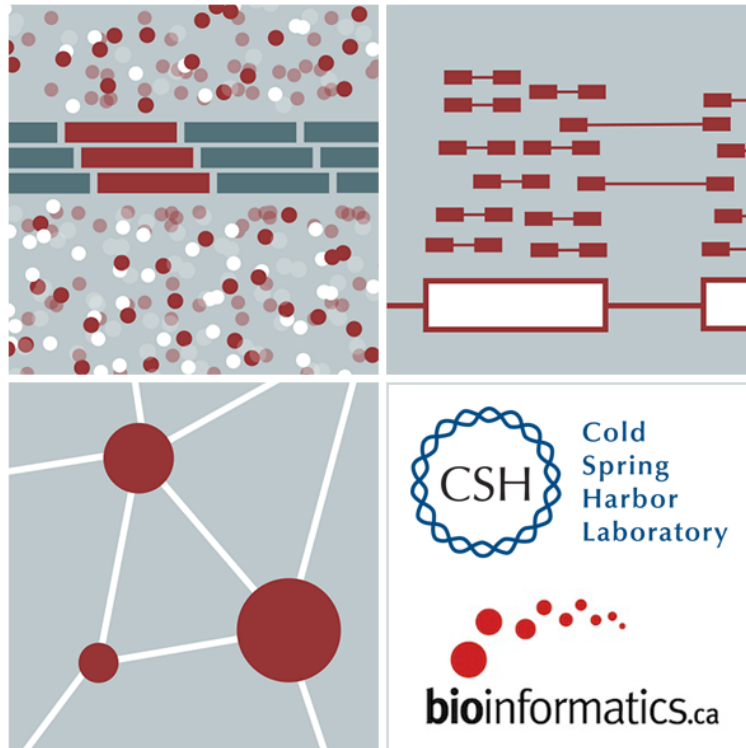


RNA-Seq Module 4

Differential Expression

Kelsy Cotto, Malachi Griffith, Obi Griffith, Megan Richters



Alternatives to FPKM

- Raw read counts for differential expression analysis
 - Assign reads/fragments to defined genes/transcripts, get “raw counts”
 - Transcript structures could still be defined by something like cufflinks

- HTSeq (htseq-count)

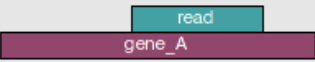
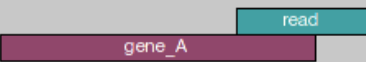

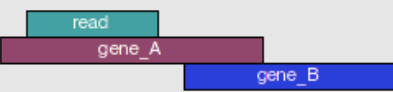
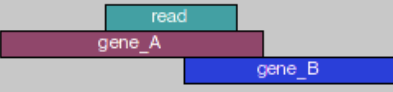
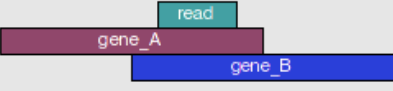
- <http://www-huber.embl.de/users/anders/HTSeq/doc/count.html>

```
htseq-count --mode intersection-strict --stranded no --minqual 1 --type  
exon --idattr transcript_id accepted_hits.sam chr22.gff >  
transcript_read_counts_table.tsv
```

- Caveats of ‘transcript’ analysis by htseq-count:

- Designed for genes - ambiguous reads from overlapping transcripts may not be handled!
- <http://seqanswers.com/forums/showthread.php?t=18068>

HTSeq-count basically counts reads supporting a feature (exon, gene) by assessing overlapping coordinates

	union	intersection_strict	intersection_nonempty
	gene_A	gene_A	gene_A
	gene_A	no_feature	gene_A
	gene_A	no_feature	gene_A
	gene_A	gene_A	gene_A
	gene_A	gene_A	gene_A
	ambiguous	gene_A	gene_A
	ambiguous	ambiguous	ambiguous

Whether a read is counted depends on the nature of overlap and “mode” selected