Nov 25, 2022

Library pooling and quality control for Illumina sequencing

DOI
dx.doi.org/10.17504/protocols.io.q26g7yp6kgwz/v1

Katherine Smollett[1], Lily Tong[1], Jenna Nichols[1], Kirsty Kwok[1], Kyriaki Nomikou[1], Ma. Jowina Galarion[1], Daniel Mair[1], Ana Filipe[1]

[1]MRC-University of Glasgow Centre for Virus Research

CVR Genomics

Katherine Smollett
University of Glasgow

OPEN ACCESS

DOI: https://dx.doi.org/10.17504/protocols.io.q26g7yp6kgwz/v1

Protocol Citation: Katherine Smollett, Lily Tong, Jenna Nichols, Kirsty Kwok, Kyriaki Nomikou, Ma. Jowina Galarion, Daniel Mair, Ana Filipe 2022. Library pooling and quality control for Illumina sequencing. **protocols.io**
https://dx.doi.org/10.17504/protocols.io.q26g7yp6kgwz/v1

Protocol status: Working
We use this protocol and it's working

Created: November 18, 2022

Last Modified: November 25, 2022

Protocol Integer ID: 72931

Keywords: viral sequencing, complicated in viral sequencing, sequencing library, pooling sample, final pool quality control, sequencing, contaminating adapter dimer, sequencing depth, accurate loading on the sequencer, procedure for pool qc, quality control for illumina, many samples on the same run, library pooling, sequencer, viral read, full length adapter sequence, many sample, data quality possible run failure, pool molarity, quantification of the final pool, containing full length adapter sequence, adapter dimer, pool molarity cause, fewer read, pool qc, underestimation of pool molarity cause, data quality, sample, specific read depth, overestimation of the pool molarity, read quality, quality control, library cluster

Abstract

The ability to multiplex many samples on the same run makes Illumina sequencing a powerful and affordable tool for many researchers. When pooling samples for sequencing it is important that they are pooled at equal molarity to prevent over or under representation of any individual sample in the same run. Care also has to be taken in deciding how many samples can be combined to achieve the required sequencing depth, a factor that is further complicated in viral sequencing where the majority of the reads may originate from the host and only a small proportion corresponding to viral reads.

Also of major importance is the final pool quality control (QC) to ensure that all contaminating adapter dimers are removed, generate accurate size distribution and quantification of the final pool. Accurate QC ensures accurate loading on the sequencer, overestimation of the pool molarity can cause under clustering resulting in fewer reads and can even cause run failure. Underestimation of pool molarity causes over clustering, lowering the read quality and risking run failure.

Adapter dimers are small fragments containing full length adapter sequences which can bind to the flow cell and cluster. Due to their small size, they cluster more efficiently than the longer library fragments and so reduce the library-specific read depth as well as causing over clustering which reduces the data quality possible run failure. In addition free adapter can be incorporated into the library clusters resulting in index hopping and incorrect assignment of library barcodes.

Here we describe our standard workflow for determining the appropriate sequencing depth and pooling by equal molarity of Illumina sequencing libraries, along with our procedure for pool QC.

Materials

Reagents:

- Agencourt AmPure XP beads **Catalog #**A63880
- Qubit® dsDNA HS Assay Kit **Thermo Fisher Scientific Catalog #**Q32854
- High Sensitivity D5000 ScreenTape **Agilent Technologies Catalog #**5067-5592
- High Sensitivity D5000 Reagents **Agilent Technologies Catalog #**5067-5593

Additional reagents:
10 mM Tris pH8
Nuclease-free water
Absolute Ethanol

QC Equipment:

Equipment	
Qubit	NAME
Flurometer	TYPE
Invitrogen	BRAND
Q33228	SKU
https://www.thermofisher.com/order/catalog/product/Q33228	LINK

Equipment

NAME	TYPE	BRAND	SKU	LINK
4200 TapeStation System	Electrophoresis tool for DNA and RNA sample quality control.	TapeStation Instruments	G2991AA	https://www.agilent.com/en/product/automated-electrophoresis/tapestation-systems/tapestation-instruments/4200-tapestation-system-228263

Protocol materials

- Agencourt AmPure XP beads **Catalog #**A63880
- Qubit® dsDNA HS Assay Kit **Thermo Fisher Scientific Catalog #**Q32854
- High Sensitivity D5000 ScreenTape **Agilent Technologies Catalog #**5067-5592
- High Sensitivity D5000 Reagents **Agilent Technologies Catalog #**5067-5593
- Agencourt AmPure XP beads **Catalog #**A63880
- Qubit® dsDNA HS Assay Kit **Thermo Fisher Scientific Catalog #**Q32854
- High Sensitivity D5000 ScreenTape **Agilent Technologies Catalog #**5067-5592
- High Sensitivity D5000 Reagents **Agilent Technologies Catalog #**5067-5593

Troubleshooting

Safety warnings

Qubit reagent is known to bind nucleic acid and is provided as a solution in DMSO. Treat the Qubit dsDNA HS Reagent with the same safety precautions as all other potential mutagens and dispose of the dye in accordance with local regulations.

Before start

It is recommended that individual libraries are cleaned up and QC'd following the protocol **Library clean up and quality control for Illumina sequencing**. Each Illumina library should have an accurate size and quantification and be free from contaminants.

Pooling

1 Before pooling the samples determine the number of samples per pool. The estimated reads per sample will depend on the required depth, genome length and expected percentage of on-target reads. The number of samples per pool will also depend on the type of Illumina sequencer/cartridge used and the read length.

You can use the attached calculator to determine the samples/cartridge and how many pools are required for your experiment.

SequencingDepth.xlsx

2 For each pool calculate the volume required so that each library is present in equal quantities.

> Note
>
> It is important to pool by molarity rather than mass as this will ensure equal numbers of molecules of each library. Otherwise longer or shorter fragment sizes will be under or over represented respectively.

Library molar concentration can be calculated from the library fragment size and mass concentration as follows:

$$Conc(nM) = \frac{sample(ng/\mu l)}{size(bp) * 660(g/mol)} * 1000000$$

You can use the attached calculator to determine the pooling volumes.

EquimolarPooling.xlsx

You may be required to pool sub-pools instead of individual libraries. If so an additional weighting by the number of samples per sub-pool is required. See the attached calculator.

EquimolarPooling_subPools.xlsx

3 Using a fresh 1.5 mL DNA LoBind tube carefully add the required volume of each library.

Note

If possible use low-retention tips to improve pipetting accuracy.

4 Optional: Concentrate pool using 1.4X Ampure XP.

Equilibrate Ampure XP beads to Room temperature and vortex well to mix.
Agencourt AmPure XP beads **Catalog** #A63880

4.1 Measure the total volume of the pool and add 1.4X volume of Ampure XP beads and mix well.

4.2 Incubate at Room temperature room temperature for 00:05:00 5m

4.3 Place on a magnetic rack for 00:05:00 until beads and solution have fully separated. 5m

Note

It is important to wait until the beads have fully separated, the time it takes will vary depending on the ratio of Ampure beads:sample and the level of DNA.

4.4 Carefully add 200 µL 80% Ethanol (freshly prepared) without disturbing the beads.

4.5 Incubate at Room temperature for 00:00:30 . 30s

4.6 Place on magnetic rack for 00:01:00 until the beads and solution have fully separated. 1m

4.7 Keeping on the magnet and carefully remove supernatant without disturbing the beads.

4.8 Repeat wash with 200 µL 80% Ethanol (freshly prepared) .

4.9 Remove all traces of ethanol.

> Note
>
> Sample can be briefly centrifuged and placed back on the magnet to remove residual ethanol.

4.10 Keeping on the magnet air dry for up to 00:03:00 . 3m

> Note
>
> Take care not to over dry the beads.

4.11 Add the required volume of 10 mM Tris pH 8.0, remove from the magnet and mix well to fully suspend the beads.

4.12 Incubate at Room temperature for at least 00:02:00 to elute the DNA. 2m

4.13 Place back on magnet until beads and solution have fully separated.

4.14 Transfer the supernatant containing the pool DNA to a fresh 1.5 mL DNA LoBind tube.

Pool QC 2m

5 It is recommended that library quantification is performed using a fluorometric method (e.g. Qubit) rather than an absorbance based method (e.g. Nanodrop) as it is more specific, sensitive and accurate.

Equipment

Qubit	NAME
Flurometer	TYPE
Invitrogen	BRAND
Q33228	SKU
https://www.thermofisher.com/order/catalog/product/Q33228	LINK

5.1 Prepare 0.5 mL thin-walled PCR tubes, including 2 tubes for standard solutions and 3 for each pool.
Label the tube lids and not the sides.

Note

For increased accuracy it is recommended that multiple readings are taken per pool.

5.2 Prepare the Qubit dsDNA High Sensitivity master mix for the total number of samples and standards with an excess.

	A	B
	Component	**Volume (µl)**
	Qubit dsDNA HS Buffer	199
	Qubit dsDNA HS Reagent	1
	Total	200

Qubit® dsDNA HS Assay Kit **Thermo Fisher Scientific Catalog #**Q32854

Safety information

Qubit reagent is known to bind nucleic acid and is provided as a solution in DMSO. Treat the Qubit dsDNA HS Reagent with the same safety precautions as all other potential mutagens and dispose of the dye in accordance with local regulations.

Note

Ensure that the Qubit standards are equilibrated to room temperature for 30 min before use.

5.3 Aliquot the Qubit dsDNA High Sensitivity master mix into the assay tubes as follows:

Standards 190 µL Qubit master mix
Samples 199 µL Qubit master mix

5.4 Add 10 µL Qubit standard to each standard assay tube.

5.5 Add 1 µL pool to each sample assay tube.

5.6 Vortex assay tubes and briefly centrifuge.

5.7 Incubate at Room temperature for 00:02:00 . 2m

5.8 Select dsDNA high Sensitivity assay on the Qubit Fluorimeter and press "Read Standards".

5.9 Insert Standard 1 and 2 into the sample chamber when prompted, close the lid and and press "Read Standard".

5.10 Select "Run samples" and select the sample volume as 1 µL .

5.11 Insert a sample tube into the sample chamber, close the lid and press "Read tube".

5.12 Record the concentration of the Qubit sample in ng/µL.

5.13 For each pool ensure all readings are within the same range.

Note

If required additional readings can be taken.

5.14 Calculate the average pool concentration from the replicates ensuring any outliers are removed.

6 It is recommended that the libraries are visualised using capillary electrophoresis, we describe visualisation with a TapeStation and High Sensitivity D5000 ScreenTape. Alternatives such as the BioAnalyzer or Fragment Analyzer can also be used.

The purpose is to provide a size for the library fragments to give accurate molar quantification and to determine the quality of the library.

Equipment

4200 TapeStation System	NAME
Electrophoresis tool for DNA and RNA sample quality control.	TYPE
TapeStation Instruments	BRAND
G2991AA	SKU
https://www.agilent.com/en/product/automated-electrophoresis/tapestation-systems/tapestation-instruments/4200-tapestation-system-228263	LINK

6.1 Ensure that the D5000 ScreenTape and Reagents are equilibrated to Room temperature at least 00:30:00 before use, vortex and briefly centrifuge. 30m

High Sensitivity D5000 ScreenTape **Agilent Technologies Catalog #**5067-5592

High Sensitivity D5000 Reagents **Agilent Technologies Catalog #**5067-5593

Note

D5000 ScreenTape has a sizing range of 100-5000 bp. The D1000 ScreenTape and Reagents could be used as an alternative which has a sizing range of 35-1000 bp.

6.2 Prepare a dilution of each pool to approximately 1 ng/µL.

6.3 In fresh PCR strip tubes prepare the ladder assay tube as follows:

	A	B
	Component	**Volume (µl)**
	D5000 sample buffer	2
	D5000 ladder	2
	Total	4

6.4 Prepare the sample assay tubes as follows:

	A	B
	Component	**Volume (µl)**
	D5000 sample buffer	2
	Diluted Pool	2
	Total	4

6.5 Spin down, using IKA vortexer mix at 2000 rpm, 00:01:00 then spin down again. 1m

6.6 Load the assay tubes and ScreenTape into the TapeStation instrument.

6.7 Select the required sample/ladder positions in the TapeStation software and click "start".

Note

The TapeStation analysis software will open automatically at the end of the run to display the results.

6.8 Analyse the results, the pool should generate a similar trace to the individual libraries.

If needed follow the same troubleshooting guidelines as in the protocol **Library clean up and quality control for Illumina sequencing**.

6.9 Determine the pool fragment peak size in bp.

Note

If your pool doesn't give a single sharp peak then a region size can be used instead.

6.10 Calculate the pool molar concentration using the fragment size (Step 6.9) and mass concentration (Step 5.14).

$$Conc(nM) = \frac{sample(ng/\mu l)}{size(bp) * 660(g/mol)} * 1000000$$

Note

The final pool should have a concentration at least 10 nM.