Part II

3. Analysis

R

R is a statistical programming language

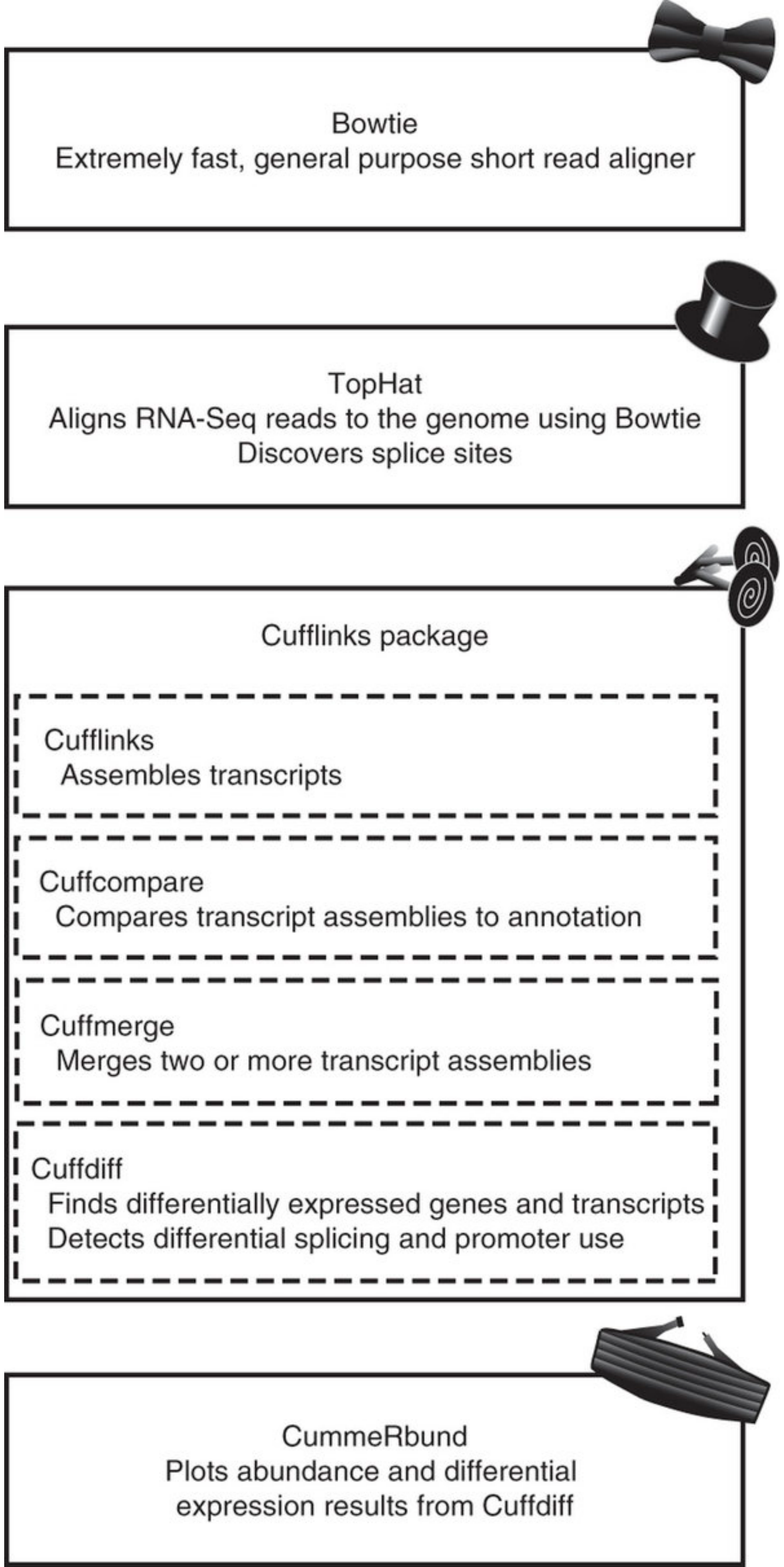
R Studio provides a GUI for R

Resources to help you learn and use R:

http://www.ats.ucla.edu/stat/r/

Cummerbund is part of the tuxedo pipeline

We will now use cummeRbund.R to analyze the output from cuffdiff from yesterday in RStudio



Bowtie
Extremely fast, general purpose short read aligner

TopHat
Aligns RNA-Seq reads to the genome using Bowtie
Discovers splice sites

Cufflinks package

Cufflinks
Assembles transcripts

Cuffcompare
Compares transcript assemblies to annotation

Cuffmerge
Merges two or more transcript assemblies

Cuffdiff
Finds differentially expressed genes and transcripts
Detects differential splicing and promoter use

CummeRbund
Plots abundance and differential expression results from Cuffdiff

edgeR

- download feature counts and unzip (http://bioinformatics.oxfordjournals.org/content/30/7/923.full.pdf?keytype=ref&ijkey=ZzPz96t2IqzAH6F):
  - http://sourceforge.net/projects/subread/files/subread-1.4.6-p5/subread-1.4.6-p5-Linux-x86_64.tar.gz

Pipeline

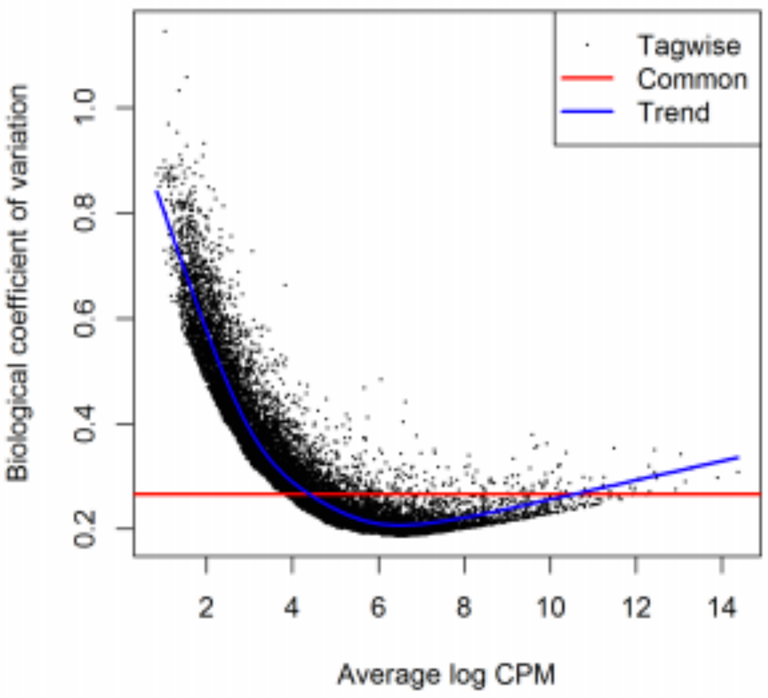
- correct gtf file (~/Desktop/ch4_demodataset/annotation):

```
sed 's/EXON/exon/g' ITAG2.3_gene_models_ch4.gtf > ITAG2.3_gene_models_ch4_ed.gtf
```

- run feature counts on each bam

```
  ~/Desktop/subread-1.4.6-p5-Linux-x86_64/bin/featureCounts -F GTF -O -s 1 -Q 10 -T 1 --primary
-p -P -d 50 -D 600 -C -a ~/Desktop/ch4_demo_dataset/annotation/ITAG2.3_gene_models_ch4_ed.gtf -o
counts_table.csv immature_fruit/SRR404331/SRR404331_ch4_thout/accepted_hits.bam immature_fruit/
SRR404333/SRR404333_ch4_thout/accepted_hits.bam breaker/SRR404334/SRR404334_ch4_thout/
accepted_hits.bam breaker/SRR404336/SRR404336_ch4_thout/accepted_hits.bam
```

- correct counts table:
    - change headers
    - remove first line
- run edge.R script on corrected feature table (edge.R) in RStudio
- compare de from cummerbund and edgeR
- plot results



Tagwise
Common
Trend
Biological coefficient of variation
0.2
0.4
0.6
0.8
1.0
2
4
6
8
10
12
14
Average log CPM