National Aeronautics and Space Administration

# Short Read Sequencing Overview

**GL4U: Introduction 2024**

Amanda M. Saravia-Butler, Ph.D.

NASA GeneLab Science Lead

Contractor: KBR

**B**iological & **P**hysical **S**ciences

# Sequencing Through the Ages



*1977*

*1990*

*Now*

∞

1st Generation Sanger Sequencing

Assembly required

2nd Generation "NextGen" Pyrosequencing

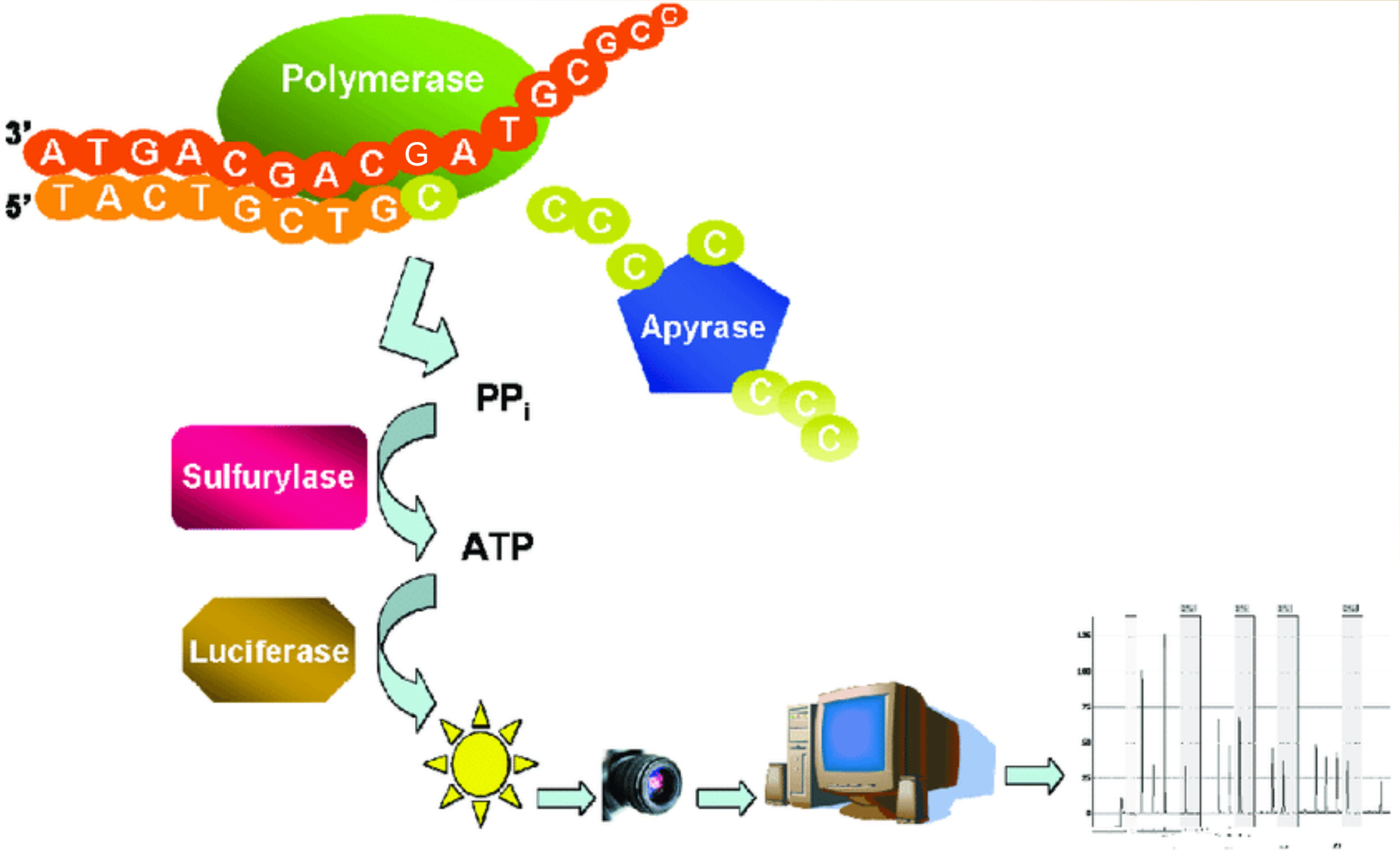
Assembly required

3rd Generation Single Molecule Nanopores

Not much assembly required

# Pyrosequencing aka Sequencing by Synthesis






https://www.youtube.com/watch?v=wY8to-_zAEo

https://www.dissertations.se/dissertation/e7acf1968c/

# Illumina Sequencing Workflow



- Illumina is a very successful biotech company specializing in next generation technology that uses the pyrosequencing method
- ~90% of all sequencing worldwide is performed on an Illumina instrument (including GeneLab)
- The Illumina sequencing workflow has the following 3 steps:

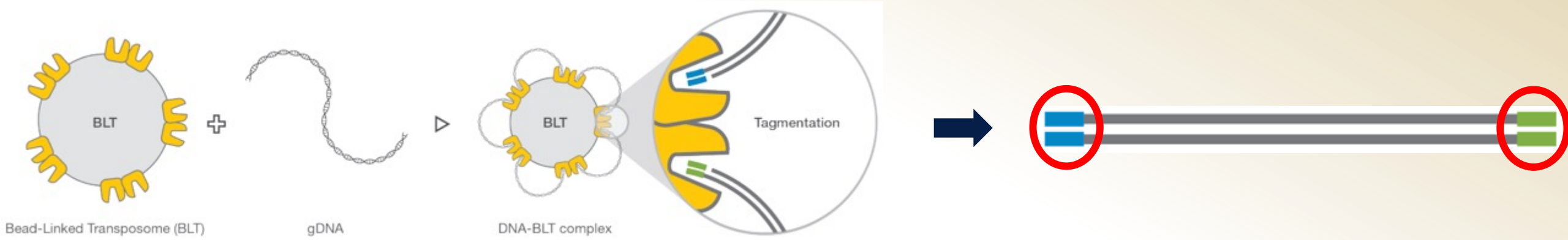
  - Library Construction
  - Cluster Formation
  - Sequencing

# Library Preparation
# DNA Sequencing

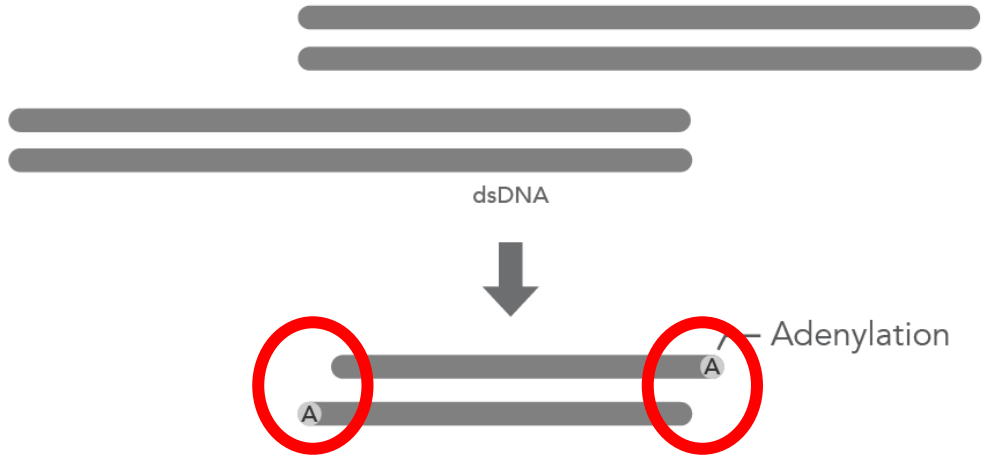
# DNAseq: Library Preparation

## Step 1: Create DNA fragments (with a means to attach adapters) from the extracted sample DNA

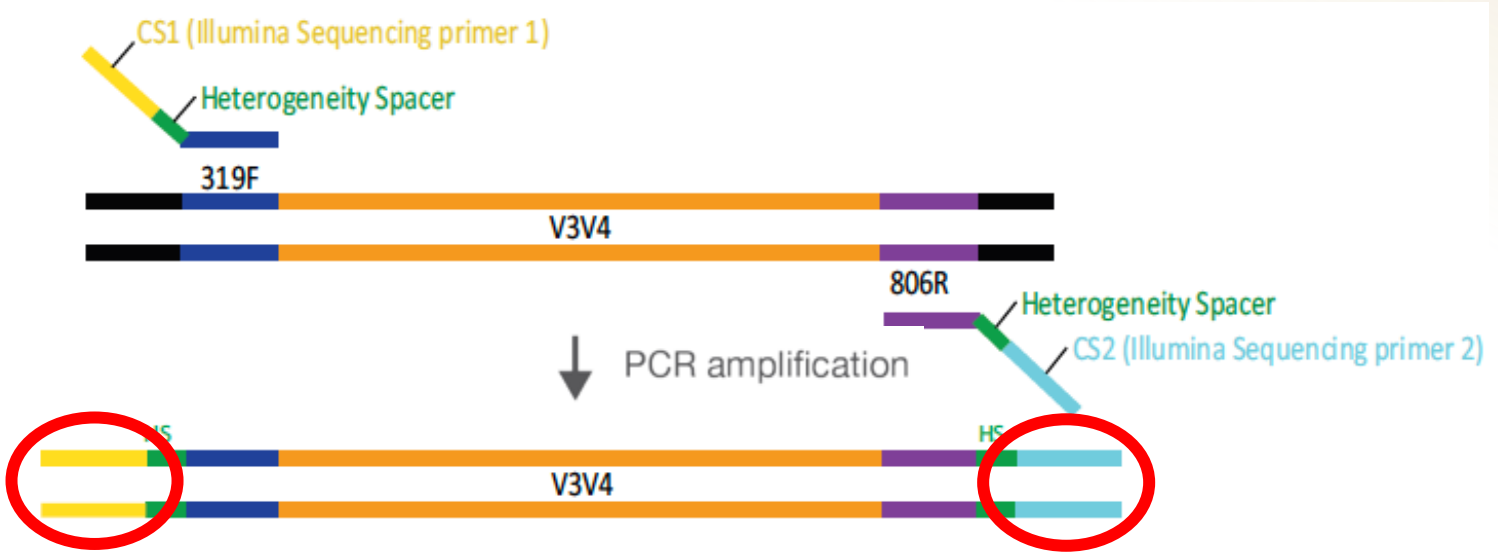
### Tagmentation: Illumina DNA Prep (formerly Nextera Flex)




### Fragmentation/A-tailing: IDT



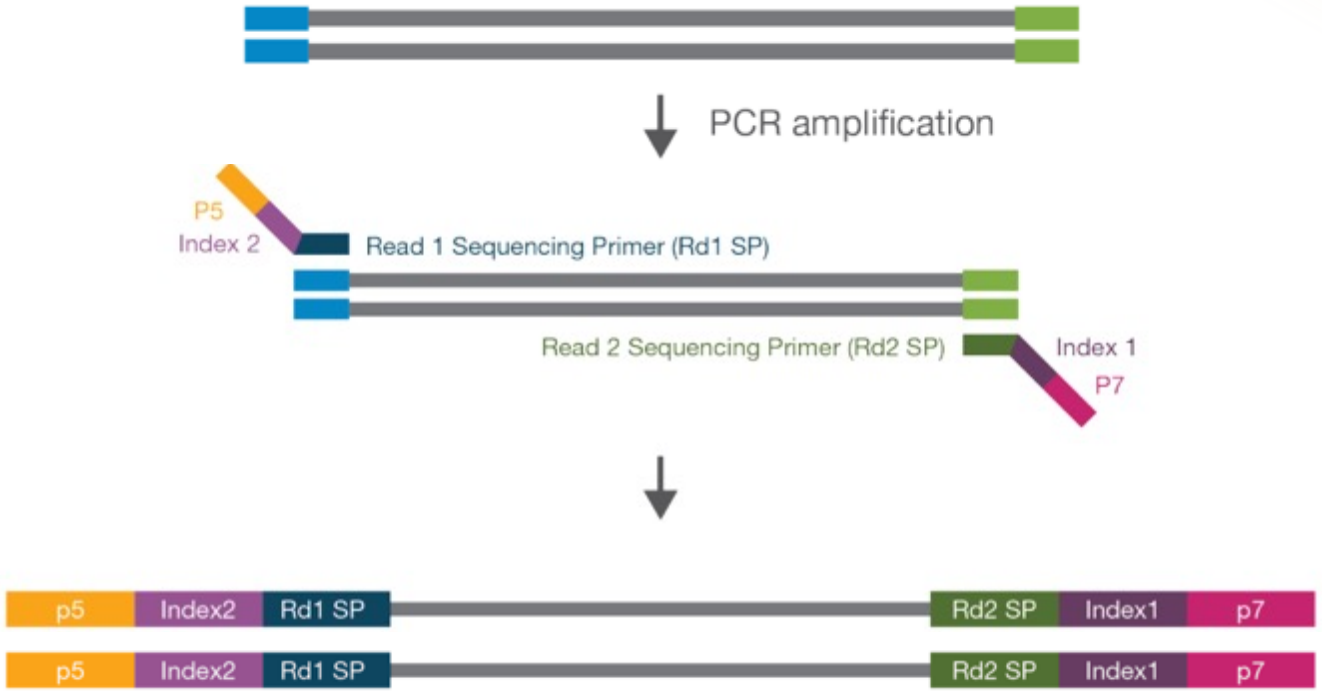

### Target PCR: Illumina Amplicon



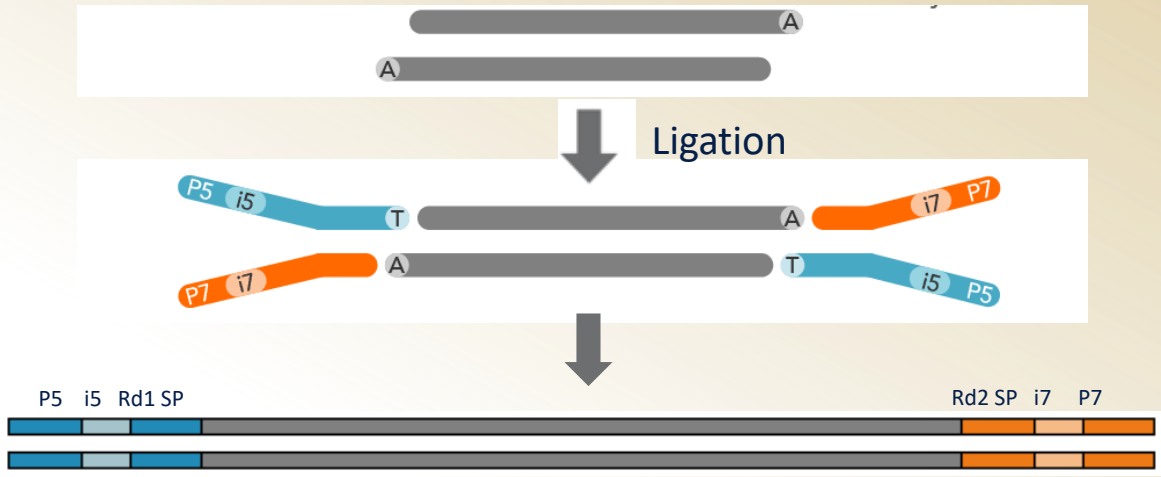
# DNAseq: Library Preparation

## Step 2: Attach adapters

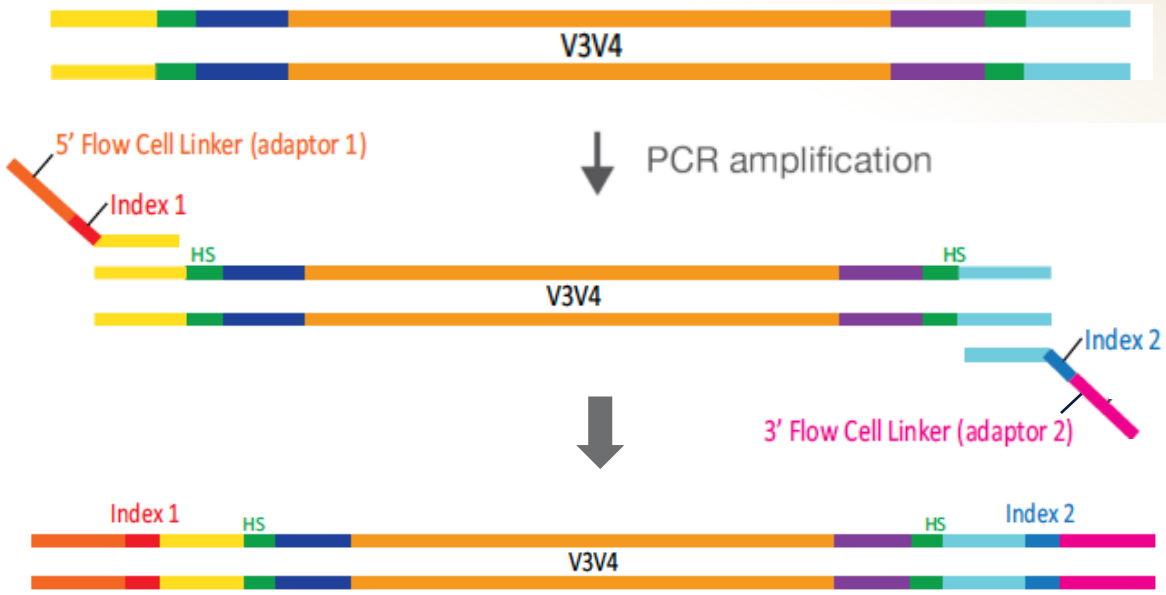
### Tagmentation: Illumina DNA Prep




### Fragmentation/A-tailing: IDT



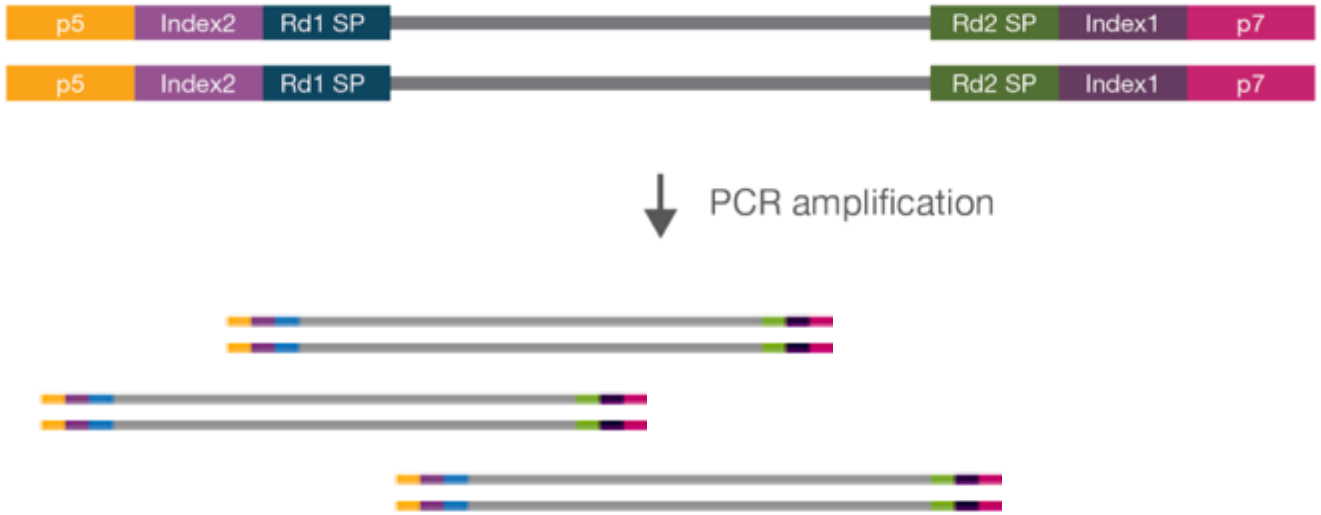

### Target PCR: Illumina Amplicon



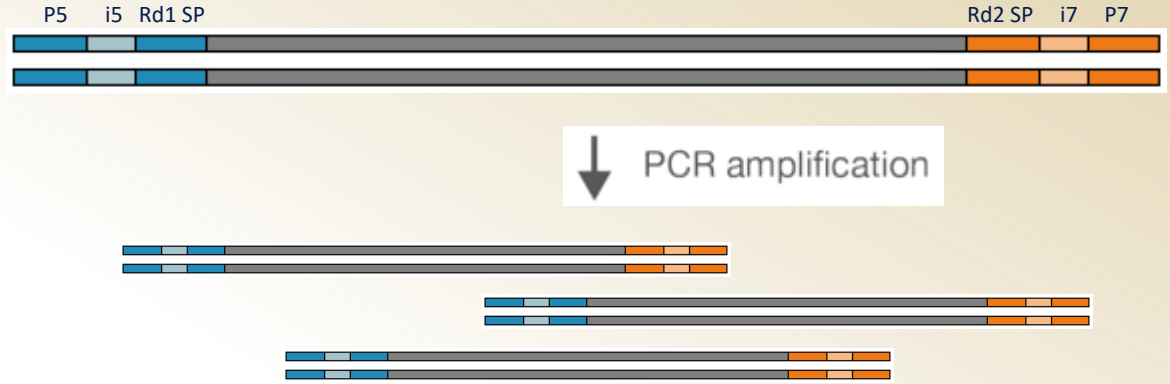
# DNAseq: Library Preparation

## Step 3: Amplify libraries

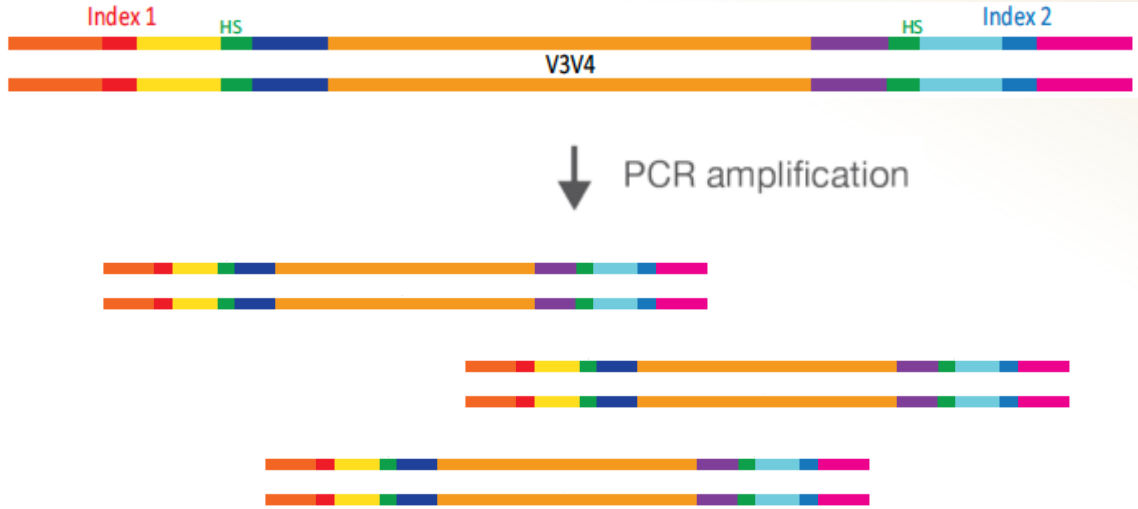
### Tagmentation: Illumina DNA Prep




### Fragmentation/A-tailing: IDT



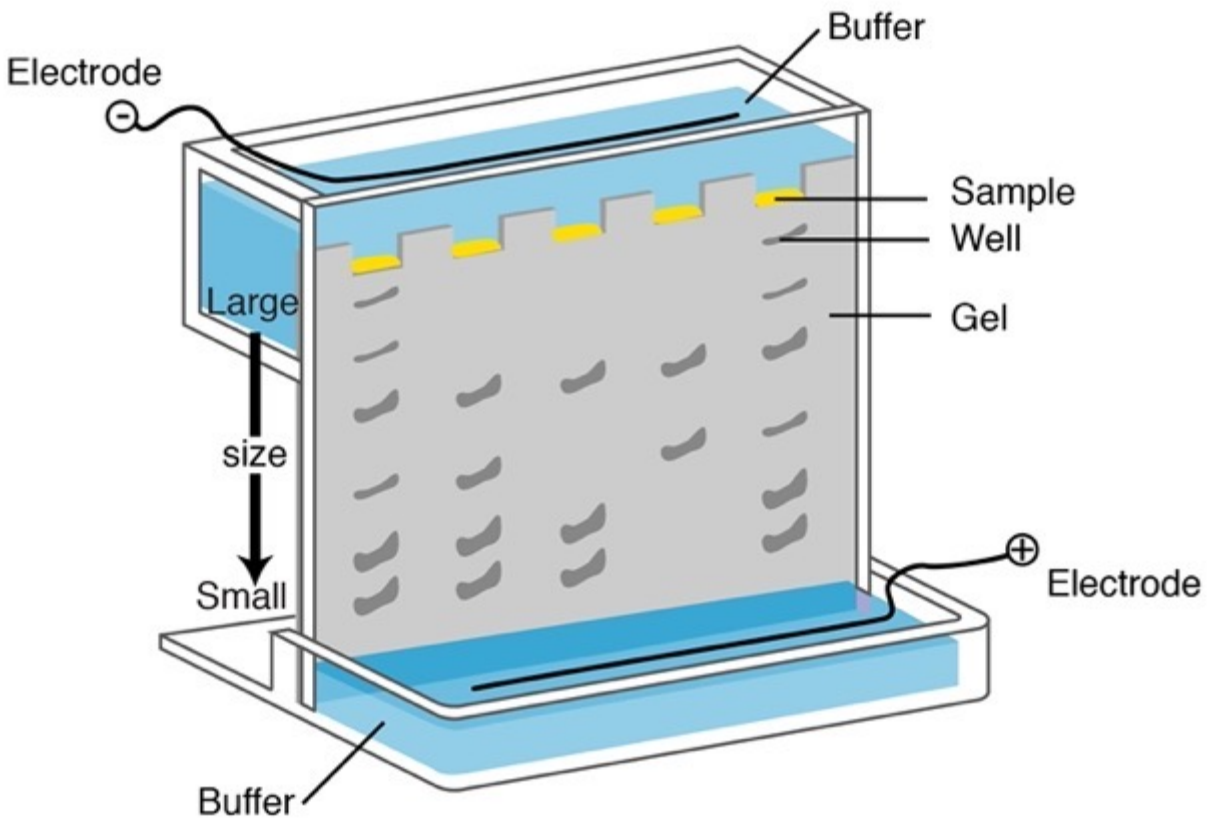

### Target PCR: Illumina Amplicon



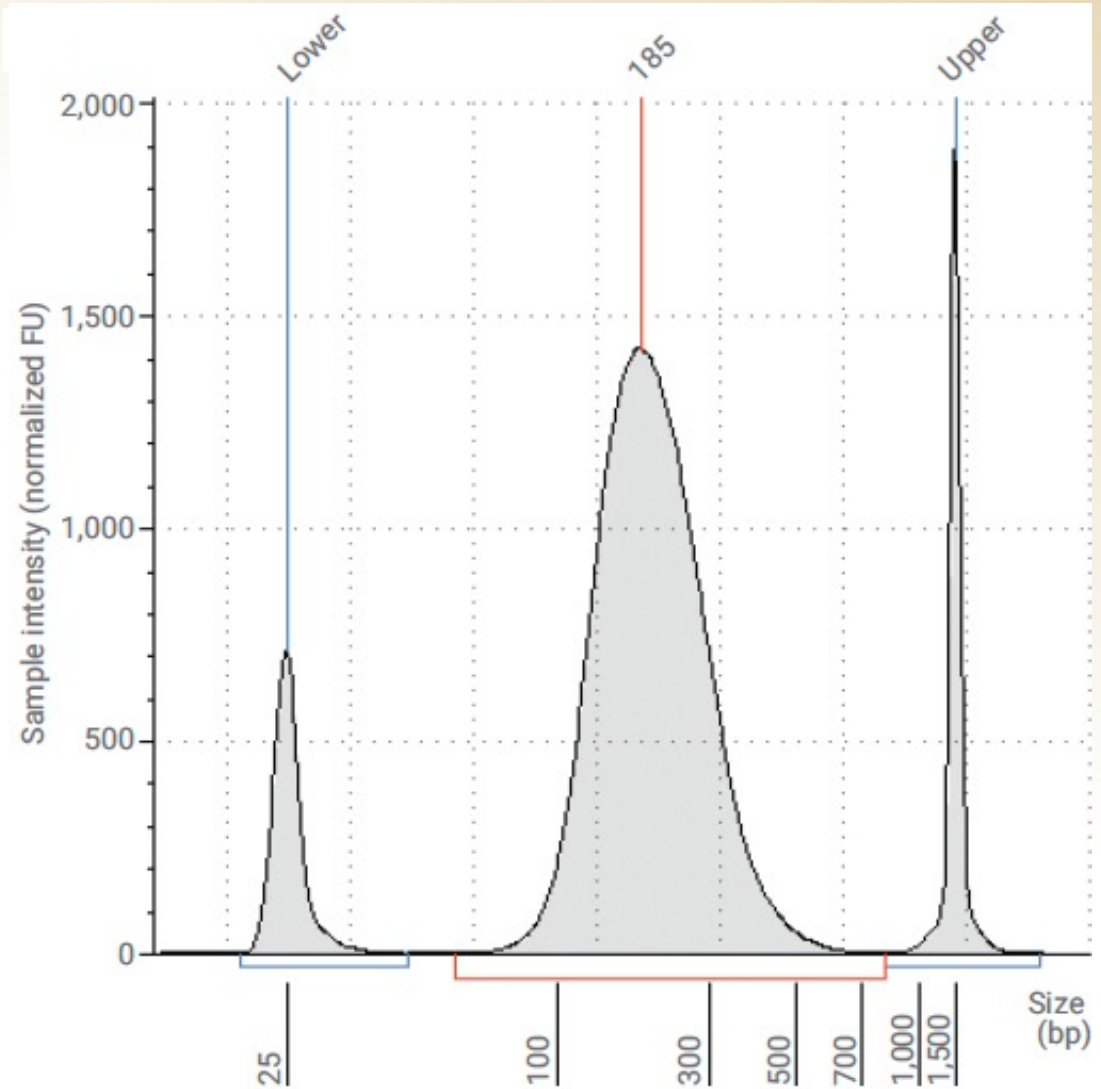
# DNA Library QC: Electrophoresis and Electropherogram



## Electrophoresis




## Electropherogram




https://www.agilent.com/cs/library/applications/application-ngs-electrophoresis-samplequalitycontrol-tapestation-5994-0127en-agilent.pdf

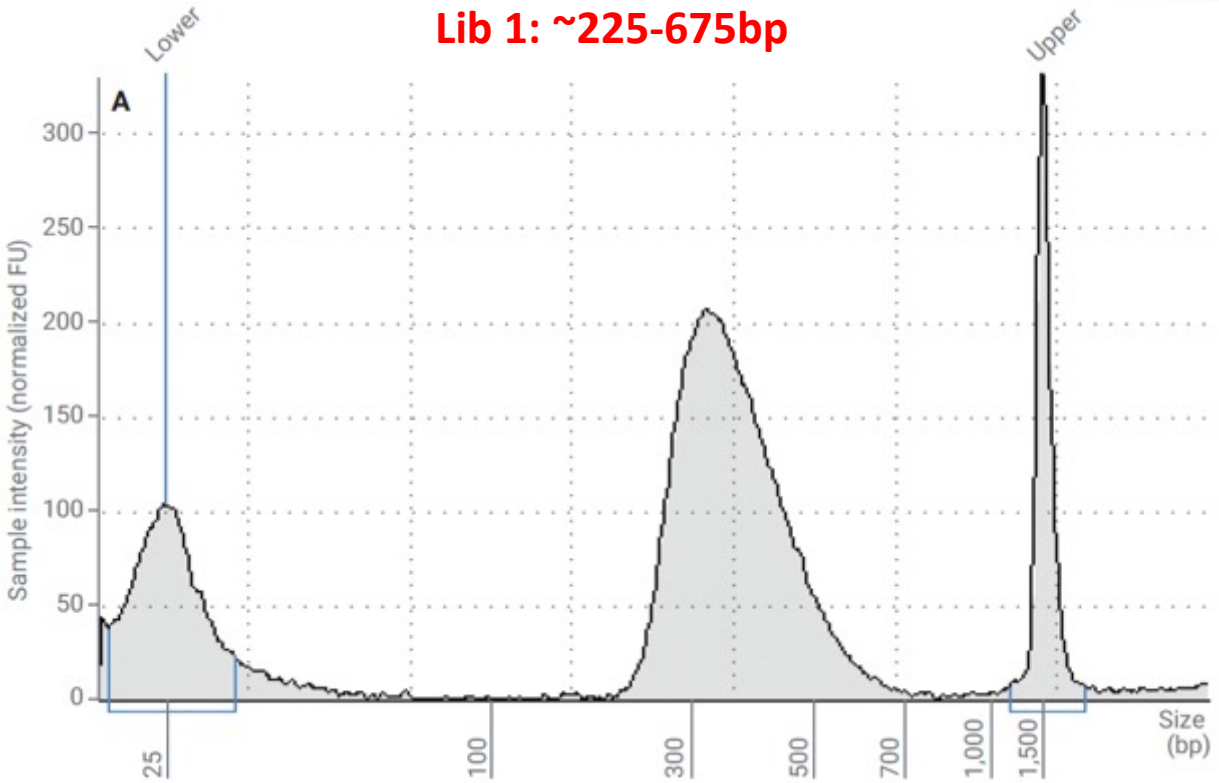
# DNAseq: Library QC



- Libraries are evaluated using a bioanalyzer or a tape station to create an electropherogram to assess quality

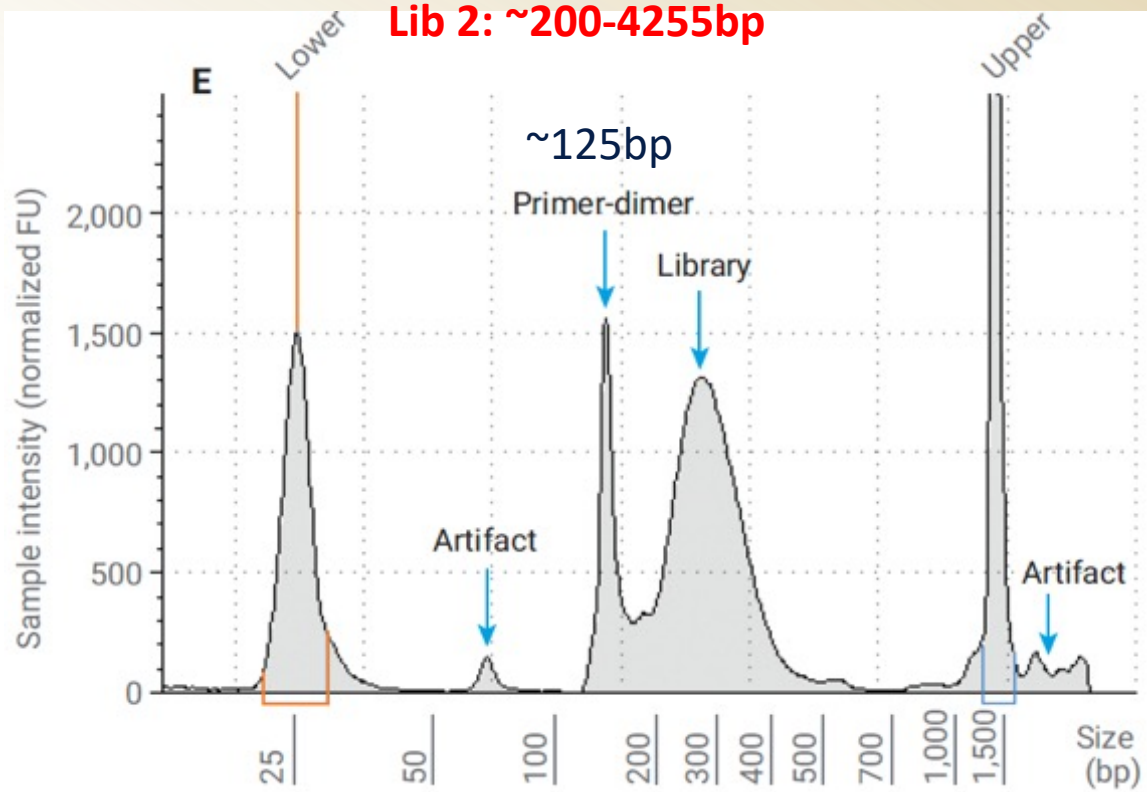
**Good** Library - 1

**Lib 1: ~225-675bp**




**Bad** Library - 2

**Lib 2: ~200-4255bp**




**When assessing library quality look for the following:**

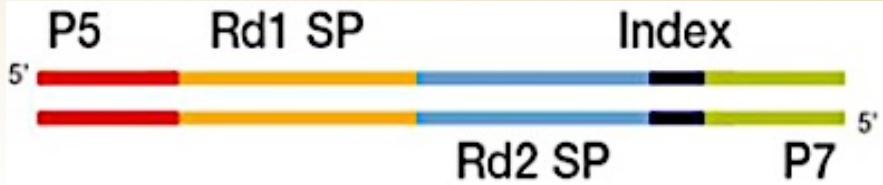
- NO adapter dimers! – Why?
- Library size is consistent with the number of desired sequencing cycles
  - If you're sequencing at PE 250, what is a good library size? (hint: library size = insert length + adapter length)

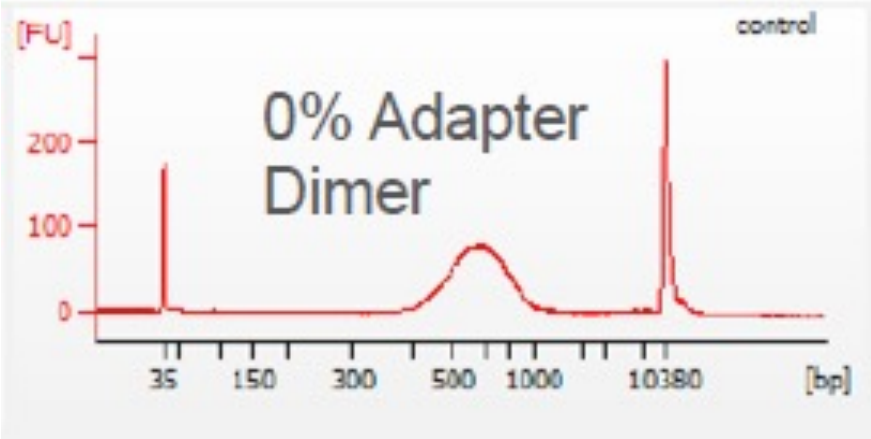
Assuming adapters are ~65bp each (130bp)
~600bp that would give an insert length of ~470,
allowing a ~30bp overlap between R1 and R2

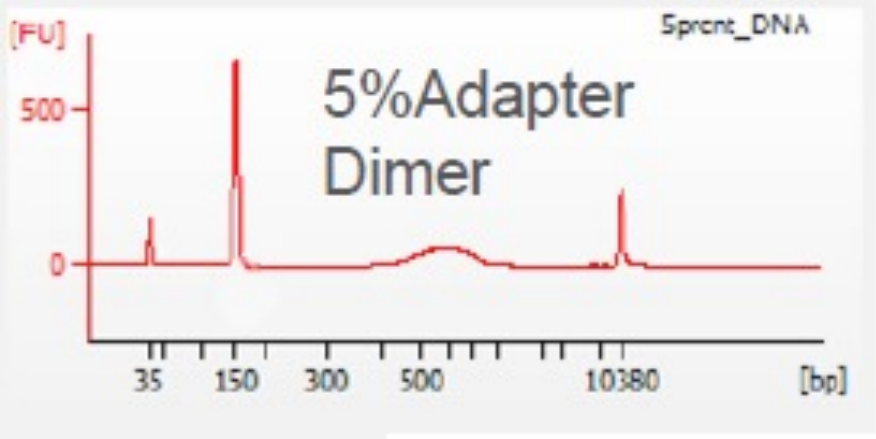
https://www.agilent.com/cs/library/applications/application-ngs-electrophoresis-samplequalitycontrol-tapestation-5994-0127en-agilent.pdf

# DNAseq: Library QC

- Assess library quality (bioanalyzer, TapeStation)
  - Adapter dimers
  - Fragment size
- Determine library quantity (Qubit, qPCR)



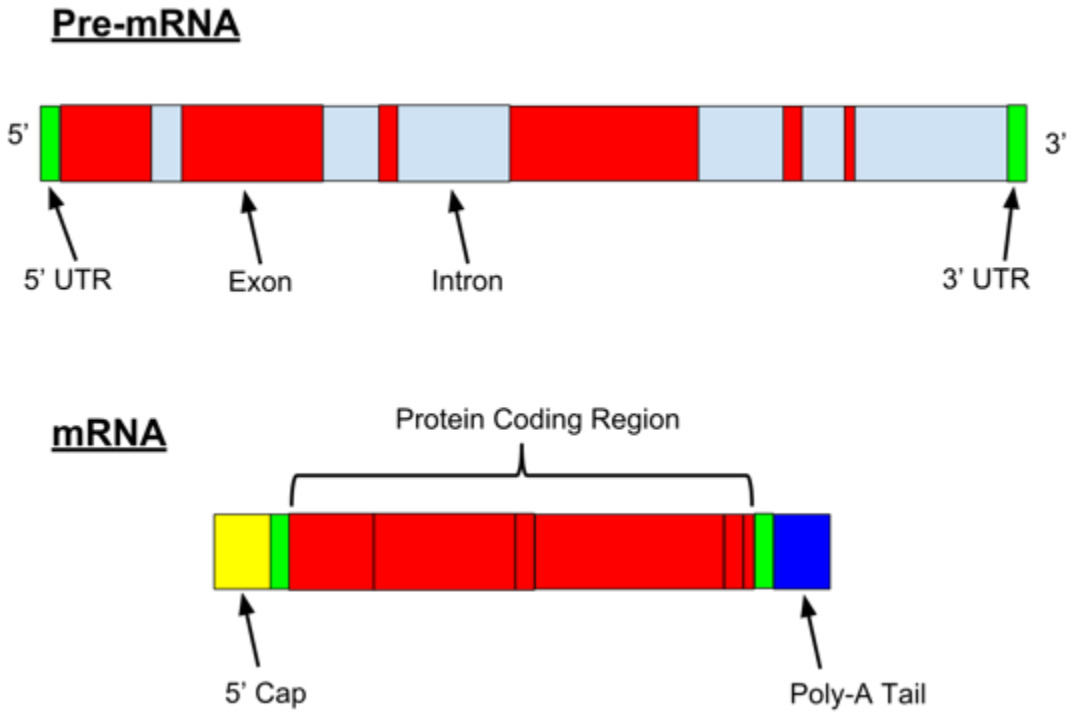

| % AD | % PF | % AD Reads |
|---|---|---|
| Control | 69.54 | 0.24 |
| 10% | 10.87 | 84.25 |
| 5% | 21.39 | 60.44 |
| 1% | 51.88 | 6.46 |

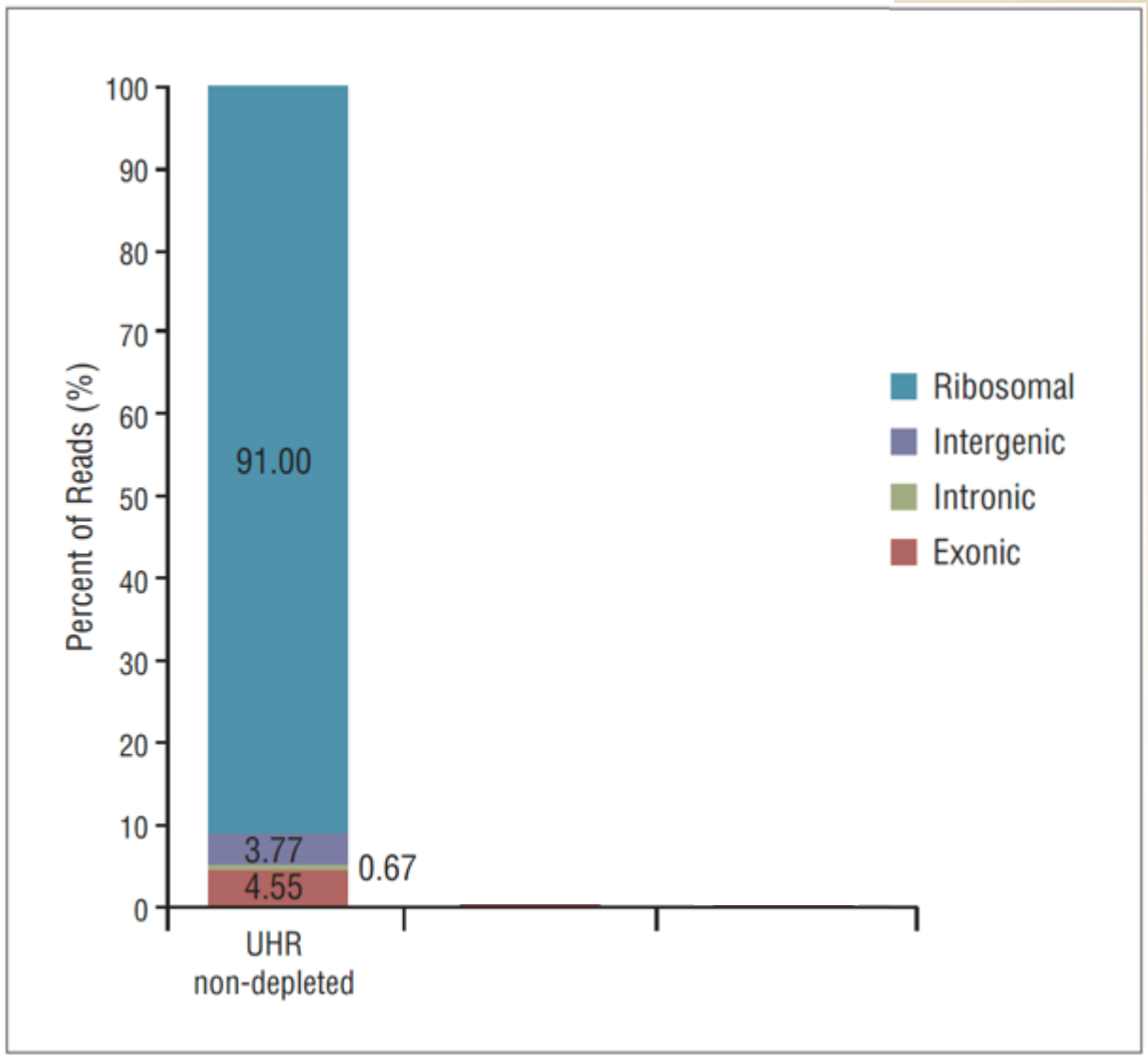
https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplepreps_truseq/truseq-stranded-total-rna-workflow/truseq-stranded-total-rna-workflow-reference-1000000040499-00.pdf

# Library Preparation
# RNA Sequencing

# rRNA Contamination

## mRNA makes up only ~2-5% of a total RNA sample




https://www.neb-online.de/wp-content/uploads/2015/04/NEB_selectivedepletionofabundantrnas_d.pdf

# RNAseq: Library Preparation



Figure 1 Ribo-Zero Depleting and Fragmenting RNA


Figure 2 Synthesizing First Strand cDNA


Figure 3 Synthesizing Second Strand cDNA

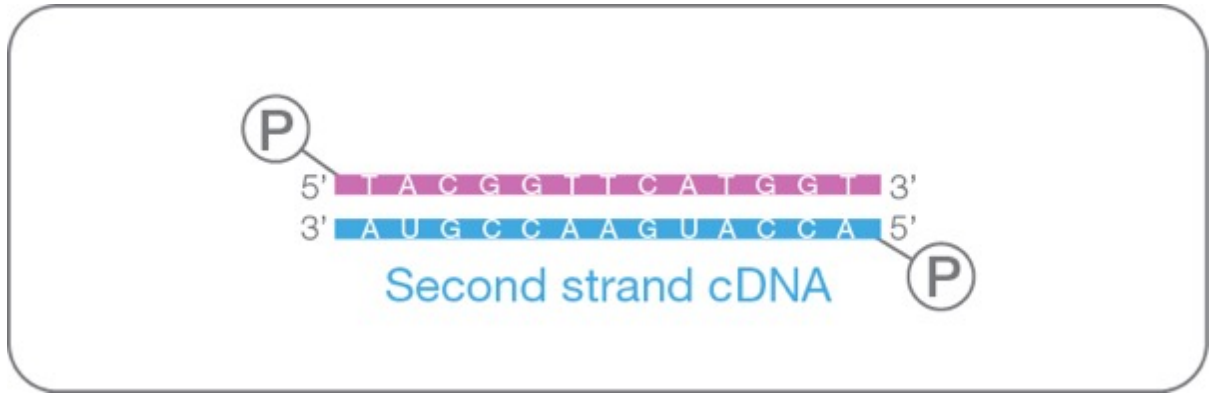
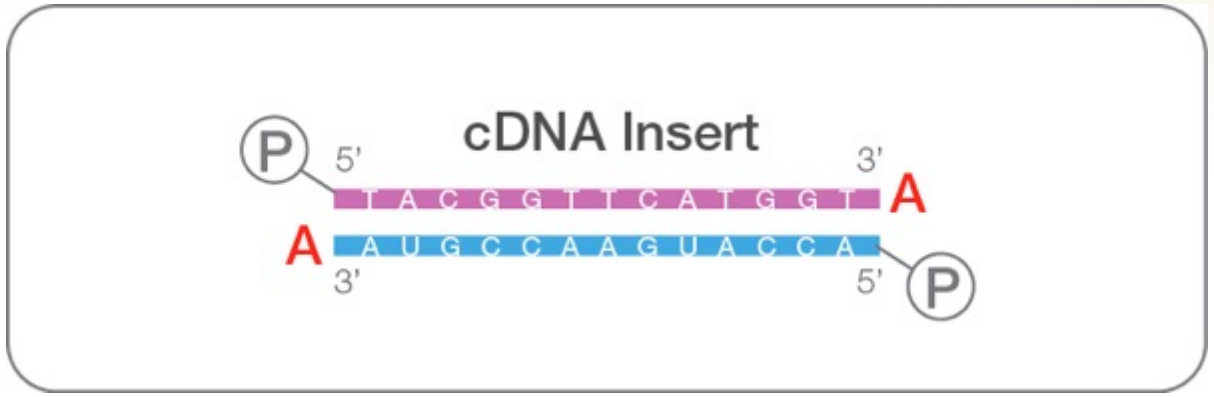
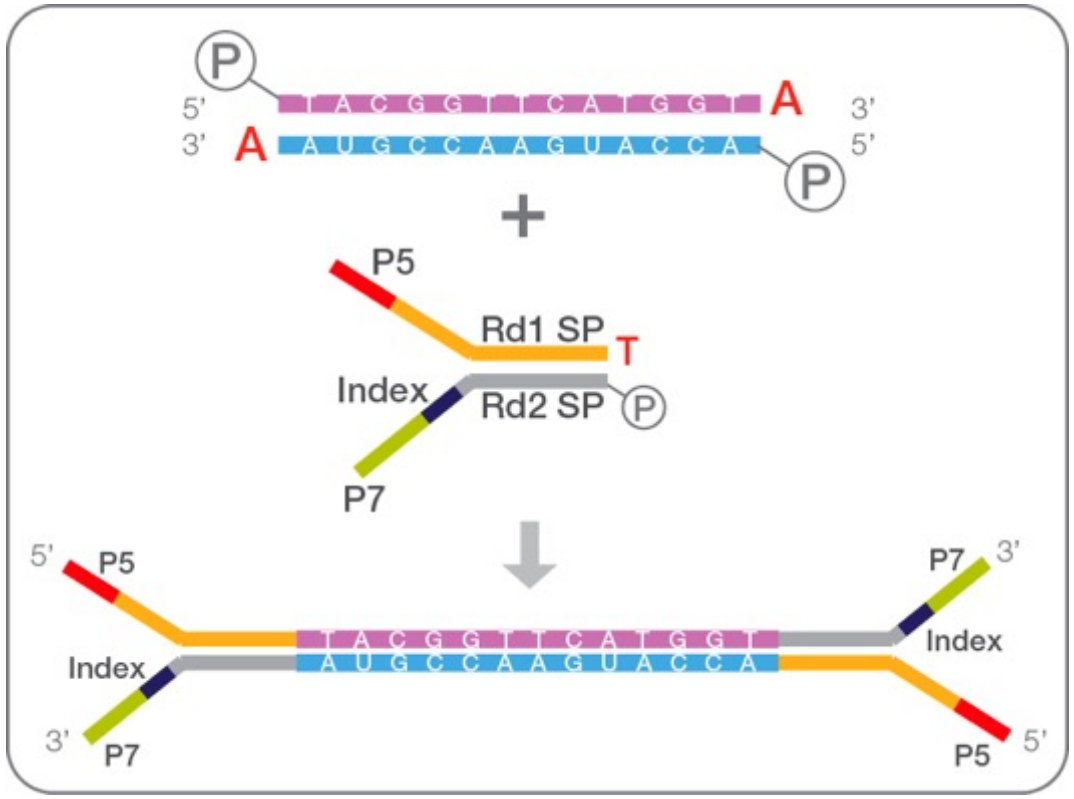



Figure 4 Adenylating 3' Ends*



# RNAseq: Library Preparation

Figure 5 Ligating Adapters





Illumina TruSeq Stranded Total RNA Reference Guide

https://www.idtdna.com/pages/products/next-generation-sequencing/adapters

# RNAseq: Library Preparation

Figure 6 Enriching DNA Fragments*

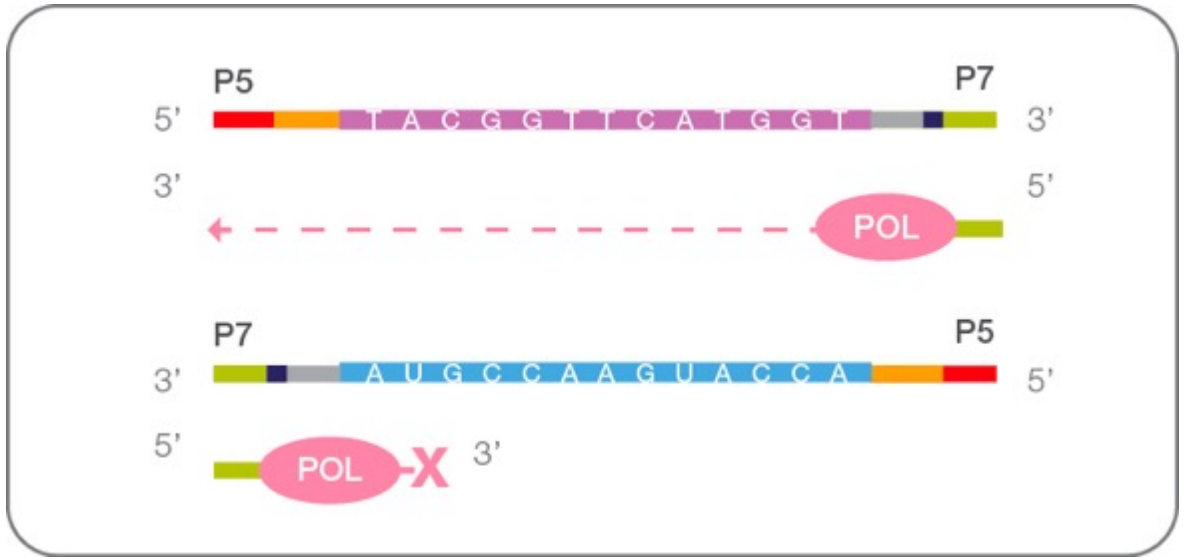
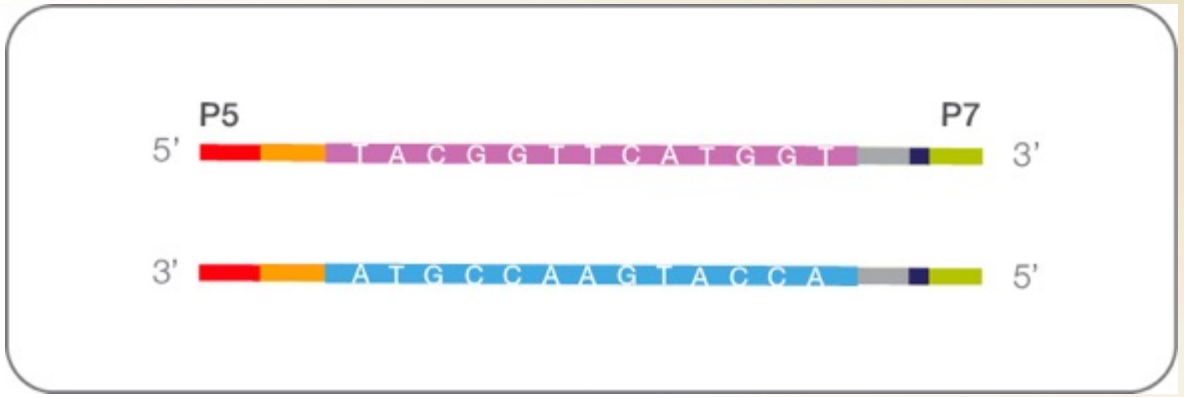


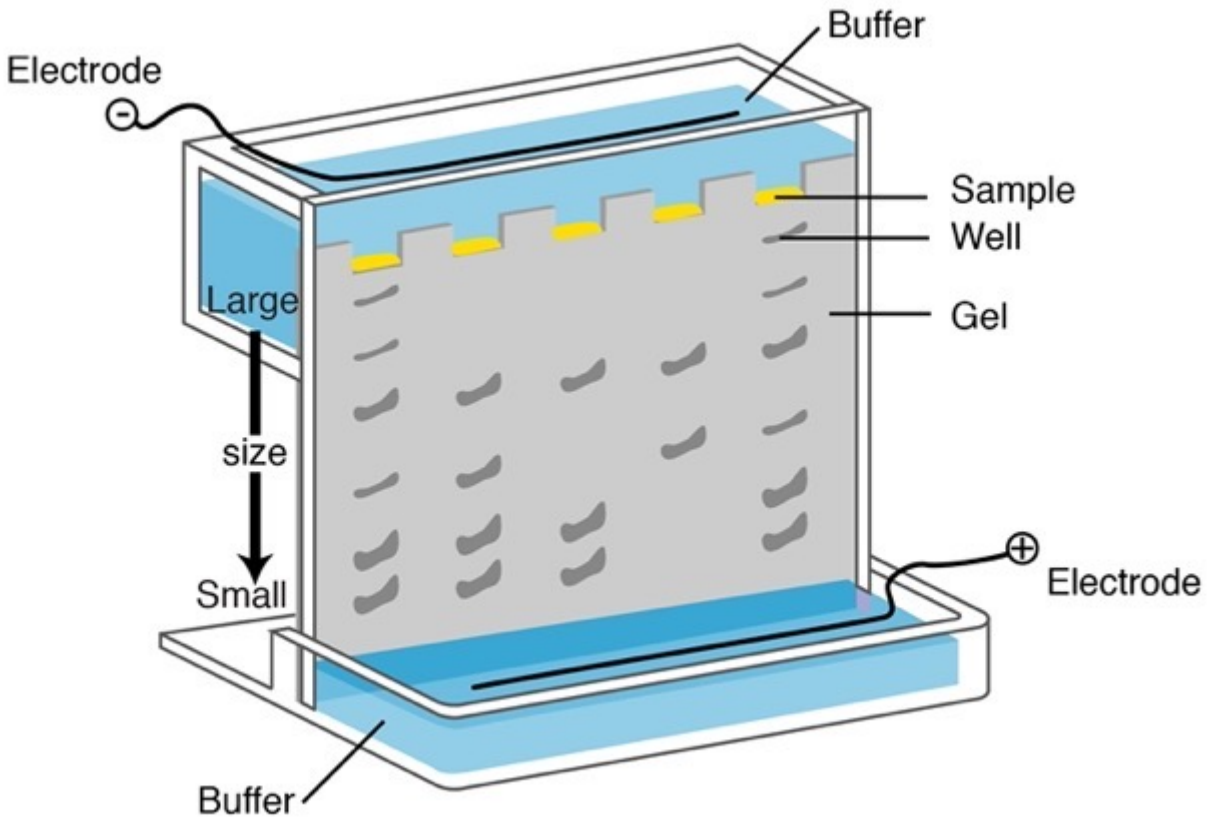

Figure 7 LS Final Library



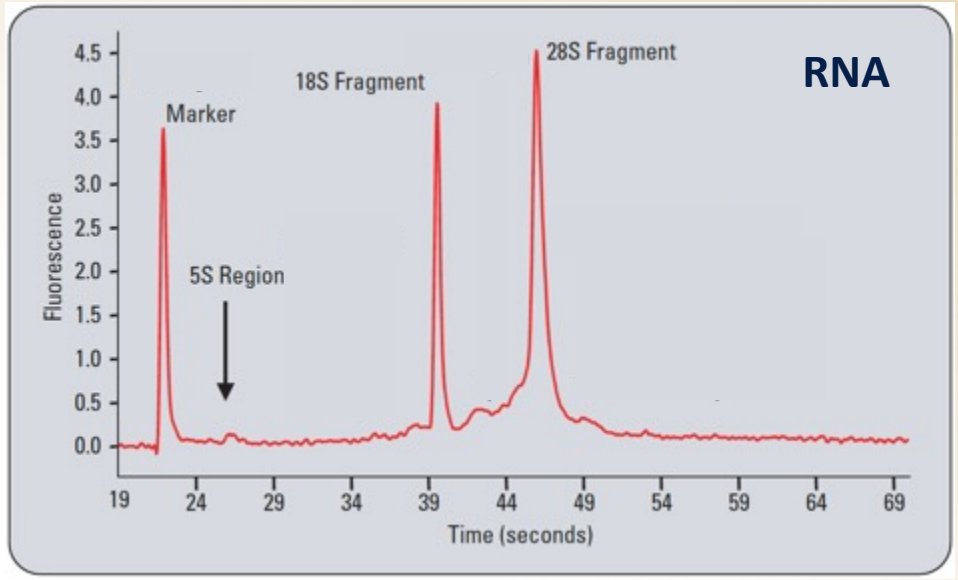
# RNA Library QC: Electrophoresis and Electropherogram



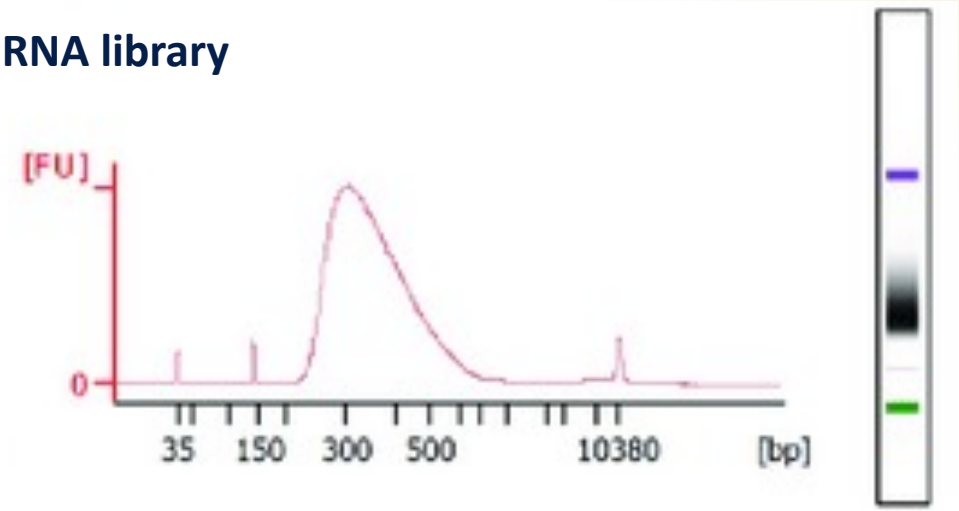
## Electrophoresis




## Electropherogram




## RNA library

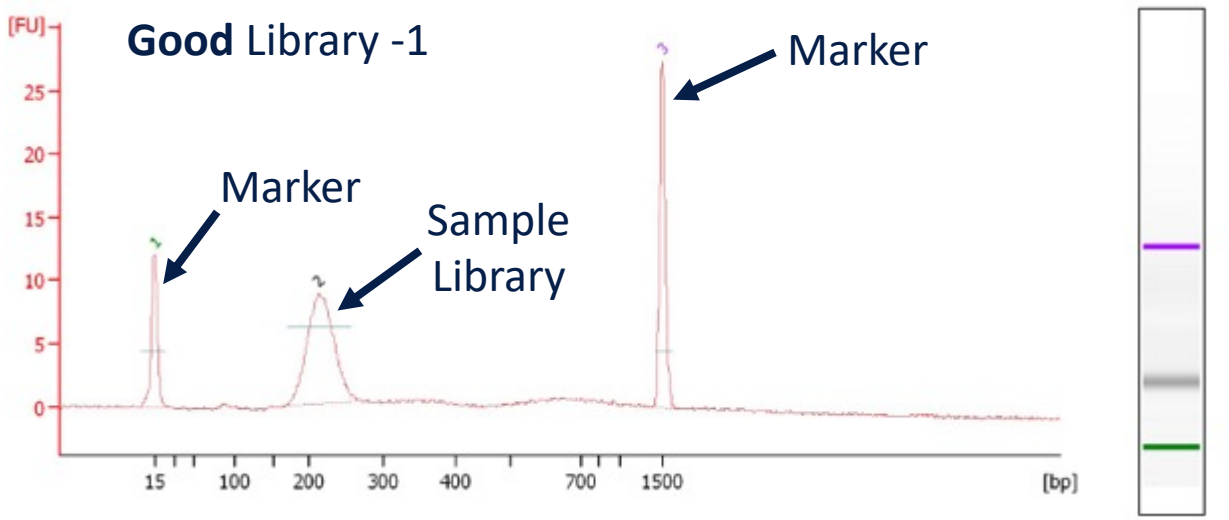



https://www.agilent.com/cs/library/applications/5989-1165EN.pdf

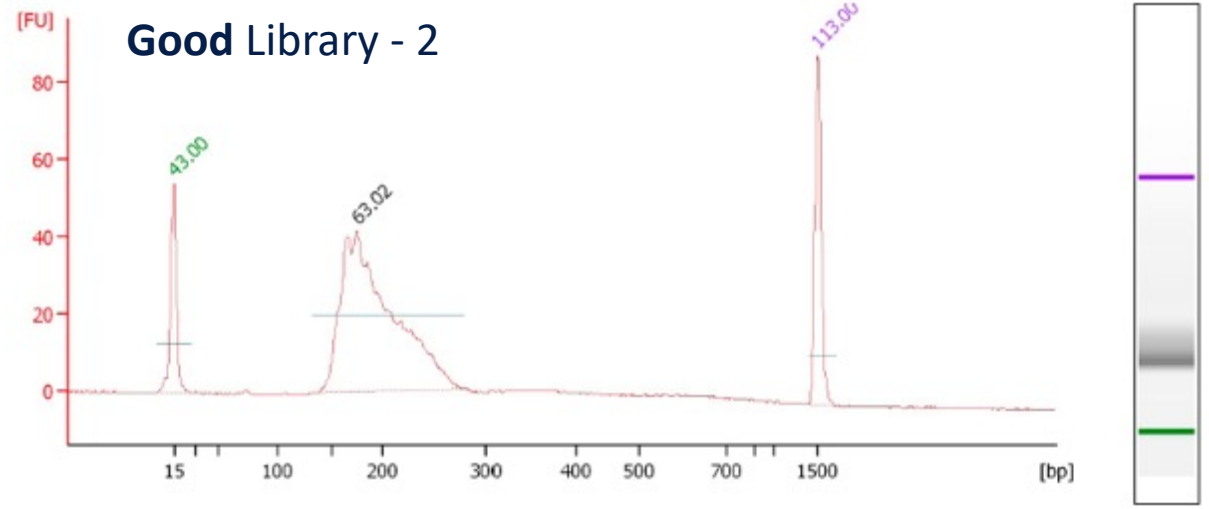
# RNAseq: Library QC

- Libraries are evaluated using a bioanalyzer or a tape station to create an electropherogram to assess quality

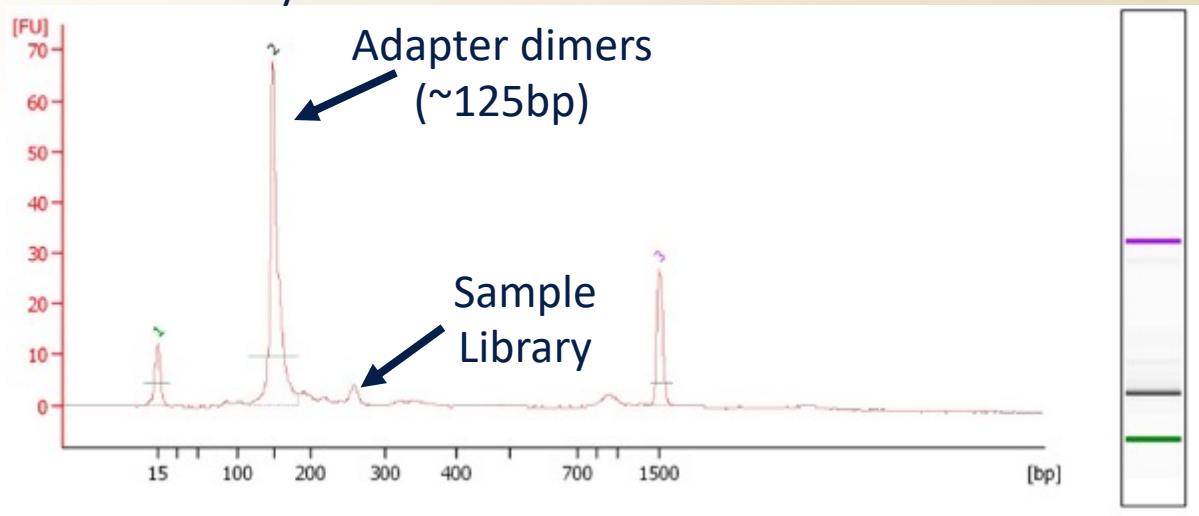
**Good** Library -1




**Good** Library - 2




**Bad** Library - 3




When assessing library quality look for the following:

- NO adapter dimers! – Why?
- Library size is consistent with the number of desired sequencing cycles
  - If you're sequencing at PE 100, what is a good library size? (hint: library size = insert length + adapter length)

Assuming adapters are ~65bp each (130bp): ~300bp that would give an insert length of ~170, allowing a ~30bp overlap between R1 and R2

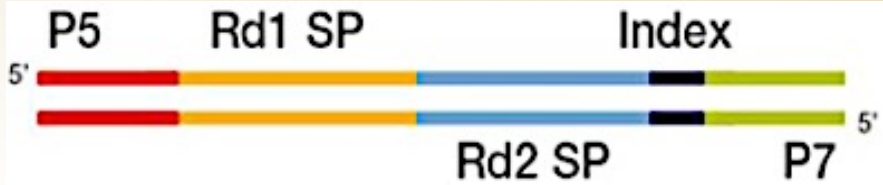
- What is the size range of library 1? Library 2?

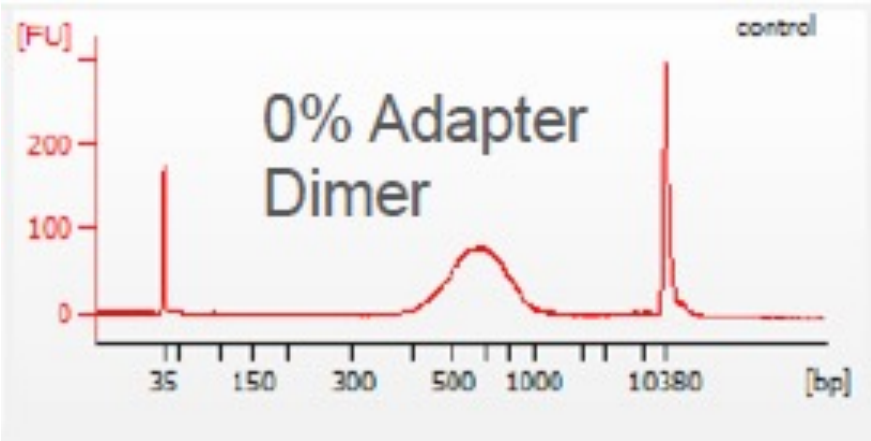
**~175-250bp; ~150-275bp**

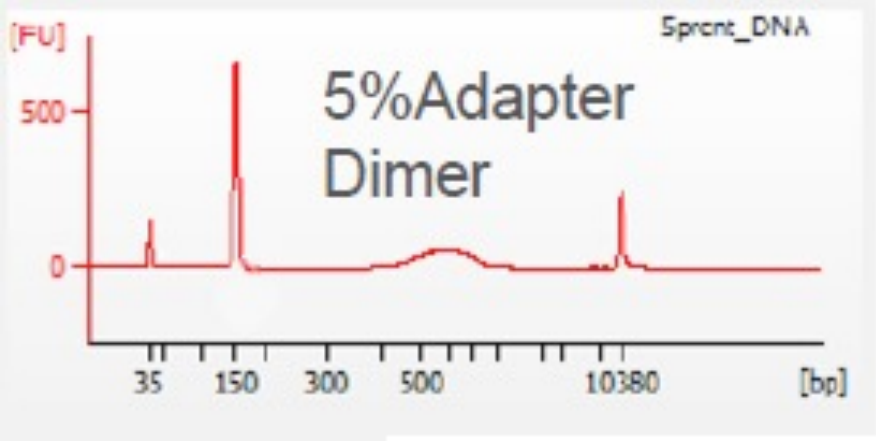
Modified from UC Davis Bioinformatics RNAseq Training

# RNAseq: Library QC

- Assess library quality (bioanalyzer, TapeStation)
  - Adapter dimers
  - Fragment size
- Determine library quantity (Qubit, qPCR)



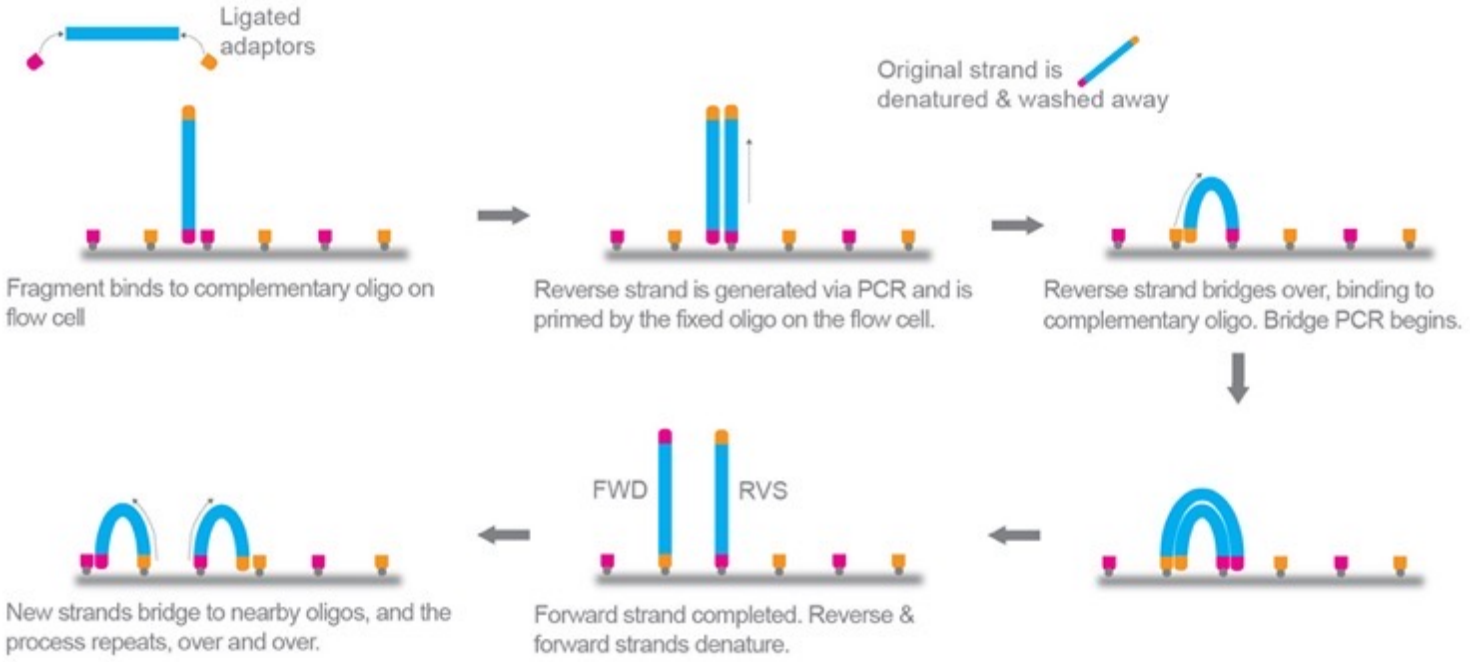

| % AD | % PF | % AD Reads |
|---|---|---|
| Control | 69.54 | 0.24 |
| 10% | 10.87 | 84.25 |
| 5% | 21.39 | 60.44 |
| 1% | 51.88 | 6.46 |

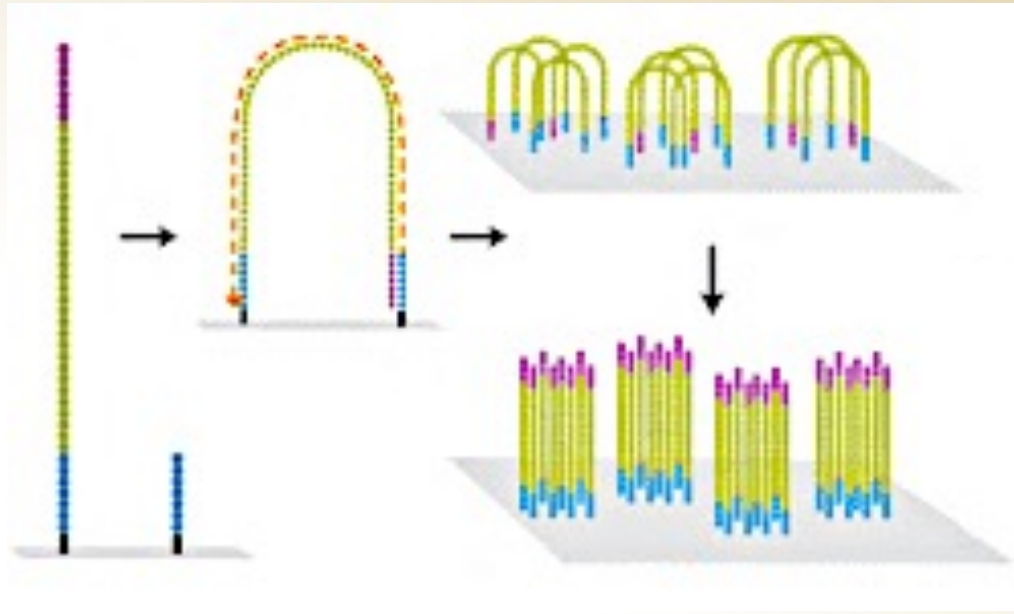
https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplepreps_truseq/truseq-stranded-total-rna-workflow/truseq-stranded-total-rna-workflow-reference-1000000040499-00.pdf

# Cluster Formation and Sequencing by Synthesis

# Cluster Formation





Attach DNA to flow cell → Bridge Amplification → Generate Clusters

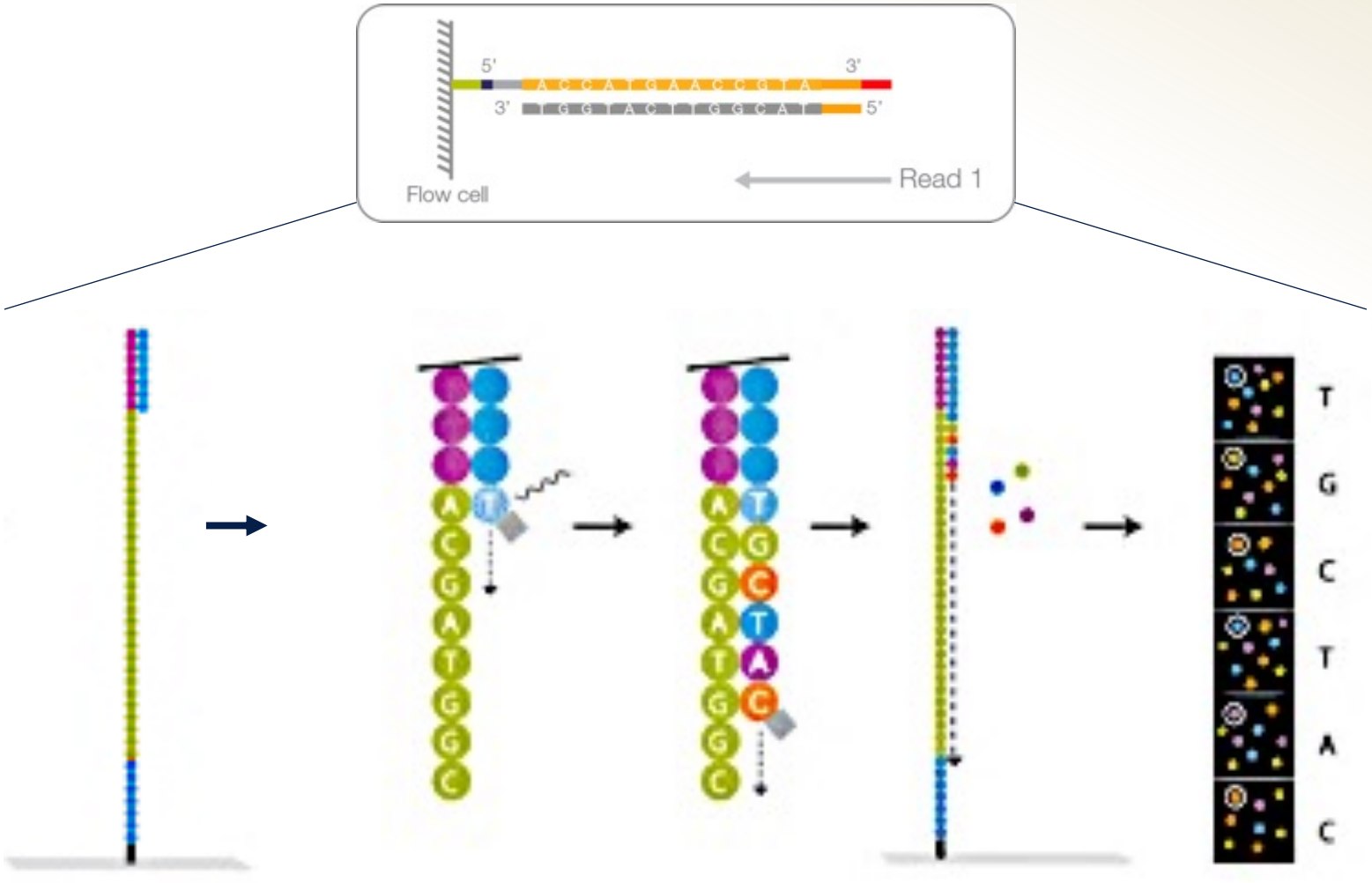
*Figure 4. Bridge PCR - a PCR method used to amplify samples for sequencing.*

https://www.goldbio.com/articles/article/Types-of-PCR-used-for-Genetic-Research

Illumina TruSeq Stranded Total RNA Reference Guide

# Illumina's High Throughput Sequencing by Synthesis

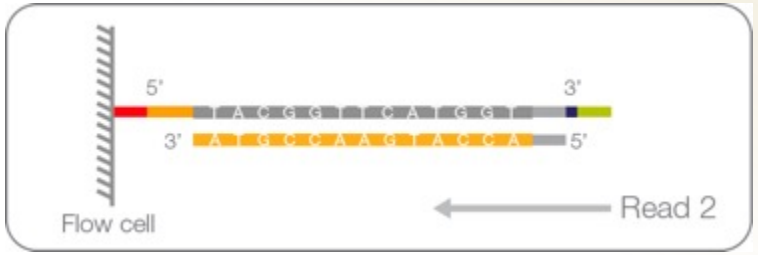
Figure 8 Read 1 Sequencing




Anneal Sequencing Primer → Extend first base, read, and deblock → Repeat previous step to extend strand → Generate base calls

- For paired-end (PE) sequencing, after read 1 is sequenced, forward strand reagents are washed
- The index read(s) are sequenced next
- Sequencing of the reverse read (read 2) is initiated after the index read(s)

Figure 9 Read 2 Sequencing



# Sequencing Parameters



| | Read length | Sequencing depth | Paired-end (PE) or Single-end (SE) |
|---|---|---|---|
| **DNAseq** | ➢ Longer reads enable greater confidence in taxonomic classifications and functional annotations<br>➢ "Assembly" is often performed with short reads to facilitate this<br>➢ GL standard is 2x250bp | ➢ Greater depth increases the likelihood of sequencing low-abundance organisms (if metagenomics) and detecting things like single-nucleotide variants and genetic rearrangements with greater confidence<br>➢ GL standards:<br>➢ Single organism or tissue:<br>➢ Re-sequencing (reference available): 10X minimum<br>➢ De novo sequencing (no reference available): 50X minimum<br>➢ Metagenomics (mixed community): 10M per sample, minimum | ➢ PE is generally preferred |
| **RNAseq** | ➢ Longer reads increase gene ID confidence<br>➢ GL standard is 2x150bp for bulk RNAseq | ➢ Greater depth increases the likelihood of sequencing low-abundant transcripts, detecting novel transcripts, and quantifying isoforms<br>➢ Greater depth is necessary for ribo-depleted samples (vs. poly-A enriched samples) – for RNAseq<br>➢ GL RNAseq standard for mammals prepared with ripo-depletion is 40-60M reads/sample<br>➢ More replicates is usually preferred over greater depth | ➢ PE is preferred |

# Illumina's High Throughput Sequencing by Synthesis



https://www.youtube.com/watch?v=fCd6B5HRaZ8

# Questions?