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

José Luiz Rybarczyk[2], Mauro A. A. Castro[1,2], Rodrigo J. S. Dalmolin[3], José C. F. Moreira[3], Leonardo G. Brunnet[2] & Rita M. C. de Almeida[1,2]

*National Institute of Science and Technology for Complex Systems[1], Instituto de Física[2], Departamento de Bioquímica[3], 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).

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

**A genome modular classification that associates cellular processes with modules could lead to a method for quantifying the variations in gene expression levels caused by different cellular stages or conditions: the *transcriptogram*, a powerful tool for assessing cell performance, would be at hand. Here we present a computational method that order genes on a line and clusters strongly interacting genes, defining functional modules associated with gene ontology terms. The starting point is a list of genes and a matrix specifying their interactions, available at large gene interaction databases. Considering the *Saccharomyces cerevisiae* genome we produced a succession of plots of gene transcription levels for a fermentation process. These plots discriminate the fermentation stage the cell is going through and may be regarded as the first versions of a *transcriptogram*. This method is useful for extracting information from cell stimuli/responses experiments, and may be applied with diagnostic purposes to different organisms.**

Determining the precise role of each biochemical agent in every relevant process occurring in the cell has turned out to be an elusive task. Local methods that focus on one or just a few components of such reactions have met their limitations and we are far from an understanding of cell metabolism that would allow preventing, for example, cancer or aging in a complex organism. In a global approach, each gene plays its role in a way that prevents the analysis of its performance as an isolated unit. However, the local analysis did pay off, although not completely, suggesting that there must be a third, intermediate route between the local and the global points of view. From this intermediate point of view, the genome is regarded as a modular network,[1-4] where genes in a module interact more strongly with one another as compared to their interactions with genes external to the module. Such a modular classification would provide an estimate of the relative influence of external genes on internal genes of a module, making possible approximate models for the module dynamics.

Genes have been classified in different ways, generating different databases. For example, similarity in nucleotide sequences generates the orthologous relationships between genes from different species.[5] However, groups of orthologs do not necessarily represent functional modules. Another way of classifying genes is by the cellular metabolic network in which they participate, as presented by KEGG´s data base[6]. A metabolic pathway is defined as a series of biochemical reactions that is responsible for modifying some set of molecules, but the rationale for identifying the pathway is a flux of mass,[7] which is also different from identifying a functional gene module. On the other hand, it would be extremely interesting to *i)* identify functional gene modules inside a genome and then *ii)* identify which modules participate in each metabolic pathway.

The point is how to produce a comparison method for the expression profiles of cells in different states (healthy vs. ill, different stages of cell cycle, etc.), yielding what we call a *transcriptogram*. We have spotted two main difficulties in producing this method. The first one has to do with the fact that identifying functional groups is not straightforward (see Asur *et. al.* and references therein).[8] Here we present a method for ordering a list of genes using the computational physics method known as Monte Carlo dynamics. The aim is to cluster on a line the strongly interacting genes. The second difficulty has to do with the wild fluctuations found in the expression profiles. Usually, fluctuations may be smoothed out by averaging the profiles values over some neighborhood, which requires a previous definition for a neighborhood. The genome ordering we propose here defines a neighborhood and a metric that correlates the mutual influence with the distance between two genes on the list. In this sense, genes that are far apart interact less strongly.

The starting point for the method is a randomly enumerated list of genes and the corresponding matrix specifying the interaction between the genes. Here we consider gene interaction as the physical and/or functional association presented by any pair of

gene products. This body of information has been produced along the years by different researchers around the world and is magnificently organized and available at STRING database (http://string.embl.de).[9] We retrieved all protein-protein interactions described in that database inferred by "experimental" and "database" evidences for the organism *Saccharomyces cerevisiae*. Our final list comprises 4655 genes and 47415 interactions.

A similar approach for genome ordering, although not for the transcriptogram, has been proposed by Barabási and collaborators.[2,10] Their ordering algorithm considers the topological overlap matrix, which measures the sharing of common neighbors by two genes. The algorithm then proceeds as a dendogram construction, where the proximity measure is taken as the topological overlap. However, this method leads to an ordering which may divide large modules, as we shall discuss in what follows.

For an ordered list with *N* genes, the interaction data may be organized in an $N \times N$ matrix $M$, where the matrix elements, $M_{ij}$, are 1 or 0 depending on whether or not the $i^{th}$ and $j^{th}$ genes on the list interact. The result is a symmetric matrix of zeroes and ones with a null diagonal. We propose here an ordering algorithm that favors the proximity of interacting genes by minimizing a virtual energy assigned to each ordering, given as

$$E = \sum_{i=1} \sum_{j=1} d_{ij} \begin{array}{l} \left| M_{i,j} - M_{i+1,j} \right| + \left| M_{i,j} - M_{i-1,j} \right| \\ + \left| M_{i,j} - M_{i,j+1} \right| + \left| M_{i,j} - M_{i,j-1} \right| \end{array} , \qquad (1)$$

where |..| stands for the positive value of the difference between the neighboring sites of the interaction matrix and $d_{ij}$ is proportional to the distance from the point *(i,j)* to the diagonal, that is, $d_{ij} = |i - j|$ .This virtual energy increases with the number of interfaces between black and white dots on the matrix and increases further when these interfaces are far from the diagonal. Details are given in Methods section.

Figure 1 presents the interaction matrices for *Saccharomyces cerevisiae,* taken for the initial random gene ordering (1a), after ordering following Barabási[2] prescription

(1b), and following the Virtual Energy Minimization (VEM) algorithm (1c). In these figures a black dot located at $(i, j)$ indicates that $M_{ij} = 1$. All three configurations present the same number of black dots. The randomly ordered gene list distributes uniformly the dots over the whole matrix surface. After Dendogram ordering, some black dots are concentrated on the main diagonal with some large clusters, while after VEM ordering the black dots concentrate even nearer the diagonal, leaving the top left and bottom right corners free of black dots. These two corners represent interactions between genes located far apart on the list. Furthermore, the black dot clusters far from the diagonal, which are present in the interaction matrix representing the Dendogram ordering, indicate a strong interaction between clusters of genes located far apart on that ordering.

To quantitatively characterize the orderings, we have defined window modularity for each gene as the ratio between the number of interactions that link any two genes in an interval (window) of size *w* of the gene list, centered at gene, and the number of interactions involving at least one gene in that window.[11] We considered periodic boundary conditions to deal with genes near the ends of the list. Window modularity strongly depends on the window size $w$. For example, for a window containing all genes of an ordered list, window modularity is one for every gene, while it decreases when the window is smaller than the cluster size, due to interactions connecting genes inside the window with genes outside the window that still belongs to the cluster. Also, genes that link different clusters present low modularity. On Figures 1d-l window modularity is represented by the gray landscapes. There we have chosen $w$=251. Plots for other values of window size are presented in Supplementary Materials On Line. Observe that window modularity in both Dendogram and VEM orderings presents well defined peaks and valleys, indicating interacting modules. The random list presents a very low modularity for all genes. Although the peaks are similar in height, in the VEM ordering the valleys are deeper and the number of peaks separated by deep

valleys is smaller. In fact, since there are valleys with different depth, the peaks may be hierarchically defined: smaller clusters composing larger clusters.

To assess the biochemical meaning of the orderings we have projected on the ordering information regarding the Biological Process terms from the Gene Ontology (GO) Database.[12] We used *DAVID* Bioinformatics Resources,[13] as described in Materials and Methods, to obtain the GO terms that best represent each window modularity peak. For each GO term we calculated the fraction of genes in windows of 251 sites that belong to the term, producing profiles that are smooth and depend on the ordering, presented on Figures 1 d-l. For the randomly ordered list, no peaks are seen and no information can be gathered from these plots. For the ordering obtained using Dendogram algorithm, some peaks appear, but the ontology terms are not as concentrated as for the VEM algorithm. Also, the VEM ordering successively locates classes of GO functions in an order that reproduces cell cycle: from right to left we first find functions associated with energy metabolism, followed by cell morphogenesis and cell communication, then GO terms related to vesicle transport and Golgi vesicle transport, then DNA replication and repair, and finally GO terms associated with RNA production and translation.

The orderings may be further characterized using the connectivity k(i) and the clustering coefficient c(i) of the $i^{th}$ gene on the ordering.[14] The interaction matrix gives information on which pairs of genes interact. The connectivity k(i) of the $i^{th}$ gene on the ordering is defined as the number of genes with which it interacts. On its turn, the clustering coefficient c(i) is defined as the fraction of existing links between any two of the k(i) neighbors of the $i^{th}$ gene, relative to the maximum possible number k(i)(k(i)-1)/2 of such links. Figure 2a and 2b presents the connectivity and clustering coefficient profiles for, respectively, the Dendogram and VEM orderings, obtained by taking the average of these quantities over windows of 251 sites. The connectivity profile of the VEM ordering shows that *i)* genes with higher connectivity are more concentrated than

the Dendogram ordering, presenting a high peak around the window modularity maximum at the region located at 0.2-0.3 on the horizontal axis, associated with the translation GO term, while the poorly connected genes are found at the ordering extremities; and *ii)* the clustering coefficient also decreases to very small values at the ordering extremities.

We now consider the evolutionary origins of these genes. By assessing information available on the Clusters of Orthologous Groups (COG) Database (updated and extended version),[9] we have built for both VEM and Dendogram orderings, a profile for the fraction of genes in a 251-gene window that belongs to a COG, presented in Figures 2c and 2d. There we also show the profiles giving the fraction of these genes in a 251-gene window whose COG presents genes from Eukarya, Bacteria and Archea, from Eukarya and Bacteria only, from Eukarya and Archea only, and from Eukarya only. Details are described in Materials and Methods. The light gray landscape represents the window modularity to guide the eye. The majority of genes with COG in (black line) correspond to COGs with genes from Eukarya, Archea and Bacteria, which in general can be associated with early evolutionary roots, but not necessarily to a common ancestor, due to the high probability of horizontal gene transfer.[15] While the profile associated with COGs presenting only Eukarya and Bacteria genes (orange line) for the Dendogram ordering spreads over all ordering, the VEM ordering shows a broad valley in the region 0.1-0.6 of the horizontal axis, associated with the GO biological process terms (see Figures 1f and 1i) translation, mRNA splicing, ribosome biogenesis and assembly, RNA processing, RNA elongation, translation from polymerase II promoter, DNA metabolic process, DNA replication, and DNA repair. Furthermore, the green profiles for the VEM ordering, associated with COGs with genes from only Eukarya, present a broad peak roughly in the same region as the valley of the Bacteria and Eukarya only COG profile. Above this peak, it is possible to spot two other higher peaks in the VEM green profile, the first around 0.3 on the horizontal axis, at the

regions associated with GO terms translation from polymerase II promoter, RNA processing, RNA elongation, mRNA polyadenylation, and mRNA 3'-end processing; and the second peak around 0.6 on the horizontal axis, associated with vesicle mediated transport and vesicle Golgi transport. That is, the regions where the green profile is higher in the VEM ordering majorly contemplate biological processes typical of Eukaryotes. The pink lines present the profiles for the Archea and Eukarya only COGs. However low these profiles are, the VEM algorithm allows the identification of three Archea peaks localized on the ordering at the regions associated with the GO biological process terms (see Figures 1f and 1i) translation, mRNA splicing, RNA processing, and ribosome biogenesis and assembly.

The most relevant difference between the results of Dendogram and VEM algorithm has to do with the correlation between interaction and proximity on the orderings at long ranges; at short range the difference between the two orderings are less marked. The evidence is given by Figures 1b and 1c, showing more black points far from the diagonal in the interaction matrix of the Dendogram ordering. We calculated the number of black dots per length on parallel lines, labeled by a parameter $d$, to the main diagonal. These lines comprehend links between genes sites located $d$ genes apart. Figures 2e and 2f shows the black dots concentration as a function of $d$. After a rapid decrease until $d=500$, the Dendogram ordering yields an interaction concentration that stabilizes at $10^{-3}$. The VEM algorithm, on the other hand, keeps decreasing at an exponential rate, indicating much less long range interactions. For very short ranges, however, Figure 2F shows that the Dendogram algorithm concentrates more the interacting genes, up to 20 genes distant; between 20 and ~600 genes apart the VEM concentrates more, between 600 an 1000 they present roughly the same black dot density and, after that, the VEM ordering presents exponentially lower dot densities. We interpret this exponentially decaying dot density profile in the VEM ordering as a correlation between interaction and localization of the genes on the ordering. This

correlation gives place to adequate window averages, allowing the smoothing out of wild fluctuations in the diverse profiles.

From now on we concentrate in analyzing the results for the VEM ordering. We sliced the VEM ordering in seven pieces, using the window modularity peaks as a guide (Fig. 3a). The genes of each piece, together with the information on the interaction between these genes, are fed to Medusa application[16] and partial network graphs were produced, shown in Figure 3. The biological functions are mapped with GO terms. Observe that in this figure we are able to discriminate gene networks of related functions.

For example, networks p1, p2 and p3 (Fig. 3b-d) are all associated with transcription and translation processes, as rRNA/mRNA processing and ribosome biogenesis and assembly. Network p4, also overlaps these functions (Fig. 3e), represented by DNA repair/replication and cell cycle regulation. All these four gene networks have in common the synthesis of biological polymers. By contrast, network p5 seems to be on single cluster, shifting the ordering to other biochemical classes (Fig. 3f), such as cell communication and morphogenesis. The last two gene networks (Figs. 3g-h) present a variety of functions, from actin cytoskeleton organization and vesicle transport to carbohydrate, lipid and amino acid metabolic processes. Supplementary Figure S3 shows the GO terms as black dots clusters located on the interaction matrix. That figure discriminates better the cytoskeleton related functions, indicating that they populate the transition of peak 5 to 6.

A remarkable feature of the right side of VEM ordering is the presence of several intermediate products and ATP-producing pathways (*e.g.* carboxylic acid cycle and cellular respiration). The network structure is enriched with highly interconnected anabolic and catabolic pathways, which is consistent with the basic strategy of central metabolism to form ATP, electron carriers and precursors for the biosynthesis of more-

complex molecules. Therefore, gene networks p6 and p7 are related to the production of both energy and the building blocks from which other molecules of the cell are made.

At the other end of the VEM ordering (the left side), the functional boundaries of the network structure seems to be better discriminated. There are sub-clusters associated with several processing steps that control the flow of genetic information in cells. As the functional chain, from DNA to RNA to protein, takes place in a more linear way comparing to central metabolism, it stands to reason that the network structure located at this ordering side is better discriminated by a one-dimensional representation.

Taken together, the metabolic pattern as organized by the VEM algorithm gives rise to a sound biochemical and functional ordering, where the closest gene networks are more interrelated than the distant ones.

Finally, we consider gene expression data for the yeast genome. We considered experimental data available at Gene Expression Omnibus database, regarding microarrays presenting probes for almost all genome components. We have then projected the expression on the VEM ordering, always considering window averages, obtaining expression profiles that we call transcriptograms. Here we show the transcriptograms for *Saccharomyces cerevisiae* using the data presented by the very nice paper by Tu *et al.*[17] As explained in that reference, the expression data were obtained from yeast continuous culture, in controlled conditions, where the concentration levels of dissolved $O_2$ is constantly measured. These levels vary periodically in time and the transcription levels were measured for 12 different stages in three different dissolved $O_2$ concentration oscillation periods, summing up 36 transcription profiles. Figure 4 presents the results concerning transcriptograms obtained using the VEM ordering. A movie presenting all 36 snapshots is available at Supplementary Materials On Line, as well as the results for the Dendogram ordering. Figure 4a presents 21 transcriptograms (7 per cycle), taken at the instants represented by

the colored (orange, blue, and purple) dots on the plot of dissolved oxygen versus time in log-linear plot (Fig. 4b). Each color is associated with one cycle. Figure 4a also presents the window modularity as a landscape, to guide the eye, and the distribution of three gene clusters as defined in Tu *et al*. paper based on sentinels genes: Ox (oxidative), R/B (reductive, building), and R/C (reductive, charging). Figures 4c to 4i present the relative expression profiles at different instants, where we divided all profiles by the expression values of the first transcriptome ($t = 0$). This stage has been arbitrarily chosen to evince time relative variations of gene expression. The expression profiles show different behaviors for the left and right hand side portions: the expression profile of left side peaks extremely abruptly at the intense burst of oxygen consumption, while the right side gradually raises when cells begin to cease oxygen consumption. According to the gene networks mapped in Figures 1 and 3, the left side embraces several energy-demanding processes, essentially represented by the synthesis of biological polymers. It requires abundant amounts of adenosine triphosphate (ATP), which is available in profusion at the respiratory phase. This interplay of metabolic pathways for energy production is compatible with the time ordering through the phases Ox, R/B, and R/C as described in the original paper.[17] Our results support the conclusion drawn by the authors based on the expression of 15 genes for each cluster, a small gene fraction available in yeast transcriptomes. Here, by the use of transcriptograms, we present the dynamic changes during the metabolic cycle assessing the complete information.

In summary, we propose here the transcriptogram as a tool for assessing cell metabolism, which is capable of discriminating the stage the cell is going through at a given instant. The reason why this is possible is based on the functional modularity of the genome, which is evinced by a hierarchical ordering of the gene list: strongly interacting genes form clusters, strongly interacting clusters form clusters of clusters, and so forth. A second requirement for a one-dimensional transcriptogram is the

correlation between distance and interaction strength not only between genes, but also between clusters of genes. In this way, the hierarchical characteristic of the ordering manifests itself on the interaction matrix by presenting a decreasing density of black dots as the distance from the main diagonal increases, such that the upper left and lower right corners are free of the black dots associated with interactions between genes far apart. Dendogram-like methods are capable of ordering the genome at gene-gene interaction level, but are less efficient in ordering at cluster-cluster interaction level, thus compromising the quality of functional information that the averages over neighboring sites may help producing. The VEM ordering algorithm contemplates the cluster-cluster interaction level by penalizing with higher virtual energy the interaction matrix configurations associated with gene orderings that locate interacting genes at distant positions. Further improvements on the algorithm should specifically contemplate this hierarchy, which ultimately reflects the functional correlation between genes. In fact, the transcriptogram opens the possibility of a tool that works as a telescope, where the focus is tunable and may be adjusted to the desired level of details: when passing from a wide genome overview to smaller functional modules analysis, the observation window may be narrowed down, discriminating more functional modules at greater detail. On the other hand, the method is readily applicable to any species, including *Homo sapiens*, which will be presented elsewhere.

## Methods

We retrieved protein-protein interactions from STRING database (http://string.embl.de/)[9] , using "experimental" and "database" (95% of these interactions) added with "neighbourhood", "fusion", "co-expression", and "co-occurrence" evidences, String-score≥0.800, comprising 4655 genes and 47415 interactions.

The VEM Algorithm minimizes the energy given by Eq. (1). Periodic boundary conditions were used, except to calculate the distance to the main diagonal. We randomly choose two genes on the ordered list, swap their positions, and re-calculate the energy. If the virtual energy decreases, the change is accepted. If the energy is increased by $\Delta E$ , the change is accepted with probability $\exp[-\Delta E /T]$, where T is a virtual temperature. We started with $T = 6 \times 10^5$ and every 100 Monte Carlo Steps (MCS) the temperature is lowered to 20% of its previous value. A MCS is a number of random choices equal to the number of elements in the system.

GO term enrichment was performed using *DAVID* bioinformatics resources (http://david.niaid.nih.gov)[13] to determine whether particular gene ontology terms occur more frequently than expected by chance in a given set of genes. We used default settings for the category GOTERM_BP_ALL, and selected those terms with $P < 0.05$ (for FDR no greater than 5%) representing central biochemical pathways/metabolic functions. From bit strings where the ith bit is set to 1(0) whenever the ith gene of an ordering is (not) listed in the GO term, we obtain smooth profiles by assigning to every gene the fraction of bits valued 1 in a window of size w, centered on the gene.

Yeast transcript expression data were obtained from YG_S98 array platforms (Affymetrix, Inc.)[17], available at GEO database, Series GSE3431 (http://www.ncbi.nlm.nih.gov/projects/geo/). The transcriptograms are obtained by assigning to the ith gene the average of the expression values of its neighbors in a window of size w centered at the gene.

Reference List


1. Rives,A.W. & Galitski,T. Modular organization of cellular networks. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 1128-1133 (2003).

2. Ravasz,E., Somera,A.L., Mongru,D.A., Oltvai,Z.N. & Barabasi,A.L. Hierarchical Organization of Modularity in Metabolic Networks. *Science* **297**, 1551-1555 (2002).

3. Hintze,A. & Adami,C. Evolution of Complex Modular Biological Networks. *PLoS Comput Biol* **4**, e23 (2008).

4. Barabasi,A.L. & Albert,R. Emergence of scaling in random networks. *Science* **286**, 509-512 (1999).

5. Koonin,E.V. ORTHOLOGS, PARALOGS, AND EVOLUTIONARY GENOMICS1. *Annual Review of Genetics* **39**, 309-338 (2005).

6. Kanehisa,M. *et al.* From genomics to chemical genomics: new developments in KEGG. *Nucl. Acids Res.* **34**, D354-D357 (2006).

7. Raman,K. & Chandra,N. Flux balance analysis of biological systems: applications and challenges. *Brief Bioinform* **10**, 435-449 (2009).

8. Asur,S., Ucar,D. & Parthasarathy,S. An ensemble framework for clustering protein protein interaction networks. *Bioinformatics* **23**, i29-i40 (2007).

9. Jensen,L.J. *et al.* STRING 8--a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.* **37**, D412-D416 (2009).

10. Barabasi,A.L. & Oltvai,Z.N. Network biology: Understanding the cell's functional organization. *Nat Rev Genet* **5**, 101-113 (2004).

11. Vinogradov,A.E. Modularity of cellular networks shows general center-periphery polarization. *Bioinformatics* **24**, 2814-2817 (2008).

12. Ashburner,M. *et al.* Gene Ontology: tool for the unification of biology. *Nature genetics* **25**, 25-29 (2000).

13. Huang,d.W., Sherman,B.T. & Lempicki,R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* **4**, 44-57 (2009).

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

15. Keeling,P.J. & Palmer,J.D. Horizontal gene transfer in eukaryotic evolution. *Nat. Rev. Genet.* **9**, 605-618 (2008).

16. Hooper,S.D. & Bork,P. Medusa: a simple tool for interaction graph analysis. *Bioinformatics* **21**, 4432-4433 (2005).

17. Tu,B.P., Kudlicki,A., Rowicka,M. & McKnight,S.L. Logic of the Yeast Metabolic Cycle: Temporal Compartmentalization of Cellular Processes. *Science* **310**, 1152-1158 (2005).

## LEGENDS

**Figure 1. Protein-protein interaction matrix analysis algorithms**. The axes relative to gene position have been divided by the total number of genes: 4655. (**a**) Random ordering. (**b**) Dendogram ordering algorithm. (**c**) Virtual Energy Minimizing (VEM) algorithm. (**d-l**) Biological function projections for different peaks and orderings. Gray landscape backgrounds: window modularity for the orderings. The maxima at the window modularity plots correspond to larger concentrations of black dots on the matrix representation, that is, intra-module interactions are more intense in these regions.

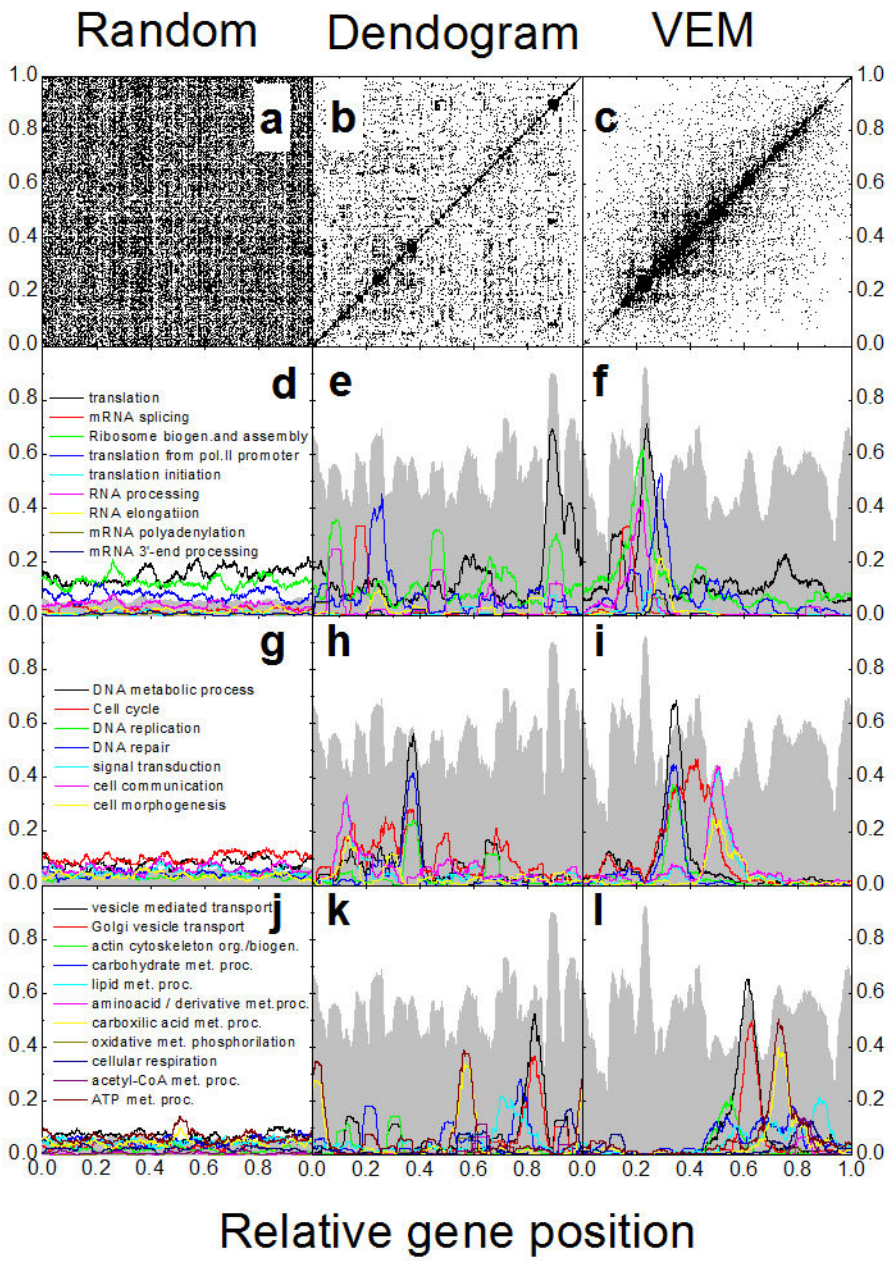
**Figure 2. Statistical and evolutionary profiles of Dendogram and VEM orderings.** The axes relative to gene position have been divided by the total number of genes: 4655. (**a-b**) Connectivity, modularity and clustering coefficient. The profiles of Dendogram and VEM orderings were obtained by taking the average of these quantities over windows of 251 sites. (**c-d**) Fraction of genes in a 251 gene window whose COG presents genes from Eukarya, Bacteria and Archea, from Eukarya and Bacteria only, from Eukarya and Archea only, and from Eukarya only. The light gray landscape background gives the window modularity to guide the eye. The gray line in both figures is the profile of the fraction of genes with COG in the window. **(e-f)** Black dot density

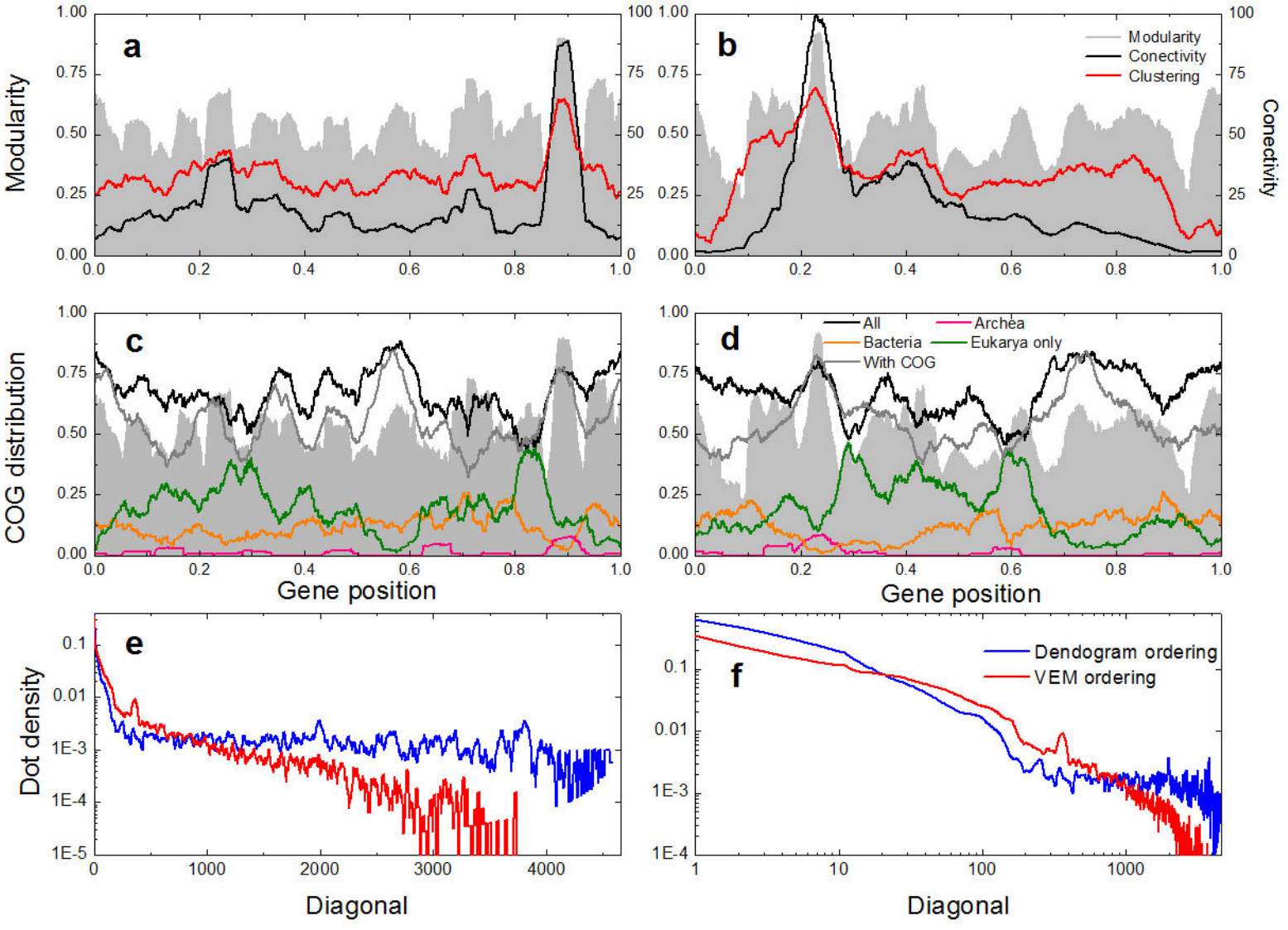
on lines parallel to the main diagonal, as a function of the distance of the line from the main diagonal. This gives information on the quantity of links between genes as a function of their distance on the ordering. (**e**) On a log-linear plot, to evince the exponential decay of dot density on the VEM plots and (**f**) on a log-log, to evince the behaviour near the main diagonal.

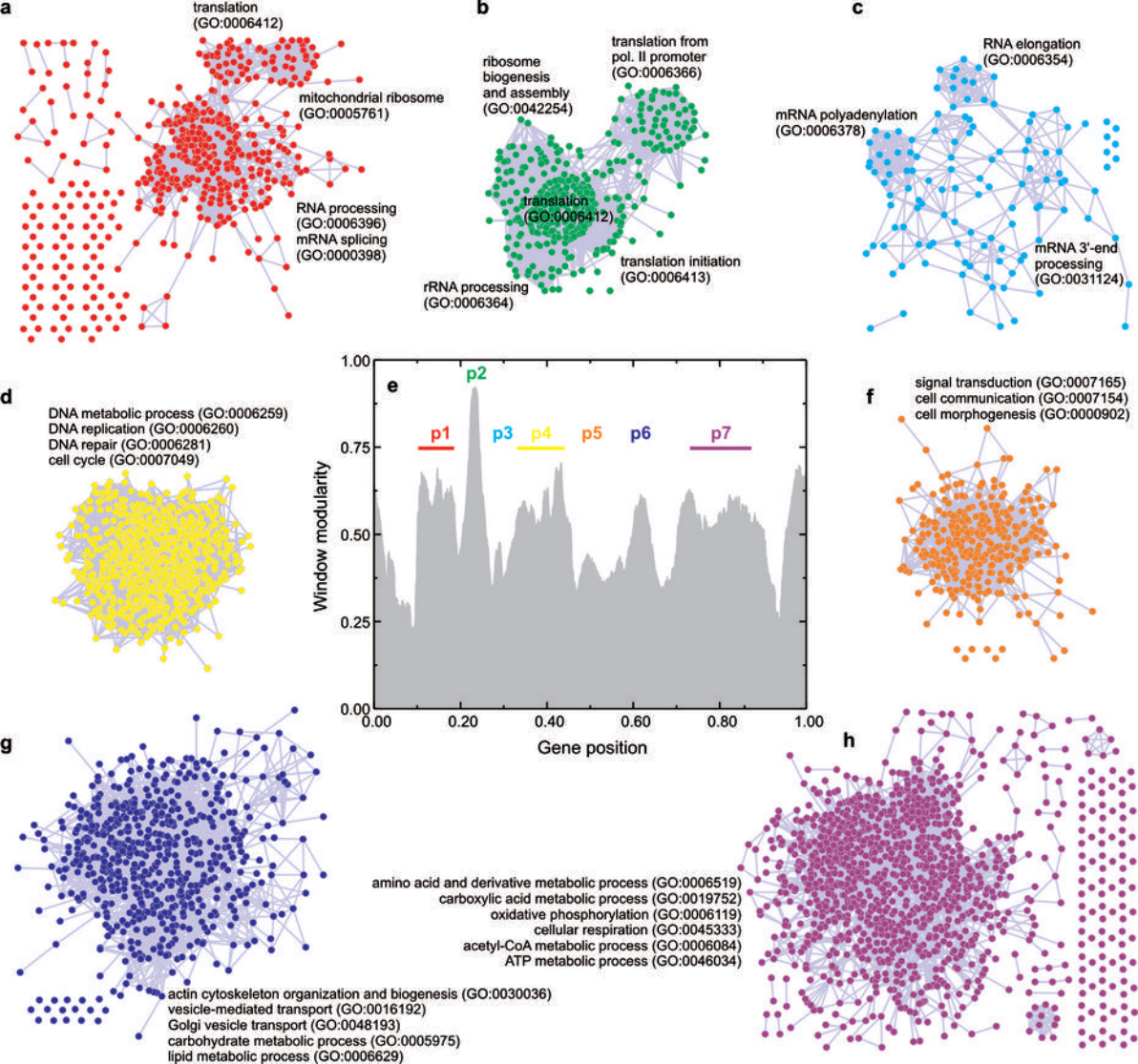
**Figure 3. Graph representation of the VEM ordering.** The axes relative to gene position have been divided by the total number of genes: 4655. (**a**) VEM ordering was sliced in seven pieces, using the window modularity peaks as a guide for this division. The genes of each piece, together with the information on the interaction between these genes, were fed to Medusa application[16] to produce the network graphs. (**b – h**) Network graphs associated with each peak, whose biological functions are mapped with GO terms using *DAVID* bioinformatics resources.[13]

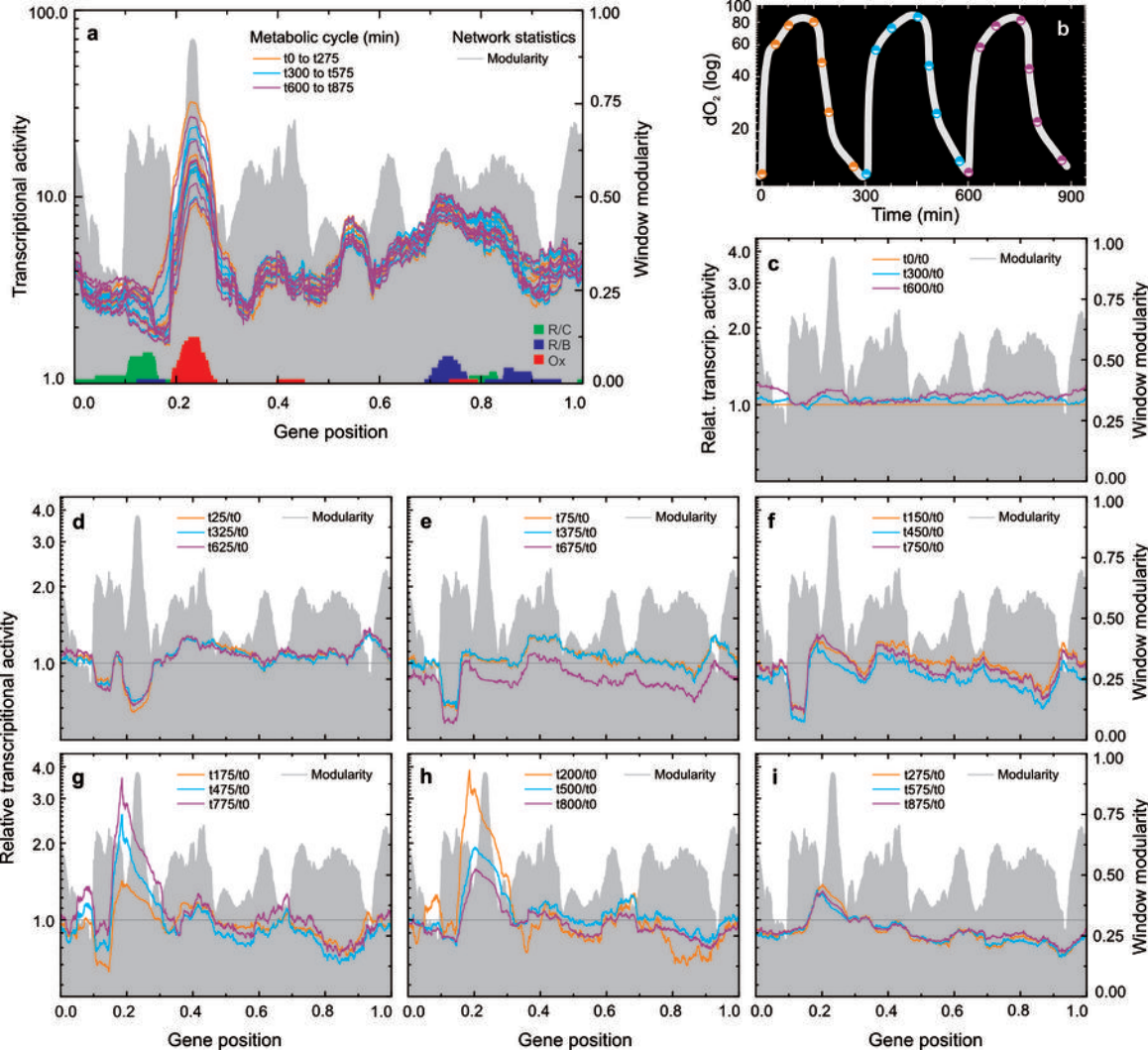
**Figure 4. *Saccharomyces cerevisiae* transcriptograms.** The axes relative to gene position have been divided by the total number of genes: 4655. (**a**) Microarray data available at Gene Expression Omnibus database[17] were projected on VEM ordering to obtain the expression profiles, or transcriptograms. Each color is associated with one cycle, as explained in B. Projections on the ordering were performed always considering window averages. To guide the eye, the window modularity is depicted as a landscape, together with the distribution of three gene clusters, as described previously[17] based on sentinels genes: Ox (oxidative), R/B (reductive, building), and R/C (reductive, charging). (**b**) Plot of dissolved Oxygen versus time in log-linear. Transcriptograms (6 per cycle), were taken at the instants represented by the colored (orange, blue, and purple) dots. (**c – i**) Relative expression profiles. Transcriptograms were divided by the expression values of the first transcriptome ($t$=0). **c**: represents the relative expression profile corresponding to the first dot of each cycle; **d**: represents the second dot of each cycle, and so on.

‘**Supplementary Information** accompanies the paper on **www.nature.com/nature**.’


Random
Dendogram
VEM
a
b
c
d
e
f
g
h
i
j
k
l
translation
mRNA splicing
Ribosome biogen.and assembly
translation from pol.II promoter
translation initiation
RNA processing
RNA elongatiion
mRNA polyadenylation
mRNA 3'-end processing
DNA metabolic process
Cell cycle
DNA replication
DNA repair
signal transduction
cell communication
cell morphogenesis
vesicle mediated transport
Golgi vesicle transport
actin cytoskeleton org./biogen.
carbohydrate met. proc.
lipid met. proc.
aminoacid / derivative met.proc.
carboxilic acid met. proc.
oxidative met. phosphorilation
cellular respiration
acetyl-CoA met. proc.
ATP met. proc.
0.0
0.2
0.4
0.6
0.8
1.0
Relative gene position


a
b
c
d
e
f
Modularity
Conectivity
Clustering
COG distribution
Gene position
All
Archea
Bacteria
Eukarya only
With COG
Dot density
Diagonal
Dendogram ordering
VEM ordering


a
translation (GO:0006412)
mitochondrial ribosome (GO:0005761)
RNA processing (GO:0006396)
mRNA splicing (GO:0000398)
b
ribosome biogenesis and assembly (GO:0042254)
translation from pol. II promoter (GO:0006366)
translation (GO:0006412)
translation initiation (GO:0006413)
rRNA processing (GO:0006364)
c
RNA elongation (GO:0006354)
mRNA polyadenylation (GO:0006378)
mRNA 3'-end processing (GO:0031124)
d
DNA metabolic process (GO:0006259)
DNA replication (GO:0006260)
DNA repair (GO:0006281)
cell cycle (GO:0007049)
e
p1
p2
p3
p4
p5
p6
p7
Window modularity
Gene position
0.00
0.25
0.50
0.75
1.00
0.20
0.40
0.60
0.80
f
signal transduction (GO:0007165)
cell communication (GO:0007154)
cell morphogenesis (GO:0000902)
g
actin cytoskeleton organization and biogenesis (GO:0030036)
vesicle-mediated transport (GO:0016192)
Golgi vesicle transport (GO:0048193)
carbohydrate metabolic process (GO:0005975)
lipid metabolic process (GO:0006629)
h
amino acid and derivative metabolic process (GO:0006519)
carboxylic acid metabolic process (GO:0019752)
oxidative phosphorylation (GO:0006119)
cellular respiration (GO:0045333)
acetyl-CoA metabolic process (GO:0006084)
ATP metabolic process (GO:0046034)


a
Metabolic cycle (min)
t0 to t275
t300 to t575
t600 to t875
Network statistics
Modularity
R/C
R/B
Ox
Transcriptional activity
Window modularity
Gene position
b
dO2 (log)
Time (min)
c
t0/t0
t300/t0
t600/t0
Modularity
Relat. transcrip. activity
Window modularity
d
t25/t0
t325/t0
t625/t0
e
t75/t0
t375/t0
t675/t0
f
t150/t0
t450/t0
t750/t0
g
t175/t0
t475/t0
t775/t0
h
t200/t0
t500/t0
t800/t0
i
t275/t0
t575/t0
t875/t0
Relative transcriptional activity
Gene position