



Deposited via The University of Leeds.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/172908/>

Version: Accepted Version

Article:

Brittin, CA, Cook, SJ, Hall, DH et al. (2021) A multi-scale brain map derived from whole-brain volumetric reconstructions. *Nature*, 591 (7848). pp. 105-110. ISSN: 0028-0836

<https://doi.org/10.1038/s41586-021-03284-x>

© The Author(s). This is an author produced version of an article published in *Nature*.
Uploaded in accordance with the publisher's self-archiving policy.

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

1 **A multiscale brain map derived from whole-brain volumetric reconstructions**

2 **Christopher A. Brittin^{1,2‡}, Steven J. Cook^{3,♦}, David H. Hall³, Scott W. Emmons^{2,3†} and Netta**
3 **Cohen^{1†*}**

4 ¹School of Computing, University of Leeds, Leeds, LS2 9JT, UK

5 ²Albert Einstein College of Medicine, Dept. Genetics, New York, NY, 10461, USA

6 ³Albert Einstein College of Medicine, Dept. Neuroscience, New York, NY, 10461, USA

7 ‡Current address: Developmental Biology Program, Sloan-Kettering Institute,
8 New York, NY 10065 USA

9 ♦Current address: Department of Biological Sciences, Columbia University,
10 New York NY 10027, USA

11 †Senior authors

12 *Corresponding author

13

14 **ABSTRACT**

15 Animal nervous system organization is crucial for all body functions and its disruption can manifest in
16 severe cognitive and behavioral impairment¹. This organization relies on features across scales, from
17 nanoscale localization of synapses, through multiplicities of neuronal morphologies and their
18 contribution to circuit organization, to stereotyped connections between different regions of the brain².
19 The sheer complexity of this organ means that, to date, we have yet to reconstruct and model the
20 structure of a complete nervous system that is integrated across all these scales. Here, we present a
21 complete structure-function model of the nematode *C. elegans* main neuropil, the nerve ring, which we
22 derive by integrating the volumetric reconstruction from two animals with corresponding³ synaptic and
23 gap junctional connectomes. Whereas previously the nerve ring was considered a densely packed tract
24 of neural processes, we uncover internal organization and show how local neighborhoods spatially
25 constrain and support the synaptic connectome. We find that the *C. elegans* connectome is not invariant,
26 but that a precisely wired core circuit is embedded in a background of variable connectivity, and
27 propose a corresponding reference connectome for the core circuit. Based on this reference, we propose
28 a modular network architecture of the *C. elegans* brain that supports sensory computation and
29 integration, sensorimotor convergence and brain-wide coordination. These findings point to scalable
30 and robust features of brain organization that are likely universal across phyla.

31 Main

32 A primary goal of systems neuroscience is to understand how the brain's structure and function combine
33 to generate behavior. Since the discovery of neurons and their connections through synapses and gap
34 junctions, a major effort has focused on characterizing these units and the micro- and macro-circuits that
35 they comprise, culminating in a growing body of high-resolution nanoconnectomic data across species³⁻
36 ¹². Naturally, data, however rich, cannot on their own provide explanatory power to address the
37 computation within circuits or to determine how these circuits communicate and coordinate information
38 flow to generate behavior. Indeed, constructing a comprehensive brain map will require a meaningful
39 strategy for integrating structure and function across scales. Achieving this feat in even a small animal
40 can provide a useful model for postulating principles of brain organization across scales².

41 The free-living nematode *C. elegans* has a small, compact nervous system^{3,5,7,13} while exhibiting a
42 range of complex, individualized behaviors, making it an ideal model system for studies of whole brain
43 organization². All 302 *C. elegans* neurons have been anatomically characterized based on serial sectioned
44 electron micrographs (EM)⁵ to produce a whole animal connectome^{3,5,13}. This animal's invariant cell-
45 lineage¹⁴ and anatomy⁵ might suggest that its connectome too is invariant¹⁵. Unfortunately, the small
46 sample size of available reconstructions has precluded a reliable estimate of reproducibility and
47 variability of the synaptic connectome. Furthermore, while the synaptic wiring has been exhaustively
48 characterized^{3,5,13,16,17}, the spatial proximity of neurons is only partially determined^{18,19}. Thus, it remains
49 to be determined whether lessons about whole brain organization in *C. elegans* can inform questions and
50 approaches for other systems.

51 We provide two complete volumetric reconstructions of the *C. elegans* nerve ring from legacy EMs⁵,
52 from one adult and one larval stage 4 (L4) animal (Methods, Supplementary Table 1, Supplementary
53 Videos 1-3, Supplementary Information 1). The two EM series (with roughly 300 sections in the L4 and
54 400 in the adult) span approximately the same 36 μm long volume, starting in the anterior and ending in
55 the ventral ganglia (Fig. 1a). Our reconstructions provide the first *contactome*: a complete,
56 nanoresolution dataset of all neuronal membrane contacts in the nerve rings of these two animals. We
57 define two neurons as immediate neighbors if the membranes along their neural processes are physically
58 adjacent in at least one EM section¹⁸. To characterize synaptic pathways within a spatial context, we
59 integrated our volumetric reconstructions with our recent rescoring of synapses on the same L4 and adult
60 animals³ (for validation and comparison with other datasets^{5,20}, see Methods).

61

62 **Conserved and variable circuits overlap**

63 Consistent with White *et al.*⁵, our volumetric reconstructions show that neural processes are bilaterally
64 (left/right) conserved (Supplementary Results, Supplementary Videos 4-7). We hypothesized that the
65 bilateral symmetry of *C. elegans* processes extends to the nanoscale to support a homology of membrane
66 contacts and synapses between cells. Homologous processes exhibit statistically high overlaps in the size
67 and composition of their immediate neighborhood (Extended Data Fig. 1 a-c) and in membrane contact
68 locations along their processes (Methods, Extended Data Fig. 1 d-f, Supplementary Information 2). In
69 contrast, the smallest 35% of membrane contacts ($< 0.4 \mu\text{m}^2$) are not reproducible (Extended Data Fig.
70 2a), account for only 2% of total membrane contact area between all neurons (Extended Data Fig. 2b,c)
71 and contain predominantly nonreproducible synaptic contacts (Extended Data Fig. 2e). As such, we
72 exclude them from our analysis. We conclude that the reproducibility of neuronal processes and their
73 immediate neighborhoods supports a stereotyped pattern of cell-cell membrane contacts.

74 The availability of two reconstructions, combined with the bilateral homology of the nerve ring,
75 naturally lends itself to establishing a reference dataset that is more likely conserved across animals,
76 providing a basis to address mechanistic questions about precision and variability of the connectome at
77 nanoscale resolution. We defined the adjacency graph, \mathbb{M}^δ , of membrane contacts across 4 datasets (adult
78 left, adult right, L4 left and L4 right), where δ labels the number of datasets in which a membrane contact
79 occurs (Supplementary Information 3). The \mathbb{M}^4 reference dataset, i.e. the most reproducible membrane
80 contacts, comprises $\sim 40\%$ of all membrane contacts (Extended Data Fig. 2g) and exhibits above average
81 membrane contact area (Extended Data Fig. 2h). Adjacency graphs of chemical synapse, \mathbb{C}^δ , and gap
82 junction, \mathbb{G}^δ , contacts are similarly defined (Supplementary Information 3). We define \mathbb{M}^4 , \mathbb{C}^4 and \mathbb{G}^4
83 contacts as reference datasets and hypothesize that the \mathbb{M}^4 set of membrane contacts is representative of
84 the conserved membrane contacts across individuals in *C. elegans* and is more likely to support a
85 conserved synaptic connectome.

86 To examine this hypothesis, we exploit the combined spatial and synaptic information across datasets
87 over the entire neuropil. We assume that stereotyped wiring patterns require precision to find target
88 neurons and specificity to avoid off-target neurons, and formulate statistical models of membrane and
89 synaptic contacts to capture their relative propensity to occur in 1, 2, 3 or all 4 of the datasets (Methods).
90 We find that a minimal model with three parameters suffices (Methods); these are the fraction of target
91 contacts, f , the precision, p , for target contacts, and the frequency to avoid off-target contacts or
92 specificity, s . Despite their parsimony, these models yield good fits for the distribution of membrane,

93 synaptic and gap junctional contacts across the 4 datasets (Methods, Fig. 2a). The high reproducibility
94 of membrane contacts across datasets (M^4 count) is consistent with our model prediction that less than
95 half of membrane contacts are actively targeted ($f = 0.44$, Fig. 2a) with high precision ($p = 0.95$). The
96 significant variability across datasets is accounted for by a non-negligible basal membrane contact rate
97 ($1-s \sim 25-30\%$). Therefore, high precision combined with basal connectivity are required to account for
98 the reproducibility and variability of membrane contacts across datasets (Fig. 2a; Supplementary Results,
99 Extended Data Fig. 3).

100 How useful is the M^4 reference in predicting conserved membrane contacts? Our model predicts that
101 $\sim 99\%$ of the M^4 contacts and 68% of the M^3 contacts together constitute the vast majority ($\geq 98\%$) of the
102 core neuronal membrane adjacency matrix of the *C. elegans* nerve ring (Methods). Furthermore, above
103 average membrane contacts ($> 1.77 \mu m^2$) comprise more than 80% of M^4 contacts (Extended Data Fig.
104 2h) and are more reproducible (with higher precision, $p = 0.98$, and larger fraction, $f = 0.77$, Extended
105 Data Fig. 3a-b). We conclude that the M^4 dataset offers an excellent candidate set of conserved membrane
106 contacts. While highly reproducible, core membrane contacts are not easily distinguished from variable
107 ones. Our model predicts that $\sim 50\%$ of membrane contacts are variable across animals. Using model-
108 generated surrogate datasets (Methods), we estimate that 20 datasets (from 10 animals, with 2 datasets
109 per bilateral reconstruction) would suffice to identify all core membrane contacts in the *C. elegans* nerve
110 ring (Fig. 2b).

111 To model synaptic and gap-junctional precision, we re-fit the model to \mathbb{C}^δ and \mathbb{G}^δ (Methods). To
112 control for synaptic variability due to differences in process placement, we restricted our analysis to M^4
113 contacts (for a more general treatment, see Extended Data Fig. 4a-c). Even among reproducible
114 membrane contacts, our model predicts that high precision ($p > 0.90$) combined with basal connectivity
115 ($1-s \sim 20-30\%$) are required to account for the reproducibility and variability of synaptic and gap
116 junctional contacts across datasets (Fig. 2a; Supplementary Results, Extended Data Fig. 4d-e). For the
117 bilateral worm, a synaptic precision of 93% implies a $\sim 99\%$ probability of a core synaptic contact
118 occurring at least once per animal (on the left, right or both sides), and $\geq 97\%$ chance to occur in at least
119 3 of 4 datasets (across 2 animals). Conversely, we predict that $\sim 98\%$ of \mathbb{C}^4 and \mathbb{G}^4 are good
120 representatives of the core circuit (as well as $> 60\%$ of \mathbb{C}^3 and \mathbb{G}^3), lending further confidence to the
121 usefulness of the reference connectome. However, the placement of the most reproducible synapses
122 along the process is not restricted to reproducible membrane contact sites (Extended Data Fig. 1g-h).
123 Thus, location along the process cannot be used to distinguish core from variable synapses. Taken

124 together, these results demonstrate that each dataset can be divided into a common, precisely targeted
125 core circuit and a variable component, and that, given additional connectomes, it should be possible to
126 distinguish between them (Extended Data Fig. 4d,e).

127 We next asked what principles of spatial organization support the reproducible, highly specified
128 neuronal placement in the nerve ring. To address this question, we noted that the observed variability of
129 membrane contacts suggests that no one animal is representative of the population at large and even core
130 contacts likely vary across individuals (Extended Data Fig. 4h,i). Hypothesizing that conserved
131 membrane contacts form the basis of the neuropil organization, we estimated the expected variability in
132 our reference contacts across a population of animals by computing the variability in M^4 contact areas
133 across the four datasets at our disposal. We used the reference membrane contact distributions and their
134 associated membrane contact area variability across the datasets to generate stochastic population models
135 of core membrane contacts from the L4 and adult bilateral datasets and the M^4 reference dataset
136 (Methods, Extended Data Fig. 5). To group together neurites with high spatial affinity, we used a multi-
137 level graph clustering algorithm²¹ on each individual in our population model (Methods). We find that 5
138 subgroups of neurons consistently emerge from the data whose processes are spatially ordered along the
139 anterior-posterior axis of the nerve ring (Fig. 1a, Supplementary Information 4). We label these clusters
140 *anterior*, *lateral*, *sublateral*, *avoidance* and *taxis* (Supplementary Results). Regionalization of processes
141 in the nerve ring into the anterior circuit (associated with mechanosensation), the posterior, amphid
142 neural circuit (associated with chemosensation and navigation) and lateral and sublateral neurons
143 (associated primarily with head motor control) has previously been highlighted⁵. Our quantitative
144 analysis is consistent with the above description but our focus on the core nanostructure reveals finer
145 organization of the nerve ring that may not be apparent from the raw volumetric data (Methods, Extended
146 Data Fig. 6).

147 We asked whether the cluster organization of the nerve ring is indicative of modularization of synaptic
148 pathways^{3,13,18,19,22}. We find that most neurons have strong membrane and synaptic contacts within a
149 single cluster, whereas others physically and synaptically contact neurons across multiple clusters (Fig.
150 3a-b, Extended Data Fig. 7). However, synaptically sparse lateral neurons and a number of neurons that
151 closely link across the lateral and sublateral neighborhoods suggest that lateral and sublateral clusters
152 may be merged for purposes of information processing analysis.

153 Neurons that synapse across clusters are often characterized by processes that change neighborhood
154 along their trajectories (Fig. 3c-e, Extended Data Fig. 8k). We identified 33 cell classes whose processes

155 synapse across different regions of the nerve ring (Methods). These cell classes use two principal
156 strategies: synapse compartmentalization (19/33 cell classes, Supplementary Information 4, Fig. 3e) and
157 flattened protrusions (23/33 cell classes, Supplementary Information 4). We hypothesize that a subset of
158 neurons synaptically link different neighborhoods of the nerve ring to support brain-wide coordinated
159 activity²³. Consistent with our hypothesis, these specialized spatial features and the synapses they support
160 are largely conserved across our 4 datasets. In summary, we find that the nerve ring obeys a consistent
161 set of spatial organization principles across scales, including a macroscopic modular neighborhood
162 organization which supports the mesoscopic organization along neurites, microscopic precision of
163 membrane contacts and nanoscopic morphological features, that together support conserved synaptic
164 wiring.

165

166 **A *C. elegans* brain map**

167 We integrate the knowledge gained to map the architecture of the *C. elegans* brain: The high-level spatial
168 organization (Fig. 1a) – the ‘macro-connectome’² – suggests modular circuits, with distinct functional
169 roles. Neuronal organization within and across spatial regions, comprising predominantly local and some
170 cross-cutting neurons (Fig. 3a-c) that exhibit micro- and nanoscale structures (Fig. 3e-h), allows us to
171 map the coordination across the nerve ring. Our reference connectome allows us to focus on reliable,
172 likely conserved connectivity (Fig. 2). Finally, classification of neurons as sensory, interneuron and
173 motoneuron allows us to trace sensorimotor pathways within and across these modules. By combining
174 these features in the data, we set out to construct a brain map of the *C. elegans* nerve ring.

175 We posit a parsimonious 3-layer architecture with parallel information processing modules and
176 assign every neuron of the nerve ring into a layer roughly corresponding to the 5 neuron clusters
177 (Methods). To achieve overall feed-forward pathways, sensory neurons all occupy the first layer whereas
178 spatially cross-cutting neurons dominate layer 3 (Methods, Fig. 4). Connectomic features, identified from
179 network analysis of the *C. elegans* connectome (such as highly connected ‘hub’ neurons, high
180 assortativity hubs known as ‘rich-club neurons’^{24,25}, network motifs^{13,26} and the small-world
181 organization¹³ as well as new features such as fan-in and fan-out motifs¹³ (characterized by higher in- or
182 out-degrees, respectively, Extended Data Fig. 9a) can now be interpreted within the context of modular,
183 brain-wide computation and information flow (Supplementary Results, Extended Data Fig. 9b-j). In
184 particular, the feed-forward loop motif, previously identified in the *C. elegans* connectome^{3,13,26},
185 reappears in our map as the skeleton of the layered synaptic pathways within each module (>50% of all

186 C^4 contacts; Fig. 4a, Extended Data Fig. 10 shows additional contacts). This system-wide feed-forward
187 connectivity is reminiscent of the layered connectivity of pyramidal neurons in the mammalian cortex
188 and its biologically inspired analogue – Residual Networks (ResNets)²⁷. Such architectures have been
189 conjectured to enhance the resilience of synaptic pathways and to support flexibility and plasticity²⁷.

190 Examination of the *C. elegans* brain map (Fig. 4b) reveals a number of features. Layer 1 separates the
191 modules (with a few notable and functional exceptions, Extended Data Fig. 10). The intra-module, intra-
192 layer connectivity indicates that sensory neurons likely perform limited sensory computation in addition
193 to sensory encoding of environmental cues, and allows the identification of sensory hub (high-degree)
194 neurons (Supplementary Results). Layer 2 largely maintains the modular synaptic information flow.
195 Convergence of sensory neurons onto this sparser layer reveals a fan-in architecture, supporting modular
196 sensory integration (Supplementary Results, Extended Data Fig. 9c-d). Layer 3 contrasts with the above.
197 Inputs are received from all three layers: Synapses from layers 1 and 2 comprise the core of each module,
198 whereas layer-3 synapses interlink and couple the modules, forming a recurrent, highly distributed circuit,
199 consistent with the dominance of spatially complex neurons in this layer and suggestive of brain-wide
200 coordination roles (Extended Data Fig. 9e-j). Outputs from the nerve ring control the pharynx, head and
201 neck muscles and the motor circuit of the ventral nerve cord (VNC). The taxis and avoidance modules
202 support distinct information pathways (Fig. 4b) despite responding to overlapping sensory cues and both
203 synapsing onto the VNC command interneurons. In contrast, the sublaterals highlight cross-connectivity
204 within the nerve ring, with all but two neuron classes occupying layer 3. Pharyngeal output is mediated
205 by layer-2 anterior neurons, indicating that the pharyngeal control is independent of the distributed layer-
206 3 circuit. In contrast, head and neck muscles are controlled by layer-3 anterior, lateral and sublateral
207 neurons and the VNC is controlled by all layer-3 modules, revealing the convergence of sensory pathways
208 and associated modular subcircuits into a small number of highly coordinated motor programs.

209 Discussion

210 The *C. elegans* connectome has been available for over 30 years, and yet the delineation of functions
211 within its main neuropil is still incomplete. By characterizing the spatial embedding of its connectome,
212 we sought insight into the structures that could support a hierarchical, modular and nested architecture
213 in the *C. elegans* brain. Previous analyses of the *C. elegans* connectome identified a common feed-
214 forward loop motif among triplets of neurons^{3,26}. Our brain map recasts this local motif as an architectural
215 motif, reminiscent of layered cortical architectures²⁸ and their artificial analogue, Residual Networks²⁷.

216 Such a ‘connectionist’ description of a biological brain provides a promising methodology for identifying
217 parallel and distributed circuits.

218 While there are no physical boundaries within the nerve ring, our analysis points to spatial clustering
219 of neural processes into five neighborhoods. The parallel pathways in our brain map largely fall into this
220 modular neighborhood organization, linking spatial and functional organization. The spatial organization
221 may also reflect developmental roles of nerve ring pioneers²⁹ and constraints on synaptic and
222 neuromuscular connectivity for motor coordination and control functions. Within the Residual-Network
223 template are intra-layer local circuits, whose neurons by-and-large lack structural or functional
224 compartmentalization. Thus, consistent with the neuron doctrine, within local subcircuits, neurons
225 represent the basic unit of computation. However, the modular architecture converges within the final
226 layer to achieve brain-wide coordination of behavior. In this distributed circuit, the nanoconnectome
227 rules: specialized subcellular structures give rise to compartmentalized dynamics and interlink distant
228 regions of the *C. elegans* brain. Similar subcellular structures performing analogous functions, found in
229 thalamic local interneurons³⁰, reveal a richness of subcellular computation. Thus, brain-wide
230 coordination may be achieved by designated processes that interface between or thread across multiple
231 subcircuits to underpin sensory convergence and sensorimotor transformations. The *C. elegans* brain
232 map and its nested architecture might suggest a much closer analogy between the *C. elegans* neuropil
233 and the coordination between the nano- and macro-connectomes of other invertebrates and even
234 vertebrates³¹.

235 The concept of a reference connectome was key to our brain map and the modeling framework we
236 used to establish this reference can easily be extended to accommodate future connectomes. In
237 vertebrates, nanoscale organization underpinning individual synapses is variable, supporting individual
238 wiring, plasticity and adaptability. In *C. elegans*, the proportion of conserved synapses was unknown.
239 We found that the connectome consists of a core, conserved circuit that is embedded in a significant
240 variable background. While pinning down the extent of the variable circuit is challenging due to the
241 technical limitations of synaptic scoring and will therefore require multiple further connectomes, it is
242 noteworthy that conserved synapses, like most variable ones, are constrained by the same connectome.
243 Thus, if the core circuit represents the baseline functionality of the animal, the variable component could
244 support redundancy, individuality³² and plasticity⁶.

245 The large number of cell classes, so densely packed in the nerve ring, presents a challenge to
246 physically achieving stereotyped connectivity. Our finding of finely orchestrated organization across

247 scales imposes spatial constraints on neurite and synaptic placement, thus restricting each neuron's
248 connectivity problem to a local neighborhood. This scalable solution is robust across a large population
249 and naturally generalizes to much larger nervous systems. Viewed differently, the spatial organization
250 reduces the required capacity for cell-cell molecular recognition machinery, while increasing the
251 complexity of mechanisms producing the cell's morphology and relative positioning in the tissue. But
252 how is the neighborhood organization developmentally orchestrated? Previous models of neuropil
253 development have proposed that pioneer neurites guide follower neurons³³. While such models could be
254 generalized to identify the pioneers of each neighborhood³⁴, the highly reproducible pattern of membrane
255 contacts indicates a more elaborate developmental mechanism. In complementary models, some
256 guidance molecules would coordinate the relative neighborhood placement and others – the placement
257 of neurites³³. Identifying key guidance molecules in early nerve ring formation may help to address such
258 predictions^{29,33}. Whatever the developmental mechanisms may be, the brain map of *C. elegans* requires
259 that these mechanisms too are nested and coordinated across scales to guide and support the modular,
260 scalable and flexible neural architecture that produces the mind and behavior of the nematode *C. elegans*.
261

262 **References**

- 263 [1] Hahamy, A., Behrmann, M. & Malach, R. The idiosyncratic brain: Distortion of spontaneous
264 connectivity patterns in autism spectrum disorder. *Nat. Neurosci.* **18**, 302-309 (2015).
- 265 [2] Swanson, L. W. & Lichtman, J. W. From Cajal to connectome and beyond. *Ann. Rev. Neurosci.*
266 **39**, 197–216 (2016).
- 267 [3] Cook, S. J. et al. Whole-animal connectomes of both *Caenorhabditis elegans* sexes. *Nature* **571**,
268 63–71 (2019).
- 269 [4] Ryan, K., Lu, Z. & Meinertzhagen, I. A. The CNS connectome of a tadpole larva of *Ciona*
270 *intestinalis* (L.) highlights sidedness in the brain of a chordate sibling. *eLife* **5**, e16962 (2016).
- 271 [5] White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of
272 the nematode *Caenorhabditis elegans*. *Phil. Trans. R. Soc. Lond. B.* **314**, 1–340 (1986).
- 273 [6] Hall, D. H. & Russell, R. L. The posterior nervous system of the nematode *Caenorhabditis*
274 *elegans*: serial reconstruction of identified neurons and complete pattern of synaptic interactions.
275 *J. Neurosci.* **11**, 1–22 (1991).
- 276 [7] Jarrell, T. A. et al. The connectome of a decision-making neural network. *Science* **337**, 437–444
277 (2012).
- 278 [8] Bumbarger, D. J., Riebesell, M., Rodelsperger, C. & Sommer, R. J. System-wide rewiring
279 underlies behavioral differences in predatory and bacterial-feeding nematodes. *Cell* **152**, 109–
280 119 (2013).
- 281 [9] Ohshima, T. et al. A multilevel multimodal circuit enhances action selection in *Drosophila*. *Nature*
282 **520**, 633–639 (2015).
- 283 [10] Zheng, Z. et al. A complete electron microscopy volume of the brain of adult *Drosophila*
284 *melanogaster*. *Cell* **174**, 730–743 (2018).
- 285 [11] Kasthuri, N. et al. Saturated reconstruction of a volume of neocortex. *Cell* **162**, 648–61 (2015).
- 286 [12] Motta, A. et al. Dense connectomic reconstruction in layer 4 of the somatosensory cortex. *Science*
287 **366**, eaay3134 (2019).
- 288 [13] Varshney, L. R., Chen, B. L., Paniagua, E., Hall, D. H. & Chklovskii, D. B. Structural properties
289 of the *Caenorhabditis elegans* neuronal network. *PLoS Comput. Biol.* **7**, 21 (2011).
- 290 [14] Sulston, J. E., Schierenberg, E., White, J. G. & Thomson, J. N. The embryonic cell lineage of the
291 nematode *Caenorhabditis elegans*. *Dev. Biol.* **100**, 64–119 (1983).
- 292 [15] Barabási, D. L. & Barabási, A-L. A genetic model of the connectome. *Neuron* **105**, 435–445
293 (2020).
- 294 [16] Albertson, D. G. & Thomson, J. N. The Pharynx of *Caenorhabditis elegans*. *Phil. Trans. R. Soc.*
295 *Lond. B.* **275**, 299–325 (1976).

- 296 [17] Cook, S. J. et al. The connectome of the *Caenorhabditis elegans* pharynx. *J. Comp. Neurol.* **528**,
297 2767–2784 (2020).
- 298 [18] White, J., Southgate, E., Thomson, J. & Brenner, S. Factors that determine connectivity in the
299 nervous system of *Caenorhabditis elegans*. *Cold Spring Harb. Symp. Quant. Biol.* **48**, 633–640
300 (1983).
- 301 [19] Durbin, R. M. *Studies on the development and organisation of the nervous system of*
302 *Caenorhabditis elegans*. Ph.D. thesis, University of Cambridge (1987).
- 303 [20] Witvliet, D. et al. Connectomes across development reveal principles of brain maturation in *C.*
304 *elegans*. Preprint at <https://www.biorxiv.org/content/10.1101/2020.04.30.066209v1> (2020).
- 305 [21] Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in
306 large networks. *J. Stat. Mech.* **2008**, P10008 (2008).
- 307 [22] Gray, J. M., Hill, J. J. & Bargmann, C. I. A circuit for navigation in *Caenorhabditis elegans*. *Proc.*
308 *Natl. Acad. Sci. USA* **102**, 3184–91 (2005).
- 309 [23] Kato, S. et al. Global brain dynamics embed the motor command sequence of *Caenorhabditis*
310 *elegans*. *Cell* **163**, 656–669 (2015).
- 311 [24] Towilson, E. K., Vertes, P. E., Ahnert, S. E., Schafer, W. R. & Bullmore, E. T. The rich club of the
312 *C. elegans* neuronal connectome. *J. Neurosci.* **33**, 6380–6387 (2013).
- 313 [25] Cohen, N. & Denham, J. E. Whole animal modeling: piecing together nematode locomotion.
314 *Curr. Opin. Syst. Biol.* **13**, 150–160 (2019).
- 315 [26] Milo, R. et al. Network motifs: simple building blocks of complex networks. *Science* **298**, 824–
316 827 (2002).
- 317 [27] He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. *Proc. IEEE*
318 *Comp. Soc. CVPR*, 770–778 (2016).
- 319 [28] Thomson, A. M. Neocortical layer 6, a review. *Front. Neuroanat.* **4** (2010).
- 320 [29] Rapti, G., Li, C., Shan, A., Lu, Y. & Shaham, S. Glia initiate brain assembly through noncanonical
321 Chimaerin–Furin axon guidance in *C. elegans*. *Nat. Neurosci.* **20**, 1350–1360 (2017).
- 322 [30] Morgan, J. L. & Lichtman, J. W. An individual interneuron participates in many kinds of
323 inhibition and innervates much of the mouse visual thalamus. *Neuron* **106**, 468–481 (2020).
- 324 [31] Chen, X. et al. Brain-wide organization of neuronal activity and convergent sensorimotor
325 transformations in larval zebrafish. *Neuron* **100**, 876–890 (2018).
- 326 [32] Stern, S., Kirst, C. & Bargmann, C. I. Neuromodulatory control of long-term behavioral patterns
327 and individuality across development. *Cell* **171**, 1649–1662 (2017).
- 328 [33] Wang, L. & Marquardt, T. What axons tell each other: Axon-axon signaling in nerve and circuit
329 assembly. *Curr. Opin. Neurobiol.* **23**, 974–982 (2013).

330 [34] Moyle, M. W. et al. Structural and developmental principles of neuropil assembly in *C. elegans*.
331 Preprint at <https://www.biorxiv.org/content/10.1101/2020.03.15.992222v1> (2020).

332

333 **Fig. 1. Five densely connected neurite clusters comprise the nerve ring neuropil. a**, The nerve ring
 334 neuropil (<4% of the worm's body length and most synaptically dense region of the nervous system)
 335 includes neurites of 181 L4 (185 adult) neurons. Complete volumetric reconstruction of the L4 neuropil
 336 spans 36 μm (Supplementary Video 3). 15 μm -long region (inset): left view, superficial neurons
 337 removed. D: dorsal, V: ventral, A: anterior. **b**, A 250 nm oblique volumetric slice at approximately the
 338 lateral midline (LM) rendered with no processes removed (right). A/P: anterior/posterior, M/L:
 339 medial/lateral, LG/VG: lateral/ventral ganglia. Scale bar: 1 μm . Neurites with relatively high spatial
 340 affinity (but no physical boundaries) form spatially ordered clusters along anterior-posterior axis. **c**,
 341 Cluster matrix: frequency that cells i and j cluster together across the population $\widetilde{\mathbb{M}}^4$: Row and column
 342 order minimized frequency variance along the diagonal. Clusters were then ordered to visually match
 343 AP ordering (original ordering in Extended Data Fig. 5i). Top: Dendrogram of the hierarchical
 344 clustering. **d**, Clustering results of model $\widetilde{\mathbb{M}}^4$, $\widetilde{\text{L4}}$ and $\widetilde{\text{Adult}}$ populations (Extended Data Fig. 5i) and
 345 consensus cluster assignment across the 3 populations. Row and column order same as rows in **c**. 7 cell
 346 classes (ADE, ALN, AVA, RID, RIR, RMD, URX) with discrepant cluster assignments among the 3
 347 populations are unclassified (gray). $n = 1000$ perturbed datasets per population (Methods).

349 **Fig. 2. The nerve ring is comprised of a core circuit embedded in a variable background. a**,
 350 Empirical data and model fits for the reproducibility, across δ datasets, of membrane, \mathbb{M}^δ (top), synaptic,
 351 \mathbb{C}^δ (middle) and gap junction, \mathbb{G}^δ (bottom) contacts. Empirical and model frequency distributions
 352 normalized by the total empirical contact count, n (e.g. for membrane contacts, $n = \sum_{\delta=1}^4 \mathbb{M}^\delta$). **b**,
 353 Surrogate data for 4, 20, 100, 1000 datasets (2, 10, 50 and 500 model animals). 4 datasets suffice to
 354 deduce that the distribution is bimodal. 20 datasets (10 animals) would suffice to completely distinguish
 355 between the core and variable subcircuits. No contact is expected to occur across 1000 datasets (500
 356 animals). Target contacts comprise $\sim 73\%$ of each dataset.

357 **Fig. 3. Nano-, micro- and meso-scale process structure supports local and nonlocal connectivity. a**,
 358 Matrix of \mathbb{C}^4 synaptic contacts (pink). Rows/Columns: pre-/post-synaptic cells. 4 zones around the main
 359 diagonal delineate growing neighborhoods around each cell with: Zone 0: average immediate
 360 neighborhood sizes in \mathbb{M}^4 . Zones 1-3: Zone 0 plus 1, 2 and 3 standard deviations, respectively; Zone 4:
 361 remaining \mathbb{C}^4 contacts not in Zones 0-3. **b**, 75% of \mathbb{C}^4 synaptic contacts form locally within Zones 0-2. n :
 362 empirical count of \mathbb{C}^4 synaptic contacts. **c**, Fraction of intra-cluster and intra-zone \mathbb{C}^4 contacts (total, n).
 363 Many nonlocal (Zones 3-4) synaptic contacts occur with neurons that traverse different neighborhoods,
 364 exhibit flattened protrusions and/or exhibit synaptic compartmentalization. **d**, Synaptic
 365 compartmentalization: RIA synaptic polarity varies with changes in cluster assignment of neighboring
 366 cells. White/black arrows label synaptic polarity (inputs/outputs). **e-g**, Volumetric rendering of selected
 367 processes shows local structures that support localized, reproducible synapses. **e**, RMDV flattened
 368 protrusions support synapses onto RMDD, diversifying synaptic polarity. **f**, RIM protrusions support
 369 synapse onto RIB. **g**, Spine-like extensions (dashed black arrows) from RMEV cell body support
 370 synapses from SMBVL (not shown) and SMBVR. All examples observed bilaterally in L4 and adult
 371 (Extended Data Fig. 9, additional examples in Supplementary Information 4). Row/Column (**a**) and cell
 372 (**d-g**) colors denote cluster assignment except **e**, RMDD (anterior cluster, yellow). Scale bar: 1 μm .

374

375 **Fig. 4. The *C. elegans* brain map. a**, A 3-layer, modular Residual Network architecture²⁷ (solid arrows
376 and recurrent connectivity in layer 3) captures 78% of \mathbb{C}^4 synaptic contacts in the nerve ring: parallel
377 feed-forward loop motifs converge onto layer 3, supporting functional sensorimotor pathways. Layer-3
378 interneurons and motoneurons (with \mathbb{C}^4 contacts across multiple zones, Fig. 3a) form a distributed
379 circuit across all modules. Dashed arrows: intra-module feedback (5% of \mathbb{C}^4). n : empirical count of \mathbb{C}^4
380 synaptic contacts. **b**, All 80 bilateral neuron classes and 11 single neurons (AVL and RID lack \mathbb{C}^4
381 contacts) overlaid on the network architecture (**a**). Sensory neurons (triangles, layer 1); interneurons
382 (ovals); motoneurons (rectangles). Except CEPD, module assignment matches cluster. CEPD (anterior
383 module, sublateral cluster) shares the same process looping trajectories as and synapses more
384 extensively with anterior sensory cells (Extended Data Fig. 10). Unclassified cells (gray) module
385 placement based on process trajectory. Black arrows: intra-module synaptic contacts (thickness
386 proportional to synapse size, aggregate number of EM sections where synapses were scored).
387

388 **Methods**

389 **Anatomical and neuron-class nomenclature**

390 The anatomy of the *C. elegans* nerve ring, associated ganglia and the delineation of 6 nerve bundles
391 entering the nerve ring, was described in detail by Ware *et al.*³⁵. Early observations, e.g. the distinction
392 between papillary and amphid sensory specializations and their postulated mechano- and chemo-sensory
393 roles have been validated since. Individual cell classes were identified and named by White *et al.*⁵. Each
394 neuron name consists of either two or three uppercase letters indicating class and in some cases a number
395 indicating the neuron number within one class (e.g. IL1, IL2). Bilaterally symmetric neurons (cell pairs)
396 have a three letter/number class name followed by L (left) or R (right). Radially symmetrical neurons
397 (with either 4 or 6 members) have a three-letter name followed by D (dorsal), or V (ventral), L (left) or
398 R (right) (e.g. SIADL, SIADR, SIAVL, SIAVR and RMDL, RMDR, RMDDL, RMDDR, RMDVL,
399 RMDVR). Unless otherwise noted, we use the term class synonymously with bilateral cell pair for
400 radially symmetric cell classes (e.g. SIAV and SIAD are treated as separate classes). Additionally, 17
401 nerve ring neurons constitute the only members of their class (ALA, ALM, ALN, AQR, AVL, AVM,
402 DVA, DVC, PVT, PVR, RID, RIH, RIR, RIS, RMED, RMEV and SABD). A small number of VNC
403 motoneurons also enter the nerve ring. These VNC motoneurons names consists of two uppercase letters
404 indicating muscle innervations (V: ventral, D: dorsal) and class (A-C) and a number indicating the neuron
405 within one class (counted from anterior to posterior). Neurons are designated as sensory neurons,
406 interneurons or motoneurons following their primary descriptions in WormAtlas³⁶ (excluding
407 proprioception from the sensory designation). We note, however, that in *C. elegans*, these designations
408 are not exclusive. Our use of the term neighborhood to describe processes that run closely together in the
409 nerve ring follows White *et al.*^{5,18}. We use the stronger term immediate neighborhood to designate neural
410 processes that make physical contact.

411

412 **Electron micrograph (EM) preparation**

413 The two legacy electron micrographs (EM) series used in this study were constructed in the MRC
414 Laboratory of Molecular Biology (Cambridge, UK) during the 1970s. Both series are of hermaphrodite
415 worms of the wild-type N2 (Bristol) strain. Worms were fixed in 1% osmium tetroxide in 0.1 M sodium
416 phosphate, pH 7.5 for 1 h at 20°C before embedding, sectioning and post-staining⁵. This method was
417 previously determined to best bring out cell membranes and synaptic structures at the expense of features
418 within the cytoplasm. The EM series are transverse to the longitudinal axis of the worm; estimated

419 section thickness is 70-90 nm, judged by silver color³⁷. The original 55 cm × 60 cm montaged prints
420 covering the nerve ring commissure and 30 cm × 40 cm covering the posterior lobe of the nerve ring
421 have since been digitized, archived in the Hall Laboratory and available at www.wormimage.org.

422 The two series reconstructed for this study include the synaptically dense nerve ring neuropil and
423 ventral ganglia regions of the anterior nervous system. One series is from a larval stage 4 (L4) worm and
424 the other series is from an adult (estimated three days from adulthood³), referred to as JSH and N2U,
425 respectively. The JSH series extends from just anterior of the nerve ring to the excretory pore. The N2U
426 series is substantially longer, extending from just anterior of the nerve ring to the vulva. We only
427 considered the section of the N2U series that physically corresponds to the JSH series. This resulted in
428 302 sections in the N2U series compared to 410 sections in the JSH series. In N2U, starting at the nerve
429 ring posterior lobe, only every other EM section was imaged (N2U EM sections 183-302). Additionally,
430 it is speculated that the JSH images may have slightly smaller section thickness. To correct for this when
431 making comparisons between the L4 and the adult, data from this region in N2U was scaled by a factor
432 of 2.

433

434 **EM segmentation**

435 EMs were manually segmented using TrakEM2 software³⁸. The software provides GUI tools to facilitate
436 the segmentation of cells across an EM stack. Within the EM series, we segmented all neuronal cell
437 bodies and processes that extend into the nerve ring (Supplementary Videos 1 and 2). Cell bodies were
438 then removed from our membrane contact analysis, because their large sizes skew the cell contact
439 distribution. We also segmented the portion of the pharynx in the nerve ring, which serves as both a
440 visual reference and spatial reference for the cylindrical coordinates. We did not segment dendrites of
441 sensory neurons, because dendrites have very few synapses and therefore were not of interest for our
442 analysis. We also did not reconstruct the sublateral cells SABVL and SABVR whose anterior processes
443 leave the ventral nerve cord via the amphid commissure⁵. Measurements of the membrane contact
444 between neurons were taken directly from the TrakEM2 XML data. We estimated each pixel to be ~5
445 nm², based on size measurements of cell bodies which are estimated to be 2-3 μm wide. In all, we
446 segmented 181 and 185 cells that innervate the L4 and adult nerve rings, respectively (the “complete
447 dataset”).

448

449 **Extracting adjacency data**

450 We developed custom software (parsetrakem2, <https://github.com/cabrittin/parsetrakem2>) to quantify
451 the pairwise membrane contacts between TrakEM2 segmented processes. In each EM, TrakEM2 stores

452 each segmented cell as a set of boundary points. For each segmented cell, i , our software defines a search
453 radius that is proportional to the diameter of the segmented cell i . Any immediately neighboring
454 segmented cell, j , that has a boundary point within the search radius is checked for adjacency to i . We
455 define adjacency for the pair (i, j) as the number of boundary points of j that are less than 10 pixels (~50
456 nm) from the boundary points of i . We found that a radius of 10 pixels was sufficiently large to ensure
457 that adjacencies were not missed. Any cell pairs erroneously identified as adjacent could easily be
458 screened out in downstream analysis based on membrane contact area.

459 To check the accuracy of the algorithm, for two EM sections, we compared the contacts scored by
460 our software to those obtained from manual scoring of membrane contacts (Supplementary Table 2). For
461 manual scoring of membrane contacts, we used the *connector* feature in TrakEM2 to generate a
462 connectivity graph of adjacent cells. An EM section with n cells has $n(n-1)/2$ possible cell pairings that
463 were then classified as either adjacent (if the cells touch) or non-adjacent (if the cells do not touch). We
464 assume the manually scored contacts to be the ‘ground truth’, which we use to define true positives (TP),
465 false positives (FP), true negatives (TN) and false negatives (FN) in our automated classification.
466 Sensitivity, defined as $TP/(TP + FN)$, measures how likely two physically touching cells are classified
467 by our software as adjacent. Specificity, defined as $TN/(TN + FP)$, measures how likely two separate
468 cells are classified by our software as non-adjacent. Aggregating results from the two manually scored
469 EM sections (JSH001 and JSH040), the sensitivity and specificity of our classification algorithm are
470 0.974 and 1.00, respectively (Supplementary Table 2). In other words, the algorithm will miss ~2.5% of
471 adjacent cell pairs within an EM section and a negligible number (<0.05%) of separate cell pairs will be
472 incorrectly classified as adjacent. We assessed the missed adjacent cells in our test set and found that
473 these adjacencies were small (tens of nanometers) and resulted primarily from poor segmentation (the
474 manual cell segmentation did not extend completely to the cell membrane). Furthermore, all 11 cell pairs
475 incorrectly classified as non-adjacent in the two test EM sections were correctly classified as adjacent in
476 subsequent EM sections. As adjacent cell pairs missed in one EM section are likely to be correctly
477 classified as adjacent in subsequent EM sections, and because most of our analyses aggregate adjacencies
478 across EM sections, any missed adjacencies within a single EM section is likely to have negligible impact
479 on our results.

480 As an additional test, we compared the adjacent cells extracted by our algorithm to the adjacent cells
481 previously reported for a small subset of neurons based on a sparse analysis of physical adjacency in the
482 L4¹⁸. White *et al.* determined the neighborhoods of cells AIAR, AIBR and AQR in the L4 (JSH) EM

483 series. Our algorithm was able to find all but 1 adjacent cell pair in the White *et al.*¹⁸ neighborhoods
484 (Supplementary Information 5). Closer inspection revealed that the cell pair does not make physical
485 contact and was thus mis-scored by White *et al.* as adjacent. Furthermore, we found an additional 69
486 adjacent cell pairs not included in the White *et al.* neighborhoods (Supplementary Information 5).
487 Therefore, our volumetric dataset is more extensive than those previously reported.

488 **EM annotation for synaptic connectivity**

489 We used our previously published connectivity data for chemical synapses and gap junctions and refer
490 the reader to Cook *et al.*³ for details on how synapses were annotated. Briefly, we used custom software³⁹
491 to aid manual annotation of chemical synapses and gap junctions. For chemical synapses, presynaptic
492 cells are identified by the presence of a presynaptic density while postsynaptic cells are identified as the
493 cells directly apposed to the presynaptic density. Most synapses are polyadic – multiple postsynaptic
494 partners are assigned to a single presynaptic cell. Gap junctions are recognized as a straightened or slightly
495 curving region of apposed membranes with increased staining and a uniform small gap. For the purpose
496 of the current study, we restrict the synaptic and gap junctional dataset to those in our volumetric
497 reconstruction (i.e. those scored in EMs that were segmented for this study). In all, the numbers of
498 synapses and gap junctions scored is larger than in the original connectome⁵, with a notable increase in
499 synapses that were scored in only 1 EM section. Within our reference \mathbb{C}^δ dataset (see *Generating reference*
500 *graphs* below), Cook *et al.* (2019)³ scored an additional 489 synaptic contacts to the White *et al.* (1986)⁵
501 connectome, of which 249 (49%) synaptic contacts only occur in 1 EM section. To control for the
502 possibility of false positives in this annotation, more restricted datasets were constructed for validation
503 (see *Validation against test datasets*, below).

504

505 **Generating reference graphs**

506 In order to control for variations in connectivity, we found it useful to map the data to a novel data
507 structure, which we call a reference graph. Reference graphs classify contacts (defined as the aggregate
508 pairwise connections over all EM sections within a dataset) based on their degree of reproducibility across
509 datasets. We took advantage of the bilateral symmetry of the worm to effectively double our sample size.
510 We therefore generated 4 datasets (adult left/right and L4 left/right) from the two reconstructed nerve
511 rings. For a sample size of $n = 4$, simply averaging across datasets is not a useful way to build a reference

512 model of the data. Instead, we segregate the contacts into 4 separate categories based on their
 513 reproducibility. To this end, we removed from our analysis a number of neurons that exhibit appreciable
 514 differences in synaptic connectivity or process morphology laterally (PLN, PVN, HSN), between the L4
 515 and adult (HSN, PVR, SABD), or those that make minimal membrane contact in the nerve ring (in VB,
 516 VC and VD classes), leaving 173 cells in 93 cell classes (the “restricted dataset”, Supplementary
 517 Information 3). The restricted dataset excludes HSNR, PLNL, PLNR, PVNL, PVR, SABD, VB01 and
 518 VD01 neurons – in both L4 and adult – and HSNL, PVNR, VB02 and VC01 – in the adult.

519 We generate reference graphs as follows. We first threshold membrane contacts by eliminating the
 520 smallest 35% of contacts in each of the adult and L4 datasets (Extended Data Fig. 2). From these, we
 521 then generate 4 datasets of membrane contacts: adult left, adult right, L4 left and L4 right. Each dataset
 522 was converted to a graph, where vertices are neurons and edges denote membrane contacts between a
 523 pair of adjacent neurons. The reference graphs \mathbb{M}^1 , \mathbb{M}^2 , \mathbb{M}^3 and \mathbb{M}^4 represent the set of membrane contacts
 524 found in $\delta = 1, 2, 3$ and all 4 datasets (see explicit calculation of reproducibility degree, δ , below).
 525 Reference graphs for chemical synapses (\mathbb{C}) and gap junctions (\mathbb{G}) were generated similarly, but with
 526 slightly different edge thresholding. Whereas for \mathbb{M} , we thresholded based on the magnitude of
 527 membrane contact, for \mathbb{C} and \mathbb{G} we only included edges that correspond to \mathbb{M}^4 contacts (or from \mathbb{M}^3 or
 528 \mathbb{M}^2 where explicitly mentioned). By only including edges in \mathbb{M}^4 , we effectively eliminate differences in
 529 synaptic connectivity due to differences in process placement. Each edge in the membrane reference
 530 graph (\mathbb{M}^δ) has an associated normalized mean contact area (across the 4 datasets). To control for slight
 531 differences in cell sizes between the larva and adult series, we normalize all membrane contact areas
 532 within each of the 4 datasets by the sum of all membrane contacts within that dataset. The normalized
 533 membrane contact area between neurons (i, j) in \mathbb{M}^δ is then the mean normalized contact area across the
 534 δ datasets in which the contact is present.

535 For bilateral cell classes, let indices, e.g. i and j , each denote some side of an animal (left or right)
 536 and let \bar{i}, \bar{j} , etc. denote the respective contralateral side. For a contact $\{X_i^1, Y_j^1\}$ made between cell X_i in
 537 class X to Y_j in class Y in animal 1, δ is defined as the number of contacts among
 538 $(\{X_i^1, Y_j^1\}, \{X_i^1, Y_j^1\}, \{X_i^2, Y_j^2\}, \{X_i^2, Y_j^2\})$ where the superscript 2 labels the other animal. For intra-class
 539 connections, δ is the number of contacts among $(\{X_i^1, X_i^1\}, \{X_i^1, X_i^1\}, \{X_i^2, X_i^2\}, \{X_i^2, X_i^2\})$, and for
 540 single cell classes, e.g. DVA connecting to some class Y (or vice versa), δ is counted among
 541 $(\{DVA^1, Y_j^1\}, \{DVA^1, Y_j^1\}, \{DVA^2, Y_j^2\}, \{DVA^2, Y_j^2\})$ (or vice versa).

542

543 **Population spatial models**

544 The observed variability in membrane contacts, both bilateral and across the two animals, indicates that
545 it is unlikely that any one animal is representative of the population at large. We generated a population
546 model of all membrane contacts, by stochastically perturbing the area associated with each membrane
547 contact, such that the overall distribution of mean membrane contact areas is preserved and that the
548 variability in membrane contact areas across datasets is also preserved. To establish the baseline
549 variability across the 4 datasets, we considered the log-normalized distribution of \mathbb{M}^4 membrane contact
550 areas (Extended Data Fig. 6a). For each contact in \mathbb{M}^4 , we computed the normalized mean membrane
551 contact area (see *Generating reference graphs*) and the standard deviation of membrane contact areas
552 across the 4 datasets. We observed no correlation between the normalized mean membrane contact area
553 and standard deviation (Extended Data Fig. 5b), indicating that the variability in membrane contact areas
554 does not depend strongly on membrane contact area (similar to immediate neighborhood sizes in
555 Extended Data Fig. 1a). Therefore, we estimated the variability in the membrane contact area by the
556 mean variability among \mathbb{M}^4 membrane contacts (Extended Data Fig. 5c).

557 To perturb each dataset, we applied multiplicative white noise to each membrane contact area, which
558 we derived from the distribution of membrane contact areas, as follows. A log-transformed (un-skewed)
559 and standardized (mean 0 and variance 1) membrane contact area y is computed from membrane contact
560 area x by

$$561 \quad y = \frac{\log(x) - \hat{\mu}}{\hat{\phi}}, \quad (1)$$

562 where $\hat{\mu}$ and $\hat{\phi}$ denote the geometric mean and standard deviations of the membrane contact areas (i.e.
563 the arithmetic mean taken in the log domain), across the 4 datasets, per cell pair. Rearranging terms gives

$$564 \quad x = e^{\hat{\mu}} e^{y\hat{\phi}}. \quad (2)$$

565 To perturb membrane contact areas ($x \rightarrow x'$), we add white noise ε in the log domain, i.e.,

$$566 \quad x' = e^{\hat{\mu}} e^{(y+\varepsilon)\hat{\phi}} = e^{\hat{\mu}} e^{y\hat{\phi}} e^{\varepsilon\hat{\phi}} = x e^{\varepsilon\hat{\phi}}. \quad (3)$$

567 Hence, we scale each membrane contact by $e^{\varepsilon\hat{\phi}}$, where $\hat{\phi}$ is determined by the membrane contact area
568 distribution of the dataset and the distribution ε is drawn randomly from a normal distribution with mean
569 0 and standard deviation σ .

570

571 The standard deviation, σ , of the ε distribution sets the amplitude of the perturbation. We determined
572 the appropriate noise amplitude by comparing the distributions of perturbed and empirical datasets. We
573 found that a noise amplitude of $\sigma = 0.23$ – roughly half of the mean standard deviation of membrane
574 contact areas (Extended Data Fig. 5c) – yields perturbed membrane contact area distributions (Extended
575 Data Fig. 5d-f) that are qualitatively similar to the empirical dataset (Extended Data Fig. 5a-c). Moreover,
576 the perturbed membrane contact areas scale linearly with membrane contact area (Extended Data Fig.
577 5g) and variability as a fraction of membrane contact area is uniform (Extended Data Fig. 5h).

578 Perturbed populations are denoted \widetilde{M}^4 , $\widetilde{L4}$ and \widetilde{Adult} . For \widetilde{M}^4 , we perturb contacts conserved across
579 the 4 datasets (L4 left, L4 right, adult left, adult right). For $\widetilde{L4}$ and \widetilde{Adult} , we perturb bilaterally conserved
580 contacts in the L4 and adult, respectively. Each population consists of 1000 perturbed datasets.

581 **Spatial modularity analysis**

582 To identify groups of neurites with high spatial affinity in the nerve ring, we performed a graph
583 modularity analysis of the membrane contact areas. Since spatial adjacencies between neurons consist of
584 both conserved and variable membrane contacts, we applied our clustering analysis to \widetilde{M}^4 , $\widetilde{L4}$ and \widetilde{Adult}
585 population models (unless otherwise stated). For clustering purposes, we reduced contralateral left/right
586 homologue vertices to a single vertex class. For example, vertices ASHL and ASHR were reduced to the
587 single vertex, ASH. The algorithm was then applied to each individual in the population.

588 The multilevel community detection algorithm yields a number of clusters of neuron classes whose
589 neurites exhibit high spatial affinity. Topological clustering methods such as modularity optimization^{21,40}
590 are well suited for characterizing the organization of a complex system from pairwise undirected linked
591 relationships^{40,41}, as is the case for characterizing spatial organization from membrane contacts between
592 neural processes. In particular, algorithms of this class are appropriate when the organization sought is
593 static^{40,41}. Other, random-walk based algorithms^{34,41} assume or impose a flow on the network and are
594 often ill-suited for characterizing spatial (i.e. static) organization, as they can introduce bias in the
595 clustering or miss static features in the organization of the system⁴¹. We applied the Louvain method, a
596 multilevel community detection algorithm using the igraph software package⁴². This topological

597 clustering algorithm is a bottom-up heuristic method based on modularity optimization²¹. Initially, every
598 vertex is placed in a separate community. Vertices are then iteratively moved between communities in a
599 way that maximizes the vertex's local contribution to the overall modularity score (the ratio of the
600 number of intra- to inter-community edges). When no vertex movement increases the modularity score,
601 communities are shrunk to a single vertex and the process is repeated.

602

603 **Cluster assignment and validation**

604 The graph clustering algorithm (see above) was applied to each individual in each population model. For
605 each population, we generated a cluster frequency matrix that counts the number of times each pair of
606 neurons is clustered together. We then sorted the rows and columns of the frequency matrix so as to
607 minimize the variance along the main diagonal of the matrix (Extended Data Fig. 5i). Sorting was
608 achieved using a hierarchical matrix clustering algorithm²¹. The resulting dendrogram assigns neurons
609 to a cluster. We obtained a set of 5 largely overlapping clusters for each of the \widetilde{M}^4 , $\widetilde{L4}$ and
610 \widetilde{Adult} population models (Extended Data Fig. 5i). Cell classes whose cluster assignment agreed across
611 the 3 population models were assigned to the consensus cluster. Seven neuron pairs (ADE, ALN, AVA,
612 RID, RIR, RMD and URX) were classified differently across the different population models, and were
613 designated 'unclassified' accordingly. To evaluate the robustness of the clusters to empirical variability
614 between the L4 and adult series, we compared clusters obtained from population models of \widetilde{M}^4 , $\widetilde{L4}$ and
615 \widetilde{Adult} (see *Population spatial model*, Fig. 1b,c, Extended Data Fig. 5i and 6b).

616 We performed four sets of validation experiments using our population models to confirm the
617 robustness of our neuron clusters. (1) As discussed above, we compared cluster assignments across \widetilde{M}^4 ,
618 $\widetilde{L4}$ and \widetilde{Adult} (Fig. 1b,c, Extended Data Fig. 5i). (2) We generated cluster assignments for \widetilde{M}^4
619 populations that were perturbed with different noise amplitudes ($\sigma = 0, 0.12, 0.23, 0.45, 0.9$; see
620 *Population spatial models*, Extended Data Fig. 5j). (3) We generated a new \widetilde{M}^4 ($\sigma = 0.23$) population
621 from membrane contacts in a more restricted volume recently used in Moyle *et al.* (2020)³⁴ which
622 consists of the anterior ~60% of our nerve ring neuropil volume (Extended Data Fig. 5k). (4) We
623 generated a new \widetilde{M}^4 ($\sigma = 0.23$) population that also includes the smallest 35% membrane contact areas
624 (Extended Data Fig. 5l; recall smallest contacts were removed from our analysis, Extended Data Fig. 2).
625 All of our validation experiments resulted in largely similar cluster assignments (Supplementary
626 Information 4). A handful of neuron classes were assigned to different clusters in different population

627 models, but the gross structure of the 5 main neuron clusters, as defined by the cluster assignments of a
628 large majority of the neurons, was consistent across the populations. Thus, our cluster assignments are
629 robust across model population datasets, the L4 and adult, a wide range of noise amplitudes, well above
630 the observed inter-animal variability, and different spatial domains.

631 Next, we validated our population model by repeating the clustering analysis on the unperturbed \mathbb{M}^4
632 reference dataset as well as on the unperturbed adult and L4 bilateral datasets (Extended Data Fig. 5k, σ
633 = 0 and Extended Data Fig. 6b). We find small differences between these datasets, but those are not
634 robust to small perturbation in our population models ($\sigma = 0.12$, i.e. below our estimated level of expected
635 biological variability in core contacts). We also validate our core assumption that the conserved structure
636 of the nerve ring requires analysis of the reproducible membrane contacts by comparing clusters from
637 unperturbed \mathbb{M}^1 - \mathbb{M}^4 datasets. We find that whereas the reproducible membrane contact datasets
638 consistently give rise to a small number of clusters, with largely similar composition, \mathbb{M}^1 - \mathbb{M}^3 membrane
639 contacts fail to reproduce these results (Extended Data Fig. 6a), suggesting that variable membrane
640 contacts may be masking the core, conserved spatial organization of the nerve ring neuropil.

641 **Mesoscale analysis of synaptic connectivity**

642 Given that the organization of the neuropil is modular, with most neurons spatially clustering within local
643 neighborhoods and others spatially interconnecting different neighborhoods, we wanted to determine if
644 synapses form local subcircuits, or to what extent synaptic circuits also span different neighborhoods of
645 the nerve ring. To assess the spatial organization of synaptic circuits, we considered the distribution of
646 conserved \mathbb{C}^4 synaptic contacts (Fig. 3a). Using the \mathbb{M}^4 dataset, we calculated the mean ($\bar{N} = 17$ cells)
647 and standard deviation ($\Delta N = 8$ cells) of immediate neighborhood sizes (*Anatomical and neuron-class*
648 *nomenclature*). We order the cells as in Fig. 1b so as to maximize the amount of physical (\mathbb{M}^4) contact
649 along the diagonal of the matrix.

650 We define 5 zones based on the size distribution of immediate neighborhoods. Each zone refers to
651 regions between two diagonals above and below the main diagonal of the \mathbb{M}^4 matrix. The inner zone
652 (labeled 0) consists of neighborhoods of size \bar{N} for each cell around the main diagonal. The next zone
653 (1) extends from the edges of the inner zone to diagonals ΔN further away from the main diagonal, and
654 zones 2-3 similarly extend by ΔN . The outermost zone (4) extends from the previous zone (3) to
655 encompass cells in the remainder of the matrix. (Formally, zones are defined by their inner and outer
656 diagonals, with the inner diagonal defined by $\bar{N}/2 + (k - 1)\Delta N$ from the main diagonal for zones 1 to

657 4 and the outer diagonals defined by $\bar{N}/2 + k\Delta N$ from the main diagonal for zones 0 to 3.) We counted
658 the number of \mathbb{C}^4 contacts in each Zone (Fig. 3b). Finally, we counted \mathbb{C}^4 contacts between pre- and
659 postsynaptic neurons that have been assigned to the same cluster (Fig. 3c). For this purpose, synapses
660 between two unclassified neurons are not considered intra-cluster.

661 **Contact localization analysis**

662 While some membrane contacts appear to be reproducible (our \mathbb{M}^4 reference dataset), contacts are
663 aggregate measures (along the entire process). To assess the reproducibility in the location of individual
664 instances of membrane adjacencies along a neurite, we assigned each EM in each process a discrete
665 coordinate, \hat{z} , from the anterior ($\hat{z} = 0$) to the posterior ($\hat{z} = 1$) of the process. This allows us to compare
666 relative locations of a contact across the four datasets (L4 left/right, adult left/right). Different
667 discretizations of \hat{z} (0.7 μm , 1.4 μm , 3.6 μm) define different resolution for the reproducibility of contacts
668 along the process. For each \mathbb{M}^4 contact, we define the spatial reproducibility count as the number of
669 datasets where the contact was observed at a given position, \hat{z} . We further define the maximum spatial
670 reproducibility count, $\max(\delta)_{\hat{z}}$, as the highest reproducibility count across all locations, \hat{z} , per cell pair
671 (i.e. given an \mathbb{M}^4 contact exists between two immediate neighbors, the highest reproducibility count of
672 instances of membrane adjacencies between the two cells). To assess synaptic localization, we similarly
673 measured the spatial reproducibility counts (and their maxima) for all \mathbb{C}^4 contacts. See also
674 Supplementary Results.

675 **Synapse compartmentalization and subcellular structures**

676 Identification of synaptic compartmentalization and subcellular structures was performed by visually
677 inspecting the volumetric reconstruction of the processes of 173 neurons (the restricted dataset) in both
678 the adult and L4 datasets (346 cells in total). To visualize synapses, we imported synapse locations³
679 (<http://wormwiring.org>) into the reconstructed TrakEM2 datasets. To facilitate visual identification, we
680 colored synapses based on whether the cell is presynaptic or postsynaptic and whether the synapses
681 occurs between cells of the same cluster or not. For each cell, we required synapse compartmentalization
682 and/or subcellular structures to be bilaterally conserved in both the L4 and the adult (Supplementary
683 Information 4). The one exception are the RMF cells, where there is clear branching in the L4 (both left
684 and right) that is not observed in the adult. However, because we are limited to 2 samples, we cannot
685 determine if these are developmental, individual differences or reconstruction error⁵.

686 We identified two types of synaptic compartmentalization: compartmentalization of synaptic inputs
687 and outputs and compartmentalization of synapses with different clusters. To identify
688 compartmentalization of inputs (outputs), we required neural segments to have ≥ 3 synaptic inputs
689 (outputs) that are spatially distinct from segments with synaptic outputs (inputs) or neural segments with
690 mixed synaptic inputs and outputs (Extended Data Fig. 9).

691 We identified flattened protrusions by looking for points along the neural processes with increased
692 surface area. We further identified flattened protrusions with mixed synaptic inputs and outputs, which
693 we interpret to be local points with diverse synaptic polarity. In some instances (Supplementary
694 Information 4), these flattened protrusions appear to extend to branches or spine-like structures
695 (Extended Data Fig. 8). Note that synaptic compartmentalization and flattened protrusions are not
696 mutually exclusive. We observed 9 cells that exhibit co-localized synaptic compartmentalization and
697 flattened protrusions. In these instances, the flattened protrusions appear to be used to compartmentalize
698 reproducible (C^4) synapses (Extended Data Fig. 8).

699 **Mapping neighborhood changes of neurites**

700 We observe that some neuron processes extend into multiple neighborhoods (Fig. 3 and Extended Data
701 Fig. 8). We manually mapped neighborhood changes along process trajectories for selected L4 left
702 neurons (as representatives of their cell class). Starting at the proximal end of the process (closest to cell
703 body), we followed the process trajectory through the EM stack. At each EM, we visually noted the
704 cluster assignments of the neighboring neurites and assigned the neighborhood of that segment of the
705 neurite accordingly. If the neighboring neurites comprised two or more clusters then we labeled the local
706 neighborhood as ‘mixed’. The sequence of local neighborhood segments along the neurite was then
707 scaled by the total length of the neurite so that all positions along the neurite range between 0 and 1. In
708 the case of AVA and RIM, which have protrusions that branch out from the main process trajectory, we
709 scaled the protrusion length by the same factor as the main neurite trajectory.

710

711 **Brain map construction**

712 We posited a 3-layer architecture as the minimal number of layers needed to capture the organizing
713 principles of the connectome. Classifications of neurons as sensory neurons, interneurons or
714 motoneurons followed WormAtlas³⁶. All sensory neurons were assigned to the first layer. SDQ, BDU
715 and ALN have been postulated to have sensory functions^{44,45} but were classified as interneurons as they
716 are not ciliated and physiological evidence for sensory function is lacking. Reclassifying them as sensory
717 neurons would not alter the high-level connectivity of the brain map. With the exception of AIY and

718 AIA, all neurons that make at least one \mathbb{C}^4 inter-cluster contact were placed in layer 3, with the remainder
719 of neurons assigned to layer 2. Placing AIY and AIA in layer 2 is consistent with functional and ablation
720 studies suggesting that these cells are first-layer amphid interneurons^{22,46}. Furthermore, AIY and AIA
721 each only make 1 inter-cluster \mathbb{C}^4 synaptic contact (Extended Data Fig. 10). AIY synapses onto the multi-
722 compartment cell RIA which traverses multiple neighborhoods (Extended Data Fig. 8a). AIA synapses
723 onto RIF whose neurite is at the interface of taxis and avoidance cells in the nerve ring posterior lobe.
724 We confirmed that our map is robust to small changes in which neurons with relatively few inter-cluster
725 synaptic contacts between layers are shifted to layer 2. However, the configuration adopted here
726 optimizes the feed-forward directionality of the synaptic circuit (from the sensory layer to layer 3).

727 Our information processing modules roughly correspond to the 5 spatially identified clusters. The
728 sublateral and lateral clusters were merged into a single module. With one exception (CEPD), cell classes
729 in the same cluster are placed within the same module. Because CEPD neurons follow the same looping
730 neurite trajectories as other papillary sensory neurons, CEPD cells, which are assigned to the sublateral
731 cluster, are more sensibly placed in the anterior module. Unclassified cells are difficult to cluster because
732 they exhibit high spatial affinity with cells from different clusters. To place the 7 unclassified cell classes
733 on the brain map, we relied on the relative placement of their process trajectories among the clusters. We
734 identified representative cells from each cluster to serve as fiducial points for process placement
735 (Anterior: RIH, Lateral: AVK and RIV, Sublateral: SIAD, Avoidance: AVB, Taxis: ASJ). Each
736 unclassified cell was then added to the module of the representative cells whose neurite most closely
737 aligned with the neurite of the unclassified cell.

738 **Statistical connectivity models**

739 We asked whether stochastic processes could account for the reproducibility and variability of contacts
740 across the 4 datasets. For parsimony, we treat all potential contacts, or graph edges, as identical and allow
741 for all-to-all connectivity. The empirical contact distributions (\mathbb{M}^δ for membrane contacts, \mathbb{C}^δ for synapses
742 and \mathbb{G}^δ for gap junctions, Fig. 2a) are all bimodal. Therefore, within the above assumptions, a single
743 stochastic process (for making, or equivalently suppressing) contacts cannot account for these
744 distributions.

745 We therefore constructed a minimal 3-parameter model combining two stochastic processes –
746 precision and specificity. Precise targeting of contacts and active avoidance of others both require us to

747 distinguish between the set of candidate target contacts and the remainder (non-targets). Accordingly,
 748 we define a fraction of target contacts (f), the probability to form a target contact (precision, p) and the
 749 probability to avoid an off-target contact (specificity, s).

750 This model suffices to define the distribution 0, 1, 2, 3 and 4 contacts. For $\mathbb{A} \in \{\mathbb{M}, \mathbb{C}, \mathbb{G}\}$, the
 751 probability of \mathbb{A}^δ is given by:

$$752 \quad \Pr[\mathbb{A}^\delta] = \sum_{\delta=0}^4 \binom{4}{\delta} (fp^\delta(1-p)^{4-\delta} + (1-f)(1-s)^\delta s^{4-\delta}), \quad (4)$$

753 where the parameters f , p and s may take on different values for different instances of $\mathbb{A} \in \{\mathbb{M}, \mathbb{C}, \mathbb{G}\}$. In
 754 the absence of empirical data for estimating the physically accessible subset of contacts, we restrict our
 755 consideration to $\delta \in \{1, 2, 3, 4\}$, or in general, for K datasets, using Bayes' theorem:

$$756 \quad \Pr[\mathbb{A}^\delta | \delta > 0] = \frac{\Pr[\mathbb{A}^\delta, \delta > 0]}{\Pr[\delta > 0]} = \frac{\Pr[\mathbb{A}^\delta]}{\sum_{\delta=1}^k \Pr[\mathbb{A}^\delta]}. \quad (5)$$

757 **Model fits**

758 We used a greedy search of the entire parameter space (with 1% resolution) to find the 3 parameter values
 759 (f , p and s) that minimize the L^1 -norm between the predicted and empirical distributions. Due to the
 760 symmetry of the equations, the model has two solutions that are equivalent up to relabeling of the nodes
 761 (and given by $f \rightarrow 1-f$, $p \rightarrow 1-s$, $s \rightarrow 1-p$) such that target and non-target populations are swapped both
 762 in size and in the probability of contacts. We choose the solution in which the target fraction, f ,
 763 corresponds to the solution with $p > 1-s$, such that precisely targeted contacts are synonymous with
 764 higher reproducibility across datasets.

765 A further equivalent reparametrization exists that replaces a specificity mechanism (acting only on
 766 non-target edges) with a uniform basal connectivity (that applies to both target and non-target edges).
 767 This variant of the model provides an alternative interpretation, in which the three parameters are the
 768 target fraction, \tilde{f} , precision, \tilde{p} , and basal activity level, \tilde{b} . The solution can be obtained with the
 769 reparametrization: $f = \tilde{f}$, $s = 1 - \tilde{b}$, $p = \tilde{p} + \tilde{b} - \tilde{p}\tilde{b}$. Imposing the condition $0 \leq \tilde{p} \leq 1$ eliminates
 770 one of the two solutions for all our model fits.

771

772 **Empirical data for fits and bias control**

773 For membrane, synaptic and gap junction contacts, we found no evidence of higher reproducibility of
 774 edges between the left sides of the L4 and adult datasets, or between the right sides of the L4 and adult,
 775 as compared to L4 left and adult right, or vice versa (Supplementary Table 3). However, for membrane
 776 contacts, development leads to an overall increase of edges between the L4 and adult (Supplementary
 777 Table 1). We considered all neuron pairs in our complete dataset (3203 edges with membrane contact
 778 areas ≥ 35 percentile, Extended Data Fig. 4b) as well as the restricted dataset (see above, 2955 edges).
 779 Models of the complete and restricted datasets yielded quantitatively similar results (Extended Data Fig.
 780 3e).

781 Our restricted dataset consists of 173 neurons. In the absence of spatial constraints, all-to-all
 782 connectivity would, in principle, allow for up to $173 \times 172 / 2 = 14,878$ edges. Conversely, using the model
 783 fit, the sum $\sum_{\delta=0}^K \mathbb{M}^{\delta}$ could provide an estimate for the size of the pool of physically accessible membrane
 784 contacts in the nerve ring. The above estimate (≈ 3500 edges for the restricted set of contacts) is about
 785 23% of the all-to-all number. This model estimate points to the strong role that spatial constraints play in
 786 the actual circuit.

787 The space of possible synaptic and gap junction contact is restricted by the existence of a physical
 788 membrane contact. Unless otherwise noted, all fits were performed on \mathbb{C}^{δ} and \mathbb{G}^{δ} that were restricted to
 789 edges from the set of \mathbb{M}^4 membrane contacts. To control for possible bias due to the subselection of \mathbb{M}^4
 790 contacts, validation plots were generated by considering \mathbb{M}^j contacts and scaling the counts $\mathbb{C}^{\delta} |_{\mathbb{M}^j} \rightarrow$
 791 $\frac{\mathbb{C} |_{\mathbb{M}^j}}{\mathbb{C} |_{\mathbb{M}^4}} \mathbb{C}^{\delta} |_{\mathbb{M}^4}$ for $j = 1, \dots, \delta$, where $\mathbb{C} = \sum_{\delta=0}^j \mathbb{C}^{\delta}$ and ' $|_{\mathbb{M}^j}$ ' denotes synaptic contacts occurring on the
 792 domain of membrane contact \mathbb{M}^j (scaling was performed in the same way for gap junction contacts, \mathbb{G} ;
 793 Extended Data Fig. 4a-c). For chemical synapses, we find good agreement with \mathbb{C}^3 when scaled by \mathbb{M}^3
 794 and \mathbb{C}^2 when scaled by either \mathbb{M}^3 or \mathbb{M}^2 . Rescaling systematically underestimated \mathbb{C}^1 across all datasets,
 795 possibly due to a subset of small synapses not accounted for by the model. Indeed, consistent with Hall
 796 and Russell⁶, we find that both \mathbb{C}^1 synapses and \mathbb{G}^1 gap junctions are significantly smaller (Extended Data
 797 Fig. 4f,g) and occur at smaller membrane contacts (Extended Data Fig. 2i, see also *Validation against*
 798 *test datasets, below*).

799 To estimate the fraction of target edges (for $\delta = 1 \dots 4$), we used Eq. (5). For example, the probability
 800 of observing $\delta = 4$ target membrane contacts is given by $f p^4 = 0.44 \times 0.95^4 = 36\%$, whereas the probability
 801 of finding 4 variable membrane contacts, $(1-f)(1-s)^4$, is negligible. Thus, the estimated fraction of \mathbb{M}^4
 802 membrane contacts in the core circuit is estimated as $f p^4 / [f p^4 + (1-f)(1-s)^4] > 99\%$, whereas, the

803 estimated fraction of \mathbb{M}^3 contacts in the core circuit is only $4fp^3(1-p) / [4fp^3(1-p) + 4(1-f)s(1-s)^3]$
804 = 68%. Finally, we estimate that in two animals (4 datasets) one would expect $p^4 + 4p^3(1-p)$ of core
805 edges to occur in at least 3 datasets (corresponding to ~99% of core membrane contacts and ~97% of
806 core synaptic edges). Additionally, we separately fit the model to intra-cluster and inter-cluster edges.
807 For each set of membrane contacts, \mathbb{M}^δ , we separated the contacts that occur between neurons with the
808 same cluster identity (intra-cluster) and contacts between neurons with different cluster identities (inter-
809 cluster). We then separately fit the model to the sets of intra- and inter-cluster edges, corresponding to
810 membrane contacts, synapses and gap junctions (Extended Data Fig. 3c-d).

811 ***Simulation and generation of surrogate data***

812 To construct each surrogate dataset, k , we set the size of the dataset, n (e.g. 2955 for membrane contacts)
813 and created an ordered list $\mathcal{L}(k)$ of edges. We generated a binary target list (the first round (fn) elements
814 in the list, $\mathcal{L}_T(k)$) and a binary non-target list $\mathcal{L}_{NT}(k)$; among target edges, contact, i.e. 1, occurs with
815 probability p and among off-target edges, contact occurs with probability $1-s$. We then aggregate the
816 counts across K surrogate datasets, $\delta_i = \sum_{k=1}^K \mathcal{L}_i(k)$, where δ_i corresponds to the number of datasets in
817 which edge i forms a contact. The list of δ_i then forms a surrogate dataset for the reproducibility of
818 contacts, e.g. \mathbb{M} .

819 ***Validation against test datasets***

820 As additional connectomes are generated and technologies change, we expect slight differences in
821 scoring of different datasets generated from different EM sets^{3,6,7,17,18,39}. These could arise from slightly
822 different demarcation of the volume being scored, different EM sectioning (or sections scored) and
823 different scoring criteria. In the absence of functional (molecular or physiological) data, it is difficult to
824 avoid some false positives (scored synapses that are not fully developed and functional) and false
825 negatives (missed synapses). Often, smaller synapses fare harder to score accurately. Furthermore, most
826 *C. elegans* synapses are polyadic and present particular challenges, especially when one of the targets
827 occurs with a considerably smaller membrane contact area. Methods and validation of synaptic scoring
828 for the dataset used here have been described by Cook *et al.*³. Here, we address complementary aspects,
829 relating to reproducibility of scores and implications for our model of core and variable circuits
830 (Extended Data Fig. 3e-i).

831 Cook *et al.*³ (the dataset used here) scored a greater number of small synapses than White *et al.*⁵
832 (Extended Data Fig. 4f). Furthermore, while this paper was under submission, additional connectomes

833 have been reported for eight hermaphrodite *C. elegans* nerve rings, including two adults²⁰. We therefore
834 validated our main results on synaptic reproducibility against the connectomes of White *et al.*⁵ and the
835 two adults in Witvliet *et al.*²⁰ (hereafter, ‘test datasets’, denoted with the subscript test). As the volumetric
836 reconstruction and hence membrane contact analysis is only available for our study, we used the \mathbb{M}^4
837 edges identified here as a common basis for comparison and validation.

838 Size dependence of synaptic reproducibility has previously been noted^{3,6,7,39}. Consistently with these
839 earlier results, Extended Data Fig. 4f shows that \mathbb{C}^4 synapses, and less so \mathbb{C}^3 synapses, have a considerably
840 higher fraction of edges associated with higher EM section counts: 87% of \mathbb{C}^4 and 37% of \mathbb{C}^3 edges are
841 observed in ≥ 5 EM sections, as compared to 13% and 21% in \mathbb{C}^1 and \mathbb{C}^2 , respectively. That said, a
842 comparison with the White *et al.* test dataset⁵ shows that the additionally scored synaptic edges are evenly
843 distributed across \mathbb{C}^1 - \mathbb{C}^4 (Extended Data Fig. 4f). To check whether different scoring criteria leading to
844 different counts of small synapses affect our conclusions, we re-fit our model to a more restricted synaptic
845 dataset in which all 1-EM section synapses were excluded. While this substantially suppresses \mathbb{C}^1 counts
846 (hence affecting the relative core and variable fractions), its effect on our model precision and specificity
847 is minor (Extended Data Fig. 3f). The scoring of polyadic synapses is also potentially challenging, if
848 synapses are formed with only a subset of co-localized postsynaptic neighbors. To check whether
849 excessive scoring of polyadic synapses might affect our results, we constructed a synaptic dataset in
850 which for every polyadic pre-synaptic site, we excluded any postsynaptic partner that is in \mathbb{C}^1 . Re-fitting
851 our model to this restricted synaptic dataset, we again find similar precision and specificity.

852 Next, we reasoned that to be reliable, our statistical model should be robust across datasets. To
853 validate this, we re-fit our model to the two test datasets (Extended Data Fig. 3h-i). Both test datasets
854 show a qualitatively similar bimodal distribution of synaptic reproducibility (\mathbb{C}^1 - \mathbb{C}^4) that is well fitted to
855 our 3-parameter model. Model fit parameters varied only slightly from our results (Fig. 2): a synaptic
856 edge precision of 92-96% and a specificity of 68-74%. For each synaptic edge scored by Cook *et al.*³, we
857 then counted the number of edges scored in the test dataset. All but 1 of our \mathbb{C}^4 edges and 93% of our \mathbb{C}^3
858 edges were scored at least once by Witvliet *et al.*²⁰ (Extended Data Fig. 4h), suggesting that some small
859 synapses are in fact highly reproducible. While slight differences in our model fits preclude automatic
860 merging of the datasets (or models), their similarity implies that it should be possible to quantitatively
861 validate the two extremes, namely non-reproducible and entirely reproducible edge counts, as those are
862 almost certain to come from the variable and core circuits, respectively.

863 To validate the scoring of postulated variable synapses, we use our model parameters and Eq. (4) to
864 estimate what number of synaptic edges in our dataset would be statistically expected to be absent from
865 two independent animals,

$$866 \quad n \frac{\Pr[\mathbb{C}^0]}{\Pr[\mathbb{C}^\delta | \delta > 0]} = n \left(\frac{1}{1 - (1 - f)s^4 - f(1 - p)^4} - 1 \right).$$

867
868 This expression yields an estimated number of ~ 330 synaptic edges. Empirically, we find that 374 of our
869 synapses were not scored in the Witvliet *et al.* test dataset²⁰ (Extended Data Fig. 3i; difference not
870 statistically significant under binomial counting statistics). This result adds confidence to the scoring of
871 variable and in particular, small synapses in our dataset. To examine the consistency of postulated
872 conserved synapses, we estimated the number of \mathbb{C}^4 synaptic edges scored by Cook *et al.*³ that would also
873 be expected to be found in two independent animals (i.e. in a new set of \mathbb{C}^4). Of our 450 \mathbb{C}^4 synaptic edges,
874 we expect a test dataset to include $\frac{n_{\text{test}}}{n} 450 \sim 380$ as \mathbb{C}^4 (also equivalent to $n_{\text{test}} f p^4 / \Pr[\mathbb{C}^\delta > 0]$).
875 Empirically, Witvliet *et al.*²⁰ score 389 \mathbb{C}^4 of our \mathbb{C}^4 contacts, consistent with our model predictions.

876

877 **Statistics and reproducibility**

878 Membrane contact datasets are derived from the EM reconstructions of the nerve rings from 2 animals at
879 different developmental stages. Each animal in our restricted dataset consists of 80 pairs of bilateral
880 homologous cells. Extended Data Fig. 1 and 2 established that bilateral homologous cells are sufficiently
881 similar. Accordingly, for the purposes of generating reference graphs and for our core-variable and
882 population models, we assume the two sides of each animal may be treated as independent, yielding 4
883 independent datasets (L4 left, L4 right, adult left and adult right) each consisting of 93 cells classes. As
884 further measures of reproducibility, we validated our core-variable synaptic and gap junction contact
885 models against data scored by different experts on the same EM series⁵ and on different EM datasets²⁰
886 (in both cases, limited to our \mathbb{M}^4 contacts). Our models yielded qualitatively similar results for the
887 different scorings and datasets (Extended Data Fig. 3h,i). Spatial population model data were drawn from
888 distributions that matched the empirical distributions of \mathbb{M}^4 membrane contact areas across the 4 datasets.

889

890 **Acknowledgements**

891 We thank Jonathan Hodgkin and John White for their help in donating archival TEM material from
892 LMB/MRC to the Hall lab for curation. Thomas Ilett, Felix Salfelder and Samuel Braunstein provided

893 useful discussion. We thank Mei Zhen for making their synaptic and gap junction data available
894 (<https://nemanode.org/>). This work was supported by NIH grant NIMH F32MH115438 (SJC), NIMH
895 R01MH11289 (SWE), the G. Harold and Leila Y. Mathers Charitable Foundation (SWE), NIH OD
896 010943 (DHH) and EPSRC EP/J004057/1 (NC). CB was supported by the Leeds International Research
897 Scholarship.

898

899 **Author Contributions**

900 CB, SJC and SWE conceived the volumetric reconstruction. CB and SJC segmented the electron
901 micrographs. DHH curated the data. CB built the software for quantifying membrane contact areas. CB
902 and NC analyzed and interpreted the data and wrote the manuscript. SJC, DHH and SWE provided
903 critical revisions.

904

905 **Competing Interests**

906 The authors declare no competing interests.

907

908 **Data Availability**

909 The volumetric datasets generated during the current study, associated connectivity databases and
910 associated analysis are available at <https://doi.org/10.5281/zenodo.4383277> and
911 <http://wormwiring.org/>. The raw data for volumetric reconstructions for Figs. 1 and 3, Extended Data
912 Fig. 8 and all Supplementary Videos is available at <https://doi.org/10.5281/zenodo.4383277>. Extracted
913 adjacency data is available in Supplementary Information 1. The reference datasets are available in
914 Supplementary Information 3. The Cytoscape files use to generate the brain map (Fig. 4 and Extended
915 Data Fig. 10) and network motifs (Extended Data Fig. 10) are available at
916 <https://doi.org/10.5281/zenodo.4383277> . See associated Source Data for the data used to generate
917 plots. The collection of *C. elegans* nervous system electron micrographs are also available at
918 <https://www.wormatlas.org/> and <https://wormimage.org>.

919

920 **Code Availability**

921 Software packages parsetrakem2 (extracting adjacency data) and elegansbrainmap (analysis and
922 visualization software) are available at <https://github.com/cabrittin/parsetrakem2> and
923 <https://github.com/cabrittin/elegansbrainmap> respectively.

924

925

926 **References**

- 927 [35] Ware, R. W., Clark, D., Crossland, K. & Russell, R. L. The nerve ring of the nematode
928 *Caenorhabditis elegans*: Sensory input and motor output. *J. Comp. Neurol.* **162**, 71–110 (1975).
- 929 [36] Altun, Z. & Hall, D. WormAtlas (2012). URL <http://www.wormatlas.org>.
- 930 [37] Peachey, L. D. Thin sections. I. A study of section thickness and physical distortion produced
931 during microtomy. *J. Biophys. Biochem. Cyt.* **4**, 233–242 (1958).
- 932 [38] Cardona, A. et al. TrakEM2 Software for Neural Circuit Reconstruction. *PLoS ONE* **7**, e38011
933 (2012).
- 934 [39] Xu, M. et al. Computer assisted assembly of connectomes from electron micrographs: Application
935 to *Caenorhabditis elegans*. *PLoS ONE* **8**, e54050 (2013).
- 936 [40] Newman, M. E. & Girvan, M. Finding and evaluating community structure in networks. *Phys.*
937 *Rev. E* **69**, 026113 (2004).
- 938 [41] Rosvall, M. & Bergstrom, C. T. Maps of random walks on complex networks reveal community
939 structure. *Proc. Natl. Acad. Sci. USA* **105**, 1118–1123 (2008).
- 940 [42] Csardi, G. C. & Nepusz, T. The igraph software package for complex network research.
941 *InterJournal Complex Sys.* 1695 (2006).
- 942 [43] Virtanen, P. et al. SciPy 1.0 Contributors. SciPy 1.0: fundamental algorithms for scientific
943 computing in Python. *Nat Methods.* **17**, 261–272 (2020).
- 944 [44] Chang, A. J., Chronis, N., Karow, D. S., Marletta, M. A. & Bargmann, C. I. A distributed
945 chemosensory circuit for oxygen preference in *C. elegans*. *PLoS Biol.* **4**, e274 (2006).
- 946 [45] Zimmer, M. et al. Neurons detect increases and decreases in oxygen levels using distinct
947 guanylate cyclases. *Neuron* **61**, 865–879 (2009).
- 948 [46] Tomioka, M. et al. The Insulin/PI 3-Kinase pathway regulates salt chemotaxis learning in
949 *Caenorhabditis elegans*. *Neuron* **51**, 613–625 (2006).
- 950 [47] Hendricks, M., Ha, H., Maffey, N. & Zhang, Y. Compartmentalized calcium dynamics in a *C.*
951 *elegans* interneuron encode head movement. *Nature* **487**, 99–103 (2012).
- 952 [48] Perkins, L. A., Hedgecock, E. M., Thomson, J. N. & Culotti, J. G. Mutant sensory cilia in the
953 nematode *Caenorhabditis elegans*. *Dev. Biol.* **117**, 456–487 (1986).
- 954 [49] Sawin, E. R., Ranganathan, R. & Horvitz, H. R. *C. elegans* locomotory rate is modulated by the
955 environment through a dopaminergic pathway and by experience through a serotonergic pathway.
956 *Neuron* **26**, 619–631 (2000).

- 957 [50] Kang, L., Gao, J., Schafer, W. R., Xie, Z. & Xu, X. Z. *C. elegans* TRP family protein TRP-4 is
958 apore-forming subunit of a native mechanotransduction channel. *Neuron* **67**, 381–391 (2010).
- 959 [51] Chalfie, M. & Sulston, J. Developmental genetics of the mechanosensory neurons of
960 *Caenorhabditis elegans*. *Dev. Biol.* **82**, 358–370 (1981).
- 961 [52] Suzuki, H. et al. In vivo imaging of *C. elegans* mechanosensory neurons demonstrates a specific
962 role for the MEC-4 channel in the process of gentle touch sensation. *Neuron* **39**, 1005–1017
963 (2003).
- 964 [53] Chalfie, M. et al. The neural circuit for touch sensitivity in *Caenorhabditis elegans*. *J. Neurosci.*
965 **5**, 956–964 (1985).
- 966 [54] Li, C. et al. The FMRamide-related neuropeptide FLP-20 is required in the mechanosensory
967 neurons during memory for massed training in *C. elegans*. *Learn. Mem.* **20**, 103–108 (2013).
- 968 [55] Hukema, R. K., Rademakers, S., Dekkers, M. P. J., Burghoorn, J. & Jansen, G. Antagonistic
969 sensory cues generate gustatory plasticity in *Caenorhabditis elegans*. *EMBO J.* **25**, 312–322
970 (2006).
- 971
- 972

973 **Extended Data Fig. 1. Neuron neighborhoods are bilaterally conserved in size, composition and**
974 **membrane contact positions. a,** Variability in immediate neighborhood size (adjacency degree) does
975 not vary with immediate neighborhood size. Immediate neighborhood sizes for each neuron in each
976 dataset (adult left, adult right, L4 left, L4 right, $n = 80$ bilateral cell classes common to L4 and adult)
977 plotted against the immediate neighborhood size of the corresponding neuron in the adult left. The inset
978 shows the immediate neighborhood size difference between homologous left/right neurons (vertical
979 spread) as a function of neighborhood size for the L4 (red) and adult (blue). **b,** Distribution of
980 immediate neighborhood size differences between homologous contralateral neurons in the same animal
981 (adult L/R and L4 L/R) are statistically indistinguishable from 0 (two-sided Wilcoxon signed-rank test
982 derived p-values: 0.07 and 0.29, respectively, $n = 80$ cell classes). Immediate neighborhood size
983 differences between homologous adult and L4 neurons on the same side of the body are statistically
984 distinguishable from 0 (two-sided Wilcoxon signed-rank test, p-value 9.2×10^{-11} , $n = 160$ cells), but the
985 difference is small (mean degree difference 3.6). **c,** Similarity between immediate neighborhood
986 compositions as quantified by the Jaccard index (Supplementary Results), shows higher compositional
987 similarity between homologous contralateral neighborhoods ($n = 80$ cell classes) than between proximal
988 ipsilateral neighborhoods (Supplementary Results, $n = 160$ cells). **d-f,** Membrane contact placement
989 along processes is highly reproducible bilaterally and across the adult and L4 datasets. For each process,
990 we mapped each M^4 contact to a position along the anterior-posterior (AP) axis, \hat{z} , (see Methods and
991 Supplementary Results). For each M^4 contact, we then counted the number of datasets where the
992 contact was observed at a given \hat{z} (reproducibility count). **d,** Demonstration of reproducibility count for
993 a single cell class (RIA): RIA has the longest process in the nerve ring and among the highest average
994 reproducibility counts. A raster plot of reproducibility counts as a function \hat{z} of all M^4 contacts made
995 with RIA. Neighboring processes: rows in alphabetical order. Color: reproducibility count. We define
996 the maximum spatial reproducibility count, $\max(\delta)_z$ as the highest reproducibility count across all
997 locations, \hat{z} , per cell pair (i.e. for every row in the raster). For rasters of all other cell classes, see
998 Supplementary Information 2. **e,** Fraction of M^4 membrane contact sites co-localized in δ datasets
999 (distribution over $n = 80$ cell classes). **f,** For each cell class, the fraction of membrane contacts achieved
1000 with a maximum spatial reproducibility count, $\max(\delta)_z$ (distribution over $n = 80$ cell classes). **g-h,**
1001 Comparatively, C^4 synaptic contact placement is less reproducible than physical adjacency. For each
1002 process, we mapped each C^4 contact along the AP axis, \hat{z} . **g,** Demonstration of synaptic spatial
1003 reproducibility count for RIA. **h,** For each cell class, the fraction of C^4 synaptic contacts achieved with a
1004 maximum spatial reproducibility count, $\max(\delta)_z$ (distribution over $n = 80$ cell classes). Box plots: center
1005 line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers.
1006

1007 **Extended Data Fig. 2. Contact sizes and reproducibility. a-f,** Small membrane contact areas are less
1008 likely to be bilaterally conserved. Membrane contacts were divided into three groups ('low', 'mid' and
1009 'high') based on their membrane contact areas (35% low, 31% mid, 34% high), see Supplementary
1010 Results). **a,** Similarity of homologous (L4 bilateral; adult bilateral; L4 and adult – same side) immediate
1011 neighborhood compositions for low, middle and high membrane contact groups, as measured by the
1012 Jaccard index (Supplementary Results, $n = 80$ cell classes). Box plot: center line, median; box limits,
1013 upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers. **b,c,** Survival (i.e.
1014 complementary cumulative) distribution of membrane contacts in **b,** the adult ($n = 5,179$) and **c,** the L4
1015 ($n = 4,744$). The pie charts show the fraction of total membrane area contact between all processes
1016 accounted for by each group. **d,** Empirical frequency distribution of synaptic ($n = 2,433$) and gap
1017 junctional ($n=573$) contacts broken down by the reproducibility of membrane contacts. The majority of
1018 synaptic contacts (77% and 85% of synaptic and gap junction contacts, respectively) occur at M^4

1019 contacts. **e,f**, Cumulative distribution of **e**, \mathbb{C}^δ synaptic contacts and **f**, \mathbb{G}^δ gap junction contacts for $\delta =$
1020 1,2,3,4 as a function of membrane contact area (in percentiles). To control for differences in neurite
1021 placement, we restrict \mathbb{C}^δ and \mathbb{G}^δ to contacts that occur on \mathbb{M}^4 membrane contacts. The smallest 35% of
1022 membrane contacts (dashed line) comprises $\sim 3\%$ of \mathbb{C}^4 synaptic contacts and $\sim 9\%$ of \mathbb{G}^4 gap junction
1023 contacts (on \mathbb{M}^4) with growing fractions for smaller δ (up to $\sim 33\%$ and $\sim 27\%$ of the more variable \mathbb{C}^1
1024 and \mathbb{G}^1 contacts). **g**, Empirical frequency distribution of membrane, synaptic and gap junctional contacts
1025 across the 4 datasets ($\delta = 1$ to 4). **h-j**, Survival distribution of contacts as a function of membrane
1026 contact area for \mathbb{M}^δ , \mathbb{C}^δ and \mathbb{G}^δ graphs (n given in **g**), plotting the probability that a
1027 membrane/synaptic/gap junction contact occur with membrane contact area $>$ some value). Membrane
1028 contact areas have been log-normalized and standardized so that the distribution is centered about 0, i.e.
1029 log-transformed, standardized (by subtracting the mean) and normalized (by dividing by the standard
1030 deviation), such that a range of ± 1 corresponds to ± 1 standard deviation of the distribution of
1031 $\log(\text{membrane contact area})$.

1032
1033 **Extended Data Fig. 3. Core and variable model validations. a,b**, Model fits for reproducibility of
1034 \mathbb{M}^δ contacts, with membrane contact areas **a**, below and **b**, above the log-normalized mean (after
1035 thresholding, Methods, Extended Data Fig. 2h). **c,d**, Reproducibility model fits of **c**, inter-cluster and **d**,
1036 intra-cluster \mathbb{M}^δ contacts. **e**, Reproducibility model fits for complete \mathbb{M}^δ , \mathbb{C}^δ and \mathbb{G}^δ datasets including
1037 membrane contact areas $< 35\%$ (results qualitatively similar to restricted dataset model fit in Figure 2a;
1038 Methods: Generating reference graphs). **f**, Reproducibility model fits for \mathbb{C}^δ excluding synaptic contacts
1039 scored in only 1 EM section (Methods). **g**, Reproducibility model fits for \mathbb{C}^δ excluding synaptic contacts
1040 derived from non-reproducible post-synaptic partners of polyadic synapses (Methods). **h,i**,
1041 Reproducibility model fits for synaptic and gap junction contact datasets scored by **h**, White *et al.*
1042 (1986)⁵ and **i**, Witvliet *et al.* (2020)²⁰ limited to our \mathbb{M}^4 contacts. \mathbb{M} : membrane, \mathbb{C} : chemical synapse
1043 and \mathbb{G} : gap junction contacts. Black bars: empirical distributions used in this study. Gray bars: other
1044 empirical distributions^{5,20}. Red bars: Model fits for the empirical distributions. All fractions of the total
1045 empirical counts (n).

1046
1047 **Extended Data Fig. 4. Validation of core-variable model and contact scoring. a-c**, Core-variable
1048 model reliably predicts the empirical synaptic and gap junction contact reproducibility (\mathbb{C}^δ and \mathbb{G}^δ) on
1049 \mathbb{M}^2 and \mathbb{M}^3 . To predict synaptic/gap junctional contact counts on $\mathbb{M}^{i < 4}$ contacts, \mathbb{C}^δ (or \mathbb{G}^δ) contact
1050 counts on \mathbb{M}^4 are scaled by the ratio of all \mathbb{C} (\mathbb{G}) on \mathbb{M}^i count : all \mathbb{C} (\mathbb{G}) on \mathbb{M}^4 count (Methods). E.g. in
1051 **a**, the model predicts a \mathbb{C}^3 count on \mathbb{M}^3 contacts as $206 \times 285 / 1474 = 40$ where 206 is the empirical \mathbb{C}^3
1052 count on \mathbb{M}^4 contacts, 285 is the total empirical synaptic contact count, \mathbb{C} , on \mathbb{M}^3 and 1474 is the total
1053 empirical count of synaptic contacts on \mathbb{M}^4 . The model prediction is consistent with the empirical \mathbb{C}^3 on
1054 \mathbb{M}^3 count (43). Error bars: $\pm \sqrt{n}$, where n is the empirical or predicted count (see Source Data for
1055 precise n values). **d**, Chemical synapses and **e**, gap junctions also consist of a core and variable circuit.
1056 Surrogate model data for \mathbb{C}^δ and \mathbb{G}^δ , generated as in Fig. 2b. Across each dataset, $\sim 62\%$ of synaptic
1057 contacts and $\sim 59\%$ of gap junction contacts consist of target contacts (given by $fp / [fp + (1-f)(1-s)]$,
1058 Methods). **f,g**, Core synaptic contacts are typically larger than variable ones in both Cook *et al.* (2019)³
1059 and White *et al.* (1986)⁵. Distribution of **f**, \mathbb{C}^δ and **g**, \mathbb{G}^δ contact counts by EM sizes (the total number of
1060 EM sections in which a contact was observed)^{3,7}. To check for biases in contact size due to possible
1061 differences in synaptic/gap junction scoring criteria, we compare the distributions of EM sizes for
1062 contacts identified by White *et al.* (1986)⁵ (orange) and those identified by Cook *et al.* (blue). Because
1063 White *et al.* (1986)⁵ does not provide EM sizes, we used the EM sizes from Cook *et al.* (2019)³ for all

1064 contacts. Although many additional synapses identified by Cook *et al.* (2019)³ occur only in 1 EM
1065 section, we find no systematic bias towards smaller synaptic contacts by Cook *et al.* (2019)³. **h,i**,
1066 Bidirectional comparison of Cook *et al.* (2019)³ and Witvliet *et al.* (2020)²⁰ synaptic contact
1067 reproducibility. **h**, Fraction of Cook *et al.* (2019)³ synaptic contacts scored by Witvliet *et al.* (2020)²⁰. **i**,
1068 Fraction of Witvliet *et al.* (2020)²⁰ synaptic contacts scored by Cook *et al.* (2019)³. **h,i**, Fractions of the
1069 total empirical count of synaptic contacts (n).
1070

1071 **Extended Data Fig. 5. Robust clustering of nerve ring processes from \mathbb{M}^4 spatial population**
1072 **models.** The variability of membrane contacts (Fig. 2, Extended Data Fig. 2) suggest that no single
1073 animal is representative of the population. We estimated the variability among membrane contact areas.
1074 **a**, The log-normalized empirical distribution of \mathbb{M}^4 membrane contact areas (mean centered at 0, STD:
1075 standard deviation, red line: normal distribution with empirical mean and standard deviation, $n = 1,258$
1076 membrane contacts). We estimated the variability across the four datasets (L4 left, L4 right, adult left
1077 and adult right). For each conserved \mathbb{M}^4 contact, we computed the mean and standard deviation of the
1078 membrane contact area across the four datasets (see Methods). **b**, Plot of the standard deviation versus
1079 mean contact area across the datasets, where each point is one \mathbb{M}^4 contact. Similar to Extended Data
1080 Fig. 1a, we find no dependence of the variability on membrane contact area. Therefore, we estimate
1081 membrane contact area variability by the mean variability among all membrane contact areas. **c**, The
1082 distribution of standard deviations of membrane contact area for all \mathbb{M}^4 contacts. Red dashed line: mean
1083 standard deviation. **d-i**, A stochastic spatial population model matches the above distributions by
1084 randomly perturbing membrane contact areas in the four datasets with multiplicative white noise with
1085 standard deviation (σ) of 0.23 (Methods). **d-f**, Spatial population data perturbs the membrane contact
1086 areas while maintaining contact area and variability distributions that are similar to the empirical \mathbb{M}^4
1087 contact area distributions. **g**, Perturbed contact areas scale linearly with the empirical contact areas. **h**,
1088 The spread of perturbed contact areas (log of the perturbed contact area as a fraction of the empirical
1089 contact area) is mostly uniform across membrane contact areas. **i-l**, Neurite clusters obtained from a
1090 population of 1,000 $\widetilde{\mathbb{M}}^4$ perturbed individuals and 1,000 $\widetilde{\text{L4}}$ and $\widetilde{\text{Adult}}$ perturbed individuals (perturbing
1091 left/right conserved contacts in the L4 and adult datasets). For each perturbed individual in each
1092 population we used a multi-level graph clustering algorithm to identify spatial clusters. Across each
1093 population, we computed the frequency that cell pairs cluster together, represented as an $n \times n$ cluster
1094 frequency matrix ($n = 93$). A hierarchical clustering algorithm is used to sort the rows and columns of
1095 the cluster frequency matrix in order to minimize variation along the diagonal. Hence, cells pairs that
1096 frequently cluster together are sorted together on the cluster frequency matrix (Methods). Five largely
1097 overlapping subgroups of neurons emerge across different perturbations (see main text). **i**, Consensus
1098 subgroups are robust across datasets. $\widetilde{\text{L4}}$ and $\widetilde{\text{Adult}}$ clusters visualized using row and column colors of
1099 the \mathbb{M}^4 population cluster assignments (dashed box). **j**, The consensus subgroups are robust across
1100 different noise amplitudes. Clustering applied to populations generated by perturbations to \mathbb{M}^4 using
1101 white noise with standard deviations 0 (empirical data), 0.12, 0.45 and 0.9. **k,l**, The consensus
1102 subgroups are robust across different spatial domains. **k**, Clustering applied to \mathbb{M}^4 populations
1103 generated from the more spatially restricted subset of the neuropil considered by Moyle *et al.* (2020)³⁴,
1104 which excluded the posterior lobe of the neuropil. **l**, Clustering applied to populations generated by
1105 perturbations to all reproducible membrane contact areas after restoring the smallest 35% contact
1106 areas to each of the L4, adult and \mathbb{M}^4 datasets (Extended Data Fig. 2). For all cluster frequency
1107 matrices: Matrix element (i, j) corresponds to the frequency that cells i and j cluster together across the

1108 1000 perturbed individuals. Row and column orders minimize variance along the diagonal (Methods).
1109 Cell cluster assignments (color) follow the perturbed $\widetilde{\mathbb{M}}^4$ dataset (Figure 1b reproduced in dashed box).
1110 Top: dendrogram of the hierarchical clustering.

1111

1112 **Extended Data Fig. 6. Variable contacts obscure the organization of the nerve ring. a,** Cluster
1113 analysis of unperturbed membrane contact datasets \mathbb{M}^1 , \mathbb{M}^2 , \mathbb{M}^3 and \mathbb{M}^4 . Clustering results for
1114 membrane contacts predicted to combine core and variable contacts (\mathbb{M}^3) and overwhelmingly variable
1115 contacts (\mathbb{M}^2 , \mathbb{M}^1) significantly and increasingly diverge from 5 consensus clusters, indicated by large
1116 numbers of small clusters. **b,** Cluster analysis of (unperturbed) L4 and adult datasets. Both the
1117 unperturbed \mathbb{M}^4 and adult datasets yield 6 clusters rather than the 5 clusters found in the perturbed
1118 population models (Figure 1c and Extended Data Fig. 5). The additional cluster results from a split of
1119 the taxis cluster into two. This split of the taxis cluster is not observed in either the perturbed \mathbb{M}^4 or the
1120 perturbed Adult dataset, even with half the noise levels observed empirically, indicating that the split is
1121 unlikely to be robust across a population of animals. For all cluster frequency matrices: Row and
1122 column ordering and colors are the same as the perturbed $\widetilde{\mathbb{M}}^4$ population dataset (Figure 5i). Matrix
1123 element (i, j) is 1 if cells i and j cluster together and 0 otherwise. Top: dendrogram of the hierarchical
1124 clustering.

1125

1126 **Extended Data Fig. 7. Distribution of core and variable synapses among neighborhoods. a,**
1127 Membrane contacts of the L4, adult and reference \mathbb{M}^4 datasets demonstrate that all three datasets have
1128 similar membrane contact profiles. For L4 and adult datasets, only bilaterally conserved contacts are
1129 included. **b,** Synaptic contacts on \mathbb{M}^4 membrane contacts broken down by degree of synaptic contact
1130 reproducibility (\mathbb{C}^1 , \mathbb{C}^2 , \mathbb{C}^3 and \mathbb{C}^4). Most (56%) of conserved synapses (\mathbb{C}^4) occur within clusters near
1131 the main diagonal, while variable synapses (\mathbb{C}^1) are spread across clusters. **c,** Gap junction contacts on
1132 \mathbb{M}^4 membrane contacts broken down by degree of reproducibility (\mathbb{G}^1 , \mathbb{G}^2 , \mathbb{G}^3 and \mathbb{G}^4). For all matrices:
1133 Row and column ordering is the same as the perturbed \mathbb{M}^4 dataset (Extended Data Fig. 5i). Row and
1134 column colors correspond to final clusters assignments (Fig. 1c), where unclassified cells are colored
1135 gray. Matrix element (i, j) corresponds to the fraction of cell i 's membrane contact with cell j , with rows
1136 normalized to sum to 1.

1137

1138 **Extended Data Fig. 8. Subcellular structures support local and nonlocal connectivity; RIA and**
1139 **AIB processes demonstrate synaptic compartmentalization. a,b,** Volumetric rendering of RIAL and
1140 its synapses (cuboids) colored by **a**, synaptic polarity or **b**, intra-/inter-cluster. Combining **a** and **b**:
1141 synaptic input and output segments correspond to changes in neighborhood composition. Changes in
1142 RIA neighborhood correspond to the 3 neurite segments (nV, nD and loop) which exhibit independent
1143 calcium dynamics that encode head movement⁶². **c,d,** AIB processes change neighborhood at the lateral
1144 midline¹⁸. The ipsilateral segment (\dagger) of the AIB process is surrounded by cells in the taxis cluster
1145 while the contralateral segment ($\dagger\dagger$) makes contact with cells in every other cluster. **c,** AIB process
1146 segments alternate between synaptic inputs on the ipsilateral side and synaptic outputs on the
1147 contralateral side. **d,** The alternating synaptic inputs and outputs correspond to a change in
1148 neighborhood occurring at the dorsal midline. **e-h,** Flattened protrusions link processes to adjacent cells
1149 in adjacent clusters. **e,** The flattened protrusion strategy as demonstrated by RIM processes (\blacklozenge). **f,** The
1150 RMDV processes demonstrate how flattened protrusions are used to locally expand synaptic polarity.
1151 On the contralateral side, the main process trajectory is postsynaptic while the contralateral protrusion is

1152 presynaptic. Both **g**, AVA and **h**, SAAV exhibit flattened protrusions that appear to turn into small
1153 branches. The small AVA branch extends into a neighborhood comprised of cells from a different
1154 cluster (*). SAAV ipsilateral branches receive synaptic inputs while its main process trajectory on the
1155 contralateral side is mostly pre-synaptic. RMEV/D processes spine-like features extend to cells in a
1156 different cluster. **i**, 2 longer RMED extensions and **j**, 3 shorter RMEV spine-like extensions are
1157 postsynaptic to the sublateral cluster. In all images, the pharynx is shown for a spatial reference. R:
1158 right, A: anterior, V: ventral. Note: for visual clarity, synapses have been offset from the cell process. **k**,
1159 Schematic of neighborhood changes of selected cells (labeled in color of cluster assignment). P:
1160 proximal and D: distal to cell body. Each trajectory scaled to the length of the reconstructed left L4
1161 process. Black boxes denote sections in which the process makes contact with at least two clusters.
1162

1163 **Extended Data Fig. 9. Network features of the brain map.** **a**, Schematics of network features (from
1164 left to right): Feed-forward loop motif (FF) defined by a triplet of nodes with connectivity: *Source* →
1165 *Intermediary* → *Target* and *Source* → *Target*; network hub (high degree node, H); fan-in (high in-
1166 degree node, FI); fan-out (high out-degree node, FO); and rich club (highly connected hubs, RC). **b**, FF
1167 triplets within the brain map support the ResNet architecture of the nerve ring. All 101 FF instances
1168 among \mathbb{C}^4 synaptic contacts (all edges in Fig. 4, Extended Data Fig. 10) are shown. Black arrows: FF
1169 synaptic contacts within the ResNet architecture (Fig. 4). Gray arrows: additional FF synaptic contacts
1170 (Extended Data Fig. 10). 72/93 cell classes participate in at least one FF motif. Prominent FF targets
1171 include: AIA, AIB, AIZ, AVA, AVB, AVE, RIA, RIC, RIM, RIP, RMDV and SMDV. Additional
1172 contacts superimposed on the ResNet come mostly from cross-sensory module connectivity (Extended
1173 Data Fig. 10b). **c**, RIP, the only synaptic link between the somatic and pharyngeal nervous systems, is a
1174 major FF target cell for papillary sensory source cells and URA intermediaries. **d**, AIA are a major taxis
1175 Layer-2 intermediary cell pair regulating information flow from Layer-1 Taxis sensory cells onto the
1176 Layer-3 AIB Taxis target cell. **e**, AIZ, major Layer-3 cells that supports nonlocal connectivity (Fig. 3a),
1177 serve both to integrate information flow from Layers-1 and 2 Taxis source cells (FI) and as an
1178 intermediary to various Layer-3 target cells in other modules (FO). **f**, Primary locomotion regulating
1179 interneurons – AVA, AVB and AVE – are major Layer-3 FF targets and connect extensively onto
1180 motoneurons of the ventral nerve cord. Connectivity among these cells occurs in the ventral nerve cord
1181 (but not observed in the nerve ring), suggesting that the regulation of locomotion down the body occurs
1182 posteriorly to the nerve ring. **g**, RIM, a major hub that support nonlocal connectivity, triples as source,
1183 intermediary and target FF cell pair within Layer-3. **h**, The nonlocal supporter, multi-compartment pair
1184 RIA are a major FF target for Layer-1 sensory (primarily Avoidance) source cells and Layer-2/3 (Taxis
1185 and Avoidance) intermediary cells as well as intermediaries that control Layer-3 head motoneurons.
1186 Additionally, RIA are major targets for feedback from lateral (RMD, RMDD, RMDV) and sublateral
1187 (SMDD, SMDDV) head motoneurons, consistent with their roles in spatially encoding dorso-ventral
1188 head movement to coordinate turning behaviors.⁴⁷ **i**, Major FF targets (11 neuron classes acting as a
1189 target of > 3 FF motifs, including 5 RC classes) form a highly interconnected subnetwork. Note the
1190 frequent representation of some cells in multiple motifs (**c-i**). **j**, Layer-3 aggregated FF synaptic
1191 contacts within and among the modules shows strong recurrence and no clear feed-forward
1192 directionality or hierarchy of Layer-3 connectivity, consistent with highly distributed computation.
1193 Sublaterals are merged into the Lateral module node. Layer-3 anterior cells form FF connectivity with
1194 only one other module (Taxis). All network schematics were generated with Cytoscape 3.7.1.
1195

1196 **Extended Data Fig. 10. 17% of \mathbb{C}^4 contacts are not accounted for by the ResNet model. a**, Layer-1
1197 synaptic connectivity across information processing modules in \mathbb{C}^4 could support distributed sensory
1198 computation and integration. 8 (2% of \mathbb{C}^4) contacts occur between sensory cells across different
1199 modules. These contacts include: (i) ADE→OLL, (ii) ALM→CEPD/V, (iii) reciprocal contacts
1200 between ASH, ADL and AFD. (i) Mechanosensitive^{48,49} anterior cell OLL loops around intermediate
1201 processes, while the processes of ADE extend toward the OLL loop, suggesting a functional role for the
1202 more elaborate loop morphology. (ii) Both CEPD and CEPV processes loop around intermediate
1203 processes and extend flattened protrusions to meet the ALM processes, where ALM are postsynaptic.
1204 CEPD and CEPV respond to head touch⁵⁰, while ALM respond to both gentle⁵¹ and harsh⁵² body touch,
1205 inhibit backward locomotion⁵³ and have been implicated in the habituation of tap response⁵⁴. (iii)
1206 Nociception: ASH, ADE and ADF may coordinate avoidance behaviors between the taxis and
1207 avoidance modules⁵⁵. **b**, Layer-1 to Layer-3 inter-module feed-forward synaptic connectivity in \mathbb{C}^4 . 54
1208 (12% of \mathbb{C}^4) contacts are inter-module, originate in Layer 1 and target Layer 3 neurons directly. A small
1209 number of taxis and avoidance sensory neurons (ADF and ADL, ASH, URX and BAG) project to all
1210 but Laterals in Layer 3; this contrasts with extensive anterior sensory neuron projections that almost
1211 exclusively target (sub)lateral Layer-3 interneurons and motoneurons, likely mediating rapid
1212 sensorimotor transformations. **c**, Layer-2 and inter-module feed-forward \mathbb{C}^4 synaptic connectivity. 3
1213 contacts (1% of \mathbb{C}^4) are inter-module and originate in Layer 2. Notably, Layer-2 taxis AIY neurons
1214 synapse onto Layer-3 anterior multi-compartment neurons RIA. **d**, Inter-module feedback synaptic
1215 connectivity in \mathbb{C}^4 . 9 (3% of \mathbb{C}^4) contacts provide inter-module feedback. Black arrows: synaptic
1216 contacts between cells in the same neighborhood. Grey arrows: synaptic contacts between layer 3 cells
1217 in different neighborhoods. Red arrows: synaptic contacts not accounted for by the ResNet model. Solid
1218 arrows: feed-forward or recurrent (intra-layer) synaptic contacts. Dashed arrows: feedback synaptic
1219 contacts.

1220

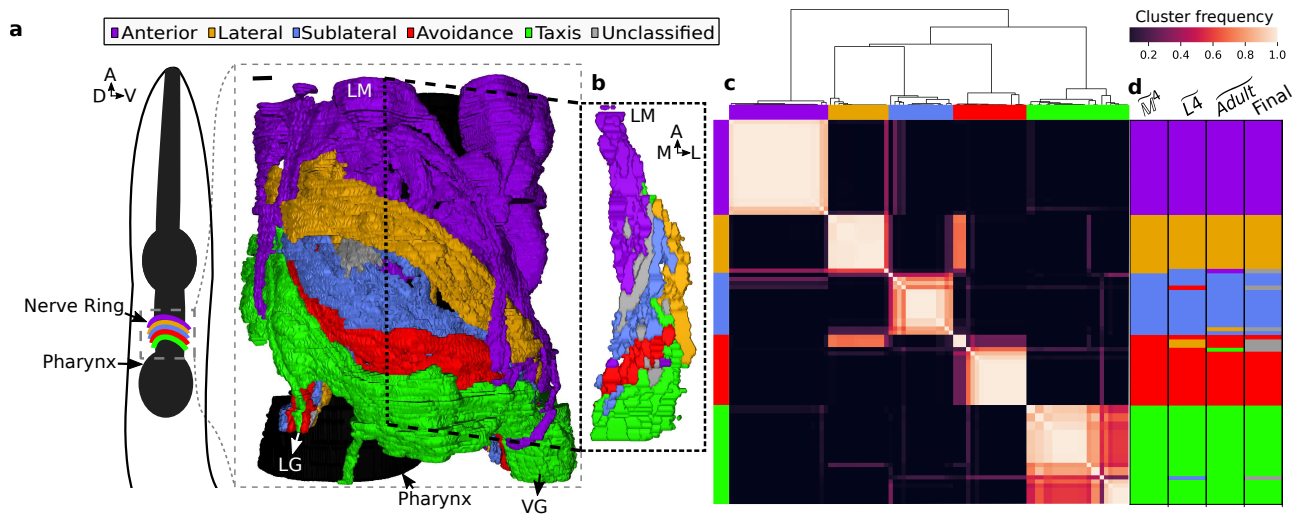


Figure 1

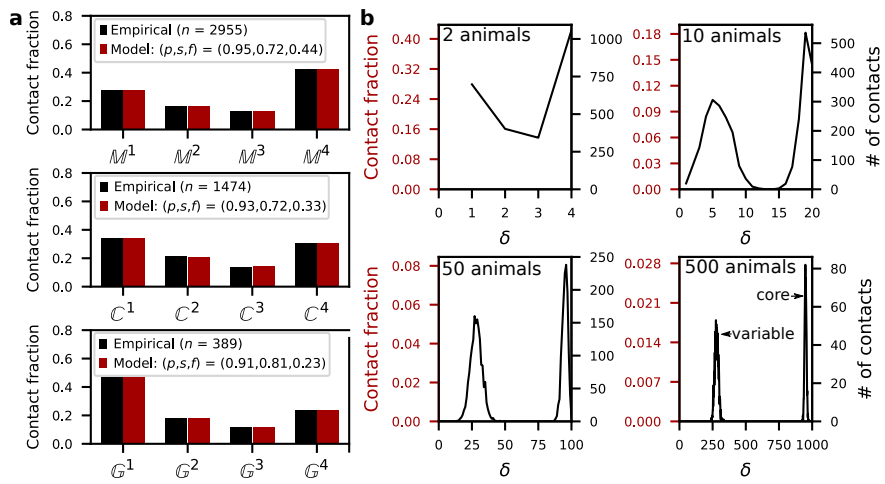


Figure 2

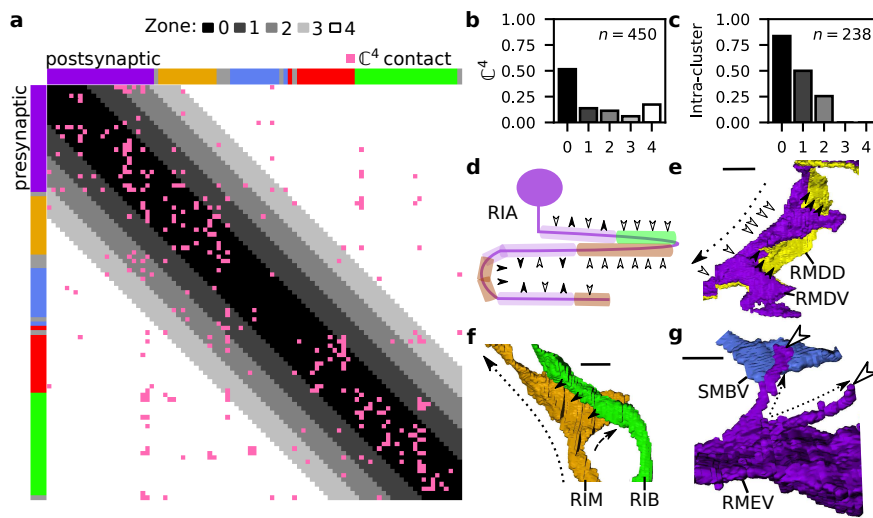


Figure 3

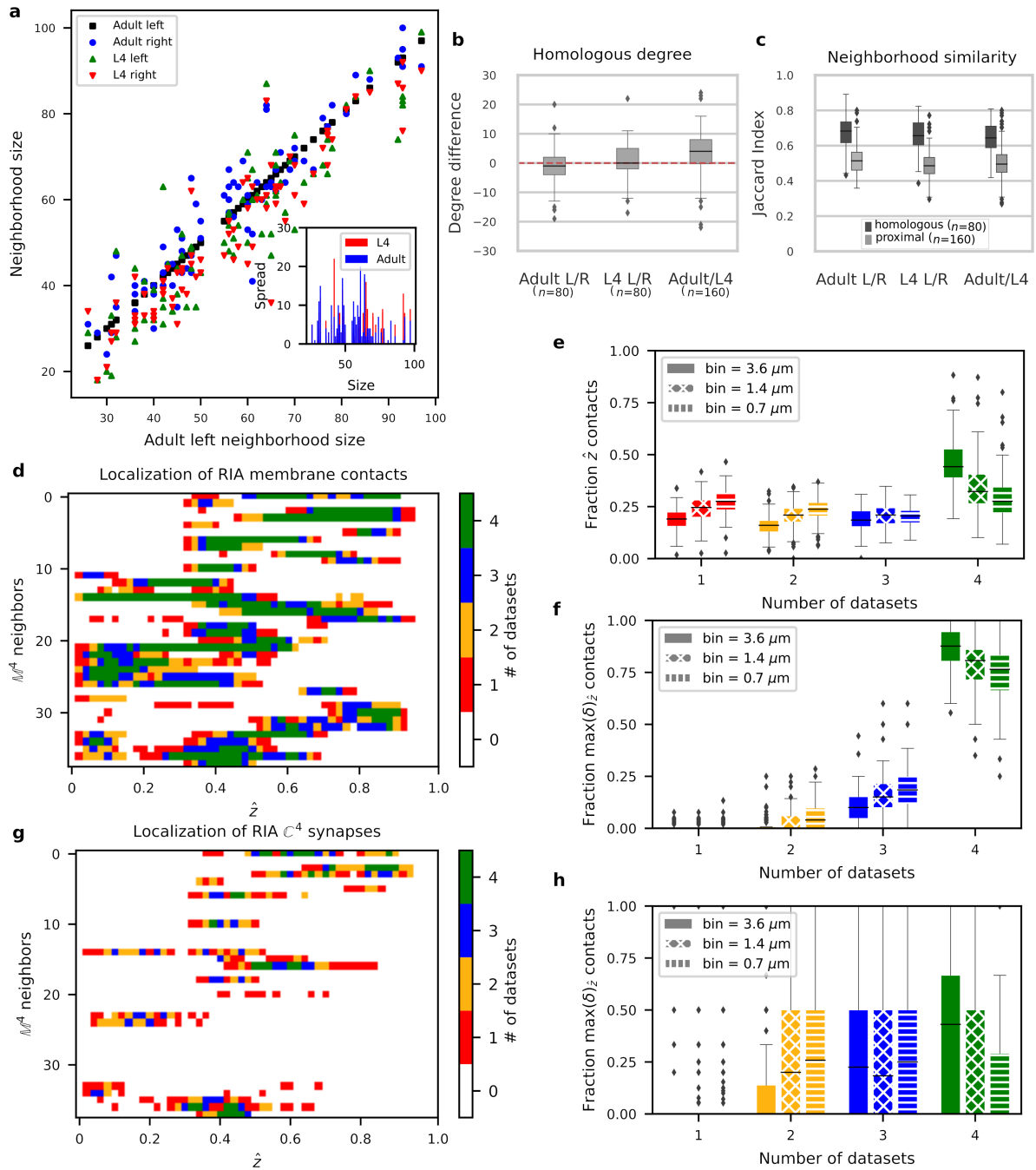


Figure Extended Data Figure 1

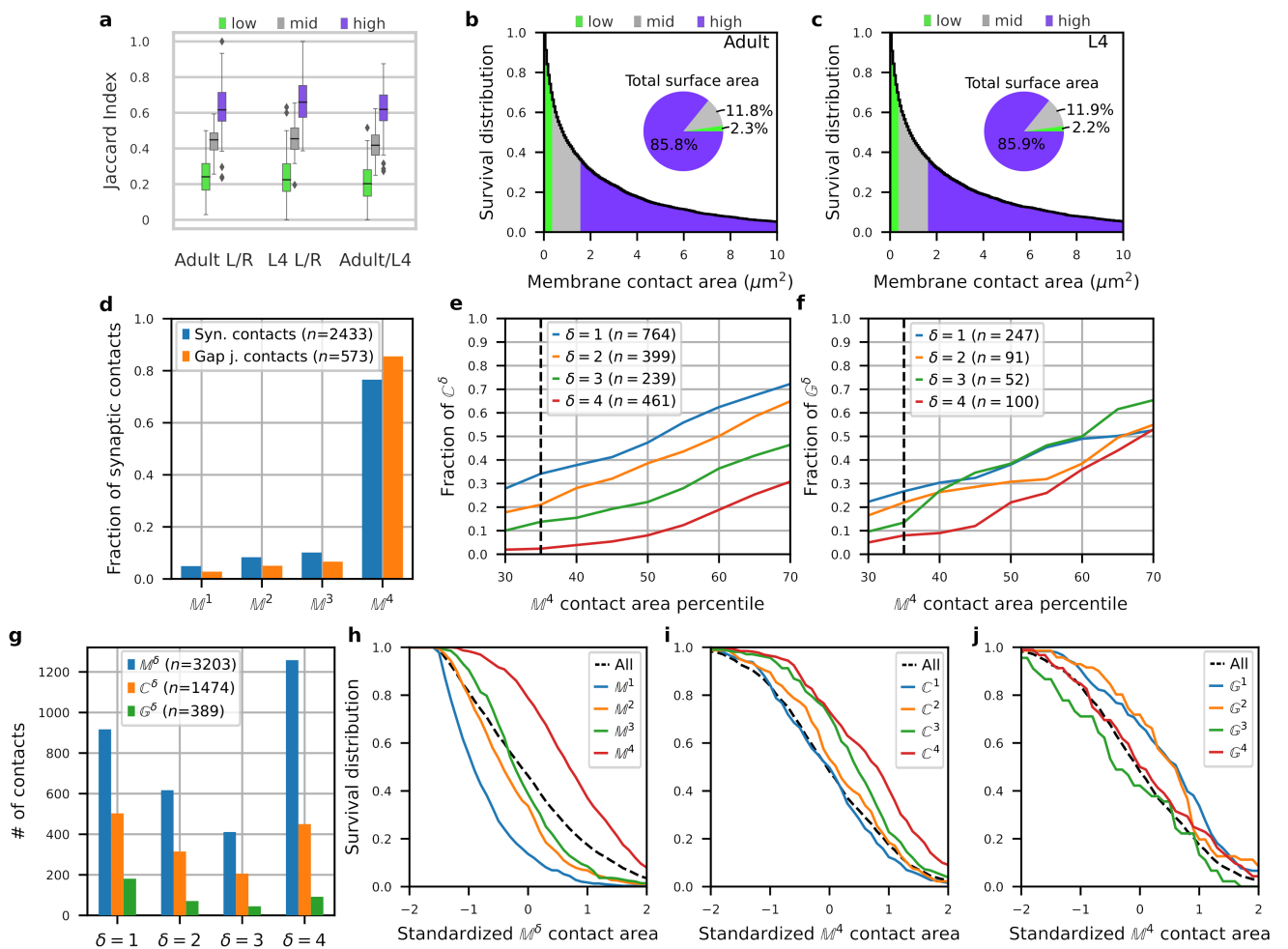


Figure Extended Data Figure 2

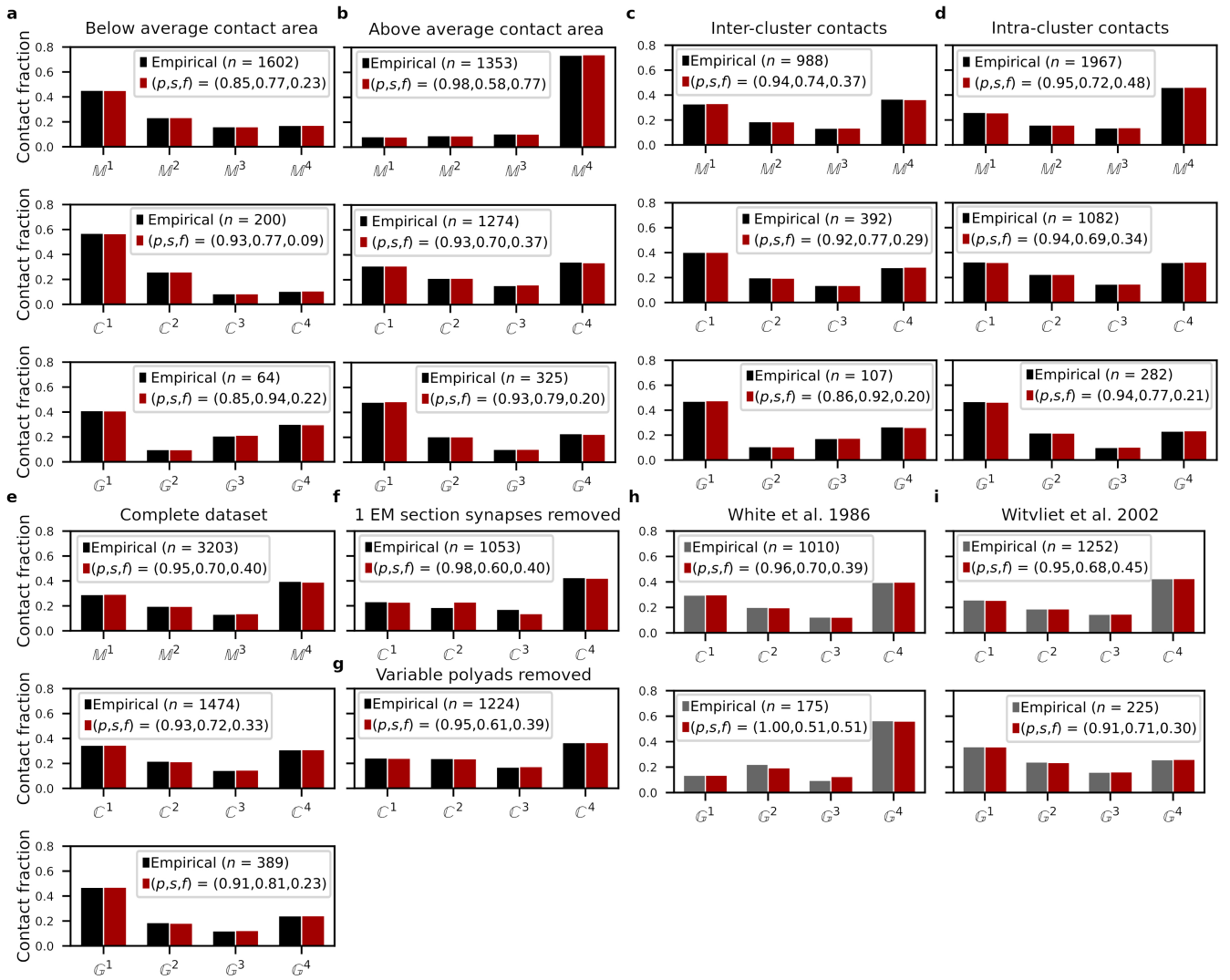


Figure Extended Data Figure 3

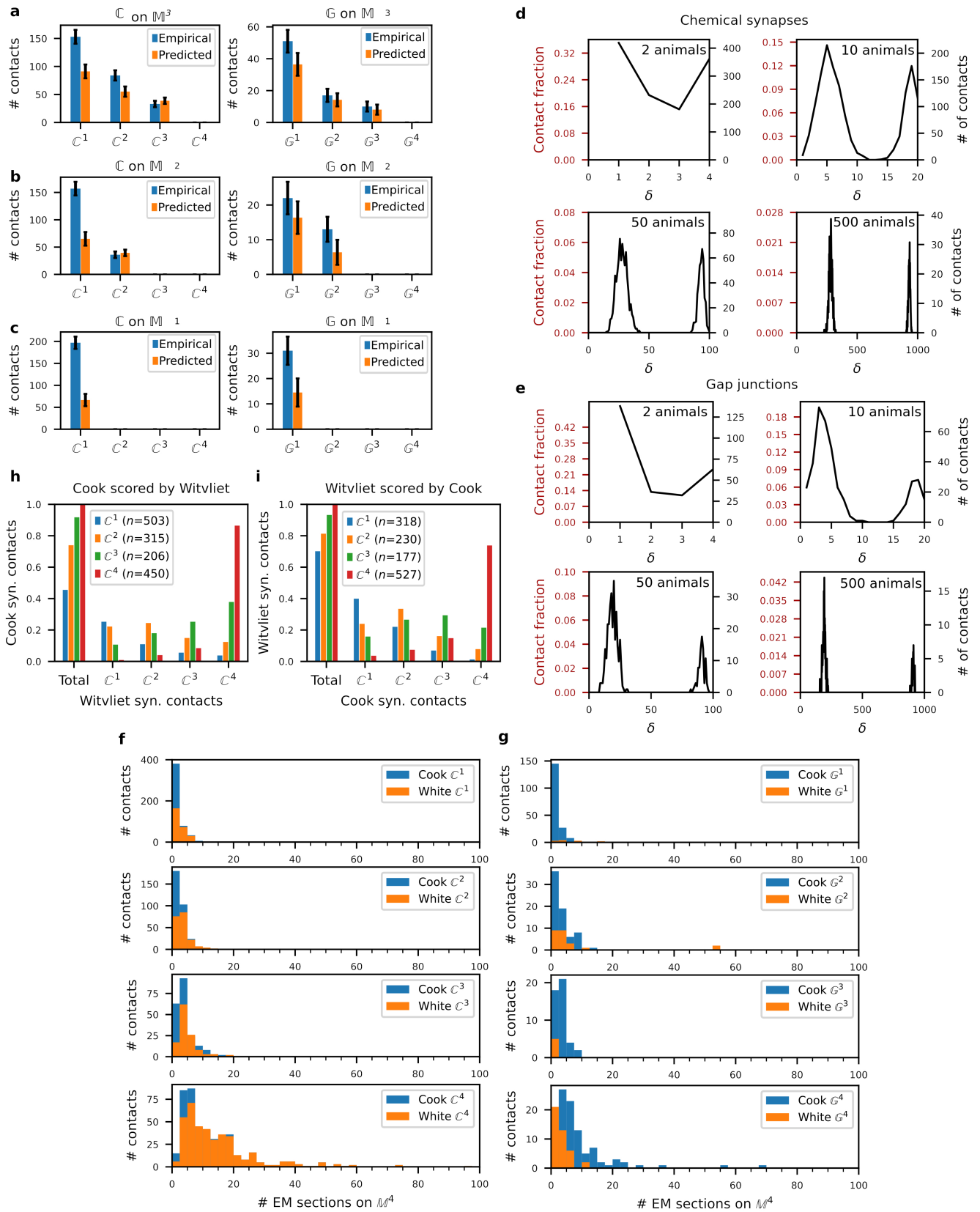


Figure Extended Data Figure 4

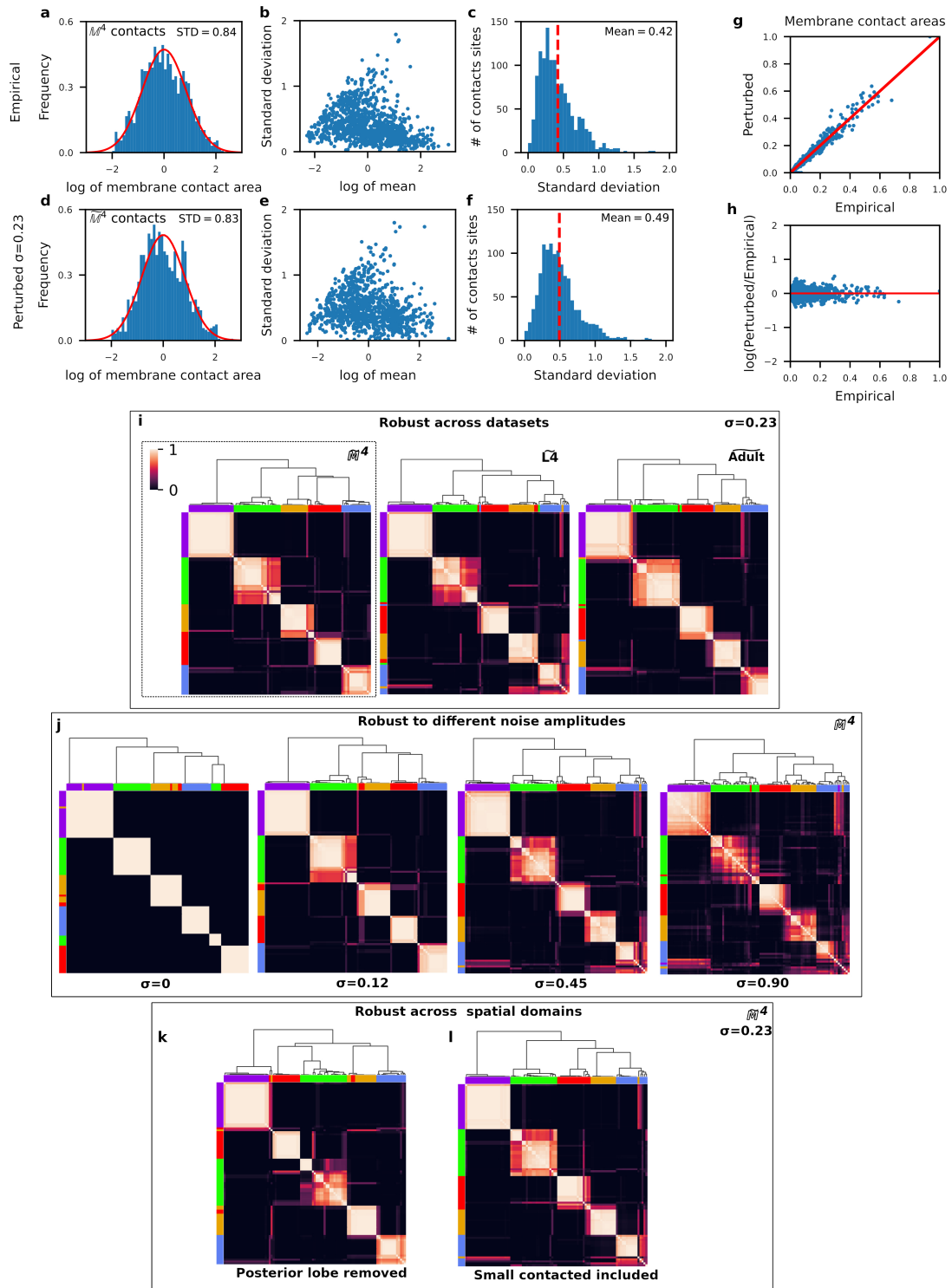


Figure Extended Data Figure 5

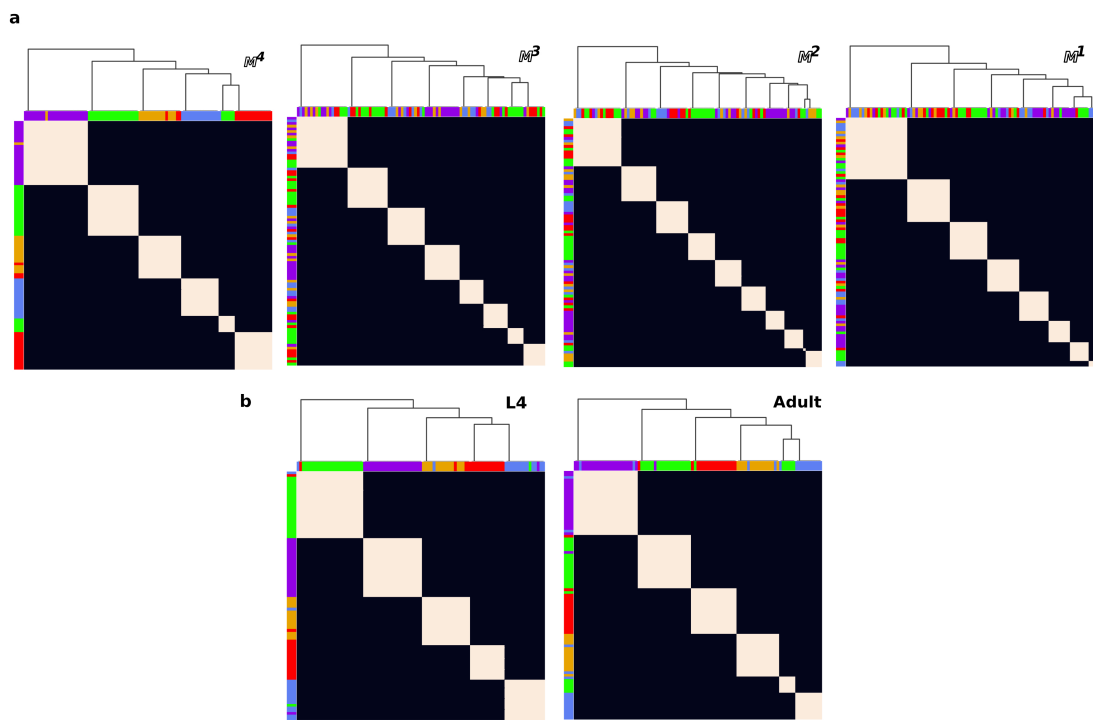


Figure Extended Data Figure 6

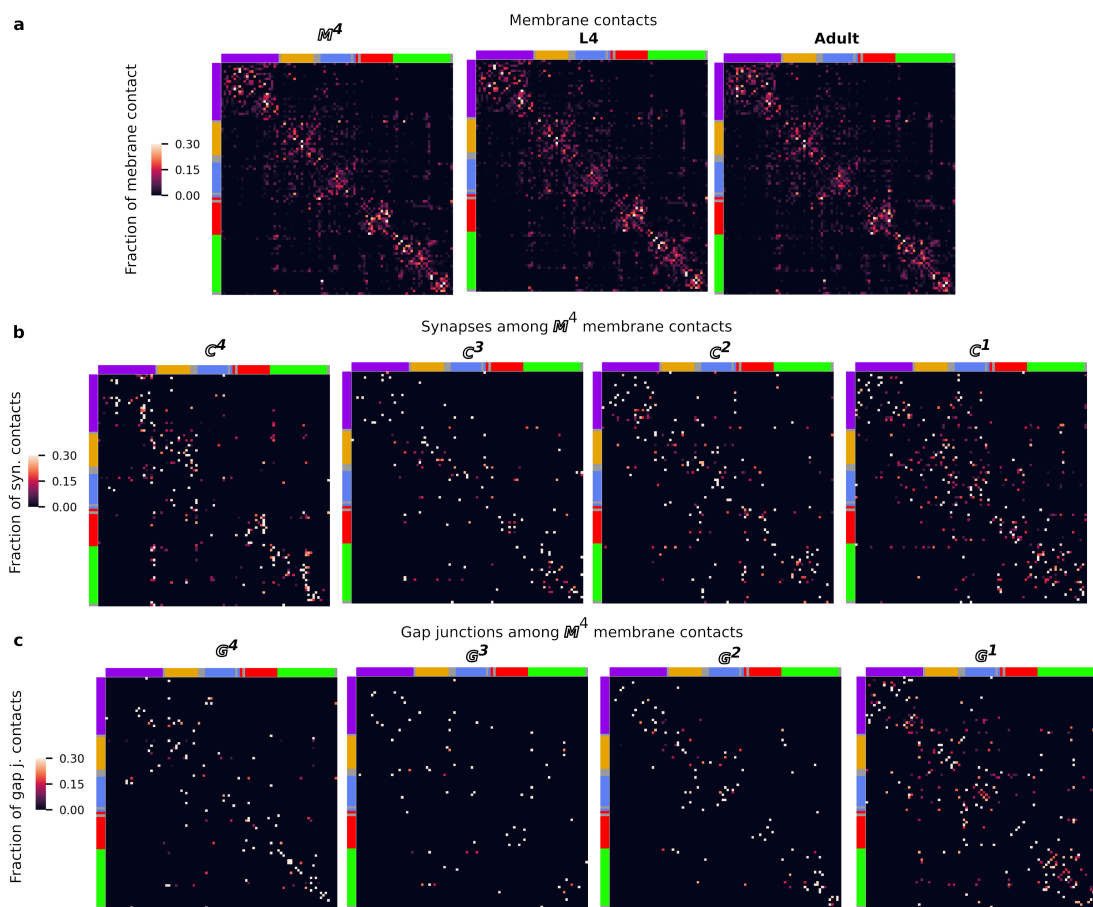


Figure Extended Data Figure 7

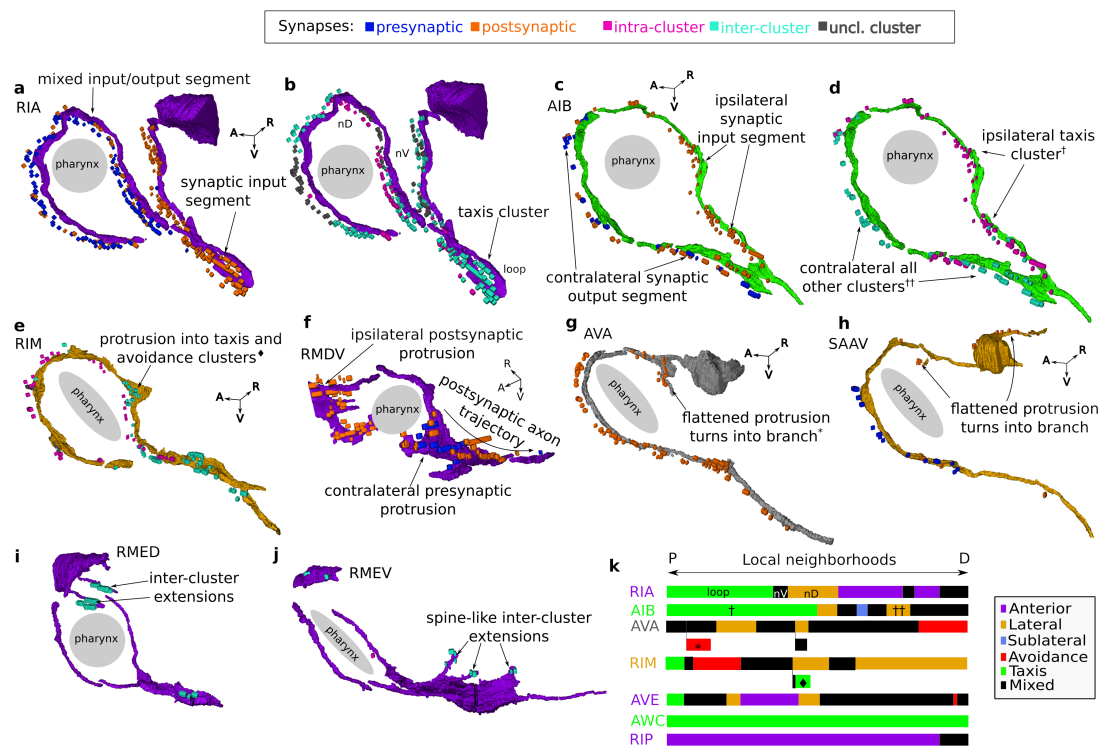


Figure Extended Data Figure 8

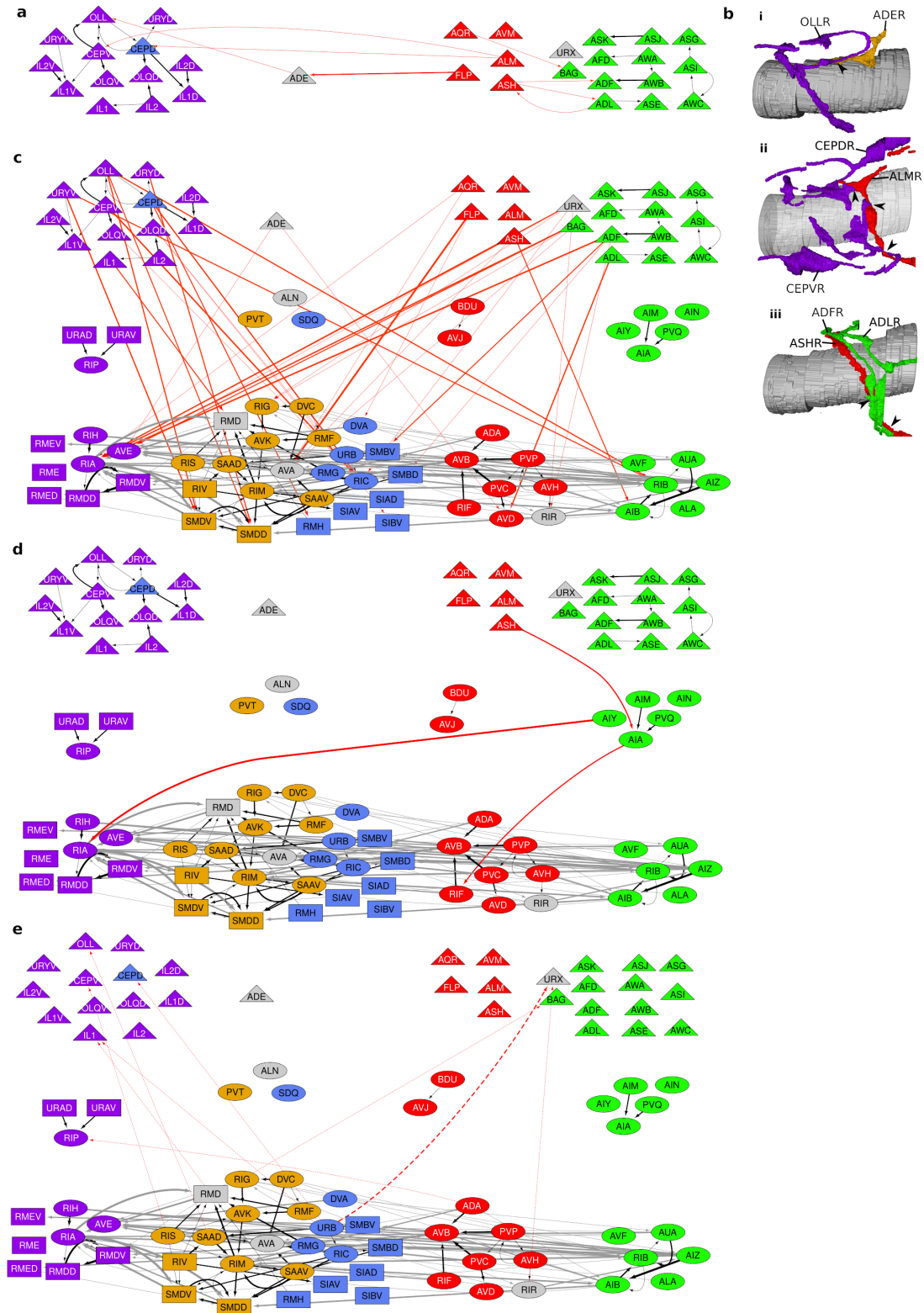


Figure Extended Data Figure 10