A robust Bayesian procedure for detecting copy number variations from sequencing read depths

Yu-Chung Wei* (魏裕中) and Guan-Hua Huang (黃冠華)
Institute of Statistics, National Chiao Tung University

Abstract

Copy number variations (CNVs) are genomic structural mutations with abnormal gene fragment copies. Read depths (RDs) signal mirrored the variants directly from the next generation sequencing data. We adopt a Bayesian hierarchical model that is suitable for both whole genome and targeted exome sequencing RD data with separate bias adjustments. This model is applied not only to a single sample for identifying the number of copies but also to case-control ratio for detecting the patient specific variations. In order to estimate the varying dimension states and breakpoints of CNV regions, we proceed to a new reversible jump Markov chain Monte Carlo (RJMCMC) algorithm with four moves including merge, split, trifid and boundary change. We demonstrate this pragmatic approach with targeted region exome sequencing data from the National Taiwan University Hospital.

Keywords: copy number variations, next generation sequencing, reversible jump Markov chain Monte Carlo, Bayesian inference