BIOINFORMATICS ORIGINAL PAPER

Vol. 23 no. 5 2007, pages 605–611
doi:10.1093/bioinformatics/btl683

Systems biology

Comparison of human protein–protein interaction maps

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

[1]Institute for Theoretical Biology, Charité, Humboldt-Universität and [2]Max-Delbrück-Centrum, Invalidenstrasse 43, 10115 Berlin, Germany

Received on November 3, 2006; revised on December 9, 2006; accepted on January 7, 2007

Advance Access publication January 19, 2007

Associate Editor: Dmitrij Frishman

ABSTRACT

Motivation: 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.

Results: We present here a first comparative analysis of eight currently available large-scale maps with a total of over 10 000 unique proteins and 57 000 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 imply considerable selection and detection biases. These results have to be taken into account for future assembly of the human interactome.

Availability: An integrated human interaction network called Unified Human Interactome (*UniHI*) is made publicly accessible at http://www.mdc-berlin.de/unihi.

Contact: m.futschik@biologie.hu-berlin.de

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 INTRODUCTION

Protein–protein interactions are essential for a wide range of cellular processes 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. First attempts to collect protein–protein interactions on a large scale were initiated for model organisms such as *S.cerevisiae*, *D.melanogaster* and *C.elegans* (Gavin *et al*., 2002; Giot *et al*., 2003; Ito *et al*., 2001; Li *et al*., 2004; Uetz *et al*., 2000). Evidently, the generated interaction maps offered a rich resource for systematic studies of molecular networks.

After these initial efforts, the focus has moved towards deciphering the human interactome. Recently, the first large-scale human protein interaction networks have been constructed using alternative strategies. Most currently available human interaction maps can be divided into three classes: (i) maps obtained from literature search (Bader *et al*., 2001; Peri *et al*., 2003; Ramani *et al*., 2005), (ii) maps derived from interactions between orthologous proteins in other organisms (Lehner and Fraser, 2004; Brown and Jurisica 2005; Persico *et al*., 2005) and (iii) maps based on large scans using yeast-two-hybrid (Y2H) assays (Rual *et al*., 2005; Stelzl *et al*., 2005). 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 be compromised by large false-positive rates. 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 and compare quality and reliability of produced maps.

For yeast interaction maps, several of such critical comparisons have been performed (Bader and Hogue, 2002; Reguly *et al*., 2006; von Mering *et al*., 2002). They revealed a surprising divergence between different interaction maps. Such comparison is still lacking for human protein interaction maps despite their expected importance for biomedical research (Goehler *et al*., 2004). 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*., 2001; Peri *et al*., 2003). 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 assembled by Ramani and co-workers using a text-mining approach (Ramani *et al*., 2005).

*To whom correspondence should be addressed.

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 human protein interaction map was proposed by Lehner and Fraser (2004). 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 the Online Predicted Human Interaction Database (OPHID) and HOMOMINT database (Brown and Jurisica, 2005; Persico *et al.*, 2005). 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 (Rual *et al.*, 2005; Stelzl *et al.*, 2005). We will refer to these maps as MDC-Y2H and CCSB-H1 in our study. Although both scans are based on Y2H assay, considerable differences exist with regard to experimental procedures.

To enable comparison, all proteins were mapped to their corresponding EntrezGene ID. The number of proteins that could not be mapped can be found in the Supplementary Materials. For efficient computational analysis, we converted all interaction maps into graphs using the Bioconductor *graph* package (Balasubramanian *et al.*, 2004; Carey *et al.*, 2005; Gentleman *et al.*, 2004).

2.2 Overlap of interaction maps

Protein interaction maps are formed by both their proteins and interactions included. Comparison of the proteins in different maps is based on the 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 with regard to the total number of proteins in A or B ($P^A_{AB} = |P_{AB}|/|P_A|$; $P^B_{AB} = |P_{AB}|/|P_B|$ where $|P|$ is the number of proteins in set P). 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 *protein overlap* O_{ij} between A and B, i.e. $O_{ij} = 0.5 \cdot (P^A_{AB} + P^B_{AB})$.

Interactions in different maps were similarly compared. For normalization of the intersections, however, only the number of interactions between common proteins is used. This avoids confounding the interaction overlap with the protein overlap, as otherwise a small protein overlap would always lead to a small interaction overlap. Thus, the *interaction overlap* is defined as the average percentage of shared interactions between common proteins.

Although intuitive, the described measure for interaction overlap has the drawback, that it only assesses concurrence of the observed interactions, but not of missing interactions. To evaluate the concurrence of maps for both observed as well as missing interactions, we used a log likelihood ratio (LLR_{AB}) score (Lee *et al.*, 2004). The LLR provides a similarity measure for two sets of interactions (A, B). It is defined as

$$\mathrm{LLR}_{AB} = \ln \frac{P(A \mid B)}{P(A \mid \sim B)}$$

where $P(A \mid B)$ is the probability of observing an interaction in A conditioned on observing the same interaction in B and, $P(A \mid \sim B)$ is the probability of observing an interaction in A conditioned on not observing the same interaction in B, respectively. Note, that we counted again only (missing) interactions between proteins that are included in both A and B. To obtain a reciprocal measure for similarity, we define here $\mathrm{LLR}\ (A, B) = 0.5 \cdot (\mathrm{LLR}_{AB} + \mathrm{LLR}_{BA})$. 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.

In addition to the LLR score, we used two permutation tests to stringently assess the statistical significance of observed concurrence of interactions (Balasubramanian *et al.*, 2004). Details regarding these tests can be found in the Supplementary Materials.

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 (see Supplementary Materials). Since we tested simultaneously for multiple gene ontology (GO) categories, obtained P-values were converted to false discovery rates applying the Benjamini–Hochberg procedure (Benjamini and Hochberg, 1995). 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:

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

Randomized networks had the same number of proteins and interactions as the corresponding maps and conserved connectivity of proteins (number of interactions per protein). Note that the analysis presented here is based on the third GO level. This might appear somewhat arbitrary, but similar results were obtained for GO levels two to five.

Alternatively, we can evaluate the tendency that proteins of similar function interact. Although difficult to define rigorously, similarity of function may be approximated by the following procedure (Jansen *et al.*, 2003): 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 path 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. Details can be found in Supplementary Materials.

2.4 Integration of protein–protein interaction maps

Although not discussed in detail here, the comparison of interaction maps can also be regarded as first step towards the compilation of an integrated human interactome. A more complete description of our efforts to combine interaction maps can be found in a separate publication (Chaurasia *et al.*, 2007). We like to note that the integrated human interaction network called Unified Human Interactome (*UniHI*) is made publicly accessible at http://www.mdc-berlin.de/unihi.

Table 1. List of compared human protein–protein interaction maps

Map	Reference	P	I	D_{av}	Method
MDC-Y2H	Stelzl *et al.* (2005)	1703	3186	3.7	Y2H-ASSAY
CCSB-H1	Rual *et al.* (2005)	1549	2754	3.5	Y2H-ASSAY
HPRD	Peri *et al.* (2003)	5908	15 658	5.2	LITERATURE REVIEW
BIND	Bader *et al.* (2001)	2677	4233	2.9	LITERATURE REVIEW
COCIT	Ramani *et al.* (2005)	3737	6580	3.5	TEXT MINING
OPHID	Brown and Jurisica (2005)	2284	8962	7.8	ORTHOLOGY
ORTHO	Lehner and Fraser (2004)	3503	9641	5.4	ORTHOLOGY
HOMOMINT	Persico *et al.* (2005)	2556	5582	4.2	ORTHOLOGY

The number of proteins P and interactions I result after mappings of proteins to their corresponding EntrezGene ID. D_{av} denotes the average degree of proteins.

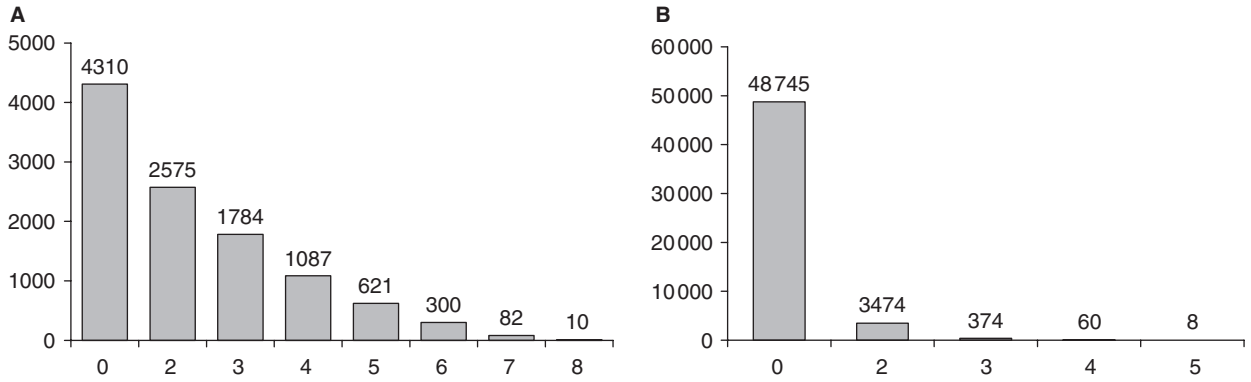

Fig. 1. Number of proteins (**A**) and interactions (**B**) common to multiple maps. The x-axis shows the number of maps in which proteins or interactions are included.

3 RESULTS

In total, we were able to map 57 095 interactions between 10 769 proteins uniquely identified by the corresponding EntrezGene IDs (Table 1). The size of the interaction maps varied between 2754 (CCSB-H1) and 15 658 (HPRD) interactions. Proteins had an average degree of 2.9–7.8 which lies well within the range of previous estimates of 3–10 interactions per protein (Bork *et al.*, 2004).

3.1 Common proteins and interactions

We examined first how many proteins and interactions were common to the different maps in our comparison (Fig. 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. 0.1% 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 eight interactions are common to five maps. Nevertheless, graphical representation shows that the overall network does not disintegrate into separate sub-networks (Fig. S1).

3.2 Protein overlap

To investigate whether some maps tend to share more proteins than others, we calculated the protein overlap for each pair of maps (Table S2). 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 $\Delta_{ij} = 1 - O_{ij}$. 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 clear pattern (Fig. 2A): Maps are grouped in accordance with the mapping approach used for their generation. We obtained two clusters including either literature-based or orthology-based map. 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 underrepresentation of protein categories in interaction maps. The categories used were based on GO

that currently represents the most comprehensive system of annotation for the human genome (Ashburner *et al*., 2000). GO assigns defined categories 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. The results of the analysis can be found in Tables S7–S12 in the Supplementary Materials. 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). Orthology-based maps showed additional enrichment in RNA-binding proteins. 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. Remarkably, 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

Next, we analyzed the shared interactions between different maps. The overlap of interactions between common proteins ranges from 1.8 to 45% (Table S3). Maps were subsequently clustered based on the interaction overlap. As before, characteristic clusters were obtained (Fig. S2). However, the detected clusters were differently composed compared to the clusters based on protein overlap. CCSB-H1 was assigned to the cluster of literature-based maps. In contrast, MDC-Y2H was located separately displaying the weakest similarity to remaining maps. Notably, using the quality score developed for MDC-Y2H resulted in an increase of interaction overlap with most other maps (Supplementary Materials).

For an additional assessment of similarity between maps, the LLR was calculated for each pair (Table S4). It ranged from 1.8 (MDC-Y2H–OPHID) to 6.4 (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 being rather small. To confirm this finding, we applied two permutation tests (described in Section 2) 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 LLRs also suggested that the interaction maps can be divided into distinct groups. 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. The derived clustering resembles closely the results obtained for interaction overlap pointing to a potential detection bias for the maps compared (Fig. 2B). The clustering results also demonstrate that maps derived by literature search or orthology are more convergent. This does not seem to be the case for the Y2H-based approach. The functional analysis of a potential detection bias revealed less clear results (Supplementary Materials).

3.4 Conservation of connectivity of proteins across maps

The question was addressed whether the connectivity of proteins is conserved across interaction maps. To measure the conservation of connectivity between pairs of networks, we correlated the number of interactions of proteins in the two networks using Spearman correlation for the set of common proteins. High correlation between two maps signifies that the interaction-rich (interaction-poor) proteins in one map are also interaction-rich (interaction-poor) in the other map. An overall weakly positive correlation (0.20) ranging from −0.07 to 0.57 was recorded (Table S6). Only 6 out of 28 pair-wise comparisons resulted in correlation coefficients larger than 0.3. Notably, all of these six moderately positive correlations found between maps were generated by similar approaches. Connectivity was less conserved between maps derived by different methods. This is also reflected in a subsequent cluster analysis based on the Spearman correlation. The interaction maps group according to their method of generation (Fig. S3).

3.5 Coherency of interaction maps

Finally, 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*., 2000; von Mering *et al*., 2002). To test whether current human interaction 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 the 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 Section 2): 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 (Fig. 3).

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

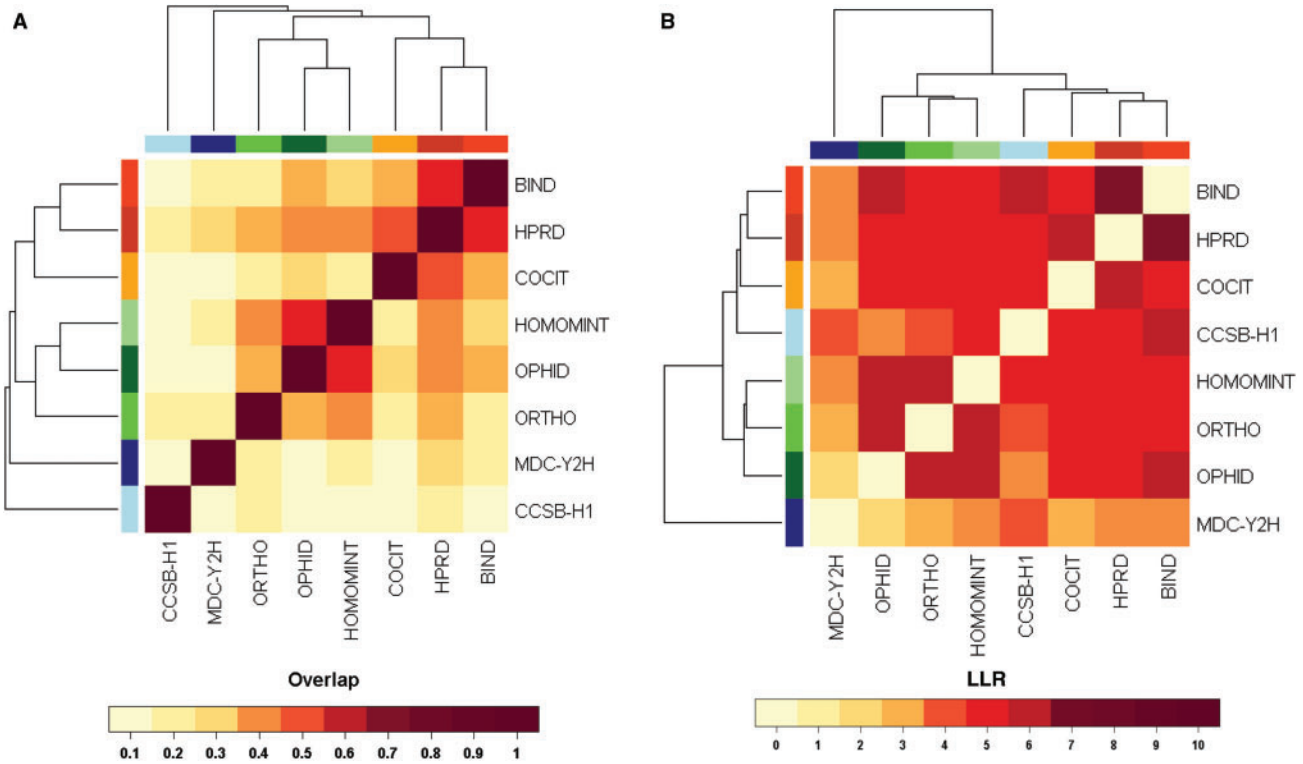

Fig. 2. Hierarchical clustering of maps based on protein overlap (**A**) and log likelihood ratio LLR (**B**) as defined in Section 2. The matrices display the protein overlap, respectively the LLR between all possible pairs of maps. Their numerical values are represented according to color bars at the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

of the interaction matrix as described in Section 2. 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.*, 2002). If only interactions of proteins of the same category occur, a diagonal pattern emerges in the corresponding interactions matrices. However, this approach 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 by comparing 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 whereas figures for MF and BP can be found in Supplementary Materials (Figs S4 and S5). 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 MDC-Y2H and OPHID. This result seems to contradict the previous observation that the coherency for location is similar in all interaction maps (Fig. 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 CCs. 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 protein–protein interaction maps promise to have a considerable impact on the revelation of molecular networks. Similar to fully sequenced genomes serving nowadays as basis for genomics, large-scale maps of the interactome might become the foundation for systematic analysis of cellular networks. However, quality and reliability of large-scale human interaction maps have to be critically assessed. 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. Since it is crucial to know their strengths and weaknesses, our aim was to examine coherency within maps as well as concurrency between maps.

The analysis showed that current maps have only a small, but statistically significant overlap. Whereas most

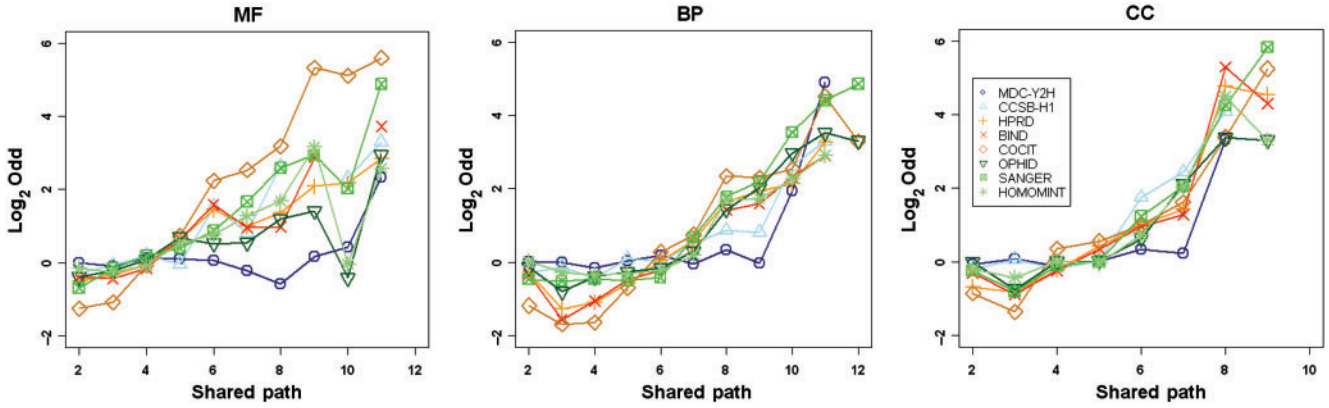

Fig. 3. Assessment of coherency based on GO annotations for molecular function (MF), biological process (BP) and cellular component (CC). For interacting proteins, the shared path lengths of GO categories were calculated as described in Section 2. The figures show the $\log_2$ odds for the observed path lengths with respect to path lengths derived for random networks. Log_2 odds are plotted as function of shared path lengths.

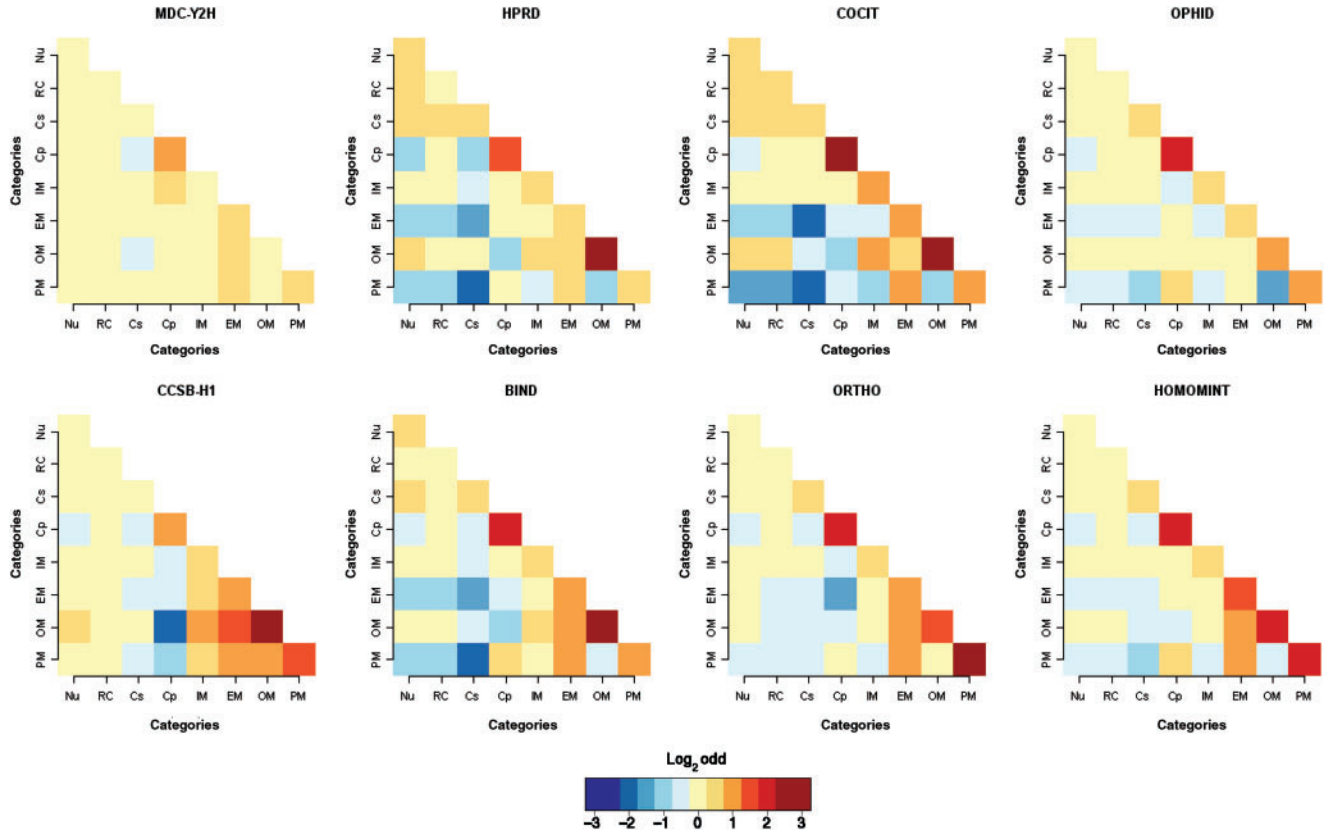

Fig. 4. Cellular components of interacting proteins. Pairs of interacting proteins were mapped to the pairs of cellular components to which the proteins are assigned in gene ontology. The plots display the $\log_2$ odds ratios of the observed distribution compared to the distribution obtained for randomized networks with conserved degree distribution. Categories of the third level of GO were chosen. The following abbreviations were used: Nu—*Nucleus*, RC—*Ribonucleoprotein complex*, Cs—*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% of the total number of proteins.

proteins can be found in multiple maps, this is only the case for <10% of the interactions making the maps largely complementary. The small number of common interactions is somewhat surprising considering the large number of shared proteins, but resembles similar observations in previous comparisons for yeast (Bader *et al.*, 2001; Bader and Hogue, 2002; von Mering *et al.*, 2002).

Naturally, the question arises why the number of common interaction between different current maps is so small. To reveal potential causes for the small overlap, we used different measures to assess the similarity of interaction maps. For literature and orthology-based maps, our study showed that maps were generally more concurrent if they are based on the same method. Less similarity was found between the Y2H-based maps. Strong sampling and detection biases could be linked to the different mapping approaches demonstrating their considerable impact on the resulting maps.

But why do the overlaps remain small even for maps derived by similar approach? For example, only 38–40% of interactions are shared between orthology-based maps. Possible causes are the use of different data sets and methods for prediction of interactions. Although both HOMOMINT and OPHID are based on the approach by Lehner and Fraser, some important deviations exist: For OPHID, interactions from mouse were utilized and a different mapping model to identify human orthologs was applied. For HOMOMINT, information about the domain architecture was additionally used to predict human interactions based on interactions in various organisms stored in the molecular interaction database. Likewise, literature-based maps have only 14–36% of their interactions in common. This might result from inspection bias, such as the focus of HPRD towards disease-related genes. Also, COCIT does not distinguish between physical and functional interactions and lacks self-interactions due to the computational approach taken.

Nonetheless, our analysis also 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. The observed differences in functional coherency also suggest that a differentiation between interaction maps might be favorable for future prediction of protein function. The attachment of larger weights to interactions from maps of large coherency might improve the prediction accuracy. However, 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.

In conclusion, this study is aimed to provide a first groundwork for future integration of large-scale human interaction maps (Chaurasia *et al*., 2007). Combination of different maps can be expected to become great assets. Nevertheless, researchers should be aware of the characteristics of the underlying mapping approaches.

ACKNOWLEDGEMENTS

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

Conflict of Interest: none declared.

REFERENCES

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. and Hogue,C.W. (2002) Analyzing yeast protein-protein interaction data obtained from different sources. *Nat. Biotechnol*., **20**, 991–997.

Bader,G.D. *et al*. (2001) BIND – The biomolecular interaction network database. *Nucleic Acids Res*., **29**, 242–245.

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 Hochberg,Y. (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 Jurisica,I. (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*. (2007) UniHI: an entry gate to the human protein interactome. *Nucleic Acids Res*., **35**, D590–594.

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. USA*, **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 Fraser,A.G. (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.

Reguly,T. *et al*. (2006) Comprehensive curation and analysis of global interaction networks in Saccharomyces cerevisiae. *J. Biol*., **5**, 11.

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.