Cat 4	6	6	904	5464	5314
Phase 2 (planned)	hs-Flp	Cat 2	6	2	1100	6700	2200
		Cat 3	9	2	>1300	>11000	>3700

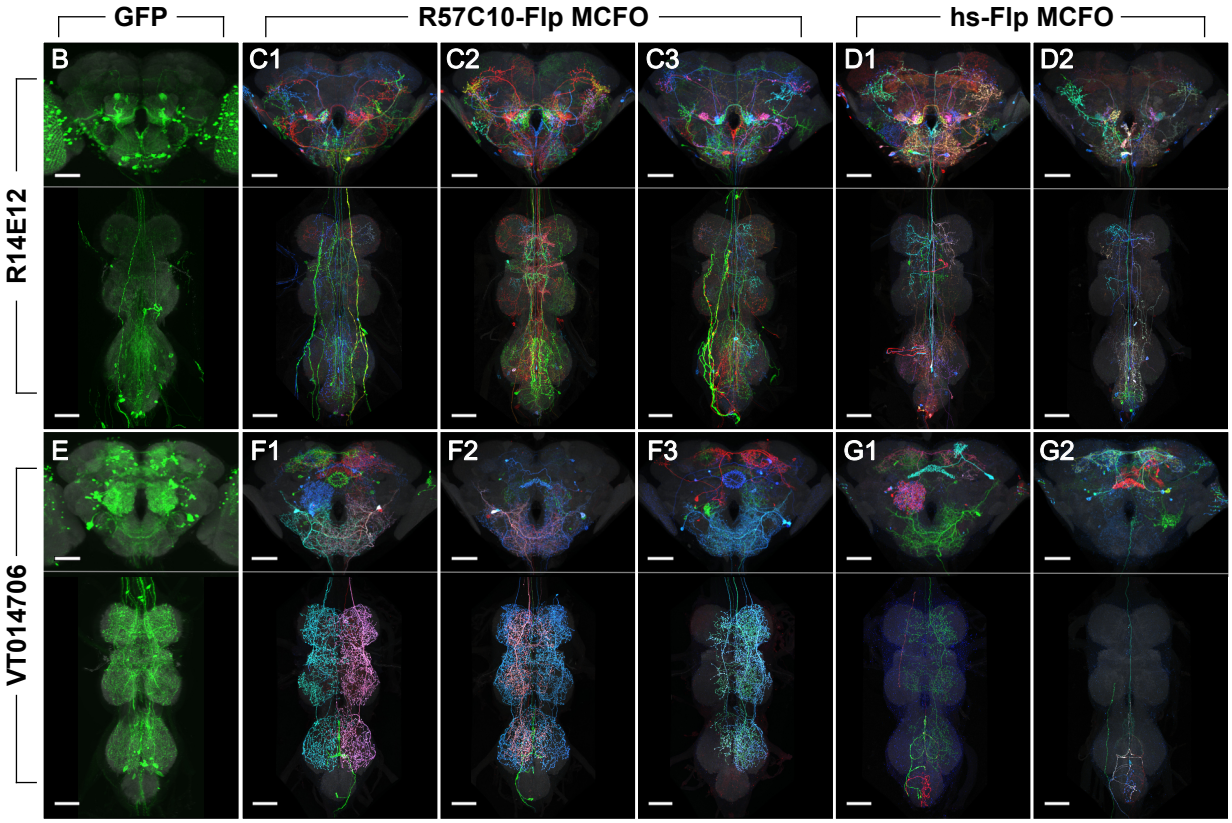

Figure 2. Phase 1 & 2 overview and labeling examples.
(A) The approaches for each phase and density category are tabulated. Phase 1 is completed. Phase 2 is in progress.
(B-D) R14E12-GAL4 in attP2 crossed to (B) pJFRC2-10XUAS-IVS-mCD8::GFP, (C) R57C10-Flp MCFO, or (D) hs-Flp MCFO. Adult CNS MIPs are shown, with neuropil reference in gray and neuronal signal in green (B) or multiple colors (C-D). Multiple examples are shown for C-D.
(E-G) VT014706-GAL4 in attP2 crossed to (E) pJFRC2-10XUAS-IVS-mCD8::GFP, (F) R57C10-Flp MCFO, or (G) hs-Flp MCFO. Adult CNS MIPs are shown, with neuropil reference in gray and neuronal signal in green (E) or multiple colors (F-G). Multiple examples are shown for F-G. Scale bars, 50 μm.

It is also not immediately obvious how useful these dense lines are as split halves. As an extreme example, a perfectly labeled pan-neuronal line would result in many MCFO neurons per sample, but it would not be a useful split half in most scenarios.

Utility of lines within the different density categories was evaluated further by examining the categories of lines previously tested to create existing FlyLight split-GAL4 lines. We considered two criteria: (1) The lines initially selected by investigators to use in previous attempts to make split-GAL4 lines provides an indication of what densities of lines were judged to be potentially useful. (2) The lines that appear in our current collections of high quality stable split-GAL4 lines provides a measure of which lines proved to be to be most useful. We compared the distribution of density categories of these two sets of lines to the overall distribution of all imaged lines to assess their relative utility. We found that Category 3 lines (not shown) were most commonly used, as expected based on their prevalence (Figure 1B). Category 2, 3, and 4 lines were used in successful (stabilized) split-GAL4 combinations, with Category 2 slightly more likely to proceed from initial screening to stabilization. The subsequent focus is therefore on Categories 2 and 3 in Phase 2 to avoid the denser Category 4 lines that present more challenges using current approaches.

Heat-shock Flp (hs-Flp) was used in Phase 2 rather than 57C10-Flp from Phase 1 (Figure 2). While both R57C10-Flp and hs-Flp are theoretically expected to label all neurons, in practice each is likely to have subtle biases as previously proposed (Nern, et al., 2015; see also below). By switching Flp enhancers in Phase 2, we attempted to mitigate the impact of these biases. The 37C heat shock duration for hs-Flp was optimized for each density category. Prior results reported by Nern, et al. (2015) indicated that heat shock effectiveness is nonlinear: limited to background activity up to ~10 minutes, a somewhat linear range between 10 and 20 minutes, and gradually diminishing returns up to ~40 minutes; heat shocks longer than an hour begin to harm fly survival. We chose a heat shock duration of 40 minutes for Category 2 lines to yield as many neurons as possible per sample. For Category 3, a 13 minute heat shock provided the desired labeling density similar to Category 3 in Phase 1.

We attempted to compensate for sex differences in a manner similar to compensating for heat shock enhancer biases. In Phase 1 this was accomplished by randomly choosing one sex for half of the lines and the other sex for the rest. In Phase 2, we switched each line to the opposite sex to increase the chance of detecting sex-specific neurons present in each line.

Coverage and diminishing returns

Every additional MCFO brain examined for a given GAL4 line results in diminishing returns of additional unique neurons labeled, making it inefficient to obtain every neuron within each line. The diminishing returns within a line are especially pronounced for the sparsest lines, for which we can use higher Flp activity to label a greater fraction of available GAL4 neurons per sample without saturating detection. Thus, in Phase 2 we processed fewer samples for Category 2 GAL4 lines than for Category 3.

In addition to diminishing returns within each GAL4 line, there are diminishing returns within each region of the CNS. The adult central brain (including subesophageal zone) is estimated to contain approximately 30,000 neurons, compared to 15,000 in the VNC (Simpson, 2009; see also Yu, et al., 2013 for a lower bound), suggesting earlier diminishing returns in the VNC. To optimize the number of brains and VNCs processed per line and category, we estimated 'coverage' of each region, inspired by the metric for sequencing data (reviewed in Sims, et al., 2014). Here we define coverage as the number of identifiable neurons labeled in a region from all samples, divided by the total neurons

present in that region of a single fly. Greater coverage of a region should be associated with greater diminishing returns, so we attempted to obtain similar coverage of each region and thus minimize overall diminishing returns.

We made rough estimates of Phase 1 central brain and VNC coverage by counting neurons labeled in a few lines per density category (3 to 5 lines per category, each with 4 to 6 brains & VNCs; Table S2; and see Materials and Methods). Similar numbers of identifiable neurons were counted in the central brain and VNC. However, because whole CNSs were processed in Phase 1 and the brain is estimated to contain double the neurons of the VNC, the resulting ~8x coverage of the brain was lower than the ~19x for the VNC. Thus, we focused Phase 2 more heavily on the brain than the VNC, imaging on average 6.0 brains in Category 2 or 9.1 brains in Category 3 and 2.5 VNCs per line. We estimate reaching approximately 22x to 30x coverage of both the central brain and VNC from both phases of the effort combined. We view the exact coverage number as less important than the general approach of balancing relative neuron yield from the different regions.

MCFO labeling observations

The large number of lines processed under mostly uniform MCFO conditions provided the opportunity to observe, at scale, some features of MCFO labeling with the specific Flp recombinase drivers used here. Similar observations were noted previously (Nern, et al., 2015).

As with R57C10-GAL4, which contains the same fragment of the *synaptobrevin* enhancer region (Pfeiffer, et al., 2008), R57C10-Flp is thought to be exclusively expressed in postmitotic neurons. In contrast, hs-Flp is expected to label most if not all cells in the fly,

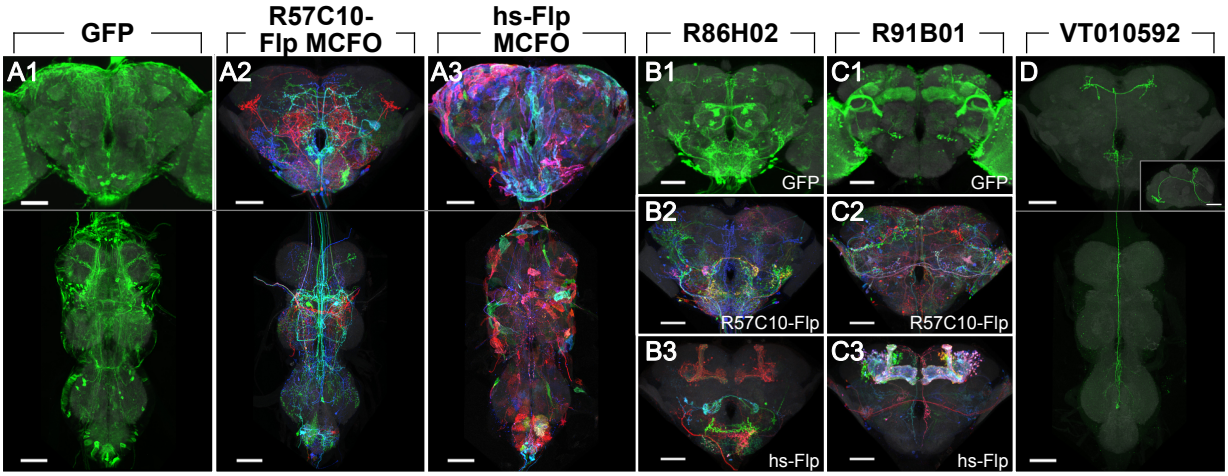

Figure 3. Comparison between Phase 1 and Phase 2 approaches.
(A) Glia are seen with VT008658-GAL4 in attP2 crossed to (A1) pJFRC2-10XUAS-IVS-mCD8::GFP and (A3) hs-Flp MCFO, but not (A2) 57C10-Flp MCFO.
(B) Mushroom body labeling is not seen with R86H02-GAL4 in attP2 crossed to (B1) pJFRC2-10XUAS-IVS-mCD8::GFP or (B2) R57C10-Flp MCFO, but is seen when crossed to (B3) hs-Flp MCFO.
(C) Mushroom body labeling is seen with R91B01-GAL4 in attP2 crossed to (C1) pJFRC2-10XUAS-IVS-mCD8::GFP and (C3) hs-Flp MCFO, but is not seen when crossed to (C2) R57C10-Flp MCFO.
(D) An ascending neuron ("sparse T") is commonly seen with many Gen1 GAL4 lines crossed to different reporters. VT010592-GAL4 in attP2 crossed to R57C10-Flp MCFO is shown as an example. A single neuron channel plus reference are shown for clarity. The inset shows a lateral (y-axis) maximum intensity projection of the brain. All scale bars, 50 µm.

including neurons, glia, and trachea, as reviewed in Ashburner and Bonner (1979), though this will depend on the specific hs-Flp transgene. Thus, glial patterns were obtained in 8% of lines (36 of 460 lines tabulated) in Phase 2 with *pBPhsFlp2::PEST in attP3*. This obscured neurons in maximum intensity projections, but typically did not impair three dimensional visualization or searching, and may prove of use for future glial studies (Figure 3A). For example, the split-GAL4 approach has also been successfully applied to several types of glia in the optic lobe (Davis, et al., 2020).

Kenyon cells making up the mushroom body were labeled at different rates with each reporter. Labeling was scored in a random sample of 10% of the total lines imaged (n=460 lines). Labeling manifested as either distinctly labeled neurons, a relatively faint hazy labeling or both. The mushroom body was much more commonly labeled using hs-Flp MCFO (430 lines, or 93%) than with R57C10-Flp (44 lines, or 10%) or GFP (111 lines, or 24%; Figure 3B-C). The most frequent combination within a line was an unlabeled mushroom body via GFP or R57C10-Flp MCFO, but with hs-Flp labeling (253 lines, or 55%; Figure 3B). Lines were also observed with labeled mushroom bodies using GFP and hs-Flp MCFO, but not R57C10-Flp (59 lines, or 13%; Figure 3C). The MCFO labeling likelihood in the mushroom body is thus on both sides of what is seen with GFP. As the Kenyon cells are well characterized (and thus an unlikely target for new split-GAL4s), compact, and easily identified, this labeling can be ignored except when substantially brighter than other neurons of interest.

A characteristic ascending neuron (sometimes referred to as "sparse T") was observed at very high frequency: in at least one sample from over 60% of lines crossed to either MCFO

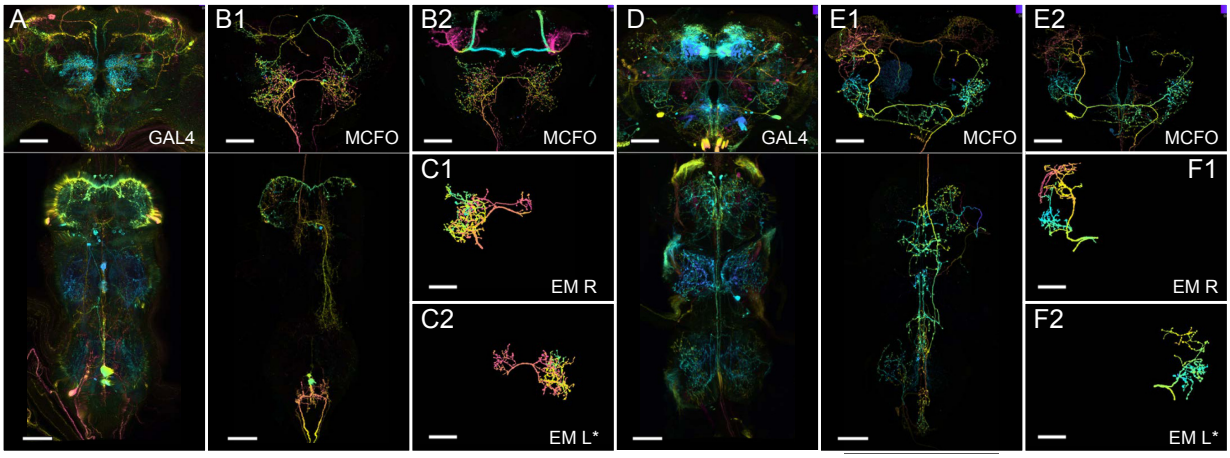

Figure 4. Matching to LM data can help interpret EM reconstructions.
(A) and (D) Overall GAL4 expression pattern of the driver lines used for (B) and (E) shown as color depth MIPs (Otsuna, et al., 2018). Original images are from published datasets (Jenett, et al., 2012; Tirian & Dickson, 2017).
(B1) and (E1) Examples of MCFO labeled cells from these driver lines in the same format. Images include a prominent ascending (B) or descending neuron (E). An additional MCFO labeled cell of these same type, but from a different line, is shown in (B2, E2). Color depth MIPs represent data from one of the three MCFO markers.
(C1, C2) and (F1, F2). Matching EM reconstructions for the two cell types. Both panels show reconstructions from the right side hemibrain; the lower panel is mirrored to facilitate comparison to the LM data.
Driver lines used are VT029752 (A, B1), R89C01(B2), R56H09 (D, E1) and R23E11(E2). Hemibrain body IDs are 1353375816 (C1), 1687321237 (C2), 571346836 (F1), and 1786496543 (F2). All scale bars, 50 μm.

reporter (67 lines in Phase 1 and 64 lines in Phase 2, out of 107 lines scored) and was likely present but obscured in other lines. The greater density of labeling when the GAL4 lines were crossed to GFP made scoring more difficult, yet a similar neuron was seen in 22 of the same 107 lines. This suggests that the high labeling frequency of this neuron in our dataset is a property of the GAL4 collections rather than an artifact of our sampling methods. The neuron(s) has a cell body near the metathoracic ganglion and projections ascending to the anterior then the posterior brain, loosely resembling the letter "T" in MIP images (Figure 3D).

Discussion

We anticipate that optimal use of the recently published FlyEM hemibrain dataset (Scheffer, et al., 2020) will include targeting EM neurons via LM/EM matching and generation of split-GAL4 lines. We therefore are releasing this initial MCFO dataset while finishing the second phase. As described, we have optimized driver line selection, sample preparation, and imaging to yield the maximum identifiable neurons per sample, per line, and across the central brain and VNC.

This image collection makes it possible to identify GAL4 driver lines expressed in identified single neurons using manual or computational searches without the need for new anatomical experiments. The cellular resolution of the data enables many analyses that are impossible with the existing libraries of full GAL4 driver expression patterns. The single cell data are particularly useful for LM/EM matches. While accurate matching of EM reconstructions with single cell LM images can be achieved by direct visual inspection (e.g. Takemura, et al., 2013), automated approaches for image alignment, segmentation, and search are essential for efficient use of these large datasets. All our samples are registered to the JRC2018 brain and VNC templates (Bogovic, et al., 2019). With this data release, we are also making the search tool NeuronBridge (Clements, et al., 2020) publicly available.

NeuronBridge allows the user to perform anatomical similarity searches between published datasets reported by Janelia's FlyLight and FlyEM Team Projects. Searching is based on the color depth MIP approach, allowing direct comparisons of expression similarity in registered images without the need for a complete skeletonization (Otsuna, et al., 2018). To improve matches for denser MCFO data, the color depth MIP approach was extended in several ways (Otsuna, et al., in preparation). These include (1) preprocessing the MCFO images with direction selective local thresholding (DSLT; Kawase, et al., 2015) 3D segmentation to create a separate color depth MIP for each fully connected component; (2) color depth searching using mirrored EM hemibrain neurons as masks and MCFO images as target libraries; and (3) weighting of match scores based on signal outside of the search masks.

These comparisons are currently pre-computed as data is added or updated in NeuronBridge, so searching is fast. Searches can begin at NeuronBridge given a GAL4 line name or EM body ID, or from FlyEM's neuPrint (Scheffer, et al., 2020) and FlyLight's Gen1 MCFO and split-GAL4 anatomy websites, leading directly to potential matches in the complementary modality. Search results are sorted by match quality and displayed for easy comparison. The color depth MIP format is also well suited for fast visual inspection of search results, simplifying the exclusion of false positives, which are difficult to avoid without compromising search sensitivity.

While LM images do not match the synaptic resolution of the EM data, they can provide additional, complementary anatomical information (Figure 4). First, identification of LM matches provides an independent quality check for EM reconstructions. Second, the LM data often includes multiple examples of a cell type and thus provide insights into variable

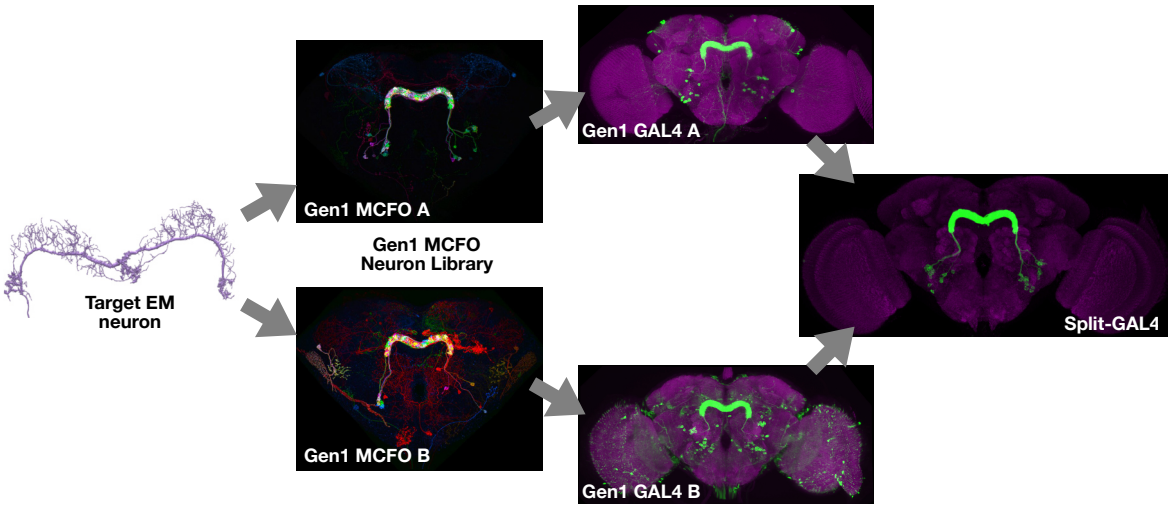

Figure 5. Schematic of anticipated workflow.
An example is shown of the anticipated search process, from a neuron identified via electron microscopy to the creation of a split-GAL4 driver. The example shown includes FlyEM hemibrain body ID 911911004 (Scheffer, et al., 2020), GAL4 lines R46G08-GAL4, R55G08-GAL4, and split-GAL4 SS30295.

features of cell shapes. Finally, except for the optic lobes, our LM data include the full brain and (for many specimens) VNC and thus provide the full shape of cells that are only partly contained in an EM volume. For example, the hemibrain does not fully include neurons that span both brain hemispheres or project to or from the VNC (Figure 4).

This study has optimized the flexible and multipurpose Gen1 MCFO data and search tools for efficiently designing split-GAL4 lines to target EM-identified neurons (Figure 5). Once a neuron of interest is identified in EM, candidate matches in Gen1 MCFO light images can be identified using NeuronBridge. Pairs of corresponding split-GAL4 AD and DBD versions can then be screened for specific split-GAL4 combinations to be used to target the neuron for additional characterization.

The split-GAL4 system has become the dominant approach for targeting individual or subsets of neurons in *Drosophila*. We anticipate that the MCFO data and search tools described here will improve the efficiency of split-GAL4 creation and other enhancer-based approaches to cell type targeting to the adult CNS. The resulting precise targeting will be especially useful for functional studies of neurons and connectomes identified by electron microscopy.

Materials and Methods

Fly stocks

The 4562 Generation 1 GAL4 stocks included in this Phase 1 release (Table S1) were from Jenett, et al. (2012) and Tirian & Dickson (2017). We focused on driver lines with available AD or DBD hemidrivers (Tirian & Dickson, 2017; Dionne, et al., 2018).
'R57C10-Flp MCFO' was *R57C10-Flp2::PEST in attp18; brp::Snap / CyO; pJFRC201-10XUAS-FRT>STOP>FRT-myr::smGFP-HA in VK00005, pJFRC240-10XUAS-FRT>STOP>FRT-myr::smGFP-V5-THS-10XUAS-FRT>STOP>FRT-myr::smGFP-FLAG in*

su(Hw)attP1 / TM6B (JRC stock 3023701; Kohl, et al., 2014; Nern, et al., 2015) for 94% of such samples, with *R57C10-FlpL in su(Hw)attP8; brp::Snap / CyO; pJFRC201-10XUAS-FRT>STOP>FRT-myr::smGFP-HA in VK00005, pJFRC240-10XUAS-FRT>STOP>FRT-myr::smGFP-V5-THS-10XUAS-FRT>STOP>FRT-myr::smGFP-FLAG in su(Hw)attP1 / TM6B* (JRC stock 3023700; Kohl, et al., 2014; Nern, et al., 2015) used for 6% of samples from sparser lines.
'hs-Flp MCFO' was *pBPhsFlp2::PEST in attP3; brp::Snap / CyO; pJFRC201-10XUAS-FRT>STOP>FRT-myr::smGFP-HA in VK00005, pJFRC240-10XUAS-FRT>STOP>FRT-myr::smGFP-V5-THS-10XUAS-FRT>STOP>FRT-myr::smGFP-FLAG in su(Hw)attP1 / TM6B* (JRC stock 3023951; Kohl, et al., 2014; Nern, et al., 2015).
pJFRC2-10XUAS-IVS-mCD8::GFP was from (Pfeiffer, et al., 2010).
Split-GAL4 stock SS30295 consists of *R55G08-p65ADZp in attP40* and *R45G06-ZpGdbd in attP2* (Turner-Evans, et al., 2020).

Fly crosses, heat shock, and dissection
Flies were raised on standard corn meal molasses food, typically in at least partial-brightness 24 hour light. All crosses were performed at 21-25C, with a few exceptions (~2.5% of all samples) performed at 18C when scheduling necessitated. Crosses with hs-Flp in particular were held at 21C until adulthood, when they were heat-shocked at 37C for 40 minutes (Category 2 lines) or 13 minutes (Category 3 lines). Flies were generally dissected at 5-14 days of adulthood, giving time for R57C10-Flp and then MCFO reporter expression.

Tagging and immunohistochemistry
After dissection of the brain or full CNS, samples were fixed for 55 minutes in 2% paraformaldehyde and washed 1 to 4 times for 15 minutes. They were tagged with 2 μM Cy2 SNAP-tag ligand to visualize the Brp-SNAP neuropil the same day, after which immunohistochemistry and DPX mounting followed Meissner, et al. (2018), based on Kohl, et al. (2014) and Nern, et al. (2015).

Imaging and image processing
Imaging was performed using Zeiss LSM 710 and 780 laser scanning confocal microscopes with Plan-Apochromat 40x/1.3 Oil DIC M27 objectives. Confocal stacks were captured at 0.44 μm isotropic resolution to maximize effective z-resolution while limiting the size of the full data set. The field of view was set to the widest 0.7 zoom to fit most central brains in a single tile without cropping. The large field of view resulted in the introduction of lens distortion at the edges of images—most noticeable when two tiles are stitched together—which was corrected before stitching (Bogovic, et al., 2019).

Four-color imaging was configured as described in Nern, et al. (2015). Briefly, two LSM confocal stacks were captured at each location, one with 488 nm and 594 nm laser lines and one with 488 nm, 561 nm, and 633 nm laser lines. Stacks were merged together after imaging. Imaging was performed using Zeiss's ZEN software with a custom MultiTime macro. The macro was programmed to automatically select appropriate laser power for each sample and region, resulting in independent image parameters between samples and between brains and VNCs. Gain was typically set automatically for the 561 nm and 633 nm channels and manually for 488 nm and 594 nm. Imaging parameters were held constant within the two tiles making up each VNC.

The central brain and two VNC tiles (where present) were captured for each sample. After merging and distortion correction, the VNC tiles were automatically stitched together, as described (Yu & Peng, 2011). Brains and VNCs were aligned to the JRC2018 sex-specific and unisex templates using CMTK software, and color depth MIPs were generated (Rohlfing & Maurer, 2003; Otsuna, et al., 2018; Bogovic, et al., 2019).

The image processing pipeline (distortion correction, normalization, merging, stitching, alignment, MIP generation, file compression) was automated using the open-sourced Janelia Workstation software (Rokicki, et al., 2019), which was also used to review the secondary results and annotate lines for publishing. Imagery for published lines was uploaded to AWS S3 (Amazon Web Services) and made available in a public bucket for download or further analysis on AWS. Original LSM (i.e. lossless TIFF) imagery is available alongside the processed (merged/stitched/aligned) imagery in H5J format. H5J is a "visually lossless" format developed at Janelia, which uses the H.265 codec and differential compression ratios on a per-channel basis to obtain maximum compression while minimizing visually-relevant artifacts (see data.janelia.org/h5j).

The open-sourced NeuronBridge tool (Clements, et al., 2020) was created as a cloud-native application which can be easily deployed to support other data sets. Multiple types of results can be represented in NeuronBridge, including precomputed matches, curated matches, and ad-hoc searches based on user data. NeuronBridge was constructed as a single-page application built on the React framework for ease of deployment. The entire application (for viewing precomputed results) can be deployed using only AWS S3 as supporting infrastructure. For ad-hoc searching, NeuronBridge uses only serverless components on AWS to minimize cost. NeuronBridge also takes advantage of the innovative "burst-parallel" compute paradigm (Fouladi, et al., 2019) to massively scale color depth MIP search by leveraging micro VMs (virtual machines) on AWS Lambda, thereby enabling rapid ad-hoc searches across a nominally petabyte-scale dataset.

Quality control and expression density categorization

Samples had to pass quality control at several stages to be included in the final set. Samples lacking visible neuron expression or too dense for IHC were excluded prior to imaging or after image review. Samples were excluded that contained damage, distortion, debris, or low neuropil reference quality causing a failure to align or an error in the image processing pipeline. Samples with minor issues in neuron channels were typically included if neurons could be distinguished.

Selected *Drosophila* lines were qualitatively grouped into Categories 1 through 5 by expression density, primarily using MCFO and less often by full GFP patterns. Category boundaries were initially established based on functional properties. Category 1 and 5 samples were excluded due to lack of information, either no unique neurons or too many to label, respectively. Categories 3 and 4 were separated based on performance of an automatic neuron segmentation algorithm combined with intuition about future segmentation difficulty, such that Category 3 lines were expected to be tractable for segmentation, whereas Category 4 could be challenging. Categories 2 and 3 were divided such that Category 2 mostly contained neurons that could easily be "segmented" by eye, whereas Category 3 had more instances of overlapping neurons that were harder to distinguish.

Coverage

To estimate central brain and VNC coverage for planning purposes, five Category 2 lines, four Category 3 lines, and three Category 4 lines had 4 to 6 samples each scored for their number of neurons (Table S2). Due to the density of Category 4 lines and their exclusion from Phase 2, Category 4 was not included in the overall coverage estimate. Counts are best treated as estimates. To roughly factor in the added difficulty of parsing dense patterns, neuron counts over 10 in each region were divided by the logarithm of the neuron count. For example, 100 neurons became 50 after the adjustment. If the count was under 10, it was assumed that all neurons could be identified. To make estimates for Phase 2, whole GAL4/GFP patterns were counted. It was assumed that Category 2 lines would

label half the total GAL4 pattern in each hs-Flp MCFO sample (with long heat shock), whereas Category 3 lines would label 10% of the total GAL4 pattern in each sample (with shorter heat shock).

Data availability

Gen1 MCFO anatomical images are available at https://data.janelia.org/ctEpc. NeuronBridge search is available at https://data.janelia.org/sk7zb.

Acknowledgements

This work is part of the FlyLight Project Team at Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA. Author order includes the following alphabetical groups: FlyLight Project Team, Janelia Scientific Computing Shared Resource, and Janelia laboratories.

During this effort, the FlyLight Project Team consisted of Gina M. DePasquale, Zachary Dorman, Kaitlyn Forster, Theresa Gibney, Joanna H. Hausenfluck, Yisheng He, Jennifer Jeter, Lauren Johnson, Kelley Lee, Oz Malkesman, Geoffrey W. Meissner, Brian Melton, Scott Miller, Alyson Petruncio, Jacquelyn Price, Sophia Stavrou, Allison Vannan, Rebecca Vorimo, and Brianna Yarbrough, with Steering Committee of Yoshinori Aso, Gwyneth M. Card, Barry J. Dickson, Reed George, Wyatt Korff, Gerald M. Rubin, and James W. Truman.

We thank Gudrun Ihrke and Project Technical Resources for management coordination and staff support.

For setting up thousands of crosses, we thank the Janelia Fly Facility: Amanda Cavallaro, Tam Dang, Guillermo Gonzalez, Scarlett Harrison, Jui-Chun Kao, Todd R. Laverty, Brenda Perez, Brandi Sharp, Viruthika Vallanadu, and Grace Zheng.

We thank Melanie Radcliff for administrative support.

We thank Ruchi Parekh and Stephen M. Plaza for EM/LM coordination.

Stocks obtained from the Bloomington Drosophila Stock Center (NIH P40OD018537) were used in this study. We thank them, especially Annette Parks, Cale Whitworth, and Sam Zheng, for the maintenance and distribution of thousands of stocks from the Janelia collection.

Funding was provided by Howard Hughes Medical Institute.

References

Ashburner M, Bonner JJ. The induction of gene activity in drosophila by heat shock. Cell. 1979 Jun;17(2):241-54.

Aso Y, Hattori D, Yu Y, Johnston RM, Iyer NA, Ngo TT, Dionne H, Abbott LF, Axel R, Tanimoto H, Rubin GM. The neuronal architecture of the mushroom body provides a logic for associative learning. Elife. 2014 Dec 23;3:e04577. doi: 10.7554/eLife.04577.

Bogovic JA, Otsuna H, Heinrich L, et al. An unbiased template of the Drosophila brain and ventral nerve cord. bioRxiv 376384; doi: https://doi.org/10.1101/376384. Posted December 21, 2019.

Brand AH, Perrimon N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development. 1993 Jun;118(2):401-15.

Chiang AS, Lin CY, Chuang CC, et al. Three-dimensional reconstruction of brain-wide wiring networks in Drosophila at single-cell resolution. Curr Biol. 2011 Jan 11;21(1):1-11. doi: 10.1016/j.cub.2010.11.056. Epub 2010 Dec 2.

Clements J, Goina C, Kazimiers A, Otsuna H, Svirskas RR, Rokicki K. NeuronBridge Codebase. doi: 10.25378/janelia.12159378.v1. Software posted on 21.04.2020.

Costa M, Manton JD, Ostrovsky AD, Prohaska S, Jefferis GS. NBLAST: Rapid, Sensitive Comparison of Neuronal Structure and Construction of Neuron Family Databases.

Neuron. 2016 Jul 20;91(2):293-311. doi: 10.1016/j.neuron.2016.06.012. Epub 2016 Jun 30.
Davis FP, Nern A, Picard S, Reiser MB, Rubin GM, Eddy SR, Henry GL. A genetic, genomic, and computational resource for exploring neural circuit function. Elife. 2020 Jan 15;9. pii: e50901. doi: 10.7554/eLife.50901.
Dionne H, Hibbard KL, Cavallaro A, Kao JC, Rubin GM. Genetic Reagents for Making Split-GAL4 Lines in Drosophila. Genetics. 2018 May;209(1):31-35. doi: 10.1534/genetics.118.300682. Epub 2018 Mar 13.
Dolan MJ, Frechter S, Bates AS, et al. Neurogenetic dissection of the Drosophila lateral horn reveals major outputs, diverse behavioural functions, and interactions with the mushroom body. Elife. 2019 May 21;8. pii: e43079. doi: 10.7554/eLife.43079.
Fischbach KF, Dittrich AP. The optic lobe of Drosophila melanogaster. I. A Golgi analysis of wild-type structure. Cell and Tissue Research. 1989 Dec 1;258(3):441-75.
Fouladi S, Romero F, Iter D, et al. From laptop to lambda: outsourcing everyday jobs to thousands of transient functional containers. In Proceedings of the 2019 USENIX Conference on Usenix Annual Technical Conference (USENIX ATC ,Äô19). 2019. USENIX Association, USA, 475,Äì488. https://dl.acm.org/doi/10.5555/3358807.3358848
Gao S, Takemura SY, Ting CY, et al. The neural substrate of spectral preference in Drosophila. Neuron. 2008 Oct 23;60(2):328-42. doi: 10.1016/j.neuron.2008.08.010.
Germani F, Bergantinos C, Johnston LA. Mosaic Analysis in Drosophila. Genetics. 2018 Feb; 208(2):473-490. doi: 10.1534/genetics.117.300256.
Guo C, Pan Y, Gong Z. Recent Advances in the Genetic Dissection of Neural Circuits in Drosophila. Neurosci Bull. 2019 Dec;35(6):1058-1072. doi: 10.1007/s12264-019-00390-9. Epub 2019 May 22.
Jenett A, Rubin GM, Ngo TT, et al. A GAL4-driver line resource for Drosophila neurobiology. Cell Rep. 2012 Oct 25;2(4):991-1001. doi: 10.1016/j.celrep.2012.09.011. Epub 2012 Oct 11.
Kawase T, Sugano SS, Shimada T, Hara-Nishimura I. A direction-selective local-thresholding method, DSLT, in combination with a dye-based method for automated three-dimensional segmentation of cells and airspaces in developing leaves. Plant J. 2015;81(2):357-366. doi:10.1111/tpj.12738
Kohl J, Ng J, Cachero S, Ciabatti E, Dolan MJ, Sutcliffe B, Tozer A, Ruehle S, Krueger D, Frechter S, Branco T, Tripodi M, Jefferis GS. Ultrafast tissue staining with chemical tags. Proc Natl Acad Sci U S A. 2014 Sep 9;111(36):E3805-14. doi: 10.1073/pnas.1411087111. Epub 2014 Aug 25.
Lai SL, Lee T. Genetic mosaic with dual binary transcriptional systems in Drosophila. Nat Neurosci. 2006 May;9(5):703-9. Epub 2006 Apr 2.
Luan H, Peabody NC, Vinson CR, White BH. Refined spatial manipulation of neuronal function by combinatorial restriction of transgene expression. Neuron. 2006 Nov 9;52(3):425-36.
Maniates-Selvin JT, Hildebrand DGC, Graham BJ, et al. Reconstruction of motor control circuits in adult Drosophila using automated transmission electron microscopy. bioRxiv 2020.01.10.902478; doi: https://doi.org/10.1101/2020.01.10.902478. Posted January 11, 2020.
Meissner GW, Grimm JB, Johnston RM, Sutcliffe B, Ng J, Jefferis GSXE, Cachero S, Lavis LD, Malkesman O. Optimization of fluorophores for chemical tagging and immunohistochemistry of Drosophila neurons. PLoS One. 2018 Aug 15;13(8):e0200759. doi: 10.1371/journal.pone.0200759. eCollection 2018.
Morante J, Desplan C. The color-vision circuit in the medulla of Drosophila. Curr Biol. 2008 Apr 22;18(8):553-65. doi: 10.1016/j.cub.2008.02.075. Epub 2008 Apr 10.
Namiki S, Dickinson MH, Wong AM, Korff W, Card GM. The functional organization of descending sensory-motor pathways in Drosophila. Elife. 2018 Jun 26;7. pii: e34272. doi: 10.7554/eLife.34272.

Nern A, Pfeiffer BD, Rubin GM. Optimized tools for multicolor stochastic labeling reveal diverse stereotyped cell arrangements in the fly visual system. Proc Natl Acad Sci U S A. 2015 Jun 2;112(22):E2967-76. doi: 10.1073/pnas.1506763112. Epub 2015 May 11.
Otsuna H, Ito M, Kawase T. Color depth MIP mask search: a new tool to expedite Split-GAL4 creation. bioRxiv 318006; doi: https://doi.org/10.1101/318006. Posted May 09, 2018.
Otsuna H, Bogovic JA, Takemura SY, Kazunori S, Saalfeld S, Kawase T. Robust search method for Drosophila neurons between electron and light microscopy. In preparation.
Pfeiffer BD, Jenett A, Hammonds AS, et al. Tools for neuroanatomy and neurogenetics in Drosophila. Proc Natl Acad Sci U S A. 2008 Jul 15;105(28):9715-20. doi: 10.1073/pnas.0803697105. Epub 2008 Jul 9.
Pfeiffer BD, Ngo TT, Hibbard KL, Murphy C, Jenett A, Truman JW, Rubin GM. Refinement of tools for targeted gene expression in Drosophila. Genetics. 2010 Oct;186(2):735-55. doi: 10.1534/genetics.110.119917. Epub 2010 Aug 9.
Pfeiffer BD, Truman JW, Rubin GM. Using translational enhancers to increase transgene expression in Drosophila. Proc Natl Acad Sci U S A. 2012 Apr 24;109(17):6626-31. doi: 10.1073/pnas.1204520109. Epub 2012 Apr 9.
Potter CJ, Tasic B, Russler EV, Liang L, Luo L. The Q system: a repressible binary system for transgene expression, lineage tracing, and mosaic analysis. Cell. 2010 Apr 30;141(3):536-48. doi: 10.1016/j.cell.2010.02.025.
Rokicki K, Bruns CM, Goina C, et al. Janelia Workstation Codebase. doi: 10.25378/janelia.8182256.v1. Software posted on 24.05.2019.
Rohlfing T, Maurer CR Jr. Nonrigid image registration in shared-memory multiprocessor environments with application to brains, breasts, and bees. IEEE Trans Inf Technol Biomed. 2003 Mar;7(1):16-25.
Scheffer LK, Xu CS, Januszewski M, et al. A Connectome and Analysis of the Adult Drosophila Central Brain. bioRxiv 2020.04.07.030213; doi: https://doi.org/10.1101/2020.04.07.030213. Posted April 09, 2020.
Simpson JH. Mapping and manipulating neural circuits in the fly brain. Adv Genet. 2009;65:79-143. doi: 10.1016/S0065-2660(09)65003-3.
Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP. Sequencing depth and coverage: key considerations in genomic analyses. Nat Rev Genet. 2014 Feb;15(2):121-32. doi: 10.1038/nrg3642.
Takemura SY, Bharioke A, Lu Z, et al. A visual motion detection circuit suggested by Drosophila connectomics. Nature. 2013 Aug 8;500(7461):175-81. doi: 10.1038/nature12450.
Takemura SY, Xu CS, Lu Z, et al. Synaptic circuits and their variations within different columns in the visual system of Drosophila. Proc Natl Acad Sci U S A. 2015 Nov 3;112(44):13711-6. doi: 10.1073/pnas.1509820112. Epub 2015 Oct 19.
Tirian L, Dickson BJ. The VT GAL4, LexA, and split-GAL4 driver line collections for targeted expression in the Drosophila nervous system. bioRxiv 198648; doi: https://doi.org/10.1101/198648. Posted October 05, 2017.
Turner-Evans DB, Jensen KT, Ali S, et al. The neuroanatomical ultrastructure and function of a biological ring attractor. bioRxiv 847152; doi: https://doi.org/10.1101/847152. Posted January 19, 2020.
Tuthill JC, Nern A, Holtz SL, Rubin GM, Reiser MB. Contributions of the 12 neuron classes in the fly lamina to motion vision. Neuron. 2013 Jul 10;79(1):128-40. doi: 10.1016/j.neuron.2013.05.024.
Wolff T, Rubin GM. Neuroarchitecture of the Drosophila central complex: A catalog of nodulus and asymmetrical body neurons and a revision of the protocerebral bridge catalog. J Comp Neurol. 2018 Nov 1;526(16):2585-2611. doi: 10.1002/cne.24512. Epub 2018 Oct 18.

Wu M, Nern A, Williamson WR, Morimoto MM, Reiser MB, Card GM, Rubin GM. Visual projection neurons in the Drosophila lobula link feature detection to distinct behavioral programs. Elife. 2016 Dec 28;5. pii: e21022. doi: 10.7554/eLife.21022.
Yu Y, Peng H. Automated high speed stitching of large 3D microscopic images. 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Chicago, IL, 2011, pp. 238-241.
Yu HH, Awasaki T, Schroeder MD, et al. Clonal development and organization of the adult Drosophila central brain. Curr Biol. 2013 Apr 22;23(8):633-43. doi: 10.1016/j.cub.2013.02.057. Epub 2013 Mar 28.
Zheng Z, Lauritzen JS, Perlman E, et al. A Complete Electron Microscopy Volume of the Brain of Adult Drosophila melanogaster. Cell. 2018 Jul 26;174(3):730-743.e22. doi: 10.1016/j.cell.2018.06.019. Epub 2018 Jul 19.

Supplement

Supplemental Table 1. Generation 1 GAL4 lines included in the study.
The 4562 listed stocks were crossed to the MCFO reporter and imaged. Available stocks with a corresponding AD or DBD split-GAL4 half (Dionne et al. 2018; Tirian and Dickson 2017) were included.

Supplemental Table 2. Estimation of brain and VNC coverage.
MCFO neurons labeled per region were counted for a small number of lines, in order to estimate ‘coverage’, defined here as the total number of neurons labeled across lines compared to the total number of target neurons. Cell counts and calculations are presented.