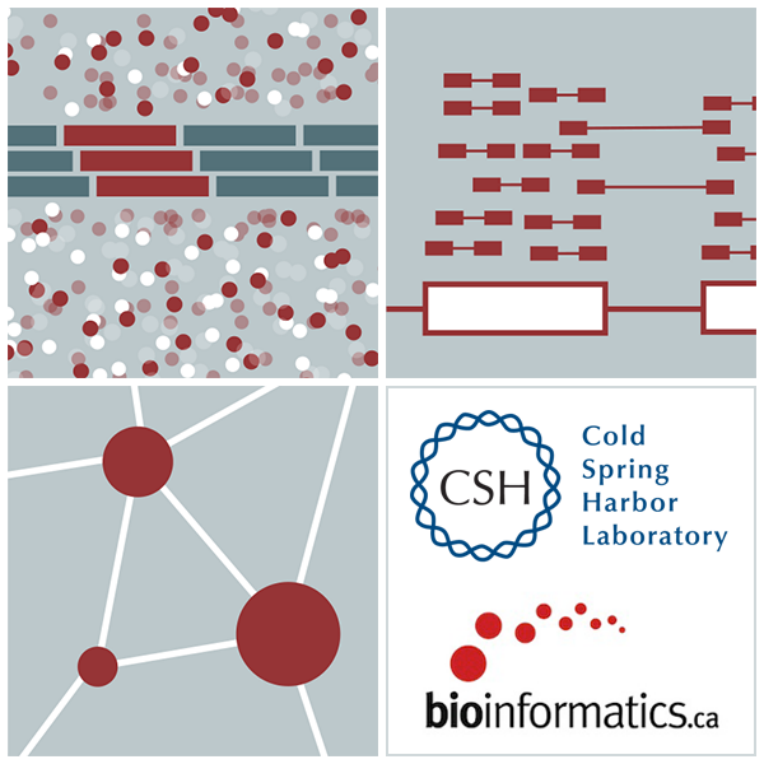
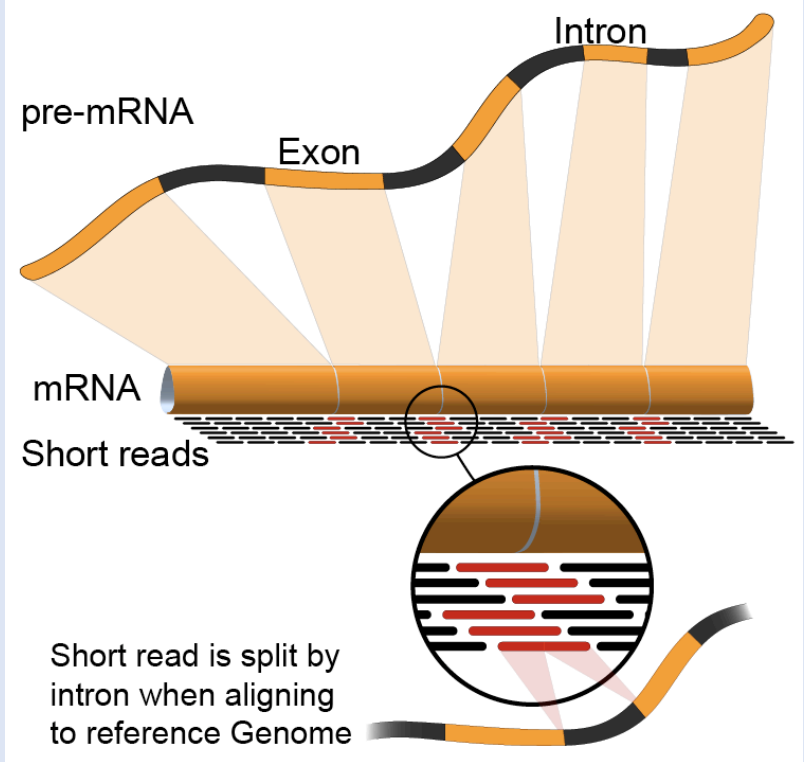
RNA-Seq Module 3
HTSeq

Kelsy Cotto, Felicia Gomez,
Obi Griffith, Malachi Griffith, Huiming Xia

Advanced Sequencing Technologies & Applications

November 5- 16, 2019





THE GENOME INSTITUTE
at Washington University

Alternatives to FPKM

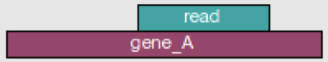
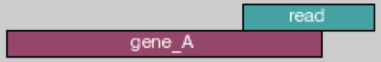
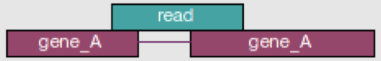
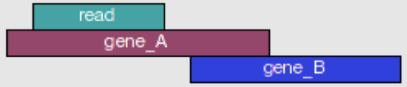
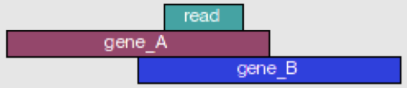
- Raw read counts for differential expression analysis
  - Assign reads/fragments to defined genes/transcripts, get “raw counts”
    - Transcript structures could still be defined by something like cufflinks

- HTSeq (htseq-count)
  - http://www-huber.embl.de/users/anders/HTSeq/doc/count.html

```
htseq-count --mode intersection-strict --stranded no --minaqual 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