

# 1 A Connectome and Analysis of the 2 Adult *Drosophila* Central Brain

3 Louis K. Scheffer<sup>1\*</sup>, C. Shan Xu<sup>1</sup>, Michal Januszewski<sup>2</sup>, Zhiyuan Lu<sup>1,3</sup>, Shin-ya  
4 Takemura<sup>1</sup>, Kenneth J. Hayworth<sup>1</sup>, Gary B. Huang<sup>1</sup>, Kazunori Shinomiya<sup>1</sup>, Jeremy  
5 Maitin-Shepard<sup>2</sup>, Stuart Berg<sup>1</sup>, Jody Clements<sup>1</sup>, Philip Hubbard<sup>1</sup>, William Katz<sup>1</sup>,  
6 Lowell Umayam<sup>1</sup>, Ting Zhao<sup>1</sup>, David Ackerman<sup>1</sup>, Tim Blakely<sup>2</sup>, John Bogovic<sup>1</sup>, Tom  
7 Dolafi<sup>1</sup>, Dagmar Kainmueller<sup>1¶</sup>, Takashi Kawase<sup>1</sup>, Khaled A. Khairy<sup>1\*\*</sup>, Laramie  
8 Leavitt<sup>2</sup>, Peter H. Li<sup>2</sup>, Larry Lindsey<sup>2</sup>, Nicole Neubarth<sup>1††</sup>, Donald J. Olbris<sup>1</sup>, Hideo  
9 Otsuna<sup>1</sup>, Eric T. Trautman<sup>1</sup>, Masayoshi Ito<sup>1,4</sup>, Jens Goldammer<sup>1,5</sup>, Tanya Wolff<sup>1</sup>,  
10 Robert Svirskas<sup>1</sup>, Philipp Schlegel<sup>9</sup>, Erika R. Neace<sup>1</sup>, Christopher J. Knecht, Jr.<sup>1</sup>,  
11 Chelsea X. Alvarado<sup>1</sup>, Dennis A. Bailey<sup>1</sup>, Samantha Ballinger<sup>1</sup>, Jolanta A Borycz<sup>3</sup>,  
12 Brandon S. Canino<sup>1</sup>, Natasha Cheatham<sup>1</sup>, Michael Cook<sup>1</sup>, Marisa Dreher<sup>1</sup>, Octave  
13 Duclos<sup>1</sup>, Bryon Eubanks<sup>1</sup>, Kelli Fairbanks<sup>1</sup>, Samantha Finley<sup>1</sup>, Nora Forknall<sup>1</sup>,  
14 Audrey Francis<sup>1</sup>, Gary Patrick Hopkins<sup>1</sup>, Emily M. Joyce<sup>1</sup>, Sungjin Kim<sup>1</sup>, Nicole A.  
15 Kirk<sup>1</sup>, Julie Kovalyak<sup>1</sup>, Shirley A. Lauchie<sup>1</sup>, Alanna Lohff<sup>1</sup>, Charli Maldonado<sup>1</sup>,  
16 Emily A. Manley<sup>1</sup>, Sari McLin<sup>3</sup>, Caroline Mooney<sup>1</sup>, Miatta Ndama<sup>1</sup>, Omotara  
17 Ogundeyi<sup>1</sup>, Nneoma Okeoma<sup>1</sup>, Christopher Ordish<sup>1</sup>, Nicholas Padilla<sup>1</sup>,  
18 Christopher Patrick<sup>1</sup>, Tyler Paterson<sup>1</sup>, Elliott E. Phillips<sup>1</sup>, Emily M. Phillips<sup>1</sup>, Neha  
19 Rampally<sup>1</sup>, Caitlin Ribeiro<sup>1</sup>, Madelaine K Robertson<sup>3</sup>, Jon Thomson Rymer<sup>1</sup>, Sean  
20 M. Ryan<sup>1</sup>, Megan Sammons<sup>1</sup>, Anne K. Scott<sup>1</sup>, Ashley L. Scott<sup>1</sup>, Aya Shinomiya<sup>1</sup>,  
21 Claire Smith<sup>1</sup>, Kelsey Smith<sup>1</sup>, Natalie L. Smith<sup>1</sup>, Margaret A. Sobeski<sup>1</sup>, Alia  
22 Suleiman<sup>1</sup>, Jackie Swift<sup>1</sup>, Satoko Takemura<sup>1</sup>, Iris Talebi<sup>1</sup>, Dorota Tarnogorska<sup>3</sup>,  
23 Emily Tenshaw<sup>1</sup>, Temour Tokhi<sup>1</sup>, John J. Walsh<sup>1</sup>, Tansy Yang<sup>1</sup>, Jane Anne Horne<sup>3</sup>,  
24 Feng Li<sup>1</sup>, Ruchi Parekh<sup>1</sup>, Patricia K. Rivlin<sup>1</sup>, Vivek Jayaraman<sup>1</sup>, Kei Ito<sup>1,4,5</sup>, Stephan  
25 Saalfeld<sup>1</sup>, Reed George<sup>1</sup>, Ian A. Meinertzhagen<sup>1,3</sup>, Gerald M. Rubin<sup>1</sup>, Harald F.  
26 Hess<sup>1</sup>, Viren Jain<sup>2</sup>, Stephen M. Plaza<sup>1</sup>

**\*For correspondence:**

[Schefferl@janelia.hhmi.org](mailto:Schefferl@janelia.hhmi.org) (LKS)

**Present address:** <sup>¶</sup>Max Delbrück  
Centre for Developmental Medicine,  
Germany; <sup>\*\*</sup>Department of  
Developmental Neurobiology, St.  
Jude Children's Research Hospital;  
<sup>††</sup>Two Six Labs

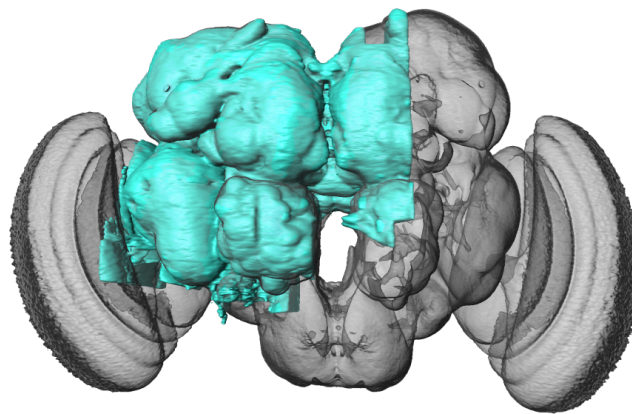
27 <sup>1</sup>Janelia Research Campus, HHMI; <sup>2</sup>Google Research; <sup>3</sup>Life Sciences Centre, Dalhousie  
28 University; <sup>4</sup>University of Tokyo, Institute for Quantitative Biosciences; <sup>5</sup>Institute of  
29 Zoology, Biocenter Cologne, University of Cologne

31 **Abstract** The neural circuits responsible for animal behavior remain largely unknown. We  
32 summarize new methods and present the circuitry of a large fraction of the brain of the fruit fly  
33 *Drosophila melanogaster*. Improved methods include new procedures to prepare, image, align,  
34 segment, find synapses in, and proofread such large data sets. We define cell types, refine  
35 computational compartments, and provide an exhaustive atlas of cell examples and types, many of  
36 them novel. We provide detailed circuits consisting of neurons and their chemical synapses for  
37 most of the central brain. We make the data public and simplify access, reducing the effort needed  
38 to answer circuit questions, and provide procedures linking the neurons defined by our analysis  
39 with genetic reagents. Biologically, we examine distributions of connection strengths, neural motifs  
40 on different scales, electrical consequences of compartmentalization, and evidence that  
41 maximizing packing density is an important criterion in the evolution of the fly's brain.

42

## 43 Introduction

44 The connectome we present is a dense reconstruction of a portion of the central brain (referred to  
45 here as the hemibrain) of the fruit fly, *Drosophila melanogaster*, as shown in Figure 1. This region  
46 was chosen since it contains all the circuits of the central brain (assuming bilateral symmetry), and  
47 in particular contains circuits critical to unlocking mysteries involving associative learning in the  
48 mushroom body, navigation and sleep in the central complex, and circadian rhythms among clock  
49 circuits. The largest dense reconstruction to date, it contains around 25,000 neurons, most of which  
50 were rigorously clustered and named, with about  $20 \cdot 10^6$  chemical synapses between them, plus  
51 portions of many other neurons truncated by the boundary of the data set (details in Figure 1  
52 below). Each neuron is documented at many levels - the detailed voxels that constitute it, a skeleton  
with segment diameters, its synaptic partners and the location of most of their synapses.



Neurons traced, most arbors in volume (uncropped)	21,662
Neurons traced, large ( $\geq 1000$ connections) but cropped by edge of volume	4,495
Remaining traced, small ( $< 1000$ connections) and cropped	67,475
Presynaptic sites (T-Bars) in uncropped/traced/total T-bars	6M/8.6M/9.5M
Postsynaptic densities (PSDs) in uncropped/traced/total	18M/23M/64M

**Figure 1.** The hemibrain and some basic statistics. The highlighted area shows the portion of the central brain that was imaged and reconstructed, superimposed on a grayscale representation of the entire *Drosophila* brain. For the table, a neuron is traced if all its main branches within the volume are reconstructed. A neuron is considered uncropped if most arbors (though perhaps not the soma) are contained in the volume. Others are considered cropped. Note: 1) our definition of cropped is somewhat subjective; 2) the usefulness of a cropped neuron depends on the application; and 3) some small fragments are known to be distinct neurons. For simplicity, we will often state that the hemibrain contains  $\approx 25K$  neurons.

53

54 Producing this data set required advances in sample preparation, imaging, image alignment, ma-  
55 chine segmentation of cells, synapse detection, data storage, proofreading software, and protocols  
56 to arbitrate each decision. A number of new tests for estimating the completeness and accuracy  
57 were required and therefore developed, in order to verify the correctness of the connectome.

58 These data describe whole-brain properties and circuits, as well as contain new methods to  
59 classify cell types based on connectivity. Computational compartments are now more carefully  
60 defined, we identify actual synaptic circuits, and each neuron is annotated by name and putative  
61 cell type, making this the first complete census of neuropils, tracts, cells, and connections in this

<i>OL(R)</i>	optic lobe	<b>CX</b>	central complex	<b>LH(R)</b>	lateral horn
<i>LA</i>	lamina	<b>FB</b>	fan-shaped body	<i>SNP(R)/(L)</i>	superior neuropils
<b>ME(R)</b>	medulla	<b>FB1</b>	fan-shaped body layer 1	<b>SLP(R)</b>	superior lateral protocerebrum
<b>AME(R)</b>	accessory medulla	<b>FB2</b>	fan-shaped body layer 2	<b>SIP(R)/(L)</b>	superior intermediate protocerebrum
<b>LO(R)</b>	lobula	<b>FB3</b>	fan-shaped body layer 3	<b>SMP(R)/(L)</b>	superior medial protocerebrum
<b>LOP(R)</b>	lobula plate	<b>FB4</b>	fan-shaped body layer 4	<i>INP</i>	inferior neuropils
<b>MB(R)/(L)</b>	mushroom body	<b>FB5</b>	fan-shaped body layer 5	<b>CRE(R)/(L)</b>	crepine
<b>CA(R)/(L)</b>	calyx	<b>FB6</b>	fan-shaped body layer 6	<b>RUB(R)/(L)</b>	rubus
<b>dACA(R)</b>	dorsal accessory calyx	<b>FB7</b>	fan-shaped body layer 7	<b>ROB(R)</b>	round body
<b>IACA(R)</b>	lateral accessory calyx	<b>FB8</b>	fan-shaped body layer 8	<b>SCL(R)/(L)</b>	superior clamp
<b>vACA(R)</b>	ventral accessory calyx	<b>FB9</b>	fan-shaped body layer 9	<b>ICL(R)/(L)</b>	inferior clamp
<b>PED(R)</b>	pedunculus	<b>EB</b>	ellipsoid body	<b>IB</b>	inferior bridge
<b>a'L(R)/(L)</b>	alpha prime lobe	<b>EBr1</b>	ellipsoid body zone r1	<b>ATL(R)/(L)</b>	antler
<b>a'1(R)</b>	alpha prime lobe compartment 1	<b>EBr2r4</b>	ellipsoid body zone r2r4	<b>AL(R)/(L)</b>	antennal lobe
<b>a'2(R)</b>	alpha prime lobe compartment 2	<b>EBr3am</b>	ellipsoid body zone r3am	<i>VMNP</i>	ventromedial neuropils
<b>a'3(R)</b>	alpha prime lobe compartment 3	<b>EBr3d</b>	ellipsoid body zone r3d	<b>VES(R)/(L)</b>	vest
<b>aL(R)/(L)</b>	alpha lobe	<b>EBr3pw</b>	ellipsoid body zone r3pw	<b>EPA(R)/(L)</b>	epaulette
<b>a1(R)</b>	alpha lobe compartment 1	<b>EBr5</b>	ellipsoid body zone r5	<b>GOR(R)/(L)</b>	gorget
<b>a2(R)</b>	alpha lobe compartment 2	<b>EBr6</b>	ellipsoid body zone r6	<b>SPS(R)/(L)</b>	superior posterior slope
<b>a3(R)</b>	alpha lobe compartment 3	<b>AB(R)/(L)</b>	asymmetrical body	<b>IPS(R)</b>	inferior posterior slope
<b>gL(R)/(L)</b>	gamma lobe	<b>PB</b>	protocerebral bridge	<i>PENP</i>	periesophageal neuropils
<b>g1(R)</b>	gamma lobe compartment 1	<b>PB(R1)</b>	protocerebral bridge glomerulus R1	<b>SAD</b>	saddle
<b>g2(R)</b>	gamma lobe compartment 2	<b>PB(R2)</b>	protocerebral bridge glomerulus R2	<b>AMMC</b>	antennal mechanosensory and motor center
<b>g3(R)</b>	gamma lobe compartment 3	<b>PB(R3)</b>	protocerebral bridge glomerulus R3	<b>FLA(R)</b>	flange
<b>g4(R)</b>	gamma lobe compartment 4	<b>PB(R4)</b>	protocerebral bridge glomerulus R4	<b>CAN(R)</b>	cantle
<b>g5(R)</b>	gamma lobe compartment 5	<b>PB(R5)</b>	protocerebral bridge glomerulus R5	<b>PRW</b>	proaw
<b>b'L(R)/(L)</b>	beta prime lobe	<b>PB(R6)</b>	protocerebral bridge glomerulus R6	<b>GNG</b>	gnathal ganglia
<b>b'1(R)</b>	beta prime lobe compartment 1	<b>PB(R7)</b>	protocerebral bridge glomerulus R7	major fiber bundles	
<b>b'2(R)</b>	beta prime lobe compartment 2	<b>PB(R8)</b>	protocerebral bridge glomerulus R8	<b>AOT(R)</b>	anterior optic tract
<b>bL(R)/(L)</b>	beta lobe	<b>PB(R9)</b>	protocerebral bridge glomerulus R9	<b>GC</b>	great commissure
<b>b1(R)</b>	beta lobe compartment 1	<b>PB(L1)</b>	protocerebral bridge glomerulus L1	<b>GF(R)</b>	giant fiber (single neuron; not a fiber bundle)
<b>b2(R)</b>	beta lobe compartment 2	<b>PB(L2)</b>	protocerebral bridge glomerulus L2	<b>mALT(R)/(L)</b>	medial antennal lobe tract
<i>LX(R)/(L)</i>	lateral complex	<b>PB(L3)</b>	protocerebral bridge glomerulus L3	<b>POC</b>	posterior optic commissure
<b>BU(R)/(L)</b>	bulb	<b>PB(L4)</b>	protocerebral bridge glomerulus L4		
<b>LAL(R)/(L)</b>	lateral accessory lobe	<b>PB(L5)</b>	protocerebral bridge glomerulus L5		
<b>GA(R)</b>	gall	<b>PB(L6)</b>	protocerebral bridge glomerulus L6		
<i>VLNP(R)</i>	ventrolateral neuropils	<b>PB(L7)</b>	protocerebral bridge glomerulus L7		
<b>AOTU(R)</b>	anterior optic tubercle	<b>PB(L8)</b>	protocerebral bridge glomerulus L8		
<b>AVLP(R)</b>	anterior ventrolateral protocerebrum	<b>PB(L9)</b>	protocerebral bridge glomerulus L9		
<b>PVLP(R)</b>	posterior ventrolateral protocerebrum	<b>NO</b>	noduli		
<b>PLP(R)</b>	posterior lateral protocerebrum	<b>NO1(R)/(L)</b>	nodulus 1		
<b>WED(R)</b>	wedge	<b>NO2(R)/(L)</b>	nodulus 2		
		<b>NO3(R)/(L)</b>	nodulus 3		

**Figure 2.** Brain regions contained and defined in the hemibrain, following the naming conventions of (Ito et al., 2014) with the addition of (R) and (L) to specify the side of the soma for that region. Gray *italics* indicate master regions not explicitly defined in the hemibrain. Region LA is not included in the volume. The regions are hierarchical, with the more indented regions forming subsets of the less indented. The only exceptions are dACA, IACA, and vACA which are considered part of the mushroom body but are not contained in the master region MB.

62 portion of the brain. We compare the statistics and structure of different brain regions, and for  
 63 the brain as a whole, without the confounds introduced by studying different circuitry in different  
 64 animals.

65 All data are publicly available through web interfaces. This includes a browser interface, Ne-  
 66 uPrint (Clements et al., 2020), designed so that any interested user can query the hemibrain con-  
 67 nectome even without specific training. NeuPrint can query the connectivity, partners, connection  
 68 strengths and morphologies of all specified neurons, thus making identification of upstream and  
 69 downstream partners orders of magnitude easier than through existing genetic methods. In addi-  
 70 tion, for those who are willing to program, the full data set - the gray scale voxels, the segmentation  
 71 and proofreading results, skeletons, and graph model of connectivity, are also available through  
 72 publicly accessible application program interfaces (APIs).

73 This effort differs from previous EM reconstructions in its social and collaborative aspects.  
 74 Previous reconstructions were either dense in much smaller EM volumes (such as (Meinertzhagen  
 75 and O'neil, 1991) (Helmstaedter et al., 2013) (Takemura et al., 2017)) or sparse in larger volumes  
 76 (such as (Eichler et al., 2017) or (Zheng et al., 2018)). All have concentrated on the reconstruction  
 77 of specific circuits to answer specific questions. When the same EM volume is used for many

78 such efforts, as has occurred in the *Drosophila* larva and the full adult fly brain, this leads to an  
79 overall reconstruction that is the union of many individual efforts(Saalfeld *et al.*, 2009). The result  
80 is inconsistent coverage of the brain, with some regions well reconstructed and others missing  
81 entirely. In contrast, here we have analyzed the entire volume, not just the subsets of interest to  
82 specific groups of researchers with the expertise to tackle EM reconstruction. We are making these  
83 data available without restriction, with only the requirement to cite the source. This allows the  
84 benefits of known circuits and connectivity to accrue to the field as a whole, a much larger audience  
85 than those with expertise in EM reconstruction. This is analogous to progress in genomics, which  
86 transitioned from individual groups studying subsets of genes, to publicly available genomes that  
87 can be queried for information about genes of choice(Altschul *et al.*, 1990).

88 One major benefit to this effort is to facilitate research into the circuits of the fly's brain. A  
89 common question among researchers, for example, is the identity of upstream and downstream  
90 (respectively input and output) partners of specific neurons. Previously this could only be addressed  
91 by genetic trans-synaptic labelling, such as trans-Tango(Talay *et al.*, 2017), or by sparse tracing in  
92 previously imaged EM volumes(Zheng *et al.*, 2018). However, the genetic methods may give false  
93 positives and negatives, and both alternatives require specialized expertise and are time consuming,  
94 often taking months of effort. Now, for any circuits contained in our volume, a researcher can  
95 obtain the same answers in seconds by querying a publicly available database.

96 Another major benefit of dense reconstruction is its exhaustive nature. Genetic methods such  
97 as stochastic labeling may miss cell types, and counts of cells of a given type are dependent on  
98 expression levels, which are always uncertain. Previous dense reconstructions have demonstrated  
99 that existing catalogs of cell types are incomplete, even in well-covered regions(Takemura *et al.*,  
100 2017). In our hemibrain sample, we have identified all the cells within the reconstructed volume,  
101 thus providing a complete and unbiased census of all cell types in the fly's central brain (at least in  
102 this single female), and a precise count of the instances of each type.

103 Another scientific benefit lies in an analysis without the uncertainty of pooling data obtained  
104 from different animals. The detailed circuitry of the fly's brain is known to depend on nutritional  
105 history, age, and circadian rhythm. Here these factors are held constant, as are the experimental  
106 methods, facilitating comparison between different fly brain regions in this single animal. Evaluating  
107 stereotypy across animals will of course eventually require additional connectomes.

108 Previous reconstructions of compartmentalized brains have concentrated on particular regions  
109 and circuits. The mammalian retina(Helmstaedter *et al.*, 2013) and cortex(Kasthuri *et al.*, 2015),  
110 and insect mushroom bodies(Eichler *et al.*, 2017)(Takemura *et al.*, 2017) and optic lobes(Takemura  
111 *et al.*, 2015) have all been popular targets. Additional studies have examined circuits that cross  
112 regions, such as those for sensory integration(Ohyama *et al.*, 2015) or motion vision(Shinomiya  
113 *et al.*, 2019).

114 So far lacking are systematic studies of the statistical properties of computational compartments  
115 and their connections. Neural circuit motifs have been studied(Song *et al.*, 2005), but only those  
116 restricted to small motifs and at most a few cell types, usually in a single portion of the brain. Many  
117 of these results are in mammals, leading to questions of whether they also apply to invertebrates,  
118 and whether they extend to other regions of the brain. While there have been efforts to build  
119 reduced, but still accurate, electrical models of neurons(Marasco *et al.*, 2012), none of these to our  
120 knowledge have used the compartments structure of the brain.

## 121 **What is included**

122 Figure 2 shows the hierarchy of the named brain regions that are included in the hemibrain. Table 1  
123 shows the primary regions that are at least 50% included in the hemibrain sample, their approximate  
124 size, and their completion percentage. Our names for brain regions follow the conventions of (Ito  
125 *et al.*, 2014) with the addition of '(L)' or '(R)' to indicate whether the region (most of which occur on  
126 both sides of the fly) has its cell bodies in the left or right, respectively. The mushroom body(Tanaka  
127 *et al.*, 2008)(Aso *et al.*, 2014) and central complex(Wolff *et al.*, 2015) are further divided into finer



128 compartments.

129 The supplementary material includes a section on known sensory input, and motor outputs,  
130 included in the volume.

### 131 **Differences from connectomes of vertebrates**

132 Most accounts of neurobiology define the operation of the mammalian nervous system with, at  
133 most, only passing reference to invertebrate brains. Fly (or other insect) nervous systems differ from  
134 those of vertebrates in several respects (*Meinertzhagen, 2016b*). Some main differences include:

- 135 • Most synapses are polyadic. Each synapse structure comprises a single presynaptic release  
136 site and, adjacent to this, several neurites expressing neurotransmitter receptors. An element,  
137 T-shaped and typically called a T-bar in flies, marks the site of transmitter release into the cleft  
138 between cells. This site typically abuts the neurites of several other cells, where a postsynaptic  
139 density (PSD) marks the receptor location.
- 140 • Most neurites are neither purely axonic or dendritic, but have both pre- and postsynaptic part-  
141 ners, a feature that may be more prominent in mammalian brains than recognized (*Morgan  
142 and Lichtman, 2020*). Within a single brain region, however, neurites are frequently predomi-  
143 nantly dendritic (postsynaptic) or axonic (presynaptic).
- 144 • Unlike some synapses in mammals, EM imagery (at least as we have acquired and analyzed it  
145 here) fails to reveal obvious information about whether a synapse is excitatory or inhibitory.
- 146 • The soma or cell body of each fly neuron resides in a rind (the cell body layer) on the periphery

Name	%inV	T-bars	comp%	Name	%inV	T-bars	comp%
PED(R)	100%	54805	85%	aL(R)	100%	95375	84%
b'L(R)	100%	67695	83%	bL(R)	100%	71112	83%
gL(R)	100%	176785	83%	a'L(R)	100%	39091	82%
EB	100%	164286	81%	bL(L)	56%	58799	80%
NO	100%	36722	79%	b'L(L)	88%	57802	78%
gL(L)	55%	133256	76%	CA(R)	100%	69515	73%
AB(R)	100%	2734	64%	aL(L)	51%	44803	62%
FB	100%	451040	61%	AL(R)	83%	501007	58%
AB(L)	100%	572	57%	PB	100%	46557	55%
AME(R)	100%	6045	47%	BU(R)	100%	9381	46%
CRE(R)	100%	137946	39%	AOTU(R)	100%	92579	37%
LAL(R)	100%	234398	36%	SMP(R)	100%	510943	33%
PVLP(R)	100%	475228	29%	ATL(R)	100%	25472	28%
SPS(R)	100%	253821	28%	ATL(L)	100%	28153	28%
VES(R)	84%	157171	27%	IB	100%	200447	27%
CRE(L)	90%	130498	27%	SIP(R)	100%	187494	26%
BU(L)	52%	7014	26%	GOR(R)	100%	27140	25%
WED(R)	100%	232901	24%	SMP(L)	100%	460793	24%
EPA(R)	100%	31439	24%	PLP(R)	100%	429106	24%
AVLP(R)	100%	630542	22%	ICL(R)	100%	202550	22%
SLP(R)	100%	475903	21%	LO(R)	64%	855261	21%
SCL(R)	100%	187674	21%	GOR(L)	60%	19558	20%
LH(R)	100%	231667	18%	CAN(R)	68%	6513	15%

**Table 1.** Regions with  $\geq 50\%$  included in the hemibrain, sorted by completion percentage. The approximate percentage of the region included in the hemibrain volume is shown as '%inV'. 'T-bars' gives a rough estimate of the size of the region. 'comp%' is the fraction of the PSDs contained in the brain region for which both the PSD, and the corresponding T-bar, are in neurons marked as 'Traced'.

147 of the brain, mostly disjoint from the main neurites innervating the internal neuropil. As  
148 a result, unlike vertebrate neurons, no synapses form directly on the soma. The neuronal  
149 process between the soma and the first branch point is the cell body fiber (CBF), which is  
150 likewise not involved in the synaptic transmission of information.  
151 • Synapse sizes are much more uniform than those of mammals. Stronger connections are  
152 formed by increasing the number of synapses in parallel, not by forming larger synapses, as in  
153 vertebrates. In this paper we will refer to the ‘strength’ of a connection as the synapse count,  
154 even though we acknowledge that we lack information on the relative activity and strength of  
155 the synapses, and thus a true measure of their coupling strength..  
156 • The brain is small, about 250  $\mu\text{m}$  per side, and has roughly the same size as the dendritic  
157 arbor of a single pyramidal neuron in the mammalian cortex.  
158 • Axons of fly neurons are not myelinated.  
159 • Some fly neurons relay on graded transmission (as opposed to spiking), without obvious  
160 anatomical distinction. Some neurons even switch between graded and spiking opera-  
161 tion(*Pimentel et al., 2016*).

## 162 **Connectome Reconstruction**

163 Producing a connectome comprising reconstructed neurons and the chemical synapses between  
164 them required several steps. The first step, preparing a fly brain and imaging half of its center,  
165 produced a dataset consisting of 26 teravoxels of data, each with 8 bits of information. We applied  
166 numerous machine learning algorithms and over 50 person-years of proofreading effort over  $\approx 2$   
167 calendar years to extract a variety of more compact and useful representations, such as neuron  
168 skeletons, synapse locations, and connectivity graphs. These are both more useful and much  
169 smaller than the raw grayscale data. For example, the connectivity could be reasonably summarized  
170 by a graph with  $\approx 25,000$  nodes and  $\approx 3$  million edges. Even when the connections were assigned to  
171 different brain regions, such a graph took only 26 MB, still large but roughly a million fold reduction  
172 in data size.

173 Many of the supporting methods for this reconstruction have been recently published. Here  
174 we briefly survey each major area, with more details reported in the companion papers. Major  
175 advances include:

- 176 • New methods to fix and stain the sample, preparing a whole fly brain with well-preserved  
177 subcellular detail particularly suitable for machine analysis.
- 178 • Methods that have enabled us to collect the largest EM dataset yet using Focused Ion Beam  
179 Scanning Electron Microscopy (FIB-SEM), resulting in isotropic data with few artifacts, features  
180 that significantly speed up reconstruction.
- 181 • A coarse-to-fine, automated flood-filling network segmentation pipeline applied to image  
182 data normalized with cycle-consistent generative adversarial networks, and an aggressive  
183 automated agglomeration regime enabled by advances in proofreading.
- 184 • A new hybrid synapse prediction method, using two differing underlying techniques, for  
185 accurate synapse prediction throughout the volume.
- 186 • New top-down proofreading methods that utilize visualization and machine learning to achieve  
187 orders of magnitude faster reconstruction compared with previous approaches in the fly’s  
188 brain.

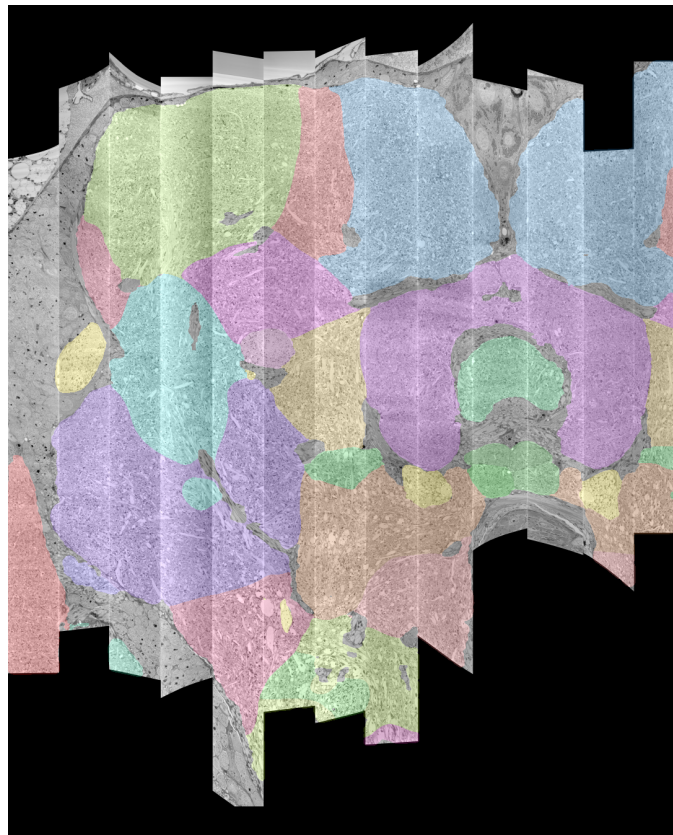
189 Each of these is explained in more detail in the following sections and, where necessary, in the  
190 Supplemental Methods.

## 191 **Image stack collection**

192 The first steps, fixing and staining the specimen, have been accomplished taking advantage of  
193 three new developments. These improved methods allow us to fix and stain a full fly’s brain  
194 but nevertheless recover neurons as round profiles with darkly stained synapses, suitable for

195 machine segmentation and automatic synapse detection. Starting with a five day old female of  
196 wild-type Canton S strain G1 x w<sup>1118</sup>, we used a custom-made jig to microdissect the central nervous  
197 system, which was then fixed and embedded in Epon, an epoxy resin. We then enhanced the  
198 electron contrast by staining with heavy metals, and progressively lowered the temperature during  
199 dehydration of the sample. Collectively these methods optimize morphological preservation, allow  
200 full-brain preparation without distortion (unlike fast freezing methods), and provide increased  
201 staining intensity that speeds the rate of FIB-SEM imaging(Lu *et al.*, 2019).

202 The hemibrain sample is roughly 250 x 250 x 250  $\mu\text{m}$ , larger than we can FIB-SEM without  
203 introducing milling artifacts. Therefore we subdivided our epoxy-embedded samples into 20  $\mu\text{m}$   
204 thick slabs, both to avoid artifacts and allow imaging in parallel (each slab imaged in a different FIB  
205 machine) for increased throughput. To be effective, the cut surfaces of the slabs must be smooth at  
206 the ultrastructural level and have only minimal material loss. Specifically, for connectomic research,  
207 all long-distance processes must remain traceable across sequential slabs. We used an improved  
208 version of our previously published 'hot-knife' ultrathick sectioning procedure(Hayworth *et al.*,  
209 2015) which uses a heated, oil-lubricated diamond knife, to section the *Drosophila* brain into 37  
210 sagittal slabs of 20  $\mu\text{m}$  thickness with an estimated material loss between consecutive slabs of only  
211  $\sim 30$  nm - sufficiently small to allow tracing of long-distance neurites. Each slab was re-embedded,  
212 mounted, and trimmed, then examined in 3-D with X-ray tomography to check for sample quality  
213 and establish a scale factor for Z-axis cutting by FIB. The resulting slabs were FIB-SEM imaged  
214 separately (often in parallel, for increased throughput) and the resulting volume datasets were  
215 stitched together computationally.



**Figure 3.** The 13 slabs of the hemibrain, each flattened and co-aligned. Colors are arbitrary and added to the monochrome data to define the brain regions, as computed in section 2.5 .

216 Connectome studies come with clearly defined resolution requirements - the finest neurites must

217 be traceable by humans and should be reliably segmented by automated algorithms(*Januszewski*  
218 *et al., 2018*). In *Drosophila*, the very finest neural processes are usually 50 nm but can be as  
219 little as 15 nm(*Meinertzhagen, 2016a*). This fundamental biological dimension determines the  
220 minimum isotropic resolution requirements for tracing neural circuits. To meet the demand for high  
221 isotropic resolution and large volume imaging, we chose the FIB-SEM imaging platform, which offers  
222 high isotropic resolution (< 10 nm in x, y, and z), minimal artifacts, and robust image alignment.  
223 The high-resolution and isotropic dataset possible with FIB-SEM has substantially expedited the  
224 *Drosophila* connectome pipeline. Compared to serial-section imaging, with its sectioning artifacts  
225 and inferior Z-axis resolution, FIB-SEM offers high quality image alignment, a smaller number of  
226 artifacts, and isotropic resolution. This allows higher quality automated segmentation and makes  
227 manual proofreading and correction easier and faster.

228 At the beginning, deficiencies in imaging speed and system reliability of any commercial FIB-SEM  
229 system capped the maximum possible image volume to less than 0.01% of a full fly brain, problems  
230 that persist even now. To remedy them, we redesigned the entire control system, improved the  
231 imaging speed more than 10x, and created innovative solutions addressing all known failure modes,  
232 which thereby expanded the practical imaging volume of conventional FIB-SEM by more than four  
233 orders of magnitude from  $10^3 \mu\text{m}^3$  to  $3 \cdot 10^7 \mu\text{m}^3$ , while maintaining an isotropic resolution of  $8 \times 8 \times$   
234  $8 \text{ nm}$  voxels(*Xu et al., 2017*)(*Xu et al., 2019*). In order to overcome the aberration of a large field of  
235 view (up to  $300 \mu\text{m}$  wide), we developed a novel tiling approach without sample stage movement, in  
236 which the imaging parameters of each tile are individually optimized through an in-line auto focus  
237 routine without overhead(*Xu et al., 2018*). After numerous improvements, we have transformed  
238 the conventional FIB-SEM from a laboratory tool that is unreliable for more than a few days of  
239 imaging to a robust volume EM platform with effective long-term reliability, able to perform years  
240 of continuous imaging without defects in the final image stack. Imaging time, rather than FIB-SEM  
241 reliability, is now the main impediment to obtaining even larger volumes.

242 In our study here, the *Drosophila* "hemibrain", thirteen consecutive hot-knifed slabs were imaged  
243 using two customized enhanced FIB-SEM systems, in which an FEI Magnum FIB column was  
244 mounted at  $90^\circ$  upon a Zeiss Merlin SEM. After data collection, streaking artifacts generated by  
245 secondary electrons along the FIB milling direction were computationally removed using a mask  
246 in the frequency domain. The image stacks were then aligned using a customized version of the  
247 software platform developed for serial section transmission electron microscopy (*Zheng et al.,*  
248 *2018*)(*Khairy et al., 2018*), followed by binning along z-axis to form the final  $8 \times 8 \times 8 \text{ nm}^3$  voxel  
249 datasets. Milling thickness variations in the aligned series were compensated using a modified  
250 version of the method described by Hanslovsky et al.(*Hanslovsky et al., 2017*), with the absolute  
251 scale calibrated by reference to the MicroCT images.

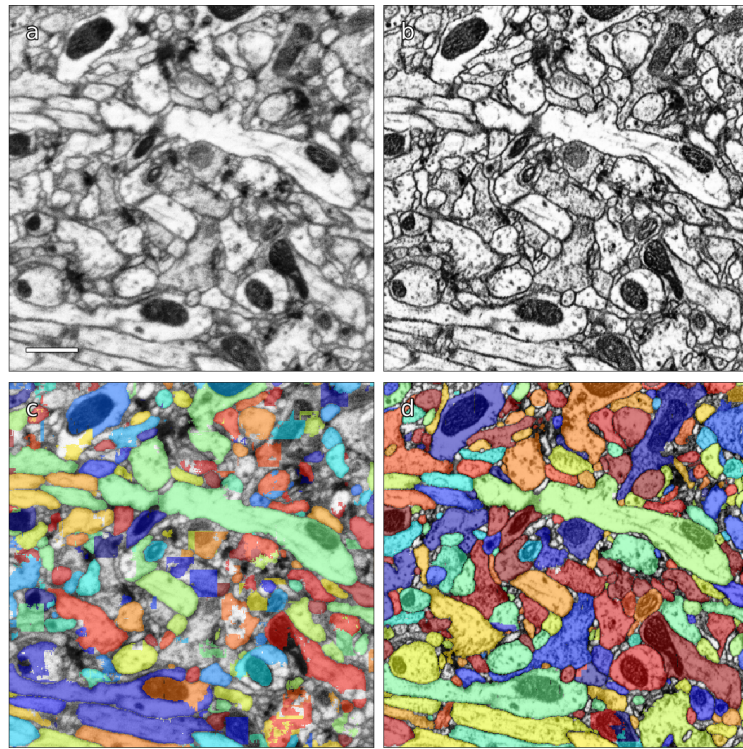
252 The  $20 \mu\text{m}$  slabs generated by the hot-knife sectioning were re-embedded in larger plastic tabs  
253 prior to FIB-SEM imaging. To correct for the warping of the slab that can occur in this process,  
254 methods adapted from Kainmueller(*Kainmueller et al., 2008*) were used to find the tissue-plastic  
255 interface and flatten each slab's image stack.

256 The series of flattened slabs was then stitched using a custom method for large scale deformable  
257 registration to account for deformations introduced during sectioning, imaging, embedding, and  
258 alignment (Saalfeld et al. in prep). These volumes were then contrast adjusted using slice-wise  
259 contrast limited adaptive histogram equalization (CLAHE)(*Pizer et al., 1987*), and converted into a  
260 versioned database(Distributed, Versioned, Image-oriented Database, or DVID), which formed the  
261 raw data for the reconstruction, as illustrated in Figure 3.

## 262 **Automated Segmentation**

263 Computational reconstruction of the image data was performed using flood-filling networks (FFNs)  
264 trained on roughly five-billion voxels of volumetric ground truth contained in two tabs of the  
265 hemibrain dataset(*Januszewski et al., 2018*). Initially, the FFNs generalized poorly to other tabs of  
266 the hemibrain, whose image content had a different appearance. Therefore we adjusted the image





**Figure 4.** (a) Original EM data from tab 34 at a resolution of 16 nm / resolution, (b) EM data after CycleGAN processing, (c-d) FFN segmentation results with the 16 nm model applied to original and processed data, respectively. Scale bar in (a) represents 1  $\mu\text{m}$ .

267 content to be more uniform using cycle-consistent generative adversarial networks (CycleGANs)(*Zhu*  
268 *et al., 2017*). Specifically, “generator” networks were trained to alter image content such that a  
269 second “discriminator” network was unable to distinguish between image patches sampled from, for  
270 example, a tab that contained volumetric training data versus a tab that did not. A cycle-consistency  
271 constraint was used to ensure that the image transformations preserved ultrastructural detail. The  
272 improvement is illustrated in Figure 4. Overall, this allowed us to use the training data from just two  
273 slabs, as opposed to needing training data for each slab.

274 FFNs were applied to the CycleGAN-normalized data in a coarse-to-fine manner at  $32 \times 32 \times 32$   
275  $\text{nm}^3$  and  $16 \times 16 \times 16 \text{ nm}^3$ , and to the CLAHE-normalized data at the native  $8 \times 8 \times 8 \text{ nm}^3$  resolution, in  
276 order to generate a base segmentation that was largely over-segmented. We then agglomerated the  
277 base segmentation, also using FFNs. We aggressively agglomerated segments despite introducing  
278 substantial numbers of erroneous mergers. This differs from previous algorithms, which studiously  
279 avoided merge errors since they were so difficult to fix. Here, advances in proofreading methodology  
280 described elsewhere in this report enabled efficient detection and correction of such mergers.

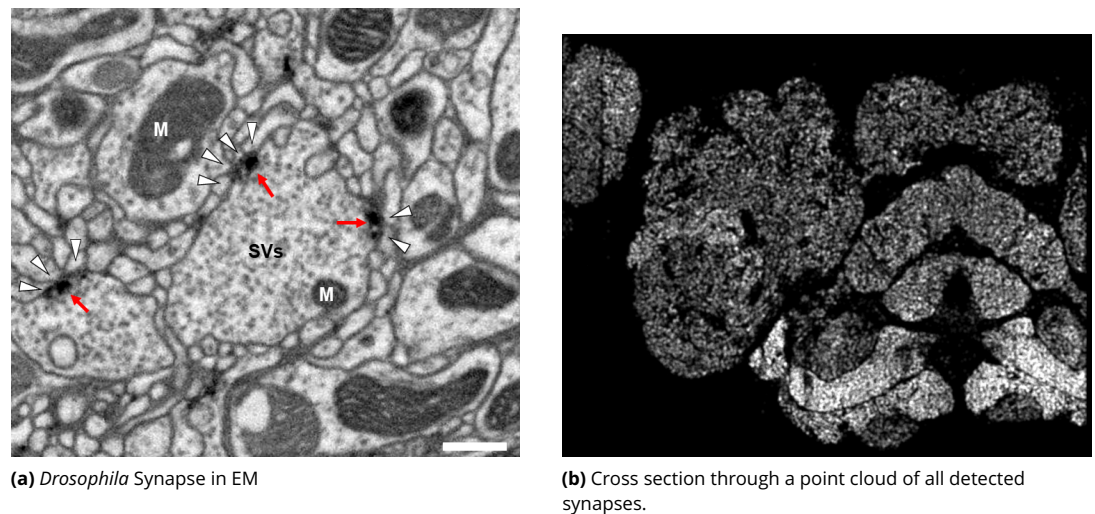
281 We evaluated the accuracy of the FFN segmentation of the hemibrain using metrics for expected  
282 run length (ERL) and merge rate(*Januszewski et al., 2018*). The base segmentation (i.e., the auto-  
283 mated reconstruction prior to agglomeration) achieved an ERL of 163  $\mu\text{m}$  with a merge rate of  
284 0.25%. After (automated) agglomeration, run length increased to 585  $\mu\text{m}$  but with a false merge  
285 rate of 27.6% (i.e., nearly 30% of the path length was contained in segments with at least one merge  
286 error). We also evaluated a subset of neurons in the volume, ~500 olfactory PN and KC cells chosen  
287 to roughly match the evaluation performed in (*Li et al., 2019*) which yielded an ERL of 825  $\mu\text{m}$  at a  
288 15.9% merge rate.



## 289 Synapse Prediction

290 Accurate synapse identification is central to our analysis, given that synapses form both a critical  
291 component of a connectome and are required for prioritizing and guiding the proofreading effort.  
292 Synapses in *Drosophila* are typically polyadic, with a single presynaptic site (a T-bar) contacted  
293 by multiple receiving dendrites (most with PSDs, postsynaptic densities) as shown in Figure 5a.  
294 Initial synapse prediction revealed that there are over 9 million T-bars and 60 million PSDs in the  
295 hemibrain. Manually validating each one, assuming a rate of 1000 connections annotated per  
296 trained person, per day, would have taken more than 230 working years. Given this infeasibility, we  
297 developed machine learning approaches to predict synapses as detailed below. The results of our  
298 prediction are shown in Fig 5b, where the predicted synapse sites clearly delineate many of the fly  
299 brain regions.

300 Given the size of the hemibrain image volume, a major challenge from a machine learning  
301 perspective is the range of varying image statistics across the volume. In particular, model per-  
302 formance can quickly degrade in regions of the data set with statistics that are not well-captured  
by the training set (Buhmann et al., 2019). To address this challenge, we took an iterative



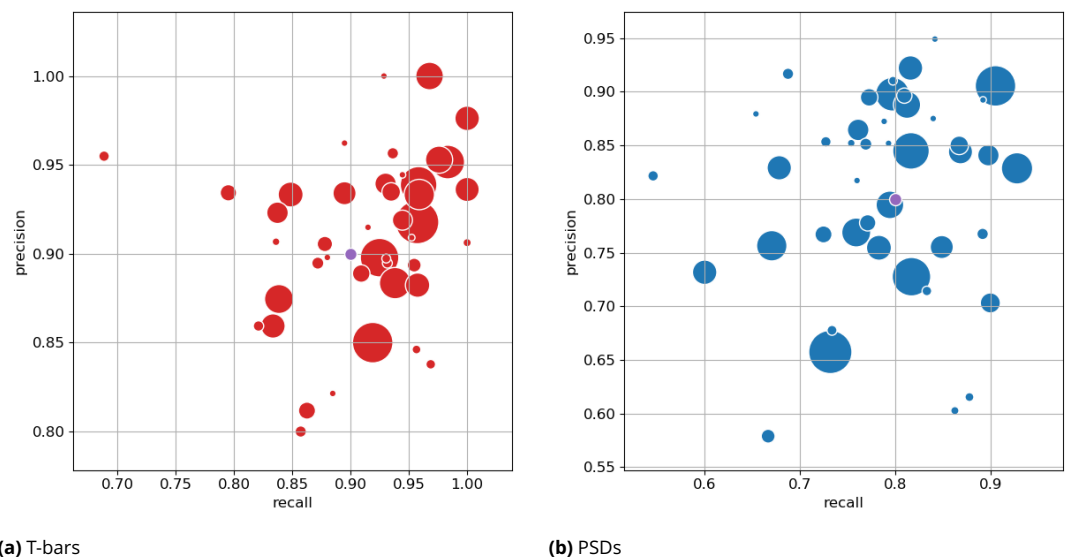
**Figure 5.** Well-preserved membranes, darkly stained synapses, and smooth round neurite profiles are characteristics of the hemibrain sample. Panel (a) shows polyadic synapses, with a red arrow indicating the presynaptic T-bar, and white triangles pointing to the postsynaptic densities. Mitochondria ('M'), synaptic vesicles ('SV'), and the scale bar (0.5  $\mu\text{m}$ ) are shown. Panel (b) shows a cross section through a point cloud of all detected synapses. This EM point cloud defines many of the compartments in the fly's brain, much like an optical image obtained using antibody nc82 antibody (against Bruchpilot, a component of T-bars) to stain synapses. This point cloud is used to generate the transformation from our sample to the standard *Drosophila* brain.

303  
304 approach to synapse prediction, interleaving model re-training with manual proofreading, all based  
305 on previously reported methods (Huang et al., 2018). Initial prediction, followed by proofreading,  
306 revealed a number of false positive predictions from structures such as dense core vesicles which  
307 were not well-represented in the original training set. A second filtering network was trained on  
308 regions causing such false positives, and used to prune back the original set of predictions. We  
309 denote this pruned output as the 'initial' set of synapse predictions.

310 Based on this initial set, we began collecting human-annotated dense ground-truth cubes  
311 throughout the various brain regions of the hemibrain, to assess variation in classifier performance  
312 by brain region. From these cubes, we determined that although many regions had acceptable  
313 precision, there were some regions in which recall was lower than desired. Consequently, a subset  
314 of cubes available at that time was used to train a new classifier focused on addressing recall in the

315 problematic regions. This new classifier was used in an incremental (cascaded) fashion, primarily by  
316 adding additional predictions to the existing initial set. This gave better performance than complete  
317 replacement using only the new classifier, with the resulting predictions able to improve recall while  
318 largely maintaining precision.

319 As an independent check on synapse quality, we also trained a separate classifier (Buhmann  
320 *et al.*, 2019), using a modified version of the 'synful' software package. Both synapse predictors give  
321 a confidence value associated with each synapse, a measure of how firmly the classifier believes the  
322 prediction to be a true synapse. We found that we were able to improve recall by taking the union of  
323 the two predictor's most confident synapses, and similarly improve precision by removing synapses  
324 that were low confidence in both predictions. Figures 6a and 6b show the results, illustrating the  
325 precision and recall obtained in each brain region.



**Figure 6.** Precision and recall for synapse prediction, on the left for T-bars, and on the right for synapses as a whole including the identification of PSDs. T-bar identification is better than PSD identification since this organelle is both more distinct and typically occurs in larger neurites. Each dot is one brain region. The size of the dot is proportional to the volume of the region. Humans proofreaders typically achieve 0.9 precision/recall on T-bars and 0.8 precision/recall on PSDs, indicated in purple.

### 326 Proofreading

327 Since machine segmentation is not perfect, we made a concerted effort to fix the errors remaining  
328 at this stage by several passes of human proofreading. Segmentation errors can be roughly grouped  
329 into two classes - "false merges", in which two separate neurons are mistakenly merged together,  
330 and "false splits", in which a single neuron is mistakenly broken into several segments. Enabled by  
331 advances in visualization and semi-automated proofreading using our Neu3 tool (Hubbard *et al.*,  
332 2020), we first addressed large false mergers. A human examined each putative neuron and  
333 determined if it had an unusual morphology suggesting that a merge might have occurred, a task  
334 still much easier for humans than machines. If judged to be a false merger, the operator identified  
335 discrete points that should be on separate neurons. The shape was then resegmented in real time  
336 allowing users to explore other potential corrections. Neurons with more complex problems were  
337 then scheduled to be re-checked, and the process repeated until few false mergers remained.

338 In the next phase, the largest remaining pieces were merged into neuron shapes using a  
339 combination of machine-suggested edits (Plaza, 2014) and manual intuition, until the main shape of  
340 each neuron emerged. This requires relatively few proofreading decisions and has the advantage

341 of producing an almost complete neuron catalog early in the process. As discussed below, in the  
342 section on validation, emerging shapes were compared against genetic/optical image libraries  
343 (where available) and against other neurons of the same putative type, to guard against large  
344 missing or superfluous branches. These procedures (which focused on higher-level proofreading)  
345 produced a reasonably accurate library of the main branches of each neuron, and a connectome  
346 of the stronger neuronal pathways. At this point there was still considerable variations among  
347 the brain regions, with greater completeness achieved in regions where the initial segmentation  
348 performed better.

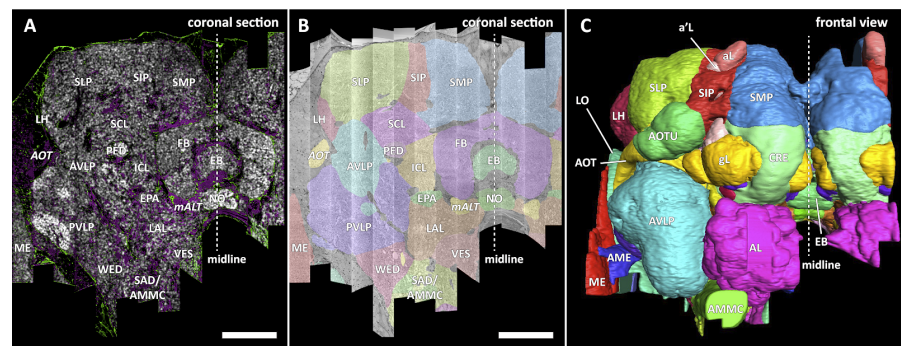
349 Finally, to achieve the highest reconstruction completeness possible in the time allotted,  
350 and to enable confidence in weaker neuronal pathways, proofreaders connected remaining iso-  
351 lated fragments (segments) to already constructed neurons, using NeuTu(Zhao *et al.*, 2018) and  
352 Neu3(Hubbard *et al.*, 2020). The fragments that would result in largest connectivity changes were  
353 considered first, exploiting automatic guesses through focused proofreading where possible. Since  
354 proofreading every small segment is still prohibitive, we tried to ensure a basic level of completeness  
355 throughout the brain with special focus in regions of particular biological interest such as the central  
356 complex and mushroom body.

### 357 Defining brain regions

358 In a parallel effort to proofreading, the sample was annotated with discrete brain regions. Our  
359 progression in mapping the cells and circuits of the fly's brain bears formal parallels to the history of  
360 mapping the earth, with many territories that are named and with known circuits, and others that  
361 still lack all or most of these. For the hemibrain dataset the regions are based on the brain atlas in  
362 Ito *et al.*(Ito *et al.*, 2014). The dataset covers most of the right hemisphere of the brain, except the  
363 optic lobe (OL), periesophageal neuropils (PENP) and gnathal ganglia (GNG), as well as part of the  
364 left hemisphere (Table 1). It covers about 36% of all synaptic neuropils by volume, and 54% of the  
365 central brain neuropils. We examined innervation patterns, synapse distribution, and connectivity  
366 of reconstructed neurons to define the neuropils as well as their boundaries on the dataset. We  
367 also made necessary, but relatively minor, revisions to some boundaries by reflecting anatomical  
368 features that had not been known during the creation of previous brain maps, while following the  
369 existing structural definitions(Ito *et al.*, 2014). We also used information from synapse point clouds,  
370 a predicted glial mask, and a predicted fiber bundle mask to determine boundaries of the neuropils  
371 (Figure 7 A). The brain regions of the fruit fly (Figure 7, B and C) include synaptic neuropils and  
372 non-synaptic fiber bundles. The non-synaptic cell body layer on the brain surface, which contains  
373 cell bodies of the neurons and glia, surrounds these structures. The synaptic neuropils can be  
374 further categorized into two groups: delineated and diffuse neuropils. The delineated neuropils  
375 have distinct boundaries throughout their surfaces, often accompanied by glial processes, and have  
376 clear internal structures in many cases. They include the antennal lobe (AL), bulb (BU), as well as  
377 the neuropils in the optic lobe (OL), mushroom body (MB), and central complex (CX). Remaining  
378 are the diffuse neuropils, sometimes referred to as *terra incognita*, since most have been less  
379 investigated than the delineated neuropils. In the previous brain atlas of 2014, boundaries of many  
380 *terra incognita* neuropils were rather arbitrarily determined, due to a lack of information then of  
381 their innervating neurons.

### 382 Diffuse (*terra incognita*) neuropils

383 In the hemibrain data, we adjusted the boundaries of some *terra incognita* neuropils using recon-  
384 structed neurons and their synaptic sites. Examples include the lateral horn (LH), ventrolateral  
385 neuropils (VLNP), and the boundary between the crepine (CRE) and lateral accessory lobe (LAL).  
386 The LH has been defined as the primary projection target of the olfactory projection neurons (PNs)  
387 from the antennal lobe (AL) via several antennal lobe tracts (ALTs)(Ito *et al.*, 2014)(Pereanu *et al.*,  
388 2010). The boundary between the LH and its surrounding neuropils is barely visible with synaptic  
389 immunolabeling such as nc82 or predicted synapse point clouds, as the synaptic contrast in these



**Figure 7.** Panel (A) A coronal section of the hemibrain dataset with synapse point clouds (white), predicted glial tissue (green), and predicted fiber bundles (magenta). (B) Grayscale image overlaid with segmented neuropils at the same level as (A). (C) A frontal view of the reconstructed neuropils. Scale bar: (A, B) 50 $\mu$ m.

390 regions is minimal. The olfactory PNs can be grouped into several classes, and the projection sites  
391 of the uniglomerular PNs that project through the medial ALT (mALT), the thickest fiber bundle  
392 between the AL and LH, give the most conservative and concrete boundary of the ‘core’ LH (Figure  
393 8A). Multiglomerular PNs, on the other hand, project to much broader regions, including the vol-  
394 umes around the core LH (Figure 8B). These regions include areas which are currently considered  
395 parts of the superior lateral protocerebrum (SLP) and posterior lateral protocerebrum (PLP). Since  
396 the “core” LH roughly approximates the shape of the traditional LH, and the boundaries given  
397 by the multiglomerular PNs are rather discrete, in this study we assumed the core to be the LH  
398 itself. Of course, the multiglomerular PNs convey olfactory information as well, and therefore the  
399 neighboring parts of the SLP and PLP to some extent also receive inputs from the antennal lobe.  
400 These regions might be functionally distinct from the remaining parts of the SLP or PLP, but they  
401 are not explicitly separated from those neuropils in this study.

402 The VLNP is located in the lateral part of the central brain and receives extensive inputs from  
403 the optic lobe through various types of the visual projection neurons (VPNs). Among them, the  
404 projection sites of the lobula columnar (LC), lobula plate columnar (LPC), lobula-lobula plate colum-  
405 nar (LLPC), and lobula plate-lobula columnar (LPLC) cells form characteristic glomerular structures,  
406 or the optic glomeruli (OG), in the AOTU, PVLVP, and PLP (Klapoetke *et al.*, 2017)(Otsuna and Ito,  
407 2006)(Panser *et al.*, 2016)(Wu *et al.*, 2016). We exhaustively identified columnar VPNS and found 23  
408 types of LC, two types of LPC, three types of LLPC, and three types of LPLC cells. The glomeruli of  
409 these pathways were used to determine the medial boundary of the PVLVP and PLP, following existing  
410 definitions (Ito *et al.*, 2014), except for a few LC types which do not form glomerular terminals. The  
411 terminals of the reconstructed LC cells and other lobula complex columnar cells (LPC, LLPC, LPLC)  
412 are shown in Figures 8C and 8D, respectively.

413 In the previous paper (Ito *et al.*, 2014), the boundary between the CRE and LAL was defined as  
414 the line roughly corresponding to the posterior-ventral surface of the MB lobes, since no other  
415 prominent anatomical landmarks were found around this region. In this dataset, we found several  
416 glomerular structures surrounding the boundary both in the CRE and LAL. These structures include  
417 the gall (GA), rubus (RUB), and round body (ROB). Most of them turned out to be projection targets  
418 of several classes of central complex neurons, implying the ventral CRE and dorsal LAL are closely  
419 related in their function. We re-determined the boundary so that each of the glomerular structures  
420 would not be divided into two, while keeping the overall architecture and definition of the CRE and  
421 LAL. The updated boundary passes between the dorsal surface of the GA and the ventral edge of  
422 the ROB. Other glomerular structures, including the RUB, are included in the CRE.



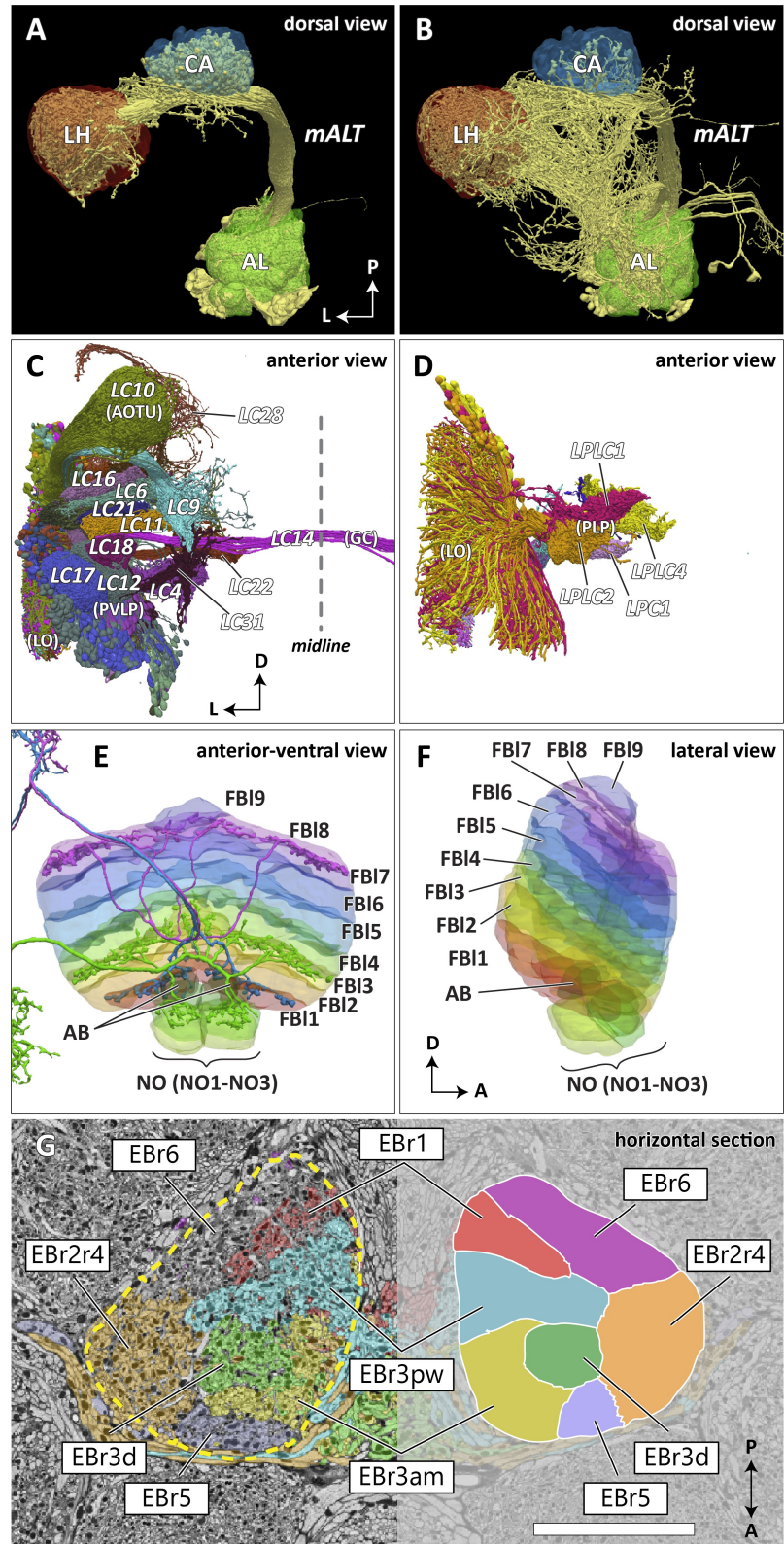


Figure 8. Caption next page.



**Figure 8.** Reconstructed brain regions and substructures. (A, B) Dorsal views of the olfactory projection neurons (PNs) and the innervated neuropils, AL, CA, and LH. Uniglomerular PNs projecting through the mAL are shown in (A), and multiglomerular PNs are shown in (B). (C, D) Columnar visual projection neurons. Each subtype of cells is color-coded. LC cells are shown in (C), and LPC, LLPC, and LPLC cells are shown in (D). (E, F) The nine layers of the fan-shaped body (FB), along with the asymmetrical bodies (AB) and the noduli (NO), displayed as an anterior-ventral view (E), and a lateral view (F). In (E), three FB tangential cells (FB1D (blue), FB3A (green), FB7L (purple)) are shown as markers of the corresponding layers (FBI1, FBI3, and FBI7, respectively). (G) Zones in the ellipsoid body (EB) defined by different types of ring neurons. In this horizontal section of the EB, the left side shows the original grayscale data, and the seven ring neuron subtypes are color-coded. The right side displays the seven segmented zones based on the innervation pattern. Scale bar: 20 $\mu$ m.

## 423 Delineated neuropils

424 Substructures of the delineated neuropils have also been added to the brain region map in the  
425 hemibrain. The asymmetrical bodies (AB) were added as the fifth independent neuropil of the  
426 CX(Wolff and Rubin, 2018). The AB is a small synaptic volume adjacent to the ventral surface of the  
427 fan-shaped body (FB) that has historically been included in FB(Ito et al., 2014). The AB has been  
428 described as a fasciculin II (fasII)-positive structure that exhibits left-right structural asymmetry  
429 by Pascual et al.(Pascual et al., 2004), who reported that most flies have their AB only in the right  
430 hemisphere, while a small proportion (7.6%) of wild type flies have their AB on both sides. In  
431 the hemibrain dataset, a pair of ABs is situated on both sides of the midline, but the left AB is  
432 notably smaller than the right AB (right: 1,467 $\mu$ m<sup>3</sup>, left: 452  $\mu$ m<sup>3</sup>), still showing an obvious left-right  
433 asymmetry. The AB is especially strongly connected to the neighboring neuropil, the FB, by neurons  
434 including Delta0A, Delta0B, and Delta0C, while it also houses postsynaptic terminals of the CX  
435 output neurons including FQ12a(Wolff and Rubin, 2018). While these anatomical observations  
436 imply that the AB is part of the central body (CB), along with the FB and the ellipsoid body (EB), this  
437 possibility is neither developmentally nor phylogenetically proven.

438 The round body (ROB) is also a small round synaptic structure situated on the ventral limit  
439 of the crepine (CRE), close to the  $\beta$  lobe of the MB (Lin et al., 2013)(Wolff and Rubin, 2018). It  
440 is a glomerulus-like structure and one of the foci of the CX output neurons, including the PFR  
441 (protocerebral bridge – fan-shaped body – round body) neurons. It is classified as a substructure  
442 of the CRE along with other less-defined glomerular regions in the neuropil, many of which also  
443 receive signals from the CX. Among these, the most prominent one is the rubus (RUB). These are  
444 two distinct structures; the RUB is embedded completely within the CRE, while the ROB is located  
445 on the ventrolateral surface of the CRE. The lateral accessory lobe (LAL), neighboring the CRE,  
446 also houses similar glomerular terminals, and the gall (GA) is one of them. While the ROB and  
447 GA have relatively clear boundaries separating them from the surrounding regions, they may not  
448 qualify as independent neuropils because of their small size and the structural similarities with the  
449 glomerulus-like terminals around them. They may be comparable with other glomerular structures  
450 such as the AL glomeruli and the optic glomeruli in the lateral protocerebrum, both of which are  
451 considered as substructures of the surrounding neuropils.

452 Substructures of independent neuropils are also defined using neuronal innervations. The  
453 five MB lobes on the right hemisphere are further divided into 15 compartments ( $\alpha$ 1-3,  $\alpha$ '1-3,  $\beta$ 1-  
454 2,  $\beta$ '1-2, and  $\gamma$ 1-5)(Tanaka et al., 2008)(Aso et al., 2014) by the mushroom body output neurons  
455 (MBONs) and dopaminergic neurons (DANs). Our compartment boundaries were defined by  
456 approximating the innervation of these neurons. Although the innervating regions of the MBONs  
457 and DANs do not perfectly tile the entire lobes, the compartments have been defined to tile the  
458 lobes, so every synapse in the lobes belongs to one of the 15 compartments. The FB is subdivided  
459 into nine horizontal layers (FBI1-9) (Figure 8E and 8F) as already illustrated(Wolff et al., 2015).  
460 They are determined by the pattern of innervation of 480 FB tangential cells, which form nine  
461 groups depending on the dorsoventral levels they innervate in the FB. While neurons innervating  
462 neighboring layers may overlap slightly, the layer boundaries were drawn so that the coverage of

463 the tangential arbors by each layer was maximized.

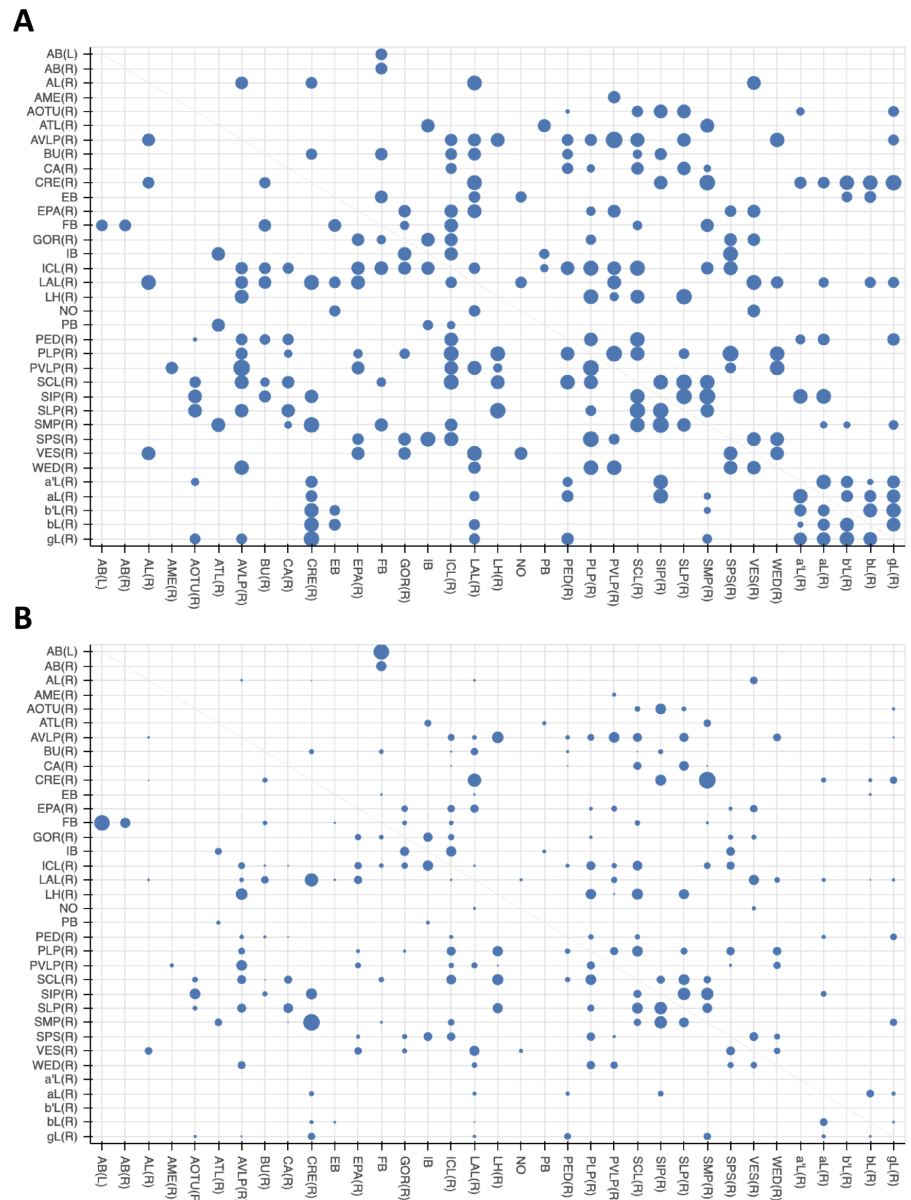
464 The EB is likewise subdivided into zones by the innervating patterns of the EB ring neurons, the  
465 most prominent class of neurons innervating the EB. The ring neurons have six subtypes, R1-R6,  
466 and each projects to specific zones of the EB. Among them, the regions innervated by R2 and R4 are  
467 mutually exclusive but highly intermingled, so these regions are grouped together into a single zone  
468 (EBr2r4). R3 has the most neurons among the ring neuron subtypes and is further grouped into  
469 five subclasses. While each subclass projects to a distinct part of the EB, the innervation patterns  
470 of the subclasses R3a and R3m, and also R3p and R3w, are very similar to each other. The region  
471 innervated by R3 is, therefore, subdivided into three zones, including EBr3am, EBr3pm, and EBr3d.  
472 Along with the other three zones, EBr1, EBr5, and EBr6, the entire EB is subdivided into seven  
473 non-overlapping zones (Figure 8G). Unlike other zones, EBr6 is innervated only sparsely by the R6  
474 cells, and the space mainly filled by synaptic terminals of other neuron types, including the extrinsic  
475 ring neurons (ExR). Omoto et al. (Omoto et al., 2017) segmented the EB into five domains (EBa, EBoc,  
476 EBop, EBic, EBip) by the immunolabeling pattern of DN-cadherin, and each type of the ring neurons  
477 may innervate more than one domain in the EB. Our results show that the innervation pattern of  
478 each ring neuron subtype is highly compartmentalized at the EM level and the entire neuropil can  
479 be sufficiently subdivided into zones based purely on the neuronal morphologies. The neuropil  
480 may be subdivided differently if other neuron types, such as the extrinsic ring neurons (ExR) (Omoto  
481 et al., 2018), are recruited as landmarks.

#### 482 Quality of the brain region boundaries

483 Since many of the *terra incognita* neuropils are not clearly partitioned from each other by solid  
484 boundaries such as glial walls, it is important to evaluate if the current boundaries reflect anatomical  
485 and functional compartments of the brain. We first measured the relative sizes of the boundaries  
486 between any two adjacent neuropil regions (Figure 9A). The map shows results for brain regions that  
487 are over 75% in the hemibrain region, restricted to right regions with exception to the asymmetric  
488 AB(L). For these regions, we counted the number of wire crossings by large traced neurons and  
489 estimated a cost. A bigger dot indicates a higher cost or a less clean boundary. We do not penalize  
490 neurons that cross a boundary once, but rather penalize when a neuron crosses the same boundary  
491 multiple times. By restricting our analysis to the right part of the hemibrain, we hopefully minimize  
492 the effect of smaller, traced-but-truncated neuron fragments on our score. Figure 9B shows the  
493 number of intersections normalized by the area of boundary. We spot checked many of the  
494 instances and in general note that the brain regions with a high cost, such as those in SNP, INP  
495 and VLNP, tend to have less well defined boundaries. In particular, the boundaries at SMP/CRE,  
496 CRE/LAL, SMP/SIP, and SIP/SLP have worse scores, indicating these boundaries may not reflect  
497 actual anatomical and functional segregation of the neuropils. These brain regions were defined  
498 based on the arborization patterns of characteristic neuron types, but because neurons in the *terra*  
499 *incognita* neuropils tend to be rather heterogeneous, there are many other neuron types that do  
500 not follow these boundaries. The boundaries between the FB and AB also give relatively bad scores,  
501 and this suggests that the AB is tightly linked to the neighboring FB.

#### 502 Insights for a whole-brain remapping

503 The current brain regions based on Ito et al. (Ito et al., 2014) contain a number of arbitrary  
504 determinations of brain regions and their boundaries in the *terra incognita* neuropils. In this study,  
505 we tried to solidify the ambiguous boundaries as much as possible using the information from  
506 the reconstructed neurons. However, large parts of the left hemisphere and the subesophageal  
507 zone (SEZ) are missing from the hemibrain dataset, and neurons innervating these regions are not  
508 sufficiently reconstructed. This incompleteness of the dataset is the main reason that we did not  
509 alter the previous map drastically and kept all the existing brain regions even if their anatomical  
510 and functional significance is not obvious. Once a complete EM volume of the whole fly brain is  
511 imaged and most of its 100,000 neurons are reconstructed, the entire brain can be re-segmented



**Figure 9.** Quality check of the brain compartments. (A) The relative sizes of the boundaries between adjacent neuropils indicated in a log scale. (B) The number of neuronal intersections normalized by the area of neuropil boundary.

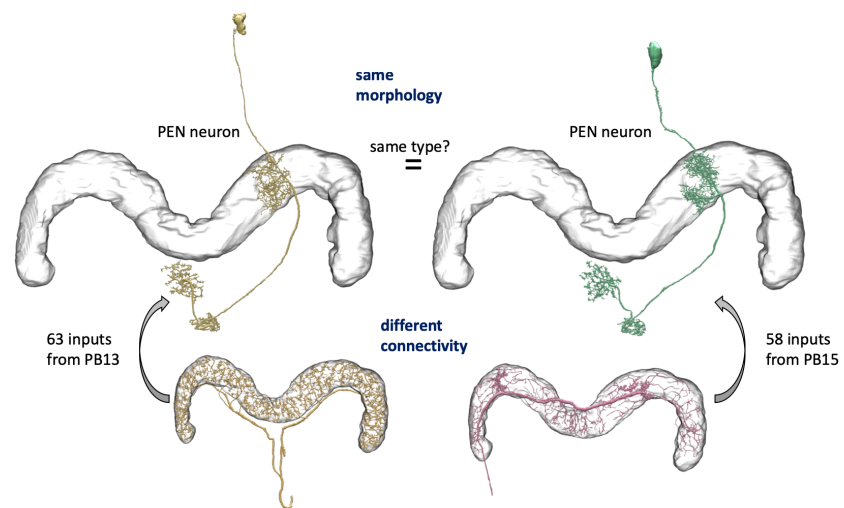
512 from scratch with more comprehensive anatomical information. Arbitrary or artificial neuropil  
 513 boundaries will thereby be minimized, if not avoided, in a new brain map. Anatomy-based neuron  
 514 segmentation strategies such as NBLAST may be used as neutral methods to revise the neuropils  
 515 and their boundaries. Any single method, however, is not likely to produce consistent boundaries  
 516 throughout the brain, especially in the *terra incognita* regions. It may be necessary to use different  
 517 methods and criteria to segment the entire brain into reasonable brain regions. Such a new map  
 518 would need discussion in a working group, and approval from the community in advance (as did the  
 519 previous map(Ito et al., 2014)), insofar as it would replace the current map and therefore require a  
 520 major revision of the neuron mapping scheme.

## 521 Cell Type Classification

522 Defining cell types for groups of similar neurons is a time-honored means to attempt to understand  
523 the anatomical and functional properties of a circuit. Presumably, neurons of the same type execute  
524 similar circuit roles. However, the definition of what is a distinct cell type and the exact delineation  
525 between one cell type and another is inherently vague and represents a classic taxonomic challenge,  
526 pitting 'lumpers' vs 'splitters'. Despite our best efforts, we recognize that our typing of cells is not  
527 exact, and expect future revisions to cell type classification.

528 One common method of cell type classification, used in flies, exploits the GAL4 system to  
529 highlight the morphology of neurons having similar gene expression (Jenett *et al.*, 2012). Since  
530 these genetic lines are imaged using fluorescence and confocal microscopy, we refer to them  
531 as 'light lines'. Where they exist and are sufficiently sparse, light lines provide a key method for  
532 identifying types by grouping morphologically similar neurons together. However, there are several  
533 limitations. There are no guarantees of coverage, and it is sometimes difficult to distinguish between  
534 neurons of very similar morphology but different connectivity.

535 We enhanced the classic view of morphologically distinct cell types by defining distinct cell types  
536 (or sub-cell types) based on morphology and connectivity. Connectivity-based clustering often  
537 serves a clear arbiter of cell type distinctions, even when genetic markers have yet to be found,  
538 or when the morphology of different types is quite similar, sometimes sufficiently similar to be  
539 indistinguishable in optical images. For example, the two PEN (protocerebral bridge - ellipsoid body  
540 - noduli) neurons have very similar forms but quite distinct inputs (Figure 10) (Turner-Evans *et al.*,  
541 2019). Confirming their differences, PEN1 and PEN2 neurons, in fact, have been shown to have  
542 different functional activity (Green *et al.*, 2017).



**Figure 10.** An example of two neurons with very similar shapes but differing connectivities.

543 Based on our previous definition of cell type, many neurons exhibit a unique morphology or  
544 connectivity pattern at least within one hemisphere of the brain (presumably with a matching  
545 type in the other hemisphere). Therefore, in our hemibrain reconstruction, many neuron types  
546 consisting of a distinct morphology and connectivity have only a single example. It is possible  
547 in principle to provide coarser groupings of neurons. For instance, most cell types are grouped  
548 by their cell body fiber representing a distinct clonal unit, which we discuss in more detail below.  
549 Furthermore, each neuron can be grouped with neurons that innervate similar brain regions. In  
550 this paper, we do not explicitly formalize this higher-level grouping, but data on the innervating  
551 brain regions can be readily mined from the dataset.

## 552 **Methodology for assigning cell types and nomenclature**

553 Assigning names and types to the more than 20,000 reconstructed cells was a difficult and con-  
554 tentious undertaking. Many of the neurons have no previously annotated type. Adding to the  
555 complexity, prior work focused on morphological similarities and differences, but here we have, for  
556 the first time, connectivity information to assist in cell typing as well.

557 Most cell types for the visual projection neurons (VPNs), mushroom body (MB) neurons and  
558 central complex (CX) neurons are already described in the literature, but the existing names can be  
559 both inconsistent and ambiguous. The same cell type is often given differing names in different  
560 publications, and conversely, the same name, such as PN for projection neuron, is used for many  
561 different cell types. Nonetheless, for cell types already named in the literature (which we designate  
562 as famous cell types), we have tried to use an existing name. We apologize in advance for any  
563 offense given by our selection of names.

564 Overall, we defined a 'type' of neurons as either a single cell or a group of cells that have a  
565 very similar cell body location, morphology, and pattern of synaptic connectivity. We found 18,478  
566 neuronal cell bodies in the hemibrain volume, most of which are located in the right side of the  
567 brain.

568 We classified these neurons in a few steps. The first step classified all cells by their lineage,  
569 grouping neurons according to their bundle of cell body fibers (CBFs). Neuronal cell bodies are  
570 located in the cell body layer that surrounds the brain, and each neuron projects a single CBF  
571 towards a synaptic neuropil. In the central brain, cell bodies of clonally related neurons deriving  
572 from a single stem cell tend to form clusters, from each of which arises one or several bundles  
573 of CBFs. We carefully examined the trajectory and origins of CBFs of the 15,532 neurons on the  
574 right central brain and identified 192 distinct CBF bundles. Among them, 154 matched the CBF  
575 bundles of 102 known clonal units (*Ito et al., 2013*) (*Lin et al., 2013*). The rest are minor populations  
576 and most likely of embryonic origin.

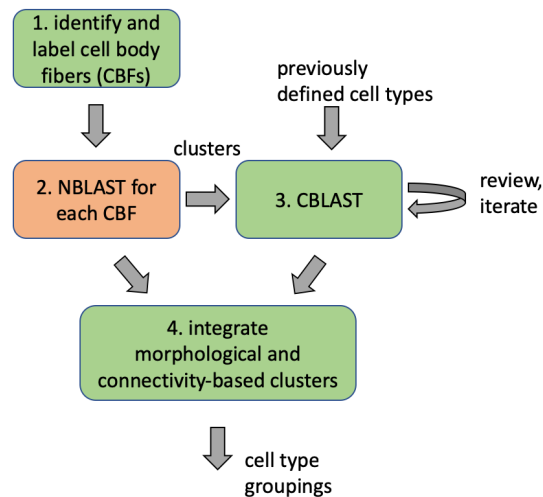
577 Different stem cells sometimes give rise to neurons with very similar morphologies. We classified  
578 these as different types because of their distinct developmental origin and slightly different locations  
579 of their cell bodies and CBFs. Thus, the next step in neuron typing was to cluster neurons within  
580 each CBF group. This process consisted of three further steps. First, we used NBLAST (*Costa et al.,*  
581 **2016**) to subject all the neurons of a particular CBF group to morphology-based clustering. Next,  
582 we used CBLAST, a new tool to cluster neurons based on synaptic connectivity (see below). This  
583 step is an iterative process, using neuron morphology as a template, to regroup neurons after  
584 more careful examination of neuron projection patterns and their connections. Finally, we validated  
585 the cell typing with extensive manual review and visual inspection. This review both allowed us to  
586 confirm cell type identity and help ensure neuron reconstruction accuracy.

587 In the hemibrain, using the defined brain regions and reference to known expression driver  
588 strains, we were able to assign a cell type to many cells. Where possible, we matched previously  
589 defined cell types with those labeled in light data using a combination of Neuprint, an interactive  
590 analysis tool (described below), and human recognition to find the matching cell types, especially  
591 in well explored neuropils such as the mushroom body (MB) and central complex (CX), where  
592 abundant cell type information was already available and where we are more confident in our  
593 anatomical expertise. Even though most of the cell types in the MB and CX were already described,  
594 we still found new cell types in these regions, an important vindication of our methods. In these  
595 cases we tried to name them using the existing schemes for these regions, and further refined  
596 these morphological groupings with relevant information on connectivity.

597 Outside the heavily studied regions, the fly's circuits are largely composed of cells of unknown  
598 type. In this case putative type names were derived from a) the CBF group, b) the morphological  
599 type, and c) the connectivity type.

- 600 · Each of the 192 CBF bundles was given an ID according to the location of the cell body  
601 cluster (split into eight sectors of the brain surface with the combination of Anterior/Posterior,





**Figure 11.** Workflow for defining cell types

602 Ventral/Dorsal, and Medial/Lateral) and a number within the sector given according to the  
603 size of cell population. Thus, a CBF group might be named ADM01, meaning a group with the  
604 largest number of neurons in the Anterior Dorsal Medial sector of the brain's surface.

- 605 • Morphological types were represented by the CBF group name followed by 1-3 lowercase  
606 letters, e.g. ADM01a.
- 607 • If neurons of near-identical morphology could be further subdivided into different connectivity  
608 types, they were suffixed with an underscore and a lowercase letter, e.g. ADM01a\_b.

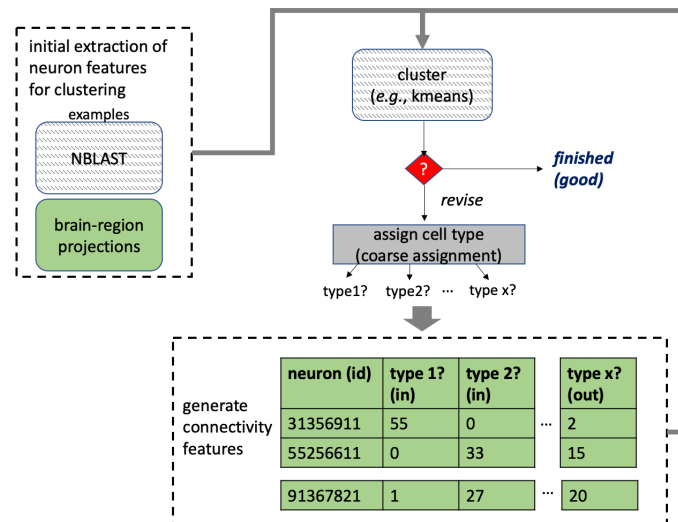
609 Finally, a suffix '\_pct', for putative cell type, was added. Thus, a full putative type name might be  
610 'ADM01a\_pct' if all the neurons of this type shared similar connectivity patterns, or 'ADM01b\_a\_pct'  
611 and 'ADM01b\_b\_pct' if there are different connectivity types within neurons having a similar form. The  
612 resulting names may lack elegance, but the process is systematic and scalable.

613 The assignment of type names to neurons is still ongoing, and we expect the names of putative  
614 cell types will be refined by the research community, including simpler names that are easier to  
615 pronounce, as new information emerges. What will not change are the unique body ID numbers  
616 given in the database that refer to a particular (traced) cell in this particular image dataset. We  
617 strongly advise that such IDs be included in any publications based on our data to avoid confusion  
618 as cell type names (and possibly instance names) evolve.

### 619 CBLAST

620 As part of our effort to assign cell types, we built a tool for cell type clustering based on neuron  
621 connectivity, called CBLAST (by analogy with the existing NBLAST(Costa *et al.*, 2016), which forms  
622 clusters based on the shapes of neurons). The tool is described in more detail in Figure 12.

623 Partitioning a network into clusters of nodes that exhibit similar connectivity is known as  
624 community detection or graph clustering(Fortunato and Hric, 2016). Numerous methods have  
625 been proposed for selecting such partitions, the best known being the stochastic block model. To  
626 non-theoreticians, the process by which most methods choose a partitioning is not intuitive, and  
627 the results are not easily interpretable. Furthermore, most approaches do not readily permit a  
628 domain expert to guide the partitioning based on her intuition or on other features of the nodes  
629 that are not evident in the network structure itself. In contrast, CBLAST is based on traditional data  
630 clustering concepts, leading to more intuitive results. Additionally, a user can apply their domain  
631 expertise by manually refining the partitioning during successive iterations of the procedure. This is  
632 especially useful in the case of a network like ours, in which noise and missing data make it difficult



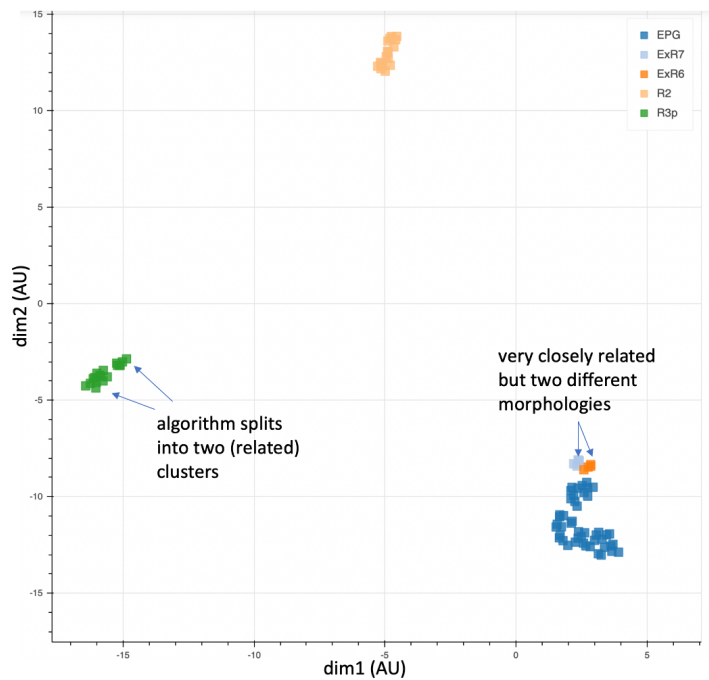
**Figure 12.** Overview of the operation of CBLAST

633 to rely solely on connectivity to find a good partitioning automatically. Additionally, other graph  
 634 clustering methods do not accommodate the notion of left-right symmetry amongst communities,  
 635 a feature that is critical for assigning cell types in a connectome.

636 CBLAST clusters neurons together using a similarity feature score defined by how the neuron  
 637 distributes inputs and outputs to different neuron types. However, this is a circular requirement  
 638 since neuron types must already be defined to use this technique. CBLAST therefore uses an  
 639 iterative approach, refining cell type definitions successively. Initial cell type groups are putatively  
 640 defined using an initial set of features based on morphological overlap as in NBLAST and/or based  
 641 on the distribution of inputs and outputs in defined brain regions. These initial groups are fed  
 642 into CBLAST in which the user can visualize and analyze the results using plots such as that in  
 643 Figure 13. Given the straightforward similarity measure, the user can look at the input and output  
 644 connections for each neuron to better understand the decision made by the clustering algorithm. As  
 645 the definitions of cell type definitions are improved, the clustering becomes more reliable. In some  
 646 cases, this readily exposes incompleteness (e.g., due to the boundary of the hemibrain sample)  
 647 in some neurons which would complicate clustering even for more computationally intensive  
 648 strategies such as a stochastic block model. Based on these interactions, the user makes decisions  
 649 and refines the clusters manually, iterating until further changes are not observed.

650 Our large, dense connectome is a key requirement for CBLAST. Unless a significant fraction of  
 651 a neuron's inputs and outputs is known, neurons that are in fact similar may not cluster together  
 652 correctly. This requirement is not absolute, as we note that CBLAST is often able to match left and  
 653 right symmetric neurons, despite some of these left side neurons being truncated by the boundaries  
 654 of the dataset. Nonetheless, reconstruction incompleteness and any noise in the reconstruction  
 655 can contribute to noise in clustering results.

656 CBLAST usually generates clusters that are consistent with the morphological groupings of the  
 657 neurons, with CBLAST often suggesting new sub-groupings as intended. This agreement serves as  
 658 some validation of the concepts behind CBLAST. In some cases it can be preferable to NBLAST, since  
 659 the algorithm is less sensitive to exact neuron location, and for many applications the connectivity  
 660 is more important than the morphology. In Figure 13, we show the results of using CBLAST on a few  
 661 neuron types extracted from the ellipsoid body. The clusters are consistent with the morphology,  
 662 with exception to a new sub-grouping for R3p being suggested as a more distinct group than type  
 663 ExR7/ExR6.



**Figure 13.** Cells of five types plotted according to their connectivities. Coordinates are in arbitrary units after dimensionality reduction using UMAP (McInnes *et al.*, 2018). The results largely agree with those from morphological clustering but in some cases show separation even between closely related types.

## 664 Results of cell typing

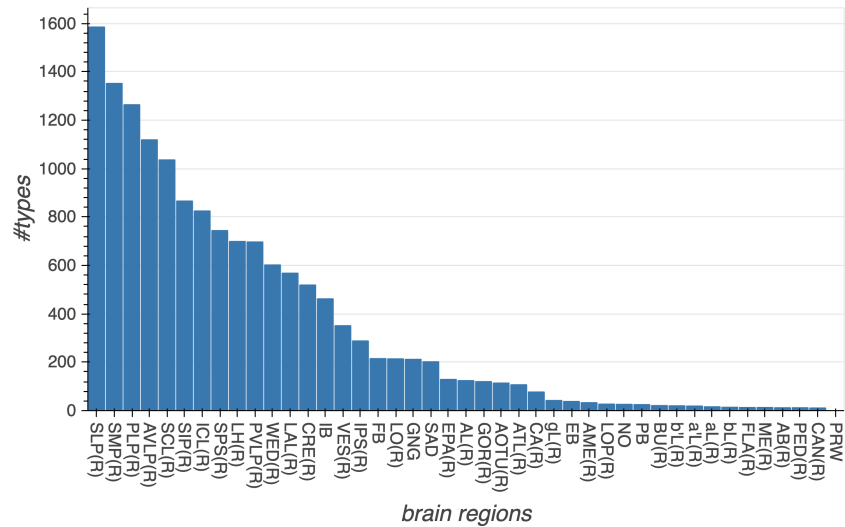
Brain Region(s)	Number of types	Number of cells
Visual Projection Neurons (VPNs)	55	3,329
Antennal Lobe	159	2,439
Mushroom Body	68	2,117
Central Complex	264	2,826
Others	4,335	10,190
Total	4,890	20,901

**Table 2.** Cell types and cell counts for well known brain regions, and totals

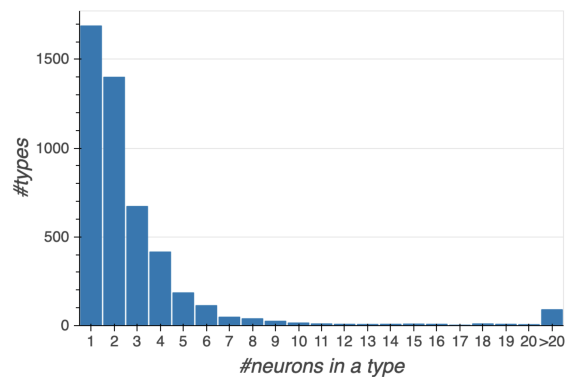
665 Using the above semi-automated procedures, we identified 55 types for VPNs, 159 types in  
 666 the antennal lobe (AL), 68 types in MB, and 264 types in CX, which in aggregate apply to a total of  
 667 10,734 neurons (note that cells in CX are counted for both right and left sides) (Table 2). For the  
 668 remaining  $\approx 10,000$  neurons in the other brain regions, over 4000 cell types were identified. Over a  
 669 thousand of these are types with only a single instance, although presumably, for a whole brain  
 670 reconstruction, most of these types would have partners on the opposite side of the brain. Figure  
 671 14 shows the number of distinct neuron types found in different brain regions. Figure 15 shows the  
 672 distribution of the number of neurons in each cell type.

## 673 Assessing Morphologies and Cell Types

674 Verifying correctness and completeness in these data is a challenging problem because no existing  
 675 full brain connectome exists against which our data might be compared. We devised a number  
 676 of tests to check the main features: Are the morphologies correct? Are the regions and cell types  
 677 correctly defined? Are the synaptic connection counts representative?



**Figure 14.** The number of cell types in each major brain region. The sum of cell types in the graph is larger than the total number of cell types, because a single cell type may contribute to many regions.



**Figure 15.** Histogram showing the number of cell types with a given number of constituent cells.

678 Assessing completeness is much easier than assessing correctness. Since the reconstruction is  
679 dense, we believe the census of cells, types, and regions should be essentially complete. The main  
680 arbors of every cell within the volume are reconstructed, and almost every cell is assigned to at  
681 least a putative cell type. Similarly, since the identified brain regions nearly tile the entire brain,  
682 these are complete as well.

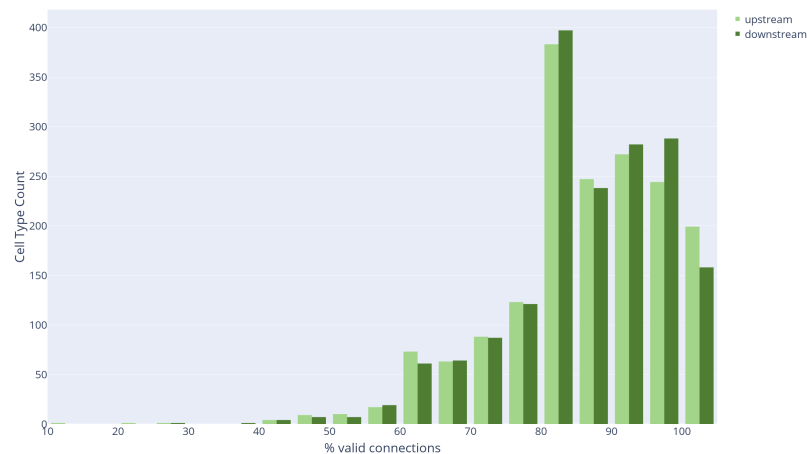
683 For checking morphologies, we searched for major missing or erroneous branches using a  
684 number of heuristics. Each neuron was reviewed by multiple proofreaders. The morphology of  
685 each neuron was compared with light microscopy data whenever it was available. When more than  
686 one cell of a given type was available (either left and right hemisphere, or multiple cells of the same  
687 type in one hemisphere), a human examined and compared them. This helped us find missing  
688 or extra branches, and also served as a double check on the cell type assignment. In addition,  
689 since the reconstruction is dense, all sufficiently large “orphan” neurites were examined manually  
690 until they were determined to form part of a neuron, or they left the volume. To help validate the  
691 assigned cell types, proofreaders did pairwise checks of every neuron with types that had been  
692 similarly scored.

693 For subregions in which previous dense proofreading was available (such as the alpha lobes  
694 of the mushroom body) we compared the two connectomes. We were also helped by research

695 groups using both sparse tracing in the full fly brain TEM dataset (Zheng *et al.*, 2018), and our  
696 hemibrain connectome. They were happy to inform us of any inconsistencies. There are limits  
697 to this comparison, as the two samples being compared were of different ages and raised under  
698 different conditions, then prepared and imaged by different techniques, but this comparison would  
699 nevertheless have revealed any gross errors. Finally, we generated a 'probabilistic connectome'  
700 based on a different segmentation, and systematically visited regions where the two versions  
701 differed.

## 702 **Assessing Synapse Accuracy**

703 As discussed in the section on finding synapses, we evaluated both precision (the fraction of found  
704 synapses that are correct) and recall (fraction of true synapses that were correctly predicted) on  
705 sample cubes in each brain region. We also double checked by comparing our findings with a  
706 different, recently published, synapse detection algorithm (Buhmann *et al.*, 2019).



**Figure 16.** Connection precision of upstream and downstream partners for  $\approx 1000$  cell types.

707 As a final check, we also evaluated the end-to-end correctness of given connections between  
708 neurons for different cell types and across brain regions. Specifically, for each neuron, we sampled  
709 25 upstream connections (T-bar located within the neuron) and 25 downstream connections (PSD  
710 located within the neuron), and checked whether the annotations were correct, meaning that the  
711 pre/post annotation was valid and assigned to the correct neuron.

712 In total, we examined 1735 traced neurons spanning 1518 unique cell types (therefore examining  
713 43k upstream connections and 43k downstream connections). The histogram of synapse accuracy  
714 (end-to-end precision of predicted synapses) is given in Figure 16. Median precision for upstream  
715 connections, as well as for downstream connections, is 88%. Additionally, 90% of cell types have  
716 an accuracy of at least 70%. For the few worst cases, we manually refined the synapse predictions  
717 afterwards. We note that the worst outlier, having an upstream connection accuracy of 12%, is  
718 both a case involving few total connections (17 T-bars), and some ambiguity in the ground-truth  
719 decisions (whether the annotated location is an actual T-bar).

720 We also evaluated single-connection pathways across each brain region. In the fly, functionally  
721 important connections are thought typically to have many synapses, with the possible exception  
722 of cases where many neurons of the same type synapse onto the same downstream partner..  
723 However, the presence of connections represented by few synapses is also well known, even if  
724 the biological importance of these is less clear. Regardless, we wanted to ensure that even single  
725 connection pathways were mostly correct. We sampled over 5500 single-connection pathways,  
726 distributed across 57 brain regions. Mean synapse precision per brain region was 76.1%, suggesting  
727 that single-connection accuracy is consistent with overall synapse prediction accuracy.



728 We also undertook a preliminary evaluation of two-connection pathways (two synapses be-  
729 tween a single pair of bodies). We sampled 100 such two-connection pathways within the FB.  
730 Overall synapse precision (over the 200 synapses) is 79%, consistent with the single-edge accuracy.  
731 Moreover, the results also suggest that synapse-level accuracy is largely uncorrelated with path-  
732 way/bodies, implying that the probability that both synapses in a two-connection pathway were  
733 incorrect is 4.4% ( $1 - 0.79^2$ ), close to the observed empirical value of 3%. (Applying a  $\chi^2$  goodness of  
734 fit test with a null hypothesis of independence gives a  $p$  value of 0.7.)

### 735 **Assessing connection completeness**

736 A synapse in the fly's brain consists of a presynaptic density (with a characteristic T-bar) and typically  
737 several postsynaptic partners (PSDs). The T-bars are contained in larger neurites, and most (>90%)  
738 of the T-bars in our dataset were contained in identified neurons. The postsynaptic densities are  
739 typically in smaller neurites, and it is these that are difficult for both machine and human to connect  
740 with certainty.

741 With current technology, tracing all fine branches in our EM images is impractical, so we sample  
742 among them (at completeness levels typically ranging from 20% to 85%) and trace as many as  
743 practical in the allotted time. The goal is to provide synapse counts that are representative, since  
744 completeness is beyond reach and largely superfluous. Provided the missing PSDs are independent  
745 (which we try to verify), then the overall circuit emerges even if a substantial fraction of the  
746 connections are missing. If a connection has a strength of 10, for example, then it will be found in  
747 the final circuit with more than 99.9% probability, provided at least half the individual synapses are  
748 traced.

749 If unconnected small twigs are the main source of uncertainty in our data (as we believe to be  
750 the case), then as proofreading proceeds existing connections should only get stronger. Of course  
751 corrections resulting in lower connection strength, such as correcting a false connection or removing  
752 an incorrect synapse, are also possible, but are considerably less likely. To see if our proofreading  
753 process worked as expected, we took a region that had been read to a lower percentage completion  
754 and then spent the manual effort to reach a higher percentage, and compared the two circuits. (A  
755 versioned database such as DVID is enormously helpful here.) If our efforts were successful, ideally  
756 what we see is that almost all connections that changed got stronger, very few connections got  
757 weaker, and no new strong connections appeared (since all strong connections should already be  
758 present even in low coverage proofreading). If this is the behavior we find, we could be reasonably  
759 certain that the circuits found are representative for all strong connections.

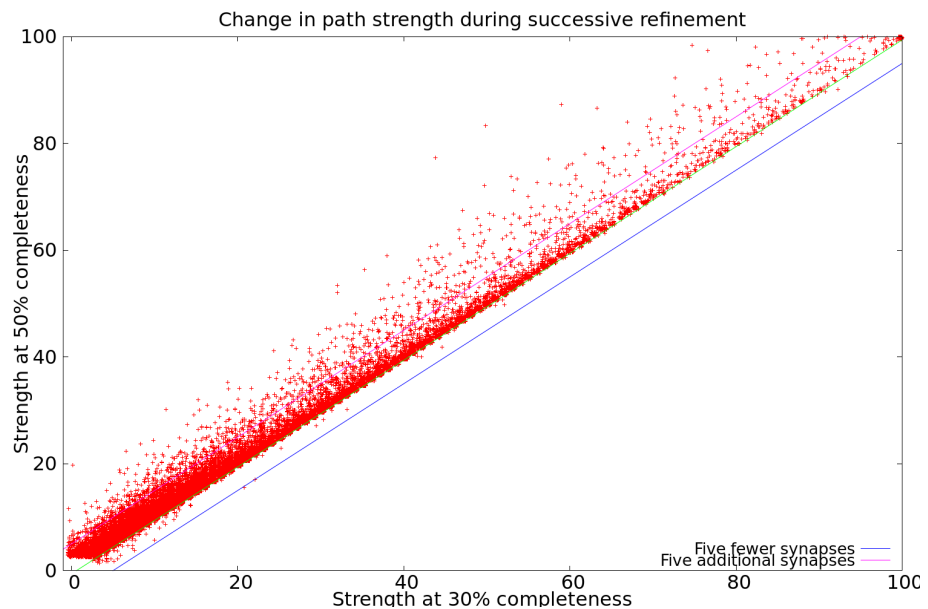
760 Figure 17 below shows such an analysis. The results support our view that the circuits we report  
761 reflect what would be observed if we extrapolated to assign all pre- and postsynaptic elements.

### 762 **Interpreting the connection counts**

763 Given the complexity of the reconstruction process, and the many different errors that could  
764 occur, how confident should the user be that the returned synapse counts are valid? This section  
765 gives a quick guide in the absence of detailed investigation. The number of synapses we return  
766 is the number we found. The true number could range from slightly less, largely due to false  
767 synapse predictions, to considerably more, in the regions with low percentage reconstructed. For  
768 connections known to be in a specific brain region, the reciprocal of the completion percentage (as  
769 shown in Table 1) gives a reasonable estimate of the undercount.

770 If we return a count of 0 (the neurons are not connected), there are two cases. If the neurons do  
771 not share any brain regions, then the lack of connections is real. If they do share a brain region or  
772 regions, then a count of 0 is suspect. It is possible that there might be a weak connection (count 1-2)  
773 and less likely there is a connection of medium strength (3-9 synapses). Strong connections can be  
774 confidently ruled out, minus the small chance of a mis- or un-assigned branch with many synapses.

775 If we report a weak connection (1-2 synapses), then the true strength might range from 0 (the  
776 connection does not exist) through a weak connection (3-9 synapses). If your model or analysis relies



**Figure 17.** Difference between connection strengths in the Ellipsoid Body with increased completeness in proofreading. Roughly 40,000 paths are shown. Almost all points fall above the line  $Y=X$ , showing that almost all paths increased in strength, with very few decreasing. In particular, no path decreased in strength by more than 5 synapses. Only two new strong (strength > 10) paths were found that were not present in the original. This should be rarer at higher levels of proofreading since neuron fragments (orphans) are added in order of decreasing size (see text).

777 on the strength of these weak connections, it is a good idea to manually check our reconstruction.  
778 If your analysis does not depend on knowledge of weak connections, we recommend ignoring  
779 connections based on 3 or fewer synapses.

780 If we report a medium strength connection (3-9 synapses) then the connection is real. The true  
781 strength could range from weak to the lower end of a strong connection.

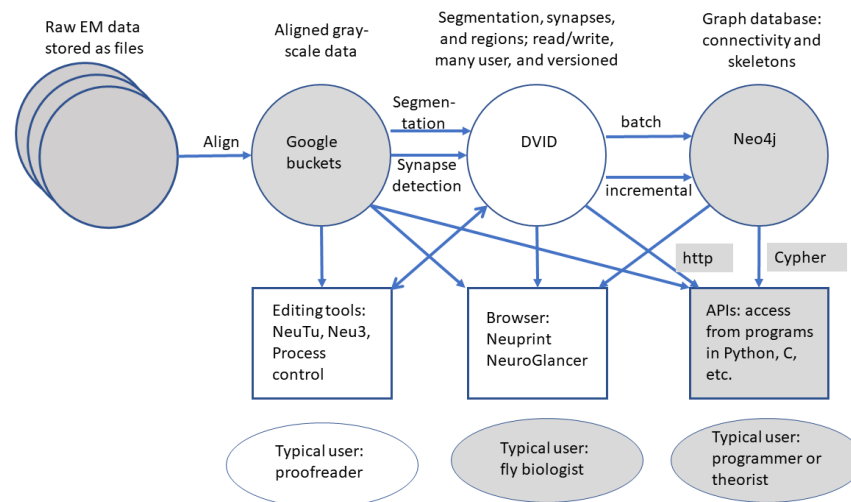
782 If we report a strong connection (10 or more synapses), the connection not only exists, but is  
783 strong. It may well be considerably stronger than we report.

## 784 Data Representation

785 The representation of connectomics data is a significant problem for all connectomics efforts. The  
786 raw image data on which our connectome is based is larger than 20 TB, and takes 2 full days to  
787 download even at a rate of 1 gigabit/second. Looking forward, this problem will only get worse.  
788 Recent similar projects are generating petabytes worth of data (Yin *et al.*, 2019), and a mouse brain  
789 of 500 mm<sup>3</sup>, at a typical FIB-SEM resolution of 8nm isotropic, would require almost 1000 petabytes.

790 In contrast, most users of connectivity information want a far smaller amount of much more  
791 specific information. For example, a common query is 'what neurons are downstream (or upstream)  
792 of a given target neuron?'. This question can be expressed in a few tens of characters, and the  
793 desired answer, the top few partners, fits on a single page of text.

794 Managing this wide range of data, from the raw gray-scale through the connectivity graph,  
795 requires a variety of technologies. An overview of the data representations we used to address  
796 these needs is shown in Figure 18. This organization offers several advantages. In most cases,  
797 instead of transferring files, the user submits queries for the portion of data desired. If the user  
798 needs only a subset of the data (as almost all users do) then they need not cope with the full size of  
799 the data set. Different versions of the data can be managed efficiently behind the scenes with a  
800 versioned database such as DVID (Katz and Plaza, 2019) that keeps track of changes and can deliver  
801 data corresponding to any previous version. The use of existing software infrastructure, such as



**Figure 18.** Overview of data representations of our reconstruction. Circles are stored data representations, rectangles are application programs, ellipses represent users, and arrows indicate the direction of data flow labeled with transformation and/or format. Filled areas represent existing technologies and techniques; open areas were developed for the express purpose of EM reconstruction of large circuits.

802 Google buckets or the graph package neo4j, which are already optimized for large data, helps with  
 803 both performance and ease of development. The advanced user is not limited to these interfaces -  
 804 for those who may wish to validate or extend our results; we have provided procedures whereby  
 805 the user can make personal copies of each representation, including the grayscale, the DVID data  
 806 storage, and our editing and proofreading software. These allow other researchers to establish  
 807 an entirely independent version of all we have done, completely under their control. Contact the  
 808 authors for the details of how to copy all the underlying data and software.

### 809 **What are the data types?**

810 Grayscale data correspond to traditional electron microscope images. This is written only once,  
 811 after alignment, but often read, because it is required for segmentation, synapse finding, and  
 812 proofreading. We store the grayscale data, 8 bits per voxel, in Google buckets, which facilitates  
 813 access from geographically distributed sites.

814 Segmentation, synapses, and identifying regions annotate and give biological meaning to the  
 815 grayscale data. For segmentation, we assign a 64 bit neuron ID to each voxel. Despite the larger  
 816 size per voxel (64 vs 8 bits) compared with the grayscale, the storage required is much smaller (by a  
 817 factor of more than 20) since segmentation compresses well. Although the voxel level segmentation  
 818 is not needed for connectivity queries, it may be useful for tasks such as computing areas and  
 819 cross-sections at the full resolution available, or calculating the distance between a feature and the  
 820 boundary.

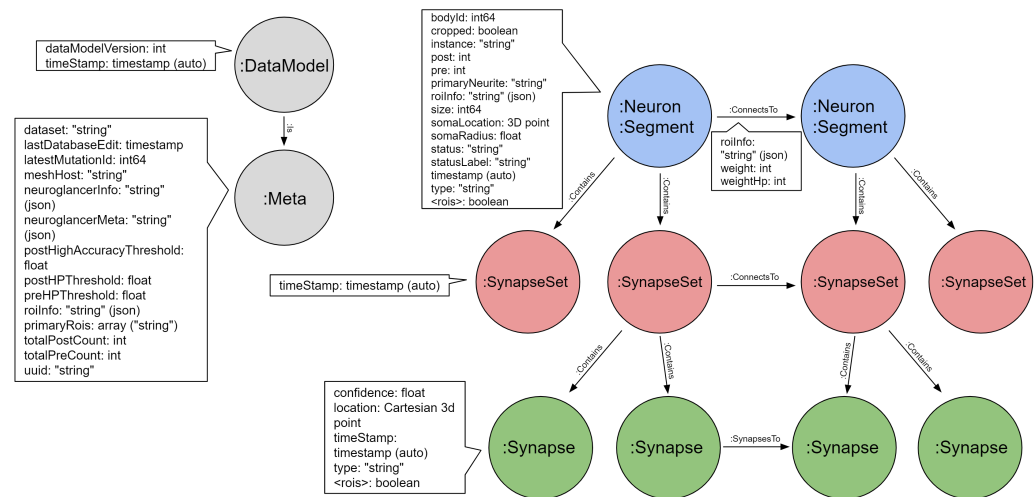
821 Synapses are stored as point annotations - one point for a presynaptic T-bar, and one point for  
 822 each of its postsynaptic densities (or PSDs). The segmentation can then be consulted to find the  
 823 identity of the neurons containing their connecting synapses.

824 The compartment map of the brain is stored as a volume specified at a lower resolution, typically  
 825 a 32x32x32 voxel grid. At 8nm voxels, this gives a 256 nm resolution for brain regions, comparable  
 826 to the resolution of confocal laser scanning microscopy.

827 Unlike the grayscale data, segmentation, synapses, and regions are all modified during proof-  
 828 reading. This requires a representation that must cope with many users modifying the data  
 829 simultaneously, log all changes, and be versioned. We use DVID(Katz and Plaza, 2019), developed  
 830 internally, to meet these requirements.

831 Neuron skeletons are computed from the segmentation (Zhao and Plaza, 2014), and not entered  
 832 or edited directly. A skeleton representation describes each neuron with (branching) centerlines  
 833 and diameters, typically in the SWC format popularized by the simulator *Neuron* (Carnevale and  
 834 Hines, 2006). These are necessarily approximations, since it normally not possible (for example) to  
 835 match both the cross sectional area and the surface area of each point along a neurite with such  
 836 a representation. But SWC skeletons are a good representation for human viewing, adequate for  
 837 automatic morphology classification, and serve as input to neural simulations such as *Neuron*. SWC  
 838 files are also well accepted as an interchange format, used by projects such as *NeuroMorpho* (Ascoli  
 839 et al., 2007) and *FlyBrain* (Shinomiya et al., 2011).

840 The connectivity graph is also derived from the data and is yet more abstract, describing only  
 841 the identity of neurons and a summary of how they connect - for example, Neuron ID1 connects  
 842 to neuron ID2 through a certain number of synapses. In our case it also retains the brain region  
 843 information and the location of each synapse. Such a connectivity graph is both smaller and faster  
 844 than the geometric data, but sufficient for most queries of interest to biologists, such as finding  
 845 the upstream or downstream partners of a neuron. A simple connectivity graph is often desired  
 846 by theorists, particularly within brain regions, or when considering neural circuits in which each  
 847 neuron can be represented as a single node.



**Figure 19.** Schema for the neo4j graph model of the hemibrain. Each neuron contains 0 or more SynapseSets, each of which contains one or more synapses. All the synapses in a SynapseSet connect the same two neurons. If the details of the synapses are not needed, the neuron to neuron weight can be obtained as a property on the “ConnectsTo” relation, as can the distribution of this weight across different brain regions (the roInfo).

848 A final, even more abstract form is the adjacency matrix: This compresses the connectivity  
 849 between each pair of neurons to a single number. Even this most economical form requires careful  
 850 treatment in connectomics. As our brain sample contains more than 25K traced neurons as well as  
 851 many unconnected fragments, the adjacency matrix has more than a billion entries (most of which  
 852 are zero). Sparse matrix techniques, which report only the non-zero coefficients, are necessary for  
 853 practical use of such matrices.

### 854 Accessing the data

855 For the hemibrain project we provide access to the data through a combination of a software  
 856 interface (Clements et al., 2020) and a server (<https://neuprint.janelia.org>). Data are available in the  
 857 form of gray-scale, pixel-level segmentation, skeletons, and a graph representation. Two previous  
 858 connectomics efforts are available as well (a 7-column optic lobe reconstruction (Takemura et al.,

859 **2015)** and the alpha lobe of the mushroom body(*Takemura et al., 2017*). These can be found at  
860 <https://neuprint-examples.janelia.org> .

861 The most straightforward way to access the hemibrain data is through the Neuprint(*Clements*  
862 *et al., 2020*) interactive browser. This is a web-based application that is intended to be usable by  
863 biologists with minimal or no training. It allows the selection of neurons by name, type, or brain  
864 region, displays neurons, their partners, and the synapses between these in a variety of forms, and  
865 provides many of the graphs and summary statistics that users commonly want.

866 Neuprint also supports queries from languages such as Python(*Sanner et al., 1999*) and R, as  
867 used by the neuroanatomy tool NatVerse(*Manton et al., 2019*). Various formats are supported,  
868 including SWC format for the skeletons. In particular, the graph data can be queried through an  
869 existing graph query language, Cypher(*Francis et al., 2018*), as seen in the example below. The  
870 schema for the graph data is shown in Figure 19.

```
871 MATCH (n:Neuron) - [c:ConnectsTo] -> (t:Neuron) WHERE t.type = 'MBON18'  
872 RETURN n.type, n.bodyId, c.weight ORDER BY c.weight DESCENDING
```

873 This query looks for all neurons that are presynaptic to any neuron of type 'MBON18'. For each  
874 such neuron it returns the types and internal identities of the presynaptic neuron, and the count of  
875 synapses between them. The whole list is ordered in order of decreasing synapse count. This is just  
876 an illustration for a particular query that is quite common and supported in Neuprint without the  
877 need for any programming language.

878 Adjacency matrices, if needed, can be derived from the graph representation. We provide a  
879 small demonstration program that queries the API and generates such matrices, either with or  
880 without the brain regions. The two matrices themselves are available in gzipped Python format. For  
881 more information on accessing data and other hemibrain updates, please see <https://www.janelia.org/project-teams/flyem/hemibrain> .  
882

### 883 **Matching EM and light microscopy data**

884 We registered the hemibrain EM data to the JRC2018 *Drosophila* template brain(*Bogovic et al.,*  
885 *2018*) using an automatic registration algorithm followed by manual correction. We began by using  
886 the automated T-bar predictions (described in section 2.3) to generate a T-bar density volume  
887 rendered at a resolution comparable to those from light microscopic images. This hemibrain  
888 synapse density volume was automatically registered to the template brain using ANTs(*Avants*  
889 *et al., 2008*), producing both a forward and inverse transform. The resulting registration was  
890 manually fine-tuned using BigWarp(*Bogovic et al., 2016*). The total transform is the composition of  
891 the ANTs and BigWarp transformations, and can be found at <https://www.janelia.org/open-science/jrc-2018-brain-templates>.  
892

893 Given a particular neuron of interest, researchers can use these resources to identify GAL4  
894 lines labeling that neuron. First the representation of the neuron must be spatially transformed  
895 into the template space that GAL4 driver line to which images have previously been registered. A  
896 mask based approach(*Otsuna et al., 2018*) enables a search for GAL4 driver line image databases  
897 for particular neurons. Skeletonizing hemibrain neurons can enable the enquirer to query GAL4  
898 neuronal skeleton databases using NBLAST(*Costa et al., 2016*).

### 899 **Longer term storage of data, and archival references**

900 Historically, archival data from biology data have been expressed as files that are included with  
901 supplementary data. However, for connectivity data this practice has two main problems. First, the  
902 data are large, and hard to store. Journals, for example, typically limit supplemental data to a few  
903 10s of megabytes. The data here are about 6 orders of magnitude larger. Second, connectome data  
904 are not static, during proofreading and even after initial publication. As proofreading proceeds, the  
905 data improve in their completeness and quality. The question then is how to refer to the data as  
906 they existed at some point in time, required for reproducibility of scientific results. If represented



907 as files, this would require many copies, checkpointed at various times - the 'as submitted' version,  
908 the 'as published' version, the 'current best version', and so on.

909 We resolve this, at least for now, by hosting the data ourselves and making them available  
910 through query mechanisms. Underlying our connectome data is a versioned database (DVID) so it  
911 is technically possible to access every version of the data as it is revised. However, as it requires  
912 effort to host and format this data for the Neuprint browser and API, only selected versions (called  
913 named versions) are available by default from the website, starting with the initial version, which is  
914 'hemibrain:v1.0.' Although this is only version currently, when reproducibility is required, such as  
915 when referencing the data in a paper, it is still best to refer explicitly to the milestone versions by  
916 name (such as 'hemibrain:v1.0.') because we expect a new milestone version every few months, at  
917 least at first. We will supply a DOI for each of these versions, and each is archived, can be viewed  
918 and queried through the web browser and APIs at any time, and will not change.

919 The goal of multiple versions is that later versions should be of higher quality. Towards this end  
920 we have implemented several systems for reporting errors so we can correct them. Users can add  
921 annotations in NeuroGlancer(*Perlman, 2019*), the application used in conjunction with Neuprint  
922 to view image data, where they believe there are such errors. To make this process easier, we  
923 provide a video explaining it. We will review these annotations and amend those that we agree are  
924 problems. Users can also contact us via email about problems they find.

925 Archival storage is an issue since, unlike genetic data, there is not yet an institutional repository  
926 for connectomics data and the data are too large for journals to archive. We pledge to keep our  
927 data available for at least the next 10 years.

## 928 **Analysis**

929 Of necessity, most previous analyses have concentrated on particular circuits, cell types, or brain  
930 regions with relevance to specific functions or behaviors. For example, a classic paper about  
931 motifs(*Song et al., 2005*) sampled the connections between one cell type (layer 5 pyramidal neurons)  
932 in one brain region (rat visual cortex), and found a number of non-random features, such as over-  
933 represented reciprocal connections and a log-normal strength distribution. However, it has never  
934 been clear which of these observations generalize to other cell types, other brain regions, and the  
935 brain as a whole. We are now in a position to make much stronger statements, ranging over all  
936 brain regions and cell types.

937 In addition, many analyses are best performed (or can only be performed) on dense connec-  
938 tomes. Type-wide observations depend on a complete census of that cell type, and depending on  
939 the observation, a complete census of upstream and downstream partners as well. Some analyses,  
940 such as null observations about motifs (where certain motifs do not occur in all or portions of the  
941 fly's brain) can only be undertaken on dense connectomes.

## 942 **Compartment statistics**

943 One analysis enabled by a dense whole-brain reconstruction involves the comparison between the  
944 circuit architectures of different brain areas within a single individual.

945 The compartments vary considerably. Table 3 shows the connectivity statistics of compartments  
946 that are completely contained within the volume, have at least 100 neurons, and have the largest  
947 or smallest value of various statistics. Across regions, the number of neurons varies by a factor of  
948 74, the average number of partners of each neuron by a factor of 36, the network diameter by a  
949 factor of 4, the average strength of connection between partner neurons by a factor of 5, and the  
950 fraction of reciprocal connections by a factor of 5. The average graph distance between neurons is  
951 more conserved, differing by a factor of only 2.

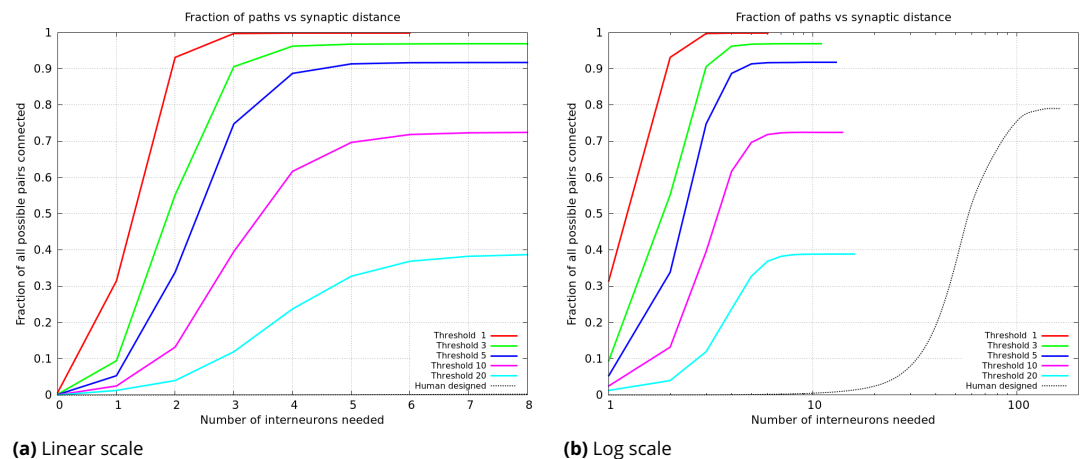
## 952 **Paths in the fly brain are short**

953 Neurons in the fly brain are tightly interconnected, as shown in Figure 20, which plots what fraction  
954 of neuron pairs are connected as a function of the number of interneurons between them. Three

Name	N	L	<k>	D	<str>	<non-r>	<r>	fracR	AvgDist
MB(R)	3430	573664	167.249	8	3.270	3.076	3.383	0.632	2.189
bL(R)	1159	108134	93.299	8	2.019	1.855	2.122	0.613	2.090
EB	518	58793	113.500	4	10.053	4.627	12.172	0.719	1.756
PLP(R)	6689	224985	33.635	16	2.692	2.403	3.685	0.226	3.170
SNP(R)	9121	775474	85.021	13	2.987	2.515	4.492	0.239	2.748
RUB(L)	123	576	4.683	6	7.682	2.852	20.686	0.271	2.743
EPA(R)	1468	18199	12.397	13	2.171	2.098	2.644	0.134	3.496

**Table 3.** Regions with minimum or maximum characteristics, picked from those regions lying wholly within the reconstructed volume and containing at least 100 neurons. Yellow indicates a minimum value; green a maximal value. N is the number of neurons in the region, L the number of connections between those neurons, <k> the average number of partners (in the region), D the network diameter, <str> the average connection strength, broken up into non-reciprocal and reciprocal. fracR is the fraction of connections that are reciprocal, and AvgDist is the average number of hops (one hop corresponding to a direct synaptic connection) between any two neurons in the compartment. Network diameter is computed on the undirected graph; all other metrics use the directed graph.

955 quarters of all possible pairs are connected by a path with fewer than three interneurons, even  
 956 when only connections with  $\geq 5$  synapses are included. If weaker connections are allowed, the  
 957 paths become shorter yet. These short paths and tight coupling are very different from human  
 958 designed systems, which have much longer path lengths connecting node pairs. As an example, a  
 959 standard electrical engineering benchmark (S38584 from (Brglez et al., 1989)) is shown alongside  
 960 the hemibrain data in Figure 20A-B. The connection graph for this example has roughly the same  
 961 number of nodes as the graph of the fly brain, but pair-to-pair connections involve paths more than  
 962 an order of magnitude longer – a typical node pair is separated by 60 intervening nodes. This is  
 963 because a typical computational element in a human designed circuit (a gate) connects only to a  
 964 few other elements, whereas a typical neuron receives input from, and sends outputs to, hundreds  
 965 of other neurons.

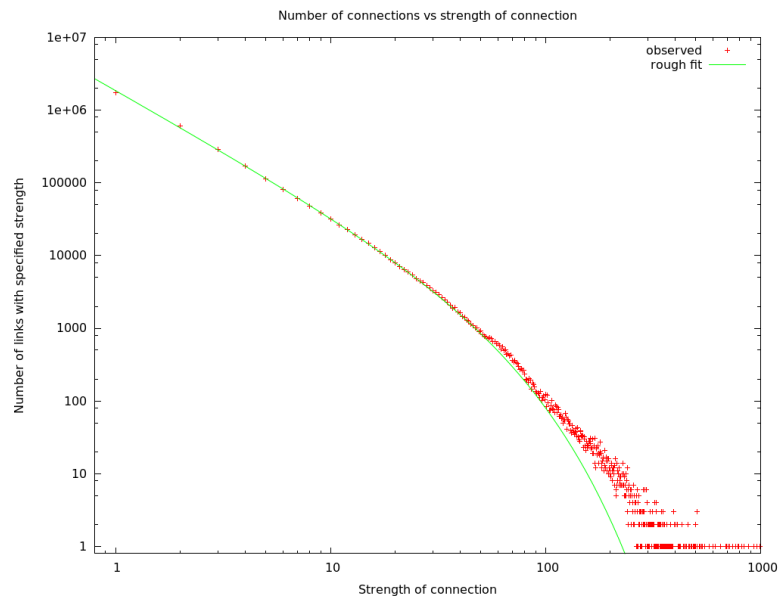


**Figure 20.** Plots of the percentage of pairs connected (of all possible) versus the number of interneurons required. (a) shows the data from the whole hemibrain, for up to 8 interneurons. (b) is a much wider view of the same data, shown on a log scale so the curve from a human designed system is visible.

### 966 Distribution of connection strength

967 The distribution of connection strengths has been studied in mammalian tissue, looking at specific  
 968 cell types in specific brain areas. These findings, such as the log-normal distribution of connection  
 969 strengths in rat cortex, do not appear to generalize to flies. Assuming the strength of a connection  
 970 is proportional to the number of synapses in parallel, we can plot the distribution of connection

971 strengths, summing over the whole central brain, as shown in Figure 21. We find a nearly pure  
972 power law with an exponential cutoff, very different from the log-normal distribution of strengths  
973 found by Song(Song *et al.*, 2005) in pyramidal cells in the rat cortex, or the bimodal distribution  
974 found for pyramidal cells in the mouse by Dorkenwald(Dorkenwald *et al.*, 2019). However, we  
975 caution that these analyses are not strictly comparable. Even aside from the very different species  
976 examined, the three analyses differ. Both Song and Dorkenwald looked at only one cell type, with  
977 excitatory connections only, but one looked at electrical strength while the other looked at synapse  
978 area as a proxy for strength. In our analysis, we use synapse count as a proxy for connection  
979 strength, and look at all cell types, including both excitatory and inhibitory synapses.



**Figure 21.** The number of connections with a given strength. Up to a strength of 100, this is well described by a power law (exponent -1.67) with exponential cutoff (at  $N=42$ ).

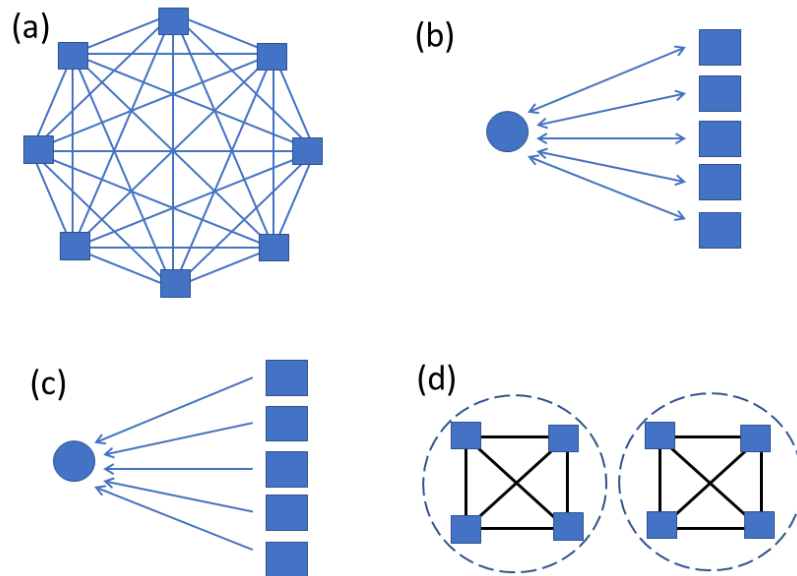
### 980 **Small Motifs**

981 As mentioned earlier, there have been many studies of small motifs, usually involving limited  
982 circuits, cell types, and brain regions. We emphatically confirm some traditional findings, such as  
983 the over-representation of reciprocal connections. We observe this in all brain regions and among  
984 all cell types, confirming similar findings in the antennal lobe(Horne *et al.*, 2018). This can now be  
985 assumed to be a general feature of the fly's brain, and possibly all brains. In the fly, the incidence  
986 varies somewhat by compartment, however, as shown in Table 3.

### 987 **Large motifs**

988 We define a large motif as a graph structure that involves every cell of an abundant type ( $N \geq 20$ ).  
989 The most tightly bound motif is a clique, in which every cell of a given type is connected to every  
990 other cell of that type, with synapses in both directions. Such connections, as illustrated in Figure  
991 22(a), are extremely unlikely in a random wiring model. Consider, for example, the clique of R4d\_b  
992 cells found in the ellipsoid body, as shown in Table 4. In the ellipsoid body, two cells are connected  
993 with an average probability of 0.19. Therefore the odds of finding all 600 possible connections  
994 between R4d\_b cells, assuming a random wiring model, is  $0.19^{600} \approx 10^{-432}$ .

995 In the fly's brain, large cliques occur in only a few cases, as shown in Table 4. All true cliques  
996 are in the central complex, with a near-clique among the KCab-p cells of the mushroom body. The  
997 cells of type PFNa form an interesting case. There are 58 such cells, 29 on each side. They do not



**Figure 22.** Large motifs searched for. Squares represent abundant types with at least 20 instances. Circles represent sparse types with at most two instances. Panel (a) shows a clique, where all possible connections are present. (b) shows bidirectional connections between a sparse type and all instances of an abundant type. (c) show unidirectional connections from all of an abundant type to a sparse type. (d) illustrates a cell type that does not form a clique overall, but does within each of two compartments.

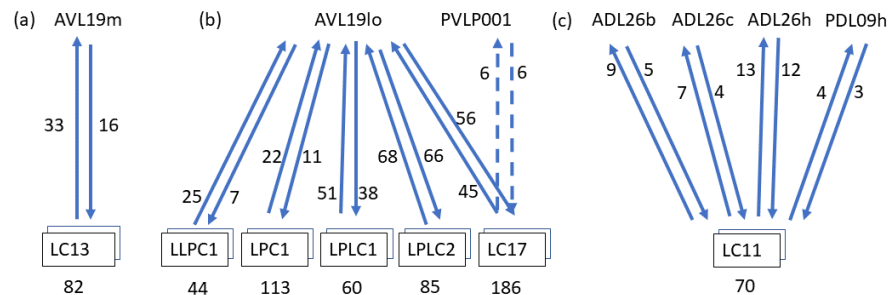
998 form a clique as shown in Figure 22(a), as there are few connections between the opposite sides.  
 999 But within each side, the 29 cells on that side form a clique, as shown in Figure 22(d). The cliques  
 1000 within the central complex, and their potential operation, are discussed in detail in a companion  
 1001 paper.

Type	Region	Cells	Coverage	Avg. Strength	Synapses
KCab-p	MB	55/62	3552/3782	5.04	17899
Delta7_a	PB, CX	32/32	992/992	13.79	13683
R4d_b	EB, CX	25/25	600/600	54.94	32961
R5	EB, CX	20/20	380/380	26.62	10114
R3m	EB, CX	22/22	462/462	24.32	11238
R3d_a	EB, CX	20/20	377/380	28.46	10729
PFNa	NO(R)	29/29	811/812	6.73	5459
PFNa	NO(L)	29/29	811/812	7.22	5858
PFNd	NO(R)	20/20	377/380	7.67	2891
PFNd	NO(L)	20/20	378/380	7.59	2869

**Table 4.** Cliques and near-cliques in the hemibrain data. To be included, a cell type must have at least 20 cell instances, 90% or more of which connect both to and from at least 90% of all cells of the same type. Coverage is the fraction of all possible edges in the clique that are present. Average strength is the average number of synapses in each connection. Synapses is the total number of synapses in the clique.

1002 The next most tightly bound motifs are individual cells that connect both to and from all cells of  
 1003 a given type, but are themselves of a different type. This is illustrated in Figure 22(b). Such a motif  
 1004 is often speculated to be a gain or sparseness controlling circuit, where the single neuron reads  
 1005 the collective activation of a population and then controls their collective behavior. A well known  
 1006 example is the APL neuron in the mushroom body, which connects both to and from all the Kenyon

1007 cells, and is thought to regulate the sparseness of the Kenyon cell activation(Lin et al., 2014).  
 1008 We search for this motif by looking at cells with few instances (one or two) connecting bidirectional-  
 1009 tionally to almost all cells (at least 90%) of an abundant type (N >= 20). We find this motif in three  
 1010 regions of the brain – it is common in the CX (73 different cells overseeing 22 cell types), the optic  
 1011 lobe circuits (19 cells overseeing 14 types), and somewhat in the MB (12 types overseeing 9 types).  
 1012 Spreadsheets containing these cell types, who they connect to, and the numbers and strengths  
 1013 of their connections are found in the supplementary data. We only analyze the optical circuits  
 1014 here, since the mushroom body and central complex are the subjects of companion papers. We  
 1015 observe three variations on this motif - a single cell connected to all of a type (Figure 23(a), found 5  
 1016 times), a single cell with bi-directional connections to many types (Figure 23(b), found once), and  
 1017 multiple cells all connected bidirectionally to a single type (Figure 23(c), found 3 times). We find  
 1018 one circuit that is a combination: There is one cell that connects bidirectionally to all the LC17  
 1019 neurons, and then a higher order cell that connects bidirectionally to a larger set (LPLC1, LPLC2,  
 1020 LLP1, LPC1, and LC17). In this case these are all looming-sensitive cells and hence these circuits  
 1021 may regulate the features of the overall looming responses. It is tempting to speculate that the  
 1022 more complex structures of Figure 23 (b) and (c) arose from the simpler structures of (a) through  
 1023 cell type duplication followed by divergence, but the connectomes of many more related species  
 1024 will be needed before this argument could be made quantitative.



**Figure 23.** One to many motifs found in the optic circuits. Individual neurons are named by unboxed text. Cell type names, in boxes, represent cells with many instances, with the number of instances shown beneath. The arrows show the average synapse count of each connection type. (a) shows an example of the most common case. Here one cell, AVL19m, has bidirectional connections to all cells of type LC13. (b) shows a single cell with exhaustive connections to several types. (c) shows an alternative motif where several cells form these one-to-many connections. For clarity the cell names have been truncated, with the suffix \_pct (for putative cell type) removed.

1025 The least tightly bound large motif is a cell that connects either to or from (but not both) all  
 1026 cells of a given type, as shown in Figure 22(c). Examples include the mushroom body output  
 1027 neurons(Takemura et al., 2017). This is a very common motif, found in many regions. We find more  
 1028 than 500 examples of this in the fly's brain.

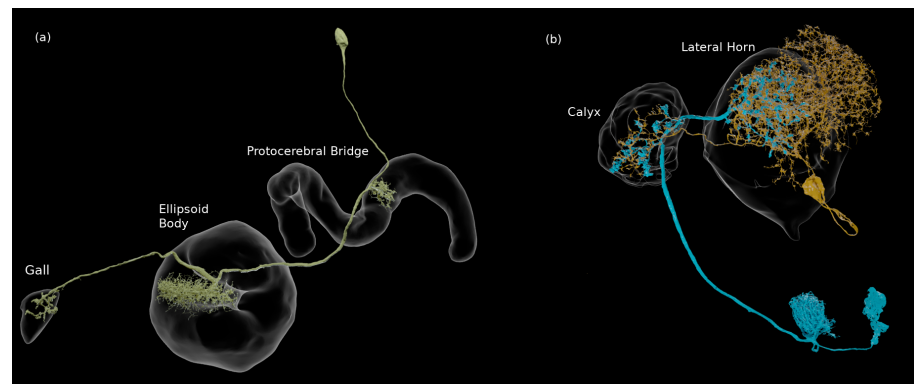
### 1029 Brain regions and electrical response

1030 How does the compartmentalization of the fly brain affect neural computation? In a few cases this  
 1031 has been established. For example, the CT1 neuron performs largely independent computations in  
 1032 each branch(Meier and Borst, 2019), whereas estimates show that within the medulla, the delays  
 1033 within each neuron are likely not significant for single column optic lobe neurons, and hence the  
 1034 neurons likely perform only a single computation(Takemura et al., 2013). Similarly, compartments  
 1035 of PEN2 neurons in the protocerebral bridge have been shown to respond entirely differently from  
 1036 their compartments in the ellipsoid body(Green et al., 2017)(Turner-Evans et al., 2019).

1037 Our detailed skeleton models allow us to construct electrical models of neurons. In particular,  
 1038 to look more generally at the issues of intra- vs inter-compartment delays and amplitudes, we can



1039 construct a linear passive model for each neuron. Our method is similar to that elsewhere (Segev  
1040 *et al.*, 1985), except that instead of using right cylinders, we represent each segment of the skeleton  
1041 as a truncated cone. This is then used to derive the axonic resistance, the membrane resistance,  
1042 and membrane capacitance for each segment. To analyze the effect of compartment structure on  
1043 neuron operation, we inject the neuron at a postsynaptic density (input) with a signal corresponding  
1044 to a typical synaptic input (1 nS conductance, 1 ms width, 0.1 ms rise time constant, 1 ms fall  
1045 time constant, 60 mV reversal potential). We then compute the response at each of the T-bar  
1046 sites (outputs). Since the synapses, both input and output, are annotated by the brain region that  
1047 contains them, this allows us to calculate the amplitudes and delays from each synapse (or a sample  
1048 of synapses) in each compartment to each output synapse in all other compartments.

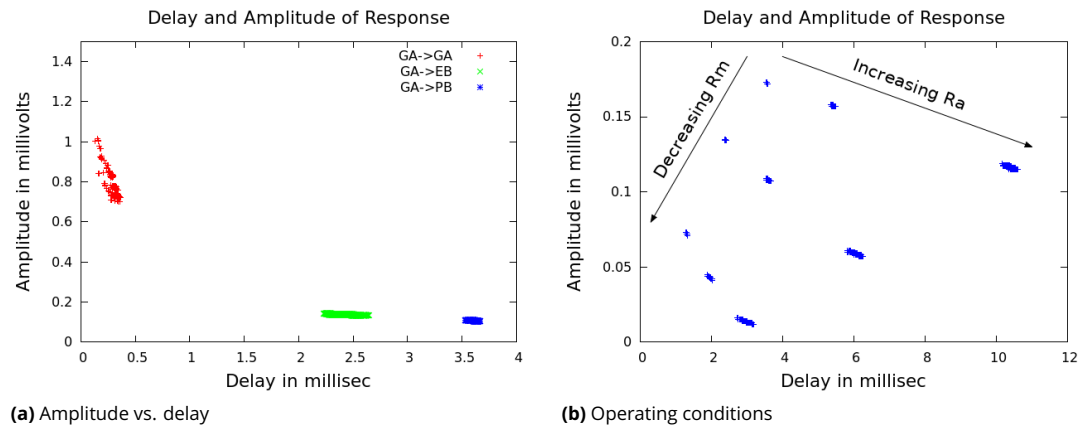


**Figure 24.** (a) An EPG neuron, with arbors in three compartments. (b) Two neurons that connect in more than one ROI, in this case the calyx and the lateral horn. They are each pre- and postsynaptic to each other in both compartments.

1049 In general, we find the ROI structure of the neuron is clearly reflected in the electrical response.  
1050 Consider, for example, the EPG neuron (Figure 24(a)) with arbors in the ellipsoid body, the protocerebral  
1051 bridge, and the gall. Figure 25(a) shows the responses to synaptic input in the gall. Within the  
1052 gall, the delays are very short, and the amplitude relatively high and variable, depending somewhat  
1053 on the input and output synapse within the gall. From the gall to other regions the delays are  
1054 longer (typically a few milliseconds) and the amplitudes much smaller and nearly constant, largely  
1055 independent of the exact transmitting and receiving synapse. There is a very clean separation  
1056 between the within-ROI and across-ROI delays and amplitudes, as shown in Figure 25(a). The same  
1057 overall behavior is true for inputs into the other regions - short delays and strong responses within  
1058 the ROI, with longer delays and smaller amplitudes to other compartments.

1059 This simple pattern motivates a model that describes delays and amplitudes not as a single  
1060 number, but as  $N \times N$  matrix, where  $N$  is the number of ROIs. Each row contains the estimated  
1061 amplitude and delay, measured in each compartment, for a synaptic input in the given compartment.  
1062 This gives a much improved estimate of the linear response. For the example EPG neuron above,  
1063 with nominal values for  $R_a$ ,  $R_m$ , and  $C_m$ , if we represent all delays by a single number then the  
1064 standard deviation of the error is 0.446 ms. If instead we represent the delays as a  $3 \times 3$  matrix  
1065 indexed by the compartment, the average error is 0.045 ms, for 10x greater accuracy. Similarly, the  
1066 average error in amplitude drops from 0.168 mv to 0.021 mv, an eightfold improvement. While the  
1067 improvement in error will depend on the neuron topology, in all cases it will be more accurate than  
1068 a point model, for relatively little increase in complexity.

1069 The absolute values of delay and amplitude are strongly dependent on the electrical parameters  
1070 of the cell, however. A wide range of electrical properties have been reported in the fly literature  
1071 (see Table 5) and it is plausible that these vary on a cell-to-cell basis. We therefore simulate with  
1072 minimum, medium, and maximal values of  $R_a$  and  $R_m$ , for a total of 9 cases, as shown in Figure



(a) Amplitude vs. delay

(b) Operating conditions

**Figure 25.** (a) The linear response to inputs in the gall(GA) for an EPG neuron, which also has arbors in the ellipsoid body(EB) and the protocerebral bridge (PB). Each point in the modeled plot shows the time each response reached its peak amplitude (the delay), and the amplitude at that time, for an input injected at one of the PSDs in the Gall. (b) Delays and amplitudes for gall to PB response, for all combinations of three values of cytoplasmic resistance  $R_A$  and three values of membrane resistance  $R_M$ .

1073 25(b). All are needed since the resistance parameters interact non-linearly. We fix the value of  $C_m$  at  
 1074 0.01 F/m<sup>2</sup> since this value is determined by the membrane thickness and is not expected to vary  
 1075 from cell to cell(Kandel *et al.*, 2000). The results over the parameter range are shown in Figure  
 1076 25(b) for the case of the EPG neuron above for delay from the gall to the PB. The intra-ROI and  
 1077 between-ROI values are well separated for any value of the parameters (not shown).

Reference	$R_a, \Omega \cdot m$	$R_m, \Omega/m^2$	$C_m, F/m^2$
Borst(Borst and Haag, 1996), CH cells	0.60	0.25	0.015
Borst(Borst and Haag, 1996), HS cells	0.40	0.20	0.009
Borst(Borst and Haag, 1996), VS cells	0.40	0.20	0.008
Gouers(Gouwens and Wilson, 2009), DM1 cell 1	1.62	0.83	0.026
Gouers(Gouwens and Wilson, 2009), DM1 cell 2	1.02	2.04	0.015
Gouers(Gouwens and Wilson, 2009), DM1 cell 3	2.66	2.08	0.008
Gouers(Gouwens and Wilson, 2009), dendrite 1	2.44	1.92	0.008
Gouers(Gouwens and Wilson, 2009), dendrite 2	2.66	2.08	0.008
Gouers(Gouwens and Wilson, 2009), dendrite 3	3.11	2.64	0.006
Cuntz(Cuntz <i>et al.</i> , 2013), HS cells	4.00	0.82	0.006
Meier(Meier and Borst, 2019), CT1 cells	4.00	0.80	0.006

**Table 5.** Values reported in the literature

1078 Programs that deduce synaptic strength and sign by fitting a computed response to a con-  
 1079 nectome and measured electrical or calcium imaging data(Tschopp *et al.*, 2018) may at some  
 1080 point require estimates of the delays within cells. If this is required, the above results suggest this  
 1081 could be accomplished with reasonable accuracy with a ROI-to-ROI delay table and 2 additional  
 1082 parameters per neuron,  $R_A$  and  $R_M$ . This is relatively few new parameters in addition to the many  
 1083 synaptic strengths already fitted.

1084 A number of neurons have parallel connections in separate ROIs (see Figure 24(b)). This motif  
 1085 is common in the fly's brain – about 5% of all connections having a strength  $\geq 6$  are spread  
 1086 across two or more non-adjacent ROIs. Given the increased delays and lower amplitudes of

1087 cross-compartment responses, this type of interaction differs electrically from those in which all  
1088 connections are contained in a single ROI. A point neuron model cannot generate an accurate  
1089 response for such connections – a synapse in region A will result in a fast response in A and a slower,  
1090 smaller response in B, and vice versa, even though both of these events involve communication  
1091 between the same two neurons. It is not known if this configuration has a significant influence on  
1092 the neurons' operation.

1093 From these models we conclude (a) the compartment structure of the fly brain shows up  
1094 directly in the electrical response of the neurons. (b) the compartment structure, though defined  
1095 anatomically, matches that of the electrical response. From the clear separation in Figure 25, it is  
1096 likely that the same compartment definitions could be found starting with the electrical response,  
1097 though we have not tried this. (c) These results suggest a low dimensional model for neural  
1098 operation, at least in the linear region. A small region-to-region matrix can represent the delays  
1099 and amplitudes well. (d) Absolute delays depend strongly (but in a very predictable manner) on the  
1100 values of axial and membrane resistance, which can vary both from animal to animal and from  
1101 cell to cell. (e) Neurons that have parallel connections in separate ROIs have a different electrical  
1102 response than they would have with the same total number of synapses in a single ROI.

### 1103 **Rent's rule analysis**

1104 Rent's rule (*Lanzerotti et al., 2005*) is an empirical observation that in human designed computing  
1105 systems, when the system is packed as tightly as possible, at every level of the hierarchy the required  
1106 communication (the number of pins) scales as a power law of the amount of contained computation,  
1107 measured in gates. Rent's rule is an observed relationship, not derived from underlying theory,  
1108 and the relationship is not exact and still contains scatter. A biological equivalent might be the  
1109 observation that brain size tends to vary as a power law of body size (*Harvey and Krebs, 1990*),  
1110 across a wide range of species occupying very different ecological and behavioral niches. Rent's rule  
1111 is roughly true over many orders of magnitude in scale, and for almost every system in which it has  
1112 been measured. Somewhat surprisingly, Rent's rule applies almost independently of the function  
1113 performed by the computation being performed, and at every level of a hierarchical system. It  
1114 also applies whether the compactness criterion is minimization of communication (partitioning) or  
1115 physical close packing.

Rent's rule is expressed as

$$Pins = a * (computation)^b$$

1116 where  $a$  is a scale factor (typically in the range 1-4), and  $b$  is the 'Rent exponent' describing how the  
1117 number of connections to the compartment varies as a function of the amount of computation  
1118 performed in the compartment. The Rent exponent has a theoretical range of 0.0 to 1.0, where 0  
1119 represents a constant number of connections, with no dependence on the amount of computation  
1120 performed, and 1.0 represents a circuit in which every computation is visible on a connection.  
1121 Human designed computational systems occupy almost the full range, from spreadsheets in which  
1122 every computation is visible, to largely serial systems in which minimizing communication (pins) is  
1123 critical. This relationship is shown in Figure 26. However, when the overriding criterion is that the  
1124 system must be packed as tightly as possible, Rent observed that the exponent of the power law  
1125 falls in a close range of roughly 0.5-0.7.

1126 For electrical circuits, the computation is measured in gates, and the connections are measured  
1127 by pin count. These ranges are shown in Figure 26 for circuits that are roughly the size of the fly's  
1128 brain, packed in either two (*Yang et al., 2001*) or three (*Das et al., 2004*) dimensions.

1129 Also shown in this plot are the values for the fly's brain computational regions. In this case,  
1130 the computation is measured as the number of contained T-bars, and the connection count is the  
1131 number of neurons that have at least one synapse both inside and outside the compartment. (Very  
1132 similar results are obtained if the computation is measured as the number of PSDs, or the number  
1133 of unique connection pairs). Almost all the fly brain compartments fall well within the range of



1134 exponents expected for packing-dominated systems, while the ellipsoid body (EB) falls just outside  
1135 the expected area. This is perhaps due to the large number of clique-containing circuits in the  
1136 ellipsoid body (see Table 4), since such circuits have few connections for the amount of synapses  
1137 they contain.

1138 Both human designed and biological systems have huge incentives to pack their computation as  
1139 tightly as possible. A tighter packing of the same computation yields faster operation, lower energy  
1140 consumption, less material cost, and lower mass. A natural speculation, therefore, is that both the  
1141 human-designed and evolved systems are dominated by packing considerations, and that both  
1142 have found similar solutions.

## 1143 **Conclusions and future work**

1144 In this work we have achieved a dream of anatomists that is more than a century old. For at least  
1145 the central brain of at least one animal with a complex brain and sophisticated behavior, we have a  
1146 complete census of all the neurons and all the cell types that constitute the brain, a definitive atlas  
1147 of the regions in which they reside, and a graph representing how they are connected.

1148 To achieve this, we have made improvements to every stage of the reconstruction process.  
1149 Better means of sample preparation, imaging, alignment, segmentation, synapse finding, and  
1150 proofreading are all summarized in this work and will form the basis of yet larger and faster  
1151 reconstructions in the future.

1152 We have provided the data for all the circuits of the central brain, at least as defined by nerve  
1153 cells and chemical synapses. This includes not only circuits of regions that are already the subject  
1154 of extensive study, but also a trove of circuits whose structure and function are yet unknown.

1155 We have provided a public resource that should be a huge help to all who study fly neural circuits.  
1156 Finding upstream and downstream partners, a task that until now has typically taken months of  
1157 challenging experiments, is now replaced by a lookup on a publicly available web site. Detailed  
1158 circuits, which used to require considerable patience, expertise, and expertise to acquire, are now  
1159 available for the cost of an internet query.

1160 More widely, a dense connectome is a valuable resource for all neuroscientists, enabling novel,  
1161 system-wide analyses, as well as suggesting roles for specific pathways. A surprising revelation is  
1162 the richness of anatomical synaptic engagements, which far exceeds pathways required to support  
1163 identified fly behaviors, and suggests that most behaviors have yet to be identified.

1164 Finally, we have started the process of analyzing the connectome, though much remains to be  
1165 done. We have quantified the difference between computational compartments, determined that  
1166 the distribution of strengths is different from that reported in mammals, discovered cliques and  
1167 other structures and where these occur, examined the effect of compartmentalization on electrical  
1168 properties, and provided evidence that the wiring of the brain is consistent with optimizing packing.

1169 Many of the extensions of this work are obvious and already underway. Not all regions of the  
1170 hemibrain have been read to the highest accuracy possible, insofar as we have concentrated first  
1171 on the regions overlapping with other projects, such as the central complex and the mushroom  
1172 body. We will continue to update other sections of the brain, and distributed circuits such as clock  
1173 neurons that are not confined to one region, but spread throughout the brain.

1174 There is much more to be learned about the graph properties of the brain, and how these relate  
1175 to its function.

1176 The two sexes of the *Drosophila* brain are known to differ (Auer and Benton, 2016). so that  
1177 reconstructing a male fly is critical to compare the circuits of the two sexes. The ventral nerve  
1178 cord (VNC) should be included since the circuits in the VNC are known to be crucial for fly motor  
1179 behavior (Yellman et al., 1997). At least one optic lobe should be included to simplify analysis of  
1180 visual inputs to the central brain. A whole brain connectome is preferable to the hemibrain, since  
1181 then most cell types would have at least two examples, left and right, which would lend increased  
1182 confidence to our reconstructions. It would also provide complete reconstruction to the many



1183 neurons that span the brain, especially the clock neurons, and are incomplete in the hemibrain.  
1184 These three goals are combined in a project that is currently underway, to image and reconstruct  
1185 an entire male central nervous system (CNS) including the VNC and optic lobes.

1186 We continue to improve sample preparation, imaging, and reconstruction both to decrease the  
1187 efforts expended on reconstruction and to speed reconstruction of more specimens. Improvements  
1188 include multi-beam imaging, etching methods(Hayworth *et al.*, 2019) that can handle larger areas,  
1189 and yet better reconstruction techniques.

## 1190 Author Contributions

1191 ZL developed sample preparation and fixed and stained the sample; CSX, KJH, HFH developed  
1192 the imaging hardware and imaged the sample; LS, ETT, DK, KAK, DA and SS aligned, flattened and  
1193 integrated the slabs into a coherent volume; WK wrote and managed the versioned data system; SB,  
1194 TZ, PH, LU, TD, DRS, DJO, NN, SP, JC, LS, ET, PS, TK wrote proofreading and analysis software; MJ, JMS,  
1195 PHL, VJain developed and applied segmentation approaches; TB, LL, JMS developed analysis, visual-  
1196 ization and pipeline software; MJ, LL, VJain developed and applied tissue classification approaches;  
1197 GH developed and applied methods to identify synapses; KS, ST, JG, MI, TW, FL, KI, RP, JAH defined  
1198 brain regions and cell types; JB and HO did the EM-optical mapping; CXA, DAB, SB, JAB, BSC NC, MC,  
1199 MD, OD, BE, KF, SF, NF, AF, GPH, EMJ, SK, NAK, JK, SAL, AL, CM, EAM, SM, CM, MN, OO, NO, Co, NP, CP,  
1200 TP, EEP, EMP, NR, CR, MKR, JTR, SMR, MS, AKS, ALS, AS CS, KS, NLS, MAS, AS, JS, ST, IT, DT, ET, TT, JJW  
1201 AND TY performed proofreading; EN and CJK provided proofreading analytics; VJay, IAM, ST, KS, PR,  
1202 RP, FL, GMR undertook biological interpretation and analysis; LS, SP did connectivity analysis and  
1203 wrote the paper; JAH, FL, PR, RP managed proofreading; RG, GMR, SP managed the overall effort.

## 1204 Acknowledgments

1205 We thank our colleagues at Janelia and the broader connectomics field for many helpful discussions  
1206 and suggestions during the course of this work. We thank David Peale, Patrick Lee, and the Janelia  
1207 Experimental Technology group for supporting the modifications of the FIB-SEM systems. Goran  
1208 Ceric and other members of the Scientific Computing Systems and Scientific Computing Software  
1209 Teams at Janelia provided critical support throughout this work. The Janelia Facilities group was  
1210 essential in proving a stable environment for image collection. We thank Julia Buhmann and Jan  
1211 Funke for help in implementing the synapse prediction algorithm described in (Buhmann *et al.*,  
1212 2019). Many colleagues at Janelia as well as Marta Costa, Greg Jefferis and others in Cambridge  
1213 tested the performance of Neuprint performance prior to its release. Major financial support for  
1214 this work was provided by the Howard Hughes Medical Institute and Google Research. Feng Li,  
1215 Philipp Schlegel and approximately 10 percent of the proofreading team were supported by a  
1216 Wellcome Trust Collaborative Award (203261/Z/16/Z).

## 1217 Supplemental Methods

### 1218 Sensory inputs and motor outputs

1219 The dataset covers most of the antennal lobe (AL) glomeruli, which house the presynaptic terminals  
1220 of the olfactory receptor neurons (ORNs) from the antennae. The ORNs are named after their  
1221 innervating glomeruli, e.g., ORN\_DA2, and the olfactory receptors they express, as well as their lig-  
1222 ands, and have been identified through various physiological studies(Couto *et al.*, 2005)(Fishilevich  
1223 and Vosshall, 2005)(Hallem and Carlson, 2006). The olfactory signals are then transmitted by the  
1224 olfactory projections neurons (PNs) to the calyx (CA) of the mushroom body, the lateral horn (LH)  
1225 and beyond.

1226 While a large part of the optic lobe (OL) neuropils are missing, more than half of the lobula  
1227 (LO) and small pieces of the lobula plate (LOP) and medulla (ME) are within the dataset. Many  
1228 neurons connecting the OL and the central brain, or the visual projection neurons (VPNs), are  
1229 identified and annotated, along with their synaptic terminals in the central brain, and in the

1230 optic lobe when possible. Among them, the columnar VPNS, including the lobula columnar (LC),  
1231 lobula plate columnar (LPC), lobula-lobula plate columnar (LLPC), and lobula plate-lobula columnar  
1232 (LPLC) neurons(Ache et al., 2019)(Fischbach and Dittrich, 1989)(Klapoetke et al., 2017)(Otsuna and  
1233 Ito, 2006)(Wu et al., 2016), account for the vast majority of the population and are more or less  
1234 densely identified. Since the distribution of the columnar neurons follows the arrangement of the  
1235 photoreceptor cells in the compound eye, the retinotopy can be traced even in their terminals in  
1236 the central brain. In most cases, these neurons terminate in synapse-rich structures called the  
1237 optic glomeruli in the ventrolateral neuropils, where they relay visual information to higher-order  
1238 neurons(Panser et al., 2016)(Wu et al., 2016).

1239 The antennal mechanosensory and motor center (AMMC) is located lateral and ventral to the  
1240 esophagus foramen. It houses terminals of the Johnston's organ neurons (JONs), the mechanosen-  
1241 sory neurons from the Johnston's organ in the second segment of the antennae, as well as their  
1242 synaptic partners. The AMMC is subdivided into five functionally and anatomically segregated  
1243 zones, A, B, C, D, and E(Kamikouchi et al., 2006). Since the neuropil is partially truncated especially  
1244 in the medial and ventral part corresponding to the zones D and E in the hemibrain dataset, only a  
1245 limited number of the JONs innervating zones A, B, and C have been annotated, as JO-A/B/C.

1246 The gustatory receptor neurons (GRNs) from the labellum and maxillary palp terminate in the  
1247 gustatory sensory centers in the gnathal ganglia (GNG) and the prow (PRW)(Hartenstein et al.,  
1248 2018)(Ito et al., 2014)(Miyazaki and Ito, 2010). Both of them are mostly out of the imaging range of  
1249 the dataset and therefore no GRNs have been identified.

1250 We have identified 60 types of descending neurons (out of a total of 98 types identified by the LM  
1251 study) that play a key role in behavior. These neurons were annotated based on the nomenclature  
1252 described in a previous study(Namiki et al., 2018), namely the classes of DN<sub>a</sub>, DN<sub>b</sub>, DN<sub>g</sub> and DN<sub>p</sub>.  
1253 Due to the lack of ventral region in the current dataset, we are not able to specify other cell types  
1254 that run in the neck connective.

## 1255 Sample Preparation

1256 We employed the Progressive Lowering of Temperature dehydration with Low temperature en bloc  
1257 Staining (PLT-LTS), a modified conventional chemical fixation and en bloc staining method. This  
1258 method, mentioned in our previous papers (Hayworth et al., 2015)(Xu et al., 2017)(Lu et al., 2019),  
1259 is here abbreviated as 'C-PLT'. PLT-LTS is an optimization method to give tissue advanced high  
1260 contrast staining and minimize artifacts: extraction, size and shape variation etc, by treating tissue  
1261 under 0°C to -25 °C in acetone or ethanol based uranyl acetate and OsO<sub>4</sub> after routine fixation.  
1262 PLT-LTS samples show highly visible membranes with fewer deflated and collapsed profiles and  
1263 conspicuous synaptic densities in FIB-SEM images.

1264 Five-day-old adult female *Drosophila*, of the genotype Canton S G1 x w<sup>1118</sup>, were used in this  
1265 experiment. Isolated whole brains were fixed in 2.5% formaldehyde and 2.5% glutaraldehyde in  
1266 0.1 M phosphate buffer at pH 7.4 for 2 hours at 22°C. After washing, the tissues were post-fixed  
1267 in 0.5% osmium tetroxide in double distilled H<sub>2</sub>O for 30 min at 4°C. After washing and en bloc  
1268 staining with 0.5% aqueous uranyl acetate for 30 min and then further washing in water, for 20min  
1269 in 0.8% OsO<sub>4</sub>, a Progressive of Lowering Temperature (PLT) procedure started from 1°C when  
1270 the tissues were transferred into 10% acetone. The temperature was progressively decreased to  
1271 -25°C while the acetone concentration was gradually increased to 97%. The tissue was incubated in  
1272 1% osmium tetroxide and 0.2% uranyl acetate in acetone for 32 hours at -25°C. After PLT and low  
1273 temperature incubation, the temperature was increased to 22°C, and tissues were rinsed in pure  
1274 acetone following by propylene oxide, then infiltrated and embedded in Poly/Bed 812 epoxy (Luft  
1275 formulation).

1276 **Hot knife cutting**

1277 Ultrathick sectioning

1278 The hemibrain is too large to image by FIB-SEM without artifacts so we used our ultrathick sectioning  
1279 'hot knife' procedure(Hayworth *et al.*, 2015) to first slice the brain into 20  $\mu\text{m}$  thick slabs which  
1280 were better suited to FIB-SEM imaging. The Epon-embedded *Drosophila* brain block's face was  
1281 trimmed to present a width of just over 1 mm to the knife during sectioning (with the brain centered  
1282 in this width). The length of the blockface was trimmed to be  $> 3$  mm so that each cut section  
1283 would have a large enough region of blank plastic surrounding the tissue to allow forceps to grasp  
1284 it during later processing steps. All sides of the block were trimmed to be perpendicular to the  
1285 face except the trailing edge which was trimmed to slope away at  $\approx 45^\circ$  (to prevent this trailing  
1286 edge from deforming during hot knife sectioning). Hot knife sectioning was performed on our  
1287 custom ultrathick sectioning testbed(Hayworth *et al.*, 2015). The block was cut at a speed of 0.1  
1288 mm/s into a total of 37 slices, each 20  $\mu\text{m}$  thick, using an oil-lubricated (filtered thread cutting oil,  
1289 Master Plumber) diamond knife (Cryo 25 $^\circ$  from Diatome). The knife temperature was adjusted  
1290 at the beginning of the run to ensure sections flowed smoothly across the knife surface without  
1291 curling (too cold) or buckling (too hot). The knife temperature was measured to be 61 $^\circ\text{C}$  at the  
1292 end of the run. The knife was forced to oscillate via a piezo at 39 kHz during sectioning. A laser  
1293 vibrometer (Polytec CLV-2534) was used to measure the amplitude of vibration at 0.5  $\mu\text{m}$  peak-to-  
1294 peak. Each thick section was collected individually from the knife surface by pressing a vacuum  
1295 aspirator (extended fine tip plastic transfer pipette, Samco Scientific, attached to lab vacuum) onto  
1296 the surface of the section. Each section was transferred to an individual well in the top of a 96-well  
1297 microplate (Costar) into an awaiting oil drop. Once all sections were collected, they were transferred  
1298 via forceps under a dissection microscope to a glass slide. The slide was placed on a hot plate  
1299 (200 $^\circ\text{C}$ ) long enough ( $\approx 10$  s) to flatten any residual curl in the sections. Each section was then  
1300 imaged in a 20x light microscope to evaluate its quality.

1301 Flat embedding

1302 Each of the 20  $\mu\text{m}$  thick Epon-embedded fly brain sections was re-embedded in Durcupan resin to  
1303 allow high quality FIB-SEM imaging. Durcupan re-embedding was required because FIB milling of  
1304 Epon-embedded tissue without a Durcupan front covering resulted in milling streaks which mar the  
1305 SEM images(Xu *et al.*, 2017). Residual oil left over from the cutting process was first removed from  
1306 each thick sections by dipping the section in Durcupan resin. Four drops of Durcupan resin were  
1307 spaced out in sequence on a fresh glass slide. Each section was manually grasped with forceps  
1308 (under a dissecting microscope) and dipped and lightly agitated sequentially in each Durcupan  
1309 drops. Sections were gently wiped against the glass slide between each dipping to remove excess  
1310 Durcupan and oil. After the final dipping, each section was placed (blockface side up) onto the  
1311 heat-sealable side of a strip of 25  $\mu\text{m}$  thick PET film (PP24I, Polymex Clear one side heat sealable/one  
1312 side untreated polyester film, Polyester Converter Ltd.). Flat embedding tissue sections against  
1313 this PET backing provided the strength needed for later mounting and handling. The PET film had  
1314 been previously affixed to a glass slide for support, separated from the slide by a thin Kapton film  
1315 designed to allow easy stripping of the PET. A gasket made from 50  $\mu\text{m}$  thick adhesive-backed  
1316 Kapton was positioned so as to surround all of the sections making a well for Durcupan resin to be  
1317 poured into. This arrangement of sections was placed in a 65 $^\circ\text{C}$  oven for  $\approx 1$  hour to partially cure  
1318 the Durcupan so as to 'tack' the sections into position against the PET film. Then fresh Durcupan  
1319 was poured to fill the well to its brim, and several large area pieces of 20  $\mu\text{m}$  thick Durcupan  
1320 (previously cut from a blank block) were placed above the tissue sections to act as spacers during  
1321 flat embedding to ensure that at least a 20  $\mu\text{m}$  layer of Durcupan would exist in front of each tissue  
1322 section during FIB milling. A piece of 25  $\mu\text{m}$  Kapton film was laid on top of the Durcupan along with  
1323 a glass slide and a weight was placed on top to press excess Durcupan out of the well. This flat  
1324 embedding stack up was cured at 65 $^\circ\text{C}$  for 2 days.

1325 Tab mounting, laser trimming, X-ray imaging  
1326 Each individual brain slab to be FIB-SEM imaged was cut out of this flat embedding using a scalpel,  
1327 and the resulting 'tab' was affixed with cyanoacrylate (Super Glue) to a metal stud. An ultraviolet  
1328 laser (LaserMill, New Wave Research) was used to trim away excess blank resin to minimize the  
1329 FIB-milling time required. An X-ray micro-CT scan (Versa 520, Zeiss) was then performed on each  
1330 tab prior to FIB-SEM imaging.

### 1331 **Imaging**

1332 For the hemibrain, thirteen such slices were imaged using two customized enhanced FIB-SEM  
1333 systems, in which an FEI Magnum FIB column was mounted at 90° onto a Zeiss Merlin SEM. Three  
1334 different imaging conditions were used for different sections with details listed in Table 6. In general,  
1335 SEM images were acquired at 8 nm XY pixel size with a 4-nA beam with 1.2 kV landing energy,  
1336 but other parameters were tuned for best imaging quality. Slices 24 to 27 were imaged with the  
1337 specimen biased at + 600 V to prevent secondary electrons from reaching the detector, so that  
1338 only backscattered electrons were collected. The electron beam energy was lowered to 600 V  
1339 accordingly to maintain the same 1.2 kV landing energy. The remaining slices were imaged with  
1340 specimen grounded at 0 V, and both secondary and backscattered electrons were collected to  
1341 improve signal-to-noise ratio. As a result, SEM scanning rates were set at 2 MHz for slabs with  
1342 specimen bias and 4 MHz for those without specimen bias. FIB milling was carried out by a 7-nA  
1343 30 kV Ga ion beam. Since optic lobes are typically more heavily stained than the central complex,  
1344 the FIB milling step size in sections 22 to 30 was set to 2 nm, while the step size on sections 31 to  
1345 34 was set at 4 nm, to compensate for staining nonuniformity while preserving throughput and  
1346 signal-to-noise ratio. The total FIB-SEM imaging time for the entire hemibrain was roughly four  
1347 FIB-SEM-years: two years of on and off operation with two machines.

Sample ID	Electron beam energy (kV)	Sample bias (kV)	Landing energy (kV)	SEM current (nA)	SEM scan rate (MHz)	x-y pixel (nm)	z-step (nm)
Z0115-22_Sec22	1.2	0	1.2	4	4	8	2
Z0115-22_Sec23	1.2	0	1.2	4	4	8	2
Z0115-22_Sec24	0.6	0.6	1.2	4	2	8	2
Z0115-22_Sec25	0.6	0.6	1.2	4	2	8	2
Z0115-22_Sec26	0.6	0.6	1.2	4	2	8	2
Z0115-22_Sec27	0.6	0.6	1.2	4	2	8	2
Z0115-22_Sec28	1.2	0	1.2	4	4	8	2
Z0115-22_Sec29	1.2	0	1.2	4	4	8	2
Z0115-22_Sec30	1.2	0	1.2	4	4	8	2
Z0115-22_Sec31	1.2	0	1.2	4	4	8	4
Z0115-22_Sec32	1.2	0	1.2	4	4	8	4
Z0115-22_Sec33	1.2	0	1.2	4	4	8	4
Z0115-22_Sec34	1.2	0	1.2	4	4	8	4

**Table 6.** FIB-SEM imaging conditions

### 1348 **Slab Alignment**

1349 From each of the flattened sections, we generated a multi-scale pyramid of the section faces.  
1350 The highest resolution pyramid level sat exactly at the plane, had a thickness of 1px and showed  
1351 a significant amount of cutting artifacts. Lower levels of the pyramid were increasingly thicker,  
1352 projecting deeper into the volume and showed larger structures.

1353 The alignment was initialized with a regularized affine alignment for the complete series of  
1354 face pairs using the feature based method by Saalfeld et al. (Saalfeld et al., 2010) The pyramid of  
1355 section face pairs was then used to robustly calculate pairwise deformations between adjacent  
1356 sections. The faces are of notable size (>30k<sup>2</sup>px) and expose many preparation artifacts such that  
1357 off the shelf registration packages failed to process them reliably. We therefore developed a custom  
1358 pipeline that was able to robustly align the complete series without manual corrections. Using

1359 the same feature based method as above, an increasingly fine grid of local affine transformations  
1360 was calculated and converted into a smooth and increasingly accurate interpolated deformation  
1361 field. The resulting deformation field was further refined using a custom hierarchical optic flow  
1362 method down to a resolution of 2px. Optic flow minimizing the normalized cross correlation (NCC)  
1363 was calculated for a pyramid of square block-sizes. For each pixel, the translation vector with the  
1364 highest number of votes from all block-sizes was selected, and the resulting flow-field was further  
1365 smoothed with an adaptive Gaussian filter that was weighted by the corresponding NCC.

1366 The deformation fields were then applied to each section volume by smoothly interpolating  
1367 between the deformation field at the top face and the affine transformation at the bottom face.

1368 The block-based N5 format (<https://github.com/saalfeldlab/n5>) was used to store volumes,  
1369 multi-scale face pyramids, deformation fields, meta-data, and to generate the final export. Apache  
1370 Spark was used to parallelize on a compute cluster. The pipeline is open source and available on  
1371 GitHub (<https://github.com/saalfeldlab/hot-knife>).

## 1372 Segmentation

### 1373 Image Adjustment with CycleGANs

1374 To reduce photometric variation, we first normalized the contrast of the aligned EM images at full  
1375 resolution ( $[8 \text{ nm}]^3$  / voxel) with CLAHE in planes parallel to the hot-knife cuts. In experiments  
1376 targeted to small subvolumes we observed that segmentation quality decreased in certain areas of  
1377 the hemibrain volume due to variations in the image content arising from, for example, fluctuations  
1378 in staining quality as well as reduced contrast near the boundaries of the physically distinct 13  
1379 hot-knife “tabs” that partitioned the original tissue volume. To compensate for these irregularities,  
1380 we trained and applied CycleGAN(*Zhu et al., 2017*) models. This unsupervised machine learning  
1381 method was originally introduced to adjust the appearance of images from one set A (e.g. photos)  
1382 to be similar to those from another set B (e.g. paintings), without being given any explicit pairings  
1383 between elements of both sets. Here we extended this method to 3d volumes, and used model  
1384 architectures and training hyperparameters as previously described(*Januszewski and Jain, 2019*),  
1385 but without utilizing the flood-filling module.

1386 We trained separate CycleGAN models to make data from every tab visually similar to that of  
1387 a reference area spanning tabs 26 and 27 at  $[32 \text{ nm}]^3$  and  $[16 \text{ nm}]^3$  voxel sizes (i.e. using 4x, and  
1388 2x downsampled images, respectively), yielding a total of 20 CycleGAN models (no model was  
1389 trained for tabs 26 and 27 at 32 nm and for tabs 23, 24, 26, and 27 at 16 nm). The reference area  
1390 was chosen based on similarity to the region in which training data for segmentation models was  
1391 located. The images in tabs 26 and 27 were sufficiently similar that no additional adjustment was  
1392 required. The bounding boxes within the hemibrain volume used for training the CycleGAN models  
1393 are specified in Supplementary Table 7.

1394 During training, a snapshot of network weights (“checkpoint”) was saved every 30 min. CycleGAN  
1395 inference was performed over a tab- and resolution-specific region of interest (ROI; see Table 8)  
1396 with every saved checkpoint from the tab- and resolution-matched model. We then segmented  
1397 the resulting volumetric images with a resolution-matched flood-filling network (FFN) model, and  
1398 screened the segmentations for merge errors. Merge errors were identified by visually inspecting  
1399 the largest objects (by the number of voxels) in the segmentations using a 3d mesh viewer (Neu-  
1400 roglancer). For every CycleGAN model, we selected checkpoints resulting in the minimum number  
1401 of mergers, and then among these, selected the checkpoint corresponding to a segmentation with  
1402 the maximum number of labeled voxels in objects containing at least 10,000 voxels.

1403 We then performed CycleGAN inference with the selected checkpoint for every tab-resolution  
1404 pair over the part of the aligned hemibrain volume corresponding to that tab. The stitched inference  
1405 results were used as input volumes for tissue classification and neuron segmentation. CycleGAN  
1406 normalization was not done at the native  $[8 \text{ nm}]^3$ /voxel resolution because there was insufficient  
1407 evidence that the 8 nm FFN model could generalize well to different tabs.



tab	start			size		
	x	y	z	x	y	z
reference	4633	3792	2000	1374	2000	2000
22	8089	4030	1744	518	2000	2000
23	7435	3925	2101	654	2000	2000
24	6713	2939	4094	722	2000	2000
25	6017	2895	3635	694	2000	2000
28	3980	4944	3495	638	2000	2000
29	3307	2414	4094	666	2000	2000
30	2649	2519	4094	657	2000	2000
31	1979	2750	4094	670	2000	2000
32	1312	3065	4094	667	2000	2000
33	668	3101	3520	663	2000	2000
34	1	3112	3520	660	2000	2000

**Table 7.** Bounding boxes within the hemibrain volume used for training CycleGAN models. Coordinates and sizes are given for  $[32 \text{ nm}]^3$  voxels. The same physical area of the hemibrain volume was used to train both 32 nm and 16 nm CycleGAN models.

#### 1408 Tissue Classification

1409 We manually labeled voxels in 4 tabs of the hemibrain volume as belonging to one of 7 classes:  
1410 'broken white tissue', trachea, cell bodies, glia, large dendrites, neuropil, or 'out of bounds'. We used  
1411 these labels to train a 3d convolutional network that receives as input a field of view of  $65 \times 65 \times 65$   
1412 voxels at  $[16 \text{ nm}]^3/\text{voxel}$  resolution. The network uses 'valid' convolution padding and 'max' pooling  
1413 operations with a kernel and striding shape of  $2 \times 2 \times 2$ , with convolution and pooling operations  
1414 interleaved in the following sequence: convolution with 64 feature maps and a  $3 \times 3 \times 3$  kernel shape,  
1415 max-pooling, convolution with 64 feature maps, max-pooling, convolution with 64 feature maps,  
1416 max-pooling, convolution with  $3 \times 3 \times 3$  kernel size and 16 feature maps, convolution with  $4 \times 4 \times 4$   
1417 kernel shape 512 feature maps (i.e., fully connected layer), and finally a logistic layer output with  
1418 8 units (the first unit was unused in the labeling scheme). The network was trained with data  
1419 augmentation in which the order of the three spatial axes was randomly and uniformly permuted  
1420 for each example during construction of the 16-example minibatch. For each example, the order  
1421 of voxels along each spatial axis was also inverted at random with 50% probability. Examples  
1422 from the 7 classes were sampled randomly with equal probability. The model was implemented in  
1423 TensorFlow and training was performed with asynchronous SGD on eight workers using NVIDIA  
1424 P100 GPUs.

1425 The resulting classifier output was, on certain slices of the hemibrain, manually proofread using  
1426 a custom tool ("Armitage"). The inference and proofreading process was then iterated seven times  
1427 in order to expand and improve the set of ground truth voxels, resulting in a final ground truth  
1428 set with the following number of examples (voxels) in each class: 9.7M broken white tissue, 22.9M  
1429 trachea, 42.1M cell bodies, 5.6M glia, 17.7M large dendrites, 71.4M neuropil, and 208.1M out of  
1430 bounds.

#### 1431 Mitochondria Classification

1432 We detected and classified mitochondria within the hemibrain volume using the same neural  
1433 network architecture and training setup as that used for tissue classification. Ground truth data  
1434 was collected through iterative annotation (2 rounds) in Armitage, in which voxels within hemibrain  
1435 were manually annotated as belonging to one of 4 classes: 'background' (33.7 Mvx), 'regular' (0.7  
1436 Mvx), 'special' (0.5 Mvx), and 'intermediate' (0.5 Mvx).

tab	voxel res. [nm]	start			size		
		x	y	z	x	y	z
22	32	8092	4392	5447	500	936	936
23	32	7435	2479	4979	500	936	936
24	32	6717	5414	4873	500	936	936
25	32	6010	3960	6235	500	936	936
28	32	3971	2591	2954	500	936	936
29	32	3471	4252	2224	500	936	936
30	32	2650	2995	4875	500	936	936
31	32	1982	3196	4875	500	936	936
32	32	1311	3141	4873	500	936	936
33	32	664	2850	4875	500	936	936
34	32	0	1900	4500	500	5000	2500
22	16	16080	8353	9871	1034	936	936
25	16	11900	12657	12636	1406	936	936
25	16	11900	5266	10578	1408	936	936
28	16	7900	9279	4613	1297	936	936
29	16	6550	8520	4613	1333	936	936
30	16	5250	7997	7510	1315	936	936
31	16	3860	7749	7510	1340	936	936
32	16	2550	9482	4225	1334	936	936
33	16	1280	7176	12265	1298	936	936
34	16	0	7587	12265	1328	936	936

**Table 8.** ROIs within the hemibrain volume used for CycleGAN checkpoint selection.

#### 1437 Automated Neuron Segmentation with FFNs

1438 We trained three FFN models composed of the same architecture as detailed in previous work  
 1439 (*Januszewski et al., 2018*) for FIB-SEM volumes, targeted specifically for 8 nm, 16 nm, and 32 nm  
 1440 voxel resolution data. For the 8 nm model we used manually generated ground truth spread over 6  
 1441 subvolumes ( $520^3$  voxels each) located within the ellipsoid body, fan-shaped body and protocerebral  
 1442 bridge. The 16 nm and 32 nm models were trained with a proofread segmentation contained within  
 1443 a  $8600 \times 3020 \times 9500$ -voxel region spanning tabs 26 and 27. For the 32 nm model, training examples  
 1444 were sampled from objects comprising 5,000 or more labeled voxels at 32 nm/voxel resolution. In  
 1445 total, 4.2 Gvx of labeled data were used for the 16 nm model and 423 Mvx for the 32 nm model.

1446 We split the training examples into ‘probability classes’ similarly to (*Januszewski et al., 2018*).  
 1447 Classes 13-17 were not sampled when training the 8 nm model in order to bias it towards small-  
 1448 diameter neurites. For 16 nm and 32 nm models fewer classes were used and the first class  
 1449 comprising all initial training examples with the fraction of voxels set to  $0.95 f_a < 0.05$ . Other than the  
 1450 changes regarding the probability classes, we followed the same procedures for training example  
 1451 sampling, seed list generation, field-of-view movement, and distributed inference as detailed  
 1452 previously (*Januszewski et al., 2018*).

1453 FFN checkpoints were selected in a screening process. We generated tab 24 segmentations at 16  
 1454 and 32 nm voxel resolution for every available checkpoint. We then screened these segmentations  
 1455 for merge errors, annotating every such error with two points, one in each distinct neurite. The  
 1456 segmentation generated with an FFN checkpoint that avoided the most errors was selected. For  
 1457 the 8 nm segmentation, we followed the same procedure but restricted to a  $500^3$  subvolume within  
 1458 tab 24, located at 23284, 1540, 12080.

## 1459 Pipeline for Segmentation of Hemibrain with Flood-filling Networks

### 1460 Multi-resolution and oversegmentation consensus

1461 We built the hemibrain segmentation with a coarse-to-fine variant of the FFN pipeline (*Januszewski*  
1462 *et al., 2018*) combining partial segmentations generated at different resolutions. First, we used the  
1463 16 nm and 32 nm FFN models to segment the dataset at the corresponding resolution, with voxels  
1464 identified by the tissue classifier as glia and out-of-bounds excluded from FFN FOV movement  
1465 ('tissue masking'), and voxels classified as 'broken white tissue' excluded from seed generation.  
1466 Voxels located within 128 nm from every hot knife plane were removed from the image data, and  
1467 segmentation proceeded as if these regions did not exist. The resulting segmentation was extended  
1468 back to the original coordinate system by nearest neighbor interpolation to fill the unsegmented  
1469 spaces.

1470 We then removed objects smaller than 10,000 voxels from the 32 nm segmentation (we will  
1471 refer to the resulting segmentation as S32), isotropically upsampled it 2x, and combined it with the  
1472 16 nm segmentation using oversegmentation consensus (*Januszewski et al., 2018*). The resulting  
1473 segmentation (S16) was used as the initial state for 8 nm FFN inference. In addition to tissue  
1474 masking which was applied in the same way as in the case of lower resolution segmentations, we  
1475 also masked areas within 32 voxels (at 8 nm/voxel resolution) from each hot-knife plane.

1476 FlyEM proofreaders analyzed the roughly 200,000 largest objects in the segmentation, and  
1477 manually split supervoxels identified as causing merge errors. This was done in 3 iterations – two  
1478 targeting neuropil supervoxels, and one targeting cell bodies. The resulting corrected segmentation  
1479 (S8) was used as the base segmentation for further work.

### 1480 Agglomeration

1481 For agglomeration, we modified the scheme described in (*Januszewski et al., 2018*) for use with  
1482 resolution-specific FFN models. First, we established a class for every segment by performing a  
1483 majority vote of the tissue classification model predictions over the voxels covered by the segment.  
1484 For every S16 segment (A, B), we also identified the maximally overlapping segment in S32 (denoted  
1485 respectively  $A_{max}$ ,  $B_{max}$  below). For each of the S32, S16, and S8 segmentations we then computed  
1486 candidate object pairs and agglomeration scores, restricting object pairs to ones involving both  
1487 segments classified as either neuropil or 'large dendrite'. For S8 the object pairs were additionally  
1488 restricted to those that included at least one object not present in S16.

1489 For every evaluated segment pair (A, B) and the corresponding segments (A\*, B\*) generated  
1490 during agglomeration, we computed the scores originally defined in (*Januszewski et al., 2018*) that  
1491 is the recovered voxel fractions ( $f_{AA}$ ,  $f_{AB}$ ,  $f_{BA}$ , and  $f_{BB}$ , where  $f_{AB}$  is the fraction of B found in A\*,  
1492 and so on), the Jaccard index JAB between A\* and B\*, and the number of voxels contained in A\* or  
1493 B\* that had been 'deleted' (i.e., during inference their value in the predicted object mask fell from  
1494  $> 0.8$  to  $< 0.5$ ) during one of the runs (dA, dB)).

1495 We then used the following criteria to connect segments A and B. In S32, we connected segments  
1496 that were scored as  $(f_{**} \geq 0.6 \wedge J_{AB} \geq 0.4) \vee (f_{A*} \geq 0.8) \vee (f_{B*} \geq 0.8)$ . In S16, we connected segments  
1497 that either a) were scored as  $d_* \leq 0.02$  or were both classified as neuropil, and  $f_{**} \geq 0.6 \wedge J_{AB} \geq 0.4$ ,  
1498 or b) were both classified as neuropil,  $A_{max} = 0$  or  $B_{max} = 0$  and  $(f_{A*} \geq 0.9) \vee (f_{B*} \geq 0.9)$ . In S8, we  
1499 connected segments that were scored as  $(d_A \leq 0.02 \vee d_B \leq 0.02) \wedge f_{**} \geq 0.6 \wedge J_{AB} \geq 0.8$ .

1500 Given the application of oversegmentation consensus in the process of building S16, objects  
1501 created in S32 could have a different shape in S8. To compensate for this possibility, when ag-  
1502 glomeration scores were being computed for S32 segments A and B, for each we computed up  
1503 to 8 maximally overlapping objects (A', B') in a downsampled version of S8 with matching voxel  
1504 resolution, subject to a minimum overlap size of 1,000 voxels and considered the agglomeration  
1505 decision to apply to all combinations of A' and B'.

1506 Agglomeration constraints

1507 From the procedure above, we used the agglomeration scores to organize segment connection  
 1508 decisions into priority groups and assign them a single numerical priority score (see Table 9). The  
 1509 decisions were then sorted in ascending order of the priority score, and sequentially processed,  
 1510 removing any decisions that would cause two cell bodies (as defined by manual annotations), or  
 1511 two segments previously separated manually in S8 proofreading to be connected was removed.  
 1512 Additionally, once all decisions with score < 10 were processed, we also disallowed any remaining  
 decisions that connected together any objects larger than 100M vx.

Group	Segmen- tation	Criterion	Score
1	S32	$(d_A \leq 0.02 \vee d_B \leq 0.02) \wedge (f_{**} \geq 0.6 \wedge J_{AB} \geq 0.8)$	$1 - J_{AB}$
2	S16	$(d_A \leq 0.02 \vee d_B \leq 0.02) \wedge (f_{**} \geq 0.6 \wedge J_{AB} \geq 0.8) \wedge (A_{max} = 0 \vee B_{max} = 0) \wedge$ A and B are classified as neuropil	$2 - J_{AB}$
3	S16	$(d_A \leq 0.02 \vee d_B \leq 0.02) \wedge (f_{**} \geq 0.6 \wedge J_{AB} \geq 0.8) \wedge (A_{max} = 0 \vee B_{max} = 0)$	$3 - J_{AB}$
4	S16	$(d_A \leq 0.02 \vee d_B \leq 0.02) \wedge (f_{**} \geq 0.6 \wedge J_{AB} \geq 0.8) \wedge$ A and B are classified as neuropil	$4 - J_{AB}$
5	S16	$(d_A \leq 0.02 \vee d_B \leq 0.02) \wedge (f_{**} \geq 0.6 \wedge J_{AB} \geq 0.8)$	$5 - J_{AB}$
6	S32	$(f_{**} \geq 0.6 \wedge J_{AB} \geq 0.4) \wedge (A_{max} = 0 \vee B_{max} = 0) \wedge$ A and B are classified as neuropil	$6 - J_{AB}$
7	S16	$(f_{**} \geq 0.6 \wedge J_{AB} \geq 0.4) \wedge (A_{max} = 0 \vee B_{max} = 0)$	$7 - J_{AB}$
8	S16	$(f_{**} \geq 0.6 \wedge J_{AB} \geq 0.4) \wedge$ A and B are classified as neuropil	$8 - J_{AB}$
9	S16	$(f_{**} \geq 0.6 \wedge J_{AB} \geq 0.4)$	$9 - J_{AB}$
10	S8	none	$11 - J_{AB}$
11	S32	none	$12 - \max(\min(f_{AA}, f_{AB}), \min(f_{BA}, f_{BB}))$
12	S16	none	$13 - \max(\min(f_{AA}, f_{AB}), \min(f_{BA}, f_{BB}))$

**Table 9.** Criteria for agglomerating priority groups. If an agglomeration decision fulfills the criteria for multiple priority groups, it is assigned to the one with the lowest resulting score.

1513

1514 Speculative agglomeration

1515 Any body (or set of segments connected by the agglomeration graph) larger than 10M vx was consid-  
 1516 ered to be an “anchor” body. We connected smaller bodies to these anchor bodies in a greedy pro-  
 1517 cedure to further reduce the total number of bodies in the agglomerated segmentation. We formed  
 1518 body pair scores using segment pair agglomeration scores as  $\max(\min(f_{AA}, f_{AB}), \min(f_{BA}, f_{BB}))$ . We  
 1519 then merged every body with its highest scoring candidate partner, as long as this would not  
 1520 connect two anchor bodies, and the body pair score was > 0.1. This procedure was repeated 7  
 1521 times.

## 1522 **Synapse Prediction**

### 1523 **Ground Truth**

1524 For training and validation, we collected dense synapse annotations within small cubes, spread  
1525 through different brain regions. In total, we collected 122 such cubes, using 25 for classifier training,  
1526 and the remaining 97 for validation. At each cube location, proofreaders manually annotated all  
1527 T-bars within a  $400^3$  window, and further annotated all PSDs attached to T-bars within a smaller  
1528  $256^3$  sub-window. In total, 7.6k T-bars were annotated, split between 1.8k for training and 5.8k for  
1529 validation, and 11.7k PSDs were annotated, split between 3k for training and 8.7k for validation.

### 1530 **Method**

1531 Details of the T-bar and PSD detection algorithms we used can be found in ([Huang et al., 2018](#)).  
1532 For reference, the T-bar classifier is a 3D CNN using a U-Net architecture([Ronneberger et al., 2015](#)),  
1533 with a receptive field size of  $40^3$  voxels and 770k parameters.

1534 At inference, we leverage the tissue classification results mentioned above by discarding any  
1535 predictions that fell outside of tissue categories of large dendrites or neuropil.

1536 As mentioned in the main text, after collecting ground-truth throughout additional brain regions,  
1537 we found that our initial T-bar classifier was giving lower than desired recall in certain areas.  
1538 Therefore, we trained a new classifier, and combined the results in a cascade fashion, which we  
1539 found gave better results than simply replacing the initial predictions. Specifically, we added any  
1540 predictions above a given confidence threshold made by the new classifier for synapses that were  
1541 not near an existing prediction, and removed any existing predictions that were far from predictions  
1542 made by the new classifier at a second lower/conservative threshold.

1543 One difficulty in placing a single T-bar annotation at each presynaptic location is a certain  
1544 ambiguity with respect to ‘multi T-bars’, cases in which two distinct T-bar pedestals lay in close  
1545 proximity, within the same neuron. Such a case can be difficult to distinguish from a single large  
1546 synapse, both for manual annotators as well as the automated prediction algorithm. To make  
1547 such a distinction reliably would require obtaining many training examples for both cases (multi  
1548 T-bar versus single large synapse), and would only have a slight effect on the final weights of the  
1549 connectome (but not the unweighted connectivity). Therefore, we make no attempt to predict  
1550 multi T-bars, and instead as a final post-processing step, collapse to a single annotation any T-bar  
1551 annotations that are in close proximity and in the same segmented body.

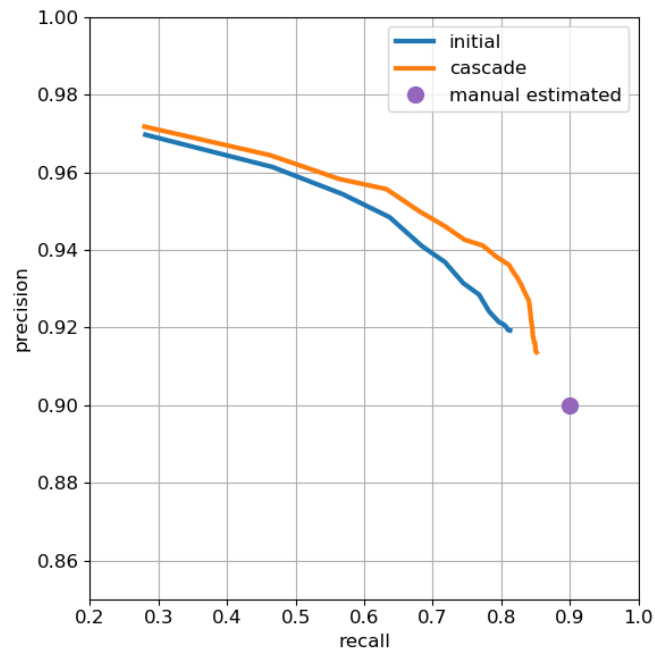
1552 Finally, we observed that in certain brain regions, there are instances of T-bars in separate  
1553 bodies but in close proximity to one another. These often form a ‘convergent T-bars’ motif, in  
1554 which multiple T-bars closely situated in distinct bodies form a synapse onto the same PSD body.  
1555 The proximity of such T-bars is often less than the distance threshold used in the non-maxima  
1556 suppression (NMS) that is applied to generate the T-bar annotations from the pixel-wise U-Net  
1557 predictions. Given the NMS, a number of these types of T-bars would be missed by our predictor.

1558 To address this issue, we modified the post-processing of pixel-wise predictions so as to use  
1559 a ‘segmentation-aware NMS’. Specifically, we constrain the NMS applied to each pixel-wise local  
1560 maxima to largely be limited to the specific segment in which the maxima occurs. Each segment is  
1561 dilated slightly to avoid additional predictions that only fall a very small number of voxels outside the  
1562 segment containing the maxima. (Note that unlike standard NMS, this procedure does require that  
1563 the automated segmentation be available prior to inference.) We apply the segmentation-aware  
1564 NMS only in brain regions where convergent T-bars were observed, as occurs in the mushroom  
1565 body and fan-shaped body.

### 1566 **Evaluation**

1567 Figure 27 gives the precision-recall plot for T-bar prediction, averaged over all the available ground-  
1568 truth validation cubes. As mentioned above, we do not attempt to predict multi T-bars; therefore,  
1569 for the purposes of evaluation, we also collapse any ground-truth T-bars within close proximity in  
1570 the same body to a single annotation. As can be seen from the figure, the cascade predictions are





**Figure 27.** Precision-recall plot of T-bar prediction. The purple intercept indicates estimated manual agreement rate of 0.9.

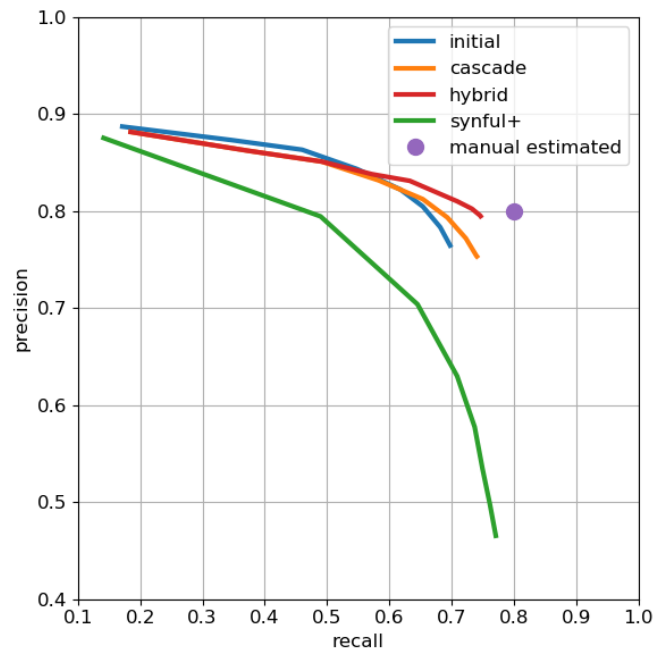
1571 able to increase recall while maintaining precision. One of the primary error modes that leads to  
1572 a difference between automated accuracy and manual agreement rate is the case of convergent  
1573 T-bars, noted above. For instance, in Figure 6 of the main text, the brain region with lowest recall  
1574 is b'L in the mushroom body; closer analysis revealed many convergent T-bars in the annotated  
1575 ground-truth cubes for b'L.

1576 Figure 28 below in the next subsection gives the corresponding precision-recall plot for end-  
1577 to-end synapse prediction, averaged over all the available ground-truth validation cubes. As with  
1578 both (*Huang et al., 2018*) and (*Buhmann et al., 2019*), we do not attempt to predict autapses, and  
1579 remove any predicted connections that lie within the same neuron. For evaluation, any occasional  
1580 ground-truth autapses are filtered out.

#### 1581 Additional Classifier

1582 As an independent check on synapse quality, we also trained a separate classifier proposed by  
1583 Buhmann (*Buhmann et al., 2019*), using the 'synful' software package provided. We additionally  
1584 made several modifications to the code, including: adding an 'ignore' region around synapse blobs  
1585 where predictions were not penalized, using focal loss (*Lin et al., 2017*) to help with class imbalance,  
1586 using batch normalization (*Ioffe and Szegedy, 2015*) and residual layers (*He et al., 2016*), and adding  
1587 explicit T-bar prediction as an additional network output. We found this multi-task learning (adding  
1588 explicit T-bar prediction to PSD prediction and partner direction prediction) to be beneficial, similar  
1589 to the use of cleft prediction in (*Buhmann et al., 2019*), most likely due to the T-bar pedestals being  
1590 a more reliable and prominent signal in our hemibrain preparation/staining than the PSDs. We  
1591 refer to this network and its resulting synapse predictions as 'synful+'.

1592 Figure 28 shows the overall end-to-end precision-recall plots for each of the classifiers. As  
1593 mentioned in the main text, we combined the predictions from the cascade and synful+ classifiers  
1594 to yield a 'hybrid' classifier that achieved both better recall and precision than the two individual  
1595 classifiers. Specifically, we modified the cascade predictions by 1) adding any PSDs that were  
1596 predicted with strong confidence by synful+ and attached to existing T-bars, and 2) removing any



**Figure 28.** Precision-recall plot of end-to-end synapse prediction. The purple intercept indicates estimated manual agreement rate of 0.8.

1597 PSDs that were predicted with weak confidence by the cascade classifier and not predicted by  
1598 synful+ even at a very low confidence threshold.

### 1599 Pathway Analysis

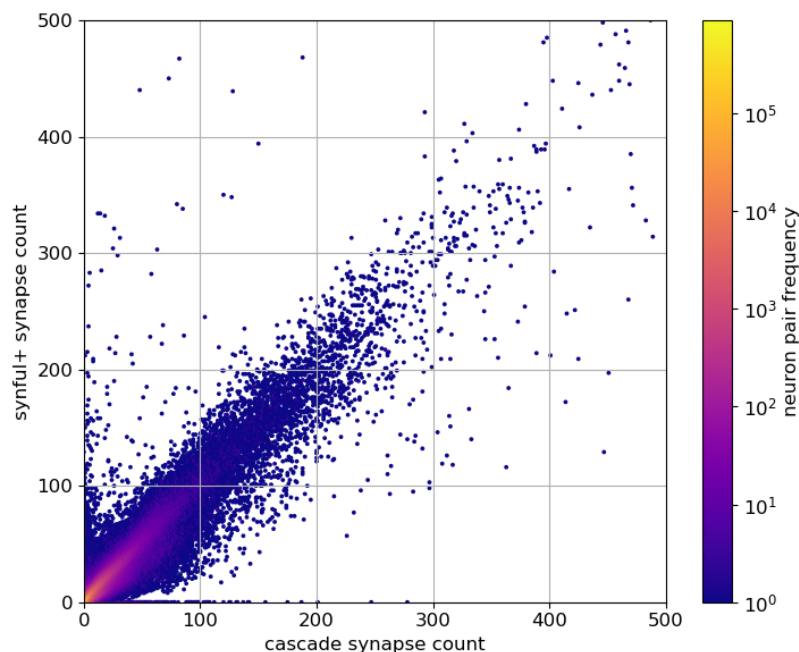
1600 Given two independent sets of synapse predictions (cascade and synful+), we further conduct  
1601 an analysis of their respective connectivity graphs. We construct connectomes from each set of  
1602 synapse predictions, limited to the 21,000+ traced bodies. At the level of individual synapses, the  
1603 two sets of predictions have an agreement rate of about 80%.

1604 However, we can look at connections of a given strength in one set of predictions, and see  
1605 whether the other set of predictions gives a corresponding connection of any strength. For instance,  
1606 among bodies that are connected with at least 5 synapses in the cascade predictions, less than  
1607 1% have no connection in the synful+ predictions, and similarly, among bodies that are connected  
1608 with at least 5 synapses in the synful+ predictions, less than 2% have no connection in the cascade  
1609 predictions. This suggests some level of stability in edges with a stronger connection, so that using  
1610 a different classifier would be still likely to maintain that edge.

1611 We also further manually assessed the small percentage of outlier edges. We sampled 100  
1612 synapses from the strongest of the edges in the cascade predictions that are not present in the  
1613 synful+ predictions, and similarly 100 synapses from the synful+ predictions. For the cascade  
1614 predictions, we find an overall accuracy of 64%, lower than the general accuracy of the cascade  
1615 predictor, but we did not observe a pathway in which all sampled synapses were false positives.  
1616 For the synful+ predictions, we found that all sampled synapses were false positives, resulting from  
1617 improper placement of the T-bar annotation, thereby assigning the T-bar to an incorrect body. This  
1618 suggests another use for such pathway analysis, in potentially discovering particular error modes  
1619 of a classifier and allowing for re-training/refining to address such errors.

1620 As a related measure of connectome stability, we also looked at how often the magnitude of the  
1621 pathway connections were comparable. For instance, we can examine connections consisting of at  
1622 least 10 synapses in one prediction set, and see how often those connections are within a factor of

1623 2 in the other prediction set. We find that this holds for 93% of the connections of strength greater  
1624 than 10. Figure 29 shows a plot comparing pathway connection strength between the two sets of  
predictions.



**Figure 29.** Comparison of synful+ connection strength versus cascade connection strength (truncated at a connection strength of 500 for clarity, omitting 40 edges from each prediction set).

1625

## 1626 **Logistics and Management**

1627 The hemibrain reconstruction required a large-scale effort involving several research labs, Janelia  
1628 shared services, about ten staff scientists, and about 60 proofreaders. The overall initiative planning,  
1629 including the choice of biological regions to image and reconstruct, timeline, and budget, was  
1630 orchestrated by the FlyEM project team at the Janelia Research Campus with a guiding steering  
1631 committee composed of several experts within the institute. The Connectomics Group at Google  
1632 Research collaborated extensively with FlyEM developing key technology to segment the hemibrain  
1633 volume.

1634 Extensive orchestration by project staff and Janelia shared services was required to manage  
1635 the team of proofreaders and the reconstruction effort. Our proofreading team consisted of  
1636 full-time technicians hired specifically for proofreading. To satisfy the ambitious reconstruction  
1637 goals of the hemibrain effort, we hired close to 30 people in a few months to augment the existing  
1638 proofreading resources, requiring a streamlined system of recruitment and training. We found that  
1639 the average proofreader required around 2 months of training to become reasonably proficient in  
1640 EM tracing, which entailed working on carefully designed training modules and iterative feedback  
1641 with more experienced proofreaders or managers. Ongoing training was necessary for both new  
1642 and experienced proofreaders to meet the needs of different reconstruction tasks. The team  
1643 of proofreaders had frequent meetings, and a Slack channel, with the software staff to improve  
1644 proofreading software. We found that for a project of this size, several additional software personnel  
1645 were required for data management, monitoring, orchestrating, and streamlining proofreading  
1646 assignments.

1647 The hemibrain reconstruction involved several different reconstruction steps or workflows,  
1648 many discussed in the paper. The primary workflows were cleaving, false split review, focused

1649 proofreading, and orphan linking. Cleaving is the task of splitting a falsely merged segment. False  
1650 split review entails examining a neuron, using 3D morphology, for potential false splits. Focused  
1651 proofreading is a “merge” or “don’t merge” protocol based on automated suggestions from the  
1652 segmentation algorithm. Orphan linking is fixing small detached segments that should either be  
1653 annotated as exiting the hemibrain dataset, or be merged to a larger, already proofread body.  
1654 Overall we estimate that we undertook  $\approx$  50-100 proofreading years of reconstruction effort.

## 1655 References

- 1656 **Ache JM**, Polsky J, Alghailani S, Parekh R, Breads P, Peek MY, Bock DD, von Reyn CR, Card GM. Neural basis  
1657 for looming size and velocity encoding in the *Drosophila* giant fiber escape pathway. *Current Biology*. 2019;  
1658 29(6):1073–1081.
- 1659 **Altschul SF**, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *Journal of molecular*  
1660 *biology*. 1990; 215(3):403–410.
- 1661 **Ascoli GA**, Donohue DE, Halavi M. NeuroMorpho. Org: a central resource for neuronal morphologies. *Journal*  
1662 *of Neuroscience*. 2007; 27(35):9247–9251.
- 1663 **Aso Y**, Hattori D, Yu Y, Johnston RM, Iyer NA, Ngo TT, Dionne H, Abbott L, Axel R, Tanimoto H, et al. The neuronal  
1664 architecture of the mushroom body provides a logic for associative learning. *Elife*. 2014; 3:e04577.
- 1665 **Auer TO**, Benton R. Sexual circuitry in *Drosophila*. *Current opinion in neurobiology*. 2016; 38:18–26.
- 1666 **Avants BB**, Epstein CL, Grossman M, Gee JC. Symmetric diffeomorphic image registration with cross-correlation:  
1667 evaluating automated labeling of elderly and neurodegenerative brain. *Medical image analysis*. 2008; 12(1):26–  
1668 41.
- 1669 **Bogovic JA**, Hanslovsky P, Wong A, Saalfeld S. Robust registration of calcium images by learned contrast  
1670 synthesis. In: *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)* IEEE; 2016. p. 1123–1126.
- 1671 **Bogovic JA**, Otsuna H, Heinrich L, Ito M, Jeter J, Meissner GW, Nern A, Colonell J, Malkesman O, Ito K, et al. An  
1672 unbiased template of the *Drosophila* brain and ventral nerve cord. *BioRxiv*. 2018; p. 376384.
- 1673 **Borst A**, Haag J. The intrinsic electrophysiological characteristics of fly lobula plate tangential cells: I. Passive  
1674 membrane properties. *Journal of computational neuroscience*. 1996; 3(4):313–336.
- 1675 **Brglez F**, Bryan D, Kozminski K. Combinational profiles of sequential benchmark circuits. In: *IEEE International*  
1676 *Symposium on Circuits and Systems*, IEEE; 1989. p. 1929–1934.
- 1677 **Buhmann J**, Sheridan A, Gerhard S, Krause R, Nguyen T, Heinrich L, Schlegel P, Lee WCA, Wilson R, Saalfeld S,  
1678 et al. Automatic Detection of Synaptic Partners in a Whole-Brain *Drosophila* EM Dataset. *bioRxiv*. 2019; .
- 1679 **Carnevale NT**, Hines ML. *The NEURON book*. Cambridge University Press; 2006.
- 1680 **Clements J**, Dolafi T, Umayam L, Neubarth NL, Berg S, Scheffer LK, Plaza SM. neuPrint: Analysis Tools for EM  
1681 Connectomics. *BioRxiv*. 2020; .
- 1682 **Costa M**, Manton JD, Ostrovsky AD, Prohaska S, Jefferis GS. NBLAST: rapid, sensitive comparison of neuronal  
1683 structure and construction of neuron family databases. *Neuron*. 2016; 91(2):293–311.
- 1684 **Couto A**, Alenius M, Dickson BJ. Molecular, anatomical, and functional organization of the *Drosophila* olfactory  
1685 system. *Current Biology*. 2005; 15(17):1535–1547.
- 1686 **Cuntz H**, Forstner F, Schnell B, Ammer G, Raghu SV, Borst A. Preserving neural function under extreme scaling.  
1687 *PloS one*. 2013; 8(8).
- 1688 **Das S**, Chandrakasan AP, Reif R. Calibration of Rent’s rule models for three-dimensional integrated circuits. *IEEE*  
1689 *Transactions on Very Large Scale Integration (VLSI) Systems*. 2004; 12(4):359–366.
- 1690 **Dorkenwald S**, Turner NL, Macrina T, Lee K, Lu R, Wu J, Bodor AL, Bleckert AA, Brittain D, Kemnitz N, et al. Binary  
1691 and analog variation of synapses between cortical pyramidal neurons. *bioRxiv*. 2019; .
- 1692 **Eichler K**, Li F, Litwin-Kumar A, Park Y, Andrade I, Schneider-Mizell CM, Saumweber T, Huser A, Eschbach C,  
1693 Gerber B, et al. The complete connectome of a learning and memory centre in an insect brain. *Nature*. 2017;  
1694 548(7666):175.

- 1695 **Fischbach KF**, Dittrich A. The optic lobe of *Drosophila melanogaster*. I. A Golgi analysis of wild-type structure.  
1696 Cell and tissue research. 1989; 258(3):441–475.
- 1697 **Fishilevich E**, Vosshall LB. Genetic and functional subdivision of the *Drosophila* antennal lobe. Current Biology.  
1698 2005; 15(17):1548–1553.
- 1699 **Fortunato S**, Hric D. Community detection in networks: A user guide. Physics reports. 2016; 659:1–44.
- 1700 **Francis N**, Green A, Guagliardo P, Libkin L, Lindaaker T, Marsault V, Plantikow S, Rydberg M, Selmer P, Taylor A.  
1701 Cypher: An evolving query language for property graphs. In: *Proceedings of the 2018 International Conference*  
1702 *on Management of Data* ACM; 2018. p. 1433–1445.
- 1703 **Gouwens NW**, Wilson RI. Signal propagation in *Drosophila* central neurons. Journal of Neuroscience. 2009;  
1704 29(19):6239–6249.
- 1705 **Green J**, Adachi A, Shah KK, Hirokawa JD, Magani PS, Maimon G. A neural circuit architecture for angular  
1706 integration in *Drosophila*. Nature. 2017; 546(7656):101.
- 1707 **Hallem EA**, Carlson JR. Coding of odors by a receptor repertoire. Cell. 2006; 125(1):143–160.
- 1708 **Hanslovsky P**, Bogovic JA, Saalfeld S. Image-based correction of continuous and discontinuous non-planar axial  
1709 distortion in serial section microscopy. Bioinformatics. 2017; 33(9):1379–1386.
- 1710 **Hartenstein V**, Omoto JJ, Ngo KT, Wong D, Kuert PA, Reichert H, Lovick JK, Younossi-Hartenstein A. Structure  
1711 and development of the subesophageal zone of the *Drosophila* brain. I. Segmental architecture, compartmen-  
1712 talization, and lineage anatomy. Journal of Comparative Neurology. 2018; 526(1):6–32.
- 1713 **Harvey PH**, Krebs JR. Comparing brains. Science. 1990; 249(4965):140–146.
- 1714 **Hayworth KJ**, Peale D, Januszewski M, Knott GW, Lu Z, Xu CS, Hess HF. Gas cluster ion beam SEM for imaging of  
1715 large tissue samples with 10 nm isotropic resolution. Nature methods. 2019; p. 1–4.
- 1716 **Hayworth KJ**, Xu CS, Lu Z, Knott GW, Fetter RD, Tapia JC, Lichtman JW, Hess HF. Ultrastructurally smooth thick  
1717 partitioning and volume stitching for large-scale connectomics. Nature methods. 2015; 12(4):319.
- 1718 **He K**, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference*  
1719 *on computer vision and pattern recognition*; 2016. p. 770–778.
- 1720 **Helmstaedter M**, Briggman KL, Turaga SC, Jain V, Seung HS, Denk W. Connectomic reconstruction of the inner  
1721 plexiform layer in the mouse retina. Nature. 2013; 500(7461):168.
- 1722 **Horne JA**, Langille C, McLin S, Wiederman M, Lu Z, Xu CS, Plaza SM, Scheffer LK, Hess HF, Meinertzhagen IA.  
1723 A resource for the *Drosophila* antennal lobe provided by the connectome of glomerulus VA1v. Elife. 2018;  
1724 7:e37550.
- 1725 **Huang GB**, Scheffer LK, Plaza SM. Fully-automatic synapse prediction and validation on a large data set.  
1726 Frontiers in neural circuits. 2018; 12.
- 1727 **Hubbard PM**, Berg S, Zhao T, Olbris DJ, Umayam L, Maitin-Shepard J, Januszewski M, Katz WT, Neace ER, Plaza  
1728 SM. Accelerated EM Connectome Reconstruction using 3D Visualization and Segmentation Graphs. BioRxiv.  
1729 2020; .
- 1730 **Ioffe S**, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift.  
1731 In: *Proceedings of the International Conference on Machine Learning*; 2015. p. 448–456.
- 1732 **Ito K**, Shinomiya K, Ito M, Armstrong JD, Boyan G, Hartenstein V, Harzsch S, Heisenberg M, Homberg U, Jenett A,  
1733 et al. A systematic nomenclature for the insect brain. Neuron. 2014; 81(4):755–765.
- 1734 **Ito M**, Masuda N, Shinomiya K, Endo K, Ito K. Systematic analysis of neural projections reveals clonal composition  
1735 of the *Drosophila* brain. Current Biology. 2013; 23(8):644–655.
- 1736 **Januszewski M**, Jain V. Segmentation-Enhanced CycleGAN. bioRxiv. 2019; p. 548081.
- 1737 **Januszewski M**, Kornfeld J, Li PH, Pope A, Blakely T, Lindsey L, Maitin-Shepard J, Tyka M, Denk W, Jain V. High-  
1738 precision automated reconstruction of neurons with flood-filling networks. Nature methods. 2018; 15(8):605.
- 1739 **Jenett A**, Rubin GM, Ngo TT, Shepherd D, Murphy C, Dionne H, Pfeiffer BD, Cavallaro A, Hall D, Jeter J, et al. A  
1740 GAL4-driver line resource for *Drosophila* neurobiology. Cell reports. 2012; 2(4):991–1001.



- 1741 **Kainmueller D**, Lamecker H, Zachow S, Heller M, Hege HC. Multi-object segmentation with coupled deformable  
1742 models. In: *Proc. of Medical Image Understanding and Analysis*; 2008. p. 34–38.
- 1743 **Kamikouchi A**, Shimada T, Ito K. Comprehensive classification of the auditory sensory projections in the brain  
1744 of the fruit fly *Drosophila melanogaster*. *Journal of Comparative Neurology*. 2006; 499(3):317–356.
- 1745 **Kandel ER**, Schwartz JH, Jessell TM, of Biochemistry D, Jessell MBT, Siegelbaum S, Hudspeth A. Principles of  
1746 neural science, vol. 4. McGraw-hill New York; 2000.
- 1747 **Kasthuri N**, Hayworth KJ, Berger DR, Schalek RL, Conchello JA, Knowles-Barley S, Lee D, Vázquez-Reina A, Kaynig  
1748 V, Jones TR, et al. Saturated reconstruction of a volume of neocortex. *Cell*. 2015; 162(3):648–661.
- 1749 **Katz WT**, Plaza SM. DVID: Distributed Versioned Image-Oriented Dataservice. *Frontiers in neural circuits*. 2019;  
1750 13.
- 1751 **Khairy K**, Denisov G, Saalfeld S. Joint Deformable Registration of Large EM Image Volumes: A Matrix Solver  
1752 Approach. arXiv preprint arXiv:180410019. 2018; .
- 1753 **Klapoetke NC**, Nern A, Peek MY, Rogers EM, Breads P, Rubin GM, Reiser MB, Card GM. Ultra-selective looming  
1754 detection from radial motion opponency. *Nature*. 2017; 551(7679):237–241.
- 1755 **Lanzerotti MY**, Fiorenza G, Rand RA. Microminiature packaging and integrated circuitry: The work of EF Rent,  
1756 with an application to on-chip interconnection requirements. *IBM journal of research and development*. 2005;  
1757 49(4.5):777–803.
- 1758 **Li PH**, Lindsey LF, Januszewski M, Zheng Z, Bates AS, Taisz I, Tyka M, Nichols M, Li F, Perlman E, et al. Automated  
1759 reconstruction of a serial-section EM *Drosophila* brain with flood-filling networks and local realignment.  
1760 bioRxiv. 2019; p. 605634.
- 1761 **Lin AC**, Bygrave AM, De Calignon A, Lee T, Miesenböck G. Sparse, decorrelated odor coding in the mushroom  
1762 body enhances learned odor discrimination. *Nature neuroscience*. 2014; 17(4):559–568.
- 1763 **Lin S**, Marin EC, Yang CP, Kao CF, Apenteng BA, Huang Y, O'Connor MB, Truman JW, Lee T. Extremes of lineage  
1764 plasticity in the *Drosophila* brain. *Current Biology*. 2013; 23(19):1908–1913.
- 1765 **Lin TY**, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. In: *Proceedings of the IEEE*  
1766 *international conference on computer vision*; 2017. p. 2980–2988.
- 1767 **Lu Z**, Xu CS, Hayworth KJ, Rivlin P, Plaza SM, Scheffer L, Rubin GM, Hess HF, Meinertzhagen IA. En bloc preparation  
1768 of *Drosophila* brains enables high-throughput FIB-SEM connectomics. bioRxiv. 2019; p. 855130.
- 1769 **Manton JD**, Bates AS, Jagannathan SR, Costa M, Schlegel P, Rohlfing T, Jefferis GS. The natverse: a versatile  
1770 computational toolbox to combine and analyse neuroanatomical data. bioRxiv. 2019; p. 006353.
- 1771 **Marasco A**, Limongiello A, Migliore M. Fast and accurate low-dimensional reduction of biophysically detailed  
1772 neuron models. *Scientific reports*. 2012; 2:928.
- 1773 **McInnes L**, Healy J, Melville J. Umap: Uniform manifold approximation and projection for dimension reduction.  
1774 arXiv preprint arXiv:180203426. 2018; .
- 1775 **Meier M**, Borst A. Extreme Compartmentalization in a *Drosophila* Amacrine Cell. *Current Biology*. 2019;  
1776 29(9):1545–1550.
- 1777 **Meinertzhagen IA**. Connectome studies on *Drosophila*: a short perspective on a tiny brain. *Journal of neuroge-*  
1778 *netics*. 2016; 30(2):62–68.
- 1779 **Meinertzhagen IA**. Morphology of invertebrate neurons and synapses. *Handbook of Invertebrate Neurobiology*.  
1780 2016; p. 1–80.
- 1781 **Meinertzhagen IA**, O'neil S. Synaptic organization of columnar elements in the lamina of the wild type in  
1782 *Drosophila melanogaster*. *Journal of comparative neurology*. 1991; 305(2):232–263.
- 1783 **Miyazaki T**, Ito K. Neural architecture of the primary gustatory center of *Drosophila melanogaster* visualized  
1784 with GAL4 and LexA enhancer-trap systems. *Journal of Comparative Neurology*. 2010; 518(20):4147–4181.
- 1785 **Morgan JL**, Lichtman JW. An individual interneuron participates in many kinds of inhibition and innervates  
1786 much of the mouse visual thalamus. *Neuron*. 2020; .

- 1787 **Namiki S**, Dickinson MH, Wong AM, Korff W, Card GM. The functional organization of descending sensory-motor  
1788 pathways in *Drosophila*. *Elife*. 2018; 7:e34272.
- 1789 **Ohyama T**, Schneider-Mizell CM, Fetter RD, Aleman JV, Franconville R, Rivera-Alba M, Mensh BD, Branson KM,  
1790 Simpson JH, Truman JW, et al. A multilevel multimodal circuit enhances action selection in *Drosophila*. *Nature*.  
1791 2015; 520(7549):633–639.
- 1792 **Omoto JJ**, Keleş MF, Nguyen BCM, Bolanos C, Lovick JK, Frye MA, Hartenstein V. Visual input to the *Drosophila*  
1793 central complex by developmentally and functionally distinct neuronal populations. *Current Biology*. 2017;  
1794 27(8):1098–1110.
- 1795 **Omoto JJ**, Nguyen BCM, Kandimalla P, Lovick JK, Donlea JM, Hartenstein V. Neuronal constituents and putative  
1796 interactions within the *Drosophila* ellipsoid body neuropil. *Frontiers in neural circuits*. 2018; 12:103.
- 1797 **Otsuna H**, Ito K. Systematic analysis of the visual projection neurons of *Drosophila melanogaster*. I. Lobula-  
1798 specific pathways. *Journal of Comparative Neurology*. 2006; 497(6):928–958.
- 1799 **Otsuna H**, Ito M, Kawase T. Color depth MIP mask search: a new tool to expedite Split-GAL4 creation. *bioRxiv*.  
1800 2018; p. 318006.
- 1801 **Panser K**, Tirian L, Schulze F, Villalba S, Jefferis GS, Bühler K, Straw AD. Automatic segmentation of *Drosophila*  
1802 neural compartments using GAL4 expression data reveals novel visual pathways. *Current Biology*. 2016;  
1803 26(15):1943–1954.
- 1804 **Pascual A**, Huang KL, Neveu J, Pr at T. Brain asymmetry and long-term memory. *Nature*. 2004; 427(6975):605–  
1805 606.
- 1806 **Pereanu W**, Kumar A, Jennett A, Reichert H, Hartenstein V. Development-based compartmentalization of the  
1807 *Drosophila* central brain. *Journal of Comparative Neurology*. 2010; 518(15):2996–3023.
- 1808 **Perlman E**. Visualizing and Interacting with Large Imaging Data. *Microscopy and Microanalysis*. 2019;  
1809 25(S2):1374–1375.
- 1810 **Pimentel D**, Donlea JM, Talbot CB, Song SM, Thurston AJ, Miesenb ock G. Operation of a homeostatic sleep  
1811 switch. *Nature*. 2016; 536(7616):333.
- 1812 **Pizer SM**, Amburn EP, Austin JD, Cromartie R, Geselowitz A, Greer T, ter Haar Romeny B, Zimmerman JB,  
1813 Zuiderveld K. Adaptive histogram equalization and its variations. *Computer vision, graphics, and image*  
1814 *processing*. 1987; 39(3):355–368.
- 1815 **Plaza SM**. Focused proofreading: efficiently extracting connectomes from segmented EM images. *arXiv preprint*  
1816 *arXiv:14091199*. 2014; .
- 1817 **Ronneberger O**, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In:  
1818 *International Conference on Medical image computing and computer-assisted intervention* Springer; 2015. p.  
1819 234–241.
- 1820 **Saalfeld S**, Cardona A, Hartenstein V, Toman ak P. CATMAID: collaborative annotation toolkit for massive  
1821 amounts of image data. *Bioinformatics*. 2009; 25(15):1984–1986.
- 1822 **Saalfeld S**, Cardona A, Hartenstein V, Toman ak P. As-rigid-as-possible mosaicking and serial section registration  
1823 of large ssTEM datasets. *Bioinformatics*. 2010; 26(12):i57–i63.
- 1824 **Sanner MF**, et al. Python: a programming language for software integration and development. *J Mol Graph*  
1825 *Model*. 1999; 17(1):57–61.
- 1826 **Segev I**, Fleshman J, Miller J, Bunow B. Modeling the electrical behavior of anatomically complex neurons using  
1827 a network analysis program: passive membrane. *Biological cybernetics*. 1985; 53(1):27–40.
- 1828 **Shinomiya K**, Huang G, Lu Z, Parag T, Xu CS, Aniceto R, Ansari N, Cheatham N, Lauchie S, Neace E, et al.  
1829 Comparisons between the ON-and OFF-edge motion pathways in the *Drosophila* brain. *eLife*. 2019; 8:e40025.
- 1830 **Shinomiya K**, Matsuda K, Oishi T, Otsuna H, Ito K. Flybrain neuron database: a comprehensive database system  
1831 of the *Drosophila* brain neurons. *Journal of Comparative Neurology*. 2011; 519(5):807–833.
- 1832 **Song S**, Sj ostr om PJ, Reigl M, Nelson S, Chklovskii DB. Highly nonrandom features of synaptic connectivity in  
1833 local cortical circuits. *PLoS biology*. 2005; 3(3):e68.

- 1834 **Takemura Sy**, Aso Y, Hige T, Wong A, Lu Z, Xu CS, Rivlin PK, Hess H, Zhao T, Parag T, et al. A connectome of a  
1835 learning and memory center in the adult *Drosophila* brain. *Elife*. 2017; 6:e26975.
- 1836 **Takemura Sy**, Bharioke A, Lu Z, Nern A, Vitaladevuni S, Rivlin PK, Katz WT, Olbris DJ, Plaza SM, Winston P, et al. A  
1837 visual motion detection circuit suggested by *Drosophila* connectomics. *Nature*. 2013; 500(7461):175.
- 1838 **Takemura Sy**, Xu CS, Lu Z, Rivlin PK, Parag T, Olbris DJ, Plaza S, Zhao T, Katz WT, Umayam L, et al. Synaptic  
1839 circuits and their variations within different columns in the visual system of *Drosophila*. *Proceedings of the*  
1840 *National Academy of Sciences*. 2015; 112(44):13711–13716.
- 1841 **Talay M**, Richman EB, Snell NJ, Hartmann GG, Fisher JD, Sorkaç A, Santoyo JF, Chou-Freed C, Nair N, Johnson  
1842 M, et al. Transsynaptic mapping of second-order taste neurons in flies by trans-Tango. *Neuron*. 2017;  
1843 96(4):783–795.
- 1844 **Tanaka NK**, Tanimoto H, Ito K. Neuronal assemblies of the *Drosophila* mushroom body. *Journal of Comparative*  
1845 *Neurology*. 2008; 508(5):711–755.
- 1846 **Tschopp FD**, Reiser MB, Turaga SC. A connectome based hexagonal lattice convolutional network model of the  
1847 *Drosophila* visual system. *arXiv preprint arXiv:180604793*. 2018; .
- 1848 **Turner-Evans DB**, Jensen K, Ali S, Paterson T, Sheridan A, Ray RP, Lauritzen S, Bock D, Jayaraman V. The  
1849 neuroanatomical ultrastructure and function of a biological ring attractor. *bioRxiv*. 2019; p. 847152.
- 1850 **Wolff T**, Iyer NA, Rubin GM. Neuroarchitecture and neuroanatomy of the *Drosophila* central complex: A GAL4-  
1851 based dissection of protocerebral bridge neurons and circuits. *Journal of Comparative Neurology*. 2015;  
1852 523(7):997–1037.
- 1853 **Wolff T**, Rubin GM. Neuroarchitecture of the *Drosophila* central complex: A catalog of nodulus and asymmetrical  
1854 body neurons and a revision of the protocerebral bridge catalog. *Journal of Comparative Neurology*. 2018;  
1855 526(16):2585–2611.
- 1856 **Wu M**, Nern A, Williamson WR, Morimoto MM, Reiser MB, Card GM, Rubin GM. Visual projection neurons in the  
1857 *Drosophila* lobula link feature detection to distinct behavioral programs. *Elife*. 2016; 5:e21022.
- 1858 **Xu CS**, Hayworth KJ, Hess HF, Office UP, editor, Enhanced FIB-SEM Systems for Large-Volume 3D Imaging. Google  
1859 Patents; 2018. US Patent App. 15/883,025.
- 1860 **Xu CS**, Hayworth KJ, Lu Z, Grob P, Hassan AM, García-Cerdán JG, Niyogi KK, Nogales E, Weinberg RJ, Hess HF.  
1861 Enhanced FIB-SEM systems for large-volume 3D imaging. *Elife*. 2017; 6:e25916.
- 1862 **Xu CS**, Pang S, Hayworth KJ, Hess HF. Enabling FIB-SEM Systems for Large Volume Connectomics and Cell  
1863 Biology. *bioRxiv*. 2019; p. 852863.
- 1864 **Yang X**, Bozorgzadeh E, Sarrafzadeh M. Wirelength estimation based on rent exponents of partitioning and  
1865 placement. In: *Proceedings of the 2001 international workshop on System-level interconnect prediction* ACM; 2001.  
1866 p. 25–31.
- 1867 **Yellman C**, Tao H, He B, Hirsh J. Conserved and sexually dimorphic behavioral responses to biogenic amines in  
1868 decapitated *Drosophila*. *Proceedings of the national academy of sciences*. 1997; 94(8):4131–4136.
- 1869 **Yin W**, Brittain D, Borseth J, Scott ME, Williams D, Perkins J, Own C, Murfitt M, Torres RM, Kapner D, et al. A  
1870 Petascale Automated Imaging Pipeline for Mapping Neuronal Circuits with High-throughput Transmission  
1871 Electron Microscopy. *bioRxiv*. 2019; p. 791889.
- 1872 **Zhao T**, Olbris DJ, Yu Y, Plaza SM. Neutu: software for collaborative, large-scale, segmentation-based connectome  
1873 reconstruction. *Frontiers in Neural Circuits*. 2018; 12.
- 1874 **Zhao T**, Plaza SM. Automatic neuron type identification by neurite localization in the *Drosophila* medulla. *arXiv*  
1875 preprint arXiv:14091892. 2014; .
- 1876 **Zheng Z**, Lauritzen JS, Perlman E, Robinson CG, Nichols M, Milkie D, Torrens O, Price J, Fisher CB, Sharifi N,  
1877 et al. A complete electron microscopy volume of the brain of adult *Drosophila melanogaster*. *Cell*. 2018;  
1878 174(3):730–743.
- 1879 **Zhu JY**, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks.  
1880 In: *Proceedings of the IEEE international conference on computer vision*; 2017. p. 2223–2232.