Introduction to Transcriptomics Analysis

Class 10 - Practice about Gene Expression Quantification.

INSTRUCTOR:
Aureliano Bombarely
Department of Bioscience
Universita degli Studi di Milano
aureliano.bombarely@unimi.it

Outline of Topics

- Exercise 1: Running Stringtie on the mapped reads.
- Exercise 2: Running HTSeq to calculate expression

Outline of Topics

- Exercise 1: Running Stringtie on the mapped reads.
- Exercise 2: Running HTSeq to calculate expression

A- RNASeq Analysis pipeline with Hisat2-StringTie-Ballgown

PROTOCOL

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

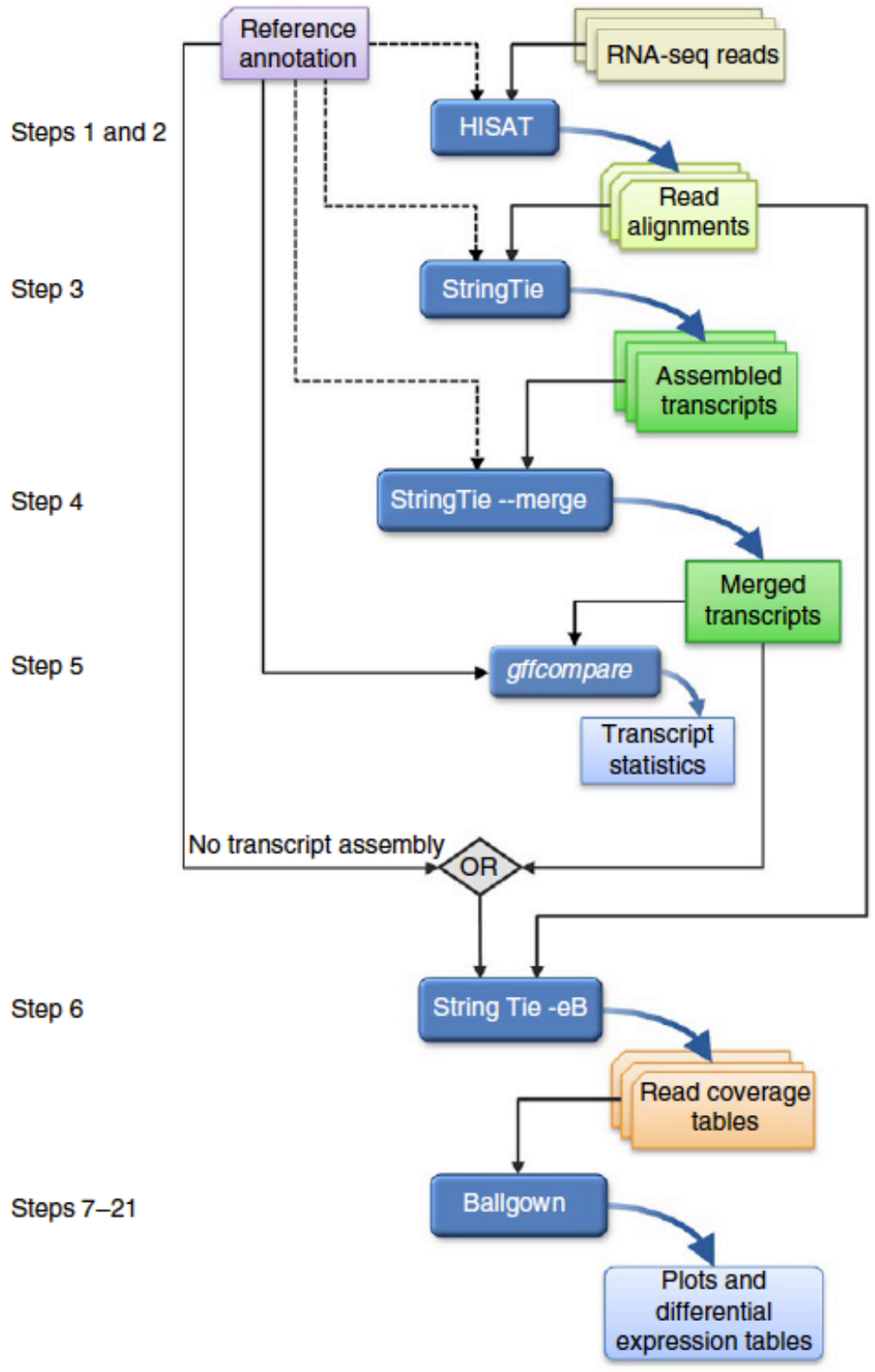
Mihaela Pertea[1,2], Daehwan Kim[1], Geo M Pertea[1], Jeffrey T Leek[3] & Steven L Salzberg[1–4]

[1]Center for Computational Biology, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, Maryland, USA. [2]Department of Computer Science, Whiting School of Engineering, Johns Hopkins University, Baltimore, Maryland, USA. [3]Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, USA. [4]Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland, USA.
Correspondence should be addressed to S.L.S. (salzberg@jhu.edu).

Published online 11 August 2016; doi:10.1038/nprot.2016.095

Pertea, Mihaela, et al. "Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown." *Nature protocols* 11.9 (2016): 1650.

A- RNASeq Analysis pipeline with Hisat2-StringTie-Ballgown



Pertea, Mihaela, et al. "Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown." *Nature protocols* 11.9 (2016): 1650.

- Exercise 1: Running Stringtie on the mapped reads.

Goal: Generate the expression tables for Stringtie. Stringtie is divided in three steps: 1- Transcript discovery; 2- Comparison and 3- Quantification

Input:

- Reads mapped in BAM format (Artha_*.bam files)
- GFF file with the annotations (Araport11_GFF3_genes_transposons.201606.gtf)

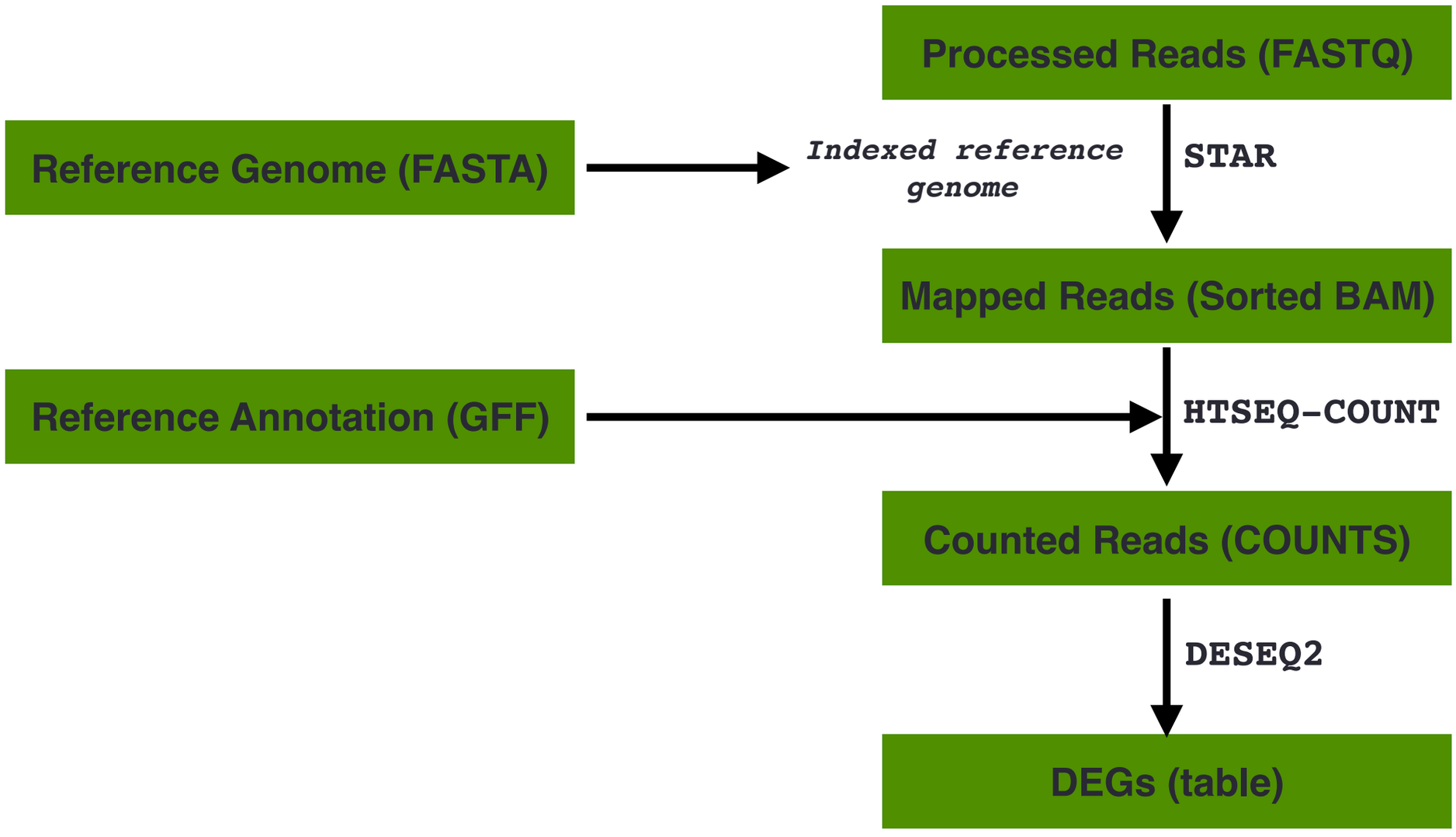
Recommended commands:

```
stringtie -p 48 -G ref.gtf -o my_sample.gtf -l my_sample_label my_sample.bam

ls | grep gtf > merged_list.txt

stringtie --merge -p 8 -G ref.gtf -o stringtie_merged.gtf merged_list.txt

gffcompare Stringtie_merged.gtf -r ref.gtf

stringtie -e -B -p 48 -G stringtie_merged.gtf -o my_sample/count.gtf
my_sample.bam
```

Outline of Topics

- Exercise 1: Running Stringtie on the mapped reads.
- Exercise 2: Running HTSeq to calculate expression

B- RNASeq Analysis pipeline with STAR-HTSeqCount—DESeq2



- Exercise 2: Running HTSeq to calculate expression

Goal: Generate the expression tables with HTSeq. It will be a single step.

Input:

- Reads mapped in BAM format (Artha_*.bam files)
- GFF file with the annotations (Araport11_GFF3_genes_transposons.201606.gff)

Recommended command:

```
htseq-count -f bam -m union -r pos -i ID -a 10 --type=gene --stranded=no
my_star_align.bam my_ref.gff > my_htseq.counts
```