RNAseq

BCMB bootcamp

RNAseq: sequencing all of the RNA in a sample

- gives a sense for what the cell is doing and how it is responding to environmental cues

- bulk RNAseq cannot distinguish single cells; it sequences a mix of all of the cells in the sample

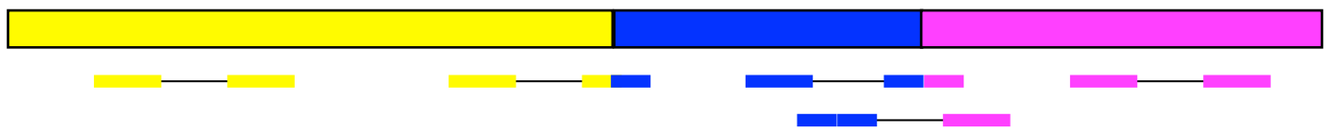
enrichment for non-ribosomal RNAs

- ribosomal RNA is 90% of the RNA in a cell so we need to enrich for mRNA
- two categories of methods for enriching mRNA:
    - poly(A) selection
    - amplification with random primers that are depleted for rRNA sequences
- with either of these it is possible to get 80-90% mRNA in the final sequences

the alignment problem

3-exon gene in genomic DNA

spliced mRNA



paired end sequencing reads don’t align concordantly to the genome

solving the alignment problem

- use a spliced aligner and align to the genome (HISAT2, TopHat, STAR)
- align to an assembled transcriptome (RSEM)

steps in RNAseq analysis using HISAT2

- index the genome using HISAT2-build
- align the sequencing reads to the genome with HISAT2
- use Cuffdiff to quantify gene expression and determine differential gene expression
- look at differentially expressed genes in IGV (http://software.broadinstitute.org/software/igv/)

samples

- paired end RNAseq from the epigenetics roadmap project
- start with HUES64, an undifferentiated cell line, and differentiate the cells into ectoderm, mesoderm, and endoderm lineages using different growth factors
- RNAseq and ChIPseq on all samples