

Research

Open Access

The MoVIN server for the analysis of protein interaction networks

Paolo Marcatili¹, Giovanni Bussotti¹ and Anna Tramontano*^{1,2}

Address: ¹Department of Biochemical Sciences, "Sapienza" University, Rome, Italy and ²Istituto Pasteur – Fondazione "Cenci Bolognetti", Rome, Italy

Email: Paolo Marcatili - paolo.marcatili@uniroma1.it; Giovanni Bussotti - giovannibussotti@gmail.com; Anna Tramontano* - anna.tramontano@uniroma1.it

* Corresponding author

from Italian Society of Bioinformatics (BITS): Annual Meeting 2007
Naples, Italy. 26-28 April 2007

Published: 26 March 2008

BMC Bioinformatics 2008, 9(Suppl 2):S11 doi:10.1186/1471-2105-9-S2-S11

This article is available from: <http://www.biomedcentral.com/1471-2105/9/S2/S11>

© 2008 Marcatili et al.; licensee BioMed Central Ltd.

This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Protein-protein interactions are at the basis of most cellular processes and crucial for many bio-technological applications. During the last few years the development of high-throughput technologies has produced several large-scale protein-protein interaction data sets for various organisms. It is important to develop tools for dissecting their content and analyse the information they embed by data-integration and computational methods.

Results: Interactions can be mediated by the presence of specific features, such as motifs, surface patches and domains. The co-occurrence of these features on proteins interacting with the same protein can indicate mutually exclusive interactions and, therefore, can be used for inferring the involvement of the proteins in common biological processes.

We present here a publicly available server that allows the user to investigate protein interaction data in light of other biological information, such as their sequences, presence of specific domains, process and component ontologies. The server can be effectively used to construct a high-confidence set of mutually exclusive interactions by identifying similar features in groups of proteins sharing a common interaction partner. As an example, we describe here the identification of common motifs, function, cellular localization and domains in different datasets of yeast interactions.

Conclusions: The server can be used to analyse user-supplied datasets, it contains pre-processed data for four yeast Protein Protein interaction datasets and the results of their statistical analysis. These show that the presence of common motifs in proteins interacting with the same partner is a valuable source of information, it can be used to investigate the properties of the interacting proteins and provides information that can be effectively integrated with other sources. As more experimental interaction data become available, this tool will become more and more useful to gain a more detailed picture of the interactome.

Background

Protein functions are mediated and regulated through a complex network of interactions [1]. In many cases proteins physically bind to each other to absolve their role, and the interaction is often mediated by the physical binding of some of their subunits, such as domains, surface patches or small regions composed of a few residues called motifs [2-4]. Although the latter is rather frequent, there have been few attempts to systematically explore the information that they provide at the genomic level. Motif recognition has proven to be very useful in many biological contexts, but is not an easy task [3,5,6]. Motifs are often short in length (three to twelve residues), they are often located in disordered regions of the proteins and their conservation is limited to closely related species. Nevertheless the identification of shared motifs has proven to be very useful to characterize protein interactions (e.g. the binding of the SH3 domain to the PxxP local sequence), function (DNA binding), localization (nuclear localization signal) and domain fingerprints (PROSITE [4]).

The recent development of high-throughput technologies for detecting protein-protein interactions (PPIs) has produced many publicly available databases [1,7-10]. Although the accuracy of the data is not always optimal [11,12], the information they provide is of primary importance for formulating biologically relevant hypotheses and it is therefore essential to develop tools for analysing and dissecting them. There are methods that make use of different biological data to assess the reliability of interactions: gene expression [13], homology [14], Gene Ontology (GO) annotations [15], phylogenetic features [16], synthetic lethality, domain interaction [17], and a combination of these [18]. PPI maps have also been mined to infer functional similarity, domain interactions and protein motifs [5,19,20].

In this work we describe a server for simplifying the analysis of the features shared by proteins interacting with the same partner. We show here its power by investigating the presence of sequence motifs in yeast PPI maps and their correlation with the presence of similar Gene Ontology annotations (process and component) [15] and Pfam domains [21]. The result of our analysis is that the information that can be gained by motif detection is relevant and coherent with functional, localization and domain data but it is not redundant with respect to these other sources of information. It is indeed possible to exploit the presence of common motifs to identify mutually exclusive interactions and to estimate the reliability of a PPI map.

Results and discussion

The MoVIN server

The MoVIN server can load experimental PPI datasets and perform an analysis of sets of interactions sharing a common partner (Figure 1). It contains pre-computed data for four dataset for *S. cerevisiae* of different size, level of curation and estimated false positive rate, which overlap only partially (see Methods).

Given a dataset, the tool automatically extracts the groups of proteins sharing a common interaction partner (see Methods). These are the sets of all and only the proteins that bind to a common protein partner. This central protein (hub) is used to identify the cluster. The user can select the minimum size of the cluster. For each cluster, MoVIN collates the corresponding set of protein sequences and searches them for the presence of common motifs using MEME [22] and MAST [23].

MEME (vers. 3.5.3) is a tool for discovering motifs in a group of related DNA or protein sequences. MEME takes as input a group of sequences (in this case the sequences of the proteins in a cluster) and outputs as many motifs as possible according to the constraints given by the user. MEME uses statistical modelling techniques to automatically choose the best length and description for each motif. In MoVIN we use as background distribution the average composition of the complete set of proteins in the map and focus on motifs with length between three and twelve residues and with an E-value (as reported by the output of the program) lower than a user-defined threshold (default is $10e-7$). To avoid that motifs with a strong signal (e.g. fingerprints of a domain shared by some proteins of the group) mask "weaker" motifs, the server repeats this step on each cluster many times, recursively eliminating the proteins for which a common motif has already been identified.

In order to estimate the specificity of the motifs found by MEME, they are mapped on all the proteins in the dataset using MAST (vers. 3.5.3). For each motif, MAST returns a list of proteins containing it, the position of the match in the sequence and the corresponding E-value. Only matches with an E-value smaller than a user-defined threshold (default is $10e-5$, i.e. 100 times larger than the E-value threshold chosen for MEME, in order to prevent overfitting of the motifs on the initial set of clustered proteins) are retained.

Next, the server assigns to each cluster C_i and each motif M_j a *motif P value* $S_{i,j}$ related to the over-representation of the motif in that cluster. If the motif has been found on x_{ij} protein in the cluster C_i of n_i proteins and on X_j proteins in the complete dataset of size N we have $S_{i,j} = \text{hgcdf}(x_{ij}, n_i, X_j, N)$ where hgcdf is the hypergeometric cumulative



Home

- Home
- Paper
- Additional Files

Server

- Submit your Data
- Pre-computed Datasets

Authors

BioComputing Group

> Submit your data

Enter your e-mail address:

Upload or paste here your PPI list (see [help](#) to see the accepted format):

Select the dataset you want to merge with your PPIs:

Min cluster size ([help](#)): Max cluster size ([help](#)):

MEME treshold ([help](#)): MAST treshold ([help](#)):

Output type([help](#)):

Figure 1
Snapshot of the input page of the MoVIN server. The user can upload a Protein-protein interaction map in any of the accepted formats (tab or comma separated) with or without merging it with existing datasets. The minimum and maximum cluster size as well as the threshold E-values for MEME and MAST can be selected as well.

distribution function, which measures the probability of finding at least as many occurrences of the motif in a cluster of similar size randomly extracted from the whole set of proteins.

Finally, the previously calculated scores are assigned to the interaction between the central protein of the cluster and each protein in the cluster containing the motif M_i . Different motifs can be present on a protein and one pro-

tein of a cluster can be, at the same time, the hub protein of another cluster, therefore different scores can be associated to the same interaction. In such cases, we assign the minimum score to the interaction.

The *process P value* S^p , the *component P value* S^c and the *domain P value* S^d are computed in a similar fashion. The GO process score and the GO component score are calculated by mapping the corresponding GO ontology terms on each protein with the program BatchGoViewer, which returns, given a list of proteins, the annotations (at any level) with the lowest P values. The Pfam domain score is calculated by analysing the presence of Pfam A Domains on each protein. The protein-domain relationship is taken from the Ensembl website [24].

The results of all the analyses are displayed with a user-friendly interface, including textual search and graphics tools. The cluster and their features are graphically visualised using GraphViz. For each displayed item there are links to several different publicly available databases.

Additionally, the user can look for the presence of known mutations and for their position in the highlighted proteins (as reported in the Protein Mutant Database) [25]. It is possible to download and visualize experimental structures or three-dimensional models for a large fraction of the yeast genome. Known structures are downloaded from the Protein Data Bank [26], models are downloaded from the ModBase database [27]. Each structure or model can be visualised in the web browser via Jmol [28] and all the motifs found by the MoVIN web server can be highlighted.

Application of the MoVIN server to the *S.cerevisiae* interactome

We used MoVIN to analyse the BioGRID, BIND, Gavin06 and Krogan datasets (Table 1). We only considered clusters containing more than four proteins and used the default values for all the parameters. A summary of the final datasets is reported in Table 1.

Table 1: Dataset Summary. The number of interactions in each dataset is between 942 (Uetz) and 51,086 (BioGRID), the number of clusters containing more than 4 proteins is between 100 (Uetz) and 3,963. The average number of proteins in each cluster ranges from 6.15 (Uetz) to 24.96 (BioGRID).

Dataset	# of interactions	# of clusters	average cluster size
BIND	8847	1271	9.60
BioGRID	51086	3963	24.96
Gavin02	3500	522	9.78
Gavin06	19973	1738	20.94
Krogan	6699	1042	10.44
Uetz	942	100	6.15

First of all, we analysed the motif P value distribution generated by the server using as input the original datasets and then compared them with the corresponding background distribution. Such background distribution is obtained -for each PPI map- by randomizing the set of interactions of the original map (written as ordered pairs) with no duplicate (i.e. if the pair (a_i, b_i) is present we can not have the pair (b_i, a_i)). We randomly shuffle all the second terms of each pair and remove duplicate interactions. By doing so, we preserve the connectivity degree of each protein and hub proteins remain such in the randomized map. The total number of interactions of each randomized map differs by less than 1% from that of the original map.

The difference between the original dataset P value distributions and their randomized version is computed by a permutation test on the means. We generate a global distribution by merging the original and the randomized dataset distribution. Next, we randomly generate 100.000 distributions of the same size as the original dataset and compute their means. The distribution of the means for the analysed datasets is normal (D'Agostino-Pearson test, significance=0.05), therefore we can compute the standard Z-score of our original dataset:

$$z = \frac{(M_d - M)}{\sigma}$$

where M_d is the mean of the distribution on the original dataset, M is the mean and σ is the standard deviation of the mean distribution. The randomization step was repeated 5 times for each dataset, yielding each time similar results (data not shown). The results (Figure 2 and Table 2) show that the motif distributions in PPI maps are far from random.

It is interesting to note that several parameters, such as the percentage of interactions that can be explained by a motif on one of the binding partners or the difference between the occurrence of the motifs in the real map and in the randomized one are consistent with the expected fraction of spurious interaction present in the datasets. As shown in Table 2, MoVIN finds more motifs in the databases BioGRID (that is the largest available repository of PPIs) and BIND (that only contains well annotated PPIs and is manually-curated), while it finds far less common motifs in the Uetz dataset (which is estimated to contain a relatively larger fraction of false positives [11]). Moreover, the effectiveness of the method increases with the dimension of the map. The more the map covers the complete interactome, the larger the number of motifs identified.

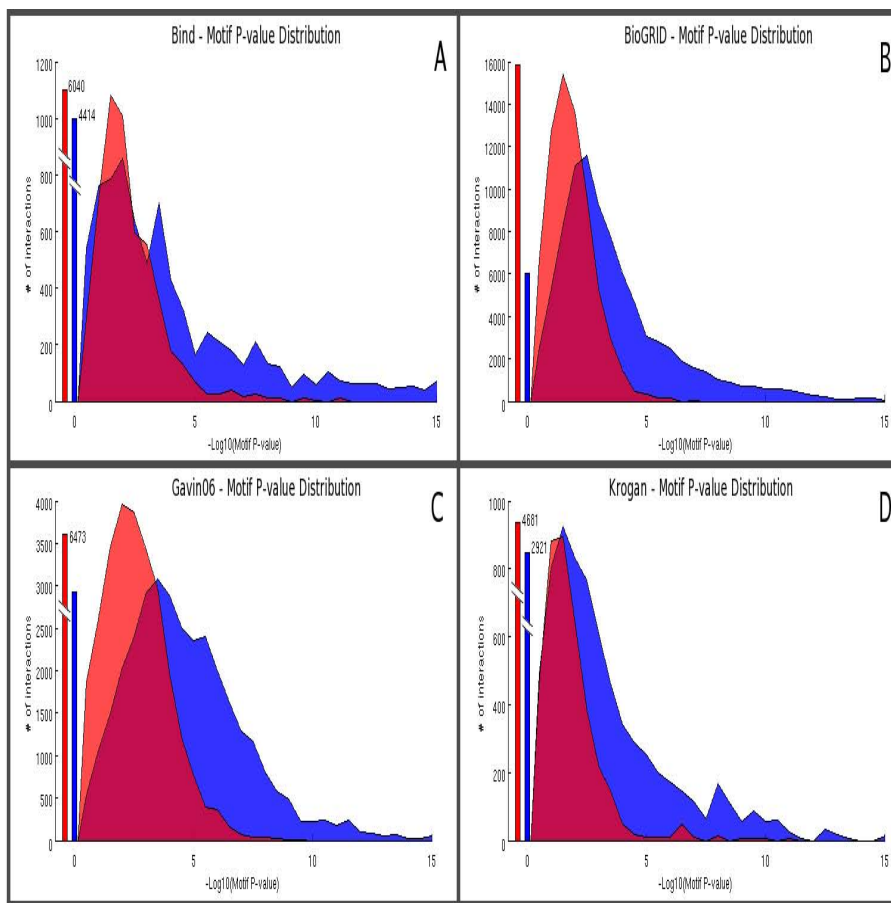


Figure 2
Motif P Value Distributions for yeast datasets. We report the number of interactions with Motif P value lower than the threshold (reported on the x axis in logarithmic scale) for the experimental datasets (blue) and for its randomized version (red). The interactions for which no motif was found are reported as bars in the origin. The Motif P value distribution for the experimental datasets contains a larger fraction of interactions with respect to the random datasets and is shifted towards lower P values. (A) BIND, (B) BioGRID, (C) Gavin06 and (D) Krogan datasets.

On the basis of this analysis, we generated a high-confidence subset of the yeast PPI datasets (see additional files 1 and 2) that contains 17.733 interactions (see Methods).

Such interactions, as previously stated, are likely to be mutually exclusive.

Table 2: Statistical Significance of the Motif P value Distributions. The table reports the Z-score of the mean of the experimental distribution with respect to the random distribution of the means. The latter was obtained by computing the means of 100,000 distributions of the same size of the experimental one obtained by randomly extracting interactions from the original and the randomized distributions.

Dataset	Motif	Process	Component	Domain
BIND	33,17	26,46	82,85	32,31

Although the motifs that we identify are not necessarily related to physical binding of the proteins (they could be functional motifs or localization signals) and the motif-selected dataset is biased towards mutually exclusive interactions, it is likely that our selected subset is enriched in true positive interactions, and can be a good candidate for applications that need a benchmark interaction dataset.

Comparison with GO annotations and Pfam domains

The MoVIN server can also use other sources of information commonly used to assess the reliability of interaction data, i.e. GO annotations (process and component) and Pfam domains. An important question is to which extent the motif information overlaps with that given by the co-occurrence of GO terms and the presence of similar Pfam domain in proteins interacting with a common partner.

Table 3: Correlation coefficient between Motif, Process, Component and Domain P values

BIND	Motif	Process	Component
Motif	1.00	0.30	0.26
Process	0.30	1.00	0.80
Component	0.26	0.80	1.00
Domain	0.57	0.40	0.37

To address the issue, we applied the method to the same yeast datasets described above. Here we only report the results for the BIND dataset, the results for the other datasets are very similar (data not shown).

The correlations between \log_{10} (Motif P value) and the P values for GO process, GO component and Pfam (Table 3) are positive in all cases and range from 0.25 (motif vs component P value) to 0.56 (motif vs domain P value). This indicates that the information obtained from the motif analysis is consistent, but not completely redundant, with the other sources of information here considered.

An example: function prediction using motif analysis

The method that we described for the analysis of PPI maps can almost naturally be extended to become a tool for protein functional assignment, on the basis of the hypothesis that two proteins interacting with the same partner and sharing a common motif are likely to have some functional similarity as well. Here we will just describe one instructive example of the potential of the approach, both in assigning functions and in detecting erroneous or outdated annotations.

We found a motif shared by 5 among 7 proteins in the YKL074C cluster (BIND dataset). An alignment of the sequences is reported in Figure 3. The motif displays several conserved residues and does not match any pattern found in the PROSITE database [4].



Figure 3 Alignment of the motif present in 5 out of the 7 proteins binding to YKL074C. The first protein (YBR172C) is annotated as “Cytoskeleton organization and biogenesis”, the second (YPL105C) does not have any GO annotation. The last three (YML046W, YLR117C and YLR357W) are annotated with terms related to the spliceosome activity.

The hub protein YKL074C and 4 proteins in the group (YDL043C, YML046W and YLR117C, where the latter two contain the motif) have a GO process annotation that is related to the spliceosome. Another protein (YLR357W) has an annotation (double-strand break repair via nonhomologous end joining) which is consistent with the splicing pathway. The two proteins YBR172C and YPL105C are homologous and both contain the motif. The second protein is not annotated, while the first has a GO process annotation “Cytoskeleton organization and biogenesis”. Interestingly, more recent experimental evidence supports the hypothesis of its involvement with RNA splicing [29]. It is reasonable to suggest that YPL105C should be tentatively annotated as involved in RNA splicing and that the GO annotation for YBR172C should be verified.

Conclusions

We have described here a tool for the analysis of protein interaction maps, able to correlate the co-occurrence of sequence motifs, common GO annotations and similar Pfam domains with interactions sharing a common partner. Furthermore, we have investigated the relationship between the presence of common motifs and the presence of shared functional, component and domain data. The information given by the presence of common motifs is coherent and complementary to that present in other data sources. All these sources could be integrated to generate a large high-confidence yeast PPI dataset.

As further developments of the server we are extending the approach to the search for discontinuous motifs brought together by the three-dimensional structures of the interacting proteins.

Methods

Data sources

We analysed yeast (*Saccharomyces Cerevisiae*) PPI maps obtained from different datasets: the databases BIND [7] and BioGRID [9], the experimental datasets Gavin02 [30], Gavin06 [31], Krogan [32] and Uetz [33]. We filtered out the data for which it was not possible to precisely identify both interacting partner (e.g. multi-protein complexes).

The BIND dataset contains 8.847 yeast interactions, manually curated and annotated. BioGRID contains 51.086 interactions extracted from the literature and obtained with different experimental methods. The datasets from Gavin and Krogan are from three different tandem affinity purification (TAP) experiments and contain respectively 3.500 (Gavin02), 19.973 (Gavin06) and 6.699 (Krogan) interactions. The Uetz dataset contains 942 interactions derived from a Yeast 2 Hybrid experiment.

Protein sequences were downloaded from the Ensembl Genome Browser website on July 7th 2006).

To assign the GO terms, we used the annotation files downloaded from the GO website and only used IDA (Inferred from Direct Assay), IGI (Inferred from Genetic Interaction), IMP (Inferred from Mutant Phenotype) and TAS (Traceable Author Statement) annotations as to avoid low-confidence or indirect annotations.

The high-confidence PPI dataset (MOVIN_1) was generated by merging all the interactions that, at least in one dataset, were reported to have a motif P value smaller than $10e-4$. Such dataset contains 17.733 interactions and can be downloaded from <http://arianna.bio.uniroma1.it/MOVIN>

A second high-confidence PPI dataset (MOVIN_2) was generated by merging all the interactions that, at least in one dataset, were reported to have a significant sequence motif and at least a GO function or a GO component P value smaller than $10e-4$. This dataset contains 14.103 interactions and can be downloaded from <http://arianna.bio.uniroma1.it/MOVIN>

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

AT conceived the study, participated in its design and coordination and helped to draft the manuscript. GB helped in data analysis. PM developed the method, implemented it and performed the statistical analysis. All authors read and approved the final manuscript

Additional material

Additional file 1 – MOVIN1.list

List of all interactions between proteins interacting with the same partner and containing a motif with a P value $<10e-4$. (CSV file).

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2105-9-S2-S11-S1.csv>]

Additional file 2 – MOVIN2.list

List of all the interactions between proteins interacting with the same partner, containing a motif with a P value $<10e-4$ and sharing a process or component GO term with a P value $<10e-4$. (CSV file).

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2105-9-S2-S11-S2.csv>]

Acknowledgements

This work was partially supported by the BioSapiens project, funded by the European Commission within its FP6 Programme, under the thematic area "Life sciences, genomics and biotechnology for health", contract number LSHG-CT-2003-503265, by the Istituto Pasteur – Fondazione Cenci Bolo-

gnetti and by FIRB projects LIBI and ITALBIONET. PM gratefully acknowledges the support of the Biocomputing Unit of the University of Bologna.

This article has been published as part of *BMC Bioinformatics* Volume 9 Supplement 2, 2008: Italian Society of Bioinformatics (BITS): Annual Meeting 2007. The full contents of the supplement are available online at <http://www.biomedcentral.com/1471-2105/9?issue=S2>

References

1. Bork P, Jensen LJ, von Mering C, Ramani AK, Lee I, Marcotte EM: **Protein interaction networks from yeast to human.** *Curr Opin Struct Biol* 2004, **14(3)**:292-299.
2. Nooren IM, Thornton JM: **Diversity of protein-protein interactions.** *Embo J* 2003, **22(14)**:3486-3492.
3. Neduva V, Russell RB: **DILIMOT: discovery of linear motifs in proteins.** *Nucleic Acids Res* 2006, **34(Web Server issue)**:W350-355.
4. Hulo N, Bairoch A, Bulliard V, Cerutti L, De Castro E, Langendijk-Genevaux PS, Pagni M, Sigrist CJ: **The PROSITE database.** *Nucleic Acids Res* 2006, **34(Database issue)**:D227-230.
5. Neduva V, Linding R, Su-Angrand I, Stark A, de Masi F, Gibson TJ, Lewis J, Serrano L, Russell RB: **Systematic discovery of new recognition peptides mediating protein interaction networks.** *PLoS Biol* 2005, **3(12)**:e405.
6. Neduva V, Russell RB: **Peptides mediating interaction networks: new leads at last.** *Curr Opin Biotechnol* 2006.
7. Alfarano C, Andrade CE, Anthony K, Bahroos N, Bajec M, Bantoft K, Betel D, Bobechko B, Bouillier K, Burgess E, et al.: **The Biomolecular Interaction Network Database and related tools 2005 update.** *Nucleic Acids Res* 2005, **33(Database issue)**:D418-424.
8. Mewes HW, Frishman D, Mayer KF, Munsterkötter M, Noubibou O, Pagel P, Rattei T, Oesterheld M, Ruepp A, Stumpflen V: **MIPS: analysis and annotation of proteins from whole genomes in 2005.** *Nucleic Acids Res* 2006, **34(Database issue)**:D169-172.
9. Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M: **BioGRID: a general repository for interaction datasets.** *Nucleic Acids Res* 2006, **34(Database issue)**:D535-539.
10. Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, Eisenberg D: **The Database of Interacting Proteins: 2004 update.** *Nucleic Acids Res* 2004, **32(Database issue)**:D449-451.
11. Mrowka R, Patzak A, Herzel H: **Is there a bias in proteome research?** *Genome Res* 2001, **11(12)**:1971-1973.
12. von Mering C, Krause R, Snel B, Cornell M, Oliver SG, Fields S, Bork P: **Comparative assessment of large-scale data sets of protein-protein interactions.** *Nature* 2002, **417(6887)**:399-403.
13. Grigoriev A: **A relationship between gene expression and protein interactions on the proteome scale: analysis of the bacteriophage T7 and the yeast *Saccharomyces cerevisiae*.** *Nucleic Acids Res* 2001, **29(17)**:3513-3519.
14. Matthews LR, Vaglio P, Reboul J, Ge H, Davis BP, Garrels J, Vincent S, Vidal M: **Identification of potential interaction networks using sequence-based searches for conserved protein-protein interactions or "interologs".** *Genome Res* 2001, **11(12)**:2120-2126.
15. **The Gene Ontology (GO) project in 2006.** *Nucleic Acids Res* 2006, **34(Database issue)**:D322-326.
16. Enright AJ, Ouzounis CA: **Functional associations of proteins in entire genomes by means of exhaustive detection of gene fusions.** *Genome Biol* 2001, **2(9)**: RESEARCH0034
17. Sprinzak E, Margalit H: **Correlated sequence-signatures as markers of protein-protein interaction.** *J Mol Biol* 2001, **311(4)**:681-692.
18. Patil A, Nakamura H: **Filtering high-throughput protein-protein interaction data using a combination of genomic features.** *BMC Bioinformatics* 2005, **6**:100.
19. Deng M, Zhang K, Mehta S, Chen T, Sun F: **Prediction of protein function using protein-protein interaction data.** *Proc IEEE Comput Soc Bioinform Conf* 2002, **1**:197-206.
20. Riley R, Lee C, Sabatti C, Eisenberg D: **Inferring protein domain interactions from databases of interacting proteins.** *Genome Biol* 2005, **6(10)**: R89
21. Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R, et al.: **Pfam:**

- clans, web tools and services. *Nucleic Acids Res* 2006, **34(Database issue):D247-251**.
22. Bailey TL, Williams N, Misleh C, Li WW: **MEME: discovering and analyzing DNA and protein sequence motifs.** *Nucleic Acids Res* 2006, **34(Web Server issue):W369-373**.
 23. Bailey TL, Gribskov M: **Methods and statistics for combining motif match scores.** *J Comput Biol* 1998, **5(2):211-221**.
 24. Hubbard T, Andrews D, Caccamo M, Cameron G, Chen Y, Clamp M, Clarke L, Coates G, Cox T, Cunningham F, et al.: **Ensembl 2005.** *Nucleic Acids Res* 2005, **33(Database issue):D447-453**.
 25. Kawabata T, Ota M, Nishikawa K: **The Protein Mutant Database.** *Nucleic Acids Res* 1999, **27(1):355-357**.
 26. Berman HM, Bhat TN, Bourne PE, Feng Z, Gilliland G, Weissig H, Westbrook J: **The Protein Data Bank and the challenge of structural genomics.** *Nat Struct Biol* 2000, **7(Suppl):957-959**.
 27. Pieper U, Eswar N, Davis FP, Braberg H, Madhusudhan MS, Rossi A, Marti-Renom M, Karchin R, Webb BM, Eramian D, et al.: **MODBASE: a database of annotated comparative protein structure models and associated resources.** *Nucleic Acids Res* 2006, **34(Database issue):D291-295**.
 28. **Jmol: an open-source Java viewer for chemical structures in 3D** [<http://www.jmol.org>]
 29. Andoh T, Azad AK, Shigematsu A, Ohshima Y, Tani T: **The fission yeast *ptr1+* gene involved in nuclear mRNA export encodes a putative ubiquitin ligase.** *Biochem Biophys Res Commun* 2004, **317(4):1138-1143**.
 30. Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon AM, Cruciat CM, et al.: **Functional organization of the yeast proteome by systematic analysis of protein complexes.** *Nature* 2002, **415(6868):141-147**.
 31. Gavin AC, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, Rau C, Jensen LJ, Bastuck S, Dumpelfeld B, et al.: **Proteome survey reveals modularity of the yeast cell machinery.** *Nature* 2006, **440(7084):631-636**.
 32. Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, Li J, Pu S, Datta N, Tikuisis AP, et al.: **Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*.** *Nature* 2006, **440(7084):637-643**.
 33. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P, et al.: **A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*.** *Nature* 2000, **403(6770):623-627**.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

