

Supplementary Material Online for

**Towards a genome-wide transcriptogram: the  
*Saccharomyces cerevisiae* case**

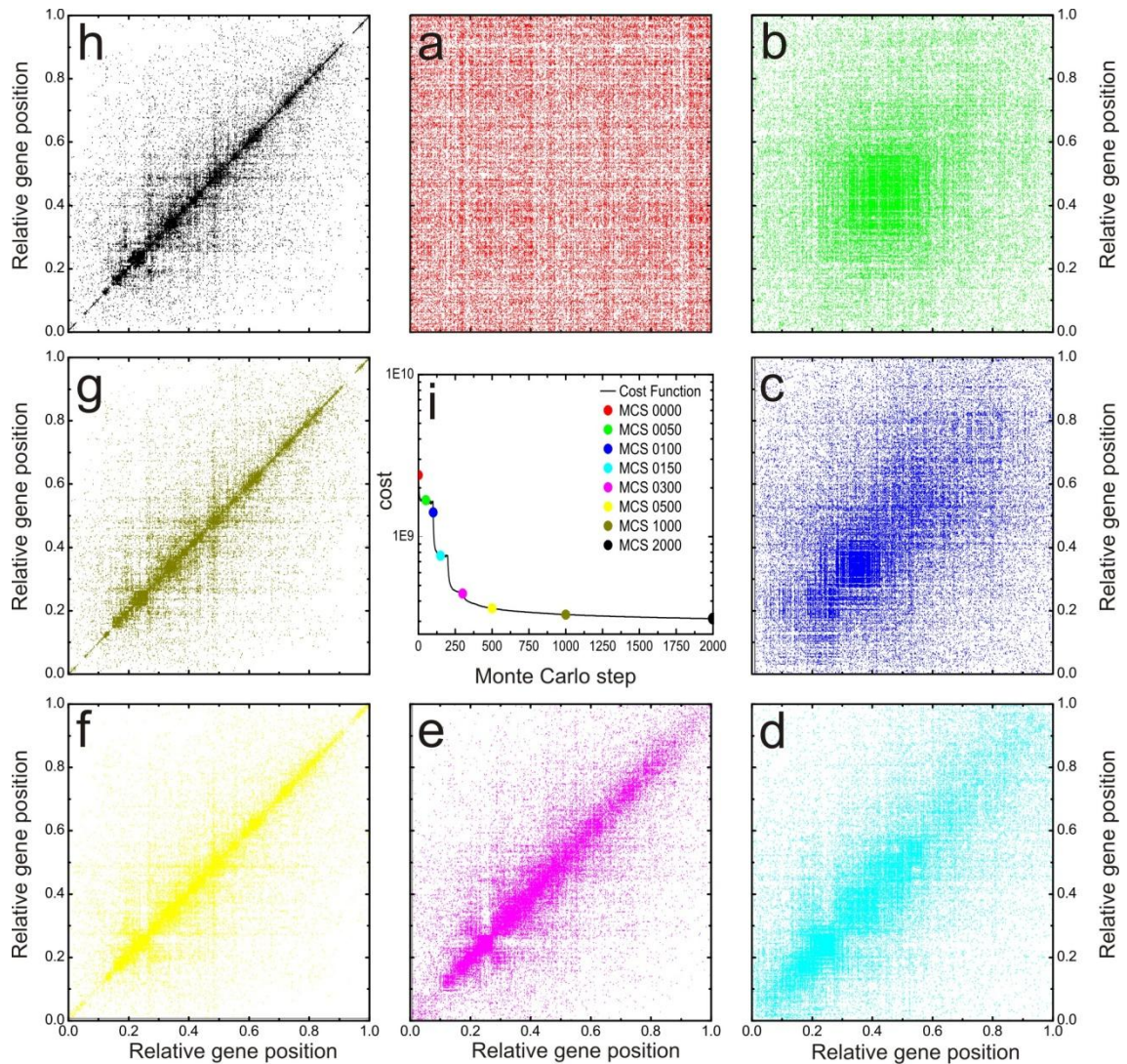
José Luiz Rybarczyk-Filho<sup>2</sup>, Mauro A. A. Castro<sup>1,2</sup>, Rodrigo J. S. Dalmolin<sup>3</sup>, José C. F.  
Moreira<sup>3</sup>, Leonardo G. Brunnet<sup>2</sup> & Rita M. C. de Almeida<sup>1,2</sup>

*National Institute of Science and Technology for Complex Systems<sup>1</sup>, Instituto de Física<sup>2</sup>,  
Departamento de Bioquímica<sup>3</sup>, Universidade Federal do Rio Grande do Sul, Av. Bento  
Gonçalves, 9500, 91051-970 C.P. 15051, Porto Alegre, Brazil.*

Correspondence to: Rita M. C. de Almeida ([rita@if.ufrgs.br](mailto:rita@if.ufrgs.br)).

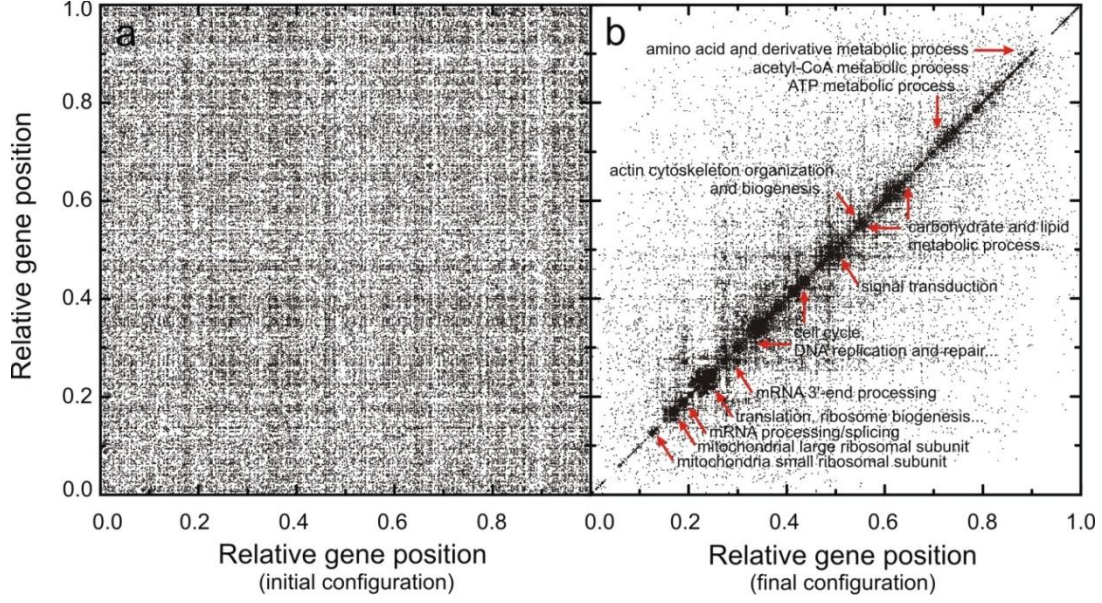
Keywords: Gene network, Interactome, Transcriptome, Transcriptogram, Protein-protein interaction matrix, Monte Carlo dynamics.

## 1. Cost function versus Monte Carlo steps



**Figure S1:** (a-h) Snapshots of the interaction matrix relative to *Saccharomyces cerevisiae* as the ordering is obtained. (i) Plot of the cost function versus Monte Carlo steps. The color of the matrices indicates the ordering stage as presented in (i).

Matrix (a) is the starting point, built on basis of a randomly ordered gene list. The ordering algorithm proceeds by randomly choosing a pair of genes and swapping their positions on the ordering. A new interaction matrix is produced for this new ordering and its cost is recalculated using cost function. (b) corresponds to 50 MCS, (c) 100 MCS, (d) 150 MCS, (e) 300 MCS, (f) 500 MCS, (g) 1000 MCS, (h) 2000 MCS. Plot (i) presents the evolution of cost function, which is minimized throughout the process.



**Figure S2:** (a) matrix for Random and (b) CFM orderings. The arrows indicate the interaction dots clusters associated to different GO terms. This same layout has already been used by Strong et al.[4] to display the Dendrogram ordering of *Mycobacterium tuberculosis*.

## 2. Dendrogram

The links between proteins of the protein network can be used to define a topological overlap for all pairs of proteins (1). Using  $M_{i,j} = M_{j,i} = 1$  if proteins associated to genes in positions  $i$  and  $j$  interact, and zero otherwise, the topological overlap  $O_T(i, j)$  between genes  $i$  and  $j$  is defined as

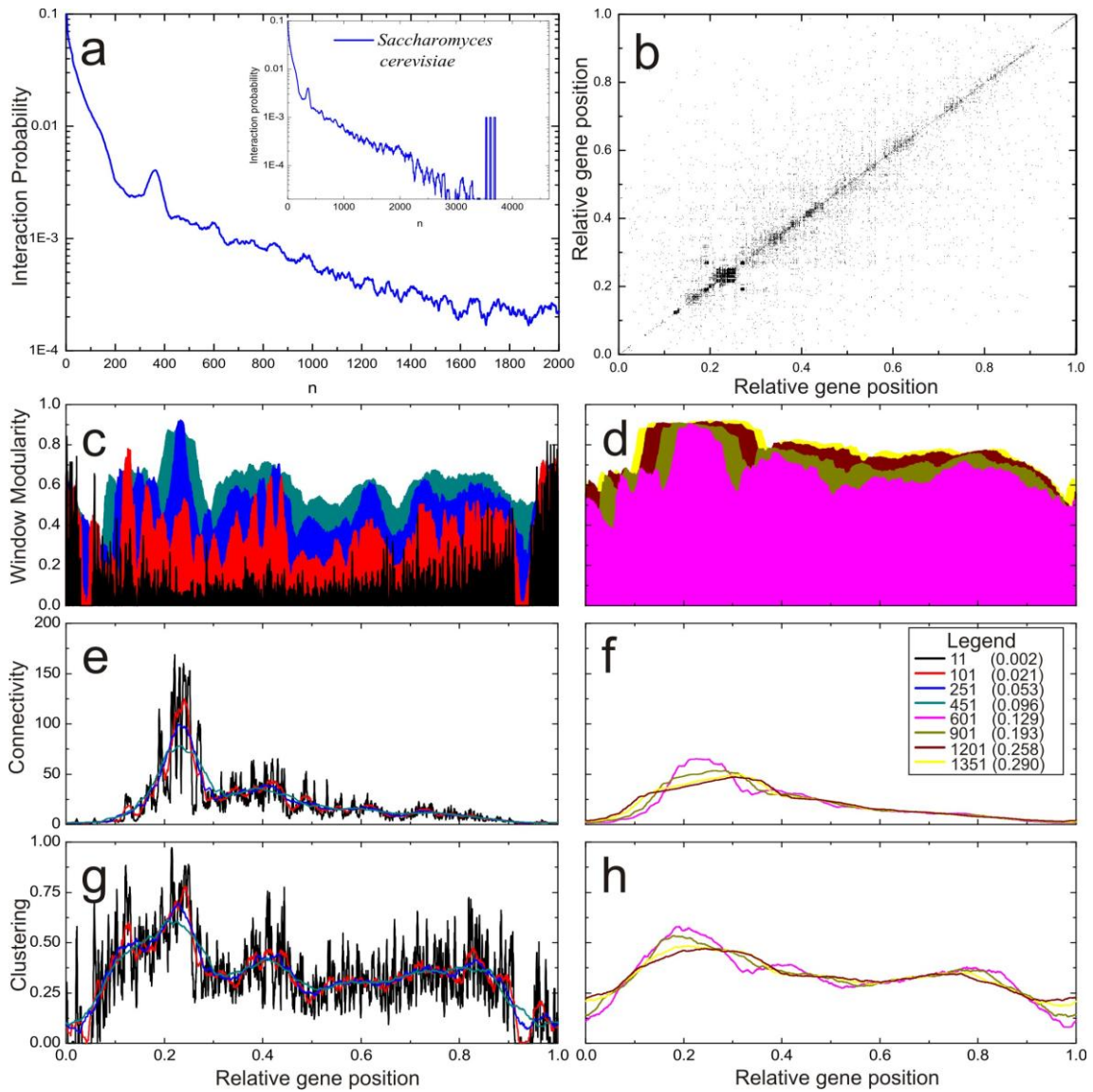
$$O_T(i, j) = \frac{M_{i,j} + \sum_{l=1}^N M_{i,l} M_{l,j}}{\min(k_i, k_j)}, \quad (\text{S1})$$

where  $k_i$  is the connectivity of the  $i^{\text{th}}$  of the network. The dendrogram ordering considers the unweighted average linkage clustering algorithm (2,3) using the topological overlap. Starting from an  $N \times N$  overlap matrix, this method merges the pair of nodes  $(i, j)$  with largest value for the overlap, producing a  $(N - 1) \times (N - 1)$  matrix, where the columns and lines corresponding to nodes  $i$  and  $j$  are replaced by a single column and a single line, corresponding to the overlap values with a new node, noted by  $\{i, j\}$ . The new overlap values of node  $\{i, j\}$  with all other, remaining nodes  $l$  of the original matrix are given as

$$O_T(\{i, j\}, l) = \frac{n_i O_T(i, l) + n_j O_T(j, l)}{n_i + n_j}, \quad (\text{S2})$$

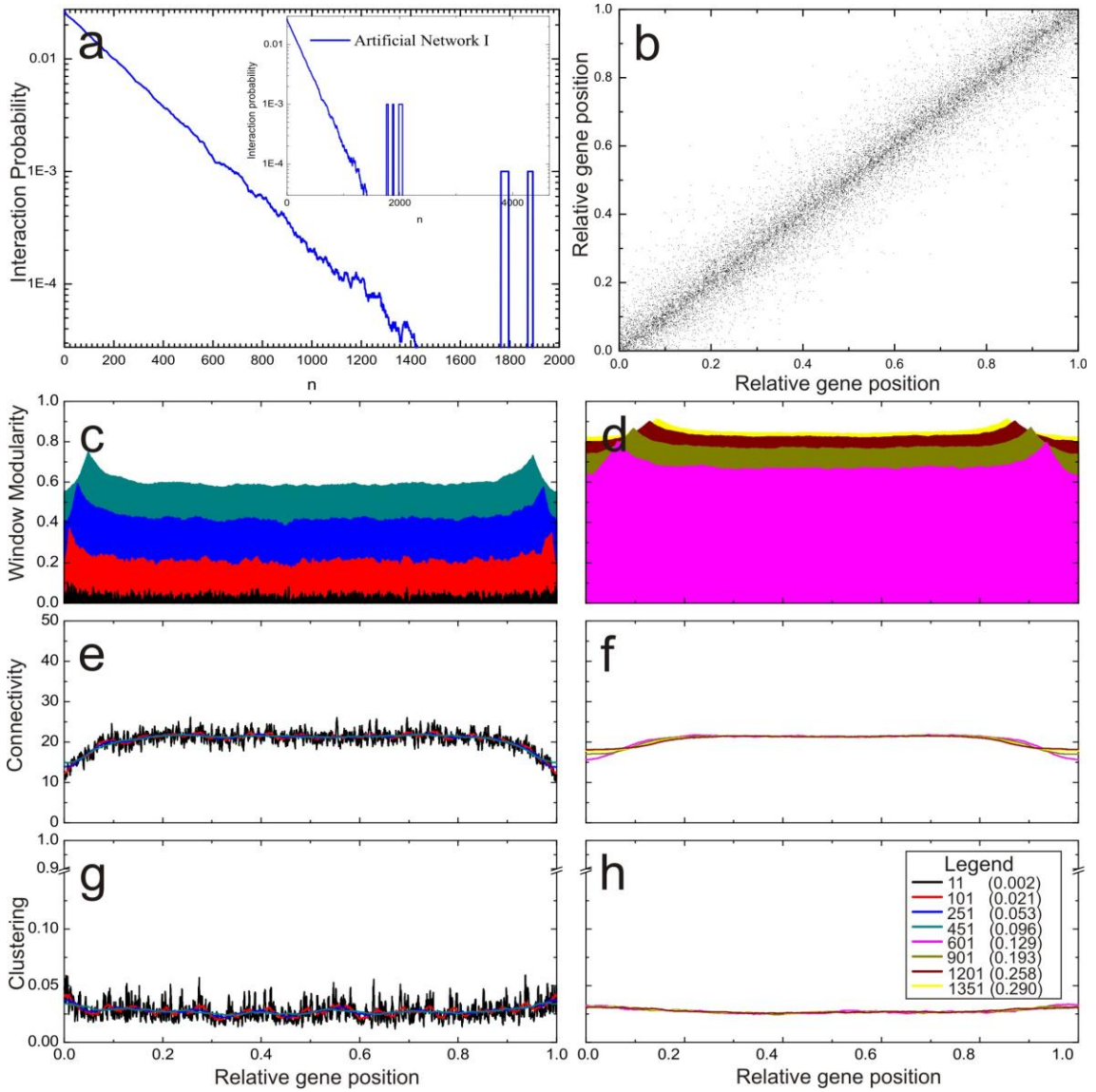
where  $n_i$  and  $n_j$  is the number of components in nodes  $i$  and  $j$ . Successive repetitions of the above procedure reduces the overlap matrix to a single value. In case of two pairs having equal maximal values, the algorithm merges the first located pair. Observe that, as the ordering of two nodes that have been merged is arbitrary, the resulting final ordering is not unique.

### 3. *Saccharomyces cerevisiae* network: different window values.



**Figure S3.** (a) Interaction probability between two genes as a function of their distance on the CFM ordering for the *Saccharomyces cerevisiae*. (b) Interaction matrix. (c and d) Window modularity using different window sizes for CFM. Different window sizes may evince modules of different sizes. (e and f) Connectivity projected on the ordering using different window sizes for window averages.

#### 4. Artificial networks

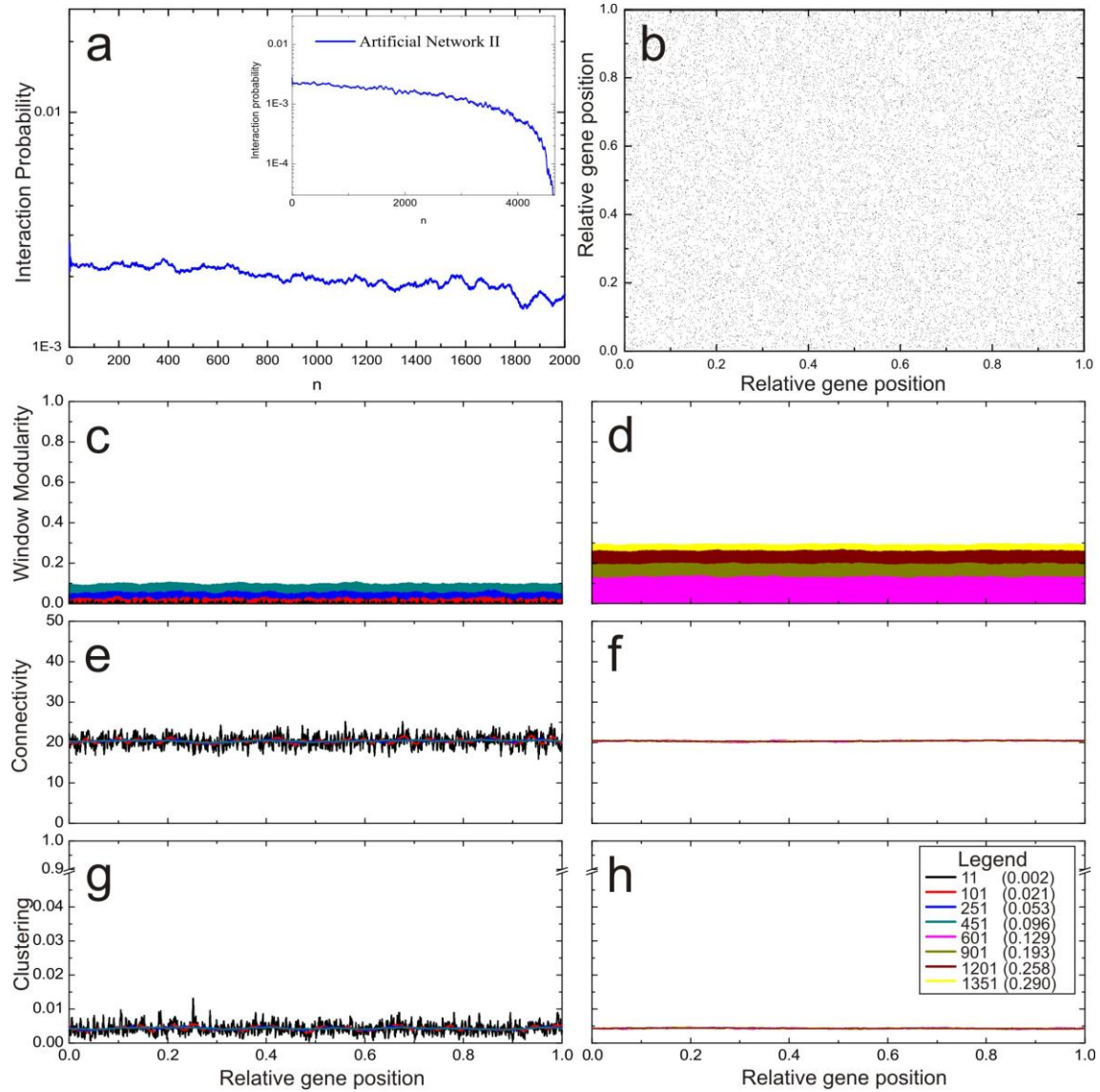


**Figure S4.** (a) Interaction probability between two genes as a function of their distance on the for an artificial network. (b) Interaction matrix. (c and d) Window modularity using different window sizes. Different window sizes may evince modules of different sizes. (e and f) Connectivity projected on the ordering using different window sizes for window averages.

The network presented in Figure S3 has been built as follows. Consider  $N = 4655$  nodes that are connected with probability  $p$ , given as :

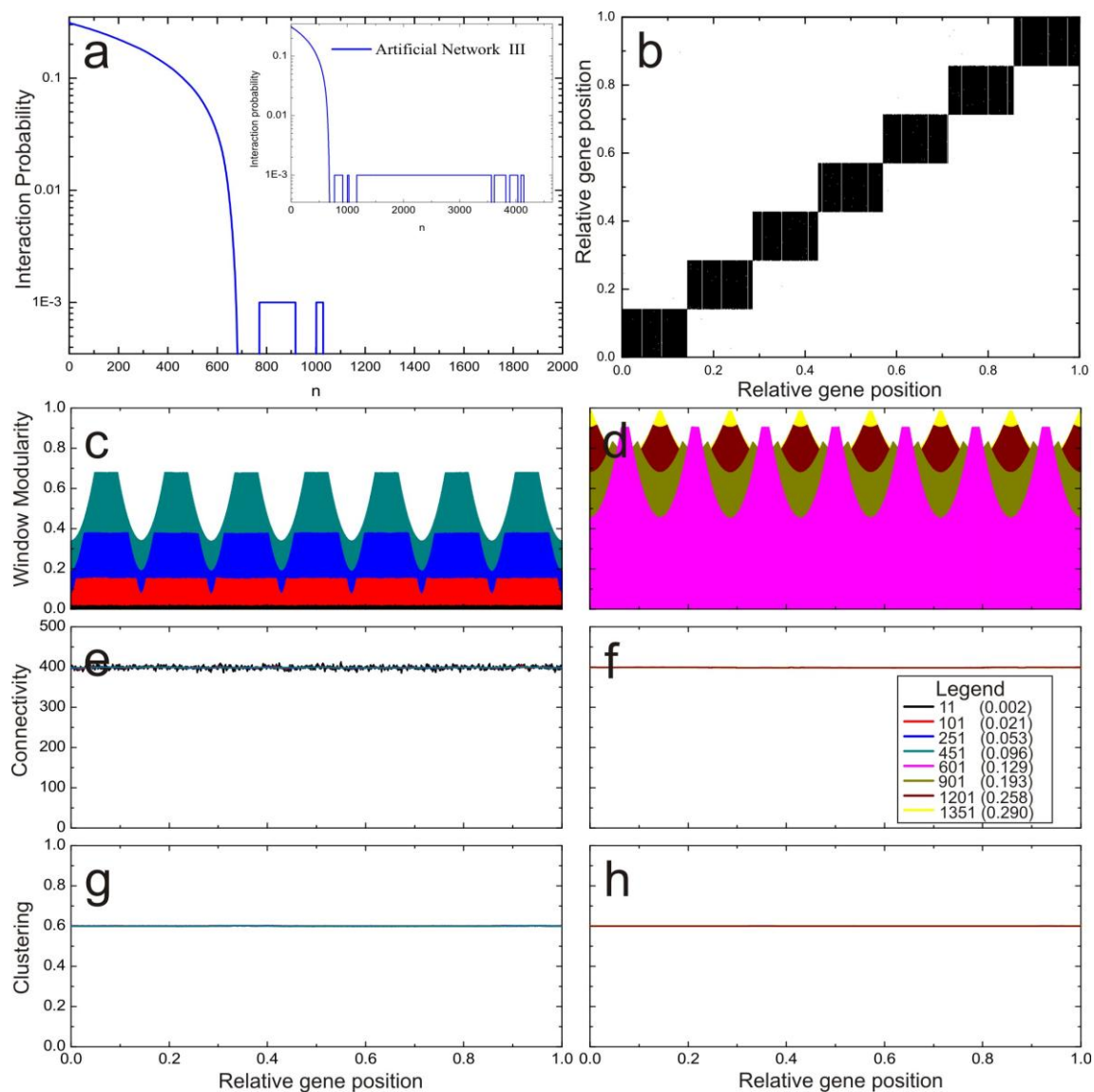
$$p = 0.05 e^{-\left(\frac{j-i}{2150}\right)}. \quad (\text{S4})$$

This interaction probability between two nodes decays exponentially with the distance on between the nodes on the axes (See Figure S3a). It obviously does not produce any module, yielding flat modularity landscapes.



**Figure S5.** (a) Interaction probability between two genes as a function of their distance on the for an artificial network. (b) Interaction matrix. (c and d) Window modularity using different window sizes. Different window sizes may evince modules of different sizes. (e and f) Connectivity projected on the ordering using different window sizes for window averages.

The network presented in Figure S4 has been built as follows. Consider  $N = 4655$ , choose at random nodes, and connect them. Altogether 94830 interactions were selected.

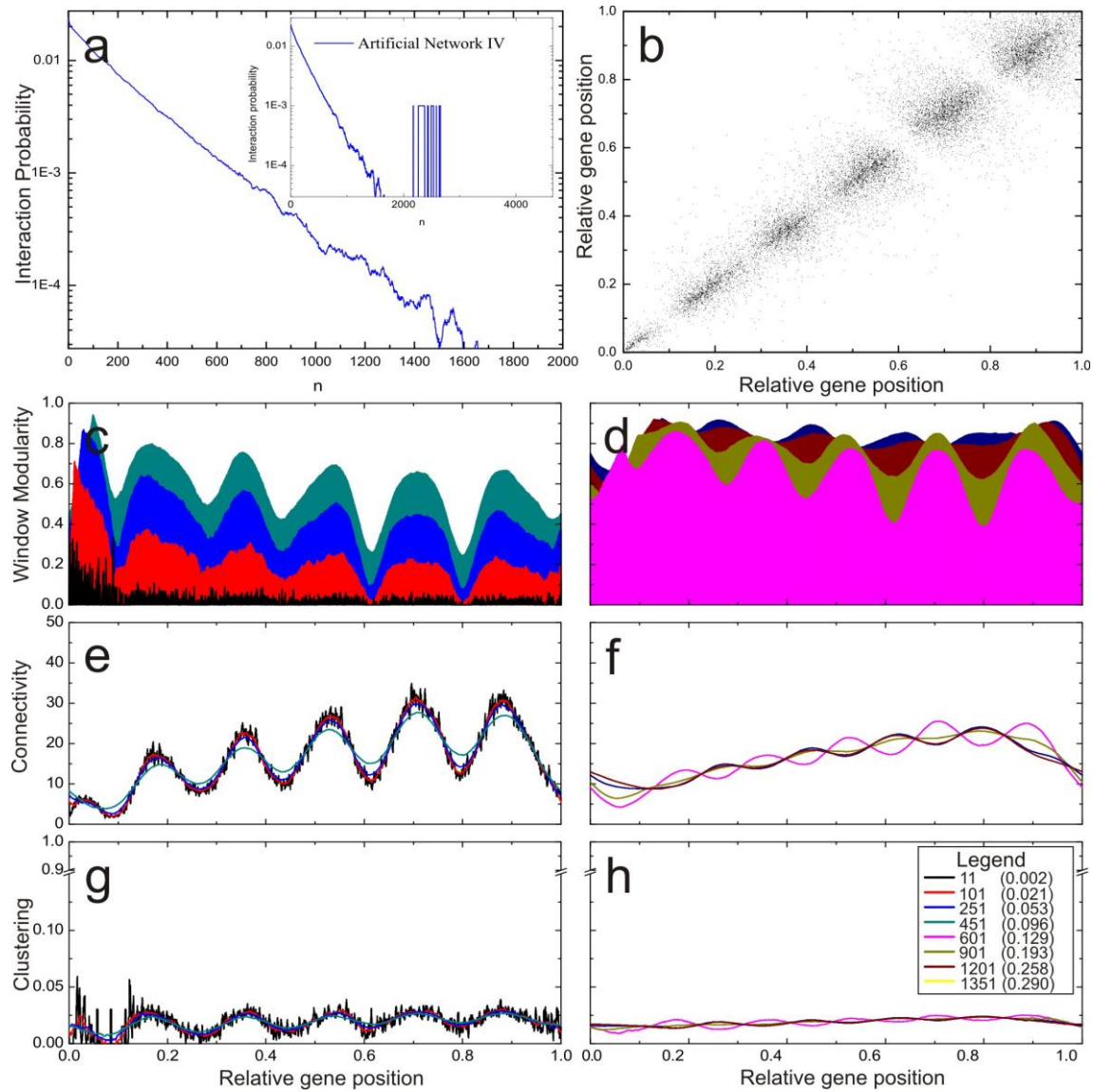


**Figure S6.** (a) Interaction probability between two genes as a function of their distance on the for an artificially modular network. (b) Interaction matrix. (c and d) Window modularity using different window sizes. Different window sizes may evince modules of different sizes. (e and f) Connectivity projected on the ordering using different window sizes for window averages.

The network presented in Figure S5 has been built as follows. Consider  $N = 4655$ , build 7 moduli dividing the nodes in seven groups with 665 nodes each. Assign random interactions between the nodes belonging to the same group, with probability 0.8. Observe that the window modularity present flat maxima when window size is smaller than the modulus (S5 c). Also, when the window size is larger than the moduli, the maxima are shifted from the moduli centers towards the moduli border. When the



window size is roughly the double of modulus size, the window modularity maxima are located at the moduli borders.



**Figure S7.** (a) Interaction probability between two genes as a function of their distance on the for an artificial network. (b) Interaction matrix. (c and d) Window modularity using different window sizes. Different window sizes may evince modules of different sizes. (e and f) Connectivity projected on the ordering using different window sizes for window averages.

The network presented in Figure S6 has been built as follows. Consider  $N = 4655$ , and assign an interaction between nodes  $i$  and  $j$  with probability  $p$  given as

$$\begin{aligned}
p = & \\
& 0.0125 \exp \left[ -\frac{j-i}{\sqrt{\frac{16(i+j)(1-i-j)}{4655^2}}} \right] \sin^2 \left[ \frac{7\pi(i+j)\sqrt{2}}{9310} \right] + \\
& 0.0063 \exp \left[ -\frac{j-i}{\sqrt{\frac{32(i+j)(1-i-j)}{4655^2}}} \right] \cos^2 \left[ \frac{3\pi(i+j)\sqrt{2}}{9310} \right] + \\
& 0.065 \exp \left[ -\frac{j-i}{\sqrt{\frac{8(i+j)(1-i-j)}{4655^2}}} \right] \cos^2 \left[ \frac{4\pi(i+j)\sqrt{2}}{9310} \right]. \quad (S5)
\end{aligned}$$

This network has modules of different sizes together with an interaction probability that decays exponentially with the distance between two nodes on the axes. Here it is also possible to observe the shift of window modularity maxima as the window size increases (**S6 c** and **d**).

## References

- [1] Ravasz,E., Somera,A.L., Mongru,D.A., Oltvai,Z.N. and Barabasi,A.L. (2002) Hierarchical Organization of Modularity in Metabolic Networks. *Science*, **297**, 1551-1555.
- [2] J. Sokal, P. Sneath, *Numerical Taxonomy*, Freeman, San Francisco (1973).
- [3] M. B. Eisen, P. T. Spellman, P. O. Brown, D. Botstein (1998). *Proc Natl Acad Sci USA* **95**, 12863-8
- [4] Strong,M. *et al.* (2003).Visualization and interpretation of protein networks in Mycobacterium tuberculosis based on hierarchical clustering of genome-wide functional linkage maps. *Nucl. Acids Res.* **31**, 7099-7109