bioinformatics.ca

Canadian Bioinformatics Workshops

www.bioinformatics.ca

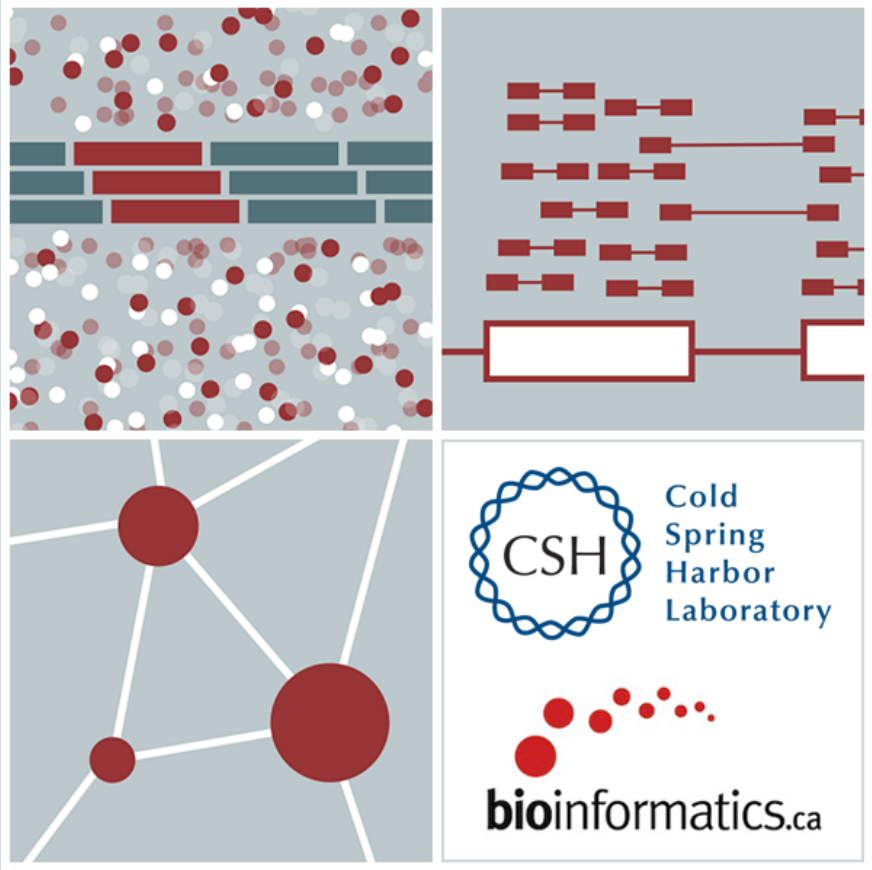
RNA-Seq Module 2
Alignment and Visualization (tutorial)

Malachi Griffith, Obi Griffith, Fouad Yousif

Informatics for RNA-seq Analysis

July 10-12, 2017

bioinformatics.ca



THE GENOME INSTITUTE at Washington University

Learning Objectives of Tutorial

- Run HISAT2 with parameters suitable for gene expression analysis
- Use samtools to demonstrate the features of the SAM/ BAM format and basic manipulation of these alignment files (view, sort, index, filter)
- Use IGV to visualize RNA-seq alignments, view a variant position, etc.
- Determine BAM-read counts at a variant position
- Use samtools flagstat, samstat, FastQC to assess quality of alignments

2-i. Adaptor trim

- Use Flexbar to trim sequence adapter from the read FASTQ files
  - The output of this step will be trimmed FASTQ files for each data set.
- Compare the FastQC reports for fastq files before and after trimming
- http://sourceforge.net/projects/flexbar/

2-ii. Align reads with HISAT2

- Align all reads in the 6 libraries of the test data
  - 6 libraries with two files each (one for each read1 and read2 of the paired-end reads)
- Use HISAT2 for the alignment
  - Supply the bowtie indexed genome obtained in section 1-iv
  - The '-dta' option tells HISAT2 to report alignments tailored for transcript assemblers
- Since there are 6 libraries in the test data set, 6 alignment commands are run
- On a test system, each of these alignments took ~4 seconds using 8 CPUs
- Each alignment job outputs a SAM/BAM file
  - http://samtools.sourceforge.net/SAM1.pdf

2-iii. Post-alignment visualization

- Create indexed versions of bam files
  - These are needed by IGV for efficient loading of alignments
- Visualize spliced alignments
  - Identify exon-exon junction supporting reads
  - Identify differentially expressed genes
- Try to find variant positions

- Create a pileup from bam file
- Determine read counts at a specific position

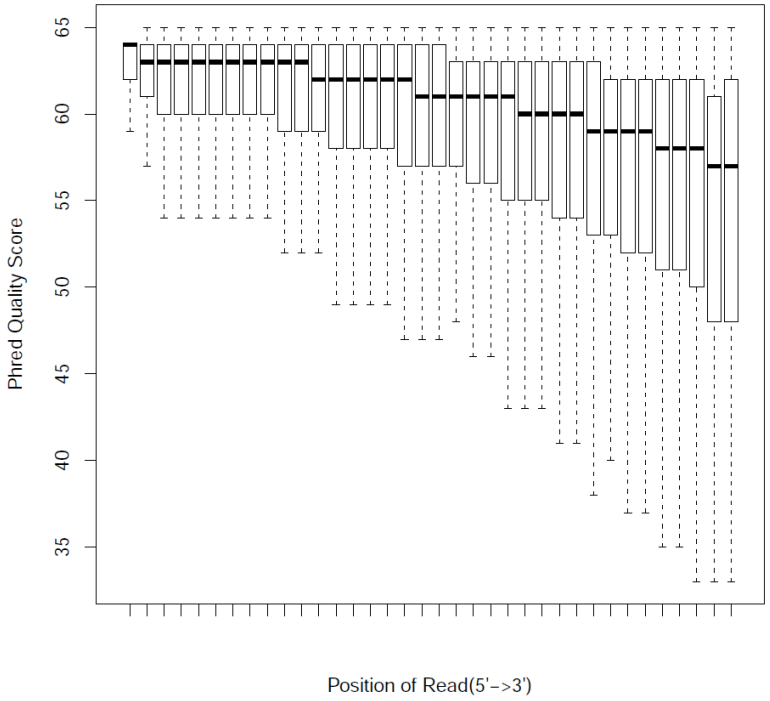
2-iii. Post-alignment visualization (IGV)

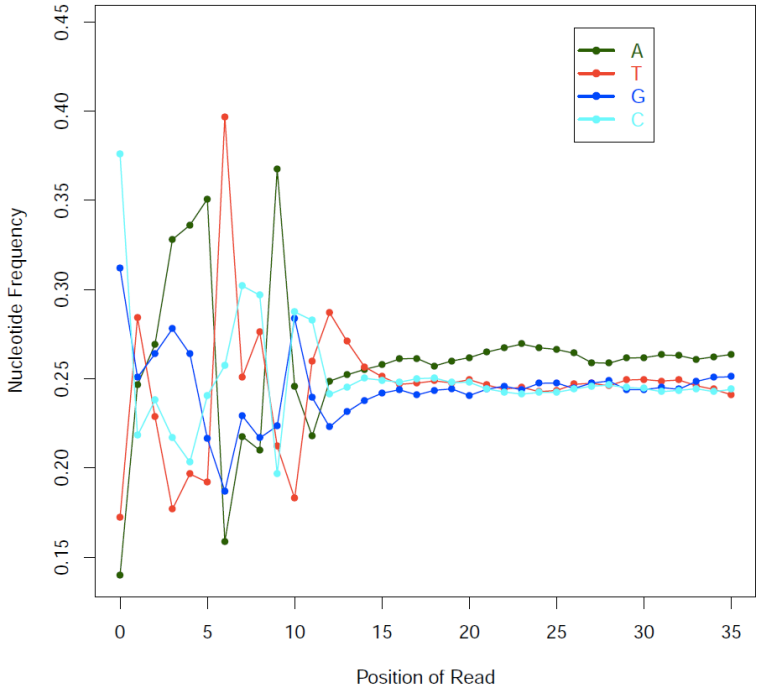
2-iv. Post-alignment QC

- Use 'samtools view' to see the format of a SAM/BAM alignment file
  - Use ‘FLAGs’ to filter out certain kinds of alignments
- Use 'samtools flagstat' to get a basic summary of an alignment
- Use FastQC to perform basic QC of your alignments

- Optional: explore RSeQC for alignment QC

2-iv. Post-alignment QC (RSeQC)





We are on a Coffee Break & Networking Session