

UCLA Institute for Quantitative and Computational Biology

March 2020

QCB Workshop W5a: RNAseq-1 (day 1)

Nicolas Rochette

(EEB/ISG)

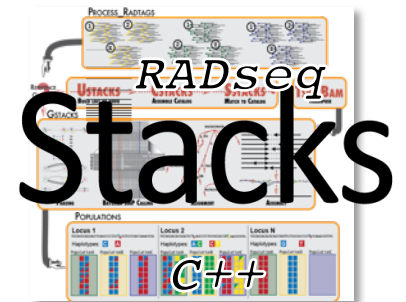
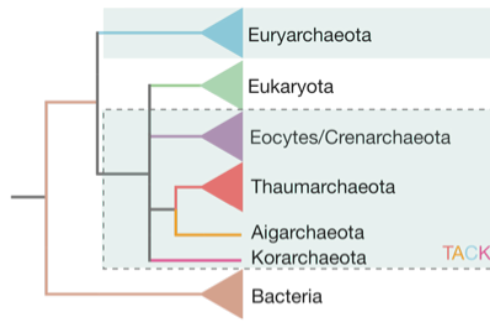
Karolina Kaczor-Urbanowicz

(Oral Biology & Medicine)

My background



Peromyscus maniculatus



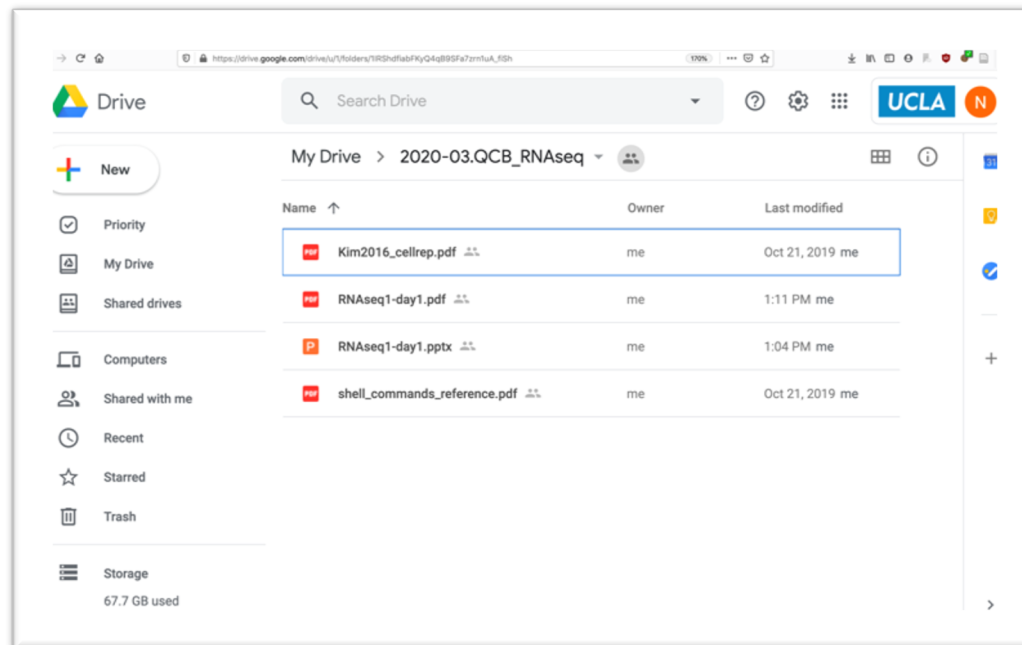
Workshops' google drive folder

Slides & supporting documents:

https://drive.google.com/drive/folders/1IRShdfiabFKyQ4qB9SFa7zrn1uA_fiSh

<https://tinyurl.com/sshouvw>

Please bookmark it / write it down!



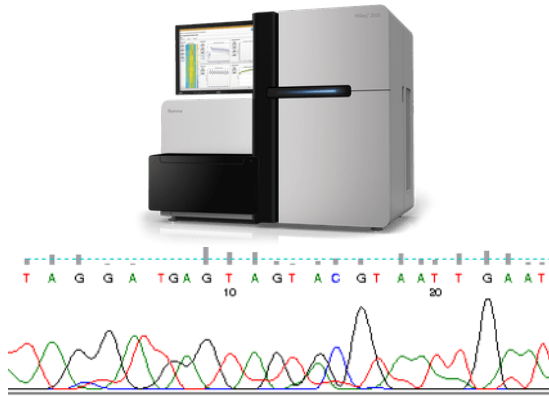
Workshop Overview

RNAseq workshops 1 & 2

- **RNAseq I: “from raw sequence data to gene counts”**
 1. Overview of RNAseq
Workshop data: *Kim JW et al. (2016), Dev Cell 37(6)*
Raw sequence reads: format, quality controls, filtering
 2. Alignments to an annotated genome: formats & software
 3. Counting gene/exon abundancies, visualizing alignments
- **RNAseq II: “from gene counts to biology”**
 - Normalization of expression counts
 - Visualization of gene expression levels (heatmaps, PCA)
 - Test for differentially expressed genes, functional enrichments

RNAseq workshops 1 & 2

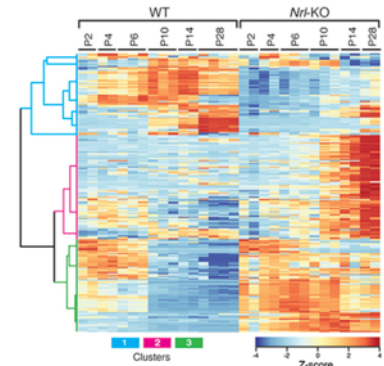
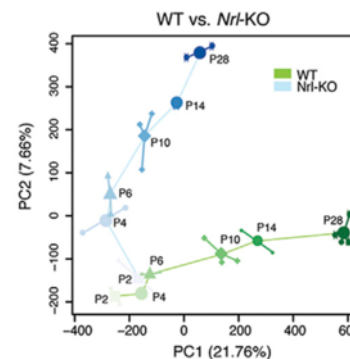
- RNAseq I: “from raw sequence data to gene counts”



	sample1	sample2	sample3	...
gene1	999	701	616	
gene2	532	520	41	
gene3	14	36	305	
...				

- RNAseq II: “from gene counts to biology”

	sample1	sample2	sample3	...
gene1	999	701	616	
gene2	532	520	41	
gene3	14	36	305	
...				

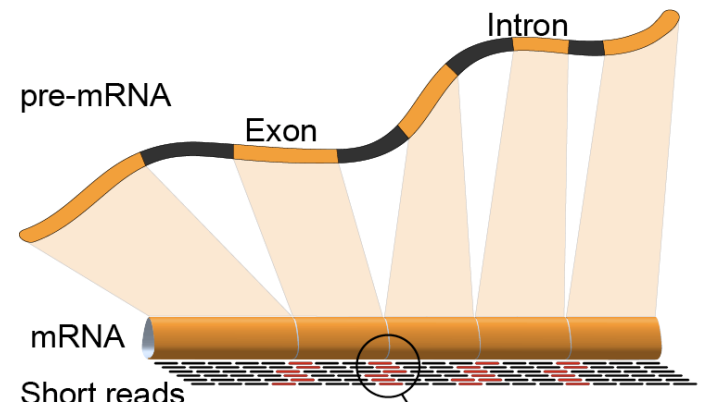
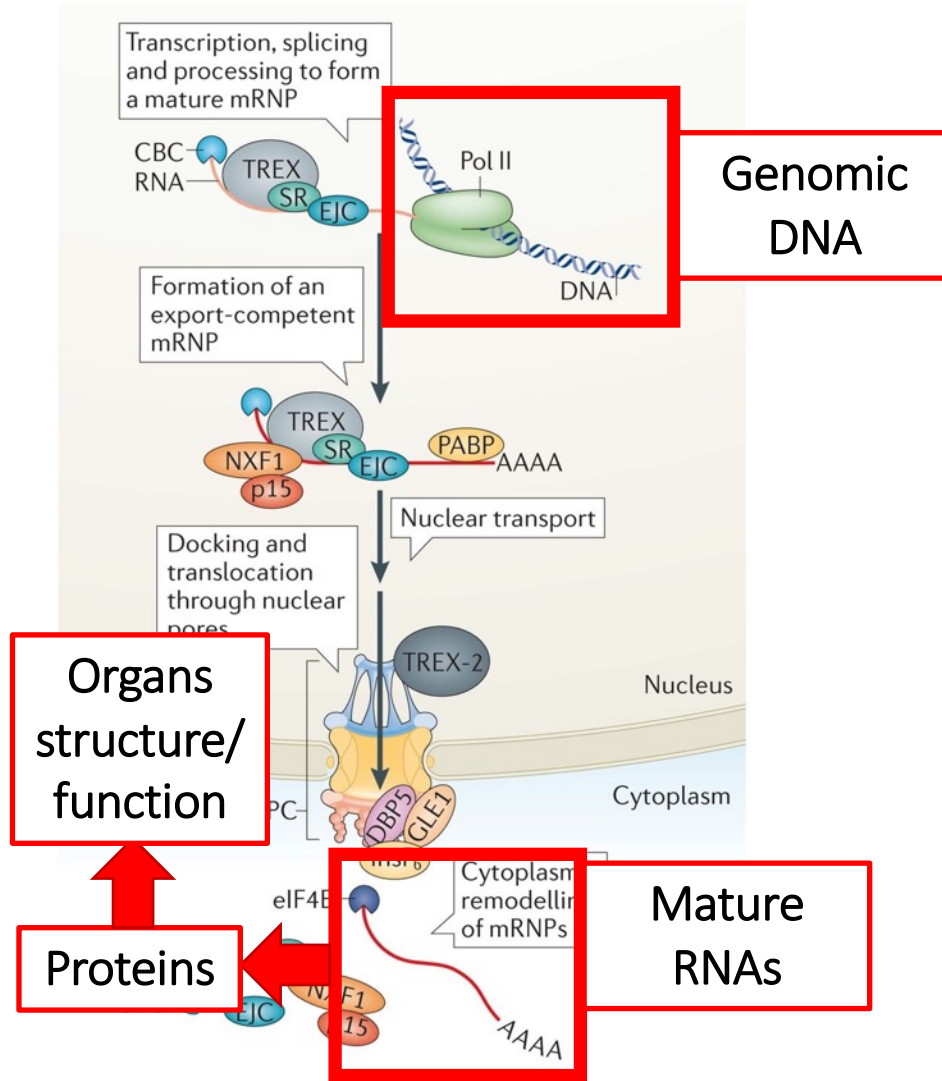


Today

- Introduction to RNAseq analysis
- Workshop data & biological context
- Working on the Hoffman2 server
- Raw sequence data:
 - General properties
 - Format
 - Quality controls, filtering

Introduction to RNAseq analysis

What is RNA-seq?



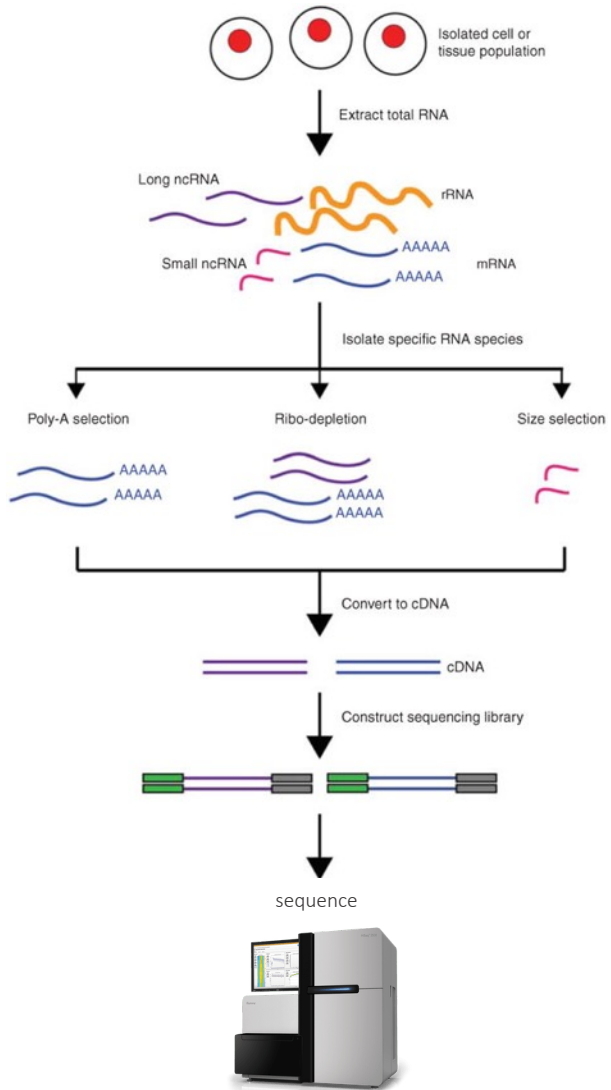
RNA-seq: Purposes

Functional genomics: high-throughput data on molecular phenotypes, such as:

epigenetics, transcriptomics (RNA-seq), proteomics, metabolomics...

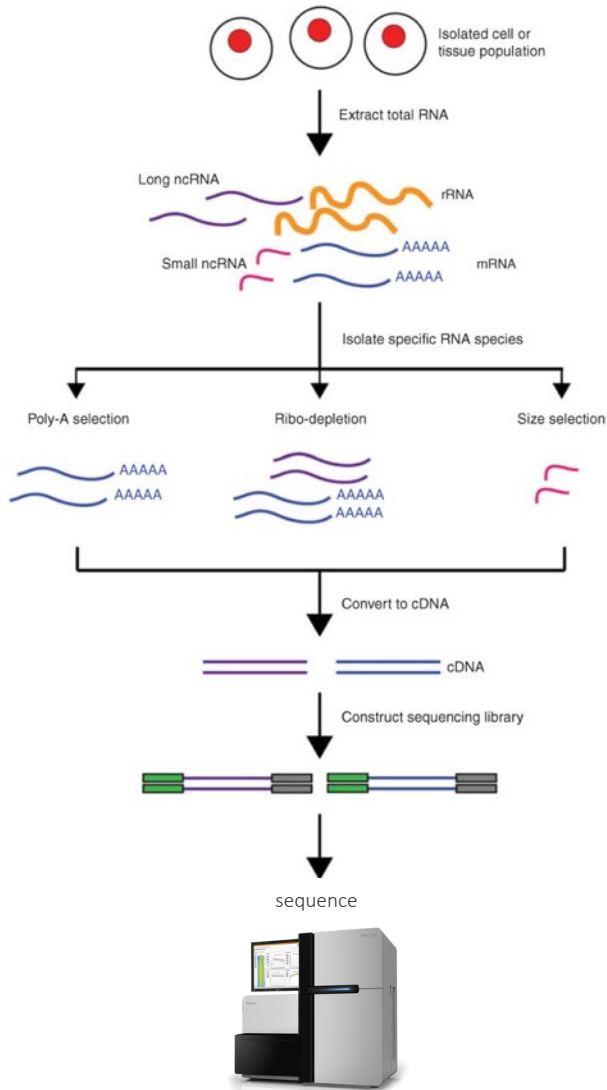
- Study of **genotype** $\leftrightarrow$ **phenotype** mapping
- Discovery/annotation of genes & transcripts
- **Which genes are expressed in which tissues & how much?**
- Gene **regulatory networks** (co-regulated genes)
- Expression variation in **response to environmental stimuli**
- Medical **diagnostic**

RNA-seq protocol

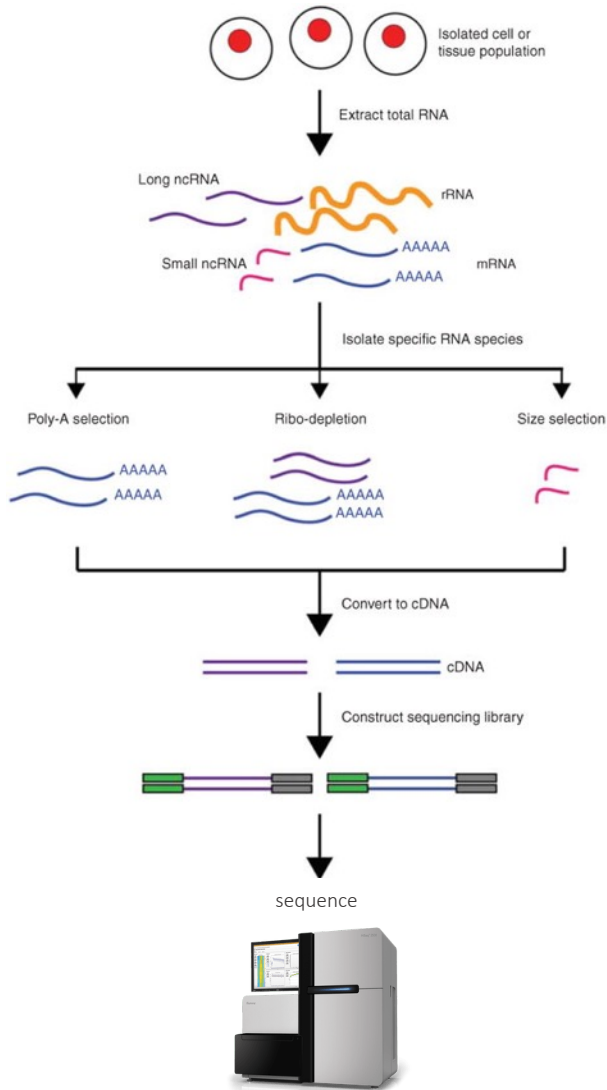


RNA-seq protocol

➤ Whole-tissue, single-cell, cell type-specific, nuclei...



RNA-seq protocol

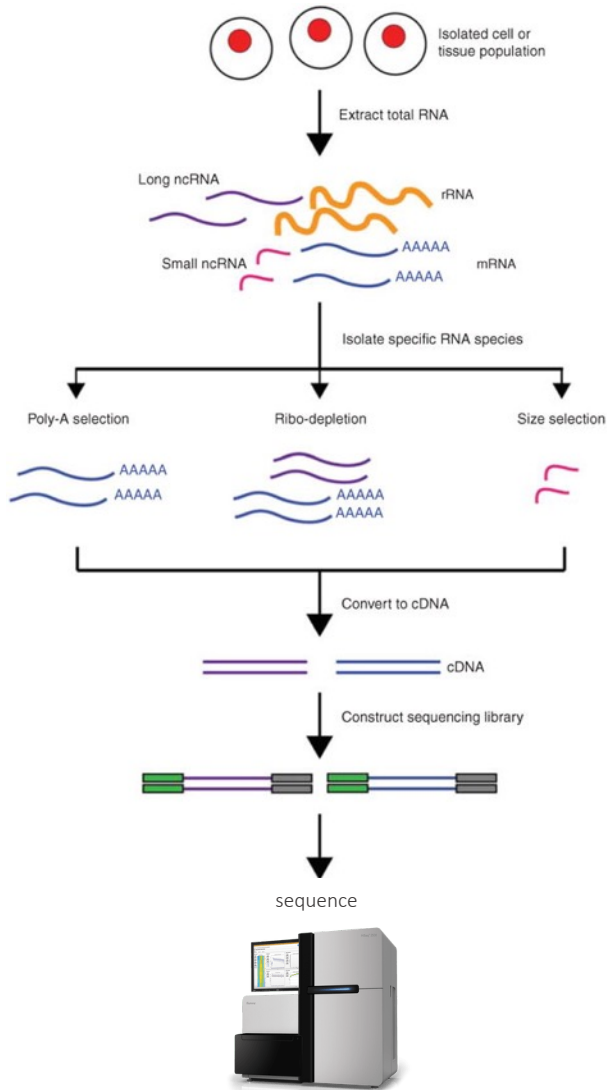


➤ RNA selection

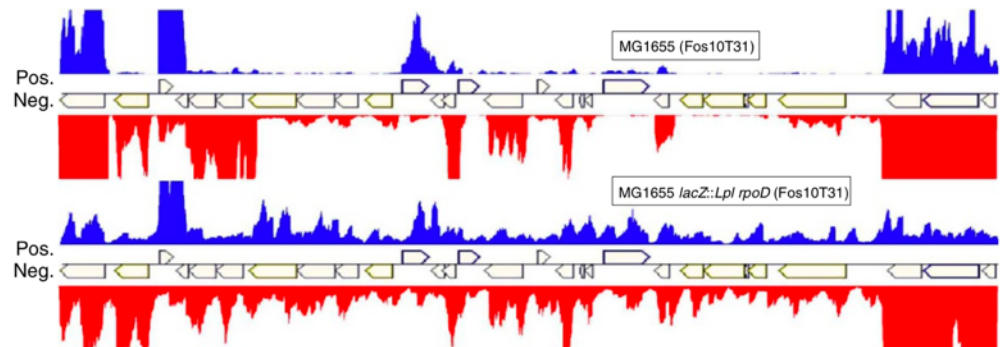
RNAs come in many flavors:

- ribosome-related RNAs:
rRNA, tRNA, snoRNA **Up to 90% of RNAs...**
- protein-coding RNAs:
mRNA (poly-adenylated), pre-mRNA
- regulatory RNAs:
miRNA, siRNA, piRNA, lncRNA, tdRNA...

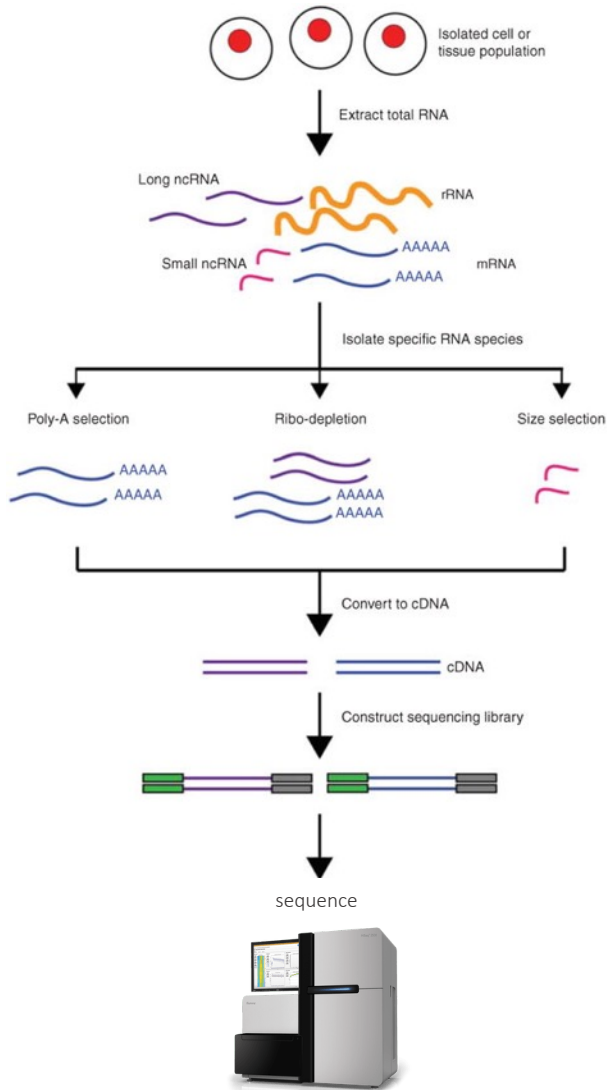
RNA-seq protocol



➤ Reverse transcription:
Stranded or unstranded



RNA-seq protocol



➤ Sequencing technology:

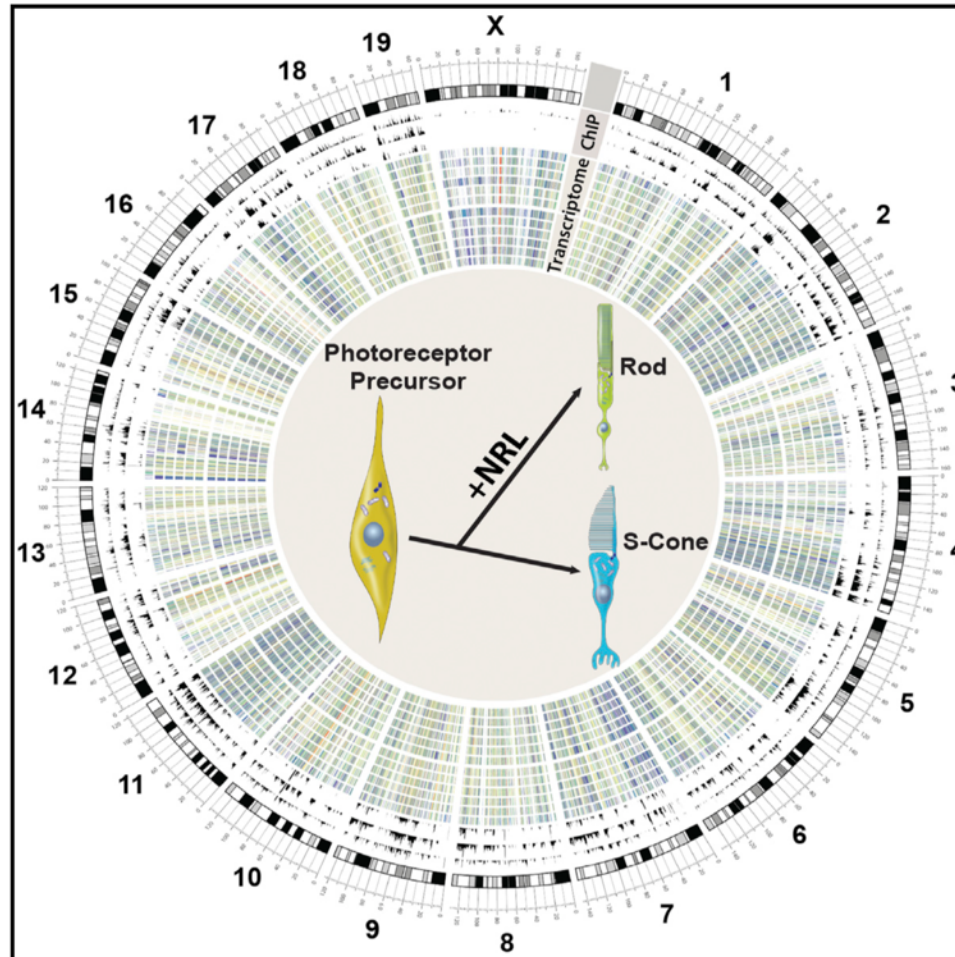
- Illumina: **short** reads; **single-end** or **paired-end**
- Pacbio, Nanopore: **long** reads

The Data

Kim JW et al. (2016), Dev Cell 37(6)

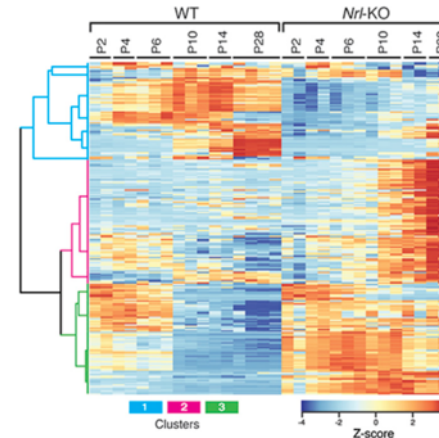
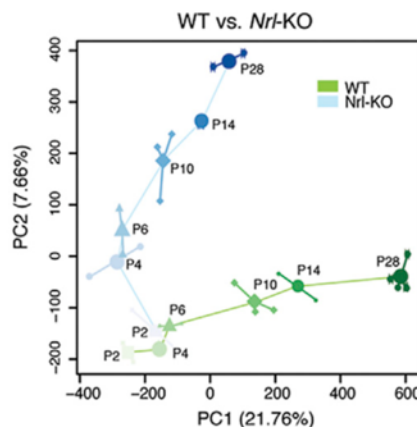
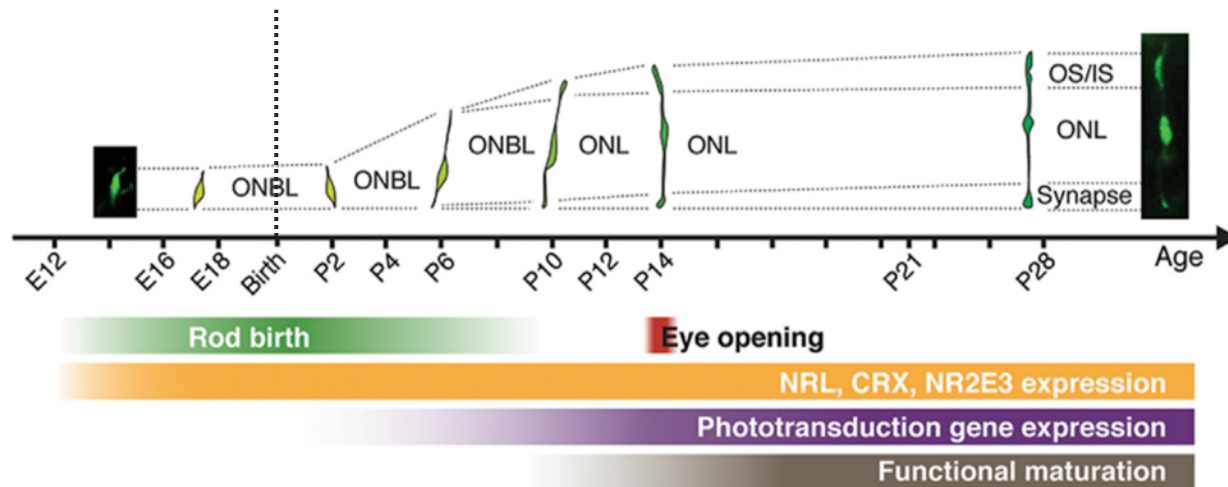
Kim JW et al. (2016) Cell Reports, 17(9)

Kim JW et al. (2016)
NRL-Regulated Transcriptome Dynamics of Developing Rod
Photoreceptors.



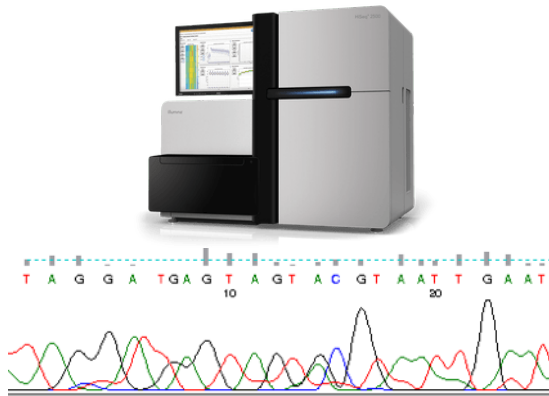
Kim JW et al. (2016). NRL-Regulated Transcriptome Dynamics of Developing Rod Photoreceptors. *Cell Reports*, 17(9).

- Analyzed the gene regulatory network activated by the NRL transcription factor, across developmental stages
- Gene expression with “stranded total RNAseq”



RNAseq workshops 1 & 2

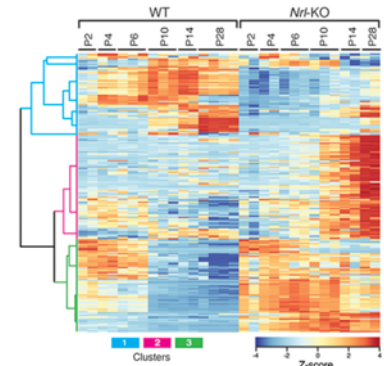
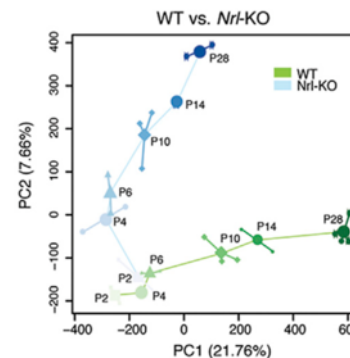
- RNAseq I: “from raw sequence data to gene counts”



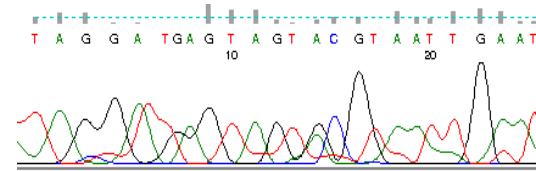
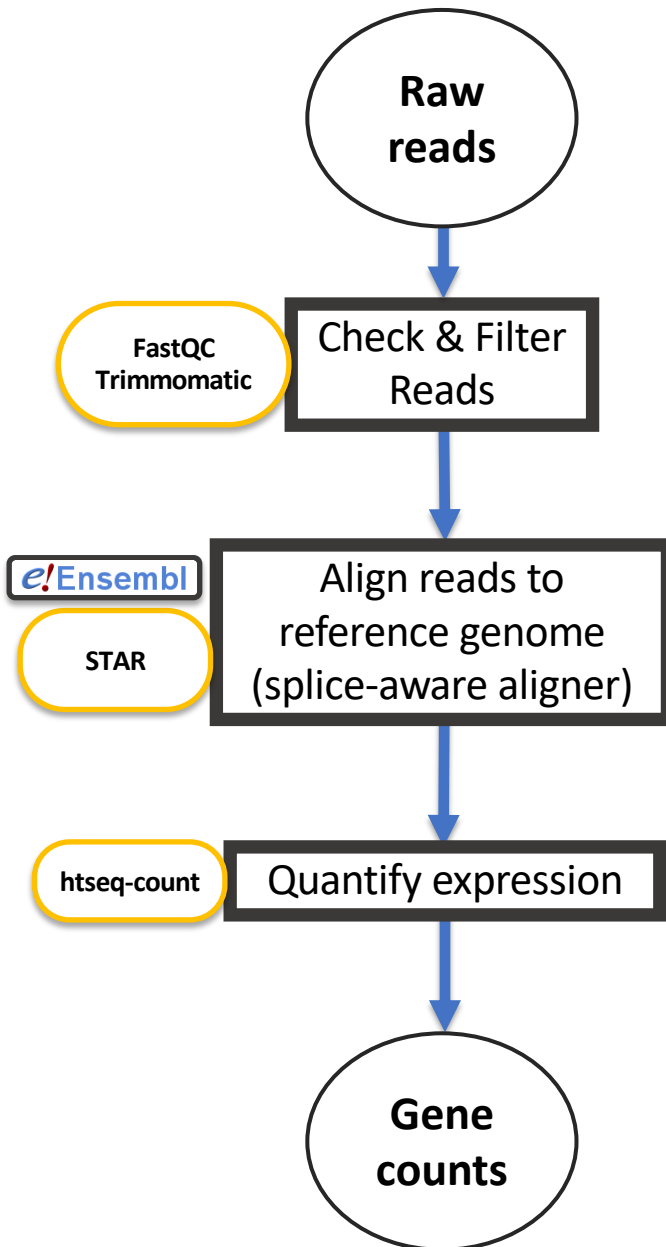
	sample1	sample2	sample3	...
gene1	999	701	616	
gene2	532	520	41	
gene3	14	36	305	
...				

- RNAseq II: “from gene counts to biology”

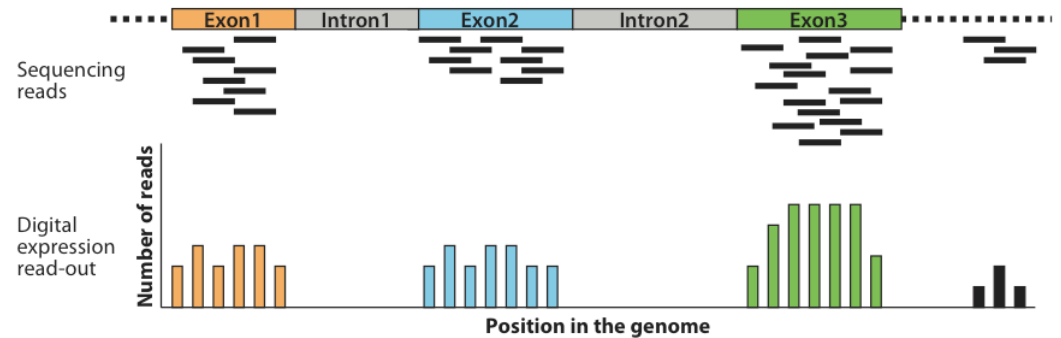
	sample1	sample2	sample3	...
gene1	999	701	616	
gene2	532	520	41	
gene3	14	36	305	
...				



RNA-seq analysis: Overview



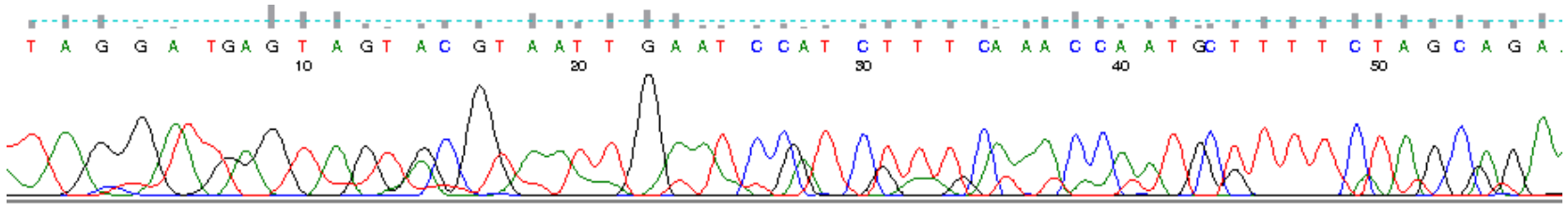
Reference genome sequence



	sample1	sample2	sample3	...
gene1	999	701	616	
gene2	532	520	41	
gene3	14	36	305	
...				

The raw output
of sequencing

What is a read?



Sanger

One chromatogram = 1 'read':

1. Sequence
2. Per-basepair confidence



Illumina

- Similar in principle...
...but higher-throughput & fully automated
- Typical RNAseq:
~10M reads of ~100bp each (=1Gbp's), per sample
⇒ Large files (even with compression)
e.g. for Kim et al. 2016, the raw data alone is 53 GB's



The FASTQ format

- “Plain text” format (could have a “.txt” extension)
- Usually compressed with GZIP (“.fastq.gz”)

```
@J00153:333:HF2VNBBXY:1:1101:1347:1086 1:N:0:NGCTCATT+NAGGACGT
ACCGGATTGCATCCCCAATCTGCTGTTTGTTCCTGCGTACGTTCC
+
9<AFAJAFJJJJJJJJ7AA-<FFF-<F-<FJJAJJJJJJJ-FFJFFJ<
@J00153:333:HF2VNBBXY:1:1101:2402:1086 1:N:0:NGCTCATT+NAGGACGT
CTGGTGGTGCCCTTCCGTCAATTCTTTAAGTTTCAGCTTTGCAACCAT
+
BAAFAJJFJJJJJJFJJJJ<A<<AJJFFJJJJJFJJJ-F7JJJJFJJJJ
(etc.)
```

ID
Seq
+
Qual

} Read 1
}
} Read 2

etc.

One read = 4 lines:

1. Read ID + properties (forward/reverse, barcodes)
2. Read sequence
3. (spacer line)
4. Base qualities according to the sequencer

Base qualities: “PHRED” scale

```
@J00153:333:HF2VNBBXY:1:1101:1347:1086 1:N:0:NGCTCATT+NAGGACGT  
ACCGGATTGCATCCCCAATCTGCTGTTTGTTCCTGCGTACGTTCC  
+  
9<AFAJAFJJJJJJJJ7AA-<FFF-<F-<FJJAJJJJJJJ-FFJFFJ<
```



Logarithmic scale to write confidence probabilities as integers:

$$Prob(WrongBase) = 10^{-Score*0.1}$$

$$Prob(CorrectBase) = 1 - Prob(WrongBase)$$

Score	Error rate	Confidence
10	$10^{-1} = 0.1$	90%
20	$10^{-2} = 0.01$	99%
30	$10^{-3} = 0.001$	99.9%
40	$10^{-4} = 0.0001$	99.99%

ASCII encoding of PHRED scores

ASCII codes (1-byte characters)

0	<NUL>	32	<SPC>	64	@	96	`
1	<SOH>	33	!	65	A	97	a
2	<STX>	34	"	66	B	98	b
3	<ETX>	35	#	67	C	99	c
4	<EOT>	36	\$	68	D	100	d
5	<ENQ>	37	%	69	E	101	e
6	<ACK>	38	&	70	F	102	f
7	<BEL>	39	'	71	G	103	g
8	<BS>	40	(	72	H	104	h
9	<TAB>	41	)	73	I	105	i
10	<LF>	42	*	74	J	106	j
11	<VT>	43	+	75	K	107	k
12	<FF>	44	,	76	L	108	l
13	<CR>	45	-	77	M	109	m
14	<SO>	46	.	78	N	110	n
15	<SI>	47	/	79	O	111	o
16	<DLE>	48	0	80	P	112	p
17	<DC1>	49	1	81	Q	113	q
18	<DC2>	50	2	82	R	114	r
19	<DC3>	51	3	83	S	115	s
20	<DC4>	52	4	84	T	116	t
21	<NAK>	53	5	85	U	117	u
22	<SYN>	54	6	86	V	118	v
23	<ETB>	55	7	87	W	119	w
24	<CAN>	56	8	88	X	120	x
25		57	9	89	Y	121	y
26	<SUB>	58	:	90	Z	122	z
27	<ESC>	59	;	91	[	123	{
28	<FS>	60	<	92	\	124	
29	<GS>	61	=	93	]	125	}
30	<RS>	62	>	94	^	126	~
31	<US>	63	?	95	_	127	

- Codes are standard since the 60s
- Lets us write letters, numbers, punctuation in one byte (ie. 8 bits, ie. a number in 0-255)

Examples:

- 72 is 'H'
- 101 is 'e'
- The byte sequence 72, 101, 108, 108, 111 is "Hello"
- **Conversely** — 'A' is 65.

ASCII codes 33-126 are proper text characters.

The **PHRED-33** encoding uses the 33-73 range to associate characters to scores 0 to 40 (confidences from 0 to 99.99%)

⇒ An 'A' quality (65) corresponds to: a PHRED score of 25, a confidence of 99.7%

FASTQ base qualities: summing up

```
@J00153:333:HF2VNBBXY:1:1101:1347:1086 1:N:0:NGCTCATT+NAGGACGT  
ACCGGATTGCATCCCCAATCTGCTGTTTGTTCCTGCGTACGTTCC  
+  
9<AFAJAFJJJJJJJJ7AA-<FFF-<F-<FJJAJJJJJJJ-FFJFFJ<
```



Qualities 9<AFAJA... correspond to:

- ASCII codes 57, 60, 65, 70, 74, 65
- Scores 24, 27, 32, 37, 41, 32 in PHRED-33
- Confidences of 99.60; 99.80; 99.94; 99.98; 99.99; 99.94 %

Exercise 1

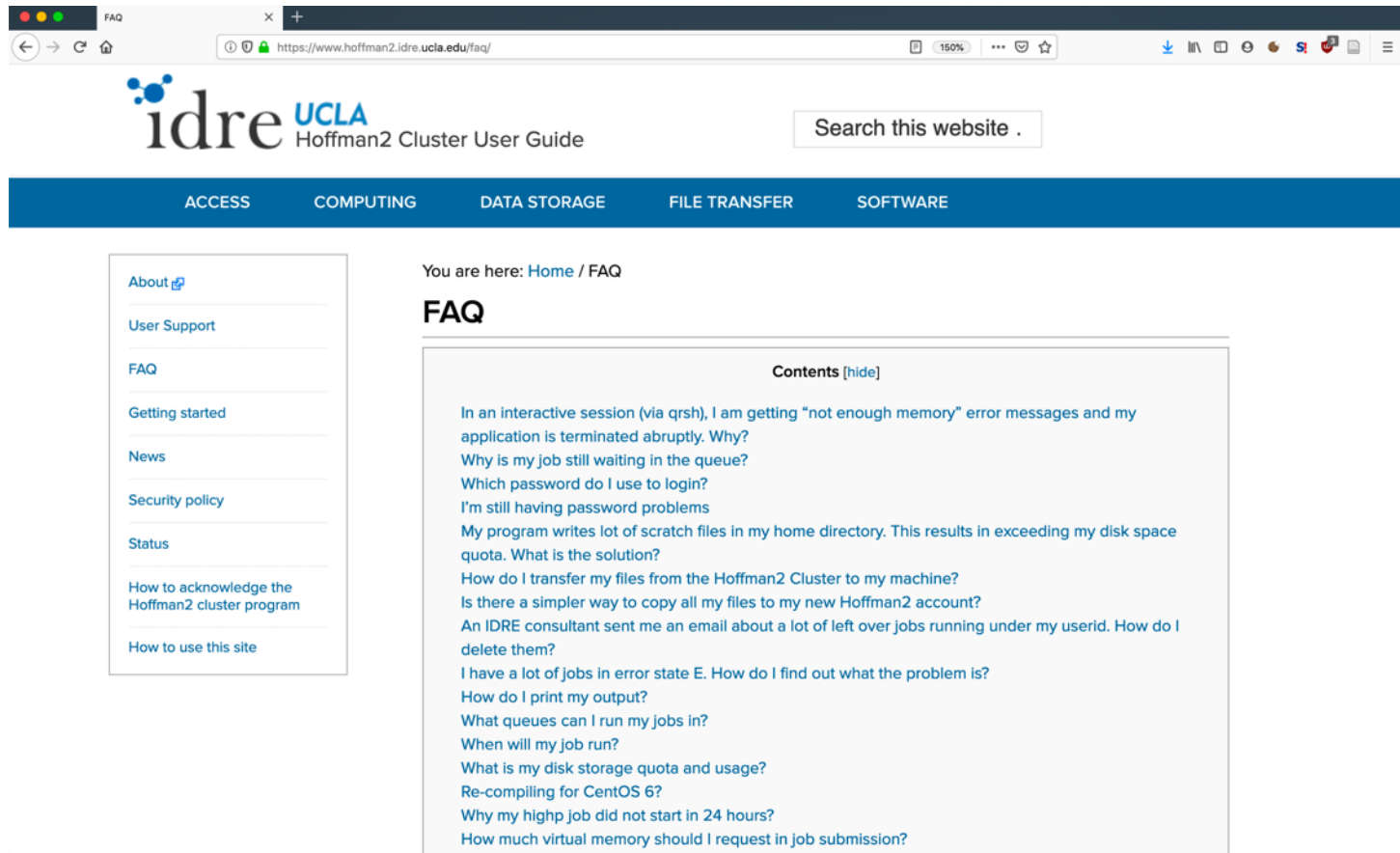
- The workshop data is in: `/u/home/n/nrochett/QCB_RNAseq`
- Copy this directory and its content to **your** home directory (note: this is only 5% of the raw data, for only two samples)
- **How many samples** are listed in `samples.txt` ?
- According to the `Kim2016.infoTable.tsv` table:
 - Has this RNAseq library been sequenced **single-end or paired-end**?
 - **How many samples** are there in the KO group (NRL^{-/-})? In the control group?
- For file `P10KO_rep1.five_percent.fastq.gz` :
 - **How many lines** are there?
 - **How many reads** does this represent?
 - **What is the confidence** for the first base call of the first read?
 - **How long** are the reads?

Working on the
Hoffman2 cluster

Hoffman2 Homepage & Resources

<https://www.hoffman2.idre.ucla.edu/>

Also, <https://idre.ucla.edu/>



The screenshot shows a web browser window displaying the Hoffman2 Cluster User Guide homepage. The browser's address bar shows the URL <https://www.hoffman2.idre.ucla.edu/faq/>. The page features the IDRE UCLA logo and the text "Hoffman2 Cluster User Guide". A search bar is located on the right side of the header. A blue navigation bar contains links for ACCESS, COMPUTING, DATA STORAGE, FILE TRANSFER, and SOFTWARE. On the left, a sidebar menu lists various resources: About, User Support, FAQ (selected), Getting started, News, Security policy, Status, How to acknowledge the Hoffman2 cluster program, and How to use this site. The main content area shows the breadcrumb "You are here: Home / FAQ" followed by the "FAQ" heading. Below this is a "Contents [hide]" section listing various troubleshooting questions and answers, such as "In an interactive session (via qssh), I am getting 'not enough memory' error messages and my application is terminated abruptly. Why?" and "Why is my job still waiting in the queue?".

FAQ

<https://www.hoffman2.idre.ucla.edu/faq/>

idre UCLA
Hoffman2 Cluster User Guide

Search this website .

ACCESS COMPUTING DATA STORAGE FILE TRANSFER SOFTWARE

About
User Support
FAQ
Getting started
News
Security policy
Status
How to acknowledge the Hoffman2 cluster program
How to use this site

You are here: Home / FAQ

FAQ

Contents [hide]

- In an interactive session (via qssh), I am getting "not enough memory" error messages and my application is terminated abruptly. Why?
- Why is my job still waiting in the queue?
- Which password do I use to login?
- I'm still having password problems
- My program writes lot of scratch files in my home directory. This results in exceeding my disk space quota. What is the solution?
- How do I transfer my files from the Hoffman2 Cluster to my machine?
- Is there a simpler way to copy all my files to my new Hoffman2 account?
- An IDRE consultant sent me an email about a lot of left over jobs running under my userid. How do I delete them?
- I have a lot of jobs in error state E. How do I find out what the problem is?
- How do I print my output?
- What queues can I run my jobs in?
- When will my job run?
- What is my disk storage quota and usage?
- Re-compiling for CentOS 6?
- Why my highp job did not start in 24 hours?
- How much virtual memory should I request in job submission?

Connecting to the cluster

- Remote shell

ssh USER@hoffman2.idre.ucla.edu

- (through WSL/Putty/Mobaxterm on Windows)
- Once connected, start an 'interactive job' requesting 4 hours and 4GB memory

qssh -l h_rt=4:00:00,h_data=4G

- Remote copy/File transfer

Download:

scp USER@hoffman2.idre.ucla.edu:/PATH/TO/FILE ~/Desktop/

Upload:

scp ~/Desktop/FILE USER@hoffman2.idre.ucla.edu:~/

- Interactive clients exist:
Cyberduck, FileZilla...

Essential Hoffman2 commands

- Request an interactive job:

qrsh

(Do not run computationally heavy commands on the login node!)

- Submit a batch/script job:

qsub -N JOBNAME script.sh

- Main resource options for jobs:

Time: **-l h_rt=4:00:00**

Memory / RAM: **-l h_data=4G**

- List current jobs: **myjobs** (or **qstat -u \$USER**)

Main job states are **r** (for running) and **qw** (for queued-waiting)

- Modules:

module load NAME

module avail

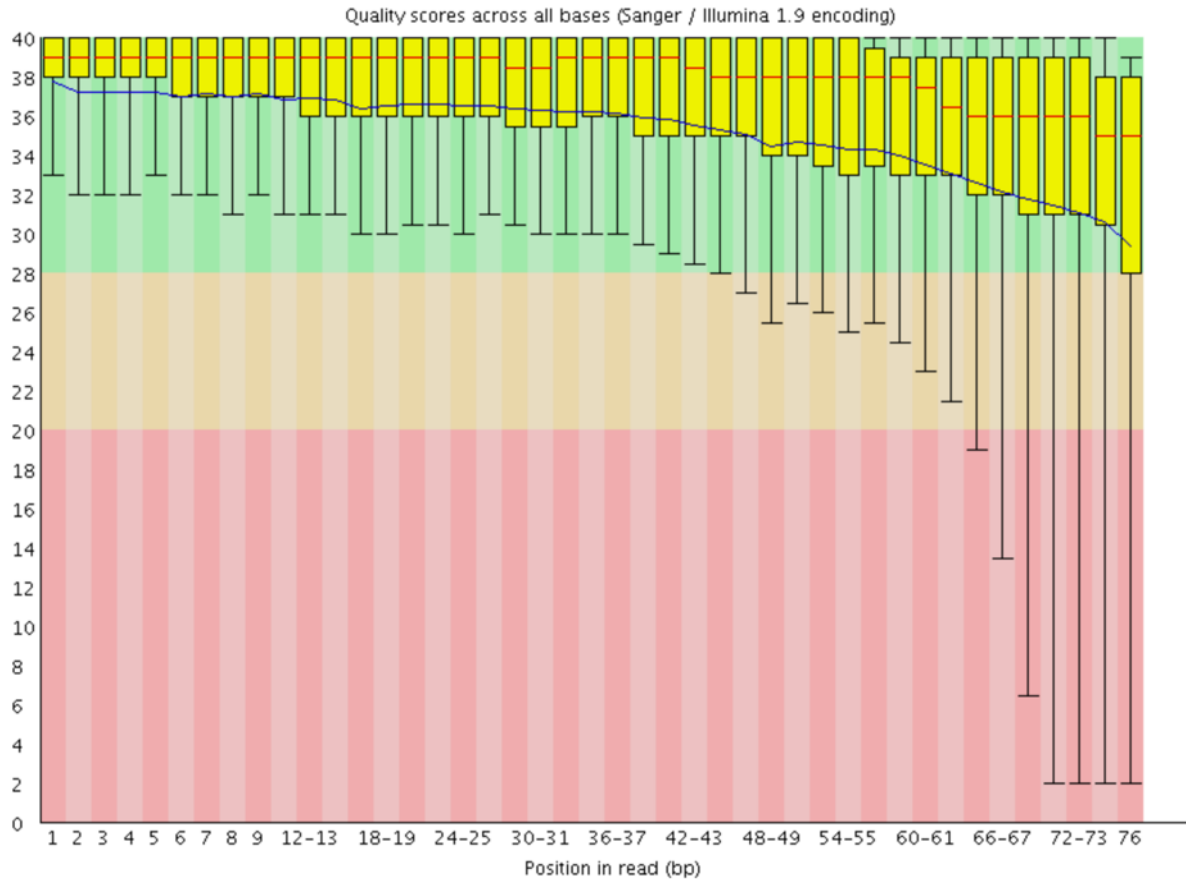
Hoffman2 disk spaces

- Home directory (20Gb max.):
/u/home/?/USER
- Scratch (2Tb max., 'temporary'):
/u/flashscratch/?/USER
N.B. Scratch data is considered abandoned after two weeks & it is not backuped
- Purchased space (cf. IDRE website):
/u/project/LABNAME/USERNAME

Checking Read qualities with FastQC

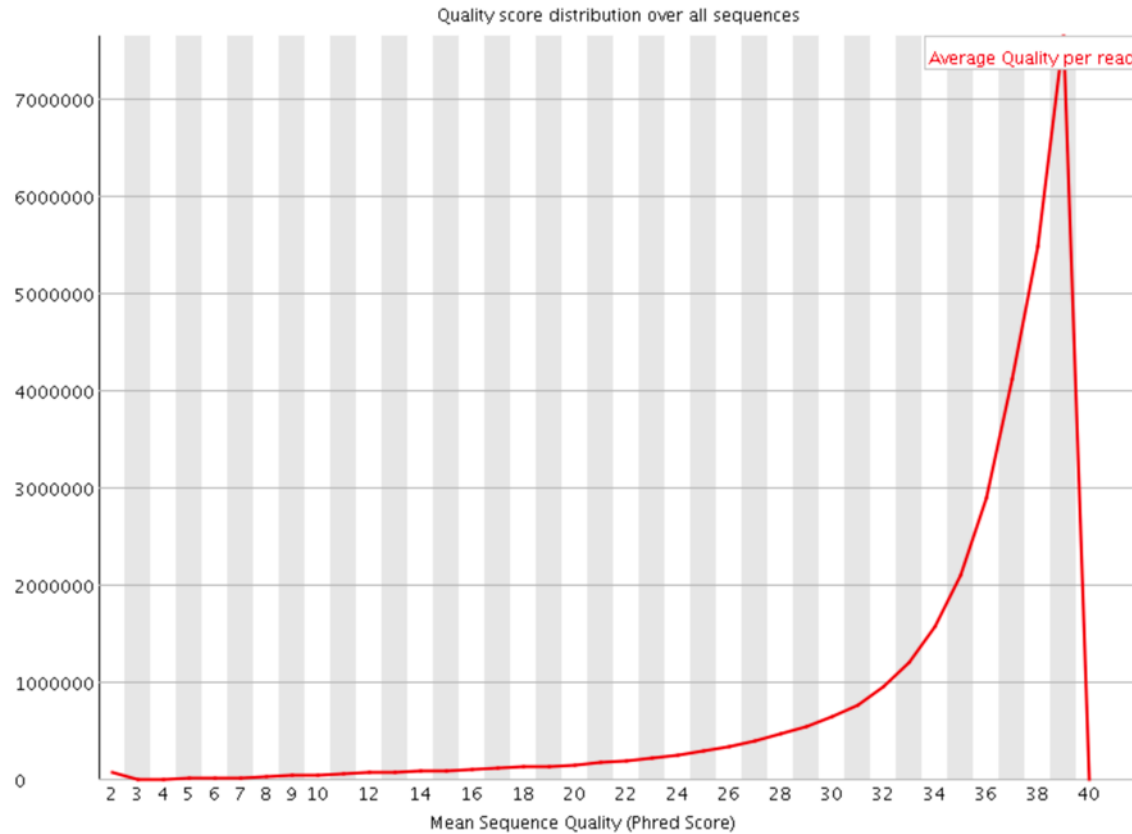
FastQC: base qualities

✓ Per base sequence quality

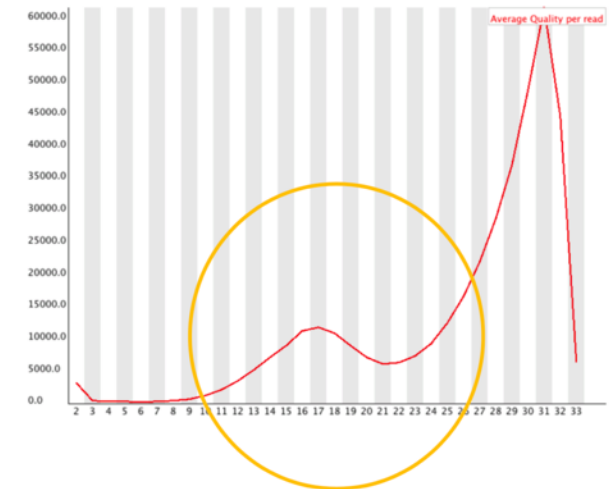


FastQC: per-read mean base qualities

✓ Per sequence quality scores

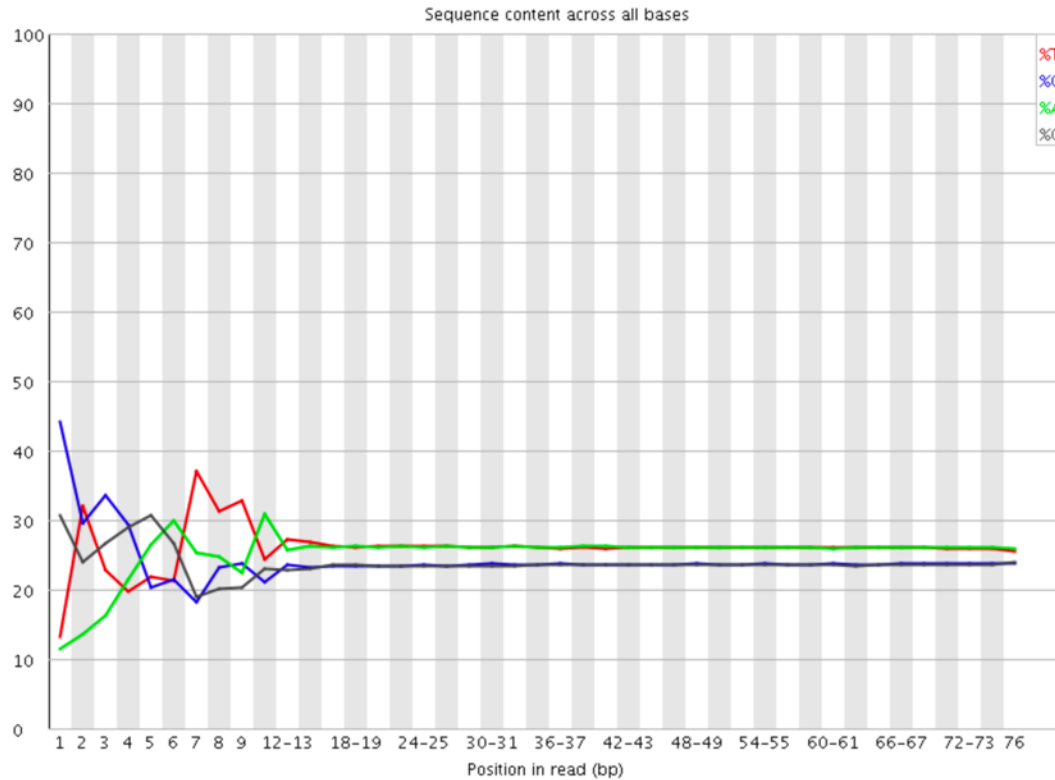


(bad)



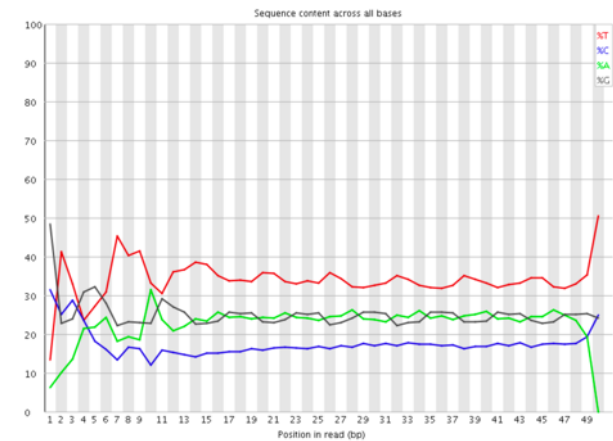
FastQC: %ACGT over read length

⚠ Per base sequence content



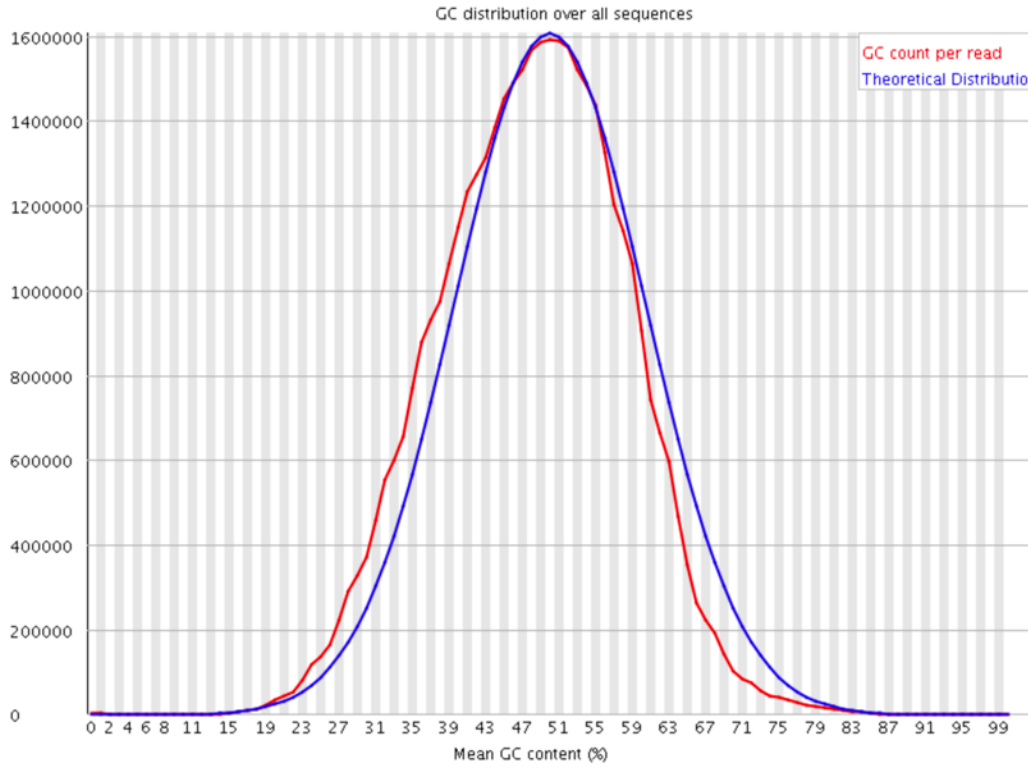
(bad)

⚠ Per base sequence content



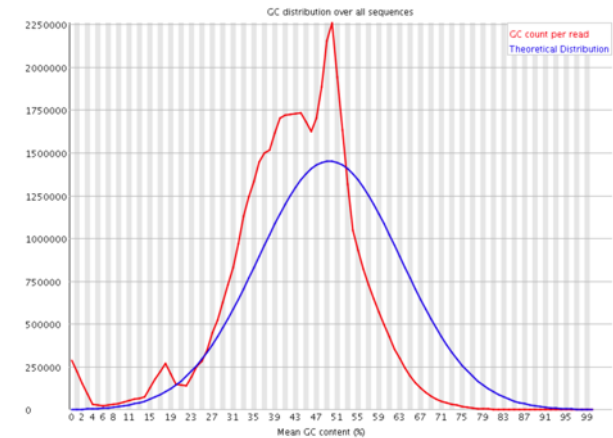
FastQC: per-read %GC

✅ Per sequence GC content



(bad)

❌ Per sequence GC content

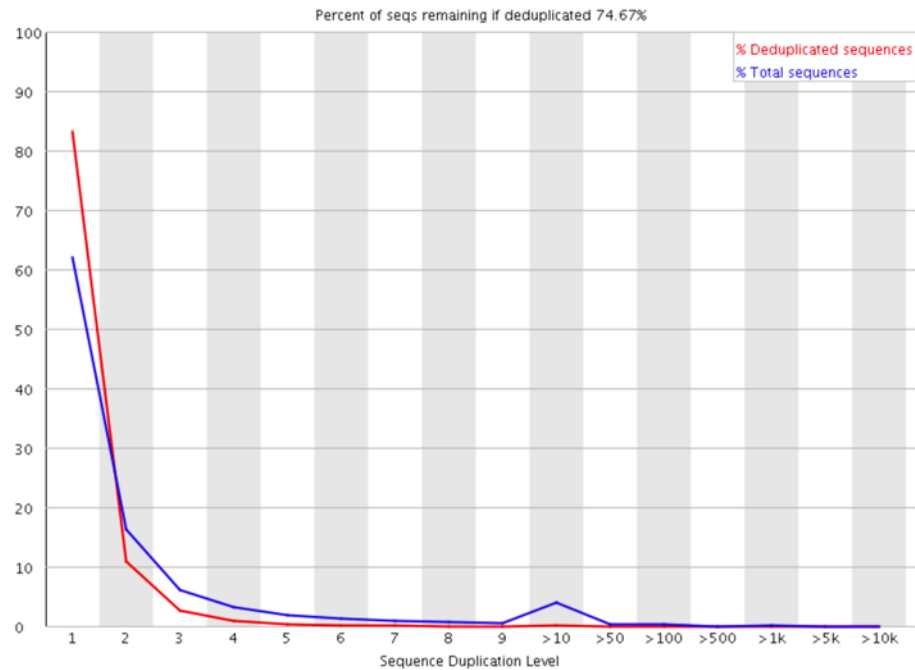


FastQC: over-represented sequences

- A little over-representation is normal in RNAseq (highly expressed genes)



Sequence Duplication Levels



Exercise 2: let's run FastQC

- (Installation demonstration)
- Run FastQC on **P10KO_rep1.five_percent.fastq.gz**
 - You will need to specify the full path to the program!
 - For most programs the **-h** option asks for the 'help'
- Download the HTML results file (*e.g.* with SCP)
& review it in your web browser

FastQC: code

- Usage is given by:

```
./programs/fastqc/fastqc --help
```

- We can run it with just:

```
./programs/fastqc/fastqc P10KO_rep1.five_percent.fastq.gz
```

- To download the HTML file (from *your* computer):

```
scp USER@hoffman2.idre.ucla.edu:~/QCB_RNAseq/P10KO_rep1.five_percent.fastq.gz ~/
```

(Or with Cyberduck/etc.)



6. Read trimming & filtering with 'Trimmomatic'

Exercise 3: let's run Trimmomatic

- Website & manual
- Run it for sample P10KO_rep1 and/or sample P10WT_rep1

```
jar_file=~ /QCB_RNAseq/programs/trimmomatic/trimmomatic-0.39.jar  
java -jar $jar_file <ARGUMENTS...>
```

- Ask for:
 - for minimum leading & trailing base qualities of 20 (**LEADING** and **TRAILING** parameters)
 - for clipping when the base quality over 5 bases drops under 20 (**SLIDINGWINDOW** parameter)
 - a minimum final length of 60 (**MINLEN** parameter)
 - *(for simplicity let's not use the ILLUMINACLIP parameter; otherwise we would have used the file TrueSeq2-SE to match the library preparation)*
- How many reads passed filters? What proportion of the raw reads have been discarded?

Trimmomatic: code

- According to the online “quick start”, the usage for single-end data is:

```
java -jar trimmomatic-0.35.jar SE -phred33 input.fq.gz  
output.fq.gz ILLUMINACLIP:TruSeq3-SE:2:30:10 LEADING:3  
TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36
```

- We adapt it to (we also need to specify to specify a single thread/processor using `-threads 1`, as the default is to use several):

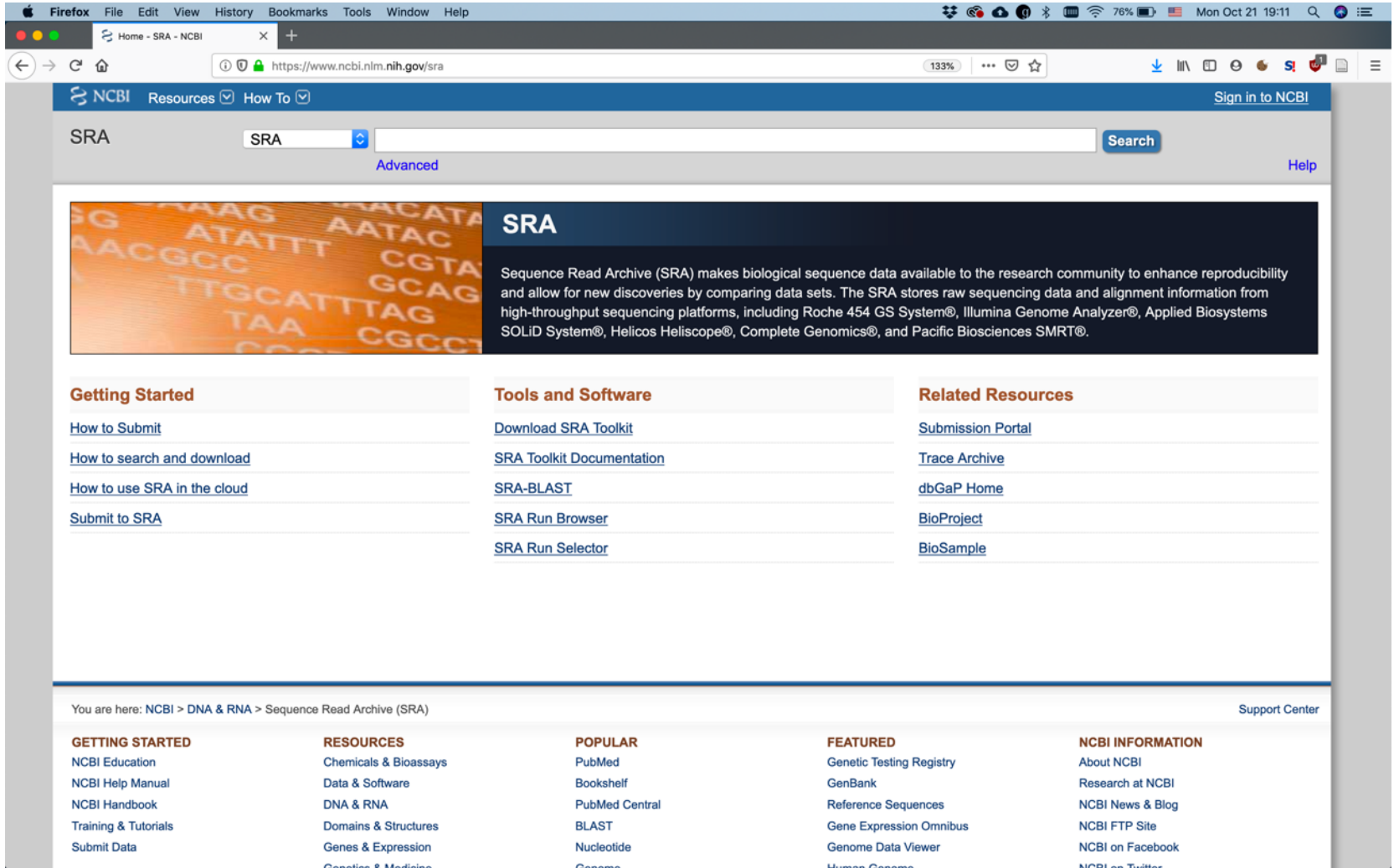
```
jar_file=~ / QCB_RNAseq / programs / trimmomatic / trimmomatic-0.39.jar
```

```
java -jar $jar_file \  
    SE -phred33 -thread 1 \  
    P10KO_rep1.five_percent.fastq.gz \  
    P10KO_rep1.five_percent.trimmo.fastq.gz \  
    LEADING:20 TRAILING:20 SLIDINGWINDOW:5:20 MINLEN:60
```

(**Note:** a backslash `\` at the end of a line means “continued line”.)

the NCBI SRA (Sequence Read Archive)

<https://www.ncbi.nlm.nih.gov/sra>



The screenshot shows the NCBI SRA website in a Firefox browser. The browser's address bar displays the URL <https://www.ncbi.nlm.nih.gov/sra>. The website's header includes the NCBI logo, navigation links for 'Resources' and 'How To', and a 'Sign in to NCBI' button. Below the header, there is a search bar with 'SRA' entered and a 'Search' button. The main content area features a large banner with the text 'SRA' and a description: 'Sequence Read Archive (SRA) makes biological sequence data available to the research community to enhance reproducibility and allow for new discoveries by comparing data sets. The SRA stores raw sequencing data and alignment information from high-throughput sequencing platforms, including Roche 454 GS System®, Illumina Genome Analyzer®, Applied Biosystems SOLiD System®, Helicos Heliscope®, Complete Genomics®, and Pacific Biosciences SMRT®.' To the left of this banner is an image showing DNA sequence data. Below the banner, the page is organized into three columns: 'Getting Started' with links like 'How to Submit', 'How to search and download', 'How to use SRA in the cloud', and 'Submit to SRA'; 'Tools and Software' with links like 'Download SRA Toolkit', 'SRA Toolkit Documentation', 'SRA-BLAST', 'SRA Run Browser', and 'SRA Run Selector'; and 'Related Resources' with links like 'Submission Portal', 'Trace Archive', 'dbGaP Home', 'BioProject', and 'BioSample'. At the bottom, a footer section contains 'You are here: NCBI > DNA & RNA > Sequence Read Archive (SRA)', a 'Support Center' link, and five columns of links under the headings 'GETTING STARTED', 'RESOURCES', 'POPULAR', 'FEATURED', and 'NCBI INFORMATION'.

Firefox File Edit View History Bookmarks Tools Window Help

Home - SRA - NCBI

<https://www.ncbi.nlm.nih.gov/sra>

133%

NCBI Resources How To Sign in to NCBI

SRA SRA Advanced Help

SRA

Sequence Read Archive (SRA) makes biological sequence data available to the research community to enhance reproducibility and allow for new discoveries by comparing data sets. The SRA stores raw sequencing data and alignment information from high-throughput sequencing platforms, including Roche 454 GS System®, Illumina Genome Analyzer®, Applied Biosystems SOLiD System®, Helicos Heliscope®, Complete Genomics®, and Pacific Biosciences SMRT®.

Getting Started

- [How to Submit](#)
- [How to search and download](#)
- [How to use SRA in the cloud](#)
- [Submit to SRA](#)

Tools and Software

- [Download SRA Toolkit](#)
- [SRA Toolkit Documentation](#)
- [SRA-BLAST](#)
- [SRA Run Browser](#)
- [SRA Run Selector](#)

Related Resources

- [Submission Portal](#)
- [Trace Archive](#)
- [dbGaP Home](#)
- [BioProject](#)
- [BioSample](#)

You are here: NCBI > DNA & RNA > Sequence Read Archive (SRA) Support Center

GETTING STARTED

- NCBI Education
- NCBI Help Manual
- NCBI Handbook
- Training & Tutorials
- Submit Data

RESOURCES

- Chemicals & Bioassays
- Data & Software
- DNA & RNA
- Domains & Structures
- Genes & Expression
- Genetics & Medicine

POPULAR

- PubMed
- Bookshelf
- PubMed Central
- BLAST
- Nucleotide
- Genome

FEATURED

- Genetic Testing Registry
- GenBank
- Reference Sequences
- Gene Expression Omnibus
- Genome Data Viewer
- Human Genome

NCBI INFORMATION

- About NCBI
- Research at NCBI
- NCBI News & Blog
- NCBI FTP Site
- NCBI on Facebook
- NCBI on Twitter

NCBI SRA & BioProjects

eg. project **PRJNA301098**

Firefox File Edit View History Bookmarks Tools Window Help

Mus musculus (ID 301098) - B X +

https://www.ncbi.nlm.nih.gov/bioproject/PRJNA301098 133% ... ☆

NCBI Resources How To Sign in to NCBI

BioProject BioProject Search

Advanced Browse by Project attributes Help

Display Settings: Send to:

Mus musculus (house mouse) Accession: PRJNA301098 ID: 301098

Recruitment of Rod Photoreceptors from Short Wavelength Sensitive Cones during the Evolution of Nocturnal Vision in Mammals

Vertebrate ancestors had only cone-like photoreceptors. The duplex retina evolved in jawless vertebrates with the advent of highly photosensitive rod-like photoreceptors. [More...](#)

See [Genome](#) Information for Mus musculus

Related information

- BioProject
- BioSample
- Full text in PMC
- Genome
- GEO DataSets
- PubMed
- SRA
- Taxonomy
- Umbrella projects

Recent activity [Turn Off](#) [Clear](#)

- Mus musculus BioProject
- SRP065767 (30) SRA
- Contrasting Frequencies and Effects of cis- and trans-Regulatory Mutations Affect PubMed
- Contrasting[Title] AND frequencies[Title] AND effects[Title] AND ... (1) PubMed
- Contrasting frequencies and effects of cis- and trans-regulatory ... (0) PubMed

[See more...](#)

Project Data:

Resource Name	Number of Links
SEQUENCE DATA	
SRA Experiments	30
PUBLICATIONS	

Project Details Table:

Accession	PRJNA301098; GEO: GSE74660
Data Type	Transcriptome or Gene expression
Scope	Multiisolate
Organism	Mus musculus [Taxonomy ID: 10090] Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Myomorpha; Muroidea; Muridae; Murinae; Mus; Mus; Mus musculus
Publications	Kim JW et al. , "Recruitment of Rod Photoreceptors from Short-Wavelength-Sensitive Cones during the Evolution of Nocturnal Vision in Mammals.", <i>Dev Cell</i> , 2016 Jun 20;37(6):520-32
Submission	Registration date: 4-Nov-2015 NNRL, NEI, NIH
Relevance	Model Organism

NAVIGATE UP

This project is a component of the Recruitment of Rod Photoreceptors from Short Wavelength Sensitive Cones during the Evolution of Nocturnal Vision in Mammals

NAVIGATE ACROSS

2 additional projects are components of the Recruitment of Rod Photoreceptors from Short Wavelength Sensitive Cones during the Evolution of Nocturnal Vision in Mammals.

31960 additional projects are related by organism.

Downloading from the SRA Via “SRA run selector”

The screenshot shows the SRA Run Selector interface. The top section, 'Common Fields', displays metadata for a sample: BioProject (PRJNA301098), Consent (PUBLIC), Assay Type (RNA-Seq), AssemblyName (GCF_000001635.20), cell_type (GFP positive retina cells), Center Name (GEO), DATASTORE filetype (SRA), DATASTORE provider (GS, NCBI, S3), and DATASTORE region (gs.US, ncbl.public, s3.us-east-1). Below this is a 'Select' table with columns for Runs, Bytes, Bases, and Download. It shows a total of 30 runs (34.74 Gb, 88.14 Gb) and 0 selected runs. A table of 30 items is displayed below, with columns for Run, BioSample, AvgSpotLen, Experiment, Genotype, GEO_Accession, MBases, MBytes, and Sample Name. The table lists 8 runs, each corresponding to a specific sample and library.

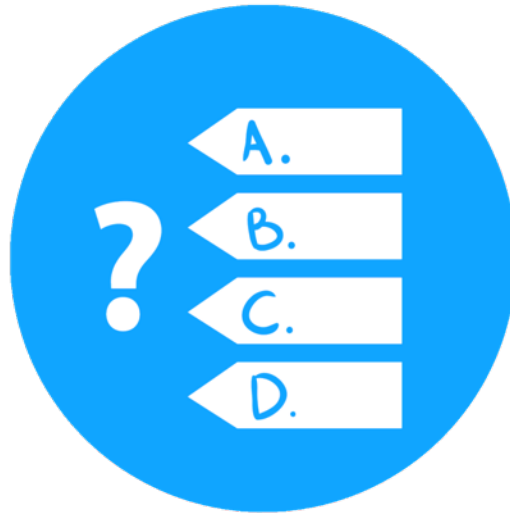
Run	BioSample	AvgSpotLen	Experiment	Genotype	GEO_Accession	MBases	MBytes	Sample Name
SRR2936836	SAMN04235754	76	SRX1411331	Nrlp-GFP	GSM1924968	2416	966	GSM1924968
SRR2936837	SAMN04235755	76	SRX1411332	Nrlp-GFP	GSM1924969	2572	1153	GSM1924969
SRR2936838	SAMN04235756	76	SRX1411333	Nrlp-GFP	GSM1924970	2463	1206	GSM1924970
SRR2936839	SAMN04235757	76	SRX1411334	Nrlp-GFP	GSM1924971	2317	1122	GSM1924971
SRR2936840	SAMN04235758	76	SRX1411335	Nrlp-GFP	GSM1924972	2562	998	GSM1924972
SRR2936841	SAMN04235759	76	SRX1411336	Nrlp-GFP	GSM1924973	2292	995	GSM1924973
SRR2936842	SAMN04235760	76	SRX1411337	Nrlp-GFP	GSM1924974	2327	1150	GSM1924974
SRR2936843	SAMN04235761	76	SRX1411338	Nrlp-GFP	GSM1924975	2807	1047	GSM1924975

Information on
samples & libraries

Sequence files
Each “Run” (SRR
identifiers) is one
sample

Reads for every sample were retrieved using the SRA toolkit, eg.:
`fastq-dump --gzip $SRR_ID`

Quizzes



<https://tinyurl.com/...>

RNA-seq analysis: Overview

