Introduction to Transcriptomics Analysis

Class 08 - Practice about Read Mapping.

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

Outline of Topics

- Exercise 1: Generating the Hisat2 index.
- Exercise 2: Mapping with Hisat2
- Exercise 3: Evaluating Hisat2 results
- Exercise 4: Generating the STAR index
- Exercise 5: Mapping with STAR.
- Exercise 6: Evaluating STAR results

A- RNASeq Analysis pipeline with Hisat2-StringTie-Ballgown

PROTOCOL

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

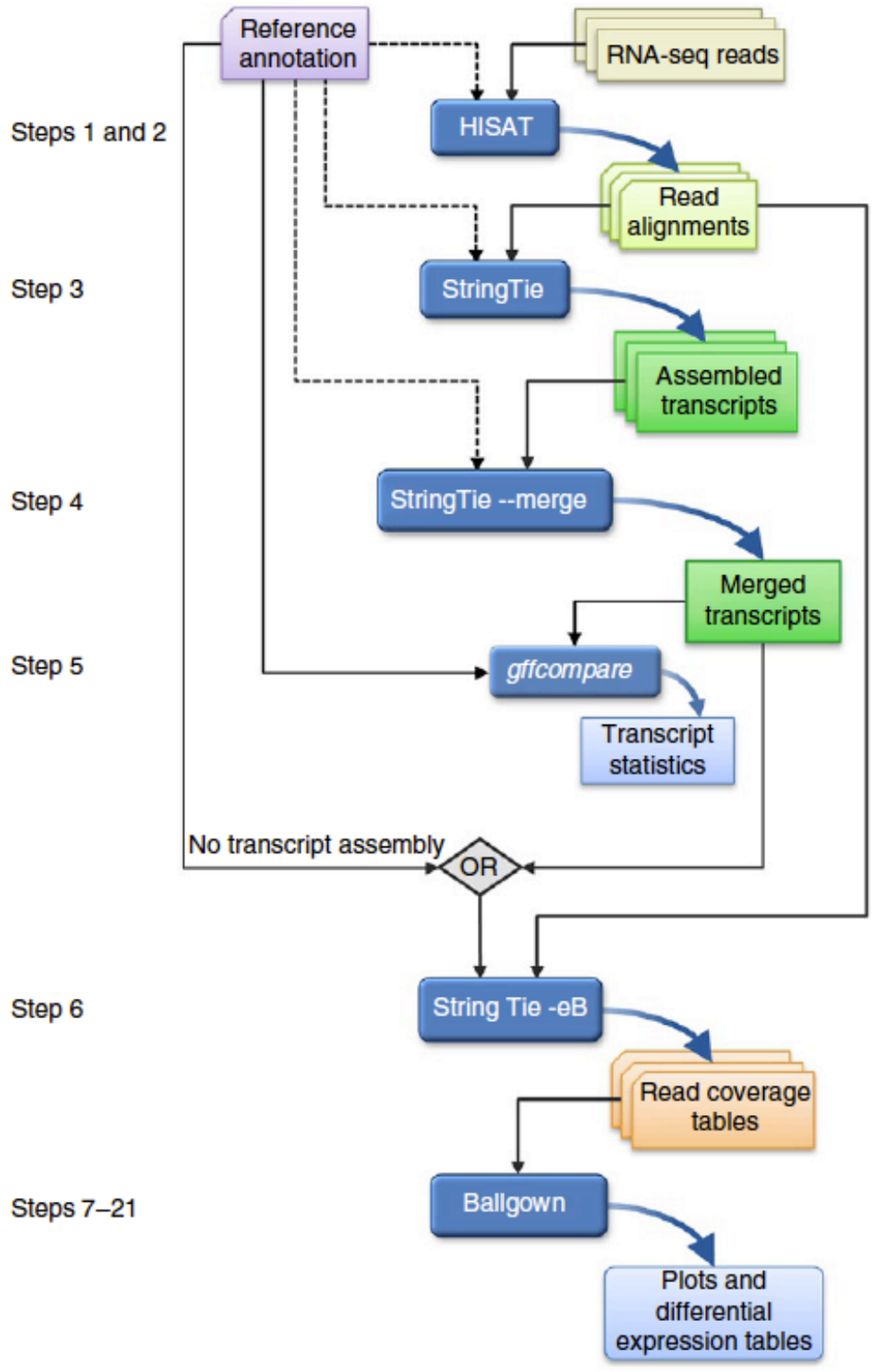
Mihaela Pertea[1,2], Daehwan Kim[1], Geo M Pertea[1], Jeffrey T Leek[3] & Steven L Salzberg[1–4]

[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.

Outline of Topics

- Exercise 1: Generating the Hisat2 index.
- Exercise 2: Mapping with Hisat2
- Exercise 3: Evaluating Hisat2 results
- Exercise 4: Generating the STAR index
- Exercise 5: Mapping with STAR.
- Exercise 6: Evaluating STAR results

- Exercise 1: Generating the Hisat2 index.

  Goal: Generate the Hisat2 index previous to the read mapping.

  Input: File “Artha_TAIR10_genome.fasta”

  Recommended command: **hisat2-build <reference.fasta> <index_name>**

Outline of Topics

- Exercise 1: Generating the Hisat2 index.
- Exercise 2: Mapping with Hisat2
- Exercise 3: Evaluating Hisat2 results
- Exercise 4: Generating the STAR index
- Exercise 5: Mapping with STAR.
- Exercise 6: Evaluating STAR results

- Exercise 2: Mapping with Hisat2.

Goal: Map the reads (by pairs) to the reference genome using Hisat2. The output will be piped into samtools to convert SAM to BAM and filter out the non-mapped reads

Input:

- Hisat2 index of “Artha_TAIR10_genome.fasta”
- Reads to map in FASTQ format (Artha_*.fastq.gz files)

Recommended commands:

```
hisat2 -p 100 --dta -x <hisat2_reference> -1 <first_pair.fq.gz> -2
<second_pair.fq.gz> --summary-file <mapping_summary.txt> | samtools view -F
4 -Sb -o <my_mapped.bam> -


samtools sort -o <my_mapped.bam> <my_mapped.bam>
```

Outline of Topics

- Exercise 1: Generating the Hisat2 index.
- Exercise 2: Mapping with Hisat2
- Exercise 3: Evaluating Hisat2 results
- Exercise 4: Generating the STAR index
- Exercise 5: Mapping with STAR.
- Exercise 6: Evaluating STAR results

- Exercise 3: Evaluating Hisat2 results.

Goal: Evaluate the mapping results for each of the mapped read files (BAM) using bamtools stats. Additionally the summary output of HISAT can be explored

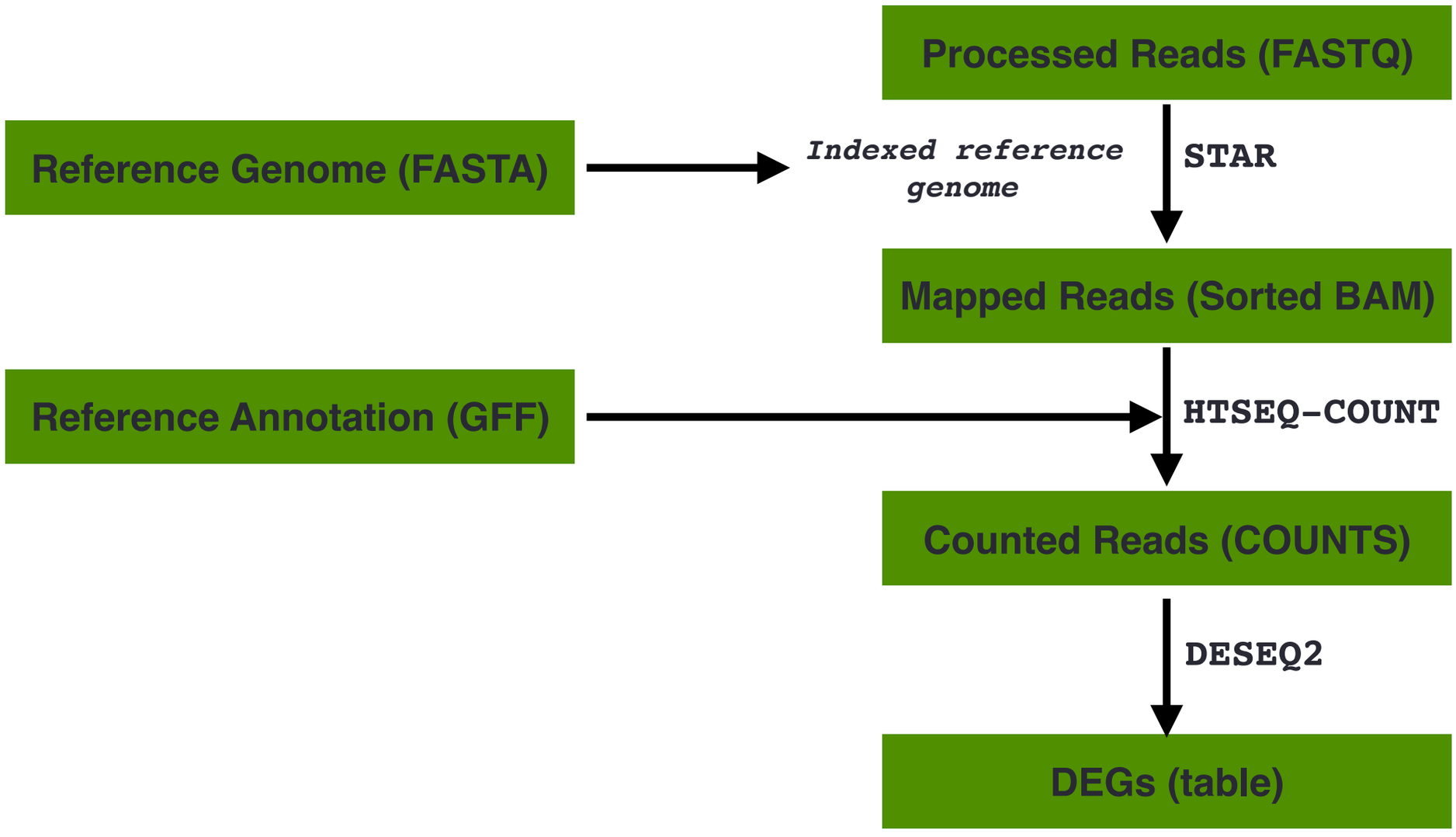
Input:

- Reads mapped in BAM format (Artha_*.bam files)

Recommended commands:

```
bamtools stats <my_mapped.bam>
```

B- RNASeq Analysis pipeline with STAR-HTSeqCount—DESeq2



Outline of Topics

- Exercise 1: Generating the Hisat2 index.
- Exercise 2: Mapping with Hisat2
- Exercise 3: Evaluating Hisat2 results
- Exercise 4: Generating the STAR index
- Exercise 5: Mapping with STAR.
- Exercise 6: Evaluating STAR results

- Exercise 4: Generating the STAR index

Goal: Generate the STAR index previous to the read mapping.

Input:

- Reference genome in FASTA format “Artha_TAIR10_genome.fasta”
- Annotation of the reference in GTF format “Araport11_GFF3_genes_transposons.201606.gtf”

Recommended commands:

```
gffread -T -o <ref.gtf> <ref.gff>

mkdir <ref_STAR_idxs>

STAR --runThreadN 8 --runMode genomeGenerate --genomeDir <ref_STAR_idxs> --
genomeFastaFiles <ref.fasta> --sjdbGTFfile <ref.gtf> –genomeSAindexNbases 12
```

Outline of Topics

- Exercise 1: Generating the Hisat2 index.
- Exercise 2: Mapping with Hisat2
- Exercise 3: Evaluating Hisat2 results
- Exercise 4: Generating the STAR index
- Exercise 5: Mapping with STAR.
- Exercise 6: Evaluating STAR results

- Exercise 5: Mapping with STAR.

Goal: Map the reads (by pairs) to the reference genome using STAR.

Input:

- STAR index of “Artha_TAIR10_genome.fasta”
- Reads to map in FASTQ format (Artha_*.fastq.gz files)

Recommended commands:

```
STAR --runThreadN 8 --genomeDir <ref_STAR_idxs> --readFilesIn
<first_pair.fq.gz> <second_pair.fq.gz> --outFileNamePrefix <sample1_rep1> --
readFilesCommand "gunzip -c" --outSAMtype BAM SortedByCoordinate
```

Outline of Topics

- Exercise 1: Generating the Hisat2 index.
- Exercise 2: Mapping with Hisat2
- Exercise 3: Evaluating Hisat2 results
- Exercise 4: Generating the STAR index
- Exercise 5: Mapping with STAR.
- Exercise 6: Evaluating STAR results

- Exercise 6: Evaluating STAR results

Goal: Evaluate the mapping results for each of the mapped read files (BAM) using bamtools stats. Additionally the summary output of STAR can be explored

Input:

- Reads mapped in BAM format (Artha_*.bam files)

Recommended commands:

```
bamtools stats <my_mapped.bam>
```