Integrated studies of copy number alterations and genotype

Wei Sun
Department of Biostatistics and Department of Genetics
University of North Carolina, Chapel Hill

We propose a statistical framework, named genoCN, to simultaneously dissect copy number states and genotypes using high density SNP arrays. There are at least two types of genomic DNA copy number differences: Copy Number Variations (CNVs) and Copy Number Aberrations (CNAs). While CNVs are naturally occurring and inheritable, CNAs are acquired somatic alterations most often observed in tumor tissues only. CNVs tend to be short and more sparsely located in the genome compared to CNAs. GenoCN consists of two components, genoCNV and genoCNA, designed for CNV and CNA studies, respectively. In contrast to most existing methods, genoCN is more flexible in that the model parameters are estimated from the data instead of being decided a priori. GenoCNA also incorporates two important strategies for CNA studies. First, the effects of tissue contamination are explicitly modeled. Second, if SNP arrays are performed for both tumor and normal tissues of one individual, the genotype calls from normal tissue are used to study CNAs in tumor tissue. We evaluated genoCN by applications to 162 HapMap individuals and a brain tumor (Glioblastoma) dataset and showed that our method can successfully identify both types of CN differences and produce high quality genotype calls.