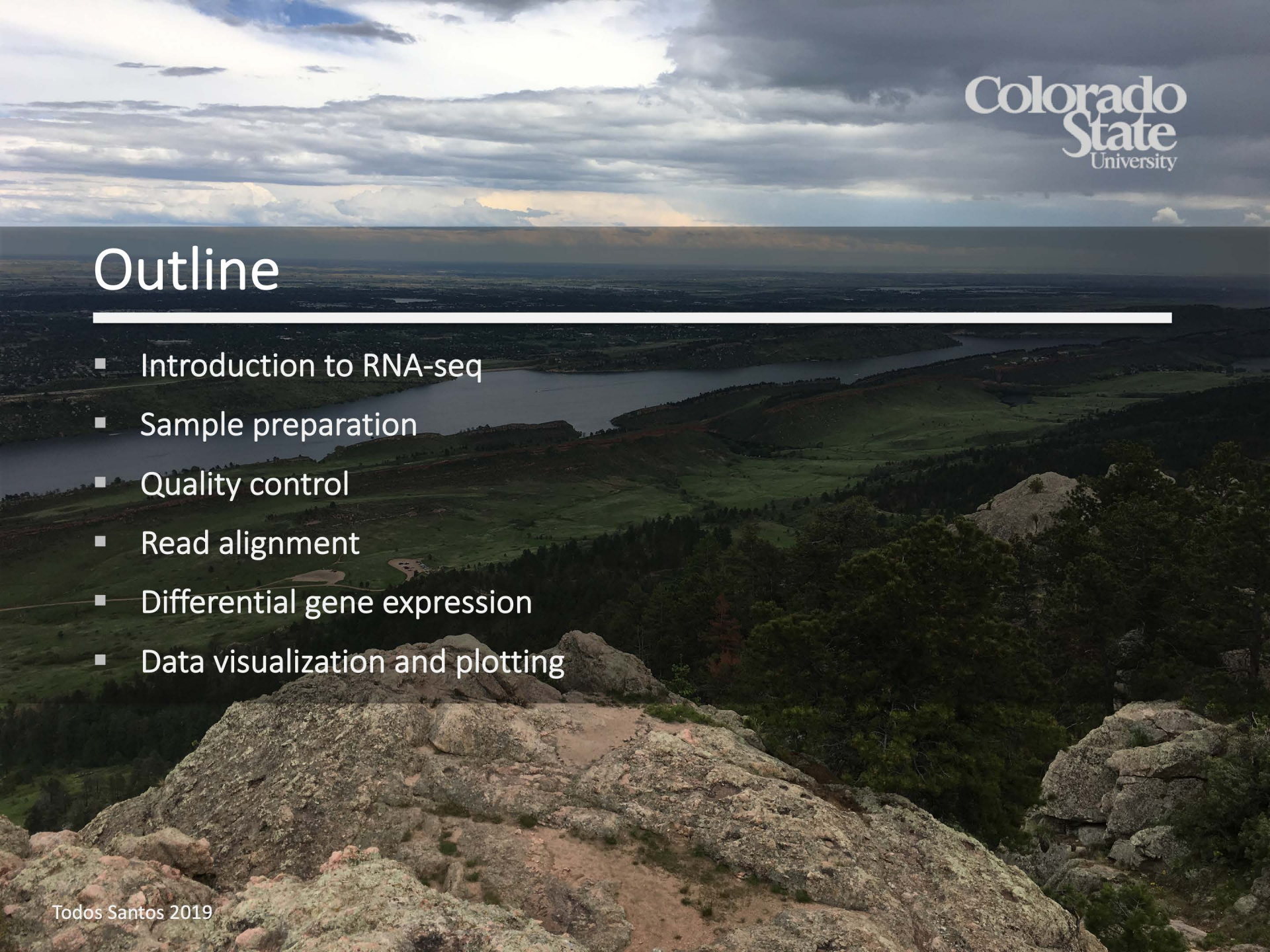


Outline

- Introduction to RNA-seq
- Sample preparation
- Quality control
- Read alignment
- Differential gene expression
- Data visualization and plotting

Regulation of gene expression

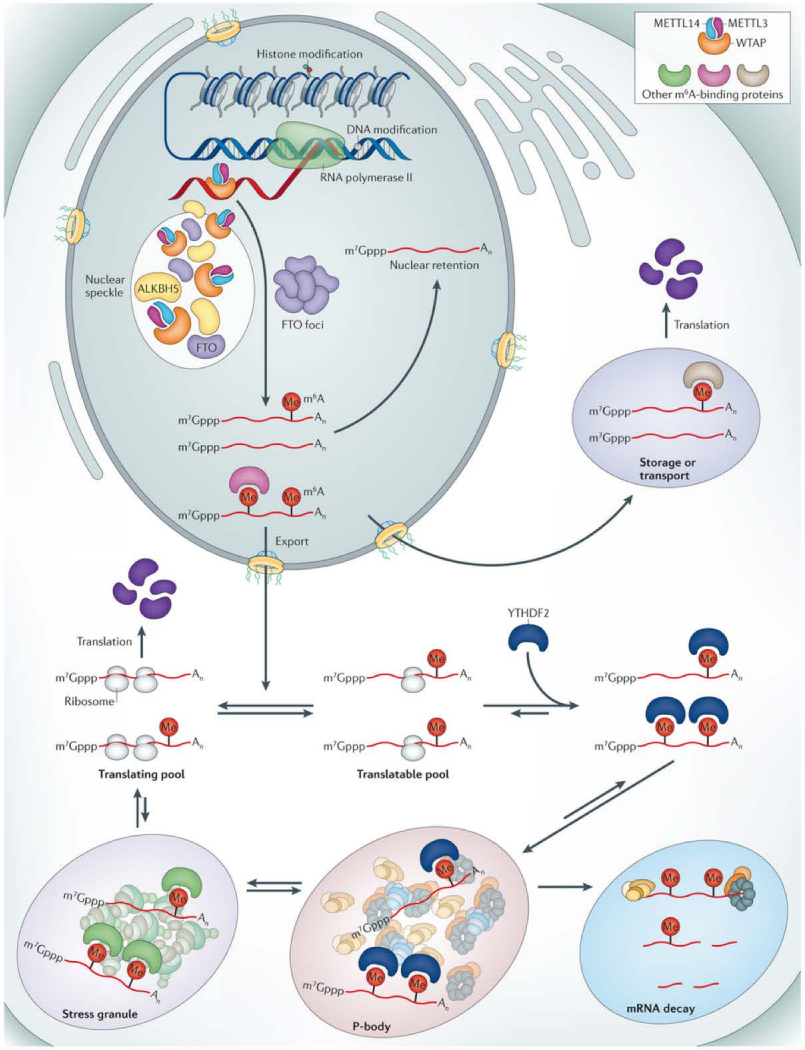
Regulation of transcription:
- Transcription factors
- Histone modifications
- DNA methylation

Regulation of RNA processing:
- Polyadenylation
- Splicing
- Capping
- RNA export

Regulation of translation:
- mRNA decay
- Translational repression
- Sequestration

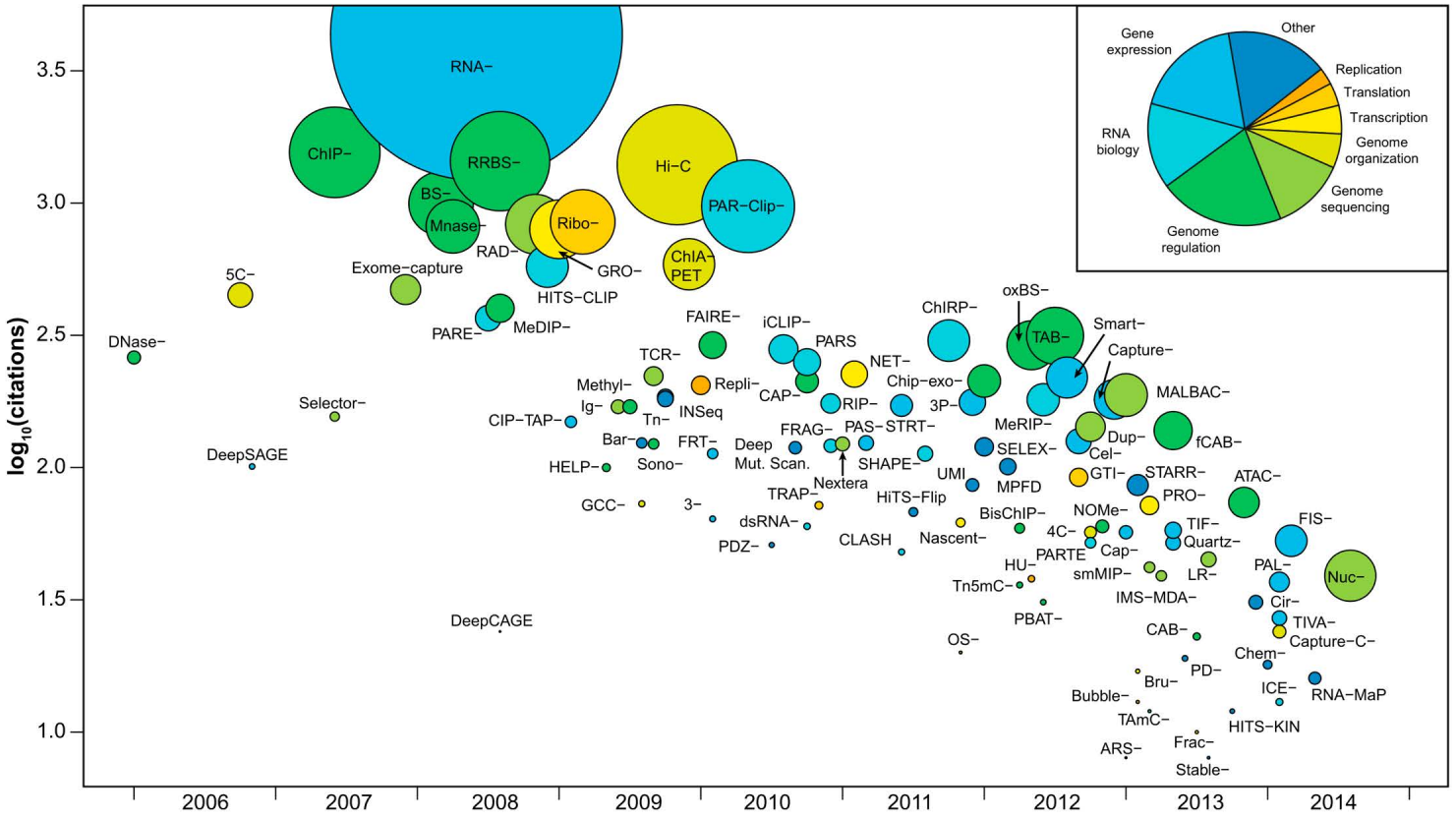
RNA-seq measures steady state mRNA levels and RNA sequence composition

Posttranslational regulation:
- Chemical modifications (e.g. phosphorylation)
- Protein turnover (proteolysis)



Fu et al. (2014)

RNA-seq is the most common HTS application



Reuter et al 2015

Sample preparation

Colorado State University

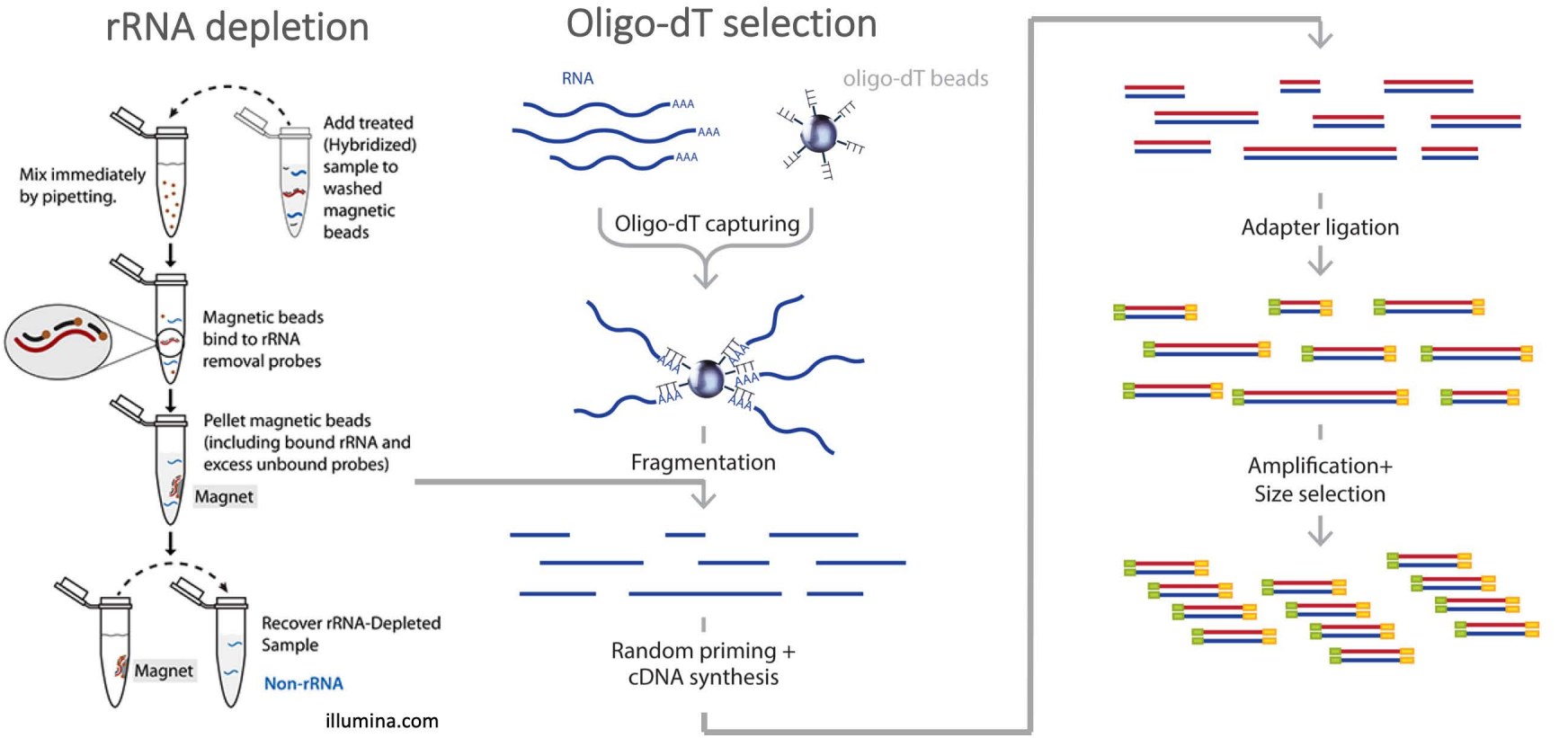
- Use high-quality RNA as starting material.
- Minor differences between samples can have a substantial impact on gene expression.
- 3-4 biological replicates is the default but not ideal for every situation.
- Some recommended kits for standard RNA-seq:
  - NEBNext Ultra II Directional RNA Library Prep Ki
  - Illumina kits

Sample preparation

Colorado State University

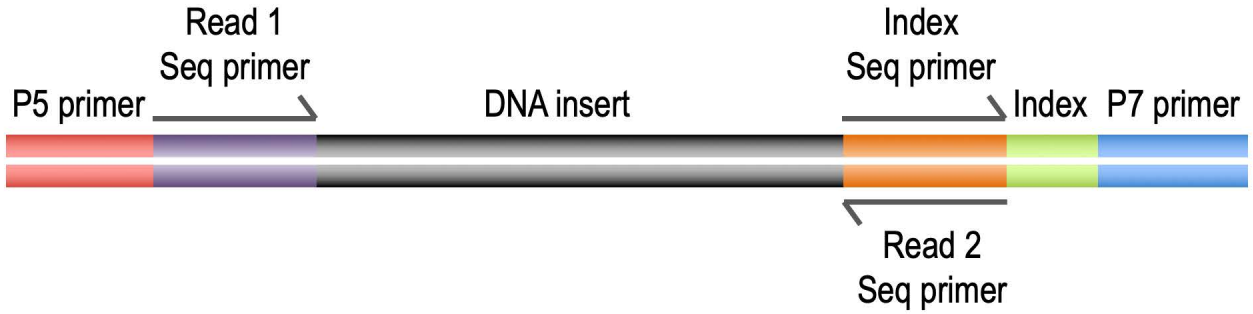
- Starting RNA
  - Typically 1-5 ug of high-quality total RNA is ideal.
- Sequencing depth
  - Typically you want 20-30 million high quality reads/library.
- Considerations
  - Strand specific (default is yes)
  - Single-end or paired-end (single-end is typically sufficient)
  - Long reads vs short reads (short Illumina reads, 50-150 nt, are usually sufficient)
  - rRNA depletion or oligo-dT
  - Low quantity/single cell

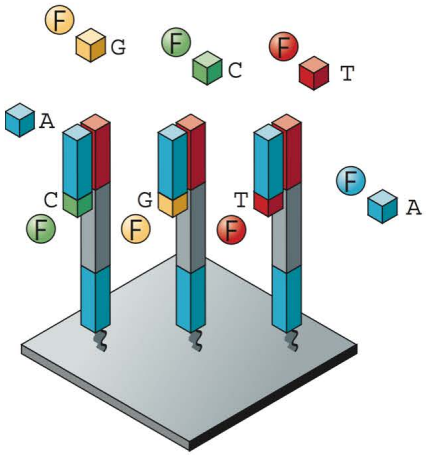
RNA-seq library preparation



Zhernakova et al. (2009)

Library composition





Metzker, M.L. (2010) NRG

HiSeq 2500

FASTQ format

Index sequence

Read 1

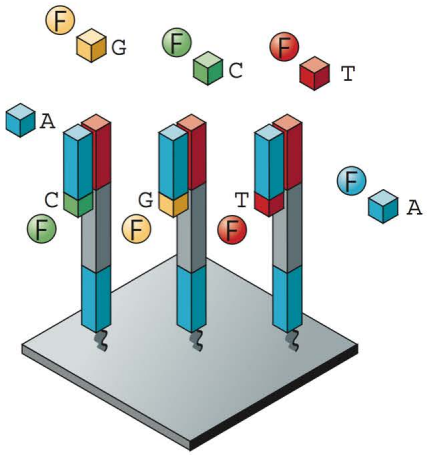
```
1 @D64TDFP1:248:C50DMACXX:5:1101:1241:2095 1:N:0:ATCACG
2 CACCGCCCGTCGCTATCCGGGACTGGAATTCTCGGGTGCCAAGGAACTCCA
3 +
4 CCCFFFFFHHHHHJIJGHJJJJIJJJJJGGGFFFFEABDHHHFHFF@@DD>
```

Read 2

```
1 @D64TDFP1:248:C50DMACXX:5:1101:1371:2154 1:N:0:ATCACG
2 TCAATATTTGCATAGGGTATCTGGAATTCTCGGGTGCCAAGGAACTCCAGT
3 +
4 CCCFFFFFHHHHHJJJJGFHIJJJJJJJJJJJJJFHHIIJJHGHJFGHJJI
```

Read 3

```
1 @D64TDFP1:248:C50DMACXX:5:1101:1461:2205 1:N:0:ATCACG
2 GAAAGACGTCTTCCTAGATTATGGAATTCTCGGGTGCCAAGGAACTCCAGT
3 +
4 CCCFFFFFHHHHHJJJJJJJJJJJIJJJJJJJJJHIJJJJJGIIJFGIJJJ
```



Metzker, M.L. (2010) NRG

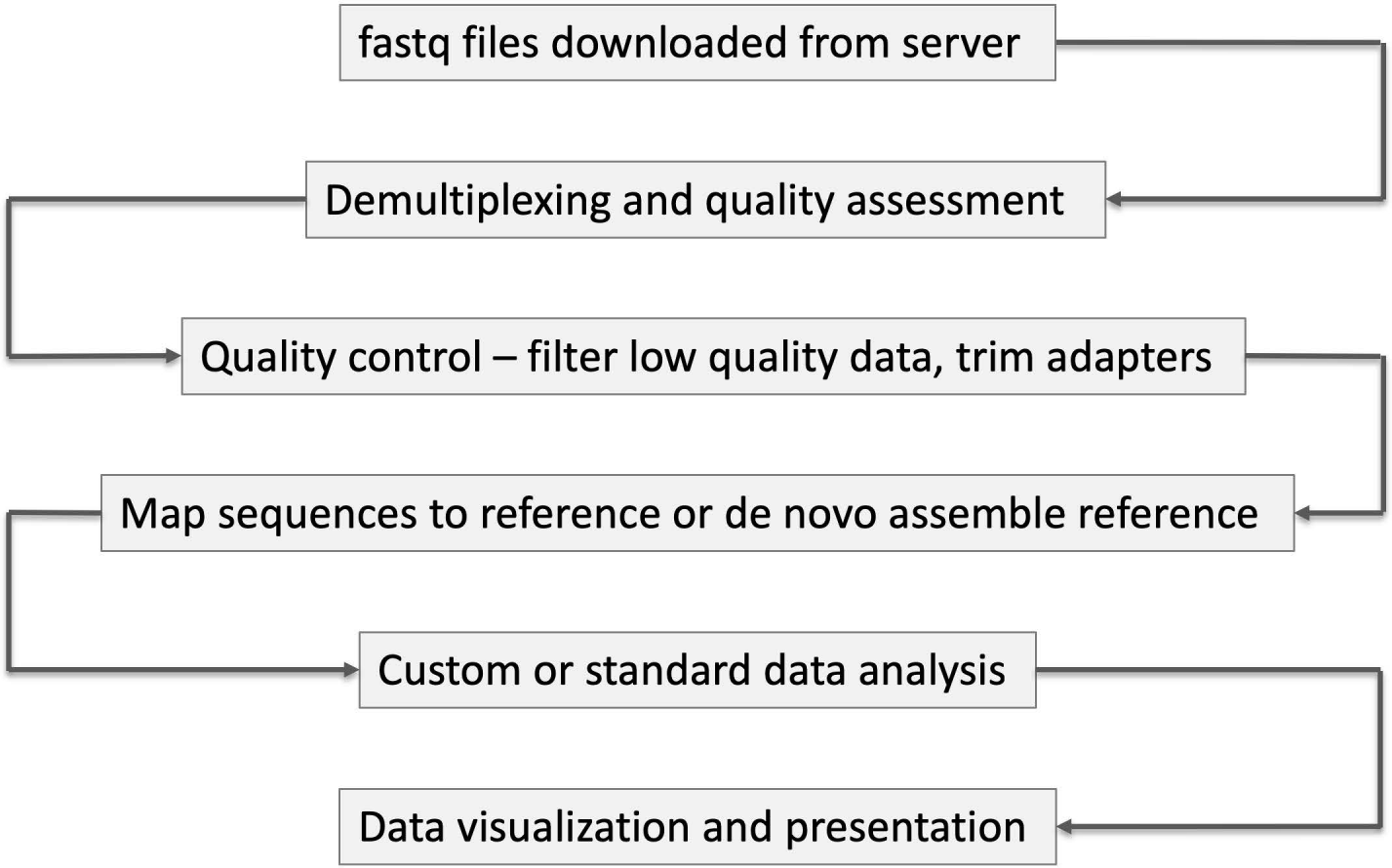
Line 1: sequence ID, description, and index; begins with @

Line 2: sequence; contains only A, C, T, G, and N

Line 3: optional sequence ID; begins with +

Line 4: signal quality of each base, cryptic code, phred 33 or 64

Data analysis workflow



Quality control

Colorado State University

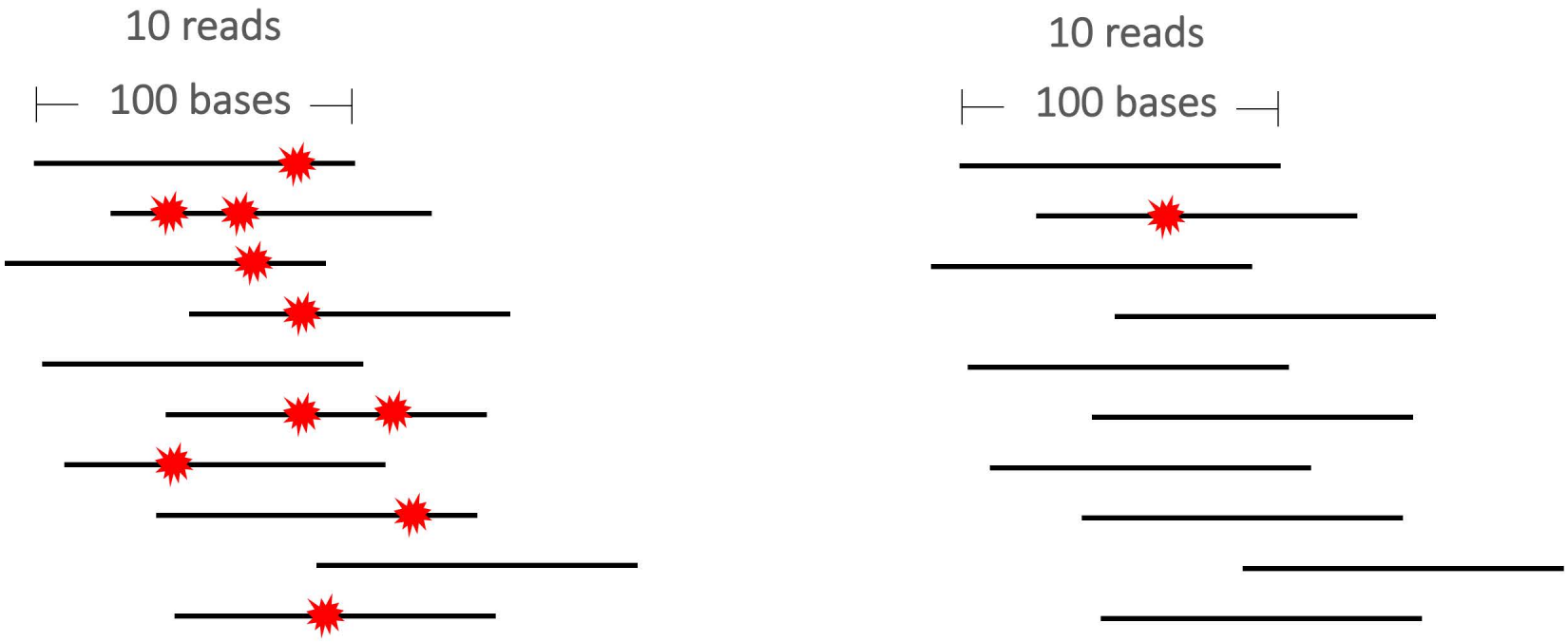
Assessing Read Quality

Phred quality score: a measure of the quality of base calling:

$Q = -10 \log(P)$ where P is the error probability

Phred Quality Score	Probability of incorrect base call	Base call accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1000	99.9%
40	1 in 10,000	99.99%
50	1 in 100,000	99.999%
60	1 in 1,000,000	99.9999%

Quality control



P = 0.01
Q = 20 (Q20)

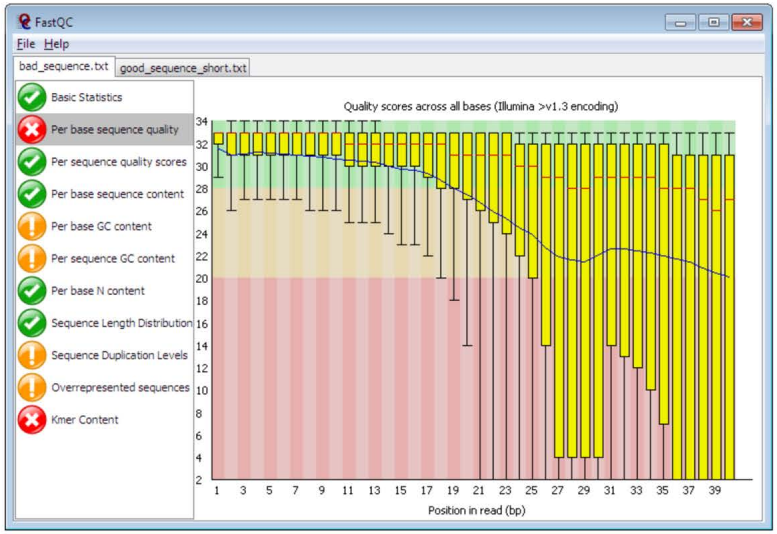
$Q = -10 \log(P)$

P = ?
Q = ?

Q30 is a common quality threshold or quality criterion

Quality control

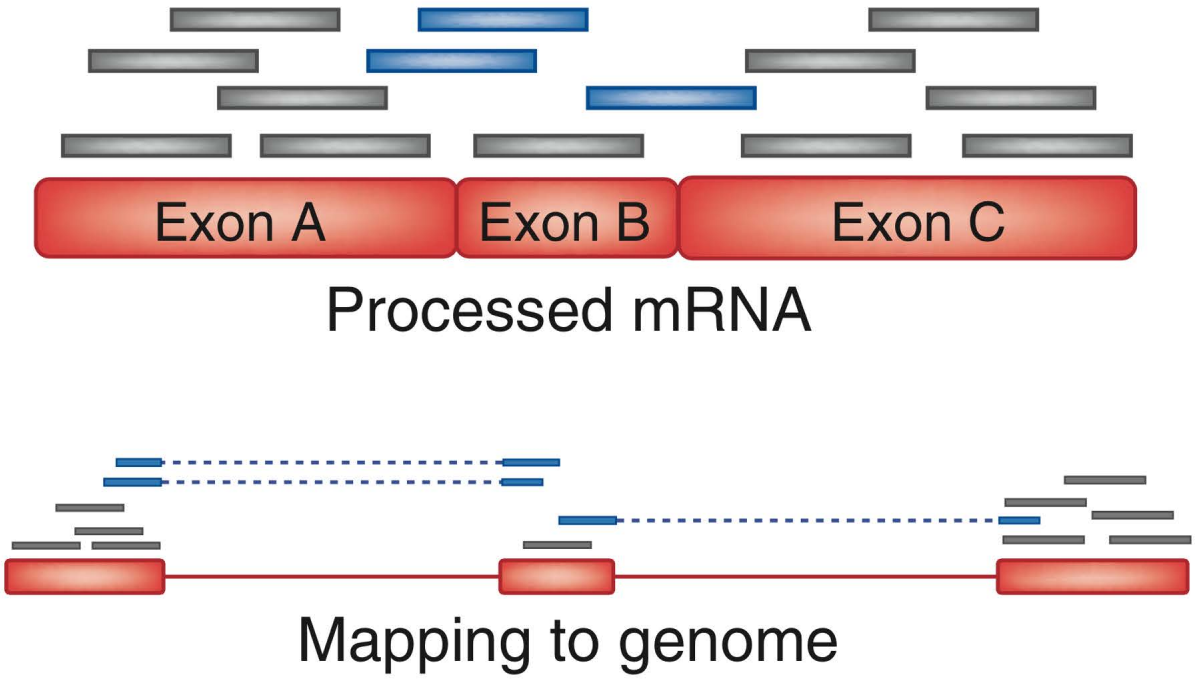
FastQC: a GUI tool for assessing the quality of high-throughput sequencing data.



Trimmomatic: software for trimming adapter sequences and low-quality bases from sequencing reads.

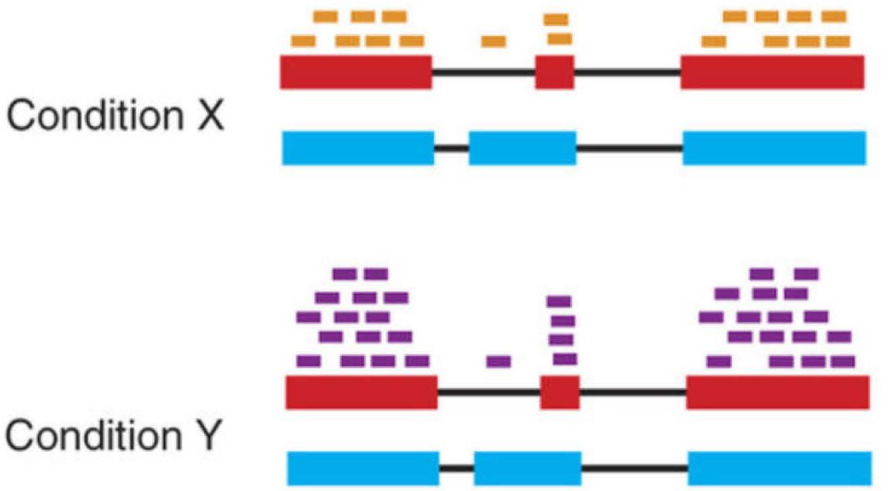
THE USADEL LAB

Aligning reads to mRNAs



Trapnell et al (2009)

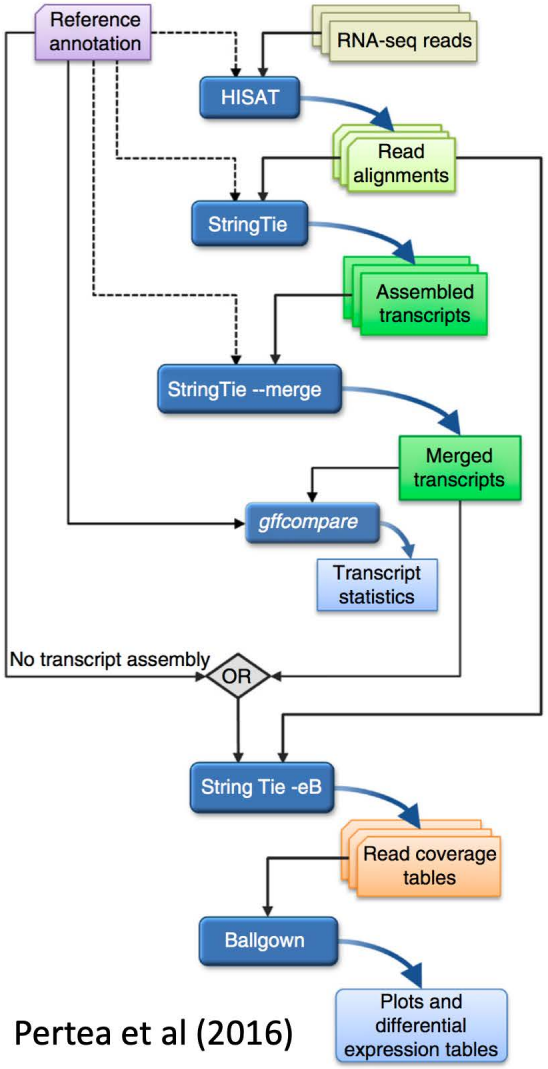
Differential gene expression



Trapnell et al (2010)

How many reads align to each gene in condition X vs condition Y?

RNA-seq pipelines



Pertea et al (2016)

No reference genome? Use Trinity to assemble transcripts

Other mRNA aligners: Star, GNSAP, Tophat2

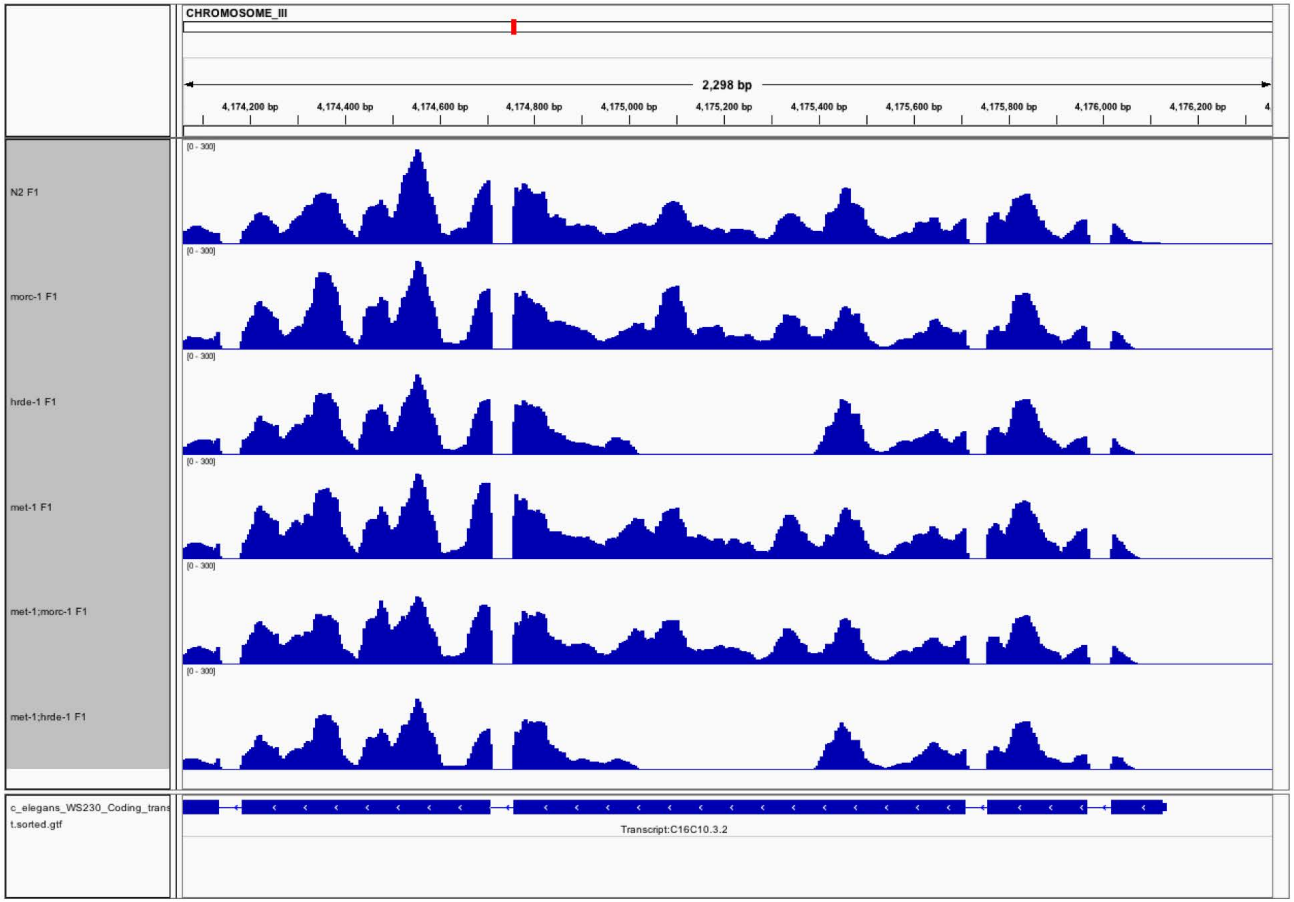
Other abundance estimators: RSEM, htseq-count, kallisto (alignment free esitimates), salmon(alignment free esitimates)

Other common DE software: DESeq2, edgeR, cuffdiff

Various GUIs and R-based tools for drawing plots

Genome browsers

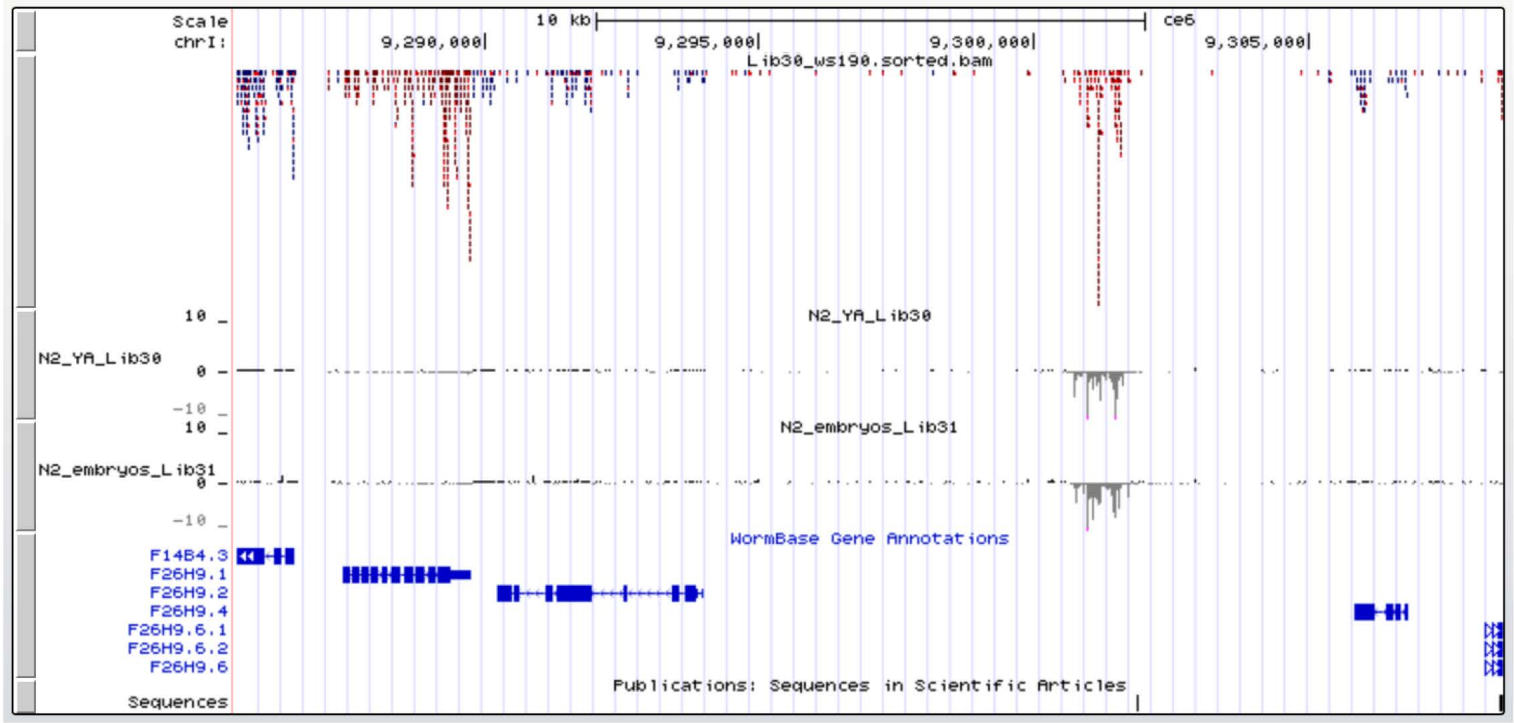
Integrative Genomics Viewer (IGV)



https://software.broadinstitute.org/software/igv/

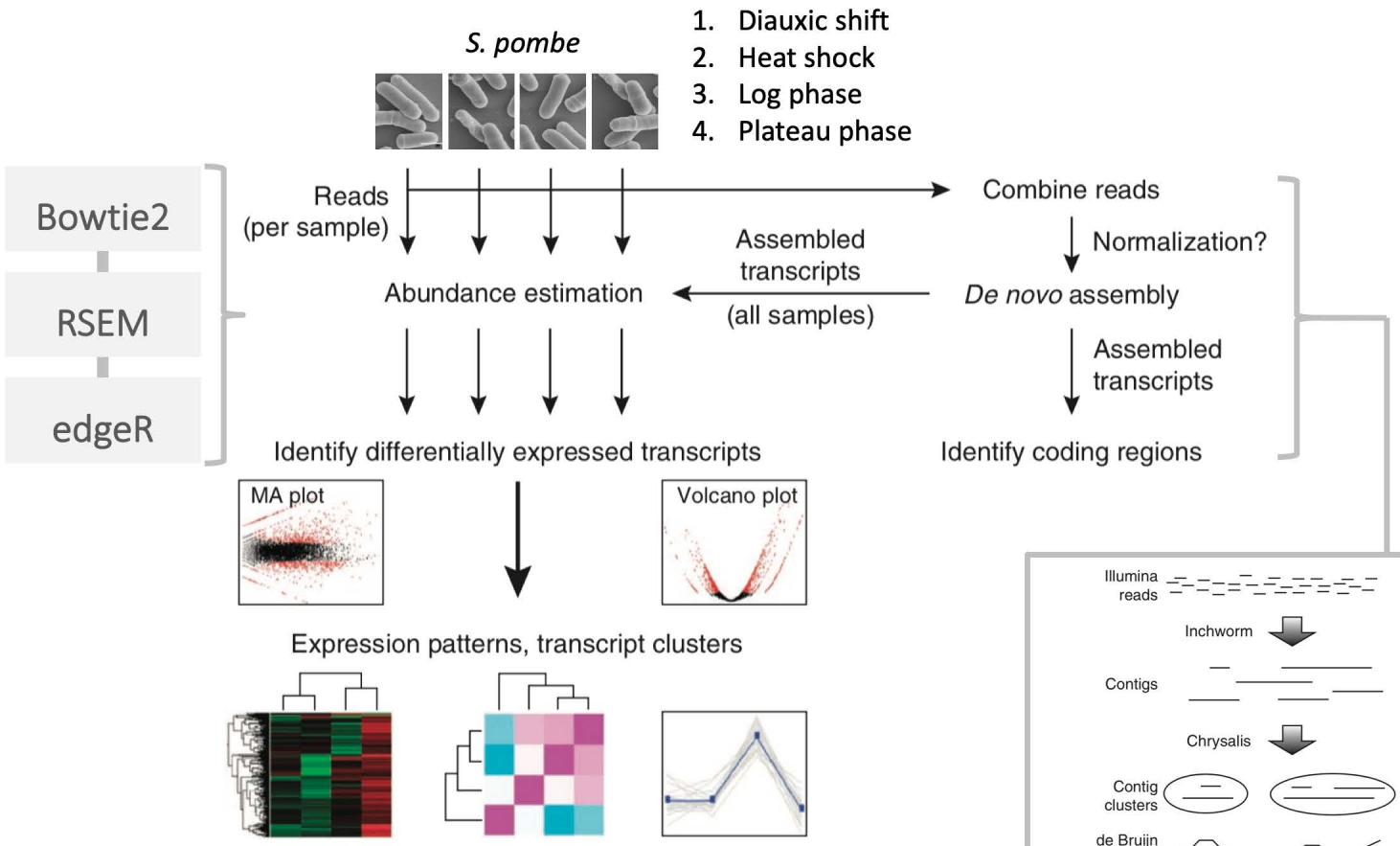
Genome browsers

UCSC Genome Browser



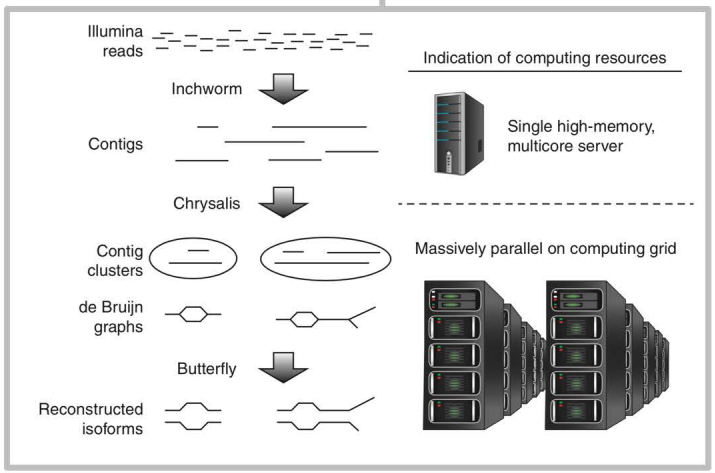
https://genome.ucsc.edu

Trinity workflow



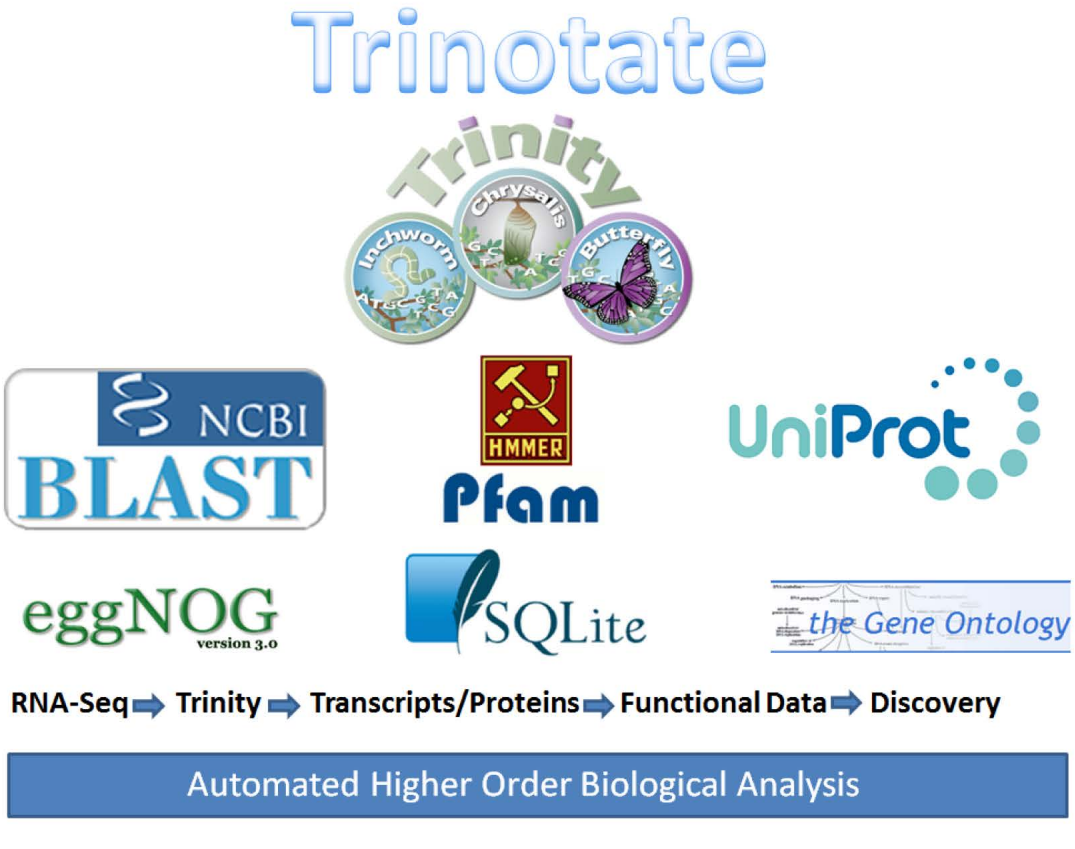
Haas et al (2013)

https://www.broadinstitute.org/videos/trinity-how-it-works



Functional annotation

Colorado State University



http://trinotate.github.io