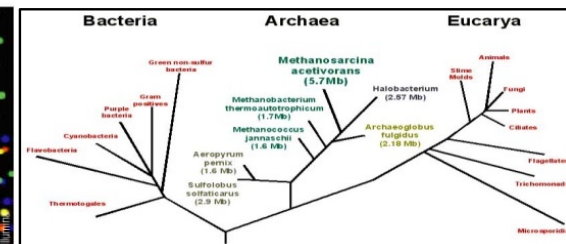
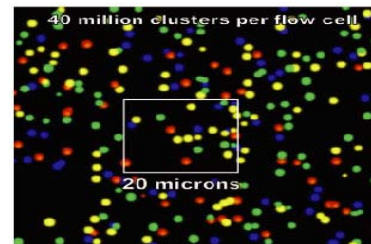




TAACCCTAACCCCTAACCCCTAACCCCTAACCCCTA
 CCTAACCCCTAACCCCTAACCCCTAACCCCTAACCC
 CCTAACCCCTAACCCCTAACCCCTAACCCCTAAC
 AACCCCTAACCCCTAACCCCTAACCCCTAACCCCTA
 ACCCTAACCCCAACCCCAACCCCAACCCCAAC
 CTACCCTAACCCCTAACCCCTAACCCCTAACCCCTA
 ACCCTAACCCCTAACCCCTAACCCCTAACCCCTAA

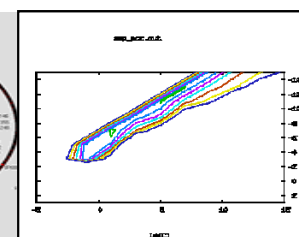
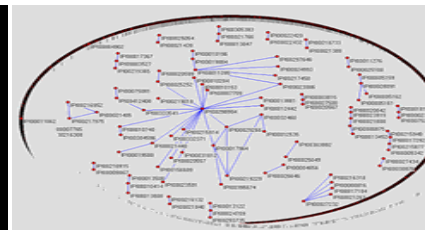
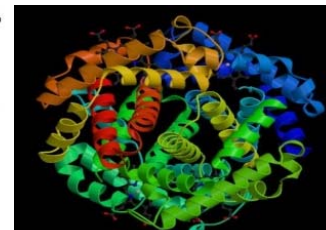
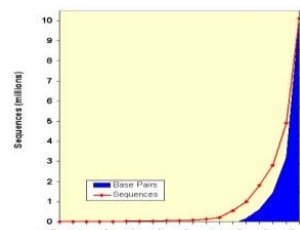
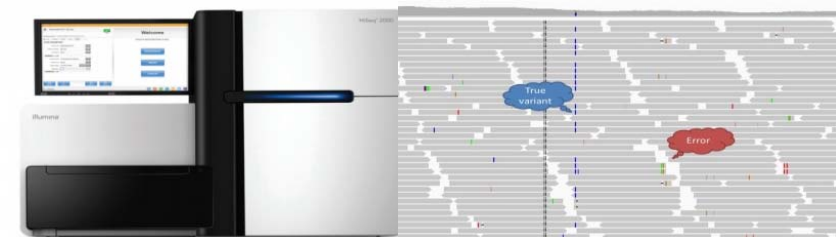


Computer Lab: TopHat-Cufflinks

侯玫 (Mei Hou) 田丰 (Feng Tian)

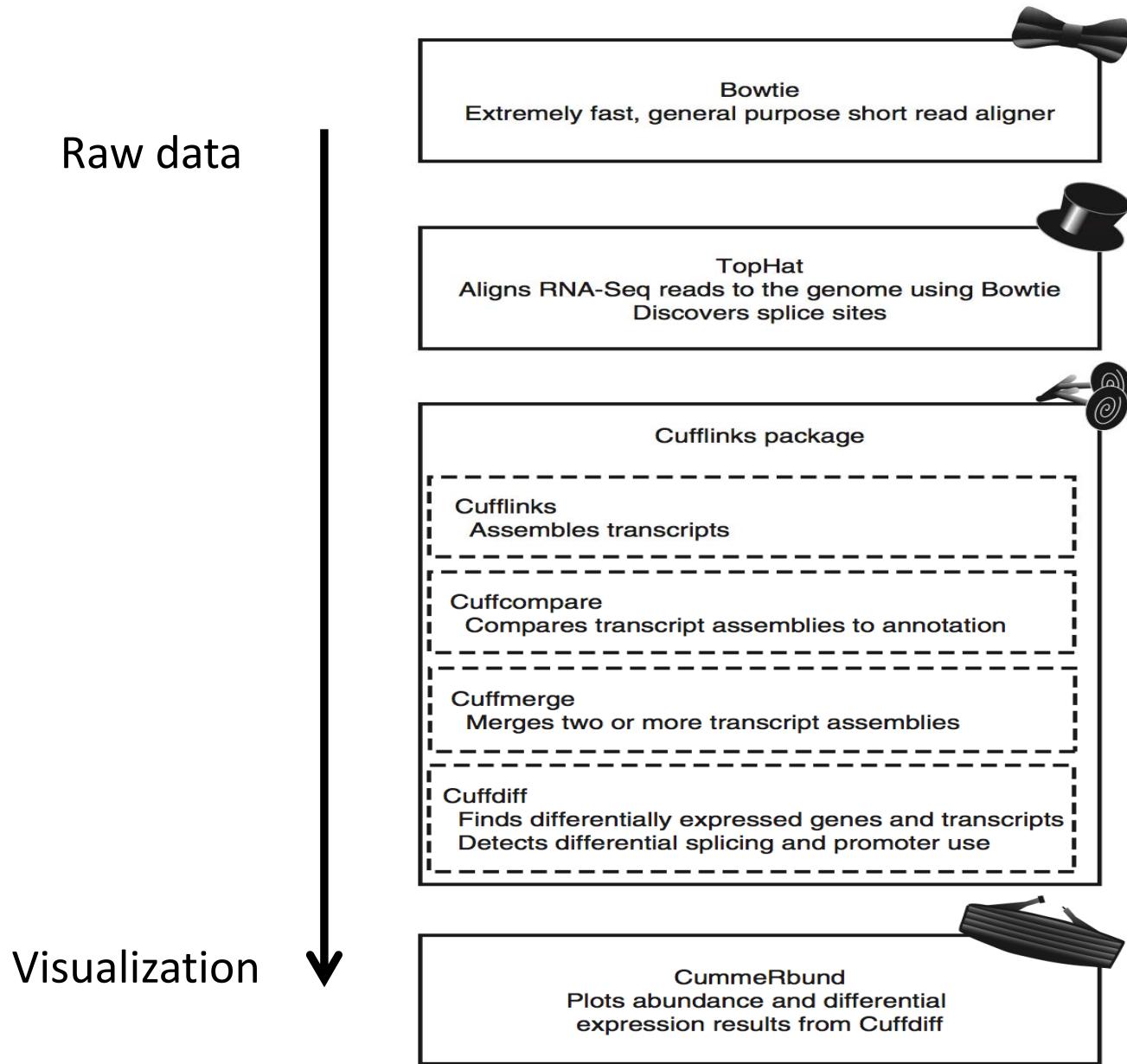
北京大学生物信息学中心

Center for Bioinformatics, Peking University



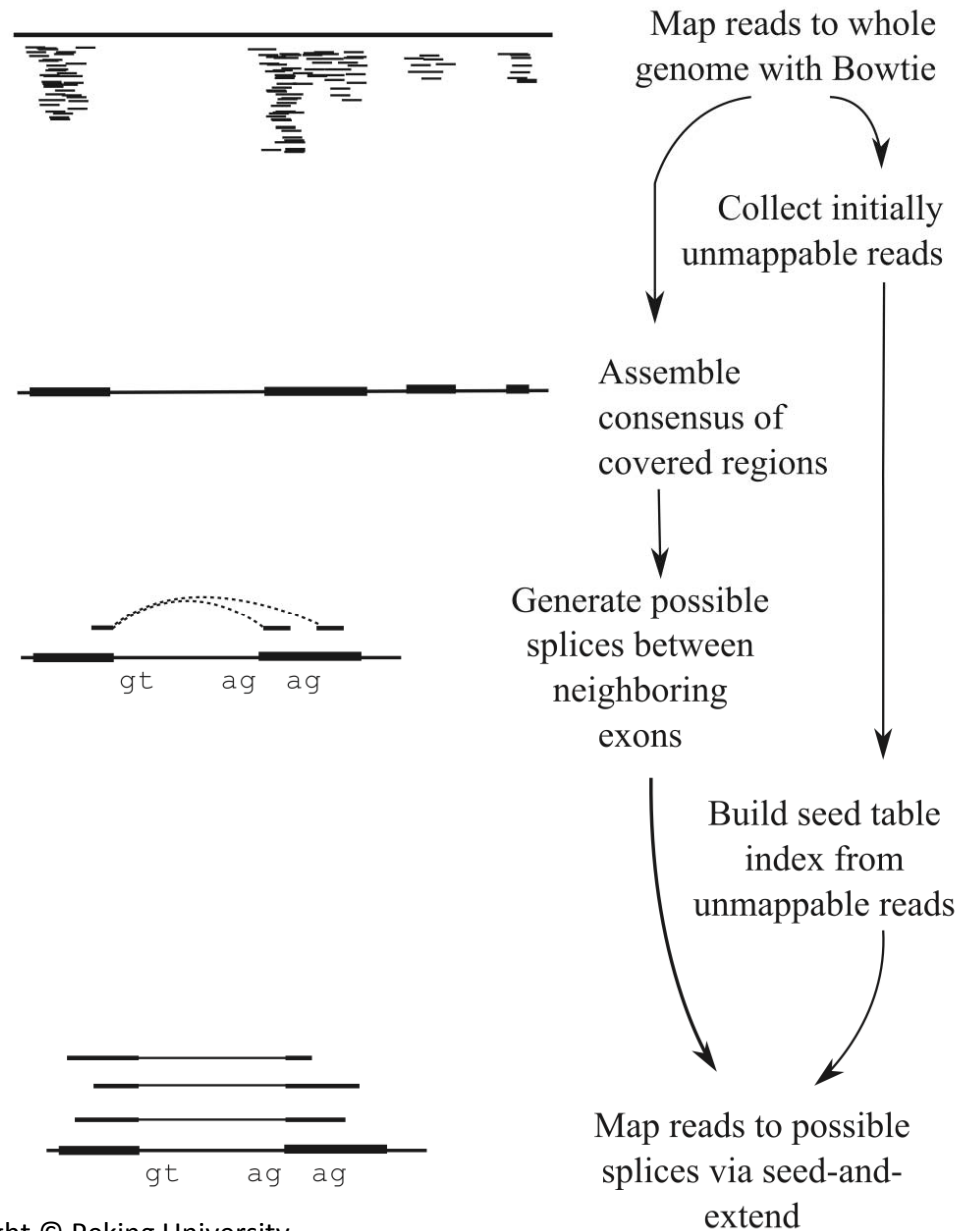
Outline

- Tophat
- Cufflinks
- Cuffmerge
- Cuffdiff
- CummeRbund



TopHat

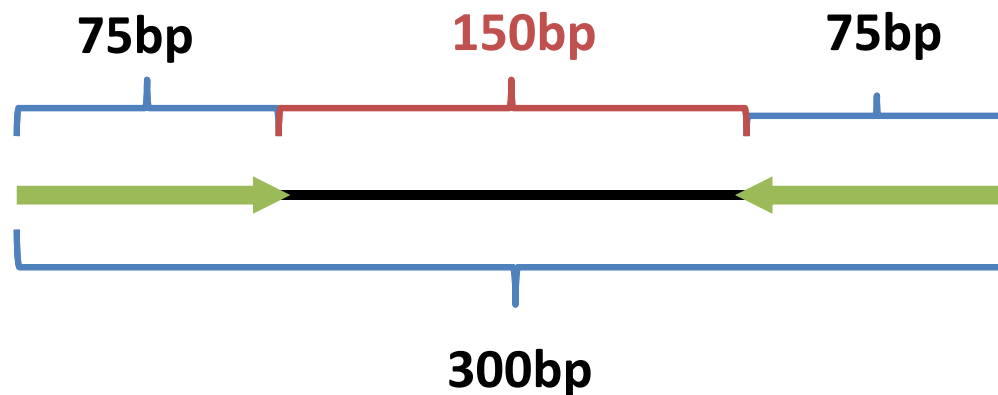
- A spliced read mapper for RNA-Seq.



Trapnell C. *et al. Bioinformatics* 2009

TopHat – Important options

- -r/--mate-inner-dist <int>
- --mate-std-dev <int>

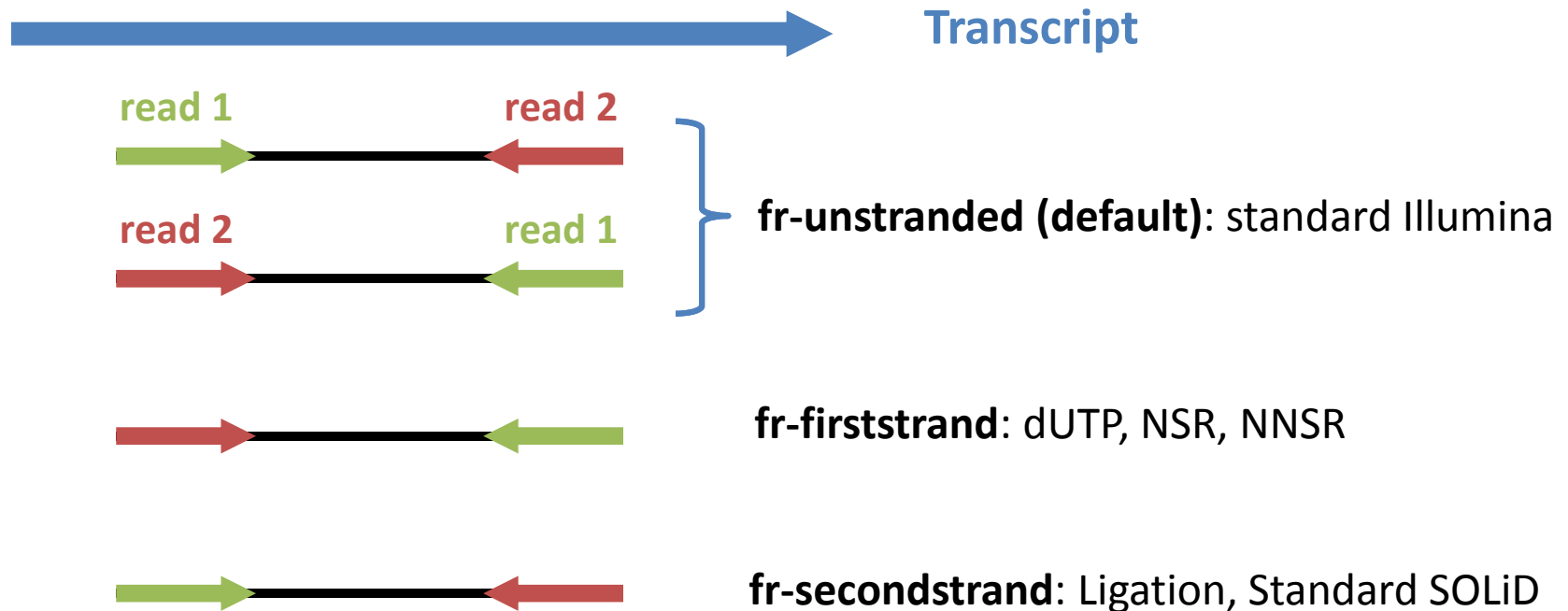


TopHat – Important options

- `-G/--GTF <GTF/GFF3 file>`
 - Provide an annotation file for more accurate and efficient mapping.

TopHat – Important options

- `--library-type`



TopHat – Demonstration

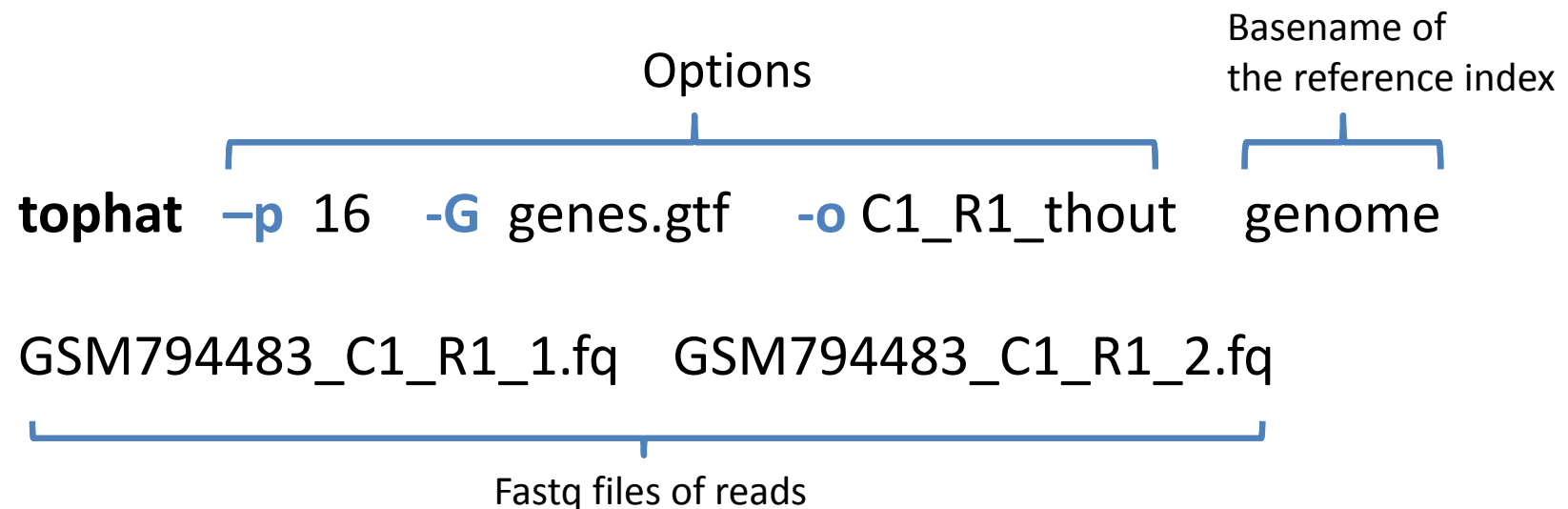
- The dataset used in this video
 - <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE32038>

tophat **-p** 16 **-G** genes.gtf **-o** C1_R1_thout **genome**

GSM794483_C1_R1_1.fq GSM794483_C1_R1_2.fq

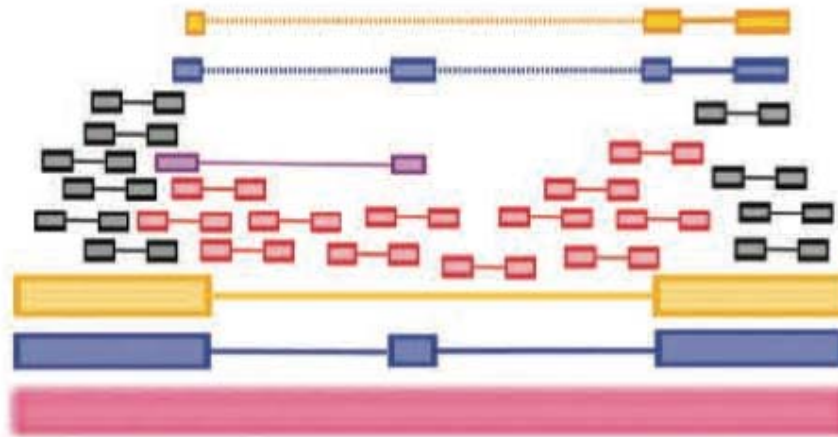
Options Basename of the reference index

Fastq files of reads



Cufflinks

- Transcript assembly, differential expression, and differential regulation for RNA-Seq



Cufflinks — Important options

- `-G/--GTF <reference_annotation.(gtf/gff)>`
 - Use the supplied annotation and not assemble novel transcripts.
- `-g/--GTF-guide <reference_annotation.(gtf/gff)>`
 - Use the supplied annotation to guide the assembly of novel transcripts.

Cufflinks — Important options

- `-u/--multi-read-correct`

Cufflinks — Important options

- `--library-type`
 - The same as TopHat.

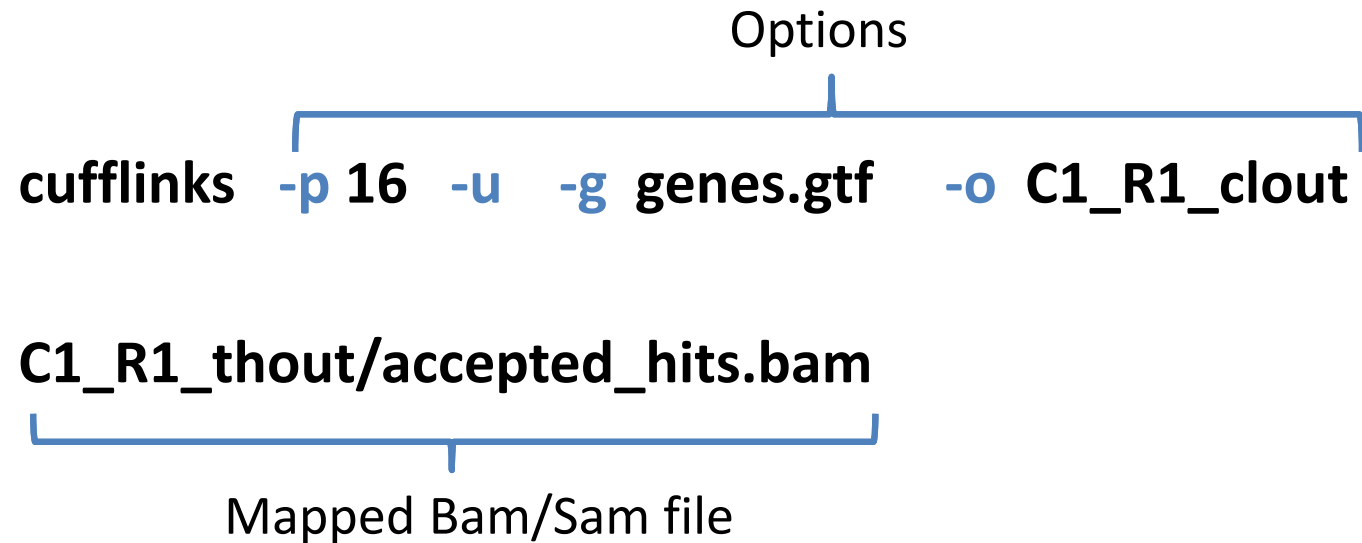
Cufflinks – Demonstration

Options

cufflinks -p 16 -u -g genes.gtf -o C1_R1_clout

C1_R1_thout/accepted_hits.bam

Mapped Bam/Sam file

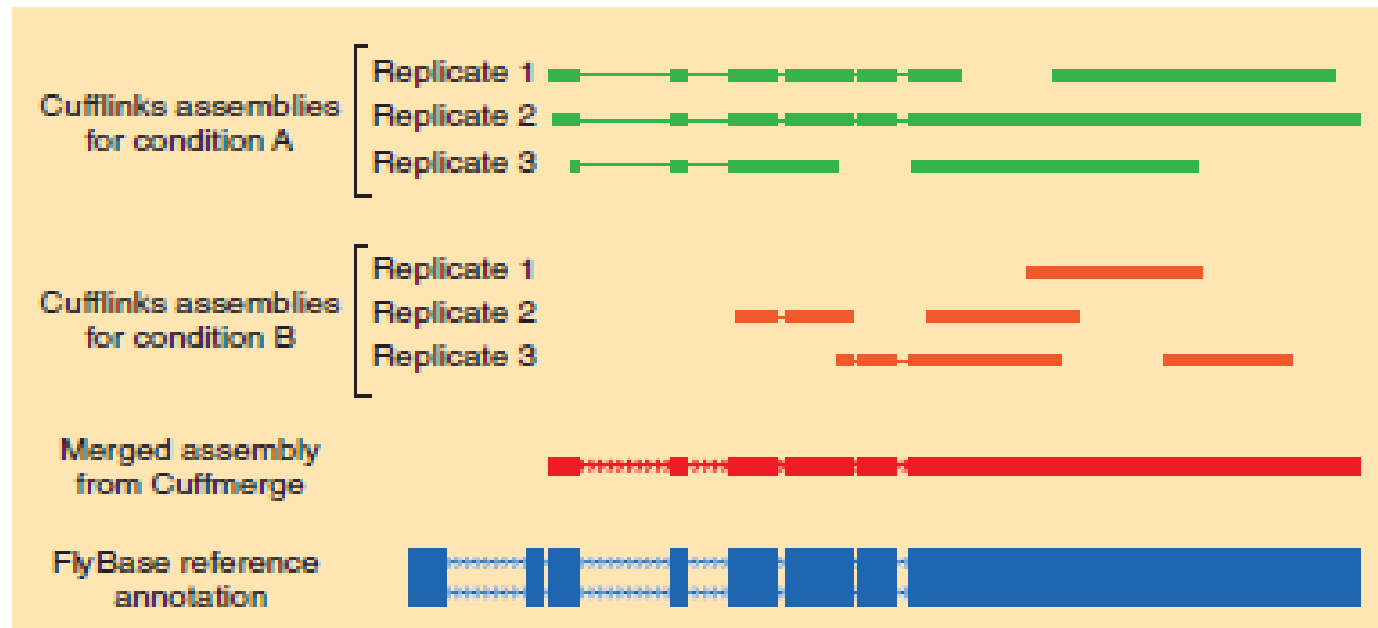


Outline

- Tophat
- Cufflinks
- Cuffmerge
- Cuffdiff
- CummeRbund

1. Cuffmerge

- merge together several Cufflinks assemblies.



Cole Trapnell. et al *Nat Protoc* 2012

Cuffmerge – Important Options

- -g / --ref-gtf # (optional)
- -p/--num-threads # default: 1
- -s / --ref-sequence <seq_dir>/<seq_fasta>

Cuffmerge – Demonstration

Options

cuffmerge -g genes.gtf -s genome.fa -p 16 -o cuffmerge_out

assemblies.txt

List of assembled transcripts file paths

2. Cuffdiff

- 1. calculates expression in two or more samples
 - **number of reads** of each transcript \propto its **abundance**
- 2. tests the statistical **significance** each observed change in expression between them

Cuffdiff – Important Options

- `-u/--multi-read-correct` # default: FALSE

(For multi-mapped read, which is described in Cufflinks)

- `-L/--labels`

- <http://cufflinks.cbc.umd.edu/manual.html#cuffdiff>

Cuffdiff – Demonstration

Options

- cuffdiff -o diff_out -p 8 -L C1,C2 -u merged_asm/merged.gtf \ → gtf file from Cuffmerge
./C1_R1_thout/accepted_hits.bam,./C1_R2_thout/accepted_hits.bam,./C1_R3_thout/accepted_hits.bam \
./C2_R1_thout/accepted_hits.bam,./C2_R3_thout/accepted_hits.bam,./C2_R2_thout/accepted_hits.bam

bam files from Tophat

3. CummeRbund

- An **R package** to aid and simplify the task of analyzing Cufflinks RNA-Seq output.

```
> source('http://www.bioconductor.org/biocLite.R')  
> biocLite('cummeRbund')
```

CummeRbund – Important Functions

- > csDensity(genes(cuff_data))
- > csScatter(genes(cuff_data), 'C1', 'C2')
- > csVolcano(genes(cuff_data), 'C1', 'C2')
- > expressionBarplot(mygene)
- > expressionBarplot(isoforms(mygene))

Reference

- Trapnell, Cole, Lior Pachter, and Steven L. Salzberg. "TopHat: discovering splice junctions with RNA-Seq." *Bioinformatics* 25.9 (2009): 1105-1111.
- Trapnell, Cole, et al. "Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation." *Nature biotechnology* 28.5 (2010): 511-515.
- Mortazavi, Ali, et al. "Mapping and quantifying mammalian transcriptomes by RNA-Seq." *Nature methods* 5.7 (2008): 621-628.
- Trapnell, Cole, et al. "Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks." *Nat Protoc.* 7.3 (2012): 562-578.
- <http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>
- <http://tophat.cbcb.umd.edu>
- <http://cufflinks.cbcb.umd.edu>
- <http://compbio.mit.edu/cummeRbund/>

Thank you for your attention



<https://www.coursera.org/course/pkubioinfo>