A network-based method for extracting the organization of brain-wide circuits from reconstructed connectome datasets

Manjunatha K.K.H.[2,#], Bruzzone M.[1,3,#], Nicoletti G.[2,3], Suweis S.[2,3,#,@] and dal Maschio M. [1,3,#,@]

[1,] Department of Biomedical Sciences, Università degli Studi di Padova, Padua, IT
[2,] Department of Physics and Astronomy, Università degli Studi di Padova, Padua, IT
[3,] Padova Neuroscience Center, Università degli Studi di Padova, Padua, IT

[#,] equal contribution
[@,] corresponding author

Lead contact: marco.dalmaschio@unipd.it

Abstract

Understanding brain mechanisms is a currently open challenge. There are several methods available to characterize the functional aspects. However, the information that one can retrieve about the organization of the wiring between the cells across the brain that could support the network organization and the circuit mechanism remains limited. This appears to be a general problem, even when the cell anatomy is well described and potentially accessible at scales up to the synaptic level across the whole brain. Typically, such high-resolution analyses focus on the reconstruction of a limited portion of the brain or of a specific circuit, while missing the general organization principles of the neuronal networks within the brain, the network components, and internetwork connection schemes. To extend the investigation on these aspects, we present a theoretical and computational framework using modularity analysis to identify from connectome datasets the brain-wide circuit organization and the underlying cell modules based on synaptic connectivity patterns. We applied this approach on an Electron Microscopy (EM) dataset with about 25-thousand annotated cell reconstructions that incorporate synaptic information from the hemibrain of adult *Drosophila melanogaster*. The analysis provided a representation of the connectome organized in a hierarchical structure and characterized by several cell modules with precise and topographically organized connection patterns. We then tested this approach on a "synthetic" connectome, one obtained from a light microscopy dataset of more than three thousand neuronal skeletonizations from the zebrafish larvae brain, where a cell-to-cell proximity rule was used to assign synaptic contacts. Despite

this approximation, the developed approach revealed reported connectivity patterns and functional connection schemes. In conclusion, we show that our framework is scalable and amenable for different types of datasets, independently from the particular anatomical reconstruction method adopted, to reveal the neuronal architectures supporting the brain mechanisms.

Introduction

The understanding of the brain mechanisms takes advantage from the continuous refinement of molecular, hardware, and software approaches. These methods allow reconstructing, in suitable model organisms, the circuit activity from synapse level to brain wide scales[1–5]. Current optical and electrophysiological methods represent invaluable tools for studying the neuronal dynamics associated with sensory information, the fine-tuning of motor programs, and spatiotemporal patterns and the networks characterizing the resting state activity[6–8]. The obtained dataset are fundamental building blocks for the formulation, and the possible falsification, of circuit models describing the functional working mechanisms[9–11]. However, even under optimal conditions, one faces a typical scenario where different models, despite their intrinsic architectural difference, number of parameters, and complexity, are similarly good in capturing the same circuit dynamics[12,13]. With the aim to better constrain this kind of inverse problem, one possible strategy could be to complement the amount of information available on the system under investigation with other functional or structural aspects[14]. In this view, intervention paradigms have been proposed and adopted to extract a set of relationships of necessity or sufficiency, linking the different neuronal components within the circuit[15,16]. Indeed, increasing or decreasing the cell firing rate enables us to modulate the activity of targeted brain regions, neuronal populations, or arbitrary ensembles of neurons, and to evaluate the impact of targeted neuronal subpopulations in the observed patterns of activity and behavior[5,17,18]. However, even in this case, the large number of degrees of freedom that characterize this approach, except for particular cases, easily results in a combinatorial explosion that limits the actual application of this paradigm to dissect the circuit mechanisms and the contribution of its neuronal components[17]. It is convenient then to try to link these two functional aspects to the wiring properties of the underlying circuit, i.e., the anatomical organization and the connectivity

patterns characterizing the neuronal substrates involved[19]. One can take advantage of dense anatomical reconstructions of large brain parcels based on Electron Microscopy (EM)[20–25], which achieves sufficient resolution levels to characterize the synaptic contacts and possibly the neurotransmitters involved between neuronal partners while covering a large field of view. Retrieving the connectome across the whole brain, along with an accessible organism, requires a relatively long acquisition time, careful cell tracing, a cell segmentation process, annotation of the synaptic contacts, and lastly a thorough control of the precision of the reconstructions. Today, this is still a time-demanding process that limits its applicability to a few exemplars and requires computational strategies to minimize human workload[22,26]. Even assuming to be able to retrieve complete information on brain networks, this mapping at the sub-cellular scale of the interactions among neurons does not provide per se an insight into the organization principles of different circuits in the brain. Moreover, this type of anatomical reconstruction frequently focuses on targeted circuit motives, specific neuronal subpopulations or particular brain regions rendering this bottom-up approach sub-optimal to unravel the general organization principle characterizing the brain circuits.

Here, we address this open challenge by presenting a theoretical and computational approach to unraveling the brain modular organization of the brain from connectome datasets. We first applied this approach on a public EM dataset that provides an open connectome of the hemibrain of an invertebrate model organism, the *Drosophila melanogaster*[27]. We then extend the same analysis on a connectome reconstructed through a general and novel method from an annotated database of more than 3000 acquisitions of neuronal cells from the zebrafish larval brains co-registered in the same reference space[28]. Finally, we tested the results of the anatomical modules obtained by our approach on actual empirical data reported in the literature.

Results

1. Analysis of the Drosophila EM connectome

We approached the analysis of the adult brain *Drosophila melanogaster* connectome from the available information on synaptic contacts from the EM reconstructions of 21739 connected neurons[27]. The corresponding network structure, where the neurons represent the network nodes and the synaptic connections the corresponding network edges, is summarized in an

asymmetric adjacency matrix A_{ij}, whose entries are the numbers of the annotated synaptic contacts between the presynaptic i-th neuron and the postsynaptic j-th neuron. From the point of view of the topological organization, the properties of the graph[29,30] returned a network density of 0.0075, a clustering coefficient of 0.0001, and an average shortest path length of 2.9775 for the weighted analysis (see the Methods section for details). The in- and out-degree distributions showed a general heterogeneity: most of the nodes are characterized by a relatively small number of interactions ($<10^3$) and few ones by a high number of connections ($>10^4$) (known as hubs, see Supplementary Figure 1).

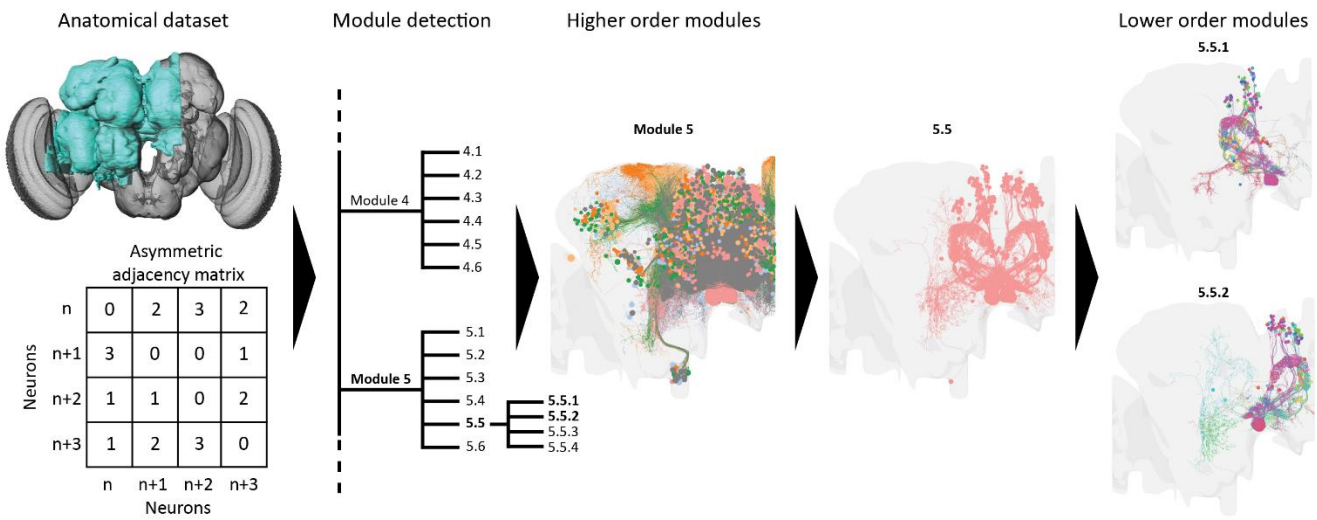


Figure 1. The general approach. *The application of the Leiden algorithm on the published anatomical dataset and its corresponding adjacency matrix enables the identification of a series of higher-order modules based on connectivity patterns. The successive iteration of the algorithm allows the identification of lower-order modules.*

Aiming at extending the analysis of the connectome dataset to understand the structure and the organization of brain-wide networks, we considered applying a network algorithm for the detection of modules of neurons within the original network that represent densely connected subgroups of nodes[31–33]. This wide class of algorithms typically works based on the optimization of a parameter associated with the network connectivity structure. One possibility is to consider network modularity (see the Methods section), a function that, by evaluating the difference between the actual number of edges in a module and the expected number of edges in a randomly connected network, provides a representation of the modular organization of the network[34]. Although this approach has been frequently used to investigate the adjacency matrix originating from a functional connectivity dataset as human brain fMRI[35,36], it has rarely been applied to characterize the brain organization from a structural point of view[33,37,38]. We adopted

here the Leiden algorithm, as it provides superior guarantees on the density of the discovered modules and effective connectivity[39]. This analysis, taking advantage of the available adjacency matrix describing the entire brain network obtained from synapses annotation (Figure 1 – Anatomical dataset), can reveal the existence of main modules (higher order modules) (Figure 1). To capture the structure of the corresponding networks, we rendered the corresponding wiring diagram of the modules in anatomical space, thus, facilitating the evaluation of the physiological relevance of the identified neuronal modules. Once a network has been identified with this approach, it is possible to re-iterate the algorithm to identify sub-components of the original module (Figure 1 Lower order modules).

This analysis, applied to the Drosophila EM dataset, revealed a hierarchical organization with eight higher-order modules (first iteration of the Leiden algorithm), each counting an average number of three thousand cells (Figure 2A). Reiterating the Leiden algorithm on each high-order module showed that these have a variable number of lower-order components (second iteration of the Leiden algorithm , Figure 3), ranging from four to nine. A further iteration on these components revealed different cell subcommunities (third Leiden algorithm iteration)

(Supplementary Figures 2 - 9, complete dataset modular structures available at https://github.com/dalMaschio-lab/Connectome_analysis).

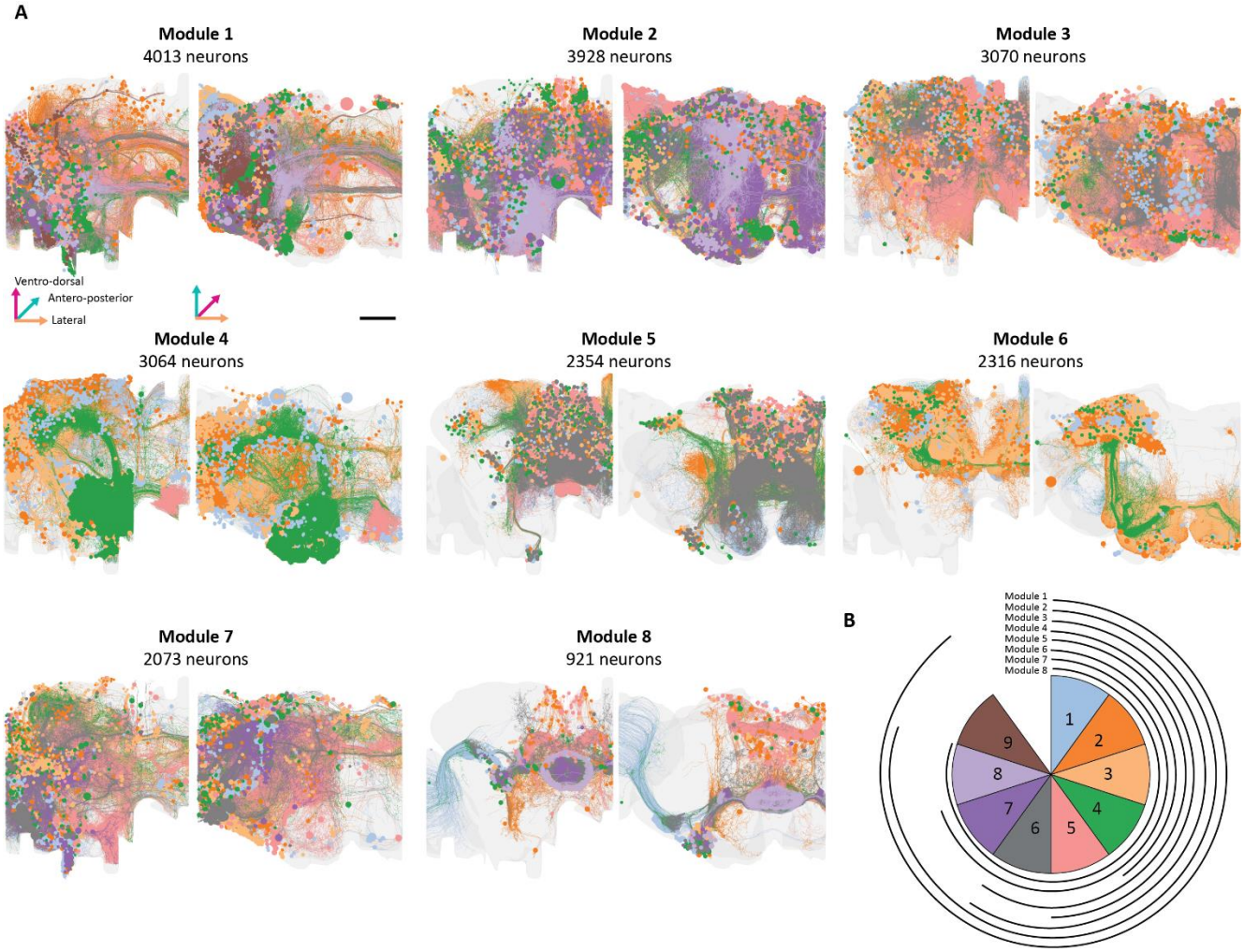


Figure 2. The Drosophila EM higher-order modules. A) *Anatomical reconstructions of the neurons belonging to the eight high-order modules found in the EM dataset overlayed on the reconstructed hemibrain. Colors are assigned according to the sub-modules described in panel B. Scale bar = 50 µm.* ***B)*** *Decomposition of the higher-order modules in submodules obtained with the second iteration of the algorithm. Colors and numbers are assigned according to the submodules. The length of the lines represents the number of submodules for each module.*

By focusing on a single higher-order module and evaluating its structure, it is possible to discriminate basic organizational schemes underlying the neuronal wiring in a certain region or nucleus of the brain. This appears as a general aspect, which is particularly evident in the case of module eight, mainly associated with the central complex (CX) and especially with the ellipsoid body (EB) network (Figure 3A). The resulting decomposition of this brain module reveals a subset of lower-order components. They can be divided into three main groups whose

components display peculiar spatial segregations that reflect different functional roles. The first group (Figure 3B, submodule 8.1) comprises neurons that project in the anterior optic tubercle (AOTU) and the early visual pathway. These are primarily medulla columnar neurons that receive visual inputs and convey them to the bulb (BU)[40]. The second group (Figure 3B, submodules 8.3, 8.4, 8.6, 8.7, 8.8) consists mainly of tangential neurons that transmit signals from the BU to the ellipsoid body (EB) forming a series of concentric lamination domains[40,41]. The third group (Figure 3B, submodules 8.2, 8.5) comprises columnar neurons projecting to the protocerebrum (PB) and the EB involved in the interpretation of spatial information[41,42]. Furthermore, analysis of the inter-module communication revealed that the majority of neurons of module eight participate in module five, which is mainly associated with the CX (Supplementary Figures 7 and 10).

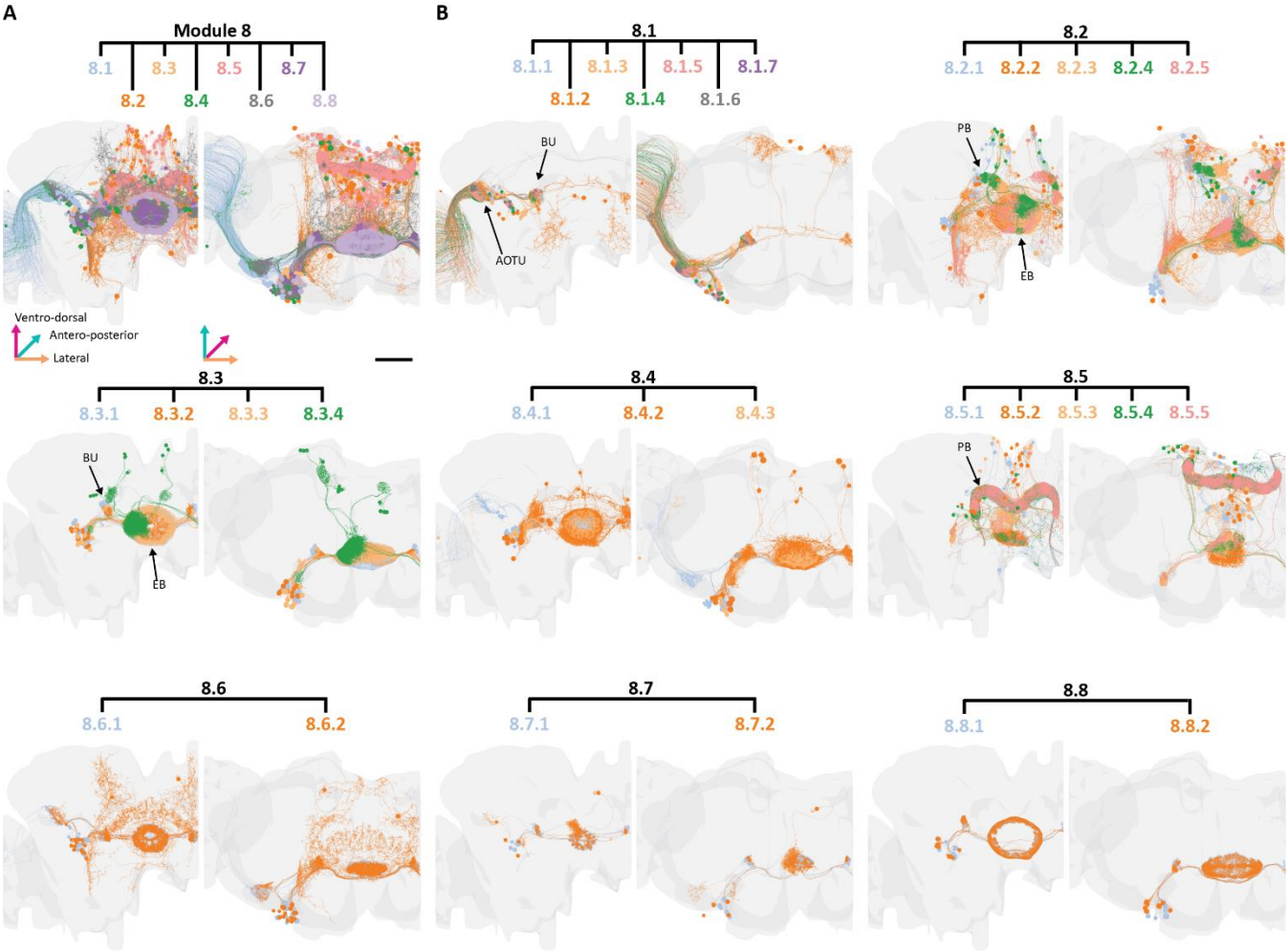


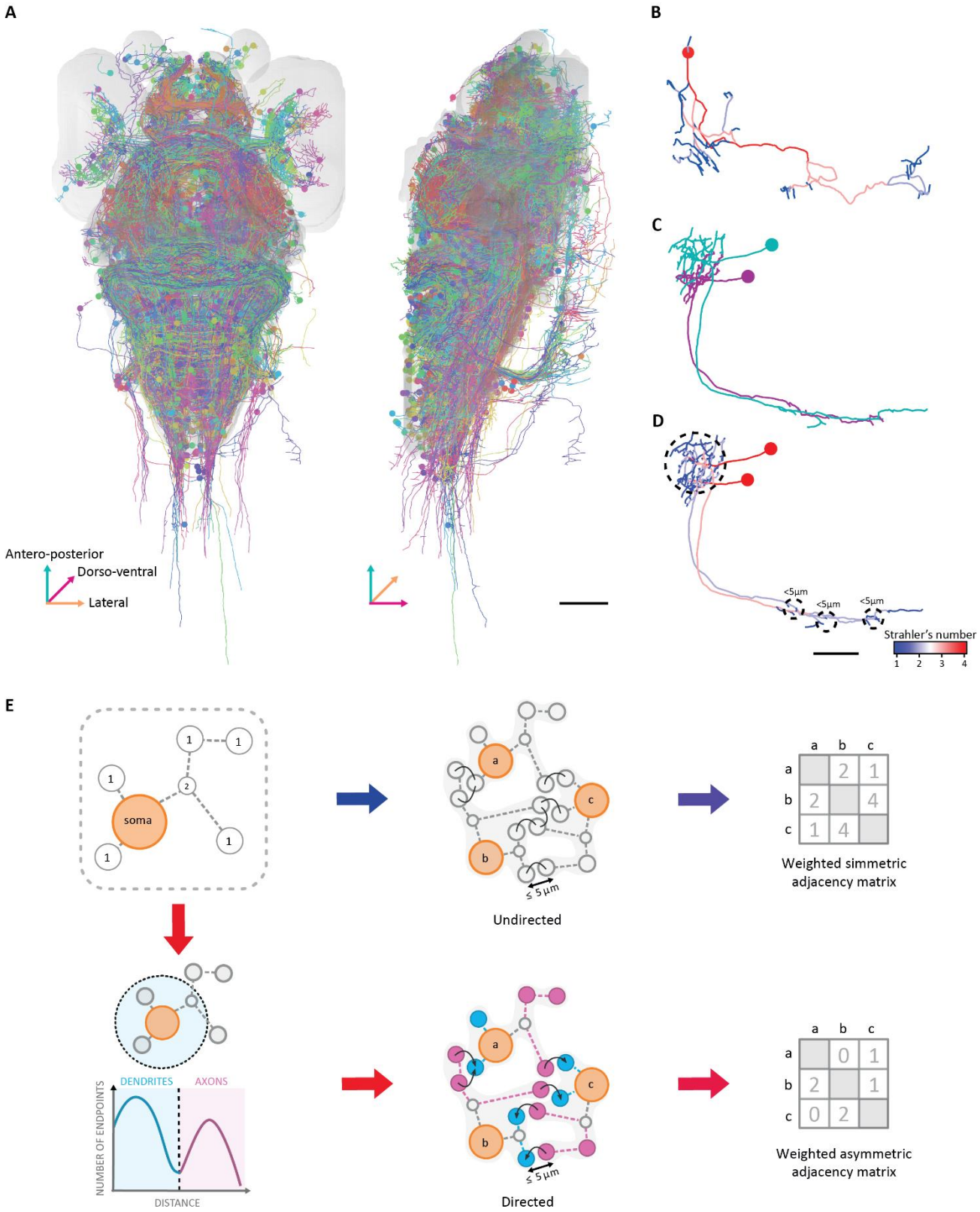
Figure 3. The Drosophila EM ellipsoid body module. A) *Anatomical reconstructions of the neurons assigned to the high-order module eight decomposed into the lower-order modules. Colors are assigned according to the order of the lower modules. Scale bar = 50 µm.* ***B)***

Sub-modules from 8.1 to 8.8 corresponding to the original module eight shown in A. AOTU: anterior optic tubercle, BU: bulb, PB: protocerebrum, EB: ellipsoid body.

It is also evident from the analysis that, within the hierarchical organization of the Drosophila EM brain modules, there exists a certain degree of heterogeneity, with the compresence in the same module of submodules. These are characterized on one side by relatively simple structures with one or two different neuronal populations, and on the other side by a more articulated organization involving from three to six neuronal components. This further level of hierarchical stratification can be assessed by re-iterating the module detection, so that to reduce the more complex structures in elementary components. This is particularly evident in the case of spatially segregated innervation patterns of the columnar neurons at the level of PB and EB. (submodules 8.2.3 and 8.2.4 Supplementary Figure 11)[24,43].

2. Building an approximated connectome of the zebrafish larva brain

Unraveling the structure of the brain network requires a suitable dataset, with sufficient sampling of the cellular morphologies and coverage across the different regions of the brain. Currently, even if there are several reported EM reconstructions[21,22,24,25,44–48], brain-wide connectome datasets that provide skeletonization of neurons and detailed annotations of their synaptic contacts are still missing for a vertebrate organism. To overcome this limitation, we considered the possibility to extend the application of the module identification algorithm on a dataset with individual cell anatomical reconstructions obtained with optical microscopy co-registered on the same reference brain[28] (Figure 4A). In this case, where actual synaptic information is lacking, we devised an approach to reconstruct the whole brain network by assigning cell-to-cell connections on the basis of a proximity rule between cells (see the Methods section). For this purpose, we calculated for each point of the cell skeleton the corresponding Strahler's number, a parameter that represents the relative level of the cell anatomy points in the cell branching hierarchy[19] (Figure 4B). Terminal points along the neurite structures, characterized by a unitary value of the Strahler number, are considered putative points of connection between different cells if they fall in a defined proximity range one from the other. Based on the typical errors in the anatomical reconstruction and co-registration procedures, we adopted a proximity range corresponding to 5 µm (Figure 4D).



190

191 ***Figure 4. Building the synthetic connectome from a mesoscale atlas. A)*** *Anatomically reconstructed neurons from* [28] *overlaid to the*
192 *zebrafish atlas. Colors are assigned randomly. Scale bar= 100 µm.* ***B)*** *Representative neuron colored according to Strahler's number.* ***C)***
193 *Two representative neurons anatomically located one close to each other.* ***D)*** *Same neurons of C colored according to the Strahler's*

number. Endpoints (Strahler's number = 1, color dark blue) closer than 5 μm are defined as connections (dashed circles). Scale bar= 50 μm. ***E)*** *Schematic representation of the pipelines for generating the undirected synthetic connectome (blue arrows) and the directed synthetic connectome (red arrows).*

As reconstructing the connectome relies on the determination of the adjacency matrix A_{ij}, we consider two different scenarios: the first one, where we assume no subcellular segregations for the input and output points of a neuron, resulting in a symmetric adjacency matrix A_{ij}^{U} and an undirected connectome (Figure 4E, Figure 5A); in the second one, we considered that most of the neuronal reconstructions are characterized by a bimodal spatial distribution of the neuronal terminals [49] (Supplementary Figure 12) and we assigned for each cell a separation in two putative components, input dendritic points in the regions proximal to the soma and a output axonal points in more distal areas. This allowed us to constrain the connectome according to an axon-to-dendrite communication scheme. In the latter case, we obtained an asymmetric adjacency matrix A_{ij}^{D} representing a directed connectome (Figure 4E, Figure 5E-H). The two connectome scenarios resulted in a rather articulated scheme of neuronal interactions between the different regions of the brain (Figure 5B, Figure 5F).

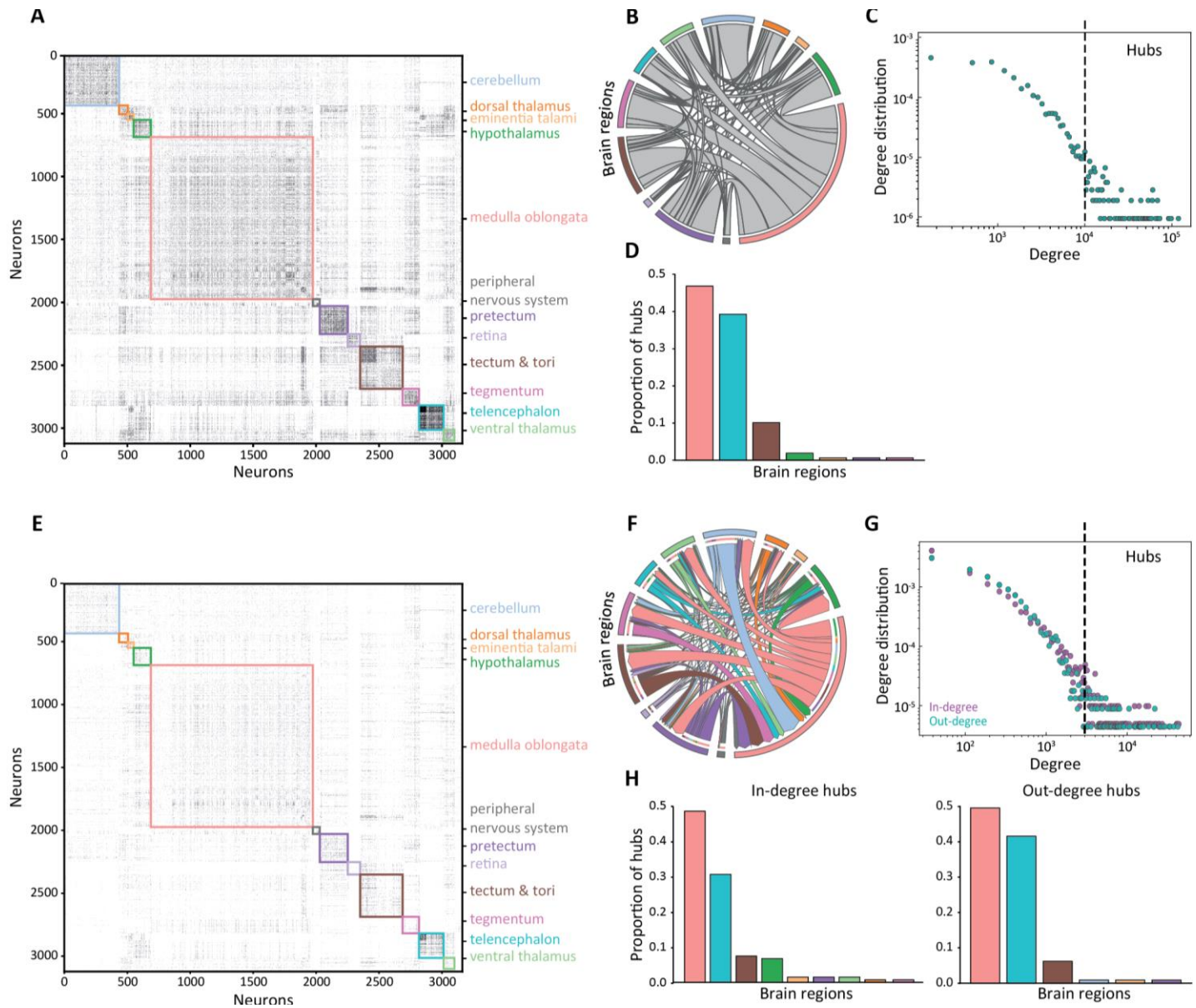


Figure 5. Topological properties of undirected and directed networks. A, E) *Binarized adjacency matrixes showing connections in the undirected (A) and directed (E) networks with neurons sorted by brain regions in which the soma is located. The matrix is binarized for visualization purposes.* ***B, F)*** *Chord plots representing inter-regions connections for the undirected (B) network and incoming – outgoing inter-regions connection for the directed network (F). Labels for the colors are reported in A and E.* ***C, G)*** *Weighted degree distributions for the undirected (C) and directed (G) network. Dashed lines depict the threshold above which neurons are defined as hubs.* ***D, H)*** *Bar plots representing the distribution of hubs in the brain regions for the undirected (D) and directed (H) networks. Labels for the colors are reported in A and E.*

From the obtained adjacency matrices, one for the undirected, A_{ij}^{U}, and the other for the directed, A_{ij}^{D}, graphs, we extracted a set of graph parameters. In general, being the non-zero entries of the directed matrix a subset of the undirected ones, A_{ij}^{D} results sparser than A_{ij}^{U}. The analysis of A_{ij}^{U} revealed the existence of 23 connected subgraphs, with a Giant Component

(GC) accounting for 99% of the edges of the original graph and a set of small, poorly populated disconnected subgraphs (accounting for 32 out of total 3163 nodes), that have been neglected from further analyses. The GC presents a network density of 0.0397, pointing to a rather limited connectivity, a small weighted clustering coefficient of 0.00203 and an average shortest path length of 3.792. The undirected graph has a heterogeneous degree distribution, with most of the nodes characterized by a relatively small number of connections ($<10^3$) and fewer ones – hubs-like - with high degrees ($>10^4$) (Figure 5C). Interestingly, in the small degree region, the distribution shows a flat trend, suggesting that this graph deviates from the power-law behavior. An assortativity analysis gave a correlation coefficient β of 0.345, meaning that the graph is assortative and that nodes tend to connect with ones with a similar degree – i.e., small-degree nodes to small-degree nodes and hubs to hubs. We mapped hubs to the corresponding anatomical regions normalizing for the relative number of cells; it appears that the majority belongs to Medulla Oblungata (Figure 5D - Pink), followed by the Telencephalon (Figure 5D - Cyan) and the Tectum Opticum (Figure 5D - Brown). On the other side, regarding A_{ij}^{D} and the directed connectome (Figure 5E-H), the GC displays a lower network density (0.01), weighted clustering coefficient (0.0005), and weighted shortest path length of 6.10 for the weighted analysis. While the in- and out-degree distributions resemble the general heterogeneity of the undirected scenario, the low-degree region is not showing a corresponding saturation (Figure 5G).

4. Analysis of brain-wide modules from the zebrafish larvae connectome

Similarly to the case of the Drosophila connectome, we applied the Leiden algorithm for network discovery on the zebrafish connectome. Running the algorithm 10,000 times on the undirected GC typically resulted in the stable and robust identification of 12 neuronal modules, with small deviations in the assignment of a few nodes from trial to trial (<0.5%). However, only eight of the identified modules are prominent and well populated, the remaining containing an extremely limited number of nodes (4 out of 3116 and 14 out of 3131 in total for directed and undirected graphs) (Figure 6A).

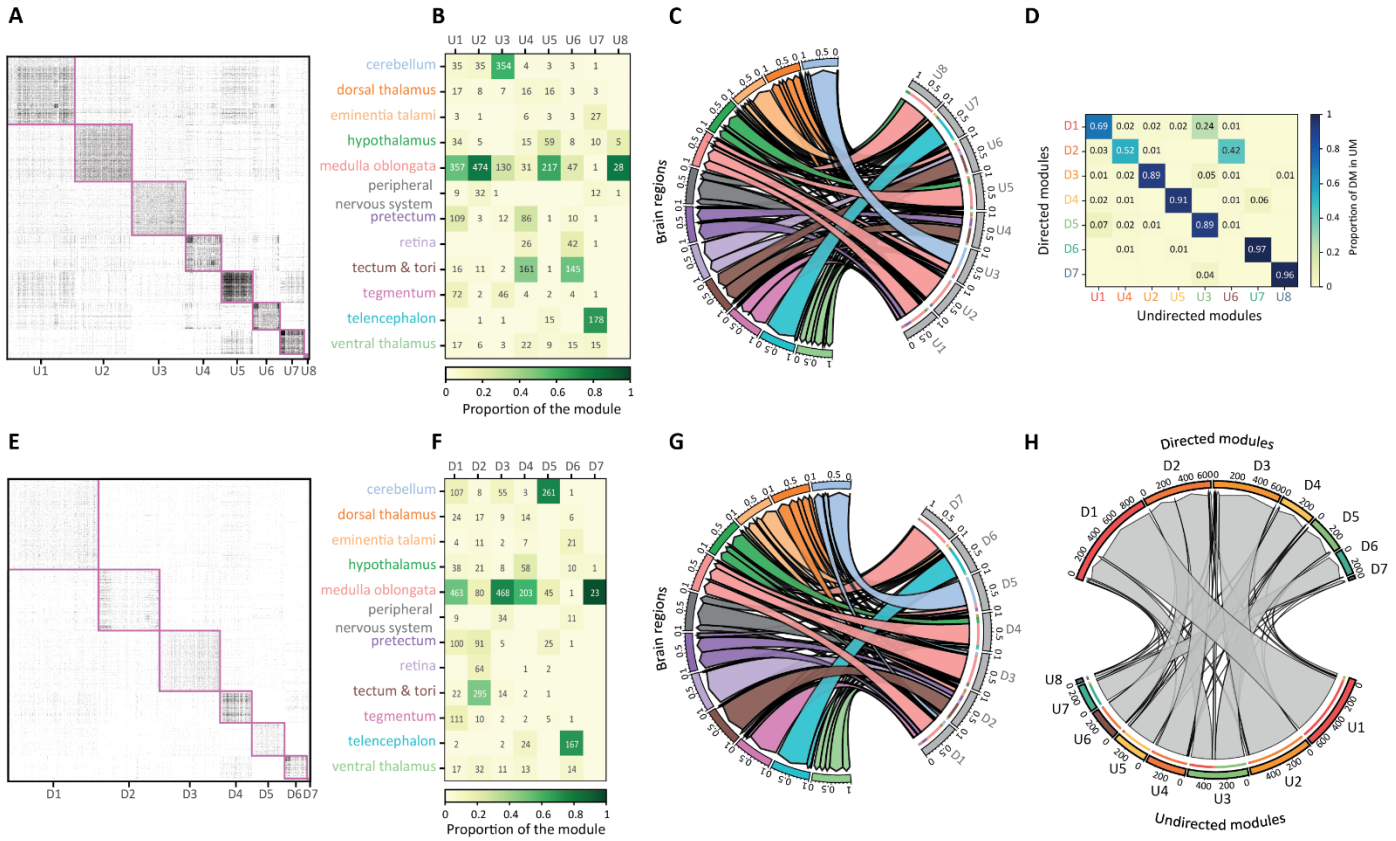


Figure 6. Distribution of the modules in the different regions for the undirected and directed networks. A,E) *Binarized adjacency matrices showing connections in the undirected (A) and directed (E) networks sorted by the identified modules. The matrix is binarized for visualization purposes.* ***B,F)*** *Distribution of the modules across the brain regions for the undirected (B) and directed (F) network. Number refers to neurons belonging to a certain module and with the soma located in the regions. Matrix cells are color-coded based on their proportion within the module. U: undirected module, D: directed module.* ***C,G)*** *Proportion of the distribution of the modules across the brain regions for the undirected (C) and the directed (G) networks. Labels for the colors are reported in B and F.* ***D)*** *Proportion of neurons of the undirected network modules present in the directed network modules.* ***H)*** *Number of neurons of the undirected network modules present in the directed network modules. Labels for the colors are reported in D.*

To characterize the resulting networks, we first mapped all identified undirected modules (U) in the brain regions containing the body of neurons belonging to each corresponding module (Figure 6B). In general, all modules present rather distributed spatial patterns, covering a number of brain regions in a non-exclusive manner. In addition, most of the brain regions in this analysis, although to different extents, participate in more than one module (Figure 6C). Similar to the undirected connectome, the analysis of the directed connectome revealed the presence of seven main modules (D) (Figure 6E-G), showing a substantial overlap, beside two directed modules. D2 splits into two undirected modules, U4 and U6, and D1 combines U1 with a fraction of U3 (Figure 6D, H). This substantial overlap between the two modularity analyses points to the possibility that the two corresponding connectomes generated in this way rely to a large

extent on a common backbone represented by a directed graph and axo-dendrite communication.

To better capture the 3D anatomical organization of the 8 identified modules and the extension of their network structures, we represented the corresponding wiring diagrams in the anatomical reference space. In some of the cases, the identified modules correspond to previously described brain-extended circuits underlying information processing or motor control functions, like the visual network module (U4 and U6) and the cerebellar network module (U3) (Figure 7).

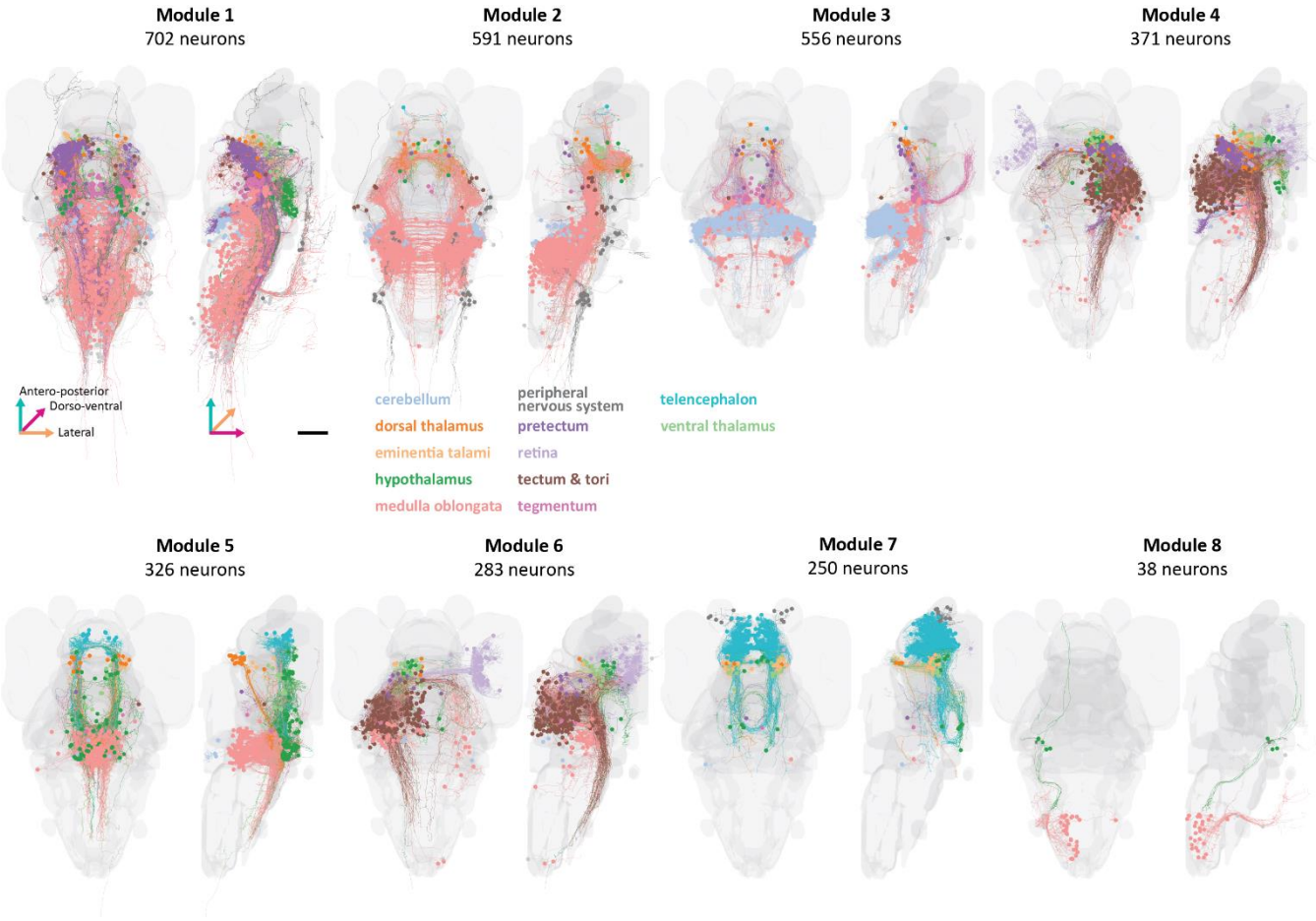


Figure 7. Modules of the zebrafish brain. *Anatomical representation of the eight high-order modules found in the zebrafish dataset overlaid on the brain silhouette. Labels indicate the most represented brain regions in the module. Colors are assigned according to the distribution of soma in the different brain regions. Scale bar = 100 µm.*

Similarly to the Drosophila EM dataset, we applied the community detection algorithm to each identified module to reveal finer patterns of network organization. This led to the finding of a

variable number of subcomponents, typically ranging between five and seven (Supplementary Figures 13 - 19). These submodule patterns, even with the limitations deriving from the approximation, captured an additional level of the circuit organization, involving aspects of spatial segregation of the neuronal populations across the brain. This is the case for the organization of the RGC projections in the tectum: an analysis of the relationship of the cell body position between the retina (Retinal Ganglion Cells, RGCs) and the contralateral tectum (PeriVentricular Neurons, PVNs) showed a retinotopic organization in module 4 (submodules 4.2 and 4.6, right tectum) and in module 6 (submodules 6.1 and 6.2, left tectum) (Supplementary Figure 20). Similarly, an articulated spatial pattern of connections involving different brain areas characterized the module U3, associated with a pretectal-cerebellar network (Figure 8 and Supplementary Figure 21). The connections formed between neurons located in the medulla oblongata, cerebellum, and tegmentum (Figure 8, submodules 3.3 and 3.4) resemble those associated with vestibulo-ocular reflex (VOR)[50]. The ascending and ascending/descending tangential neurons are located in the medial octavolateral nucleus (MON) of the medulla (Figure 8, submodules 3.3 and 3.4), the recipient of utricular inputs. These neurons project to the oculomotor nucleus (nIII) where extraocular motor neurons mediate gaze stabilization[50]. Furthermore, connections between neurons located in the cerebellum have been shown to be involved in VOR[51,52]. The cerebellum displays a spatially segregated pattern of distinct cells, with lateral Purkinje cells projecting to the MON (Figure 8, submodule 3.2 and Supplementary Figure 21) and eurydendroid cells located in the middle with less specific projections[28,53] (Figure 8, submodules 3.1 and 3.5 and Supplementary Figure 21).

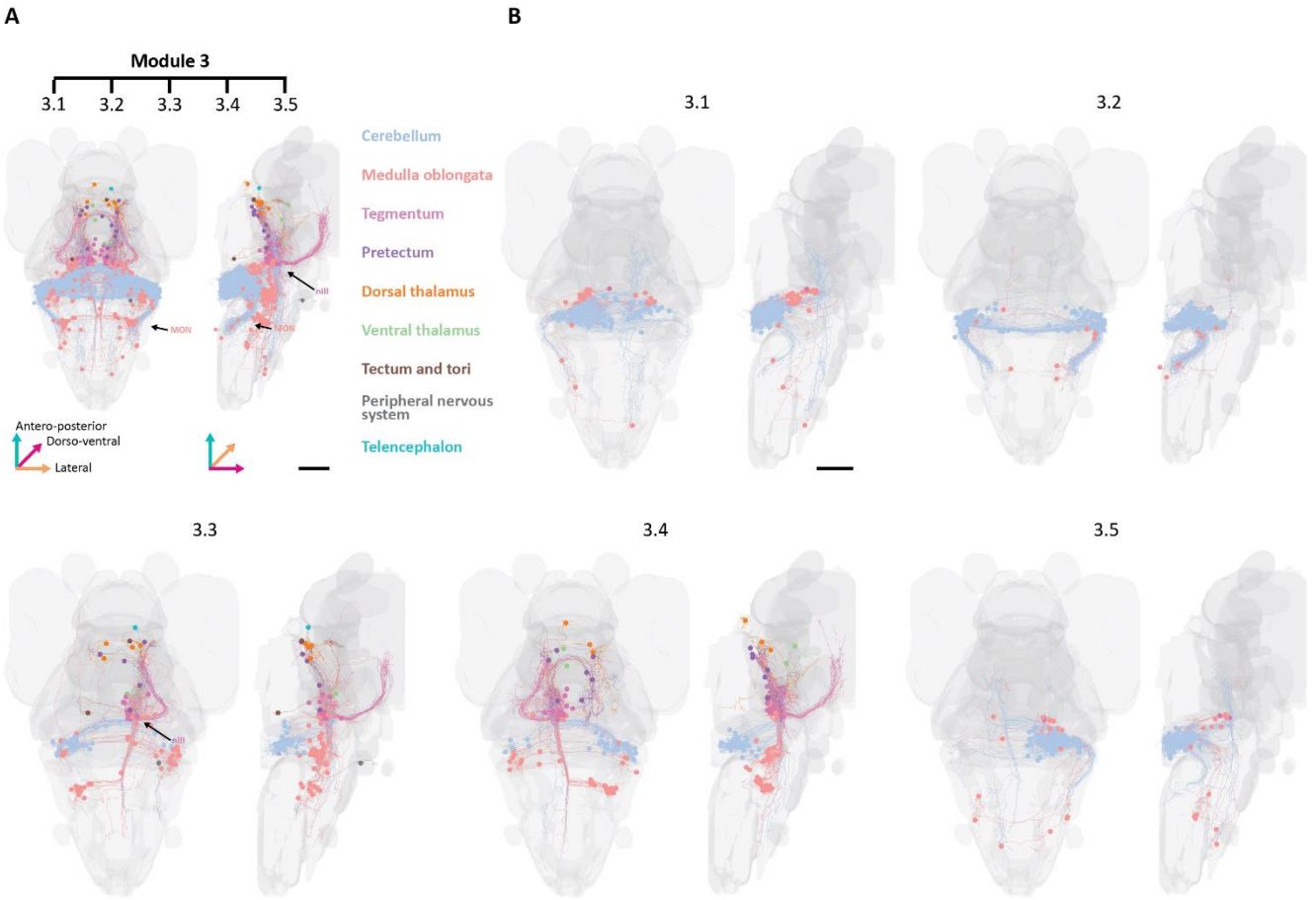


Figure 8. The zebrafish dataset cerebellar module. A) *Anatomical representation of the high-order module U3 as composed by the five lower-order modules (3.1-3.5). Colors are assigned according to the soma distribution in the brain regions. Scale bar = 100 µm. MON: medial octavolateral nucleus, nIII: oculomotor nucleus.* ***B)*** *Anatomical representation of the individual lower-order modules. Colors are assigned according to the location of the soma. Scale bar = 100 µm.*

The existence of principal networks supported by modules of neurons showing a relatively higher connection degree leaves open the question of how the different identified networks could interconnect with one another. To evaluate this aspect, we evaluated the links formed across different modules (Supplementary Figure 22). This analysis shows a rather heterogeneous scenario, with some modules characterized by a relevant number of connections, while others with a very limited number of inter-module links.

CONCLUSIONS

We have analyzed the graph properties of two brain connectome datasets, one obtained directly from the annotation of EM reconstructions of an invertebrate organism, *Drosophila melanogaster*, comprising about 25-thousand neurons each with its input-output synaptic

contacts; the other one, reconstructed from a mesoscale atlas of about 3200 single cell reconstructions obtained from a vertebrate organism, *Danio rerio*, where the input-output relationships were estimated. We then have applied on the corresponding adjacency matrices an algorithm based on the network modularity to identifying, according to a top-down process, those neuronal modules characterized by a relatively higher degree of connections, underlying network-like structures. With this approach and its reiterations, we have obtained a decomposition of the connectomes of two phylogenetically distant organisms on a set of basic components that form a sort of neuronal backbone of the brain wiring diagram. For most of the identified modules, it is possible to associate a functional role, either based on already reported anatomical characterizations of the circuits involved or on the evidences provided by functional analyses. In conclusion, the organization obtained of the modules of the Drosophila hemibrain EM and the Zebrafish larva light microscopy brain dataset show that the application of a top-down approach of community detection is able to capture relevant circuit motives.

Methods

Dataset and data extraction

Drosophila. We used the published *Drosophila melanogaster* adult hemibrain connectome[27]. The achieved isotropic resolution was 8x8x8 nm (X,Y,Z) and by automated segmentation algorithms more than 25000 neurons (approximately 25% of the total neurons in the Drosophila brain) were reconstructed and their synapses (10^6) annotated. Skeletonized neurons have been downloaded in the .swc[54] format from the NeuPrint[55] database using the 'Navis' Python library(https://github.com/navis-org/navis).

Zebrafish. We used a whole-brain atlas of the zebrafish larva[28]. The atlas comprises more than 3000 neurons, approximately 3% of the total neurons in the zebrafish larvae brain, covering almost the entire brain, anatomically reconstructed by random and sparse Green Fluorescent Protein (GFP) labelling using BGUG[56] or with other labelling techniques. With a typical resolution of 300x300x1000 nm (X, Y, Z), the size of the skeletonized dataset was approximately 60 MB. Skeletonized neurons have been downloaded in the .swc format[54] from the mapzebrain website (https://mapzebrain.org/home).

Computing the Strahler ordering for the zebrafish neuron reconstructions

For a given neuron skeleton i from the original dataset of 3163 reconstructions, we imported the .swc file and used the "networkx" library (https://networkx.org/)[57] to convert the corresponding collection of ordinated endpoints to a tree-like directed graph representing its anatomical structure with the soma considered as a root and the connected endpoints as children. In order to build an approximated connectome from this dataset, we first classified each skeleton endpoint within its branching hierarchy using Strahler ordering[19]. In this representation, terminal endpoints, those without branching children, obtain a value one. For the other endpoints in the .swc file, the integer value of the Strahler progressively increases with the number of upstream branches along the anatomical structure to reach the maximum at the level of the soma. For the calculation of the putative synaptic contacts and hence for the calculation of the connectome, we considered only endpoints with unitary Strahler number (below referred to as endpoint unless otherwise specified), which we denoted as E_{μ}^{i}, where the italic superscript refers to *i*-th neuron within the dataset and the Greek subscript stands for the μ-th endpoint of unitary Strahler number in the .swc file of *i*-th neuron.

Zebrafish undirected network reconstruction

We assume that all endpoints between two neurons form connections with each other. We introduce a proximity range u that represents the maximal effective distance at which two endpoints can form a connection. We set $u = 5\ \mu m$ based on the typical reconstruction error in optical microscopy[28,38,58]. Then, a link between two neurons is present if at least two of their endpoints are at a distance smaller than u. That is, given two neurons i, j and their endpoints $\{E_{\mu}^{(i)}\}$ and $\{E_{\nu}^{(j)}\}$, we compute the Euclidean distance matrix $d_{\mu\nu}^{(ij)} = d(E_{\mu}^{(i)}, E_{\nu}^{(j)})$, for all $\mu = 1, \dots, N_i$ and $\nu = 1, \dots, N_j$. The elements of the resulting adjacency matrix A_{ij}^{U} are the number of entries of $d_{\mu\nu}^{(ij)}$ such that $d_{\mu\nu}^{(ij)} \leq u$. Since $d_{\mu\nu}^{(ij)}$ is symmetric in the i, j indexes, A_{ij}^{U} is symmetric as well.

Zebrafish directed network reconstruction

The morphological data we considered do not contain information between the dendritic and axonal endpoints. Therefore, given the endpoints $\{E_{\mu}^{(i)}\}$ of a neuron, we computed the radial

distance $r_\mu^{(i)}$ between each endpoint $E_\mu^{(i)}$ and the soma. The associated probability distribution $p_i(r)$ resulted often bimodal, pointing to the fact that a large number of the endpoints are concentrated either close to the soma or far from it. We introduce a threshold T which corresponds to the antimode of $p_i(r)$, i.e., the local minimum between the two modes. If $r_\mu^{(i)} < T$, we assume that the corresponding synaptic endpoint is a dendritic endpoint for the neuron i, whereas $E_\mu^{(i)}$ is an axonal endpoint if $r_\mu^{(i)} \geq T$.

Then, the connections from the i-th neuron to the j-th one are obtained from the distances $d_{\mu\nu}^{(ij)} < u$, similarly to the undirected reconstruction. However, we now impose that a connection is counted only for those elements of $d_{\mu\nu}^{(ij)}$ such that μ is an axonal endpoint for neuron i, and ν is a dendritic endpoint for neuron j. Clearly, the resulting adjacency matrix A_{ij}^D is not symmetric, since an axonal endpoint connecting to a dendritic endpoint does not imply that the reverse connection is present.

Calculation of the network properties

The analysis of network parameters[59,60] was performed using dedicated Python libraries. The following functions from the 'networkx' (https://networkx.org/)[57] library have been used: 'connected_components' to identify the network connected components, 'average_shortest_path_length' to compute the average shortest path length, 'density' to compute the density of the network, 'average_clustering' to compute the average clustering coefficient and 'degree_assortativity_coefficient' to compute assortativity correlation coefficient.

Leiden algorithm for neuronal module detection

We adopted Leiden algorithm[39] to extract neuronal modules. Briefly, this algorithm consists of three main steps which are iterated until convergence of a quality function, in this case the network modularity. In our case we have chosen the well-known quality function for the undirected case[29] $Q_{undir} = \frac{1}{2m}\sum_{ij}\left(A_{ij} - \gamma\frac{k_i k_j}{2m}\right)\delta(\sigma_i, \sigma_j)$ and for the directed case[61] $Q_{dir} = \frac{1}{m}\sum_{ij}\left(A_{ij} - \gamma\frac{{k_i}^{out}\,{k_j}^{in}}{m}\right)\delta(\sigma_i, \sigma_j)$, where m is total number of edges or total edge weight, A is the corresponding adjacency matrix for directed or undirected graph, $\gamma > 0$ is the linear resolution

parameter[62] and is assumed equal to 1 in our analysis. Higher value of this parameter leads to more communities k_i and k_j are the weighted node degree of node i node j for the undirected case while for the directed it is mentioned in superscript to specify whether it is weighted out-degree ${k_i}^{out} = \sum_j A_{ij}$ or weighted in-degree ${k_j}^{in} = \sum_i A_{ij}$ and finally σ_i denotes the community of node i with δ representing the Kronecker delta.

In the fast local moving step, the algorithm starts from a singleton partition, i.e., each node is considered as its own community. Nodes are then moved among communities to maximize the quality function. Then, the obtained partition P is further randomly split into multiple communities, defining a new partition P* (refinement step). Initially, P* is considered as a singleton partition. A node of such P* is then merged with a randomly selected community such that the increase in the quality function is maximal, and communities remain connected. Mergers that lead to a decrease in the quality function are rejected. Finally (aggregation step), the communities of the refined partition P* are aggregated, meaning that each community in P* is considered a node in an aggregated graph. An edge between such nodal communities, say C and D, exists only when there already exists edges between the nodes of C and D [39].

Code and data availability

The scripts and data used to produce the results of the manuscript are available at https://github.com/dalMaschio-lab/Connectome_analysis.

Funding

The funding was provided by Horizon 2020 Framework Programme (734862) to M.D.M.

Contributions

K.K.H.M. : conceptualization and data analysis.

M.B.: conceptualization, data analysis, anatomical analysis, supervision.

G.N. : conceptualization, data analysis, supervision.

S.S.: conceptualization and supervision.

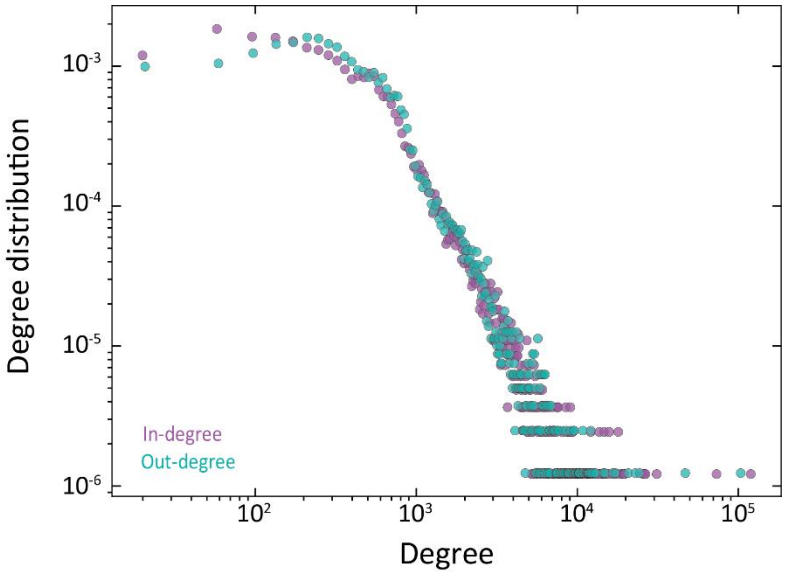
M.D.M: conceptualization, supervision and funding acquisition.

All the authors participated in the writing, review and editing of the manuscript.

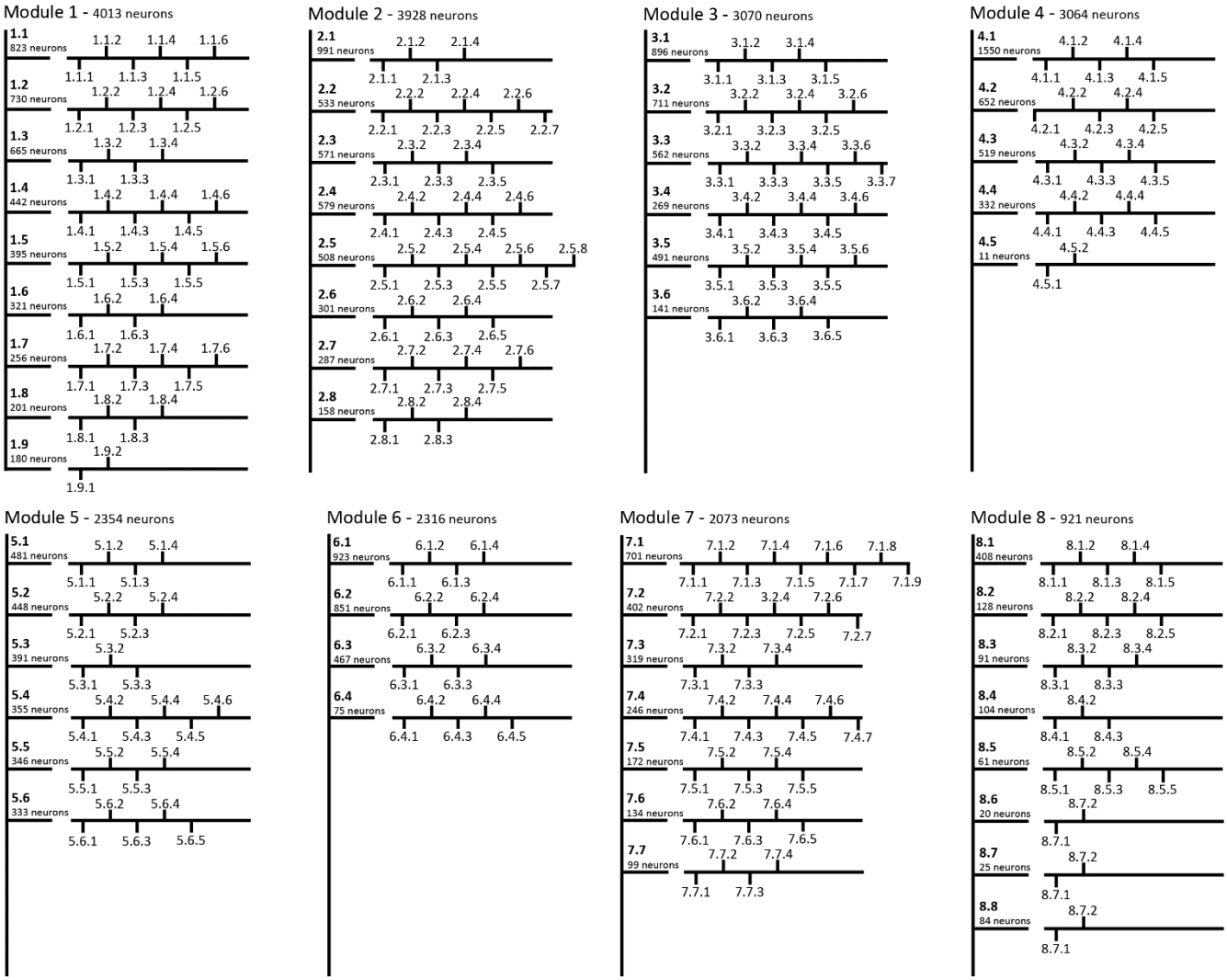
Aknowledgments

We thank Delfina Iriarte and Nora Nikoloska from the Physics Department of Padua for their support in preliminary analyses; we thank Aram Megighian from the Biomedical Department of Padua for support in anatomical analysis of Drosophila dataset; we thank Tomaso Erseghe from the Information Engineering Department of Padua for useful discussion.

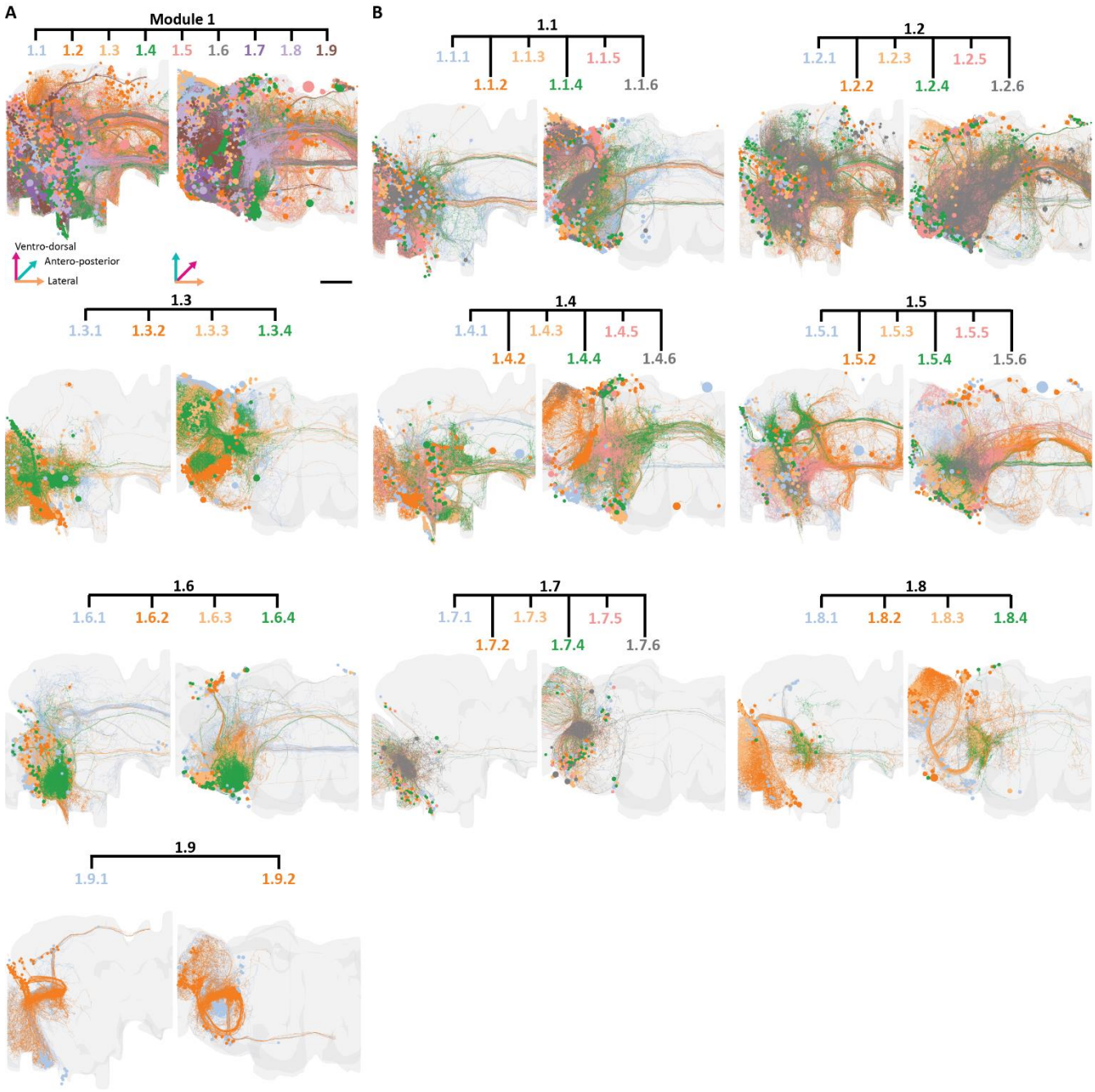
SUPPLEMENTARY FIGURES



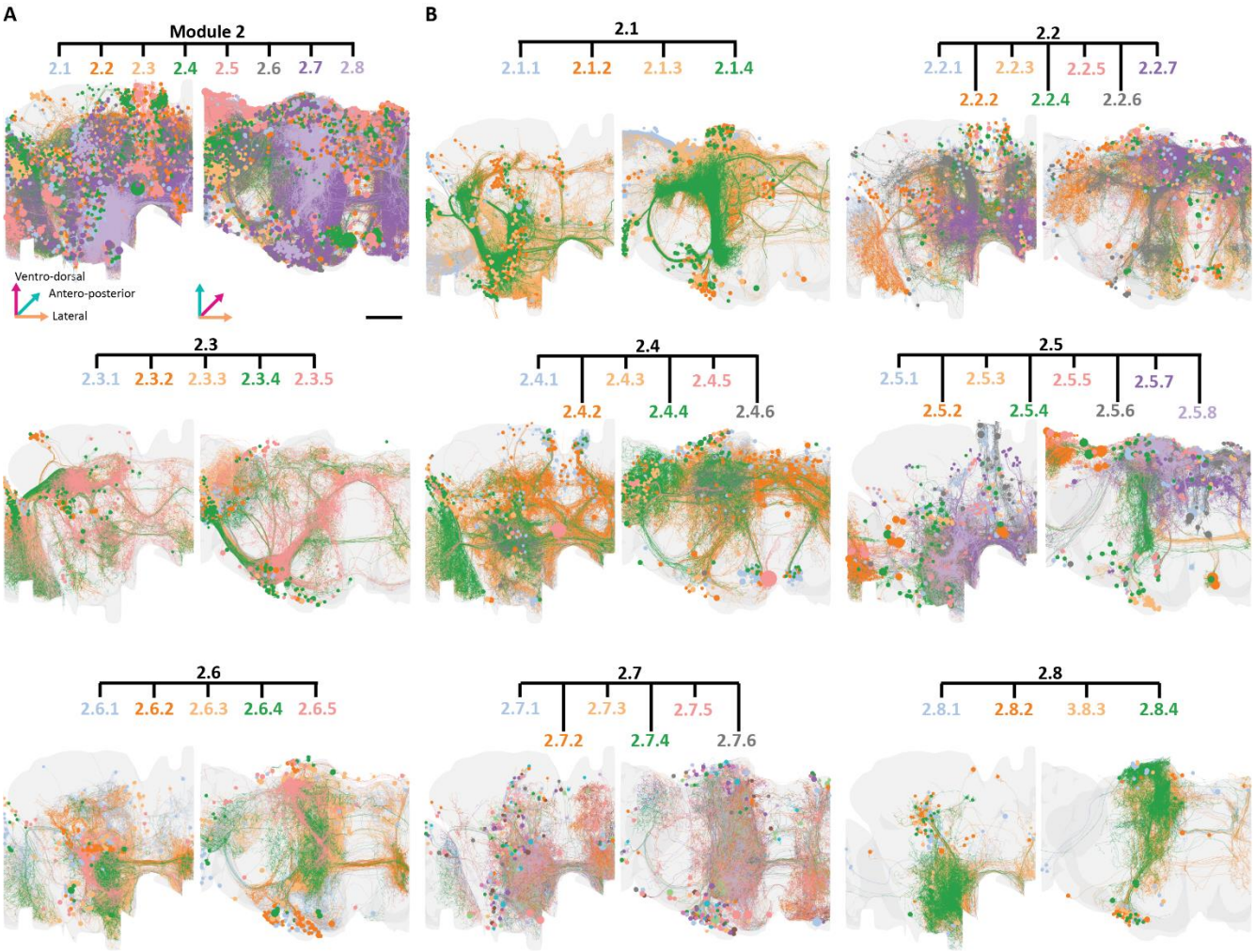
Supplementary Figure 1. Weighted degree distributions for the Drosophila hemibrain EM dataset network.



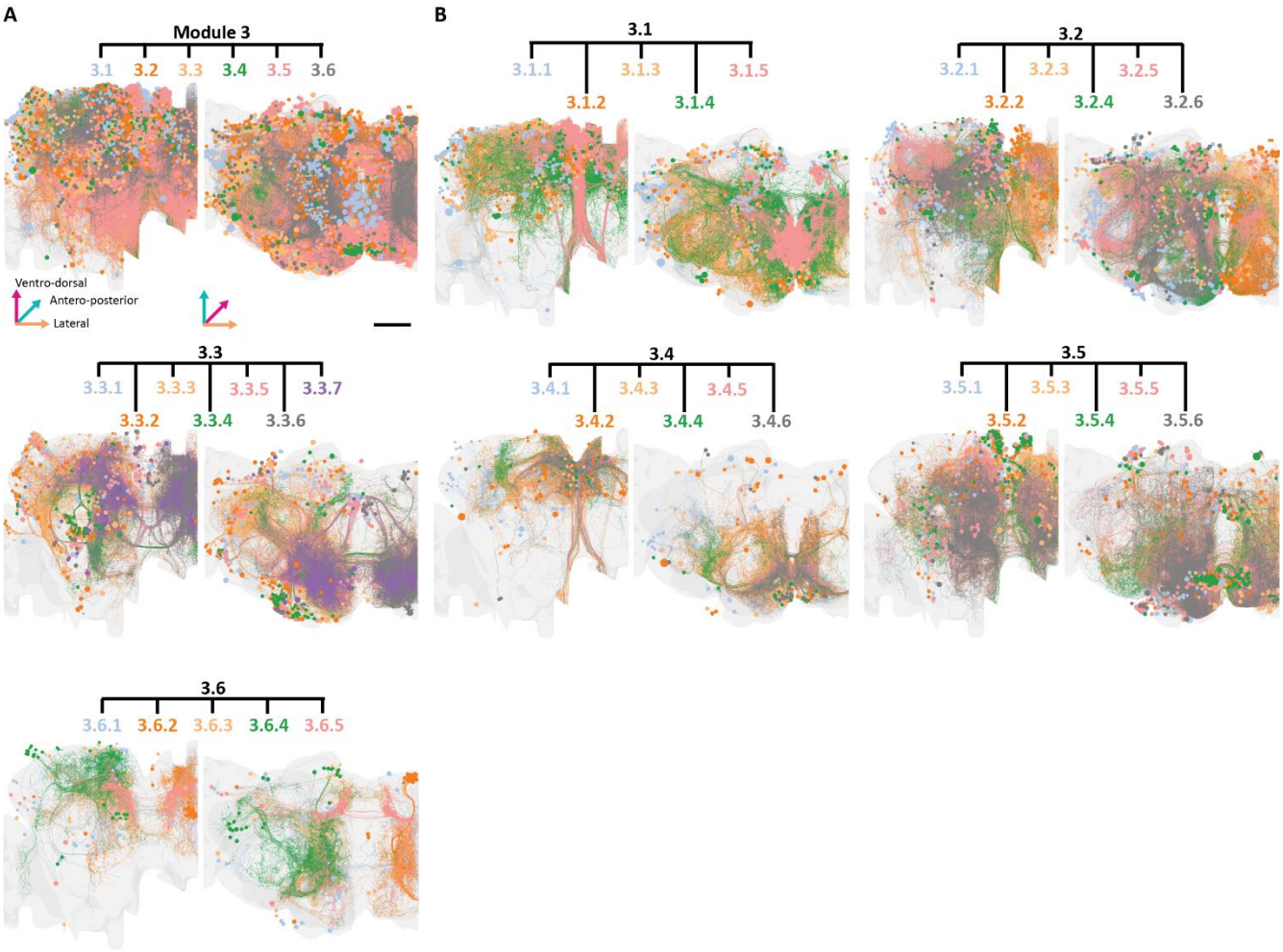
Supplementary Figure 2 . *Drosophila melanogaster* modules hierarchy. Tree diagram reporting the structure of the Drosophila melanogaster network.



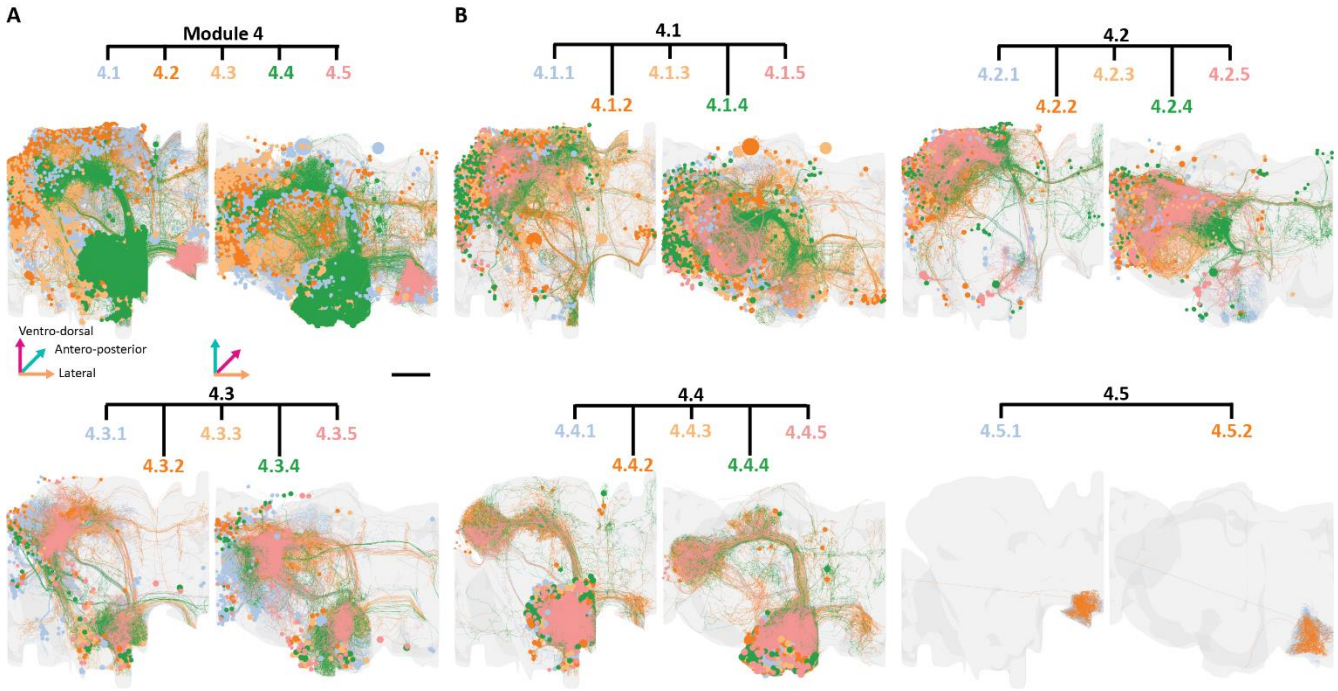
Supplementary Figure 3. **The Drosophila module 1**. A) Anatomical reconstructions of neurons assigned to the high-order module one decomposed into the lower-order modules. Colors are assigned according to the order of the lower modules. Scale bar = 50 µm.



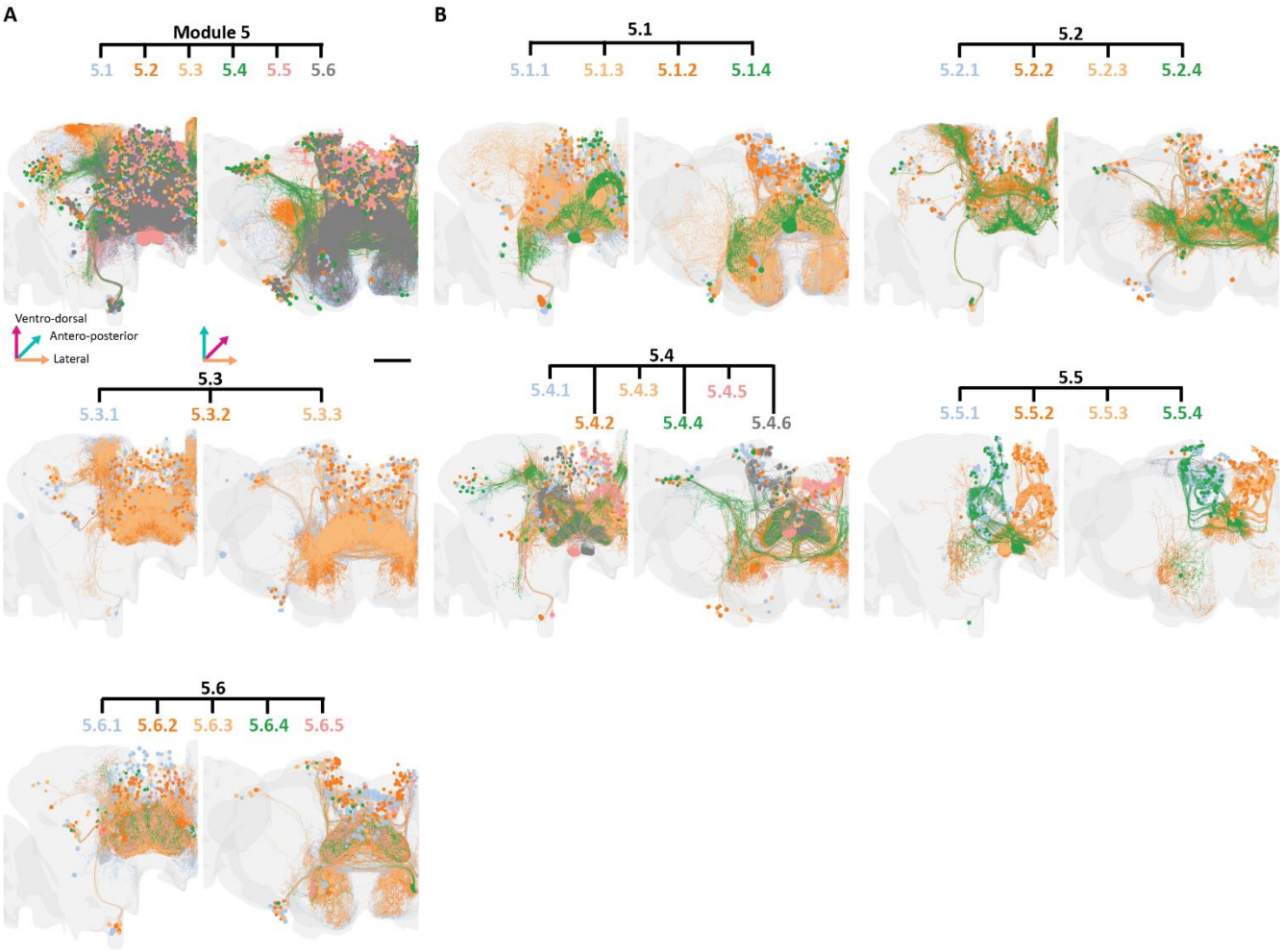
Supplementary Figure 4. **The Drosophila module 2.** A) Anatomical reconstructions of neurons assigned to the high-order module two decomposed into lower-order modules. Colors are assigned according to the order of the lower modules. Scale bar = 50 µm.



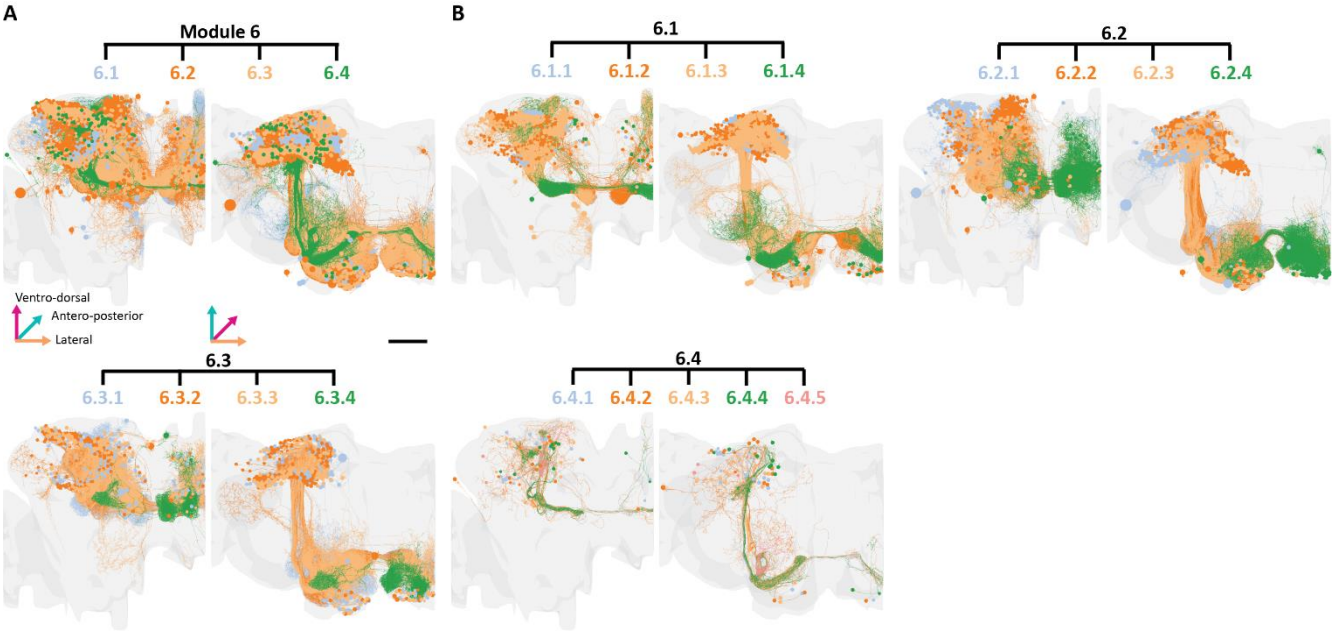
Supplementary Figure 5. **The Drosophila module 3.** A) Anatomical reconstructions of neurons assigned to the high-order module three decomposed into the lower-order modules. Colors are assigned according to the order of the lower modules. Scale bar = 50 µm.



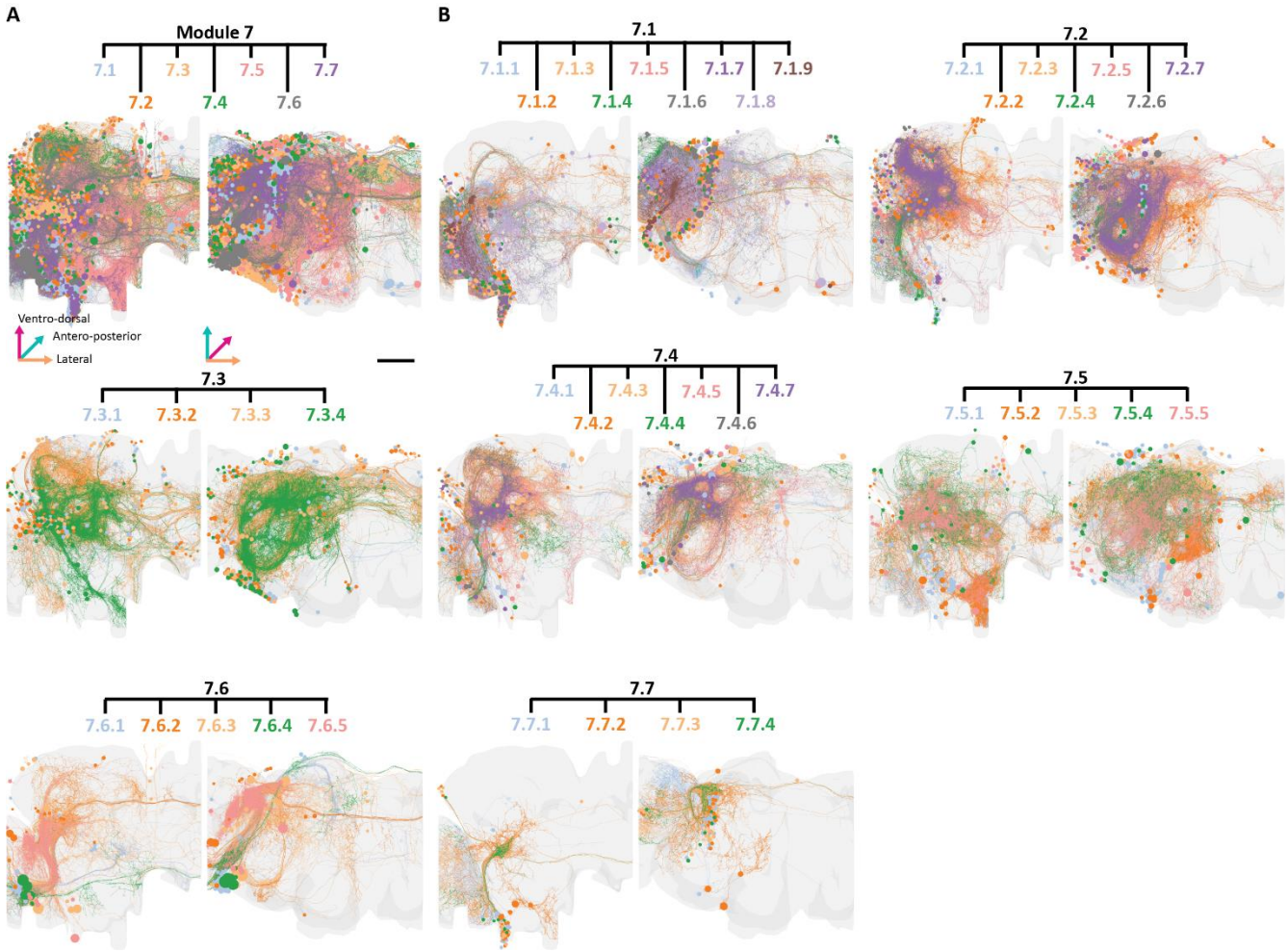
Supplementary Figure 6. **The Drosophila module 4.** A) Anatomical reconstructions of neurons assigned to the high-order module four decomposed into the lower-order modules. Colors are assigned according to the order of the lower modules. Scale bar = 50 µm.



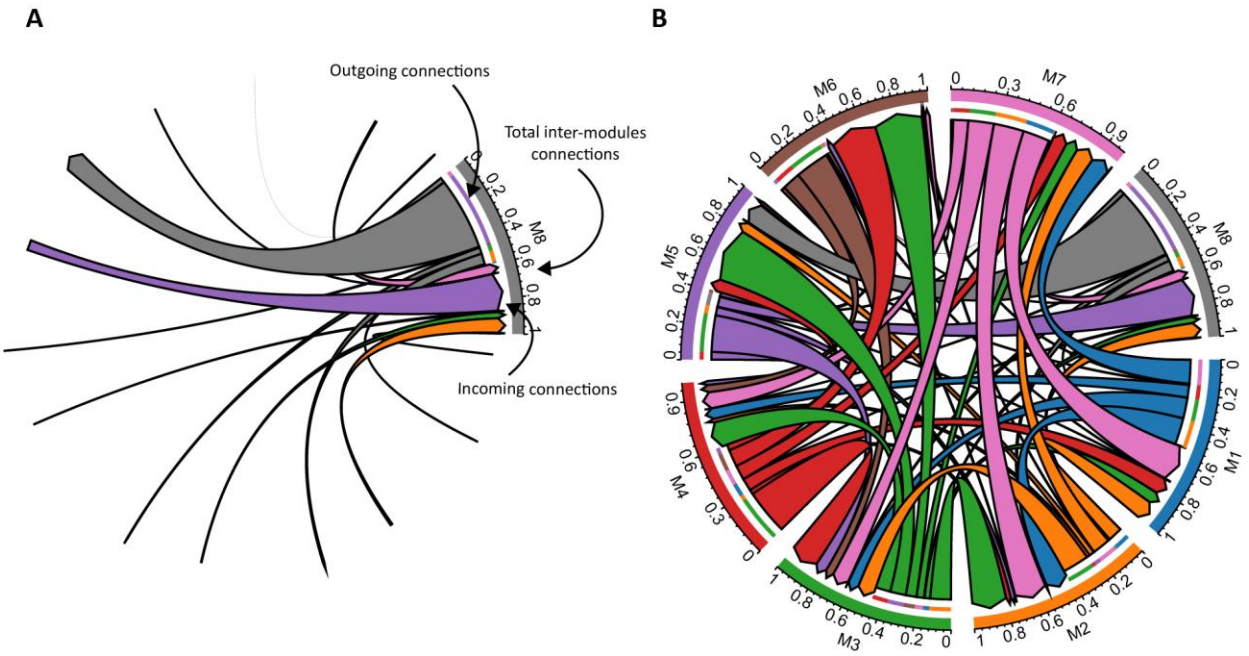
Supplementary Figure 7. **The Drosophila module 5.** A) Anatomical reconstructions of neurons assigned to the high-order module five decomposed into the lower-order modules. Colors are assigned according to the order of the lower modules. Scale bar = 50 µm.



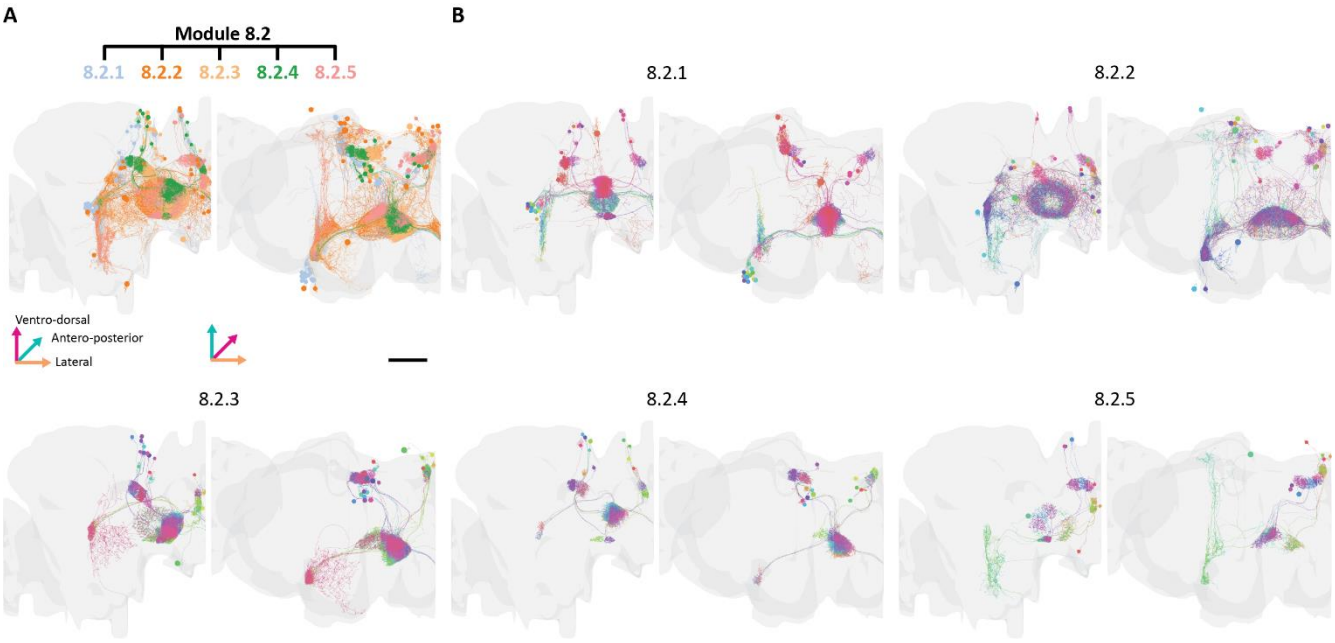
Supplementary Figure 8. **The Drosophila module 6.** A) Anatomical reconstructions of neurons assigned to the high-order module six decomposed into the lower-order modules. Colors are assigned according to the order of the lower modules. Scale bar = 50 μm.



Supplementary Figure 9. **The Drosophila module 7.** A) Anatomical reconstructions of neurons assigned to the high-order module seven decomposed into the lower-order modules. Colors are assigned according to the order of the lower modules. Scale bar = 50 µm.

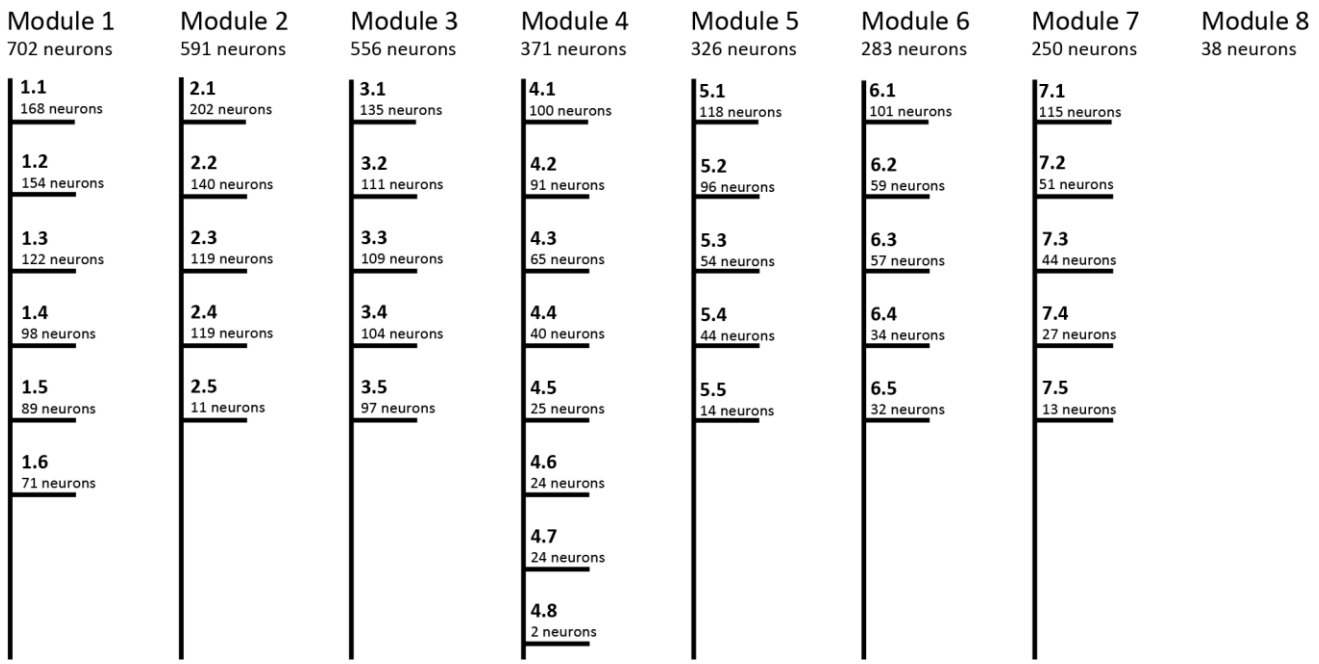


Supplementary Figure 10. Relationships between Drosophila higher level modules. A) Distribution of inter-modules connections of the higher-level module 8 of the Drosophila connectome. **B)** Distribution of connections between the higher-level modules of the Drosophila connectome. M: module.

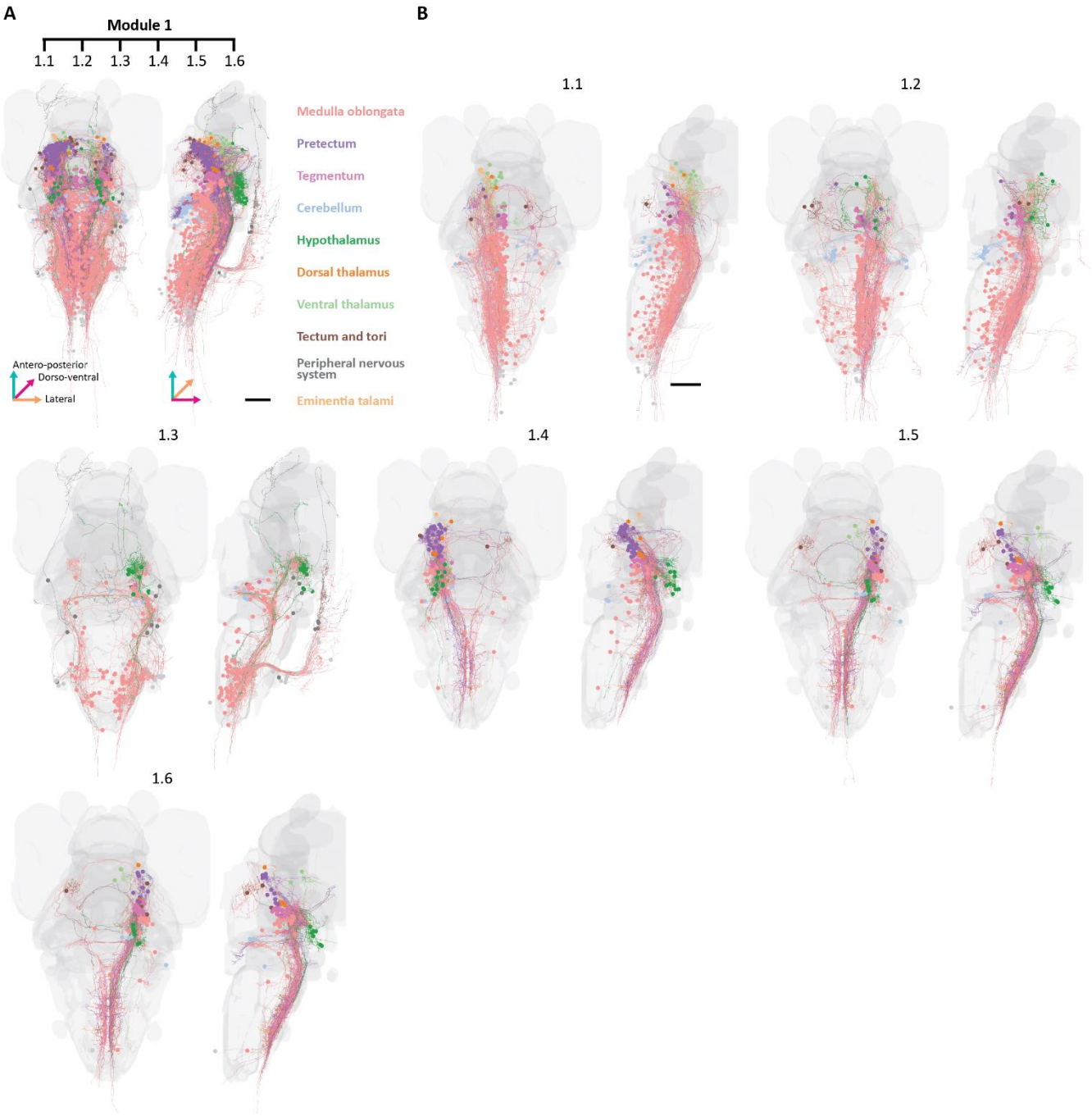


Supplementary Figure 11. Decomposition of the lower-order Drosophila module. A) Anatomical reconstructions of neurons assigned to one of the submodules belonging to module 8 subdivided by the five lower-order modules found. Colors are assigned according to the lower-order modules. Scale bar = 50 µm. **B)** Anatomical reconstructions of the neurons assigned to the five different lower-order modules shown in A. Colors are assigned randomly.

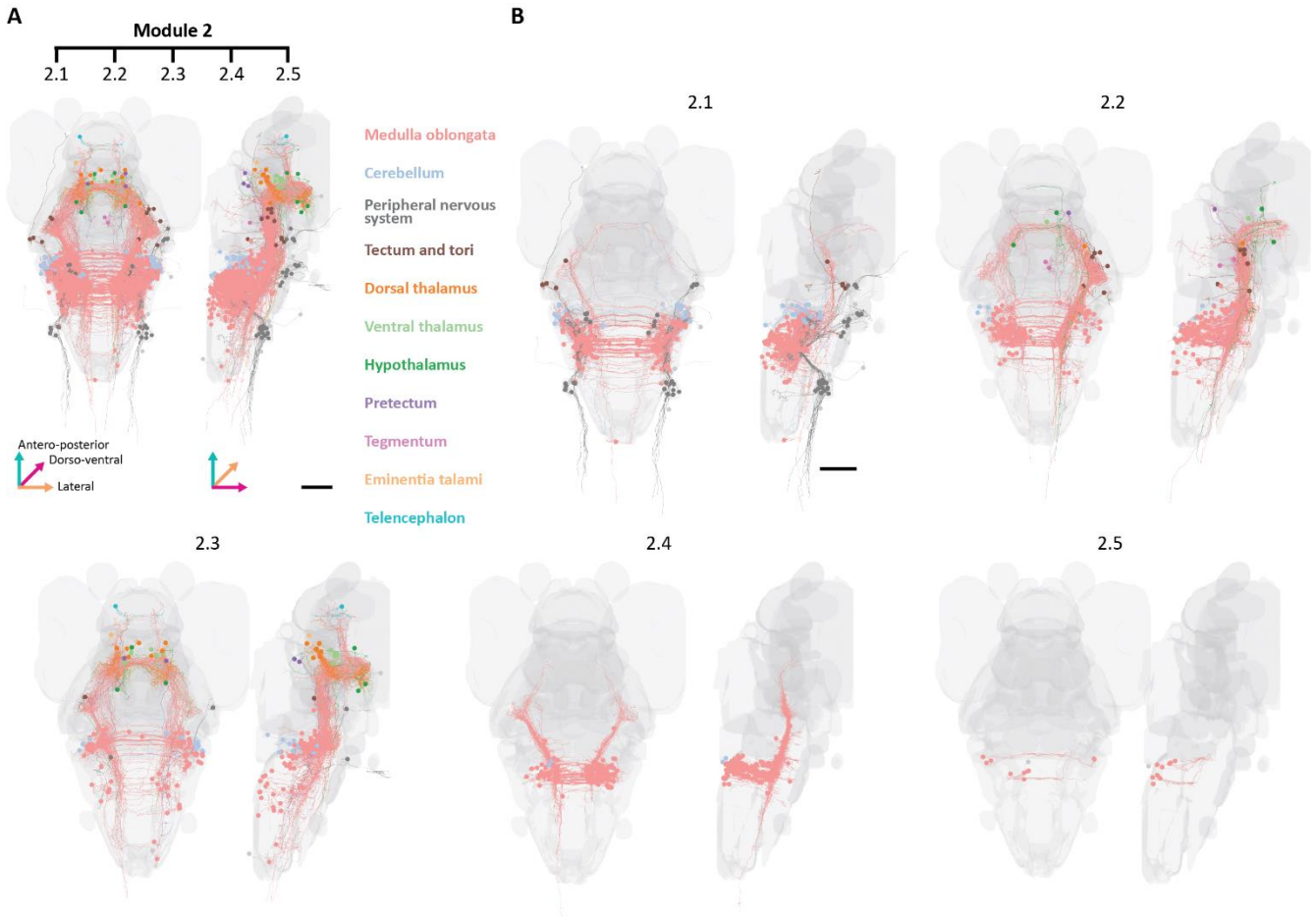
Supplementary Figure 12. Distribution of endpoints of neurons. A) Typically the distance between the endpoints of a neuron and the soma is distributed accordingly to two main distribution classes, following the Galtung classification[49]. **B**) Bimodal class S. The antimode of the distribution is used to set a threshold (dashed line) to distinguish between dendrites and axon. C) Bimodal class U. The antimode of the distribution is used to set a threshold (dashed line) to distinguish between dendrites and axon.



Supplementary Figure 13. Zebrafish modules hierarchy. Tree diagram reporting the structure of the zebrafish network.

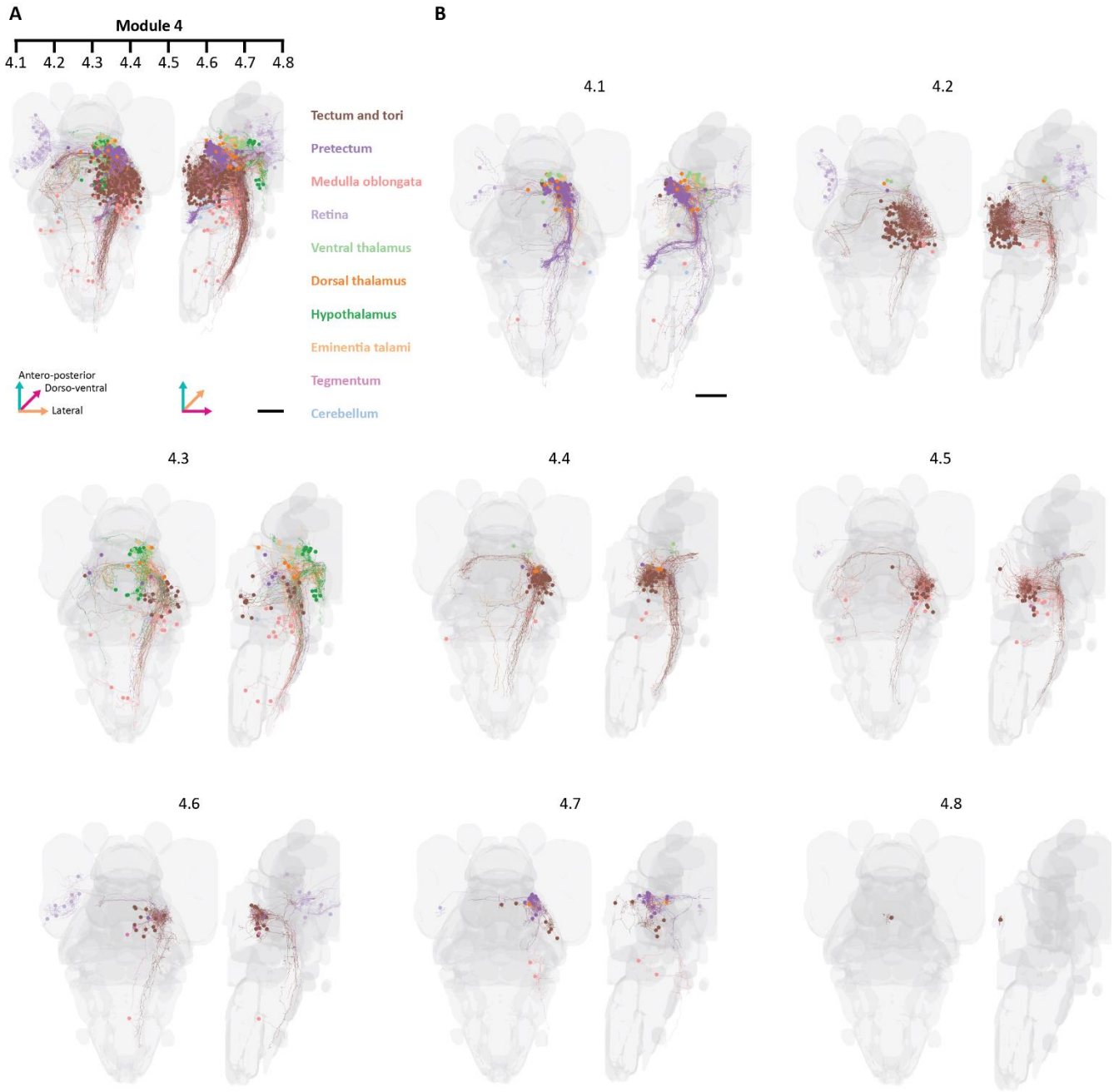


Supplementary Figure 14. The zebrafish module 1. A) Anatomical representation of the high-order module U1. Colors are assigned according to the soma distribution in the brain regions. Scale bar = 100 µm. **B)** Anatomical representation of the individual lower-order modules. Colors are assigned according to the location of the soma. Scale bar = 100 µm.

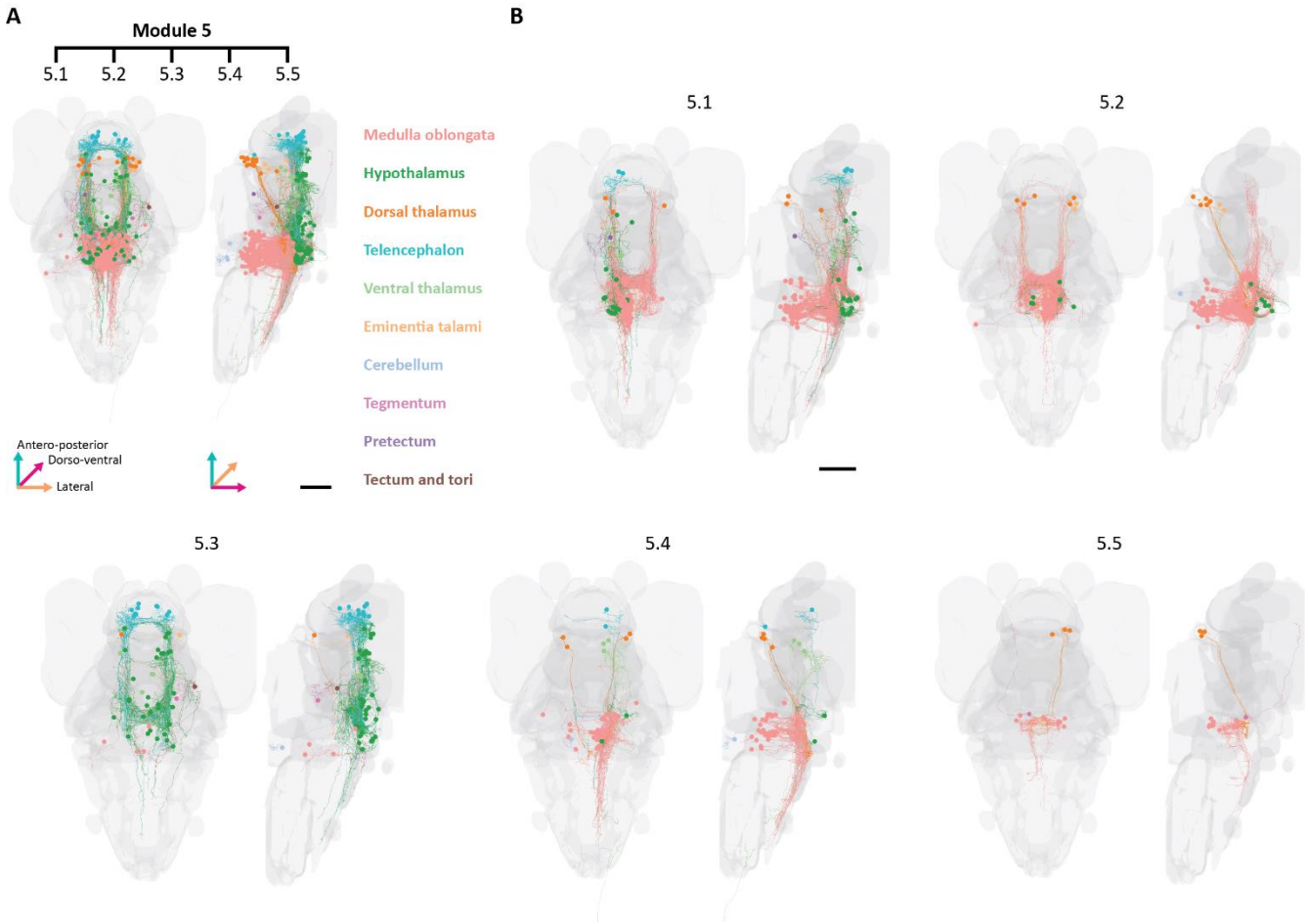


506

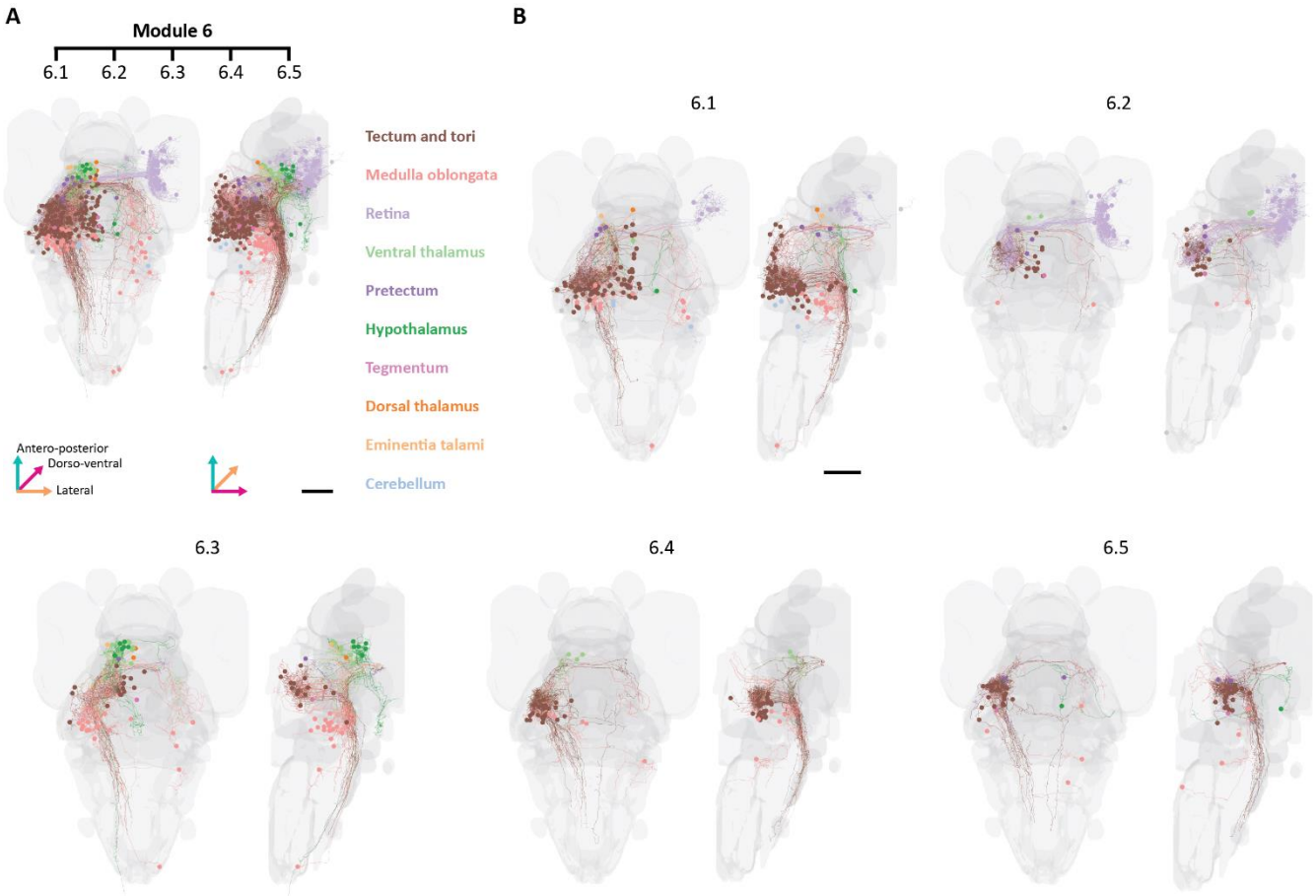
507 **Supplementary Figure 15. The zebrafish module 2. A)** Anatomical representation of the high-order module U2. Colors are
508 assigned according to the soma distribution in the brain regions. Scale bar = 100 µm. **B)** Anatomical representation of the
509 individual lower-order modules. Colors are assigned according to the location of the soma. Scale bar = 100 µm.



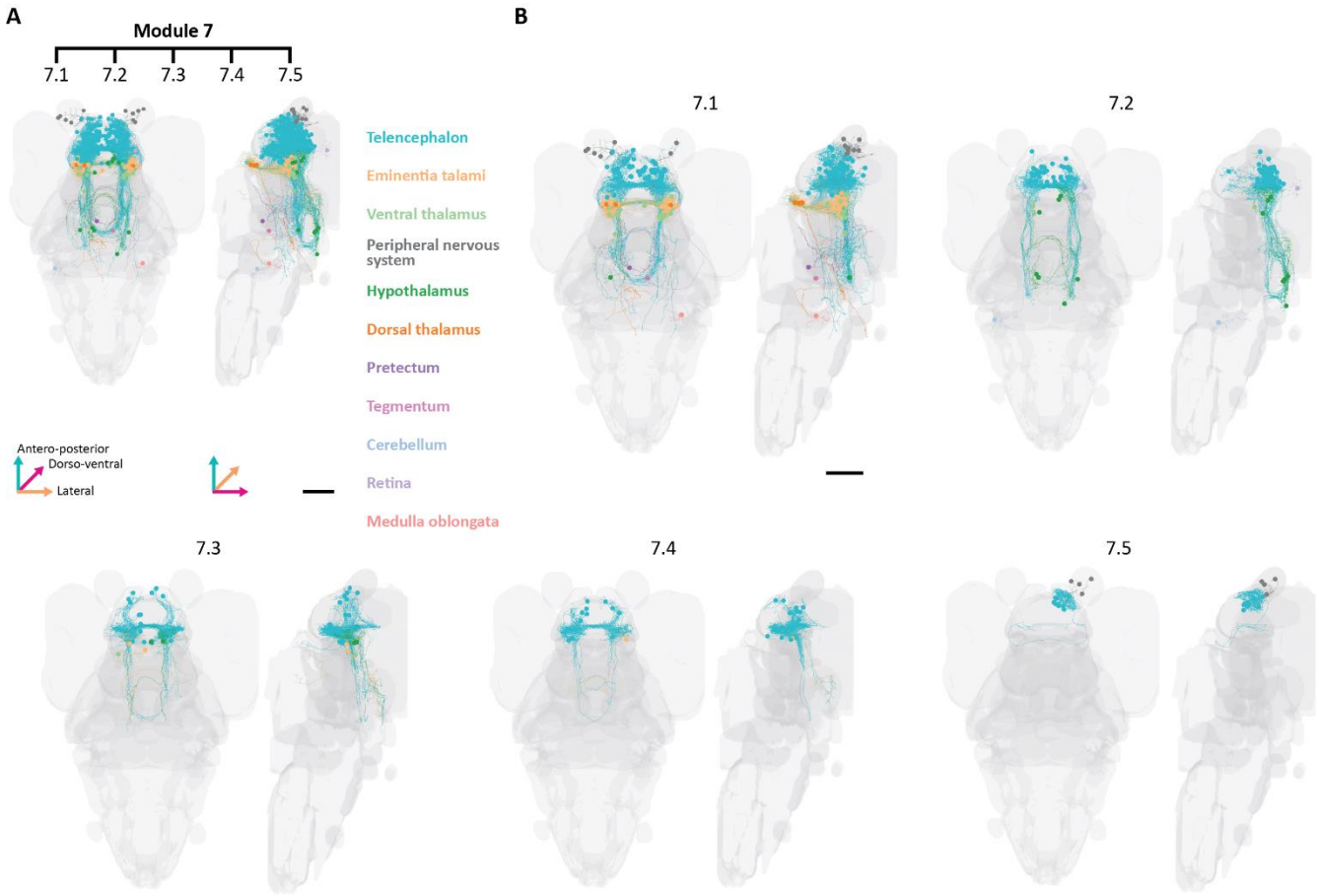
Supplementary Figure 16. The zebrafish module 4. A) Anatomical representation of the high-order module U4. Colors are assigned according to the soma distribution in the brain regions. Scale bar = 100 µm. **B)** Anatomical representation of the individual lower-order modules. Colors are assigned according to the location of the soma. Scale bar = 100 µm.



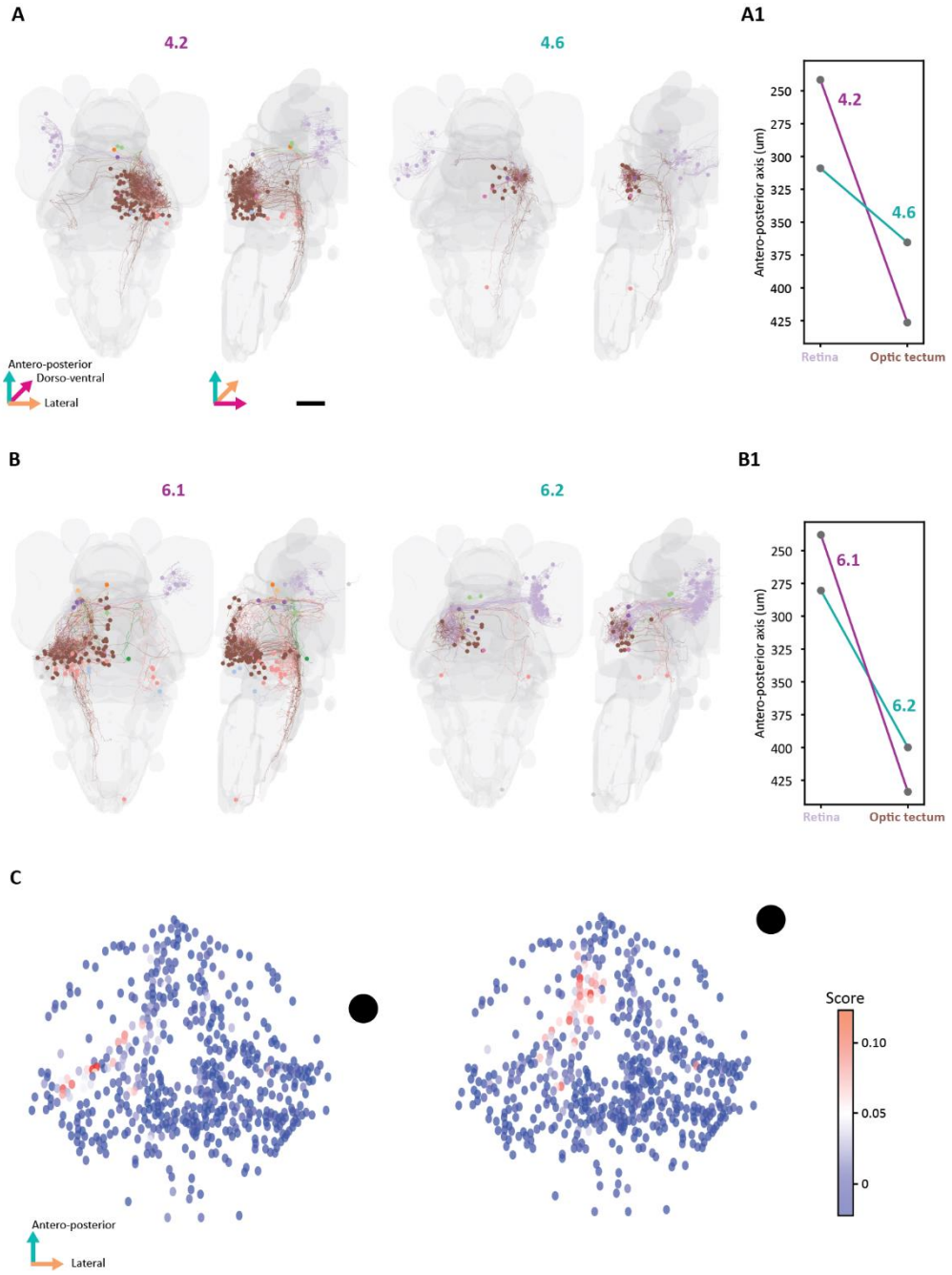
Supplementary Figure 17. The zebrafish module 5. A) Anatomical representation of the high-order module U5. Colors are assigned according to the soma distribution in the brain regions. Scale bar = 100 µm. **B)** Anatomical representation of the individual lower-order modules. Colors are assigned according to the location of the soma. Scale bar = 100 µm.



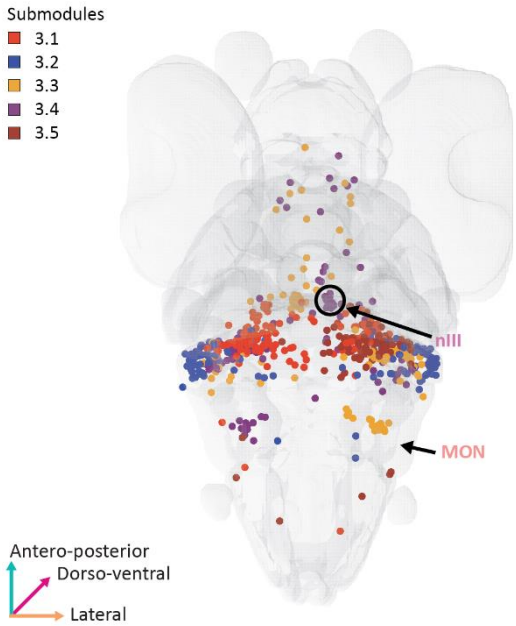
Supplementary Figure 18. The zebrafish module 6. A) Anatomical representation of the high-order module U6. Colors are assigned according to the soma distribution in the brain regions. Scale bar = 100 µm. **B)** Anatomical representation of the individual lower-order modules. Colors are assigned according to the location of the soma. Scale bar = 100 µm.



Supplementary Figure 19. The zebrafish module 7. A) Anatomical representation of the high-order module U7. Colors are assigned according to the soma distribution in the brain regions. Scale bar = 100 µm. **B)** Anatomical representation of the individual lower-order modules. Colors are assigned according to the location of the soma. Scale bar = 100 µm.



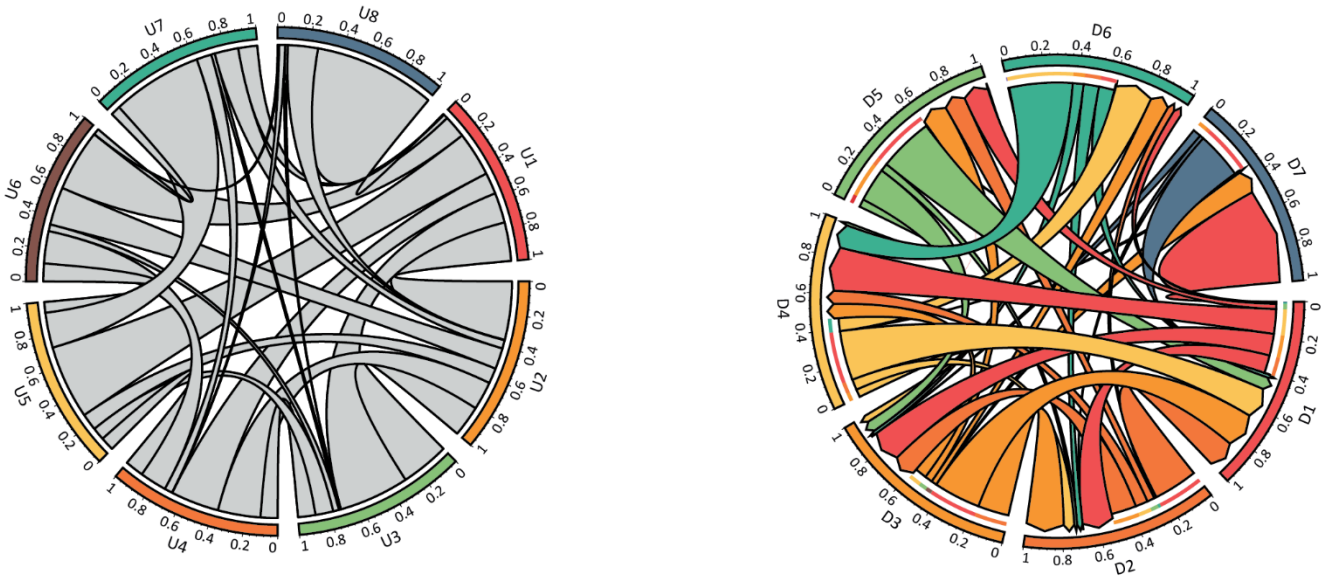
Supplementary Figure 20. **Retinotopy in optic tectum modules. A)** Sub-modules associated with the right Tectum opticum involving Peri Ventricular Neurons and Retinal Ganglion Cells, highlighting two different sub-populations of neurons (4.2 caudal tectum, 4.6 medial tectum). **A1)** Median of the distributions along the antero-posterior axis of retina and optic tectum neurons of modules 4.2 (magenta) and 4.6 (teal). **B)** Sub-modules associated with the left Tectum opticum involving Peri Ventricular Neurons and Retinal Ganglion Cells, highlighting two different sub-populations of neurons (6.1 caudal tectum, 6.2 medial tectum). **B1)** Median of the distributions along the antero-posterior axis of retina and optic tectum neurons of modules 6.1 (magenta) and 6.2 (teal). **C)** Functional correlate of the retinotopic organization: different positions of the stimulus (black circle) elicit activity with different spatial distribution of stimulus-responsive cells in the contralateral tectum (red cells).



541

542 **Supplementary Figure 21. Submodules of the zebrafish dataset cerebellar module.** Anatomical representation
543 of the high-order module U3 as composed by the five lower-order modules (3.1-3.5). Colors are assigned
544 according to the submodules. MON: medial octavolateral nucleus, nIII: oculomotor nucleus.

545



Supplementary Figure 22. Relationships between zebrafish modules. A) Distribution of connections between the modules of the undirected network. U: undirected modules. B) Distribution of connections between the modules of the directed network. D: directed modules.

552 **REFERENCES**

553 1. Brunner, C. *et al.* Whole-brain functional ultrasound imaging in awake head-fixed mice. *Nat Protoc* **16**,
554 3547–3571 (2021).
555 2. Ahrens, M. B., Orger, M. B., Robson, D. N., Li, J. M. & Keller, P. J. Whole-brain functional imaging at cellular
556 resolution using light-sheet microscopy. *Nat Methods* **10**, 413–420 (2013).
557 3. Bruzzone, M. *et al.* Whole brain functional recordings at cellular resolution in zebrafish larvae with 3D
558 scanning multiphoton microscopy. *Sci Rep* **11**, 11048 (2021).
559 4. Schrödel, T., Prevedel, R., Aumayr, K., Zimmer, M. & Vaziri, A. Brain-wide 3D imaging of neuronal activity in
560 Caenorhabditis elegans with sculpted light. *Nat Methods* **10**, 1013–1020 (2013).
561 5. Brondi, M., Bruzzone, M., Lodovichi, C. & Dal Maschio, M. Optogenetic Methods to Investigate Brain
562 Alterations in Preclinical Models. *Cells* **11**, 1848 (2022).
563 6. Ahrens, M. B. *et al.* Brain-wide neuronal dynamics during motor adaptation in zebrafish. *Nature* **485**, 471–
564 477 (2012).
565 7. Chen, X. *et al.* Brain-wide Organization of Neuronal Activity and Convergent Sensorimotor Transformations
566 in Larval Zebrafish. *Neuron* **100**, 876-890.e5 (2018).
567 8. Ahrens, M. B. & Engert, F. Large-scale imaging in small brains. *Curr Opin Neurobiol* **32**, 78–86 (2015).
568 9. Naumann, E. A. *et al.* From Whole-Brain Data to Functional Circuit Models: The Zebrafish Optomotor
569 Response. *Cell* **167**, 947-960.e20 (2016).
570 10. Mariani, B. *et al.* Disentangling the critical signatures of neural activity. *Sci Rep* **12**, 10770 (2022).
571 11. Barzon, G., Nicoletti, G., Mariani, B., Formentin, M. & Suweis, S. Criticality and network structure drive
572 emergent oscillations in a stochastic whole-brain model. *J. Phys. Complex.* **3**, 025010 (2022).
573 12. Panzeri, S., Harvey, C. D., Piasini, E., Latham, P. E. & Fellin, T. Cracking the Neural Code for Sensory
574 Perception by Combining Statistics, Intervention, and Behavior. *Neuron* **93**, 491–507 (2017).

575 13. Schaeffer, R., Khona, M. & Fiete, I. R. *No Free Lunch from Deep Learning in Neuroscience: A Case Study*
576 *through Models of the Entorhinal-Hippocampal Circuit*.
577 http://biorxiv.org/lookup/doi/10.1101/2022.08.07.503109 (2022) doi:10.1101/2022.08.07.503109.
578 14. Vanwalleghem, G. C., Ahrens, M. B. & Scott, E. K. Integrative whole-brain neuroscience in larval zebrafish.
579 *Current Opinion in Neurobiology* **50**, 136–145 (2018).
580 15. Beltramo, R. *et al.* Layer-specific excitatory circuits differentially control recurrent network dynamics in the
581 neocortex. *Nat Neurosci* **16**, 227–234 (2013).
582 16. Rajasethupathy, P., Ferenczi, E. & Deisseroth, K. Targeting Neural Circuits. *Cell* **165**, 524–534 (2016).
583 17. Dal Maschio, M., Donovan, J. C., Helmbrecht, T. O. & Baier, H. Linking Neurons to Network Function and
584 Behavior by Two-Photon Holographic Optogenetics and Volumetric Imaging. *Neuron* **94**, 774-789.e5
585 (2017).
586 18. D, F., M, D. M., E, L. & H, B. An optogenetic toolbox for unbiased discovery of functionally connected cells
587 in neural circuits. *Nature communications* vol. 8 https://pubmed.ncbi.nlm.nih.gov/28740141/ (2017).
588 19. Helmbrecht, T. O., Dal Maschio, M., Donovan, J. C., Koutsouli, S. & Baier, H. Topography of a Visuomotor
589 Transformation. *Neuron* **100**, 1429-1445.e4 (2018).
590 20. Wanner, A. A. & Friedrich, R. W. Whitening of odor representations by the wiring diagram of the olfactory
591 bulb. *Nat Neurosci* **23**, 433–442 (2020).
592 21. Wanner, A. A., Genoud, C., Masudi, T., Siksou, L. & Friedrich, R. W. Dense EM-based reconstruction of the
593 interglomerular projectome in the zebrafish olfactory bulb. *Nat Neurosci* **19**, 816–825 (2016).
594 22. Svara, F. *et al.* Automated synapse-level reconstruction of neural circuits in the larval zebrafish brain. *Nat*
595 *Methods* **19**, 1357–1366 (2022).
596 23. White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of the
597 nematode Caenorhabditis elegans. *Philos Trans R Soc Lond B Biol Sci* **314**, 1–340 (1986).
598 24. Takemura, S. *et al.* The comprehensive connectome of a neural substrate for 'ON' motion detection in
599 Drosophila. *eLife* **6**, e24394 (2017).

600 25. Zheng, Z. *et al.* A Complete Electron Microscopy Volume of the Brain of Adult Drosophila melanogaster.
601 *Cell* **174**, 730-743.e22 (2018).
602 26. Schubert, P. J. *et al.* SyConn2: dense synaptic connectivity inference for volume electron microscopy. *Nat*
603 *Methods* **19**, 1367–1370 (2022).
604 27. Scheffer, L. K. *et al.* A connectome and analysis of the adult Drosophila central brain. *Elife* **9**, e57443
605 (2020).
606 28. Kunst, M. *et al.* A Cellular-Resolution Atlas of the Larval Zebrafish Brain. *Neuron* **103**, 21-38.e5 (2019).
607 29. Newman, M. E. J. & Girvan, M. Finding and evaluating community structure in networks. *Phys. Rev. E* **69**,
608 026113 (2004).
609 30. Newman, M. & Newman, M. *Networks*. (Oxford University Press, 2018).
610 31. Fortunato, S. Community detection in graphs. *Physics Reports* **486**, 75–174 (2010).
611 32. Kim, J. S. & Kaiser, M. From Caenorhabditis elegans to the human connectome: a specific modular
612 organization increases metabolic, functional and developmental efficiency. *Philos Trans R Soc Lond B Biol*
613 *Sci* **369**, 20130529 (2014).
614 33. Moreira, C. A. & de Aguiar, M. A. M. Modular structure in C. elegans neural network and its response to
615 external localized stimuli. *Physica A: Statistical Mechanics and its Applications* **533**, 122051 (2019).
616 34. Sporns, O. Graph theory methods: applications in brain networks. *Dialogues Clin Neurosci* **20**, 111–121
617 (2018).
618 35. Meunier, D., Lambiotte, R., Fornito, A., Ersche, K. & Bullmore, E. Hierarchical modularity in human brain
619 functional networks. *Frontiers in Neuroinformatics* **3**, (2009).
620 36. Vértes, P. E. *et al.* Simple models of human brain functional networks. *Proceedings of the National*
621 *Academy of Sciences* **109**, 5868–5873 (2012).
622 37. Garcia, J. O., Ashourvan, A., Muldoon, S. F., Vettel, J. M. & Bassett, D. S. Applications of community
623 detection techniques to brain graphs: Algorithmic considerations and implications for neural function. *Proc*
624 *IEEE Inst Electr Electron Eng* **106**, 846–867 (2018).

38. Vishwanathan, A. *et al.* Predicting modular functions and neural coding of behavior from a synaptic wiring diagram. 2020.10.28.359620 Preprint at https://doi.org/10.1101/2020.10.28.359620 (2022).
39. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep* **9**, 5233 (2019).
40. Omoto, J. J. *et al.* Visual Input to the Drosophila Central Complex by Developmentally and Functionally Distinct Neuronal Populations. *Current Biology* **27**, 1098–1110 (2017).
41. 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).
42. Kandimalla, P., Omoto, J. J., Hong, E. J. & Hartenstein, V. Lineages to circuits: the developmental and evolutionary architecture of information channels into the central complex. *J Comp Physiol A* (2023) doi:10.1007/s00359-023-01616-y.
43. Hulse, B. K. & Jayaraman, V. Mechanisms Underlying the Neural Computation of Head Direction. *Annual Review of Neuroscience* **43**, 31–54 (2020).
44. Ohyama, T. *et al.* A multilevel multimodal circuit enhances action selection in Drosophila. *Nature* **520**, 633–639 (2015).
45. Randel, N. *et al.* Neuronal connectome of a sensory-motor circuit for visual navigation. *eLife* **3**, e02730 (2014).
46. Ryan, K., Lu, Z. & Meinertzhagen, I. A. The CNS connectome of a tadpole larva of Ciona intestinalis (L.) highlights sidedness in the brain of a chordate sibling. *eLife* **5**, e16962 (2016).
47. Consortium, T. Mic. *et al.* Functional connectomics spanning multiple areas of mouse visual cortex. 2021.07.28.454025 Preprint at https://doi.org/10.1101/2021.07.28.454025 (2023).
48. The connectome of an insect brain | Science. https://www.science.org/doi/10.1126/science.add9330.
49. Theory and Methods of Social Research. - Galtung, Johan: 9780231030885 - AbeBooks. https://www.abebooks.it/9780231030885/Theory-Methods-Social-Research-Galtung-0231030886/plp.

50. Bianco, I. H. *et al.* The tangential nucleus controls a gravito-inertial vestibulo-ocular reflex. *Curr Biol* **22**, 1285–1295 (2012).
51. Functional regionalization of the teleost cerebellum analyzed in vivo | PNAS. https://www.pnas.org/doi/10.1073/pnas.1403105111.
52. Favre-Bulle, I. A., Vanwalleghem, G., Taylor, M. A., Rubinsztein-Dunlop, H. & Scott, E. K. Cellular-Resolution Imaging of Vestibular Processing across the Larval Zebrafish Brain. *Current Biology* **28**, 3711-3722.e3 (2018).
53. Bae, Y.-K. *et al.* Anatomy of zebrafish cerebellum and screen for mutations affecting its development. *Dev Biol* **330**, 406–426 (2009).
54. Cannon, R. C., Turner, D. A., Pyapali, G. K. & Wheal, H. V. An on-line archive of reconstructed hippocampal neurons. *Journal of Neuroscience Methods* **84**, 49–54 (1998).
55. Clements, J. *et al.* neuPrint: Analysis Tools for EM Connectomics. 2020.01.16.909465 Preprint at https://doi.org/10.1101/2020.01.16.909465 (2020).
56. Scott, E. K. *et al.* Targeting neural circuitry in zebrafish using GAL4 enhancer trapping. *Nat Methods* **4**, 323–326 (2007).
57. Hagberg, A. A., Schult, D. A. & Swart, P. J. Exploring Network Structure, Dynamics, and Function using NetworkX. (2008).
58. Stepanyants, A. & Chklovskii, D. B. Neurogeometry and potential synaptic connectivity. *Trends Neurosci* **28**, 387–394 (2005).
59. *Network Science by Albert-László Barabási*.
60. Preface. in *Networks: An Introduction* (ed. Newman, M.) 0 (Oxford University Press, 2010). doi:10.1093/acprof:oso/9780199206650.002.0005.
61. Leicht, E. A. & Newman, M. E. J. Community Structure in Directed Networks. *Phys. Rev. Lett.* **100**, 118703 (2008).
62. Reichardt, J. & Bornholdt, S. Statistical mechanics of community detection. *Phys. Rev. E* **74**, 016110 (2006).