2- Good Laboratory Practices for Next-Generation Sequencing

DARRELL L. DINWIDDIE, PHD

DARYL B. DOMMAN, PHD

Updates/Reminders

▶ **Optional Office Hours**

    ▶ **Thursday Morning, 1 hour**

    ▶ **Questions & Discussions on previous or upcoming topics**

▶ **Upcoming Topics**

    ▶ **Week 3- Introduction to NGS Data and File types- Dr. Domman**

    ▶ **Week 4- Introduction to working on the command line and virtual machine- Dr. Domman**

Which Protocols to Cover in Depth?

Illumina

- SARS-CoV-2
  - AmpliSeq
  - COVIDSeq
  - Respiratory virus oligo panel v2
  - SWIFT/IDT, Qiagen
- Bacterial
  - Nextera XT/DNA Prep
  - 16S rRNA

ONT

- SARS-CoV-2
  - ARTIC/Midnight- Ligation
  - Midnight-Rapid Barcoding
- Bacterial
  - Shotgun whole genome-short & long read
  - 16S rRNA
- Direct RNA
  - RNA viruses

Schedule Changes?

- **Current Schedule**
- **Week 5 (March 15)- Overview of laboratory protocols for pathogen sequencing (non SARS-CoV-2)**
- **Week 7 (March 29)- Illumina based SARS-CoV-2 genome sequencing protocols**
- **Week 9 (April 12)- Nanopore based SARS-CoV-2 genome sequencing protocols**

- **Proposed Schedule**
- **Week 5 (March 15)- Illumina based SARS-CoV-2 genome sequencing protocols**
- **Week 7 (March 29)- Nanopore based SARS-CoV-2 genome sequencing protocols**
- **Week 9 (April 12)- Overview of laboratory protocols for pathogen sequencing (non SARS-CoV-2)**

Abbreviations

- ONT- Oxford Nanopore Technologies
- PCR- Polymerase Chain Reaction
- VTM- Viral Transport Media
- RIN- RNA Integrity Number
- QC- Quality Control

Sequencing Read Length

▶ **The number of nucleotides that are being sequenced.**

    ▶ **For Illumina sequencing this is determined by the number of cycles of sequencing performed & the sequencing kit. Read lengths are all the same.**

    ▶ **For ONT sequencing this is dependent on the nucleotide length of the sample, library prep method, & nanopore function. Read length can be different for each sequencing read.**

17bp   TTCGGCTAGCTTGCAGC

23bp   AGCTTCATGATGGGCCAAATTTT

31bp   CCAAATTTTCTAGAGTAGTCACTAGCTTCCG

Reference-based Alignment (Mapping)

▶ Matching of sequencing reads to a reference sequence.

Reference

ATTTCGGCTAGCTTGCAGCTTCATGATGGGCCAAATTTTCTAGAGTAGTCACTAGCTTCCG

ATTTCGGCTAGCTTGCA   AGTCACTAGCTTCCG

 TTCGGCTAGCTTGCAGC   GAGTAGTCACTAGCTTCCG

             GCTTCATGATGGGCCAAATTTTC

             AGCTTCATGATGGGCCAAATTTT

                               CCAAATTTTCTAGAGTAGTCACTAGCTTCCG

             AGCTTCATGATGGGCCAAATTTT

             AGCTTCATGATGGGCCAAATTTT

             AGCTTCATGATGGGCCAAATTTT

Depth of Coverage

▶ The number of independent sequencing reads that a nucleotide is sequenced.

Reference

ATTTCGGCTAGCTTGCAGCTTCATGATGGGCCAAATTTTCTAGAGTAGTCACTAGCTTCCG

ATTTCGGCTAGCTTGCA

TTCGGCTAGCTTGCAGC

GCTTCATGATGGGCCAAATTTTC

AGCTTCATGATGGGCCAAATTTT

CCAAATTTTCTAGAGTAGTCACTAGCTTCCG

AGCTTCATGATGGGCCAAATTTT

AGCTTCATGATGGGCCAAATTTT

AGCTTCATGATGGGCCAAATTTT

AGTCACTAGCTTCCG

GAGTAGTCACTAGCTTCCG

2X

6X

3X

Variant/Mismatch
Single Nucleotide Polymorphism (SNP)

▶ **A nucleotide sequence from a sequencing read that is different from the reference.**

Mismatch, possible sequencing error

Reference

ATTTCGGCTAGCTTGCAGCTTCATGATGGGCCAAATTTTCTAGAGTAGTCACTAGCTTCCG
ATTTCGG**A**TAGCTTGCA
 TTCGG**A**TAGCTTGCAGC

AGTCACTAGCC**C**TCCG
GAGTAGTCACTAGCC**C**TCCG

     GCTTCATG**T**TGGGCCA**C**ATTTTC
    AGCTTCATGATGGGCCAAATTTT

Variant

CCAAATTTTCTAGAGTAGTCACTAGCC**C**TCCG

AGCTTCATGATGGGCCAAATTTT
AGCTTCATGATGGGCCAAATTTT
AGCTTCATGATGGGCCAAATTTT

Variant

Genome Assembly

▶ **Matching reads together based off sequence similarity to build a larger sequence.**

ATTTCGGCTAGCTTGCA
  TTCGGCTAGCTTGCAGC
         TTGCAGCTTCATGATGGGCCAAATTTT
                CCAAATTTTCTAGAGTAGTCACTAGCTTCCG

ATTTCGGCTAGCTTGCAGCTTCATGATGGGCCAAATTTTCTAGAGTAGTCACTAGCTTCCG

Consensus Sequence

▶ Sequence that includes the most common nucleotide at each position after reference guided alignment or assembly.

Reference

```
ATTTCGGCTAGCTTGCAGCTTCATGATGGGCCAAATTTTCTAGAGTAGTCACTAGCTTCCG
ATTTCGGATAGCTTGCA                                    AGTCACTAGCCTCCG
  TTCGGATAGCTTGCAGC                              GAGTAGTCACTAGCCTCCG
            GCTTCATGTTGGGCCACATTTTC
            AGCTTCATGATGGGCCAAATTTT
                      CCAAATTTTCTAGAGTAGTCACTAGCCTCCG
            AGCTTCATGATGGGCCAAATTTT
            AGCTTCATGATGGGCCAAATTTT
            AGCTTCATGATGGGCCAAATTTT
```

Consensus

```
ATTTCGGATAGCTTGCAGCTTCATGATGGGCCAAATTTTCTAGAGTAGTCACTAGCCTCCG
```

Sample Types

Sample Type- Pure Culture/Isolate

- **High Concentration of DNA/RNA**
    - **Suitable for protocols that require high input amounts (>μg)**
- **High Quality DNA/RNA**
    - **Suitable for long sequencing read protocols (ONT & PacBio)**
- **Isolation Protocols May Need to Consider Hard to Lyse Bacteria**
    - **Bead-bashing/stringent lysis buffers**
- **Viral Cultures**
    - **Mixture of cells & virus (typically significantly higher viral load than clinical samples)**
    - **Cell culture supernatants may contain more virus and less cells**
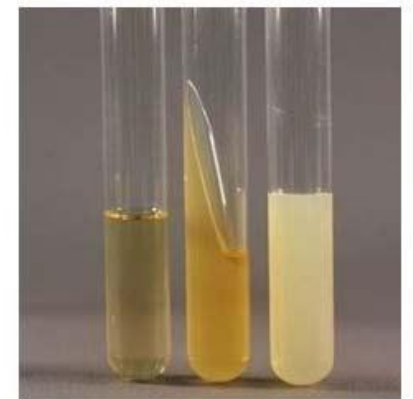
Pure Population of Bacteria

Shotgun Whole Genome Sequencing
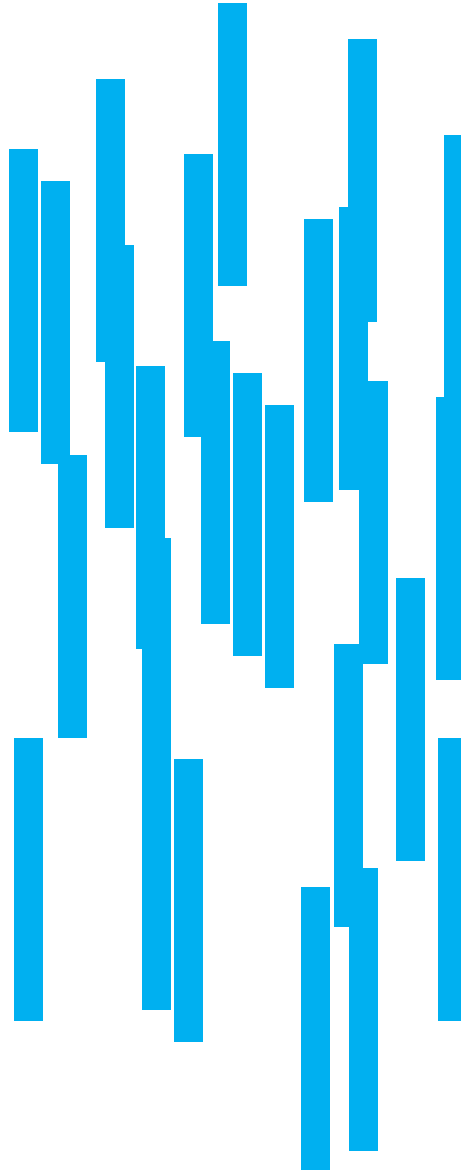
Bioinformatic Assembly of Bacteria

Cultured Isolate
Genome of ~4.9 Mbp
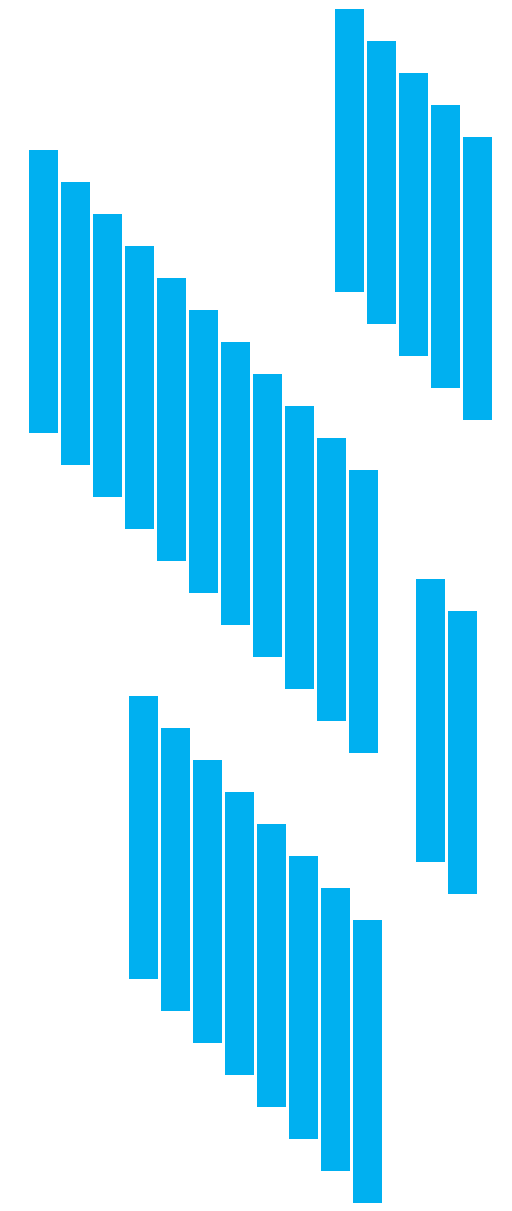
Sample Type- Clinical Sample

- **Types**
  - Nasal/Oral Swab, Stool, Blood, Puss, CSF, Urine, Sputum
- **Mixed Sample**
  - Includes human cells (DNA/RNA), host microbial flora, pathogen
- **Inhibitors**
  - Can include inhibitors of nucleic acid Isolation & PCR (e.g. High salt concentration in urine)
  - Consider what sample is being stored in (VTM, lysis buffer, proprietary reagents)
- **Variable Amounts of Pathogen**
  - Sampling inconsistencies
  - Location of infection

Sample Type- Environmental

- **Types**
  - **Swabs**
    - **Hospital equipment**
  - **Soil or water**
  - **Food**
    - **Contaminated, food-borne illness**
  - **Wastewater**
    - **Surveillance**
  - **Veterinary**
    - **Animal (domesticated or wild)**

- **Mixed samples**
  - **Multiple types of bacterial, fungal, human, animal, virus may be present**
  - **Can include inhibitors of nucleic acid Isolation & PCR**
- **May be hard to lyse**

Mixed Population of Bacteria

5 Dark Blue
4 Purple
3 Red
2 Green
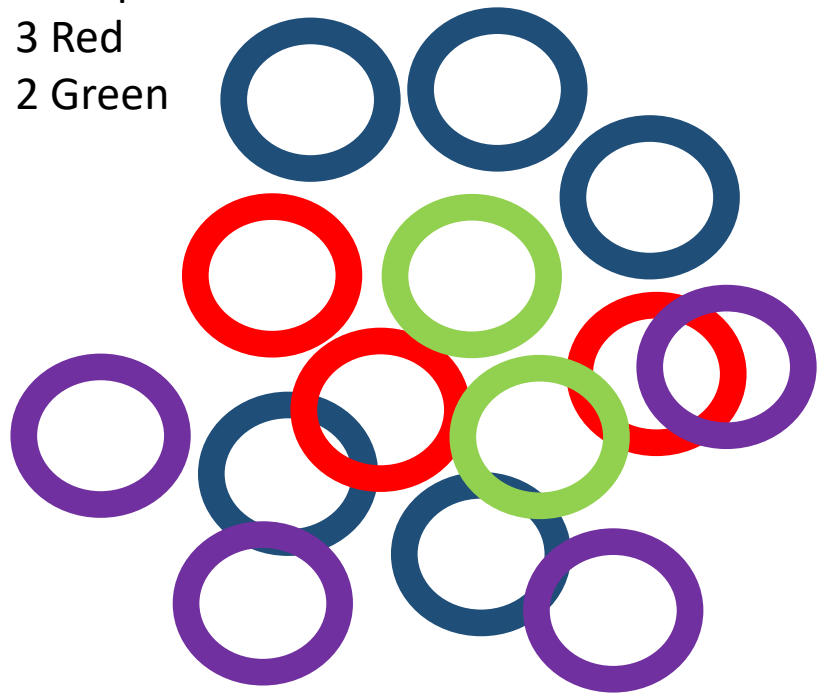
Environmental Swab
Clinical Swab

Shotgun Metagenomic Sequencing
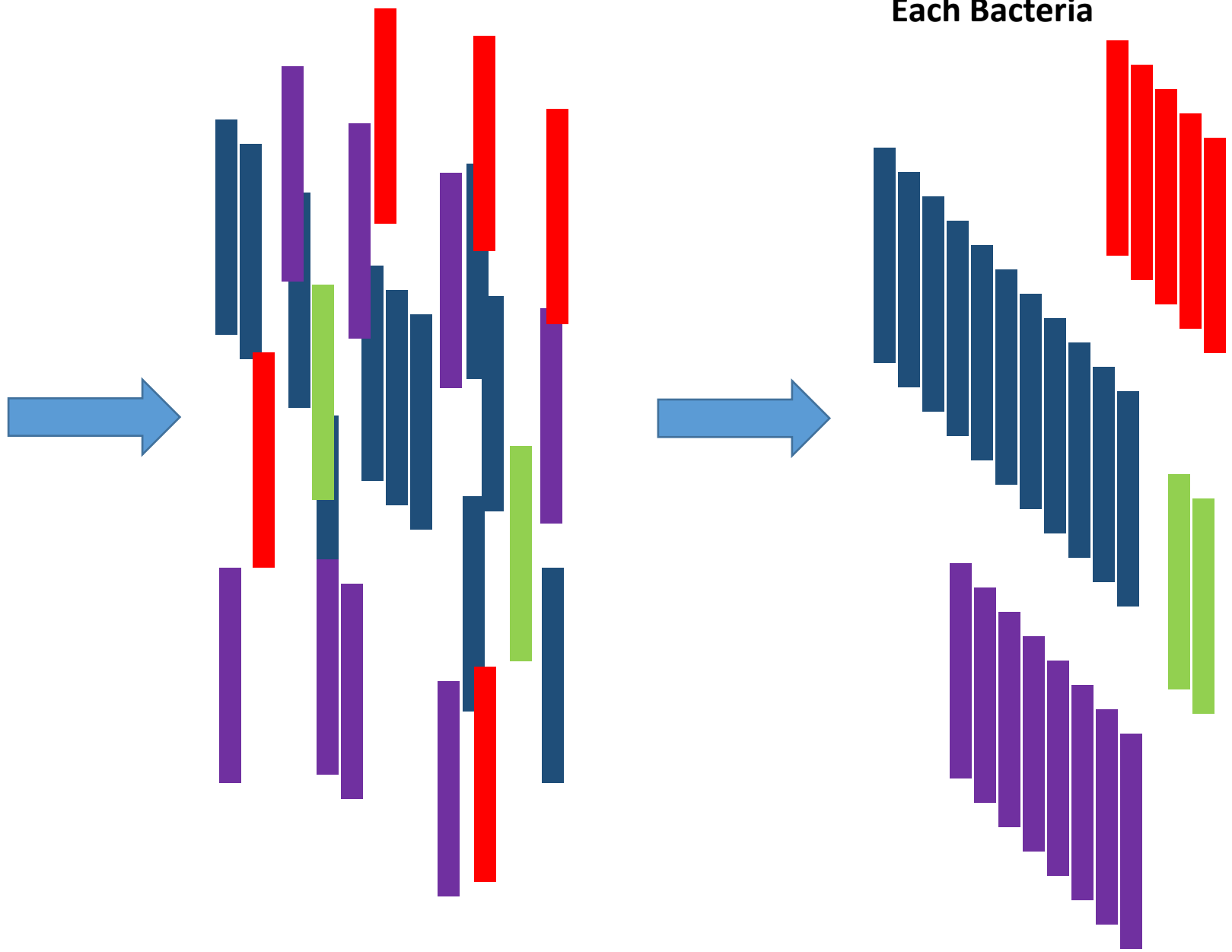
Bioinformatic Assembly of Each Bacteria

Genome Coverage Calculations

4.9 Mbp Genome
Pure Isolate
2x150bp
100X Average Coverage
2% Duplications

	MiSeq	MiSeq	NextSeq 2000	NextSeq 2000	NextSeq 2000
Reagents	v2	v3	P1	P2	P3
Read Length	2x150	2x150	2x150	2x150	2x150
Samples/run	9	15	60	240	720
Cost	$1,200	$1,750	$1,250	$3,630	$6,150
Cost/sample	$133.33	$116.67	$20.83	$15.13	$8.54

Mixed metagenomic samples will require more sequencing

https://support.illumina.com/downloads/sequencing_coverage_calculator.html

Nucleic Acid Isolation

DNA & RNA Isolation

Methods

- **Organic**
  - **Phenol-Chloroform**
- **Column-based**
  - **Qiagen, Zymo**
- **Bead-Based**
  - **ThermoFisher Dynabeads, MagMax**

- **Available from a variety of manufacturers with different capacities & elution volumes**

Considerations

- **DNA**
  - **Fragmentation**
  - **Lysis of bacterial cells**

- **RNA**
  - **Size (miRNA)**
  - **Fragmentation**
  - **DNA contamination (DNAse treatment)**
  - **Carrier RNA (total RNA & target enrichment)**

Organic DNA/RNA Isolation

Altayari W. (2016) DNA Extraction: Organic and Solid-Phase. In: Goodwin W. (eds) Forensic DNA Typing Protocols. Methods in Molecular Biology, vol 1420. Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-3597-0_5
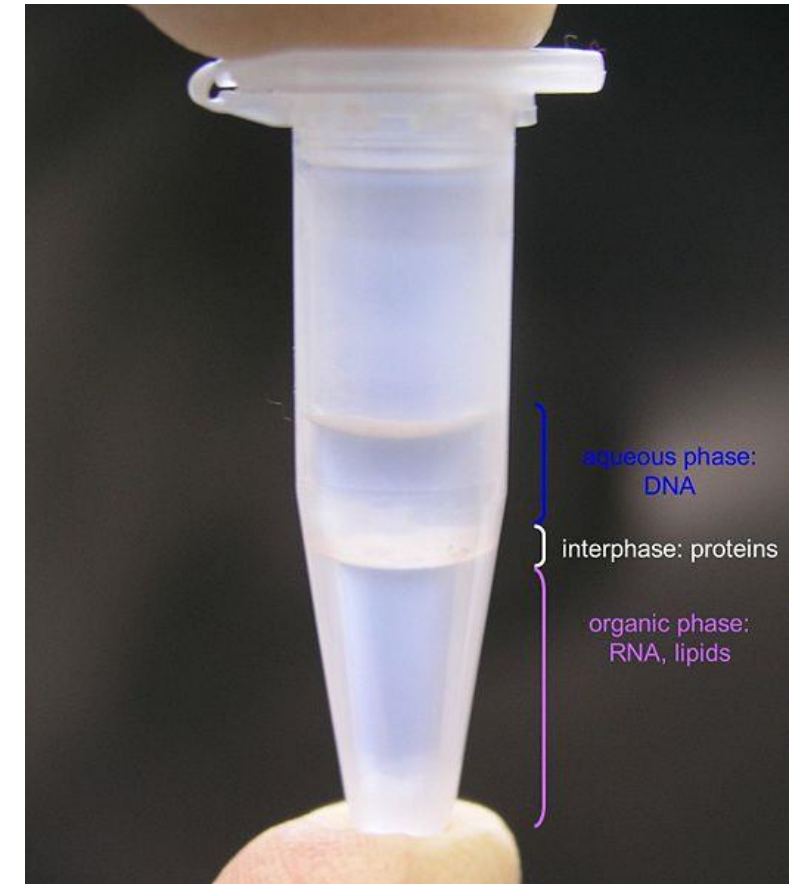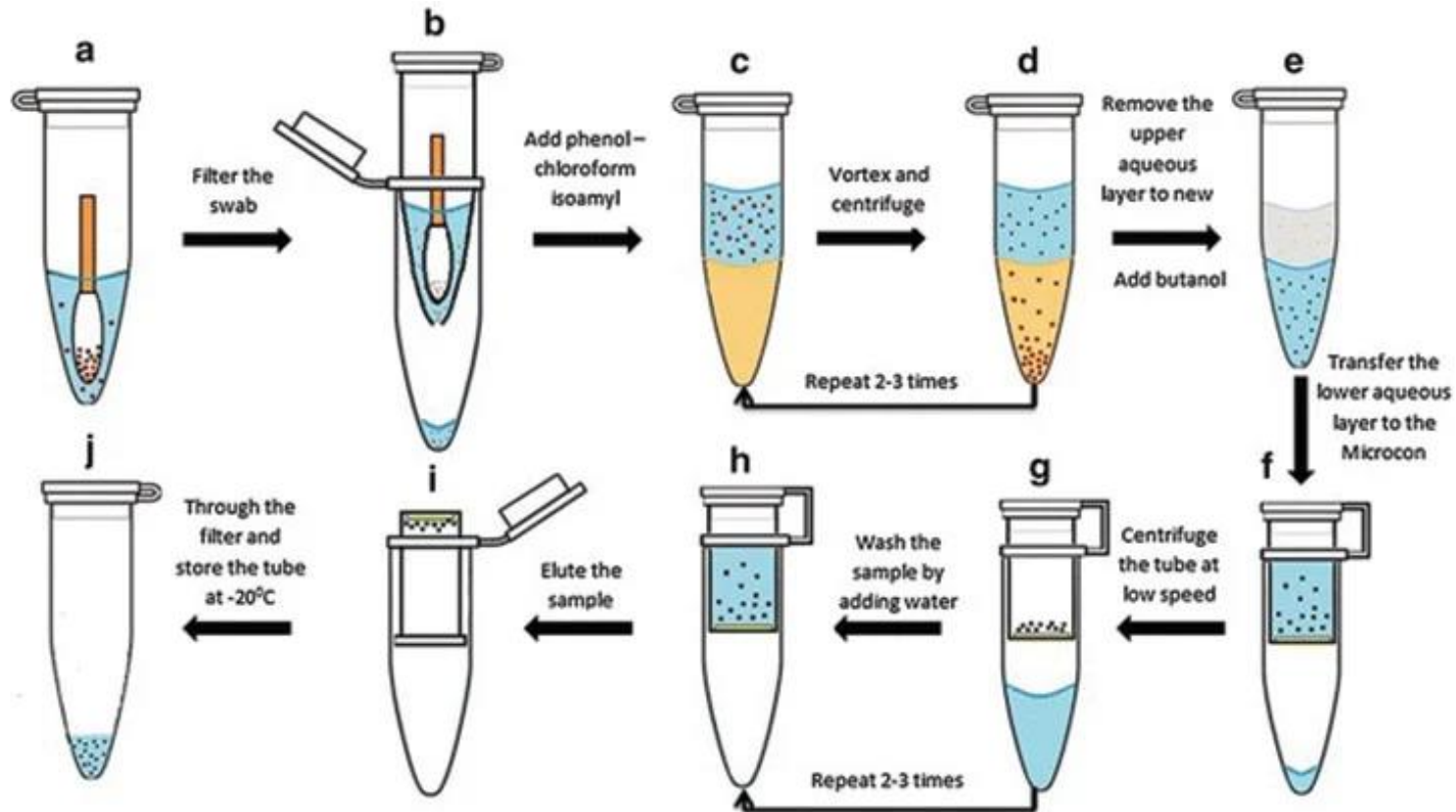
Column-Based DNA/RNA Isolation

Lee, H., Na, W., Park, C. *et al. Sci Rep* **8,** 5467 (2018).

Bead-Based Isolation

Cell Harvest → Lysis → Add Beads → Wash → Separation → Release

Magnetic
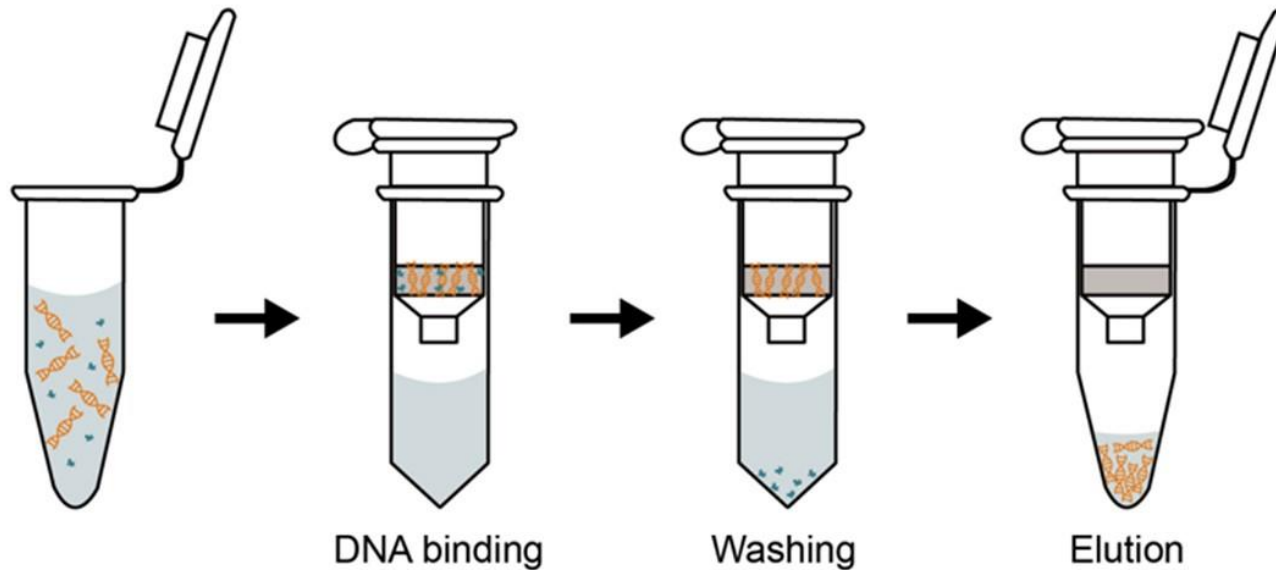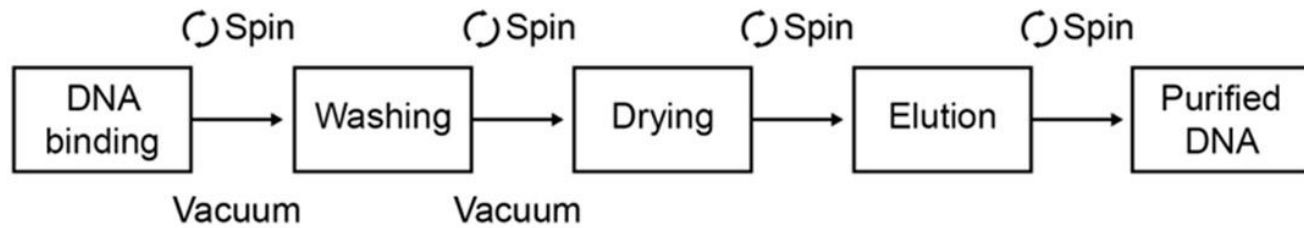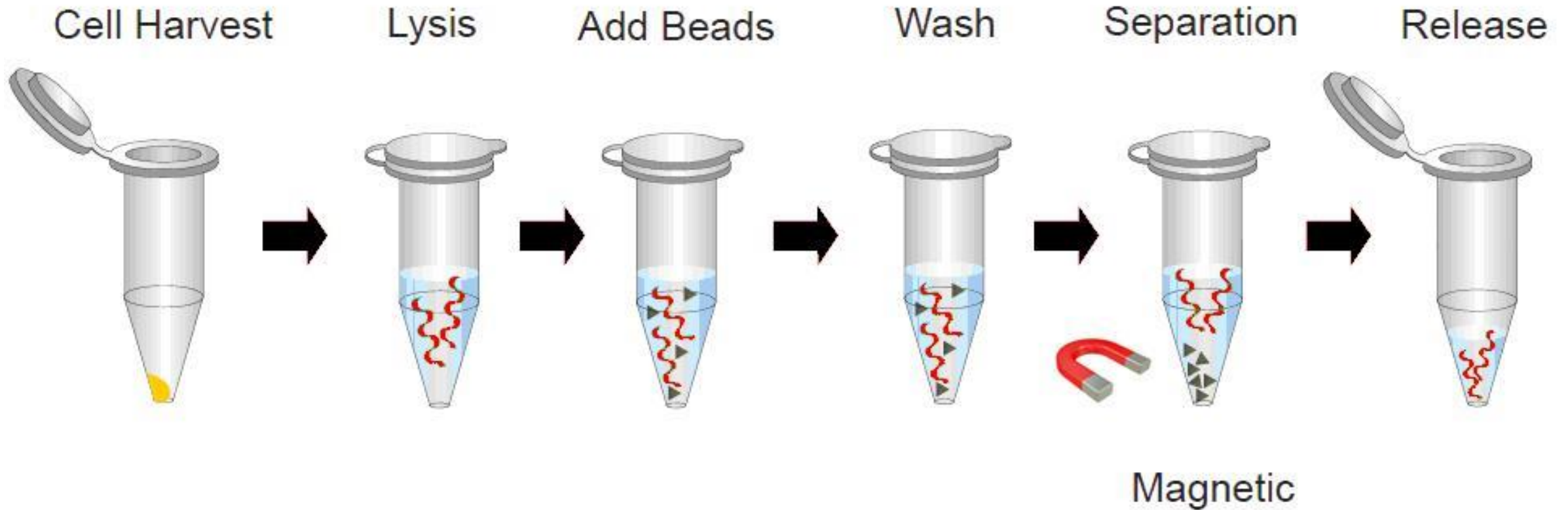
DNA/RNA Quality Control

- **Concentration**
  - **Absorption- Nanodrop**
  - **Fluorometric- Qubit**
  - **Electrophoresis- Bioanaylzer**

- **Size**
  - **Electrophoresis- Gel or Bioanaylzer**

- **Molarity**
  - **Size + concentration**

- **Integrity**
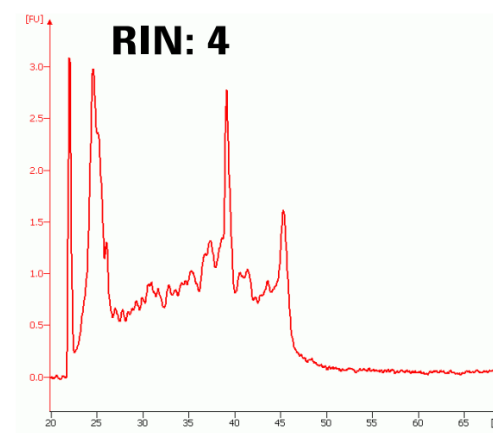 - **RNA integrity number (RIN) - Bioanalyzer**
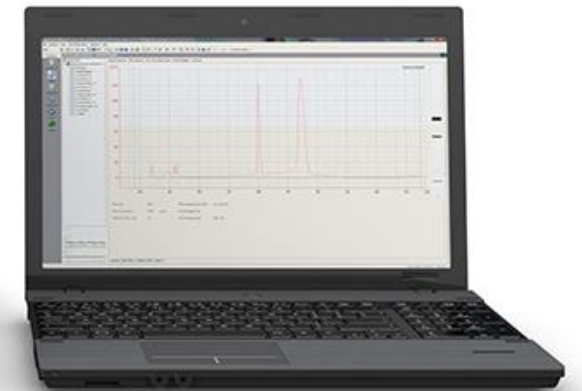
Nanodrop
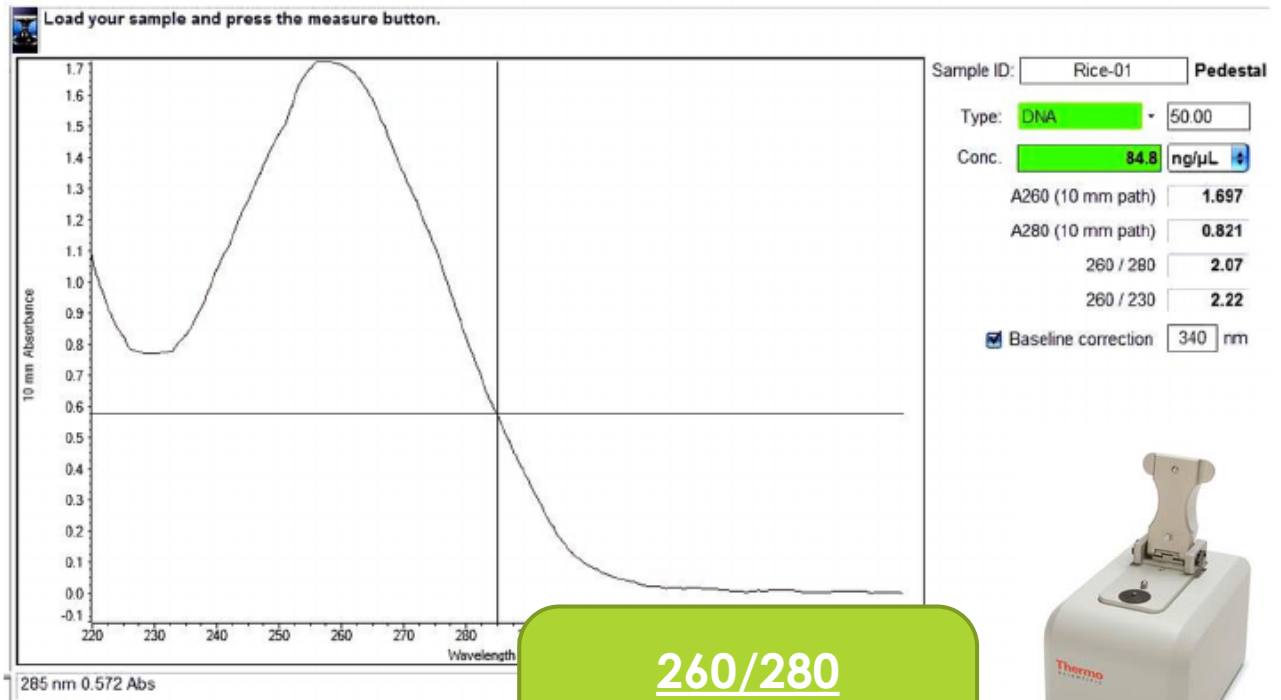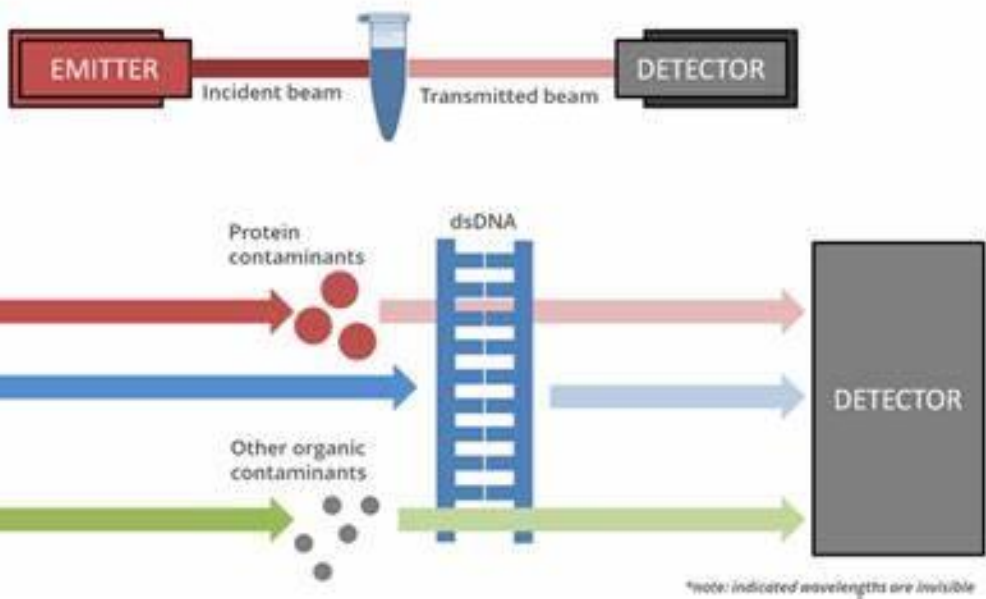
260/280
Pure DNA ~1.8
Pure RNA ~2.0

Qubit- Fluorometric Quantification

https://upcvmda-pl480.weebly.com/educational-articles/

Qubit- Fluorometric Quantification

Standards from kit — 10 μL, 10 μL — Add Qubit working solution for a total volume of 200 μL

User samples — 1–20 μL, 1–20 μL, 1–20 μL — Add Qubit working solution for a total volume of 200 μL

Vortex tubes for 2–3 sec

Incubate at room temperature for 2 min (DNA, RNA) or 15 min (protein)

Read tubes in Qubit Fluorometer
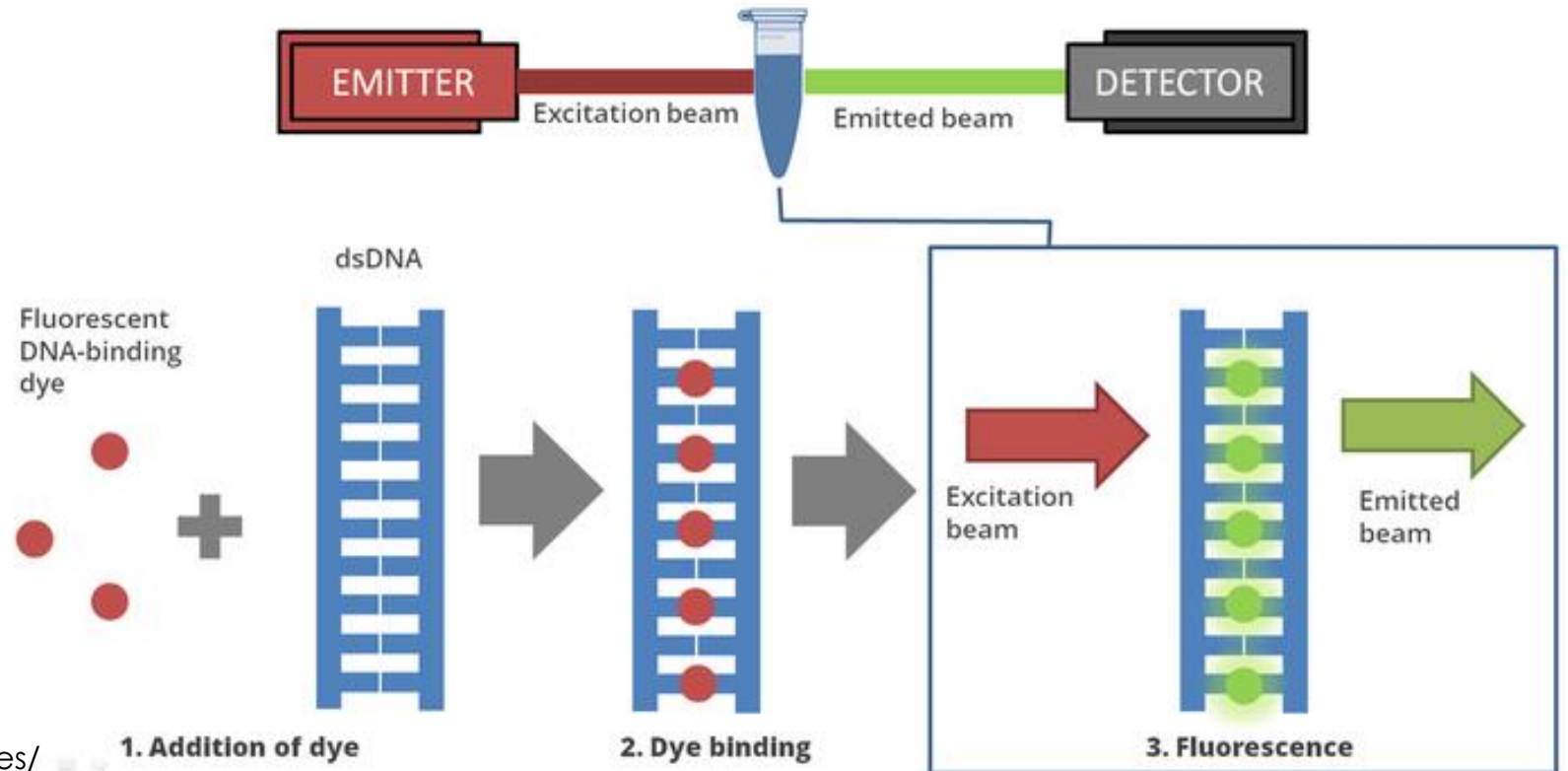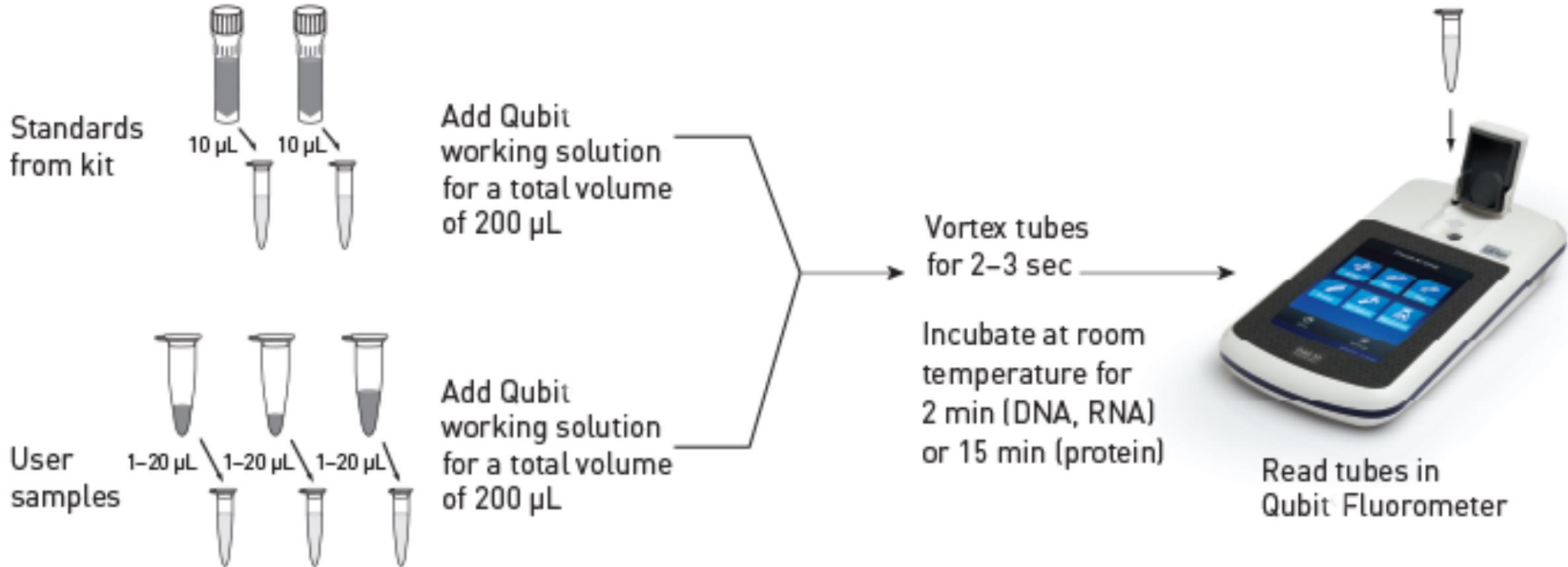
Bioanalyzer- Agilent
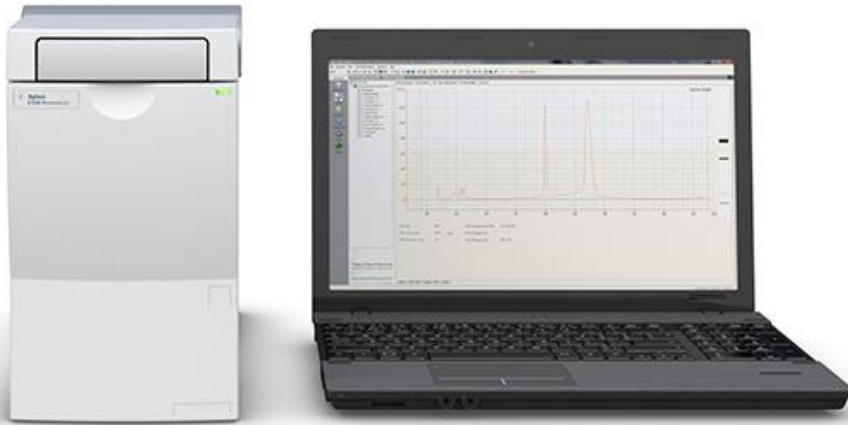
Chip-based gel electrophoresis
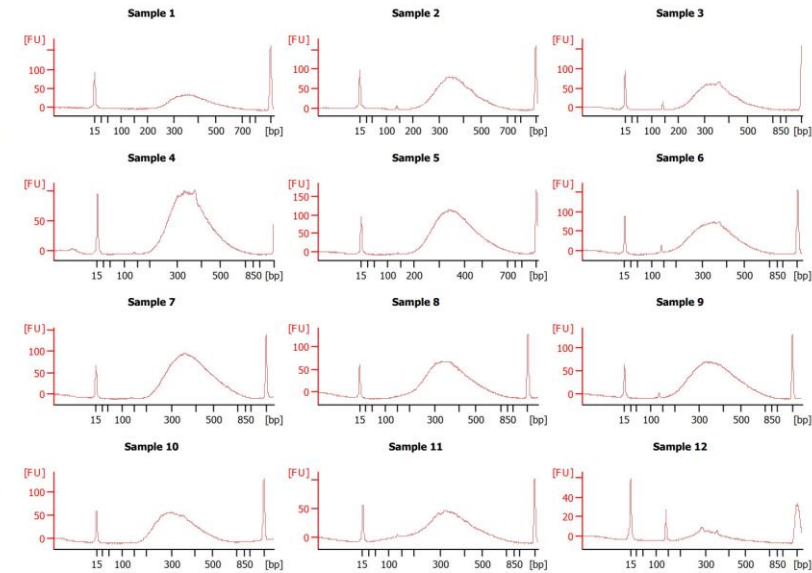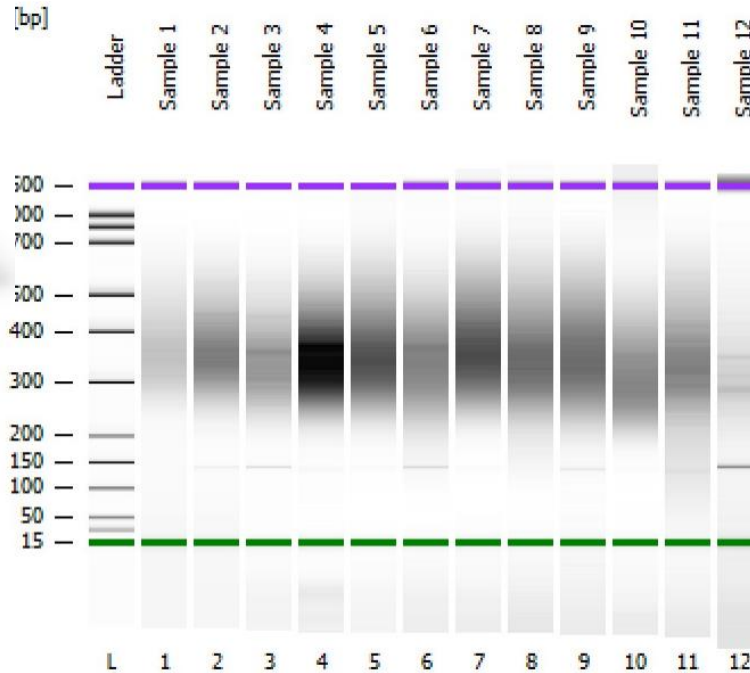
https://www.agilent.com/

Genomic DNA

Agilent
DNA 7500 and
DNA 12000

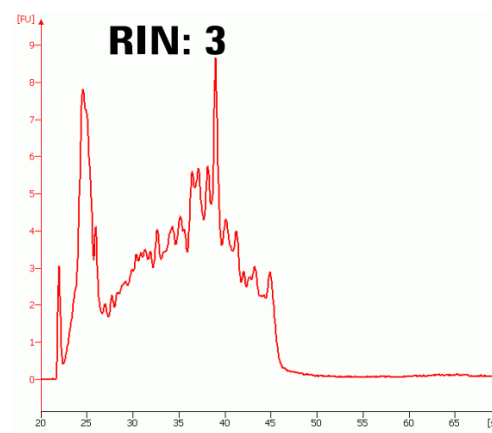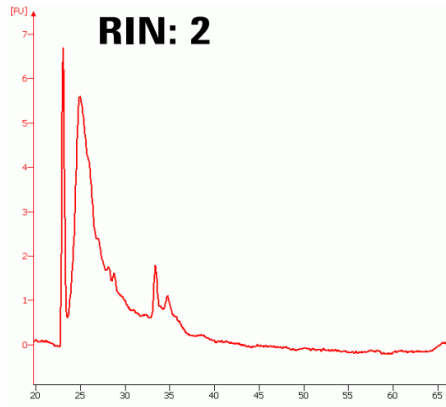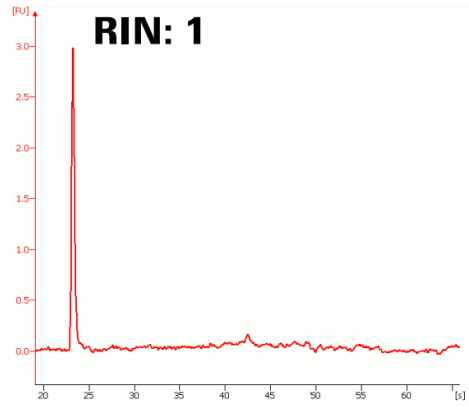Standards in Genomic Sciences. 2017. 12. 27. 10.1186/s40793-017-0239-1.
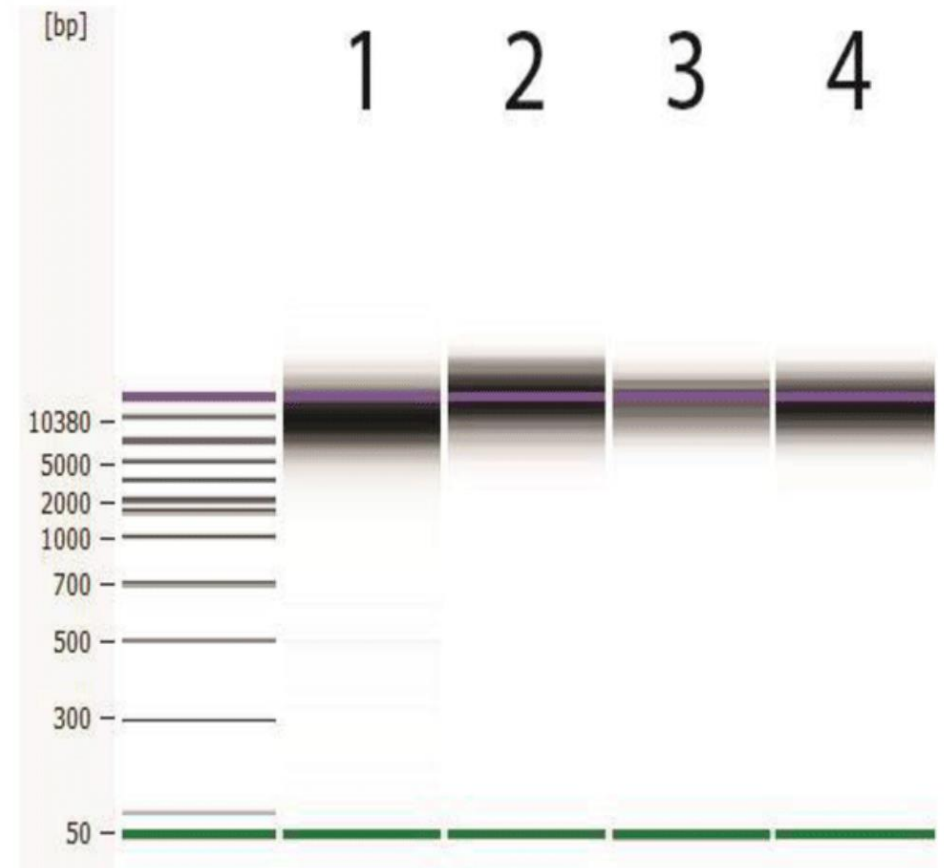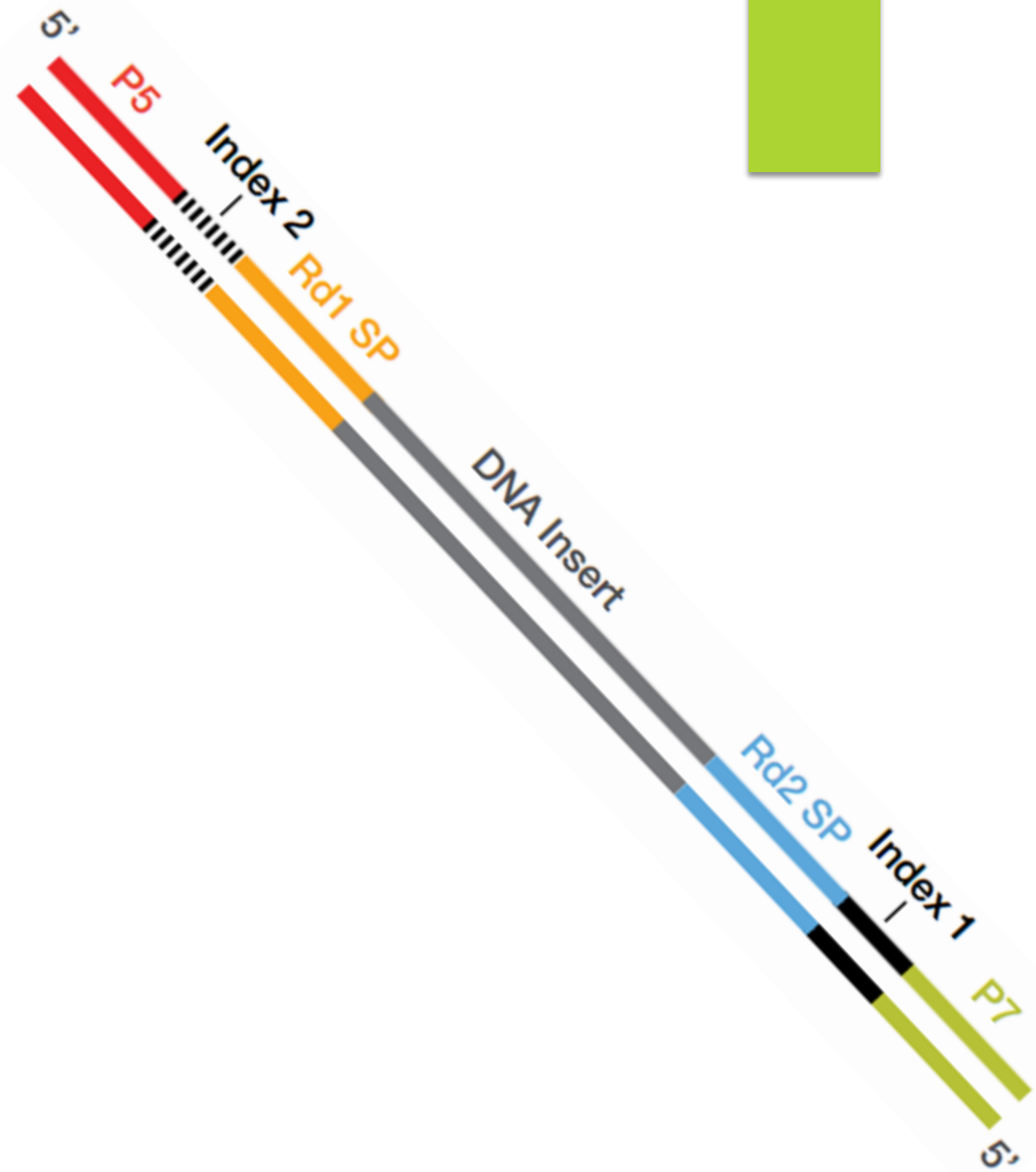
RNA Integrity Number- RIN

- The RNA integrity number (RIN) is a tool designed to help scientists estimate the integrity of total RNA samples.

- The RIN extension automatically assigns an integrity number to a eukaryote total RNA sample analyzed on the 2100 Bioanalyzer system.

- Sample integrity determined by the ratio of the ribosomal bands & the entire electrophoretic trace of the RNA sample, including the presence or absence of degradation products.
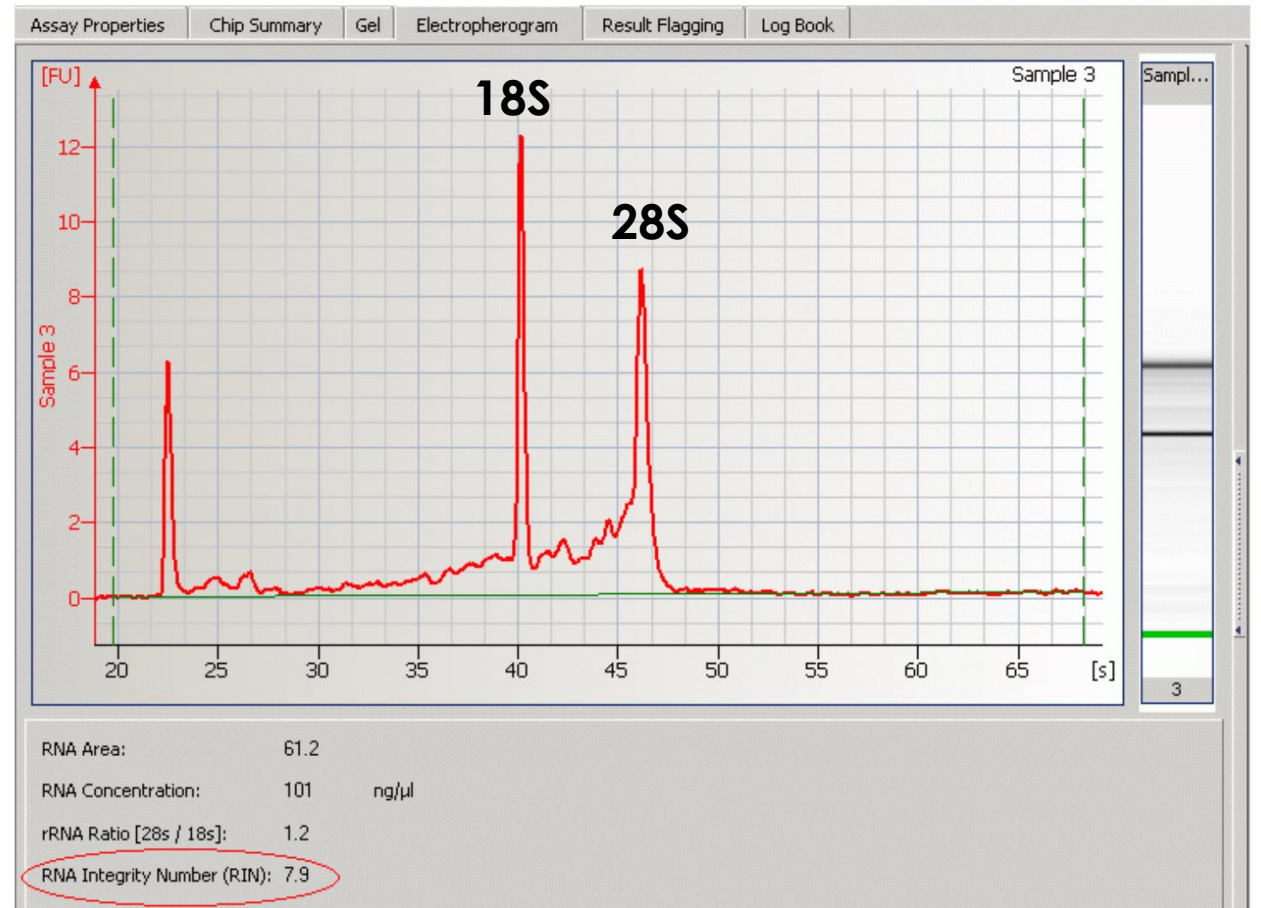
RNA Integrity Number- RIN

Library Prep

Illumina

- https://www.illumina.com/techniques/sequencing/ngs-library-prep.html
- https://www.illumina.com/library-prep-array-kit-selector.html
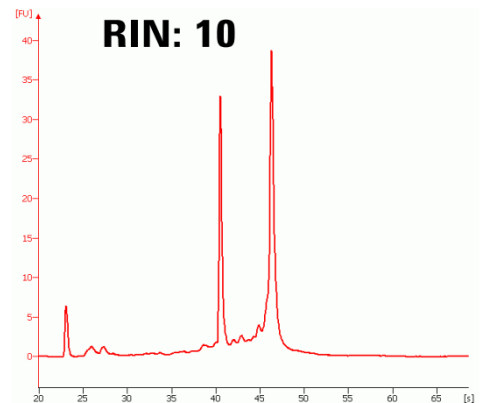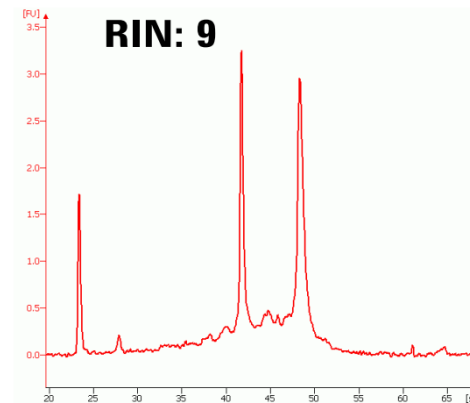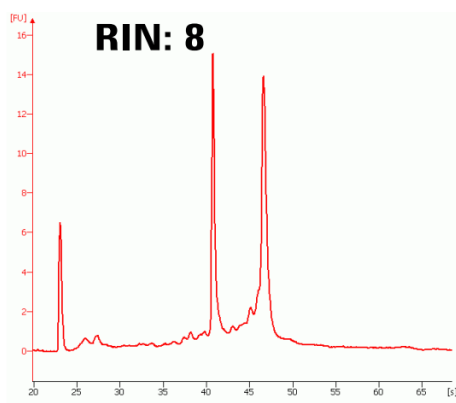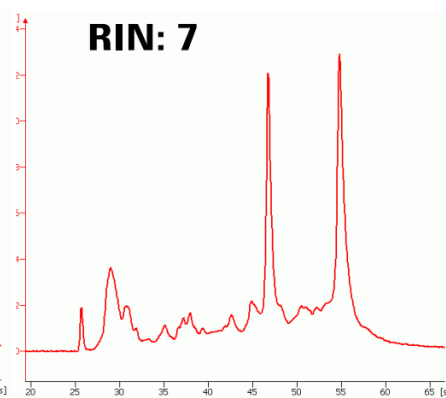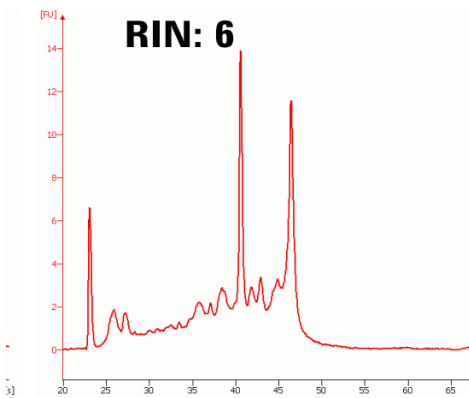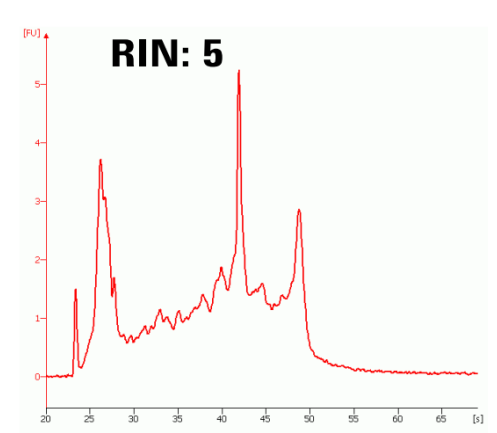
Library preparation kits for diverse methods

Ask virtually any question related to the genome, transcriptome, or epigenome of any organism with NGS library prep kits optimized for Illumina sequencers

ONT

▶ **https://store.nanoporetech.com/us/sample-prep.html**

Prepare ›

Extract nucleic acid from your sample and prepare it for sequencing
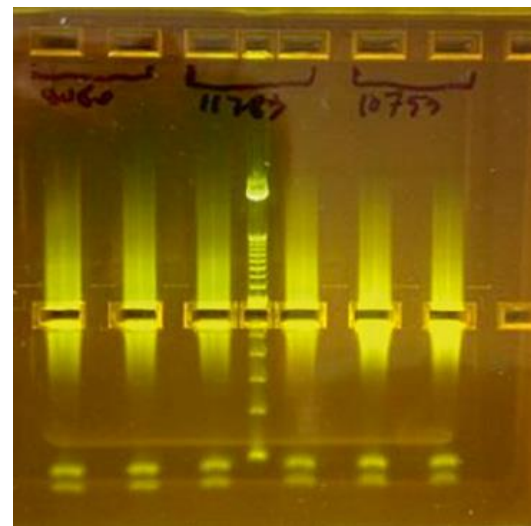
Sequence ›

Sequence your library on a device that suits your needs

Analyse ›

Analyse your data from as soon as you start sequencing

Purification & Size Selection

Purification & Size Selection Methods

- **Column-based Purification**
  - **Removal of nucleotides, enzymes, buffers**
- **Gel Purification**
  - **Extract and purify DNA from an agarose gel of a specific size**
- **Bead - (SPRI AMPure XP, Axygen AxyPrep MAG)**
  - **SPRI technology uses paramagnetic beads to selectively bind nucleic acids by type and size, and are used for high-performance isolation, purification, and cleanup protocols.**
- **Pippin Prep- Sage Science**
  - **Automated gel-cassette size selection**

Column-based PCR Clean Up

https://www.qiagen.com/
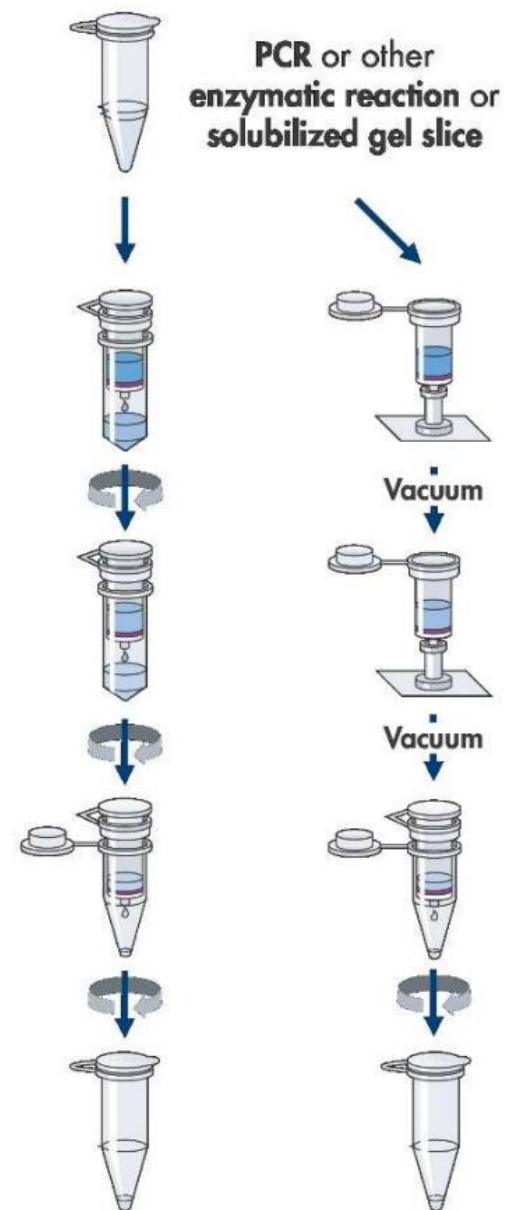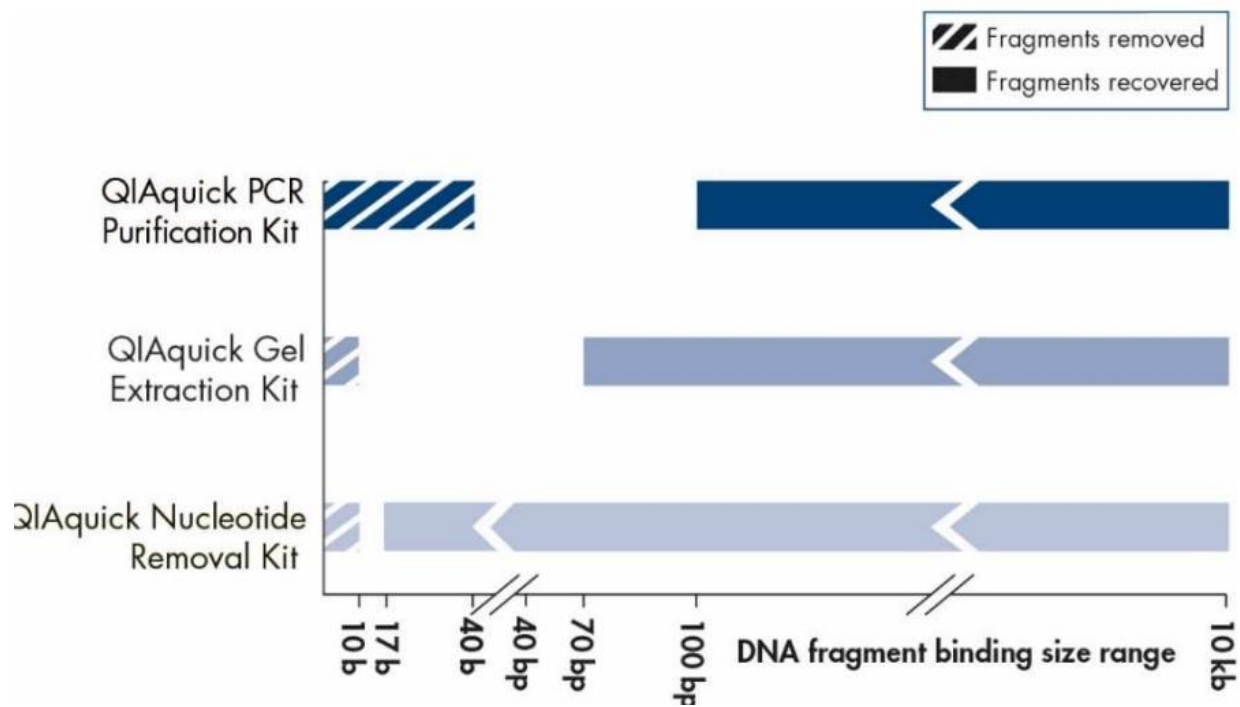
Column-based PCR Clean Up

	QIAquick PCR Purification	QIAquick Nucleotide Removal	QIAquick Gel Extraction
Maximum binding capacity	10 µg	10 µg	10 µg
Maximum weight of gel slice	–	–	400 mg
Minimum elution volume	30 µl	30 µl	30 µl
Capacity of column reservoir	800 µl	800 µl	800 µl
Typical recoveries			
Recovery of DNA	90–95% (100 bp – 10 kb)	80–95% (40 bp – 10 kb)	70–80% (70 bp – 10 kb)
Recovery of oligonucleotides (17–40mers)	0	60–80%	10–20%
Recovered			
Oligonucleotides	–	17–40mers	–
dsDNA	100 bp – 10 kb	40 bp – 10 kb	70 bp – 10 kb
Removed			
<10mers	**YES**	**YES**	**YES**
17–40mers	**YES**	No	No

Table 1. QIAquick DNA cleanup guide

	From solutions			From gels
	QIAquick PCR Purification Kit	QIAquick Nucleotide Removal Kit	QIAquick Gel Extraction Kit	QIAquick Gel Extraction Kit
Alkaline phosphatase	**YES**	**YES**	**YES**	**YES**
cDNA synthesis	**YES**	No	No	**YES**
DNase, nuclease digestion	**YES**	**YES**	**YES**	**YES**
Kinase				
DNA fragments	**YES**	**YES**	**YES**	**YES**
Oligonucleotides	No	**YES**	No	No
Ligation	**YES**	**YES**	**YES**	**YES**
Nick translation	**YES**	**YES**	**YES**	**YES**
PCR	**YES**	No	No	**YES**
Random priming	**YES**	**YES**	**YES**	**YES**
Restriction digestion	**YES**	**YES**	**YES**	**YES**
Tailing				
DNA fragments	**YES**	**YES**	**YES**	**YES**
Oligonucleotides	No	**YES**	No	No

https://www.qiagen.com/

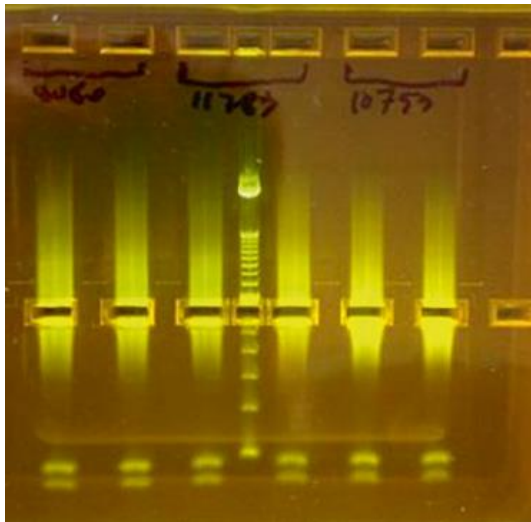# Bead-based PCR Clean up

- **SPRI AMPure XP PCR Clean-Up (1.8-2.0x volume beads:PCR product)**

Figure 1 Workflow for PCR Purification
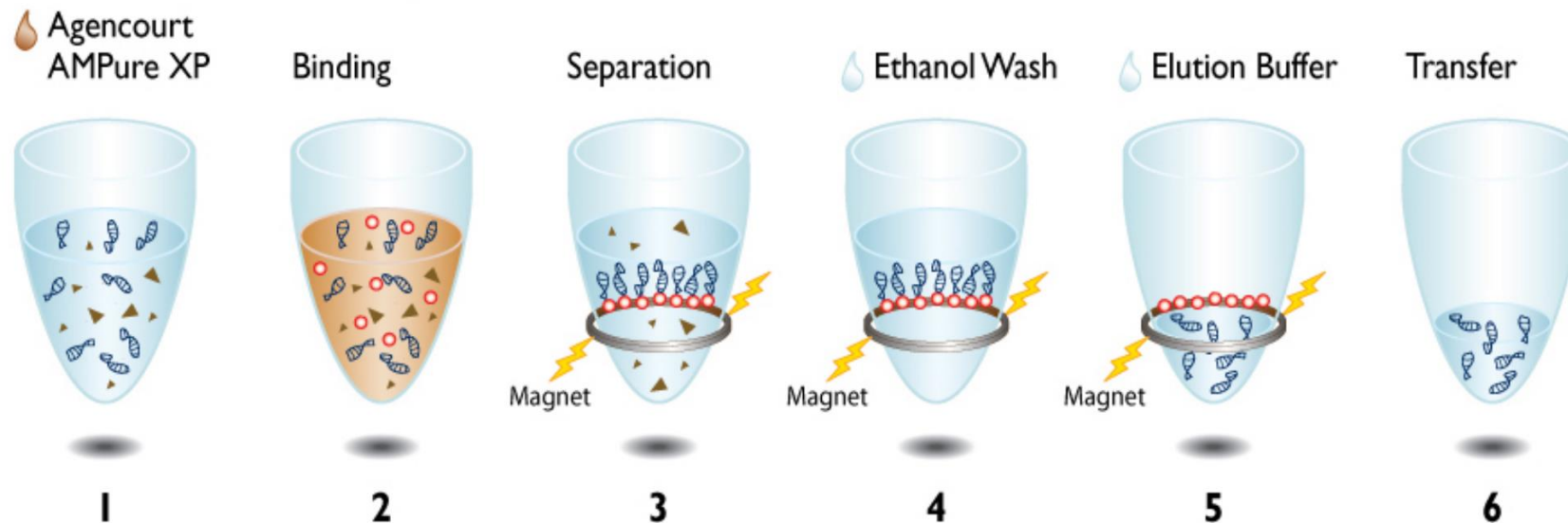
Gel Purification & Size Selection

- Agarose gel electrophoresis with DNA ladder

- Cut out desired size range
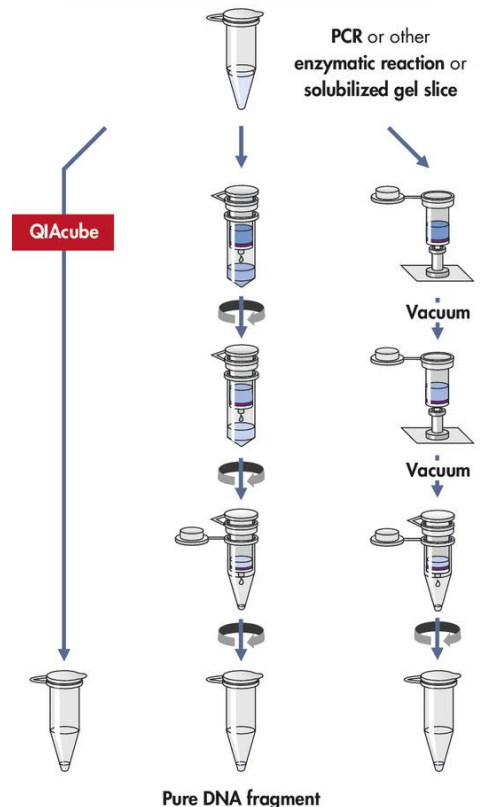
- Significant loss of DNA/library

Solubilize Gel

Optimal pH   pH too High

QIAquick and MinElute Procedure

PCR or other enzymatic reaction or solubilized gel slice

QIAcube

Vacuum

Vacuum

Pure DNA fragment

From solutions	5min
From gels	15min

https://www.qiagen.com/us/products/discovery-and-translational-research/dna-rna-purification/dna-purification/dna-clean-up/qiaquick-gel-extraction-kit/

Bead-based Size Selection

- SPRI AMPure XP (Axygen AxyPrep MAG)

  - Ratio of volume of beads to volume of suspended DNA/library

  - As a general rule, increasing the ratio of SPRIselect volume to sample volume will increase the efficiency of binding smaller fragments.

  - Can complete 2 captures to select size
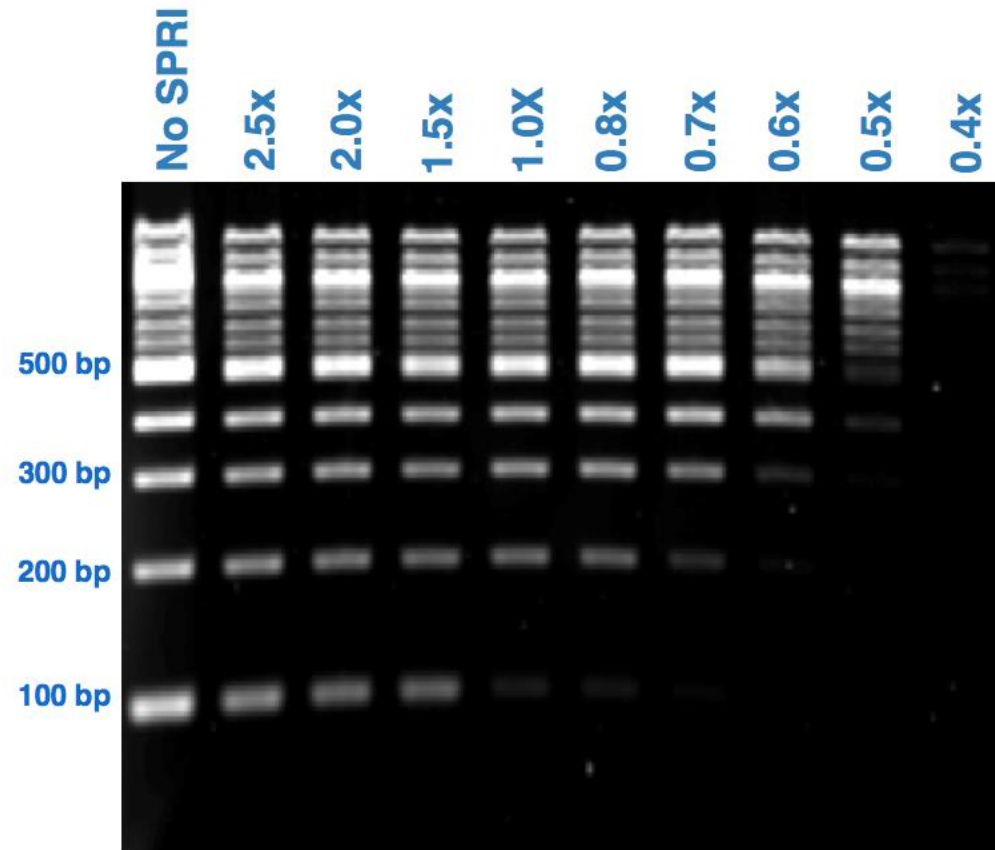
Bead-based Size Selection- Left Side

Left Side Size Selection Process Overview
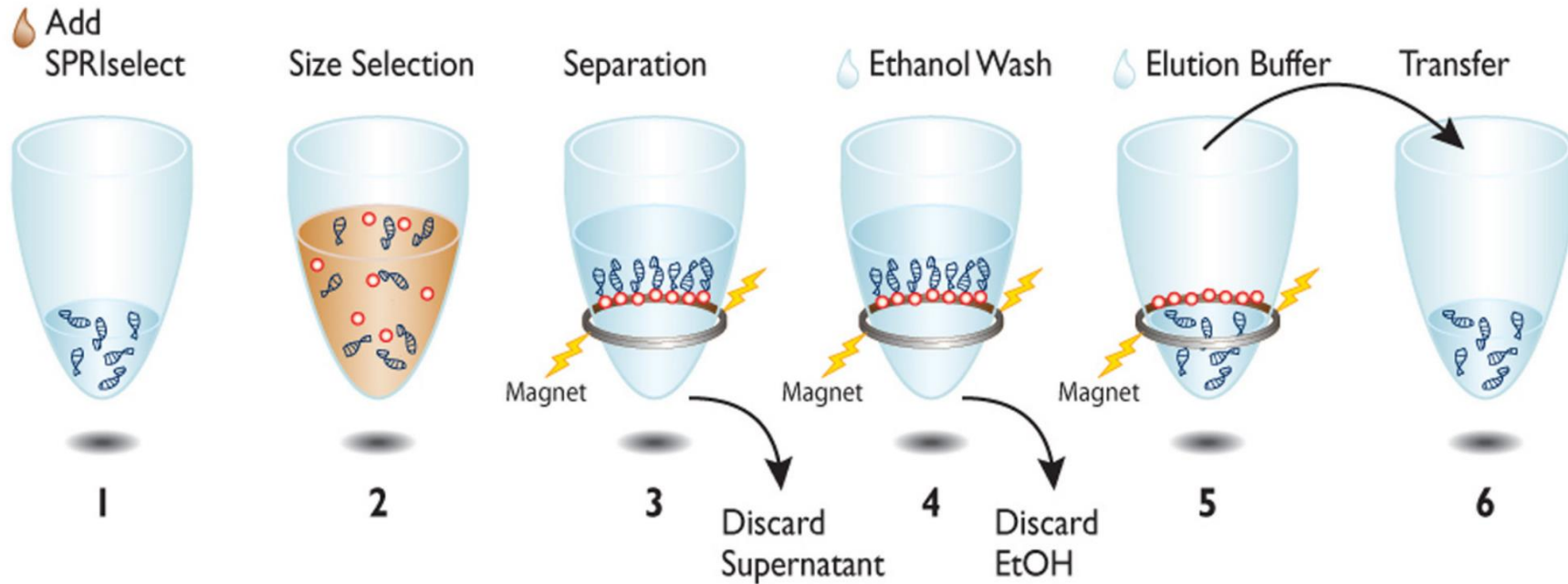
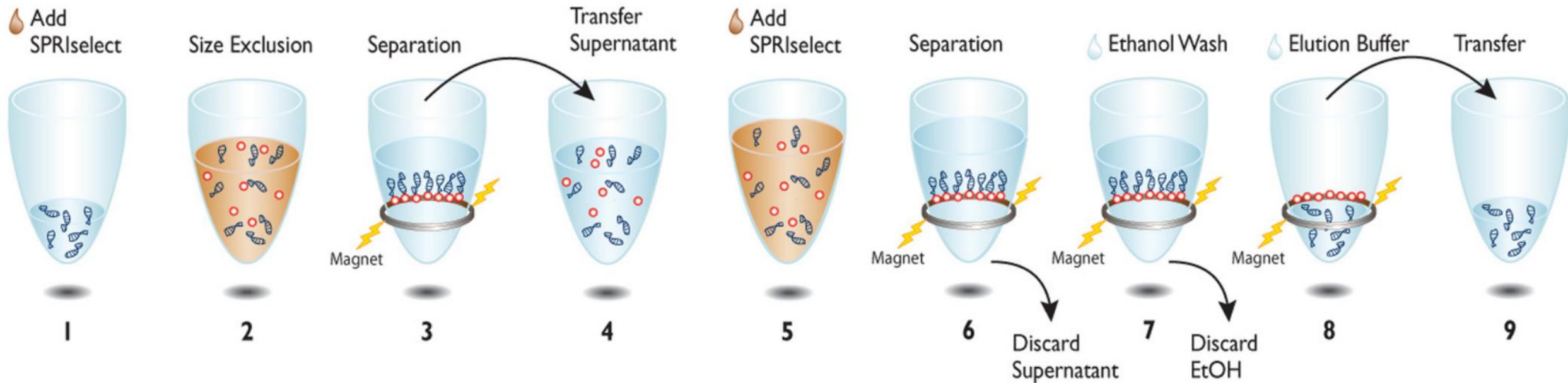**Same procedure as PCR Clean-Up**

Bead-based Size Selection- Left Side

Left Side Size
Selection- defines
the new start point
of the sample's
size distribution

Figure 1 Agilent High Sensitivity DNA chip Electropherogram.
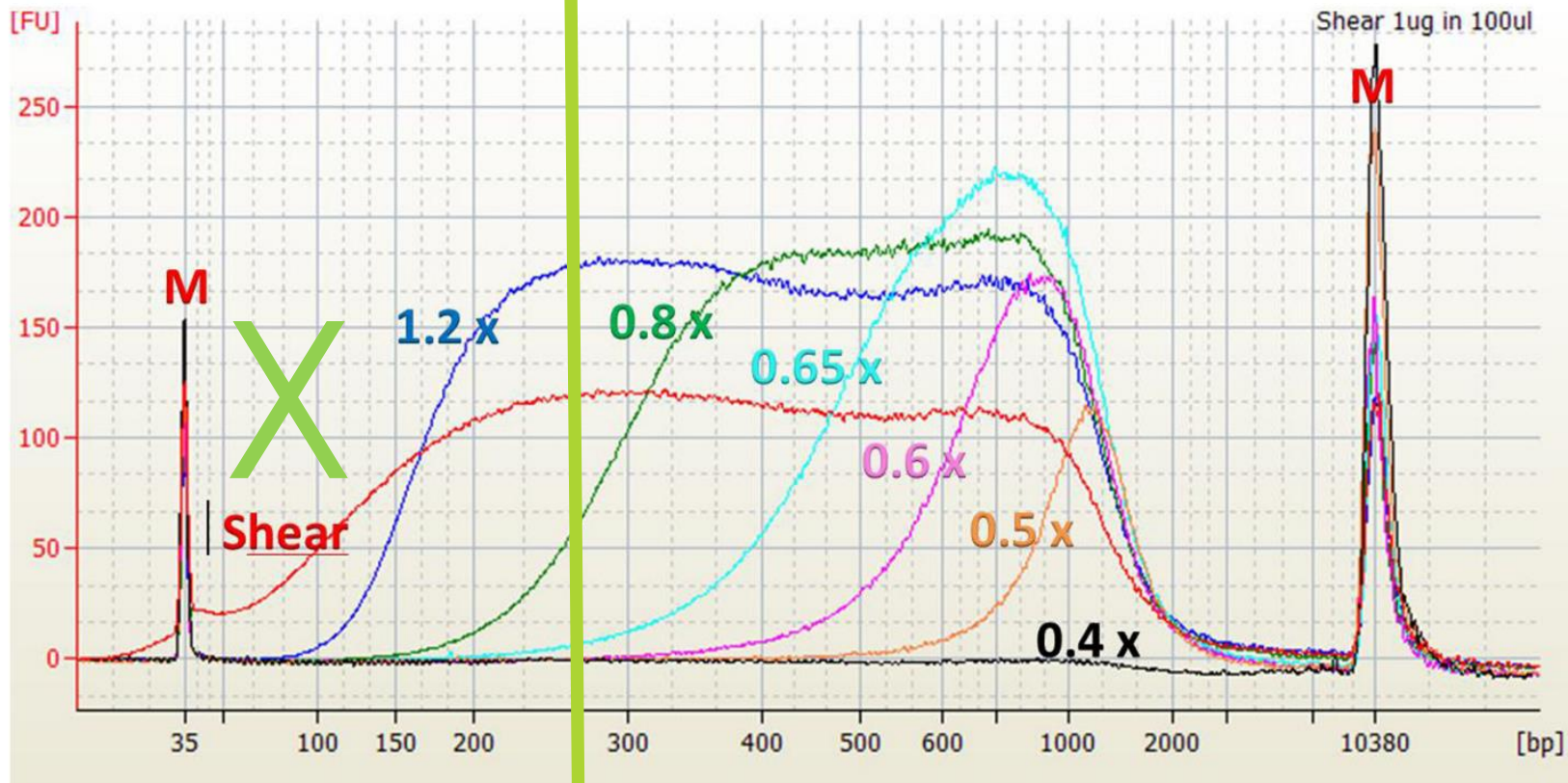
Bead-based Size Selection- Right Side

Right Side Size Selection Process Overview

Bead-based Size Selection- Right Side

Figure 3 Agilent High Sensitivity DNA chip Electropherogram.
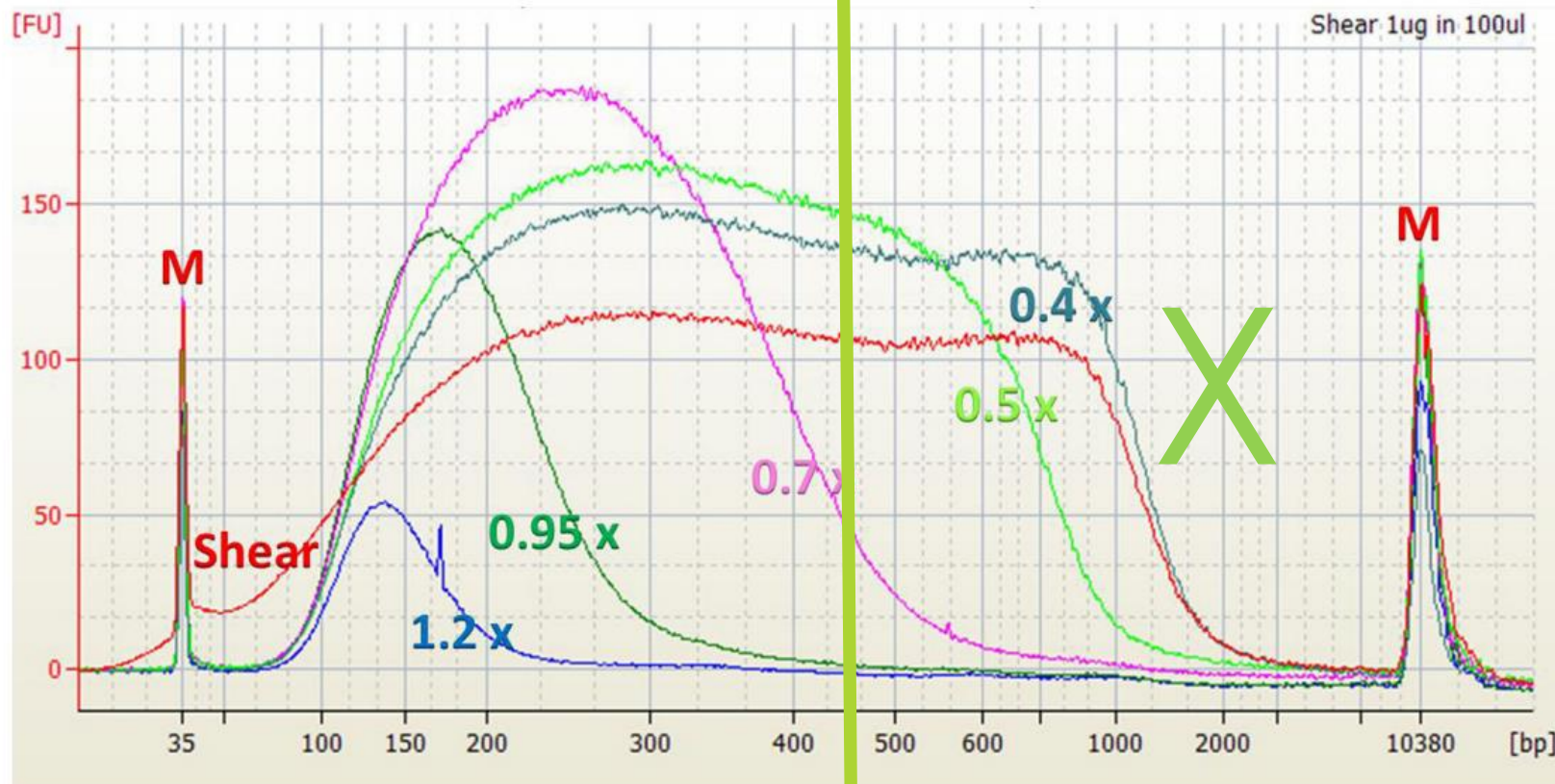
Right Side Size Selection- defines the new end point of the sample's size distribution

Double-Sided Size Selection

Left Side Size Selection- defines the new start point of the sample's size distribution

Right Side Size Selection- defines the new end point of the sample's size distribution

Complete both left & right side selection

Figure 5 Agilent High Sensitivity DNA chip Electropherogram.
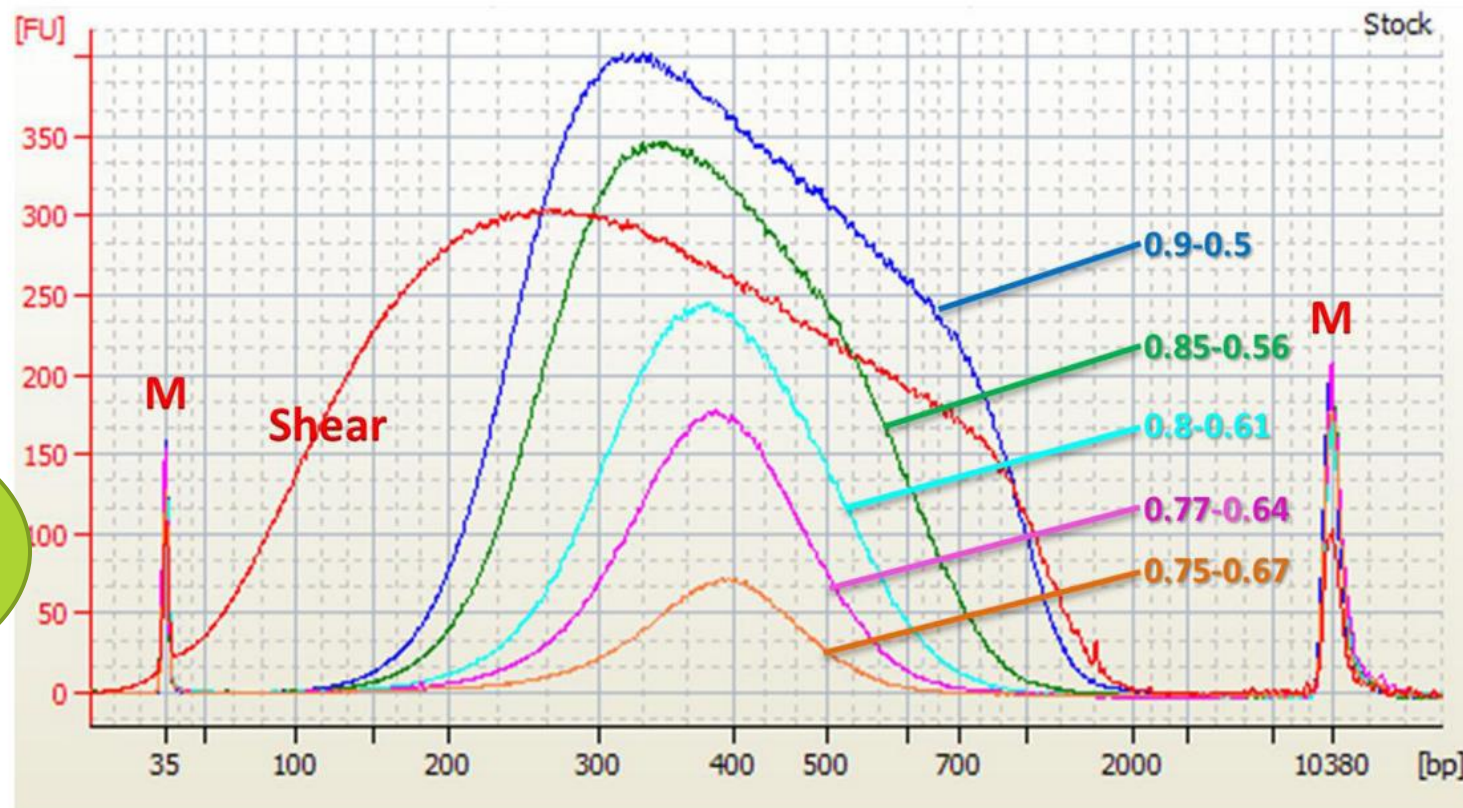
Pippen Prep- Sage Sciences

▶ **Cassette-based gel electrophoresis.**

▶ **Target sizes or ranges of sizes are entered in software, and fractions are collected in buffer.**

▶ **Up to 5 samples per gel cassette may be run, with no possibility of cross contamination.**

▶ **Requires expensive machine & reagents.**

https://sagescience.com/products/pippin-prep/
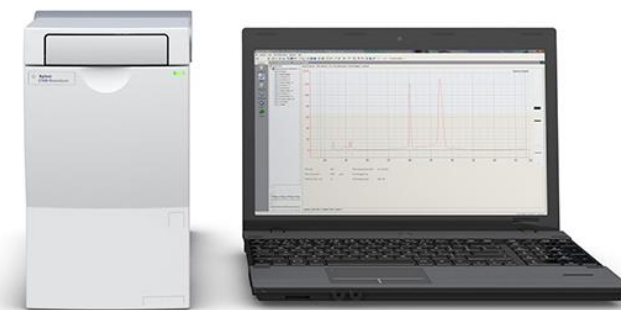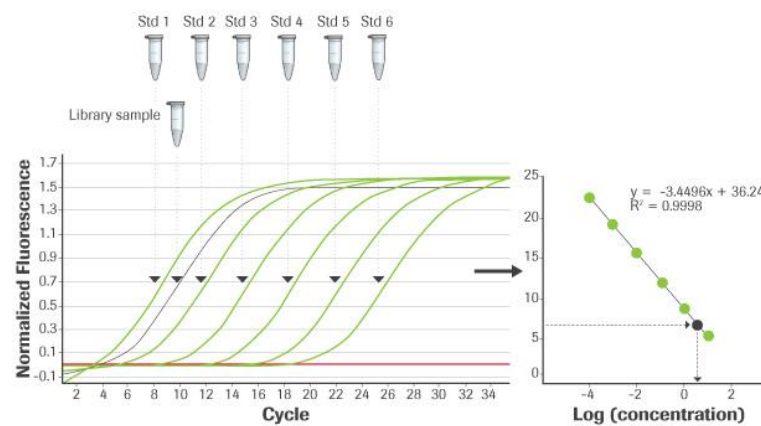
Normalization

Sample/Library Normalization

Non-Normalized Libraries
Different concentrations

A   B   C

Pool →

Raw Sequence Data:
20% Sample A
50% Sample B
30% Sample C

Non-uniform data
distribution- X

Sequence

Normalized Libraries
Same concentrations

A   B   C

Pool →

Raw Sequence Data:
33.3% Sample A
33.3% Sample B
33.3% Sample C

Data OK
Equal distribution-✓

Sequence
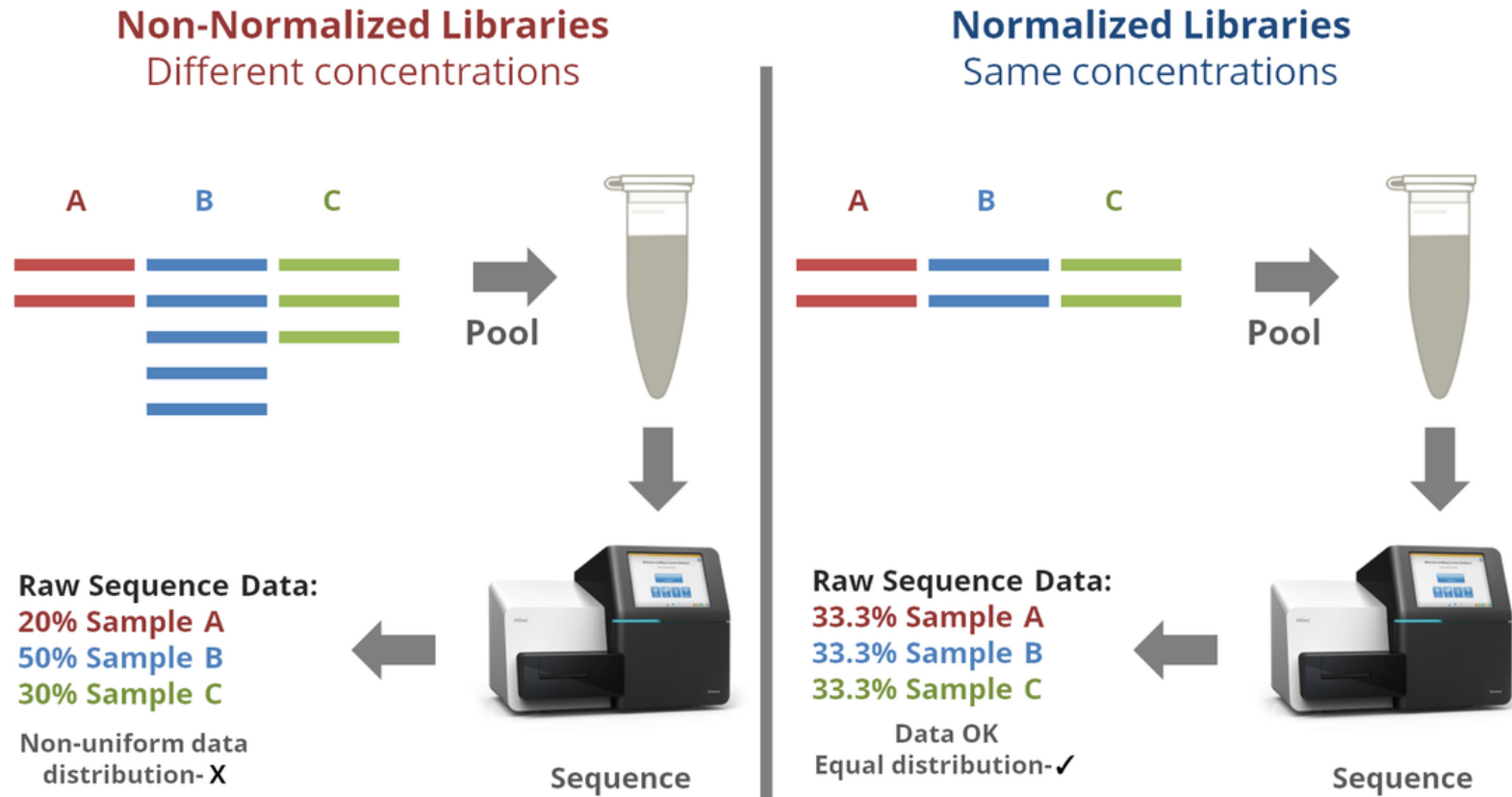
Sample/Library Normalization

- **Mix equal amounts by concentration**
  - **Assumes similar size**
- **Mix equal amounts by molarity**
  - **Requires size determination**
- **qPCR**
  - **Requires standard curve that is a library of similar type to protocol (insert size, PCR vs fragmented DNA)**
- **Bead-based**
  - **Saturation of DNA binding beads**

Sequencing Reads per Sample

Poor Normalization

% Reads Identified (PF) Per Index
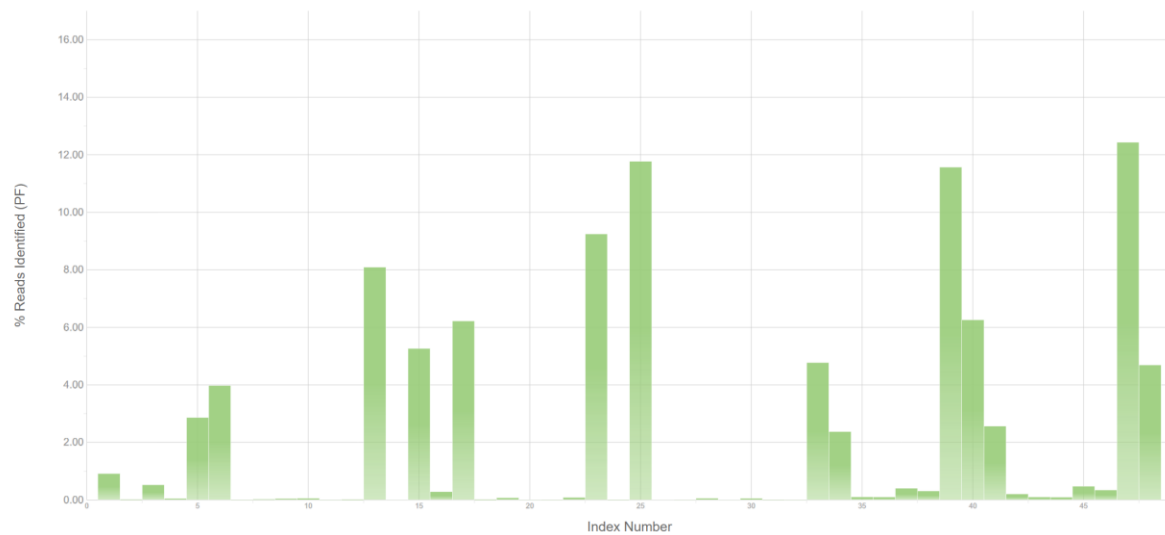
Good Normalization

% Reads Identified (PF) Per Index
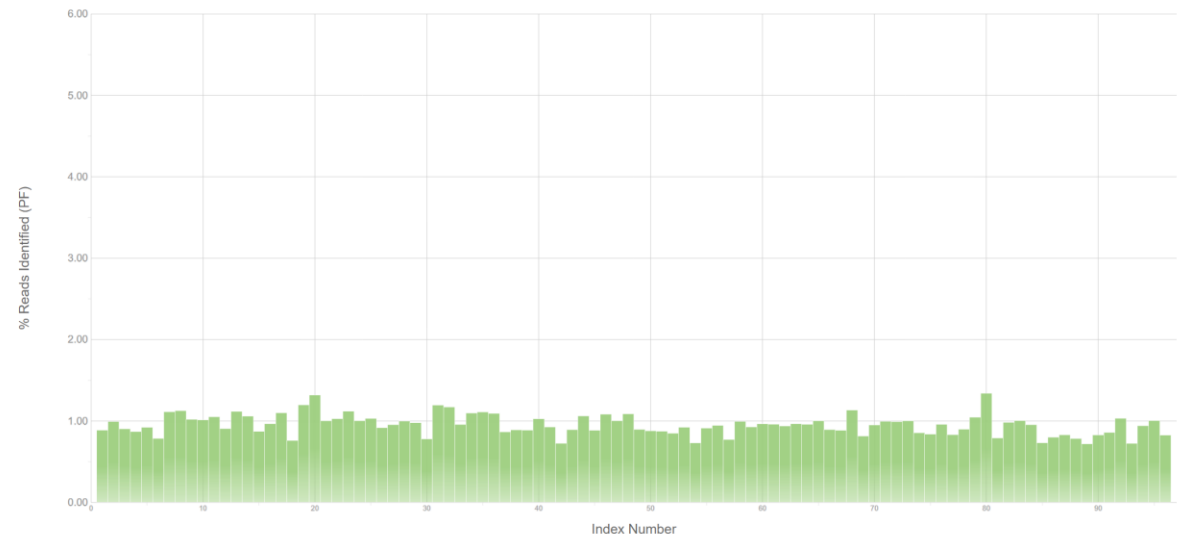
Qubit Normalization

- Same library prep

- Similar starting concentration

- Determine concentration by Qubit

- Pool 25 ng of each library together

Sample	Concentration (ng/ul)	Desired Amount (ng)	Volume (ul)
A1	5.8	25	4.31
A2	11.4	25	2.19
A3	2.6	25	9.62
A4	20.1	25	1.24
A5	14.5	25	1.72

Desired amount (ng)/Concentration (ng/μl)= Volume to Pool

Bioanalyzer
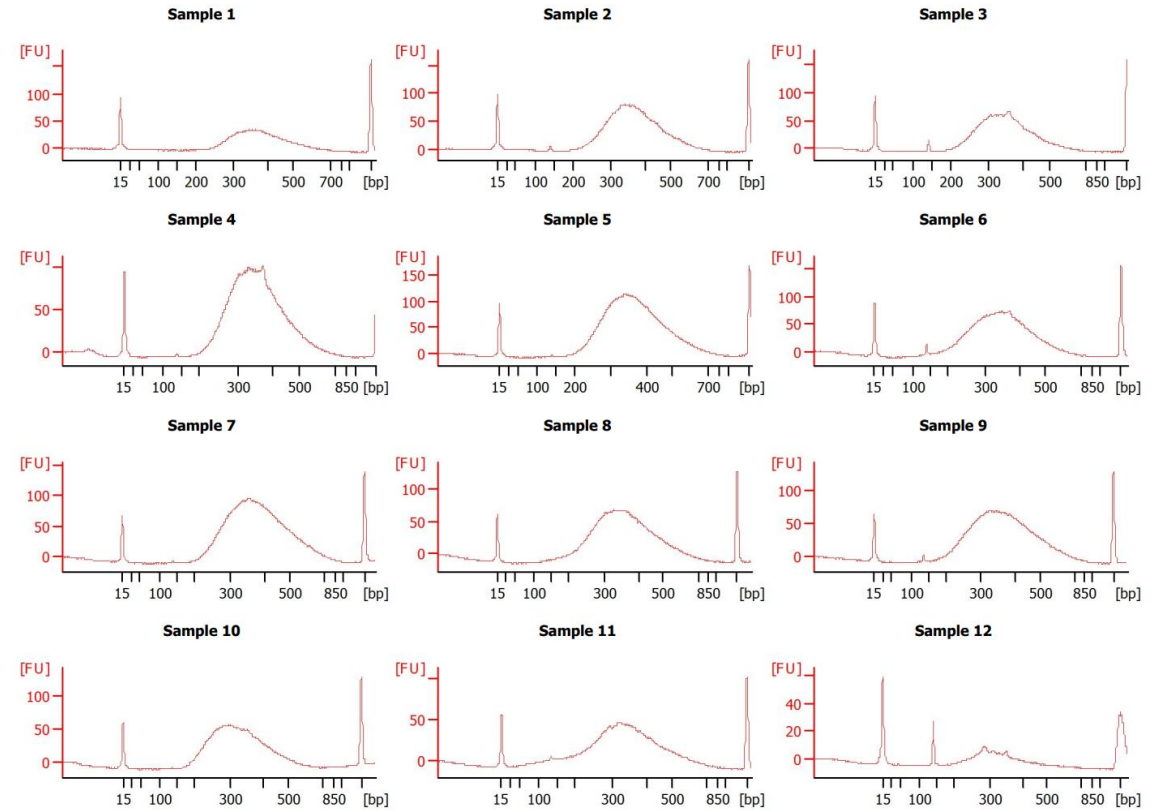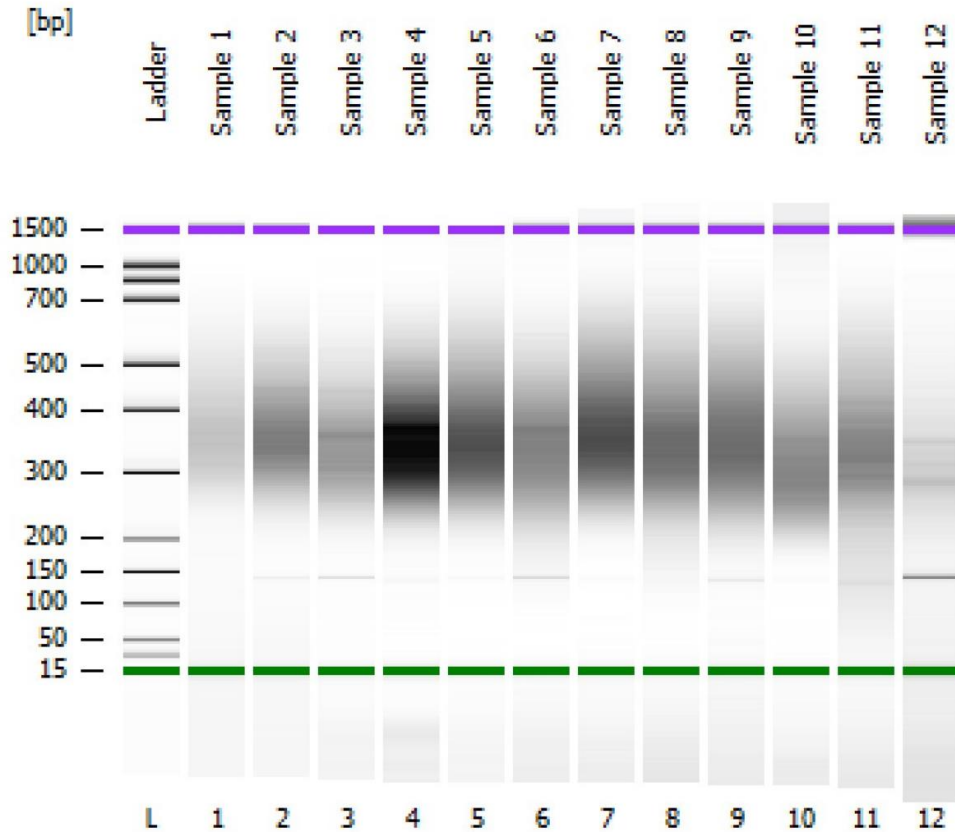
qPCR Normalization

▶ Six pre-diluted DNA Standards and appropriately diluted NGS libraries are amplified using platform-specific qPCR primers that target adapter sequences.

▶ The average Cq value for each DNA Standard is plotted against its known concentration to generate a standard curve.

▶ The standard curve is used to convert the average Cq values for diluted libraries to concentration, from which the working concentration of each library is calculated.
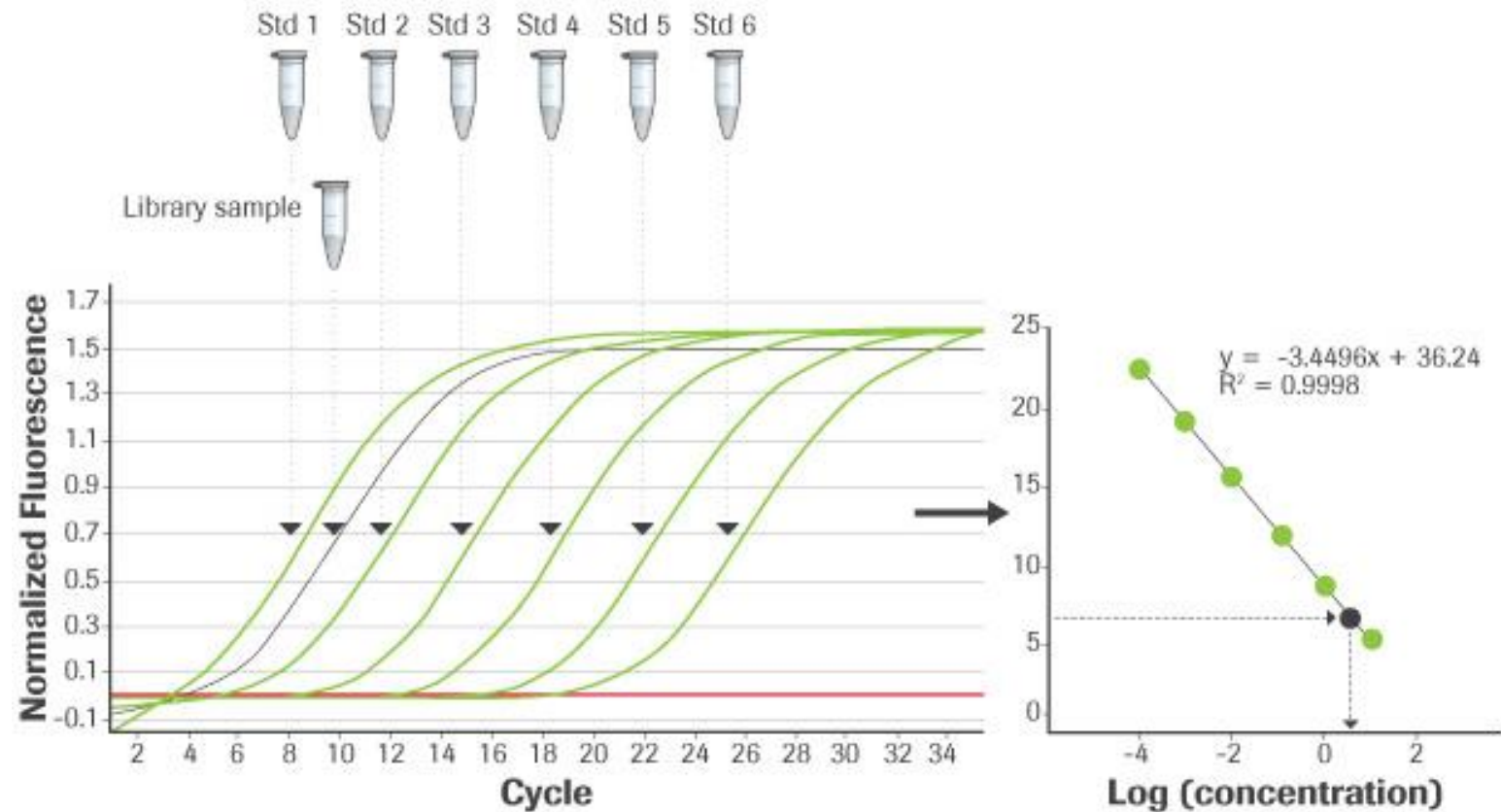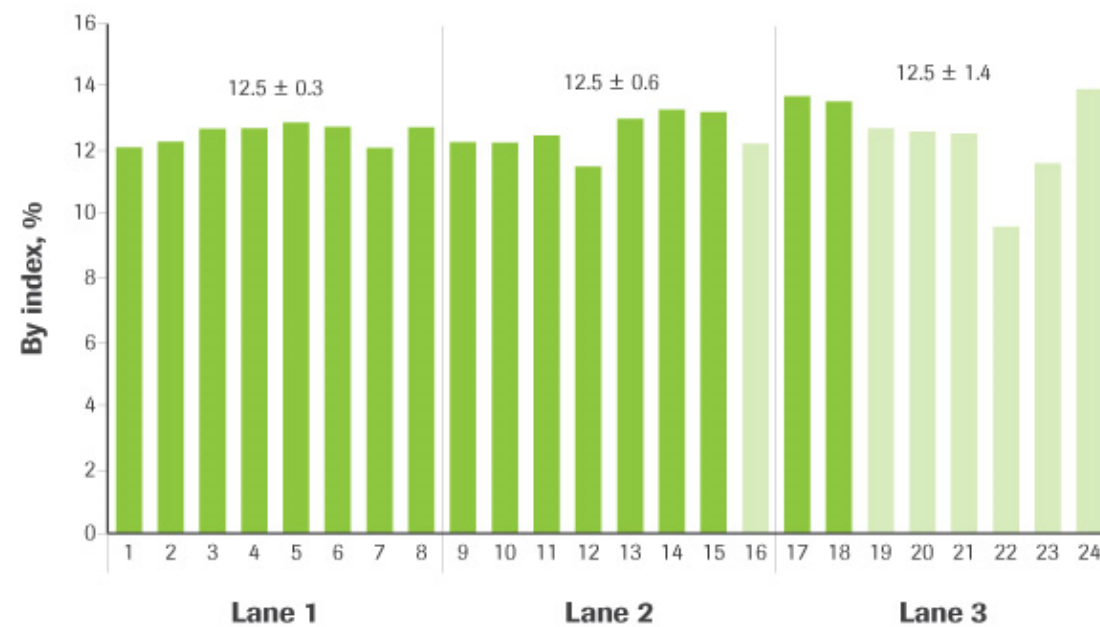
qPCR Normalization

Final library concentration
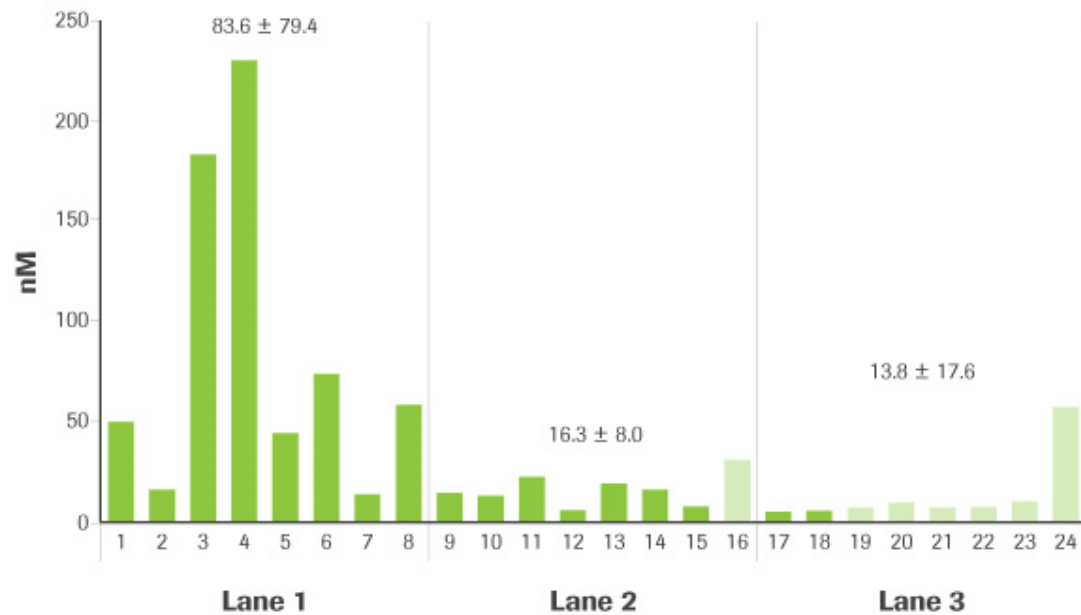
Total assigned reads in lane

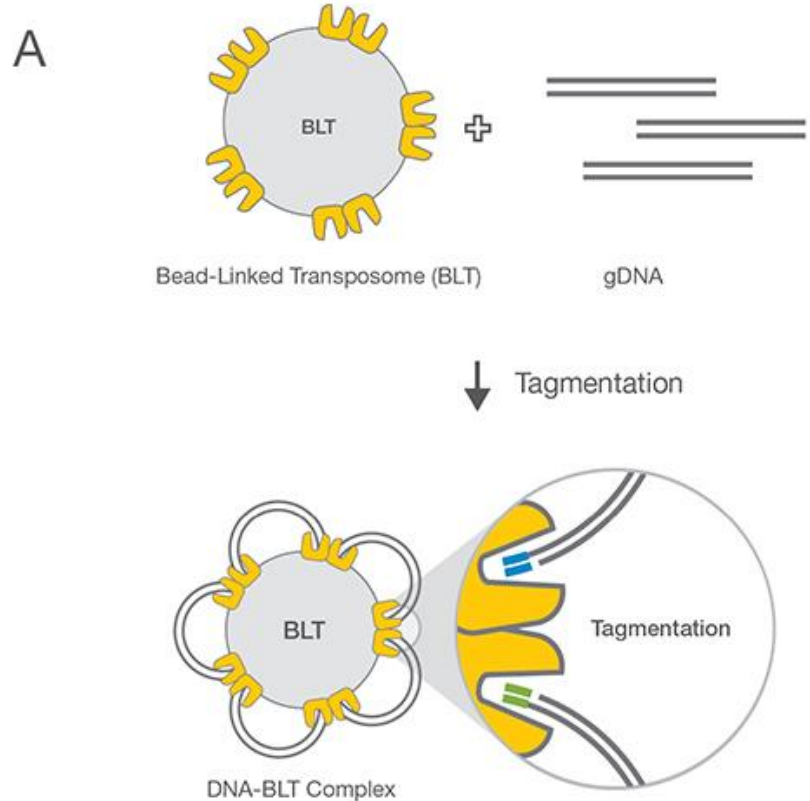# Bead Normalization

- DNA binding beads are mixed into a library with different concentrations of DNA.

- Equal amount of beads/sample.

- Beads are saturated with maximum DNA binding capacity.

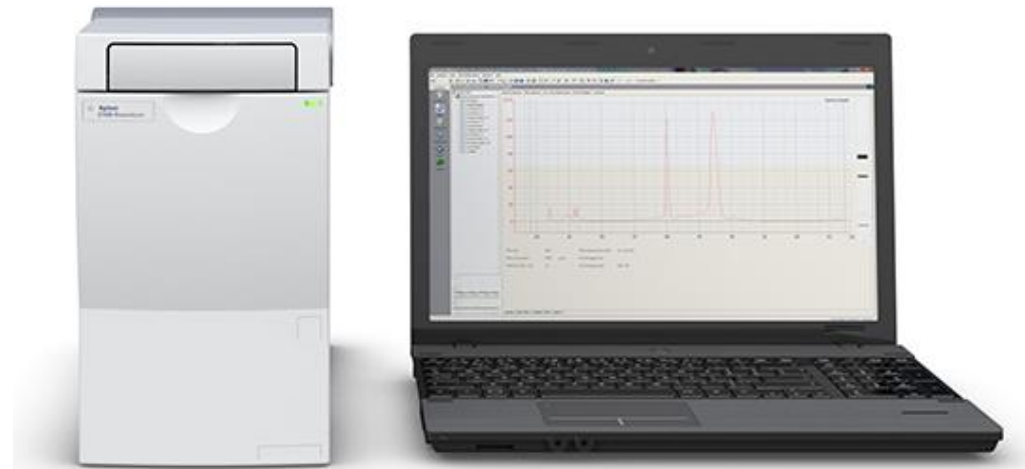- Beads captured with magnet & non-bound DNA library moved in supernatant.

https://upcvmda-pl480.weebly.com/educational-articles/
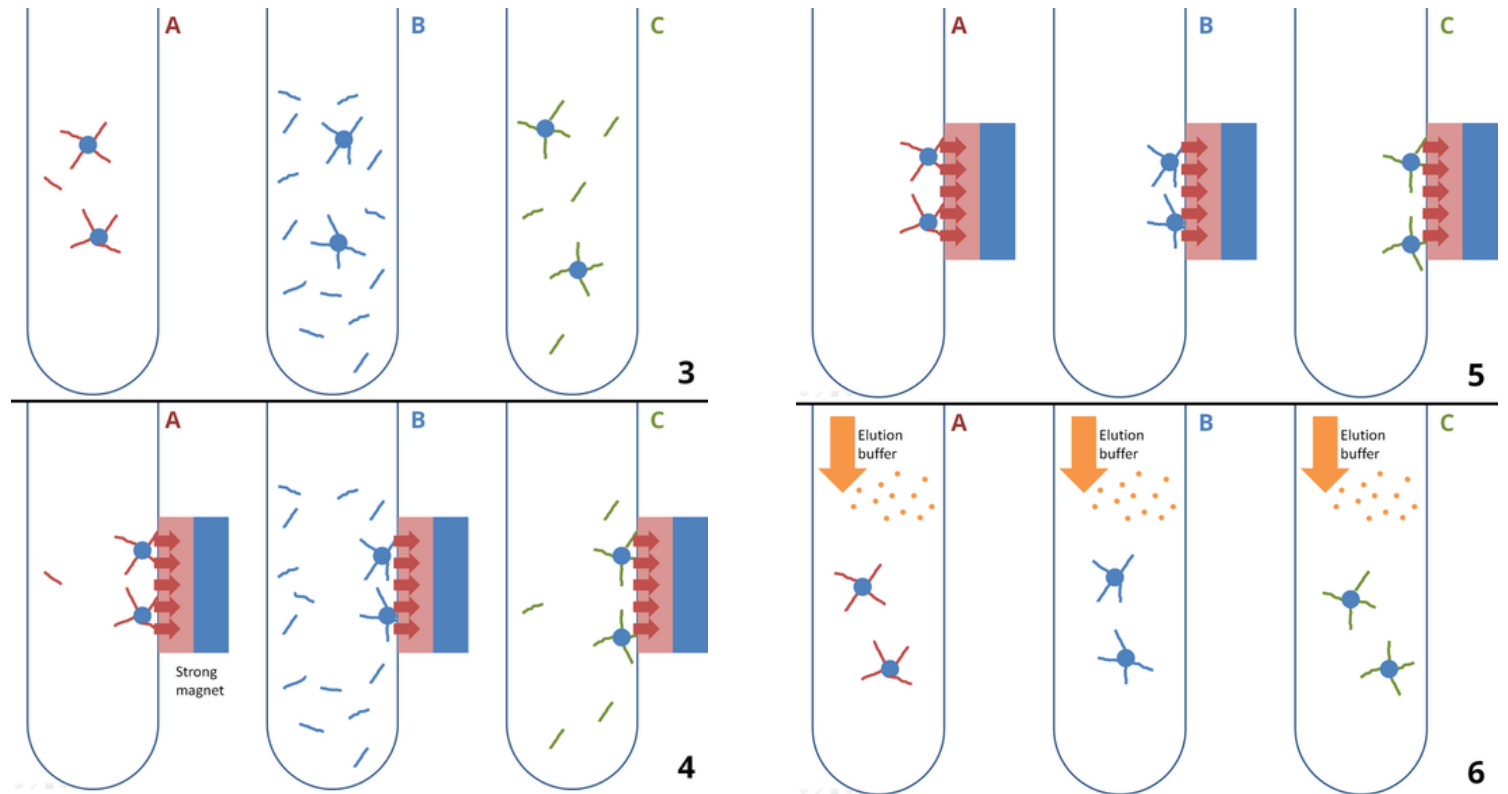
Bead-Linked Transposome Normalization
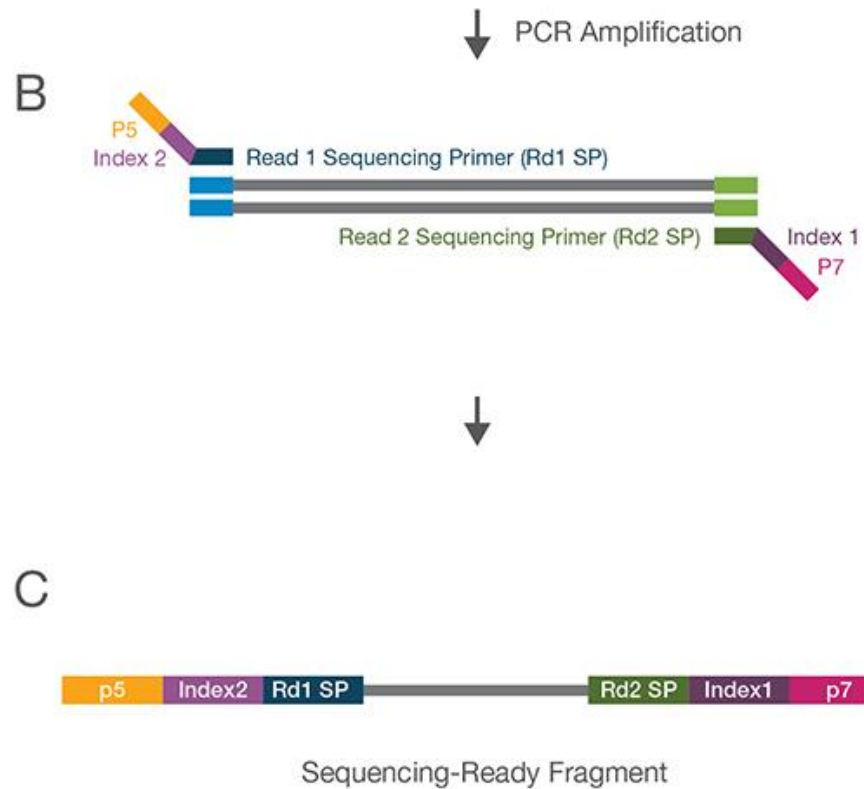
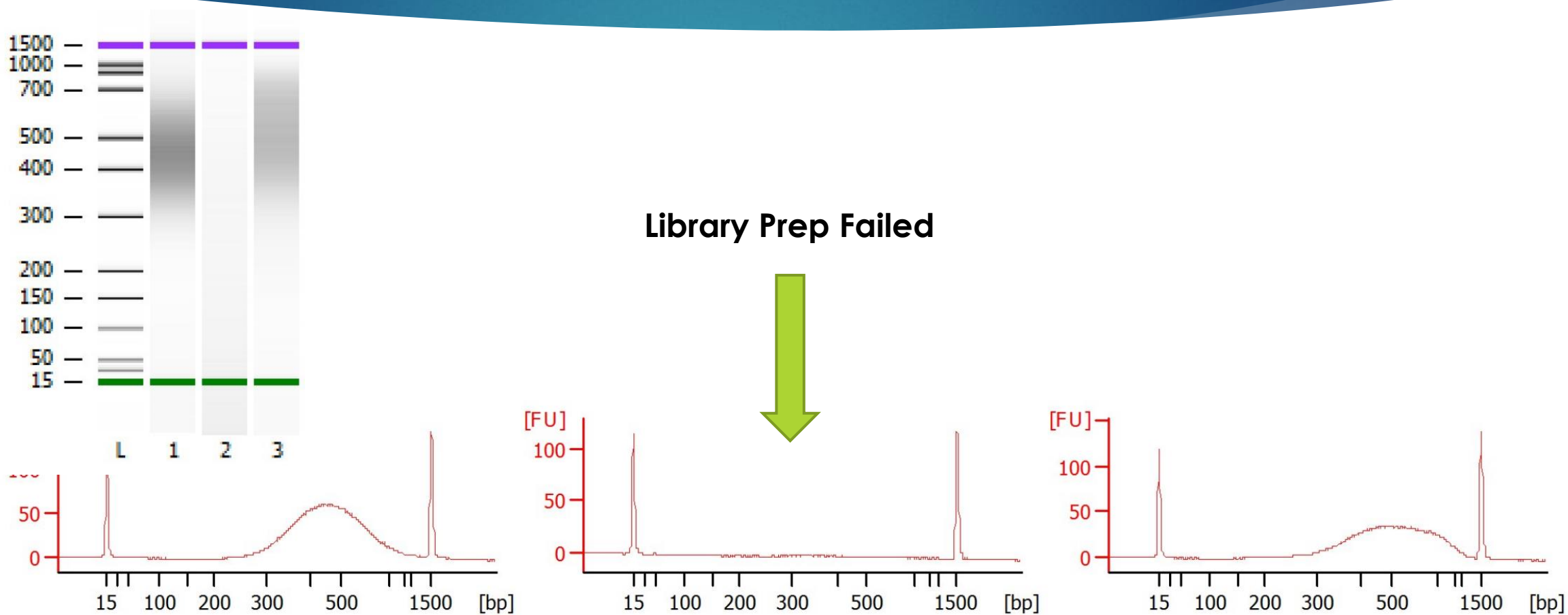Occurs during library prep not afterwards
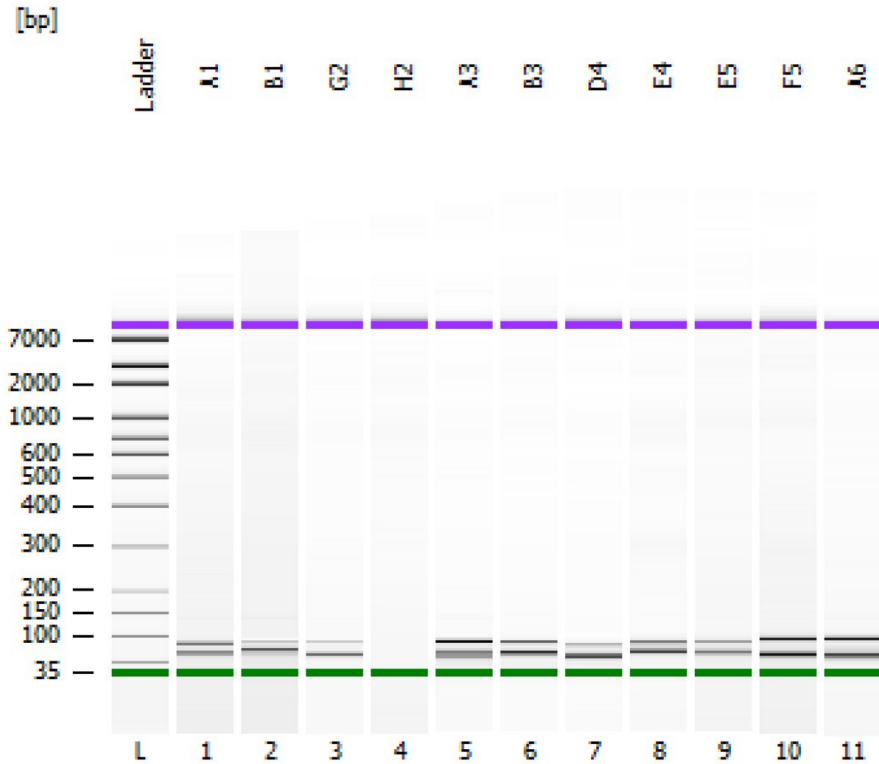
Library Quality Control

Library Quality Control

► **Ensure successful generation of library**

► **Check for Adapter/PCR Primer Dimers**

► **Determine library size & accurate molarity**
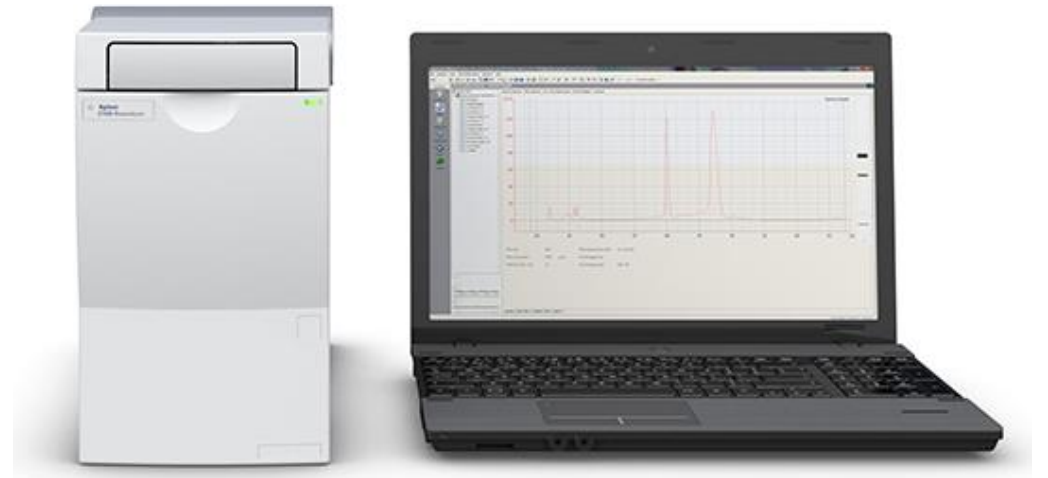
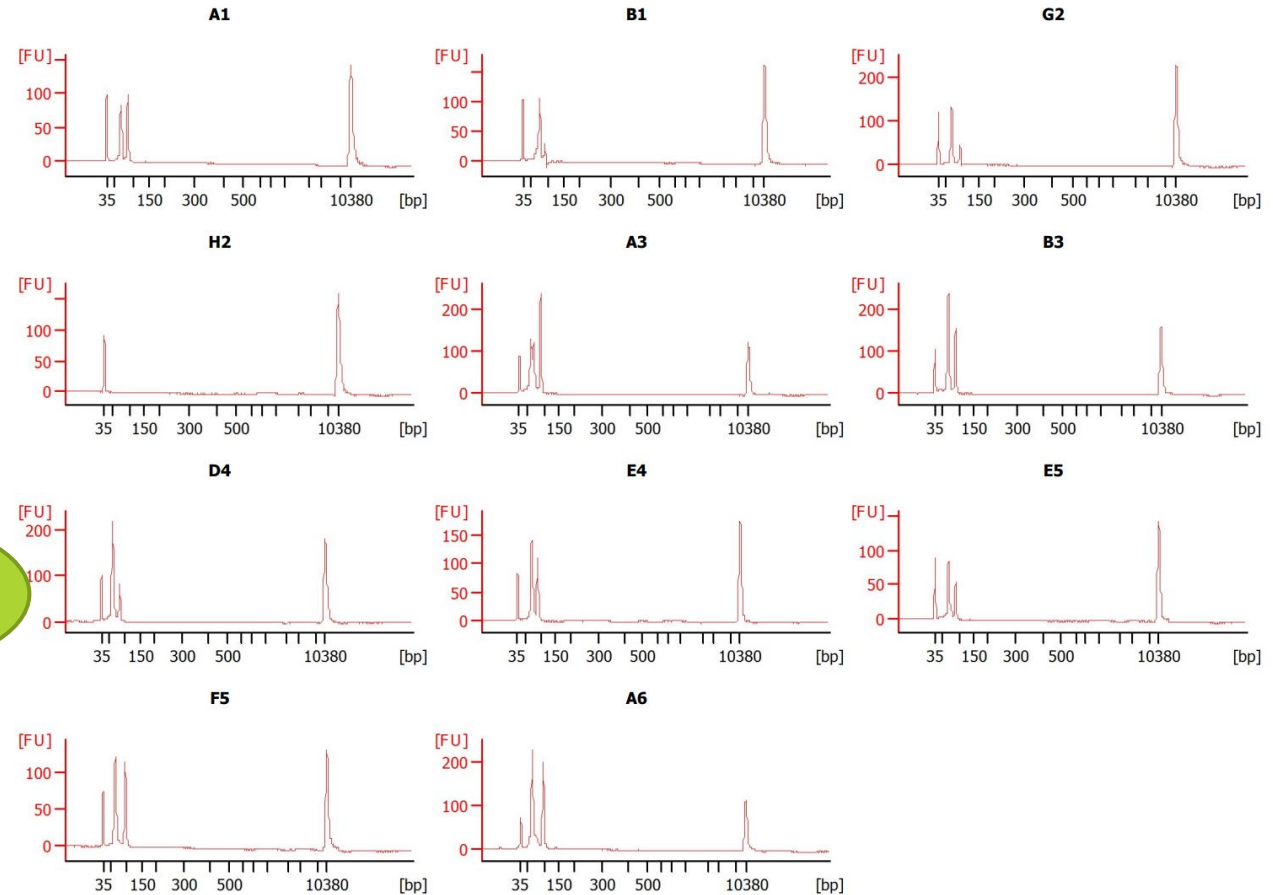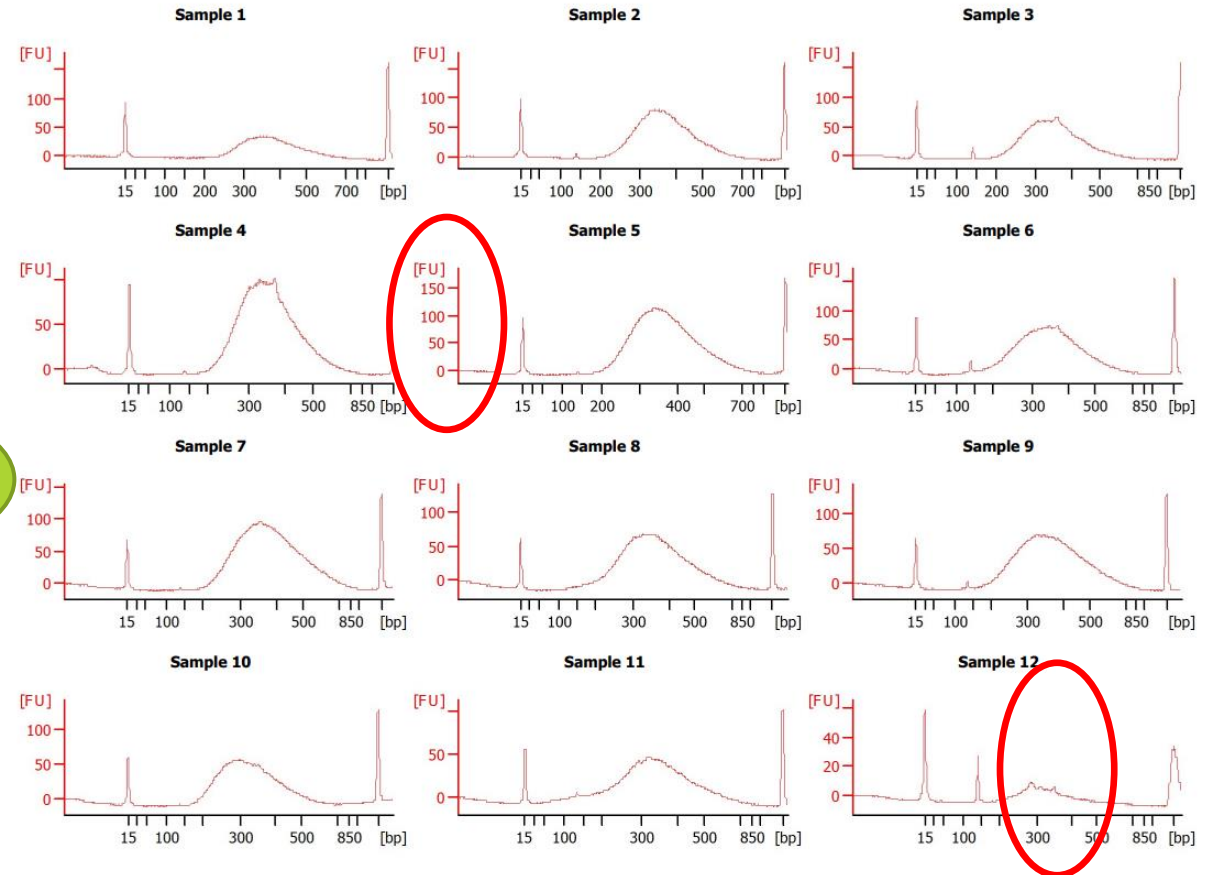# Verify Completed Library
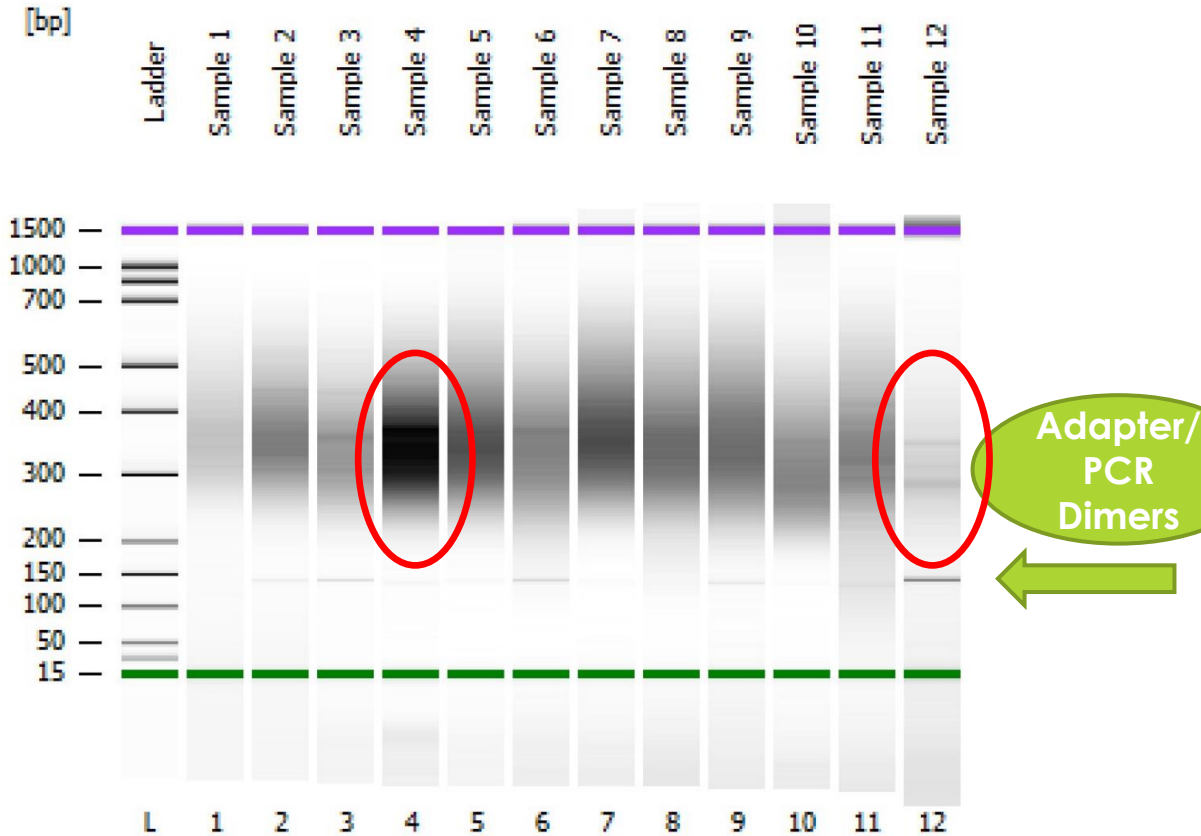

Library Prep Failed

Failed Libraries with Adapter/PCR Dimers

Completed Libraries

Summary

- Sample type can impact the choice of sequencing library prep and sequencing method.

- DNA/RNA isolation methods may impact sequence quality, however, most methods are sufficient for NGS applications.

- Purification & size selection of sequencing libraries improves sequencing quality.

- Library normalization ensures that each sample in a pool of sequencing libraries receives a similar amount of sequencing.

- For optimal sequencing results and troubleshooting library prep protocols, libraries should be checked for quality prior to sequencing.