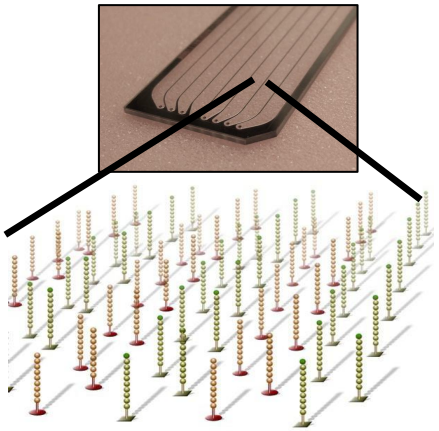
H3ABioNet
Pan African Bioinformatics Network for H3Africa

wellcome connecting science

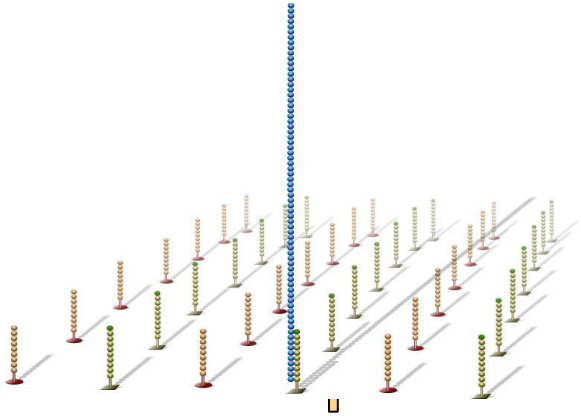
Next Generation Sequencing Bioinformatics Course 2021

Module 2: Session 3
Introduction to NGS
Illumina Sequencing

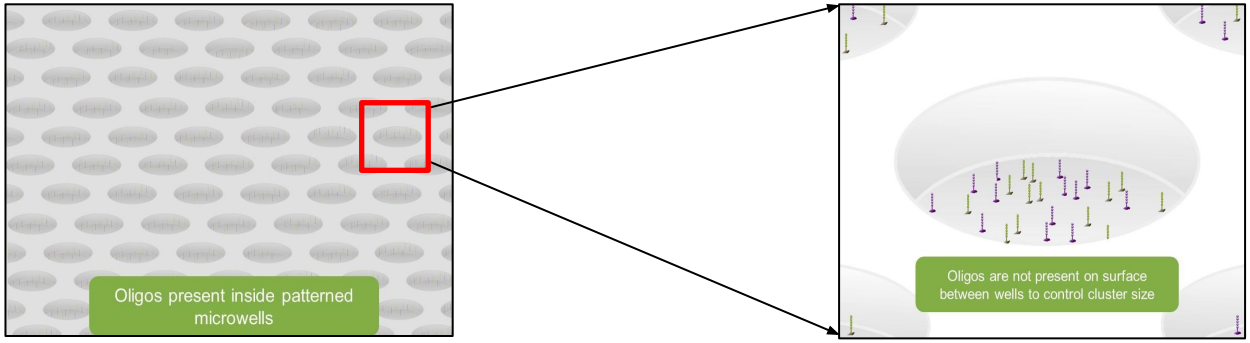
Illumina flow cells



Glass flow cell

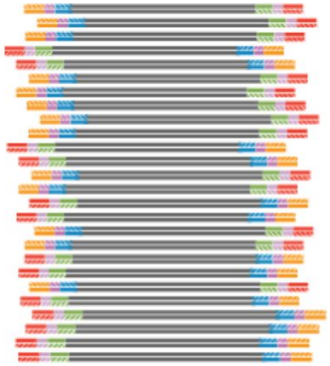


Glass flow cell

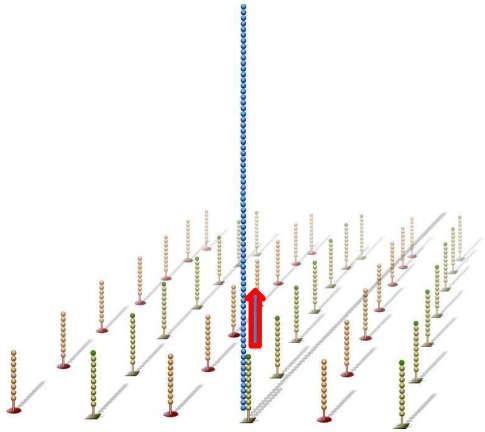


Patterned flow cell

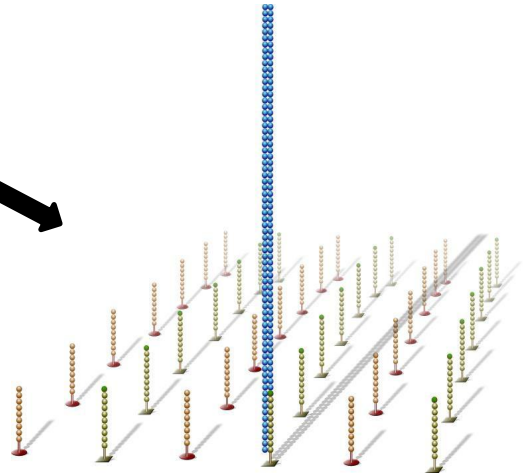
Hybridization and amplification



Prepared library

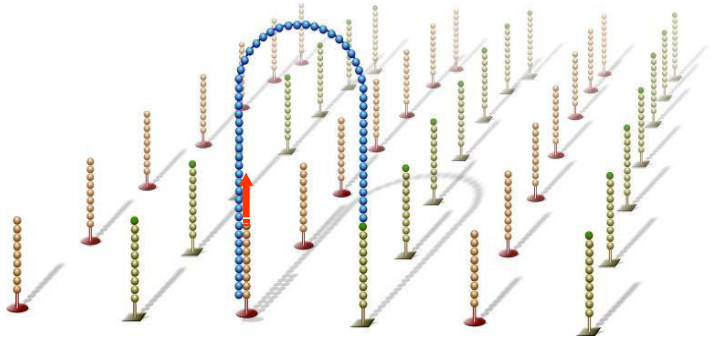


Denatured and loaded on to flow cell

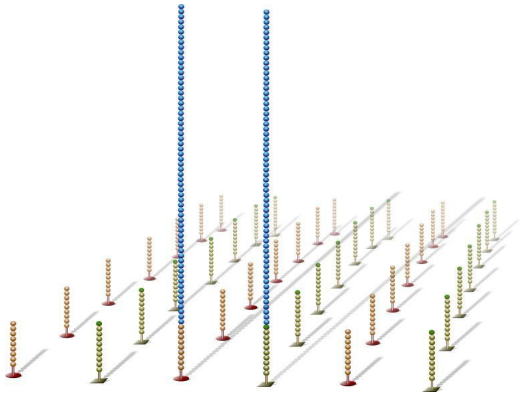


Synthesis of template strand

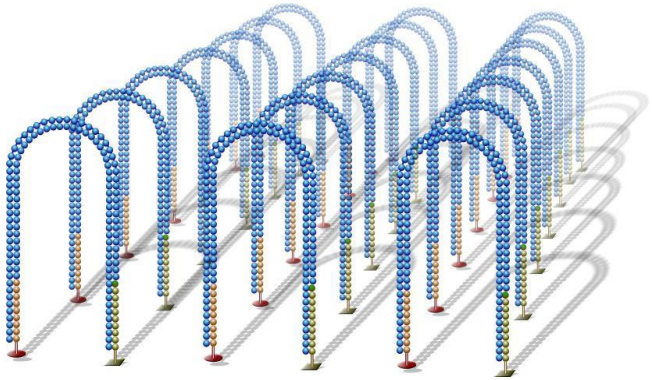
Solid-phase amplification (Bridge amplification)



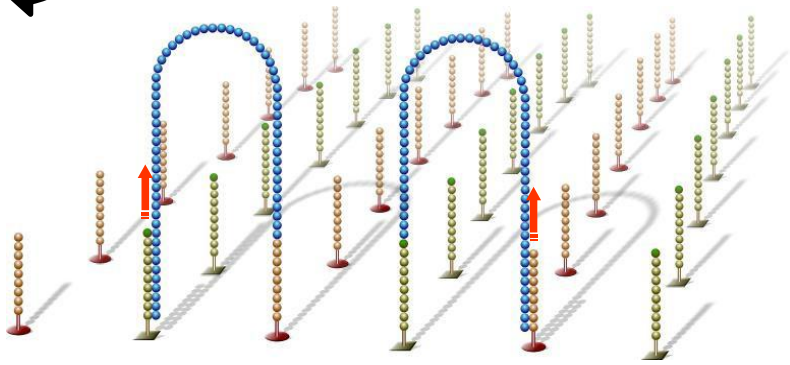
Bridge formation and extension



denaturation



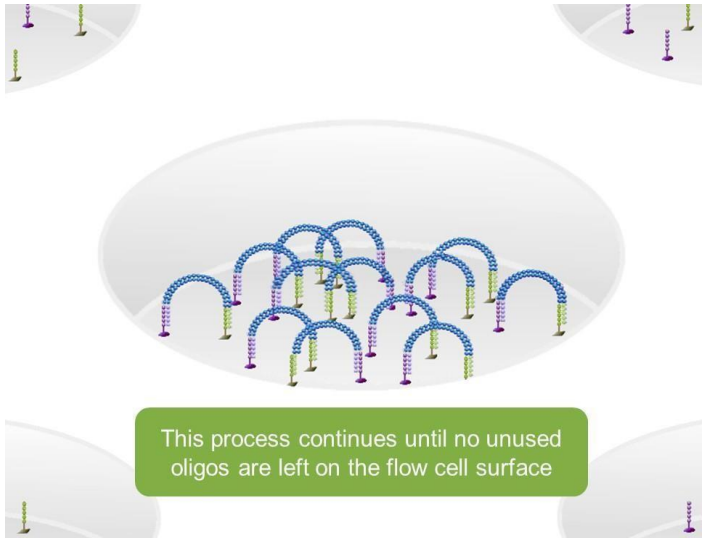
Clonal amplification

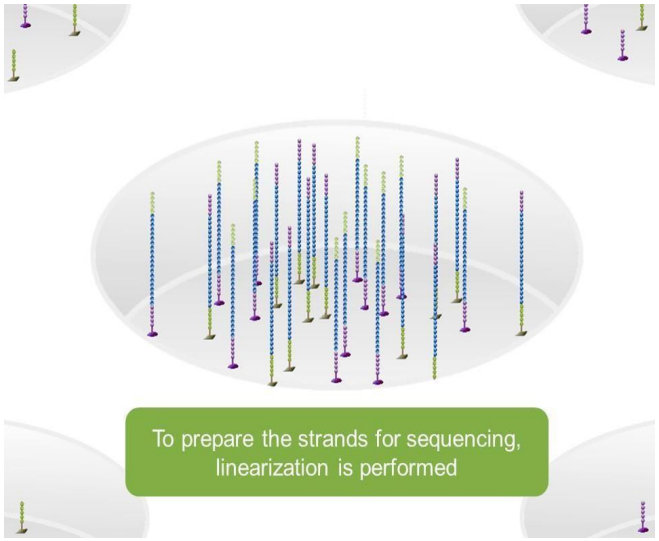


Bridge formation and extension

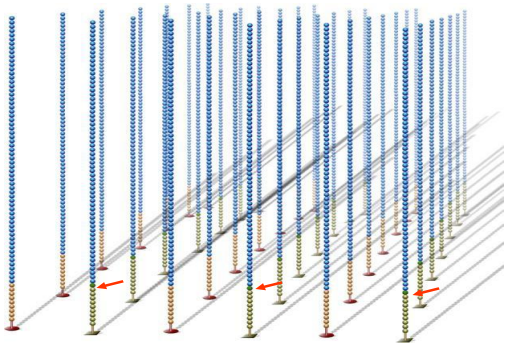
Solid-phase amplification: Cluster formation

~1000 copies of single fragment within 1um diameter

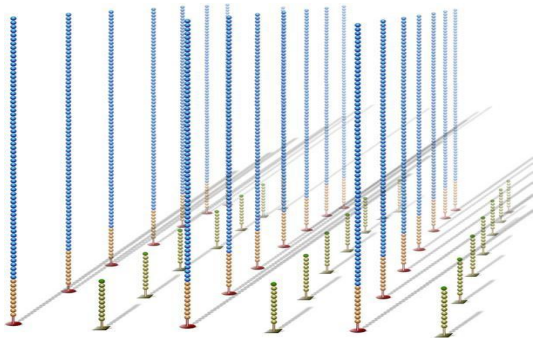




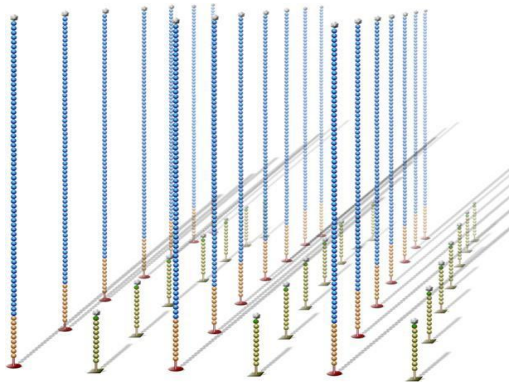
Sequence determination



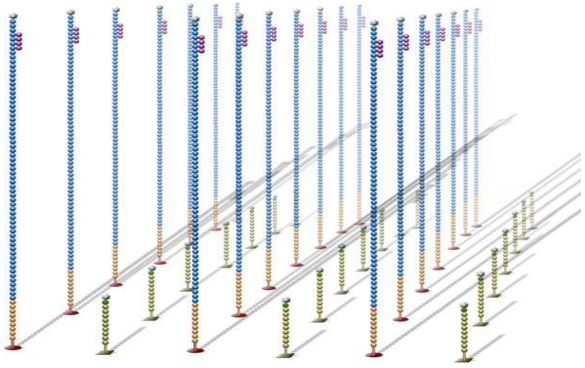
Linearization



Reverse strand cleavage

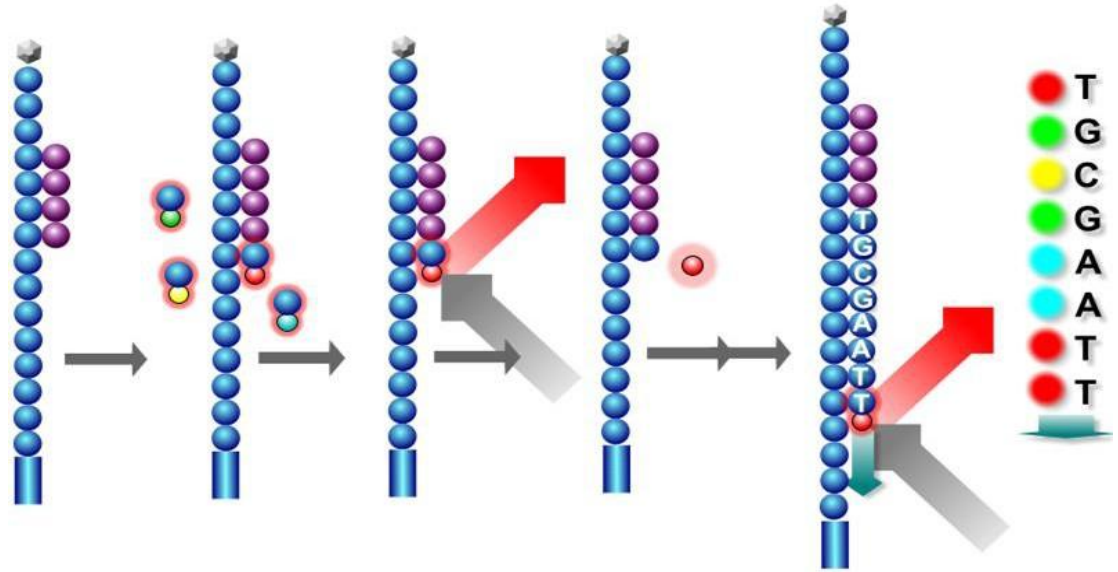


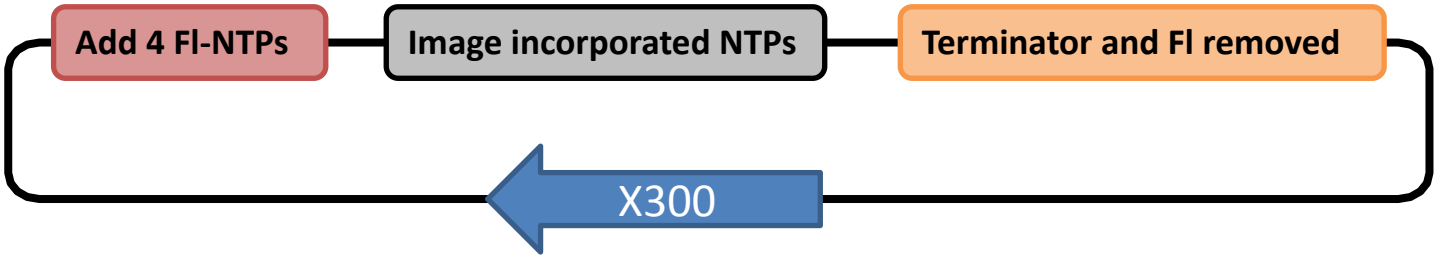
Blocking



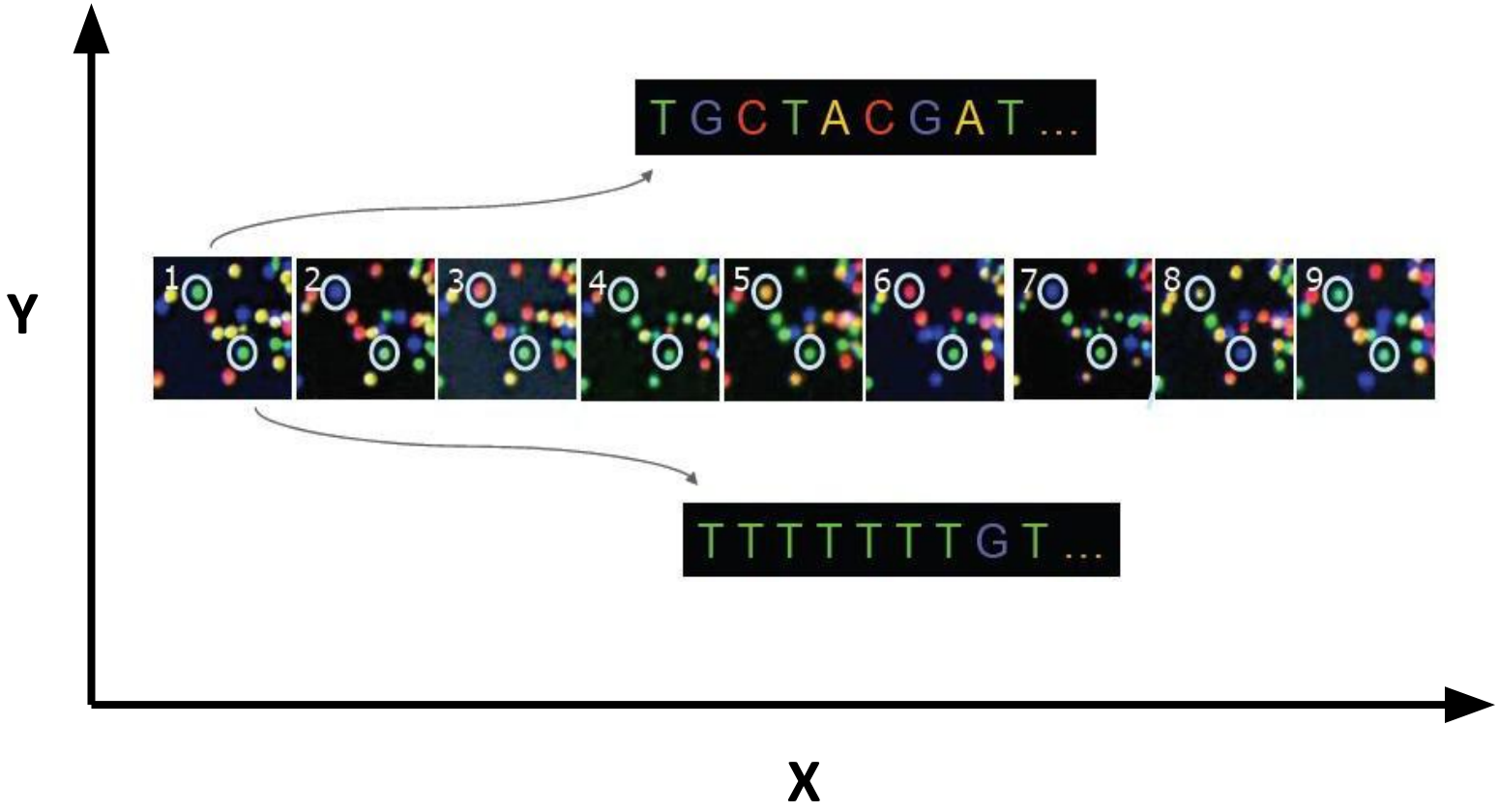
Read 1 sequencing primer

Base identification: a closer look





Assigning bases to cluster (read)



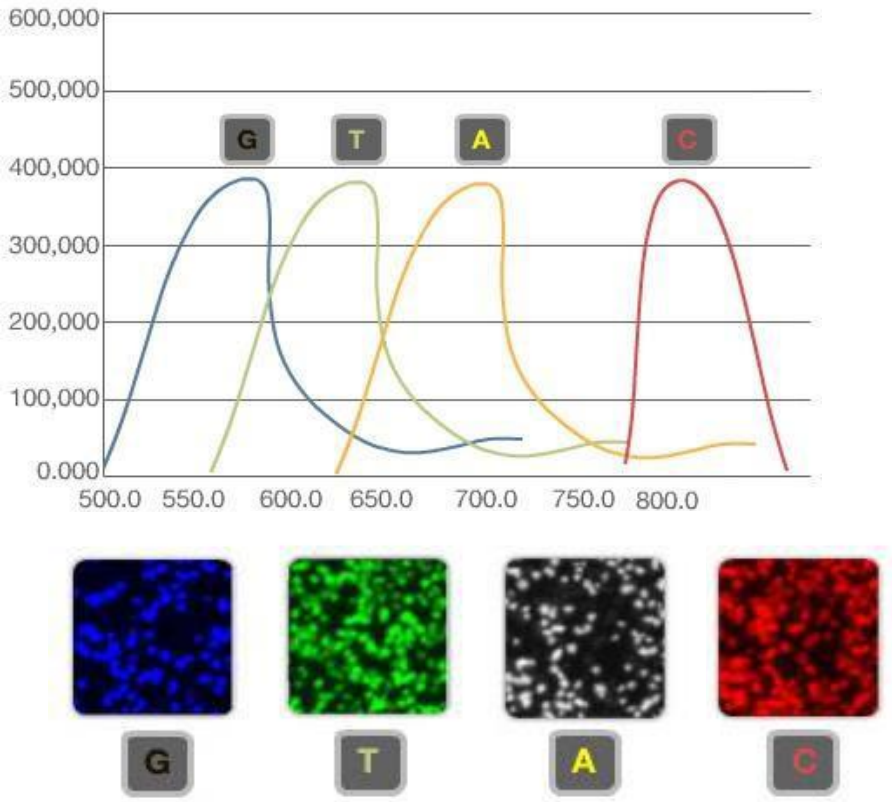
4 - Channel chemistry

Each base labelled with unique fluorophore

Each base emits light of unique intensity

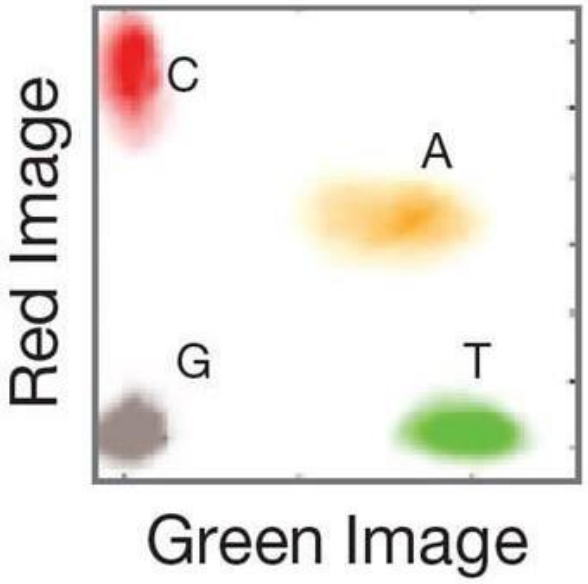
4 images are captured for each cycle

Used in GA, HiSeq and MiSeq machines



2 - Channel chemistry

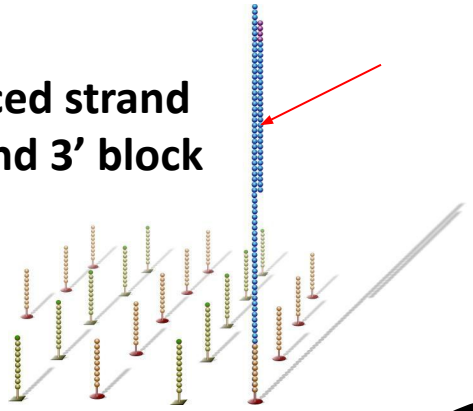
- Two channel SBS uses two images
- Clusters appearing in green only are T
- Clusters appearing in red only are C
- Clusters appearing in both images are A
- Clusters not present in either green nor red are G
- Cluster intensities are plotted and bases are called accordingly



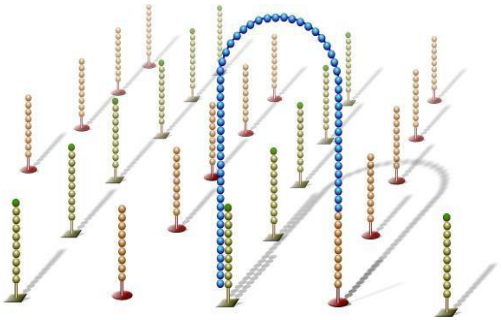
illumina

Paired-end sequencing

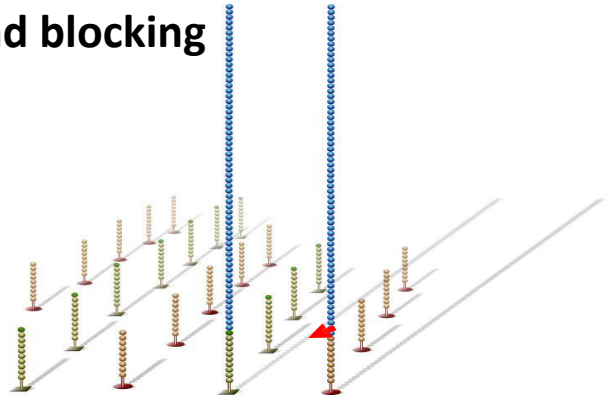
1. Sequenced strand stripped and 3’ block removed



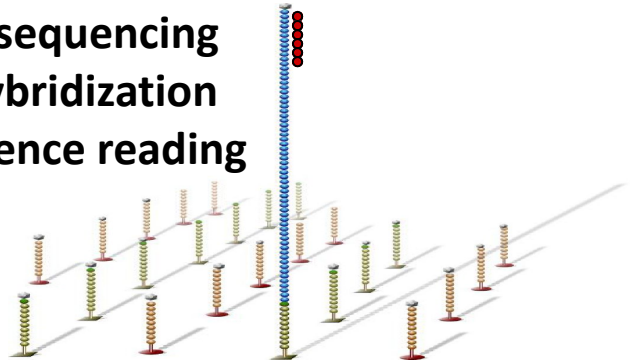
2. Bridge amplification



3. Forward strand cleavage and blocking

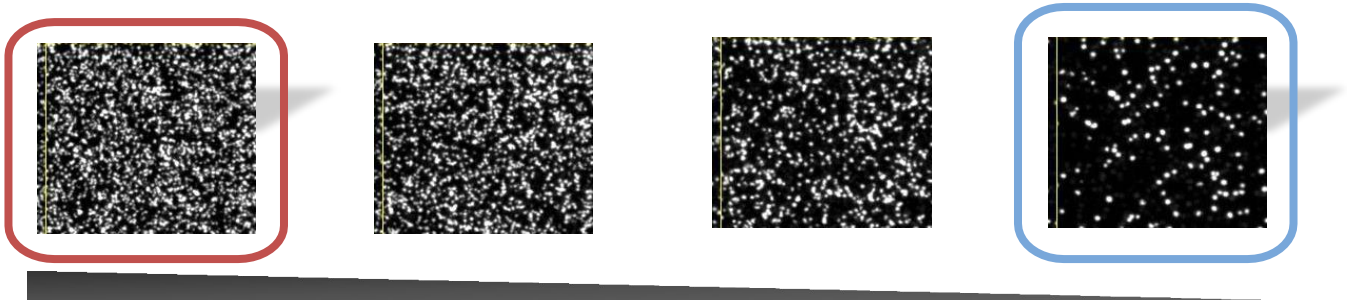


4. Read2 sequencing primer hybridization and sequence reading



Troubleshooting a sequence run

Optimized flow cell clustering determines data quality and overall data yield



Library Concentration

Over-clustering can result in:

- Loss of data quality and data output
- Reduced base calls and Q30 scores
- Complete run failure
- Loss of focus

Under-clusterring can result in:

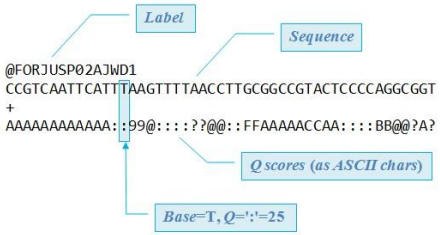
- Loss of time and money
- Loss of focus
- Complete run failure

MiSeq Sequence v3 run metrics

MiSeq Run

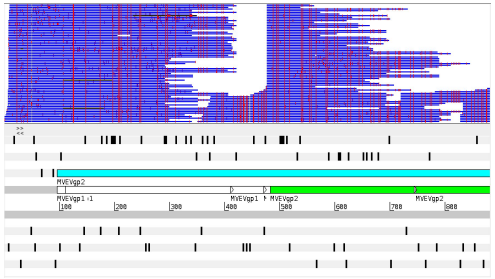
Cluster density	>900 K/mm^2
Cluster passing	>85%
Q30 output	>85%
Data output	10-15 GB

Raw sequence data



HQ reads (q30)	>80%
Length	>60b

Reference Alignment



Reads mapped	>95%
Ref cov. (>20x)	>95%

Applications

Platform	iSeq 100	MiniSeq	MiSeq	NextSeq	HiSeq	NovaSeq
Large Genomes				●	●	●
Small Genomes	●	●	●	●	●	●
Exome Sequencing				●	●	●
Targeted Resequencing	●	●	●	●	●	●
Transcriptome Sequencing				●	●	●
Gene Expression Profiling				●	●	●
miRNAs	●	●	●	●	●	●
DNA-Protein Interactions			●	●	●	●
Methylation Sequencing				●	●	●
16S Metagenomic sequencing		●	●	●	●	●

Limitations

- Some systematic errors:
  - Difficult to spot rare variants (<1%).
- Low complexity templates.
  - Add complex library to 30%, phase ensure variation at start of read.
- Sequencing short fragments doesn't give any long range information.
- Index hopping
- Transposase based library prep can introduce bias leading to uneven coverage
- Difficult to assemble complete genome

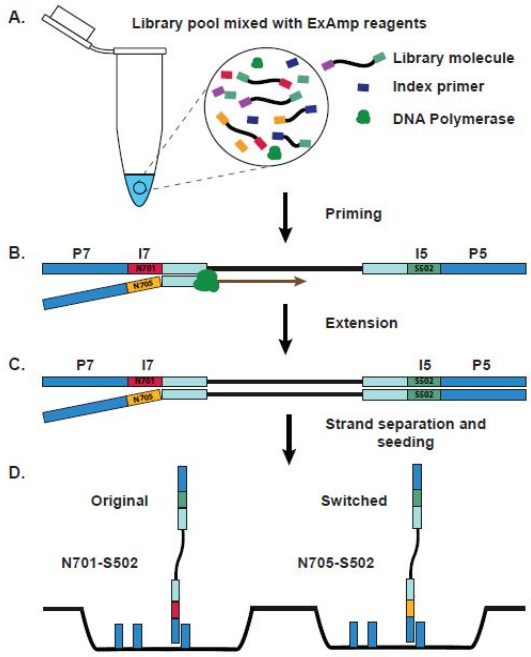


Figure 1

Thank you