

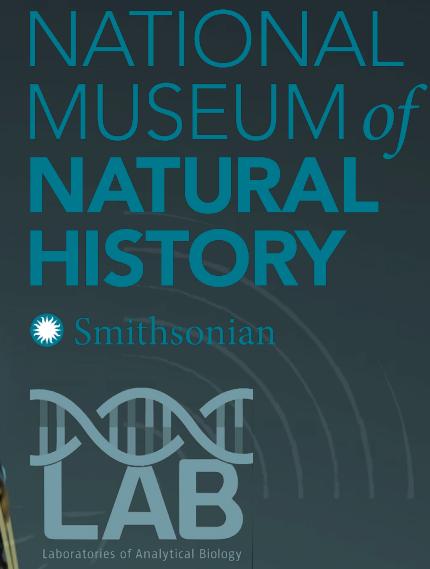
Genome Skimming: Library Pooling, qPCR, and Sequencing

Katie Murphy

Laboratories of Analytical Biology

NMNH

March 2024



Code of Conduct (Carpentries)



We are following the Carpentries organization's Code of Conduct



https://docs.carpentries.org/topic_folders/policies/code-of-conduct.html

Code of Conduct (excerpts)

https://docs.carpentries.org/topic_folders/policies/code-of-conduct.html

We are dedicated to providing a welcoming and supportive environment for all people, regardless of background or identity.

- Use welcoming and inclusive language
- Be respectful of different viewpoints and experiences
- Gracefully accept constructive criticism
- Focus on what is best for the community
- Show courtesy and respect towards other community members

Topics for today



Library Pooling 101



Quantifying with qPCR



Basics of Illumina sequencing



Submitting a sequencing run to LAB

Library Pooling 101



Library molarity & pooling

- Make sure all libraries in a pool have different indices!
- Libraries should be pooled in equimolar amounts.
- Why pool based on molarity instead of mass (ng/ μ L)?
- Example: You have two libraries, each at a concentration of 2 ng/ μ L. If you pool 5 μ L of each together (10 ng of each), how would each end up represented in your final pool if:
 - One library has a length of 300 bp, and the second has a length of 600 bp?
 - What about if they were 200 bp and 1000 bp?
 - What if they were both the same size?
- Now, what if you had two libraries at 2 nM each, and pooled 5 μ L together. Would the same size differences described above make a difference?

Library molarity & pooling

- We need to know our libraries' molarities, but yesterday we measured ng/ μ L! So now what do we do??
 - We can't directly measure nM. Instead, we calculate it based on the ng/ μ L from Quant-iT and the average size (bp) from the TapeStation.

$$\frac{(\text{concentration in ng}/\mu\text{L})}{(660 \text{ g/mol}) \times (\text{average library size in bp})} \times 10^6 = \text{concentration in nM}$$

Average library size should be the total length (including index/adapters), not just your insert length.

Yay more math! ☺
(thank goodness for Excel)

AutoSave Home Insert Draw Page Layout Formulas Data Review View Automate Picture Format Tell me

Cut Copy Wrap Text Paste Conditional Formatting

Format <input type="button" value="A<sup>5

Questions on pooling?



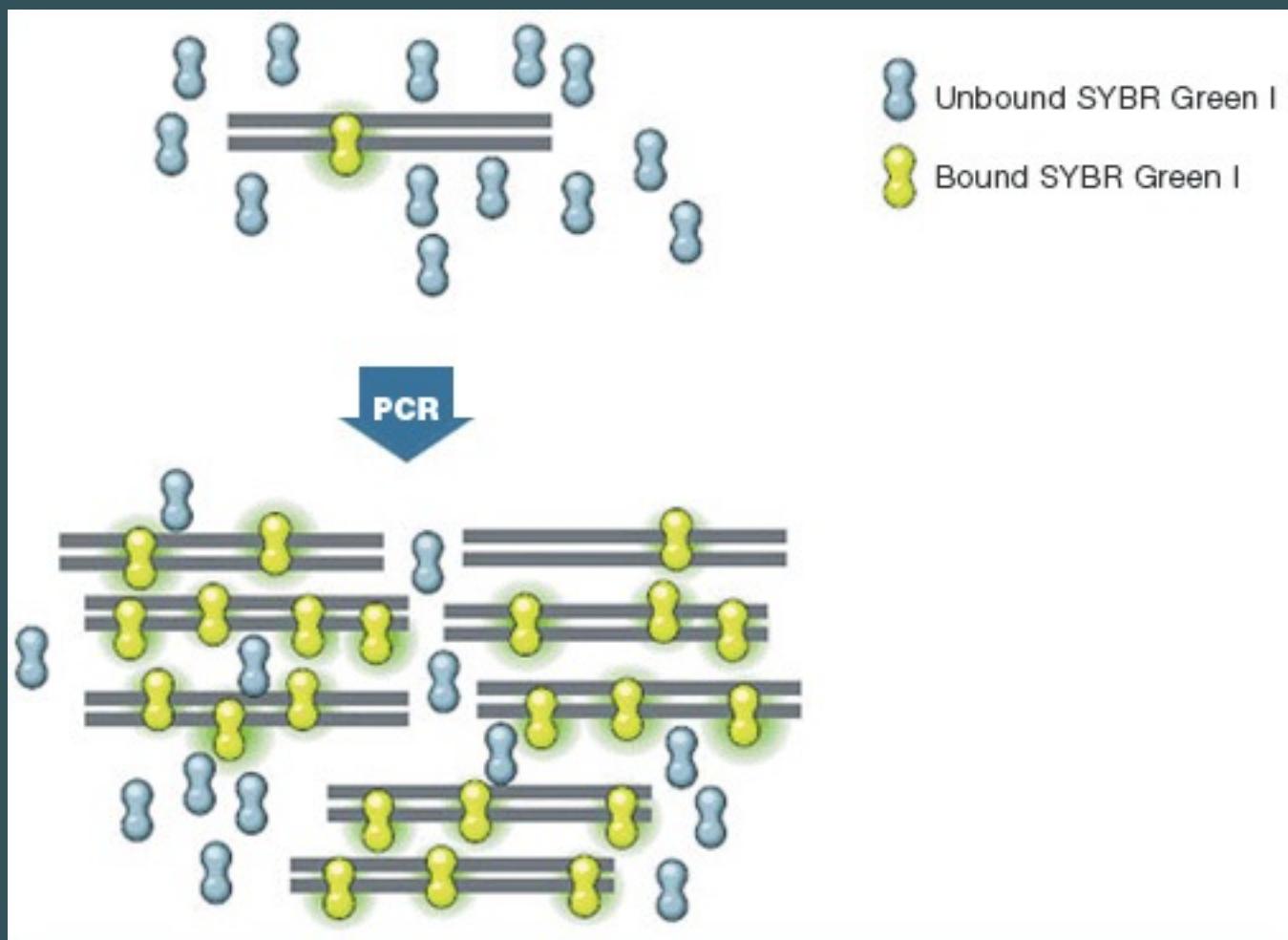
Quantifying libraries with qPCR



qPCR

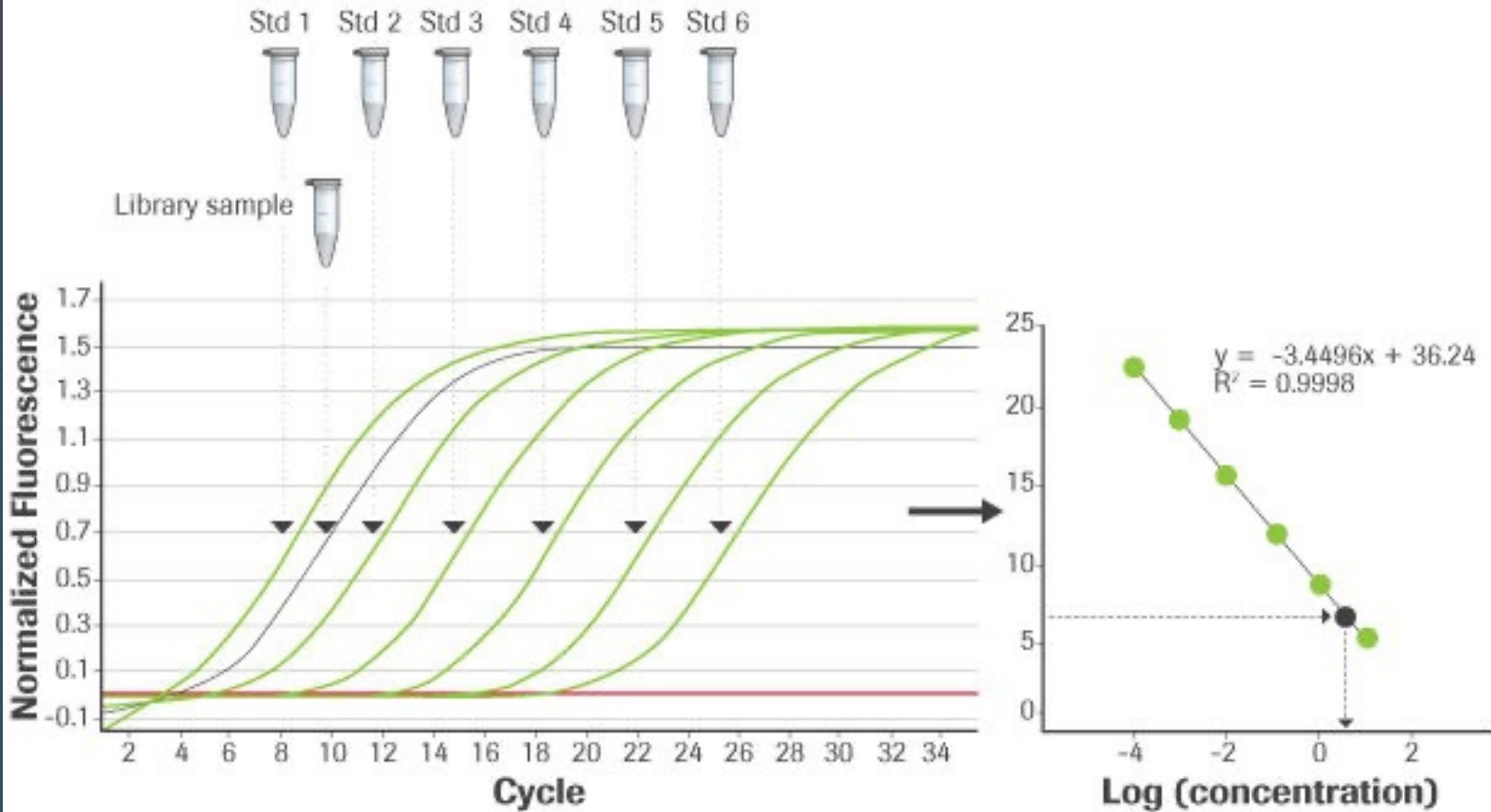
- qPCR is the most accurate way to quantify your libraries, as it is a better measure of **sequenceable** DNA
 - Quant-iT measured **total** dsDNA
- Labor-intensive and can get expensive, so often only used on final sequencing pools, not individual libraries
- Sequencing cores often perform this service for you
- Recommended assay kit depends on the qPCR machine
 - For LAB's qPCR machines, we use: KAPA Library Quant Kit for Illumina - ROX Low

How does qPCR work?



- <https://www.bio-rad.com/webroot/web/images/tlp/real-time-pcr-detection-standard-pcr-primer-and-DNA-binding-dye.jpg>

How does qPCR work?



Let's head to the lab...

- I'll show you the supplies, our qPCR machine (QuantStudio 6 Pro), and how to set up a run*.
- We'll then take a look at some qPCR data, so you see how to perform the analysis.

unfortunately, we're not setting up an actual run today



Basics of Illumina Sequencing

How does Illumina sequencing work?

- [Illumina Sequencing Video](#)
- SBS = "Sequencing By Synthesis"
- Short-read sequencing
 - Longest Illumina read length is 300 bp
 - Read length is defined by the number of sequencing cycles, and also limited by the sequencing chemistry
- Additional information: <https://www.illumina.com/science/technology/next-generation-sequencing.html>



Illumina Sequencing: Clusters on a flow cell - the images

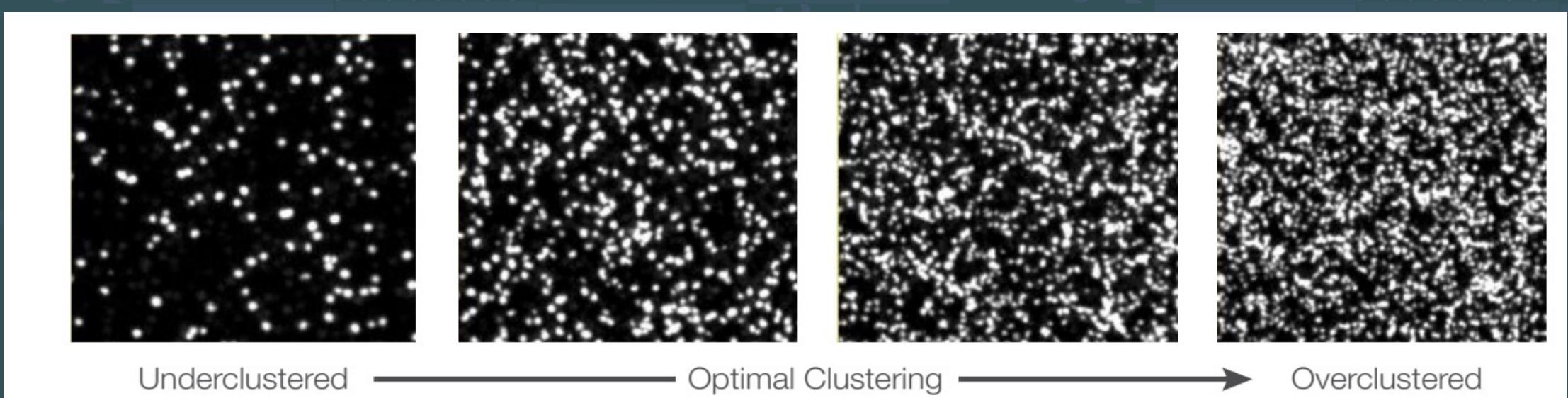
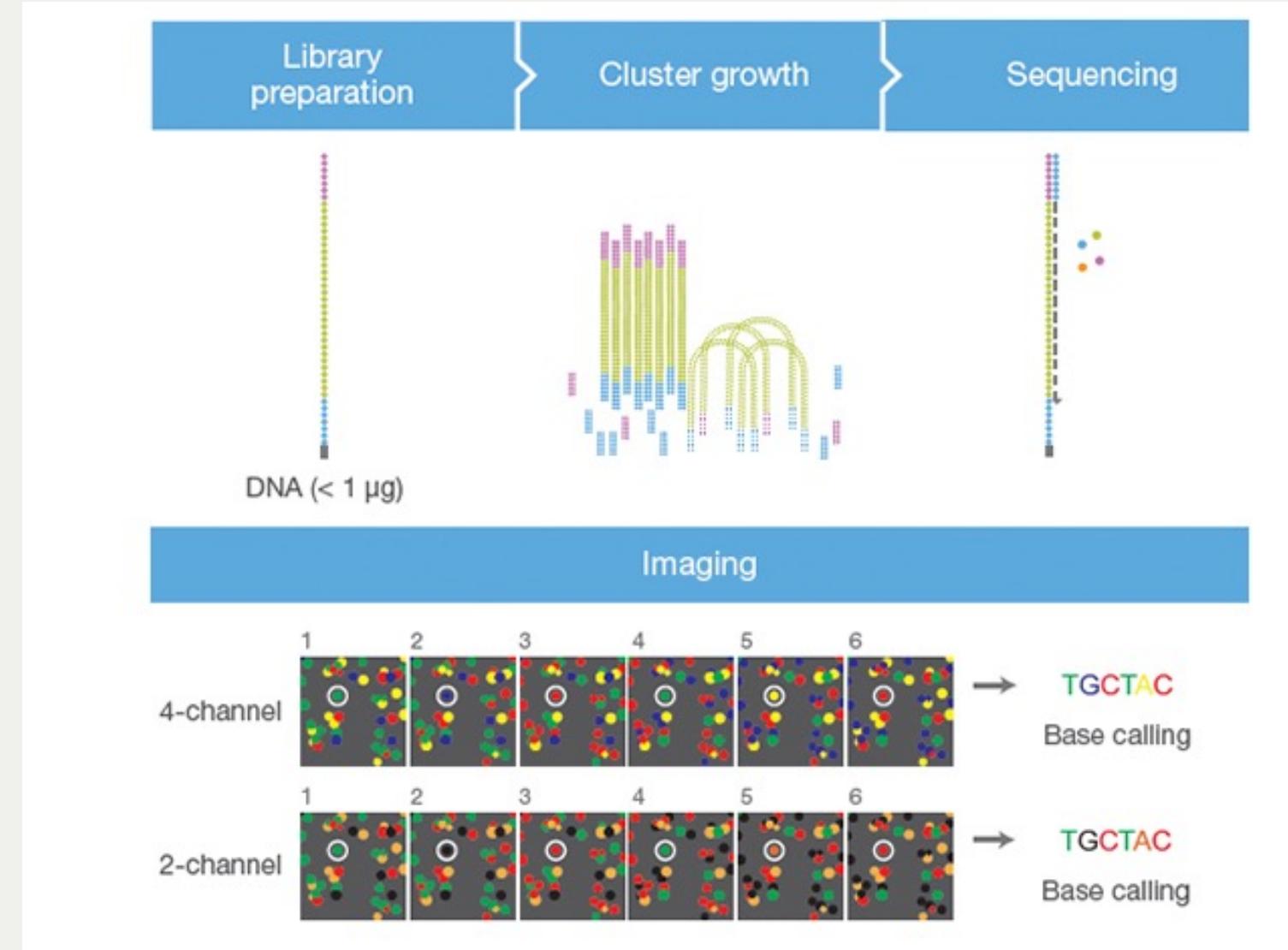


Figure 5: Thumbnail Images. Thumbnail images showing cluster densities ranging from underclustered to overclustered.

Illumina Sequencing: Two vs Four Channel Chemistry



Sequencing Considerations

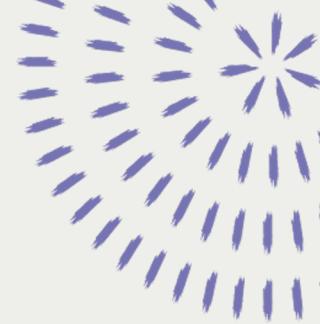
- Read depth - how many do you need?
 - Read depth = the average number of reads at any given base position.
- Read length - how long (bp) do they need to be?
- Which specific sequencing instrument is most appropriate for your needs?



How many reads do you need?

- Targeted sequencing coverage often ranges from 0.1X - 10X of genome size.
- E.g. if genome is 400 Mb, then a 0.5X genome skimming study would want 200 Mb worth of data.
 - number of required reads = $\frac{\text{total bp of data needed}}{\text{read length}}$
 - If read length is 150bp, you'd need to target $\frac{200,000,000 \text{ bp}}{150 \text{ bp}} \rightarrow 1,333,333$ reads (1.3M)
 - If read length is 300bp, you'd only need to target $\frac{200,000,000 \text{ bp}}{300 \text{ bp}} \rightarrow 666,667$ reads (667K)
- Recommended: perform a literature search to see what coverage is likely most appropriate for you

Illumina Instruments



	iSeq 100	MiniSeq	MiSeq Series +	NextSeq 550 Series +
Run Time	9.5–19 hrs	4–24 hours	4–55 hours	12–30 hours
Maximum Output	1.2 Gb	7.5 Gb	15 Gb	120 Gb
Maximum Reads Per Run	4 million	25 million	25 million [†]	400 million
Maximum Read Length	2 × 150 bp	2 × 150 bp	2 × 300 bp	2 × 150 bp

LAB currently has
a MiSeq



LAB also has access to this instrument at
MSC through our colleagues at WRBU

Illumina Instruments

LAB is getting a
NextSeq 1000
in 2024



NextSeq 1000 & 2000



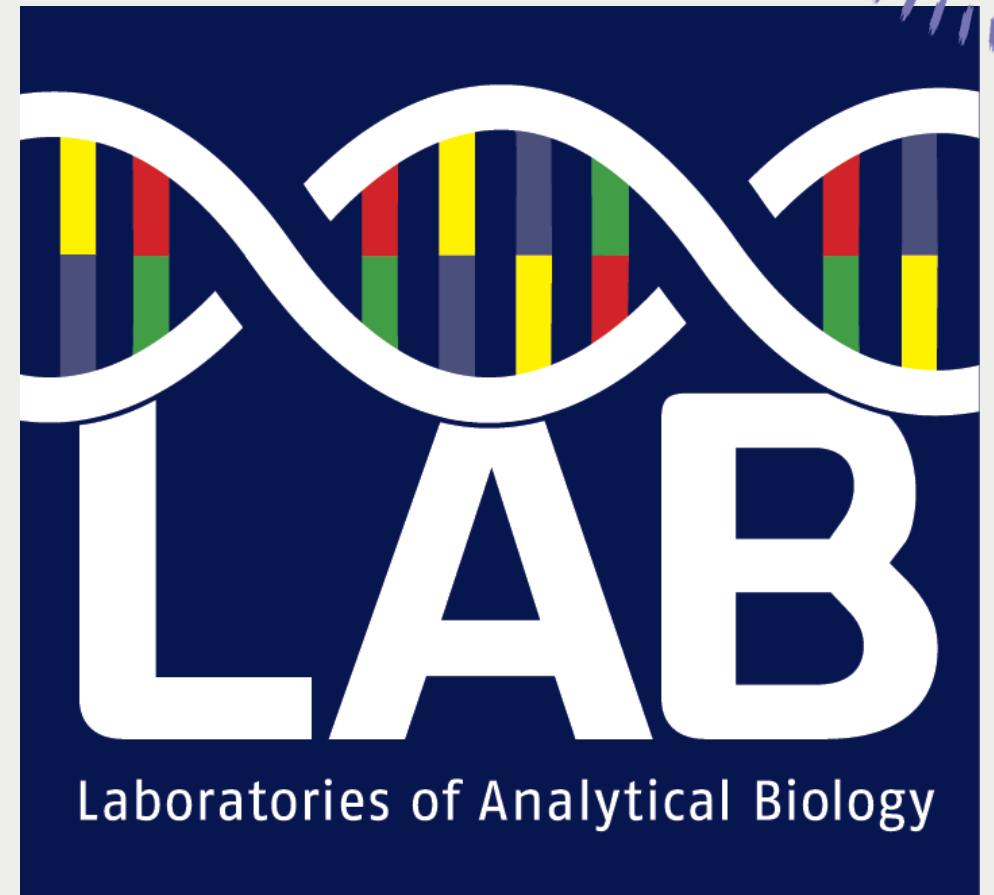
NovaSeq 6000 Series +



NovaSeq X Series

Run Time	11-48 hours	~13-38 hours (dual SP flow cells) ~13-25 hours (dual S1 flow cells) ~16-36 hours (dual S2 flow cells) ~44 hours (dual S4 flow cells)	~15-21 hours (1.5B flow cells [‡]) ~18-25 hours (10B flow cells [‡]) ~48 hours (25B flow cells [‡])
Maximum Output	360 Gb *	6000 Gb	16 Tb
Maximum Reads Per Run	1.2 billion *	20 billion	26 billion (single flow cells) 52 billion (dual flow cells)
Maximum Read Length	2 x 150 bp	2 x 250 bp**	2 x 150 bp

How to submit a library pool for sequencing at LAB



Submitting for a MiSeq run at LAB

1

Place an order for your MiSeq reagent kit on labdb.si.edu.

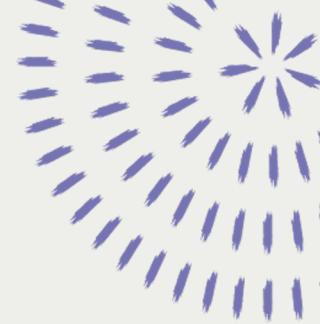
- If you aren't sure what kit you need, please discuss w/ me.

2

Fill out a MiSeq submission form with your library and run specifics.

3

Email me (murphykr@si.edu) about scheduling your sequencing run. Make sure to attach your submission form!



View our "MiSeq submission guidelines" on LAB's website (on Darwin/Sharepoint) for full details, and/or contact me with questions. To access Darwin, you must be on the SI network (or be connected via VPN or telework.si.edu)

General Guidelines for your library pool

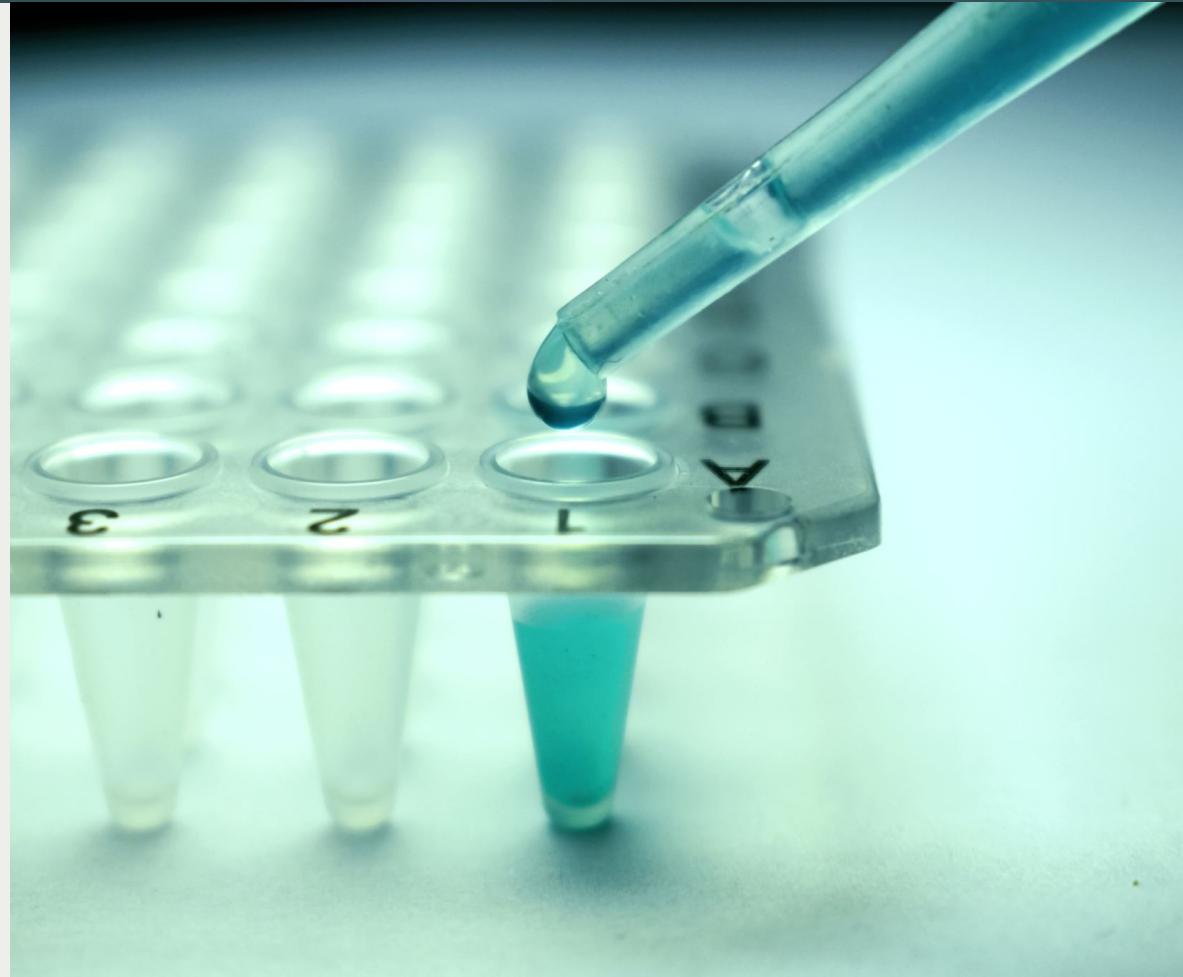
When
submitting
a library
pool to
LAB,
please do
the
following:

Submit at least 10 uL of your pool at a concentration of ≥ 4 nM for v3 kits or ≥ 2 nM for v2 kits; higher volume is preferred.

- Final sample (pool) must be quantified by fluorimetry (Qubit or Quant-iT) or qPCR after pooling and any subsequent dilutions.
- Library size should be determined by Tapestation/Bioanalyzer (best options) or agarose gel electrophoresis.
- The combination of TapeStation/Bioanalyzer and qPCR will give us a good idea what your library quality is. Insufficient library quality is the most common reason for failed runs!
- If your final pool molarity is less than 2 nM, please let me know ahead of time to verify if your pool can still be properly sequenced.

If you need to outsource a sequencing run...

- We highly recommend using Genohub.com
 - You'll usually get better pricing and turnaround times than trying to find a sequencing core yourself
 - If using this site, LAB can help handle the PO so that you just get a LAB bill instead. Up to you if you want to include us!



**What
questions do
you have?**

