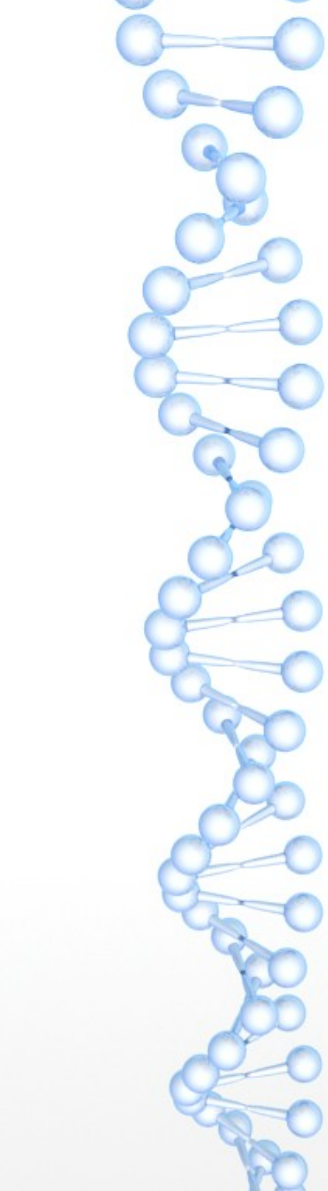


RNA-Seq Analysis of Gene Expression: A Walk-Thru and Tutorial

Helen Nigussie, Michael Mayhew, Dina Machuve

June 4, 2019

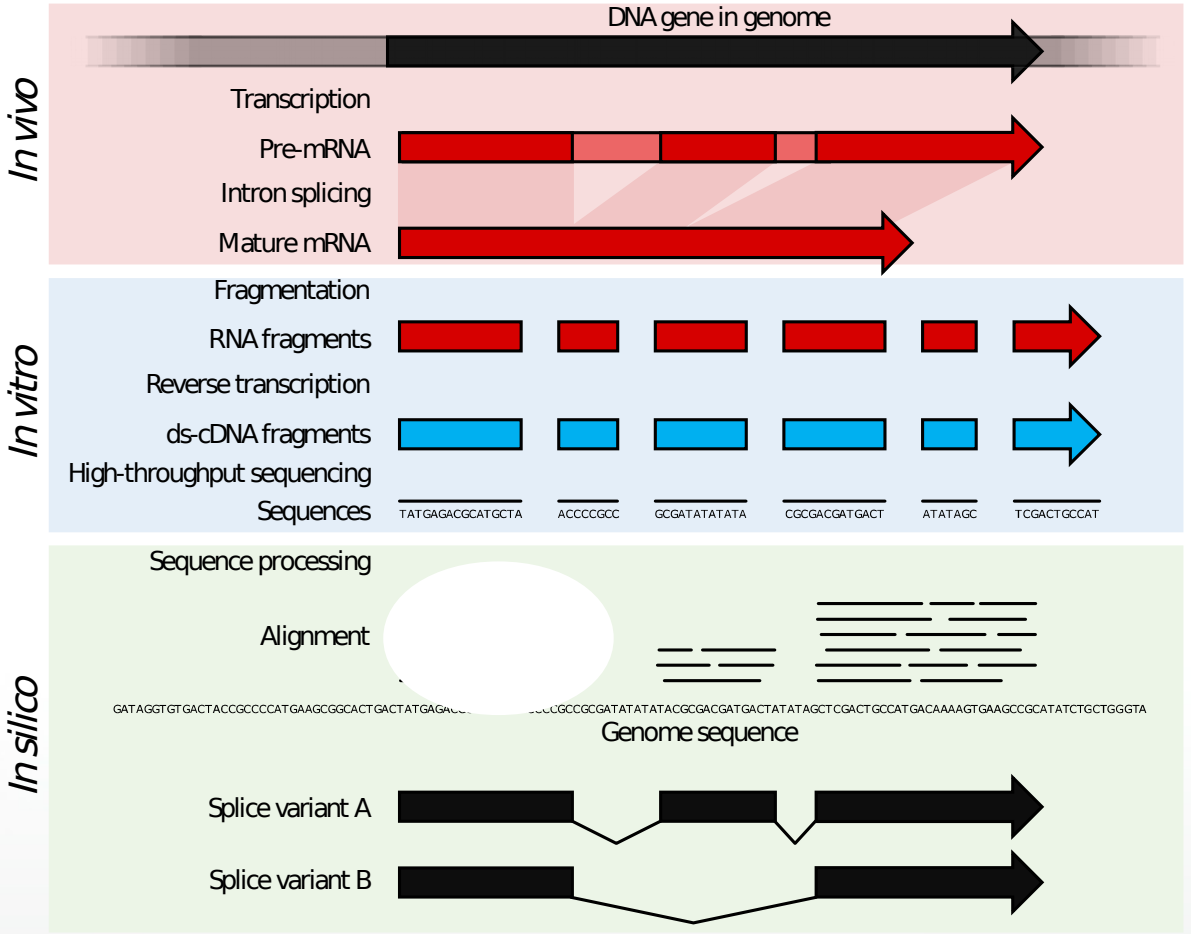
Data Science Africa 2019

Addis Ababa, Ethiopia

What is RNA-Seq analysis?

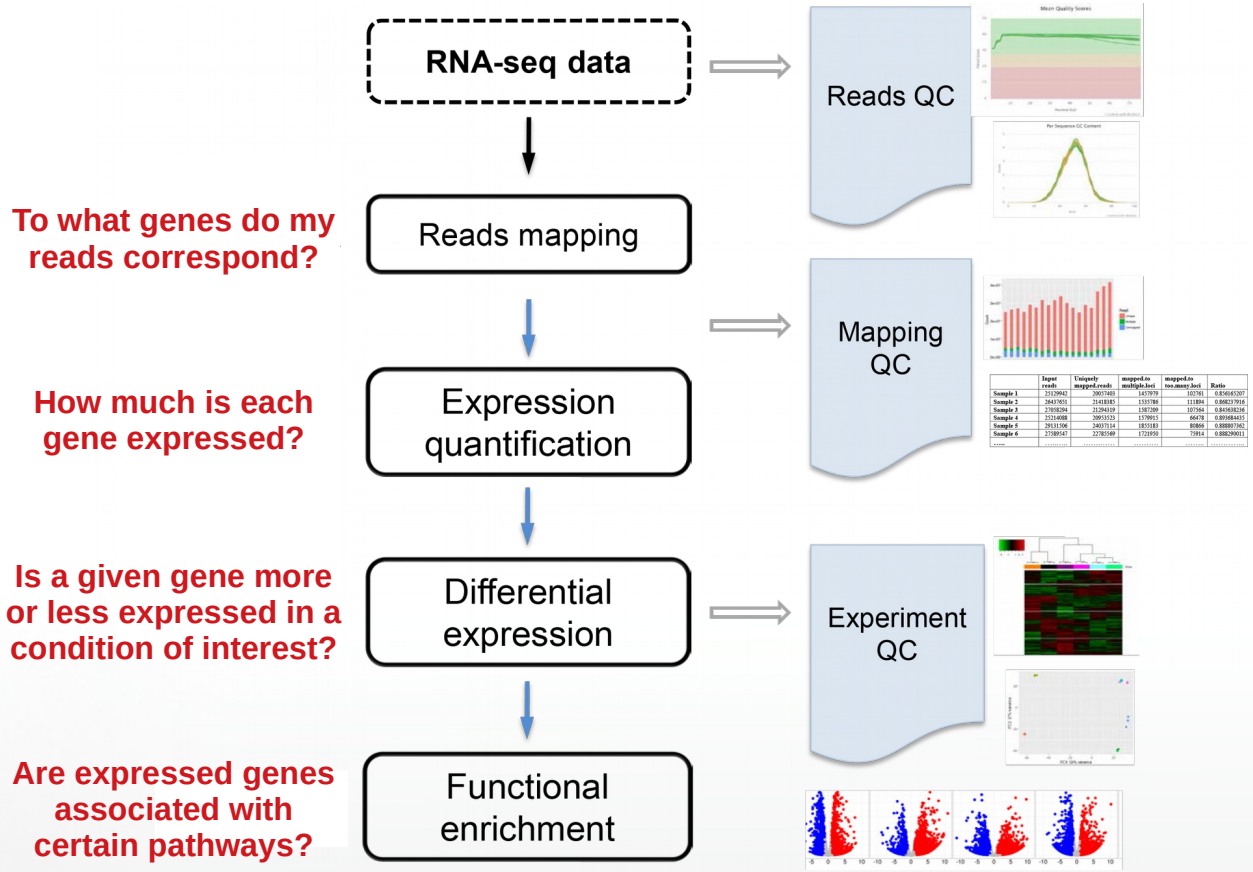
RNA sequencing *(*RNA-Seq for short) is a process of assessing the ***expression of genes*** across a genome by ***sequencing the RNA transcripts*** from a collection of cells

What is RNA-Seq analysis?



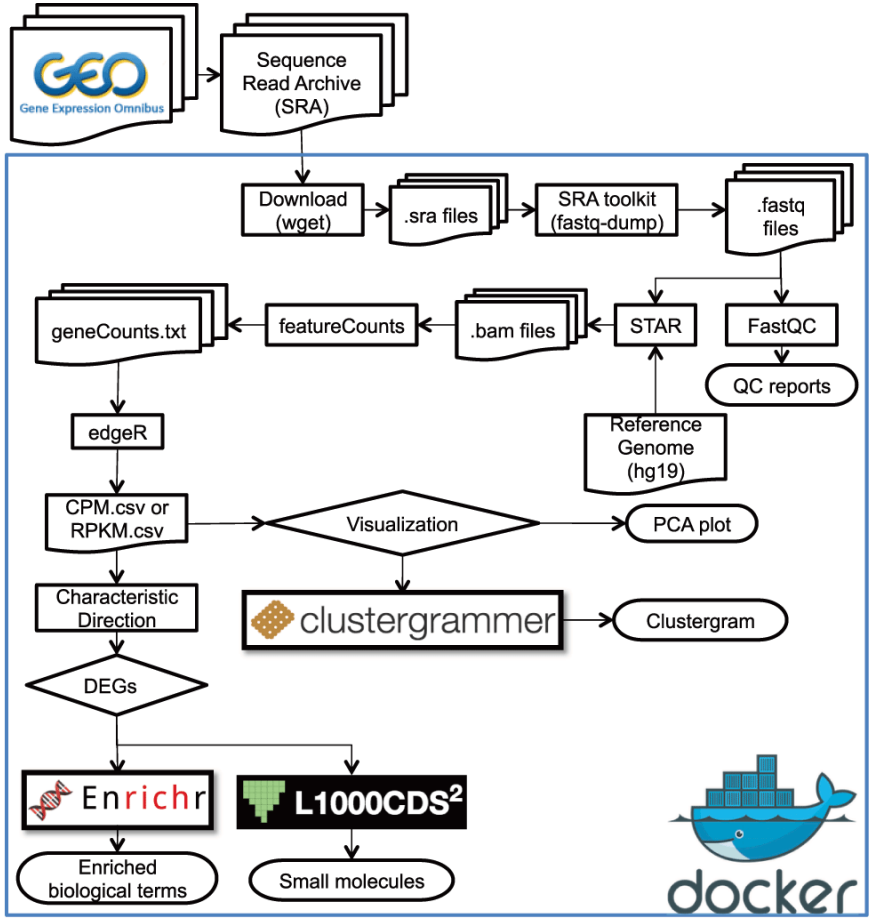
https://en.wikipedia.org/wiki/RNA-Seq

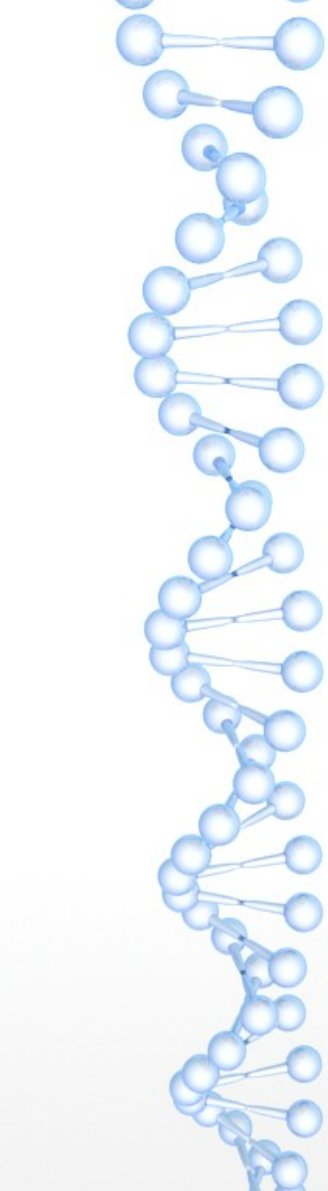
What are the different stages of RNA-Seq analysis?



http://bioinfo.vanderbilt.edu/vangard/services-rnaseq.html

What are the different stages of RNA-Seq analysis?





Stage 1: Processing and quality control of raw sequencing reads

Stage 1: Processing and quality control of raw sequencing reads (cont'd)

- Insert snapshot here of FastQC report

Stage 2: Alignment of sequencing reads to genome

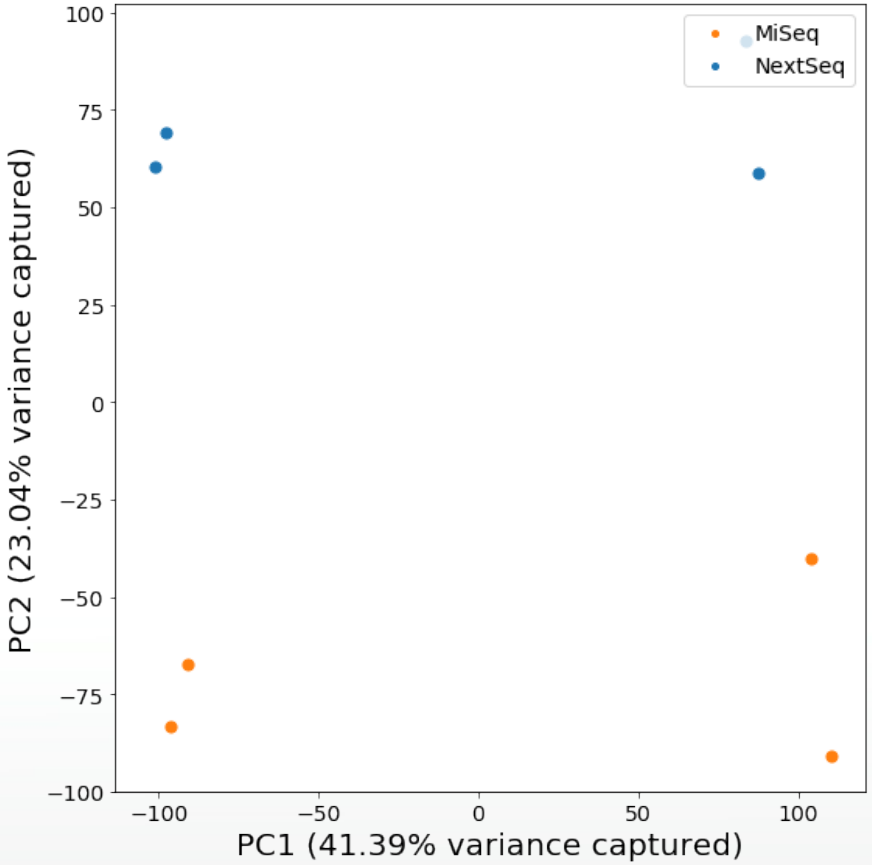
- Insert image of aligned reads to the genome

Stage 3: Mapping of reads to individual genes to attain expression measurements

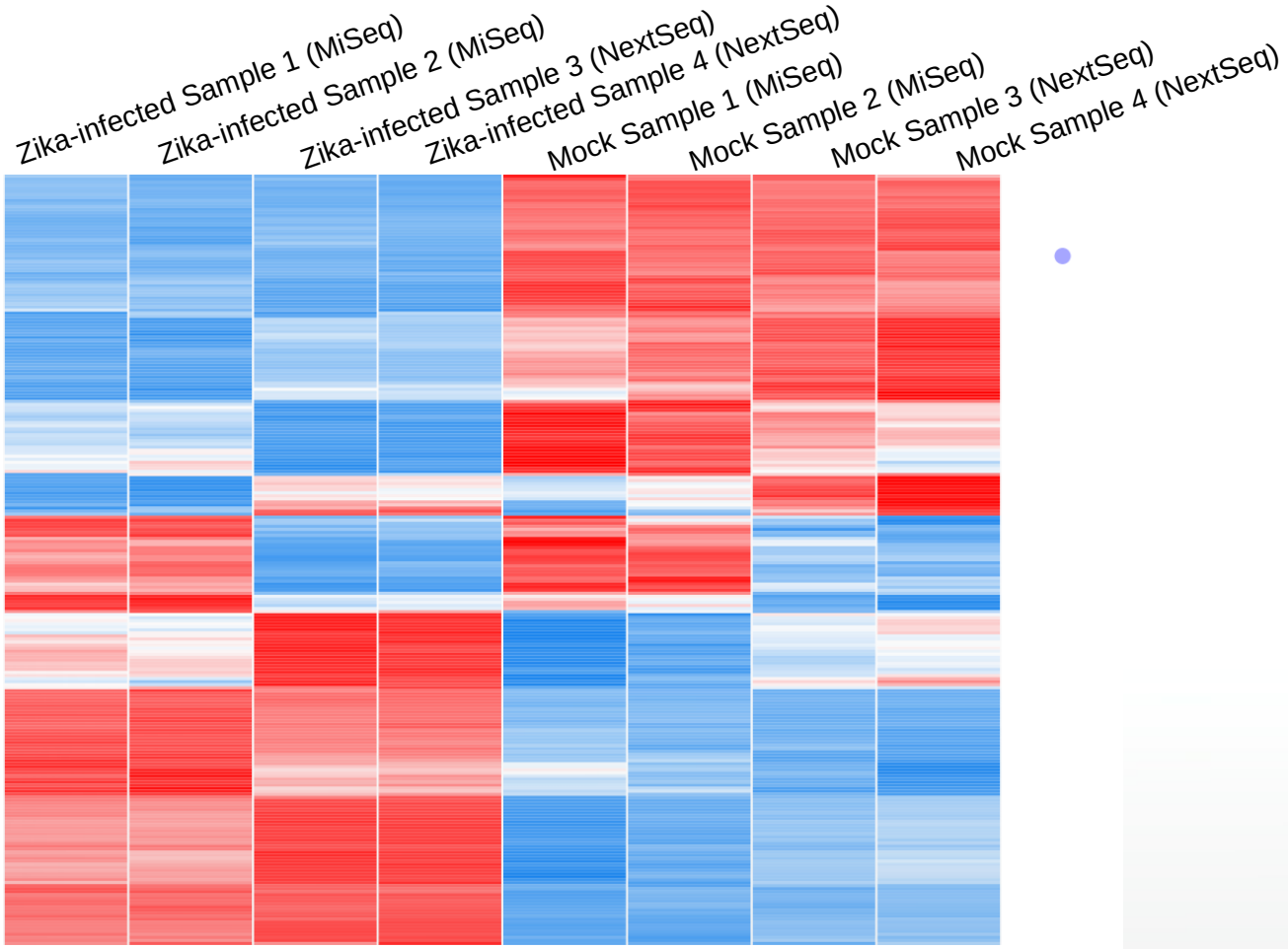
- Sequencing reads are aligned (‘mapped’) to a reference genome in which locations of genes are known
- Algorithms (like featureCounts) assign the aligned reads to each gene
  - Results in ‘digital’ measures of expression – one unit of expression per mapped read
- Counts are then normalized according to sequencing depth and/or gene length
  - Two common normalized expression measures are:
    - CPM – transcripts or counts per million
      - [(Read count)/(Gene length in kb)] / (
    - RPKM – reads per kilobase per million

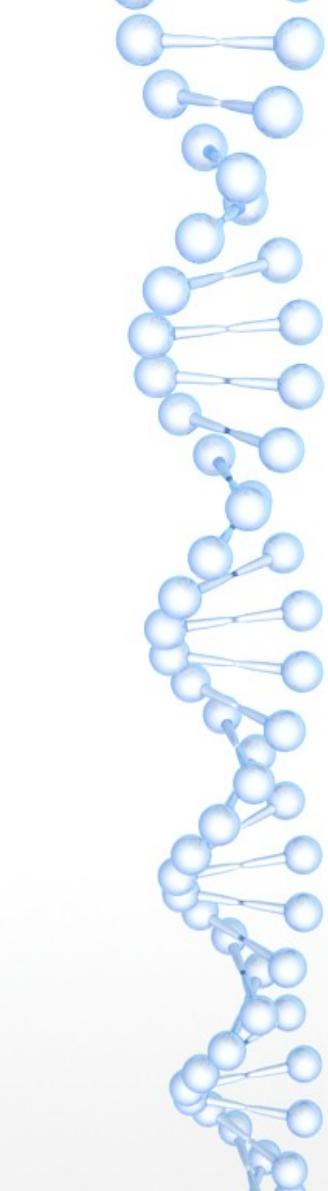
Important considerations when performing an RNA-Seq analysis

- Should I consider all genes in my analysis? What about those with low or no expression across all conditions/platforms?
- Are the expression differences I'm seeing solely due to the condition? Or some other factor?

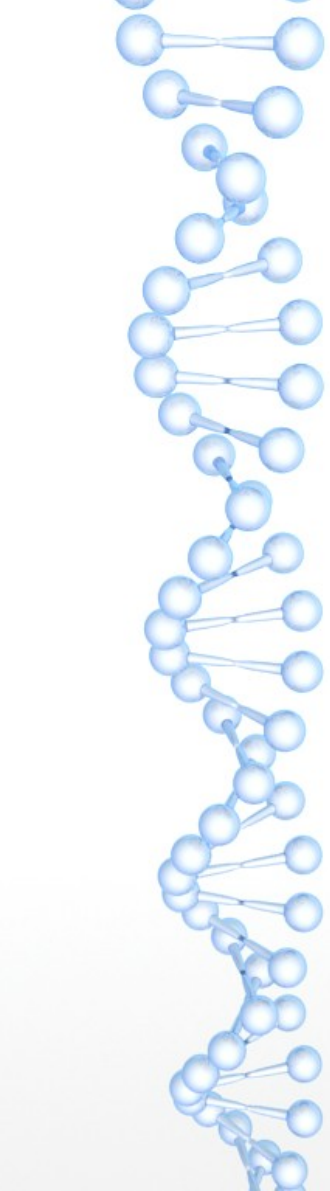


What is the structure in my expression data?





What genes show different expression patterns in my conditions of interest?



Are differentially expressed genes enriched for any biological processes or pharmacological targets?

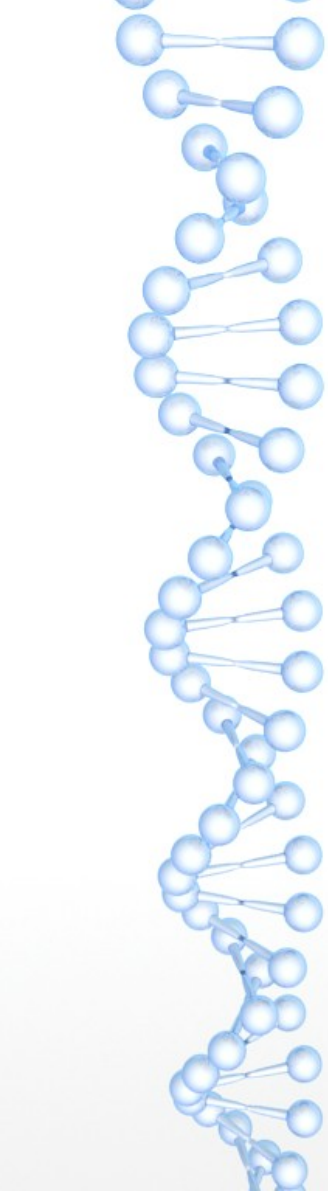
An unsolicited advertisement

Oral Presentation Submission Deadline: September 13, 2019
Poster Presentation Submission Deadline: October 15, 2019

https://www.iscb.org/iscbafrica2019

Additional resources

- Galaxy Community Hub's RNA-Seq Introduction: https://galaxyproject.org/tutorials/rb_rnaseq/
-



Thanks for your attention and see you at the workshop!

Any questions?