

Predicting Gene Functions: Departing From Single Gene – Single Function Paradigm

Daisuke Kihara

Elucidating gene function is a central problem in modern biology, including molecular biology, genetics, and bioinformatics. Ultimately, biological function of genes needs to be characterized by experimental methods; however, with a wealth of data of known genes, now it is feasible to predict function of many genes by computational methods.

We will first introduce sequence-based function prediction methods we have developed, named PFP and ESG. These two methods considers weakly similar sequences to achieve higher coverage in genome-scale function assignments and also to improve accuracy. PFP an ESG (1,2) were ranked among the top in recently held community-based function prediction contest, CAFA (3).

Next, we will extend function prediction to predict proteins that have two or more independent functions (moonlighting proteins). We recently develop a computational prediction model, MPFit (4), for automatic identification of MPs using a diverse range of protein association information. we applied the method on three genomes: *Saccharomyces cerevisiae*, *Caenorhabditis elegans* and *Homo sapiens*, and found that about 2-10% of proteins in the genomes are potential MPs. Finally, we will introduce our current effort to predict function of genes in the context of biological systems. Function of a gene is determined not only by the gene itself; but also by the other genes in the same pathways or by the cellular environment where the gene appear. We present our recent results toward the systems level gene function prediction.

References:

1. PFP/ESG: Automated protein function prediction servers enhanced with Gene Ontology visualization tool. Khan, Q. Wei, M. Chitale, & D. Kihara, *Bioinformatics*, 31: 271-272 (2015)
2. PFP and ESG protein function prediction methods in 2014: Effect of database updates and ensemble approaches. I.K. Khan, Q. Wei, S. Chapman, Dukka, B.K.C., & D. Kihara, *GigaScience*, 4: 43. (2015)
3. An expanded evaluation of protein function prediction methods shows an improvement in accuracy. Y. Jian et al., *Genome Biology.*, 17: 184 (2016).
4. Genome-scale prediction of moonlighting proteins using diverse protein association information. I. K. Khan & D. Kihara. *Bioinformatics*, 32: 2281-2288 (2016).