Towards a transcriptogram

Instituto de Física UFRGS

Team

- José Cláudio Moreira (Biochemistry-UFRGS)
- Rita M. C. de Almeida (Physics-UFRGS)
- Leonardo Brunnet (Physics-UFRGS)
- José Carlos M. Mombach (Physics-UFSM)
- Mauro Castro (Physics-UFRGS)
- Marialva Sinigaglia (Genetics-UFRGS)
- José Luiz Rybarczyk (Physics-UFRGS)
- Ricardo Mello (Physic´-UFRGS)
- Rodrigo Dalmolin (Biochemistry-UFRGS)
- Melissa Vanzella (Physics-UFRGS)

Motivation: tumors

- Highly heterogeneous, intratumoral heterogeneous
- Cellular growth, invasion and metastasis
- Neoplasia sólida (ou tumores sólidos)
- Cytogenetic and molecular alterations

disfunction in genome stability and apoptosis

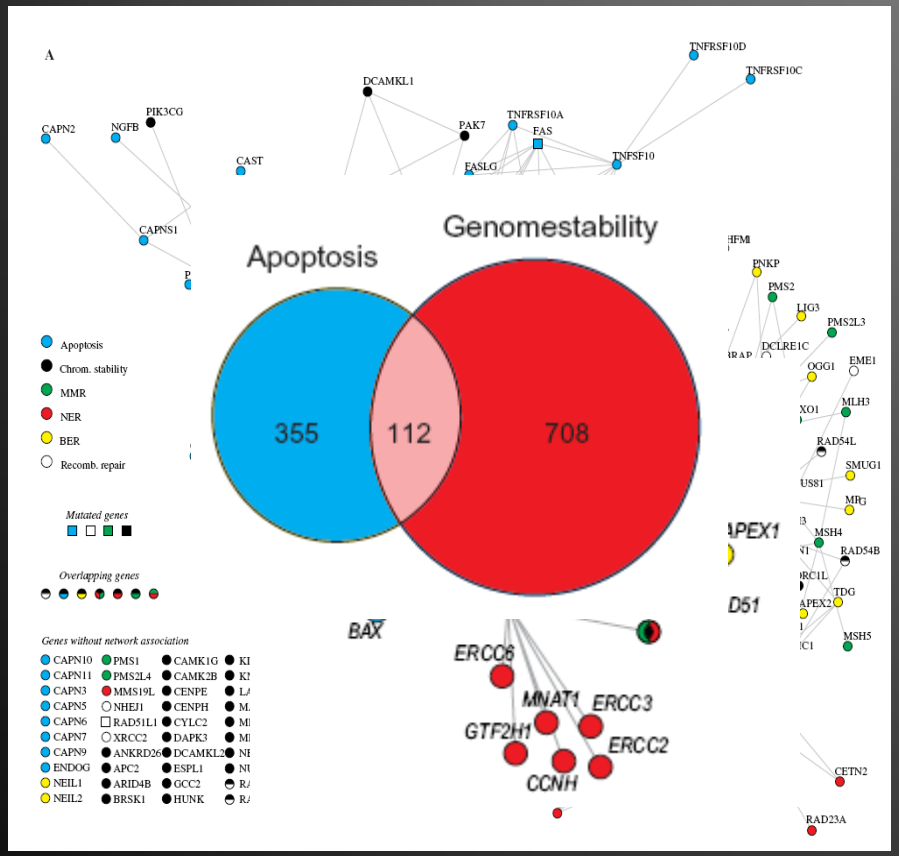
Published online 1 March 2007 *Nucleic Acids Research, 2007, Vol. 35, No. 6* ***1859–1867***
doi:10.1093/nar/gkm061

Impaired expression of NER gene network in sporadic solid tumors

Mauro A. A. Castro[1,2,3,*], José C. M. Mombach[2,4], Rita M. C. de Almeida[2] and José C. F. Moreira[1]

[1]Departamento de Bioquímica, Universidade Federal do Rio Grande do Sul, Rua Ramiro Barcelos 2600-anexo, Porto Alegre 90035-003, Brazil, [2]Instituto de Física, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves 9500, Porto Alegre 91501-970, Caixa Postal 15051, Brazil, [3]Universidade Luterana do Brasil, Av. Itacolomi 3600, Gravataí 94170-240, Brazil and [4]Centro de Ciências Rurais, Unipampa/São Gabriel - Pós-Graduação em Física, Prédio 13, Universidade Federal de Santa Maria, Santa Maria 97105-900, Brazil

Rede de Interações



Castro, M. A. A. et al. Nucl. Acids Res. 2007 35:1859-1867;

- `Data selection:
  - Recombination repair (RER): 25 genes
  - Mismatch repair (MMR): 25 genes
  - Base excision repair (BER): 17 genes
  - Nucleotide excision repair (NER): 28 genes
  - Chromosome stability : 77 genes
  - Apoptosis: 98 genes
- Sources:
  - KEGG (http://www.genome.jp/kegg/)
  - STRING (http://string.embl.de/)
  - HPRD (http://www.hprd.org/)
  - CGAP (http://cgap.nci.nih.gov/Pathways)
  - HGNC (http://www.genenames.org/)
  - PUBMED

Nucleic Acids Research Advance Access published October 2, 2008

Nucleic Acids Research, 2008, 1–15
doi:10.1093/nar/gkn636

Evolutionary origins of human apoptosis and genome-stability gene networks

Mauro A. A. Castro[1,2,*], Rodrigo J. S. Dalmolin[1], José C. F. Moreira[1], José C. M. Mombach[3] and Rita M. C. de Almeida[4]

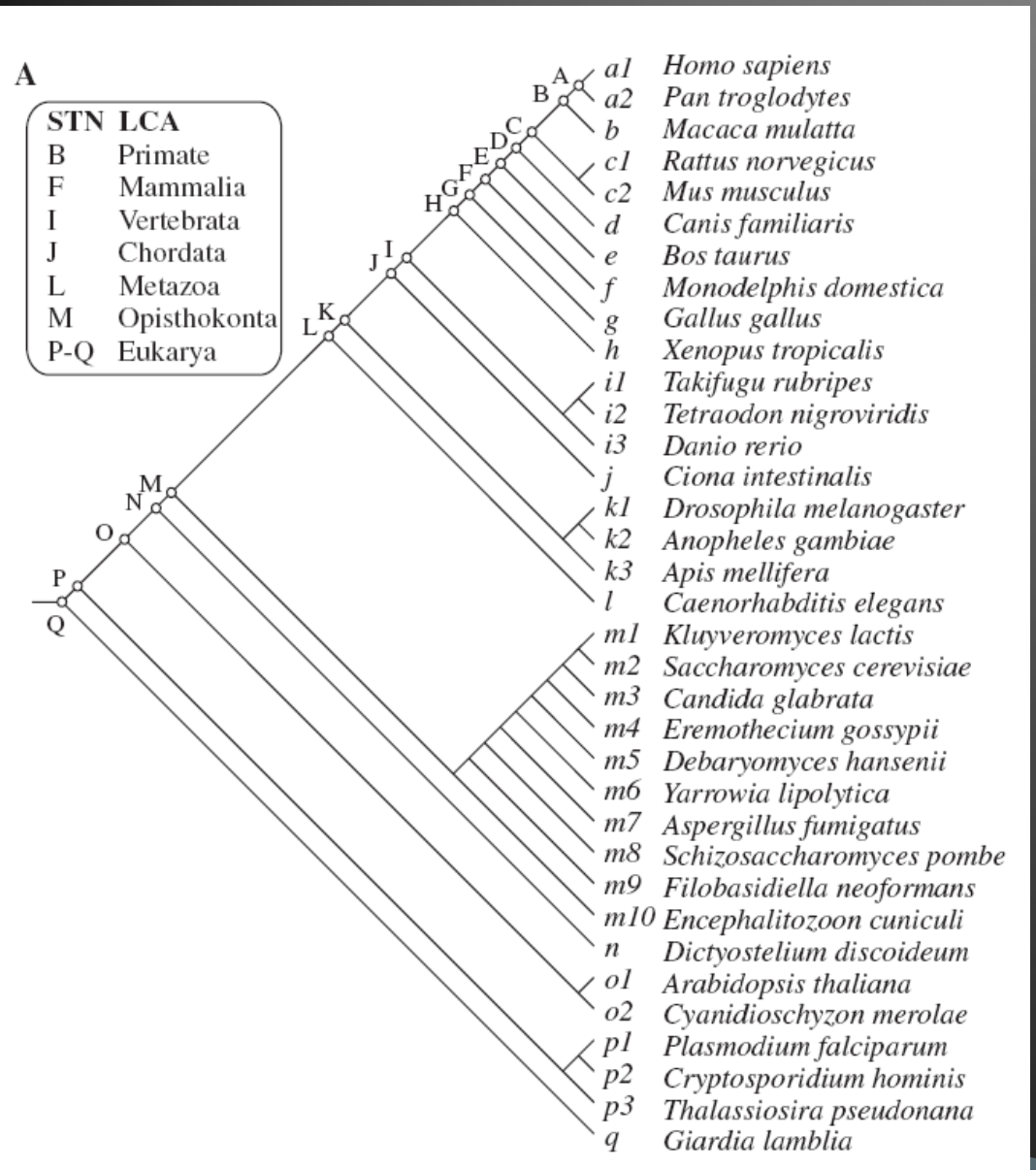
[1]Bioinformatics Unit, Department of Biochemistry, Federal University of Rio Grande do Sul (UFRGS), Rua Ramiro Barcelos 2600-anexo, Porto Alegre 90035-003, [2]Department of Biological Sciences, Lutheran University of Brazil, Gravataí 94170-240, [3]Department of Physics, Federal University of Santa Maria (UFSM), Santa Maria 97105-900 and [4]Institute of Physics, Federal University of Rio Grande do Sul (UFRGS), Avenida Bento GonÓalves 9500, Porto Alegre 91501-970, Caixa Postal 15051, Brazil

Received May 23, 2008; Revised September 14, 2008; Accepted September 15, 2008

Objectives

- Identify the evolutionary root of each component
- Map the orthology to transfer functional information

Infering Evolutionary Roots

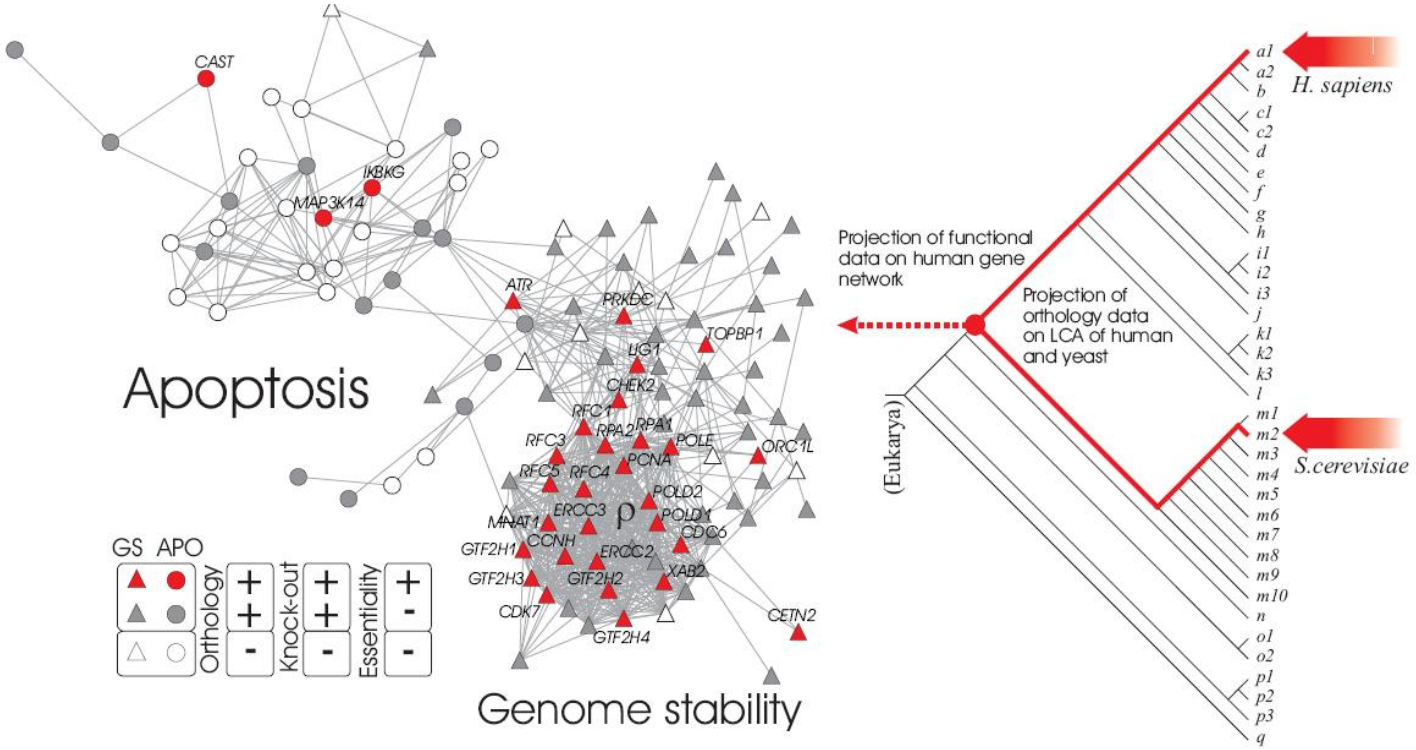


Ortholog groups:

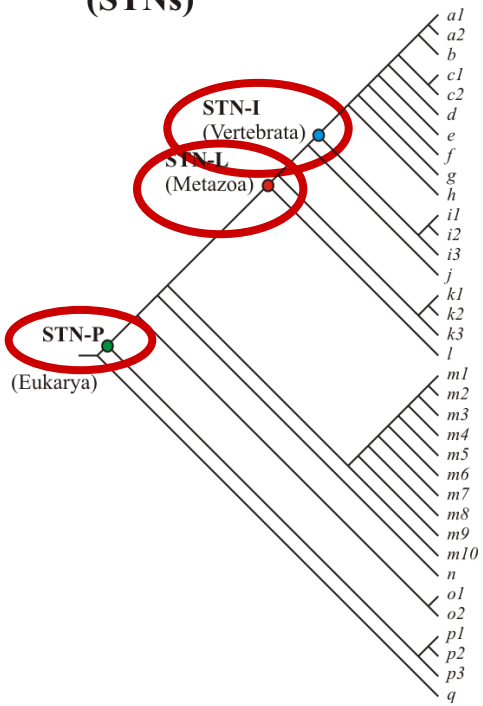
- STRING and InParanoid
- 35 eukaryotes
- Map the components onto the philogenetic tree
- Penalties system

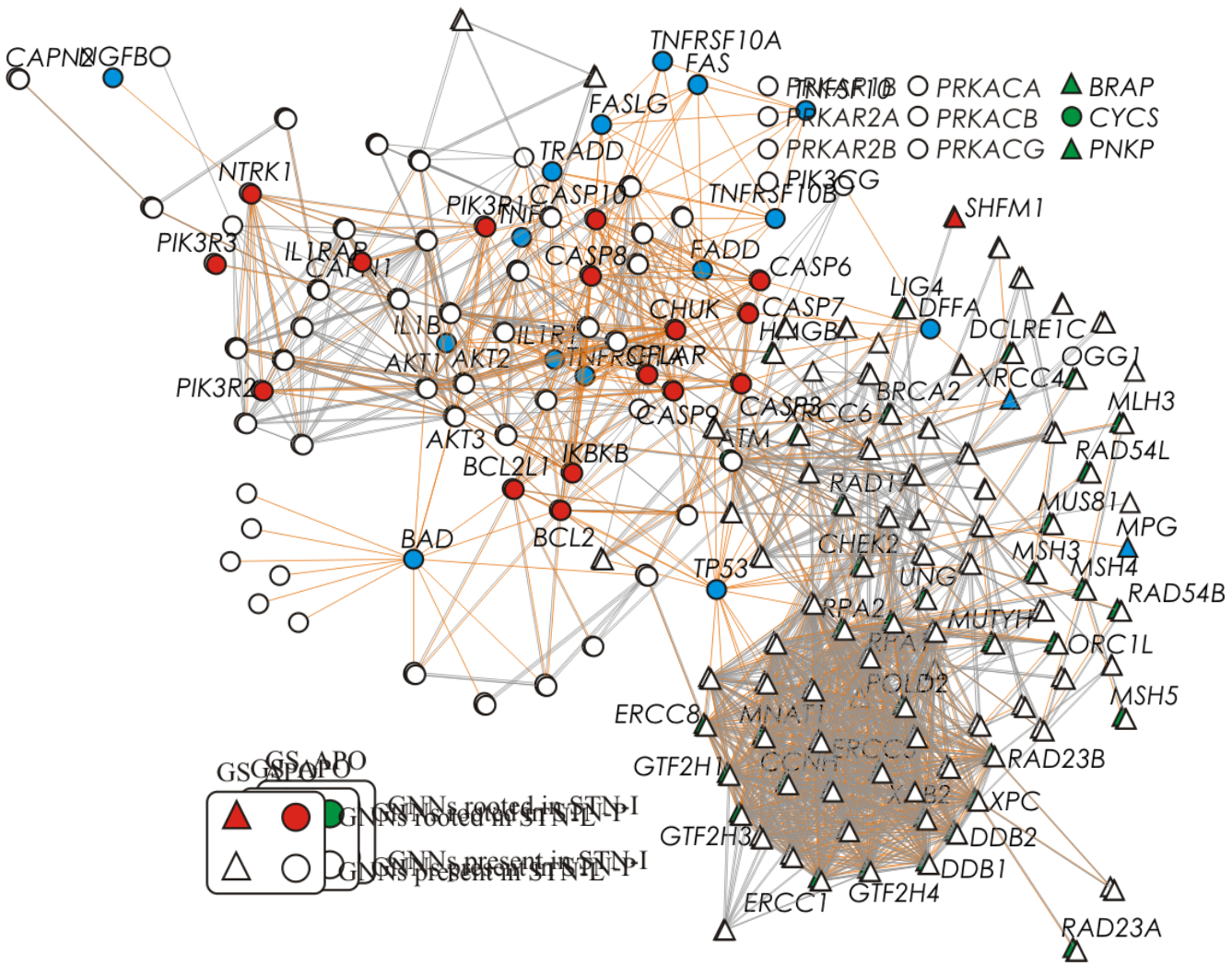
Mirkin, BG et al. *BMC Evol. Biol.* (2003) 3, 2.

Transfering functional data

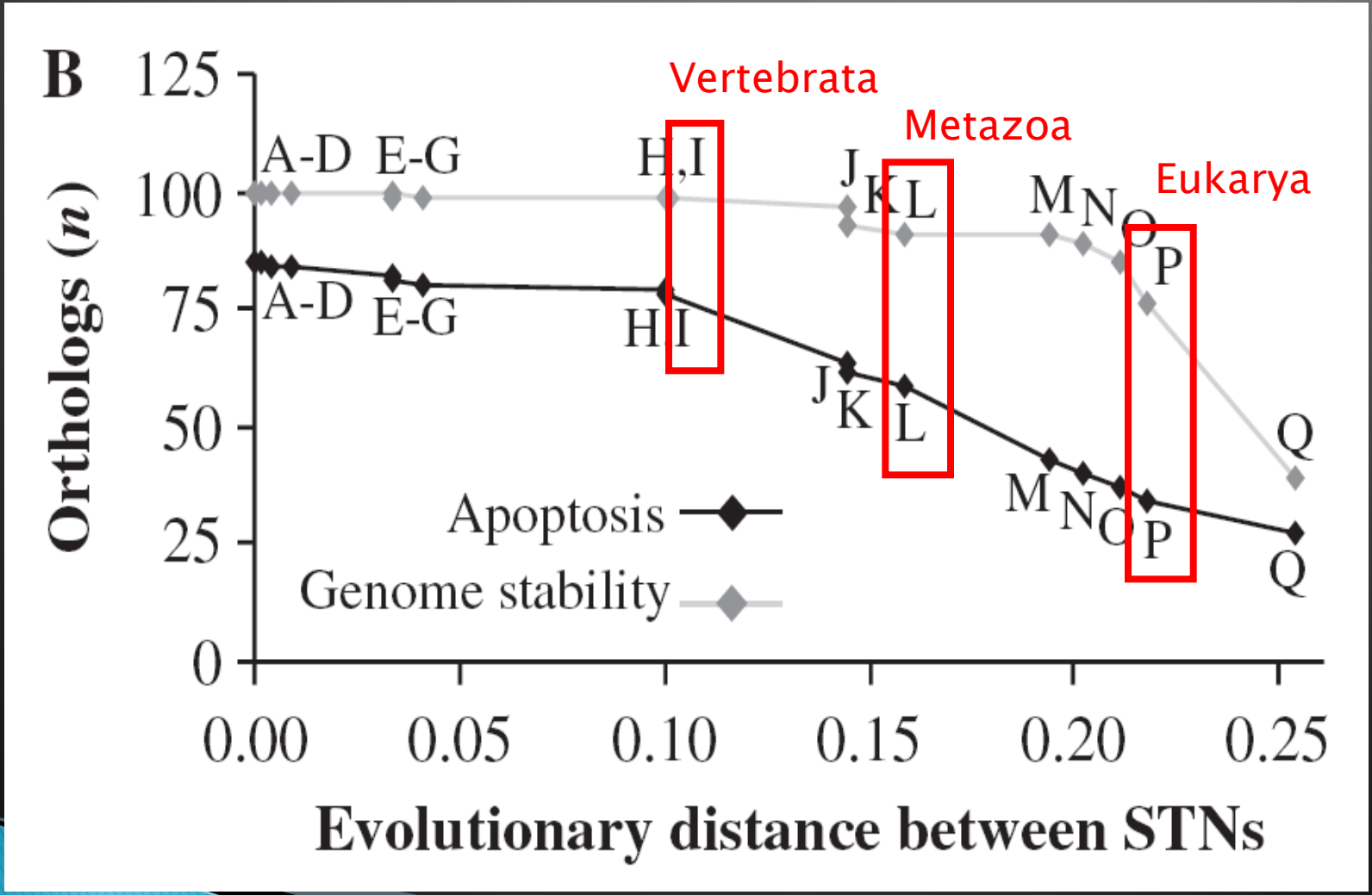


Species-tree nodes (STNs)

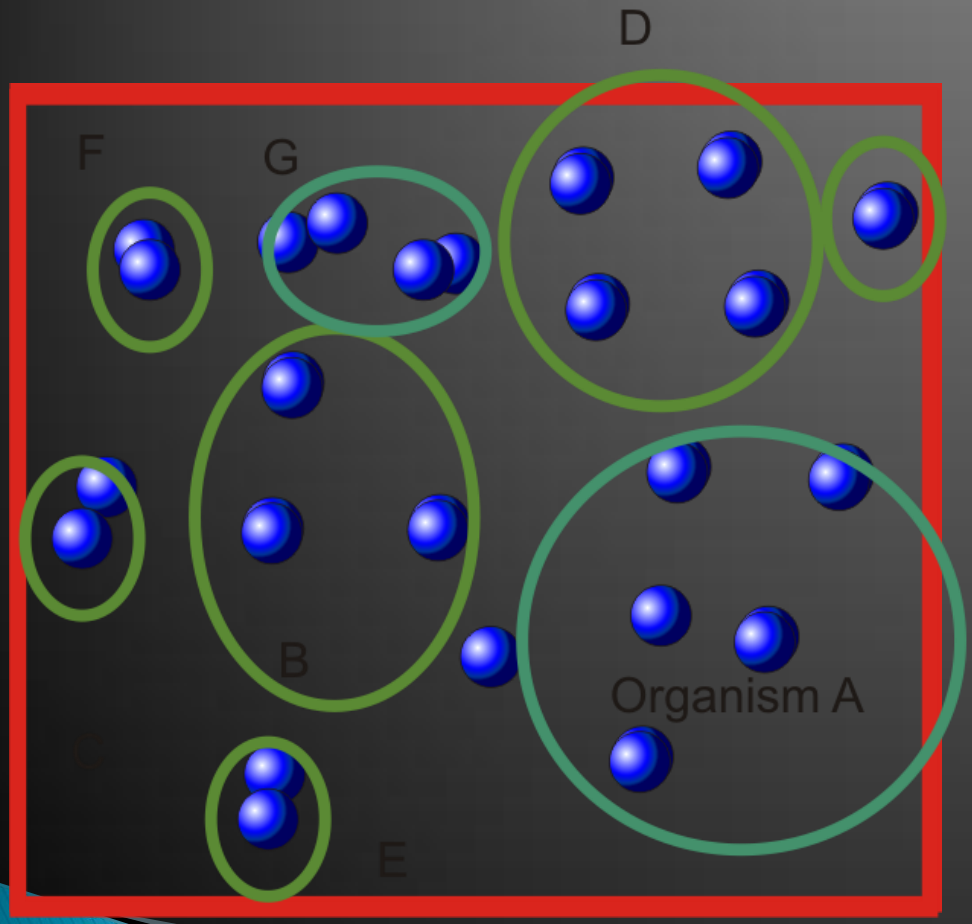




Inferência da Raiz Evolutiva



Diversity and Abundance



In a COG:

Number of orthologs

Number of organisms

Number of orthologs per organisms

Diversity: High number of organisms

Abundance: High number of orthologs

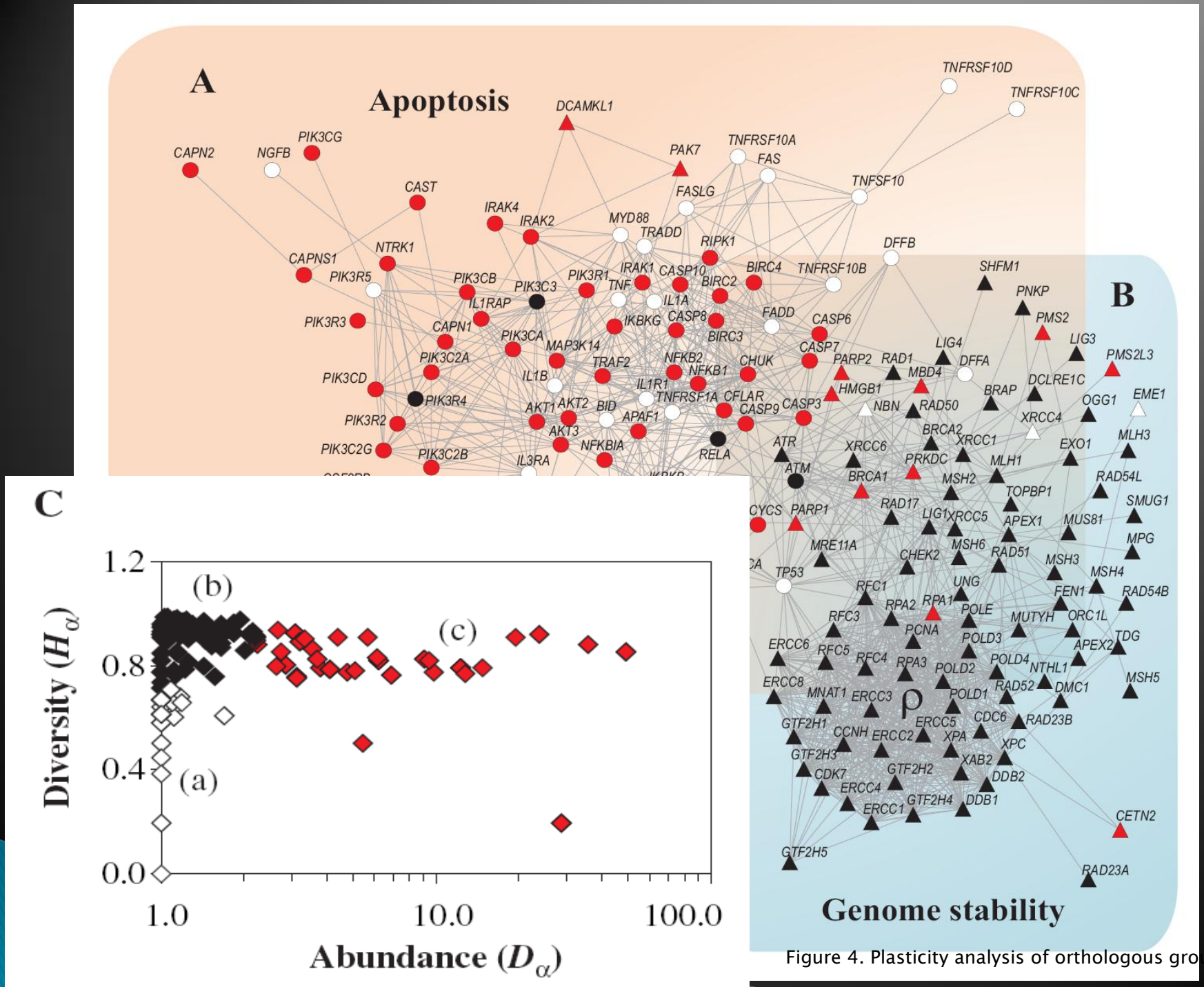


Figure 4. Plasticity analysis of orthologous groups.

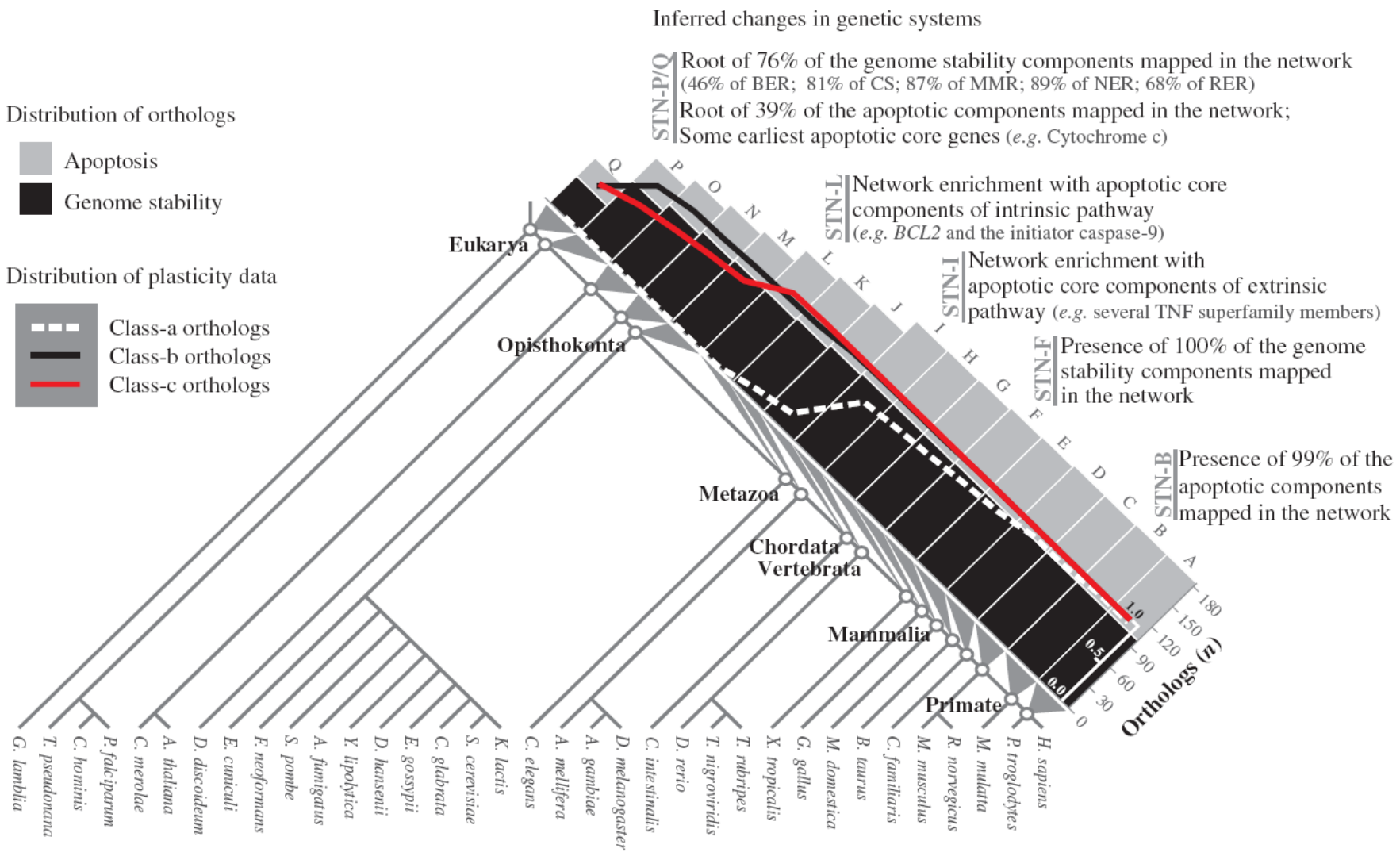
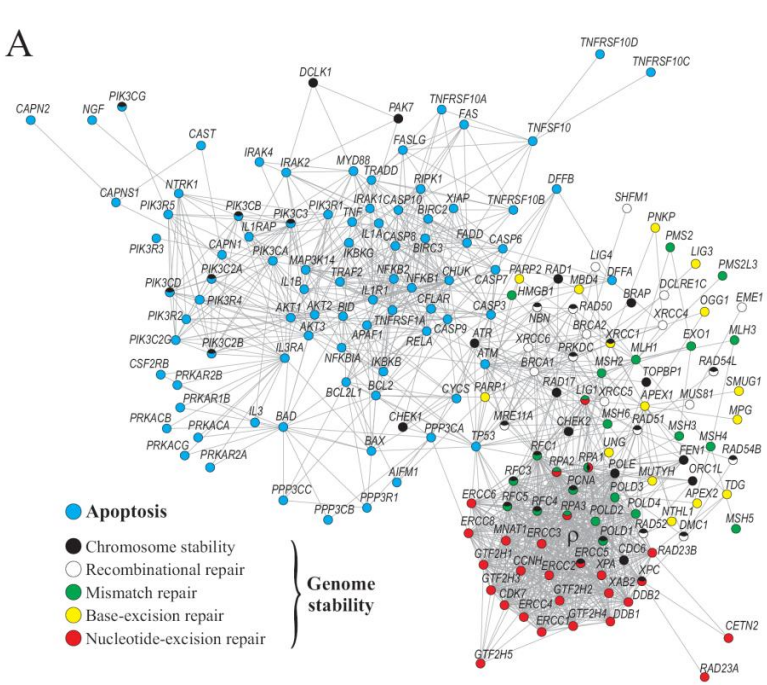
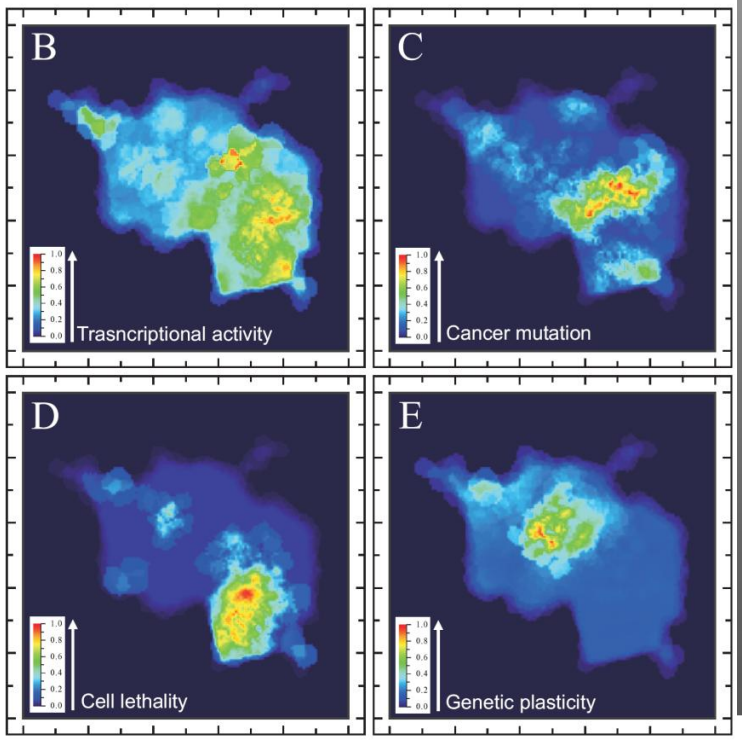


Figure 6. Summary of the inferred changes in genetic systems. The histograms show the distribution of 180 human orthologs according to the roots inferred in the eukaryote species tree (for details, see Figures 2 and 3). STNs and the corresponding LCA are indicated. Inset graph shows the presence fraction of orthologs of each STNs (for details, see Figure 4D). Diverse important events related to the roots of sets of genes are pointed along the STNs. Chromosome stability (CS).





BIOINFORMATICS APPLICATIONS NOTE

2009, pages 1–2
doi:10.1093/bioinformatics/btp246

Systems biology

ViaComplex: software for landscape analysis of gene expression networks in genomic context

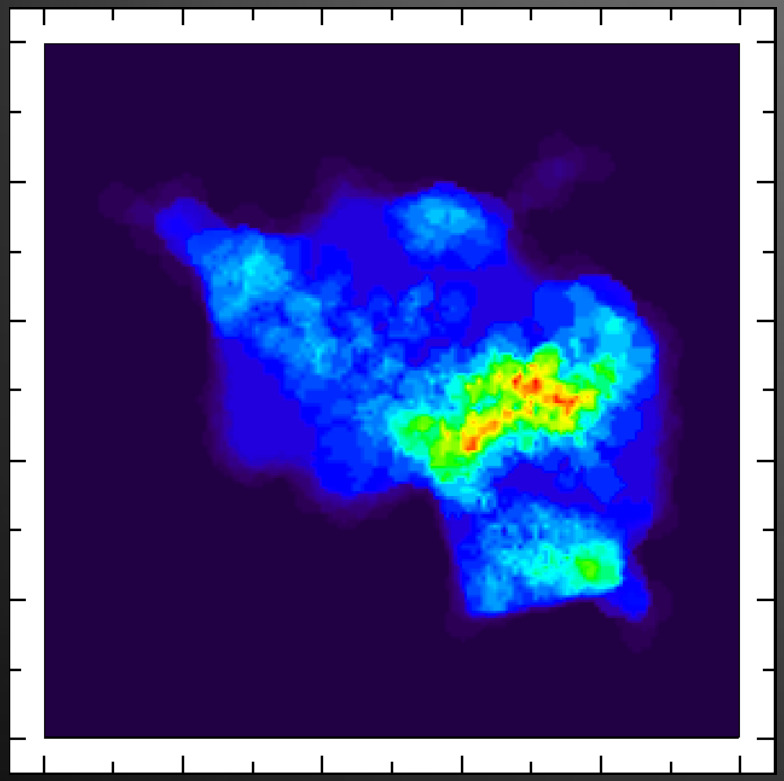
Mauro A. A. Castro[1,*], José L. Rybarczyk Filho[2], Rodrigo J. S. Dalmolin[1], Marialva Sinigaglia[3], José C. F. Moreira[1], José C. M. Mombach[4] and Rita M. C. de Almeida[2]

[1]Unidade de Bioinformática, Departamento de Bioquímica, [2]Instituto de Física, [3]Núcleo de Bioinformática, Departamento de Genética, Universidade Federal do Rio Grande do Sul, Rua Ramiro Barcelos 2600-anexo, Porto Alegre 90035-003 and [4]Departamento de Física, Universidade Federal de Santa Maria, Santa Maria 97105-900, Brazil

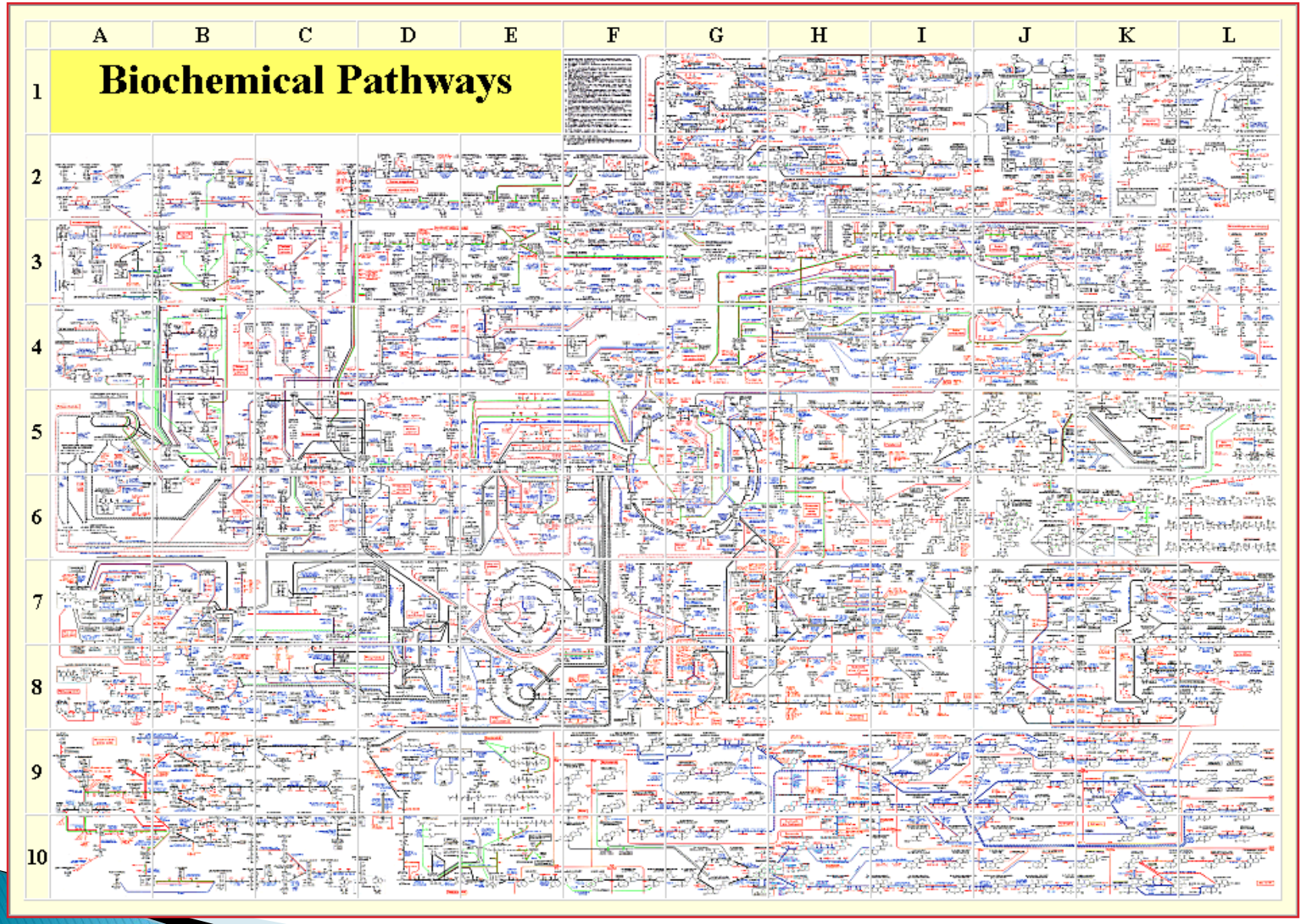
Received on January 13, 2009; revised on March 22, 2009; accepted on April 6, 2009

Associate editor: Thomas Lengauer

Results

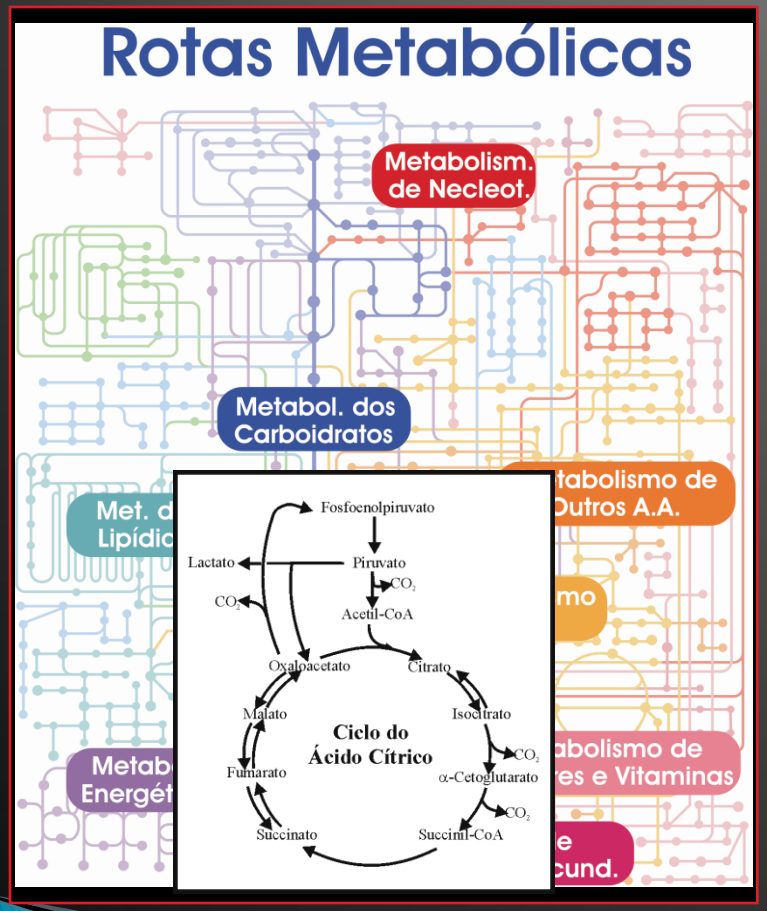


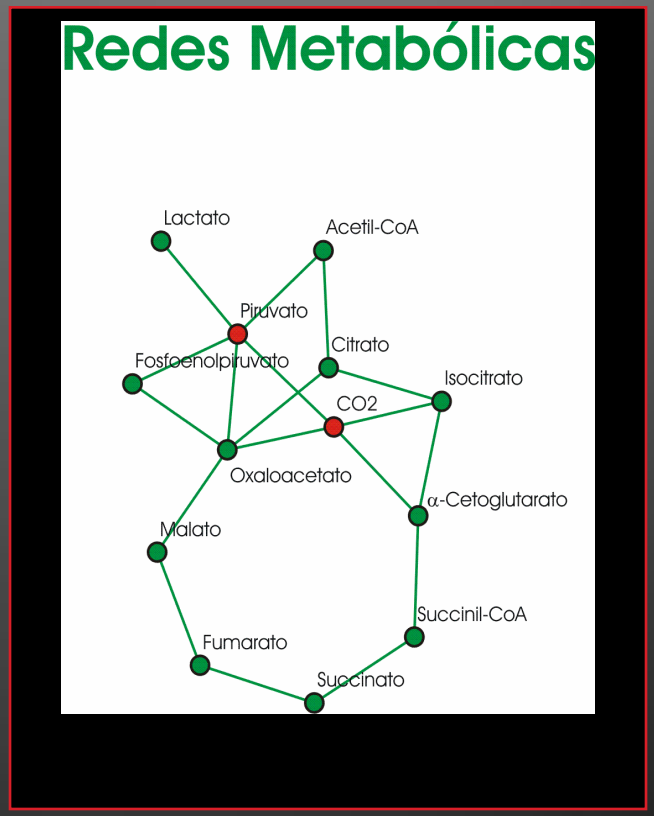
- Ancestrality
- Cellular essentiality
- Embryonary lethality
- Gene causally associated to cancer



Biochemical Pathways - Boehring-Mennheim

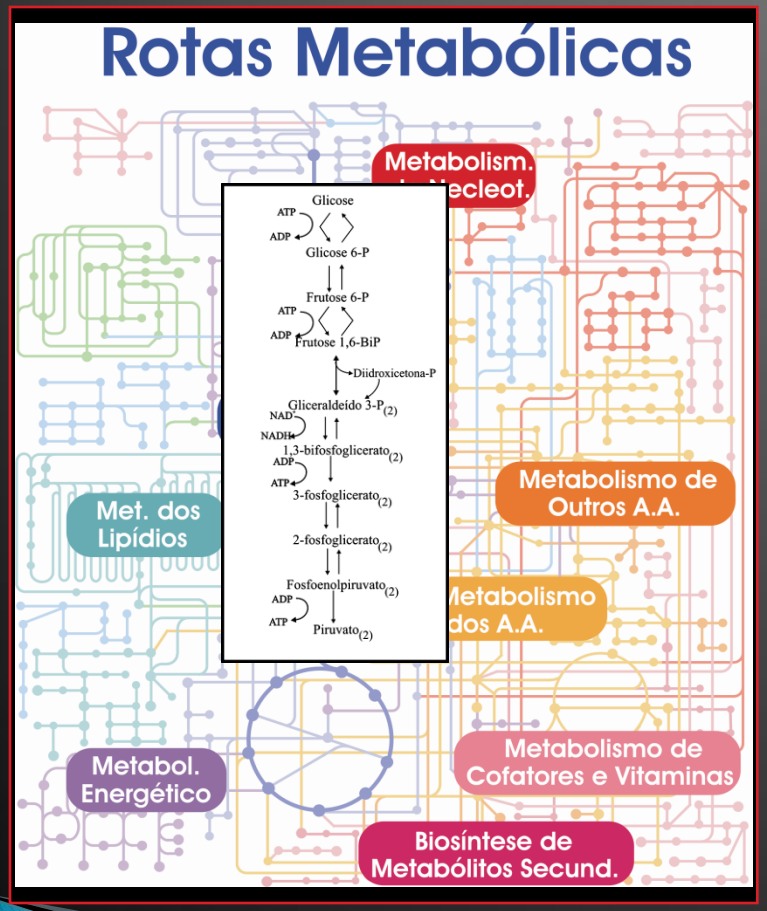
Definitions

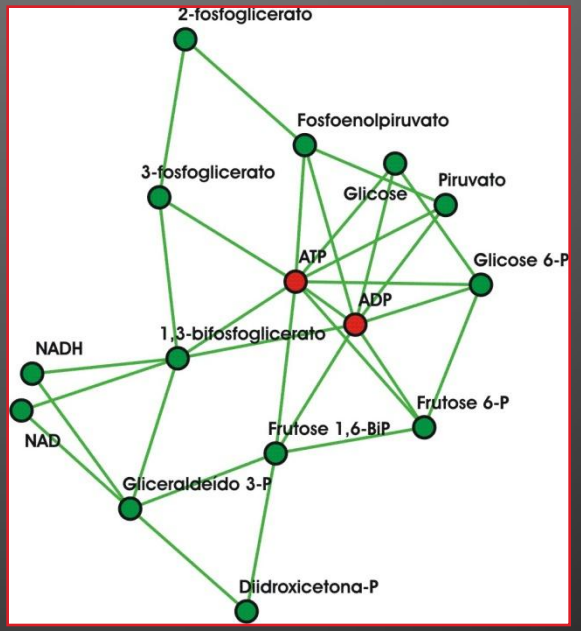




Adaptado de *KEEG database* (www.genome.jp\keeg\)

Definições





Adaptado de *KEEG database* (www.genome.jp\keeg\)

Gene Ontology Database

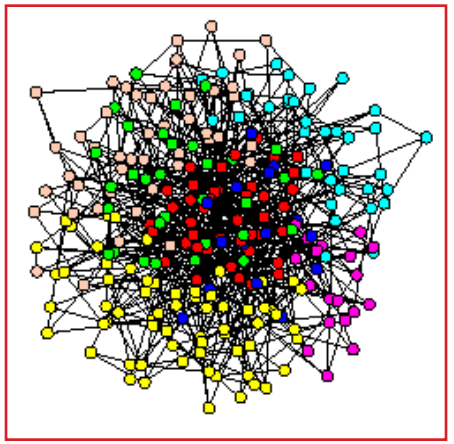
- 3 classifications – ontologies
  - Biologic Process
  - Cellular Component
  - Molecular molecular functions
- Hierarchical structure

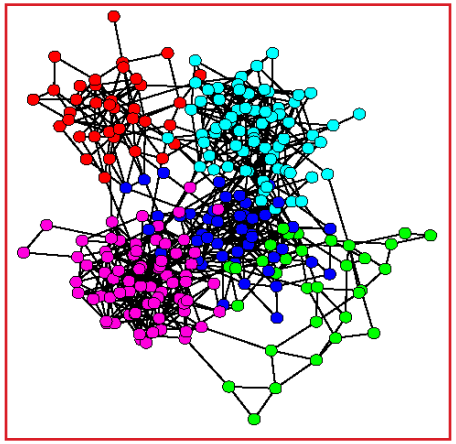
Which one?

- Metabolic pathways → function based
- How isolate systems? How estimate errors?

Modularity

- Scale-free
- Modular





Ravasz E. *Phys. Rev. E*, 67: 026112 (2003)

How to choose the protein set?

- Functions?
- Interactions?
- Can we define clusters?
- Error estimatives
- What about evolution?

The interaction matrix

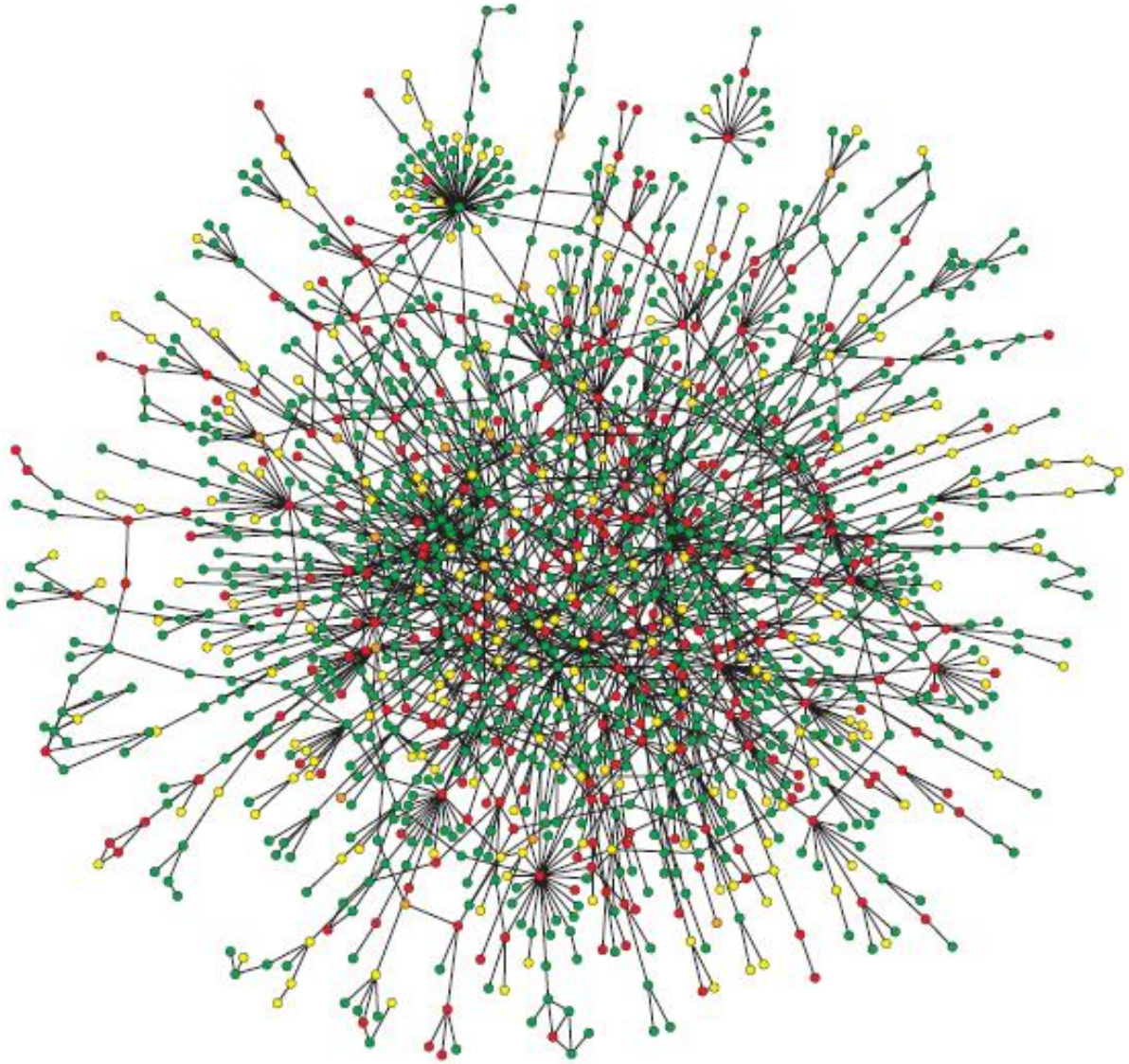
- Enumerate the proteins in random list.
- Build an interaction matrix M_{ij} such that

$M_{ij}=1$ if i and j interact

$M_{ij}=0$ if i and j do not interact

STRING for interaction information

A matriz para o *Saccharomyces*

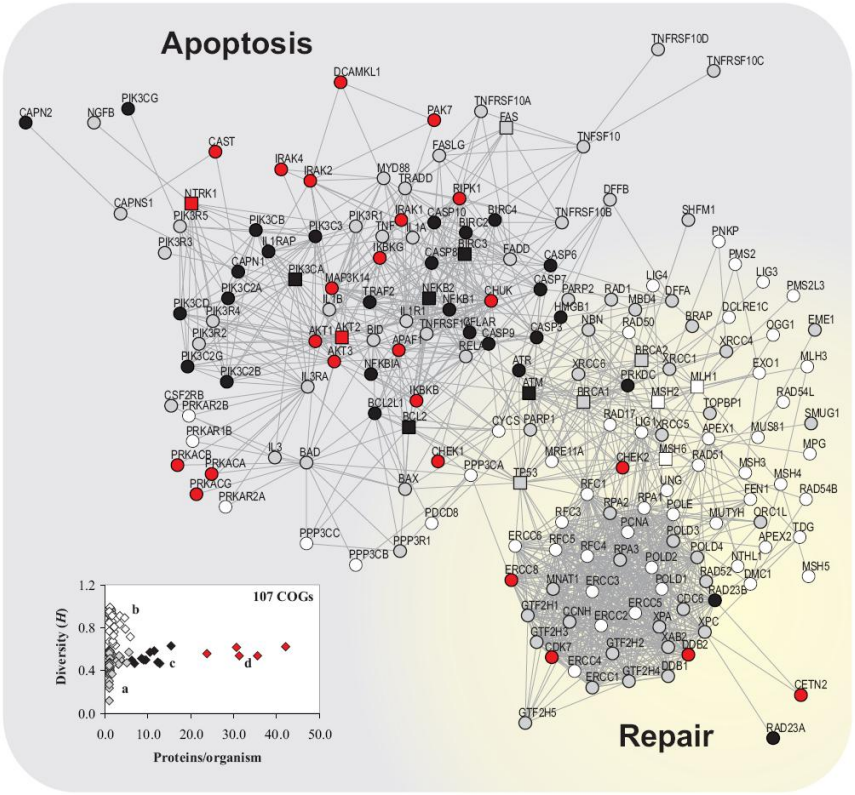


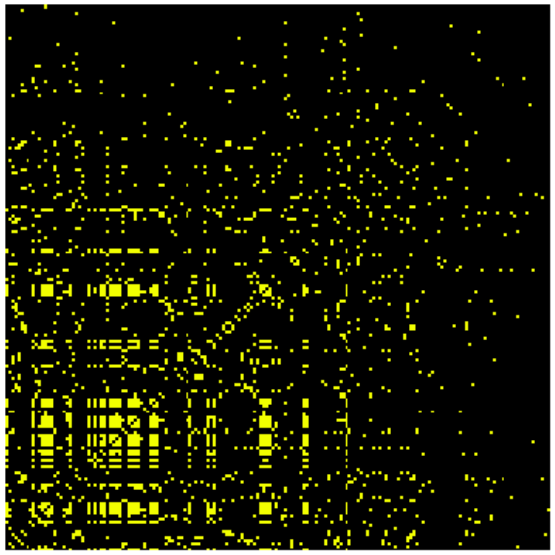
s genes
tão
eatoriamente
denados.

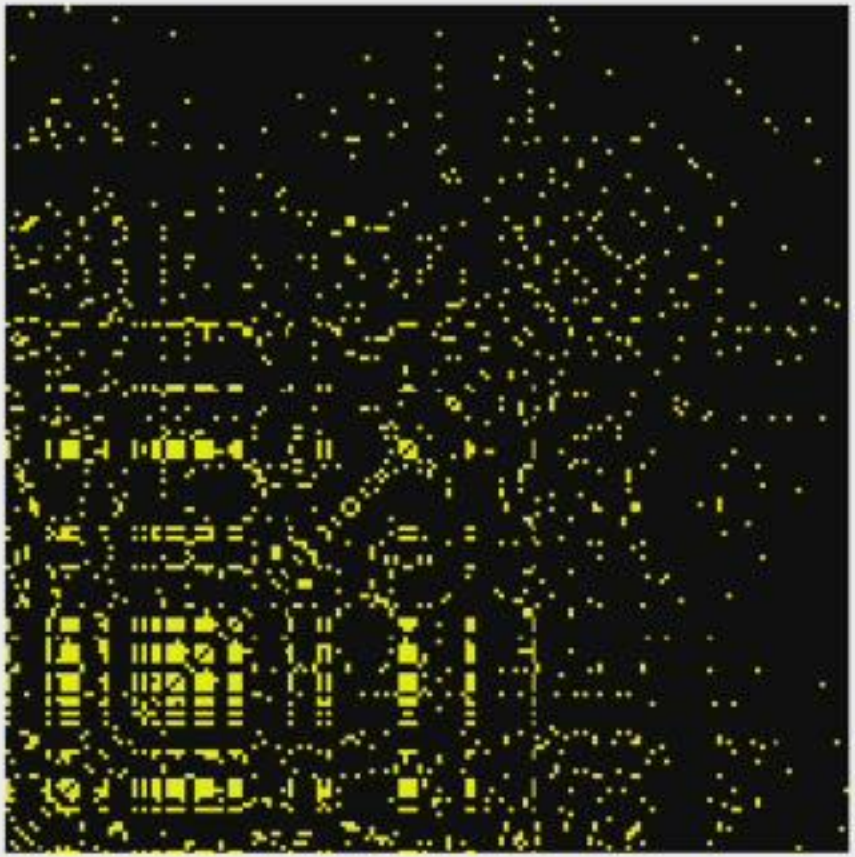
0.0

What looks like an interaction matrix for a modular case

- Apoptosis and DNA repair 180 genes

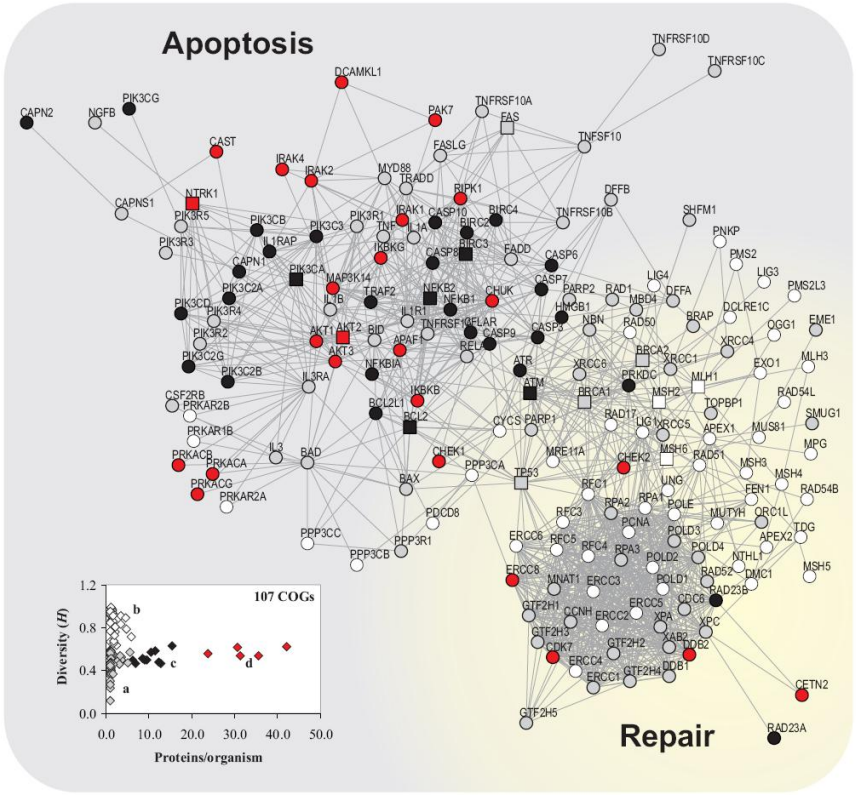


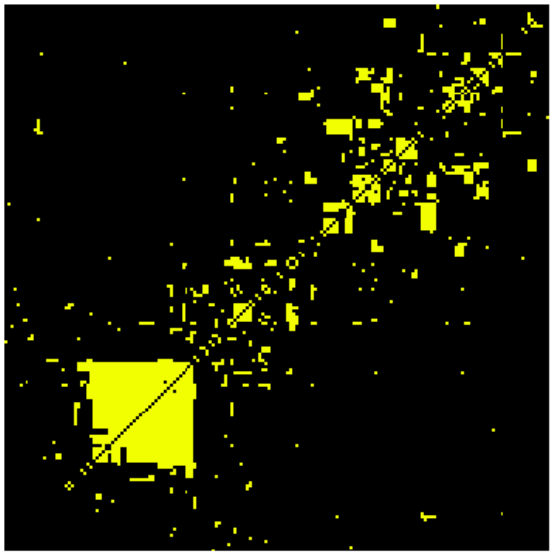




Como seria uma matriz para um caso modular?

- Redes apoptosis e repara de DNA: 180 genes.



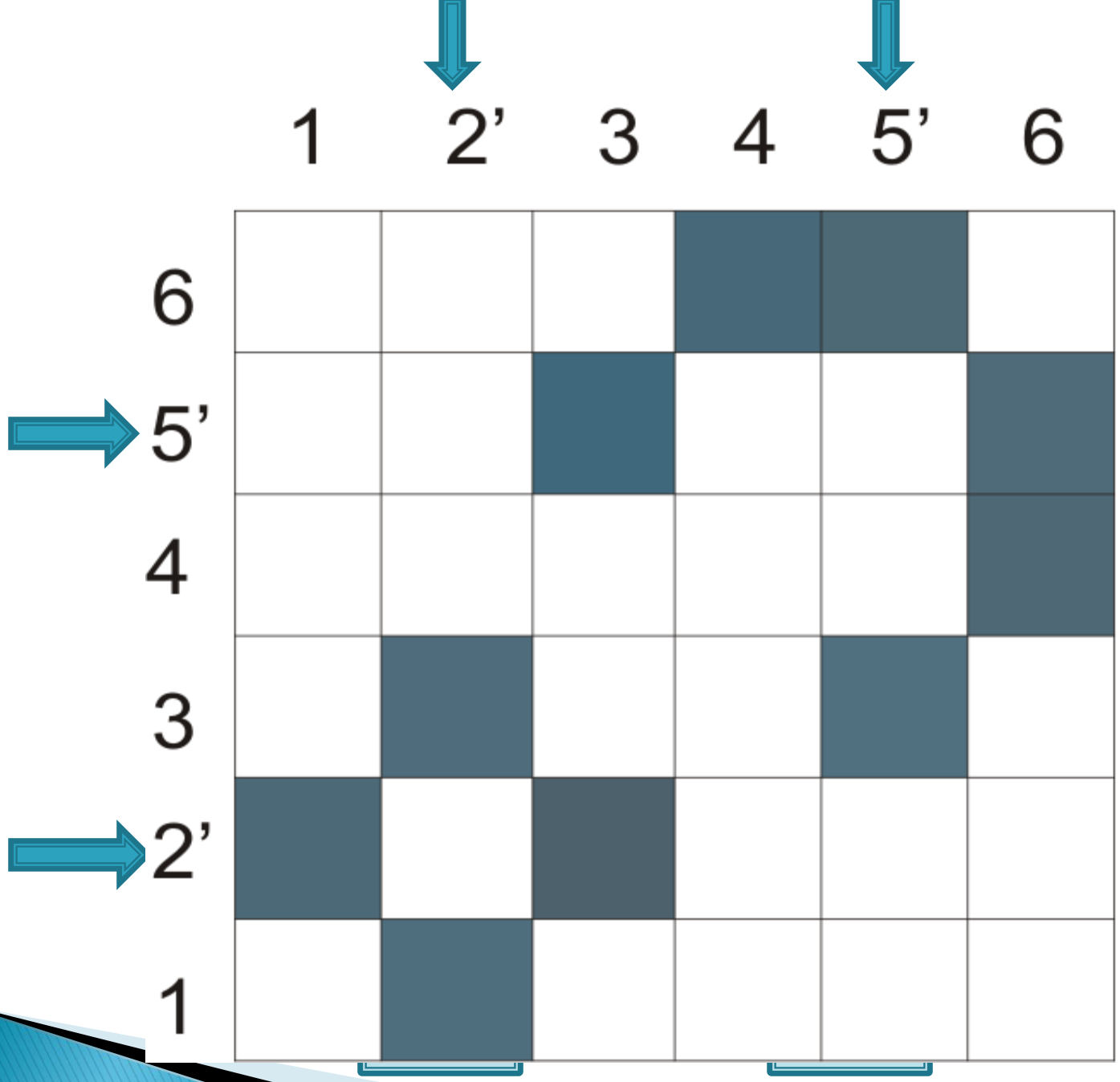


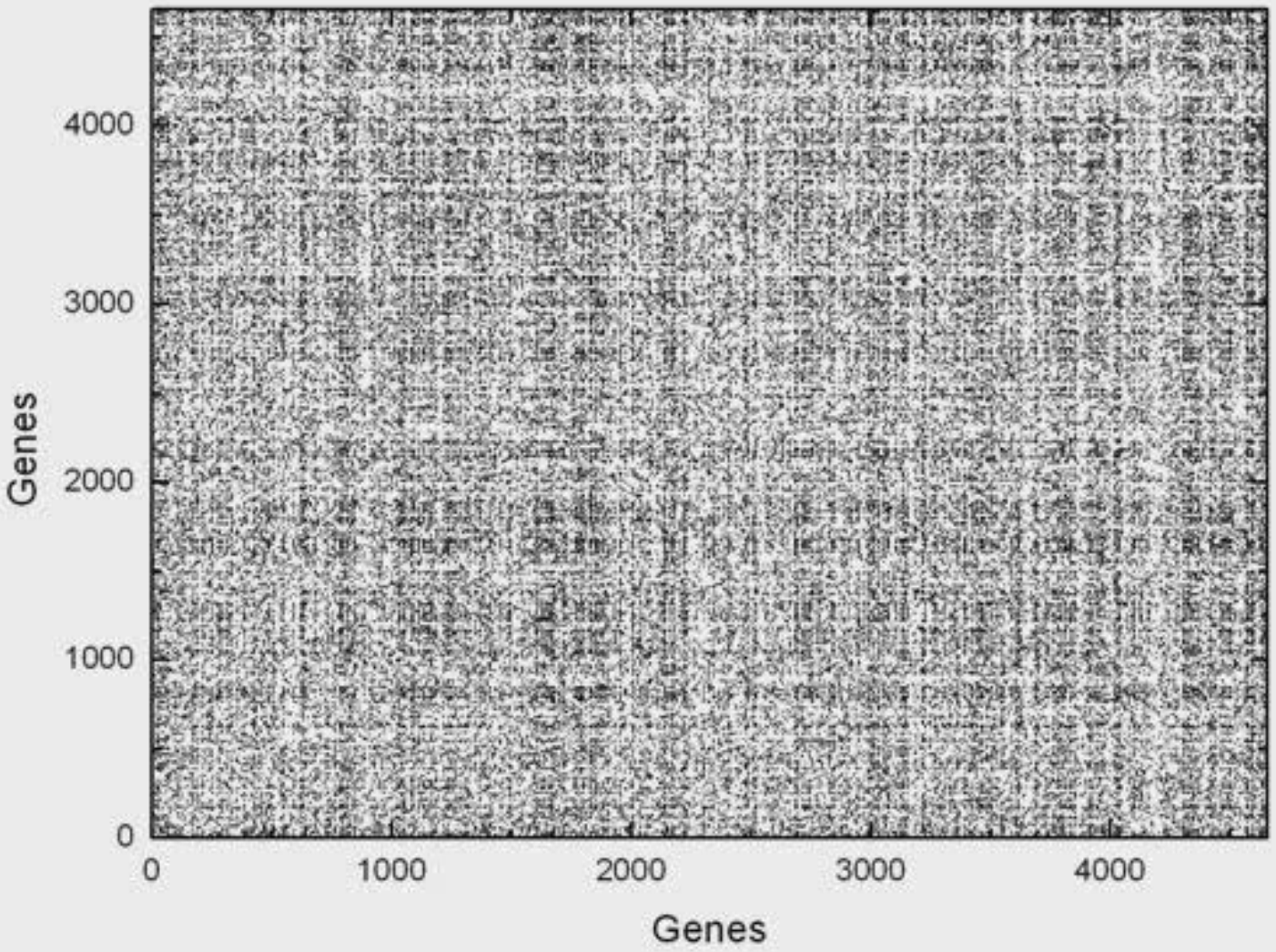
Strategy

- Energy of a matrix configuration:

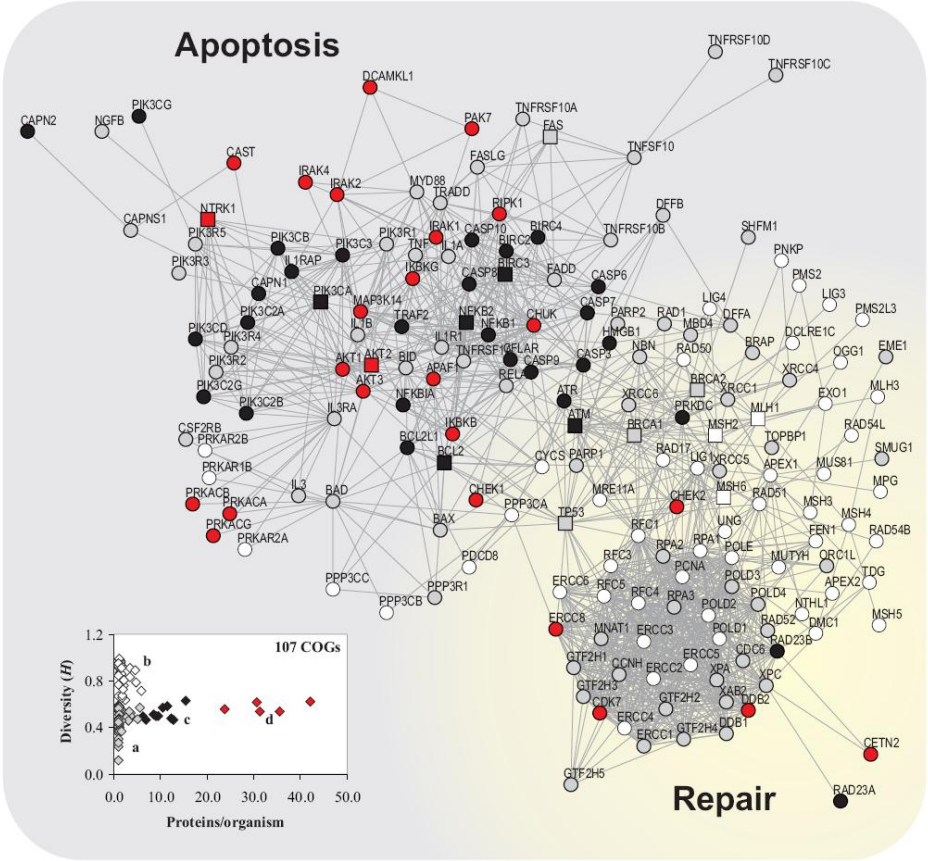
$$E = \sum_{i=1} \sum_{j=1} d_{ij} \{ \left| M_{i,j} - M_{i+1,j} \right| + \left| M_{i,j} - M_{i-1,j} \right| + \left| M_{i,j} - M_{i,j+1} \right| + \left| M_{i,j} - M_{i,j-1} \right| \}$$

- Randomly choose two proteins, swap their positions and recalculate the energy.
- If the swapping lowers the energy, accept the change
- Repeat.


1
2’
3
4
5’
6
6
5’
4
3
2’
1


4000
3000
2000
1000
0
Genes
0
1000
2000
3000
4000
Genes

Modularity



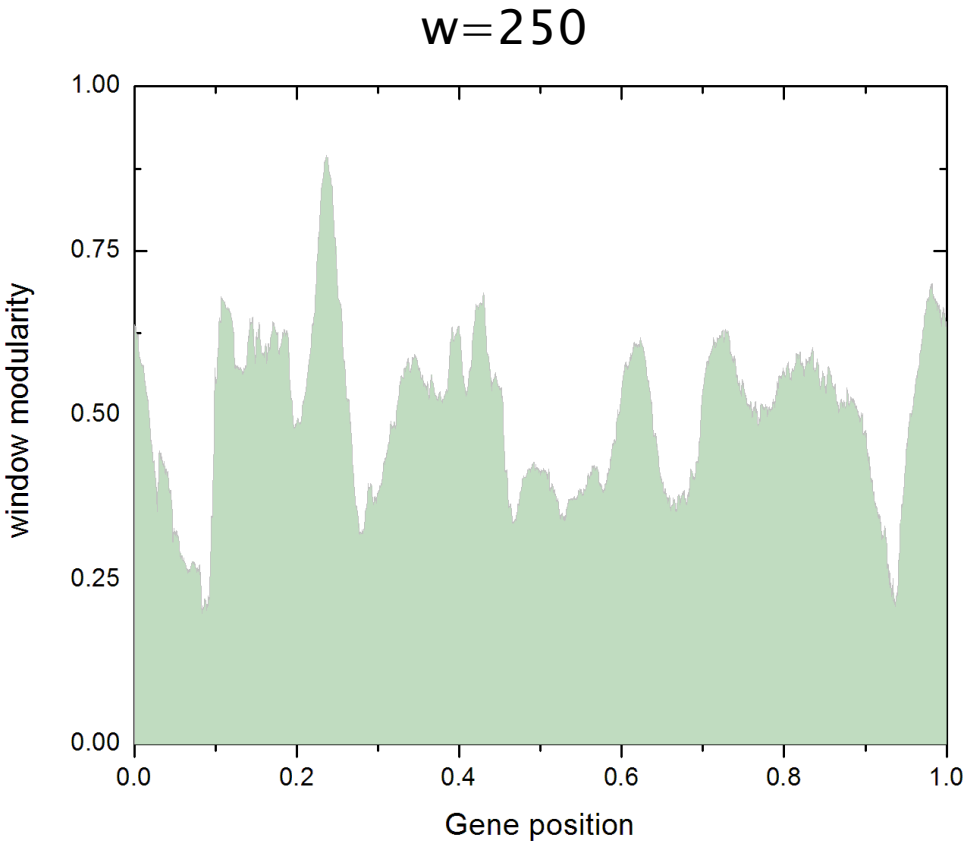
Given a gene set, Modularity is defined as the fraction

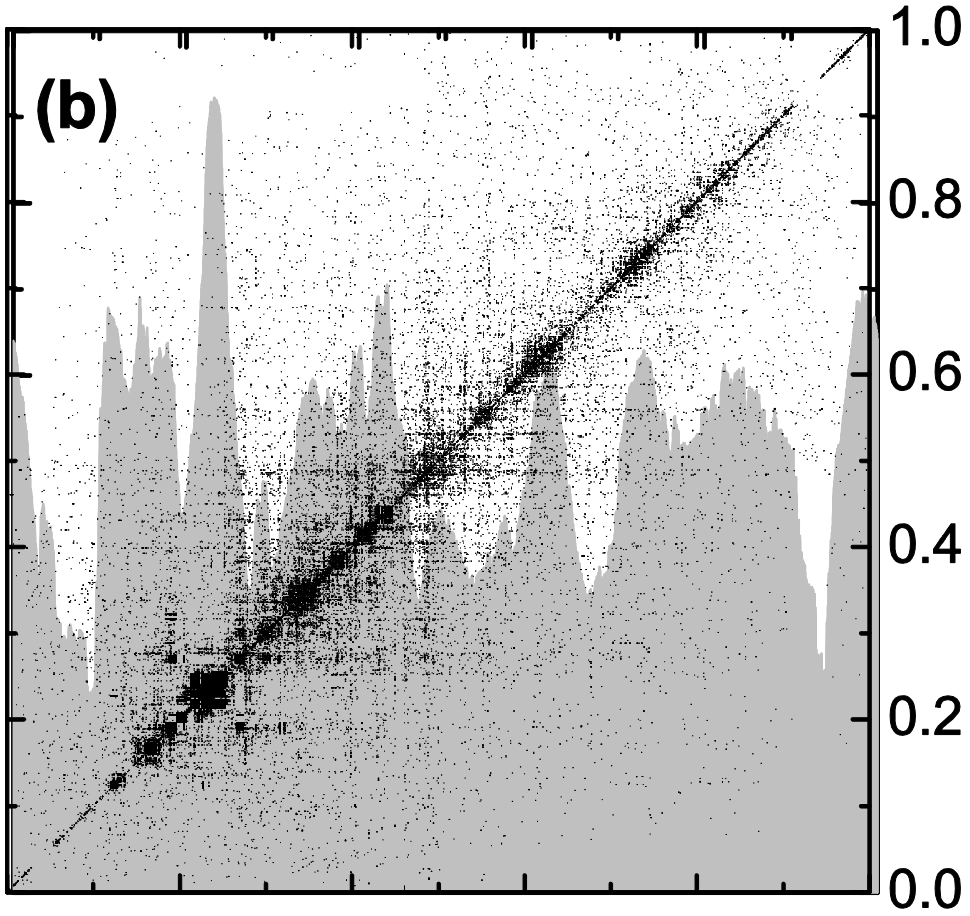
Number of links between two genes in the set

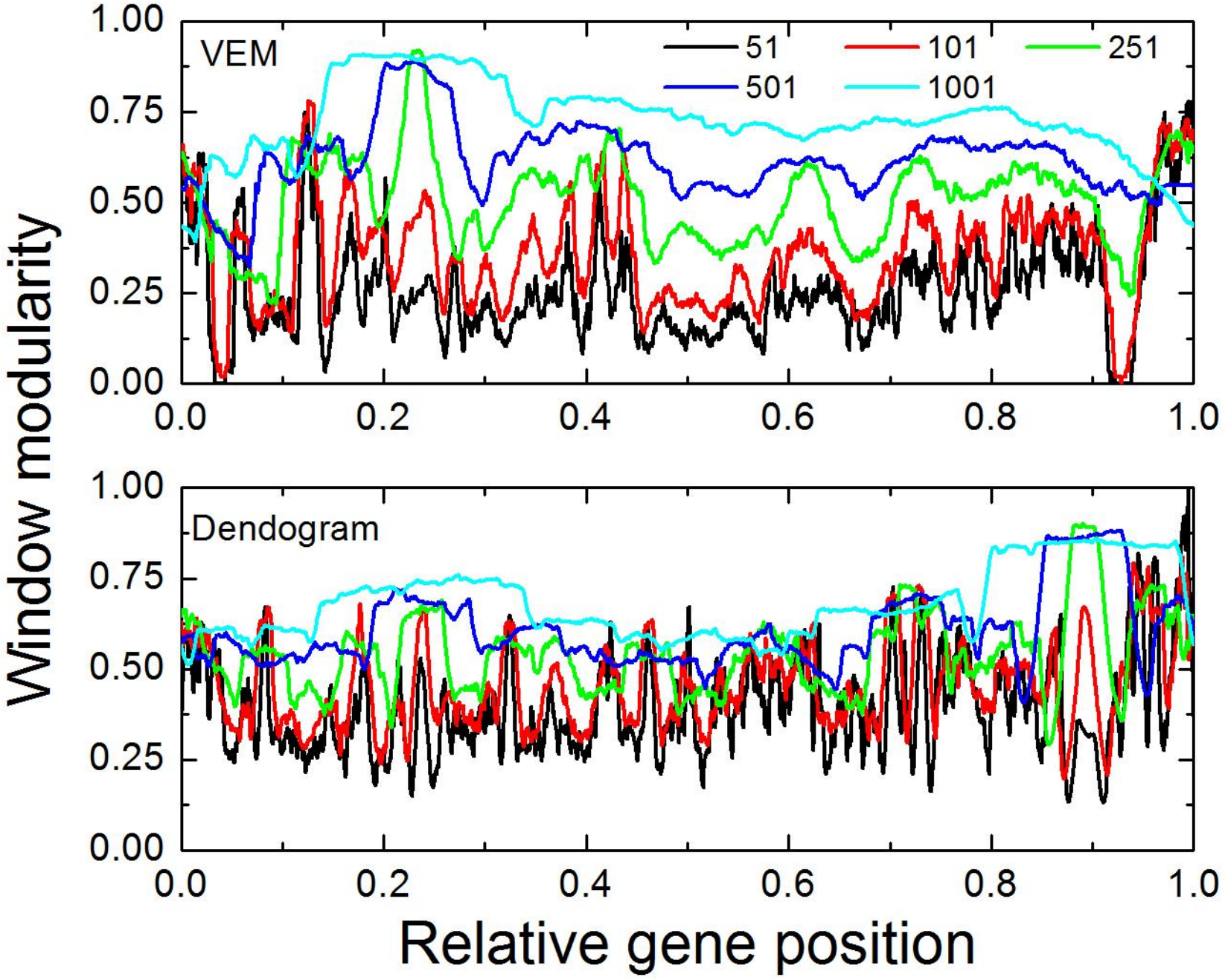
Number of links involving at least one gene in the set

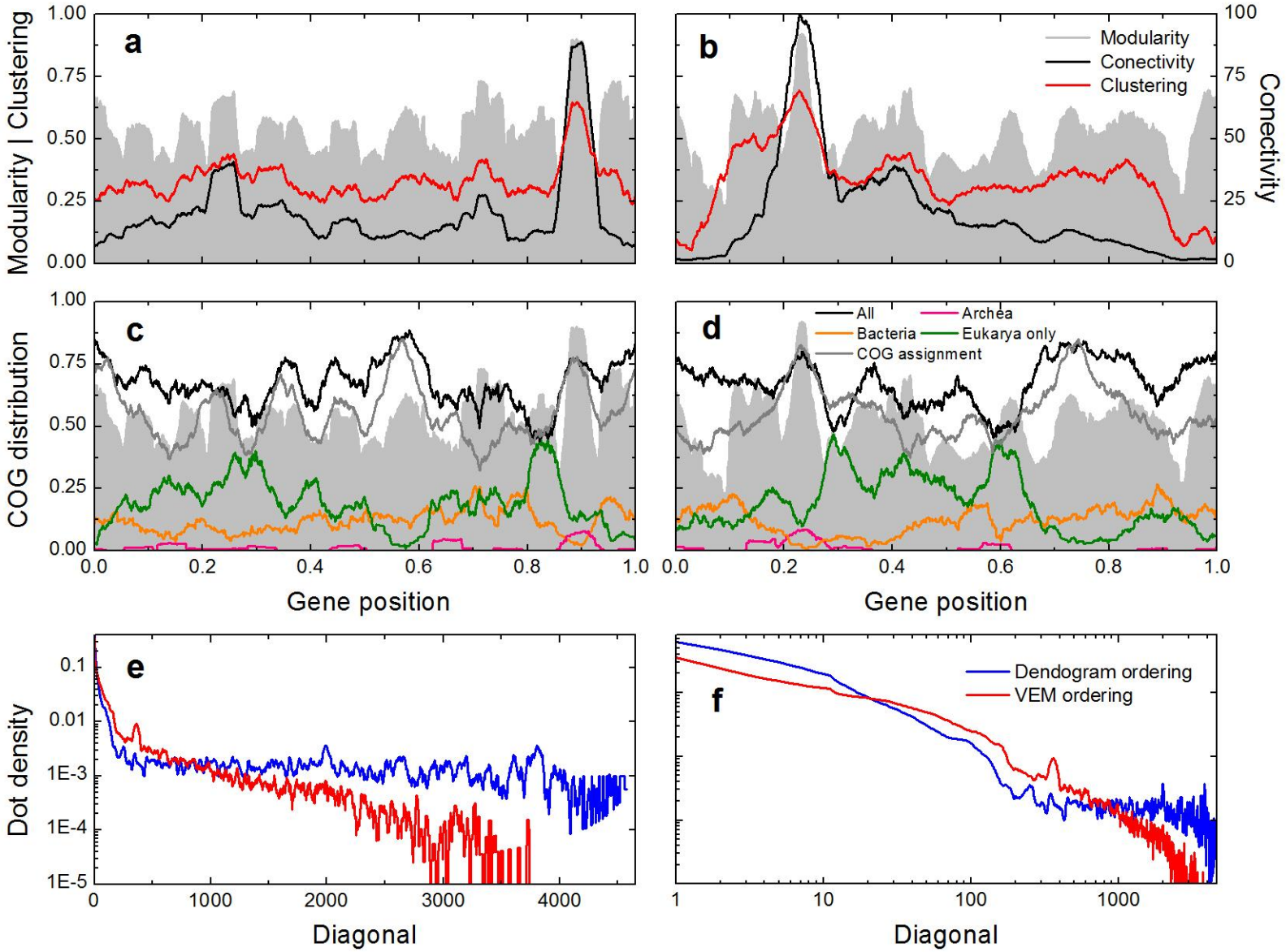
Window modularity

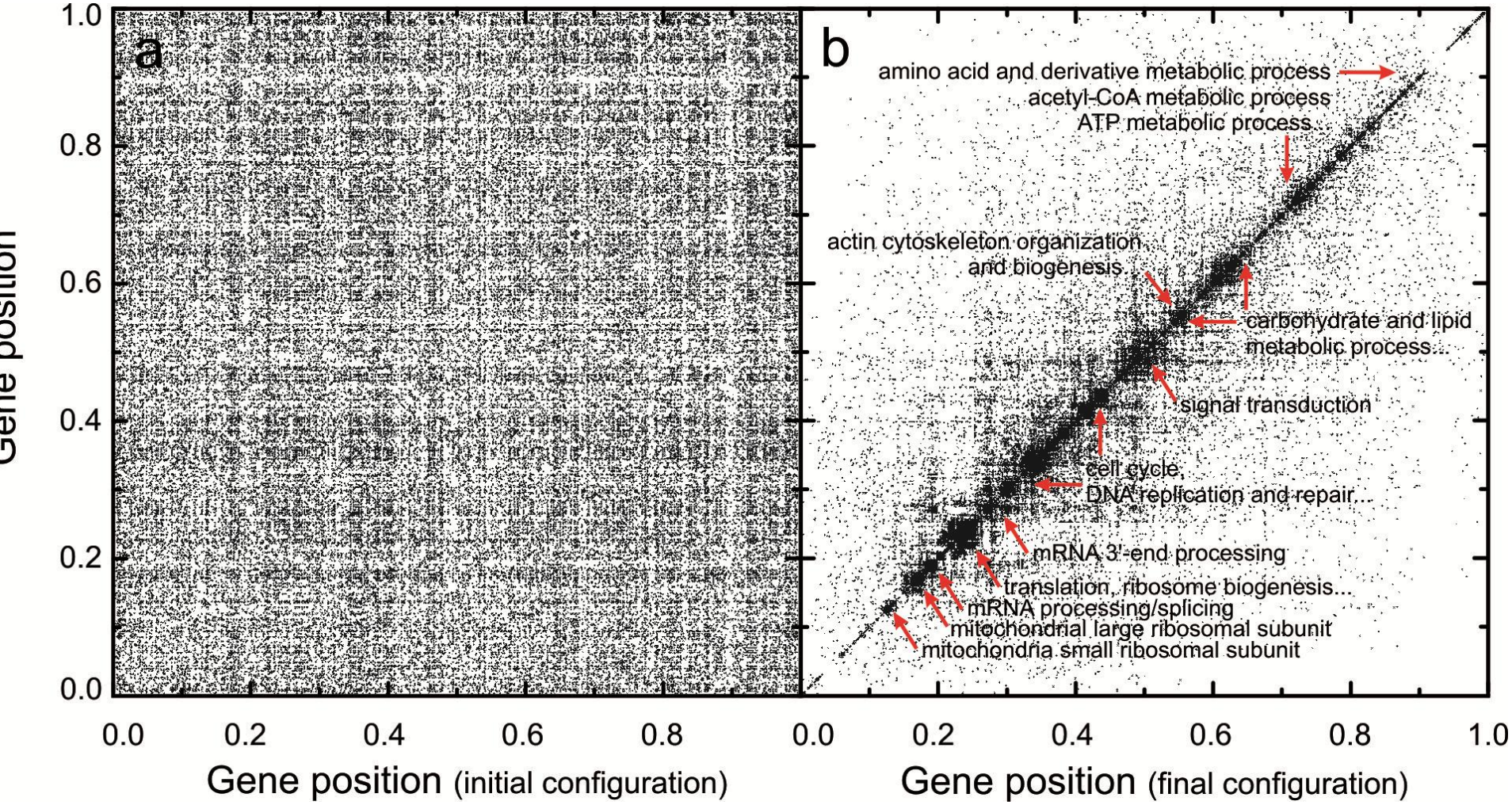
- The gene set is given by the w/2 neighbors to the right and to the lef of each gene.

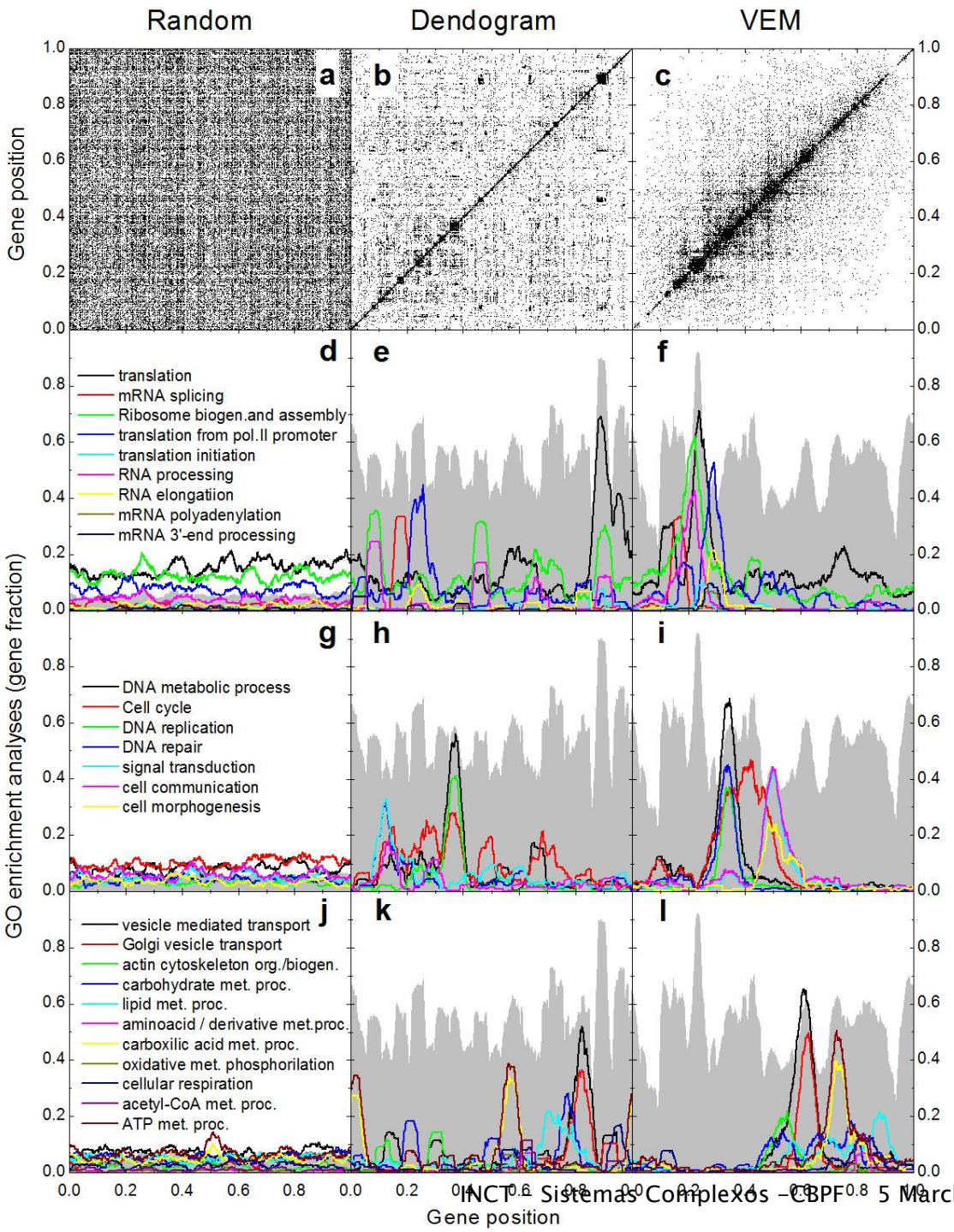


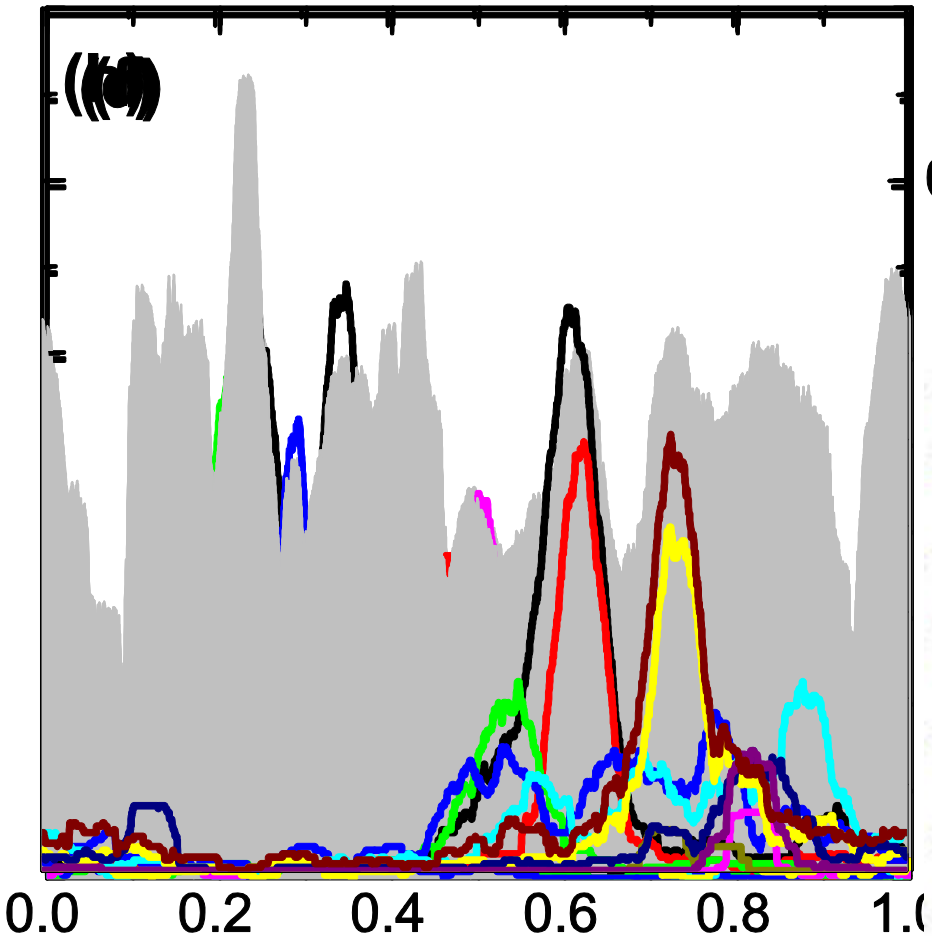

(b)
1.0
0.8
0.6
0.4
0.2
0.0

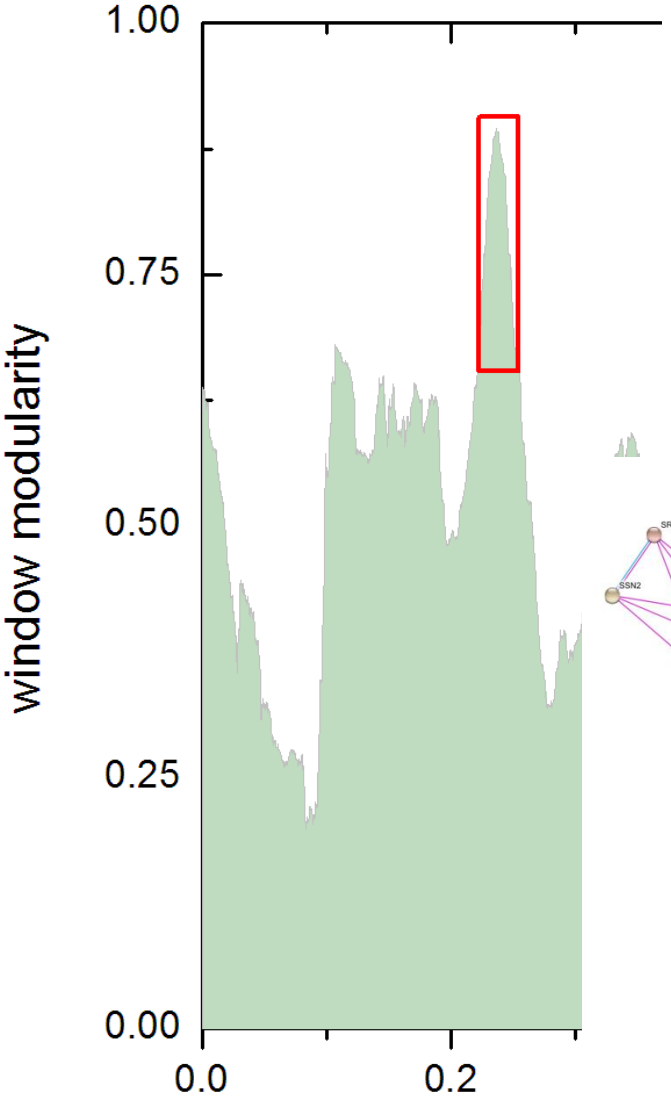

VEM
51
101
251
501
1001
Dendogram
Window modularity
Relative gene position
1.00
0.75
0.50
0.25
0.00
0.0
0.2
0.4
0.6
0.8
1.0

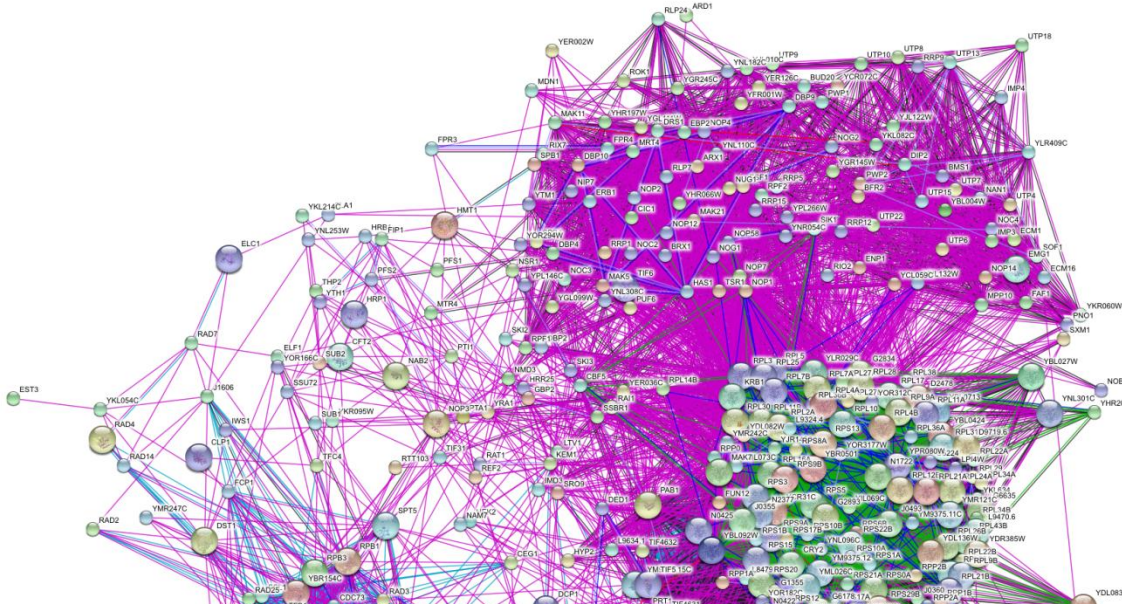

a
b
c
d
e
f
Modularity | Clustering
Conectivity
Modularity
Conectivity
Clustering
COG distribution
Gene position
Gene position
All
Archea
Bacteria
Eukarya only
COG assignment
Dot density
Diagonal
Diagonal
Dendogram ordering
VEM ordering

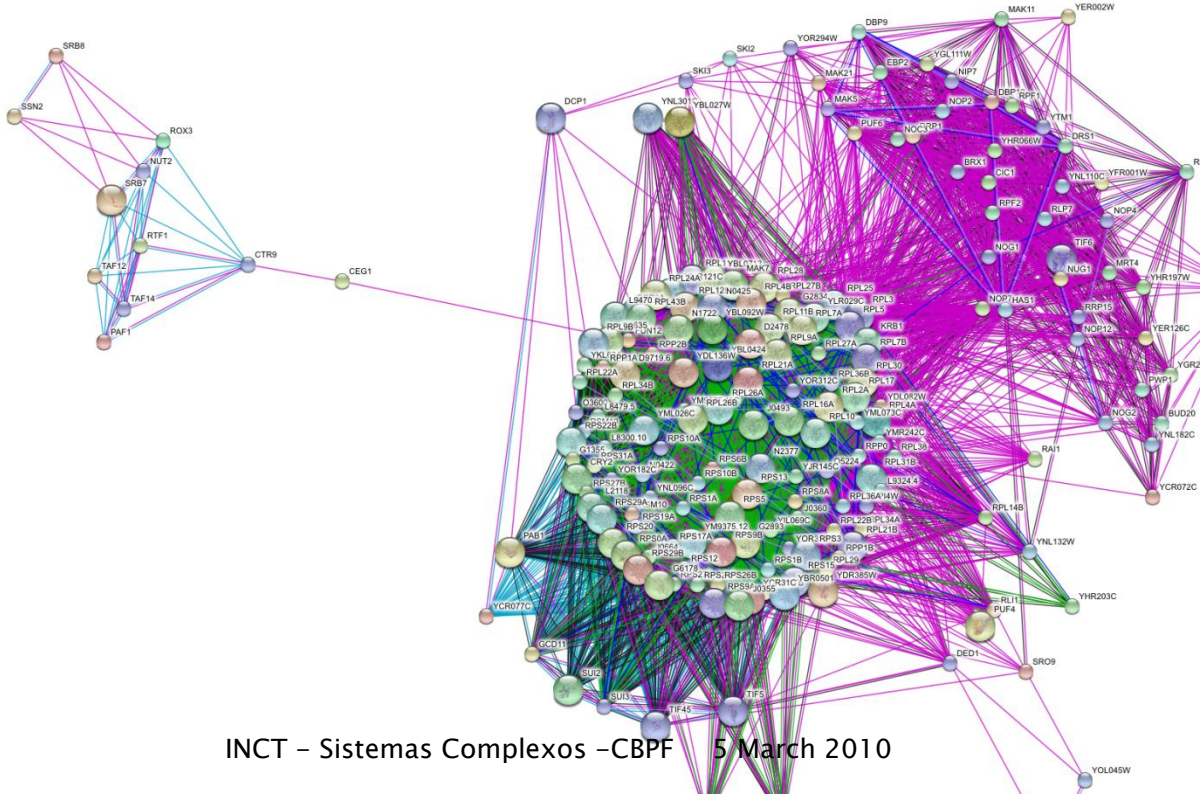

a
b
amino acid and derivative metabolic process
acetyl-CoA metabolic process
ATP metabolic process
actin cytoskeleton organization and biogenesis
carbohydrate and lipid metabolic process
signal transduction
cell cycle, DNA replication and repair...
mRNA 3'-end processing
translation, ribosome biogenesis...
mRNA processing/splicing
mitochondrial large ribosomal subunit
mitochondria small ribosomal subunit
Gene position (initial configuration)
Gene position (final configuration)
Gene position
0.0
0.2
0.4
0.6
0.8
1.0

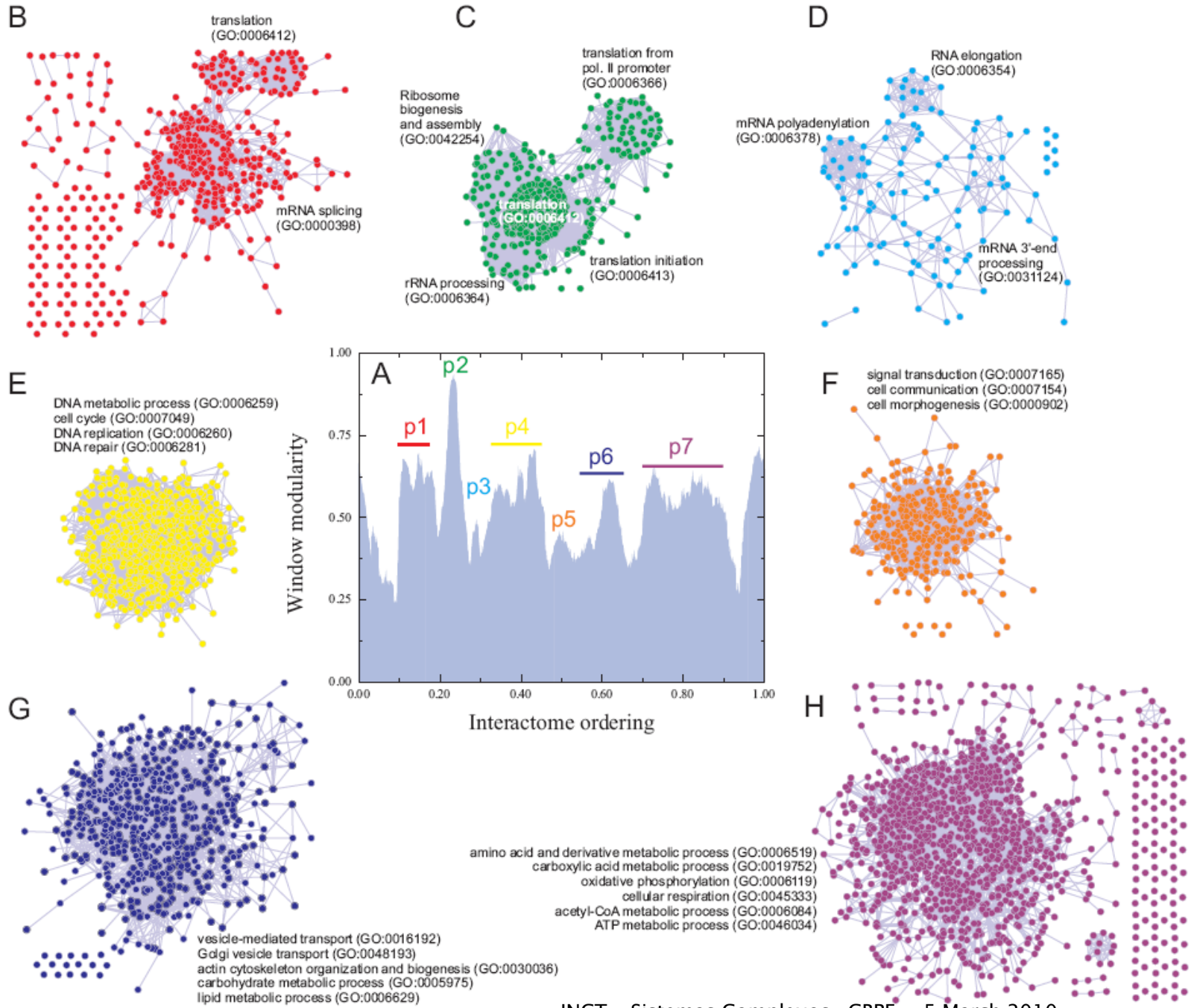

Random
Dendogram
VEM
a
b
c
d
e
f
g
h
i
j
k
l
Gene position
GO enrichment analyses (gene fraction)
translation
mRNA splicing
Ribosome biogen.and assembly
translation from pol.II promoter
translation initiation
RNA processing
RNA elongatiion
mRNA polyadenylation
mRNA 3'-end processing
DNA metabolic process
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
Gene position

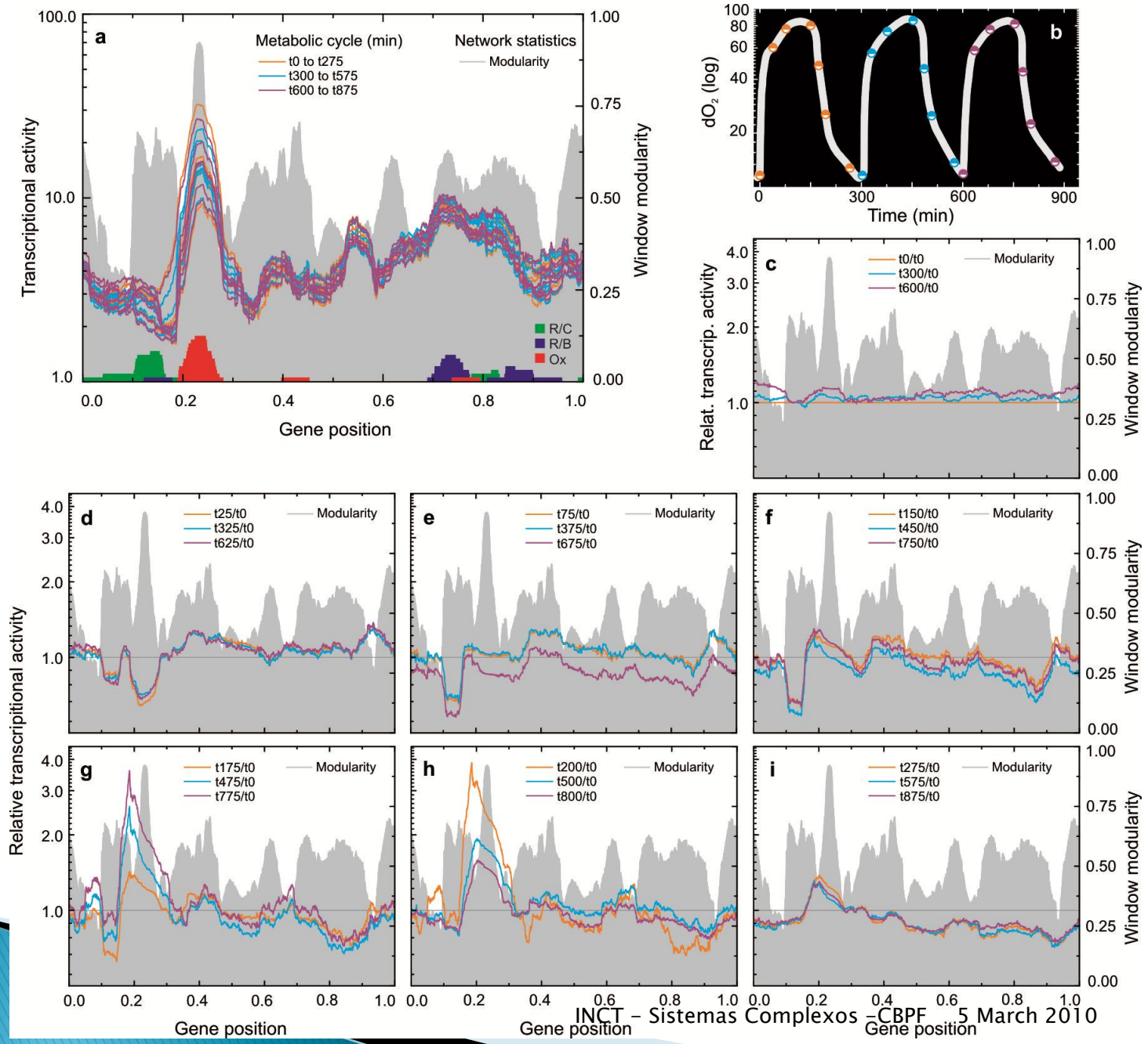

0.0
0.2
0.4
0.6
0.8
translation
mRNA splicing
Ribosome biogen.and assembly
translation from pol.II promoter
translation initiation
RNA processing
vesicle mediated transport
Golgi vesicle transport
actin cytoskeleton org./biogen.
carbohydrate met. proc.
lipid met. proc.
aminoacid / derivative met.proc
carboxilic acid met. proc.
oxidative met. phosphorilation
cellular respiration
acetyl-CoA met. proc.
ATP met. proc.

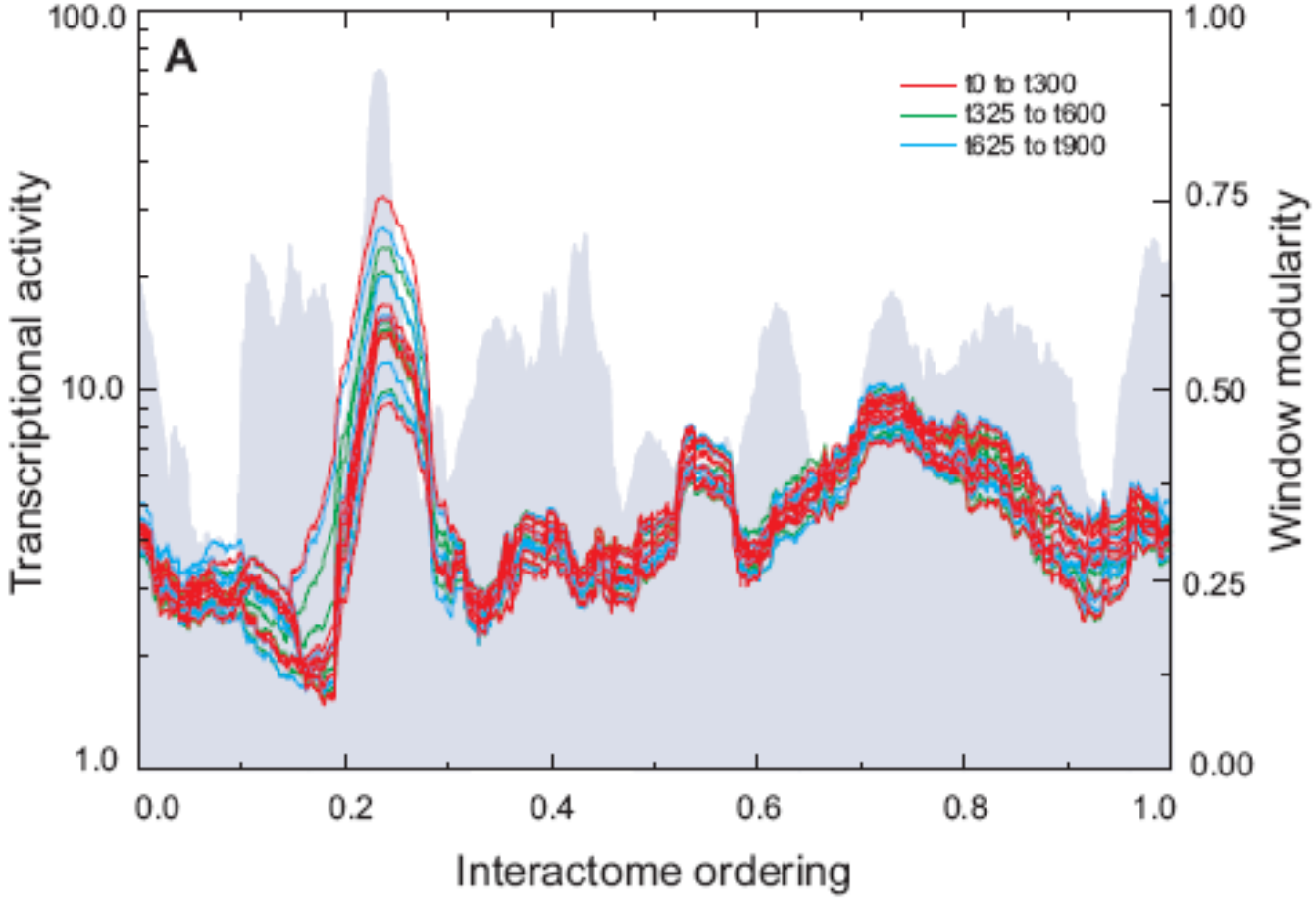

window modularity
1.00
0.75
0.50
0.25
0.00
0.0
0.2

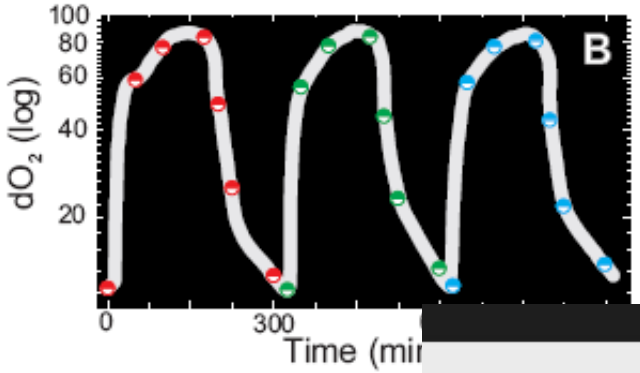




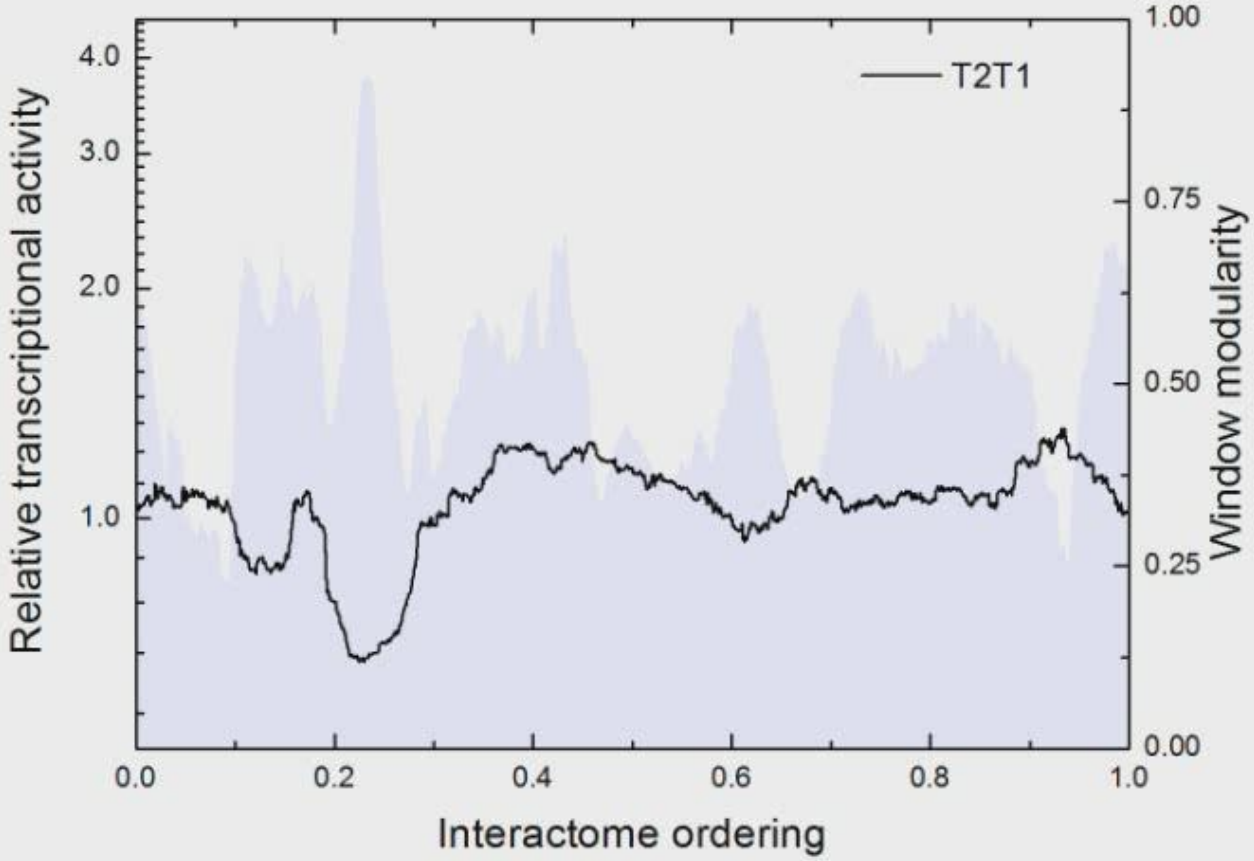

B
translation (GO:0006412)
mRNA splicing (GO:0000398)
C
translation from pol. II promoter (GO:0006366)
Ribosome biogenesis and assembly (GO:0042254)
translation (GO:0006412)
translation initiation (GO:0006413)
rRNA processing (GO:0006364)
D
RNA elongation (GO:0006354)
mRNA polyadenylation (GO:0006378)
mRNA 3'-end processing (GO:0031124)
E
DNA metabolic process (GO:0006259)
cell cycle (GO:0007049)
DNA replication (GO:0006260)
DNA repair (GO:0006281)
A
p1
p2
p3
p4
p5
p6
p7
Window modularity
Interactome ordering
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
vesicle-mediated transport (GO:0016192)
Golgi vesicle transport (GO:0048193)
actin cytoskeleton organization and biogenesis (GO:0030036)
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)

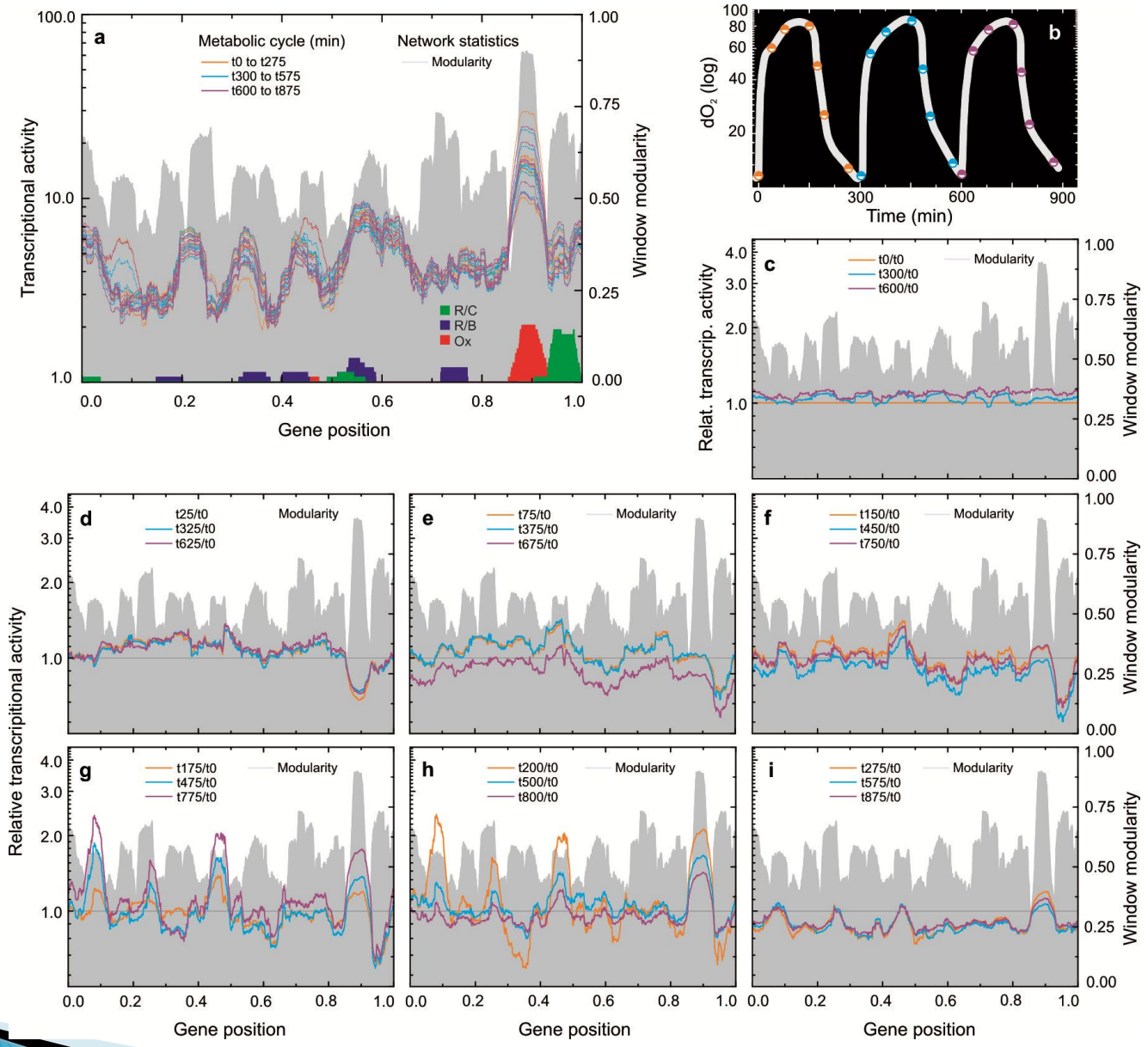

a
Metabolic cycle (min)
t0 to t275
t300 to t575
t600 to t875
Network statistics
Modularity
R/C
R/B
Ox
Transcriptional activity
Window modularity
Gene position
b
dO2 (log)
Time (min)
c
t0/t0
t300/t0
t600/t0
Modularity
Relat. transcrip. activity
Window modularity
d
t25/t0
t325/t0
t625/t0
e
t75/t0
t375/t0
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 transcripitional activity
Gene position


A
t0 to t300
t325 to t600
t625 to t900
100.0
10.0
1.0
Transcriptional activity
1.00
0.75
0.50
0.25
0.00
Window modularity
0.0
0.2
0.4
0.6
0.8
1.0
Interactome ordering



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).




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
Modularity
e
t75/t0
t375/t0
t675/t0
Modularity
f
t150/t0
t450/t0
t750/t0
Modularity
g
t175/t0
t475/t0
t775/t0
Modularity
h
t200/t0
t500/t0
t800/t0
Modularity
i
t275/t0
t575/t0
t875/t0
Modularity
Relative transcripitional activity
Window modularity
Gene position

GNATT: A tool to analyze gene networks

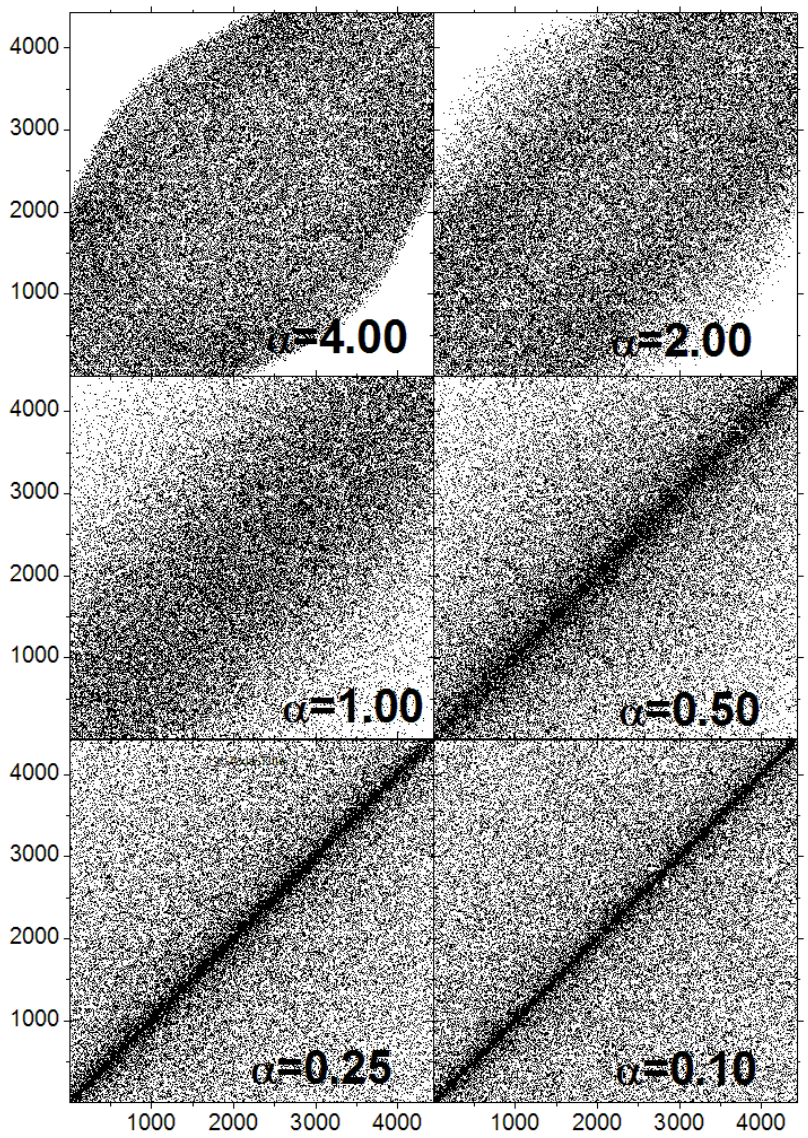
- Considers different databases: NIH/NCBI, KEGG, EMBL, String, etc.
  - Deals with different nomenclatures
  - Constructs partial gene networks
  - Analyzes expression data
  - Searches for functional modules
  - Identifies different functions
  - Globally organizes gene networks
  - Network characterization

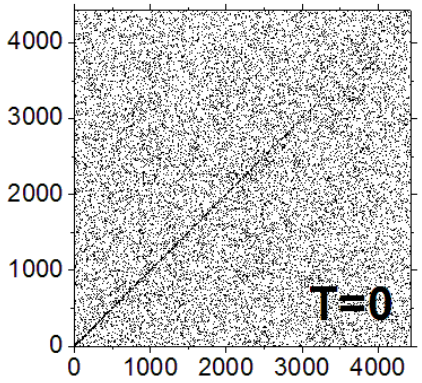
What´s next?

$$E = \sum_{i=1}\sum_{j=1} d_{ij}^{\alpha} \left\{ \left| M_{i,j} - M_{i+1,j} \right| + \left| M_{i,j} - M_{i-1,j} \right| \right.$$

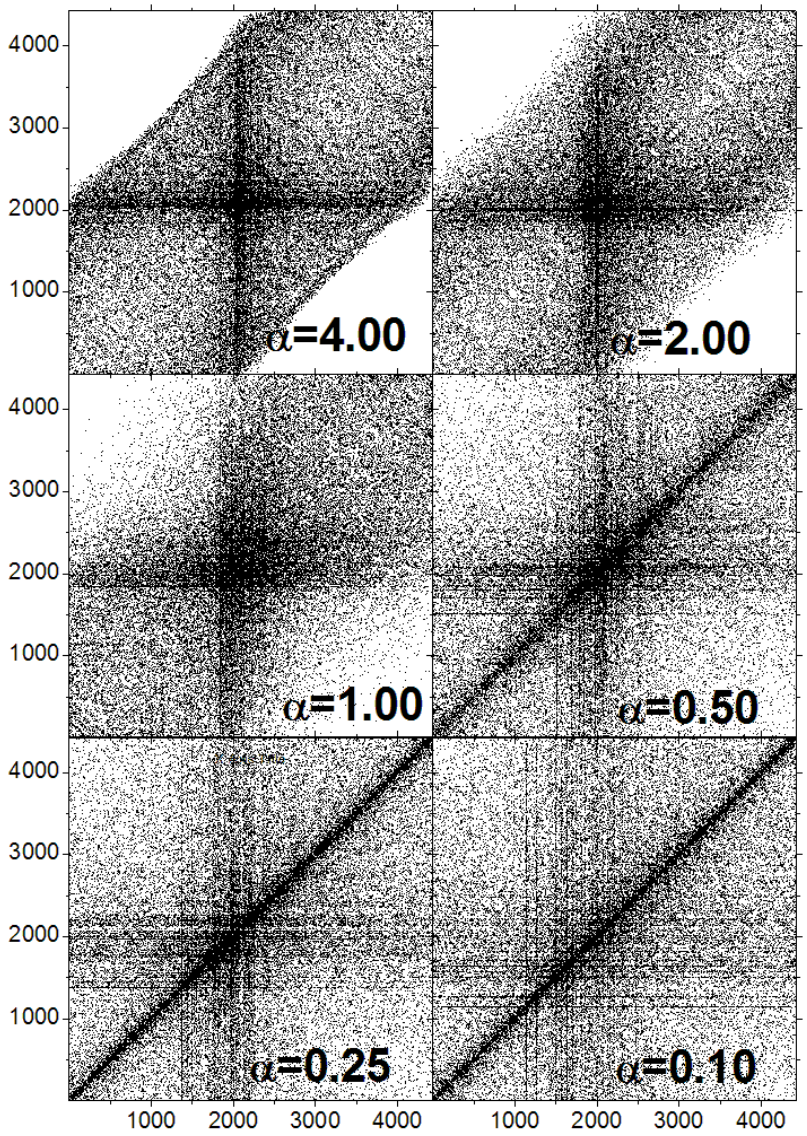
$$\left. + \left| M_{i,j} - M_{i,j+1} \right| + \left| M_{i,j} - M_{i,j-1} \right| \right\}$$

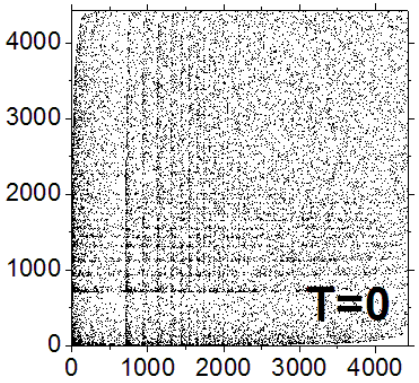
Random


4000
3000
2000
1000
α=4.00
α=2.00
α=1.00
α=0.50
α=0.25
α=0.10
1000
2000
3000
4000

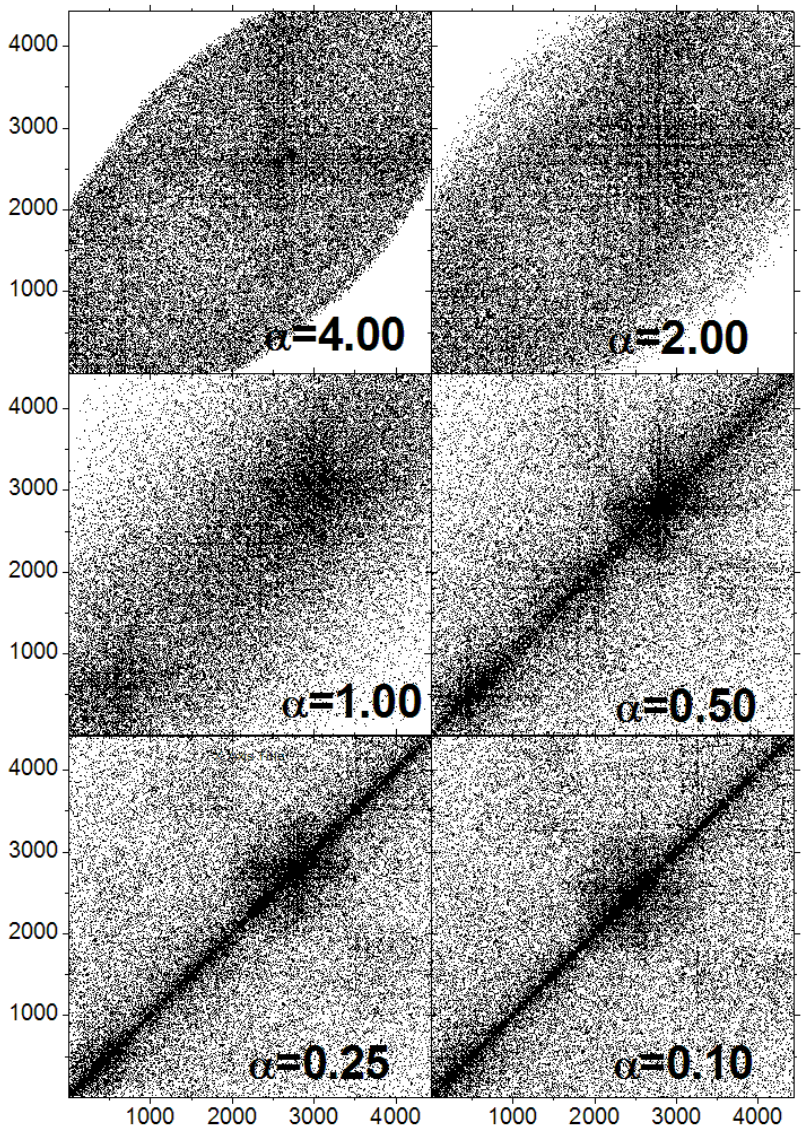

4000
3000
2000
1000
0
T=0
0
1000
2000
3000
4000

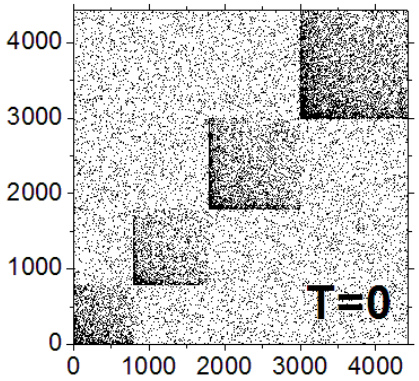
Barabasi


α=4.00
α=2.00
α=1.00
α=0.50
α=0.25
α=0.10
1000
2000
3000
4000

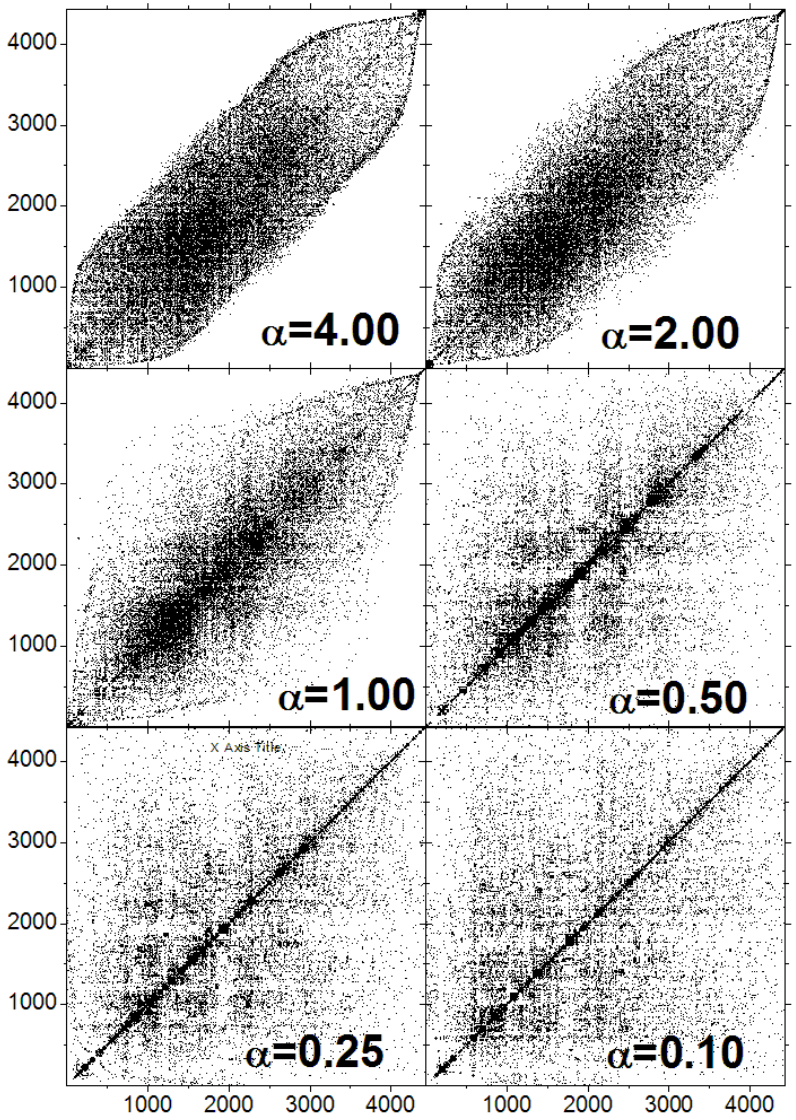

T=0
0
1000
2000
3000
4000

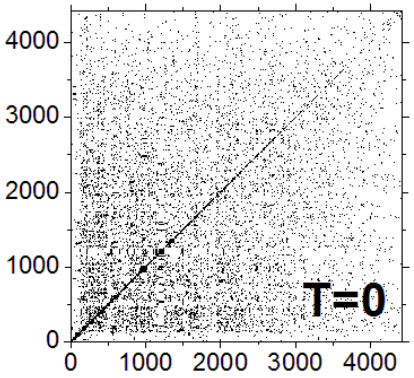
Barabasi-Randomico



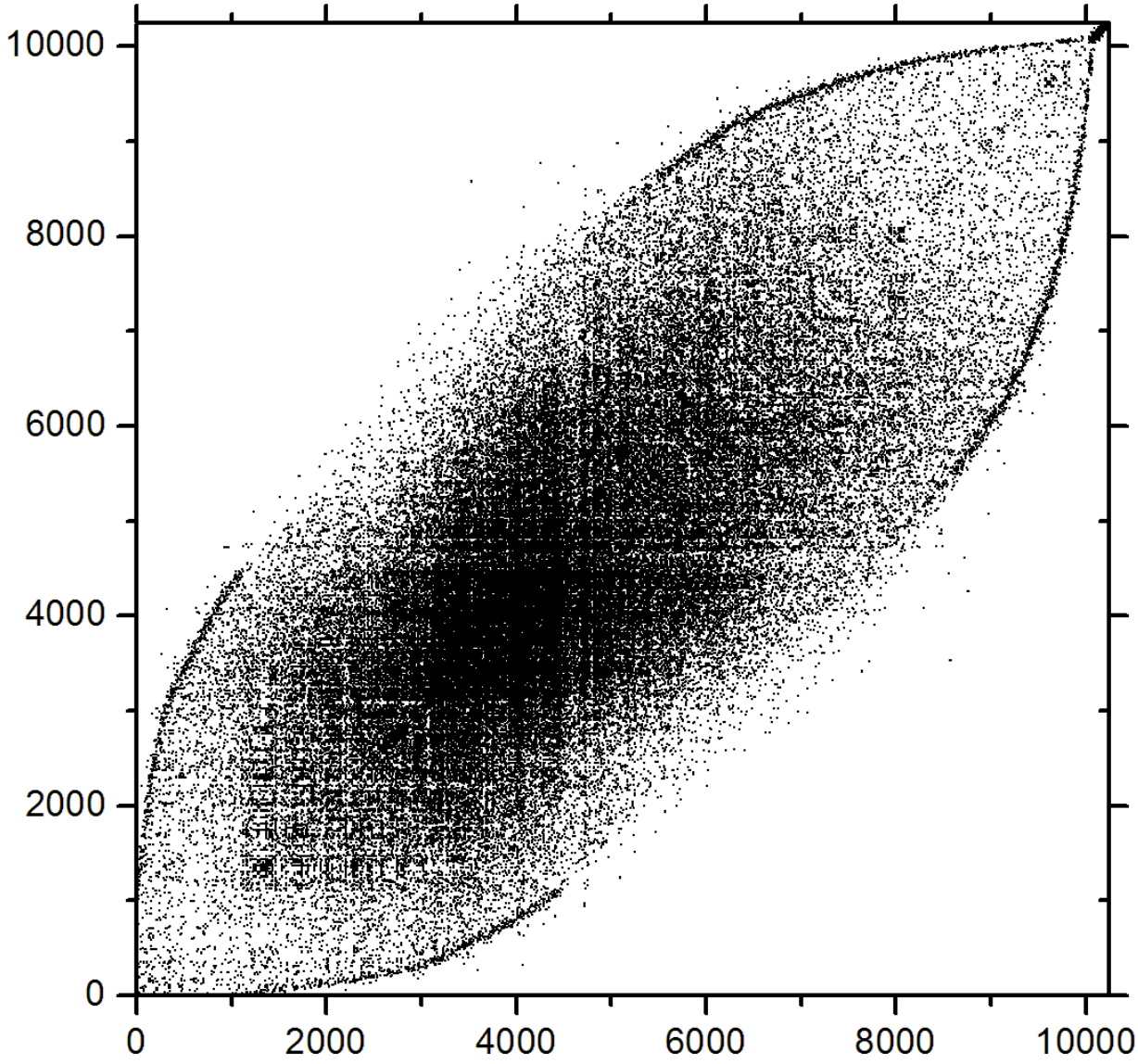


Prod-Saccharomyces

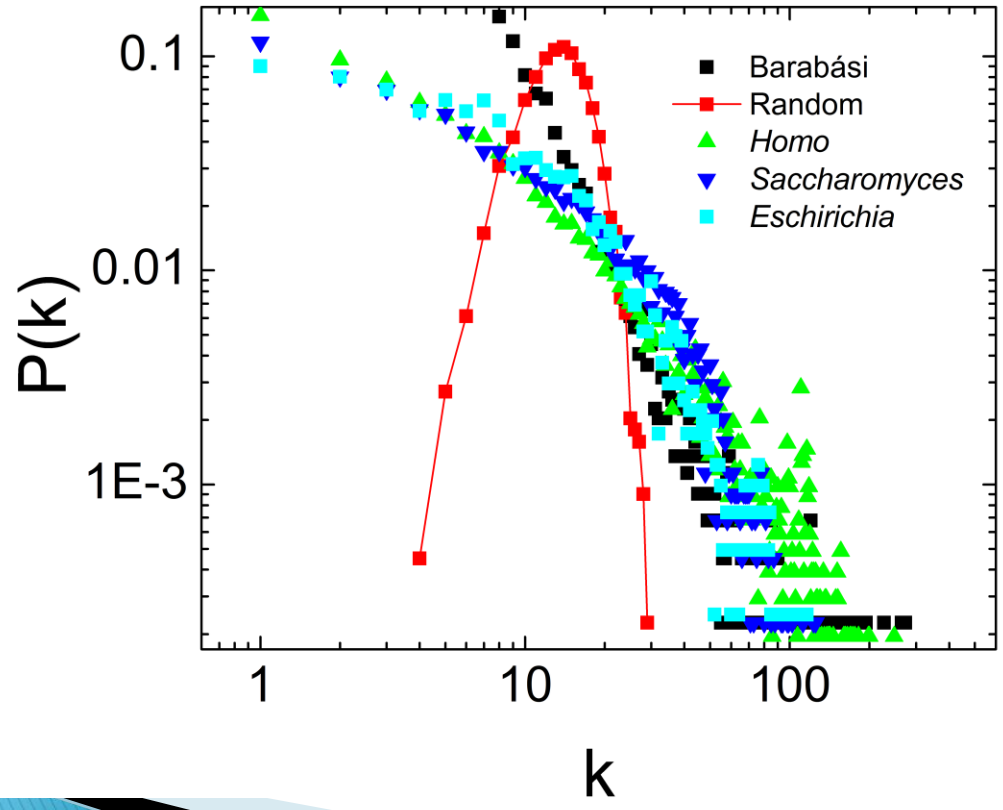




Homo sapiens



Distribuição de grau de conectividade



References

- Castro MAA, Onsten TTG, de Almeida RMC, Moreira JCF. Journal of Theoretical Biology 234:487-495. 2005.
- Mauro A.A. Castro; Verônica A. Grieneisen; Rita M. C.de Almeida, Cell Biology International v. 29, n. 11, p. 929-931 (2005)
- M. A. A. Castro, T. T. G. Onsten, J. C. F Moreira, and R. M. C de Almeida. Mutation Research. Fundamental and Molecular Mechanisms of Mutagenesis, v. 600, p. 150-164, 2006.
- Mauro A. A. Castro, José C. M. Mombach, Rita M. C. de Almeida, and José C. F. Moreira, Nucleic Acids Res. 2007; **35** (6):1859-67. Epub 2007 Mar 1.
- Mauro A. A. Castro[1,3*], Rodrigo J. S. Dalmolin[1*], José C. F. Moreira[1], José C. M. Mombach[4] & Rita M. C. de Almeida[2] , *Evolutionary origins of human apoptosis and genome stability gene networks,* to appear.
- ViaComplex: software for landscape analysis of gene expression networks in genomic context, Mauro A. A. Castro, José L. Rybarczyk Filho, Rodrigo J. S. Dalmolin, Marialva Sinigaglia, José C. F. Moreira, José C. M. Mombach e Rita M. C. de Almeida, Bionformatics, accepted for publication (2009).

Thank you !

rita@if.ufrgs.br

http://lief.if.ufrgs.br/pub/biosoftwares/viacomplex

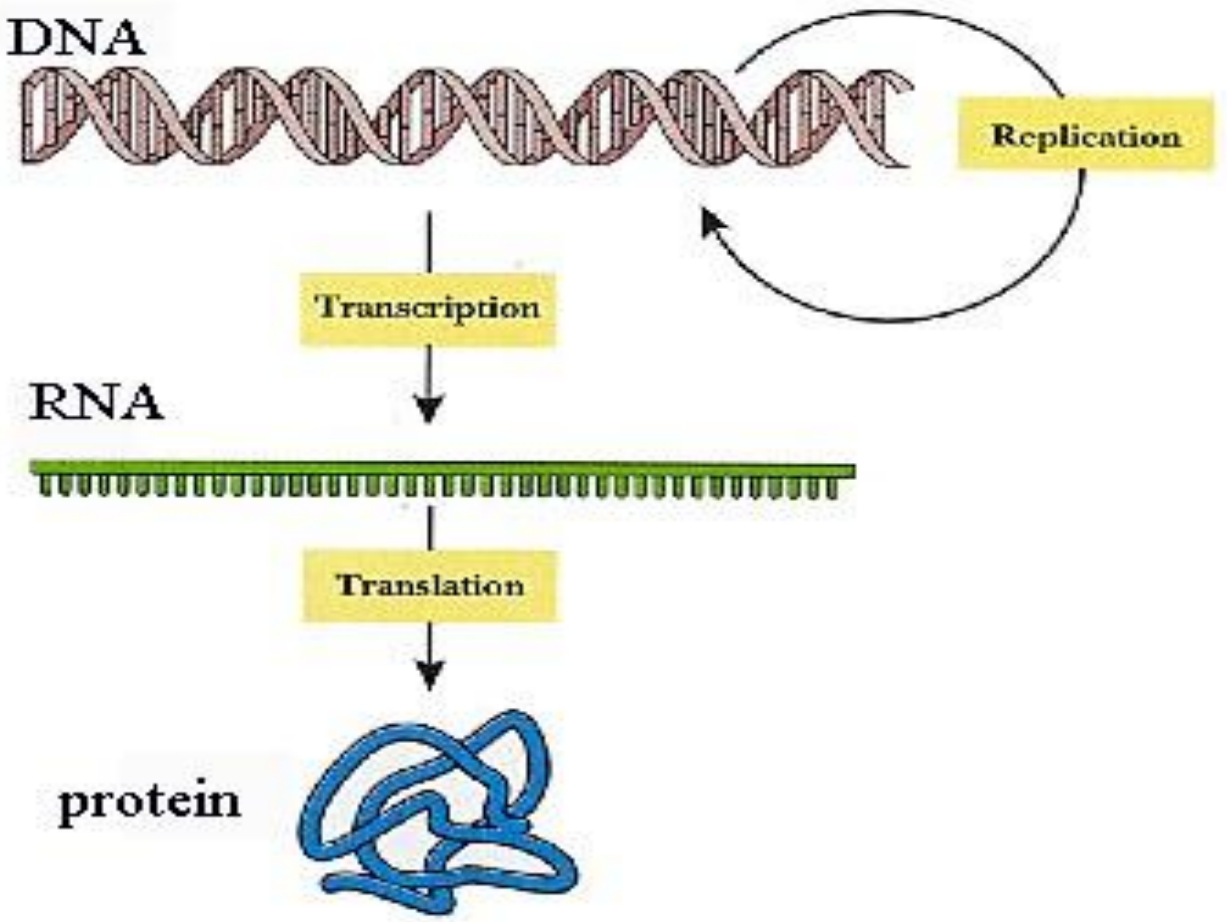
Transcript_mov_fundo_0001.wmv

TRANSCRIPTION

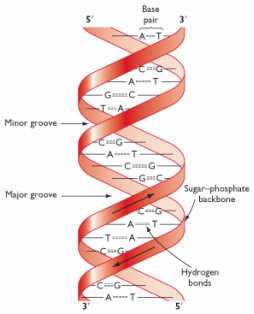
DNA

RNA

PROTEÍNA

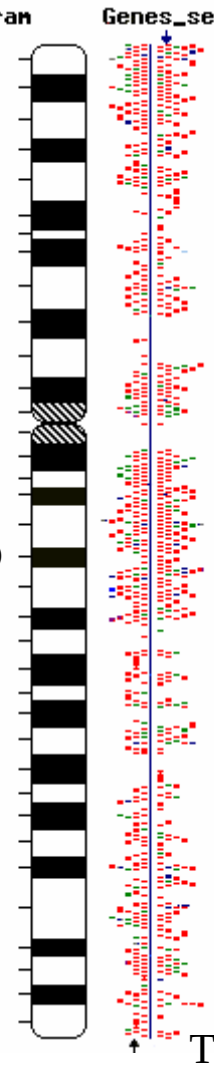


GENOMA PROTEOMA METABOLOMA



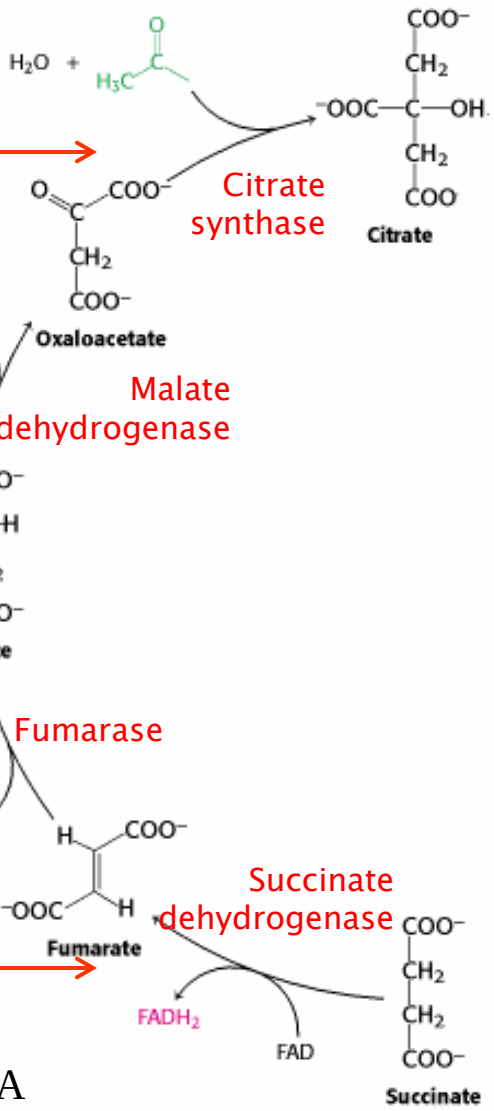
Genoma:
~24.000 genes
(ORFs aprovadas)

Crom. 11:
1543 ORFs



interação prot-gene

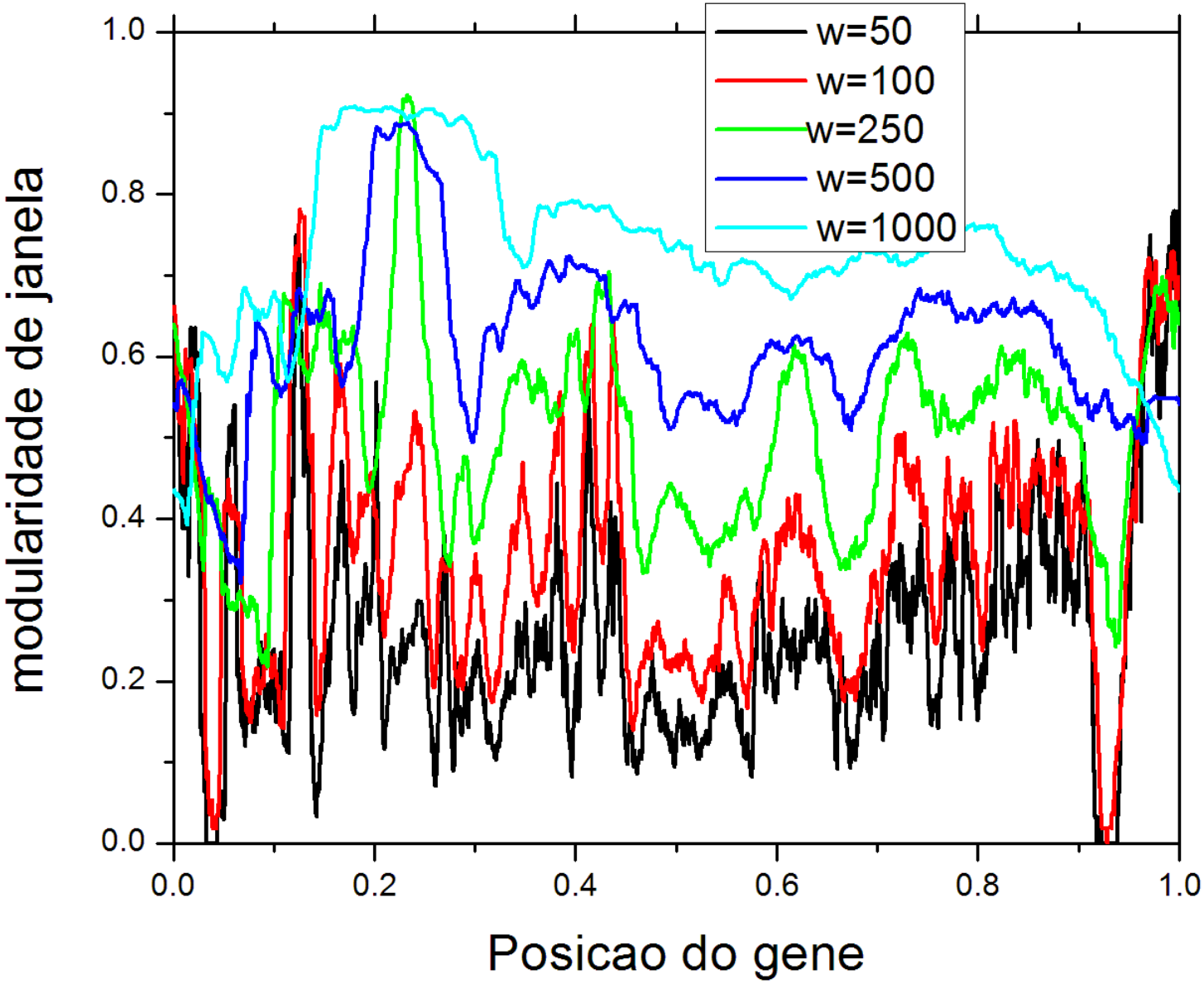
interação prot-prot

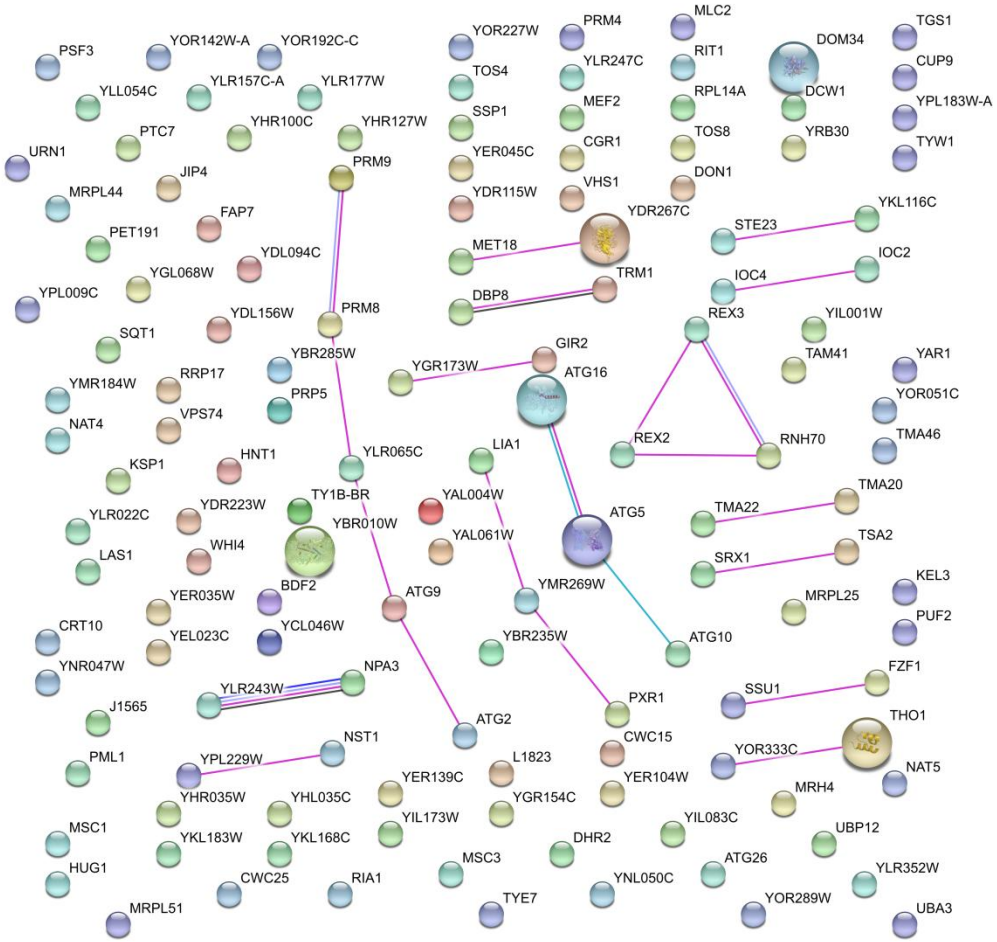
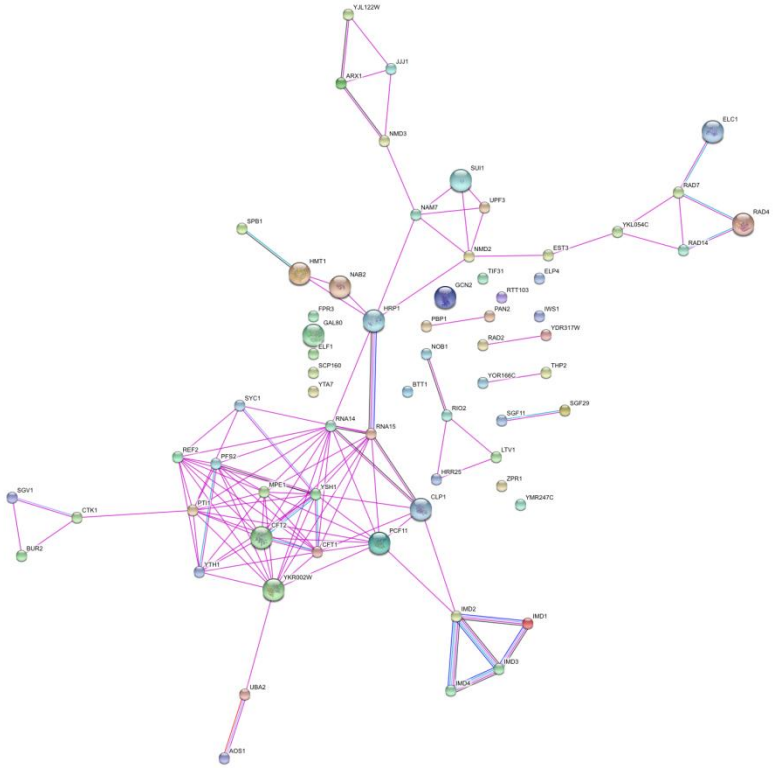
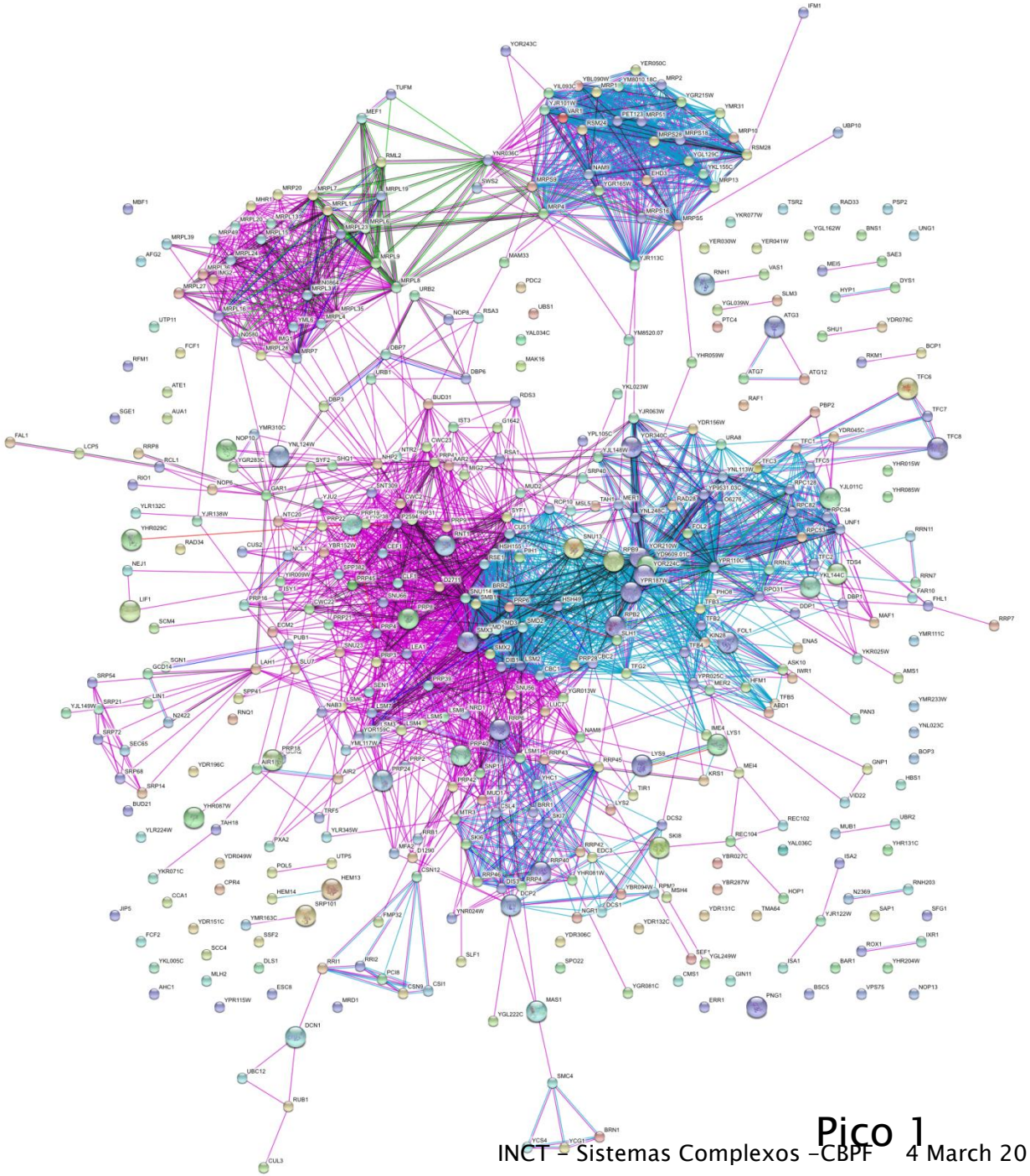


TRANSCRIPTOMA INTERATOMA

Experimental databases.

- Ensembl (Proteins and Genes)
- Uniprot EBI –
- Entrez NCBI
- HUGO
- Saccharomyce cerevisae
  - Ecocyc
  - STRING (EMBL)
  - Affymetrix (microarray)
  - SAGE
  - KEGG
  - Gene Ontology
  - Gene Expression Omnibus


w=50
w=100
w=250
w=500
w=1000
modularidade de janela
Posicao do gene
0.0
0.2
0.4
0.6
0.8
1.0


PSF3
YOR142W-A
YOR192C-C
YLL054C
YLR157C-A
YLR177W
PTC7
YHR100C
YHR127W
URN1
JIP4
PRM9
MRPL44
FAP7
PET191
YDL094C
YGL068W
YPL009C
YDL156W
PRM8
SQT1
YBR285W
YMR184W
RRP17
PRP5
NAT4
VPS74
KSP1
HNT1
YLR065C
YLR022C
YDR223W
TY1B-BR
YBR010W
LAS1
WHI4
BDF2
YER035W
YEL023C
YCL046W
CRT10
YNR047W
NPA3
YLR243W
J1565
PML1
NST1
YPL229W
YHR035W
YHL035C
MSC1
YKL183W
YKL168C
HUG1
CWC25
RIA1
MRPL51
YOR227W
PRM4
MLC2
TOS4
YLR247C
RIT1
SSP1
MEF2
RPL14A
YER045C
CGR1
TOS8
YDR115W
VHS1
DON1
YDR267C
MET18
TRM1
DBP8
YGR173W
GIR2
ATG16
LIA1
YAL004W
YAL061W
ATG5
YMR269W
ATG9
ATG2
YBR235W
PXR1
CWC15
L1823
YER139C
YER104W
YGR154C
YIL173W
MSC3
TYE7
DHR2
YNL050C
ATG26
YOR289W
YIL083C
UBP12
MRH4
YOR333C
THO1
NAT5
SSU1
FZF1
ATG10
TMA22
TMA20
SRX1
TSA2
MRPL25
KEL3
PUF2
REX2
RNH70
REX3
YIL001W
TAM41
YAR1
YOR051C
TMA46
DOM34
TGS1
CUP9
DCW1
YPL183W-A
YRB30
TYW1
YKL116C
STE23
IOC2
IOC4