



QUALITY CONTROL IN SARS-CoV-2 NGS workflow

Author name: Shebbar Osiany

Position: Research Associate

Program and/or organization name: Animal and Human Health

Viral Pathogen Genome Sequencing and Bioinformatics Analysis.

Nairobi, Kenya



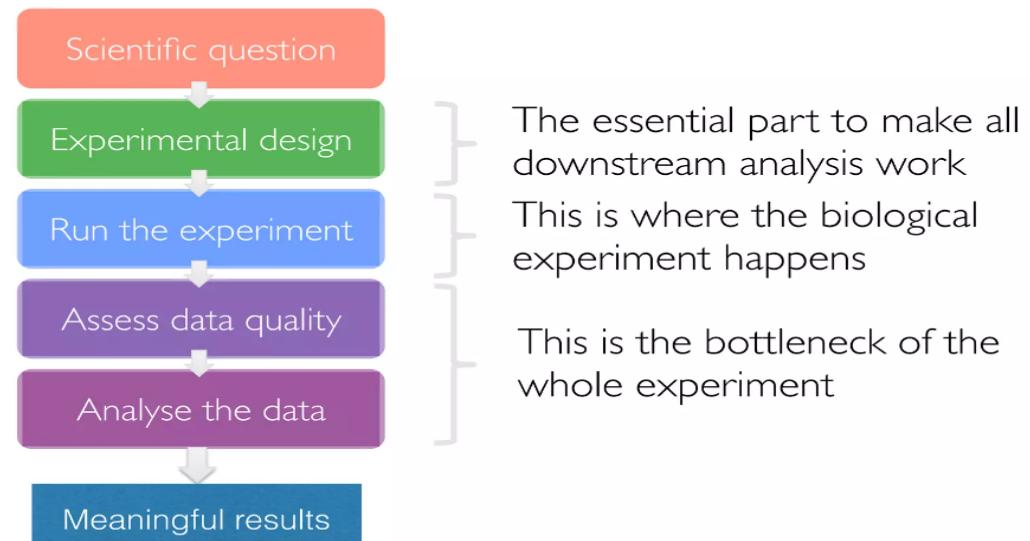
AfricaCDC
Centres for Disease Control
and Prevention





Better lives through livestock

Typical Genomics Workflow



Outline:

- What is Quality Control (QC)?
- Why Quality Control?
- QC methods used in library preparation workflow



AfricaCDC
Centres for Disease Control
and Prevention



Quality Control.

Quality control is an essential step in any NGS workflow, allowing the integrity and quality of the samples and data to be checked before downstream analysis and interpretation.

Why Quality Control?

- Human errors
- Operator-to-operator variations
- Tools imprecisions/misuse
- Environmental changes
- Sample to sample variations

**Risks of degrading sample quality
and generating results of poor quality**



- Sample Quality?
- Purity?
 - Quantity?
 - Integrity?
 - Sequence?

**Are these parameters in the range of my
experiment's requirements for
generating high-quality results?**

QC IN NGS WORKFLOW

Nucleic acid extraction



- Samples collection
- Nucleic acid extraction
- Quality Control

Library preparation



- Adaptor ligation/barcoding
- Size selection
- Amplification/purification
- Quality control

Sequencing and analysis

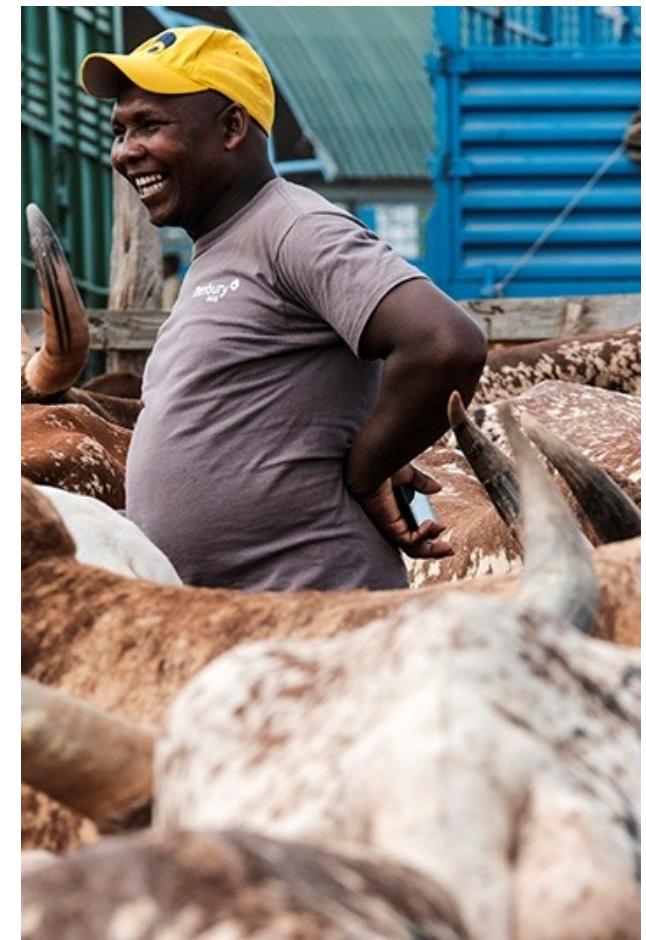


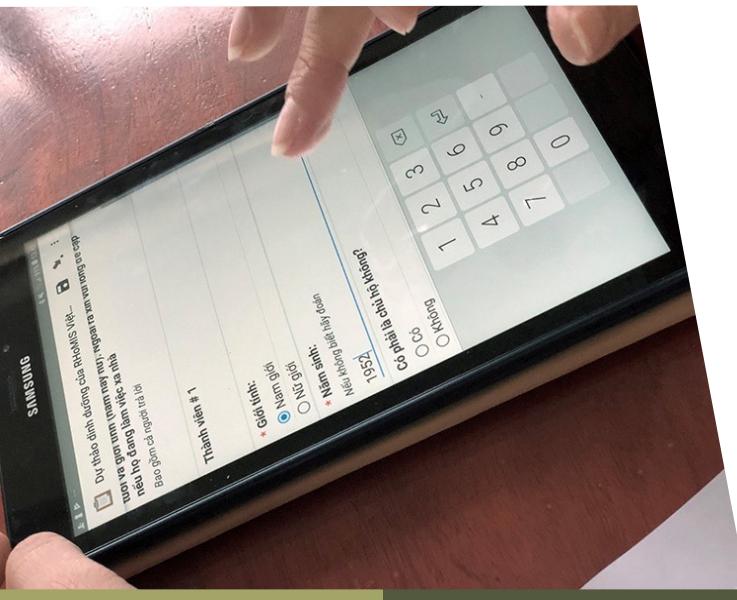
- Sequencing
- Data analysis
 - Base calling
 - Read alignment
 - Variant calling
 - Variant annotation



QC at sample reception

- Correct sample received
- Proper storage and transport conditions
- Correct metadata of samples received





Controls in RNA extraction and RT-PCR

NTC (Non template control); Nuclease free water
NC (Negative control); Known negative sample
PC (Positive control); Known positive sample



Click and delete to replace
this photo with another one

QC AT RNA EXTRACTION

3 controls included in the extraction process;

- I. No template control(NTC) normally NFW used which should not amplify any genes to ascertain that extraction process was smooth. Amplification of a target in NTC is an indicator of contamination in the extraction process
 - II. A known negative sample that should only amplify the internal control
 - III. A known positive sample that should amplify all target gene for a positive sample
- NB; These controls are given same treatment as normal sample



QC AT RT-PCR

Kit used has an NC and PC provided that is included in the run

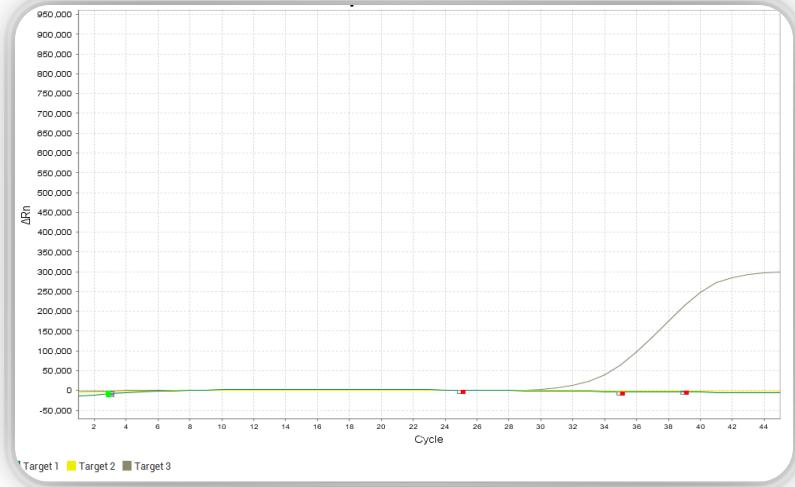
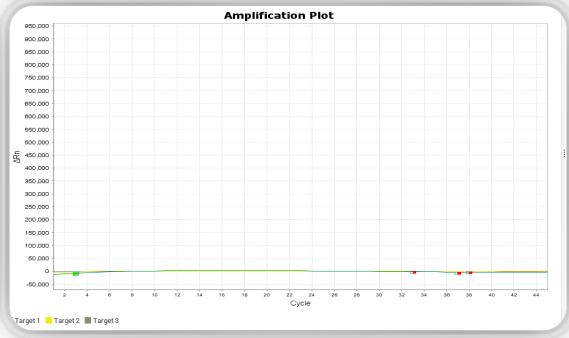
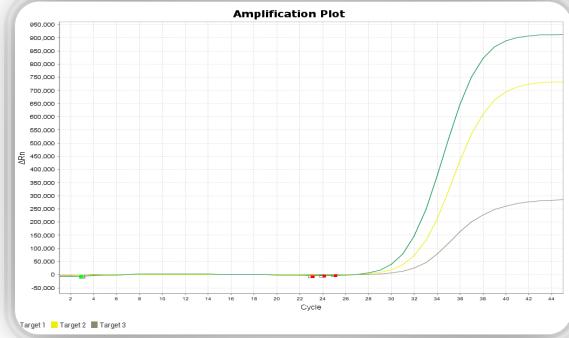
According to kit specifications the NC amplifies only the internal control and the 2 targets cts remain undetermined

The PC amplifies both target genes and cts have a value

If controls pass QC both for extraction and RT-PCR samples are analyzed based on CT values to determine which samples proceed for sequencing

Only samples with a CT value of 30 and below proceed to cDNA and tiling

CONTROL AMPLIFICATION PLOTS AND cts



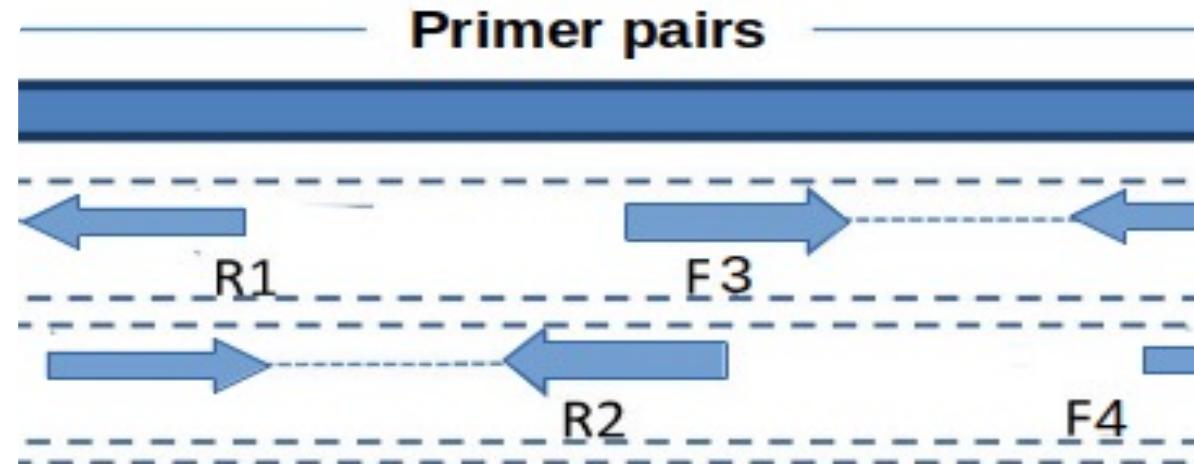
40 D4	FALSE	NC	Target 1	NTC	FAM	None	Undetermined
40 D4	FALSE	NC	Target 2	NTC	VIC	None	Undetermined
40 D4	FALSE	NC	Target 3	NTC	CY5	None	33.213
41 D5	FALSE	PC	Target 1	STANDARD	FAM	None	29.475
41 D5	FALSE	PC	Target 2	STANDARD	VIC	None	29.092
41 D5	FALSE	PC	Target 3	STANDARD	CY5	None	32.023
42 D6	FALSE	NTC	Target 1	NTC	FAM	None	Undetermined
42 D6	FALSE	NTC	Target 2	NTC	VIC	None	Undetermined
42 D6	FALSE	NTC	Target 3	NTC	CY5	None	Undetermined



QC AT GENOME AMPLIFICATION (TILING)

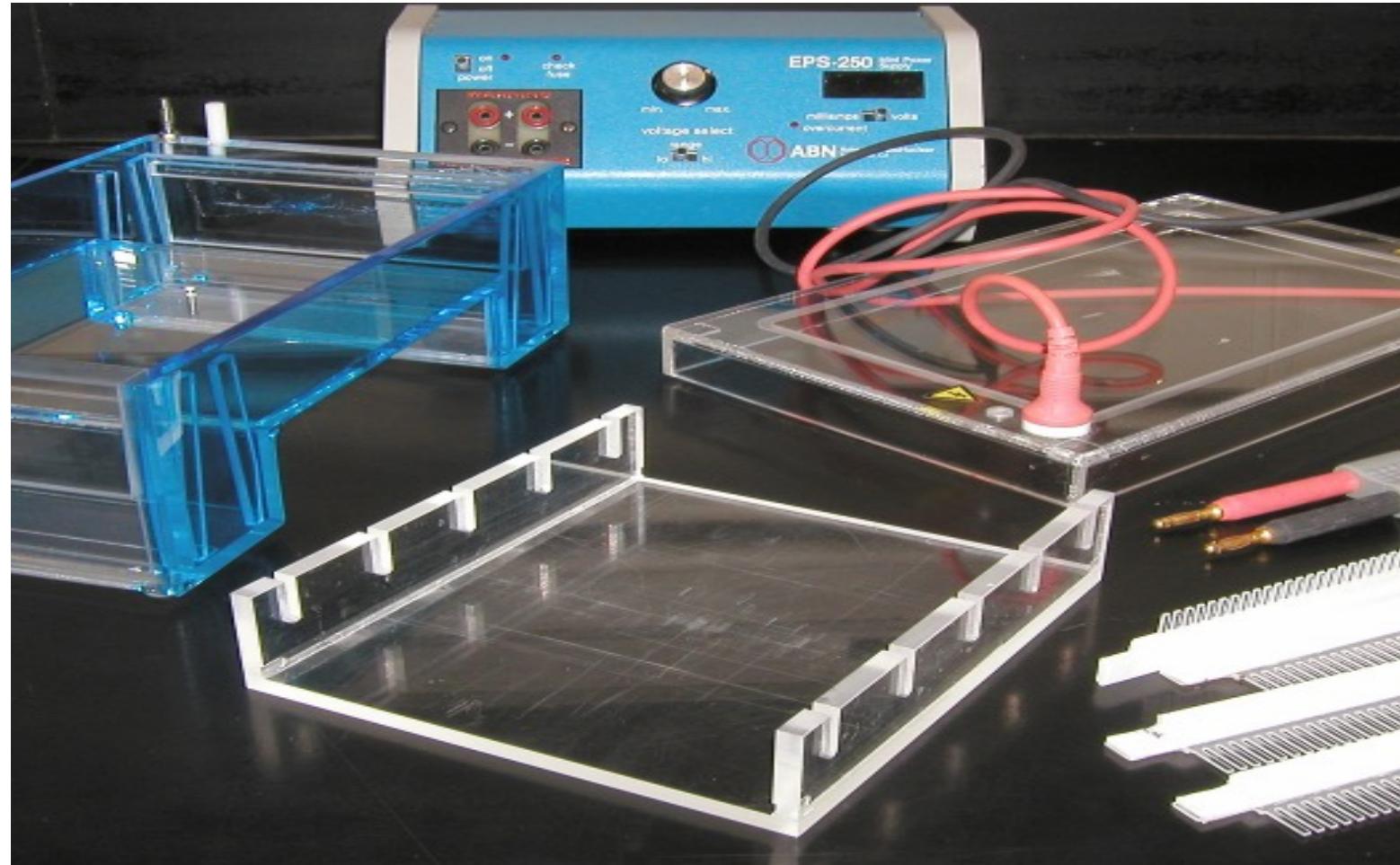
Gel electrophoresis

Is a qualitative method





Gel electrophoresis set up





Genome amplification QC: GEL electrophoresis

Why gel?

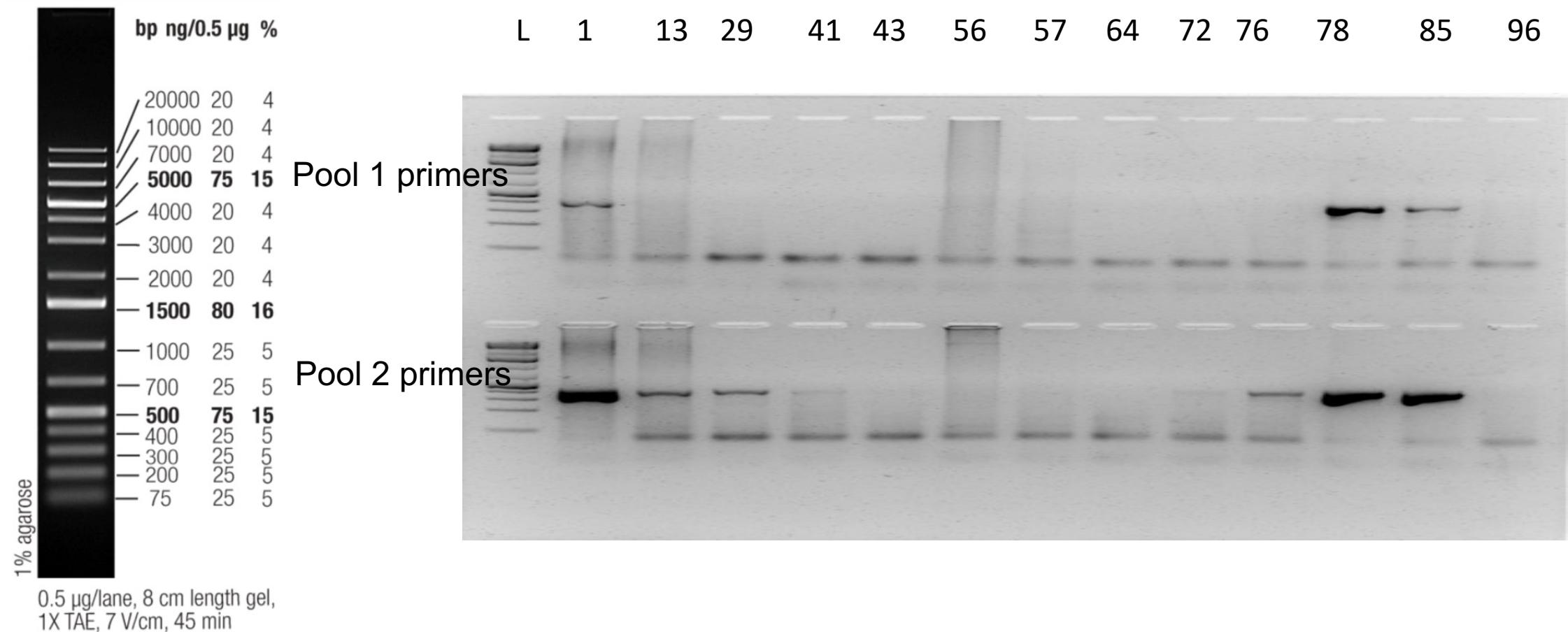
Check for success of pool amplification

Check for sample purity

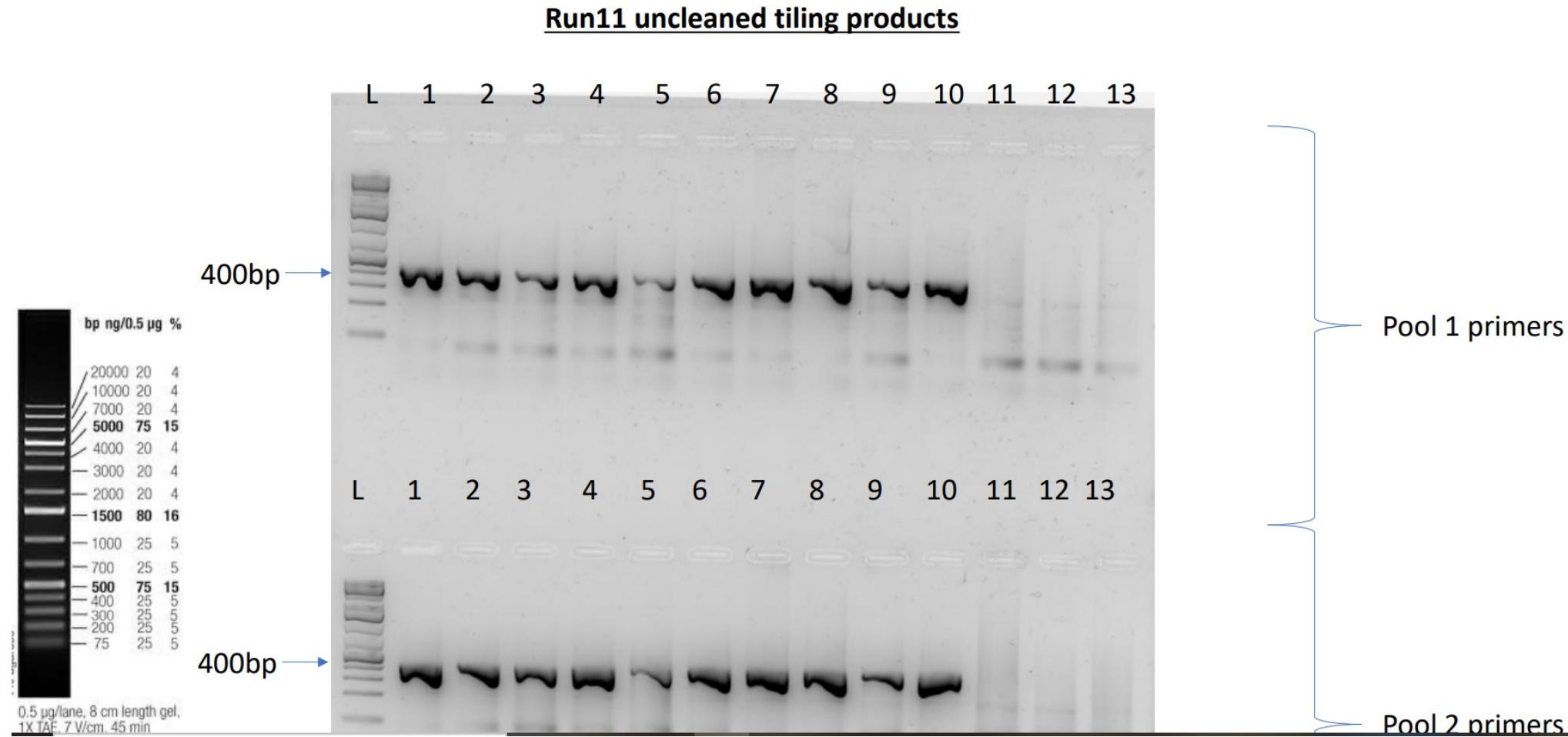
Establish expected fragment size

**NB:- For high throughput (96-384 samples) only,
representative samples used for QC.**

Gel to determine tiling amplification.

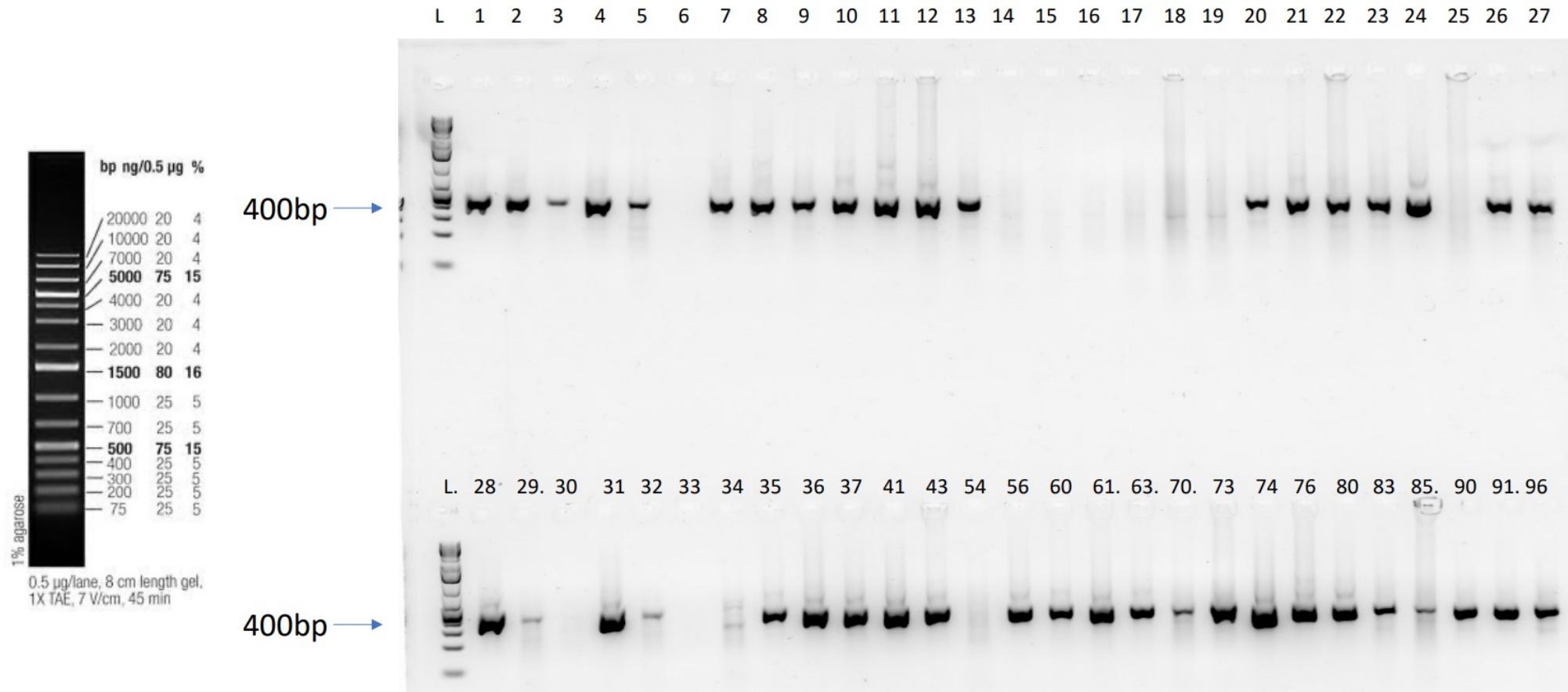


Gels to check for EXPECTED FRAGMENT SIZE AND purity

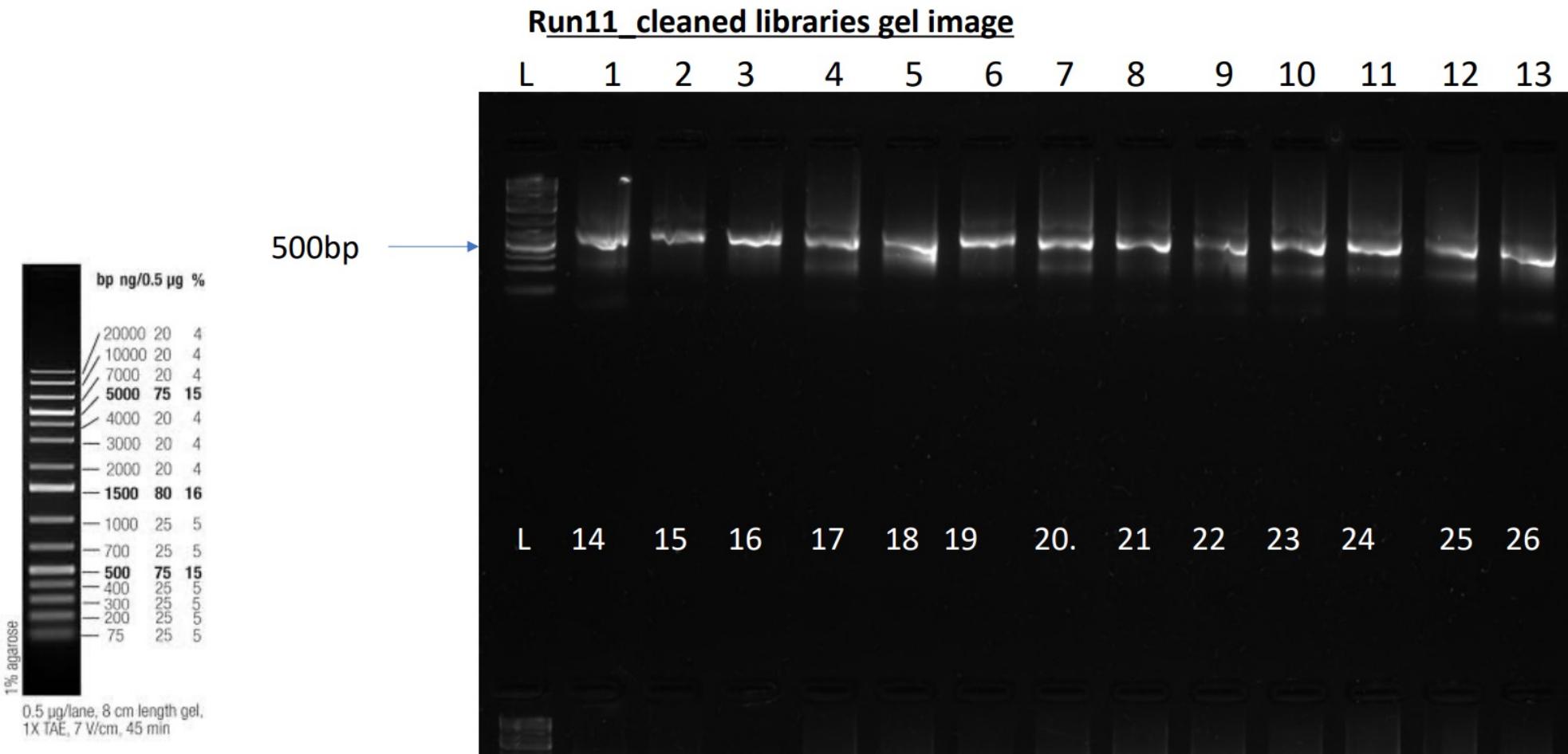


CLEANED TILED PRODUCTS

Run11 cleaned tiling products



EXPECTED SIZE AFTER ADAPTOR LIGATION



QC IN LIBRARY PREPARATION

Quantitation

There are 3 main methods of quantifying NGS library

- I. qPCR
- II. Fluorometry and spectrophotometry
- III. Electrophoresis

qPCR to quantify final library products before sequencing

- Main purpose is to accurately quantify each library before pooling in order to obtain optimal cluster density when sequencing
- It amplifies complete library DNA fragments i.e; fragments with adaptors ligated to ends
- It is a relative quantification based on known standards

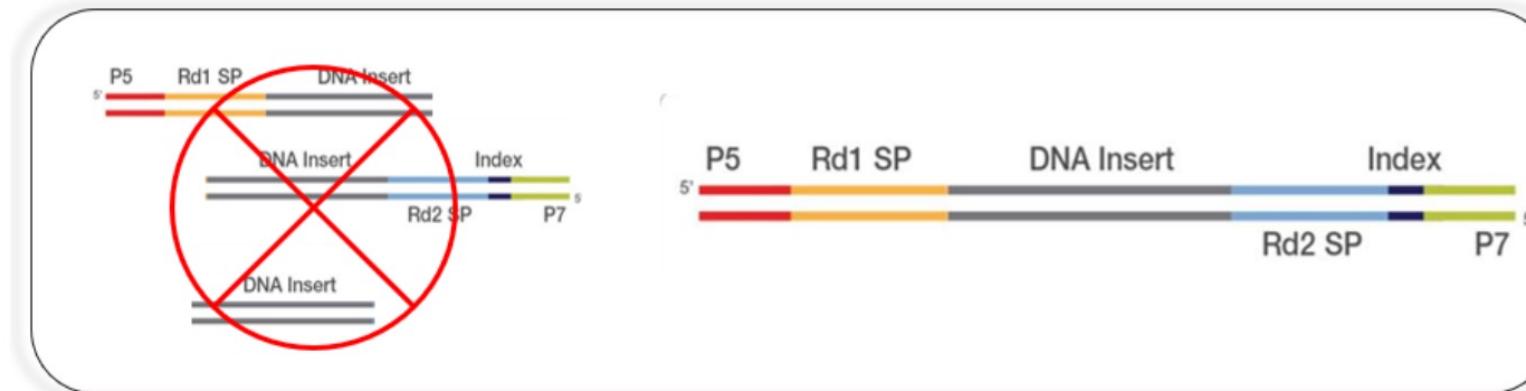


Figure 1: Only full-length libraries (on the right) are quantified by qPCR. Incomplete libraries (on the left) are not quantified.

Fluorescence

Qubit

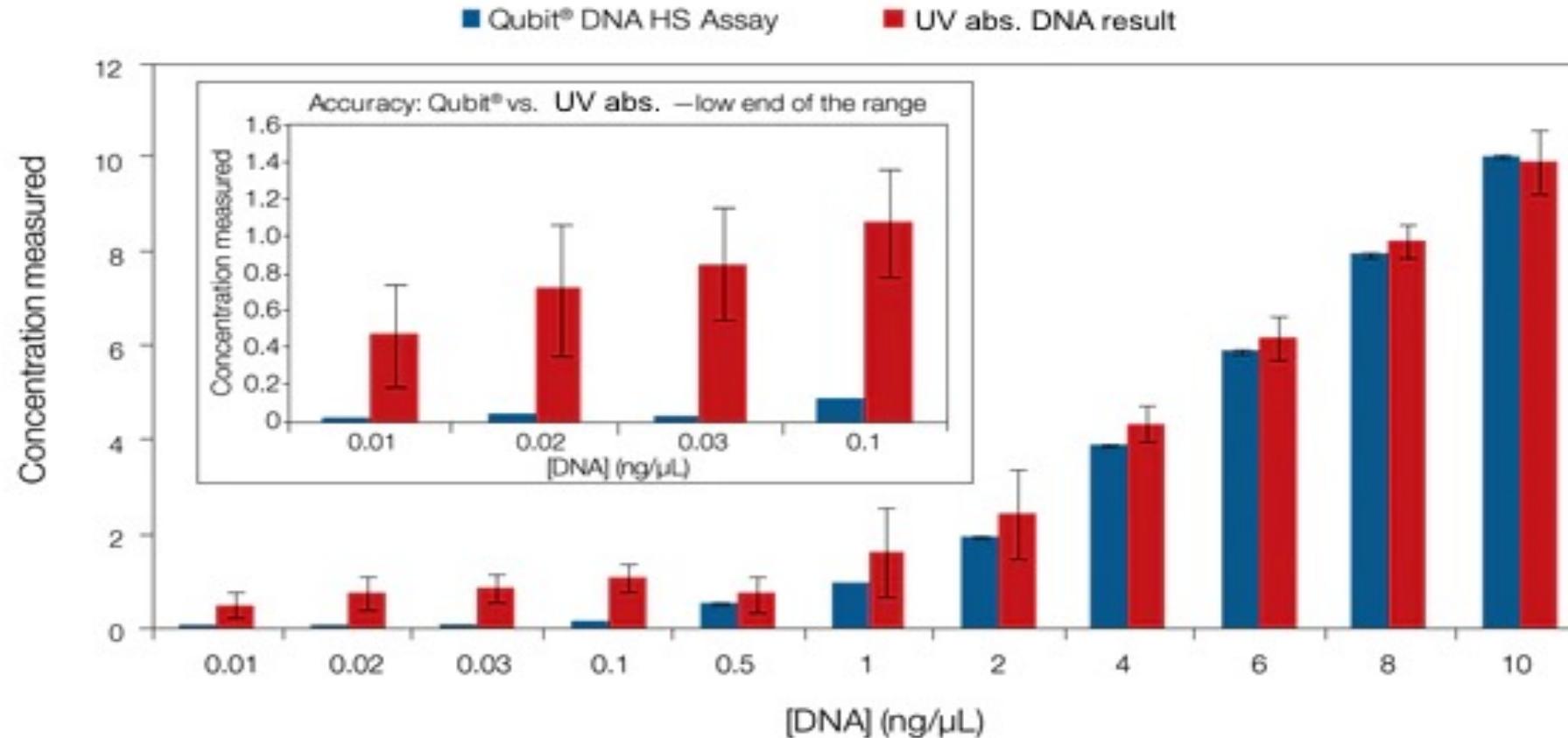
- This method involves using an intercalating fluorescent dye that specifically binds with high affinity to DNA or RNA.
- It is very accurate because only target-bound dye emits fluorescent signal
- Concentration of unknown sample is calculated based on comparison to a standard curve generated from samples of known DNA concentration

Spectrophotometry/U.V Absorbance

Nanodrop

- Measures anything absorbing at 260nm-DNA, RNA, protein, free nucleotides or excess salts
- Absorbance ratio of N.A to contaminants provide estimation of the sample purity
- Absorbance at 280nm and 230nm can be used to access level of contaminating proteins or chemicals respectively

Comparison of qubit and Nanodrop.



Continuation....

Qubit platforms give sensitive, accurate and precise measurements compared to UV absorbance instruments.

Electrophoresis.

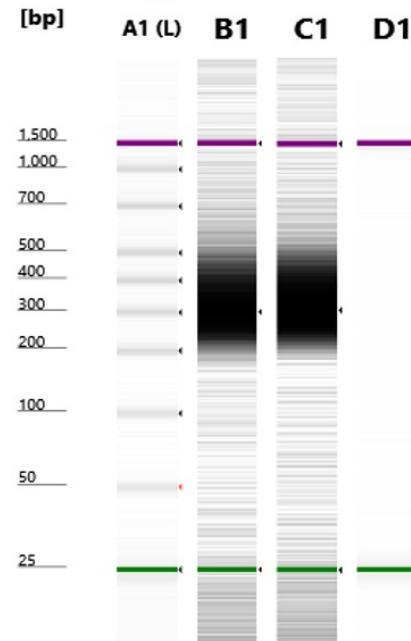
Agilent Tape station

- Integrates an instrument, data processing software, reagents and screen tape devices specific for DNA or RNA
- It analyzes size, quantity and integrity of samples
- Ladder is included in each run containing 2 internal standards (lower and upper markers) to align the ladder data with samples to determine sizing
- Concentration of upper marker is known and is used to determine sample concentration
- DNA Integrity number assess DNA degeneration within a scale of 0-10 with 1 indicating severe degradation

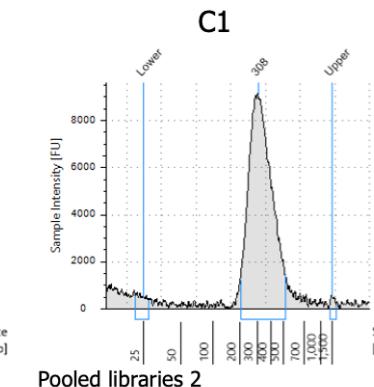
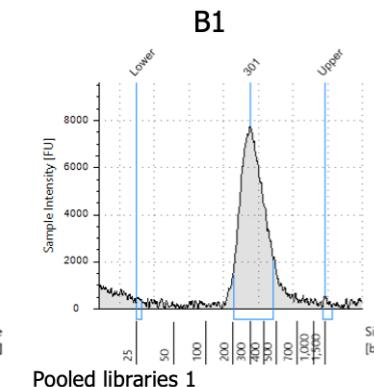
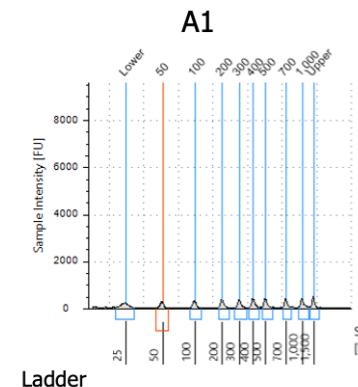
Data output of tape station.

High Sensitivity D1000 ScreenTape®

Gel Image



High Sensitivity D1000 ScreenTape®



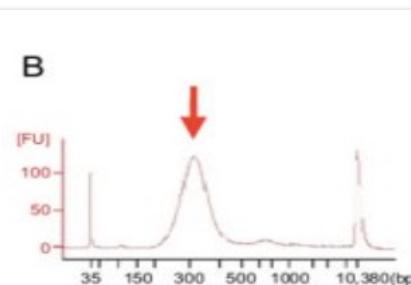
HS D1000

Contrast: 0.50

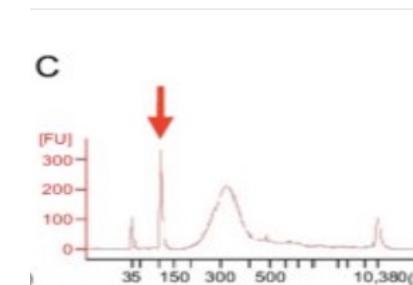
Agilent Bioanalyzer



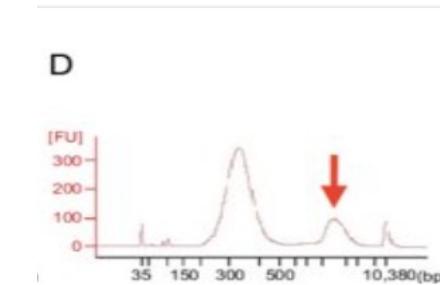
Data represented as gel image.



Bioanalyzer trace showing DNA fragments present in library



Sample containing adaptor dimers; need for further cleanup

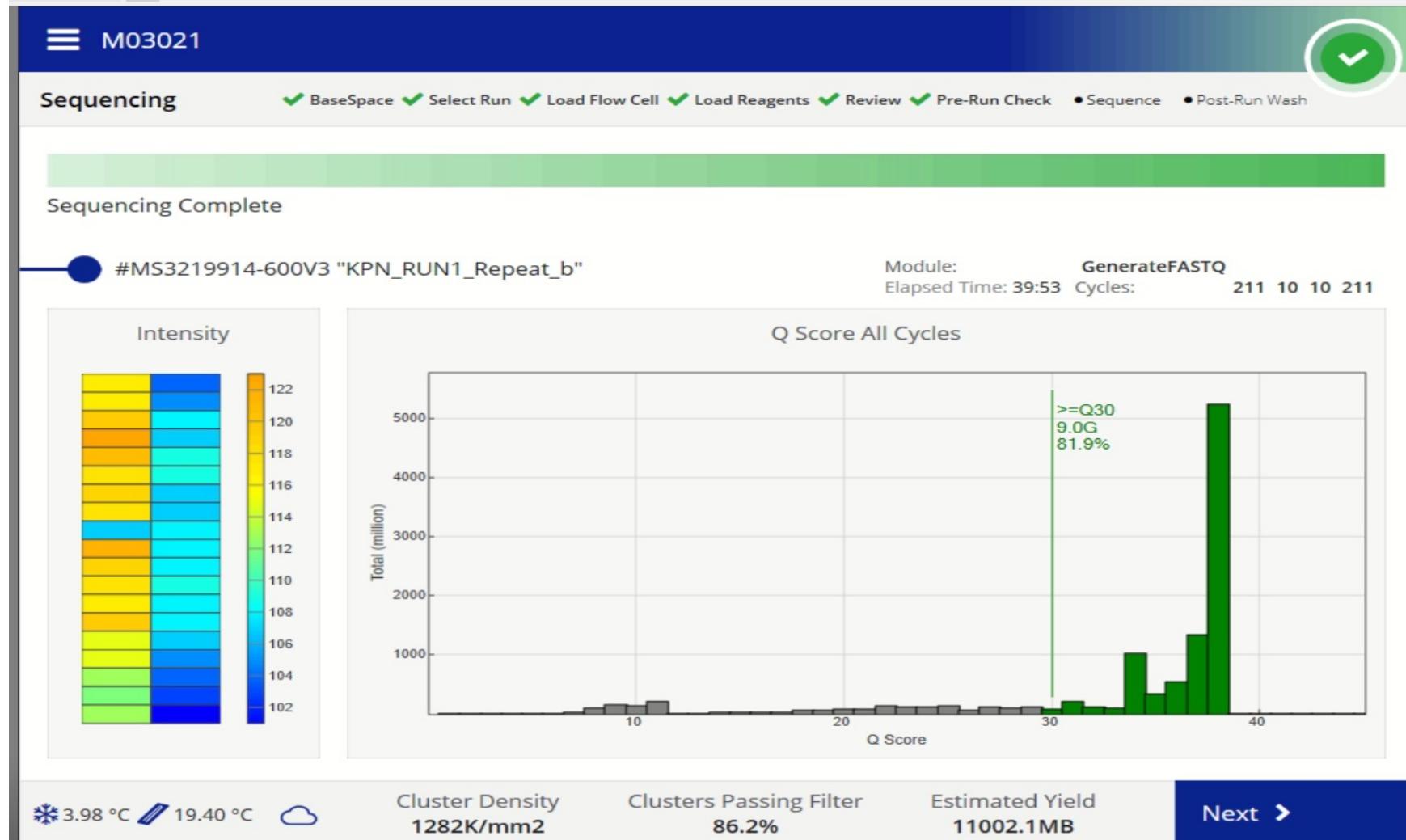


Library with extra high molecular band; indication of over amplification of sample.

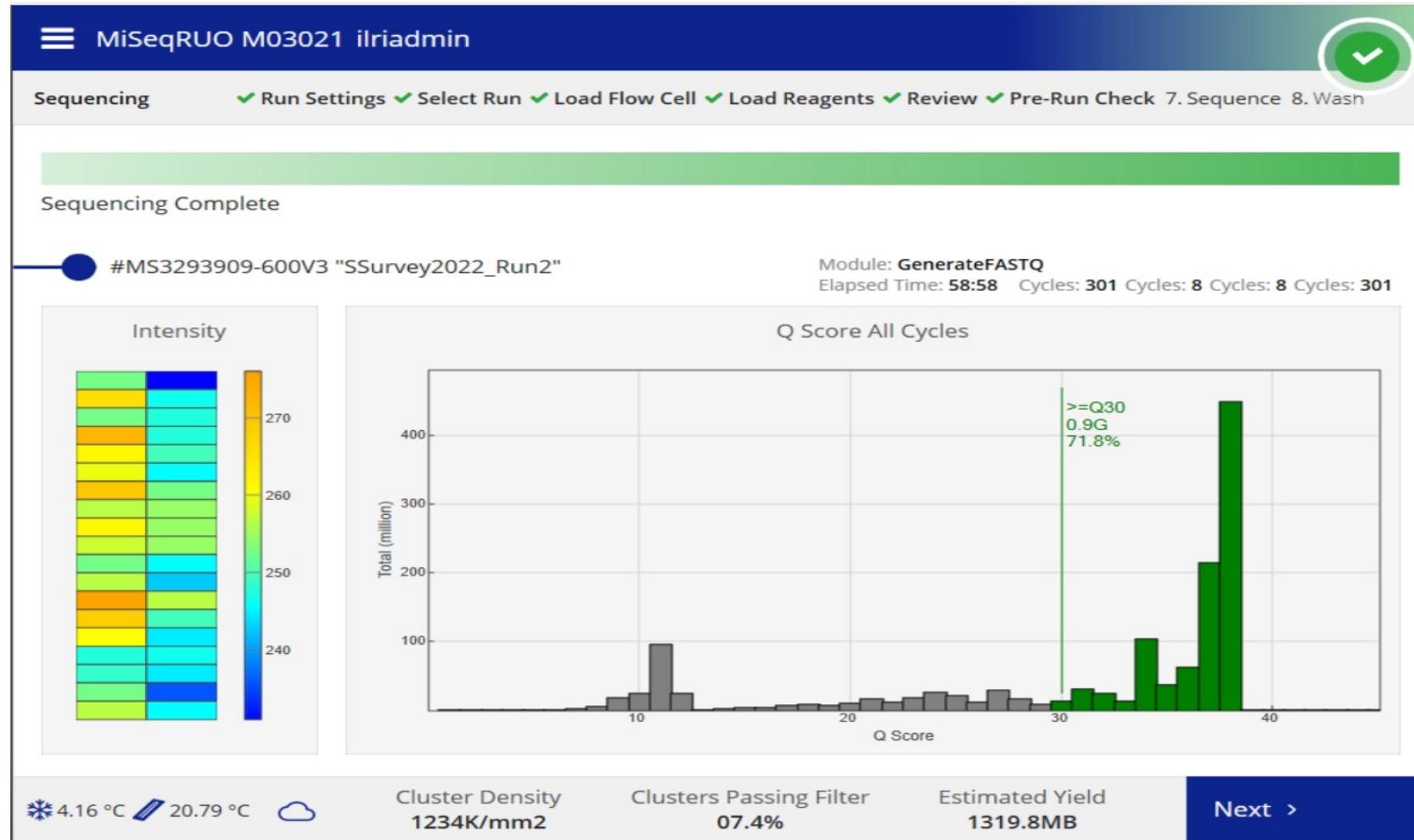
QC Metrics of Raw Reads.

- 1.Yield:** total number of reads per run.
- 2.Read analysis:** Read length, GC content, adapter content, duplication to give indication of data quality.
- 3.Q-score:**determines probability that an incorrect base was called during a run. A score of above 30 is considered a good run.
- 4.Cluster passing filter:** refers to number of clusters that passed the chastity filter of the machine. A lower PF is associated with low yield.
- 5.Cluster density:** refers to the density of clusters on flow cell. Is assessed in tandem with the PF as the 2 together diagnose problem with over or under loading your library.

Good Run...



Bad Run...



Average run...

M03021

Sequencing ✓ BaseSpace ✓ Select Run ✓ Load Flow Cell ✓ Load Reagents ✓ Review ✓ Pre-Run Check • Sequence • Post-Run Wash

Sequencing Complete

#MS3141827-600V3 "16s_WMBuru_Run8"

Module: GenerateFASTQ
Elapsed Time: 64:27 Cycles: 301 8 8 301

Intensity

Q Score All Cycles

Q Score Range	Total (million)
0-5	~100
5-10	~400
10-15	~200
15-20	~100
20-25	~100
25-30	~100
30-35	~150
35-40	~1200

4.11 °C 20.72 °C
Cluster Density 979K/mm²
Clusters Passing Filter 65.1%
Estimated Yield 10453.6MB
Next >

LIVESTOCK RESEARCH INSTITUTE
CGIAR

Summary

Good QC will:

- I. Improve efficiency of your workflow
- II. Save time; detect errors early, so you don't have to repeat the whole process
- III. Help you understand your results better
- IV. Ensure your data is of quality standard, dependable and reproducible



Thank you....

