The natverse: a versatile computational toolbox to combine and analyse neuroanatomical data

James D. Manton,[1,†] Alexander Shakeel Bates,[1,†] Sridhar R. Jagannathan,[2] Marta Costa,[1,2] Philipp Schlegel,[2] Torsten Rohlfing,[3] and Gregory S. X. E. Jefferis[1,2,*]

[1] MRC Laboratory of Molecular Biology, Cambridge, CB2 0QH, UK
[2] Drosophila Connectomics Group, Department of Zoology, University of Cambridge, Downing Street, Cambridge, CB2 3EJ UK
[3] SRI International, Neuroscience Program, Center for Health Sciences, Menlo Park, CA, USA
[†] These authors contributed equally to this work
[*] jefferis@mrc-lmb.cam.ac.uk

Abstract

To analyse neuron data at scale, neuroscientists expend substantial effort reading documentation, installing dependencies and moving between analysis and visualisation environments. To facilitate this, we have developed a suite of interoperable open source R packages called the natverse. The natverse allows users to read local and remote data, perform popular analyses including visualisation, clustering and graph theoretic analysis of neuronal branching. Unlike most tools, the natverse enables comparison of morphology and connectivity across many neurons after imaging or co-registration within a common space. The natverse also enables transformations between different template spaces and imaging modalities. We demonstrate tools that integrate the vast majority of Drosophila neuroanatomical light microscopy and electron microscopy connectomic datasets. The natverse is an easy-to-use environment for neuroscientists to solve complex, large-scale analysis challenges as well as an open platform to create new code and packages to share with the community.

1 Introduction

Neuroanatomy has become a large-scale, digital and quantitative discipline. Improvements in sample preparation and imaging increasingly enable the collection of large 3D image volumes containing complete neuronal morphologies in the context of whole brains or brain regions. Neuroscientists therefore need to tackle large amounts of morphological data, often writing custom code to enable repeated analysis using their specific requirements. They also need to analyse neuronal morphology and connectivity in the context of whole nervous systems or sub-regions. However, it is desirable not to rewrite basic functionalities such as reading various types of data file, representing neurons in different data structures, implementing spatial transforms between samples, or performing popular analyses from scratch. Scaling up or developing custom analysis strategies is simpler and more feasible for researchers if they can reuse existing infrastructure. This has been amply demonstrated by flexible but open-source platforms such as ImageJ/Fiji for image analysis (Schindelin et al., 2012) or Bioconductor for bioinformatics (Huber et al., 2015). One important consequence of these free and open-source tools is that they aid collaboration and reproducibility. Together, these considerations have motivated us to create the NeuroAnatomy Toolbox (nat) and its extensions, which we detail in this paper.

A number of software tools are already available to analyse neuronal data (Billeci et al., 2013; Glaser and Glaser, 1990; Schmitz et al., 2011; Rasband, 2011; Meijering et al., 2004; Pool et al., 2008; Ho et al., 2011; Narro et al., 2007; Brown et al., 2005; Wearne et al., 2005; Gensel et al., 2010; Cuntz et al., 2010; Kim et al., 2015; Saalfeld et al., 2009; Katz and Plaza, 2019; Feng et al., 2015; Myatt et al., 2012; Peng et al., 2014). However, most focus on image processing and the morphological analysis options available are fairly basic, such as examining arbour lengths or performing Sholl analyses (Sholl, 1953). Of these, the trees toolbox (Cuntz et al., 2010) has particularly strong support for morphological analysis of neurons, but focuses on individual neurons in isolation rather than neurons within the volume of the brain as a whole.

Recent technological advances have made acquiring large amounts of neuronal morphology data in their whole-brain contexts feasible across phyla (Chiang et al., 2011; Zheng et al., 2018; Li et al., 2019; Kunst et al., 2019; Jenett et al., 2012; Cook et al., 2019; Economo et al., 2016; Oh et al., 2014; Ohyama et al., 2015; Winnubst et al., 2019; Ryan et al., 2016). Image data are typically registered against a template space, allowing one to compare data from many brains directly and quantitatively. This significantly aids the classification of neuronal cell types because it allows type classification relative to the arbours of other neuronal types (Sümbül et al., 2014) and anatomical subvolumes. However, while this enables the comparison of data within a given study, template spaces are often different across studies or laboratories, hindering data integration.

This paper describes the Neuroanatomy Toolbox (nat), a general purpose open source R-based package for quantitative neuroanatomy, and a suite of extension R packages that together we call the natverse. A distinctive feature of the natverse, as compared with other available tools, is to analyse neurons within and across template spaces and to simplify access to a wide range of data sources. Neurons can be read from local files or from online repositories (Ascoli et al., 2007; Jenett et al., 2012; Kunst et al., 2019; Chiang et al., 2011; Economo et al., 2016; Winnubst et al., 2019) and web-based reconstruction environments (Saalfeld et al., 2009;

Schneider-Mizell et al., 2016; Katz and Plaza, 2019). The natverse can be installed in one line of code as described on the project website (https://natverse.org). Every function is documented with a large number of examples based on bundled or publicly available data. The code to generate the figures in this manuscript is available through https://github.com/jefferislab/nat.examples. We provide online community support through our nat-user mailing list: https://groups.google.com/forum/#!forum/nat-user.

We now give an overview of the natverse and showcase a number of common applications. The natverse has recently been employed for large scale analysis of zebrafish data (Kunst et al., 2019) and we provide examples across a range of invertebrate and vertebrate species. We then give more specific examples focussing on cell type identification across Drosophila datasets. Using the natverse, we have created bridging registrations that transform data from one template to another along with mirroring registrations (e.g. left-to-right hemisphere) and made these easily deployable. This unifies all publicly available Drosophila neuroanatomical datasets, including those image data for genetic resources and whole brain connectomics.

2 Results

2.1 Software packages for neuroanatomy

We have opted to develop our software in R, a leading platform for bioinformatics and general data analysis. R is free and open source, and is supported by high quality integrated development environments (e.g. RStudio). It features a well-defined system for creating and distributing extension packages that bundle code and documentation. These can easily be installed from high-quality curated repositories (CRAN, Bioconductor) as well as via GitHub. R supports a range of reproducible research strategies including reports and notebooks and integrates with the leading cross-platform tools in this area (jupyter, binder).

The core package of the natverse is the Neuroanatomy Toolbox, nat. It supports 3D visualisation and analysis of neuroanatomical data (Figure 1a), especially tracings of single neurons (Figure 1b). nat allows a user to read neuronal data from a variety of popular data formats produced by neuron reconstruction tools (Figure 1a). Typical image analysis pipelines include imaging neurons with confocal microscopy, reconstructing them using ImageJ's Simple Neurite Tracer (Longair et al., 2011) then saving them as SWC files (Cannon et al., 1998); nat can read a collection of such files with a single command. In addition, a user can, for example, mark the boutons on each neuron using ImageJ's point tool and export that as a CSV file, load this into nat and then analyse the placement of these synaptic boutons with respect to the originally traced neuron (Supplementary Figure 1).

We have extended nat by building the natverse as an ecosystem of interrelated R packages, each with a discrete purpose (Figure 1a). The natverse is developed using modern software best practices including revision control, code review, unit testing, continuous integration, and code coverage. Developing sets of functions in separate packages helps compartmentalise development, ease troubleshooting and divides the natverse into documentated units that users can search to find the more specific code examples or functions that they need. To the casual user, these divisions may initially be of little consequence. We therefore provide a single wrapper package, natverse; installing this results in the installation of all packages and their dependencies, immediately giving the user all the capabilities described in this paper (Figure 1a). natverse packages have already been used in recent publications from our lab (Frechter et al., 2019; Dolan et al., 2019, 2018; Huoviala et al., 2018; Jefferis et al., 2007; Cachero et al., 2010; Grosjean et al., 2011; Costa et al., 2016) and others (Felsenberg et al., 2018; Kunst et al., 2019; Zheng et al., 2018; Clemens et al., 2018; Eichler et al., 2017; Saumweber et al., 2018; Jeanne et al., 2018; Clemens et al., 2015). Confirmed stable versions of nat, nat.templatebrains, nat.nblast, nat.utils and nabor can be downloaded from the centralised R package repository, CRAN, with development versions available from our GitHub page (https://github.com/natverse/).

In brief, natverse packages can be thought of as belonging to four main groups (Figure 1a). The first two support obtaining data, either by a) interacting with repositories and software primarily used for neuron reconstructions from electron micrograph (EM) data, including CATMAID, NeuPrint and DVID or b) interacting with repositories for light-level data, including MouseLight, FlyCircuit, Virtual Fly Brain, NeuroMorpho, the InsectBrainDB and the FishAtlas projects. Additional R packages help with c) manipulating and deploying registrations to move data between brainspaces, and d) data analysis and visualisation (see Materials and methods for additional details).

2.2 Manipulating neuroanatomical data

2.2.1 Neuron skeleton data

Raw 3D images enable true to life visualisation but simplified representations are usually required for data analysis. For example, neurons can be traced to generate a compact 3D skeleton consisting of 3D vertices joined by edges. The natverse provides functions for morphological and graph-theoretic analyses of neurons, collections of neurons, neurons as vector clouds and neurons as tree graphs (Figure 2a). The natverse represents skeletonised neurons as neuron objects, with the neuron's geometry in the standard SWC format where each node in the skeleton has its own integer identifier. There are additional data fields (Supplementary Figure 2): the treenode IDs for branch points, the location of its synapses in 3D space and their polarity, including the source file, leaf nodes and series of IDs that belong to continuous non-branching segments of the neuron (Supplementary Figure 2).

Neurons have tree like structures that are critical to their function (Cuntz et al., 2010). `ngraph` data objects represent a neuron as a graph originating at a root (usually the soma) with directed edges linking each of the neuron's tree nodes (Figure 2a). This representation provides a bridge to the rich and efficient graph theory analysis provided by the igraph package (Csardi and Nepusz, 2005).

Objects of class `neuronlist` are collections of `neuron` objects, representing one or more neurons, with some at-

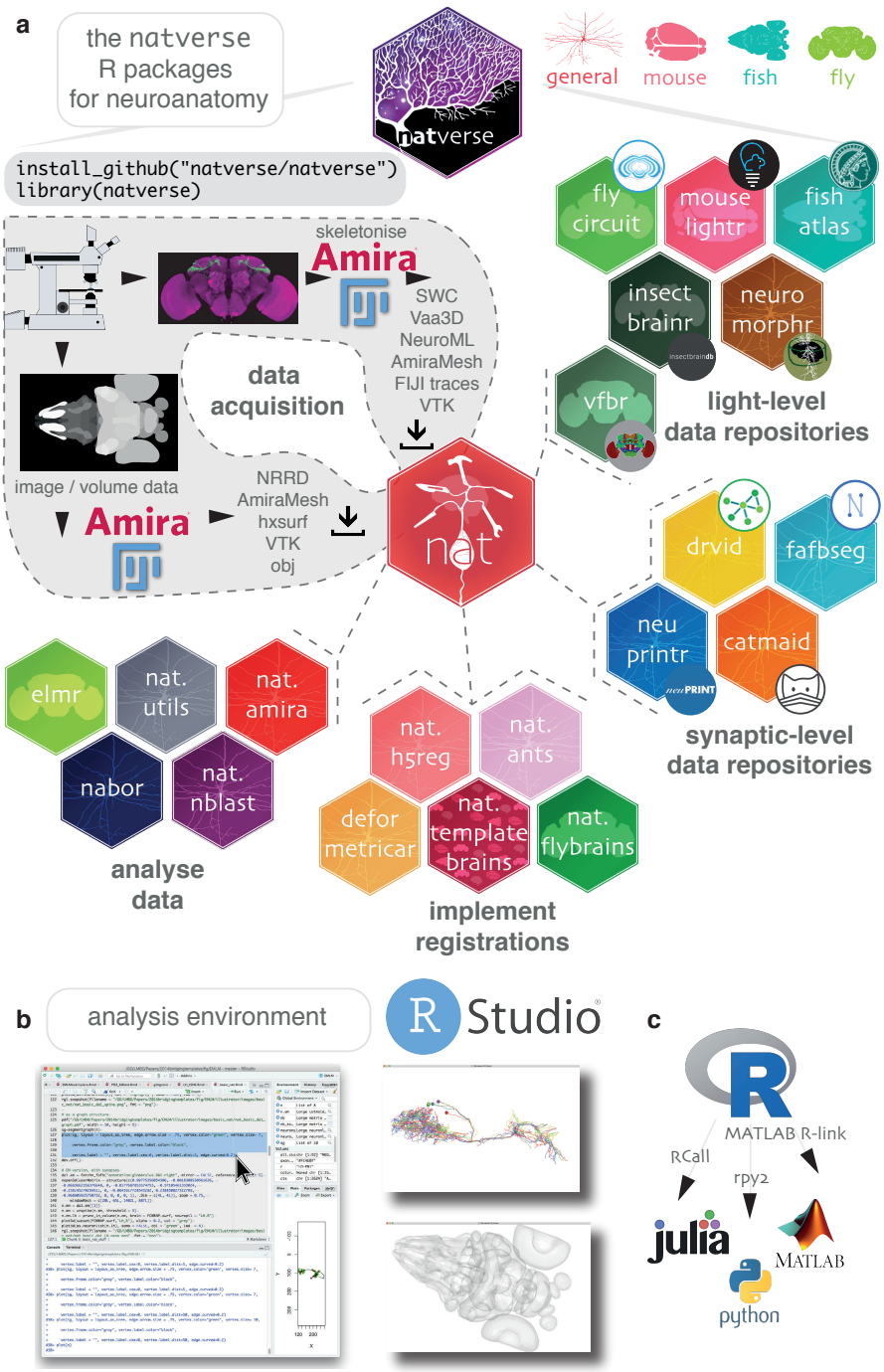


Figure 1: The natverse. **a** R packages that constitute the natverse. Packages are coloured by whether they are general purpose, or cater specifically for *Mus musculus*, *Danio rerio* or *Drosophila melanogaster* datasets. Coarse division into packages for fetching remote data, implementing registrations and analysing data are shown. Data, as output by most reconstruction pipelines, can also be read by nat. **b** The natverse is designed to work best in the RStudio environment, by far the most popular environment in which to execute and write R code. 3D visualisation is based on the R package rgl, which is based on OpenGL. It runs on Mac, Linux and Windows platforms. **c** R functions can be called by other popular scientific programming languages; some example packages/libraries are shown. P. Schlegel has developed Python code inspired by and parallel with the natverse for analysing neuron morphology, NAVIS (https://github.com/schlegelp/navis), and talking to the CATMAID API, PyMaid (https://github.com/schlegelp/pyMaid). These python libraries transform natverse-specific R objects into Python data types and can call natverse R functions like `nblast`.

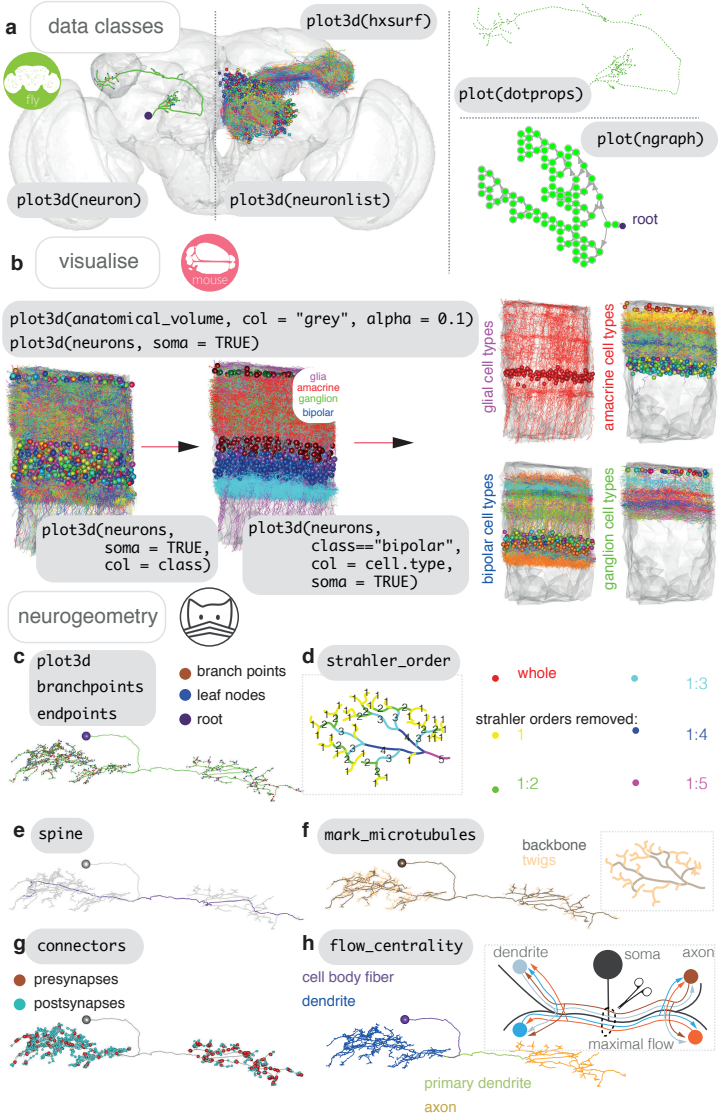


Figure 2: Neurons in nat. **a** Data classes defined by nat. A *D. melanogaster* DA1 olfactory projection neuron (Costa et al., 2016) is shown as part of four different data types, useful for different sorts of analyses: as a `neuron` object (left), as part of a larger `neuronlist` of multiple olfactory projection neurons (middle), as a vector-cloud `dotprops` object (right, upper) and an `ngraph` object (right, lower). In grey, the FCWB template brain, an `hxsurf` object from the package nat.flybrains, is shown. Generic code for visualizing these data types is shown is grey boxes. **b** Visualisation, with generic sample code, of connectomic data from a dense reconstruction inner plexiform layer of the mouse retina is shown, coloured by the cell class and then cell type annotations given by their reconstructors (Helmstaedter et al., 2013). Because this dataset contains many neuron fragments that have been severely transected, we only consider skeletons of a total cable length greater than 500 μm using functions `summary` and `subset`. Somata are shown as spheres. **c** A synaptic-resolution neuron reconstruction for a *D. melanogaster* lateral horn neuron (Dolan et al., 2018) has been read from a live CATMAID project hosted by Virtual Fly Brain (https://fafb.catmaid.virtualflybrain.org/) using catmaid, and plotted as a `neuron` object. It is rooted at the soma, consistent with the convention. **d** Boxed, Strahler order is a measure of branching complexity for which high Strahler order branches are more central to a neuron's tree structure, and the lower order ones more peripheral, such that branches with leaf nodes are all Strahler order 1. Main, the same neuron which has had its lower Strahler order branches progressively pruned away. **e** We can extract the longest path through a neuron, its 'spine', purple, a method that could help define the tracts that a neuron might traverse. **f** Boxed, in insect neurons, the main structure of the neuron is supported by a microtubular backbone, but as it branches its more tortuous, smaller caliber neurites loose the microtubule, and make more postsynapses (Schneider-Mizell et al., 2016). Main, in CATMAID users can tagged the tree nodes that mark the position where neurite loses its microtubular backbone, so an R user can use prune family functions to remove, add or differentially colour microtubular backbone versus twigs. **g** Both presynapses and postsynapses can been manually annotated in CATMAID, and be visualised in R. Because neurons from CATMAID have synaptic data, they are treated as a slightly different class by the natverse, called `catmaidneuron`. A `neuronlist` can also contain many `catmaidneuron` objects. **h** Boxed, neurons can be split into putative dendrite and axon using the flow centrality algorithm from Schneider-Mizell et al. (2016). In short, a path is drawn from every post- to presynapse, and the point in the neuron where the most paths cross is used to split the neuron in two. Main, exemplar neuron algorithmically split into compartments.

tached metadata. This attached metadata can give information such as a neuron's name, some unique identifier, its cell type, etc (Supplementary Figure 2). An `ngraph`, `neuron` or `neuronlist` can be passed to many functions in the natverse, and also to other functions familiar to R users for which we have written specific methods. For example, users can call `subset` on a `neuronlist` to select only those neurons with a certain entry in their corresponding metadata, e.g. all amacrine cells. Methods passed to `plot3d` enable a `neuronlist` to be plotted with colours based on its metadata entries (Figure 2b); in this case connectomic data from the inner plexiform layer of the mouse retina (Helmstaedter et al., 2013) is shown coloured by cell type. Many functions are built to work with `neuron` objects, but will also have a method that allows them to be applied to every `neuron` in a given `neuronlist` via the nat function `nlapply`. R users will be familiar with this logic from using the base R function `lapply`.

2.2.2 Basic analysis

A useful function with methods for `neuron` objects and `neuronlist` objects is `summary`. This gives the user counts for tree nodes, branch points, leaf nodes and the total combined cable length of a neuron (Supplementary Figure 3a). We can further use the natverse to identify points on a neuron that have particular properties based on the neuron's skeleton structure (Figure 2c-e) or because we have some other data that identifies the position of some biological feature (Figure 2f-g), or both (Figure 2h). Branching can be assessed by a Sholl analysis (`sholl_analysis`) (Supplementary Figure 3b) or decomposed by Strahler order (Supplementary Figure 3c). Geodesic distances can be calculated between any tree node in the graph (Supplementary Figure 3c), with the help of functions from the R package igraph (Csardi and Nepusz, 2005). Nearest neighbour distances in Euclidean space can be calculated using our R package nabor.

Some reconstruction environments allow tree nodes to be tagged with extra information, e.g. CATMAID. This can include microtubules (Figure 2e) and pre- and postsynapses (Figure 2f). This information is fetched when the catmaid package reads a neuron. It can be used by a graph theoretic algorithm (Schneider-Mizell et al., 2016) (Figure 2g, inset) to divide a neuron into its dendrites, axon and intervening cable (Figure 2g). We put this information together in the example in Supplementary Figure 3c, which shows the geodesic distribution of pre- and postsynapses along three neurons arbors, split by axon and dendrite, then further by Strahler order, then further by presence or absence of microtubule. Here, for our three exemple neurons, presynapses only exist on the larger calibre, microtubule containing backbone of the neuron, and are present in high number except at the highest Strahler orders. In contrast postsynapses are mainly on twigs (i.e. fine terminal branches) at Strahler order 1–2. We can also identify connected neurons using catmaid functions, and see that the dendrites of these cells receive distinct olfactory and memory inputs.

2.2.3 Neuroanatomical volumes

The natverse also helps users analyze neuronal skeletons with respect to volume objects that might represent neuroanatomical structures on a scale from a whole brain to neuropil subvolumes. 3D objects from diverse sources can be visualized and analyzed with nat, and we can calculate their volumes (Figure 3a). By using the nat function `make_model`, a user can interactively create their own 3D objects from, for example, points from a neuron's cable or its synapses (Figure 3b). This could be used to look at the volume enclosing a dendritic arbour, which may correlate with other features of a neuron (Figure 3b). The nat function `pointsinside` can tell a user which synapses lie within a 3D mesh object such as a standard neuropil segmentation while `prune_in_volume` can cut a skeleton to include or exclude the cable within a given volume (Figure 3c).

2.2.4 Advanced analysis

Because the natverse is a flexible platform that allows users to easily write their own custom R code, very specific analyses can be performed. For example, we might be interested in using skeletons to define anatomical subvolumes and analyzing the logic of neuronal projections between these subvolumes. For Supplementary Figure 4 we developed custom code on top of core natverse functionality to examine light-level *D. melanogaster* olfactory projections to, and target neurons with dendrites in a subregion of the brain called the lateral horn (Grosjean et al., 2011; Chiang et al., 2011; Frechter et al., 2019). We voxelized the lateral horn as well as its target regions into overlapping kernel density estimates based on agglomerating similarly shaped subbranches for projection neuron axons. This analysis reveals a structure of distinct subregions in a classic neuropil, and the 3D locations that are likely to receive input from these new subregions (Supplementary Figure 4d). The natverse contains other functions to infer connectivity from light-level data, including `potential_synapses`, an implementation of a synapse prediction algorithm that makes use of spatial proximity and approach angle (Stepanyants and Chklovskii, 2005), and `overlap`, a simpler algorithm that measures the putative overlap in Euclidean space between neuron pairs (Frechter et al., 2019).

2.2.5 Cell typing neurons

Neuronal cell type is a useful classification in neuroscience. It constitutes a testable hypothesis, that shared morphology indicates functional equivalency (Bates et al., 2019). Neuronal cell typing can be done by expert examination (Helmstaedter et al., 2013), purely by morphological clustering (Jeanne et al., 2018), or a combination of both (Frechter et al., 2019). Many neurogeometric algorithms for assessing similarity exist. Some are invariant to the 3D embedding space (Sholl, 1953; Wan et al., 2015; Li et al., 2017), but those that are dependent on neurons' relative positioning in a template space (Cardona et al., 2010; Costa et al., 2016; Zhao and Plaza, 2014) can be more performant when they can be applied (Zhao and Plaza, 2014; Li et al., 2017).

NBLAST (Costa et al., 2016) is a recent morphological similarity algorithm that has been used in a range of publications

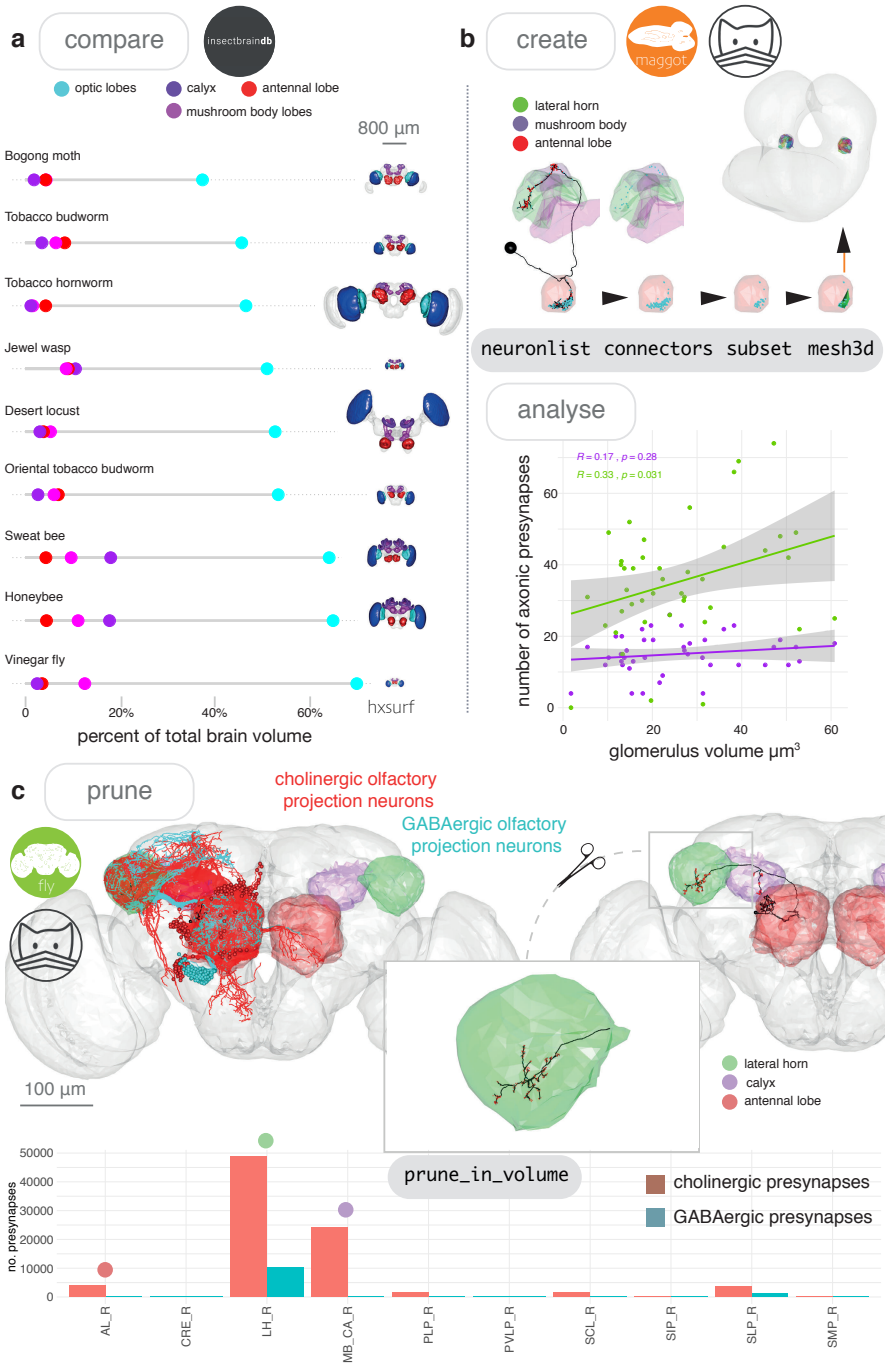


Figure 3: Neuroanatomical models with nat. **a** We accessed the InsectBrainDB.org via insectbrainr to obtain template brains for different species of insect (Brandt et al., 2005; de Vries et al., 2017; El Jundi et al., 2018; Heinze and Reppert, 2012; Kurylas et al., 2008; Løfaldli et al., 2010; Stone et al., 2017; Zhao and Plaza, 2014). The package insectbrainr converts retrieved OBJ files into `hxsurf` objects, which contain one set of 3D points for each whole brain, and then different sets of edges between these points to form 3D neuropil subvolumes. These subvolumes were already defined by expert annotators. Their volume are compared across insect brain, normalised by total brain size. Insect template brain data curated by: S. Heinze, M. Younger, J. Rybak, G. Pfuhl, B. Berg, B. el Jundi, J. Groothuis and U. Homberg. **b** We can create our own subvolumes by pulling synaptic neuron reconstructions (Berck et al., 2016) from a first-instar larva EM dataset (Ohyama et al., 2015) (a public Virtual Fly Brain CATMAID instance), extracting dendritic post synapses from olfactory projections neurons, and using synapse clouds from neurons of the same cell type, to define glomerular volumes by creating an α-shape. The volumes of these α-shapes can then be calculated, and correlated with the number of presynapses the same neurons make in two higher-order brain regions, the lateral horn and the mushroom body calyx. **c** Volumes can be used to analyze skeleton data. In Figure 3c we look again at olfactory projection neurons, this time from an adult fly EM dataset (Zheng et al., 2018) and use the nat function `pointsinside` with standard neuropil volumes (Ito et al., 2014) to find the numbers of presynapses GABAergic and cholinergic olfactory projection neurons from the antennal lobe make in different neuropils. These neuropils exist as an `hxsurf` object in our R package nat.flybrains.

(Kohl et al., 2013; Masse et al., 2012; Jeanne et al., 2018; Kunst et al., 2019; Zheng et al., 2018; Frechter et al., 2019). NBLAST is included in the natverse in our nat.nblast package (Costa et al., 2016). In many parts of mammalian nervous systems, morphologically similar neurons are repeated in space, and so aligning neurons to one another, without a specified template space, is sufficient for quantitative comparison (Figure 4a). NBLAST scores can be hierarchically clustered in R, plotted as a dendrogram, and used to visualize morphological groups at a defined group number or cut height (Figure 4a). Often, this forms a good starting point for cell typing, but might not be in exact agreement with manually defined cell types (Figure 4b). This can be due to neuron reconstructions being differently severed by the field of view or size of the tissue sample collected (Helmstaedter et al., 2013), or due to registration offsets between registered neuronal skeletons (Kunst et al., 2019; Chiang et al., 2011). The natverse includes interactive functions, such as `nlscan`, that allow users to visually scan neurons and identify mis-assignments (Figure 4c), or `find.neuron` and `find.soma`, that allow users to select a neuron of interest from a 3D window (Figure 4c).

In smaller brains, like theinsect or larval zebrafish central brains, the overlap of both axons and dendrites in 3D space is an excellent starting point for defining a neuronal type, since neurite apposition is suggestive of synaptic connectivity (Rees et al., 2017) and they are highly stereotyped (Pascual et al., 2004; Jenett et al., 2012). If they have been registered to whole brain templates (Costa et al., 2016; Kunst et al., 2019; Chiang et al., 2011) it is desirable to choose a canonical brain hemisphere and standardise such that all neurons are mapped onto this side to approximately double the neurons available for clustering and assign the same cell types on both hemispheres (Supplementary Figure 5 and Supplementary Figure 6).

2.3 Comparing disparate datasets

2.3.1 Template brains in *D. melanogaster*

It is also highly desirable to compare neurons between different datasets within a single template space. Considering just the case of *D. melanogaster*, separate template brains 'contain' many large and useful but disparate datasets (see Table 1 on page 9): ~23,000 single light-level neuronal morphologies, hundreds of neuronal tracings from dye fills, a collection of ~11,000 genetic driver lines, ~100 neuroblast clones, and connectomic data, including a brainwide draft connectome on the horizon (Zheng et al., 2018; Scheffer and Meinertzhagen, 2019). Because of the wealth of data available for *D. melanogaster*, we focus on its brain for our registration examples.

Two approaches have been taken in specifying template spaces: a) choosing a single brain avoids any potential artifacts generated by the averaging procedure, but b) an average brain can reduce the impact of biological variation across individuals and deformations introduced during sample preparation, thus increasing the likelihood of successful and accurate registration (Bogovic et al., 2018). Quantitative neuroanatomical work requires images to be spatially calibrated, but such calibrations are not present in all template brains.

Table 2 on page 10 lists the template brains for *D. melanogaster* considered in this work and details the resources available for each; some are shown in Figure 6. Unregistered raw data, along with two template brains (one for each sex) are publicly available for FlyCircuit (Chiang et al., 2011). The FlyLight project also provides only raw image data. Template brains and registered data are publicly available for the Vienna Tiles GAL4 libraries (Tirian and Dickson, 2017), but are not distributed in bulk form. We created an intersex reference brain for the FlyCircuit dataset and added spatial calibrations and re-registered data to our new template brains as necessary (see Materials and methods) before constructing bridging registrations. We have deposited all template brain images, in NRRD format (http://teem.sourceforge.net/nrrd/) at http://zenodo.org to ensure long-term availability.

2.3.2 Mirroring data in *D. melanogaster*

Whilst the Drosophila brain is highly symmetric it is not perfectly so and the physical handling of brains during sample preparation introduces further non-biological asymmetries. A simple 180° flip about the medio-lateral axis is therefore insufficient. To counter this, we have constructed non-rigid warping registrations for a number of template spaces that apply the small displacements required to fix the mapping from one hemisphere to the other (Supplementary Figure 7, see Methods).

Our mirroring registrations can be applied using the function `mirror_brain`. They can be used to counter non-biological asymmetries, allowing the investigation of relevant similarities and differences in morphology between the two sides of the brain (Supplementary Figure 7a). We found that the mirroring procedure (see Materials and methods) does not introduce any systematic errors into neuron morphology. NBLAST was used to calculate morphologically-determined similarity scores between DL2d olfactory projection neurons taken from the same side of the brain and compare them with those calculated between DL2d projection neurons taken from alternate sides of the brain (Figure 5b). We do not find the distributions of scores (Figure 5c) to be significantly different (D=0.025, p=0.094, two-sample Kolmogorov-Smirnov test). Extending this, we have used these scores to classify neurons based on their bilateral symmetry. Figure 5d shows 12 example neurons, taken from the bilateral subset of the FlyCircuit dataset, spanning the range of similarity scores from most asymmetric **A** to most bilaterally symmetric **L**. Interestingly, the distribution of scores suggest that most bilateral neurons are reasonably symmetric.

It is also possible to use our mirroring registrations to test the degree of symmetry for sections of neurons. We take segments of a neuron and use our similarity metric to compute a score between the segment and the corresponding segment in the mirrored version of the neuron. This allows differences in innervation and axonal path between the two hemispheres to be clearly visualised (Figure 5e).

2.3.3 Bridging template spaces in *D. melanogaster*

Simply rescaling a sample image to match a reference brain usually fails due to differences in position and rotation (Sup-

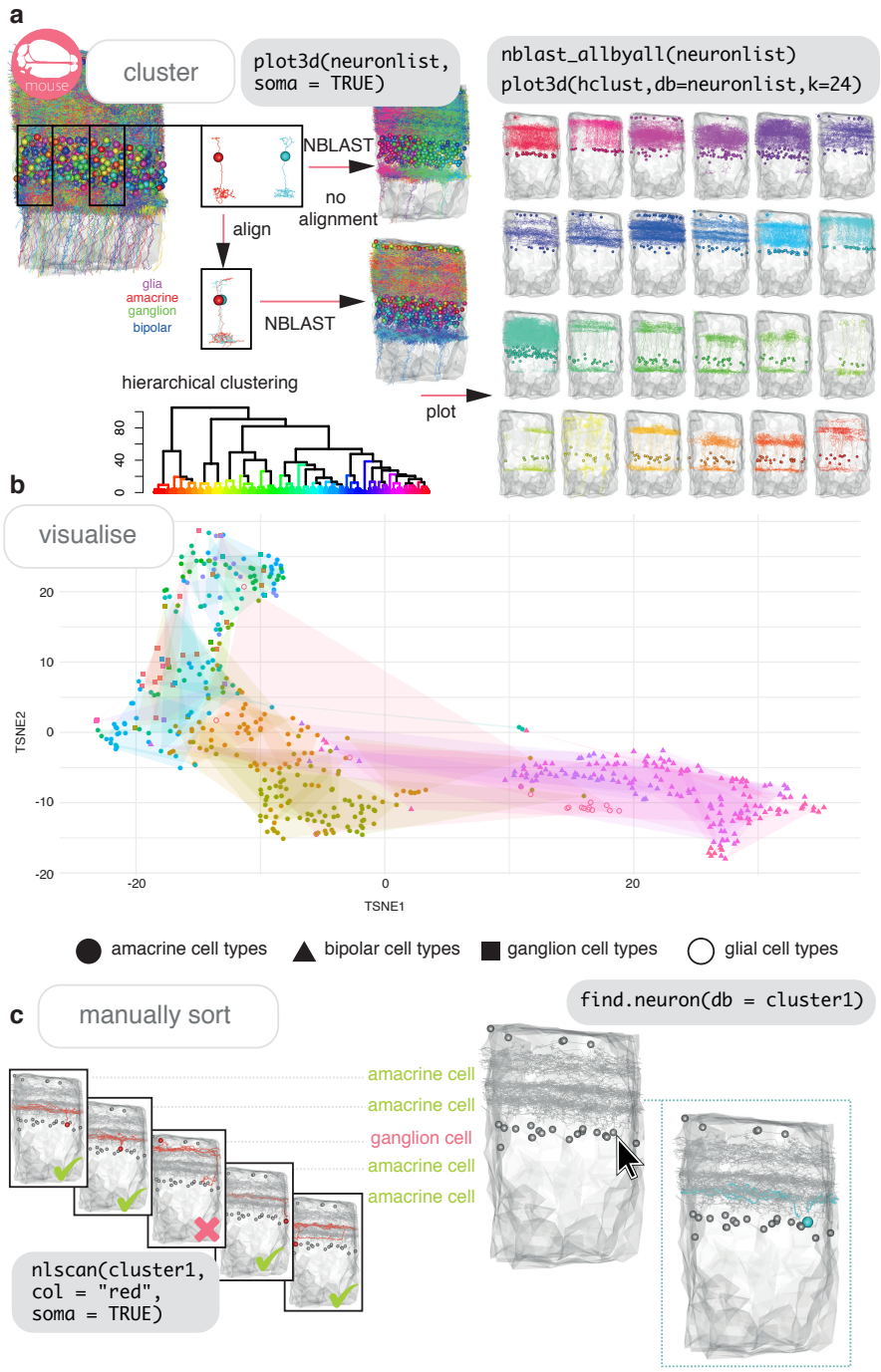


Figure 4: Cell typing with nat. **a** Neurons from a dense reconstruction from EM data of the mouse inner plexiform layer (Helmstaedter et al., 2013) can either be NBLAST-ed in situ (upper) or after alignment by their principal axes in 3D space (lower) in order to make a first pass at defining, finding or discovering morphological neuronal cell types using NBLAST. **b** A tSNE plot visualising the results of an aligned NBLAST of neurons in A, coloured by the manually annotated cells types seen in Figure 2c, with shapes indicating the cell class. **c** Manual sorting using the interactive nat functions `nlscan` and `find.soma` or `find.neuron` can help identify misclassifications and make assignments.

Dataset	Type	Count	References
Flycircuit	Single neuron morphologies stochastically labeled from dense transmitter-related lines	~23,000 neurons	Shih et al. (2015); Chiang et al. (2011)
FlyLight GMR collection	Collection of genetic driver lines, drive by orthogonal transcription factors GAL4 (Brand and Perrimon, 1993) or LexA (Lai and Lee, 2006)	~3500 GAL4 lines ~1500 LexA lines	Pfeiffer et al. (2008); Jenett et al. (2012)
Vienna Tiles collection	Collection of genetic driver lines, driven by orthogonal transcription factors GAL4 or LexA	~8000 GAL4 lines ~3000 LexA lines	Kvon et al. (2014); Tirian and Dickson (2017)
FlyLight split-GAL4 collection	Genetic driver lines labelling small constellations of neurons using the split-GAL4 system	~400 sparse lines covering the mushroom body, lobula plate and column, visual projection neurons, ellipsoid body, descending neurons , central complex, olfactory projection neurons (Y. Aso, personal communication, 2019) and lateral horn	Aso et al. (2014); Aso and Rubin (2016); Dolan et al. (2019); Wu et al. (2016); Robie et al. (2017); Klapoetke et al. (2017); Namiki et al. (2018); Wolff and Rubin (2018)
K. Ito, T. Lee and V. Hartenstein	Neuroblast clones for the central brain & larval-born neurons, generated using the MARCM method (Lee and Luo, 2001)	~100 neuroblast clones	Yu et al. (2013); Ito et al. (2013); Wong et al. (2013)
FlyEM and Harvard Medical School	Volume-restricted connectomes	Hundreds of neurons from the mushroom body alpha lobe, two antennal lobe glomeruli and several columns of the optic medulla	Takemura et al. (2013, 2015, 2017); Tobin et al. (2017); Horne et al. (2018)
FAFB project	Serial section transmission electron microscopy data for a single, whole adult female fly brain (Zheng et al., 2018), that has a partial automatic segmentation available (Li et al., 2019)	Raw image data for ~150,000 neurons of which several hundred have been partially reconstructed for recent publications, 7 thousand more unpublished; an estimated ~5 % of neurons have some level of reconstruction	Sayin et al. (2019); Felsenberg et al. (2018); Frechter et al. (2019); Dolan et al. (2018); Huoviala et al. (2018); Zheng et al. (2018); Dolan et al. (2019)
Various laboratories	Single neuron morphologies extracted from dye-filling (e.g. with biocytin) neurons	Hundreds across a range of studies, some cited here	Jeanne and Wilson (2015); Frechter et al. (2019); Grosjean et al. (2011); Jefferis et al. (2007)

Table 1: Neuron morphology resources currently available for the adult *D. melanogaster* brain.

Template	Description	Resources	DOI	Citation
Wuerzburg	Single nc82-stained female brain	Wild-type CantonS nc82 stained reference brain	—	Rein et al. (2002)
Cell07	Partial intersex nc82-stained averaged brain (14 f, 2 m)	~240 lateral horn projection neuron tracings	10.5281/zenodo.10570	Jefferis et al. (2007)
T1	Intersex nc82-stained averaged brain	The Vienna Tiles collection	10.5281/zenodo.10590	Yu et al. (2010)
IS2	Intersex nc82-stained averaged brain	1018 3D confocal images of fruitless neurons	10.5281/zenodo.10595	Cachero et al. (2010)
FCWB	Intersex Dlg-stained averaged brain (17 f, 9 m)	Good for FlyCircuit data, ~16,000 neurons re-registered	10.5281/zenodo.10568	Costa et al. (2016)
JFRC	Single nc82-stained female brain	The FlyLight collection	—	Jenett et al. (2012)
JFRC2	Spatially calibrated copy of JFRC	The FlyLight collection	10.5281/zenodo.10567	This study
IBN	Tri-labelled half brain, with n-syb-GFP	Neuropil and tract segmentations (half-brain)	—	Ito et al. (2014)
IBNWB	Synthetic whole-brain version of IBN	Neuropil and tract segmentations (whole-brain)	10.5281/zenodo.10569	This study
FAFB14	An aligned volume for a single whole female fly brain from EM data	Thousands of single neuron partial manual reconstructions and fragmented automatic segmentation Li et al. (2019)	—	Zheng et al. (2018)
JRC2018F	A symmetrised high-quality template from FAFB extracted synapse data (Heinrich et al., 2018)	—	—	Bogovic et al. (2018)

Table 2: *D. melanogaster* template brains.

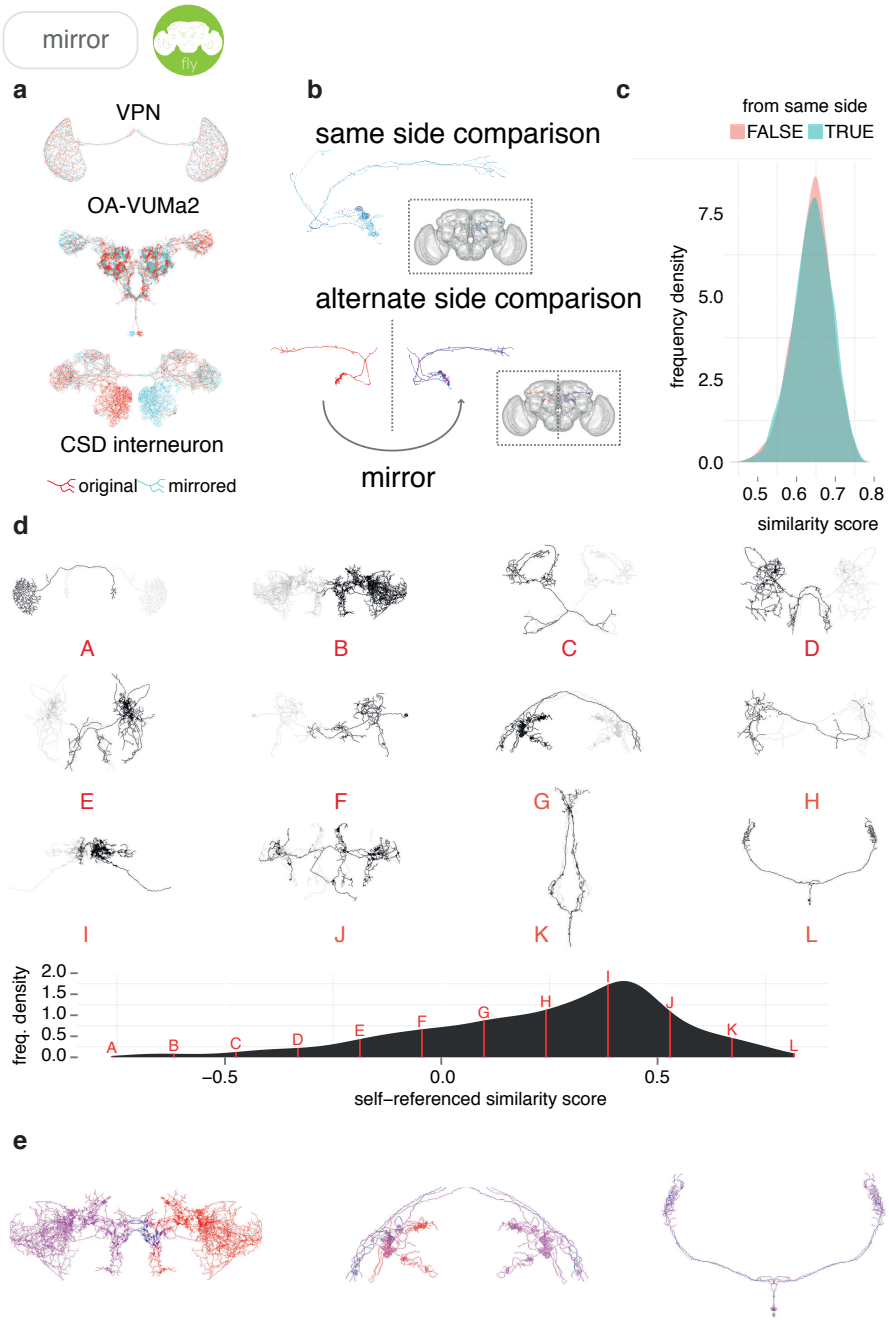


Figure 5: Sample applications of mirroring registrations. **a** Three FlyCircuit neurons along with mirrored versions; a visual projection neuron, OA-VUMa2 (Busch et al., 2009) and the CSD interneuron (Dacks et al., 2006). Co-visualisation facilitates the detection of differences in innervation, such as the higher density of innervation for the CSD interneuron in the lateral horn on the contralateral side compared to the ipsilateral lateral horn. **b** Neurons from the same side of the brain and alternate side of brain are compared and a similarity score generated. **c** Distributions of similarity scores for comparisons within the same brain hemisphere and across brain hemispheres. **d** Sequence of 12 example neurons (black) with mirrored counterparts (grey), having equally spaced similarity scores. Below, full distribution of scores for all neurons in FlyCircuit dataset. **e** Segment-by-segment measures of neuron similarity. Redder shades correspond to low degrees of symmetricity, bluer shades higher. Flipped version of neuron in gray.

plementary Figure 8a). An affine transformation can account for these differences, but not for differences in shape that may be of biological or experimental origin. To correct for these, we use a full non-rigid warping deformation, as described previously (Rueckert et al., 1999; Rohlfing and Maurer, 2003; Jefferis et al., 2007) (also see Materials and methods). Briefly, a regular lattice of control points is created in the reference brain and corresponding control points in the sample brain are moved around to specify the deformations required to take the sample data into the reference space (Figure 6c-g). Deformations between control points are interpolated using B-splines, which define a smooth deformation of sample to reference (Figure 6f). The use of a mutual information metric based on image intensity avoids the requirement for landmarks to be added to each image a time-consuming task that can often introduce significant inaccuracies. Our approach allows for the unsupervised registration of images and the independent nature of each registration allows the process to be parallelized across CPU cores. By utilizing a high-performance computational cluster, we re-registered, with high accuracy, the entire FlyCircuit dataset within a day.

Our bridging registrations can be deployed on any 3D natverse-compatible data (e.g. points, neurons, surfaces, images) using the function `xform_brain`. A successful and accurate bridging registration will result in two template spaces being well co-localized (Figure 6). After visually inspecting co-localized template spaces to check for any obvious defects, we find it helpful to map a standard neuropil segmentation (Ito et al., 2014) into the space of the new brain to check for more subtle defects (Figure 6d). If the registration passes these checks it can then be used to combine data from multiple datasets.

The creation of a bridge between a GAL4 expression library, such as the GMR collection (Jenett et al., 2012), and images of single neurons, such as those of FlyCircuit (Chiang et al., 2011), facilitates the decomposition of an expression pattern into its constituent neurons, allowing the correct assessment of innervation density on, for example, ipsilateral and contralateral sides (Figure 6e). Similarly, correspondences between neuroblast clones can be identified with co-visualisation. We bridge Fru+ clones (Cachero et al., 2010) from IS2 space into the JFRC2 space of elav clones (Ito et al., 2013) and hence determine subset relations (Supplementary Figure 9b). Furthermore, we can bridge the single neuron FlyCircuit data (Chiang et al., 2011) from the FCWB space into the IS2 space of the Fru+ clones and use the known sexual dimorphisms of Fru clones to predict which neurons may be sexually dimorphic (Supplementary Figure 9c).

The ability to bridge segmentations from one space to another is useful for checking innervation across datasets. While FlyCircuit single neurons (Chiang et al., 2011) were provided along with information on innervation density based on their own neuropil segmentation, this segmentation is not the same as the canonical one (Ito et al., 2014). We have bridged the latter segmentation into FCWB space and recalculated innervation for all the FlyCircuit neurons, providing a more standardized measure (Figure 6f). Further, we can compare neurons from Flycircuit with those for which we have electrophysiological data (Kohl et al., 2013; Frechter et al., 2019), enabling us to suggest a functional role for unrecorded neurons based on their morphological similarity to recorded neurons (Figure 6g).

Both the FlyLight (Jenett et al., 2012) and Vienna Tiles libraries (Tirian and Dickson, 2017) contain a wealth of GAL4 lines amenable to intersectional strategies (Luan et al., 2006). However, as the two libraries are registered to different template spaces, it is difficult to predict which combinations of a FlyLight GMR line with a Vienna Tiles line would produce a good intersection (split-GAL4, targeting one cell type present in both parent lines) from the raw images provided by both. Bridging one library into the space of another (Figure 6h) enables direct co-visualisation. This could be used manually or computationally to identify combinations that could potentially yield useful intersectional expression patterns (Venken et al., 2011).

It is also possible to warp 3D neuropils and neuron skeletons onto some target, without using landmark pairs. For this, Deformetrica (Bône et al., 2018; Durrleman et al., 2014) can be used to compute many pairwise registrations at once for different kinds of 3D objects to produce a single deformation of ambient 3D space describing a registration (Supplementary Figure 10). This is a generic method that does not require landmark correspondences to be manually assigned. We give a simple example in Supplementary Figure 10a, symmetrizing a wonky brain and making a LM-EM bridge for first-instar larva, for which there is a nascent connectome (Berck et al., 2016; Eichler et al., 2017; Ohyama et al., 2015; Schneider-Mizell et al., 2016). With such a method it should be possible to bridge EM or LM data between developmental stages for a nervous system to make comparisons or identify neurons.

2.3.4 EM to LM and back again

Finding neurons of the same cell type between a high-resolution EM dataset and light-level images of neurons (Figure 7a) is an essential step in identifying neurons and their genetic resources. So doing links connectivity and detailed morphological information acquired at the nanometer resolution to other forms of data. This can most easily be done by finding corresponding landmarks in EM data and a LM template space to build a registration (Supplementary Figure 8).

In Figure 7 and Figure 8 we give the general pipeline we used in recent publications (Dolan et al., 2018, 2019; Frechter et al., 2019; Li et al., 2019) to connect neurons sparsely labeled in a split-GAL4 line (registered to the template space JFRC2) to sparsely reconstructed neurons from an EM dataset (FAFB14). Neurons can be manually reconstructed (Schneider-Mizell et al., 2016) or, more recently, partially reconstructed by machine learning methods (Januszewski et al., 2018) as segments that can be manually concatenated (Li et al., 2019). A thin plate spline bridging registration between JFRC2 and FAFB14 can be built by finding ~100 corresponding landmarks between the two brainspaces e.g. the location of turns in significant tracts, the boundaries of neuropils, the location of easily identifiable single neurons. This registration can be deployed using `xform_brain` and our `elmr` package.

By bridging multiple other light-level datasets into JFRC2 (Figure 6), candidate neurons from the EM brainspace can be co-visualised (Figure 8c) and NBLAST-ed against light-level datasets in order to confirm their cell type identity and

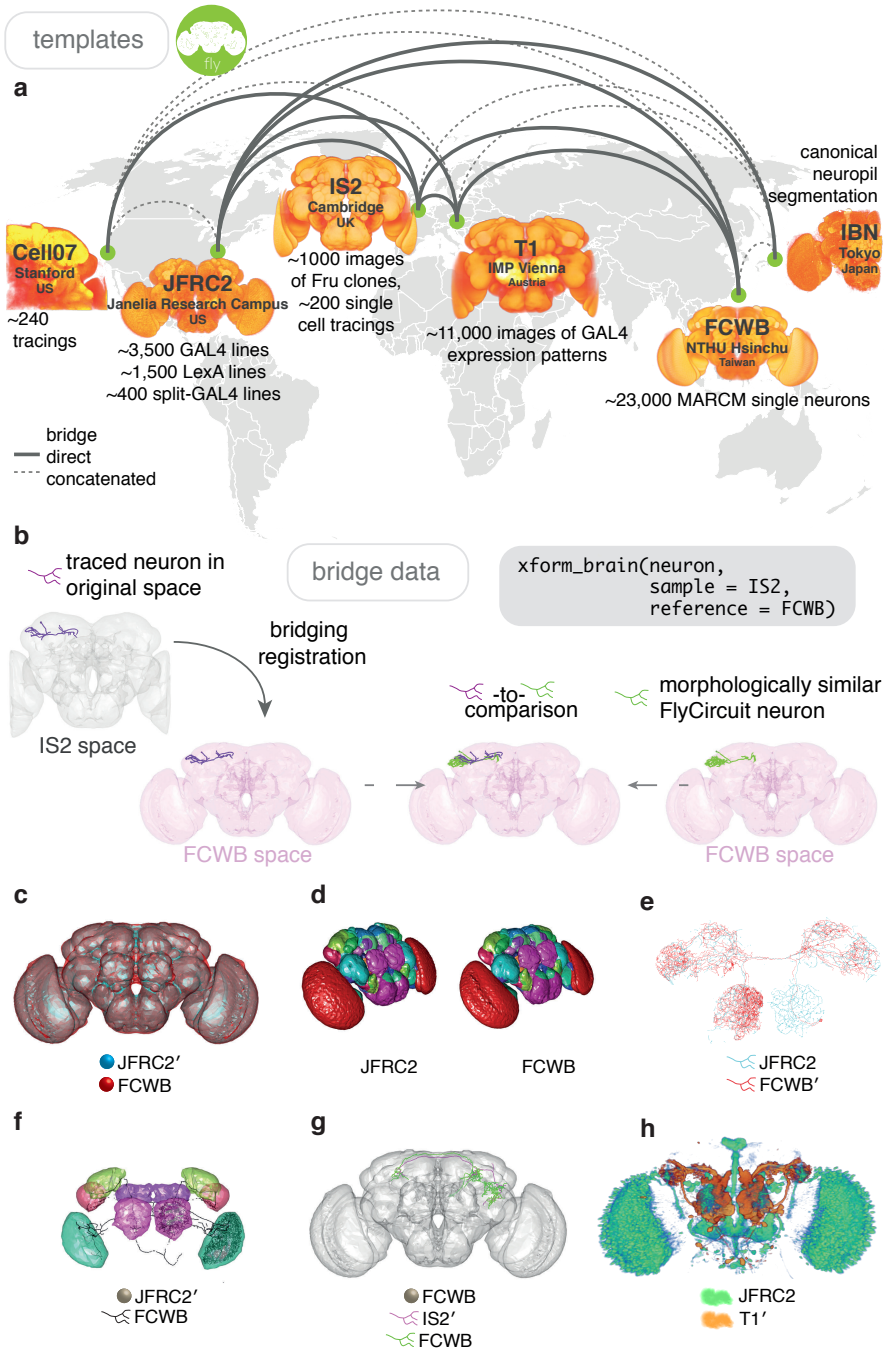


Figure 6: Bridging registrations for brain templates. **a** A small sample of Drosophila template brains used around the world are shown. **b** A partial neuron tracing (purple) made using Simple Neurite Tracer (Longair et al., 2011) being transformed (`xform_brain`) from the IS2 standard brainspace to FCWB, where it can be co-visualised with a more complete, matching morphology from FlyCircuit, using our R package flycircuit. **c** Outermost neuropil boundaries for FlyCircuit (red) and FlyLight (cyan) template brains. Primed space names indicate original spaces for data. Unprimed space names indicate current space of image. **d** Neuropil segmentation from JFRC2 space alongside FCWB reformatted version. **e** CSD interneuron from FlyCircuit (red) and FlyLight (cyan). **f** Neuropil segmentation from JFRC2 (Ito et al., 2014) space that has been bridged into FCWB space, so it can be seen along with selected neurons from FlyCircuit. **g** A traced neuron in FCWB space alongside morphologically similar neuron from FlyCircuit. **h** Expression pattern of Vienna Tiles line superimposed on expression pattern of FlyLight line.

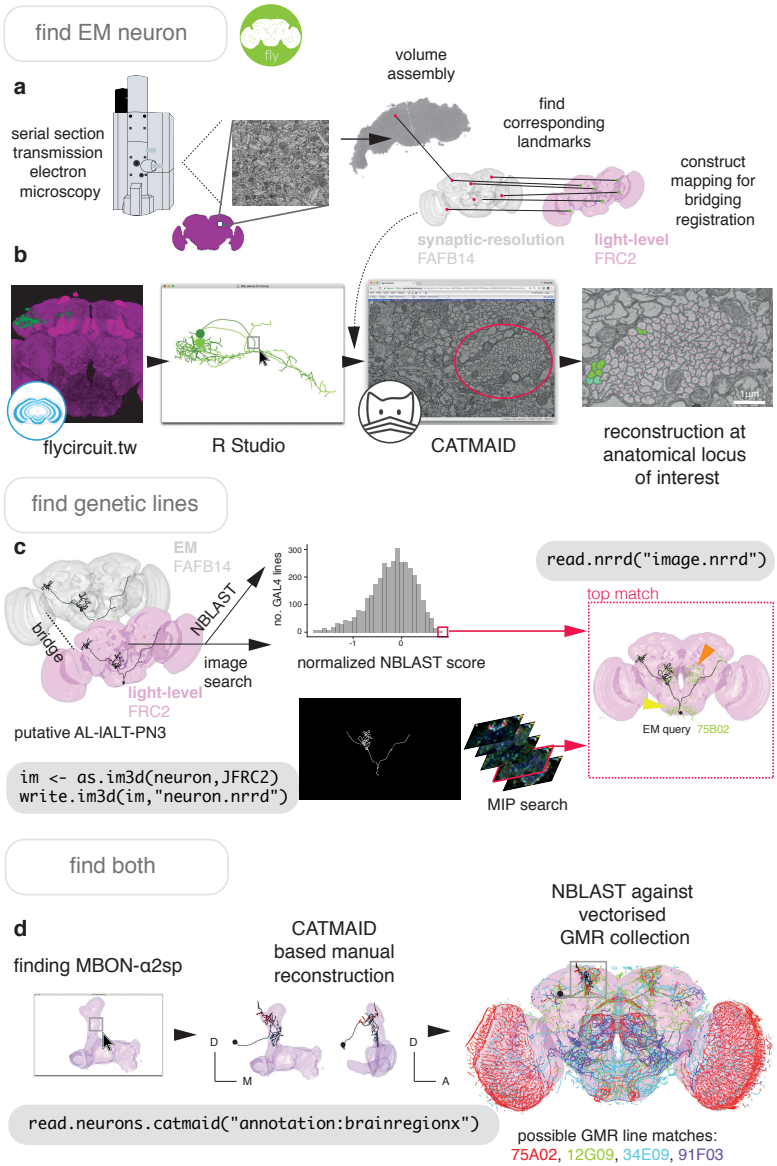


Figure 7: Finding specific neurons in EM and LM data. **a** Pipeline for acquiring EM neuron data. Serial section transmission EM at high speed with a TEM camera array (Bock et al., 2011) produced several micrographs per section at 4 × 4 nm resolution, ~40 nm thick. These were, per section, stitched into mosaics which were, across sections, registered to create the female adult fly brain v.14 template space (FAFB14, grey) (Zheng et al., 2018). Corresponding landmarks between FAFB14 and JFRC2 can be found and used to build a bridge. **b** The R package elmr can be used to select an anatomical locus, here the PD2 primary neurite tract (Frechter et al., 2019), from 3D plotted light-level neurons, taken from FlyCircuit, and generate a URL that specifies its correct coordinates in a FAFB14 CATMAID instance. Candidates (185) may then be coarsely traced out until they deviate from the expected light-level morphologies (178 pink dotted profiles, often a few minutes to an hour of manual reconstruction time to rule out neurons of dissimilar cell types sharing a given tract, similar cell types are more subtly different and might need to be near completely reconstructed). Those that remain largely consistent were fully reconstructed (green profiles, ~7-12 person-hours per neuron) (Li et al., 2019). **c** Close matches reveal likely morphology of non-reconstructed branches (orange arrow) but also contain off-target expression (yellow arrow). Identification of multiple candidate lines enables split-GAL4 line generation aimed at retaining common neurons in two GAL4 patterns. MultiColor FlpOut of resultant splits can be compared with the EM morphology. Here, a candidate GAL4 line is found for AL-lALT-PN3 (Frechter et al., 2019; Tanaka et al., 2012) using NBLAST and a MIP search (Otsuna et al., 2018). **d** A recent dense, but volume-restricted reconstruction of the mushroom body α-lobe discovered a 'new' mushroom body output neuron type (MBON-α2sp) (Takemura et al., 2017). By bridging from the correct mushroom-body compartment using a mushroom body mesh (Ito et al., 2014) to the FAFB14 EM data's equivalent space in CATMAID using the R package elmr, an experienced tracer can easily identify dendrites and find MBON-α2sp. By doing so, we found its previously unreported axon-morphology. We then imported the skeleton into R, bridged MBON-α2sp into the JFRC2 template space where it could be NBLAST-ed against GMR GAL4 lines to identify candidate lines containing the MBON.

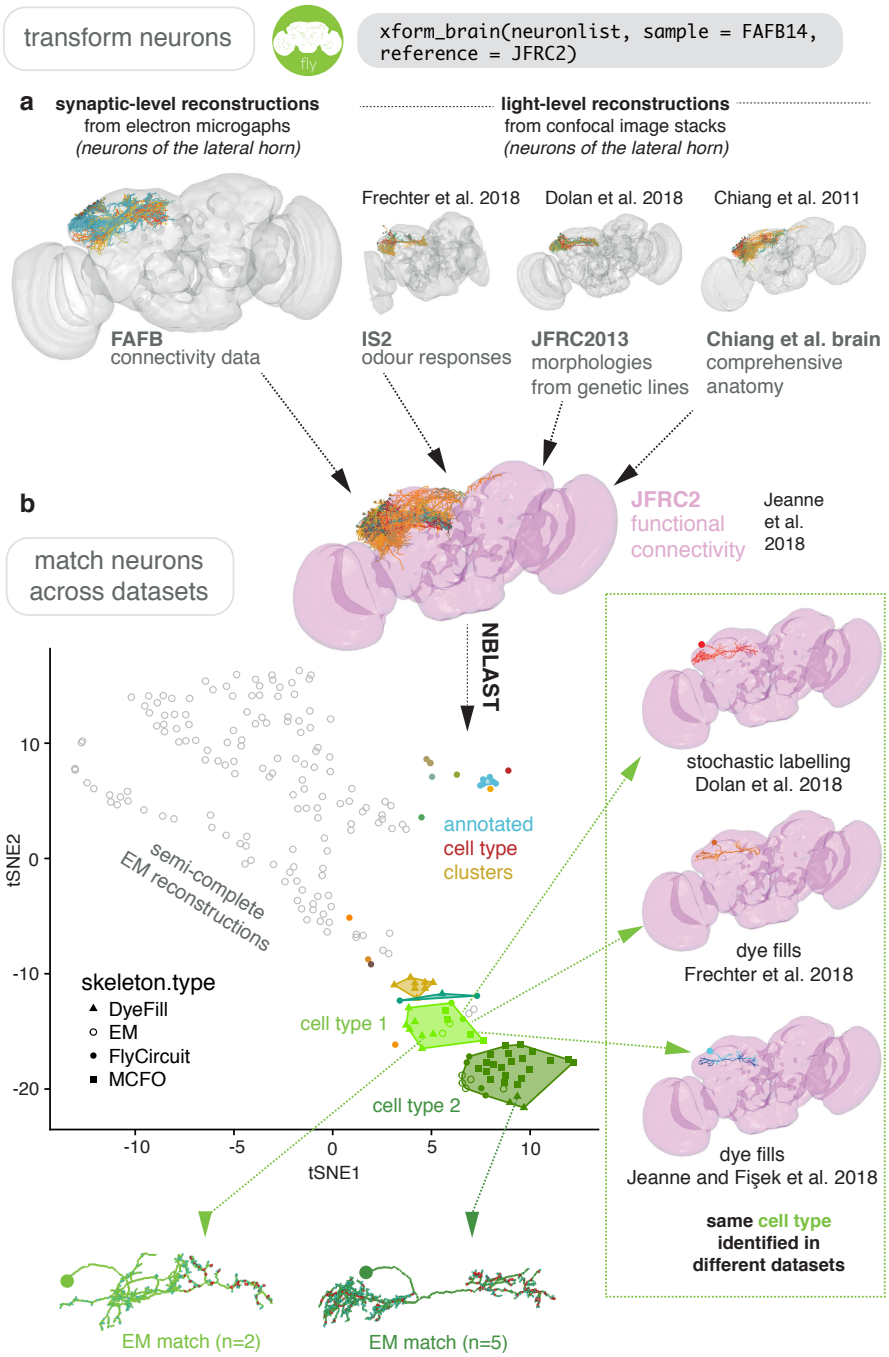


Figure 8: Bridging EM and LM data. **a** Sparse EM reconstruction providing a database of non-comprehensive, partial morphologies that can be searched using NBLAST. Candidate neurons from the EM brainspace can be NBLAST-ed against MCFO data and other light-level datasets in order to connect them to cell-type specific information, such as odour responses and functional connectivity (Chiang et al., 2011; Dolan et al., 2019; Frechter et al., 2019; Jeanne et al., 2018), by bridging these datasets into the same brainspace. **b** An all-by-all NBLAST of all neurons in the PD2 primary neurite cluster (Frechter et al., 2019) in multiple datasets can be shown as a tSNE plot. EM cell type matches can easily be found, as well as other correspondences between the light level datasets.

consider results from different studies (Jeanne et al., 2018; Frechter et al., 2019; Chiang et al., 2011; Dolan et al., 2019) (Figure 8d). However, FAFB14 contains unannotated image data for ~150,000 neurons (Bates et al., 2019), each requiring hours of manual reconstruction time, and person-power is limited. To find specific neurons in this volume, we can use the R package elmr to select a distinctive anatomical locus, for example the cell body fiber tract (Frechter et al., 2019) from 3D plotted neurons, and jump to its approximate coordinates in FAFB14 in a supported CATMAID instance using the generated URL (Figure 7b). Reconstruction efforts can then be focused at this location, being aware that the jump is not always completely accurate despite a good bridging registration as some light-level datasets can be ill-registered (Figure 7b). In the absence of an extant light-level reconstruction, candidate neurons can be found by identifying distinctive anatomical loci in the EM volume that correspond to the anatomy of the cell type in question (Figure 7d).

A user may also want to work the opposite way, and connect an interesting EM reconstruction to light-level data, e.g. to identify a genetic resource that targets that neuron. In this situation, a similar pipeline can be used. For *D. melanogaster*, a reconstruction can be bridged into JFRC2 and NBLAST-ed against GAL4 lines (Jenett et al., 2012; Tirian and Dickson, 2017) read from image data and represented as vector clouds (Costa et al., 2016). Alternatively, image matching tools can be used, such as the recent colour depth MIP mask search (Otsuna et al., 2018), which operates as an ImageJ plug-in (Figure 7c).

Further, because close light-level matches for in-progress EM reconstructions reveal the likely morphology of non-reconstructed branches (Figure 7c) this process can help human annotators reconstruct neurons accurately and in a targeted manner, which may be desirable given how time intensive the task is. In order to reduce this burden further we make use of a recent attempt to automatically segment neurites in FAFB14 using a flood filling approach (Li et al., 2019), which produces volumetric fragments of neurites, where segments may be fairly large, ~100 μm in cable length.

Our fafbseg package includes functions to implement improved up-/downstream sampling of neurons based on these segments, which we have recently discussed elsewhere (Li et al., 2019). We can also generate volumetric reconstructions of manually traced neurons by mapping them onto volumetric data (Supplementary Figure 11b), hosted by a brainmaps server and visible through a Neuroglancer instance (Supplementary Figure 11a). Currently, ~500 such segments will map onto one accurately manually traced neuron but only ~20 segments may constitute the highest Strahler order branches meaning that manual concatenation of these fragments speeds up discovery of coarse morphologies by ~10× (Li et al., 2019). These fragments can be used to identify the neuron in question by NBLAST-ing against light-level data. Twigs and small-calibre, lower Strahler order branches are more difficult to automatically segment (Supplementary Figure 11c). Nevertheless, matching tracings to segmentations allows us to estimate the volume of neurons that we have previously manually reconstructed (Dolan et al., 2019, 2018) by only tracing the neurites' midline (i.e. skeletonisation). We can therefore observe that superior brain neurons' axons are slightly thicker than their dendrites and that neuronal volumes correlates strongly with its total cable length (Supplementary Figure 11d).

3 Discussion

The shape of a neuron is of major functional significance. Morphology is driven by and constrains connectivity. It is also the primary means by which neuroscientists have historically identified neuron classes. There have been three main drivers behind the recent emphasis on quantitative neuroanatomy: a) the ever increasing scale of new approaches for acquiring image data and reconstructing neurons, b) a drive to formalise descriptions of the spatial properties of neurons and networks at various scales, and c) a desire to intuit the organisational principles behind different nervous tissues and correlate these findings with dynamic data on neuron activity.

With the natverse, a suite of R packages for neuroanatomy with well documented code and detailed installation and tutorials available online, we aim to expedite analysis of these data in a flexible programming environment. The natverse allows a user to read data from local or remote sources into R, and leverage both natverse functions and the >10,000 R packages on CRAN (and more on Bioconductor, Neuroconductor, GitHub, etc.) to aid their data analysis. Users may also call natverse R functions from other languages such as Python, Julia and MATLAB. We have provided detailed examples to analyse skeleton and volume data from various sources and have made the code available with both R and Python examples, at https://github.com/jefferislab/nat.examples. These examples demonstrate how to obtain skeleton and volume data, calculate basic metrics for neurons, examine synapses and other tagged biological features like microtubules, analyse morphology as a graph or through Strahler order and NBLAST searches, prune neurons, semi-manually cell type neurons, spatially transform neurons and create subvolumes using neurons. We have also given an example of building a more complex analysis, based on natverse tools but making use of other available R packages.

We hope that the natverse becomes a collaborative platform for which users can contribute to existing R packages or link to their own. We note that the natverse is an actively developing project and also anticipate a) an increasing interest in dealing with neurons as volumes as automatic segmentation of datasets becomes commonplace, b) the need to expand on our bridging tools for other species and between similar species and developmental stages, c) writing libraries to facilitate the use of the natverse in other programming languages and toolboxes besides Python, and d) expanding the range of neurogeometric analysis algorithms readily available in the natverse.

In addition to general purpose natverse tools, we have generated some specific R packages to support ongoing projects in the *D. melanogaster* brain. We have constructed high quality registrations for the bridging of data from one template space to another, along with registrations for mirroring data across brain hemispheres. In two of the largest cases, only raw unregistered data were available, so we began by registration to an appropriate template space. This has allowed us to deposit ~20,000 co-registered images from different sources in the virtualflybrain.org project. Averaged intersex template

spaces can form high quality registration templates for both sexes and we recommend the use of averaged brains to reduce the effects of sample-to-sample variation. We propose using a small number of template spaces, particularly those that are already associated with the most data (JFRC2) or of highest quality (Bogovic et al., 2018), as a hub. High quality bridging registrations would be created between new template spaces and brains in the hub, ensuring that any template could be bridged to any other via appropriate concatenations and inversions of these registrations.

Using these resources, it is now possible to co-visualise and analyze more than 23,000 single neuron images (Chiang et al., 2011), expression patterns of >9,500 GAL4 lines (Kvon et al., 2014; Jenett et al., 2012; Tirian and Dickson, 2017) and a near complete set of ~100 adult neuroblast clone lineage data (Ito et al., 2013; Yu et al., 2013) and easily combine these data with the standard insect brain name nomenclature system (Ito et al., 2014). For example we have calculated the neuropil overlap between single neurons in the FlyCircuit data, which we have deposited with virtualflybrain.org so they can be queried online. It will soon be possible to identify split-GAL4 lines, a synaptic EM reconstruction and the developmental clone of origin for any given neuron or neuronal cell type for *D. melanogaster*. We anticipate such mappings to become publicly available and easy to use via resources such as https://v2.virtualflybrain.org/.

The near future will see generation of EM data for multiple whole adult Dipteran brains and larval zebrafish, possibly from different sexes and species, as well as quality automatic segmentations for such data's neurites (Januszewski et al., 2018; Funke et al., 2019) and synapses (Heinrich et al., 2018), even from anisotropic serial section transmission EM data (Li et al., 2019).Interpreting high resolution EM connectomic data will be sped up and enriched by making links to light level data (Schlegel et al., 2017), and it is further possible that connectomes and transcriptomes may be linked on a cell type basis, using neuron morphology as a bridge (Bates et al., 2019). The natverse provides extensible functionality for easily combining and analysing all these data.

4 Materials and methods

4.1 R packages for Neuroanatomy

The R programming language (Development Core Team, 2011) is perhaps the premier environment for statistical data analysis, is well supported by the integrated development environment RStudio and is a strong choice for data visualisation (Wickham, 2016). It already hosts a wealth of packages for general morphometric and graph theoretic analysis (Schlager, 2017; Lafarge et al., 2014; Csardi and Nepusz, 2005; Duong, 2007). An R package is a bundle of functions, documentation, data, tests and example code (Wickham, 2015). R packages are discrete, standardized and highly shareable units of code. They are primarily installed either from the Comprehensive R Archive Network (CRAN, >14,000 packages, curated), Bioconductor (>1,700 packages, curated) or GitHub (larger, uncurated), using just one or two function calls and an Internet connection. The R packages behind the natverse can be divided into four groups (Figure 1a):

4.1.1 Working with synaptic resolution data in nat

Group a) obtains synaptic-level data required for connectomes and includes catmaid, neuprintr, drvid and fafbseg. The package catmaid provides application programming interface (API) access to the CATMAID web image annotation tool (Schneider-Mizell et al., 2016; Saalfeld et al., 2009). CATMAID is a common choice for communities using terabyte-scale EM data to manually reconstruct neuron morphologies and annotate synaptic locations (Ohyama et al., 2015; Berck et al., 2016; Dolan et al., 2018; Frechter et al., 2019; Eichler et al., 2017; Zheng et al., 2018). Users can use catmaid to read CATMAID neurons into R including the locations and associations of their synapses, and other tags that might identify biological entities such as somata, microtubules or gap junctions. Users can also leverage CATMAID's infrastructure of flexible hierarchical semantic annotations to make queries for neurons e.g. in a brain region of interest. Further catmaid can edit CATMAID databases, e.g. adding annotations, uploading neurons, synapses and meshes. Some CATMAID instances are kept private by a community before data publication. In this case catmaid can enable a user to send authenticated requests to a CATMAID server, i.e. data can be kept private but still be read into R over an Internet connection. The packages neuprintr and drvid are very similar, except that they interact with API endpoints for different distributed annotation tools, the NeuPrint connectome analysis service (https://github.com/connectome-neuprint/neuPrint) and DVID (Katz and Plaza, 2019) and can retrieve neurons as volumes as well as skeletons. The package fafbseg aims to make use of the results of automatic segmentation attempts for large, dense brain volumes. It includes support for working with Google's BrainMaps and NeuroGlancer (https://github.com/google/neuroglancer). Automatic segmentation of EM data is a rapidly-developing field and this package is currently in active development; at present it onlys supports auto-segmentation (Li et al., 2019) of a single female adult fly brain (FAFB) dataset (Zheng et al., 2018).

4.1.2 Working with light-resolution data projects in nat

Group b) is targeted at light microscopy and cellular resolution atlases, or mesoscale projectomes. Its packages, neuromorphr, flycircuit, vfbr, mouselight, insectbrainr and fishatlas can read from large repositories of neuron morphology data, many of which are co-registered in a standard brain space. neuromorphr provides an R client for the NeuroMorpho.org API (Ascoli et al., 2007; Halavi et al., 2008; Nanda et al., 2015), a curated inventory of reconstructed neurons (n = 107395, 60 different species) that is updated as new reconstructions are collected and published. Since its neurons derive from many different systems and species, there is no 'standard' orientation, and so they are oriented by placing the soma at the origin and aligning neurons by their principal components in Euclidean space. insectbrainr can retrieve neurons and brain region surface models from InsectBrainDB.org (n = 139 neurons, 14 species). Similarly flycircuit interacts with the flycircuit.tw project (Chiang et al., 2011; Shih et al., 2015), which contains >23,000 registered and skeletonised *D. melanogaster* neurons. The vfbr package can pull image data from virtualflybrain.org, which hosts

registered stacks of central nervous system image data for *D. melanogaster*, including image stacks for the major GAL4 genetic driver line collections (Jenett et al., 2012), neuroblast clones (Yu et al., 2013; Ito et al., 2013) and FlyCircuit's stochastically labelled neurons (Chiang et al., 2011). This non-skeleton data can be read into R as point clouds. The fishatlas package interacts with FishAtlas.neuro.mpg.de, which contains 1,709 registered neurons from the larval Danio rerio (Kunst et al., 2019), while mouselightr does the same for the MouseLight project at Janelia Research Campus (Economo et al., 2016), which has generated >1,000 morphologies. In both cases, investigators have acquired submicron single neuron reconstructions from datasets of whole brains using confocal (Kunst et al., 2019) or two-photon microscopy (Economo et al., 2016), modified tissue clearing techniques (Treweek et al., 2015), and generated a template brain with defined subvolumes.

4.1.3 Working with registrations in nat

Group c) helps users make use of registration and bridging tools. The package nat.ants wraps the R package ANTsRCore (Kandel et al., 2019) with a small number of functions to enable nat functions to use Advanced Normalisation Tools (ANTs) registrations (Avants et al., 2009). The R package deformetricar does the same for the non-image (e.g. mesh or line data) based registration software Deformetrica (Bône et al., 2018; Durrleman et al., 2014) without the need for landmark correspondences. The nat package already contains functions to support CMTK registrations (Rohlfing and Maurer, 2003). The nat.templatebrains package extends nat to explicitly include the notion of each neuron belonging to a certain template space, as well as functions to deploy bridging and mirroring registrations. Additionally, nat.flybrains contains mesh data describing commonly used template spaces for *D. melanogaster* as well as CMTK bridging and mirror deformations discussed in the latter half of the results section.

4.1.4 Analyzing data in nat

Group d) contains functions that help users to easily analyze neuron data as both skeletons and volumes. Its biggest contributor is still nat, but nat.nblast allows users to deploy the NBLAST neuron similarity algorithm (Costa et al., 2016), by pairwise comparison of vector clouds describing these neurons in R. Our nabor package is a wrapper for libnabo (S. Magnenat), a k-nearest neighbour library which is optimised for low dimensional (e.g. 3D) spaces. The package elmr is another fly focused package that has been born out of a specific use case. Currently, ~22 laboratories and ~100 active users worldwide are engaged with reconstructing *D. melanogaster* neurons from EM data (Zheng et al., 2018) using CATMAID (Saalfeld et al., 2009; Schneider-Mizell et al., 2016) in order to build a draft, sparse connectome. The package elmr allows users to read neurons from this environment, transform them into a template space where they can be compared with light-level neurons for which the community may have some other information (e.g. gene expression, functional characterisation, presence in genetic drive lines, etc.), then visualised and/or NBLAST-ed; all with only a few lines of code. This process enables CATMAID users to perform interim analyses as they reconstruct neurons, helping them to choose interesting targets for reconstruction and identify manually traced or automatically reconstructed neuron fragments (Dolan et al., 2019) or anatomical landmarks such as fiber tracts (Frechter et al., 2019), and so improve the efficiency of their targeted circuit reconstructions (Dolan et al., 2018; Huoviala et al., 2018; Felsenberg et al., 2018).

4.2 Building mirroring registrations

A simple 180° flip about the medio-lateral axis is insufficient to generate a left-right mirror for most neuroanatomical volumes; after flipping, the brain will not be perfectly centered in the image. It is first necessary to apply an affine registration to roughly match the flipped brain to the same location as the original. This results in a flipped brain with the correct gross structure (i.e. large structures such as neuropils align) but with mismatched fine details (e.g. bilaterally symmetric neurons may appear to innervate slightly different regions on either side (Figure 5a). For example, for the JFRC2 template space we found that points are, on average, displaced by 4.8 μm from their correct position, equivalent to 7–8 voxels of the original confocal image. The largest displacements, of the order of 10–15 μm, are found around the esophageal region (Supplementary Figure 7b) and are likely due to specimen handling when the gut is removed during dissection. An ideal mirroring registration would result in zero total displacement after two applications of the mirroring procedure, i.e. a point would be mapped back to exactly the same location in the original brain hemisphere. Our constructed mirroring registrations have, on average, a round-trip displacement of less than a quarter of a micron – i.e. about the diffraction limit resolution of an optical microscope and less than half of the sample spacing of the original confocal image (Supplementary Figure 7c).

4.3 Building bridging registrations

Given a bridging registration A $\mapsto$ B, an attempt to produce the registration B $\mapsto$ A can be made via numerical inversion of the original registration. This is a computationally intensive process but we find it to be useful for neuroanatomical work as the inaccuracies are set by numerical error, which is much smaller than registration error. As the registration A $\mapsto$ B may be injective (i.e. points within brain A may map to a subset of the points within brain B), there may be some points in B, particularly near the boundaries of the brain, for which this inversion will not map them into A. To counter this we have, for some brains, constructed a new registration B $\mapsto$ A by explicitly registering B onto A, rather than relying on numerical inversion. Full details of the building of bridging registrations and their directions are shown in Supplementary Figure 8. Here, the arrows indicate the direction of the forward transformation but, due to the ability to numerically invert the transformations, it is possible to travel 'backwards' along an arrow to transform in the opposite direction. While the inversion takes an appreciable time to calculate, the resulting errors are extremely small, far below the resolution of the original images, and only exist due to the finite precision with which the floating-point numbers are manipulated. By inverting and concatenating bridging registrations as appro-

priate, it is possible to transform data registered to any of the template spaces to any of the other template spaces.

4.4 Creating accurate registrations

Full, non-rigid warping registrations were computed using the Computational Morphometry Toolkit (CMTK), as described previously (Jefferis et al., 2007). An initial rigid affine registration with twelve degrees of freedom (translation, rotation and scaling of each axis) was followed by a non-rigid registration that allows different brain regions to move somewhat independently, subject to a smoothness penalty (Rueckert et al., 1999). In the non-rigid step, deformations between the independently moving control points are interpolated using B-splines, with image similarity being computed through the use of a normalized mutual information metric (Studholme et al., 1999). The task of finding an accurate registration is treated as an optimisation problem of the mutual information metric that, due to its complex nature, has many local optima in which the algorithm can become stuck. To help avoid this, a constraint is imposed to ensure the deformation field is spatially smooth across the brain, as is biological reasonable. Full details of the parameters passed to the CMTK tools are provided in the 'settings' file that accompanies each registration. To create mirroring registrations, images were first flipped horizontally in Fiji before being registered to the original template spaces using CMTK. For convenience, we also encoded the horizontal flip as a CMTK-compatible affine transformation, meaning that the entire process of mirroring a sample image can be carried in single step with CMTK.

4.5 Construction of new template spaces

The template space provided by the FlyLight project (JFRC) is not spatially calibrated and so we added spatial calibration to a copy named JFRC2. Similarly, FlyCircuit images are registered to male and female template spaces and so we created an intersex template space from 17 female and 9 male brains to bring all FlyCircuit neurons into a common space, irrespective of sex. The IS2, Cell07 and T1 template spaces were left unaltered.

As the neuropil and tract masks provided by the Insect Brain Name working group (Ito et al., 2014) only cover half a brain (IBN), we extended the IBN template space into a new whole brain template (named IBNWB) to improve the quality of the bridging registration between the IBN files and the other whole brain templates. The green channel (n-syb-GFP) of the tricolour confocal data provided was taken, duplicated and flipped about the medio-lateral axis using Fiji (Schindelin et al., 2012). The Fiji plugin 'Pairwise stitching' (Preibisch et al., 2009) was used to stitch the two stacks together with an offset of 392 pixels. This offset was chosen by eye as the one from the range of offsets 385–400 pixels that produced the most anatomically correct result. The overlapping region's intensity was set using the 'linear blend' method. We attempted improving on this alignment using the Fourier phase correlation method that the plugin also implements, but this gave poor results — the algorithm favoured overlapping the optic lobes, with a half central brain being present on each of the left and right sides.

As the template space is synthesized from an affine transformation of the original IBN template, we only considered an affine bridging registration between IBN and IBNWB. The n-syb-GFP labelling used in the IBN template strongly labels a large collection of cell bodies close to the cortex, posterior of the superior lateral protocerebrum and lateral horn, that are not labelled by nc82 or Dlg and hence the warping registrations from IBNWB to the other whole brain templates are less accurate in this region.

4.6 Construction of averaged template spaces

CMTK's `avg_adm` tool was used to iteratively produce new averaged seed brains given a set of template spaces and an initial seed brain drawn from the set. In each round, template spaces are registered to the seed brain and averaged to produce a new seed brain. After all rounds are complete, a final affine registration between the latest seed brain and a flipped version is calculated and then halved, resulting in a final brain that is centered in the middle of the image. The FCWB template was produced in this manner using 17 female and 9 male brains. We have developed documented tools to help users make average templates, available via https://github.com/jefferislab/MakeAverageBrain.

4.7 Application of registrations to images, traced neurons and surface data

CMTK provides two commands, `reformatx` and `streamxform` that will use a registration to reformat images and transform points, respectively. The R package nat (NeuroAnatomy Toolbox) wraps these commands and can use them to transform neuroanatomical data, stored as objects in the R session, between template spaces. A 3D surface model of the standard neuropil segmentation (Ito et al., 2014) was generated from the labelled image stack, using Amira, read into R using nat, transformed into the different template template spaces, via JFRC2, and saved as new 3D surfaces. These can then be used to segment neurons in their original space, providing interesting volumetric data for a neuron such as the relative density of neuropil innervation.

4.8 Flies

Wild-type (Canton S, Bloomington Stock Center, Indiana University) and transgenic strains were kept on standard yeast/agar medium at 25 °C. Transgenics were a GH146-lexA line and the dFasciculin-II-GFP protein trap line (courtesy of M. Landgraf). Lines were balanced with CyO, Dfd-GMR-YFP or TM6b, Sb, Dfd-GMR-YFP balancer chromosomes (Bloomington Stock Center, Indiana University).

4.9 Larval dissection, immunohistochemistry and imaging

Flies were mated a day before dissection and laid eggs on apple-juice based media with a spot of yeast paste overnight at 25°C. Adults and large hatched larvae were subsequently removed, and small embryos (approx. the length of an egg) were dissected in Sorensen's saline (pH 7.2, 0.075 M). A hypodermic needle (30 G × 0.5′; Microlance) was used to sever the

mouth hooks of each larva, at which point the CNS extruded along with viscera, and was gently separated and stuck to a cover glass that has been coated with poly-L-lysine (SigmaAldrich) in a bubble of solution. The CNS were then fixed in 4% formaldehyde (Fisher Scientific) in Sorensen's saline for 15 min at room temperature, and subsequently permeabilised in PBT (phosphate buffer with 0.3% Triton-X-100, SigmaAldrich). Incubated overnight in primary antibodies at 4°C and, after washes in PBT, in secondary antibodies for 2 hrs at room temperature. Washes took place in either a bubble of fluid or shallow dish filled with solution to prevent collapse of brain lobes into the VNC. For this reason also, confocal stacks were acquired with a 40× dipping lens on a Zeiss LSM 710, voxel resolution 0.2×0.2×0.5 microns^3. Primary antibodies used were Chicken anti-GFP (Invitrogen), 1: 10000, mouse IgG1 anti-FasciclinII (DSHB), 1:10, rat N-Cadherin (DSHB) and mouse IgG1 Discs large-1, 1:50. Secondaries used were goat anti-mouse CF568, 1:600, goat anti-Chicken Alexa488, goat anti-mouse CF647, 1:600. Some antibodies and dissection training were kindly supplied by M. Landgraf.

4.10 Visualisation

The majority of images shown in this manuscript were generated in RStudio. 3D images were plotted with natverse functions that depend on the R package rgl (Murdoch, 2001), 2D plots were generated using ggplot2 (Wickham, 2016). 3D images of confocal data were visualized using Amira 6.0, and Paraview. Figures were generated using Adobe Illustrator.

4.11 Data Availability

The bridging and mirroring registrations are deposited in two version controlled repositories at http://github.com/jefferislab with revisions uniquely identified by the SHA-1 hash function. As some template spaces may have multiple versions, we identify each version by its SHA-1 hash as this is uniquely dependent on the data contained in each file. Since we use the distributed version control system, git, any user can clone a complete, versioned history of these repositories. We have also taken a repository snapshot at the time of the release of this paper on the publicly funded http://zenodo.org site, which associates the data with a permanent digital object identifiers (DOI). To simplify data access for colleagues, we have provided spatially calibrated template spaces for the main template spaces in use by the Drosophila community in a single standard format, NRRD. These brain images have permanent DOIs listed in Table 2 on page 10. We have also generated registrations for the entire FlyCircuit single neuron and FlyLight datasets. The registered images have been deposited at http://virtualflybrain.org. The R packages nat.flybrains and elmr in the natverse also contain easy-to-use functions for deploying these registrations. The complete software toolchain for the construction and application of registrations consists exclusively of open source code released under the GNU Public License and released on GitHub and Sourceforge. A full listing of these resources is available at http://jefferislab.org/si/bridging. All of these steps will ensure that these resources will be available for many years to come, as has been recommended (Ito, 2010).

Acknowledgements

We are very grateful to the original data providers including Ann-Shyn Chiang, Gerry Rubin, Arnim Jenett, Tzumin Lee, Kazunori Shinomiya and Kei Ito for generously sharing their image data with the research community. We specifically thank Arnim Jenett, Kazunori Shinomiya and Kei Ito for sharing the nc82-based *D. melanogaster* neuropil segmentation. We thank M-J. Dolan for providing confocal microscopy exemplar images. Images from FlyCircuit were obtained from the NCHC (National Center for High-performance Computing) and NTHU (National Tsing Hua University), Hsinchu, Taiwan. We thank the Virtual Fly Brain team including MC, David Osumi-Sutherland, Robert Court, Cahir O'Kane and Douglas Armstrong for making some of our processed data available online through http://virtualflybrain.org. We note that data integration work with the virtualflybrain.org website was supported in part by an award from the Isaac Newton Trust to MC and Dr Cahir O'Kane. We thank Tom Kazimiers for help navigating the CATMAID API. We thank Alex Vourvoukelis, Alex von Klemperer, and Colin J. Akerman for sharing sample data.

We thank members of the Jefferis laboratory and the Drosophila Connectomics group for comments on this manuscript along with Jake Grimmett and Toby Darling for assistance with the LMB's computer cluster. This work made use of the Computational Morphometry Toolkit, supported by the National Institute of Biomedical Imaging and Bioengineering (NIBIB). We thank early users of the natverse for their help finding bugs and suggesting features, including but not limited to: István Taisz, Shanice Bailey, William Morris, Kathi Eichler, Dana Gallii, Sebastian Cachero, Erika Dona, Shahar Frechter, Konrad Heinz, Fiona Love, Paavo Huoviala, Amelia Edmondson-Stait and Lisa Marin.

This work was supported by a Starting Investigator Grant from the European Research Council, the Wellcome Trust (203261/Z/16/Z), an ERC Consolidator grant (649111), the Boehringer Ingelheim Fonds and Herchel Smith Studentship (A.S.B.), a Fitzwilliam College Research Fellowship (J.D.M.) and core support from the MRC (MC-U105188491) to G.S.X.E.J.

References

K. Anderson, B. Bones, B. Robinson, C. Hass, H. Lee, K. Ford, T.-A. Roberts, and B. Jacobs. The morphology of supragranular pyramidal neurons in the human insular cortex: a quantitative golgi study. *Cereb. Cortex*, 19(9):2131–2144, Sept. 2009.

K. Anderson, E. Yamamoto, J. Kaplan, M. Hannan, and B. Jacobs. Neurolucida lucivid versus neurolucida camera: A quantitative and qualitative comparison of three-dimensional neuronal reconstructions. *J. Neurosci. Methods*, 186(2):209–214, Feb. 2010.

G. A. Ascoli, D. E. Donohue, and M. Halavi. NeuroMorpho.Org: a central resource for neuronal morphologies. *J. Neurosci.*, 27(35):9247–9251, Aug. 2007.

Y. Aso and G. M. Rubin. Dopaminergic neurons write and

update memories with cell-type-specific rules. *Elife*, 5, July 2016.

Y. Aso, D. Hattori, Y. Yu, R. M. Johnston, N. A. Iyer, T.-T. Ngo, H. Dionne, L. F. Abbott, R. Axel, H. Tanimoto, and Others. The neuronal architecture of the mushroom body provides a logic for associative learning. *Elife*, 3:e04577, 2014.

B. B. Avants, N. Tustison, and G. Song. Advanced normalization tools (ANTS). *Insight J.*, 2:1–35, 2009.

A. S. Bates, J. Janssens, G. S. X. E. Jefferis, and S. Aerts. ScienceDirect Neuronal cell types in the fly : single-cell anatomy meets single-cell genomics. *Curr. Opin. Neurobiol.*, 56:125–134, 2019.

M. E. Berck, A. Khandelwal, L. Claus, L. Hernandez-Nunez, G. Si, C. J. Tabone, F. Li, J. W. Truman, R. D. Fetter, M. Louis, A. D. Samuel, and A. Cardona. The wiring diagram of a glomerular olfactory system. *Elife*, 5, May 2016.

L. Billeci, C. Magliaro, G. Pioggia, and A. Ahluwalia. NEuronMOrphological analysis tool: open-source software for quantitative morphometrics. *Front. Neuroinform.*, 7:2, Feb. 2013.

D. D. Bock, W.-C. A. Lee, A. M. Kerlin, M. L. Andermann, G. Hood, A. W. Wetzel, S. Yurgenson, E. R. Soucy, H. S. Kim, and R. C. Reid. Network anatomy and in vivo physiology of visual cortical neurons. *Nature*, 471(7337):177–182, Mar. 2011.

J. A. Bogovic, H. Otsuna, L. Heinrich, M. Ito, J. Jeter, G. W. Meissner, A. Nern, J. Colonell, O. Malkesman, K. Ito, and S. Saalfeld. An unbiased template of the drosophila brain and ventral nerve cord. July 2018.

A. Bône, M. Louis, B. Martin, and S. Durrleman. Deformetrica 4: An Open-Source software for statistical shape analysis. In *Shape in Medical Imaging*, pages 3–13. Springer International Publishing, 2018.

A. H. Brand and N. Perrimon. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development*, 118(2):401–415, June 1993.

R. Brandt, T. Rohlfing, J. Rybak, S. Krofczik, A. Maye, M. Westerhoff, H.-C. Hege, and R. Menzel. Three-dimensional average-shape atlas of the honeybee brain and its applications. *J. Comp. Neurol.*, 492(1):1–19, Nov. 2005.

K. M. Brown, D. E. Donohue, G. D'Alessandro, and G. A. Ascoli. A cross-platform freeware tool for digital reconstruction of neuronal arborizations from image stacks. *Neuroinformatics*, 3(4):343–360, 2005.

S. Busch, M. Selcho, K. Ito, and H. Tanimoto. A map of octopaminergic neurons in the Drosophila brain. *J. Comp. Neurol.*, 513(6):643–667, 2009.

S. Cachero, A. D. Ostrovsky, J. Y. Yu, B. J. Dickson, and G. S. X. E. Jefferis. Sexual dimorphism in the fly brain. *Curr. Biol.*, 20(18):1589–1601, Sept. 2010.

R. C. Cannon, D. A. Turner, G. K. Pyapali, and H. V. Wheal. An on-line archive of reconstructed hippocampal neurons. *J. Neurosci. Methods*, 84(1-2):49–54, Oct. 1998.

A. Cardona, S. Saalfeld, I. Arganda, W. Pereanu, J. Schindelin, and V. Hartenstein. Identifying neuronal lineages of drosophila by sequence analysis of axon tracts. *J. Neurosci.*, 30(22):7538–7553, June 2010.

A.-S. Chiang, C.-Y. Lin, C.-C. Chuang, H.-M. Chang, C.-H. Hsieh, C.-W. Yeh, C.-T. Shih, J.-J. Wu, G.-T. Wang, Y.-C. Chen, C.-C. Wu, G.-Y. Chen, Y.-T. Ching, P.-C. Lee, C.-Y. Lin, H.-H. Lin, C.-C. Wu, H.-W. Hsu, Y.-A. Huang, J.-Y. Chen, H.-J. Chiang, C.-F. Lu, R.-F. Ni, C.-Y. Yeh, and J.-K. Hwang. Three-dimensional reconstruction of brain-wide wiring networks in drosophila at single-cell resolution. *Curr. Biol.*, 21(1):1–11, Jan. 2011.

J. Clemens, C. C. Girardin, P. Coen, X.-J. Guan, B. J. Dickson, and M. Murthy. Connecting neural codes with behavior in the auditory system of drosophila. *Neuron*, 87(6):1332–1343, Sept. 2015.

J. Clemens, P. Coen, F. A. Roemschied, T. D. Pereira, D. Mazumder, D. E. Aldarondo, D. A. Pacheco, and M. Murthy. Discovery of a new song mode in drosophila reveals hidden structure in the sensory and neural drivers of behavior. *Curr. Biol.*, 28(15):2400–2412.e6, Aug. 2018.

S. J. Cook, T. A. Jarrell, C. A. Brittin, Y. Wang, A. E. Bloniarz, M. A. Yakovlev, K. C. Q. Nguyen, L. T.-H. Tang, E. A. Bayer, J. S. Duerr, H. E. Bülow, O. Hobert, D. H. Hall, and S. W. Emmons. Whole-animal connectomes of both caenorhabditis elegans sexes. *Nature*, 571(7763):63–71, July 2019.

M. Costa, J. D. Manton, A. D. Ostrovsky, S. Prohaska, and G. S. X. E. Jefferis. NBLAST: Rapid, sensitive comparison of neuronal structure and construction of neuron family databases. *Neuron*, 91(2):293–311, July 2016.

G. Csardi and T. Nepusz. *The Igraph Software Package for Complex Network Research*, volume Complex Systems. Nov. 2005.

H. Cuntz, F. Forstner, A. Borst, and M. Häusser. One rule to grow them all: a general theory of neuronal branching and its practical application. *PLoS Comput. Biol.*, 6(8), Aug. 2010.

A. M. Dacks, T. A. Christensen, and J. G. Hildebrand. Phylogeny of a serotonin-immunoreactive neuron in the primary olfactory center of the insect brain. *J. Comp. Neurol.*, 498(6):727–746, 2006.

L. de Vries, K. Pfeiffer, B. Trebels, A. K. Adden, K. Green, E. Warrant, and S. Heinze. Comparison of Navigation-Related brain regions in migratory versus Non-Migratory noctuid moths. *Front. Behav. Neurosci.*, 11:158, Sept. 2017.

R. R. Development Core Team. R: A language and environment for statistical computing. 2011.

M.-J. Dolan, G. Belliart-Guérin, A. S. Bates, S. Frechter, A. Lampin-Saint-Amaux, Y. Aso, R. J. V. Roberts, P. Schlegel, A. Wong, A. Hammad, D. Bock, G. M. Rubin, T. Preat, P.-Y. Plaçais, and G. S. X. E. Jefferis. Communication from learned to innate olfactory processing centers is required for memory retrieval in drosophila. *Neuron*, Sept. 2018.

M.-J. Dolan, S. Frechter, A. S. Bates, C. Dan, P. Huoviala, R. J. Roberts, P. Schlegel, S. Dhawan, R. Tabano, H. Dionne, C. Christoforou, K. Close, B. Sutcliffe, B. Giuliani, F. Li, M. Costa, G. Ihrke, G. W. Meissner, D. D. Bock, Y. Aso, G. M. Rubin, and G. S. Jefferis. Neurogenetic dissection of the drosophila lateral horn reveals major outputs, diverse behavioural functions, and interactions with the mushroom body. *Elife*, 8, May 2019.

T. Duong. ks: Kernel density estimation and kernel discriminant analysis for multivariate data in R. *J. Stat. Softw.*, 2007.

S. Durrleman, M. Prastawa, N. Charon, J. R. Korenberg, S. Joshi, G. Gerig, and A. Trouvé. Morphometry of anatomical shape complexes with dense deformations and sparse parameters. *Neuroimage*, 101:35–49, Nov. 2014.

M. N. Economo, N. G. Clack, L. D. Lavis, C. R. Gerfen, K. Svoboda, E. W. Myers, and J. Chandrashekar. A platform for brain-wide imaging and reconstruction of individual neurons. *Elife*, 5:e10566, Jan. 2016.

K. Eichler, F. Li, A. Litwin-Kumar, Y. Park, I. Andrade, C. M. Schneider-Mizell, T. Saumweber, A. Huser, C. Eschbach, B. Gerber, R. D. Fetter, J. W. Truman, C. E. Priebe, L. F. Abbott, A. S. Thum, M. Zlatic, and A. Cardona. The complete connectome of a learning and memory centre in an insect brain. *Nature*, 548(7666):175–182, Aug. 2017.

B. El Jundi, E. J. Warrant, K. Pfeiffer, and M. Dacke. Neuroarchitecture of the dung beetle central complex. *J. Comp. Neurol.*, 526(16):2612–2630, Nov. 2018.

R. Farhoodi, B. J. Lansdell, and K. P. Kording. Quantifying how staining methods bias measurements of neuron morphologies. *Front. Neuroinform.*, 13:36, May 2019.

J. Felsenberg, P. F. Jacob, T. Walker, O. Barnstedt, A. J. Edmondson-Stait, M. W. Pleijzier, N. Otto, P. Schlegel, N. Sharifi, E. Perisse, C. S. Smith, J. S. Lauritzen, M. Costa, G. S. X. E. Jefferis, D. D. Bock, and S. Waddell. Integration of parallel opposing memories underlies memory extinction. *Cell*, 175(3):709–722.e15, Oct. 2018.

L. Feng, T. Zhao, and J. Kim. neutube 1.0: A new design for efficient neuron reconstruction software based on the SWC format. *eNeuro*, 2(1), Jan. 2015.

S. Frechter, A. S. Bates, S. Tootoonian, M.-J. Dolan, J. D. Manton, A. R. Jamasb, J. Kohl, D. Bock, and G. S. Jefferis. Functional and anatomical specificity in a higher olfactory centre. *Elife*, 8, May 2019.

J. Funke, F. Tschopp, W. Grisaitis, A. Sheridan, C. Singh, S. Saalfeld, and S. C. Turaga. Large scale image segmentation with structured loss based deep learning for connectome reconstruction. *IEEE Trans. Pattern Anal. Mach. Intell.*, 41(7):1669–1680, July 2019.

J. C. Gensel, D. L. Schonberg, J. K. Alexander, D. M. McTigue, and P. G. Popovich. Semi-automated sholl analysis for quantifying changes in growth and differentiation of neurons and glia. *J. Neurosci. Methods*, 190(1):71–79, June 2010.

J. R. Glaser and E. M. Glaser. Neuron imaging with neurolucida–a PC-based system for image combining microscopy. *Comput. Med. Imaging Graph.*, 14(5):307–317, Sept. 1990.

Y. Grosjean, R. Rytz, J.-P. Farine, L. Abuin, J. Cortot, G. S. X. E. Jefferis, and R. Benton. An olfactory receptor for food-derived odours promotes male courtship in drosophila. *Nature*, 478(7368):236–240, Oct. 2011.

M. Halavi, S. Polavaram, D. E. Donohue, G. Hamilton, J. Hoyt, K. P. Smith, and G. A. Ascoli. NeuroMorpho.Org implementation of digital neuroscience: dense coverage and integration with the NIF. *Neuroinformatics*, 6(3):241–252, Sept. 2008.

L. Heinrich, J. Funke, C. Pape, J. Nunez-Iglesias, and S. Saalfeld. Synaptic cleft segmentation in Non-Isotropic volume electron microscopy of the complete drosophila brain. May 2018.

S. Heinze and S. M. Reppert. Anatomical basis of sun compass navigation i: the general layout of the monarch butterfly brain. *J. Comp. Neurol.*, 520(8):1599–1628, June 2012.

M. Helmstaedter, K. L. Briggman, S. C. Turaga, V. Jain, H. S. Seung, and W. Denk. Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature*, 500(7461): 168–174, Aug. 2013.

S.-Y. Ho, C.-Y. Chao, H.-L. Huang, T.-W. Chiu, P. Charoenkwan, and E. Hwang. NeurphologyJ: an automatic neuronal morphology quantification method and its application in pharmacological discovery. *BMC Bioinformatics*, 12:230, June 2011.

J. A. Horne, C. Langille, S. McLin, M. Wiederman, Z. Lu, C. S. Xu, S. M. Plaza, L. K. Scheffer, H. F. Hess, and I. A. Meinertzhagen. A resource for the drosophila antennal lobe provided by the connectome of glomerulus VA1v. *Elife*, 7, Nov. 2018.

W. Huber, V. J. Carey, R. Gentleman, S. Anders, M. Carlson, B. S. Carvalho, H. C. Bravo, S. Davis, L. Gatto, T. Girke, R. Gottardo, F. Hahne, K. D. Hansen, R. A. Irizarry, M. Lawrence, M. I. Love, J. MacDonald, V. Obenchain, A. K. Oleś, H. Pagès, A. Reyes, P. Shannon, G. K. Smyth, D. Tenenbaum, L. Waldron, and M. Morgan. Orchestrating high-throughput genomic analysis with bioconductor. *Nat. Methods*, 12(2):115–121, Feb. 2015.

P. Huoviala, M.-J. Dolan, F. Love, S. Frechter, R. J. V. Roberts, Z. Mitrevica, P. Schlegel, A. S. S. Bates, Y. Aso, T. Rodrigues, H. Cornwall, M. Stensmyr, D. Bock, G. M. Rubin, M. Costa, and G. S. X. E. Jefferis. Neural circuit basis of aversive odour processing in drosophila from sensory input to descending output. *bioRxiv*, 2018.

K. Ito. Technical and organizational considerations for the long-term maintenance and development of the digital brain atlases and web-based databases. *Front. Syst. Neurosci.*, 4(26), 2010.

K. Ito, K. Shinomiya, M. Ito, J. D. Armstrong, G. Boyan, V. Hartenstein, S. Harzsch, M. Heisenberg, U. Homberg, A. Jenett, H. Keshishian, L. L. Restifo, W. Rössler, J. H. Simpson, N. J. Strausfeld, R. Strauss, L. B. Vosshall, and Insect Brain Name Working Group. A systematic nomenclature for the insect brain. *Neuron*, 81(4):755–765, Feb. 2014.

M. Ito, N. Masuda, K. Shinomiya, K. Endo, and K. Ito. Systematic analysis of neural projections reveals clonal composition of the drosophila brain. *Curr. Biol.*, 23(8):644–655, Apr. 2013.

B. Jacobs, L. Driscoll, and M. Schall. Life-span dendritic and spine changes in areas 10 and 18 of human cortex: a quantitative golgi study. *J. Comp. Neurol.*, 386(4):661–680, Oct. 1997.

B. Jacobs, M. Schall, M. Prather, E. Kapler, L. Driscoll, S. Baca, J. Jacobs, K. Ford, M. Wainwright, and M. Treml. Regional dendritic and spine variation in human cerebral cortex: a quantitative golgi study. *Cereb. Cortex*, 11(6):558–571, June 2001.

B. Jacobs, J. Lubs, M. Hannan, K. Anderson, C. Butti, C. C. Sherwood, P. R. Hof, and P. R. Manger. Neuronal morphology in the african elephant (loxodonta africana) neocortex. *Brain Struct. Funct.*, 215(3-4):273–298, Jan. 2011.

B. Jacobs, T. Harland, D. Kennedy, M. Schall, B. Wicinski, C. Butti, P. R. Hof, C. C. Sherwood, and P. R. Manger. The neocortex of cetartiodactyls. II. neuronal morphology of the visual and motor cortices in the giraffe (giraffa camelopardalis). *Brain Struct. Funct.*, 220(5):2851–2872, Sept. 2015.

B. Jacobs, L. Lee, M. Schall, M. A. Raghanti, A. H. Lewandowski, J. J. Kottwitz, J. F. Roberts, P. R. Hof, and C. C. Sherwood. Neocortical neuronal morphology in the newborn giraffe (giraffa camelopardalis tippelskirchi) and african elephant (loxodonta africana). *J. Comp. Neurol.*, 524(2):257–287, Feb. 2016.

B. Jacobs, M. E. Garcia, N. B. Shea-Shumsky, M. E. Tennison, M. Schall, M. S. Saviano, T. A. Tummino, A. J. Bull, L. L. Driscoll, M. A. Raghanti, A. H. Lewandowski, B. Wicinski, H. Ki Chui, M. F. Bertelsen, T. Walsh, A. Bhagwandin, M. A. Spocter, P. R. Hof, C. C. Sherwood, and P. R. Manger. Comparative morphology of gigantopyramidal neurons in primary motor cortex across mammals. *J. Comp. Neurol.*, 526 (3):496–536, Feb. 2018.

M. Januszewski, J. Kornfeld, P. H. Li, A. Pope, T. Blakely, L. Lindsey, J. Maitin-Shepard, M. Tyka, W. Denk, and V. Jain. High-precision automated reconstruction of neurons with flood-filling networks. *Nat. Methods*, 15(8):605–610, Aug. 2018.

J. M. Jeanne and R. I. Wilson. Convergence, divergence, and reconvergence in a feedforward network improves neural speed and accuracy. *Neuron*, 88(5):1014–1026, 2015.

J. M. Jeanne, M. Fişek, and R. I. Wilson. The organization of projections from olfactory glomeruli onto Higher-Order neurons. *Neuron*, 98(6):1198–1213.e6, June 2018.

G. S. X. E. Jefferis, C. J. Potter, A. M. Chan, E. C. Marin, T. Rohlfing, C. R. J. Maurer, and L. Luo. Comprehensive maps of drosophila higher olfactory centers: spatially segregated fruit and pheromone representation. *Cell*, 128(6):1187–1203, 2007.

A. Jenett, G. M. Rubin, T.-T. Ngo, D. Shepherd, C. Murphy, H. Dionne, B. D. Pfeiffer, A. Cavallaro, D. Hall, J. Jeter, and Others. A GAL4-Driver line resource for drosophila neurobiology. *Cell Rep.*, 2(4):991–1001, 2012.

B. M. Kandel, P. A. Cook, N. J. Tustison, and J. Muschelli. ANTsRCore: Core software infrastructure for 'ANTsR', 2019.

W. T. Katz and S. M. Plaza. DVID: Distributed versioned Image-Oriented dataservice. *Front. Neural Circuits*, 13:5, Feb. 2019.

K.-M. Kim, K. Son, and G. T. R. Palmore. Neuron image analyzer: Automated and accurate extraction of neuronal data from low quality images. *Sci. Rep.*, 5:17062, Nov. 2015.

N. C. Klapoetke, A. Nern, M. Y. Peek, E. M. Rogers, P. Breads, G. M. Rubin, M. B. Reiser, and G. M. Card. Ultra-selective looming detection from radial motion opponency. *Nature*, 551(7679):237–241, Nov. 2017.

J. Kohl, A. D. Ostrovsky, S. Frechter, and G. S. X. E. Jefferis. A bidirectional circuit switch reroutes pheromone signals in male and female brains. *Cell*, 155(7):1610–1623, Dec. 2013.

M. Kunst, E. Laurell, N. Mokayes, A. Kramer, F. Kubo, A. M. Fernandes, D. Förster, M. Dal Maschio, and H. Baier. A Cellular-Resolution atlas of the larval zebrafish brain. *Neuron*, May 2019.

A. E. Kurylas, T. Rohlfing, S. Krofczik, A. Jenett, and U. Homberg. Standardized atlas of the brain of the desert locust, schistocerca gregaria. *Cell Tissue Res.*, 333(1):125–145, July 2008.

E. Z. Kvon, T. Kazmar, G. Stampfel, J. O. Yáñez-Cuna, M. Pagani, K. Schernhuber, B. J. Dickson, and A. Stark. Genome-scale functional characterization of drosophila developmental enhancers in vivo. *Nature*, 512(7512):91–95, Aug. 2014.

T. Lafarge, B. Pateiro-López, A. Possolo, and J. Dunkers. R implementation of a polyhedral approximation to a 3D set of points using the α-Shape. *J. Stat. Softw.*, 056(i04), 2014.

S.-L. Lai and T. Lee. Genetic mosaic with dual binary transcriptional systems in drosophila. *Nat. Neurosci.*, 9(5):703–709, May 2006.

T. Lee and L. Luo. Mosaic analysis with a repressible cell marker (MARCM) for drosophila neural development. *Trends Neurosci.*, 24(5):251–254, May 2001.

P. H. Li, L. F. Lindsey, M. Januszewski, Z. Zheng, A. S. Bates, I. Taisz, M. Tyka, M. Nichols, F. Li, E. Perlman, J. Maitin-Shepard, T. Blakely, L. Leavitt, Gregory S X, D. Bock, and V. Jain. Automated reconstruction of a Serial-Section EM drosophila brain with Flood-Filling networks and local realignment. Apr. 2019.

Y. Li, D. Wang, G. A. Ascoli, P. Mitra, and Y. Wang. Metrics for comparing neuronal tree shapes based on persistent homology. *PLoS One*, 12(8):1–24, 2017.

B. B. Løfaldli, P. Kvello, and H. Mustaparta. Integration of the antennal lobe glomeruli and three projection neurons in the standard brain atlas of the moth heliothis virescens. *Front. Syst. Neurosci.*, 4:5, Feb. 2010.

M. H. Longair, D. A. Baker, and J. D. Armstrong. Simple neurite tracer: open source software for reconstruction, visualization and analysis of neuronal processes. *Bioinformatics*, 27(17):2453–2454, Sept. 2011.

H. Luan, N. C. Peabody, C. R. Vinson, and B. H. White. Refined spatial manipulation of neuronal function by combinatorial restriction of transgene expression. *Neuron*, 52(3): 425–436, 2006.

N. Y. Masse, S. Cachero, A. D. Ostrovsky, and G. S. X. E. Jefferis. A mutual information approach to automate identification of neuronal clusters in drosophila brain images. *Front. Neuroinform.*, 6:21, June 2012.

E. Meijering, M. Jacob, J.-C. F. Sarria, P. Steiner, H. Hirling, and M. Unser. Design and validation of a tool for neurite tracing and analysis in fluorescence microscopy images. *Cytometry A*, 58(2):167–176, Apr. 2004.

D. Murdoch. RGL: An R interface to OpenGL. In *Proceedings of DSC*, page 2. pdfs.semanticscholar.org, 2001.

D. R. Myatt, T. Hadlington, G. A. Ascoli, and S. J. Nasuto. Neuromantic - from semi-manual to semi-automatic reconstruction of neuron morphology. *Front. Neuroinform.*, 6: 4, Mar. 2012.

S. Namiki, M. H. Dickinson, A. M. Wong, W. Korff, and G. M. Card. The functional organization of descending sensory-motor pathways in drosophila. *Elife*, 7, June 2018.

S. Nanda, M. M. Allaham, M. Bergamino, S. Polavaram, R. Armañanzas, G. A. Ascoli, and R. Parekh. Doubling up on the fly: NeuroMorpho.Org meets big data. *Neuroinformatics*, 13(1):127–129, Jan. 2015.

M. L. Narro, F. Yang, R. Kraft, C. Wenk, A. Efrat, and L. L. Restifo. NeuronMetrics: software for semi-automated processing of cultured neuron images. *Brain Res.*, 1138:57–75, Mar. 2007.

S. W. Oh, J. A. Harris, L. Ng, B. Winslow, N. Cain, S. Mihalas, Q. Wang, C. Lau, L. Kuan, A. M. Henry, M. T. Mortrud, B. Ouellette, T. N. Nguyen, S. A. Sorensen, C. R. Slaughterbeck, W. Wakeman, Y. Li, D. Feng, A. Ho, E. Nicholas, K. E. Hirokawa, P. Bohn, K. M. Joines, H. Peng, M. J. Hawrylycz, J. W. Phillips, J. G. Hohmann, P. Wohnoutka, C. R. Gerfen, C. Koch, A. Bernard, C. Dang, A. R. Jones, and H. Zeng. A mesoscale connectome of the mouse brain. *Nature*, 508 (7495):207–214, Apr. 2014.

T. Ohyama, C. M. Schneider-Mizell, R. D. Fetter, J. V. Aleman, R. Franconville, M. Rivera-Alba, B. D. Mensh, K. M. Branson, J. H. Simpson, J. W. Truman, A. Cardona, and M. Zlatic. A multilevel multimodal circuit enhances action selection in drosophila. *Nature*, 520(7549):633–639, 2015.

H. Otsuna, M. Ito, and T. Kawase. Color depth MIP mask search: a new tool to expedite Split-GAL4 creation. May 2018.

A. Pascual, K.-L. Huang, J. Neveu, and T. Préat. Neuroanatomy: brain asymmetry and long-term memory. *Nature*, 427(6975):605–606, 2004.

H. Peng, A. Bria, Z. Zhou, G. Iannello, and F. Long. Extensible visualization and analysis for multidimensional images using Vaa3D. *Nat. Protoc.*, 9(1):193–208, Jan. 2014.

B. D. Pfeiffer, A. Jenett, A. S. Hammonds, T.-T. B. Ngo, S. Misra, C. Murphy, A. Scully, J. W. Carlson, K. H. Wan, T. R. Laverty, C. Mungall, R. Svirskas, J. T. Kadonaga, C. Q. Doe, M. B. Eisen, S. E. Celniker, and G. M. Rubin. Tools for neuroanatomy and neurogenetics in drosophila. *Proc. Natl. Acad. Sci. U. S. A.*, 105(28):9715–9720, July 2008.

M. Pool, J. Thiemann, A. Bar-Or, and A. E. Fournier. NeuriteTracer: a novel ImageJ plugin for automated quantification of neurite outgrowth. *J. Neurosci. Methods*, 168(1):134–139, Feb. 2008.

S. Preibisch, S. Saalfeld, and P. Tomancak. Globally optimal stitching of tiled 3D microscopic image acquisitions. *Bioinformatics*, 25(11):1463–1465, 2009.

W. S. Rasband. Imagej, us national institutes of health, bethesda, maryland, usa. *http://imagej. nih. gov/ij/*, 2011.

C. L. Rees, K. Moradi, and G. A. Ascoli. Weighing the evidence in peters' rule: Does neuronal morphology predict connectivity? *Trends Neurosci.*, 40(2):63–71, 2017.

K. Rein, M. Zockler, M. T. Mader, C. Grubel, and M. Heisenberg. The drosophila standard brain. *Curr. Biol.*, 12(3):227–231, 2002.

L. D. Reyes, T. Harland, R. L. Reep, C. C. Sherwood, and B. Jacobs. Golgi analysis of neuron morphology in the presumptive somatosensory cortex and visual cortex of the florida manatee (trichechus manatus latirostris). *Brain Behav. Evol.*, 87(2):105–116, May 2016.

A. A. Robie, J. Hirokawa, A. W. Edwards, L. A. Umayam, A. Lee, M. L. Phillips, G. M. Card, W. Korff, G. M. Rubin, J. H. Simpson, M. B. Reiser, and K. Branson. Mapping the neural substrates of behavior. *Cell*, 170(2):393–406.e28, July 2017.

T. Rohlfing and J. Maurer, C. R. Nonrigid image registration in shared-memory multiprocessor environments with application to brains, breasts, and bees. *IEEE Trans. Inf. Technol. Biomed.*, 7(1):16–25, Apr. 2003.

D. Rueckert, L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. Hawkes. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans. Med. Imaging*, 18(8):712–721, Aug. 1999.

K. Ryan, Z. Lu, and I. A. Meinertzhagen. The CNS connectome of a tadpole larva of ciona intestinalis (l.) highlights sidedness in the brain of a chordate sibling. *Elife*, 5, Dec. 2016.

S. Saalfeld, A. Cardona, V. Hartenstein, and P. Toman????k. CATMAID: Collaborative annotation toolkit for massive amounts of image data. *Bioinformatics*, 25(15):1984–1986, 2009.

T. Saumweber, A. Rohwedder, M. Schleyer, K. Eichler, Y.-C. Chen, Y. Aso, A. Cardona, C. Eschbach, O. Kobler, A. Voigt, A. Durairaja, N. Mancini, M. Zlatic, J. W. Truman, A. S. Thum, and B. Gerber. Functional architecture of reward learning in mushroom body extrinsic neurons of larval drosophila. *Nat. Commun.*, 9(1):1104, Mar. 2018.

S. Sayin, J.-F. De Backer, K. P. Siju, M. E. Wosniack, L. P. Lewis, L.-M. Frisch, B. Gansen, P. Schlegel, A. Edmondson-Stait, N. Sharifi, C. B. Fisher, S. A. Calle-Schuler, J. Scott Lauritzen, D. D. Bock, M. Costa, Gregory S X, J. Gjorgjieva, and I. C. Grunwald Kadow. A neural circuit arbitrates between persistence and withdrawal in hungry drosophila. *Neuron*, 0(0), Aug. 2019.

L. K. Scheffer and I. A. Meinertzhagen. The fly brain atlas. *Annu. Rev. Cell Dev. Biol.*, July 2019.

J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A. Cardona. Fiji: an open-source platform for biological-image analysis. *Nat. Methods*, 9(7): 676–682, June 2012.

S. Schlager. *Morpho and Rvcg - Shape Analysis in R*, pages 217–256. Dec. 2017.

P. Schlegel, M. Costa, and G. S. Jefferis. Learning from connectomics on the fly. *Curr Opin Insect Sci*, 24:96–105, Dec. 2017.

S. K. Schmitz, J. J. J. Hjorth, R. M. S. Joemai, R. Wijntjes, S. Eijgenraam, P. de Bruijn, C. Georgiou, A. P. H. de Jong, A. van Ooyen, M. Verhage, L. N. Cornelisse, R. F. Toonen, and W. J. H. Veldkamp. Automated analysis of neuronal morphology, synapse number and synaptic recruitment. *J. Neurosci. Methods*, 195(2):185–193, Feb. 2011.

C. M. Schneider-Mizell, S. Gerhard, M. Longair, T. Kazimiers, F. Li, M. F. Zwart, A. Champion, F. M. Midgley, R. D. Fetter, S. Saalfeld, and A. Cardona. Quantitative neuroanatomy for connectomics in drosophila. *Elife*, 5, Mar. 2016.

C.-T. Shih, O. Sporns, S.-L. Yuan, T.-S. Su, Y.-J. Lin, C.-C. Chuang, T.-Y. Wang, C.-C. Lo, R. J. Greenspan, and A.-S. Chiang. Connectomics-based analysis of information flow in the drosophila brain. *Curr. Biol.*, 25(10):1249–1258, May 2015.

D. A. Sholl. Dendritic organization in the neurons of the visual and motor cortices of the cat. *J. Anat.*, 87(4):387–406, Oct. 1953.

A. Stepanyants and D. B. Chklovskii. Neurogeometry and potential synaptic connectivity. *Trends Neurosci.*, 28(7):387–394, 2005.

T. Stone, B. Webb, A. Adden, N. B. Weddig, A. Honkanen, R. Templin, W. Wcislo, L. Scimeca, E. Warrant, and S. Heinze. An anatomically constrained model for path integration in the bee brain. *Curr. Biol.*, 27(20):3069–3085.e11, Oct. 2017.

C. Studholme, D. L. G. Hill, and D. J. Hawkes. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognit.*, 32(1):71–86, 1999.

U. Sümbül, S. Song, K. McCulloch, M. Becker, B. Lin, J. R. Sanes, R. H. Masland, and H. S. Seung. A genetic and computational approach to structurally classify neuronal types. *Nat. Commun.*, 5:3512, 2014.

S.-Y. Takemura, A. Bharioke, Z. Lu, A. Nern, S. Vitaladevuni, P. K. Rivlin, W. T. Katz, D. J. Olbris, S. M. Plaza, P. Winston, and Others. A visual motion detection circuit suggested by drosophila connectomics. *Nature*, 500(7461):175–181, 2013.

S.-Y. Takemura, C. S. Xu, Z. Lu, P. K. Rivlin, T. Parag, D. J. Olbris, S. Plaza, T. Zhao, W. T. Katz, L. Umayam, C. Weaver, H. F. Hess, J. A. Horne, J. Nunez-Iglesias, R. Aniceto, L.-A. Chang, S. Lauchie, A. Nasca, O. Ogundeyi, C. Sigmund, S. Takemura, J. Tran, C. Langille, K. Le Lacheur, S. McLin, A. Shinomiya, D. B. Chklovskii, I. A. Meinertzhagen, and L. K. Scheffer. Synaptic circuits and their variations within different columns in the visual system of drosophila. *Proc. Natl. Acad. Sci. U. S. A.*, 112(44):13711–13716, Nov. 2015.

S. y. Takemura, Y. Aso, T. Hige, A. Wong, Z. Lu, C. S. Xu, P. K. Rivlin, H. Hess, T. Zhao, T. Parag, S. Berg, G. Huang, W. Katz, D. J. Olbris, S. Plaza, L. Umayam, R. Aniceto, L. A. Chang, S. Lauchie, O. Ogundeyi, C. Ordish, A. Shinomiya, C. Sigmund, S. Takemura, J. Tran, G. C. Turner, G. M. Rubin, and L. K. Scheffer. A connectome of a learning and memory center in the adult drosophila brain. *Elife*, 6:1–43, 2017.

N. K. Tanaka, K. Endo, and K. Ito. Organization of antennal lobe-associated neurons in adult drosophila melanogaster brain. *J. Comp. Neurol.*, 520(18):4067–4130, 2012.

L. Tirian and B. Dickson. The VT GAL4, LexA, and split-GAL4 driver line collections for targeted expression in the drosophila nervous system. *bioRxiv*, 2017.

W. F. Tobin, R. I. Wilson, and W.-C. A. Lee. Wiring variations that enable and constrain neural computation in a sensory microcircuit. *Elife*, 6, May 2017.

K. Travis, K. Ford, and B. Jacobs. Regional dendritic variation in neonatal human cortex: a quantitative golgi study. *Dev. Neurosci.*, 27(5):277–287, 2005.

J. B. Treweek, K. Y. Chan, N. C. Flytzanis, B. Yang, B. E. Deverman, A. Greenbaum, A. Lignell, C. Xiao, L. Cai, M. S. Ladinsky, P. J. Bjorkman, C. C. Fowlkes, and V. Gradinaru. Whole-body tissue stabilization and selective extractions via tissue-hydrogel hybrids for high-resolution intact circuit mapping and phenotyping. *Nat. Protoc.*, 10(11):1860–1896, Nov. 2015.

K. J. T. Venken, J. H. Simpson, and H. J. Bellen. Genetic manipulation of genes and cells in the nervous system of the fruit fly. *Neuron*, 72(2):202–230, Oct. 2011.

Y. Wan, F. Long, L. Qu, H. Xiao, M. Hawrylycz, E. W. Myers, and H. Peng. BlastNeuron for automated comparison, retrieval and clustering of 3D neuron morphologies. *Neuroinformatics*, 13(4):487–499, Oct. 2015.

S. L. Wearne, A. Rodriguez, D. B. Ehlenberger, A. B. Rocher, S. C. Henderson, and P. R. Hof. New techniques for imaging, digitization and analysis of three-dimensional neural morphology on multiple scales. *Neuroscience*, 136(3):661–680, 2005.

H. Wickham. *R Packages: Organize, Test, Document, and Share Your Code*. "O'Reilly Media, Inc.", Mar. 2015.

H. Wickham. ggplot2: Elegant graphics for data analysis, 2016.

J. Winnubst, E. Bas, T. A. Ferreira, Z. Wu, M. N. Economo, P. Edson, B. J. Arthur, C. Bruns, K. Rokicki, D. Schauder, D. J. Olbris, S. D. Murphy, D. G. Ackerman, C. Arshadi, P. Baldwin, R. Blake, A. Elsayed, M. Hasan, D. Ramirez, B. Dos Santos, M. Weldon, A. Zafar, J. T. Dudman, C. R. Gerfen, A. W. Hantman, W. Korff, S. M. Sternson, N. Spruston, K. Svoboda, and J. Chandrashekar. Reconstruction of 1,000 projection neurons reveals new cell types and organization of Long-Range connectivity in the mouse brain. *Cell*, Aug. 2019.

T. Wolff and G. M. Rubin. 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.*, 526(16):2585–2611, Nov. 2018.

D. C. Wong, J. K. Lovick, K. T. Ngo, W. Borisuthirattana, J. J. Omoto, and V. Hartenstein. Postembryonic lineages of the drosophila brain: II. identification of lineage projection patterns based on MARCM clones. *Dev. Biol.*, 384(2):258–289, Dec. 2013.

M. Wu, A. Nern, W. R. Williamson, M. M. Morimoto, M. B. Reiser, G. M. Card, and G. M. Rubin. Visual projection neurons in the drosophila lobula link feature detection to distinct behavioral programs. *Elife*, 5, Dec. 2016.

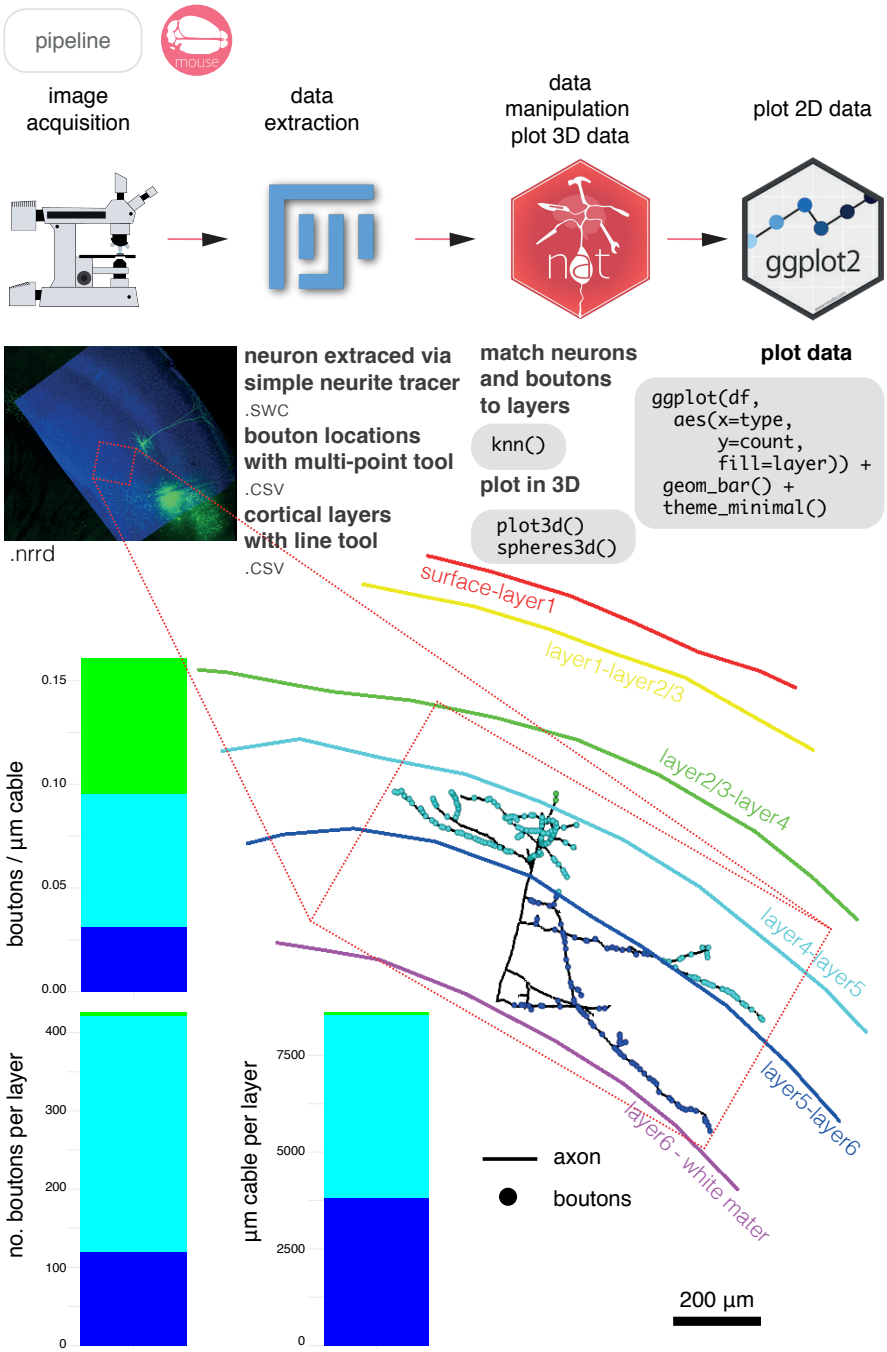
H.-H. Yu, T. Awasaki, M. D. Schroeder, F. Long, J. S. Yang, Y. He, P. Ding, J.-C. Kao, G. Y.-Y. Wu, H. Peng, G. Myers, and T. Lee. Clonal development and organization of the adult drosophila central brain. *Curr. Biol.*, 23, Mar. 2013.

J. Y. Yu, M. I. Kanai, E. Demir, G. S. X. E. Jefferis, and B. J. Dickson. Cellular organization of the neural circuit that drives drosophila courtship behavior. *Curr. Biol.*, 20(18): 1602–1614, Sept. 2010.

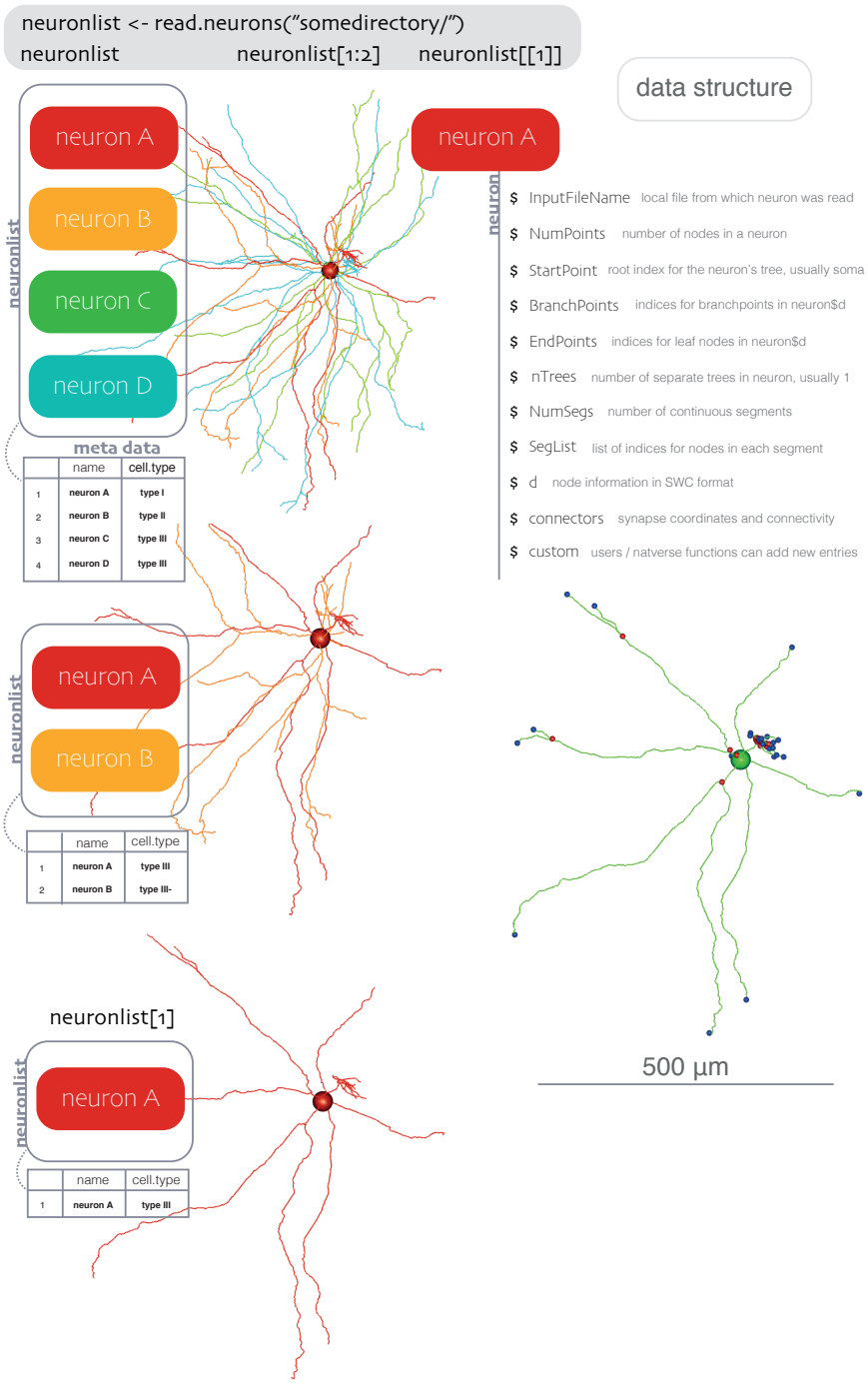
T. Zhao and S. M. Plaza. Automatic neuron type identification by neurite localization in the drosophila medulla. Sept. 2014.

Z. Zheng, J. S. Lauritzen, E. Perlman, C. G. Robinson, M. Nichols, D. Milkie, O. Torrens, J. Price, C. B. Fisher, N. Sharifi, S. A. Calle-Schuler, L. Kmecova, I. J. Ali, B. Karsh, E. T. Trautman, J. A. Bogovic, P. Hanslovsky, G. S. X. E. Jefferis, M. Kazhdan, K. Khairy, S. Saalfeld, R. D. Fetter, and D. D. Bock. A complete electron microscopy volume of the brain of adult drosophila melanogaster. *Cell*, 174(3):730–743.e22, July 2018.

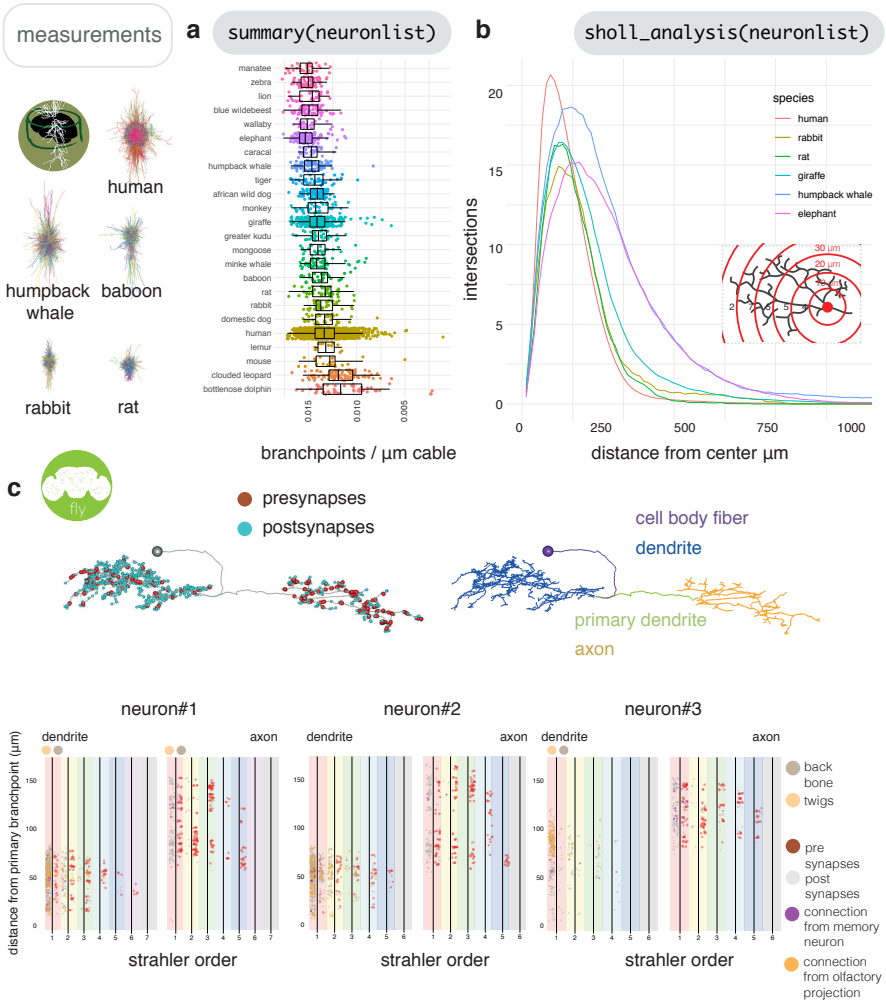
Supplementary Information



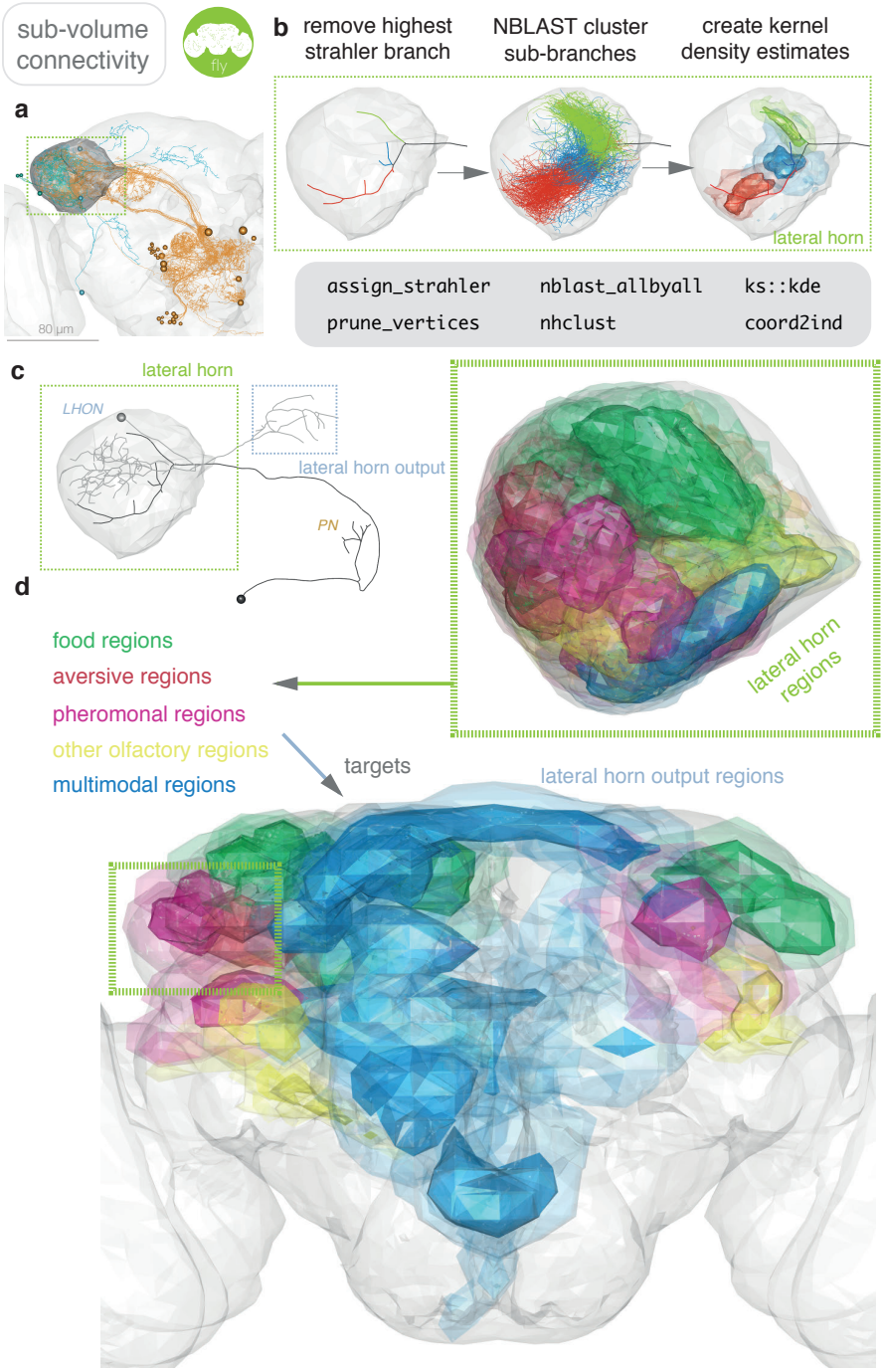
Supplementary Figure 1: A simple pipeline for neuron analysis with nat. Bouton placement for a layer 5 pyramidal neuron from the mouse primary somatosensory cortex is examined relative to the neurons' cable length and position within the layered structure of the cortex. Coloured names indicate layer transitions. Data courtesy of A. Vourvoukelis, A. von Klemperer and C.J. Akerman.



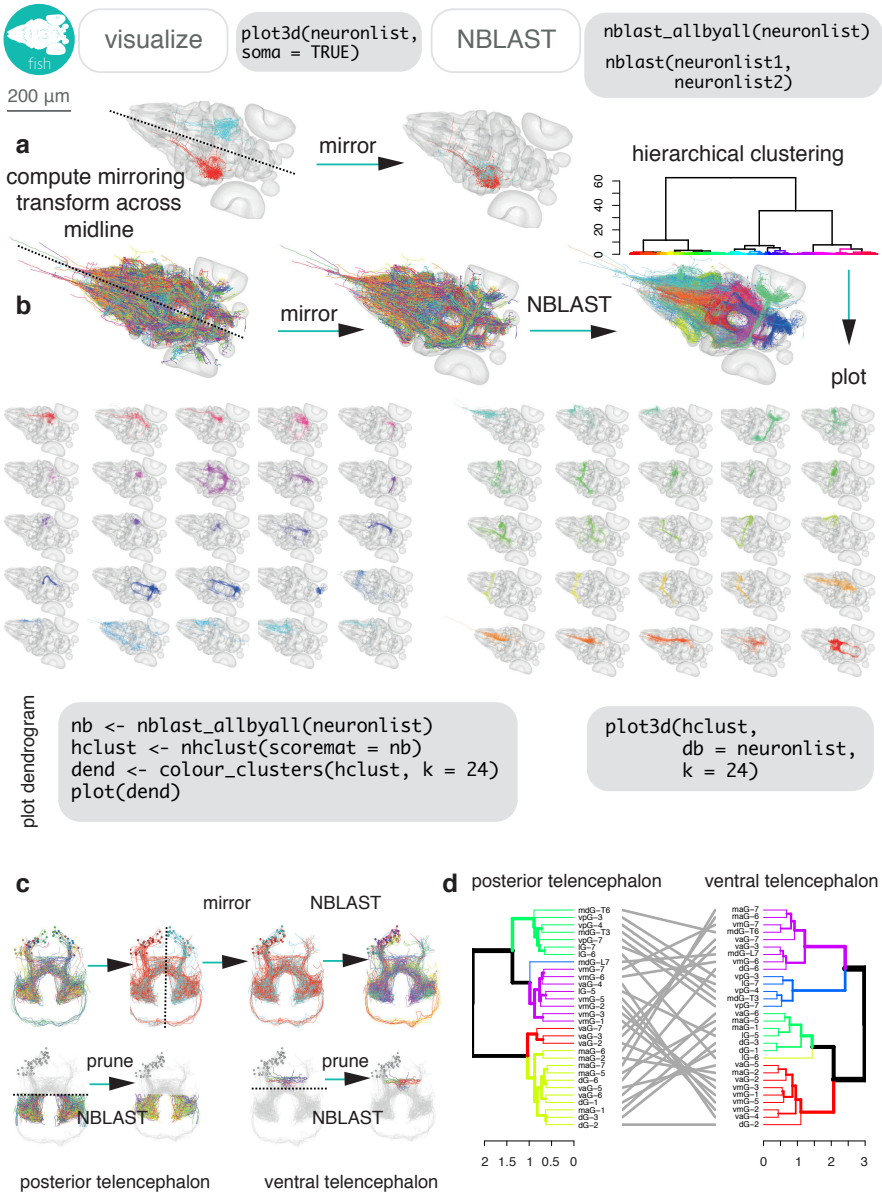
Supplementary Figure 2: Schematic representation of the data structure behind `neuron` objects and `neuronlist` objects. Objects of class `neuronlist` are essentially lists of `neuron` objects, representing one or more neurons, with some attached metadata.



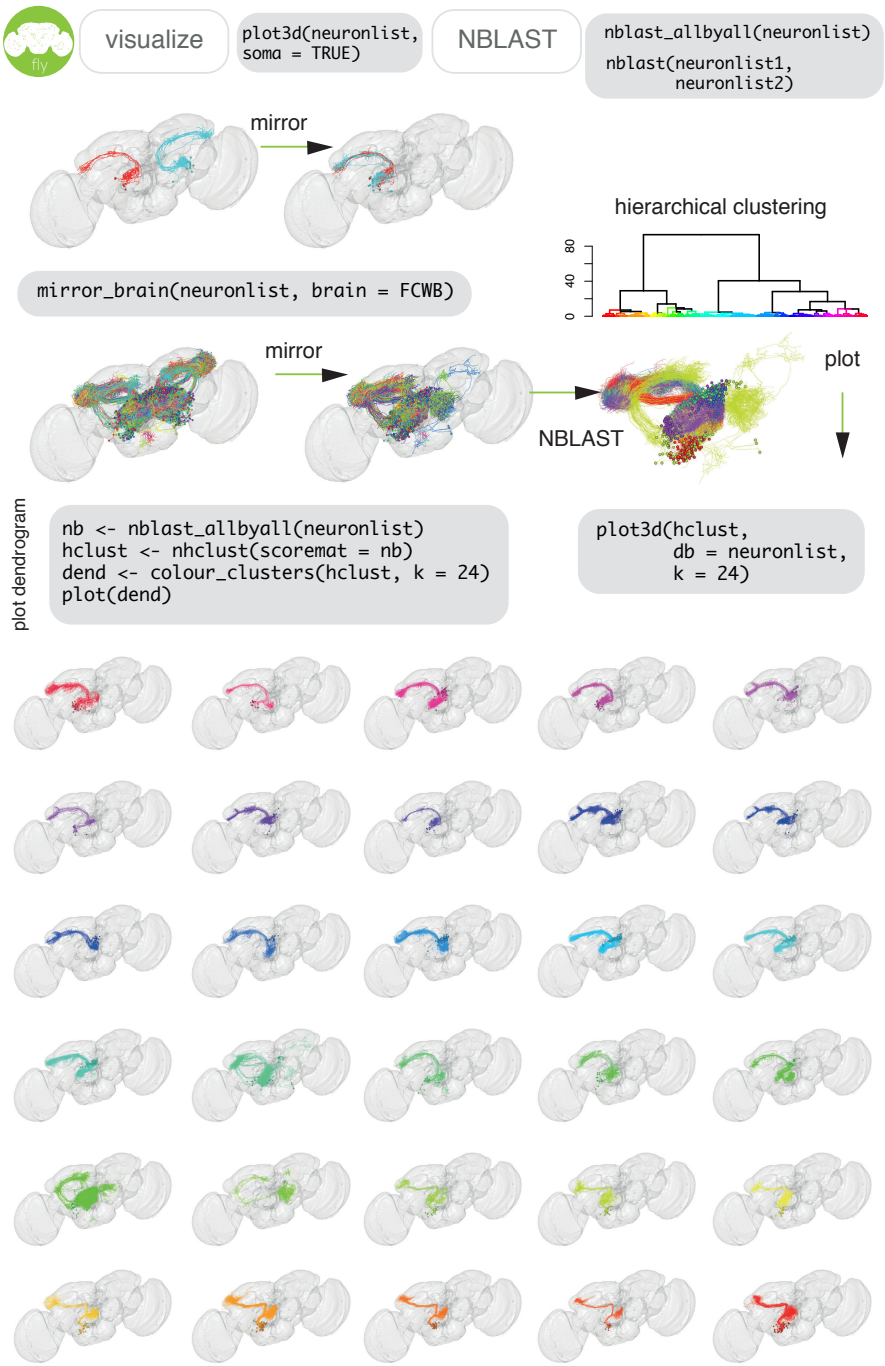
Supplementary Figure 3: a We used neuromorphr to obtain information about 40599 mammalian neurons from NeuroMorpho.org, in the brain region 'neocortex', from 193 different laboratories. However, it is quite difficult to compare morphologies acquired from different laboratories, that have used different imaging pipelines, neuron reconstruction tools and staining protocols (Farhoodi et al., 2019). We can use neuromorphr to access neuron metadata on NeuroMorpho.org, and choose only those neurons from the same laboratory (Bob Jacobs'), that have been obtained using a Golgi stain and the reconstruction software NeuroLucida, and that are classed as a principal cell (Anderson et al., 2009; Jacobs et al., 2015, 1997; Travis et al., 2005; Jacobs et al., 2016; Anderson et al., 2010; Jacobs et al., 2001, 2018, 2011; Reyes et al., 2016), giving us 3174 neurons. Branch points per micron cable can then be plotted for 24 different species' cortical principal neurons using ggplot2. **b** Inset, a Sholl analysis (Sholl, 1953) obtains the number of intersections the neurons' morphologies make with radii of increasing size centered at their roots. Main, the mean intersections for a subset of the species' cortical principal cells is shown. **c** Right, using synaptic information, it is possible to use a graph theoretic approach which divides the neuron at the point of maximum 'flow' having 'drawn' a path between each input synapse and each output synapse that pass through every node on the skeleton (Schneider-Mizell et al., 2016). This helps divide a neuron into its dendrites, axon, intervening cable (maximum flow, the primary dendrite) and its cell body fiber (no flow). In insects, the cell body lies outside the neuropil and is connected to its arbour by a single fiber. Left, axon-dendrite split shown for exemplary neuron using `seesplit3d`. **d** The distribution of pre- and post-synapses along three neurons arbors, split by axon and dendrite, then further by Strahler order, then further by presence (backbone) or absence (twigs) of microtubule. Olfactory related neurons (gold), associative memory related neurons (purple) and unknown inputs (grey). In this case, neuron#1 and neuron#2 are of the same cell type and look similar (PD2a1), but neuron#3 differs, indeed it belongs to a different cell type (PD2b1) (Dolan et al., 2018).



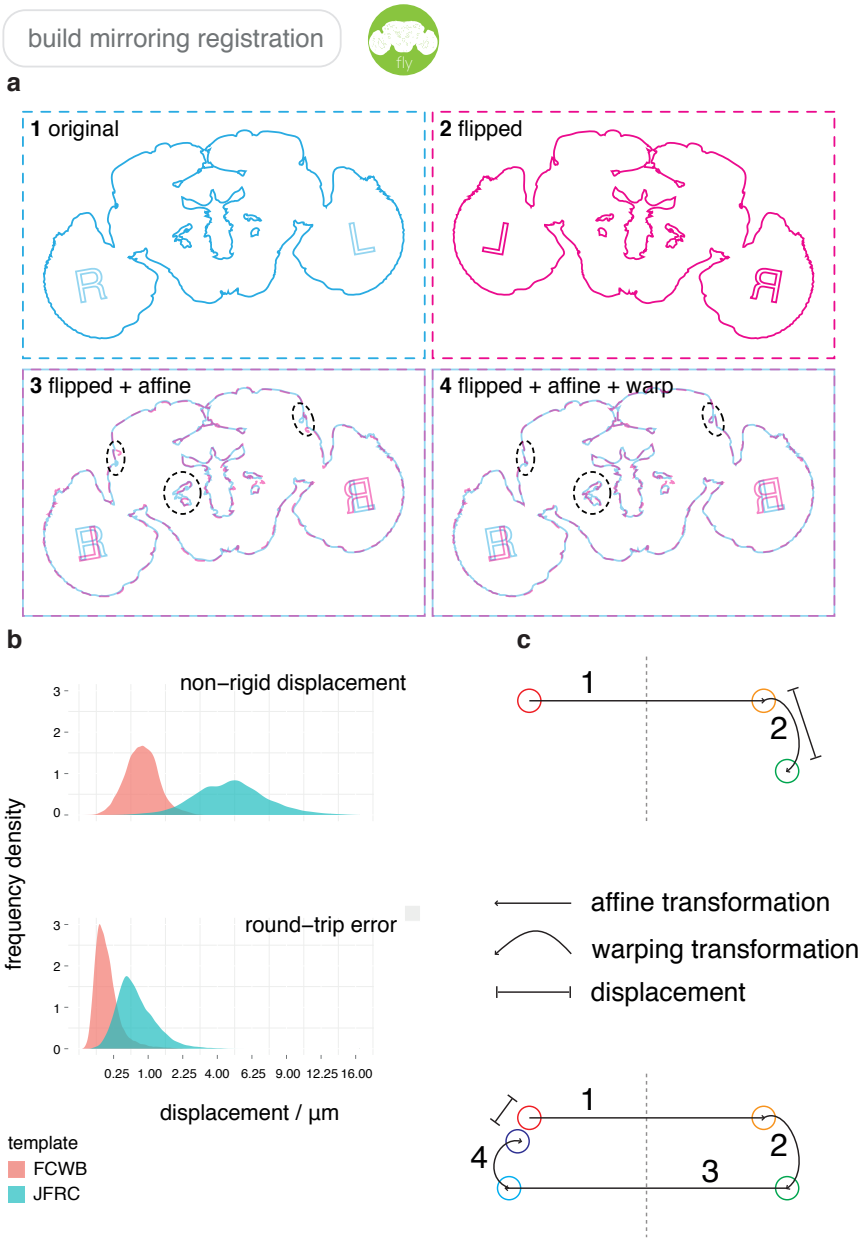
Supplementary Figure 4: a The basic connectivity scheme of the lateral horn, a second order olfactory centre in insects. Olfactory projection neurons (in orange) connect to lateral horn neurons (in cyan) (Frechter et al., 2019). **b** To create anatomically meaningful continuous voxels for the lateral horn, rather than random contiguous partitions of our standard neuropil space (Ito et al., 2014), we first removed the highest Strahler order branch (`assign_strahler`) from projection neuron axons' so that their sub-branches could be clustered into 25 separate groups (`nblast`). For each cluster, a 3D weighted kernel density estimate was generated based on 3D points (`xyzmatrix`) extracted from clustered sub-branches, using the R package ks (Duong, 2007). Points spaced on neurites at 1 μm intervals (`resample`) and weighted as 1 / total number of points in the cluster, so that supervoxels could be directly compared. **c** An 'inclusion' score for each neuron was calculated for each supervoxel by summing the density estimate for each point in the chosen arbor, again sampled at 1 μm intervals, and normalized by the total number of points in each arbor. An atlas of the lateral horn, colouring supervoxels by the modality/valence of their strongest input neurons. **d** A 'projection' was calculated between each lateral horn voxel and each lateral horn output voxel based on the number of neuronal cell types that have processes in both and the density of this arborisation. An atlas of the lateral horn output regions can then be made, colouring supervoxels by the modality/valence of their strongest input lateral horn supervoxels.



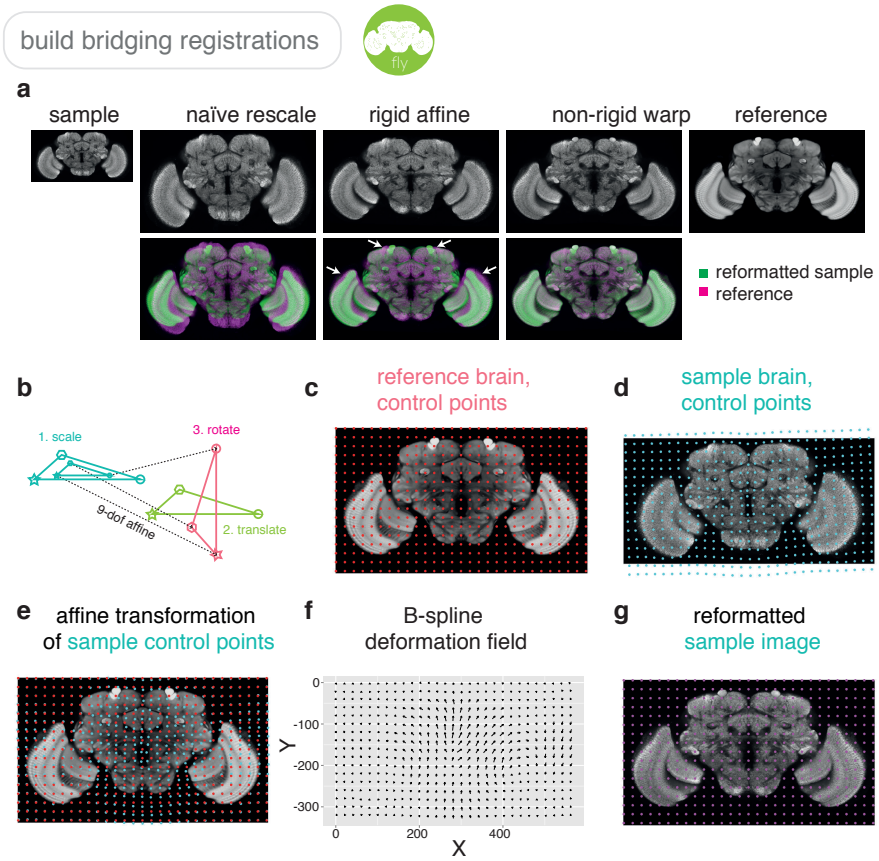
Supplementary Figure 5: a Light-level neurons registered to a standard brain for the larval zebrafish (Kunst et al., 2019) can be read into R using fishatlas, and all transformed onto the right hemisphere using the mirroring registration generated by Kunst et al. **b** They can then be NBLAST-ed to discover new cell types. 25 NBLAST clusters are shown, generic exemplary code shown in monospaced font. **c** Upper, shows this process for a small group of neurons from the telencephalon. Mirroring neurons to the same side produces cohesive clusters. Lower, neurons can be cut up (pruned) relative to user defined coordinates to, in this case, divide them by telencephalon region. NBLAST can be run separately on these two groups, and the results were compared using a tanglegram to see whether neurons that cluster together in the posterior telencephalon also cluster together in the ventral telencephalon.



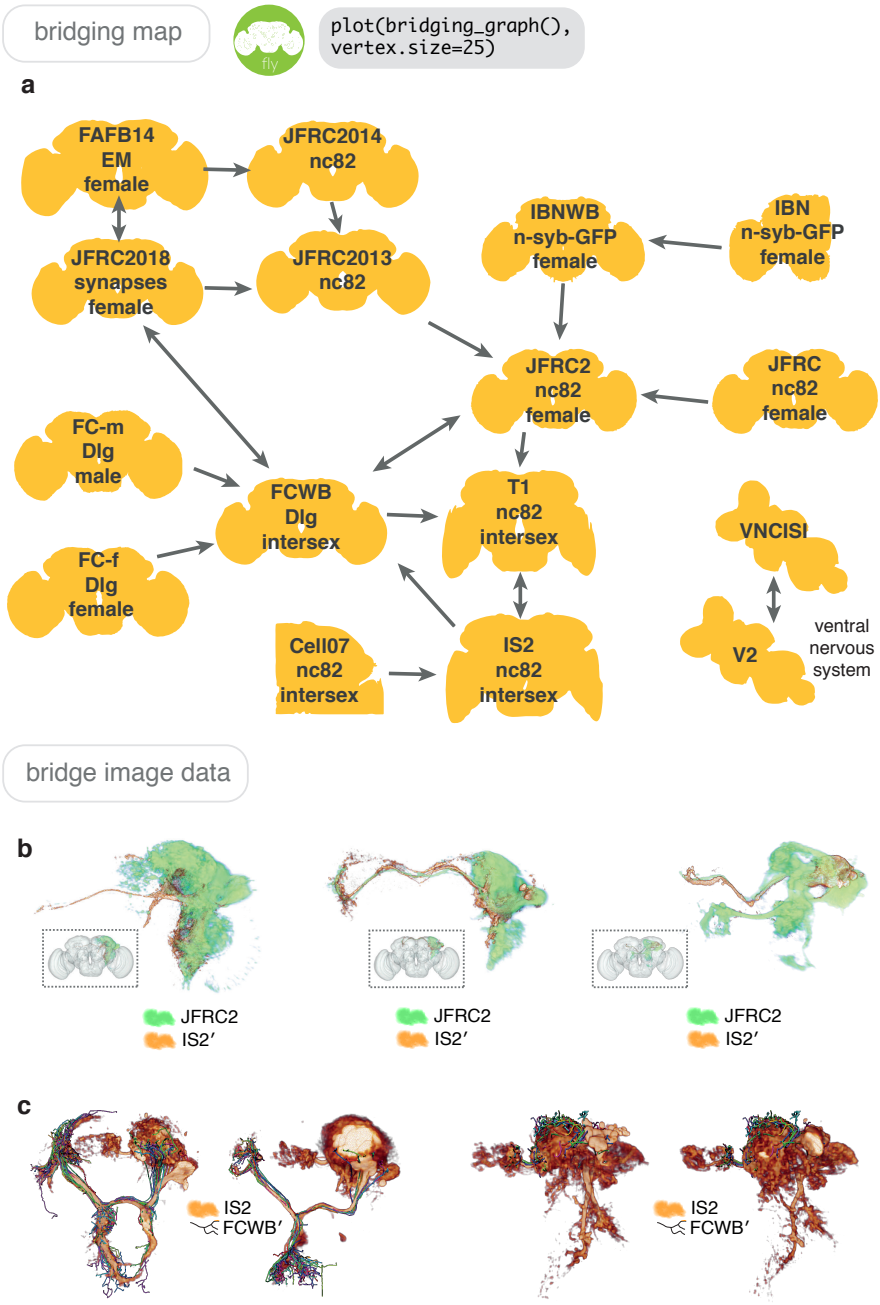
Supplementary Figure 6: The same basic process can be done for FlyCircuit neurons (Chiang et al., 2011), here subsetted (`in_volume`) to those neurons that have arbor in both the antennal lobe and lateral horn (i.e. olfactory projection neurons).



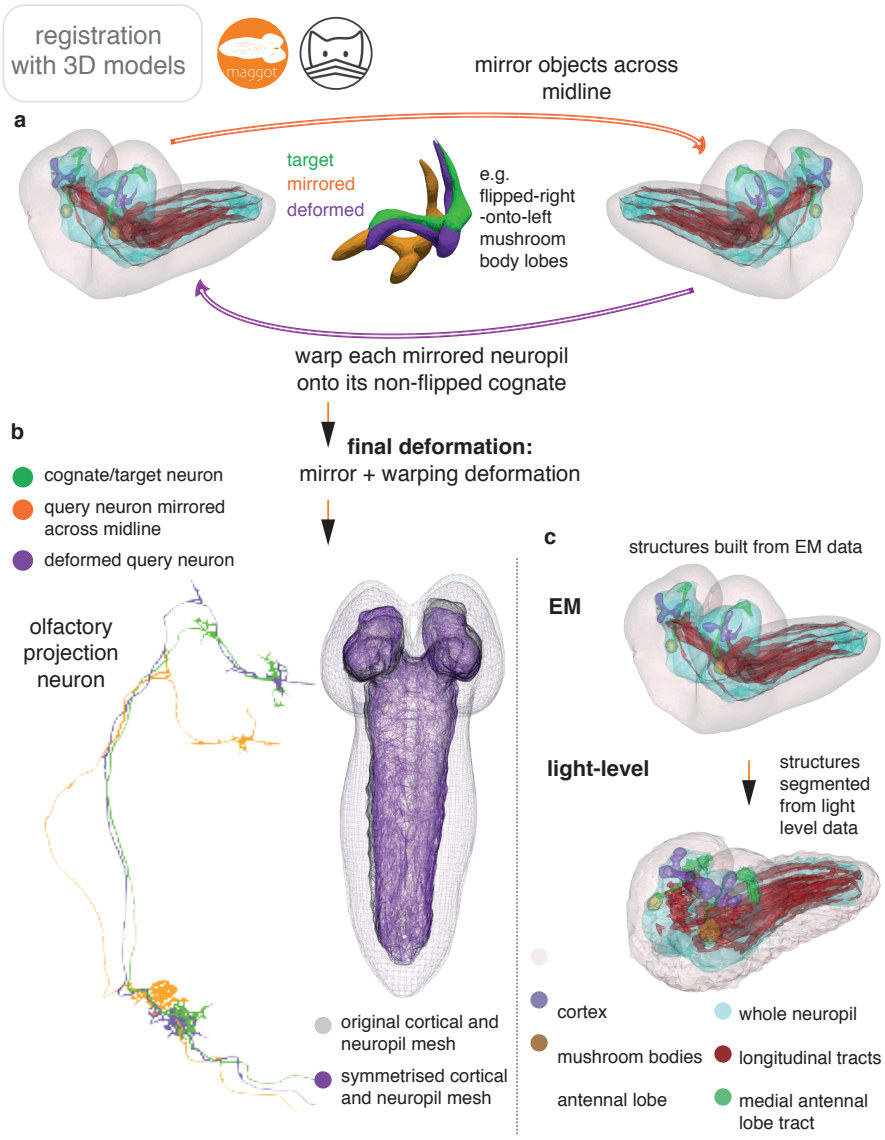
Supplementary Figure 7: a The full process undergone by an image during a mirroring registration. (1) Original image. (2) Flipped 180° around the medio-lateral axis. (3) Affinely transformed. (4) Non-rigidly warped. **b** Heatmaps of deformation magnitude fields for mirroring FlyCircuit and FlyLight reference brains. **c** Distribution of deformation displacements for both brains, in a single mirroring operation and a full round-trip. Illustrations show the transformations that points undergo.



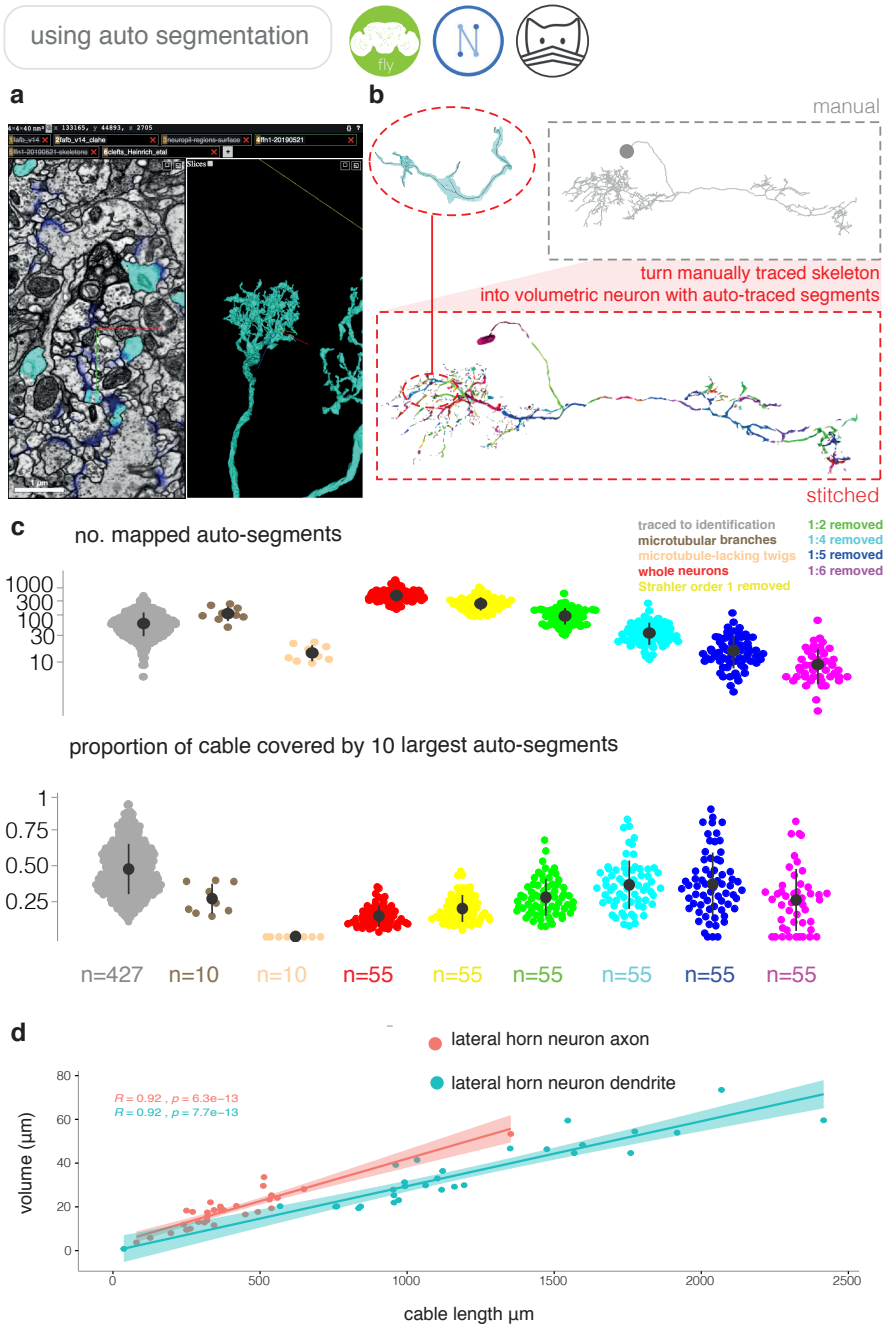
Supplementary Figure 8: a Increasing levels of registration complexity give increasingly good registration results. **b** Four composite transformations of a 12-degree-of-freedom affine transformation. **c** Regularly spaced grid of control points (red dots) in the reference brain. **d** Control points (cyan dots) in the sample image. **e** Sample brain control points affinely transformed into image space of reference brain, along with original control points. **f** Deformation field interpolated using B-splines. **g** Reformatted sample image.



Supplementary Figure 9: a A sub-section of the bridging registrations available through nat.flybrains, via which a neuroanatomical entity in any brainspace can eventually be placed into any other brainspace, by chaining bridging registrations. Arrows indicate the existence and direction of a bridging registration. Registrations are numerically invertible and can be concatenated, meaning that the graph can be traversed in all directions. **b** Fru+ neuroblast clones (orange) transformed from IS2 space into JFRC2 space of elav neuroblast clones (green). **c** Sexually dimorphic Fru+ neuroblast clone (male on left, female on right) along with traced neurons from FlyCircuit.



Supplementary Figure 10: a A left-right registration for the lop-sided nascent L1 larval connectome (Ohyama et al., 2015). Neuroanatomical models created from CATMAID data (upper) and 644 manually left-right paired neurons can be used as a basis to generate a single deformation of L1 space using Deformetrica (via deformetricar), that describes a left-right mapping in a single deformation of ambient space. **b** Symmetrisation (purple) of the lop-sided cortex and neuropil in the L1 EM dataset. **c** Registration of segmented light-level volumes onto the EM dataset. Immunohistochemical staining of dissected L1 central nervous system was needed to make 3D models for coarse neuroanatomical volumes, which could be matched to those in the L1 EM; namely, the pattern of longitudinal ventral nervous system axon tracts and the mushroom bodies, visible in a dFasciculin-II-GFP protein trap line, the laddered ventral nervous system pattern and olfactory tracts visible in a GH146-lexA line, and a Discs large-1/dN-Cadherin stain to visualize the neuropil and cortex of the brain.



Supplementary Figure 11: a A NeuroGlancer window open on a web-browser, showing an example of an automatic reconstruction. **b** Automatically reconstructed segments can be mapped onto extant manual tracing in FAFB14 using our R package fafbseg, allowing neurons to be easily volumetrically reconstructed. **c** Upper, automatically traced segments are mapped onto 55 manually reconstructed neurons from the lateral horn of *D. melanogaster* in the FAFB14 dataset (Bates & Schlegel, in prep.), broken down by Strahler order, 10 by microtubular backbone versus small calibre twigs, to simulate different levels of 'completeness' and 427 traced to 'identification' (major branches only). Lower, proportion of cable at different levels of pruning, that are covered by the 10 largest auto-segmented fragments for each neuron. The traced to identification category comprises neurons reconstructed by expert annotators sufficiently for them to be identifiable in light level data, using the pipeline shown in Figure 8. **d** Correlation between cable length and volume for axons and dendrites (Schneider-Mizell et al., 2016) for a selection of central brain neurons.