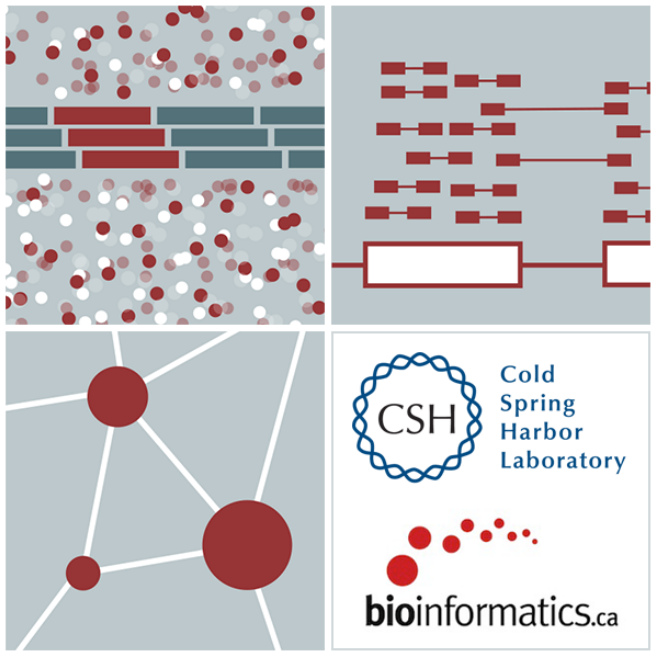
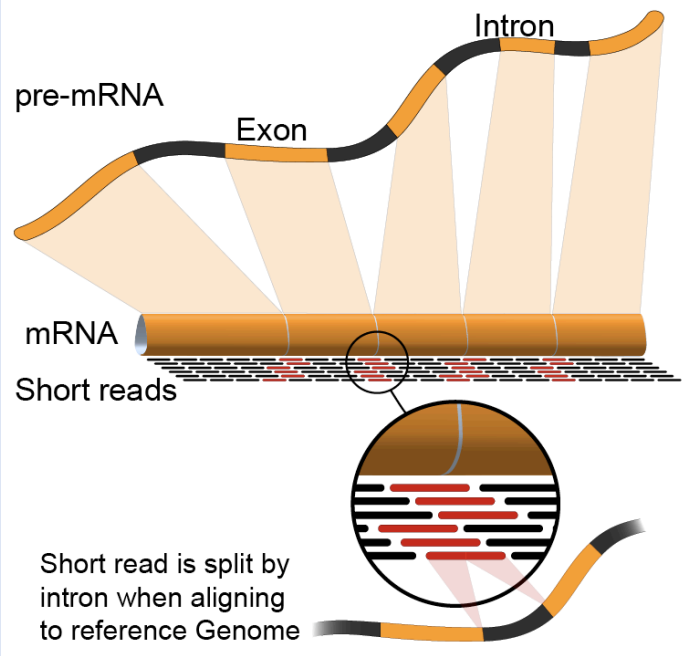
RNA-Seq Module 2
Alignment vs Assembly vs Pseudoalignment

Kelsy Cotto, Arpad Danos, Harriet Dashnow, Felicia Gomez, Sharon Freshour, Obi Griffith, Malachi Griffith, Jason Kunisaki, Chris Miller, Jonathan Preall, Aaron Quinlan

Advanced Sequencing Technologies & Bioinformatics Analysis November 11-19, 2021

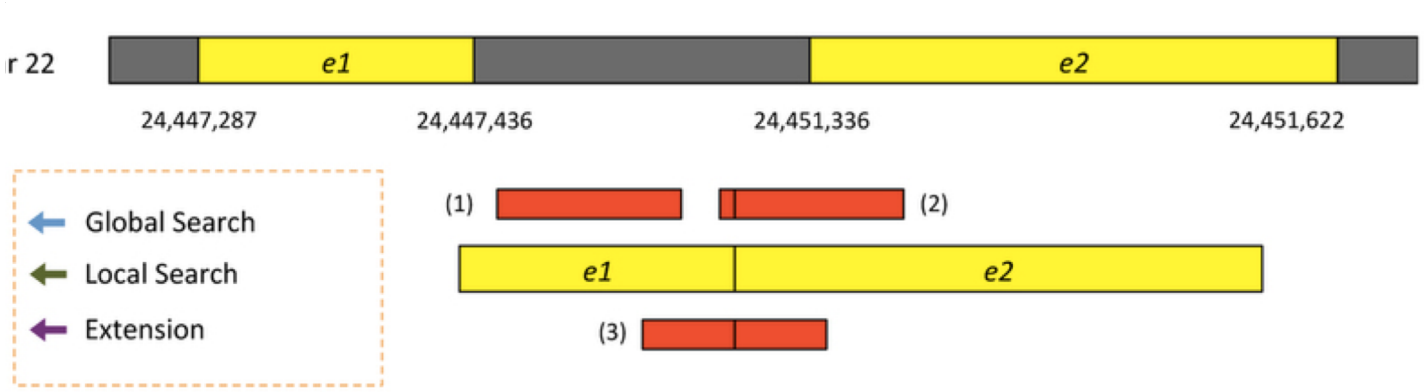




Washington University in St. Louis
School of Medicine

Alignment

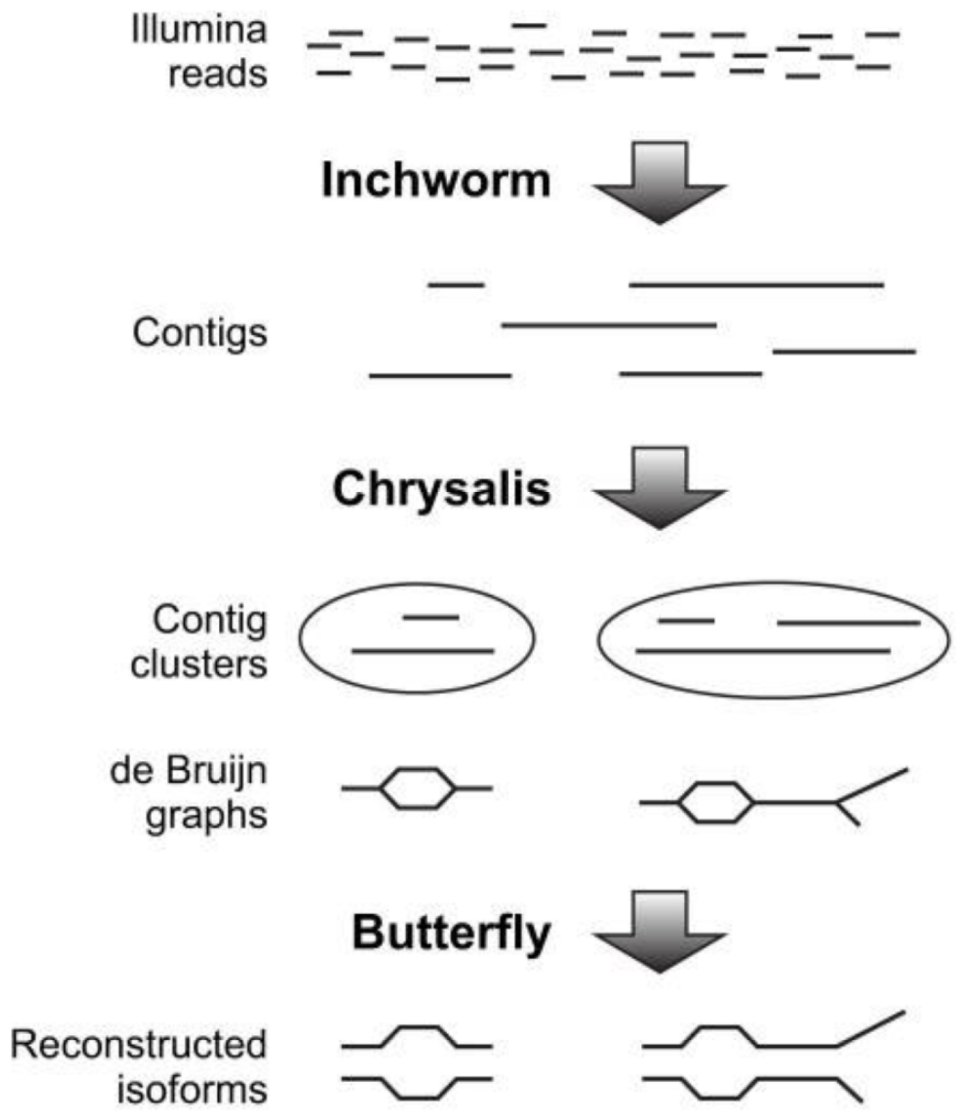
- Uses a reference genome/transcriptome to map reads
- Capable of some novel transcript inference
- Relatively fast runtime
- Tools: HISAT2, STAR, GSNAP



Kim et al. 2015. Nat Methods 12:357–360

Assembly

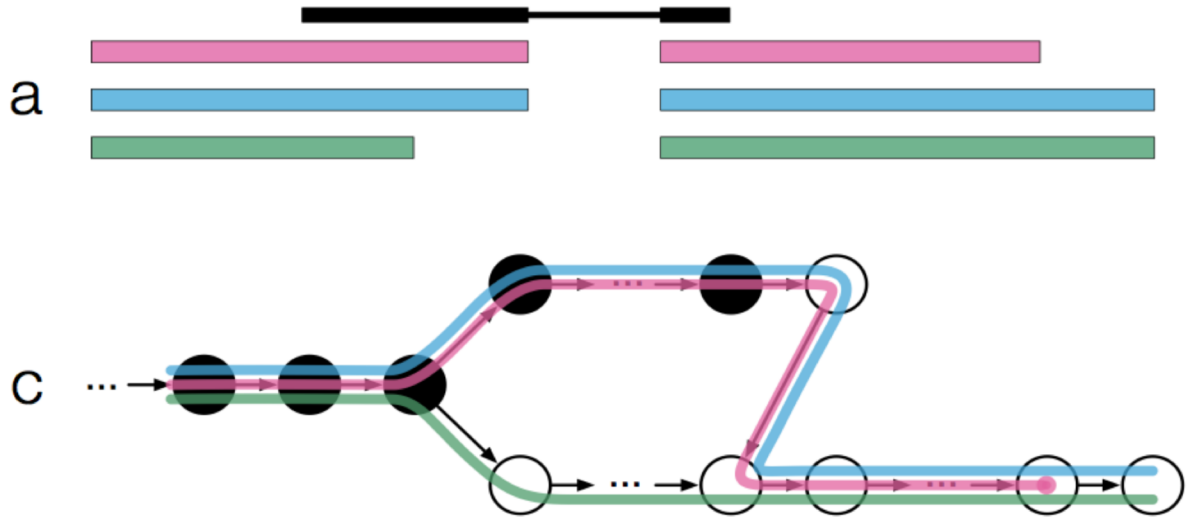
- Infer transcript structure directly from the data
- Useful when you do not have a reference sequence
- Other uses – highly rearranged genomes (some cancers)
- Computationally expensive
- Tools: Trinity, Velvet, SPAdes



Haas, et al (2013) doi: 10.1038/nprot.2013.084

Pseudoalignment

- Does not determine where in the genome a read lies, only which transcripts it is compatible with
- Very fast!
- Does not produce a bam by default (though pseudo-bams can be created), not useful for variant detection.
- Tools: Kallisto, Sailfish



Bray, 2016 doi:10.1038/nbt.3519

https://tinyheero.github.io/2015/09/02/pseudoalignments-kallisto.html

We are on a Coffee Break & Networking Session