

Statistical methods for integrative clustering analysis of multi-omics data

Qianxing Mo

Department of Biostatistics & Bioinformatics, H. Lee Moffitt Cancer Center & Research Institute

E-mail: Qianxing.Mo@moffitt.org

Abstract: In an effort to identify clinically relevant tumor molecular subtypes and driver omics signatures, we developed integrative clustering methods to jointly model multi-omics data. These methods can model multi-omics data with continuous, count, binary and multi-categorical data types. In the integrative clustering framework, latent variables are used to capture the inherent structure of multi-omics data and variable selection algorithms are used to select important omics features. As a result, the samples can be clustered in the low dimensional latent variable space and driver features contributing to the sample clustering are identified through variable selection. Several multi-omics data sets from large scale genomics studies will be used to illustrate the methods.