

Emerging evidence suggests that copy number variation (CNV), as a type of genetic polymorphism, influences risks of many diseases such as autism, psoriasis and schizophrenia. With the rapid growth of modern genotyping technology, massive high-throughput genetic datasets have been generated that provide a great resource for CNV studies. Although many statistical approaches are developed at a fast pace, it remains persistently difficult to accurately identify disease associated CNVs. The problem is mainly due to high dimensionality, technical noise present in the data and intrinsic characteristics of different platforms. The difficulties in high dimensionality and technical noise demand efficient and accurate segmentation procedures. In this talk, we will discuss the current existing algorithms and computational tools for CNV detection for both array-based and sequencing-based data, and the challenge arising from the application of these methods. Then we will talk about an ongoing project of methodology development for accurate and powerful CNV detection. Among the existing segmentation algorithms for CNV analysis, it is usually assumed that neighboring loci across the genome are independent. The existence of inter-correlation among genomic positions, such as linkage disequilibrium (LD), is therefore ignored. To allow for more accurate statistical modeling, we propose a novel change-point algorithm assuming dependency of genetic locations related to physical distance or genetic distance (e.g., LD). The new method will have the potential to be implemented in the segmentation procedures for developing analytical tools for both array-based data and sequencing-based data.