A searchable image resource of *Drosophila* GAL4-driver expression patterns with single neuron resolution

Geoffrey W. Meissner[1*], Aljoscha Nern[1*], Zachary Dorman[1], Gina M. DePasquale[1], Kaitlyn Forster[1], Theresa Gibney[1], Joanna H. Hausenfluck[1], Yisheng He[1], Nirmala Iyer[1], Jennifer Jeter[1], Lauren Johnson[1], Rebecca M. Johnston[1,3], Kelley Lee[1], Brian Melton[1], Brianna Yarbrough[1], Christopher T. Zugates[1], Jody Clements[1], Cristian Goina[1], Hideo Otsuna[1], Konrad Rokicki[1], Robert R. Svirskas[1], Yoshinori Aso[1*], Gwyneth M. Card[1*], Barry J. Dickson[1*,4], Erica Ehrhardt[1,5], Jens Goldammer[1,5], Masayoshi Ito[1], Dagmar Kainmueller[2], Wyatt Korff[1*], Lisa Mais[2], Ryo Minegishi[1], Shigehiro Namiki[1,6], Gerald M. Rubin[1*], Gabriella R. Sterne[1,7], Tanya Wolff[1], Oz Malkesman[1], for FlyLight Project Team[1]

[1]Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, United States; [2]Max-Delbrueck-Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany

***For correspondence:**
meissnerg@janelia.hhmi.org (GWM); nerna@janelia.hhmi.org (AN); 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: [3]Friday Harbor Laboratories, University of Washington, Friday Harbor, WA, United States; [4]Queensland Brain Institute, The University of Queensland, Brisbane, Australia; [5]Institute of Zoology, University of Cologne, Cologne, Germany; [6]Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Japan; [7]University of California, Berkeley, Berkeley, CA, United States

Abstract Precise, repeatable genetic access to specific neurons via GAL4/UAS and related methods is a key advantage of *Drosophila* neuroscience. Neuronal targeting is typically documented using light microscopy of full GAL4 expression patterns, which generally 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 74,000 such adult central nervous systems. An anticipated use of this resource is to bridge the gap between neurons identified by electron or light microscopy. Identifying individual neurons that make up each GAL4 expression pattern improves the prediction of split-GAL4 combinations targeting particular neurons. To this end we have also developed the NeuronBridge search tool. NeuronBridge rapidly and effectively identifies neuron matches based on morphology across imaging modalities and datasets.

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 and Perrimon, 1993***; ***Lai and 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 cre-

ated, 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 and 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 with expression 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 typically expresses 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 the 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.*** (***2014a***); ***Wu et al.*** (***2016***); ***Namiki et al.*** (***2018***); ***Wolff and Rubin*** (***2018***); ***Dolan et al.*** (***2019***); ***Davis et al.*** (***2020***); ***Sterne et al.*** (***2021***)), 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 directly distinguished by visual inspection, 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.*** (***2018***)). 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 identify single cells of interest using 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. 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 formats 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 (Maximum Intensity Projection; together abbreviated CDM) search (***Otsuna et al., 2018***), which efficiently compares bitmap images using color to represent depth within the samples. NBLAST was recently expanded upon with the combination of PatchPerPix neuron segmentation (***Hirsch et al., 2020***) and PatchPerPixMatch search (PPPM; ***Mais et al.*** (***2021***)). PPPM identifies neuron segments with similar color and high NBLAST scores that best cover a target neuron of interest, allowing the use of partial segments from densely labeled MCFO samples. Advanced anatomical templates such as JRC2018 improve point-to-point mapping between samples and modalities (***Bogovic et al., 2020***). 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 5155 Gen1 GAL4 lines, most of which have been converted into split-GAL4 hemidrivers, performing three rounds of MCFO labeling to improve coverage of neurons. The resource includes images of 74,337 fly samples, with an average of 14 brain and 7 ventral nerve cord (VNC) images per line. We have released the data, along with the NeuronBridge tool to search between the FlyEM hemibrain, Gen1 MCFO data, and published split-GAL4 data.

Results

We used the MCFO approach on Generation 1 GAL4 lines (***Figure 1***A) to visualize individual neurons (***Figure 1***B) making up the GAL4 expression pattern. These neurons can be matched to EM neurons (***Figure 1***C-D) in order to predict split-GAL4 combinations for an EM neuron of interest (***Figure 1***E). We generated two collections of Gen1 MCFO images. The collection imaged with 20x and 63x microscope objectives targeted particular neurons of interest to collaborators annotating regions primarily in the brain and optic lobes. The collection imaged with 40x objectives broadly canvassed neurons in the central brain and VNC.

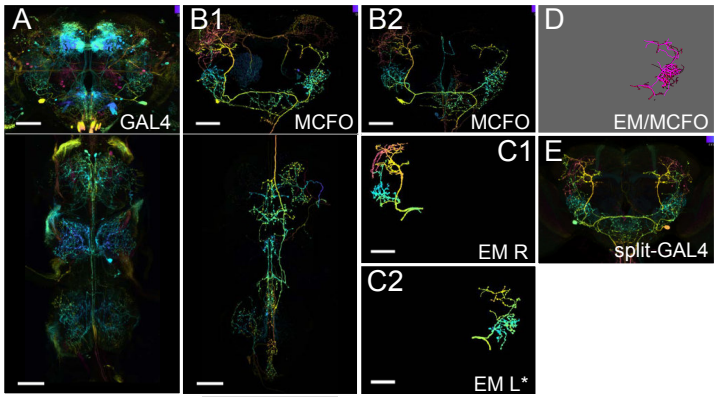
A challenge with any stochastic neuron labeling approach is to optimize the number of identifiable neurons in each sample: too sparse and samples are empty or have few labeled neurons; too dense and the neurons overlap, making it difficult to fully isolate individual neurons even if they are labeled in different colors. MCFO allows for control of labeling density by optimizing the amount of Flp activity, either by selecting different Flp drivers, or altering heat shock duration for hs-Flp (***Nern et al., 2015***). GAL4 lines with broader expression typically require lower Flp activity to yield isolated neurons. In the 20x/63x MCFO collection, labeling density was customized for regions of interest, iterating on prior results (***Nern et al., 2015***). In the 40x MCFO collection, labeling density was initially standardized (Phase 1), then optimized based on overall GAL4 expression density (Phase 2; ***Figure 1–Figure Supplement 1***A). For many lines there is no globally ideal level of Flp activity, as they have varying levels of expression density in different central nervous system (CNS) regions.

The 20x/63x and 40x datasets differed in several other respects (***Figure 1***F). The 20x/63x collection was imaged with 20x objectives, followed by 63x imaging of specific regions of interest, whereas the 40x collection was uniformly imaged at 40x. The 20x/63x collection was focused on a smaller set of lines visualized primarily in female brains (94.2%), whereas the 40x collection covered more lines (4575 vs. 2463), a mixture of male and female samples (44.9% female), and both brains and VNCs (7.1 VNCs per line vs. 0.9 in the 20x/63x dataset).

Finally, as the 20x/63x dataset and existing publications (e.g. ***Fischbach and Dittrich*** (***1989***); ***Morante and Desplan*** (***2008***); ***Takemura et al.*** (***2013***); ***Nern et al.*** (***2015***); ***Takemura et al.*** (***2015***)) effectively documented the largely repetitive structure of the optic lobes, the 40x dataset excluded them. Collections of split-GAL4 driver lines for many optic lobe cell types are already available (***Tuthill et al., 2013***; ***Wu et al., 2016***; ***Davis et al., 2020***). Many neurons that connect the optic lobe with the central brain can still be identified in the 40x 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.

40x Gen1 MCFO collection

After performing extensive MCFO labeling together with collaborators focused on annotating particular CNS regions, we performed comprehensive MCFO mapping of Gen1 GAL4 lines across most



F

	20x/63x collection	40x collection	Total
Gen1 GAL4 Lines	2463	4575	5155
Samples	27546	46791	74337
Average Samples/Line	11.2	10.2	14.4
Std. Dev. Samples/Line	7.6	4.6	8.7
Average Brain/Line	11.2	10.1	14.3
Average VNC/Line	0.9	7.1	6.8
Female %	94.2%	44.9%	63.2%
20x/40x image tiles	29784	111380	141164
63x image tiles	22775	-	22775
Lines with 63x images	1748	-	1748
Samples with 63x images	8447	-	8447

Figure 1. Generation 1 MCFO and EM/LM comparison overview.
(A) Overall GAL4 expression pattern of a driver line containing a cell type of interest, shown as a color depth MIP (***Otsuna et al., 2018***). Original images are from published datasets (***Jenett et al., 2012***).
(B1) Example MCFO labeled cells from the driver line in (A). MCFO labeling reveals a prominent descending neuron. (B2) An additional MCFO labeled cell of the same type but from a different line. The color depth MIPs in B1 & B2 represent data from one of the three MCFO markers, so color changes indicate changes in the z-dimension rather than differential MCFO labeling.
(C1, C2). Matching EM reconstructions for the cell type. Both panels show reconstructions from the right side Hemibrain; the lower panel is mirrored to facilitate comparison to the LM data.
(D) PPPM overlay of MCFO from (B1) and EM reconstruction from (C2).
(E) Split-GAL4 made from split hemidrivers derived from GAL4 lines in A and B.
Driver lines used are R56H09 (A, B1), R23E11 (B2), and SS01588 (E). Hemibrain body IDs are 571346836 (C1) and 1786496543 (C2). All scale bars, 50 μm.
(F) Statistics for each part of the Gen1 MCFO collection are tabulated.

Figure 1–Figure supplement 1. Generation 1 MCFO expression density categories.

of the CNS. 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 CNS 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, 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.

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 1–Figure Supplement 1***A). 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 1–Figure Supplement 1***B).

In order to increase the number of identifiable neurons, a subset of lines was re-examined with altered parameters. Phase 2 of the 40x pipeline generated images of an additional 18,894 central brains and 6,235 VNCs from 19,062 flies (***Figure 1***). 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. Phase 2 focused on Category 2 and 3 lines as most likely to be useful for split-GAL4 creation. Category 1 and 5 lines were outside our effective labeling range 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 is low (although future improvements to neuron segmentation approaches are expected to improve yields).

Heat-shock Flp (hs-Flp) was used in Phase 2 rather than 57C10-Flp (***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. To increase the chance of obtaining sex-specific neurons and neuronal morphology, we randomly choose one sex for each half of the lines in Phase 1 and then in Phase 2 switched them to the opposite sex.

As the number of MCFO samples for a given GAL4 line increases, the probability of labeling additional unique neurons diminishes until every neuron labeled by that GAL4 line is represented within the MCFO dataset. Sparser lines approach saturation more rapidly, especially because we can use higher Flp activity to label a greater fraction of available GAL4 neurons per sample without overwhelming 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 *Drosophila* 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. Thus, we focused Phase 2 more heavily on the brain than the VNC, which together with the above density adjustment led to imaging on average 6.0 brains in Category 2 or 9.1 brains in Category 3, and 2.5 VNCs per line across both categories.

MCFO labeling observations

The large number of lines processed under mostly uniform MCFO conditions provided an 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, including neurons, glia, and trachea, as reviewed in ***Ashburner and Bonner*** (***1979***). 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 2***). 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 of the mushroom body were labeled at different rates with each reporter. We scored for the presence of Kenyon cell labeling 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. Kenyon cells were much more commonly labeled using hs-Flp MCFO (430 lines, or 93%) than with R57C10-Flp MCFO (44 lines, or 10%) or UAS-GFP (111 lines, or 24%; ***Figure 2***). Most frequently lines had unlabeled Kenyon cells with GFP and R57C10-Flp MCFO and labeled Kenyon cells with hs-Flp (253 lines, or 55%; ***Figure 2***E). Lines were also observed with labeled Kenyon cells using GFP and hs-Flp MCFO, but not R57C10-Flp (59 lines, or 13%; ***Figure 2***F). 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. 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

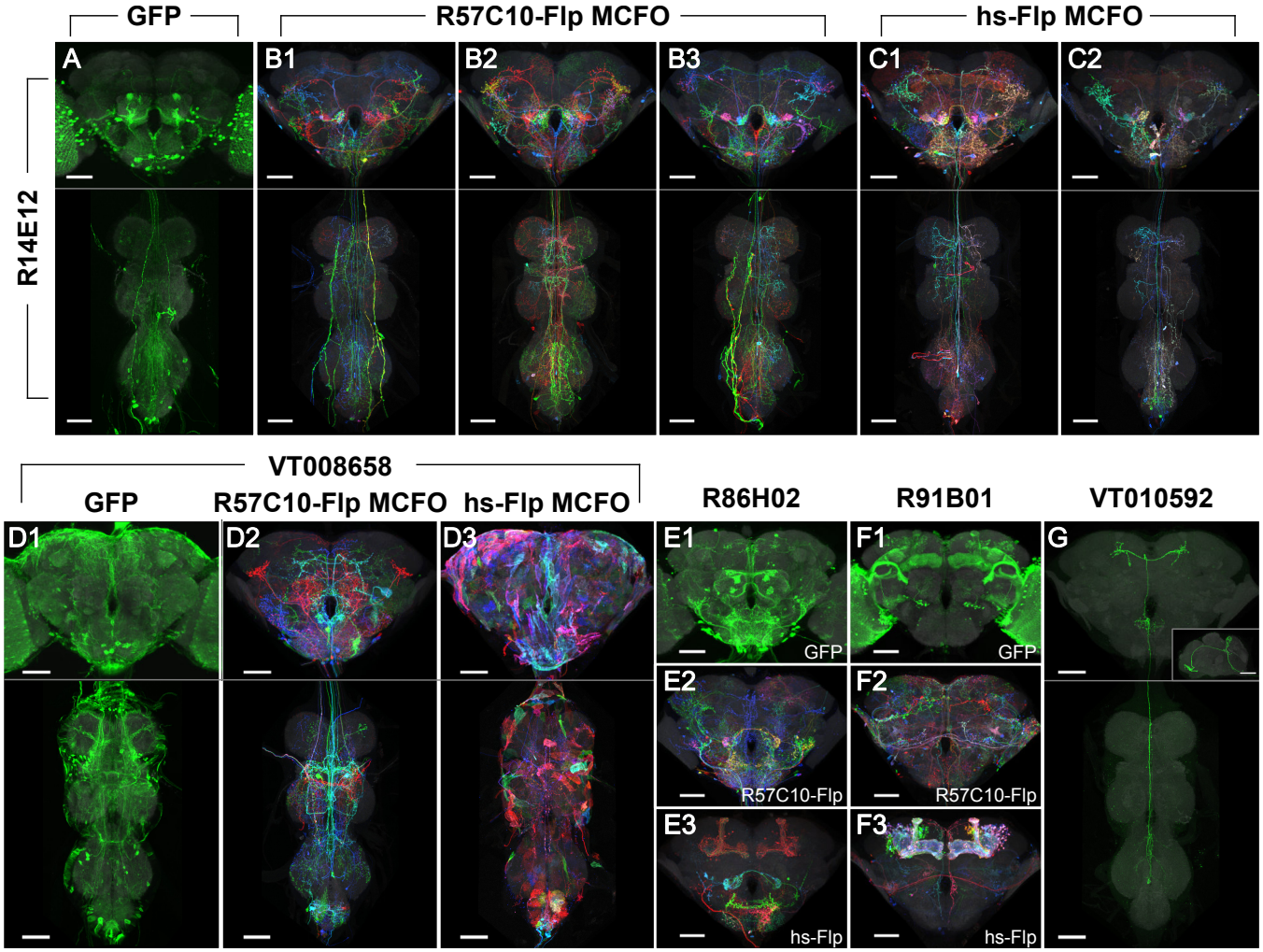


Figure 2. Phase 1 & 2 overview and labeling examples.
(A-C) R14E12-GAL4 in attP2 crossed to (A) pJFRC2-10XUAS-IVS-mCD8::GFP, (B) R57C10-Flp MCFO, or (C) hs-Flp MCFO. Adult CNS MIPs are shown, with neuropil reference in gray and neuronal signal in green (A) or full MCFO colors (B-C). Multiple examples are shown for B-C. Scale bars, 50 μm.
(D) Glia are seen with VT008658-GAL4 in attP2 crossed to (D1) pJFRC2-10XUAS-IVS-mCD8::GFP and (D3) hs-Flp MCFO, but not (D2) 57C10-Flp MCFO.
(E) Kenyon cell labeling is not seen with R86H02-GAL4 in attP2 crossed to (E1) pJFRC2-10XUAS-IVS-mCD8::GFP or (E2) R57C10-Flp MCFO, but is seen when crossed to (E3) hs-Flp MCFO.
(F) Kenyon cell labeling is seen with R91B01-GAL4 in attP2 crossed to (F1) pJFRC2-10XUAS-IVS-mCD8::GFP and (F3) hs-Flp MCFO, but is not seen when crossed to (F2) R57C10-Flp MCFO.
(G) 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.

(*Figure 2*G). It was observed in at least one sample from over 60% of lines crossed to either MCFO 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 in full GAL4 patterns (when crossed to UAS-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.

Neuron searching across image collections

This image collection makes it possible to identify GAL4 driver lines with expression 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 identifying a neuron in both EM and LM datasets.

Although LM images do not match the synaptic resolution of EM data, they can provide additional, complementary anatomical information. First, identification of LM matches provides an independent quality check for EM reconstructions (e.g. ***Scheffer et al.*** (***2020***); ***Phelps et al.*** (***2021***)). Second, the LM data often includes multiple examples of a cell type and thus provide insights into variable 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 current EM volumes. For example, the Hemibrain dataset does not fully include neurons that span both brain hemispheres or project to or from the VNC (see ***Figure 1***). It is thus important to be able to perform EM/LM matching.

While accurate matching of EM reconstructions with single cell LM images can sometimes 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. Alignment here was accomplished by registering all samples to JRC2018 brain and VNC templates (***Bogovic et al., 2020***). We have also made the neuron search tool NeuronBridge (***Rokicki et al., 2022***) publicly available.

NeuronBridge currently allows the user to perform anatomical similarity searches between published datasets reported by Janelia's FlyLight and FlyEM Team Projects. Searching is based on two approaches: (1) Color Depth MIP (CDM), which allows direct comparisons of expression similarity in registered images without the need for a complete skeletonization (***Otsuna et al., 2018***); and (2) PatchPerPixMatch (PPPM), which enhances NBLAST to find groups of neuron segments that best match a target neuron (***Costa et al., 2016***; ***Mais et al., 2021***).

The basic strategy of CDM searching is to represent neuronal expression with a two-dimensional maximum intensity projection (MIP), using color to indicate the third depth dimension. Two aligned brain images can then be compared by looking for pixels of similar color at similar x-y coordinates of their color depth MIPs. The color depth MIP search approach was extended in several ways to improve matches for denser MCFO data (***Otsuna et al., 2022***). 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.

PPPM searching is based on the evaluation of fully (but often imperfectly) segmented neurons (***Hirsch et al., 2020***; ***Mais et al., 2021***). The underlying NBLAST algorithm compares the similarity in 3D location and neuronal arbor orientation at many points along two neuron segments. PPPM looks for an optimal combination of neuron segments that together maximize an NBLAST-derived similarity score for the target neuron. It includes optimizations for identifying non-overlapping segments that tile a target, along with positive weighting for segments of similar color, as would be expected from a MCFO neuron broken into multiple segments.

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.

In addition to pre-computed search results for published data sets, we have also made custom search capability available in NeuronBridge. An unaligned 3D image stack can be uploaded, and the service will register it to the JRC2018 standard reference template (***Bogovic et al., 2020***). CDMs are automatically generated from the aligned image, and an interactive selection tool allows the user to choose a channel and mask a target neuron for search. Targets can be searched against either the EM or LM image database, in a highly parallel (~3000 threads) cloud-based implementation that completes within a few minutes. Custom search results are browsed in the same way as pre-computed results.

Search approach evaluation

We evaluated search performance of CDM & PPPM approaches between the EM Hemibrain (***Scheffer et al., 2020***) and the Gen1 MCFO dataset in the context of making split-GAL4 lines specifically targeting EM bodies of interest (***Figure 3***).

Search performance can be evaluated in several ways depending on the application (***Costa et al., 2016***; ***Otsuna et al., 2018***; ***Mais et al., 2021***). We refer here to "forward" and "reverse" analysis in the context of split-GAL4 creation. Forward analysis consisted of direct qualitative evaluation of EM to LM search results, determining whether top LM results appeared to contain the searched for EM body. Forward analysis is best performed with detailed knowledge of the examined neurons to avoid false positives. Reverse analysis made use of previously documented associations between split-GAL4 lines and EM bodies. If a split-GAL4 line labels a neuron, its constituent split hemidrivers should as well, as should some MCFO of Gen1 GAL4 lines with the same enhancers. We thus evaluated whether known EM/LM matches were highly ranked within the search results. Due to the stochastic nature of MCFO, not every sample of a valid matching GAL4 line will contain the target neuron.

Evaluation of the search approaches also addressed neuron coverage of the Gen1 MCFO dataset. For both search directions the total number of correct matching samples and GAL4 lines gave a measure of how completely the Gen1 MCFO dataset labels each queried neuron.

We performed forward analysis on the top 100 CDM and PPPM Phase 1 Gen1 MCFO search results for ten Hemibrain bodies (***Figure 4***). Both CDM & PPPM correctly identified many matches in the dataset for each EM body in the top search results. CDM identified 17.6 ± 8.3 (average ± standard deviation) correct lines per Hemibrain body, whereas PPPM identified 20.1 ± 10.6.

For cell type LC18, PPPM outperformed CDM, with 24 and 13 correct matches in the top 100, repspectively (***Figure 4***B). For cell type CT1, on the other hand, CDM correctly found 8 results in the top 100, compared to 3 for PPPM (***Figure 4***C). More generally, CDM and PPPM each identified many lines in the top 100 results that were not identified by the other search approach (***Figure 4–Figure Supplement 1***). CDM uniquely identified 8.2 ± 6.1 and PPPM uniquely identified 10.7 ± 8.7 lines, respectively.

Thus, at least for this limited set of neurons, the Gen1 MCFO collection isolates enough examples of each neuron to likely create a split-GAL4 combination. CDM & PPPM successfully identify these correct matches, although they are interspersed with a larger number of false matches. Both approaches varied widely by neuron, without obvious correlation to neuron morphology (***Figure 4–Figure Supplement 1***). Although all ten neurons examined here yielded at least nine matching lines, we do not expect this to hold for every neuron. It remains likely that that expanding the MCFO collection with more samples or more drivers would improve the chances of obtaining a good set of matches.

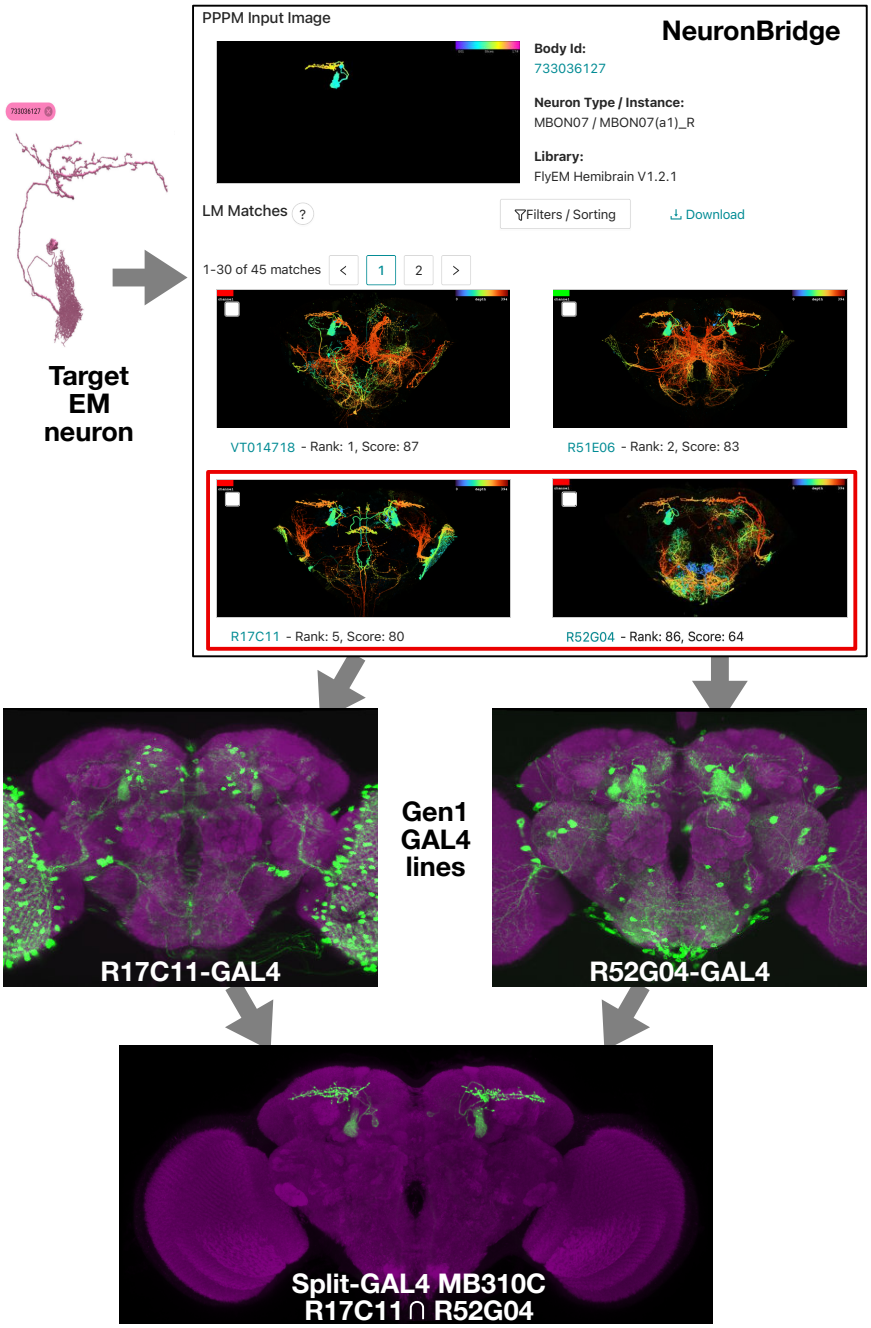


Figure 3. EM/LM search for split-GAL4 creation.
Neuron search techniques allow for the identification of Gen1 MCFO images containing an EM body of interest. The corresponding Gen1 GAL4 lines should label the same neuron with other UAS reporters, as should split-GAL4 hemidrivers constructed with the same enhancer fragment. The two hemidrivers can then be combined into a split-GAL4 with the aim of generating a driver that specifically targets that neuron. An example is shown of the anticipated search process, from a neuron identified via electron microscopy to the creation of a split-GAL4 driver. As in ***Figure 1*** NeuronBridge displays color depth MIPs of single MCFO markers rather than the full MCFO image, so color changes indicate depth rather than different neurons. NeuronBridge result order was reformatted for display purposes. The example shown includes FlyEM Hemibrain body ID 733036127 (***Scheffer et al., 2020***), Generation 1 GAL4 lines R17C11-GAL4, R52G04-GAL4, and split-GAL4 MB310C (MBON07) (***Jenett et al., 2012***; ***Aso et al., 2014b***).

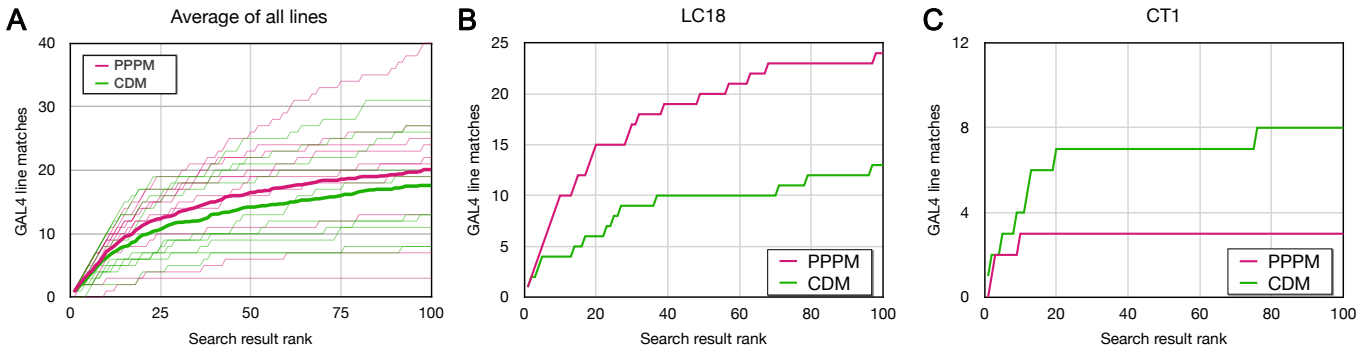


Figure 4. Forward analysis: direct evaluation of CDM & PPPM search results.
EM bodies were searched for in Phase 1 40x Gen1 MCFO light images using CDM & PPPM approaches. Search results were qualitatively evaluated by an anatomical expert for the presence of the sought neuron. Most results were scored based on color depth MIP images. Full image stacks were used to score about 20% of samples, including the majority of samples scored as containing the sought neuron. The cumulative number of correct matches found is plotted against the depth of searching for CDM (green) & PPPM (magenta).
(A) Average results for each search approach are plotted in bold on top of individual results.
(B) Cell type LC18 (Hemibrain body 1722342048) search result evaluation.
(C) Cell type CT1 (Hemibrain body 1311993208) search result evaluation.

Figure 4–Figure supplement 1. Table of all forward analysis results by cell type.

Figure 4–Figure supplement 2. Forward analysis summary and individual plots for CDM & PPPM.

We extended the PPPM reverse analysis in ***Mais et al.*** (***2021***) with a comparison to CDM (***Figure 5***). We examined nine Hemibrain bodies, each with 2 to 13 published split-GAL4 associations (***Schretter et al., 2020***; ***Wang et al., 2020b***,a). The best rank of each known-matching line was recorded, with ***Figure 5*** showing the median line rank and the percentage of lines with ranks in the top 50 results. PPPM and CDM both had median line ranks under 100 for most EM bodies. PPPM was somewhat more consistent, with 33% to 80% of known matches in the top 50 results, compared to 0% to 100% for CDM. As with the forward analysis, each approach performed better on some neurons than the other approach.

Discussion

We have described an extensive MCFO image resource from Generation 1 GAL4 lines, providing single-cell-level resolution of the neurons labeled by each line. The NeuronBridge website allows rapid searching of this resource from published EM datasets or uploaded images. CDM and PPPM search approaches both find valid EM/LM matches, demonstrating their effectiveness and the good coverage of the brain by the Gen1 MCFO collection. NeuronBridge has already seen frequent usage (***Bidaye et al., 2020***; ***Morimoto et al., 2020***; ***Nojima et al., 2021***; ***Sareen et al., 2021***; ***Zolin et al., 2021***; ***Israel et al., 2022***; ***Tanaka and Clark, 2022***). Together these tools allow for the rapid determination of likely split-GAL4 lines and other enhancer-based approaches to target most neurons initially in the FlyEM Hemibrain and eventually in the full *Drosophila* CNS.

While performing these analyses and practically applying the tools to screen split-GAL4 combinations, we made some qualitative observations: (1) In general, both CDM and PPPM are complimentary and best used in combination, although PPPM tended to bring good matches closer to the top of search results. (2) CDM occasionally struggled with occluded neurons and benefited from examination of full 3D stacks of matching MCFO samples. (3) PPPM correspondingly showed the most improvement in samples with occluded neurons. (4) Both techniques return some highly-ranked false positives with clear flaws, such that rankings alone are insufficient for algorithmic association of EM and LM neurons. (5) We estimate the image collection and search techniques can lead to good split combinations for 50-80% of cell types, depending on how clean a combination is needed. More split hemidrivers would likely be needed to increase this rate. The search techniques don't significantly change which cell types can be targeted, but greatly simplify identifying candidate split

name	hemibrain body id	Known-matching lines	PPPM median line rank	CDM median line rank	PPPM % in top 50	CDM % in top 50
pC1e	**514850616**	13	14	58	69%	46%
pC1d	**5813063587**	12	28	41	58%	50%
aIPg	**645456880**	5	6	3	80%	100%
oviDN	**550655668**	4	70	42	50%	75%
oviDN	**519949044**	4	95	41	50%	50%
SAG	**517587356**	2	49	78	50%	0%
SAG	**5812981862**	2	44	118	50%	0%
vpoDN	**5813057864**	4	NA*	95	50%	50%
DNp13	**887195902**	3	84	53	33%	33%

* Line ranks: 7, 49, >400, >400

Figure 5. Reverse analysis: scoring known match search ranks in CDM & PPPM results.
PPPM and CDM search results on 9 Hemibrain bodies were scored for the presence of known GAL4 matches from the literature (***Schretter et al., 2020***; ***Wang et al., 2020b***,a). Only the top-ranking sample for each line and EM body comparison was considered. Searches were performed across only Phase 1 40x Gen1 MCFO collection data. Results for bodies 514850616 and 5813063587 are reformatted from ***Mais et al.*** (***2021***) Figure 9.

combinations without requiring as much anatomical expertise.

There are several caveats for why close EM/LM matches don't always lead to successful split-GAL4 combinations: (1) Many CNS cell types contain multiple neurons that are indistinguishable based on morphology. Thus, two matches for a cell type may label different neurons within the cell type and fail as a split combination. Information from connectomic approaches and other modalities are also continuing to refine cell type definitions. (2) Although split-GAL4 hemidrivers are made with the same enhancer fragments as Gen1 GAL4 lines, they can differ in vector sequence and genomic insertion site. These differences can alter expression patterns and hence split-GAL4 effectiveness. (3) UAS reporters can vary in genomic insertion site, number of UAS elements, and other factors that affect how well they label particular cell types. MCFO reporters in particular can tend to brightly label neurons that are weakly labeled by reporters for the full GAL4 pattern. An examination of the full Gen1 GAL4 patterns (if not too dense) can help predict likely effectiveness of a split combination. (4) GAL4 driver expression can vary temporally, so there could be spatial but not temporal overlap between two split hemidrivers.

In creating the image resource, 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. For the search resource, we have implemented two complementary search approaches that effectively identify neuron matches in an easy to use interface. The image resource should be amenable to analysis with future search approaches as they continue to develop.

While our focus has been on the EM to split-GAL4 use case, we described other uses, including guiding EM proofreading and extending EM analyses beyond limited regions or sample sizes currently available. We anticipate other uses will be found for this resource.

Materials and Methods

Fly stocks

The 5155 Generation 1 GAL4 stocks included in this resource (Table S1) were from ***Jenett et al.*** (***2012***); ***Tirian and Dickson*** (***2017***). Lines in the 20x/63x ("Annotator") collection were selected by collaborators for individual projects. For the 40x collection we focused on driver lines with available AD or DBD hemidrivers (***Tirian and Dickson, 2017***; ***Dionne et al., 2018***). Split-GAL4 stock MB310C consists of *R52G04-p65ADZp in VK00027* and *R17C11-ZpGdbd in attP2* (***Aso et al., 2014b***).

MCFO UAS reporters are described in ***Table 1***. *pJFRC2-10XUAS-IVS-mCD8::GFP* was from (***Pfeiffer et al., 2010***).

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.

For the 40x pipeline 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 was not as bright as Brp antibody staining with nc82, but was more consistent and had lower background. 40x pipeline samples were washed 1 to 4 times for 15 minutes and then tagged with 2 µM Cy2 SNAP-tag ligand to visualize the Brp-SNAP neuropil the same day, after which immunohistochemistry and DPX mounting followed.

20x/63x samples used nc82 for neuropil reference labeling, as in ***Nern et al.*** (***2015***), and typically received 4 washes of 10 minutes each after fixation. See https://www.janelia.org/project-team/flylight/protocols for full MCFO protocols with either nc82 or hybrid Brp-SNAP neuropil labeling.

Imaging and image processing

Imaging was performed using eight Zeiss LSM 710 or 780 laser scanning confocal microscopes over a combined capture time of 11 years. 20x/63x imaging was performed with 20x air and 63x oil objectives to combine rapid scanning of all samples with detailed scanning of regions of interest. 40x imaging was performed with 40x oil objectives to cover the central brain and VNC with good axial resolution in a single pass. Confocal stacks were captured at 0.52x0.52x1.00 micron (20x objective), 0.19x0.19x0.38 micron (63x), or 0.44 µm isotropic resolution (40x). 40x resolution was selected to maximize effective z-resolution while limiting the size of the full data set (about 100 TB combined). The field of view was set to the widest 0.7 zoom for 40x & 63x objectives, resulting in heightened lens distortion at the edges of images, which was corrected before stitching (***Bogovic et al., 2020***). The whole brain and VNC (where present) were captured in separate 20x tiles for 20x/63x samples, followed by selected 63x tiles of regions of interest. The central brain and two VNC tiles (where present) were captured for each 40x sample. After merging and distortion correction, overlapping 40x/63x tiles were automatically stitched together, as described (***Yu and 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 and Maurer, 2003***; ***Otsuna et al., 2018***; ***Bogovic et al., 2020***).

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 tiles covering a single brain or VNC.

The image processing pipeline (distortion correction, normalization, merging, stitching, alignment, MIP generation, file compression) was automated using the open-source Janelia Workstation software (***Rokicki et al., 2019***), which was also used to review the secondary results and annotate lines for publishing. Images for published lines were 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 http://data.janelia.org/h5j).

The open-source NeuronBridge tool (***Rokicki et al., 2022***) is a web application designed for ease of use and accessibility to neuron mappings across large multi-modal data sets. It hosts precomputed matches for publicly available EM and LM data sets originating at Janelia, and also supports ad-hoc searches against those data sets based on user data. NeuronBridge was constructed as a single-page application built on the React framework for fast performance, responsiveness, and ease of deployment. The web app and backend services are both deployed to AWS to ensure scalability and reliability, and they use only serverless components to minimize costs. 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 collection. Samples lacking visible neuron expression or too dense for IHC were in most cases excluded prior to imaging. 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. Every effort was made to accurately track and correct line and sample metadata, but the dataset may still contain occasional errors.

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 are expected to be tractable for segmentation, whereas Category 4 lines are more 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.

Search approach evaluation

For the forward analysis the top 100 NeuronBridge search results were examined for one hemibrain body in each cell type. About 20% of the samples were checked by opening the image stacks, including the majority of the samples annotated as including the cell type in question.

Reverse analysis was performed as in ***Mais et al.*** (***2021***).

Data availability

Gen1 MCFO anatomical images are available at https://gen1mcfo.janelia.org.
NeuronBridge search is available at https://neuronbridge.janelia.org.

Acknowledgments

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 contributing laboratories.

During this effort, the FlyLight Project Team included Megan Atkins, Shelby Bowers, Kari Close, Gina DePasquale, Zack Dorman, Kaitlyn Forster, Jaye Anne Gallagher, Theresa Gibney, Asish Gulati, Joanna H. Hausenfluck, Yisheng He, Kristin Hendersen, Hsing Hsi Li, Nirmala Iyer, Jennifer Jeter, Lauren Johnson, Rebecca Johnston, Rachel Lazarus, Kelley Lee, Hua-Peng Liaw, Oz Malkesman, Geoffrey Meissner, Brian Melton, Scott Miller, Reeham Motaher, Alexandra Novak, Omatara Ogundeyi, Alyson Petruncio, Jacquelyn Price, Sophia Protopapas, Susana Tae, Athreya Tata, Jennifer Taylor, Allison Vannan, Rebecca Vorimo, Brianna Yarborough, Kevin Xiankun Zeng, and Chris Zugates, with Steering Committee of Yoshinori Aso, Gwyneth Card, Barry Dickson, Reed George, Wyatt Korff, Gerald Rubin, and James Truman.

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

We thank Melanie Radcliff for administrative support.

We thank Barret Pfeiffer for his early work in developing the MCFO method. We thank Teri Ngo for her early collaborations with FlyLight. We thank Kei Ito, Kristin Scott, and Michael H. Dickinson for contributions to visitor and team projects.

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 Karen Hibbard for establishing the brp-SNAP MCFO reporter stocks.

We thank Mark Bolstad, Tom Dolafi, Leslie L. Foster, Sean Murphy, Donald J. Olbris, Todd Safford, Eric Trautman, and Yang Yu for their work on software infrastructure. 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 stocks from the Janelia collection.

Funding was provided by Howard Hughes Medical Institute.

This article is subject to HHMI's Open Access to Publications policy. HHMI lab heads and project team leads 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.

Additional information

Author ORCIDs

Geoffrey W. Meissner: https://orcid.org/0000-0003-0369-9788
Aljoscha Nern: https://orcid.org/0000-0002-3822-489X
Zachary Dorman: https://orcid.org/0000-0001-9933-7217
Theresa Gibney: https://orcid.org/0000-0001-5461-724X
Christopher T. Zugates: https://orcid.org/0000-0003-1882-3665
Jody Clements: https://orcid.org/0000-0003-3932-8188
Cristian Goina: https://orcid.org/0000-0003-2835-7602
Hideo Otsuna: https://orcid.org/0000-0002-2107-8881
Konrad Rokicki: https://orcid.org/0000-0002-2799-9833
Robert R. Svirskas: https://orcid.org/0000-0001-8374-6008
Barry J. Dickson: https://orcid.org/0000-0003-0715-892X
Erica Ehrhardt: https://orcid.org/0000-0002-9252-1414

Jens Goldammer: https://orcid.org/0000-0002-5623-8339
Dagmar Kainmueller: https://orcid.org/0000-0002-9830-2415
Wyatt Korff: https://orcid.org/0000-0001-8396-1533
Ryo Minegishi: https://orcid.org/0000-0002-2895-9438
Gerald M. Rubin: https://orcid.org/0000-0001-8762-8703
Gabriella R. Sterne: https://orcid.org/0000-0002-7221-648X
Tanya Wolff: https://orcid.org/0000-0002-8681-1749
Oz Malkesman: https://orcid.org/0000-0003-2219-7476

References

Ashburner M, Bonner JJ. The induction of gene activity in drosophilia by heat shock. Cell. 1979 Jun; 17(2):241–54. doi: 10.1016/0092-8674(79)90150-8.

Aso Y, Hattori D, Yu Y, Johnston RM, Iyer NA, Ngo TTB, 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; 3:e04577. doi: 10.7554/eLife.04577.

Aso Y, Sitaraman D, Ichinose T, Kaun KR, Vogt K, Belliart-Guérin G, Plaçais PY, Robie AA, Yamagata N, Schnaitmann C, Rowell WJ, Johnston RM, Ngo TTB, Chen N, Korff W, Nitabach MN, Heberlein U, Preat T, Branson KM, Tanimoto H, et al. Mushroom body output neurons encode valence and guide memory-based action selection in Drosophila. Elife. 2014 Dec; 3:e04580. doi: 10.7554/eLife.04580.

Bidaye SS, Laturney M, Chang AK, Liu Y, Bockemühl T, Büschges A, Scott K. Two Brain Pathways Initiate Distinct Forward Walking Programs in Drosophila. Neuron. 2020 11; 108(3):469–485.e8. doi: 10.1016/j.neuron.2020.07.032.

Bogovic JA, Otsuna H, Heinrich L, Ito M, Jeter J, Meissner G, Nern A, Colonell J, Malkesman O, Ito K, Saalfeld S. An unbiased template of the Drosophila brain and ventral nerve cord. PLoS One. 2020; 15(12):e0236495. doi: 10.1371/journal.pone.0236495.

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, Chang HM, Hsieh CH, Yeh CW, Shih CT, Wu JJ, Wang GT, Chen YC, et al. Three-Dimensional Reconstruction of Brain-wide Wiring Networks in Drosophila at Single-Cell Resolution. Curr Biol. 2011; 21(1):1–11.

Costa M, Manton JD, Ostrovsky AD, Prohaska S, Jefferis GSXE. NBLAST: Rapid, Sensitive Comparison of Neuronal Structure and Construction of Neuron Family Databases. Neuron. 2016 07; 91(2):293–311. doi: 10.1016/j.neuron.2016.06.012.

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 01; 9. 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.

Dolan MJ, Frechter S, Bates AS, Dan C, Huoviala P, Roberts RJ, Schlegel P, Dhawan S, Tabano R, Dionne H, Christoforou C, Close K, Sutcliffe B, Giuliani B, Li F, Costa M, Ihrke G, Meissner GW, Bock DD, Aso Y, et al. Neurogenetic dissection of the Drosophila lateral horn reveals major outputs, diverse behavioural functions, and interactions with the mushroom body. Elife. 2019 05; 8. doi: 10.7554/eLife.43079.

Fischbach KF, Dittrich A. The optic lobe of Drosophila melanogaster. I. A Golgi analysis of wild-type structure. Cell and tissue research. 1989; 258(3):441–475.

Fouladi S, Romero F, Iter D, Li Q, Chatterjee S, Kozyrakis C, Zaharia M, Winstein K. From laptop to lambda: Outsourcing everyday jobs to thousands of transient functional containers. In: *2019 {USENIX} Annual Technical Conference ({USENIX}{ATC} 19)*; 2019. p. 475–488.

Gao S, Takemura SY, Ting CY, Huang S, Lu Z, Luan H, Rister J, Thum AS, Yang M, Hong ST, Wang JW, Odenwald WF, White BH, Meinertzhagen IA, Lee CH. The neural substrate of spectral preference in Drosophila. Neuron. 2008 Oct; 60(2):328–42. doi: 10.1016/j.neuron.2008.08.010.

Germani F, Bergantinos C, Johnston LA. Mosaic Analysis in Drosophila. Genetics. 2018 02; 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.

Hirsch P, Mais L, Kainmueller D. PatchPerPix for Instance Segmentation. arXiv. 2020 01; https://arxiv.org/pdf/2001.07626.pdf.

Israel S, Rozenfeld E, Weber D, Huetteroth W, Parnas M. Olfactory stimuli and moonwalker SEZ neurons can drive backward locomotion in Drosophila. Curr Biol. 2022 03; 32(5):1131–1149.e7. doi: 10.1016/j.cub.2022.01.035.

Jenett A, Rubin GM, Ngo TT, Shepherd D, Murphy C, Dionne H, Pfeiffer BD, Cavallaro A, Hall D, Jeter J, et al. A GAL4-Driver Line Resource for Drosophila Neurobiology. Cell reports. 2012; 2(4):991–1001.

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 Jan; 81(2):357–66. 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 GSXE. Ultrafast tissue staining with chemical tags. Proc Natl Acad Sci U S A. 2014 Sep; 111(36):E3805–14. doi: 10.1073/pnas.1411087111.

Lai SL, Lee T. Genetic mosaic with dual binary transcriptional systems in Drosophila. Nat Neurosci. 2006 May; 9(5):703–9. doi: 10.1038/nn1681.

Luan H, Peabody NC, Vinson CR, White BH. Refined spatial manipulation of neuronal function by combinatorial restriction of transgene expression. Neuron. 2006 Nov; 52(3):425–36. doi: 10.1016/j.neuron.2006.08.028.

Mais L, Hirsch P, Managan C, Wang K, Rokicki K, Svirskas RR, Dickson BJ, Korff W, Rubin GM, Ihrke G, Meissner GW, Kainmueller D. PatchPerPixMatch for Automated 3d Search of Neuronal Morphologies in Light Microscopy. bioRxiv. 2021 jul; https://doi.org/10.1101%2F2021.07.23.453511, doi: 10.1101/2021.07.23.453511.

Maniates-Selvin JT, Hildebrand DGC, Graham BJ, Kuan AT, Thomas LA, Nguyen T, Buhmann J, Azevedo AW, Shanny BL, Funke J, Tuthill JC, Lee WCA. Reconstruction of motor control circuits in adult Drosophila using automated transmission electron microscopy. bioRxiv. 2020; https://www.biorxiv.org/content/early/2020/01/11/2020.01.10.902478, doi: 10.1101/2020.01.10.902478.

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; 13(8):e0200759. doi: 10.1371/journal.pone.0200759.

Morante J, Desplan C. The color-vision circuit in the medulla of Drosophila. Curr Biol. 2008 Apr; 18(8):553–65. doi: 10.1016/j.cub.2008.02.075.

Morimoto MM, Nern A, Zhao A, Rogers EM, Wong AM, Isaacson MD, Bock DD, Rubin GM, Reiser MB. Spatial readout of visual looming in the central brain of Drosophila. Elife. 2020 11; 9. doi: 10.7554/eLife.57685.

Namiki S, Dickinson MH, Wong AM, Korff W, Card GM. The functional organization of descending sensory-motor pathways in Drosophila. Elife. 2018 06; 7. 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; 112(22):E2967–76. doi: 10.1073/pnas.1506763112.

Nojima T, Rings A, Allen AM, Otto N, Verschut TA, Billeter JC, Neville MC, Goodwin SF. A sex-specific switch between visual and olfactory inputs underlies adaptive sex differences in behavior. Curr Biol. 2021 03; 31(6):1175–1191.e6. doi: 10.1016/j.cub.2020.12.047.

Otsuna H, Bogovic JA, ya Takemura S, Shinomiya K, Saalfeld S, Kawase T. Robust search method for Drosophila neurons between electron and light microscopy; 2022, in preparation.

Otsuna H, Ito M, Kawase T. Color depth MIP mask search: a new tool to expedite Split-GAL4 creation. bioRxiv. 2018 may; https://doi.org/10.1101%2F318006, doi: 10.1101/318006.

Pfeiffer BD, Jenett A, Hammonds AS, Ngo TTB, Misra S, Murphy C, Scully A, Carlson JW, Wan KH, Laverty TR, Mungall C, Svirskas R, Kadonaga JT, Doe CQ, Eisen MB, Celniker SE, Rubin GM. Tools for neuroanatomy and neurogenetics in Drosophila. Proc Natl Acad Sci U S A. 2008 Jul; 105(28):9715–20. doi: 10.1073/pnas.0803697105.

Pfeiffer BD, Ngo TTB, 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.

Phelps JS, Hildebrand DGC, Graham BJ, Kuan AT, Thomas LA, Nguyen TM, Buhmann J, Azevedo AW, Sustar A, Agrawal S, Liu M, Shanny BL, Funke J, Tuthill JC, Lee WCA. Reconstruction of motor control circuits in adult Drosophila using automated transmission electron microscopy. Cell. 2021 02; 184(3):759–774.e18. doi: 10.1016/j.cell.2020.12.013.

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; 141(3):536–48. doi: 10.1016/j.cell.2010.02.025.

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.

Rokicki K, Bruns CM, Goina C, Schauder D, Olbris DJ, Trautman ET, Svirskas R, Clements J, Ackerman D, Kazimiers A, Foster LL, Dolafi T, Bolstad M, Otsuna H, Yu Y, Safford T, Murphy SD. Janelia Workstation Codebase. Janelia Research Campus. 2019 May; doi: 10.25378/janelia.8182256.v1.

Rokicki K, Clements J, Goina C, Kawase T, Olbris D, Otsuna H, Svirskas R. NeuronBridge: an intuitive web application for neuronal morphology search across large data sets; 2022, in preparation.

Sareen PF, McCurdy LY, Nitabach MN. A neuronal ensemble encoding adaptive choice during sensory conflict in Drosophila. Nat Commun. 2021 07; 12(1):4131. doi: 10.1038/s41467-021-24423-y.

Scheffer LK, Xu CS, Januszewski M, Lu Z, Takemura SY, Hayworth KJ, Huang GB, Shinomiya K, Maitlin-Shepard J, Berg S, Clements J, Hubbard PM, Katz WT, Umayam L, Zhao T, Ackerman D, Blakely T, Bogovic J, Dolafi T, Kainmueller D, et al. A connectome and analysis of the adult Drosophila central brain. Elife. 2020 09; 9. doi: 10.7554/eLife.57443.

Schretter CE, Aso Y, Robie AA, Dreher M, Dolan MJ, Chen N, Ito M, Yang T, Parekh R, Branson KM, Rubin GM. Cell types and neuronal circuitry underlying female aggression in Drosophila. Elife. 2020 11; 9. doi: 10.7554/eLife.58942.

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.

Sterne GR, Otsuna H, Dickson BJ, Scott K. Classification and genetic targeting of cell types in the primary taste and premotor center of the adult Drosophila brain. Elife. 2021 09; 10. doi: 10.7554/eLife.71679.

Takemura Sy, Bharioke A, Lu Z, Nern A, Vitaladevuni S, Rivlin PK, Katz WT, Olbris DJ, Plaza SM, Winston P, Zhao T, Horne JA, Fetter RD, Takemura S, Blazek K, Chang LA, Ogundeyi O, Saunders MA, Shapiro V, Sigmund C, et al. A visual motion detection circuit suggested by Drosophila connectomics. Nature. 2013 Aug; 500(7461):175–81. doi: 10.1038/nature12450.

Takemura Sy, Xu CS, Lu Z, Rivlin PK, Parag T, Olbris DJ, Plaza S, Zhao T, Katz WT, Umayam L, Weaver C, Hess HF, Horne JA, Nunez-Iglesias J, Aniceto R, Chang LA, Lauchie S, Nasca A, Ogundeyi O, Sigmund C, 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; 112(44):13711–6. doi: 10.1073/pnas.1509820112.

Tanaka R, Clark DA. Neural mechanisms to exploit positional geometry for collision avoidance. Curr Biol. 2022 Apr; doi: 10.1016/j.cub.2022.04.023.

Tirian L, Dickson BJ. The VT GAL4, LexA, and split-GAL4 driver line collections for targeted expression in the Drosophila nervous system. bioRxiv. 2017 oct; https://doi.org/10.1101%2F198648, doi: 10.1101/198648.

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; 79(1):128–40. doi: 10.1016/j.neuron.2013.05.024.

Wang F, Wang K, Forknall N, Parekh R, Dickson BJ. Circuit and Behavioral Mechanisms of Sexual Rejection by Drosophila Females. Curr Biol. 2020 10; 30(19):3749–3760.e3. doi: 10.1016/j.cub.2020.07.083.

Wang F, Wang K, Forknall N, Patrick C, Yang T, Parekh R, Bock D, Dickson BJ. Neural circuitry linking mating and egg laying in Drosophila females. Nature. 2020 03; 579(7797):101–105. doi: 10.1038/s41586-020-2055-9.

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 11; 526(16):2585–2611. doi: 10.1002/cne.24512.

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; 5. doi: 10.7554/eLife.21022.

Yu HH, Awasaki T, Schroeder MD, Long F, Yang JS, He Y, Ding P, Kao JC, Wu GYY, Peng H, Myers G, Lee T. Clonal development and organization of the adult Drosophila central brain. Curr Biol. 2013 Apr; 23(8):633–43. doi: 10.1016/j.cub.2013.02.057.

Yu Y, Peng H. Automated high speed stitching of large 3D microscopic images. In: *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro* IEEE; 2011. p. 238–241.

Zheng Z, Lauritzen JS, Perlman E, Robinson CG, Nichols M, Milkie D, Torrens O, Price J, Fisher CB, Sharifi N, Calle-Schuler SA, Kmecova L, Ali IJ, Karsh B, Trautman ET, Bogovic JA, Hanslovsky P, Jefferis GSXE, Kazhdan M, Khairy K, et al. A Complete Electron Microscopy Volume of the Brain of Adult Drosophila melanogaster. Cell. 2018 07; 174(3):730–743.e22. doi: 10.1016/j.cell.2018.06.019.

Zolin A, Cohn R, Pang R, Siliciano AF, Fairhall AL, Ruta V. Context-dependent representations of movement in Drosophila dopaminergic reinforcement pathways. Nat Neurosci. 2021 11; 24(11):1555–1566. doi: 10.1038/s41593-021-00929-y.

Supplement

Supplemental Table 1. Generation 1 MCFO samples included in the study.

Metadata for the included 74363 MCFO samples from 5155 Gen1 GAL4 lines is tabulated, including line name, landing site, effector, slide code, creation date, GUID, gender, heat shock duration, objectives, release name, and contributing annotator. See ***Table 1*** for effector codes.

Supplemental Table 2, related to Figure 4. Forward analysis individual scores for CDM & PPPM search results.

Table 1. Genotypes of MCFO UAS reporters

Bloomington ID (Janelia ID)	Reporter name	Genotype
64085 (1117734)	MCFO-1 (hsPESTOPT_attP3_ 3stop1_X_0036)	w, pBPhsFlp2::PEST in attP3 ; ; 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 / TM3,Sb
64086 (3022015)	MCFO-2 (pBPhsFLP_PEST_ HAV5_FLAG_OLLAS_ X3_0095)	w, pBPhsFlp2::PEST in attP3 ; ; pJFRC210-10XUAS-FRT>STOP>FRT-myr::smGFP-OLLAS in attP2, pJFRC201-10XUAS-FRT>STOP>FRT-myr::smGFP-HA in VK0005, pJFRC240-10XUAS-FRT>STOP>FRT-myr::smGFP-V5-THS-10XUAS-FRT>STOP>FRT-myr::smGFP-FLAG in su(Hw)attP1/TM2
64088 (1116898)	MCFO-4 (57C10wt_attp8_ 3stop1)	w, R57C10-Flp2 in su(Hw)attP8 ; ; 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
64089 (1116876)	MCFO-5 (57C10PEST_attp8_ 3stop1)	w, R57C10-Flp2::PEST in su(Hw)attP8 ; ; 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 / TM2
64090 (1116894)	MCFO-6 (57C10L_attp8_ 4stop1)	w, R57C10-FlpL in su(Hw)attp8 ; ; pJFRC210-10XUAS-FRT>STOP>FRT-myr::smGFP-OLLAS in attP2, 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 / TM2
64091 (1116875)	MCFO-7 (57C10PEST_attp18_ 4stop1)	w, R57C10-Flp2::PEST in attp18 ; ; pJFRC210-10XUAS-FRT>STOP>FRT-myr::smGFP-OLLAS in attP2, 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 / TM2
(3023700)	MCFO-3 derivative (57C10L_brp_SNAP_ MCFO_X23_0117)	w, 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
(3023701)	(57C10PEST_brp_SNAP_ MCFO_X23_0099)	w, 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
(3023951)	MCFO-1 derivative (pBPhsFlp2_PEST_ brp_SNAP_ MCFO_0128)	w, 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

MCFO-x reporters and names are described in ***Nern et al.*** (***2015***) and available from Bloomington Drosophila Stock Center or Janelia. brp::SNAP is described in ***Kohl et al.*** (***2014***). 'R57C10-Flp MCFO' in the text was JRC stock 3023701 for 94% of such samples, and JRC stock 3023700 for 6% of samples from sparser lines. 'hs-Flp MCFO' was JRC stock 3023951. See Table S1 for details of individual samples.

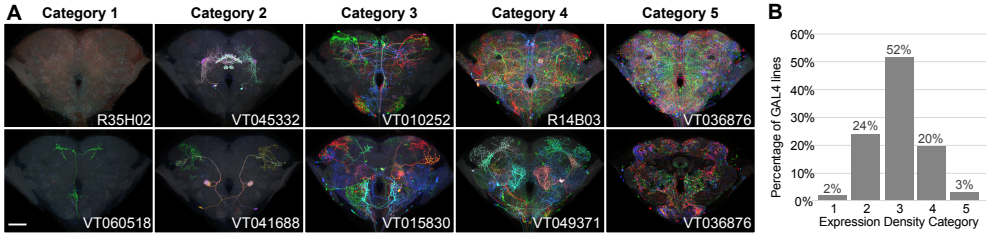


Figure 1–Figure supplement 1. Generation 1 MCFO expression density categories.

683 (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 of the 40x pipeline. 95% of lines were within the desired range.

684

Cell type	Body ID	CDM	PPPM	CDM only	PPPM only	CDM & PPPM	Total
LC12	1563393697	20	19	8	7	12	27
LPLC1	1220257051	27	13	21	7	6	34
LC11	1188885499	11	40	3	32	8	43
LC21	1281316775	19	22	8	11	11	30
LPLC2	1437850908	26	27	7	8	19	34
LT33	1818696317	8	7	3	2	5	10
LC4	1189559257	31	25	17	11	14	42
CT1	1311993208	8	3	6	1	2	9
DCH	1466485353	13	21	4	12	9	25
LC18	1722342048	13	24	5	16	8	29
	Average:	**17.6**	**20.1**	**8.2**	**10.7**	**9.4**	**28.3**
	Standard deviation:	8.3	10.6	6.1	8.7	4.9	11.5

Figure 4–Figure supplement 1. Table of all forward analysis results by cell type.

Table shows the number of lines independently identified by CDM & PPPM, number only identified by one approach (XOR), number identified by both approaches (AND), and total number identified (OR).

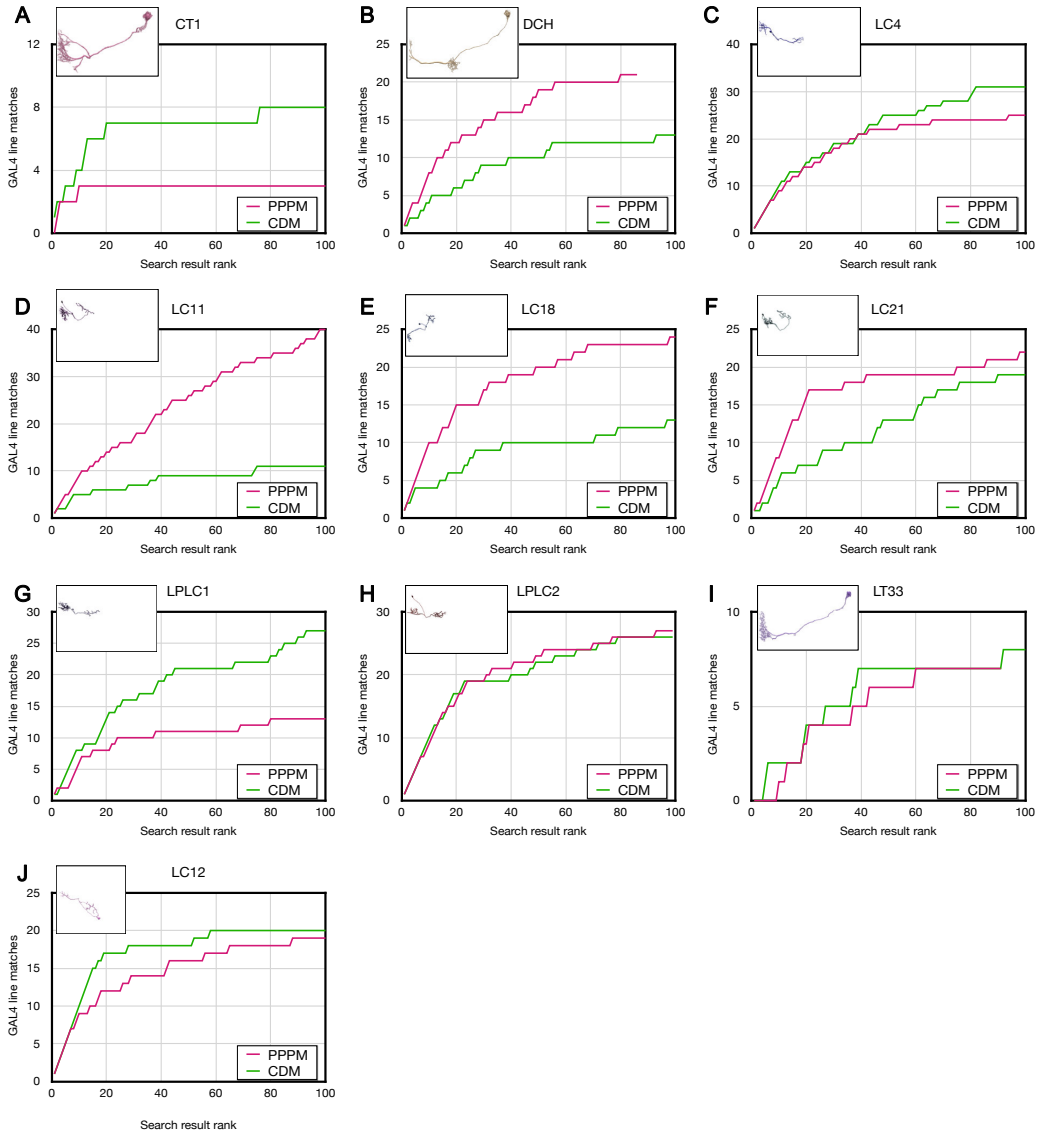


Figure 4–Figure supplement 2. Forward analysis individual plots for CDM & PPPM.
(A-J) Individual CDM & PPPM results for the indicated cell types. Includes duplicated images from Figure 4. EM images are from https://neuprint.janelia.org.