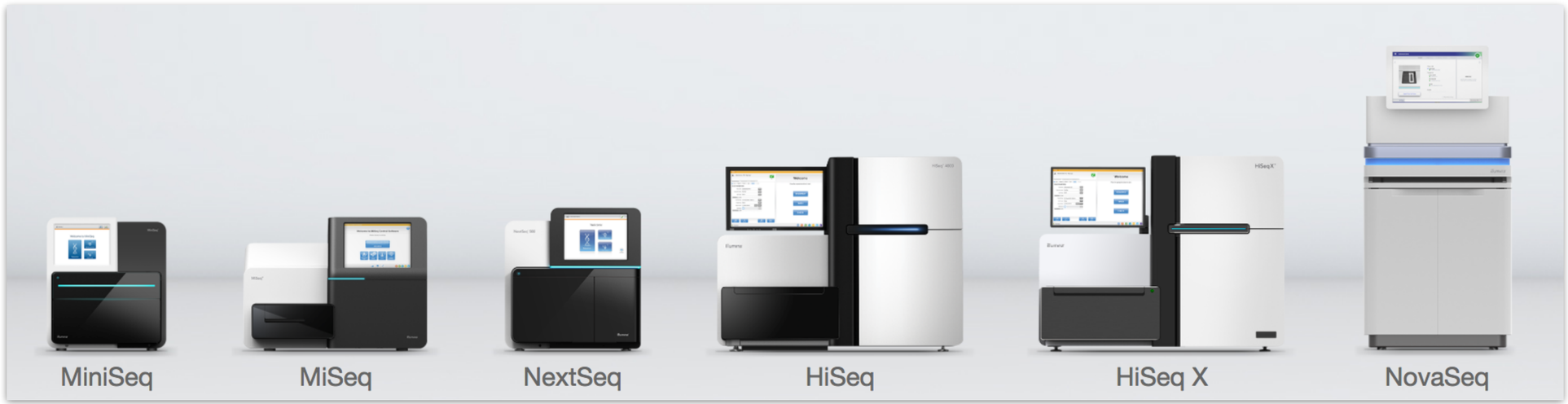
Sequencing technologies



Illumina: Sequencing Platforms

https://www.illumina.com/systems/sequencing-platforms.html

Benchtop

iSeq 100 System

MiniSeq System

MiSeq Series ⊕

NextSeq Series ⊕

Run Time	9–17.5 hours	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

Production-Scale

NextSeq Series ⊕

HiSeq Series ⊕

HiSeq X Series‡

NovaSeq 6000 System

Run Time	12–30 hours	< 1–3.5 days (HiSeq 3000/HiSeq 4000) 7 hours–6 days (HiSeq 2500)	< 3 days	16–36 hours (Dual S2 flow cells) 44 hours (Dual S2 flow cells)
Maximum Output	120 Gb	1500 Gb	1800 Gb	6000 Gb
Maximum Reads Per Run	400 million	5 billion	6 billion	20 billion
Maximum Read Length	2 × 150 bp	2 × 150 bp	2 × 150 bp	2 × 150 bp

Flow cell

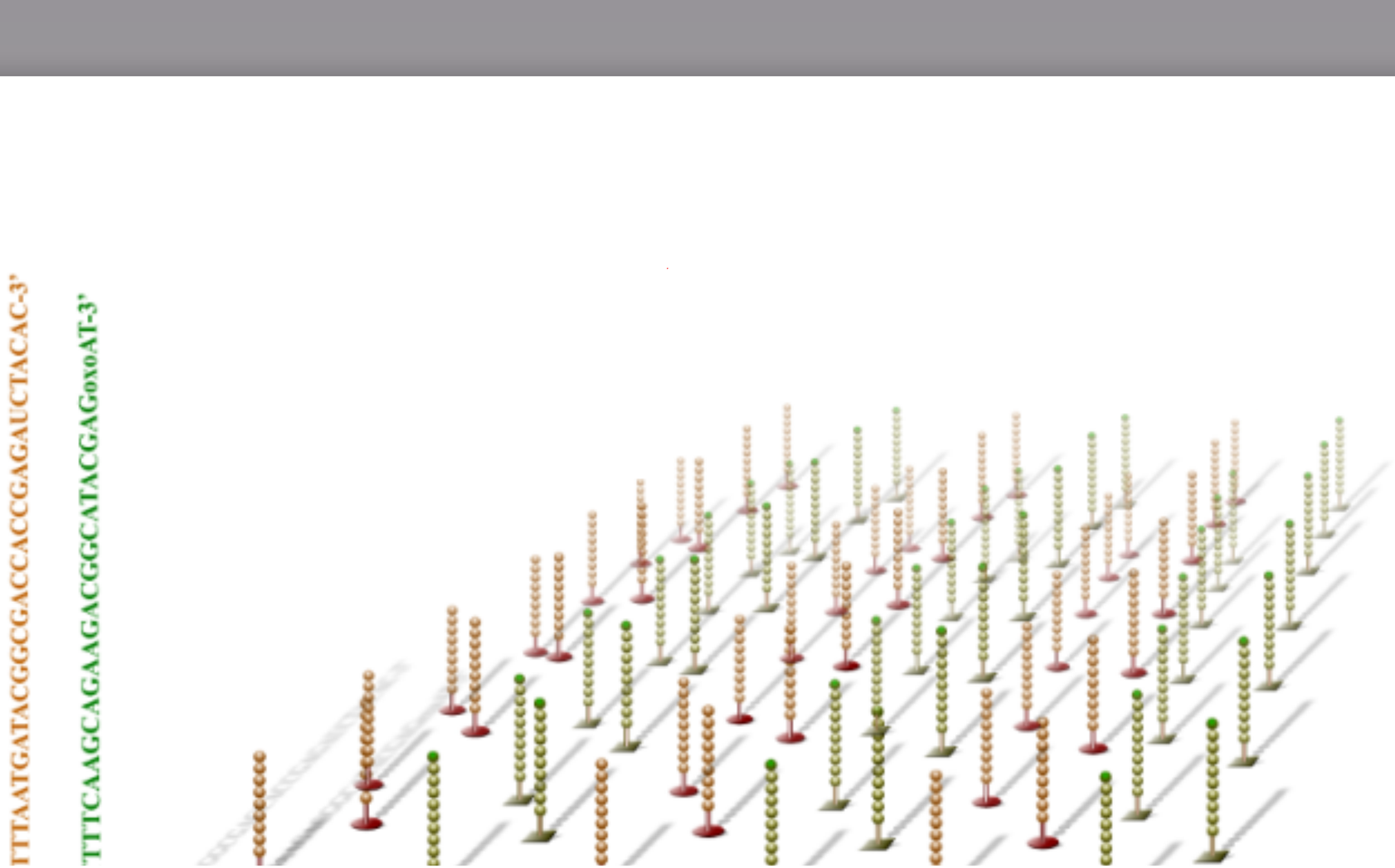
- A glass surface coated with an arrangement of paired oligos that are complementary to the adapters added to your template molecules.
- A flow cell can have varying numbers of lanes, depending on the sequencing machine.
http://training.bioinformatics.ucdavis.edu/docs/2014/09/september-2014-workshop/Monday_JF_HTS_lecture.html

HiSeq 2500 (2 lane)

HiSeq 3000/4000

NextSeq 500

http://www.illumina.com



Illumina: flow cell

http://training.bioinformatics.ucdavis.edu/docs/2014/09/september-2014-workshop/Monday_JF_HTS_lecture.html

Introduction to Sequencing by Synthesis

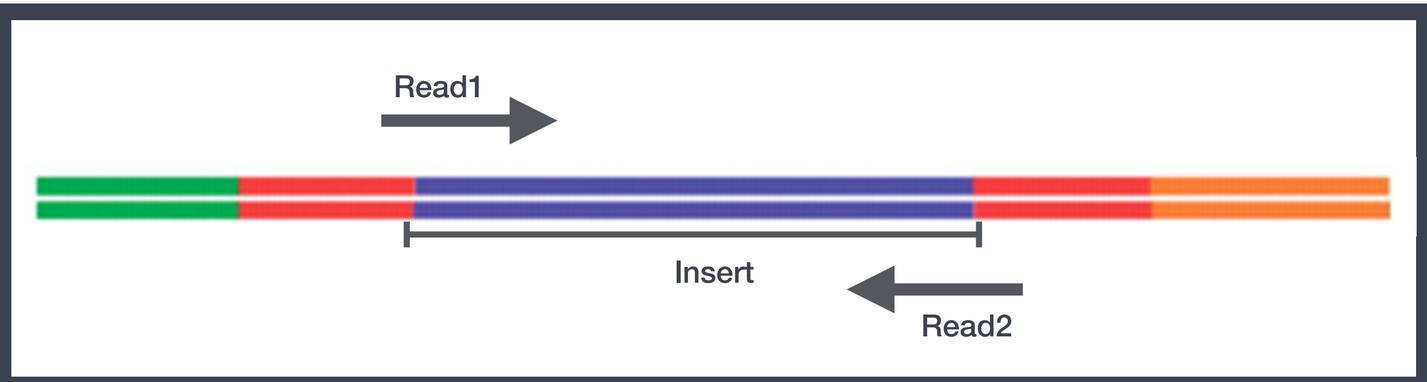
Video courtesy of Illumina YouTube channel

Number of clusters ~= Number of reads

Number of sequencing cycles ~= Length of reads

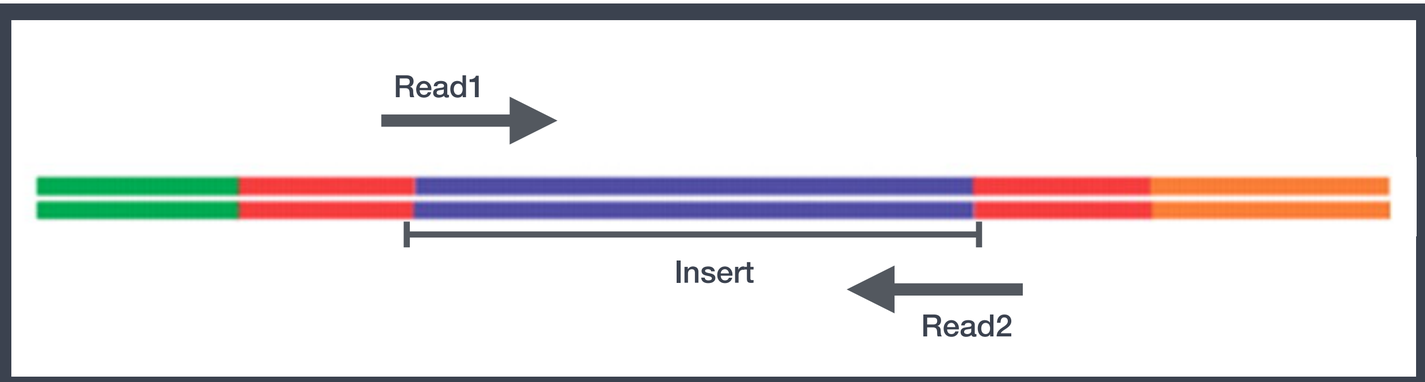
Single end or Paired end?

Depends on your goals, paired-end reads are better for reads that map to multiple locations, for assemblies and for splice isoform differentiation.



- ✓ SE - Single end dataset => Only Read1
- ✓ PE - Paired-end dataset => Read1 + Read2
  - can be 2 separate FastQ files or just one with interleaved pairs

Options for sequencing



- ✓ SE - Single end dataset => Only Read1
- ✓ PE - Paired-end dataset => Read1 + Read2
  - can be 2 separate FastQ files or just one with interleaved pair
- ✓ Fragment length: ~300-500bp
- ✓ Read length: 50bp - 250bp, depends on the sequencer (HiSeq2500, MiSeq, NextSeq)

Options for sequencing

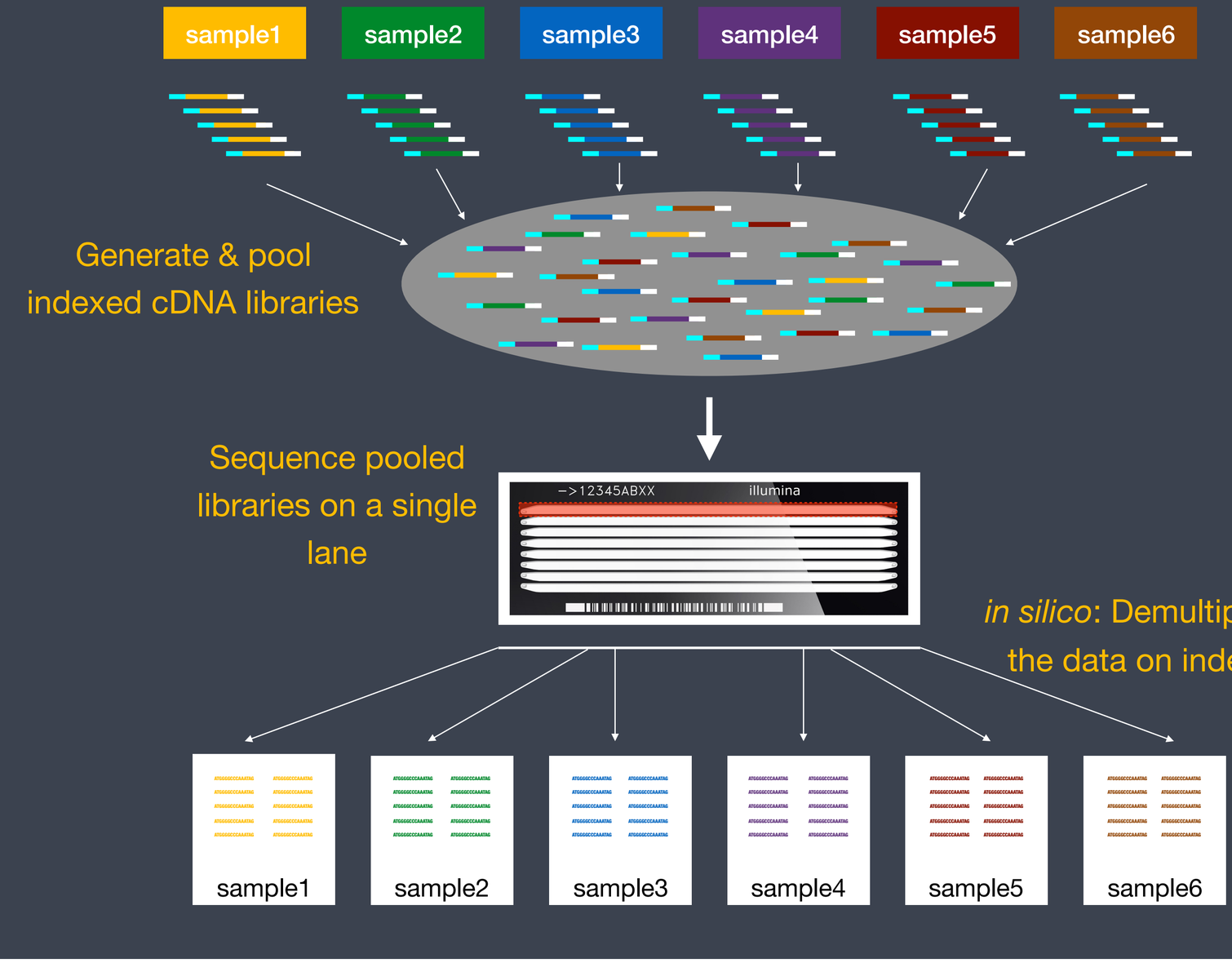
Single end or Paired end?

For differential gene expression, which one you pick depends on-

- If you are specifically interested in isoform-level differences
- The abundance of paralogous genes in your system of interest
- Your budget, paired-end data is usually 2x more expensive

Multiplexing (with barcodes and indices)

- Charges for sequencing are usually per lane of the flow cell, and usually you don’t need one lane per sample
- Multiplexing allows you to sequence multiple samples per lane with addition of indices (within the Illumina adapter) or special barcodes (outside the Illumina adapter).


sample1
sample2
sample3
sample4
sample5
sample6
Generate & pool indexed cDNA libraries
Sequence pooled libraries on a single lane
–>12345ABXX
illumina
in silico: Demultiplex the data on index
sample1
sample2
sample3
sample4
sample5
sample6

Pacific Biosciences: http://www.pacb.com/

Oxford Nanopore (MinION): https://nanoporetech.com/

10X Genomics: https://www.10xgenomics.com/

Other Sequencing Platforms

	Advantages	Disadvantages
Pacific Biosciences	Iso-Seq protocol for transcripts up to 10Kb, high base calling accuracy	High cost, large machines
Oxford Nanopore	Accurate quantitative data for short transcripts (< 700bp), portable, high yield	High errors rate affects assembling de novo transcripts, higher amount of cDNA input
10X Genomics	Low cost (integrated with short-read technology), barcoding for accurate isoform detection, low error rates	Extra preparation step (barcode), extra computational step

Transcriptomics with long read technologies

	Advantages	Disadvantages
Pacific Biosciences	Iso-Seq protocol for transcripts up to 10Kb, high base calling accuracy	High cost, large machines
Oxford Nanopore	Accurate quantitative data for short transcripts (< 700bp), portable, high yield	High errors rate affects assembling de novo transcripts, higher amount of cDNA input
10X Genomics	Low cost (integrated with short-read technology), barcoding for accurate isoform detection, low error rates	Extra preparation step (barcode), extra computational step

Transcriptomics with long read technologies