

Enabling comparative research through multi-omics data harmonization

The biological data generated in the last few years from omics-based approaches have grown exponentially which have led to a high dimensional and unbalanced nature of the data. This brings new challenges in terms of data management, query, and analysis. One of the challenges that still need to be addressed is the incompleteness inherent in these data, i.e., several types of genomic/phenotypic information covering only a few of the genotypes under study. In this talk, we address the complex issue of utilizing the high dimensional and unbalanced omics data by combining the relationship information from multiple data sources, and how we can facilitate data integration from interdisciplinary research. More specifically, we propose a covariance-based method for combining partial datasets in the genotype to phenotype spectrum. Several examples will be shown to demonstrate the application ability of the method.