# A method for extracting an approximated connectome from libraries of single cell reconstructions

Manjunatha K.K.H.[2,4#], 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
[4,] Modelling and Engineering Risk and Complexity, Scuola Superiore Meridionale, Napoli, IT

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

Lead contact: marco.dalmaschio@unipd.it

**Abstract**

Understanding how the neuronal circuit organization supports the spatio-temporal patterns that characterize the brain's neuronal activity is still an open challenge. Despite a large number of approaches available to record and modulate neuronal activity at cell resolution and in living animals, limited data are available to map the functional information into a circuit wiring diagram supporting possible circuit working mechanisms. When available, the analysis of electron microscopy based high-resolution connectomes, leveraging synapse annotation, allows the unraveling of portions of the nervous system wiring diagram or specific circuit motives. However, even assuming a complete annotation of the synapses, extracting the general organization principles of the neuronal networks across the brain remains a challenging effort. In order to extend the available methods, we present an approach to reconstruct an approximated brain connectome starting from libraries of single cell reconstructions belonging to or co-registered in the same anatomical space. By leveraging both the Strahler numbering of the nodes characterizing the cell morphological reconstructions and a proximity range criterion, we inferred the general connectivity structures between the different cells, bypassing the need for synapse annotation. We applied this approach to extract an approximated connectome of the zebrafish larvae brain from a light microscopy-based dataset of about 3-thousand co-registered neuronal skeletonizations. Modularity analysis of the retrieved connectome provided a representation of the resulting graph organized in hierarchical structures, with neuronal modules capturing precise and

topographically organized connection patterns mirroring identified functional circuit motives. In conclusion, we present a scalable, from-circuit-to-brain range approach amenable to revealing the neuronal architectures supporting brain mechanisms.

## Introduction

The understanding of the brain mechanisms takes advantage of the continuous refinement of molecular, hardware, and software approaches. These methods allow reconstructing, in suitable model organisms, the brain dynamics 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 characterizing brain activity[6–8]. The obtained dataset becomes 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, perform similarly well in capturing the same circuit dynamics[12,13]. With the aim to better constrain this kind of problem, one possible strategy is offered by the availability of the anatomical information characterizing the circuit wiring diagram supporting a certain functional mechanism. Then, it is convenient to try to refine a circuit working model combining functional aspects to the wiring properties of the underlying circuit. For this purpose, 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, supervised or partially-supervised cell tracing, a cell segmentation process, annotation of the synaptic contacts, and lastly a thorough control of the precision of the reconstructions. Today, even with the support of AI-based algorithms, 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]. Once the information on the neuronal connectivity schemes, i.e. the organization of the input-output contacts characterizing the

cells involved, has been retrieved, it becomes possible to reconstruct an adjacency matrix representing a connectome and to identify the circuit wiring diagrams. Because of the amount and the complexity of the data one has to face, this type of anatomical reconstruction frequently focuses on targeted circuit motifs, specific neuronal subpopulations, or particular brain regions. On the other hand, unraveling the anatomical organization of brain-wide networks requires suitable datasets, with a sufficient sampling of the cellular morphologies, coverage across different regions of the brain, and, ideally, based on the evaluation of a large number of exemplars. Currently, even if there are several reported whole-brain EM reconstructions[21,22,24,25,27–31], partial or whole-brain connectomes that provide skeletonization of neurons along with detailed annotations of their synaptic contacts are extremely limited. In order to extend the approaches currently available, here, we present a method to generate an approximated connectome starting from an anatomical dataset with neuronal cell skeletonizations, belonging to or co-registered in the same reference space, where the actual synaptic information is not accessible or not yet available. This is the typical setting of mesoscale atlases based on optical microscopy acquisitions, as well as electron microscopy datasets not yet fully annotated with the neuronal input-output features[32–35]. For this type of dataset, here, we devised an approach to infer brain network organization by assigning cell-to-cell connections, which are subsequently translated into adjacency matrices. These matrices represent connections between nodes along the reconstructed neurites. We applied this approach to a zebrafish single-cell dataset, which comprises more than three thousand neuronal reconstructions. Through this, we obtained adjacency matrices that summarize the connectivity properties of the dataset, both for directed and undirected graphs[36].

We took advantage of the calculated adjacency matrices to evaluate, along with topological features of associated graphs, a modular organization of the brain networks covered by the dataset. This analysis returned a decomposition of the connectome in several modules and submodules, reflecting known neuronal circuit structures and their functional organization..

## Results

87 **1. Building an approximated connectome for the larval zebrafish brain based on Strahler numbering**
88 **and spatial proximity**

89 Aiming to build an approximated connectome for the zebrafish larval brain, we took advantage of an
90 available dataset with anatomical reconstructions of individual neurons obtained by means of optical
91 microscopy z-stack acquisitions co-registered on the same reference brain[36] (Figure 1A). In this case,
92 where actual synaptic information is unavailable, we devised an approach to infer brain network
93 structure by assigning cell-to-cell connections on the basis of a proximity rule between the nodes along
94 the reconstructed cell skeleton (see the Methods section). For this purpose, we calculated for each
95 point of the cell skeleton (or neurite) the corresponding Strahler's number, a parameter that represents
96 the relative level of the cell skeleton points in the cell branching hierarchy[19] (Figure 1B). Reflecting
97 Peters's rule postulate[37–39], terminal points (endpoints) along the neurite structures, characterized by
98 a unitary value of the Strahler number, are considered putative points of connection between different
99 cells if they fall within a defined proximity range one from the other. For our light microscopy dataset,
100 considering the intrinsic resolution of the acquisition process, the typical errors in the anatomical
101 reconstruction, and co-registration procedures, we adopted a proximity range corresponding to 5 µm
102 (Figure 1D).

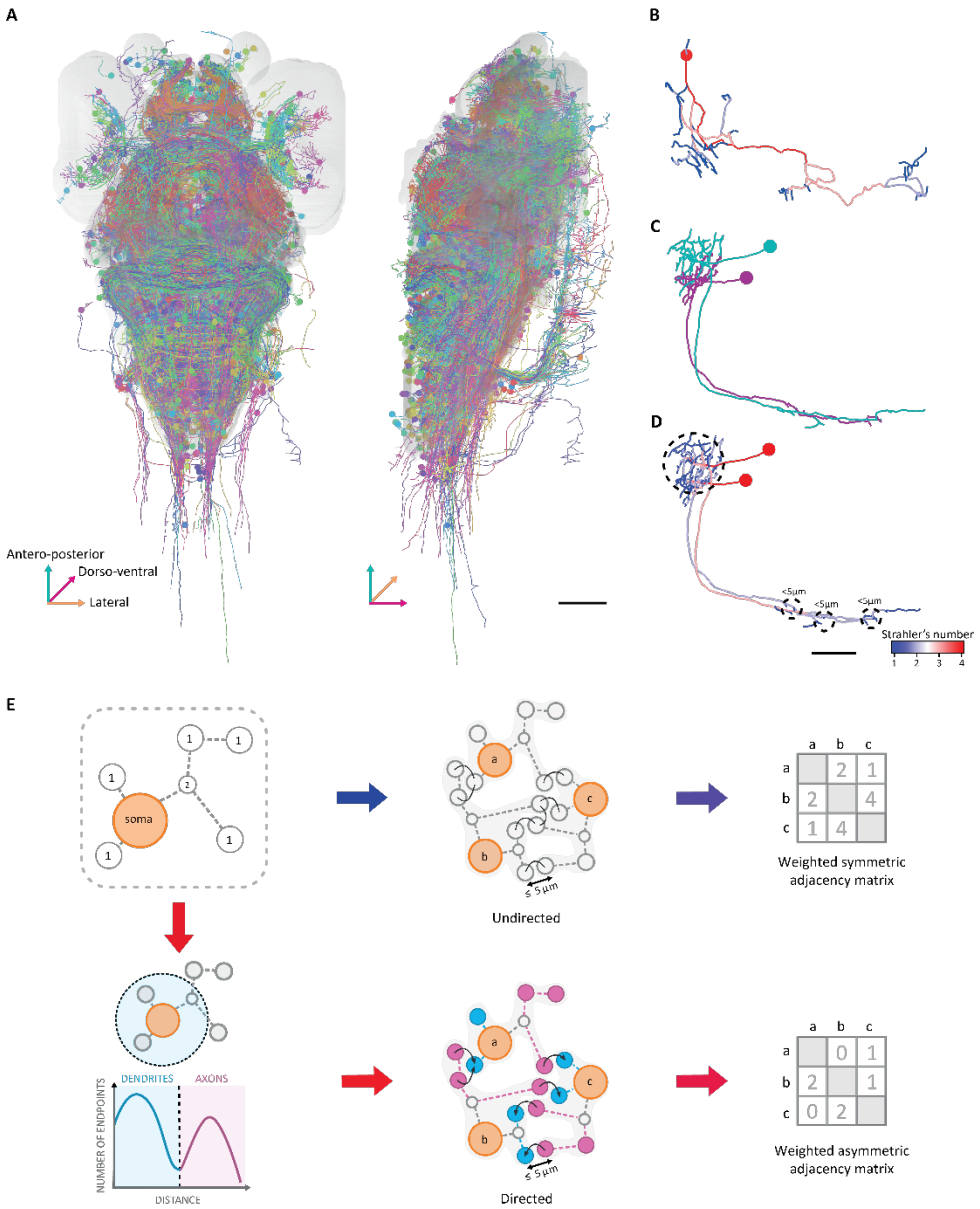



***Figure 1. Building the approximated connectome from a mesoscale cell atlas. A)*** *Anatomically reconstructed neurons from* [36] *overlaid to the zebrafish atlas. Colors are assigned randomly. Scale bar= 100 µm.* ***B)*** *Representative neuron with nodes color-coded according to the corresponding Strahler's number.* ***C)*** *Two representative neurons anatomically located one close to the other.* ***D)*** *Same neurons as in C with nodes color-coded 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 approximated connectome (blue arrows) and the directed approximated connectome (red arrows).*

As reconstructing the connectome relies on the determination of the adjacency matrix $A_{ij}$, we considered two different scenarios. In the first one, we assume no subcellular segregations for the input and output points of a neuron (or in other words, dendrites and axon terminals), resulting in a symmetric adjacency matrix $A_{ij}^{U}$ corresponding to an undirected connectome (Figure 1E, Figure 2A). In the second one, as most of the neuronal reconstructions are characterized by a bimodal spatial

distribution of the neuronal endpoints[40] (Supplementary Figure 1), we assigned for each cell a separation in two putative components: input dendritic points in the regions proximal to the soma, and output axonal points in more distal areas. This allowed us to constrain the connectome according to an axon-to-dendrite synaptic scheme. In the latter case, we obtained an asymmetric adjacency matrix $A_{ij}^{D}$ representing a directed connectome (Figure 1E, Figure 2E-H). The two connectome scenarios resulted in a rather articulated scheme of neuronal interactions between the different regions of the brain (Figure 2B, Figure 2F).

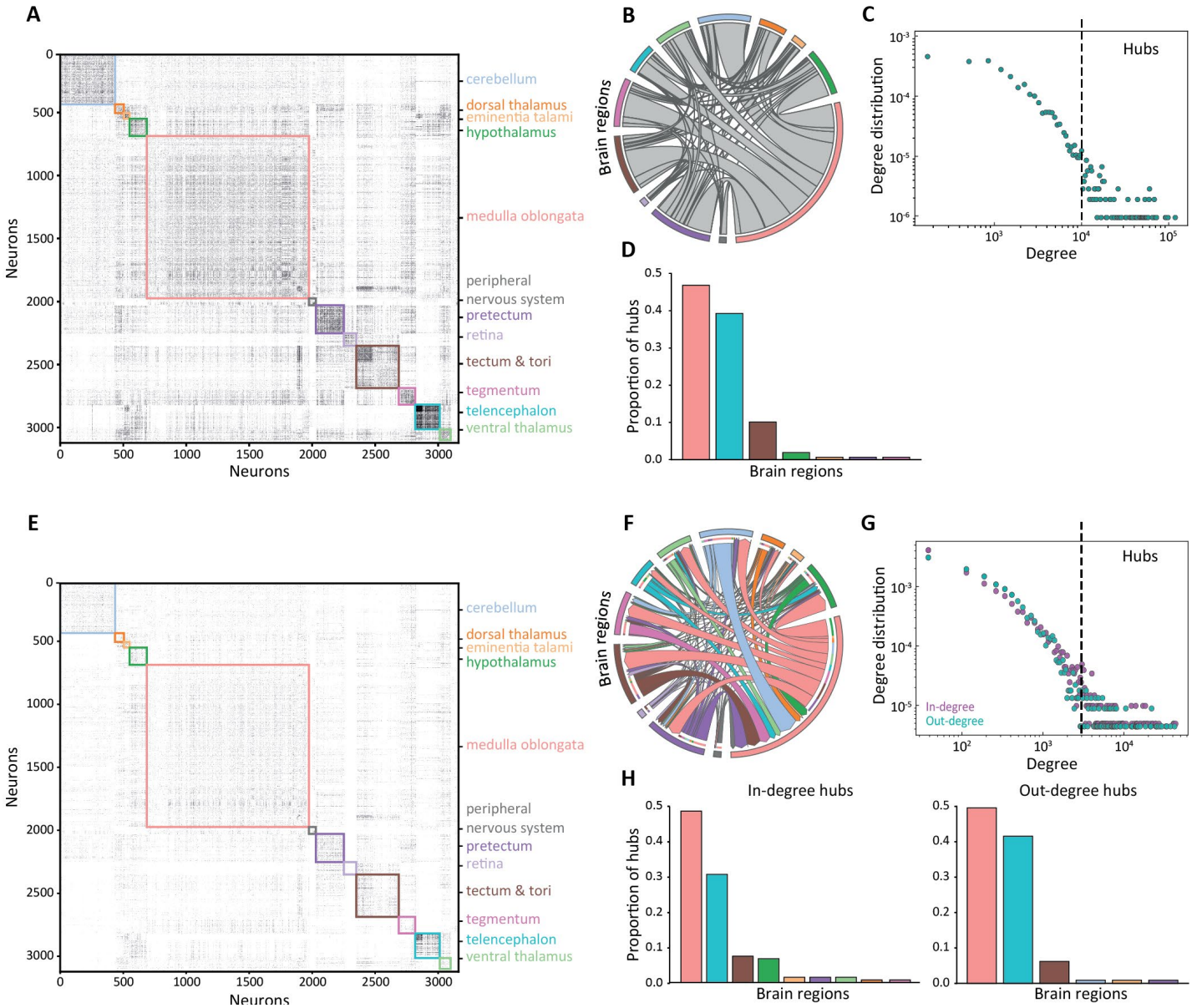



***Figure 2. Topological properties of undirected and directed networks. A, E)*** *Binarized adjacency matrixes showing connections in the undirected (A) and directed (E) networks with neurons sorted by brain regions in which the soma is located. The matrix is binarized for*

*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, $A_{ij}^{U}$, and $A_{ij}^{D}$, we extracted a set of graph parameters. In general, being the non-zero entries of the directed matrix a subset of the undirected ones, $A_{ij}^{D}$ is sparser than $A_{ij}^{U}$. The analysis of the undirected graph, $A_{ij}^{U}$, 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 cells), 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 0.002 and an average shortest path length of 3.792. The undirected graph shows 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 2C). Interestingly, for low-degree nodes, the distribution shows a flat trend, suggesting deviations from the power-law behavior. An assortativity analysis gives a correlation coefficient $\beta$ =0.345, meaning that the graph is assortative and that nodes tend to connect with those of 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 2D - Pink), followed by the Telencephalon (Figure2D - Cyan) and the Tectum Opticum (Figure 2D - Brown). On the other hand, regarding the directed graph, $A_{ij}^{D}$ (Figure 2E), we have found a low clustering coefficient (0.0005) and a path length of 6.10 when considering the weighted adjacency matrix. 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 2G).

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

One of the most appealing use of a brain wide connectome is unravelling possible network architecture and circuit organization motives that characterize an organism brain. We considered to use an established network community discovery strategy based on the modularity optimization, i.e. the Leiden algorithm[41], to analyze the network organization of 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 GC graphs) (Figure 3A).

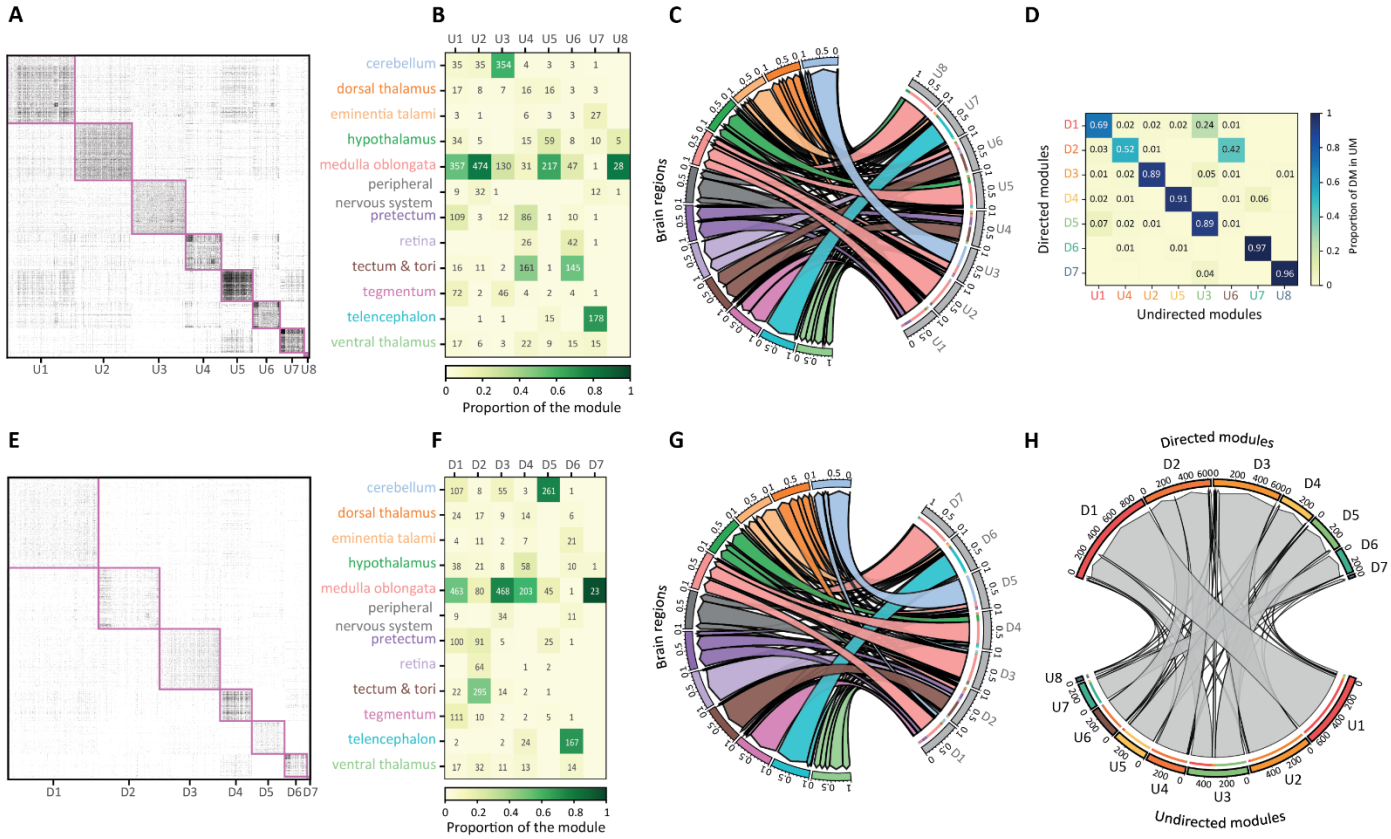



***Figure 3. Distribution of the modules in the different regions for the undirected and directed networks. A,E)*** *Adjacency matrices recapitulating neuronal connections in the undirected (A) and directed (E) networks sorted by the identified modules. The matrix is binarized for visualization purposes only.* ***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) on the brain regions containing the body of neurons belonging to each corresponding module (Figure 3B). 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 3C). Similar to the undirected connectome, the analysis of the directed connectome revealed the presence of seven main modules (D) (Figure 3E-G),

175 showing a substantial overlap, beside two directed modules. D2 includes two undirected modules, U4
176 and U6, and D1 combines U1 with a fraction of U3 (Figure 3D, H). This substantial overlap between the
177 two modularity analyses, directed and undirected cases, points to the possibility that the two
178 corresponding connectomes generated in this way rely to a large extent on a common backbone
179 represented by a directed graph and axo-dendrite communication as reported for other datasets with
180 synapse annotation [42] .

181 **3. Hierarchical anatomical organization of the modules reflects circuit functional role**

182 To better capture the 3D anatomical organization of the 8 identified modules in the undirected graph,
183 we represented the wiring diagrams corresponding to each module in the anatomical reference space.
184 In some of the cases, the identified modules correspond to previously described brain-wide circuits
185 underlying information processing or motor control functions, like the visual network module[19] (Figure
186 4D and Figure 4F) and the cerebellar network module (Figure 4C).

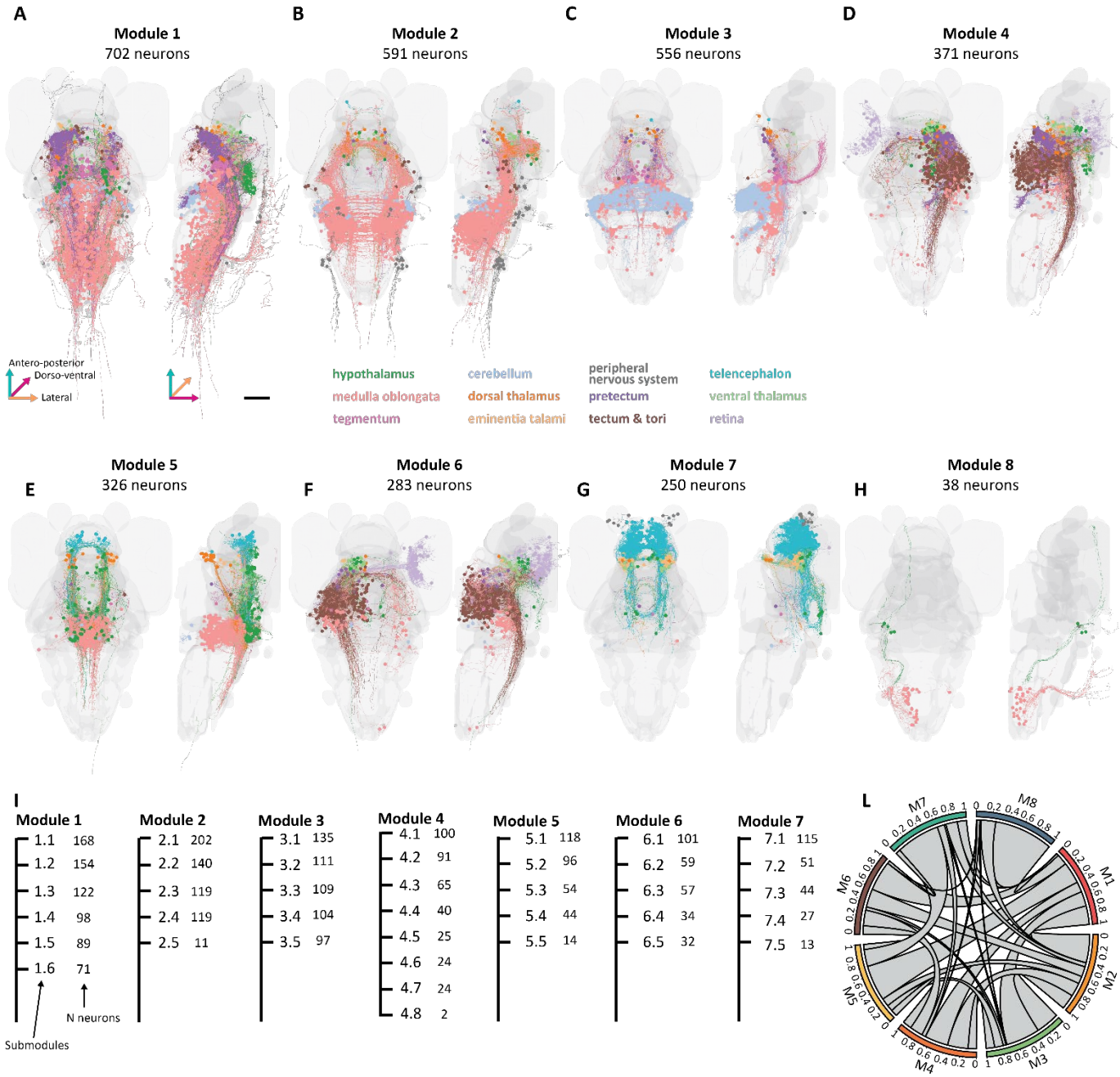
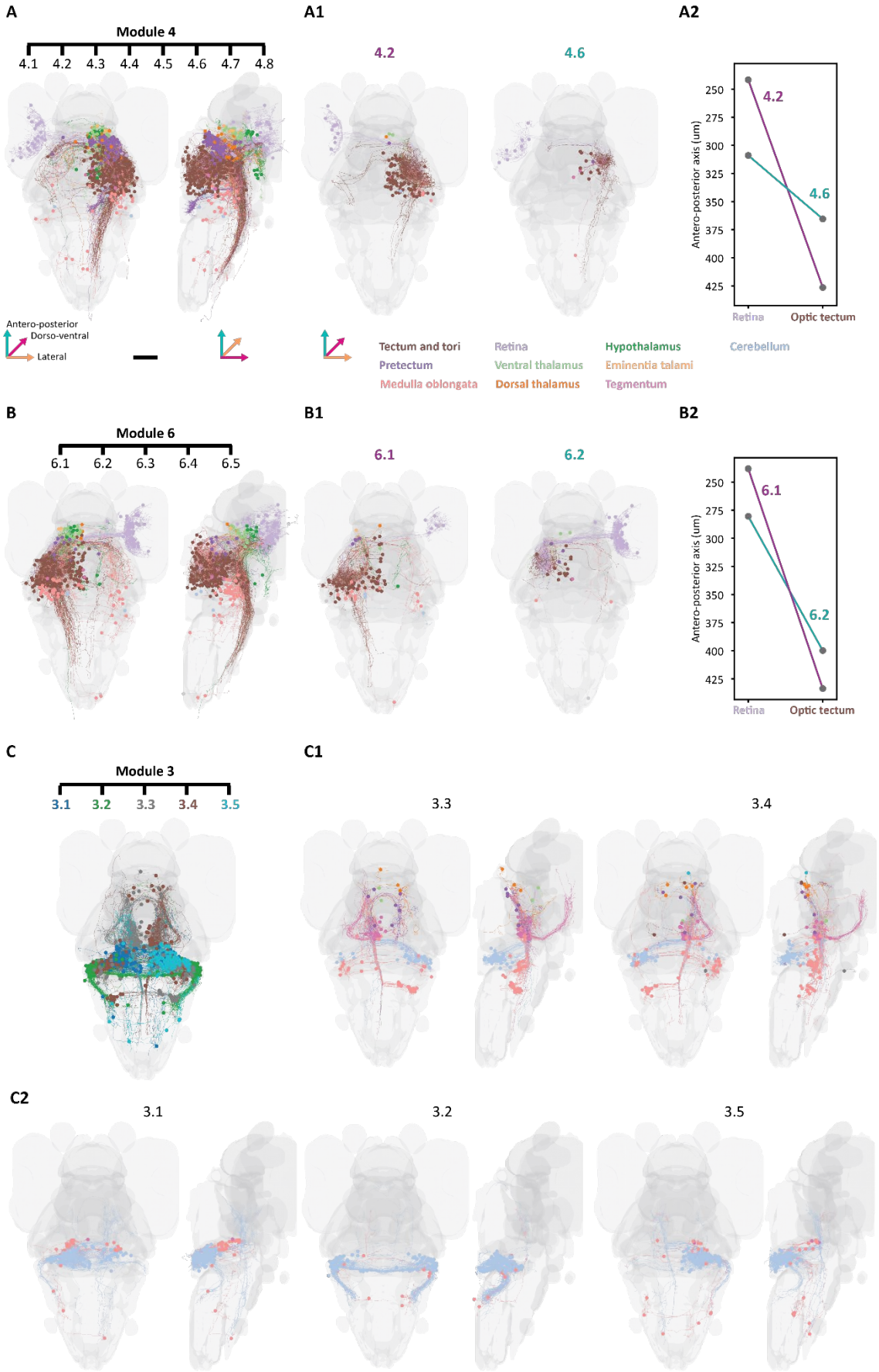



***Figure 4. Modules of the zebrafish brain. A-H)*** *Anatomical representation of the eight high-order modules with the corresponding submodules found in the zebrafish dataset overlaid on the brain silhouette. Colors are assigned to the cells according to the soma position in the different brain regions. Scale bar = 100 µm. I) Tree diagram reporting the structure of the zebrafish network. L) Distribution of connections between the modules.*

To unravel finer structures in the identified modules, we re-iterated the Leiden algorithm in each main module. This approach, where each module is considered an independent subnetwork, led to the finding of a variable number of subcomponents, typically ranging between five and eight (Figure 4 and

196 Supplementary Figures 2 - 7). These submodule patterns, even with the limitations deriving from the
197 neuronal contact approximation employed here, captured an additional level of the circuit
198 organization, involving features of spatial organization of the neuronal populations across the brain.
199 First, we considered the modules including the organization of the Retinal Ganglion Cells, (RGCs),
200 projections to the tectum (modules 4 and 6). The analysis of the relationship of the RGC body position
201 in the retina and the connected PeriVentricular Neurons (PVNs) in the contralateral tectum
202 (PeriVentricular Neurons, PVNs) showed a retinotopic organization in submodules 4.2 and 4.6 (right
203 tectum, Figure 5A), and in the sub modules 6.1 and 6.2 (left tectum) (Figure 5B). Indeed, these
204 submodules show projections from RGC cells with soma located in the caudal part of the retina to
205 connect with PVNs in the rostral part of the contralateral periventricular region at the level of the
206 tectum and, conversely, for the RGC connections involving the rostral part of the retina and the caudal
207 part of the contralateral retino-recipient area. Similarly, our analysis revealed an articulated spatial
208 pattern of connections involving different brain areas associated with the module 3, representing a
209 pretectal-cerebellar network (Figure 5C). The connections formed between neurons located in the
210 medulla oblongata, cerebellum, and tegmentum (Figure 5C1, submodules 3.3 and 3.4) resemble those
211 associated with vestibulo-ocular reflex (VOR)[43]. In this case, the ascending and ascending/descending
212 tangential neurons are located in the medial octavolateral nucleus (MON) of the medulla (Figure 5C1,
213 submodules 3.3 and 3.4), the recipient of utricular inputs. These neurons project to the oculomotor
214 nucleus (nIII) where extraocular motor neurons mediate gaze stabilization[43]. Furthermore, connections
215 between neurons located in the cerebellum have been shown to be involved in VOR[44,45]. The
216 cerebellum displays a spatially segregated pattern of distinct cells, with lateral Purkinje cells projecting
217 to the MON (Figure 5C2, submodule 3.2) and eurydendroid cells located in the middle with less specific
218 projections[36,46] (Figure 5C2, submodules 3.1 and 3.5).


A
Module 4
4.1 4.2 4.3 4.4 4.5 4.6 4.7 4.8
Antero-posterior
Dorso-ventral
Lateral
A1
4.2
4.6
A2
Antero-posterior axis (um)
250 275 300 325 350 375 400 425
4.2
4.6
Retina
Optic tectum
Tectum and tori
Pretectum
Medulla oblongata
Retina
Ventral thalamus
Dorsal thalamus
Hypothalamus
Eminentia talami
Tegmentum
Cerebellum
B
Module 6
6.1 6.2 6.3 6.4 6.5
B1
6.1
6.2
B2
Antero-posterior axis (um)
250 275 300 325 350 375 400 425
6.1
6.2
Retina
Optic tectum
C
Module 3
3.1 3.2 3.3 3.4 3.5
C1
3.3
3.4
C2
3.1
3.2
3.5

220 ***Figure 5. Functional description of the modules. A)*** *Anatomical representation of the high-order module U4. Colors are assigned according*
221 *to the soma distribution in the brain regions. Scale bar = 100 µm.* ***A1)*** *Sub-modules associated with the right Tectum opticum involving*
222 *Peri Ventricular Neurons and Retinal Ganglion Cells, highlighting two different sub-populations of neurons (4.2 caudal tectum, 4.6 medial*
223 *tectum).* ***A2)*** *Median of the distributions along the antero-posterior axis of retina and optic tectum neurons of modules 4.2 (magenta) and*
224 *4.6 (teal).* ***B)*** *Anatomical representation of the high-order module U6. Colors are assigned according to the soma distribution in the brain*
225 *regions.* ***B1)*** *Sub-modules associated with the left Tectum opticum involving PeriVentricular Neurons and Retinal Ganglion Cells,*
226 *highlighting two different sub-populations of neurons (6.1 caudal tectum, 6.2 medial tectum).* ***B2)*** *Median of the distributions along the*
227 *antero-posterior axis of retina and optic tectum neurons of modules 6.1 (magenta) and 6.2 (teal).* ***C)*** *Anatomical representation of the*
228 *high-order module U3 as composed by the five lower-order modules (3.1-3.5). Colors are assigned according to the lower-order modules.*
229 *MON: medial octavolateral nucleus, nIII: oculomotor nucleus.* ***C1-C2)*** *Anatomical representation of the individual lower-order modules.*
230 *Colors are assigned according to the location of the soma. Scale bar = 100 µm.*

231

232 Even though our approach highlighted the existence of principal networks supported by modules of
233 neurons characterized by a relatively higher connection degree, it is clear that the different identified
234 subnetworks show mutual intermodule connection schemes. In order to quantify this aspect, we
235 evaluated the links formed across different modules (see Supplementary Figure 8). This analysis
236 showed a rather heterogeneous scenario, with some modules characterized by a relevant number of
237 intermodular connections, while others with a very limited number of intermodules links.

238 **CONCLUSIONS**

239 We have devised a method to generate approximated connectomes from datasets of single cell
240 reconstruction, covering neuronal circuits composed of different neuronal components represented in
241 terms of single cell morphologies and extending across the complete brain. The proposed approach,
242 inferring an approximated cell-to-cell contact structure on the basis of a spatial proximity rule, can be
243 applied to all datasets with single cell reconstructions where synapse direct annotation is not
244 accessible/available or is not yet complete. This typically is the case of anatomical reconstructions of
245 sparse cells based on light microscopy[32–35]. Within the characteristic resolution limits dictated by the
246 optical constraints, using laser scanning or light-sheet microscopy in combination with restricted or
247 sparse fluorescence labelling, z-stacks of large portions of the brain or even of the complete brain are
248 acquired to map the neuronal cell interregional connectivity or projection patterns[32,47]. In such
249 scenarios, even considering a synapse tagging strategy in combination with pre- and post- synaptic
250 partner labelling[48,49], the synaptic information retrieved could be partial or biased, so overlooking the
251 involvement of other potential players. On the other hand, considering high-resolution, so-called
252 "dense" tissue reconstructions based on serial block electron microscopy, the amount of data to
253 process, even with artificial intelligence approaches, becomes easily relevant even for small brain

254 sections[50]. While automatically reconstructing cell anatomies and their manual validation could
255 become possible for small datasets, a complete synapse annotation may still represent a prohibitive
256 challenge. In this case, one must either focus on a specific circuit or brain region, or use the lower-
257 hanging information available about the cell skeletons to extract approximated information about
258 brain-wide circuit organization. The two approaches are not mutually exclusive, but rather answer two
259 different types of questions. The first aims at characterizing the microcircuit structure or the circuit
260 working motifs[51]. The second, instead, seeks to understand the general organization of brain
261 networks[52].

262 We have applied our approach to a mesoscale atlas of about 3200 single cell reconstructions obtained
263 from a vertebrate organism, *Danio rerio,* where the input-output relationships were estimated based
264 on a proximity rule, and we have analyzed the graph properties of the inferred connectome. From the
265 corresponding adjacency matrices, we extracted the network communities by modularity
266 maximization. In this way, we were able to identify, according to a top-down process, those neuronal
267 modules characterized by a relatively higher degree of connections. With this approach and its
268 reiterations, we have obtained a decomposition of the connectomes of two phylogenetically distant
269 organisms on a set of basic components that form a sort of neuronal backbone of the brain wiring
270 diagram. For most of the identified modules, it is possible to associate a functional role, either based
271 on known anatomical characterizations of the circuits involved, or on the evidence provided by
272 functional analyses[19,36,43,46]. In conclusion, the organization of the modules obtained from the zebrafish
273 larva light microscopy brain dataset shows that the application of such top-down approach of
274 community detection is able to capture relevant circuit motifs and may be applied to several datasets
275 at both the circuit and whole-brain scale.

276 **Methods**

277 **Dataset and data extraction**

278 We used a whole-brain atlas of the zebrafish larva[36]. The atlas comprises more than 3000 neurons,
279 approximately 3% of the total neurons in the zebrafish larvae brain, covering almost the entire brain,
280 anatomically reconstructed by random and sparse Green Fluorescent Protein (GFP) labelling using
281 BGUG[53] 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 neuronal skeleton $i$ from the original dataset of 3163 reconstructions, we imported the .swc file and used the "networkx" library (https://networkx.org/)[55] to convert the corresponding collection of ordinated endpoints to a tree-like directed graph representing its anatomical structure. The soma is 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 numbers (below referred to as endpoint unless otherwise specified), which we denoted as $E_{\mu}^{i}$, where the italic superscript refers to *i*-th neuron within the dataset and the Greek subscript stands for the $\mu$-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[36,56,57]. 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 approximately 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 $\mu$ is an axonal endpoint for neuron $i$, and $\nu$ 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.

**Network properties**

The analysis of network parameters[58,59] was performed using dedicated Python libraries. The following functions from the 'networkx' [55] (https://networkx.org/) 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[41] to extract neuronal modules. The analysis of network parameters was performed in Python using the 'leidenalg' package[60] (https://leidenalg.readthedocs.io/en/stable/index.html) .

The Leiden community detection algorithm iteratively refines an initial partition of nodes in a network to optimize a quality function called modularity Q. It has three main steps, which are fast local movement of nodes to communities, refinement of partition and lastly aggregation.

1. **Initialization**. The algorithm starts with a singleton partition which means each node is considered as its own community. The quality function Q is computed using the formulas below. We have chosen the well-known quality function for the undirected case[61] $Q_{undir} = \frac{1}{2m}\sum_{ij}\left(A_{ij} - \gamma\frac{k_i k_j}{2m}\right)\delta(\sigma_i, \sigma_j)$ and for the directed case[62] $Q_{dir} = \frac{1}{m}\sum_{ij}\left(A_{ij} - \gamma\frac{{k_i}^{out}\,{k_j}^{in}}{m}\right)\delta(\sigma_i, \sigma_j)$, 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[63] 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 $\sigma_i$ denotes the community of node $i$ with $\delta$ representing the Kronecker delta.

2. **Fast local movement of nodes.** A set of all nodes is created. The following steps are repeated until there are no further nodes in a list of nodes: (i) randomly select and remove a node from this set; (ii) for each community, calculate the gain in modularity as if the node were to move to that community; (iii) consider the move that results in the largest modularity gain, or no move if the modularity cannot be improved; (iv) update the partition by moving the node to the selected community; (v) all the neighboring nodes of the selected node which are not in the new community are added to the set of nodes. The steps are repeated until there is no node in the set of nodes.

3. **Refinement phase.** The partition P obtained from the fast local moving phase is refined as follows. Let the refined partition P* be initialized to a singleton partition, where each node represents a community. Randomly select one of the singleton communities in P*, and merge it to another community C* from P* with a probability that depends on the increase of the quality function. The higher the increase, the higher the probability of the merging. Repeat till all

singleton community of P* are merged. Iterate these steps over all the communities of C belonging to P and obtain a new refined partition P*.

4. **Aggregation phase.** In this phase, each community in P* is considered as a node, while an edge between such communities is present if any of their nodes are connected. In this aggregated graph, communities are formed by merging all the aggregated nodes which are also subsets of the same community in the original partition.

5. **Convergence.** Return to step 2 of fast local moving of nodes and then terminate if there is no further change in the quality function.

## 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.

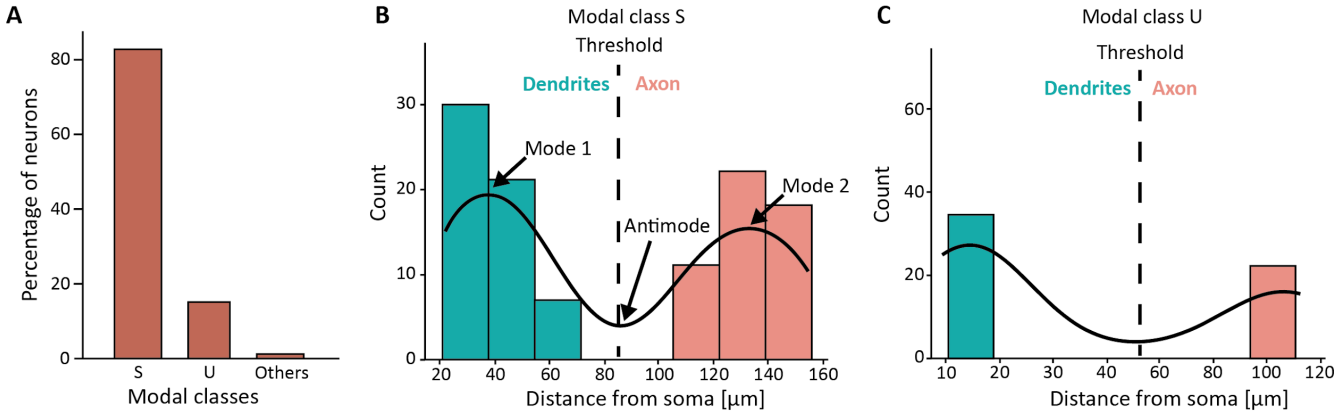
M.D.M: conceptualization, supervision and funding acquisition.

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

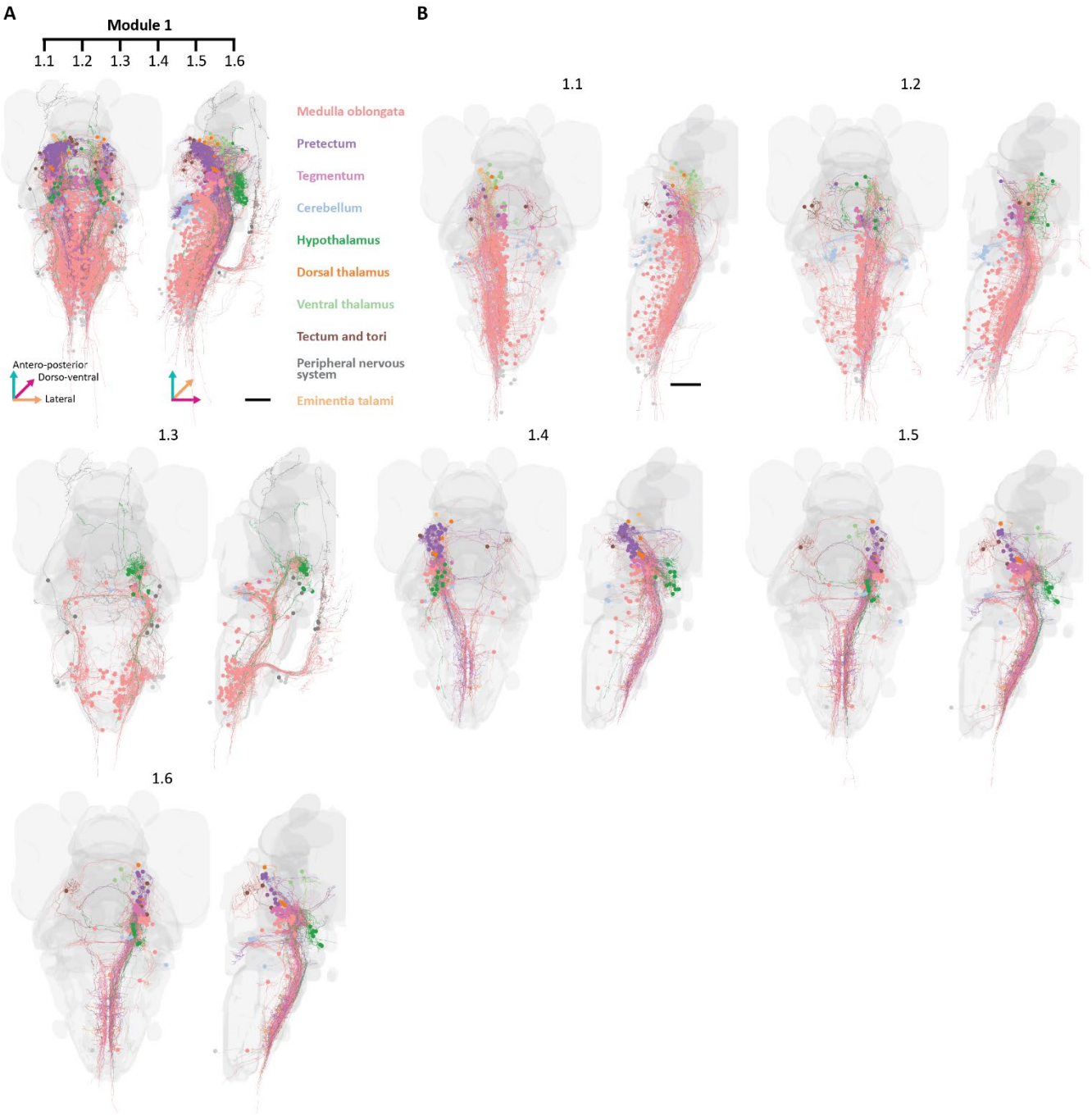
**Acknowledgments**

We thank Delfina Iriarte, Physics department of University of Padua and Nora Nikoloska, Mathematics Department of University of Padua for their support in preliminary analyses; we thank Prof. Tomaso Erseghe from the Information Engineering Department of University of Padua for useful discussion.

SUPPLEMENTARY FIGURES

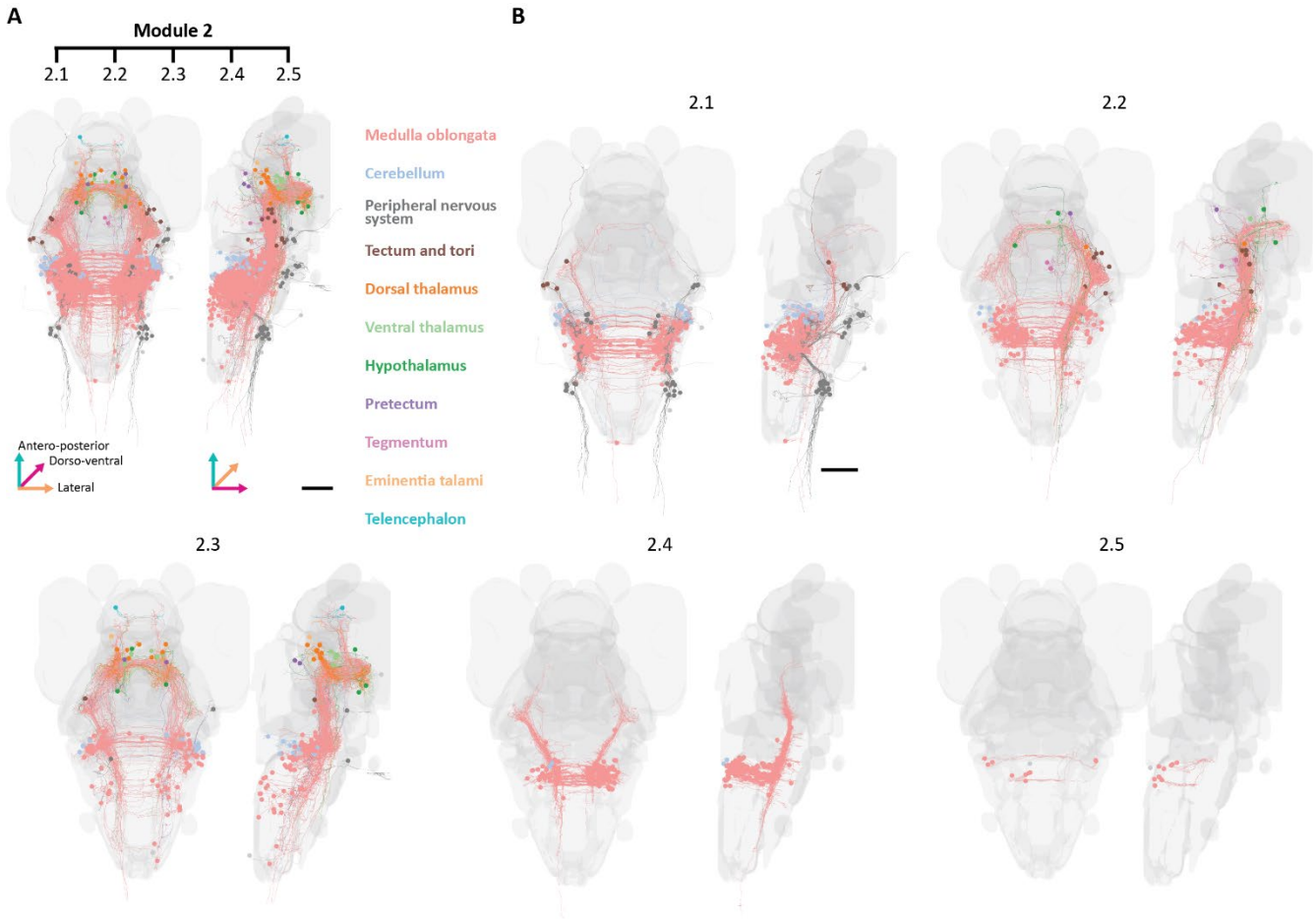



**Supplementary Figure 1. Distribution of endpoints of neurons. A**) Typically the distance between the endpoints of a neuron and the soma is distributed accordingly to two main distribution classes, following the Galtung classification[40]. **B**) Bimodal class S. The antimode of the distribution is 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.

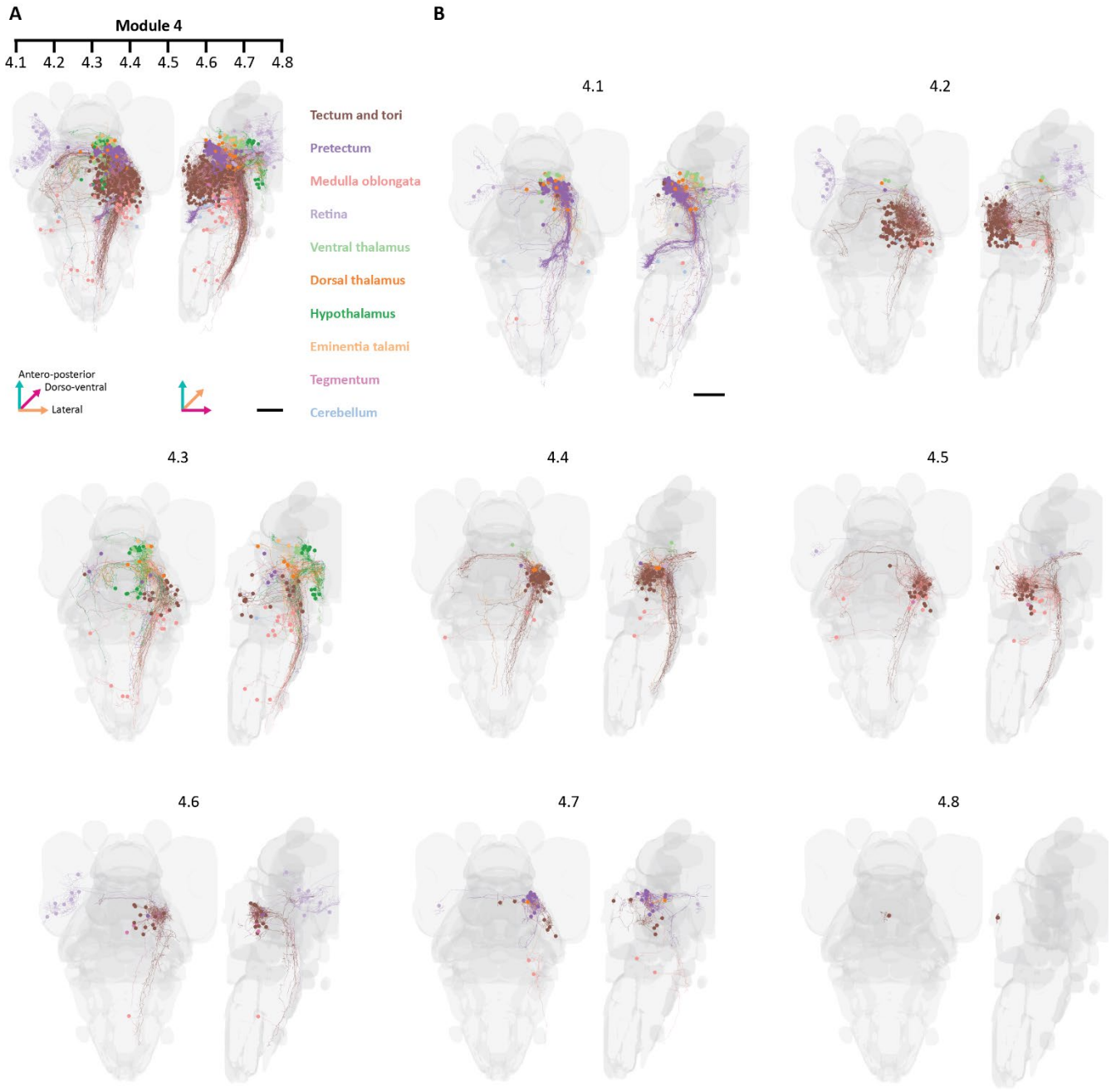



402

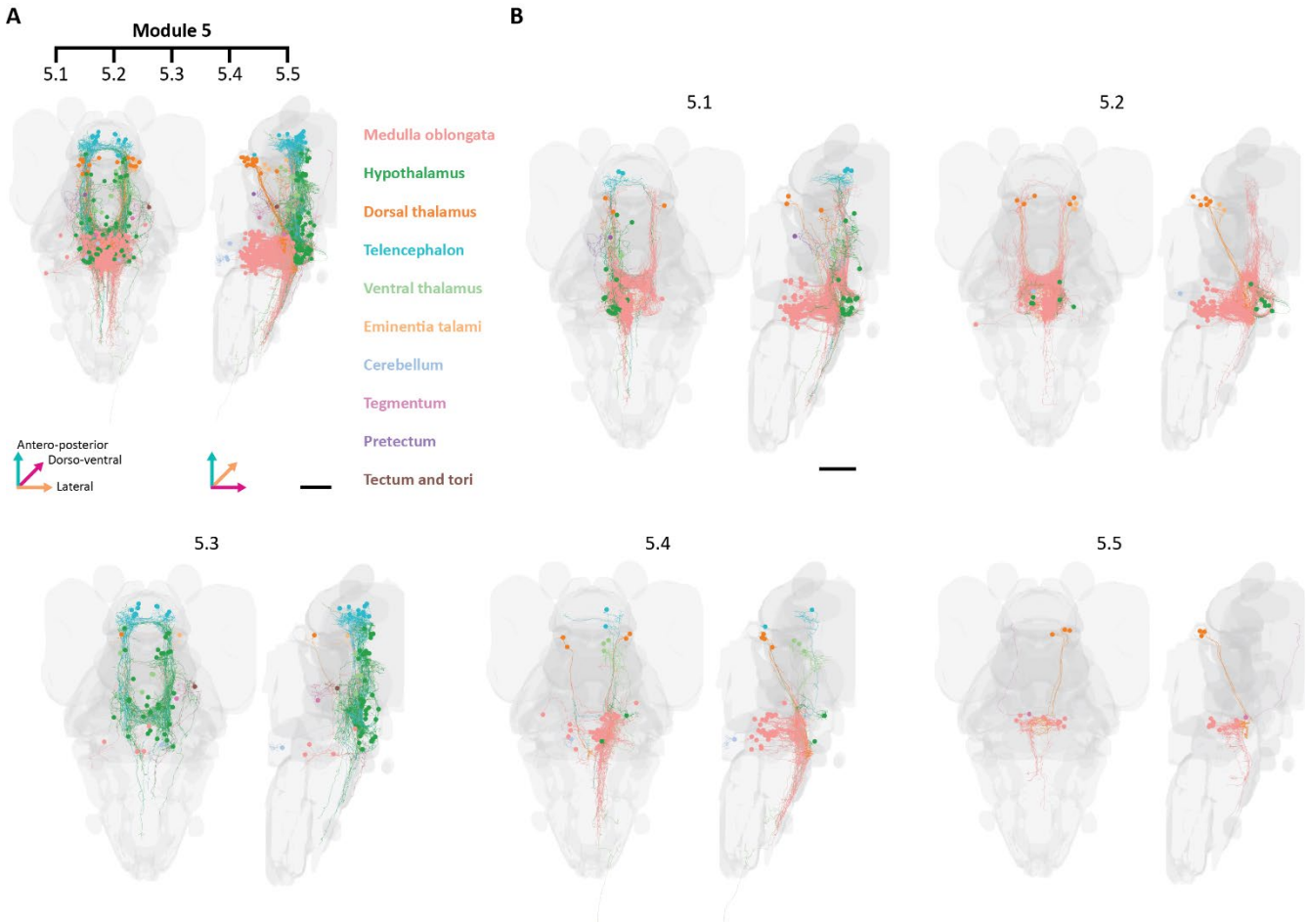
403 **Supplementary Figure 2. The zebrafish module 1. A)** Anatomical representation of the high-order module U1. Colors are
404 assigned according to the soma distribution in the brain regions. Scale bar = 100 µm. **B)** Anatomical representation of the
405 individual lower-order modules. Colors are assigned according to the location of the soma. Scale bar = 100 µm.



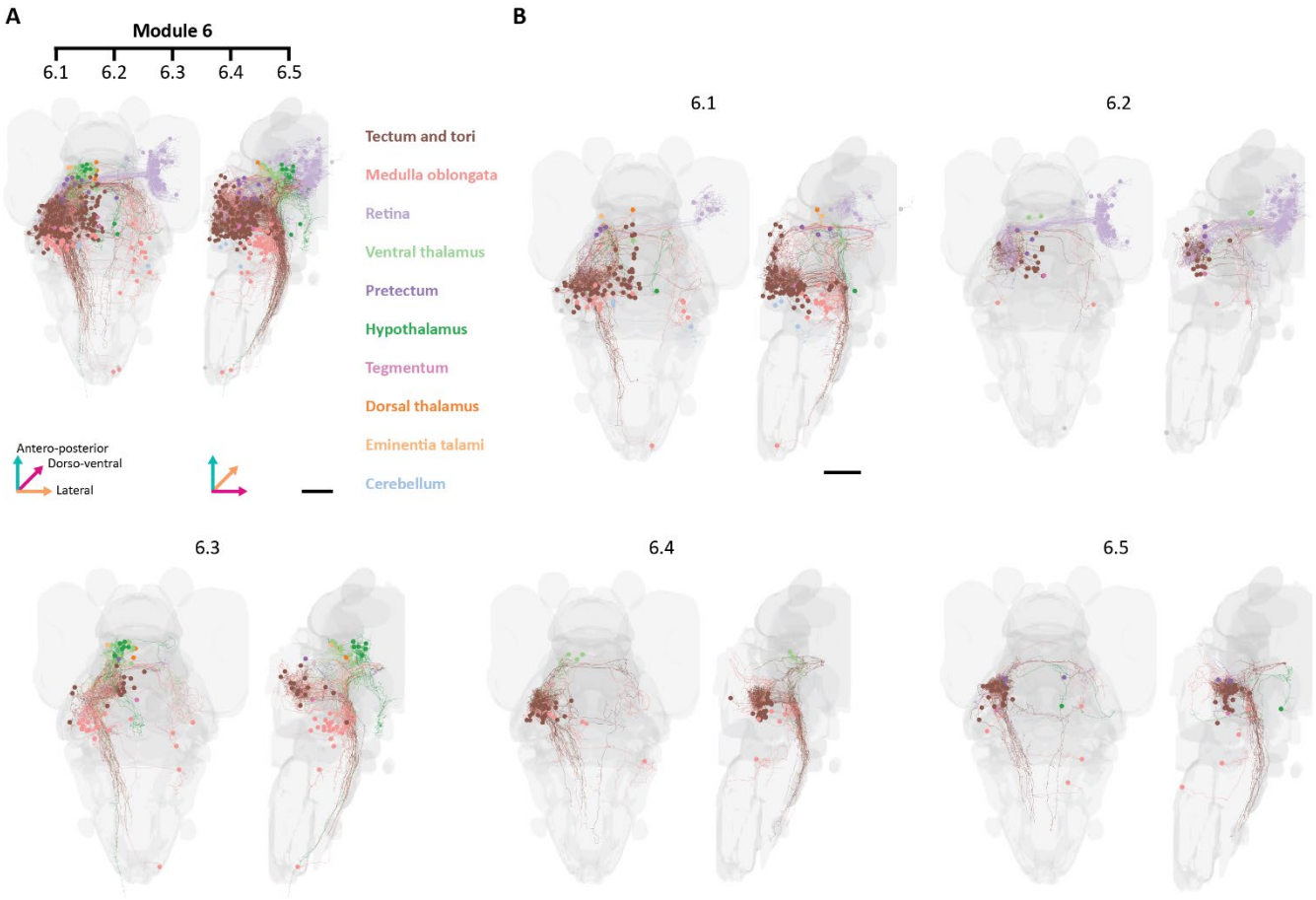

**Supplementary Figure 3. The zebrafish module 2. A)** Anatomical representation of the high-order module U2. Colors are assigned according to the soma distribution in the brain regions. Scale bar = 100 µm. **B)** Anatomical representation of the individual lower-order modules. Colors are assigned according to the location of the soma. Scale bar = 100 µm.



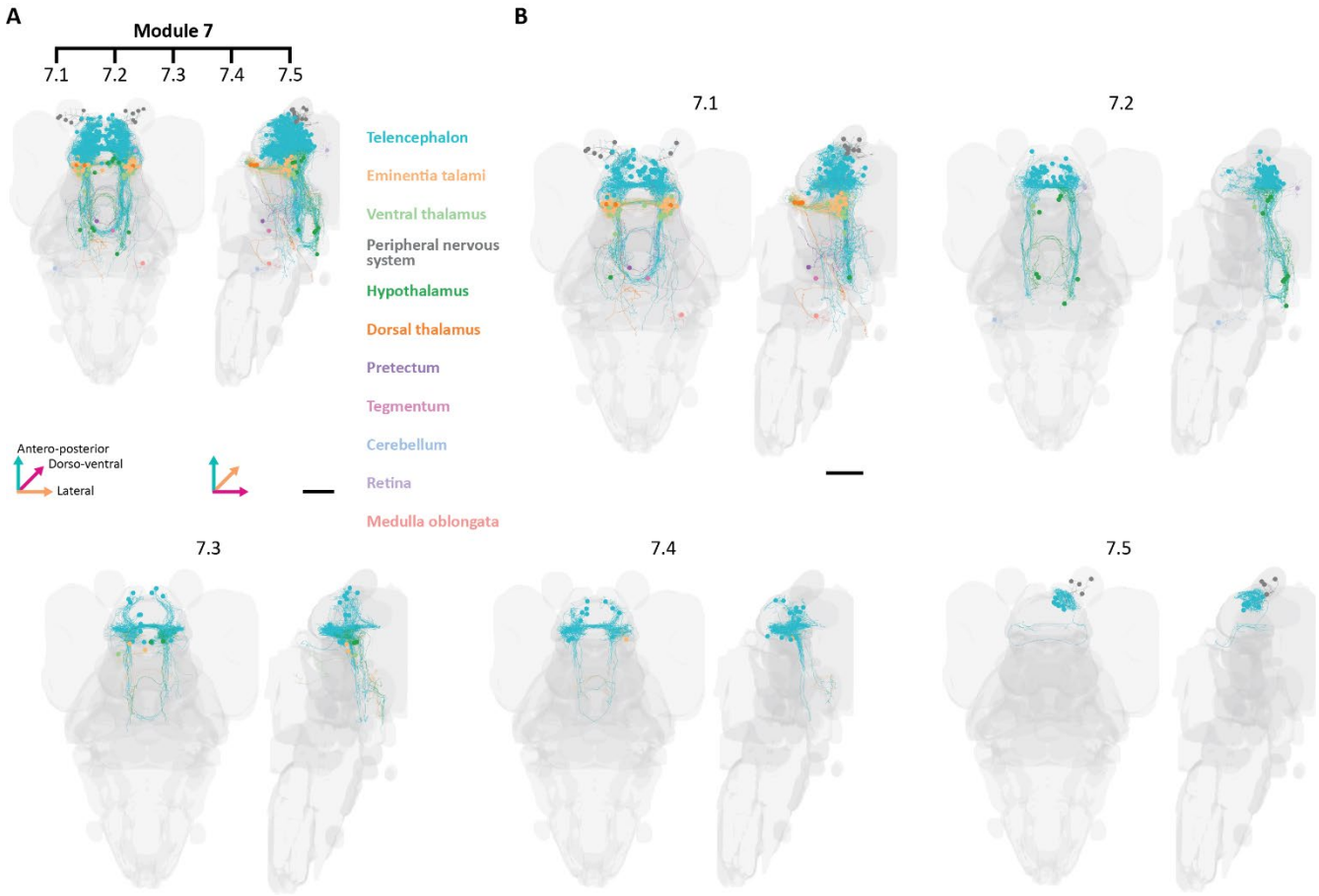

**Supplementary Figure 4. The zebrafish module 4. A)** Anatomical representation of the high-order module U4. Colors are assigned according to the soma distribution in the brain regions. Scale bar = 100 µm. **B)** Anatomical representation of the individual lower-order modules. Colors are assigned according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



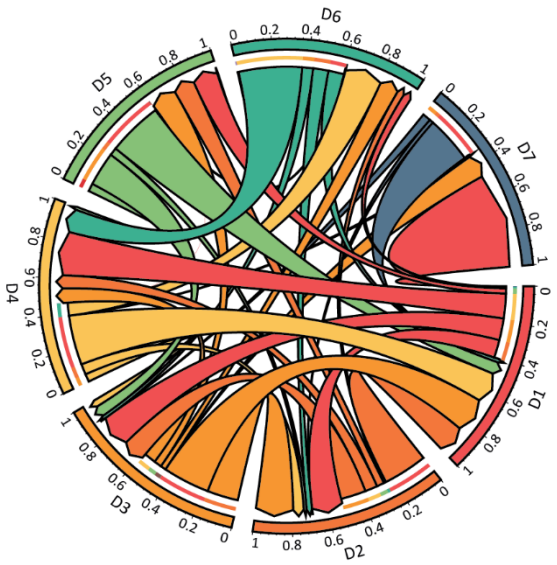

**Supplementary Figure 5. The zebrafish module 5. A)** Anatomical representation of the high-order module U5. Colors are assigned according to the soma distribution in the brain regions. Scale bar = 100 µm. **B)** Anatomical representation of the individual lower-order modules. Colors are assigned according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




**Supplementary Figure 6. 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 7. 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 8. Relationships between zebrafish modules.** Distribution of connections between the modules of the directed network. D: directed modules.

433 **REFERENCES**


434 1. Brunner, C. *et al.* Whole-brain functional ultrasound imaging in awake head-fixed mice. *Nat Protoc* **16**,
435 3547–3571 (2021).
436 2. Ahrens, M. B., Orger, M. B., Robson, D. N., Li, J. M. & Keller, P. J. Whole-brain functional imaging at cellular
437 resolution using light-sheet microscopy. *Nat Methods* **10**, 413–420 (2013).
438 3. Bruzzone, M. *et al.* Whole brain functional recordings at cellular resolution in zebrafish larvae with 3D
439 scanning multiphoton microscopy. *Sci Rep* **11**, 11048 (2021).
440 4. Schrödel, T., Prevedel, R., Aumayr, K., Zimmer, M. & Vaziri, A. Brain-wide 3D imaging of neuronal activity in
441 Caenorhabditis elegans with sculpted light. *Nat Methods* **10**, 1013–1020 (2013).
442 5. Brondi, M., Bruzzone, M., Lodovichi, C. & Dal Maschio, M. Optogenetic Methods to Investigate Brain
443 Alterations in Preclinical Models. *Cells* **11**, 1848 (2022).
444 6. Ahrens, M. B. *et al.* Brain-wide neuronal dynamics during motor adaptation in zebrafish. *Nature* **485**, 471–
445 477 (2012).
446 7. Chen, X. *et al.* Brain-wide Organization of Neuronal Activity and Convergent Sensorimotor Transformations
447 in Larval Zebrafish. *Neuron* **100**, 876-890.e5 (2018).
448 8. Ahrens, M. B. & Engert, F. Large-scale imaging in small brains. *Curr Opin Neurobiol* **32**, 78–86 (2015).
449 9. Naumann, E. A. *et al.* From Whole-Brain Data to Functional Circuit Models: The Zebrafish Optomotor
450 Response. *Cell* **167**, 947-960.e20 (2016).
451 10. Mariani, B. *et al.* Disentangling the critical signatures of neural activity. *Sci Rep* **12**, 10770 (2022).
452 11. Barzon, G., Nicoletti, G., Mariani, B., Formentin, M. & Suweis, S. Criticality and network structure drive
453 emergent oscillations in a stochastic whole-brain model. *J. Phys. Complex.* **3**, 025010 (2022).
454 12. Panzeri, S., Harvey, C. D., Piasini, E., Latham, P. E. & Fellin, T. Cracking the Neural Code for Sensory
455 Perception by Combining Statistics, Intervention, and Behavior. *Neuron* **93**, 491–507 (2017).

456 13. Schaeffer, R., Khona, M. & Fiete, I. R. *No Free Lunch from Deep Learning in Neuroscience: A Case Study*
457 *through Models of the Entorhinal-Hippocampal Circuit*.
458 http://biorxiv.org/lookup/doi/10.1101/2022.08.07.503109 (2022) doi:10.1101/2022.08.07.503109.
459 14. Vanwalleghem, G. C., Ahrens, M. B. & Scott, E. K. Integrative whole-brain neuroscience in larval zebrafish.
460 *Current Opinion in Neurobiology* **50**, 136–145 (2018).
461 15. Beltramo, R. *et al.* Layer-specific excitatory circuits differentially control recurrent network dynamics in the
462 neocortex. *Nat Neurosci* **16**, 227–234 (2013).
463 16. Rajasethupathy, P., Ferenczi, E. & Deisseroth, K. Targeting Neural Circuits. *Cell* **165**, 524–534 (2016).
464 17. Dal Maschio, M., Donovan, J. C., Helmbrecht, T. O. & Baier, H. Linking Neurons to Network Function and
465 Behavior by Two-Photon Holographic Optogenetics and Volumetric Imaging. *Neuron* **94**, 774-789.e5
466 (2017).
467 18. D, F., M, D. M., E, L. & H, B. An optogenetic toolbox for unbiased discovery of functionally connected cells
468 in neural circuits. *Nature communications* vol. 8 https://pubmed.ncbi.nlm.nih.gov/28740141/ (2017).
469 19. Helmbrecht, T. O., Dal Maschio, M., Donovan, J. C., Koutsouli, S. & Baier, H. Topography of a Visuomotor
470 Transformation. *Neuron* **100**, 1429-1445.e4 (2018).
471 20. Wanner, A. A. & Friedrich, R. W. Whitening of odor representations by the wiring diagram of the olfactory
472 bulb. *Nat Neurosci* **23**, 433–442 (2020).
473 21. Wanner, A. A., Genoud, C., Masudi, T., Siksou, L. & Friedrich, R. W. Dense EM-based reconstruction of the
474 interglomerular projectome in the zebrafish olfactory bulb. *Nat Neurosci* **19**, 816–825 (2016).
475 22. Svara, F. *et al.* Automated synapse-level reconstruction of neural circuits in the larval zebrafish brain. *Nat*
476 *Methods* **19**, 1357–1366 (2022).
477 23. White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of the
478 nematode Caenorhabditis elegans. *Philos Trans R Soc Lond B Biol Sci* **314**, 1–340 (1986).
479 24. Takemura, S. *et al.* The comprehensive connectome of a neural substrate for 'ON' motion detection in
480 Drosophila. *eLife* **6**, e24394 (2017).

481 25. Zheng, Z. *et al.* A Complete Electron Microscopy Volume of the Brain of Adult Drosophila melanogaster.
482 *Cell* **174**, 730-743.e22 (2018).
483 26. Schubert, P. J. *et al.* SyConn2: dense synaptic connectivity inference for volume electron microscopy. *Nat*
484 *Methods* **19**, 1367–1370 (2022).
485 27. Ohyama, T. *et al.* A multilevel multimodal circuit enhances action selection in Drosophila. *Nature* **520**, 633–
486 639 (2015).
487 28. Randel, N. *et al.* Neuronal connectome of a sensory-motor circuit for visual navigation. *eLife* **3**, e02730
488 (2014).
489 29. Ryan, K., Lu, Z. & Meinertzhagen, I. A. The CNS connectome of a tadpole larva of Ciona intestinalis (L.)
490 highlights sidedness in the brain of a chordate sibling. *eLife* **5**, e16962 (2016).
491 30. Consortium, T. Mic. *et al.* Functional connectomics spanning multiple areas of mouse visual cortex.
492 2021.07.28.454025 Preprint at https://doi.org/10.1101/2021.07.28.454025 (2023).
493 31. The connectome of an insect brain | Science. https://www.science.org/doi/10.1126/science.add9330.
494 32. Oh, S. W. *et al.* A mesoscale connectome of the mouse brain. *Nature* **508**, 207–214 (2014).
495 33. Jenett, A. *et al.* A GAL4-Driver Line Resource for Drosophila Neurobiology. *Cell Reports* **2**, 991–1001 (2012).
496 34. Arganda-Carreras, I. *et al.* A Statistically Representative Atlas for Mapping Neuronal Circuits in the
497 Drosophila Adult Brain. *Front. Neuroinform.* **12**, (2018).
498 35. Bogovic, J. A. *et al.* An unbiased template of the Drosophila brain and ventral nerve cord. *PLOS ONE* **15**,
499 e0236495 (2020).
500 36. Kunst, M. *et al.* A Cellular-Resolution Atlas of the Larval Zebrafish Brain. *Neuron* **103**, 21-38.e5 (2019).
501 37. Peters, A. & Feldman, M. L. The projection of the lateral geniculate nucleus to area 17 of the rat cerebral
502 cortex. I. General description. *J Neurocytol* **5**, 63–84 (1976).
503 38. Cook, S. J., Kalinski, C. A. & Hobert, O. Neuronal contact predicts connectivity in the C. elegans brain.
504 *Current Biology* **33**, 2315-2320.e2 (2023).

39. Onuchin, A. A., Chernizova, A. V., Lebedev, M. A. & Polovnikov, K. E. Communities in C. elegans connectome through the prism of non-backtracking walks. *Sci Rep* **13**, 22923 (2023).
40. Theory and Methods of Social Research. - Galtung, Johan: 9780231030885 - AbeBooks. https://www.abebooks.it/9780231030885/Theory-Methods-Social-Research-Galtung-0231030886/plp.
41. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep* **9**, 5233 (2019).
42. Winding, M. *et al.* The connectome of an insect brain. *Science* **379**, eadd9330 (2023).
43. Bianco, I. H. *et al.* The tangential nucleus controls a gravito-inertial vestibulo-ocular reflex. *Curr Biol* **22**, 1285–1295 (2012).
44. Functional regionalization of the teleost cerebellum analyzed in vivo | PNAS. https://www.pnas.org/doi/10.1073/pnas.1403105111.
45. 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).
46. Bae, Y.-K. *et al.* Anatomy of zebrafish cerebellum and screen for mutations affecting its development. *Dev Biol* **330**, 406–426 (2009).
47. Yao, S. *et al.* A whole-brain monosynaptic input connectome to neuron classes in mouse visual cortex. *Nat Neurosci* **26**, 350–364 (2023).
48. Shearin, H. K., Quinn, C. D., Mackin, R. D., Macdonald, I. S. & Stowers, R. S. t-GRASP, a targeted GRASP for assessing neuronal connectivity. *J Neurosci Methods* **306**, 94–102 (2018).
49. Kim, J. *et al.* mGRASP enables mapping mammalian synaptic connectivity with light microscopy. *Nat Methods* **9**, 96–102 (2012).
50. Scheffer, L. K. *et al.* A connectome and analysis of the adult Drosophila central brain. *Elife* **9**, e57443 (2020).

51. Eschbach, C. *et al.* Recurrent architecture for adaptive regulation of learning in the insect brain. *Nat Neurosci* **23**, 544–555 (2020).
52. van der Plas, T. L. *et al.* Neural assemblies uncovered by generative modeling explain whole-brain activity statistics and reflect structural connectivity. *eLife* **12**, e83139.
53. Scott, E. K. *et al.* Targeting neural circuitry in zebrafish using GAL4 enhancer trapping. *Nat Methods* **4**, 323–326 (2007).
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. Hagberg, A. A., Schult, D. A. & Swart, P. J. Exploring Network Structure, Dynamics, and Function using NetworkX. (2008).
56. 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).
57. Stepanyants, A. & Chklovskii, D. B. Neurogeometry and potential synaptic connectivity. *Trends Neurosci* **28**, 387–394 (2005).
58. *Network Science by Albert-László Barabási*.
59. Preface. in *Networks: An Introduction* (ed. Newman, M.) 0 (Oxford University Press, 2010). doi:10.1093/acprof:oso/9780199206650.002.0005.
60. Traag, V. *et al.* vtraag/leidenalg: 0.10.0. Zenodo https://doi.org/10.5281/zenodo.8147844 (2023).
61. Newman, M. E. J. & Girvan, M. Finding and evaluating community structure in networks. *Phys. Rev. E* **69**, 026113 (2004).
62. Leicht, E. A. & Newman, M. E. J. Community Structure in Directed Networks. *Phys. Rev. Lett.* **100**, 118703 (2008).
63. Reichardt, J. & Bornholdt, S. Statistical mechanics of community detection. *Phys. Rev. E* **74**, 016110 (2006).