A method for extracting an approximated connectome from libraries of single anatomically co-registered 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

Mapping functional brain information onto a circuit wiring diagram supporting possible brain circuit working mechanisms remains challenging. When available, the analysis of electron microscopy based high-resolution connectomes, leveraging synapse annotation, allows for unraveling portions of the nervous system's wiring diagram or specific circuit motifs. However, obtaining a complete brain reconstruction with annotation of the synapses and extracting the general organization principles of the neuronal networks across the brain remains a demanding effort. To extend the available methods, we present here an approach to obtain an approximate brain connectome starting from libraries of single cell reconstructions co-registered in a reference anatomical space. Our approach relies on both the Strahler numbering of the nodes characterizing the cell morphological reconstructions and a proximity range criterion, to infer the general connectivity structures between the different cells, in the absence of synapse annotation. First, we confronted this approach on a ground truth derived from a fully annotated EM dataset of *Drosophila melanogaster*. The connectivity modules identified with our approach show generally a similar organization compared to those resulting from the EM annotated dataset, in terms of Normalized Mutual Information and Adjusted Rand Index, always with high significance with respect to random null models. Building on this, we then applied this approach to extract an approximate connectome of the zebrafish larvae brain starting from a light microscopy-based dataset of about 3000 co-registered neuronal skeletonizations. Modularity analysis of the retrieved

directed and undirected connectomes revealed a hierarchical organization of the resulting graph. The identified architecture encompasses a series of high-order brain-wide neuronal modules composed of a number of lower-order submodules. Importantly, we show that this approximated representation is sufficient for capturing precise and topographically organized connection patterns as well as known functional circuit architectures. In conclusion, we present a scalable, from-circuit-to-brain range, approach to reveal neuronal architectures supporting brain mechanisms, potentially suitable for hypothesis drawing and as for guiding the exploration of and integration with EM connectomes.

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 of better constraining 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 with 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)[14–18], 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 step, 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

61 and requires computational strategies to minimize human workload[15,19]. Once the information on the
62 neuronal connectivity schemes i.e. the organization of the input-output contacts characterizing the cells
63 involved- has been retrieved, it becomes possible to reconstruct an adjacency matrix representing a
64 connectome and identify the circuit wiring diagrams. Because of the amount and the complexity of the
65 data one has to face, this type of anatomical reconstruction frequently focuses on targeted circuit
66 motifs, specific neuronal subpopulations, or particular brain regions. On the other hand, unraveling the
67 anatomical organization of brain-wide networks requires suitable datasets, with a sufficient sampling
68 of the cellular morphologies, coverage across different regions of the brain, and, ideally, based on the
69 evaluation of a large number of exemplars. Currently, even if there are several reported whole-brain
70 EM reconstructions[15,16,18,20–22], partial or whole-brain connectomes that provide skeletonization of
71 neurons along with detailed annotations of their synaptic contacts are extremely limited. To extend
72 and complement the methods currently available, here, we present a computational approach to
73 generate an approximated connectome starting from an anatomical dataset with neuronal cell
74 skeletonizations, belonging to or co-registered in the same reference space, where the actual synaptic
75 information is not accessible or not yet available. This is the typical setting of mesoscale atlases based
76 on optical microscopy acquisitions, as well as electron microscopy datasets not yet fully annotated at
77 the detail level corresponding to the neuronal input-output features[23–28]. For this type of dataset, here,
78 we devised an approach to infer brain network organization by assigning cell-to-cell connections, which
79 are subsequently translated into adjacency matrices, representing the brain wide cell resolution
80 connectivity scheme. Following its evaluation on EM-based ground truth, we applied this approach to
81 a zebrafish single-cell dataset, which comprises more than three thousand neuronal reconstructions.

82 **Results**

83 **1. Building an approximated connectome based on Strahler numbering and spatial proximity**

84 For the cases where actual synaptic information is unavailable, we evaluated an approach to infer an
85 approximated brain network structure by assigning cell-to-cell connections on the basis of a proximity
86 rule between the nodes along the reconstructed cell skeleton (see the Methods section). For this

purpose, we devised to start from the information of the cell anatomy and calculated for each point of

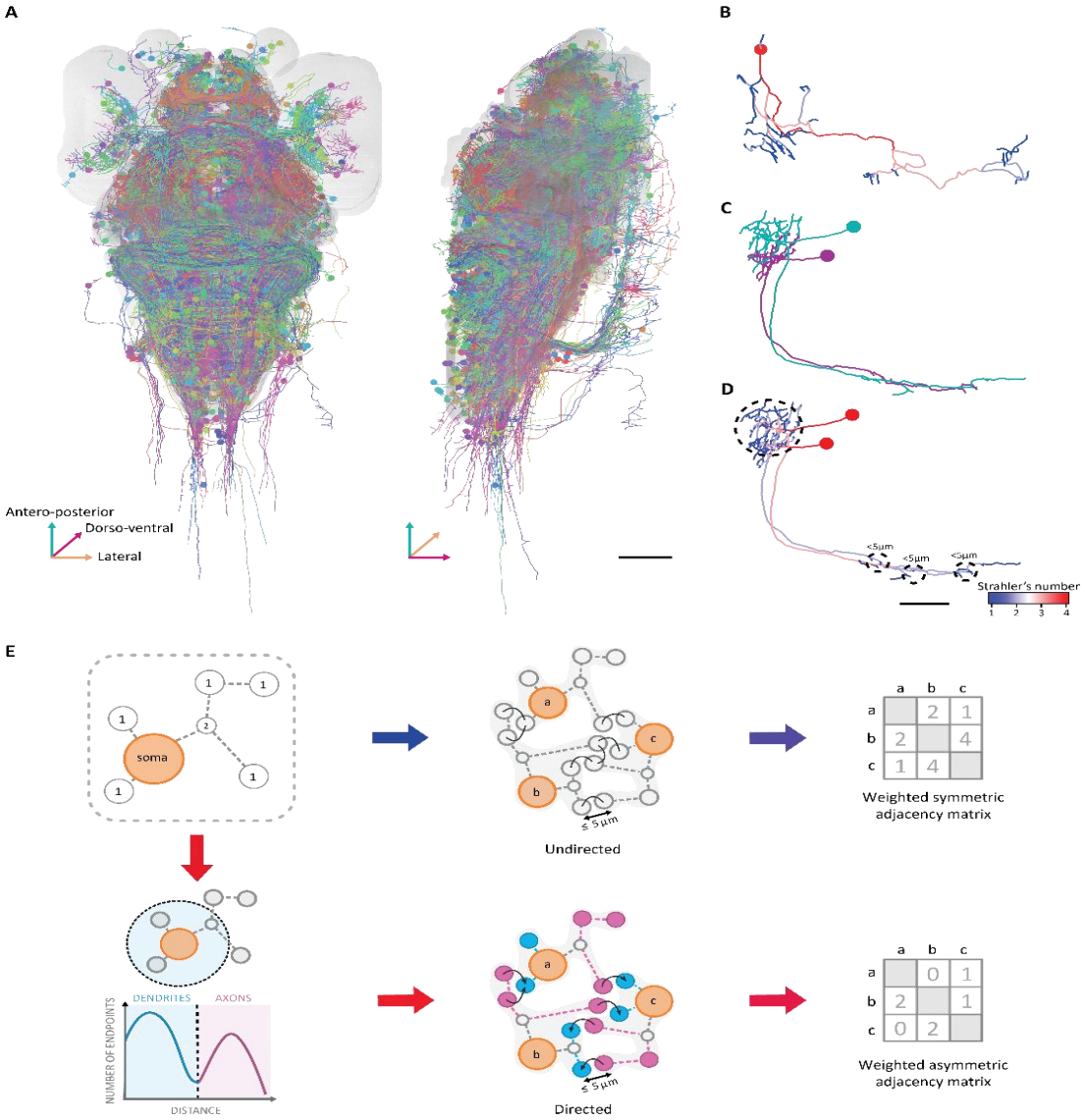


Figure 1. Building the approximated connectome from a mesoscale cell atlas. A) *Anatomically reconstructed neurons from [27] overlaid to the zebrafish atlas. Colors are assigned randomly. Scale bar= 100 µm.* ***B)*** *Representative neuron with nodes color-coded according to the corresponding Strahler's 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).*

the cell skeleton (or neurite) the corresponding Strahler number. This is a parameter that represents the relative level of the cell skeleton points in the cell branching hierarchy[28] (Figure 1A,B). Terminal points (endpoints) along the neurite structure, characterized by a unitary value of the Strahler number, are considered putative points of connection between different cells if they fall within a defined proximity range one from the other. For a typical light microscopy dataset, considering the intrinsic

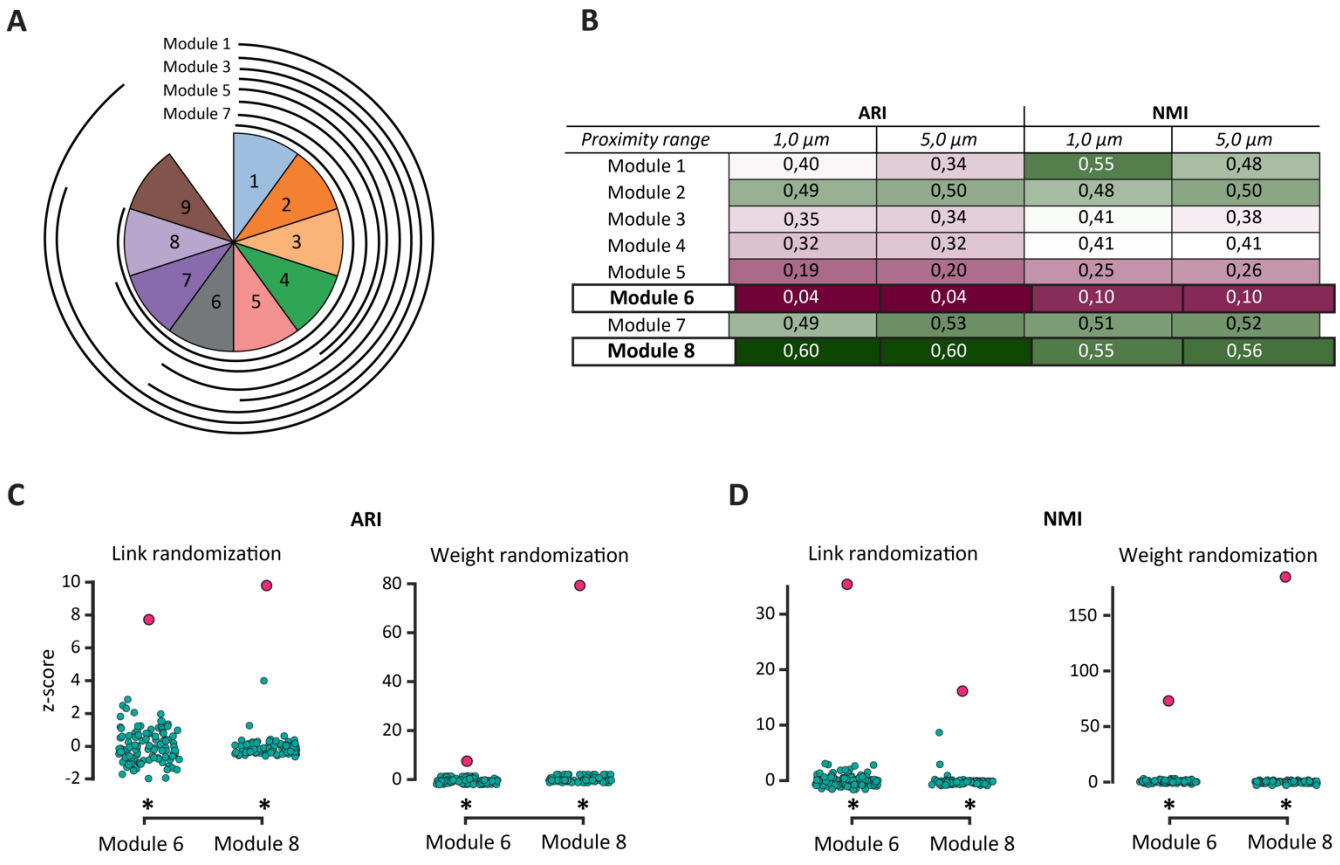
resolution of the acquisition process, the typical errors in the anatomical reconstruction, and co-registration procedures, we adopted a proximity range between 0 and 5 µm (Figure 1D).

As reconstructing the connectome relies on the determination of the adjacency matrix A_{ij}, we envisaged two different possible scenarios. In the first one, we assume no subcellular segregations for the input and output points of a neuron, resulting in a symmetric adjacency matrix A_{ij}^{U} corresponding to an undirected connectome (Figure 1E). In the second one, as most of the neuronal reconstructions available are characterized by a bimodal spatial distribution of the neuronal endpoints[29] (Supplementary Figure 1), we assign 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 allows us to constrain the connectome according to an axon-to-dendrite synaptic scheme. In this latter case, we obtain an asymmetric adjacency matrix A_{ij}^{D} representing a directed connectome (Figure 1E panels at the bottom).

2. Evaluating the connectome reconstruction strategy against an EM-based ground truth

To assess the actual applicability of the approach and its capability to capture brain circuit organization patterns, we took advantage from the synapse-annotated connectome of the Drosophila melanogaster hemibrain, composed by a library of 21739 connected neurons[30]. Starting from the available information on the input-output contacts obtained from the EM analysis, we identified using a modularity analysis algorithm a hierarchical organization with eight main modules (module 1-8), each accounting on average for about three thousand cells distributed in a variable number of submodules, from four to nine, associated with a same main module (Figure 2A). We then considered the same dataset and calculated the adjacency matrix and the modularity organization, but this time without considering the available synapse annotation, but relying solely on the Strahler numbering of the cell skeleton and on the proximity range definition to build an undirected graph (see Methods section). The pairs of adjacency matrices obtained for each main module, one based on the synapse annotation (EM-based) and one relying on the Strahler numbering (Strahler-based) , present Pearson correlation coefficients ranging from 0.05 to 0.215 always with p-value smaller than 0.001 (see Methods), with

126 minimal variations considering the 1 μm and 5 μm proximity range (0.146 ± 0.053 and 0.14 ± 0.052,
127 average ± SD, respectively). In order to evaluate the capability of the approximated approach to actually
128 capture the expected wiring, we calculated for each of the main modules identified in the EM-based
129 analysis the Adjusted Rand Index (ARI)[31–33] based on the cell grouping associated with the organization
130 of the corresponding submodules. This metric is frequently used for comparing different partitions or
131 clustering patterns and returns a value between -1 and 1 to quantify the level of overlap in the
132 contingency matrix. We obtained for the 1 μm proximity range an average ARI of 0.36 ± 0.17 (average
133 ± SD, covering a range between 0.04 and 0.6), pointing to a general overlap of the Strahler-based
134 approach with the ground truth sub-communities. A similar trend resulted for the 5 μm proximity
135 range, 0.38 ± 0.17 (average ± SD, Figure 2B, covering a range between 0.04 and 0.6). To compare the
136 community detection accuracy between the two datasets, we also computed the Normalized Mutual
137 Information (NMI)[34,35]. This metric returns a parameter between 0 and 1, depending on the level of
138 overlap between the communities using adjacency matrices and modularity analysis obtained with the
139 two approaches. NMI ranged between 0.1 and 0.55 (0.38 ± 0.17 average ± SD) and between 0.1 and
140 0.56 (0.38 ± 0.17 average ± SD) for 1 μm and 5 μm, respectively (Figure 2B). Then, to corroborate this
141 outcome, we evaluated the significance of the calculated ARI and NMI scores, with a comparison against
142 two types of random null models. We calculated the Z-score for NMI and ARI between the Strahler-
143 based connectome modules and the EM-based connectome modules, focusing here on two modules in
144 particular, module 6 and module 8, those characterized by the lowest and the highest, ARI and NMI
145 values. We constructed 100 random networks for the two Strahler-based connectome modules either
146 by randomizing the links or the weights of the links between the neurons (see Methods). Then, we
147 quantified the p-value by counting the number of random networks that had its own z-scores (for ARI
148 and NMI) higher than that one of the corresponding Strahler-based modules[36]. For each module we
149 found $p < 0.05$ highlighting that the Strahler-based model can capture connectivity patterns in the
150 networks better than models with the same number of nodes, but different links or link strength
151 organizations.



	ARI		NMI	
Proximity range	*1,0 µm*	*5,0 µm*	*1,0 µm*	*5,0 µm*
Module 1	0,40	0,34	0,55	0,48
Module 2	0,49	0,50	0,48	0,50
Module 3	0,35	0,34	0,41	0,38
Module 4	0,32	0,32	0,41	0,41
Module 5	0,19	0,20	0,25	0,26
Module 6	0,04	0,04	0,10	0,10
Module 7	0,49	0,53	0,51	0,52
Module 8	0,60	0,60	0,55	0,56

Figure 2: Comparison with EM-based ground truth and null models. A) *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. Length of the lines represents the number of submodules for each module.* ***B)*** *ARI and NMI computed between the EM ("ground truth") and Strahler reconstructed network at different proximity ranges.* ***C, D)*** *Z-scores computed over the ARI (panel* ***C****) and NMI (panel* ***D****) between Module 6 (or 8) of EM and Strahler reconstructed network (red point) and z-scores computed over the same mentioned similarity measures between EM and 100 random null models of Strahler reconstructed network (green points). The larger the gaps between the red point and the green points, the farther is the Strahler based networks from random behavior and closer is to the ground truth (EM based networks). *: p<0.05.*

3. Building an approximated connectome of the zebrafish larval brain

Aiming to build an approximated connectome for the zebrafish larval brain, we took advantage of an available dataset with anatomical reconstructions of individual neurons obtained by means of optical microscopy z-stack acquisitions co-registered on the same reference brain[27] (Figure 1A). Applying the proposed method, we retrieved two connectome scenarios showing a rather articulated scheme of neuronal interactions between the different regions of the brain (Figure 3B, Figure 3F). From the obtained adjacency matrices, A_{ij}^{U}, and A_{ij}^{D} (Figure 3A and Figure 3B), 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} resulted 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 3C). 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 β =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 3D - Pink), followed by the Telencephalon (Figure3D - Cyan) and the Tectum Opticum (Figure 3D - Brown). On the other hand, regarding the weighted directed graph, A_{ij}^{D} (Figure 3E), we found a low clustering coefficient (0.0005) and a path length of 6.10. While the in- and out-degree distributions resemble the general heterogeneity of the undirected scenario, the low-degree region did not show a corresponding saturation (Figure 3G).

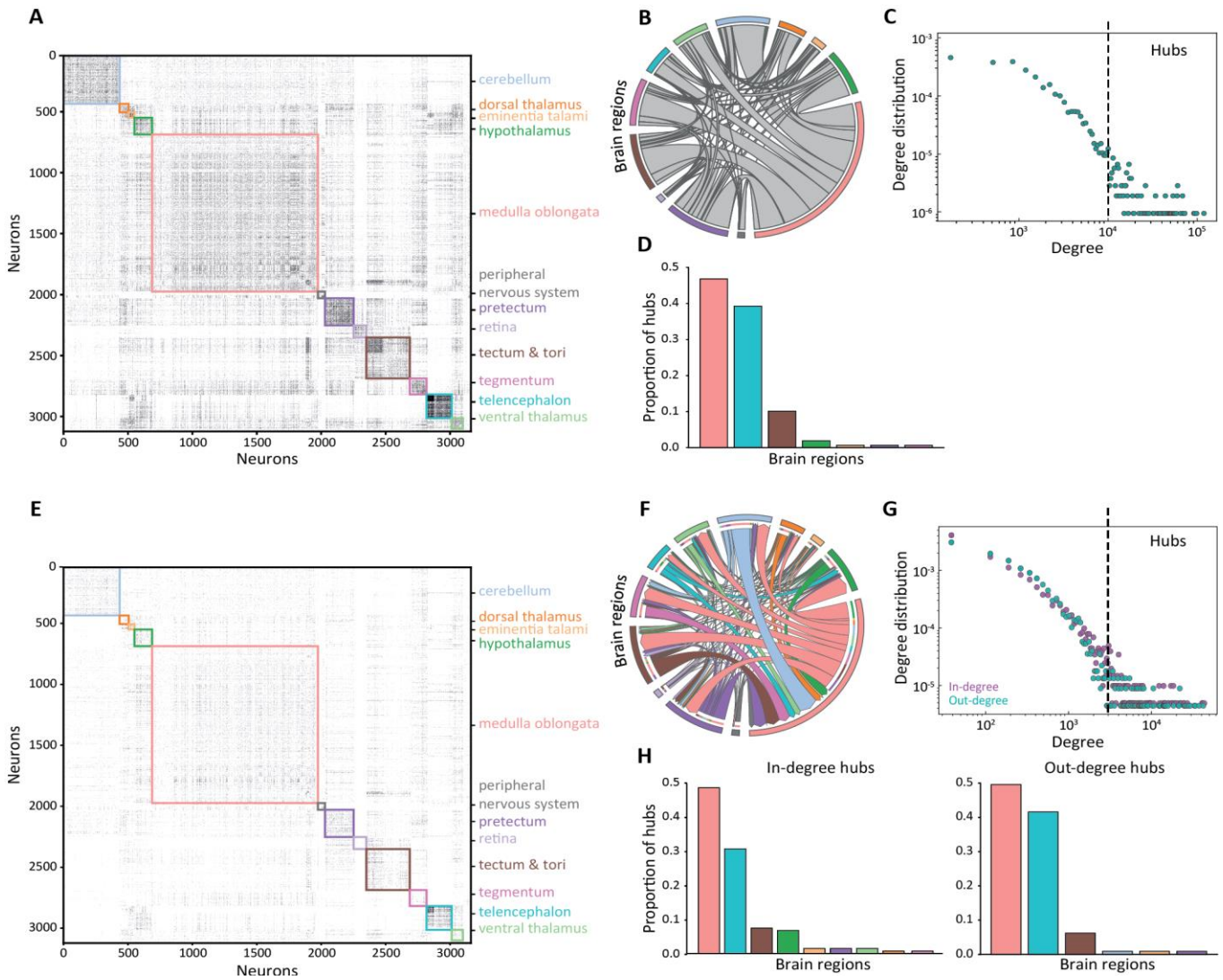
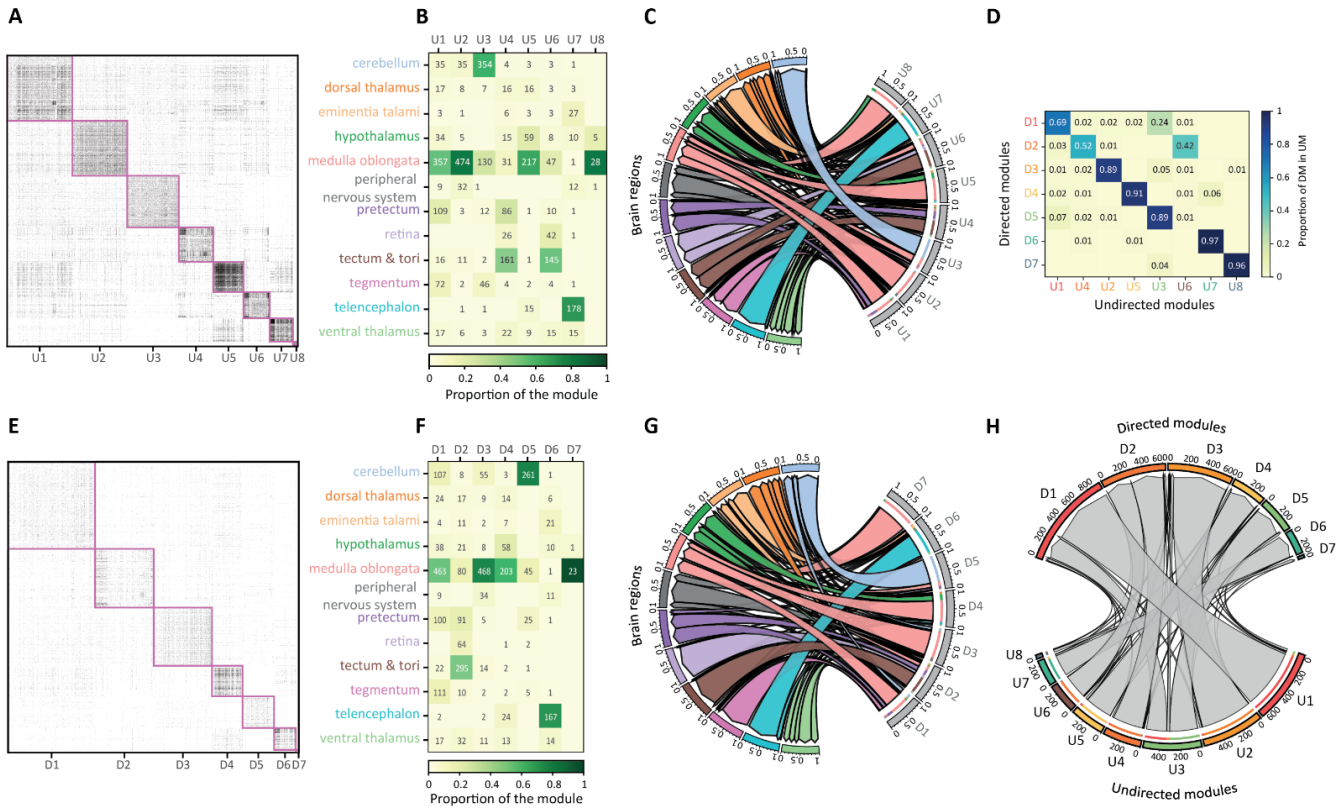


Figure 3. Topological properties of undirected and directed networks. A, E) *Binarized adjacency matrixes showing connections in the undirected (A) and directed (E) networks with neurons sorted by brain regions in which the soma is located. The matrix is binarized for visualization purposes.* ***B, F)*** *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.*

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

One of the most appealing uses 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[37], 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 4A). 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 4B). 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 4C). Similar to the undirected connectome, the analysis of the directed connectome revealed the presence of seven main modules (D) (Figure 4E-G), showing a substantial overlap, beside two directed modules. D2 includes two undirected modules, U4 and U6, and D1 combines U1 with a fraction of U3 (Figure 4D, H). This substantial overlap between the two modularity analyses, directed and undirected cases, 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 as reported for other datasets with synapse annotation[21].



217

218 ***Figure 4. Distribution of the modules in the different regions for the undirected and directed networks.*** ***A,E)*** *Adjacency matrices*
219 *recapitulating neuronal connections in the undirected (A) and directed (E) networks sorted by the identified modules. The matrix is*
220 *binarized for visualization purposes only.* ***B,F)*** *Distribution of the modules across the brain regions for the undirected (B) and directed (F)*
221 *network. Number refers to neurons belonging to a certain module and with the soma located in the regions. Matrix cells are color-coded*
222 *based on their proportion within the module. U: undirected module, D: directed module.* ***C,G)*** *Proportion of the distribution of the modules*
223 *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*
224 *of neurons of the undirected network modules present in the directed network modules.* ***H)*** *Number of neurons of the undirected network*
225 *modules present in the directed network modules. Labels for the colors are reported in D.*

226

227 **5. Hierarchical anatomical organization of the modules reflects circuit functional role**

228

229 To better capture the 3D anatomical organization of the 8 identified modules in the undirected graph,
230 we represented the wiring diagrams corresponding to each module in the anatomical reference space
231 (Figure 5). In some of the cases, the identified modules correspond to previously described brain-wide
232 circuits underlying information processing or motor control functions, like the visual network module[19]
233 (Figure 5D and Figure 5F) and the cerebellar network module (Figure 5C). To unravel finer structures in
234 the identified modules, we re-iterated the Leiden algorithm in each main module. This approach, where
235 each module is considered an independent subnetwork, led to the finding of a variable number of
236 subcomponents, typically ranging between five and eight (Figure 5 and Supplementary Figures 2 - 7).

These submodule patterns, even with the limitations deriving from the neuronal contact approximation employed here, captured an additional level of the circuit organization, involving features of spatial organization of the neuronal populations across the brain.

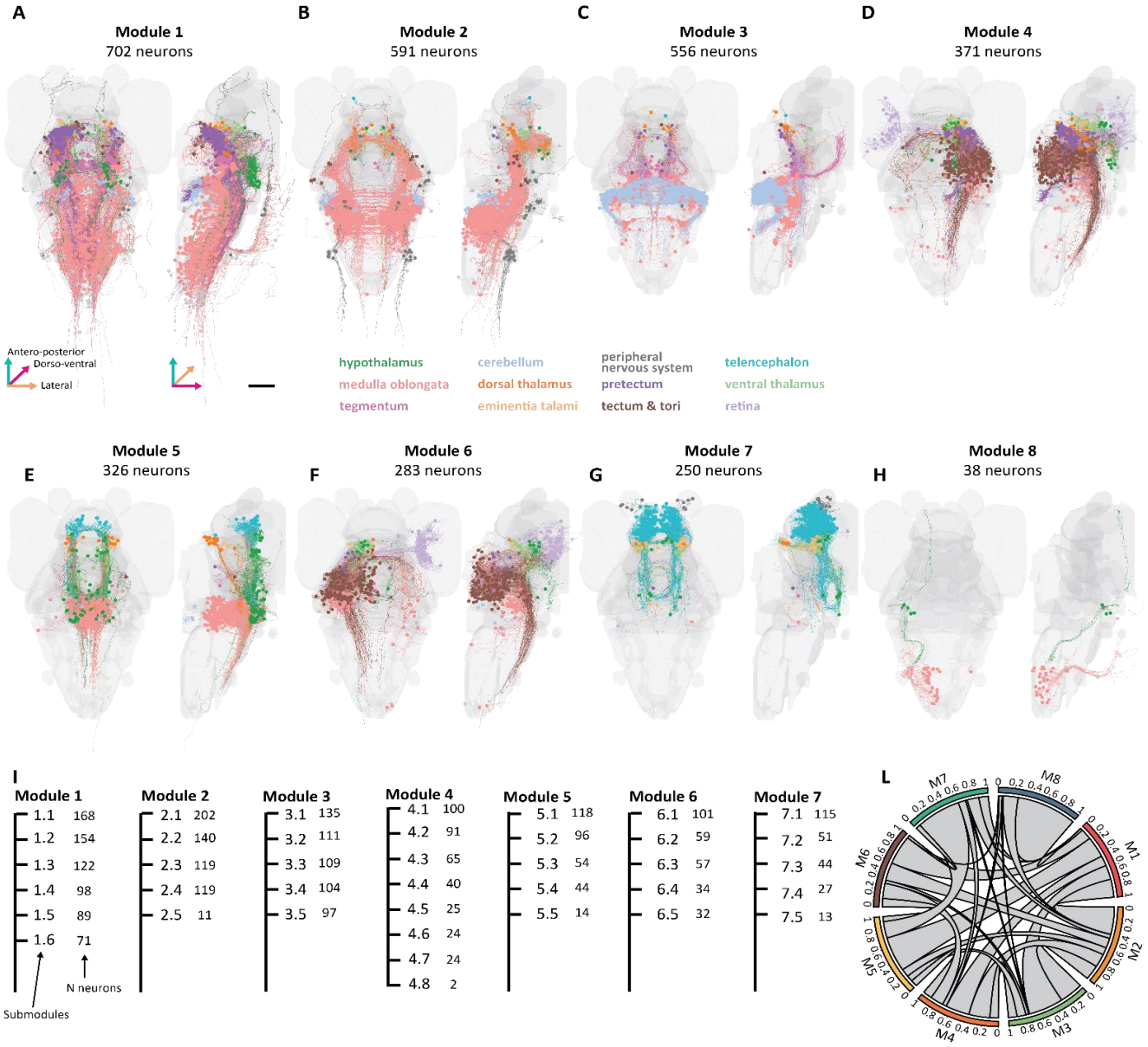


Figure 5. Modules of the zebrafish brain. A-H) *Anatomical representation of the eight high-order modules with the corresponding submodules found in the zebrafish dataset overlaid on the brain silhouette. Colors are assigned to the cells according to the soma position in the different brain regions.* Scale bar = 100 µm. ***I)*** *Tree diagram reporting the structure of the zebrafish network with the organization of the submodules for the different modules.* ***L)*** *Distribution of connections between the different modules.*

First, we considered the modules including the organization of the Retinal Ganglion Cells, (RGCs), projections to the tectum (modules 4 and 6). The analysis of the relationship of the RGC body position

244 in the retina and the connected PeriVentricular Neurons (PVNs) in the contralateral tectum showed a
245 retinotopic organization in submodules 4.2 and 4.6 (right tectum, Figure 5A), and in the sub modules
246 6.1 and 6.2 (left tectum) (Figure 5B). Indeed, these submodules show projections from RGC cells with
247 soma located in the caudal part of the retina to connect with PVNs in the rostral part of the contralateral
248 periventricular region at the level of the tectum and, conversely, for the RGC connections involving the
249 rostral part of the retina and the caudal part of the contralateral retino- recipient area. Similarly, our
250 analysis revealed an articulated spatial pattern of connections involving different brain areas associated
251 with the module 3, representing a pretectal-cerebellar network (Figure 6C). The connections formed
252 between neurons located in the medulla oblongata, cerebellum, and tegmentum (Figure 6C1,
253 submodules 3.3 and 3.4) resemble those associated with vestibulo-ocular reflex (VOR)[32]. In this case,
254 the ascending and ascending/descending tangential neurons are located in the medial octavolateral
255 nucleus (MON) of the medulla (Figure 6C1, submodules 3.3 and 3.4), the recipient of utricular inputs.
256 These neurons project to the oculomotor nucleus (nIII) where extraocular motor neurons mediate gaze
257 stabilization[38]. Furthermore, connections between neurons located in the cerebellum have been shown
258 to be involved in VOR[39,40]. The cerebellum displays a spatially segregated pattern of distinct cells, with
259 lateral Purkinje cells projecting to the MON (Figure 6C2, submodule 3.2) and eurydendroid cells located
260 in the middle with less specific projections[27,41] (Figure 6C2, submodules 3.1 and 3.5). Even though our
261 approach highlighted the existence of principal networks supported by modules of neurons
262 characterized by a relatively higher connection degree, it is clear that the different identified
263 subnetworks show mutual intermodule connection schemes. In order to quantify this aspect, we
264 evaluated the links formed across different modules (see Supplementary Figure 8). This analysis showed
265 a rather heterogeneous scenario, with some modules characterized by a relevant number of
266 intermodular connections, while others with a very limited number of intermodules links.

267

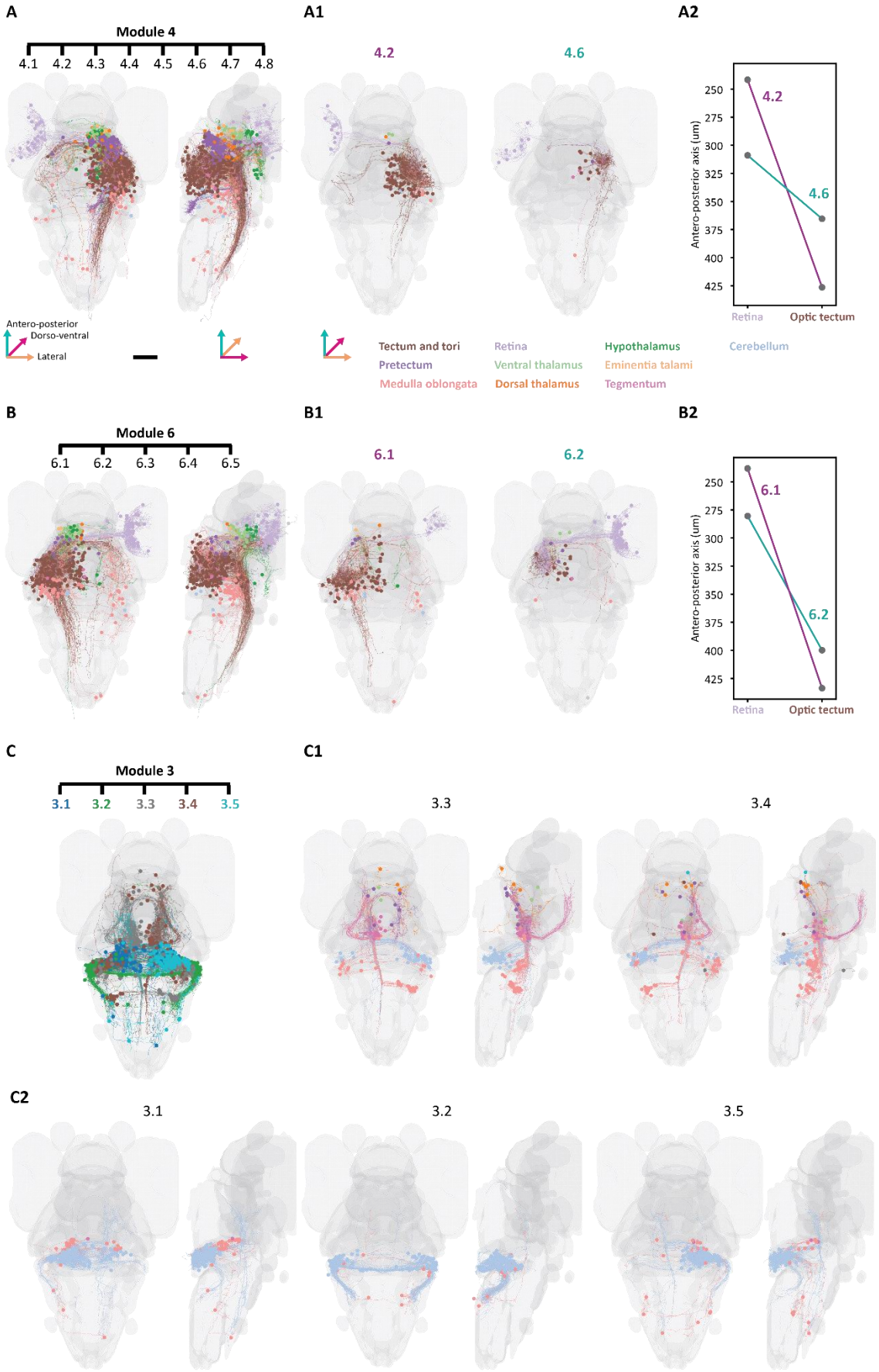

A
Module 4
4.1 4.2 4.3 4.4 4.5 4.6 4.7 4.8
Antero-posterior
Dorso-ventral
Lateral
A1
4.2
4.6
A2
Antero-posterior axis (um)
250 275 300 325 350 375 400 425
4.2
4.6
Retina
Optic tectum
Tectum and tori
Pretectum
Medulla oblongata
Retina
Ventral thalamus
Dorsal thalamus
Hypothalamus
Eminentia talami
Tegmentum
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

Figure 6. Functional description of the modules. 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.* ***A1)*** *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).* ***A2)*** *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)*** *Anatomical representation of the high-order module U6. Colors are assigned according to the soma distribution in the brain regions.* ***B1)*** *Sub-modules associated with the left Tectum opticum involving PeriVentricular Neurons and Retinal Ganglion Cells, highlighting two different sub-populations of neurons (6.1 caudal tectum, 6.2 medial tectum).* ***B2)*** *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)*** *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 lower-order modules. MON: medial octavolateral nucleus, nIII: oculomotor nucleus.* ***C1-C2)*** *Anatomical representation of the individual lower-order modules. Colors are assigned according to the location of the soma. Scale bar = 100 µm.*

CONCLUSIONS

We have devised a method to generate approximated connectomes from datasets of single cell reconstruction, covering neuronal circuits composed of different neuronal components represented in terms of single cell morphologies and extending across the complete brain. The proposed approach, inferring an approximated cell-to-cell contact structure on the basis of a spatial proximity rule, can be applied to all datasets with single cell reconstructions where synapse direct annotation is not accessible/available or is not yet complete. This typically is the case of anatomical reconstructions of sparse cells based on light microscopy[23–27]. Within the characteristic resolution limits dictated by the optical constraints, using laser scanning or light-sheet microscopy in combination with restricted or sparse fluorescence labelling, z-stacks of large portions of the brain or even of the complete brain are acquired to map the neuronal cell interregional connectivity or projection patterns[24,42]. In such scenarios, even considering a synapse tagging strategy in combination with pre- and post-synaptic partner labelling[43,44], the synaptic information retrieved could be partial or biased, so overlooking the involvement of other potential players. On the other hand, considering high-resolution, so-called "dense" tissue reconstructions based on serial block electron microscopy, the amount of data to process, even with artificial intelligence approaches, becomes easily relevant even for small brain sections[30]. While automatically reconstructing cell anatomies and their manual validation could become possible for small datasets, a complete synapse annotation may still represent a prohibitive challenge. In this case, one must either focus on a specific circuit or brain region or use the lower-hanging information available about the cell skeletons to extract approximated information about brain-wide circuit organization. The two approaches are not mutually exclusive, but rather answer two different types of questions. The first aims at characterizing the microcircuit structure or the circuit

303 working motifs[45]. The second, instead, seeks to understand the general organization of brain
304 networks[46].

305 We have validated the applicability of our approach by leveraging the synapse-annotated connectome
306 of the Drosophila melanogaster hemibrain, containing 21,739 connected neurons. Through modularity
307 analysis, we uncovered a hierarchical organization comprising eight main modules, each with several
308 submodules. When comparing the adjacency matrices generated from synapse annotations (EM-based)
309 and those derived using Strahler numbering and proximity range definitions (Strahler-based), we
310 observed significant correlations with respect two null models where links or weights have been re-
311 shuffled. Importantly, Adjusted Rand Index (ARI) and Normalized Mutual Information (NMI) scores
312 confirmed a general overlap between the Strahler-based and EM-based community structures, with
313 significant p-values supporting the robustness of the approach.

314 We then have applied our approach to a mesoscale atlas of about 3200 single cell reconstructions
315 obtained from a vertebrate organism, *Danio rerio,* where the input-output relationships were
316 estimated based on a proximity rule, and we have analyzed the graph properties of the inferred
317 connectome. From the corresponding adjacency matrices, we extracted the network communities by
318 modularity maximization. In this way, we were able to identify, according to a top-down process, those
319 neuronal modules characterized by a relatively higher degree of connections. With this approach and
320 its reiterations, we have obtained a decomposition of the connectomes of two phylogenetically distant
321 organisms on a set of basic components that form a sort of neuronal backbone of the brain wiring
322 diagram. For most of the identified modules, it is possible to associate a functional role, either based
323 on known anatomical characterizations of the circuits involved, or on the evidence provided by
324 functional analyses[27,28,38,41]. In conclusion, these results highlight the capability of our Strahler-based
325 method to approximate neuronal circuit architectures, demonstrating its potential as a scalable tool for
326 capturing connectivity patterns in large-scale brain networks. Moreover, the organization of the
327 modules obtained from the zebrafish larva light microscopy brain dataset shows that the application of
328 such top-down approach of community detection is able to capture relevant circuit motifs and may be
329 applied to several datasets at both the circuit and whole-brain scale.

330 **Methods**

331 **Dataset and data extraction**

332 We used a whole-brain atlas of the zebrafish larva[27]. The atlas comprises more than 3000 neurons,
333 approximately 3% of the total neurons in the zebrafish larvae brain, covering almost the entire brain,
334 anatomically reconstructed by random and sparse Green Fluorescent Protein (GFP) labelling using
335 BGUG[47] or with other labelling techniques. With a typical resolution of 300x300x1000 nm (X, Y, Z), the
336 size of the skeletonized dataset was approximately 60 MB. Skeletonized neurons have been
337 downloaded in the .swc format[48] from the mapzebrain website (https://mapzebrain.org/home).

338 **Computing the Strahler ordering for the neuronal anatomy reconstructions**

339 For a given neuronal skeleton i from the original dataset of 3163 reconstructions, we imported the
340 .swc file and used the "networkx" library (https://networkx.org/)[49] to convert the corresponding
341 collection of ordinated endpoints to a tree-like directed graph representing its anatomical structure.
342 The soma is considered as a root and the connected endpoints as children. In order to build an
343 approximated connectome from this dataset, we first classified each skeleton endpoint within its
344 branching hierarchy using Strahler ordering[28]. In this representation, terminal endpoints, those without
345 branching children, obtain a value one. For the other endpoints in the .swc file, the integer value of the
346 Strahler progressively increases with the number of upstream branches along the anatomical structure
347 to reach the maximum at the level of the soma. For the calculation of the putative synaptic contacts
348 and hence for the calculation of the connectome, we considered only endpoints with unitary Strahler
349 numbers (below referred to as endpoint unless otherwise specified), which we denoted as E_{μ}^{i}
350 where the italic superscript refers to *i*-th neuron within the dataset and the Greek subscript stands for
351 the μ-th endpoint of unitary Strahler number in the .swc file of *i*-th neuron.

352 **Comparison with the ground truth in Drosophila EM**

353 We considered the EM annotated connectome of the Drosophila hemibrain as ground-truth and
354 compared how close the reconstruction results when we applied on the same dataset our simplified
355 method based on Strahler numbering and proximity range (SP method). For that, we identified eight
356 modules using a community discovery approach on the EM connectome and considered the resulting
357 partitions in submodules, $P^{EM} = \{P_i^{EM} | i \in (1,..,8)\}$ where $P_i^{EM} = \{P_{i,j}^{EM} | j \in (1,..,..)\}$ represents the

358 *i*-module and $P_{i,j}^{EM}$ the *j*-sub-module associated with the module. Each P_i^{EM} was considered for
359 evaluating the sub-module partitions obtained based on the EM annotation ($P_{i,j}^{EM}$) or retrieved with the
360 approximated approach, $P_i^{EM} = \{P_{i,q}^{SP} | q \in (1,..,..)\}$, where $P_{i,q}^{SP}$ is the *q*-sub-module obtained with
361 the SP method in association with the reference module P_i^{EM}. For each module P_i^{EM}, we calculated the
362 corresponding ARI_i according to the following equation, as reported[31,50].

363
$$ARI_i\left(P_{i,j}^{EM}, P_{i,q}^{SP}\right) = \frac{\sum_{j,q}\binom{n_{jq}}{2} - [\sum_j \binom{a_j}{2} \sum_q \binom{b_q}{2}]/\binom{n}{2}}{\frac{1}{2}[\sum_j \binom{a_j}{2} + \sum_q \binom{b_q}{2}] - [\sum_j \binom{a_j}{2} \sum_q \binom{b_q}{2}]/\binom{n}{2}}$$

364 where n_{jq} represents the entries of the contingency matrix associated with the *i*-module, i.e. the
365 number of neurons that are belonging to $P_{i,j}^{EM} \cap P_{i,q}^{SP}$, a_j and b_q are the sum of n_{jq} along *q* and *j* indices
366 respectively, and *n* is the total number of partition elements. As for the NMI, we started by calculating
367 the probability $\pi_{jq} = \frac{n_{jq}}{n}$, as the probability to belong simultaneously to $P_{i,j}^{EM}$and $P_{i,q}^{SP}$, and the marginal
368 distributions computed as $\pi_{j*} = \sum_q \pi_{jq}$ and $\pi_{*q} = \sum_j \pi_{jq}$. We could then calculate the NMI as:

369
$$NMI_i\left(P_{i,j}^{EM}, P_{i,q}^{SP}\right) = \frac{-2\sum_{j,q} \pi_{jq} \log(\frac{\pi_{jq}}{\pi_{j*}\pi_{*q}})}{\sum_j \pi_{j*} \log(\pi_{j*}) + \sum_q \pi_{*q} \log(\pi_{*q})}$$

370

371 **Validation of ARI and NMI scores using Random null models**

372 For validating the NMI and ARI parameters, we constructed 100 random networks for each of the
373 mentioned 8 networks obtained with the SP method using two different approaches. In a first approach,
374 called "Link randomization or directed configuration", we used directed_edge_swap function from the
375 Python NetworkX library to swap the edges of the graph while keeping the node degree fixed. In a
376 second approach, called Weight Randomization, we kept the original links between the nodes
377 unchanged but changed the weight of the connection by randomly sampling from the original weight
378 distribution. The z-score was calculated for ARI and NMI considering two different proximity range, 1
379 and 5 μm, and the two types of random models as the difference of the between the actual value with
380 respect to the average value obtained from the 100 null-model networks, divided by the corresponding
381 standard deviation.

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[27,51,52]. 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 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.

Network properties

The analysis of network parameters[53] was performed using dedicated Python libraries. The following functions from the 'networkx'[49] (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[54] (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[55] $Q_{undir} = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \gamma \frac{k_i k_j}{2m} \right) \delta(\sigma_i, \sigma_j)$ and for the directed case[56] $Q_{dir} = \frac{1}{m} \sum_{ij} \left(A_{ij} - \gamma \frac{{k_i}^{out} {k_j}^{in}}{m} \right) \delta(\sigma_i, \sigma_j)$
, 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[57] 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.

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

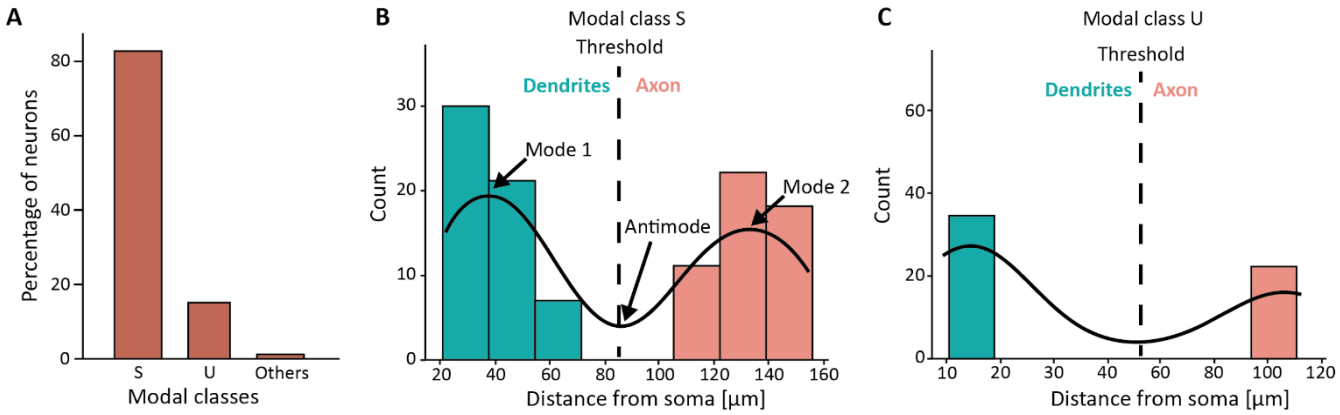
We thank Delfina Iriarte, Physics department of University of Padua and Nora Nikoloska, Mathematics Department of University of Padua for their support in preliminary analyses; we thank Prof. Tomaso Erseghe from the Information Engineering Department of University of Padua for useful discussion.

473 SUPPLEMENTARY FIGURES

474

475

476



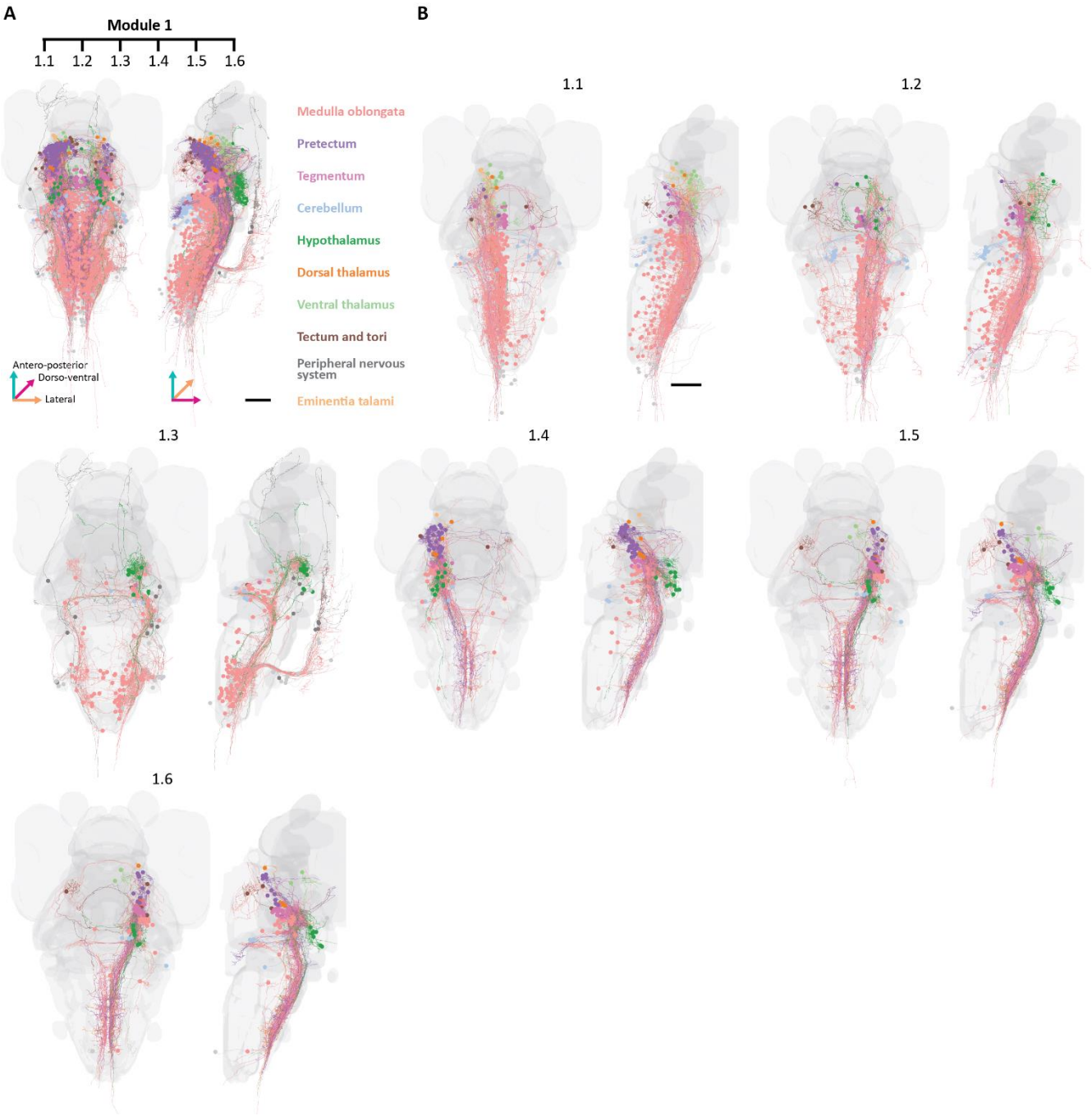
477

478 **Supplementary Figure 1. Distribution of endpoints of neurons. A**) Typically, the distance between the
479 endpoints of a neuron and the soma is distributed accordingly to two main distribution classes, following the
480 Galtung classification[40]. **B**) Bimodal class S. The antimode of the distribution is used to set a threshold (dashed
481 line) to distinguish between dendrites and axon. C) Bimodal class U. The antimode of the distribution is used to
482 set a threshold (dashed line) to distinguish between dendrites and axon.

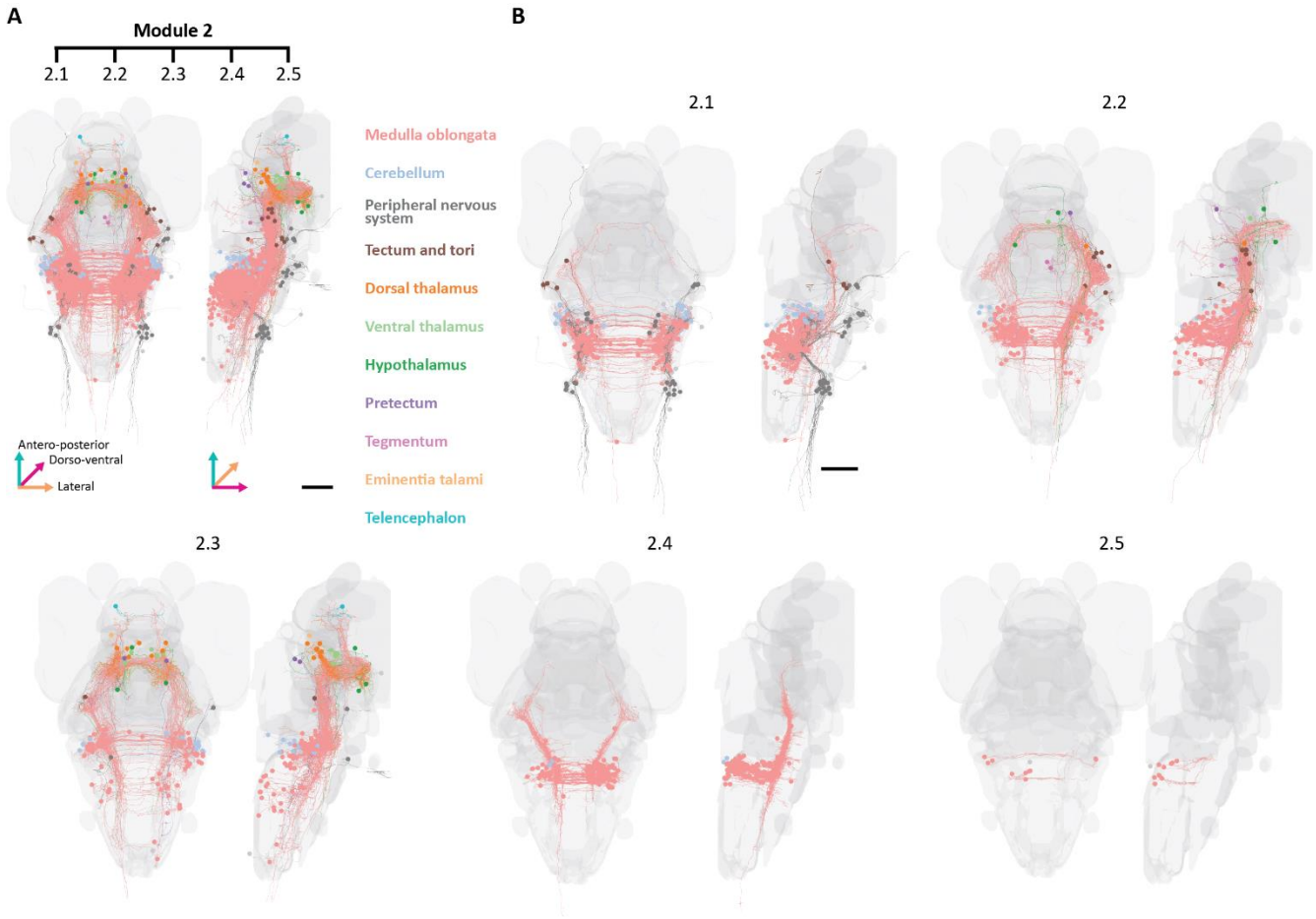
483

484

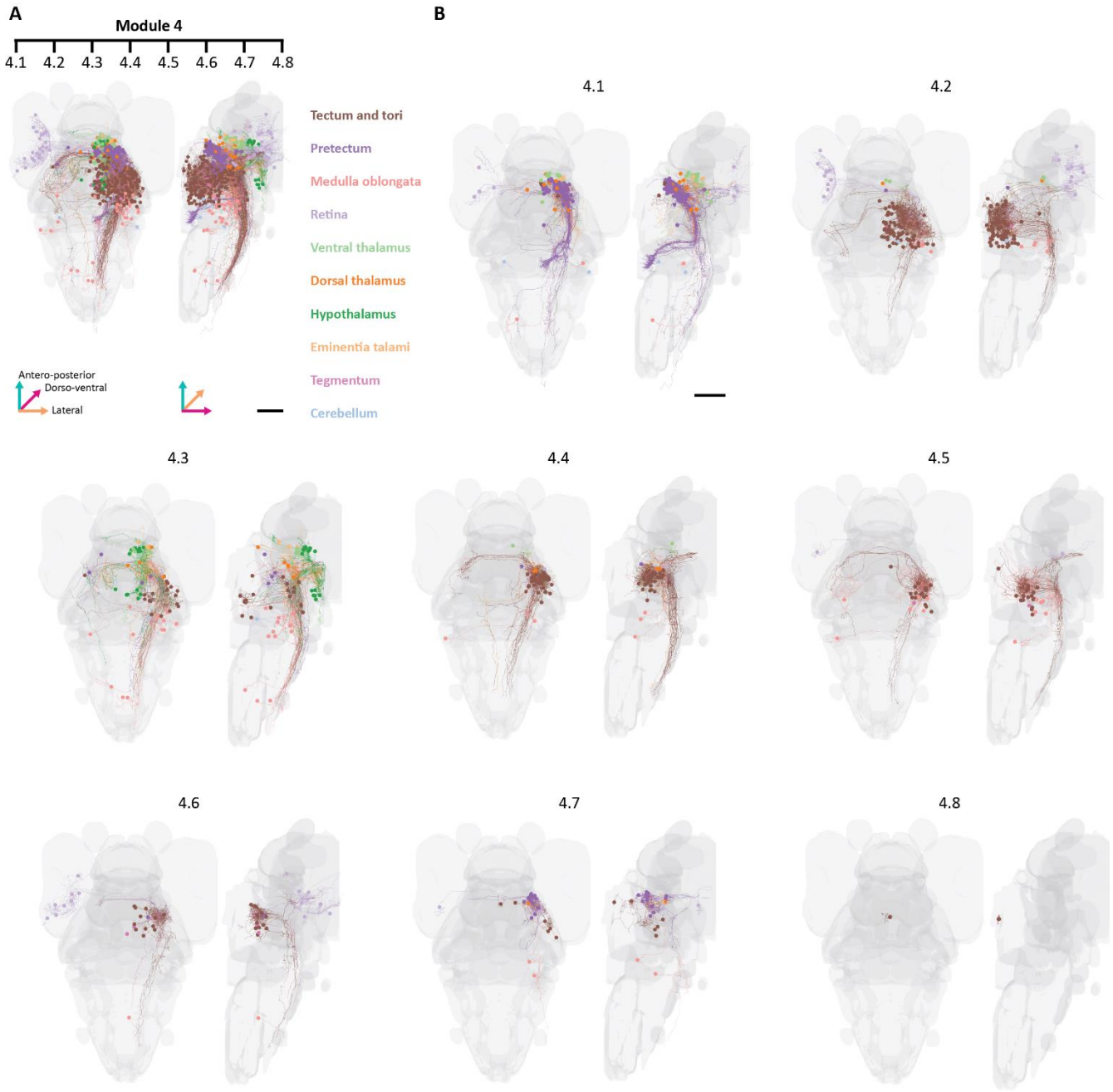
485



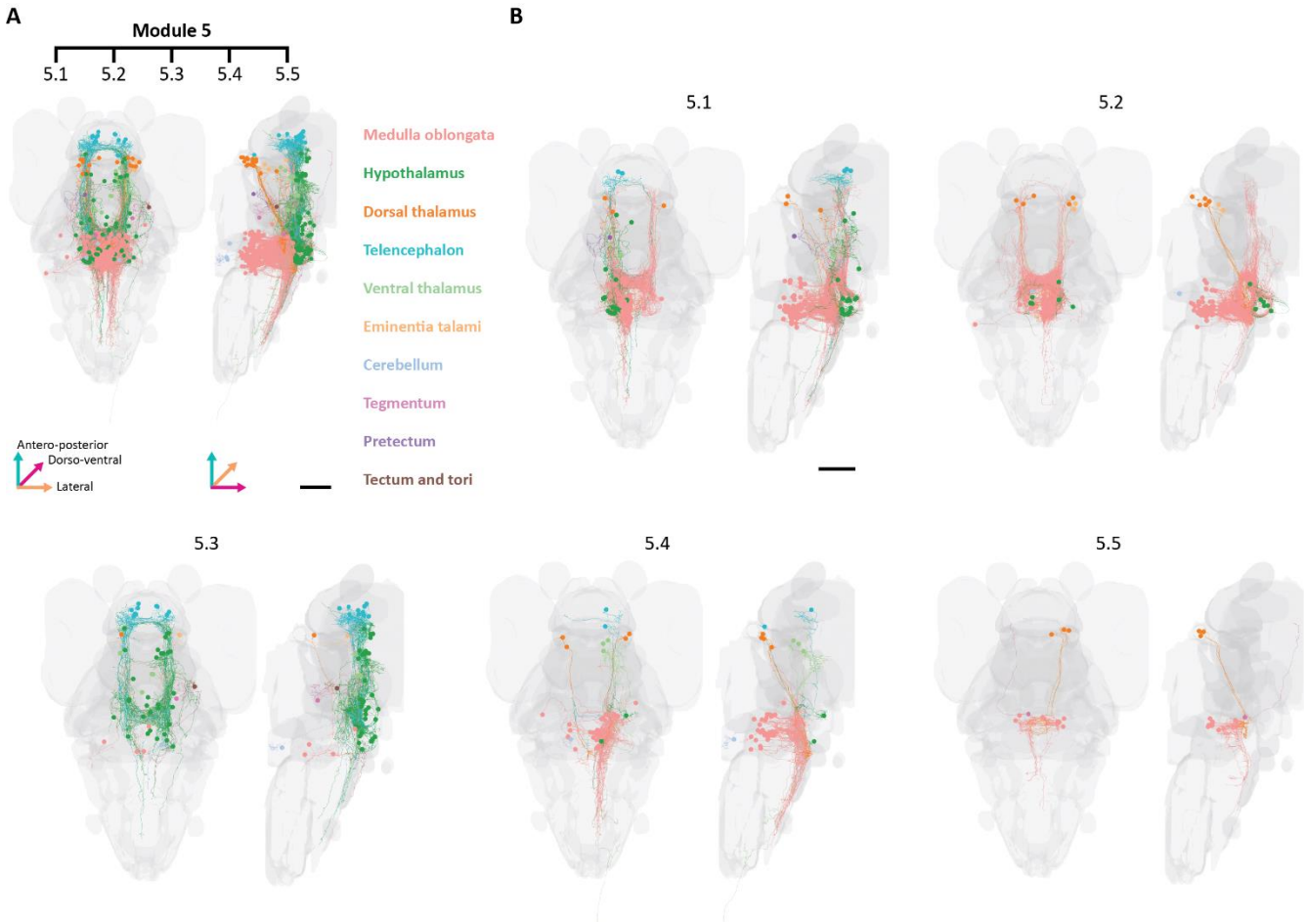
Supplementary Figure 2. The zebrafish module 1. A) Anatomical representation of the high-order module U1. Colors are assigned according to the soma distribution in the brain regions. Scale bar = 100 µm. **B)** Anatomical representation of the individual lower-order modules. Colors are assigned according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



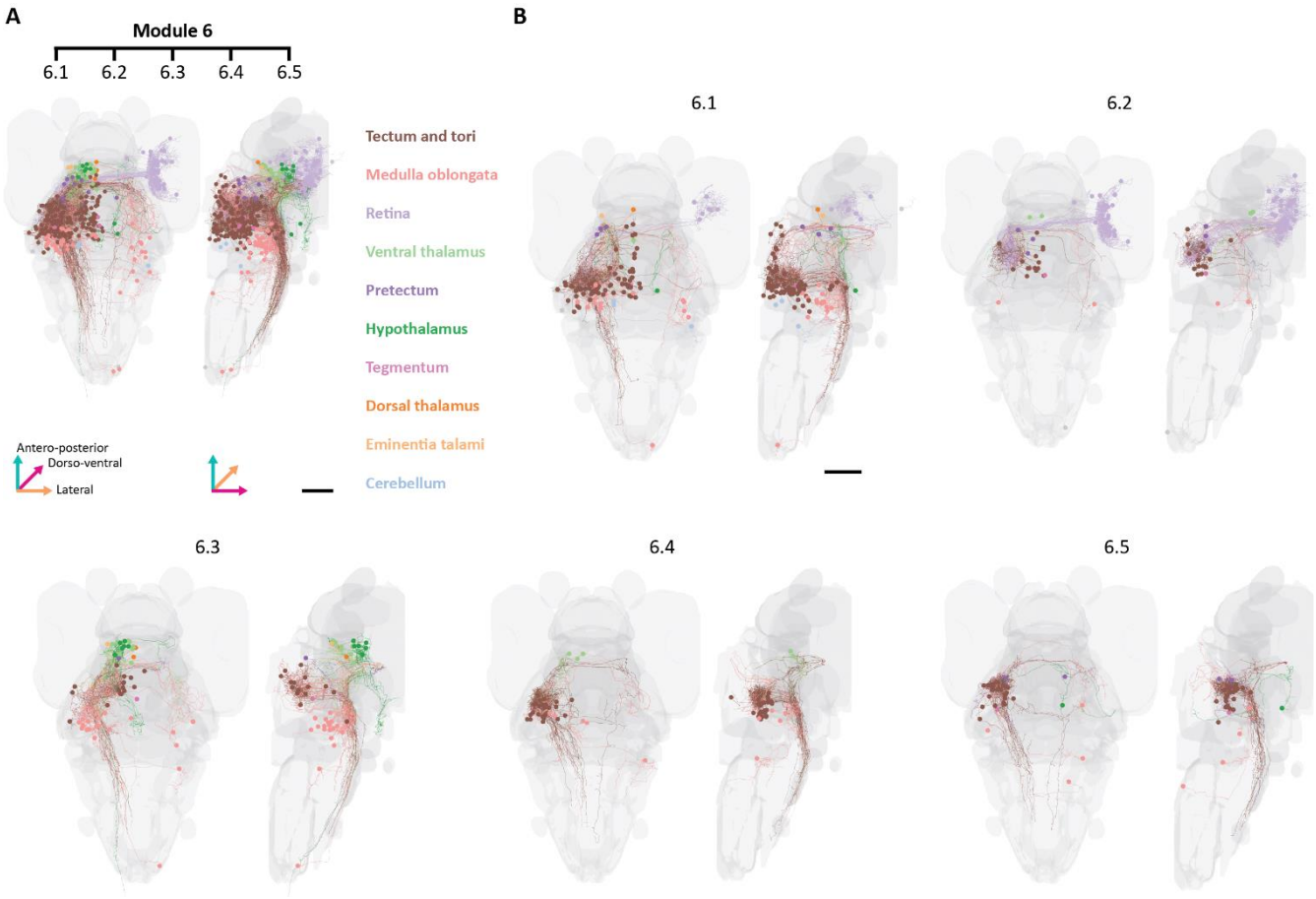
Supplementary Figure 3. The zebrafish module 2. A) Anatomical representation of the high-order module U2. Colors are assigned according to the soma distribution in the brain regions. Scale bar = 100 µm. **B)** Anatomical representation of the individual lower-order modules. Colors are assigned according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



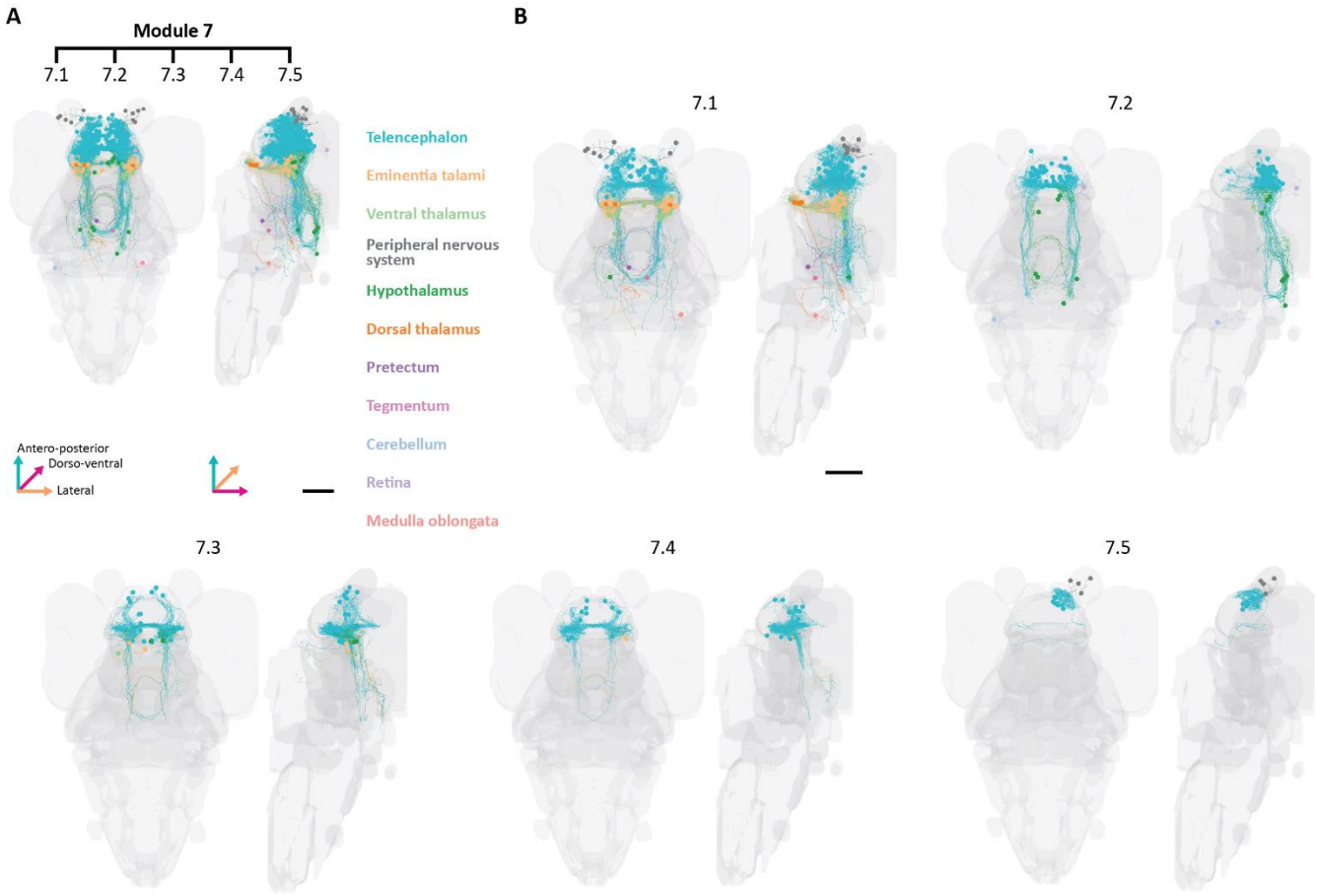
Supplementary Figure 4. The zebrafish module 4. A) Anatomical representation of the high-order module U4. Colors are assigned according to the soma distribution in the brain regions. Scale bar = 100 µm. **B)** Anatomical representation of the individual lower-order modules. Colors are assigned according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



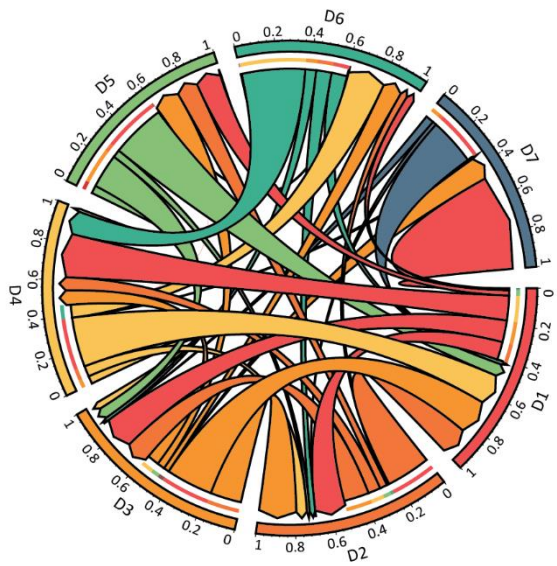
Supplementary Figure 5. The zebrafish module 5. A) Anatomical representation of the high-order module U5. Colors are assigned according to the soma distribution in the brain regions. Scale bar = 100 µm. **B)** Anatomical representation of the individual lower-order modules. Colors are assigned according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

Supplementary Information

Validation of ARI and NMI scores using Random null models, details

We considered the 8 networks from the SP method and let's represent them as $\{M_i^{SP} : 1 \leq i \leq 8\}$. Before proceeding to compute the z-scores, we constructed 100 random networks for each of the mentioned 8 networks of SP method using two unique random null models:

1. **Link randomization or directed configuration**[58]**:** In this method, we use *directed_edge_swap()* function of NetworkX library of python to Swap three edges in a directed graph while keeping the node degrees fixed. This method considers randomly chosen three connected edges such that $a \rightarrow b \rightarrow c \rightarrow d$ and after the swap is applied it becomes $a \rightarrow c \rightarrow b \rightarrow d$. In this way, we reach all possible states with the same in- and out-degree distribution in a directed graph and finally giving us directed configuration random null model.

2. **Weight randomization:** In this method, while keeping the links between the nodes unchanged or unswapped, we assign weight to each edge by sampling from weight distribution of all existing edge weights.

We computed the z-score for ARI and NMI for the 8 networks reconstructed with 1um proximity range and 5 μm proximity range using SP method. The mathematical formulation is as follows:

$$\text{z-score}(ARI,\ M_i,\ PR,\ \text{randomization type}) = \frac{ARI(M_i, PR) - \mu_{100}(M_i, PR, \text{randomization type})}{\sigma_{100}(M_i, PR, \text{randomization type})} \qquad [1]$$

where the notations refer to:

- M_i - refers to the module(set of nodes belonging to community i) for which we compute the similarity measures $ARI(P_i^{EM}, P_i^{SP})$ or $NMI(P_i^{EM}, P_i^{SP})$.
- PR - refers to proximity range which is either 1 μm or 5 μm used for reconstructing the network M_i^{SP} using SP method.
- randomization type - refers to the type of random null model used to randomize the network M_i^{SP} i.e. directed configuration model or weight randomization.
- μ_{100} and σ_{100} are the mean and standard deviations over ARI or NMI computed on 100 random networks. Specifically, it is $ARI(P_i^{EM}, P_i^{SP,rand})$ or $NMI(P_i^{EM}, P_i^{SP})$

The above formula similarly can be written by replacing it with NMI measure.

p-values computation

We have computed the p-values for the z scores(Eq. [1]) obtained for ARI or NMI in two ways:

$$\text{p-value}(M_i) = 1 - \Phi(\text{z-score}) \qquad [2]$$

where $\Phi(\quad)$ refers to cumulative distribution function of standard normal distribution.

$$\text{p-value}(M_i) = \frac{\sum_j 1_{\text{z-score}(rand_j) > \text{z-score}(M_i)} : 1 \le j \le 100}{\text{No. of realizations of random networks(i.e. 100)}} \qquad [3]$$

where $1 \le i \le 8$.

REFERENCES

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

582 13. Schaeffer, R., Khona, M. & Fiete, I. R. *No Free Lunch from Deep Learning in Neuroscience: A Case Study*
583 *through Models of the Entorhinal-Hippocampal Circuit*.
584 http://biorxiv.org/lookup/doi/10.1101/2022.08.07.503109 (2022) doi:10.1101/2022.08.07.503109.
585 14. Wanner, A. A., Genoud, C., Masudi, T., Siksou, L. & Friedrich, R. W. Dense EM-based reconstruction of the
586 interglomerular projectome in the zebrafish olfactory bulb. *Nat Neurosci* **19**, 816–825 (2016).
587 15. Svara, F. *et al.* Automated synapse-level reconstruction of neural circuits in the larval zebrafish brain. *Nat*
588 *Methods* **19**, 1357–1366 (2022).
589 16. White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of the
590 nematode Caenorhabditis elegans. *Philos Trans R Soc Lond B Biol Sci* **314**, 1–340 (1986).
591 17. Takemura, S. *et al.* The comprehensive connectome of a neural substrate for 'ON' motion detection in
592 Drosophila. *eLife* **6**, e24394 (2017).
593 18. Zheng, Z. *et al.* A Complete Electron Microscopy Volume of the Brain of Adult Drosophila melanogaster.
594 *Cell* **174**, 730-743.e22 (2018).
595 19. Schubert, P. J. *et al.* SyConn2: dense synaptic connectivity inference for volume electron microscopy. *Nat*
596 *Methods* **19**, 1367–1370 (2022).
597 20. Ryan, K., Lu, Z. & Meinertzhagen, I. A. The CNS connectome of a tadpole larva of Ciona intestinalis (L.)
598 highlights sidedness in the brain of a chordate sibling. *eLife* **5**, e16962 (2016).
599 21. Winding, M. *et al.* The connectome of an insect brain. *Science* **379**, eadd9330 (2023).
600 22. Dorkenwald, S. *et al.* Neuronal wiring diagram of an adult brain. *Nature* **634**, 124–138 (2024).
601 23. Arganda-Carreras, I. *et al.* A Statistically Representative Atlas for Mapping Neuronal Circuits in the
602 Drosophila Adult Brain. *Front Neuroinform* **12**, 13 (2018).
603 24. Oh, S. W. *et al.* A mesoscale connectome of the mouse brain. *Nature* **508**, 207–214 (2014).
604 25. Jenett, A. *et al.* A GAL4-driver line resource for Drosophila neurobiology. *Cell Rep* **2**, 991–1001 (2012).
605 26. Bogovic, J. A. *et al.* An unbiased template of the Drosophila brain and ventral nerve cord. *PLOS ONE* **15**,
606 e0236495 (2020).

27. Kunst, M. *et al.* A Cellular-Resolution Atlas of the Larval Zebrafish Brain. *Neuron* **103**, 21-38.e5 (2019).
28. Helmbrecht, T. O., dal Maschio, M., Donovan, J. C., Koutsouli, S. & Baier, H. Topography of a Visuomotor Transformation. *Neuron* **100**, 1429-1445.e4 (2018).
29. Galtung, J. *Theory and Methods of Social Research*. (Oslo, Universitetsforlaget; New York, Columbia University Press [etc.], 1967).
30. Scheffer, L. K. *et al.* A connectome and analysis of the adult Drosophila central brain. *Elife* **9**, e57443 (2020).
31. Hubert, L. & Arabie, P. Comparing partitions. *Journal of Classification* **2**, 193–218 (1985).
32. Rand, W. M. Objective Criteria for the Evaluation of Clustering Methods. *Journal of the American Statistical Association* **66**, 846–850 (1971).
33. Vinh, N. X., Epps, J. & Bailey, J. Information Theoretic Measures for Clusterings Comparison: Variants, Properties, Normalization and Correction for Chance. *Journal of Machine Learning Research* **11**, 2837–2854 (2010).
34. McDaid, A. F., Greene, D. & Hurley, N. Normalized Mutual Information to evaluate overlapping community finding algorithms. Preprint at https://doi.org/10.48550/arXiv.1110.2515 (2013).
35. Lancichinetti, A., Fortunato, S. & Kertész, J. Detecting the overlapping and hierarchical community structure in complex networks. *New J. Phys.* **11**, 033015 (2009).
36. Suárez, L. E. *et al.* Connectome-based reservoir computing with the conn2res toolbox. *Nat Commun* **15**, 656 (2024).
37. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep* **9**, 5233 (2019).
38. Bianco, I. H. *et al.* The Tangential Nucleus Controls a Gravito-inertial Vestibulo-ocular Reflex. *Current Biology* **22**, 1285–1295 (2012).
39. Matsui, H., Namikawa, K., Babaryka, A. & Köster, R. W. Functional regionalization of the teleost cerebellum analyzed in vivo. *Proceedings of the National Academy of Sciences* **111**, 11846–11851 (2014).

632 40. Favre-Bulle, I. A., Vanwalleghem, G., Taylor, M. A., Rubinsztein-Dunlop, H. & Scott, E. K. Cellular-Resolution
633 Imaging of Vestibular Processing across the Larval Zebrafish Brain. *Current Biology* **28**, 3711-3722.e3
634 (2018).
635 41. Bae, Y.-K. *et al.* Anatomy of zebrafish cerebellum and screen for mutations affecting its development. *Dev*
636 *Biol* **330**, 406–426 (2009).
637 42. Yao, S. *et al.* A whole-brain monosynaptic input connectome to neuron classes in mouse visual cortex. *Nat*
638 *Neurosci* **26**, 350–364 (2023).
639 43. Kim, J. *et al.* mGRASP enables mapping mammalian synaptic connectivity with light microscopy. *Nat*
640 *Methods* **9**, 96–102 (2011).
641 44. Shearin, H. K., Quinn, C. D., Mackin, R. D., Macdonald, I. S. & Stowers, R. S. t-GRASP, a targeted GRASP for
642 assessing neuronal connectivity. *J Neurosci Methods* **306**, 94–102 (2018).
643 45. Eschbach, C. *et al.* Multilevel feedback architecture for adaptive regulation of learning in the insect brain.
644 649731 Preprint at https://doi.org/10.1101/649731 (2019).
645 46. van der Plas, T. L. *et al.* Neural assemblies uncovered by generative modeling explain whole-brain activity
646 statistics and reflect structural connectivity. *Elife* **12**, e83139 (2023).
647 47. Scott, E. K. *et al.* Targeting neural circuitry in zebrafish using GAL4 enhancer trapping. *Nat Methods* **4**, 323–
648 326 (2007).
649 48. Cannon, R. C., Turner, D. A., Pyapali, G. K. & Wheal, H. V. An on-line archive of reconstructed hippocampal
650 neurons. *Journal of Neuroscience Methods* **84**, 49–54 (1998).
651 49. Hagberg, A. A., Schult, D. A. & Swart, P. J. Exploring Network Structure, Dynamics, and Function using
652 NetworkX. (2008).
653 50. Labatut, V. Generalized Measures for the Evaluation of Community Detection Methods. Preprint at
654 https://doi.org/10.48550/arXiv.1303.5441 (2016).
655 51. Vishwanathan, A. *et al.* Predicting modular functions and neural coding of behavior from a synaptic wiring
656 diagram. 2020.10.28.359620 Preprint at https://doi.org/10.1101/2020.10.28.359620 (2022).

52. Stepanyants, A. & Chklovskii, D. B. Neurogeometry and potential synaptic connectivity. *Trends Neurosci* **28**, 387–394 (2005).
53. Preface. in *Networks: An Introduction* (ed. Newman, M.) 0 (Oxford University Press, 2010). doi:10.1093/acprof:oso/9780199206650.002.0005.
54. Traag, V. *et al.* vtraag/leidenalg: 0.10.0. Zenodo https://doi.org/10.5281/zenodo.8147844 (2023).
55. Newman, M. E. J. & Girvan, M. Finding and evaluating community structure in networks. *Phys. Rev. E* **69**, 026113 (2004).
56. Leicht, E. A. & Newman, M. E. J. Community Structure in Directed Networks. *Phys. Rev. Lett.* **100**, 118703 (2008).
57. Reichardt, J. & Bornholdt, S. Statistical mechanics of community detection. *Phys. Rev. E* **74**, 016110 (2006).
58. Erdős, P. L., Miklós, I. & Toroczkai, Z. A simple Havel-Hakimi type algorithm to realize graphical degree sequences of directed graphs. Preprint at https://doi.org/10.48550/arXiv.0905.4913 (2010).