Article

Network statistics of the whole-brain connectome of *Drosophila*

https://doi.org/10.1038/s41586-024-07968-y

Received: 23 August 2023

Accepted: 20 August 2024

Published online: 2 October 2024

Open access

Check for updates

Albert Lin[1,2,8], Runzhe Yang[1,3,8], Sven Dorkenwald[1,3], Arie Matsliah[1], Amy R. Sterling[1], Philipp Schlegel[4,5], Szi-chieh Yu[1], Claire E. McKellar[1], Marta Costa[5], Katharina Eichler[5], Alexander Shakeel Bates[4,5,6], Nils Eckstein[7], Jan Funke[7], Gregory S. X. E. Jefferis[4,5] & Mala Murthy[1]✉

Brains comprise complex networks of neurons and connections, similar to the nodes and edges of artificial networks. Network analysis applied to the wiring diagrams of brains can offer insights into how they support computations and regulate the flow of information underlying perception and behaviour. The completion of the first whole-brain connectome of an adult fly, containing over 130,000 neurons and millions of synaptic connections[1–3], offers an opportunity to analyse the statistical properties and topological features of a complete brain. Here we computed the prevalence of two- and three-node motifs, examined their strengths, related this information to both neurotransmitter composition and cell type annotations[4,5], and compared these metrics with wiring diagrams of other animals. We found that the network of the fly brain displays rich-club organization, with a large population (30% of the connectome) of highly connected neurons. We identified subsets of rich-club neurons that may serve as integrators or broadcasters of signals. Finally, we examined subnetworks based on 78 anatomically defined brain regions or neuropils. These data products are shared within the FlyWire Codex (https://codex.flywire.ai) and should serve as a foundation for models and experiments exploring the relationship between neural activity and anatomical structure.

Network theory has been applied to connectomes at multiple scales to understand brain-wide organization[6–12]. Such analyses quantify the interconnectivity and robustness of a network[13] and can identify highly connected nodes that may act as hubs[14]. They also serve as a basis for comparison across brain regions, individuals, developmental stages or species, enabling researchers to identify commonalities and differences in brain organization.

Mesoscale connectomes have been constructed for the brains of humans and other mammals from MRI and MEG data[6,15,16], relying on functional activity correlations to infer connectivity at the millimetre scale. Rich-club organization has been observed in several mesoscale connectomes: in humans, other mammals[8,9,17] and light-level (not synaptic resolution) connectomes in *Drosophila*[18,19]. Rich-club organization has been proposed to contribute to the ability of brains to efficiently integrate and disseminate information. Advancements in electron microscopy and dense volumetric reconstruction have enabled researchers to examine increasingly larger brain networks at the microscale. These methods do not make assumptions about the relationship between neuron connectivity and functional correlations; nodes and edges can be directly related to neurons and synaptic connections. For example, rich-club neurons in *Caenorhabditis elegans* are part of well-studied motor control circuits[7,20,21]. Patterns of neuronal connectivity that reoccur in different brain areas, known as motifs, have been proposed as 'building blocks' of networks[22,23], and their prevalence in neural networks has been studied to identify organizational principles[7,10–12,24–26]. Motifs such as reciprocal connections[7,11,12,21,27], feedforward loops[7,24,25] and 3-unicycles[12] have received considerable attention in neuroscience owing to their implications for local computation and information flow.

In this study, we characterize the network properties of the whole-brain, synapse-resolution connectome generated by the Flywire Consortium[3,5], the first complete wiring diagram of an adult brain, that of a female *Drosophila melanogaster*[1], and we lay the groundwork for future experimental and theoretical studies in this model system. A summary of data products can be found in Table 1.

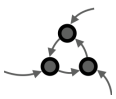
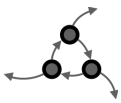
Summary of the dataset and definitions

Here we analyse the v630 snapshot of the *Drosophila* connectome as a directed weighted graph (information on the differences between v630 and v783 is provided in the Methods). Synapses were detected automatically[28] and, for each $\alpha \rightarrow \beta$ neuron pair, the total number of synapses between two neurons defines the connection weight (which we consider to be a proxy for connection strength). Consistent with the flagship connectome paper[3], we applied a threshold of five synapses per connection, unless otherwise noted (Methods); however,

[1]Princeton Neuroscience Institute, Princeton University, Princeton, NJ, USA. [2]Center for the Physics of Biological Function, Princeton University, Princeton, NJ, USA. [3]Computer Science Department, Princeton University, Princeton, NJ, USA. [4]Neurobiology Division, MRC Laboratory of Molecular Biology, Cambridge, UK. [5]Drosophila Connectomics Group, Department of Zoology, University of Cambridge, Cambridge, UK. [6]Centre for Neural Circuits and Behaviour, University of Oxford, Oxford, UK. [7]Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA. [8]These authors contributed equally: Albert Lin, Runzhe Yang. ✉e-mail: mmurthy@princeton.edu

Table 1 | Data availability

Computed network statistics		
Connected components	Strongly connected components	Fig. 1d
	Weakly connected components	Fig. 1e
Path length analysis	Directed shortest path lengths	Fig. 1d
	Undirected shortest path lengths	Fig. 1e
Survival analysis	Neuron removal survival curves	Fig. 1f,g, Extended Data Fig. 2a–d
	Edge removal survival curves	Extended Data Fig. 1a
Rich club analysis	Total degree rich club	Figs. 1h, 4, Extended Data Fig. 4
	In-degree rich club	Extended Data Fig. 2e
	Out-degree rich club	Extended Data Fig. 2e
Small-world analysis	Clustering coefficient	Table 2
	Small-worldness	Equation 1
Large-scale connectivity	Degree distribution	Fig. 1c
	Cell categories	Fig. 4
Spectral analysis	Forward random walk	Extended Data Fig. 1f
	Reversed random walk	Extended Data Fig. 1g
2-neuron motifs	Reciprocity	Table 2
	Connection strength	Fig. 2a,d, Extended Data Fig. 3a,c
	Neurotransmitter types	Fig. 2c,e,f, Extended Data Fig. 3b,d
3-neuron motifs	Motif frequencies	Fig. 3a
	Motif strength	Fig. 3b
	Neurotransmitter types	Fig. 3c–e
Neuropils	Internal/external connection weights	Extended Data Fig. 4
	Reciprocity	Fig. 5, Extended Data Fig. 6, Extended Data Fig. 7
	3-neuron motifs	Fig. 6, Extended Data Fig. 8
	Inter-neuropil reciprocal connections	Fig. 5h, Extended Data Fig. 7e,f

	Neuron lists available on Codex		# of neurons
2-neuron motifs	Reciprocal connection participants		77,607
3-neuron motifs	Feedforward loop participants		113,978
	3-unicycle participants		66,835
N-neuron motifs	Highly reciprocal neurons		2,183
	Neuropil-specific highly reciprocal neurons (NSRNs)		704
Rich club analysis	Rich-club neurons		40,218
	Broadcasters		676
	Integrators		638
Spectral analysis	Attractors		3,469
	Repellers		3,469

List of data products, including computed statistics (left) and neuron populations (right). Interactive neuron lists are available online via Codex (codex.flywire.ai). Definitions for each of these neuron populations can be found in the text, and in Extended Data Table 1.

we also examined the robustness of several results to variation in the synapse threshold (Extended Data Fig. 1b,c and Extended Data Table 2). This gives us a graph with 127,978 neurons and 2,613,129 connections[3] (Fig. 1a and Methods). The connectome also contains synapse-level neurotransmitter predictions[4]. The classifier discriminates between six neurotransmitters: the fast-acting classical neurotransmitters acetylcholine (ach), GABA and glutamate (glut), and the monoamines dopamine (da), octopamine (oct) and serotonin (ser). In *Drosophila*, ach is excitatory and GABA is inhibitory. Glutamate can be either excitatory or inhibitory, but within the brain of the fly, it has largely been observed to be inhibitory[29,30].

The distribution of degrees reflects the amount of connectivity between neurons. A neuron's in-degree is defined as the number of presynaptic partners, and the out-degree is defined as the number of postsynaptic partners. The average in/out-degree of an intrinsic neuron in the brain is 20.5 (ref. 3), but in-degree and out-degree are not highly correlated (Pearson $R = 0.76$, $P < 0.001$; Fig. 1c). The average connection consists of 12.6 synapses[3], and the connection probability is 0.000161. This makes the fly brain a very sparse matrix when compared with the *C. elegans* nervous system or the partial wiring diagrams of larval zebrafish and mouse, with or without a threshold applied (Table 2, Extended Data Table 2 and Extended Data Fig. 1c). This sparsity is due in part to the size of the fly brain. The connection probability is highest among neurons with nearby arbours. Only 3% of neuron pairs have arbours within 50 µm of each other, but such pairs make 71% of all connections (Extended Data Fig. 1d). However, even in this close regime, the connection probability (the likelihood that any two neurons are connected) in the fly is lower than in other wiring diagrams.

Neurons form one connected component

To assess interconnectivity, we searched for strongly (SCCs) and weakly (WCCs) connected components (Methods). Despite its sparsity, the brain forms a highly connected network—93.3% of neurons are contained in a single SCC, while 98.8% of neurons are contained in a single WCC (Fig. 1d,e). These giant connected components, which contain the majority of neurons, persist when either the strongest or weakest connections are pruned (Extended Data Fig. 1a) and are robust to threshold choice (Extended Data Fig. 1b), indicating that connectivity is robust: many paths connect neuron pairs. Within the giant SCC, the average shortest directed path length between neuron pairs is 4.42 hops, with every neuron reachable within 13 hops (Fig. 1d). In the giant WCC, the average shortest undirected path length between neuron pairs is 3.91 hops, with every neuron reachable within 11 hops (Fig. 1e). These numbers are comparable to those found in a similar analysis of the hemibrain dataset[31]. The short path lengths show that the fly brain is relatively shallow compared with artificial networks. The similarities in size and path length distribution between the giant SCC and WCC highlight the prevalence of recurrent connections in the brain: in a mostly feedforward network, one would expect a smaller first SCC with longer path lengths.

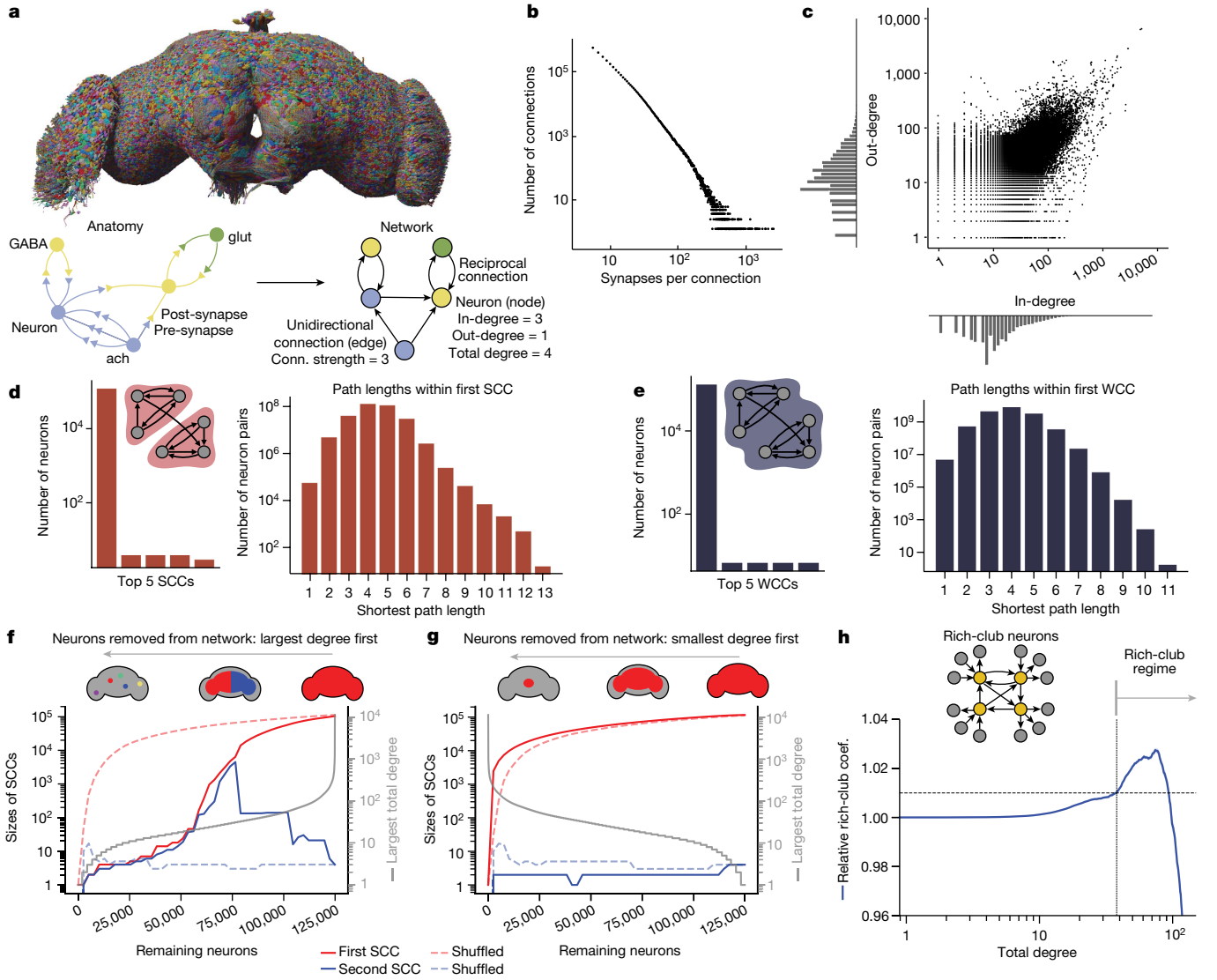


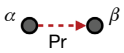
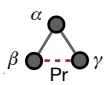
Fig. 1 | Whole-brain network properties. a, The FlyWire dataset[1–3] is an electron microscopy reconstruction of the whole brain of an adult female *Drosophila*. Conn., connection. **b**, The distribution of synapses per connected neuron pair. **c**, In-degrees plotted against out-degrees, with log-scale *x* and *y* axes. **d**, A giant SCC contains 93.3% of all neurons. The distribution of path lengths within this SCC is plotted. **e**, A giant WCC contains 98.8% of all neurons. The distribution of path lengths between neuron pairs within this WCC is plotted. **f**, The sizes of the two largest SCCs as neurons are removed by total degree (2,500 neurons per step). The brain splits into two SCCs when neurons of approximately degree 50 start to be removed, a deviation from when neurons are removed randomly (dotted lines). The largest surviving total degree as a function of the number of remaining neurons is plotted in grey. **g**, Removal of neurons starting with the smallest degree results in a single giant SCC until most neurons are removed (2,500 neurons per step). The smallest surviving total degree as a function of the number of remaining nodes is plotted in grey. For each graph in **f** and **g**, the red and blue curves are read with the left *y* axis and the grey curve is read with the right *y* axis. **h**, The relative rich-club coefficient (coef.) as a function of total degree, computed relative to CFG models. We take neurons with total degree >37 to be within the rich-club regime.

To assess whether high interconnectivity is a consequence of a relatively large number of interconnected neurons, or a small number of very highly connected neurons, we constructed survival curves, observing whether large connected components persist when neurons are removed. Removing neurons from the directed network, starting with those of largest total degree, we find that the first SCC persists until a total degree of 50, at which point the network splits into two SCCs of roughly equal size (Fig. 1f). These two SCCs correspond to the left and right hemispheres, and do not split into separate networks until about 60% of neurons are removed. Removing neurons from the network by smallest total degree does not result in division of the first SCC (Fig. 1g). These results highlight the brain's robust interconnectivity—it is not dependent on a small population of neurons. We observed similar behaviour in the undirected network (Extended Data Fig. 2a,b). These results remain qualitatively consistent when neurons are removed by in-degree or out-degree alone (Extended Data Fig. 2c,d).

Spectral analysis of the brain

To better understand the brain's topology, we performed a spectral analysis of a random walk in the giant SCC (Methods), which converges to a stationary distribution over all neurons in which 3% of neurons were visited 61.2% of the time (Extended Data Fig. 1f). These top-visited neurons are classified as attractor nodes. These attractors often make connections in the gnathal ganglia (GNG), a large midline neuropil that sends and receives information from the periphery and contains many connections to the ventral nerve cord (VNC). We also performed a reverse walk within the giant SCC, which converges to a stationary

Table 2 | Connection probabilities, reciprocity, and clustering coefficiens

Neuronal wiring diagrams	Fly *Drosophila melanogaster* (Dorkenwald et al.[3])	Nematode *Hermaphrodite C. elegans* (Cook et al.[21])	Nematode *Male C. elegans* (Cook et al.[21])	Zebrafish (sub-vol.) *Larval Danio rerio (hindbrain)* (Yang et al.[12])	Mouse (sub-vol.) *Mus musculus (V1 L2/3)* (Turner et al., 2022)
					
Network size 	127,978 neurons 2,613,129 connections	302 neurons 3,242 connections	364 neurons 3,467 connections	419 neurons 5,605 connections	111 neurons 659 connections
Avg. connection strength 	12.61 synapses 5~2358	3.15 synapses 1~36	3.59 synapses 1~63	1.69 synapses 1~21	1.14 synapses 1~5
Connection probability 	0.000160 x1	0.0356 x222 denser than fly	0.0262 x164 denser than fly	0.0320 x200 denser than fly	0.0540 x360 denser than fly
Connection reciprocity 	0.138 x858 than ER x43.8 than CFG x45.9 than NND x7.22 than NPC	0.372 x10.4 than ER x5.03 than CFG	0.386 x14.7 than ER x6.02 than CFG	0.113 x3.53 than ER x2.64 than CFG	0.088 x1.63 than ER x1.33 than CFG
Clustering coefficient 	0.0463 x144 than ER x7.57 than CFG x10.9 than NND x2.88 than NPC	0.284 x4.06 than ER x1.86 than CFG	0.331 x6.39 than ER x2.40 than CFG	0.182 x2.89 than ER x1.90 than CFG	0.159 x1.51 than ER x1.06 than CFG

The probability that any two neurons is connected is 0.000160. Connection reciprocity in the fly is 0.138, larger than in Erdős–Rényi (ER), configuration (CFG), neuron-neuron distance (NND), or neuropil connection (NPC) models (Methods). The clustering coefficient in the fly is 0.0463. Both reciprocity and clustering coefficient are higher than expected with ER, CFG, NND, and NPC models. Statistics for *C. elegans* were computed for the chemical networks of neurons in hermaphrodite and male worms[21]. Statistics for larval zebrafish hindbrain[12] and mouse visual cortex[11] were computed excluding any truncated neurons. Note that direct comparisons of graph metrics must be made with caution (see Discussion).

distribution in which 3% of neurons were visited 42.4% of the time (Extended Data Fig. 1g). These neurons are repeller nodes, and include many with synapses in the antennal lobes (AL) and medullae (ME), brain regions that are close to the olfactory and visual periphery, respectively.

The fly brain has a large rich club

Many networks exhibit the rich-club property[8,14,20], in which well-connected nodes are preferentially connected to other well-connected nodes (Methods). Comparing the observed network to a degree-preserving configuration (CFG) random model, there exists a rich-club regime in the fly brain, with a cut-off of total degree = 37 (Fig. 1h and Extended Data Fig. 2e–g). This regime contains 40,218 neurons, approximately 30% of neurons in the brain (Extended Data Fig. 2e). The connection probability within the rich club is 0.000870, 5.4 times higher than the overall connection probability. The size of the rich club suggests that the network topology of the fly brain is fairly distributed, consistent with the connected component results. A rich-club analysis considering in-degree alone finds a rich-club regime with an in-degree threshold of 10, while no rich club is observed when considering out-degree alone (Extended Data Fig. 2f). Note that, as we continue to remove nodes, we enter a regime (beyond total degree 92) where the remaining neurons are no longer preferentially connected to each other (Fig. 1h). We further consider this very-high-degree regime in the Discussion.

To examine the extent to which long-distance connections contribute to the rich-club property that we observe in the fly brain, we constructed an extension of the CFG model in which we constrain the random network by enforcing the measured connection probabilities between the 78 neuropils. By fixing the number of connections between and within neuropils, this neuropil connection (NPC) model implicitly contains mesoscale spatial information (Methods). We find that the observed network is not more highly connected than the NPC model (Extended Data Fig. 2g), indicating that interneuropil connections contribute to the rich-club effect.

The fraction of neurons in the rich-club regime in the fly is substantially larger than in *C. elegans*, which has a rich club of 11 neurons (4% of the neurons in the worm)[20]. Note that the differences in rich-club size are sensitive to the criteria used to determine the rich-club cut-off, and may also be a consequence of the different scales of these two networks. Nonetheless, it is interesting to note that, while the worm rich club contains known hub neurons, such as the command neurons AVA and AVB, such highly connected hub neurons do not seem to be present in the fly brain—although there are neurons with very high degrees, there also exist alternative paths between most neuron pairs.

Recurrent motifs are over-represented

Connection reciprocity is a measure of the amount of direct feedback in the brain. Across the brain, this connection reciprocity probability is 0.138 (Table 2), higher than in both the Erdős–Rényi (ER) and CFG random models (Methods). The over-representation of reciprocal connections in brains is well established, and our results are consistent with previous observations both in *Drosophila*[31–33] and in other species[7,11,21,24,25,34]. The clustering coefficient, a metric that assesses the prevalence of triplet structures in the network, is 0.0477 in the fly brain (Table 2), higher than in both the ER and CFG models. The high clustering coefficient demonstrates that the network is highly connected and is non-random in its structure.

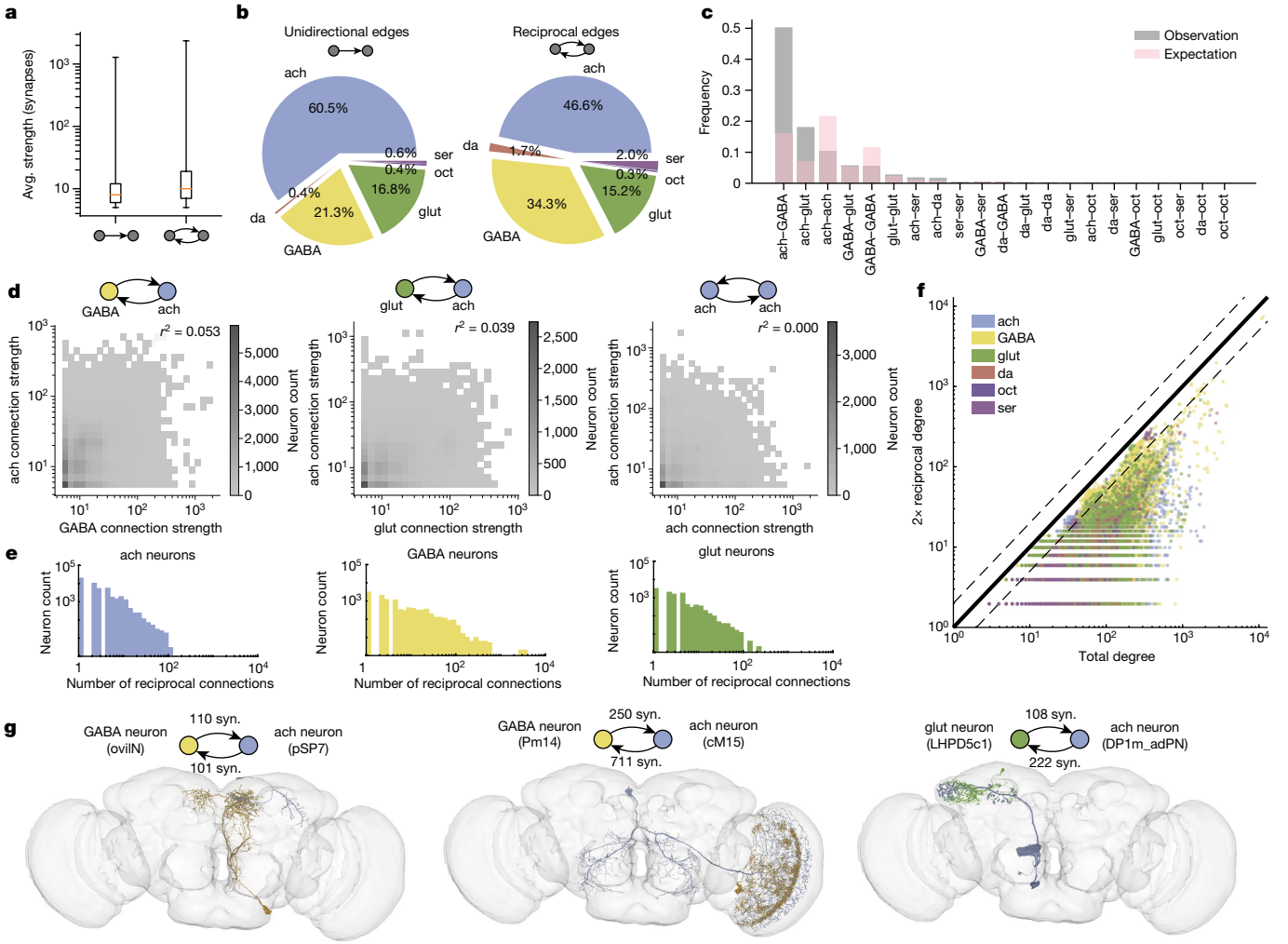


Fig. 2 | Characterizing reciprocal connections in the brain. a, The average (avg.) reciprocal and unidirectional edge weights. **b**, Breakdown of unidirectional and reciprocal edges by neurotransmitter. **c**, The frequency of neurotransmitter pairs forming reciprocal connections (grey) compared with the expected frequency of neurotransmitter pairs under the assumption of independent neurotransmitter choice (pink). **d**, The relative strengths (synapse (syn.) counts) of the two connections forming ach–GABA reciprocal pairs (left), ach–glutamate reciprocal pairs (centre) and ach–ach reciprocal pairs (right). **e**, The distributions of reciprocal degree for cholinergic neurons (left), GABAergic neurons (middle) and glutamatergic neurons (right). **f**, Scatter plot of 2× reciprocal degree against total degree. The dotted lines indicate a factor of 2 around the $x = y$ line. **g**, Visualizations of exemplar reciprocal neuron pairs.

We present these metrics alongside those for two existing whole-animal connectomes, the hermaphrodite and male *C. elegans*[7,21,35], and two subvolume wiring diagrams, the hindbrain of a larval zebrafish[12] and a region of L2/3 mouse visual cortex[11] (Table 2). Despite differences in the sparsity of the different brain networks, the values of reciprocity and clustering coefficient are comparable across all five datasets. The ER and CFG models do not contain any spatial information, instead assuming that any neuron pair may randomly connect. Our NPC model contains mesoscale spatial information but does not make distance comparisons for each neuron pair. We therefore constructed a neuron–neuron distance (NND) model with two different spatial zones of connection probability, informed by the distribution of connections as a function of distance (Methods and Extended Data Fig. 1e). We computed the reciprocity and clustering coefficient for both the NND and NPC models and found that the metrics for the real network were higher than in both spatial null models, suggesting that the non-random connectivity observed in the brain is not solely a consequence of spatial or morphological constraints.

Direct comparisons of networks of different sizes should be interpreted with caution[36–38]. While the fly and worm datasets are complete brains, the zebrafish and mouse datasets are derived from brain subvolumes with truncated neurons. Moreover, differences in synapse detection and thresholding will impact topological metrics such as connection probability and reciprocity (Discussion). Varying the synapse threshold in the fly did not substantially change these metrics (Extended Data Fig. 1c and Extended Data Table 2).

Small-worldness of the fly brain

A small-world network is one in which nodes are highly clustered and path lengths are short[13]. High small-worldness coefficients are associated with efficient communication between nodes[39]. The small-worldness coefficient of the whole-brain fly connectome is $S^{\Delta} = 141$, higher than that of the *C. elegans* connectome ($S^{\Delta} = 3.21$) and close to that of the internet ($S^{\Delta} = 98.1$)[13], implying highly effective global communication among neurons in the brain.

Neurotransmitter makeup of reciprocal connections

The average strength of edges participating in reciprocal connections is higher than the average strength of unidirectional connections (Fig. 2a). The majority of unidirectional connections are cholinergic (excitatory),

while the edges participating in reciprocal connections contain fewer cholinergic neurons and more GABAergic neurons (Fig. 2b). Inhibitory connections in the brain have more synapses on average compared with excitatory connections[3], which may partially explain the higher average strength of reciprocal connections. The most common reciprocal pairing is ach–GABA, and the second most common pairing is ach–glutamate (Fig. 2c). Both of these reciprocal motifs are excitatory–inhibitory, and both are over-represented when compared to the neurotransmitter frequencies observed for reciprocal connections (Fig. 2b). By contrast, excitatory–excitatory ach–ach pairs are under-represented, as are inhibitory–inhibitory GABA–GABA pairs. We observed reciprocal excitatory–inhibitory (ach–GABA and ach–glut) connection strengths to be only weakly correlated, whereas ach–ach pairs were uncorrelated (Fig. 2d). Examples of reciprocal neuron pairs are shown in Fig. 2g.

Reciprocal degree across the population

Of the 127,978 neurons in the brain, 77,607 participate in at least one reciprocal connection: approximately 2 in every 3 neurons (Methods). Many neurons participate in multiple reciprocal connections. We define reciprocal degree as the number of reciprocal connections made by a given neuron (Extended Data Fig. 3a). Plotting the distributions of reciprocal degree by neurotransmitter, we observe that the majority of neurons with a high reciprocal degree ($d^{rec} > 100$) are GABAergic (Fig. 2e and Extended Data Fig. 3b), while, at lower reciprocal degrees ($d^{rec} < 100$), all three primary neurotransmitter types are well represented.

For most neurons, the fraction of reciprocal connections is low—on average 23% of incoming and 18% of outgoing connections are reciprocal. Plotting the fraction of reciprocal incoming connections against the fraction of reciprocal outgoing connections, we observed only a weak correlation (Extended Data Fig. 3c), suggesting that the reciprocal degree is not strongly coupled to either in-degree or out-degree. Similarly, there is no relationship between the total degree and the reciprocal degree (Fig. 2f). However, dividing the neuron population by neurotransmitter, we find that neurons of high total degree are mostly GABAergic and, for many of these neurons, more than half of their total connections are reciprocal (Extended Data Fig. 3d).

Neurotransmitter composition of three-node motifs

The high clustering coefficient of the brain implies an over-representation of triplet structures. We determined the frequency at which each of the 12 directed three-node motifs occur in the brain (Fig. 3a). Feedforward motifs (motifs 1–3) are under-represented when compared to both ER and CFG models, while all others, including the highly recurrent motifs (motifs 7–13), are over-represented. The strengths of edges participating in three-node motifs are higher than the average edge strength (Fig. 3b). Complex three-node motifs that contain reciprocal connections tend to be stronger than feedforward motifs.

Examining the neurotransmitter composition of two of these motifs, feedforward loops (motif 4) and 3-unicycles (motif 7) (Fig. 3c), we found that edges that participate in feedforward loops were predominantly cholinergic, and that the most common neurotransmitter composition is three cholinergic neurons, a feedforward excitatory configuration (Fig. 3d). The next most common compositions contain either one or two inhibitory (GABAergic or glutamatergic) edges. Feedforward loops with one inhibitory edge are probably feedforward inhibition motifs, whereas loops with two inhibitory edges are probably disinhibition motifs. By contrast, 3-unicycles contain a higher proportion of inhibitory GABAergic and glutamatergic neurons, and the three most common 3-unicycle compositions all contain at least one inhibitory neuron (Fig. 3e). These cycles may act as indirect feedback inhibition circuits. Notably, observed neurotransmitter composition frequencies are closer to what may be expected by chance for feedforward loops than for 3-unicycles. Examples of neurons that form three-node motifs are shown in Fig. 3f.

A high clustering coefficient and an over-representation of highly connected motifs suggest that the local structure of the brain displays a high degree of non-randomness, consistent with previous studies in *C. elegans*[7,21] and in the mouse cortex[11,24,25]. The over-representation of feedforward loops has been widely observed in other biological networks, such as in the rat cortex and *C. elegans*[7,21,24,25]. This over-representation is present in most neuropils in the brain.

Large-scale connectivity in the brain

Neurons with few inputs and many outputs may function as broadcasters of signals, while those with many inputs and few outputs may act as integrators. To examine such neurons, we divided the intrinsic rich-club neurons into three categories on the basis of their relative in-degree and out-degree (Fig. 4a), defining broadcasters as those for which out-degree ≥ 5 × in-degree and integrators as those for which in-degree ≥ 5 × out-degree. These definitions, while arbitrary, enable us to describe neuron populations with unbalanced in-/out-degrees. Given these definitions, we find 676 broadcasters and 638 integrators. The remaining intrinsic rich-club neurons (37,093), including most highly reciprocal neurons, fall into the balanced category.

In comparison to the overall population, rich-club neurons are less likely to be cholinergic and more likely to be GABAergic (Fig. 4b and Extended Data Fig. 4a). Integrators are even less likely to be cholinergic (49%), and include a large fraction of dopaminergic neurons that may be engaged during learning. By contrast, broadcasters are predominantly cholinergic (75%). Central brain neurons are substantially over-represented in the rich club, while optic lobe intrinsic neurons are under-represented (Fig. 4c and Extended Data Fig. 4b). Many integrators are either central-brain-intrinsic neurons or visual projection neurons, while few broadcasters are intrinsic to the central brain—many are visual centrifugal neurons or optic lobe intrinsic neurons. Broadcasters include a large number of Mi1 and Tm3 neurons—excitatory cells in the ME known to have key roles in the motion detection circuit[32,40]. Most neurons are restricted to a single hemisphere—11% have inputs in both hemispheres and 11% have outputs in both hemispheres[3] (Extended Data Fig. 4c). Rich-club neurons are more likely to have inputs or outputs spanning both hemispheres: 18% and 17%, respectively. This is more common for integrators (23%) than for broadcasters (16%). Some examples of broadcasters, integrators and balanced neurons are shown in Fig. 4d.

The rich club is close to sensory inputs

To assess the distance of the rich-club neurons from sensory inputs, we used a probabilistic information flow model to determine the relative distance of each neuron from a set of seed neurons[3,41] (Methods). The model was run with different sets of seed neurons, each corresponding to a specific set of sensory inputs to the brain (olfactory, gustatory and so on). This includes all sensory receptors projecting to the central brain as well as the population of visual projection neurons (which provide visual input to the central brain and were used here instead of the photoreceptor neurons, inputs to the optic lobe, because photoreceptor proofreading was incomplete in v630 of the data). Ranking these distances and normalizing returned the percentile rank of each neuron with respect to each modality. Neurons with percentile rank < 50% are closer than average to the given input, while neurons with percentile rank > 50% are farther.

The rich-club neurons have a mean percentile rank of 44% relative to the set of all sensory inputs (Fig. 4e). Integrators have a mean percentile rank of 43%, while broadcasters have a mean percentile rank of 53%. The distribution of broadcasters is bifurcated, with one peak closer to inputs

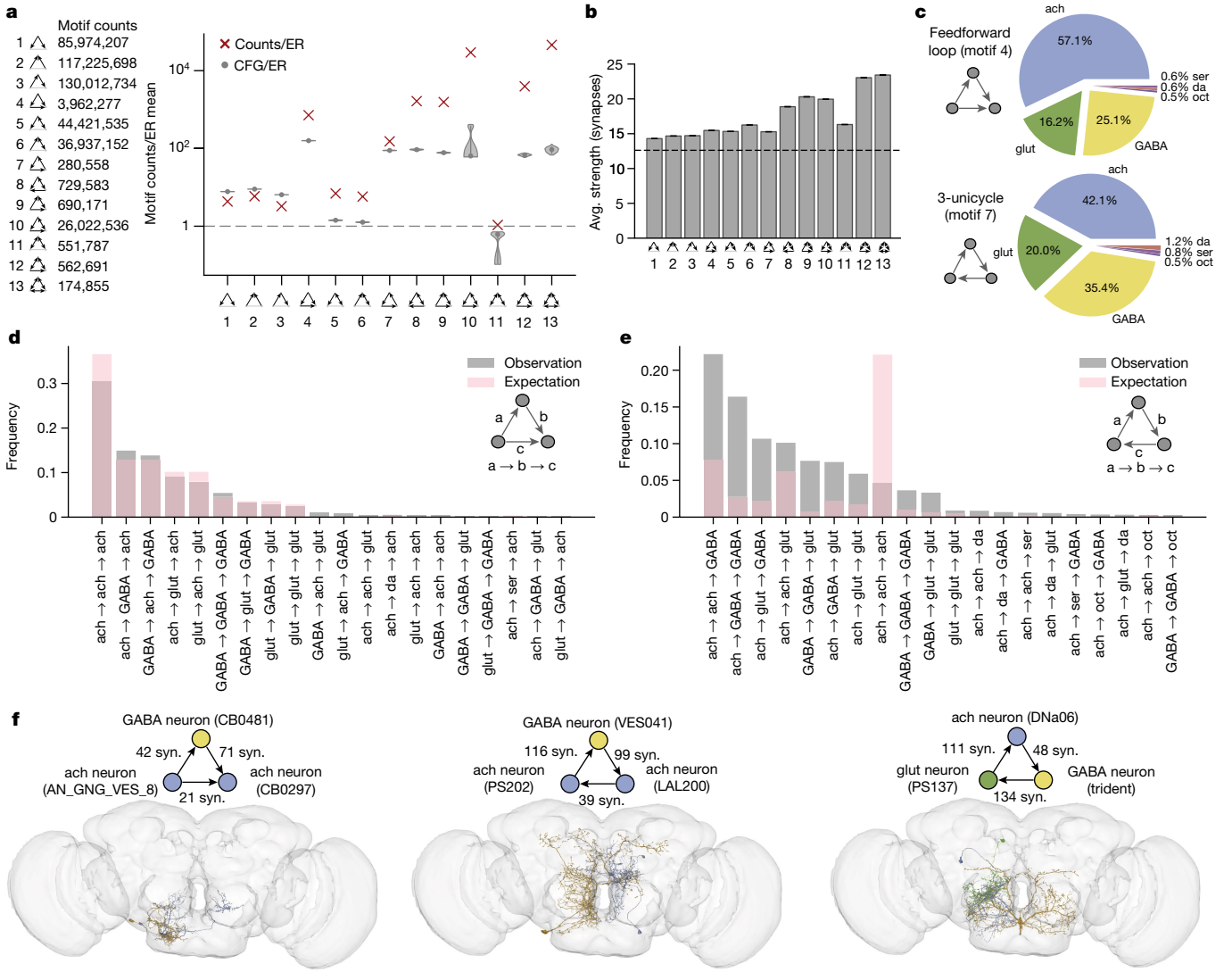


Fig. 3 | Examining three-node motifs. a, The distribution of three-node motifs across the whole brain. Absolute counts of each motif are shown on the left, and the frequency of each motif relative to that in an ER null model is plotted on the right, together with the average motif frequencies of 100 CFG models (grey violin plots). **b**, The average strength of three-node motif edges. The dotted line is the average connection strength in the brain. **c**, Breakdown by neurotransmitter of edges participating in two motifs: feedforward loops (motif 4) and 3-unicycles (motif 7). **d**,**e**, Neurotransmitter compositions of feedforward loops (motif 4) (**d**) and 3-unicycles (motif 7) (**e**). **f**, Visualizations of exemplar three-node motifs.

and another peak far from inputs. The second peak consists primarily of broadcasters in the optic lobe, which are far from sensory inputs to the central brain (Extended Data Fig. 4d). Examining the ranks with respect to individual sensory modalities, we find that rich-club neurons are again closer than average to each modality (Fig. 4f and Extended Data Fig. 4e). Examining the distance of neurons to multiple sensory modalities, we find that broadcasters and integrators, while scattered throughout these distributions, tend to be closer than average to multiple sensory inputs (Extended Data Fig. 4f). Integrators and broadcasters that are low in rank relative to multiple modalities may be candidate sites of multisensory integration and information propagation.

Network differences across brain regions

The *Drosophila* connectome has been segmented into 78 neuropils (Fig. 5a), each with different average connection strengths[3]. To understand information flow between neuropils, we used a fractional weighting method accounting for each neuron's projections to and from every neuropil[3] (Methods). We computed for each neuropil the relative fraction of internal, external incoming and external outgoing connection weights (Extended Data Fig. 5a,b). These fractions reflect, respectively, the net number of connections within, being received and being sent from each neuropil.

We find differences in these fractions across brain regions: the ellipsoid body (EB) and fan-shaped body (FB) of the central complex have the highest fraction of internal connections, while in other regions, such as the mushroom body (MB), the majority of connections are external (Extended Data Fig. 5b). Some regions, such as the lateral horn, send more connections than they receive, while others, such as the lobula plate, receive more connections than they send. The fraction of internal connection weights is not correlated with neuropil size: while large neuropils such as the anterior and posterior ventrolateral protocerebra (AVLP and PVLP) have sizable fractions of internal weights, they do not rank the highest. Note that internal weights include any neurons with endings outside the brain, such as sensory, ascending and descending neurons, accounting for the high fraction of internal weights in regions such as the ME, which receive inputs from R7 and R8 photoreceptors, and the GNG, which connects with large numbers of ascending and

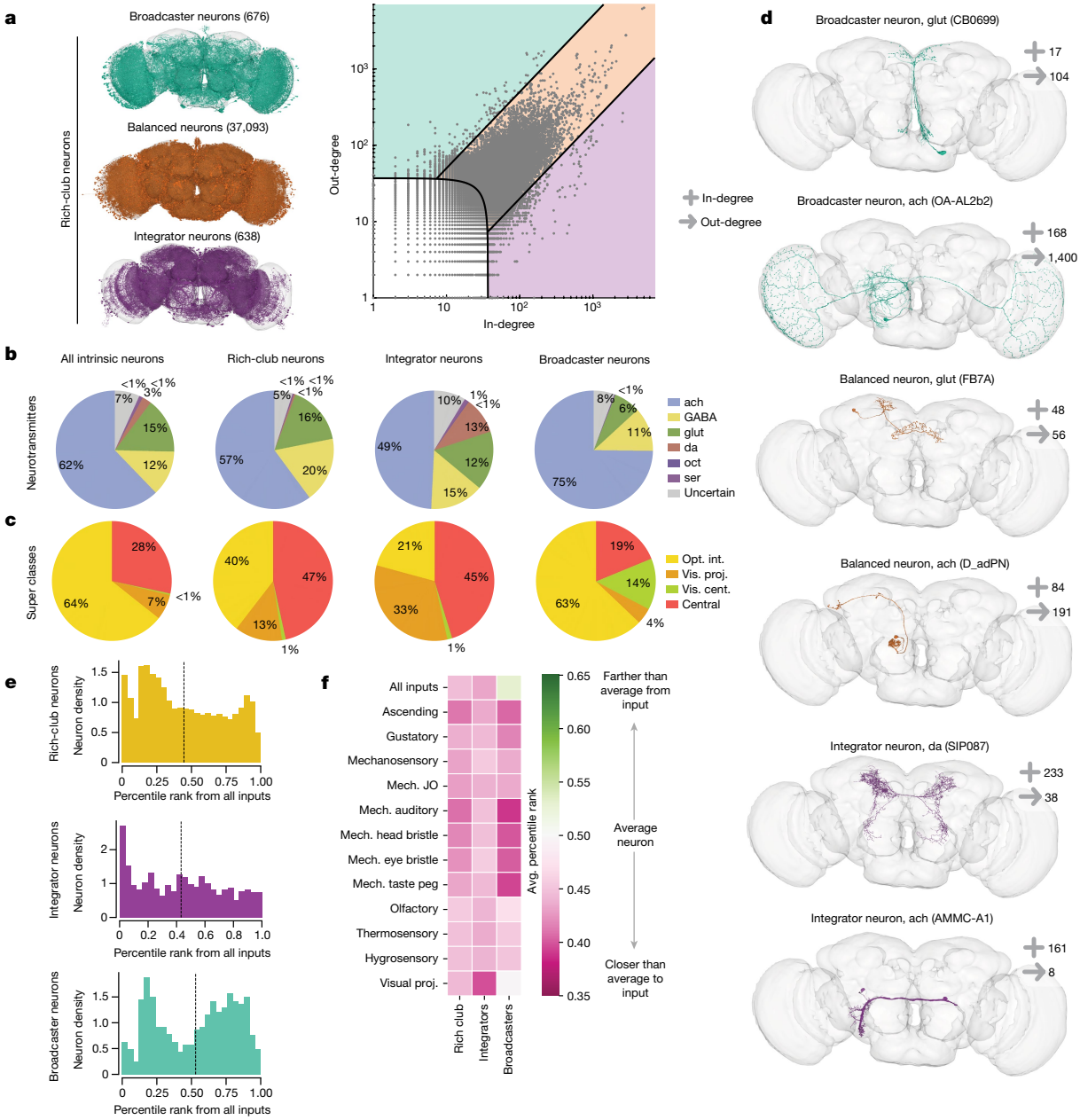


Fig. 4 | Large-scale neuron connectivity in the brain. a, We grouped the intrinsic rich-club neurons into three categories by in-degree/out-degree ratio: broadcasters, integrators and large balanced neurons. **b**,**c**, Comparison of the prevalence of neurotransmitters (**b**) and super classes (**c**) of all intrinsic neurons, rich-club neurons, integrators and broadcasters. opt. int., optic lobe intrinsic; vis. cent., visual centrifugal; vis. proj., visual projection. **d**, Examples of rich-club neurons in these three categories. **e**, Applying the information flow model from refs. 3,41, we determined the percentile rank distributions of rich-club, integrator and broadcaster neuron populations from all inputs to the brain, as well as to specific modalities (Extended Data Fig. 4e). **f**, The average percentile rank of rich-club, integrator and broadcaster neurons for different modalities. JO, Johnston's organ; Mech., mechanosensory.

descending neurons. Across the brain, 52% of all connection weights can be classified as internal. Comparing the neurotransmitters of the neurons contributing connection weights, we see that internal connections are more likely than external ones to be inhibitory (GABAergic or glutamatergic) (Extended Data Fig. 5c). We also see differences in the neurotransmitter composition across regions (Extended Data Fig. 5d).

Reciprocal connections in neuropils

To perform motif analyses within each neuropil, we identified neuropil subnetworks that treat all connections within a neuropil as edges (Fig. 5b and Extended Data Fig. 6a). Different neuropils differ in connection strength and density (Extended Data Fig. 6b). Computing the reciprocity in each neuropil (Fig. 5c and Extended Data Fig. 6c), we found regions with high reciprocity probabilities, including the central complex (FB, EB and noduli (NO)) and the ALs. The relative number of reciprocal connections is high in the MB and ME (Extended Data Fig. 6b). Note that, for these motif analyses, the results for small neuropils such as the cantles, bulbs, galls, accessory ME and ocellar ganglion are less interpretable owing to the small number of neurons.

In most neuropils, as in the whole brain, reciprocal connections are stronger than unidirectional connections (Extended Data Fig. 6d). Exceptions include the protocerebral bridge, MB calyces and bulbs, which have stronger unidirectional than reciprocal connections. There are differences in the relative prevalence of each neurotransmitter in reciprocal and unidirectional connections across neuropils (Fig. 5d,e

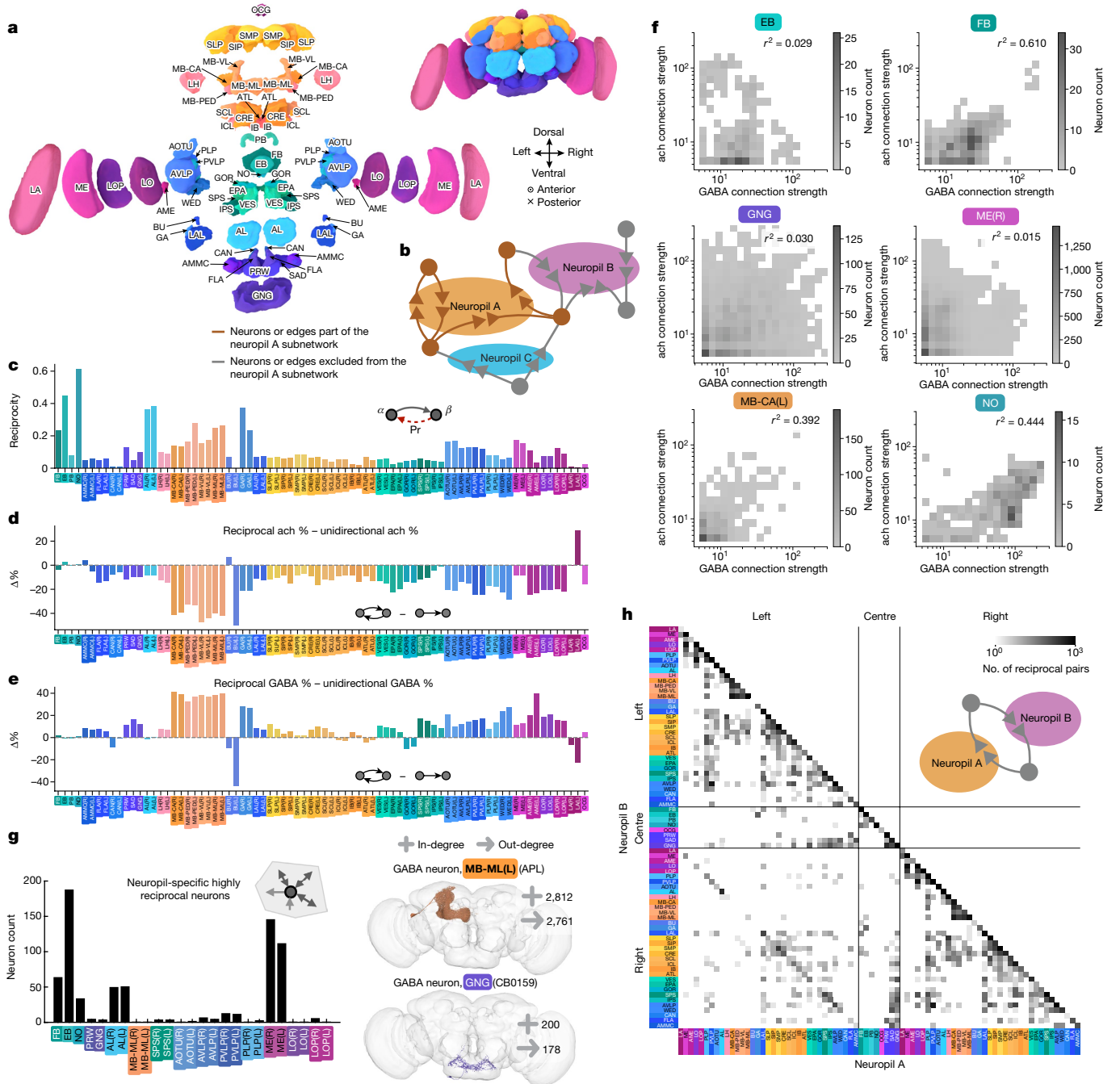


Fig. 5 | Neuropil-specific differences in connectivity. a, An exploded view of the brain showing the neuropils of the *Drosophila* brain. Each synapse is assigned to a neuropil based on its location. **b**, Schematic of how neuropil subnetworks are identified for motif analyses. **c**, Reciprocity within each neuropil subnetwork. **d**,**e**, The differences in the percentage of cholinergic (**d**) and GABAergic (**e**) edges between reciprocal and unidirectional connections across different neuropils. The absolute percentages are shown in Extended Data Fig. 6. **f**, The relationship between excitatory and inhibitory connection strengths in reciprocal connections in different brain regions. **g**, The number of NSRNs in different neuropils. Examples of NSRNs are shown to the right. **h**, Map of the total number of reciprocal pairs between different neuropils. Examples of such pairs are shown in Extended Data Fig. 7e. Definitions for neuropils are provided in Supplementary Fig. 1. L and R in parentheses indicate the left and right side of the brain.

and Extended Data Fig. 6d–h): while, in most neuropils, reciprocal connections contain fewer cholinergic and more GABAergic edges, there are notable exceptions, such as in the neuropils of the central complex (FB, EB, protocerebral bridge and NO). In the MB, we find especially large differences in the neurotransmitter composition between unidirectional and reciprocal connections. Comparing the edge strengths of reciprocal excitatory–inhibitory (ach–GABA) connections within neuropil subnetworks, we observe that excitatory–inhibitory connection strengths are more strongly correlated in some neuropils (such as the FB and NO) than in others (Fig. 5f and Extended Data Fig. 7a,b). These correlations do not appear to be dependent on neuropil size (Extended Data Fig. 7c).

Identifying NSRNs

We searched for intrinsic highly reciprocal rich-club neurons that make the majority of their connections within a single neuropil, and found 1,863 neurons that meet these criteria (Fig. 5g). These neuropil-specific

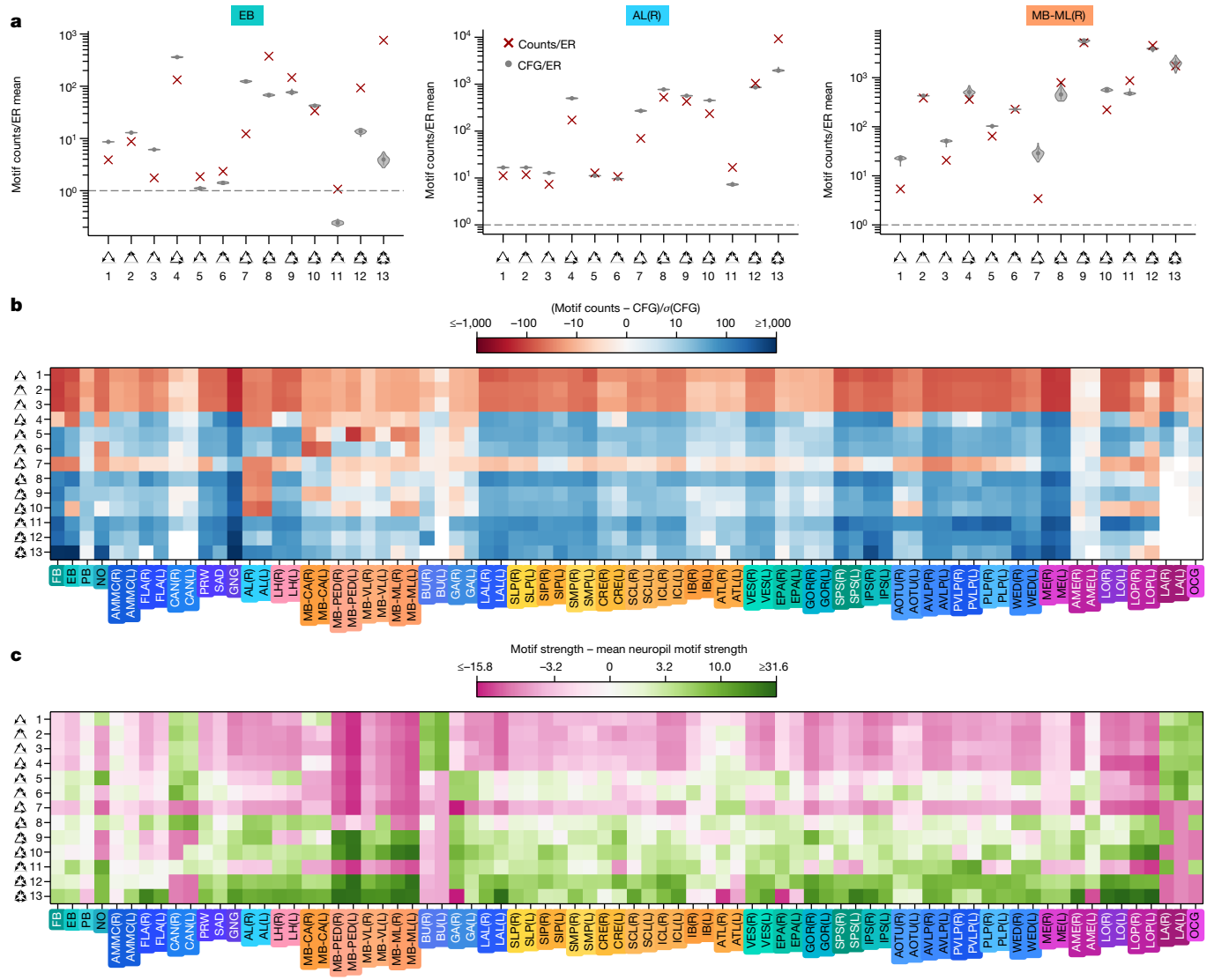


Fig. 6 | Differences in three-node motifs across neuropils. a, Three-node motif distributions for three example neuropils: the EB, AL(R) and right mushroom body medial lobe (MB-ML(R)). The frequency of each motif relative to that in an ER null model is plotted on the right, together with the average motif frequencies of 100 CFG models (grey violin plots). Further examples of other neuropils are shown in Extended Data Fig. 8a. **b**, The motif frequencies for the three-node motifs across all 78 neuropil subnetworks, normalized to their respective CFG models. **c**, The average strengths of edges participating in three-node motifs in the different neuropil subnetworks relative to the average three-node motif strength in each subnetwork. Average strengths relative to the average neuropil subnetwork edge strength are shown in Extended Data Fig. 8b.

highly reciprocal neurons (NSRNs) are predominantly inhibitory: 54% are GABAergic and another 10% are glutamatergic (Extended Data Fig. 7d). In some neuropils, such as the AL and EB, there are many NSRNs, while in other neuropils, there exist only a handful of such neurons. Some NSRNs, like the APL neurons in the MB[42], CT1 neurons in the lobula (LO)[43] or AL local neurons (ALLNs)[44], are known to provide global feedback inhibition in these regions. These neurons tend to be highly branched, with individual processes making reciprocal connections with different feedforward neurons. Some have been shown to have compartmentalized activity, raising the possibility of local computation[45,46]. NSRNs that have yet to be characterized may have similar roles in other circuits—for example, it is likely that NSRNs found in the AVLP provide feedback to the auditory circuits in this region[47].

Interneuropil reciprocal connections

While many reciprocal connections occur within single neuropils, 12.1% of all reciprocal pairs are formed by connections in two neuropils (Methods). Mapping these reciprocal connections, we find off-diagonal terms that represent pairs connecting across neuropils (Fig. 5h). The compartments of the MB are linked by many reciprocal connections, while the neuropils of the suboesophageal zone, including the GNG, saddle (SAD) and prow (PRW), form a connected block. Strong reciprocal connectivity also occurs across the midline, such as between the left and right ALs. Neuropils close to the midline, such as superior medial protocerebrum (SMP), superior posterior slope (SPS) and inferior posterior slope (IPS), tend to have many cross-hemispheric reciprocal connections. There also exist reciprocal connections that span the central brain, such as those between the left and right anterior optic tubercles (AOTUs) or the left and right lateral accessory lobes (LALs), demonstrating that recurrent motifs are not limited to local connections—they also exist at large spatial scales. Examples of such pairs are shown in Extended Data Fig. 7e.

We also compared the map of reciprocal connections to the projectome map of all connections (figure 4 of ref. 3). For example, the superior lateral protocerebrum (SLP) and superior intermediate protocerebrum

(SIP) are connected to the FB in the projectome but have no reciprocal connections. Similarly, the laminae has many internal connections but few reciprocal ones. Examining ach–GABA reciprocal connections highlights deviations from symmetry that represent a net imbalance of excitatory–inhibitory reciprocal connections (Extended Data Fig. 7f). For example, between the LO and PVLP, most ach–GABA reciprocal connections share directionality: the ach connections are in the LO and the GABA connections are in the PVLP.

Three-node motifs differ across regions

We computed the prevalence of three-node motifs in each neuropil subnetwork relative to ER and CFG null models (Fig. 6a and Extended Data Fig. 8a). Across most neuropils, we observed the same trend as we do across the entire brain: an under-representation of feedforward motifs (1–3) and an over-representation of complex motifs (Fig. 6b). However, there are notable differences between neuropils. In the cantles, epaulettes and gorgets, for example, the frequency of three-node motifs was closer to that expected in a CFG null model, while, in the FB and EB, complex motifs are particularly over-represented, consistent with their high reciprocity (Fig. 6b).

Feedforward loops (motif 4) are over-represented in most neuropils, except for the FB, EB, NO and MB compartments. This suggests a relative under-representation of both feedforward excitatory and feedforward inhibitory circuits in these brain regions. 3-unicycles (motif 7), an indirect feedback inhibition circuit, are over-represented across the entire brain (Fig. 3c) but are under-represented in most neuropils. The notable exceptions, the ME and GNG, are very large neuropils with many sensory inputs. The over-representation of 3-unicycles in the ME implies the existence of localized cyclic structures in early visual circuitry. Notably, this motif is also over-represented in the zebrafish oculomotor circuit[12]. Motifs 7–10 are under-represented in the ALs, perhaps a result of the small number of unidirectional edges in these regions.

In most neuropils, we find that edges participating in under-represented motifs are weaker on average than edges participating in over-represented motifs (Fig. 6c), and that edges participating in three-node motifs are stronger than average (Extended Data Fig. 8b). This is broadly consistent with the whole-brain three-node motif strength results. A notable exception is the laminae, where feedforward connections are strong despite being under-represented.

Discussion

The completion of a wiring diagram of the full brain of *D. melanogaster* enabled us to characterize its network properties, laying the groundwork for connecting circuit motifs to biological function. Our analysis leverages spatial information, neuron class distinctions (for example, sensory neurons versus descending neurons), cell type labels (annotations) and neurotransmitter predictions to interpret the network features we uncovered. Despite its sparsity, the fly brain is a robust and highly interconnected network, with over-represented reciprocal and recurrent motifs spanning multiple brain regions. These long-range connections may explain results from recent brain-wide neural recording experiments in fly[48–50] and other species[51–53] that have uncovered widespread activity related to both sensory processing and behaviour (beyond canonical pathways and brain regions). We provide searchable and filterable lists of the neurons discussed in this study, as well as a motif search and visualization tool to facilitate hypothesis generation and model development online through Codex (https://codex.flywire.ai).

Limitations

The availability of neurotransmitter predictions[4], for at least ach, GABA, glut, da, ser and oct, greatly enhanced our ability to interpret the circuit motifs we identified in the connectome. However, although neurotransmitter predictions are 94% accurate when compared to a set of ground truth neurons, there are cases in which the predicted neurotransmitter does not align with the known transmitter. Here we manually corrected the Kenyon cells to be cholinergic (Methods). There may exist other populations that are also systematically misidentified, but which currently lack ground truth neurotransmitter information. While we assume that neurons obey Dale's law (each releasing only one neurotransmitter), there are several known examples of co-transmission in *Drosophila*[54]. It remains unclear how widespread neurotransmitter co-transmission is. Note also that the synaptic connectome does not provide a complete picture of information flow in the brain. We do not currently have a map of gap junction synapses in the fly or of the extent of neuropeptide signalling[55].

Some analyses, particularly those dependent on network topology, are sensitive to our choice of synapse threshold. While connectomes in *C. elegans* have been proofread to the level of individual synapses[21,27,35], this is not feasible in larger datasets[2,3,31]. We therefore rely on automated synapse detection algorithms with a non-negligible error rate[28]. Not all synapses are successfully attached to neurons, and this completion rate varies across brain regions[3]. Although some of these low-synapse-number connections may be spurious, it is likely that some weak connections persist across individuals, as has been found in *C. elegans*[27]. Our companion paper on the fly optic lobe has found that a lower synapse threshold is important for correct cell type identification in those brain regions, further suggesting the importance of some low-synapse connections[56]. Here we applied a consistent and conservative threshold of five synapses per connection between neurons (analysing 2.6 million connections instead of 14.7 million unthresholded connections; Extended Data Fig. 1b,c and Extended Data Table 2), and we demonstrated that our interpretations of connected components, sparsity, reciprocity and clustering coefficients are not dependent on this threshold. It is likely that the fly brain is more strongly interconnected than the results here indicate.

Local circuit motifs are often inferred to be feedforward or feedback connections. However, it is difficult to place local circuits in the context of global directionality from sensory input to motor output. In shallow networks, the directionality of the wiring diagram is clear. However, the larger the network becomes, the more difficult it becomes to establish directionality. In this study, we used an information flow method to rank neurons relative to various sensory modalities[3]. Ultimately, directionality of information flow in particular circuits, especially those in regions of the brain far from sensory inputs or motor outputs, must be determined experimentally.

Neuropil-specific differences

Different neuropils in *Drosophila* are known to have different functions. Studying network topology at the brain scale enables us to highlight how these functional differences are subserved by network differences. For example, the MBs are known to be a centre for learning and memory in the fly brain[32]. We find in the MB compartments the largest fraction of dopaminergic connections (Extended Data Fig. 5d), the biggest differences between reciprocal and unidirectional neurotransmitter percentages (Fig. 5d,e), and a three-node motif distribution close to that of null models (Fig. 6a,b). This combination of network characteristics is unique across the neuropils of the brain. Another well studied neuropil, the EB, is known to be involved in navigation, supporting persistent activity correlated to heading direction[57–59]. We observed that the EB has one of the highest reciprocity rates (Fig. 5c), one of the highest average connection strengths (Extended Data Fig. 6b) and very large over-representations of highly recurrent three-node motifs (Fig. 6a,b). This represents a unique combination of neuropil-scale network properties. By contrast, while the GNG has a similar three-node motif profile (Fig. 6b and Extended Data Fig. 8a), it has much lower reciprocity (Fig. 5c).

The rich club and anatomical bottlenecks

The anatomy of the fly brain suggests two potential bottlenecks: one between left hemisphere and right hemisphere and one between the central brain and optic lobes. Only 12% of neurons cross hemispheres and 6% of neurons cross between the central brain and optic lobes[3,5]. Despite these bottlenecks, the brain is still robustly interconnected with short path lengths. The large rich-club regime in the fly brain may explain these short path lengths. When compared to the average neuron in the brain, rich-club neurons are more likely to contain synapses in both hemispheres, and are also more likely to connect the optic lobes to the central brain. The broad spatial reach of these rich-club neurons also keeps path lengths short across these bottlenecks. In the human brain, it has been proposed that rich-club hubs in functional connectomes help keep path lengths short. In the fly brain, we see that the rich-club neuron population indeed contributes to short path lengths. Future functional experiments in the fly focusing on the rich-club population may shed light on whether this is the case at the neuron scale.

We also expect that the ascending and descending neurons, which form a bottleneck between the brain and the VNC, will be part of the central nervous system rich club. Many ascending and descending neurons appear to have high degrees when examined within either the brain or the VNC[60]. We await the completion of a complete CNS connectome to verify the rich-club membership of such neurons.

The very-high-degree regime

Our rich-club analysis showed that we enter a rich-club regime above total degree 37. However, as we continue to remove neurons, we find that the population of neurons above degree ~100 are no longer preferentially connected relative to either CFG or NPC random networks (Fig. 1h and Extended Data Fig. 2e–g). On the basis of the SCC and WCC survival curves (Fig. 1g and Extended Data Fig. 2b), these very-high-degree neurons do not form multiple islands of connectivity. Rather, the first connected component persists until nearly all neurons are removed, and the second connected component never approaches the size of the first. The under-representation of connections between the very highest degree neurons is correlated with neurotransmitter composition—most are GABAergic (Extended Data Figs. 2e,f,h and 3d). These very large GABAergic neurons are less likely than chance to form connections with each other.

Comparing connectomes across animals

Comparing network properties across different species has the potential to uncover conserved and divergent properties of brain organization. Similarities in reciprocity and clustering coefficient across animals, which vary substantially in both size and connection density (Table 2), hint at the possibility that some features of circuit architecture may be broadly conserved across nervous systems (or that brains optimize for these features). However, comparisons of graph metrics across networks of different sizes, even for neuron-normalized metrics such as reciprocity and clustering coefficient, must be made with care[36–38]. Moreover, as mentioned previously, differences in proofreading and thresholding will impact topological metrics, which treat all edges as equivalent. Applying the same threshold across datasets does not resolve this, as a given number of synapses may have different biological implications across species. It has also been observed, both in this work and in past studies, that different parts of the brain of the fly differ in their connectivity properties[31,33]. It is likely that the same is true in larger brains, meaning that statistics derived from partial wiring diagrams may not be representative.

In *C. elegans*, there is substantial variability in the connectomes of genetically identical individuals[27]. Comparisons between the FlyWire and hemibrain datasets have already revealed similarities and differences between individual flies, as outlined in our companion paper[5], but more datasets will be needed before we fully understand the variability between individuals in *Drosophila*. Fortunately, more whole-brain connectomes are on the horizon. The network analysis of the fly brain presented should serve as a valuable baseline for comparison, both to the connectomes of other *Drosophila* individuals and to the connectomes of other species, to identify which network features are common and which are species- or individual-specific.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-024-07968-y.

1. Zheng, Z. et al. A complete electron microscopy volume of the brain of adult *Drosophila melanogaster*. *Cell* **174**, 730–743 (2018).
2. Dorkenwald, S. et al. Flywire: online community for whole-brain connectomics. *Nat. Methods* **19**, 119–128 (2022).
3. Dorkenwald, S. et al. Neuronal wiring diagram of an adult brain. *Nature* https://doi.org/10.1038/s41586-024-07558-y (2023).
4. Eckstein, N. et al. Neurotransmitter classification from electron microscopy images at synaptic sites in *Drosophila melanogaster*. *Cell* **187**, 2574–2594 (2024).
5. Schlegel, P. et al. Whole-brain annotation and multi-connectome cell typing of *Drosophila*. *Nature* https://doi.org/10.1038/s41586-024-07686-5 (2024).
6. Sporns, O., Tononi, G. & Kötter, R. The human connectome: a structural description of the human brain. *PLoS Comput. Biol.* **1**, e42 (2005).
7. Varshney, L. R., Chen, B. L., Paniagua, E., Hall, D. H. & Chklovskii, D. B. Structural properties of the *Caenorhabditis elegans* neuronal network. *PLoS Comput. Biol.* **7**, e1001066 (2011).
8. van den Heuvel, M. P. & Sporns, O. Rich-club organization of the human connectome. *J. Neurosci.* **31**, 15775–15786 (2011).
9. Kim, D. J. & Min, B. K. Rich-club in the brain's macrostructure: Insights from graph theoretical analysis. *Comput. Struct. Biotechnol. J.* **18**, 1761–1773 (2020).
10. Udvary, D. et al. The impact of neuron morphology on cortical network architecture. *Cell Rep.* **39**, 110677 (2022).
11. Turner, N. L. et al. Reconstruction of neocortex: organelles, compartments, cells, circuits, and activity. *Cell* **185**, 1082–1100 (2022).
12. Yang, R. et al. Cyclic structure with cellular precision in a vertebrate sensorimotor neural circuit. *Curr. Biol.* **33**, 2340–2349 (2023).
13. Humphries, M. D. & Gurney, K. Network 'small-world-ness': a quantitative method for determining canonical network equivalence. *PLoS ONE* **3**, e0002051 (2008).
14. Colizza, V., Flammini, A., Serrano, M. A. & Vespignani, A. Detecting rich-club ordering in complex networks. *Nat. Phys.* **2**, 110–115 (2006).
15. Li, L. et al. Mapping putative hubs in human, chimpanzee (*Pan troglodytes*) and rhesus macaque (*Macaca mulatta*) connectomes via diffusion tractography. *NeuroImage* **80**, 462 (2013).
16. Elam, J. S. et al. The human connectome project: a retrospective. *NeuroImage* **244**, 118543 (2021).
17. Betzel, R. F., Gu, S., Medaglia, J. D., Pasqualetti, F. & Bassett, D. S. Optimally controlling the human connectome: the role of network topology. *Sci. Rep.* https://doi.org/10.1038/srep30770 (2016).
18. Shih, C. T. et al. Connectomics-based analysis of information flow in the *Drosophila* brain. *Curr. Biol.* **25**, 1249–1258 (2015).
19. Worrell, J. C., Rumschlag, J., Betzel, R. F., Sporns, O. & Misic, B. Optimized connectome architecture for sensory-motor integration. *Netw. Neurosci.* **1**, 415 (2017).
20. Towlson, E. K., V.rtes, P. E., Ahnert, S. E., Schafer, W. R. & Bullmore, E. T. The rich club of the *C. elegans* neuronal connectome. *J. Neurosci.* **33**, 6380–6387 (2013).
21. Cook, S. J. et al. Whole-animal connectomes of both *Caenorhabditis elegans* sexes. *Nature* **571**, 63–71 (2019).
22. Milo, R. et al. Network motifs: simple building blocks of complex networks. *Science* **298**, 824–827 (2002).
23. Sporns, O. & Kötter, R. Motifs in brain networks. *PLOS Biol.* **2**, e369 (2004).
24. Song, S., Sjöström, P. J., Reigl, M., Nelson, S. & Chklovskii, D. B. Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biol.* **3**, e68 (2005).
25. Perin, R., Berger, T. K. & Markram, H. A synaptic organizing principle for cortical neuronal groups. *Proc. Natl Acad. Sci. USA* **108**, 5419–5424 (2011).
26. Winding, M. et al. The connectome of an insect brain. *Science* **379**, eadd9330 (2023).
27. Witvliet, D. et al. Connectomes across development reveal principles of brain maturation. *Nature* **596**, 257–261 (2021).
28. Buhmann, J. et al. Automatic detection of synaptic partners in a whole-brain *Drosophila* electron microscopy data set. *Nat. Methods* **18**, 771–774 (2021).
29. Liu, W. W. & Wilson, R. I. Glutamate is an inhibitory neurotransmitter in the *Drosophila* olfactory system. *Proc. Natl Acad. Sci. USA* **110**, 10294–10299 (2013).
30. Molina-Obando, S. et al. On selectivity in the *Drosophila* visual system is a multisynaptic process involving both glutamatergic and gabaergic inhibition. *eLife* **8**, e49373 (2019).
31. Scheffer, L. K. et al. A connectome and analysis of the adult *Drosophila* central brain. *eLife* **9**, e57443 (2020).
32. Takemura, S. Y. et al. A connectome of a learning and memory center in the adult *Drosophila* brain. *eLife* **6**, e26975 (2017).
33. Horne, J. A. et al. A resource for the *Drosophila* antennal lobe provided by the connectome of glomerulus va1v. *eLife* **7**, e37550 (2018).

34. Rall, W., Shepherd, G. M., Reese, T. S. & Brightman, M. W. Dendrodendritic synaptic pathway for inhibition in the olfactory bulb. *Exp. Neurol.* **14**, 44–56 (1966).
35. White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Philos. Trans. R. Soc. Lond. B* **314**, 1–340 (1986).
36. van Wijk, B. C., Stam, C. J. & Daffertshofer, A. Comparing brain networks of different size and connectivity density using graph theory. *PLoS ONE* **5**, e13701 (2010).
37. Smith, A., Calder, C. A. & Browning, C. R. Empirical reference distributions for networks of different size. *Soc. Netw.* **47**, 24–37 (2016).
38. Wills, P. & Meyer, F. G. Metrics for graph comparison: a practitioner's guide. *PLoS ONE* **15**, e0228728 (2020).
39. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world'networks. *Nature* **393**, 440–442 (1998).
40. Arenz, A., Drews, M. S., Richter, F. G., Ammer, G. & Borst, A. The temporal tuning of the *Drosophila* motion detectors is determined by the dynamics of their input elements. *Curr. Biol.* **27**, 929–944 (2017).
41. Schlegel, P. et al. Information flow, cell types and stereotypy in a full olfactory connectome. *eLife* **10**, e66018 (2021).
42. Liu, X. & Davis, R. L. The gabaergic anterior paired lateral neuron suppresses and is suppressed by olfactory learning. *Nat. Neurosci.* **12**, 53–59 (2008).
43. Shinomiya, K. et al. A common evolutionary origin for the on- and off-edge motion detection pathways of the *Drosophila* visual system. *Front. Neural Circ.* **9**, 152337 (2015).
44. Olsen, S. R. & Wilson, R. I. Lateral presynaptic inhibition mediates gain control in an olfactory circuit. *Nature* **452**, 956–960 (2008).
45. Meier, M. & Borst, A. Extreme compartmentalization in a *Drosophila* amacrine cell. *Curr. Biol.* **29**, 1545–1550 (2019).
46. Schenk, J. E. & Gaudry, Q. Nonspiking interneurons in the *Drosophila* antennal lobe exhibit spatially restricted activity. *eNeuro* **10**, ENEURO.0109-22.2022 (2023).
47. Baker, C. A. et al. Neural network organization for courtship song feature detection in *Drosophila*. *Curr. Biol.* **32**, 3317–3333 (2022).
48. Pacheco, D. A., Thiberge, S. Y., Pnevmatikakis, E. & Murthy, M. Auditory activity is diverse and widespread throughout the central brain of *Drosophila*. *Nat. Neurosci.* **24**, 93–104 (2021).
49. Turner, M. H., Mann, K. & Clandinin, T. R. The connectome predicts resting-state functional connectivity across the *Drosophila* brain. *Curr. Biol.* **31**, 2386–2394 (2021).
50. Brezovec, B. E. et al. Mapping the neural dynamics of locomotion across the *Drosophila* brain. *Curr. Biol.* **34**, 710–726.e4 (2024).
51. Lin, A. et al. Imaging whole-brain activity to understand behaviour. *Nat. Rev. Phys.* https://doi.org/10.1038/s42254-022-00430-w (2022).
52. Musall, S., Kaufman, M. T., Juavinett, A. L., Gluf, S. & Churchland, A. K. Single-trial neural dynamics are dominated by richly varied movements. *Nat. Neurosci.* **22**, 1677–1686 (2019).
53. West, S. L. et al. Wide-field calcium imaging of dynamic cortical networks during locomotion. *Cereb. Cortex* **32**, 2668–2687 (2022).
54. Sherer, L. M. et al. Octopamine neuron dependent aggression requires dvglut from dual-transmitting neurons. *PLoS Genet.* **16**, e1008609 (2020).
55. Reinhard, N. et al. Synaptic connectome of the *Drosophila* circadian clock. Preprint at *bioRxiv* https://doi.org/10.1101/2023.09.11.557222 (2023).
56. Matsliah, A. et al. Neuronal parts list and wiring diagram for a visual system. *Nature* https://doi.org/10.1038/s41586-024-07981-1 (2024).
57. Seelig, J. D. & Jayaraman, V. Neural dynamics for landmark orientation and angular path integration. *Nature* **521**, 186–191 (2015).
58. Green, J. et al. A neural circuit architecture for angular integration in *Drosophila*. *Nature* **546**, 101–106 (2017).
59. Hulse, B. K. et al. A connectome of the *Drosophila* central complex reveals network motifs suitable for flexible navigation and context-dependent action selection. *eLife* **10**, e66039 (2021).
60. Cheong, H. S. J. et al. Transforming descending input into behavior: the organization of premotor circuits in the *Drosophila* male adult nerve cord connectome. *eLife* **13**, RP96084 (2024).

Methods

Dataset

The connectome reconstructed by the FlyWire Consortium is that of a 7-day-old adult female *D. melanogaster* from a [iso] w1118 × [iso] Canton-S G1 cross[1]. The EM images were aligned and neurons were automatically reconstructed using deep learning and computer vision methods, then proofread by the community[2,3]. Neuron cell types and community labels were also attached to these data[5,56]. The brain is divided into 78 distinct anatomical brain regions, or neuropils[61].

All analyses presented in this paper were performed on the v630 snapshot of the dataset. The v630 snapshot contains 127,978 neurons and 2,613,129 thresholded connections. The central brain of the fly was fully proofread, with the optic lobes around 80% complete. Most of the neurons missing from the v630 snapshot were photoreceptors, and we do not expect that the addition of these neurons would significantly alter our whole-brain network results. At the time of publication, the most up-to-date version of the dataset is the v783 snapshot, containing 139,255 neurons, 2,701,601 thresholded connections and completed optic lobes. Both data snapshots are available at Codex (https://codex.flywire.ai).

Synaptic connections and thresholding

Synapses were detected algorithmically[28,62], with each synapse receiving a confidence score. We then removed synapses if either the pre- or postsynaptic location of the synapse was not assigned to a segment or if the synapse had a confidence score of less than 50. We then set a threshold of five synapses per connection between neurons for most of our analyses to reduce the impact of spurious connections. This threshold is also consistent across our companion papers on the whole-brain *Drosophila* connectome[3,5]. We used a threshold because manual proofreading of the dataset did not extend to individual synapses[3]. Thresholding connections by synapse number was previously implemented in the hemibrain connectome, with similar rationale[31]. This is a conservative threshold and is likely to result in an undercounting of true connections. We assessed key statistics as a function the threshold to ensure that our qualitative observations hold over a range of threshold choices (Extended Data Fig. 1b,c).

Neurotransmitter assignments

The neurotransmitter at each synapse was predicted directly from the EM images using a trained convolutional neural network with a per-synapse accuracy of 87% (refs. 3,4). The algorithm returns a 1 × 6 probability vector containing the odds that a given synapse is each of the six primary neurotransmitters in *Drosophila*: ach, GABA, glut, da, oct or ser. We then averaged these probabilities across all of a neuron's outgoing synapses, under the assumption that each neuron expresses a single outgoing neurotransmitter, to obtain a 1 × 6 probability vector representing the odds that a given neuron expresses a given neurotransmitter. We then assigned the highest-probability neurotransmitter as the putative neurotransmitter for that neuron. The per-neuron accuracy is 94%. In cases in which the highest probability $p_1 < 0.2$ and the difference between the top two probabilities $p_1 - p_2 < 0.1$, we classified the neuron as having an uncertain neurotransmitter. In the approximately 1,600 Kenyon cells, for which the neurotransmitter of a neuron is known to be ach but the algorithm often returned erroneous predictions, the neurotransmitter prediction associated with that neuron was overwritten by the known neurotransmitter.

Cell classifications and labels

84% of neurons are intrinsic to the brain, meaning that their projections are fully contained in the brain volume[3]. Central brain neurons are fully contained in the central brain, while optic lobe intrinsic neurons are fully contained in the optic lobes. Visual projection neurons have inputs in the optic lobes and outputs in the central brain. Visual centrifugal neurons have inputs in the central brain and outputs in the optic lobe. Sensory neurons are those that are entering the brain from the periphery, and are divided into classes by modality. Refer to our companion paper for more details on the classification criteria[5]. We also used annotation labels contributed by the FlyWire community[3].

Connected components

SCCs are defined as subnetworks in which all neurons are mutually reachable through directed pathways[63]. WCCs are a relaxed criterion in which all neurons are mutually reachable, ignoring the directionality of connections.

Degree definitions

For a given neuron i, the in-degree d_i^+ is the number of incoming synaptic partners the neuron has and the out-degree d_i^- is the number of outgoing synaptic partners the neuron has. The total degree of a neuron i is the sum of in-degree and out-degree:

$$d_i^{\mathrm{tot}} := d_i^+ + d_i^-.$$

The reciprocal degree d_i^{rec} is the number of partners that a given neuron forms reciprocal connections with. As each reciprocal connection consists of two edges, we can determine the fraction of reciprocal inputs and outputs as d_i^{rec}/d_i^+ and d_i^{rec}/d_i^-, respectively (Extended Data Fig. 3c).

Definitions of connectivity metrics

Given the observed wiring diagram as a simple (no self-edges) directed graph $G(V,E)$, the connection probability or density is the probability that, given an ordered pair of neurons α and β, a directed connection exists from one to the other:

$$p^{\mathrm{conn}} := P[\alpha \to \beta] = \frac{|E|}{|V|\ (|V|-1)}.$$

The reciprocity is the probability that, given a pair of neurons which are connected α to β, there exists a returning β to α connection:

$$p^{\mathrm{rec}} := P[\beta \to \alpha | \alpha \to \beta].$$

The (global) clustering coefficient is the probability that for three neurons α, β and γ, given that neurons α and β are connected and neurons α and γ are connected (regardless of directionality), neurons β and γ are connected:

$$C^{\Delta} := P[\beta - \gamma | \alpha - \beta \wedge \alpha - \gamma].$$

We computed these metrics both across the whole-brain and within-brain-region (neuropil) subnetworks. We also systematically quantified the occurrence of distinct directed three-node motifs within the network, ensuring that duplicates are eliminated: any subgraph involving three unique nodes is counted only once in our analysis. To compute the expected prevalence of specific neurotransmitter motifs (Figs. 2c and 3d,e), we multiplied the relevant neurotransmitter probabilities for the motif of interest, under the assumption the neurons connect independent of neurotransmitter. We then compared this expectation to the true frequency of motifs with these neurotransmitter combinations.

ER and CFG null models

We probed different statistics of the wiring diagram $G(V,E)$ by comparing them with the statistics of various null models. The simplest null model that we used was a directed version of the ER model $\mathcal{G}(V,p)$, where all edges are drawn independently at random, and the connection probability p is set such that the expected number of edges in the

ER model equals that observed in the wiring diagram[64]. For any nodes $i,j \in V$, the connection probability is constant:

$$P[i \to j] = p = \frac{|E|}{|V|\ (|V|-1)}.$$

As reciprocal edges in the wiring diagram are over-represented when compared to a standard ER model, we adopted a generalized ER model, which preserves the expected number of reciprocal edges. The generalized ER model $\mathcal{G}(V, p^{\text{uni}}, p^{\text{bi}})$ has two parameters, unidirectional connection probability p^{uni} and bidirectional connection probability p^{bi}, both of which are set to match the wiring diagram. To do this, we defined the sets of unidirectional and bidirectional edges as:

$$E^{\text{uni}} := \{(i,j) | (i,j) \in E \wedge (j,i) \notin E\},$$
$$E^{\text{bi}} := \{(i,j) | (i,j) \in E \wedge (j,i) \in E\}.$$

For any nodes i and j:

$$P[i \overset{\rightarrow}{\nleftrightarrow} j] = P[i \overset{\leftarrow}{\nrightarrow} j] = p^{\text{uni}} = \frac{|E^{\text{uni}}|}{|V|\ (|V|-1)},$$
$$P[i \overset{\leftarrow}{\rightarrow} j] = p^{\text{bi}} = \frac{|E^{\text{bi}}|}{|V|\ (|V|-1)},$$
$$[i \overset{\leftarrow}{\nrightarrow} j] = 1 - 2p^{\text{uni}} - p^{\text{bi}}.$$

All edges between unordered node pairs were drawn independently and at random.

Consistent with previous work[12,49], we also used a directed configuration model (CFG), $\mathcal{G}(V, \{d_i^+\}, \{d_i^-\})$, which preserves degree sequences during random rewiring. We sampled 1,000 random graphs uniformly from a configuration space of graphs with the same degree sequences as the observed graph by applying the switch-and-hold algorithm[65], where we randomly select two edges in each iteration and swap their target endpoints (switch), or else keep them unchanged (hold), under the conditions that doing so does not introduce self-loops or multiple edges. With these conditions, this CFG model is mathematically equivalent to the Maslov–Sneppen edge-swapping null model[66–68].

Pairwise distances between neurons

To determine the connection probability distribution as a function of distance between neurons, we first had to distil the available spatial information into a handful of points. This was the only practical way to enable distance comparisons between all neurons—a total of 14 billion pairs. For each neuron, we defined two coordinates based on the location of their incoming and outgoing synapses. We computed the average 3D position of all of the neuron's incoming synapses to approximate the position of the neuron's dendritic arbour, and did the same to approximate the position of the neuron's axonal arbour. We then computed for all neuron pairs the pairwise distances between the axonal arbour of neuron A and the dendritic arbour of neuron B. Binning by distance and comparing the number of true connections to the number of neuron pairs allowed us to compute connection probability as a function of space (Extended Data Fig. 1d).

NND model

Informed by the distribution of connection probability as a function of distance, we constructed a NND model with two zones of probability—a 'close' zone (0 to 50 μm) where connections are possible with a relatively high probability ($p_{\text{close}} = 0.00418$) and a 'distant' zone (more than 50 μm) where connections occur with lower probability ($p_{\text{distant}} = 0.00418$) (Extended Data Fig. 1e). The probabilities in these two zones were derived from the real network. Then, for every neuron pair (around 14 billion pairs), we generate a random number drawn uniformly from between 0 and 1. The distance between the two neuronal arbours sets the probability of forming an edge between the pairs, p_{close} or p_{distant}. If the random number is below the probability threshold, a connection is formed between these two neurons in the model.

NPC model

To provide a more tractable spatial null model while preserving degree sequences, we developed the NPC model. This model is a degree-corrected stochastic block model (DC-SBM)[69]. We assigned each neuron to one of the 78 'neuropil blocks' based on the neuropil in which the neuron has the most outgoing synapses. During random rewiring, the inter- and intra-neuropil connection probabilities are preserved. Moreover, like in the CFG model, we keep the degree sequences unchanged during randomization and prohibit self-loops and multiple edges. The interneuropil connection densities implicitly contain mesoscale spatial information. These constraints also mean that the total number of internal edges in each neuropil remains the same after reshuffling.

Spectral analysis

Given a strongly connected graph $G(V,E)$ and its 0–1 adjacency matrix $A \in \mathbb{R}_{\geq 0}^{n \times n}$, where A_{ij} corresponds to a connection from neuron i to neuron j, one can construct an irreducible Markov chain on the strongly connected graph with a transition matrix $P_{ij} := A_{ij} / \sum_k A_{kj}$ giving the transition probability from j to i. This corresponds to a random walk, where the transition probability from neuron α to neuron β is $p_{\alpha \to \beta} = \delta_{\alpha \to \beta} / d_\alpha^-$, where d_α^- is the out-degree of neuron α, and $\delta_{\alpha \to \beta} \in \{0,1\}$ indicates the existence of a connection.

The Perron–Frobenius theorem guarantees that P has a unique positive right eigenvector π with eigenvalue 1, and therefore that π is the stationary distribution of the Markov chain. We constructed such a transition matrix for the connectome and determined the eigenvector π.

We also defined a 'reverse' Markov chain with a transition matrix $P_{ij}^{\text{rev}} := A_{ji} / \sum_k A_{jk}$ giving the transition probability from j to i. P^{rev} also has a unique positive right eigenvector π^{rev} with eigenvalue 1. This corresponds to a reverse random walk with the transition probability from neuron α to neuron β set as $p_{\alpha \to \beta}^{\text{rev}} = \delta_{\alpha \to \beta} / d_\alpha^+$, where d_α^+ is the in-degree of neuron α. Extended Data Fig. 1f,g shows the stationary distribution of forward and reversed Markov chains, respectively.

The normalized symmetric Laplacian of the Markov chain P is:

$$\mathcal{L} = I - \frac{1}{2}\left(\Pi^{\frac{1}{2}} P_\alpha \Pi^{-\frac{1}{2}} + \Pi^{-\frac{1}{2}} P^\top \Pi^{\frac{1}{2}}\right),$$

where $\Pi := \text{Diag}(\pi)$ and I is the identity matrix. We similarly defined $\mathcal{L}^{\text{rev}}$ for the reverse Markov chain. The eigen-spectra of $\mathcal{L}$ and $\mathcal{L}^{\text{rev}}$ are shown in Extended Data Fig. 1f,g, respectively. The gaps between eigenvalues indicate the conductance properties of the graph.

Finding rich-club neurons

We used the standard rich-club formulation to quantify the rich-club effect[14]. The rich-club coefficient $\Phi(d)$ at a given degree value d, with all nodes with degree < d pruned, is the number of existing connections in the surviving subnetwork divided by the total possible connections in the surviving subnetwork:

$$\Phi(d) = \frac{M_d}{N_d(N_d - 1)},$$

where N_d is the number of neurons in the network with degree ≥ d and M_d is the number of connections between such neurons. To control for the fact that high-degree nodes have a higher probability of connecting to each other by chance, we normalized the rich club coefficient to the average rich club value of 100 samples from a CFG null model (Fig. 1h and Extended Data Fig. 2e,f):

$$\Phi_{\mathrm{norm}}(d) = \frac{\Phi(d)}{\langle \Phi_{\mathrm{CFG}}(d) \rangle}.$$

The standard method of determining the rich-club threshold is to look for values of k for which $\Phi_{\mathrm{norm}}(d) > 1 + n\sigma$, where σ is the s.d. of $\Phi_{\mathrm{CFG}}(d)$ and n is chosen arbitrarily[20]. However, as the s.d. from our samples is extremely small (approaches 0) near the bump in relative rich-club coefficient, we chose instead to define the onset threshold of the rich club as $\Phi_{\mathrm{norm}}(d) > 1.01$ (1% denser than the CFG random networks).

We computed the rich-club coefficient in three different ways, by sweeping by total degree (Fig. 1h and Extended Data Fig. 2e), in-degree and out-degree (Extended Data Fig. 2f), progressively moving from small to large values. As we observed, when the total degrees of the remaining nodes surpass 37, the network becomes denser compared with randomized networks. The peak occurs at total degree = 75. 38.9% of neurons have degree ≥ 75, and they are 2.76% more dense than predicted by the CFG model. Once the minimal total degree reaches 93, the network becomes as sparse as the randomized counterpart. We classified neurons with total degrees above 37 as rich-club neurons because they exhibit denser interconnections when considered as a subnetwork. In terms of in-degree, the range for denser-than-random connectivity is between 10 and 54. Considering out-degree alone did not reveal any specific onset or offset threshold for rich-club behaviour, as the subnetwork always remains sparser than random. The rich-club coefficient relative to the NPC model (Extended Data Fig. 2g) was computed similarly, with 100 samples of the null model:

$$\Phi_{\mathrm{norm}}(d) = \frac{\Phi(d)}{\langle \Phi_{\mathrm{NPC}}(d) \rangle}.$$

Small-worldness coefficient

We quantified the small-worldness of the connectome by comparing it to an ER graph. The average undirected path length in the ER graph, denoted as ℓ_{rand}, is estimated to be 3.57 hops, similar to the observed average path length in the fly brain's WCC ($\ell_{\mathrm{obs}} = 3.91$). The clustering coefficient ($C^{\Delta}_{\mathrm{rand}}$) of the ER graph is only 0.0003, much smaller than the observed clustering coefficient ($C^{\Delta}_{\mathrm{obs}} = 0.0463$) (Table 2). The small-worldness coefficient of the whole-brain fly connectome is:

$$S^{\Delta} = \frac{C^{\Delta}_{\mathrm{obs}} / C^{\Delta}_{\mathrm{rand}}}{\ell_{\mathrm{obs}} / \ell_{\mathrm{rand}}} = 141, \quad (1)$$

Definitions of highly connected neurons

To identify broadcaster neurons, we filtered the intrinsic rich-club neurons ($d^{\mathrm{tot}} > 37$) for those that had an out-degree that was at least five times higher than their in-degree:

$$d^{-} \geq 5 \times d^{+}.$$

Similarly, we identified integrator neurons by filtering the intrinsic rich-club neurons for those that had an in-degree that was at least five times higher than their out-degree:

$$d^{+} \geq 5 \times d^{-}.$$

Rich-club neurons that did not fall into either category were defined as 'large balanced' neurons. This analysis was limited to intrinsic neurons—those that have all of their inputs and outputs within the brain—to avoid spurious identification of afferent or efferent neurons as broadcasters or integrators.

When identifying large recurrent neuropil-specific neurons (Fig. 5g) we applied the following criteria. First, the neurons were intrinsic and met the rich-club criteria. Second, at least 50% of the neuron's incoming connections were contained within the subnetwork of a single neuropil. Third, at least 50% of the neuron's outgoing connections were contained within the same neuropil.

Neuron ranking

We used a probabilistic connectome flow model previously published[41] to determine the ranking of neurons relative to various sensory neuron populations[3,41]. This method ignores the sign of connections. Starting from a set of user-defined seed neurons, the model traverses the wiring diagram probabilistically: in each iteration, the chance that a neuron is added to the traversed set increases linearly with the fractions of synapses that it is receiving from neurons already in the traversed set. When this likelihood reaches 30%, the neuron is guaranteed to be added to the traversed set. The process is then repeated until the entire network graph has been traversed. The iteration in which a neuron was added to the traversed set corresponds to the distance in hops it was from the seed neurons. For each set of seed neurons, the model was run 10,000 times. The distance used to determine the rank of any given neuron was the average iteration in which it was added to the traversed set.

We ran this model using the following subsets of sensory neurons as seeds: olfactory receptor neurons, gustatory receptor neurons, mechanosensory Johnston's Organ (auditory) neurons, head and neck bristle mechanosensory neurons, thermosensory neurons, hygrosensory neurons, visual projection neurons, visual photoreceptors, ocellar photoreceptors and ascending neurons. We also ran the model using the set of all of the input neurons as seed neurons. All neurons in the brain were then ranked by their traversal distance from each set of starting neurons, and this ranking was normalized to return a percentile rank. We rank from the visual projection neurons as a proxy for visual sensory inputs to the central brain, but note that this is not a true sensory population.

Information flow between neuropils

To determine the contributions that a single neuron makes to information flow between neuropils, we first applied two simplifying assumptions: first, that information flow through the neuron can be approximated by the fraction of synapses in a given region; and, second, that inputs and outputs can be treated independently. Using these two assumptions, we constructed a matrix representing the projections of a single neuron between neuropils. The fractional inputs of a given neuron are a $1 \times N$ vector containing the fraction of incoming synapses the neuron has in each of the N neuropils, and the fractional outputs are a similar vector containing the fraction of outgoing synapses in each of the N neuropils. We multiplied these vectors against each other to generate the $N \times N$ matrix of the neuron's fractional weights, with a total weight of one. Summing these matrices across all neurons produced a matrix of neuropil-to-neuropil connectivity, or projectome (see figure 4 of ref. 3).

From the neuropil-to-neuropil connectivity matrix, we determined the total weight of internal connections—those within a given neuropil—by identifying the neurons which contribute to the diagonal of the matrix. We likewise determined the weight external connections—either incoming to the neuropil or outgoing from the neuropil—by looking at the off-diagonals. These data were used to construct the analyses in Extended Data Fig. 6a–c.

Identifying neuropil subnetworks

Most of the neurons in the *Drosophila* brain have soma at the surface of the brain. Therefore, they cannot be associated to neuropils (brain regions) based on their soma locations. Synapses, however, can be associated with neuropils. To perform motif analyses at the level of individual neuropils, we identified neuropil subnetworks based on the connections made by the synapses contained within each neuropil volume. All connections within the neuropil of interest are taken as

edges of this subnetwork, and all neurons connected to these edges are included (Fig. 5b). The number of neurons associated with each neuropil subnetwork is plotted in Extended Data Fig. 6d. Note that if two neurons both in a given neuropil subnetwork share a connection that occurs in a different neuropil, that connection is not included as an edge in the given subnetwork.

Identifying interneuropil reciprocal pairs

We constructed a map of reciprocal connections between neuropils in the form of a triangular matrix with the neuropils as axes. For clarity, here we will refer to a unidirectional connection as an edge. A reciprocal connection contains two opposing edges. While some edges are composed of synapses in multiple neuropils, the majority of edges are composed of synapses in a single neuropil after thresholding. We therefore applied a winner-take-all approach to assigning edges to neuropils. Given two reciprocally connected neurons *X* and *Y*, let us call the edge from *X* to *Y* edge 1, and the edge from *Y* to *X* edge 2. If the synapses that form edge 1 are in neuropil A, and the synapses that form edge 2 are in neuropil B, then we assign this reciprocal pair to the neuropil A to neuropil B square of the matrix. This was done for all reciprocal pairs, with each reciprocal pair is counted as 1 in the matrix. Note that this means that a given neuron can be represented multiple times if it has multiple reciprocal partners.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The complete connectome dataset is available online at Codex (Connectome Data Explorer; https://codex.flywire.ai). Neuron annotations, neurotransmitter information and compact data downloads are available through Codex, along with neuron lists generated in this work, including neurons participating in two-node and selected three-node motifs, rich-club neurons, broadcaster and integrator neurons, and neuropil-specific reciprocal neurons. The connectome data are available for download at Zenodo[70] (https://doi.org/10.5281/zenodo.10676865).

Code availability

The analyses presented in this paper were performed in Python with the numpy and graph-tool[71] packages, and in MATLAB (standard toolboxes). Software written for this publication is available at GitHub (https://github.com/murthylab/flywire-network-analysis) and at Zenodo[72] (https://zenodo.org/records/12572930). Some 3D renders were generated in Cinema4D.

61. Ito, K. et al. A systematic nomenclature for the insect brain. *Neuron* **81**, 755–765 (2014).
62. Heinrich, L. Funke, J., Pape, C., Nunez-Iglesias, J. & Saalfeld, S. Synaptic cleft segmentation in non-isotropic volume electron microscopy of the complete *Drosophila* brain. In *Proc. 21st International Conference on Medical Image Computing and Computer Assisted Intervention—MICCAI 2018* (eds Frangi, A. F. et al.) 317–325 (Springer, 2018).
63. Tarjan, R. Depth-first search and linear graph algorithms. *SIAM J. Comput.* **1**, 146–160 (1972).
64. Gilbert, E. N. Random graphs. *Ann. Math. Stat.* **30**, 1141–1144 (1959).
65. Artzy-Randrup, Y. & Stone, L. Generating uniformly distributed random networks. *Phys. Rev. E* **72**, 056708 (2005).
66. Maslov, S. & Sneppen, K. Specificity and stability in topology of protein networks. *Science* **296**, 910–913 (2002).
67. Milo, R., Kashtan, N., Itzkovitz, S., Newman, M. E. J. & Alon, U. On the uniform generation of random graphs with prescribed degree sequences. Preprint at arxiv.org/abs/cond-mat/0312028 (2003).
68. Fosdick, B. K., Larremore, D. B., Nishimura, J. & Ugander, J. Configuring random graph models with fixed degree sequences. *SIAM Rev.* **60**, 315–355 (2016).
69. Karrer, B. & Newman, M. E. J. Stochastic block models and community structure in networks. *Phys. Rev. E* **83**, 016107 (2011).
70. Lin, A. et al. Data for 'Network statistics of the whole-brain connectome of *Drosophila*'. *Zenodo* https://doi.org/10.5281/zenodo.10676865 (2024).
71. Peixoto, T. P. The graph-tool Python library. *Figshare* https://doi.org/10.6084/m9.figshare.1164194 (2014).
72. Lin, A. et al. Code for 'Network statistics of the whole-brain connectome of *Drosophila*'. *Zenodo* https://doi.org/10.5281/zenodo.12572929 (2024).

Acknowledgements We thank S. Seung, D. Bock, T. Clandinin and the members of the Murthy, Seung, Jefferis and Clandinin laboratories for their advice on the project and comments on the manuscript; B. Mimica for providing the mouse drawing; the staff at the Princeton Neuroscience Institute for support; and staff at Google for assistance. A.L. was supported by the NSF through the Center for the Physics of Biological Function (PHY-1734030); G.S.X.E.J. by Wellcome Trust Collaborative Awards 203261/Z/16/Z and 220343/Z/20/Z, the Neuronex2 award (MRC_MC_EX_MR/T046279/1) and the MRC (MC-U105188491); and M.M. by NIH BRAIN Initiative grants RF1 MH117815, RF1 MH129268 and U24 NS126935.

Author contributions A.L., R.Y. and S.D. analysed the data. S.D., P.S. and A.M. curated the data and made it available for download. A.M. developed software and data analysis tools. A.M. and A.R.S. built the Codex online platform. S.-c.Y., C.E.M., M.C., K.E. and P.S. trained and managed Flywire proofreaders. A.S.B., N.E., G.S.X.E.J. and J.F. provided neurotransmitter information. C.E.M. and A.R.S. provided Flywire community support and training. A.L., R.Y., S.D. and A.R.S. generated the figure panels. A.L., R.Y. and M.M. wrote the manuscript with feedback from all of the other authors. M.M. supervised the project.

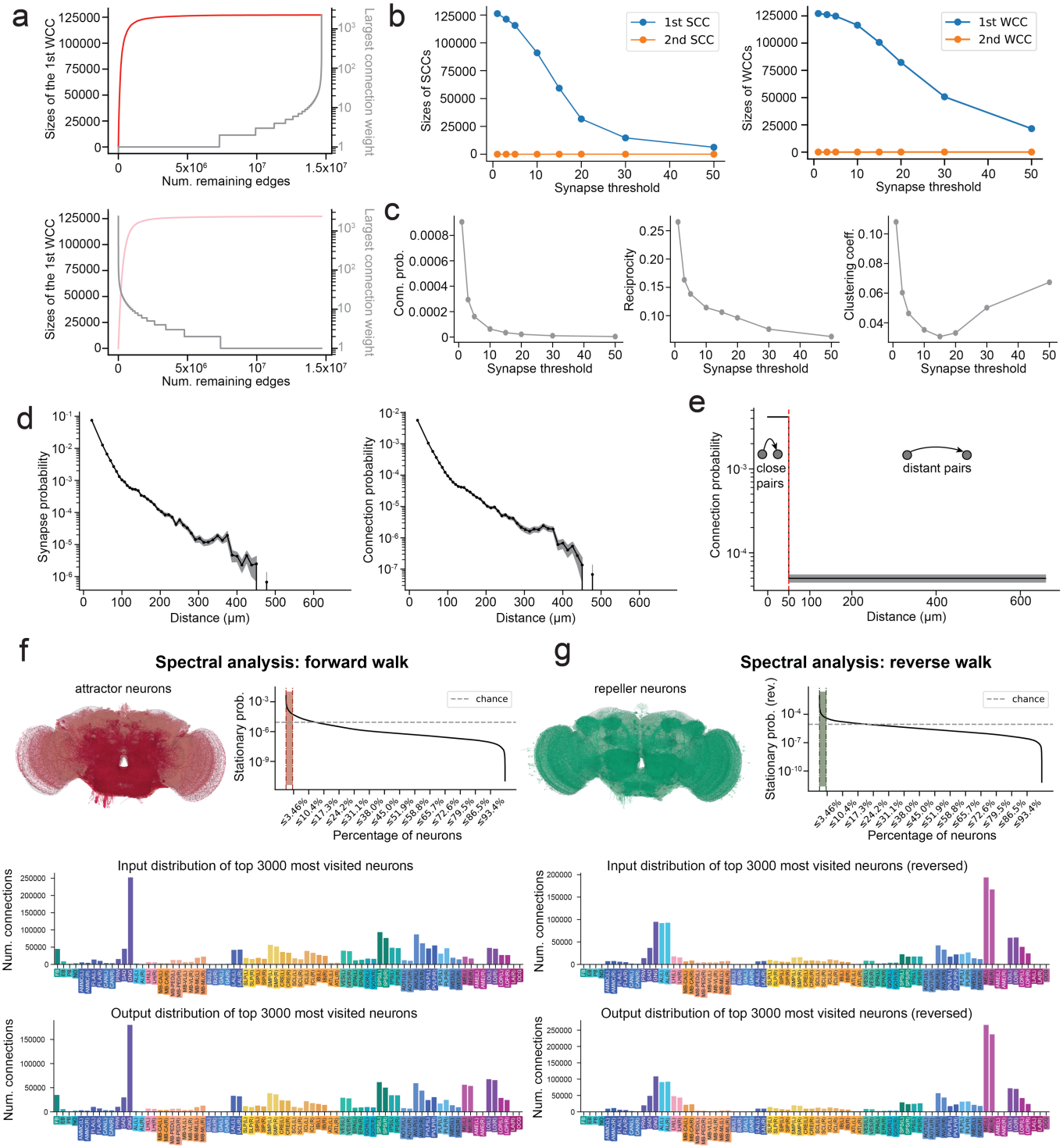
Competing interests The authors declare no competing interests.

Additional information

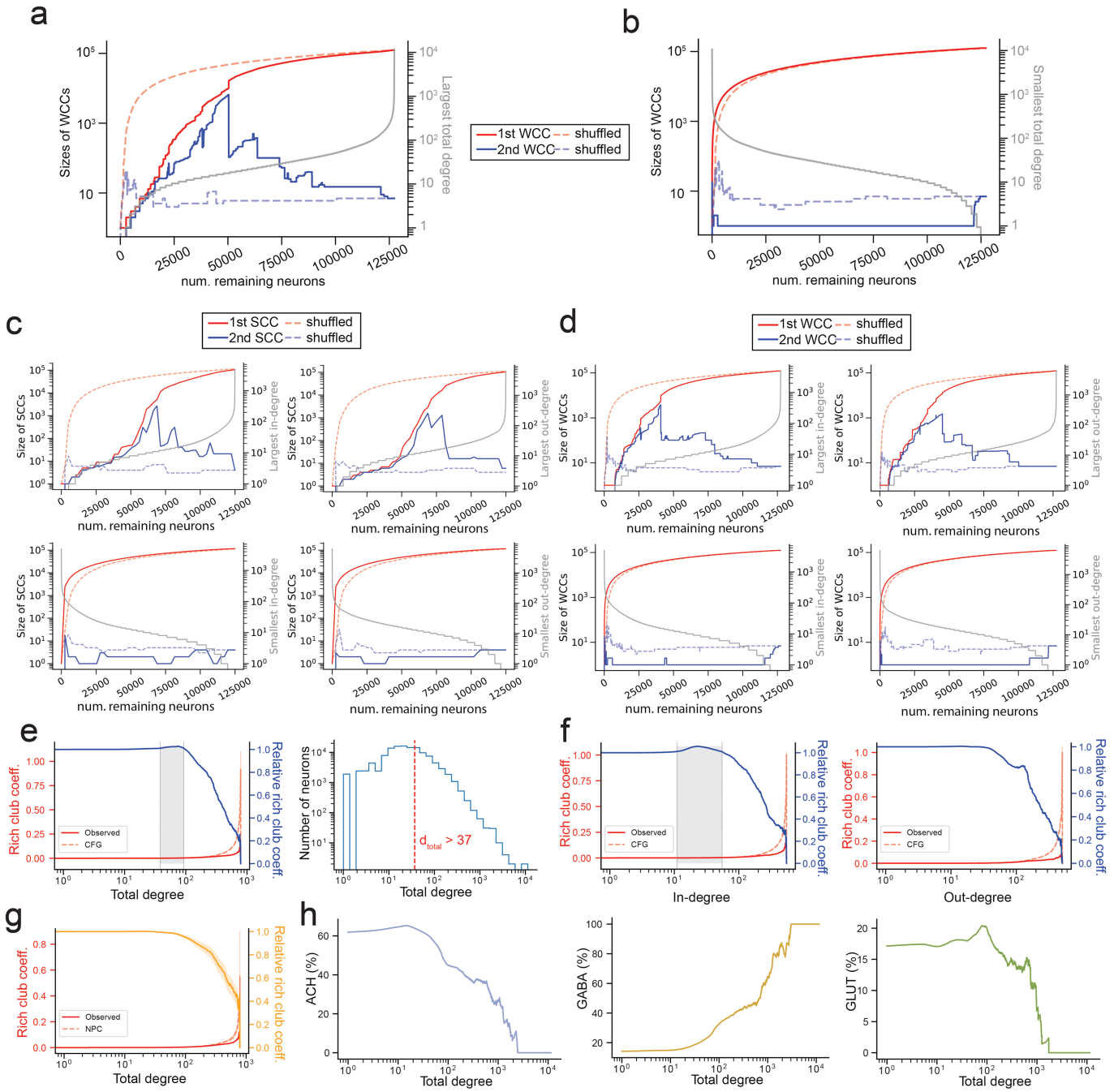
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-024-07968-y.

Correspondence and requests for materials should be addressed to Mala Murthy.

Peer review information *Nature* thanks Martijn van den Heuvel and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

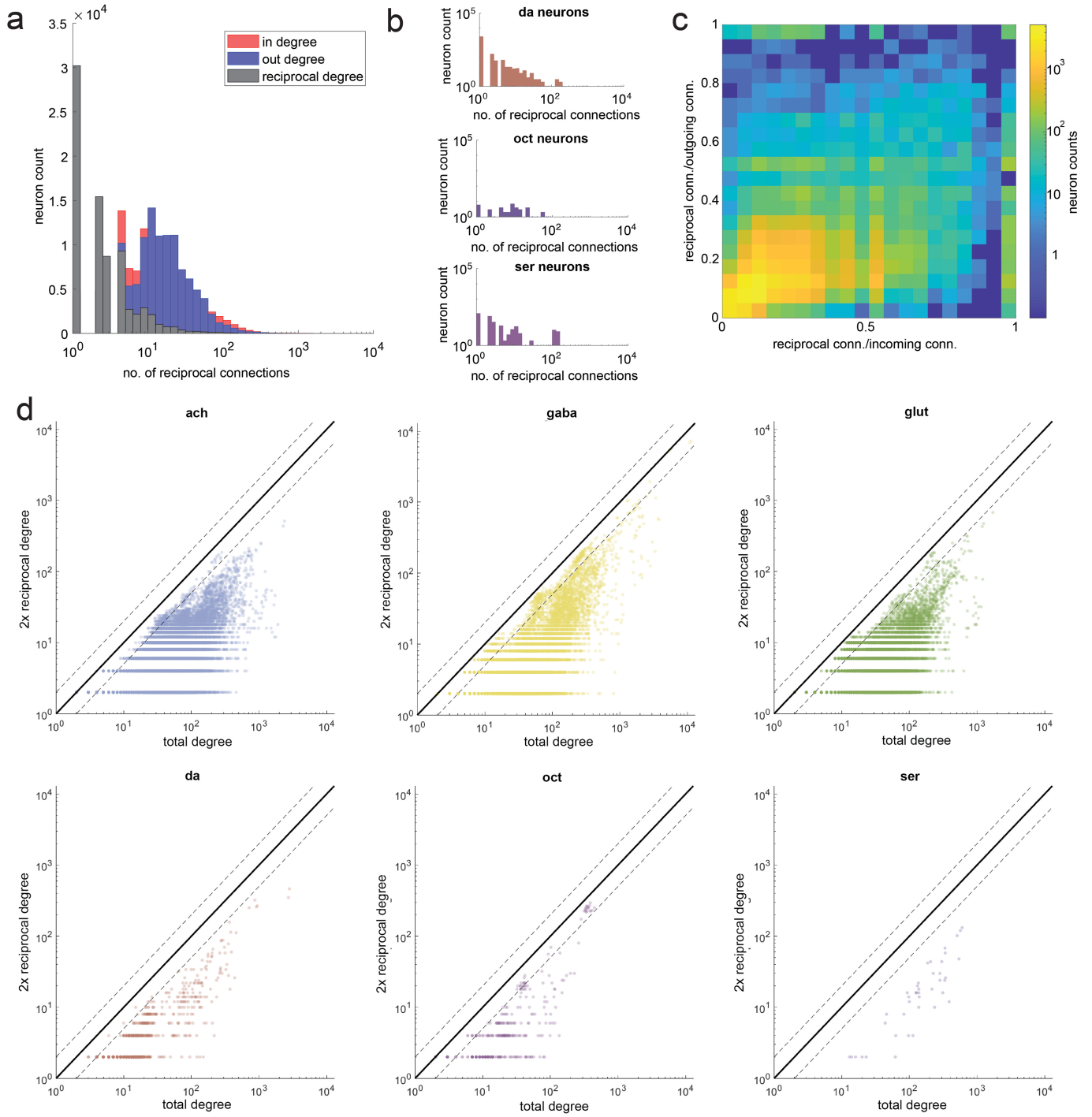


Extended Data Fig. 1 | Thresholding, connections as a function of distance, and spectral analyses. The effects of edge percolation on the size of the largest WCC when (a) large connections are removed first and when (b) small connections are removed first. (c) The sizes of the first two SCCs as a function of the synapse threshold. (d) Synapse probability (left) and connection probability (right) as a function of the average distance between neuronal arbours. Plots are of a drawn from a subsample of 700 million pairs (5% of the total 14 billion pairs). (e) The probability of random connection of the two-zone spatial null model, with one close regime with high connection probability and a distant regime with low connection probability. Spectral analysis of the whole-brain network with (f) forward and (g) reverse walks. In each case, the stationary probability distributions are shown, as well as the distribution of neuropils in which the inputs and outputs of the top 3000 most visited neurons are located. Renders of the top 3% attractor (red) and repeller (green) neurons are also shown. The top 0.3% are rendered in darker colours.

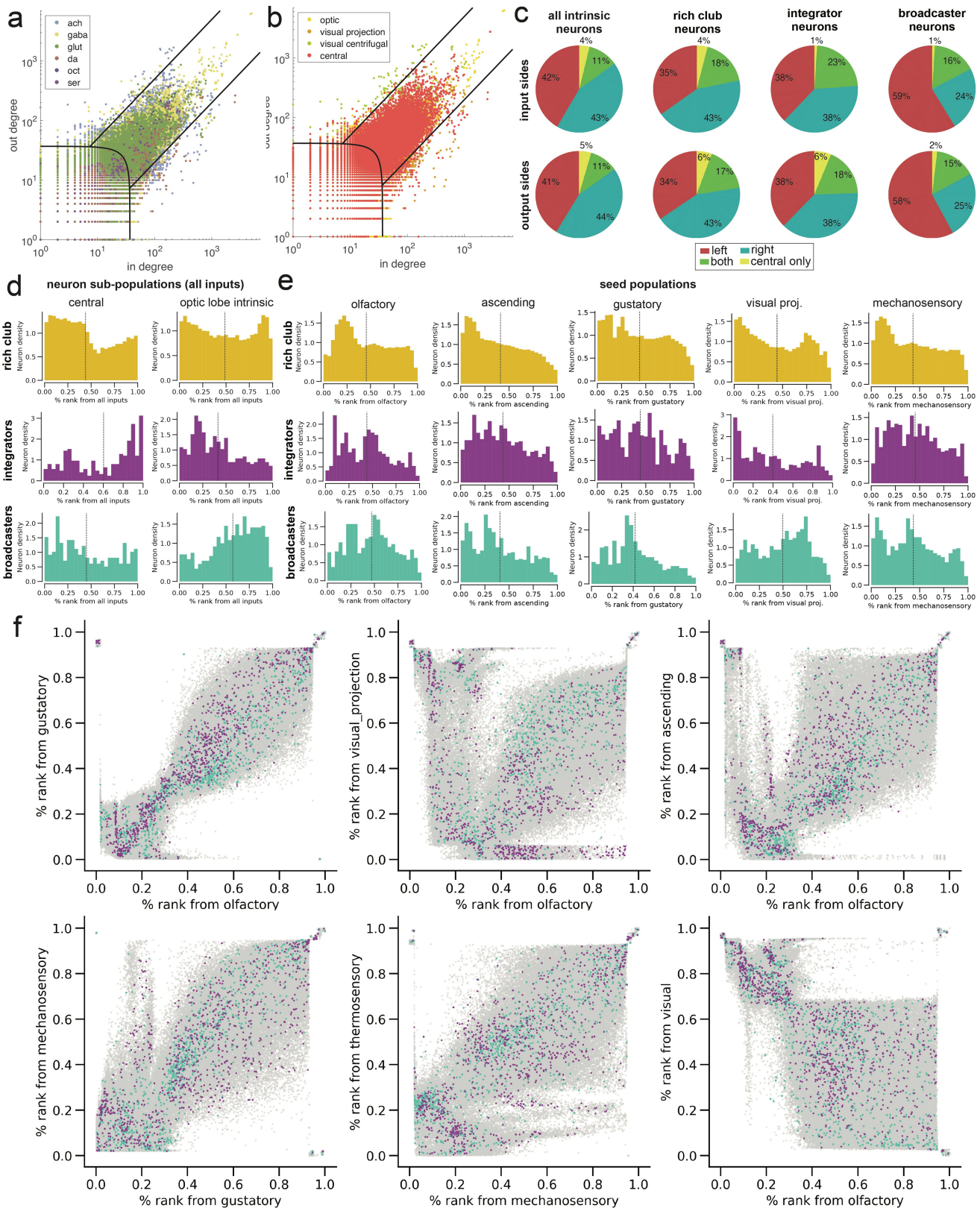


Extended Data Fig. 2 | See next page for caption.

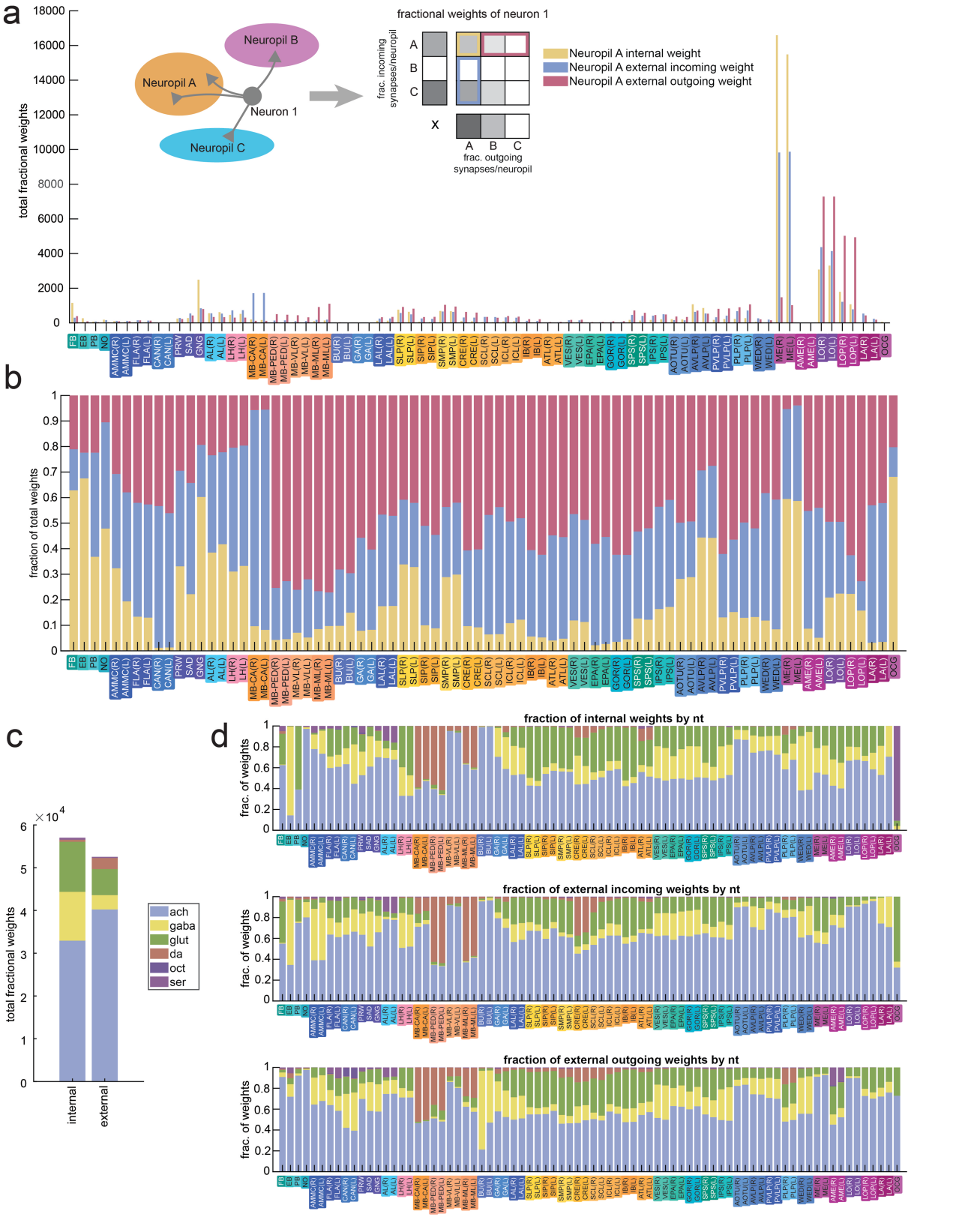
Extended Data Fig. 2 | Additional connected components and rich club analyses. (a) The sizes of the first two weakly connected components (WCCs) as nodes are removed by total degree (1 neuron per step). Removal of neurons starting with those with largest degree results in the brain splitting into two WCCs when neurons of approximately degree 50 start to be removed, a deviation from when neurons are removed in a random order (dotted lines). The largest surviving total degree as a function of the number of remaining nodes is plotted in grey. (b) Removal of neurons starting with those with smallest degree results in a single giant WCC until all neurons are removed. The smallest surviving total degree as a function of the number of remaining nodes is plotted in grey. (c) The sizes of the first two strongly connected components (SCCs) as nodes are removed by in-degree or out-degree (2500 neurons per step). Removal of neurons starting with those with largest in-degree (top left) or largest out-degree (top right) result in the brain splitting into two SCCs when neurons of approximately degree 50 start to be removed, a deviation from when neurons are removed in a random order (dotted lines). Removal of neurons starting with those with smallest in-degree (bottom left) or smallest out-degree (bottom right) results in a single giant SCC until all neurons are removed. (d) The sizes of the first two weakly connected components (WCCs) as nodes are removed by in-degree or out-degree (1 neuron per step). Removal of neurons starting with those with largest in-degree (top left) or largest out-degree (top right) result in the brain similarly splitting into two WCCs when neurons of approximately degree 50 start to be removed, a deviation from when neurons are removed in a random order (dotted lines). Removal of neurons starting with those with smallest in-degree (bottom left) or smallest out-degree (bottom right) results in a single giant WCC until all neurons are removed. (e) The relative rich club coefficient ($\Phi_{norm} = \Phi/\Phi_{CFG}$, blue) as a function of total degree (left). The absolute observed rich-club coefficient (Φ) is plotted in solid red, and the rich-club coefficient from a CFG model (Φ_{CFG}) is plotted in dotted red. To the right is plotted the degree distribution showing the rich club regime cutoff. (f) The rich club coefficient plotted as a function of in-degree (left) and out-degree (right), relative to CFG models. (g) The relative rich club coefficient ($\Phi_{norm(NPC)} = \Phi/\Phi_{NPC}$, orange) as a function of total degree. The absolute observed rich-club coefficient (Φ) is plotted in solid red, and the rich-club coefficient from a NPC model (Φ_{NPC}) is plotted in dotted red. (h) Survival curves for the percentages of ach, GABA, and glut neurons remaining as a function of total degree.



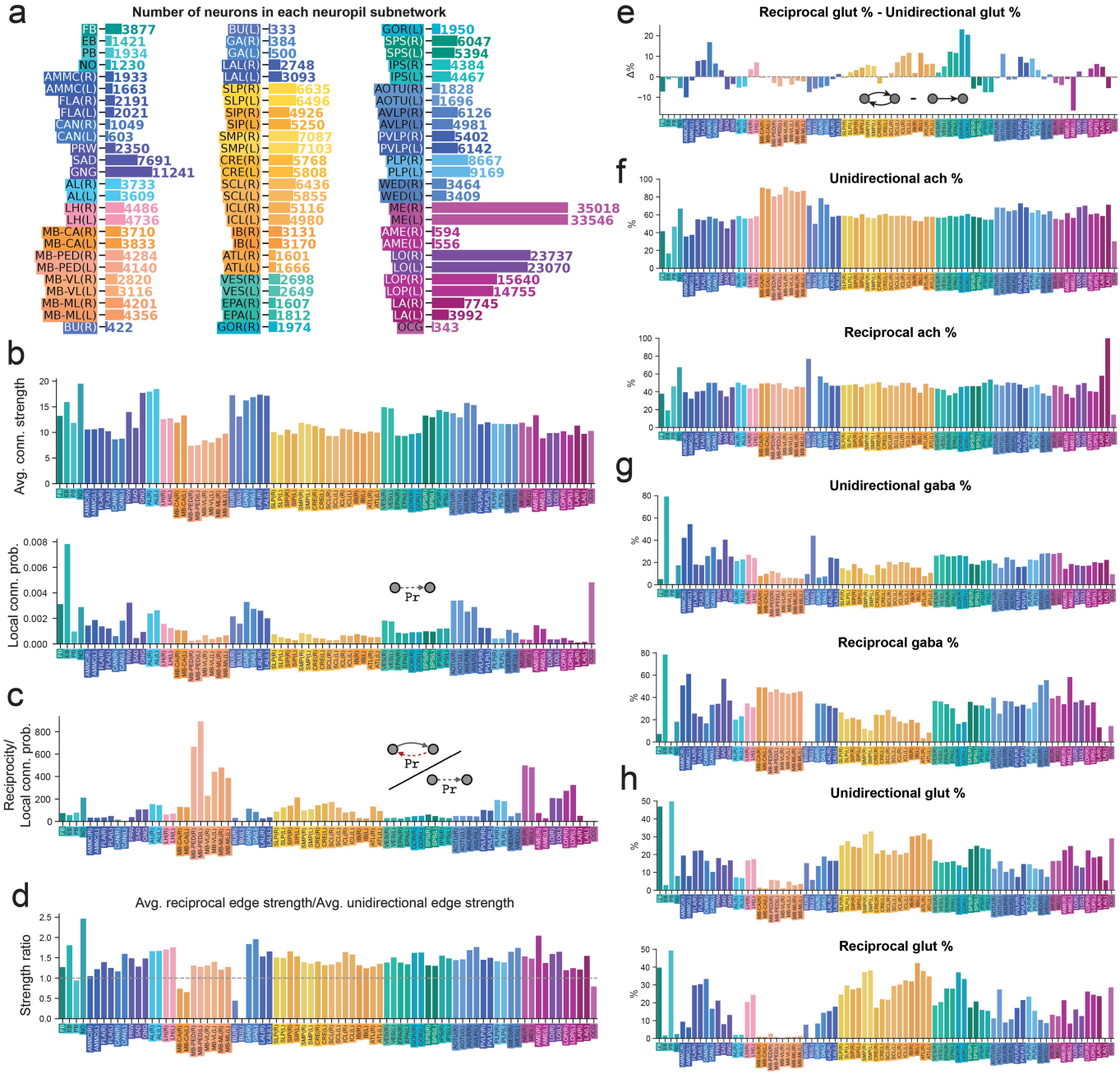
Extended Data Fig. 3 | Reciprocal connectivity and degree. (a) Distribution of reciprocal degree (grey) alongside distributions of in-degree (red) and out-degree (blue). (b) Distributions of reciprocal degree for glut, da, oct, and ser neurons. (c) Heatmap showing the fraction of reciprocal incoming connections versus the fraction of reciprocal outgoing connections. (d) Scatterplots of 2 times the reciprocal degree of neurons versus their total degree, divided by neurotransmitter. Dotted lines indicate a factor of 2 around the $x = y$ line.



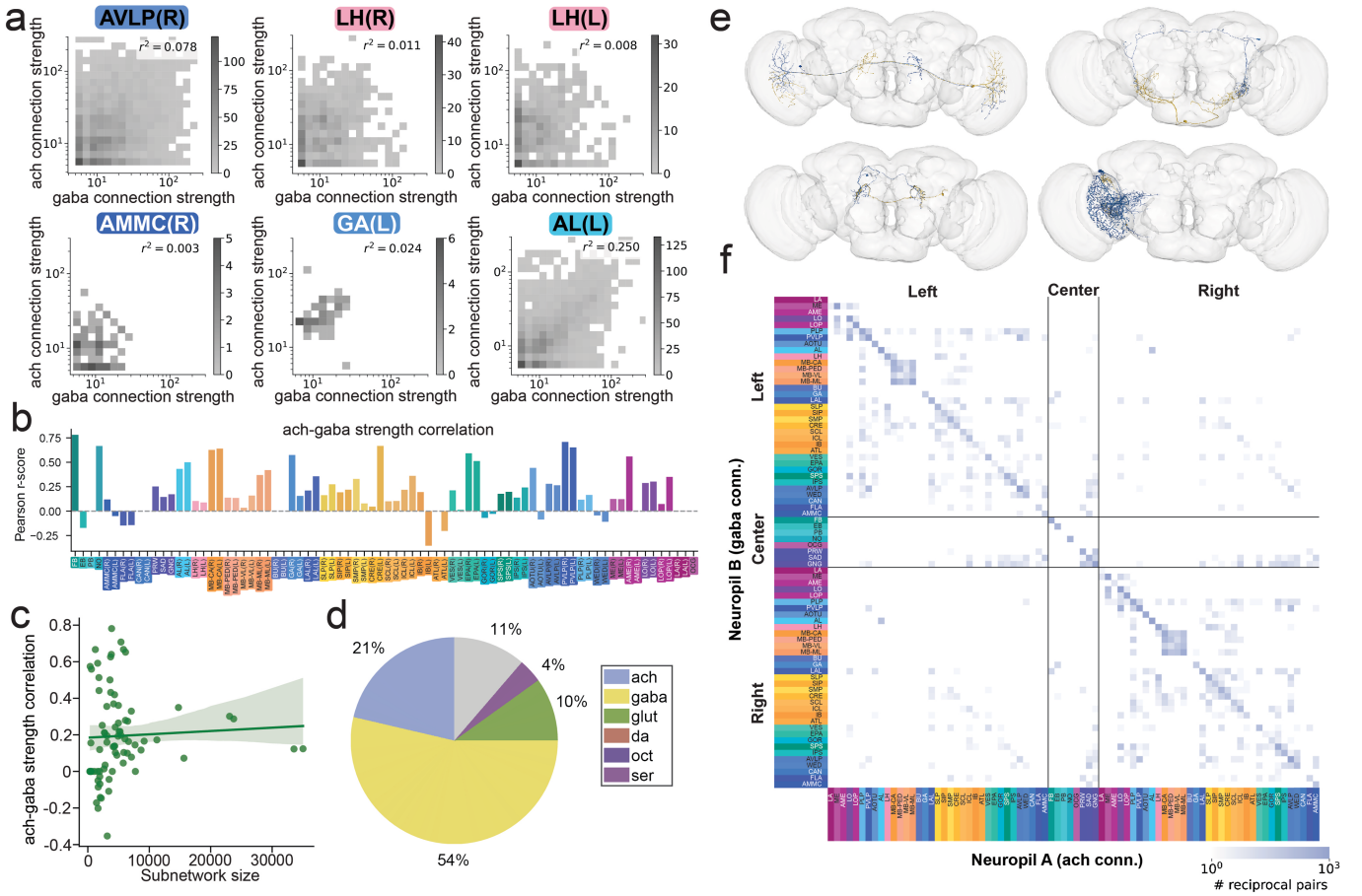
Extended Data Fig. 4 | Additional large-scale connectivity analyses. In-degree vs. out-degree scatterplots showing broadcaster, rich balanced, and integrator regimes, with neurons plotted by (a) the putative neurotransmitter of each neuron and (b) the super class of each neuron. (c) Comparing the input and output sides of all intrinsic neurons, rich club neurons, integrators, and broadcasters. The asymmetry in L/R percentages for broadcaster neurons is due to the large number of medulla-intrinsic broadcasters which connect with photoreceptors (Proofreading of photoreceptors was incomplete in Snapshot v630). (d) Percentile rank distributions of central and optic lobe intrinsic neuron populations to all inputs. (e) Percentile rank distributions of rich club, integrator, and broadcaster neuron populations from various input modalities. (f) Scatterplots of percentile rank from one sensory modality on each axis. Broadcaster neurons are highlighted in teal and integrator neurons are highlighted in purple.



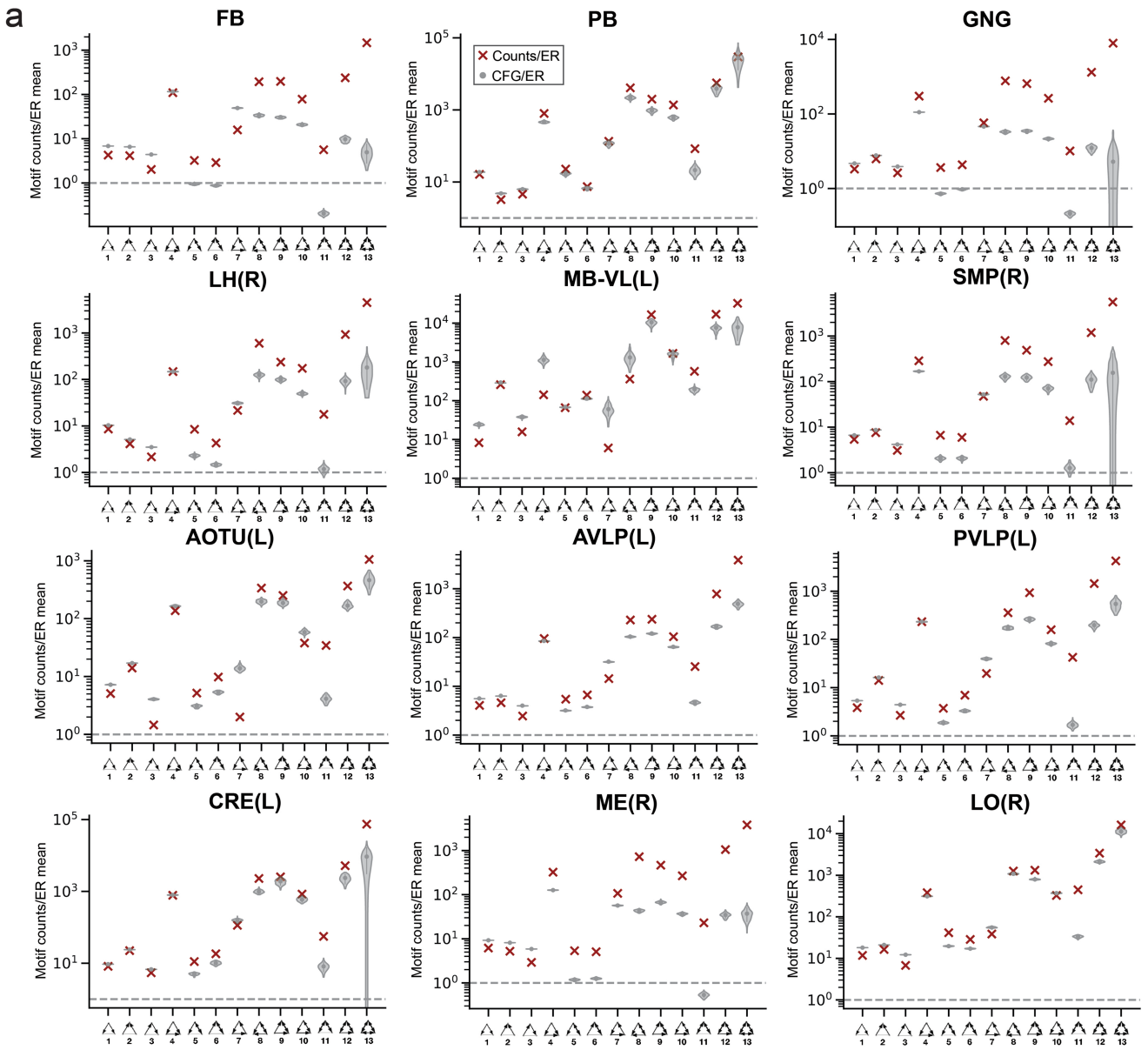
Extended Data Fig. 5 | Internal and external connections across neuropils. (a) The number and (b) relative fraction of neuron weights in each neuropil making connections internal to that neuropil, external incoming connections, and external outgoing connections. Each neuron contributes a total weight of 1, computed based on the fraction of incoming and outgoing synapses the neuron has in each neuropil. (c) Comparing the neurotransmitter composition of all internal and all external neuron weights across the whole brain and (d) by neuropil.



Extended Data Fig. 6 | Additional neuropil-specific connectivity differences. (a) The number of neurons included in each neuropil subnetwork. (b) The average connection strength (synapse threshold of 5 synapses/connection applied) of connections made in each neuropil (above), and the connection probability of each neuropil (below). (c) Reciprocity normalized by connection density for all 78 neuropils. (d) Average reciprocal connection strength normalized by average unidirectional connection strength in all neuropils. (e) The relative fraction of glutamatergic neurons participating in reciprocal and unidirectional connections. Absolute percentages of (f) acetylcholine, (g) GABA, and (h) glutamate occurrence in unidirectional and reciprocal connections within each neuropil subnetwork.

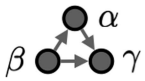
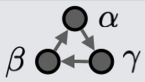
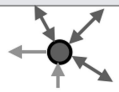
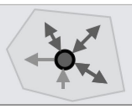
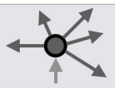
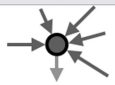
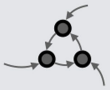
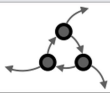


Extended Data Fig. 7 | Reciprocal connections within and between neuropils. (a) Heatmaps showing the relationship between excitatory (ach) and inhibitory (GABA) connection strengths in reciprocal connections in different brain regions. (b) Ach-GABA reciprocal connection strength correlations (Pearson r-score) for all neuropils. (c) These correlations do not appear to be correlated with neuropil subnetwork size. (d) The neurotransmitter composition of the population of neuropil-specific highly reciprocal neurons (NSRNs). (e) Examples of inter-neuropil reciprocal neuron pairs, one neuron in blue and one neuron in gold. (f) Map of the total number of ach-GABA reciprocal pairs between different neuropils.

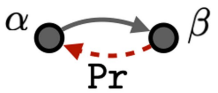
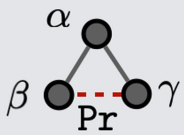


Extended Data Fig. 8 | Additional differences in three-node motifs across neuropils. (a) Three-node motif distributions for additional neuropils. The frequency of each motif relative to that in an ER null model is plotted to the right, together with the average motif frequencies of 100 CFG models (grey violin plots). (b) Average strengths of edges participating in 3-node motifs in the different neuropil subnetworks relative to the average edge strength in each subnetwork.

Extended Data Table 1 | Definitions for all neuron populations identified in this paper

	Neuron lists available on Codex		Definitions
2-neuron motifs	reciprocal connection participants		all neurons that participate in reciprocal connections.
3-neuron motifs	feedforward loop participants		neurons that participate in feedforward loop motifs consisting of unidirectional connections: β→α, α→ γ and β→γ, precisely.
	3-unicycle participants		neurons that participate in 3-unicycles consisting of unidirectional connections: β→α, α→ γ and γ→β, precisely.
N-neuron motifs	highly reciprocal neurons		neurons with the numbers of reciprocal edges ≥ 0.5 x total-degrees.
	neuropil-specific highly reciprocal neurons (NSRNs)		intrinsic rich-club and highly reciprocal neurons with ≥ 50% of incoming connections, and ≥ 50% outgoing connections contained in the same neuropils, respectively.
Rich-club analysis	rich-club neurons		high-degree neurons that are densely connected with other high-degree neurons (total-degree is higher than 37).
	broadcasters		intrinsic rich-club neurons with out-degrees ≥ 5 x in-degree.
	integrators		intrinsic rich-club neurons with in-degrees ≥ 5 x out-degrees.
Spectral analysis	attractors		top 3% most visited neurons in a forward random walk over the largest strongly connected component.
	repellers		top 3% most visited neurons in a reversed random walk over the largest strongly connected component.

Extended Data Table 2 | Network statistics of the fly connectome with no threshold on the number of synapses per connection (left) and a threshold of 5 synapses per connection

Neuronal wiring diagrams	Fruit fly (no threshold) *Drosophila melanogaster* (Dorkenwald *et al.*, 2024) 	Fruit fly (≥ 5 synapses) *Drosophila melanogaster* (Dorkenwald *et al.*, 2024) 
Network size 	127,978 neurons 14,680,950 connections	127,978 neurons 2,613,129 connections
Avg. connection strength 	3.59 synapses 1 ~ 2358	12.61 synapses 5 ~ 2358
Connection probability 	0.000896 x5.62	0.000160 x1
Connection reciprocity 	0.265 x293 than ER x46.1 than CFG	0.138 x858 than ER x43.8 than CFG
Clustering coefficient 	0.108 x59.7 than ER x9.17 than CFG	0.0463 x144 than ER x7.57 than CFG

nature portfolio

Corresponding author(s): Mala Murthy

Last updated by author(s): Jun 13, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	*Provide a description of all commercial, open source and custom code used to collect the data in this study, specifying the version used OR state that no software was used.*
Data analysis	Data was analyzed using custom MATLAB and Python scripts which are publicly available on Github.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The full connectome data has been made publicly available on codex.flywire.ai. We have also prepared a Zenodo repository for publication.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The statistics presented here describe the first (and currently only) complete EM whole-brain connectome in Drosophila. As such, the population statistics have an N=1 In these cases, comparisons to statistical null models are made wherever possible. However, for analyses involving neuron subpopulations and motifs, we were able to report averages over large populations of neurons (order 10,000 or more).
Data exclusions	No data were excluded from the whole-brain analyses.
Replication	N/A. Analysis was done on the first (and currently only) densely reconstructed whole-brain connectome in Drosophila melanogaster.
Randomization	N/A
Blinding	N/A

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	Drosophila melanogaster
Wild animals	N/A
Reporting on sex	Female fly (this individual was first reported in Zheng et al., 2014).
Field-collected samples	*For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.*
Ethics oversight	*Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*