

# HOW MUCH SEQUENCING DO I NEED?

KEVIN CHILDS

GENOMICS CORE

# HOW MUCH SEQUENCING?

Really three questions

How much sequence is required for good experimental design?

What type of sequencing run is best?

How many lanes of sequencing?

**\*\*All based on Illumina sequencing options\*\***

# EXPERIMENTAL DESIGN

What are you sequencing?

Genome

- de novo assembly
- resequencing project

Transcriptome

- de novo assembly
- gene expression project

Whole Meta-Genomes, Small RNAs, CHIP-Seq, Exome Capture, Amplicon Sequencing

# WHAT TYPE OF SEQUENCING RUN

Single end or paired end?

What read length?

35 bp, 50 bp, 75 bp, 150 bp, 250 bp, 300 bp

Not all read lengths available on all machines

Assembly of genome or transcriptome?

PE 150 bp, 250 bp, 300 bp

Counting experiment?

SE 35 bp, 50 bp, 75 bp

# HOW MANY LANES OF SEQUENCING

## For genome assembly

- answer depends on desired coverage
- new assembly 75X – 100X
- resequencing 10X – 20X
- long-read error correction 20X – 30X

## For transcriptome assembly

- number of genes in the genome
- complexity of the transcriptome

# lanes required =

desired Gbp / expected Gpb per lane

# HOW MANY LANES OF SEQUENCING

## For gene expression analysis

- counting experiment
- what is typical in your field
- consider ploidy
- how many replicates
- account for variability between samples

# lanes required =

minimum reads per sample X # replicates X  
# samples X fudge factor / reads per lane

# GENOME SEQUENCING EXAMPLE #1

## New eukaryotic genome

- 1.2 Gbp genome
- target 80X coverage
- PE 150 reads
- HiSeq 4000 averages 350 million reads/lane

# lanes required =

desired Gbp / expected Gpb per lane

Resequencing – What changes?

# GENOME SEQUENCING EXAMPLE #2

## New prokaryotic genomes

- 12 different bacterial isolates
- 8 Mbp genome
- target 40X coverage
- PE 150 reads
- HiSeq 4000 averages 350 million reads/lane

# lanes required =

desired Gbp / expected Gpb per lane



# GENOME SEQUENCING EXAMPLE #2

## New prokaryotic genomes

- 12 different bacterial isolates
- 8 Mbp genome
- target 40X coverage
- ~~PE 150 reads~~
- ~~HiSeq 4000 averages 350 million reads/lane~~
- MiSeq v2 Standard PE 250 give ~12-15 Gbp

# lanes required =

desired Gbp / expected Gpb per lane

# GENOME SEQUENCING EXAMPLE #3

## Whole genome metagenomic sequencing

- unknown number of fungal, bacterial & other species
- unknown genome sizes
- PE 150 reads
- HiSeq 4000 averages 350 million reads/lane

# lanes required =

desired Gbp / expected Gpb per lane

# TRANSCRIPT SEQUENCING EXAMPLE

Goal is transcript assembly

- 25,000 genes
- target of 60 million reads per sample
- PE 150
- HiSeq 4000 averages 350 million reads/lane

# lanes required =

desired Gbp / expected Gpb per lane

Bonus – How many different mRNA samples can be prepared and loaded?

# GENE EXPRESSION EXAMPLE

Gather counts for differential expression analysis

- Mammals: 30 to 50 million reads per sample
- Plants: 25 million reads per sample
- Replicates: 3 to 5
- # samples is experiment-dependent
- SE 50
- HiSeq 4000 averages 350 million reads/lane

# lanes required =

minimum reads per sample X # replicates X  
# samples X fudge factor / reads per lane

# AMPLICON SEQUENCING EXAMPLE

Sequencing same target from multiple samples

- metagenomic survey
- specific target from many individuals
- barcoding required
- PE 250
- coverage dependent on number of samples
- variation between samples is very large

# lanes required =

desired coverage X # samples X  
fudge factor / reads per lane

# SMALL RNA SEQUENCING EXAMPLE

Gather counts for differential expression analysis

- for miRNAs, 10 million reads are common
- replicates 3 to 5
- SE 50

# lanes required =

minimum reads per sample X # replicates X  
# samples X fudge factor / reads per lane