Transcriptome Quantification

PRACTICAL

Transcript quantification - featureCounts

Basic steps to analyse RNA-seq data

Mapping on reference genome

Transcript discovery
and/or
known transcripts

Quantification of gene/transcript expression

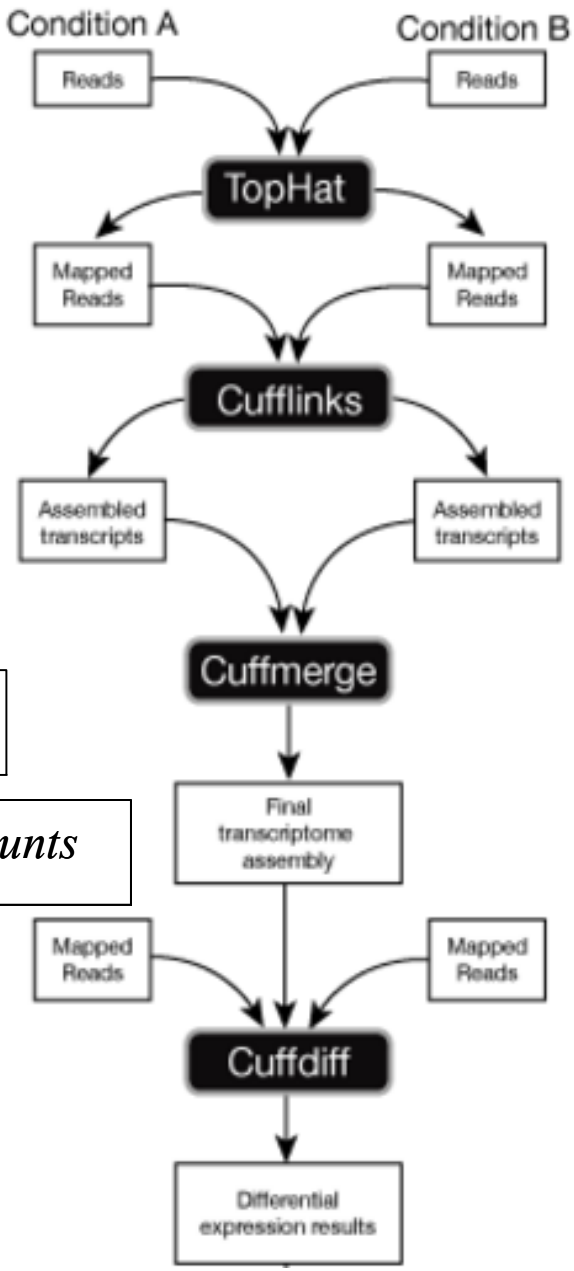
Differential analysis

featureCounts

raw counts

EdgeR,
DESeq,

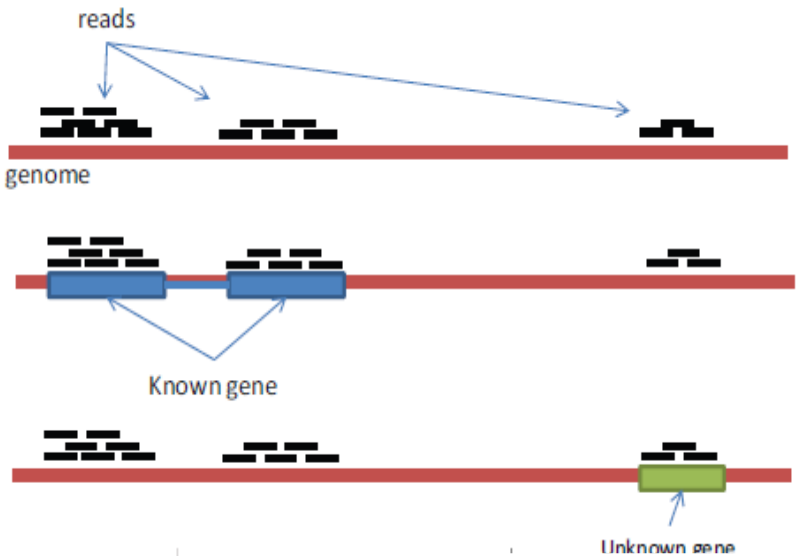
Tuxedo Suite



Transcript quantification - featureCounts

Transcript quantification is based on the number of reads (raw counts) that maps on it.

The simplest approach to quantification is to aggregate raw counts of mapped reads using programs such as HTSeq-count or **featureCounts**



Transcript quantification - featureCounts

featureCounts is a program suitable for counting reads generated from either RNA or genomic DNA sequencing experiments

faster than existing methods and requires far less computer memory

single or paired-end reads (in the second case it counts fragments rather than reads)

supports strand-specific read counting

possible to specify a minimum mapping quality score that the assigned reads must satisfy

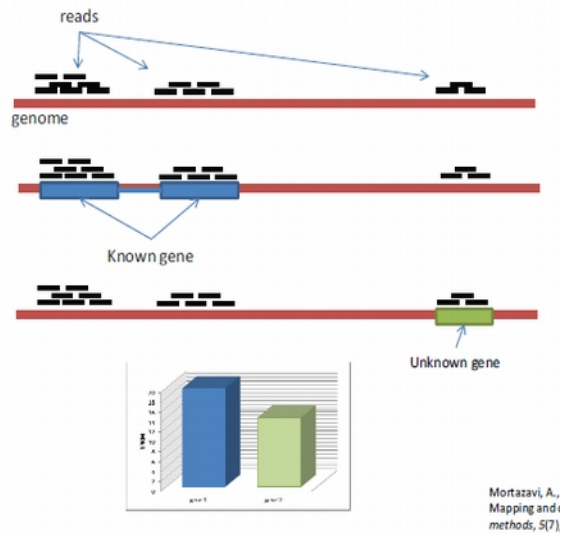
it is part of the Subread (http://subread.sourceforge.net) or Rsubread (R package)

Transcript quantification at gene or exon level featureCounts

a feature is an interval (range of positions) on one of the reference sequences

one approach is counting reads overlapping each annotated exon in a GTF:
can be used to test for alternative splicing between experimental conditions

another approach is counting reads at the gene level in a GTF:
all reads that overlap any exon for each gene



Mortazavi, A.,
Mapping and
methods, 5(7)

PRACTICAL - featureCounts

```
cd $HOME/tutorial
mkdir featureCounts_output

/home/studente/Scrivania/Elixir-RNA-Seq-Tools/subread-1.5.3-Linux-
x86_64/bin/featureCounts -t exon -g gene_id -a
/home/studente/Scrivania/Dataset_Corso/Danio_rerio.Zv9.66.gtf -o
featureCounts_output/counts.txt /home/studente/tutorial/tophat_out/2cells/accepted_hits.bam


awk '{print $1, $NF}' counts.txt | sort -gr -k 2 > 2cells_ordered_counts.txt
# awk prints only the gene id ($1: the first field of counts.txt file) and the counts ($NF: the last
field)
# the output of awk is piped to be ordered (by sort command) with respect to the number of the
reads (it is the second field in the output of awk)
# the output of sort is written into a file
```

→ Do the same for the 6h samples
(warning: create a new output directory, e.g.: featureCounts_6h_output)