

Title: Network-guided prediction of function-driven domains

Authors: Jung Eun Shim, Insuk Lee

Affiliation: Yonsei University, Korea

E-mail: jjuggeuni@hanmail.net, insuksysbio@gmail.com\*

Identifying the functions of biological units such as proteins, genes and protein domains is important for the understanding the biological processes within a cell. In particular, protein domain is useful feature, because it is the structural, functional and evolutionary units of proteins. However, even if the functions of many genes (or proteins) have been known, it is quite difficult to identify the functions of domains that they include.

In this study, we developed a network-guided prediction model to identify the function-driven domains which play a critical role to drive functions of their proteins, using *Domain Information Content Score (DomICS)*. In addition, we also constructed the domain co-occurrence network to infer the functional probability of each domain, scored by *Weighted Mutual Information (WMI)* in which we give larger weights to rare domains.

The performance of the proposed network was analyzed for the two organisms: microbe yeast (*Saccharomyces cerevisiae*) and multi-cellular human (*Homo sapiens*). To evaluate the predicted function-driven domains, we performed the survey of literatures and the enrichment analysis with known domains assigned in Gene Ontology (GO) term and Interpro2GO.