Sequencing Tools

SAGESA

Dr Taya Forde

Feb 25th, 2022

University of Glasgow

Contents

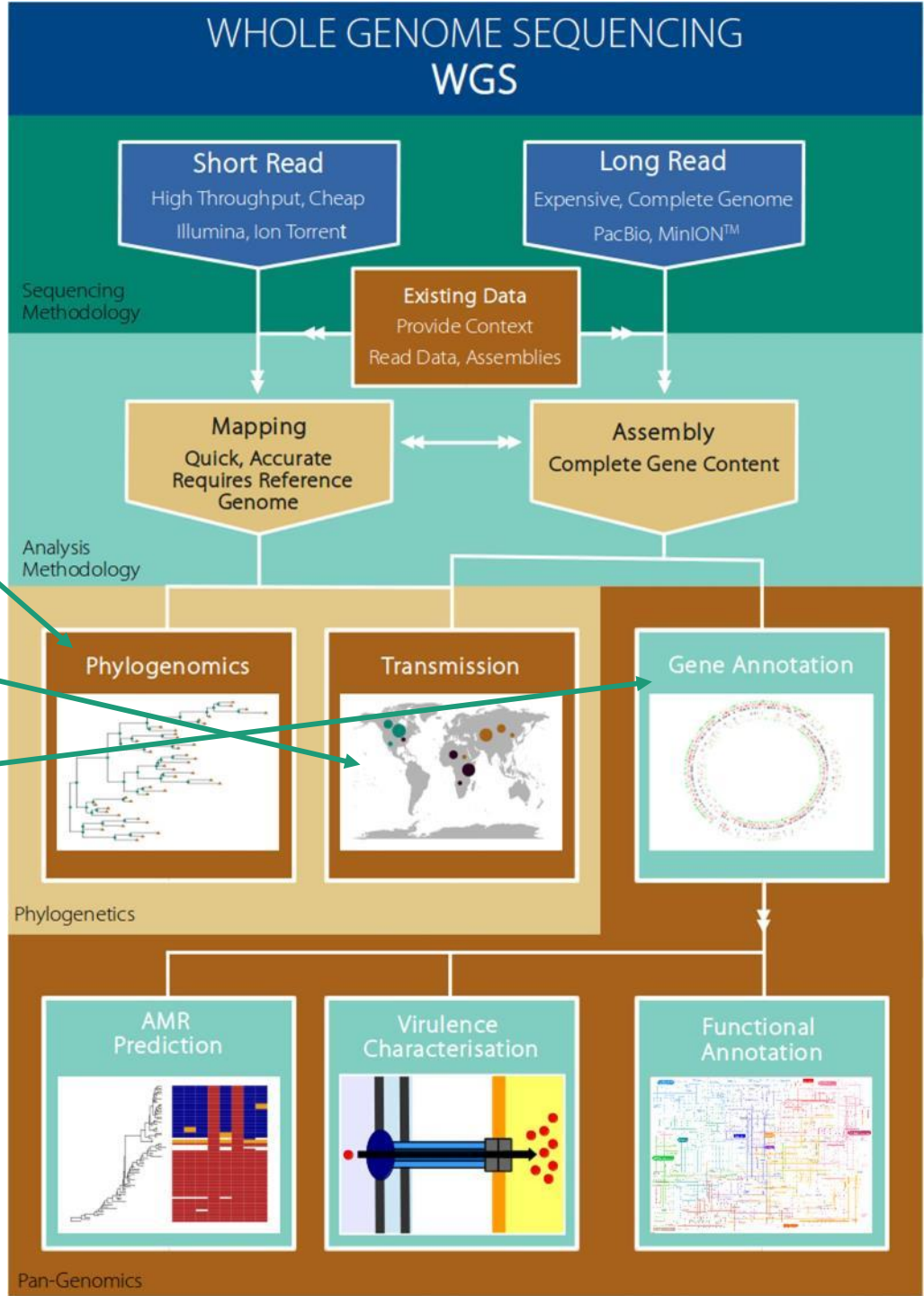
1. Preparing for sequencing
   - DNA extraction, quantification, library prep…
2. Different types of sequencing
   - Sanger, NextGen (Illumina), 3rd gen (MinION)
3. A few alternatives
   - Targeted capture
   - Metagenomics

Definitions

- **Sequencing**: determining the order of nucleotides in a DNA (or RNA) molecule
- **Whole Genome Sequencing (WGS)**: determining the 'complete' nucleic acid sequence of an organism
- **Genome**: ensemble of genetic material of an organism
  - For viruses, may be DNA or RNA
  - For bacteria, may include plasmids in addition to chromosomal DNA
  - For parasites, multiple diploid chromosomes
- **Genomics**: the branch of molecular biology concerned with the structure, function, evolution and mapping of genomes

Why do WGS?

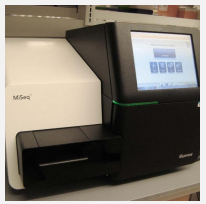
- Phylogenomics: Determine genetic relationships
- Molecular epidemiology: understand transmission patterns at various scales
- Determine what genes are present: functional and phenotypic predictions



Bayliss *et al.* Front. Microbiol. 2017;8:121.

The ‘evolution’ of sequencing technology

First Generation (Sanger)

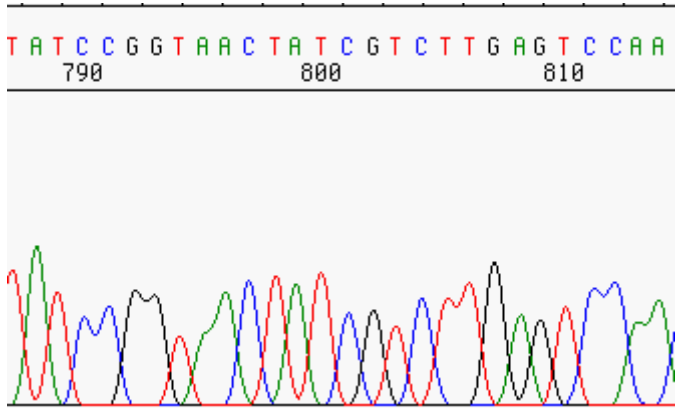


Next Generation (NGS)

Third Generation
(single molecule)

Preparing for Sanger sequencing

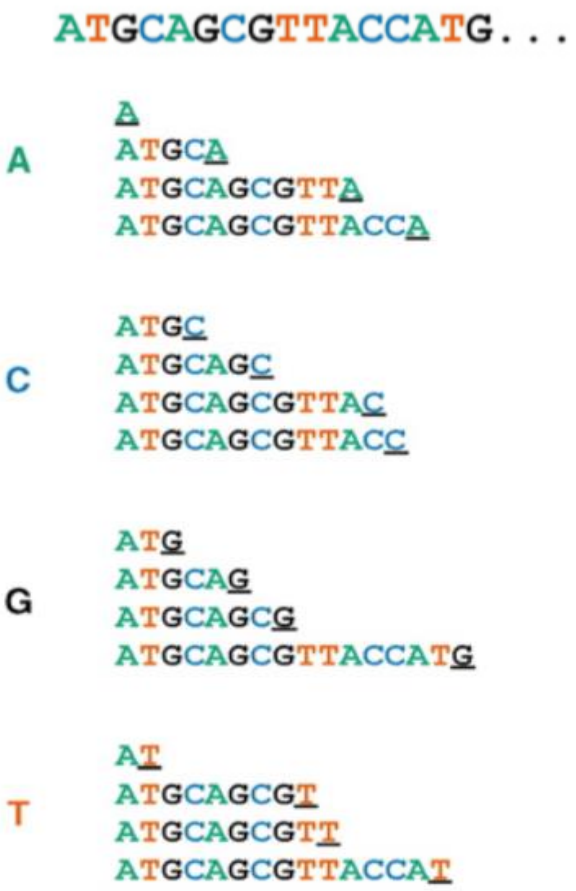
1. Nucleic acid extraction
2. Nucleic acid amplification (PCR)
3. Quality checks

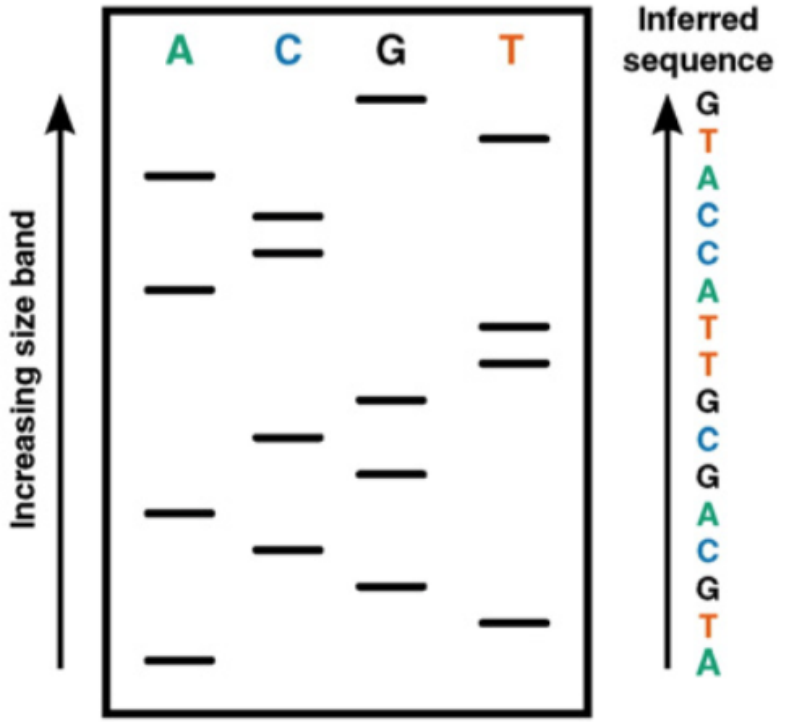


https://www.youtube.com/watch?v=ONGdehkB8jU

First Generation Sequencing: Sanger

- Also known as ‘standard’ sequencing, sequencing by chain termination, capillary sequencing



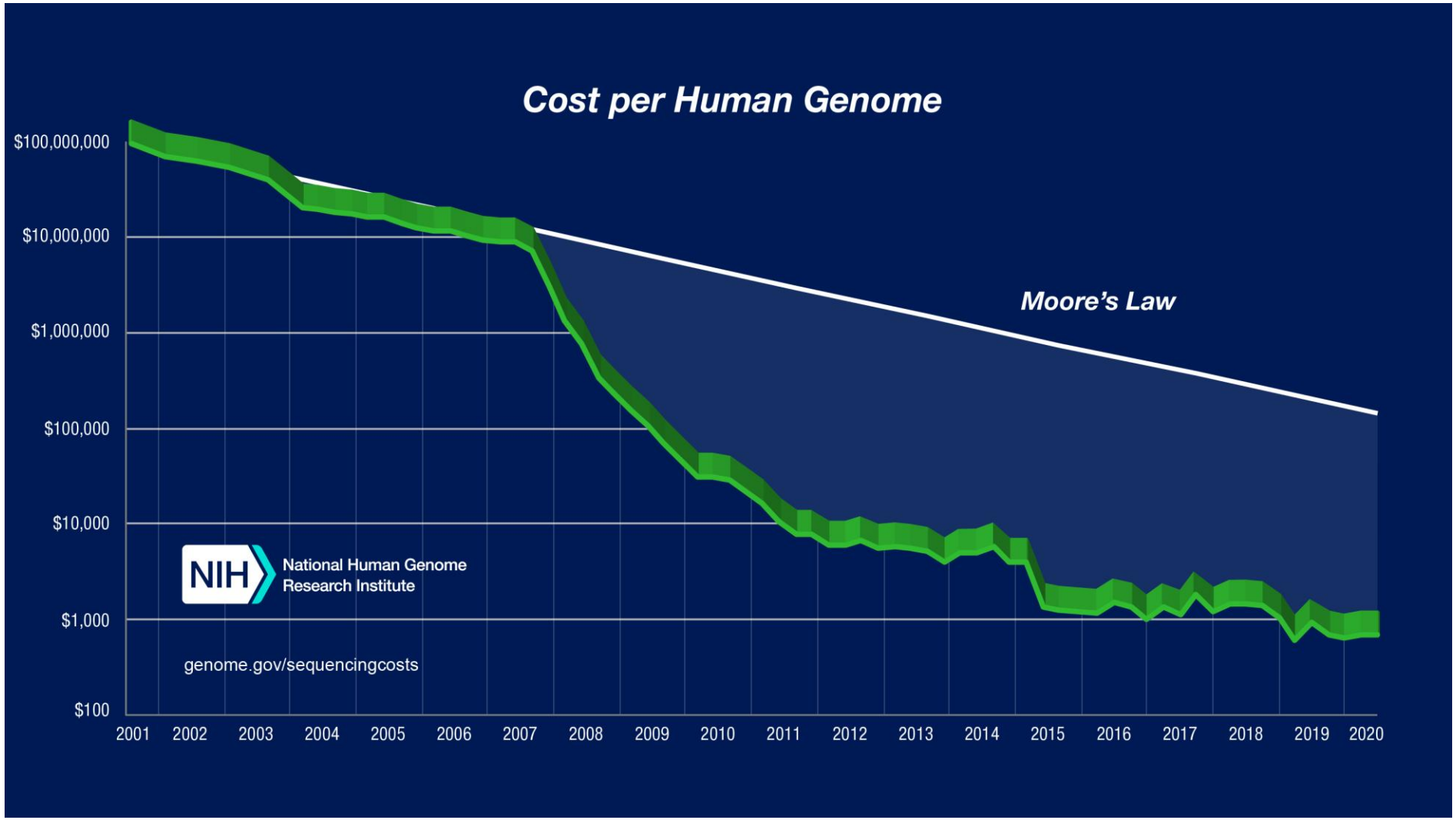


Heather and Chain, Genomics, 2016

‘Next Generation’ Sequencing (NGS): Basic Principles

- aka ‘Second generation’, ‘massively parallel’, ‘deep sequencing’, ‘high-throughput’
- Cell-free system (doesn’t rely on bacterial cloning of DNA fragments)
- Template is attached or immobilized to a solid surface or support for amplification (different varieties)
- Many reactions in parallel (thousands to millions!) = inexpensive, high throughput

The decreasing cost of genome sequencing



By National Human Genome Research Institute (NHGRI) Web site genome.gov is in the public domain. - https://www.genome.gov/sequencingcosts/, CC0, https://commons.wikimedia.org/w/index.php?curid=30648381

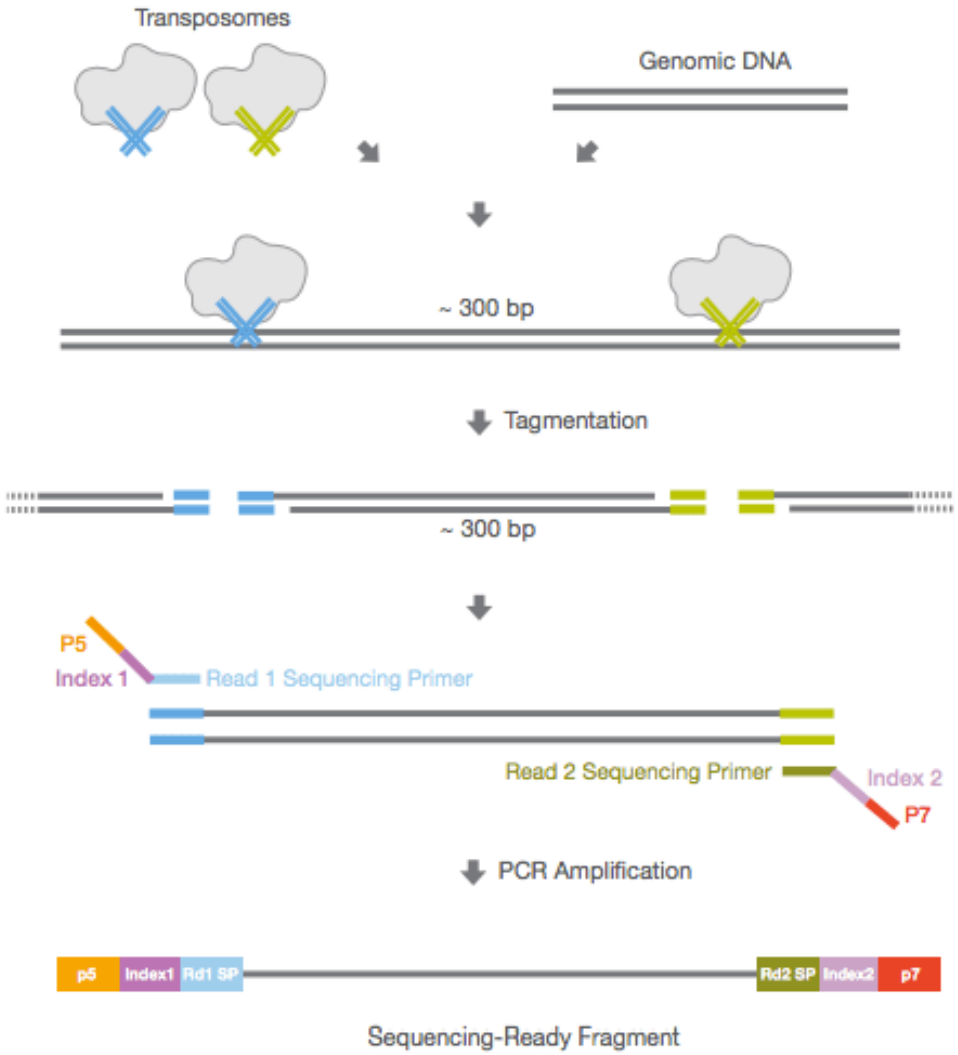
Preparing for Next Generation WGS

1. Culture bacterial isolate
2. Nucleic acid extraction
3. Nucleic acid quantification
4. Library preparation

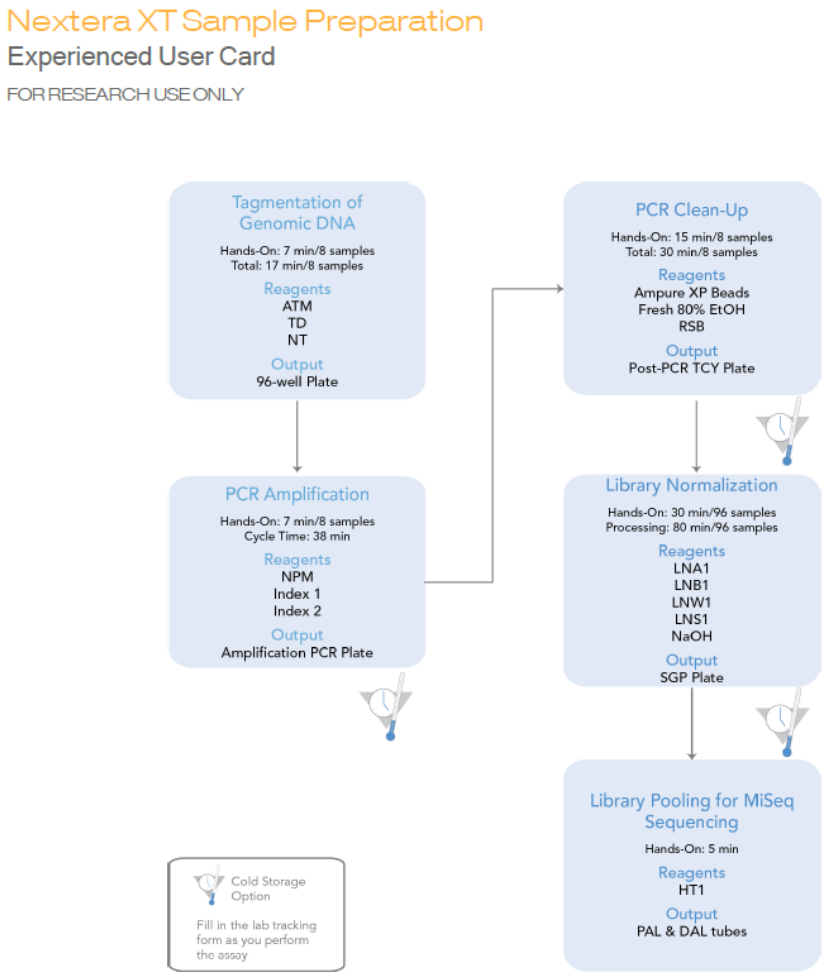
Library preparation example: Nextera XT (Illumina)

- "Tagmentation" (DNA fragmented and adaptors added)

- Addition of sample-specific barcodes (indices) through PCR amplification
  - Each sample has unique pair of p5 and p7 indices = pool multiple samples in single run



Illumina Library Preparation Workflow

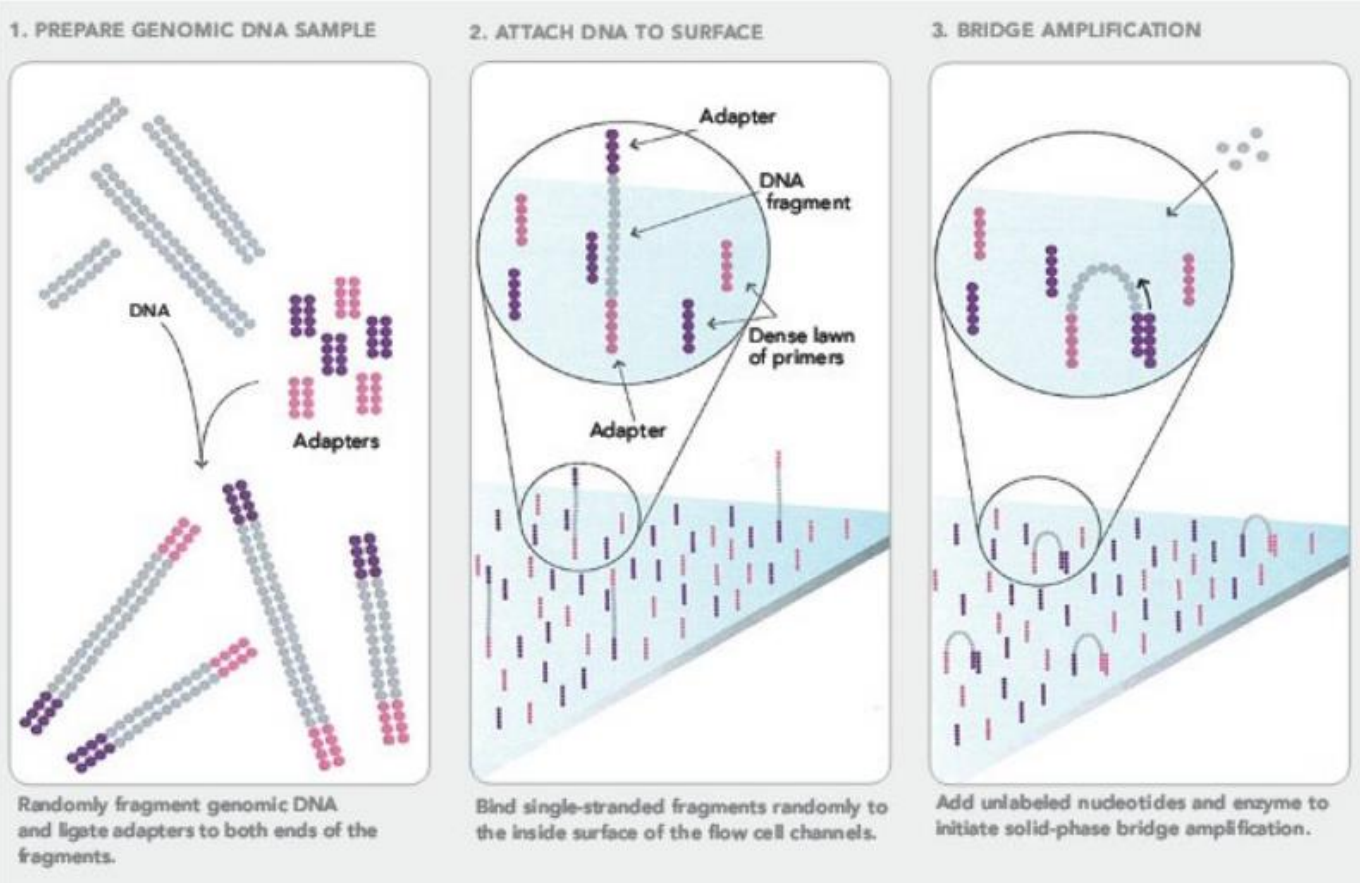


- *Tagmentation*
- *PCR amplification (add barcodes)*
- PCR Clean-up
- Library Normalization
- Library Pooling

https://www.youtube.com/watch?v=9YxExTSwgPM

Next Generation Sequencing in detail: Illumina Sequencing (NextSeq/HiSeq/MiSeq) as an example

1. Illumina- solid phase amplification

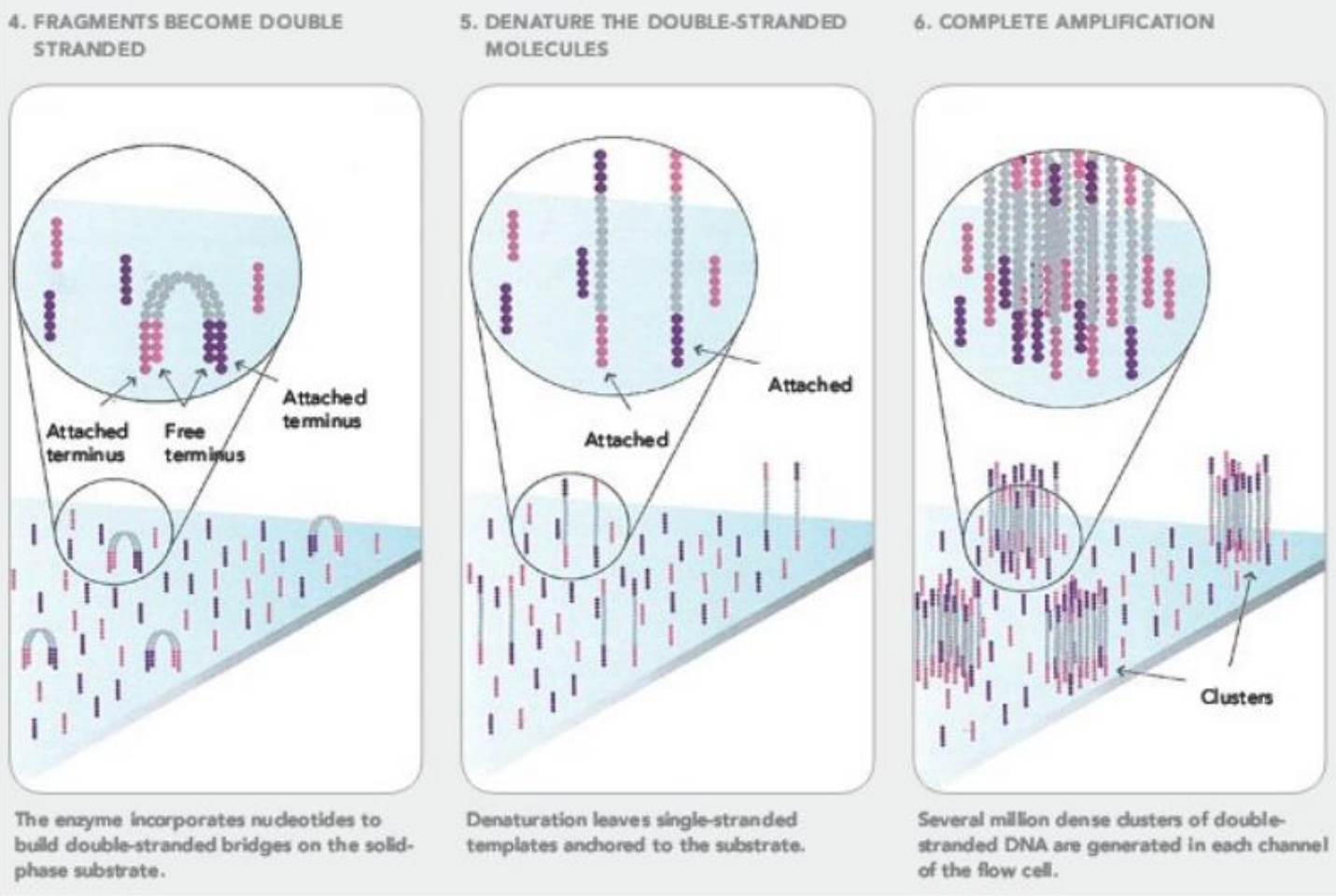


1. Prepare genomic DNA sample by fragmenting DNA and ligating adapters
2. Attach DNA to surface of flow cell based on primers complementary to those on the adapter sequence
3. Bridge amplification through the addition of unlabelled nucleotides

http://seqanswers.com

Tech Summary: Illumina's Solexa Sequencing Technology

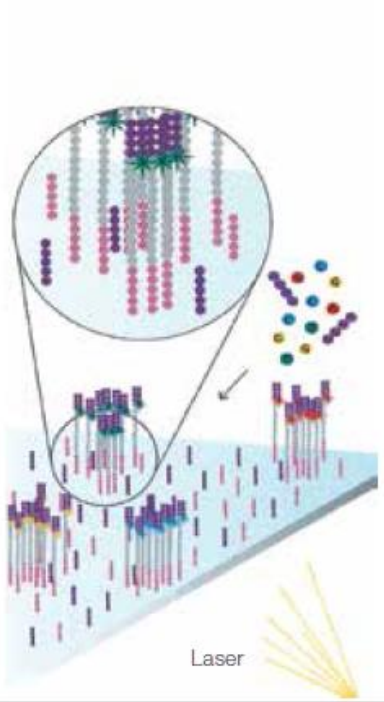
Illumina- solid phase amplification (cntd.)



4. Fragments become double stranded
5. Double stranded DNA molecules are denatured
6. Amplification is continued to generate clusters

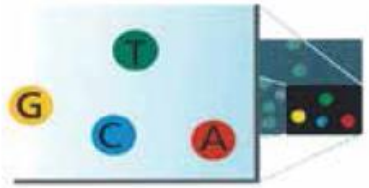
2. Illumina- base determination with reversibly terminated dNTPs and fluorescent capture

Figure 8: Determine First Base



The first sequencing cycle begins by adding four labeled reversible terminators, primers, and DNA polymerase.

Figure 9: Image First Base



After laser excitation, the emitted fluorescence from each cluster is captured and the first base is identified.

- Sequencing begins by adding labelled reversable terminators, primers and polymerase
- Bases from each cluster are imaged after each nucleotide addition

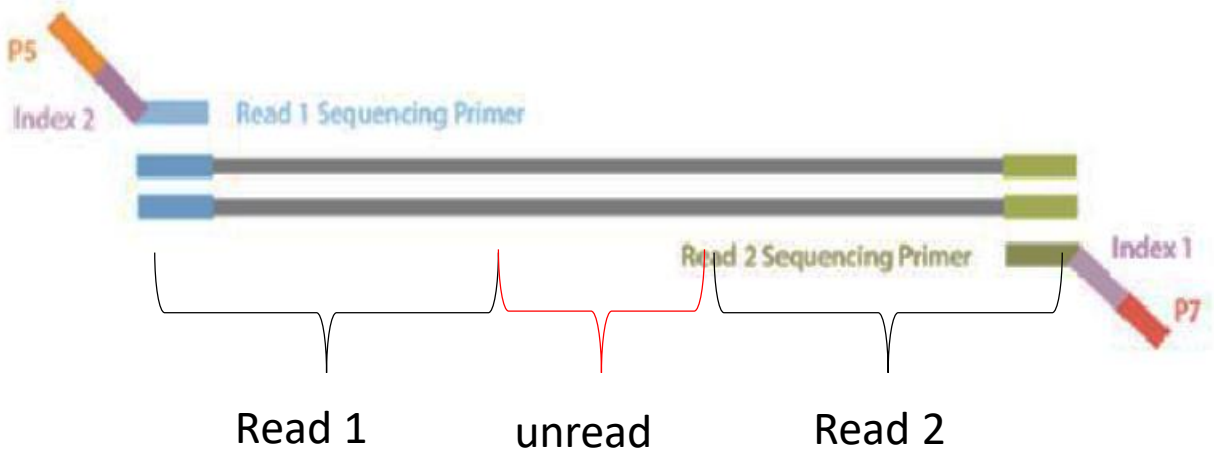
‘Next Generation’ Sequencing Platforms: A Comparison

	Roche 454 *pyrosequencing*	**SOLiD**	**Ion Torrent** *semi-conductor*	**Illumina** *Next/Hi/Mi-Seq*
amplification	bead-support; oil emersion (“emulsion”)	bead-support; oil emersion (“emulsion”)	bead-support; oil emersion (“emulsion”)	bridge amplification
sequencing	by synthesis	by ligation	by synthesis	by synthesis
detection	light intensity, corresponding to number of bases incorporated	fluorescence emission	ion sensors capture proton release during nucleotide incorporation	fluorescence emission

- Platforms vary in methods used to amplify and ‘read’ the sequence

Paired-end reads

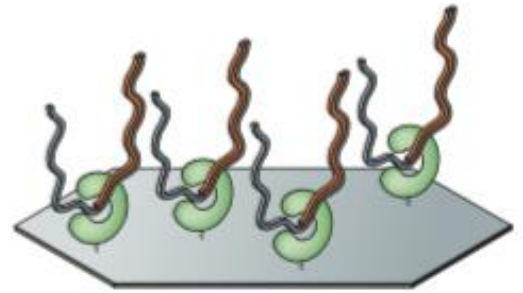
- End of each fragment (~200-800 bp) sequenced from both directions
- Read length depends on kit (now often 250-300 bp)
- Segment in middle may be ‘unread’, but we know the two reads are ‘linked’



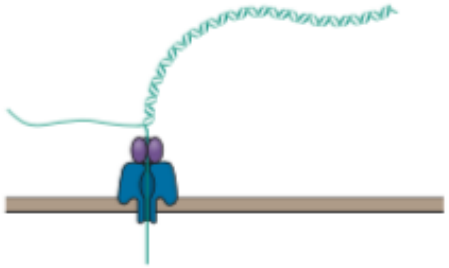
Third Generation sequencing platforms

https://www.youtube.com/watch?v=GUb1TZvMWsw

- Single molecule template
- Sequencing occurs in real time
- Pac-Bio (SMRT): polymerase immobilized
  - Sequencing by synthesis
- Oxford Nanopore (minION)
  - Disruption of electrical current



Pac Bio SMRT sequencing



Oxford Nanopore sequencing

- Increased read length
- Less accurate base calling

If you want to know more…

- **Heather and Chain (2016) The sequencing of sequencers: the history of sequencing DNA. Genomics 107(1):1-8.**
- Loman and Pallen (2015) Twenty years of bacterial genome sequencing. Nat Rev Microbiol 13(12):787-94.

Targeted capture

- Culture-free alternative
- Selective capture of DNA of interest from a complex sample (e.g. tissue, blood, swab)
- Could be applied to multiple genes of interest

Metagenomics

Sequencing of all contents in a sample (mixed community)

E.g. Microbiome, environmental samples

16S rRNA (targeted) vs. shotgun (unbiased)

+/- selective depletion/enrichment

Questions?