bioinformatics.ca

Canadian Bioinformatics Workshops

www.bioinformatics.ca

bioinformaticsdotca.github.io

Supported by

McGill UNIVERSITY


Creative Commons
This page is available in the following languages:
Afrikaans български Català Dansk Deutsch Ελληνικά English English (CA) English (GB) English (US) Esperanto
Castellano Castellano (AR) Español (CL) Castellano (CO) Español (Ecuador) Castellano (MX) Castellano (PE)
Euskara Suomeksi français français (CA) Galego עברית hrvatski Magyar Italiano 日本語 한국어 Macedonian Melayu
Nederlands Norsk Sesotho sa Leboa polski Português română slovenski jezik српски srpski (latinica) Sotho svenska
中文 華語 (台灣) isiZulu
creative commons
Attribution-Share Alike 2.5 Canada
You are free:
to Share — to copy, distribute and transmit the work
to Remix — to adapt the work
Free Cultural
APPROVED FOR
Works
Under the following conditions:
Attribution. You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).
Share Alike. If you alter, transform, or build upon this work, you may distribute the resulting work only under the same or similar licence to this one.
For any reuse or distribution, you must make clear to others the licence terms of this work.
Any of the above conditions can be waived if you get permission from the copyright holder.
The author's moral rights are retained in this licence.
Disclaimer
Your fair dealing and other rights are in no way affected by the above.
This is a human-readable summary of the Legal Code (the full licence) available in the following languages:
English French
Learn how to distribute your work using this licence

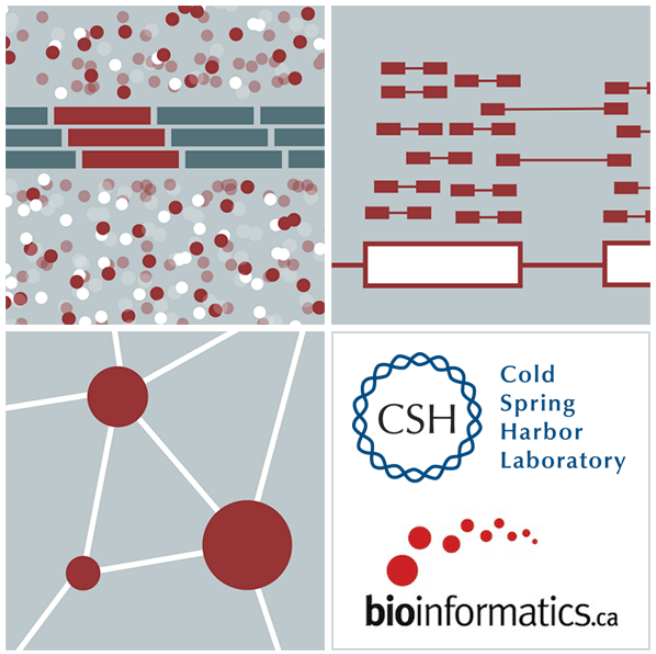
HT-Seq

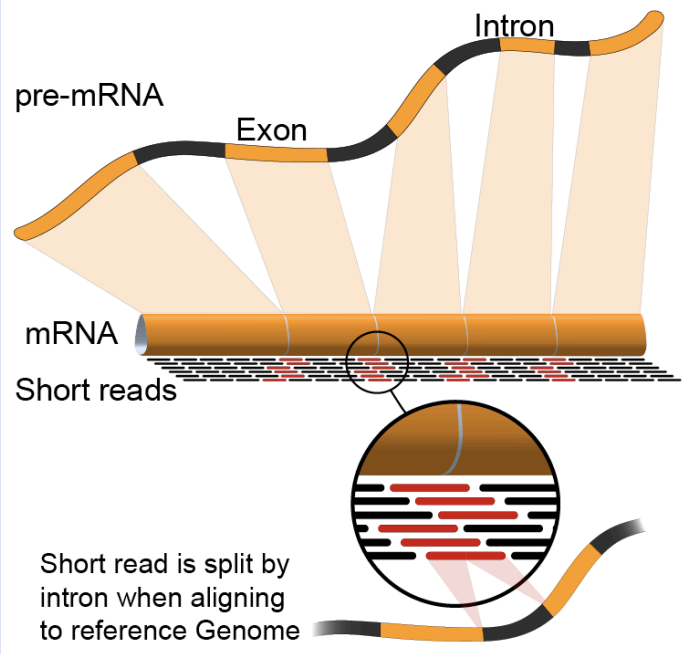
Emma Bell, Felicia Gomez, Obi Griffith, Malachi Griffith, Huiming Xia

RNA-Seq Analysis

Sep 8th-10th, 2021

bioinformatics.ca





Washington University in St. Louis
School of Medicine

Alternatives to FPKM

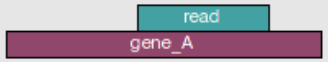
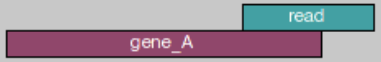
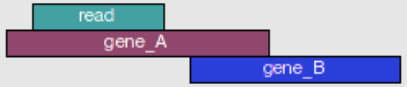
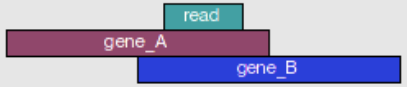
- Raw read counts for differential expression analysis
  - Assign reads/fragments to defined genes/transcripts, get “raw counts”
    - Transcript structures could still be defined by something like cufflinks

- 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

We are on a Coffee Break & Networking Session

Workshop Sponsors:

Canadian Centre for Computational Genomics

neuro

Institut-Hôpital neurologique de Montréal

Montreal Neurological Institute-Hospital

McGill initiative in Computational Medicine

HPC4Health