Network Inference Using High Dimensional Genomics Data

by

Pei Wang Division of Public Health Science, Fred Hutchinson Cancer Research Center 1100 Fairview Ave N. Seattle, WA, 98107 pwang@fhcrc.org

Abstract

In this talk, I will introduce a new method remMap — REgularized Multivariate regression for identifying MAster Predictors — for fitting multivariate response regression models under the high-dimension-low-sample-size setting. remMap is motivated by investigating the regulatory relationships among different biological molecules based on multiple types of high dimensional genomic data. Particularly, we are interested in studying the influence of DNA copy number alterations on RNA transcript levels. For this purpose, we model the dependence of the RNA expression levels on DNA copy numbers through multivariate linear regressions and utilize proper regularizations to deal with the high dimensionality as well as to incorporate desired network structures. Criteria for selecting the tuning parameters are also discussed. The performance of the proposed method is illustrated through extensive simulation studies. Finally, remMap is applied to a breast cancer study, in which genome wide RNA transcript levels and DNA copy numbers were measured for 172 tumor samples. We identify a tran-hub region in cytoband 17q12-q21, whose amplification influences the RNA expression levels of more than 30 unlinked genes. These findings may lead to a better understanding of breast cancer pathology.