

# **GRABBAG!**

**STEPHANIE J SPIELMAN, PHD**

**BIO5312, FALL 2017**

# REGULAR EXPRESSIONS

- *Pattern-based* search and replace
- **Extremely powerful beyond all reason**
- **Excellent for text (file) manipulation!**

# CRITICAL PSA: TEXT EDITORS

- **Microsoft Word is not a text editor!!!!!! I'm so serious!!!**
- **GUI**
  - ~~TextEdit and Notepad~~
  - Textwrangler/BBEdit for Macs
  - Sublime 3 for everyone else
  - Newer, awesome one called Atom
- **CLI**
  - Vim/vi, emacs, nano, pico (b/c puns)
  - [https://en.wikipedia.org/wiki/Editor\\_war](https://en.wikipedia.org/wiki/Editor_war)

# REGULAR EXPRESSIONS

String: Mus musculus

Regex: Mus

Match: **Mus** musculus

# REGULAR EXPRESSIONS

String: Mus musculus

Regex: Mus musculus

Match: Mus musculus

# REGULAR EXPRESSIONS

String: Mus musculus

Regex: `[mM]us`

Match: **Mus** **mus**culus

# REGULAR EXPRESSIONS

String: Mus musculus

Regex: **[A-Za-z]**us

Match: **Mus** **mus**cu**lus**

# REGULAR EXPRESSIONS

String: Mus musculus

Regex: `\wus`

Match: **Mus** **mus**cu**lus**



# REGULAR EXPRESSIONS

String: Mus musculus

Regex: `\w+`

Match: `Mus` `musculus`

# REGULAR EXPRESSIONS

String: Mus musculus

Regex: `[A-Z]\w+ \w+`

Match: `Mus musculus`

# REGULAR EXPRESSIONS

String: Mus musculus

Regex: `([A-Z])\w+ (\w+)`

Replace: `\1. \2`

New string: M. musculus

# REGULAR EXPRESSIONS

String: 85.34 cm

Regex: `\d+`

Match: **85**.**34** cm

# REGULAR EXPRESSIONS

String: 85.34 cm

Regex: `\d+\.\d+`

Match: `85.34` cm

# REGULAR EXPRESSIONS

String: 85.34 cm

Regex: `\d+\.\d+ \w+`

Match: `85.34 cm`

# REGULAR EXPRESSIONS

String: 85 cm

Regex: `\d+\.\d+ \w+`

Match: 85 cm

# REGULAR EXPRESSIONS

String: 85 cm

Regex: `\d+\.\d*\s\d+`

Match: `85 cm`



# REGULAR EXPRESSIONS

String: 85 cm

Regex: `^\d`

Match: **8**5 cm

# REGULAR EXPRESSIONS

String: 85 cm

Regex: `\w$`

Match: 85 cm

# REGULAR EXPRESSIONS

String: 85.341234 cm

Regex: `(\d+\.\d{3})\d+ cm`

Replace: `\1`

New string: 85.341

# REGULAR EXPRESSIONS

String: 85.34 cm

Regex: `(\d+\.\d{3})\d+ cm`

Replace: `\1`

New string: ?????

# GROUP EXERCISE

Come up with a regular expression to convert the following text:

85.34 cm

85.678 cm

923.1115 cm

1.95 cm

6 cm

85.3 cm

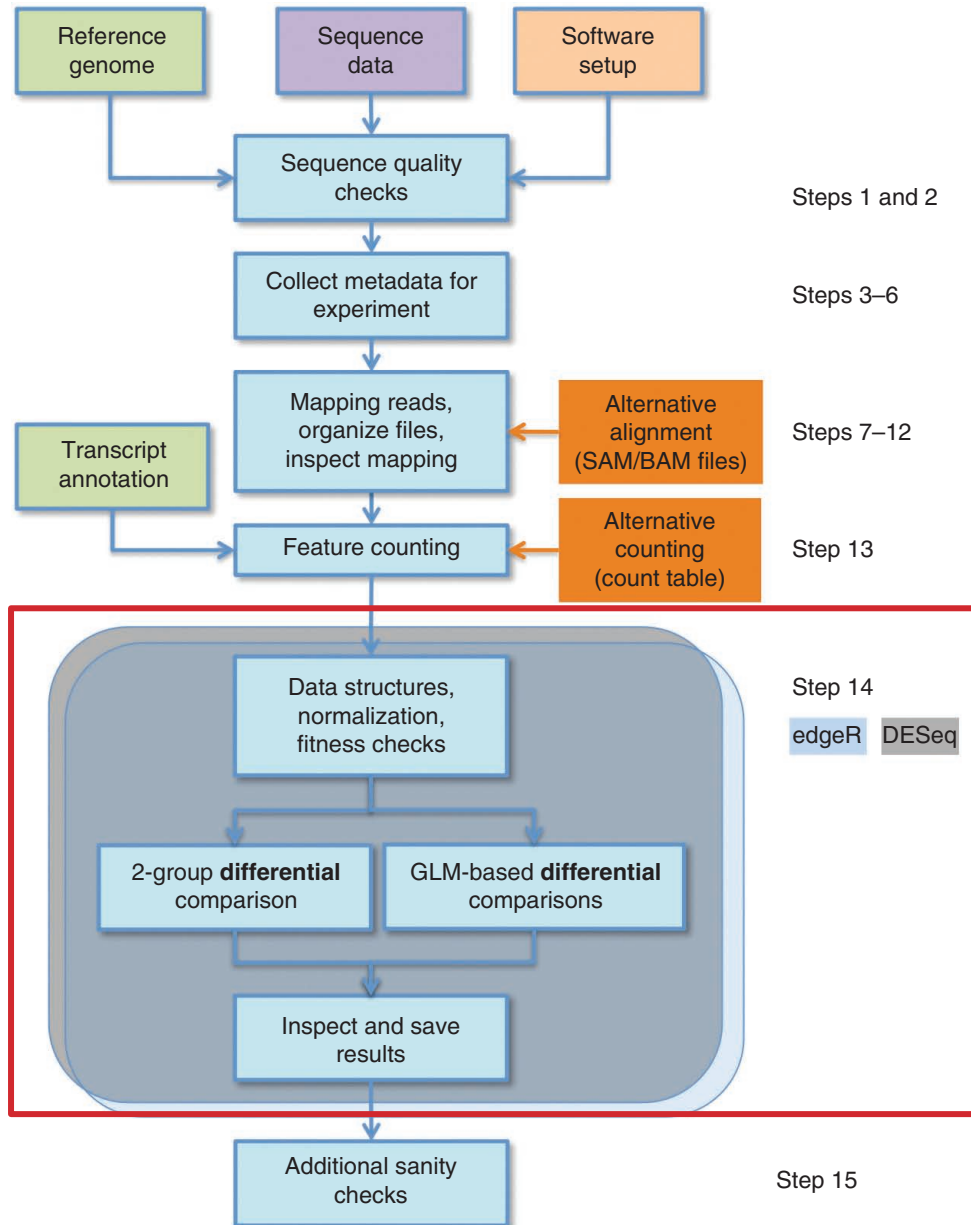
85.6 cm

923.1 cm

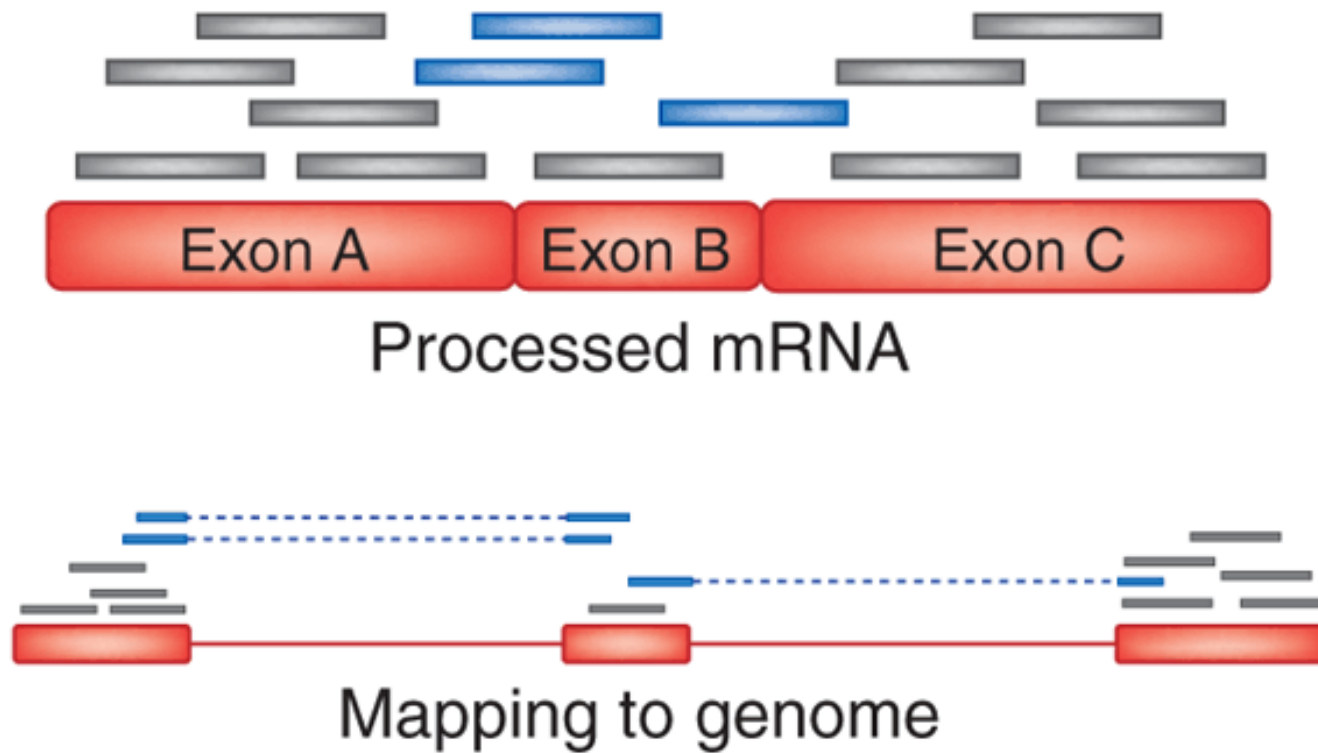
1.9 cm

6 cm

**BREAK**



# USE A SPLICE-AWARE ALIGNER





# ALIGNERS AND PSEUDO-

## PROTOCOL

### Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown

Mihaela Pertea<sup>1,2</sup>, Daehwan Kim<sup>1</sup>, Geo M Pertea<sup>1</sup>, Jeffrey T Leek<sup>3</sup> & Steven L Salzberg<sup>1-4</sup>

#### STAR: ultrafast universal RNA-seq aligner

Alexander Dobin<sup>1,\*</sup>, Carrie A. Davis<sup>1</sup>, Felix Schlesinger<sup>1</sup>, Jorg Drenkow<sup>1</sup>, Chris Zaleski<sup>1</sup>, Sonali Jha<sup>1</sup>, Philippe Batut<sup>1</sup>, Mark Chaisson<sup>2</sup> and Thomas R. Gingeras<sup>1</sup>

<sup>1</sup>Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA and <sup>2</sup>Pacific Biosciences, Menlo Park, CA, USA

Associate Editor: Inanc Birol

#### Near-optimal probabilistic RNA-seq quantification

Nicolas L Bray<sup>1</sup>, Harold Pimentel<sup>2</sup>, Páll Melsted<sup>3</sup>  
& Lior Pachter<sup>2,4,5</sup>

We present kallisto, an RNA-seq quantification program that is two orders of magnitude faster than previous approaches and achieves similar accuracy. Kallisto pseudoaligns reads to a reference, producing a list of transcripts that are compatible with each read while avoiding alignment of individual bases. We use kallisto to analyze 30 million unaligned paired-end RNA-seq reads in <10 min on a standard laptop computer. This removes a major computational bottleneck in RNA-seq analysis.

THIS IS  
THE NEW  
TOPHAT2