Supplemental Figure 1. Analysis of conserved co-expression using data from human microarray experiments and four other organisms ((A) Arabidopsis, (B) fly, (C) mouse, and (D) yeast). Each of the panels is calculated and colored as in Figure 2C and shows that tendency to physically associate is strongest when the human gene pairs and the ortholog pairs are both strongly positively correlated in their expression.

Supplemental Figure 2. Direct measurement of DNA microarray cross-hybridization for removal of cross-hybridization-induced artifacts in the analysis of conserved co-expression. A microarray-based analysis of four yeast genes (YPL274W, YLR467W, YIR039C, and YKL224) reveals the relationship between DNA sequence identity and mRNA cross-hybridization on cDNA microarrays: when the DNA sequence identity between probe genes and homologous genes on the array is 70% or lower, the cross-hybridization is negligible. Adapted from (Carlson 2002).

Supplemental Figure 3. Reproducibility of sucrose gradient separations and mass spectrometry spectral counts abundance estimates. (A) BY4741 yeast cell lysate was prepared as described in Methods and fifteen OD260 units of the supernatant loaded onto three independent continuous 7% to 47% sucrose density gradients. After centrifuging for 2.5-hr at 40,000 rpm in a Beckman SW40 rotor, each of the three sucrose gradients was fractionated and absorbance at 254 nm was measured (ISCO fractionator). To correct for trivial shifts in the gradients, the resulting polysome profiles were aligned by correlation optimized warping (COW; available as open source Matlab code (Giorgio Tomasi 2004)). Average reproducibility of the gradients is high ($<r> \sim 1$ for the three gradients). (B) Two biological replicate HeLa S3 human cell culture whole cell lysates were prepared as in Methods, then trypsin digested and analyzed by tandem mass spectrometry as described in Methods, measuring the frequency of MS/MS spectral counts for each protein identified. The spectral frequencies observed in the two replicates both correlate well (r = 0.95, 543 proteins) and agree in absolute values, falling on the plot diagonal (solid line).

Supplemental Figure 4. Ribosomal protein mass spectrometry-based elution profiles are consistent with nucleic acid concentration profiles. The top panel indicates nucleic acid concentrations measured as absorbance at 254 nm across the cytosolic HeLa sucrose gradient fractionation of Figure 4, with locations of the 40S, 60S, 80S and polysomes marked. In the middle and bottom panels, specific protein elution profiles are plotted as measured by mass spectrometry. Elution profiles of known 40S and 60S ribosomal subunits (bottom panel) match those expected from the nucleic acid distributions; similarly, known 40S biogenesis factor BMS1 elutes with 40S cytosolic ribosomes; newly implicated 60S biogenesis factor BCCIP (ortholog of yeast BCP1; Figure 8) co-elutes with cytosolic ribosomal subunits.

Supplemental Figure 5. Proteasomal protein mass spectrometry-based elution profiles are consistent with Western blotting of independently fractionated biological replicate samples. Elution profiles measured by mass spectrometry (as in Methods, Figure 4) are plotted in the top panel for 3 proteasomal protein subunits. The profiles are generally consistent with those observed from a biological replicate HeLa cell sample, prepared in similar fashion, separated independently into cytosolic and nuclear/mitochondrial preparations and fractionated on replicate sucrose gradients, with proteins detected by Western blotting.

Supplemental Figure 6. Models of error for the accuracy benchmark based on yeast network path length. Results are plotted for replicate trials evaluating the accuracy of sets composed of known proportions of true and false positive associations (true positives are protein pairs known to belong to the same complexes from (Joshi-Tope et al. 2005); false positives are random associations chosen from the same set of proteins, omitting true positive associations). The test is performed as in Methods, employing YeastNet v. 2 (Lee et al. 2007), with all linkages derived exclusively from mRNA co-expression omitted. Each curve plots the mean +/- 1 s.d. for 10 random selections of associations with the indicated percentage of true positives. These plots verify that the benchmark correctly estimates accuracy across the full range of values. The mean of the standard deviation of the random trials for all tested error rates (doubling the standard deviations

from tests of 0% and 100%, which fall at the accuracy boundaries) is 3.4 %, which we use to estimate the variation expected in measurements of other association sets using these benchmarks.

Supplemental Figure 7. The overall distribution of KOG functional annotations among the 2,348 unique proteins in the 7,000 conserved co-expression associations, measured using DAVID (Dennis et al. 2003). The number of proteins with each function is indicated around the perimeter; proteins for which only general function is known form the largest set, followed by proteins in post-translational modification, turnover, and folding.

Supplemental Figure 8. Statistical enrichment for specific protein functions and localizations among the 2,348 unique proteins in the 7,000 conserved co-expression associations, measured using DAVID (Dennis et al. 2003) and summarizing enrichment for specific Gene Ontology (A) biological processes, (B) cellular components, and (C) molecular functions. The plots indicate categories in each analysis that are enriched with $p < 10^{-4}$. Pie charts display the enriched categories ordered by the number of proteins in each of the significantly-enriched categories; histograms plot the corresponding *p*-values. In each pair of graphs (*i.e.*, in A, B or C), the same colors are used for the pie chart and histogram.

Supplemental Figure 9. Statistical enrichment for specific protein sequence families among the 2,348 unique proteins in the 7,000 conserved co-expression associations, measuring significantly enriched InterPro protein domain families using DAVID (Dennis et al. 2003). The pie chart and histogram represent the number of proteins in each domain family and the *p*-value associated with each domain family, respectively.

Supplemental Figure 10. (A) Histogram of BIOS scores for the 7,000 CCE associations. 4,354 (62%) of the 7,000 associations can be supported by at least one of the additional lines of evidence summarized in the BIOS score. (B) The LLR scores of CCE associations generally increase with increasing BIOS scores, although high-

confidence (high LLR) CCE associations are identified with no supporting BIOS evidence. The bottom, middle, and top horizontal lines of the box-and-whisker plots represent the 1st quartile, the median, and the 3rd quartile of AUCs, respectively; whiskers indicate 1.5 times the interquartile range. Symbols represent individual outliers.

References

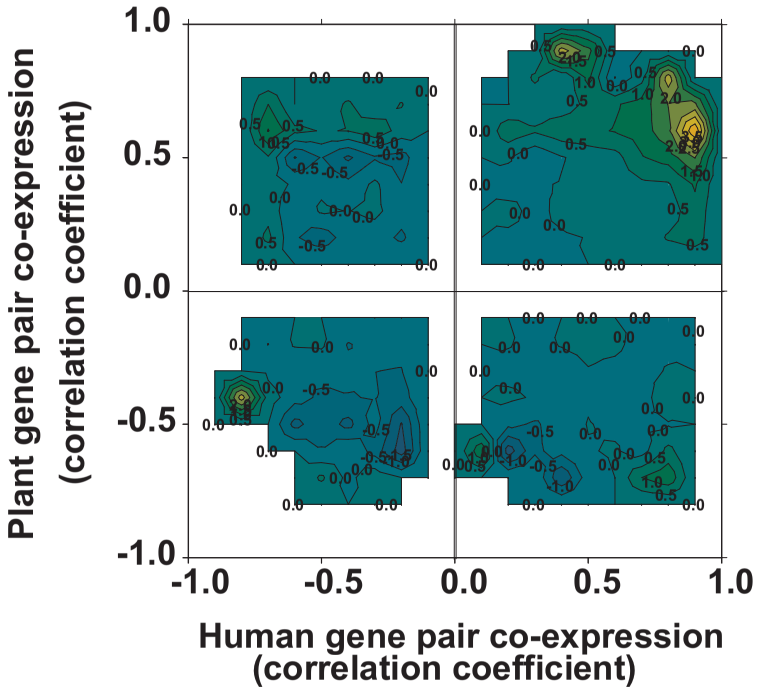
Carlson, M.W. 2002. Surveying yeast genomic diversity using cDNA microarrays. In *Masters Thesis, Biomedical Engineering*, pp. 60. University of Texas, Austin.

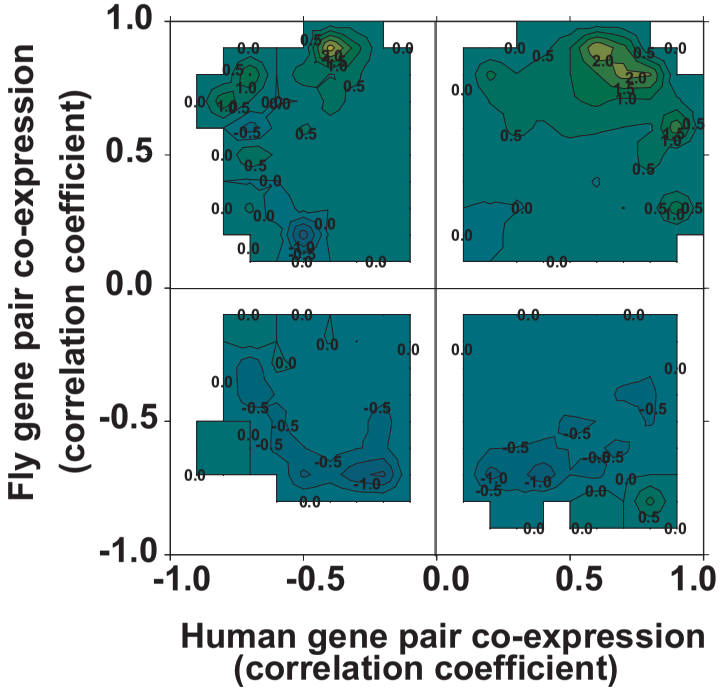
Dennis, G., Jr., B.T. Sherman, D.A. Hosack, J. Yang, W. Gao, H.C. Lane, and R.A. Lempicki. 2003. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* **4:** P3.

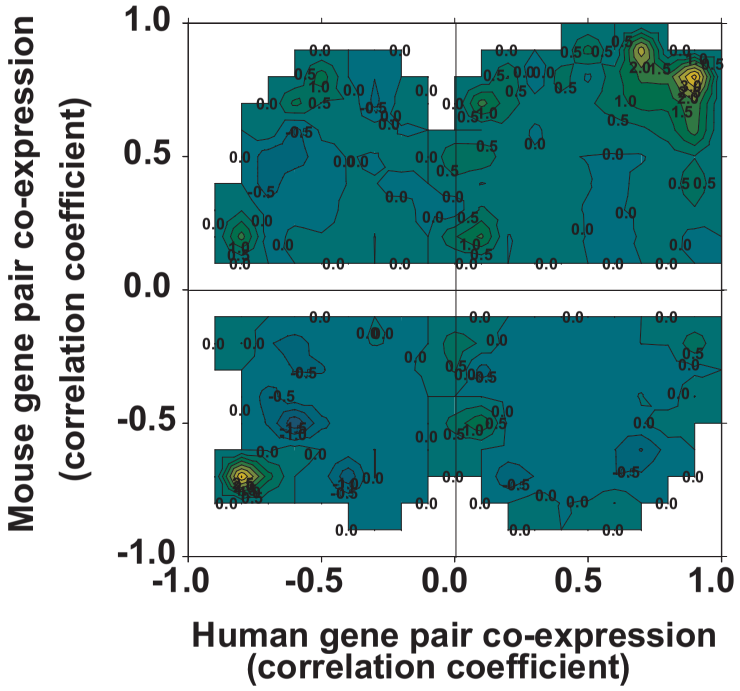
Giorgio Tomasi, F.v.d.B.C.A. 2004. Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data. *Journal of Chemometrics* **18:** 231-241.

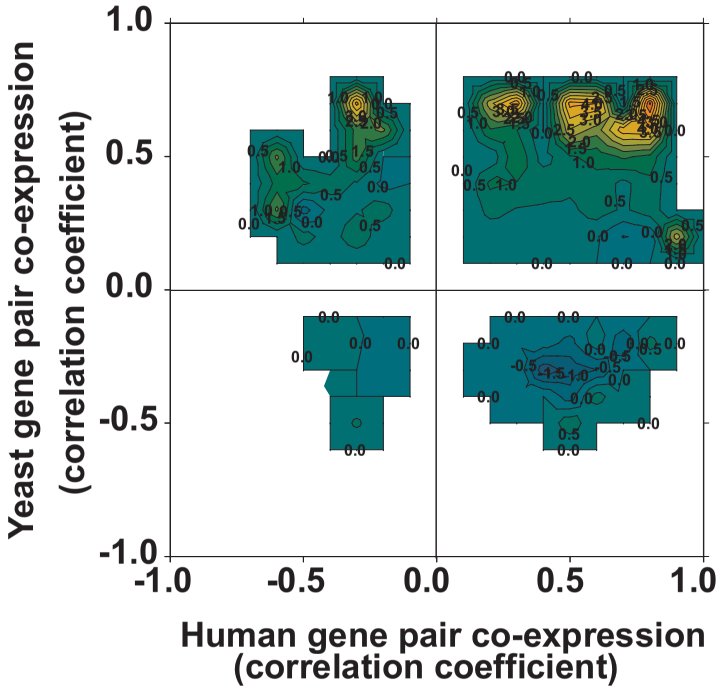
Joshi-Tope, G., M. Gillespie, I. Vastrik, P. D'Eustachio, E. Schmidt, B. de Bono, B. Jassal, G.R. Gopinath, G.R. Wu, L. Matthews, S. Lewis, E. Birney, and L. Stein. 2005. Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res* **33 Database Issue:** D428-432.

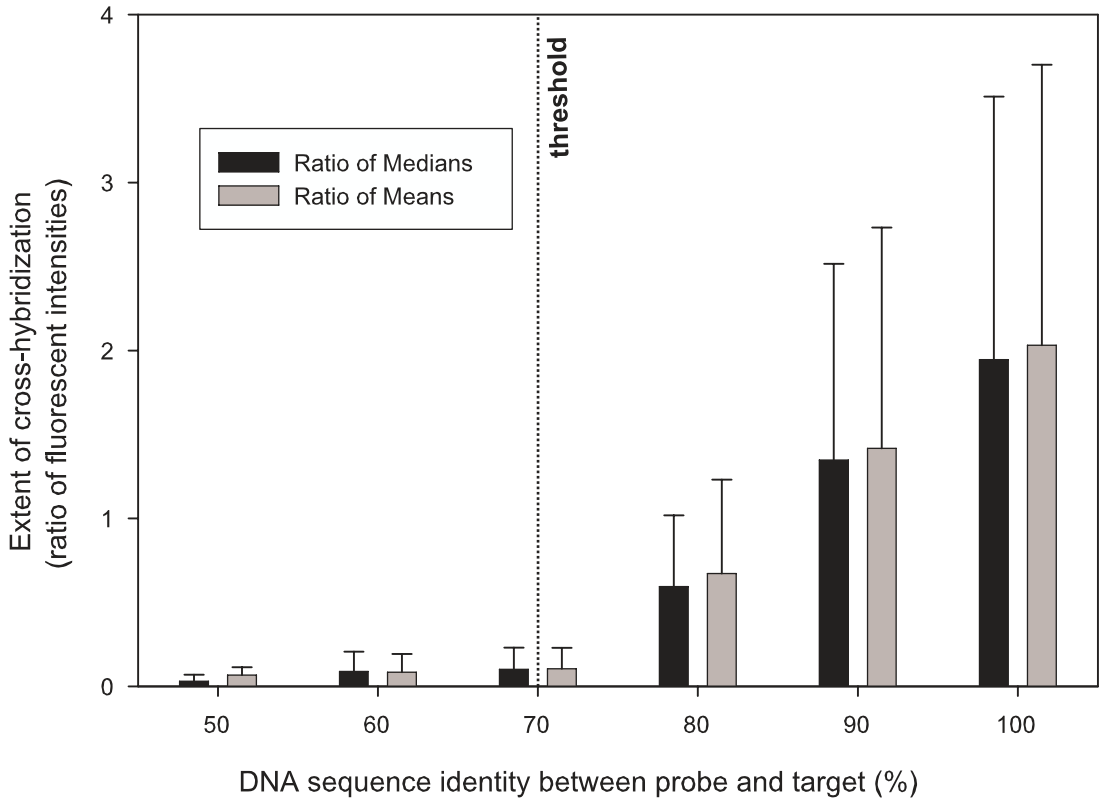
Lee, I., Z. Li, and E.M. Marcotte. 2007. An Improved, Bias-Reduced Probabilistic Functional Gene Network of Baker's Yeast, Saccharomyces cerevisiae. *PLoS ONE* **2:** e988.

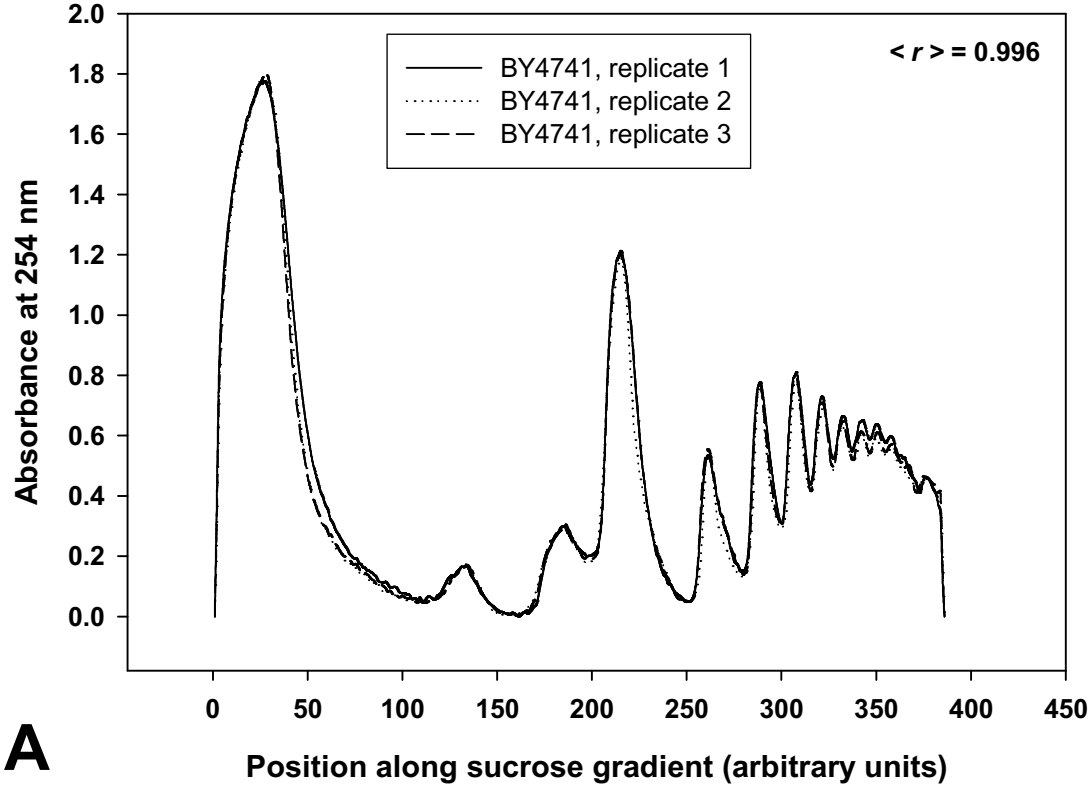

Plant gene pair co-expression
(correlation coefficient)
Human gene pair co-expression
(correlation coefficient)
1.0
0.5
0.0
-0.5
-1.0
-1.0
-0.5
0.0
0.5
1.0


Fly gene pair co-expression
(correlation coefficient)
Human gene pair co-expression
(correlation coefficient)
1.0
0.5
0.0
-0.5
-1.0
-1.0
-0.5
0.0
0.5
1.0

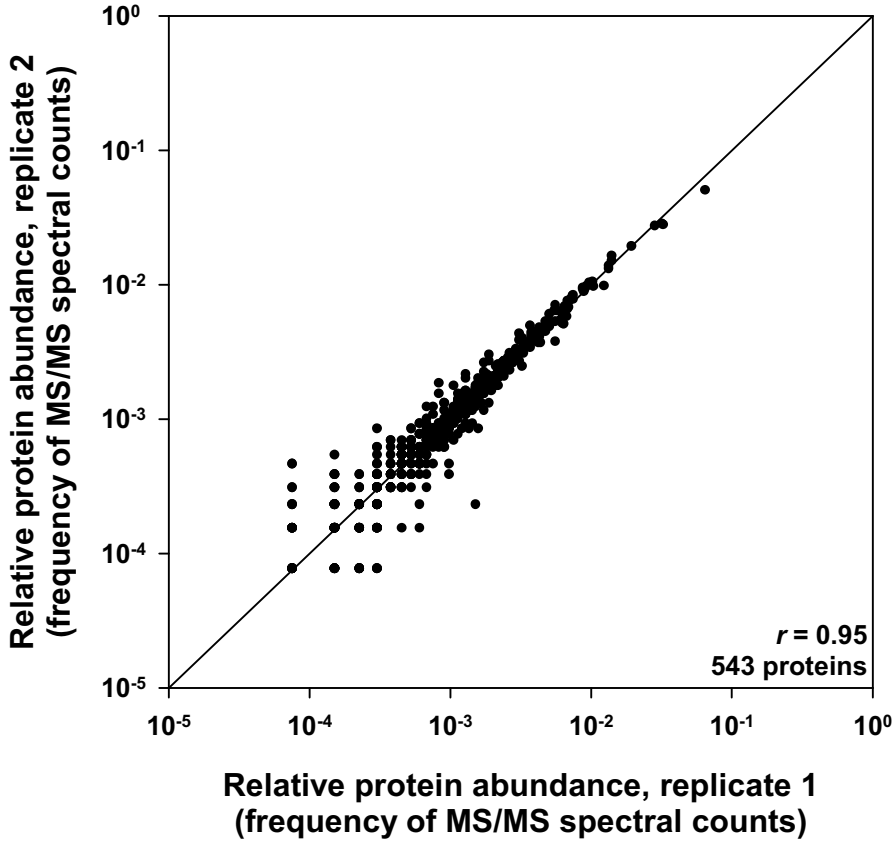

Mouse gene pair co-expression
(correlation coefficient)
Human gene pair co-expression
(correlation coefficient)
1.0
0.5
0.0
-0.5
-1.0
-1.0
-0.5
0.0
0.5
1.0


Yeast gene pair co-expression
(correlation coefficient)
Human gene pair co-expression
(correlation coefficient)
1.0
0.5
0.0
-0.5
-1.0
-1.0
-0.5
0.0
0.5
1.0

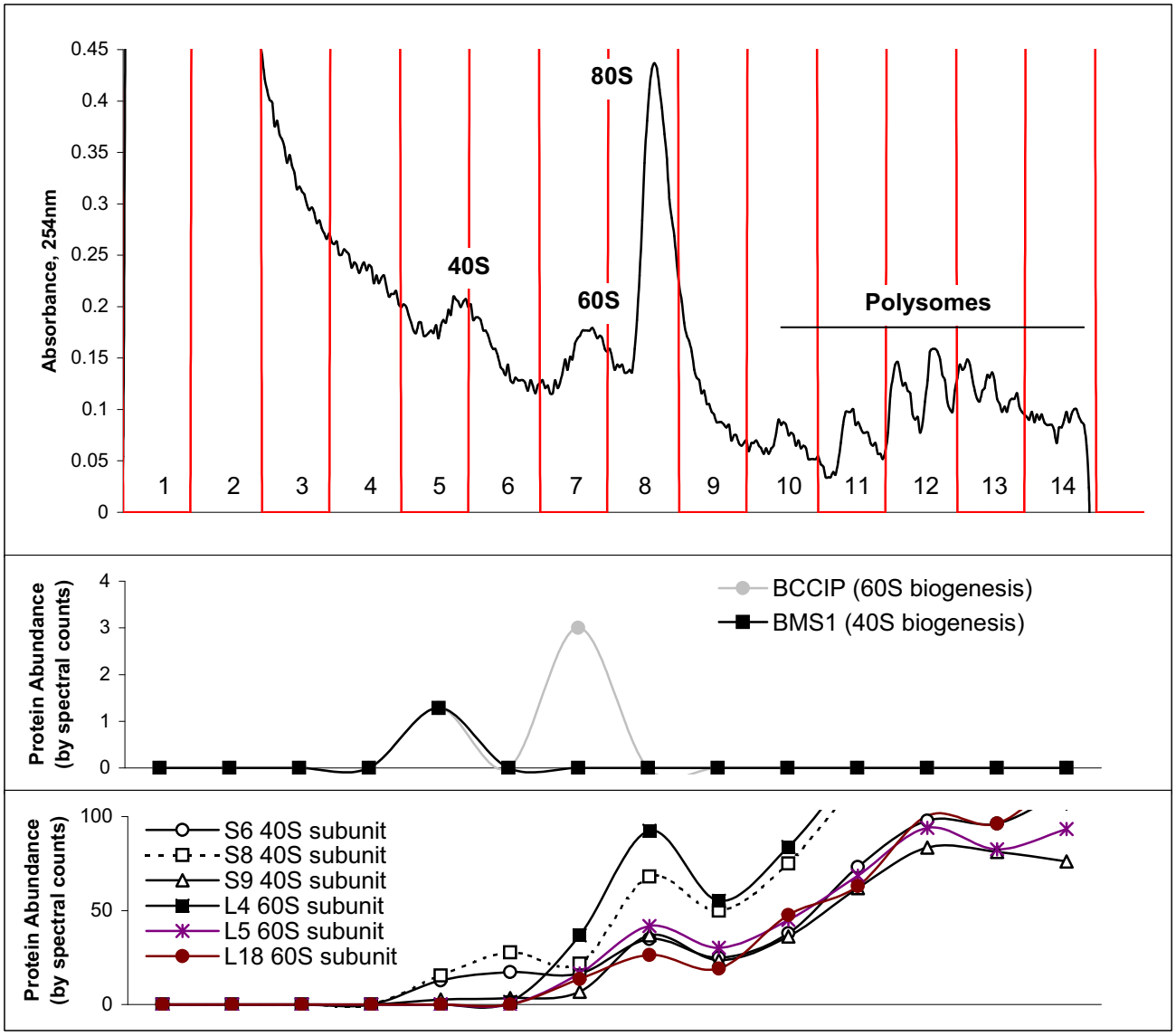

Extent of cross-hybridization
(ratio of fluorescent intensities)
4
3
2
1
0
threshold
Ratio of Medians
Ratio of Means
50
60
70
80
90
100
DNA sequence identity between probe and target (%)

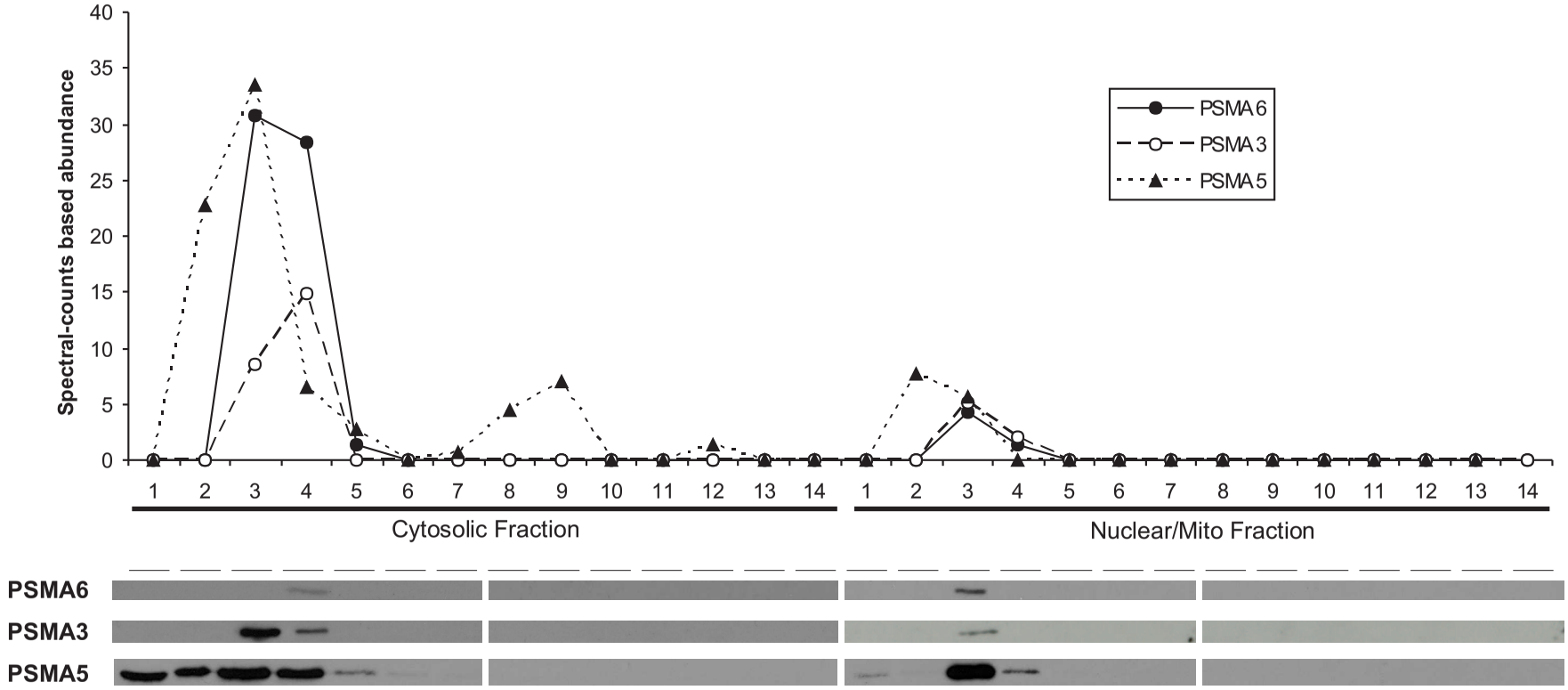

BY4741, replicate 1
BY4741, replicate 2
BY4741, replicate 3
< r > = 0.996
Absorbance at 254 nm
Position along sucrose gradient (arbitrary units)

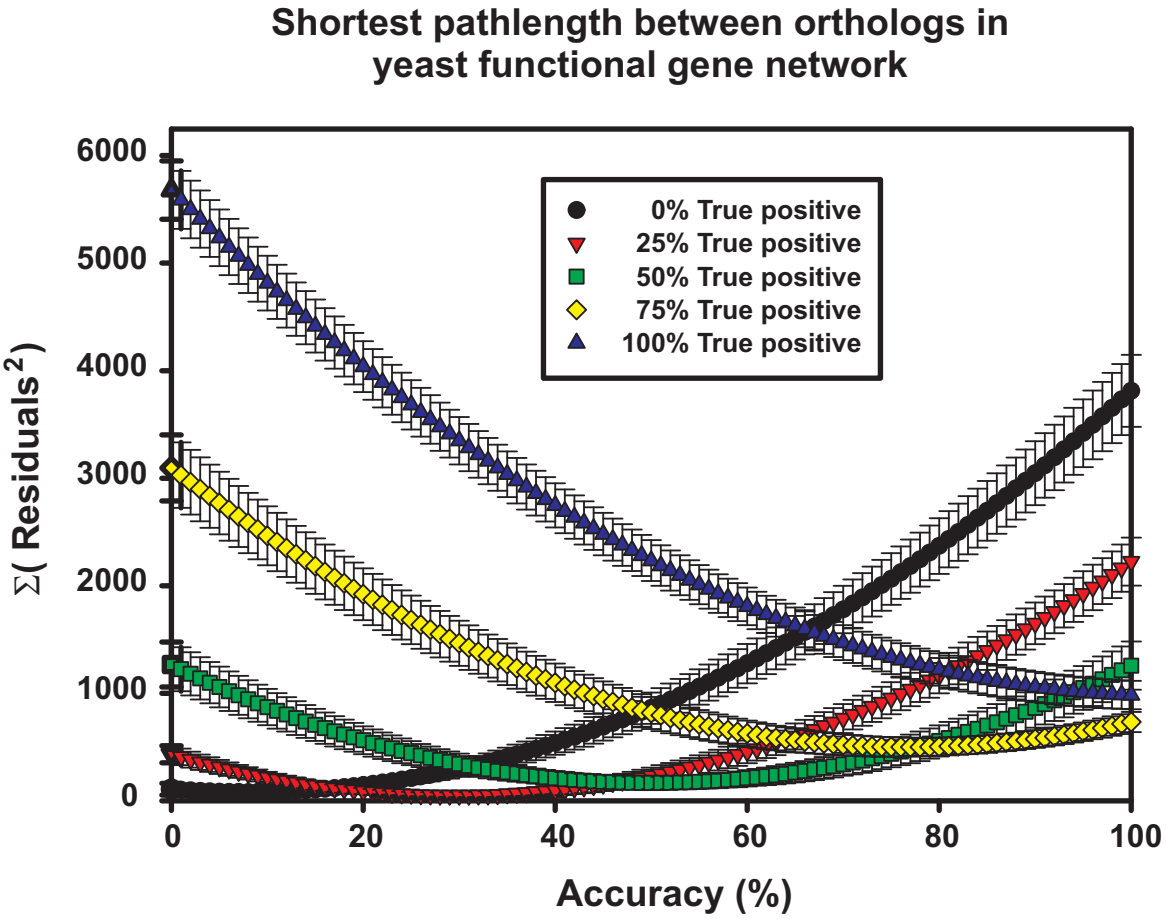
A

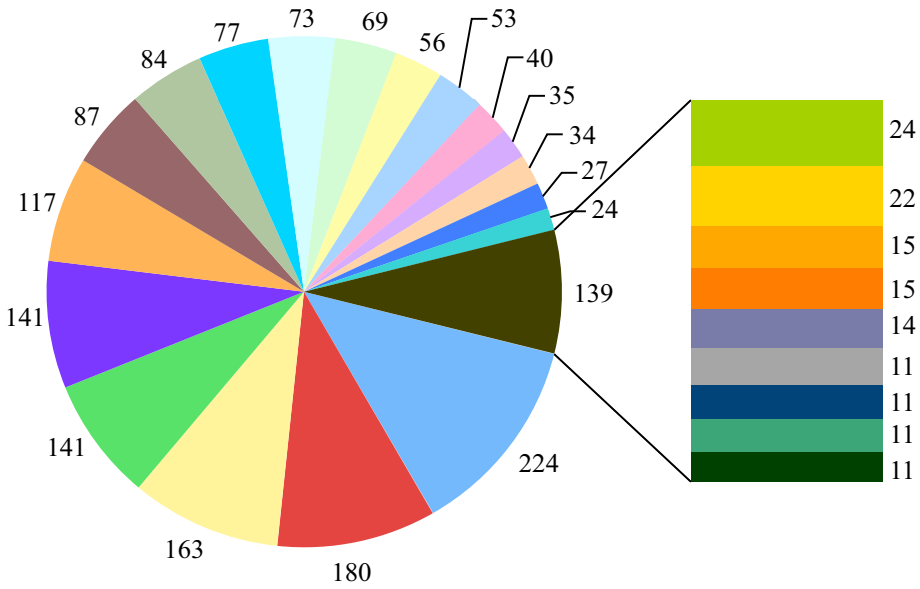

Relative protein abundance, replicate 2
(frequency of MS/MS spectral counts)
r = 0.95
543 proteins
Relative protein abundance, replicate 1
(frequency of MS/MS spectral counts)

B


Absorbance, 254nm
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
40S
60S
80S
Polysomes
1
2
3
4
5
6
7
8
9
10
11
12
13
14
Protein Abundance (by spectral counts)
BCCIP (60S biogenesis)
BMS1 (40S biogenesis)
Protein Abundance (by spectral counts)
S6 40S subunit
S8 40S subunit
S9 40S subunit
L4 60S subunit
L5 60S subunit
L18 60S subunit

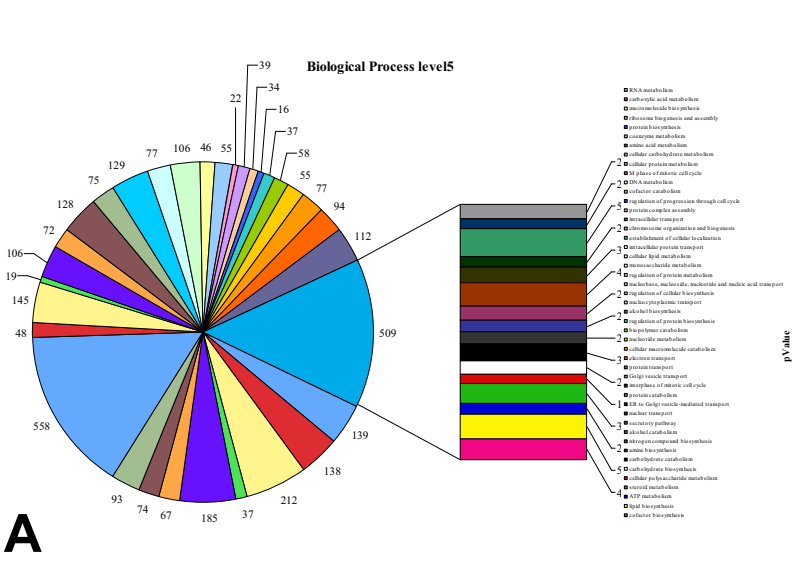

Spectral-counts based abundance
40
35
30
25
20
15
10
5
0
PSMA6
PSMA3
PSMA5
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Cytosolic Fraction
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Nuclear/Mito Fraction
PSMA6
PSMA3
PSMA5

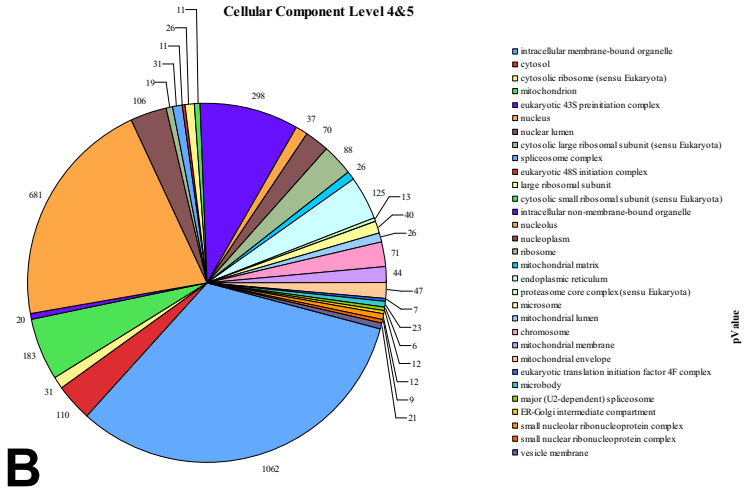

Shortest pathlength between orthologs in yeast functional gene network
0% True positive
25% True positive
50% True positive
75% True positive
100% True positive
Σ(Residuals2)
6000
5000
4000
3000
2000
1000
0
0
20
40
60
80
100
Accuracy (%)

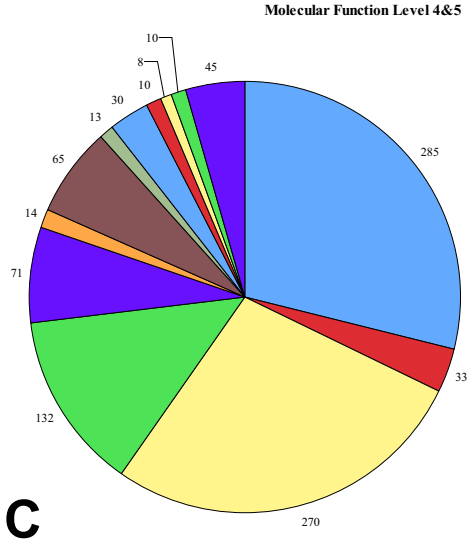


General function prediction only
Posttranlational modification, protein turnover, chaperones
Signal transduction mechanisms
Translation, Ribosome structure and biogenesis
Transcription
RNA processing and modification
Function unknown
Energy production and conversion
Carbohydrate transport and metabolism
Cytoskeleton
Intracellular trafficking, secretion and vesicular transport
Amino acid transport and metabolism
Lipid transport and metabolism
Inorganic ion transport and metabolism

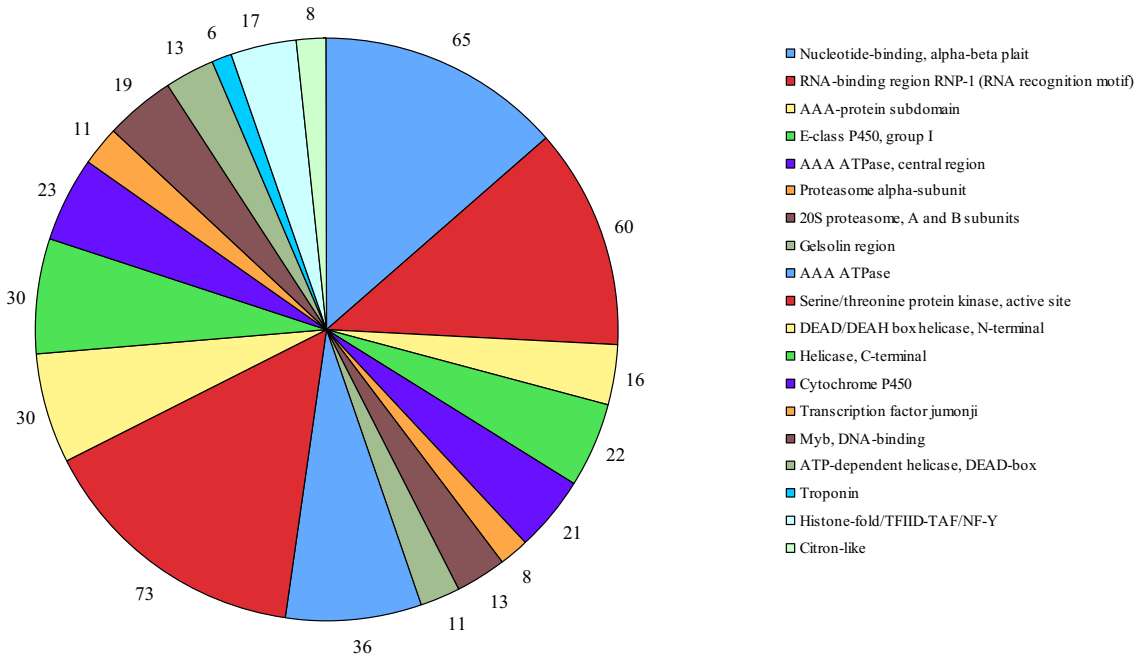
Replication, recombination and repair
Cell cycle control, cell division, chromosome partitioning
Chromatin structure and dynamics
Nucleotide transport and metabolism
Extracellular structures
Secondary metabolites biosynthesis, transport and catabolism
Lipid metabolism
Cell division and chromosome partitioning
Signal transduction mechanisms, general function prediction only
Defense mechanisms
Coenzyme transport and metabolism
Signal transduction mechanisms, intracellular trafficking, secretion and vesicular transport
Chromatin structure and dynamics, transport


Biological Process level5
39
34
22
16
37
58
55
77
94
112
509
139
138
212
37
185
67
74
93
558
48
145
19
106
72
128
75
129
77
106
46
55
RNA metabolism
carboxylic acid metabolism
macromolecule biosynthesis
ribosome biogenesis and assembly
protein biosynthesis
coenzyme metabolism
amino acid metabolism
cellular carbohydrate metabolism
cellular protein metabolism
M phase of mitotic cell cycle
DNA metabolism
cofactor catabolism
regulation of progression through cell cycle
protein complex assembly
intracellular transport
chromosome organization and biogenesis
establishment of cellular localization
intracellular protein transport
cellular lipid metabolism
monosaccharide metabolism
regulation of protein metabolism
nucleobase, nucleoside, nucleotide and nucleic acid transport
regulation of cellular biosynthesis
nucleocytoplasmic transport
alcohol biosynthesis
regulation of protein biosynthesis
biopolymer catabolism
nucleotide metabolism
cellular macromolecule catabolism
electron transport
protein transport
Golgi vesicle transport
interphase of mitotic cell cycle
protein catabolism
ER to Golgi vesicle-mediated transport
nuclear transport
secretory pathway
alcohol catabolism
nitrogen compound biosynthesis
amine biosynthesis
carbohydrate catabolism
carbohydrate biosynthesis
cellular polysaccharide metabolism
steroid metabolism
ATP metabolism
lipid biosynthesis
cofactor biosynthesis
A
Biological Process Level 5
pV'alue
1.00E-01
1.00E-04
1.00E-07
1.00E-10
1.00E-13
1.00E-16
1.00E-19
RNA metabolism
carboxylic acid metabolism
macromolecule biosynthesis
ribosome biogenesis and assembly
protein biosynthesis
coenzyme metabolism
amino acid metabolism
cellular carbohydrate metabolism
cellular protein metabolism
M phase of mitotic cell cycle
DNA metabolism
cofactor catabolism
regulation of progression through cell cycle
protein complex assembly
intracellular transport
chromosome organization and biogenesis
establishment of cellular localization
intracellular protein transport
cellular lipid metabolism
monosaccharide metabolism
regulation of protein metabolism
nucleobase, nucleoside, nucleotide and nucleic acid transport
regulation of cellular biosynthesis
nucleocytoplasmic transport
alcohol biosynthesis
regulation of protein biosynthesis
biopolymer catabolism
nucleotide metabolism
cellular macromolecule catabolism
electron transport
protein transport
Golgi vesicle transport
interphase of mitotic cell cycle
protein catabolism
ER to Golgi vesicle-mediated transport
nuclear transport
secretory pathway
alcohol catabolism
nitrogen compound biosynthesis
amine biosynthesis
carbohydrate catabolism
carbohydrate biosynthesis
cellular polysaccharide metabolism
steroid metabolism
ATP metabolism
lipid biosynthesis
cofactor biosynthesis

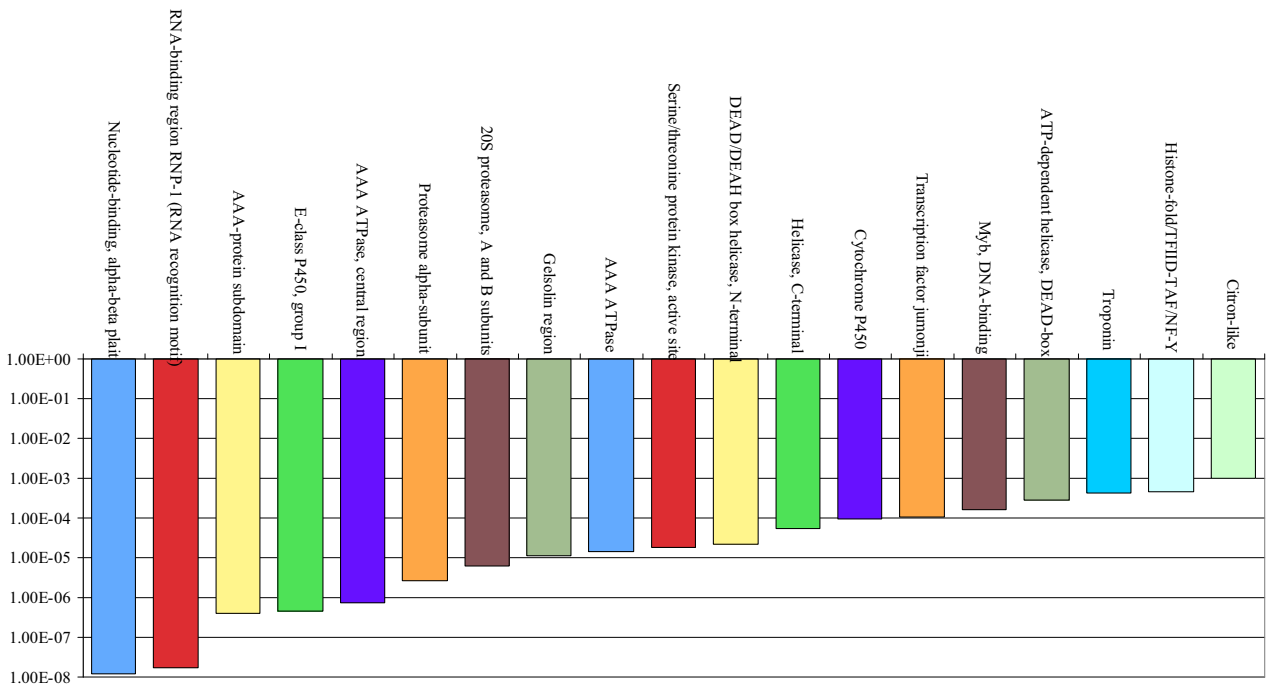

Cellular Component Level 4&5
11
26
11
31
19
106
298
37
70
88
26
125
13
40
26
71
44
47
7
23
6
12
12
9
21
1062
110
31
183
20
681
intracellular membrane-bound organelle
cytosol
cytosolic ribosome (sensu Eukaryota)
mitochondrion
eukaryotic 43S preinitiation complex
nucleus
nuclear lumen
cytosolic large ribosomal subunit (sensu Eukaryota)
spliceosome complex
eukaryotic 48S initiation complex
large ribosomal subunit (sensu Eukaryota)
cytosolic small ribosomal subunit (sensu Eukaryota)
intracellular non-membrane-bound organelle
nucleolus
nucleoplasm
ribosome
mitochondrial matrix
endoplasmic reticulum
proteasome core complex (sensu Eukaryota)
microsome
mitochondrial lumen
chromosome
mitochondrial membrane
mitochondrial envelope
eukaryotic translation initiation factor 4F complex
microbody
major (U2-dependent) spliceosome
ER-Golgi intermediate compartment
small nucleolar ribonucleoprotein complex
small nuclear ribonucleoprotein complex
vesicle membrane
B
Cellular Component Level 4&5
pV'alue
1.00E+02
1.00E-02
1.00E-06
1.00E-10
1.00E-14
1.00E-18
1.00E-22
1.00E-26
1.00E-30
1.00E-34
1.00E-38
1.00E-42
intracellular membrane-bound organelle
cytosol
cytosolic ribosome (sensu Eukaryota)
mitochondrion
eukaryotic 43S preinitiation complex
nucleus
nuclear lumen
cytosolic large ribosomal subunit (sensu Eukaryota)
spliceosome complex
eukaryotic 48S initiation complex
large ribosomal subunit (sensu Eukaryota)
cytosolic small ribosomal subunit (sensu Eukaryota)
intracellular non-membrane-bound organelle
nucleolus
nucleoplasm
ribosome
mitochondrial matrix
endoplasmic reticulum
proteasome core complex (sensu Eukaryota)
microsome
mitochondrial lumen
chromosome
mitochondrial membrane
mitochondrial envelope
eukaryotic translation initiation factor 4F complex
microbody
major (U2-dependent) spliceosome
ER-Golgi intermediate compartment
small nucleolar ribonucleoprotein complex
small nuclear ribonucleoprotein complex
vesicle membrane

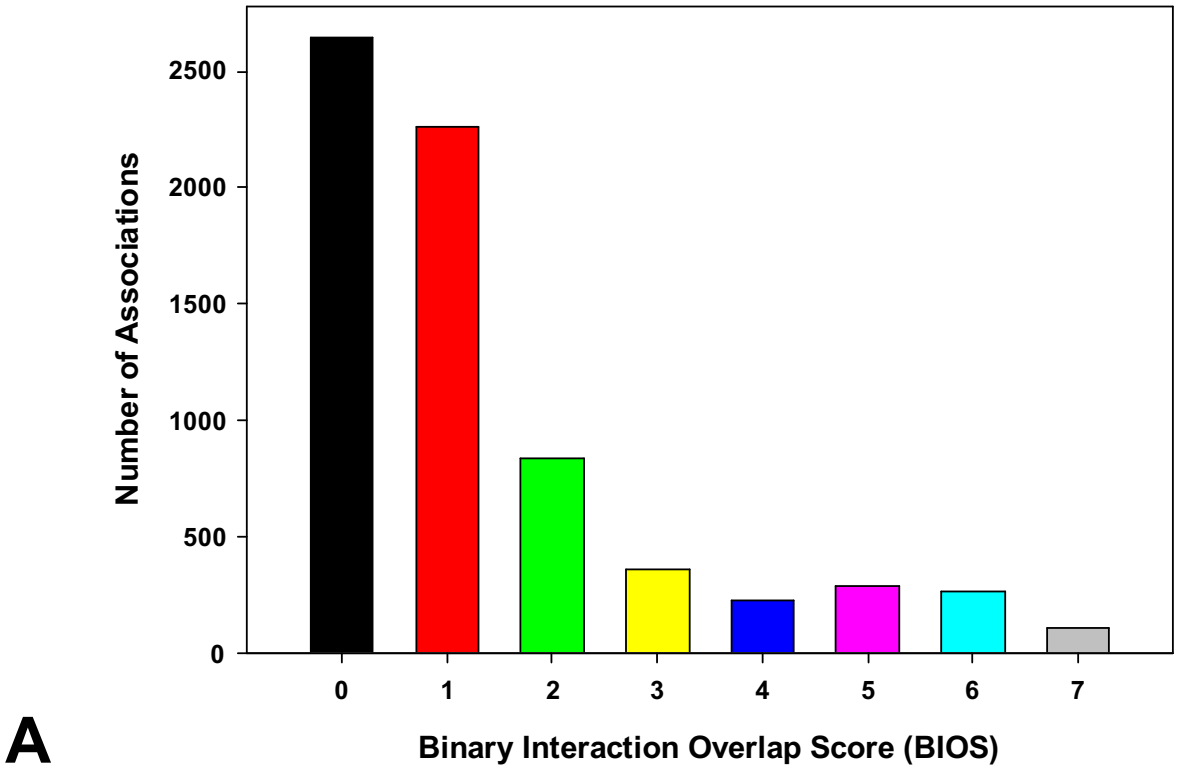

Molecular Function Level 4&5
10
8
10
30
13
65
14
71
45
285
33
270
132
adenyl nucleotide binding
translation initiation factor activity
ATP binding
pyrophosphatase activity
iron ion binding
unspecific monooxygenase activity
magnesium ion binding
threonine endopeptidase activity
oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor
RNA splicing factor activity, transesterification mechanism
tRNA binding
protein phosphatase type 2A regulator activity
monovalent inorganic cation transporter activity
C
Molecular Function Level 4&5
1.00E+00
1.00E-03
1.00E-06
1.00E-09
1.00E-12
1.00E-15
adenyl nucleotide binding
translation initiation factor activity
ATP binding
pyrophosphatase activity
iron ion binding
unspecific monooxygenase activity
magnesium ion binding
threonine endopeptidase activity
oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor
RNA splicing factor activity, transesterification mechanism
tRNA binding
protein phosphatase type 2A regulator activity
monovalent inorganic cation transporter activity

Interpro Domains

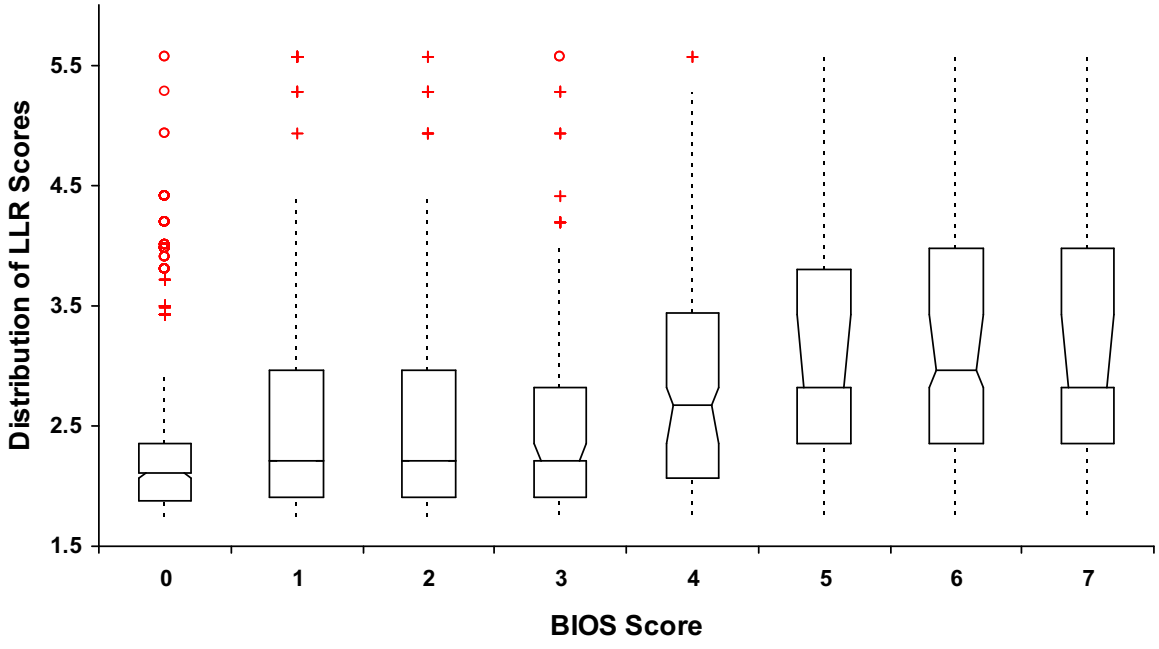


Interpro Domains




Number of Associations
2500
2000
1500
1000
500
0
0
1
2
3
4
5
6
7
Binary Interaction Overlap Score (BIOS)

A


Distribution of LLR Scores
5.5
4.5
3.5
2.5
1.5
0
1
2
3
4
5
6
7
BIOS Score

B