A multiscale brain map derived from whole-brain volumetric reconstructions

Christopher A. Brittin, Steven J. Cook, David H. Hall, Scott W. Emmons and Netta Cohen

Supplementary Results

1 Analysis to establish bilateral homology

Having systematically observed that process morphologies are bilaterally conserved[5] (Supplementary Videos 4-7), we asked if relative process placement is also bilaterally conserved. We define two neurons as immediate neighbors if they are physically adjacent in at least one EM section. The set of immediate neighbors for neuron i is the immediate neighborhood of i and the size of the immediate neighborhood is measured by the number of neuron i's immediate neighbors. We first assessed the differences in immediate neighborhood sizes between contralateral homologous neurons. Thus, we can split our two datasets (adult and L4) into four datasets: adult left, adult right, L4 left and L4 right. To assess if immediate neighborhood size and its variation is conserved, we plotted immediate neighborhood sizes for each neuron in each dataset (adult left, adult right, L4 left, L4 right) against the immediate neighborhood size of the corresponding neuron in the adult left (Extended Data Fig. 1a). There are relatively small and even fluctuations in immediate neighborhood size around the diagonal, indicating that the range in immediate neighborhood sizes does not increase with immediate neighborhood size. Furthermore, we find that in both the L4 and adult nerve rings, contralateral immediate neighborhood sizes are typically within ± 5 cells and are statistically indistinguishable from 0 (two-sided Wilcoxon signed-rank test derived p-values: 0.29 and 0.07, respectively, $n = 80$ bilateral cell classes, Extended Data Fig. 1b). Thus despite considerable variability in immediate neighborhood sizes between different neurons, the differences between homologous cells are relatively small.

Of course, similar neighborhood sizes do not necessarily imply similar identities of the neighbors. To test the latter, we measured the similarity between the sets of immediate neighbors of contralateral homologous neurons, here called the homologous similarity, and compared it to the similarity between immediate neighborhoods of spatially proximal ipsilateral neighboring neurons (Extended Data Fig. 1c). For each neuron, the immediate neighbor with the most similar immediate neighborhood (i.e. smallest composition difference) is used to define the proximal similarity (see next section). We use the Jaccard index to measure the similarity between immediate neighborhood compositions, which equals 1 if two immediate neighborhoods are equivalent and 0 if they have no common immediate neighbors (see next section). If neuron processes were randomly placed, then we would expect proximal ipsilateral neighbors to have a larger Jaccard index relative to contralateral homologous neighborhoods. For both the L4 and the

adult, we find that the average homologous similarity is 0.7 while the average proximal similarity is 0.5 (Extended Data Fig. 1c). The relatively high similarity between contralateral immediate neighborhoods, combined with the considerable variability between proximal immediate neighborhoods, is indicative of tissue level specification of process placement.

We next assessed whether reproducible process placement supports bilaterally conserved positions of individual membrane contacts. To facilitate comparisons between datasets, we mapped neurite positions to an effective z coordinate along the AP axis, which we denote $\hat{z}$, such that all 4 datasets are spatially aligned. The alignment step is critical, first, because each worm has a slight bend in the head, and second, because the adult is slightly longer than the L4 but has fewer EM sections in our reconstruction (as only every other EM section in the ventral ganglia was digitized.

To align the 4 datasets, we identified three fiducial points along the AP axis which we defined to be spatially equivalent across datasets: the point where the dorsal body wall muscles enter the nerve ring, the point where the ventral body wall muscles enter the nerve ring and the location of the RMEV cell body. Collectively, these three fiducial points span most of the reconstruction length (along the AP axis). For each of the 4 datasets, we then mapped the z coordinate to an effective position, $\hat{z}$, such that the coordinates of the fiducial points were equivalent across the 4 datasets. For simplicity, we scaled $\hat{z}$ to the range [0,1] and discretized it into 50 segments of equal length, each corresponding to $\sim 0.7\mu$m. As a basis for comparison of contacts between neighboring processes, we only considered membrane contacts that occur across all 4 datasets. To that end, we determined the $\hat{z}$ coordinates of all $\mathbb{M}^4$ contacts across the 4 datasets. For each $\mathbb{M}^4$ contact, we define the reproducibility count as the number of datasets where the contact was observed at a given $\hat{z}$ (Extended Data Fig. 1d).

Formally, for each cell l, we define a matrix $D^{(l)}_{ijk}$ where i denotes the position $\hat{z}_i$, j denotes the neighbor $j \neq l$ and k denotes the datasets (L4 left, L4 right, adult left, or adult right). $D^{(l)}_{ijk} = 1$ if the cell l contacts cell j at $\hat{z}_i$ in the k^{th} dataset and 0 otherwise. The reproducibility count, $S^{(l)}_{ij}$, for a given coordinate i and contact between cells l and j is given by

$$S^{(l)}_{ij} = \begin{cases} \sum_k D^{(l)}_{ijk}, & \text{if } \sum_k D^{(l)}_{ijk} > 0, \\ \text{undefined}, & \text{otherwise,} \end{cases}$$

which takes on values of $\{1,2,3,4\}$. The raw counts for each reproducibility count value (aggregated across all $\mathbb{M}^4$ contacts and all $\hat{z}$ coordinates) is shown in Extended Data Fig. 1e). We also computed the fraction of each reproducibility count value per cell, l; the mean and standard deviation of these fractions is given in Extended Data Fig. 1f).

We find that for every effective location of membrane contact between $\mathbb{M}^4$ partners in which at least one contact is observed, there is a high ($\sim 45\%$) likelihood to find co-localized contacts in all four datasets and $\sim 70\%$ likelihood of finding at least 3 contacts at the same effective location (Extended Data Fig. 1e,f).

Finally, homologous similarities are correlated with the membrane contact area between immediate neighbors. We measured the membrane contact area between neighboring cells and found that 95% of measured membrane contact areas ranged from 1 nm^2 to 10 μm^2 (Extended Data Fig. 2a). The largest membrane contact area is over 60 μm^2 between the sublateral SMB motoneurons. Notably, the upper 33% of membrane contact areas ($>$1.77 μm^2) account for 86% of the total membrane contact area between all cells in the nerve ring, while the lower 35% of membrane contact areas ($<$0.4 μm^2) account for only 2% of the total surface area contact between all neurons (Extended Data Fig. 2a). Based on this, we hypothesized that smaller membrane contact areas are likely noise, i.e. random contacts between cell processes. We calculated the homologous similarity where membrane contact areas were restricted to either *low* ($<$ 35%), *medium* (35%-67%) or *high* ($>$67%) membrane contact areas. The average homologous similarity for the high, medium and low membrane contact areas is 0.65, 0.45 and 0.25, respectively (Extended Data Fig. 2b,c). Thus, homology similarity increases with membrane contact area.

2 Similarity between homologous and proximal immediate neighborhoods

To quantify the variation in neighborhood composition, we compared immediate neighborhoods of homologous neurons. Most nerve ring cell classes are by and large bilaterally symmetric, which allowed us to compare homologous neighborhood on the right and left side of the animal. Additionally, we could compare immediate neighborhood compositions of equivalent neurons between the L4 and adult. Let i and i' be homologous neurons with neighborhoods $N(i)$ and $N(i')$, respectively. A popular metric for the similarity of two sets is the Jaccard index, which is computed as

$$J(N(i),N(i')) = \frac{|N(i) \cap N(i')|}{|N(i) \cup N(i')|}. \tag{6}$$

We refer to this as the homologous similarity between immediate neighborhoods.

To establish a meaningful baseline for the homologous similarity between immediate neighborhood compositions, we computed the Jaccard index between neighborhoods that might be expected to be most similar. We reasoned that two neighboring neurons with processes that innervate the nerve ring together can be expected to have many common immediate neighbors. We say that the two neurons have overlapping proximal immediate neighborhoods. Let $N(i)$ be the set of immediate neighbors for neuron i and let neuron j be one of these immediate neighbors, i.e. $j \in N(i)$. Then in practice, it is typically the case that $N(i) \cap N(j) \neq \emptyset$, i.e. the immediate neighborhoods of i and j overlap. We define the proximal similarity as the Jaccard index, $J(N(i),N(j))$, of the two neighborhoods. In order to find pairs of neurons with the largest proximal similarity, for each neuron i we compute the maximum Jaccard index over its set of immediate neighbors,

$$\Delta_i^{\max} = \max_{j \in N(i)} \{J(N(i),N(j))\} \tag{7}$$

By computing $\Delta^{\max}$ for each neuron, we get a distribution of the similarity between the most proximal immediate neighborhoods.

3 Validation of core-variable synaptic model

Our model predicts that in each animal, the conserved core connectome is supplemented by a large set of variable contacts. Even restricting our count to reproducible membrane contacts, each dataset consists of over $\sim 40\%$ variable synapses and gap junctions (Extended Data Fig. 4d,e). We find that most but not all variable contacts appear in < 5 EM sections, although some small synapses are highly reproducible (Extended Data Fig. 4f). To validate the robustness of our results, we compared our datasets with those of White *et al.*[5] and Witvliet *et al.*[20] (Extended Data Fig. 4) and repeated our analysis for those datasets, as well as for more restricted, more conservative scoring of synaptic contacts (thresholding EM sections, or eliminating the most variable polyadic synapses, see Methods, Extended Data Fig. 3). We conclude that while the exact delineation of core and variable contacts would require additional nerve ring reconstructions, the existence of a considerable variable component (comprising $\gtrsim 50\%$ of synapses across different datasets) is robust. These results cannot provide evidence of the functionality of synapses in the variable circuit, but they hint at considerably more individual variability and redundancy in wiring than previously expected[6,18].

4 Functional identification of neuron process clusters

The taxis cluster (green) has amphid sensory neuron pairs that modulate a number of attractant behaviors: ASE (NaCl chemotaxis[56]), AWA, AWB and AWC (volatile odors chemotaxis[57,58]), AFD (thermotaxis[59]) and ADF (aerotaxis[44]). Additionally, the taxis cluster includes the major amphid interneurons of classes AIA, AIB, AIY and AIZ[46]. The avoidance cluster (red) consists of the major polymodal nociceptive sensory neuron pairs ASH[60], mechanosensory neurons of class AVM and ALM[52,61] and O_2 and CO_2 sensing cell AQR. Additionally, the avoidance cluster includes premotor locomotion command interneurons AVB, AVD and PVC that directly control forward and reverse locomotion[62,63]. The anterior cluster consists of cells with cell bodies that are anterior to the nerve ring (e.g. the papillary sensory cells). The defining functional feature of the anterior cluster is a set of mechanosensory neurons (CEP, IL, OLL)[64,65]. The remaining sublateral and lateral clusters mostly consist of inter- and motoneurons that regulate posture, head movement and locomotion. Notable in the sublateral cluster are a set of motoneurons that innervate the dorsal and ventral sublateral nerve cords (SIA and SIB) and which pioneer nerve ring development[34,66]. We refer to the final cluster as the lateral cluster because several of the neurons (RIV, SAAV and SMDV classes) have cell bodies located in the lateral ganglia and most have processes that extend contralaterally around the nerve ring.

5 Rich-club neurons exhibit meso-connectivity properties

The so-called "rich club" of the *C. elegans* neural circuit (DVA, AIB, RIB, RIA, AVE, AVD, AVA, PVC), refers to a set of 9 cell classes that have both very high degree (i.e. 'hubs') and are highly connected to each other (high assortativity)[24] (Extended Data Fig. 9a). We find that all "rich club" cells make $\mathbb{C}^4$ contacts with cells in 2 or more distinct clusters and 4 of the cell classes (AIB, RIB, RIA and AVA) exhibit subcellular specializations. These neurons appear, from their connectivity, to mediate sensory-motor transformations[25] and have been linked to whole-brain dynamics in the worm[23], which could be indicative of a broader role meso-connectivity plays in synaptically linking and coordinating spatially disparate parts of the nerve ring.

6 Network motifs and architectural features of the brain map

Network features, such as hub and rich club neurons and the feed-forward loop network motif (see schematics in Extended Data Fig. 9a) have previously been identified in the *C. elegans* connectome. To frame these observations in the context of our brain map, we considered the $\mathbb{C}^4$ synaptic contacts. The resulting network has a variable degree distribution with a median of 4 and a maximum of 24. 20 high-degree (defined here as $\geq$10) hub neurons are found in Layer 1 (OLL, CEPD and CEPV) and, more prevalently, in Layer 3. To gain a better understanding of the role of hub neurons, we wish to distinguish between high in-degree (fan-in) nodes, indicative of integration roles, and high out-degree (fan-out) nodes with likely coordinating roles (Extended Data Fig. 9a). We noted that (due to the high assortativity of the *C. elegans* connectome[67,68]) high-degree nodes are also likely to be the main sources, intermediaries and targets of instances of the feed-forward loop network motif[26]. This motif is defined by a triplet of nodes with directed connectivity in which one node (the source) connects to the other two (intermediary and target) and the intermediary also connects to the target. The feed-forward loop motif is also the key macro-level organizing principle that emerges from our brain map (Layer 1 projects to Layers 2 and 3, and Layer 2 projects to Layer 3). We therefore asked whether triplets of nodes obeying feed-forward loop motif connectivity can be said to form the skeleton of the brain map, or rather to complement it.

To address this question, we combined $\mathbb{C}^4$ contacts (across Fig. 4 and Extended Data Fig. 10). We used the FANMOD algorithm[69] to identify all connected triplet of nodes and identified the feed-forward loop motifs among those. $>$ 80% of all $\mathbb{C}^4$ contacts participate in at least one feed-forward loop motif. Within this network of exclusive feed-forward loop motifs, a number of fan-out (typically feed-forward source nodes: ADA, ADF, ADL, AIB, AIZ, AQR, ASH, ASI, AWA, CEPD, CEPV, FLP, IL2V, RIF, RIS, URX, URYV), fan-in (typically target nodes in Layer 3: AIA, AIB, AIZ, AVA, AVB, AVE, RIA, RIC, RIM, RIP, RMDV, SMDV) and fan-in-fan-out nodes (typically intermediaries in Layer 2 or 3, ADE, AIA, AIB, AIZ, ASG, AUA, AVJ, AWC, IL1V, PVC, PVP, RIB, RIG, RIM, SAAD, URAV), are highlighted (Extended Data Fig. 9c-h) to shed light on their putative information processing roles. The feed-forward loop motifs sub-network is also consistent with the brain-map result of highly distributed (both intra-cluster

and inter-cluster) connectivity of Layer 3, that lacks any obvious internal feed-forward directionality (Extended Data Fig. 9i-j). We observe that the most prominent feed-forward connectivity outside the strict template of the brain map (Extended Data Fig. 9) corresponds to extensive connections from anterior sensory neurons (Layer 1) to lateral interneurons and head and neck motoneurons in Layer 3 (Extended Data Fig. 9b).

References

[56] Bargmann, C. I. & Horvitz, H. R. Chemosensory neurons with overlapping functions direct chemotaxis to multiple chemicals in *C. elegans*. *Neuron* **7**, 729–42 (1991).

[57] Bargmann, C. I., Hartwieg, E. & Horvitz, H. R. Odorant-selective genes and neurons mediate olfaction in *C. elegans*. *Cell* **74**, 515–527 (1993).

[58] Troemel, E. R., Kimmel, B. E. & Bargmann, C. I. Reprogramming chemotaxis responses: Sensory neurons define olfactory preferences in *C. elegans*. *Cell* **91**, 161–169 (1997).

[59] Beverly, M., Anbil, S. & Sengupta, P. Degeneracy and neuromodulation among thermosensory neurons contribute to robust thermosensory behaviors in *Caenorhabditis elegans*. *J. Neurosci.* **31**, 11718–11727 (2011).

[60] Kaplan, J. M. & Horvitz, H. R. A dual mechanosensory and chemosensory neuron in *Caenorhabditis elegans*. *Proc. Natl. Acad*. Sci. *USA* **90**, 2227–2231 (1993).

[61] Pirri, J. K., McPherson, A. D., Donnelly, J. L., Francis, M. M. & Alkema, M. J. A tyramine-gated chloride channel coordinates distinct motor programs of a *Caenorhabditis* elegans escape response. *Neuron* **62**, 526–38 (2009).

[62] Piggott, B. J., Liu, J., Feng, Z., Wescott, S. A. & Xu, X. Z. The neural circuits and synaptic mechanisms underlying motor initiation in *C. elegans*. *Cell* **147**, 922–933 (2011).

[63] Zhen, M. & Samuel, A. D. *C. elegans* locomotion: Small circuits, complex functions. *Curr. Opin. Neurobiol*. **33**, 117–126 (2015).

[64] Hart, A. C., Sims, S. & Kaplan, J. M. Synaptic code for sensory modalities revealed by *C. elegans* GLR-1 glutamate receptor. *Nature* **378**, 82–85 (1995).

[65] Goodman, M. B. Mechanosensation. WormBook (2006).

[66] Kennerdell, J. R., Fetter, R. D. & Bargmann, C. I. Wnt-Ror signaling to SIA and SIB neurons directs anterior axon guidance and nerve ring placement in *C. elegans*. *Dev.* **136**, 3801–3810 (2009).

[67] Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442 (1998).

[68] Bassett, D. S. & Bullmore, E. T. Small-World Brain Networks Revisited. *Neurosci*. **23**, 499–516 (2017).

[69] Wernicke, S. Efficient detection of network motifs. *IEEE/ACM Trans. on Comp. Biol. Bioinfor.* **3**, 347–359 (2006).