Abstract

Computational Biology deals with problems arising from the interactions between genes and proteins and how it could be resolved with the help of computer scientists. Information from genes get transformed into proteins which are actually involved in monitoring the different traits in an individual. This process is known as *Central Dogma*. Proteins become functional only by interacting with other proteins. This interaction leads to the different functionalities in the living organism. Interaction between proteins is represented in the form of a network called Protein Protein Interaction Network (PPIN). Analyzing such a network in terms of similar proteins, technically known as protein complexes has been known to occupy the research domain for years. These complexes are then used in studying the behavior of proteins which may be involved in causing certain kinds of diseases. This PhD work reports four such protein complex finding methods namely CNCM, DCRS, CSC and ComFiR. All these methods have been shown to extract biologically relevant complexes. The quality of complexes obtained using these methods have been compared and reported w.r.t. benchmark complexes. The protein complexes obtained using CSC and ComFiR have been extended towards analyzing their association with Alzheimer's Disease.

In order to get more realistic correlation between genes and its role in diseases, one can analyze their activity under different conditions. Genes become active only by expressing themselves and their activity is recorded in the form of gene expression data. Analyzing this expression data is of utmost importance as they can yield interesting findings. Two such works are discussed in this thesis. The first work deals with finding gene modules across two stages of breast cancer, i.e., metastasis and non-metastasis stage. It also highlights some novel biomarkers associated during the progression of disease. The next work discusses gene module finding from a subset of active conditions. This work has been analyzed in case of Parkinson's Disease. It discusses the correspondence between gene modules in a healthy person w.r.t. a diseased person.

Keywords — Protein protein interaction network, protein complex, gene network, gene module, subspace network, pathway, biomarkers, Topological subspace overlap matrix, module correspondence, hub gene.