

SOFTWARE

Open Access



# NeuronBridge: an intuitive web application for neuronal morphology search across large data sets

Jody Clements<sup>1†</sup>, Cristian Goina<sup>1†</sup>, Philip M. Hubbard<sup>1†</sup>, Takashi Kawase<sup>1†</sup>, Donald J. Olbris<sup>1†</sup>, Hideo Otsuna<sup>1†</sup>, Robert Svirskas<sup>1†</sup> and Konrad Rokicki<sup>1\*</sup>

<sup>†</sup>Jody Clements, Cristian Goina, Philip M. Hubbard, Takashi Kawase, Donald J. Olbris, Hideo Otsuna and Robert Svirskas: These authors are listed in alphabetical order.

\*Correspondence: rokickik@janelia.hhmi.org

<sup>1</sup> Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, USA

## Abstract

**Background:** Neuroscience research in *Drosophila* is benefiting from large-scale connectomics efforts using electron microscopy (EM) to reveal all the neurons in a brain and their connections. To exploit this knowledge base, researchers relate a connectome's structure to neuronal function, often by studying individual neuron cell types. Vast libraries of fly driver lines expressing fluorescent reporter genes in sets of neurons have been created and imaged using confocal light microscopy (LM), enabling the targeting of neurons for experimentation. However, creating a fly line for driving gene expression within a single neuron found in an EM connectome remains a challenge, as it typically requires identifying a pair of driver lines where only the neuron of interest is expressed in both. This task and other emerging scientific workflows require finding similar neurons across large data sets imaged using different modalities.

**Results:** Here, we present NeuronBridge, a web application for easily and rapidly finding putative morphological matches between large data sets of neurons imaged using different modalities. We describe the functionality and construction of the NeuronBridge service, including its user-friendly graphical user interface (GUI), extensible data model, serverless cloud architecture, and massively parallel image search engine.

**Conclusions:** NeuronBridge fills a critical gap in the *Drosophila* research workflow and is used by hundreds of neuroscience researchers around the world. We offer our software code, open APIs, and processed data sets for integration and reuse, and provide the application as a service at <http://neuronbridge.janelia.org>.

**Keywords:** *Drosophila*, Connectomics, Cloud computing, Serverless, Web service, Database, Open data API

## Background

Advances in serial-section electron microscopy technologies have enabled the generation of comprehensive nervous systems maps, known as connectomes, of the *Drosophila melanogaster* central nervous system [1–7]. These connectomes are derived from segmented EM images which are reconstructed to describe the precise structure of the neurons in a brain, as well as the connections (i.e., synapses) between them. Neuroscientists



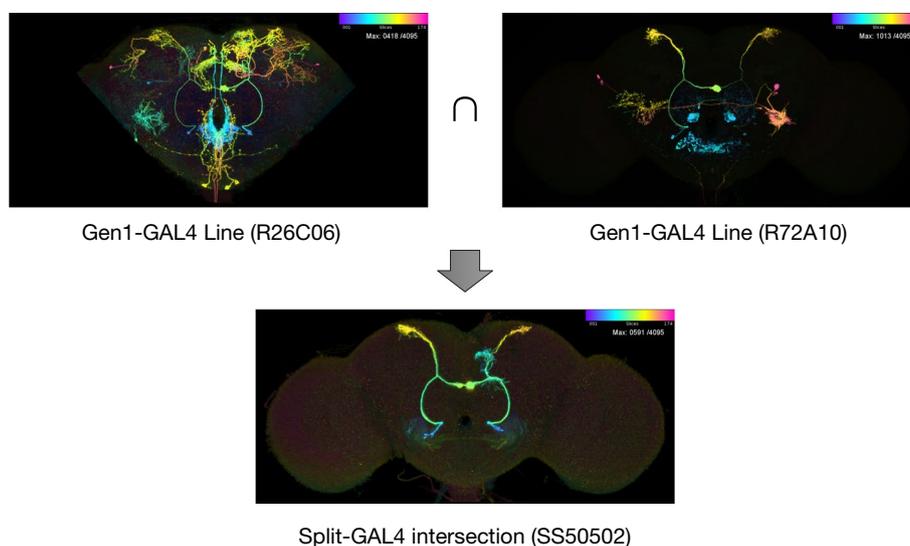
© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

use connectomic information to relate neuronal connectivity structure with how neurons function *in vivo*. Here we focus on two large connectomes produced by Janelia's FlyEM project team: the FlyEM Hemibrain data set [4], which covers more than half of a female adult fly brain, and the FlyEM Male Adult Nerve Cord (MANC) data set [6]. Together these data sets cover a large portion of the fly nervous system.

Importantly, neuron morphology is highly conserved across different specimens in *Drosophila* [8], enabling the study of selected neuronal circuits across individuals. Researchers target one or more carefully selected neurons for visualization, neuronal activity measurement, genetic modification, ablation, or stimulation, and combine these tools with animal behavior studies. Janelia's FlyLight project team produced a large driver line library known as the Generation 1 (Gen1) GAL4 Collection [9], which provides the starting point for targeting specific neurons. Each driver line expresses the GAL4 protein under control of a genomic enhancer fragment. When crossed to another line with a UAS (Upstream Activating Sequence) reporter gene, GAL4 binds to UAS and activates reporter expression in defined neurons [10].

To provide greater specificity and target individual neurons, the two-component Split-GAL4 method [11, 12] is used to produce an expression pattern reflecting only the common elements in the patterns of two driver lines; i.e., an intersection between two expression patterns is performed. Ideally, this results in a new driver line that targets a single neuron or cell type (Fig. 1). Subsequently, the GAL4/UAS system [10] allows further genetic modifications of the targeted neurons (fluorescence, ablation, etc.).

Creating an effective Split-GAL4 line requires the careful identification of two parent Gen1 GAL4 lines which have expression in the same neuron of interest with no other expression in common. Finding such candidate lines in a collection of thousands of



**Fig. 1** Genetic tools for *Drosophila*. Two Gen1 driver lines with broad expression are crossed using Split-GAL4 to create a new driver line with the intersection of the parent lines' expression pattern. Images are color depth maximum intensity projections (CDM), where the blue color indicates the anterior of the brain and red the posterior. Upper section shows one of three MCFO channels as a CDM. Lower section shows the full Split-GAL4 expression pattern from driver line SS50502 [24]. Original 3D images are available at <https://gen1mcf0.janelia.org> and <https://splitgal4.janelia.org>

images is a labor-intensive process and benefits from a computationally assisted workflow [13, 14]. The FlyLight Gen1 GAL4 driver collection has been characterized recently [15] with confocal imaging of neurons isolated by stochastic multi-color labeling (MCFO; [16]). This allows for the visualization and identification of individual neurons and provides a basis for a computational workflow for Split-GAL4 creation [15].

Increasing numbers of neuron and cell-type specific Split-GAL4 driver lines are being characterized [17–36]. FlyLight recently released a large collection of cell-type-specific Split-GAL4 driver lines [37]. As the number of published Split-GAL4 driver lines increases, a second workflow will become important: searching these lines to find ready-made drivers for neurons identified in a connectome.

In other common workflows, researchers traversing circuits in a connectome want to verify cell type identity by referencing known driver lines [26, 33, 38], or by finding additional neurons of the same cell type with a similar morphology to a known neuron of interest [28]. In yet another workflow, a researcher starts with a neuron imaged in LM and looks for that neuron in a connectome [36].

All of the above workflows require identifying similar neuron morphology across large data sets of EM and LM images. Registration of neurons to a common reference alignment space enables spatial matching across biological samples and modalities [39]. Two current solutions to the computational problem of matching neurons across registered imaging modalities are Color Depth MIP Search (abbreviated as CDM Search; [13]) and PatchPerPixMatch (PPPM; [14]). CDM Search represents location in the projection dimension as color in a maximum intensity projection (MIP) and thus enables efficient 3D structure comparison through simple 2D similarity between overlapping pixels. PPPM leverages deep-learning segmentation (PPP; [40]) of the LM images and an algorithm based on NBLAST [41] to find the best matching neuron fragments (from LM) that fit a reconstructed neuron skeleton (from EM) in 3D space. A comparison of CDM Search and PPPM is available elsewhere [15].

However, the CDM and PPPM algorithms are not readily accessible to experimentalists. CDM Search was initially implemented as a Fiji plugin [42] for executing searches locally. Therefore, it requires the user to download large data sets and wait for each search to execute using local compute resources. PPPM provides precomputed matches for a subset of FlyLight Gen1 MCFO images against the FlyEM Hemibrain and MANC, but the algorithm is expensive to run (~3 h per LM image on a single GPU) and is not easily usable for custom searching with user data. To make these algorithms more accessible, we built a web application that experimentalists can use to rapidly identify and view similar neurons in published EM and LM data sets, or to perform searches on their own data. The only comparable software that we know of is the NBLAST functionality in Virtual Fly Brain ([43]), which allows users to find similar neurons across the FlyCircuit data set. However, it is currently limited to data sets where the neurons have been traced as skeletons, so it does not address the important use case of searching large amounts of LM data containing neurons which are difficult to trace automatically.

## Implementation

To address the problem of finding similar neurons across large multi-modal data sets, we developed NeuronBridge, an easy-to-use web application. It provides instant access to neuron morphology matches for published EM and LM data sets, as well as rapid custom searching against those data sets. Our initial implementation was based on the CDM Search algorithm and PPPM results were added later. Implementing CDM Search in a publicly available web application with interactive response times presented several major challenges: (a) the CDM Search tools were only available within Fiji [42] and lacked programmatic APIs for reuse; (b) when centralized and scaled to hundreds of users, CDM searching becomes expensive to run *de novo* for each query; (c) CDM Search relies on fast local disk access and multithreading, which are expensive resources to leave idle at the scale needed to support unpredictable usage access patterns; and (d) users seeking to search using their own data first need to align it to the standard template and generate aligned CDM images for their neurons.

We first extracted the CDM Search algorithm from its Fiji plugin, refactored it into a Java library ([Software Availability](#)), and made it available for programmatic reuse. To enable large-scale computation on a high performance computing (HPC) cluster, we used the CDM Search library to create an Apache Spark application for running the CDM Search algorithm in a distributed manner.

Second, we generated collections of CDM images, henceforth referred to as “CDM libraries”, from *Janelia’s* published EM and LM data sets including the FlyEM Hemibrain [4], FlyEM MANC [6], FlyLight Gen1 MCFO and Annotator Gen1 MCFO [15], and the Split-GAL4 driver line collection (Fig. 2). To populate the CDM libraries, we generated aligned CDM images ([Image Alignment](#)) from FlyEM’s published EM neuron skeletons and from FlyLight’s published 3D image volumes. We then used our Apache Spark CDM Search implementation to precompute neuron matches between EM and LM libraries ([Neuron Morphology Matching](#)). Each EM CDM was compared with each LM CDM, a total of over 33 billion image-to-image comparisons for FlyEM Hemibrain, and 10 billion comparisons for the FlyEM MANC. Matches were recorded in both directions, to allow bi-directional searching. This step obviated the need to run *de novo* searches for matches

CDM Library	Anatomical Area	Samples	Neuron Images	Searchable Files	Input Size (GB)	Data Size (GB)
FlyEM Hemibrain	Brain	1	30,790	44,477	24,000	3.90
FlyEM MANC	VNC	1	23,271	29,326	30,000	3.40
FlyLight Gen1 MCFO	Brain	46,031	136,381	349,405	89,822	85.50
	VNC	32,025	98,072	292,146	108,597	66.60
FlyLight Annotator Gen1 MCFO	Brain	34,724	101,215	355,179	37,214	61.90
	VNC	3,652	10,765	28,919	2,962	7.00
FlyLight Split-GAL4 Drivers	Brain	16,229	24,560	47,753	11,573	10.50
	VNC	7,841	10,373	10,721	6,219	1.90
TOTAL		234,976	435,427	1,158,468	310,387	240.70

**Fig. 2** Data available in NeuronBridge. CDM libraries are loaded into NeuronBridge with precomputed results and are available for custom searching. *CDM Library* indicates the CDM library being described. *Anatomical Area* shows the region of the fly central nervous system covered by the library. *Samples* shows the number of biological samples which form the original data set. *Neuron Images* indicates the total number of individual images containing neurons derived from these samples. *Searchable Files* indicates the number of CDM images in the library that can be searched. *Input Size (GB)* indicates the total uncompressed size of the original data set. *Data Size (GB)* indicates the total size of the CDM images used for searching

between published data sets, mitigating the ongoing costs of running the shared service, and allowing us to make the service freely available to the community.

Third, we developed a data model ([Data Model](#)) for representing matches that we define as putatively similar neuronal morphologies identified with respect to two images that are registered to the same template. Our data model allows matches from CDM Search, PPPM, and potentially any future algorithms to be represented with the same entity classes, as well as allowing for an extensible set of imaging modalities, making it easier to add new data in the future while maintaining a consistent user experience.

Then we imported the published PPPM results [14] for the FlyEM Hemibrain, FlyEM MANC, and FlyLight Gen1 MCFO data sets into our data model alongside the CDM Search results. The PPPM results feature additional images for visualizing matches, including overlays of EM skeletons on top of masked LM signals, which we were able to import into our data model as new file types.

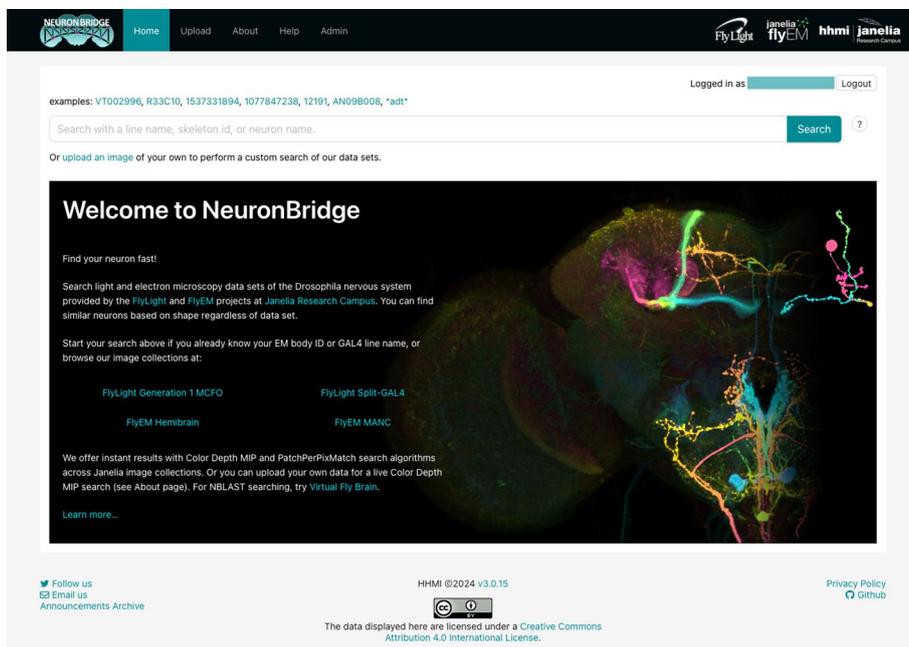
Next, we leveraged the Amazon Web Services (AWS) Open Data program to make all of the images and match metadata available on AWS S3 as a robust public data API ([Open Data API](#)). AWS S3 is a serverless object storage service with high reliability, low latency, access control, and tight integration with other AWS services. Using a serverless solution allowed us to keep costs low compared to a traditional relational database deployment. We also developed a simple Python module ([Software Availability](#)) on top of our API, to enable ad hoc data analysis through the use of Python libraries, and to serve as a client reference implementation.

Next, we created an online custom search service for user-uploaded data ([Software Architecture](#)) that (1) registers uploaded confocal images to a reference brain using the Computational Morphometry Toolkit (CMTK, [44]) running on AWS Batch, (2) generates a CDM image for each channel of the uploaded stack, (3) allows the user to select their neurons of interest with an interactive browser-based image masking tool, and (4) runs a massively parallel CDM Search to find neuron matches in available CDM libraries. It's also possible for the user to directly upload their own aligned CDM image to use as the search target. Importantly, the custom search service is entirely serverless, which minimizes costs by not idling compute resources in the cloud. To make this possible we developed a burst-parallel compute engine ([Software Architecture](#)) to achieve on-demand, 3000-fold parallelism using the AWS Lambda service, which allows a search of the EM CDM libraries to complete in seconds and a search of the LM CDM libraries to complete in under two minutes.

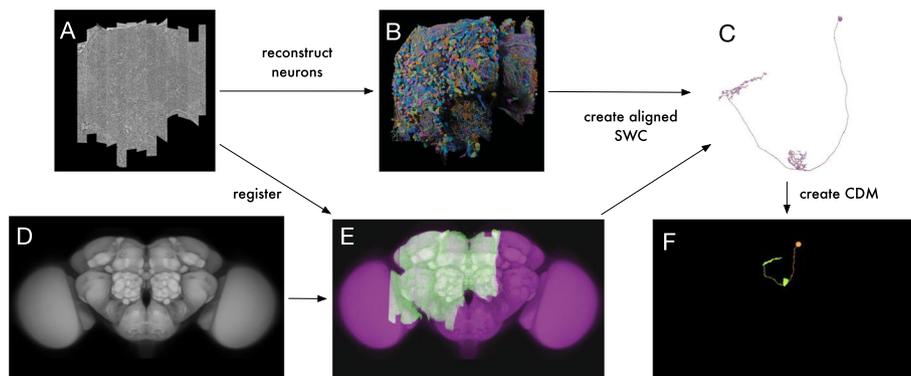
Finally, we built a user-friendly, single-page application (SPA) using the React framework ([User Experience](#)), enabling end-users to browse precomputed matches and run searches with their own data. The graphical user interface (GUI) of the application (Fig. 3) was designed to be simple and intuitive, focusing solely on the task of neuron matching, providing contextual help information, and including usage examples where appropriate.

### Image alignment

For the FlyEM Hemibrain data set we relied on the published registration to the JRC2018 *Drosophila* female brain template [39] (Fig. 4A, D, E). For the FlyEM MANC, we used the published registration to the JRC2018 *Drosophila* male VNC template [6]. For the



**Fig. 3** NeuronBridge home page. The search box enables quick access to EM and LM CDM images using common identifiers such as neuron identifiers and driver line names



**Fig. 4** Pipeline for EM CDM generation. **A** FlyEM Hemibrain stitched data set; screenshot from Neuroglancer. **B** FlyEM Hemibrain reconstruction by Janelia FlyEM and Google. **C** Example neuron reconstruction (ID 1537331894) from the FlyEM Hemibrain; screenshot from neuPrint. **D** JRC2018 Unisex brain template. **E** Registration of FlyEM Hemibrain to the JRC2018 Unisex template. **F** CDM of the neuron shown in **C**

FlyLight data sets, we used CMTK [44] to register all LM images (Fig. 5A, B) to the sex-appropriate JRC2018 template [39]. In all cases we subsequently used a bridging transform to move the data into the alignment space of the JRC2018 Unisex template, where neuron matching could be performed across all samples, irrespective of sex.

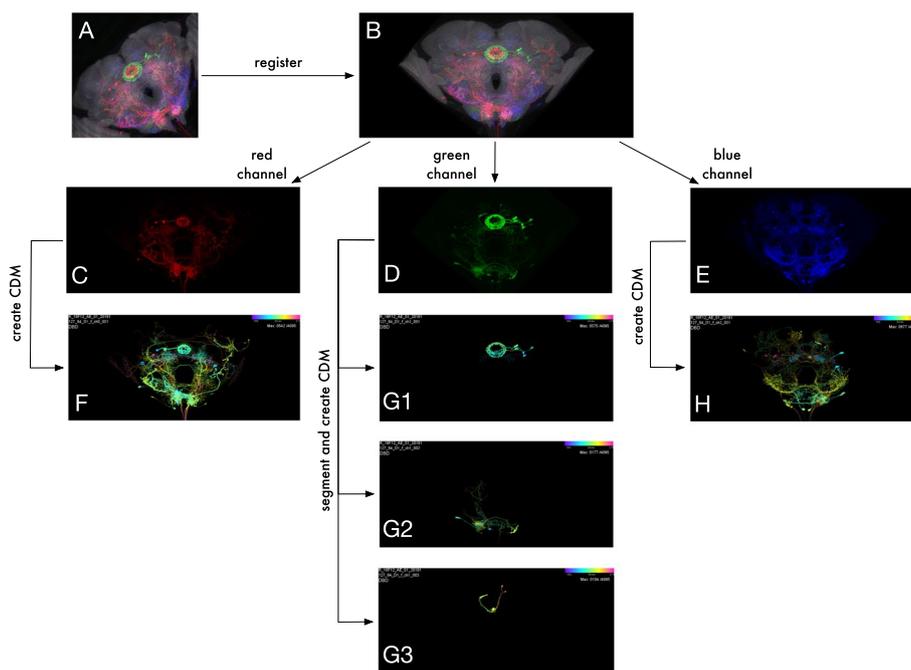
For online custom searching, the alignment code is packaged as a Docker container and run using AWS Batch. For better performance, online searches skip the alignment to sex-specific templates and align directly to the JRC2018 Unisex template.

### Neuron morphology matching

The CDM Search algorithm [13] was designed for interactive use and required several modifications for reliable batch execution on large data sets [15]. These changes will be described fully in a future paper [45] but we provide a brief outline here.

First, we generated CDM libraries from all of the input data sets. To create the EM CDM libraries we started with neuron skeletons (Fig. 4C), which are approximations of neuronal morphology computed from the proofread segmentation of the EM image data [4]. Compared to the full neuron reconstructions, using the skeletons allowed us to control the diameter of the neurons for better CDM search results. We transformed the skeletons into the aligned space (Image Alignment) and generated a CDM image for each neuron (Fig. 4F).

While EM segmented neurons represent individual cells, biological GAL4 expressions typically exhibit bilateral symmetry [4]. To enhance matching between the FlyEM Hemibrain and LM data, we generated additional artificial CDM images. This was done by mirroring and combining the image with the original CDM for any neurons crossing the midline. These ‘flipped’ neurons expand the CDM image set, resulting in higher matching scores for bilateral neurons that express GAL4 on both sides of the brain in LM images. Nonetheless, due to the stochastic nature of MCFO, many MCFO samples express GAL4 on only one side of the brain. This type of asymmetric expression tends to yield higher matching scores with non-flipped, ‘original’ EM single neurons. To maximize the probability of matches in both cases, we compared the EM-LM matching



**Fig. 5** Pipeline for LM CDM generation. **A** MIP of an example LM MCFO image (slide code 20181127\_64\_D1) of a brain sample from the R16F12 driver line. **B** LM image registration to JRC2018 Unisex template. **C–E** Aligned MIPs of individual channels. **F–H** CDM images for each channel. **G1–G3** Segmentation of green channel using modified DSLT yields three voxel sets for the green channel. Each voxel set is converted into a separate CDM

scores between ‘flipped’ and ‘original’ versions for the same EM neuron and LM sample, selecting the higher score as the final matching result in NeuronBridge.

To create LM CDM libraries, we started with registered LM confocal images and generated multiple CDMs for each channel of every aligned LM image (Fig. 5C–H). LM images may contain tissue background signal which interferes with CDM Search matching. Also, denser fluorescence labeling in MCFO samples can produce occlusions of neurons. To address these issues, we performed a 3D neuronal segmentation of each channel using an algorithm based on Direction Selective Local Thresholding (DSLTL, [46]), which resulted in a handful of voxel sets per channel (Fig. 5G). We eliminated smaller “junk” voxel sets based on manually chosen Fiji shape descriptor thresholds and generated CDM images of the remaining voxel subsets.

When running CDM Search, we used the images in the EM CDM library as search targets and compared each one with every image in the LM CDM libraries. We limited the number of LM matches to a maximum of 300 driver lines. The CDM matching was initially conducted only from the EM to the LM direction. Nevertheless, NeuronBridge enables users to conduct searches from LM to EM because we recorded the matching results for each LM sample across all EM neurons, based on the matching score.

To maximize the likelihood of the best matches rising to the top of the search results, we created a uni-directional shape-matching algorithm that is run after CDM Search to re-sort the results. The algorithm penalizes EM CDM pixels that are not within approximately  $\pm 12$  microns of LM CDM pixels in the XY dimensions. It also penalizes more than 40 microns of unmatched EM CDM pixels in the Z dimension, within the  $\pm 12$  micron XY region. It does not similarly penalize LM CDM pixels because LM samples typically express GAL4 in multiple neurons simultaneously, leading to expression patterns that do not align with single EM neurons. Ideally, if an EM neuron is contained within an LM sample, the LM image should cover all corresponding EM branches depicted in the LM data. However, if an incorrect EM neuron is matched with the LM sample, this will result in EM branches that are not represented in the LM sample, penalizing the match score. The final score for each match was computed as a ratio of positive matching (i.e., number of matched pixels) to negative matching (i.e., the penalties).

### User experience

NeuronBridge is an end-user web application designed for accessibility and ease-of-use by wet-lab biologists. Using any web browser, a user can log into the service’s website with their Google account credentials or create a separate NeuronBridge account. The user is presented with a simple search interface for looking up their neuron or GAL4 driver line of interest (Fig. 3). Searching for a neuron identifier returns CDM representations of EM neuron reconstructions, in the JRC2018 Unisex alignment space. Searching for a driver line returns CDM representations of that driver line. These initial search results are displayed in a tabular, paginated format, allowing the user to select an image of interest to begin neuron matching. Each result has buttons corresponding to the types of match algorithms that were used to compare it to the other modality, typically both CDM Search and PPPM. These buttons let the user view putative neuron matches with other images (LM matches for EM targets, EM matches for LM targets).

**CDM Input Image**

Line Name: MB543B  
 Slide Code: 20140407\_33\_G4  
 Channel: 1  
 Library: FlyLight Split-GAL4 Drivers  
 Gender: Female

Anatomical Area: Brain  
 Mounting Protocol: DPX PBS Mounting  
 Alignment Space: JRC2018\_Unisex\_20x\_HR  
 Magnification: 20x  
 Links: [FlyLight Split-GAL4](#), [Virtual Fly Brain](#)

**EM Matches** ?

Filters / Sorting | Download (30+) | Clear Selected | Grid

1-30 of 1524 matches < 1 2 3 ... 51 > 30 / page

hemibrain:v1.2.1:456833888 (Score: 47203)	hemibrain:v1.2.1:5812981264 (Score: 43750)	hemibrain:v1.2.1:449209971 (Score: 42715)	hemibrain:v1.2.1:640963528 (Score: 42308)
hemibrain:v1.2.1:5812981543 (Score: 38148)	hemibrain:v1.2.1:703387200 (Score: 37654)	hemibrain:v1.2.1:519581562 (Score: 37438)	hemibrain:v1.2.1:449205588 (Score: 37057)

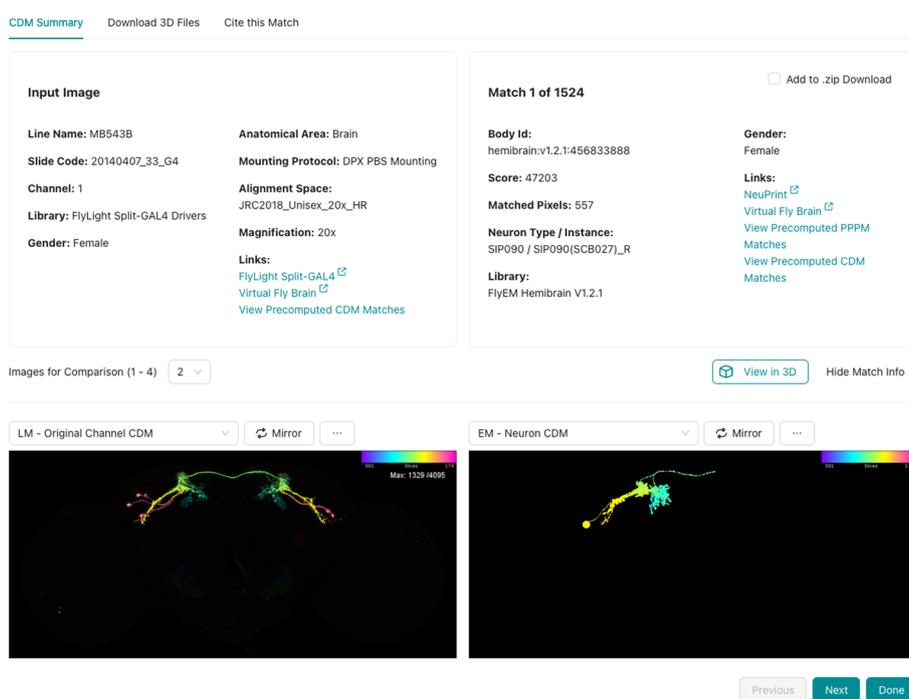
1-30 of 1524 matches < 1 2 3 ... 51 > 30 / page

**Fig. 6** Browsing EM matches for an LM image. Paginated EM matches are displayed for the CDM input image shown at the top. The results can be filtered and sorted, or shown as a list. Checkboxes allow users to select promising matches for download and further analysis. Screenshot was truncated to only show the first two rows of results

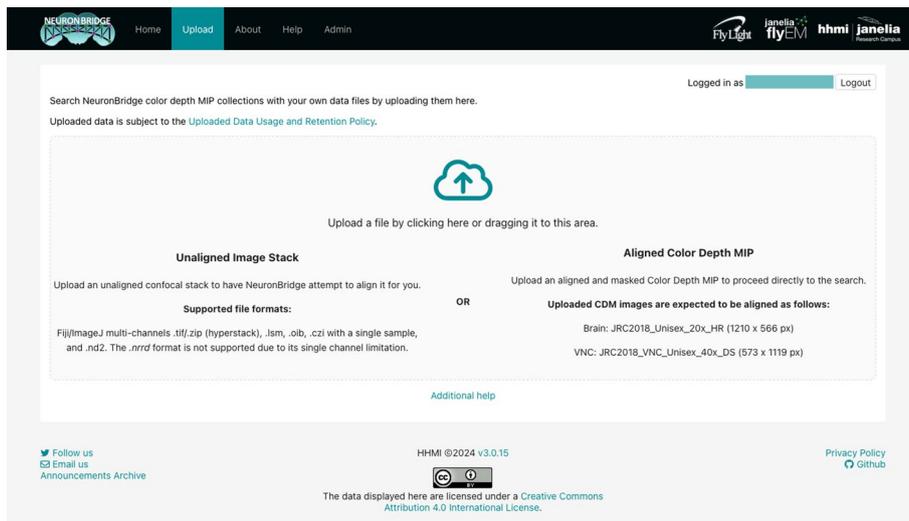
After selecting a target image and an algorithm, precomputed neuron matches are immediately displayed as a list or a grid (Fig. 6) and paginated for rapid browsing. They can be filtered by various criteria, such as the CDM library that the image belongs to, or the driver line name. When searching for LM matches, all of the matches for a single driver line are grouped and sorted together by the score of the highest ranking representative of that line. These results can be filtered to display a maximum number of representatives per driver line. A setting of 1 is useful when the user is interested in finding candidate lines for creating a Split-GAL4 driver line.

Clicking one of the matches brings up a detail interface (Fig. 7) which allows the user to compare the match to the target using a mouse cursor synchronized across multiple images. Various accessory images can be displayed to better characterize the match and the driver line or connectome where it originated. Matches can be selected for download, from either the overview page or the detail page. The user can download all the metadata for the selected matches as a CSV file, and all of the matching images as a ZIP file.

In addition to the precomputed match results between published images, NeuronBridge supports custom CDM Search against any of the EM and LM image libraries. Users begin a custom search by uploading their own image (Fig. 8), which can be an unaligned image stack in a variety of standard microscopy image formats (TIFF, ZIP, LSM, OIB, CZI, ND2), or an aligned CDM mask in the JRC2018 Unisex alignment space. NeuronBridge attempts to align uploaded images and generates a CDM for each channel. This process may take several minutes, and the user may wait for the alignment task to complete, or come back later to find their aligned CDMs. Next,

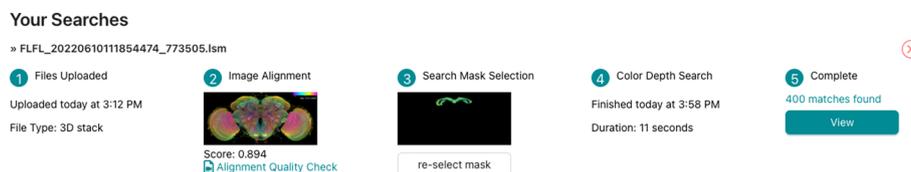


**Fig. 7** Details for a single EM/LM match. LM input image shown on the left and the EM match on the right. This view is customizable such that the user can choose how many images to display and select which representation to show in each screen location



**Fig. 8** Uploading data for custom search. The upload GUI allows the user to drag and drop a file to begin the custom search workflow. Input data requirements are described in detail

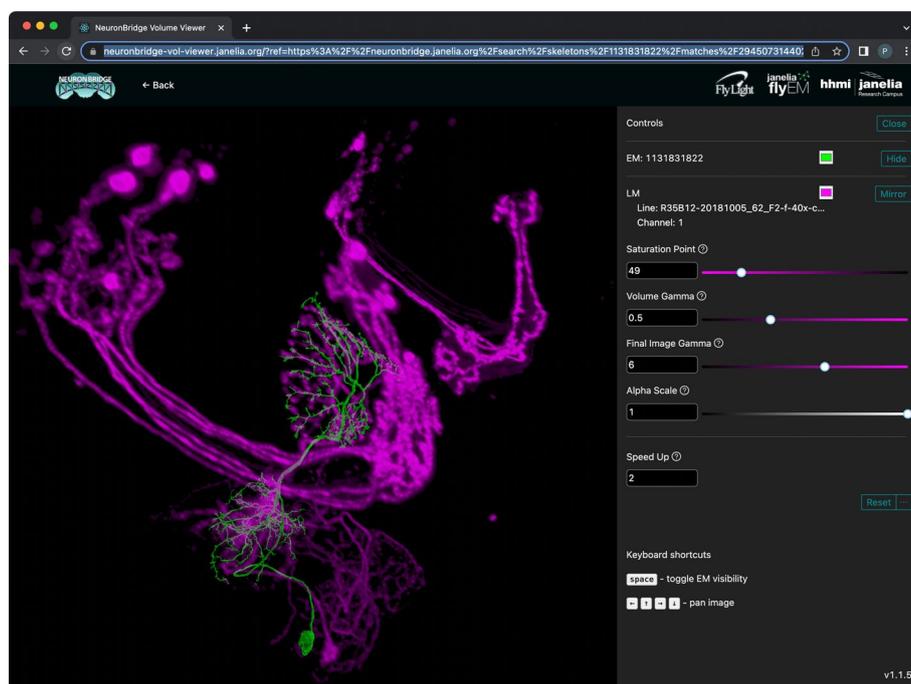
the user is asked to choose a channel and create a search mask. The user may select the CDM libraries to search, and optionally specify other search parameters before invoking the search. A custom search typically takes less than a minute to process and



**Fig. 9** Custom search progress steps. The workflow can be branched by re-selecting the mask so that a different search can be run with the same input stack. This saves a considerable amount of time by reusing the alignment result

progress is displayed in a step-wise workflow GUI (Fig. 9). Results are presented identically to how precomputed results are displayed.

The web GUI was developed using the React framework in order to make the GUI components reusable and more responsive than a traditional web application, by never fully reloading the page as one moves around the site. Instead, data is loaded via asynchronous HTTP requests, and the GUI is updated as data becomes available. By using AWS AppSync, we also extended this idea to asynchronous updates using Web Sockets, so that long-running back-end operations such as alignments can be monitored without polling. The single-page application (SPA) paradigm also allowed us to serve the website as static content, with caching via the AWS CloudFront Content Distribution Network (CDN) that speeds up initial loading.

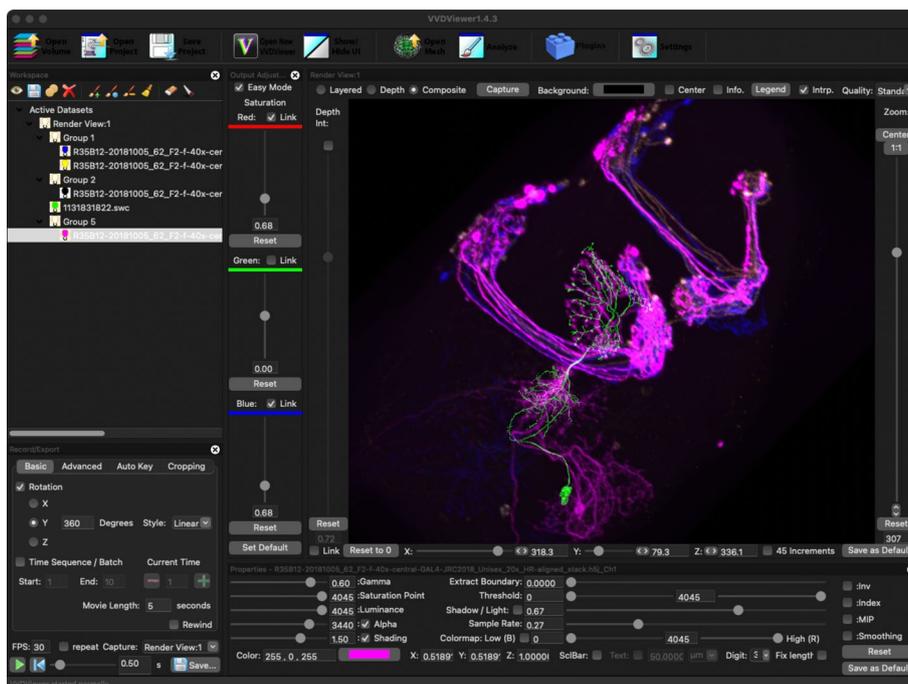


**Fig. 10** 3D visualization of matches in the browser. One channel of the LM image volume is rendered in magenta, and the matching EM neuron skeleton is rendered in green. Controls on the right affect qualities of the rendering like the relative transparency in the LM image volume

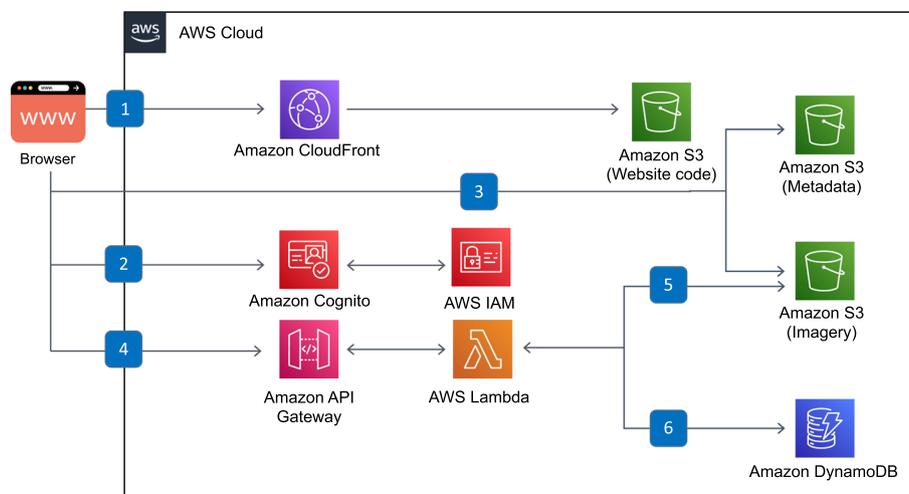
### 3D visualization

NeuronBridge provides two ways a user can manually verify a putative match by viewing the EM neuron skeletons together with LM images from arbitrary viewpoints in 3D. The simplest approach is to click the “View in 3D” button in the detail interface for a search result (Fig. 7). Doing so launches a new browser page that loads the EM neuron skeleton and LM image and renders them in 3D with interactive camera controls (Fig. 10, see also the video in Additional file 1). For LM images we use direct volume rendering, which samples the image volume along rays from the camera, applies a transfer function and lighting model to compute color and opacity at each sample, and composites the results to produce pixels. The EM neuron skeleton is rendered as an opaque surface and blended into the volume rendering with proper occlusion based on its depth map. The renderer uses WebGL2 and achieves interactive performance on modern desktop and laptop computers. The URL for the renderer page is continually updated to encode the details of the current view, so a live rendering can be shared with a collaborator by sharing the URL.

A more sophisticated rendering is available by clicking the “Download 3D Files” tab and loading the data into a desktop rendering application like VVDViewer [47] (Fig. 11, see also the video in Additional file 2). VVDViewer renders with slightly higher quality because it uses more bits per voxel of the LM data (12 instead of 8). It also adds several



**Fig. 11** 3D visualization of matches in VVDViewer. All three channels of the LM image volume are rendered, in red, green and blue, and the EM neuron skeleton is rendered in aquamarine



**Fig. 12** Serverless back-end for precomputed result browsing. The user’s browser communicates with (1) the CloudFront Content Distribution Network (CDN) to retrieve the web client code which is cached from AWS S3. (2) Authentication is done using federated identity providers through Cognito, supporting both email-based accounts and Google accounts, and delegating to IAM for authorization. Precomputed results are (3) statically loaded from public buckets on S3. Dynamic features use (4) an API Gateway endpoint to run a Lambda function, for example to (5) create a ZIP archive of multiple files for download or (6) run prefix searches on DynamoDB

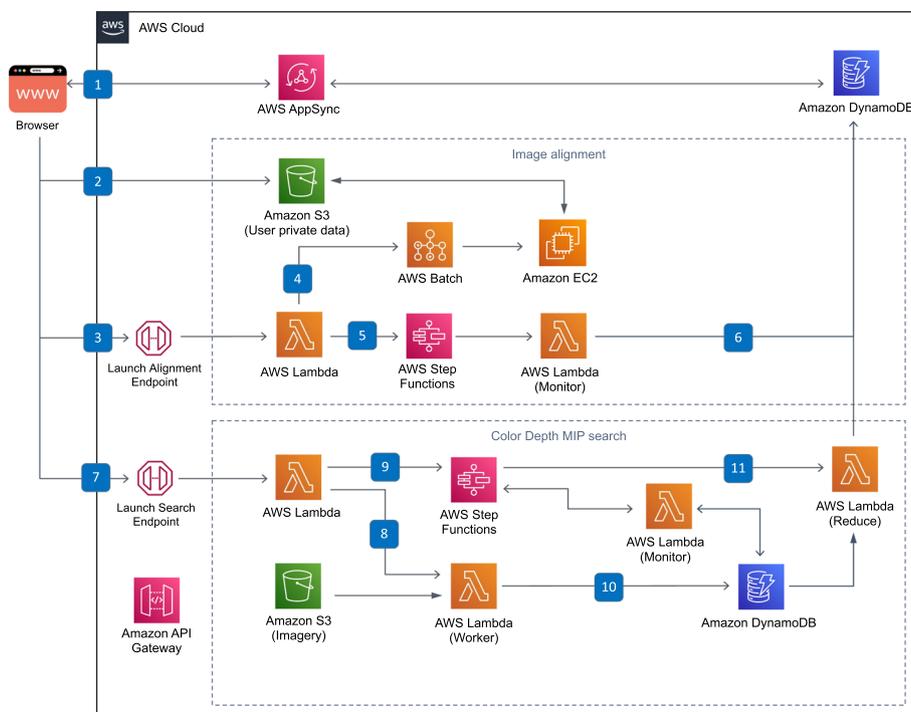
capabilities beyond the browser-based renderer. It can render multiple channels of an LM image simultaneously, better representing the data used by the PPPM matching algorithm. It can render multiple EM skeletons simultaneously, which can be useful for verifying fragmented neurons or for comparing several match results at once. It can also generate animated videos, showing changes to the camera position and other viewing parameters.

**Software architecture**

We used fully-managed, serverless AWS services to build the back-end (Fig. 12) of the web application, including Cognito identity federation for authentication, IAM for authorization, Lambda for compute, S3 for storage, and DynamoDB as the database. We used only serverless services to keep operational costs low while allowing us to focus on application logic instead of spending time on server administration.

Custom search (Fig. 13) consists of two services: image alignment which runs on AWS Batch, and distributed CDM Search which runs on AWS Lambda. The aligner runs on AWS Batch as a Docker container and is monitored by a Step Function which notifies the client when the alignment is completed. To enable efficient searching across large image data sets while keeping costs low, we implemented a burst-parallel [48] search engine built on Lambda and Step Functions [49]. To reduce cold starts when executing searches on AWS Lambda, we also translated the core CDM Search algorithm from Java to JavaScript.

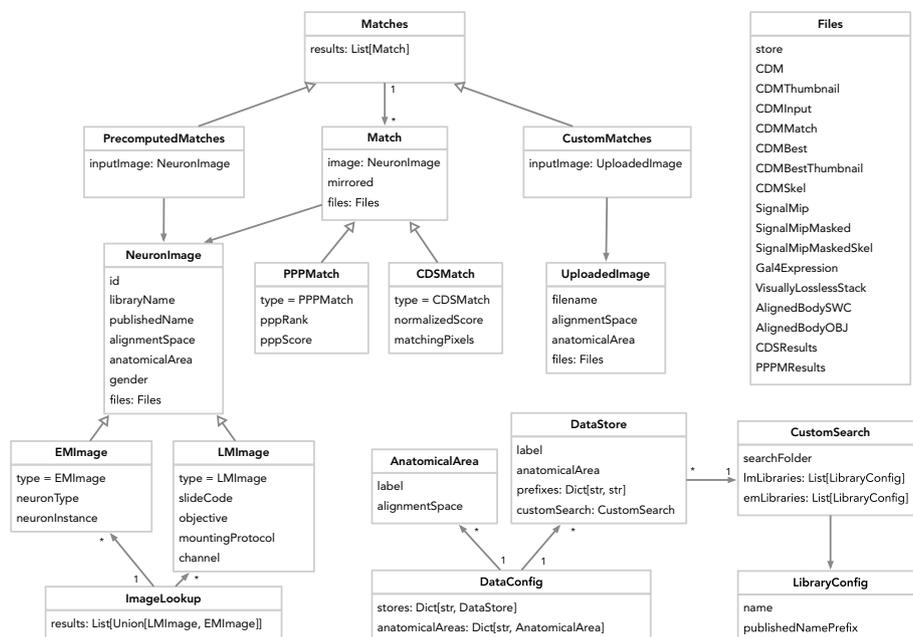
The serverless model is a particularly good fit for the burst-parallel implementation for several reasons. First, massive parallelism is required to quickly search large image libraries, and the ideal level of parallelism is prohibitively expensive in a traditional server-based architecture. The 3000-fold parallelism we achieved with AWS Lambda



**Fig. 13** Serverless back-end for custom search. The user initiates a search by uploading an image, and three messages are sent to the back-end: (1) Using AppSync, a new record is created in DynamoDB to track the search workflow, (2) the image is uploaded to a private user folder on S3, and (3) the client starts the alignment process by invoking a Lambda function. (4) A Batch job is created, which allocates an EC2 node and runs the alignment Docker container, fetching the uploaded file from S3, and depositing the aligned CDM images back on S3. (5) A Step Function periodically runs a Lambda to monitor the progress of the Batch job. (6) Once the alignment job is finished, it updates the workflow state in DynamoDB, which notifies the user via AppSync. Back in the browser, the user is asked to create a search mask from one of the resulting CDM images, and this mask is uploaded to S3. (7) The client calls a second endpoint to begin the search, which calls a Lambda that (8) recursively starts the burst-parallel search workers as well as (9) the monitoring Step Function state machine. (10) When a worker finishes, it records its result in a DynamoDB table. Once the monitoring Lambda detects that all workers are finished, (11) it runs the reducer which produces a search result that is updated in the DynamoDB workflow table, notifying the user again through AppSync that the search is complete

for a few cents would require one hundred 30-core dedicated servers in a traditional data center deployment. Second, the service has unpredictable usage and can sit idle for extended periods, while at other times has bursts of high activity. This type of usage pattern is directly addressed by a serverless model that can scale to zero. The costs to run this service are low because serverless services bill for compute/storage that is utilized, similar in concept to a traditional HPC compute cluster. In addition, the serverless approach reduced the amount of ongoing maintenance by eliminating server administration. Finally, using serverless Lambda functions allowed us to fully leverage the concurrent throughput of the S3 buckets hosting the images.

The software is reliably deployed through the use of the Serverless Framework, which generates low-level AWS CloudFormation instructions for deployment. Reproducibility and isolation of the deployment is important to being able to run multiple versions of the system at the same time. In a single AWS account, we run a separate instance for each developer, a validation instance for testing before production deployments, a



**Fig. 14** Data model. The DataConfig describes the metadata accessible through the API. Image lookup queries return an ImageLookup object containing neuron images. A NeuronImage has a GUID called id which can be queried for Matches

pre-publication instance for internal use, and the production instance that is publicly accessible.

### Data model

The data model (Fig. 14) is anchored by a NeuronImage, representing a set of neurons in an alignment space that can be compared with other images in that alignment space, and a set of Matches between NeuronImages, generated by either CDM Search or PPPM. Image metadata is denormalized wherever an image is referenced, to allow clients to fetch a single JSON file instead of making multiple requests. The DataConfig provides summary information, such as the list of possible anatomical areas, as well as providing constants which allow long, common values to be interpolated into JSON values as shorter keys, reducing disk space usage and transfer time. Importantly, this model is trivially extensible with new matching algorithms and new imaging modalities.

All data served by NeuronBridge is available for programmatic access via open data APIs (Open Data API) allowing for reuse and integration. These APIs have already enabled the creation of several third-party applications including the Neuronbridger R API (<http://natverse.org/neuronbridger>) and FlyBrainLab NeuroNLP [50]. The serverless back-end automatically scales to meet additional load from these API users and imposes rate limits to avoid impacts on overall service performance.

## Results and discussion

NeuronBridge currently serves over 500 million precomputed matches between 6773 driver lines and 54,061 EM neuron skeletons. In addition, CDM images of these driver lines and neurons have been made available for online custom searching, comprising a data set of 1.2 million CDM images. These searchable 2D images efficiently represent signals extracted from approximately 256 TB of confocal light microscopy and 54 TB of electron microscopy data (Fig. 2) in 241 GB of searchable images, a three orders of magnitude compression which operationalizes the data for neuron searching. We additionally provide aligned EM neuron skeletons in SWC format [51], a standardized format (<https://incf.org/swc>) for neuron morphologies. We also link to aligned LM images [15] in H5J format (<https://data.janelia.org/h5j>), so that putative matches can be validated manually in 3D using an external system like VVDViewer [47], to supplement the 3D viewing supported directly in the GUI (3D Visualization).

We intentionally focused NeuronBridge on the use case of finding similar neuron morphologies, so it contains no features allowing users to search or view other types of data such as neuronal circuits, neurotransmitter information, or detailed metadata. The quality of the matches presented by NeuronBridge is bounded by the effectiveness of the underlying matching algorithms. The CDM Search and PPPM algorithms both produce useful and often complementary sets of putative matches [15]. CDM Search struggles with occlusions in the projection dimension, while PPPM has difficulty with segmentation of dense samples, which can cause false positives during search [14]. PPPM results are currently only available for a subset of the Gen1 MCFO data set and exclude the Gen1 Annotator MCFO and Split-GAL4 data sets. NeuronBridge also does not currently load reverse matches (LM→EM) for PPPM.

The architecture is scalable but limits some of the functionality in ways we anticipated. The choice of using S3 as a match database limits the queries that can be done to lookups by identifier, so users must rely on other websites for metadata-based search. Also, implementing the custom search on top of AWS Lambda limits us to a 15 min execution time per function, though that can be worked around by using a smaller batch size and launching more Lambda functions.

In the future, NeuronBridge could benefit from improved and/or new neuron matching algorithms, additional functionality for match inspection, and any other improvements that address match quality or the core workflows of match browsing, verification, and export. We also expect this resource to grow with additional data over time, including additional EM data sets (e.g. the FANC [5] and FlyWire [7] connectomes), as well as new driver line collections.

We intend for NeuronBridge to remain a very focused tool, and we have therefore intentionally constrained its architecture and implementation to solving the neuron matching use cases. We provide contextual links in cases where other web applications already provide related functionality. For access to source LM images, we link to FlyLight's anatomy websites (<https://flweb.janelia.org>, <https://gen1mcfo.janelia.org>, <https://splitgal4.janelia.org>). For EM neuron reconstructions we link to neuPrint ([52], <https://neuprint.janelia.org>). We also provide cross reference links to the Virtual Fly Brain [43] for both EM and LM results. All of these websites link back to NeuronBridge, forming a synergistic ecosystem of tools and data.

All of the software code and data underlying the NeuronBridge service is available for reuse. Up-to-date deployment instructions are maintained on each of our GitHub repositories ([Software Repositories](#)) and the entire service could be duplicated elsewhere, for example to serve matches for other data sets. The code is also highly modular so that parts of it can be reused without deploying the entire system. For example, the back-end AWS service implementation could be deployed to integrate CDM Search into another web application. The Apache Spark CDM Search could be reused in a similar manner for an on-premise search solution. The data model and accompanying Python API could be reused for providing programmatic access to neuron morphology matches, including matches from novel matching algorithms that might be developed in the future.

NeuronBridge is an open source project that we developed for the open science community. We are continuing to add new features, most often driven by additional data sets including upcoming FlyLight/FlyEM data sets and connectomes originating outside of Janelia. We also welcome open source community engagement, including code reuse, issue reporting, and code contributions through pull-requests.

We hope that the cloud-based serverless architecture of NeuronBridge, uncommon in open source scientific research software, inspires architectural decisions or code reuse in other tools and platforms for large scientific data analysis. In particular, the burst-parallel image search balances large data analysis with online query capability on interactive timescales, and could be reused to perform many other types of analysis where a large data volume must be fully traversed while a user is waiting.

## Conclusion

The NeuronBridge web application fills a gap created by recently published large data sets including the FlyEM Hemibrain and MANC connectomes and FlyLight image sets for characterizing driver lines. Researchers making use of these connectomes can now efficiently find driver lines that target their neurons of interest, as well as search these data sets to verify cell type identity and driver line expression. The software code underlying the service is modular and can be reused in different ways. In addition, the data powering the web application can be reused either by download or through integration with our Open Data APIs. We believe that these properties will make NeuronBridge an indispensable tool for *Drosophila* neuroscience research.

## Availability and requirements

**Project name** NeuronBridge

**Home page** <http://neuronbridge.janelia.org>

**Operating system(s)** Platform independent

**Programming language** Java, Javascript, Python

**Other requirements** W3C-compliant web browser

**License** BSD-3-Clause license

**Any restrictions to use by non-academics** None.

## Additional material

### Open data API

The S3 bucket containing the NeuronBridge matches forms an open REST API. Key structures are arranged such that they can be queried predictably. We use the standard JSON format and publish a schema for each document type. The schemas are versioned so that future changes are not breaking. All of the endpoints below are relative to the bucket root at `s3://janelia-neuronbridge-data-prod`

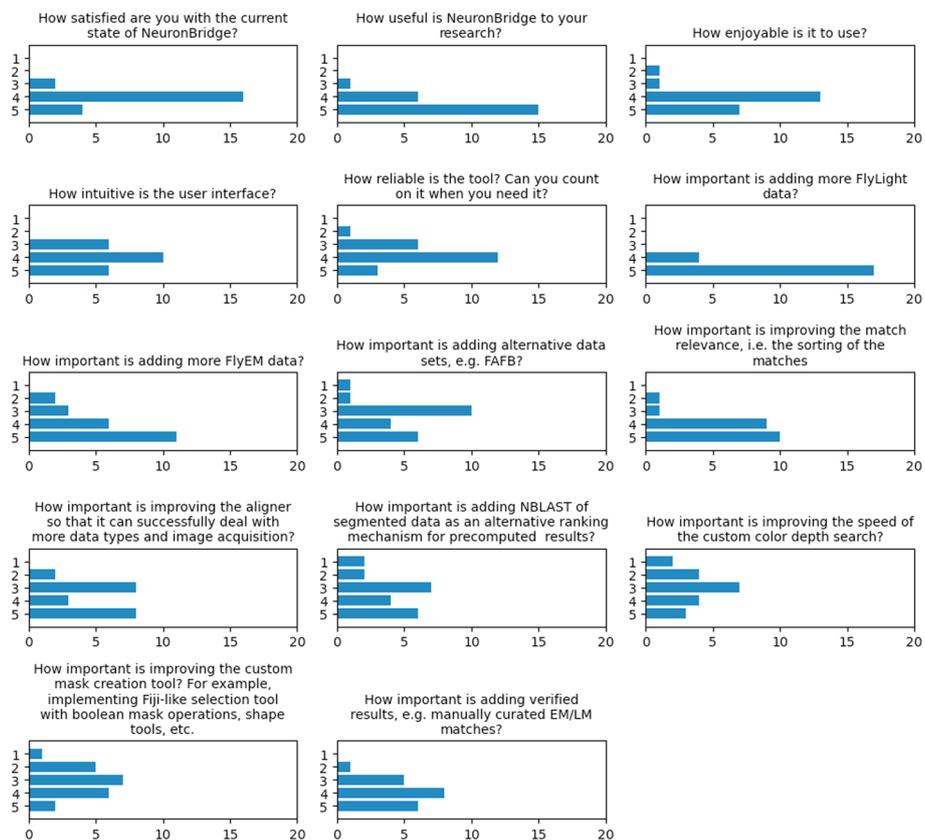
- `current.txt`

Returns the current version, to be used in subsequent API calls as `<VER>`. Older versions of the metadata are preserved so a client can choose to request older data for compatibility or other reasons.

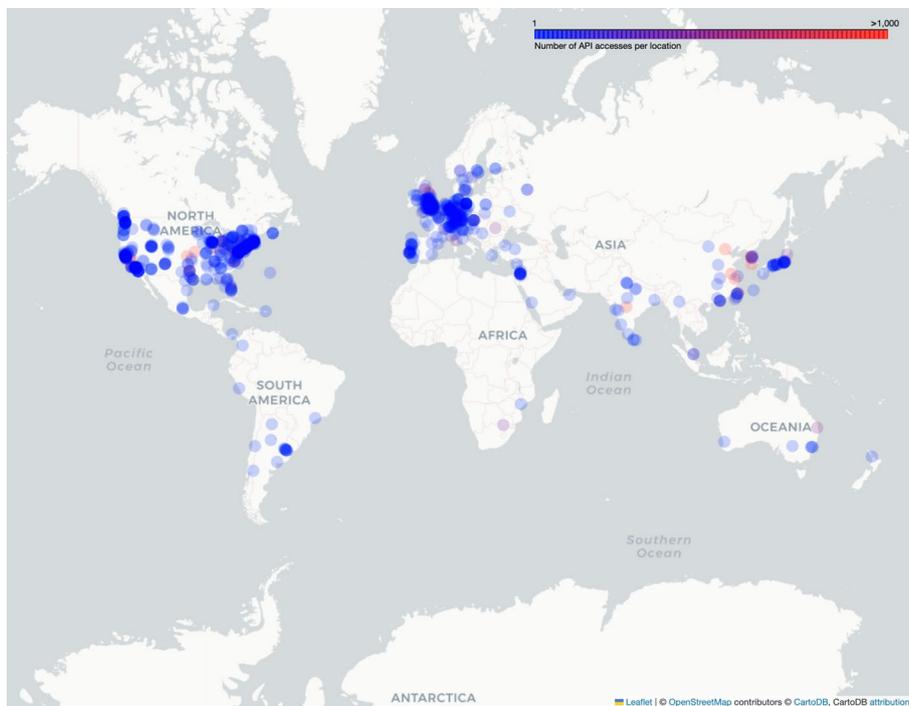
- `<VER>/schemas/`  
A prefix containing the version-specific JSON schemas for all objects in the data model.
- `<VER>/config.json`  
Returns a configuration object containing base URL prefixes and other metadata necessary for programmatic use of the following API. Follows the JSON schema defined by `DataConfig.json`.
- `<VER>/DATA_NOTES.md`  
Returns a Markdown document containing the data release notes.
- `<VER>/metadata/by_body/<body_id>.json`  
For a given body ID (i.e., the identifier for a specific EM neuron reconstruction), returns metadata such as a path to a representative image, neuron names, etc. Follows the JSON schema defined by `ImageLookup.json`.
- `<VER>/metadata/by_line/<line_id>.json`  
For a given driver line identifier, returns metadata including a list of images from the line (most lines have been characterized multiple times), representative images, and other attributes. Follows the JSON schema defined by `ImageLookup.json`.
- `<VER>/metadata/cdsresults/<image_id>.json`  
For a CDM neuron image provided by either the `by_body` or `by_line` endpoints, returns a list of matching images as computed by CDM Search. Neuron images have Globally Unique Identifiers (GUID) which persist across data versions, making it easier to link to data and results. Follows the JSON schema defined by `PrecomputedMatches.json`.
- `<VER>/metadata/pppmresults/<image_id>.json`  
For a CDM neuron image provided by either the `by_body` or `by_line` endpoints, returns a list of matching images as computed by PPPM search. Follows the JSON schema defined by `PrecomputedMatches.json`.

### User survey results

We conducted a survey to assess users' satisfaction with the service and determine which features and improvements should be prioritized. We received 22 responses to the survey during the 2021 calendar year (Fig. 15). By far, additional data was the most



**Fig. 15** User survey responses. Aggregated responses to the user survey where a value of 1 indicates “Least” and 5 indicates the “Most” response for each question



**Fig. 16** Access map. Depicts access requests to the s3://janelia-neuronbridge-data-prod and s3://janelia-neuronbridge-web-prod buckets during the two year period between 2020-05-08 and 2022-06-08

important aspect for users, which led us to prioritize the addition of matches for the FlyEM MANC and additional FlyLight data sets as they were released.

### Usage statistics

As of January 2024, there are over 1600 registered users of NeuronBridge. During the calendar year of 2023, the site received 75,256 page views, users uploaded 880 images, and ran 933 custom searches against the uploaded data. The S3 API received 328.4 million requests from 477 distinct organizations, excluding our own. Since its initial deployment, NeuronBridge has been accessed from 72 countries around the world (Fig. 16).

### Data repositories

The data is available in the following public S3 buckets:

- `s3://janelia-neuronbridge-data-prod` CDM Search and PPM match metadata in standard JSON format. The bucket structure is fully described in the Open Data API section. The JSON metadata refers to images in the buckets below.
- `s3://janelia-flylight-color-depth`  
CDM images in PNG format for display are found within each `<AlignmentSpace>/<Library>` folder. This bucket also contains the same images in TIFF format for searching under `<AlignmentSpace>/<Library>/searchable_neurons`. The TIFF files are grouped into prefix groups for better scalability when searching with thousands of AWS Lambda functions.
- `s3://janelia-ppp-match-prod`  
PPPM result images in PNG format. Imported from the published PPPM results [14].

### Software repositories

- Web client implementation
  - Source code repository: <https://github.com/JaneliaSciComp/neuronbridge>
  - Archived code at time of publication: <https://doi.org/10.5281/zenodo.10541060>
- Back-end implementation
  - Source code repository: <https://github.com/JaneliaSciComp/neuronbridge-services>
  - Archived code at time of publication: <https://doi.org/10.5281/zenodo.10541063>
- Color Depth MIP search library and Apache Spark implementation
  - Source code repository: <https://github.com/JaneliaSciComp/colormipsearch>
  - Archived code at time of publication: <https://doi.org/10.5281/zenodo.10541073>
- Scripts related to precomputed CDM results

- Source code repository: <https://github.com/JaneliaSciComp/neuronbridge-precompute>
- Archived code at time of publication: <https://doi.org/10.5281/zenodo.10541087>
- Aligner implementation
  - Source code repository: <https://github.com/JaneliaSciComp/neuronbridge-aligners>
  - Archived code at time of publication: <https://doi.org/10.5281/zenodo.10541088>
- Python API
  - Source code repository: <https://github.com/JaneliaSciComp/neuronbridge-python>
  - Archived code at time of publication: <https://doi.org/10.5281/zenodo.10541244>
- 3D visualization
  - Source code repositories: <https://github.com/JaneliaSciComp/neuronbridge-vol-viewer><https://github.com/JaneliaSciComp/web-vol-viewer><https://github.com/JaneliaSciComp/web-h5j-loader>
  - Archived code at time of publication: <https://doi.org/10.5281/zenodo.10541250><https://doi.org/10.5281/zenodo.10530403><https://doi.org/10.5281/zenodo.10541255>

#### Abbreviations

API	Application Programming Interface
AWS	Amazon Web Services
CDM	Color Depth MIP
CDN	Content Distribution Network
CMTK	Computational Morphometry Toolkit
DSLTL	Direction Selective Local Thresholding
EM	Electron microscopy
FIB-SEM	Focused Ion Beam Scanning Electron Microscope
GB	Gigabyte
GPU	Graphics Processing Unit
GUI	Graphical User Interface
H5J	File format for light microscopy images
HPC	High Performance Computing
HTTP	Hyper-Text Transfer Protocol
JSON	JavaScript Object Notation
LM	Light microscopy
MANC	FlyEM Male Adult Nerve Cord
MIP	Max intensity projection
MCFO	Multi-Color Flip Out
PPPM	PatchPerPix Match
REST	Representational State Transfer
SWC	File format for reconstructed neuron morphologies
TB	Terabyte
UAS	Upstream Activating Sequence
VNC	Ventral nerve cord of <i>Drosophila</i>
ZIP	Compressed file format

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12859-024-05732-7>.

**Additional file 1.** Screen capture movie showing a user clicking on the “View in 3D” button to inspect a putative match between a EM neuron skeleton and LM image in 3D.

**Additional file 2.** Screen capture movie showing a user inspecting a putative match in VVDViewer after downloading the EM neuron skeleton and LM image using the “Download 3D Files” tab.

### Acknowledgements

We are sincerely grateful to Geoffrey Meissner for considerable input into both the NeuronBridge application design and the draft manuscript. We are likewise very grateful to Stephan Preibisch for considerable input in reviewing and editing of the draft manuscript. We would like to thank the FlyLight and FlyEM project teams for our long-term collaborations without which NeuronBridge would not have been possible. Thank you to Wyatt Korff, Reed George, and Gerry Rubin for supporting this project. Thanks to Stephen Plaza for proposing the initial idea. Thanks to Dagmar Kainmueller and Lisa Mais for help with integrating their PPPM results. Thanks to Scott Glasser and Ray Chang for valuable advice and prototyping of the burst-parallel search implementation on AWS. Thanks to Antje Kazimiers for early prototyping work on the web GUI. Special thanks to Brianna Yarbrough and Yisheng He for designing the NeuronBridge logo. This software makes use of the Computational Morphometry Toolkit (<http://www.nitrc.org/projects/cmtk>), supported by the National Institute of Biomedical Imaging and Bioengineering. Cloud storage for images was generously provided by Amazon Web Services (AWS) as part of the AWS Open Data Sponsorship Program. The software architecture diagrams use icons from AWS Architecture Icons and flaticon.com.

### Authors' contribution

JC designed and implemented the front-end web client. CG and DJO implemented the back-end data processing system and generated the precomputed matches. CG and JC implemented the serverless back-end. HO designed and optimized the modified CDM Search algorithm and created the image alignment tools. TK implemented the CDM Search algorithms and optimized the custom search. RS uploaded and managed data on S3, and created the DynamoDB database for lookups. PMH implemented the browser-based 3D visualization. KR conceptualized the project, designed the software architecture, managed the engineering effort, implemented the burst-parallel search framework, and wrote the draft manuscript. All authors reviewed and approved the final version of the manuscript.

### Funding

This work was funded by the Howard Hughes Medical Institute.

### Availability of data and materials

All public data in NeuronBridge (i.e., excluding data uploaded by users) is available for access on AWS S3. In addition to acting as a database back-end for the NeuronBridge browser application, the data on S3 serves as an Open Data API (Open Data API). All of the data is licensed under the Creative Commons Attribution 4.0 International (CC BY 4.0) license.

### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

Received: 17 January 2023 Accepted: 6 March 2024

Published online: 15 March 2024

### References

1. Takemura S-Y, Bharioke A, Lu Z, Nern A, Vitaladevuni S, Rivlin PK, Katz WT, Olbris DJ, Plaza SM, Winston P, Zhao T, Horne JA, Fetter RD, Takemura S, Blazek K, Chang L-A, Ogundeyi O, Saunders MA, Shapiro V, Sigmund C, Rubin GM, Scheffer LK, Meinertzhagen IA, Chklovskii DB. A visual motion detection circuit suggested by *Drosophila* connectomics. *Nature*. 2013;500(7461):175–81. <https://doi.org/10.1038/nature12450>.
2. Xu CS, Hayworth KJ, Lu Z, Grob P, Hassan AM, García-Cerdán JG, Niyogi KK, Nogales E, Weinberg RJ, Hess HF. Enhanced FIB-SEM systems for large-volume 3D imaging. *eLife*. 2017;6:25916. <https://doi.org/10.7554/eLife.25916>.
3. Zheng Z, Lauritzen JS, Perlman E, Robinson CG, Nichols M, Milkie D, Torrens O, Price J, Fisher CB, Sharifi N, Calle-Schuler SA, Kmcocva L, Ali IJ, Karsh B, Trautman ET, Bogovic JA, Hanslovsky P, Jefferis GSXE, Kazhdan M, Khairy K, Saalfeld S, Fetter RD, Bock DD. A complete electron microscopy volume of the brain of adult *Drosophila melanogaster*. *Cell*. 2018;174(3):730–74322. <https://doi.org/10.1016/j.cell.2018.06.019>.
4. Scheffer LK, Xu CS, Januszewski M, Lu Z, Takemura S-Y, Hayworth KJ, Huang GB, Shinomiya K, Maitlin-Shepard J, Berg S, Clements J, Hubbard PM, Katz WT, Umayam L, Zhao T, Ackerman D, Blakely T, Bogovic J, Dolafi T, Kainmueller D, Kawase T, Khairy KA, Leavitt L, Li PH, Lindsey L, Neubarth N, Olbris DJ, Otsuna H, Trautman ET, Ito M, Bates

- AS, Goldammer J, Wolff T, Svirskas R, Schlegel P, Neace E, Knecht CJ, Alvarado CX, Bailey DA, Ballinger S, Borycz JA, Canino BS, Cheatham N, Cook M, Dreher M, Duclos O, Eubanks B, Fairbanks K, Finley S, Forknall N, Francis A, Hopkins GP, Joyce EM, Kim S, Kirk NA, Kovalyak J, Lauchie SA, Lohff A, Maldonado C, Manley EA, McLin S, Mooney C, Ndama M, Ogundeyi O, Okeoma N, Ordish C, Padilla N, Patrick CM, Paterson T, Phillips EE, Phillips EM, Rampally N, Ribeiro C, Robertson MK, Rymer JT, Ryan SM, Sammons M, Scott AK, Scott AL, Shinomiya A, Smith C, Smith K, Smith NL, Sobeski MA, Suleiman A, Swift J, Takemura S, Talebi I, Tarnogorska D, Tenshaw E, Tokhi T, Walsh JJ, Yang T, Horne JA, Li F, Parekh R, Rivlin PK, Jayaraman V, Costa M, Jefferis GS, Ito K, Saalfeld S, George R, Meinertzhagen IA, Rubin GM, Hess HF, Jain V, Plaza SM. A connectome and analysis of the adult *Drosophila* central brain. *eLife*. 2020;9:57443. <https://doi.org/10.7554/eLife.57443>.
5. Phelps JS, Hildebrand DGC, Graham BJ, Kuan AT, Thomas LA, Nguyen TM, Buhmann J, Azevedo AW, Sustar A, Agrawal S, Liu M, Shanny BL, Funke J, Tuthill JC, Lee W-CA. Reconstruction of motor control circuits in adult *Drosophila* using automated transmission electron microscopy. *Cell*. 2021;184(3):759–77418. <https://doi.org/10.1016/j.cell.2020.12.013>.
  6. Takemura S-y, Hayworth KJ, Huang GB, Januszewski M, Lu Z, Marin EC, Preibisch S, Xu CS, Bogovic J, Champion AS, Cheong HS, Costa M, Eichler K, Katz W, Knecht C, Li F, Morris BJ, Ordish C, Rivlin PK, Schlegel P, Shinomiya K, Stürner T, Zhao T, Badalamente G, Bailey D, Brooks P, Canino BS, Clements J, Cook M, Duclos O, Dunne CR, Fairbanks K, Fang S, Finley-May S, Francis A, George R, Gkantia M, Harrington K, Hopkins GP, Hsu J, Hubbard PM, Javier A, Kainmueller D, Korff W, Kovalyak J, Krzemiński D, Lauchie SA, Lohff A, Maldonado C, Manley EA, Mooney C, Neace E, Nichols M, Ogundeyi O, Okeoma N, Paterson T, Phillips E, Phillips EM, Ribeiro C, Ryan SM, Rymer JT, Scott AK, Scott AL, Shepherd D, Shinomiya A, Smith C, Smith N, Suleiman A, Takemura S, Talebi I, Tamimi IF, Trautman ET, Umayam L, Walsh JJ, Yang T, Rubin GM, Scheffer LK, Funke J, Saalfeld S, Hess HF, Plaza SM, Card GM, Jefferis GS, Berg S. A connectome of the male *Drosophila* ventral nerve cord. *bioRxiv*. 2023.06.05.543757 Section: New Results. 2023. <https://doi.org/10.1101/2023.06.05.543757>
  7. Dorkenwald S, Matsliah A, Sterling AR, Schlegel P, Yu S-c, McKellar CE, Lin A, Costa M, Eichler K, Yin Y, Silversmith W, Schneider-Mizell C, Jordan CS, Brittain D, Halageri A, Kuehner K, Ogedengbe O, Morey R, Gager J, Kruk K, Perlman E, Yang R, Deutsch D, Bland D, Sorek M, Lu R, Macrina T, Lee K, Bae JA, Mu S, Nehoran B, Mitchell E, Popovych S, Wu J, Jia Z, Castro M, Kemnitz N, Ih D, Bates AS, Eckstein N, Funke J, Collman F, Bock DD, Jefferis GSXE, Seung HS, Murthy M, Consortium TF. Neuronal wiring diagram of an adult brain. *bioRxiv*. 2023.06.27.546656 Section: New Results (2023). <https://doi.org/10.1101/2023.06.27.546656>
  8. Schlegel P, Bates AS, Stürner T, Jagannathan SR, Drummond N, Hsu J, Serratos Capdevila L, Javier A, Marin EC, Barth-Maron A, Tamimi IF, Li F, Rubin GM, Plaza SM, Costa M, Jefferis GSXE. Information flow, cell types and stereotypy in a full olfactory connectome. *eLife*. 2021;10:66018. <https://doi.org/10.7554/eLife.66018>.
  9. Jenett A, Rubin GM, Ngo T-TB, Shepherd D, Murphy C, Dionne H, Pfeiffer BD, Cavallaro A, Hall D, Jeter J, Iyer N, Fetter D, Hausenfluck JH, Peng H, Trautman ET, Svirskas RR, Myers EW, Iwinski ZR, Aso Y, DePasquale GM, Enos A, Hulamm P, Lam SCB, Li H-H, Laverty TR, Long F, Qu L, Murphy SD, Rokicki K, Safford T, Shaw K, Simpson JH, Sowell A, Tae S, Yu Y, Zugates CT. A GAL4-driver line resource for *Drosophila* neurobiology. *Cell Rep*. 2012;2(4):991–1001. <https://doi.org/10.1016/j.celrep.2012.09.011>.
  10. Brand AH, Perrimon N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development (Cambridge, England)*. 1993;118(2):401–15. <https://doi.org/10.1242/dev.118.2.401>.
  11. Luan H, Diao F, Scott RL, White BH. The *Drosophila* split Gal4 system for neural circuit mapping. *Front Neural Circuits*. 2020;14: 603397. <https://doi.org/10.3389/fncir.2020.603397>.
  12. Pfeiffer B, Ngo B, Hibbard K, Murphy C, Jenett A, Truman J, Rubin G. Refinement of tools for targeted gene expression in *Drosophila*. *Genetics*. 2010;186:735–55. <https://doi.org/10.1534/genetics.110.119917>.
  13. Otsuna H, Ito M, Kawase T. Color depth MIP mask search: a new tool to expedite Split-GAL4 creation. Technical report. 2018. <https://doi.org/10.1101/318006>, Company: Cold Spring Harbor Laboratory Distributor: Cold Spring Harbor Laboratory Label: Cold Spring Harbor Laboratory Section: New Results Type: article. <https://www.biorxiv.org/content/10.1101/318006v1>. Accessed 12 Nov 2021.
  14. Mais L, Hirsch P, Managan C, Wang K, Rokicki K, Svirskas RR, Dickson BJ, Korff W, Rubin GM, Ihrke G, Meissner GW, Kainmueller D. PatchPerPixMatch for automated 3d search of neuronal morphologies in light microscopy. Technical report. 2021. <https://doi.org/10.1101/2021.07.23.453511>, Company: Cold Spring Harbor Laboratory Distributor: Cold Spring Harbor Laboratory Label: Cold Spring Harbor Laboratory Section: New Results Type: article. <https://www.biorxiv.org/content/10.1101/2021.07.23.453511v1>. Accessed 12 Nov 2021
  15. Meissner GW, Nern A, Dorman Z, DePasquale GM, Forster K, Gibney T, Hausenfluck JH, He Y, Iyer N, Jeter J, Johnson L, Johnston RM, Lee K, Melton B, Yarbrough B, Zugates CT, Clements J, Goina C, Otsuna H, Rokicki K, Svirskas RR, Aso Y, Card GM, Dickson BJ, Ehrhardt E, Goldammer J, Ito M, Kainmueller D, Korff W, Mais L, Minegishi R, Namiki S, Rubin GM, Sterne GR, Wolff T, Malkesman O, Team FP. A searchable image resource of *Drosophila* GAL4-driver expression patterns with single neuron resolution. Publisher: eLife Sciences Publications Limited. 2023. <https://doi.org/10.7554/eLife.80660>. <https://elifesciences.org/articles/80660>. Accessed 7 Mar 2024.
  16. Nern A, Pfeiffer BD, Rubin GM. Optimized tools for multicolor stochastic labeling reveal diverse stereotyped cell arrangements in the fly visual system. *Proc Natl Acad Sci*. 2015;112(22):2967–76. <https://doi.org/10.1073/pnas.1506763112>.
  17. Gao S, Takemura S-Y, Ting C-Y, Huang S, Lu Z, Luan H, Rister J, Thum AS, Yang M, Hong S-T, Wang JW, Odenwald WF, White BH, Meinertzhagen IA, Lee C-H. The neural substrate of spectral preference in *Drosophila*. *Neuron*. 2008;60(2):328–42. <https://doi.org/10.1016/j.neuron.2008.08.010>.
  18. Tuthill JC, Nern A, Holtz SL, Rubin GM, Reiser MB. Contributions of the 12 neuron classes in the fly lamina to motion vision. *Neuron*. 2013;79(1):128–40. <https://doi.org/10.1016/j.neuron.2013.05.024>.
  19. Aso Y, Hattori D, Yu Y, Johnston RM, Iyer NA, Ngo T-T, Dionne H, Abbott L, Axel R, Tanimoto H, Rubin GM. The neuronal architecture of the mushroom body provides a logic for associative learning. *eLife*. 2014;3:04577. <https://doi.org/10.7554/eLife.04577>.

20. Wu M, Nern A, Williamson WR, Morimoto MM, Reiser MB, Card GM, Rubin GM. Visual projection neurons in the *Drosophila* lobula link feature detection to distinct behavioral programs. *eLife*. 2016;5:21022. <https://doi.org/10.7554/eLife.21022>.
21. Aso Y, Rubin GM. Dopaminergic neurons write and update memories with cell-type-specific rules. *eLife*. 2016;5:16135. <https://doi.org/10.7554/eLife.16135>.
22. Robie AA, Hirokawa J, Edwards AW, Umayam LA, Lee A, Phillips ML, Card GM, Korff W, Rubin GM, Simpson JH, Reiser MB, Branson K. Mapping the neural substrates of behavior. *Cell*. 2017;170(2):393–40628. <https://doi.org/10.1016/j.cell.2017.06.032>.
23. Namiki S, Dickinson MH, Wong AM, Korff W, Card GM. The functional organization of descending sensory-motor pathways in *Drosophila*. *eLife*. 2018;7:34272. <https://doi.org/10.7554/eLife.34272>.
24. Wolff T, Rubin GM. Neuroarchitecture of the *Drosophila* central complex: a catalog of nodulus and asymmetrical body neurons and a revision of the protocerebral bridge catalog. *J Comp Neurol*. 2018;526(16):2585–611. <https://doi.org/10.1002/cne.24512>.
25. ...Dolan M-J, Frechter S, Bates AS, Dan C, Huovalia P, Roberts RJ, Schlegel P, Dhawan S, Tabano R, Dionne H, Christoforou C, Close K, Sutcliffe B, Giuliani B, Li F, Costa M, Ihrke G, Meissner GW, Bock DD, Aso Y, Rubin GM, Jefferis GS. Neurogenetic dissection of the *Drosophila* lateral horn reveals major outputs, diverse behavioural functions, and interactions with the mushroom body. *eLife*. 2019;8:43079. <https://doi.org/10.7554/eLife.43079>.
26. Bidaye SS, Laturney M, Chang AK, Liu Y, Bockemühl T, Büschges A, Scott K. Two brain pathways initiate distinct forward walking programs in *Drosophila*. *Neuron*. 2020;108(3):469–4858. <https://doi.org/10.1016/j.neuron.2020.07.032>.
27. Davis FP, Nern A, Picard S, Reiser MB, Rubin GM, Eddy SR, Henry GL. A genetic, genomic, and computational resource for exploring neural circuit function. *eLife*. 2020;9:50901. <https://doi.org/10.7554/eLife.50901>.
28. Morimoto MM, Nern A, Zhao A, Rogers EM, Wong AM, Isaacson MD, Bock DD, Rubin GM, Reiser MB. Spatial readout of visual imaging in the central brain of *Drosophila*. *eLife*. 2020;9:57685. <https://doi.org/10.7554/eLife.57685>.
29. Schretter CE, Aso Y, Robie AA, Dreher M, Dolan M-J, Chen N, Ito M, Yang T, Parekh R, Branson KM, Rubin GM. Cell types and neuronal circuitry underlying female aggression in *Drosophila*. *eLife*. 2020;9:58942. <https://doi.org/10.7554/eLife.58942>.
30. Feng K, Sen R, Minegishi R, Dübbert M, Bockemühl T, Büschges A, Dickson BJ. Distributed control of motor circuits for backward walking in *Drosophila*. *Nat Commun*. 2020;11(1):6166. <https://doi.org/10.1038/s41467-020-19936-x>.
31. Wang K, Wang F, Forknall N, Yang T, Patrick C, Parekh R, Dickson BJ. Neural circuit mechanisms of sexual receptivity in *Drosophila* females. *Nature*. 2021;589(7843):577–81. <https://doi.org/10.1038/s41586-020-2972-7>.
32. Sterne GR, Otsuna H, Dickson BJ, Scott K. Classification and genetic targeting of cell types in the primary taste and premotor center of the adult *Drosophila* brain. *eLife*. 2021;10:71679. <https://doi.org/10.7554/eLife.71679>.
33. Nojima T, Rings A, Allen AM, Otto N, Verschut TA, Billeter J-C, Neville MC, Goodwin SF. A sex-specific switch between visual and olfactory inputs underlies adaptive sex differences in behavior. *Curr Biol*. 2021;31(6):1175–11916. <https://doi.org/10.1016/j.cub.2020.12.047>.
34. Tanaka R, Clark DA. Neural mechanisms to exploit positional geometry for collision avoidance. *Curr Biol*. 2022. <https://doi.org/10.1016/j.cub.2022.04.023>.
35. Namiki S, Ros IG, Morrow C, Rowell WJ, Card GM, Korff W, Dickinson MH. A population of descending neurons that regulates the flight motor of *Drosophila*. *Curr Biol*. 2022;32(5):1189–11966. <https://doi.org/10.1016/j.cub.2022.01.008>.
36. Israel S, Rozenfeld E, Weber D, Huetteroth W, Parnas M. Olfactory stimuli and moonwalker SEZ neurons can drive backward locomotion in *Drosophila*. *Curr Biol*. 2022;32(5):1131–11497. <https://doi.org/10.1016/j.cub.2022.01.035>.
37. Meissner, GW, Vannan, A, Jeter, J, Atkins, M, Bowers, S, Close, K, DePasquale, GM, Dorman, Z, Forster, K, Beringer, JA, Gibney, TV, Gulati, A, Hausenfluck, JH, He, Y, Henderson, K, Johnson, L, Johnston, RM, Ihrke, G, Iyer, N, Lazarus, R, Lee, K, Li, H-H, Liaw, H-P, Melton, B, Miller, S, Motaher, R, Novak, A, Ogundeyi, O, Petrucio, A, Price, J, Protopapas, S, Tae, S, Tata, A, Taylor, J, Vorimo, R, Yarbrough, B, Zeng, KX, Zugates, CT, Dionne, H, Angstadt, C, Ashley, K, Cavallaro, A, Dang, T, Gonzalez, GA, Hibbard, KL, Huang, C, Kao, J-C, Laverty, T, Mercer, M, Perez, B, Pitts, S, Ruiz, D, Vallanadu, V, Zheng, GZ, Goina, C, Otsuna, H, Rokicki, K, Svirskas, RR, Cheong, HS, Dolan, M-J, Ehrhardt, E, Feng, K, Galfi, BE, Goldammer, J, Hu, N, Ito, M, McKellar, C, Minegishi, R, Namiki, S, Nern, A, Schretter, CE, Sterne, GR, Venkatasubramanian, L, Wang, K, Wolff, T, Wu, M, George, R, Malkesman, O, Aso, Y, Card, GM, Dickson, BJ, Korff, W, Ito, K, Truman, JW, Zlatic, M, Rubin, GM, Team, FP: A split-GAL4 driver line resource for *Drosophila* CNS cell types. *bioRxiv*. Section: New Results. 2024. <https://doi.org/10.1101/2024.01.09.574419>. <https://www.biorxiv.org/content/10.1101/2024.01.09.574419v1>.
38. Sareen PF, McCurdy LY, Nitabach MN. A neuronal ensemble encoding adaptive choice during sensory conflict in *Drosophila*. *Nat Commun*. 2021;12(1):4131. <https://doi.org/10.1038/s41467-021-24423-y>.
39. Bogovic JA, Otsuna H, Heinrich L, Ito M, Jeter J, Meissner G, Nern A, Colonell J, Malkesman O, Ito K, Saalfeld S. An unbiased template of the *Drosophila* brain and ventral nerve cord. *PLoS ONE*. 2020;15(12):0236495. <https://doi.org/10.1371/journal.pone.0236495>. (Publisher: Public Library of Science. Accessed 2021-11-19).
40. Hirsch P, Mais L, Kainmueller D, PatchPerPix for instance segmentation. *arXiv:2001.07626* [cs] (2020). Accessed 8 Apr 2022
41. Costa M, Manton J, Ostrovsky A, Prohaska S, Jefferis GXE. NBLAST: rapid, sensitive comparison of neuronal structure and construction of neuron family databases. *Neuron*. 2016;91(2):293–311. <https://doi.org/10.1016/j.neuron.2016.06.012>. (Accessed 2022-01-10).
42. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez J-Y, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A. Fiji: an open-source platform for biological-image analysis. *Nat Methods*. 2012;9(7):676–82. <https://doi.org/10.1038/nmeth.2019>. (Number: 7 Publisher: Nature Publishing Group. Accessed 2024-01-09).
43. Milyaev N, Osumi-Sutherland D, Reeve S, Burton N, Baldock RA, Armstrong JD. The Virtual Fly Brain browser and query interface. *Bioinformatics*. 2012;28(3):411–5. <https://doi.org/10.1093/bioinformatics/btr677>. (Accessed 2022-01-11).

44. Rohlfing T, Maurer CR. Nonrigid image registration in shared-memory multiprocessor environments with application to brains, breasts, and bees. *IEEE Trans Inf Technol Biomed Publ IEEE Eng Med Biol Soc.* 2003;7(1):16–25. <https://doi.org/10.1109/titb.2003.808506>.
45. Bogovic J, Otsuna H, Takemura S-Y, Shinomiya K, Saalfeld S, Kawase T. Robust search method for *Drosophila* neurons between electron and light microscopy. 2024 (**in preparation**).
46. Kawase T, Sugano SS, Shimada T, Hara-Nishimura I. A direction-selective local-thresholding method, DSLT, in combination with a dye-based method for automated three-dimensional segmentation of cells and airspaces in developing leaves. *Plant J Cell Mol Biol.* 2015;81(2):357–66. <https://doi.org/10.1111/tpj.12738>.
47. Kawase T, Otsuna H. VVD viewer codebase. GitHub. 2022. <https://doi.org/10.5281/zenodo.6762216>.
48. Fouladi S, Romero F, Iter D, Li Q, Chatterjee S, Kozyrakis C, Zaharia M, Winstein K. From laptop to lambda: outsourcing everyday jobs to thousands of transient functional containers, pp. 475–488. 2019. <https://www.usenix.org/conference/atc19/presentation/fouladi>. Accessed 25 Mar 2022.
49. Rokicki K. Scaling Neuroscience Research on AWS. 2021. <https://aws.amazon.com/blogs/architecture/scaling-neuroscience-research-on-aws/>.
50. Lazar AA, Liu T, Turkcan MK, Zhou Y. Accelerating with FlyBrainLab the discovery of the functional logic of the *Drosophila* brain in the connectomic and synaptic era. *eLife.* 2021;10:62362. <https://doi.org/10.7554/eLife.62362>. (**Publisher: eLife Sciences Publications, Ltd. Accessed 2021-12-03**).
51. Cannon RC, Turner DA, Pyapali GK, Wheal HV. An on-line archive of reconstructed hippocampal neurons. *J Neurosci Methods.* 1998;84(1):49–54. [https://doi.org/10.1016/S0165-0270\(98\)00091-0](https://doi.org/10.1016/S0165-0270(98)00091-0).
52. Plaza SM, Clements J, Dolafi T, Umayam L, Neubarth NN, Scheffer LK, Berg S. neuPrint: An open access tool for EM connectomics. *Front Neuroinform.* 2022;16:896292. <https://doi.org/10.3389/fninf.2022.896292>.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.