ABSTRACT
INTEGRATION OF MULTI-OMICS DATA TO COMPUTE GENE REGULATORY NETWORKS AND TO PREDICT PERSONALIZED DRUG RESPONSE IN CANCER

Banabithi Bose

Marquette University, 2021

With the advancement of high throughput technologies, many repositories of the genome, proteome, transcriptome, and epigenome datasets, called multiple "omes" or multi-omics, are now accessible to researchers worldwide. There are several omics layers, such as gene expression, mutation, copy number aberration, DNA methylation, microRNA (miRNA) expression, and transcription factor binding that are generated from large sets of biospecimens and describe different aspects of the underlying biology. Integration of these multi-omics datasets along with clinical information opens up an opportunity to utilize the complementary information from several individuals in the understanding of human diseases and paves the path towards personalized medicine. The overall goal of this dissertation was to leverage the integration of multi-omics datasets to compute a regulatory gene network and to develop a prediction model of patient-specific drug response in cancer. To this end, we designed a computational tool, miRDriver, which computes miRNA-gene interaction network based on frequently aberrated chromosomal regions of tumor samples utilizing a LASSO-based regression model. We carried out a pan-cancer-wide rigorous analysis with miRDriver and predicted several known and novel cancer-related miRNAs and genes in multiple cancer types. These findings were associated with patients' survival and disease progression significantly. Studies found that pre-clinical testing of cancer drugs on cell lines (i.e., tumor-derived cells grown *in vitro*) may deviate from actual patients' drug response leading towards adverse drug reaction among cancer patients. Hence, it is crucial to understand the drug response in cancer patients accurately before providing the drug therapy. To this end, we designed a computational tool, PDDRNet-MH, which utilizes multiple networks of patient-drug associations along with the multi-omics interplay between patient samples and cancer cell lines to capture patient-specific anti-cancer drug response. To relate patient samples and cancer cell lines in PDDRNet-MH, we developed another computational tool, CTDPathSim, which computes similarity scores between patient samples and cancer cell lines utilizing a biological pathway activity-based approach integrating multi-omics datasets. These computed scores recapitulate drug response phenotypes between patient samples and cell lines successfully. All these three tools outperformed several state-of-the-art-methods and were able to capture biologically meaningful results along with clinically proven findings. Furthermore, we implemented our tools as R packages to provide the opportunity to reproduce our results and integrate these tools into relevant use cases.