Article

A multi-scale brain map derived from whole-brain volumetric reconstructions

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

Received: 23 May 2020

Accepted: 25 January 2021

Published online: 24 February 2021

Check for updates

Christopher A. Brittin[1,2,4], Steven J. Cook[3,5], David H. Hall[3], Scott W. Emmons[2,3,6] & Netta Cohen[1,6]✉

Animal nervous system organization is crucial for all body functions and its disruption can lead to severe cognitive and behavioural impairment[1]. This organization relies on features across scales—from the localization of synapses at the nanoscale, through neurons, which possess intricate neuronal morphologies that underpin circuit organization, to stereotyped connections between different regions of the brain[2]. The sheer complexity of this organ means that the feat of reconstructing and modelling the structure of a complete nervous system that is integrated across all of these scales has yet to be achieved. Here we present a complete structure–function model of the main neuropil in the nematode *Caenorhabditis elegans*—the nerve ring—which we derive by integrating the volumetric reconstructions from two animals with corresponding[3] synaptic and gap-junctional connectomes. Whereas previously the nerve ring was considered to be a densely packed tract of neural processes, we uncover internal organization and show how local neighbourhoods spatially constrain and support the synaptic connectome. We find that the *C. elegans* connectome is not invariant, but that a precisely wired core circuit is embedded in a background of variable connectivity, and identify a candidate reference connectome for the core circuit. Using this reference, we propose a modular network architecture of the *C. elegans* brain that supports sensory computation and integration, sensorimotor convergence and brain-wide coordination. These findings reveal scalable and robust features of brain organization that may be universal across phyla.

A primary goal of systems neuroscience is to understand how the structure and function of the brain combine to generate behaviour. Since the discovery of neurons and their connections through synapses and gap junctions, major efforts have been devoted to characterize these units and the micro- and macro-circuits that they comprise, culminating in a growing body of nanoconnectomic[2] (synaptic resolution) data across species[3–12]. Nonetheless, structural data alone—however rich—cannot provide explanatory power to address the computation within circuits or to determine how these circuits communicate and coordinate information flow to generate behaviour. Indeed, constructing a comprehensive brain map will require a meaningful strategy for the multi-scale integration of structure and function. Achieving this feat in even a small animal can provide a useful model for inferring principles of brain organization[2].

The free-living nematode *C. elegans* has a small, compact nervous system[3,5,7,13] and yet exhibits a range of complex, individualized behaviours, making it an ideal model system for studies of whole-brain organization[2]. All 302 neurons in *C. elegans* have been anatomically characterized from serial-sectioned electron micrographs[5] to produce a whole-animal connectome[3,5,13]. The invariant cell lineage[14] and anatomy[5] of *C. elegans* might suggest that its connectome too is invariant[15]. Unfortunately, the small sample size of available reconstructions has precluded a reliable estimate of the reproducibility and variability of the synaptic connectome. Furthermore, whereas the synaptic wiring has been exhaustively characterized[3,5,13,16,17], the spatial proximity of neurons is only partially described[18,19]. Thus, it remains to be determined whether lessons about whole-brain organization in *C. elegans* can inform questions and approaches for other systems.

We provide two complete volumetric reconstructions of the *C. elegans* nerve ring from previously published electron micrographs[5], one from an adult and one from a larval stage 4 (L4) worm (Methods, Supplementary Table 1, Supplementary Videos 1–3, Supplementary Data 1. The two series of electron micrographs (with roughly 300 sections in the L4 worm and 400 sections in the adult) span approximately the same volume with a length of around 36 µm, starting in the anterior and ending in the ventral ganglia (Fig. 1a). To our knowledge, our reconstructions of these two nerve rings provide the first complete, nanoscale-resolution datasets of all neuronal membrane contacts of any neuropil. We dub the set of membrane contacts the 'contactome' of the brain. We define two neurons as immediate neighbours if the membranes along their neural processes are physically adjacent in at least one electron micrograph[18]. To characterize synaptic pathways within a spatial context, we integrated our volumetric reconstructions with our recent rescoring of synapses on the same L4 and adult

[1]School of Computing, University of Leeds, Leeds, UK. [2]Department of Genetics, Albert Einstein College of Medicine, New York, NY, USA. [3]Department of Neuroscience, Albert Einstein College of Medicine, New York, NY, USA. [4]Present address: Developmental Biology Program, Sloan Kettering Institute, New York, NY, USA. [5]Present address: Department of Biological Sciences, Columbia University, New York, NY, USA. [6]These authors jointly supervised this work: Scott W. Emmons, Netta Cohen. ✉e-mail: N.Cohen@leeds.ac.uk

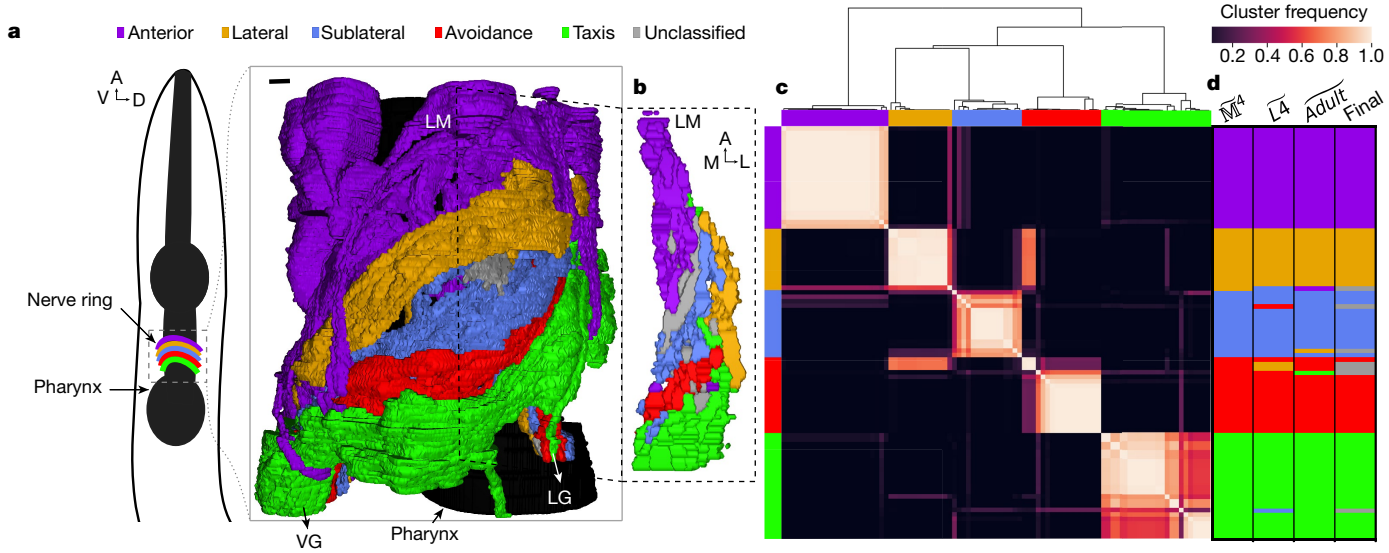


Fig. 1 | The nerve-ring neuropil comprises five densely connected neurite clusters. a, The nerve-ring neuropil (less than 4% of the body length and the most synaptically dense region of the *C. elegans* nervous system) includes neurites of 181 L4 (185 adult) neurons. The complete volumetric reconstruction of the L4 neuropil spans a length of 36 µm (Supplementary Video 3). Inset, a 15-µm-long region, view of the left side with the superficial neurons removed. A, anterior; D, dorsal; V, ventral; LG, lateral ganglia; LM, lateral midline; VG, ventral ganglia. Scale bar, 1 µm. **b**, A 250-nm oblique volumetric slice at approximately the lateral midline, rendered with no processes removed. Neurites with relatively high spatial affinity (but no physical boundaries) form spatially ordered clusters along the anterior–posterior axis. L, lateral; M, medial. **c**, Cluster matrix showing the frequency with which cells *i* and *j* cluster together across the population $\widetilde{\mathbb{M}^4}$. The order of rows and columns minimizes the variance in frequency along the diagonal. Clusters were then ordered to visually match anterior–posterior ordering (original ordering in Extended Data Fig. 5i). Top, dendrogram of the hierarchical clustering. **d**, Clustering results of model $\widetilde{\mathbb{M}^4}$, $\widetilde{L4}$ and $\widetilde{Adult}$ populations (Extended Data Fig. 5i) and consensus cluster assignment across the three populations. The order of rows and columns is the same as in **c**. Seven cell pairs (ADE, ALN, AVA, RID, RIR, RMD, URX, see Methods) with discrepant cluster assignments among the three populations are unclassified (grey). $n = 1{,}000$ perturbed datasets per population (Methods).

worms[3] (for validation and comparison with other datasets[5,20], see Methods).

Conserved and variable circuits overlap

Consistent with previous work[5], our volumetric reconstructions show that neural processes are bilaterally (left–right) conserved (Supplementary Results, Supplementary Videos 4–7). We hypothesized that the bilateral symmetry of *C. elegans* processes extends to the nanoscale to support a homology of membrane contacts and synapses between cells. Homologous processes exhibit statistically significant similarity in the size and composition of their immediate neighbourhoods (Extended Data Fig. 1a–c) and in the locations of membrane contacts along their processes (Methods, Extended Data Fig. 1d–f, Supplementary Data 2). By contrast, the smallest 35% of membrane contacts (less than 0.4 µm^2) are not reproducible (Extended Data Fig. 2a), account for only 2% of total membrane contact area between all neurons (Extended Data Fig. 2b, c) and contain synaptic contacts that are predominantly non-reproducible (Extended Data Fig. 2e). As such, we exclude them from our analysis. We conclude that the reproducibility of neuronal processes and their immediate neighbourhoods supports a stereotyped pattern of cell–cell membrane contacts.

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

To examine this hypothesis, we exploit the combined spatial and synaptic information across datasets over the entire neuropil. We assume that stereotyped wiring patterns require precision to find target neurons and specificity to avoid off-target neurons, and formulate statistical models of membrane and synaptic contacts to capture their relative propensity to occur in one, two, three or all four of the datasets (Methods). We find that a minimal model with three parameters suffices (Methods); these are the fraction of target contacts (f); the precision (p) for target contacts; and the frequency to avoid off-target contacts, or specificity (s). Despite their parsimony, these models yield good fits for the distribution of membrane, synaptic and gap-junctional contacts across the four datasets (Methods, Fig. 2a). The high reproducibility of membrane contacts across datasets ($\mathbb{M}^4$ count) is consistent with our model prediction that less than half of membrane contacts are actively targeted ($f = 0.44$, Fig. 2a) with high precision ($p = 0.95$). The significant variability across datasets is accounted for by a non-negligible basal membrane contact rate (with $1 - s$ in the range of 25–30%). Therefore, high precision combined with basal connectivity are required to account for the reproducibility and variability of membrane contacts across datasets (Fig. 2a, Supplementary Results, Extended Data Fig. 3).

One might ask how useful the $\mathbb{M}^4$ reference is in predicting conserved membrane contacts. Our model predicts that around 99% of the $\mathbb{M}^4$ contacts and 68% of the $\mathbb{M}^4$ contacts together constitute the vast majority (around 98% or more) of the core neuronal membrane adjacency matrix of the *C. elegans* nerve ring (Methods). Furthermore,

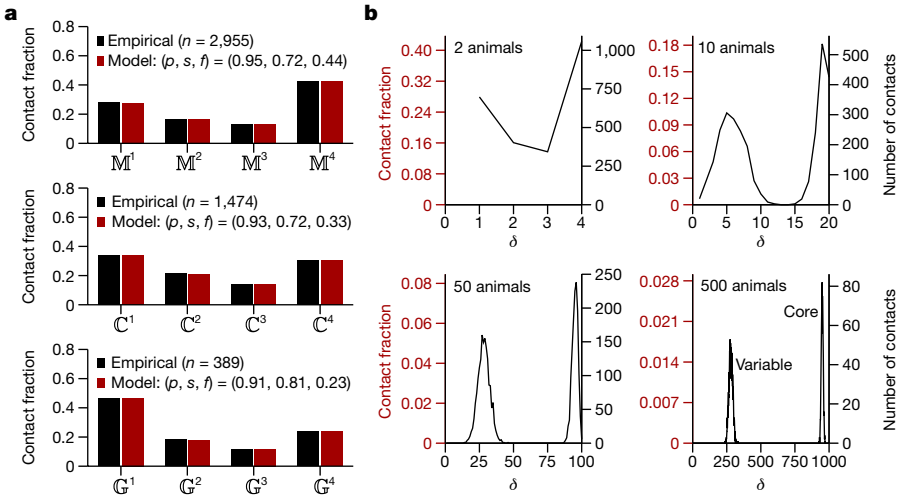


Fig. 2 | The nerve ring is composed of a core circuit embedded in a variable background. a, Empirical data and model fits for the reproducibility, across δ datasets, of membrane, $\mathbb{M}^{\delta}$ (top), synaptic, $\mathbb{C}^{\delta}$ (middle) and gap junction, $\mathbb{G}^{\delta}$ (bottom) contacts. Empirical and model frequency distributions normalized by the total empirical contact count, n (for example, for membrane contacts, $n = \sum_{\delta=1}^{4} \mathbb{M}^{\delta}$). **b**, Surrogate data for 4, 20, 100 and 1,000 datasets (2, 10, 50 and 500 model animals). Four datasets are sufficient to deduce that the distribution is bimodal. Twenty datasets (10 animals) would be sufficient to completely distinguish between the core and variable subcircuits. No contact is expected to be perfectly reproducible across 1,000 datasets (500 animals). Target contacts comprise around 73% of each dataset.

larger-than-average membrane contacts (greater than 1.77 μm^2) comprise more than 80% of $\mathbb{M}^4$ contacts (Extended Data Fig. 2h) and are more reproducible (with higher precision, $p = 0.98$, and a larger fraction of target contacts, $f = 0.77$; Extended Data Fig. 3a, b). We conclude that the $\mathbb{M}^4$ reference offers an excellent candidate set of conserved membrane contacts. Despite being highly reproducible, core membrane contacts are not easily distinguishable from variable ones. Our model predicts that about 50% of membrane contacts are variable across different worms. Using model-generated surrogate datasets (Methods), we estimate that 20 datasets (from 10 worms, with 2 datasets per bilateral reconstruction) would suffice to identify all core membrane contacts in the *C. elegans* nerve ring (Fig. 2b).

To model synaptic and gap-junctional precision, we refitted the model to $\mathbb{C}^{\delta}$ and $\mathbb{G}^{\delta}$ (Methods). To control for synaptic variability due to differences in process placement, we restricted our analysis to $\mathbb{M}^4$ contacts (for a more general treatment, see Extended Data Fig. 4a–c). Even among reproducible membrane contacts, our model predicts that high precision ($p > 0.90$) combined with basal connectivity (with $1 - s$ in the range of 20–30%) is required to account for the reproducibility and variability of synaptic and gap-junctional contacts across datasets (Fig. 2a, Supplementary Results, Extended Data Fig. 4d, e). For the bilateral worm, a synaptic precision of 93% implies that the probability of a core synaptic contact occurring at least once per worm (on the left, right or both sides) is around 99%, and that the chance of this occurring in at least three of four datasets (across two worms) is at least 97%. Conversely, we predict that about 98% of the $\mathbb{C}^4$ and $\mathbb{G}^4$ contacts are good representatives of the core circuit (as well as more than 60% of $\mathbb{C}^3$ and $\mathbb{G}^3$ contacts), lending further confidence to the usefulness of the reference connectome. However, the placement of the most reproducible synapses along the process is not restricted to reproducible membrane contact sites (Extended Data Fig. 1g, h). Thus, location along the process cannot be used to distinguish core from variable synapses. Together, these results demonstrate that each dataset can be divided into a common, precisely targeted core circuit and a variable component, and that, given additional connectomes, it should be possible to distinguish between them (Extended Data Fig. 4d, e).

We next asked what principles of spatial organization support the reproducible, highly specified neuronal placement in the nerve ring. To address this question, we noted that the observed variability of membrane contacts suggests that no one nematode is representative of the population at large, and even core contacts probably vary across individuals (Extended Data Fig. 4h, i). Hypothesizing that conserved membrane contacts form the basis of the neuropil organization, we sought to identify robust features of spatial organization across a population of animals. To this end, we generated population models of membrane contacts by perturbing the four datasets at our disposal. For the reference $\mathbb{M}^4$ contacts (and similarly for the L4 bilateral and adult bilateral membrane contact sets) we used the distributions of membrane contact areas and their associated variability across the datasets to generate stochastic population models of core membrane contacts (Methods, Extended Data Fig. 5). Given our spatial population models, we then grouped together neurites with high spatial affinity, using a multi-level graph-clustering algorithm[21] on each individual in our population model (Methods). We find that five subgroups of neurons, with processes that are spatially ordered along the anterior–posterior axis of the nerve ring, consistently emerge from the data (Fig. 1c, d, Supplementary Data 4). We label these clusters 'anterior', 'lateral', 'sublateral', 'avoidance' and 'taxis' (Supplementary Results). Regionalization of processes in the nerve ring into the anterior circuit (associated with mechanosensation), the posterior, amphid neural circuit (associated with chemosensation and navigation) and lateral and sublateral neurons (associated primarily with head motor control) has previously been highlighted[5]. Our quantitative analysis is consistent with the above description but our focus on the core nanostructure reveals a finer organization of the nerve ring that may not be apparent from the raw volumetric data (Methods, Extended Data Fig. 6).

We asked whether the cluster organization of the nerve ring is indicative of modularization of synaptic pathways[3,13,18,19,22]. We find that most neurons have strong membrane and synaptic contacts within a single cluster, whereas others physically and synaptically contact neurons across multiple clusters (Fig. 3a, b, Extended Data Fig. 7). However, synaptically sparse lateral neurons and a number of neurons that closely link across the lateral and sublateral neighbourhoods suggest that lateral and sublateral clusters may be merged when considering information processing in the nerve ring.

Neurons that synapse across clusters are often characterized by processes that change neighbourhood along their trajectories (Fig. 3c, d, Extended Data Fig. 8k). We identified 33 cell classes with processes that

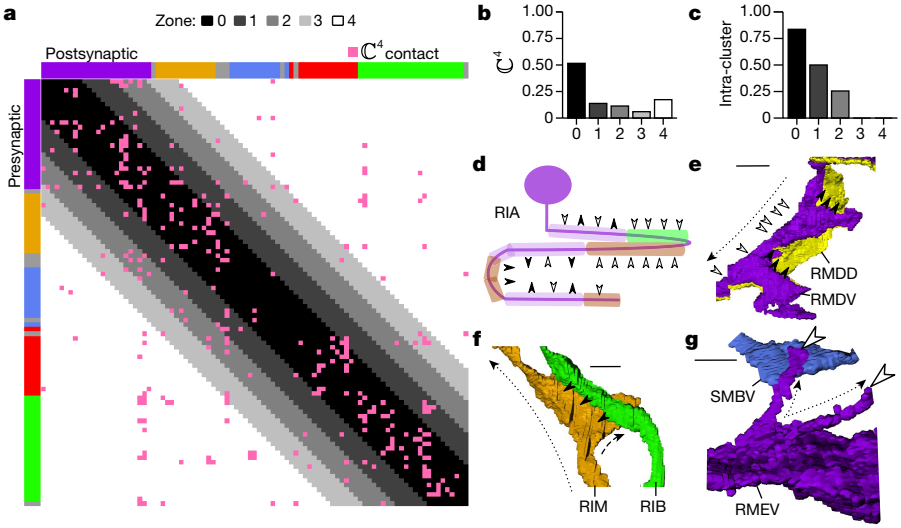


Fig. 3 | Nano-, micro- and mesoscale structure of processes supports local and nonlocal connectivity. a, Matrix of $\mathbb{C}^4$ synaptic contacts (pink). Rows, presynaptic cells; columns, postsynaptic cells. Five zones around the main diagonal delineate growing neighbourhoods around each cell. Zones 0–3, boundaries at ±1–4 standard deviations in the size of the immediate neighbourhood in $\mathbb{M}^4$, respectively (see Methods); zone 4, remaining $\mathbb{C}^4$ contacts not in zones 0–3. **b**, Seventy-five per cent of $\mathbb{C}^4$ synaptic contacts form locally within zones 0–2 ($n = 450$, empirical count of $\mathbb{C}^4$ contacts). **c**, Fraction of intra-cluster $\mathbb{C}^4$ contacts ($n = 238$, total number of these contacts). Many nonlocal (zones 3–4) synaptic contacts occur with neurons that traverse different neighbourhoods, exhibit flattened protrusions and/or exhibit synaptic compartmentalization. **d**, Synaptic compartmentalization: the synaptic polarity in RIA neurons varies with changes in the cluster assignment of neighbouring cells. White and black arrows label synaptic polarity (inputs, white; outputs, black). **e**–**g**, Volumetric rendering of selected processes shows nano- and microscale structures that support localized, reproducible synapses. **e**, The flattened protrusions of RMDV neurons support synapses onto RMDD neurons, diversifying synaptic polarity. **f**, Protrusions of RIM neurons (dashed arrow) extend from the main neurite trajectory (dotted arrow) to support synapses onto RIB neurons. **g**, Spine-like extensions (dotted arrows) from an RMEV cell body support synapses from SMBVL (not shown) and SMBVR neurons. All examples were observed bilaterally in L4 and adult neurons (Extended Data Fig. 9; additional examples in Supplementary Data 4). Colours of rows and columns (**a**) and cells (**d**–**g**) denote cluster assignment except for **e**, RMDD (anterior cluster, yellow). Scale bars, 1 µm (**e**–**g**).

synapse across different regions of the nerve ring (Methods). These cell classes use two principal strategies: synapse compartmentalization (19 out of 33 cell classes; Supplementary Data 4, Fig. 3d, e) and flattened protrusions (23 out of 33 cell classes; Supplementary Data 4, Fig. 3f). We hypothesize that a subset of neurons synaptically link different neighbourhoods of the nerve ring to support brain-wide coordinated activity[23]. Consistent with our hypothesis, these specialized spatial features and the synapses they support are largely conserved across our four datasets. In summary, we find that the nerve ring obeys a consistent set of spatial organization principles across scales—including a macroscopic modular neighbourhood organization that supports the mesoscopic organization along neurites, microscopic precision of membrane contacts and nanoscopic morphological features—that together support conserved synaptic wiring.

A *Caenorhabditis elegans* brain map

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

We posit a parsimonious three-layer architecture with parallel information-processing modules and assign every neuron of the nerve ring into a layer that roughly corresponds to the five neuron clusters (Methods). To achieve overall feed-forward pathways, sensory neurons all occupy the first layer, whereas spatially cross-cutting neurons dominate layer 3 (Methods, Fig. 4). Connectomic features identified from network analysis of the *C. elegans* connectome, such as highly connected 'hub' neurons, high assortativity hubs known as 'rich-club neurons'[24,25], network motifs[13,26] and small-world organization[13], as well as new features such as fan-in and fan-out motifs[13] (characterized by higher in- or out-degrees, respectively; Extended Data Fig. 9a), can now be interpreted within the context of modular, brain-wide computation and information flow (Supplementary Results, Extended Data Fig. 9b–j). In particular, the feed-forward loop, which was previously identified in the *C. elegans* connectome[3,13,26], reappears in our map as the foundational architecture (Fig. 4a) and dominant connectivity motif of the layered synaptic pathways within each module (more than 50% of all $\mathbb{C}^4$ contacts; Fig. 4b; see Extended Data Fig. 10 for additional contacts). This system-wide feed-forward connectivity is reminiscent of the layered connectivity of pyramidal neurons in the mammalian cortex and its biologically inspired artificial analogue, 'residual networks' (ResNets)[27]. Such architectures have been conjectured to enhance the resilience of synaptic pathways and to support flexibility and plasticity[28].

Examination of the *C. elegans* brain map (Fig. 4b) reveals a number of features. Layer 1 separates the modules (with a few notable and functional exceptions; Extended Data Fig. 10). The intra-module, intra-layer connectivity indicates that sensory neurons probably perform limited sensory computation in addition to sensory encoding of environmental cues, and allows the identification of sensory hub (high-degree) neurons (Supplementary Results). Layer 2 largely maintains the modular synaptic information flow. Convergence of sensory neurons onto this sparser layer reveals a fan-in architecture, supporting modular sensory integration (Supplementary Results, Extended Data Fig. 9c, d). Layer 3 contrasts with the above. Inputs are received from all three layers: synapses from layers 1 and 2 comprise the core of each module, whereas

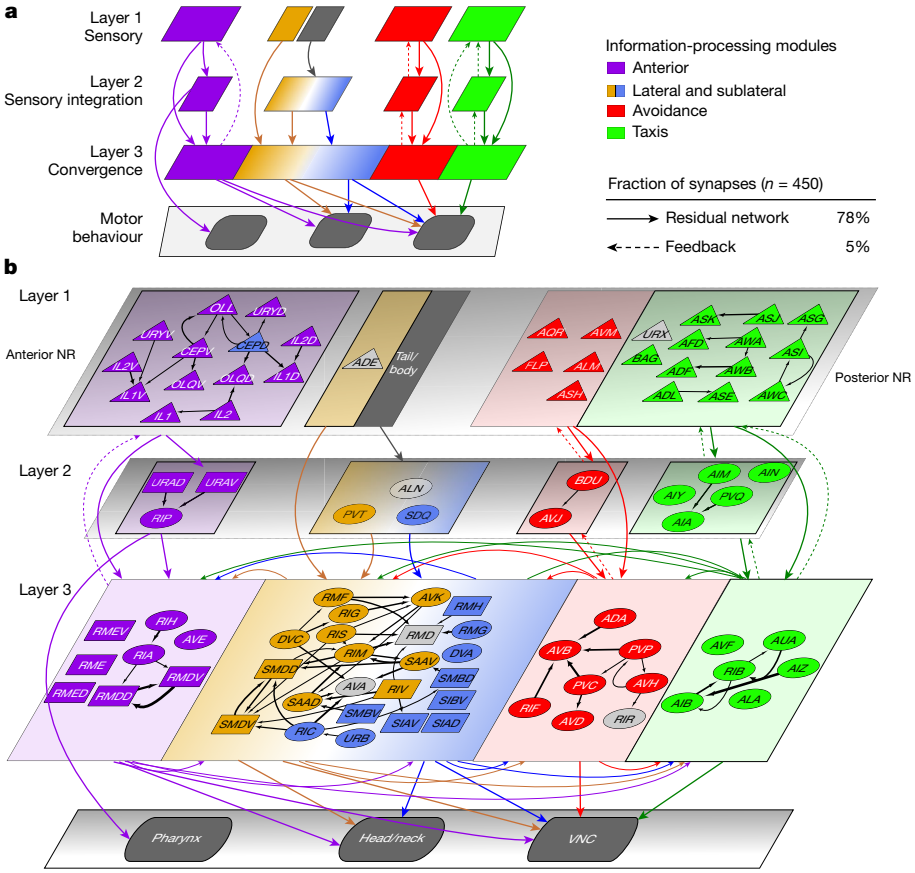


Fig. 4 | The *Caenorhabditis elegans* brain map. a, A three-layer, modular residual network architecture[27] (solid arrows and recurrent connectivity in layer 3) captures 78% of $\mathbb{C}^4$ synaptic contacts in the nerve ring: parallel feed-forward loop motifs converge onto layer 3, supporting functional sensorimotor pathways. Layer-3 interneurons and motor neurons (with $\mathbb{C}^4$ contacts across multiple zones; Fig. 3a) form a distributed circuit across all modules. Dashed arrows indicate intra-module feedback (5% of $\mathbb{C}^4$) (*n*, empirical count of $\mathbb{C}^4$ synaptic contacts). **b**, All 80 bilateral classes of neurons and 11 single neurons (AVL and RID lack $\mathbb{C}^4$ contacts) that innervate the nerve ring (NR), overlaid on the network architecture shown in **a**. Sensory neurons, triangles (layer 1); interneurons, ovals; motor neurons, rectangles. Except for CEPD neurons, the module assignment matches the cluster. CEPD neurons (anterior module, sublateral cluster) share the same looping trajectories as anterior sensory cells, and synapse more extensively with them (Extended Data Fig. 10). The module placement of unclassified cells (grey) was based on process trajectory. Black arrows indicate intra-module synaptic contacts (thickness proportional to aggregate synapse size; that is, the aggregate number of electron micrographs in which synapses were scored).

layer-3 synapses interlink and couple the modules, forming a recurrent, highly distributed circuit, consistent with the dominance of spatially complex neurons in this layer and suggestive of brain-wide coordination roles (Extended Data Fig. 9e–j). Outputs from the nerve ring control the pharynx, head and neck muscles and the motor circuit of the ventral nerve cord (VNC). The taxis and avoidance modules support distinct information pathways (Fig. 4b) despite responding to overlapping sensory cues and both synapsing onto the VNC command interneurons. Unlike other modules, the sublaterals highlight cross-connectivity within the nerve ring, with all but two neuron classes occupying layer 3. Pharyngeal output is mediated by layer-2 anterior neurons, indicating that pharyngeal control is independent of the distributed layer-3 circuit. In contrast, head and neck muscles are controlled by layer-3 anterior, lateral and sublateral neurons and the VNC is controlled by all layer-3 modules, revealing the convergence of sensory pathways and associated modular subcircuits into a small number of highly coordinated motor programs.

Discussion

The *C. elegans* connectome has been available for over 30 years[5] and yet the delineation of functions within its main neuropil is still incomplete. By characterizing the spatial embedding of its connectome, we sought insight into the structures that could support a hierarchical, modular and nested architecture in the *C. elegans* brain. Previous analyses of the *C. elegans* connectome identified a common feed-forward loop motif among triplets of neurons[3,26]. Our brain map recasts this local motif as an architectural motif, reminiscent of layered cortical architectures[28] and their artificial analogue, residual networks[27]. Such a 'connectionist' description of a biological brain provides a promising methodology for identifying parallel and distributed circuits.

Although there are no physical boundaries within the nerve ring, our analysis points to the spatial clustering of neural processes into five neighbourhoods. The parallel pathways in our brain map largely fall into this modular neighbourhood organization, linking spatial and functional organization. The spatial organization may also reflect developmental roles of nerve-ring pioneers[29] and constraints on synaptic and neuromuscular connectivity for motor coordination and control functions. Within the residual-networks template are intra-layer local circuits, comprising neurons that by-and-large lack structural or functional compartmentalization. Thus, consistent with the neuron doctrine, within local subcircuits, neurons represent the basic unit of computation. However, the modular architecture converges within the final layer to achieve brain-wide coordination of behaviour. In this distributed circuit, the nanoconnectome rules: specialized subcellular structures give rise to compartmentalized dynamics and interlink distant regions of the *C. elegans* brain. Similar subcellular structures that perform analogous functions are found in thalamic local interneurons[30]

and reveal a richness of subcellular computation. Thus, brain-wide coordination may be achieved by designated processes that interface between or thread across multiple subcircuits to underpin sensory convergence and sensorimotor transformations. The *C. elegans* brain map and its nested architecture might suggest a much closer analogy between the *C. elegans* neuropil and the coordination between the nano- and macroconnectomes of other invertebrates and even vertebrates[31].

The concept of a reference connectome was key to our brain map and the modelling framework we used to establish this reference can easily be extended to accommodate future connectomes. In vertebrates, the nanoscale organization that underpins individual synapses is variable, supporting individual wiring, plasticity and adaptability[2]. In *C. elegans*, the proportion of conserved synapses was unknown. We found that the connectome consists of a core, conserved circuit that is embedded in a considerably variable background. Although determining the extent of the variable circuit is challenging due to the technical limitations of synaptic scoring and will therefore require multiple further connectomes, it is noteworthy that conserved synapses—like most variable ones—are constrained by the same contactome. Thus, if the core circuit represents the baseline functionality of the animal, the variable component could support redundancy, individuality[32] and plasticity[6].

The dense packing of a large number of neurons in the nerve ring presents a challenge to physically achieving stereotyped connectivity. Our finding of finely orchestrated organization across scales imposes spatial constraints on neurite and synaptic placement, thus restricting the connectivity problem of each neuron to its local neighbourhood. This scalable solution is robust across a large population and naturally generalizes to much larger nervous systems. Viewed differently, the spatial organization reduces the required capacity for cell–cell molecular recognition machinery, while increasing the complexity of mechanisms that produce the morphology and relative positioning of cells in the tissue. How the neighbourhood organization is developmentally orchestrated remains an open question. Previous models of neuropil development have proposed that pioneer neurites guide follower neurons[33]. Although such models could be generalized to identify the pioneers of each neighbourhood[34], the highly reproducible pattern of membrane contacts indicates a more elaborate developmental mechanism. In complementary models, some guidance molecules would coordinate the relative neighbourhood placement, while others would control the placement of neurites[33]. The identification of key guidance molecules in the early stages of nerve-ring formation could help to address such predictions[29,33]. Whatever the developmental mechanisms may be, the brain map of *C. elegans* requires that these mechanisms too are nested and coordinated across scales to guide and support the modular, scalable and flexible neural architecture that produces the mind and behaviour of the nematode *C. elegans*.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-021-03284-x.

1. Hahamy, A., Behrmann, M. & Malach, R. The idiosyncratic brain: distortion of spontaneous connectivity patterns in autism spectrum disorder. *Nat. Neurosci.* **18**, 302–309 (2015).
2. Swanson, L. W. & Lichtman, J. W. From Cajal to connectome and beyond. *Annu. Rev. Neurosci.* **39**, 197–216 (2016).
3. Cook, S. J. et al. Whole-animal connectomes of both *Caenorhabditis elegans* sexes. *Nature* **571**, 63–71 (2019).
4. Ryan, K., Lu, Z. & Meinertzhagen, I. A. The CNS connectome of a tadpole larva of *Ciona intestinalis* (L.) highlights sidedness in the brain of a chordate sibling. *eLife* **5**, e16962 (2016).
5. White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Phil. Trans. R. Soc. B* **314**, 1–340 (1986).
6. Hall, D. H. & Russell, R. L. The posterior nervous system of the nematode *Caenorhabditis elegans*: serial reconstruction of identified neurons and complete pattern of synaptic interactions. *J. Neurosci.* **11**, 1–22 (1991).
7. Jarrell, T. A. et al. The connectome of a decision-making neural network. *Science* **337**, 437–444 (2012).
8. Bumbarger, D. J., Riebesell, M., Rödelsperger, C. & Sommer, R. J. System-wide rewiring underlies behavioral differences in predatory and bacterial-feeding nematodes. *Cell* **152**, 109–119 (2013).
9. Ohyama, T. et al. A multilevel multimodal circuit enhances action selection in *Drosophila*. *Nature* **520**, 633–639 (2015).
10. Zheng, Z. et al. A complete electron microscopy volume of the brain of adult *Drosophila melanogaster*. *Cell* **174**, 730–743 (2018).
11. Kasthuri, N. et al. Saturated reconstruction of a volume of neocortex. *Cell* **162**, 648–661 (2015).
12. Motta, A. et al. Dense connectomic reconstruction in layer 4 of the somatosensory cortex. *Science* **366**, eaay3134 (2019).
13. Varshney, L. R., Chen, B. L., Paniagua, E., Hall, D. H. & Chklovskii, D. B. Structural properties of the *Caenorhabditis elegans* neuronal network. *PLOS Comput. Biol.* **7**, e1001066 (2011).
14. Sulston, J. E., Schierenberg, E., White, J. G. & Thomson, J. N. The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev. Biol.* **100**, 64–119 (1983).
15. Barabási, D. L. & Barabási, A.-L. A genetic model of the connectome. *Neuron* **105**, 435–445 (2020).
16. Albertson, D. G. & Thomson, J. N. The pharynx of *Caenorhabditis elegans*. *Phil. Trans. R. Soc. B* **275**, 299–325 (1976).
17. Cook, S. J. et al. The connectome of the *Caenorhabditis elegans* pharynx. *J. Comp. Neurol.* **528**, 2767–2784 (2020).
18. White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. Factors that determine connectivity in the nervous system of *Caenorhabditis elegans*. *Cold Spring Harb. Symp. Quant. Biol.* **48**, 633–640 (1983).
19. Durbin, R. M. *Studies on the Development and Organisation of the Nervous System of Caenorhabditis elegans*. PhD thesis, Univ. Cambridge (1987).
20. Witvliet, D. et al. Connectomes across development reveal principles of brain maturation in *C. elegans*. Preprint at https://doi.org/10.1101/2020.04.30.066209 (2020).
21. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech.* **2008**, P10008 (2008).
22. Gray, J. M., Hill, J. J. & Bargmann, C. I. A circuit for navigation in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **102**, 3184–3191 (2005).
23. Kato, S. et al. Global brain dynamics embed the motor command sequence of *Caenorhabditis elegans*. *Cell* **163**, 656–669 (2015).
24. Towlson, E. K., Vértes, P. E., Ahnert, S. E., Schafer, W. R. & Bullmore, E. T. The rich club of the *C. elegans* neuronal connectome. *J. Neurosci.* **33**, 6380–6387 (2013).
25. Cohen, N. & Denham, J. E. Whole animal modeling: piecing together nematode locomotion. *Curr. Opin. Syst. Biol.* **13**, 150–160 (2019).
26. Milo, R. et al. Network motifs: simple building blocks of complex networks. *Science* **298**, 824–827 (2002).
27. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proc. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 770–778 (IEEE, 2016).
28. Thomson, A. M. Neocortical layer 6, a review. *Front. Neuroanat.* **4**, 13 (2010).
29. Rapti, G., Li, C., Shan, A., Lu, Y. & Shaham, S. Glia initiate brain assembly through noncanonical Chimaerin-Furin axon guidance in *C. elegans*. *Nat. Neurosci.* **20**, 1350–1360 (2017).
30. Morgan, J. L. & Lichtman, J. W. An individual interneuron participates in many kinds of inhibition and innervates much of the mouse visual thalamus. *Neuron* **106**, 468–481 (2020).
31. Chen, X. et al. Brain-wide organization of neuronal activity and convergent sensorimotor transformations in larval zebrafish. *Neuron* **100**, 876–890.e5 (2018).
32. Stern, S., Kirst, C. & Bargmann, C. I. Neuromodulatory control of long-term behavioral patterns and individuality across development. *Cell* **171**, 1649–1662.e10 (2017).
33. Wang, L. & Marquardt, T. What axons tell each other: axon-axon signaling in nerve and circuit assembly. *Curr. Opin. Neurobiol.* **23**, 974–982 (2013).
34. Moyle, M. W. et al. Structural and developmental principles of neuropil assembly in *C. elegans*. *Nature* https://doi.org/10.1038/s41586-020-03169-5 (2021).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2021

Methods

Anatomical and neuron-class nomenclature

The anatomy of the *C. elegans* nerve ring and associated ganglia, and the delineation of six nerve bundles entering the nerve ring was described by Ware et al.[35]. Early observations—for example, the distinction between papillary and amphid sensory specializations and their postulated mechano- and chemosensory roles—have been validated since. Individual cell classes were identified and named by White et al.[5]. Each neuron name consists of either two or three upper-case letters indicating class and in some cases ends in a number indicating the neuron number within the class (for example, IL1, IL2). Bilaterally symmetric neurons (cell pairs) are labelled by their class name followed by L (left) or R (right). Radially symmetrical neurons (with either four or six members) have a three-letter name followed by D (dorsal), or V (ventral), L (left) or R (right) (for example, SIADL, SIADR, SIAVL, SIAVR and RMDL, RMDR, RMDDL, RMDDR, RMDVL, RMDVR). Unless otherwise noted, we use the term class synonymously with bilateral cell pair for radially symmetric cell classes (for example, SIAV and SIAD are treated as separate classes). In addition, 17 nerve-ring neurons constitute the only members of their class (ALA, ALM, ALN, AQR, AVL, AVM, DVA, DVC, PVT, PVR, RID, RIH, RIR, RIS, RMED, RMEV and SABD). A small number of VNC motor neurons also enter the nerve ring. The names of these VNC motor neurons consist of two upper-case letters indicating muscle innervations (V, ventral; D, dorsal) and class (A–C) and a number indicating the neuron within one class (counted from anterior to posterior). Neurons are designated as sensory neurons, interneurons or motor neurons according to their primary descriptions in WormAtlas (http://www.wormatlas.org). We note, however, that in *C. elegans*, these designations are not exclusive, and that here proprioception is not designated as sensory. Our use of the term neighbourhood to describe processes that run closely together in the nerve ring follows White et al.[5,18]. We use the stronger term 'immediate neighbourhood' to designate neural processes that make physical contact.

Preparation of electron micrographs

The two legacy electron micrograph series used in this study were constructed in the MRC Laboratory of Molecular Biology during the 1970s[5]. Both series are of hermaphrodite nematodes of the wild-type N2 (Bristol) strain. One series is from an L4 nematode and the other series is from an adult (estimated three days from adulthood[3]), referred to as JSH and N2U, respectively. The specimens were fixed in 1% osmium tetroxide in 0.1 M sodium phosphate, pH 7.5 for 1 h at 20 °C before embedding, sectioning and post-staining[5]. This method was previously determined to best bring out cell membranes and synaptic structures at the expense of features within the cytoplasm. The electron micrograph (EM) series are transverse to the longitudinal axis of the body; estimated EM section thickness is 70–90 nm, judged by silver colour[36]. The original 55 cm × 60-cm montaged prints covering the nerve-ring commissure and 30 cm × 40-cm prints covering the posterior lobe of the nerve ring have since been digitized and archived in the Hall Laboratory, and are available at www.wormimage.org.

The two series reconstructed for this study include the synaptically dense nerve-ring neuropil and ventral ganglia regions of the anterior nervous system. The JSH series extends from just anterior of the nerve ring to the excretory pore. The N2U series is substantially longer, extending from just anterior of the nerve ring to the vulva. We only considered the section of the N2U series that physically corresponds to the JSH series. This resulted in 302 sections in the N2U series compared to 410 sections in the JSH series. In N2U, starting at the posterior lobe of the nerve ring, only every other EM section was imaged (N2U sections 183–302). In addition, it is speculated that the JSH images have a slightly smaller section thickness. To correct for this when making comparisons between the L4 and the adult reconstructions, data from this region in N2U were scaled by a factor of 2.

Segmentation of electron micrographs

Electron micrographs were manually segmented using TrakEM2 software[37]. The software provides GUI tools to facilitate the segmentation of cells across a stack of electron micrographs. Within each series, we segmented all neuronal cell bodies and processes that extend into the nerve ring (Supplementary Videos 1, 2). Cell bodies were then removed from our membrane contact analysis, because their large sizes skew the cell contact distribution. We also segmented the portion of the pharynx in the nerve ring, which serves as a visual reference. We did not segment dendrites of sensory neurons, because dendrites have very few synapses and therefore were not of interest for our analysis. We also did not reconstruct the sublateral cells SABVL and SABVR, because their anterior processes leave the ventral nerve cord via the amphid commissure[5]. Measurements of the membrane contact between neurons were taken directly from the TrakEM2 XML data. We estimated each pixel to be around 5 nm^2, on the basis of size measurements of cell bodies that are estimated to be 2–3 μm wide. In all, we segmented 181 and 185 cells that innervate the L4 and adult nerve rings, respectively (the 'complete dataset').

Extracting adjacency data

We developed custom software (parsetrakem2, https://github.com/cabrittin/parsetrakem2) to quantify the pairwise membrane contacts between TrakEM2 segmented processes. In each electron micrograph, TrakEM2 stores each segmented cell as a set of boundary points. To avoid unnecessary pairwise comparisons, for each cell, i, our software defines a search radius that is proportional to the diameter of the segmentation of that cell. Any segmented cell, j, that has a boundary point within the search radius is checked for adjacency to cell i. Given a pair of boundary points of cells i and j, we say that the points are adjacent if the distance between them is less than 10 pixels (around 50 nm). We found that a distance of 10 pixels was sufficiently large to ensure that adjacent boundary points were not missed. Let $B_{ij}^{(s)}$ be the set of boundary points of cell i that are adjacent to cell j in the same electron micrograph s, and $|B_{ij}^{(s)}|$ denote the number of boundary points in this set. For a single EM section, we define the contact length between cells i and j as $\beta_{ij}^{(s)} = \min(|B_{ij}^{(s)}|, |B_{ji}^{(s)}|)$. For two cells, i and j, the total membrane contact area between the two cells is given by the sum over contact lengths across all EM sections, $\alpha_{ij} = \sum_s \beta_{ij}^{(s)}$. To limit the number of cell pairs erroneously identified as adjacent (false positives), we say that cells i and j are immediate neighbours if α_{ij} is greater than 200 boundary points (approximately 90 nm^2).

To check the accuracy of the algorithm, for two electron micrographs, we compared the contacts scored by our software to those obtained from manual scoring of membrane contacts (Supplementary Table 2). For manual scoring of membrane contacts, we used the 'connector' feature in TrakEM2 to generate a connectivity graph of adjacent cells. An electron micrograph with n cells has $n(n-1)/2$ possible cell pairings that were then classified as either adjacent (if the cells touch) or non-adjacent (if the cells do not touch). We assume the manually scored contacts to be the 'ground truth', which we use to define true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN) in our automated classification. Sensitivity, defined as TP/(TP + FN), measures how likely it is that two physically touching cells are classified by our software as adjacent. Specificity, defined as TN/(TN + FP), measures how likely it is that two separate cells are classified by our software as non-adjacent. Aggregating results from the two manually scored EM sections (JSH001 and JSH040), the sensitivity and specificity of our classification algorithm are 0.974 and 1.00, respectively (Supplementary Table 2). In other words, the algorithm will miss around 2.5% of adjacent cell pairs within an electron micrograph and a negligible number (less than 0.05%) of separate cell pairs will be incorrectly classified as adjacent. We assessed the missed adjacent cells in our test set and found that these adjacencies were small (tens of nanometres) and

resulted primarily from poor segmentation (the manual cell segmentation did not extend completely to the cell membrane). Furthermore, all 11 cell pairs that were incorrectly classified as non-adjacent in the two test EM sections were correctly classified as adjacent in subsequent sections. As adjacent cell pairs missed in one section are likely to be correctly classified as adjacent in subsequent sections, and because most of our analyses aggregate adjacencies across sections, any missed adjacencies within a single section are likely to have a negligible effect on our results.

The immediate neighbourhoods of three neurons (AIAR, AIBR and AQR) in the L4 (JSH) series were previously determined by White et al.[18] by analysing every fifth electron micrograph in the reconstructed series[18,19]. As an additional test, we compared the adjacent cells extracted by our algorithm to these previous results for the same set of L4 neurons. Our algorithm was able to find all but one adjacent cell pair in the immediate neighbourhoods reported by White et al.[18] (Supplementary Data 5). Closer inspection revealed that this cell pair does not make physical contact and was thus mis-scored previously as adjacent. Furthermore, we found an additional 69 adjacent cell pairs that were not included in the previously reported immediate neighbourhoods (Supplementary Data 5). Therefore, our volumetric dataset is more extensive than that previously reported.

Annotation of electron micrographs for synaptic connectivity

We used our previously published connectivity data for chemical synapses and gap junctions and refer the reader to Cook et al. (2019)[3] for details on how synapses were annotated. In brief, we used custom software[38] to aid manual annotation of chemical synapses and gap junctions. For chemical synapses, presynaptic cells are identified by the presence of a presynaptic density, whereas postsynaptic cells are identified as the cells directly apposed to the presynaptic density. Most synapses are polyadic—multiple postsynaptic partners are assigned to a single presynaptic cell. Gap junctions are recognized as a straightened or slightly curving region of apposed membranes with increased staining and a uniform small gap. For the purpose of the current study, we restrict the synaptic and gap-junctional dataset to those in our volumetric reconstruction (that is, those scored in EM sections that were segmented for this study). In all, the numbers of synapses and gap junctions scored are larger than in the original connectome[5], with a notable increase in synapses that were scored in only one EM section. Across our $\mathbb{C}^{\delta}$ reference graphs (see 'Generating reference graphs'), Cook et al. (2019)[3] scored an additional 489 synaptic contacts to the White et al. (1986)[5] connectome, of which 249 (49%) synaptic contacts only occur in one EM section. To control for the possibility of false positives in this annotation, a number of test datasets were constructed for validation (see 'Validation against test datasets').

Generating reference graphs

We took advantage of the bilateral symmetry of the *C. elegans* neuropil to effectively double our sample size. We therefore generated four datasets (adult, left and right; and L4, left and right) from the two reconstructed nerve rings. To control for variations in connectivity, we found it useful to map the data to a novel data structure, which we call a reference graph. Reference graphs classify contacts (defined as the aggregate pairwise connections over all EM sections within a dataset) by their degree of reproducibility across datasets. For a sample size of $n = 4$, simply averaging across datasets is not a useful way to build a reference model of the data. Instead, we segregate the contacts into four separate categories on the basis of their reproducibility. To this end, we removed from our analysis a number of neurons that exhibit appreciable differences in synaptic connectivity or process morphology laterally (PLN, PVN, HSN), between the L4 and adult nerve rings (HSN, PVR, SABD), or those that make minimal membrane contact in the nerve ring (in VB, VC and VD classes), leaving 173 cells in 93 cell classes (the 'restricted dataset'; Supplementary Data 3). The restricted dataset excludes HSNR, PLNL, PLNR, PVNL, PVR, SABD, VB01 and VD01 neurons (in both the L4 and the adult) and HSNL, PVNR, VB02 and VC01 (in the adult).

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

For bilateral cell classes, let indices, for example, i and j, each denote one side of a neuropil (left or right) and let $\bar{i}, \bar{j}$, denote the respective contralateral side. For a contact $\{X_i^1, Y_j^1\}$ made between cell X_i in class X to Y_j in class Y in animal 1, δ is defined as the number of contacts among $(\{X_i^1, Y_j^1\}, \{X_i^1, Y_{\bar{j}}^1\}, \{X_i^2, Y_j^2\}, \{X_i^2, Y_{\bar{j}}^2\})$, where the superscripts 1 and 2 label the animals. For intra-class connections, δ is the number of contacts among $(\{X_i^1, X_{\bar{i}}^1\}, \{X_i^1, X_{\bar{i}}^1\}, \{X_i^2, X_{\bar{i}}^2\}, \{X_i^2, X_{\bar{i}}^2\})$, and for single cell classes, for example, DVA connecting to some class Y (or vice versa), δ is counted among $(\{DVA^1, Y_j^1\}, \{DVA^1, Y_{\bar{j}}^1\}, \{DVA^2, Y_j^2\}, \{DVA^2, Y_{\bar{j}}^2\})$ (or vice versa).

Population spatial models

The observed variability in membrane contacts, both bilateral and across the two reconstructions, indicates that it is unlikely that any one contactome is representative of the population at large. We generated a population model of all membrane contacts, by stochastically perturbing the area associated with each membrane contact, such that the overall distribution of mean membrane contact areas is preserved and that the variability in membrane contact areas across datasets is also preserved. To establish the baseline variability across the four datasets, we considered the log-normalized distribution of $\mathbb{M}^4$ membrane contact areas (Extended Data Fig. 5a). For each contact in $\mathbb{M}^4$, we computed the normalized mean membrane contact area (see 'Generating reference graphs') and the standard deviation of membrane contact areas across the four datasets. We observed no correlation between the normalized mean membrane contact area and standard deviation (Extended Data Fig. 5b), indicating that the variability in membrane contact areas does not depend strongly on membrane contact area (similar to immediate neighbourhood sizes in Extended Data Fig. 1a). Therefore, we estimated the variability in the membrane contact area by the mean variability among $\mathbb{M}^4$ membrane contacts (Extended Data Fig. 5c).

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

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

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

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

To perturb membrane contact areas ($x \rightarrow x'$), we add white noise ε in the log domain, that is,

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

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

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

Perturbed populations are denoted $\widetilde{\mathbb{M}^4}$, $\widetilde{\text{L4}}$ and $\widetilde{\text{Adult}}$. For $\widetilde{\mathbb{M}^4}$, we perturb contacts that are conserved across the four datasets (L4 left, L4 right, adult left, adult right). For $\widetilde{\text{L4}}$ and $\widetilde{\text{Adult}}$, we perturb bilaterally conserved contacts in the L4 and adult, respectively. Each population consists of 1,000 perturbed datasets.

Spatial modularity analysis

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

Topological clustering methods such as modularity optimization[21,39] are well suited for characterizing the organization of a complex system from pairwise undirected linked relationships[39,40], as is the case for characterizing spatial organization from membrane contacts between neural processes. In particular, algorithms of this class are appropriate when the organization sought is static[39,40]. Other, random-walk-based algorithms[34,40] assume or impose a flow on the network and are often ill-suited for characterizing spatial (that is, static) organization, as they can introduce bias in the clustering or miss static features in the organization of the system[40]. We applied the Louvain method[21], a multi-level community detection algorithm using the igraph software package[41]. This topological clustering algorithm is a bottom-up heuristic method based on modularity optimization. At first, every vertex is placed in a separate community. Vertices are then iteratively moved between communities in a way that maximizes the local contribution of the vertex to the overall modularity score (the ratio of the number of intra- to inter-community edges). When no vertex movement increases the modularity score, communities are shrunk to a single vertex and the process is repeated.

Cluster assignment and validation

The graph-clustering algorithm (see above) was applied to each individual in each population model. For each population, we generated a cluster frequency matrix that counts the number of times each pair of neurons is clustered together. We then sorted the rows and columns of the frequency matrix so as to minimize the variance along the main diagonal of the matrix (Extended Data Fig. 5i). Sorting was achieved using a hierarchical matrix clustering algorithm[42]. The resulting dendrogram assigns neurons to a cluster. We obtained a set of five largely overlapping clusters for each of the $\widetilde{\mathbb{M}^4}$, $\widetilde{\text{L4}}$ and $\widetilde{\text{Adult}}$ population models (Extended Data Fig. 5i). Cell classes were assigned to their consensus cluster if their cluster assignment agreed across all three population models. Seven neuron pairs (ADE, ALN, AVA, RID, RIR, RMD and URX) were classified differently across the different population models, and were designated 'unclassified' accordingly. To evaluate the robustness of the clusters to empirical variability between the L4 and adult series, we compared clusters obtained from population models of $\widetilde{\mathbb{M}^4}$, $\widetilde{\text{L4}}$ and $\widetilde{\text{Adult}}$ (see 'Population spatial models'; Fig. 1c, d, Extended Data Figs. 5i, 6b).

We performed four sets of validation experiments using our population models to confirm the robustness of our neuron clusters. (1) As discussed above, we compared cluster assignments across $\widetilde{\mathbb{M}^4}$, $\widetilde{\text{L4}}$ and $\widetilde{\text{Adult}}$ (Fig. 1c, d, Extended Data Fig. 5i). (2) We generated cluster assignments for $\widetilde{\mathbb{M}^4}$ populations that were perturbed with different noise amplitudes ($\sigma = 0, 0.12, 0.23, 0.45, 0.9$; see 'Population spatial models'; Extended Data Fig. 5j). (3) We generated a new $\widetilde{\mathbb{M}^4}$ ($\sigma = 0.23$) population from the membrane contacts used in a different analysis of our volumetric reconstructions[34], which consists of the same scoring of membrane contacts as this study, but restricted to a volume occupying roughly the anterior 60% of the nerve-ring neuropil (Extended Data Fig. 5k). (4) We generated a new $\widetilde{\mathbb{M}^4}$ ($\sigma = 0.23$) population that also includes the smallest 35% membrane contact areas (Extended Data Fig. 5l; recall that the smallest contacts were removed from our analysis, see 'Generating reference graphs' and Extended Data Fig. 2). All of our validation experiments resulted in largely similar cluster assignments (Supplementary Data 4). A handful of neuron classes were assigned to different clusters in different population models, but the gross structure of the five main neuron clusters, as defined by the cluster assignments of a large majority of the neurons, was consistent across the populations. Thus, our cluster assignments are robust across model population datasets, the L4 and adult, a wide range of noise amplitudes—well above the observed inter-animal variability—and different spatial domains.

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

Mesoscale analysis of synaptic connectivity

Given the modular organization of the neuropil, with most neurons spatially clustering within local neighbourhoods and others spatially interconnecting different neighbourhoods, we wanted to determine whether synapses form local subcircuits, or to what extent synaptic circuits also span different neighbourhoods of the nerve ring. To assess the spatial organization of synaptic circuits, we considered the distribution of conserved $\mathbb{C}^4$ synaptic contacts (Fig. 3a). Using the $\mathbb{M}^4$

reference, we calculated the mean ($\overline{N}$ = 17 cells) and standard deviation (ΔN = 8 cells) of immediate neighbourhood sizes (see 'Anatomical and neuron-class nomenclature'). We order the cells as in Fig. 1b so as to maximize the amount of physical ($\mathbb{M}^4$) contact along the diagonal of the matrix.

We define five zones based on the size distribution of immediate neighbourhoods. Each zone refers to regions between two diagonals above and below the main diagonal of the $\mathbb{M}^4$ matrix. The inner zone (labelled 0) consists of neighbourhoods of size $\overline{N}$ for each cell around the main diagonal. The next zone (1) extends from the edges of the inner zone to diagonals ΔN further away from the main diagonal, and zones 2–3 similarly extend by ΔN. The outermost zone (4) extends from the previous zone (3) to encompass cells in the remainder of the matrix. (Formally, zones are defined by their inner and outer diagonals, with the inner diagonal defined by $\overline{N}/2 + (k-1)\Delta N$ from the main diagonal for zones 1 to 4 and the outer diagonals defined by $\overline{N}/2 + k\Delta N$ from the main diagonal for zones 0 to 3.) We counted the number of $\mathbb{C}^4$ contacts in each zone (Fig. 3b). Finally, we counted $\mathbb{C}^4$ contacts between pre- and postsynaptic neurons that have been assigned to the same cluster (Fig. 3c). For this purpose, synapses between two unclassified neurons are not considered intra-cluster.

Contact localization analysis

Although some membrane contacts appear to be reproducible (our $\mathbb{M}^4$ reference), contacts are aggregate measures (along the entire process). To assess the reproducibility in the location of individual instances of membrane adjacencies along a neurite, we assigned each electron micrograph in each process a discrete coordinate, $\hat{z}$, from the anterior ($\hat{z} = 0$) to the posterior ($\hat{z} = 1$) of the process. This allows us to compare relative locations of a contact across the four datasets (L4 left and right; adult left and right). Different discretizations of $\hat{z}$ (0.7 µm, 1.4 µm, 3.6 µm) define different resolutions for the reproducibility of contacts along the process. For each $\mathbb{M}^4$ contact, we define the spatial reproducibility count as the number of datasets in which the contact was observed at a given position, $\hat{z}$. We further define the maximum spatial reproducibility count, $\max(\delta)_{\hat{z}}$, as the highest reproducibility count across all locations, $\hat{z}$, per cell pair (that is, given an $\mathbb{M}^4$ contact exists between two immediate neighbours, the highest reproducibility count of instances of membrane adjacencies between the two cells). To assess synaptic localization, we similarly measured the spatial reproducibility counts (and their maxima) for all $\mathbb{C}^4$ contacts. See also Supplementary Results.

Synapse compartmentalization and subcellular structures

Identification of synaptic compartmentalization and subcellular structures was performed by visually inspecting the volumetric reconstruction of the processes of 173 neurons in each of the adult and the L4 nerve rings (346 cells in total). To visualize synapses, we imported synapse locations[3] (http://wormwiring.org) into the reconstructed TrakEM2 datasets. To facilitate visual identification, we coloured synapses based on whether the cell is presynaptic or postsynaptic and whether the synapses occur between cells of the same cluster or not. For each cell, we required synapse compartmentalization and/or subcellular structures to be bilaterally conserved in both the L4 and the adult (Supplementary Data 4). The one exception are the RMF cells, in which there is clear branching in the L4 (both left and right) that is not observed in the adult. However, because we are limited to two samples, we cannot determine whether these differences are developmental or individual in nature, or whether they are due to reconstruction error[5].

We identified two types of synaptic compartmentalization: compartmentalization of synaptic inputs and outputs; and compartmentalization of synapses with different clusters. To identify compartmentalization of inputs (outputs), we required neural segments to have at least three synaptic inputs (outputs) that are spatially distinct from segments of the neurite with synaptic outputs (inputs) or neural segments with mixed synaptic inputs and outputs (Extended Data Fig. 8).

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

Mapping neighbourhood changes of neurites

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

Brain map construction

We posited a three-layer architecture as the minimum number of layers needed to capture the organizing principles of the connectome. Classifications of neurons as sensory neurons, interneurons or motor neurons followed WormAtlas (http://www.wormatlas.org). All sensory neurons were assigned to the first layer. SDQ, BDU and ALN have been postulated to have sensory functions[43,44] but were classified as interneurons as they are not ciliated and physiological evidence for sensory function is lacking. Reclassifying them as sensory neurons would not alter the high-level connectivity of the brain map. With the exception of AIY and AIA, all neurons that make at least one $\mathbb{C}^4$ inter-cluster contact were placed in layer 3, with the remainder of neurons assigned to layer 2. Placing AIY and AIA in layer 2 is consistent with functional and ablation studies suggesting that these cells are first-layer amphid interneurons[22,45]. Furthermore, AIY and AIA each only make one inter-cluster $\mathbb{C}^4$ synaptic contact (Extended Data Fig. 10). AIY synapses onto the multi-compartment cell RIA, which traverses multiple neighbourhoods (Extended Data Fig. 8a). AIA synapses onto the RIF neurite which lies at the interface of taxis and avoidance cells in the nerve-ring posterior lobe. We confirmed that our map is robust to small changes in which neurons with relatively few inter-cluster synaptic contacts between layers are shifted to layer 2. However, the configuration adopted here optimizes the feed-forward directionality of the synaptic circuit (from the sensory layer to layer 3).

Our information-processing modules roughly correspond to the five spatially identified clusters. The sublateral and lateral clusters were merged into a single module. With one exception (CEPD), cell classes in the same cluster are placed within the same module. Because CEPD neurons follow the same looping neurite trajectories as other papillary sensory neurons, CEPD cells, which are assigned to the sublateral cluster, are more sensibly placed in the anterior module. Unclassified cells are difficult to cluster because they exhibit high spatial affinity with cells from different clusters. To place the seven unclassified cell

classes on the brain map, we relied on the relative placement of their process trajectories among the clusters. We identified representative cells from each cluster to serve as fiducial points for process placement (anterior, RIH; lateral, AVK and RIV; sublateral, SIAD; avoidance, AVB; taxis, ASJ). Each unclassified cell was then added to the module of the representative cells whose neurite most closely aligned with the neurite of the unclassified cell.

Statistical connectivity models

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

We therefore constructed a minimal three-parameter model that combines two stochastic processes—precision and specificity. Precise targeting of contacts and active avoidance of others both require us to distinguish between the set of candidate target contacts and the remainder (non-targets). Accordingly, we define a fraction of target contacts (f), the probability to form a target contact (precision, p) and the probability to avoid an off-target contact (specificity, s).

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

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

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 the absence of empirical data for estimating the physically accessible subset of contacts, we restrict our consideration to $\delta \in \{1,2,3,4\}$, or in general, for K datasets, using Bayes' theorem:

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

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

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

Empirical data for fits and bias control. For membrane, synaptic and gap-junction contacts, we found no evidence of higher reproducibility of edges between the left sides of the L4 and adult datasets, or between the right sides of the L4 and adult, as compared to L4 left and adult right, or vice versa (Supplementary Table 3). However, for membrane contacts, development leads to an overall increase in the number of edges between the L4 and adult (Supplementary Table 1). We considered all neuron pairs in our complete dataset (3,203 edges with membrane contact areas ≥ 35 percentile, Extended Data Fig. 2g) as well as the restricted dataset (see above, 2,955 edges). Models of the complete and restricted datasets yielded quantitatively similar results (Extended Data Fig. 3e).

Our restricted dataset consists of 173 neurons. In the absence of spatial constraints, all-to-all connectivity would, in principle, allow for up to $173 \times 172/2 = 14{,}878$ edges. Conversely, using the model fit, the sum $\sum_{\delta=0}^{K} \mathbb{M}^\delta$ could provide an estimate for the size of the pool of physically accessible membrane contacts in the nerve ring. The latter estimate (≈3,500 edges for the restricted set of contacts) is about 23% of the all-to-all number. This model estimate points to the strong role that spatial constraints have in the actual circuit.

The set of possible synaptic and gap-junction contacts is restricted to existing physical membrane contacts. Unless otherwise noted, all fits were performed on $\mathbb{C}^\delta$ and $\mathbb{G}^\delta$ that were restricted to edges from the set of $\mathbb{M}^4$ membrane contacts. To control for possible bias due to the subselection of $\mathbb{M}^4$ contacts, validation plots were generated by considering $\mathbb{M}^j$ contacts and scaling the counts $\mathbb{C}^\delta|_{\mathbb{M}^j} \to \frac{\mathbb{C}|_{\mathbb{M}^j}}{\mathbb{C}|_{\mathbb{M}^4}} \mathbb{C}^\delta|_{\mathbb{M}^4}$ for $j = 1,\ldots,\delta$, where $\mathbb{C} = \sum_{\delta=0}^{j} \mathbb{C}^\delta$ and '$|_{\mathbb{M}^j}$' denotes synaptic contacts occurring on the domain of membrane contacts $\mathbb{M}^j$ (scaling was performed in the same way for gap-junction contacts, $\mathbb{G}$; Extended Data Fig. 4a–c). For chemical synapses, we find good agreement with $\mathbb{C}^3$ when scaled by $\mathbb{M}^3$ and $\mathbb{C}^2$ when scaled by either $\mathbb{M}^3$ or $\mathbb{M}^2$. Rescaling systematically underestimated $\mathbb{C}^1$ across all datasets, possibly owing to a subset of small synapses not accounted for by the model. Indeed, consistent with previous work[6], we find that both $\mathbb{C}^1$ synapses and $\mathbb{G}^1$ gap junctions are significantly smaller (Extended Data Fig. 4f, g) and occur at smaller membrane contacts (Extended Data Fig. 2i; see also 'Validation against test datasets').

To estimate the fraction of target edges (for $\delta = 1\ldots4$), we used equation (5). For example, the probability of observing $\delta = 4$ target membrane contacts is given by $fp^4 = 0.44 \times 0.954 = 36\%$, whereas the probability of finding 4 variable membrane contacts, $(1-f)(1-s)^4$, is negligible. Thus, the estimated fraction of $\mathbb{M}^4$ membrane contacts in the core circuit is estimated as $fp^4/[fp^4 + (1-f)(1-s)^4] > 99\%$, whereas the estimated fraction of $\mathbb{M}^3$ contacts in the core circuit is only $4fp^3(1-p)/[4fp^3(1-p) + 4(1-f)s(1-s)^3] = 68\%$. Finally, we estimate that in two animals (four datasets) one would expect $p^4 + 4p^3(1-p)$ of core edges to occur in at least three datasets (corresponding to around 99% of core membrane contacts and around 97% of core synaptic edges). In addition, we separately fit the model to intra-cluster and inter-cluster edges. For each set of membrane contacts, $\mathbb{M}^\delta$, we separated the contacts that occur between neurons with the same cluster identity (intra-cluster) and contacts between neurons with different cluster identities (inter-cluster). We then separately fit the model to the sets of intra- and inter-cluster edges, corresponding to membrane contacts, synapses and gap junctions (Extended Data Fig. 3c, d).

Simulation and generation of surrogate data. To construct each surrogate dataset, k, we set the size of the dataset, n, to the number of edges (for example, 2,955 for membrane contacts) and created an ordered list $\mathcal{L}(k)$ of edges. We generated a binary target list (the first round(fn) elements in the list), $\mathcal{L}_T(k)$, and a binary non-target list $\mathcal{L}_{NT}(k)$. Among target edges, a contact occurs with probability p, and among off-target edges, a contact occurs with probability $1-s$. We then aggregate the number of contacts across K surrogate datasets, $\delta_i = \sum_{k=1}^{K} \mathcal{L}_i(k)$, where δ_i corresponds to the number of datasets in which edge i forms a contact. The list of δ_i then forms a surrogate dataset for the reproducibility of contacts, for example, $\mathbb{M}$.

Validation against test datasets. As additional connectomes are generated and technologies change, we expect slight differences in the scoring of different datasets generated from different sets of EM

sections[3,6,7,17,18,38]. These could arise from a slightly different demarcation of the volume being scored on an electron micrograph, different sectioning of a specimen (or sections scored on an EM series) and different scoring criteria. In the absence of functional (molecular or physiological) data, it is difficult to avoid some false positives (scored synapses that are not fully developed and functional) and false negatives (missed synapses). Often, smaller synapses are harder to score accurately. Furthermore, most *C. elegans* synapses are polyadic and present particular challenges, especially when one of the targets occurs with a considerably smaller membrane contact area. Methods and validation of synaptic scoring for the dataset used here have been described by Cook et al.[3]. Here, we address complementary aspects, relating to the reproducibility of scores and implications for our model of core and variable circuits (Extended Data Fig. 3e–i).

Cook et al.[3] (the dataset used here) scored a greater number of small synapses than White et al.[5] (Extended Data Fig. 4f). Furthermore, while this paper was under submission, additional connectomes have been reported for eight hermaphrodite *C. elegans* nerve rings, including two adults[20]. We therefore validated our main results on synaptic reproducibility against the connectomes of White et al.[5] and the two adults in Witvliet et al.[20] (hereafter, 'test datasets', denoted with the subscript test). As the volumetric reconstruction and hence membrane contact analysis is only available for our study, we used the M^4 edges identified here as a common basis for comparison and validation.

Size dependence of synaptic reproducibility has previously been noted[3,6,7,38]. Consistently with these earlier results, Extended Data Fig. 4f shows that $\mathbb{C}^4$ synapses, and less so $\mathbb{C}^3$ synapses, have a considerably higher fraction of contacts associated with higher EM section counts: 87% of $\mathbb{C}^4$ and 37% of $\mathbb{C}^3$ contacts are observed in five or more EM sections, as compared to 13% and 21% in $\mathbb{C}^1$ and $\mathbb{C}^2$, respectively. That said, a comparison with the White et al. test dataset[5] shows that the additionally scored synaptic contacts are evenly distributed across $\mathbb{C}^1$–$\mathbb{C}^4$ (Extended Data Fig. 4f). To check whether different scoring criteria leading to different counts of small synapses affect our conclusions, we refitted our model to a more restricted synaptic reference graph in which all one-EM-section synapses were excluded. Although this substantially suppresses $\mathbb{C}^1$ counts (hence affecting the relative core and variable fractions), the effect on our model precision and specificity is minor (Extended Data Fig. 3f). The scoring of polyadic synapses is also potentially challenging, if synapses are formed with only a subset of co-localized postsynaptic neighbours. To check whether excessive scoring of polyadic synapses might affect our results, we constructed a synaptic reference graph in which for every polyadic presynaptic site, we excluded any postsynaptic partner that is in $\mathbb{C}^1$. Refitting our model to this restricted synaptic reference graph, we again find similar precision and specificity.

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

To validate the scoring of postulated variable synapses, we use our model parameters and equation (4) to estimate what number of synaptic edges in our reference graph would be statistically expected to be absent from two independent animals,

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

This expression yields an estimated number of around 330 synaptic edges. Empirically, we find that 374 of our synapses were not scored in the Witvliet et al. test dataset[20] (Extended Data Fig. 3i; difference not statistically significant under binomial counting statistics). This result adds confidence to the scoring of variable and, in particular, small synapses in our dataset. To examine the consistency of postulated conserved synapses, we estimated the number of $\mathbb{C}^4$ synaptic edges scored by Cook et al.[3] that would also be expected to be found in two independent animals (that is, in a new set of $\mathbb{C}^4$). Of our 450 $\mathbb{C}^4$ synaptic edges, we expect a test dataset to include $\frac{n_{\text{test}}}{n}450\sim380$ as $\mathbb{C}^4$ (also equivalent to $n_{\text{test}}fp^4/\Pr[\mathbb{C}^\delta>0]$). Empirically, 389 of our $\mathbb{C}^4$ were scored in the Witvliet et al.[20] set of $\mathbb{C}^4$ contacts, consistent with our model predictions.

Statistics and reproducibility

Membrane contact datasets are derived from the reconstructions of the nerve rings of two animals at different developmental stages. Our restricted dataset (see 'Generating reference graphs') consists of 93 cell classes per neuropil, of which 80 correspond to pairs of bilateral homologous cells. Extended Data Figures 1 and 2 established that bilateral homologous cells are sufficiently similar. Accordingly, for the purposes of generating reference graphs and for our core-variable and population models, we assume the two sides of each nerve ring may be treated as independent, yielding four independent datasets (L4 left, L4 right, adult left and adult right). We apply this independence assumption to all 93 cell classes. As further measures of reproducibility, we validated our core-variable synaptic and gap-junction contact models against data scored by different experts on the same EM series[5] and on different electron microscopy datasets[20] (in both cases, limited to our $\mathbb{M}^4$ contacts). Our models yielded qualitatively similar results for the different scorings and datasets (Extended Data Fig. 3h, i). Spatial population model data were drawn from distributions that matched the empirical distributions of $\mathbb{M}^4$ membrane contact areas across the four datasets. No statistical methods were used to predetermine sample size.

Reporting summary

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

Data availability

The volumetric datasets generated during the current study, associated connectivity databases and associated analysis are available at https://doi.org/10.5281/zenodo.4383277 and http://wormwiring.org/. The raw data for volumetric reconstructions for Figs. 1, 3, Extended Data Fig. 8 and all Supplementary Videos are available at https://doi.org/10.5281/zenodo.4383277. Extracted adjacency data are available in Supplementary Data 1. The reference datasets are available in Supplementary Data 3. The Cytoscape files used to generate the brain map (Fig. 4, Extended Data Fig. 10) and network motifs (Extended Data Fig. 10) are available at https://doi.org/10.5281/zenodo.4383277. The collection of *C. elegans* nervous system electron micrographs is also available at https://www.wormatlas.org/ and https://wormimage.org. Source data are provided with this paper.

Code availability

The software packages parsetrakem2 (extracting adjacency data) and elegansbrainmap (analysis and visualization software) are available at

https://github.com/cabrittin/parsetrakem2 and https://github.com/cabrittin/elegansbrainmap, respectively.

35. Ware, R. W., Clark, D., Crossland, K. & Russell, R. L. The nerve ring of the nematode *Caenorhabditis elegans*: sensory input and motor output. *J. Comp. Neurol.* **162**, 71–110 (1975).
36. Peachey, L. D. Thin sections. I. A study of section thickness and physical distortion produced during microtomy. *J. Biophys. Biochem. Cytol.* **4**, 233–242 (1958).
37. Cardona, A. et al. TrakEM2 software for neural circuit reconstruction. *PLoS One* **7**, e38011 (2012).
38. Xu, M. et al. Computer assisted assembly of connectomes from electron micrographs: application to *Caenorhabditis elegans*. *PLoS One* **8**, e54050 (2013).
39. Newman, M. E. & Girvan, M. Finding and evaluating community structure in networks. *Phys. Rev. E* **69**, 026113 (2004).
40. Rosvall, M. & Bergstrom, C. T. Maps of random walks on complex networks reveal community structure. *Proc. Natl Acad. Sci. USA* **105**, 1118–1123 (2008).
41. Csardi, G. C. & Nepusz, T. The igraph software package for complex network research. *InterJournal Complex Systems* 1695 (2006).
42. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
43. Chang, A. J., Chronis, N., Karow, D. S., Marletta, M. A. & Bargmann, C. I. A distributed chemosensory circuit for oxygen preference in *C. elegans*. *PLoS Biol.* **4**, e274 (2006).
44. Zimmer, M. et al. Neurons detect increases and decreases in oxygen levels using distinct guanylate cyclases. *Neuron* **61**, 865–879 (2009).
45. Tomioka, M. et al. The insulin/PI 3-kinase pathway regulates salt chemotaxis learning in *Caenorhabditis elegans*. *Neuron* **51**, 613–625 (2006).
46. Hendricks, M., Ha, H., Maffey, N. & Zhang, Y. Compartmentalized calcium dynamics in a *C. elegans* interneuron encode head movement. *Nature* **487**, 99–103 (2012).
47. Perkins, L. A., Hedgecock, E. M., Thomson, J. N. & Culotti, J. G. Mutant sensory cilia in the nematode *Caenorhabditis elegans*. *Dev. Biol.* **117**, 456–487 (1986).
48. Sawin, E. R., Ranganathan, R. & Horvitz, H. R. *C. elegans* locomotory rate is modulated by the environment through a dopaminergic pathway and by experience through a serotonergic pathway. *Neuron* **26**, 619–631 (2000).
49. Kang, L., Gao, J., Schafer, W. R., Xie, Z. & Xu, X. Z. *C. elegans* TRP family protein TRP-4 is a pore-forming subunit of a native mechanotransduction channel. *Neuron* **67**, 381–391 (2010).
50. Chalfie, M. & Sulston, J. Developmental genetics of the mechanosensory neurons of *Caenorhabditis elegans*. *Dev. Biol.* **82**, 358–370 (1981).
51. Suzuki, H. et al. In vivo imaging of *C. elegans* mechanosensory neurons demonstrates a specific role for the MEC-4 channel in the process of gentle touch sensation. *Neuron* **39**, 1005–1017 (2003).
52. Chalfie, M. et al. The neural circuit for touch sensitivity in *Caenorhabditis elegans*. *J. Neurosci.* **5**, 956–964 (1985).
53. Li, C. et al. The FMRFamide-related neuropeptide FLP-20 is required in the mechanosensory neurons during memory for massed training in *C. elegans*. *Learn. Mem.* **20**, 103–108 (2013).
54. Hukema, R. K., Rademakers, S., Dekkers, M. P. J., Burghoorn, J. & Jansen, G. Antagonistic sensory cues generate gustatory plasticity in *Caenorhabditis elegans*. *EMBO J.* **25**, 312–322 (2006).

Acknowledgements We thank J. Hodgkin and J. White for their help in donating archival transmission electron microscopy material from the MRC Laboratory of Molecular Biology to the Hall laboratory for curation. T. Ilett, F. Salfelder and S. L. Braunstein provided useful discussion. We thank M. Zhen for making their synaptic and gap junction data available (https://nemanode.org/). This work was supported by NIH grant NIMH F32MH115438 (S.J.C.), NIHD grant P30HD071593 (S.W.E.), NIMH grant R01MH112689 (S.W.E.), the G. Harold and Leila Y. Mathers Charitable Foundation (S.W.E.), NIH OD 010943 (D.H.H.) and EPSRC EP/J004057/1 (N.C.). C.A.B. was supported by the Leeds International Research Scholarship.

Author contributions C.A.B., S.J.C. and S.W.E. conceived the volumetric reconstruction. C.A.B. and S.J.C. segmented the electron micrographs. D.H.H. curated the data. C.A.B. built the software for quantifying membrane contact areas. C.A.B. and N.C. analysed and interpreted the data and wrote the manuscript. S.J.C., D.H.H. and S.W.E. provided critical revisions.

Competing interests The authors declare no competing interests.

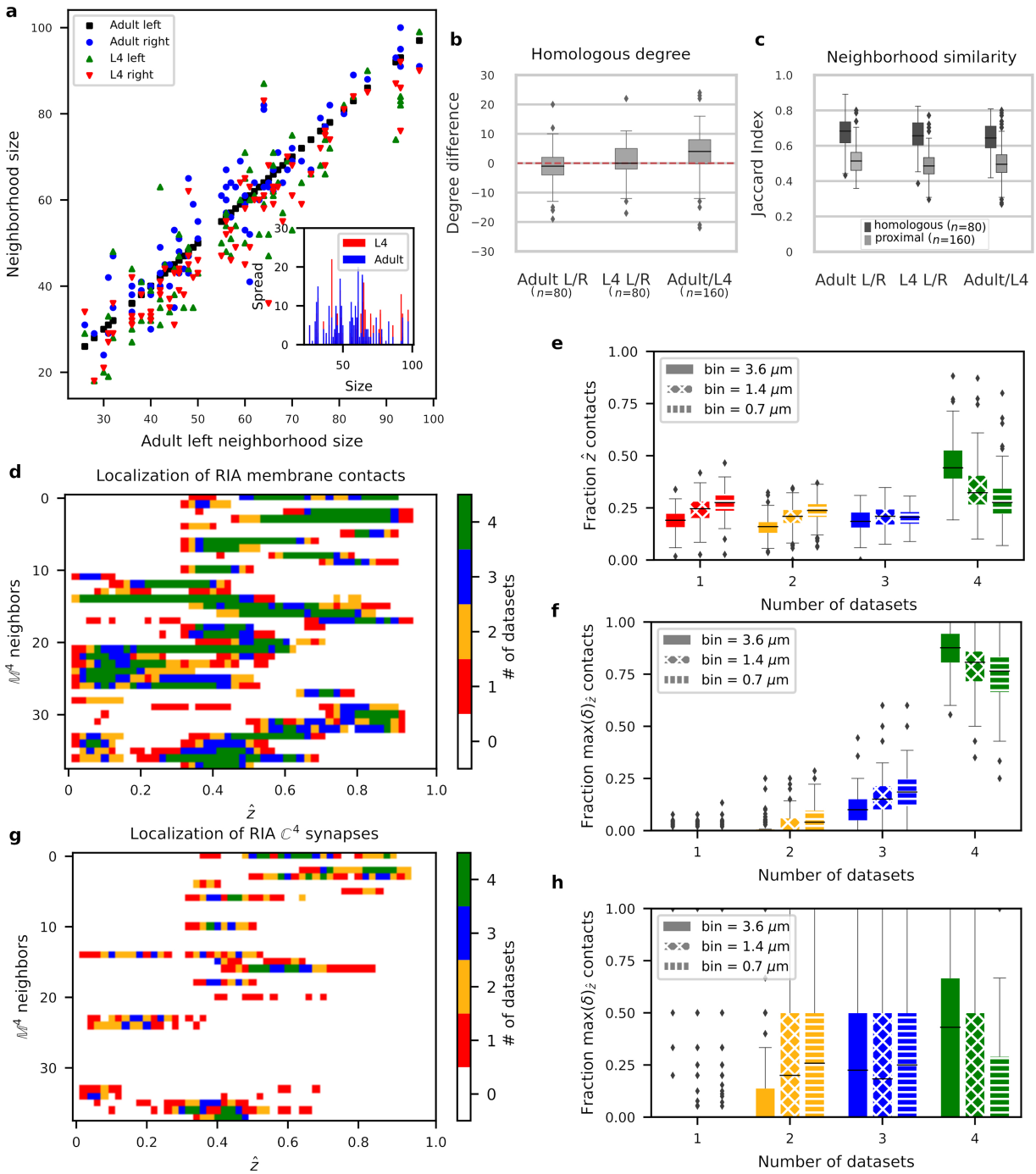
Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-021-03284-x.

Correspondence and requests for materials should be addressed to N.C.

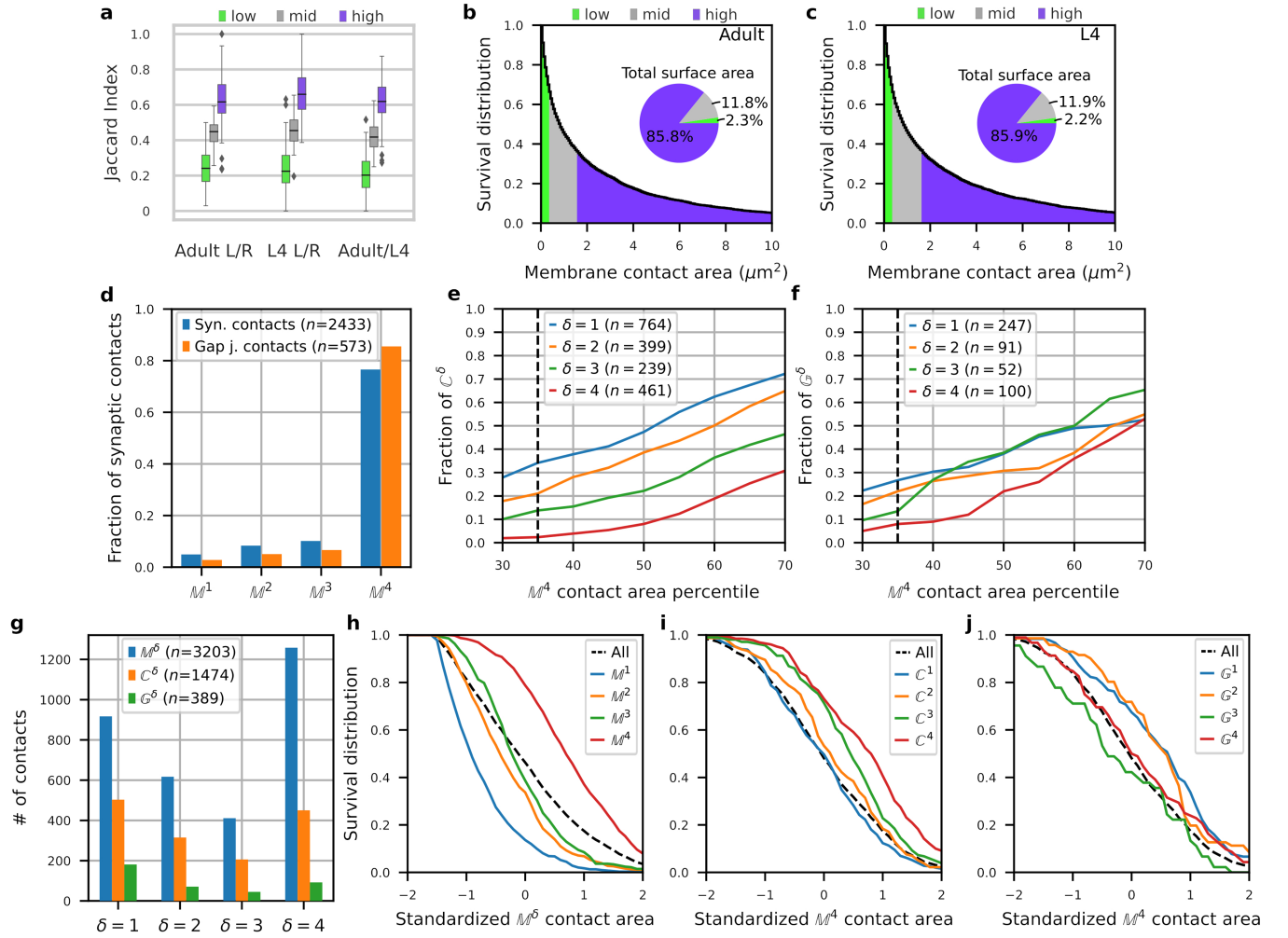
Peer review information *Nature* thanks the anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at http://www.nature.com/reprints.

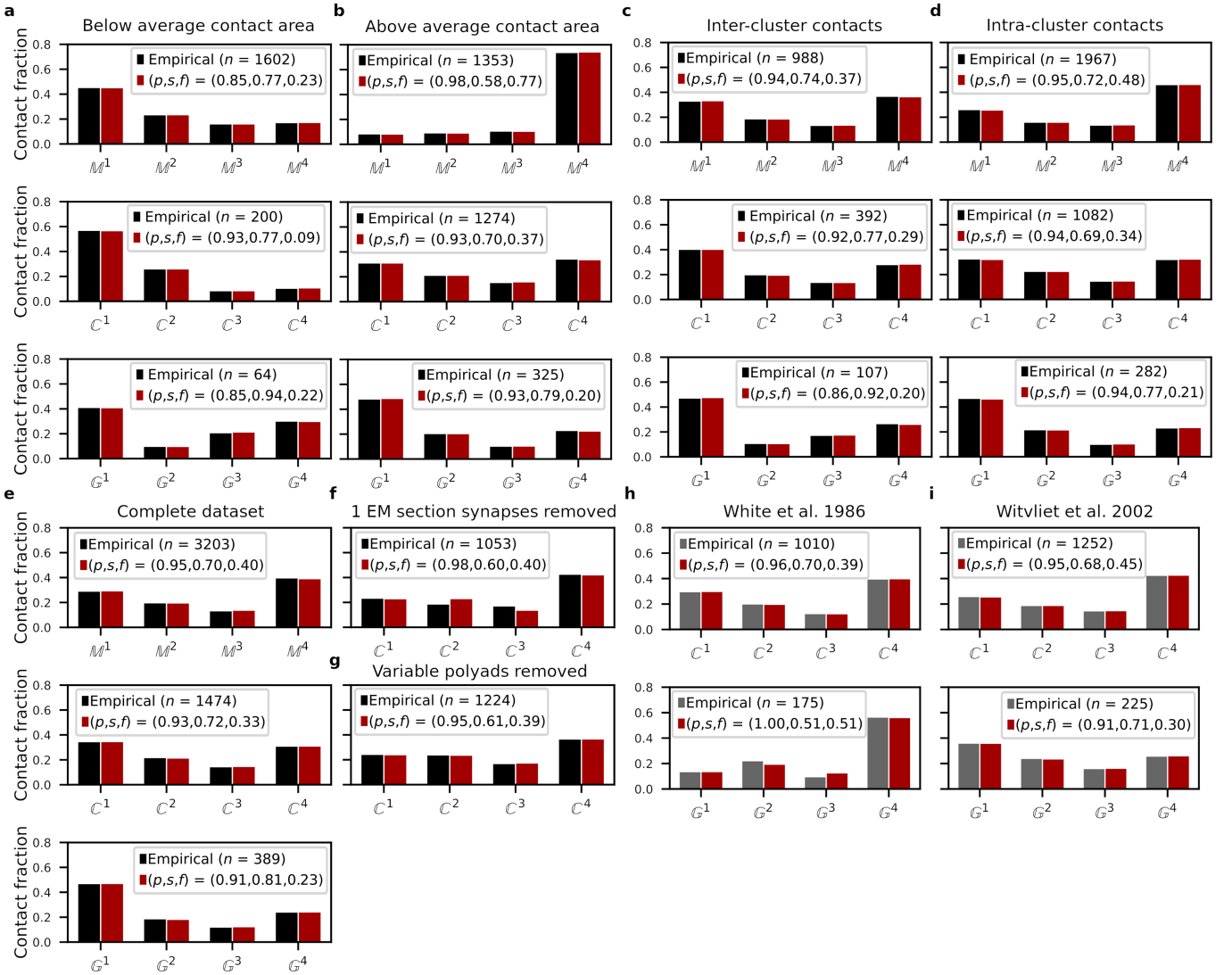


Extended Data Fig. 1 | See next page for caption.

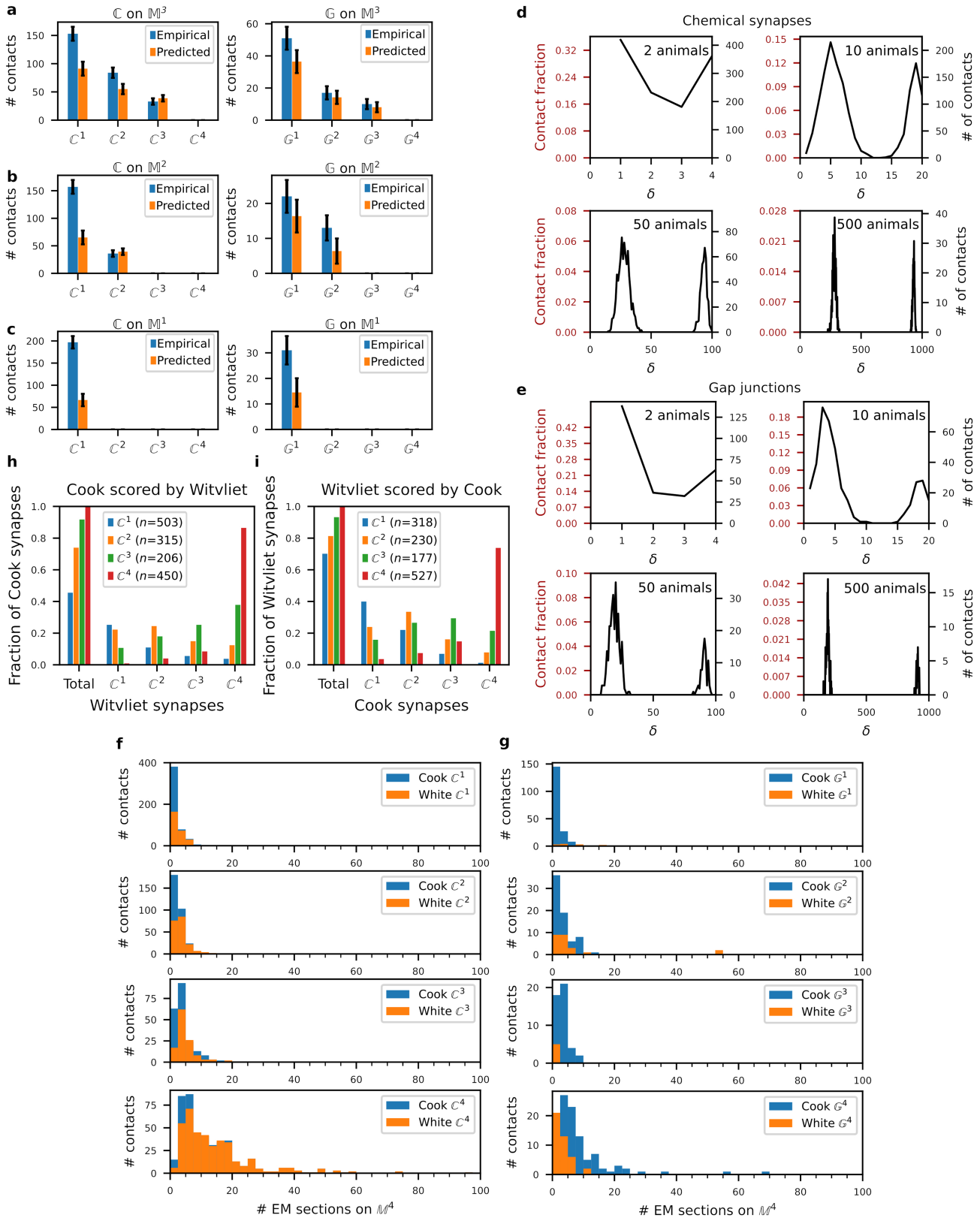
Extended Data Fig. 1 | Neuron neighbourhoods are bilaterally conserved in size, composition and membrane contact positions. a, Variability in immediate neighbourhood size (adjacency degree) does not vary with immediate neighbourhood size. Immediate neighbourhood sizes for each neuron in each dataset (adult left, adult right, L4 left, L4 right, $n = 80$ bilateral cell classes common to L4 and adult) plotted against the immediate neighbourhood size of the corresponding neuron in the adult left. The inset shows the immediate neighbourhood size difference between homologous left and right neurons (vertical spread) as a function of neighbourhood size for the L4 (red) and adult (blue) nerve ring. **b**, The distributions of immediate neighbourhood size differences between homologous contralateral neurons in the same animal—adult left and right (L/R) and L4 L/R— are statistically indistinguishable from 0 (*P* values by two-sided Wilcoxon signed-rank test: 0.07 and 0.29, respectively; $n = 80$ cell classes). Immediate neighbourhood size differences between homologous adult and L4 neurons on the same side of the body are statistically distinguishable from 0 ($P = 9.2 \times 10^{-11}$ by two-sided Wilcoxon signed-rank test; $n = 160$ cells, but the difference is small—mean degree difference 3.6). **c**, Similarity between immediate neighbourhood compositions as quantified by the Jaccard index (Supplementary Results), shows higher compositional similarity between homologous contralateral neighbourhoods ($n = 80$ cell classes) than between proximal ipsilateral neighbourhoods (Supplementary Results; $n = 160$ cells). **d**–**f**, Membrane contact placement along processes is highly reproducible bilaterally and across the adult and L4 datasets. For each process, we mapped each $\mathbb{M}^4$ contact to a position along the anterior–posterior (AP) axis, $\hat{z}$ (see Methods and Supplementary Results). For each $\mathbb{M}^4$ contact, we then counted the number of datasets in which the contact was observed at a given $\hat{z}$ (reproducibility count). **d**, Demonstration of reproducibility count for a single cell class (RIA): RIA has the longest process in the nerve ring and among the highest average reproducibility counts. Shown is a raster plot of reproducibility counts as a function of $\hat{z}$, of all $\mathbb{M}^4$ contacts made with RIA. Neighbouring processes: rows in alphabetical order. Colour: reproducibility count. We define the maximum spatial reproducibility count, $\max(\delta)_{\hat{z}}$, as the highest reproducibility count across all locations, $\hat{z}$, per cell pair (that is, for every row in the raster). For rasters of all other cell classes, see Supplementary Data 2. **e**, Fraction of $\mathbb{M}^4$ membrane contact sites co-localized in δ datasets (distribution over $n = 80$ cell classes). **f**, For each cell class, the fraction of membrane contacts achieved with a maximum spatial reproducibility count, $\max(\delta)_{\hat{z}}$ (distribution over $n = 80$ cell classes). **g**, **h**, Comparatively, $\mathbb{C}^4$ synaptic contact placement is less reproducible than physical adjacency. For each process, we mapped each $\mathbb{C}^4$ contact along the AP axis, $\hat{z}$. **g**, Demonstration of synaptic spatial reproducibility count for RIA neurons. **h**, For each cell class, the fraction of $\mathbb{C}^4$ synaptic contacts achieved with a maximum spatial reproducibility count, $\max(\delta)_{\hat{z}}$ (distribution over $n = 80$ cell classes). Box plots: centre line, median; box limits, upper and lower quartiles; whiskers, 1.5 × interquartile range; points, outliers.



Extended Data Fig. 2 | Contact sizes and reproducibility. a–**f**, Small membrane contact areas are less likely to be bilaterally conserved. Membrane contacts were divided into three groups ('low', 'mid' and 'high') on the basis of their membrane contact areas (35% low, 31% mid, 34% high; see Supplementary Results). **a**, Similarity of homologous (L4 bilateral; adult bilateral; L4 and adult–same side) immediate neighbourhood compositions for low, mid and high membrane contact groups, as measured by the Jaccard index (Supplementary Results; $n = 80$ cell classes). Box plots: centre line, median; box limits, upper and lower quartiles; whiskers, 1.5 × interquartile range; points, outliers. **b**, **c**, Survival (that is, complementary cumulative) distribution of membrane contacts in the adult nerve ring (**b**, $n = 5{,}179$) and the L4 nerve ring (**c**, $n = 4{,}744$). The pie charts show the fraction of total membrane area contact between all processes accounted for by each group. **d**, Empirical frequency distribution of synaptic ($n = 2{,}433$) and gap-junctional ($n = 573$) contacts broken down by the reproducibility of membrane contacts. The majority of synaptic contacts (77% and 85% of synaptic and gap-junction contacts, respectively) occur at $\mathbb{M}^4$ contacts. **e**, **f**, Cumulative distribution of $\mathbb{C}^\delta$ synaptic contacts (**e**) and $\mathbb{G}^\delta$ gap-junction contacts (**f**) for $\delta = 1, 2, 3, 4$ as a function of membrane contact area (in percentiles). To control for differences in neurite placement, we restrict $\mathbb{C}^\delta$ and $\mathbb{G}^\delta$ to contacts that occur on $\mathbb{M}^4$ membrane contacts. The smallest 35% of membrane contacts (dashed line) encompass around 3% of $\mathbb{C}^4$ synaptic contacts and around 9% of $\mathbb{G}^4$ gap-junction contacts (on $\mathbb{M}^4$) with growing fractions for smaller δ (up to around 33% and around 27% of the more variable $\mathbb{C}^1$ and $\mathbb{G}^1$ contacts). **g**, Empirical frequency distribution of membrane, synaptic and gap-junctional contacts across the four datasets ($\delta = 1$ to 4). **h**–**j**, Survival distribution of contacts as a function of membrane contact area for $\mathbb{M}^\delta$ (**h**), $\mathbb{C}^\delta$ (**i**) and $\mathbb{G}^\delta$ (**j**) graphs (n given in **g**), plotting the probability that a membrane, synaptic, or gap-junction contact occurs with a membrane contact area that exceeds some value. Membrane contact areas have been log-normalized and standardized so that the distribution is centred about 0, that is, log-transformed, standardized (by subtracting the mean) and normalized (by dividing by the standard deviation), such that a range of ±1 corresponds to ±1 standard deviation of the distribution of log(membrane contact area).

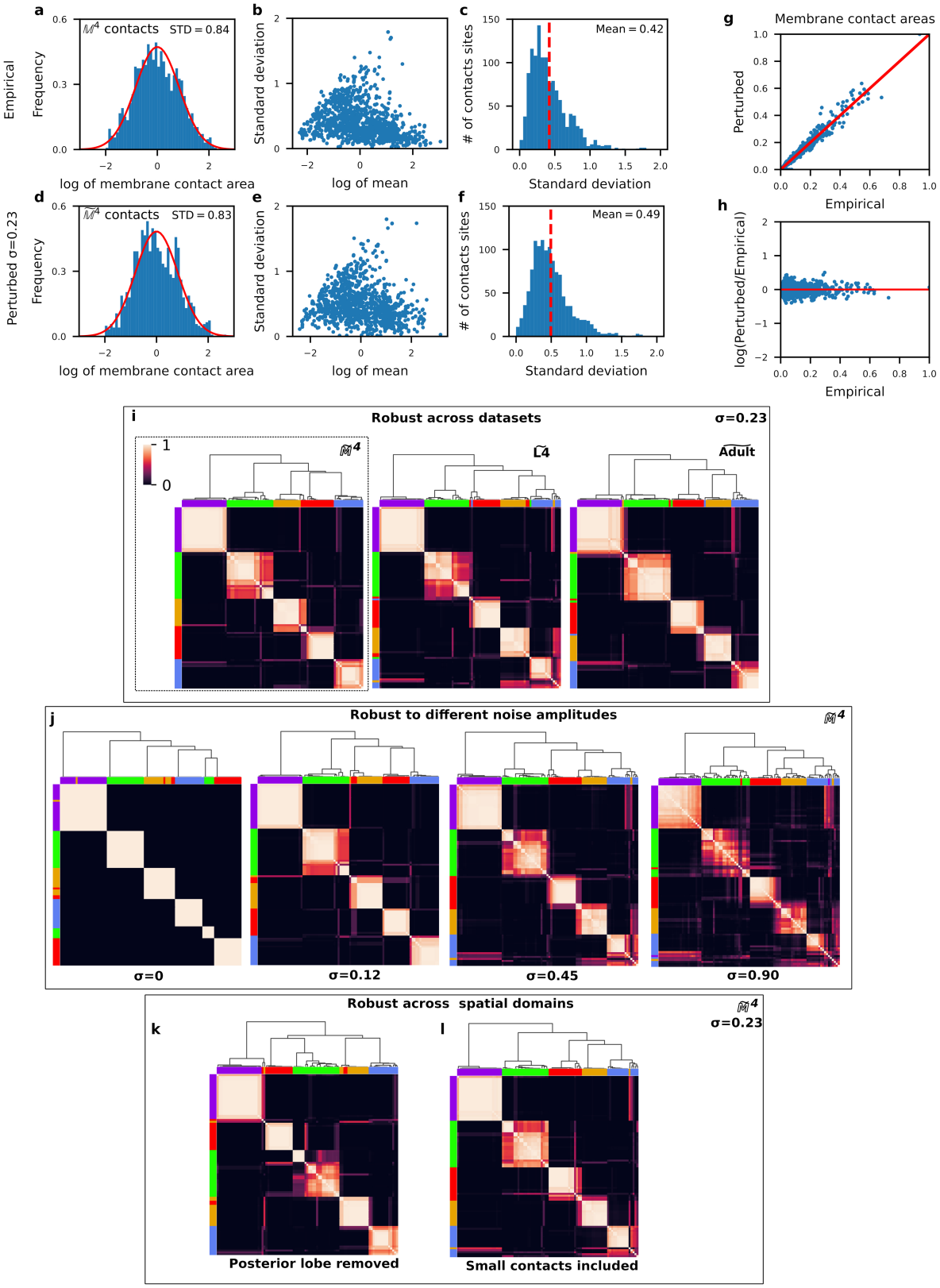


Extended Data Fig. 3 | Core and variable model validations. a, **b**, Model fits for the reproducibility of $\mathbb{M}^{\delta}$, $\mathbb{C}^{\delta}$ and $\mathbb{G}^{\delta}$ contacts, with membrane contact areas below (**a**) and above (**b**) the log-normalized mean (after thresholding; see Methods, Extended Data Fig. 2h). **c**, **d**, Reproducibility model fits of inter-cluster (**c**) and intra-cluster (**d**) $\mathbb{M}^{\delta}$, $\mathbb{C}^{\delta}$ and $\mathbb{G}^{\delta}$ contacts. **e**, Reproducibility model fits for the complete $\mathbb{M}^{\delta}$, $\mathbb{C}^{\delta}$ and $\mathbb{G}^{\delta}$ sets including contacts with the smallest 35% of membrane contact areas (results qualitatively similar to restricted dataset model fit in Fig. 2a; see Methods, 'Generating reference graphs'). **f**, Reproducibility model fits for $\mathbb{C}^{\delta}$ excluding synaptic contacts scored in only one electron micrograph (Methods). **g**, Reproducibility model fits for $\mathbb{C}^{\delta}$ excluding synaptic contacts derived from non-reproducible postsynaptic partners of polyadic synapses (Methods). **h**, **i**, Reproducibility model fits for synaptic and gap-junction contact datasets scored by White et al.[5] (**h**) and Witvliet et al.[20] (**i**) limited to our $\mathbb{M}^4$ contacts. Black bars, empirical distributions used in this study; grey bars, other empirical distributions[5,20]; red bars, model fits for the empirical distributions. All data are presented as fractions of the empirical counts (n).



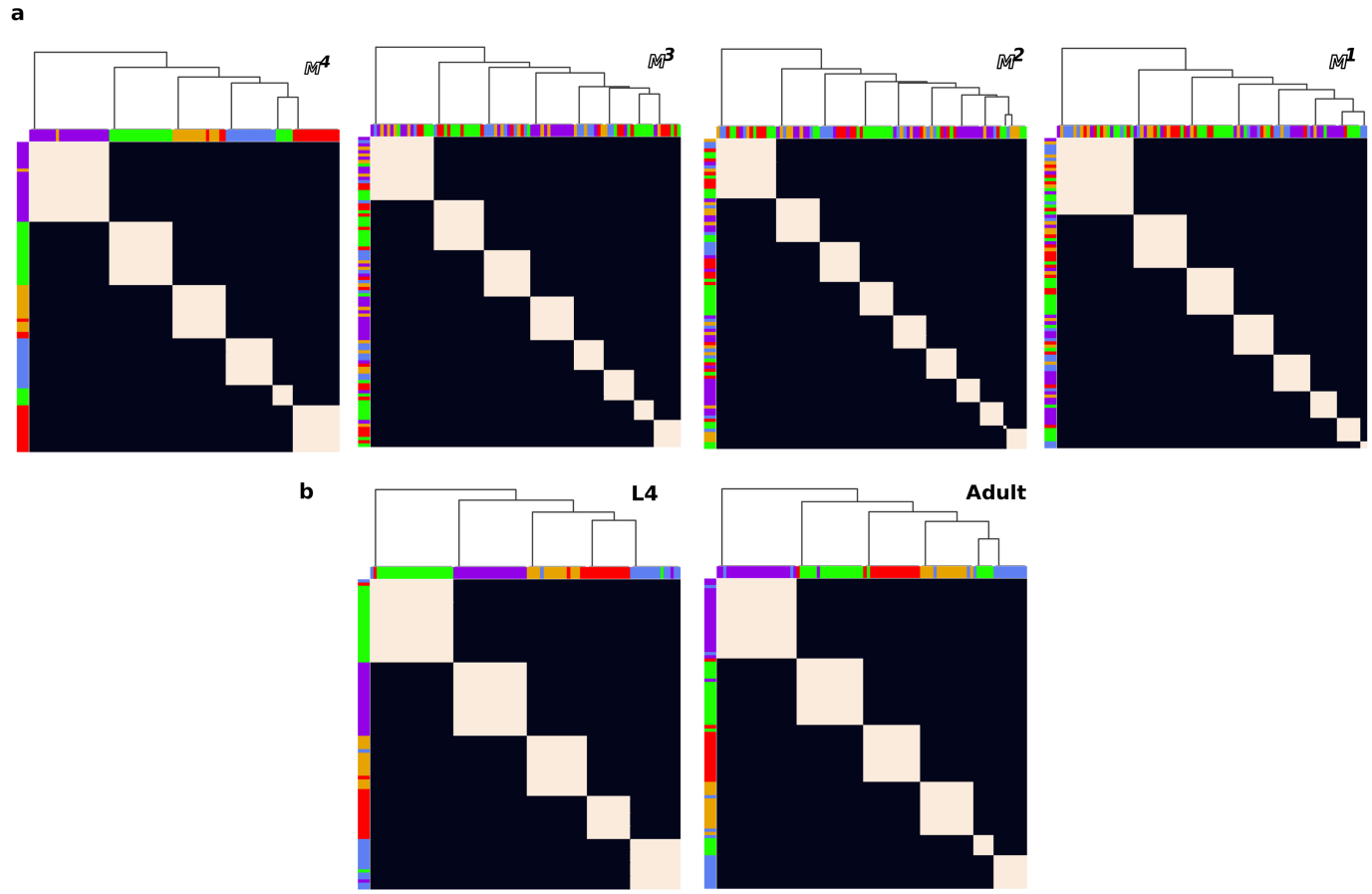
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Validation of core-variable model and contact scoring. a–**c**, The core-variable model reliably predicts the empirical synaptic and gap-junction contact reproducibility ($\mathbb{C}^{\delta}$ and $\mathbb{G}^{\delta}$) on $\mathbb{M}^2$ and $\mathbb{M}^3$. To predict synaptic and gap-junctional contact counts on $\mathbb{M}^{j<4}$ contacts, $\mathbb{C}^{\delta}$ (or $\mathbb{G}^{\delta}$) contact counts on $\mathbb{M}^4$ are scaled by the ratio of 'all $\mathbb{C}$ ($\mathbb{G}$) on $\mathbb{M}^j$ count':'all $\mathbb{C}$ ($\mathbb{G}$) on $\mathbb{M}^4$ count' (Methods). For example, in **a**, the model predicts a $\mathbb{C}^3$ count on $\mathbb{M}^3$ contacts as 206 × 285/1,474 = 40, where 206 is the empirical $\mathbb{C}^3$ count on $\mathbb{M}^4$ contacts, 285 is the total empirical synaptic contact count, $\mathbb{C}$, on $\mathbb{M}^3$ and 1,474 is the total empirical count of synaptic contacts on M^4. The model prediction is consistent with the empirical $\mathbb{C}^3$ on $\mathbb{M}^3$ count (43). Error bars: $\pm\sqrt{n}$, where n is the empirical or predicted count (see Source Data for precise n values). **d**, **e**, Chemical synapses (**d**) and gap junctions (**e**) also consist of a core and variable circuit. Surrogate model data for $\mathbb{C}^{\delta}$ and $\mathbb{G}^{\delta}$, generated as in Fig. 2b. Across each dataset, around 62% of synaptic contacts and around 59% of gap-junction contacts consist of target contacts (given by $fp/[fp + (1-f)(1-s)]$, Methods). **f**, **g**, Core synaptic contacts are typically larger than variable ones in both Cook et al.[3] and White et al.[5]. Distribution of $\mathbb{C}^{\delta}$ (**f**) and $\mathbb{G}^{\delta}$ (**g**) contact counts by electron micrograph sizes (the total number of electron microscopy sections in which a contact was observed)[3,7]. To check for biases in contact size due to possible differences in synaptic or gap-junction scoring criteria, we compare the distributions of electron micrograph sizes for contacts identified by White et al.[5] (orange) and those identified by Cook et al.[3] (blue). Because the White et al. report[5] does not provide electron micrograph sizes, we used the sizes from Cook et al.[3] for all contacts. Although many additional synapses identified by Cook et al.[3] occur only in one section, we find no systematic bias towards smaller synaptic contacts by Cook et al.[3]. **h**, **i**, Bidirectional comparison of Cook et al.[3] and Witvliet et al.[20] synaptic contact reproducibility. **h**, Fraction of Cook et al.[3] synaptic contacts scored by Witvliet et al.[20]. **i**, Fraction of Witvliet et al.[20] synaptic contacts scored by Cook et al.[3]. In **h**, **i**, data are presented as fractions of the total empirical count of synaptic contacts (n).

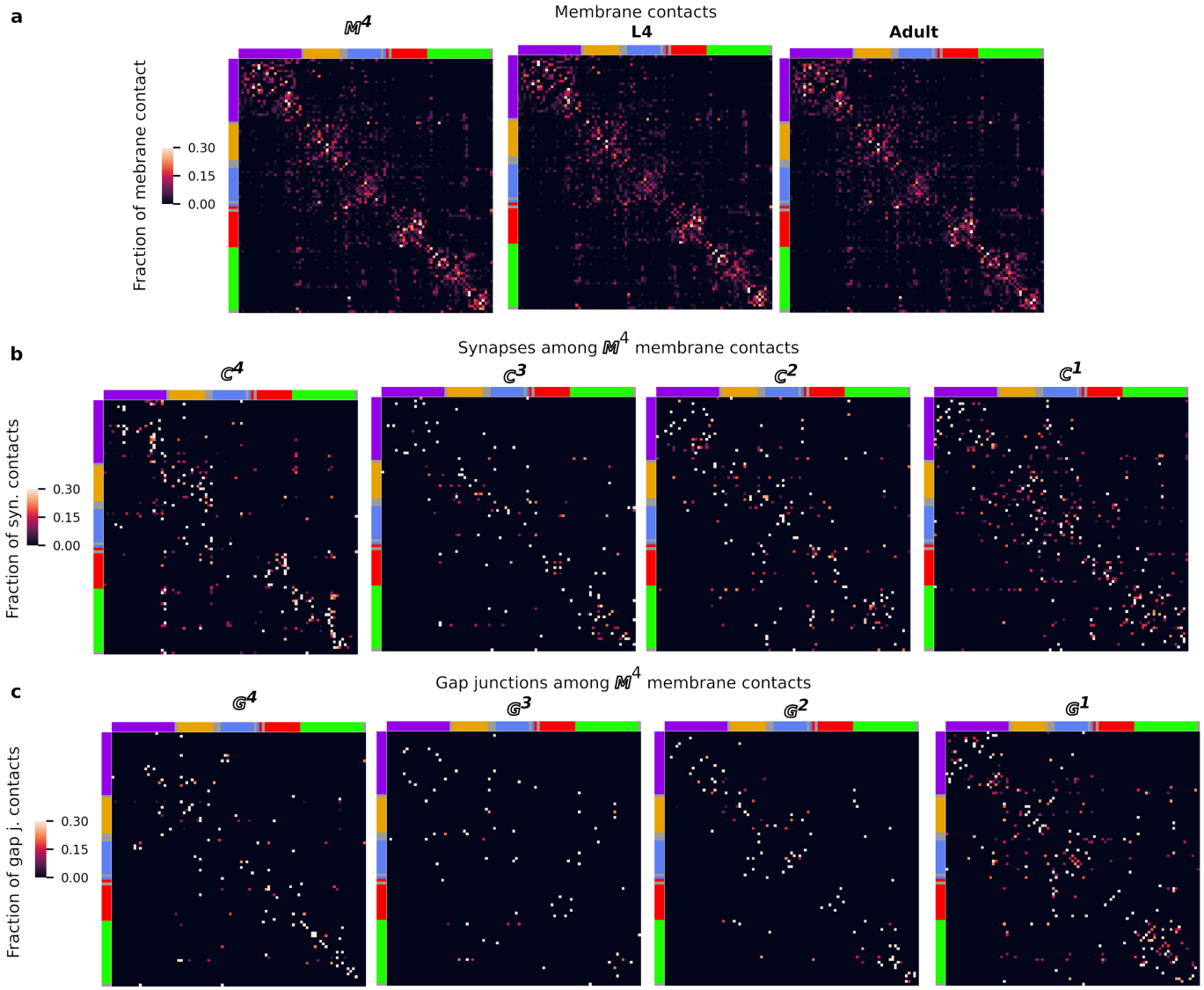


Extended Data Fig. 5 | See next page for caption.

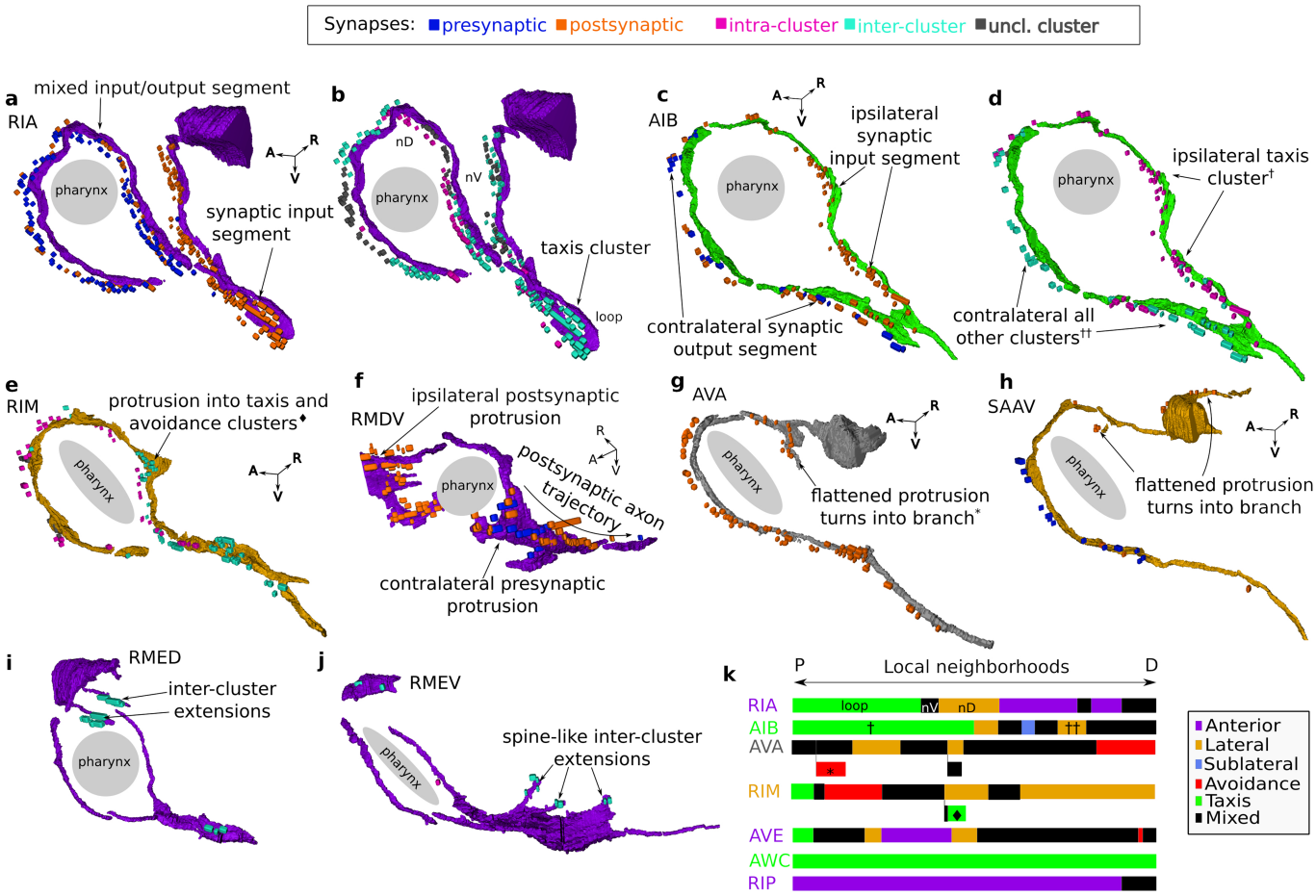
Extended Data Fig. 5 | Robust clustering of nerve-ring processes from $\mathbb{M}^4$ spatial population models. The variability of membrane contacts (Fig. 2, Extended Data Fig. 2) suggests that no single contactome is representative of the population. We estimated the variability among membrane contact areas. **a**, The log-normalized empirical distribution of $\mathbb{M}^4$ membrane contact areas (mean centred at 0; STD, standard deviation; red line, normal distribution with empirical mean and standard deviation; $n = 1{,}258$ membrane contacts). We estimated the variability across the four datasets (L4 left, L4 right, adult left and adult right). For each conserved $\mathbb{M}^4$ contact, we computed the mean and standard deviation of the membrane contact area across the four datasets (see Methods). **b**, Plot of the standard deviation versus the mean contact area across the datasets, where each point is one $\mathbb{M}^4$ contact. Similar to Extended Data Fig. 1a, we find no dependence of the variability on membrane contact area. Therefore, we estimate membrane contact area variability by the mean variability among all membrane contact areas. **c**, The distribution of standard deviations of membrane contact area for all $\mathbb{M}^4$ contacts. Red dashed line indicates mean standard deviation. **d**–**i**, A stochastic spatial population model matches the above distributions by randomly perturbing membrane contact areas in the four datasets with multiplicative white noise with standard deviation (σ) of 0.23 (Methods). **d**–**f**, Spatial population data perturb the membrane contact areas while maintaining contact area and variability distributions that are similar to the empirical $\mathbb{M}^4$ contact area distributions. **g**, Perturbed contact areas scale linearly with the empirical contact areas. **h**, The spread of perturbed contact areas (log of the perturbed contact area as a fraction of the empirical contact area) is mostly uniform across membrane contact areas. **i**–**l**, Neurite clusters obtained from a population of 1,000 $\widetilde{\mathbb{M}^4}$ perturbed individuals and 1,000 $\widetilde{\text{L4}}$ and $\widetilde{\text{Adult}}$ perturbed individuals (perturbing left–right conserved contacts in the L4 and adult contact sets). For each perturbed individual in each population we used a multi-level graph-clustering algorithm to identify spatial clusters. Across each population, we computed the frequency that cell pairs cluster together, represented as an $n \times n$ cluster-frequency matrix ($n = 93$). A hierarchical clustering algorithm is used to sort the rows and columns of the cluster-frequency matrix to minimize variation along the diagonal. Hence, cell pairs that frequently cluster together are sorted together on the cluster-frequency matrix (Methods). Five largely overlapping subgroups of neurons emerge across different perturbations (see Fig. 1 and 'Cluster assignment and validation'). **i**, Consensus clusters are robust across contact sets. $\widetilde{\text{L4}}$ and $\widetilde{\text{Adult}}$ clusters visualized using row and column colours of the $\widetilde{\mathbb{M}^4}$ population cluster assignments (dashed box). **j**, The consensus clusters are robust across different noise amplitudes. Clustering applied to populations generated by perturbations to $\mathbb{M}^4$ using white noise with standard deviations 0 (empirical data), 0.12, 0.45 and 0.9. **k**, **l**, The consensus clusters are robust across different spatial domains. **k**, Clustering applied to $\widetilde{\mathbb{M}^4}$ populations generated from a more spatially restricted volume of the neuropil[34], which excludes its posterior lobe. **l**, Clustering applied to populations generated by perturbations to all reproducible membrane contact areas after restoring the smallest 35% contact areas to each of the L4, adult and $\mathbb{M}^4$ datasets (Extended Data Fig. 2). For all cluster-frequency matrices, matrix element (i, j) corresponds to the frequency that cells i and j cluster together across the 1,000 perturbed individuals. Row and column orders minimize variance along the diagonal (Methods). Cell cluster assignments (colour) follow the perturbed $\widetilde{\mathbb{M}^4}$ dataset (Fig. 1b reproduced in dashed box). Top, dendrogram of the hierarchical clustering.



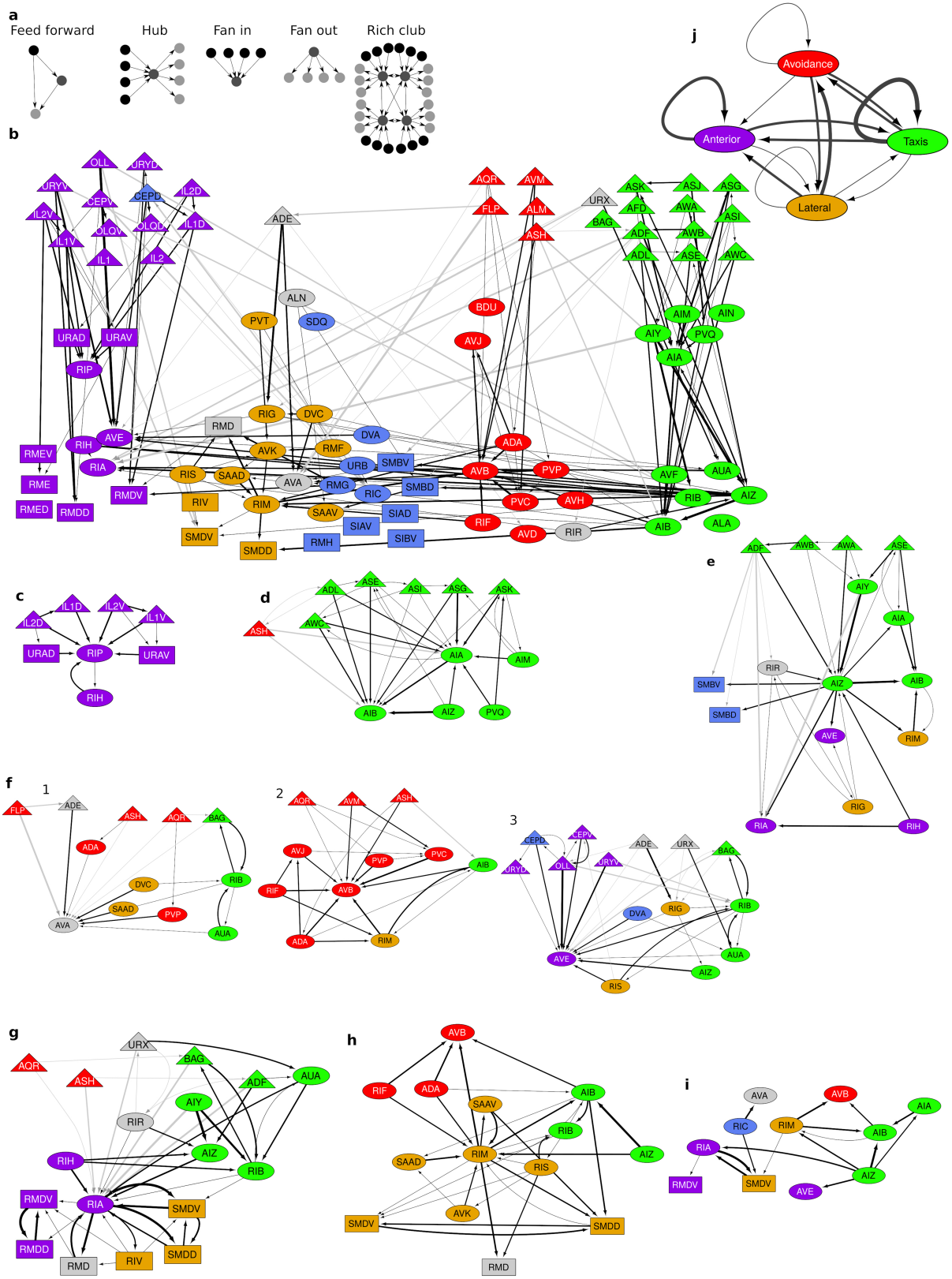
Extended Data Fig. 6 | Variable contacts obscure the organization of the nerve ring. a, Cluster analysis of unperturbed membrane contact datasets $\mathbb{M}^1$, $\mathbb{M}^2$, $\mathbb{M}^3$ and $\mathbb{M}^4$. Clustering results for membrane contacts predicted to combine core and variable contacts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



Extended Data Fig. 7 | Distribution of core and variable synapses among neighbourhoods. a, Membrane contacts of the L4, adult and $\mathbb{M}^4$ reference all have similar membrane contact profiles. For L4 and adult, only bilaterally conserved contacts are included. **b**, Synaptic contacts on $\mathbb{M}^4$ membrane contacts broken down by degree of synaptic contact 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 the main diagonal, whereas variable synapses ($\mathbb{C}^1$) are spread across clusters. **c**, Gap-junction contacts on $\mathbb{M}^4$ membrane contacts broken down by degree of reproducibility ($\mathbb{G}^1$, $\mathbb{G}^2$, $\mathbb{G}^3$ and $\mathbb{G}^4$). For all matrices, row and column ordering is the same as the perturbed $\widetilde{\mathbb{M}}^4$ dataset (Extended Data Fig. 5i). Row and column colours correspond to final clusters assignments (Fig. 1d), where unclassified cells are coloured grey. Matrix element (i,j) corresponds to the fraction of cell i's membrane contact with cell j, with rows normalized to sum to 1.

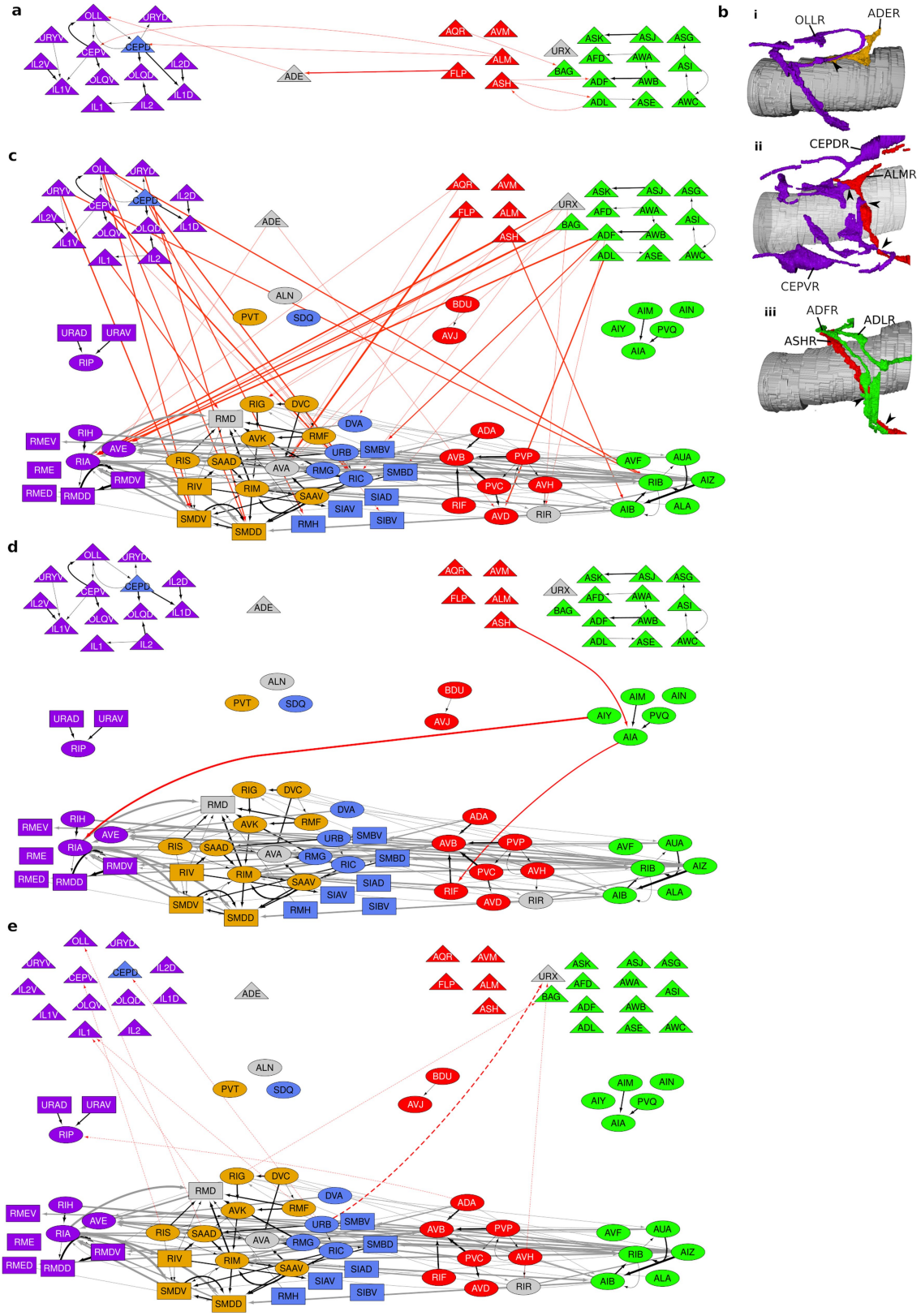


Extended Data Fig. 8 | Subcellular structures support local and nonlocal connectivity, and RIA and AIB processes exhibit synaptic compartmentalization. a, **b**, Volumetric rendering of RIAL and its synapses (cuboids) coloured by synaptic polarity (**a**) or intra-/inter-cluster (**b**). The synaptic input and output segments in **a** correspond to changes in neighbourhood composition in **b**. Changes in RIA neighbourhood correspond to the three neurite segments (nV, nD and loop) which exhibit independent calcium dynamics that encode head movement[46]. **c**, **d**, AIB processes change neighbourhood at the dorsal midline[18]. The ipsilateral segment (†) of the AIB process is surrounded by cells in the taxis cluster, whereas the contralateral segment (††) makes contact with cells in every other cluster. **c**, AIB-process segments alternate between synaptic inputs on the ipsilateral side and synaptic outputs on the contralateral side. **d**, The alternating synaptic inputs and outputs correspond to a change in neighbourhood occurring at the dorsal midline. **e**–**h**, Flattened protrusions link processes to adjacent cells in adjacent clusters. **e**, The flattened protrusion strategy as demonstrated by RIM processes (♦). **f**, The RMDV processes show how flattened protrusions are used to locally expand synaptic polarity. On the contralateral side, the main process trajectory is postsynaptic whereas the contralateral protrusion is presynaptic. Both AVA (**g**) and SAAV (**h**) neurons exhibit flattened protrusions that appear to turn into small branches. The small AVA branch extends into a neighbourhood of cells from a different cluster (*). SAAV ipsilateral branches receive synaptic inputs, whereas its main process trajectory on the contralateral side is mostly presynaptic. Spine-like features extend from RMEV/D processes to cells in a different cluster. **i**, **j**, Two longer RMED extensions (**i**) and three shorter RMEV spine-like extensions (**j**) are postsynaptic to the sublateral cluster. In all images, the pharynx is shown for spatial reference. R, right; A, anterior; V, ventral. Note that for visual clarity, synapses have been offset from the cell process. **k**, Schematic of neighbourhood changes of selected cells (labelled in the colour of the cluster assignment). P, proximal and D, distal to cell body. Each trajectory is scaled to the length of the reconstructed left L4 process. Black boxes denote sections in which the process makes contact with at least two clusters.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Network features of the brain map. a, Schematics of network features (from left to right): Feed-forward loop (FF) motif defined by a triplet of nodes with connectivity: Source → Intermediary → Target and Source → Target; network hub (high-degree node); fan-in (high in-degree node); fan-out (high out-degree node); and rich club (highly connected hubs). **b**, FF triplets within the brain map support the ResNet architecture of the nerve ring. All 101 FF instances among $\mathbb{C}^4$ synaptic contacts (all edges in Fig. 4, Extended Data Fig. 10) are shown. Black arrows, synaptic contacts forming FF motifs within the ResNet architecture (Fig. 4); grey arrows, additional synaptic contacts forming FF motifs (Extended Data Fig. 10). A total of 72 out of 93 cell classes participate in at least one FF motif. Prominent FF targets include AIA, AIB, AIZ, AVA, AVB, AVE, RIA, RIC, RIM, RIP, RMDV and SMDV. Additional contacts superimposed on the ResNet come mostly from cross-sensory module connectivity (Extended Data Fig. 10b). **c**, RIP, the only synaptic link between the somatic and pharyngeal nervous systems, is a major FF target cell for papillary sensory source cells and URA intermediaries. **d**, AIA is a major taxis layer-2 intermediary cell pair regulating information flow from layer-1 taxis sensory cells onto the layer-3 AIB taxis target cell. **e**, AIZ, major layer-3 cells that support nonlocal connectivity (Fig. 3a), serve both to integrate information flow from layer-1 and layer-2 taxis source cells (fan-in) and as intermediary connections to various layer-3 target cells in other modules (fan-out). **f**, Primary locomotion-regulating interneurons—AVA, AVB and AVE—are major layer-3 FF targets and connect extensively onto motor neurons of the ventral nerve cord. Connectivity among these cells occurs in the ventral nerve cord (but is not observed in the nerve ring), suggesting that the regulation of locomotion down the body occurs posteriorly to the nerve ring. **g**, The cell pair RIM, a major hub that supports nonlocal connectivity, serves a triple purpose as a source, intermediary and target of FF motifs within layer 3. **h**, The nonlocal supporter, multi-compartment cell pair RIA is a major FF target for layer-1 sensory (primarily avoidance) source cells and for layer-2 and layer-3 (taxis and avoidance) intermediary cells as well as intermediaries that control layer-3 head motor neurons. In addition, RIA neurons are major targets for feedback from lateral (RMD, RMDD, RMDV) and sublateral (SMDD, SMDV) head motor neurons, consistent with their roles in spatially encoding dorso-ventral head movement to coordinate turning behaviours[46]. **i**, Major targets of FF motifs (11 neuron classes acting as a target of more than three FF motifs, including five classes of rich club neurons) form a highly interconnected subnetwork. Note the frequent representation of some cells in multiple motifs (**c**–**i**). **j**, Aggregated synaptic contacts of layer-3. Layer-3 FF motifs within and among the modules show strong recurrence and no clear feed-forward directionality or hierarchy of layer-3 connectivity, consistent with highly distributed computation. Sublaterals are merged into the lateral module node. Layer-3 anterior cells form FF motifs with only one other module (taxis). All network schematics were generated with Cytoscape v.3.7.1.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Around 17% of $\mathbb{C}_4$ contacts are not accounted for by the ResNet model. a, Layer-1 synaptic connectivity across information-processing modules in $\mathbb{C}^4$ could support distributed sensory computation and integration. Eight (2% of $\mathbb{C}^4$) contacts occur between sensory cells across different modules. **b**, These contacts include: (i) ADE→OLL; (ii) ALM→CEPD/V; and (iii) reciprocal contacts between ASH, ADL and ADF. (i) The mechanosensitive[47,48] anterior neurons OLL loop around intermediate processes, whereas the processes of ADE extend towards the OLL loop, suggesting a functional role for the more elaborate loop morphology. (ii) Both CEPD and CEPV processes loop around intermediate processes and extend flattened protrusions to meet the ALM processes, in which ALM are postsynaptic. CEPD and CEPV neurons respond to head touch[49], whereas ALM neurons respond to both gentle[50] and harsh[51] body touch, inhibit backward locomotion[52] and have been implicated in the habituation of tap response[53]. (iii) Nociception: ASH, ADE and ADF neurons may coordinate avoidance behaviours between the taxis and avoidance modules[54]. **c**, Layer-1 to layer-3 inter-module feed-forward synaptic connectivity in $\mathbb{C}^4$. A total of 54 (12% of $\mathbb{C}^4$) contacts are inter-module, originate in layer 1 and target layer-3 neurons directly. A small number of taxis and avoidance sensory neurons (ADF, ADL, ASH, URX and BAG) project to all but laterals in layer 3; this contrasts with extensive anterior sensory neuron projections that almost exclusively target (sub)lateral layer-3 interneurons and motor neurons, probably mediating rapid sensorimotor transformations. **d**, Layer-2 and inter-module feed-forward $\mathbb{C}^4$ synaptic connectivity. Three contacts (1% of $\mathbb{C}^4$) are inter-module and originate in layer 2. Notably, the layer-2 taxis AIY neurons synapse onto RIA—layer-3 anterior multi-compartment neurons. **e**, Inter-module synaptic feedback connectivity in $\mathbb{C}^4$. Nine (3% of $\mathbb{C}^4$) contacts provide inter-module feedback. Black arrows, synaptic contacts between cells in the same neighbourhood; grey arrows, synaptic contacts between layer-3 cells in different neighbourhoods; red arrows, synaptic contacts not accounted for by the ResNet model; solid arrows, feed-forward or recurrent (intra-layer) synaptic contacts; dashed arrows: feedback synaptic contacts.

nature research

Corresponding author(s): Christopher Brittin

Last updated by author(s): Dec 21, 2020

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	TrakEM2 part of Fiji version date 2014-06-02
Data analysis	Cytoscape 3.7.1, Python v3.6.7, igraph (v0.7.1) implementation of Louvain clustering algorithm, custom code at https://github.com/cabrittin/parsetrakem2 and https:/github.com/cabrittin/elegansbrainmap

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

Data

Policy information about availability of data

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

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

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

Field-specific reporting

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

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

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

Life sciences study design

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

Sample size	We used two legacy serial section EM series, which provided 4 bilateral datasets. No sample-size calculation was performed. Our core-variable model analysis (Fig. 2) indicates that 4 datasets is sufficient to decipher that there is a core component and that the core component can be estimated.
Data exclusions	Our segmentation was restricted to the nerve ring. Cell bodies were excluded from our segmentation because their large sizes skew membrane contact distributions. Dendrites were excluded from our analysis because the contain few synapses. we removed from our analysis a number of neurons that exhibit appreciable differences in synaptic connectivity or process morphology laterally (PLN, PVN, HSN), between the L4 and adult (HSN, PVR, SABD), or those that make minimal membrane contact in the nerve ring (in VB, VC and VD classes), leaving 173 cells in 93 cell classes (the "restricted dataset", Supplementary Information 1). For most of our analyses, we excluded the smallest 35% of membrane contacts because these smaller membrane contacts appear to be highly variable (Extended Data Fig. 2a).
Replication	We exploited the bilateral symmetry of the datasets as a means of replication. We also applied our models to datasets from White et al. (1986) and Witvliet et al. (2020). All attempts at replication were successful.
Randomization	Allocation into experimental groups was not random. In our assessment of the variability of synaptic contacts, we restricted our analysis to synaptic contacts on core (M4) contacts in order to control for differences in process placement. To determine that our core-variable results were not an artifact of synaptic scoring used in this study (Cook et al. 2019), we repeated our analysis using different synaptic scoring (White et al. 1986). and on different datasets (Witvliet et al., 2020). To assess that our clustering is not sensitive to natural variability in membrane contact areas, we derived our clusters from population spatial models (Fig 1c), which yielded consistent results across different datasets (Extended Data Fig. 5i), biological noise levels (Extended Data Fig. 5j) and spatial domains (Extended Data Fig. 5k,l).
Blinding	Blinding is not relevant to the study because test groups were not randomly assigned. Moreover, our analyses depend on knowing and comparing specific neurons between datasets.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

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

Methods

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