Integrative Analysis of Heterogeneous Genomics Data

Xianghong Jasmine Zhou[1],*

[1]University of Southern California, Los Angeles, CA 90089-1340, USA

Abstract   The rapid accumulation of genomics data in public repositories poses new challenges on powerful methods for integrative analysis. In this talk, we will report our recent progresses in addressing some of these challenges. Firstly, we developed a tensor computation framework for pattern mining across many massive weighted biological networks. Specifically, we formulated the recurrent heavy subgraph identification problem into a heavy 3D subtensor discovery problem with sparse constraints. We applied our method to 129 coexpression networks, and identified 4,327 recurrent heavy subgraphs. We demonstrated that the identified subgraphs represent meaningful biological modules by validating against a large set of compiled biological knowledge base. We show that the likelihood for a heavy subgraph to be meaningful increases significantly with its recurrence in multiple networks, highlighting the importance of the integrative approach. Secondly, we developed a joint matrix factorization approach to perform analysis of multi-dimensional genomics profiling of the same set of samples. We apply the method to the mRNA, miRNA, and methylation profiles of 385 ovarian tumor samples from the Cancer Genome Atlas Project to study the coordination of those different regulatory levels. Our study has the potential in uncovering the cross-layer coordinated modules and implications across multiple heterogenous 'omic' datasets.

---

*E-mail: xjzhou@college.usc.edu