Comparison of Human Protein-Protein Interaction Maps

Matthias E. Futschik[1], Gautam Chaurasia[1,2], Erich Wanker[2] and Hanspeter Herzel[1]

[1]Institute for Theoretical Biology, Charité, Humboldt-Universität and [2]Max-Delbrück-Centrum,
Invalidenstrasse 43
10115 Berlin, Germany
m.futschik@biologie.hu-berlin.de

Abstract: Large-scale mappings of protein-protein interactions have started to give us new views of the complex molecular mechanisms inside a cell. After initial projects to systematically map protein interactions in model organisms such as yeast, worm and fly, researchers have begun to focus on the mapping of the human interactome. To tackle this enormous challenge, different approaches have been proposed and pursued. While several large-scale human protein interaction maps have recently been published, their quality remains to be critically assessed. We present here a first comparative analysis of eight currently available large-scale maps with a total of over 10000 unique proteins and 57000 interactions included. They are based either on literature search, orthology or by yeast-two-hybrid assays. Comparison reveals only a small, but statistically significant overlap. More importantly, our analysis gives clear indications that all interaction maps suffer under selection and detection biases. These results have to be taken into account for future assembly of the human interactome.

1 Introduction

Interactions between proteins underlie the vast majority of cellular processes. They are essential for a wide range of tasks and form a network of astonishing complexity. Until recently, our knowledge of this complex network was rather limited. The emergence of large scale protein-protein interaction maps has given us new possibilities to systematically survey and study the underlying biological system. The first attempts to collect protein-protein interactions on large scale were initiated for model organisms such as *S. cerevisiae*, *D. melanogaster* and *C. elegans* [Gavin *et al.* '02,Giot *et al.* '03,Ito *et al.* '01,Li *et al.* '04,Uetz *et al.* '00]. Evidently, the generated interaction maps offered a rich resource for systematic studies.

After these initial efforts, the focus has moved towards deciphering the human interactome. Recently, the first large-scale human protein interaction network has been constructed following alternative mapping strategies. Most currently available human interaction maps can be divided into three classes: i) maps obtained from literature search [Bader *et al.* '01,Peri *et al.* '03,Ramani *et al.* '05], ii) maps derived from interactions between orthologous proteins in other organisms [Brown and Jurisica '05,

Lehner and Fraser '04, Persico *et al.* '05] and iii) maps based from large scans using yeast-two-hybrid (Y2H) assays [Rual *et al.* '05, Stelzl *et al.* '05]. All of these different mapping strategies have their obvious advantages as well as disadvantages. For example, Y2H-based mapping approaches offer rapid screens between thousands of proteins, but might produce a high false positive rate. The extent, however, how much the resulting interaction maps are influenced by the choice of mapping strategy, is less clear. Thus, it is important to critically assess the quality and reliability of produced maps.

For yeast interaction maps, several of such critical comparisons have been performed [Bader and Hogue '02, von Mering *et al.* '02]. They revealed a surprising divergence between different interaction maps. They also indicated that functional coherency of maps is severely influenced by the choice of mapping scheme. Such comparison is still lacking for human protein interaction maps despite their expected importance for biomedical research [Goehler *et al.* '04]. Therefore, we compared several currently available large-scale interactions maps regarding their concurrence and divergence. We assess especially potential selection and detection biases as they might interfere with future applications of these maps.

2 Materials and Methods

2.1 Assembly of Protein-Protein Interaction Maps

To evaluate the different mapping approaches listed above, we selected eight publicly available large-scale interaction maps: three literature-based, three orthology-based and two Y2H-based maps. We restricted further our analysis to binary interactions in order to compare Y2H-based maps directly with the remaining interaction maps.

Two literature-based interaction maps were derived from the Human Protein Reference Database (HPRD) and Biomolecular Interaction Network Database (BIND) [Bader *et al.* '01, Peri *et al.* '03]. These manually curated databases are mainly based on literature reviews performed by human experts. At the time of analysis, interactions included in these databases were predominantly from small scale experiments. As third literature-based interaction map, we used the set of interactions found by Ramani and co-workers using a text-mining approach [Ramani *et al.* '05]. As HPRD and BIND, it is based on literature, but computationally generated. In our study, we will refer to it as the COCIT map.

The first orthology-based protein interaction map was proposed by Lehner and Fraser [Lehner *et al.* '04]. Interactions included were predicted based on interactions observed between orthologous proteins in yeast, worm and fly. We used only interactions that were assigned to core map by Lehner and Fraser, as these were identified with high confidence. Besides this map (here referred to as the ORTHO map), we included two alterative orthology-based large-scale maps from in the Online Predicted Human Interaction Database (OPHID) and HOMOMINT database [Brown *et al.* '05, Persico *et al.* '05]. Both mappings were derived following the approach by Lehner and Fraser with some deviations. We extracted from the two databases only the interactions that were based on orthology assignment to ensure conformity of the resulting maps.

The Y2H-based interaction maps included in our comparison were generated in the recent large-scale scans by Stelzl *et al*. and Rual *et al*. [Rual *et al.* '05, Stelzl *et al.* '05] We will refer to these maps as MDC-Y2H and CCSB-H1 in our study. Although both scans are based on Y2H-assay, it should be noted that considerable differences exist in regard to experimental procedures.

To enable comparison, all proteins were mapped to their corresponding EntrezGene ID. For efficient computational analysis, we converted all interaction maps into graphs using the Bioconductor *graph* package [Balasubramanian *et al.* '04, Carey *et al.* '05, Gentleman *et al.* '04].

2.2 Overlap of Interaction Maps

Protein interaction maps are formed by both their proteins and interactions included. Thus, any comparison of maps should assess the concurrence of proteins as well as of interactions in different maps.

Comparison of the proteins in different maps is based on following procedure: Given the sets of proteins (P_A, P_B) in map A and B, their intersection is $P_{AB} = P_A \cap P_B$. To facilitate assessment, the intersection was normalized in regard to the total number of proteins in A or B ($P^A_{AB} = P_{AB} / P_A$; $P^B_{AB} = P_{AB} / P_B$). Thus, the normalized intersection is simply the percentage of proteins that can be found in the other map. In our study, we will refer to the average of P^A_{AB} and P^B_{AB} as the *(relative) protein overlap* between A and B.

For the comparison of interactions, we could proceed similarly by counting common interactions in two maps. However, it is important to note that network structure is not only determined by existing interactions, but also by missing interactions. As we want to assess the concurrence of maps for both observed as well as missing interactions, we used a log-likelihood ratio (*LLR*) score [Lee *et al.* '04]. The *LLR* provides a similarity measure for two sets of interactions (I_1,I_2). It is defined as

$$LLR(I_1, I_2) = \ln\left(\frac{P(I_1 \mid I_2)}{P(I_1 \mid \sim I_2)}\right)$$

where $P(I_1|I_2)$ is the probability of observing an interaction in I_1 conditioned on observing the same interaction in I_2. Respectively, $P(I_1|\sim I_2)$ is the probability of observing an interaction in I_1 conditioned on not observing the same interaction in I_2. For highly similar interaction networks, *LLR* produces large scores. For absence of similarity, the *LLR* score is zero. The latter is the case if random interactions networks are compared.

Additionally to the *LLR* score, we used two permutation tests to stringently assess the statistical significance of observed concurrence of interactions [Balasubramanian *et al.* '04]. For both tests, a large set of random networks are generated based on the original networks, either by re-labelling of nodes (*node label permutation*) or by randomly permuting the edges (*edge permutation*). In contrast to node label permutation, the implemented scheme for edge permutation does not conserve the degree distribution i.e. the number of interactions of proteins. Subsequently, the number of common interactions between the original networks is compared to the corresponding number for randomized networks. The probability of observing at least the same number of common interactions for random networks determines the significance. Although their permutation schemes are different, the two tests usually produce similar results.

2.3 Gene Ontology Analyses

Protein interaction maps can be compromised by several types of biases. For example, selection bias arises if certain protein categories are over- or underrepresented in a chosen map. To assess stringently the significance of such potential biases, we utilized Fisher's exact test. It is based on the hypergeometric distribution and delivers the probability P to observe k or more proteins of chosen category in case of random drawings:

$$P = 1 - \sum_{i=0}^{k-1} \frac{\binom{M}{i}\binom{N-M}{l-i}}{\binom{N}{l}}$$

where M is the total number of proteins attributed to the category, N is the total number of proteins annotated and l is the number of proteins in the corresponding map. The significance of under-represented GO categories in maps can be calculated accordingly. Since we tested simultaneously for multiple GO categories, the p-values were converted to false discovery rates applying the Benjamini-Hochberg procedure [Benjamini and Hochberg '95]. As reference, the set of all proteins tested for interactions could be used. However, such sets are explicitly known only for Y2H-based maps comprising the proteins in a matrix screen. For literature- and orthology-based maps, these sets are not available. Hence, we used the set of all genes annotated in GO as reference to facilitate direct comparison.

We also assessed whether interactions between protein classes were overrepresented. We determined the number of interactions k_{mn} between proteins of GO category m and proteins of GO category n. Log_2-odds were calculated to assess deviation of the observed number of interactions k_{mn} with the number k^0_{mn} of interactions expected for randomized networks:

$$LOD_{mn} = \log_2 \frac{k_{mn}}{k^0_{mn}}$$

MAP	REFERENCE	*P*	*I*	D_{AV}	METHOD
MDC-Y2H	Stelzl *et al.* 2005 *Cell*	1703	3186	1.9	Y2H-ASSAY
CCSB-H1	Rual *et al.* 2005 *Nature*	1549	2754	1.8	Y2H-ASSAY
HPRD	Peri *et al.* 2003 *Genome Res*	5908	15658	2.7	LITERATURE
BIND	Bader *et al.* 2001 NAR	2677	4233	1.7	LITERATURE
COCIT	Ramani *et al.* 2004 *Genome Biology*	3737	6580	1.8	LITERATURE
OPHID	Brown and Jurisica 2005 *Bioinformatics*	2284	8962	3.9	ORTHOLOGY
ORTHO	Lehner and Fraser 2003 *Genome Biology*	3503	9641	2.8	ORTHOLOGY
HOMOMINT	Persico *et al.* 2005 *BMC Bioinformatics*	2556	5582	2.3	ORTHOLOGY

Table 1: List of human protein-protein interactions maps compared in this study. The number of proteins *P* and interactions *I* result after mappings of proteins to their corresponding Entrez ID. D_{av} denotes the average number of interactions per protein.

The randomized networks had the same number of proteins and interactions as the corresponding maps. The proteins' connectivity (number of interactions per protein) was also conserved.

Alternatively, we can evaluate the tendency that proteins of similar function interact. Although difficult to define rigorously, similarity of function may be approximated by following procedure [Jansen *et al.* '03]: After mapping proteins to their GO terms (categories), their functional similarity is determined by the positions of corresponding GO terms within the GO graph. Similar GO terms are expected to be located in proximity to each other. Measuring the shared paths to the GO terms (from the root term), we would expect that similar GO terms have larger shared paths than unrelated GO terms. Thus, if proteins of similar function tend to interact in a network, the average shared paths lengths will be larger than random networks. To test the significance, we compared therefore the distribution of shared path lengths to those measured for randomized networks. Note that we counted the largest shared path length in case of multiple GO assignments for proteins.

3. Results

In total, we were able to map 57095 interactions between 10769 proteins uniquely identified by the corresponding Entrez IDs (table 1). The size of the interactions maps varied between 2754 (CCSB-H1) and 15658 (HPRD) interactions. Proteins had an average number of 1.8 to 3.8 interactions. Considering previous estimates of an average of 3-10 interactions per proteins, the result indicates that interactions maps are currently still highly unsaturated [Bork *et al.* '04].

3.1 Common Proteins and Interactions

We examined first how many proteins and interactions were common to the different

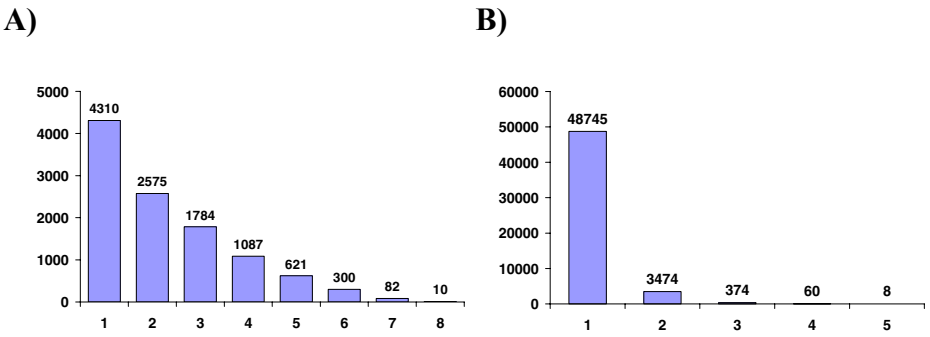

Figure 1: Number of proteins (A) and interactions (B) common to multiple maps. The x-axis shows the number of maps in which proteins or interactions are included.

maps in our comparison (figure 1). We found that a large part (60%) of all proteins can be found in at least two maps. The number of proteins included in all eight maps, however, is diminishingly small: Only 10 proteins (i.e. 0.001% of all proteins) fulfill this criterion. Even more striking were the small numbers of common interactions. The vast majority of interactions (85%) are cataloged in only a single map. No interaction can be found in six or more maps; and just 8 interactions are common to five maps.

3.2 Protein Overlap

To investigate whether some maps tend to share more proteins than others, we calculated the relative protein overlap for each pair of maps. We detected considerable variation of protein overlap ranging from 16% to 58%. Comparison of overlaps gave us first indications that maps could be ordered into distinct groups. To examine this possibility, a clustering approach was applied. First, we converted protein overlaps O_{ij} (between maps i and j) into distances Δ_{ij} defined as $\Delta_{ij} = 1\text{-}O_{ij}$. Thus, maps having large protein overlap are assigned a small distance between each other. After conversion, the interaction maps were hierarchically clustered. The resulting cluster structure showed a surprisingly clear pattern: All maps are grouped in accordance to the mapping approach used for their generation. We obtained two clusters that either included only literature-based or orthology-based maps. The Y2H-based maps formed own clusters. The CCSB-H1 has the most distinguished set of proteins, whereas MDC-Y2H is placed closer to the remaining maps. These observations indicate that all mapping approaches show their own characteristic preference for proteins included or, in others words, a prominent selection bias.

We verified this conjecture by testing systematically for over- and under-representation of protein categories in interaction maps. The categories used were based on Gene Ontology (GO) that currently represents the most comprehensive system of annotation for the human genome [Ashburner *et al.* '00]. Gene Ontology assigns defined categories

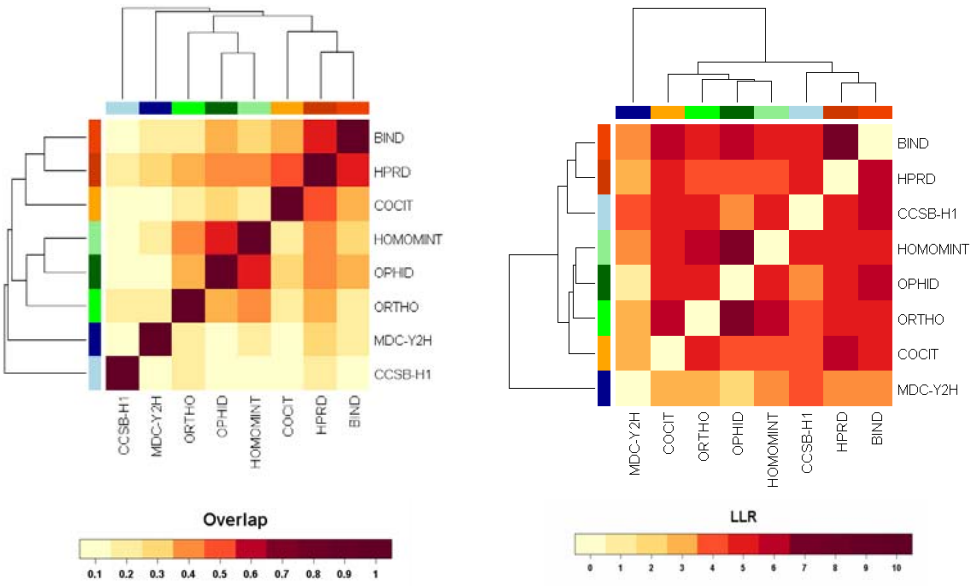

Figure 2: Hierarchical clustering of maps based on protein overlap (A) and log likelihood ratio *LLR* (B) as defined in *Materials and Methods*. The matrices display the (relative) protein overlap, respectively the *LLR* between all possible pairs of maps. Their numerical values are represented according to color-bars at the bottom. On top and right side of each matrices, dendrograms resulted from the clustering are shown. Clustering of protein overlap was based on the distance Δ between map *i* and *j* defined as $\Delta_{ij} = 1\text{-}O_{ij}$ where O_{ij} is protein overlap between maps *i* and *j*. For clustering of *LLR*, the distance Δ was defined as $\Delta_{ij} = 1/LLR(I_i, I_j)$ where $I_{i,j}$ are the sets of interactions included in map *i* or *j*. For both cluster analysis, average linkage was used

to genes according to their molecular function (MF), biological process (BP) or cellular component (CC). First, we tested whether proteins of MF categories are overrepresented in maps using Fisher's test (FDR = 0.01). As reference, the set of all annotated human genes in GO was used. Most maps showed significant enrichment for proteins involved in nucleotide binding (all maps except CCSB-H1) and protein binding (all except ORTHO). Likewise, all maps were found to be enriched by proteins related to metabolism and cell cycle (BP categories) or located in the nucleus (CC category). Interestingly, signal transducers are significantly underrepresented in Y2H- and orthology-based maps, whereas they are significantly overrepresented in literature-based maps. Whereas the reasons for the observed underrepresention are less clear, a possible explanation for the overrepresentation in literature-based maps is the existence of an inspection bias towards 'popular' signaling proteins in the literature. Surprisingly, we detected a highly significant depletion of membrane proteins in all maps including pharmaceutically important classes as the G-protein coupled receptors.

3.3 Concurrence of Interactions

After the comparison of maps based on proteins included, we focused on the concurrence of interactions. To assess the similarity between maps, the *LLR* was calculated for each pair. It ranged from 1.5 (MDC-Y2H- OPHID) to 7.1 (BIND-HPRD) having an average value of 4.6. For all comparisons, it was notably larger than zero, which is the expected value for comparison of random maps. This signifies that the observed concurrence of interaction maps did not occur merely by chance despite of being rather small. To confirm this finding, we applied two permutation tests (described the *Materials and Methods*) for pair-wise comparison of graphs. These results showed that the observed overlap of interactions is highly significant for all comparisons ($p < 0.01$).

Inspection of the *LLR*s also suggested that the interaction maps can be divided into distinct groups. Similarly as before, we subsequently clustered interaction maps to detect common tendencies. The distance was defined as the reciprocal *LLR*. Similar maps score a large *LLR* resulting in a small distance. Hierarchical clustering resulted again in the formation of distinct cluster. However, the detected clusters were differently composed compared to the clusters based on protein overlap. This time, COCIT was found in the group of orthology-based clusters, whereas CCSB-H1 was assigned to the cluster of literature–based HPRD and BIND. MDC-Y2H formed its own separate cluster displaying the weakest similarity to remaining maps. Interestingly, the two large clusters follow exactly the division into computationally generated maps (COCIT, ORTHO, OPHID, HOMOMINT) and maps based on experiments (HPRD, BIND, CCSB-H1). An explanation for this observation is still lacking.

3.4 Coherency of Interaction Maps

Next, we examine the functional coherency of maps. The observation that interacting proteins tend to have common functions has previously been utilized for assessing the quality of interaction maps as well as for *de novo* prediction [Schwikowski *et al.* '00,von Mering *et al.* '02]. To test whether current human interactions maps also display such functional coherency, we employed the gene annotations available in GO. We followed two alternative approaches: First, we assessed the similarity of GO annotations of interacting proteins. In case that interacting proteins have similar functions, their MF annotations should be more similar than expected for random pairs of proteins. This can be measured by the shared path length of GO categories for interacting proteins (see *Materials and Methods*): Assuming a strong correlation between function and interaction (i.e. large functional coherency), we would observe that short shared path length are less likely and long shared path length are more likely than expected. The results of this analysis are shown in figure 3. Indeed, all maps follow this pattern. However, considerable differences can be observed. COCIT showed the largest functional coherency of all maps whereas MDC-Y2H and OPHID showed only modest coherency. A similar analysis was performed for maps with regard to shared process (BP) and location (CC) of interacting proteins. Here, all maps displayed large coherency with only minor differences between maps (figure 3).

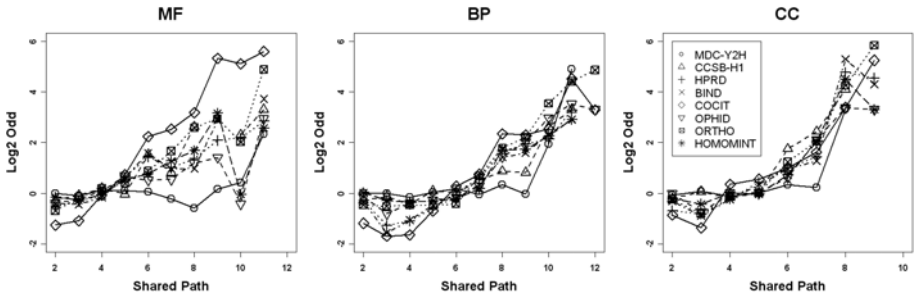

Figure 3: Assessment of coherency based on GO annotations for molecular function (MF), biological process (BP) and cellular component (CC). For interacting proteins, the shared path lengths of GO categories were calculated as described in *Materials and Methods*. The figures show the $\log_2$ odds for the observed path lengths with respect to path lengths derived for random networks. Log_2 odds are plotted as function of shared path lengths.

An alternative approach to study the coherency of interaction maps is the examination whether interacting proteins share a common location. It is based on inspection of the interaction matrix as described in *Materials and Methods*. A similar strategy was introduced by von Mering and co-workers counting the interactions within and between functional categories for yeast interaction maps [von Mering *et al.* '02]. If only interactions of proteins of the same category occur, a diagonal pattern emerges in the corresponding interactions matrices. However, this assumes that proteins are assigned to a single category and not to multiple categories as it is frequently the case for GO annotations. Thus, we modified the approach and compared the observed interaction matrices to matrices of the corresponding randomized networks. Figure 4 displays the log odds for interactions between CC categories of the third level. Interestingly, some compartments (e.g. cytoskeleton) are enriched by internal interactions independently of the map chosen. Generally, however, literature-based networks displayed most prominently enrichment of interactions within proteins of the same component. Less clear patterns for enrichment were found for MC-Y2H and OPHID. This result seems to contradict the previous observation that the coherency for location is similar in all interaction maps (figure 3). However, it is important to note that the interaction matrix approach only assesses the coherency at one particular level of the GO hierarchy. This is contrasted by the previous approach that integrates over all levels. Moreover, overrepresentation of interactions between different categories might not always derive from poor quality of interaction maps, but may point to true biological coupling of cellular compartments. For example, the repeatedly observed enrichment in protein interactions between endomembrane and plasma membrane most likely reflects the close biological connection of both membrane systems.

4. Discussion and Conclusions

Large-scale maps of protein-protein interactions promise to have a considerable impact on the revelation of molecular networks. Similar to fully sequenced genomes serving nowadays the base for genomics, large-scale maps of the interactome might become the foundation for any systematic approach to model cellular networks. Thus, they are likely

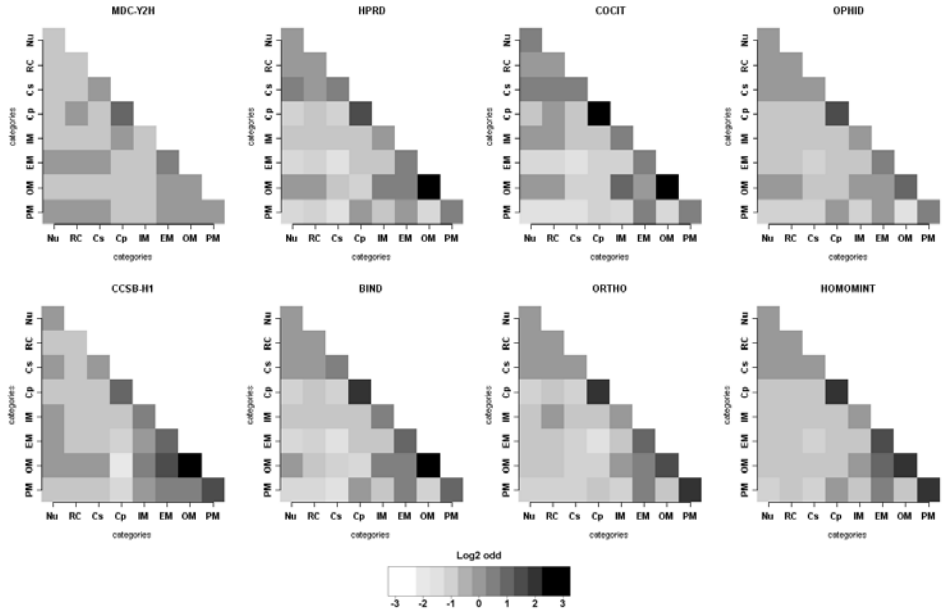

Figure 4: Cellular components of interacting proteins. Pairs of interacting proteins were mapped to the pairs of cellular components to which the proteins are assigned in Gene Ontology. The plots display the $\log_2$ odds ratios of the observed distribution compared to distribution obtained for randomized networks with conserved degree distribution. Categories of the third level of GO were chosen. The following abbreviations were used: Nu – *Nucleus*, RC – *Ribonucleoprotein complex*, Ck – *Cytoskeleton*, Cp – *Cytoplasm*, IM – *Intrinsic to membrane*, EM – *Endomembrane system*, OM – *Organelle membrane* and PM – *Plasma membrane*. For simplicity, only GO categories are shown including more than 2% percent of total number of proteins.

to be of substantial benefit for biomedical researchers. However, a requisite for future application of large-scale human interaction maps is a critical assessment of their quality and reliability. Therefore, we presented here a first comparison of eight currently available large-scale interaction maps. Our comparison is distinguished from previous studies, as it includes all three main approaches currently used for assembly of the human interactome.

A general analysis showed a distinct picture for the concurrence of proteins and interactions in different maps. While a large part of proteins are shared between maps, the interactions included are largely complementary. Only a small percentage of all interactions can be found in multiple maps. This finding has two direct consequences for the integration of maps: The previously proposed approach of assigning higher confidence to interactions found in multiple maps is strongly restricted by the low number of shared interactions [von Mering *et al.* '02]. At the same time, however, the complementary of interactions based on a large overlapping set of proteins indicates that unifying interaction maps will be highly beneficial.

We detected strong sampling and detection biases linked to the approaches used for the generation of the maps. This is reflected by the appearance of distinct groups when

cluster analysis was applied to interaction maps. Such biases have to be observed when interaction maps are utilized. Nonetheless, our analysis showed that most interaction maps display a high internal coherency regarding function, process and location of proteins. This result gives justification for future de novo annotation of proteins based on interaction maps. We like to note that the use of GO for assessment might lead to overestimation of the coherency of literature-based maps, as GO annotations are frequently also based on literature reviews and, thus, do not represent a truly independent benchmark set. In this case, the apparent lack of coherency in other maps could be interpreted that these maps may provide more novel information about the observed interactions.

Although the overlap of protein interactions is statistically significant, it remains small even for maps derived by similar approach. Only 12% - 36% of interactions are shared between orthology-based maps. Possible causes are the use of different data sets and methods for prediction of interactions. Likewise, literature-based maps have only 10% - 28% of their interactions in common. This might result from inspection bias, such as the focus of HPRD towards disease-related genes. Notably, two earlier studies reported contradicting findings for the overlaps between HPRD and BIND. Whereas our study agrees well with results by Ramani and co-workers detecting an overlap of 25%, we cannot confirm the results by Gandhi and colleagues claiming that 85% of interactions in BIND were included in HPRD [Gandhi *et al.* '06, Ramani *et al.* '05]. Finally, a more worrying finding is the minute overlap of mere 1% between interactions in Y2H-based maps underlining the importance of stringent validation of high-throughput data.

In conclusion, this study is aimed to provide a first groundwork for future integration of large-scale human interaction maps [Chaurasia *et al.*]. As we saw, the combination of different maps can be expected to offer great assets. Nevertheless, researchers should be aware of the shortcomings of the underlying mapping approaches.

Acknowledgements

The work presented was supported by the *Deutsche Forschungsgemeinschaft (DFG)* by the SFB 618 grant.

Literature

Ashburner, M. et al. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25:** 25-29.
Bader, G.D. et al. 2001. BIND--The Biomolecular Interaction Network Database. *Nucleic Acids Res* **29:** 242-245.
Bader, G.D. and C.W. Hogue. 2002. Analyzing yeast protein-protein interaction data obtained from different sources. *Nat Biotechnol* **20:** 991-997.
Balasubramanian, R. et al. 2004. A graph-theoretic approach to testing associations between disparate sources of functional genomics data. *Bioinformatics* **20:** 3353-3362.
Benjamini, Y. and Y. Hochberg. 1995. Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J Roy Stat Soc B* **57:** 289-300.
Bork, P. et al. 2004. Protein interaction networks from yeast to human. *Curr Opin Struct Biol* **14:** 292-299.

Brown, K.R. and I. Jurisica. 2005. Online predicted human interaction database. *Bioinformatics* **21:** 2076-2082.
Carey, V.J. et al. 2005. Network structures and algorithms in Bioconductor. *Bioinformatics* **21:** 135-136.
Chaurasia, G. et al. UniHI: An entry gate to the Human Protein Interactome. *submitted*.
Gavin, A.C. et al. 2002. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415:** 141-147.
Gentleman, R.C. et al. 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5:** R80.
Giot, L. et al. 2003. A protein interaction map of Drosophila melanogaster. *Science* **302:** 1727-1736.
Goehler, H. et al. 2004. A protein interaction network links GIT1, an enhancer of huntingtin aggregation, to Huntington's disease. *Mol Cell* **15:** 853-865.
Ito, T. et al. 2001. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci U S A* **98:** 4569-4574.
Jansen, R. et al. 2003. A Bayesian networks approach for predicting protein-protein interactions from genomic data. *Science* **302:** 449-453.
Lee, I. et al. 2004. A probabilistic functional network of yeast genes. *Science* **306:** 1555-1558.
Lehner, B. and A.G. Fraser. 2004. A first-draft human protein-interaction map. *Genome Biol* **5:** R63.
Li, S. et al. 2004. A map of the interactome network of the metazoan C. elegans. *Science* **303:** 540-543.
Peri, S. et al. 2003. Development of human protein reference database as an initial platform for approaching systems biology in humans. *Genome Res* **13:** 2363-2371.
Persico, M. et al. 2005. HomoMINT: an inferred human network based on orthology mapping of protein interactions discovered in model organisms. *BMC Bioinformatics* **6 Suppl 4:** S21.
Ramani, A.K. et al. 2005. Consolidating the set of known human protein-protein interactions in preparation for large-scale mapping of the human interactome. *Genome Biol* **6:** R40.
Rual, J.F. et al. 2005. Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **437:** 1173-1178.
Schwikowski, B. et al. 2000. A network of protein-protein interactions in yeast. *Nat Biotechnol* **18:** 1257-1261.
Stelzl, U. et al. 2005. A human protein-protein interaction network: a resource for annotating the proteome. *Cell* **122:** 957-968.
Uetz, P. et al. 2000. A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature* **403:** 623-627.
von Mering, C. et al. 2002. Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* **417:** 399-403.