

# Steps in an RNA-seq analysis

---

Jeff Leek

@jtleek

[www.jtleek.com](http://www.jtleek.com)

**DNA**

ACTGACCTAGATCAGTGTAGCGATCGTATACGAGACCGATTTCATCGGCAT



**transcription**

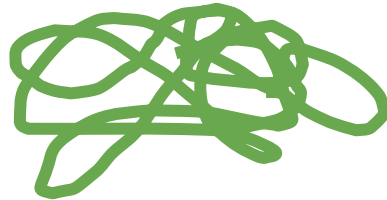
**RNA**

AUCAGUCGAUCACCGAU



**translation**

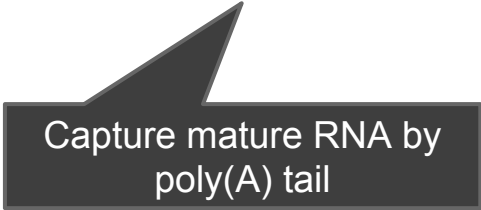
**protein**



Fragmented RNA  
molecule

AUGGGAAUUCACGAAUUCUAGAAAAAA


AUGGGAUUCACGAAUUCUAGAAAAAA



Capture mature RNA by  
poly(A) tail

AUGGGAUUCACGAAUUCUAGAAAAAAAAA

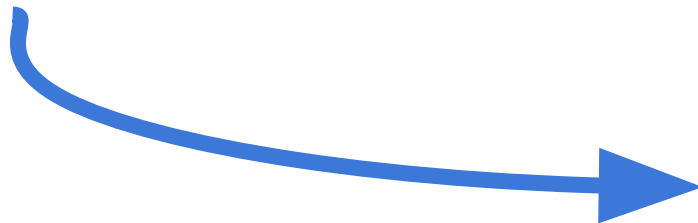
ATGGGAATTCACGAATTCCTAG

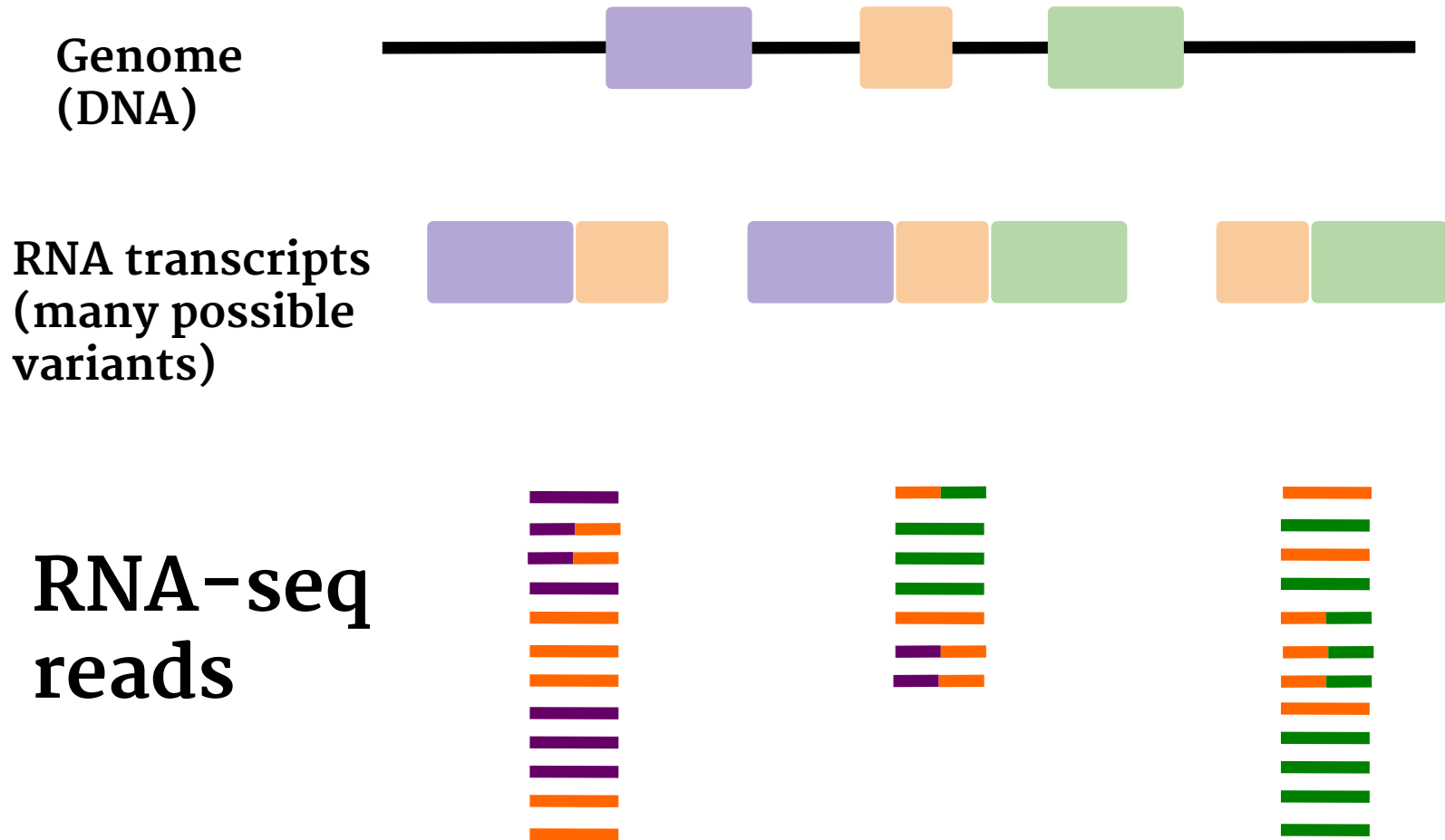


Reverse transcribe into  
complementary DNA  
(cDNA)

AUGGGAUUCACGAAUUCCUAGAAAAAAAAA

ATGGGAATTCACGAATTCCTAG



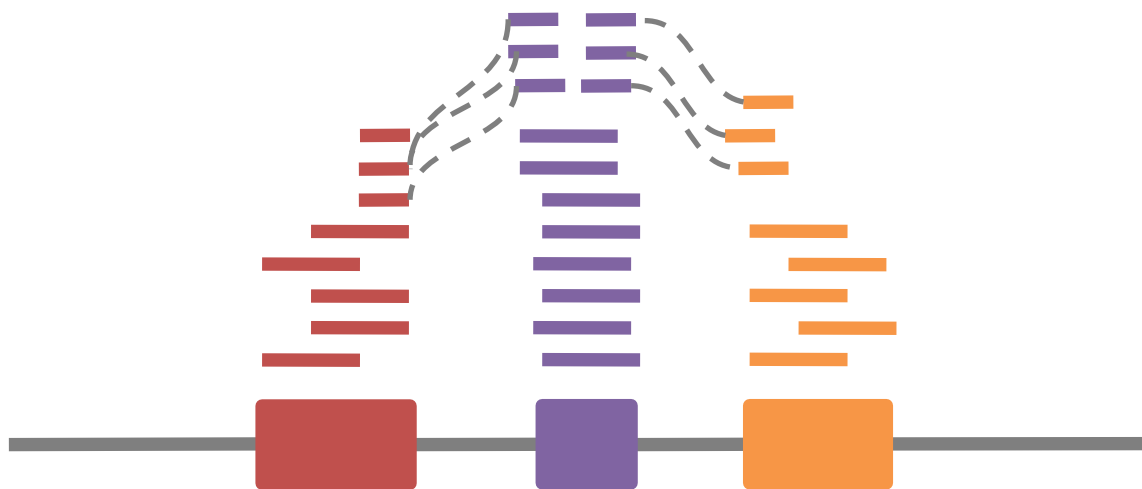


# Step 1: Align

Software:

- [HiSat](#)
- [Rail](#)
- [Star](#)
- [Tophat2](#)

Genome



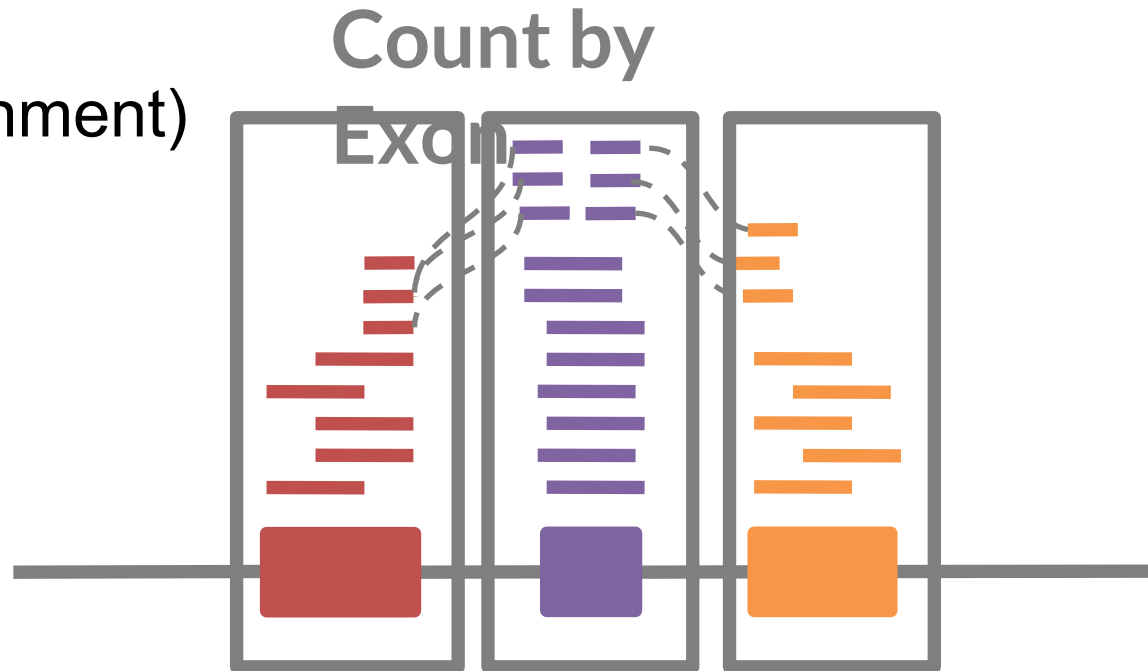


# Step 2: Count

Software:

- [HTSeq](#)
- [featureCounts](#)
- [kallisto](#) (no alignment)
- [derfinder](#)

Genome

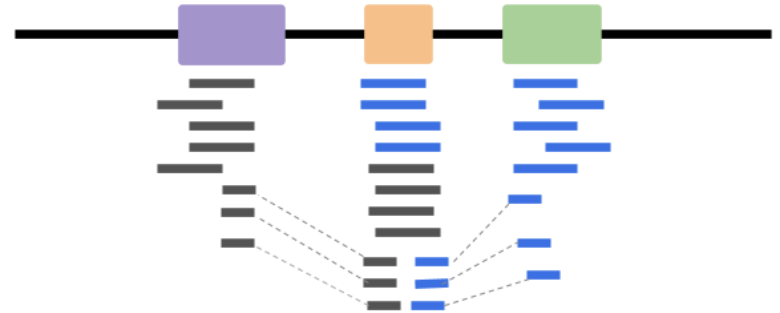


# Step 2: Assemble and quantify

Software:

- [StringTie](#)
- [Cufflinks](#)
- [Trinity](#)
- [RSEM](#)

Genome



expression  $\approx 12$  for both  
assembled transcripts

Estimated  
Transcripts



# Step 3: Normalize

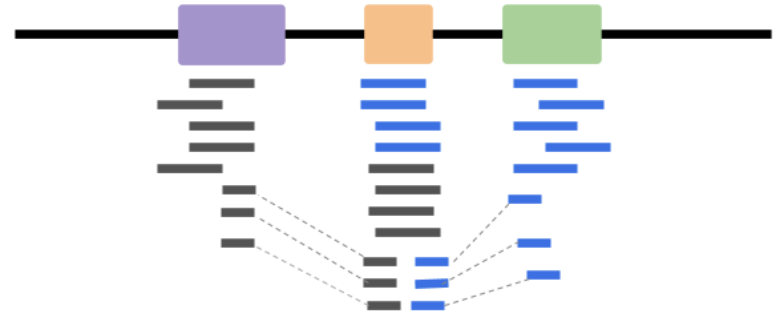
Software Normalize:

- [EDAseq](#)
- [cqn](#)
- [DESeq2/edgeR](#)
- [Ballgown](#)
- [derfinder](#)

Software Batch Effects:

- [sva](#)
- [RUVseq](#)

Genome



**expression  $\approx 12$  for both  
assembled transcripts**

Estimated  
Transcripts

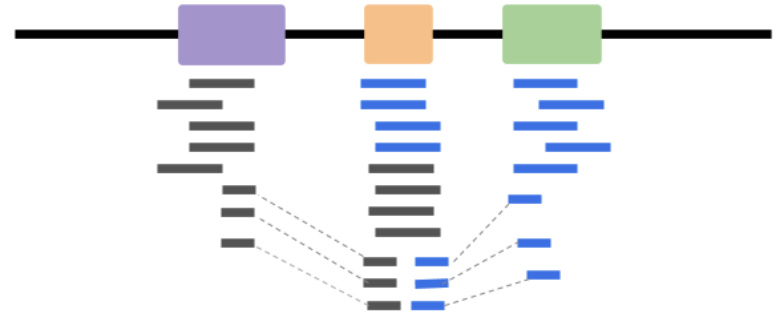


# Step 4: Statistical tests

Software:

- [DESeq2/edgeR](#)
- [Ballgown](#)
- [derfinder](#)

Genome



**expression  $\approx 12$  for both  
assembled transcripts**

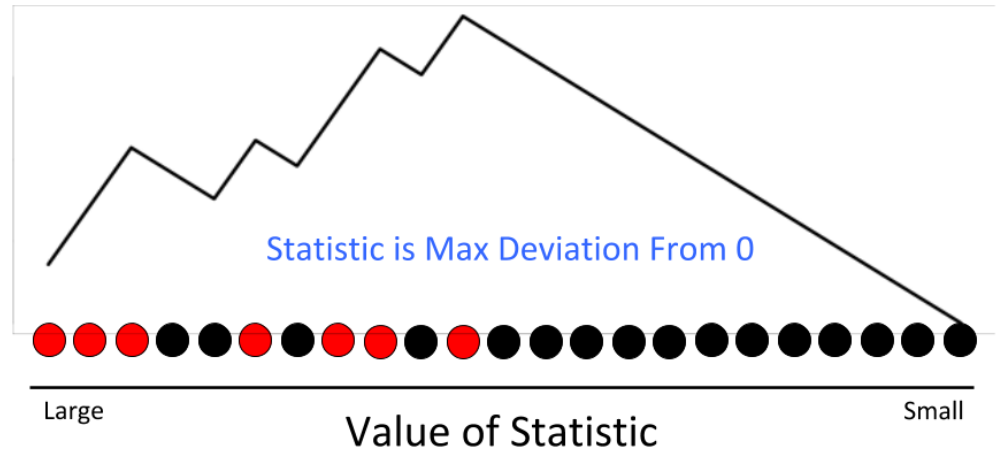
Estimated  
Transcripts



# Step 5: Gene set enrichment

Software:

- [goseq](#)
- [SeqGSEA](#)



- Gene In A Relevant Set
- Gene Not In The Set