Introduction

1. Illustration of the importance of protein-protein interactions

It is well known that protein-protein interactions are the fundamental essential to explain virtually all the cellular processes (#cite#). Investigating the protein-protein interaction network promotes better understanding towards the protein functions (#cite#) as well as deeper cognition to the regulation bioprocess (#cite#), which leads to a high-level causality between molecular mechanisms and certain phenotypes (#cite#).

2. Review of the experimental resource of protein-protein interactions.

Fortunately, with the fast growth of the new technologies of high-throughput experiments, large-scale genome-wide profiling researches of many most studied species have been reported such as Homo sapiens, Arabidopsis thaliana, Drosophila melanogaster, Saccharomyces cerevisiae and Caenorhabditis elegans (#cite#). Meanwhile, with a number of interactome databases established such as Intact (#cite#), BIND (#cite#), TAIR (#cite#), BioGRID (#cite#), the number of curated experimental interactions keeps growing explosively, with speciﬁc examples from Arabidopsis thaliana, we 6000 trustable experimental interactions (#table#).

However, the known experimental interactions are insufficient to describe the entire protein-protein interaction networks that truly explain biological process both in quantity and quality. Take Yeast for example, the number of PPI is quite limited compared to the estimated Yeast interactome size of 2.85 \* 10^5(#cite# Yu et al., 2008). Also, one considerable side-effect of detecting PI by using high-throughput method is that, the reported results might have a high false positive rate of PI for that the detected physical interactions between proteins are not sufficient to infer PI truly exists without the similar biochemical state.(#cite#)

3. Review of the old PAIR.

Therefore, PPINs (Predicted Protein Interaction Networks) have been calculated to provide an additional approach to obtain more knowledge of protein-protein interactions. PAIR is a protein-protein network, which is designed as a ‘Function Network’, rather than an ‘Interaction Network’. PAIR assumes a conceptual “homogenous functional interactome” which means that the functional interactions in the interactome have similar strengths and represent an unbiased sample of the true (complete) functional interactome (#cite# 2013 Zhou). Consequently, in training the PPFN (predicted protein functional network) in PAIR, the reported physical interaction between proteins is nonequivalent to the “Functional interactions” but merely an intimidation to it (#cite# 2011 Lin). Nevertheless, PAIR might be the best PPIN in recent years (#cite#) if employed as an “Interaction Network” with relatively low false positive rate (#data#) and high coverage (#data#) compared to other PPINs (#cite#).

4. Introduction of new features of the update of PAIR.

Compared to PAIR v3.0, in PAIR v4.0, we managed a more wide-ranged full-profiling data collection from newly updated database till the end of 2013 (#table#) including TAIR (#cite#), SUBA3 (#cite#), GOC (#cite#), RoudUp (#cite#), Inparanoid (#cite#), Weigel World (#cite#). Meanwhile, we attempted to extract features from original data by several new statistics methods whose power and effect would have been tested by their AUC of ROC. With all these new methods applied in PAIR v4.0, we’ve calculated a new PPIN by SVM with a wider coverage (#data#), and a lower false positive rate (#data#).

5. Demonstration of the applications of PAIR.