



Sequencing Technologies

Michael Quail
mq1@sanger.ac.uk



Sequencers



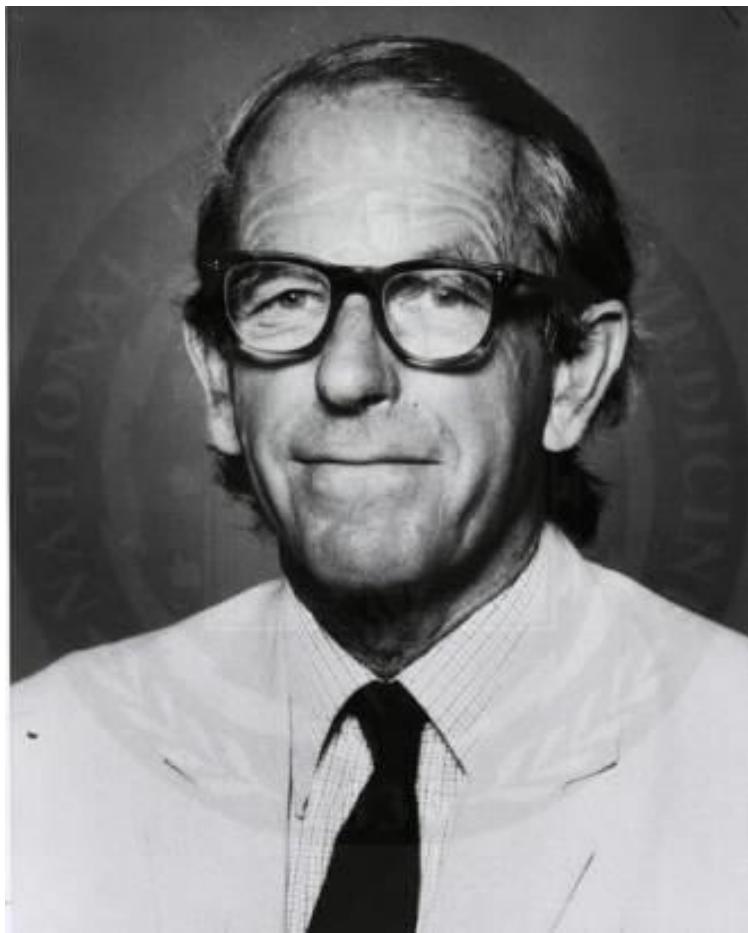


Talk outline

- » History of sequencing
- » Description of different technologies
- » Outline of advantages and disadvantages
- » Look to the future

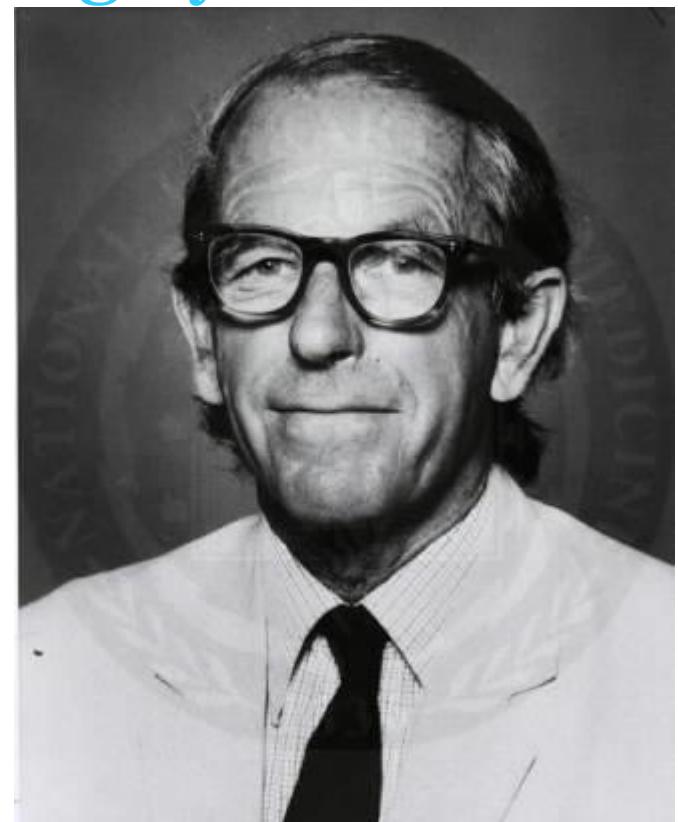


Sanger Sequencing

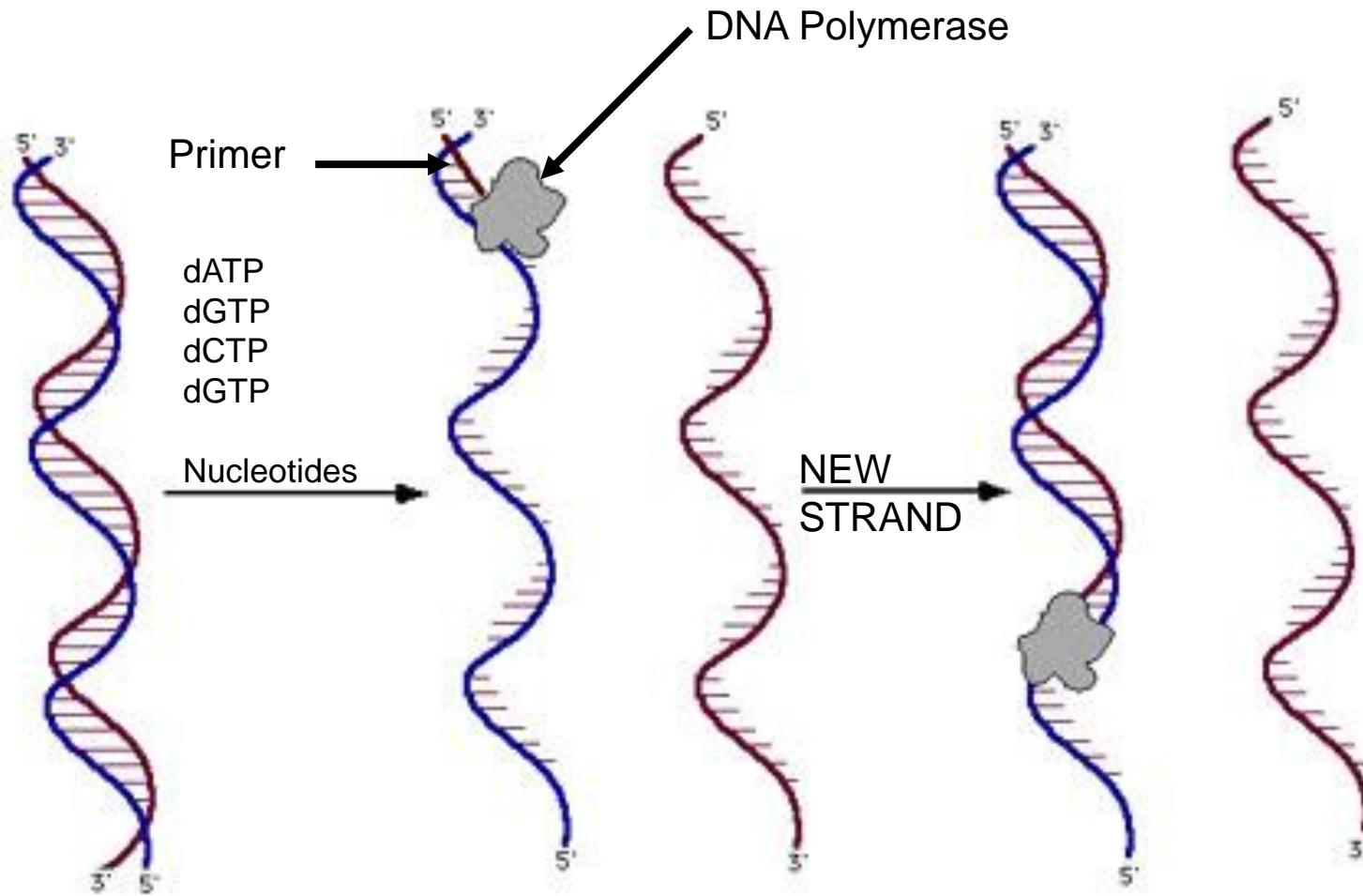


Frederick Sanger

- » Discovered DNA sequencing by chain termination method
- » Nobel Prize 1 (1958)
 - » Complete amino acid sequence of insulin
- » Nobel Prize 2 (1980)
 - » For DNA sequencing



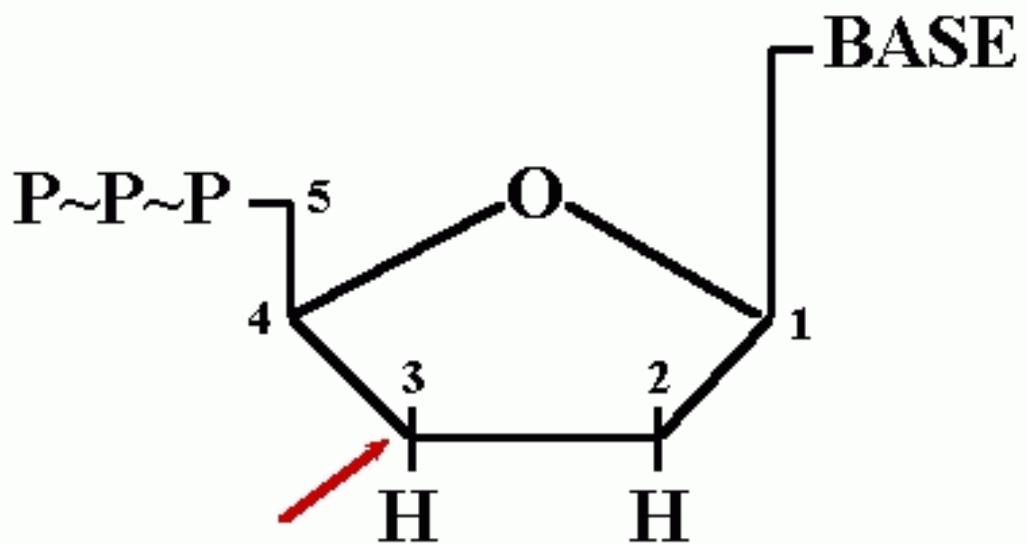
Primer Extension



Template DNA

Dideoxy Nucleotides

- Lack an -OH group at the 3-carbon position
- Cannot add another nucleoside at that position
- Prevent further DNA synthesis



All Possible Terminations

DNA Polymerase reads the template strand and synthesizes a new second strand to match:

5' - TACGCGGTAAACGGTATGTTGACCCGTTAGCTACCGAT
3' - ATGCGCCATTGCCATACAGCTGGCATAATCGATGGCTAGAGATCCAA - 5'



IF 5% of the T nucleotides are actually dideoxy T, then each strand will terminate when it gets a ddT on its growing end:

5' - TACGCGGTAAACGGTATGTTGACCCGTTAGCTACCGAT•

5' - TACGCGGTAAACGGTATGTTGACCCGTTAGCT•

5' - TACGCGGTAAACGGTATGTTGACCCGTT•

5' - TACGCGGTAAACGGTATGTTGACCCGTT•

5' - TACGCGGTAAACGGTATGTTGACCCGT•

5' - TACGCGGTAAACGGTATGTT•

5' - TACGCGGTAAACGGTATGT•

5' - TACGCGGTAAACGGTAT•

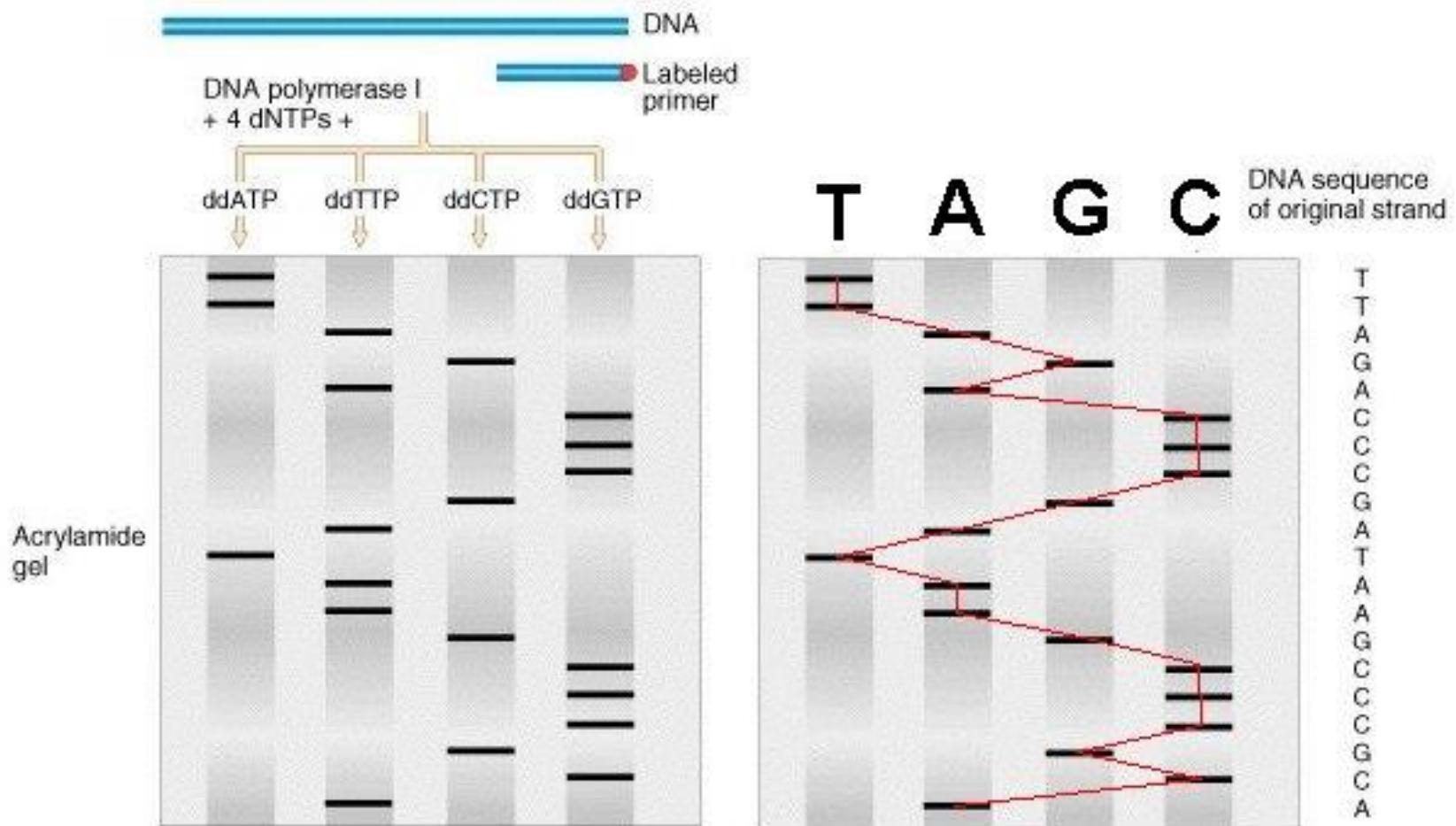
5' - TACGCGGTAAACGGT•

5' - TACGCGGT•

Original Sanger Sequencing

- » 4 sequencing reactions performed for each template, each with different terminator
- » Radioactive primer or nucleotide used
- » Sequencing reactions run on <1mm polyacrylamide gel cast between two glass plates to separate fragments according to size
- » After run gel exposed to film and developed to reveal image

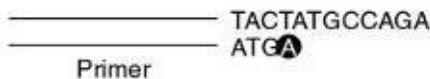
Sequencing gel autorad



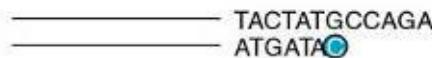
Fluorescent Terminators

Primer extension reactions:

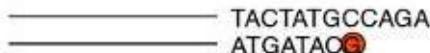
ddA reaction:



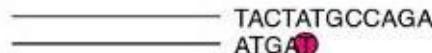
ddC reaction:



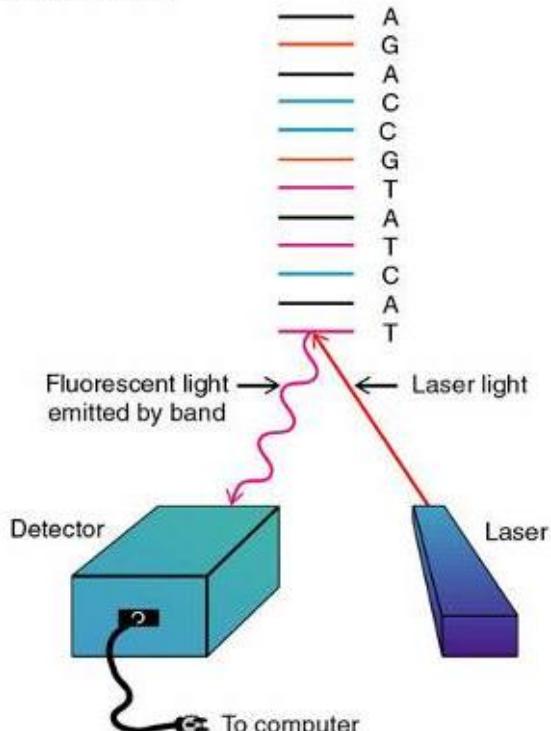
ddG reaction:



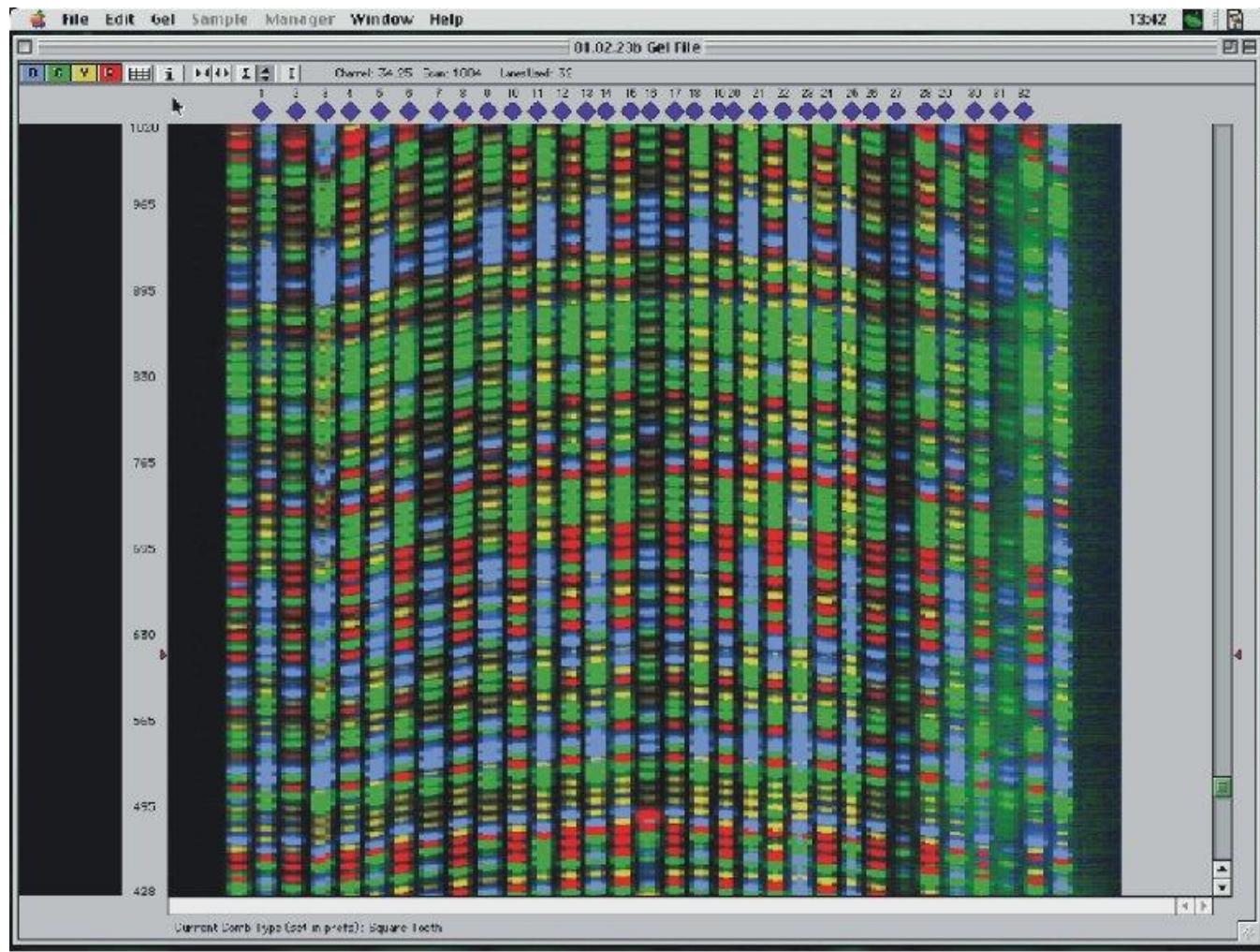
ddT reaction:

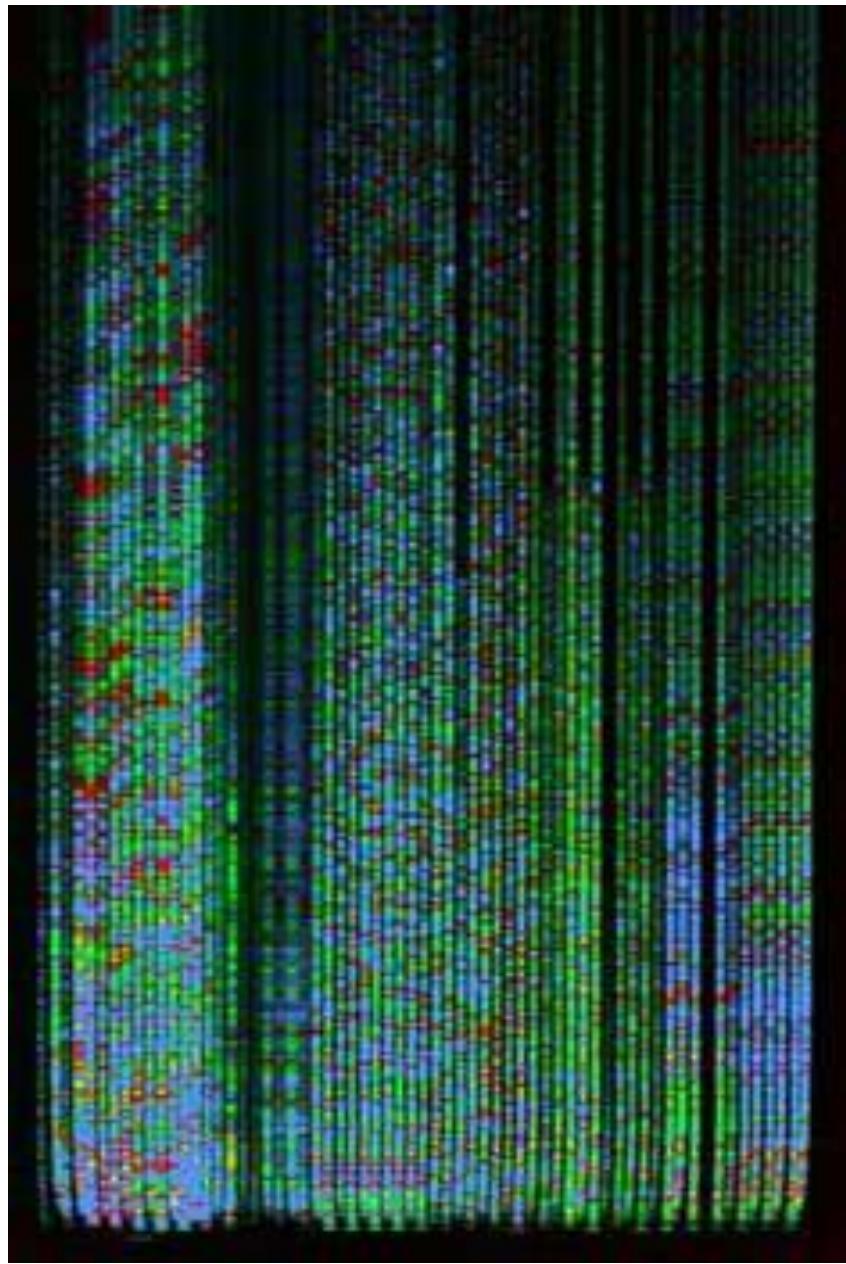


Electrophoresis:



Fluorescent Gel Sequencing



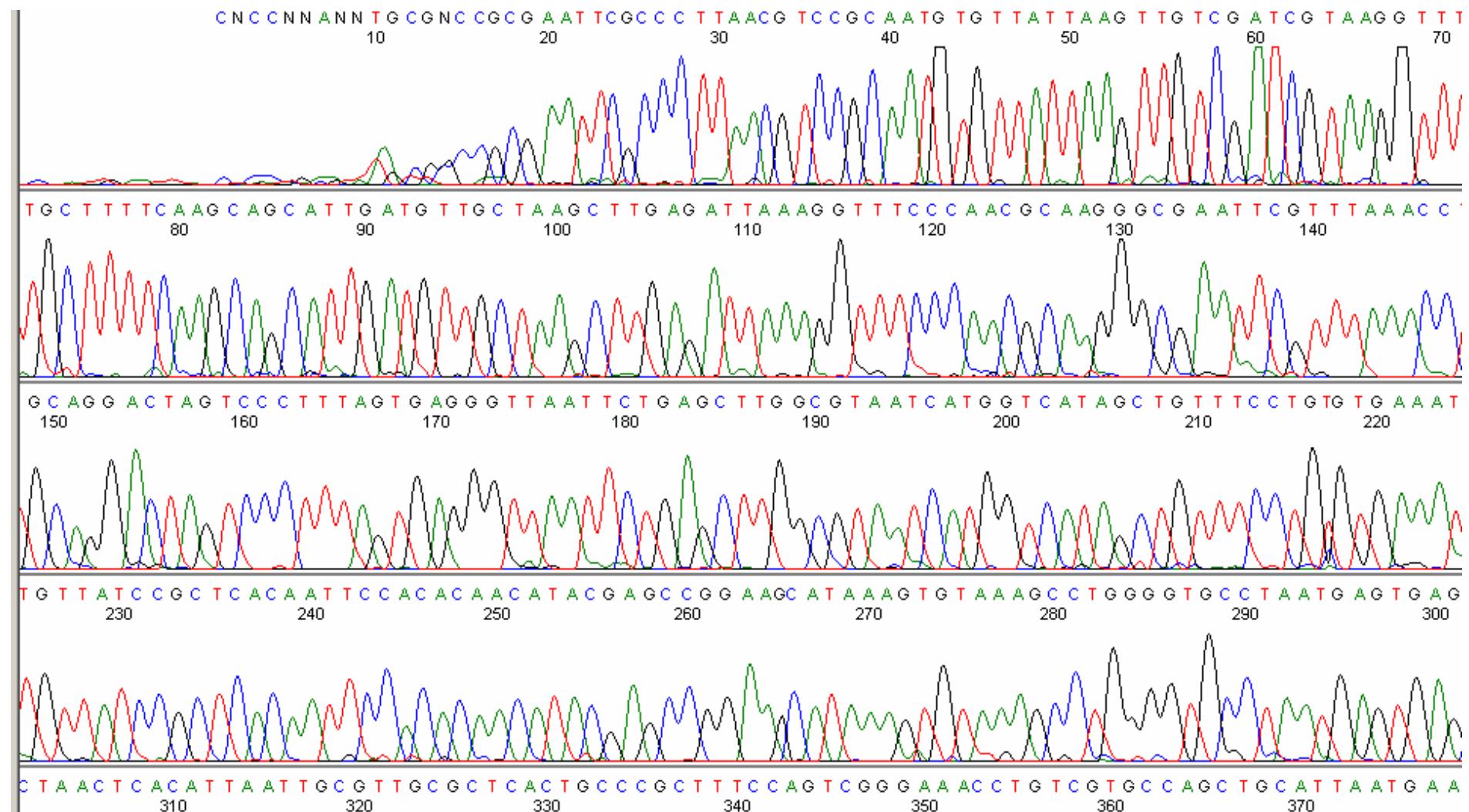




ABI 3730



DNA Sequence Files





ABI Capillary 3730

- » Individual reactions -> 96 capillary array
- » Accurate, Q30
- » PCR errors
- » Cloning bias
- » 1000-base reads
- » 1-2 hour run time





Error

- » Sequence quality Q is reported on a log scale
- » Q₁₀ is 1 error in 10
- » Q₂₀ is 1 error in 100
- » Q₃₀ is 1 error in 1000
- » Q₄₀ is 1 error in 10000
- » Q₅₀ is 1 error in 100000



ABI Capillary 3730

- » Fast + easy for individual samples
- » Robust technology
- » .0001 Gb / run
- » £150k instrument
- » \$1,000,000 / Gb





NGS

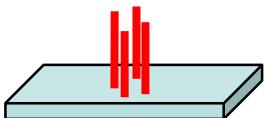
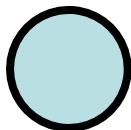
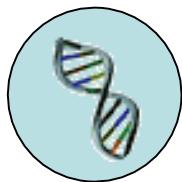
2005 - present



Next Generation sequencing

Is massively parallel
Not limited to few reactions per run

Next-Generation Sequencing



1 feature
1 template



1 chip, thousands or millions of features
Output Mb-Tb

454

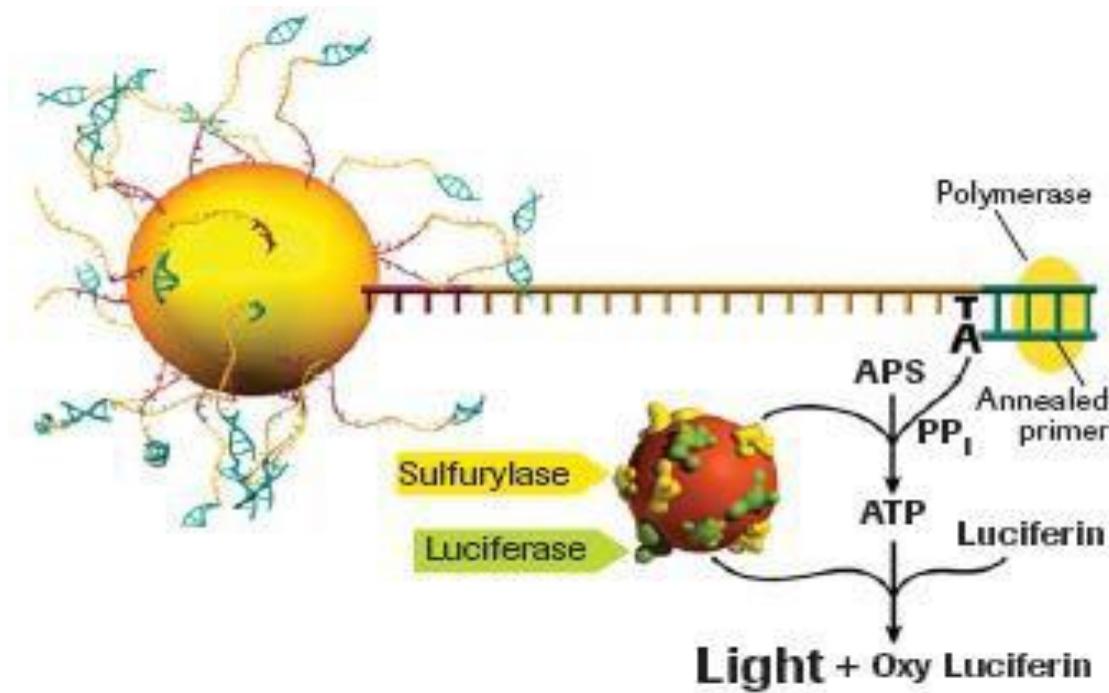


454

- » Started NGS (2005)
- » First massively parallel sequencer
- » Bought by Roche in 2007
- » Based on pyrosequencing of bead-bound DNA in microwells

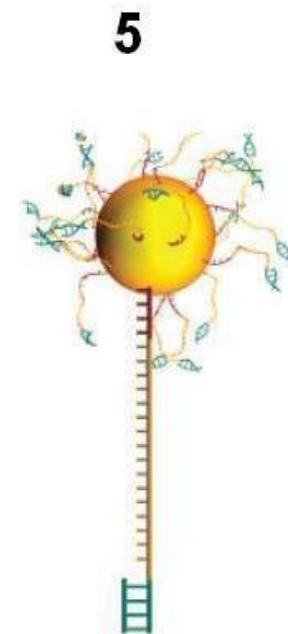
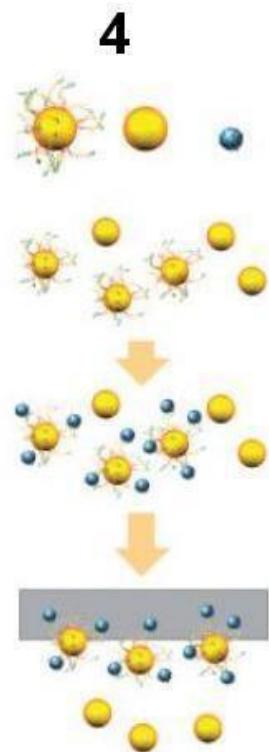
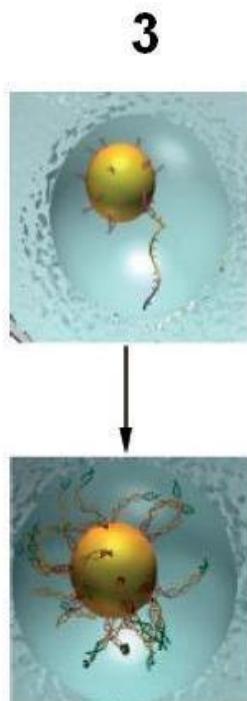
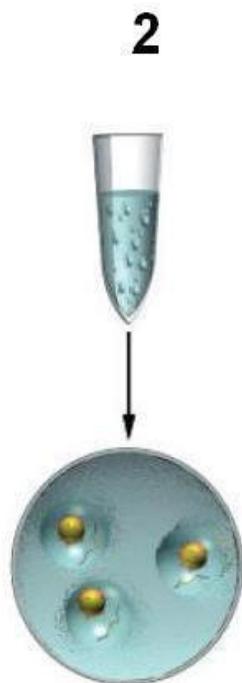
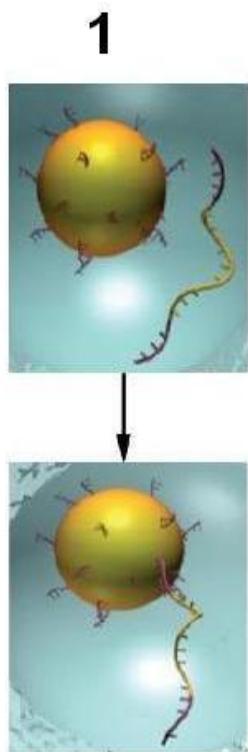


Pyrosequencing



- » The incorporation of new bases releases inorganic pyrophosphate
- » A chemical cascade converts luciferin to oxy-luciferin + light

454 Emulsion PCR



Single DNA molecule anneals to new bead via adapter sequences

Beads + PCR reagents are formed into 1-bead droplets

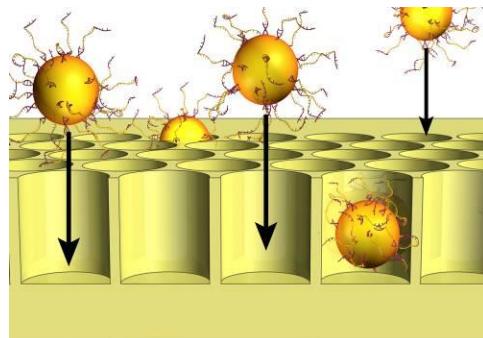
emPCR w/ biotinylated primer + lots of Taq polymerase

Emulsion is broken w/ I.P. alcohol. No-DNA beads removed

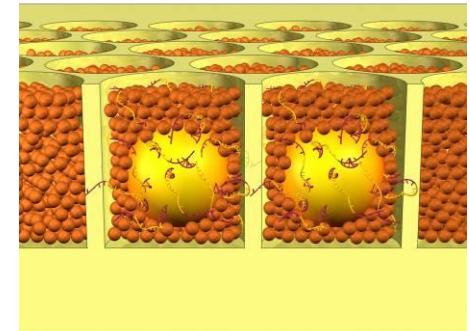
Primer is annealed ready for pyrosequencing

Depositing DNA Beads into the PicoTiter™Plate

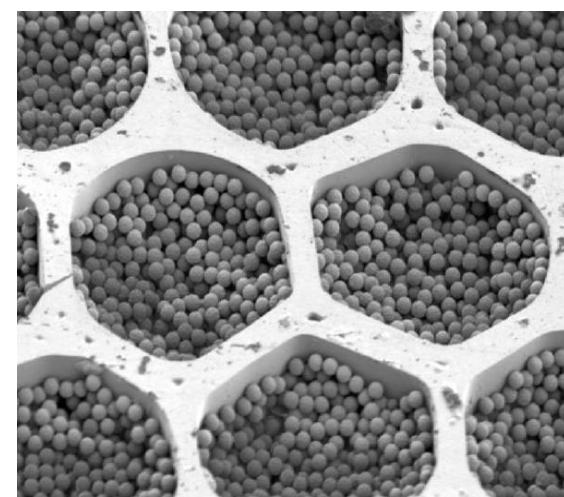
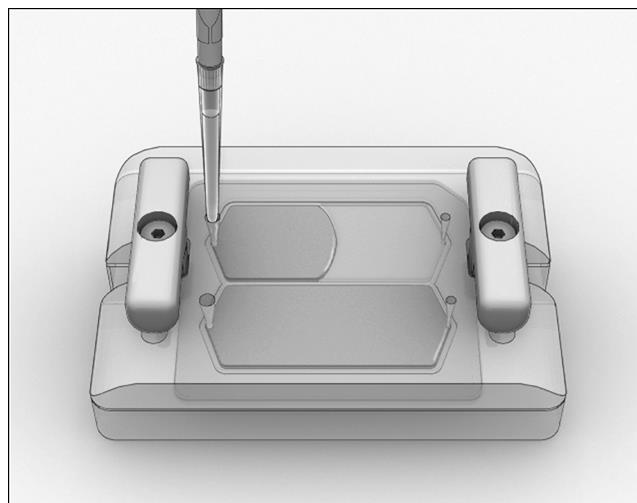
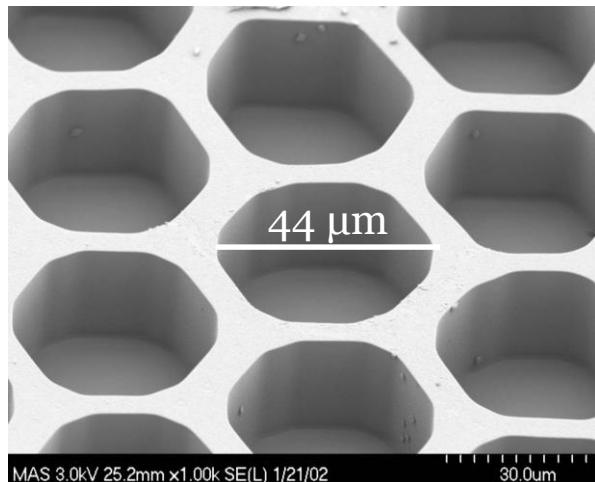
Load beads into
PicoTiter™Plate



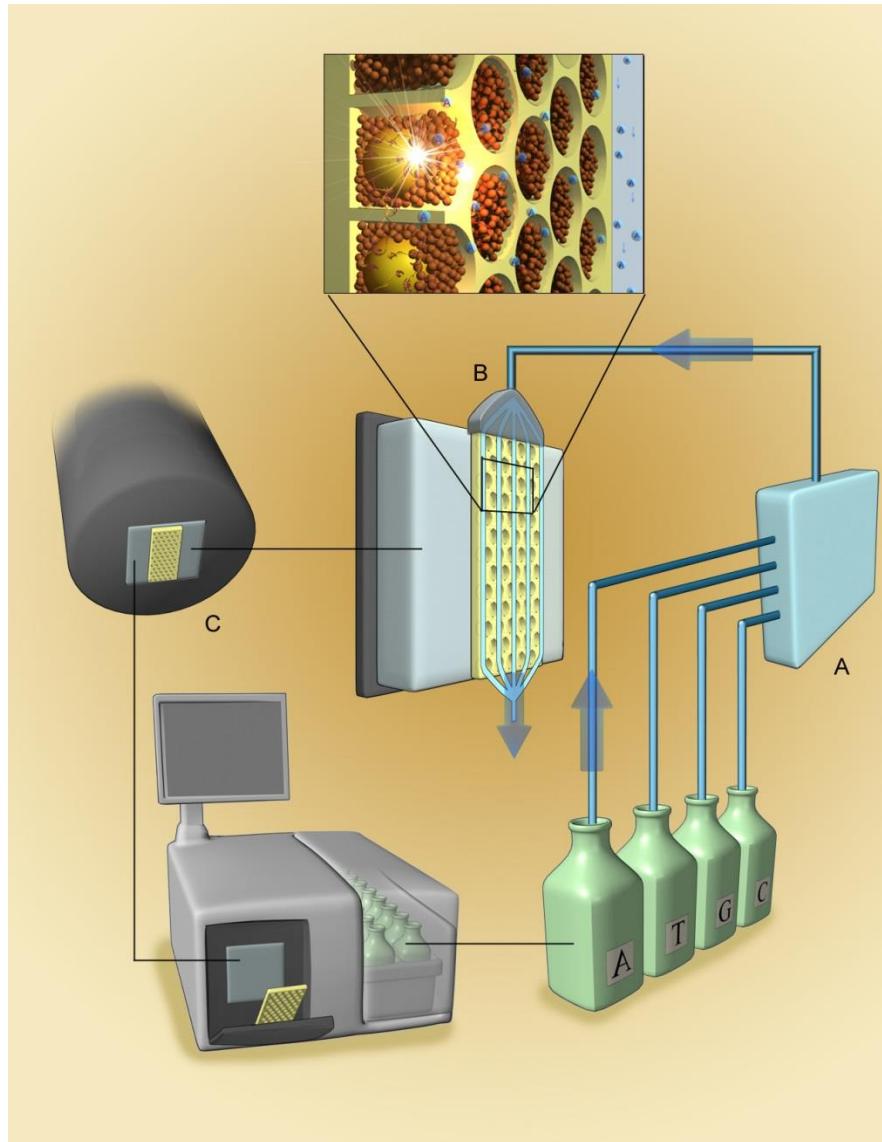
Load Enzyme Beads



Centrifugation

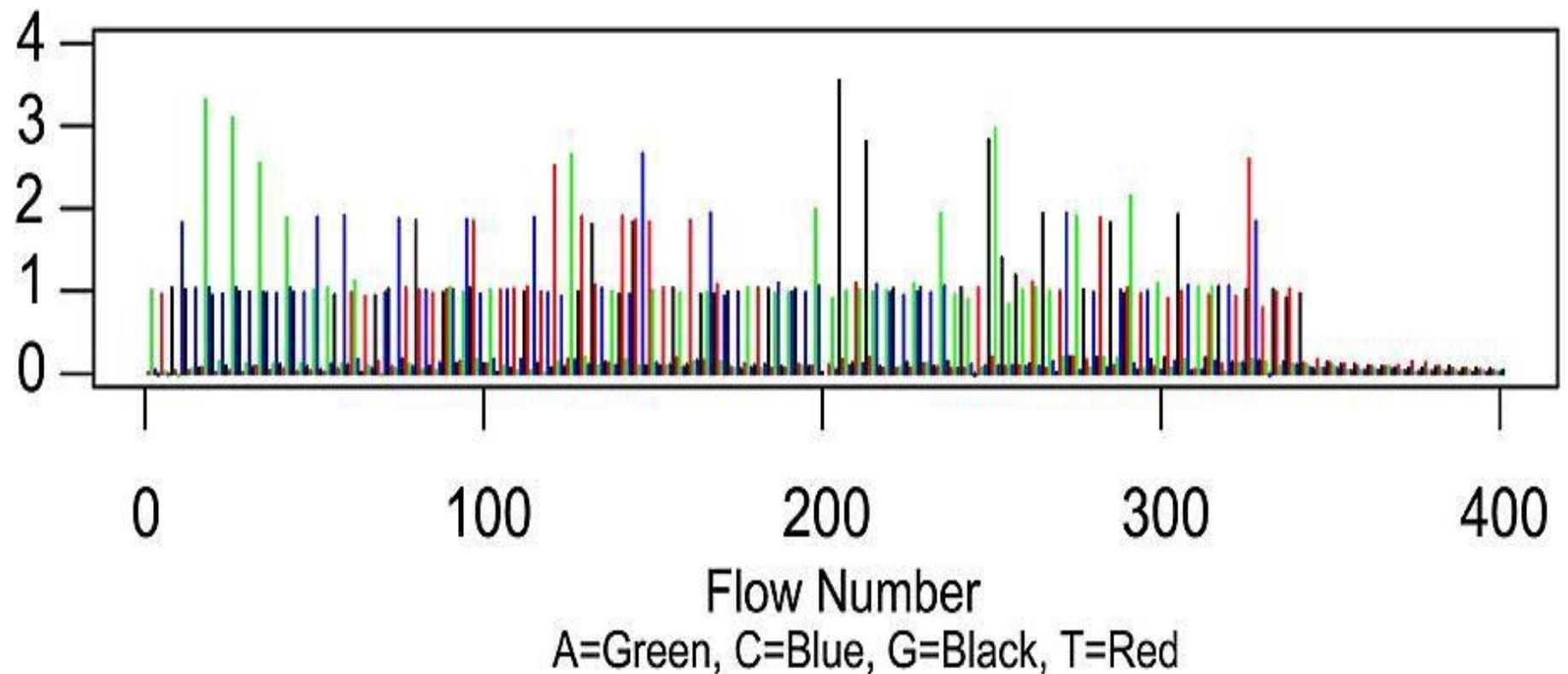


454 Technology - Sequencing Instrument





454 Data Example



454/Roche Summary

- » Long read lengths – good for amplicons and *de-novo* sequencing
- » High error rate near homopolymers
- » Single end only



454/Roche Summary

- » .7 Gb / run
- » 700 base reads
- » <24 hour run time
- » \$7,000 / Gb



Roche discontinued in 2016 – Too expensive



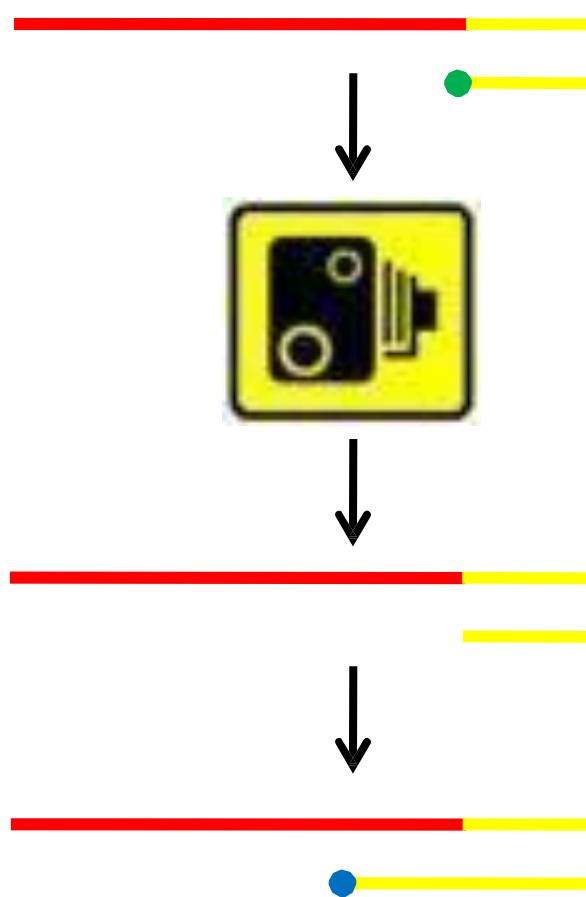
Illumina



Solexa

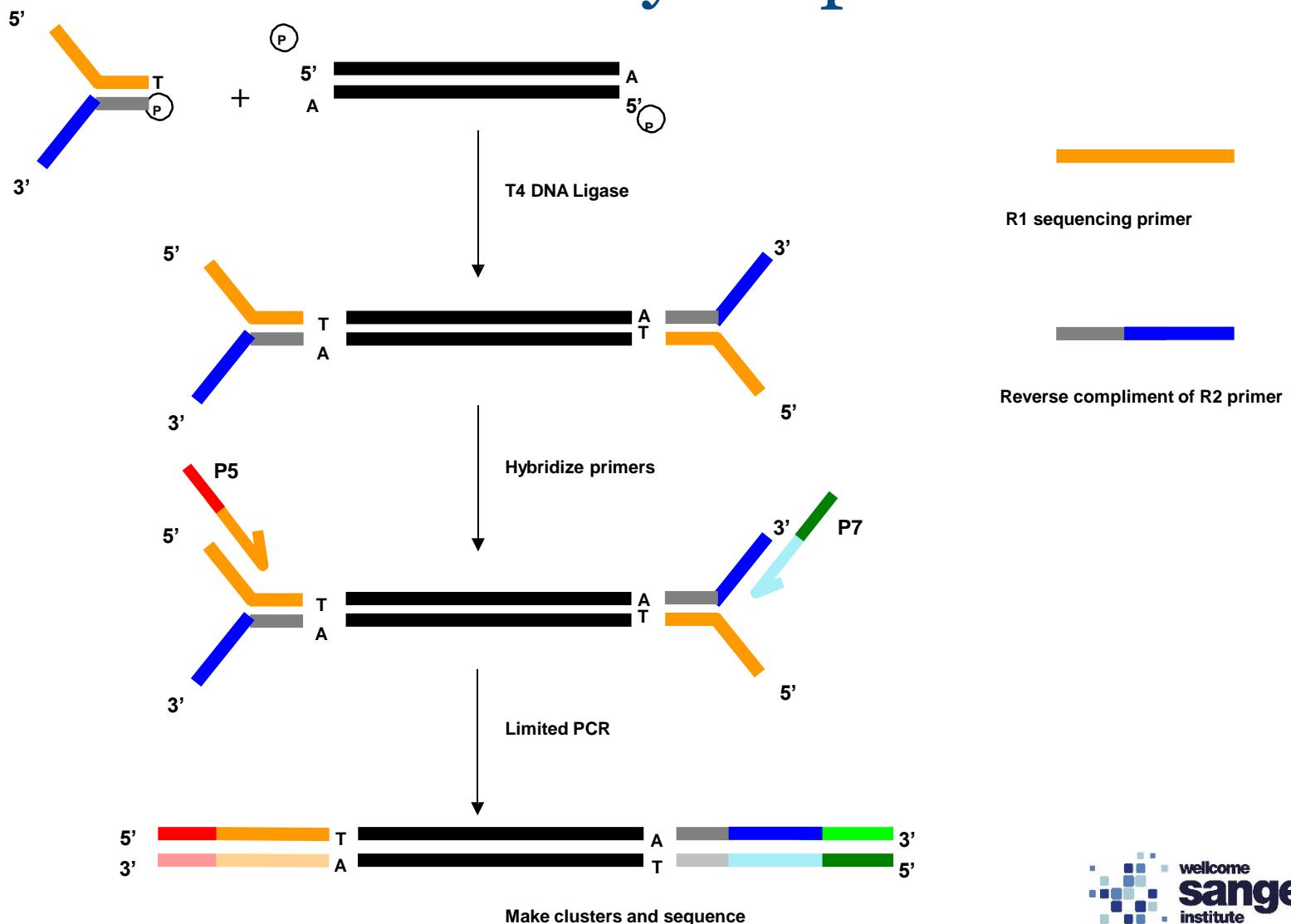
- » Launched their Genome Analyzer in 2006
- » Spinout from Cambridge University, set up at Gt. Chesterford in 2000
- » Genome Analyzer; 1Gb/run
- » Acquired by Illumina in 2007

Sequencing by Synthesis

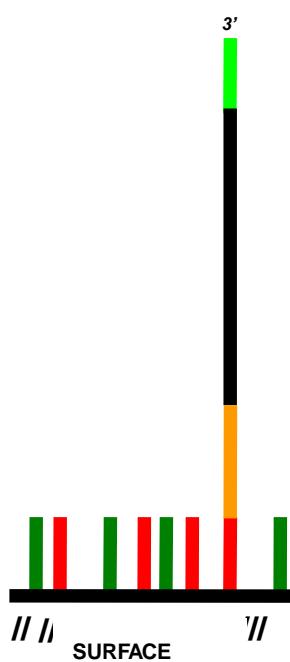


- » Extend by 1 base
- » Image
- » Reverse termination
- » Repeat

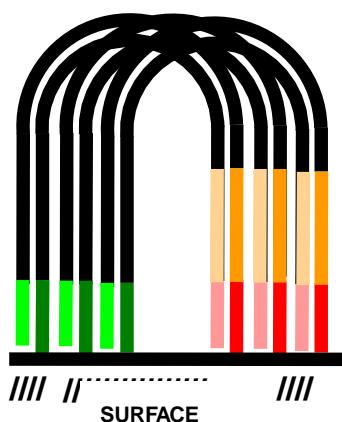
Illumina Paired End Library Prep



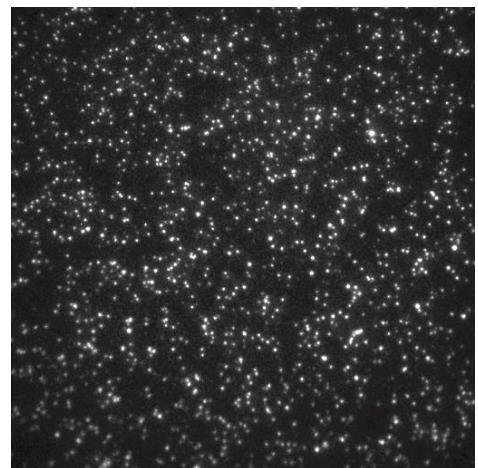
Cluster Amplification



Single-molecule
array



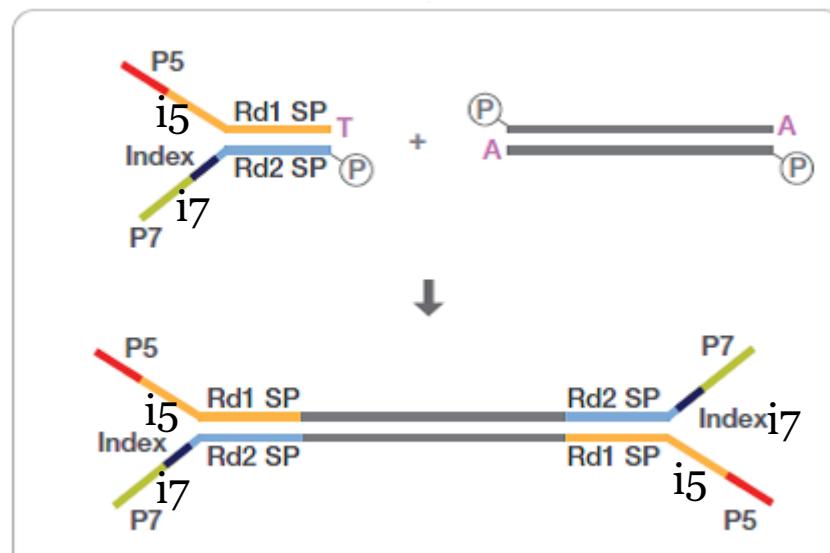
Cluster
~1000
molecules



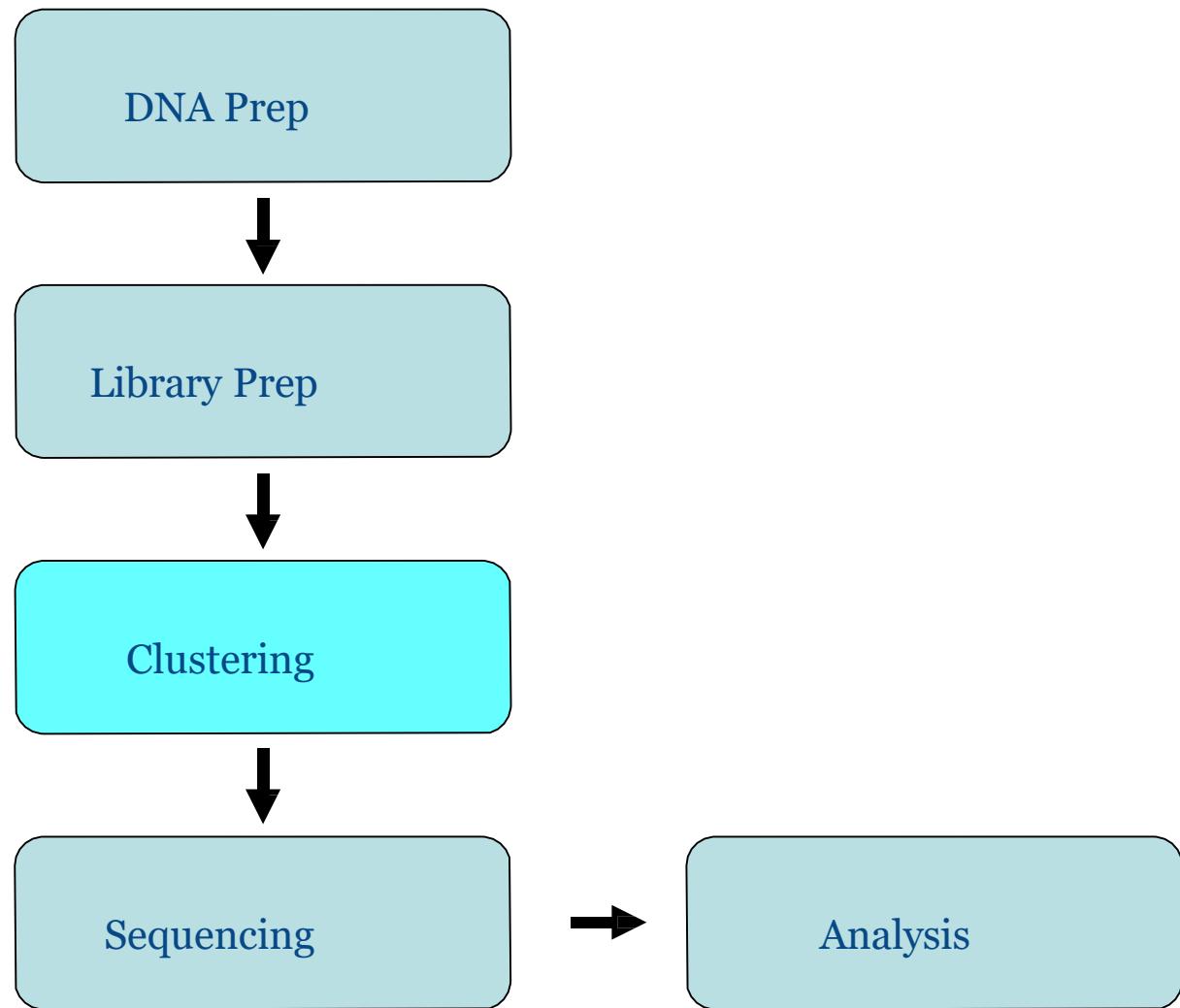
1.5 Billion
clusters on a
single glass chip

Illumina Truseq Library Prep

Dual indexing

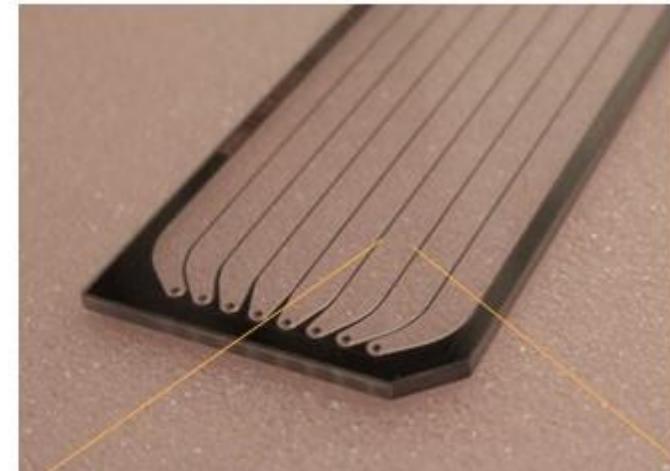


Illumina workflow



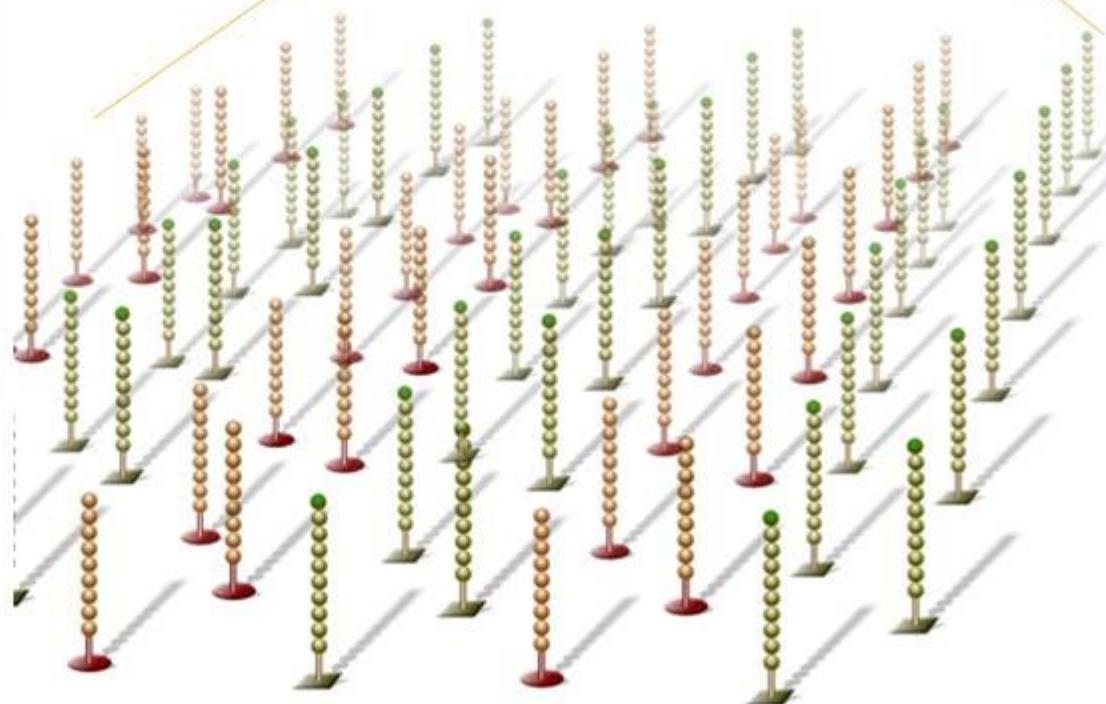
What is a Flow Cell?

Cluster generation occurs on a flow cell



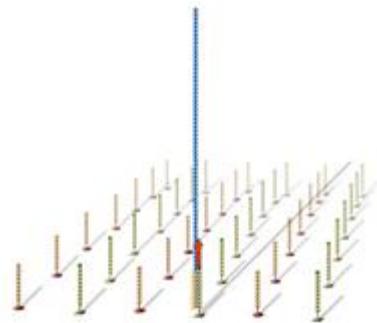
A flow cell is a thick glass slide with channels or lanes

Each lane is randomly coated with a lawn of oligos that are complementary to library adapters

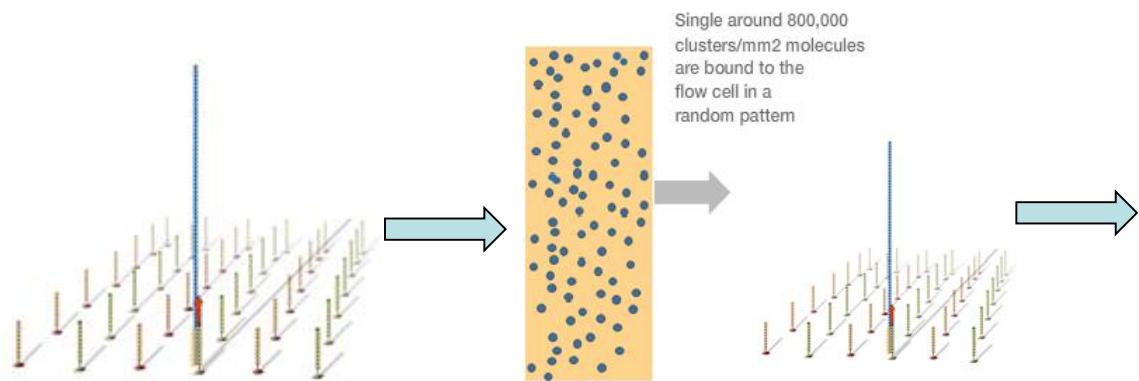




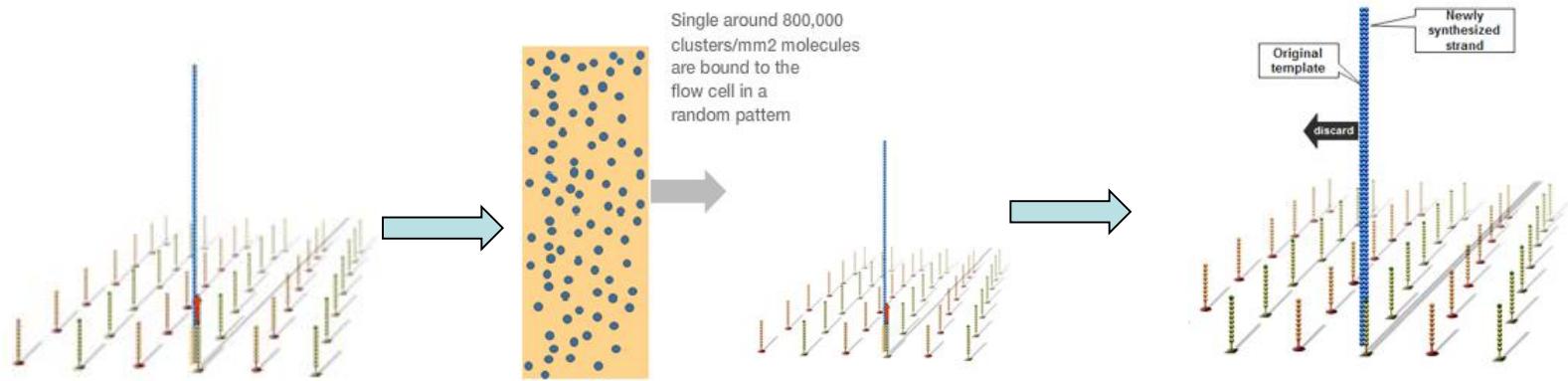
Illumina Sequencing methodology



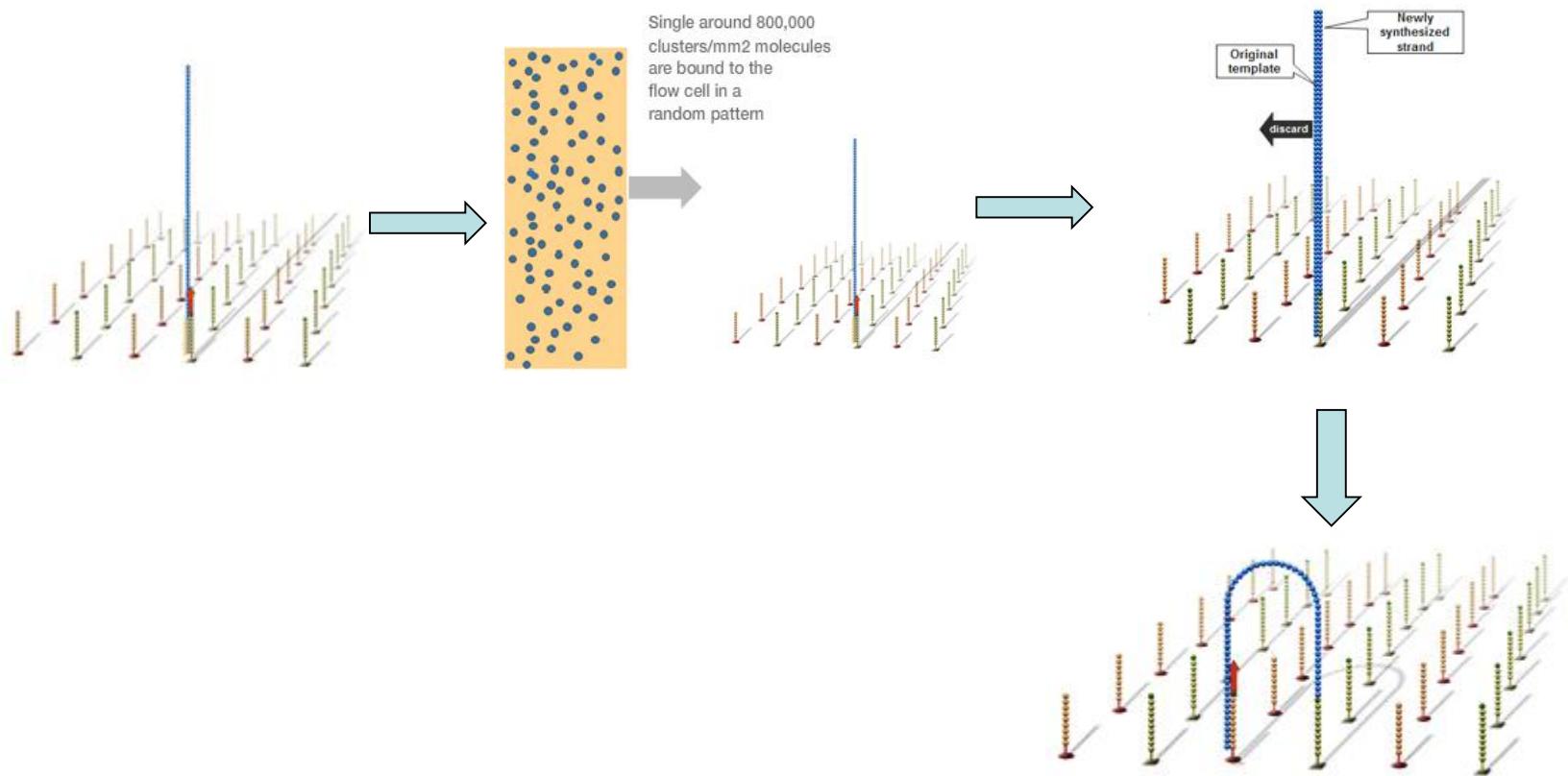
Illumina Sequencing methodology



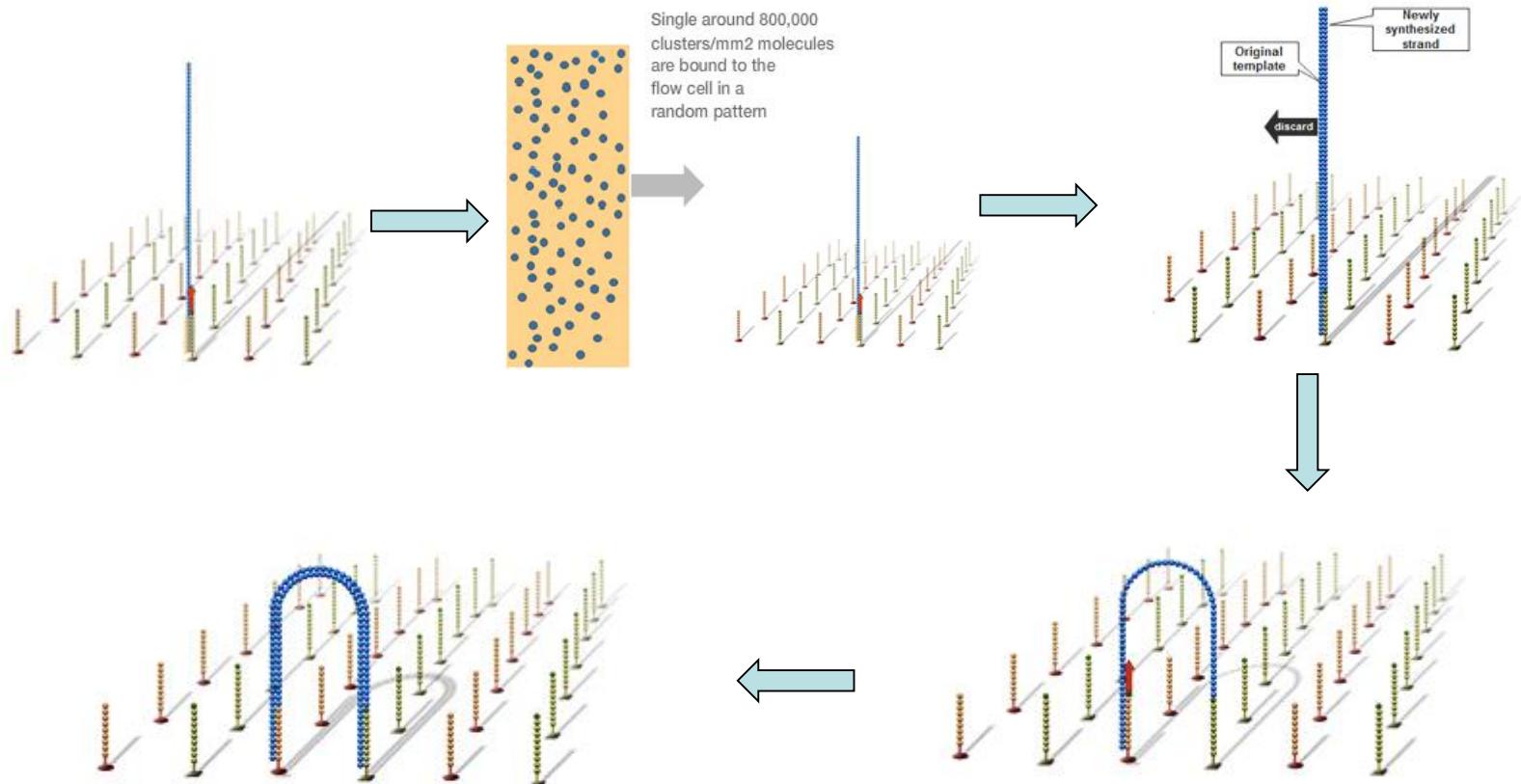
Illumina Sequencing methodology



Illumina Sequencing methodology

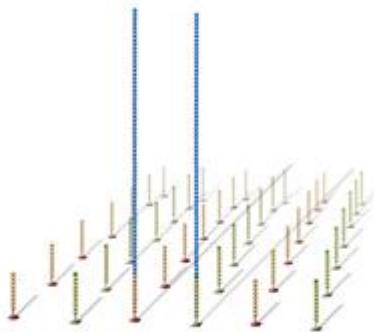


Illumina Sequencing methodology

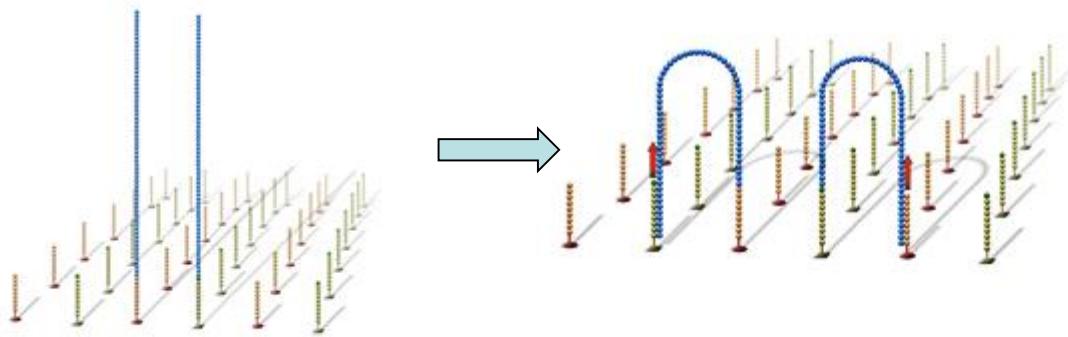




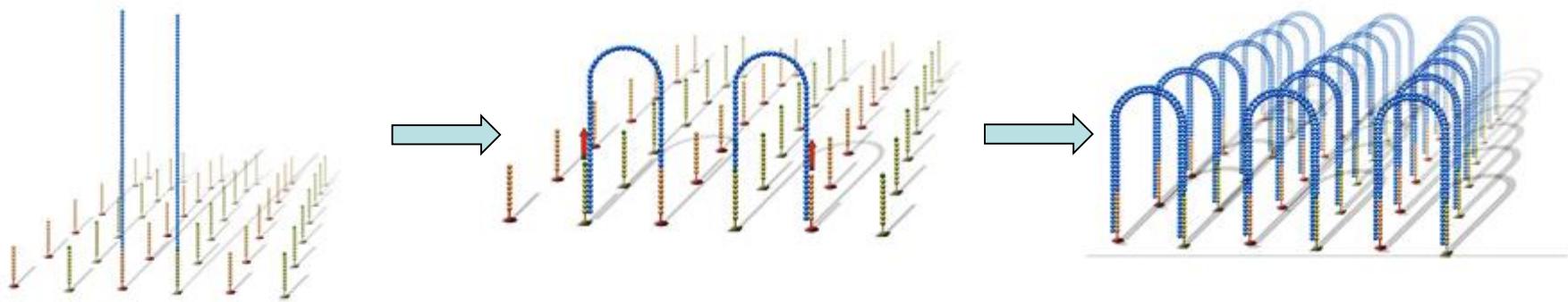
Illumina Sequencing methodology



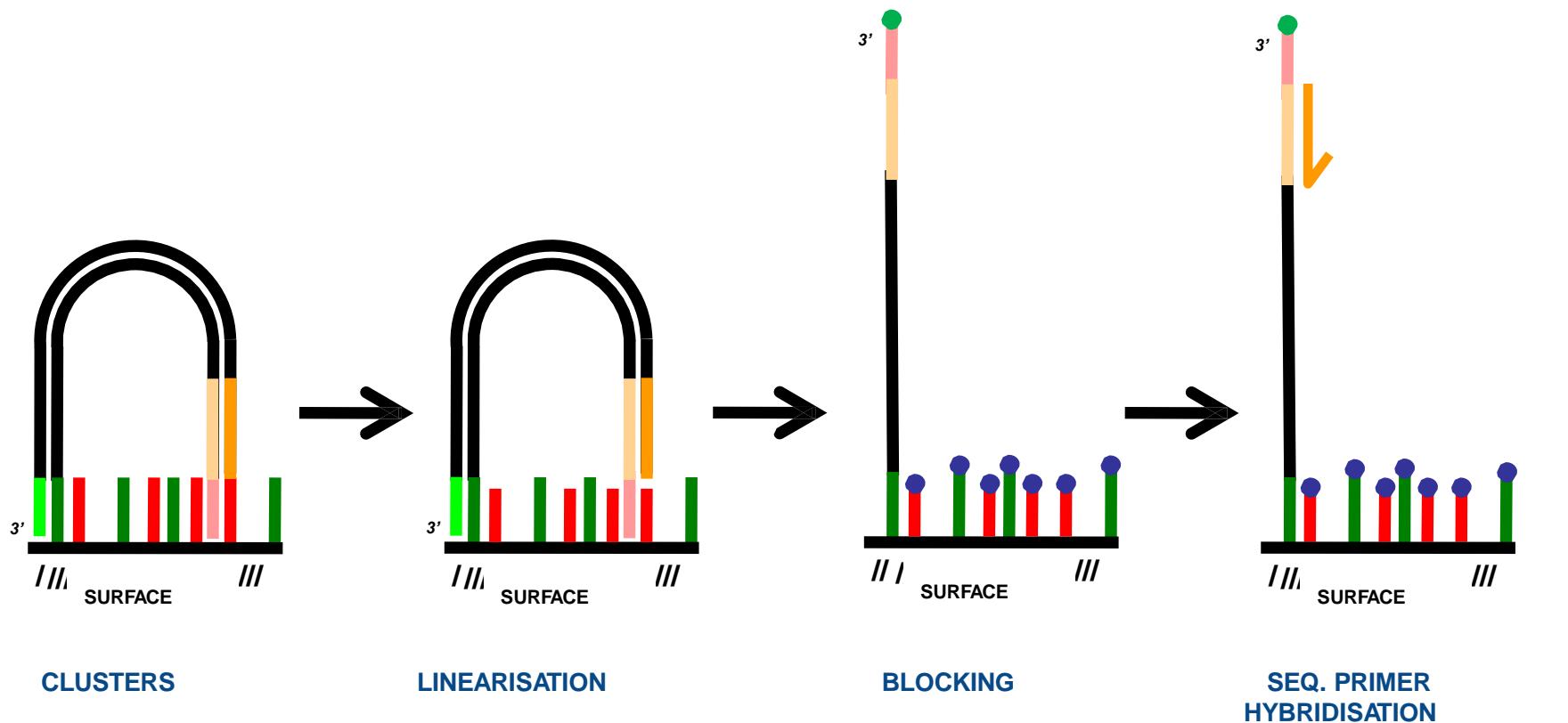
Illumina Sequencing methodology



Illumina Sequencing methodology

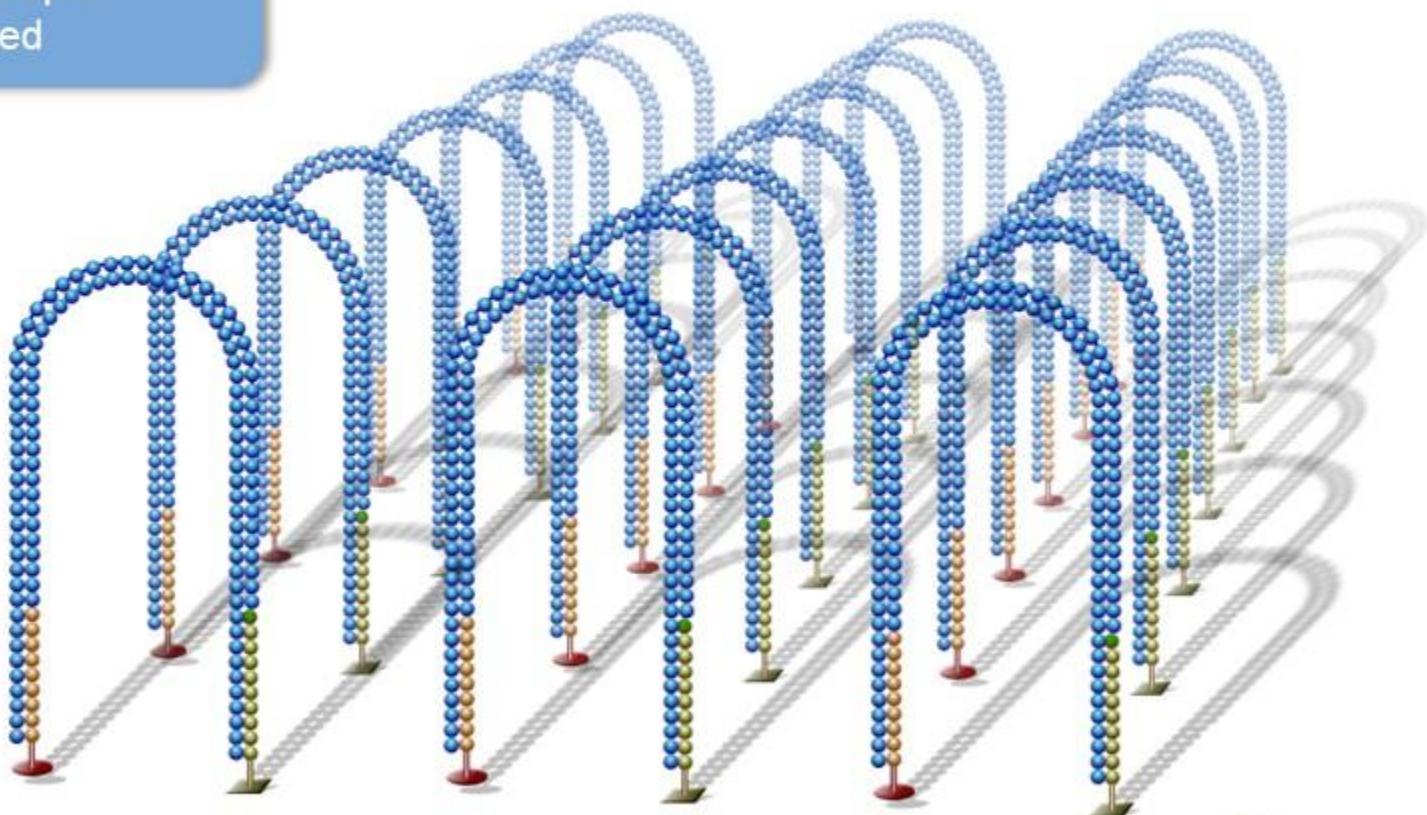


Preparing Clusters for Sequencing read 1



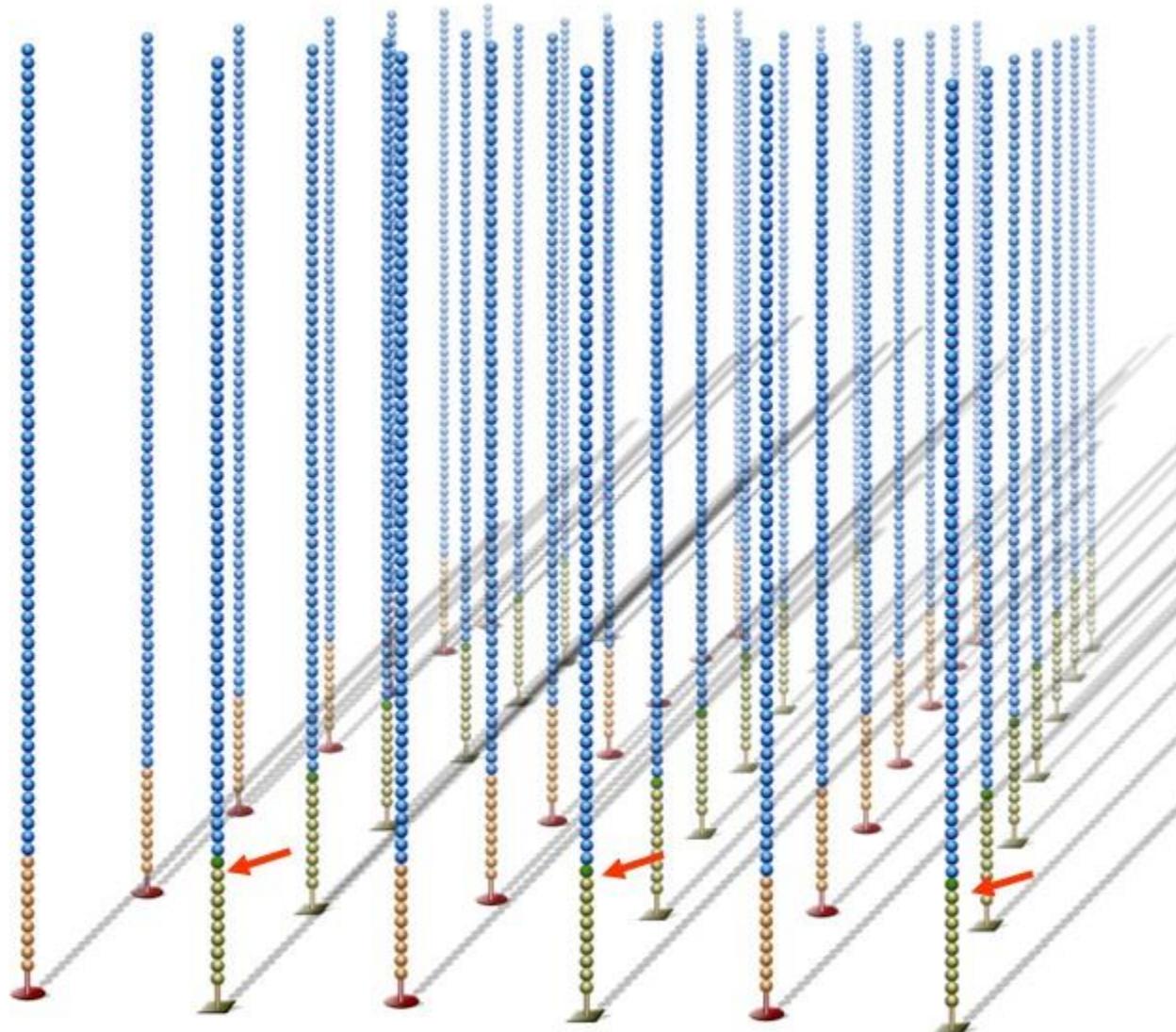
Bridge Amplification

Bridge amplification cycle is repeated until multiple bridges are formed



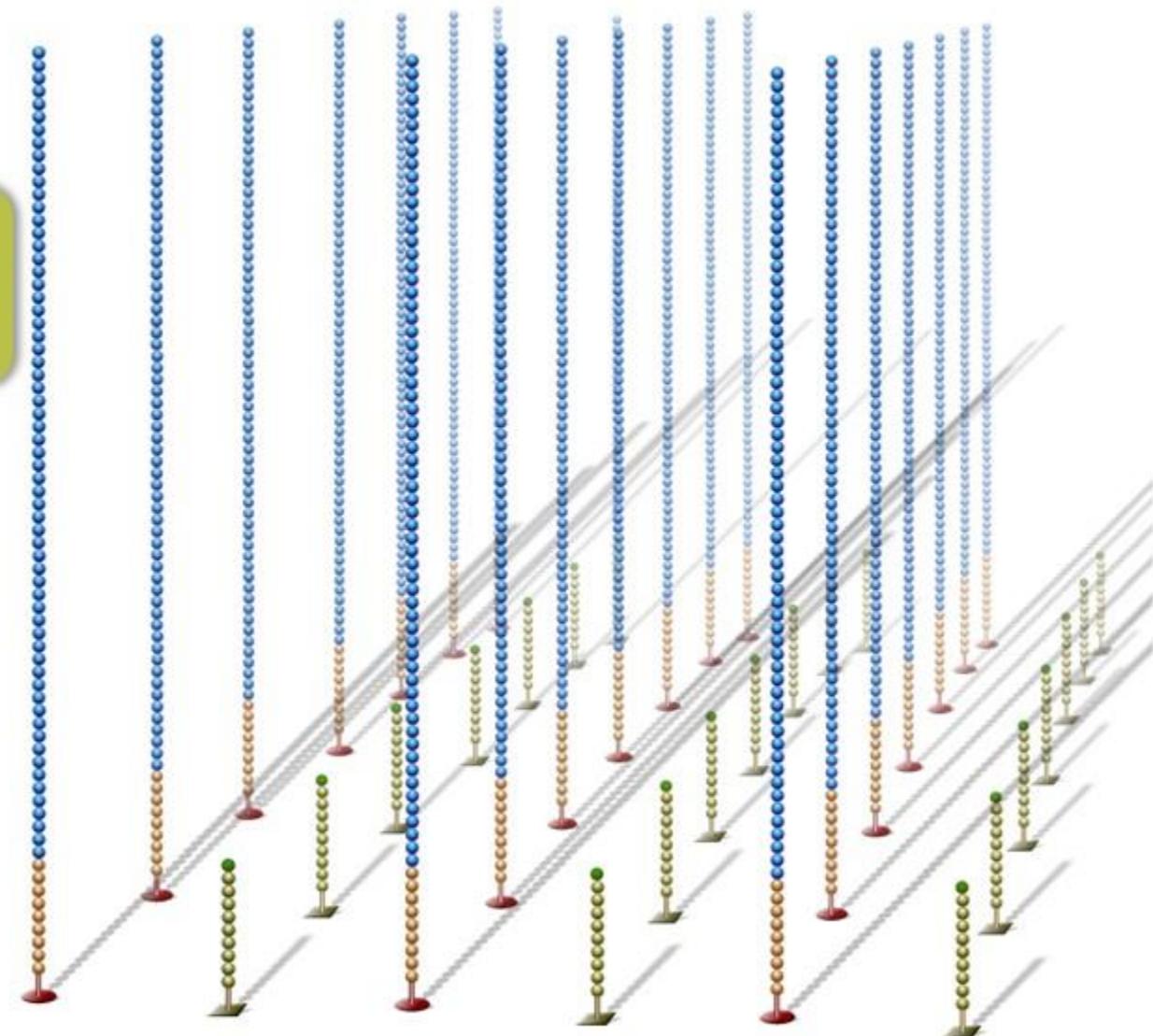
Linearization

dsDNA bridges are denatured



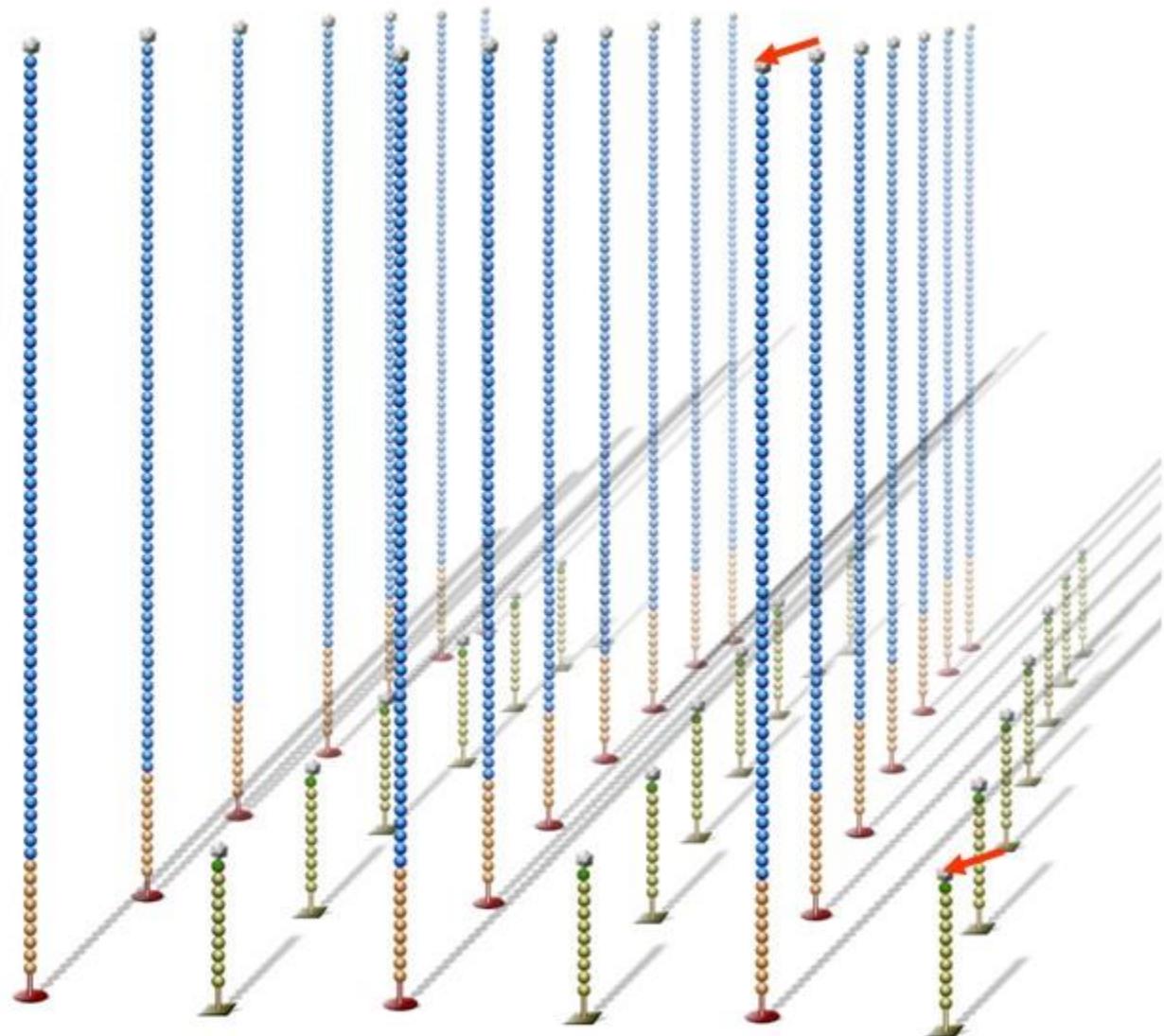
Reverse Strand Cleavage

Reverse strands are cleaved and washed away, leaving a cluster with forward strands only



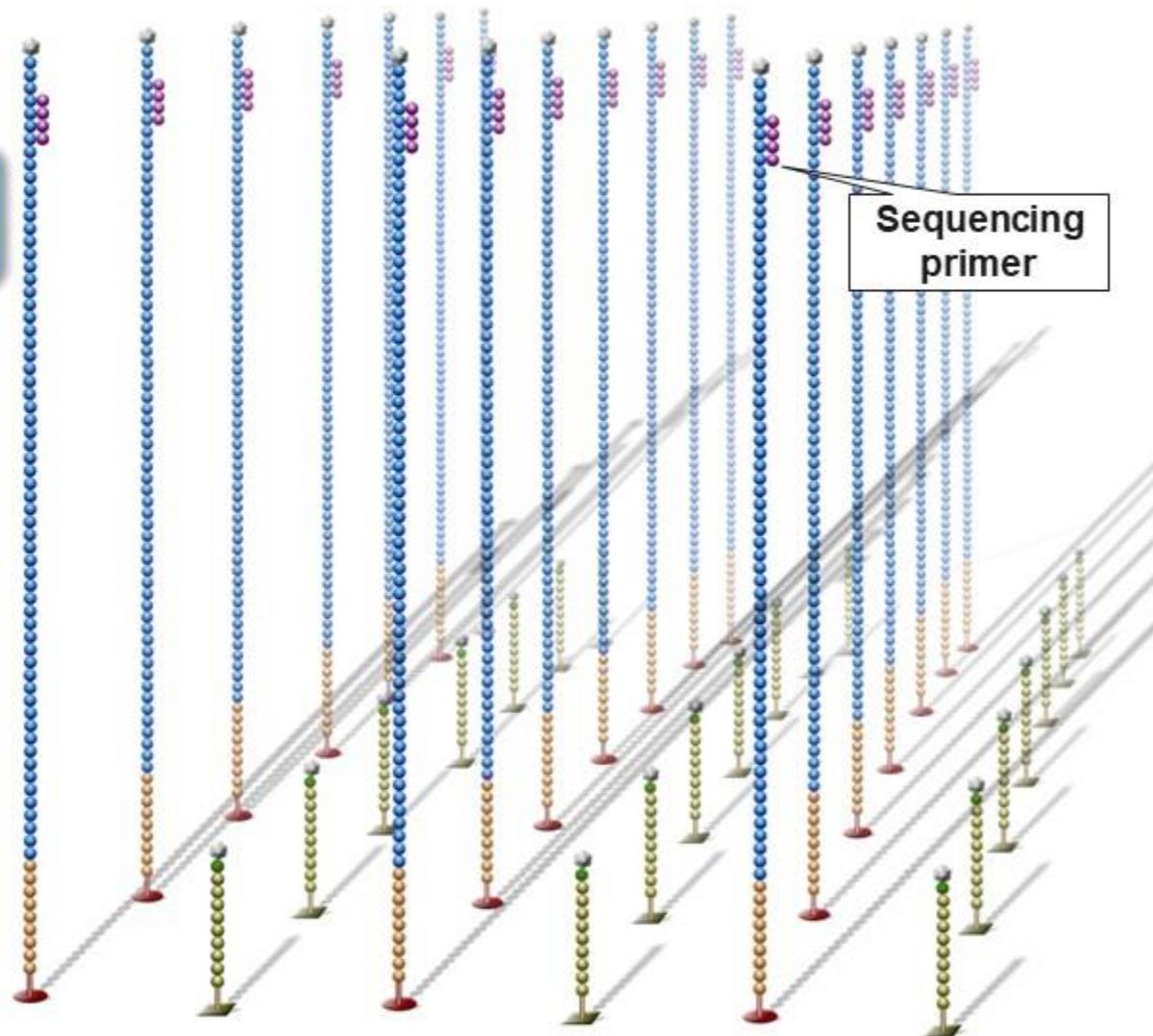
Blocking

Free 3' ends are blocked to prevent unwanted DNA priming



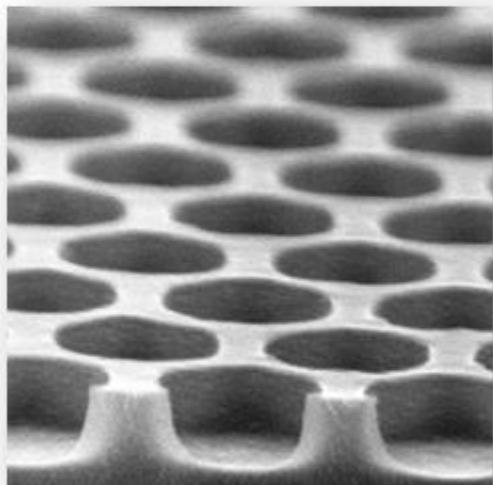
Read 1 Primer Hybridization

Sequencing primer is hybridized to adapter sequence



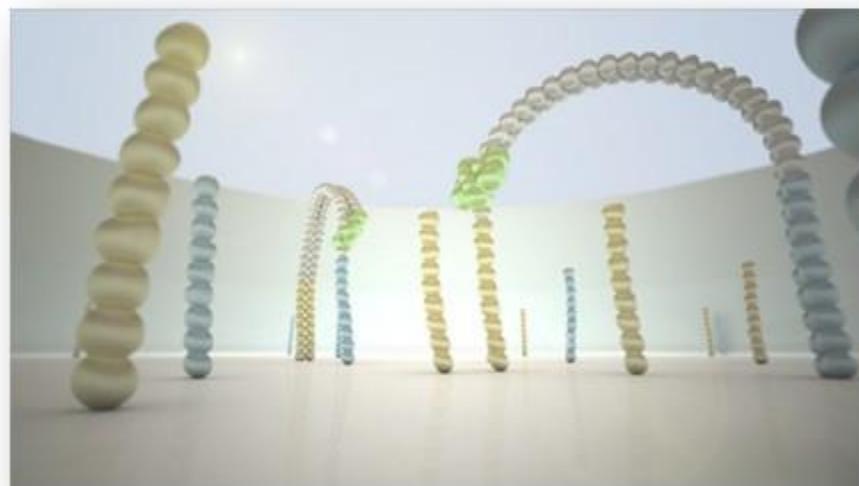
Patterned Flow Cell Technology

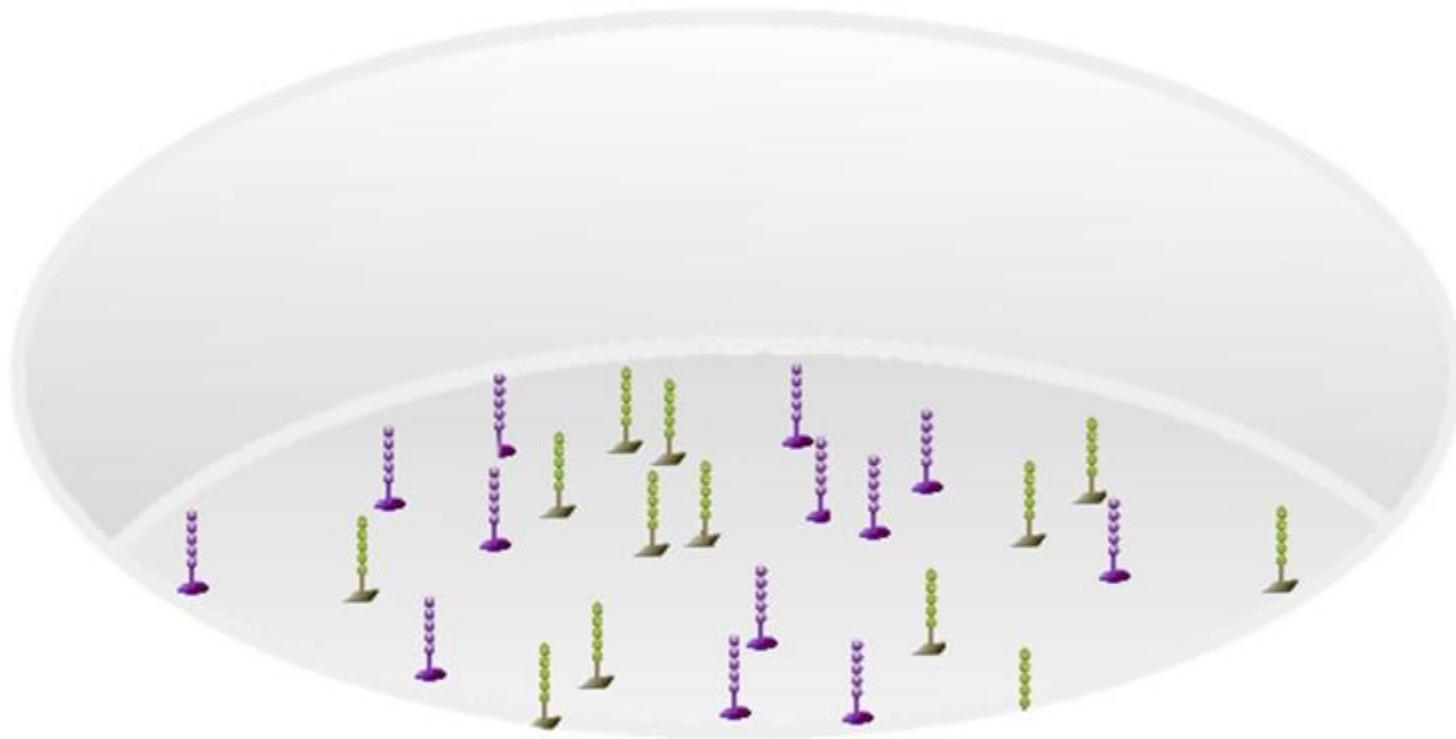
iSeq100, HiSeq4000, Xten, NovaSeq



Billions of ordered wells:

- Clusters contained within ordered wells
- Defined cluster size and spacing
- Increased cluster density
- Simplified imaging
- No pause to complete template generation



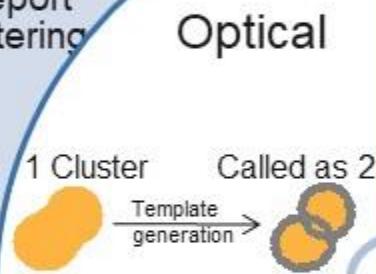


Oligos are not present on surface
between wells to control cluster size

A Review of Sequencing Duplicate Types

- A single cluster that has falsely been called as two by RTA
- Third party tools may report patterned flow cell clustering duplicates as optical duplicates

Not on Patterned Flow Cells

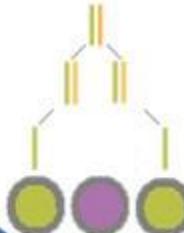


- Duplicates in nearby wells on Patterned Flow Cells
 - During cluster generation a library occupies two adjacent wells

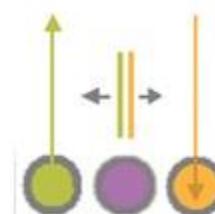
Unique to Patterned Flow Cells

- Duplicate molecules that arise from over amplification during sample prep

PCR



Sister



Complement strands of same library form independent clusters

- Treated as duplicates by some informatic pipelines

Present on all systems

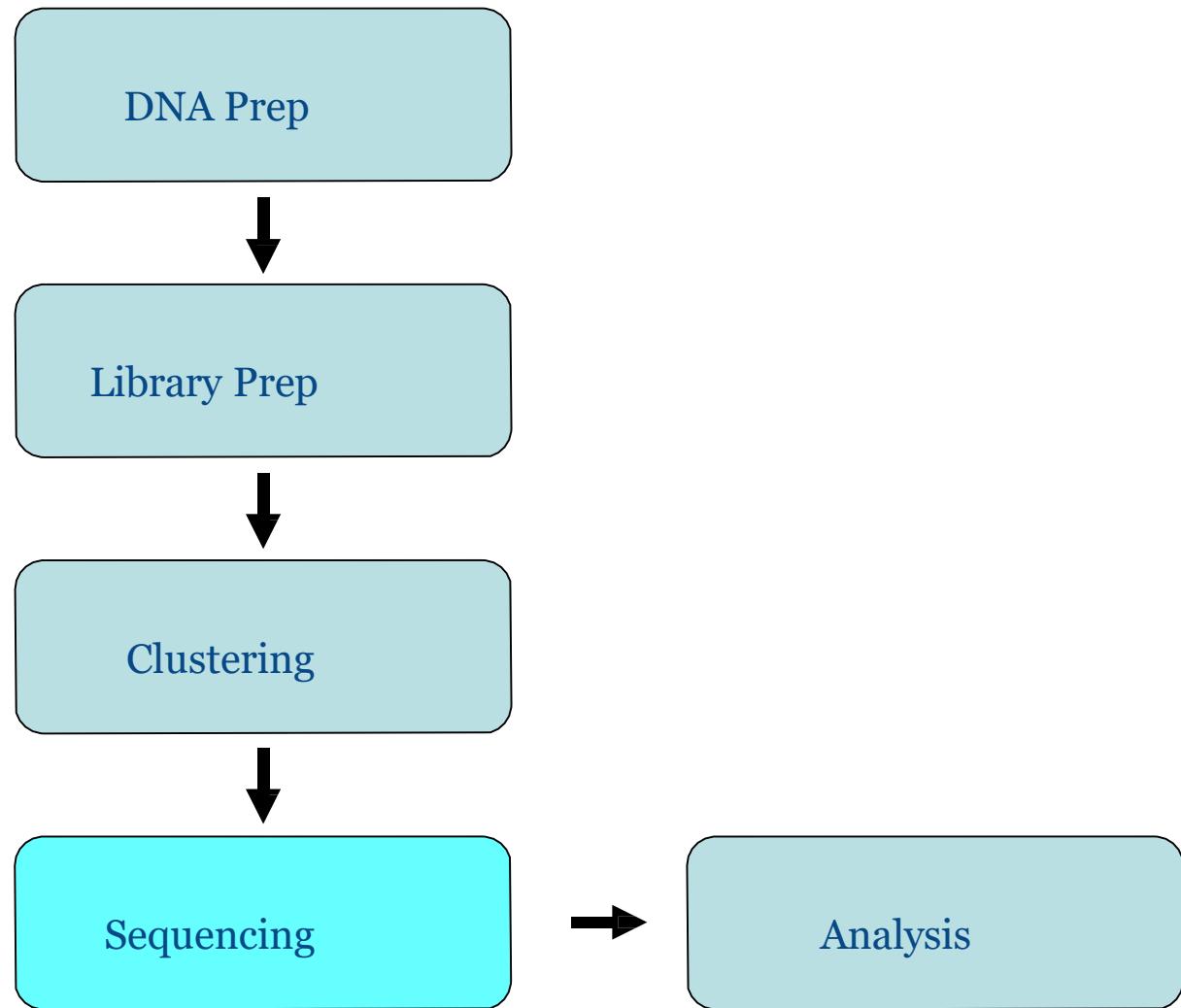


The cBOT



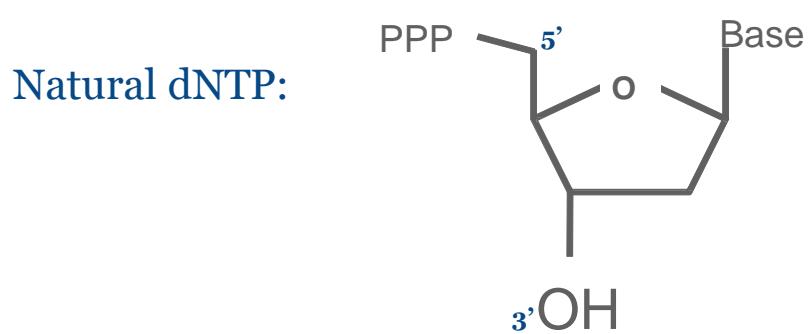


Illumina workflow



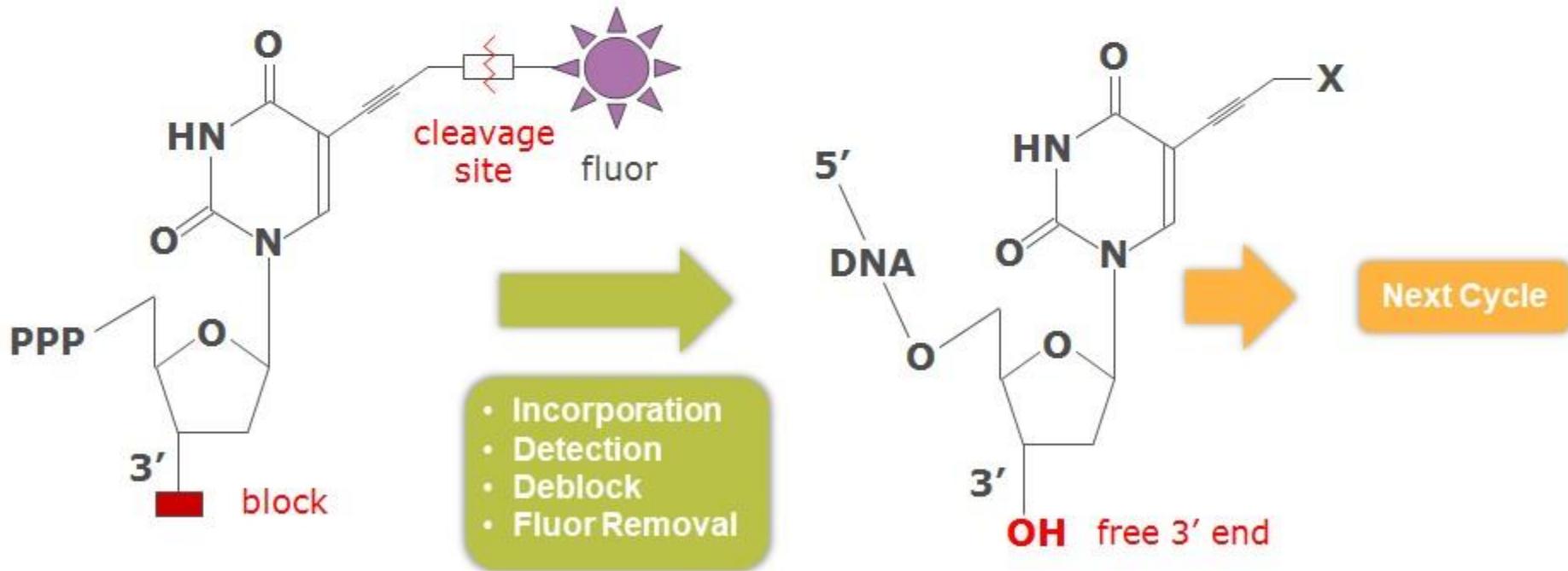


Illumina Modified Nucleotides



Reversible Terminator Chemistry

- All 4 nucleotides in 1 reaction
- Higher accuracy
- No problems with homopolymer repeats

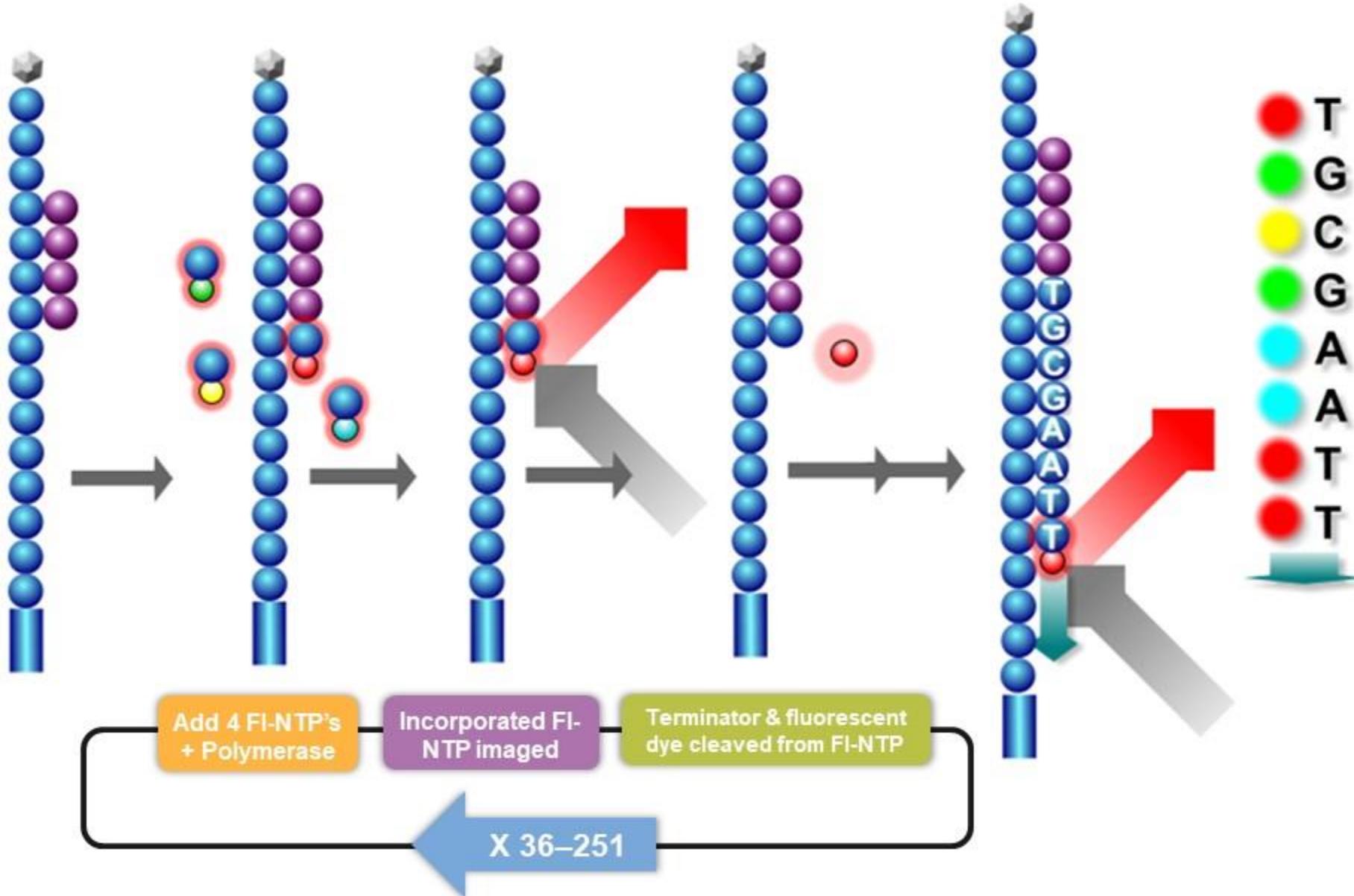




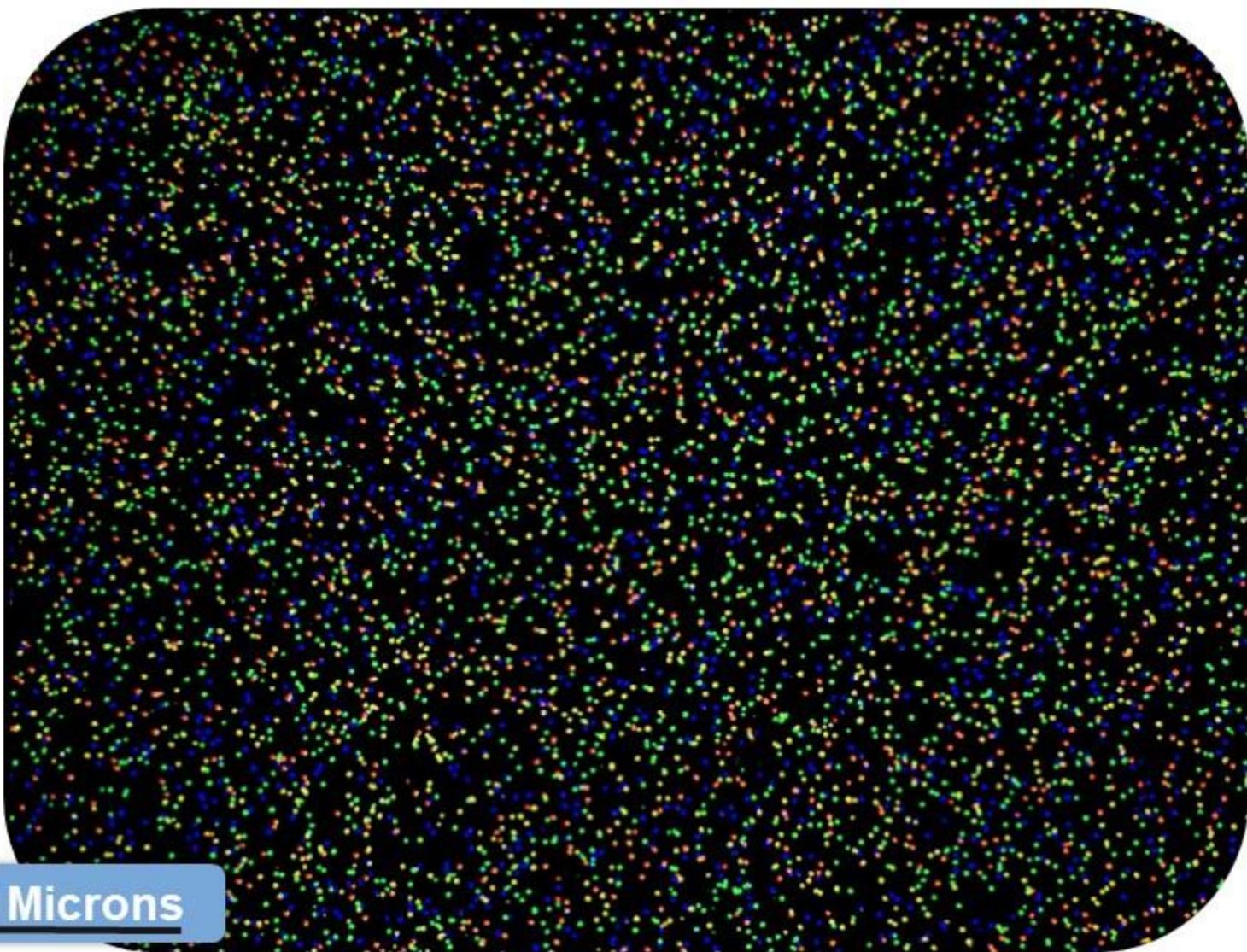
Advantages of True Terminators

- » All four labelled nucleotides in one reaction (faster, simpler)
- » Higher accuracy due to competition
- » Able to push reactions to completion (longer reads)
- » Base-by-base sequencing (reads homopolymer repeats correctly)
- » Requires modified polymerase

Sequencing By Synthesis (SBS)

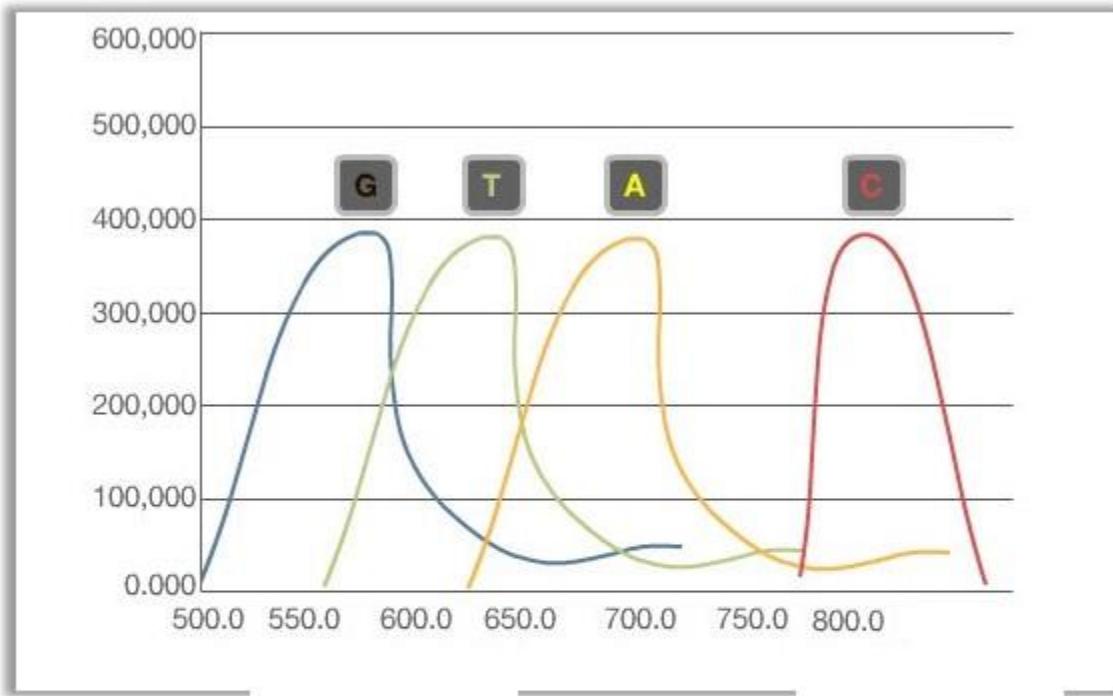


Clusters



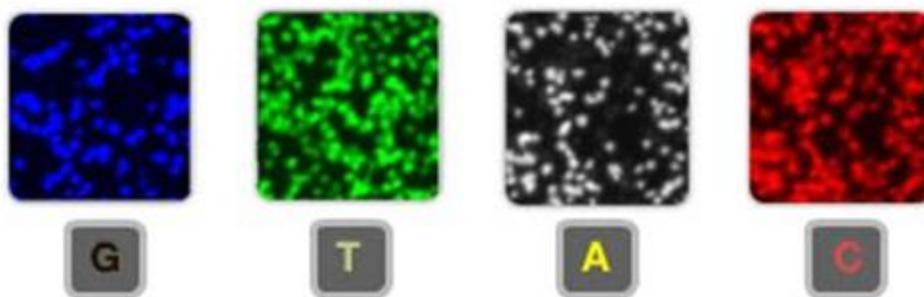
100 Microns

Four Channel SBS Chemistry: GA, HiSeq, MiSeq



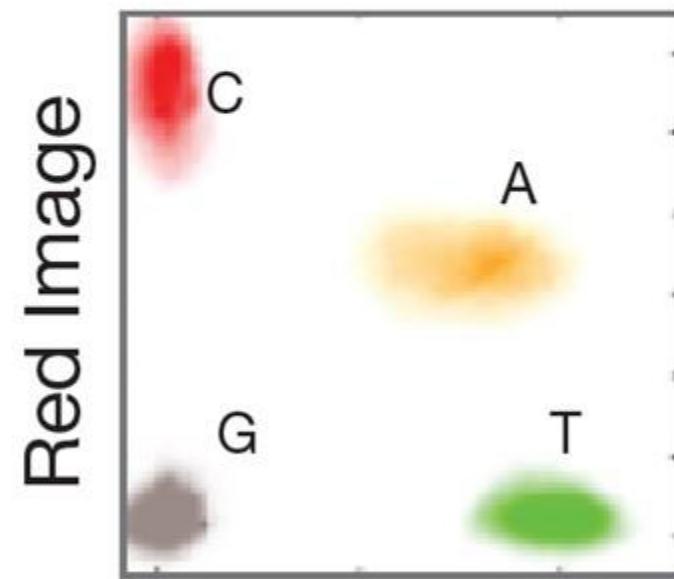
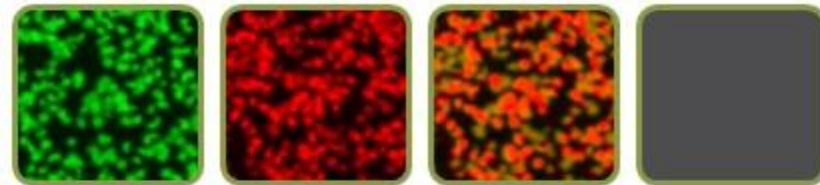
Each of the four DNA bases emit an intensity of a unique wavelength

- Collects four images:
- During each cycle, each cluster appears in only one of four images



Two Channel SBS – NextSeq 500 & Novaseq

- 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



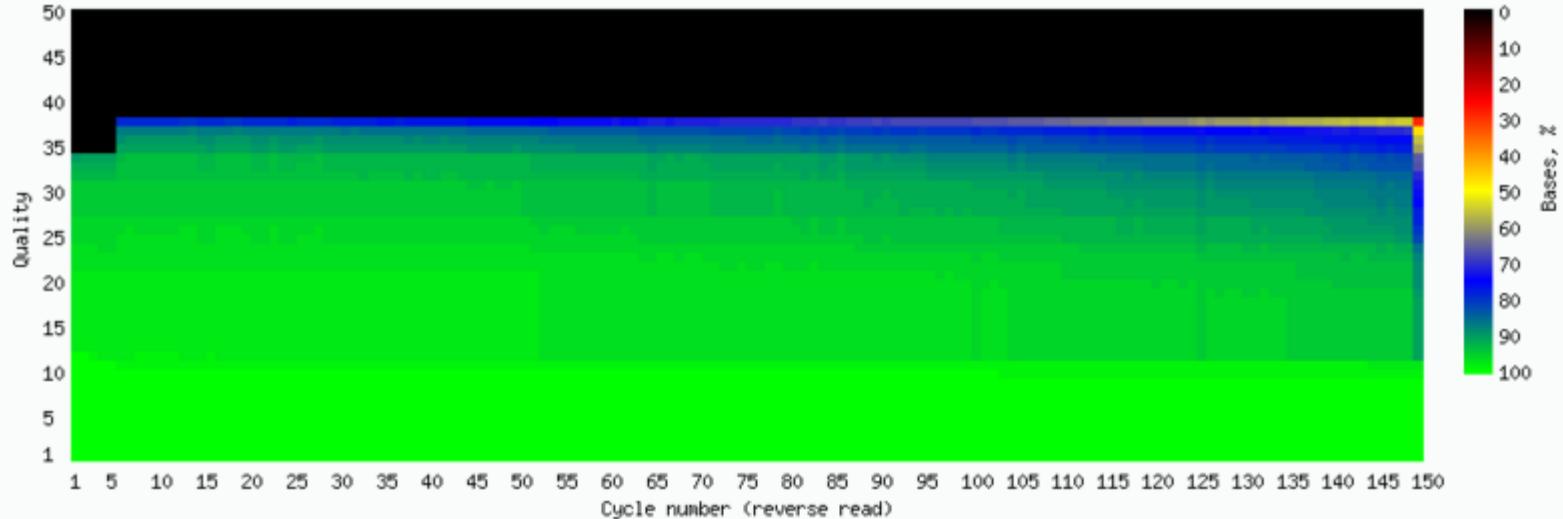
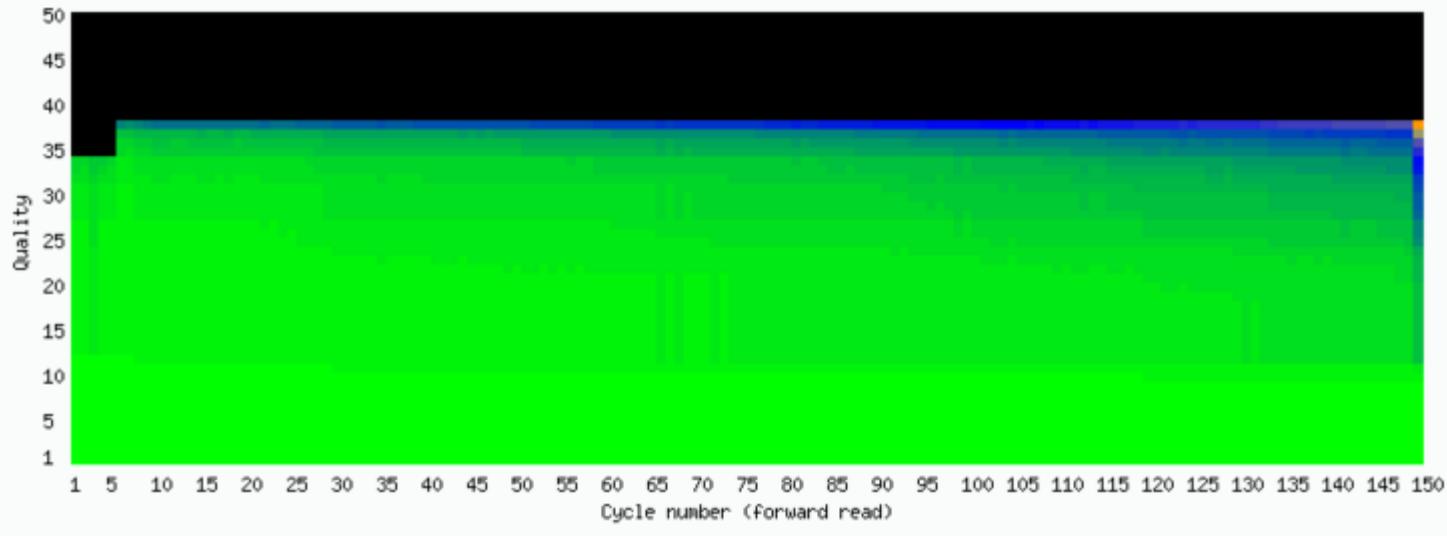
Green Image



Error

- » Sequence quality Q is reported on a log scale
- » Q10 is 1 error in 10
- » Q20 is 1 error in 100
- » Q30 is 1 error in 1000
- » Q40 is 1 error in 10000
- » Q50 is 1 error in 100000

Error limits read length



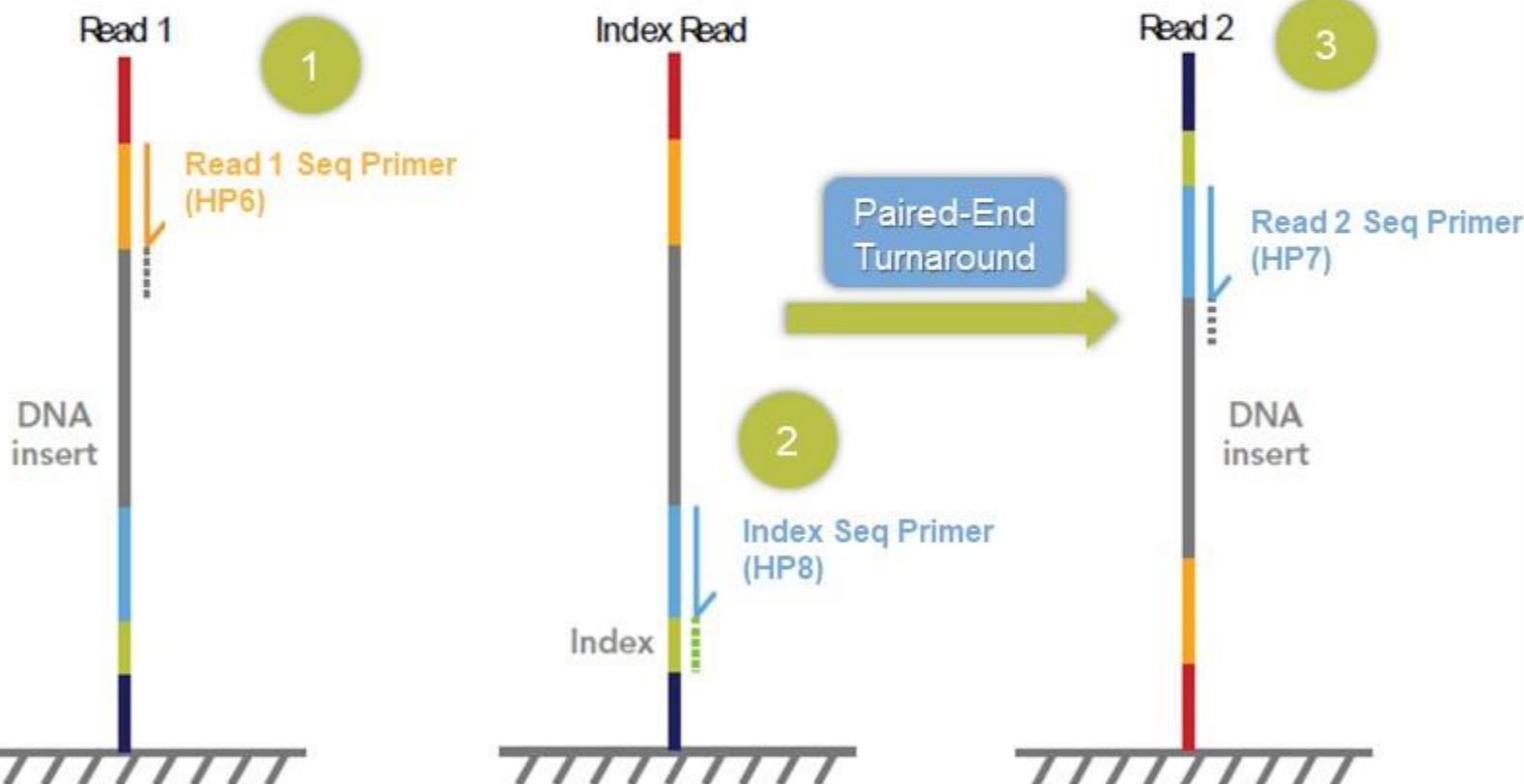


Illumina sequencing with indexing

Single Index Reads

All Platforms

Single indexed sequencing utilizes 3 sequencing reads



Sequencing with Paired-Ends



Reference This is really the best way to do sequencing

Single-reads This is

... is really

... really the

... the best

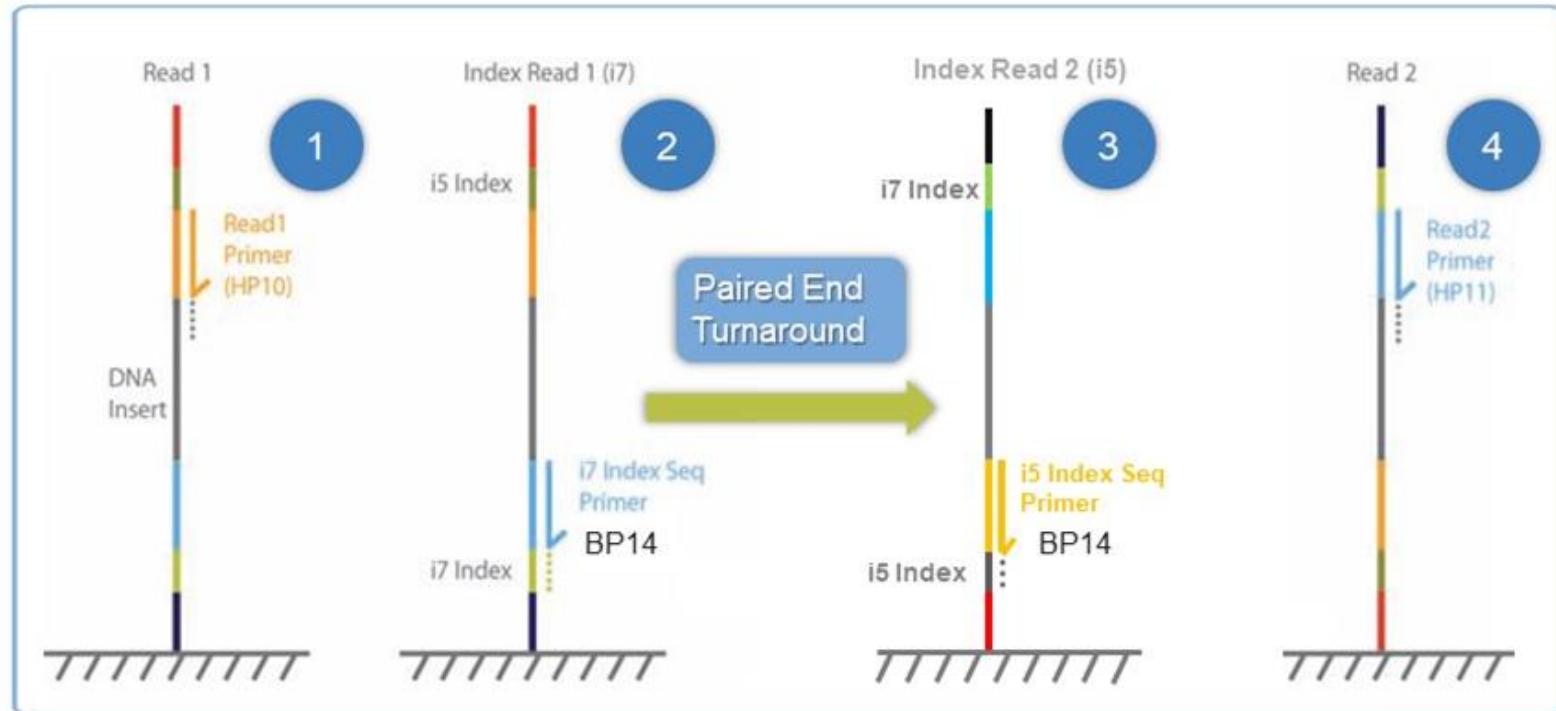
... sequencing

Paired-reads This is (----100 characters-----)sequencing

Assembly becomes easier!!

Dual indexing method B

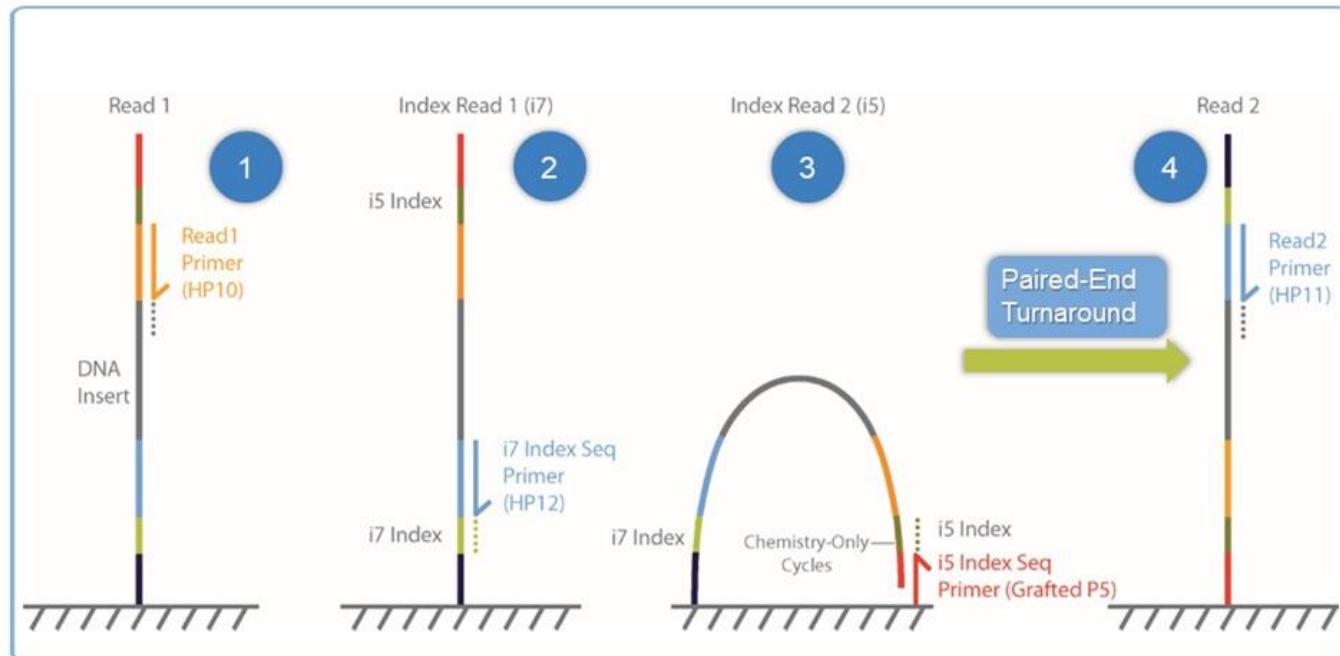
The Index 2 sequencing primer is part of the dual-indexing primer mix for iSeq 100, MiniSeq, and NextSeq. For HiSeq X, HiSeq 4000, and HiSeq 3000, the Index 2 sequencing primer is part of HP14, an indexing primer mix that contains primers for both index reads.



Dual indexing method A

Workflow A

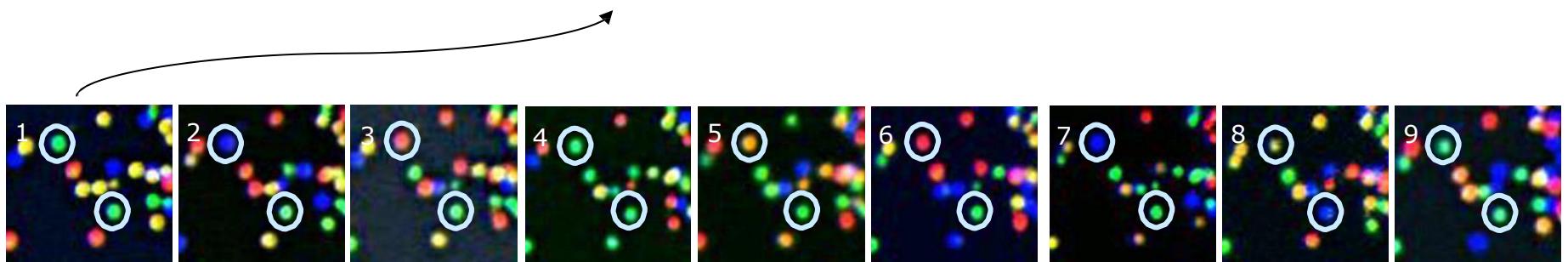
The chemistry applied to the Index 2 Read during a paired-end dual-indexed run on the NovaSeq 6000, MiSeq, HiSeq 2500, or HiSeq 2000 is specific to the paired-end flow cell. Seven additional chemistry-only cycles are required to read the i5 index. This step uses the resynthesis mix, a paired-end reagent, during the Index 2 Read process.





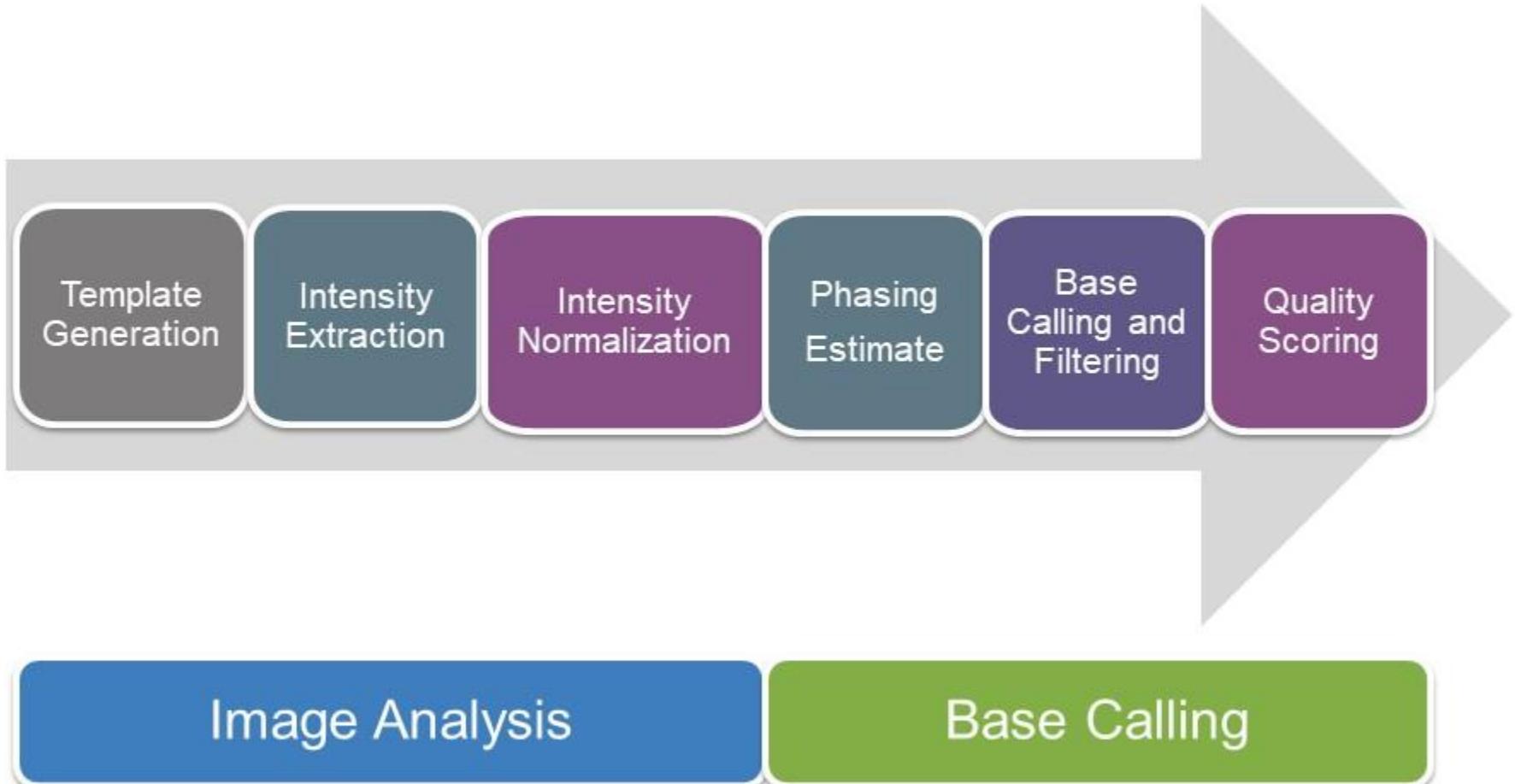
Base Calling From Raw Data

T G C T A C G A T ...



T T T T T T G T ...

Primary Data Analysis Workflow

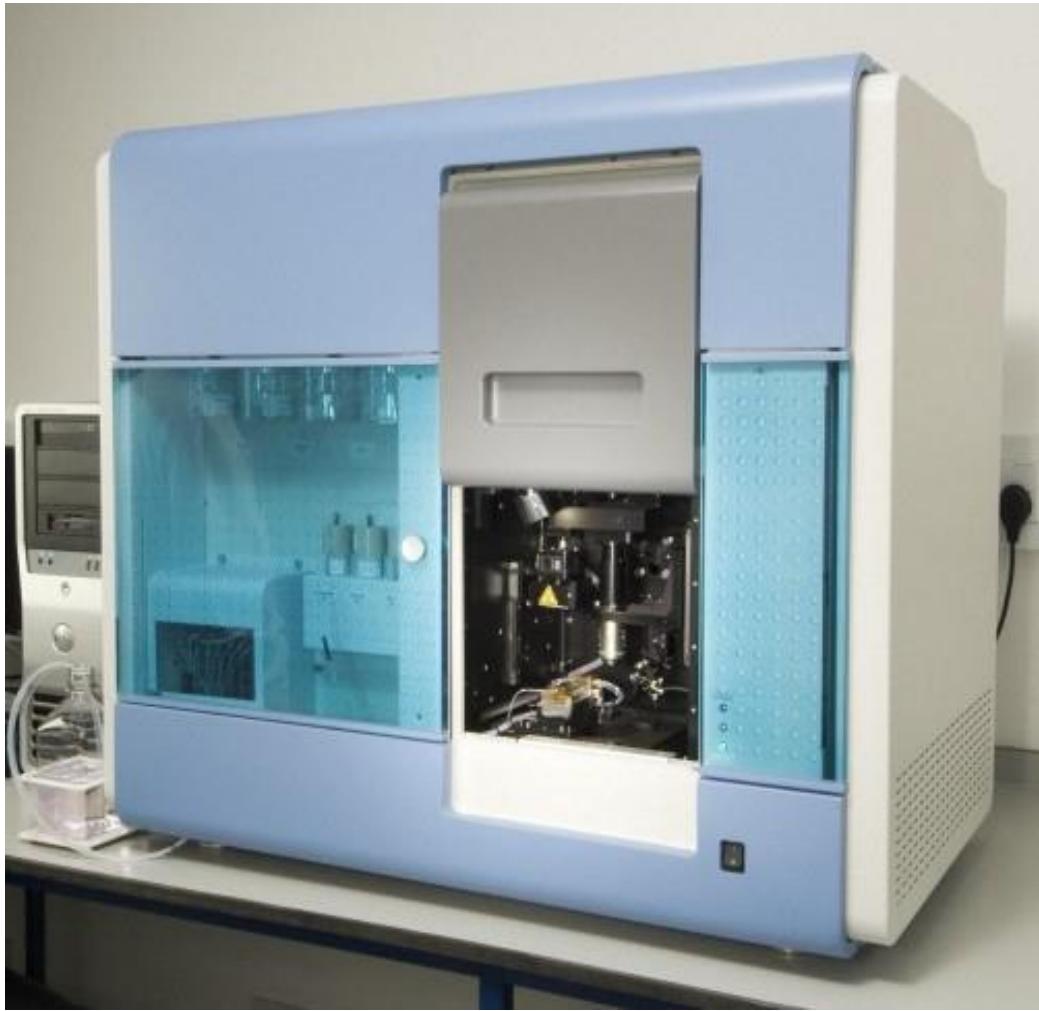




Instruments



GAII Instrument





Genome Analyzer Instrument Overview

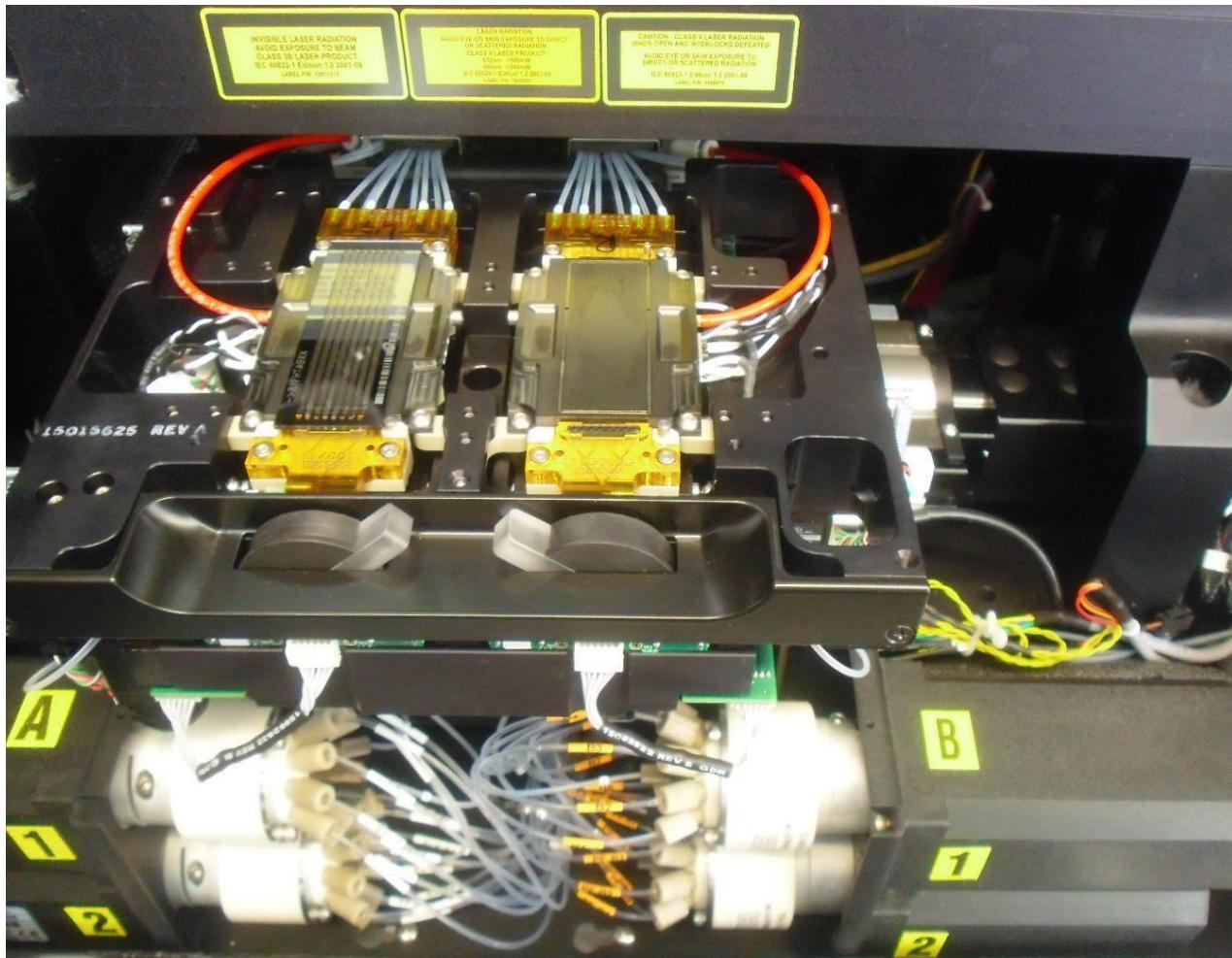
- » Microscope, pump and heater
- » Microscope designed for high-sensitivity detection:
 - » 2 lasers, 4 colours
 - » very accurate positioning
- » 8 fluidics channels per chip
- » Temperature control for enzymology



The HiSeq 2000



HiSeq Flowcell Tray



The Illumina Sequencing Portfolio

Personal Scale



iSeq™ 100

High Throughput

Large WGS | WES | T-OME | Many Samples



NextSeq™ 500



HiSeq™ Series

MiniSeq™

MiSeq™

NextSeq™ 500



Low Throughput

Targeted Sequencing | WGS (Microbes)

Production Scale



NovaSeq™



Benchtop Sequencers



iSeq 100 System



MiniSeq System



Production-Scale Sequencers



NextSeq Series

Popular Applications & Methods

	iSeq 100 System	MiniSeq System	MiSeq Series	NextSeq Series
Key Application				
Large Whole-Genome Sequencing (human, plant, animal)				●
Small Whole-Genome Sequencing (microbe, virus)	●	●	●	●
Exome Sequencing				●
Targeted Gene Sequencing (amplicon, gene panel)	●	●	●	●
Whole-Transcriptome Sequencing				●
Gene Expression Profiling with mRNA-Seq				●
Targeted Gene Expression Profiling	●	●	●	
Long-Range Amplicon Sequencing*	●	●	●	
miRNA & Small RNA Analysis	●	●	●	●
DNA-Protein Interaction Analysis			●	●
Methylation Sequencing				●
16S Metagenomic Sequencing		●	●	●



Benchtop Sequencers



NextSeq Series +



HiSeq Series +



HiSeq X Series[‡]



**NovaSeq 6000
System**

Popular Applications & Methods

Key Application

Key Application

Key Application

Key Application

Large Whole-Genome Sequencing (human, plant, animal)



Small Whole-Genome Sequencing (microbe, virus)



Exome Sequencing



Targeted Gene Sequencing (amplicon, gene panel)



Whole-Transcriptome Sequencing



Gene Expression Profiling with mRNA-Seq



miRNA & Small RNA Analysis



DNA-Protein Interaction Analysis



Methylation Sequencing



Shotgun Metagenomics



iSeq™ 100

Your new lab partner



Complement your **growing** NGS instrument **fleet**

Multiple applications on a low-throughput **Illumina NGS platform**

Affordable NGS system

iSeq™ 100 System

\$19,900 USD List

iSeq™ 100 i1 Reagent
Four Million Read Sequencing Kit

\$625 per kit
USD List or **\$2,375** four pack
USD List



iSeq™ 100 Specifications

1.2 Gigabases 1.2 billion nucleotides sequenced per run	4 Million 4 million fragments sequenced per run	9–17 Hours Total time sequencing	300 Base Pairs Up to 300 bp fragments sequenced per read
--	--	---	---

Run Configuration	Reads (M)	Output	Run Time
1x36 bp	4	144 Mb	9 hrs
1x50 bp	4	200 Mb	9 hrs
1x75 bp	4	300 Mb	10 hrs
2x75 bp	4	600 Mb	13 hrs
2x150 bp	4	1.2 Gb	17 hrs

A Closer Look At 1-Dye Chemistry

Cluster Generation and SBS

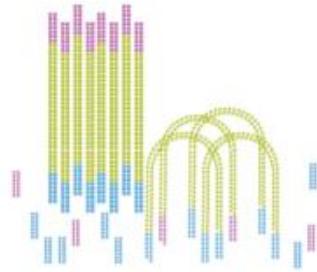


What's the Same?

Library Preparation



Cluster Growth



Sequencing



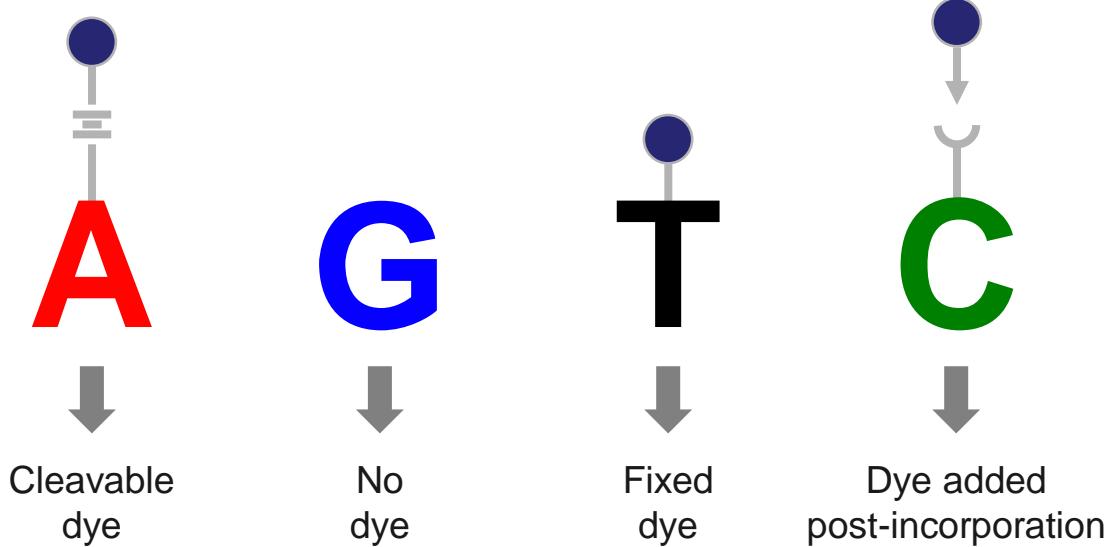
The iSeq™ 100 System **still utilizes the Illumina SBS chemistry**,
where each base is added one at a time

A Closer Look At 1-Dye Chemistry

Cluster Generation and SBS



What's the Different?



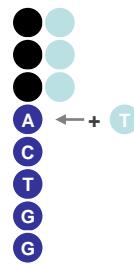
Nucleotides are labeled with a single dye, with the exception of the **G nucleotide**

A Closer Look At 1-Dye Chemistry

SBS and Imaging

Sequencing by Synthesis

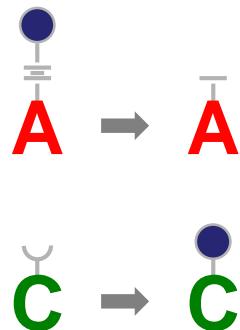
Sequencing Cycle



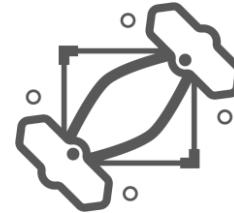
Incorporation



Imaging



Chemistry

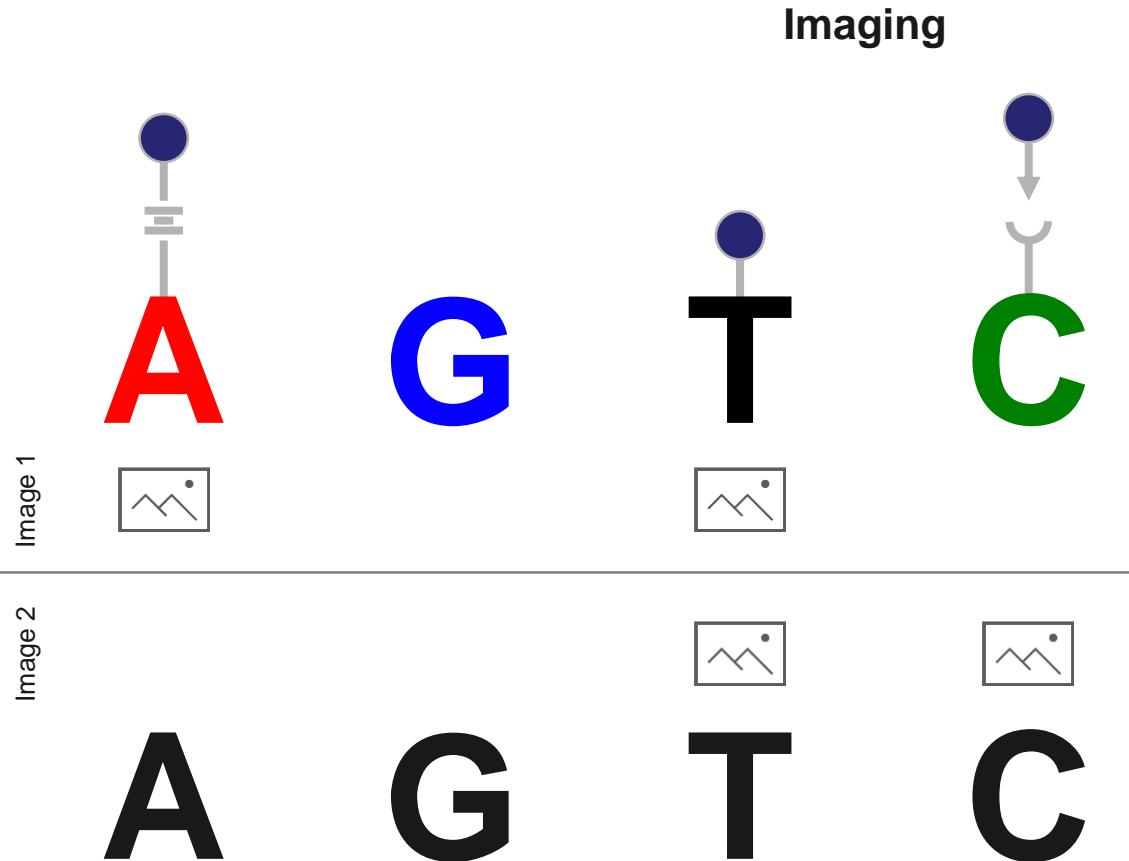


Imaging

An intermediate chemistry step, which **removes the dye from the A nucleotide** and **adds a dye to the C nucleotide**, separates the two images

A Closer Look At 1-Dye Chemistry

SBS and Imaging



Using the two images, the iSeq™ 100 innovative data processing approach uniquely determines **which nucleotide was added to the growing template strand**

MiniSeq



- » Single lane, 2x 150 reads, upto 7.5Gb
- » Two output modes 8M or 25 M reads
- » Less than 24 hour run time
- » Cheapest Illumina sequencer \$49.5K
- » Runs from \$500 to \$1500
- » 2 Colour chemistry



Illumina MiSeq

- » Single-lane version of HiSeq
- » Read length: 2 x 300 (1 -2 days)
- » Error rate 0.1%
- » Yield: > 10 Gb / run
- » \$99,000 capital cost
- » \$100 / Gb



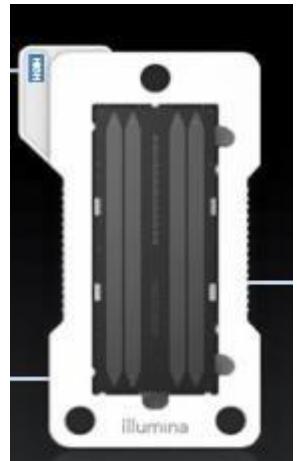
Illumina MiSeq

- » Single-lane version of HiSeq
- » Read length: 2 x 300 (1 -2 days)
- » Error rate 0.1%
- » Yield: > 10 Gb / run
- » \$99,000 capital cost
- » \$100 / Gb



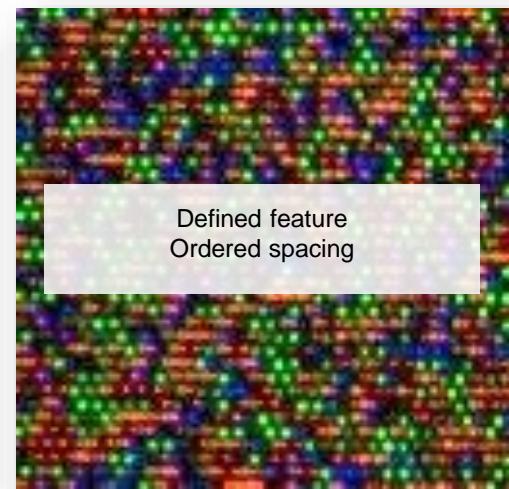
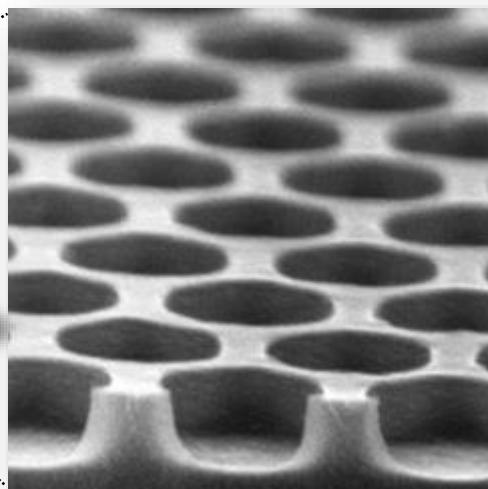
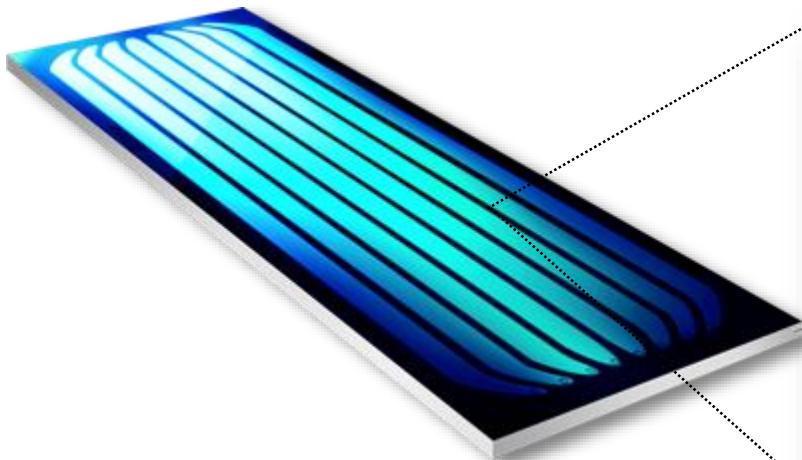
Illumina NextSeq 500

- » Uses only 2 dyes / 2 images not 4
- » Read length: 2 x 150 (1.25 days)
- » Yield: 120 Gb / run
- » \$250,000 capital cost
- » \$30 / Gb



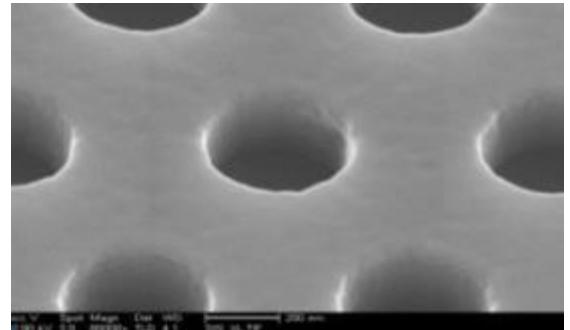
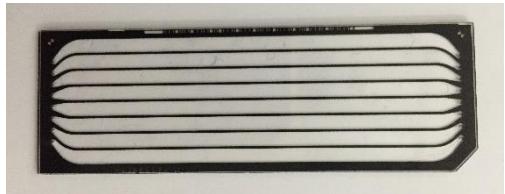
Hiseq 3000/4000

- » Single or double flowcell instrument
- » Suitable for all applications inc. exomes
- » Patterned flowcell (aka “PFCT”)
- » \$20/Gb



Illumina HiSeq X

- » Ultra-high throughput
- » Read length: 2x150 (3 days)
- » Yield: 1.6 Tb / run
- » \$1,000,000 capital cost x 10
- » \$ 7 / Gb
- » The \$1000 Genome



HiSeq X Ten Performance



1.8T | 6B READS | PE150 | <3 DAYS

UP TO 18,000 GENOMES* | YEAR | X TEN

\$1,000 GENOME* | \$800 CONSUMABLES



2017: NovaSeq

NovaSeq System Configurations

Max Output / Flow Cell: **0.5 Tb** **1 Tb** **2 Tb** **3 Tb**

NovaSeq 5000
\$850K USD

NovaSeq 6000
\$985K USD

NovaSeq 5000 Flow Cells

NovaSeq 6000 Flow Cells

NovaSeq 5000

NovaSeq 6000

**wellcome
sanger
institute**



NovaSeq Series

Any Genome. Any Method. Any Scale.

NovaSeq 5000



PE 150 | Q30 \geq 75%

167 – 2000 Gb



NovaSeq 6000



167 – 6000 Gb

1.6 – 6.6B

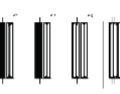


Fast



Fastest (40 Hr. for 2T Run)

2 Types



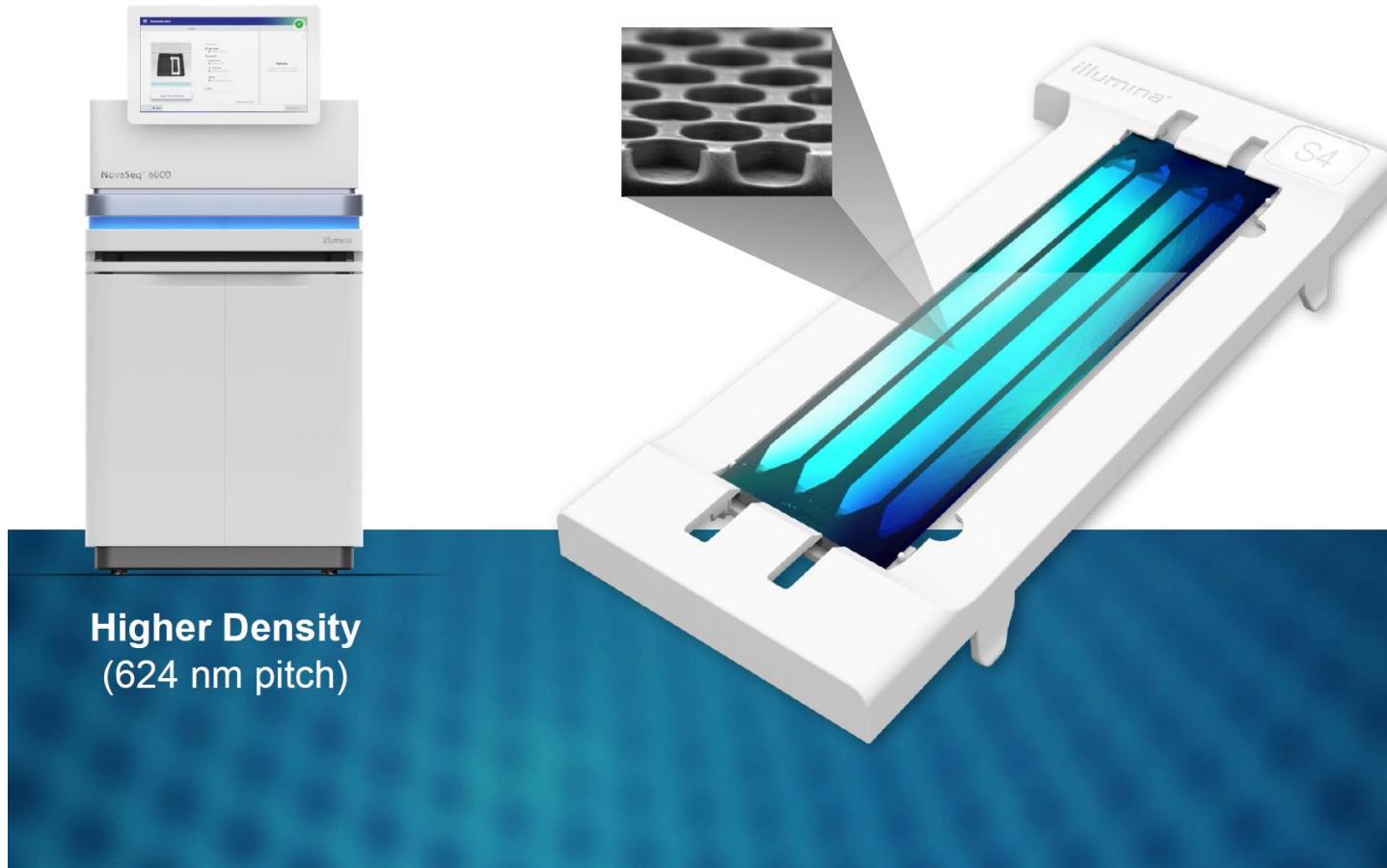
4 Types

For Research Use Only. Not for use in diagnostic procedures.

Output range shown based on currently unreleased flow cells



Nanofabricated Flow Cells

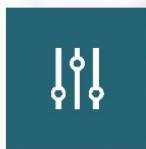




Scalable Throughput

Complete studies faster and more economically

Single flow cell output (1 or 2 can run simultaneously)



Run times:
<1 to ~2.5d
based on
system, FC
and read
length

Configure
output to
match your
application
and study
size

Flow Cell Type	NovaSeq 5000	NovaSeq 6000	Reads per Flow Cell	Output (Gb) per Flow Cell		
				100 cycles	200 cycles	300 cycles
S4*		✓	10B			3000
S3*		✓	6.6B			2000
S2	✓	✓	3.3 B	333	666	1000
S1*	✓	✓	1.6 B	167	333	500

*S1, S2 and S4 flow cells not currently released



Streamlined Operation

Increase lab efficiency with a simplified workflow



Cartridge based reagents reduce hands on time and prevent misloading



RFID encoded consumables provide traceability and ensure compatibility



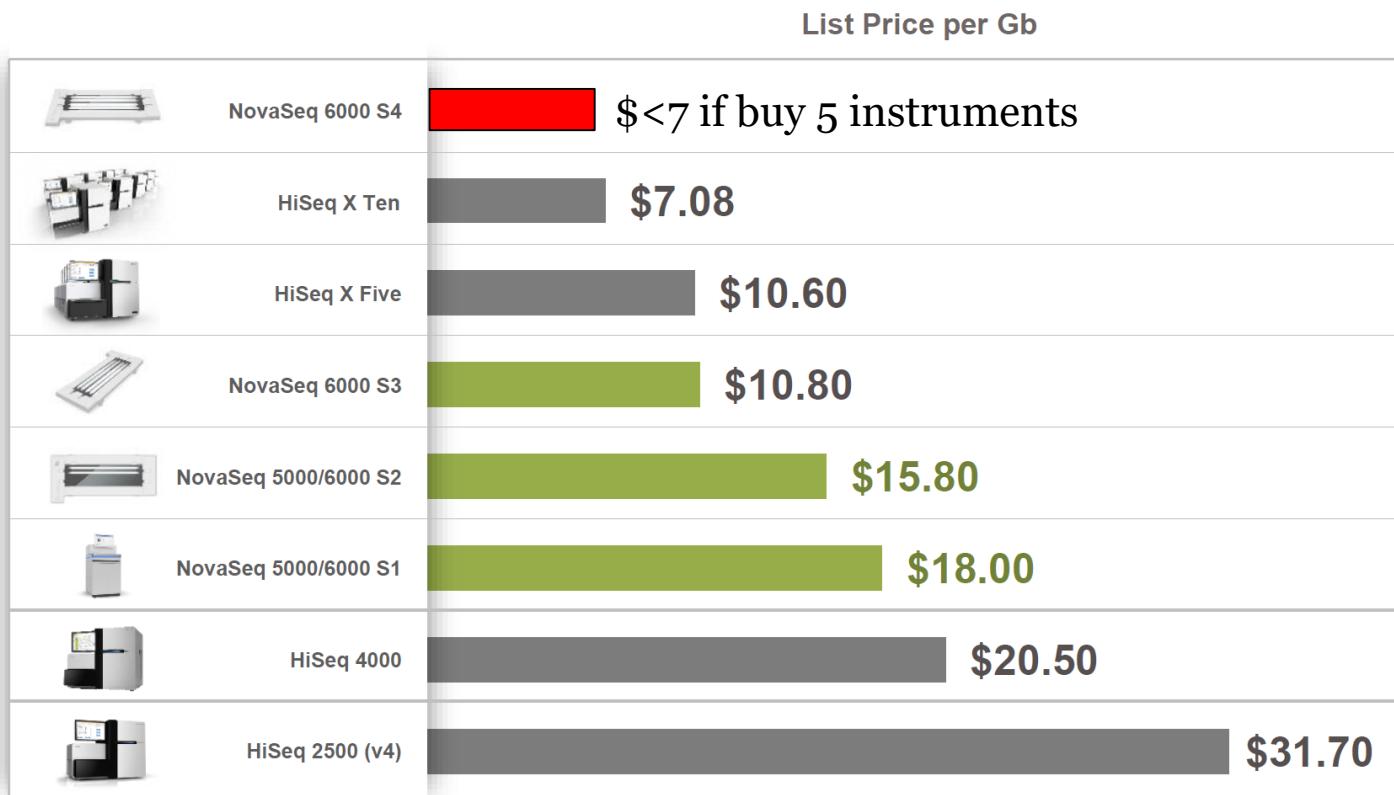
Onboard cluster gen reduces hands on time and run variability





NovaSeq Series

Compelling price per data point enables highly-powered studies



HiSeq 2500 based on 250 cycle kit. all others based on 300 cycle kit

NovaSeq Performance and Data Quality

- Data quality as good as HiSeq with initial release – significant opportunity for further improvements
 - Major R&D focus on further optimizing 2-channel chemistry with patterned flow cells
- High data quality enabled by superior optics and reformulated chemistry
 - Diffraction-limited performance optics
 - New surface chemistry, dye-sets and enzymes

*NovaSeq Prototype Instrument running S2 flow cell



Sequencing kits



iSEQ100

» 2 x 150



MiniSeq kits

- » 25M reads High output
 - » 75 cycle
 - » 150 cycle
 - » 300 cycle
- » 8M reads
 - » 300 cycle



MiSeq Kits

MiSeq System Performance Parameters

MiSeq Reagent Kit v2

Read Length	Total Time*	Output
1 × 36 bp	~4 hours	540–610 Mb
2 × 25 bp	~5.5 hours	750–850 Mb
2 × 150 bp	~24 hours	4.5–5.1 Gb
2 × 250 bp	~39 hours	7.5–8.5 Gb

Reads Passing Filter†

Single Reads	12–15 M
Paired-End Reads	24–30 M

Quality Scores‡‡

- > 90% bases higher than Q30 at 1 × 36 bp
- > 90% bases higher than Q30 at 2 × 25 bp
- > 80% bases higher than Q30 at 2 × 150 bp
- > 75% bases higher than Q30 at 2 × 250 bp

MiSeq Reagent Kit v3

Read Length	Total Time*	Output
2 × 75 bp	~21 hours	3.3–3.8 Gb
2 × 300 bp	~56 hours	13.2–15 Gb

Reads Passing Filter†

Single Reads	22–25 M
Paired-End Reads	44–50 M

Quality Scores‡‡

- > 85% bases higher than Q30 at 2 × 75 bp
- > 70% bases higher than Q30 at 2 × 300 bp

* Total times include cluster generation, sequencing, and basecalling on a MiSeq system enabled with dual surface scanning.

† Install specifications based on Illumina PhiX control library at supported cluster densities between 467–583 k/mm² clusters passing filter for v2 chemistry and 727–827 k/mm² clusters passing filter for v3 chemistry. Actual performance parameters may vary based on sample type, sample quality, and clusters passing filter.

‡‡ The percentage of bases > Q30 is averaged across the entire run.

bp = base pairs, Mb = megabases, Gb = gigabases, M = millions

Nextseq 500 kits

NextSeq 500 Kits	High Output† up to 400M Clusters	Mid Output† up to 132M Clusters
300 cycles	<ul style="list-style-type: none">▪ Whole Genome▪ <i>de novo</i>▪ Exome	<ul style="list-style-type: none">▪ Targeted Panels▪ Exome
150 cycles	<ul style="list-style-type: none">▪ Whole Transcriptome▪ mRNA-Seq▪ Exome	<ul style="list-style-type: none">▪ Whole Transcriptome▪ mRNA-Seq▪ Exome
75 cycles	<ul style="list-style-type: none">▪ Gene Expression Profiling▪ NIPT	



HiSeq kits

KIT NAME	OUTPUT MAX (PER 2 FLOW CELL)	NO. OF READS	MAX READ LENGTH	TIME
HiSeq SBS V4 Kits	Up to 1 Tb	Up to 4 billion	2 x 125 bp	6 days

- 10 genomes
- 150 Nextera Rapid Capture exomes
- 80 whole transcriptome RNA samples

HISEQ X TEN	DUAL FLOW CELL	SINGLE FLOW CELL
Output/Run	1.6-1.8 Tb	800-900 Gb
Read Passing Filter†	5.3-6 billion	2.6-3 billion
Supported Read Length	2x150	
Run Time	< 3 days	
Quality	≥ 75% of bases above Q30 at 2 x 150 bp	



HiSeq kits 1

HiSeq System Performance Parameters

High Output Run Mode*

	HISEQ SBS V4 SPECIFICATIONS			TRUSEQ SBS V3		
Read length	Dual Flow Cell	Single Flow Cell	Dual Flow Cell Run Time	Dual Flow Cell	Single Flow Cell	Dual Flow Cell Run Time
1x36	128-144 Gb	64-72 Gb	29 hrs	95-105 Gb	47-52 Gb	2 days
2x50	360-400 Gb	180-200 Gb	2.5 days	270-300 Gb	135-150 Gb	5.5 days
2x100	720-800 Gb	360-400 Gb	5 days	540-600 Gb	270-300 Gb	11 days
2x125	900-1 Tb	450-500 Gb	6 days	N/A	N/A	N/A
Reads Passing Filter (8 lanes per flow cell)	Up to 4 billion single read or 8 billion paired-end reads	Up to 2 billion single read or 4 billion paired-end reads		Up to 3 billion single read or 6 billion paired-end reads	Up to 1.5 billion single read or 3 billion paired-end reads	
Quality	Greater than 85% of bases above Q30 at 2x50 bp Greater than 80% of bases above Q30 at 2x100 bp Greater than 80% of bases above Q30 at 2x125 bp			Greater than 85% of bases above Q30 at 2x50 bp Greater than 80% of bases above Q30 at 2x100 bp		

*Install specifications based on Illumina PhiX control library at supported cluster densities (between 610-678 K clusters/mm² passing filter using TruSeq v3 Kits or 870-930 K clusters/mm² passing filter using HiSeq v4). Run times for high output mode correspond to sequencing only. Performance may vary based on sample quality, cluster density, and other experimental factors.



HiSeq kits 2

Rapid Run Mode*

HISEQ RAPID SBS KIT V2 SPECIFICATIONS

Read length	Dual Flow Cell	Single Flow Cell	Dual Flow Cell Run Time
1x36	18-22 Gb	9-11 Gb	7 hr
2x50	50-60 Gb	25-30 Gb	16 hr
2x100	100-120 Gb	50-60 Gb	27 hr
2x150	150-180 Gb	75-90 Gb	40 hr
2x250	250-300 Gb	125-150 Gb	60 hr
Reads Passing Filter (2 lanes per flow cell)	Up to 600 million single read or 1.2 billion paired-end reads	Up to 300 million single read or 600 million paired-end reads	
Quality	Greater than 85% of bases above Q30 at 2x50 bp Greater than 80% of bases above Q30 at 2x100 bp Greater than 75% of bases above Q30 at 2x250 bp		

*Install specifications based on Illumina PhiX control library at supported cluster densities (between 700-820 K clusters/mm² passing filter using HiSeq Rapid v2 Kits). Run times for rapid run mode correspond to on-board cluster generation (1.5 hr) and sequencing. Performance may vary based on sample quality, cluster density, and other experimental factors. Early HiSeq 2000 instruments will run slightly slower when upgraded to a HiSeq 2500.

HiSeq 3000/4000

Performance Parameters

	HISEQ 3000 SYSTEM	HISEQ 4000 SYSTEM
No. of Flow Cells per Run	1	1 or 2
Data Yield: 2 x 150 bp 2 x 75 bp 1 x 50 bp	630-750 Gb 315-375 Gb 105-125 Gb	1300-1500 Gb 650-750 Gb 215-250 Gb
Clusters Passing Filter (Single Reads) (8 lanes per flow cell)	2.1-2.5 billion	4.3-5 billion
Quality Scores	≥ 75% bases above Q30 at 2 x 150 bp	≥ 75% bases above Q30 at 2 x 150 bp
Daily Throughput	> 200 Gb	> 400 Gb
Run Time	< 1-3.5 days	< 1-3.5 days
Human Genomes per Run*	up to 6	up to 12
Exomes per Run**	up to 90	up to 180
Transcriptomes per Run***	up to 50	up to 100
Supported Library Prep Kits	DNA: TruSeq Nano DNA, TruSeq DNA PCR-Free RNA: TruSeq RNA v2, TruSeq Stranded mRNA, TruSeq Stranded Total RNA, TruSeq RNA Access Exome: Nextera Rapid Capture Exome	

Install specifications based on Illumina PhiX control library at supported cluster densities (between 1265-1422 K clusters/mm² passing filter). Run times correspond to sequencing only. Performance may vary based on sample quality, cluster density, and other experimental factors.

*Assumes >30x coverage of a human genome.

**Assumes 4 Gb per exome and 2 x 75 bp reads using Nextera Rapid Capture Exome.

***Assumes 50 million reads per sample.



Costs

- » 1st human genome:
~ \$500 Million
- » Current capillary cost :
~ \$10 Million
- » Illumina cost:
~ \$10,000 all-in
~ \$3,000 reagents only
~ \$1,000 on Xten





Future

- » Lower cost/higher throughput/faster
- » Increases in density readily achievable
- » Single molecule sensing

- » Sample to answer
- » Clinical solutions
- » Long reads?
- » Cloud computing

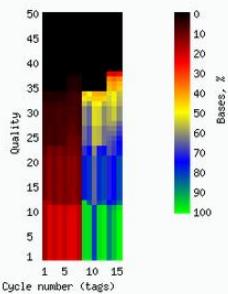
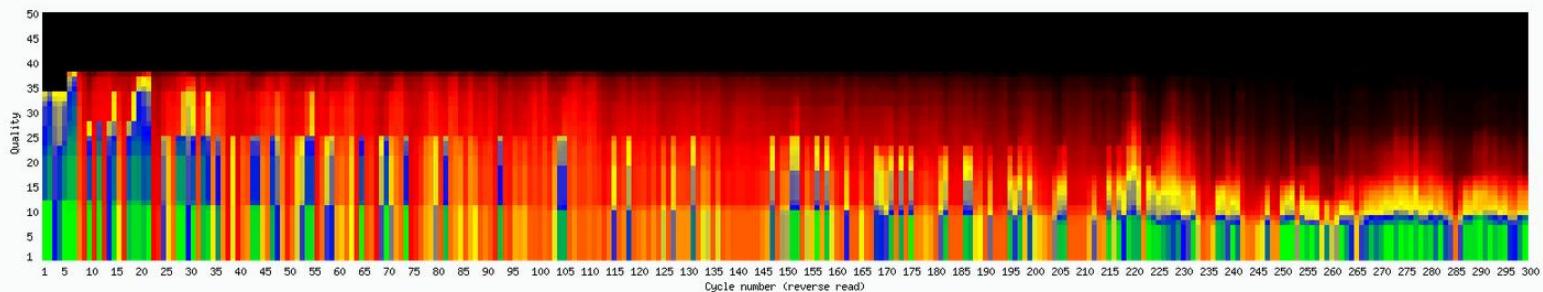
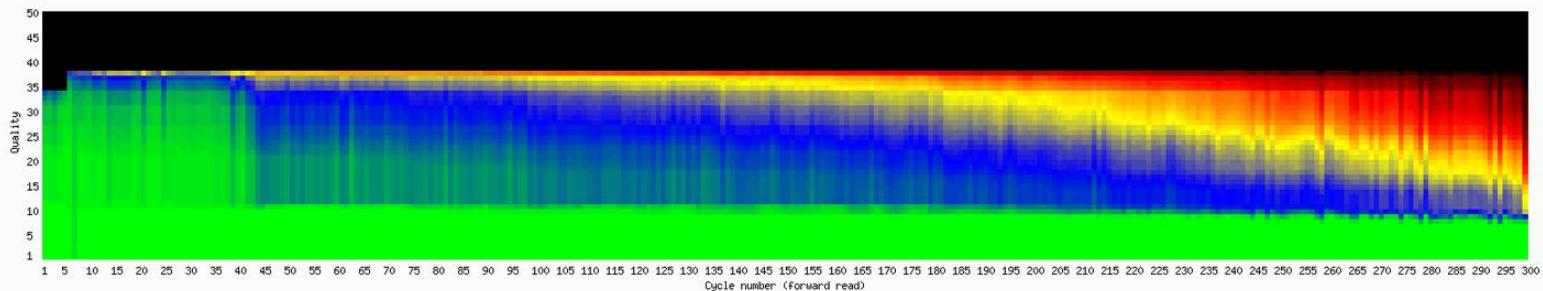
Limitations of Illumina Sequencing Technology

- » Some systematic errors
 - » Difficult to spot rare variants (<1%)
- » Low complexity templates
- » Sequencing short fragments doesn't give any long range information

Limitations of Illumina Sequencing Technology

- » Some systematic errors
 - » Difficult to spot rare variants (<1%)
 - » See Duplex seq by Scmitt et al.,
- » 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

Low quality 16S run



Limitations of Illumina Sequencing Technology

- » Some systematic errors
 - » Difficult to spot rare variants (<1%)
 - » See Duplex seq by Scmitt et al.,
- » 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



10X Genomics

<http://10xgenomics.com>



GemCode Instrument



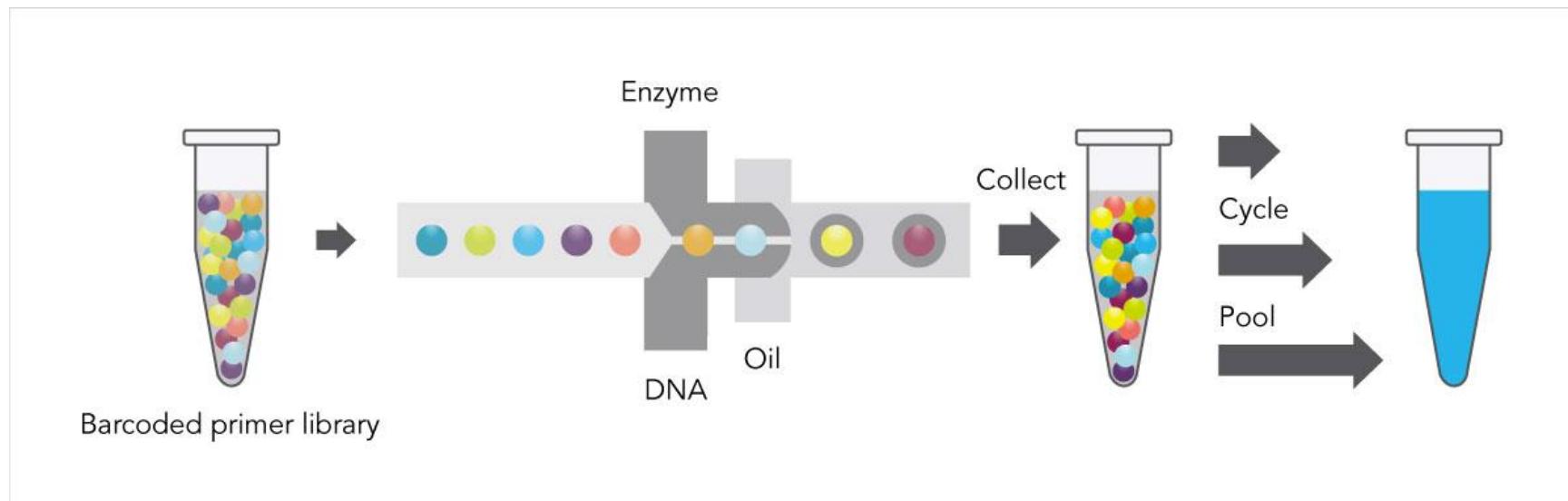


What is it?

- » The platform **upgrades** short read sequencers,
 - » Enables the capture of long range information
 - » 10 - >100 kb molecule sequencing
- » i.e., gives >PacBio / ONT read-lengths on ILMN

How does it do that?!?

- » ...partitions long DNA molecules (incl. >100 kb) and prepares sequencing libraries... **all fragments produced within a partition share a common barcode.**



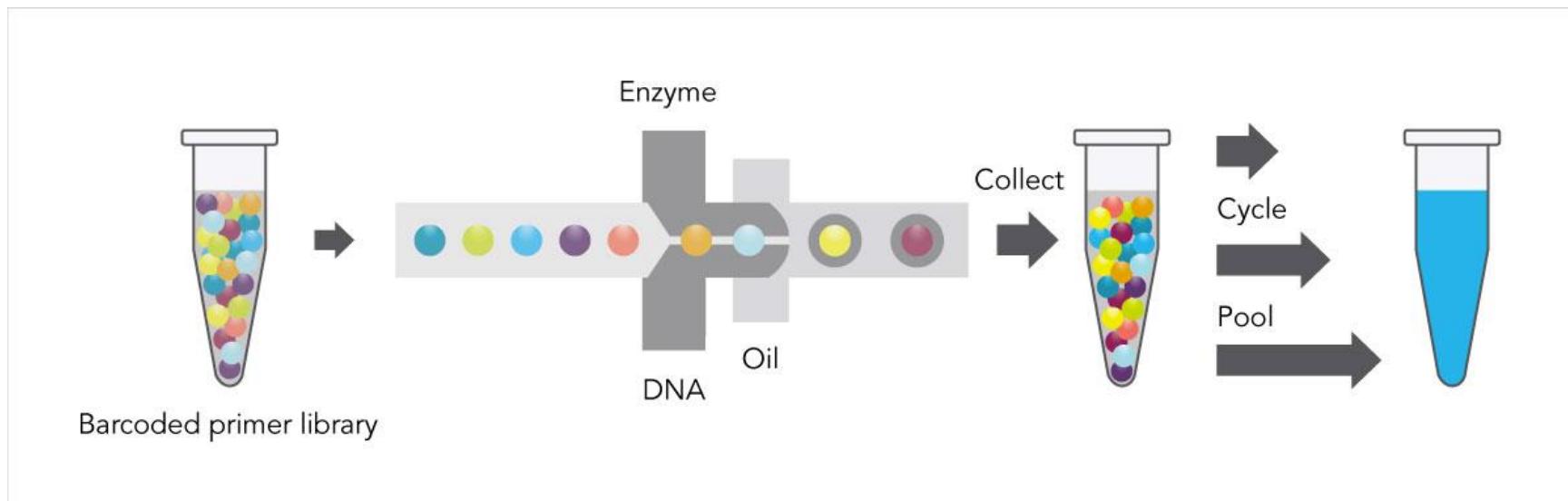


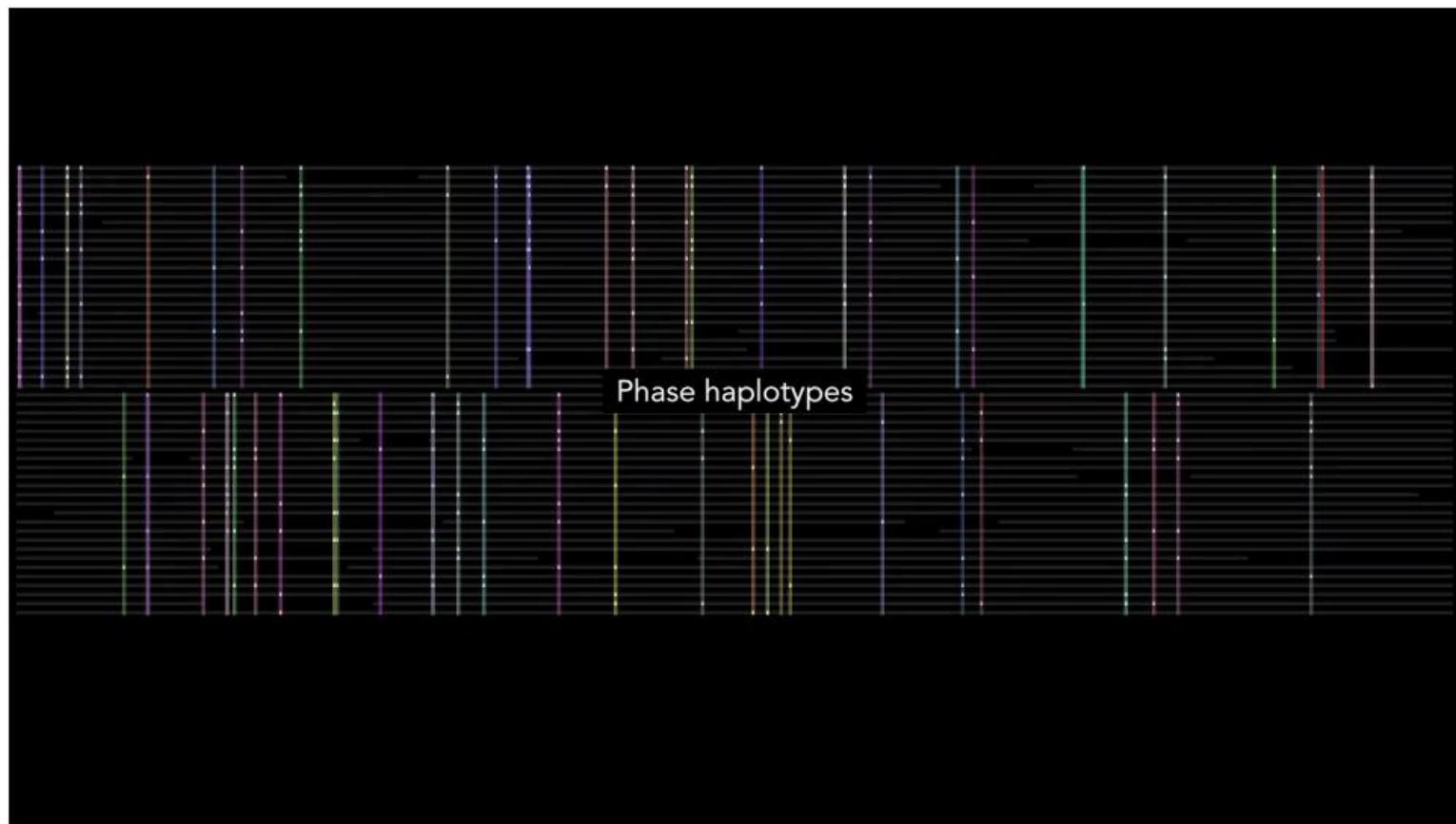
10 x genomics chip

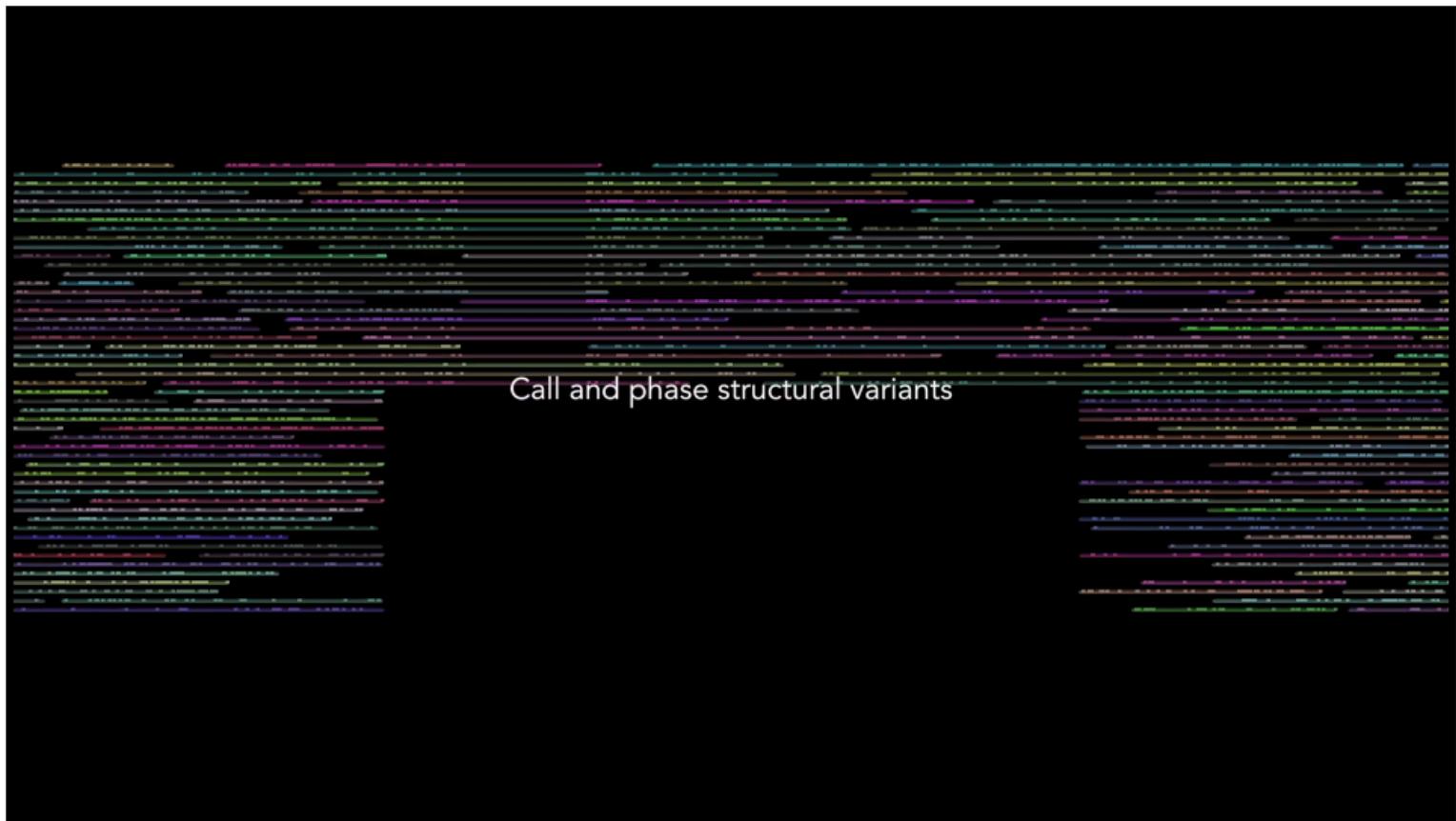


How does it do that?!?

- » ...partitions long DNA molecules (incl. >100 kb) and prepares sequencing libraries... **all fragments produced within a partition share a common barcode.**

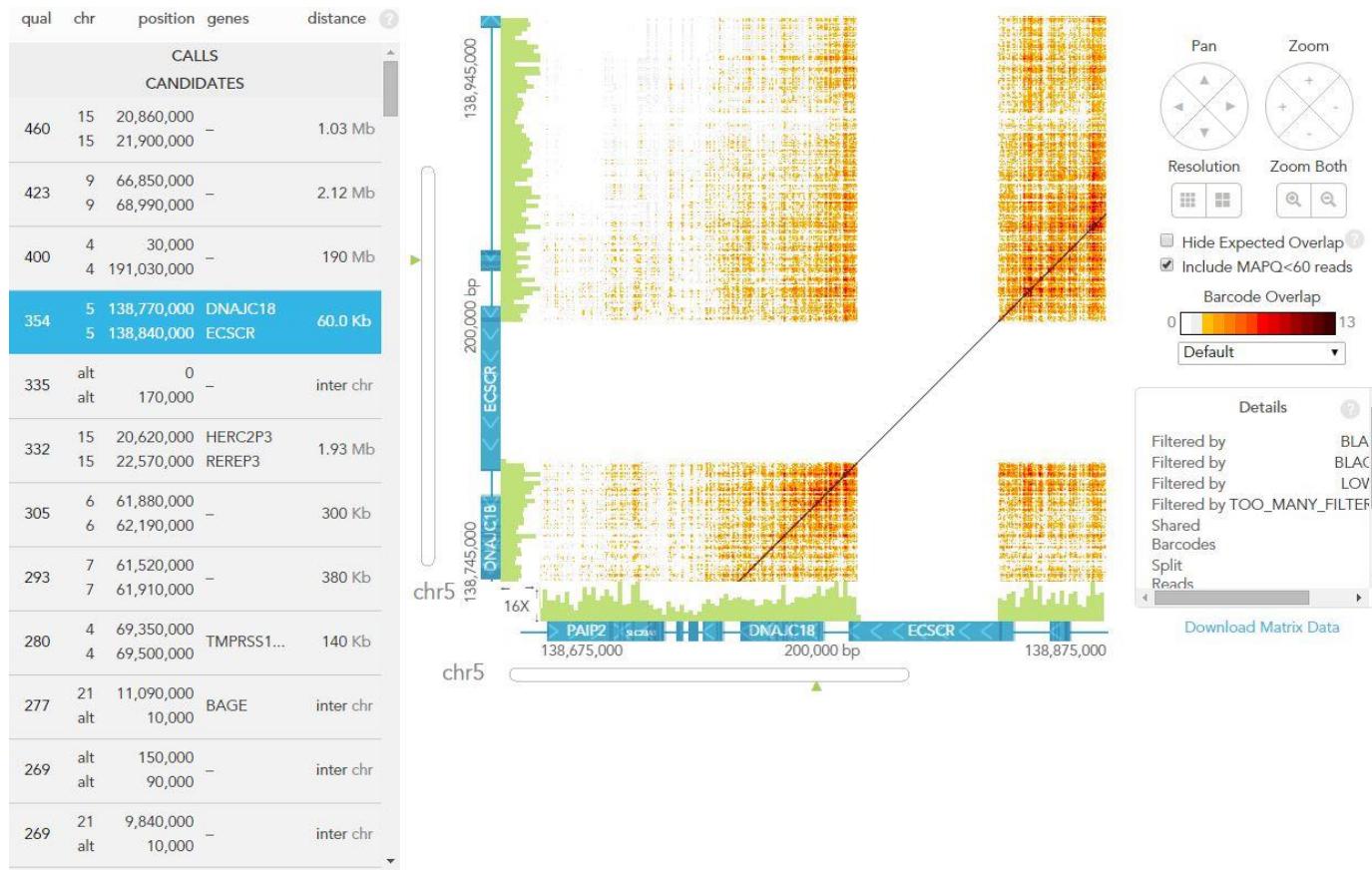




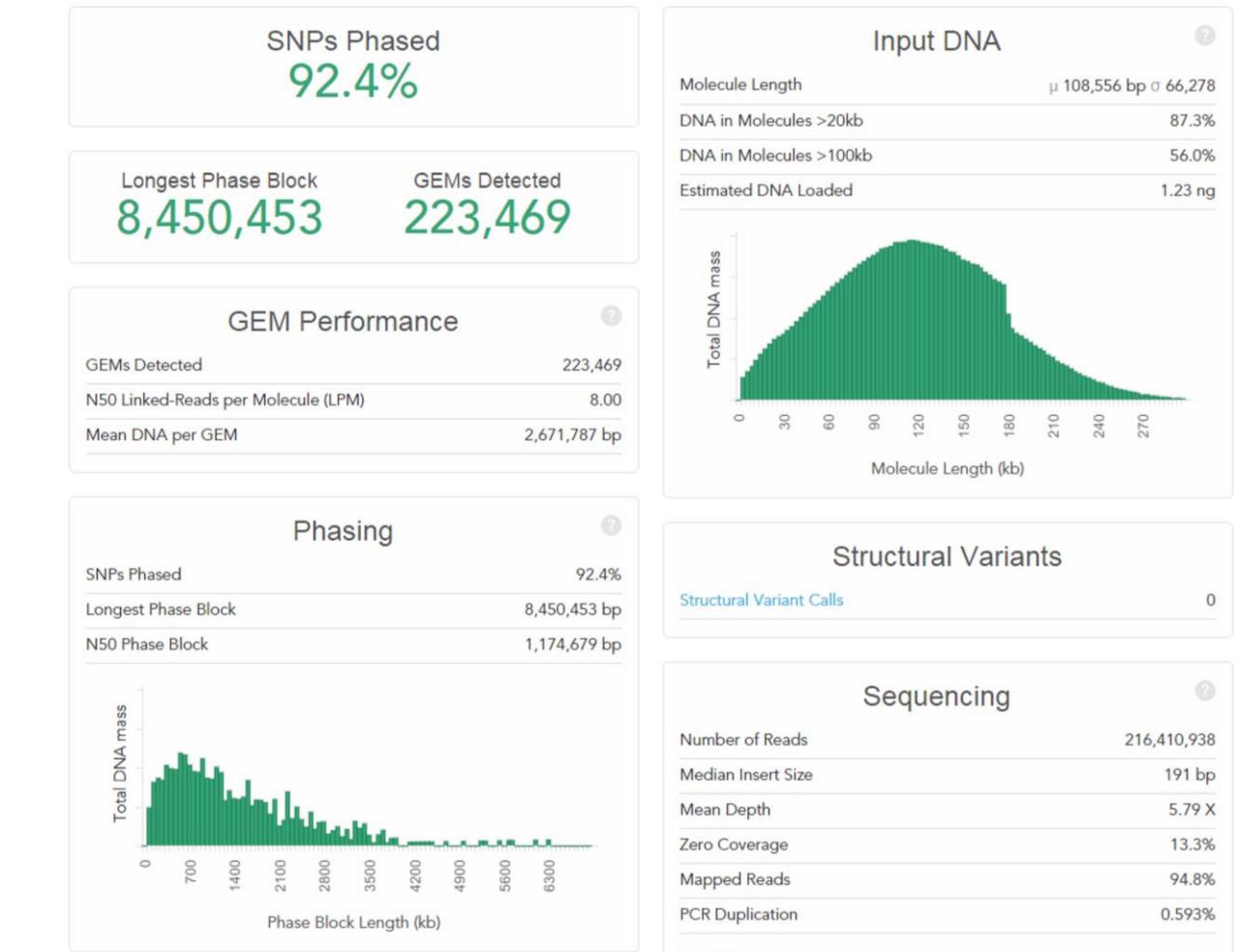


10 x Genomics Loupe

Example of a homozygous deletion



10 x Genomics Loupe Summary





10x genomics

- » \$125 K per instrument
- » \$450 per sample
- » Multiplex Kit
 - » Contains the sample indexes to enable sample multiplexing up to **96-plex** on an Illumina sequencer.
- » Phases 93-95% of SNPs into Mb size blocks
- » SuperNova genome assembly pipeline



Limitations of Illumina Sequencing Technology

» Index Hopping

Limitations of Illumina Sequencing Technology

» Index Hopping

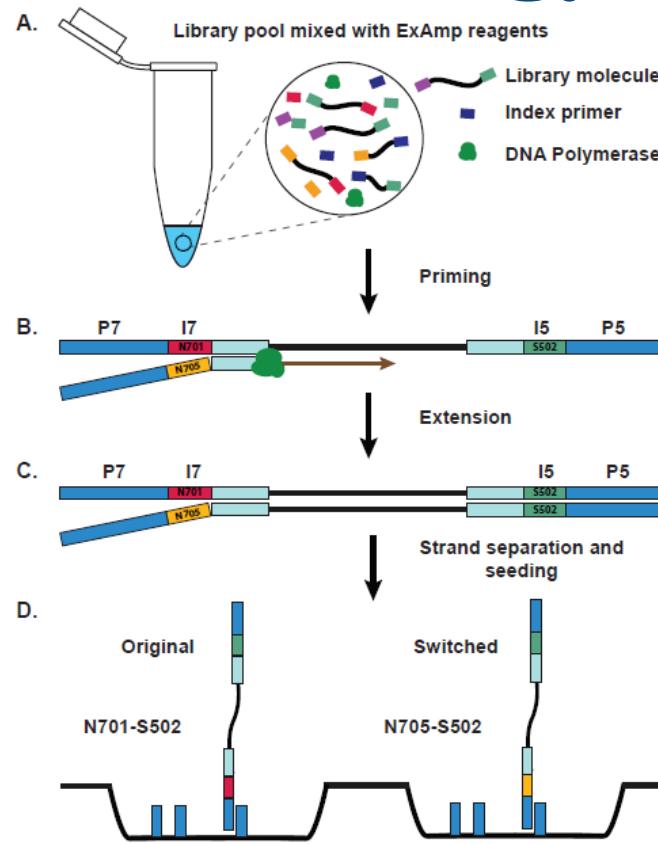


Figure 1

See Sinha et al BioRxIV 2017. <http://biorxiv.org/content/early/2017/04/09/125724>



Illumina

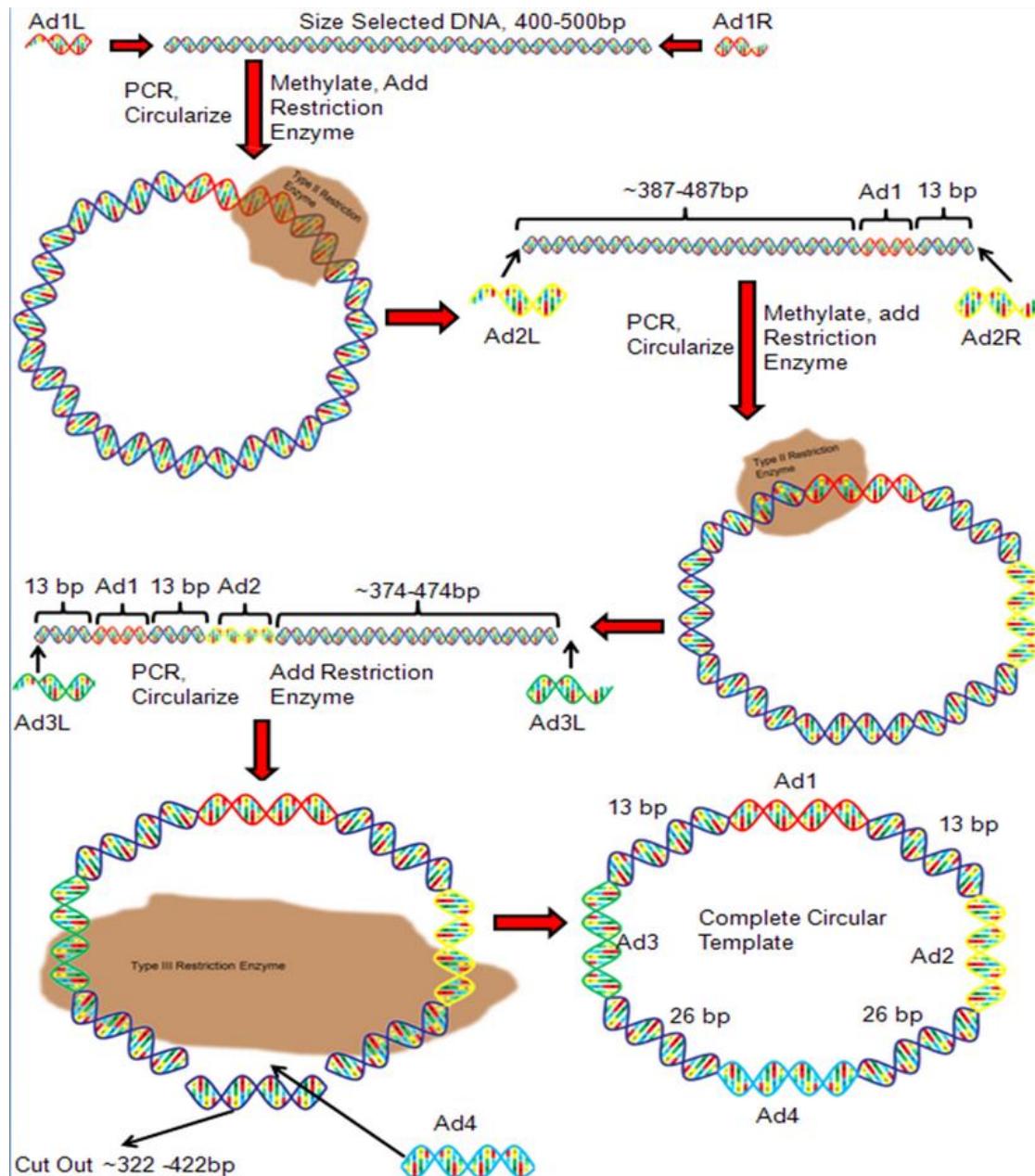
- » Cheap \$6-\$30/Gb
- » Highly accurate data mostly Q30
- » Massively parallel. Millions of reads
- » Short read

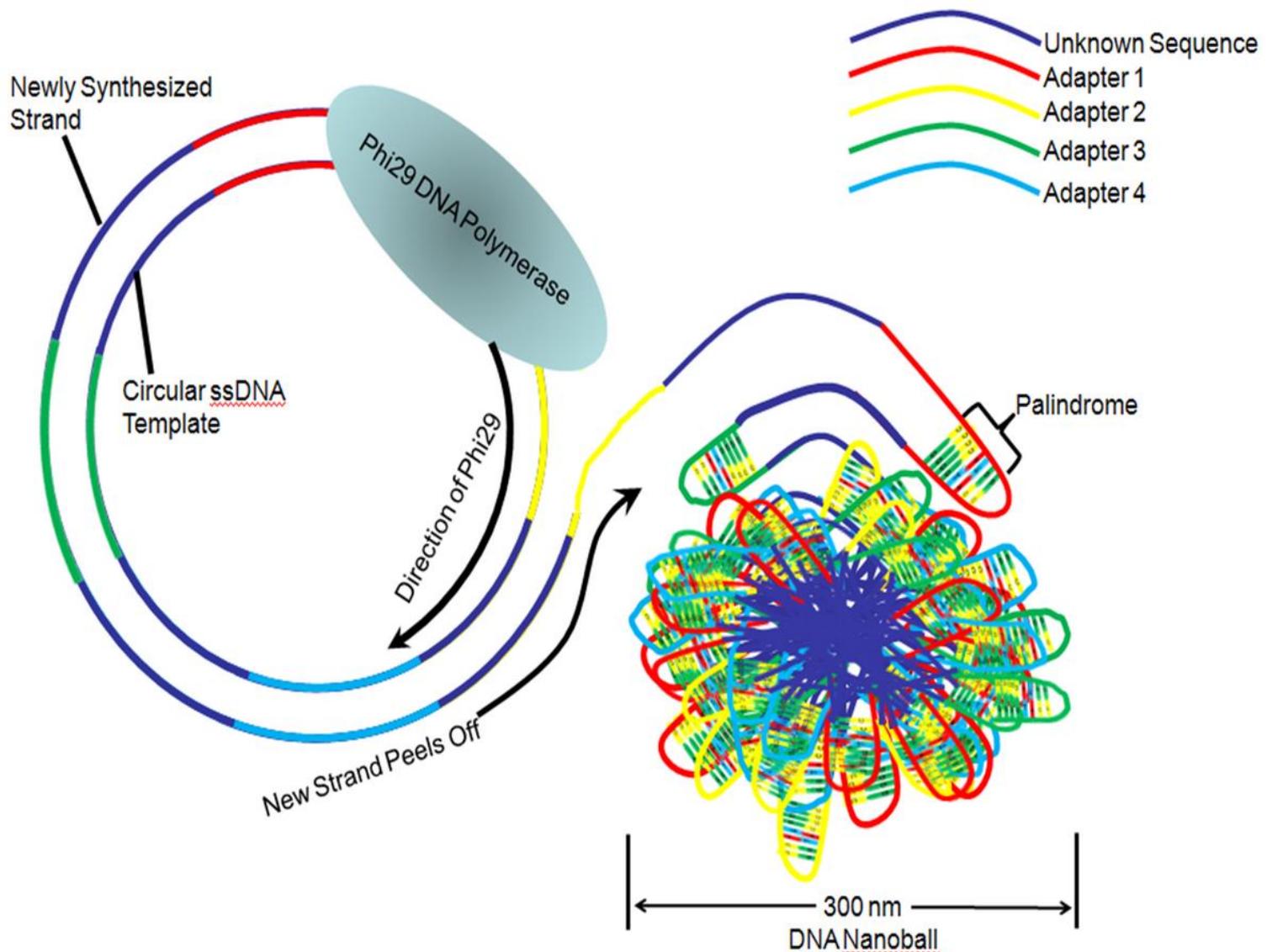
Complete Genomics



Complete Genomics

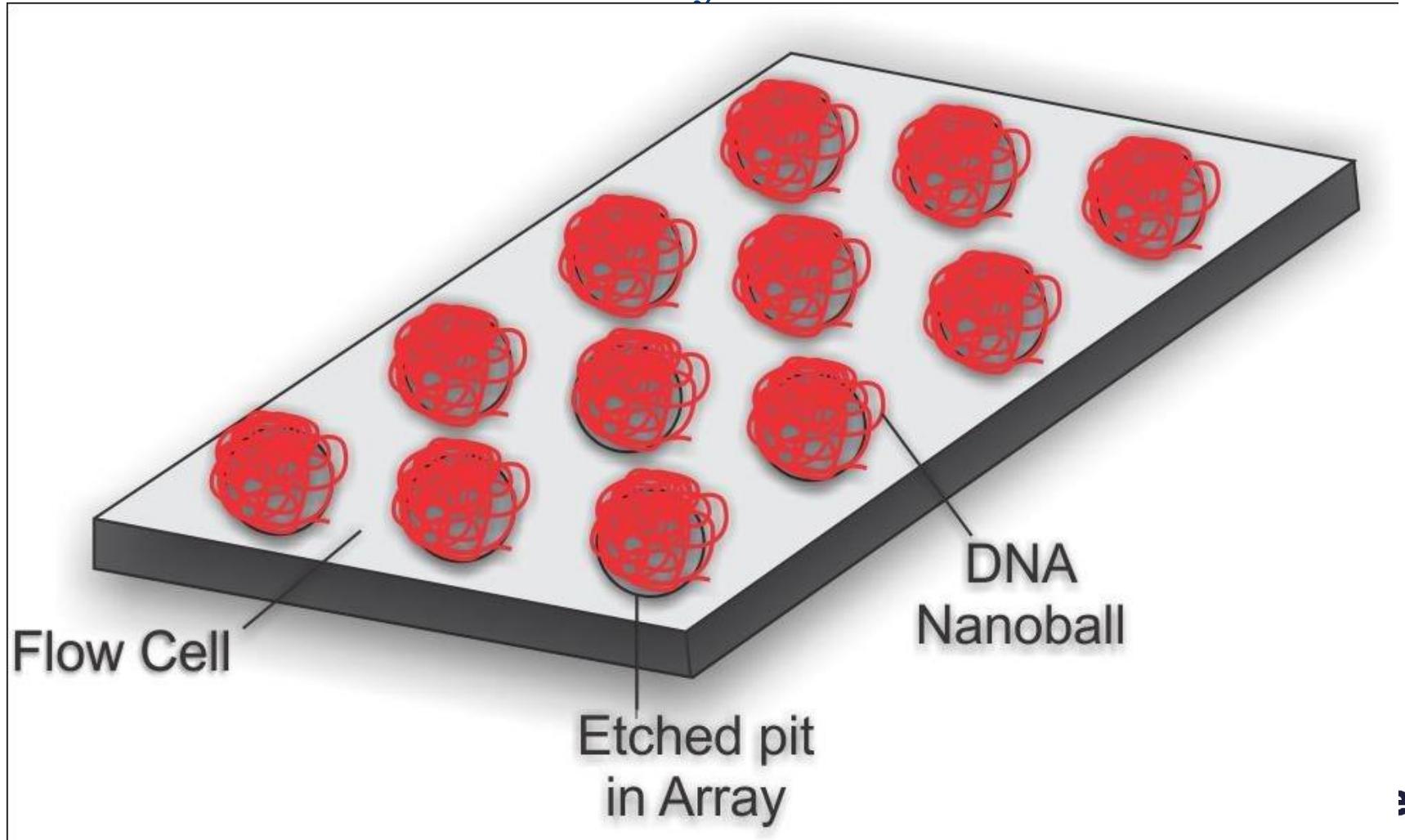


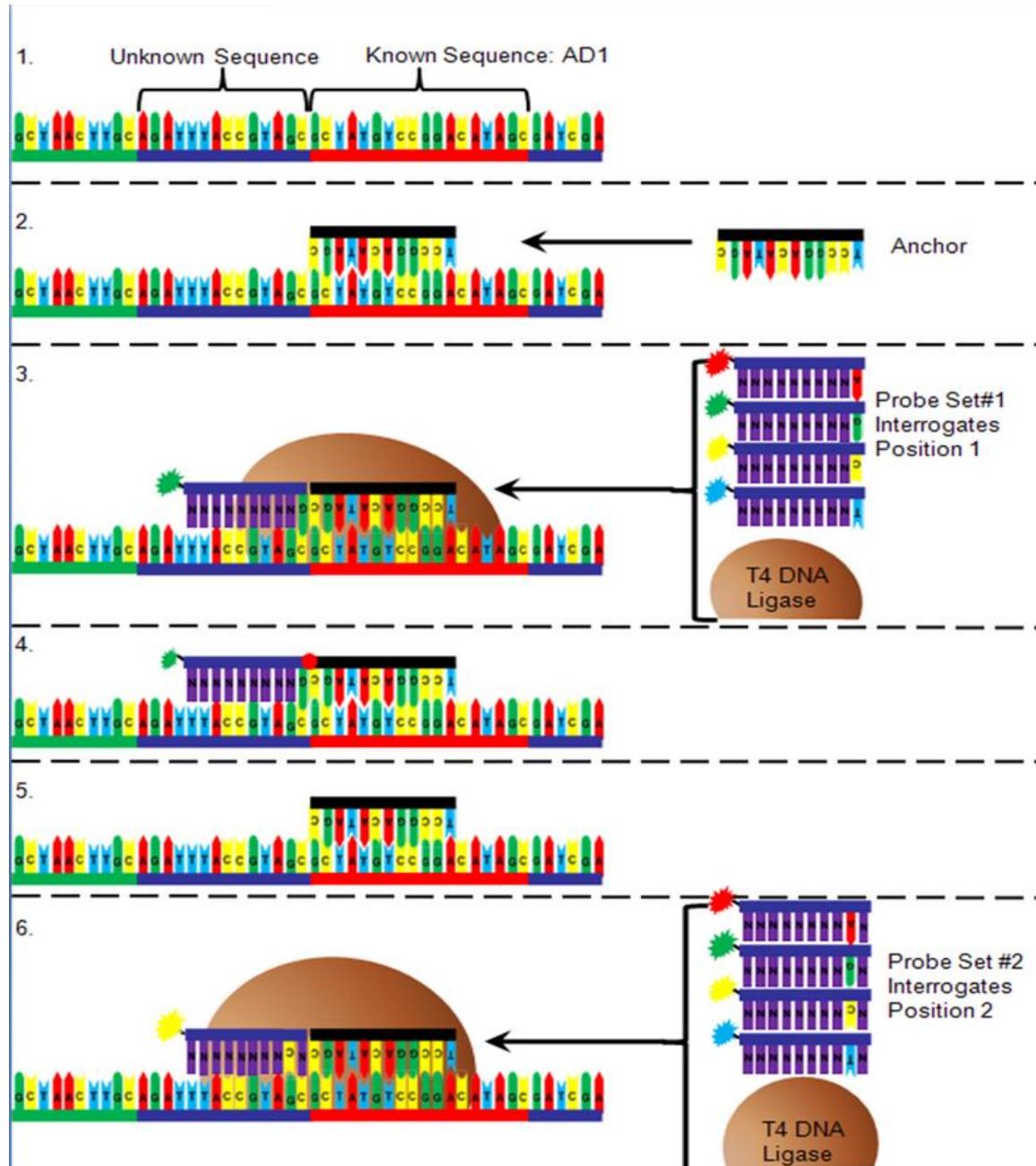




Complete Genomics

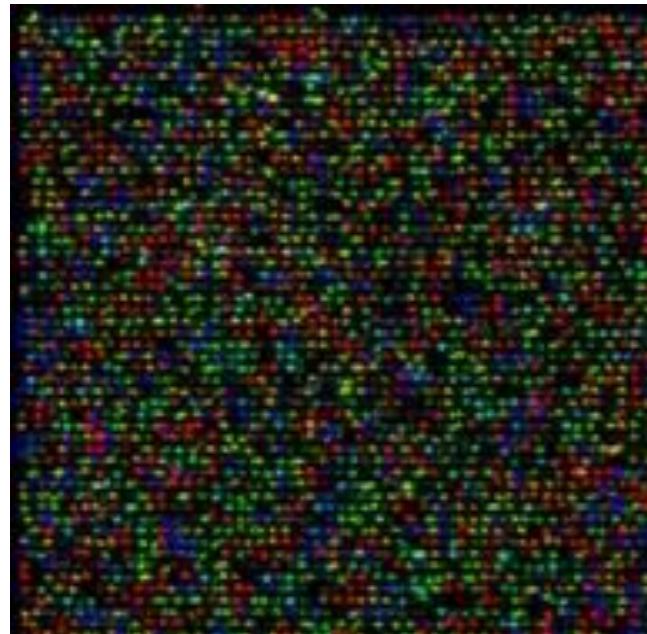
Nanoball arrays on flowcell





Complete Genomics

- » DNA nano-balls, ordered array
- » Ligation-based sequencing
- » 210 Gb / slide (18 slides)
- » 35-base paired reads
- » 1 week run time
- » \$50 / Gb (finished)



Complete Genomics

- » Raw error rate 0.5%
- » Consensus accuracy 10^{-5}
- » Short read lengths
- » Service business model
- » 500 Genomes/month cap.
- » Only 40x human genomes
 - » ~\$5,000 per genome





Complete Genomics Revolocity



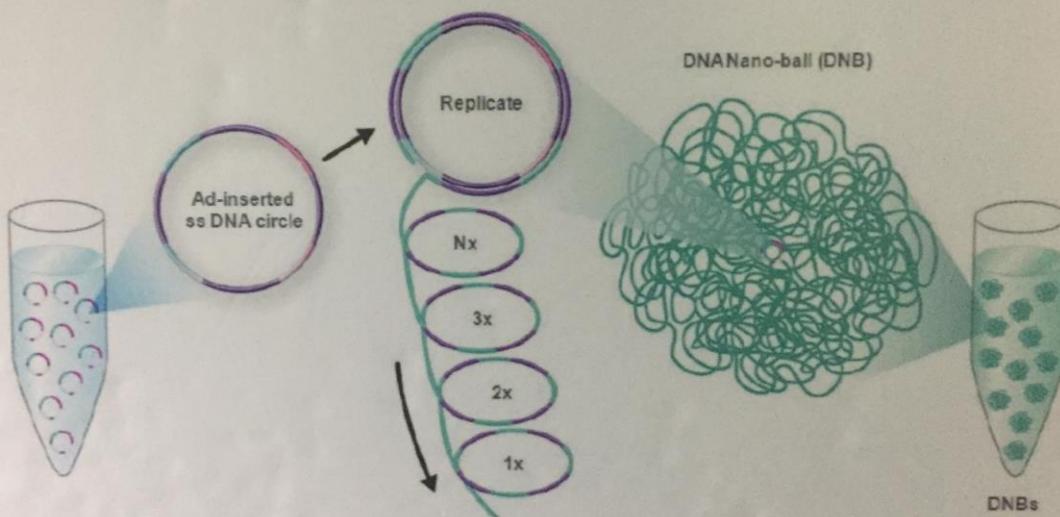
- » 10,000 genomes/year
- » \$12M
- » 10-120 samples at a time
- » Sample to answer
- » 8 day turnaround
- » 96% genome coverage
- » 1 error in a million (raw 0.5%)
- » 300 bp insert. 2 x 28bp reads



BGI

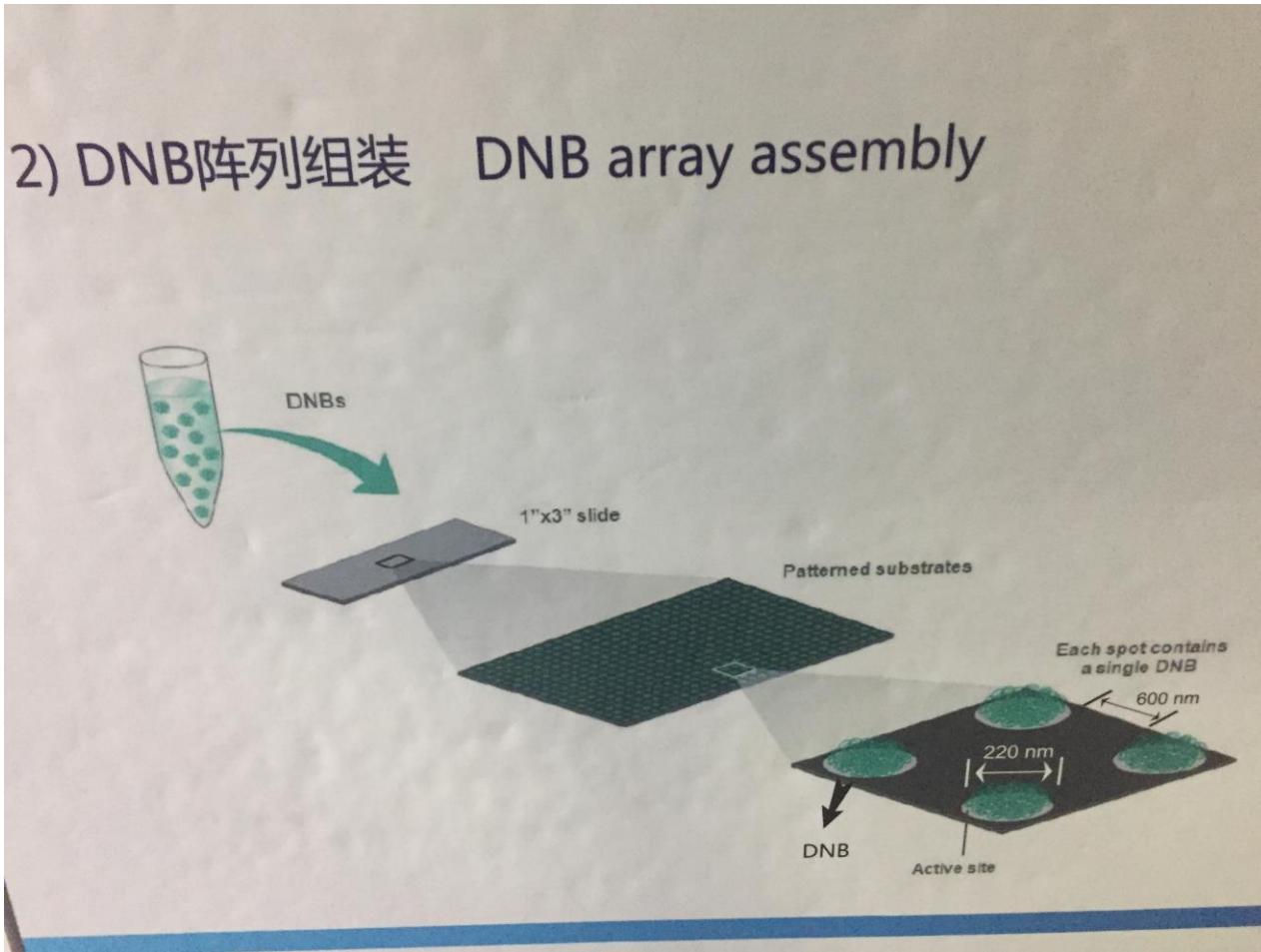
Core Technologies

1) DNA 纳米球制作 DNA nanoball fabrication



2) DNB阵列组装 DNB array assembly

Core Technologies





Sequencing

- » DNBs loaded on silicon wafer with photomasking to create patterned array of binding sites
- » Primers anchored to DNBs
- » SBS using fluorescent reversible terminators
- » 2 or 4 colour chemistry
- » After sequencing use MDA to produce DNB from second strand allowing paired end sequencing
- » Claim 5% run failure rate





Advantages of MGIs sequencers

- » Instruments cheaper than Illumina
- » Sequencing costs cheaper
- » Consistent yields
- » No index hopping. No fundamental Bias. No fragment size dependent representation bias.
- » Quality similar to Illumina
- » 0.2% raw error



BGI 500

- Launched 2015
- 520Gb/run
- 1300M reads
- PE100 or PE50
- SE 35, 50 or 100
- CFDA licensed
- 1-9 day run
- 85% > Q30



BGI 50

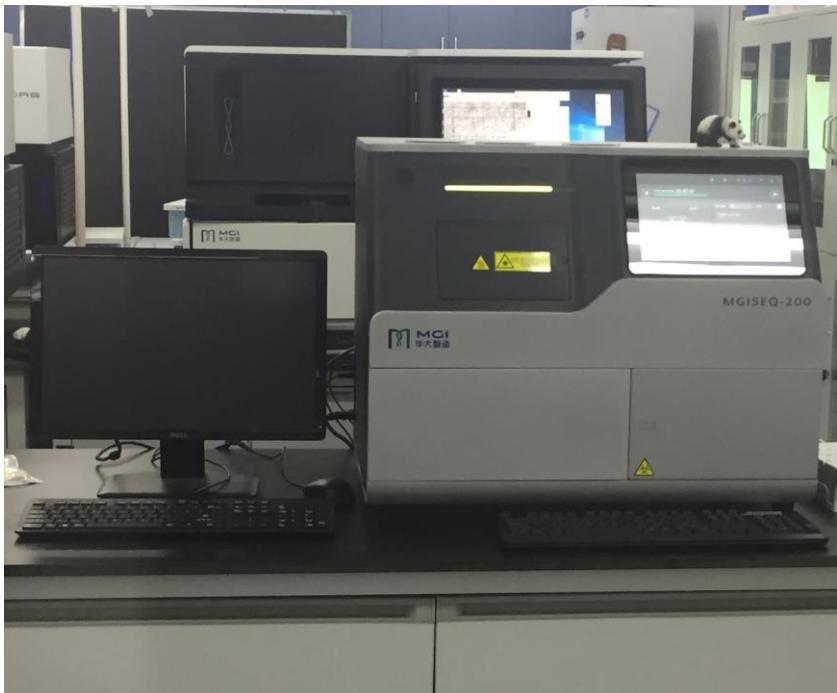
BGISEQ-50具有专业、小巧、精简等特点，读取50bp, 通量8Gb。

BGISEQ-50 is a sharp, delicate and smart high-throughput sequencing platform. Read 50bp, throughput: 8Gb per



MGI 200

- 2 colour SBS



小型化测序仪，读长200bp，
支持多种读长，通量15-
60Gb每Run。

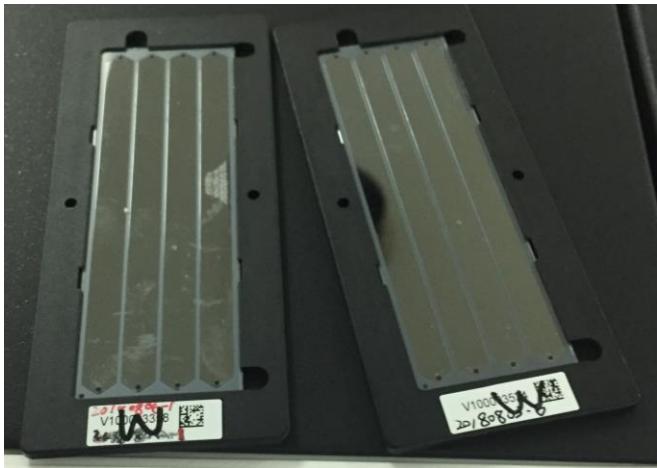
MGISEQ-200 is a miniature sequencer that transcends geographical restrictions. MGISEQ-200 adopts a whole new biochemical system design to support rapid sequencing in various different reads modes. Read length: 200bp Throughput: 15 to 60Gb per run.



测序仪型号	最大通量	价格/Gb
MGISEQ-200	60Gb	\$37



MGI 2000



MGISEQ-2000采用全新的芯片设计，能够灵活支持多种不同的测序模式，并采用优化设计的光学及生化系统，能够在较短时间内完成完整的测序流程，带给使用者更加精简流畅的测序体验。通量高达1080Gb每Run。

MGISEQ-2000 adopts a novel chip system that supports various modes of sequencing and is integrated with optimally designed optical and biochemical systems to execute the sequencing process within a relatively shorter time for a more streamlined sequencing experience for the user.

Throughput: 1080Gb per run.



测序仪型号	最大通量	价格/Gb
MGISEQ-2000	1080Gb	\$7



gbyd



MGI Seq2000

- » Approx £225k (2M Yuan)
- » \$600 per human genome as a service
- » \$300 if had own instrument (\$3/Gb)
- » 1080 Gb output per run across 8 lanes (2 x 4 lane flowcells)
- » <48 hours for PE100
- » About to launch PE150. PE300 and 400 being run in dev
- » 2 flowcell sizes. 375M or 1800M reads
- » 85%>Q30
- » Sold 2 units recently to Riken
- » 4 colour SBS

Pacific Biosciences



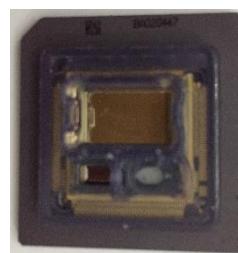
RSII : 1800 lbs. and ~11 feet long !

Pacific Biosciences

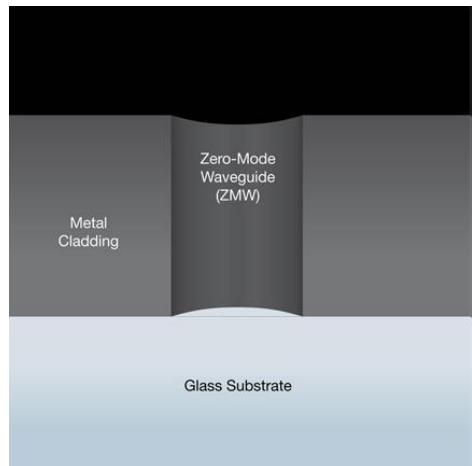
Sequel



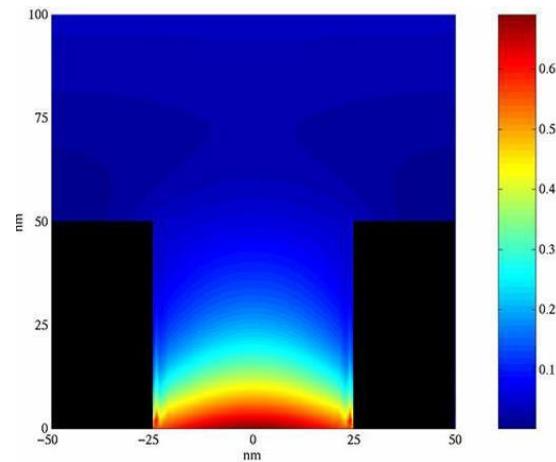
- » 2016 release
- » 1 million ZMWs/SMRT
- » 40Gb/SMRT cell
- » ~\$12/Gb
- » \$350K instrument



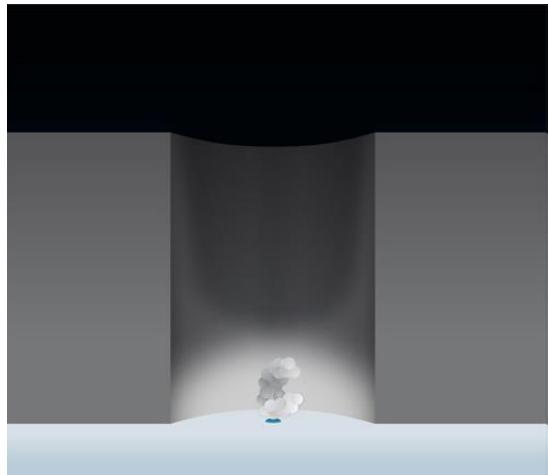
Pac Bio Technology



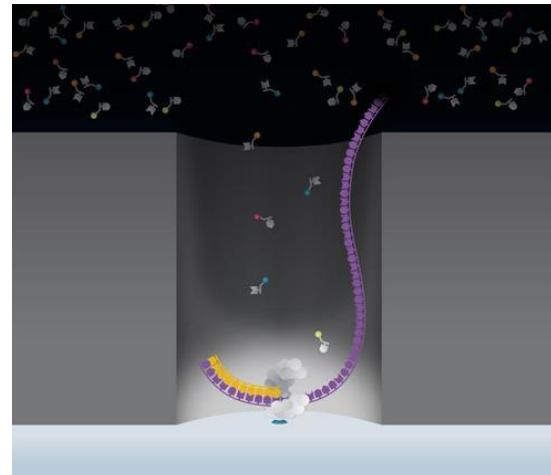
Individual ZMW



Laser light illuminates the ZMW

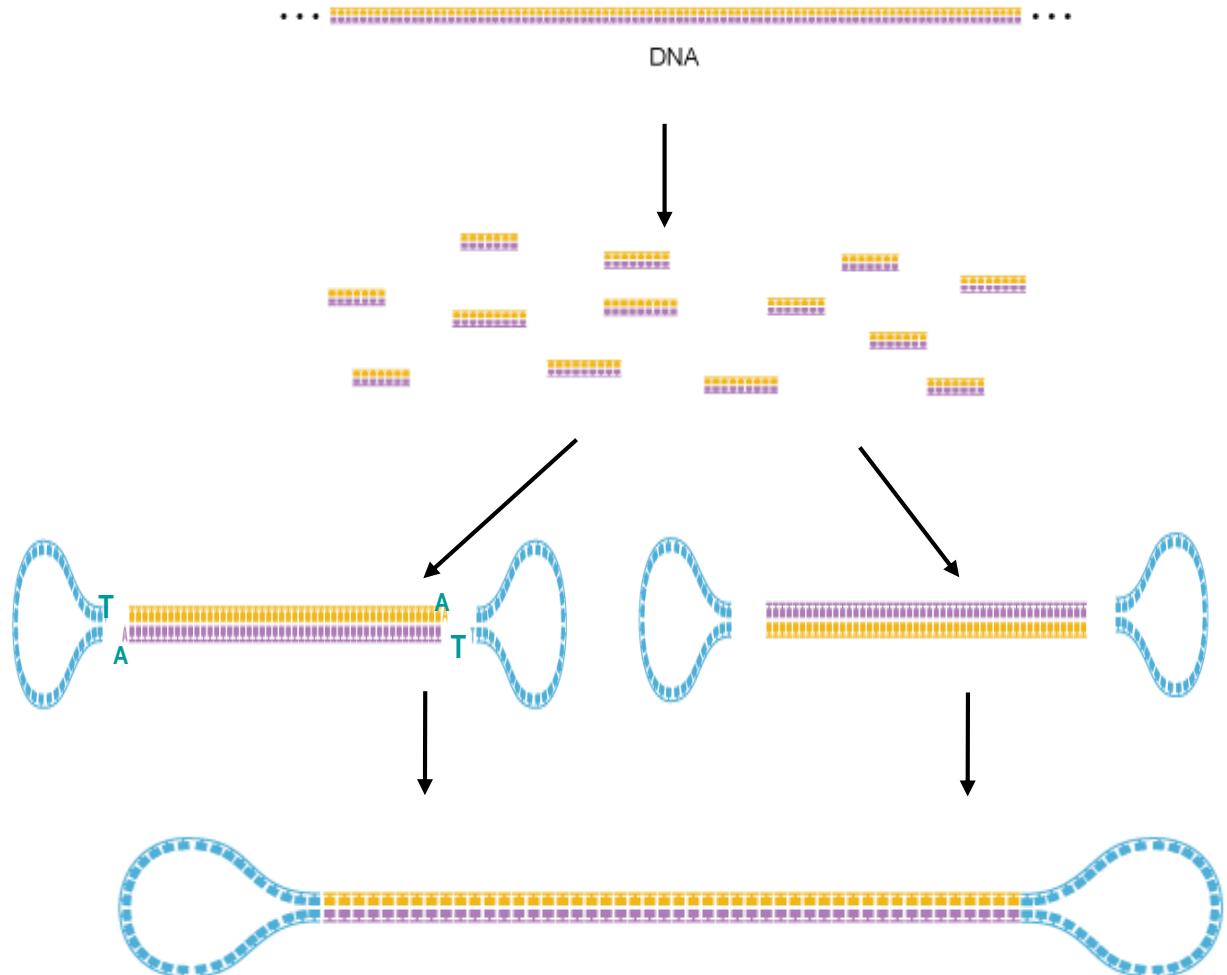
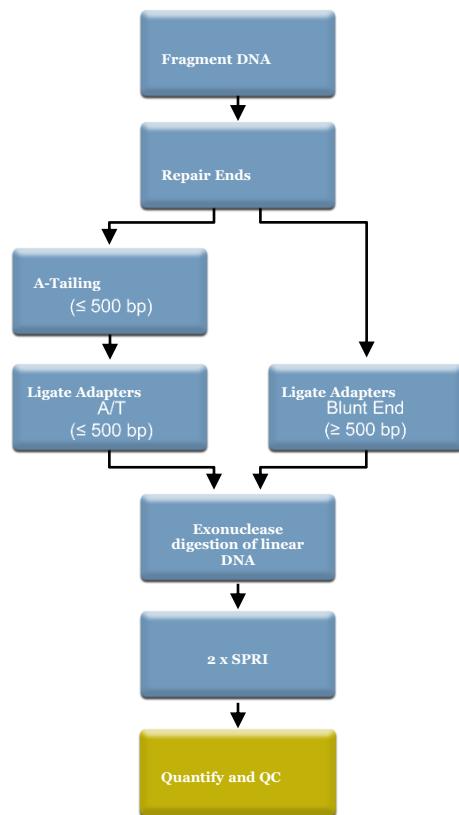


ZMW with DNA polymerase

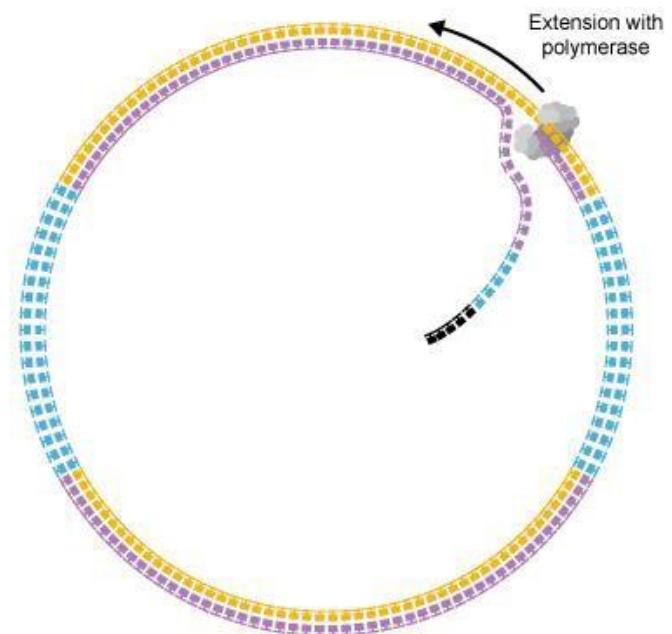
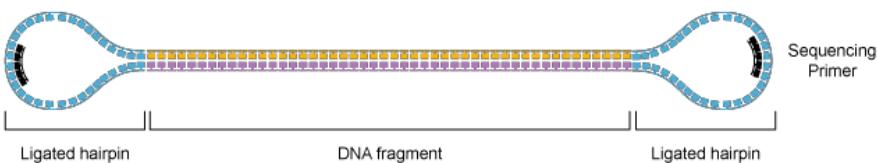


ZMW with polymerase + nucs.

PacBio Library Prep

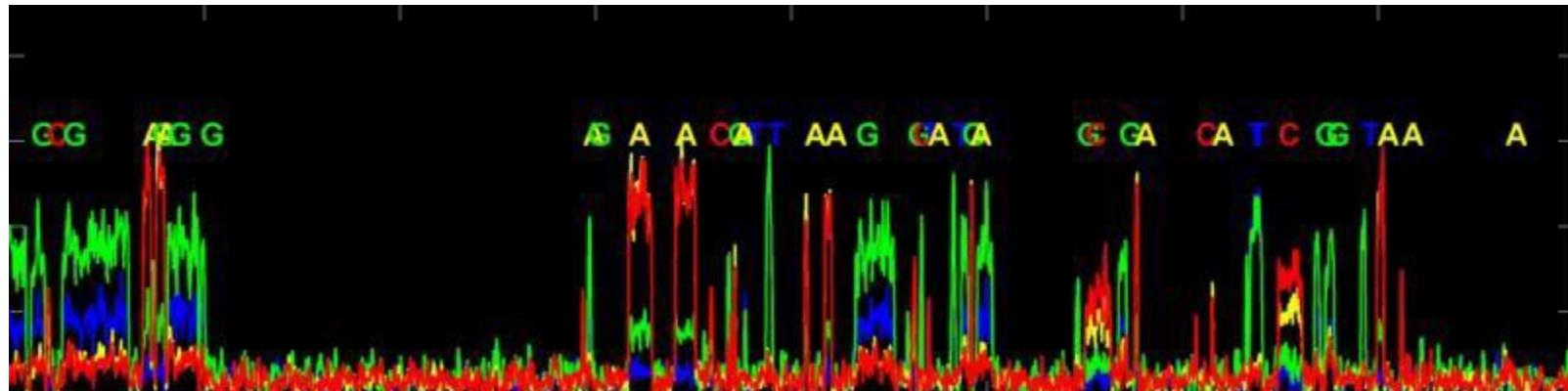


PacBio Template Preparation

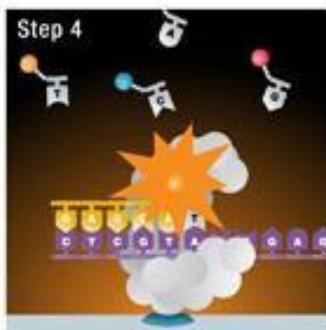
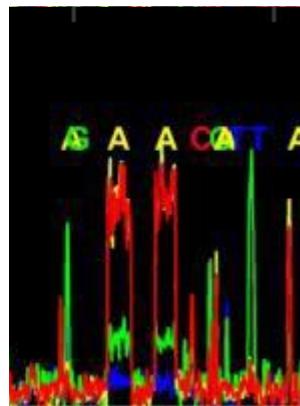


- 1 Anneal primer**
- 2 Bind polymerase**
- 3 Sequence**

PacBio Sequencing

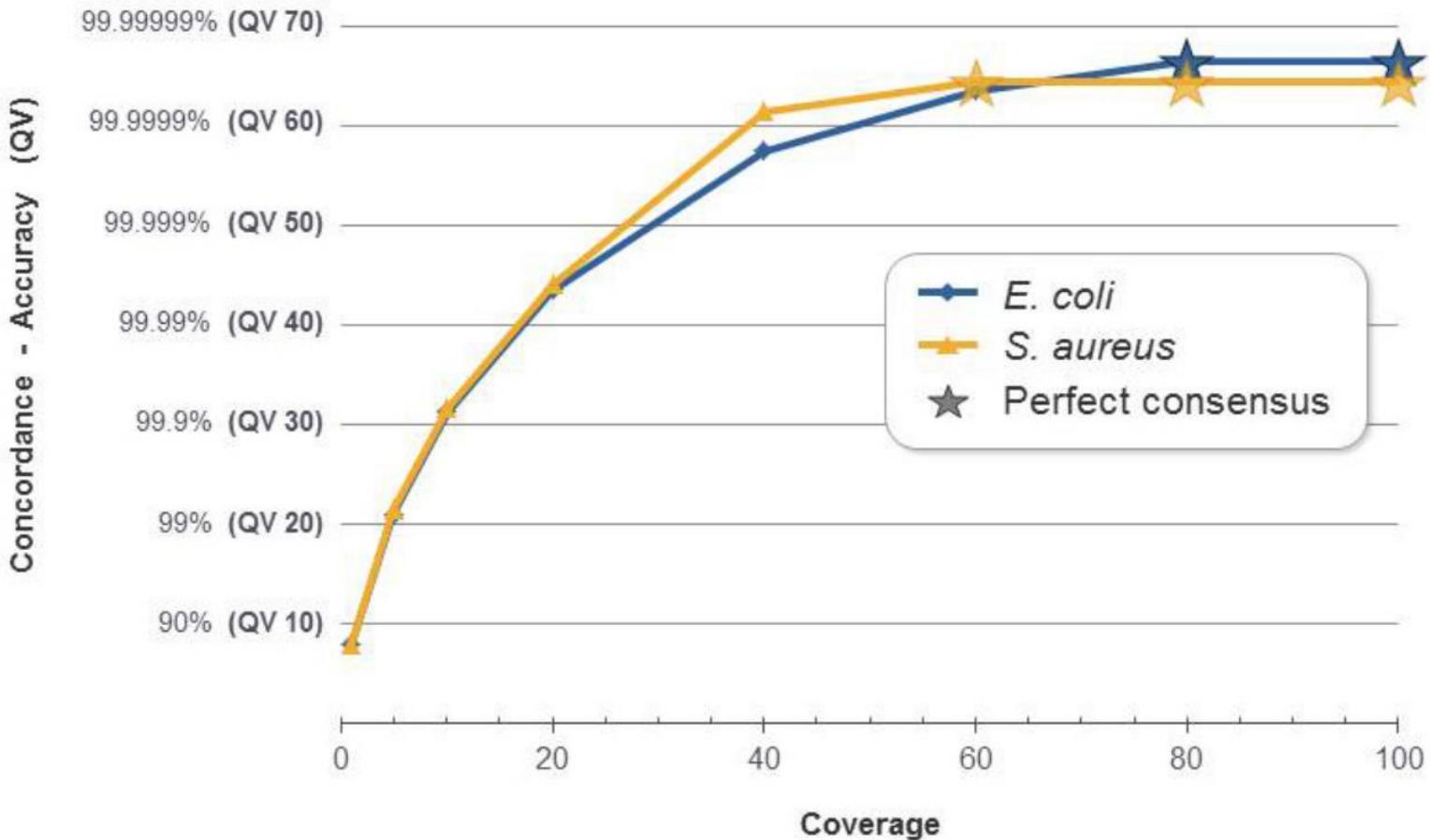


PacBio Sequencing



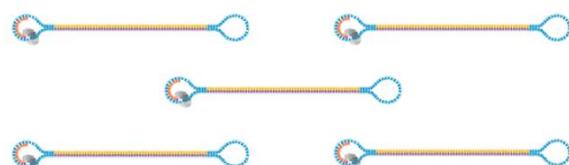
- » Some bases added very quickly and missed
- » Some wrong bases flirt with active site and go away
- » SMRT cell has 1million ZMWs
- » \$500 each

Accuracy



UNTIL NOW: 2 MODES OF SMRT SEQUENCING

- #### - Long-insert Genome Sequencing:



Molecule 1

Consensus sequence

Genotyping, SV detection, *de novo* assembly

- Circular Consensus Sequencing (CCS):



Subread 1

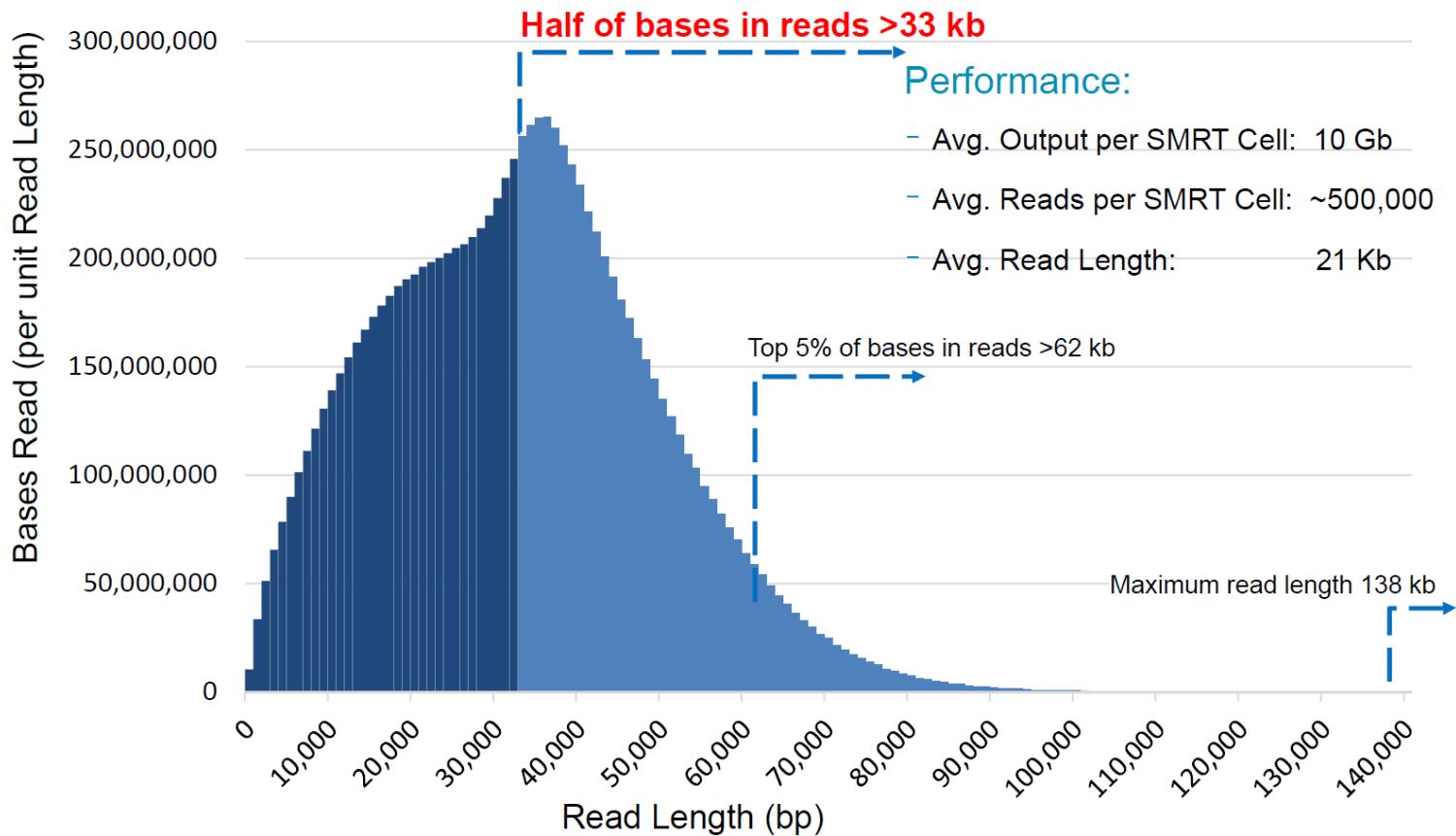
Subread n

Consensus sequence

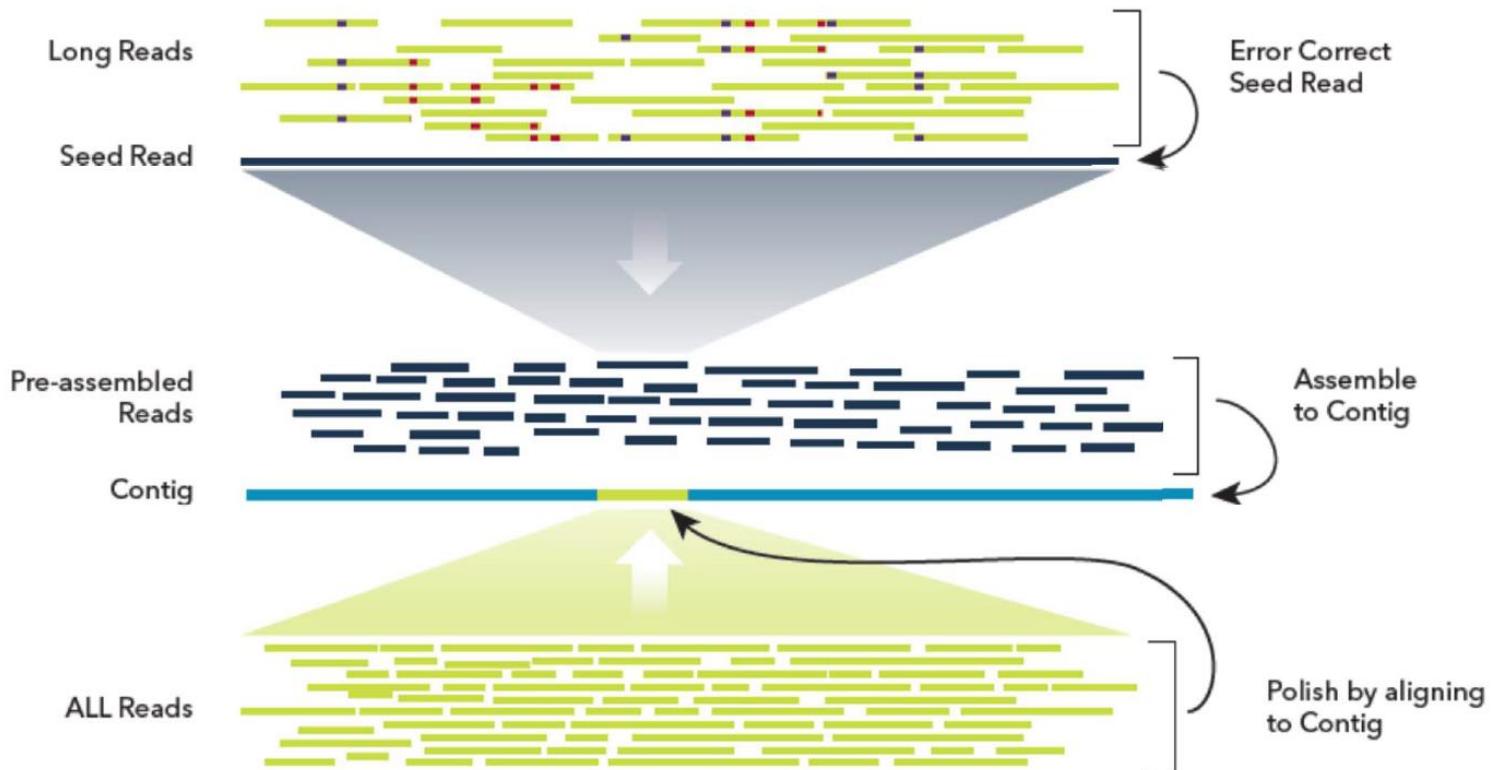
Amplicons, Minor variant detection, Metagenomics



SEQUEL SYSTEM V5.1 PERFORMANCE: HG00733 LIBRARY

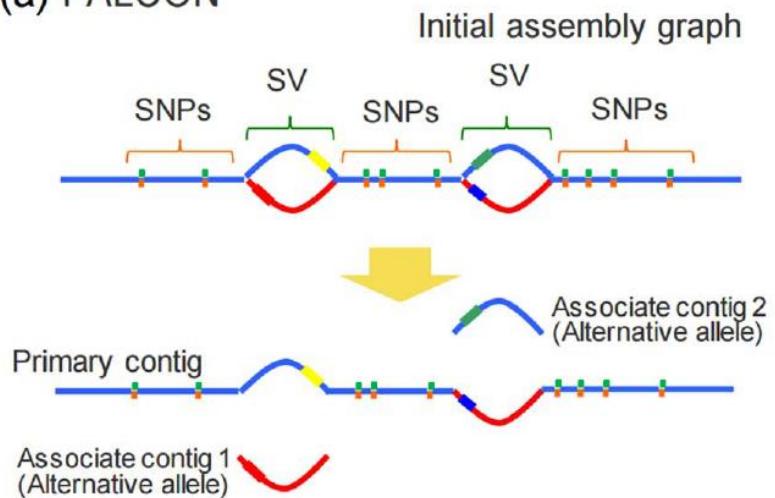


HIERARCHICAL GENOME ASSEMBLY PROCESS (HGAP) & POLISHING

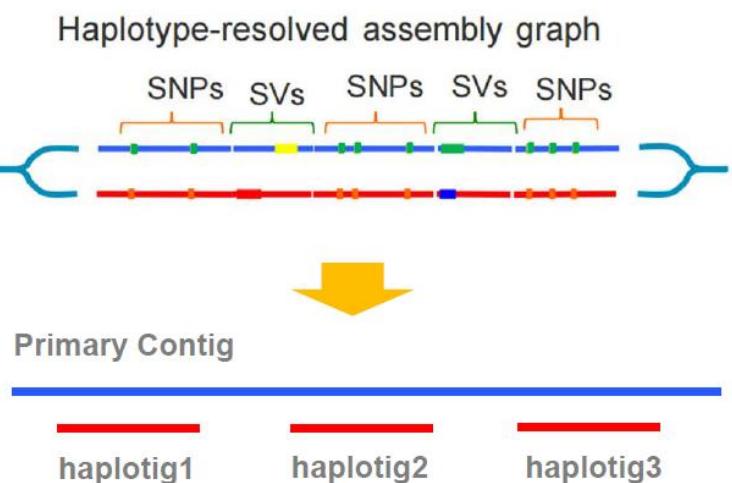


DIPLOID ASSEMBLY WITH FALCON-UNZIP

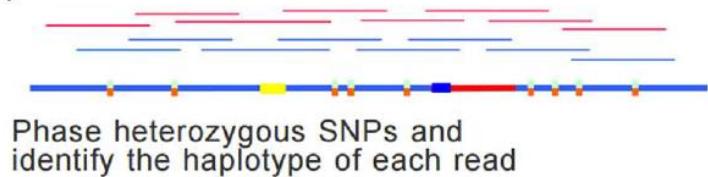
(a) FALCON



(c) FALCON-Unzip



(b)





SEQUEL SYSTEM V6.0 RELEASE

Current Commercial Release (5.1)

- Yield per SMRT Cell:
 - Up to 12 Gb for long-insert genomic libraries
 - Up to 20 Gb for amplicons
- Average read lengths up to 20kb
- 20-hour movies

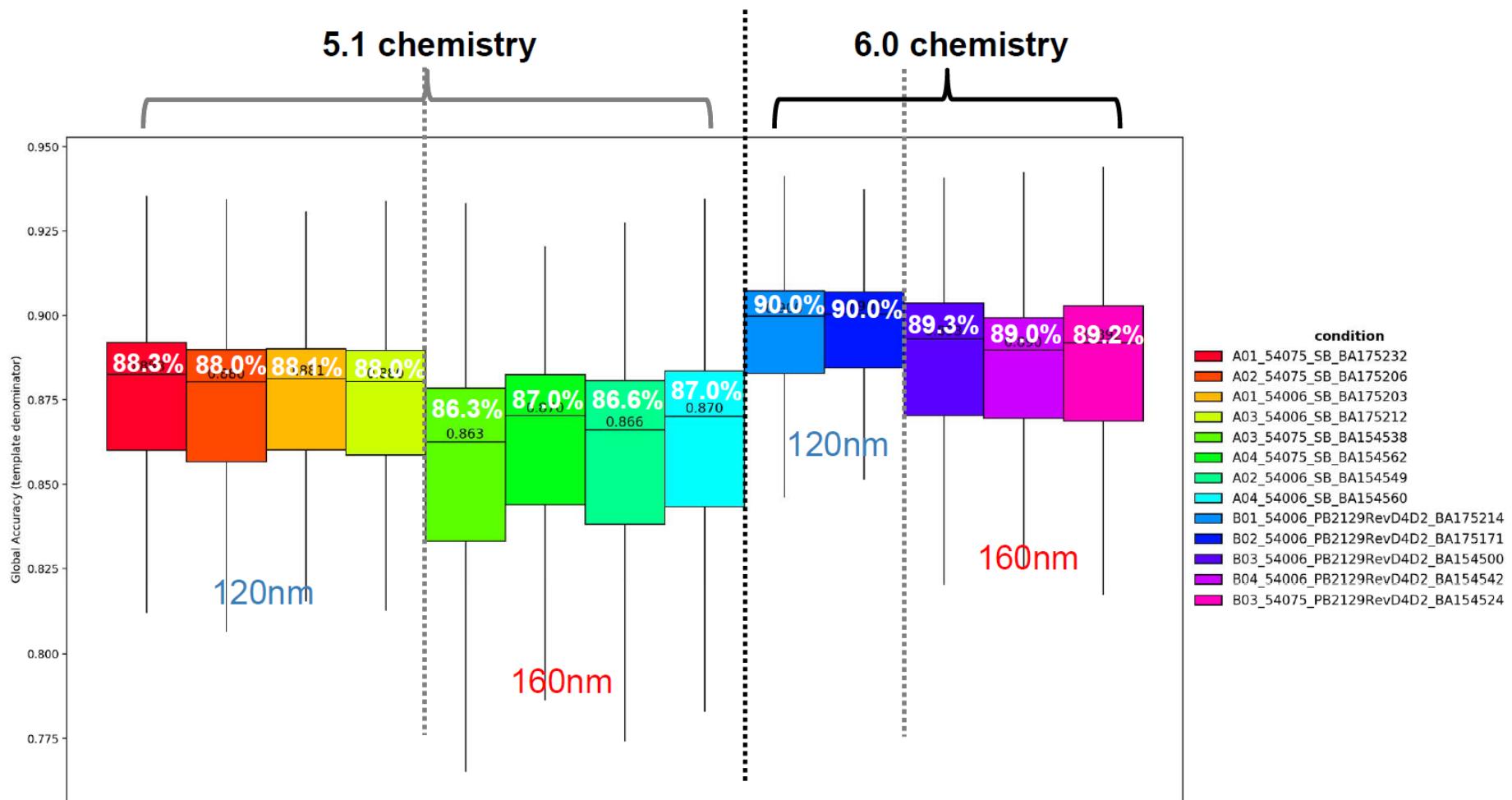
In-house 6.0 Release Performance

- Yield per SMRT Cell:
 - Up to 20 Gb for long-insert genomic libraries
 - Up to 50 Gb for 10kb insert genomic libraries
- Average read lengths up to 100kb
- 20-hour movies
- Early access in September 2018

2-4fold increase in total base yield per SMRT Cell – primarily due to significantly increased polymerase read length



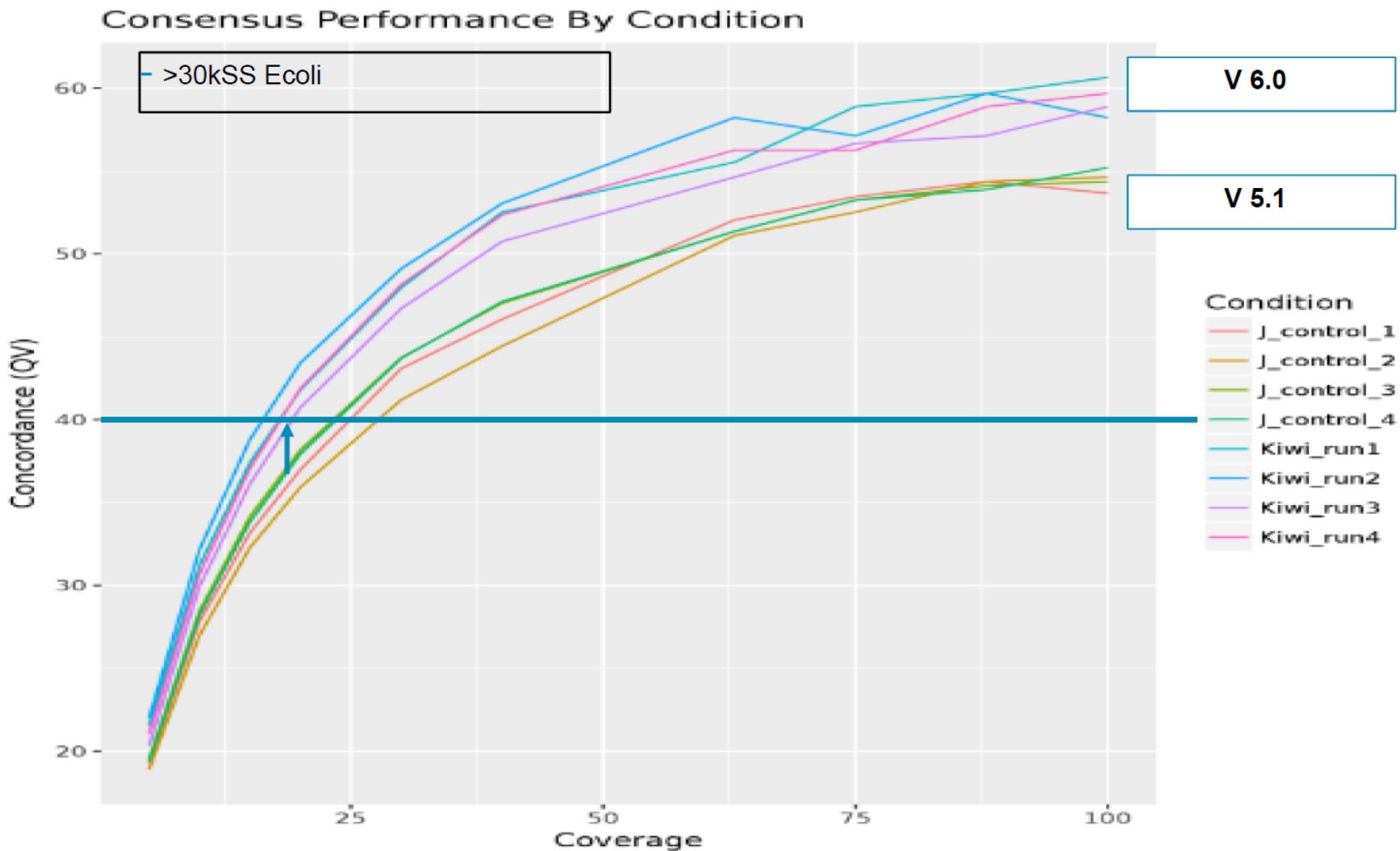
IMPROVED ACCURACY WITH SYSTEM V6.0 AND 120NM ESL CHIPS



Both the 120nm ESL chip and 6.0 chemistry contribute to a substantial gain in raw accuracy.

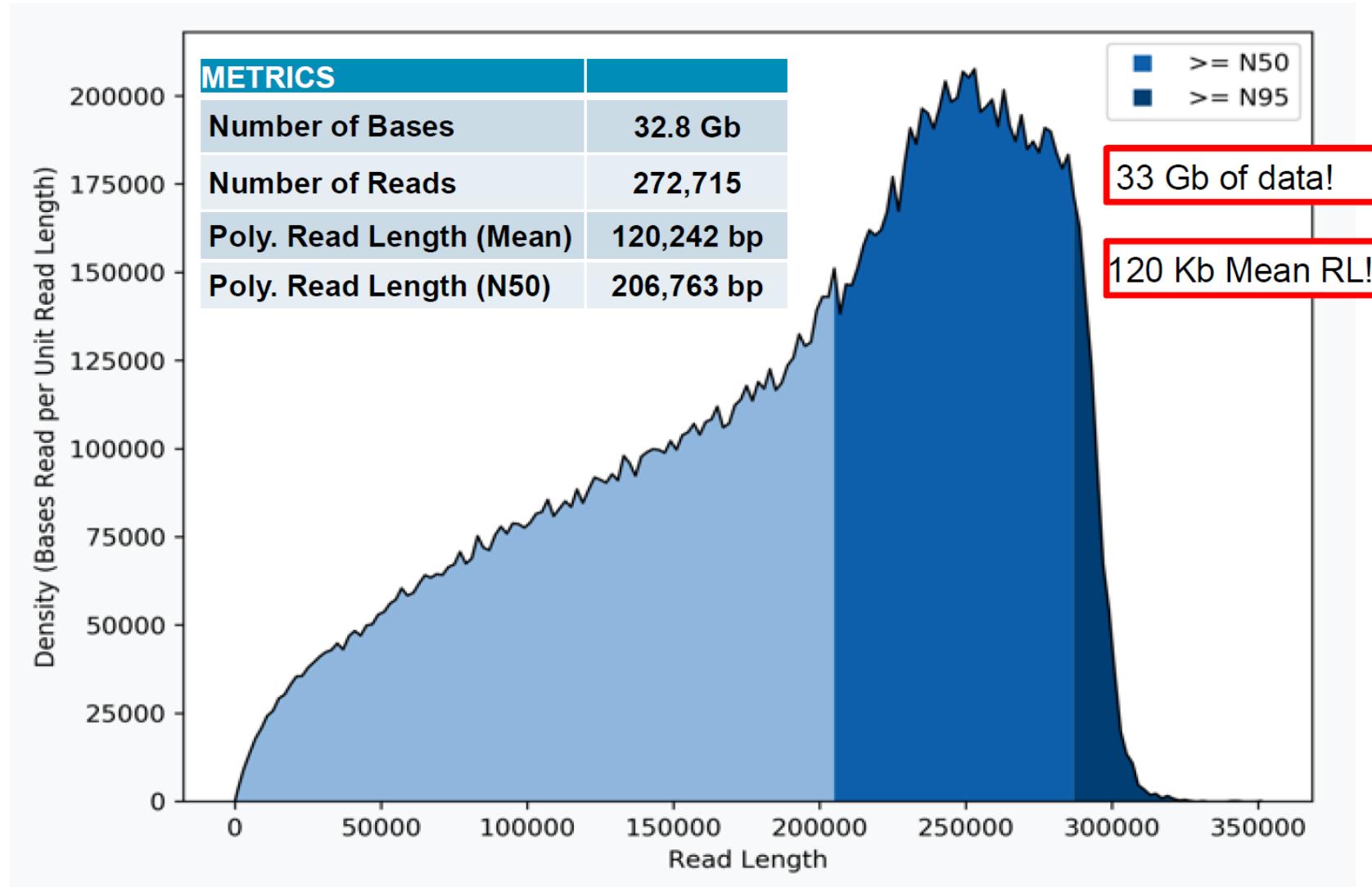


SYSTEM V6.0 IMPROVED LONG INSERT CONSENSUS





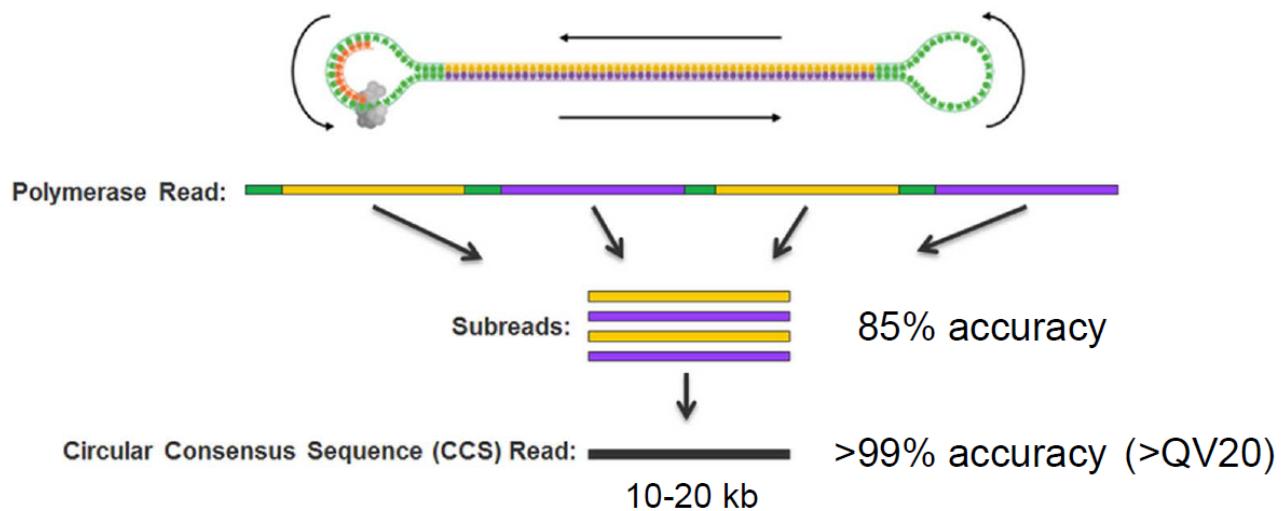
EXAMPLE SYSTEM V6.0 DATA – 14-16 KB GRAPE GENOME LIBRARY





HIGHLY ACCURATE, SINGLE MOLECULE LONG READS

Significant increase in polymerase read length in the V6.0 release allows for generation of high Q value CCS reads from 10-20 kb genomic insert libraries



HG002 CCS FALCON ASSEMBLY



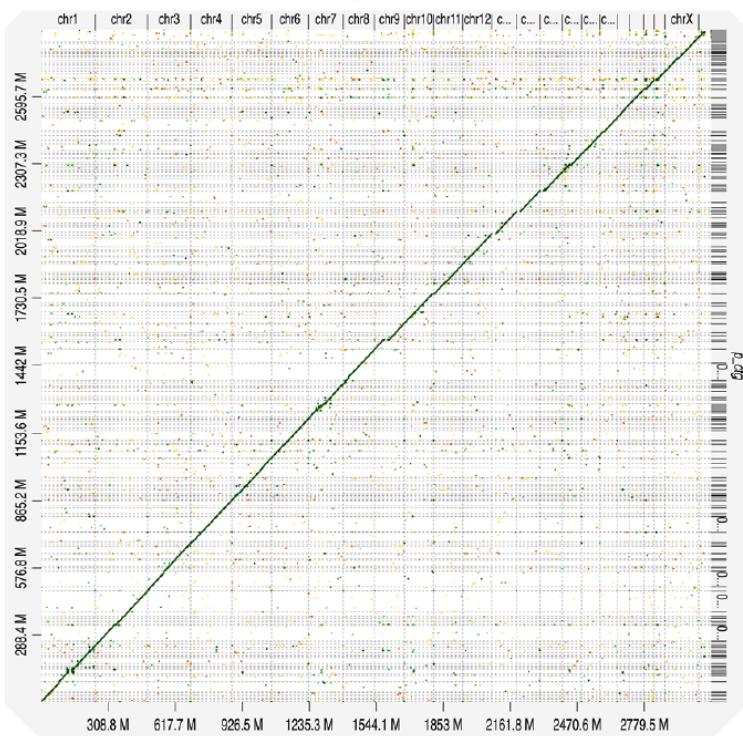
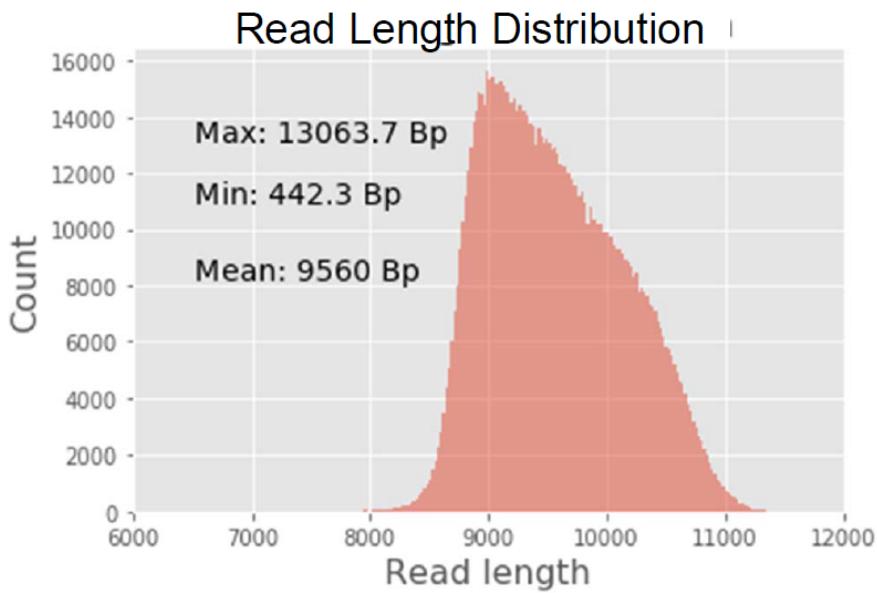
Input Data

Reads	91,811,068
Raw BP	87.8 Gbp
Coverage	30x
Average	9,567 Bp

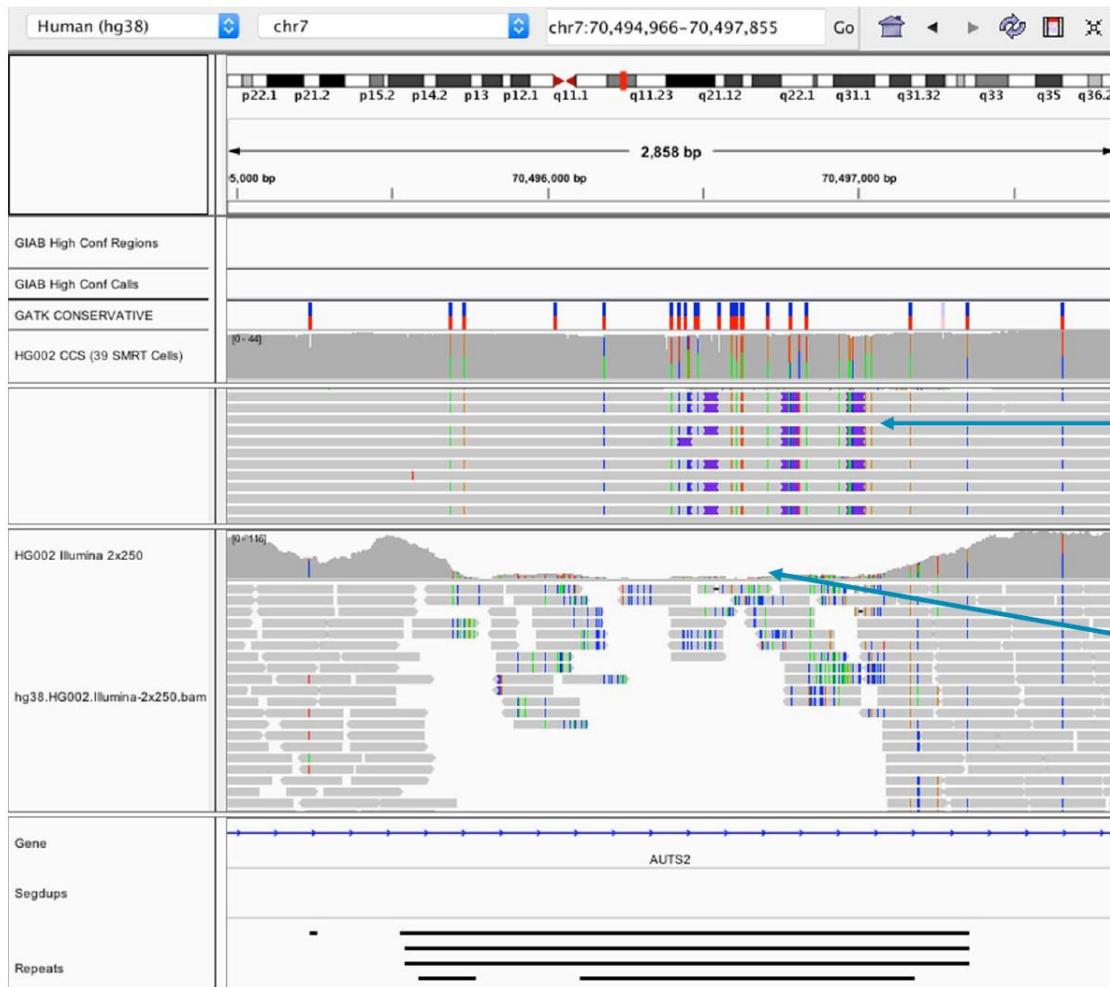
Assembly

# Contigs	4,364
Max contig	89.1 Mb
Total size	2.884 Gbp
N50	18.1 Mb

hg38



INTRONIC INSERTION IN AUTS2 (AUTISM)



PacBio reads
show insertions

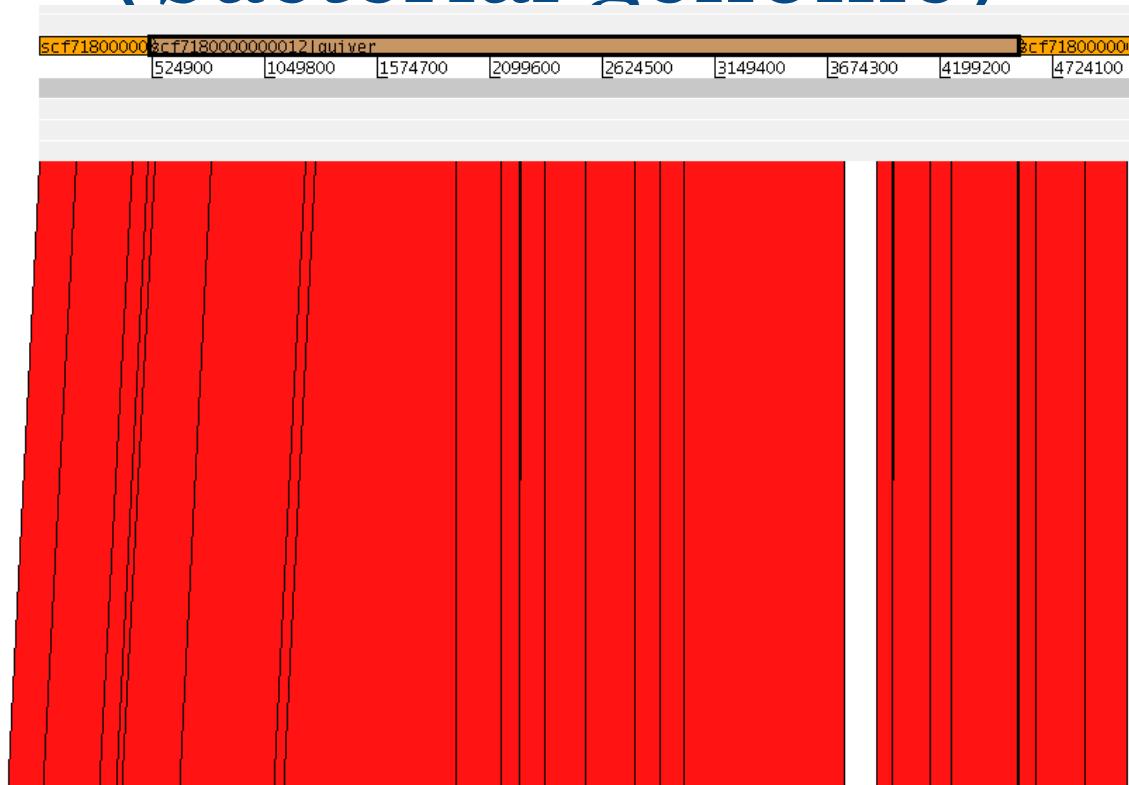
Illumina reads
do not map

Tandem repeat



Example assembly (bacterial genome)

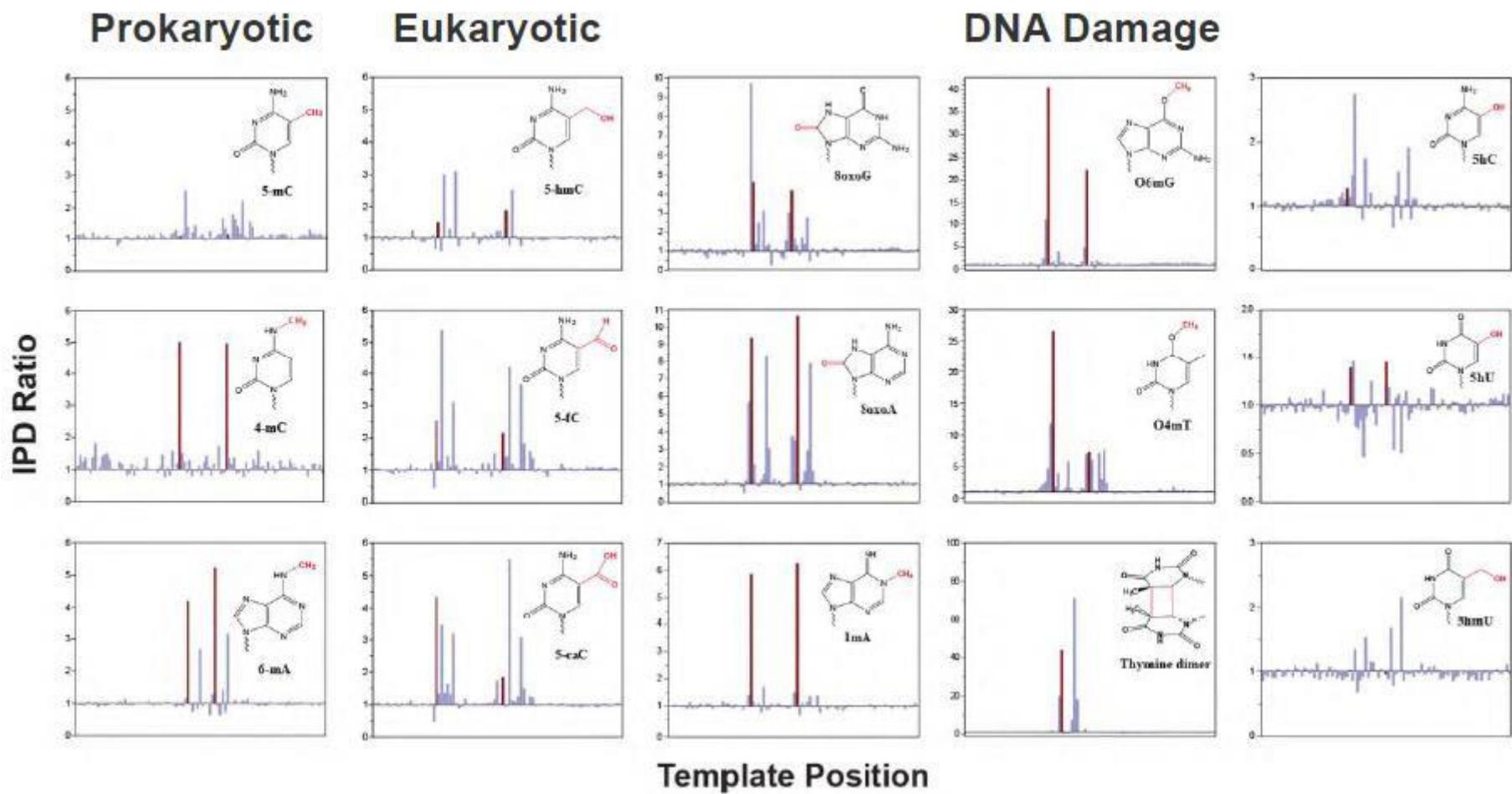
pacbio



illumina



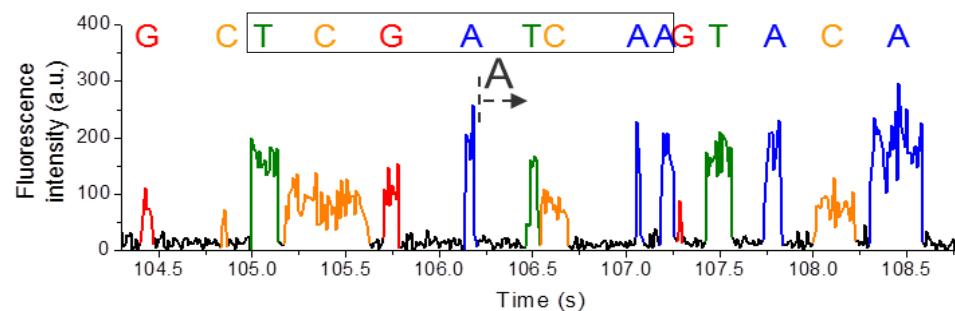
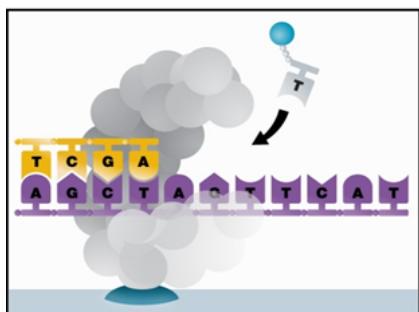
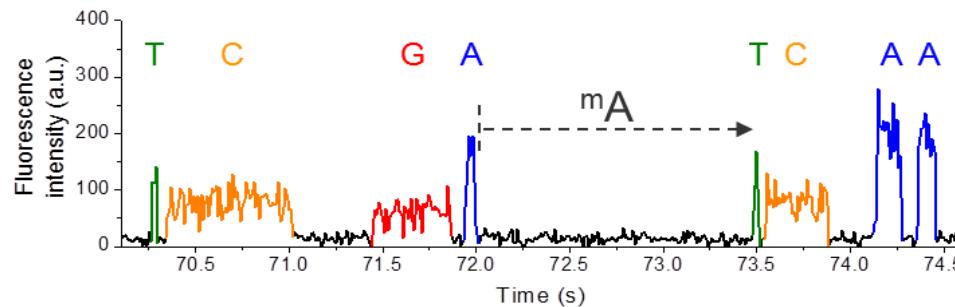
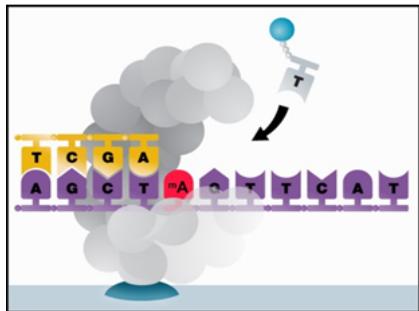
Base Modifications





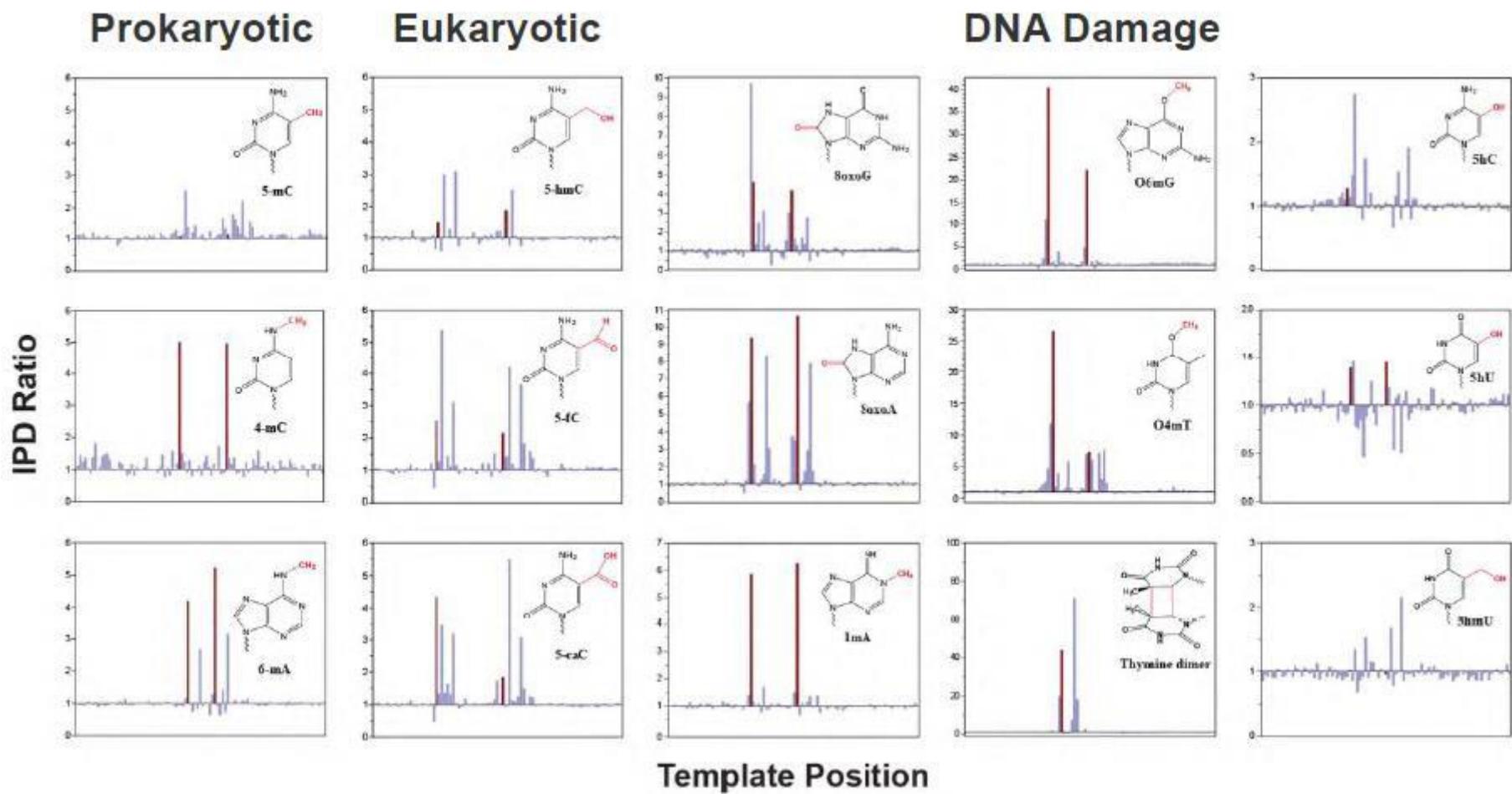
DETECTION OF DNA BASE MODIFICATIONS USING KINETICS

Example: N⁶-methyladenine

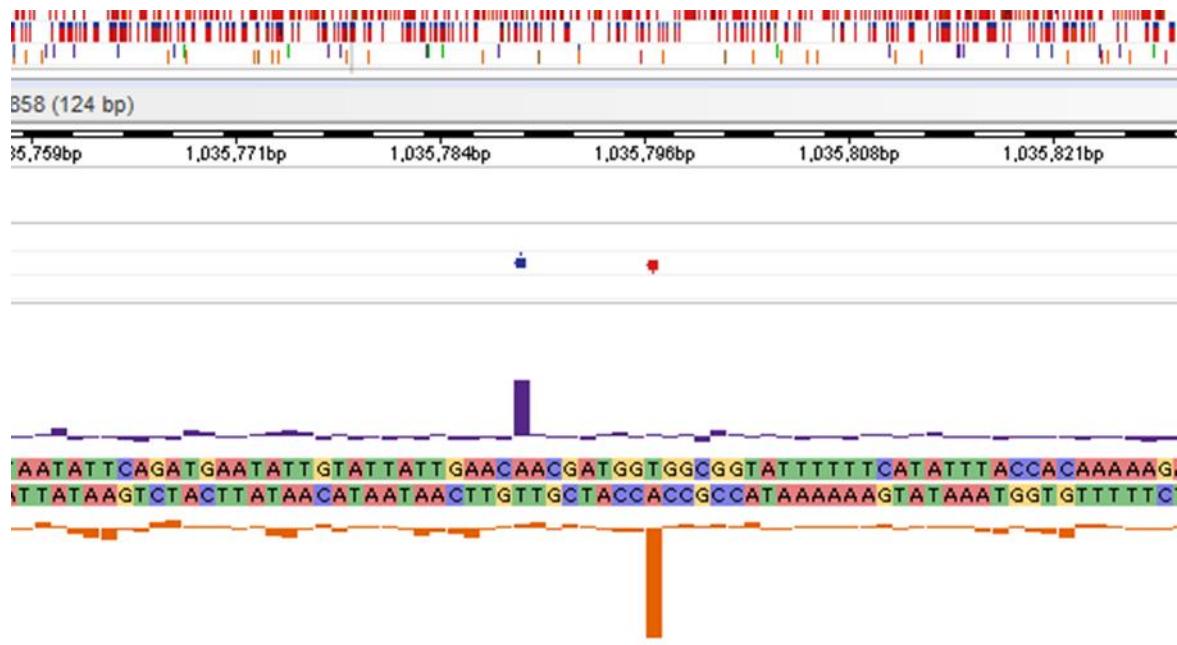


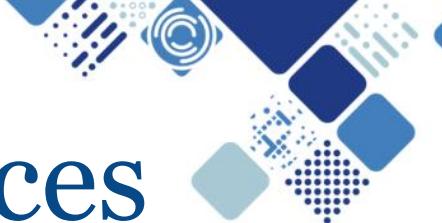
- SMRT Sequencing uses kinetic information from each nucleotide addition to call bases
- This same information can be used to distinguish modified and native bases by comparing results of SMRT Sequencing to an *in silico* kinetic reference for incorporation dynamics without modifications.

Base Modifications



An example of signal strength with m6A





Pacific BioSciences

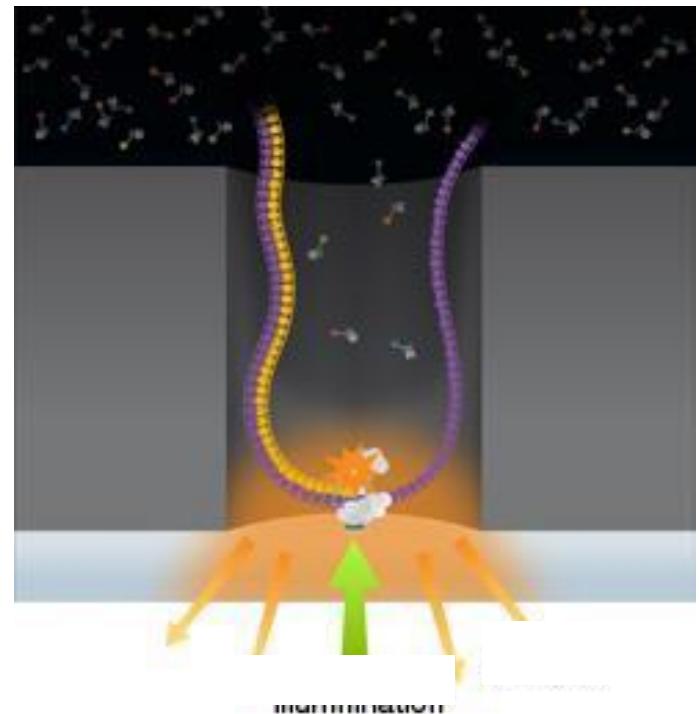
Applications

- » Long read applications
- » De Novo sequencing
- » Full length cDNA sequencing
- » Haplotyping

- » DNA modification studies

Pacific Biosciences

- » Single polymerase mol. in a 20nm hole
- » Watch incorporation in real time
- » ~2 bases per second
- » Yield 40Gb
- » Some reads 80Kb +
- » With 8M well chip in 2019 will give \$1000 genome





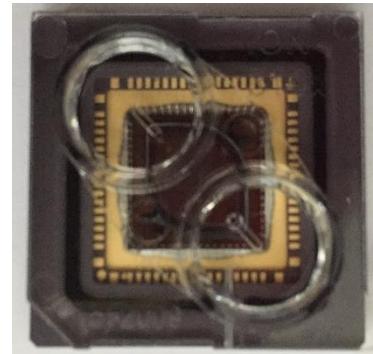
Semiconductor Sequencing



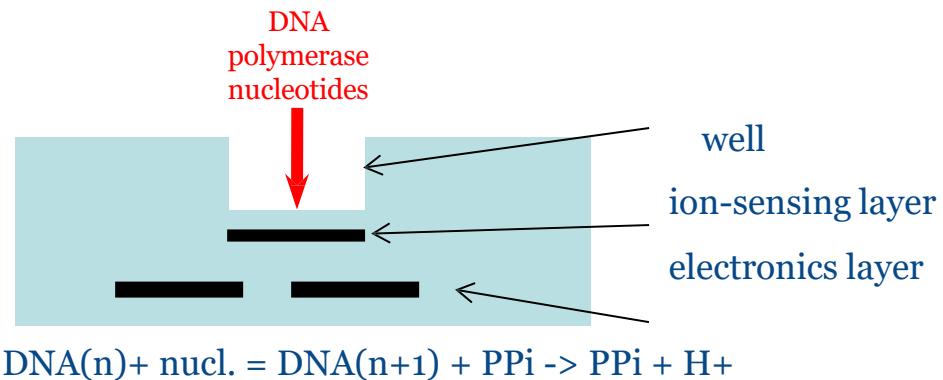
Ion Torrent's PGM



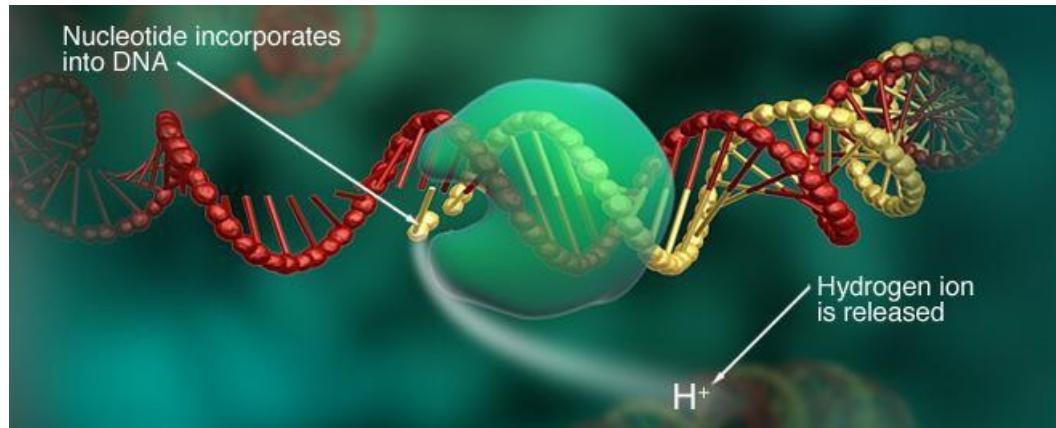
Capital cost \$50,000



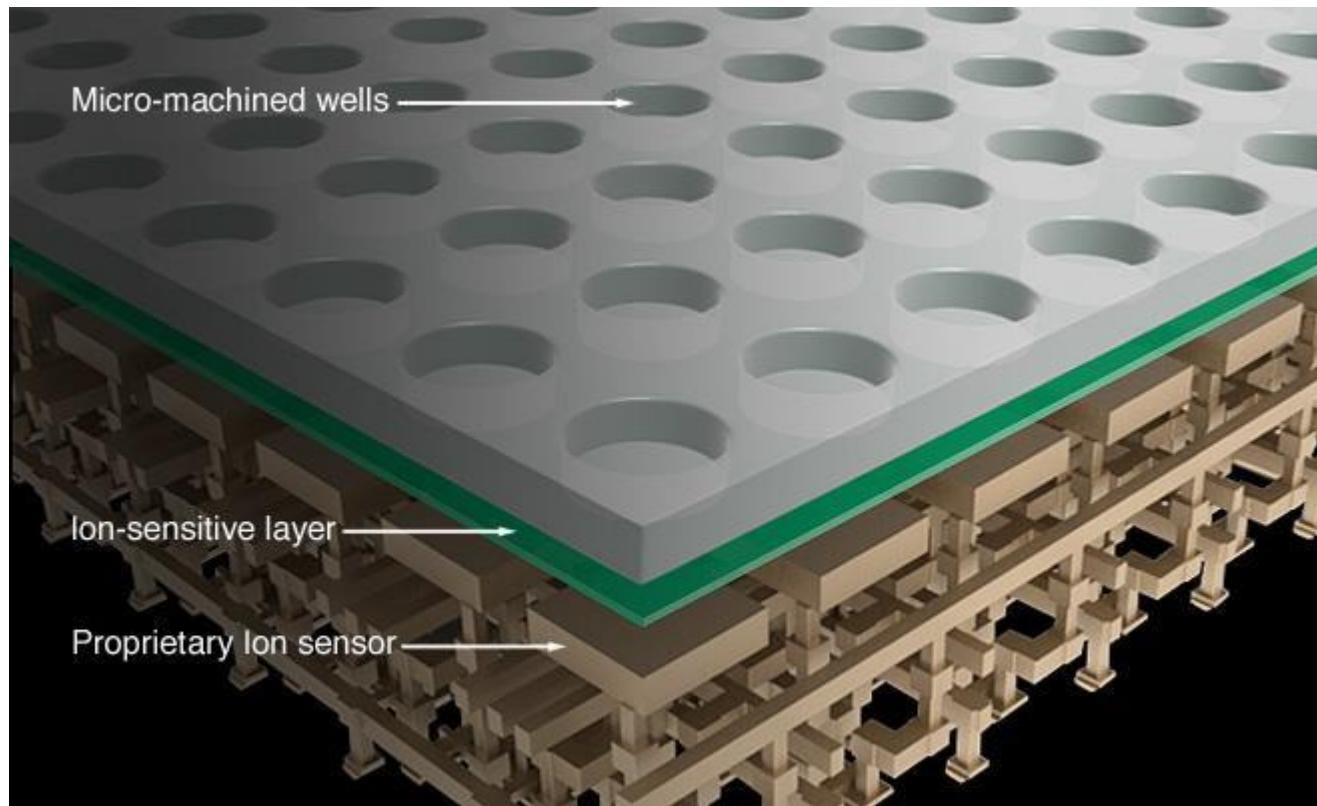
Ion Torrent's Technology



Flow through A then T then ... like 454



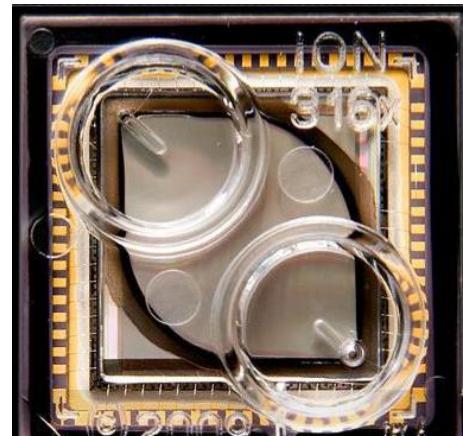
Ion Torrent Chip Drawing





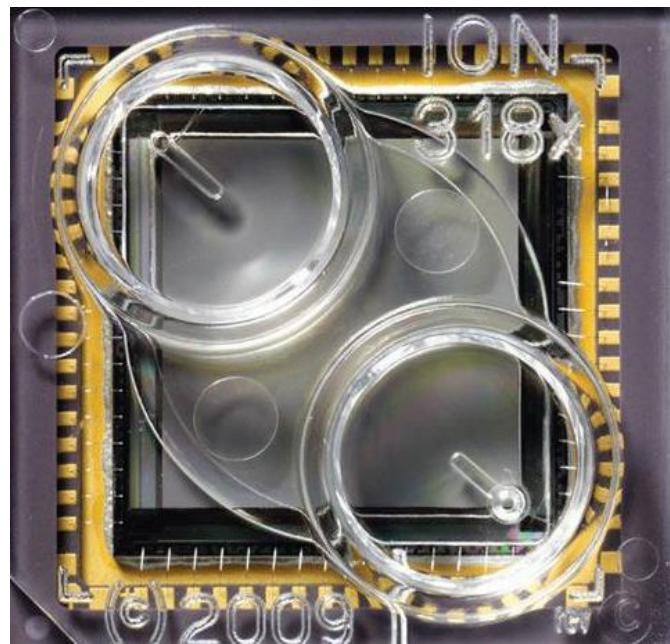
Ion Torrent

- » Library prep like original 454
- » Amplification on beads by emPCR
- » CMOS chip detection
- » Cyclic addition sequencing - pH changes
- » Not single-molecule



Ion Torrent 318 (PGM)

- » 11 million wells
- » 1 Gb / run
- » Read length ~400
- » Error rate 1-2%
- » 2-4 hr. run time
- » \$1000 / Gb (\$75K)





Ion AmpliSeq™ technology: As Simple As PCR Your Targets, Your Genome, Your Panel

The most comprehensive gene coverage
with the lowest amount of DNA or RNA Input

Simple

- 10 ng of DNA per pool
- FFPE-compatible
- PCR-based target selection



Scalable

- Up to 24,000 primers per pool
- 1–1000s of genes
- 96 barcodes for multiplexing



Fast

- 1 day from DNA to results
- 2 hours to design custom panels
- 3.5 hours for target selection and library preparation



Coming Soon: Ion AmpliSeq HD Technology



Technology exclusively available for Ion Torrent™ customers

Novel core technology with the ability to process mixed or challenging sample types—extendable to multiple applications

Customizable
Design flexibility. Add or remove content
as your biomarker of relevance changes

Low sample input down to 1ng
To identify rare variants

Achieve down to 0.1% LOD
With custom panels

Process multiple sample types
Ideally suited for cfDNA and
FFPE samples.

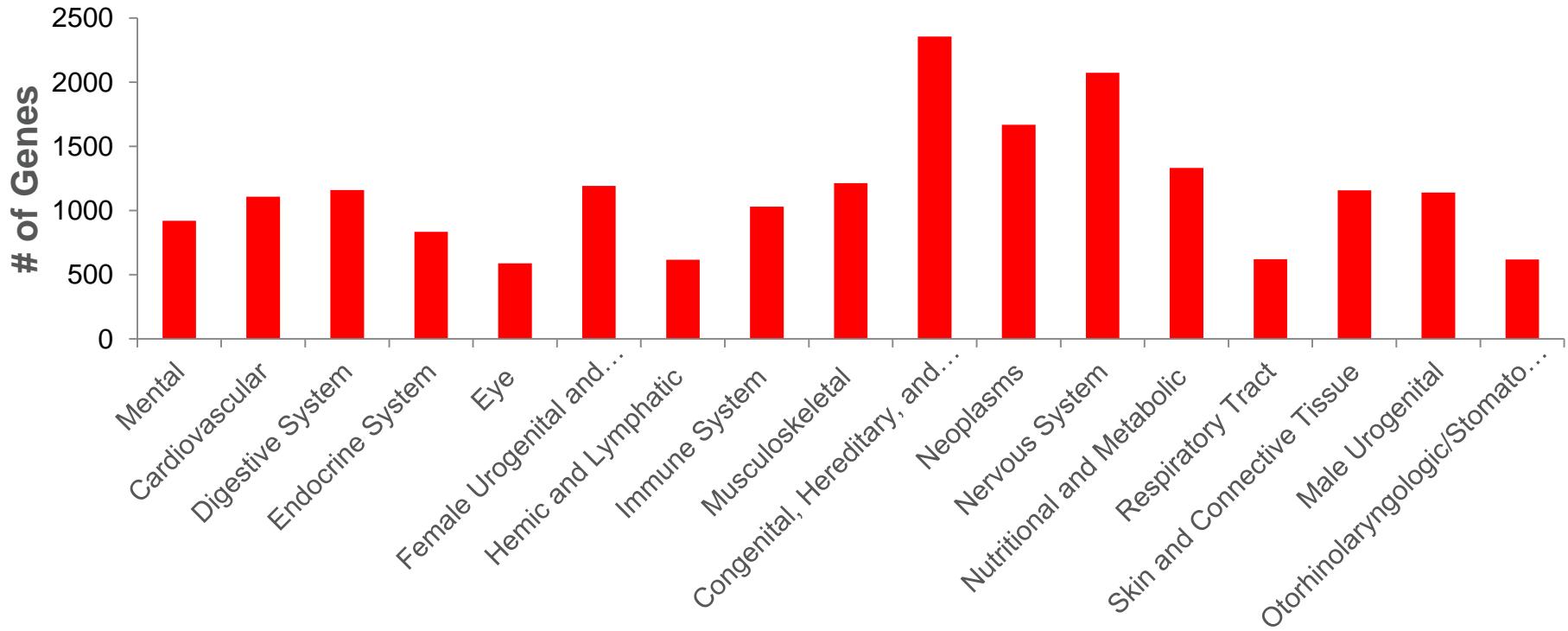
2–3 hour prep time
Shorter turn-around time

Scalable technology
For all variant types found in DNA and RNA

The content provided herein may relate to products that have not been officially released and are subject to change without notice.

For Research Use Only. Not for use in diagnostic procedures.

Expanded Gene Content Across Disease Research Areas



Average of 1154 genes per major UMLS disease research area category

For Research Use Only. Not for use in diagnostic procedures.

Ion GeneStudio S5 Series | Flexible Portfolio Configurable to Your Needs



Ion GeneStudio™ S5



Fast.



Ion 510™
Chip
2–3 M reads
Up to 400 bp



Ion 520™
Chip
3–6 M reads
Up to 600 bp

Ion GeneStudio™ S5 Plus



New

Flexible.

Ion GeneStudio™ S5 Prime



New

Powerful

				New



Ion 530™ Chip
15–20 M reads
Up to 600 bp



Ion 540™ Chip
60–80 M reads
Up to 200 bp



Ion 550™ Chip
100–130 M reads
Up to 200 bp

For Research Use Only. Not for use in diagnostic procedures. * Throughputs based on 200bp sequencing

Output and Turn-Around Time to Meet Your Lab's Peak Volume Needs



Ion GeneStudio™ S5



Ion GeneStudio™ S5 Plus



Ion GeneStudio™ S5 Prime



Speed*	19 hrs	10 hrs	6.5 hrs
Output (max/day):	15 Gb/80 M	30 Gb/160 M	50 Gb/260 M
Chips (max/day):	1 x 540	<u>2 x 540</u> or 1 x 550	2 x 550

* Based off 540 chip – sequencing (2.5 hours) and analysis (varies) time

For Research Use Only. Not for use in diagnostic procedures.



Automated Ion AmpliSeq™ library construction

Automated template preparation

Sequencing

Pipet
sample and primer
pools into Ion Chef™
cartridge for library
prep

Load cartridge
onto Ion Chef System

Pipet
library into Ion Chef
cartridge for templating
and chip loading

Load cartridge
onto Ion Chef System

Load reagents
onto Ion GeneStudio™
S5 System

Transfer chip
to Ion GeneStudio S5
System for sequencing

15 min
hands-on time

15 min
hands-on time

15 min
hands-on time

For Research Use Only. Not for use in diagnostic procedures.

List prices 2018

- Ion GeneStudio S5™ System A38194 Runs 510, 520, 530 and 540 chip. **50,528 GBP**
- Ion GeneStudio S5™ Plus System A38195 Runs 510, 520, 530, 540 and 550 chip. **104,942 GBP**
- Ion GeneStudio S5™ Prime System A38196 Runs 510, 520, 530, 540 and 550 chip. **132,150 GBP**
- Ion Chef(TM) System 4484177 **45,240 GBP**

All instruments include 12 months warranty

Promo: Trade-in any current NGS or CE instrument for up to 50% discount.

For Research Use Only. Not for use in diagnostic procedures.

World's first complete Sample to Insight NGS solution

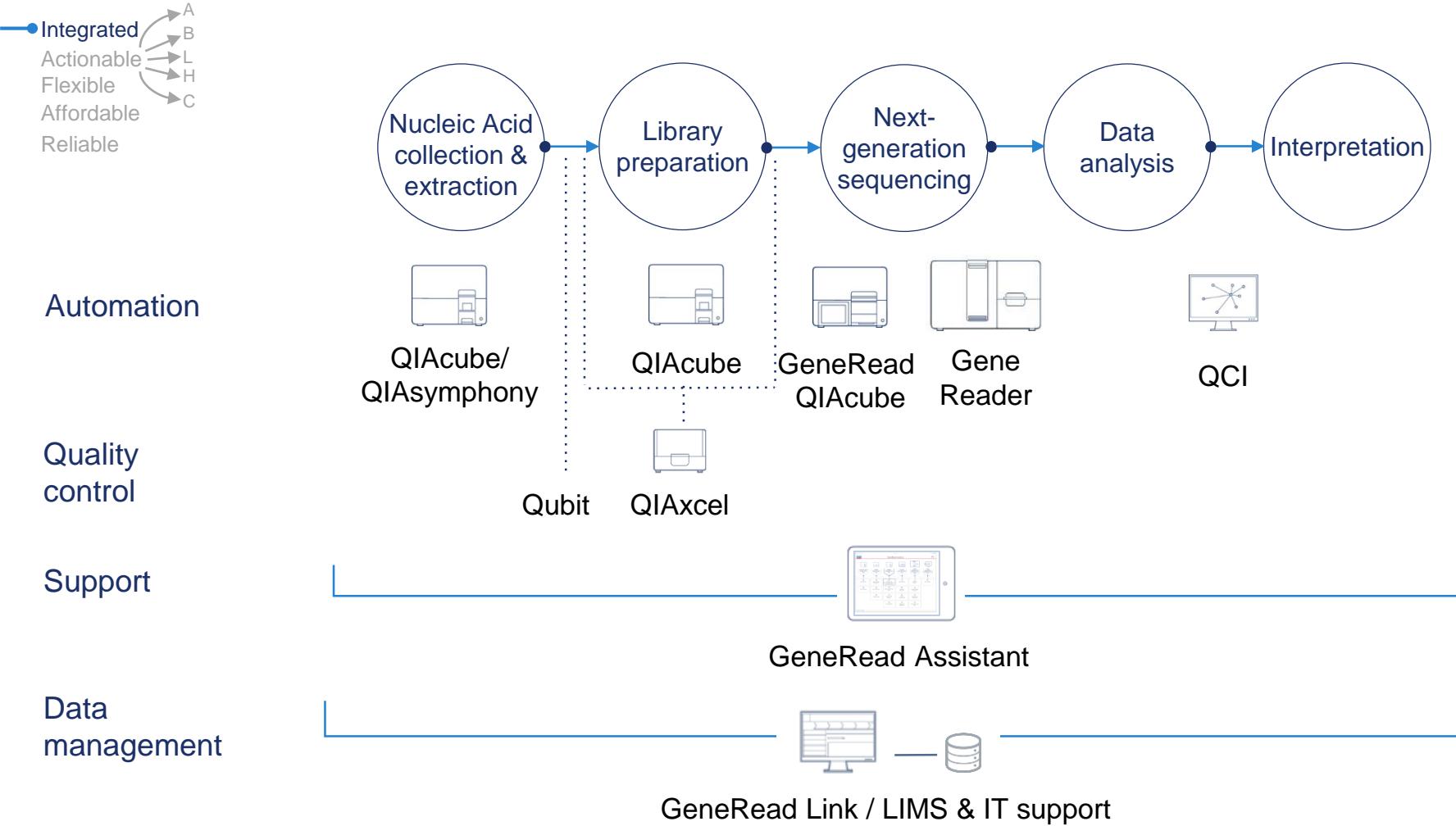
Designed to deliver actionable insights



- SBS chemistry developed by intelligent biosystems
- Sequencing system that can prep DNA, do targeted library prep and run upto 20 x 1Gb yield flowcells
- Sequencing by synthesis but one base at a time



First truly complete NGS solution



Offering everything needed from one vendor for NGS solutions

Best amplicon & insight coverage, plus best batching flexibility



Integrated
 Actionable
 Flexible
 Affordable
 Reliable

A
B
L
H
C

	GeneRead Actionable Insights Tumor Panel	ILMN Tumor 15 Panel	AmpliSeq™ HotSpot Tumor Panel V2
Panel size	12 genes / 16.7 kb	15 genes / 44 kb	50 genes / Not provided
Insight size	773 unique variant positions	Not available	2855 variants
Amplicons	330	250	207
Average amplicon size	134 bp	150–175 bp	154 bp
DNA input	10 ng x 4	10 ng x 2	10 ng x 1
Throughput	1–48 samples per run	8 samples per run	2/8/16 samples per chip
Variant frequency	5%	5%	5%
Amplicon coverage	>500x: 96.4% >200x: 98.5% Average: ~5000x	>500x: 93.5% >200x: not available Average: n/a	>500x: not available >200x: not available Average: 1000-4000x
Variant insight coverage	>500x: 99.8% >200x: 99.9%	Not available	Not available

Note: Coverage is average of 12 NA12878 datasets

The ultimate objective of NGS in molecular pathology research

AnyGenomics Lab  **QIAact Actionable Insights Tumor Panel
For the GeneReader NGS System (RUO)**

Report Date Jan 1, 2015
Report Status FINAL

Sample Metadata		Laboratory Information		Specimen	
Sample ID	Sample 1	Lab	FacilityName	Specimen Type	Biopsy
Date of birth	Aug 5, 1967	Technician	Technician ID	Specimen ID	Specimen ID
Ethnicity	African	Scientist	Scientist ID	Collection Date	Jan 1, 2015
Sex	Male			Accession Date	Jan 3, 2015
Accession	1somatic_max_api_test			Primary Tumor Site	Blood Vessel
				Diagnosis	stage 42 glioblastoma

Interpretation

2 Clinically Significant Variants Reported 2 Approved Therapies 2 Potential Clinical Trials

2 alterations were identified that may potentially be responsive to other treatments.
2 clinical trials were identified that target the detected alterations. One alteration is associated with resistance

Summary of Clinically Significant Variants

Variants Reported	FDA Approved Therapies for Indication	FDA Approved Therapies for Other Indications	Therapies Associated with Resistance	Potential Clinical Trials
EGFR p.E746_A750del				2 potential trials

Variant Details

Gene	Exon #	Nucleotide Change	Amino Acid Change	Effect on Protein
EGFR	19	NM_005228.3: c.2235_2249delGG	p.E746_A750	Gain of Function

EGFR is an oncogenic receptor tyrosine kinase involved in cell survival by regulating the RAS and PI3K pathways. Gain of function mutations and amplification cause EGFR activation. Somatic mutations have been reported in 36% of NSCLCs, 8.6% of gliomas and 2% of breast cancers [PMID: 15938733]. EGFRvii allele is associated with poor prognosis in patients with glioblastoma. EGFR amplification in patients with EGFR mutation-positive NSCLC is associated with increased resistance to osimertinib, erlotinib and gefitinib therapies. In over 40% of patients with NSCLC, EGFR amplification coexists with EGFR mutations [PMID: 17666241]. TKI-sensitive EGFR mutations are more common in female never-smokers and in patients with adenocarcinomas [PMID: 14555582].

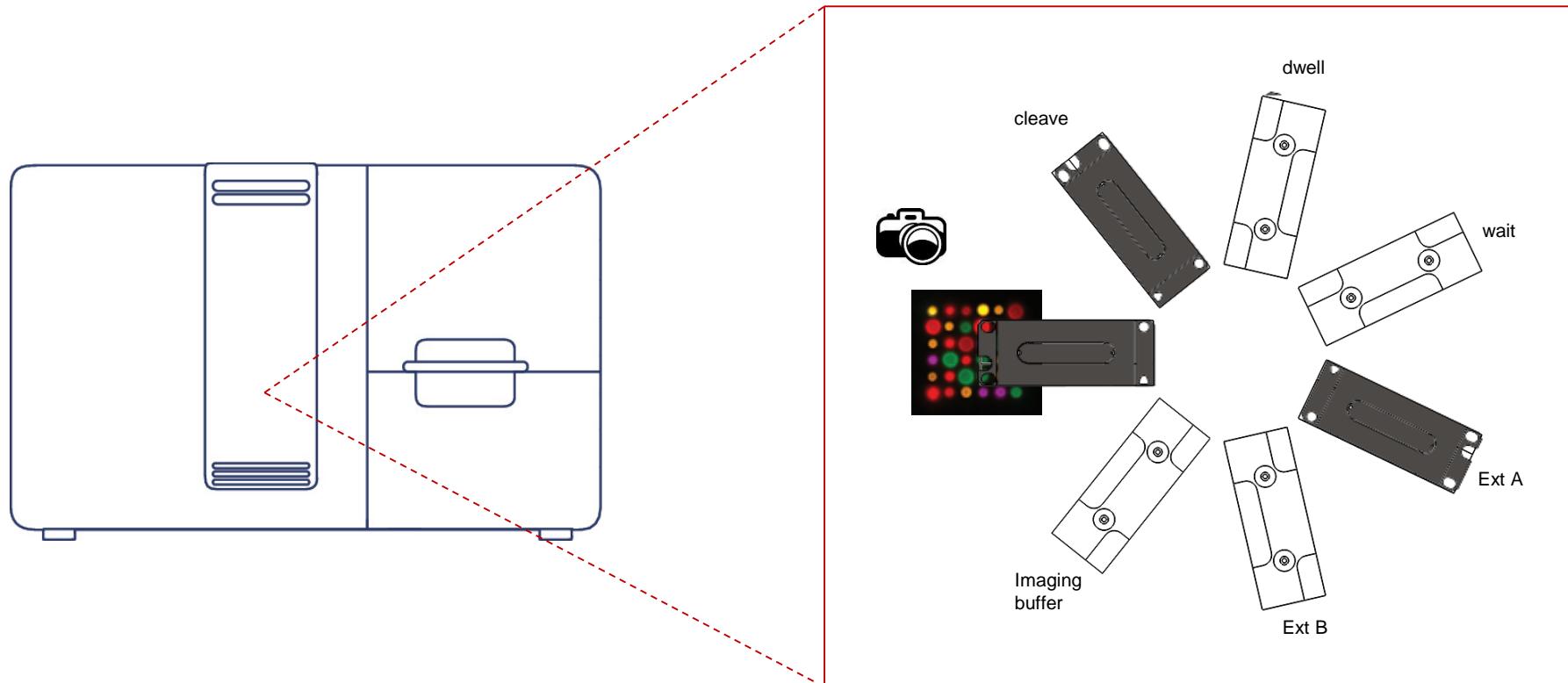
Page 1 of 3
QIAGEN | support.qiagen.com | QIAGEN.com
The QIAGEN GeneReader® is intended for research use only (RUO). This product is not intended for the diagnosis, prevention or treatment of a disease.

Sequencer behaving like 3 sequencers in 1

Integrated
Actionable
Flexible
Affordable
Reliable

A
B
L
H
C

Turntable design



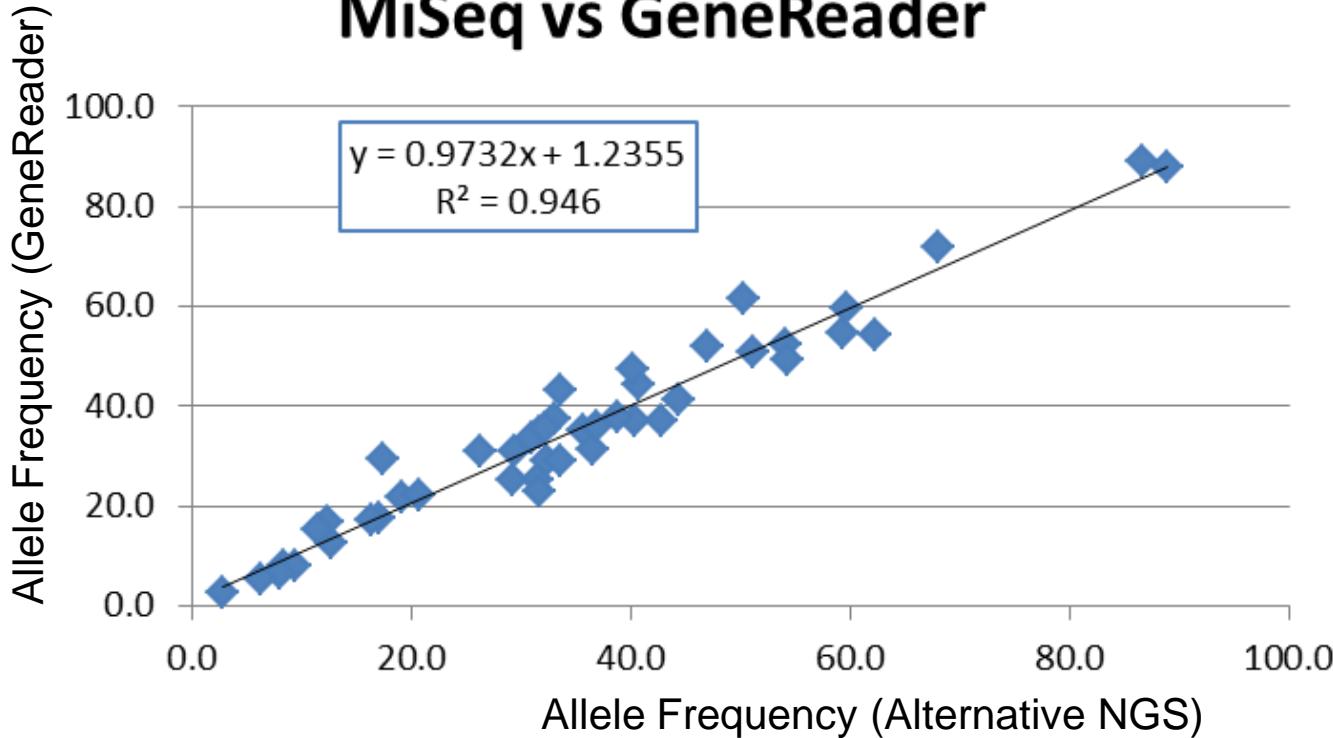
- Scaling up and down depending on number of samples

Rigorous testing by customers: Case 2

- Integrated A
- Actionable B
- Flexible L
- Affordable H
- Reliable C



MiSeq vs GeneReader



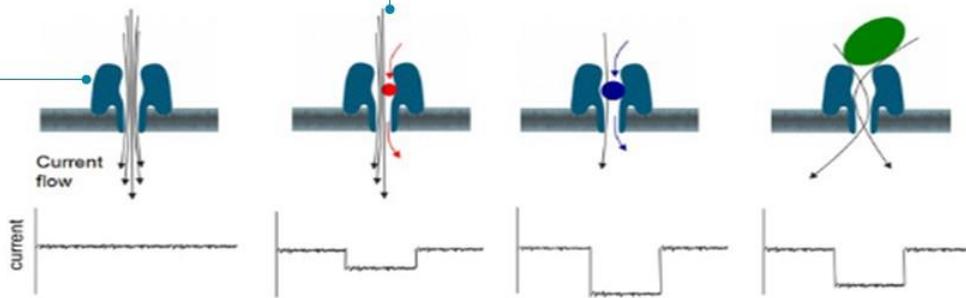


Nanopores

NANOPORE SENSING

1

Nanopore creates hole in membrane
Current passes through nanopore



2

As analyte passes through or near the nanopore,
this creates characteristic disruptions in the current

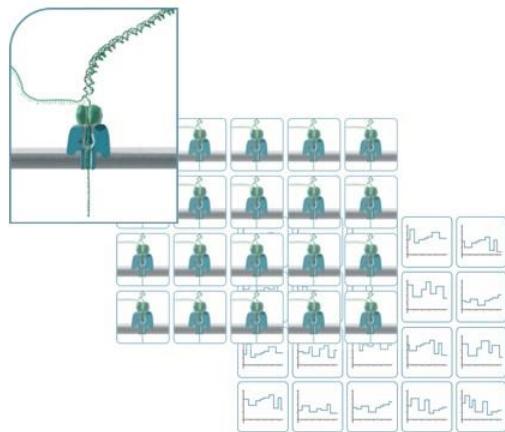
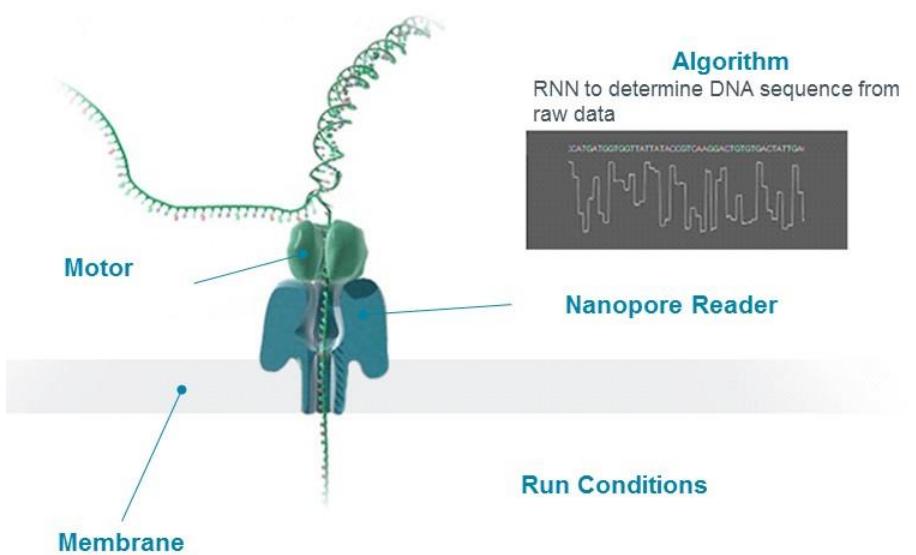
3

Current disruption is interpreted to understand the
identity of the analyte

An explanation from  Oxford
NANOPORE
Technologies™

NANOPORE DNA SEQUENCING

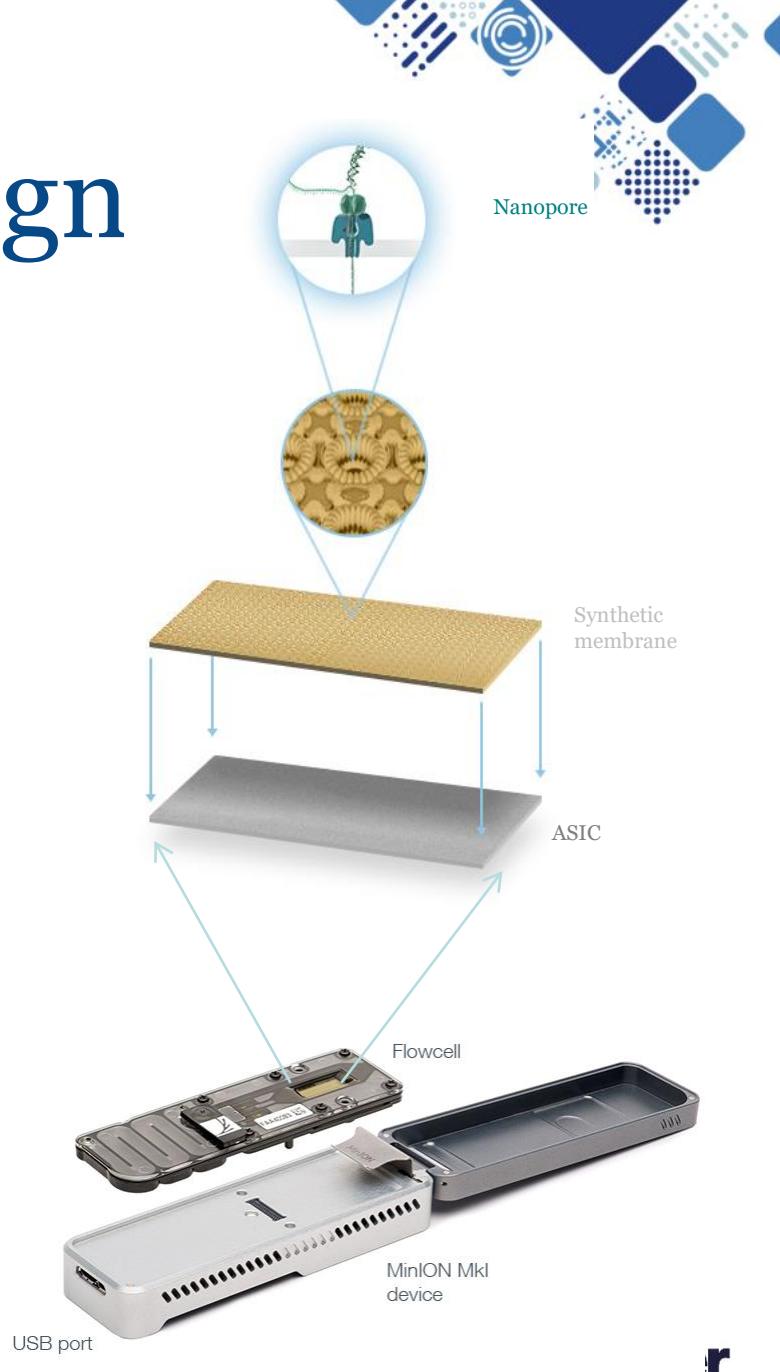
How does it work?



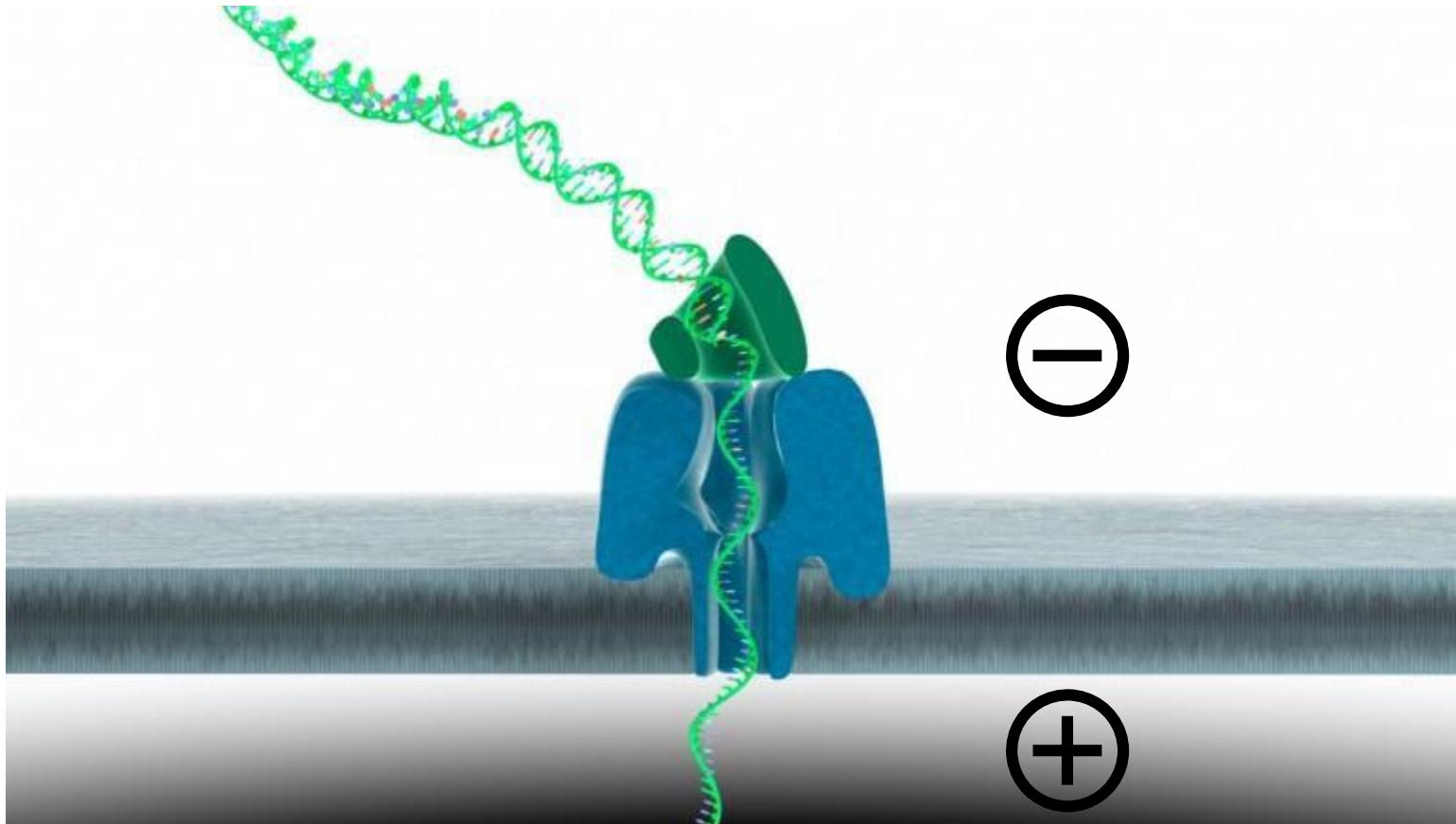
An explanation from  Oxford **NANOPORE**
Technologies™

Flow cell design

- » Application-Specific Integrated Circuits (ASICs) contains 512 channels
- » Each channel is surrounded by 4 pores & records only 1 at the tie
- » 512 pores max recorded at the time
- » Scan for “fresh” active pores automatically every 24h or when manual restarted



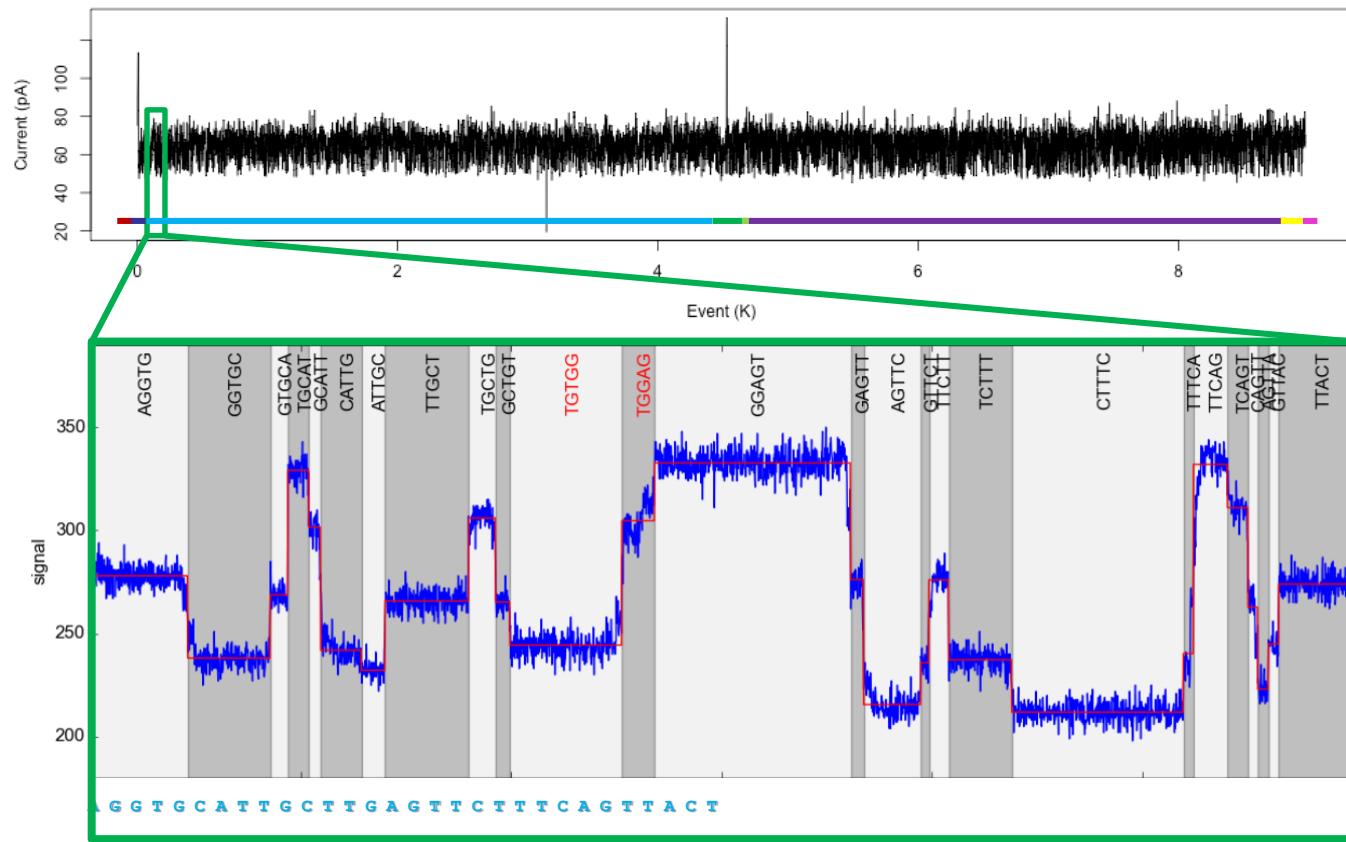
Oxford Nanopore Technologies



5-base words = 1024 current levels

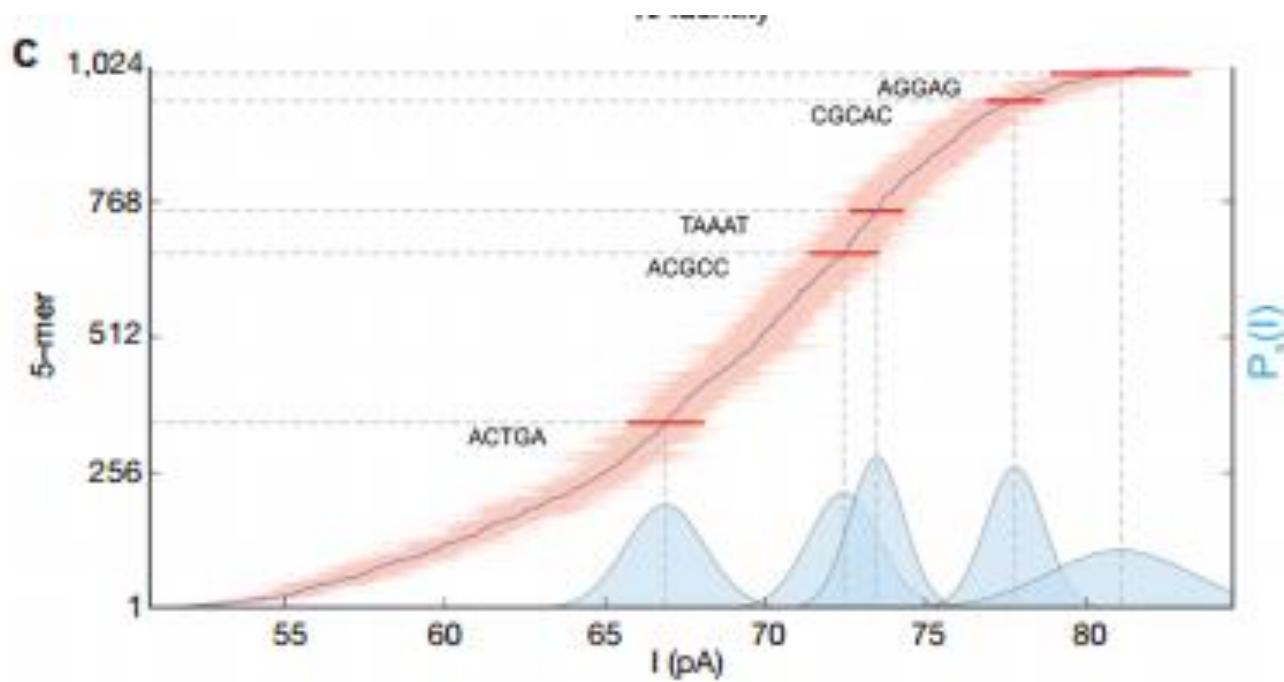


ONT: The squiggles



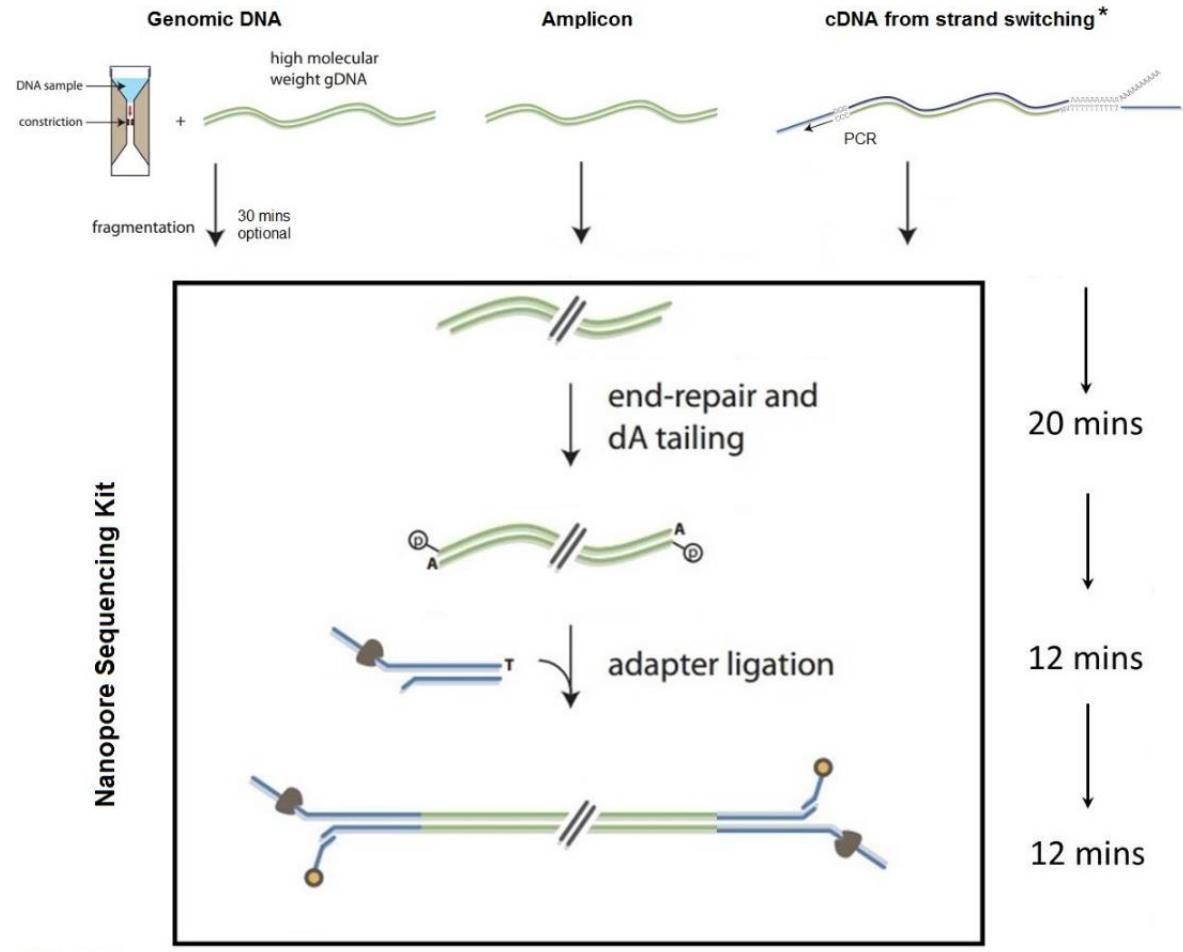
Slide courtesy of David Buck. WTCGH

5mer current signals



From Szalay & Golovchenko, Nat. Biotech., 2015.

ONT 1D Library prep



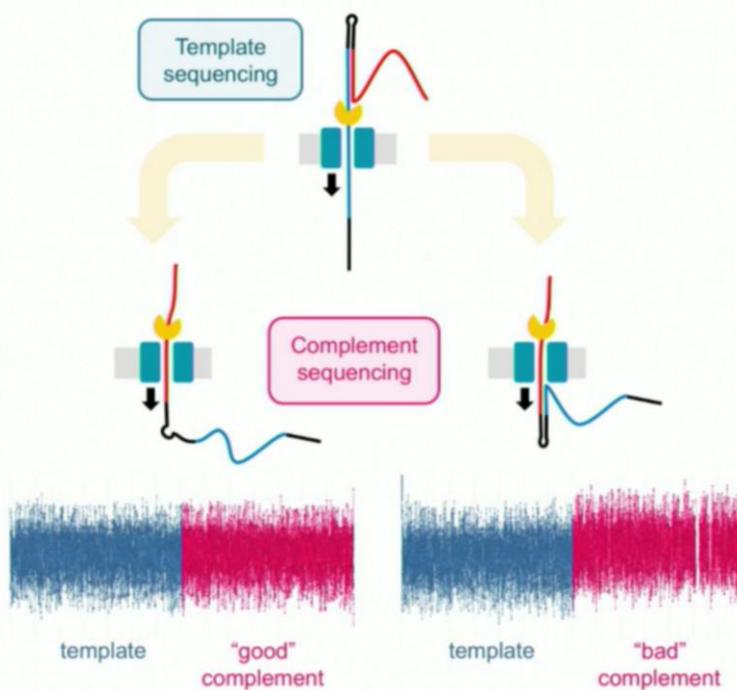
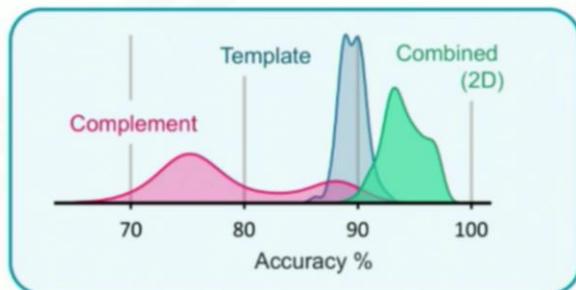
1D²

1D²

Problems with 2D

2D HAS BEEN CHALLENGING FOR MANY YEARS

- 50% of sample is lost in library preparation
- Signal problems arise from 2° structure under the pore
 - Hairpins formed change the DNA signal ("uplift")
 - Secondary structure is variable, broad distribution of accuracies
 - "Uplifted" signal results in lower accuracy
 - Missing sections also present from enzyme slips
- Problems much worse with 450 bps chemistry



1D²

Nanopore Live

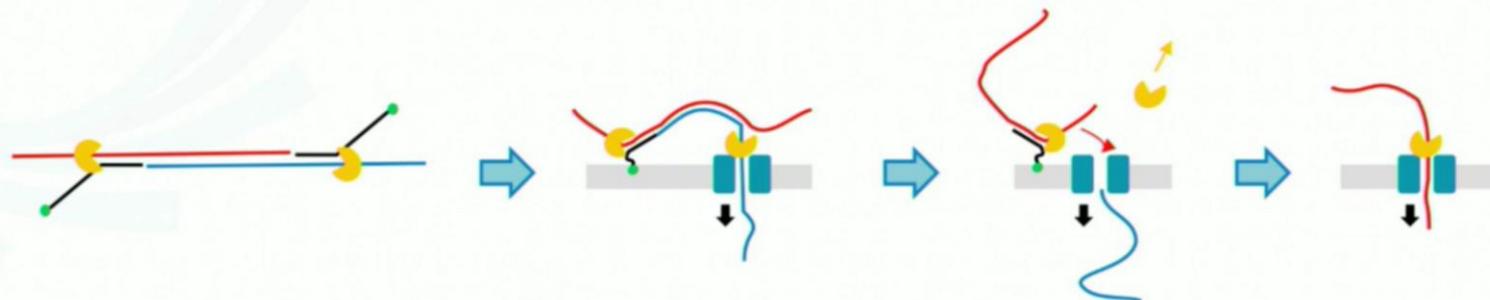
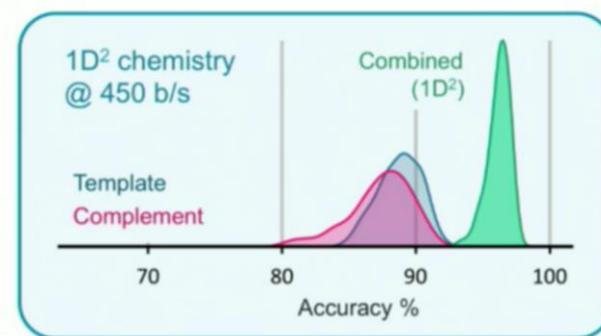
LIVE

1D²

Improved template – complement data

SEQUENCING SCHEME WHERE STRANDS ARE NOT JOINED

- Complement follows template as separate independent strand
 - Each molecule has its own motor-adapter
 - Each individual strand has high 1D accuracy
 - No secondary structure problems
- Simple library preparation, compatible addition to 1D methods
 - Compatible with E8 and 450 bps



Oxford

1D²

1D²

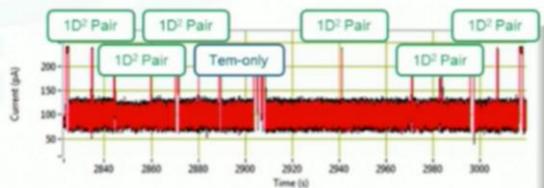
Efficiency

BUILDING ON 1D CHEMISTRY

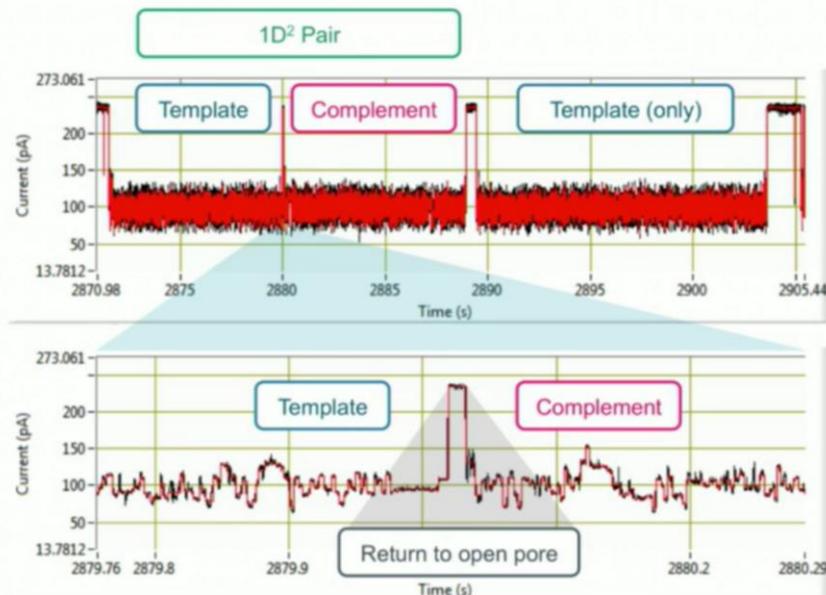
- Targeting same throughput as 1D @ 450 bps
 - 1D² is free information
- Typically the second strand immediately follows the template, with a short return to open-pore

EFFICIENCY

- 1D² occurs naturally at < 1% efficiency
- Changes to the R9.4 pore have improved efficiency
- For R9.5 > 60% of strands are 1D² pairs



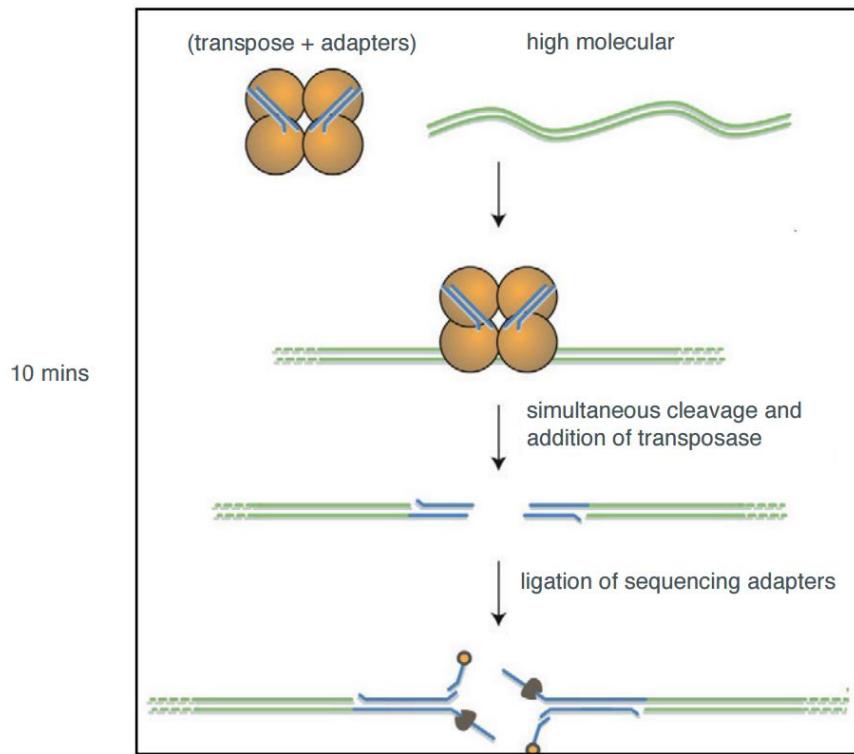
© Copyright 2017 Oxford Nanopore Technologies | 47





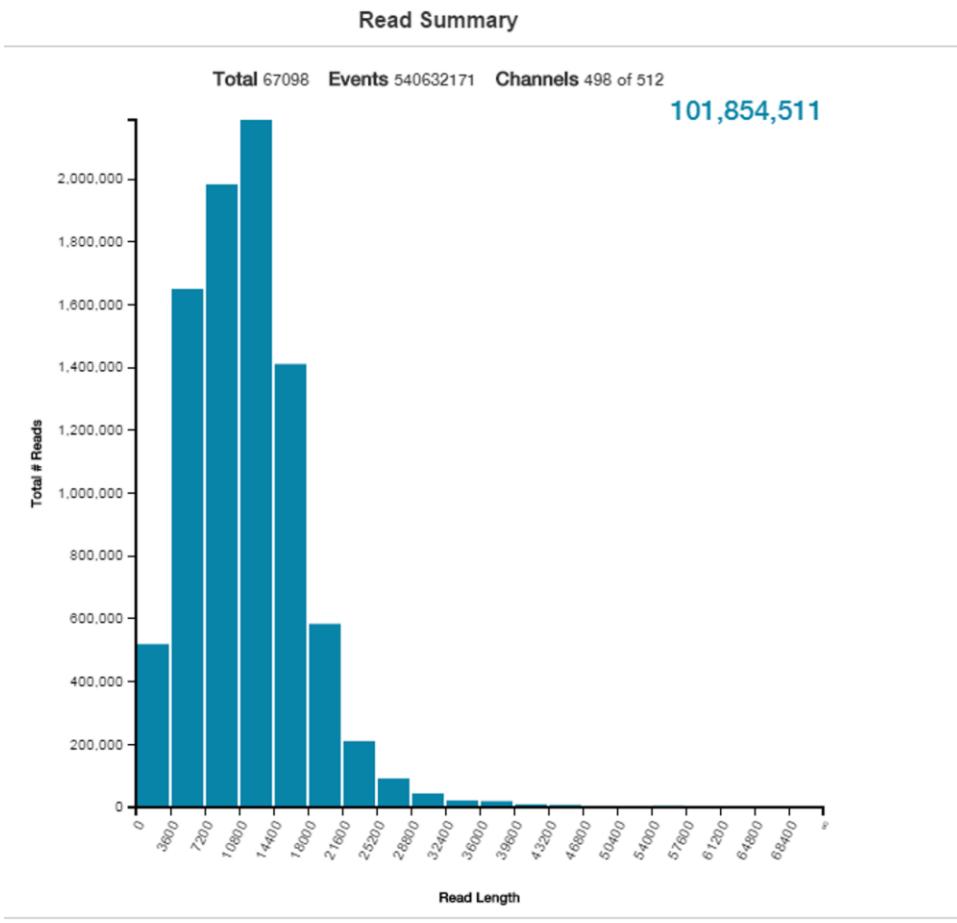
RAPID SEQUENCING KIT

A two-step, 10 minute protocol



Starting material will be fragmented; recommended starting size >30 kb for genomic DNA

MinKNOW Interface read distribution graph





Data Quality

2D error rate 2-3%, 1D error rate 8-9%

Read lengths as long as template. Record >1Mb but shorter fragments give higher yield

Some AT bias? Low coverage at high AT and GC.
Systemmatic errors in and around homopolymers

Yield per minION flowcell dependent on DNA
1-17Gb.

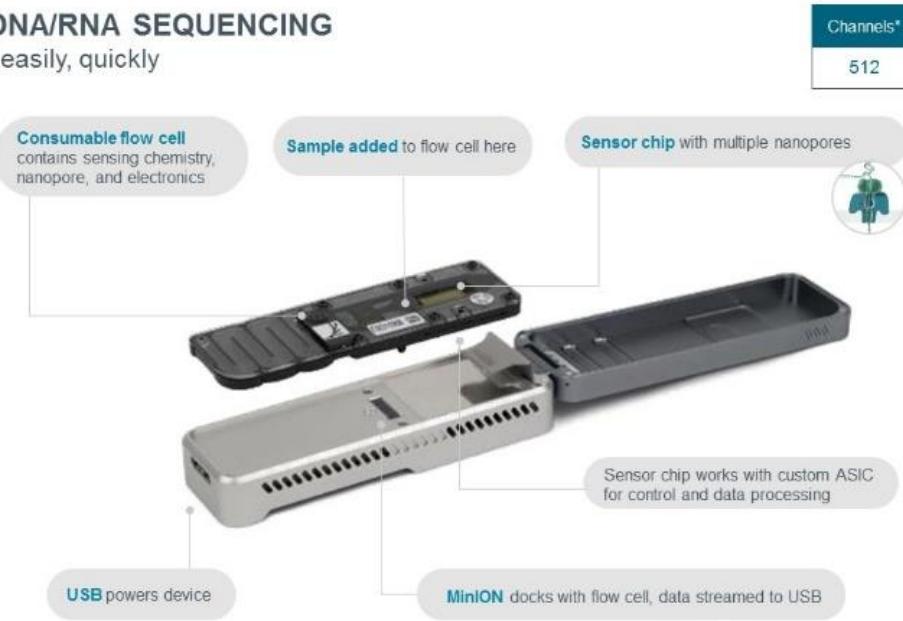
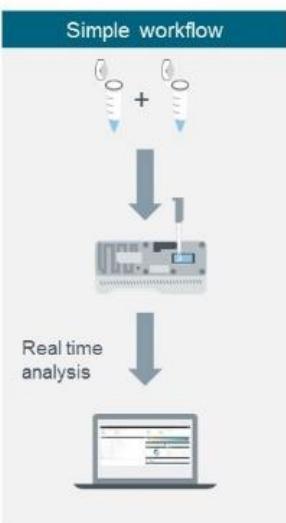


MinION: PORTABLE DNA/RNA SEQUENCING

Sample to scientific insight easily, quickly

Channels*

512



An explanation from **Oxford NANOPORE Technologies™**

GridION. For service sector

Nanopore Live LIVE

GridION X5

Bench top sequencing device

SEQUENCING

- 5 individually addressable flow cells
- Based on current MinION flow cell design
- Road map to on board Run Until... and Read Until...

COMPUTE

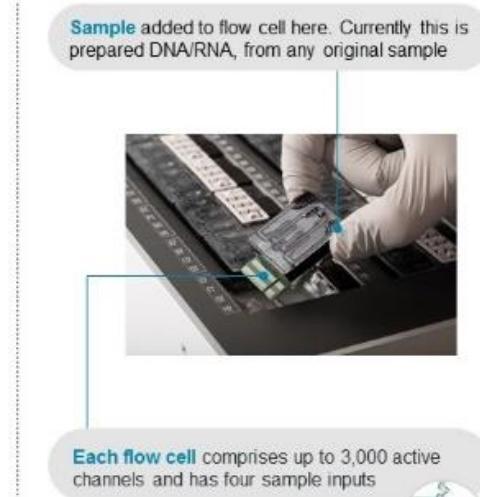
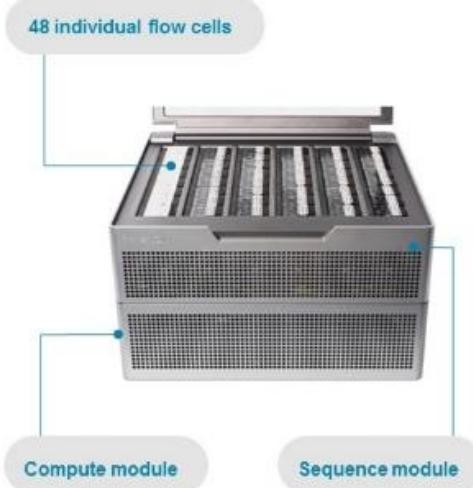
- Embedded high performance compute
- Full Real time basecalling and data analysis in the box
- Simple user interface and single ethernet for data transfer

Component	Specification
Size and weight	H200 x W 360x D 360 mm , 10 kg
Power	600 W
Compute spec:	8 TB SSD Storage, 64 GB RAM Latest Gen CPU for OS and orchestration FPGA processor
Pre-loaded software	Linux OS, MinKNOW
Connections	5x USB 3. 1x USB-C 1 x HDMI. 1 x Ethernet



Oxford

PromethION: ON-DEMAND, HIGH-THROUGHPUT



Channels*
48 x
3,000
=144,000

Instrument and flow cells

- Modular: 48 flow cells can be used individually/together for on-demand sequencing

Integrated compute module

- Real time compute provided
- Web based/remote, real time administration & monitoring

An explanation from  Oxford Nanopore Technologies

NANOPORE SYSTEMS

Summary



	MinION	GridION X5	PromethION
Sequencer type	Mobile	Benchtop	Benchtop
System Price	Starter pack of \$1000	\$0 when ordering 300 flow cells	Starter pack of \$135,000
Data produced by starter pack <i>(based on internal test Mar 17)</i>	Up to 40GB	Up to 6TB	Coming soon
Fee For Service available	No	Yes	Yes
Specifications based on internal test Mar 2017			
Run Time	1 min – 48 hours	1 min – 48 hours	1 min – 48 hours
Yield per flow cell	20GB	20GB	50GB*
Yield per Instrument run	20GB	100GB	2.4TB*

* PromethION yield still in development through the PromethION Early Access Programme

Oxford

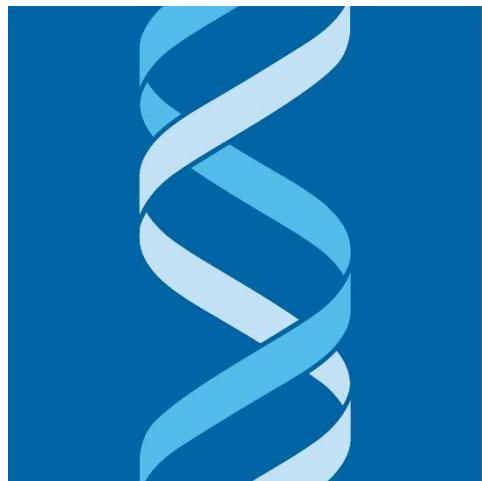


ONT applications

- » Super-long reads (>1Mb)
- » “Run until” done. W.I.M.P
- » Selective reads
- » Cas9 mediated enrichment
- » Mobile sequencing
- » Direct methylation detection?
- » Direct RNA sequencing

ONT Predicted Performance

- » MinION: 100 Mb/hr. 1-20Gb
- » PromethION: >1Tb/day
- » Read length potentially as long as molecule
- » Run until done
- » \$60-600/Gb minION, no capital cost
- » \$15-25/Gb GridION
- » \$2-10/Gb promethION



wellcome trust
sanger
institute

Current Sanger Platforms

- » ABI 3730 Capillary (outsource)
- » Illumina Xten + 5 HS2500 + 2 HS4000
- » 10 Novaseq
- » MiSeq (6)
- » Sequel(3)
- » Oxford Nanopore minION (4)
- » Oxford Nanopore gridION (1)
- » Oxford Nanopore promethION (1)



Future Sequencers

- » Roswell technologies
- » Base4
- » Quantapore
- » Gnubio
- » Genia



KEEP
CALM
ITS
**COMPETITION
TIME!**



who wants
to win...

Identify the Sequencer

1)



2)



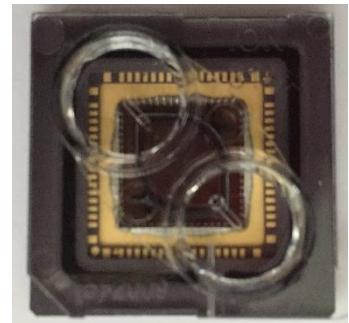
3)



4)



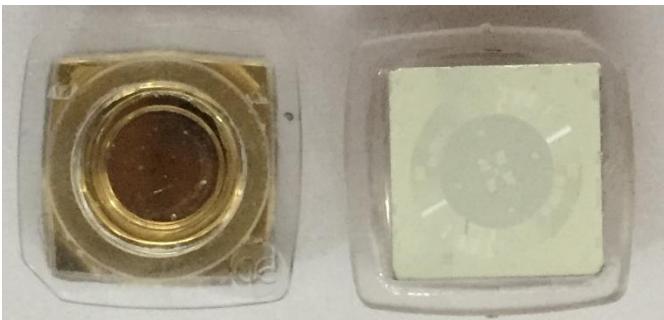
5)



- 
6. How accurate is Q20?
 7. What technology features nanoballs
 8. What does SBS stand for
 9. What does SMRT stand for
 10. Name 2 technologies that use sequencing by ligation

Identify the Sequencer

Pacific Biosciences RSII



minION

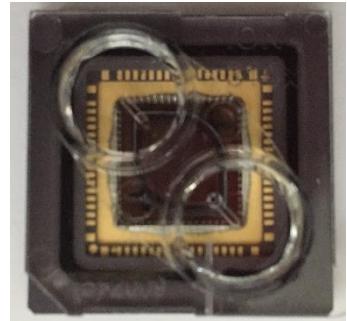


Illumina NovaSeq

454



Ion Torrent PGM





6. How accurate is Q20?

1 error in 100; 99%

7. What technology features nanoballs

Complete Genomics (BGI)

8. What does SBS stand for

Sequencing by Synthesis

9. What does SMRT stand for

Single Molecule Real Time

10. Name 2 technologies that use sequencing by ligation

Solid and Complete Genomics

The future



STRATOS
genomics inc.



QuantuMD ζ

Nanostring Hyb n Seq targeted sequencer

Details on James Hadfield's blog:

<http://enseqlopedia.com/2017/02/nanostring-hybseq/>





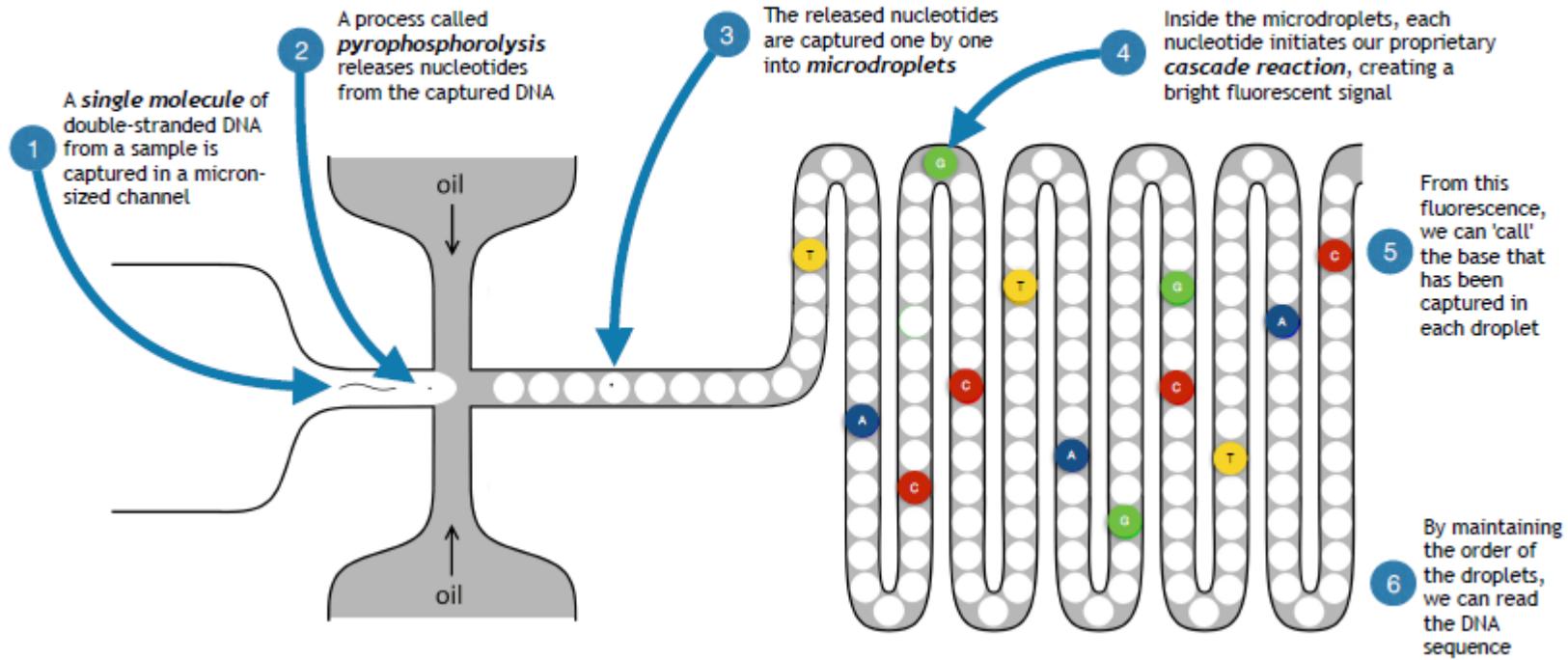
Base4

- » Based in Cambridge, UK.
- » Developing microfluidics based sequencer
- » Individual bases are cleaved from DNA one at a time, captured into droplets in oil
- » Presence of particular base in droplet causes characteristic fluorescence. Order of droplets = order of bases

Base4



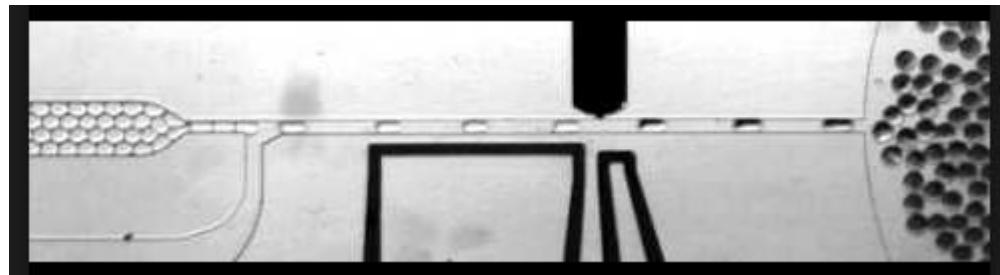
Microdroplet Sequencing Overview



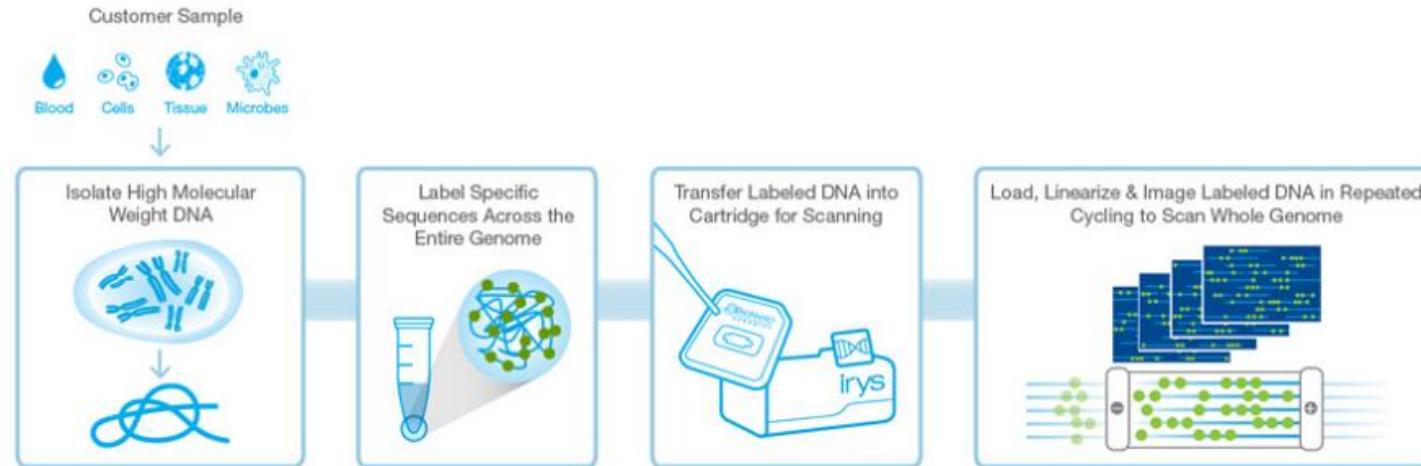


GnuBio

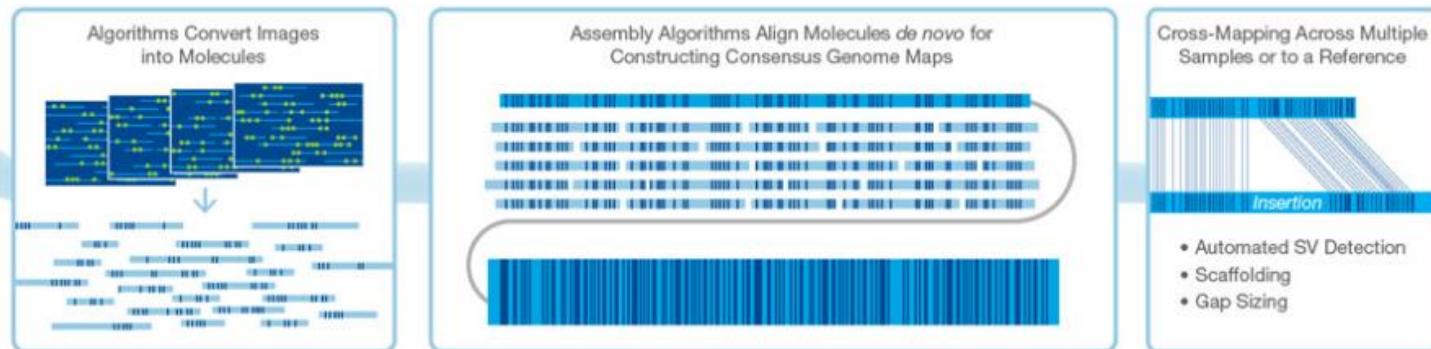
- » Acquired by BioRad
- » Sequencing by hybridisation in microfluidic oil droplets
- » Extremely accurate
- » \$200 for ~100 amplicons of 1kb



BioNanogenomics

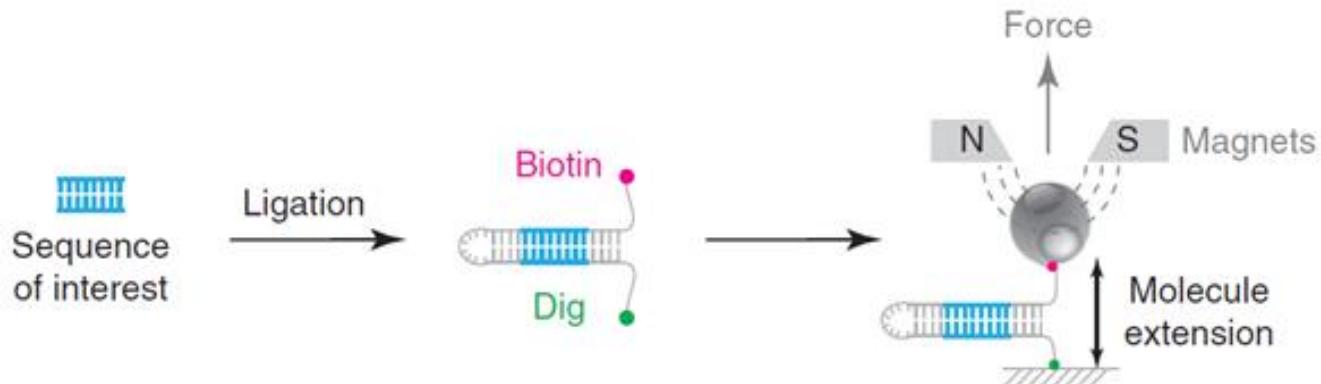


High-Throughput, High-Resolution Imaging Gives Contiguous Reads up to Mb Length



PicoSeq

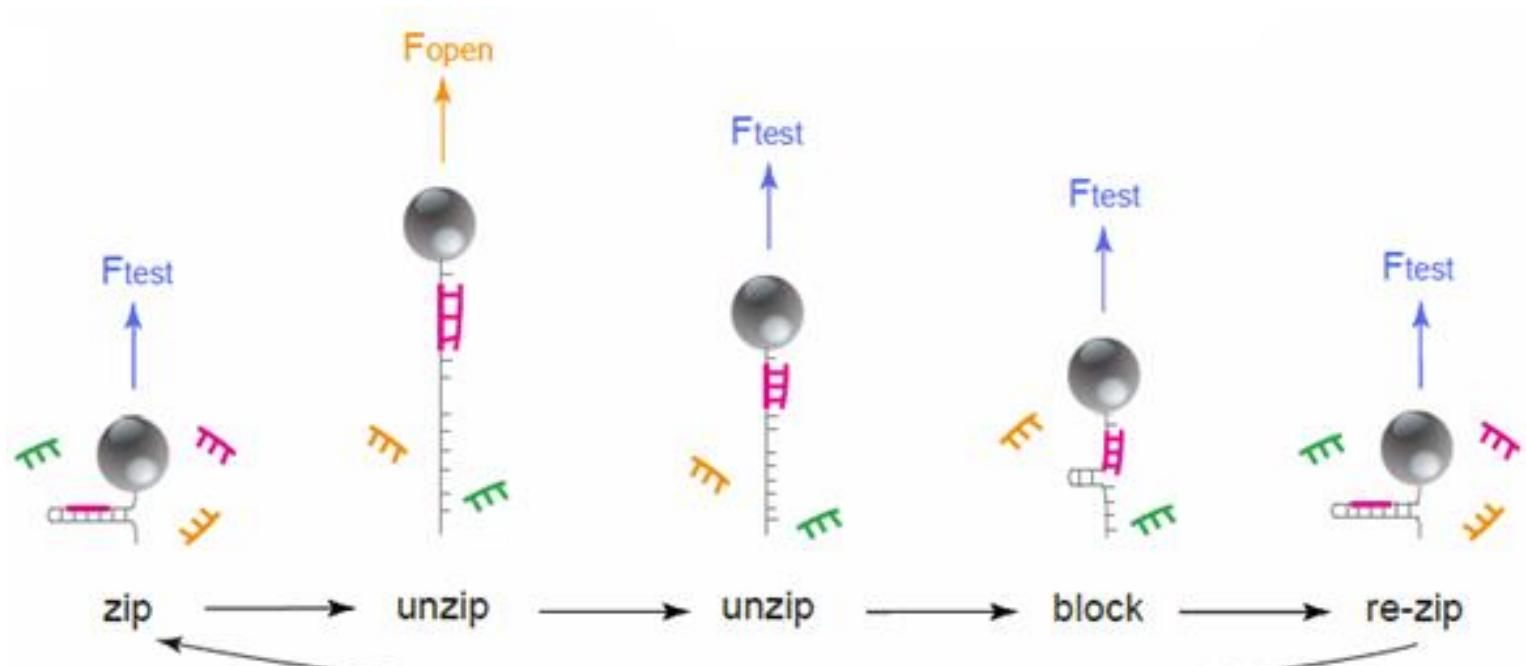
Depixus



“Single-molecule mechanical identification and sequencing” Ding F., Manosas M., Spiering M.M., Benkovic S.J., Bensimon D., Allemand J-F., & Croquette V. Nature Methods, published online: 11 March 2012 | doi:10.1038/nmeth.1925

Depixus

Parisian company using magnetic tweezers and SPR to detect binding to surface bound DNA

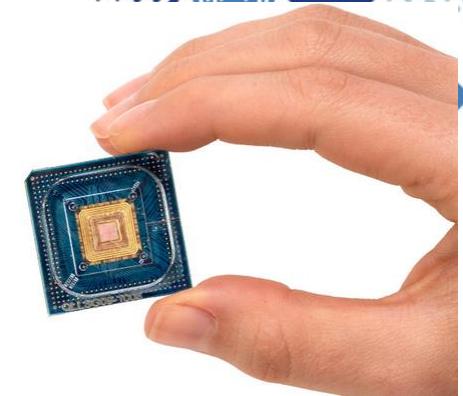


QuantMDx



- » UK company based in Newcastle. Currently sell microfluidics genotyping device (mPOC). Disease specific cartridges. Uses probes on nanowires. Hybridisation to target gives signal in < 5mins.
- » SBS. Base incorporation changes charge of DNA fragments. This can be detected with nanowires

Genapsys



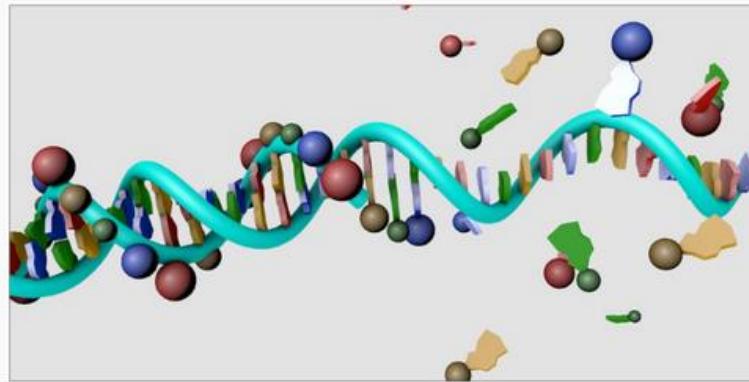
- » Small benchtop sequencer. Works in similar way to Ion Torrent
- » Eventually may be 10x cheaper than Illumina
- » 1kb reads



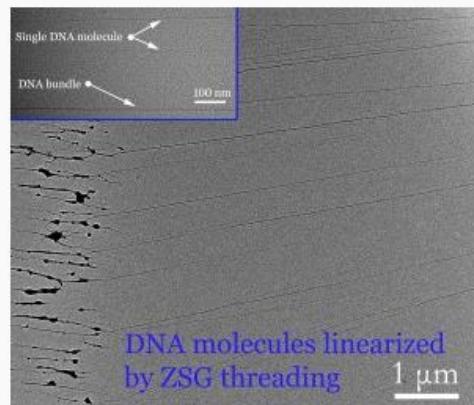


ZS Genetics

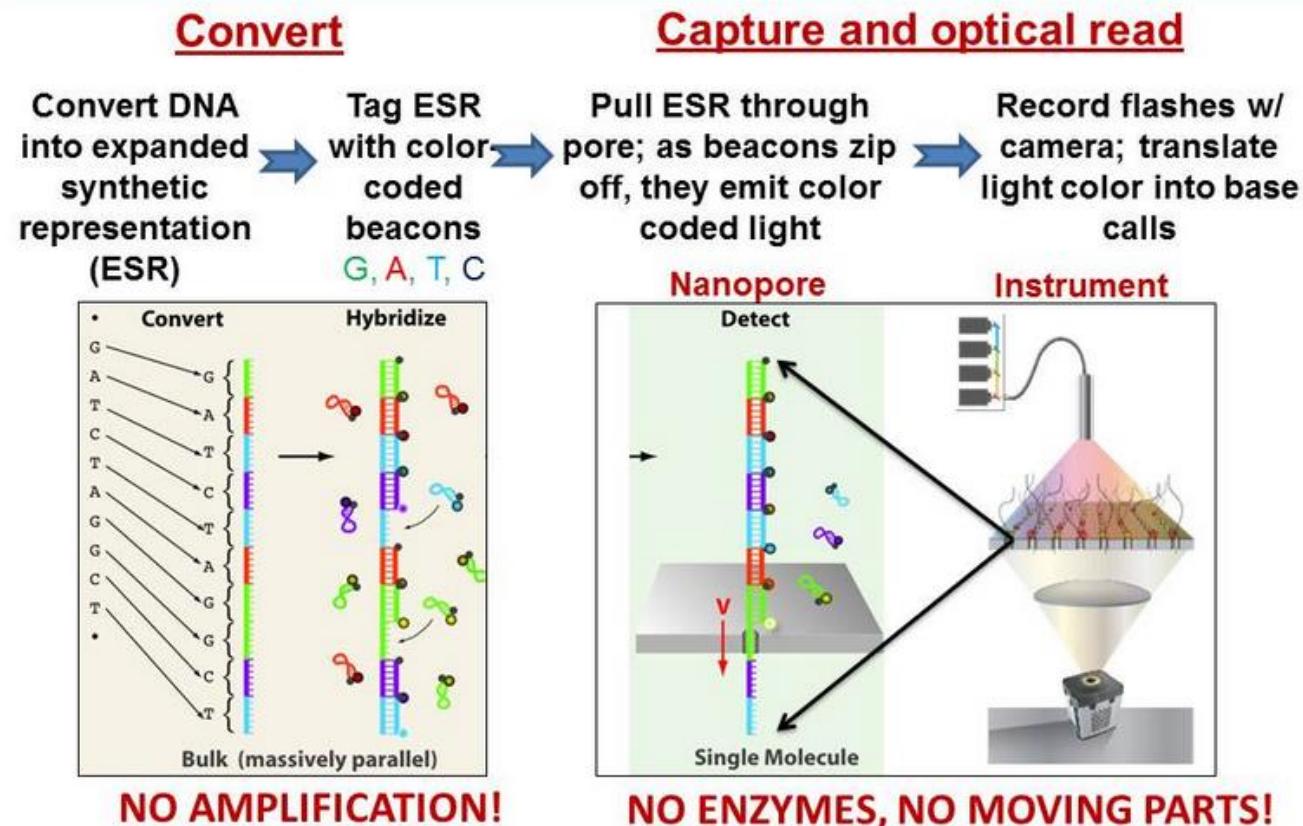
The single-stranded DNA is made into double-stranded DNA using heavy-atoms/modified dNTPs.



Each modified dNTP-type contains a unique label that is separately identifiable in the electron microscope-generated image, for example by "large" black dots, smaller black dots and large grey dots.



Noblegen



Automated, scalable workflow w/ small instrument

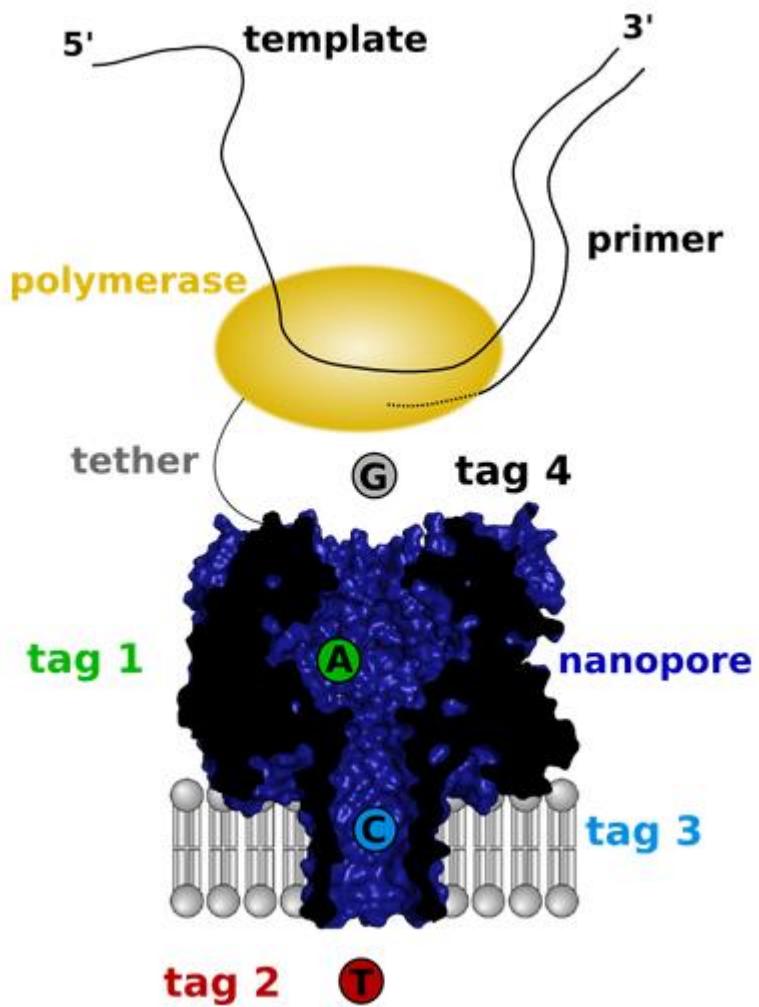
Noblegen
BIOSCIENCES

Conversion: 5h

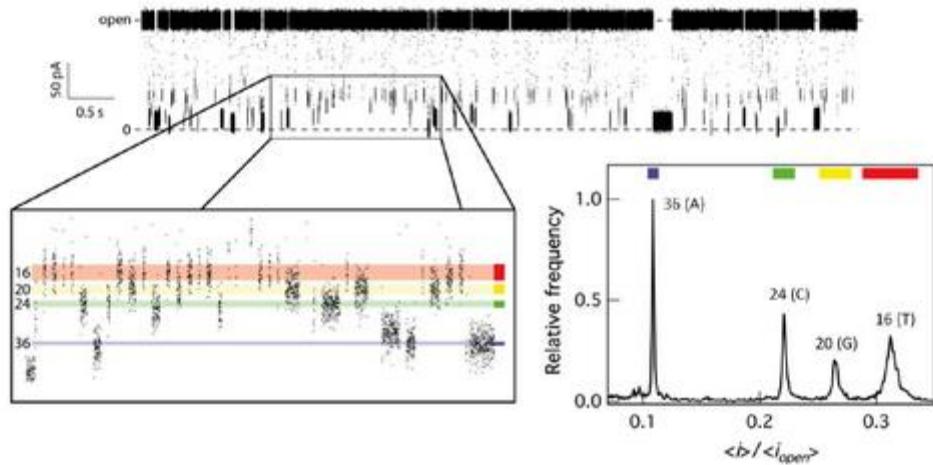
Read 96 genomes: 17h
1 genome < 30min

me
nger
insuute

Genia



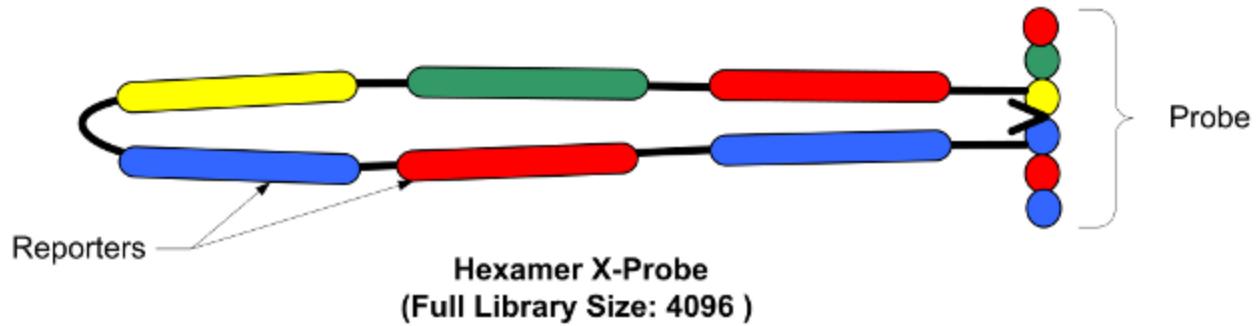
Genia



Potential for \$100 genome

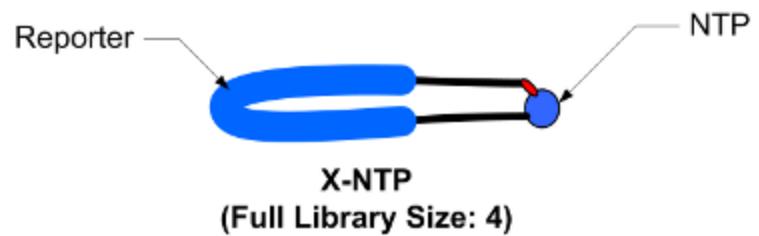


Stratos





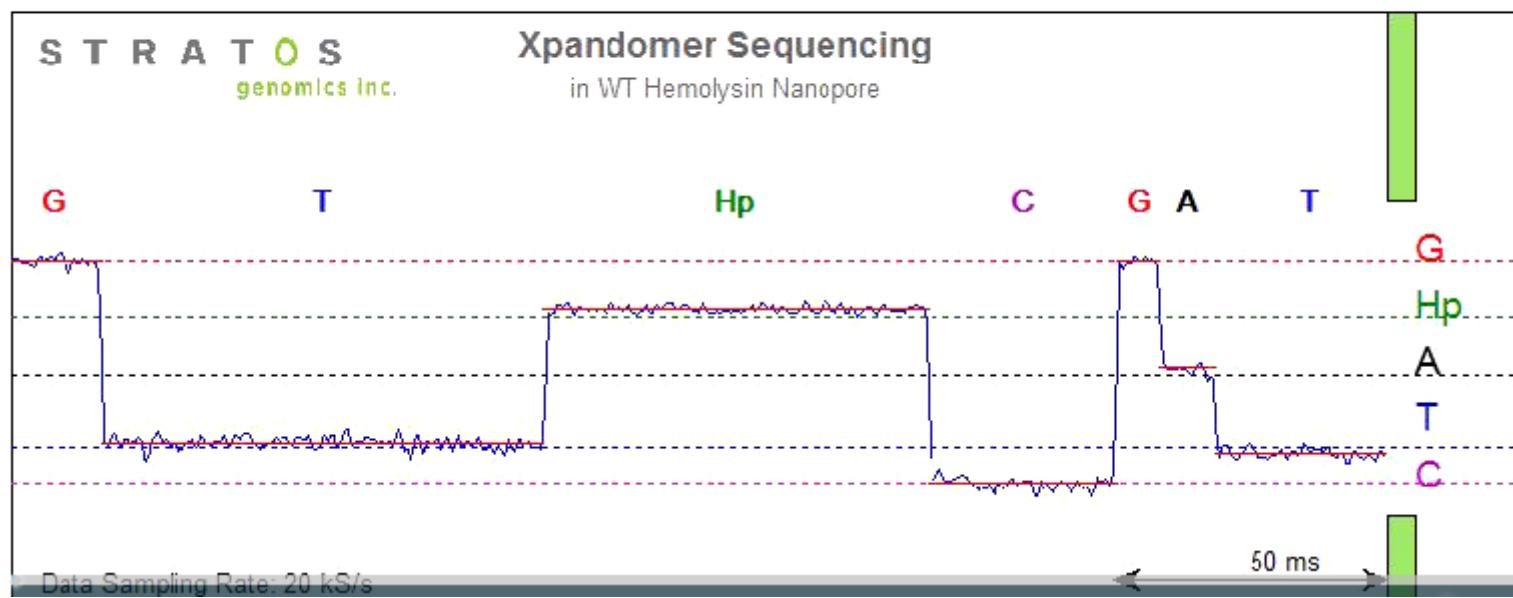
Stratos





Stratos

SBX SEQUENCING DEMONSTRATION



George Church - Harvard

» FISSEQ fluorescent in situ Sequencing:
Lee, Daugherty et al Science 2014

Cyclic in situ sequencing with reversible terminators.

Can sequence a few molecules per cell.

Takes several days.



Joakim Lundeberg - Spacial Transcriptomics (ST) At SciLifeLab

- » Histochemistry/RNA seq on single cell layer
- » Have spacially segregated barcoded oligo dT primers on glass slides



Joakim Lundeberg -Spacial Transcriptomics (ST) At SciLifeLab

- » Lyse, capture and do cDNA synthesis on surface. Cut off cDNA and make library. Sequence and sort according to barcode.
- » High density array. 135k unique Barcodes .These are 3' tag libraries. Sequence 50 K reads per feature
- » Each 13um square approximates a single cell. All cells on array treated equally

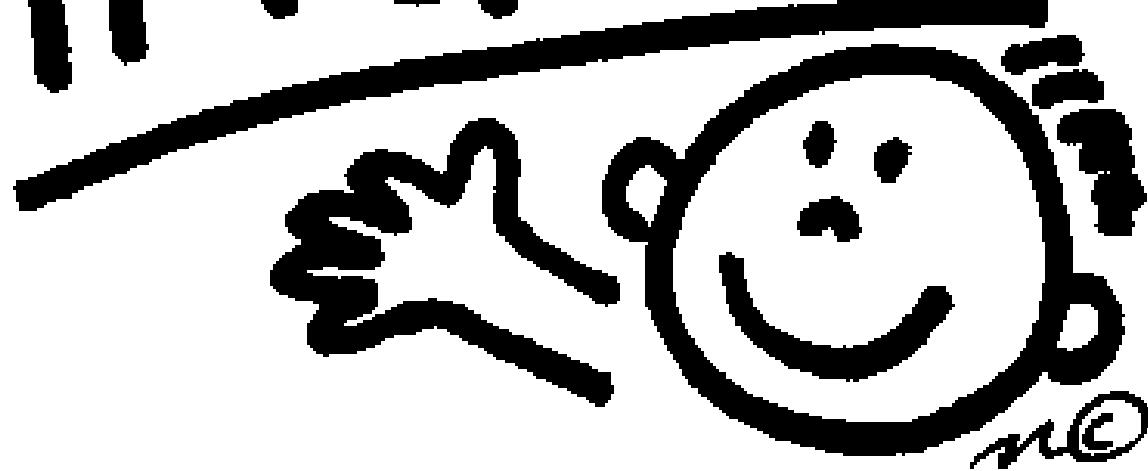


Joakim Lundeberg -Spacial Transcriptomics (ST) At SciLifeLab

- » Have software to display gene list for each cell
- » Costly and challenging to do deep Sequencing. 40k cells at 50k per run
- » Can couple barcode to gene specific probe to target capture. Also thinking about how would sequence on slide.



Thanks!





Any Questions ?

mq1@sanger.ac.uk