Dr. Martin Morgan, Fred Hutchinson Cancer Research Center, Washington

R / Bioconductor: Challenges and Rewards from Integrative Analysis of High-Throughput Genomic Data

Abstract:

It used to be that high-throughput genomic data sets were large, and this was a challenge. Increasingly, our experiments involve several different types of data (RNA-seq or microarray gene expression, copy number variation, called variants, regulatory marks...) from our own work and from reference 'consortia' like TCGA, ENCODE, PanCan, CCLE, and Roadmap Epigenomics. These data offer very rich possibilities for insight, and our software should help us gain these insights in an easy and reproducible way. Bioconductor is in many ways ideally suited to this task. Integrative analysis is often on data sets that have been processed to moderate size, rather than on very large primary data. Key insights necessarily come from statistical analysis, some of these analyses are quite nuanced and supported by years of research and application. Managing different data sets requires careful book-keeping and strategies for reproducibility, both areas in which Bioconductor has invested considerable effort. This talk highlights the ways in which Bioconductor can be used to harness integrative analysis of high-throughput genomic data.