Title: Bioconductor for high-throughput sequence analysis

Contact:
Dr. Martin Morgan
Bioconductor / Program in Computational Biology
Fred Hutchinson Cancer Research Center
Seattle, WA USA

Overview

DNA sequence analysis generates large volumes of data presenting challenging bioinformatic and statistical problems. This tutorial introduces established and new Bioconductor (http://bioconductor.org) 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 'RNAseq', variant annotation, and other experiments. The emphasis is on exploratory analysis, and the analysis of designed experiments.The workshop emphasizes orientation within the Bioconductor milieu; we will touch on the Biostrings, ShortRead, GenomicRanges, DESeq2, and VariantAnnotation, and other packages, with short exercises to illustrate the functionality of each package.

Goals

- Gain overall familiarity with Bioconductor packages for high-throughput sequence analysis, including Bioconductor vignettes and classes.

- Obtain experience running bioninformatic work flows for data quality assessment, RNA-seq differential expression, and manipulating variant call format files

- Appreciate the importance of ranges and range-based manipulation for modern genomic analysis

- Learn 'best practices' for working with large data

Outline

- Introduction to Bioconductor -- packages and classes

- Short work flows
  - Exploring sequences and alignments
  - RNA-seq: a high-level tour
  - Annotating variants

Prerequisites

The workshop assumes an intermediate level of familiarity with R, and basic understanding of biological and technological aspects of high-throughput sequence analysis. Participants should come prepared with a modern wireless-enabled laptop and web browser installed.

Intended Audience

This workshop is for professional bioinformaticians and statisticians intending to use R / Bioconductor for analysis and comprehension of high-throughput sequence data.