How many reads/reps/samples

• For *de novo* transcriptome assembly, diff expression
  • SNP calling
• For reference generation?? (20-30M? Francis et al 2013)
How many reads/reps/samples
How many reads/reps/samples

- For \textit{de novo} transcriptome assembly, diff expression
  - SNP calling
- For reference generation??
- What is the n treatments are very different?
- Now you have a reference, how many reads for replicates?
- How to sequence these samples?
Now we have sequence reads…

- Adapter trimming (Trimmomatic)
- Quality trimming
Quality trimming of NGS data

- Universal practice
Quality trimming of NGS data

Phred=5
Quality trimming of NGS data
Quality trimming of NGS data

Phred=20

Probability of nucleotide error vs. Nucleotide Position
Trimming Experiment

• 2 Illumina datasets > adapter trimmed.
• Subsampled to 10M, 20M, 50M, 75M, 100M PE reads.
• Trimmed at Phred 0,2,5,10,20
• Assembled using Trinity and SOAPdenovo-Trans
• Developed metrics for evaluating transcriptome assemblies.

MacManes, Frontiers in Genetics 2014
Quality trimming reduces error

Number of nucleotide errors per Mb of assembly

MacManes, Frontiers in Genetics 2014
Quality trimming reduces error

Number of nucleotide errors per Mb of assembly

No Trim | Phred=2 | Phred=5 | Phred=10 | Phred=20

- **SOAP10M**
- **SOAP20M**

Quality trimming reduces error.
Quality trimming reduces BLAST hits

Percent diff in number of unique BLAST hits

MacManes, Frontiers in Genetics 2014
Quality trimming reduces BLAST hits

Percent diff in number of unique BLAST hits

No Trim | Phred=2 | Phred=5 | Phred=10 | Phred=20
-------|--------|--------|----------|----------
10M     |        |        |          |          
20M     |        |        |          |          
50M     |        |        |          |          
75M     |        |        |          |          
100M    |        |        |          |          

Quality trimming reduces BLAST hits

Percent diff in number of unique BLAST hits

No Trim | Phred=2 | Phred=5 | Phred=10 | Phred=20
-------|--------|--------|----------|----------
SOAP10M|        |        |          |          
SOAP20M|        |        |          |          

Quality trimming reduces BLAST hits
Quality trimming reduces complete CDS

MacManes, Frontiers in Genetics 2014
Quality trimming reduces complete CDS
Summary

- Trimming does reduce assembly error, but at the cost of content & contiguity.
- **Proposed guidelines.**
  1. To max assembly content and contiguity $\rightarrow$ Trim at 0 or 2ish
  2. If concerned about error $\rightarrow$ Trim at Phred=5
  3. Usually probably never trim at Phred $\geq$ 10

MacManes, Frontiers in Genetics 2014
Links

- https://www.youtube.com/watch?v=5NiFibnbE8o
- http://liorpachter.wordpress.com/