Hub connectivity and gene expression in the worm connectome

Aurina Arnatkevičiūtė[1☯], Ben D. Fulcher[1☯], Alex Fornito[1]

1 Brain and Mental Health Laboratory, Monash Institute of Cognitive and Clinical Neurosciences, School of Psychological Sciences, Monash University, 770 Blackburn Rd, Clayton, 3168, VIC, Australia.

☯These authors contributed equally to this work.
* correspondingauthor@institute.edu

Abstract

Text

Author summary

Text

Introduction

1. Structural network properties
    - rich-club, hubs, conservation cross-scale/species/methods
    - Prior work in worm: different types of connectomes - wired/unwired, hubs
2. Structure-expression
    - Datasets in different scales: mouse/rat, human (also heritability), developmental, large-scale datasets
    - Kaufman example [1], Baruch 2008 [2], Varadan 2006 [3] also relevant.
3. Hubs plus gene expression
    - groundbreaking (& breathtaking) work by Ben and Alex [4]
    - Different types of connections - different expression
4. Summary

The complex pattern of axonal connections between neural elements of nervous systems provides key insights into the physical constraints underlying a functionally optimized system. Mapping the network of structural connections between neural elements, from the scale of neurons in model organisms like *C. elegans*, through to mesoscale brain regions using tract tracing in rodents, through to non-invasive MRI imaging in human subjects, has revealed a surprising consistency of organizational properties, suggesting a common set of selection pressures under which efficient neural systems have evolved. In this work, the brain network, or 'connectome', is abstracted to

a graph representation, in which neural elements are represented as nodes (e.g., neurons in *C. elegans*, or macroscopic brain regions in human), with edges connecting pairs of elements that have an axonal connection; in this way allowing very different species, measurements, and scales of neuronal networks to be represented in a consistent and unified way. As well as showing a modular network structure, that can often be interpreted in terms of distinct functional brain networks, connectomes have been consistently shown to exhibit a non-uniform distribution of connectivity across the network, leading to the existence of highly connected 'hub' nodes. Hubs have also been shown to be strongly interconnected, forming an integrated 'rich club' that may be crucial for facilitating efficient integration of information between more specialized processing modules. Prior work has characterized the rich club organization of the neuronal connectome of the *C. elegans* nervous system [5], *Drosophila* [6], mouse [7], rat [8], 53 [9] or 65 areas [10] of the cat cortex, 242 macaque cortical regions [11], and in 82 brain regions [12] and 1 170 human cortical areas [13].

With the availability of gene expression data in the worm,

Connectivity in brain networks is not uniformly distributed. Network elements with high node degree – i.e., a large number of connections to other areas – are called 'hubs'. When hubs are more densely interconnected than expected by chance they form a 'rich-club', the idea being that the richest members of the network (in terms of connections) are tightly connected to each other, thus forming a club. These densely interconnected hubs are thought to promote efficient integration between anatomically distinct areas and play an important role in brain functioning. It has been shown that hubs exhibit distinct transcriptional signatures in both humans [14] and mice [4]. According to **(author?)** [4], connections involving rich club hubs carry a distinctive genetic signature, which is driven by genes regulating the synthesis and breakdown of adenosine triphosphate (ATP) – the primary energetic substrate of neuronal signaling [4]. These findings highlight a close relationship between metabolic expenditure and the high signaling load of hub regions in the brain, as has been previously proposed [15].

We therefore have preliminary indications that the transcriptional signature of hubs may be a consistent feature of mammalian brain networks, but it is not known how distinctive this expression signature is; and in particular, whether it holds true for networks resolved at the scale of individual neurons and synapses. To test this possibility, we aimed to replicate findings presented in [4] using microscale connectivity data in *C. elegans* and gene expression data from *WormBase*. We sought to determine whether hubs in the *C. elegans* connectome exhibit distinctive gene expression patterns.

Materials and methods

Neurons making up the nervous system of the nematode worm *C. elegans* can be divided into groups, commonly as: sensory neurons (support receptive function), motor neurons (cells containing neuromuscular junctions), interneurons (all other neurons) and polymodal neurons (performing more than one type of circuit function) [16].

- Connectivity data: types of synapses, weighted/binary, degree; Rich club analysis, how neurons are annotated
- Expression data: where data comes from, processing (options: annotations, qualifiers)
- Lineage data: where and when downloaded from, how defined. What for used in here

- Coexpression metric: how did we choose? How different measures depend on the number of genes; The effect of coexpression/space: can not be corrected for; Excluding left/right homolog gene expression from calculations
- Enrichment: software (ermineJ), how to score genes (maybe define gene scoring within results as scoring for connected and unconnected and rich/feeder vs peripheral is different)

To investigate the relationship between hub connectivity and gene expression in a micro-scale network we coupled two publicly available datasets containing synaptic-level connectivity network and gene expression signatures for the somatic nervous system of the *C. elegans* hermaphrodite.

Neuronal connectivity data

The C. elegans nervous system consists of two distinct parts: a small pharyngeal nervous system (20 neurons), responsible for the feeding behaviour of the animal and a large somatic nervous system (282 neurons). Previous work has commonly divided them into three groups as: sensory neurons (support receptive function), motor neurons (cells containing neuromuscular junctions) and interneurons (all other neurons) [16].

Classification of neurons to 'sensory', 'motor', 'interneuron' types, and as 'head' or 'tail', was done using the anatomical hierarchy defined in WormBase [17], which we retrieved using their API, and propagated these (and other child terms of 'neuron', WB:0003679) to all individual neurons. [[TODO: how and where from hierarchy file was downloaded?]]

The neurotransmitter types of the majority of neurons in the hermaphrodite C. elegans nervous system has been mapped, revealing the dominance of cholinergic signaling [18]. We matched data from Table 2 in **(author?)** [18] to label the neurotransmitter type of all neurons.

Lineage data are based on previously published embryonic and post-embryonic lineage trees [19, 20]. The dataset used for the analysis was downloaded from WormAtlas (`http://www.wormatlas.org/neuronalwiring.html#Lineageanalysis`). In this dataset for each pair of neurons a common ancestor cell was identified using and a total number of cell divisions from the common progenitor was calculated. Downloaded data were arranged into a matrix format where lineage distance for each pair of neurons was defined.

Neuronal connectivity data was obtained for 279 somatic neurons (282 nonpharyngeal neurons, excluding CANL/R and VC6, for which connectivity data is unavailable) of the C. elegans nervous system, as described in **(author?)** [21]. Data were downloaded from WormAtlas (`http://www.wormatlas.org/neuronalwiring.html#NeuronalconnectivityII`). [[TODO: Summarize briefly how the data were originally collected, using EM, from White and then updated by Varshney etc., and what connections were recorded and how categorized – i.e., both electrical gap junctions and chemical synapses]]

The data distinguish presynaptic and postsynaptic neurons (providing connection direction information) as well as the number of synapses (providing connection weight information). The result can be represented as a weighted, directed connectivity matrix, with weights given by the number of synapses. By contrast, the directionality of electrical gap junctions remains unknown, and have thus been represented as bidirectional connections.

To avoid any possible differences in gene expression between chemical synapses and gap junctions, our analysis focuses just on the synaptic connectivity network, which contains a total of 1 961 connections (6394 synapses). Note that this is in contrast to

previous analysis by **(author?)** [5], which symmetrized chemical connections (ignoring directed information) and added electrical gap junctions as undirected connections. [[TODO: contextualize this section by adding references to past work – e.g., when we say we just look at synaptic network – has this been done in the past?]]

Two dimensional spatial co-ordinates for 277 neurons (including VC6) were obtained from `www.biological-networks.org/?page_id=25`. Positions for 3 neurons not included in this dataset (AIBL, AIYL, SMDVL) were reconstructed based on the positions of the contralateral neurons (AIBR, AIYR, SMDVR) by assigning identical coordinates for the contralateral neurons in these three pairs according to Varier (2011) [22].

Network analysis

In this section, we describe the methods used to characterize the C. elegans structural connectome, represented as a graph with neurons as nodes and synaptic connections as edges [23], which will later allow us to compare gene coexpression as a function of connectivity structure.

The number of regions that a neuron projects to is its out-degree, k_{out}, and the number of regions that project to a given neuron is its in-degree, k_{in}. The total number of connections involving a given neuron is defined by its degree, $k = k_{\text{in}} + k_{\text{out}}$.

At a given k threshold, neurons were classified as either 'hub' (degree $> k$) or 'non-hub' (degree $\leq k$). All edges could subsequently be classified as either 'rich' (hub $\rightarrow$ hub), 'feeder' (hub $\rightarrow$ non-hub or non-hub $\rightarrow$ hub), or 'peripheral' (non-hub $\rightarrow$ non-hub). For simplicity, hub $\rightarrow$ non-hub and non-hub $\rightarrow$ hub connections were grouped as a single 'feeder' category for simplicity, [[TODO: check whether gene coexpression patterns were similar for connections running in both directions. DONE - outgoing feeder links are more similar then incoming feeder links see: (`https://www.evernote.com/shard/s94/nl/2147483647/756bddde-7bf3-4aae-8a8c-b94bdd9b0e57/`)].

In order to assess rich-club organization of our networks, we used the rich-club coefficient, defined as

$$\Phi(k) = \frac{2M_{>k}}{N_{>k}(N_{>k} - 1)}, \tag{1}$$

where $N_{>k}$ is the number of nodes with degree $> k$ and $M_{>k}$ is the number of edges between them.

When hubs in the network are more densely interconnected than expected by chance, they form a rich-club - a sub-network of high degree nodes that share a disproportionally high number of connections between themselves.

In order to examine whether high degree nodes then to preferentially connect to other high degree nodes, at each degree k we calculated a rich-club coefficient (see eq. 1).

It quantifies the tendency of nodes with degree $\geq k$ to preferentially connect to each other by calculating how many edges are present between nodes of degree at least k, normalized by how many edges there could be between these nodes in a complete graph. Rich club coefficient monotonically increases with degree as only nodes with degree ¿k are retained and this alone yields a higher expected density of a sub-graph, therefore a normalized rich-club coefficient (2) is used in order to define, whether this increase in the empirical network is higher that could be expected by chance alone.

$$\Phi_{\text{norm}}(k) = \frac{\Phi(k)}{\Phi_{\text{random}}(k)}, \tag{2}$$

where $\Phi_{\text{random}}(k)$ is the average value of $\Phi(k)$ across random networks. $\Phi_{\text{norm}} > 1$ indicates the rich-club organization of the network.

In order to calculate the normalized rich club coefficient Φ_{norm} we computed a 1 000 random networks rewiring each edge an average of 50 times per null network and evaluated the ratio between rich club coefficient in the empirical $\Phi(k)$ and random $\Phi_{random}(k)$ resulting in $\Phi_{\text{norm}}(k)$ at each degree threshold. Statistical significance for Φ_{norm} was estimated by computing P value directly from the empirical null distribution under the null hypothesis $\Phi(k) \leq \Phi_{\text{random}}(k)$.

Rich club coefficient and null networks were generated using the rich_club_bd and randmio_dir functions from the Brain Connectivity Toolbox [24] respectively.

Modularity Modular structure of the connectome was determined applying Louvain community detection algorithm using a standard practice, implemented in Brain Connectivity Toolbox [24]. To identify the optimal modular assignment, the algorithm was repeated 1000 times resulting in a thousand community affiliation vectors. Then, we calculated the agreement matrix using modularity index of each run as a weight. Each element in the agreement matrix defines a relative probability for a pair of nodes to be assigned to the same module, where more weight is added for higher modularity partitions. In the next step, we performed consensus clustering on the agreement matrix using 1000 repeats ($\tau = 0.1$). The agreement matrix is thresholded at a level τ to remove an weak elements. τ threshold controls the resolution of the reclustering with low values corresponding to lower resolution. Afterwards, The resulting matrix is partitioned a thousand times using the Louvain algorithm. Second run of clustering produces a new agreement matrix and the process is repeated until partitions converge into a single representative partition. According to this implementation, any singeleton communities are not reconnected to the network.

Gene expression

Gene expression signatures for each neuron were obtained from the publicly available WormBase database (release WS256 downloaded on 6th Feb 2017 from `ftp://ftp.wormbase.org/pub/wormbase/releases/WS256/ONTOLOGY/anatomy_association.WS256.wb`). The dataset contains binary expression profiles for 948 genes annotated to at least one neuron. To minimize noise in the dataset we selected genes that were directly annotated to a specific neuron and labelled with the following qualifiers in the anatomy association file: certain, enriched, partial, blank. Each neuron on average is associated with the expression of 30 genes (range: 3 to 138 genes), while each gene is expressed on average in 9 out of 279 neurons (range: 1 to 148 neurons). It is important to note that, because the absence of expression data not recorded in the database, it is not possible to distinguish between the following two cases: (i) "gene is not expressed" and (ii) "there is no information on whether gene is expressed". Both cases are recorded and analyzed as "0" expression here.

Gene expression was assigned to neurons based on anatomical hierarchy obtained from... [[TODO BEN: PROBABLY SHOULD ADD INFORMATION ABOUT HIERARCHY FILE HERE as it's also from the wormbase (how/when downloaded/what queries were used]]

Gene coexpression

As shown in Fig. 1A, we have a gene expression profile for each neuron that indicates which of 948 genes have been observed to be expressed in that neuron. Given the pairwise nature of connectivity patterns, in this work we analyzed patterns of coupled gene expression between pairs of neurons. To this end, we computed the similarity in

gene expression profiles between a pair of neurons (i,j) using a binary analogue of the linear Pearson correlation coefficient, the mean square contingency coefficient, r_ϕ [25].

$$r_\phi = \frac{n_{11}n_{00} - n_{10}n_{01}}{\sqrt{n_{1\bullet}n_{0\bullet}n_{\bullet 0}n_{\bullet 1}}}, \tag{3}$$

for two vectors x, y, of length N, where n_{xy} count the number of observations of each of the four outcomes (e.g., $n_{10} = \sum_i \delta_{x_i,1}\delta_{y_i,0}$ counts the number of times $x = 1$ and $y = 0$ across the length of x and y), while the symbol $\bullet$ sums across a given variable (e.g., $n_{\bullet 0} = \sum_i \delta_{y_i,0}$ counts the number of times $y = 0$). A maximum value $r_\phi = 1$ when x and y are identical such that $n_{11} + n_{00} = N$, and a minimum value $r_\phi = -1$ when x and y are always mismatched such that $n_{10} + n_{01} = N$.

Note that, in addition to r_ϕ defined above, a range of other similarity measures for binary strings exist, including the Jaccard index, $n_{11}/(n_{10} + n_{01} + n_{11})$, Yule's Q coefficient, $(n_{00}n_{11} - n_{01}n_{10})/(n_{00}n_{11} + n_{01}n_{10})$, and the χ^2 index, $N(n_{00}n_{11} - n_{01}n_{10})/(n_{1\bullet}n_{0\bullet}n_{\bullet 0}n_{\bullet 1})$ [1].

Since the gene expression data analyzed here is sparse (max 15% expression), we needed to ensure that our similarity metric was not sensitive to the proportion of genes expressed in a given neuron. To evaluate this we used numerical simulations on randomly generated binary vectors to perform the comparison between different correlation measures. The length of each vector was set to match then number of genes in the data (948) and the number of ones ranged from 1 to 150 according to the observed number of genes expressed in each neuron resulting in a 150x948 binary matrix. Vector pairwise correlations were calculated using different correlation measures and an average of 1 000 runs was taken to summarize the behavior of each measure. Mean square contingency coefficient ϕ met this criterion best. In contrast, other measures such as Jaccard and Yule's index are highly dependent on this ratio (see S1 Fig.).We therefore selected r_ϕ to analyze in this work, which was insensitive to the proportion of genes expressed in each neuron.

In the binary and sparse gene expression data analyzed here (with a minority of genes being positively expressed in each neuron), we wanted to distinguish biologically relevant cases where genes are expressed together, n_{11}, from the dominant case where a gene is not expressed, n_{00}. To this end, we developed a new index based around that probability, $P(m)$, that two binary strings will contain m 'positive matches' (i.e., matching 1 with 1),

$$P(m) = \binom{n_2}{m}\binom{N - n_2}{n_1 - m}\Big/\binom{N}{n_1}, \tag{4}$$

for two binary expression vectors, x_i, y_i, of length N, containing n_1 and n_2 1s, respectively ($n_2 \leq n_1$), with m matches (n_{11}). Our index, r_ξ, computes the probability of at least this many matches as $r_\xi = 1 - \sum_{x=0}^{m-1} P(x)$. This showed ([[biases?]]) which gave similar results [[TODO: check that behavior is indeed similar]].

Previous work has emphasized the importance of spatial effects in driving patterns of gene expression, for example, with gene coexpression exhibiting spatial autocorrelation in the mouse brain [7] and human cortex [26, 27]. Given that connection probability is also dependent on spatial separation in mammalian brains [28, 29], it is important to correct for spatial effects to ensure that relationships between gene expression and connectivity are not simply the result of both properties showing spatial autocorrelation. In *C. elegans*, we found a semi-exponential relationship between connection probability and Euclidean separation distance (based on their two dimensional coordinates) between a pair of neurons, but no correlation between the Euclidean separation distance between a pair of neurons and their gene coexpression, r_ϕ ($r = X$, $p = Y$, cf. (see **??**)).

A total of 92 anatomically distinct neuron classes in the hermaphrodite C. elegans have a bilateral representation, meaning that each of those neuron classes have both left

and right counterparts (e.g., AVAL/AVAR). These neurons exhibit highly correlated gene expression patterns, with a mean coexpression value $\langle r_{\phi}^{\text{intraclass}} \rangle = 0.98(SD \pm 0.0.25)$. [[CODE used - testLeftRightCoexpression.m. Distribution for these links is not normal. Median is 1;]] To ensure that our analyses are not influenced by this symmetry, we excluded coexpression values within neuron classes from all analyses reported here.

Gene enrichment analysis

To interpret differences in gene coexpression between different types of node pairs, we used enrichment analysis to identify functionally annotated groups of genes from the Gene Ontology (GO) [30]. In previous work involving continuous expression data, we developed a score for each gene that quantified its contribution, at each edge, (i, j), to the overall Pearson correlation across all genes [7]. However, with the current binary expression data, expression of a given gene, a, across brain regions, i: $g_i^{(a)}$, can take three possible combinations of coexpression at each edge, (i, j): (i) *both expressed*: $g_i^{(a)} = 1$ and $g_j^{(a)} = 1$; (ii) *mismatched*: $g_i^{(a)} = 1$ and $g_j^{(a)} = 0$ or $g_i^{(a)} = 0$ and $g_j^{(a)} = 1$; or (iii) *neither expressed*: $g_i^{(a)} = 0$ and $g_j^{(a)} = 0$. Our measure attempts to score individual genes by their coexpression type in different types of inter-region pairs (e.g., connected versus unconnected pairs, or rich/feeder/peripheral connections). In particular, we were interested in quantifying increases in a gene, a, being expressed in both regions of a given class of inter-region pairs, which we computed as a cumulative binomial probability of obtaining at least m matches ($g_i^{(a)} = 1$ and $g_j^{(a)} = 1$) being observed empirically:

$$\nu^{(a)} = P(X > m) = 1 - \sum_{i=0}^{m} \binom{n}{i} p^i (1 - p^{n-i}), \tag{5}$$

where p is the probability of the given class of inter-region pairs ($p = n_{\text{class}}/M$ for n_{class} connections out of M possible pairs), and n is the maximum number of possible matches. This provides us with a score, $\nu^{(a)}$ for each gene, a, that gives higher values to genes that shows a greater number of matches in the inter-region class of interest than expected by chance. For example, when comparing differences in binary coexpression between pairs of brain regions that are structurally connected and pairs that are unconnected, Eq. (5) has p as the proportion of all possible connections that exist, and captures the probability of genes being expressed together in connected pairs of brain regions at least as frequently as observed. When comparing links involving hubs compared to those between nonhubs, p is the proportion of connections involving hubs.

To ensure that each gene contributed a meaningful score, we required more than one pair of regions $m > 1$ to exhibit matching expression ($g_i^{(a)} = 1$ and $g_j^{(a)} = 1$) in the inter-region class of interest (577 genes out of 948 satisfied this criterion for connected/unconnected analysis; 390 out of 948 genes satisfied this criterion for hub connectivity analysis). [[AURINA: is this correct? So *more than 1 match* means we require at least 2 matches??IT IS CORRECT - there are quite a few genes that have only one match on mask. Enrichment with them seems to be more noise than signal.]]

For a given enrichment analysis, after computing a probability score for each genes using Eq. (5), we then took logarithms of the probability scores, $\log_{10} \nu^{(a)}$, and used them to perform gene score resampling (GSR) analysis using version 3.0.2 of *ErmineJ* software [31] using biological process Gene Ontology (GO) annotations [30] obtained from GEMMA [32] (`Generic_worm_noParents.an.txt.gz` downloaded on February 10 2017). [[AURINA: did we only look for biological processes, or also cellular components, or molecular pathways? were doing only ONLY BIOLOGICAL PROCESSES]JUST

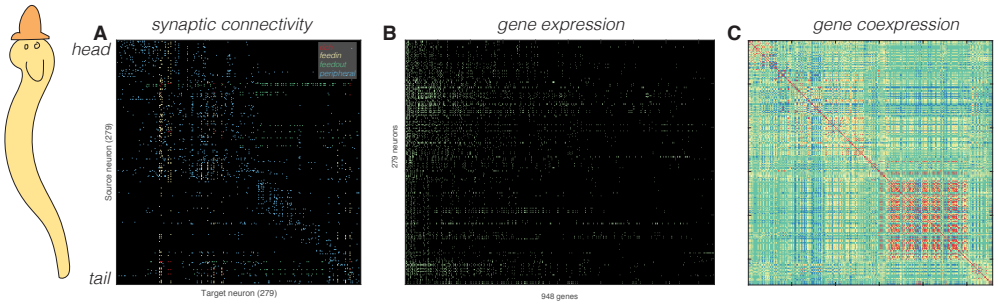


Fig 1. Schematic data representation A Binary synaptic connectivity between all pairs of neurons, colored according to type: rich, feeder, and peripheral. **B** Gene expression across the 948 genes that are expressed in at least one neuron. **C** Gene coexpression matrix. [[TODO: I think it would be nice to put neuronal metadata on the edge]] Neurons have been ordered by their position, from head (upper) to tail (lower).

DOWNLOADED NEW annotation file (uploaded 2017 Feb 27 - slightly more genes are included, enrichment results change!!! for BIOLOGICAL PROCESSES only 2 glutamate related categories) see: https://www.evernote.com/shard/s94/nl/2147483647/0dad1deb-c858-4611-b2da-8337ebf85567/] Rather than setting an arbitrary threshold on the score, this analysis allows us to use the continuous log-transformed probability scores, and compute the probability that the mean score in each GO group is larger than expected from a random assignment of scores to the genes in the list. Gene Ontology terms and definitions were obtained in RDF XML file format downloaded from `archive.geneontology.org/latest-termdb/go_daily-termdb.rdf-xml.gz` on February 10 2017. We included ontology terms containing between 5 and 100 annotations, using full resampling with 10^7 iterations.

Results

We aim to quantify patterns in the pairwise synaptic connectivity of C. elegans that relate to patterns in the pairwise similarity in gene expression (gene coexpression). A schematic overview of our data is shown in Fig. 1, which includes the directed binary synaptic connectome (Fig. 1A), binary gene expression across 948 genes (Fig. 1B), and resulting gene coexpression between all pairs of neurons (Fig. 1C). Edges in the connectome are colored according to how they connect hubs ($k > 42$) and non-hubs ($k \leq 42$), as 'rich' (hub $\rightarrow$ hub), 'feed-in' (nonhub $\rightarrow$ hub), 'feed-out' (nonhub $\rightarrow$ hub), and 'peripheral' (nonhub $\rightarrow$ nonhub).

We present our analysis in three parts: (i) we analyze the synaptic connectivity properties, demonstrating a rich-club organization of hubs, despite their increased separation distances, (ii) we show that gene coexpression is increased in connected pairs of neurons (relative to unconnected pairs), and varies also as a function of edge type, with the most similar gene expression patterns found in pairs of connected hub neurons, an effect that is specific to interneurons (mirroring previous results in mouse [7]), (iii) despite relatively incomplete gene annotations, we show that these transcriptional signatures may be driven by glutamate receptor signaling, and that the results are not driven simply by stereotypical interneuron expression or modular organization.

Hub connectivity in the C. elegans connectome

First, we analyze the topological properties of the C. elegans connectome, represented here as a directed, binary connectivity matrix of the 1961 chemical synapses between

279 non-pharingeal neurons [21], focusing particularly on hub connectivity. The degree distribution of the C. elegans chemical connectome is shown in Fig. 2A, where neurons have been distinguish according to type: 68 sensory neurons, 85 interneurons, 108 motor neurons, and 18 neurons that have multiple annotations (annotations from WORMBASE[[TODO: add reference]]). Consistent with previous work [[]], we see a positively-skewed degree distribution containing an extended tail of high-degree hubs, the majority of which are interneurons [[characterized in previous work as control INs]].

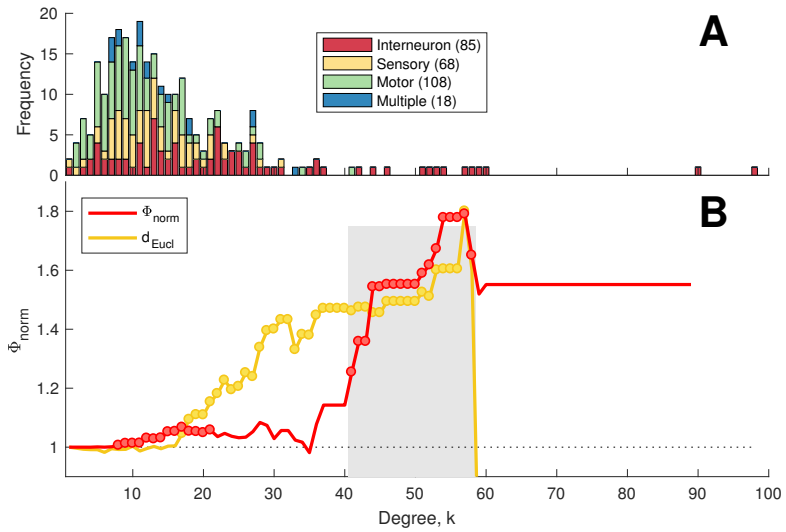


Fig 2. Rich-club organization of the connectome. **A** Degree distribution of the binary chemical connectome, where neurons are labeled according to four categories: (i) interneuron (85 neurons, red), (ii) sensory (68 neurons, yellow), (iii) motor (108 neurons, green), or (iv) multiple assignments (18 neurons, blue). An extended tail of high-degree neurons can be seen, which are mostly interneurons. **B** Normalized rich club coefficient, Φ_{norm} (red), as a function of the degree, k, at which hubs are defined (as neurons with degree $> k$). Also shown is the mean Euclidean separation distance (yellow) between connected hub regions (across degree thresholds, k). $\Phi_{\text{norm}} > 1$ indicates that hubs are more densely interconnected among each other than expected by chance, with red circles indicate values of Φ_{norm} that are significantly higher than an ensemble of 1 000 degree-matched null networks ($P < 0.05$). Yellow circles indicate where the Euclidean distance between connected pairs of hubs is significantly greater than the Euclidean distance for all other pairs of connected regions (Welch's t-test, $P < 0.05$).

We quantified the extend to which hubs are densely interconnected by computing the normalized rich-club coefficient, Φ_{norm}, with $\Phi_{\text{norm}} > 1$ indicating rich-club organization of the network. The variation in Φ_{norm} across a range of thresholds, k (at which hubs are defined, as neurons with degree $> k$), is shown in Fig. 2B, with red circles indicating a significant increase in link density among hubs relatively to 1 000 degree-preserving nulls (permutation test, $P < 0.05$) [[check whether randmio_dir only preserves degree]]. The plot reveals rich-club organization at the upper tail of the degree distribution, particularly for thresholds $41 < k < 58$, referred to here as the 'topological rich-club regime', shaded gray in Fig.2B. Most of the analyses performed here compare variation as a function of the hub threshold, k, but for analyses requiring a fixed hub definition, we define hubs as the 13 neurons with $k > 41$ [[TODO: compare to Towlson]].

[[maybe include a list with degrees and definitions in SUPP (like in Towlson's paper?). Are our hubs similar of different? Why?).]]

Recent work has shown that many aspects of connectome organization can be partially accounted for by geometric effects, that treat the connectome not just as an abstracted graph topology, but as a spatially-embedded object [28, 33, 29]. In Fig. 2B, we plot the mean distance between all pairs of connected hubs at each degree threshold, k, finding an increase in mean hub-hub connection distance with k, through to the extent of the topological rich-club regime. This increase can be attributed to a relative increase in connections between the head and tail in hub-hub connections relative to other connection types [[TODO: should we show evidence in the supps?]]. Despite an increase in hub-hub connection density across longer distances, connection probability decreases with distance in C. elegans (see Fig. 6A, particularly for head-head and tail-tail connections, up to $\approx 0.2\,\text{mm}$, and body-body connections up to $\approx 0.4\,\text{mm}$, and head-tail and tail-head connections, from $\approx 1.0 - 1.2\,\text{mm}$), mirroring recent results in mouse [34, 7], and other rodents and primates [29]. Taken together, our results demonstrate that hub-hub connections in C. elegans are dense, and extend over significantly longer distances than other types of connections in the synaptic connectome, consistent with the idea of the rich club as a costly but central backbone for neuronal communication [13].

Coexpression and connectivity

Having characterized the topological and geometric properties of connectivity, in particular the rich-club organization of hub neurons, we next investigate how connectivity properties relate to patterns of neuronal gene expression. Being a pairwise analysis, we compare pairwise synaptic connectivity to pairwise similarity in gene expression (using a binary Pearson correlation measure, r_ϕ, to define gene coexpression, and excluding coexpression values between homologous left/right neuron pairs, see Methods). We first investigated whether gene coexpression relates to synaptic connectivity by comparing the distribution of r_ϕ computed for all connected pairs of neurons, and for all unconnected pairs of neurons, as shown in Fig. 3A. We find that connected pairs of neurons have more similar expression profiles than unconnected pairs (Wilcoxon rank-sum test, $P < 10^{-50}$). Using directed connectivity information to split connections into unidirectional ($X \to Y$) and reciprocal ($X \leftrightarrow Y$) subtypes, we found no difference in coexpression as a function of connection reciprocity, but both groups showed an increase in coexpression relative to unconnected pairs of neurons (see Fig. [[]]).

Having established an increase in gene coexpression in connected pairs of brain regions (relative to unconnected pairs), we next investigated whether coexpression varied between different types of connections, focusing particularly on the role of densely interconnected hub nodes characterized above. Given the importance of hub neurons across species [[Ref]], and previous results in the mesoscale mouse connectome [7], we expect increased coexpression in connections involving hubs. To investigate this effect, we first labeled each neuron as either a hub (nodes with degree $> k$) or a nonhub (degree $\leq k$), for a given hub threshold, k, and then labeled each connection as either 'rich' (hub $\to$ hub), 'feed-in' (nonhub $\to$ hub), 'feed-out' (hub $\to$ nonhub), or 'peripheral' (nonhub $\to$ nonhub). [[TODO: Establish whether we're computing pairs of hubs (e.g., for 'rich'), or for every connection between any pair of hubs – i.e., do reciprocal connections contribute twice to the distribution at each k?]] The median coexpression for each of these four connection types described above is plotted in Fig. 3B, with circles indicating statistically significant increases of a given connection type (relative to all other connections) [[TODO: Why are we suddenly doing medians? subplot A is mean, right? Is median a better summary of the distribution than mean? Depending on distributions, the mean MAY raise fewer flags if results are similar using mean]]. We see that gene coexpression in rich connections increases with degree,

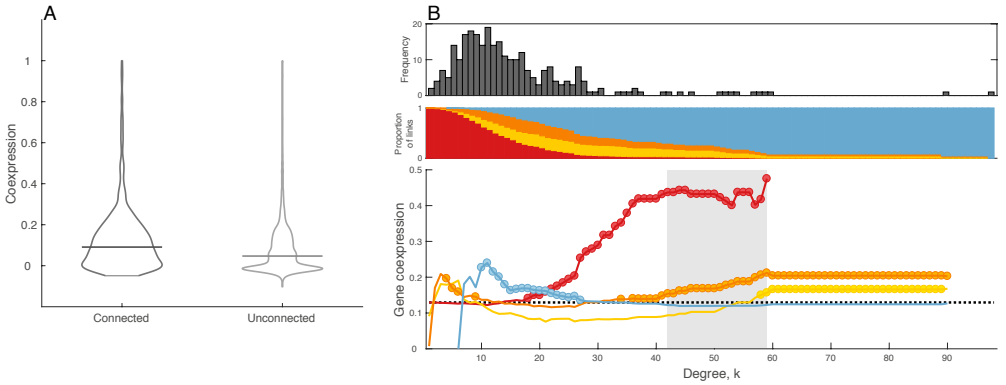


Fig 3. Gene coexpression varies as a function of connectivity, and as a function of connection type, with connections involving hubs being most similar. A Distribution of transcriptional similarity between connected ([[XX]] pairs) and unconnected ([[XX]] pairs) pairs of neurons as a violin plot, with the mean of each distribution shown as a horizontal line. Coexpression, r_ϕ is increased in connected pairs of neurons ($P < 10^{-50}$, Wilcoxon rank sum test). **B** *Top*: Degree distribution. *Middle*: proportion of connections that are 'rich' (hub→hub, red), 'feed-in' (nonhub→hub, yellow), 'feed-out' (hub→nonhub, orange), and 'peripheral' (nonhub→nonhub, blue) as a function of the degree threshold, k, used to define hubs. Note that at high k, most neurons are labeled as nonhubs, and hence the vast majority of connections are 'peripheral'. *Lower*: Median Pearson gene coexpression, r_ϕ, for each connection type as a function of k. The median coexpression across all network links shown as a dotted black line; the topological rich-club regime (determined from the network topology, cf. Fig. 2) is shaded gray. Circles indicate a statistically significant increase in gene coexpression in a given link type relative to the rest of the network (Wilcoxon rank sum test; $P < 0.05$).
[TODO: A: add r_ϕ to ylabel. Perhaps we want to use different colors (confusing: we shouldn't use red for both connected and rich within the same plot; same for blue, should be just peripheral, not peripheral and unconnected). B: Remove box from degree distribution, Change y-label to 'Median gene coexpression', fonts are small, and inconsistent sizes between subplots. I think we should crop the curves when there are fewer than X connections remaining – in this case it's a bit ridiculous to show that constant feed-out/feed-in/peripheral coexpression when there are only two neurons left as hubs]

reaching a plateau through the topological rich-club regime where hubs are densely interconnected (shaded in Fig. 3B). Interestingly, we see a split between feed-in and feed-out connections, with feed-out connections showing increased gene coexpression through the topological rich-club regime, while feed-in and peripheral connections show the lowest levels of coexpression.

These results, using partial binary expression estimates in 948 genes in a neuronal connectome, are broadly consistent with results in 213 regions of the mesoscale mouse connectome with over 17 000 genes, in which we see an increase in coexpression in connected pairs of neurons, and the greatest coexpression amongst hub-hub connections and the lowest coexpression amongst nonhub-nonhub connections. In contrast to the mouse connectome, however, in which hubs were distributed across a broad range of anatomical divisions, hubs in C. elegans are [[mostly. There is one motor neuron in there, right?]] interneurons [5] (cf. Fig. 2A). To determine the specificity of the effect between different neuron types, we split our coexpression analysis into connections

involving: (i) interneurons, (ii) sensory neurons, and (iii) motor neurons, computing the median coexpression within each class of connections as a function of the degree of the given neuron type [[TODO: is this true? It's kind of complicated to explain – is it right that for a given k threshold we include, say, sensory neurons of degree greater than k and their connections to all other neurons in the 'sensory' category? Or do we include all neurons with degree $> k$ and compute the median of those connections that involve both the high-degree neurons AND neurons of the given class? This would be another choice. Perhaps it's the latter that's plotted?]], as shown in Fig. 4A. Note that the three categories are not mutually exclusive; for example, the 'interneuron' class includes connections between any interneuron and any other neuron. We find that connections involving interneurons are unique in their increasing [[average]][[median? See comments above about whether to use mean throughout... Need to inspect the distributions to motivate the right choice]] coexpression with increasing degree, k. Connections involving motor neurons, and connections involving sensory neurons, do not show any clear increase with degree.

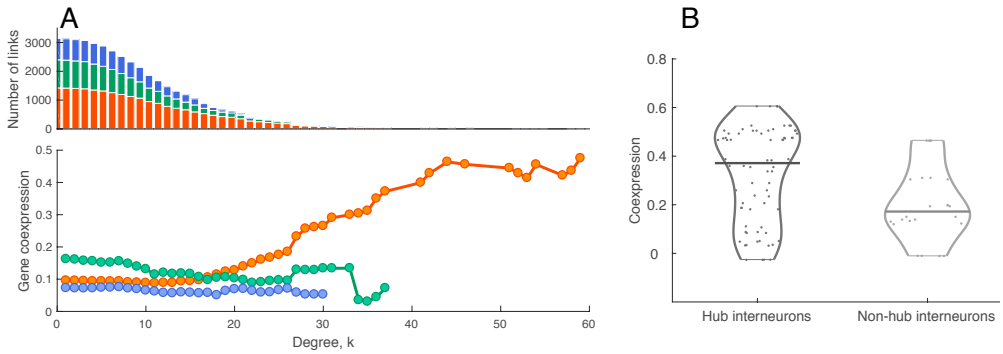


Fig 4. The increase in coexpression with degree is specific to high-degree interneurons, and hub interneurons have higher coexpression than nonhub interneurons A *Upper*: The number of connections in each category for a range of degrees. *Lower*: Average gene coexpression as a function of degree for connections involving different types of neurons. **B** Coexpression distributions for hub and non-hub interneurons. [[TODO: Did we exclude multimodal neurons from this analysis?? Can we do a statistical comparison between these distributions and add something here on that?]] [[TODO: A: the vertical axis in lower subplot should state 'median gene coexpression']]

Given the distinction of interneurons in driving the gene coexpression relationship with degree, k, we next verified whether hubs showed a unique transcriptional signature among this neuronal subtype. We compared gene coexpression between all pairs of hub interneurons [[TODO: what k threshold?]] and between all pairs of nonhub interneurons (AVFL, AVFR, AVHL, AVHR, AVKL, AVKR, AVJL, AVJR) [[TODO: what does this list represent? Not clear from the sentence structure]], as shown in Fig. 4B. Pairs of hub interneurons display significantly higher gene coexpression than pairs of nonhub interneurons ([[TODO: statistical test, p-value, etc.]]). Thus, the increase in coexpression amongst hub interneurons is not simply due to most hubs being interneurons; even amongst interneurons, [[TODO: isn't this obvious from plot A? It's very similar – at high degree, we're only looking at hub interneuron pairs, which are more similar than others. Not sure B adds much – we could just quantify the difference at a given k threshold to put a statistic to it...?]].

Hubs and modules

In the following paragraphs discuss several aspects that could have lead to the increased coexpression for hub-related links.

Distance plays and important role in gene coexpression in mammalian brains, where proximal brain areas exhibit more similar gene expression patterns [4, 35]. While we found no correlation between Euclidean distance between a pair of neurons and their gene coexpression, we performed an additional analysis to ensure that the increase in coexpression for hub links is not influenced by any distance effects. Implementing the same methods as previously used in gene coexpression analysis at each degree threshold for each link type (rich, feeder, peripheral) we calculated the mean connection distance. Connection distance increased with increasing degree for both rich and feeder links while peripheral connections demonstrated no increase through the whole range of thresholds (Fig. **??**). Given this result, it is safe to say that increased coexpression for hub-related links is not determined by lower connection distance between them as the contrary is shown to be true.

Modular organization of the connectome could potentially drive gene coexpression patters, where more densely interconnected neurons within modules would also be more similar in their gene expression patters. Utilizing Louvain community detection algorithm (see *Methods*), four sparsely interconnected modules (M1, M2, M3, M4) were detected (Fig. 5). [[NOTE: maybe not worth characterizing modules, just say: coexpression within modules is not higher than coexpression between modules on one sentence and distributions? When hub neurons are excluded, M2 and M3 coexpression is higher than M1, but no difference between M2 and M3]]) While first three modules (M1-M3) were diverse in neuron types and comparable in size, the fourth module (M4) consisted of only 13 densely interconnected motor neurons. The majority (11 out of 13) of hubs were assigned to the M3 which spanned through the whole length of the worm body as pictured in Fig. 5. This particular module was also rich in motor neurons which is in line with our expectations as 10 out of 11 hub neurons in this module are directly responsible for forward and backward locomotion (command interneurons). The main goal of this analysis was to compare coexpression within and between modules in order to examine if neurons assigned to the same module are more genetically similar than neurons in other modules. Coexpression within modules did not exceed coexpression between modules, demonstrating that neurons that belong to the same module do not share any particular genetic similarity. Therefore, higher coexpression in rich and feeder links can not be attributed to the modular organization of the connectome despite most hubs being assigned to the same module.

Lineage distance between pairs of neurons presents another source of information about the genetic makeup of the *C. elegans* nervous system. NOTE: more details needed. While the relationship between cell lineage and gene expression was studied in other species [36, 37]. It has been shown in *C. elegans* that cells with identical fates can be formed by different gene regulatory pathways [38], however, to the best of knowledge, no lineage distance analysis was done in relation to the network properties of the *C. elegans* connectome. Implementing an analogous strategy as in our previous analysis of gene coexpression, we show that lineage distance between pairs of neurons increases with degree for feeder links while no significant change is observed for rich links (Fig. **??**). This finding suggests that the difference in lineage distance between hub and non-hub neurons increases with degree, therefore in a way confirming the relevance of this classification. Contrary, peripheral links manifest stably lower lineage distance through the range of degree thresholds wit (Fig. **??**). In addition, this shows that increase in gene coexpression with degree for rich links can not be attributed to their similarity in lineage.

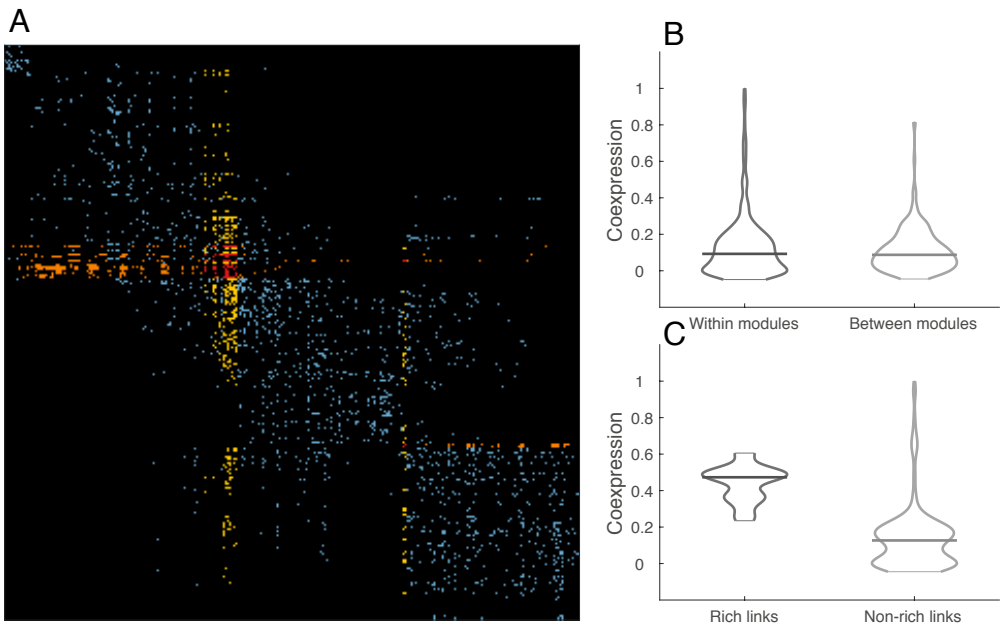


Fig 5. Plots for extras
B: Modular organisation.

Functional enrichment

To investigate which functional groups of genes contributed to the difference in transcriptional coupling between connected and unconnected neurons, we used a method of assigning each gene coexpression contribution (GCC) score that quantifies the contribution of each gene to the overall coexpression between neuron pairs (see Methods).

The majority of the significant ($P < 0.05$) GO categories are related to neuronal connectivity and communication, as listed in Table 1 (1st column: GO category). These findings are in line with our expectations as they confirm that genes that are more likely to be expressed in a pair of connected neurons are in fact related to neuronal connectivity and communication.

To investigate which functional groups of genes contributed to the increased coexpression for links involving hubs (rich and feeder) we again we used a method of scoring genes according to their contribution towards increased coexpression (see Methods).

Enrichment analysis show that the majority of the significant ($P < 0.05$) GO categories are related to glutamate signaling, neuronal connectivity and communication. It is line with the previous analysis where same GO categories were shown to be significant in differentiating connected from unconnected neurons.

These findings are in line with our expectations as high degree neurons are highly involved in signaling and the majority of hubs while being cholinergic, receive input from sensory neurons the majority of which is glutamatergic.

[NOTE: NOTE HERE ABOUT REDUCED LIST OF GENES IN EACH ANALYSIS (577 and 390 genes for connected and R/F respectively).]

Discussion

- First coexpression analysis in worm. Despite noisy gene expression data we get

Table 1. Enrichment.

GO category	Connected *vs* unconnected: pvalue	Rich and feeder *vs* peripheral
glutamate receptor signalling pathway	10^{-7}	10^{-5}
ionitropic glutamate receptor signalling	10^{-7}	10^{-5}
synaptic transmission, glutamatergic	10^{-7}	10^{-5}
neuron-neuron synaptic transmission	10^{-5}	0.00031
chemical synaptic transmission	0.00057	10^{-5}
anterograde trans-synaptic signalling	0.00057	10^{-5}
synaptic signalling	0.00057	10^{-5}
trans-synaptic signalling	0.00057	10^{-5}
cell-cell signalling	-	0.00086
cell surface receptor signalling pathway	-	0.00107

Table notes what GO categories were overrepresented in driving the difference between different types of links.

some insights into the genetic basis of connectivity on a neuronal level

Discussion text

Limitations

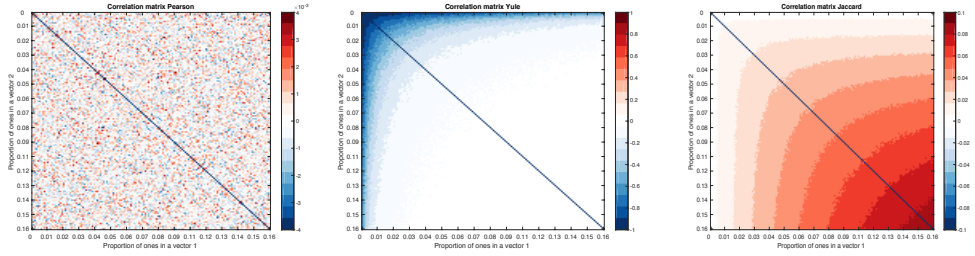
- Binary gene expression data
- No way of discriminating between missing data and the absence of expression
- only around 5 percent of genes in the genome available
- annotation problems: different qualifiers, loosing sensitivity/specificity if including too much or too little - need to balance

Conclusion

Conclusion text.

Supporting information

S1 Fig. Comparing different coexpression measures.



S2 Fig. Coexpression as a function of distance

S3 Fig. Coexpression for rich, feeder and peripheral links as a function of degree.

S5 Fig. Weighted and mixed rich club

S7 Fig. Average coexpression for rich, feeder and peripheral links as a function of degree. Top: degree distribution. Average gene coexpression for rich, feeder, and peripheral connections as a function of k, with the mean across all network links shown as a dashed black line and the topological rich club regime shaded grey. Circles indicate a statistically significant increase in gene coexpression in a given link type relative to the rest of the network (one-sided Welch's t-test; $P < 0.05$)

Acknowledgments

Thanks to Alex Fornito, for being a big deal.

References

1. Kaufman A, Dror G, Meilijson I, Ruppin E. Gene expression of Caenorhabditis elegans neurons carries information on their synaptic connectivity. PLoS computational biology. 2006;2(12):e167. doi:10.1371/journal.pcbi.0020167.
2. Baruch L, Itzkovitz S, Golan-Mashiach M, Shapiro E, Segal E. Using Expression Profiles of Caenorhabditis elegans Neurons To Identify Genes That Mediate Synaptic Connectivity. PLoS computational biology. 2008;4(7):e1000120. doi:10.1371/journal.pcbi.1000120.
3. Varadan V, Miller DM, Anastassiou D. Computational inference of the molecular logic for synaptic connectivity in C. elegans. Bioinformatics. 2006;22(14):e497–506. doi:10.1093/bioinformatics/btl224.
4. Fulcher BD, Fornito A. A transcriptional signature of hub connectivity in the mouse connectome. Proceedings of the National Academy of Sciences of the United States of America. 2016;113(5):1513302113–. doi:10.1073/pnas.1513302113.
5. Towlson EK, Vértes PE, Ahnert SE. The rich club of the C. elegans neuronal connectome. J Neurosci. 2013;33(15):6380–6387.
6. Shih CT, Sporns O, Yuan SL, Su TS, Lin YJ, Chuang CC, et al. Connectomics-Based Analysis of Information Flow in the Drosophila Brain. Curr Biol. 2015;.
7. Fulcher BD, Fornito A. A transcriptional signature of hub connectivity in the mouse connectome. Proc Natl Acad Sci USA. 2016;113(5):1435–1440.
8. van den Heuvel MP, Scholtens LH, de Reus MA. Topological organization of connectivity strength in the rat connectome. Brain Struct Funct. 2015; p. 1–18.
9. Zamora-López G, Zhou C, Kurths J. Cortical Hubs Form a Module for Multisensory Integration on Top of the Hierarchy of Cortical Networks. Front Neuroinf. 2010;4:1.
10. de Reus MA, van den Heuvel MP. Rich Club Organization and Intermodule Communication in the Cat Connectome. J Neurosci. 2013;33(32):12929–12939.

11. Harriger L, van den Heuvel MP, Sporns O. Rich Club Organization of Macaque Cerebral Cortex and Its Role in Network Communication. PLoS ONE. 2012;7(9):e46497.

12. van den Heuvel MP, Sporns O. Rich-Club Organization of the Human Connectome. J Neurosci. 2011;31(44):15775–15786.

13. van den Heuvel MP, Kahn RS, Goñi J, Sporns O. High-cost, high-capacity backbone for global brain communication. Proc Natl Acad Sci USA. 2012;109(28):11372–11377.

14. Vértes PE, Rittman T, Whitaker KJ, Romero-Garcia R, Váša F, Kitzbichler MG, et al. Gene transcription profiles associated with inter-modular hubs and connection distance in human functional magnetic resonance imaging networks. Philosophical Transactions of the Royal Society B. 2016;371(1705):735–769. doi:10.1098/rstb.2015.0362.

15. Bullmore E, Sporns O. The economy of brain network organization. Nature Reviews Neuroscience. 2012;13(5):336–49. doi:10.1038/nrn3214.

16. White JG, Southgate E, Thomson JN. The structure of the nervous system of the nematode Caenorhabditis elegans. ... Trans R Soc Lond B Biol .... 1986;.

17. Harris TW, Antoshechkin I, Bieri T, Blasiar D, Chan J, Chen WJ, et al. WormBase: a comprehensive resource for nematode research. Nucleic Acids Research. 2009;38:D463–D467.

18. Pereira L, Kratsios P, Serrano-Saiz E, Sheftel H, Mayo AE, Hall DH, et al. A cellular and regulatory map of the cholinergic nervous system of C. elegans. eLife. 2015;4:e12432.

19. Sulston JE, Horvitz HR. Post-embryonic cell lineages of the nematode, Caenorhabditis elegans. Developmental biology. 1977;56(1):110–56.

20. Sulston JE, Schierenberg E, White JG, Thomson JN. The embryonic cell lineage of the nematode Caenorhabditis elegans. Developmental biology. 1983;100(1):64–119.

21. Varshney LR, Chen BL, Paniagua E, Hall DH, Chklovskii DB. Structural properties of the Caenorhabditis elegans neuronal network. PLoS computational biology. 2011;7(2):e1001066. doi:10.1371/journal.pcbi.1001066.

22. Varier S, Kaiser M, Koh W, Meuli R, Honey C. Neural Development Features: Spatio-Temporal Development of the Caenorhabditis elegans Neuronal Network. PLoS Computational Biology. 2011;7(1):e1001044. doi:10.1371/journal.pcbi.1001044.

23. Schröter M, Paulsen O, Bullmore E. Micro-connectomics: probing the organization of neuronal networks at the cellular scale. Nat Rev Neurosci. 2017; p. 1–16.

24. Rubinov M, Sporns O. Complex network measures of brain connectivity: Uses and interpretations. NeuroImage. 2010;52(3):1059–1069. doi:10.1016/j.neuroimage.2009.10.003.

25. Warrens MJ. On Association Coefficients for 2x2 Tables and Properties That Do Not Depend on the Marginal Distributions. Psychometrika. 2008;73(4):777–789. doi:10.1007/s11336-008-9070-3.

26. Krienen FM, Yeo BTT, Ge T, Buckner RL, Sherwood CC. Transcriptional profiles of supragranular-enriched genes associate with corticocortical network architecture in the human brain. Proc Natl Acad Sci USA. 2016;113(4):E469–78.

27. Pantazatos SP, Li X. Commentary: BRAIN NETWORKS. Correlated gene expression supports synchronous activity in brain networks. Science 348, 1241-4. bioRxiv. 2016; p. 079202.

28. Henderson JA, Robinson PA. Relations Between the Geometry of Cortical Gyrification and White-Matter Network Architecture. Brain Conn. 2014;4(2):112–130.

29. Horvát S, Gămănuț R, Ercsey-Ravasz M, Magrou L, Gămănuț B, Van Essen DC, et al. Spatial Embedding and Wiring Cost Constrain the Functional Layout of the Cortical Network of Rodents and Primates. PLoS Biol. 2016;14(7):e1002512.

30. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: Tool for The Unification of Biology. NATURE GENETICS. 2000;25(1):25–29. doi:10.1038/75556.

31. Gillis J, Mistry M, Pavlidis P. Gene function analysis in complex data sets using ErmineJ. Nature protocols. 2010;5(6):1148–59. doi:10.1038/nprot.2010.78.

32. Zoubarev A, Hamer KM, Keshav KD, Luke Mccarthy E, Santos JRC, Van rossum T, et al. Gemma: A resource for the reuse, sharing and meta-analysis of expression profiling data. Bioinformatics. 2012;28(17):2272–2273. doi:10.1093/bioinformatics/bts430.

33. Roberts JA, Perry A, Lord AR, Roberts G, Mitchell PB, Smith RE, et al. The contribution of geometry to the human connectome. NeuroImage. 2016;124(Pt A):379–93. doi:10.1016/j.neuroimage.2015.09.009.

34. Goulas A, Uylings HBM, Hilgetag CC. Principles of ipsilateral and contralateral cortico-cortical connectivity in the mouse. Brain Struct Funct. 2016;252:1–15.

35. Krienen FM, Yeo BTT, Ge T, Buckner RL, Sherwood CC. Transcriptional profiles of supragranular-enriched genes associate with corticocortical network architecture in the human brain. Proceedings of the National Academy of Sciences. 2016;113(4):E469–78. doi:10.1073/pnas.1510903113.

36. Cui Q, Yu Z, Purisima EO, Wang E, Li X, Schwarz P, et al. MicroRNA regulation and interspecific variation of gene expression. Trends in Genetics. 2007;23(8):372–375. doi:10.1016/j.tig.2007.04.003.

37. Kluger Y, Tuck DP, Chang JT, Nakayama Y, Poddar R, Kohya N, et al. Lineage specificity of gene expression patterns. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(17):6508–6513. doi:10.1073/pnas.0401136101.

38. Liu X, Long F, Peng H, Aerni SJ, Jiang M, Sánchez-Blanco A, et al. Analysis of Cell Fate from Single-Cell Gene Expression Profiles in C. elegans. Cell. 2009;139(3):623–633. doi:10.1016/j.cell.2009.08.044.

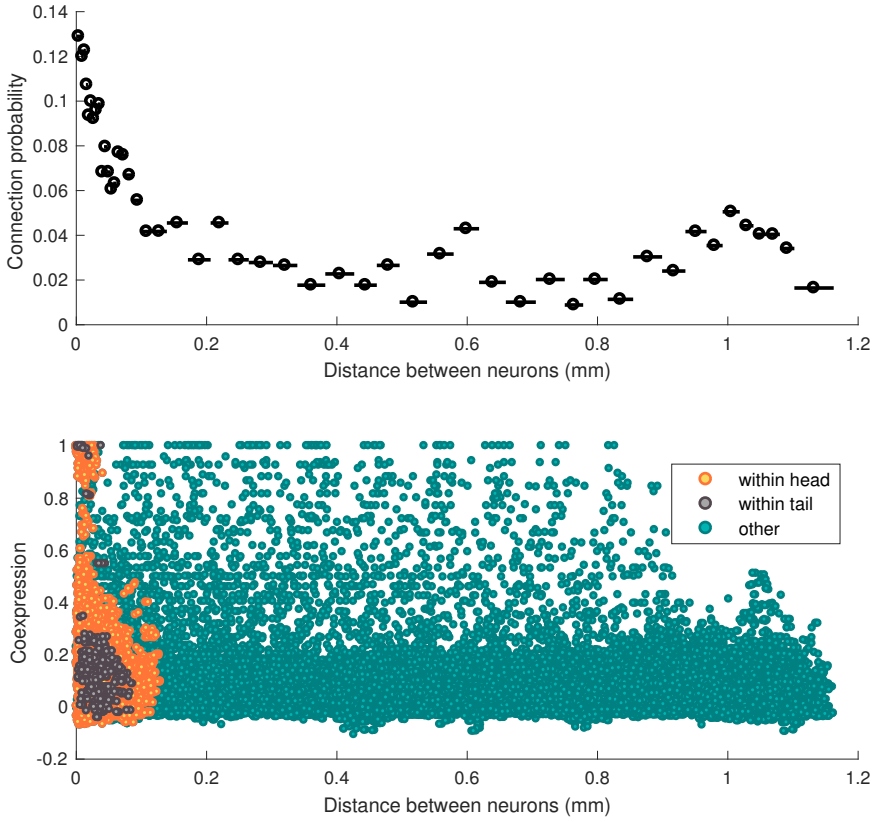
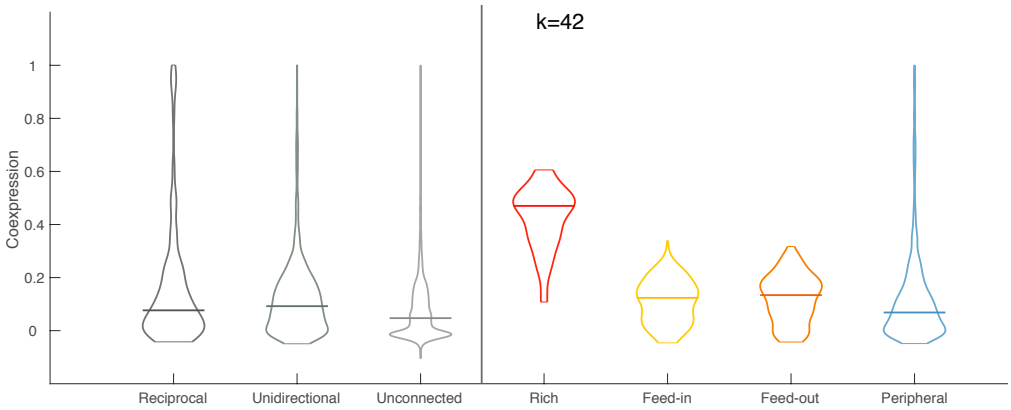
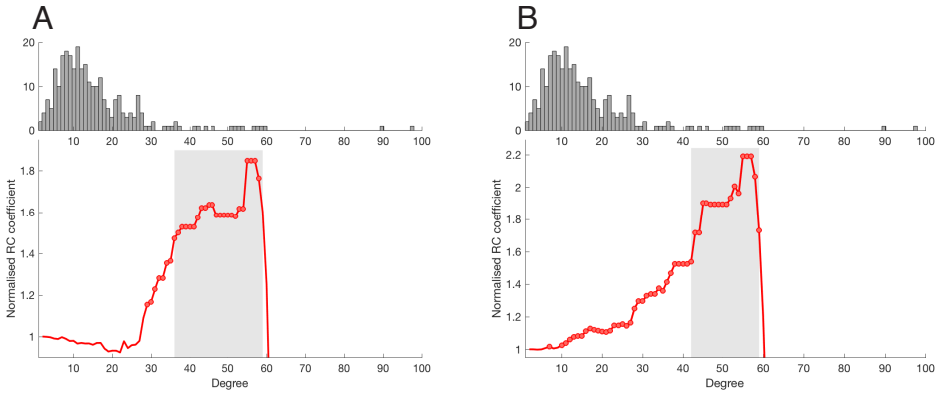
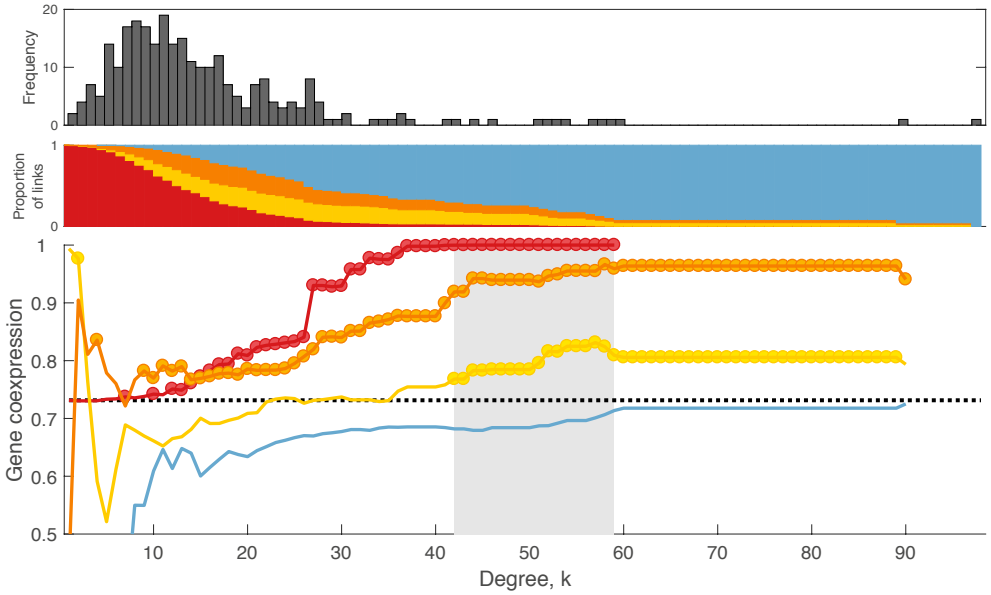


Fig 6. Separation distance affects connection probability but not gene coexpression A The connection probability is estimated in 50 equiprobable distance bins, shown as circles (bin centers) and lines (bin extent). A decreasing relationship is evident up to a separation distance of approximately 0.2 mm (driven by tail-tail and head-head connections), with a longer-range decrease up to ≈0.5 mm due to body-body connections. The increase at large distances (≈ 1 mm) is due to head-tail connections, which also show a decay with distance, from ≈ 1 mm through to ≈1.2 mm. connections between body and head and body and tail showed no clear distance dependence. [[should we show relationships by type here? It's in the Evernote: https://www.evernote.com/l/AF5h5EZ0dxZBU6RiKJk3BtNFs36tLrE4h0U]]. **B** Pearson gene coexpression, r_ϕ, is plotted as a function of separation distance for all pairs of neurons, labeled according to within-head (orange), within-tail (gray), and all other connections (aqua). Unlike data from macroscopic mammalian brains of mouse [7] and human ([26]), there is no clear spatial dependence in these neuronal data. [[TODO: Did we check this is really true when we actually look across distance bins?]]


k=42
Coexpression
1
0.8
0.6
0.4
0.2
0
Reciprocal
Unidirectional
Unconnected
Rich
Feed-in
Feed-out
Peripheral


A
B
Normalised RC coefficient
Degree


Frequency
Proportion of links
Gene coexpression
Degree, k