Indirect Relations in Yeast Protein Interactome

Akihiro Nakaya nakaya@kuicr.kyoto-u.ac.jp
Akiyasu C. Yoshizawa acyshzw@kuicr.kyoto-u.ac.jp
Susumu Goto goto@kuicr.kyoto-u.ac.jp
Minoru Kanehisa kanehisa@kuicr.kyoto-u.ac.jp
Bioinformatics Center, Institute for Chemical Research, Kyoto University, Uji, Kyoto 611-0011, Japan

Keywords: yeast two-hybrid interaction, indirect interaction, autophagy, spindle pole body

1 Introduction

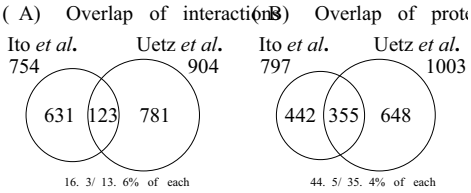

Figure 1: Overlap of the datasets.

As pointed out frequently, there is very little overlap of observed interactions among yeast proteins when more than one method [4] is combined or when more than one experiment with the same method [1] is compared. As an example, focusing on the two yeast two-hybrid (Y2H) interaction datasets by Ito *et al.* ("core data") [1] and Uetz *et al.* [3], the overlap of their interactions corresponds to only 16.3% and 13.6% of each dataset (Fig. 1A).[1] On the other hand, the overlap of proteins appearing in each dataset is larger than that of interactions, corresponding to 44.5% and 35.4% of the proteins of each dataset (Fig. 1B). This means that there is a good possibility that some portions of proteins have different network topologies in respective datasets, giving a clue to a further examination of this kind of protein networks that are seemingly diverging from experiment to experiment.

Besides protein pairs that have direct interactions both in the datasets, one can find proteins that indirectly interact through intermediate proteins in the datasets. Fig. 2 shows examples of such protein pairs. Prs2 (*YER099C*) and Prs5 (*YOL061W*) directly interact both in the two datasets (Fig. 2A). Two MAP kinase-associated proteins Dig2 (*YDR480W*) and Dig1 (*YPL049C*) directly interact in Ito's dataset but indirectly through *YBL016W* in Uetz's dataset (Fig. 2B). Similarly, two proteins that are involved in autophagy Aut7 (*YBL978C*) and Aut1 (*YNR007C*) directly interact in Uetz's dataset but indirectly through *YHR171W* in Ito's dataset (Fig. 2C). According to the descriptions of genes, those intermediate proteins have functional relations with the protein pairs, therefore, such indirect interactions can be one reason why the overlap of interactions in the two datasets is small.

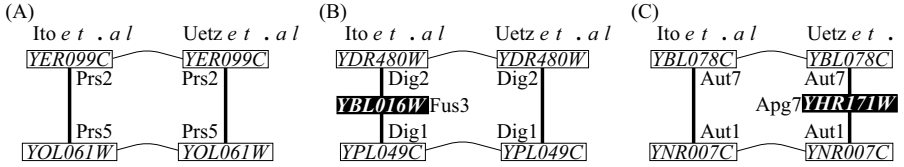

Figure 2: Direct and indirect interactions.

2 Method and Results

Based on the observations above, we first redefined the edges of the protein networks so that the indirect interactions can be explicitly evaluated. Assigning one's to the weights of all the edges of the original network one, we calculated the shortest path length l between two proteins. Then, letting g be a given threshold value, when the shortest path length l between a protein pair is less than $g+1$, we

[1] We excluded redundancy caused by multiple and self-interactions in the data.

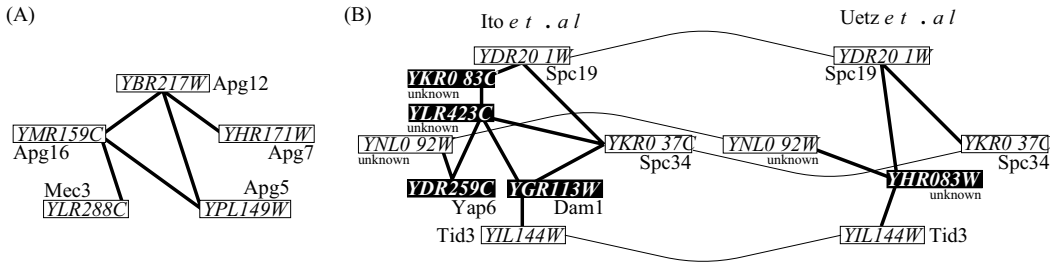

Figure 3: Proteins that are directly or indirectly interacting in both datasets.

defined an edge with weight l between the protein pair. Table 1 shows the effects of this redefinition on the number of edges. By increasing the value of g (the first column), the numbers of edges of Ito's dataset (the second column) and Uetz's dataset (the third column) also increased. The fourth column shows the number of common interactions of the two datasets with the new edge definition. For example, when $g = 3$, *i.e.*, edges are defined between two proteins that interact through at most three proteins, the number increased up to 618 from 123 (= that of the original network).

Using the redefined networks and tools we developed [2], we clustered the proteins so that the proteins belonging to the same cluster are connected through at most g proteins both in the datasets. When $g = 3$, we obtained 90 clusters.[2] For example, five proteins in Fig. 3A are involved in the autophagy process, and directly interact as presented in this figure in the two datasets. Fig. 3B shows that four proteins (open boxes) that are related to the spindle pole body function indirectly interact in both datasets through intermediate proteins (filled boxes).[3] The fifth column of Table 1 shows the total number of the protein pairs in the same cluster. By considering N-to-N relationships among proteins instead of one-to-one relationships, it evaluates the indirect interactions in a more strict manner than that shown in the fourth column, showing that this strategy could extract clusters of proteins indirectly interacting each other.

Table 1: Effects of indirect interactions.

	Indirect interactions		Common	Interactions
g	Ito's dataset	Uetz's dataset	interactions	in the same cluster
0	754	904	123	86
1	4,426	3,153	221	170
2	10,682	6,817	342	233
3	22,071	14,371	618	274

3 Discussion

The advantages of the method presented are: (1) latent interactions can be extracted by considering intermediate proteins, (2) more reliable interactions can be obtained by comparing multiple datasets, (3) obtained results are not dependent on prediction methods and probability models, and (4) those are automatically carried out without human interactions.

References

[1] Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y., A comprehensive two-hybrid analysis to explore the yeast protein interactome, *Proc. Natl. Acad. Sci. USA*, 98(8):4569–4574, 2001.

[2] Nakaya, A., Goto, S., and Kanehisa, M., Extraction of correlated gene clusters by multiple graph comparison, *Genome Informatics*, 12:44–53, 2001.

[3] Uetz, P. *et al.*, A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*, *Nature*, 403:623–627, 2000.

[4] Von Mering, C., Krause, R., Snel, B., Cornell, S., Fields, S., and Bork, P., Comparative assessment of large-scale data sets of protein-proein interactions, *Nature*, 417:399–403, 2002.

[2]The total list of clusters is available from `http://web.kuicr.kyoto-u.ac.jp/~nakaya/pub/giw02/Y2H.html`

[3]These figures correspond to Fig. 3A and 3B of Ito *et al.* [1].