## Bioinformatics Lessons Schedule

Date	Subject
01-07	Process RNA-seq
01-14	Process RNA-seq, continued
01-21	Process RNA-seq, continued
01-28	Analyze RNA-seq
02-04	Analyze RNA-seq, continued
02-11	Analyze RNA-seq, continued
02-18	skipped
02-25	Process RRBS
03-03	Process RRBS, continued
03-10	Process RRBS, continued
03-17	Analyze RRBS
03-24	Analyze RRBS, continued
03-31	???

1. Everyone works through R for Data Science to hone their R skills with help available if they have questions.

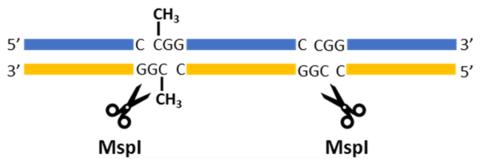
- 1. Everyone works through R for Data Science to hone their R skills with help available if they have questions.
- 2. Meetup 1 or a few people show something casually (could be less frequent than every week)
  - Code review
  - Demonstrations of cool code or packages
  - How people have used the code they've learned

- 1. Everyone works through R for Data Science to hone their R skills with help available if they have questions.
- 2. Meetup 1 or a few people show something casually (could be less frequent than every week)
  - Code review
  - Demonstrations of cool code or packages
  - How people have used the code they've learned
- 3. Nothing

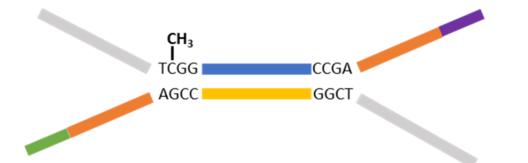
# Process Reduced Representation Bisulfite Sequencing (RRBS)

2020-02-24

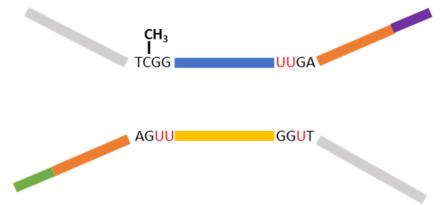
### 1. Mspl digestion



- 2. End repair
- 3. Adaptor Ligation and size selection

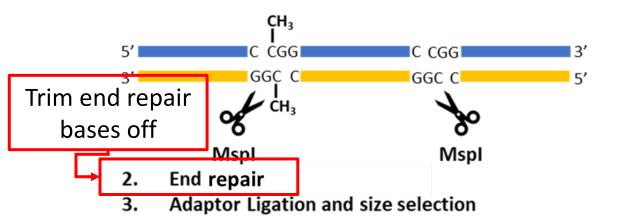


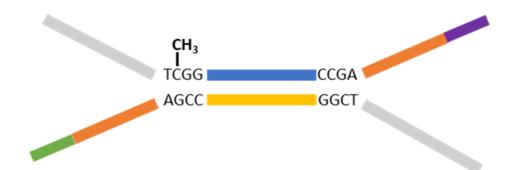




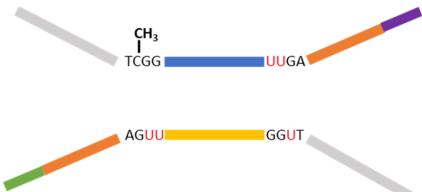
- 5. PCR amplification
- 6. Library preparation & sequencing

## 1. Mspl digestion



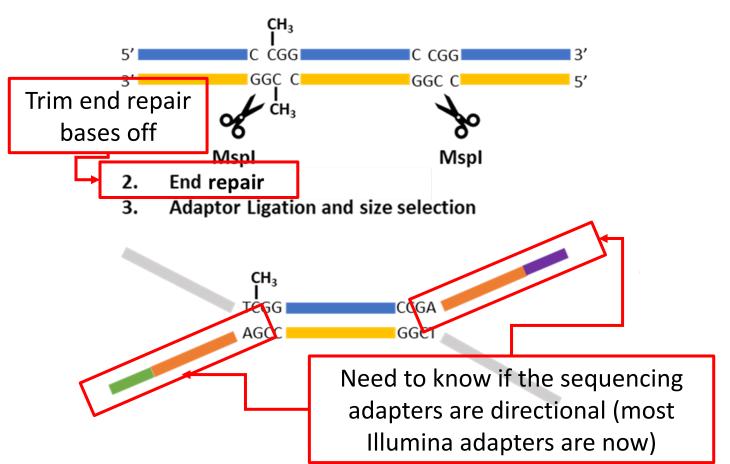


#### 4. Bisulfite conversion

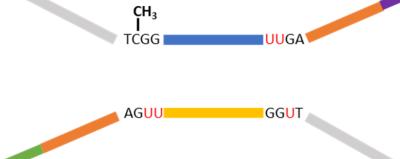


- 5. PCR amplification
- 6. Library preparation & sequencing

### 1. Mspl digestion

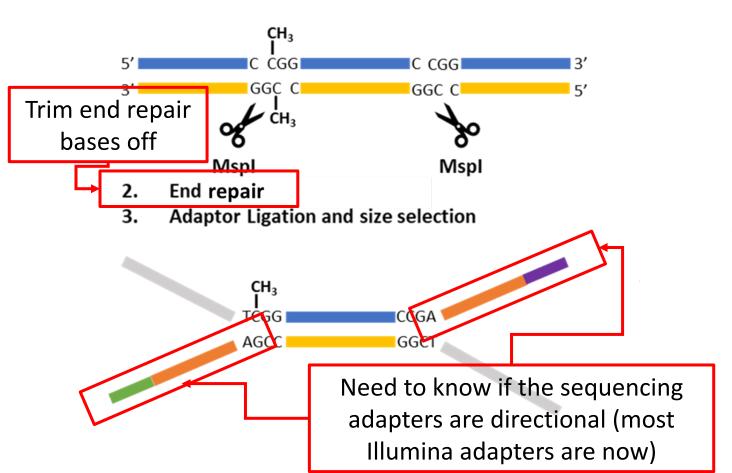


4. Bisulfite conversion



- 5. PCR amplification
- 6. Library preparation & sequencing

### 1. Mspl digestion



- A Bisulfite conversion

  CH<sub>3</sub>

  Need to align twice, to the regular genome and a C-to-T converted version of the genome
- 5. PCR amplification
- 6. Library preparation & sequencing

## The Data

- Blood from healthy controls from the National Institute of Neurological Disorders and Stroke (NINDS) repository here at Coriell
- Woonbok did the RRBS as part of a larger study to identify CpGs that are good biomarkers of aging (we've talked about it at length in lab meetings)
- I subsampled 10,000 reads from every file so we could run the commands quickly

File Prefix	Sex	Age	
F22_sub10000	female	22	
F24_sub10000	female	24	
F25_sub10000	female	25	
F77_sub10000	female	77	
F78_sub10000	female	78	
F80_sub10000	female	80	

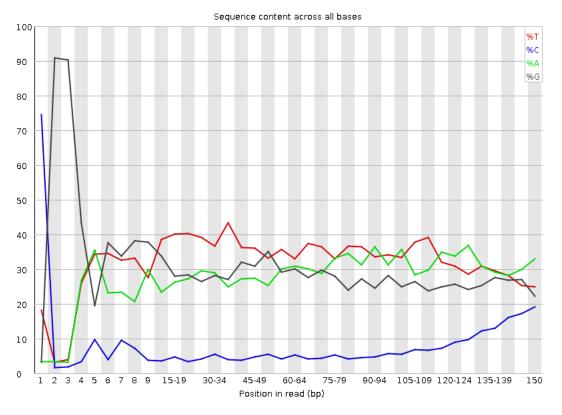
# Step 1: Quality Check w/ FastQC

fastqc \*.fq.qz -o fastqc/

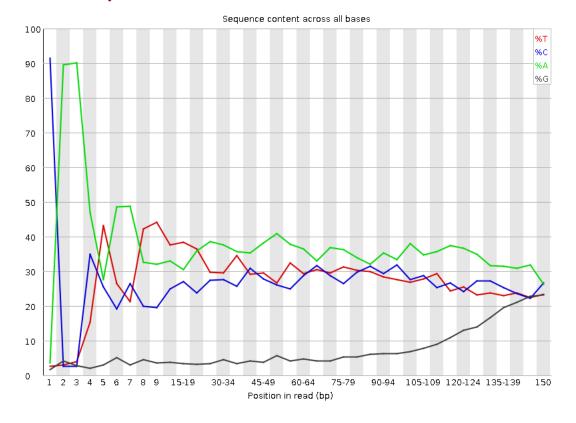
# FastQC: Per base sequence content

Read 1 Read 2





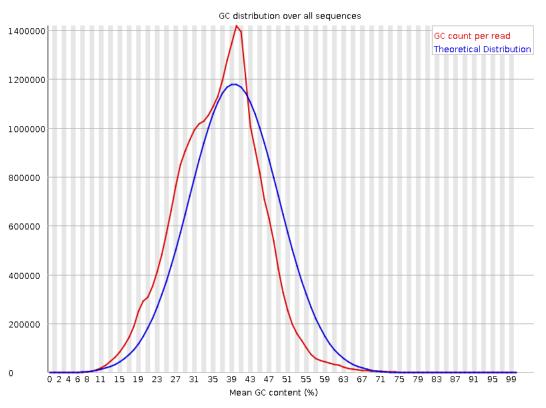
## ②Per base sequence content



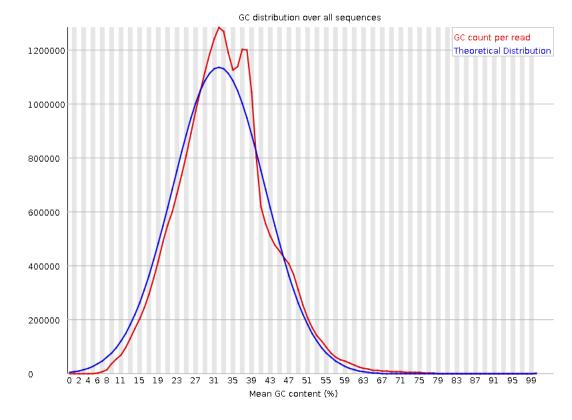
# FastQC: Per sequence GC content

Read 1 Read 2





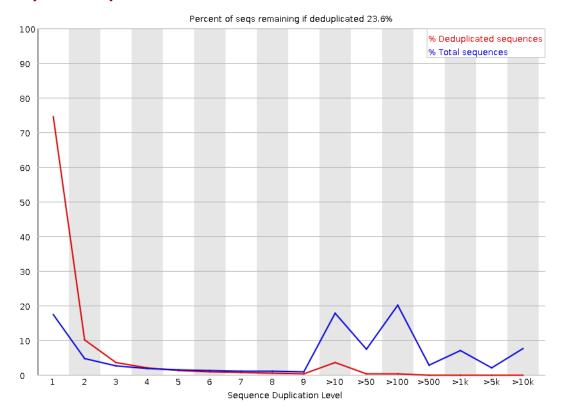
## Per sequence GC content



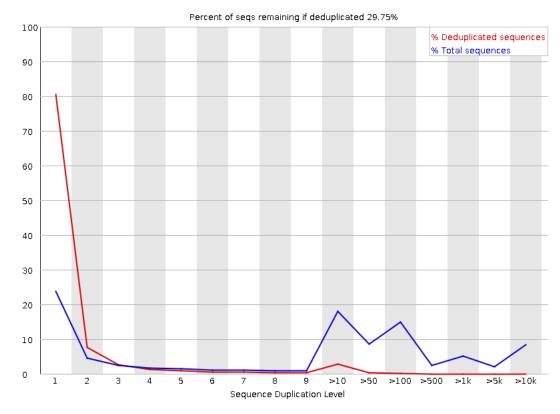
# FastQC: Per sequence GC content

Read 1 Read 2

## **Sequence Duplication Levels**



## Sequence Duplication Levels



# FastQC: Overrepresented sequences

Read 1 Read 2

## **Overrepresented sequences**

Sequence	Count	Percentage	Possible Source
Sequence	Count	reiceillage	rossible source
${\tt CGGAATAGAATGGAATGGAATGGAATGGAATGGAATGGA$	285907	1.008042814151108	No Hit
$\tt CGGGCGGGTGGTTTACGTTTGTAATTTTAGTATTTTGGGAGGTCGAGGC$	207030	0.7299405184682567	No Hit
${\tt CGGGTGGAGTGGAATGGAATGTAATGGAGTGGAATGTAATGGAATTTAGT}$	96288	0.3394895070389388	No Hit
${\tt CGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGA$	93849	0.3308901498223804	No Hit
${\tt CGGAATAGAATGGAATGGAATGGAATGGAATGGAATGGA$	81960	0.2889722498848395	No Hit
${\tt CGGAATGGAATGGAATGGAATGGAATGGAATGGAATTAAT}$	70703	0.2492826376721304	No Hit
${\tt CGGAACGGAATGGAATGGAATGGAATTAATTCGATTGTAATGGAA}$	65183	0.22982037779701675	No Hit
${\tt CGGTGGATTTTCGGTTTAAGTTTTGGTAATACGGTGAAATTTCGTTTTA}$	61728	0.2176388365164882	No Hit
${\tt CGGAATAGAATGGAATGGAATGGAATGGAATGGAATGGA$	53744	0.18948907513190352	No Hit
${\tt CGGAATGGAATGGAATGGAATGGAATGAAATGTAATGGATTTAAT}$	49586	0.17482891633467118	No Hit
${\tt CGGGCGGGTGGTTTACGTTTGTAATTTTAGTATTTTGGGAGGTCGAGGT}$	47280	0.16669848675640814	No Hit
${\tt CGGGAGGCGGAGTTGTAGTGAGTCGAGATCGCGTTATTGTATTTTAGTT}$	41059	0.1447646609080237	No Hit
${\tt CGGAATGGAATGGAATGGAATTAATTTTATTGTAATGGAATGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGAATGGAATGGAATGGAATGGAATGGAATGAATGGAATGGAATGGAATGA$	31979	0.11275065372214838	No Hit
${\tt CGGGAGGCGGAGTTTGTAGTGAGTCGAGATTTCGTTATTGTATTTTAGTT}$	31331	0.11046595364985244	No Hit
${\tt CGGAATAGAATGGAATGGAATGGAACGGAATGGAATGGA$	29976	0.10568853297398667	No Hit
${\tt CGGAATGGATTGGAATAAAACGGATTCGAATGTAAAGTATTGTAA}$	28528	0.10058321552848587	No Hit

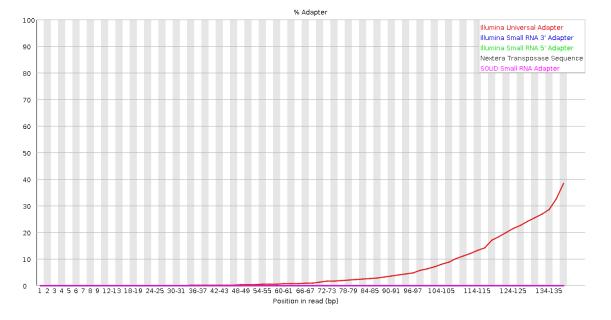
## **Overrepresented sequences**

Sequence	Count	Percentage	Possible Source
${\tt CAAATAATTCCATTCCATTACATTCCATTCCATTCCCTACACTC}$	627432	2.2121819996308516	No Hit
${\tt CAAATAATTCCATTCCATTACATTCCGTTCCATTACATTCCCCTACACTC}$	120581	0.42514107934802137	No Hit
${\tt CAACTAATTTTTATATTTTTAATAAAAACGAAATTTCACCGTATTAACC}$	112378	0.3962191739575218	No Hit
${\tt CAAAAAATTCCATTCCATTACATTCCATTCCATTCCCTACACTC}$	96550	0.3404132592286634	No Hit
${\tt CAACCTATTTATCTATTTATTAACTTTAAATCCAAATTATAAAACCAATT}$	85170	0.300289977094824	No Hit
${\tt CAACTAATTTTTATATTTTAATAAAAACGAAATTTCACCTTATTAACC}$	68317	0.24087014635654683	No Hit
${\tt CAAATAATTCCATTCCATTAAATTCCATTCCATTCCAT$	56942	0.20076449308129002	No Hit
${\tt CAAACGCGATAACTCACGCCTATAATCCCAACACTTTAAAAAAACCGAAAC}$	50011	0.17632736931418627	No Hit
${\tt CAAATTCACGCCATTCTCCTACCTCAACCTCCCGAATAACTAAAACTACA}$	45832	0.1615931693109073	No Hit
${\tt CAAAAAATTCCATTCCATTACATTCCATTCCATTCCAT$	39358	0.13876732321824686	No Hit
${\tt CAACTAATTTTTATATTTTTAATAAAAACGAAATTTCACCATATTAACC}$	36925	0.13018912063198754	No Hit
${\tt CAAATAATTCCATTCCATTACATTCCATTCCATTCCCCAACACTC}$	34825	0.12278500002732473	No Hit
${\tt CAAATAATCAATACTATATCCGACAATATTCGTATTTTATTTTCTTCCTA}$	32826	0.11573698236602904	No Hit
${\tt CAATTAATATTTACTATCCTTCATCATAACATCTTATTTTCTAAAAATCC}$	32734	0.11541261136811048	No Hit
${\tt CAAATACTATAAACTTTTTCTTTAACTTTTTACTATTTCTTTCT}$	28468	0.1003716692254955	No Hit

# FastQC: Adapter Content

Read 1 Read 2





## **⊘**Adapter Content

