

ontology (GO) and disease ontology (DO), protein structure data, drug chemical structure and drug–target interaction (DTI) data. These data represent a valuable complement to the molecular networks and they are often incorporated into various data integration frameworks to increase reliability of newly discovered knowledge.

Data Integration Approaches

- **Network inference and functional linkage network (FLN) construction.** Network inference is one of the major problems in systems biology aiming to understand GIs and their mutual influence. It aims to construct network topology (or wiring between genes) based on the evidence from different data types. The special focus in the literature has been on the inference of gene regulatory networks (GRNs), whose nodes are genes and edges are regulatory interactions. Standard methods for GRN inference have mostly been based on gene expression data.
- **Protein function and PPI prediction.** Protein function (also known as protein annotation) prediction has been demonstrated to be a good alternative to the time-consuming experimental protein function characterization. It aims to computationally assign molecular functions to unannotated proteins. The accuracy of these methods has largely improved with the use of integration methods that can incorporate multiple different biological data conveying complementary information about protein functions. Also, interactions between proteins are important for understanding intracellular signalling pathways and other cellular processes.
- **Disease gene prioritization and disease–disease association prediction.** Prioritization of disease-causing genes is a problem of great importance in medicine. It deals with the identification of genes involved in a specific disease and providing a better understanding of gene aberrations and their roles in the formation of diseases. However, a majority of diseases are characterized with a small number of known associated genes and experimental methods for discovering new disease-causing genes are expensive and time-consuming. Therefore, computational methods for prioritization of disease genes have been proposed.
- **Drug repurposing:** It aims to find new uses for existing drugs and thereby drastically reduce the cost and time for new drug discovery. Accumulation of various biological data involving interactions between drugs, diseases and genes, and protein structural