Bioconductor for High-Throughput Sequence Analysis

Martin Morgan[1,*]

1. Fred Hutchinson Cancer Research Center, Seattle, WA
[*]Contact author: mtmorgan@fhcrc.org

Keywords: Bioinformatics, DNA Sequence, RNA-seq, ChIP-seq

DNA sequence analysis generates large volumes of data presenting challenging bioinformatic and statistical problems. This tutorial introduces *Bioconductor* packages and work flows for the analysis of sequence data. We learn about approaches for efficiently manipulating sequences and alignments, and introduce common work flows and the unique statistical challenges associated with 'RNA-seq', 'ChIP-seq' and variant annotation experiments. The emphasis is on exploratory analysis, and the analysis of designed experiments. The workshop assumes an intermediate level of familiarity with R, and basic understanding of biological and technological aspects of high-throughput sequence analysis. The workshop emphasizes orientation within the *Bioconductor* milieu; we will touch on the **Biostrings**, **ShortRead**, **GenomicRanges**, **edgeR**, and **DiffBind** packages, with short exercises to illustrate the functionality of each package. Participants should come prepared with a modern laptop with current *R* installed.