

Steps in an RNA-seq analysis

Jeff Leek

@jtleek

www.jtleek.com

Background on RNA-seq

DNA

ACTGACCTAGATCAGTGTAGCGATCGTATACGAGACCGATTTCATCGGCAT



transcription

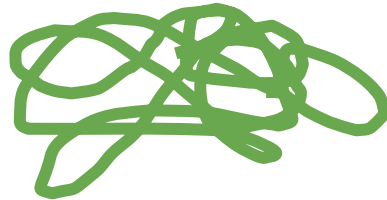
RNA

AUCAGUCGAUCACCGAU



translation

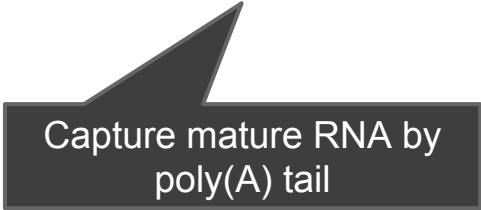
protein



Fragmented RNA
molecule

AUGGGAAUUCACGAAUUCUAGAAAAAA


AUGGGAAUUCACGAAUUCUAGAAAAAA



Capture mature RNA by
poly(A) tail

AUGGGAUUCACGAAUUCUAGAAAAAAA

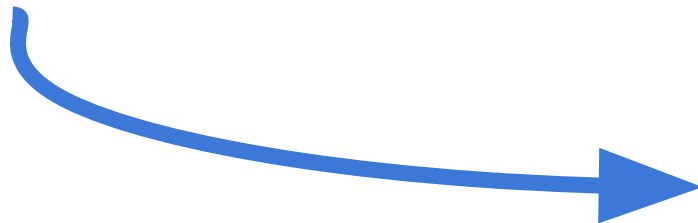
ATGGGAATTCACGAATTCCTAG



Reverse transcribe into
complementary DNA
(cDNA)

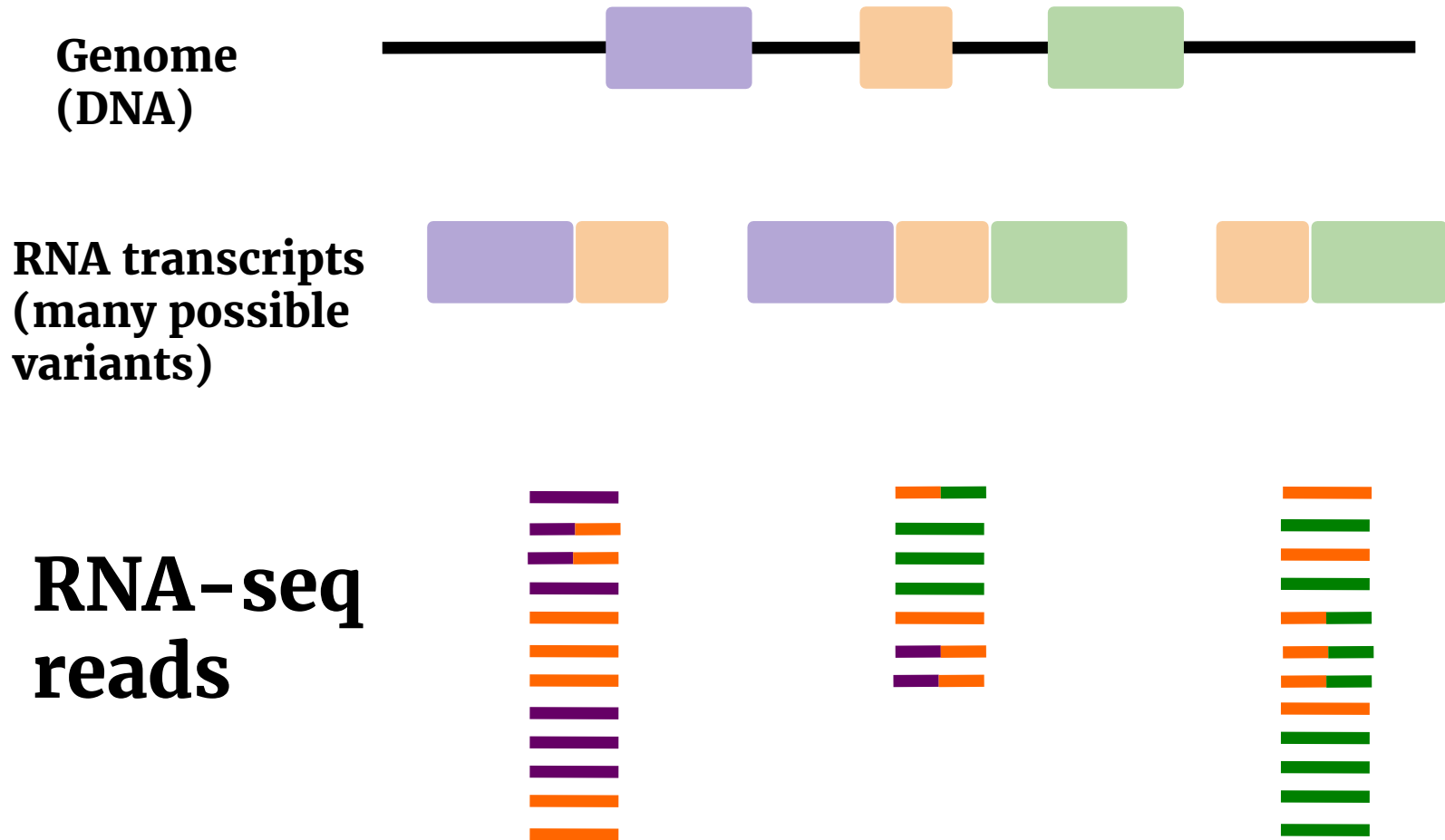
AUGGGAUUCACGAAUUCUAGAAAAAAAAA

ATGGGAATTCACGAATTCCTAG



Steps

1. Align
2. Count or quantify
3. Normalize
4. Statistical test
5. Gene set enrichment

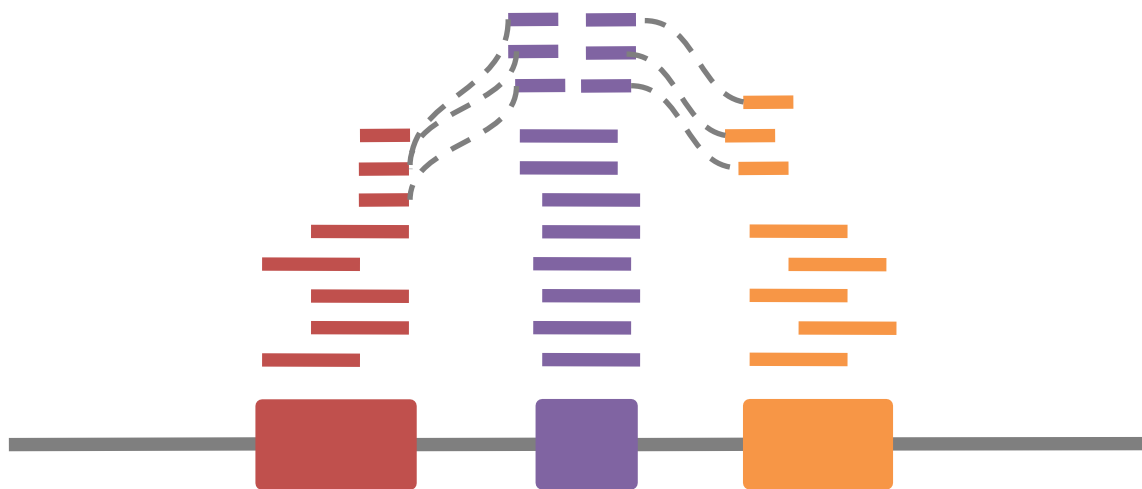


Step 1: Align

Software:

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

Genome

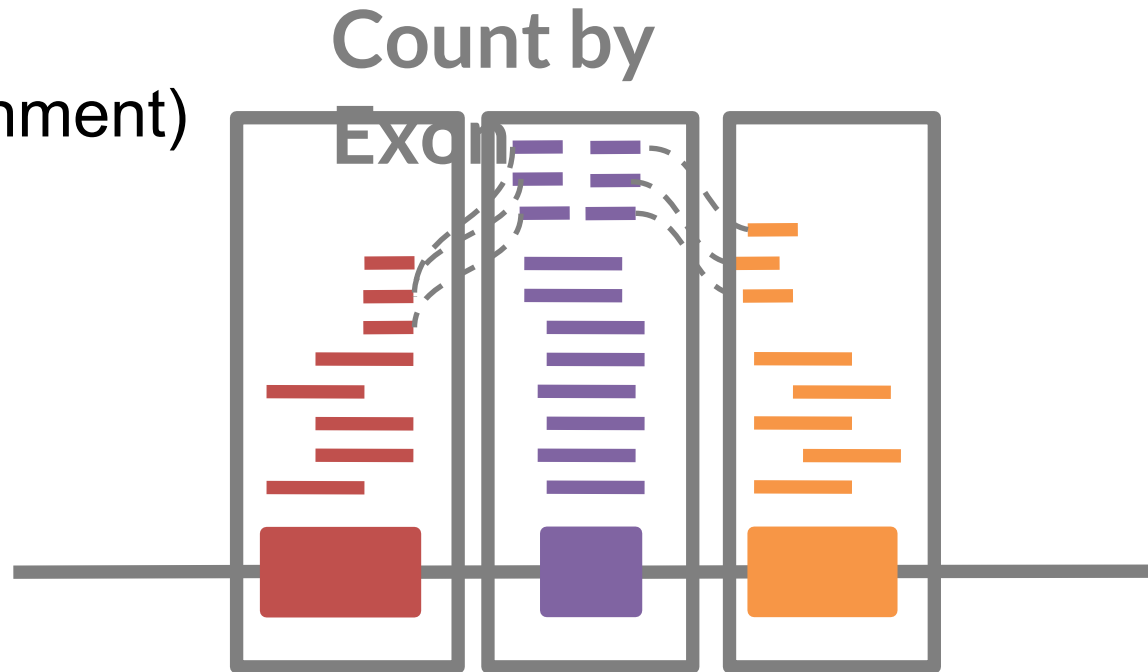


Step 2: Count

Software:

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

Genome

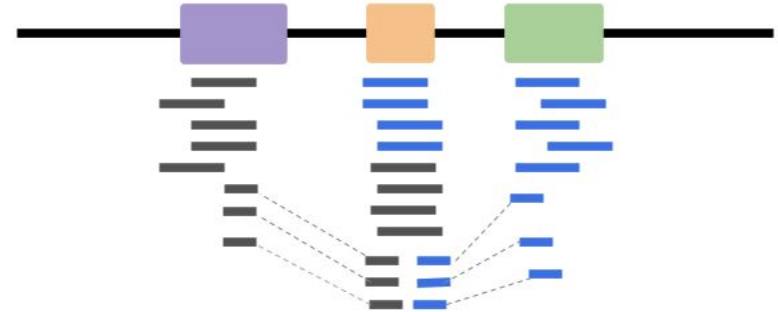


Step 2: Assemble and quantify

Software:

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

Genome



expression ≈ 12 for both
assembled transcripts

Estimated
Transcripts



Step 3: Normalize

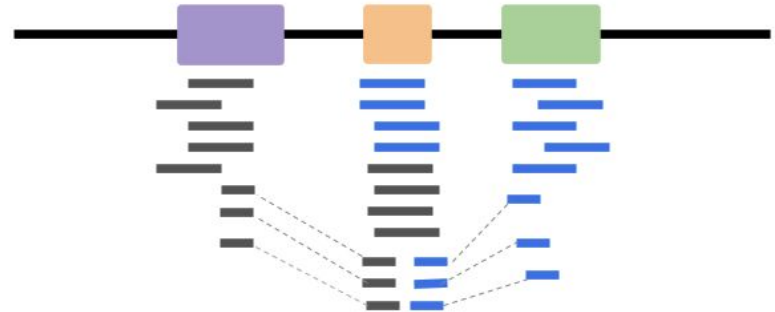
Software Normalize:

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

Software Batch Effects:

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

Genome



**expression ≈ 12 for both
assembled transcripts**

Estimated
Transcripts

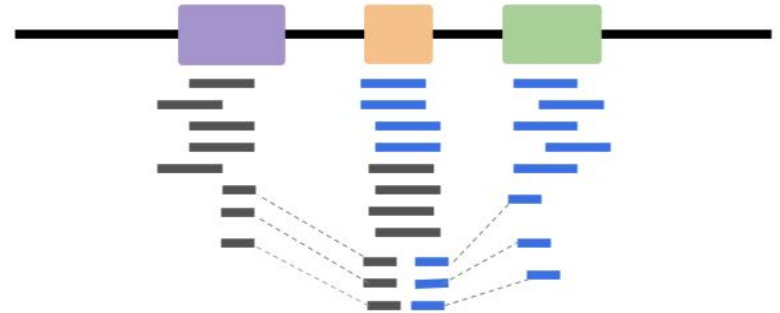


Step 4: Statistical tests

Software:

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

Genome



**expression ≈ 12 for both
assembled transcripts**

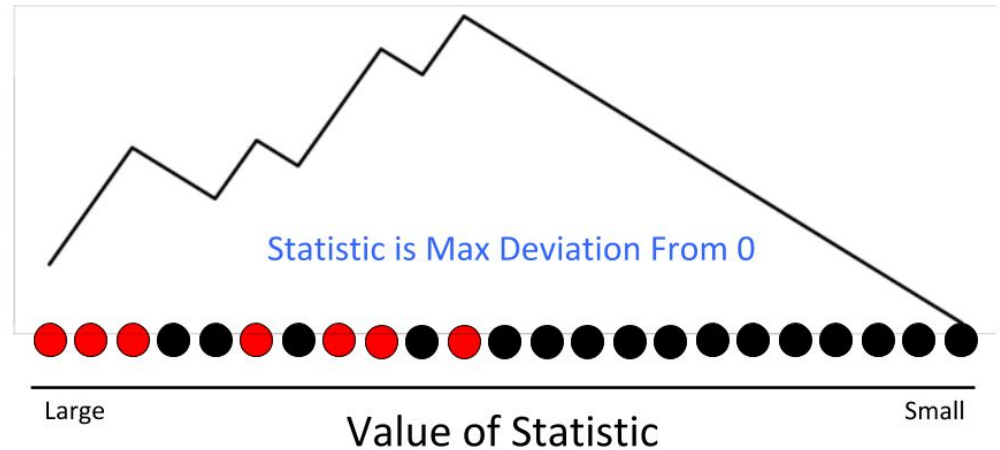
Estimated
Transcripts



Step 5: Gene set enrichment

Software:

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



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