

# lab seminar

Session2

JUN SOUNG KWAK

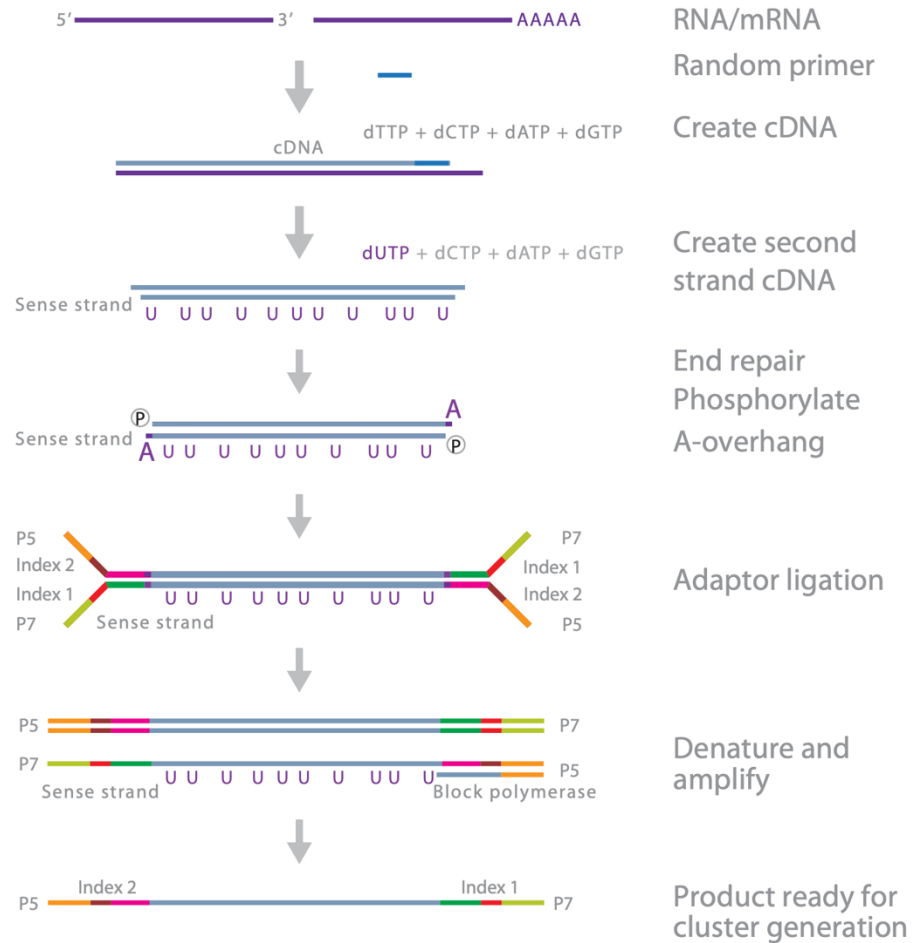


# General NGS workflow

1. library preparation
2. Sequencing library
3. Raw sequencing data
4. Alignment
5. Quantification or variant calling
6. Bioinformatics analysis



# TruSeq™ Stranded RNA

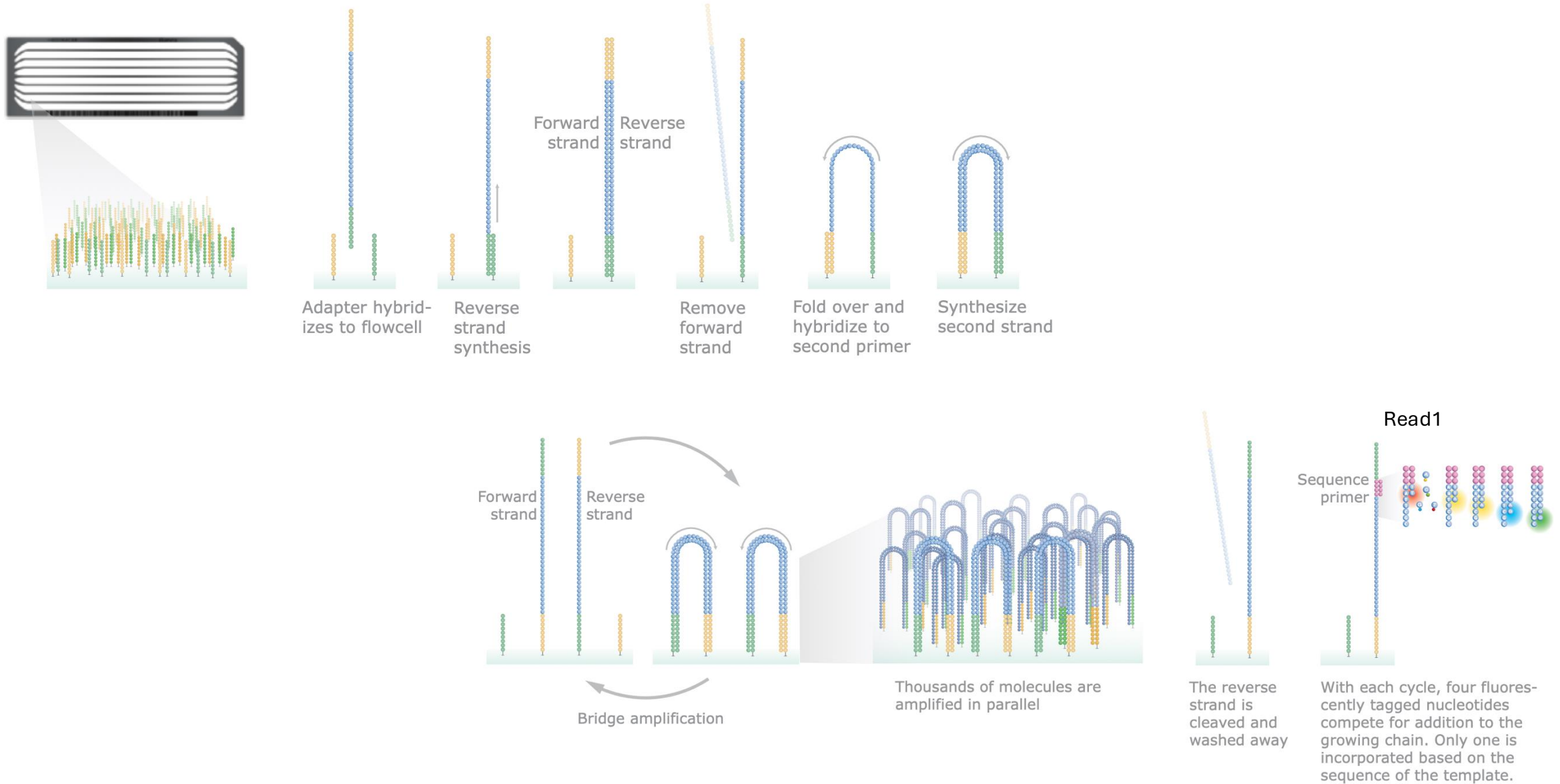


common sequence between Rd1 SP and Rd2 SP

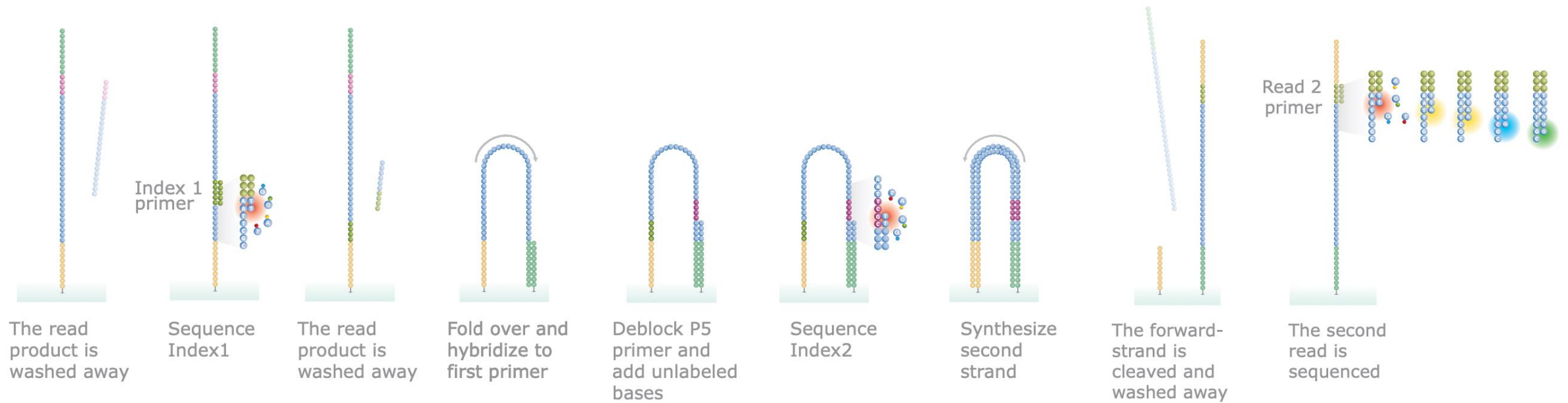
5' – AATGATACGGCGACCACCGAGATCTACACCAATTAACACACTCTTTCCCTACACGACGCTCTTCCGATCT **INSERT** AGATCGGAAGAGCACACGTCTGAACTCCAGTCACATATCTCGATCTCGTATGCCGTCTTCTGCTTG –3'

P5 Index Rd1 SP Rd2 SP Index P7

# Sequencing by synthesis



# RNA-seq output

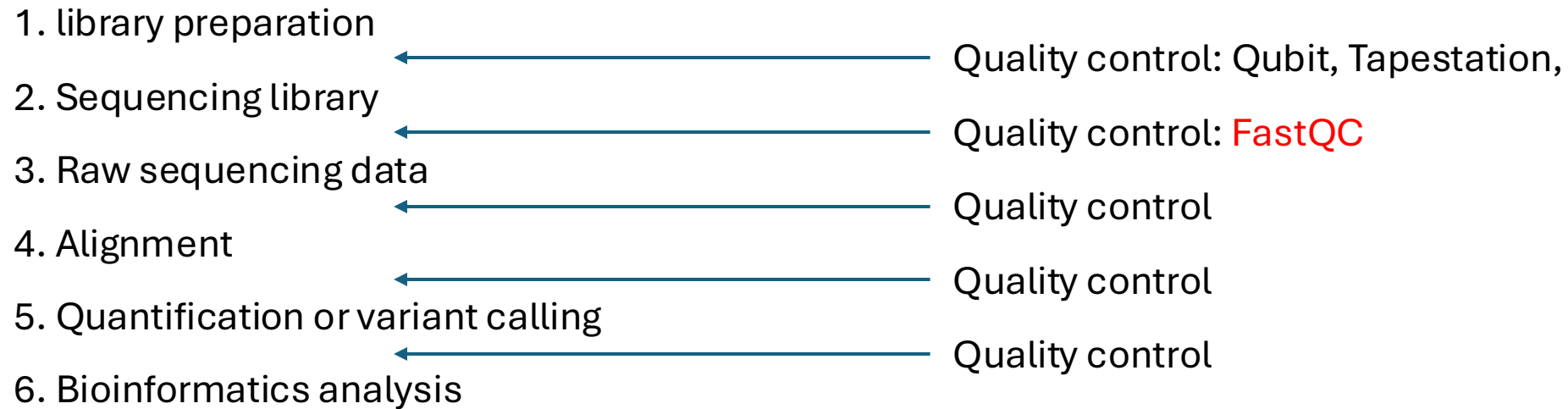


common sequence between Rd1 SP and Rd2 SP

5' – AATGATACGGCGACCACCGAGATCTACACCAATTAACACACTCTTTCCCTACACGACGCTCTTCCGATCT **INSERT** TAGATCGGAAGAGCACACGTCTGAACTCCAGTCACATATCTCGATCTCGTATGCCGTCTTCTGCTTG -3'

P5                      Index                      Rd1 SP                      Rd2 SP                      Index                      P7

# General NGS workflow



# FASTQ files

Line1: Sequence identifier  
Line2: Raw sequence  
Line3: Quality Score indicator  
Line4: Quality Score

```
Junsoungs-Mac-mini:bin jun$ head C1_R1.fastq
@A00939:243:HJTTFDSX3:4:1101:4327:1000 1:N:0:CAATTAAC+ATATCTCG
GTCCTCGTCTCCAGCTCCCTTACCCTGGGACTTCAACTGCGCTGCCTTCTTCACCTCCTCGAACGCAGCCAAGTCCACGGCCATGCCTTTCTCCTCTGCGA
+
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
@A00939:243:HJTTFDSX3:4:1101:5647:1000 1:N:0:CAATTAAC+ATATCTCG
ATCAAAAACAGCAAGTAGCCGGCGGTCAGATCCAGAATGAGGCCGCTCCGTGCACCACCAGCAGACTCACTAACTCCACAGGAAGAATCACTTTGAACG
+
FFFFFFFF:FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
@A00939:243:HJTTFDSX3:4:1101:10673:1000 1:N:0:CAATTAAC+ATATCTCG
CTCTACTACATTACACTTACAGCTCATACGTTTGTATGTATATACAGTAAAGAACAGACATGAGTAGATTCTGTTTGGATCAGTTCAAGCCAGCGATCTTA
Junsoungs-Mac-mini:bin jun$ head C1_R2.fastq
@A00939:243:HJTTFDSX3:4:1101:4327:1000 2:N:0:CAATTAAC+ATATCTCG
CAGCTTCTGCAGTGGCTAGCGGAGGTGGAGTGCCAGCAGGAAGTATGGTTTCTGCTTTTCAATCAAAGGTGCGGCCGCGTTGAAGGAGCTGTTGTTCTCG
+
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFF,FFFFFFFFFFFFFF:FFFFFFFFFFFFFFFFFFFFFF,FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
@A00939:243:HJTTFDSX3:4:1101:5647:1000 2:N:0:CAATTAAC+ATATCTCG
TGAATCAAAAATTGGATGCTGATTGGGTGGAGGGATACTCCATGTCCTACCTGGCACATCACTGGCTTTTGGATCCGTTCAAAGTGATTCTTCCTGTGG
+
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
@A00939:243:HJTTFDSX3:4:1101:10673:1000 2:N:0:CAATTAAC+ATATCTCG
ATTTTACATGTGAAAATCTGTAAATAAGATCGCTGGCTTGAAGTATCCAAACAGAATCTACTCATGTCTGTTCTTTACTGTATATACATACAAACGTAT
```

# FASTQ file QC (Quality Control)



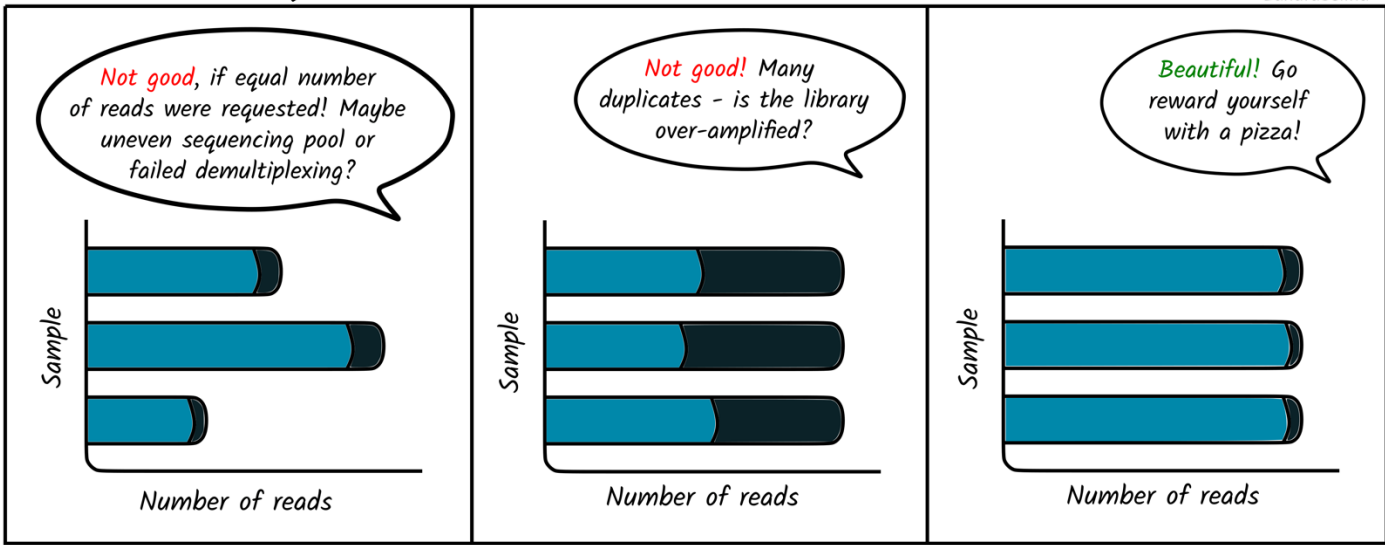
FASTQC:

<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>



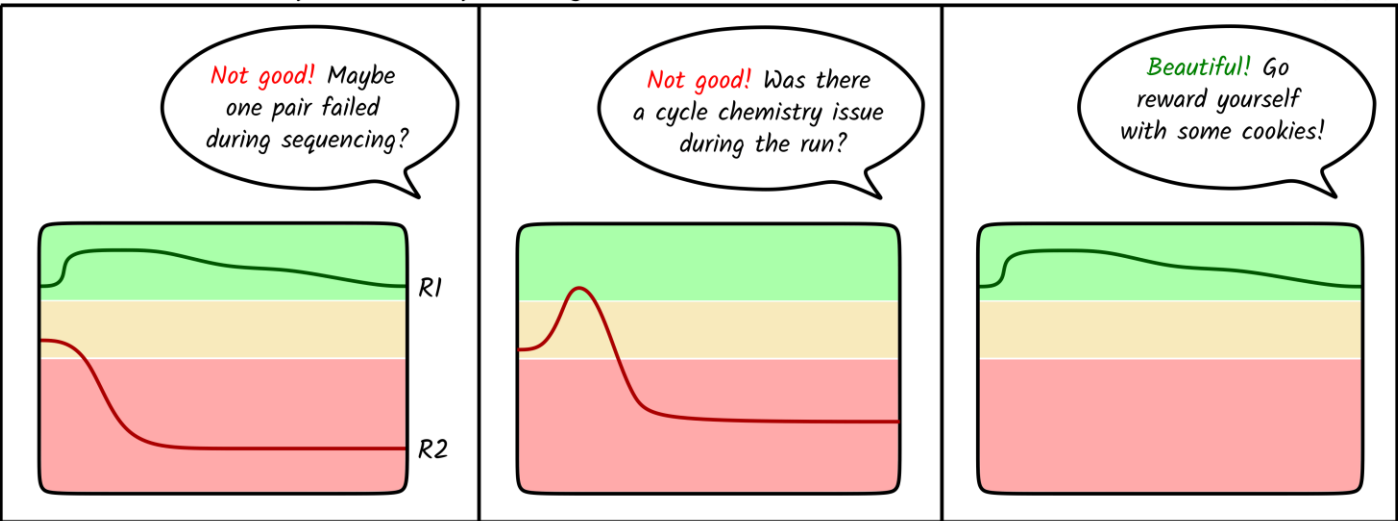
# Sequence count, Sequence Quality

## FASTQC - Sequence count



**Beautiful**  
: Equal number of reads with low duplication

## FASTQC - Sequence quality

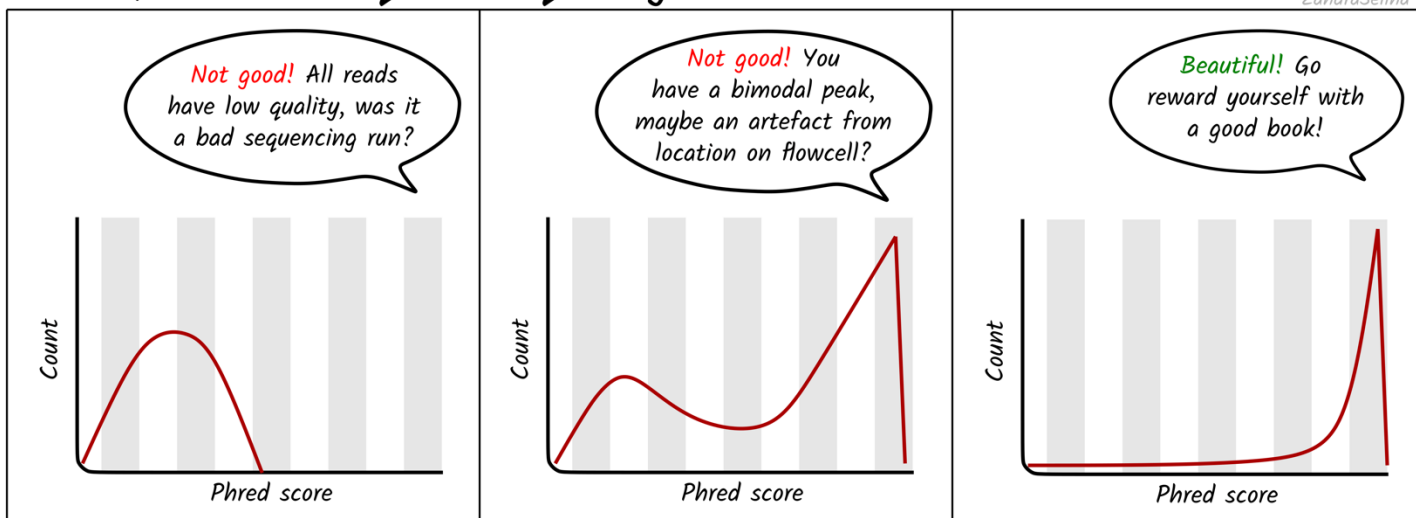


**Beautiful**  
: Stay above 35

# Per Sequence Quality Score, Per Base Sequencing Content

## FASTQC - Per sequence quality score

ZandraSelina



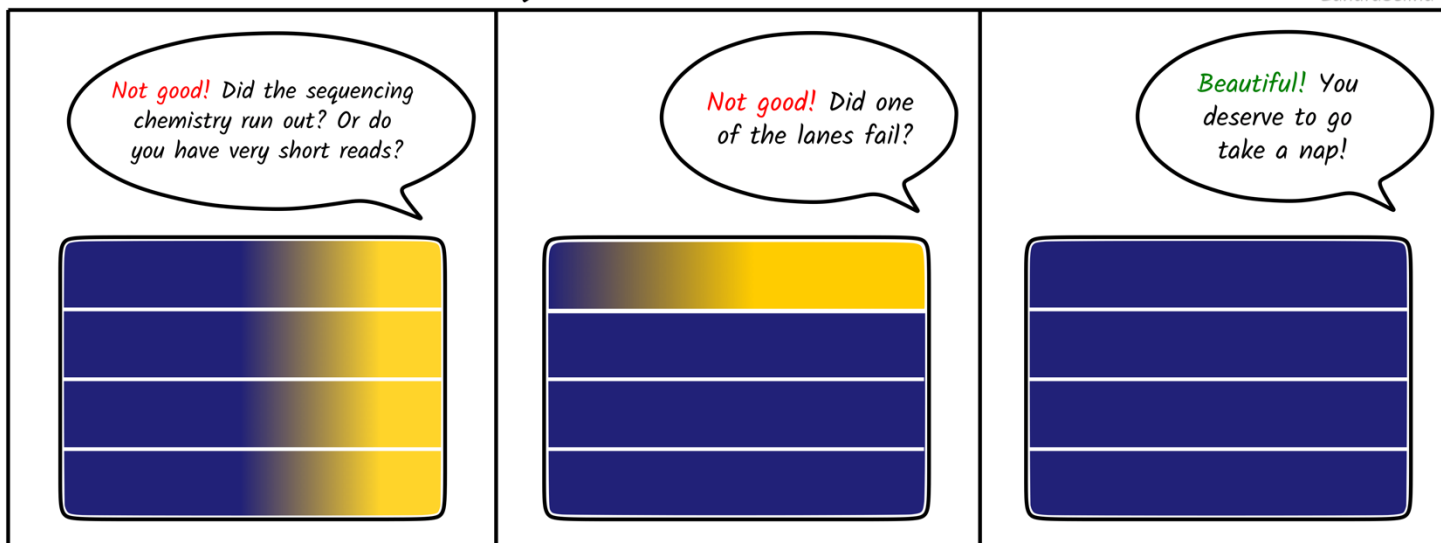
CC BY 4.0

**Beautiful**

: Equal number of reads with low duplication

## FASTQC - Per base sequence content

ZandraSelina



CC BY 4.0

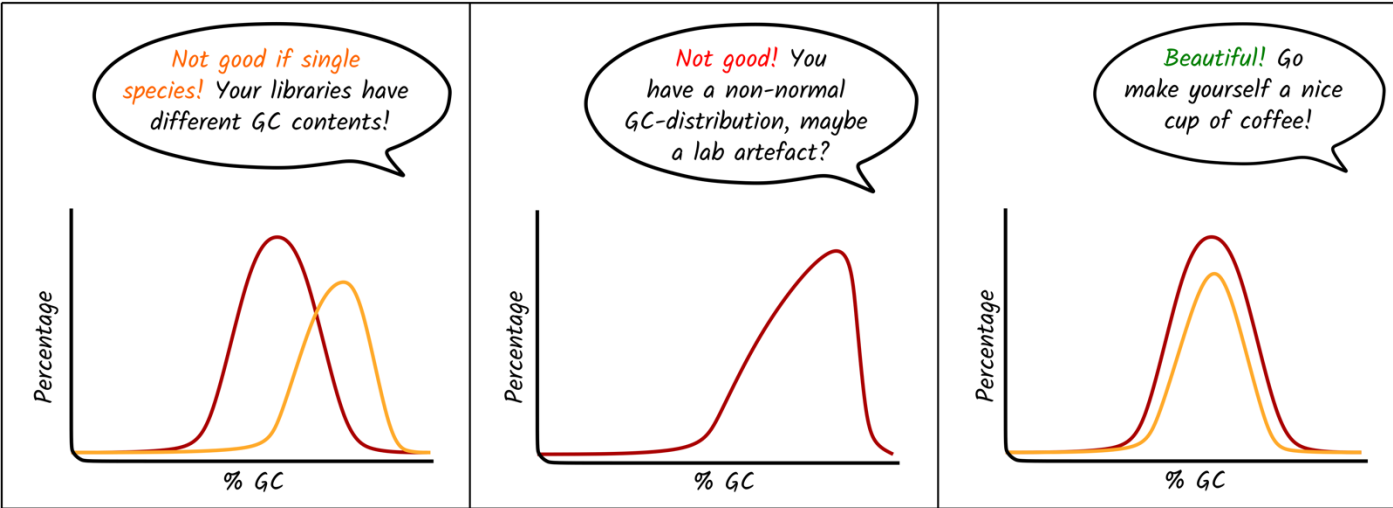
**Beautiful**

: Stay above 35

# Per Sequence GC-content, Per Base N Content

## FASTQC - Per sequence GC-content

ZandraSelina

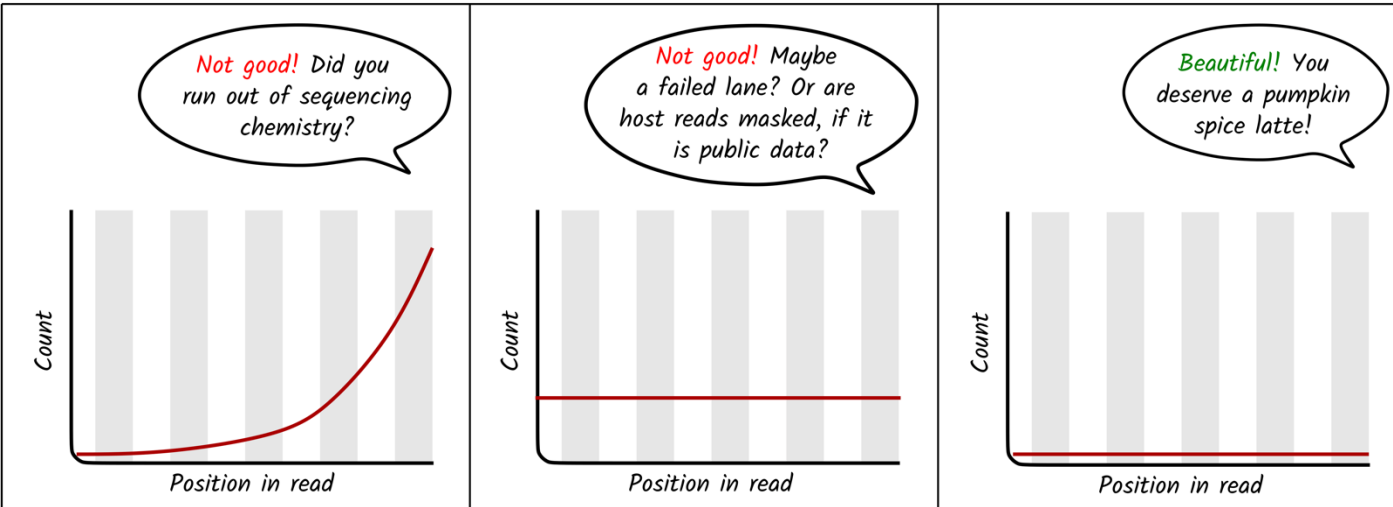


### Beautiful

: Peak around the average percent GC content of the reference genome (mean around 50%)

## FASTQC - Per base N content

ZandraSelina



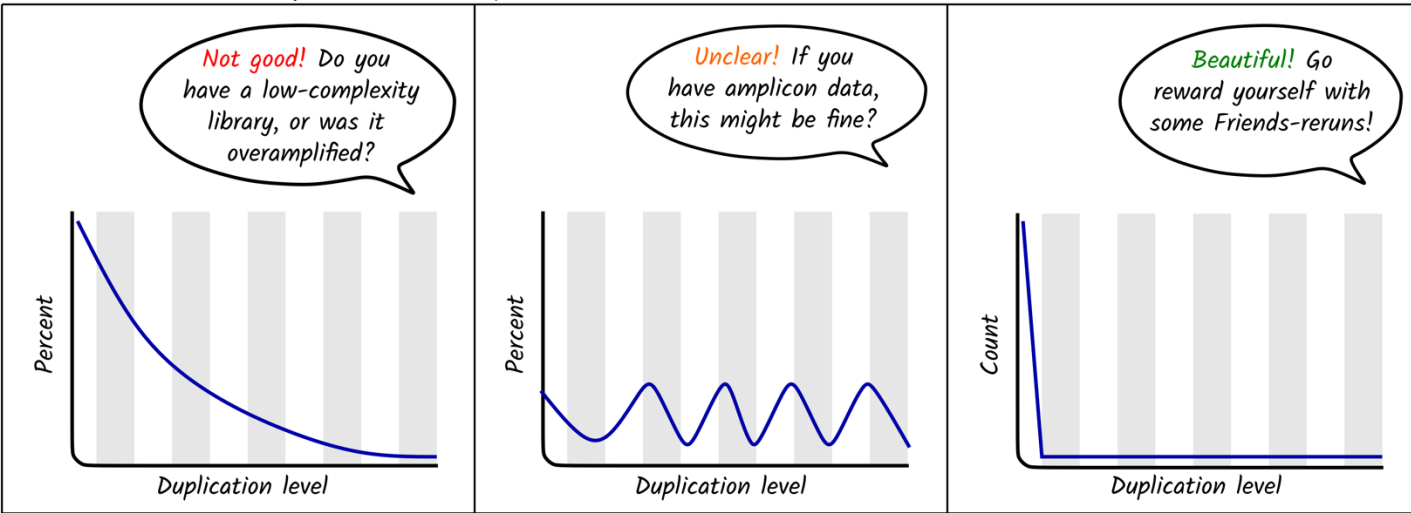
### Beautiful

: Close to 0 as possible

# Sequence duplication levels, Adapter content

## FASTQC - Sequence duplication levels

ZandraSelina



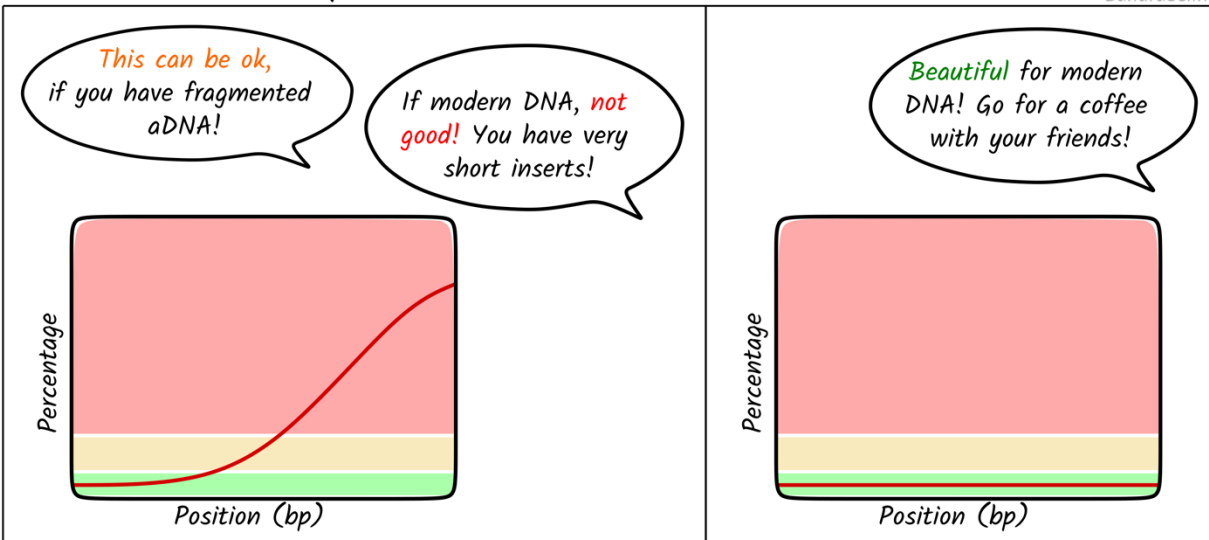
CC BY 4.0

**Beautiful**

: high complexity or lots of unique reads

## FASTQC - Adapter content

ZandraSelina



CC BY 4.0

**Beautiful**

: Close to 0 as possible

# FASTQ file QC (Quality Control)



FASTQC:

<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

The logo for usegalaxy.org, featuring the text 'usegalaxy.org' in white on a dark blue background. The 'u' is stylized with a yellow horizontal bar.

<https://usegalaxy.org/>