Steps in an RNA-seq analysis

Jeff Leek

@jtleek

www.jtleek.com

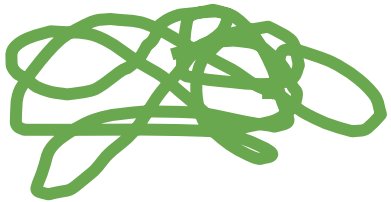
Background on RNA-seq

DNA ACTGACCTAGATCAGTGTAGCGATCGTATACGAGACCGATTCATCGGCAT

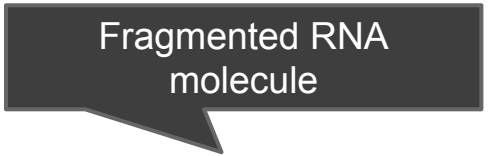
↓ transcription

RNA AUCAGUCGAUCACCGAU

↓ translation

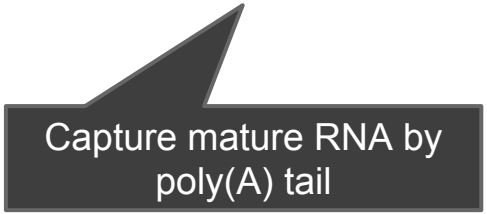


protein



AUGGGAAUUCACGAAUUCCUAGAAAAAAAAA

AUGGGAAUUCACGAAUUCCUAGAAAAAAAAA



AUGGGAAUUCACGAAUUCCUAGAAAAAAAA

Reverse transcribe into complementary DNA (cDNA)

ATGGGAATTCACGAATTCCTAG

AUGGGAAUUCACGAAUUCCUAGAAAAAAAA

ATGGGAATTCACGAATTCCTAG

Slide courtesy: Alyssa Frazee

Genome (DNA)

RNA transcripts (many possible variants)

RNA-seq reads

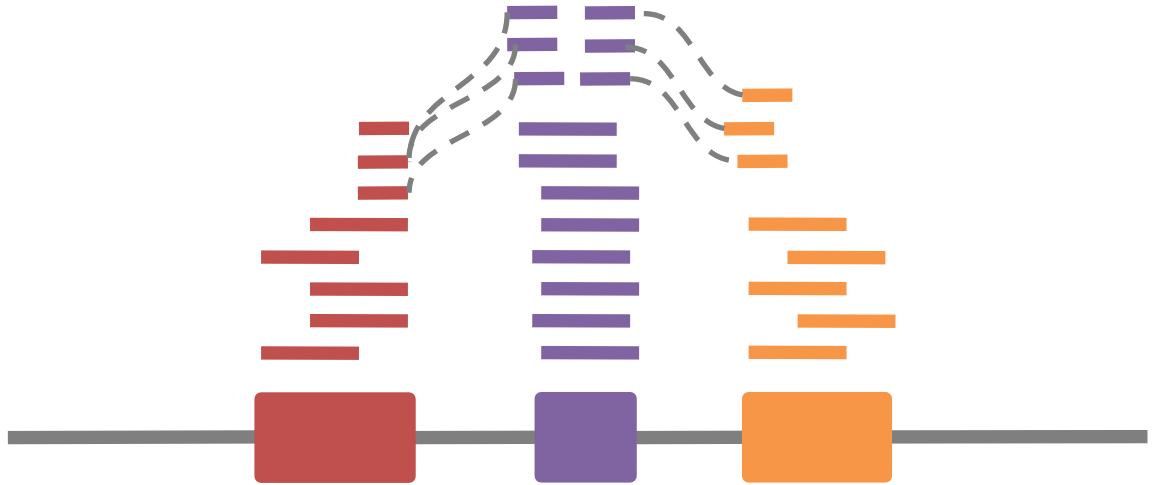
Steps

1. Align
2. Count, assemble, quantify
3. Normalize
4. Statistical tests
5. Gene set enrichment

Step 1: Align

Software:

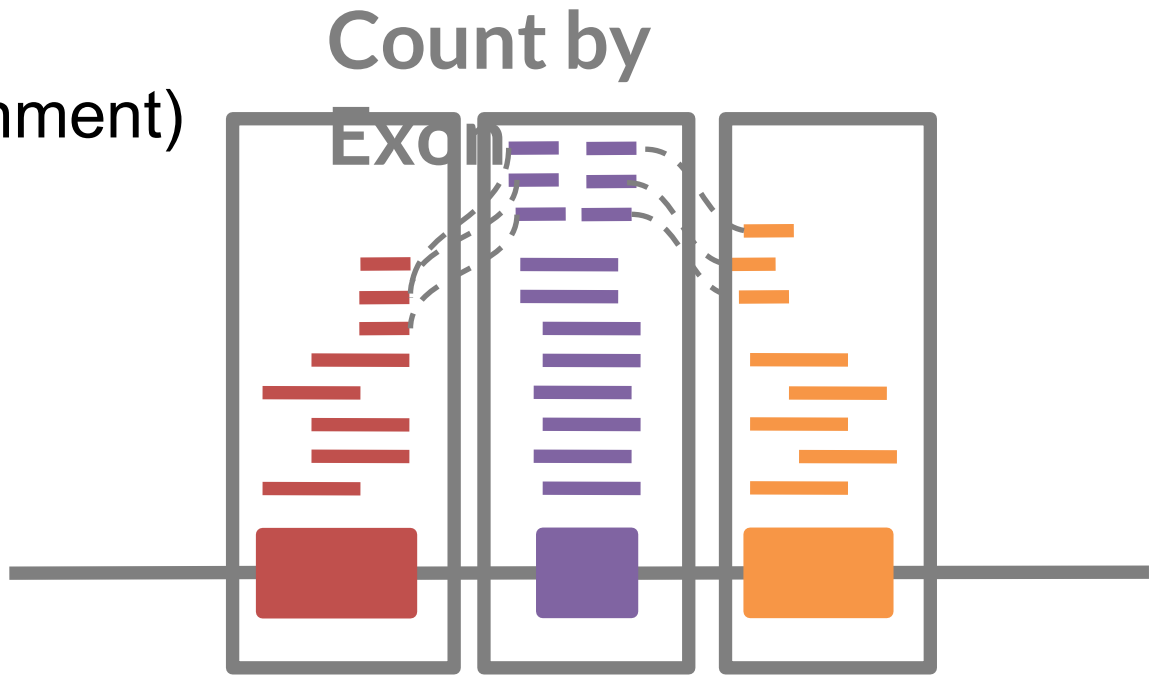
- HiSat
- Rail
- Star
- Tophat2



Step 2: Count

Software:

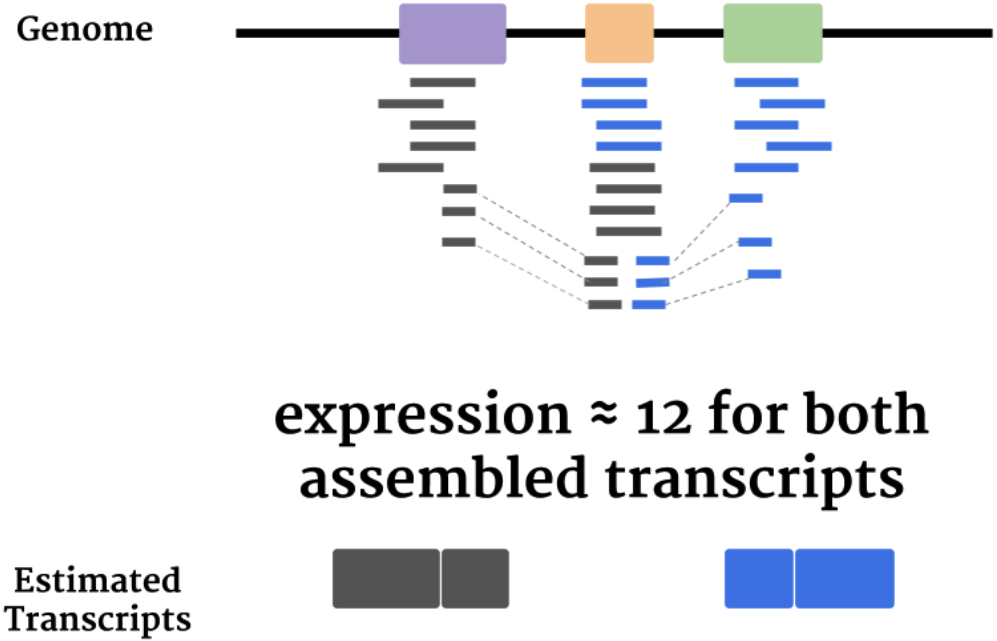
- HTSeq
- featureCounts
- kallisto (no alignment)
- derfinder



Step 2: Assemble and quantify

Software:

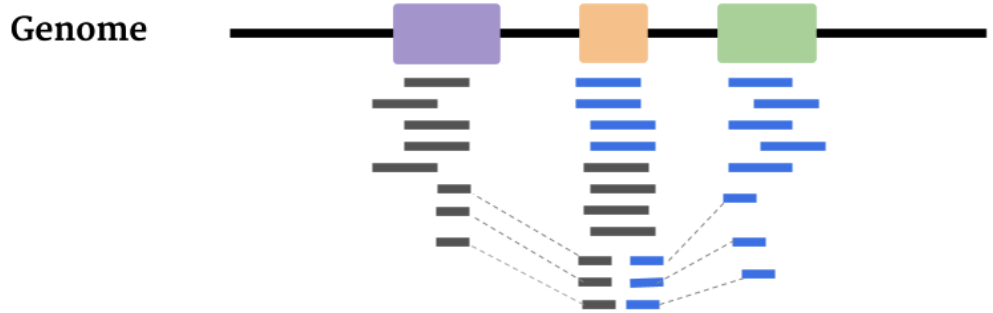
- StringTie
- Cufflinks
- Trinity
- RSEM



Step 3: Normalize

Software Normalize:

- EDAseq
- cqn
- DESeq2/edgeR
- Ballgown
- derfinder



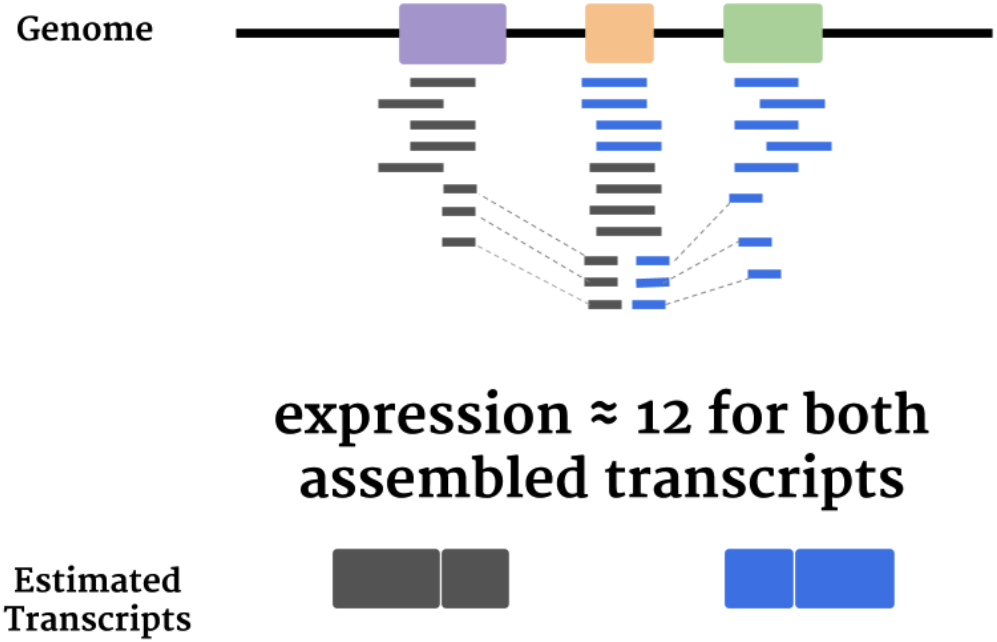
Software Batch Effects:

- sva
- RUVseq

Step 4: Statistical tests

Software:

- DESeq2/edgeR
- Ballgown
- derfinder



Step 5: Gene set enrichment

Software:

- goseq
- SeqGSEA