Week 13 RNASeq analysis with BioConductor

Introduction

Ultra-high-throughput sequencing of cDNA samples (RNA-Seq) was first introduced in this paper from Barbara Wold's lab at CalTech:

- Mortazavi (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq
    - http://www.ncbi.nlm.nih.gov/pubmed/18516045

The paper is not open access and UNCC does not have an electronic subscription to the journal. However, you can get a copy from the library. You might also be able to locate a copy by google searching the title plus the string "PDF."

The paper, although ground-breaking, is out of date in that the read lengths are much shorter than read lengths that are possible now, which means many of the concerns regarding multi-reads are a non-issue these days.

Since the paper was published, many groups have begun to use RNA-Seq to assay differential gene expression (DGE) on a genome-wide scale.

There are two major methods for assaying DGE used RNA-Seq data.

One of these methods, which we won't cover in this class, uses read alignments to build transcript models and then associates read counts to exon "fragments." This approach is implemented in the CuffLinks suite of programs, related to TopHat.

The other approach, which you'll use in the assignment, counts the number of reads that overlap each gene and then applies a statistical test to determine if the number of reads per gene is significantly different as a result of a treatment, genotype, or other experimental factors.

This approach is simpler because it doesn't depend on correctly grouping reads into individual transcripts and allows testing of the gene as a whole.

Also, the results are much more comparable to microarray experiments, where the methods for interpretation and downstream testing are well-established.

One of the major challenges with DGE analysis in RNA-Seq data is that read counts are not normally distributed, which means we need to use a different kind of test in order to detect differentially expressed genes. In the assignment, you'll learn how to use the EdgeR software to detect DE genes from RNA-Seq data, using a not-yet published data set from the Loraine lab.

References

- **edgeR at Bioconductor Web site**
    - http://www.bioconductor.org/packages/release/bioc/html/edgeR.html

To view the edgeR User's Guide, click the link labeled User's Guide at the above Web page.

Assignments

- Statistical analysis of RNA-Seq data due by 5 pm Wed Dec 5