

Statistical Challenges in Microarray Data Analysis: Past, Present, and Future.

Sandrine Dudoit

*Division of Biostatistics and Department of Statistics University of California, Berkeley
www.stat.berkeley.edu/~sandrine*

Microarray-based assays have become standard in high-throughput studies of gene expression. Microarray data analysis exemplifies the inherent and crucial role of statistical and computational methods in biomedical and genomic research. This talk will provide an overview of past, present, and future statistical challenges in microarray data analysis, from pre-processing (image processing and normalization) to higher-level analyses, such as differential expression and biological annotation metadata analysis. A brief review of the early days of microarray data analysis will be given. The present state of the field will be illustrated using a number of case studies, each investigating different aspects of gene expression using different microarray platforms: tests of association between Gene Ontology annotation and differential gene expression in acute lymphoblastic leukemia; study of gene regulation by alternative splicing and nonsense mediated decay in *Drosophila melanogaster* using exon and splice junction microarrays; genome annotation in *Saccharomyces cerevisiae* using tiling microarrays; bacterial metagenomics using 16s small-subunit rRNA microarrays (PhyloChip). Finally, questions about the future of microarray data analysis will be raised.