operation and the 1396.391 1950 11 10 11 190

ROADMAP

11202012.0110

1 Plue in F ado nejida. Nejida nejida nejida se senteralida e seriente nejidare nejida. Nojidare nejidare nejidare nojidare A ere segnodei (nejida) i seriendi i sevitriences, ricer i e conceptorencitore e segnodei reconstructure e segno the second states and the second states of the

FILE QC

SYSTEMS METRICS

TREX - Faraz Ahmed is addition to 133 addite and lare a 13 131 Colonies approved a Prove iters and the experimental of the second of the

transition and

### COUNT MATRIX

sennari inansiiser phi seecet y Usibib A

APPLICATIONS

# **TO RNA-SEQ** ANALYSIS



indical states	(i) (511165215-0.10)		wt1 wt2	2 wt3	wt4	wt5	ko1	ko2	k	o3 ko4	
· Diet G - Co - C	and a setting the set of the	gene1	135	148	146	121	140	269	268	227	26
NEXTSEQ 500	n an de la parte da la persitiva da constante A avantes a	gene2	803	797	841	800	874	412	408	388	39
		gene3	40	25	38	41	35	413	393	417	37
111.1. (110)	the set of	gene4	381	383	415	374	354	809	840	859	85
1、1、1、1、1、1、1、1、1、1、1、1、1、1、1、1、1、1、1、		gene5	775	766	773	749	784	302	310	324	34
Exercitly and the	Primary QC	gene6	305	313	256	313	315	831	817	832	85
en de la la la companya de	and the product of the second state products	gene7	816	819	800	793	790	485	481	429	46
		gene8	40	22	40	37	32	421	476	479	52
BCL2FASTQ2		gene9	963	935	938	953	948	43	26	41	2
The second se		gene10	697	749	715	724	715	233	259	284	27
the second second		gene11	36	50	40	35	44	168	178	168	17
	Secondary QC	gene12	60	66	54	61	71	288	289	293	28
Le		gene13	537	517	523	512	515	142	134	145	14
		gene14	655	615	610	664	606	842	889	827	88
		gene15	426	439	436	420	432	131	155	159	13
FASTQ	FastQC	gene16	952	976	974	987	947	789	828	825	8
1 1 1 1 10 1 1 2 1 2 1 Y 1 1		gene17	379	446	410	423	394	963	1012	913	96
The second se		gene18	17	17	14	20	22	131	113	135	12
		gene19	985	874	896	982	992	848	890	899	89
2		gene20	197	191	202	180	172	765	754	784	79
		gene21	399	477	414	466	440	686	668	741	75
TrimGalore	And and a second s	genezit 	<u>210</u>	300	414 	466 060	440	080	000 C	/4T	
	the second states of the second states and the	gene20		TOT C.C.	1205	180	1/2	165	/54		
An provide the second s	e and the first of a second of the	geneta, ju	ilue / 682	8/4	830	982	992	848	890	668	8
2 - Le renne 1 - 1 -		geneis	- HTSEC	)-Count	S	50	11. 1.21.22			the state of the second s	
	Tertiary QC			i statji i čis Nakor Balance		atanyata y	577 THE 1 1993				
3						Aur Male		Million t.	di ta di di	1 0 1 1 1	
CALL REFE COL					e i t	0			dreating	1.efter:	
AR*/HISAT2/		1 1 C 1 1 L 1	net de la companya de	tite	nettion	e("shiph				. 15	
				)ESeq2	+		ante (		e hen s	t. *irs ta.	
SWA/BOWTIEZ	LETTERS CONTRACTOR CONTRACTOR				a na na	nsitar e	nd -			Farîsî Arv	
	a chid		revent in				udo filma y				
				11121	5-5 C.III	ent faur				(11.5	
				(A*02)	1119	0 =	, 5.5 IC 1				
and the second second				JFFDIFF	1161 5415						
			1 1							6 G G G G G G G G G G G G G G G G G G G	





SEQ READS

.1

BCL2FASTQ2

STRINGENTS MARKET

stations fifii epage of the state of Arts 1511 1 1 1 1 1 1 1 1 1 diffusive salaries. Types

and a second second

1 591 . 191. 1.1011

### Multiplex Run = Sequencing of Multiple Samples in one RUN

INELS CONTROL

a cale and a second second



Le ement

\_\_\_\_\_\_listerid\_\_\_\_\_\_incod\_"urendus\_\_\_\_\_\_listerid\_lenenduone("slip



HICKINE (CECC)

1 ALIGN

are to be an and a and a state of the second

REPART OF STREPT OF STREPT OF STREPT

olisistenene titititotti (1997年) ノリーバヨチム 自由する) い dinas talan digar

e is an and in a subdivide a state

11000111101112VC BCL2FASTQ2

- BREED - CONCENCE

### Faster Sequencing and Data Processing Times



Figure 1. 4-Channel vs. 2-Channel SBS Technology Image Detection and Base Calling. In 4-channel SBS, 4 images are necessary to capture the unique florescent dyes for each base. In contrast, 2-channel SBS requires only 2 images to determine all 4 base calls.

es especialista de la substancia de la companya de

-Seconderst full-lengther in suite - seconders - settemet sub Eit. to 



Provide States and the second s as extremes intraminant and a single contrasy francians and



### Faster Sequencing and Data Processing Times



Figure 1. 4-Channel vs. 2-Channel SBS Technology Image Detection and Base Calling. In 4-channel SBS, 4 images are necessary to capture the unique florescent dyes for each base. In contrast, 2-channel SBS requires only 2 images to determine all 4 base calls.

X

nno rigulaist 2811 - 1 1 1 Ving

alta 190 Alta 190 Alta 190 Alta 190 Alta 190 Alta 190 Alta 190

1 4 7 7 7 7 7 4 4 - 1 - 1 - 1 - 1 - 1 - 1 - 1



drath, Ercheren

O the EUCLOSE

tellens f. fil



Figure 1. 4-Channel vs. 2-Channel SBS Technology Image Detection and Base Calling. In 4-channel SBS, 4 images are necessary to capture the unique florescent dyes for each base. In contrast, 2-channel SBS requires only 2 images to determine all 4 base calls.

# Demultiplexing; **Typically done by the** sequencing facility

Distinction









FASTQ

a caller and deal of the

# Light intensities translated to TEXT;

**Derived from FASTA;** 

### Very large file, contains every single sequenced read for a given sample;

### Number of FASTQ files = Number of Samples

TOR PEXZ The second second states of the second states of th

FASTQ

### @<machine\_id>:<run\_number>:<flowcell\_ID>:<lane>:<tile>:<x-pos>:<y-pos> <read>:<is filtered>:<control number>:<index sequence>

### @NB500995:63:HC5GHBGX5:1:11101:22946:1063 1:N:0:CAACTAAT+AGATCTCG GCAACNTTGATCAGTTCTGACACAGTGTTTTGAACCATATCAGGATCCCTCACATCACACTGAATTGCATGAACCT

### 



nretrated to the

ALEN

1.11 FASTQ

A Diel GLOBACIA

@NB50099 GCAACNT

AAAAA#/EEEE

Mett 11 ....tens t.^irs t Caster is the state transference

of Franks studios en la parej

**Per base sequence quality** 

up of the Lot City



SCATGAACCT





e ce server le subdes d'éléget d'éléget d'éléget de la servere de la servere de la servere de la servere de la

# A poor RNA-seq run is characterized by:

# PCR duplicates

FastQC

Adapter contamination 

## rRNA and tRNA reads

**Unmappable reads** 



inger a low and a set a set of the low approximation of the started of the started of the set of the started of the



TrimGalore

10 i e[11,3

# Trim for noisy short fragments;

## 2 color chemistry bias;

Trim for low quality reads; **Trim for adapter sequences;** an sola bilas. Sola mai de seconda en el contra de la sola de la contra de la contra de la contra de la contra A e provisión de la definita de la contra de l A sola de la contra contra de la dela contra de la contra d alless salare ly account of the second contract the second contract of the second contract and second and a and the second of the

<pre>clip Site Disconstruction and product on the clip clip is</pre>			- + 6			14							
TrimGalore	• : • : r •			54 06 91	11 24 0 - 1 0		1						
Charles (1) 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	26									Quali	ty scor	res aci	ross
	30 34 32									$\bigvee$			$\wedge$
	30											V	
	28												
	26												
الاست مستحصية في المائين المستحدة المستحد في المال المراكز والمارية. مورية: المارية مرتبع في في أن أنه المراكز مارية المراجع المارية المراجع المارية المراجع المارية المراجع المارية	24												
entered for a first set of the se	22												
and a set of a set of a state of a	20												
	18												
4em 185	16												
	14												
	12												
<pre>int brink Eff end ( ) = A subset ( ) b is</pre>	10												
the sense of the present of the presents	8												
tions stalprettic southers of	б												
1	4												
ALLENS TO A DESCRIPTION OF A DESCRIPTION	2												
<pre>cdate / i mistaristratarist / i inte change / i mistaristration / i inte change / i international / internation change / i international / internation change / i internation chan</pre>	0	1	2	3	4	5	67	8	9	15-19	30	-34	45
e in elthiù tobargai													F

dissistant lyaness distributed for all sectors as 11 g or and sectors of dissistant lyanes. As a discussion of the a calendary in the construction of the constru and a second of the second of the

LE LERONLAGE (FYTO CONTRACTOR CONTRACTOR DE LE CREALAGEROR) energienen tione ("slut" - - - - - E) dirum de la service of the manager by the second functions of the second process of the second of the second se





188913110112910 turbino, turu, i juuroskinsu lansen stranger of conditions of a constraint of bases in the constraint and the second stranger of the stranger of the

oppendation and the Seconstruct further to HECADLE (ELC) 11201012/0150 









3.11. RNA

toria i in in in-

Consistent in the second statement of the statement of the statement of the second statement of the se (a) Frank's stationers: Equators will be interested as write (1) and (1) or or (1) is write the

STAR\*/HISAT2/ BWA/Bowtie2

# MAPPING/ALIGNMENT:

**Assignment of FASTQ reads to most likely** locus of origin in the REFERENCE GENOME

### This step has to be done for all FASTQ files

\*--quantMode FOR RNA SEQ READS





4120221210110 · 51. 16:5 / STAR\*/HISAT2/

BWA/Bowtie2

THE STREET 1.1011 1396.196.1 to to 112 CRONE (0150 . Stt Discertaneers one("slight 1264-111 Co. 1240 Transition and Se ement 's m · (1), }

### (a) Aligning to the transcriptome



### (b) Aligning to the genome



up constantit all issues of a first programmer and the . at the state of to . Hite Hader tements one ("slap" A DIST DIST OF STATE TRADUCED CALENCE usuellens fundil: A PLANE T. TIL 1.5.5.0.0 \*--quantMode FOR RNA SEQ READS

### transcript

### chromosome

### Exon 3 AG

### Intron

. 513 and of all i 相同で見りした。

STAR\*/HISAT2/ BWA/Bowtie2

# **Preliminary Step for all Aligners:**

Generate genome index (per genome type)

\*--quantMode FOR RNA SEQ READS

Exon 1 GT AG Exon 2 GT AG Exon 3	
Intron Intron	

### Allows for computationally efficient mapping



11/01/01/01/01

STAR\*/HISAT2/ BWA/Bowtie2

STAR \

--runThreadN 12 \ --runMode genomeGenerate \ --genomeDir /path/to/genomeIndex/ \ --sjdbGTFfile referenceAnnotation.gtf \ −-sjdb0verhang 100 \ --limitGenomeGenerateRAM 152003700778

appendiate and the

1396.191.1

111.105.

and bareards

C 198813103240 stations funding for a constitution of pusicipal contraction of the \*--quantMode FOR RNA SEQ READS

i fit to to 1201012/01566 sites e eners che("site" 1840 Contraction BRIESS 25 STREETERS TURNED FOR DEVELOPMENT AND A STREETERS

· · · · · · · · · · · · · · · · · · ·	THEFT THE STATE OF THE		, el idine :
List Second from	TROVING HESTER	(a) Aligning to the transcriptome	
areas and and any or a	ALL REAL STR	Read 1 Read 2 Exon 1 Exon 2 Exon 3	
provide the second received	-1		
	المبالاي ومرتاف الماتي ال	(b) Aligning to the genome	
	as the second	Read 2	
	n harang ang ng manakari sara	Exon 1 GT AG Exon 2 GT	AG Exon 3
	to a storage of the storage	Intron Intro	on
	$\alpha \to \alpha M \Lambda \to \alpha \Lambda$		

# --genomeFastaFiles referenceGenome.fasta \



––runThreadN 12 ∖

--genomeDir /path/to/genomeIndex

![](_page_18_Picture_18.jpeg)

ALL'A DELLA RECENCIÓN DE LA CONTRACTÓN DE LA CONTRACTÍN D	<pre>is not set of a set of a</pre>
STAR*/HISAT2/	<pre>is evened trone("south</pre>
BWA/Bowtie2	(a) Aligning to the transcriptome         Read 1       Read 2         Exon 1       Exon 2       Exon 3
	(b) Aligning to the genome Read 2 Read 1 Exon 1 GT AG Exon 2 GT AG Exon 3 Intron Intron
	STAR \
	runThreadN 12 ∖
en del la constanta de la const La constanta de la constanta de	<pre>genomeDir /path/to/genomeIndex \</pre>

STAR ∖

- --runThreadN 12 \
- --runMode genomeGenerate \
- --genomeDir /path/to/genomeIndex/ \
- --genomeFastaFiles referenceGenome.fasta \
- --sjdbGTFfile referenceAnnotation.gtf \
- −−sjdb0verhang 100 \
- --limitGenomeGenerateRAM 152003700778

1211315 2151. \*--quantMode FOR RNA SEQ READS

÷ 1.	STAR	$\setminus$	
: 1 			<pre>runThreadN 12 \</pre>
			<pre>genomeDir /path/to/genomeIndex \</pre>
			<pre>readFilesIn sample.fastq.gz \</pre>
			<pre>readFilesCommand gunzip -c \</pre>
			<pre>outSAMstrandField intronMotif \</pre>
			outFilterIntronMotifs RemoveNoncanonica
			<pre>outSAMtype BAM SortedByCoordinate \</pre>
			––outFileNamePrefix prefix. ∖
			––limitBAMsortRAM 61675612266 ∖
			quantMode GeneCounts

![](_page_19_Picture_18.jpeg)

## **Consensus format to store alignment records;**

# format;

### There are 2 sections in this file:

**Header Section** 

# Alignment Section

# SEQUENCE ALIGNMENT MAP: SAM/BAM

All aligners will generate results in the SAM

### **SEQUENCE ALIGNMENT MAP: SAM/BAM** VN: @HD **Consensus format** @SQ LN: SN: to store alignment @RG ID: SM: records; @PG ID: All aligners will @CO generate results in З 4 the SAM format; QNAME FLAG RNAME POS Paired read? There are 2 Unmapped? Mapped to rev. sections in this file: strand? 1<sup>st</sup> in pair? 2<sup>nd</sup> in pair? Failed QC?

QNAME	FLAG	RNAME	POS	MAPQ	CIGAR	RNEXT	PNEXT	TLEN	SEQ	QUAL	OPT
QNAME	FLAG	RNAME	POS	MAPQ	CIGAR	RNEXT	PNEXT	TLEN	SEQ	QUAL	OPT
QNAME	FLAG	RNAME	POS	MAPQ	CIGAR	RNEXT	PNEXT	TLEN	SEQ	QUAL	OPT
QNAME	FLAG	RNAME	POS	MAPQ	CIGAR	RNEXT	PNEXT	TLEN	SEQ	QUAL	OPT

Section Alignment Section tellens f. fil 21 6 5

Header

(theoretically) optional

113 CLOVE (0150

6 8 9 10 11 >11 MAPQ CIGAR RNEXT PNEXT TLEN | SEQ QUAL OPT <TAG>:<TYPE>:<VALUE> M (mis)match AS insertion BC ALIGNMENT D deletion NH SECTION N skipped NM S soft clipped 1 line per locus H hard clipped P padding

![](_page_21_Picture_6.jpeg)

413 0303 S.e.Lens

5-21

appearance in the

Pos.	Field	Example entry	Description	NA value
1	QNAME	Read1	Query template (= read) name (PE: read pair name)	required
2	FLAG	833 (IIII) (IIII) (IIII) (IIII) (IIIII) (IIIII) (IIIII) (IIIIII) (IIIIII) (IIIIIII) (IIIIIII) (IIIIIIII	Information about the read's mapping properties encoded as bit-wise flags (see next section and Table $5$ ).	required
3	RNAME	chrI	Reference sequence name. This should match a <b>@SQ</b> line in the header.	*
4	POS	15364	1-based leftmost mapping position of the first matching base. Set as 0 for an unmapped read without coordi- nates.	0
5	MAPQ	30	Mapping quality of the alignment. Should be a Phred- scaled posterior probability that the position of the read is incorrect, but the value is completely dependent on the alignment program. Some tools set this to 0 if mul- tiple alignments are found for one read.	0
6	CIGAR	$51\mathrm{M}$	Detailed information about the alignment (see below).	*
7	RNEXT	—	PE reads: reference sequence name of the next read. Set to "=" if both mates are mapped to the same chromosome.	*
8	PNEXT	15535	PE reads: leftmost mapping position of the next read.	0
9	TLEN	232	PE reads: inferred template length (fragment size).	0
10	$\operatorname{SEQ}$	CCAGGC	The sequence of the aligned read on the forward strand (not including indels).	*
11	QUAL	BBH1+B	Base quality (same as the quality string in the FASTQ format, but always in Sanger format $[ASCII+33]$ ).	*
12ff	OPT	NM:i:0	Optional fields (format: ${\tt TAG\!>\!:\!{\tt TYPE\!>\!:\!{\tt VALUE\!>}};$ see below).	

![](_page_22_Picture_14.jpeg)

### areas t e Liema

and the set of the

Pos.	Field	Example Description NA entry value	idnos Nez ( m Le china Le china
1	QNAME	Read1 Query template (= read) name (PE: read pair name) required	
2	FLAG	83 SAM Flag: 4 Explain	
3	RNAME	Switch to mate     Toggle first in pair / second in pair	
4	POS	1536	
		Find SAM flag by property: Summary:	
5	MAPQ	30       To find out what the SAM flag value would be for a given combination of properties, tick the boxes for those that you'd like to include. The flag value will be shown in the SAM Flag field       read unmapped (0x4)         above       above	
6	CIGAR	<ul> <li>read paired</li> <li>51M read mapped in proper pair</li> </ul>	
7	RNEXT	<ul> <li>read unmapped</li> <li>mate unmapped</li> </ul>	
8	PNEXT	1553 read reverse strand	≠dio inf trans
9	TLEN	<ul> <li>232</li> <li>mate reverse strand</li> <li>first in pair</li> </ul>	
10	$\operatorname{SEQ}$	CCA second in pair	
11	QUAL	BBH       not primary alignment         read fails platform/vendor quality checks	alig starg
12ff	OPT	NM: read is PCR or optical duplicate supplementary alignment	
			4 4 P

![](_page_23_Picture_20.jpeg)

413 0303 S.e.Lens

5-21

appearance in the

Pos.	Field	Example entry	Description	NA value
1	QNAME	Read1	Query template (= read) name (PE: read pair name)	required
2	FLAG	Band and a set of the	Information about the read's mapping properties encoded as bit-wise flags (see next section and Table $5$ ).	required
3	RNAME	chrI	Reference sequence name. This should match a <b>@SQ</b> line in the header.	*
4	POS	15364	1-based leftmost mapping position of the first matching base. Set as 0 for an unmapped read without coordi- nates.	0
5	MAPQ	30	Mapping quality of the alignment. Should be a Phred- scaled posterior probability that the position of the read is incorrect, but the value is completely dependent on the alignment program. Some tools set this to 0 if mul- tiple alignments are found for one read.	0
6	CIGAR	51M	Detailed information about the alignment (see below).	*
7	RNEXT	—	PE reads: reference sequence name of the next read. Set to "=" if both mates are mapped to the same chromo- some.	*
8	PNEXT	15535	PE reads: leftmost mapping position of the next read.	0
9	TLEN	232	PE reads: inferred template length (fragment size).	0
10	SEQ	CCAGGC	The sequence of the aligned read on the forward strand (not including indels).	*
11	QUAL	BBH1+B	Base quality (same as the quality string in the FASTQ format, but always in Sanger format $[ASCII+33]$ ).	*
12ff	OPT	NM:i:0	Optional fields (format: ${\tt TAG>::\!{\tt VALUE>};\!$ see below).	

![](_page_24_Picture_15.jpeg)

### H2 CRD1 Sellent.

### <QNAME> <FLAG> <RNAME> <POS> <MAPQ> <CIGAR> <MRNM> <MPOS> <ISIZE> <SEQ> <QUAL>

Pos.	Field	Example entry	Description	NA value	<pre>child in a strict of the contract of the</pre>
1	QNAME	Read1	Query template (= read) name (PE: read pair name)	required	enderse ender sollt begehörteren inner die die eine eine eine eine die eine eine eine eine eine eine e
2	FLAG	83	Information about the read's mapping properties en- coded as bit-wise flags (see next section and Table 5).	required	
3	RNAME	chrI	Reference sequence name. This should match a <b>@SQ</b> line in the header.	*	
4	POS	15364	1-based leftmost mapping position of the first matching base. Set as 0 for an unmapped read without coordinates.	0	(1) A second to the second of the second
5	MAPQ	30	Mapping quality of the alignment. Should be a Phred- scaled posterior probability that the position of the read is incorrect, but the value is completely dependent on the alignment program. Some tools set this to 0 if mul- tiple alignments are found for one read.	0	

spectration th

5-25

	C	CICAD	EIN/																
	0	CIGAR	DIM				Re	feren	ce sec	uence	e with a	aligne	d reads			CIG	AR string	Explanation	
	7	RNEXT		сте	i C A	ΑTG	ітт /	AG/	АТА	A *	* G A	ΤА	GСТ	GTG	СТА				
			2003 2010							AA	<mark>G</mark> GA	ΤА	* СТ	G		1M2	214M1D3M	Insertion & De	le
	8	PNEXT	15535					G	АТА	A * (	<mark>G</mark> GA	ΤА				5M1	P <b>1I4M</b>	Padding & Ins	e
	9	TLEN	232			ΤG	тт,	4						ΤG	СТА	5M1	5N5M	Spliced read	
	10	SEQ	CCA	a a a	C A	ΑΤ G	ітт /	A G								358	BM	Soft clipping	
	11	OUAL	BBH 4	4 A 4	C A	A T G	ітт /	A G								3H8	BM	Hard clipping	
		QUIIL	DDII		forma	at, but	always in	Sange	r format	[ASCII	[+33]).		tie en Station	nent és	gz 1 1 tou	Service Provide Service Provide	i i s∏tibe Seuris	i intern Statistics in its	
	$12 \mathrm{ff}$	OPT	NM:i:0		Optic	onal fiel	ds (forma	t: <tag< td=""><td>&gt;:<typh< td=""><td>E&gt;:<val< td=""><td>UE&gt;; see</td><td>be-</td><td></td><td>• • • • • • • • •</td><td></td><td></td><td></td><td></td><td></td></val<></td></typh<></td></tag<>	>: <typh< td=""><td>E&gt;:<val< td=""><td>UE&gt;; see</td><td>be-</td><td></td><td>• • • • • • • • •</td><td></td><td></td><td></td><td></td><td></td></val<></td></typh<>	E>: <val< td=""><td>UE&gt;; see</td><td>be-</td><td></td><td>• • • • • • • • •</td><td></td><td></td><td></td><td></td><td></td></val<>	UE>; see	be-		• • • • • • • • •					
1			a = 0 10.00	The second second	low).		and the second		a i sa a s Na sa ta ta sa sa		(	n a se Antonio a tra		1000					

![](_page_25_Picture_15.jpeg)

413 0303 S.e.Lens

5-21

appearance in the

Pos.	Field	Example entry	Description	NA value
1	QNAME	Read1	Query template (= read) name (PE: read pair name)	required
2	FLAG	833	Information about the read's mapping properties encoded as bit-wise flags (see next section and Table $5$ ).	required
3	RNAME	chrI	Reference sequence name. This should match a <b>@SQ</b> line in the header.	*
4	POS	15364	1-based leftmost mapping position of the first matching base. Set as 0 for an unmapped read without coordinates.	0
5	MAPQ	30	Mapping quality of the alignment. Should be a Phred- scaled posterior probability that the position of the read is incorrect, but the value is completely dependent on the alignment program. Some tools set this to 0 if mul- tiple alignments are found for one read.	0
6	CIGAR	51M	Detailed information about the alignment (see below).	*
7	RNEXT	=	PE reads: reference sequence name of the next read. Set to "=" if both mates are mapped to the same chromo- some.	*
8	PNEXT	15535	PE reads: leftmost mapping position of the next read.	0
9	TLEN	232	PE reads: inferred template length (fragment size).	0
10	SEQ	CCAGGC	The sequence of the aligned read on the forward strand (not including indels).	*
11	QUAL	BBH1+B	Base quality (same as the quality string in the FASTQ format, but always in Sanger format $[ASCII+33]$ ).	*
12ff	OPT	NM:i:0	Optional fields (format: ${\tt TAG\!>\!:\!{\tt TYPE\!>\!:\!{\tt VALUE}\!>};$ see below).	

![](_page_26_Picture_14.jpeg)

STAR\*/HISAT2/ BWA/Bowtie2

### **Could most reads be aligned?**

### Are there any obvious biases of the read distributions?

### Are the replicate samples as similar to each other as

expected?

\*--quantMode FOR RNA SEQ READS

![](_page_27_Picture_16.jpeg)

STAR\*/HISAT2/ BWA/Bowtie2

### **Could most reads** be aligned?

Are there any obvious biases of the read distributions?

Are the replicate samples as similar to each other as expected?

![](_page_28_Figure_5.jpeg)

2151. 1211315 r ans 11 1.7 \*--quantMode FOR RNA SEQ READS

RQN\_7.1 RQN\_9.6 RQN 7.

RON 7

BON 9 RQN 8.6 RQN\_8.3

te ement

Showing <sup>12</sup>/<sub>12</sub> rows and <sup>2</sup>/<sub>2</sub> columns. Sopy table

Sample Name	% Aligned -	M Aligned
RQN_10	88.7%	32.9
RQN_9.6_2	88.3%	29.6
RQN_8.3	87.7%	31.6
RQN_7.6	87.4%	32.1
RQN_8.6	87.3%	20.9
RQN_9.1	87.1%	28.8
RQN_9.6_1	85.0%	32.0
RQN_7.4	83.0%	29.9
RQN_9.5	80.2%	27.7
RQN_7.1_1	77.3%	26.6
RQN_6.8	70.7%	24.8
RQN_7.1_2	68.1%	23.9

![](_page_28_Figure_12.jpeg)

include the terms	VELLESITE C. F.C.	wt1		wt2 w	t3 wt4	wt5	ko1	ko	v2 ko3	ko4	ko5	
- 12651 F. C. C. C. C. C.	.elt.e/* -:- *	gene1	135	148	146	121	140	269	268	227	263	259
NEXTSEO 500	. Tensora de socio como das	gene2	803	797	841	800	874	412	408	388	393	398
		gene3	40	25	38	41	35	413	393	417	374	415
Ithin treation		gene4	381	383	415	374	354	809	840	859	856	845
a service advances a service adv	Consideration and a second second	gene5	775	766	773	749	784	302	310	324	342	314
the contract of the second		gene6	305	313	256	313	315	831	817	832	859	869
and street of the completes subtra-		gene7	816	819	800	793	790	485	481	429	461	508
		gene8	40	22	40	37	32	421	476	479	528	483
BCL2FASTQ2		gene9	963	935	938	953	948	43	26	41	28	39
		gene10	697	749	715	724	715	233	259	284	277	26
the second of the second		gene11	36	50	40	35	44	168	178	168	170	187
		gene12	60	66	54	61	71	288	289	293	289	330
for an and the second states of		gene13	537	517	523	512	515	142	134	145	145	14
		gene14	655	615	610	664	606	842	889	827	885	83
	Factor	gene15	426	439	436	420	432	131	155	159	139	15
FASIQ	Faslyc	gene16	952	976	974	987	947	789	828	825	850	79
the second state of the second states		gene17	379	446	410	423	394	963	1012	913	968	98
		gene18	17	17	14	20	22	131	113	135	127	11
		gene19	985	874	896	982	992	848	890	899	896	873
and the second		gene20	197	191	202	180	172	765	754	784	791	79
		gene21	399	477	414	466	440	686	668	741	754	71
TrimGalore		86UCZ T	<del> </del>			+00	4004		++0	• • • •	<u>↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ </u>	
		genezo	300	ZZV	111	190		102	124	7/1	751	71
the state of the s		Reveral M	202	-701 -70			172	765	250	201	701	- 0 V
Sugersoner et l'av base		Relieto	HTSI	EO-Cou	nts		ni: 12: 22					
and the second sec			: 1	1 + 1 + 1 + 1 = 1	to the state							
States and the state from the state						10	etti (	. 11 1 ***	n in States	er fielde		
added a contract of the second s				101110		a geological des		0.016 (110	all's satur			
		al agent the reaction of			1				1120020000	1949 - 1949 - 1949 1949 - 1949		
				DESeq	2		1111 (F)		selens t.*	in ta.		
BWA/Bowtie2	reduction of the second		:		nnn - Srai		na)			ter and		
selling a line of the selling s		st jite stit nast					1401 10A				્યાંતુ	
(i) i tar				: () () <del>(</del> <del>)</del>	5-2 C.D.O	and fisher			- i 6.[40]		totar,	
	and should be a company of the second				di ng	0 =	, 5.5 IC 1					
As the large ly are the	<pre></pre>			CUFFDI	£.E.							
	and the second second second second											

![](_page_29_Picture_2.jpeg)

![](_page_29_Picture_3.jpeg)

![](_page_29_Picture_4.jpeg)

![](_page_30_Picture_0.jpeg)

# Number of reads mapped to a gene depends on:

### Its own expression level; Ø

### Its length;

## The sequencing depth;

Normalization is done to eliminate systematic effects; DESeq2::estimateSizeFactors( )

Expression of all other genes within the sample;

![](_page_31_Picture_0.jpeg)

![](_page_31_Figure_12.jpeg)

1 2

**Raw counts distribution** 

![](_page_31_Figure_18.jpeg)

.c.H. it to 112 01012 (0150 LEITERS extended trone("silt" 1264 C 1 C 1 C 2 Y DRUGSE ransities. 21.1 · [1], 1

### Normalized counts distribution

![](_page_31_Picture_22.jpeg)

![](_page_32_Picture_0.jpeg)

# Exploration of Global normalized read count

# patterns:

5E

Hamplehor much spire Hitscore of the Substration of the Active King of the strateHitscore of the assessment of the Human of the Human of the constraint of the strategies and an interval of the Human of the strategies affective and the second of the Human of the Human of the Hitscore of the strategies of the Human of the Human of the strategies of the strategies of the Human of the Human of the strategies of the strategies of the Human of the Human of the strategies of the strategies of the Human of the Human of the strategies of the strategies of the Human of the Human of the strategies of the strategies of the Human of the Human of the strategies of the strategies of the Human of the Human of the strategies of the Human of the Human of the Human of the Human of the strategies of the Human of the Human of the Human of the Human of the strategies of the Human of the Human of the Human of the

interview distribution in a figuration in the contract of the contract of

### **Cluster dendrogram**

![](_page_32_Figure_10.jpeg)

Method: Euclidean distance - Ward criterion hclust (\*, "ward.D")

![](_page_32_Picture_12.jpeg)

![](_page_33_Picture_0.jpeg)

# Exploration of Global normalized read count

# patterns:

5E

Hamplehor much spire Hitscore of the Substration of the Active King of the strateHitscore of the assessment of the Human of the Human of the constraint of the strategies and an interval of the Human of the strategies affective and the second of the Human of the Human of the Hitscore of the strategies of the Human of the Human of the strategies of the strategies of the Human of the Human of the strategies of the strategies of the Human of the Human of the strategies of the strategies of the Human of the Human of the strategies of the strategies of the Human of the Human of the strategies of the strategies of the Human of the Human of the strategies of the strategies of the Human of the Human of the strategies of the Human of the Human of the Human of the Human of the strategies of the Human of the Human of the Human of the Human of the strategies of the Human of the Human of the Human of the

interview distribution in a figuration in the contract of the contract of

![](_page_33_Figure_8.jpeg)

PC1 (58.24%)

![](_page_33_Picture_10.jpeg)

RNA SEQ DATA SETS

# How do we best model this data (?) ???

### **Count based (Discrete) RNA-seq data, suffers**

# from non-uniform mean-variance relationships

# Heteroscedasticity;

![](_page_35_Figure_0.jpeg)

 Constant from the first of star with the second shift constraint news to statute the content of the close of the spectra state to the content have function I ANT, ARTER FORTER ASTREE ASSOCANTED TO DECAMPLE TO DECAMPLE TO DECAMPLE TO DECAMPLE TO DECAMPLE TO DECAMPLE HECHINEROLD I I SELEMORTO UN U and a second the second s Lisebens funding funding of a province else concorrections and a second s - styles" rest in nuore por usaligno o potresidenteside distreside de la secondada. Alla ellebra destación e el celebra de la completa and Fridde mail and a fighter second seco

 $(x, z) \in \{x, y, z\} \in \{y, z\} \in \{z, z\}$ A 16 STE TS contact to the endertione ("sluck") on the state of the second of the state of the s alema funite de l'unavier de secces de la construction de la compete de la maisse 2, la salema funite des public L'une salema de la contraction de la compete de la compete transition and l'unavier d'une salema de la contracti

![](_page_35_Figure_3.jpeg)

![](_page_35_Picture_4.jpeg)

. 213

![](_page_36_Picture_0.jpeg)

# Wald test is used to report Differentially Expressed

Genes!

# **Once, mean-variance relationship is modeled,**

log2 fold ch	ange (MAP): condi	tion treated vs untre	ated		
Wald test p-	value: condition	treated vs untreated			
DataFrame wi	th 9921 rows and	5 columns			
	baseMean	log2FoldChange	lfcSE		
	<numeric></numeric>	<numeric></numeric>	<numeric></numeric>		
FBgn0000008	95.1442917575889	0.00119919675662286	0.151896997597845		
FBgn0000014	1.05652281859341	-0.00473412281044922	0.205467617376393		
FBgn0000017	4352.55356876647	-0.189899902298335	0.120376617165947		
FBgn0000018	418.61048415965	-0.0699575311158887	0.123900600388886		
FBgn0000024	6.406199980976	0.0175271520689073	0.198632752197541		
FBgn0261570	3208.38861003698	0.241102900991117	0.124446879845224		
FBgn0261572	6.19718814545467	-0.0657617344183244	0.2141351371368		
FBgn0261573	2240.97951122377	0.0100061908254208	0.0993764053703328		
FBgn0261574	4857.68037348332	0.00843552221427279	0.140826652378679		
FBgn0261575	10.6825203335563	0.00809100502438704	0.201470391594341		
	pvalu	e padj			
	<numeric< td=""><td>&gt; <numeric></numeric></td><td></td><td></td><td></td></numeric<>	> <numeric></numeric>			
FBgn0000008	0.99188165684825	4 0.99721076667093			
FBgn0000014	0.81729868295179	8 NA			
FBgn0000017	0.057559105908221	2 0.288001711413016			
FBgn0000018	0.48085581535312	4 0.826833683766374			
FBgn0000024	0.75978793648838	4 0.943501114514859			
FBgn0261570	0.020307013775005	1 0.144240002513885			
FBgn0261572	0.21620263778915	7 0.607847805203262			
FBgn0261573	0.91061455016716	6 0.982656666760864			
FBgn0261574	0.93629077250126	1 0.988179230260622			
FBgn0261575	0.8605216031793	7 0.96792800379094			
	log2 fold ch Wald test p- DataFrame wi FBgn0000014 FBgn0000017 FBgn0000018 FBgn0261570 FBgn0261572 FBgn0261573 FBgn0261574 FBgn0261574 FBgn0000018 FBgn0000018 FBgn0000018 FBgn0000018 FBgn0000018 FBgn0000018 FBgn0000018 FBgn0261572 FBgn0261572 FBgn0261573 FBgn0261574 FBgn0261574	<pre>log2 fold change (MAP): condi Wald test p-value: condition DataFrame with 9921 rows and baseMean</pre>	log2 fold change (MAP): condition treated vs untreated Wald test p-value: condition treated vs untreated DataFrame with 9921 rows and 5 columns baseMean log2FoldChange <numeric> <numeric> FBgn000008 95.1442917575889 0.00119919675662286 FBgn000014 1.05652281859341 -0.00473412281044922 FBgn000017 4352.55356876647 -0.189899902298335 FBgn0000018 418.61048415965 -0.0699575311158887 FBgn0000024 6.406199980976 0.0175271520689073 </numeric></numeric>	log2 fold change (MAP): condition treated vs untreated Wald test p-value: condition treated vs untreated DataFrame with 9921 rows and 5 columns baseMean log2FoldChange lfcSE <numeric> <numeric> <numeric> <numeric> FBgn000008 95.1442917575889 0.00119919675662286 0.151896997597845 FBgn0000014 1.05652281859341 -0.00473412281044922 0.205467617376393 FBgn0000017 4352.55356876647 -0.189899902298335 0.120376617165947 FBgn0000018 418.61048415965 -0.0699575311158887 0.123900600388886 FBgn0000024 6.406199980976 0.0175271520689073 0.198632752197541  FBgn0261570 3208.38861003698 0.241102900991117 0.124446879845224 FBgn0261573 2240.97951122377 0.0100061908254208 0.0993764053703328 FBgn0261573 240.97951122377 0.0100061908254208 0.0993764053703328 FBgn0261573 240.97951122377 0.0100061908254208 0.0993764053703328 FBgn0261574 4857.68037348332 0.00809100502438704 0.201470391594341 pvalue padj <numeric> <numeric> FBgn000008 0.991881656848254 0.99721076667093 FBgn0000014 0.817298682951798 NA FBgn000014 0.817298682951798 NA FBgn000014 0.817298682951798 NA FBgn000014 0.480855815353124 0.826833683766374 FBgn000014 0.480855815353124 0.826833683766374 FBgn000018 0.440855815353124 0.826833683766374 FBgn000018 0.480855815353124 0.826833683766374 FBgn000018 0.4203070137750051 0.144240002513885 FBgn0261570 0.0203070137750051 0.144240002513885 FBgn0261570 0.0203070137750051 0.144240002513885 FBgn0261570 0.0203070137750051 0.144240002513885 FBgn0261573 0.910614550167166 0.982656666760864 FBgn0261573 0.910614550167166 0.982656666760864 FBgn0261574 0.936290772501261 0.988179230260622 FBgn0261575 0.8605216031793 0.96792800379094</numeric></numeric></numeric></numeric></numeric></numeric>	log2 fold change (MAP): condition treated vs untreated Wald test p-value: condition treated vs untreated DataFrame with 9921 rows and 5 columns baseMean log2FoldChange lfcSE <numeric> <numeric> <numeric> FBgn000008 95.144291757889 0.00119919675662286 0.151896997597845 FBgn0000014 1.05652281859341 -0.00473412281044922 0.205467617376393 FBgn0000017 4352.55356876647 -0.189899902298335 0.120376617165947 FBgn0000018 418.61048415965 -0.0699575311158887 0.123900600388886 FBgn000024 6.406199980976 0.0175271520689073 0.198632752197541  FBgn0261570 3208.38861003698 0.241102900991117 0.124446879845224 FBgn0261572 6.19718814545467 -0.0657617344183244 0.2141351371368 FBgn0261573 2240.97951122377 0.0100061908254208 0.0993764053703328 FBgn0261573 10.682520335563 0.0080100502438704 0.201470391594341 pvalue padj <numeric> numeric&gt; FBgn0261575 10.682520335563 0.99721076667093 FBgn000018 0.48055815353124 0.826833683766374 FBgn000018 0.48055815353124 0.826833683766374 FBgn000014 0.817298682951798 NA FBgn000014 0.817298682951798 NA FBgn000014 0.817298682951798 NA FBgn000014 0.817298682951798 NA FBgn000014 0.817298682951798 NA FBgn000014 0.81729780510.218001711413016 FBgn000014 0.759787936488384 0.943501114514859  FBgn0261570 0.0203070137750051 0.144240002513885 FBgn000024 0.759787936488384 0.943501114514859  FBgn0261573 0.910614550167166 0.9882656666760864 FBgn0261574 0.936290772501261 0.988179230260622 FBgn0261575 0.86052160317937 0.96792800379094</numeric></numeric></numeric></numeric>

## log2 fold change (MAP): condition treated vs untreated ## Wald test p-value: condition treated vs untreated ## DataFrame with 9921 rows and 5 columns log2FoldChange baseMean <numeric> <numeric> ## FBgn0000008 95.1442917575889 0.00119919675662286 ## FBgn0000014 1.05652281859341 -0.00473412281044922 ## FBgn0000017 4352.55356876647 -0.189899902298335 ## FBgn0000018 418.61048415965 -0.0699575311158887 ## FBgn0000024 6.406199980976 0.0175271520689073 • • • . . . ## FBgn0261570 3208.38861003698 0.241102900991117

ingle of the contribution

-Seclasmer Suth, Screating

VALUERSITE U. 1 U

## FBgn0261572 6.19718814545467 ## FBgn0261573 2240.97951122377 ## FBgn0261574 4857.68037348332 ## FBgn0261575 10.6825203335563 0.00809100502438704 ## pvalue

### <numeric> ## FBgn0000008 0.991881656848254 0.99721076667093 0.817298682951798 ## FBgn0000017 0.0575591059082212 0.288001711413016 ## FBgn0000018 0.480855815353124 0.826833683766374 ## FBgn0000024 0.759787936488384 0.943501114514859 • • •

## FBgn0261570 0.0203070137750051 0.144240002513885 ## FBgn0261572 0.216202637789157 0.607847805203262 ## FBgn0261573 0.910614550167166 0.9826566666760864 0.936290772501261 0.988179230260622 ## FBgn0261574 ## FBgn0261575 0.86052160317937 0.96792800379094

padj <numeric> NA . . .

-0.0657617344183244

.s.elent f.≜ii

DESeq2

##

##

## ...

##

## ...

## FBgn0000014

Adversion lane

a Charles have been a started where

1201012/01560 i to . Siters extends trone("sitt" -to possibilities and a second state of the second se

lfcSE

• • •

----

- <numeric>
- 0.151896997597845
- 0.205467617376393
- 0.120376617165947
- 0.123900600388886
- 0.198632752197541
- 0.124446879845224
- 0.2141351371368
- 0.0100061908254208 0.0993764053703328
- 0.00843552221427279 0.140826652378679
  - 0.201470391594341

1 . . . . I ..

## Differentially

MA-plot

DESeq2

### expressed

# features are

highlighted in

### red

Stelprettin and DIRECTORY STREET (1991) 110 (191 tellen: f. fil

![](_page_38_Figure_7.jpeg)

![](_page_38_Figure_9.jpeg)

Mean of normalized counts

126-121.0.122 District County

![](_page_39_Picture_0.jpeg)

LISTENING

### **Contact Info:**

### http://rnasegcore.vet.cornell.edu/

E-mail List Serve: TREX-GENEREG-L jgrenier@cornell.edu

### **Jen Grenier:**

### Faraz Ahmed:

**Christine Butler:** 

cab18@cornell.edu

r. Let f. In the rest. THANK YOU FOR 12:5:13.0.0

fahmed@cornell.edu

Ann Tate: aef93@cornell.edu Scan me e i ellis interval e l'anne de la company à de la company à travaire de la company à travaire de la company à

![](_page_40_Picture_14.jpeg)

![](_page_40_Picture_15.jpeg)

a success of the second second to the second terms of the second terms of the second second terms of the second terms of terms o

istor