Cell Reports
Resource

A GAL4-Driver Line Resource for *Drosophila* Neurobiology

Arnim Jenett,[1] Gerald M. Rubin,[1,*] Teri-T.B. Ngo,[1] David Shepherd,[2] Christine Murphy,[1] Heather Dionne,[1] Barret D. Pfeiffer,[1] Amanda Cavallaro,[1] Donald Hall,[1] Jennifer Jeter,[1] Nirmala Iyer,[1] Dona Fetter,[1] Joanna H. Hausenfluck,[1] Hanchuan Peng,[1] Eric T. Trautman,[1] Robert R. Svirskas,[1] Eugene W. Myers,[1] Zbigniew R. Iwinski,[3] Yoshinori Aso,[1] Gina M. DePasquale,[1] Adrianne Enos,[1] Phuson Hulamm,[1] Shing Chun Benny Lam,[1] Hsing-Hsi Li,[1] Todd R. Laverty,[1] Fuhui Long,[1] Lei Qu,[1] Sean D. Murphy,[1] Konrad Rokicki,[1] Todd Safford,[1] Kshiti Shaw,[1] Julie H. Simpson,[1] Allison Sowell,[1] Susana Tae,[1] Yang Yu,[1] and Christopher T. Zugates[1]

[1]Janelia Farm Research Campus, Howard Hughes Medical Institute, 19700 Helix Drive, Ashburn, VA 20147, USA
[2]School of Biological Sciences, Bangor University, Deiniol Road, Bangor LL57 2UW, UK
[3]Carl Zeiss Microscopy, LLC, United States, 1 Zeiss Drive, Thornwood, NY 10594, USA
*Correspondence: rubing@janelia.hhmi.org
http://dx.doi.org/10.1016/j.celrep.2012.09.011

SUMMARY

We established a collection of 7,000 transgenic lines of *Drosophila melanogaster.* Expression of GAL4 in each line is controlled by a different, defined fragment of genomic DNA that serves as a transcriptional enhancer. We used confocal microscopy of dissected nervous systems to determine the expression patterns driven by each fragment in the adult brain and ventral nerve cord. We present image data on 6,650 lines. Using both manual and machine-assisted annotation, we describe the expression patterns in the most useful lines. We illustrate the utility of these data for identifying novel neuronal cell types, revealing brain asymmetry, and describing the nature and extent of neuronal shape stereotypy. The GAL4 lines allow expression of exogenous genes in distinct, small subsets of the adult nervous system. The set of DNA fragments, each driving a documented expression pattern, will facilitate the generation of additional constructs for manipulating neuronal function.

INTRODUCTION

The ability to express a gene of interest in a spatially restricted manner in a transgenic animal has greatly contributed to the successful use of *Drosophila* in a wide variety of biological studies. This is especially true for the nervous system, where experiments often call for manipulating the activity of small, reproducible sets of neurons (for review, see Venken et al., 2011; Griffith, 2012). Our objective was to generate and characterize a set of transgenic lines that can each drive expression of a reporter gene in a distinct, small subset of the ~150,000 neurons that comprise the adult central nervous system. The set of lines also needed to be large enough to ensure that all, or nearly all, neurons are represented in at least one line.

Transgenic animals in which the yeast transcriptional activator GAL4 is placed under the control of endogenous regulatory elements have proven to be powerful and versatile tools for manipulating gene expression (Fischer et al., 1988; Brand and Perrimon, 1993). Analysis of confocal imagery of whole mount preparations from animals in which a reporter gene is expressed under the control of GAL4 has been widely used for identifying and characterizing the morphology of neuronal populations, an approach pioneered and applied most extensively by Ito et al. (2003) (see also Otsuna and Ito, 2006; Tanaka et al., 2012).

Prior studies have used collections of GAL4-expressing lines based on enhancer traps. In an enhancer trap, a transposon carrying a GAL4 gene is inserted at a large number of random genomic locations; at each insertion site, the pattern of GAL4 expression comes under the influence of local transcriptional regulatory elements (O'Kane and Gehring, 1987). Enhancer traps have a number of properties that limit their use: the precise nature of the sequence elements driving expression is unknown, the varied genomic locations complicate genetic manipulations and, in general, GAL4 is expressed in more cells than optimal.

These limitations can often be overcome for anatomic studies. All that is required for such studies is the ability to recognize different neuronal populations under the microscope; stochastic labeling methods can be used to generate animals in which only a fraction of the cells in the parent GAL4 pattern express the reporter gene (for review, see Venken et al., 2011). But this is not the case when one intends to use the GAL4 driver to manipulate the activity of a specific population of neurons to study physiology or behavior. For such applications, the ability to direct expression of an exogenous protein to reproducible, small subsets of neurons in all animals in a population is critical.

The collection of GAL4 lines we generated for this project, based on the experimental design of Pfeiffer et al. (2008), begins to address those requirements. In these lines, a short fragment of genomic DNA controls GAL4 expression. As a result, GAL4 is, on average, expressed in fewer cells than in enhancer trap lines (Pfeiffer et al., 2008). Because the transgenes are all inserted at a known location by site-specific integration, genomic position effects on expression are held constant. Moreover, because the DNA fragments driving expression in each line are completely defined, one can make constructs that reuse the same enhancer fragment to drive the expression of another protein. Importantly, the expression pattern of that protein will

be predictable from the image data obtained with the original GAL4 line. Pfeiffer et al. (2010) show examples of such enhancer reuse in making constructs expressing the transcriptional activator LexA and the repressor GAL80.

The ability to reuse the enhancer fragments greatly enhances the value of the image database we report here. One can use our database not only for neuroanatomy or to select a GAL4 line of interest for further studies, but also as a resource to select enhancer fragments to generate the constructs required for more sophisticated applications. For example, LexA and GAL4 expressing constructs can be combined in the same animal to separately control expression of two reporters in different cell populations. This is required for a variety of experiments used to determine neuronal connectivity in circuit mapping (for review, see Griffith, 2012).

Although the GAL4 lines we present here are expressed more sparsely than enhancer trap lines, in most cases even more restricted expression will be required for behavioral studies where the effects of altering the activity of a very small population of neurons is desired. For example, GAL4 expression can be restricted to cells in which expression of two enhancer fragments overlap by using the split-GAL4 method, developed by Luan et al. (2006) and improved upon by Pfeiffer et al. (2010). This approach appears to be sufficient to reach single "cell-type" specificity. The modular nature of our constructs enables the facile construction of the required split-GAL4 transgenic lines.

With 6,650 lines successfully imaged, this is the largest data set of GAL4-driven expression patterns in the adult brain and the only large data set available of expression patterns in the adult ventral nerve cord (VNC). The expression patterns generated by these same lines in the embryonic nervous system and in third instar imaginal discs are described in accompanying articles (Manning et al., 2012 [this issue of *Cell Reports*]; Jory et al., 2012 [this issue of *Cell Reports*]). These combined data sets should be especially powerful for efforts to understand the logic underlying the *cis*-regulatory code.

The GAL4 lines have been deposited in the Bloomington *Drosophila* Stock Center for distribution and the expression patterns of the lines in several tissues are documented on a web-accessible database. We also developed a number of software tools to support our production pipeline and to analyze the large image data set we generated; we expect these tools and methods to be useful for other large-scale imaging projects.

RESULTS

Overall Strategy

We constructed transgenic lines in which the expression of the transcriptional activator GAL4 is driven by a defined DNA fragment that contains one or more transcriptional enhancers, using methods we have described previously (Pfeiffer et al., 2008). We then examined the expression patterns that these GAL4 drivers generated, in combination with a UAS-GFP reporter construct, using confocal microscopy of whole mounts of adult brains and VNCs. We developed a pipeline for dissection, immunohistochemistry, and imaging that enabled us to produce high quality data at scale. We also developed a variety of tools for browsing and annotating the image data. Our initial intent was to use expert anatomists to describe the expression patterns. Although an effective strategy for a small number of lines, or to identify neurons in a particular brain region in all the lines, such an approach did not scale well for a project of this size where our goal was to fully annotate thousands of lines. To overcome this limitation, we developed methods for machine-assisted annotation of the adult central brain. The patterns of expression of a subset of the lines were registered (Peng et al., 2011) on a standard model of the adult brain (K. Ito, personal communication) using features visible in a reference stain. Using manually constructed 3D masks, we computationally assigned aspects of the expression pattern in each line to one of the 68 major brain regions and then quantified the intensity and distribution of expression within each region. A human expert then vetted those annotations. The VNCs were annotated by an expert anatomist using a controlled vocabulary. These annotations permit text-based searching of a portion of the GAL4 image collection. The image database also allows a user to simply browse the data to identify lines that express GAL4 in a particular neuronal cell type, to uncover cell types not previously described, or to examine the patterns generated by all the DNA fragments surrounding a gene of interest.

Generation of Transgenic Lines and Image Data

We selected approximately 1,200 genes for which available expression data or predicted function implied expression in a subset of cells in the adult brain: for example, genes encoding transcription factors, neuropeptides, cell surface proteins, ion channels, transporters, and receptors. We spanned the flanking upstream and downstream intergenic regions of these genes, as well as any of their introns larger than 300 base pairs (bp), with fragments of DNA that averaged 3 kb in length and overlapped (in regions that could not be covered by a single fragment) by about 1 kb. These putative enhancer fragments were cloned into a vector containing the GAL4 coding region and a core promoter and then each construct was integrated into the same location in the fly genome by site-specific integration (Groth et al., 2004). We generated ~7,000 GAL4 lines. See Pfeiffer et al. (2008) and Experimental Procedures for details.

To determine the expression patterns generated by each GAL4 driver line, we developed a high throughput pipeline for brain dissection, immunohistochemistry, and imaging (see Experimental Procedures). Each GAL4 line was crossed to a UAS-GFP reporter line placing the expression of GFP under the positive control of GAL4 in the progeny, which were heterozygous for both the GAL4 driver and UAS reporter. Brains and VNCs were dissected from 3–5-day-old adult females, stained with an anti-GFP antibody and with a reference marker to allow identification of different brain regions and to provide the features used to align images obtained from different brains. We obtained high quality image data from 6,647 GAL4 lines and collected a total of over 47,000 confocal stacks. We developed several enhancements to methods previously employed for *Drosophila* brain and VNC dissection and immunohistochemistry, as well as software tools for data tracking, which are described in Experimental Procedures.

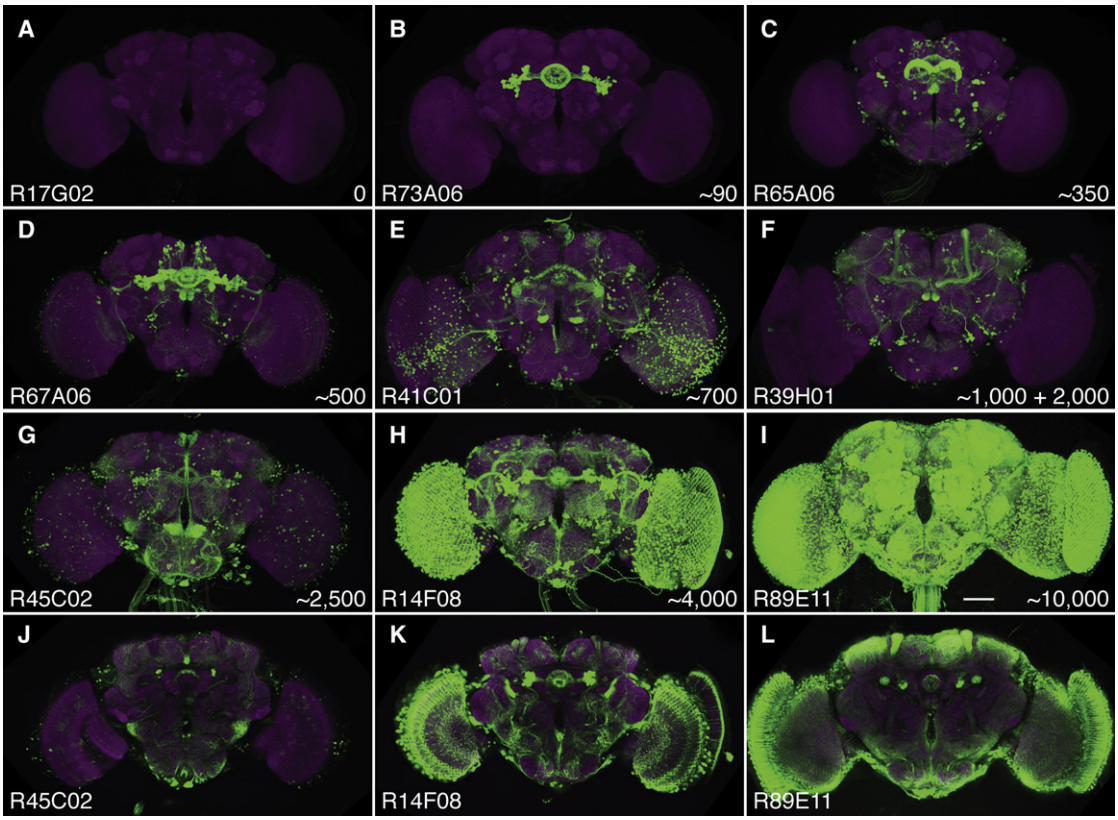

Figure 1. Maximum-Intensity Projections of Expression Patterns Seen in a Set of GAL4 Lines

(A) A line with no brain expression.

(B–I) Lines showing expression in the central complex sorted by increasing number of overall stained cell bodies. GAL4-driven GFP expression is shown in green. The nc82 reference stain, displayed at one-third intensity to prevent obstruction of the GFP signal, is shown in magenta. Line names are given in the lower left of each panel and the estimated number of cells expressing GFP in the central brain is shown in the lower right. For R39H01, shown in (F), 2,000 of the 3,000 estimated cells are MB Kenyon cells. Scale bar represents 100 μm (I).

(J–L) Individual optical sections from the specimens above (G–I) taken at the level of the frontal ellipsoid body. Note that, even in lines of the cell density shown in (J) and (K), fine details of the expression pattern can be seen.

Initial Characterization and Triage

Based on visual inspection of the confocal stacks (Figure 1), we judged that just over half of the lines had a density of labeling—as well as intense, crisp expression patterns—that would make them useful for future anatomical and behavioral experiments. These ranged from expression in ~20 (0.02%) to 5,000 (3%) neurons, excluding Kenyon cells, present in the central brain. We excluded Kenyon cells because the 5,000 Kenyon cells are tightly clustered in the mushroom body (MB) and do not obscure expression patterns elsewhere in the brain.

A total of 1,200 lines (17%) showed no obvious staining in the brain, although many of these showed expression in other developmental stages or tissues. Indeed, DNA fragments that show no enhancer activity in any tissue are rare; while we did not directly attempt to determine this number, a simple comparison of our data with the expression patterns driven by the same fragments in the third instar larva (J. Truman, personal communication; Jory et al., 2012) and embryo (Manning et al., 2012) indicates that this fraction is below 10% and would likely decrease further if more tissues were examined.

A total of 850 lines (12%) contain glial expression; in 320 lines expression was predominantly in glia, with little or no neuronal expression. All known glial cell types were observed, as well as several types not previously described (M. Kremer, personal communication). A total of 1,050 lines (15%) showed broad neuronal expression that we estimated to be in >5,000 cells in the central brain (excluding Kenyon cells; see for example, Figures 1I and 1L).

Some lines displayed obvious stochastic expression, where not all cells in the pattern showed expression in each brain. This was most easily observed in lines expressing in the repeated columnar structures of the optic lobe, where gaps in the pattern are readily recognized. An example of a nonstochastic line is shown in Figure 2G; note the even array of cell bodies in the medulla, which has a repeating columnar structure.

We chose not to further analyze broadly expressed lines, lines with no expression, lines with extensive glial expression, or those showing obvious stochastic expression. We present image data on all 6,647 lines at the FlyLight Image Database (http://www.janelia.org/gal4-gen1), but focused our more detailed analyses

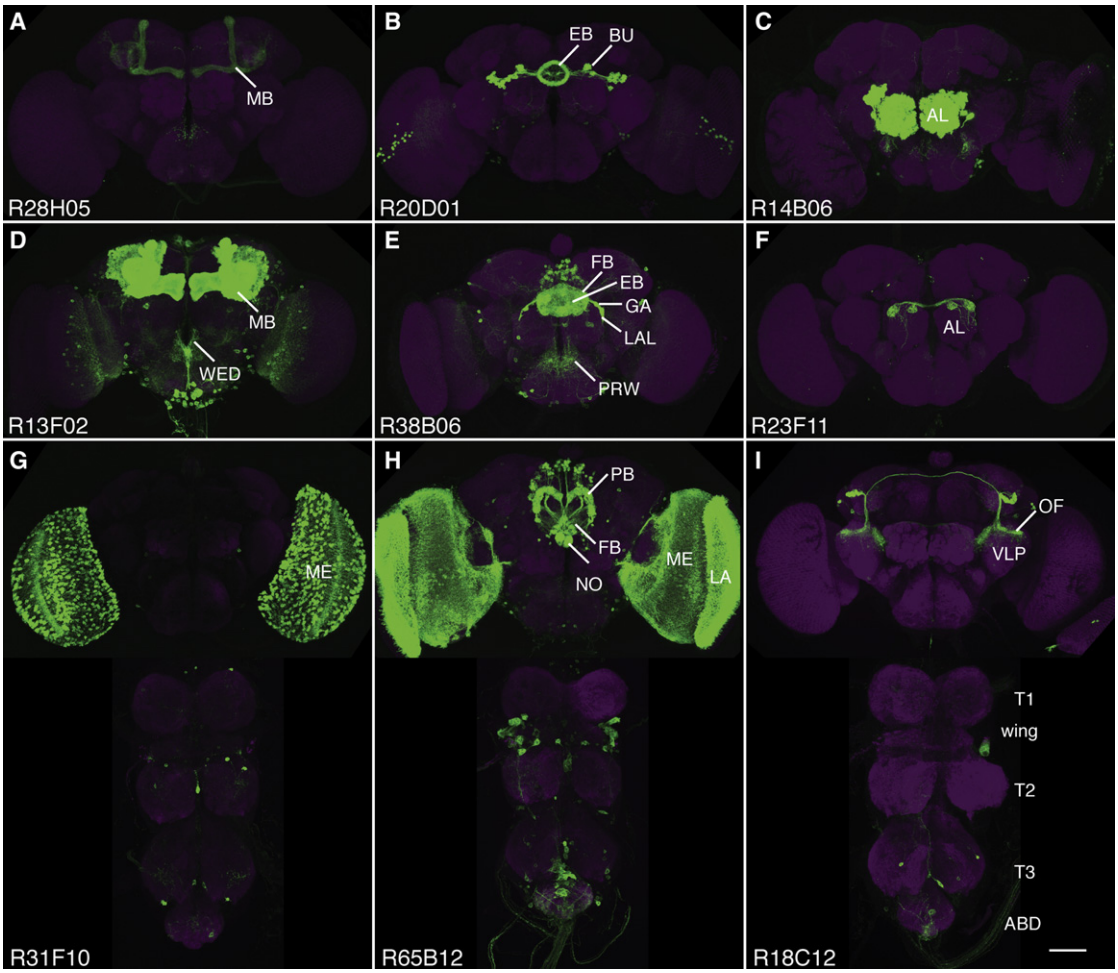

Figure 2. Maximum Intensity Projections of a Selected Set of Sparsely Expressing GAL4 Lines

(A–F) Only the central brain and optic lobes are shown.

(G–I) VNCs are also shown.

GAL4-driven GFP expression is shown in green. The nc82 reference stain, displayed at one-third intensity to prevent obstruction of the GFP signal, is shown in magenta. Line names are given in the lower left of each panel. Scale bar represents 100 μm (I). ABD, abdominal ganglia; AL, antennal lobe; BU, lateral bulb (lateral triangle); EB, ellipsoid body; FB, fan-shaped body; GA, gall; LA, lamina; LAL, lateral accessory lobe; MB, mushroom body; ME, medulla; NO, noduli; OF, optic focus; OL, optic lobe; PB, protocerebral bridge; PRW, prow; T1, thoracic ganglia 1; T2, thoracic ganglia 2; T3, thoracic ganglia 3; VLP, ventrolateral protocerebrum; wing, wing neuropil; WED, wedge.

on the remaining ~3,850 lines. Examples of what we considered very high quality lines, corresponding to the best few hundred lines in the collection, are shown in Figure 2.

Using the Image Database to Discover Neurons and to Evaluate Neuronal Stereotypy

A detailed mining of the image database is beyond the scope of this report. Here we present three examples to illustrate the types of studies that our data enable: (1) the identification of a new class of antennal lobe projection neurons; (2) a confirmation that the fly brain displays left-right asymmetry (albeit very limited); and (3) the use of a line with sparse expression to evaluate the degree of neuronal stereotypy. The transgenic GAL4 lines and the database of images we are providing should enable these types of studies to be performed by any skilled anatomist.

Many types of the antennal lobe projection neurons have been described by Golgi impregnation, stochastic labeling, and mosaic analysis with a repressible cell marker analysis of GAL4 drivers, and screening of sparsely labeled enhancer trap GAL4 drivers (reviewed in Vosshall and Stocker, 2007; see, for example, Stocker et al., 1990; Jefferis et al., 2007; Tanaka et al., 2008, 2012). Despite this intensive study, we were able to identify novel projection neurons in several GAL4 lines, including R19B06, R49A01, and R60H12. In Figure 3, we describe in detail the projection neurons from the VP3 glomerulus observed in R60H12.

Our initial visual screening of the expression patterns revealed one—but only one—clear case of an innervated structure that displayed left-right asymmetry in the adult brain. This structure appears to correspond to what Pascual et al. (2004) described

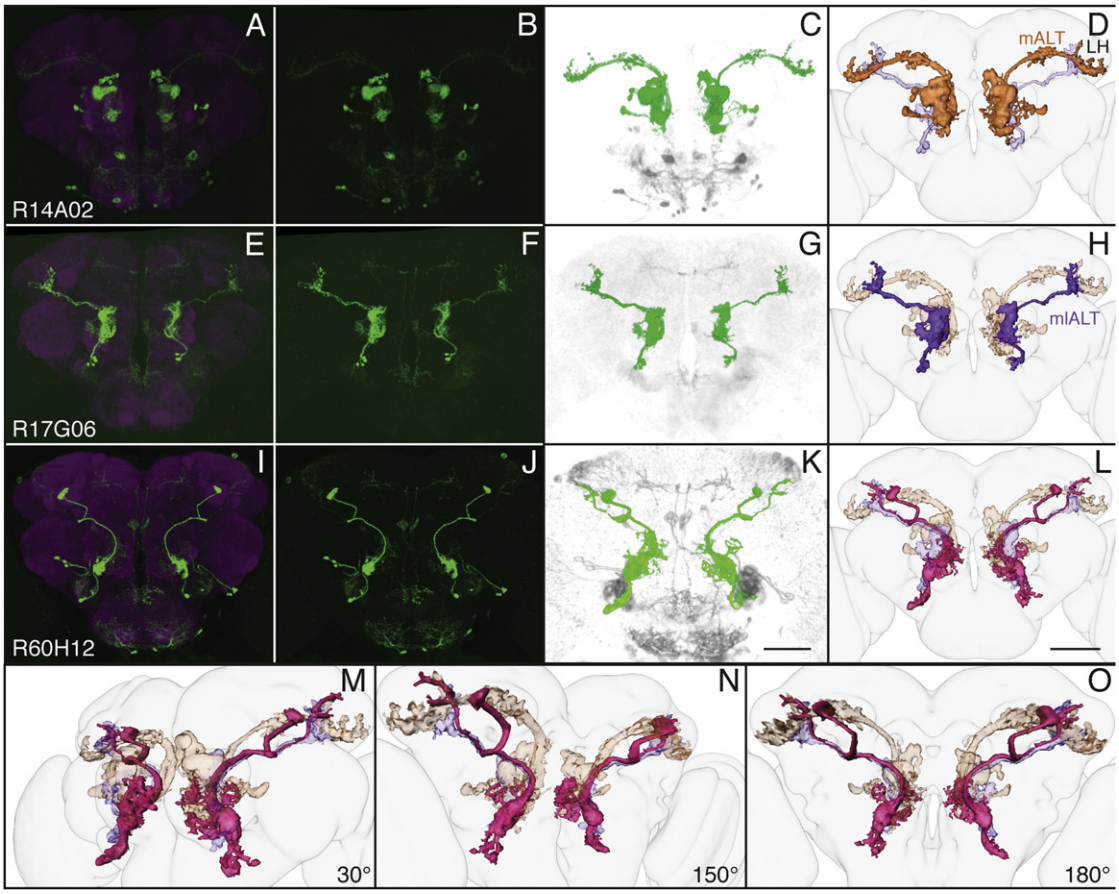

Figure 3. Examples of Antennal Lobe Projection Neurons

(A–L) R14A02 (A–D) and R17G06 (E–H) show examples of the two axon tracts used by most projection neurons, the medial antennal lobe tract (mALT) and the medio-lateral antennal lobe tract (mlALT), respectively. Both lines express in central AL glomeruli (R14A02: DL1, D1DA4m, VA7m, VA6; R17G06: DL1, DL5). In R14A02, the axons follow the mALT to the MB calyx and the LH, whereas in R17G06 they reach the LH via the mlALT. These patterns are displayed with increased transparency in (D), (H), and (L)–(O) where they serve to indicate the positions of the mALT (orange) and the mlALT (purple). R60H12 (I–L) shows a novel population of transverse antennal lobe projection neurons (ALtPN). The cell bodies of these neurons are located ventrally to the AL and project into the basal VP3 glomerulus establishing very dense arborizations in the dorsolateral part of this region. From there they follow the path of the mlALT until they reach the level of the fan-shaped body, where they turn medially, corkscrew halfway around the MB peduncle, and arborize densely in a very confined, bar shaped area of the anterior dorsolateral MB calyx. This region appears not to be the accessory calyx, which is described to be located more medial. Therefore we suggest naming this area lateral accessory calyx (lACA). Earlier descriptions of ALtPNs (Stocker et al., 1990; Tanaka et al., 2008, 2012) show a certain similarity to these neurons but the detailed morphology of the cells in R60H12 differ from those in earlier descriptions. Following the nomenclature of Tanaka et al. (2012), we suggest calling these cells AL-t5PN1. In addition to these neurons, a subpopulation in this expression pattern projects along the mlALT to the LH. With the original GAL4 pattern, it is not clear if they are a different population or if the same neurons also bifurcate to the MB calyx and the LH. However, the difference in the diameter of the proximal tracts suggests this expression pattern consists of two different cell types. (A, B, E, F, I, and J) Maximum intensity projections of GAL4-driven GFP expression patterns; including the reference staining (A, E, and I), or showing only the GFP signal (B, F, and J). (C, D, G, H, K, and L) Images were generated using Amira from the original confocal stacks by segmentation of the neurons of interest using local thresholding, followed by direct volume rendering or surface reconstruction. Color-enhanced direct volume renderings of the signal channel where the signal of interest is false-colored in green and the rest of the pattern is in gray (C, G, and K). Surface reconstructions of the signals of interest in relation to a surface model of the brain (D, H, and L). Scale bars represent 100 μm (K and L).
(M–O) Three-dimensional surface renderings from three different angles to display the morphologies of the neurons shown in (L).

as the asymmetric body (AB), a small area embedded on the right side of the central complex, ventrally in layer 1 of the fan-shaped body. They identified the AB in brains stained with an antibody against Fasciclin 2, a broadly expressed cell surface protein. They reported that a small fraction (7.6%) of wild-type flies displayed Fasciclin 2 positive structures in both hemispheres and that these animals had a defect in long-term memory. Because neurons contributing to the AB had not been identified, the opportunities for further mechanistic experiments were limited. We identified several GAL4 lines that innervate the AB (R38D01, R42C09, R52H03, R70H05, and R72A10) and show an anatomical analysis of R72A10 in Figure 4. In 22 of the 23 R72A10 brains examined, innervation of the AB was detected exclusively in the right side of the brain; in one brain, staining was detectable bilaterally, but clearly more extensive on the right side. In lines R38D01, R42C09, and R70H05 a majority of brains

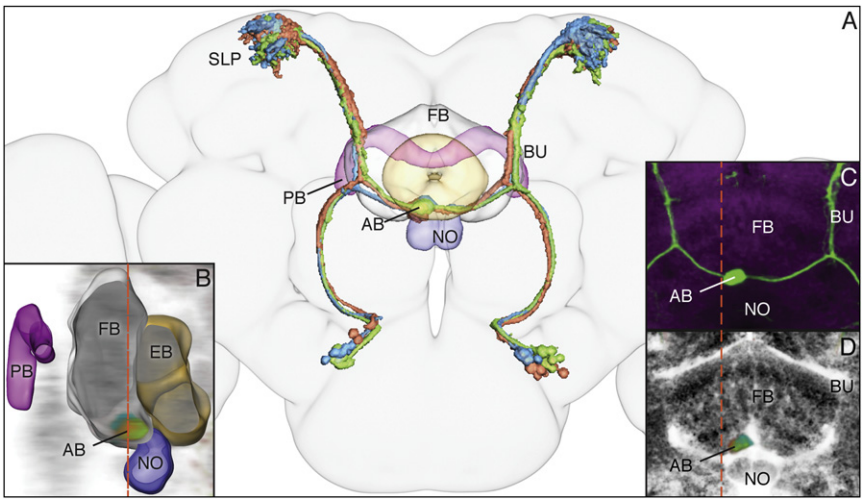

Figure 4. Brain Asymmetry Revealed by a Neuronal Population in Line R72A10

(A) A 3D surface reconstruction showing a superposition of this neuronal population in three brains; the same false colors are used in all panels.

(B) Sagittal section (3D direct volume, cut surface rendering) though the central complex showing the three color-coded specimens from (A) at the level of the AB.

(C) A maximum intensity projection of a 65-μm substack through the FB at the level of the AB showing the tracts from one specimen (shown in green in the other panels) innervating the AB. The center of projected volume corresponds to the dashed orange line in (B). The nc82 reference pattern, displayed at two-thirds intensity, is shown in magenta; GFP expression is in green.

(D) A maximum intensity projection of 10 μm of the three specimens shown in (A), illustrating the position of the AB relative to the FB and NO; the nc82 reference stain is shown in gray. The center of projected volume corresponds to the dashed orange line in (B). Note how the aborizations of the neurons from the three specimens, shown in different false colors, heavily overlap.

The dashed orange line in (C) and (D) indicates the section plane shown in (B). AB, asymmetric body; BU, lateral bulb; FB, fan-shaped body; NO, noduli; PB, protocerebral bridge; SLP, superior lateral protocerebrum.

displayed similarly asymmetric innervation; that is, in these GAL4 lines expression was usually bilateral, but staining on the right side was considerably more prominent than on the left side (data not shown).

Most neurons show reproducible overall morphologies, but the precise branching patterns of their fine arbors differ between animals. Examination of multiple brains from a GAL4 line that expresses sparsely enough to permit the identification of individual cell classes provides a way to estimate the extent and nature of such morphological variability. Chou et al. (2010) performed a study of this type of local interneurons in the antennal lobe (AL) and found a high degree of variability, but such studies have been limited. In Figure 5, we show an example of a population of about seven olfactory projection neurons that innervate glomerulus DL3. Consistent with earlier studies (Jenett et al., 2006), we observed a high degree of variability in the position of the cell bodies; the detailed morphology of the AL glomeruli is also known to vary somewhat between animals. We were, however, surprised by the high level of variability in the arborization patterns of these neurons in the areas of their synaptic targets in the MB calyx and the lateral horn. Differences could be seen not only between animals, but also between the two brain hemispheres of the same animal (Figure 5).

We only imaged the brains and VNCs of females in a systematic manner and so our data do not address the interesting question of the extent of sexual dimorphism. From a cursory comparison of expression patterns of 400 GAL4 lines in males and females, our impression is that <10% of lines show sexual dimorphism. Previous studies have shown that subsets of neurons that express *fruitless* (Cachero et al., 2010) or *doublesex* (Sanders and Arbeitman, 2008) are dimorphic.

Annotation

We developed a machine-assisted annotation process that abstracts the expression patterns displayed in the GAL4 lines in a way that they can be evaluated computationally. The process is initiated by aligning the confocal stacks for each brain to a standard brain model. This process transfers the expression pattern in each brain to a standard 3D coordinate system. Expression in any volume of interest (VOI) can then be quantitated. A 3D mask corresponding to that VOI is made manually. The mask specifies the voxels that correspond to the VOI. Although a VOI can be freely defined, we chose to use the standard regions defined by the Insect Brain Name Working Group (K. Ito, personal communication). We describe the expression pattern in a VOI with two numbers, one reflecting the intensity of the expression and one the staining distribution, or what fraction of the voxels show expression. Figure 6 illustrates these metrics as applied to a set of volumes derived from the confocal images of different GAL4 lines. In this case, the VOI was the fan-shaped body, one of the 68 standard brain regions. Our database provides for searching expression patterns, using these intensity and distribution values, of a selected subset (~3,800) of the lines.

We do not yet have computational tools capable of aligning the imaged VNCs into a standard framework. As this is a requirement for machine-assisted annotation, VNCs were annotated manually. The VNC expression patterns for 2,500 selected lines were described using a controlled vocabulary indicating the position of cell bodies, neuronal arbors, and afferent and efferent axons.

Data and Material Distribution

Image data on a representative brain and VNC of each of the 6,650 lines can be viewed at the FlyLight Image Database (http://www.janelia.org/gal4-gen1). We present maximum intensity projections of the confocal stacks as well as MPEG movies in which each frame corresponds to an ~1 μm optical section moving through the confocal stack along the z axis. The site also enables searching the image data by line name, the name

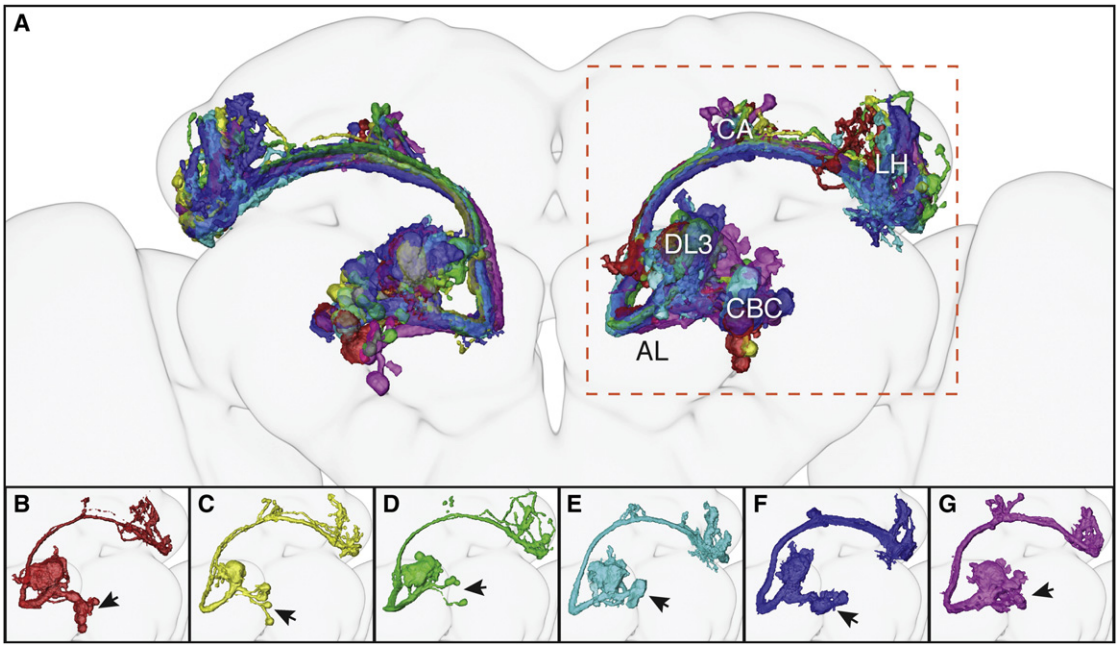

Figure 5. Three-Dimensional Surface Reconstructions of Six Specimens of Line R26B04 Illustrating Variability of the Same Small Cell Population between Individual Flies

(A) Although establishing the same overall connectivity between the DL3 glomerulus of the AL, the mushroom body calyx (CA) and the LH, the cells in this population (7 ± 1 cells) show extensive variability in fine morphology and cell body cluster (CBC) position when the brains of different animals are compared. Neurons from different animals are shown in different false colors.

(B–G) The area in the dashed box in (A) is shown separately for each specimen. The arrows indicate the position of the CBC.

of the gene from which the enhancer fragment was derived, or for expression in a given brain area. In addition to the data described in this article, data on the expression of the lines in third instar imaginal discs described in Jory et al. (2012) and expression in the embryonic CNS described in Manning et al. (2012) are included. The large size of the original confocal stacks (~500 megabytes each) limits on-line distribution. It is possible to download a small number of the original confocal stacks (laser scanning microscope [LSM] files) from the web server. We are also able to provide access to the original confocal stacks of the most useful 3,850 lines (~5 terabytes) on hard disk; requests should be sent to gal4-gen1@janelia.hhmi.org. The GAL4 lines have been deposited in the Bloomington *Drosophila* Stock Center for distribution (http://flystocks.bio.indiana.edu/Browse/misc-browse/Janelia.php).

DISCUSSION

The collection of GAL4 driver lines we describe here should be particularly useful for identifying and characterizing the morphology of neuronal populations in the adult fly for several reasons: (1) the lines express, on average, in fewer cells than the enhancer trap lines used in previous studies; (2) the number of lines we have imaged is larger than in any previous study; (3) we have imaged not only the brain but also the VNC, an important part of the central nervous system that has not generally been included in other studies; (4) we provide for convenient on-line browsing as well as an initial annotation of the expression patterns of the lines; and (5) we are making the GAL4 lines and our primary image data freely available.

We estimate that our collection contains ~3,850 lines in which the number of labeled central-brain neurons is in the range of 20 to 5,000. We expect that these lines will be adequate to represent the vast majority, if not all, cells in a variety of overlapping patterns. Establishing the completeness of the coverage in a rigorous way is not currently possible, but we note that for several well studied brain areas, such as the MB and its extrinsic neurons (Y. A., unpublished data) and the optic lobes (A. Nern, personal communication), expert annotators have been able to identify more than 90% of previously described cell types as well as many novel cell types.

We report the identification of a type of antennal lobe projection neuron not previously described, a cell class that has been heavily studied in previous work. We also present data establishing that the fly brain is asymmetric. However, asymmetry is rare, as we only observed one obviously asymmetric brain structure during our initial analysis of the collection. We believe that one of the major uses of the data we have generated will be further neuroanatomical studies, which can be carried out by simply browsing and mining the existing images. The intent of the work presented in this article was not to carry out such analyses, but to create a widely distributed resource that enables this activity.

The large number of lines in our collection that express in sparse populations of cells allows for a detailed examination of the degree of stereotypy in neuronal shape: the extent to

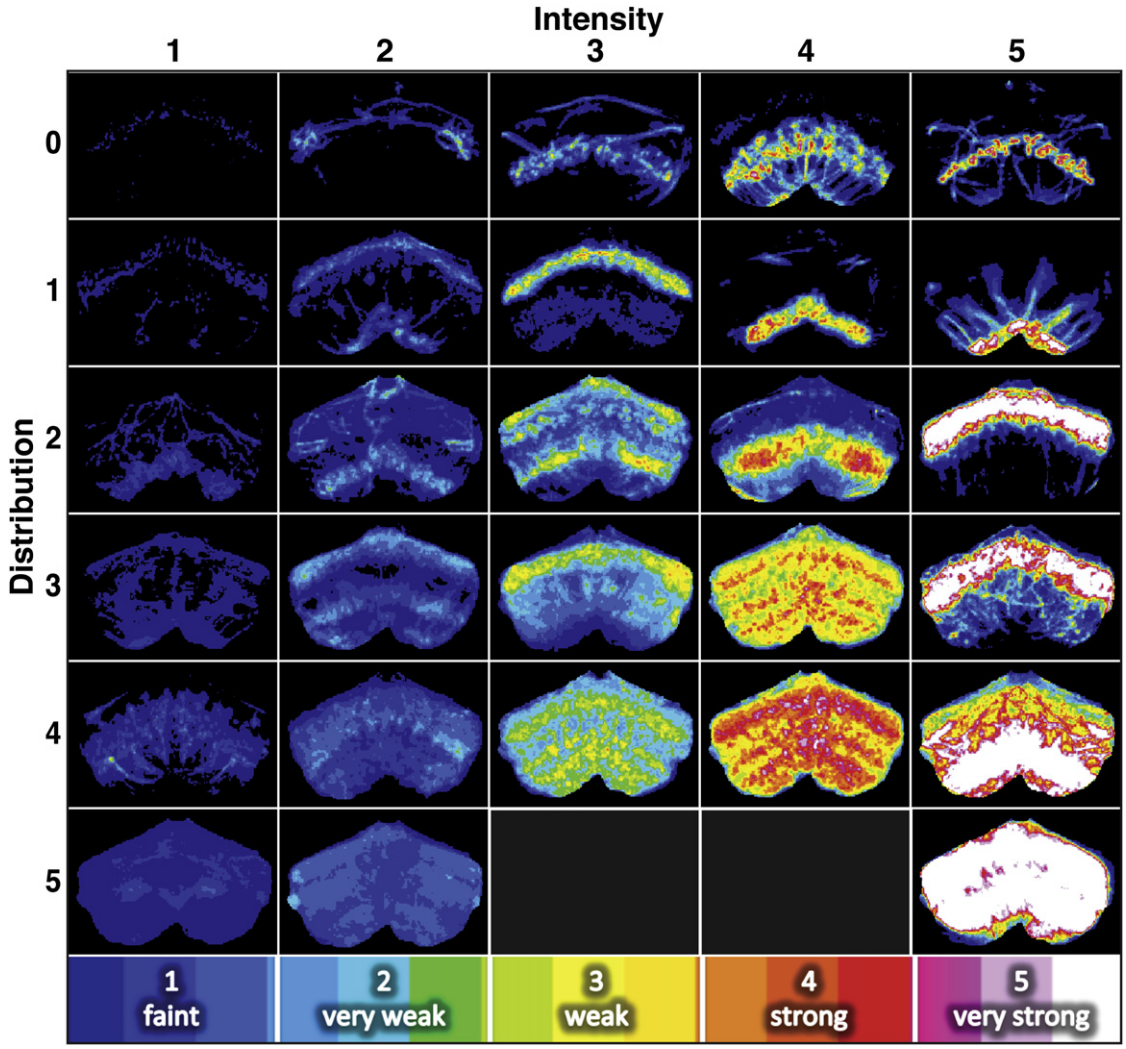

Figure 6. Examples of the Numerical Coding of Expression Intensity and Distribution Scales in the Fan-Shaped Body from Different GAL4 Lines

These maximum intensity projections of isolated FB volumes display false-color maps generated to facilitate the proofreading process. A false-color scale bar is shown.

which the same genetically defined cell population displays the same morphology from animal to animal. Initial neuronal wiring must be established to a large extent by a program specified in the genome. But how much variability these ''molecular algorithms'' generate in practice, and the extent to which this differs between cell types, has not yet been studied in detail. We present one example in this article for a population of antennal lobe projection neurons. Their display of extensive variability in the positions of axonal termini in the lateral horn implies that highly dynamic mechanisms are used for circuit formation. The resources we provide here should enable a much more comprehensive study of this question in diverse cell types.

Our analysis of the expression patterns generated by 6,650 DNA fragments confirms the initial impression from the small sample (45 fragments) described in Pfeiffer et al. (2008) about the density of enhancers (also called *cis*-regulatory modules or CRMs) in the *Drosophila melanogaster* genome. The genome contains >50,000 enhancers and multiple enhancers drive distinct subsets of expression of a gene in each tissue and developmental stage. But the GAL4 lines we describe are not individually adequate to provide cell-type specific expression. Even when single 3-kb DNA fragments are used to drive GAL4 expression, a tiny fraction—perhaps less than one in a thousand of patterns—is limited to a single neuronal population in the adult brain. The analyses of expression patterns generated by these same GAL4 lines in the embryo (Manning et al., 2012) and third instar imaginal discs (Jory et al., 2012) come to a similar conclusion. Taken together, these observations demonstrate that single enhancers only very rarely produce expression patterns limited to a single cell-type in a complex tissue.

EXPERIMENTAL PROCEDURES

Generation of Transformant Lines

We generated fragments by PCR from genomic DNA which were then cloned, sequence verified, and inserted upstream of a synthetic core promoter and the GAL4 coding region as described in Pfeiffer et al. (2008). In ~200 cases where the upstream intergenic region was small, we generated PCR fragments that also contained the start site of transcription and used them to create transcriptional fusion constructs. Constructs were inserted into the attP2 integration site using the phiC31 site-specific integration system (Groth et al., 2004) and homozygous stocks were generated. Embryo injections and screening for transformant flies were performed by GSI (Cambridge, MA); see Pfeiffer et al. (2008) for additional details. The primers used for each fragment and the molecular coordinates of the fragments on the *Drosophila* genome sequence can be downloaded at http://flystocks.bio.indiana.edu/Browse/misc-browse/Janelia.php and are also given at http://www.janelia.org/gal4-gen1.

Dissection and Immunohistochemistry

To achieve consistent image quality while producing a high number of specimens we developed standardized processes for dissection, immunohistochemistry, and confocal laser scanning microscopy. Males from each GAL4 line were crossed to females homozygous for the pJFRC2-10XUAS-IVS-mCD8::GFP reporter inserted at attP2 (Pfeiffer et al., 2010) and 3–5-day-old female adults heterozygous for the GAL4 driver and UAS reporter were dissected. Dissections were performed using Leica MZ9-5 stereomicroscopes. Flies were anesthetized with carbon dioxide gas, rinsed once in 70% ethanol, twice in S2 media (Schneider's Insect Medium, Sigma), submerged in cold S2 media on a dish coated with silicone (Sylgard 184, Dow Corning), and dissected with Dumont #5 stainless steel forceps. Dissection in S2 media resulted in better preservation of neuronal morphology than dissection in PBS. When dissecting complete central nervous systems, we first separated the VNC from the thorax and abdomen before dissecting the head. Once the nervous system was exposed, we removed membranes, trachea, and other tissues, taking care not to touch the brain, VNC, or neck connective. A well-dissected nervous system sinks in the media and has an intact connection between the VNC and the brain. A video of our dissection procedure can be found at http://www.janelia.org/team-project/fly-light#5064.

The nervous systems were transferred within 20 min of dissection to 1.5 ml Eppendorf Protein LoBind microcentrifuge tubes containing 1% paraformaldehyde (Electron Microscopy Science, Hatfield, PA) in S2 media on ice. After ensuring that the samples settled to the bottom of the tube, the tubes were placed on a rotating, rocking mixer (Fisher Scientific Nutating Mixer) at 4°C overnight. The next morning, the fixative solution was aspirated and replaced with cold PAT (0.5% Triton X-100, 0.5% bovine serum albumin in PBS) at 4°C. Samples were warmed to room temperature in a rack on the lab bench and were washed three times at room temperature with 1 ml of room temperature PAT for 1 hr per wash. They were then incubated in blocking buffer (3% normal goat serum [Invitrogen] in PAT) for 1 hr at room temperature after which the blocking buffer was replaced with 1 ml of a mixture of two primary antibodies in blocking buffer: rabbit anti-GFP (Invitrogen A11122) at 1:1,000 and mouse nc82 supernatant (Developmental Studies Hybridoma Bank, Iowa City, IA) at 1:50; mAb nc82 is directed against the Bruchpilot protein, a component of presynaptically-located T-bars (Wagh et al., 2006). Primary antibody incubations were overnight at 4°C with rocking.

After the primary antibody incubation, samples were returned to room temperature and were given three 1 hr washes in PAT. The last wash solution was replaced with 1 ml of a secondary antibody solution consisting of a 1:800 dilution of Alexa Fluor 488 goat anti-rabbit "green" (Invitrogen A11034) and a 1:400 dilution of Alexa Fluor 568 goat anti-mouse "red" (Invitrogen A11031) in blocking buffer and the samples were incubated for 4–5 days with rocking at 4°C in the dark. Following the secondary antibody incubation, tissues were washed several times with PAT as in previous steps. Just prior to mounting on a microscope slide, tissues were washed once for 20 min at room temperature in PBS.

An iterative process was used to optimize the dissection and staining protocol in which the quality of the computational alignment obtained from each imaged brain to a standard brain model was used as a metric. The BrainAligner (Peng et al., 2011) attempts to identify landmarks within the nc82 staining pattern of each brain and then uses the landmarks it can identify to warp the image to a collection of ~280 features in the standard brain. The two key factors in obtaining a good alignment were achieving high quality nc82 staining that penetrated the entire brain and avoiding damage to the brains during the dissection, staining, or mounting processes. The quality score, Qi, produced by the BrainAligner reports the fraction of features that were not found by the program. Scores of <0.30 correlated with high quality alignment as subjectively judged by visual inspection. Greater than 22,000 of the brains we imaged in our screen of the GAL4 lines aligned with a score <0.30. For over 5,500 different GAL4 lines, we have at least one brain with a score <0.30.

Imaging

For confocal imaging, samples were mounted on a 75 mm × 25 mm microscope slide (Fisherbrand SuperFrost Plus) to which we had applied an 8- or 10-well silicone adhesive spacer (custom ordered from Grace Biolabs). Approximately 40–50 samples were mounted per slide. The samples were mounted in a 9:1 mixture of Vectashield mounting solution (Vector Laboratories) and PBS. Wells were covered with a 10 mm round cover glass (Carl Zeiss, High Performance cover glass, #1.5 thickness, 0.17 ± 0.005 mm) held in place by nail polish. A video of our mounting procedure is posted at http://www.janelia.org/team-project/fly-light#5064.

Imaging was done on two Zeiss LSM 510 instruments with 20×, 0.80 numerical aperture plan-apochromat objectives; the motorized stage was modified to carry a custom two-slide holder. The pixel size was 0.62 μm × 0.62 μm with a z-section thickness of ~1 μm; the pixel dimensions of the stacks are 1024 × 1024 × ~150. We imaged with 12 bits of dynamic range and averaged four successive scans. Laser illumination was increased with z-position to counter attenuation. Fluorescence emission from the 488 nm excitation passed through a 505–550 nm bandpass filter and emission from the 561 nm excitation passed through a 575 nm longpass filter.

In collaboration with Carl Zeiss Microscopy, LLC, we developed a custom version of the Zeiss MultiTime software to enable the scanning of large numbers of selected samples in sequence. The "Add Next" feature simplified the setup of complex multiple channel Z stack image acquisitions over multiple locations. For improved image quality through Z stacks, the configuration of laser ramp points and laser power setup was simplified. In addition to other small usability enhancements, automated file management was added. Finally, MultiTime was made more robust so that it could seamlessly recover position and configuration information in the event of computer crash, power loss, or other unexpected interruption.

Our database shows image data for 6,647 lines. Although we generated ~7,000 transgenic lines, the remaining lines did not yield images of sufficiently quality in our first imaging attempt and their expression patterns were not sufficiently interesting to warrant re-imaging.

Data Handling

We developed a number of software tools to manage the volume of image data generated during the project. The large size of each confocal stack (~500 megabytes) necessitated the generation of smaller representations of the original data to facilitate analysis by enabling rapid data loading over network connections. Three types of derived representations were generated for each confocal stack: (1) a maximum intensity projection of the full stack as an ~300 kilobyte JPEG file (http://www.w3.org/Graphics/JPEG/jfif3.pdf); (2) a series of ~10 μm substack maximum intensity projections as JPEG files; and (3) a movie in which each slice of a stack is turned into a movie frame. By using the MPEG-4 compression algorithm (http://mpeg.chiariglione.org/standards/mpeg-4/mpeg-4.htm) we were able to produce movies with file sizes of ~5 megabytes, ~100 times smaller than the original confocal stack, while introducing only minor compression artifacts. During the production of the movies, the image data in each frame were contrast optimized to improve the ability to see weak signals. A calibration bar is included in each frame, which displays the maximum and minimum intensities in the original image. The calibration bars should be used when judging the strength of an expression pattern. We produced all projections in two versions, one with the reference channel (nc82 staining) and one without. The former facilitates the localization of expression pattern elements relative to the reference channel in which the

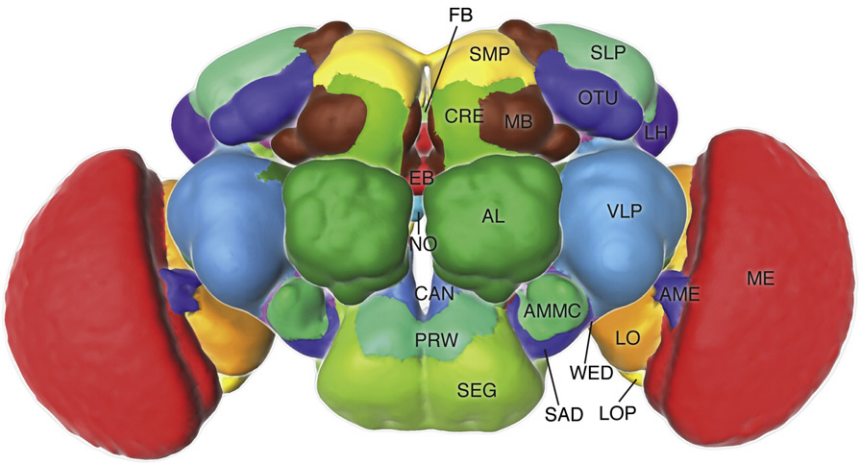

Figure 7. The 68 Major Brain Regions Shown in False Color

Additional details of the nomenclature for brain areas will be published elsewhere (K. Ito, personal communication). AL, antennal lobe; AME, accessory medulla; AMMC, antenna-mechanosensory and motor center; CAN, cantle; CRE, crepine; EB, ellipsoid body; FB, fan-shaped body; LH, lateral horn; LO, lobula; LOP, lobula plate; MB, mushroom body; ME, medulla; NO, noduli; OTU, optic tubercle; PRW, prow; SAD, saddle; SLP, superior lateral protocerebrum; SMP, superior medial protocerebrum; SEG, subesophageal ganglion; VLP, ventrolateral protocerebrum; WED, wedge.

various neuropil of the brain can be identified. The latter permits the detection of faint aspects of the expression pattern, which might otherwise be masked by the reference channel.

In the course of our analysis, we noted that the imaging process produced a primary image file (LSM format) that was inverted in the left-right axis relative to the specimen. Because we wanted to provide primary data with minimal manipulation, and we wanted the derived data files to be consistent with the primary data, we did not correct for this inversion. The sole exception was in the case of the asymmetric body data shown in Figure 4, where left-right orientation was central to data description.

We developed a standardized file name for the image files that provides human readable information about the file contents. Another benefit of the standardized file name is that it is very easy to parse, which allows the file system to function as a simple machine searchable image database. The structure of this naming convention is shown in Figure S1 and the renaming software, and the file structure of the data are described in Extended Experimental Procedures.

In order to distribute the data to the research community we developed a web interface (http://www.janelia.org/gal4-gen1) written in Perl and PHP, and fed by a MySQL relational database. The data provided in this way are primarily the derived data sets, as the large size of the original confocal stacks limits on-line distribution.

Annotation

We developed a number of software tools and scripts, which we refer to collectively as the Fiji Annotation Suite (FAS), to facilitate manual annotation of the images by a human expert. These ImageJ macros, shell scripts, and helper applications work inside of Fiji (http://fiji.sc/) to access files without requiring the user to interact with the Unix file system. They facilitate adding descriptive metadata (annotations) to image data by streamlining and organizing the workflow; for example, accelerating data entry while analyzing an image or video by assigning annotation terms from a controlled vocabulary to specific keys of the keyboard. Although initially developed for this project, these tools are flexible enough to be used for the annotation of any structured collection of image data. The FAS will be described in detail elsewhere; it is open source (GLP) and is available upon request.

For fast visual screening and sorting selected subsets of the collection, we also developed a shell script that generates JavaScript-enabled HTML pages from a list of TransformantIDs or SpecimenIDs. For each entry on an input list, these pages display the maximum intensity projections and a JavaScript test field that can be used to describe the expression pattern in a freeform text.

The expression patterns in the VNC were annotated manually using a controlled vocabulary to describe the position of cell bodies, neuronal arbors, and afferent and efferent axons. The terminology for all peripheral nerves and most of the neuropil regions has been previously described in detail (Power, 1948). An article describing all the neuropil regions and peripheral nerves of the VNC is in preparation (D.A., R. Court, and D. Shepherd, unpublished data).

Machine-assisted annotation was performed on image data that had been aligned to a standard brain model using the BrainAligner software described by Peng et al. (2011). Because of the requirement that expression patterns need to be transferred to the standard brain map with a high degree of precision, it was only possible to apply this approach to specimens that could be aligned with high quality (Qi scores below 0.30). After alignment, the expression patterns generated by each GAL4 line can be described using the same coordinate system. By defining the volumes that correspond to the brain areas in the standard map, it was possible to computationally perform annotation in a semiquantitative way. A set of 3D masks was generated manually using Amira (Visage Imaging GmbH) corresponding to the volume associated with each of 68 major brain areas as defined in the standard brain map of *D. melanogaster* (Figure 7; K. Ito, personal communication).

To abstract the intensity and distribution of the expression within the VOI from the details of the morphology of the GAL4 positive neurons, we generated a 4-bit histogram (16 bins) of all voxels within the VOI. This number of bins was determined empirically; it proved to be a good compromise between preserving data density and maximizing the visual perceptibility of brightness-values expressed in false colors. To assist in communicating these results we generated a version of the data in which the intensity and the distribution within each brain region (what fraction of the voxels within that brain region show expression) within each brain region are each represented by an integer from 0 to 5 (see Figure 6). We found that these numerical representations reflect human perception of the distribution and intensity of the expression patterns and allow efficient querying of the database for lines that express in a given brain region. The result of the machine-assisted annotation is an abstraction of the expression patterns displayed by each GAL4 line into 136 scores representing the intensity and distribution of expression in each of 68 brain areas.

ACCESSION NUMBERS

The GAL4 lines have been deposited in the Bloomington *Drosophila* Stock Center for distribution (http://flystocks.bio.indiana.edu/Browse/misc-browse/Janelia.php).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures and one figure and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2012.09.011.

LICENSING INFORMATION

Cell PRESS
Open ACCESS

ACKNOWLEDGMENTS

We thank Hua-Peng Liaw, Omotara Ogundeyi, Nicholas Abel, Emily Willis, Ying-Jou Lee, and Rebecca Vorimo for assistance with data acquisition; Karen Hibbard, Jessica Keating, James McMahon, Megan Hong, Monti Mercer, Grace Zheng, Jui-Chun Kao, and Danielle Ruiz for fly husbandry; Christopher Bruns, Nikolay Kladt, Frank Midgley, Lowell Umayam, and Charlotte Weaver for software development; Geoffrey Meissner and Eric Hoopfer for help in evaluating sexual dimorphism; and Bruce Kimmel and Crystal Sullivan for administrative support.

Received: August 16, 2012
Revised: September 14, 2012
Accepted: September 17, 2012
Published online: October 11, 2012

WEB RESOURCES

The URLs for data presented herein are as follows:

Bloomington *Drosophila* Stock Center, http://flystocks.bio.indiana.edu/Browse/misc-browse/Janelia.php
Fiji Image Processing Software, http://fiji.sc/
FlyLight Image Database, http://www.janelia.org/gal4-gen1
JPEG File Interchange Format, http://www.w3.org/Graphics/JPEG/jfif3.pdf
Overview of the MPEG-4 Standard, http://mpeg.chiariglione.org/standards/mpeg-4/mpeg-4.htm
Videos of Brain Dissection and Mounting Processes, http://www.janelia.org/team-project/fly-light#5064

REFERENCES

Brand, A.H., and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development *118*, 401–415.

Cachero, S., Ostrovsky, A.D., Yu, J.Y., Dickson, B.J., and Jefferis, G.S. (2010). Sexual dimorphism in the fly brain. Curr. Biol. *20*, 1589–1601.

Chou, Y.H., Spletter, M.L., Yaksi, E., Leong, J.C., Wilson, R.I., and Luo, L. (2010). Diversity and wiring variability of olfactory local interneurons in the *Drosophila* antennal lobe. Nat. Neurosci. *13*, 439–449.

Fischer, J.A., Giniger, E., Maniatis, T., and Ptashne, M. (1988). GAL4 activates transcription in *Drosophila*. Nature *332*, 853–856.

Griffith, L.C. (2012). Identifying behavioral circuits in *Drosophila melanogaster*: moving targets in a flying insect. Curr. Opin. Neurobiol. *22*, 609–614.

Groth, A.C., Fish, M., Nusse, R., and Calos, M.P. (2004). Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31. Genetics *166*, 1775–1782.

Ito, K., Okada, R., Tanaka, N.K., and Awasaki, T. (2003). Cautionary observations on preparing and interpreting brain images using molecular biology-based staining techniques. Microsc. Res. Tech. *62*, 170–186.

Jefferis, G.S., Potter, C.J., Chan, A.M., Marin, E.C., Rohlfing, T., Maurer, C.R., Jr., and Luo, L. (2007). Comprehensive maps of *Drosophila* higher olfactory centers: spatially segregated fruit and pheromone representation. Cell *128*, 1187–1203.

Jenett, A., Schindelin, J.E., and Heisenberg, M. (2006). The Virtual Insect Brain protocol: creating and comparing standardized neuroanatomy. BMC Bioinformatics *7*, 544.

Jory, A., Estella, C., Giorgianni, M.W., Slattery, M., Laverty, T.R., Rubin, G.M., and Mann, R.S. (2012). A survey of 6300 genomic fragments for cis-regulatory activity in the imaginal discs of Drosophila melanogaster. Cell Rep. *2*. Published online October 11, 2012. http://dx.doi.org/10.1016/j.celrep.2012.09.011.

Luan, H., Peabody, N.C., Vinson, C.R., and White, B.H. (2006). Refined spatial manipulation of neuronal function by combinatorial restriction of transgene expression. Neuron *52*, 425–436.

Manning, L., Purice, M.D., Roberts, J., Pollard, J.L., Bennett, A.L., Kroll, J.R., Dyukareva, A.V., Doan, P.N., Lupton, J.R., Strader, M.E., et al. (2012). Annotated embryonic CNS expression patterns of 5,000 GMR GAL4 lines: a resource for manipulating gene expression and analyzing cis-regulatory modules. Cell Rep. *2*. Published online October 11, 2012. http://dx.doi.org/10.1016/j.celrep.2012.09.011.

O'Kane, C.J., and Gehring, W.J. (1987). Detection in situ of genomic regulatory elements in *Drosophila*. Proc. Natl. Acad. Sci. USA *84*, 9123–9127.

Otsuna, H., and Ito, K. (2006). Systematic analysis of the visual projection neurons of *Drosophila melanogaster*. I. Lobula-specific pathways. J. Comp. Neurol. *497*, 928–958.

Pascual, A., Huang, K.L., Neveu, J., and Préat, T. (2004). Neuroanatomy: brain asymmetry and long-term memory. Nature *427*, 605–606.

Peng, H., Chung, P., Long, F., Qu, L., Jenett, A., Seeds, A.M., Myers, E.W., and Simpson, J.H. (2011). BrainAligner: 3D registration atlases of *Drosophila* brains. Nat. Methods *8*, 493–500.

Pfeiffer, B.D., Jenett, A., Hammonds, A.S., Ngo, T.T., Misra, S., Murphy, C., Scully, A., Carlson, J.W., Wan, K.H., Laverty, T.R., et al. (2008). Tools for neuroanatomy and neurogenetics in *Drosophila*. Proc. Natl. Acad. Sci. USA *105*, 9715–9720.

Pfeiffer, B.D., Ngo, T.T., Hibbard, K.L., Murphy, C., Jenett, A., Truman, J.W., and Rubin, G.M. (2010). Refinement of tools for targeted gene expression in Drosophila. Genetics *186*, 735–755.

Power, M.E. (1948). The thoracico-abdominal nervous system of an adult insect, *Drosophila melanogaster*. J. Comp. Neurol. *88*, 347–409.

Sanders, L.E., and Arbeitman, M.N. (2008). Doublesex establishes sexual dimorphism in the Drosophila central nervous system in an isoform-dependent manner by directing cell number. Dev. Biol. *320*, 378–390.

Stocker, R.F., Lienhard, M.C., Borst, A., and Fischbach, K.F. (1990). Neuronal architecture of the antennal lobe in Drosophila melanogaster. Cell Tissue Res. *262*, 9–34.

Tanaka, N.K., Tanimoto, H., and Ito, K. (2008). Neuronal assemblies of the *Drosophila* mushroom body. J. Comp. Neurol. *508*, 711–755.

Tanaka, N.K., Endo, K., and Ito, K. (2012). The organization of antennal lobe-associated neurons in the adult *Drosophila melanogaster* brain. J. Comp. Neurol. Published online May 17, 2012. http://dx.doi.org/10.1002/cne.23142.

Venken, K.J., Simpson, J.H., and Bellen, H.J. (2011). Genetic manipulation of genes and cells in the nervous system of the fruit fly. Neuron *72*, 202–230.

Vosshall, L.B., and Stocker, R.F. (2007). Molecular architecture of smell and taste in *Drosophila*. Annu. Rev. Neurosci. *30*, 505–533.

Wagh, D.A., Rasse, T.M., Asan, E., Hofbauer, A., Schwenkert, I., Dürrbeck, H., Buchner, S., Dabauvalle, M.C., Schmidt, M., Qin, G., et al. (2006). Bruchpilot, a protein with homology to ELKS/CAST, is required for structural integrity and function of synaptic active zones in *Drosophila*. Neuron *49*, 833–844.