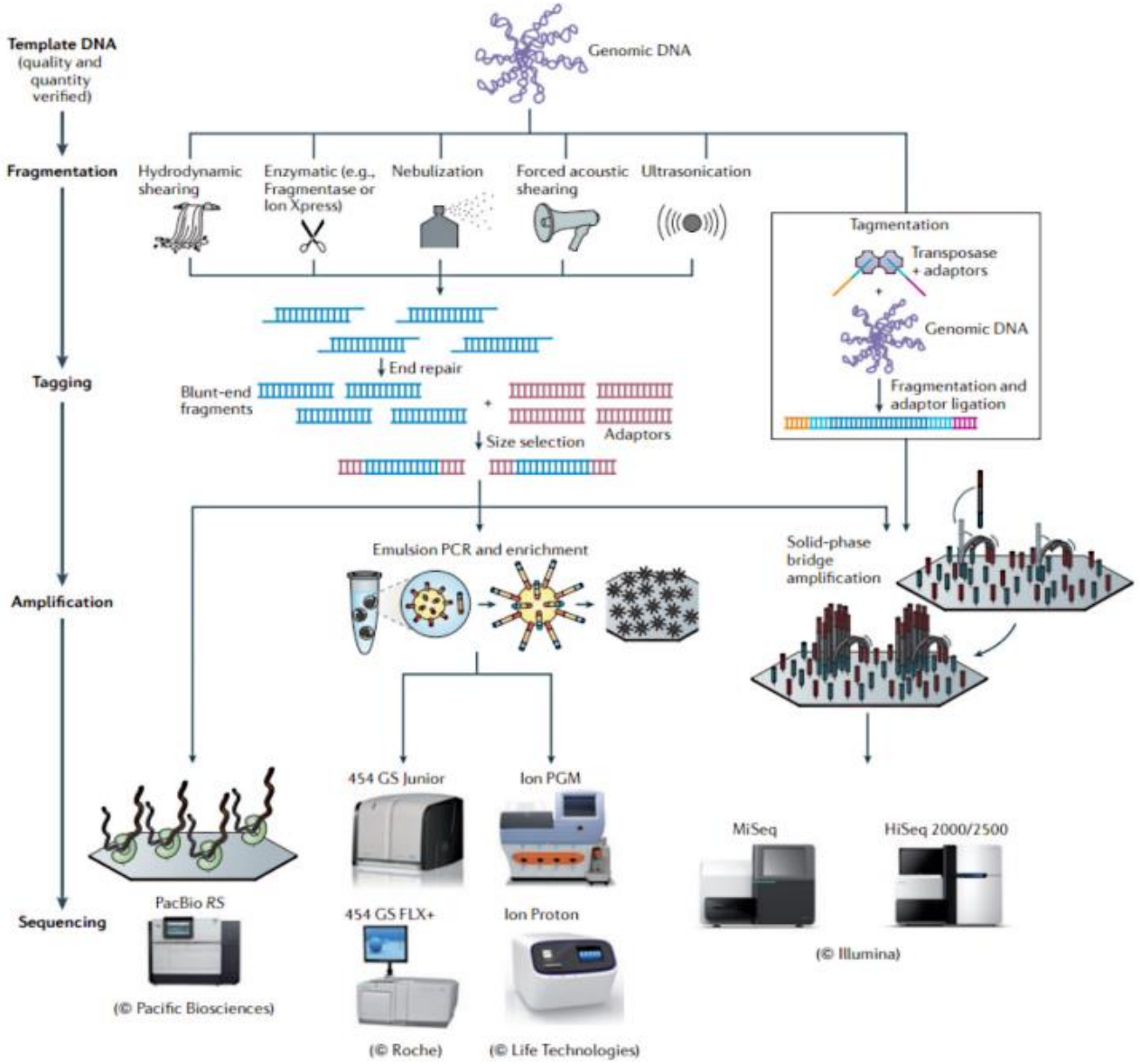
UK Health Security Agency

Short Read Sequencing Practical - 2025

Dr Jon Hubb
Head of Colindale Sequencing Laboratory – Genomics Service Delivery and Quality Division

Next Generation Sequencing methodologies

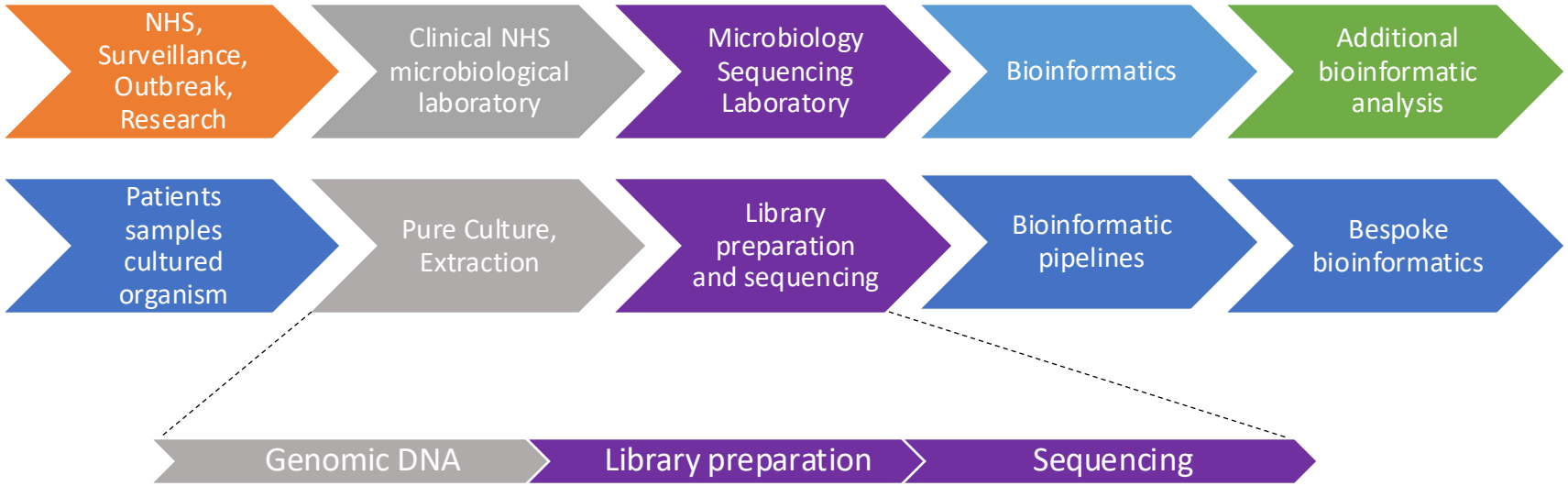
- Next Generation Sequencing
  - Catch all term
  - High throughput sequencing
  - Massively parallel sequencing
  - Terabytes of data
- Differing sequencing chemistries and engineering principles
- Sequencing of spatially separated, clonally amplified DNA.



American Academy of Microbiology, Applications of Clinical Microbial Next-Generation Sequencing, 2016

Patient sample pathways

- HCAI practical sessions:
  - Illumina short read sequencing (today)
  - Oxford Nanopore technologies long read sequencing (tomorrow)



Preparation of Genomic DNA

Genomic DNA | Library preparation | Sequencing

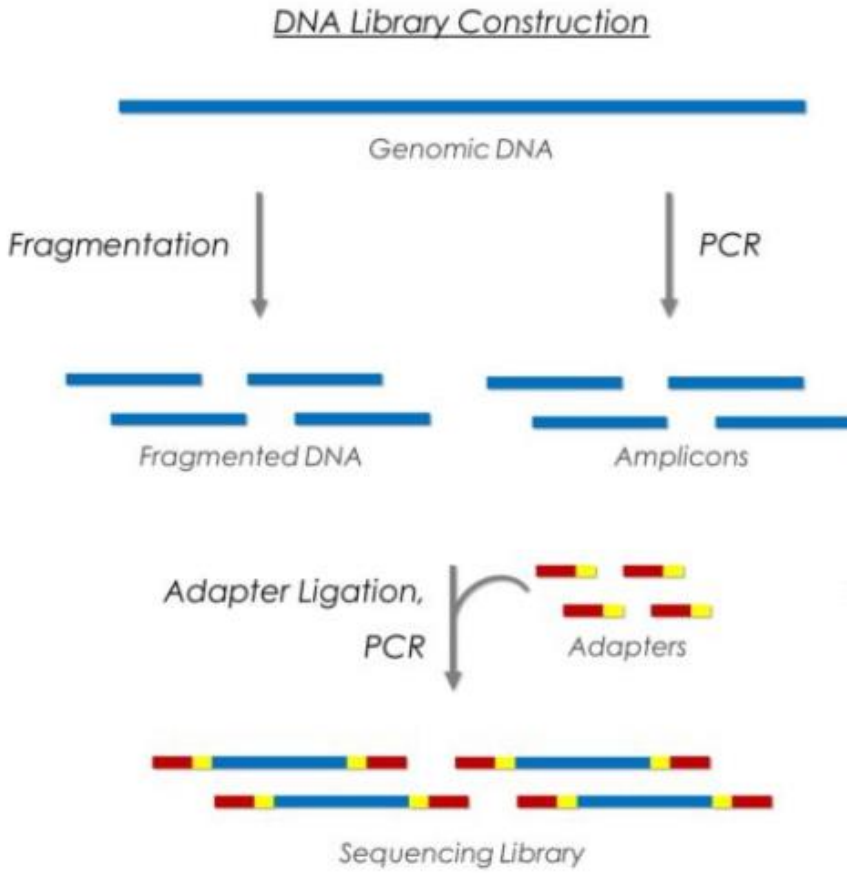
- Purification of nucleic acid (genomic DNA and extra chromosomal)
- Extraction methods vary
  - One solution may not fit all
  - Often require pre-extraction preparation
- To name a few suppliers:
  - Qiagen - Qiasymphony or EZ1
  - Thermo - Kingfisher
  - Perkin Elmer - Chemagic
  - Roche - MagnaPure
- Which platforms is best?
  - Batch size
  - Sensitivity
  - Pre-extraction preparation
  - Cost
  - Run time
  - Spatial requirements
  - Established gold standard

Library preparation

Library Preparation (short read sequencing)

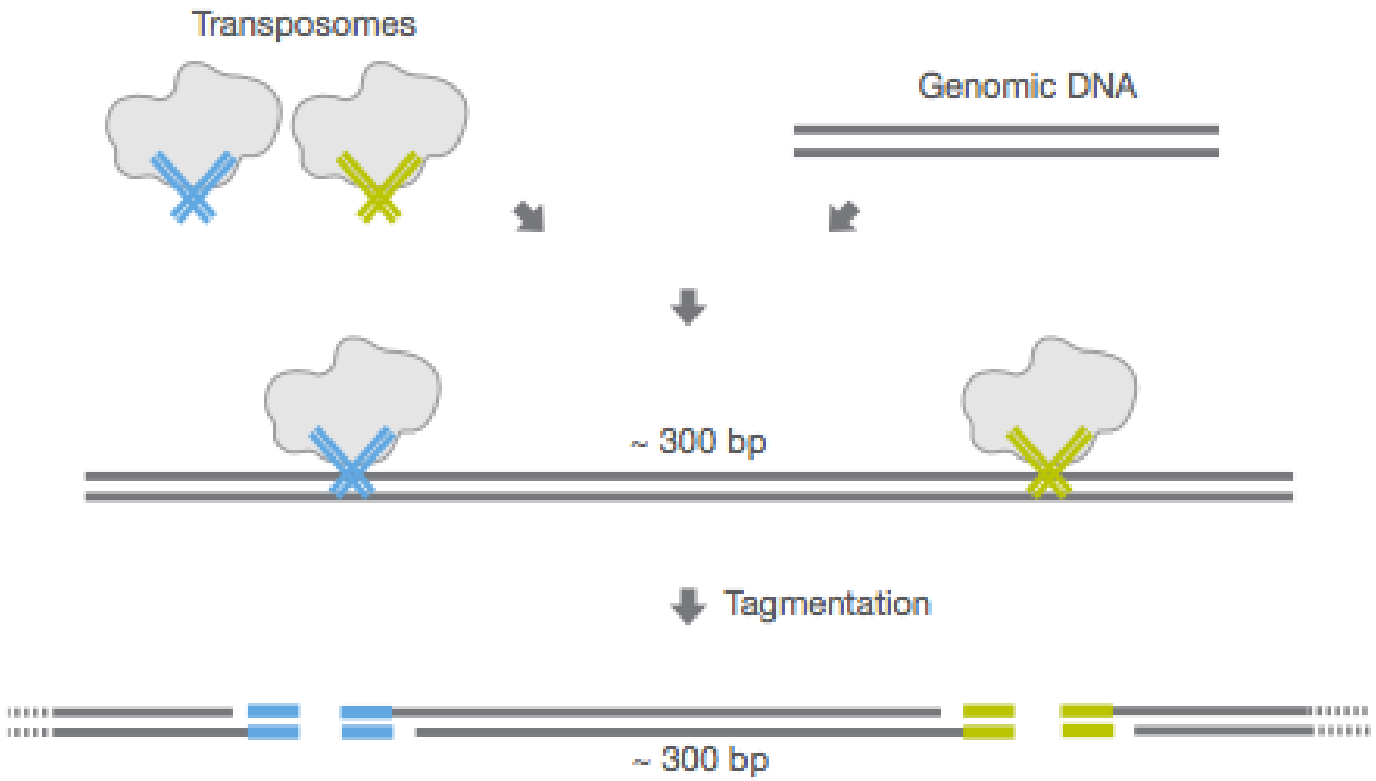
Genomic DNA | Library preparation | Sequencing

- Preparing the nucleic acid for sequencing
- DNA or RNA starting material
  - RNA needs to be reverse transcribed in cDNA
- Library preparation process
  - Fragmentation
  - End repair
  - dA-tailing 3'overhang
  - Ligation of adapter
  - PCR enrichment
  - Clean-up
  - Normalise
- Many commercially available kits for this process



Illumina Nextera XT Method

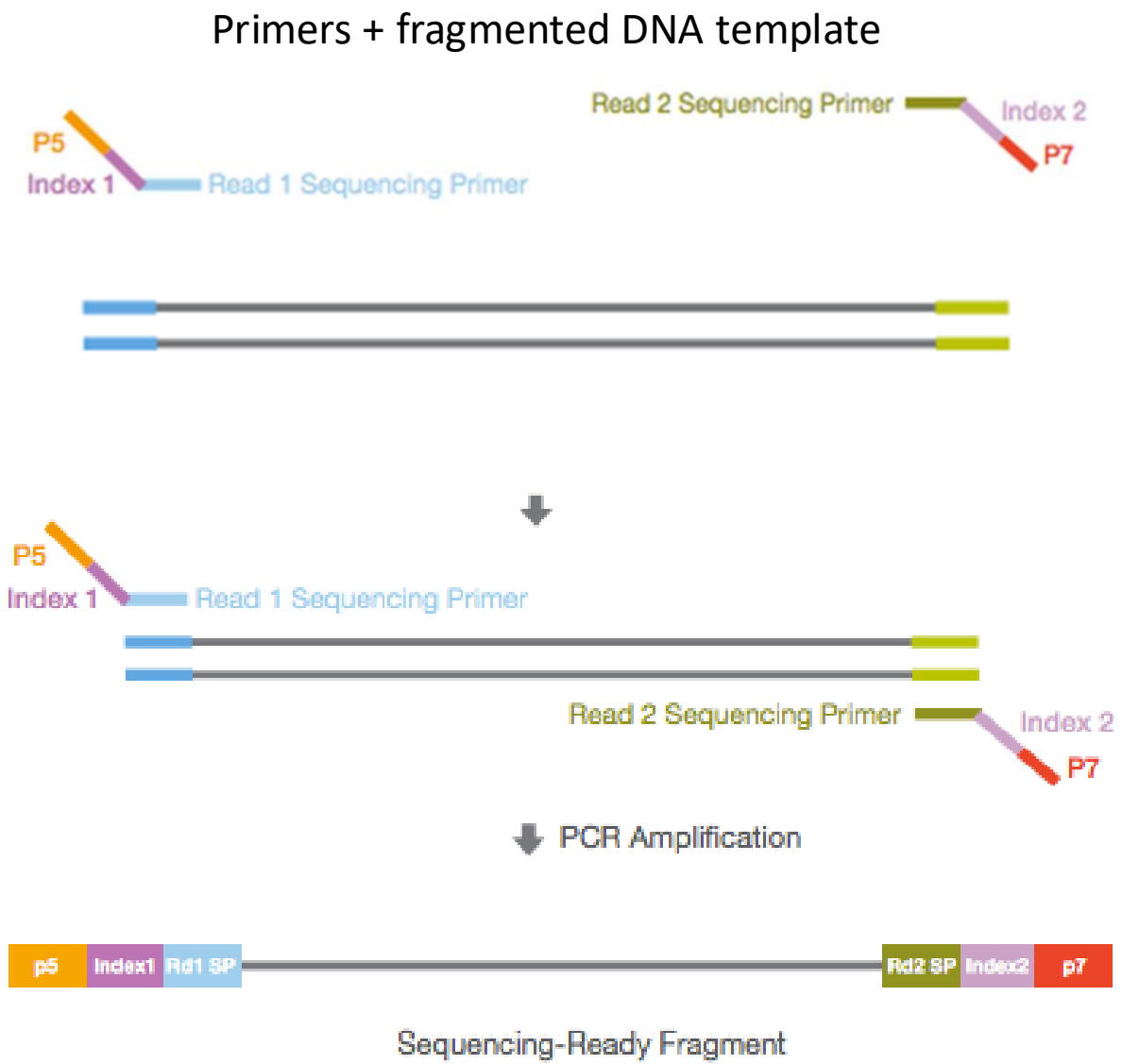
- **Tagmentation (Section 2.0)**
  - Starting material: Diluted Genomic or PCR amplified DNA (0.2ng/uL)
  - Modification of DNA using transposomes enzymes
  - Fragments DNA through enzymatic cleavage with addition of short adapter sequences that correspond to the sequencing primer sites
  - Mix transposome, buffer and DNA and incubate for 5mins at 55℃
    - Reaction stopped by adding stopping buffer



Illumina Nextera XT Method

PCR Amplification (Section 3.0)

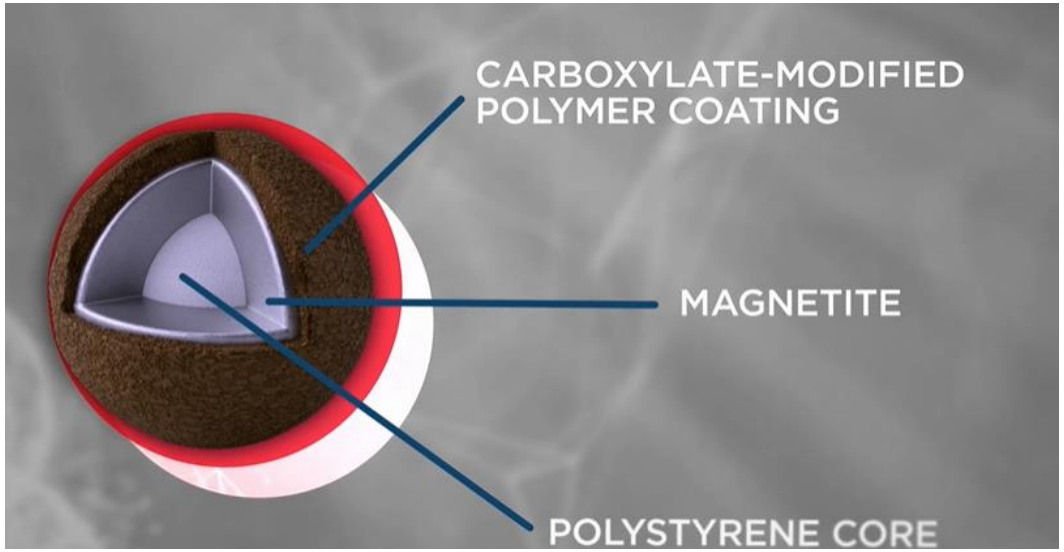
- PCR amplification of library
  - Incorporates terminal modification library with the incorporation of additional sequencer
- PCR primers are tripartite sequence structure
  - Complimentary adapter sequence: correspond to sequence primer binding site
  - Index: Adds a unique genetic tag onto the library for each samples eg.
    - UDP0001
      - i7 CGCTCAGTTC
      - i5 TCGTGGAGCG
  - Capture sequences: Perform PCR o modified DNA to include indices and P5 and P7 sequences

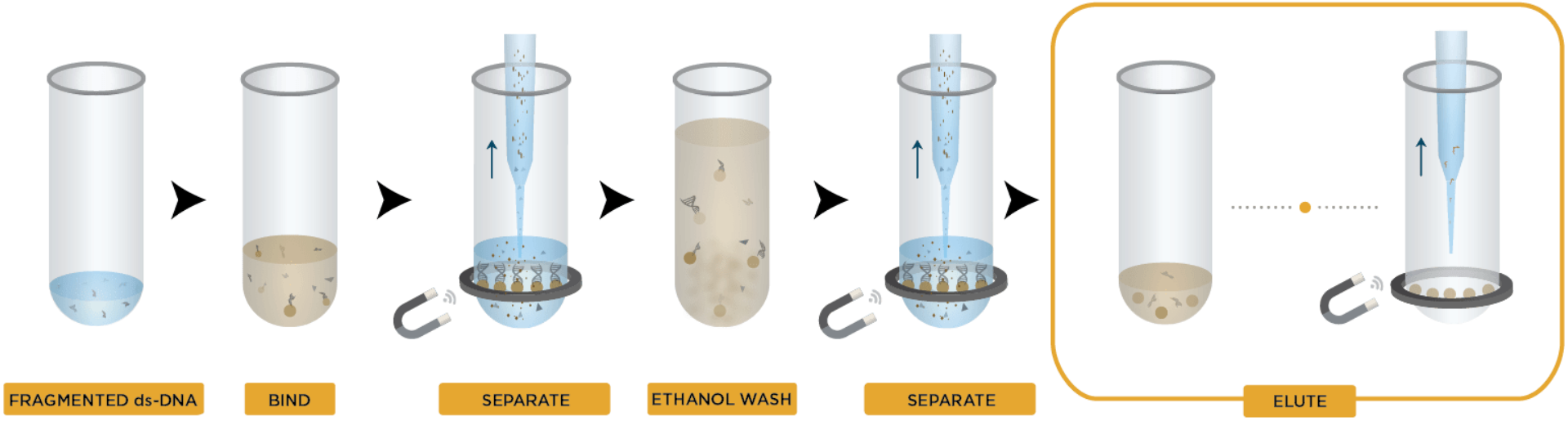


Illumina Nextera XT Method

PCR purification (Section 4.0)

- Purification of dsDNA using AMPure XP bead-based reagent
- Negatively charge DNA binds beads
- Purification of beads using magnet
- Wash impuritites
- Low salt elution buffer dissociates DNA



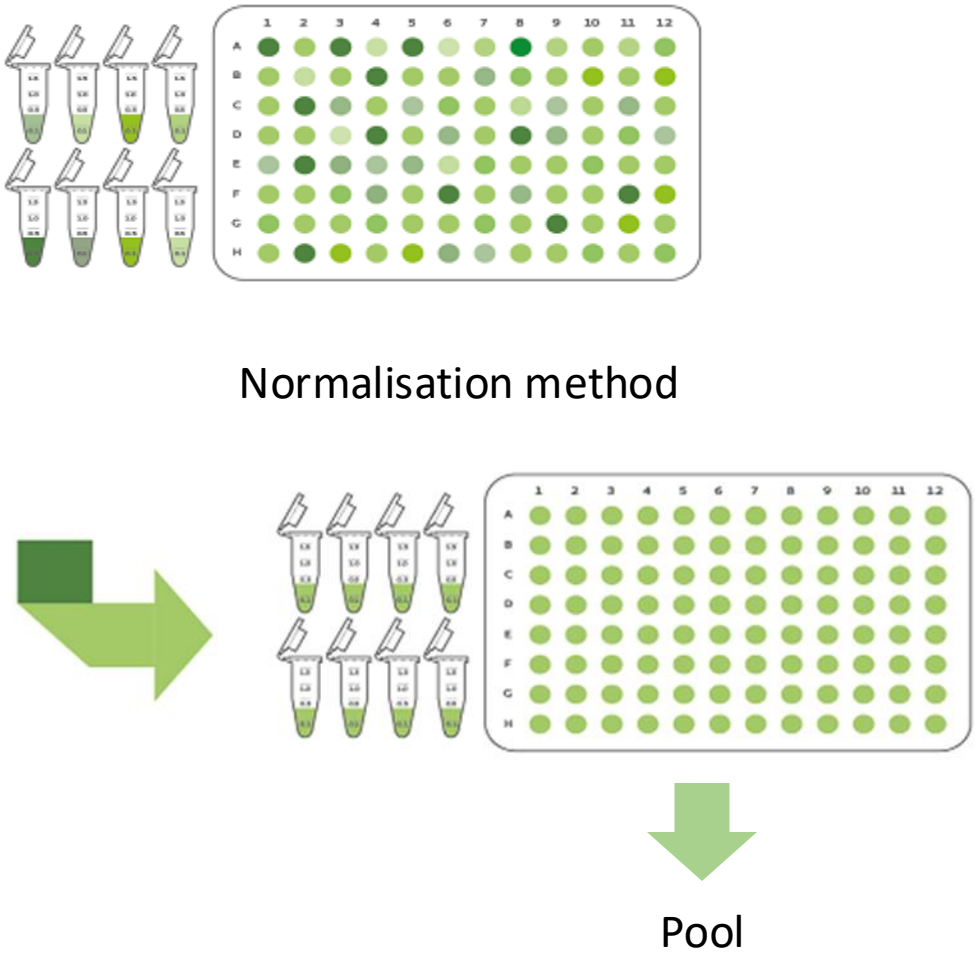


Illumina Nextera XT Method

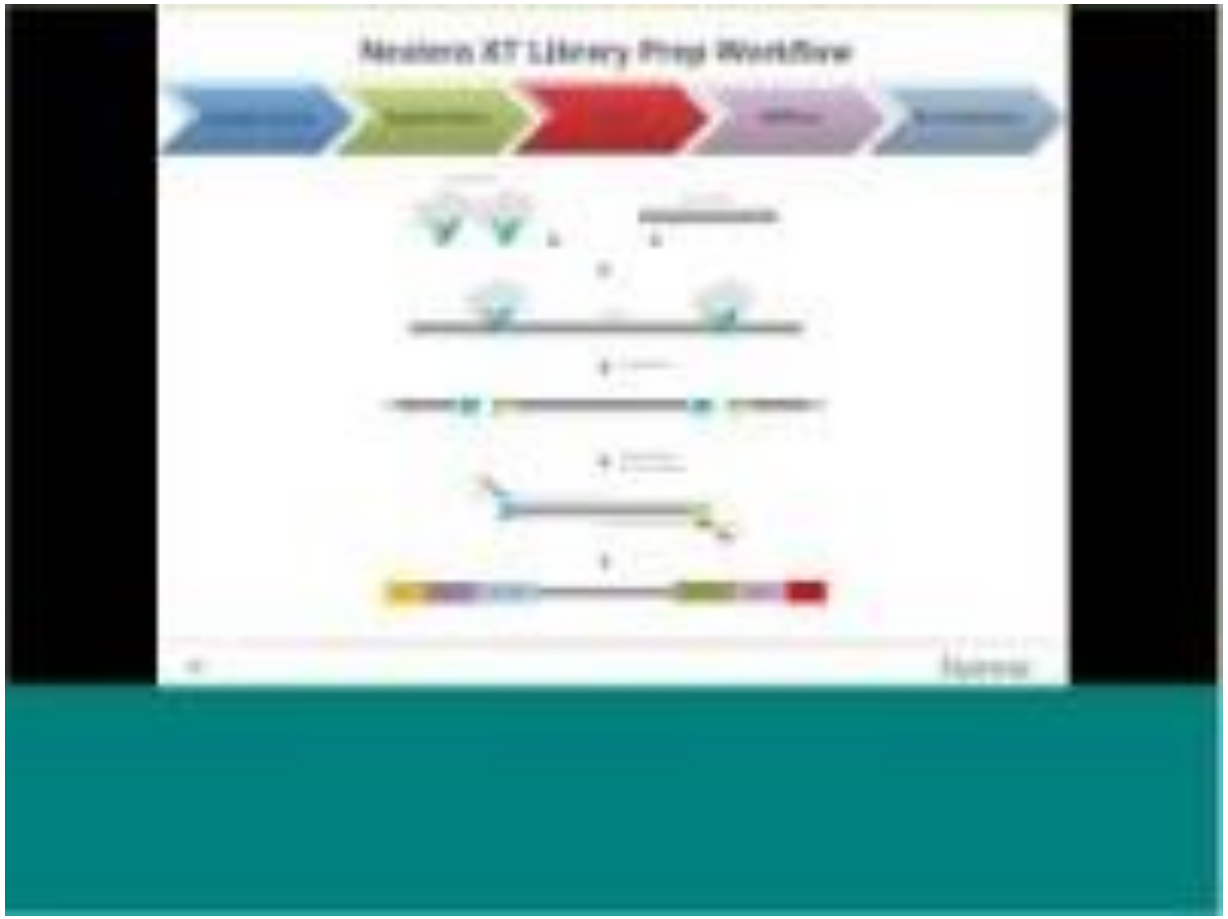
Normalisation (Section 5.0)

- Bringing all individual libraries in plate to same molarity prior to sequencing

- 4 main methods
  - Bead based:
    - Illumina library preparation (Nextera XT and Illumina DNA preparation)
  - Manual
    - Determine molarity using concentration of library and library size
  - Real time PCR
    - KAPA real time SYBR green method
    - Technically challenging
  - Enzymatic – Normalase
    - Enzymatic method
    - 2-step normalase procedure
    - Rapid, simple method



Nextera XT library preparation method



SPRI beads

Illumina Sequencing

NGS sequencing technologies

Genomic DNA → Library preparation → Sequencing

454

	GS Jr.	GS FLX+
Amplification Method	Emulsion PCR on beads	
Chemistry	Synthesis (pyrosequencing)	
Read length (bp)	400	700
Yield/run (Gb)	0.05	0.9
Primary Error	Indel	
Error rate	~1%	
Run time (hours)	10	20
Virus-related Publications	187	
Advantage(s)	Long reads, maturity	
Disadvantage(s)	Homopolymer misreads, high cost/Mb	

Illumina

	MiSeq	HiSeq
Amplification Method	Bridge PCR in situ	
Chemistry	Synthesis (reversible termination)	
Read length (bp)	250	125
Yield/run (Gb)	8	1,000
Primary Error	Substitution	
Error rate	~0.1%	
Run Time (hours)	39	276
Virus-related Publications	129	
Advantage(s)	Easy work flow, maturity	
Disadvantage(s)	Shortest reads, long run	

Ion Torrent

	PGM	Proton
Amplification Method	Emulsion PCR on beads	
Chemistry	Synthesis (H^+ detection)	
Read length (bp)	400	200
Yield/run (Gb)	2	10
Primary Error	Indel	
Error rate	~1%	
Run time (hours)	7	4
Virus-related Publications	13	
Advantage(s)	Low cost, fast run	
Disadvantage(s)	Homopolymer misreads	

PacBio

	RS II
Amplification Method	No PCR
Chemistry	Single-molecule real-time sequencing
Read length (bp)	8,500
Yield/run (Gb)	0.15
Primary Error	Indel
Error rate	~13%
Run time (hours)	2
Virus-related Publications	6
Advantage(s)	Longest reads
Disadvantage(s)	High error rate, expensive

Illumina Sequencers

Genomic DNA | Library preparation | Sequencing

- MiniSeq, MiSeq, NextSeq (550/1000), HiSeq 2500, NovaSeq 6000

- Related methods of sequencing
  - Subtle difference in clustering and sequencing chemistries
  - Vary in amount of data and run time considerably

Which sequencer or kit?

Genomic DNA | Library preparation | Sequencing

- Overwhelming amount of Illumina sequencers and kits available
- Key questions:
  - What are you trying to do?
  - How big is your organism?
  - What coverage is required?
  - What's my budget?

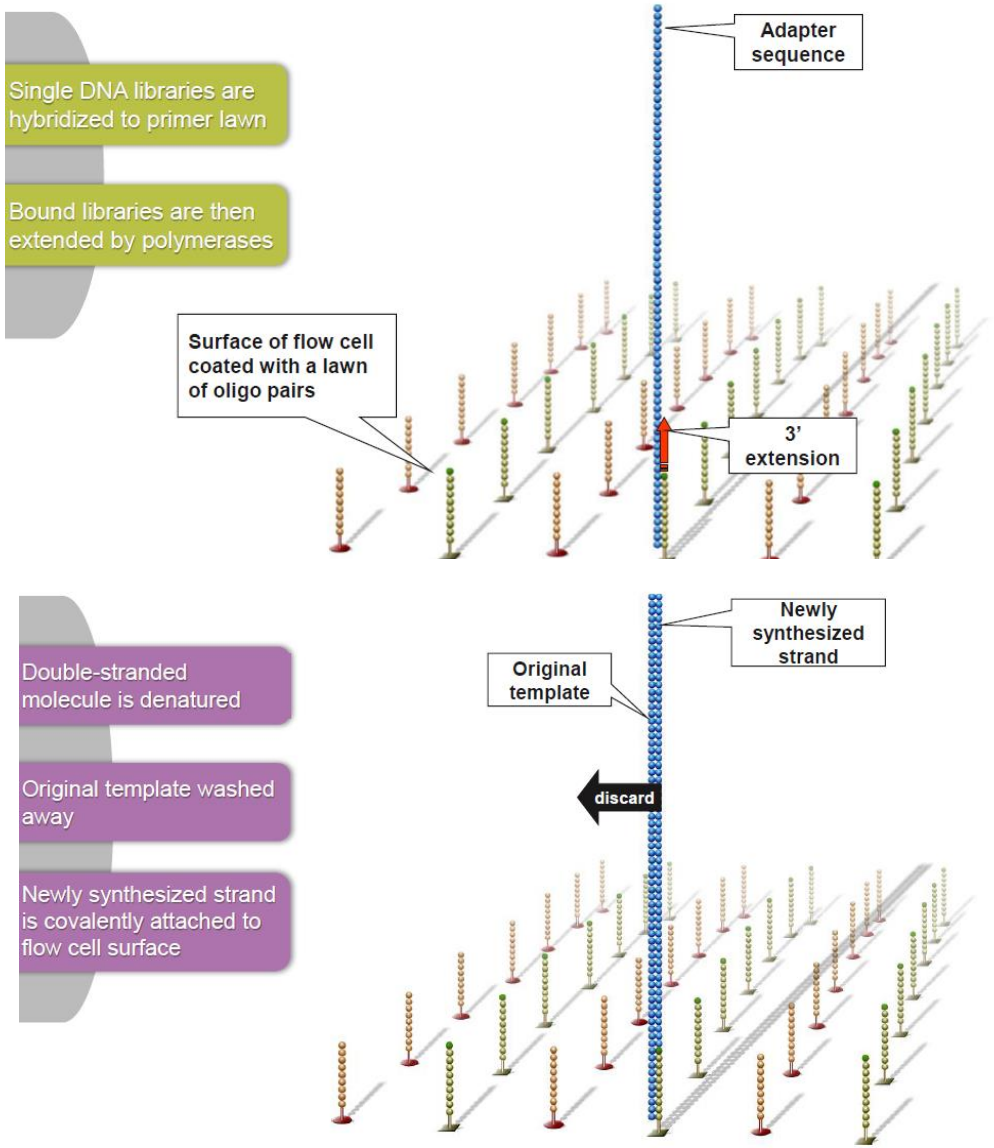
	 iSeq 100*	 MiniSeq*	 MiSeq* †	 NextSeq 1000 & 2000*
Output Range	144 Mb - 1.2 Gb	1.65–7.5 Gb	0.3–15 Gb	30-360 Gb***
Run Time	9–19 hours	4-24 hours	5–55 hr	11-48 hours
Reads Per Run	4 million	7–25 million	1–25 million	100 million–1.2 billion***
Max Read Length	2 × 150 bp	2 × 150 bp	2 × 300 bp	2 × 150 bp
Samples Per Run§	1-8	50	1–384	30-90
Relative Price Per Sample§	Higher Cost	Mid Cost	Mid Cost	Low Cost
Instrument Price	Lowest Cost	Low Cost	Low Cost	Mid Cost

https://emea.illumina.com/systems/sequencing-platforms/comparison-tool.html#/research-use-only/microbiology/small-whole-genome-sequencing
https://emea.support.illumina.com/downloads/sequencing_coverage_calculator.html

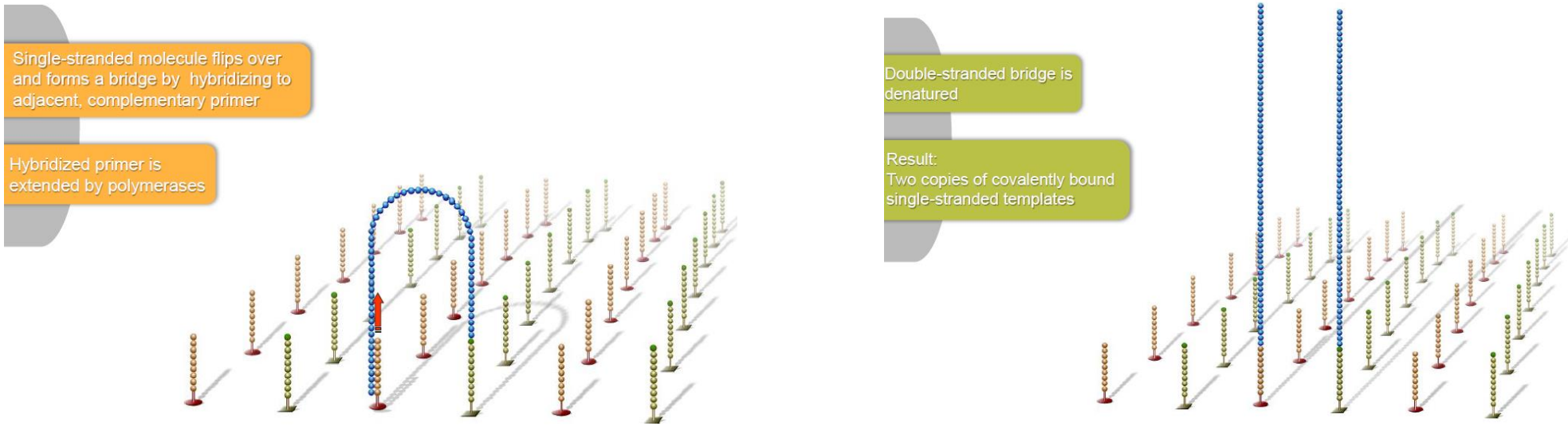
Cluster Generation

Genomic DNA | Library preparation | Sequencing

- Flow cell has P5 and P7 oligonucleotide lawn
  - Flow cell has single stranded complementary sequences
- P5 and P7 sequences on library termini bind to complementary sequences on flow cell
- Immobilised to discreet locations on flow cell



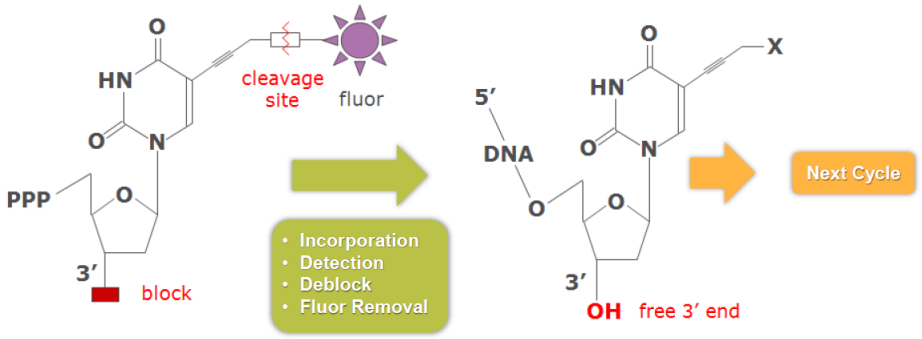
Cluster generation

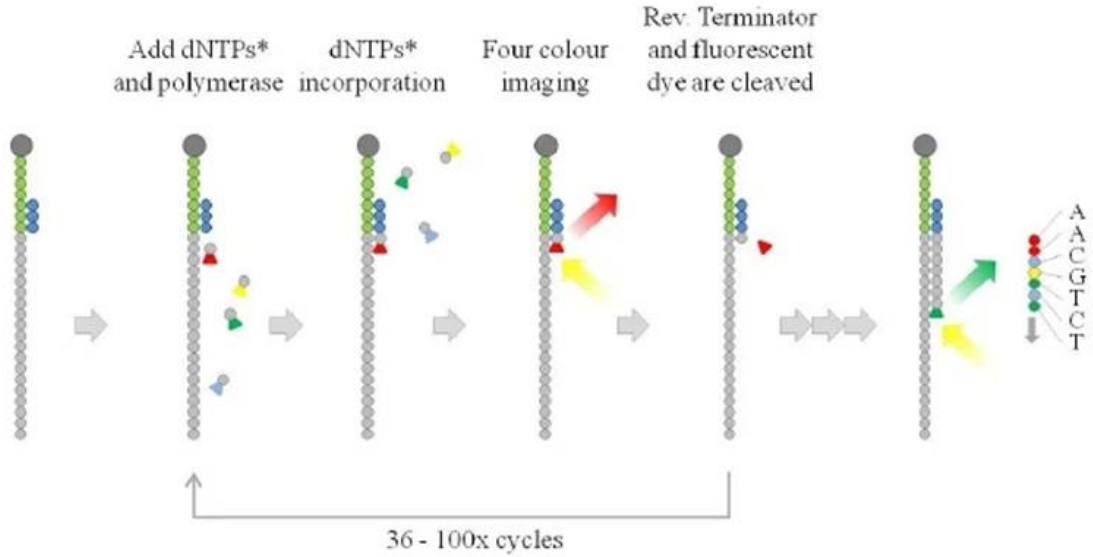


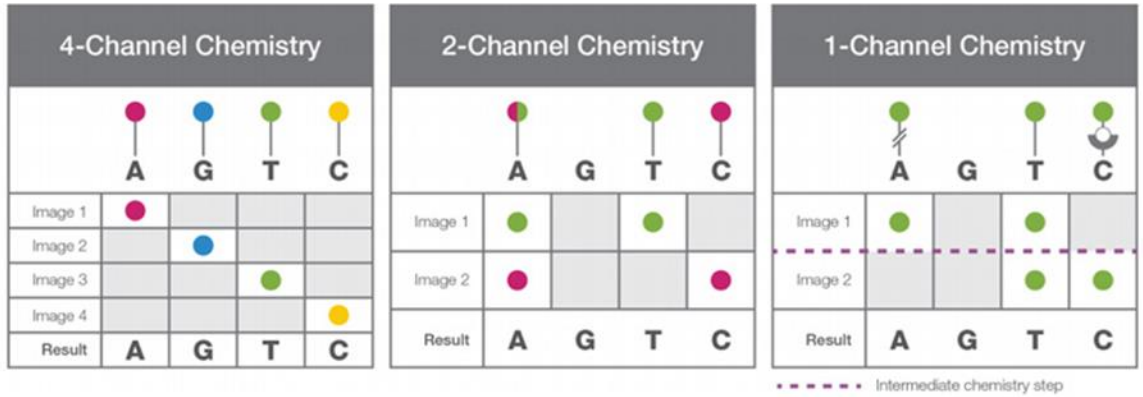
- Bridge amplification - Amplification of these single stranded fragments allows immobilised cluster generation
- Cluster generation can be random or more organised (patterned)
- Cluster generation - ~1000 clonal library copies per cluster

Sequencing By Synthesis

- Sequencing by synthesis
- All 4 nucleotides are included
- Labelled with a different fluorophore
- Contains a block on 3’ carbon
- Each base incorporation = 1 cycle

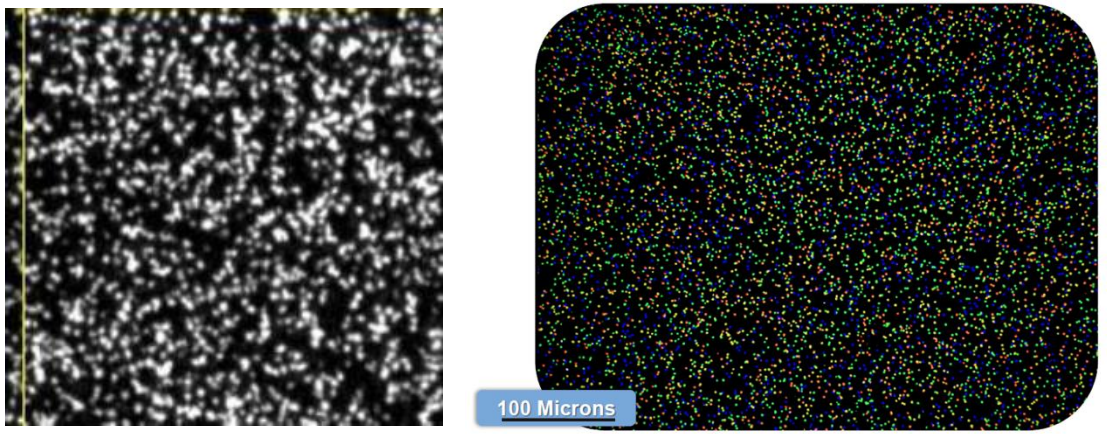


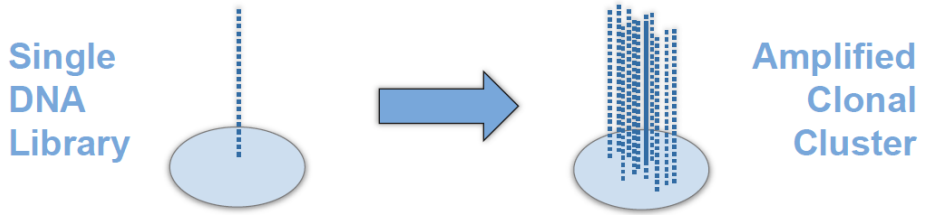




Sequence determination

- During sequencing emission of light to indicate which base has been incorporated
- Sequencer camera records the fluorescence emission from the flow cell
- Each cluster appears as pin-prick of light
- Cluster generation allows:
  - Amplification of library which improves signal acquisition
  - Reduces impact of phasing





Illumina