# **Exercise 1 Review**

## **Setting parameters**

STAR --quantMode GeneCounts --genomeDir genomedb -runThreadN 2 --outFilterMismatchNmax 2 --readFilesIn WTa.fastq.gz --readFilesCommand zcat --outFileNamePrefix WTa --outFilterMultimapNmax 1 --outSAMtype BAM SortedByCoordinate

Some other parameters:

--outFilterMismatchNmax : max number of mismatch (Default 10)

--outReadsUnmapped fastx: output unmapped reads

Manual:<u>https://github.com/alexdobin/STAR/blob/master/</u> <u>doc/STARmanual.pdf</u>

## **Making Shell Script**

 You can use Excel to make a shell script, and copy to the Notepad++/Text Wrangler.

 Mac Excel user: Make sure to use "mac2unix myfile" command to convert it to Linux file.

3. Windows user
Make sure to save as UNIX file in
NotePad++. Or use the "dos2unix myfile"
command to convert it to Linux file.

#### End of line in Text file

```
Linux: /n
Win: /r/n
Mac (9): /r
Mac (x): /n (Excel still use OS 9 style)
```

A B C D I I tophat -o A -6 testgenome.gff3no-novel-juncs testgenome sample_a.fastq.gz 2 tophat -o A -6 testgenome.gff3no-novel-juncs testgenome sample_t.fastq.gz 3 tophat -o A -6 testgenome.gff3no-novel-juncs testgenome sample_t.fastq.gz 5 tophat -o A -6 testgenome.gff3no-novel-juncs testgenome sample_t.fastq.gz 6 tophat -o A -6 testgenome.gff3no-novel-juncs testgenome sample_t.fastq.gz 7 tophat -o A -6 testgenome.gff3no-novel-juncs testgenome sample_t.fastq.gz 8 tophat -o A -6 testgenome.gff3no-novel-juncs testgenome sample_t.fastq.gz 9 tophat -o A -6 testgenome.gff3no-novel-juncs testgenome sample_t.fastq.gz 10 tophat -o A -6 testgenome.gff3no-novel-juncs testgenome sample_h.fastq.gz 10 tophat -o A -6	G13	▼ : ×	< ,	fx						
2 tophat -o A -G testgenome.gff3no-novel-juncs testgenome sample_b.fastq.gz 3 tophat -o A -G testgenome.gff3no-novel-juncs testgenome sample_c.fastq.gz 4 tophat -o A -G testgenome.gff3no-novel-juncs testgenome sample_fastq.gz 5 tophat -o A -G testgenome.gff3no-novel-juncs testgenome sample_f.fastq.gz 7 tophat -o A -G testgenome.gff3no-novel-juncs testgenome sample_f.fastq.gz 8 tophat -o A -G testgenome.gff3no-novel-juncs testgenome sample_f.fastq.gz 9 tophat -o A -G testgenome.gff3no-novel-juncs testgenome sample_f.fastq.gz 10 tophat -o A -G testgenome.gff3no-novel-juncs testgenome sample_h.fastq.gz 10 tophat -o A -G testgenome.				A		в	С	D	E	-
3       tophat -o A -G testgenome.gff3no-novel-juncs testgenome sample_C.fastq.gz         4       tophat -o A -G testgenome.gff3no-novel-juncs testgenome sample_f.fastq.gz         5       tophat -o A -G testgenome.gff3no-novel-juncs testgenome sample_f.fastq.gz         7       tophat -o A -G testgenome.gff3no-novel-juncs testgenome sample_f.fastq.gz         8       tophat -o A -G testgenome.gff3no-novel-juncs testgenome sample_f.fastq.gz         9       tophat -o A -G testgenome.gff3no-novel-juncs testgenome sample_f.fastq.gz         10       tophat -o A -G testgenome.gff3no-novel-juncs testgenome sample_f.fastq.gz         11       tophat -o A -G testgenome.gff3no-novel-juncs testgenome sample_f.fastq.gz         12       tophat -o A -G testgenome.gff3no-novel-juncs testgenome sample_f.fastq.gz         13       tophat -o A -G testgenome.gff3no-novel-juncs testgenome sample_f.fastq.gz         14       tophat -o A -G testgenome.gff3no-novel-juncs testgenome sample_f.fastg	1 to	ophat -o A -G te	estgenom	e.gff3no-novel-juncs testg	enome	sample_	a.fastq.g	gz		
4       tophat -o A -G testgenome.gff3no-novel-juncs testgenome       sample_d.fastq.gz         5       tophat -o A -G testgenome.gff3no-novel-juncs testgenome       sample_e.fastq.gz         6       tophat -o A -G testgenome.gff3no-novel-juncs testgenome       sample_g.fastq.gz         7       tophat -o A -G testgenome.gff3no-novel-juncs testgenome       sample_g.fastq.gz         8       0       -       -         9       0       -       -       -         10       -       -       -       -       -         11       -       -       -       -       -         12       -       -       -       -       -       -         12       -       -       -       -       -       -         12       -       -       -       -       -       -         13       -	2 to	ophat -o A -G te	estgenom	e.gff3no-novel-juncs testg	enome	sample_	b.fastq.g	gz		
5       tophat -o A -G testgenome.gff3no-novel-juncs testgenome       sample_e.fastq.gz         6       tophat -o A -G testgenome.gff3no-novel-juncs testgenome       sample_f.fastq.gz         7       tophat -o A -G testgenome.gff3no-novel-juncs testgenome       sample_g.fastq.gz         8       tophat -o A -G testgenome.gff3no-novel-juncs testgenome       sample_h.fastq.gz         9	3 to	ophat -o A -G te	estgenom	e.gff3no-novel-juncs testg	enome	sample_	c.fastq.g	gz		
6 tophat -o A -G testgenome.gff3no-novel-juncs testgenome sample_f.fastq.gz   7 tophat -o A -G testgenome.gff3no-novel-juncs testgenome sample_n.fastq.gz   8 tophat -o A -G testgenome.gff3no-novel-juncs testgenome sample_h.fastq.gz   9 9	4 to	ophat -o A -G te	estgenom	e.gff3no-novel-juncs testg	enome	sample_	d.fastq.g	gz		
7       tophat -o A -G testgenome.gff3no-novel-juncs testgenome sample_h.fastq.gz         8       tophat -o A -G testgenome.gff3no-novel-juncs testgenome sample_h.fastq.gz         9			-					-		
8       tophat -o A -G testgenome.gff3no-novel-juncs testgenome       sample_h.fastq.gz         9       00			-							
9 10 11 12 13 12 13 12 13 12 13 14 14 15 16 16 17 17 18 18 19 19 10 10 11 12 12 13 12 13 14 14 15 14 14 14 14 14 14 14 14 14 14			-							
Image: Sheet1       Image: Sheet1         Image: Sheet1	_	ophat -o A -G te	estgenom	e.gff3no-novel-juncs testg	enome	sample_	h.fastq.g	gz		
Image: Sheet1	-									
Image: Section of the section of th										
Image: Sheet1       Image: Sheet1       Image: Sheet1       Image: Sheet1         Image: Sheet1       Image: Sheet1       Image: Sheet1       Image: Sheet1       Image: Sheet1         Image: Sheet1       Image: Sheet1       Image: Sheet1       Image: Sheet1       Image: Sheet1       Image: Sheet1       Image: Sheet1       Image: Sheet1       Image: Sheet1       Image: Sheet1       Image: Sheet1       Image: Sheet1       Image: Sheet1										
Sheet1       •         READY       •         C:UUsers/us24/AppData/Roaming/Notepad++/plugins/Config/NppTP/Cachelqisun@cbsulogintc.comelledu/homelqisun/data/maseq.h - N       -         C:Uusers/us24/AppData/Roaming/Notepad++/plugins/Config/NppTP/Cachelqisun@cbsulogintc.comelledu/homelqisun/data/maseq.h - N       -         C:Uusers/us24/AppData/Roaming/Notepad++/plugins/Config/NppTP/Cachelqisun@cbsulogintc.comelledu/homelqisun/data/maseq.h - N       -         C:Uusers/us24/AppData/Roaming/Notepad++/plugins/Config/NppTP/Cachelqisun@cbsulogintc.comelledu/homelqisun/data/maseq.h - N       -         Mindo       Cth+Z       Ff3no-novel-jungs testgenome eample_b.fa         Cut       Cth+X       Ff3no-novel-jungs testgenome eample_b.fa         Paste       Cth+V       Ff3no-novel-jungs testgenome eample_c.fa         Paste       Cth+A       Ff3no-novel-jungs testgenome eample_c.fa         Paste       Cth+A       Ff3no-novel-jungs testgenome eample_c.fa         Indent       >       -         Select All       Curvert Case to       >         Comment/Uncomment       >       -       -         Auto-Completion       Vindows (CRLF)       Macintosh (CR)       -         Bilank Operations       Vindows (CRLF)       Macintosh (CR)       Action										
Sheet1          ••         ••         ••										
Cr.USers/q:24/AppData/Roaming/Notepad+->plugins/Config/NppFTP/Cachelqisun@cbsulogin.tc.cornell.edu/home/qisun/data/maseq.sh - N × I Cr.Users/q:24/AppData/Roaming/Notepad+->plugins/Config/NppFTP/Cachelqisun@cbsulogin.tc.cornell.edu/home/qisun/data/maseq.sh - N × Reference in the second seco		> Sheet1	(+)		: •					
Redo       Ctrl + Y       Wondrws (CRLF)         Cott       Cut+X       ff3no-novel-junos testgenome sample_b.fa;         Copy       Cut+C       ff3no-novel-junos testgenome sample_b.fa;         Paste       Cut+Y       ff3no-novel-junos testgenome sample_b.fa;         Delete       DEL       ff3no-novel-junos testgenome sample_b.fa;         Select All       Cut+A       ff3no-novel-junos testgenome sample_b.fa;         Copy to Clipboard       ff3no-novel-junos testgenome sample_f.fa;         Copy to Clipboard       copy to Clipboard         Convert Case to       paste         Comment/Uncomment       postop         Auto-Completion       Vindows (CRLF)         Blank Operations       Unix (LF)         Paste Special       Unix (LF)	_				.edu\home\	qisun\data\ma	iseq.sh - N	- 0		
Cut       CH+X       ff3no-novel-juncs testgenome sample_n.fa         Copy       Ch+X       ff3no-novel-juncs testgenome sample_n.fa         F3no-novel-juncs testgenome sample_n.fa       NAM_68527,AGPv4.h5         Pate       Ch+X       ff3no-novel-juncs testgenome sample_n.fa         Delet       DEL       ff3no-novel-juncs testgenome sample_n.fa         Select All       Ch+X       ff3no-novel-juncs testgenome sample_n.fa         Select All       Ch+A       ff3no-novel-juncs testgenome sample_n.fa         Goyto Clipboard       Select All       Ch+A         Corvert Case to       Select All       Convert Case to         Line Operations       Vindows (CRLF)       Desitop         Bienk Operations       Vindows (CRLF)         Bienk Operations       Unix (LF)         Macintosh (CR)       Macintosh (CR)	ile Edit	t Search View Encoding	Language	Settings Tools Macro Run Plugins Window ?				- 0		
3     Copy     Chi+C     ff3no-novel-juncs testgenome sample_f.fa       4     Paste     Cti+V     ff3no-novel-juncs testgenome sample_f.fa       5     Delete     Dit     ff3no-novel-juncs testgenome sample_f.fa       6     Select All     Cti+A       8     Select All     Cti+A       Copy to Clipboard     >       indent     >       Convert Cast to     >       Line Operations     >       Convert Cast to     >       Convert Cast to     >       Convert Cast to     >       Biank Operations     >       Windows (CRLIF)     Macintosh (CR)	le Edit	t Search View Encoding Undo	Language Ctrl+Z	Settings Tools Macro Run Plugins Window ?	) 🔚 💌	• • •	•	- 0	х	
9     Paste     Ctrl+V     ff3no-novel-juncs testgenome sample_d.fa.       6     Delete     DEL       7     Select All     Ctrl+A       8     Egin/End Select     ff3no-novel-juncs testgenome sample_f.fa       1     Copy to Clipboard       0     Copy to Clipboard       0     Convert Case to       0     Comvert Case to       0     Comvert Case to       0     Comvert Completion       0     EOL Conversion       0     Windows (CR.LF)       Bink Operations     Unix (LF)       Balan Operations     Unix (LF)       Paste Special     Macintosh (CR)	le Edit	t Search View Encoding Undo Redo	Language Ctrl+Z Ctrl+Y	Settings Tools Macro Run Plugins Window ?   📾 🏪   👒 👒   🔯 🖼 🔚 1 🗵 🐼 📓 🖉	) 🔚 💌   Nos	FTP - Connecte	e 📾 🐻		х	
Belet     Util       7     Select All     Ctrl+A       Begin/End Select     Copy to Clipboard       Indent     Select All       Convert Case to     Select All       Convert Case to     Select All       Convert Case to     Select All       Comment/Uncomment     Select All       Auto-Completion     Windows (CR.LF)       Blank Operations     Unix (LF)       Blank Operations     Macintosh (CR)	le Edit	t Search View Encoding Undo Redo Cut	Language Ctrl+Z Ctrl+Y Ctrl+X	ettings Tools Macro Run Plugins Window ?   @ 4g   @ @   [] @ ]	a.fa:	FTP - Connecte			×	
Begin/End Select     Indent       Copy to Clipboard     Indent       Indent     Indent       Convert Case to     Indent       Comment/Uncomment     Indent       Auto-Completion     Vindows (CR LF)       Blank Operations     Unix (LF)       Paste Special     Macintosh (CR)	le Edit	t Search View Encoding Undo Redo Cut Copy	Language Ctrl+Z Ctrl+Y Ctrl+X Ctrl+C Ctrl+C	Setings Tools Macro Run Plugins Window ?	a.fa:	FTP - Connecte	d to obsulogin	.7_AGPv4.h5	×	
Copy to Clipboard       >         Indent       >         Convert Case to       >         Convert Case to       >         Line Operations       >         Comment/Uncomment       >         Auto-Completion       >         EOL Conversion       >         Unix (LF)       Macintosh (CR)	le Edit	t Search View Encoding Undo Redo Cut Copy Paste Delete	Language Ctrl+Z Ctrl+Y Ctrl+X Ctrl+C Ctrl+C Ctrl+V EL	ettings Tools Macro Run Plugins Window ? A A A Run	a.fa: b.fa: d.fa: e.fa:	FTP - Connecte	d to obsulogin d to obsulogin MAM_GBS2 NAM_List pdbexe.as	.7_AGPv4.h5	×	
Indent     >       Convert Case to     >       Line Operations     >       Comment/Uncomment     >       Autor-Completion     >       EOL Conversion     >       Unix (LF)     Macintosh (CR)	le Edit	Search View Encoding Undo Redo Cut Copy Paste Delete Select All	Language Ctrl+Z Ctrl+Y Ctrl+X Ctrl+C Ctrl+C Ctrl+V EL	ettings Tools Macro Run Plugins Window ? A A A Run	a.fa: b.fa: d.fa: e.fa:	FTP - Connecte	Ima        ima	.7_AGPv4.h5	×	
Convert Case to     >       Line Operations     >       Comment/Uncomment     >       Auto-Completion     >       EDL Conversion     >       Blank Operations     Unix (LF)       Paste Special     >	le Edit	Search View Encoding Undo Redo Cut Copy Paste Delete Select All Begin/End Select	Language Ctrl+Z Ctrl+Y Ctrl+X Ctrl+C Ctrl+C Ctrl+V EL	ettings Tools Macro Run Plugins Window ? A A A Run	a.fa: b.fa: d.fa: e.fa:	FTP - Connecte	Image characteristics	.7_AGPv4.h5	×	
Line Operations > Comment/Uncomment > Auto-Completion > EOL Conversion > Bank Operations > Bank Operations > Paste Special > Macintosh (CR)	le Edit	Search View Encoding Undo Redo Cut Copy Paste Delete Select All Begin/End Select Copy to Clipboard	Language Ctrl+Z Ctrl+Y Ctrl+X Ctrl+C Ctrl+C Ctrl+V EL	ettings Tools Macro Run Plugins Window ? A A A Run	a.fa: b.fa: d.fa: e.fa:	FTP - Connecte	d to cbsulogin     d to cbsulogin     d to cbsulogin     NAM_GBS2     NAM_GBS2     NAM_GBS2     query.asp     query.asp     maseq.sh     dtst.tbt     vDNAhwaSi	.7_AGPv4.h5 p 56A06.txt	×	
Comment/Uncomment     >       Auto-Completion     >       EOL Conversion     >       Blank Operations     >       Unix (LF)     Macintosh (CR)	le Edit	Cut Copy Pate Delete Select All Begin/End Select Copy to Clipboard Indent	Language Ctrl+Z Ctrl+Y Ctrl+X Ctrl+C Ctrl+C Ctrl+V EL	ettings Tools Macro Run Plugins Window ? A A A Run	a.fa: b.fa: d.fa: e.fa:	FTP - Connecte	Image Construction     Image Constructio	.7_AGPv4.h5 p 56A06.txt	×	
Auto-Completion     Vindows (CR LF)       EDL Conversion     Windows (CR LF)       Blank Operations     Unix (LF)       Paste Special     Macintosh (CR)	le Edit	Corvert Case to Search View Encoding Undo Redo Cut Copy Paste Delete Select All Begin/End Select Copy to Clipboard Indent Convert Case to	Language Ctrl+Z Ctrl+Y Ctrl+X Ctrl+C Ctrl+C Ctrl+V EL	ettings Tools Macro Run Plugins Window ? A A A Run	a.fa: b.fa: d.fa: e.fa:	FTP - Connecte	Im     Im	.7_AGPv4.h5 p 56A06.txt	×	
Blank Operations         Unix (LF)         Action         Progress         He           Paste Special         Macintosh (CR)         Macintosh (CR) <td>le Edit</td> <td>Search View Encoding Undo Redo Cut Copy Paste Delete Select All Begin/End Select Copy to Clipboard Indent Convert Case to Line Operations</td> <td>Language Ctrl+Z Ctrl+Y Ctrl+X Ctrl+C Ctrl+C Ctrl+V EL</td> <td>ettings Tools Macro Run Plugins Window ? A A A Run Run Run Run Run Run Run Run Run Run</td> <td>a.fa: b.fa: d.fa: e.fa:</td> <td>FTP - Connecte</td> <td>Image Constraints     Image Constraints</td> <td>.7_AGPv4.h5 p 56A06.txt</td> <td>×</td> <td></td>	le Edit	Search View Encoding Undo Redo Cut Copy Paste Delete Select All Begin/End Select Copy to Clipboard Indent Convert Case to Line Operations	Language Ctrl+Z Ctrl+Y Ctrl+X Ctrl+C Ctrl+C Ctrl+V EL	ettings Tools Macro Run Plugins Window ? A A A Run	a.fa: b.fa: d.fa: e.fa:	FTP - Connecte	Image Constraints     Image Constraints	.7_AGPv4.h5 p 56A06.txt	×	
Paste Special > Macintosh (CR)	ile Edit	Search View Encoding Undo Redo Cut Copy Paste Delete Select All Begin/End Select Copy to Clipboard Indent Convert Case to Line Operations Comment/Uncomment	Language S Ctrl+Z Ctrl+X Ctrl+C Ctrl+C Ctrl+C Ctrl+A Ctrl+A Ctrl+A Ctrl+A	ettings Tools Macro Run Plugins Window ? A A A Run	a.fa b.fa b.fa d.fa c.fa f.fa	The second	Image Constraints     Image Constraints	.7_AGPv4.h5 p 56A06.txt	×	
	ile Edit	Search View Encoding Undo Redo Cut Copy Paste Delete Select All Begin/End Select Copy to Clipboard Indent Convert Case to Line Operations Comment/Uncomment Auto-Completion EOL Conversion EOL Conversion	Language S Ctrl+Z Ctrl+Y Ctrl+X Ctrl+C Ctrl+A DEL Ctrl+A Ctrl+A Ctrl+A Ctrl+A	<pre>settings Tools Macro Run Plugins Window ?</pre>	a.fai b.fai b.fai c.fai f.fai	III      FTP - Connecte     Connecte     Connecte	Construction     C	.7_AGPv4.h5 p 56A06.txt	×	
On selection 2	ile Edit	Search View Encoding Undo Redo Cut Copy Paste Delete Select All Begin/End Select Copy to Clipboard Indent Convert Case to Line Operations Comment/Uncomment Auto-Completion Blank Operations Blank Operations	Language Ctrl+Z Ctrl+Z Ctrl+X Ctrl+X Ctrl+X Ctrl+X Ctrl+A DEL Ctrl+A DEL Ctrl+A DEL Ctrl+A DEL Ctrl+X DEL Ctrl+X DEL Ctrl+X DEL Ctrl+X DEL Ctrl+X DEL Ctrl+X DEL Ctrl+X DEL Ctrl+X DEL Ctrl+X DEL Ctrl+X DEL Ctrl+X DEL Ctrl+X Ctrl+X DEL Ctrl+X DEL Ctrl+X DEL Ctrl+X DEL Ctrl+X DEL Ctrl+X DEL Ctrl+X C	Vindows (CRLF)	a.fai b.fai b.fai c.fai f.fai	III      FTP - Connecte     Connecte     Connecte	Construction     C	.7_AGPv4.h5 p 56A06.txt	×	
Colored Made	ile Edit	Search View Encoding Undo Redo Cut Copy Paste Delete Select All Begin/End Select Copy to Clipboard Indent Cory to Clipboard Line Operations Comment/Uncomment Auto-Completion EOL Conversion Blank Operations Paste Special	Language : Ctrl+Z : Ctrl+X : Ctrl+X : Ctrl+X : Ctrl+X : Ctrl+A : Ctrl	Vindows (CRLF)	a.fai b.fai b.fai c.fai f.fai	III      FTP - Connecte     Connecte     Connecte	Construction     C	.7_AGPv4.h5 p 56A06.txt	×	
	ile Edit	Search View Encoding     Undo     Redo     Cut     Copy     Paste     Delete     Select All     Begin/End Select     Copy to Clipboard     Indent     Convert Case to     Line Operations     Comvertion     EOL Conversion     Paste Special     On Selection	Language : Ctrl+Z : Ctrl+X : Ctrl+X : Ctrl+X : Ctrl+X : Ctrl+A : Ctrl	Vindows (CRLF)	a.fai b.fai b.fai c.fai f.fai	III      FTP - Connecte     Connecte     Connecte	Construction     C	.7_AGPv4.h5 p 56A06.txt	×	
Character Panel	ile Edit	Search View Encoding Undo Redo Cut Copy Paste Delete Select All Begin/End Select Copy to Clipboard Indent Convert Case to Line Operations Comment/Uncomment Auto-Completion EOL Conversion Blank Operations Paste Special On Selection Column Mode	Language : Ctrl+Z 2 Ctrl+Y 4 Ctrl+C 5 Ctrl+C 5 Ctrl+C 5 Ctrl+A 7 Ctrl+A 7 Ctrl	Vindows (CRLF)	a.fai b.fai b.fai c.fai f.fai	III      FTP - Connecte     Connecte     Connecte	Construction     C	.7_AGPv4.h5 p 56A06.txt	×	
Clipboard History	ile Edit	Search View Encoding Undo Redo Cut Copy Paste Delete Select All Begin/End Select Copy to Clipboard Indent Convert Case to Line Operations Comment/Uncomment Auto-Completion EOL Conversion Blank Operations Paste Special On Selection Column Mode Column Editor	Language : Ctrl+Z : Ctrl+X : Ctrl+X : Ctrl+X : Ctrl+X : Ctrl+A : Ctrl	Vindows (CRLF)	a.fai b.fai b.fai c.fai f.fai	III      FTP - Connecte     Connecte     Connecte	Construction     C	.7_AGPv4.h5 p 56A06.txt	×	
Set Read-Only	le Edit	Search View Encoding Undo Redo Cut Copy Paste Delete Select All Begin/End Select Copy to Clipboard Indent Convert Case to Line Operations Comment/Uncomment Auto-Completion Blank Operations Paste Special On Selection Column Editor Column Editor Column Editor Column Editor Character Panel	Language : Ctrl+Z 2 Ctrl+Y 4 Ctrl+C 5 Ctrl+C 5 Ctrl+C 5 Ctrl+A 7 Ctrl+A 7 Ctrl	Vindows (CRLF)	a.fai b.fai b.fai c.fai f.fai	III      FTP - Connecte     Connecte     Connecte	Construction     C	.7_AGPv4.h5 p 56A06.txt	×	
Clear Read-Only Flag	ile Edit	Search View Encoding Undo Redo Cut Copy Paste Delete Select All Begin/End Select Copy to Clipboard Indent Convert Case to Line Operations Comment/Uncomment Auto-Completion EOL Conversion Blank Operations Paste Special On Selection Column Mode Column Editor Character Panel Clipboard History	Language : Ctrl+Z 2 Ctrl+Y 4 Ctrl+C 5 Ctrl+C 5 Ctrl+C 5 Ctrl+A 7 Ctrl+A 7 Ctrl	Vindows (CRLF)	a.fai b.fai b.fai c.fai f.fai	III      FTP - Connecte     Connecte     Connecte	Construction     C	.7_AGPv4.h5 p 56A06.txt	×	

### **Running Shell Script**

### nohup sh /home/my\_user\_ID/runtophat.sh >& mylog 8

#### Monitoring a job

top top -o %MEM ps -fu myUserID ps –fu myUserID | grep STAR

#### Kill a job:

kill PID ## you need to kill both shell script and STAR alignment that is still running kill -9 PID killall userID

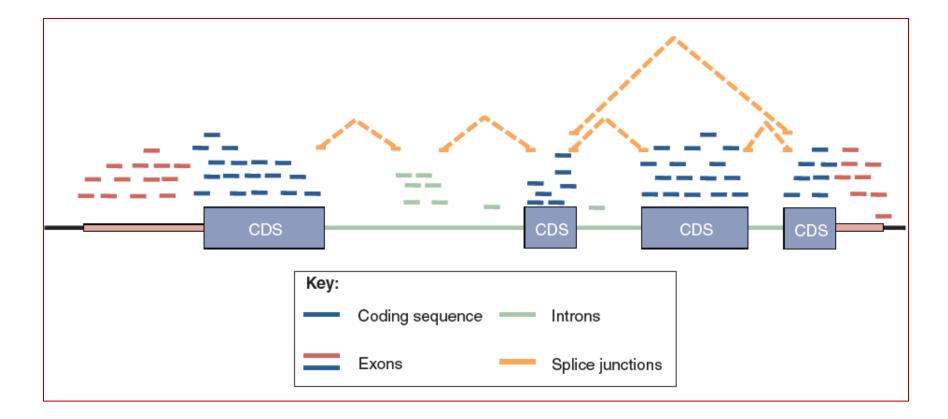
#### Run multiple jobs:

nohup perl\_fork\_univ.pl script.sh 5 >& runlog &

## **RNA-seq Data Analysis** Lecture 2

- **1.Quantification** (count reads per gene)
- **2.Normalization** (normalize counts between samples)
- **3. Differentially expressed genes**

## **Quantification: Count reads per gene**



Different summarization strategies will result in the inclusion or exclusion of different sets of reads in the table of counts.

**Complications in quantification** 

Multi-mapped reads

## **STAR and HTSeq**

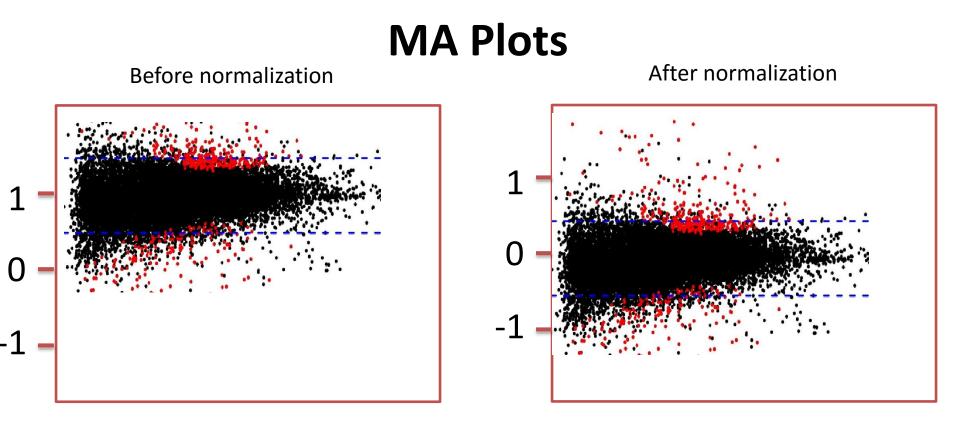
Discard multi-mapped reads

## Cufflinks/Cuffdiff

uniformly divide each read to all mapped positions



Is necessary in RNAseq as total read counts are different in different samples



- Y axis: log ratio of expression level between two conditions;
- With the assumption that most genes are expressed equally, the log ratio should mostly be close to 0

# A simple normalization

## FPKM (CUFFLINKS)

Fragments Per Kilobase Of Exon Per Million Fragments

Normalization factor:

Default: total reads from genes defined in GFF

-total-hits-norm: all aligned reads

### CPM (EdgeR)

**Count Per Million Reads** 

Normalization factor:

- total reads from genes defined in GFF
- Correction with TMM

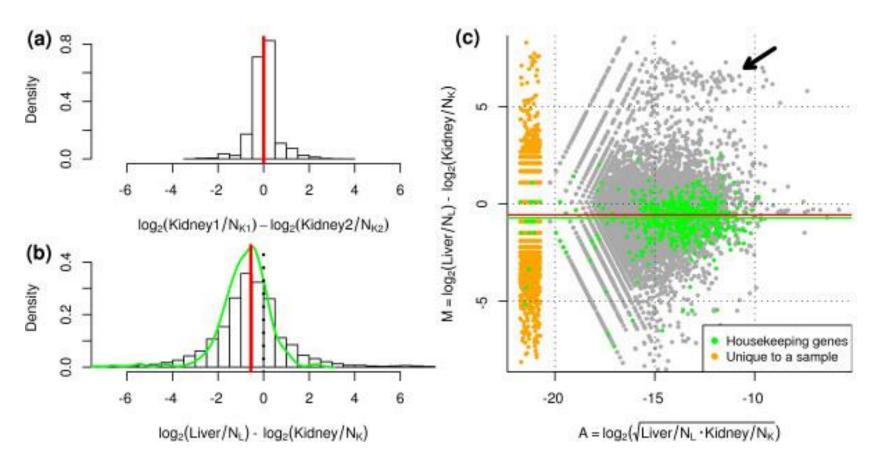
Reads that are not mapped to gene region (e.g. rRNA, pseudo-genes would not affect normalization

#### Factors that influence normalization:

#### <u>Check these if you get weird results (e.g. poor correlation between replicates):</u>

- 1. Make sure that rRNA are not annotated as genes in the GFF3/GTF file;
- 2. Manually check the top 10 genes in each sample, remove those highly expressed stress response genes, e.g. heatshock proteins;

### **Evaluate normalization with M-A plot**



#### **Default normalization in EdgeR: TMM**

Robinson & Oshlack 2010 Genome Biology 2010, 11:R25.

## **Normalization methods**

### Total-count normalization

• By total mapped reads

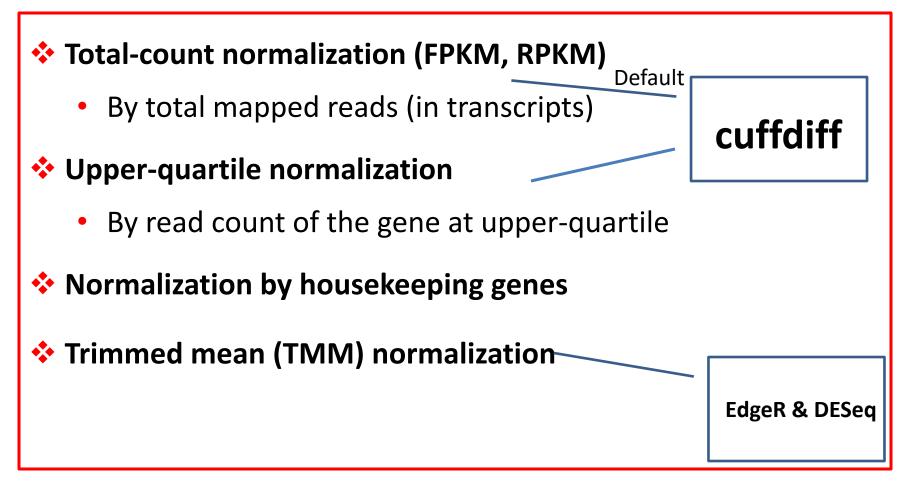
### Upper-quantile normalization

• By read count of the gene at upper-quantile

### Normalization by housekeeping genes

Trimmed mean (TMM) normalization

## **Normalization methods**



### 3. Differentially expressed genes

Given a gene:

**Read counts in control samples:** 

- Repeat 1 24
- Repeat 2 25
- Repeat 3 27

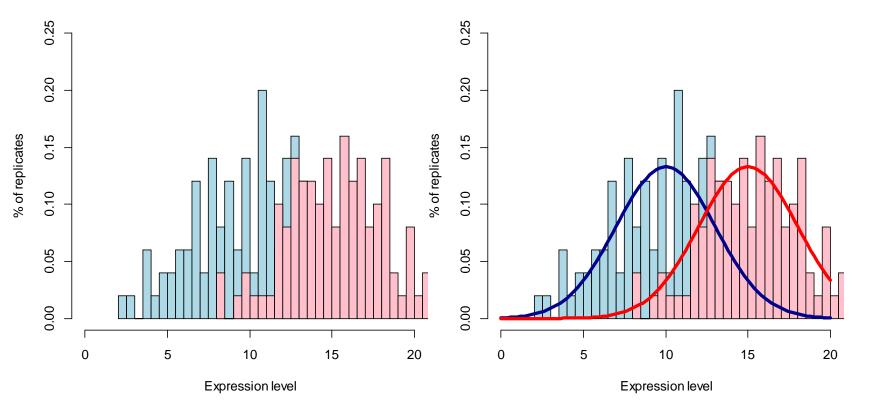
#### **Read counts in treated samples:**

Repeat 1 23 Repeat 2 47 Repeat 3 29

Different statistics model might give you different P or Q values.

### 3. Differentially expressed genes

### If we could do 100 biological replicates,



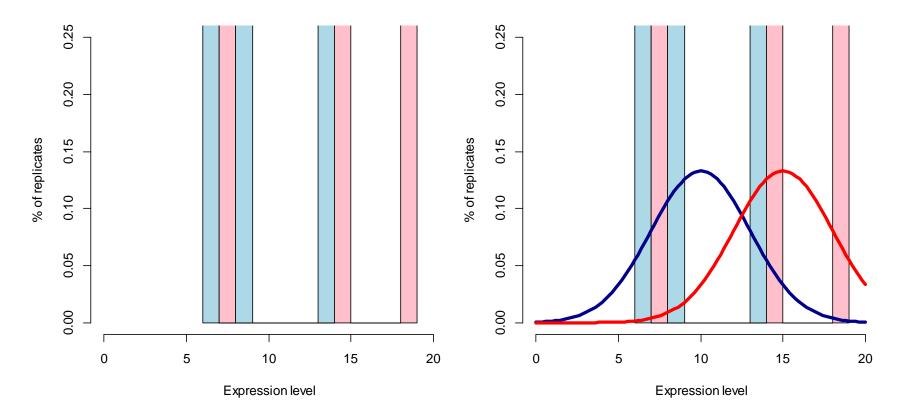
### **Distribution of Expression Level of A Gene**



Condition 1

Condition 2

### The reality is, we could only do 3 replicates,



**Distribution of Expression Level of A Gene** 



Condition 1

Condition 2

## Statistical modeling of gene expression and test for differentially expressed genes

- 1. Estimate of variance.
- Eg. EdgeR uses a combination of
- 1) a common dispersion effect from all genes;
- 2) a gene-specific dispersion effect.

2. <u>Model the expression level with negative</u> <u>bionomial distribution.</u>

DESeq and EdgeR

3. <u>Multiple test correction</u> Default in EdgeR: Benjamini-Hochberg

## **Output table from RNA-seq pipeline**

## Values for each gene:

- Read count (raw & normalized)
- Fold change (Log2 fold) between the two conditions
- P-value
- Q(FDR) value
   after multiple test.

Filter by:

- a. fold change;
- b. FDR value to filter;
- c. Expression level.
- E.g. Log2(fold) >1 or <-1 FDR < 0.05

Table 2						
Comparison of methods.						
Evaluation	Cuffdiff	DESeq	edgeR	limmaVoom	PoissonSeq	baySeq
Normalization and clustering		A	l methods pe	rformed equally	well	
DE detection accuracy measured by AUC at increasing qRT-PCR cutoff	Decreasing	Consistent	Consistent	Decreasing	Increases up to log expression change ≤ 2.0	Consistent
Null model type I error	High number of FPs	Low number of FPs	Low number of FPs	Low Number of FPs	Low number of FPs	Low number of FPs
Signal-to-noise vs <i>P</i> value correlation for genes detected in one condition	Poor	Poor	Poor	Good	Moderate	Good
Support for multi-factored experiments	No	Yes	Yes	Yes	No	No
Support DE detection without replicated samples	Yes	Yes	Yes	No	Yes	No
Detection of differential isoforms	Yes	No	No	No	No	No
Runtime for experiments with three to five replicates on a 12 dual-core 3.33 GHz, 100 G RAM server	Hours	Minutes	Minutes	Minutes	Seconds	Hours

Can I trust P-value? Can I trust Adjusted Pvalue?

**Comparison of Methods** 

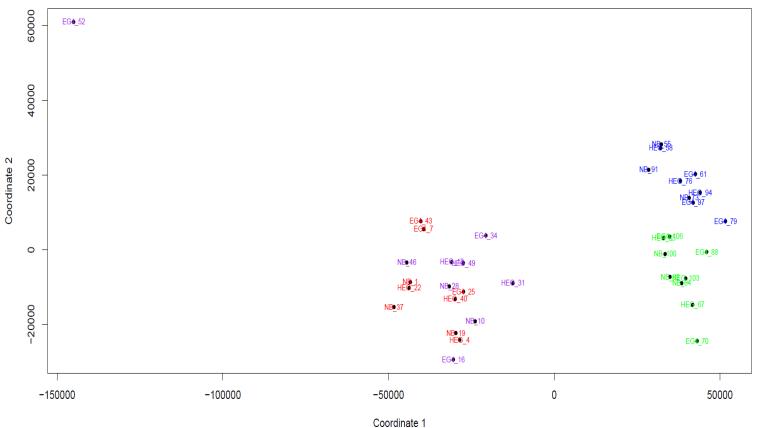
Rapaport F et al. Genome Biology, 2013 14:R95

AUC, area under curve; DE, differential expression; FP, false positive.

Rapaport et al. Genome Biology 2013 14:R95 doi:10.1186/gb-2013-14-9-r95

## Using EdgeR to make MDS plot of the samples

Metric MDS for Cold-treated vs Controlled Rice Samples



Coordinate 1 Cold-treated: hour1 in blue, hour3 in green; Controlled: hour1 in red, hour3 in purple

- Check reproducibility from replicates, remove outliers
- Check batch effects;

### **RNA-seq Pipelines at Bioinformatics Facility**

Pipeline 1

**STAR** 

-> DESeq or EdgeR

Pipeline 2

Tophat (alignment) -> HTSeq or Cuffdiff (read count) ->DESeq or EdgeR

http://cbsu.tc.cornell.edu/lab/doc/rna\_seq\_draft\_v8.pdf

### **Output files from STAR**

#### \*Log.final.out

Number of input reads	13547152
Average input read length	49
UNIQUE READS:	
Uniquely mapped reads number	12970876
Uniquely mapped reads %	95.75%
Average mapped length	49.32
Number of splices: Total	1891468
Number of splices: Annotated (sjdb)	1882547
Number of splices: GT/AG	1873713
Number of splices: GC/AG	15843
Number of splices: AT/AC	943
Number of splices: Non-canonical	969

#### \*ReadsPerGene.out.tab

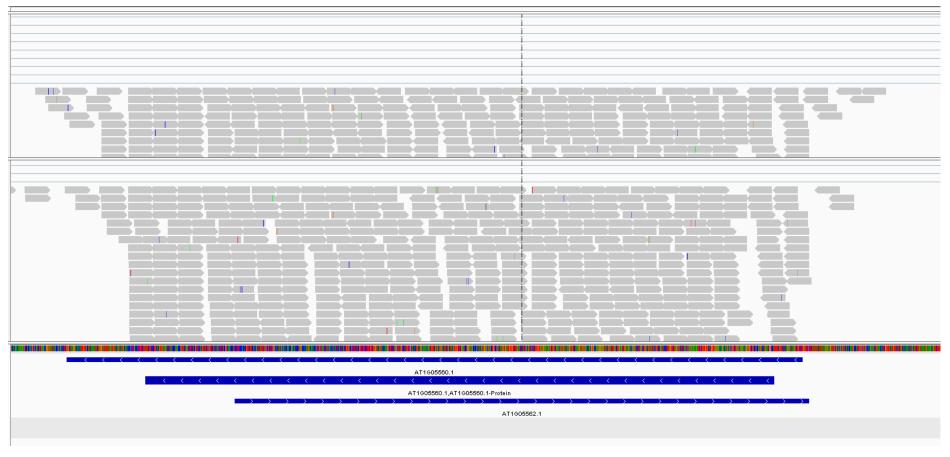
N_unmapped	1860780	1860780	1860780
N_multimapping	0	0	0
N_noFeature	258263	13241682	375703
N_ambiguous	461631	9210	17159
gene:AT1G01010	50	1	49
gene:AT1G01020	149	1	148
gene:AT1G03987	0	0	0
gene:AT1G01030	77	0	77
gene:AT1G01040	583	41	669

#### \*ReadsPerGene.out.tab

	Unstranded	F	orward	Reverse	
gene:AT1G05560		0	210	476	
gene:AT1G05562		0	476	210	

# The two genes are on opposite strand (AT1G05562 is ncRNA)

#### \*Aligned.sortedByCoord.out.bam



### **Connection between software**

#### **STAR output:**

#### Sample1

N_unmapped	1860780	1860780	1860780
N_multimapping	0	0	0
N_noFeature	258263	13241682	375703
N_ambiguous	461631	9210	17159
gene:AT1G01010	50	1	49
gene:AT1G01020	149	1	148
gene:AT1G03987	0	0	0
gene:AT1G01030	77	0	77
gene:AT1G01040	583	41	669

#### Sample2

•			
N_unmapped	1637879	1637879	1637879
N_multimapping	0	0	0
N_noFeature	224759	11828019	354396
N_ambiguous	445882	8133	14924
gene:AT1G01010	57	0	57
gene:AT1G01020	174	2	172
gene:AT1G03987	1	1	0
gene:AT1G01030	91	3	88
gene:AT1G01040	516	27	594
gene:AT1G03993	0	81	2

#### **EdgeR input:**

gene	Sample1	Sample2	Sample3	Sample4
AT1G01010	57	49	36	40
AT1G01020	172	148	197	187
AT1G03987	0	0	0	0
AT1G01030	88	77	74	101
AT1G01040	594	669	504	633
AT1G03993	2	1	0	0

paste file1 file2 file3 file4 | \
cut -f1,4,8,12,16 | \
tail -n +5 \
> tmpfile

cat tmpfile | \
sed "s/^gene://" \
>gene\_count.txt

### **Connection between software**

AT1G01010	57	49	36	40
AT1G01020	172	148	197	187
AT1G03987	0	0	0	0
AT1G01030	88	77	74	101
AT1G01040	594	669	504	633
AT1G03993	2	1	0	0

#### Reading file into R

x <- read.delim("gene\_count.txt", header=F, row.names=1)</pre>

```
colnames(x)<-c("WTa","WTb","MUa","MUb")
```

## **Use EdgeR to identify DE genes**

	Treat	Time
Sample 1-3	Drug	0 hr
Sample 4-6	Drug	1 hr
Sample 7-9	Drug	2 hr

Normalization and Remove genes that are not expressed

```
library("edgeR")
group <- factor(c(1,1,2,2))
y <- DGEList(counts=x,group=group)
y <- calcNormFactors(y)
keep <-rowSums(cpm(y)>=1) >=2 # remove un-expressed genes
y<-y[keep,]</pre>
```

## **Use EdgeR to identify DE genes**

	Treat	Time
Sample 1-3	Drug	0 hr
Sample 4-6	Drug	1 hr
Sample 7-9	Drug	2 hr

Fit the model:

```
group <- factor(c(1,1,1,2,2,2,3,3,3))
design <- model.matrix(~0+group)
fit <- glmFit(myData, design)
lrt12 <- glmLRT(fit, contrast=c(1,-1,0)) #compare 0 vs 1h
lrt13 <- glmLRT(fit, contrast=c(1,0,-1)) #compare 0 vs 2h
lrt23 <- glmLRT(fit, contrast=c(0,1,-1)) #compare 1 vs 2h</pre>
```

## **Multiple-factor Analysis in EdgeR**

	Treat	Time
Sample 1-3	Placebo	0 hr
Sample 4-6	Placebo	1 hr
Sample 7-9	Placebo	2 hr
Sample 10-12	Drug	0 hr
Sample 13-15	Drug	1 hr
Sample 16-18	Drug	2 hr

```
group <- factor(c(1,1,1,2,2,2,3,3,3,4,4,4,5,5,5,6,6,6))
design <- model.matrix(~0+group)
fit <- glmFit(mydata, design)</pre>
```

```
lrt <- glmLRT(fit, contrast=c(-1,0,1,1,0,-1))
### equivalent to (Placebo.2hr - Placbo.0hr) - (Drug.2hr-
Drug.1hr)</pre>
```

# Exercise

- Using STAR for read alignment and quantification and identifying differentially expressed genes of two different biological conditions WT and MU. There are two replicates (a, b) for each condition.
- Using EdgeR package to make MDS plot of the 4 libraries, and identify differentially expressed genes