UNDERSTANDING GENOMIC ABERRATIONS THROUGH ARRAY CGH DATA

C. Myers

Princeton University, Princeton, NJ, USA

† E-mail: clmyers@princeton.edu

Array-CGH technology has enabled genome-wide, high-resolution studies of copy number changes, which will undoubtedly play a key role in answering important medical and biological questions. However, statistically rigorous methods are required to take full advantage of the information contained in array-CGH data. We have developed ChARM (Chromosomal Aberration Region Miner), an approach for automated statistical identification of segmental copy changes. ChARM makes use of a differential filter, a formulation of the expectation maximization (EM) algorithm, and non-parametric statistical tests to achieve accurate breakpoint prediction as well as an assignment of statistical significance to predicted regions. Our approach is particularly suited for detecting subtle changes (e.g. those present in subpopulations of cells) and has been successfully applied in both array CGH studies of Saccharomyces cerevisiae and human cancer data. Another important challenge in analyzing array-CGH data is understanding which genomic changes are most functionally relevant (i.e. responsible for a particular phenotype). We have developed two useful approaches to this problem: pattern identification in multiple CGH arrays sharing a phenotype, and incorporation of parallel gene expression data as a measure of functional relevance. I will discuss the details and examples of these approaches as well as important characteristics of array-CGH data we have encountered along the way.