# Search for computational modules in the *C. elegans* brain

Markus Reigl[1], Uri Alon[2], and Dmitri B. Chklovskii[1]

[1]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724

[2]Weizmann Institute, Israel

Are there stereotypical multi-neuron computational modules in the *C. elegans* nervous system? We attempt to answer this question by performing a systematic statistical analysis of the *C. elegans* wiring diagram. We search for multi-neuron inter-connectivity patterns that are significantly over-represented, i.e. more common than expected based on the statistics of smaller inter-connectivity patterns. Our statistical analysis reveals over-represented patterns (or motifs) containing two, three and four neurons. No over-represented five-neuron motifs were found. We suggest that the discovered motifs may serve as computational modules performing stereotypical functions. These modules need to be investigated further both anatomically and physiologically.

## INTRODUCTION

There is little doubt that neurons are elementary building blocks of the nervous system (Cajal, 1999). It is less clear, however, whether multi-neuron modules (smaller than invertebrate ganglia or vertebrate nuclei and cortical columns) can be meaningfully defined, either anatomically (Peters and Sethares, 1996) or physiologically (Abeles and Gerstein, 1988). The existence of such multi-neuron modules would greatly simplify the description of the nervous system structure and function. An example of such simplification can be found in electrical engineering. An electronic circuit is often represented in terms of modules such as operational amplifiers, logical gates, and memory registers rather than as a wiring diagram showing each transistor, resistor and diode. However, unlike electrical engineers who designed these modules themselves, neurobiologists did not design the brain, and evolution rarely leaves records of its experimentation. Therefore, if multi-neuron modules have indeed evolved they need to be discovered.

In this paper, we search for anatomically defined multi-neuron modules by performing statistical analysis of the *C. elegans* wiring diagram, obtained from 3D electron microscopic reconstructions (White et al., 1986). The idea is that if a certain multi-neuron module performs some stereotypical function it may appear in the nervous system repeatedly (Milo et al., 2002). Therefore, search for multi-neuron connectivity patterns (or motifs) that appear more often than by "chance" (as represented by the null-hypothesis defined below) may yield these multi-neuron modules. Of course, there may be functionally important modules that appear infrequently and would be missed by our analysis. In the electronic circuit analogy, our approach would discover logical gates in a processor-wiring diagram but not a rectifier in a power supply, which is essential but appears only once.
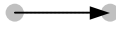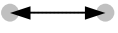
We begin our search by considering doublets (or neuronal pairs, $N=2$) and then increase the number $N$ of neurons per motif sequentially up to quintuplets *($N=5$)*. At each step we incorporate the $N$-neuron statistics into the null-hypothesis used for the search of $N+1$-neuron motifs (Milo et al., 2002). For the 2-neuron motif search we construct the following ensemble of random matrices as the null-hypothesis. Each random matrix preserves the number of incoming and outgoing connections for each neuron but swaps the synaptic partners. By comparing doublet counts in this ensemble to those in *C. elegans*, we find that bi-directionally connected doublets (out of three possible doublet classes) are over-represented. For the 3-neuron motif search, we use an ensemble of random matrices that preserves the number of bi-directional (as well as uni-directional) connections per neuron. Several triplet classes (out of all thirteen connected classes) are over-represented. For the 4-neuron motif search, we construct an ensemble of random

matrices that preserve the numbers of triplets in addition to previously listed constraints. By comparing quadruplet counts in this ensemble to *C. elegans*, we find that several quadruplet classes (out of all 199 connected classes) are over-represented. These motifs may correspond to modules that perform stereotypical functions and need to be investigated.

## RESULTS

We perform our statistical analysis on three available datasets (Achacoso and Yamamoto, 1992; White et al., 1986) of the *C. elegans* nervous system represented in connectivity matrices. Although two neurons may be connected by more than one synapse (multiplicity of connection), the small size of the data set compels us to use the binary representation of these connections (connected or unconnected). In order to obtain binary connectivity matrices, we threshold the multiplicity of connections at various values $\Theta$. Such procedure is justified by the intuition that more than a single synapse is necessary for an observable physiological effect of one neuron on another. Since we don't know the physiologically relevant count of synapses, we repeat our calculation for $1 \le \Theta \le 7$.

**1. Doublets ($N$=2).** We classify all possible doublets (or pairs) of the *C. elegans* neurons into three classes: unconnected, uni-directionally connected, bi-directionally connected, and compare the number of doublets in each class to that expected in a random network (Table 1). The random network ensemble consists of matrices that preserve the numbers of incoming and outgoing synapses for each neuron but not the identities of the partners. The choice of this random matrix ensemble (rather than Erdős-Rényi random graph, for example) is motivated by the observation that the distribution of the connection number per neuron is far from Poisson.

| Motif: $i =$ | | 1 | 2 | 3 |
|---|---|---|---|---|
| Actual # of doublets in *C. elegans* | $c_{2,i}(A)$ | 13,902 | 435 | 28 |
| Avg. # of doublets in random networks | $\overline{c_{2,i}(R_k)}$ | 13,878.8 | 481.3 | 4.8 |
| Standard deviation | $\sigma(R_k)$ | 2.048 | 4.096 | 2.048 |
| Relative difference between A and $R_k$ | $\Delta_{2,i}(A, \overline{R_k})$ | 0.002 | -0.096 | 4.802 |
| | $p$ | <= 0.001 | <= 0.001 | <= 0.001 |

**Table 1:** Doublet counts in the actual *C. elegans* network *A* are significantly different from the mean counts in the random matrix ensemble $R_k$. Counts shown are for dataset 2, threshold $\Theta$=3, number of random matrices $n$=1000. Other datasets and thresholds give similar results.

Although the data in Table 1 indicate that the actual number of doublets in each class is significantly different from the mean, we assign most significance to the reciprocally connected doublets (motif 3). This assignment is based on two considerations. First, minor but inevitable mistakes in the datasets of the *C. elegans* wiring diagram can affect the number of doublets in each class. Therefore, only the differences greater than the typical error in the number of doublets are significant. Assuming that the error is proportional to the doublet count, we introduce a relative excess measure

$$\Delta_{N,i}(A, \overline{R}) = \frac{c_{N,i}(A) - \overline{c_{N,i}(R_k)}}{c_{N,i}(R_k)}, \quad 2 \le N \le 5, \ i \text{ is the index of the } N\text{-neuron motif}. \qquad \textbf{(1)}$$

The high value of this measure for reciprocally connected neurons indicates the significance of this finding compared to the unconnected or uni-directional motif. Second, if the total number of doublets in a network is kept the same then the over-representation of one class inevitably leads to the under-representation of another. Thus it is reasonable to assign most significance to the class that is relatively most over-represented. We repeated the above calculations for other datasets and threshold values and concluded that bi-directionally connected doublets are consistently significant over-represented.

The over-representation of reciprocally connected doublets has been reported previously on the qualitative level (White et al., 1986) and explained (Durbin, 1987) as a consequence of correlation between adjacency and connectivity of neurons. Interestingly, an over-representation of bi-directionally connected doublets as also found in mammalian neocortex (Markram et al., 1997; Sjostrom et al., 2001), although Erdős-Rényi graph was used as a null-hypothesis. This suggests that the over-represented motifs may be evolutionarily conserved.

**2. Triplets ($N=3$).** We classify all connected triplets in the *C. elegans* wiring diagram into 13 classes and count the number of triplets in each class. We compare the actual numbers of triplets in each class to the null-hypothesis random matrix ensemble defined as follows. In order to include the observed over-representation of reciprocally connected doublets, we fix the number of bi-directional connections per neuron as well as the number of uni-directional connections per neuron. Based on above arguments we plot the relative number of triplets in each class.
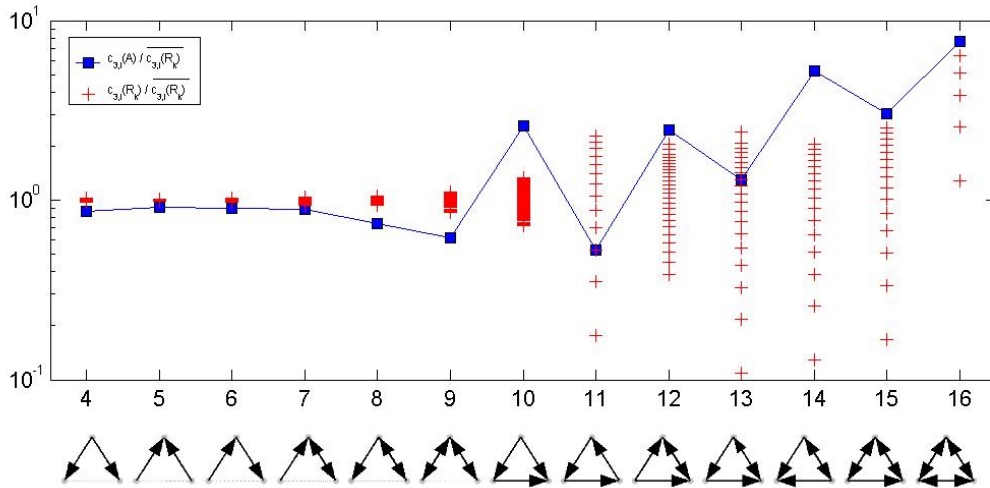


**Figure 1:** Triplet counts for the actual network (dataset 2, threshold $\Theta=2$) and for random connectivity matrices relative to the mean counts for the random matrix ensemble ($n=1000$). All matrices in this ensemble preserve the counts of uni- and bi- directional connections per neuron.

For a given threshold we find that several motifs are significantly different from the mean of the random matrix ensemble, Figure 1. To ensure that these results are robust we repeat the calculation for different values of the threshold (within the biologically plausible range) Figure 2.

The significance of the obtained results may be characterized by the probability of obtaining a given number of triplets by chance (asterisks in Figure 2 reflect a *p*-value < 0.001) as defined by the chosen random matrix ensemble. However, there are two problems with this measure. First, as discussed above, this measure is not robust against inevitable errors in the existing data sets. Second, because we perform multiple hypotheses testing (search for over-representation in multiple classes simultaneously), the significance level is over-estimated. To overcome these problems, we introduce a different significance measure $P_m$, which is based on the relative excess $\Delta_{N,i}\left(A, \overline{R}\right)$ and the following procedure. We want to

calculate the probability to find a random matrix in our ensemble to the same relative excess value. To obtain that we calculate the relative excess values of the observed number of triplets in each single random matrix compared to the average number in the other matrices of our ensemble. The probability that this value is greater than $\Delta_{N,i}\left(A, \overline{R}\right)$ for any motif $i$ will define the multiple hypotheses testing corrected $p$-value $P_m$, Figure 3.
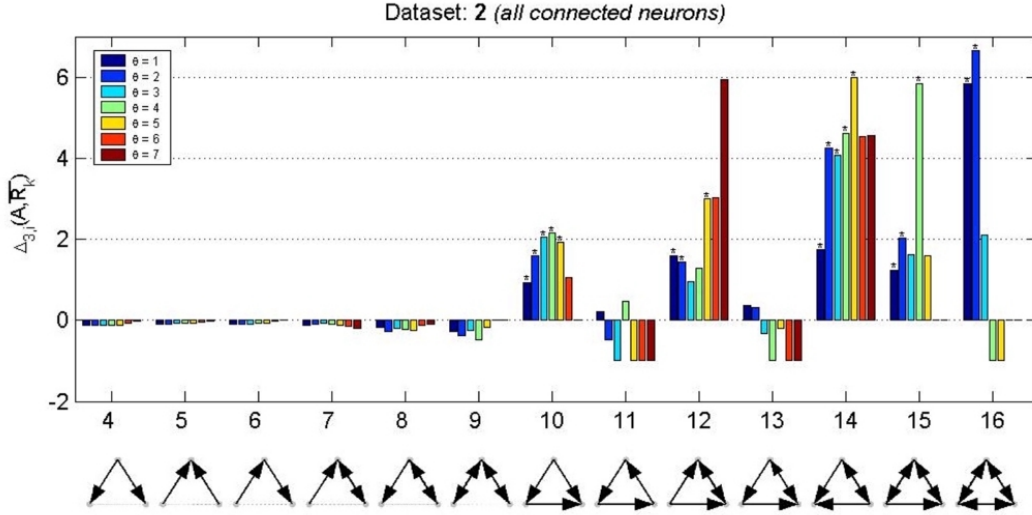


**Figure 2:** Relative excess in triplet counts for the various thresholds $\Theta$ (dataset 2). All random matrices in the ensemble preserve the number of uni- and bi-directional connections per neuron, $n$=1000, (*): $c_{3,i}\left(A\right) > c_{3,i}\left(R_k\right)$ for all random matrices in the ensemble.

One of the most consistently over-represented motifs is the feedforward loop (triplet class #10), previously noticed in *C. elegans* (White et al., 1986) and other networks (Milo et al., 2002).

In addition to the feedforward loop, we find that two other (both symmetric) motifs are consistently over-represented: triplet class #12 and triplet class #14 (Figure 3).
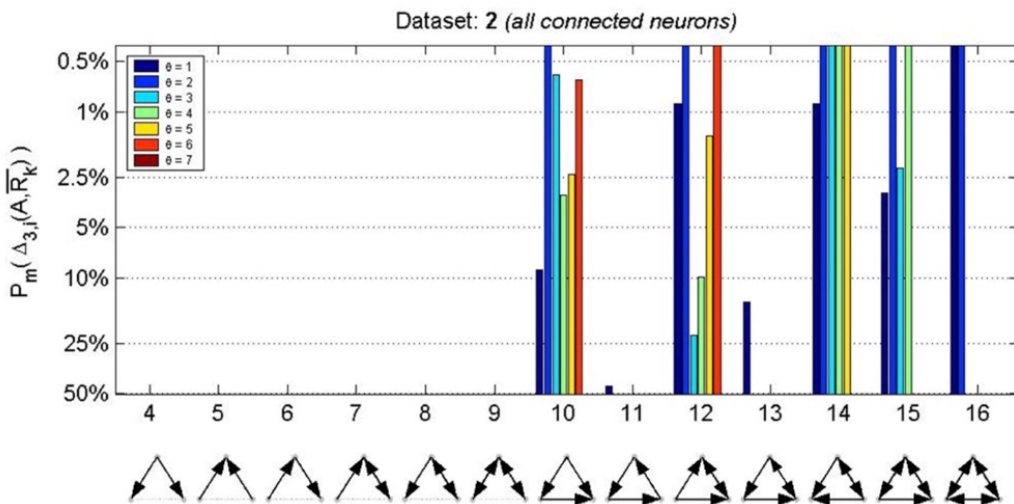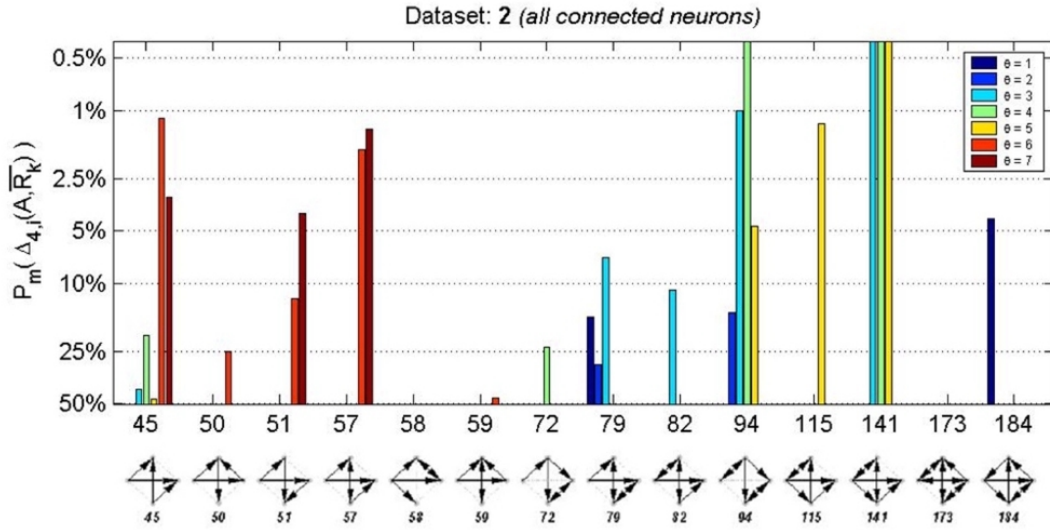


**Figure 3:** Significance of quadruplet over-representation. Multiple hypothesis testing corrected $p$-values for the triplet over-representation (datasets 2, $n$=1000). To suppress motifs with a rare appearance we calculated $P_m$ values only if the count $c_{3,i}(A)$ of this triplet in *C. elegans* was at least 5.

**3. Quadruplets (*N*=4).** We classify all connected 4-neuron pattern into 199 classes and count the number of quadruplets in each class. Then we compare the actual counts of quadruplets in each class to the mean counts of quadruplets in a random matrix ensemble. Random matrices preserve the numbers of uni-directional and bi-directional connections per neuron and, in addition, the numbers of triplets. We define over-represented motifs as the ones whose multiple hypothesis testing corrected significance values are

$$P_m\left(\Delta_{N,i}\left(A,\overline{R_k}\right)\right) <= 0.1 \quad \text{and} \quad c_{4,i}(A) >= 5 \text{ for over-represented or } \overline{c_{4,i}\left(R_k\right)} >= 5 \text{ for under-represented.}$$

With the addition of the last 2 conditions we suppress the appearance of motifs with a very small quadruplet count as over-represented or under-represented. We plot the multiple hypothesis testing corrected significance values in Figure 4.

We find two consistently over-represented motifs: quadruplet class #45 and #51 (Milo et al., 2002).



**Figure 4:** Significance of quadruplet over-representation. Multiple hypothesis testing corrected *p*-values of quadruplet over-representation (datasets 2, *n*=1000). Motifs shown satisfy the selection criteria: $P_m < 0.1$ for at least one $\Theta$ and $c_{4,i}(A) >= 5$.

**4. Quintuplets (*N*=5).** We classify all connected quintuplets into 9364 classes (out of 9608 motifs total) and count the actual number of quintuplets in each class. We compare these counts with the mean of the random matrix ensemble. The random matrices preserve the numbers of uni- and bi-directional connections per neuron and, in addition, keep the numbers of all triplets and quadruplets in a 10% range of the actual network. We do not find any significantly over-represented quintuplets, $P_M$. This could happen for two reasons. First, there may be no significantly over-represented quintuplets with a given number of quadruplets. Second, the *C. elegans* data set may be too small to detect statistical outliers.

**DISCUSSION**

Our motif search algorithm is different from previous attempts to find modules. For example, traditional clustering approaches look for the subsets of nodes, which are connected with their own subset stronger than with other subsets. In our algorithm, we consider all the connections between the neurons in a motif (unlike (Achacoso and Yamamoto, 1992),who considered only some connections within the motif) but ignore the connections with neurons outside the motif. One could question the expediency of ignoring multiple possible inputs to the neurons in a module since those inputs could influence the operation of that module. To counter this, we point out that the nervous system is capable of performing many different functions under different circumstances and neurons active in one case may be silent in another. Therefore, in any particular case, many of the anatomical inputs to the module may remain silent and can

be safely ignored. This speculation may be verified experimentally by simultaneous monitoring of neuronal activity in different neurons.

Since our algorithm requires a relatively complete knowledge of the wiring diagram it was natural to choose the *C. elegans* nervous system, which represents the biggest wiring diagram mapped entirely. Unfortunately, *C. elegans* has some disadvantages when it comes to the interpretation of the results: the polarity of a synapse (excitatory vs. inhibitory) in *C. elegans* is often unknown, electrophysiological investigations are still difficult in *C. elegans* (Lockery and Goodman, 1998), and the whole network contains only 302 neurons limiting the statistical power of the approach. Yet we hope that the recent technological developments (Kerr et al., 2000) will eliminate the first two disadvantages and allow functional analysis of the discovered modules. Moreover, we expect that our results have implications for understanding nervous system structure and function beyond *C. elegans*. The modules we identify in *C. elegans* may be evolutionarily conserved, and once identified can be searched for in other systems. Finally, the algorithm itself can be applied to other networks (Milo et al., 2002) once they become available.

As in any other theoretical analysis, we made several simplifications. For example, we thresholded the multiplicity of connections, ignored the polarity of the synapses and the existence of gap junctions. Yet, our results are robust to the inclusion of these factors in the future. This because if we find an over-represented class, it will remain over-represented even if divided into smaller sub-classes. It would be interesting to see whether the inclusion of the above factors will reveal specific over-represent sub-classes.

## REFERENCES

Abeles, M., and Gerstein, G. L. (1988). Detecting spatiotemporal firing patterns among simultaneously recorded single neurons. J Neurophysiol *60*, 909-924.

Achacoso, T. B., and Yamamoto, W. S. (1992). Neuroanatomy of C. elegans for computation (Boca Raton, FL, CRC Press).

Cajal, S. R. y. (1999). Texture of the Nervous System of Man and the Vertebrates, Vol 1 (New-York, Springer).

Durbin, R. M. (1987) Studies on the Development and Organisation of the Nervous System of Caenorhabditis elegans, Ph.D>, Cambride University, Cambridge.

Kerr, R., Lev-Ram, V., Baird, G., Vincent, P., Tsien, R. Y., and Schafer, W. R. (2000). Optical imaging of calcium transients in neurons and pharyngeal muscle of C. elegans. Neuron *26*, 583-594.

Lockery, S. R., and Goodman, M. B. (1998). Tight-seal whole-cell patch clamping of Caenorhabditis elegans neurons. Methods Enzymol *293*, 201-217.

Markram, H., Lubke, J., Frotscher, M., Roth, A., and Sakmann, B. (1997). Physiology and anatomy of synaptic connections between thick tufted pyramidal neurones in the developing rat neocortex. J Physiol *500 ( Pt 2)*, 409-440.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network motifs: simple building blocks of complex networks. Science *298*, 824-827.

Peters, A., and Sethares, C. (1996). Myelinated axons and the pyramidal cell modules in monkey primary visual cortex. J Comp Neurol *365*, 232-255.

Sjostrom, P. J., Turrigiano, G. G., and Nelson, S. B. (2001). Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. Neuron *32*, 1149-1164.

White, J. G., Southgate, E., Thomson, J. N., and Brenner, S. (1986). The structure of the nervous system of the nematode Caenorhabditis elegans. Philos Trans R Soc Lond B Biol Sci *314*, 1-340.