



# Sequencing Technologies

Michael Quail  
[mq1@sanger.ac.uk](mailto:mq1@sanger.ac.uk)



# Sequencers



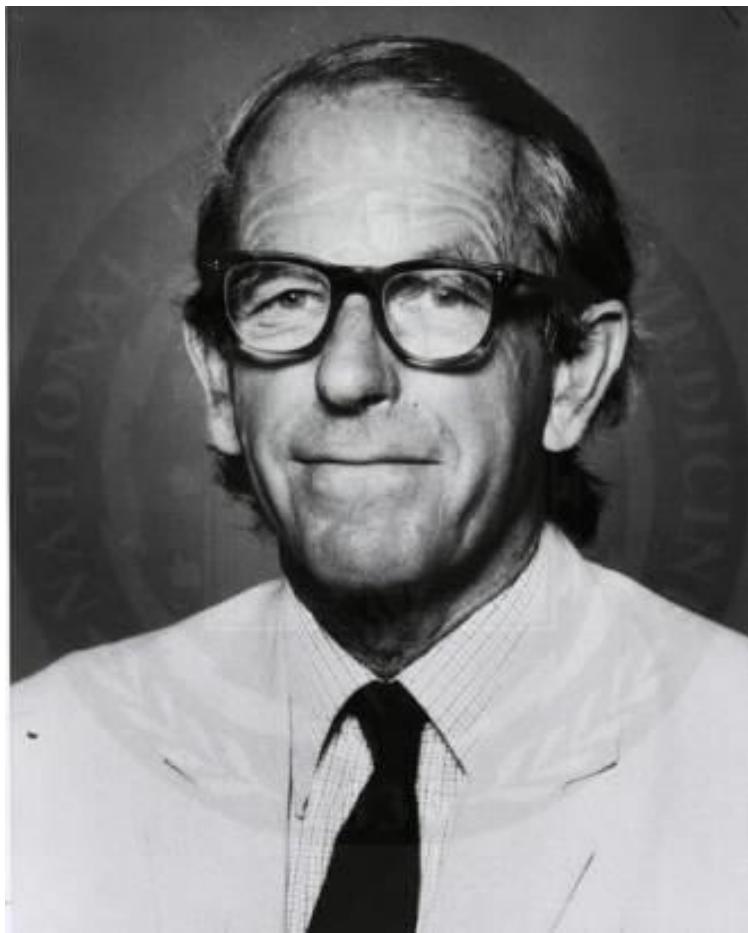


# Talk outline

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

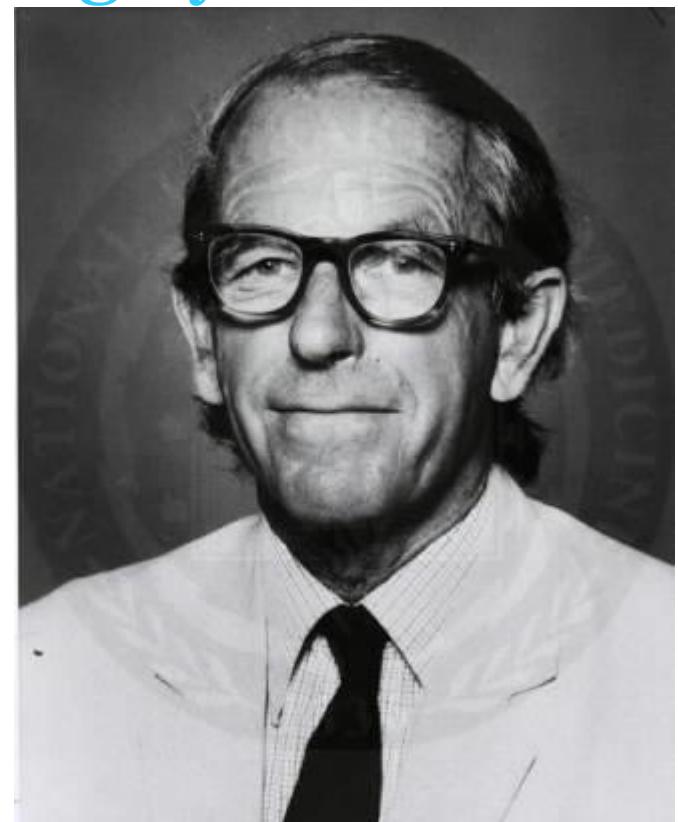


# Sanger Sequencing

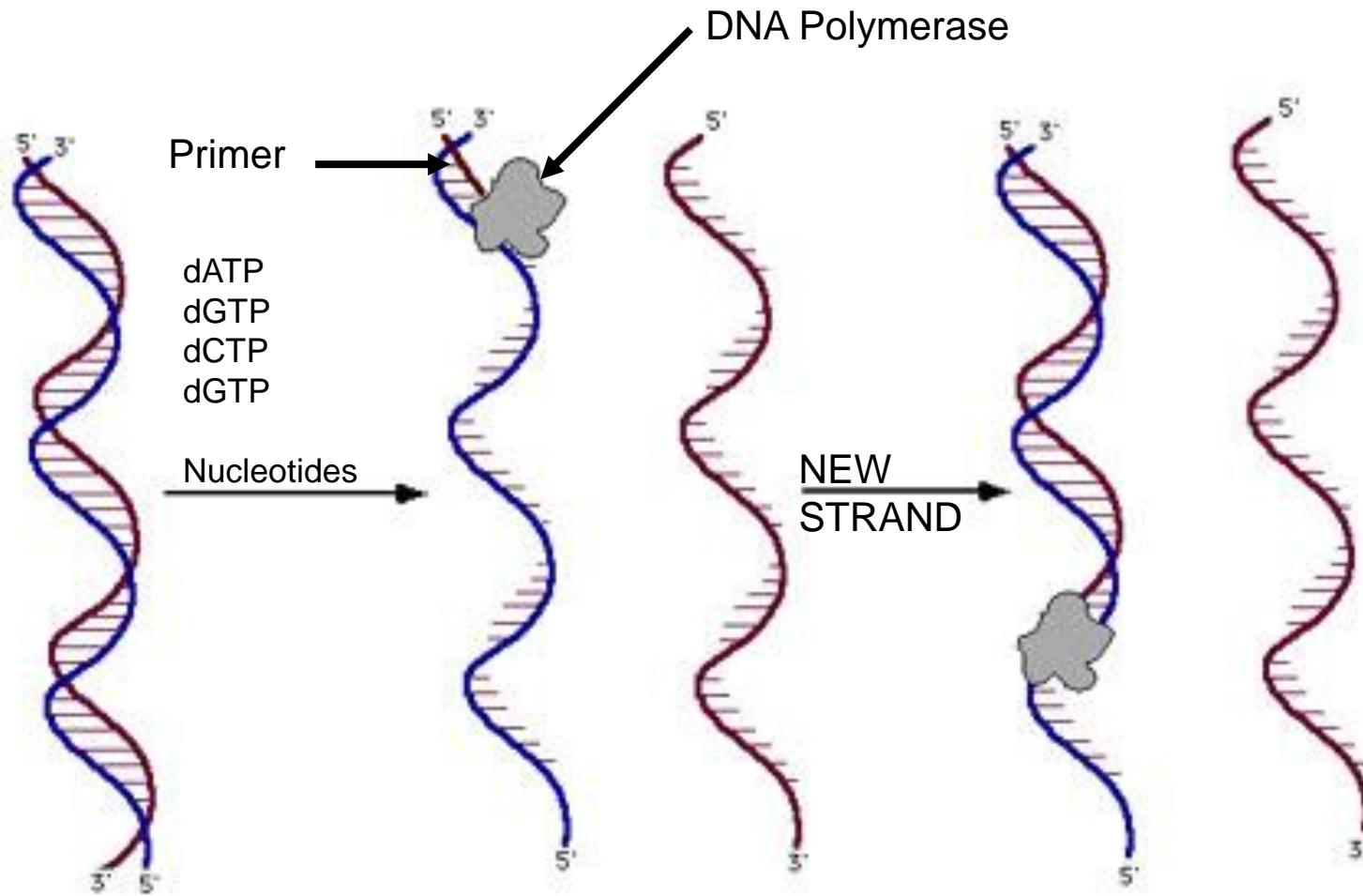


# Frederick Sanger

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



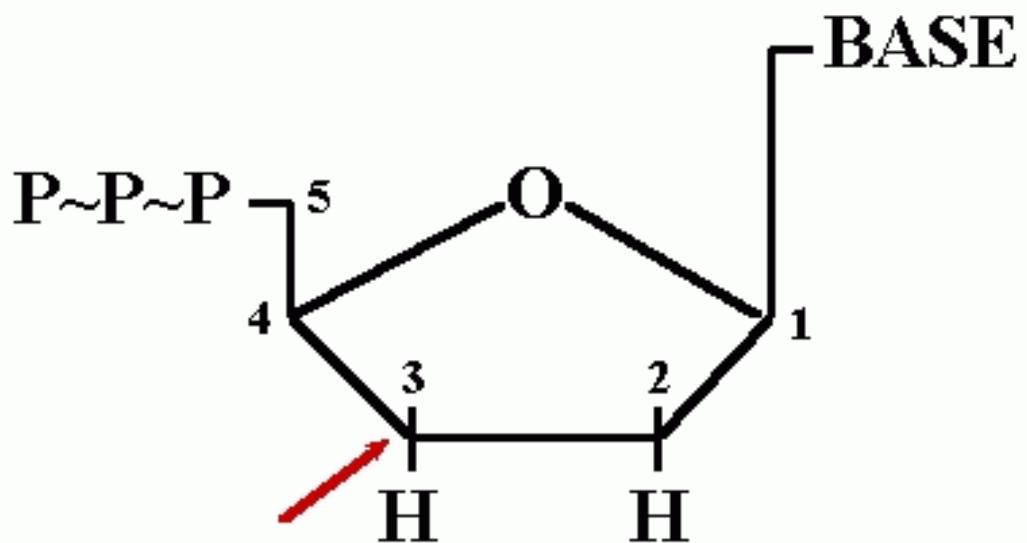
# Primer Extension



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•

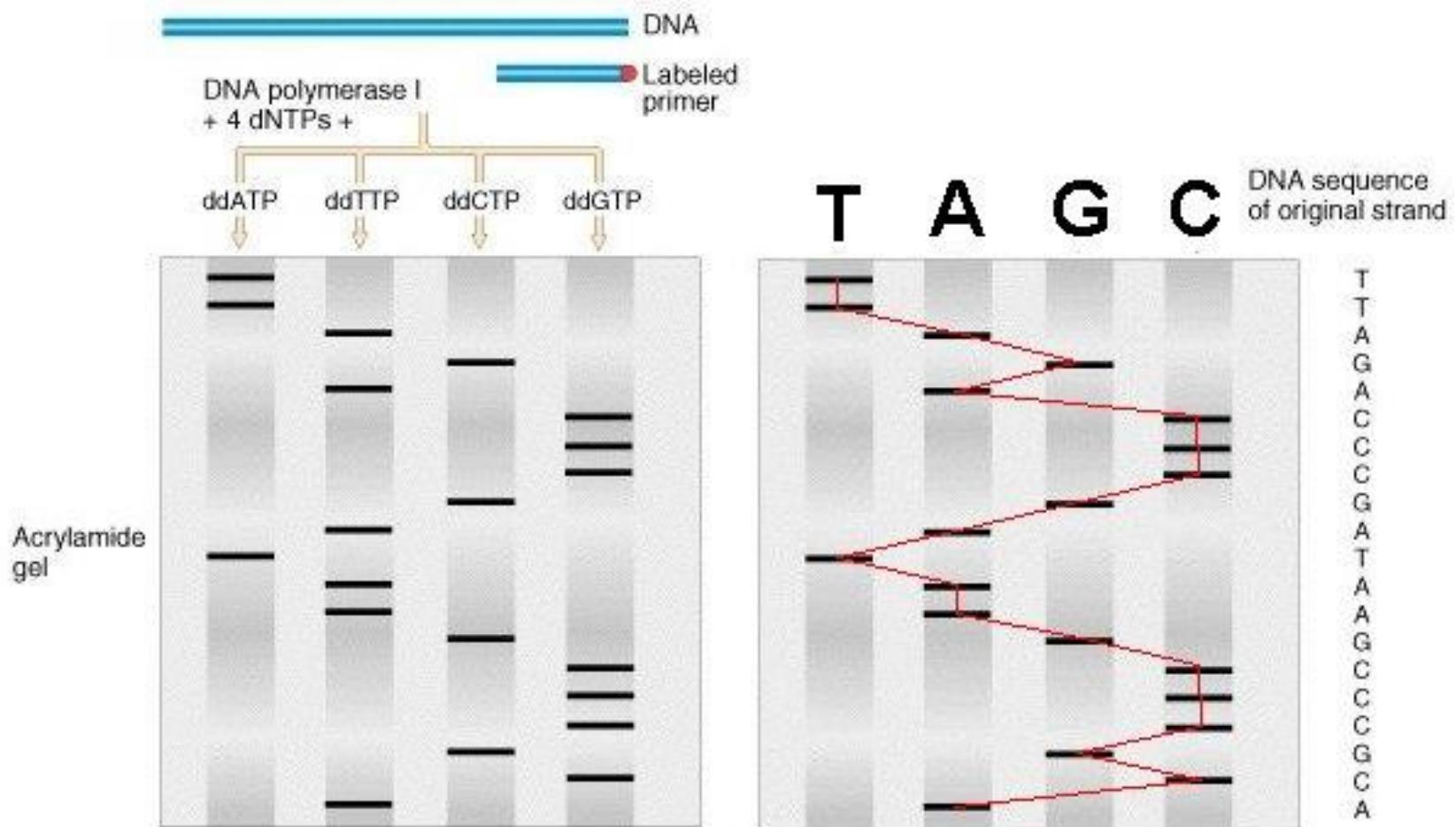
5' - TACGCGGTAAACGGT•

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

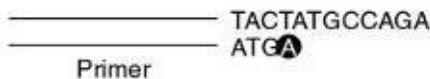
# Sequencing gel autorad



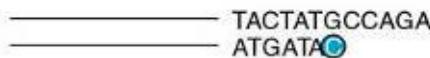
# Fluorescent Terminators

Primer extension reactions:

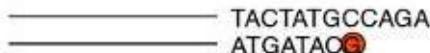
ddA reaction:



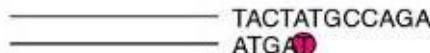
ddC reaction:



ddG reaction:

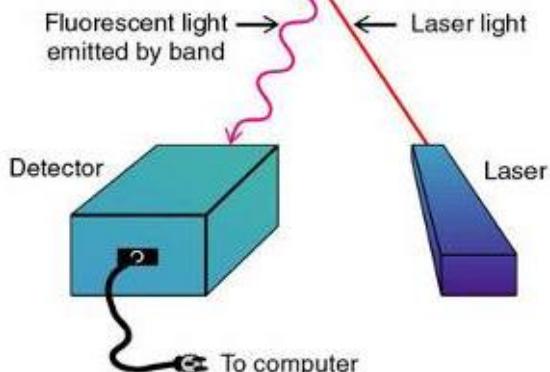


ddT reaction:

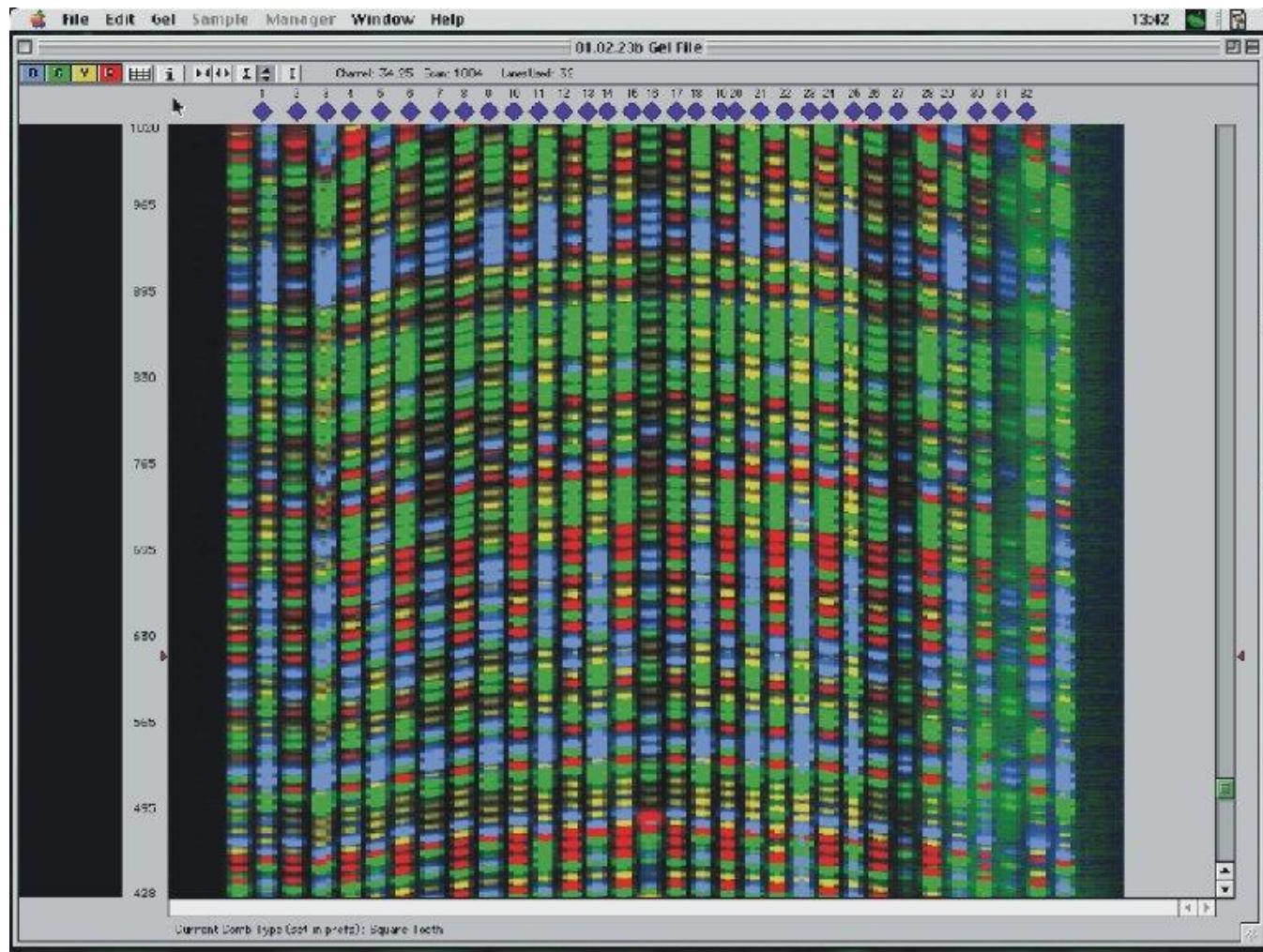


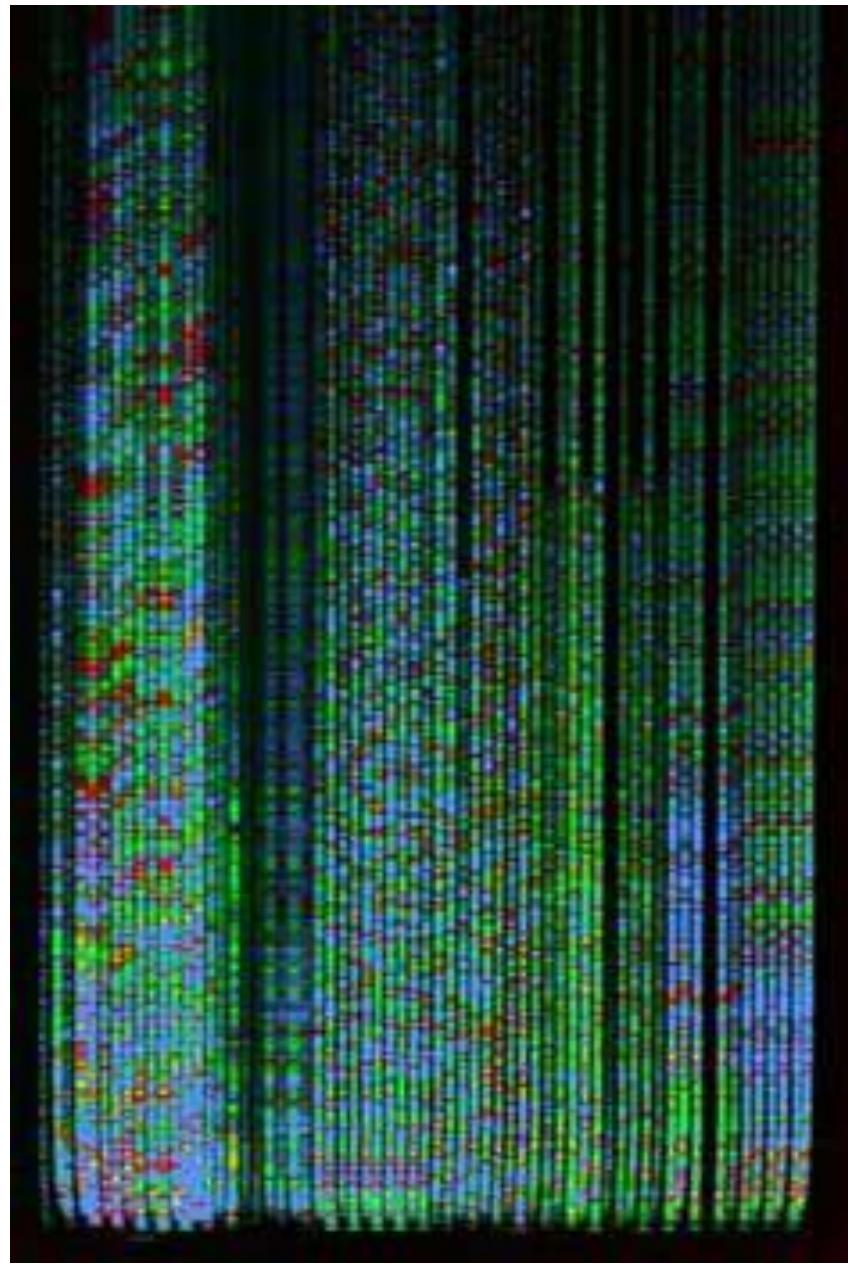
Electrophoresis:

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



# Fluorescent Gel Sequencing



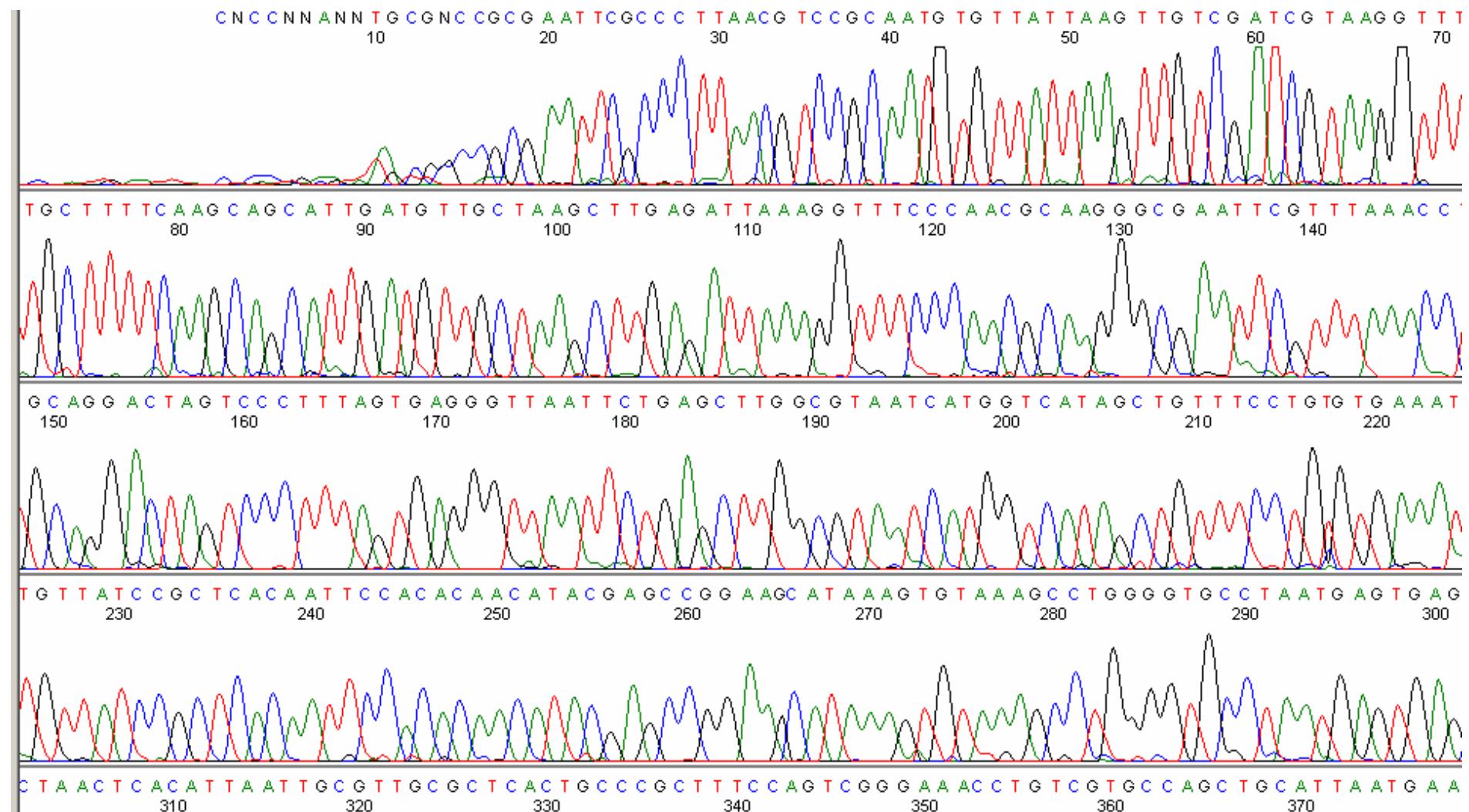




# ABI 3730



# DNA Sequence Files



# ABI Capillary 3730

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

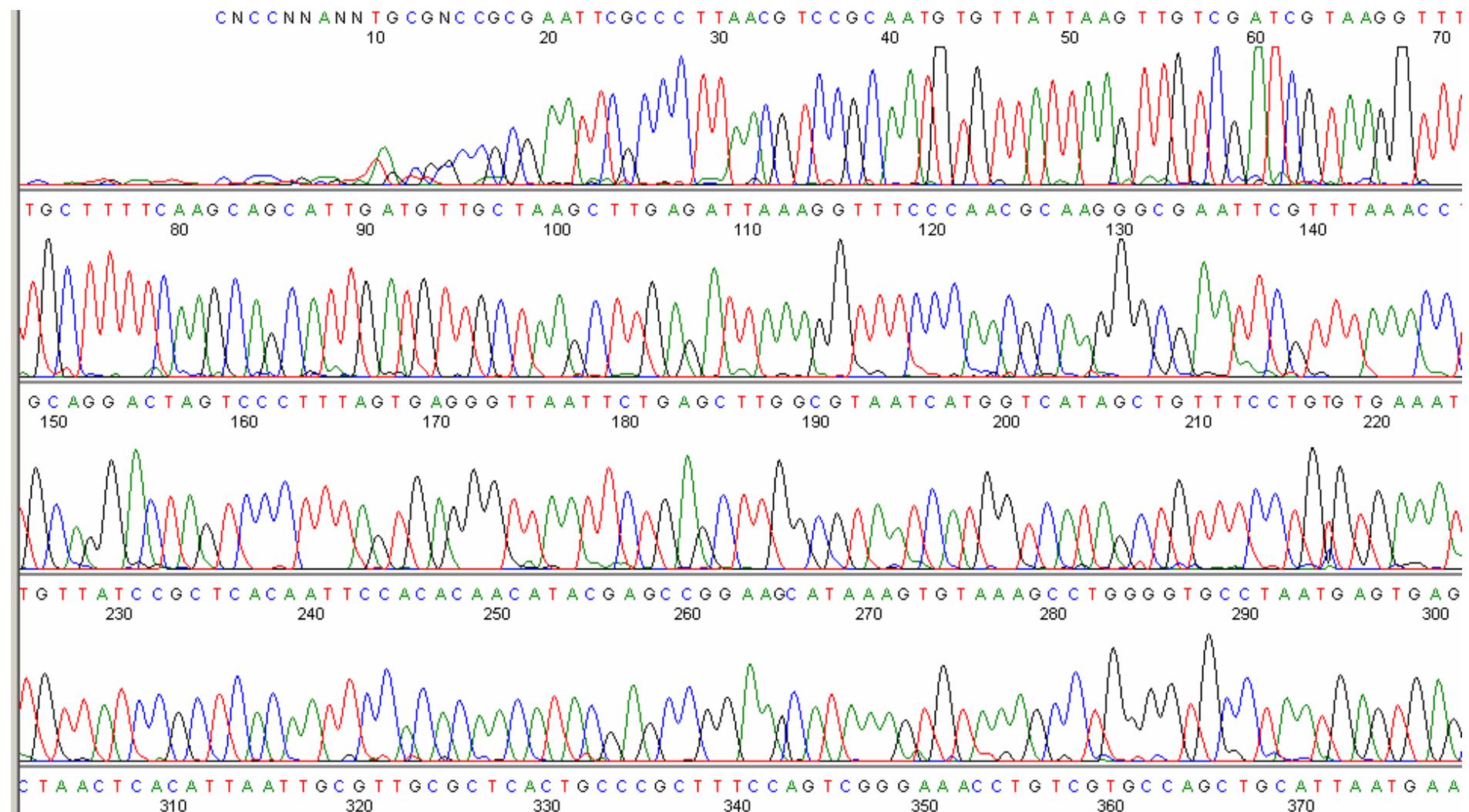




# Error

- » Sequence quality Q is reported on a log scale
- » Q<sub>10</sub> is 1 error in 10
- » Q<sub>20</sub> is 1 error in 100
- » Q<sub>30</sub> is 1 error in 1000
- » Q<sub>40</sub> is 1 error in 10000
- » Q<sub>50</sub> is 1 error in 100000

# DNA Sequence Files





# ABI Capillary 3730

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





# NGS

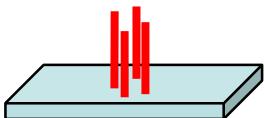
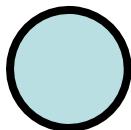
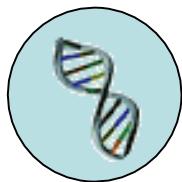
2005 - present



# 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

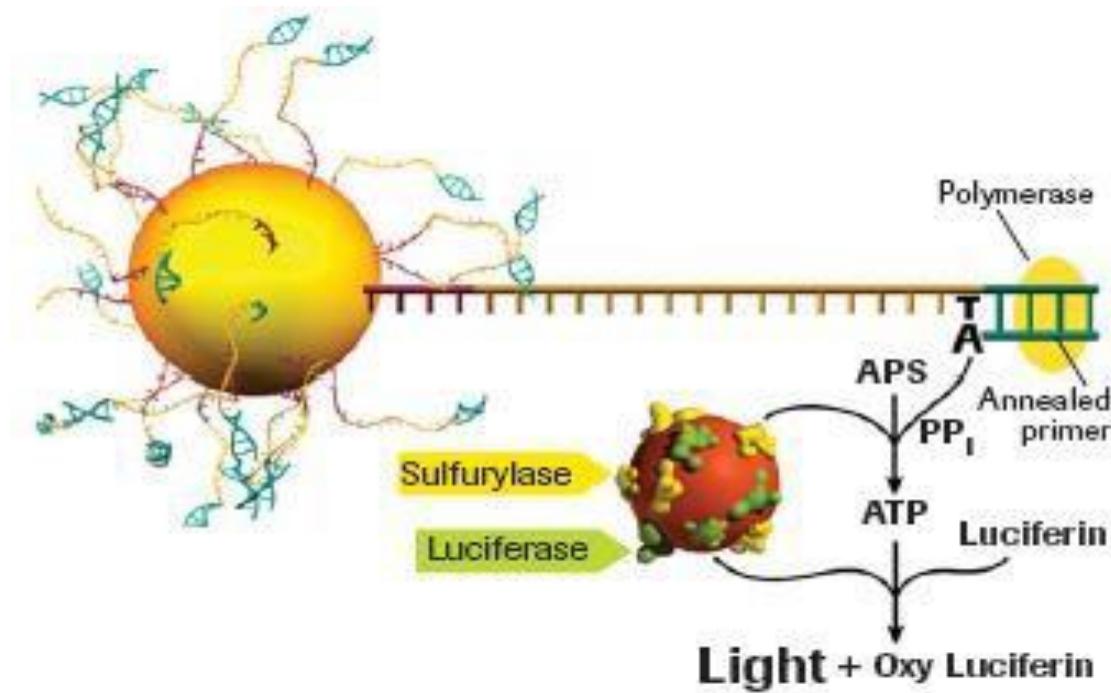


# 454

- » Started NGS (2005)
- » First massively parallel sequencer
- » Bought by Roche in 2007. Now discontinued
- » Based on pyrosequencing of bead-bound DNA in microwells
- » Fore-runner of Ion Torrent



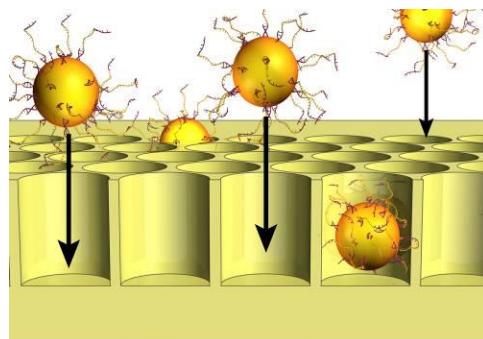
# Pyrosequencing



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

# Depositing DNA Beads into the PicoTiter™Plate

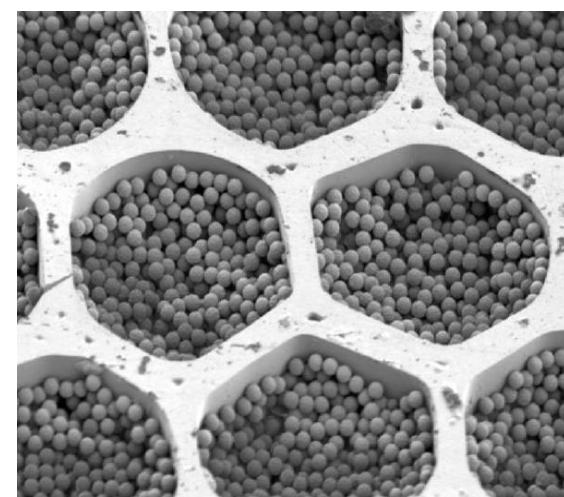
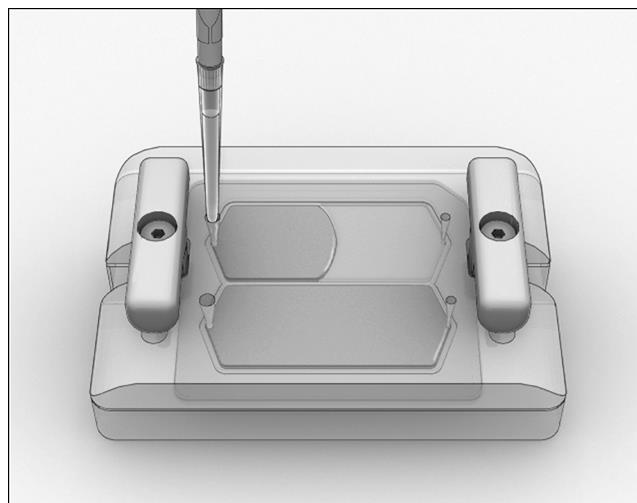
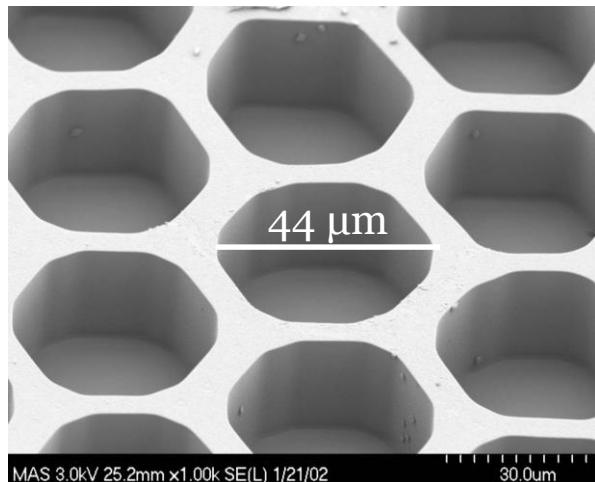
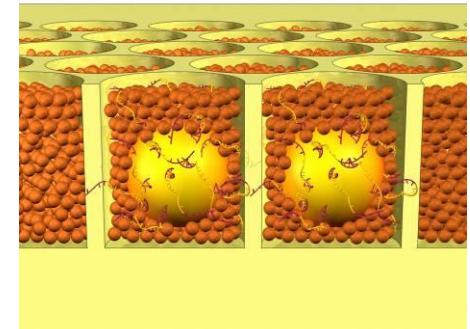
Load beads into  
PicoTiter™Plate



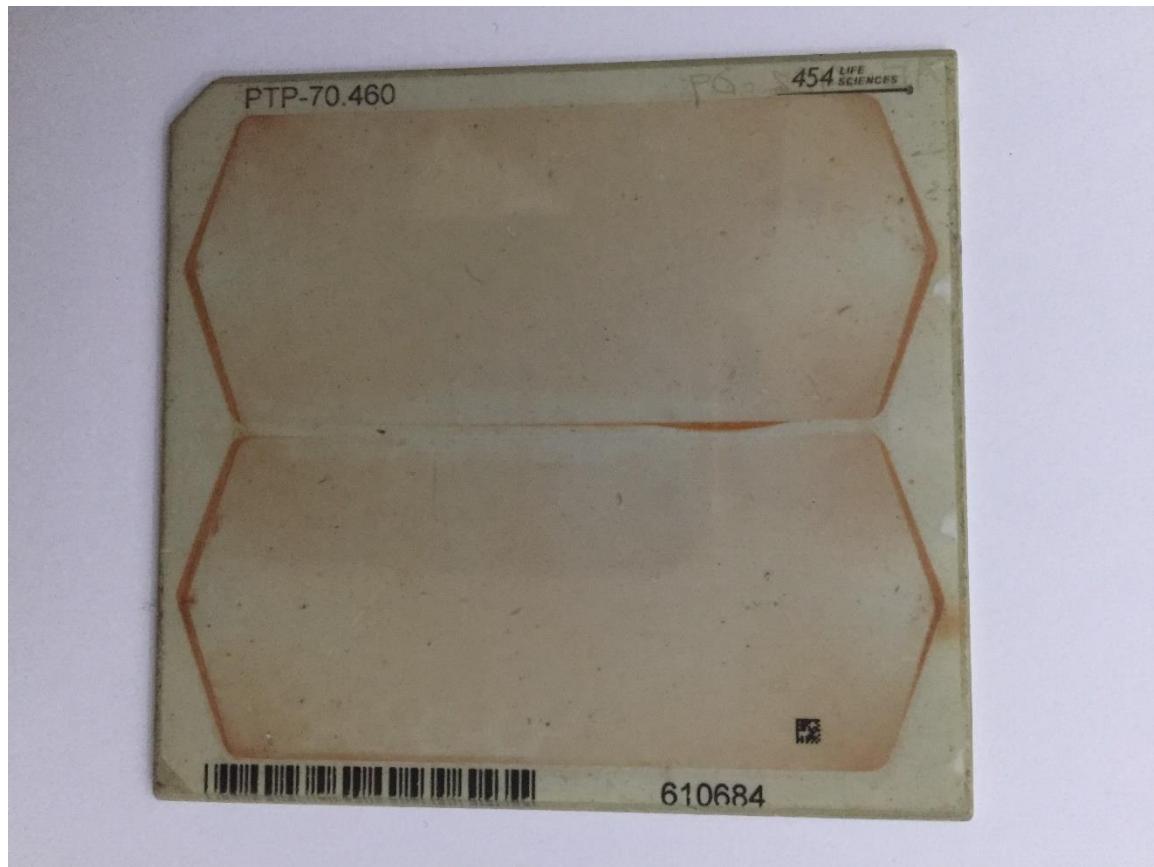
Load Enzyme Beads



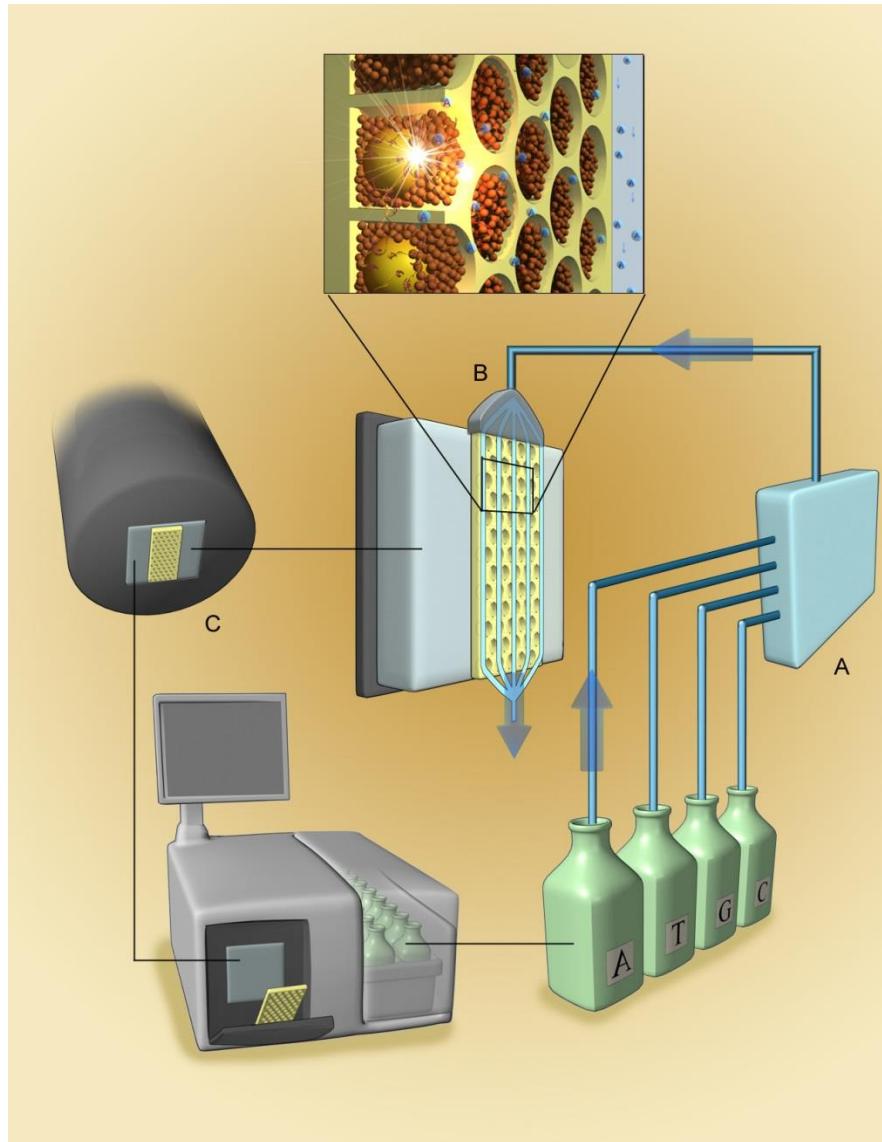
Centrifugation



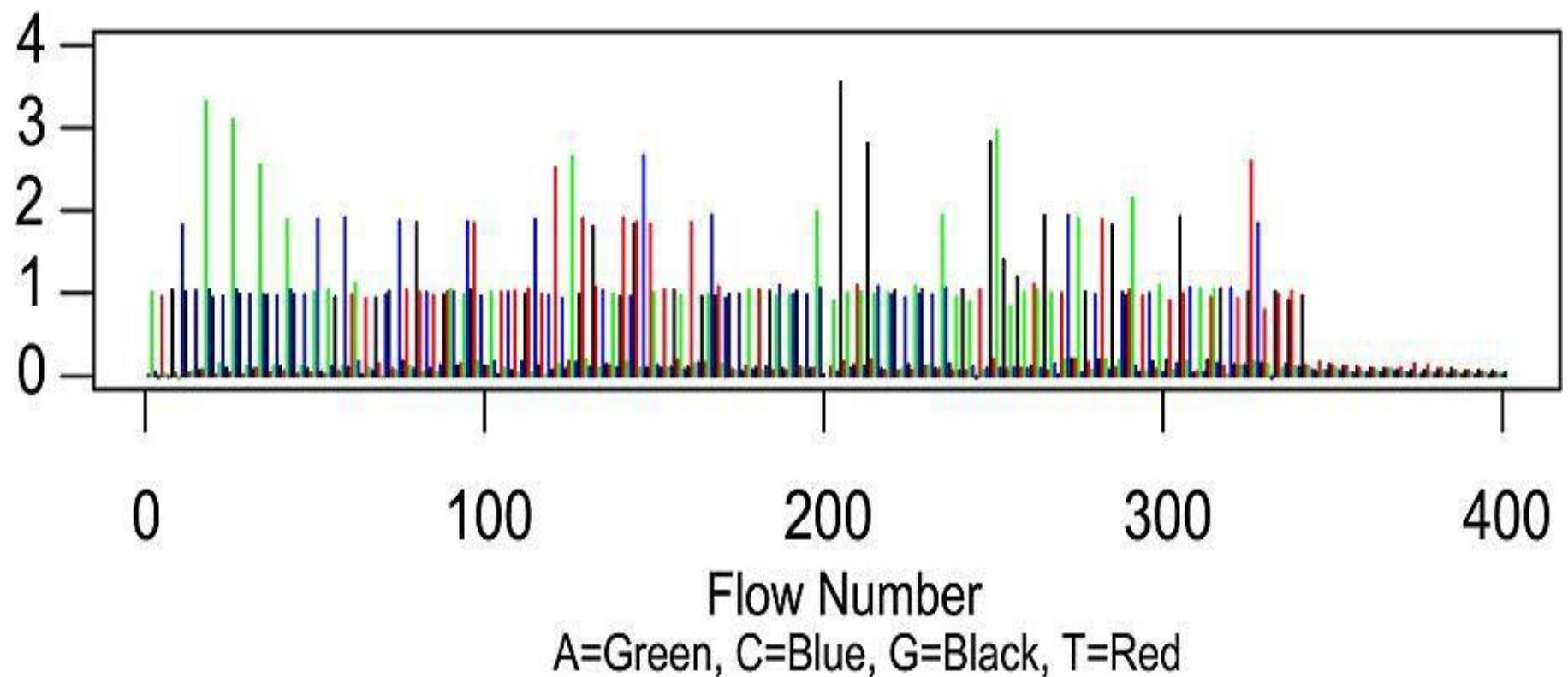
# 454 Sequencing Slide



# 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



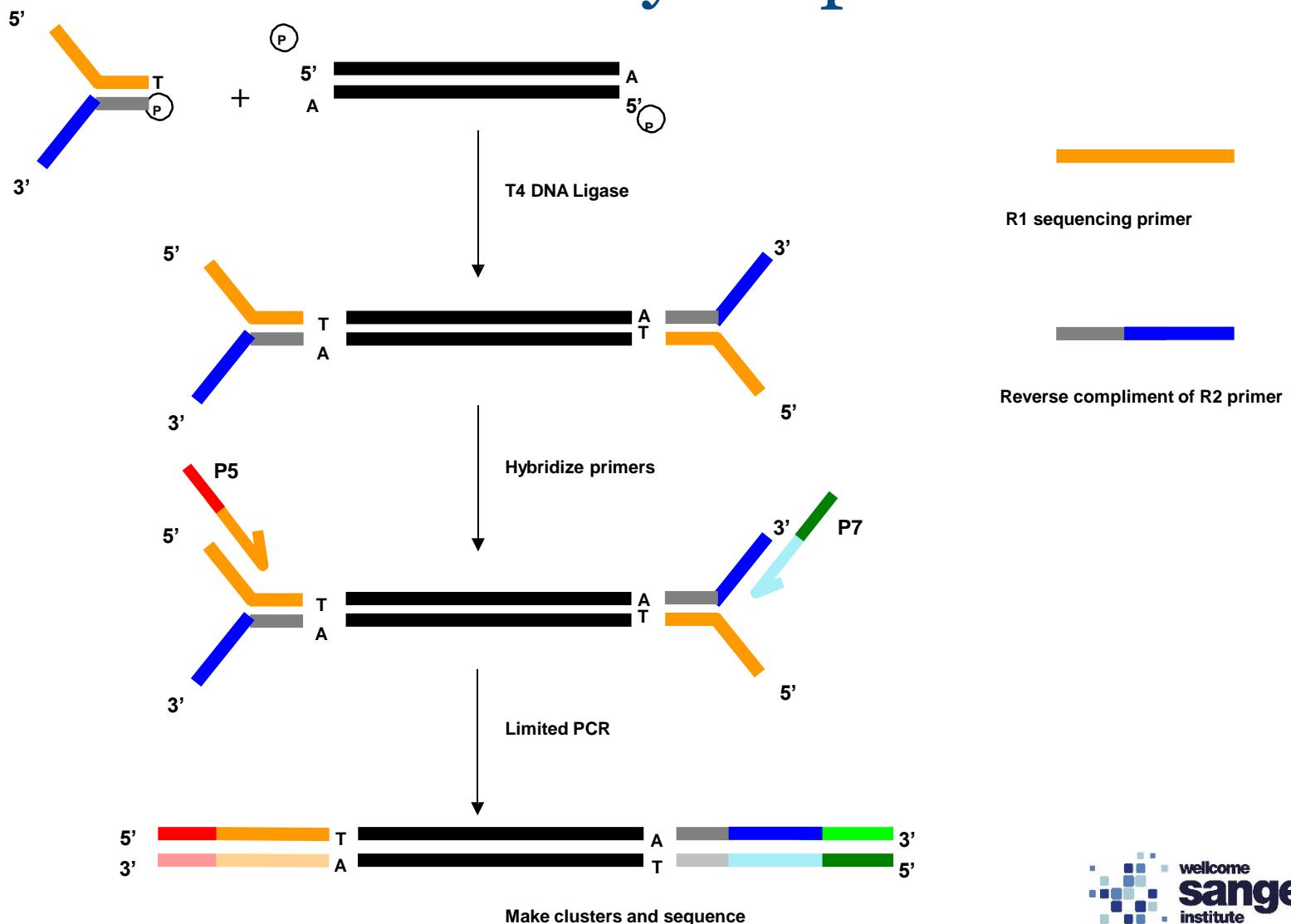
# Illumina



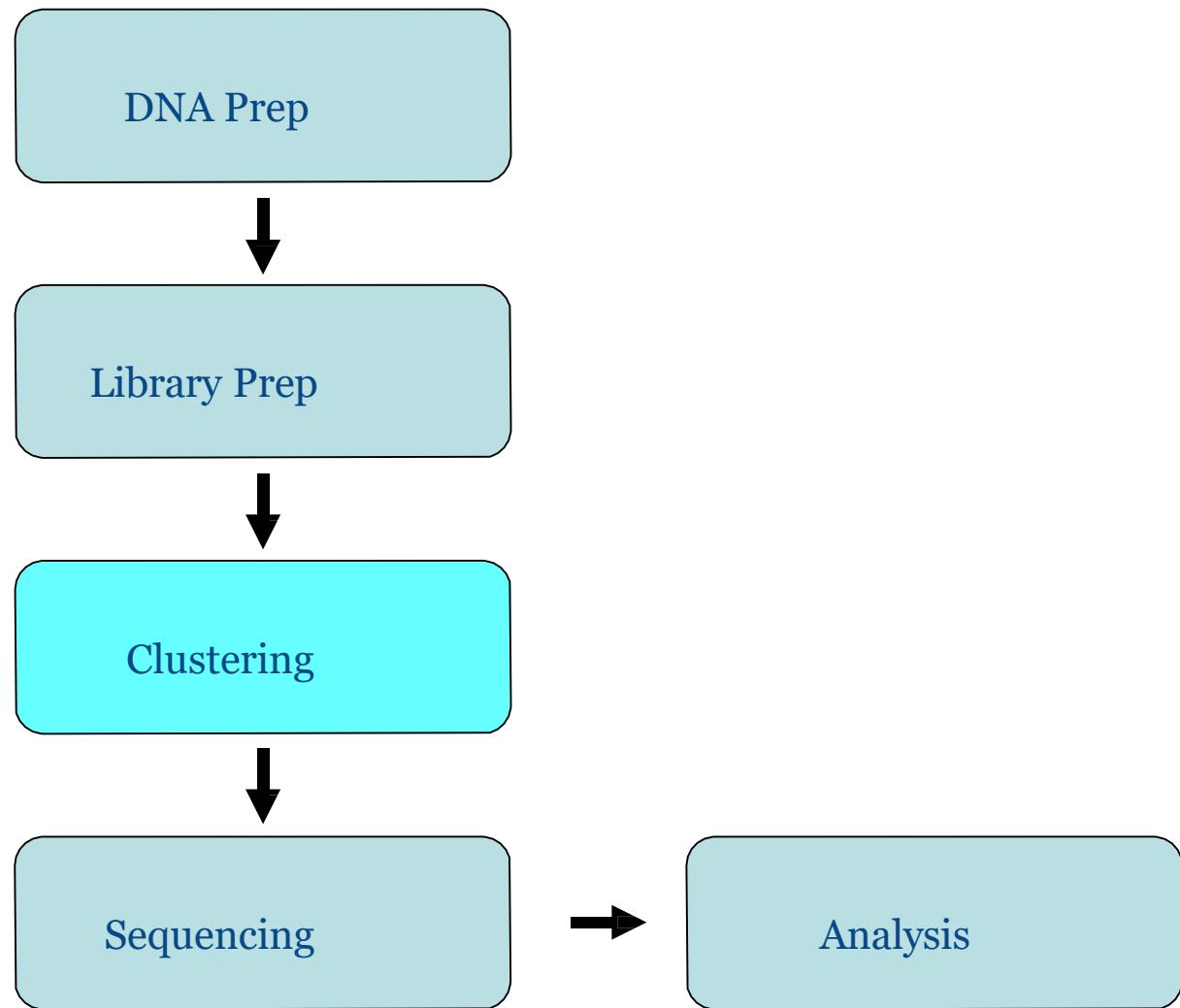
# Solexa

- » Launched their Genome Analyzer in 2006
- » Spinout from Cambridge University, set up at Gt. Chesterford in 2000
- » At launch GA gave 1Gb/run. Now upto 6Tb.
- » Acquired by Illumina in 2007
- » Short read sequencing
- » Accurate (0.2-0.4% error)
- » Market leader.
- » \$5-\$50/Gb

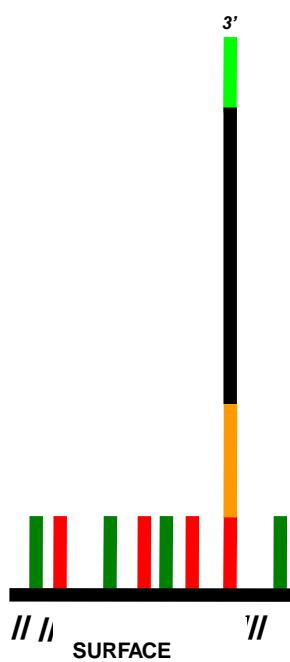
# Illumina Paired End Library Prep



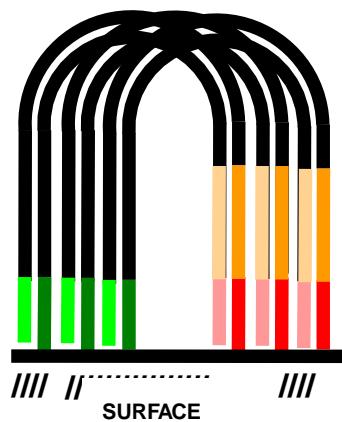
# Illumina workflow



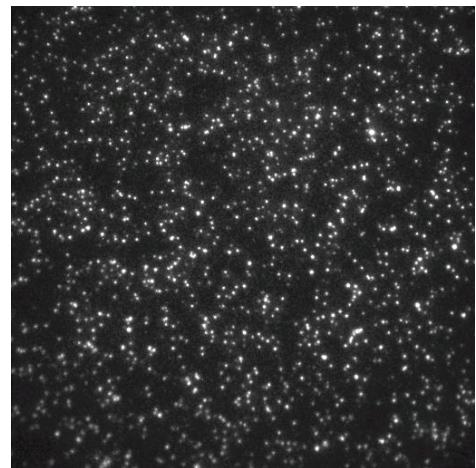
# Cluster Amplification



Single-molecule  
array



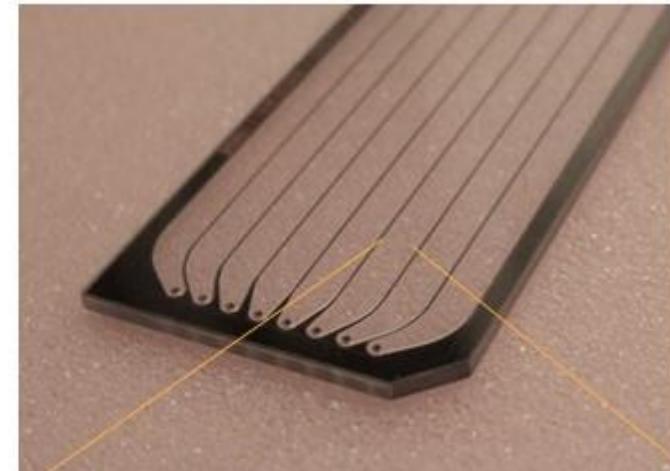
Cluster  
~1000  
molecules



1.5 Billion  
clusters on a  
single glass chip

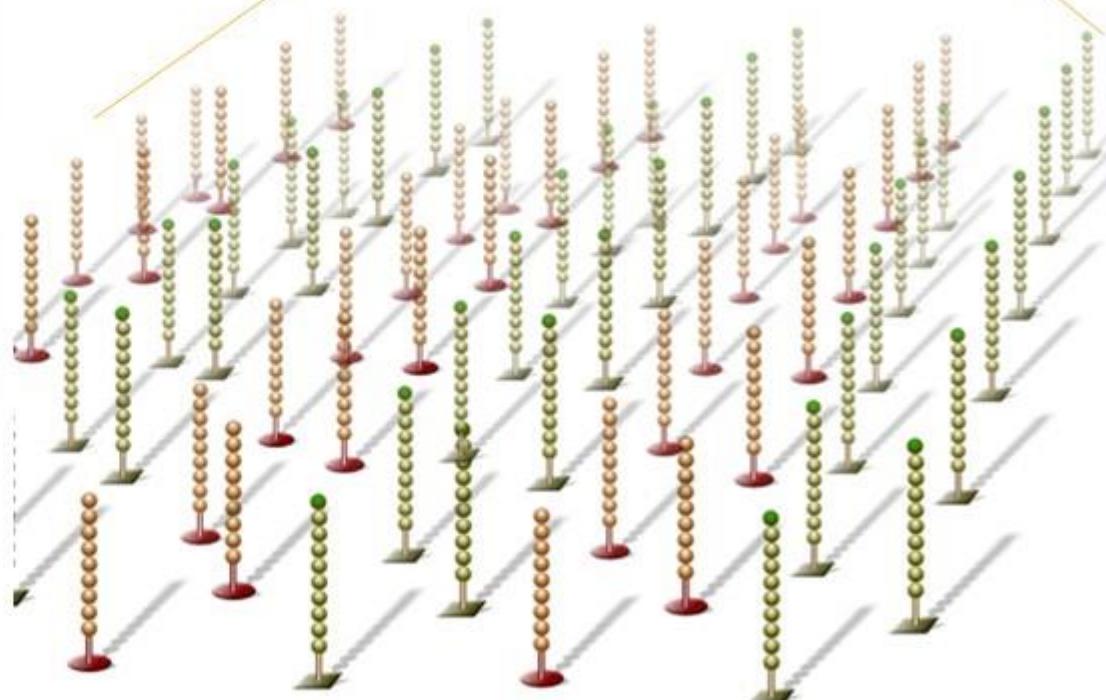
# What is a Flow Cell?

Cluster generation occurs on a flow cell



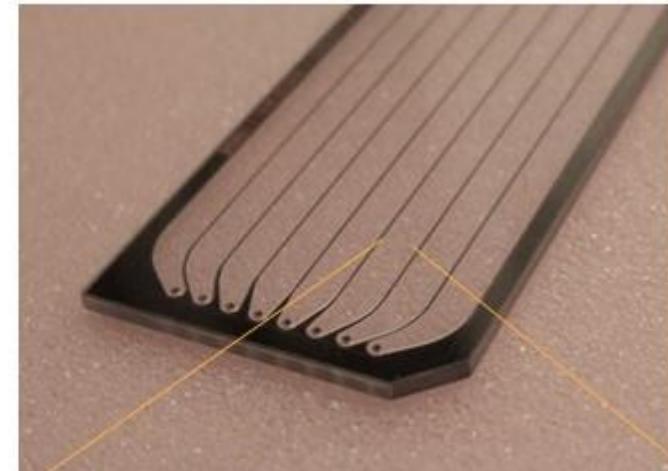
A flow cell is a thick glass slide with channels or lanes

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

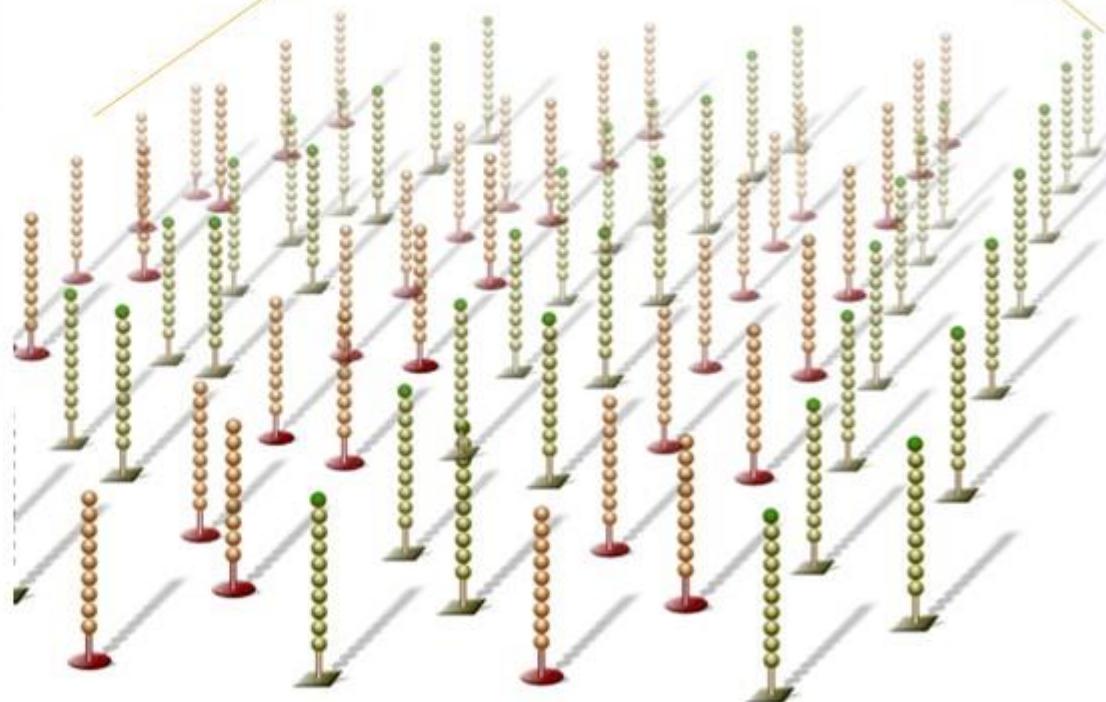


# What is a Flow Cell?

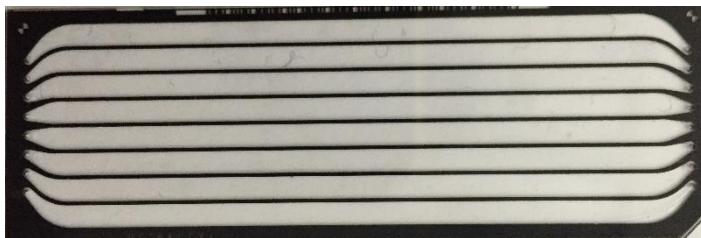
Cluster generation occurs on a flow cell

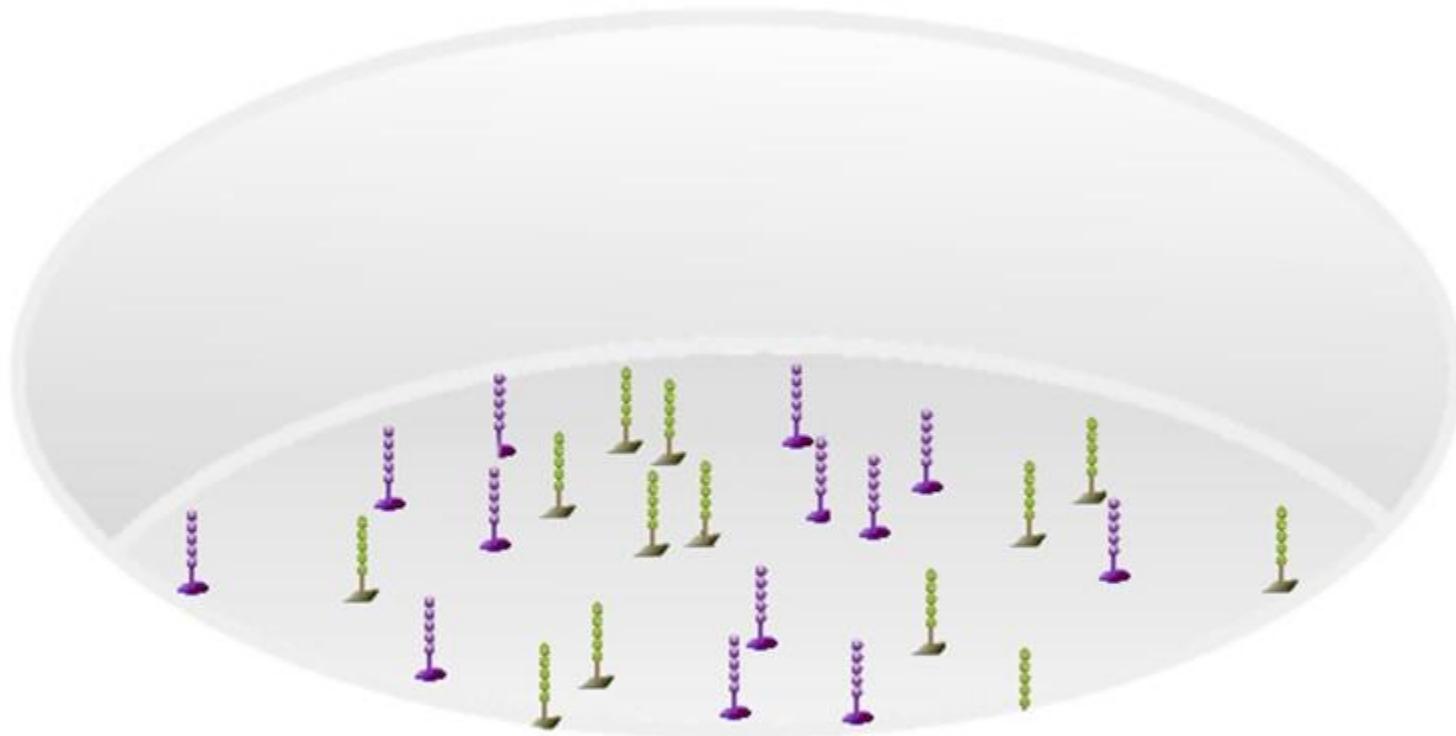


A flow cell is a thick glass slide with channels or lanes

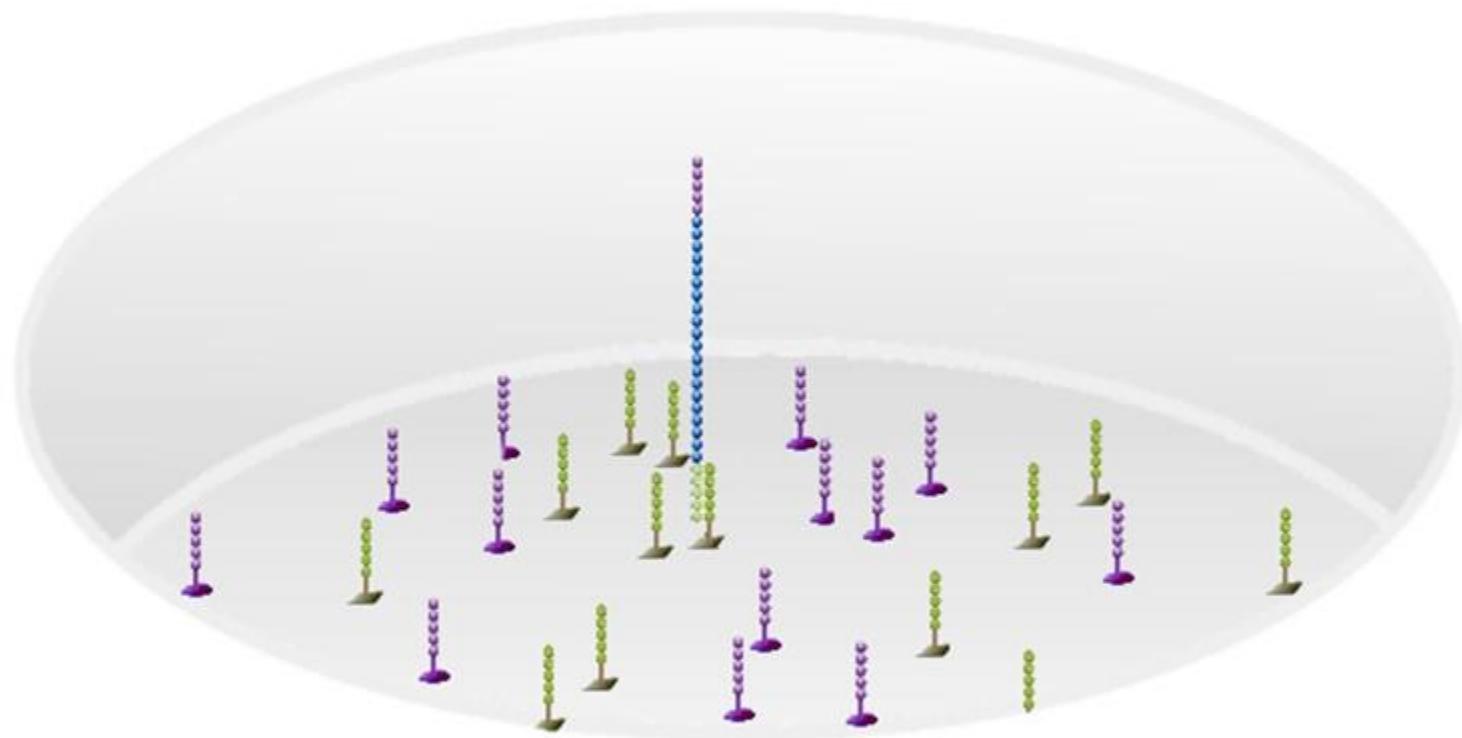


Each lane is randomly coated with a lawn of oligos that are complementary to library adapters



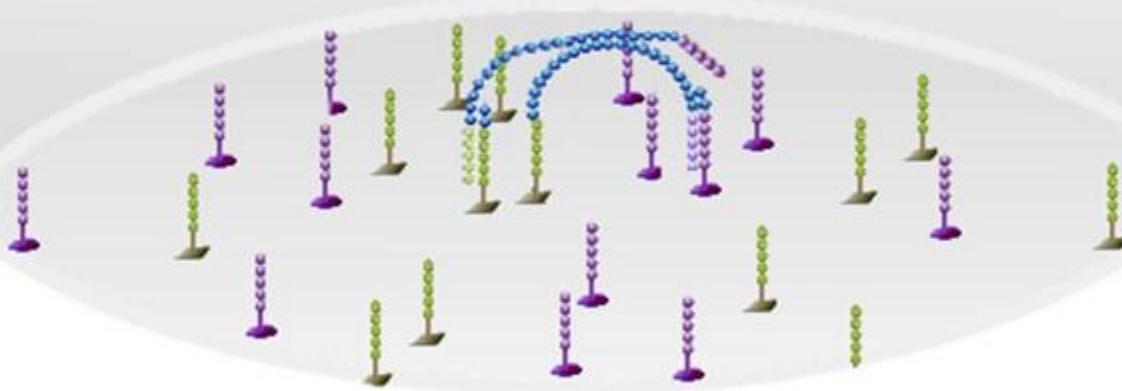


Oligos are not present on surface  
between wells to control cluster size

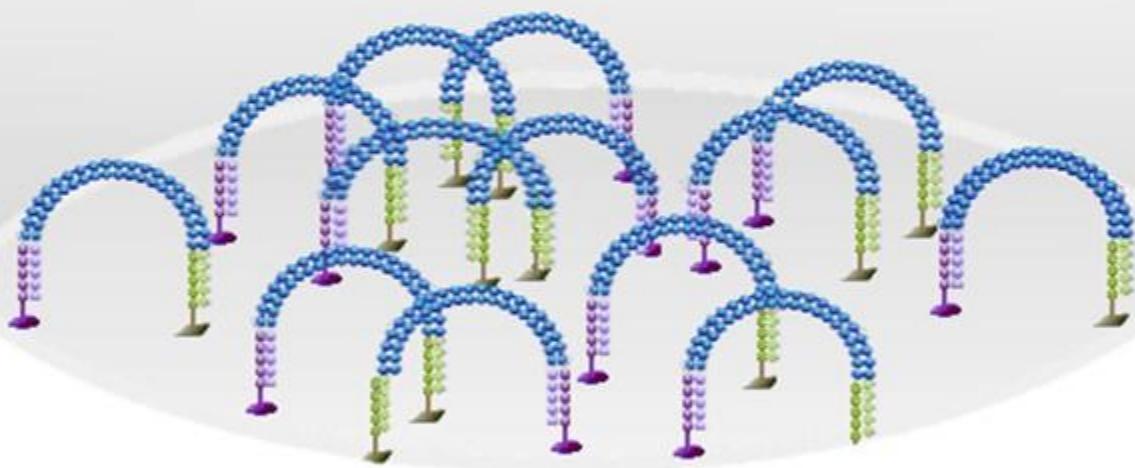


Single stranded DNA is introduced to the flow cell, which hybridizes to an oligo present in the microwell

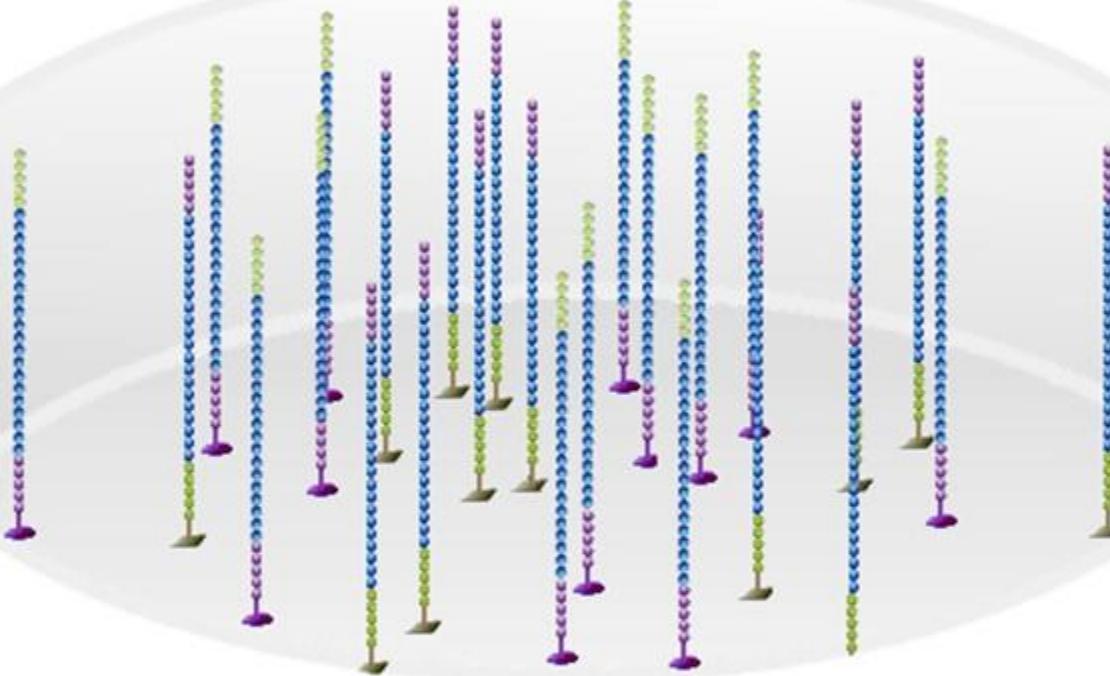
## ExAmp



Amplification occurs on both strands simultaneously

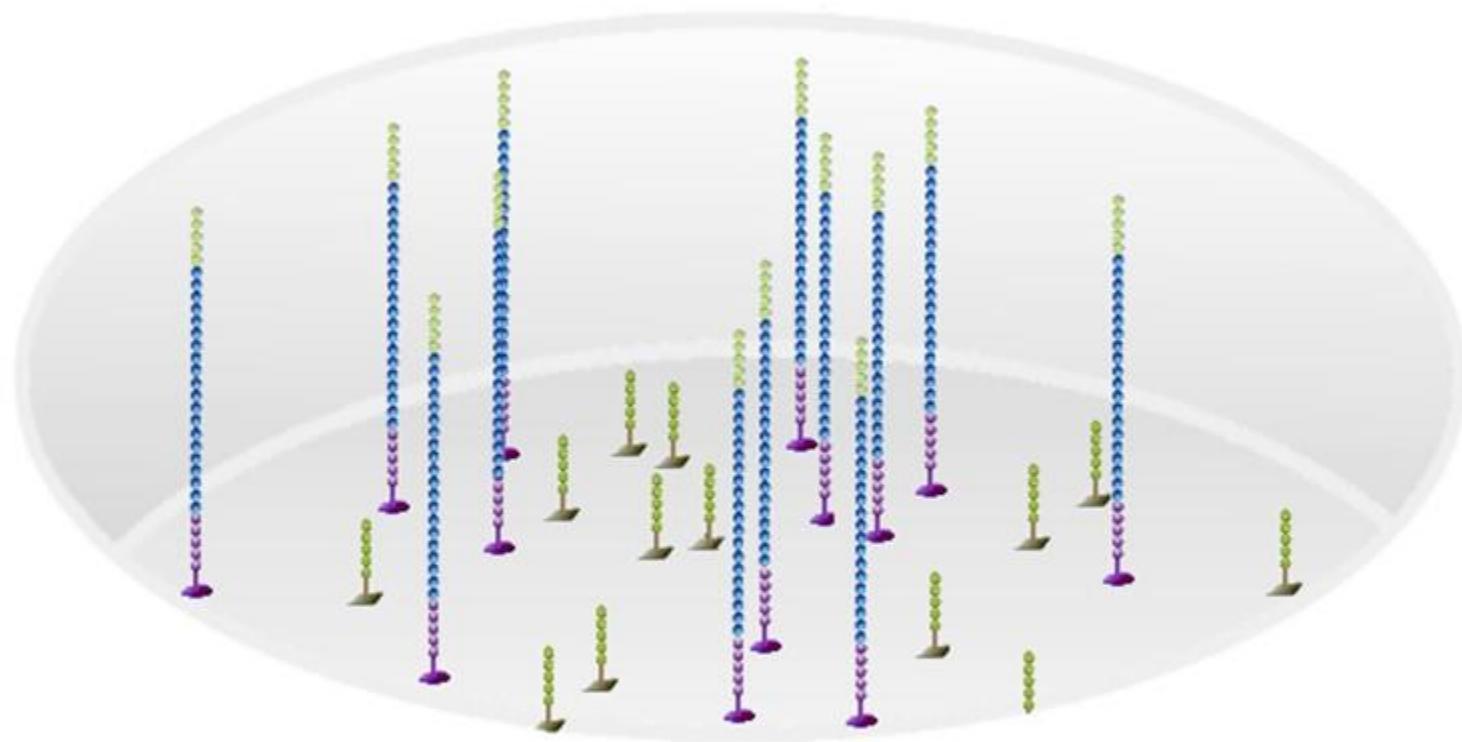


This process continues until no unused oligos are left on the flow cell surface

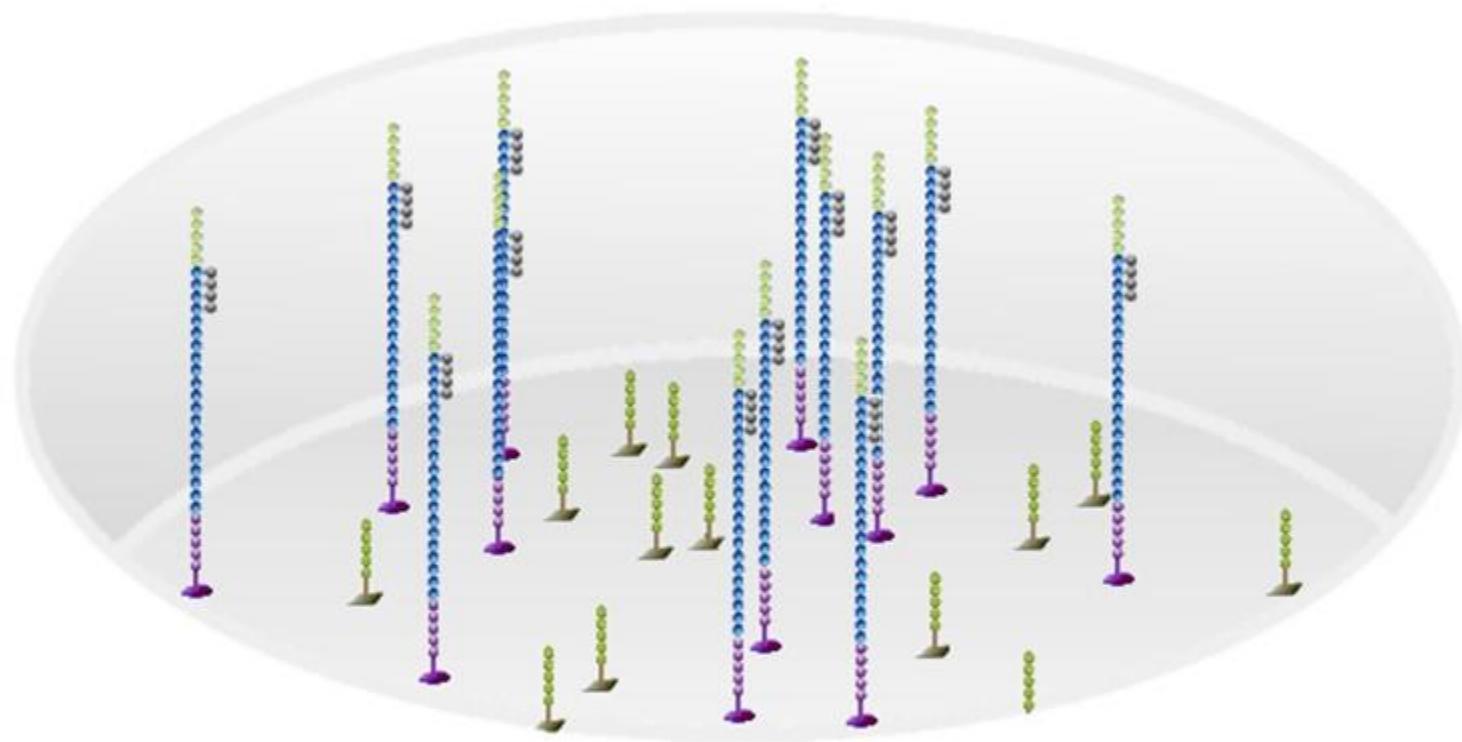


The diagram illustrates a collection of DNA molecules that have been converted into linear strands. These linear strands are represented by vertical lines composed of small, colored circular segments. The segments are colored in four distinct hues: blue, green, yellow, and purple. Each strand begins with a small, dark brown, V-shaped base at its bottom end. The strands are arranged in a somewhat scattered fashion within a large, light gray oval. Some strands are relatively short, while others extend significantly upwards, creating a sense of depth and variation.

To prepare the strands for sequencing,  
linearization is performed



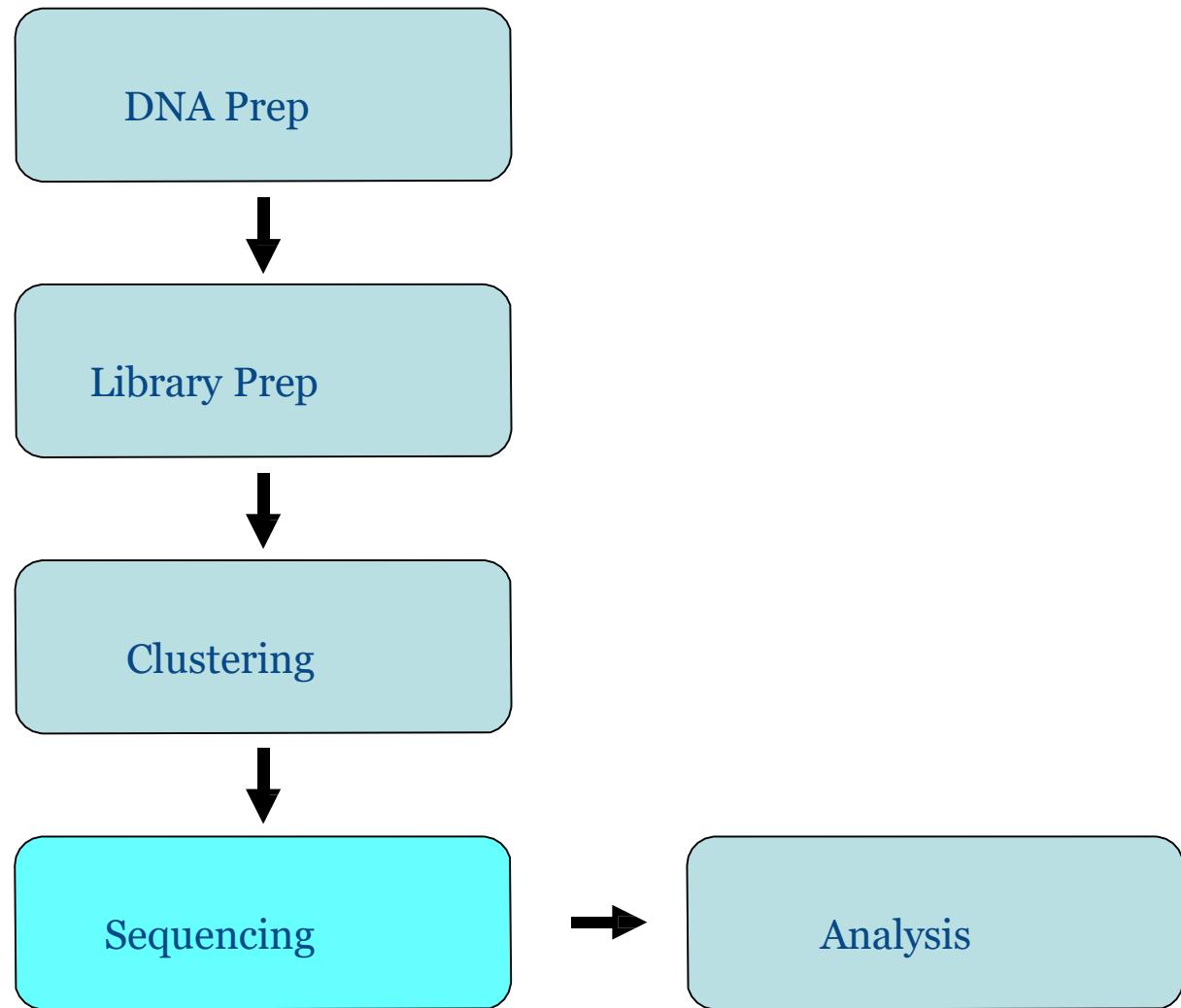
Reverse strands are cleaved



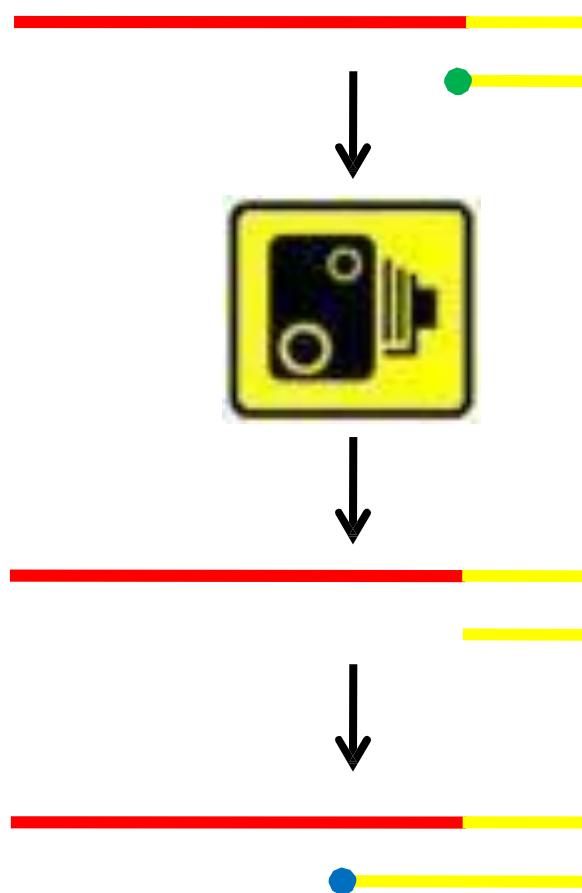
The Read 1 sequencing primer is  
hybridized



# Illumina workflow



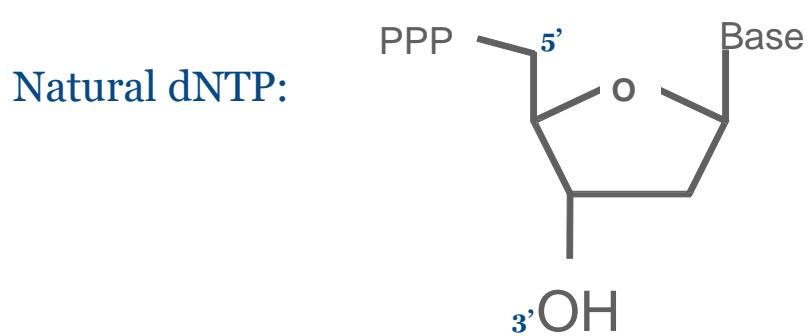
# Sequencing by Synthesis



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

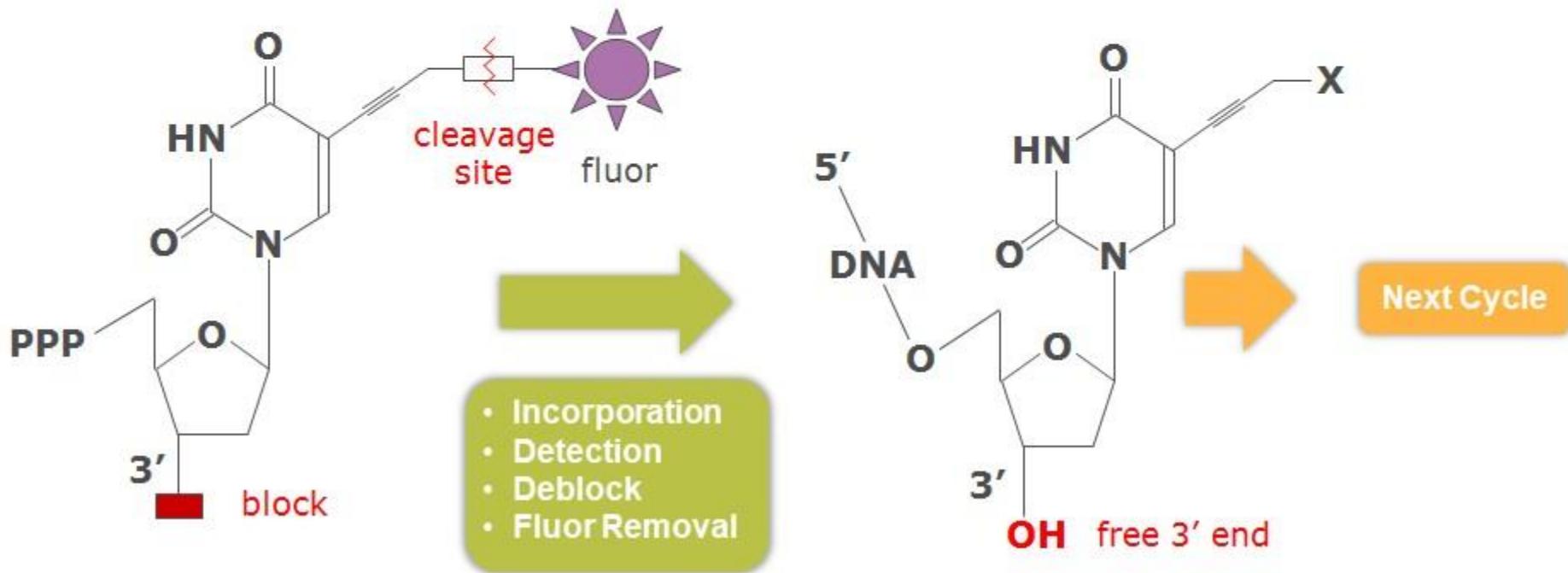


# Illumina Modified Nucleotides

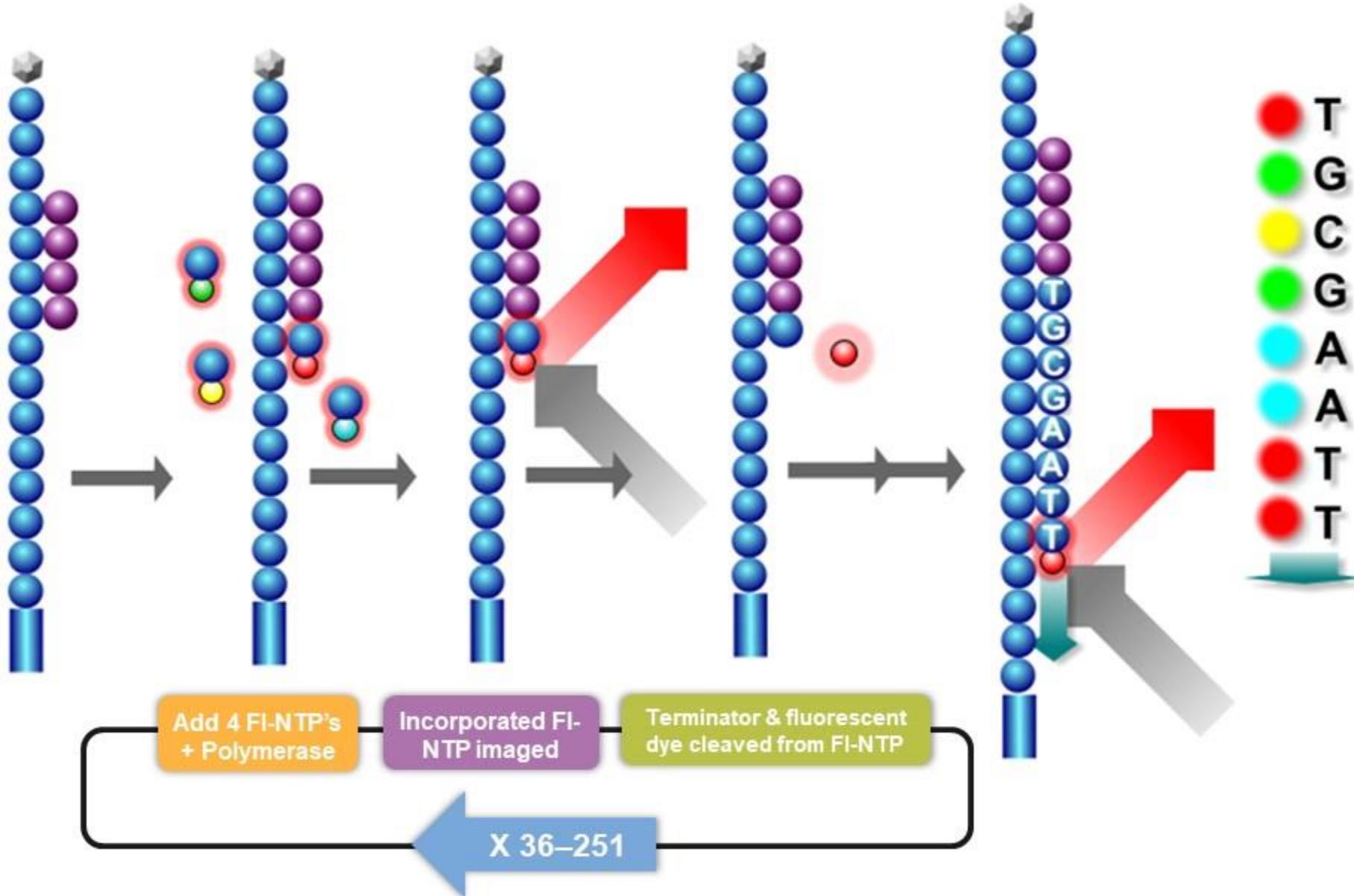


# Reversible Terminator Chemistry

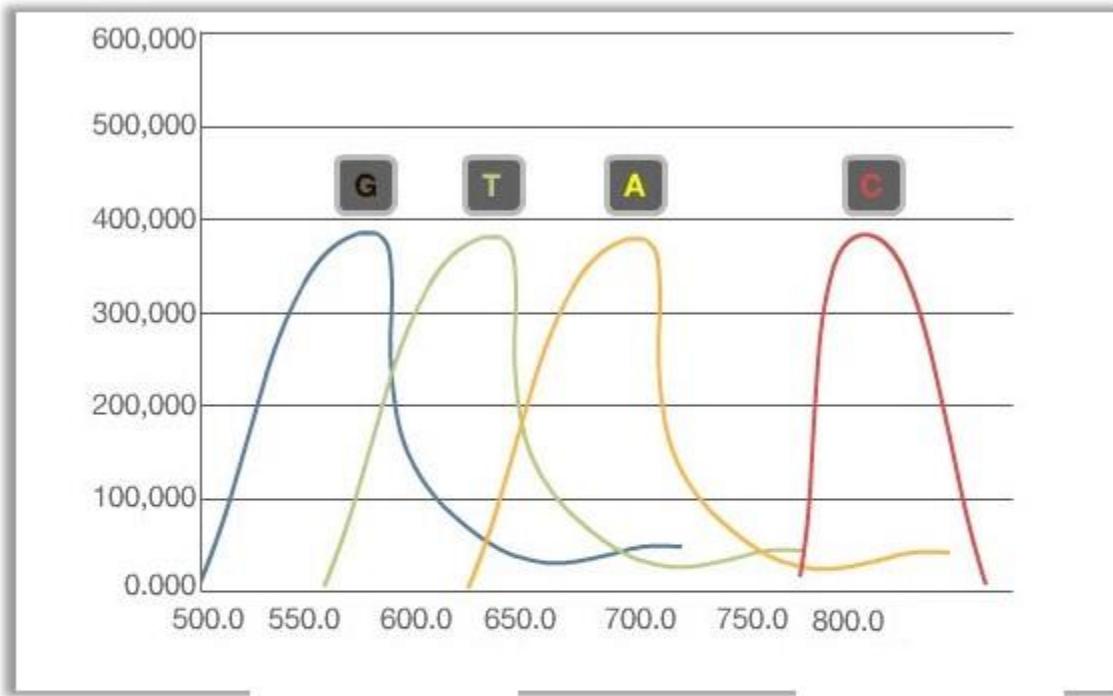
- All 4 nucleotides in 1 reaction
- Higher accuracy
- No problems with homopolymer repeats



# Sequencing By Synthesis (SBS)



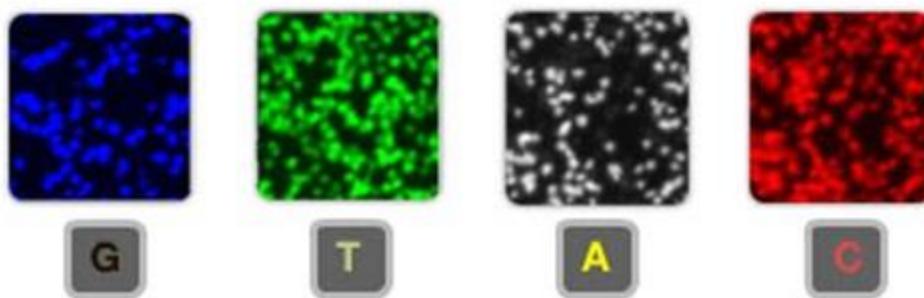
# Four Channel SBS Chemistry: GA, HiSeq, MiSeq



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

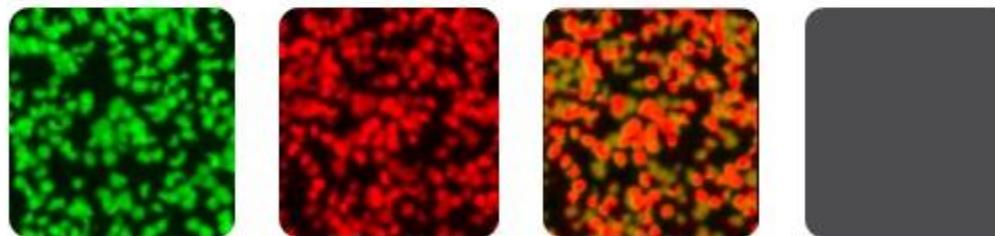
Collects four images:

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



# 2-Channel Sequencing Imaging Cycles

- 2 Channel Sequencing only requires 2 images and hence all the data from the 4 DNA bases are encoded in these 2 images.
- Uses the same sequencing by synthesis (SBS) method as 4 channel sequencing but allows more efficient acquisition of the data.
- Used by NovaSeq, NextSeq and MiniSeq



T

C

A

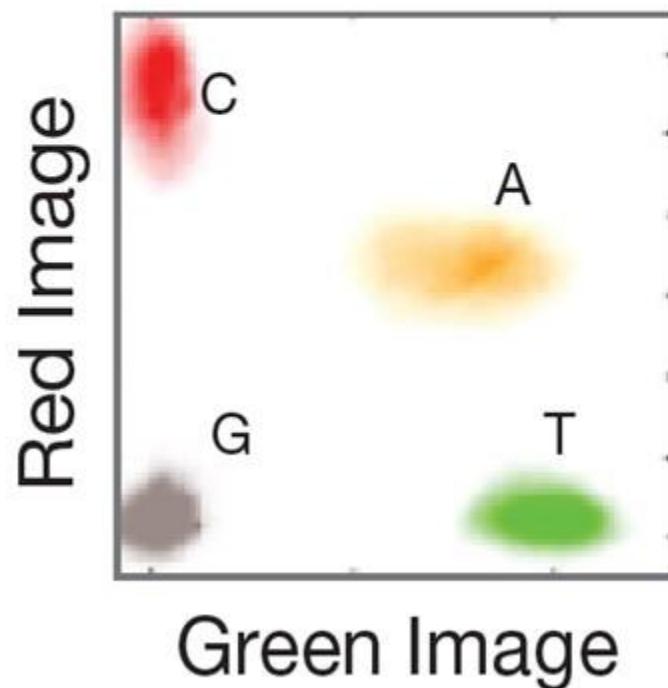
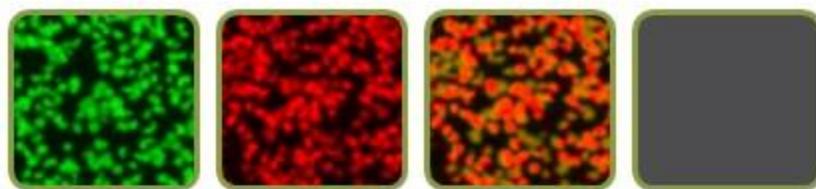
G

Channel	Green	Red	Green and Red	Dark - Neither
---------	-------	-----	---------------	----------------

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

# Two Channel SBS

- 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

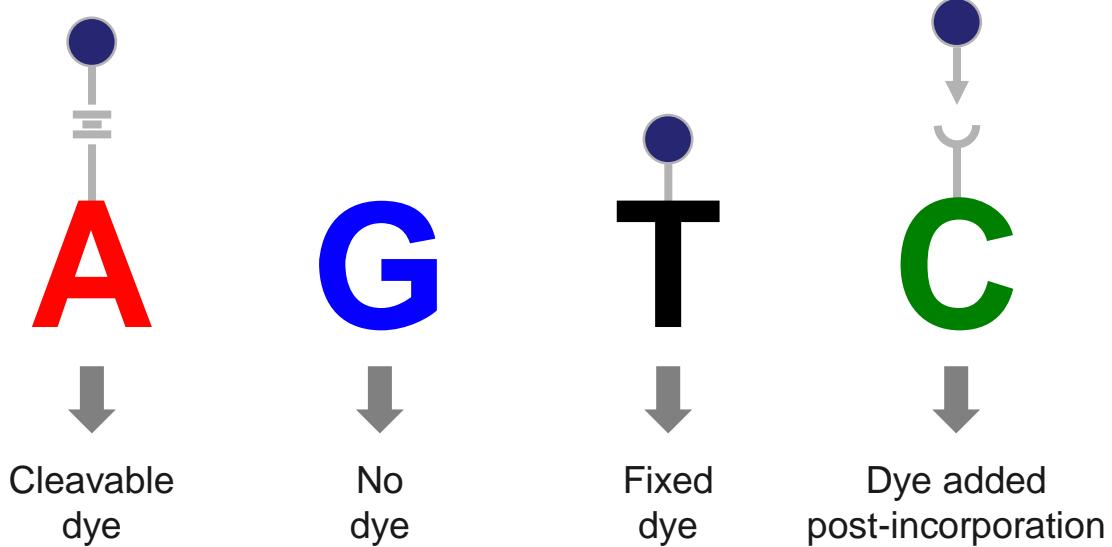


# A Closer Look At 1-Dye Chemistry

## *Cluster Generation and SBS*



What's the Different?



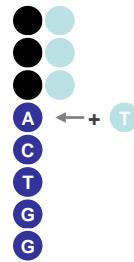
**Nucleotides** are labeled with a single dye, with the exception of the **G nucleotide**

# A Closer Look At 1-Dye Chemistry

## *SBS and Imaging*

### Sequencing by Synthesis

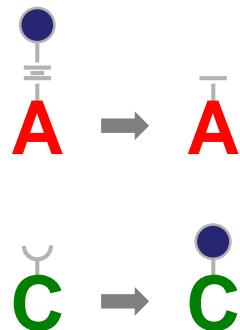
#### Sequencing Cycle



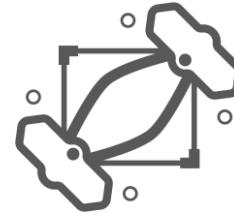
Incorporation



Imaging



Chemistry

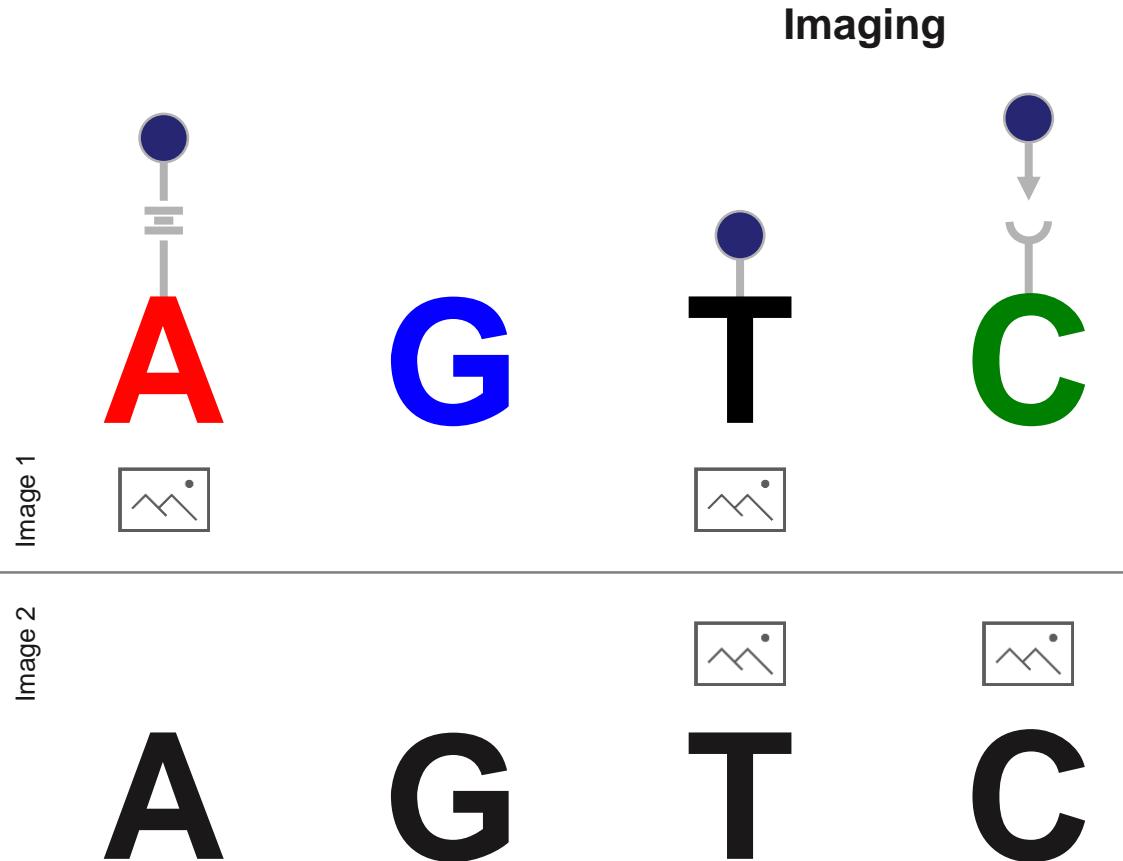


Imaging

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*



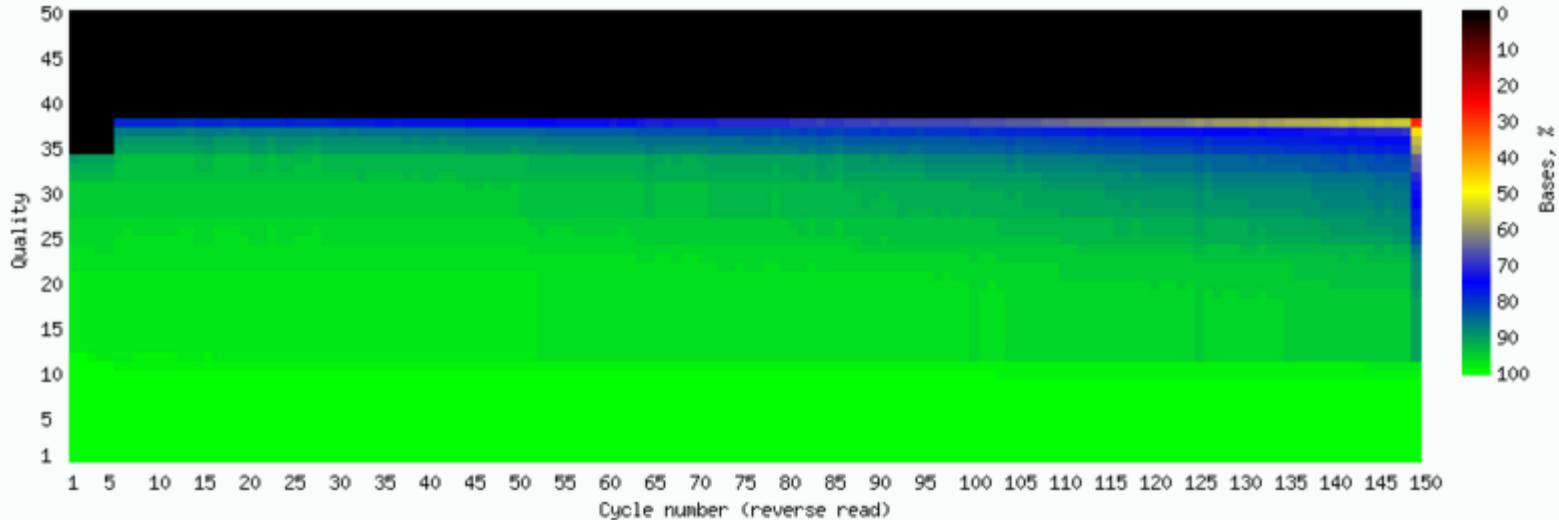
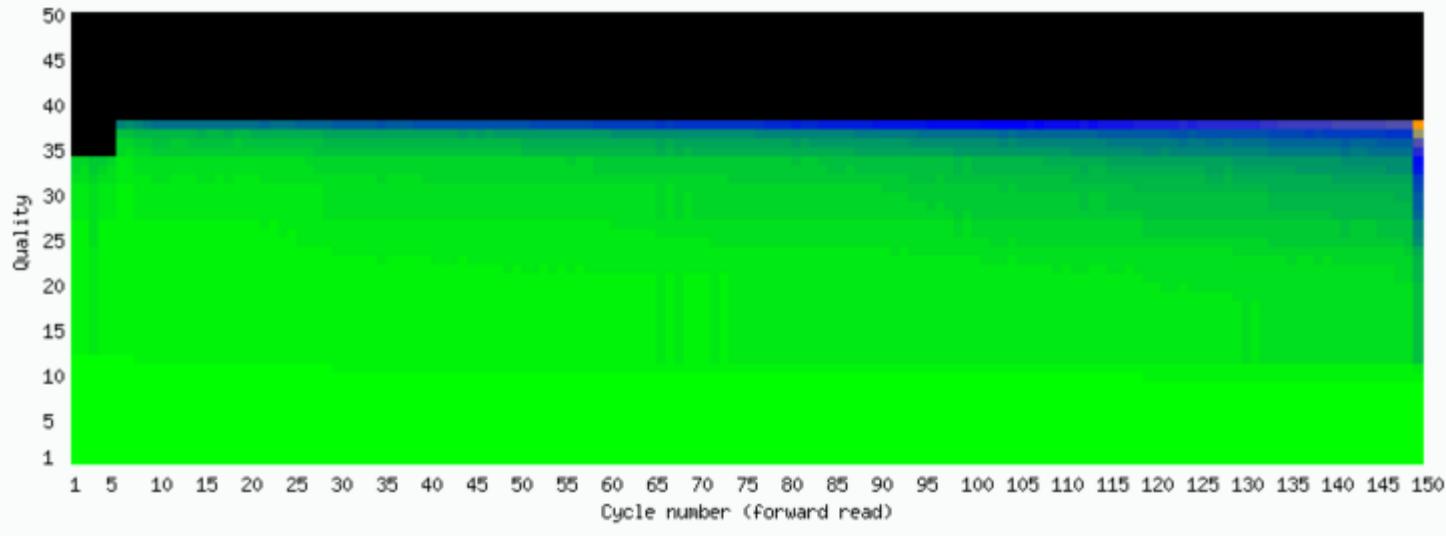
Using the two images, the iSeq™ 100 innovative data processing approach uniquely determines **which nucleotide was added to the growing template strand**



# Error

- » Sequence quality Q is reported on a log scale
- » Q<sub>10</sub> is 1 error in 10
- » Q<sub>20</sub> is 1 error in 100
- » Q<sub>30</sub> is 1 error in 1000
- » Q<sub>40</sub> is 1 error in 10000
- » Q<sub>50</sub> is 1 error in 100000

# Error limits read length



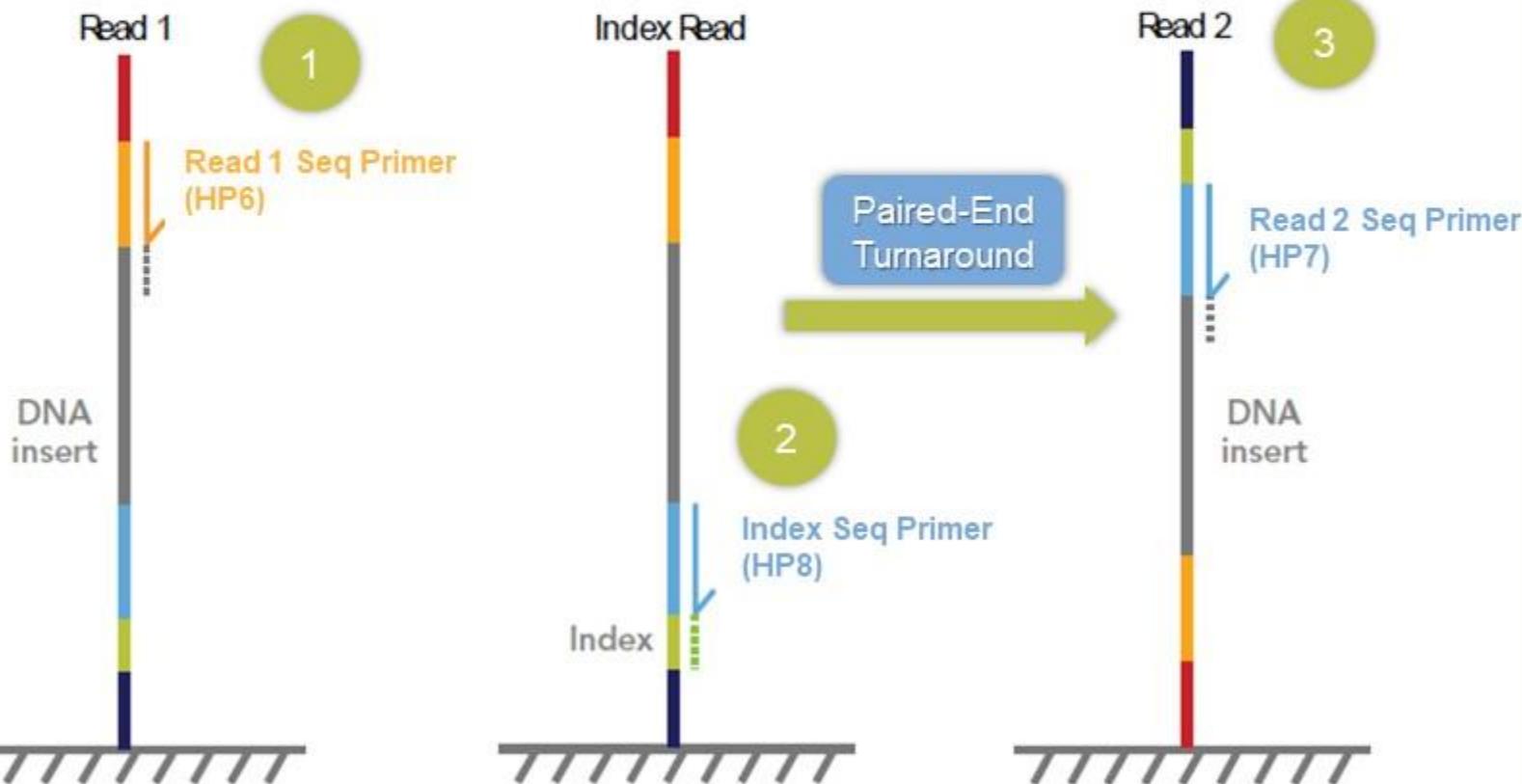


# Illumina sequencing with indexing

# Single Index Reads

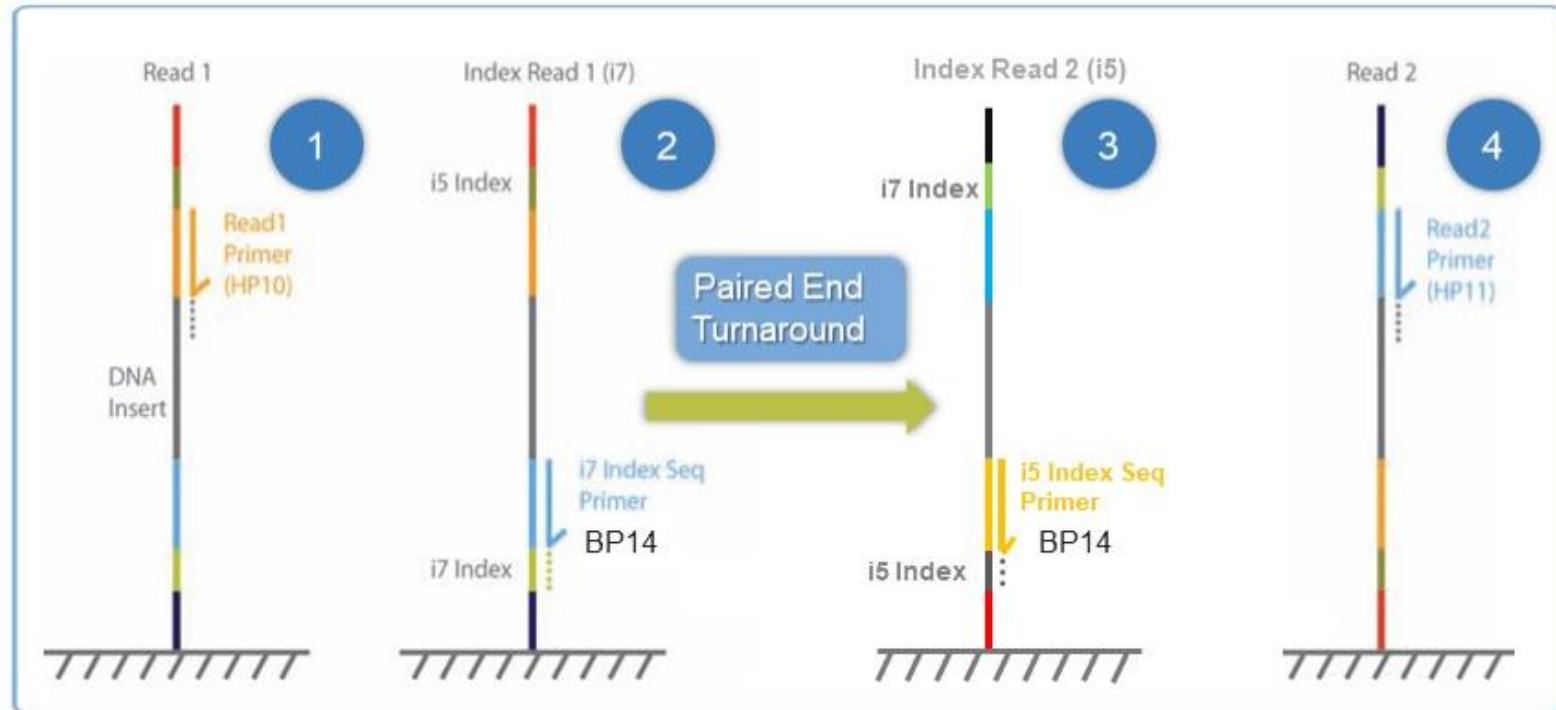
All Platforms

Single indexed sequencing utilizes 3 sequencing reads



# Dual indexing method B

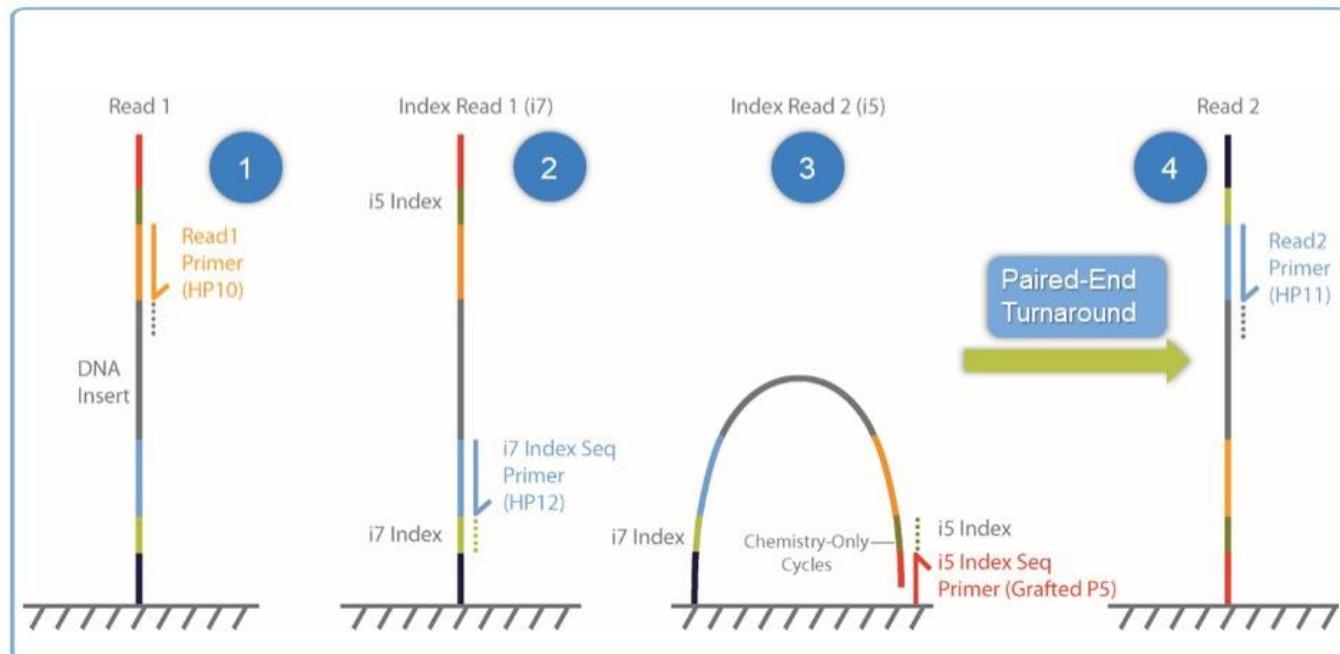
The Index 2 sequencing primer is part of the dual-indexing primer mix for iSeq 100, MiniSeq, and NextSeq. For HiSeq X, HiSeq 4000, and HiSeq 3000, the Index 2 sequencing primer is part of HP14, an indexing primer mix that contains primers for both index reads.



# 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.





# Instruments



# The Illumina Range 2022





Benchtop Sequencers		Production-Scale Sequencers				
		iSeq 100	MiniSeq	MiSeq Series +	NextSeq 550 Series +	NextSeq 1000 & 2000
Popular Applications & Methods		Key Application	Key Application	Key Application	Key Application	Key Application
Large Whole-Genome Sequencing (human, plant, animal)						
Small Whole-Genome Sequencing (microbe, virus)	●	●	●	●	●	●
Exome & Large Panel Sequencing (enrichment-based)					●	●
Targeted Gene Sequencing (amplicon-based, gene panel)	●	●	●	●	●	●
Single-Cell Profiling (scRNA-Seq, scDNA-Seq, oligo tagging assays)					●	●
Transcriptome Sequencing (total RNA-Seq, mRNA-Seq, gene expression profiling)					●	●
Targeted Gene Expression Profiling	●	●	●	●	●	●
miRNA & Small RNA Analysis	●	●	●	●	●	●
DNA-Protein Interaction Analysis (ChIP-Seq)				●	●	●
Methylation Sequencing					●	●
16S Metagenomic Sequencing		●	●	●	●	●
Metagenomic Profiling (shotgun metagenomics, metatranscriptomics)					●	●
Cell-Free Sequencing & Liquid Biopsy Analysis					●	●



## Benchtop Sequencers

## Production-Scale Sequencers



NextSeq 550 Series +



NextSeq 1000 & 2000



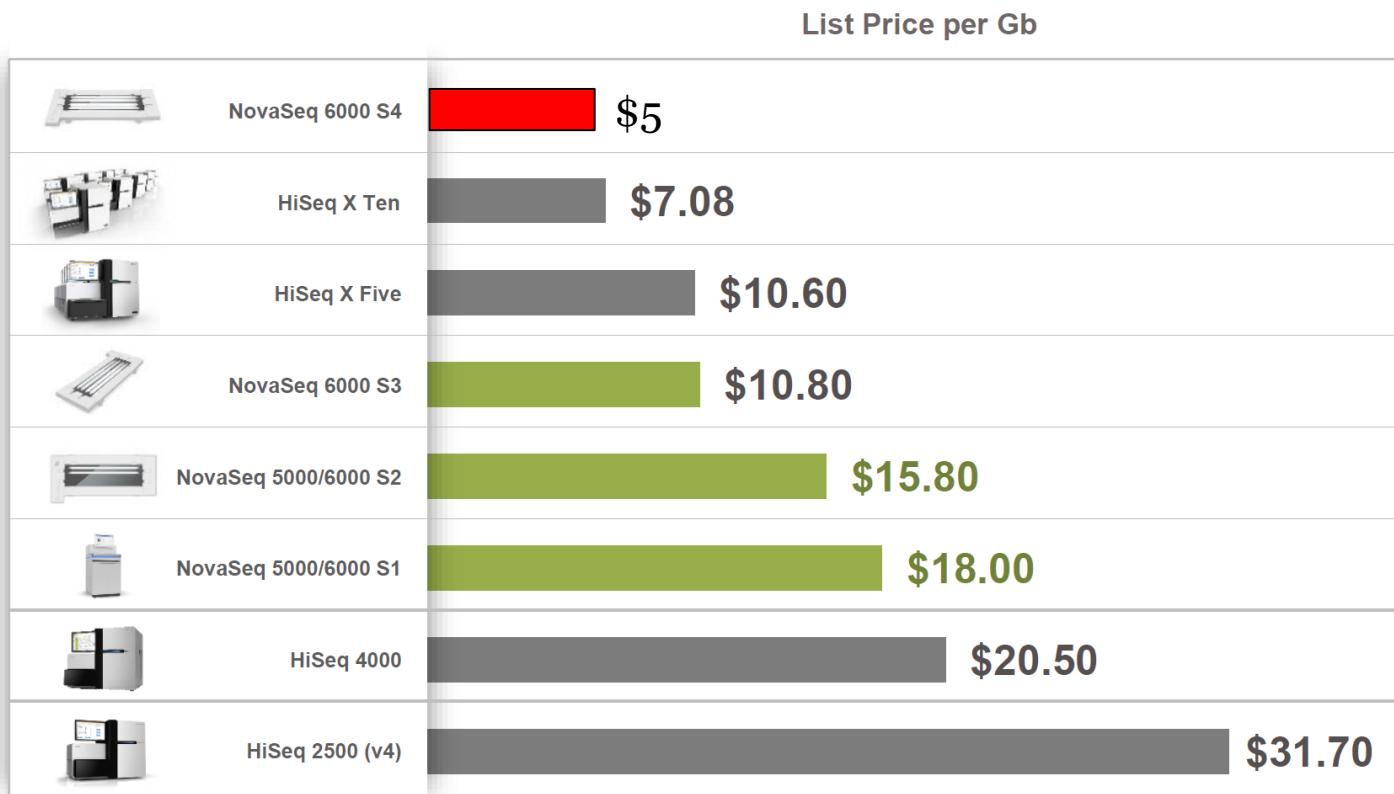
NovaSeq 6000

Popular Applications & Methods	Key Application	Key Application	Key Application
Large Whole-Genome Sequencing (human, plant, animal)			●
Small Whole-Genome Sequencing (microbe, virus)	●	●	●
Exome & Large Panel Sequencing (enrichment-based)	●	●	●
Targeted Gene Sequencing (amplicon-based, gene panel)	●	●	●
Single-Cell Profiling (scRNA-Seq, scDNA-Seq, oligo tagging assays)	●	●	●
Transcriptome Sequencing (total RNA-Seq, mRNA-Seq, gene expression profiling)	●	●	●
Chromatin Analysis (ATAC-Seq, ChIP-Seq)	●	●	●
Methylation Sequencing	●	●	●
Metagenomic Profiling (shotgun metagenomics, metatranscriptomics)	●	●	●
Cell-Free Sequencing & Liquid Biopsy Analysis	●	●	●



# NovaSeq Series

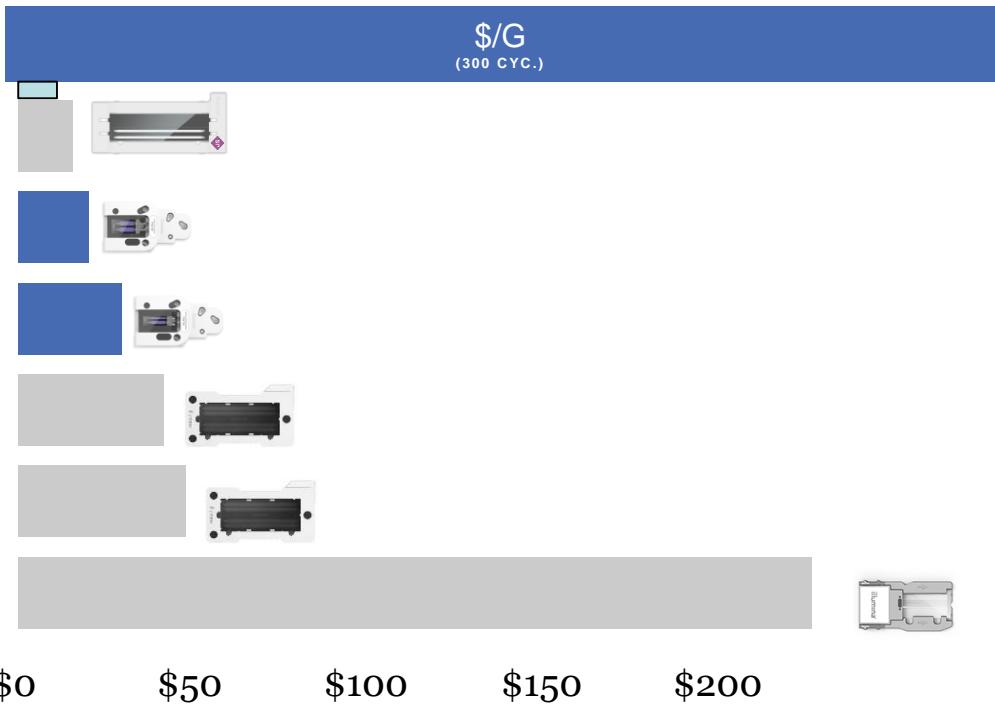
*Compelling price per data point enables highly-powered studies*



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

# Economics of Illumina's Sequencing Portfolio

Platform	Flow Cell	\$/G (300 CYC.)
<b>NovaSeq 6000</b>	S4 SP	\$5 \$15.40
<b>NextSeq 2000</b>	P3	\$20.00
<b>NextSeq 1000/2000</b>	P2	\$29.50
<b>NextSeq 550</b>	HO	\$41.38
<b>NextSeq 550</b>	MO	\$47.50
<b>MiSeq</b>	v2	\$236.00





# XLEAP SBS

illumina J.P. MORGAN HEALTHCARE CONFERENCE 28

## Breakthrough Chemistry X Sets New Industry Benchmark

2x Faster 2x Longer reads 3x Accuracy increase

New IP filings New large-scale manufacturing facility

GROUNDBREAKING IMPROVEMENTS FOR ALL FUTURE PLATFORMS

# September 2022

## Illumina Announce Novaseq X \$2/Gb with 25Billion read flowcell



### NovaSeq X series specifications

Output Range ~165 Gb - 16 Tb

Single reads per run 1.6 billion - 52 billion

Read length 2 × 150 bp

Run time ~13 hr - 48 hr

[View All NovaSeq X Specs](#)

 [View AR](#)

# 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



# Illumina Infinity

Illumina

J.P. MORGAN HEALTHCARE CONFERENCE

31

## Announcing Infinity High Performance Long Read Technology on Existing Illumina Platforms

**10x throughput**  
versus legacy long  
read technology

**Accelerate access**  
to remaining 5% of genic regions

**Up to 10kb**  
read lengths

**90% less DNA input**  
required compared to current  
long read technology

2H 2022  
EARLY ACCESS LAUNCH

**Euan Ashley, MB ChB, Ph.D.**  
Stanford Medicine

"This changes the short vs. long read debate and more importantly, changes how researchers and patients can get critical answers for rare disease."

**Shawn Levy, Ph.D.**  
Discovery Life Sciences / Hudson Alpha

"This technology can take us the rest of the way to the full genome, and with the speed, scale, and compatibility with existing Illumina, we could rapidly onboard this workflow."

ILLUMINA  
ON-MARKET





# Illumina

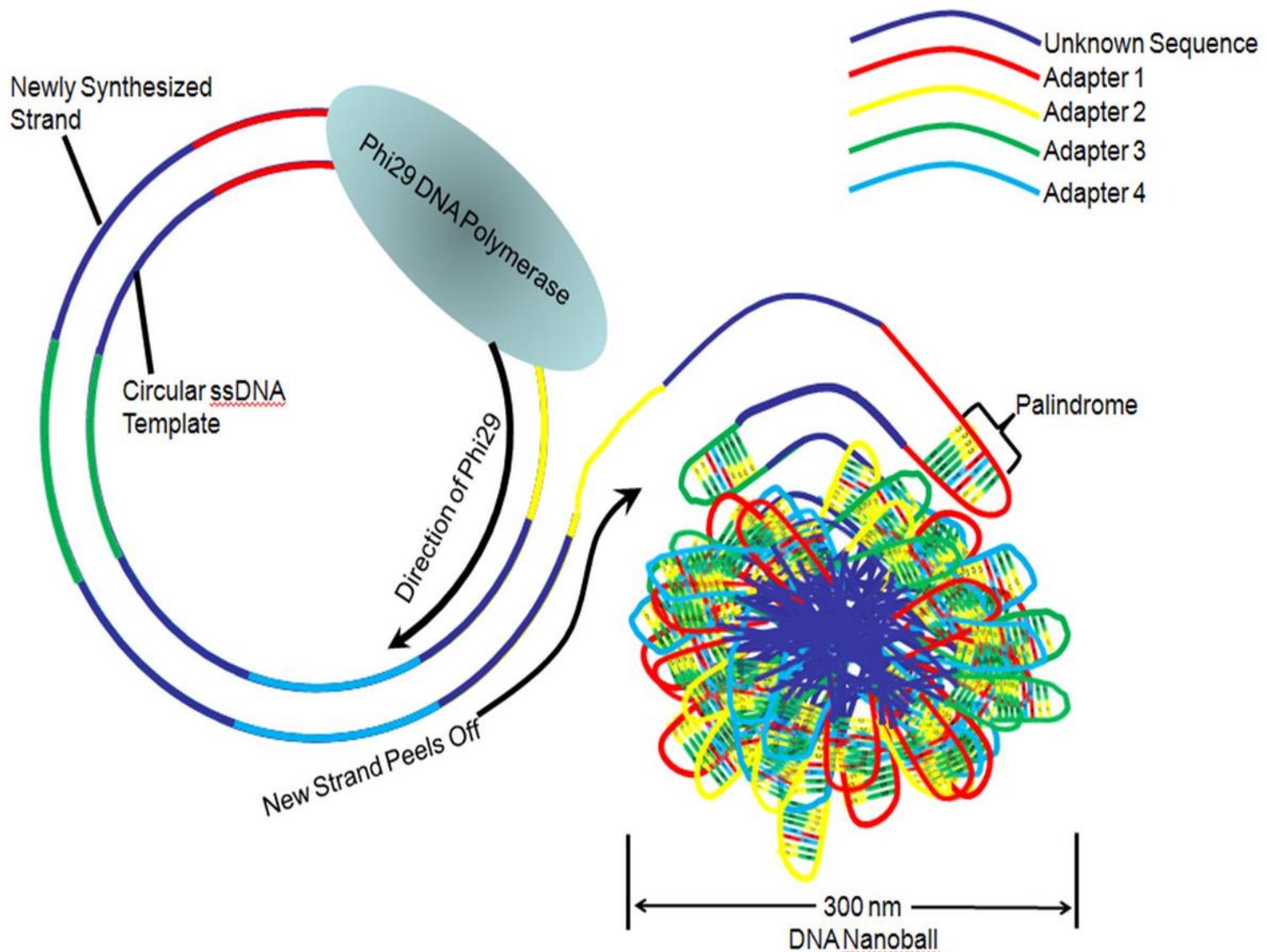
- » Cheap \$2-\$300/Gb
- » Highly accurate data mostly Q30
- » Massively parallel. Millions/billions of reads
- » Short read

# Complete Genomics



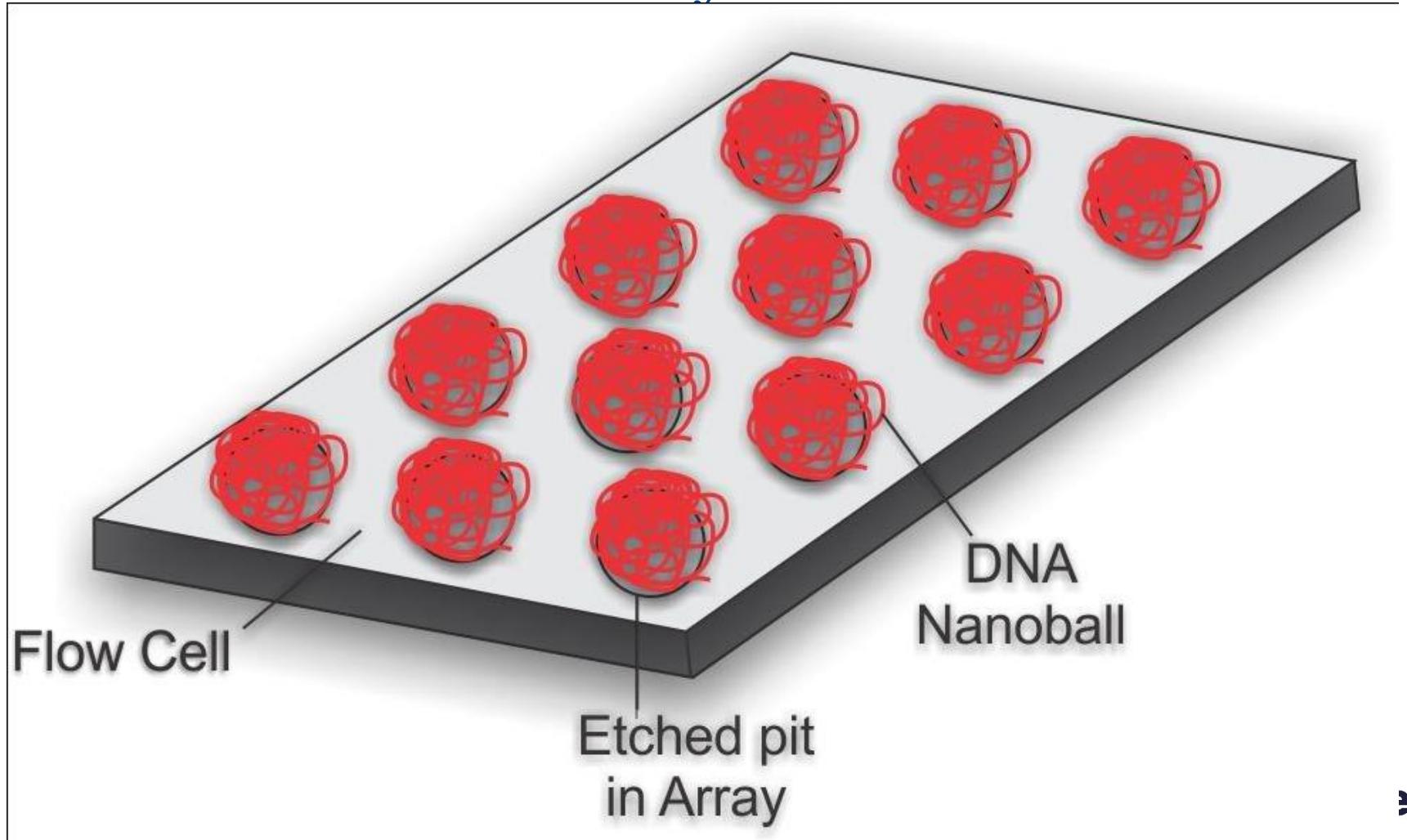
# Complete Genomics

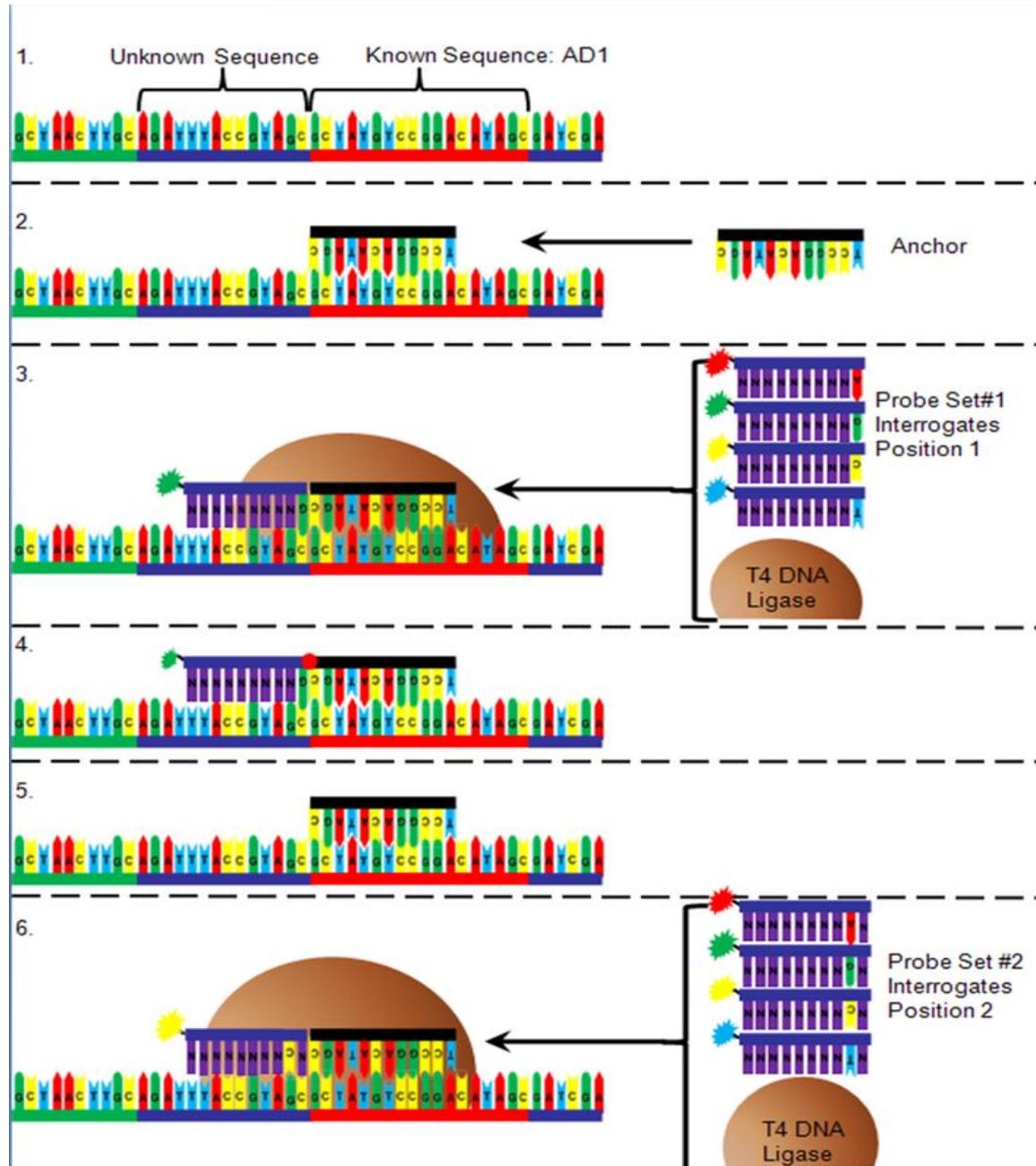




# Complete Genomics

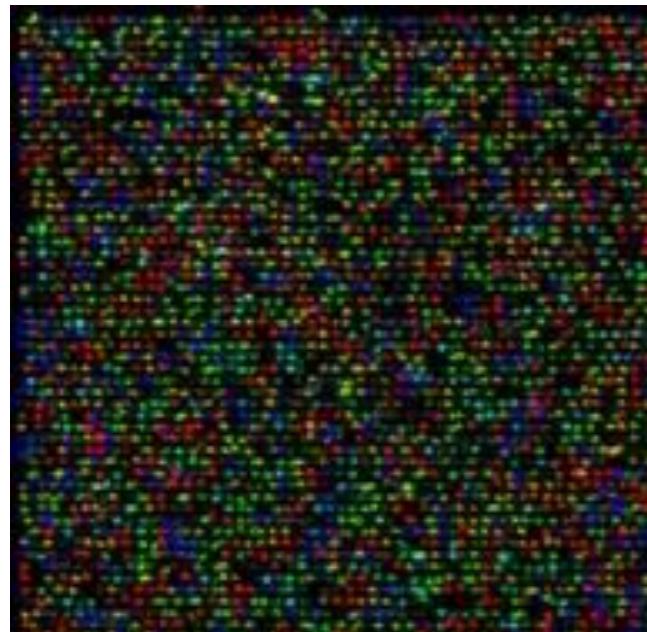
## Nanoball arrays on flowcell





# 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)
- » Raw error rate 0.5%
- » Consensus accuracy  $10^{-5}$





# Complete Genomics Revolocity



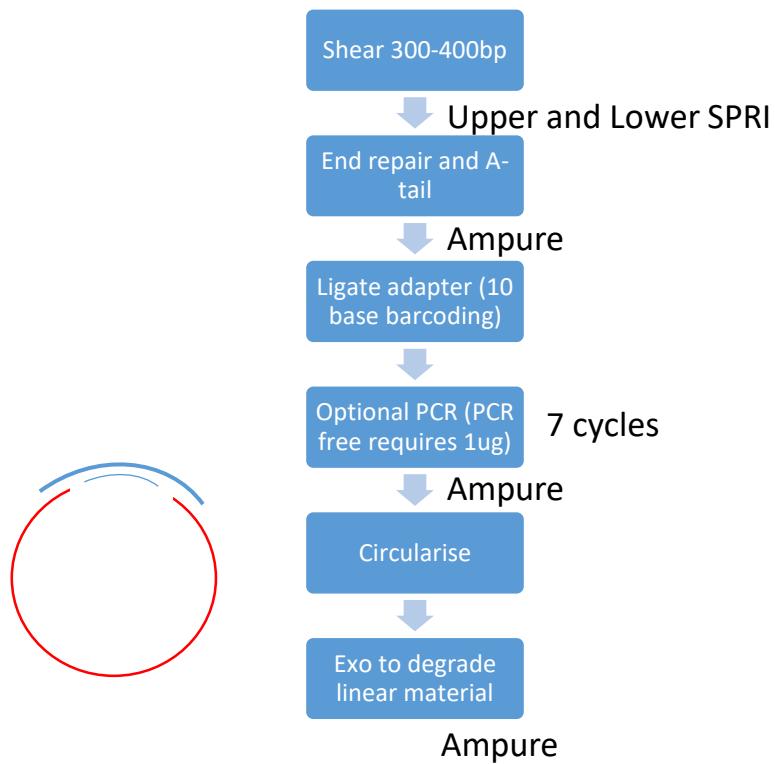
- » 10,000 genomes/year
- » \$12M
- » 10-120 samples at a time
- » Sample to answer
- » 8 day turnaround
- » 96% genome coverage
- » 1 error in a million (raw 0.5%)
- » 300 bp insert. 2 x 28bp reads



# BGI (MGI)

- » Acquired Complete Genomics
- » Making Sequencers for Chinese market
- » Short read sequencers
- » Accurate
- » Cheaper than Illumina

# Basic MGI library prep schema



# DNA Nanoball Generation

## *DNB Generation*

Low amplification bias

No amplification error accumulation

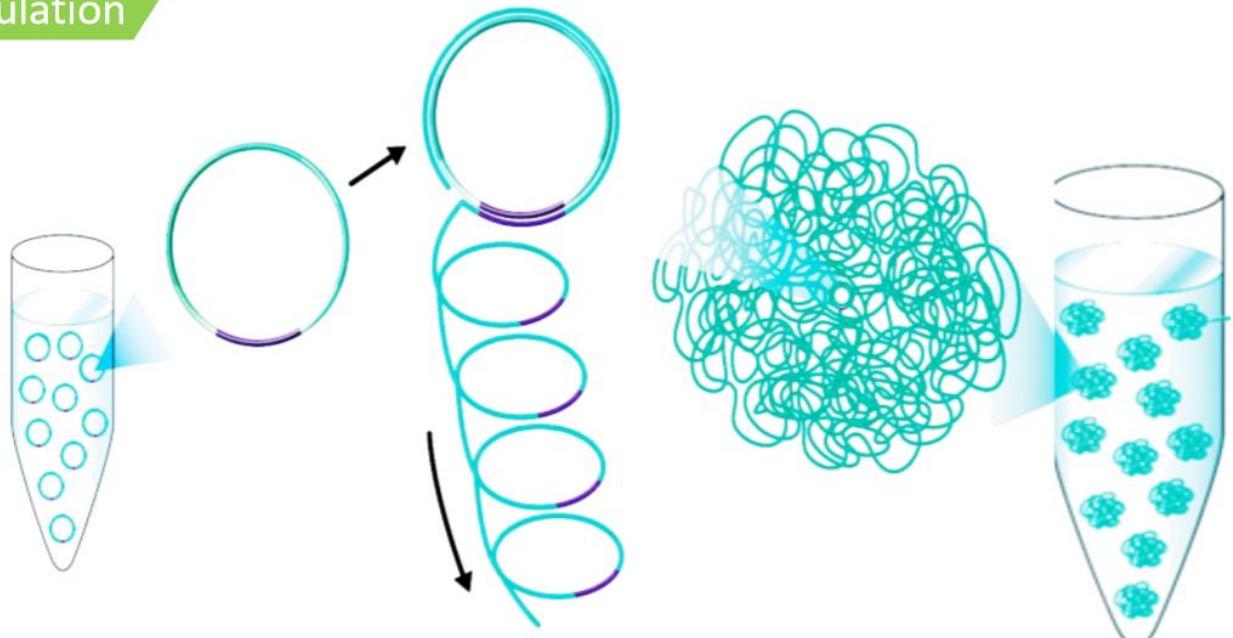
Linear Amplification

## RCA

From the ssCirDNA library to DNBs.

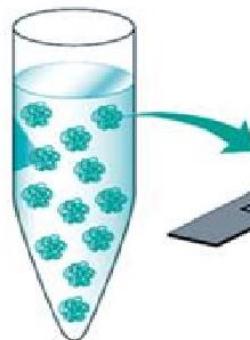
From 1 copy to 300-500 copies .

## Make DNA Nanoballs (DNB)

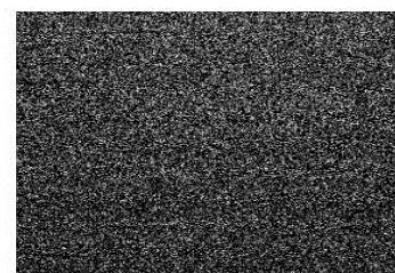
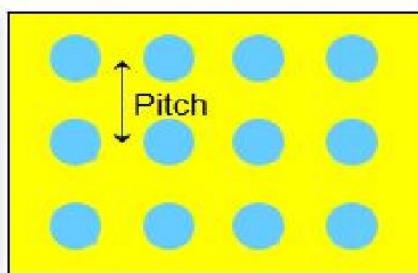
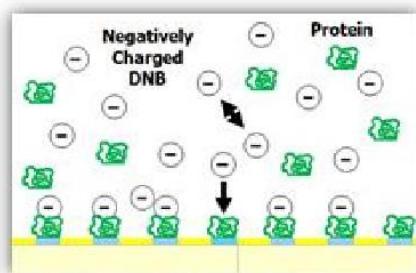


# Sequencing Chip loading

*How do our NGS platforms work?*



## DNB Loading



Pattern array flow cell

**DNBs loading**  
electrostatic interaction  
+  
Protein embedding

**Pattern array flowcell**  
Pitch: ~715 nm  
Diameter of DNBs: ~220 nm



# Advantages of MGI 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/MGI/Complete Genomics

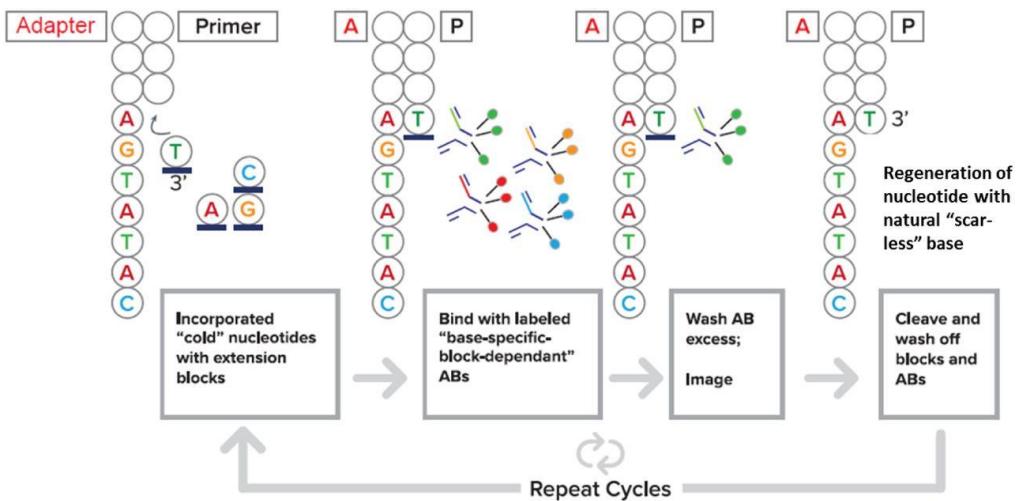
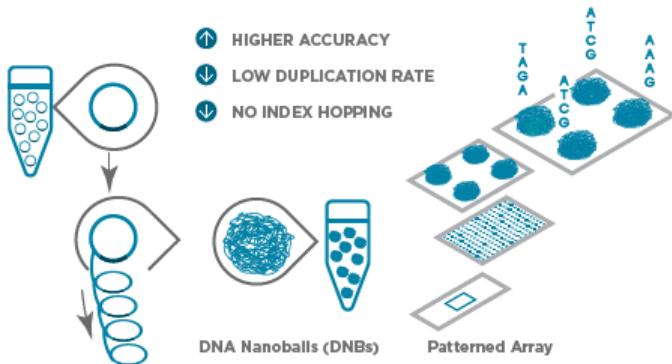
## Also StFLR and StereoSeq

### Full Portfolio of DNBSEQ™ Platform

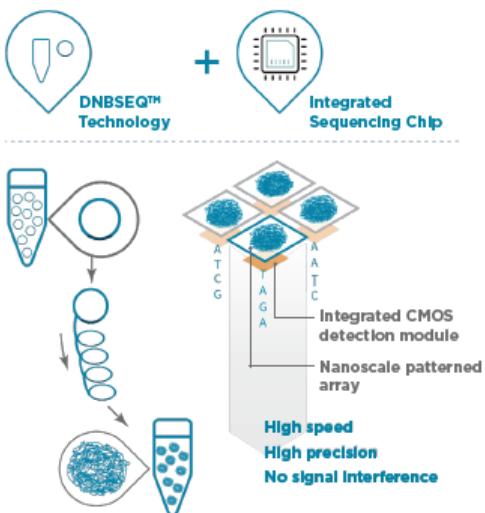


# MGI sequencing chemistry DNA nanoballs and cool MPS

## Proprietary DNBSEQ™ Technology



# G25

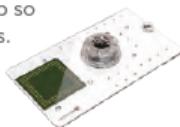


## PERFORMANCE Parameters

Reads	Read Length	Data Output	Run Time*	Q30
25M	SE100	2.5Gb	-5 hr	>90%
25M	PE150	7.5Gb	-20 hr	>80%

- **No Wash In Between Runs**

Sequencing reagents flow directly from the cartridge to the microfluidic sequencing chip so no wash is needed in between runs.



- **Plug and Go**

Replacing the traditional optical system to capture the signal image with the CMOS detection method greatly reduces the cost and weight of the sequencer. The traditional optical system tends to be very sensitive to operating temperature and humidity.

Adopting self-luminous dye to generate a signal eliminates the need for lasers which can be costly and requires very strict ambient temperature and humidity control, resulting in lower cost.



# DNB seq G99

- » 8-48Gb
- » 2 x 150 in 12 hours



# MGI DNB Seq G400

- » Approx £225k (2M Yuan)
- » \$600 per human genome as a service
- » Cheaper if had own instrument?
- » 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 to Riken, several others in Germany and UK
- » 4 colour SBS

# G400



## PERFORMANCE Parameters

Flow Cell Type	Reads	Read Length	Data Output*	Run Time**	Q30
	550M	SE100	55Gb	13 hr	>85%
		PE100	110Gb	26 hr	>85%
		PE150	165Gb	37 hr	>80%
	300M	PE300	180Gb	98 hr	>80%
	1500-1800M	SE50	75-90Gb	14 hr	>85%
		SE100	150-180Gb	25 hr	>85%
		PE100	300-360Gb	38 hr	>85%
		PE150	450-540Gb	56 hr	>80%
		SE400	600-720Gb	109 hr	>70%
		PE200	600-720Gb	107 hr	>75%



# DNBseq T7 (\$1.5/Gb)

15-60 genomes/day



**High Speed**

24-30 hours for  
PE150



**High Flexibility**

Run 4 independent  
flow cells at any time



**Ultra-high  
Throughput**

20,000 of 30x WGS  
in one year

## PERFORMANCE Parameters

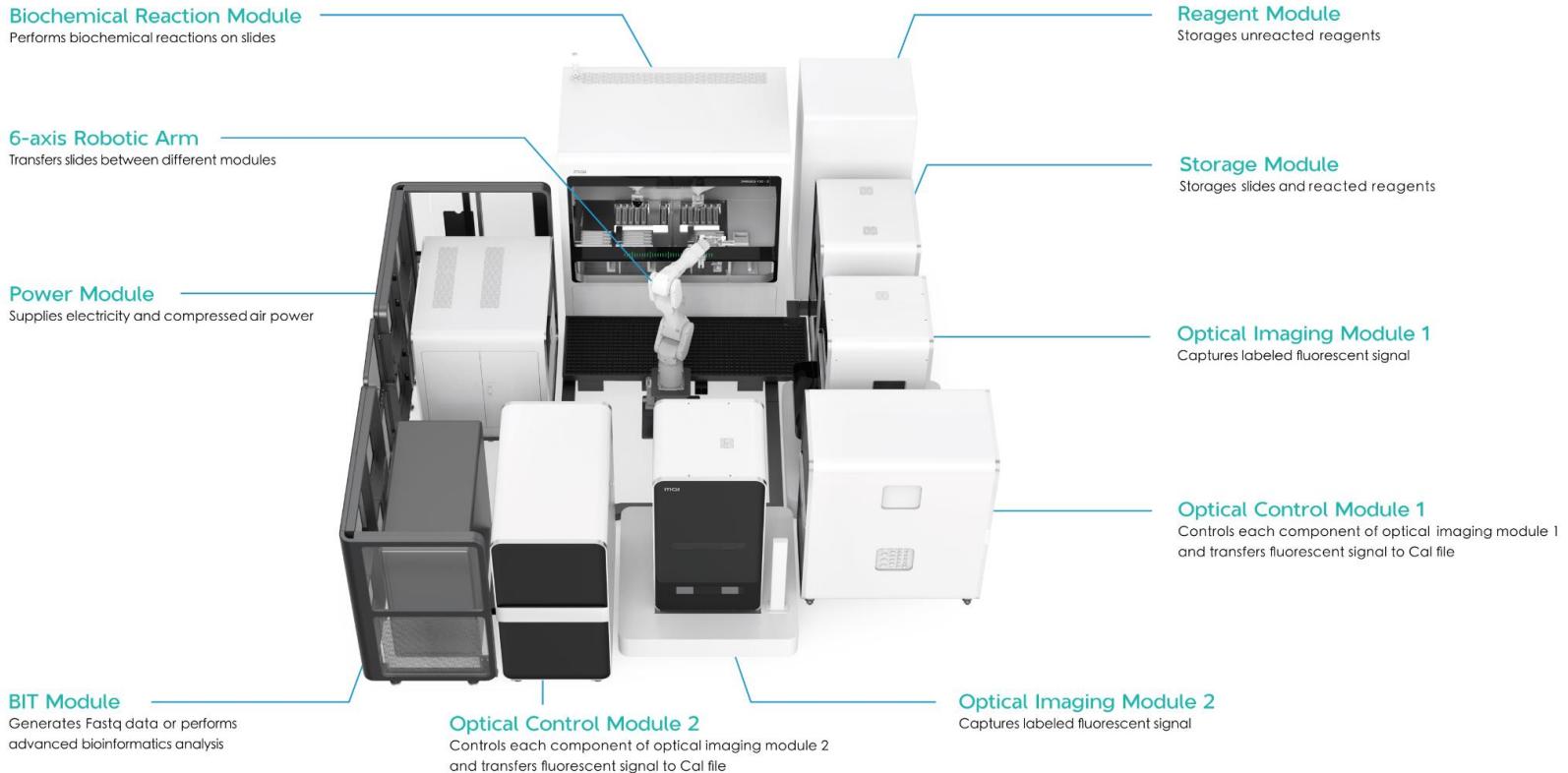
# of Flow Cells	Reads	Read Length	Data Output*	Run Time**	Q30
1-4	5800M*	SE50	290 - 1160 Gb	5.5 hr (1 FC) 6 hr (4 FC)	>85%
		PE100	1 - 5 Gb	17.5 hr (1 FC) 20.5 hr (4 FC)	>85%
		PE150	1.7 - 7 Gb	24.5 hr (1 FC) 29.5 hr (4 FC)	>80%



# T20x2

- » \$0.99/Gb inc instrument depreciation
- » 72Tb/run
- » 8 slides, 150 genomes one each
- » 2 x 100, 2 x 150 (Q3 2023)
- » 3.5 day run time

# T20 x2





# Pacific Biosciences



RSII : 1800 lbs. and ~11 feet long !



# PacBio

- » Long Read sequencer
- » Single Molecule Real Time
- » Current technology of choice for de-novo sequencing projects

# Pacific Biosciences

## Sequel II/I<sup>e</sup>



- » 8 million ZMWs/SMRT
- » 100Gb/SMRT cell
- » ~\$9/Gb
- » \$495K/\$525 instrument



# » Revio system

## Throughput



**360 Gb**  
HiFi yield per run



**24 hour**  
Sequencing time



**1,300 genomes / year**  
30× human genome equivalent



List price, USD

**\$779,000**

**\$995\*** /genome

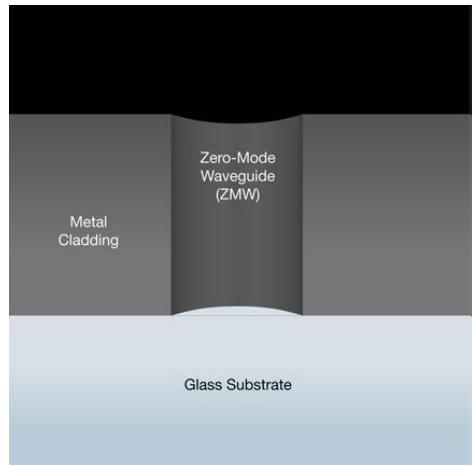
\* \$995 for sequencing reagents for one Revio SMRT Cell, which has an expected yield of 90 Gb, equivalent to a 30× human genome. Expected pricing subject to change, your local sales representative can provide detailed pricing in your currency.

PacBio

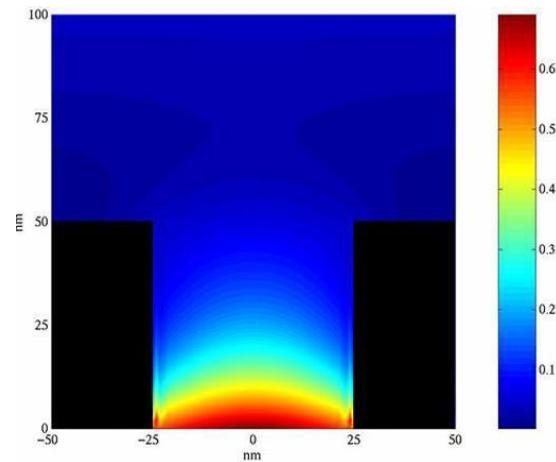
Learn more:



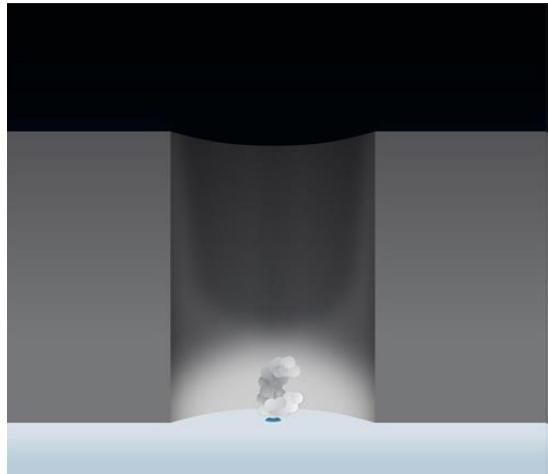
# Pac Bio Technology



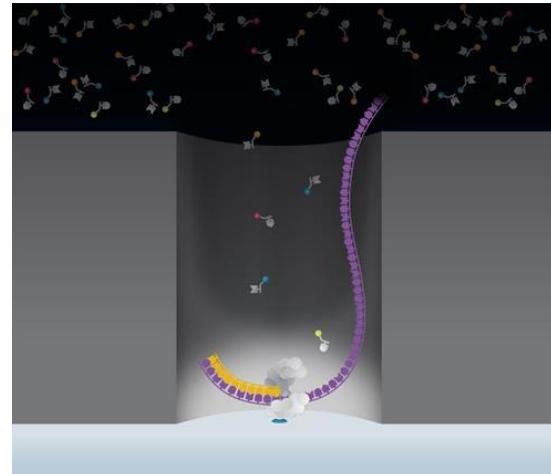
Individual ZMW



Laser light illuminates the ZMW

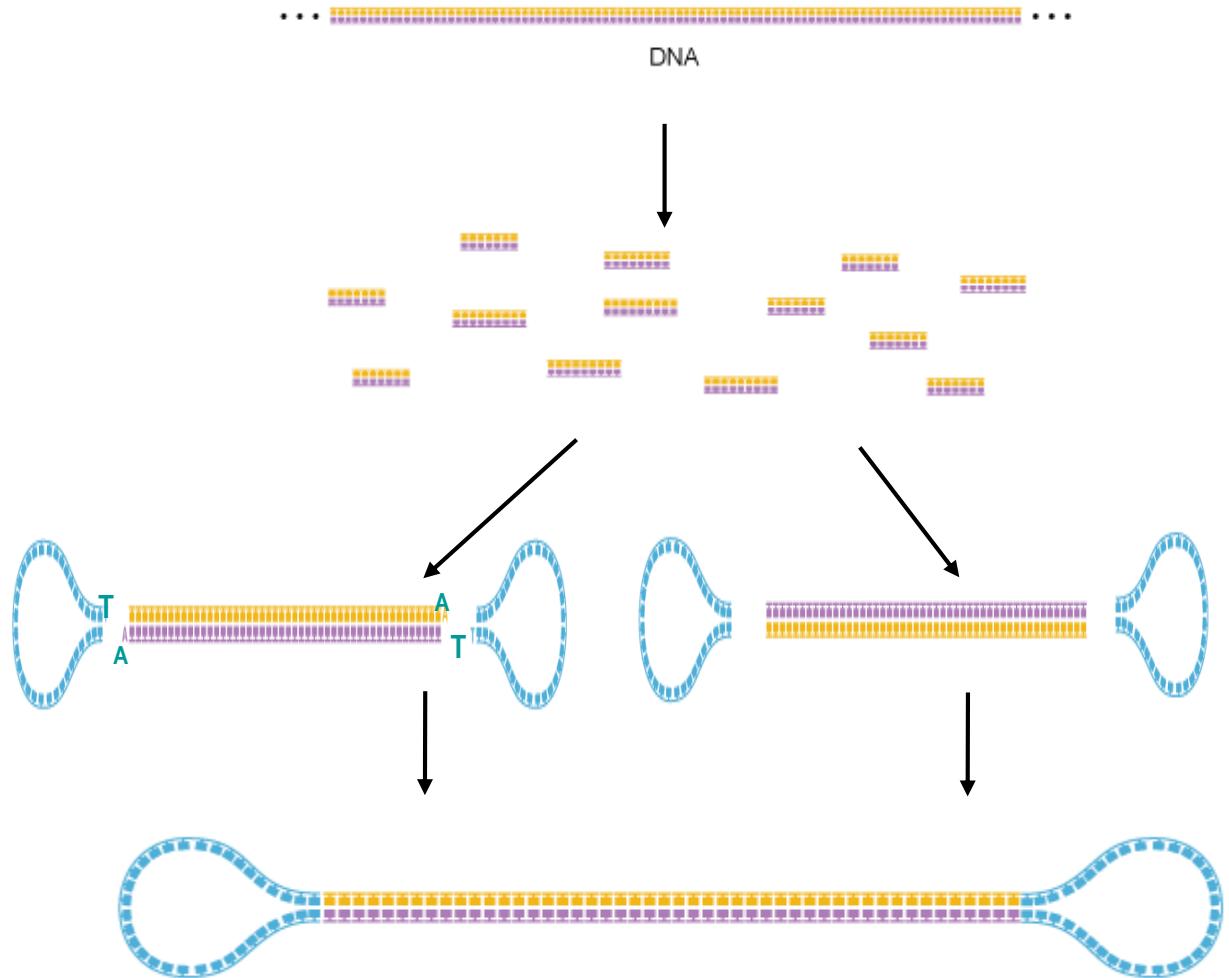
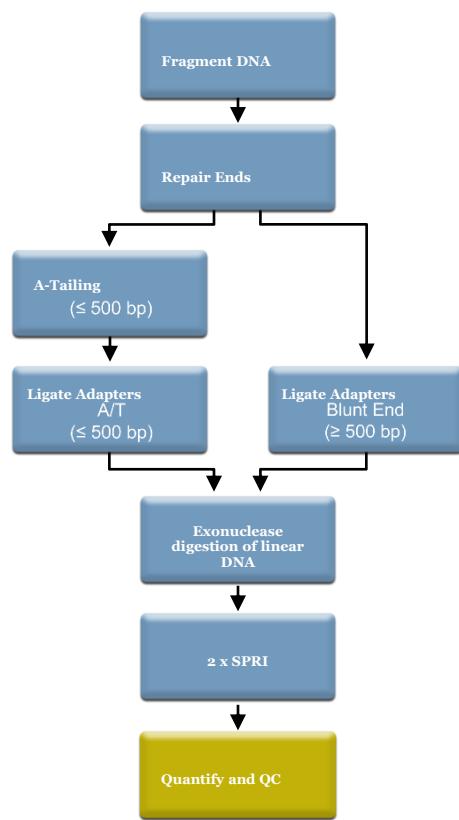


ZMW with DNA polymerase

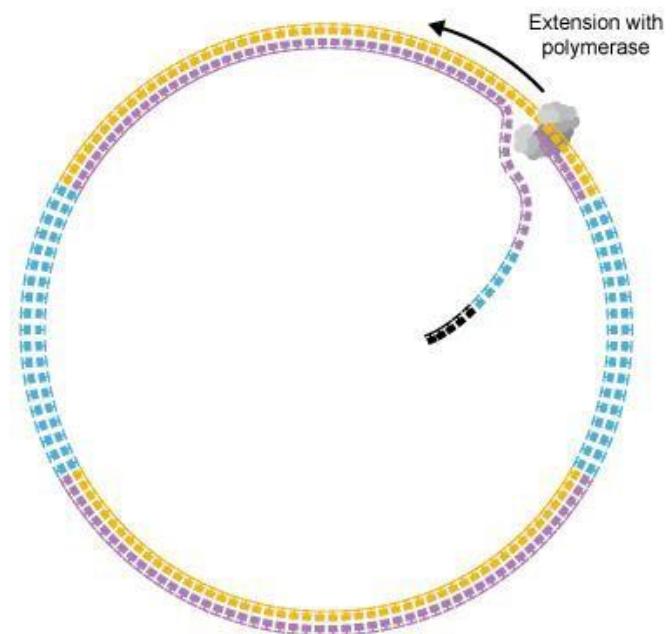
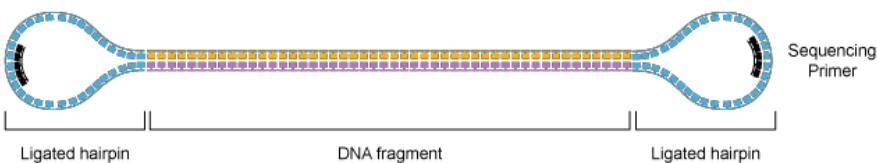


ZMW with polymerase + nucs.

# PacBio Library Prep

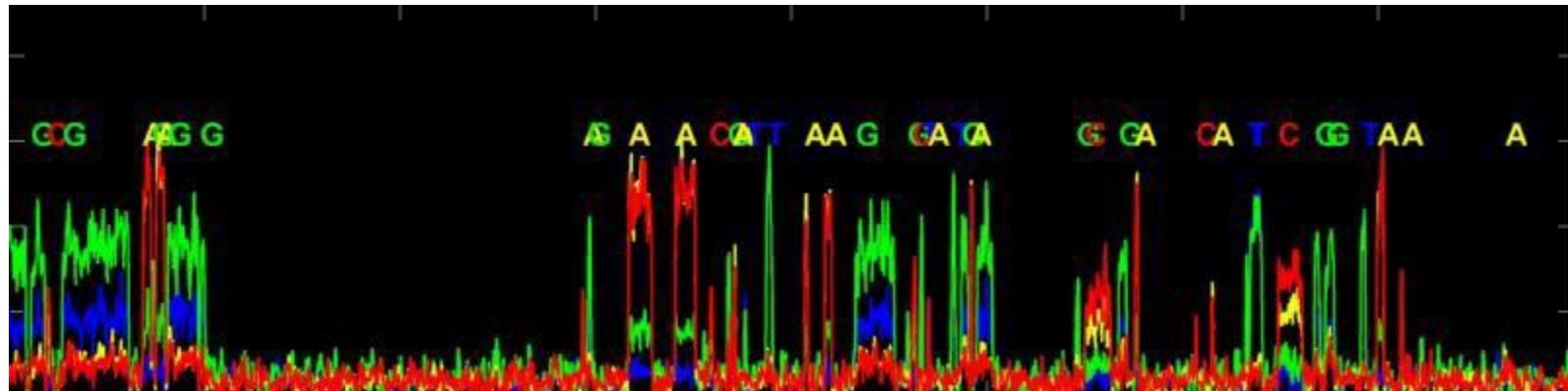


# PacBio Template Preparation



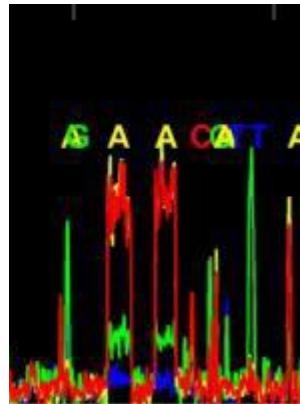
- 1 Anneal primer**
- 2 Bind polymerase**
- 3 Sequence**

# PacBio Sequencing

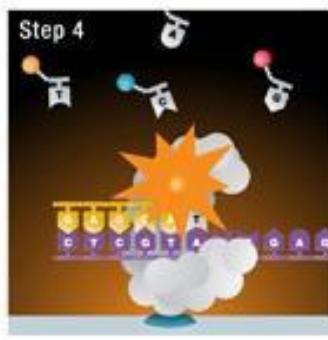




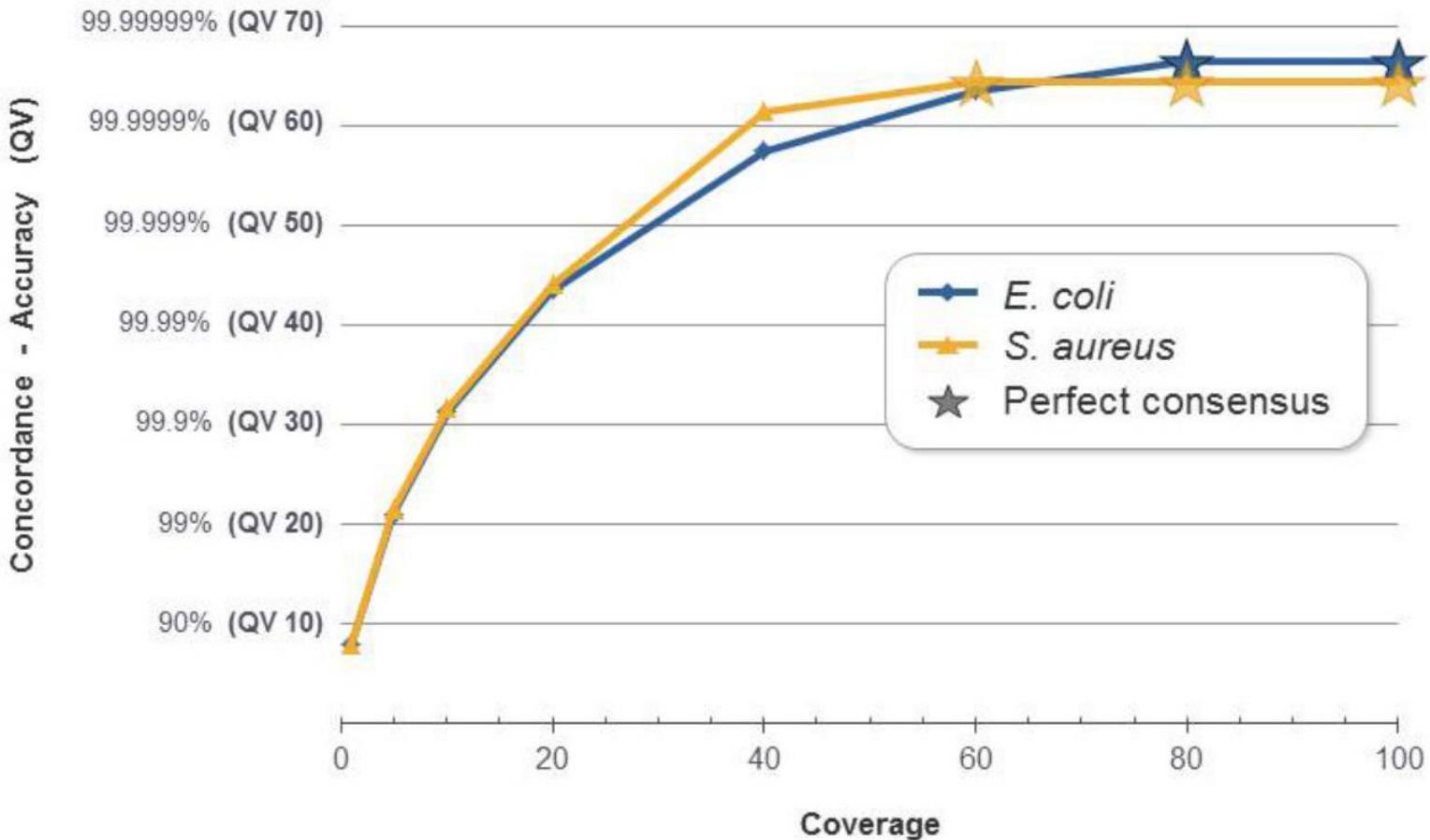
# PacBio Sequencing



- » Some bases added very quickly and missed
- » Some wrong bases flirt with active site and go away
- » SMRT cell has 8million ZMWs

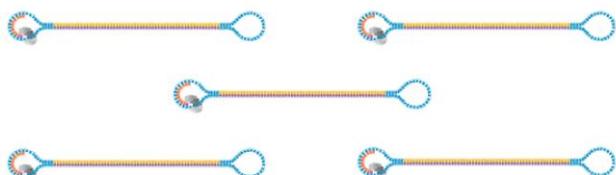


# Accuracy



## 2 MODES OF SMRT SEQUENCING

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



Molecule 1

Molecule n

---

## Genotyping, SV detection, *de novo* assembly

- Circular Consensus Sequencing (CCS):

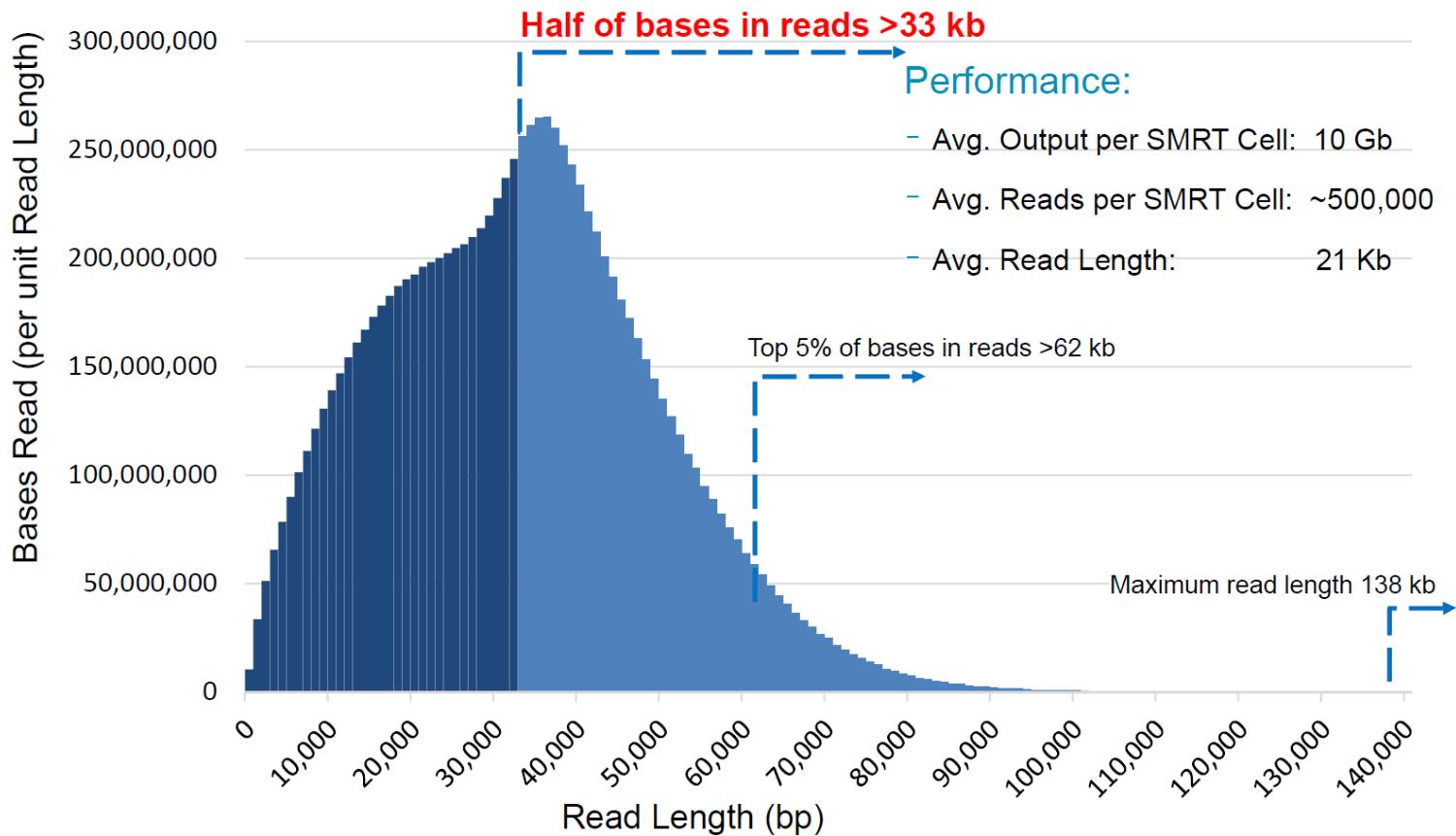


## *Consensus sequence*

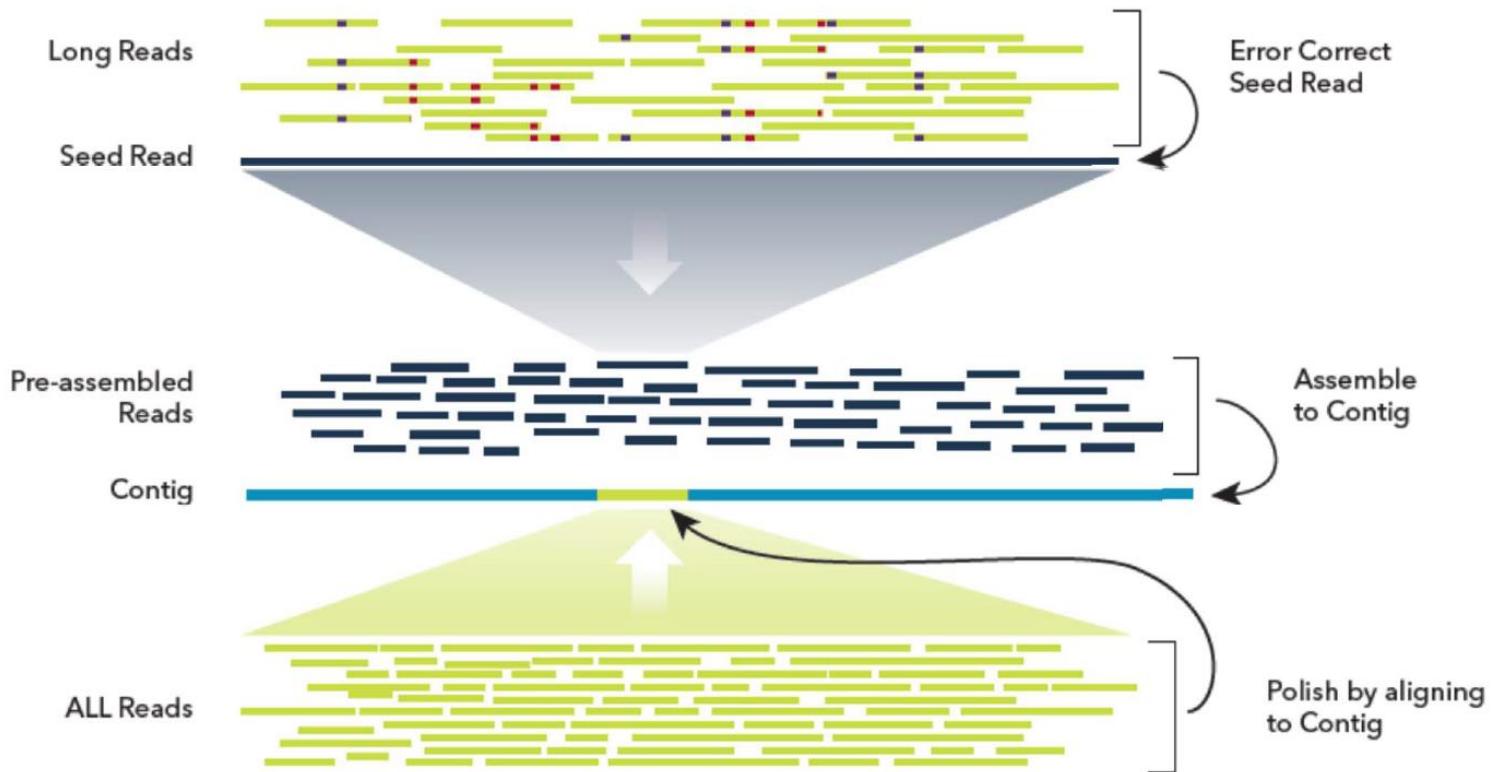
## Amplicons, Minor variant detection, Metagenomics



## SEQUEL SYSTEM V5.1 PERFORMANCE: HG00733 LIBRARY



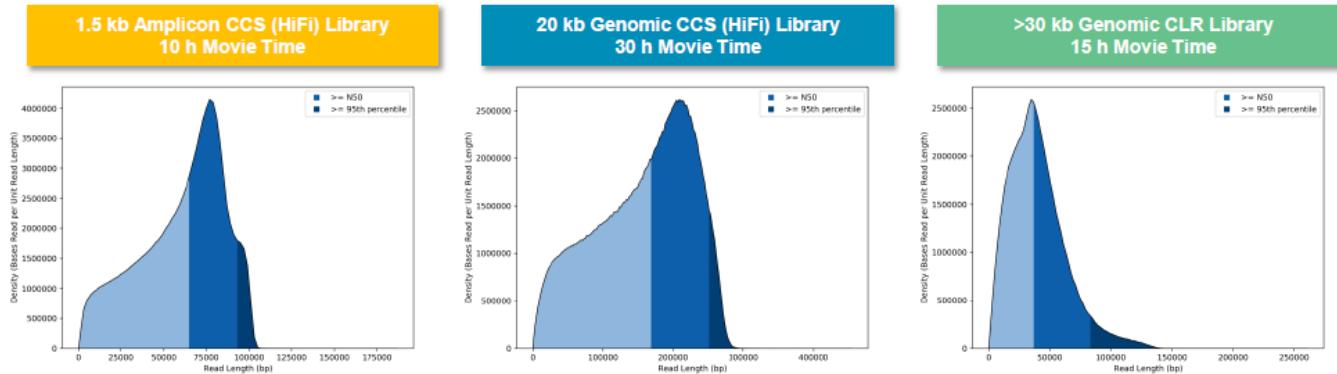
## HIERARCHICAL GENOME ASSEMBLY PROCESS (HGAP) & POLISHING



# Sequel II yields and base yield density

Sample Information >		Run Settings >	
Name	Movie Time (hrs)	Status	Total Bases (Gb)
1.5 kb Amplicon CCS (HiFi) Library	10	Complete	200.51
20 kb Genomic CCS (HiFi) Library	30	Complete	447.81
>30 kb Genomic CLR Library	15	Complete	129.97

## Base Yield Density

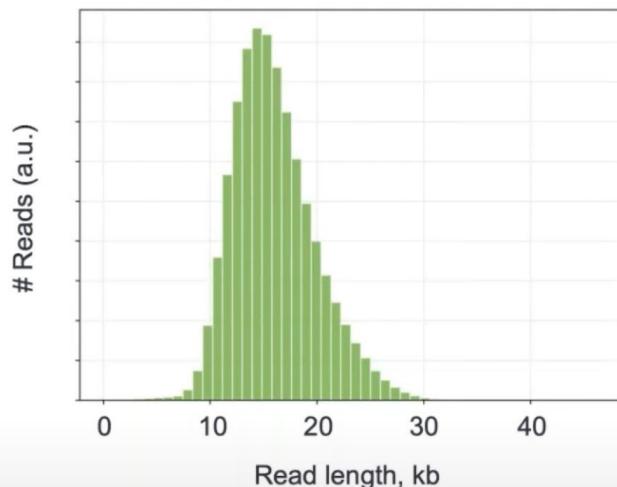




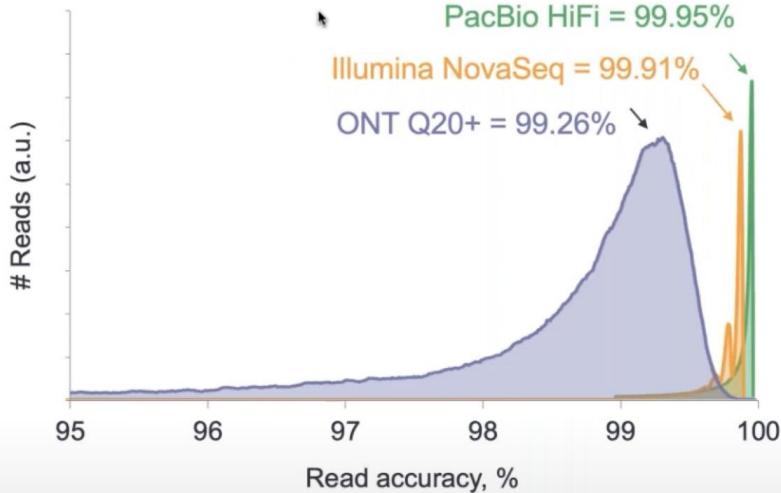
# PacBio HiFi

HiFi reads are long and accurate... over the 4-base DNA alphabet

15 – 25 kb read length



>99.9% read accuracy



GCCCAAGGCCAGGAGAGGGCAGGGCTTTCAAGCAGAGGGGGCTATTGGCCTACCTGGACTCTGTTCTCTCGCTGCTGCCCTCCAAATCAGGAGGTCTTGAAGCAGCTGCCCTACCCACAGGCCAGAAGTTCTGGTTCTCACCAAGAGAACATCAGCATTCCTGTCCTCCCTCCCACCTCCCTCTCCAGGGACAGTGAAGGTCAGGCCACACCCATGGAACCTGGAGGCCAGCAGCTGCTTGAGCCAACACGTGCAAGGCCAGCCTGGAGCCAGGGTCCCCACAGGATGAACACAGTAAGTTGGAGGGGAGGGGTCCTGAGGGACATTGGAGAGAAAAGGTGAGGGCTTCCGGGTGGCGTGCAGCTGAGACGCCCCTAGGGACTTCTGACAGAAGCAGACAGAAAACACAGAGAGACGAGGTACTTCAGACATGGAGCGCTCTGAGTTACAGTGGGCAATTAAAGTAAGGGTGTGTTGCTGGGATCTGAGAAGTCGATCTTGGACTGAGCTGAGCCTGTTGAAGGAGAAACAGCATGGAGAGGGCAAGGAAACAAACCCAGCAAGGCCCTTGTGAGGGATTAGGGAGCTGGAGGGATTGGAGCAGAGGGACATAGGTTGTTGAGCAGCAGGGCTGGAGGGCTGGAGAGGGCAAGGAAACAAACCCAGCAAGGCCCTTGTGAGGG...

PacBio HiFi: HG003 18 kb library, Sequel II System Chemistry 2.0, [precisionFDA Truth Challenge V2](#)

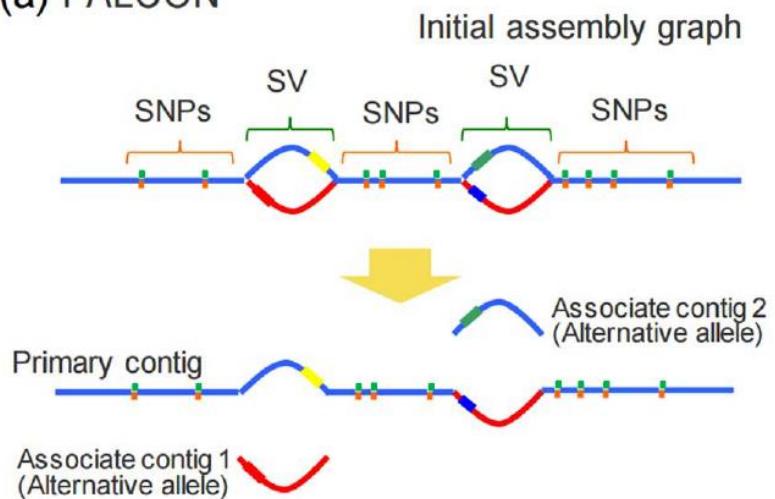
Illumina: HG002 2×150 bp NovaSeq library, [precisionFDA Truth Challenge V2](#)

ONT: Q20+ chemistry (R10.4, Kit 12), Oct 2021 GM24385 Q20+ Simplex Dataset Release

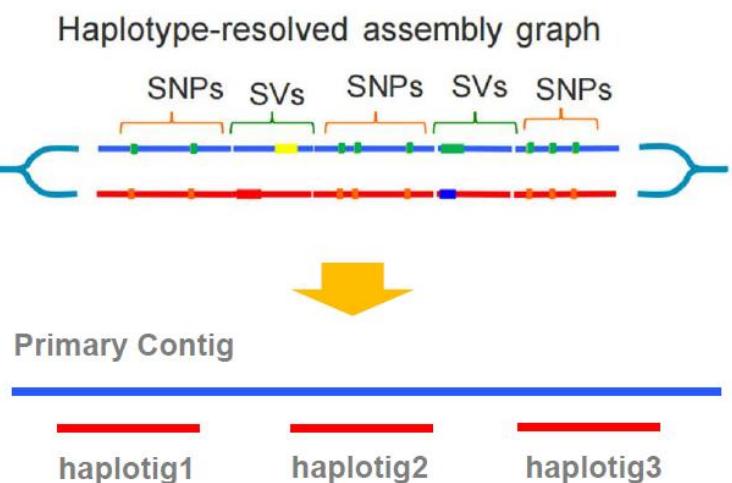


# DIPLOID ASSEMBLY WITH FALCON-UNZIP

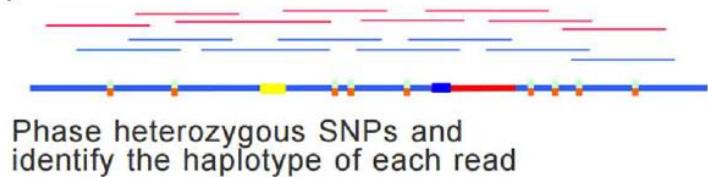
(a) FALCON



(c) FALCON-Unzip



(b)



Long read can access 193 medically relevant genes that short read cannot

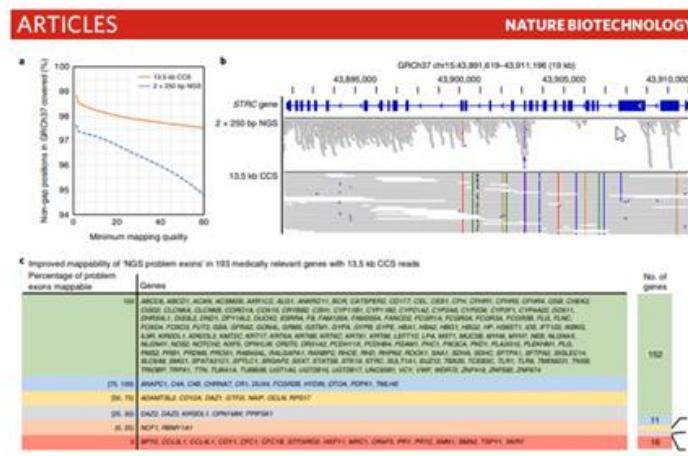
## PacBio Claims Differentiation Through Ability to Address Challenging and Medically Relevant Genomic Regions



## Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome

Aaron M. Wenger<sup>✉1,4</sup>, Paul Peluso<sup>1,4</sup>, William J. Rowell<sup>3</sup>, Pi-Chuan Chang<sup>2</sup>, Richard J. Hall<sup>1</sup>, Gregory T. Concepcion<sup>1</sup>, Jana Ebler<sup>1,4</sup>, Arkarachai Fungtammasan<sup>1</sup>, Alexey Kolesnikov<sup>4</sup>, Nathan D. Olson<sup>2</sup>, Armin Töpfer<sup>2</sup>, Michael Alongé<sup>2</sup>, Medhat Mahmoud<sup>1</sup>, Yufeng Qian<sup>1</sup>, Chen-Shan Chin<sup>2</sup>, Adam M. Phillippe<sup>1</sup>, Michael C. Schatz<sup>2</sup>, Gene Myers<sup>1</sup>, Mark A. DePristo<sup>1</sup>, Jue Ruan<sup>1,2</sup>, Tobias Marschall<sup>1,4</sup>, Fritz J. Sedlacek<sup>3,8</sup>, Justin M. Zook<sup>1</sup>, Heng Li<sup>1,3</sup>, Serzey Konor<sup>1</sup>, Andrew Carroll<sup>1</sup>, David R. Rankin<sup>1,4</sup> & Michael W. Hunkapiller<sup>1,4\*</sup>

The DNA sequencing technologies in use today produce either highly accurate short reads or less-accurate long reads. We report the optimization of circular consensus sequencing (CCS) to improve the accuracy of single-molecule real-time (SMRT) sequencing (PacBio) and generate highly accurate (99.8%) long high-quality (HQ) reads with an average length of 13.5 kilobases (kb). CCS also minimizes the bias in the whole genome assembly by increasing the number of CCS reads generated per base pair and read length at least 99 times for single-nucleotide variants (SNVs), 95.98% for insertions and deletions (INDELs) and 95.99% for structural variants. Our CCS method matches or exceeds the ability of short-read sequencing to detect small variants and structural variants. We estimate that 2,434 discrepancies are correctable mistakes in the genome in a botany (GRCh38) benchmark set. Near 99.4% variation was phased into haplotypes, further improving variant detection. Despite generating only 10% of the CCS reads, our CCS method generated a complete genome assembly with near 95.0% of 15 megabase (Mb) and concordance of 99.99%, substantially outperforming assembly with less-accurate long reads.



**Fig. 2 | Mappability of the human genome with CCS reads.** a, Percentage of the nongap GRCh37 human genome covered by at least ten reads from 28-fold coverage NGS ( $2 \times 250$  bp, Hs65250) and CCS ( $13.5$  kb) datasets at different mapping quality thresholds. b, Coverage of the congenital deafness gene STRC in HGO02 with  $2 \times 250$  bp NGS reads and  $13.5$  kb CCS reads at a mapping quality threshold of  $10 \pm 4$ . Improvement in mappability with  $13.5$  kb CCS reads for 193 human genes previously reported as medically relevant and problematic to map with NGS reads<sup>12</sup>.



## INTRONIC INSERTION IN AUTS2 (AUTISM)



PacBio reads  
show insertions

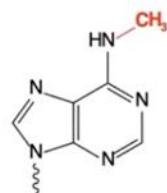
Illumina reads  
do not map

Tandem repeat

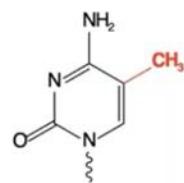
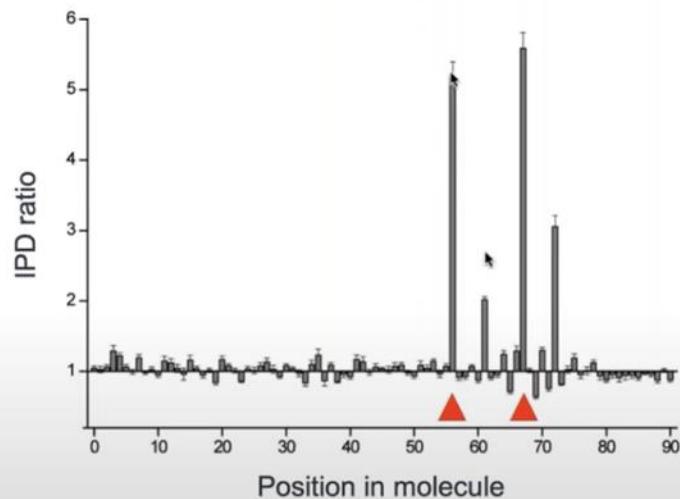


# Epigenetics

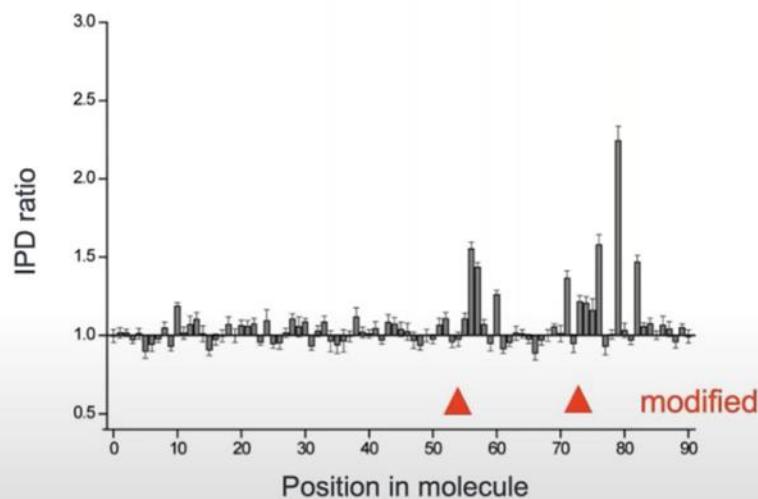
Different modifications have distinct kinetic signatures



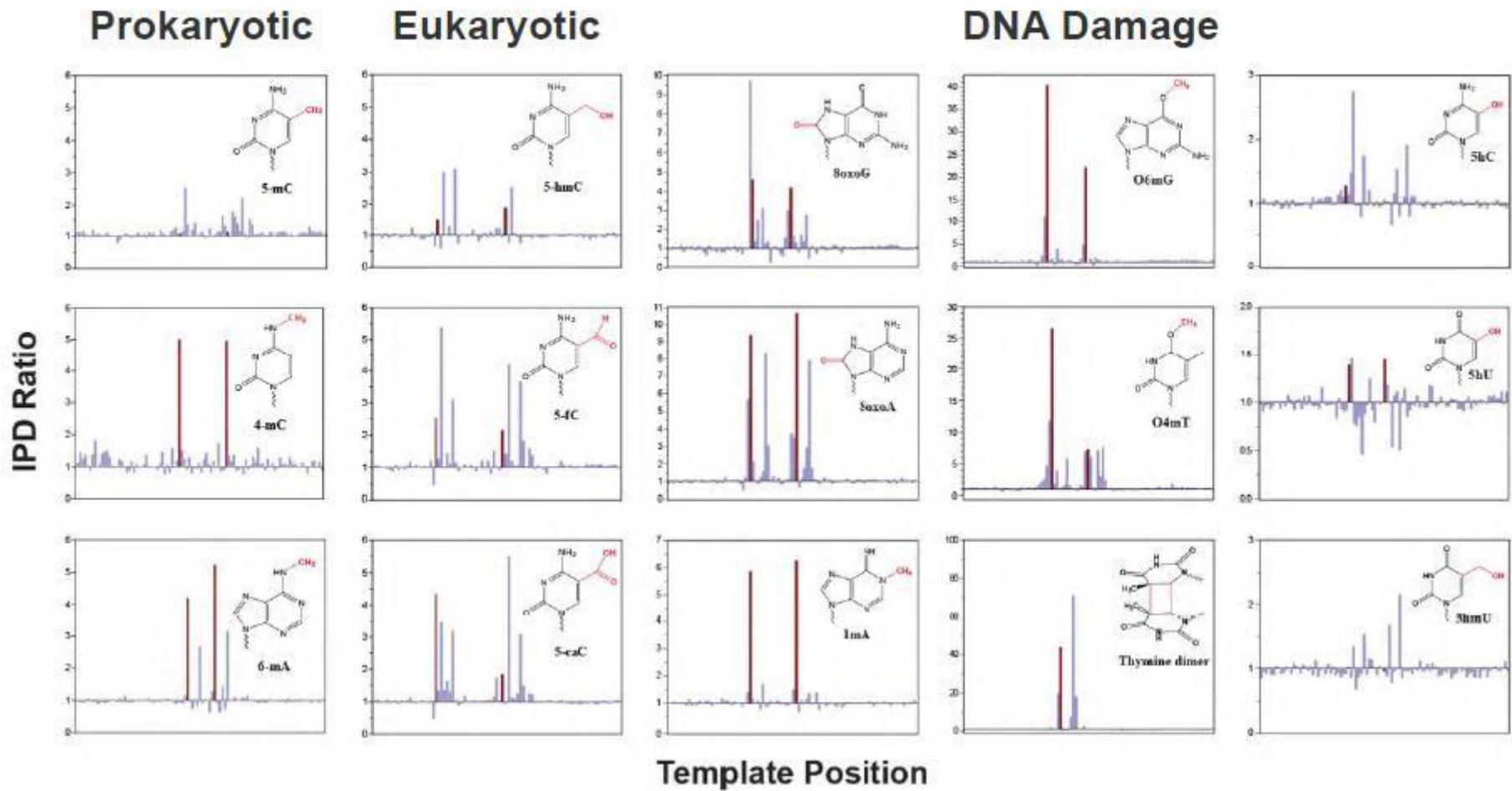
6mA  
strong, localized signal



5mC  
diffuse signal



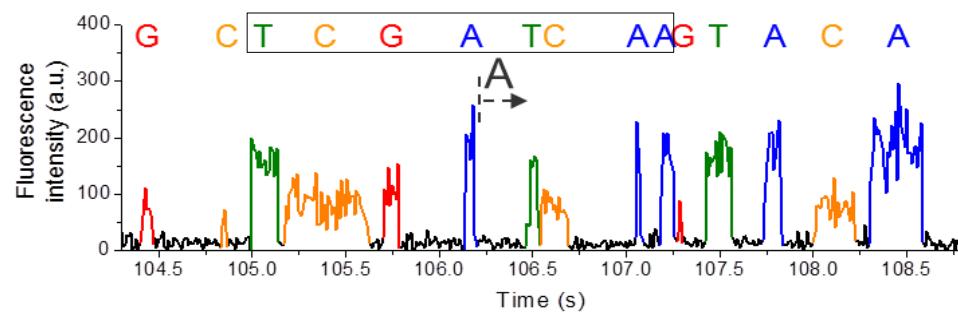
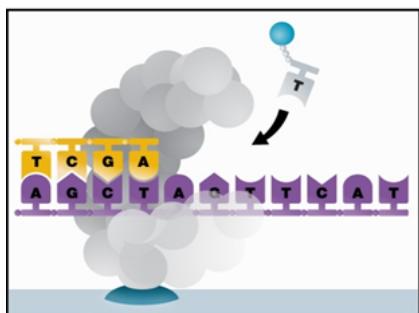
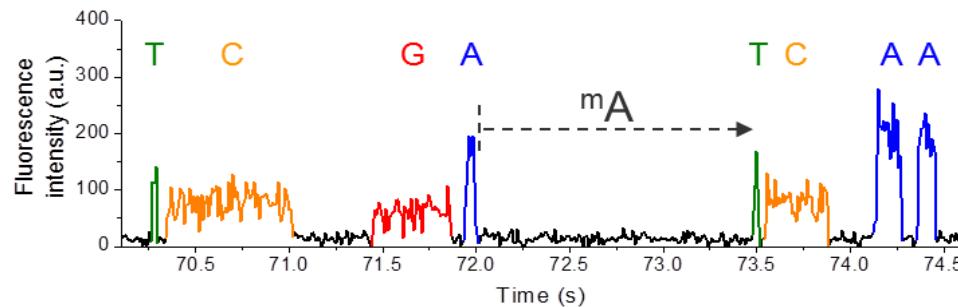
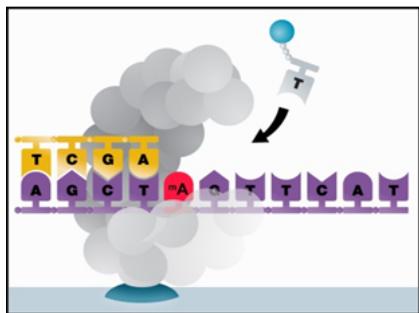
# Base Modifications





# DETECTION OF DNA BASE MODIFICATIONS USING KINETICS

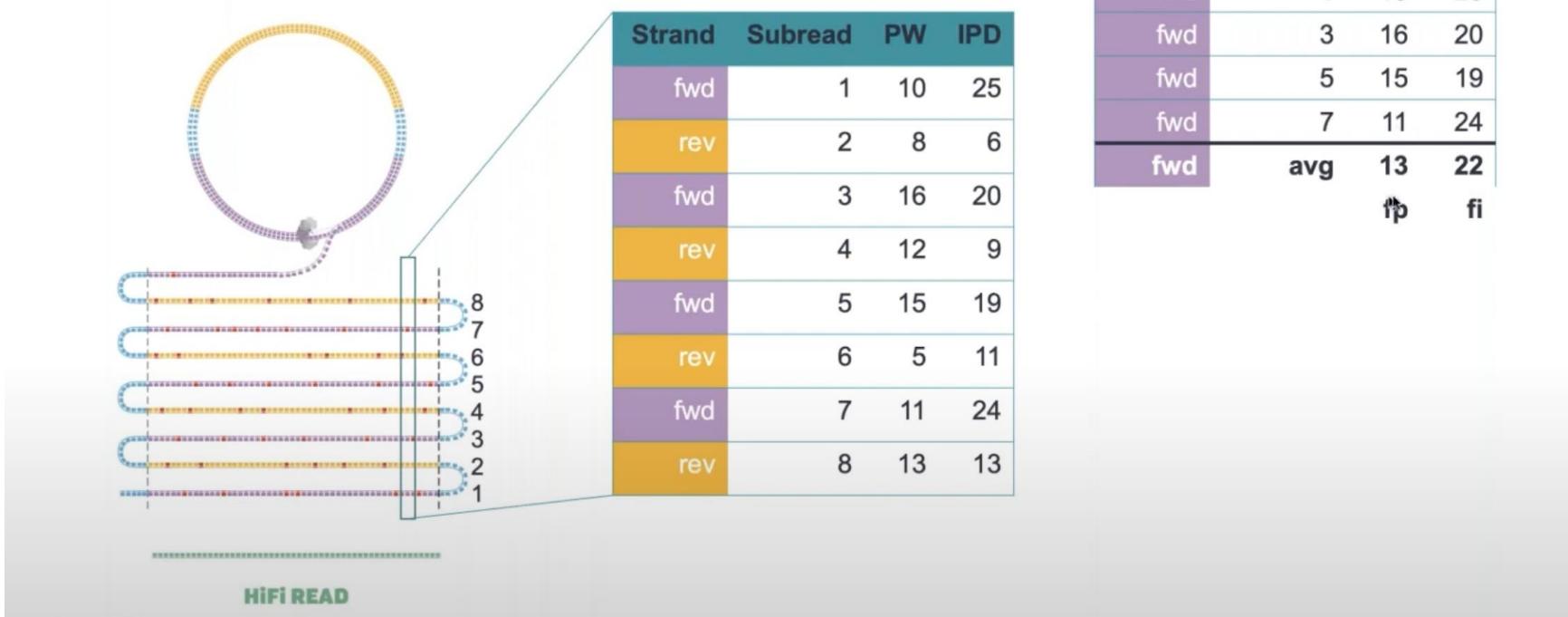
Example: N<sup>6</sup>-methyladenine



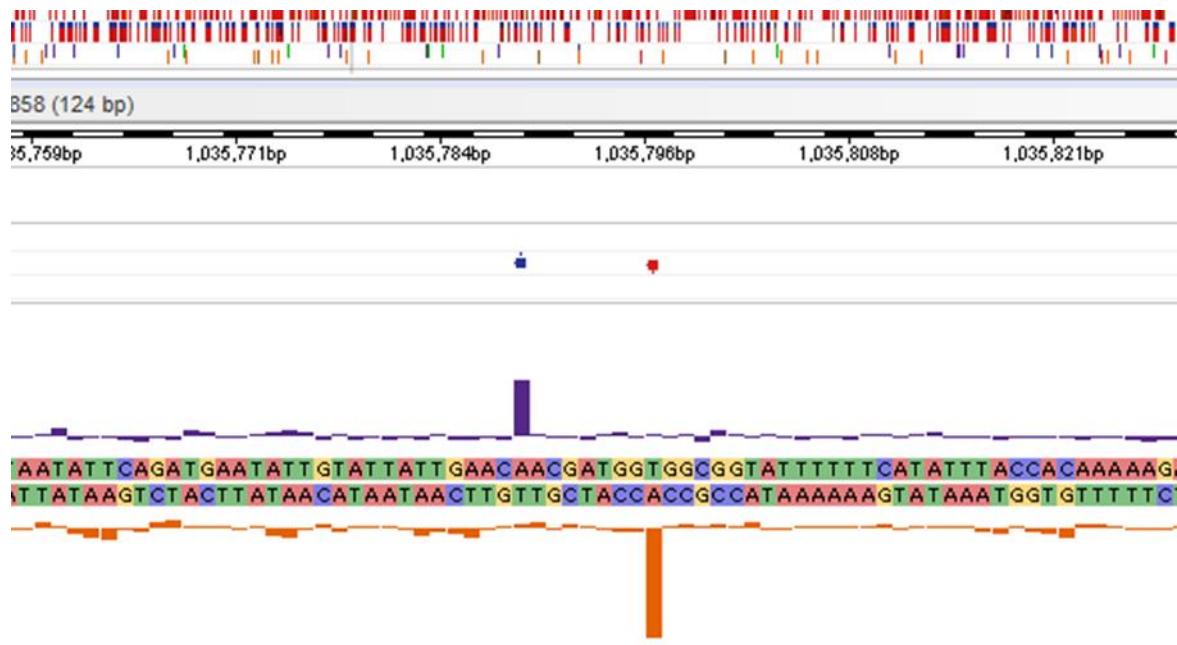
- SMRT Sequencing uses kinetic information from each nucleotide addition to call bases
- This same information can be used to distinguish modified and native bases by comparing results of SMRT Sequencing to an *in silico* kinetic reference for incorporation dynamics without modifications.

## Single-molecule methylation calling with HiFi reads

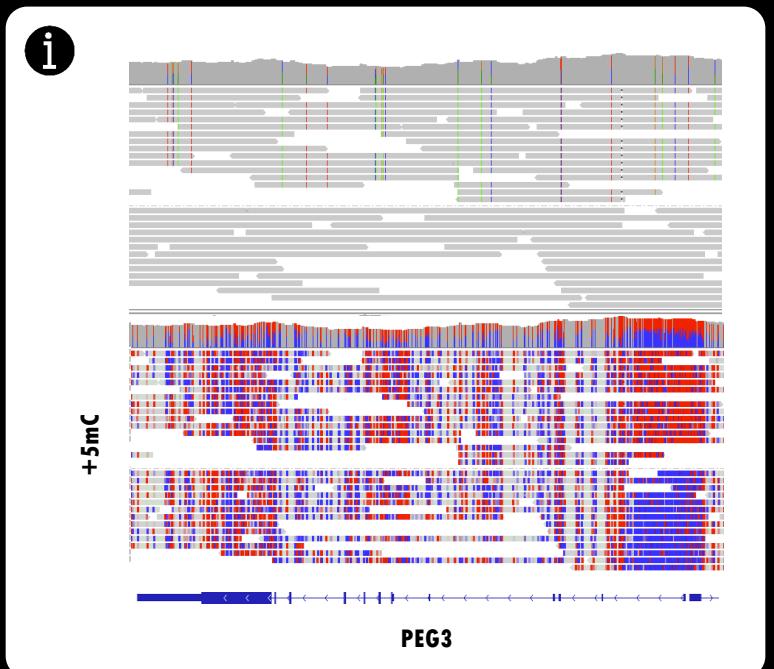
ccs software outputs summary kinetics as average of subread kinetics



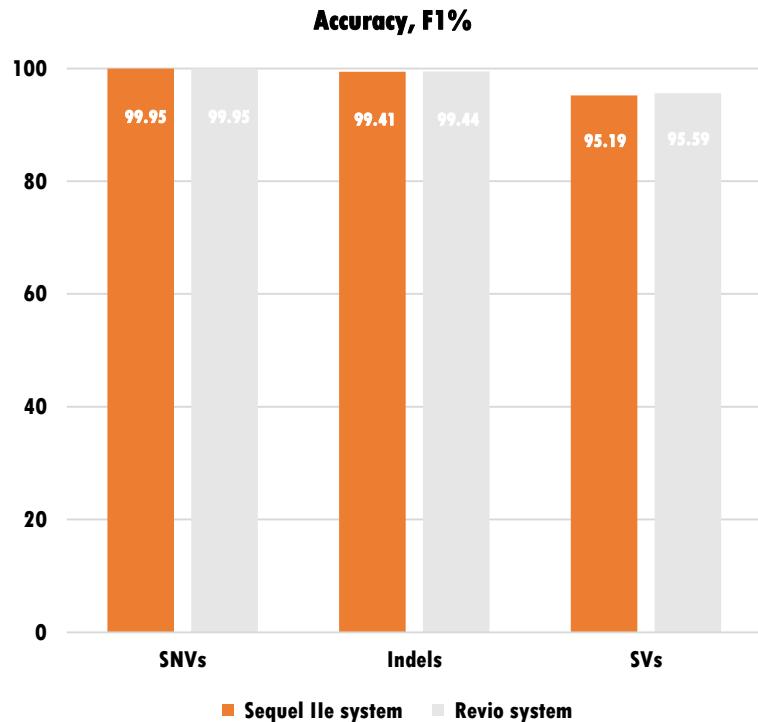
# An example of signal strength with m6A



# Revio variant calling and methylation performance



PacBio



HG002 at maternally imprinted PEG3 locus  
Data for internal beta testing [https://downloads.pacbcloud.com/public/revio/2022Q4/HG002-rep3/analysis/HG002.m84005\\_220827\\_014912\\_s1.GRCh38.bam](https://downloads.pacbcloud.com/public/revio/2022Q4/HG002-rep3/analysis/HG002.m84005_220827_014912_s1.GRCh38.bam)  
Variant calling measured against Genome in a Bottle benchmarks, HG002

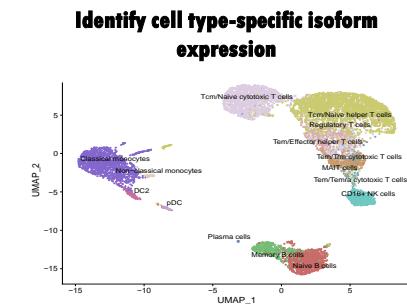
## • Extending MAS-Seq technology to bulk Iso-Seq and 16S rRNA

New MAS-Seq kits to support bulk Iso-Seq and 16S rRNA amplicons on the Sequel II/Ile and Revio systems



### MAS-Seq for 10x Single Cell

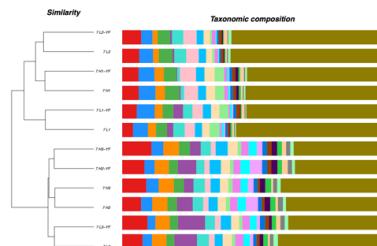
- ✓ Launched Oct 2022
- ✓ 16-fold concatenation
- ✓ 40 million reads (Sequel II/Ile)
- ✓ 80 million reads (Revio)



### MAS-Seq for 16S rRNA

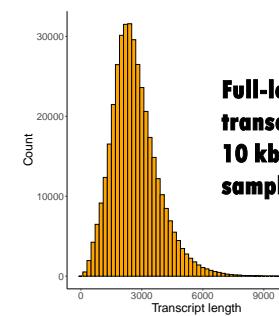
- Expected H2 2023
- 12-fold concatenation
- Support on Sequel II/Ile and Revio with multiplexing

### Taxonomic profiling for human, animal, and environmental samples



### MAS-Seq for bulk Iso-Seq

- Expected H2 2023
- 6-fold concatenation
- Support on Sequel II/Ile and Revio with multiplexing





# Pacific BioSciences

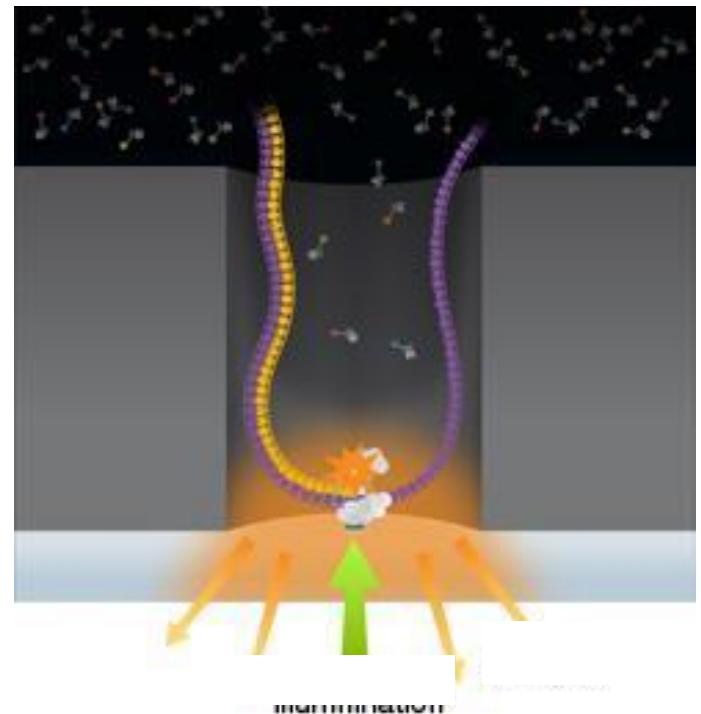
## Applications

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



# Pacific Biosciences

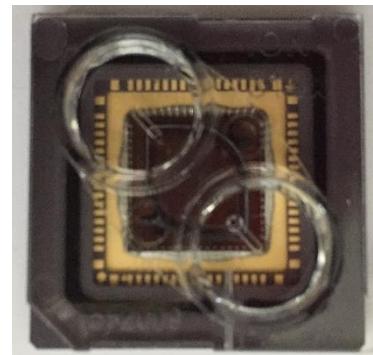
- » Single polymerase mol. in a 20nm hole
- » Watch incorporation in real time
- » ~2 bases per second
- » Yield 100Gb
- » ~\$10/Gb
- » Some reads 100Kb +



# Ion Torrent's PGM



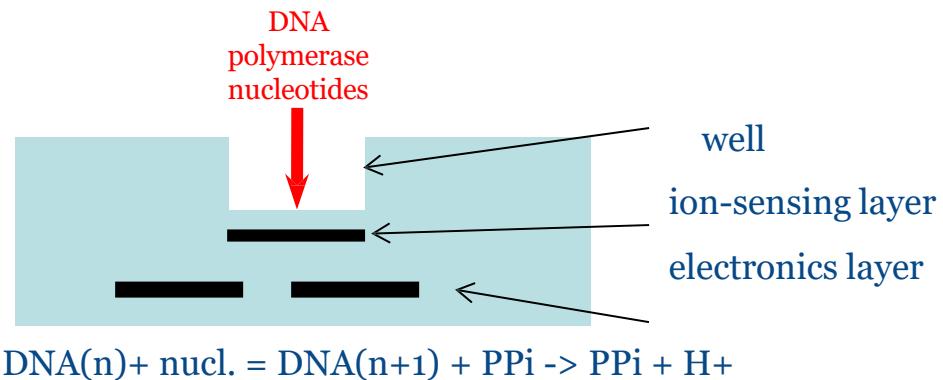
Capital cost \$50,000



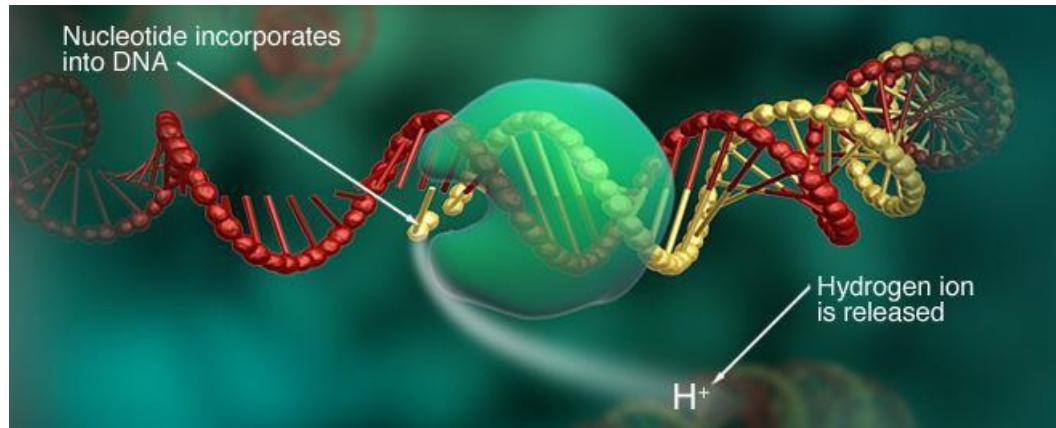
# Ion Torrent

- » Similar to 454 but detects H<sup>+</sup> released as a base is added
- » Prone to errors near homopolymers
- » Not good for whole genome sequencing but useful for targeted sequencing as run times are short
- » Used in a lot of clinical settings for disease panel sequencing

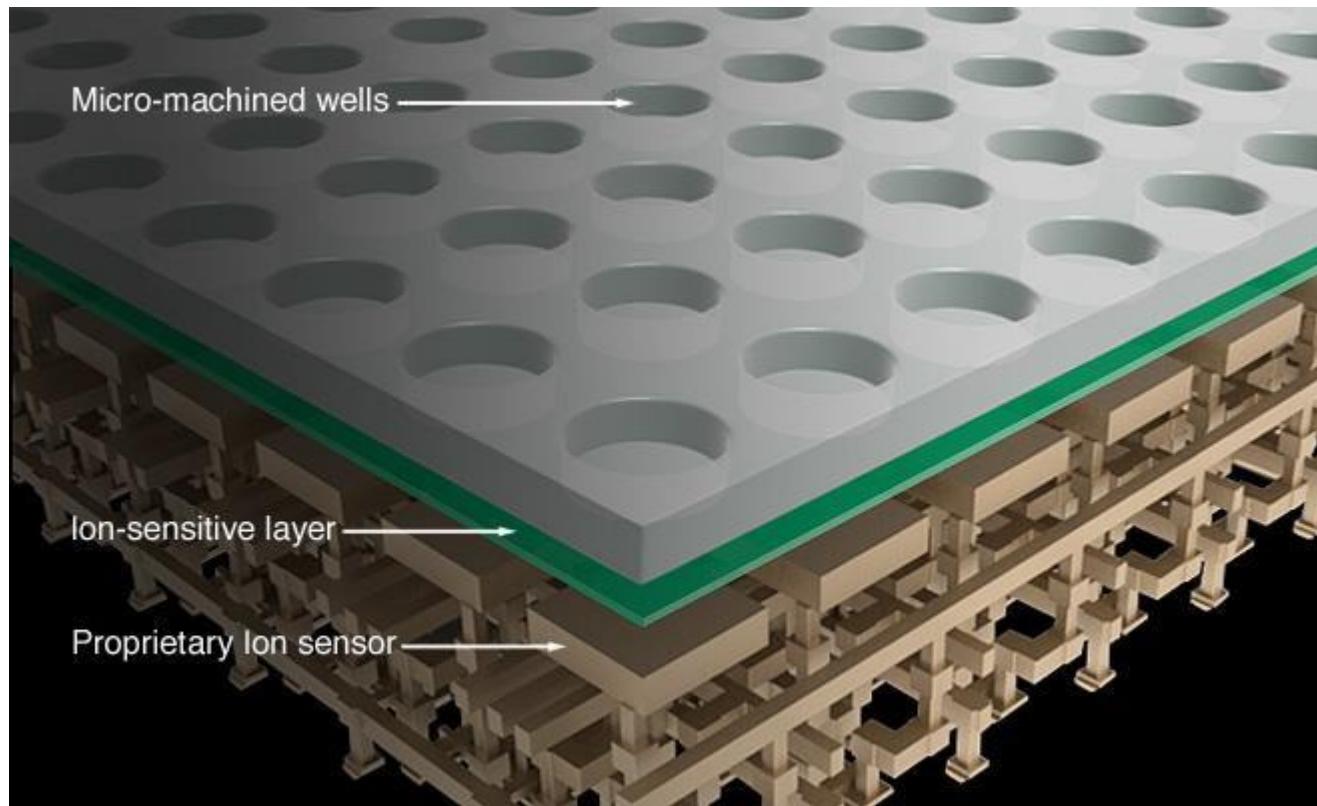
# Ion Torrent's Technology



Flow through A then T then ... like 454



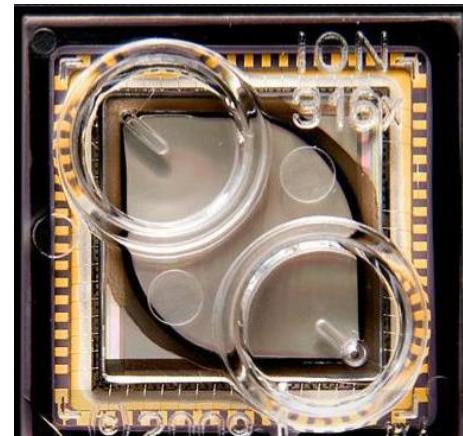
# Ion Torrent Chip Drawing





# Ion Torrent

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





## Ion 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.

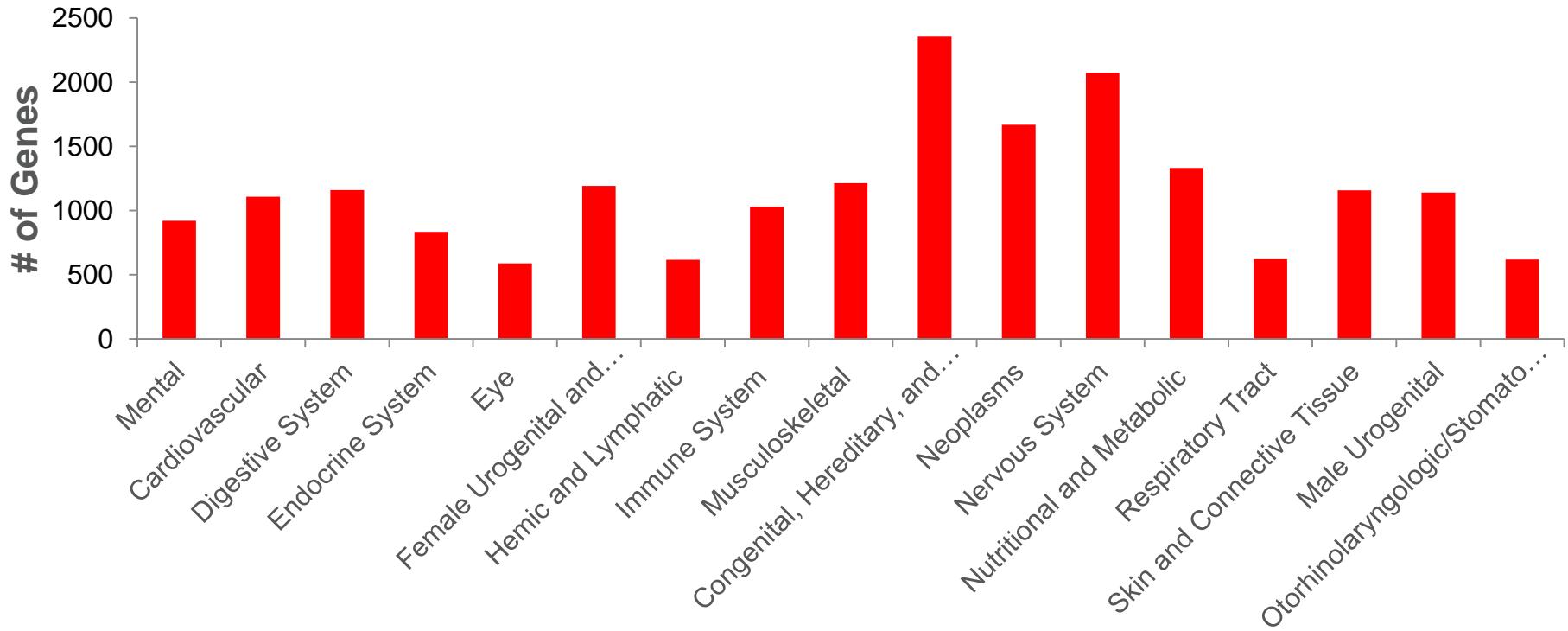
**2–3 hour prep time**  
Shorter turn-around time

**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



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



<b>Speed*</b>	19 hrs	10 hrs	6.5 hrs
<b>Output (max/day):</b>	15 Gb/80 M	30 Gb/160 M	50 Gb/260 M
<b>Chips (max/day):</b>	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.

## Genexus System—Tomorrow's Specimen-to-Report NGS Workflow

### Genexus Software

\$299k  
12-15M reads

- FFPE tissue
- Frozen tissue
- Bone marrow
- Whole blood
- PBL
- Urine
- Saliva

#### Nucleic acid purification and quantitation\*

**Ion Torrent™ Genexus™**  
Purification System (Available 2020).



Up to 32 FFPE tissue samples with DNA OR RNA only input

2 hr TAT

#### Library preparation to variant interpretation

**Ion Torrent™ Genexus™**  
Integrated Sequencer (Available Nov 2019)

**Ion Torrent™ GX5™ Chip:**  
12–15M reads/lane



14 hours for a single-lane run (approx. 24 to 30 hours for full chip)  
Up to 32 Samples per run

Report\*

- \* Specimen-to-report workflow available after Ion Torrent Genexus Purification System launches in 2020.  
The content provided herein may relate to products that have not been officially released and is subject to change without notice.

## 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.

## Oncomine Precision Assay on Ion Torrent Genexus System

Maximizes your ability to detect relevant variants

### Curated pan-cancer content



- Mutations, CNVs, and fusion variant types across 50 key genes
- Tumor suppressors, drivers, and resistance variants

### Tissue and plasma samples



- One test, one workflow, multiple sample types
- Maximizes the number of tumors that can be profiled

### Molecular tagging



- Enhanced low-level variant detection
- Key for liquid biopsy testing

### FusionSync™ Detection Technology



- Sensitive and specific—targeted isoform designs
- Novel fusion detection

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



# Nanopores



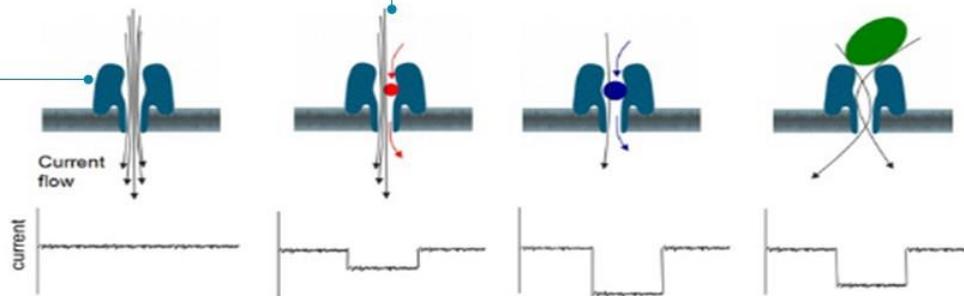
# Oxford nanopore

- » Long read sequencing
- » Sequences single molecules of DNA as travel through pore
- » Reads typically 5-50kb but some much longer (4Mb record)
- » Fast moving technology that may soon be cheaper and have higher throughput than Illumina
- » Zero capital cost price structures available

## 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

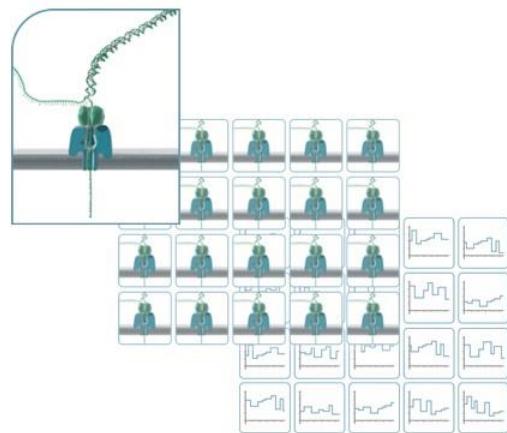
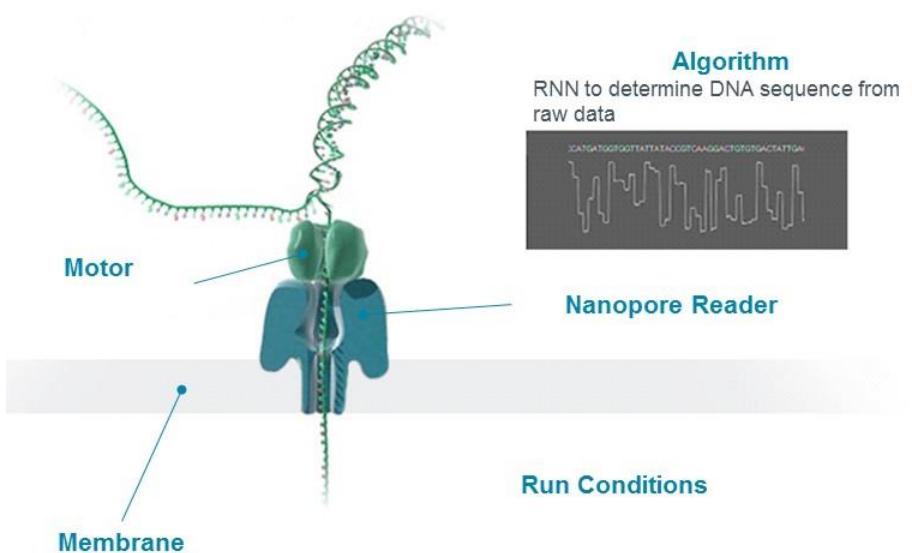
3

Current disruption is interpreted to understand the  
identity of the analyte

An explanation from  Oxford  
**NANOPORE**  
Technologies™

## NANOPORE DNA SEQUENCING

How does it work?



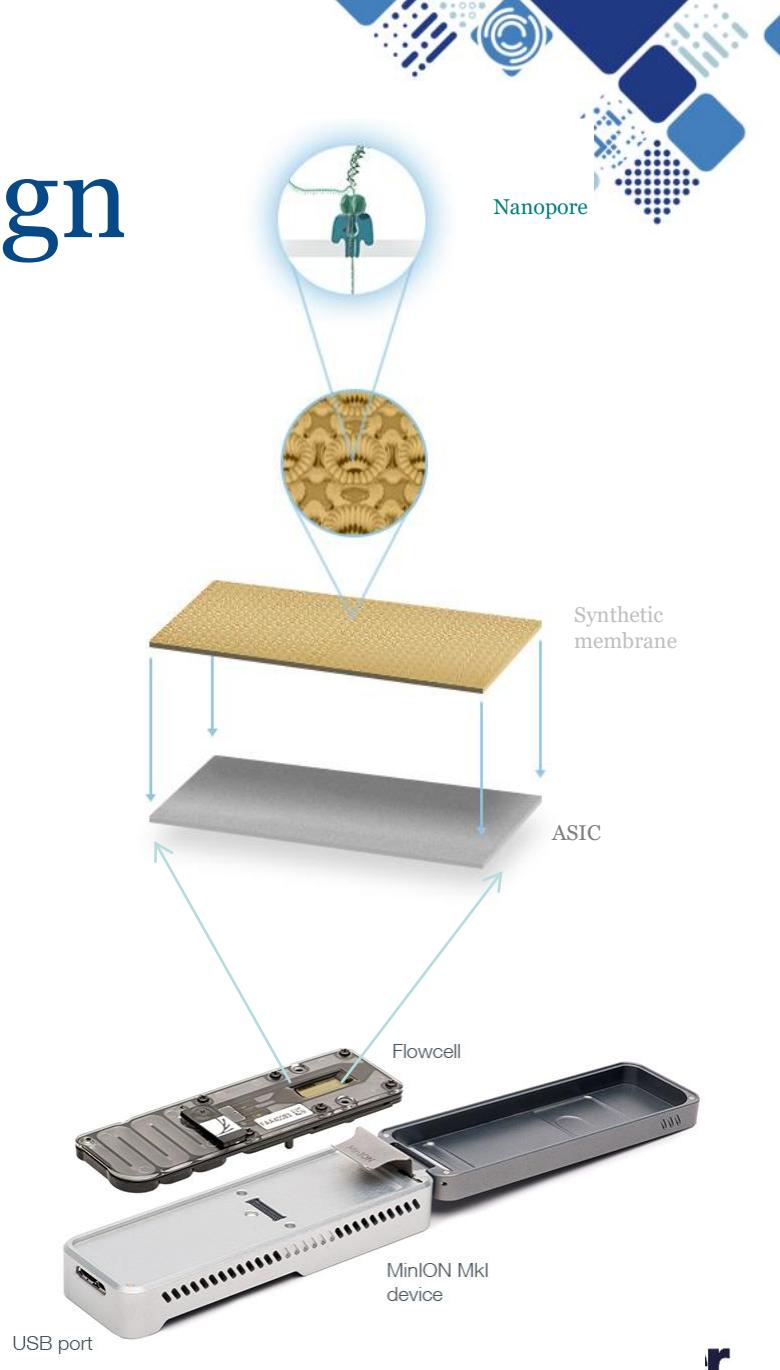
Multiple nanopore sensors arrayed in one device

Operate independently but at the same time

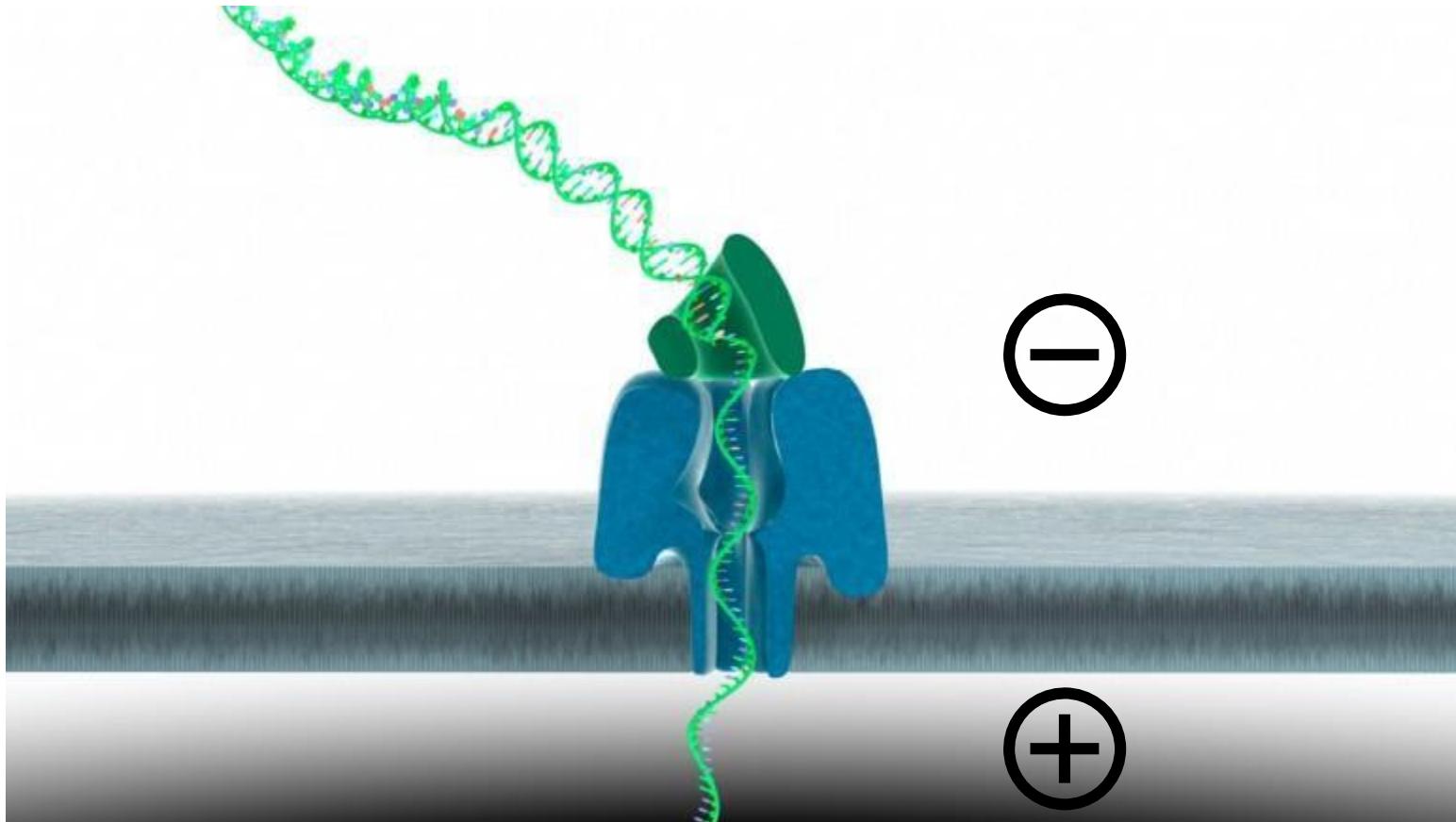
An explanation from  Oxford **NANOPORE** Technologies™

# Flow cell design

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

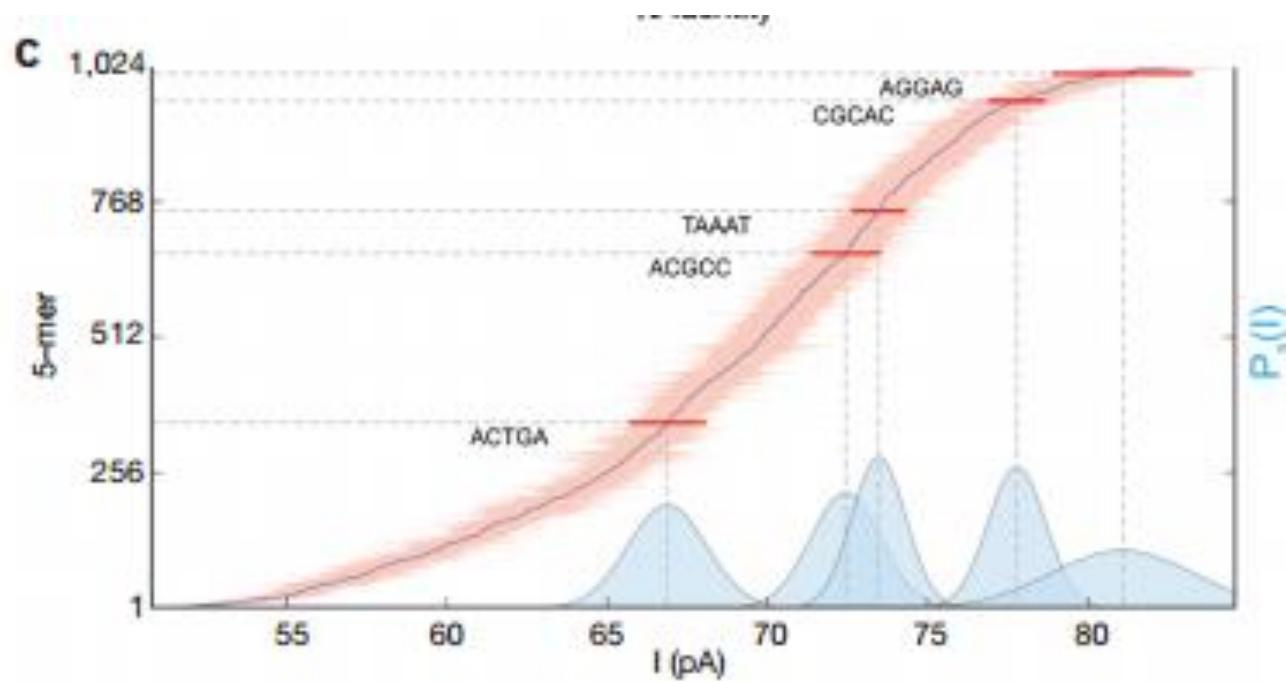


# Oxford Nanopore Technologies



Crudely 5-base words = 1024 current levels

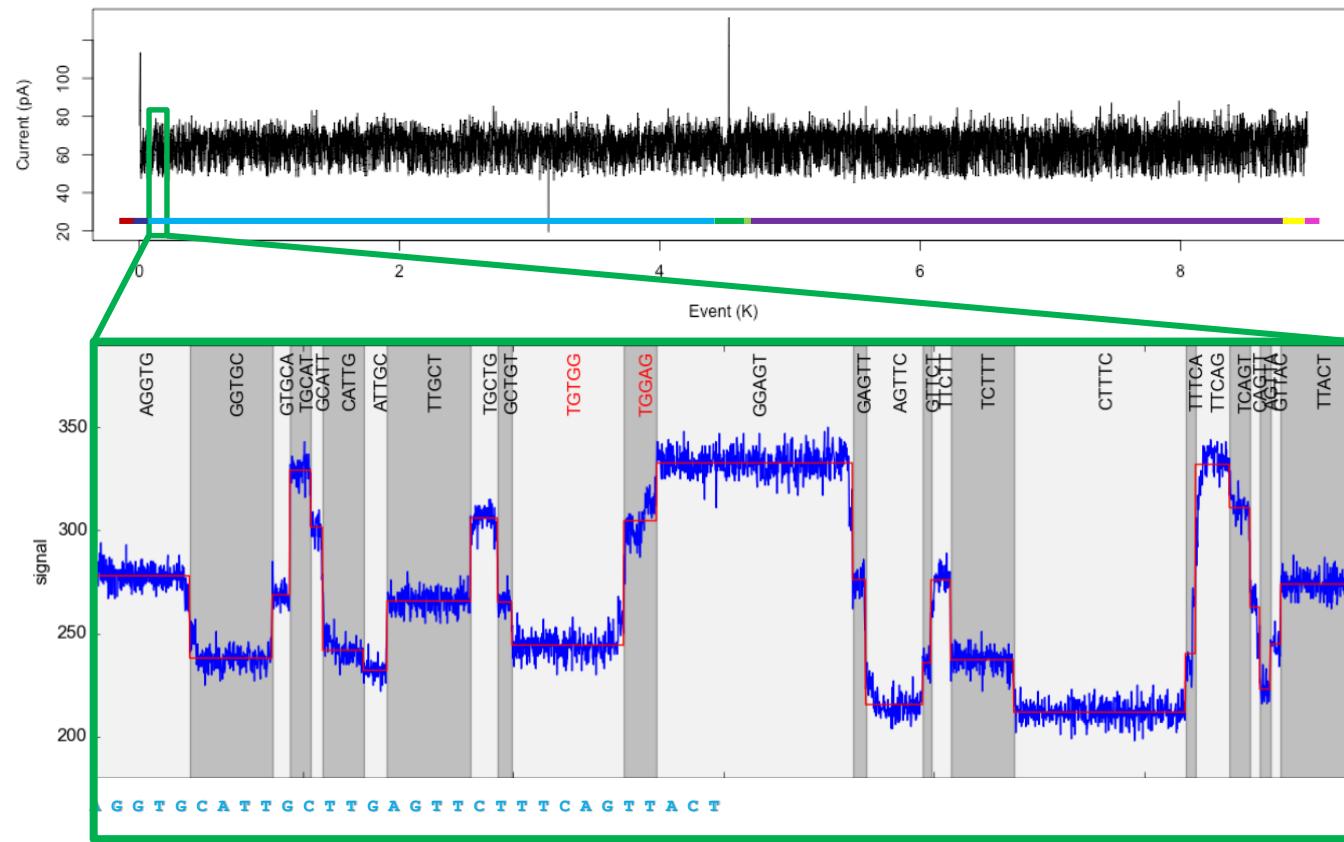
# Assuming 5mer current signals



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

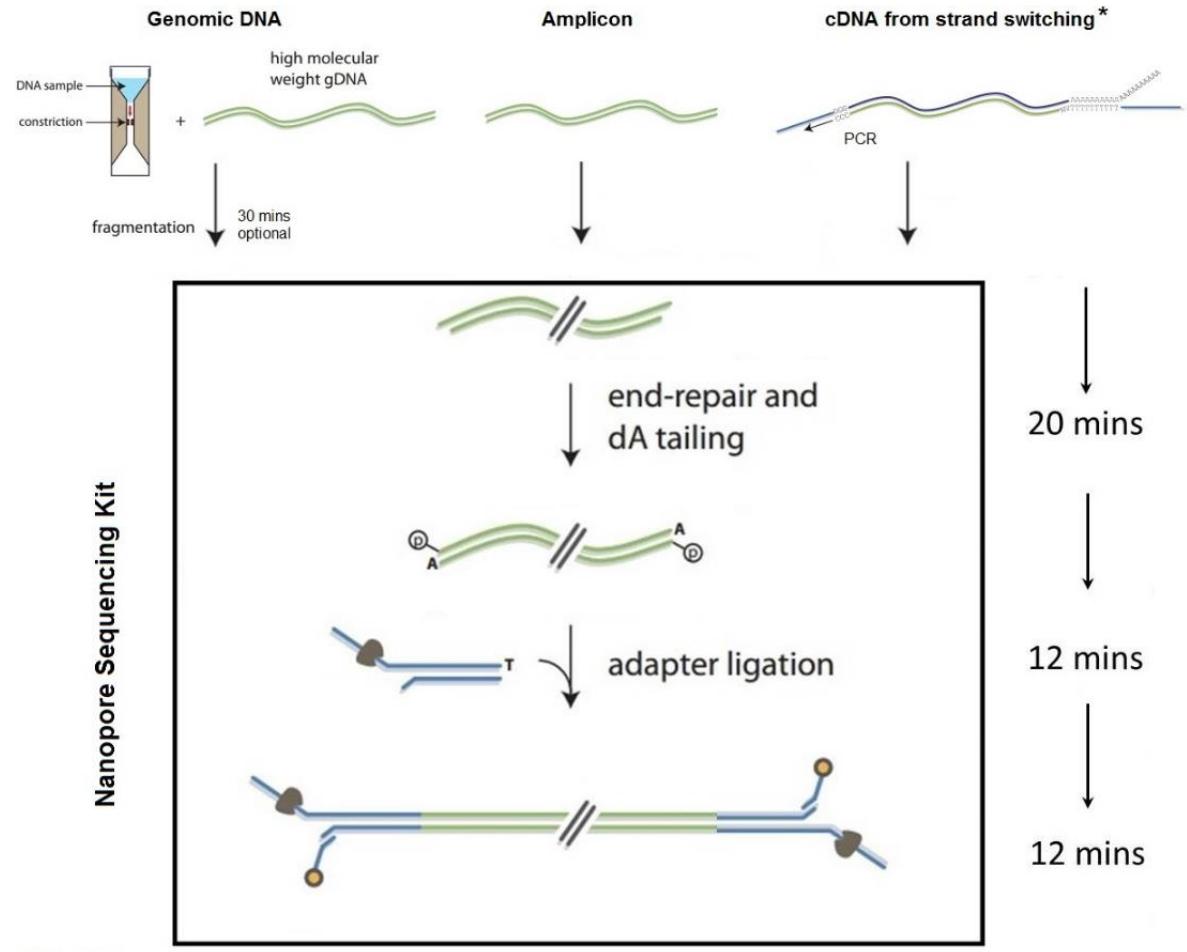


# ONT: The squiggles



Slide courtesy of David Buck. WTCGH

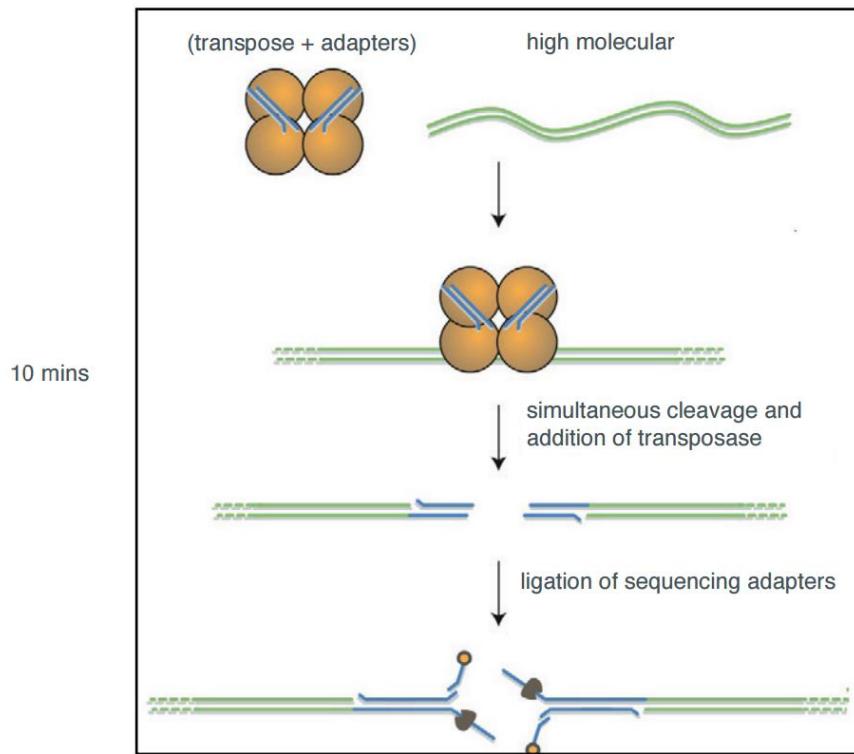
# ONT 1D Library prep





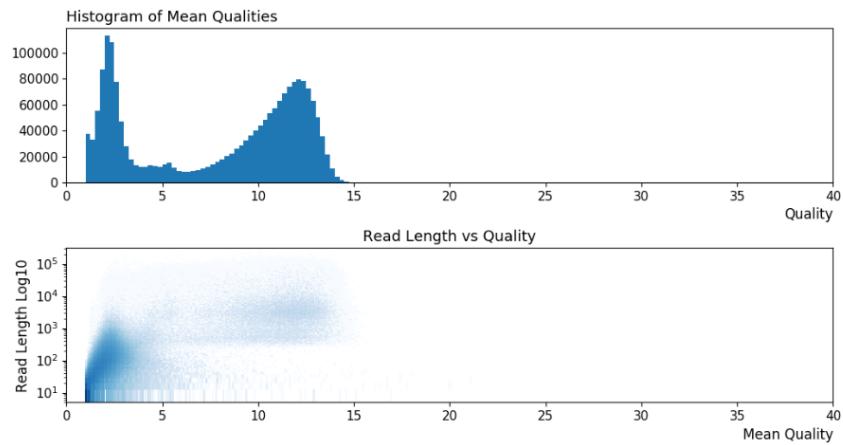
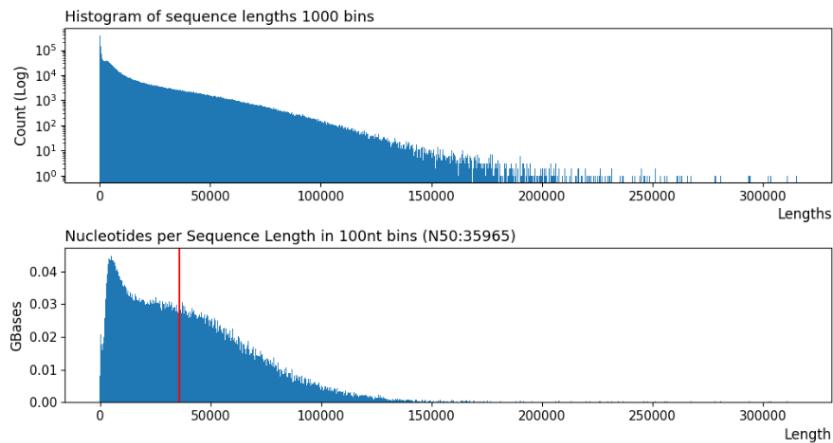
## RAPID SEQUENCING KIT

A two-step, 10 minute protocol



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

# ONT Sequencing performance graphs





# Data Quality

Single read raw error rate 1-2%

Read lengths as long as template. Record >4Mb  
but shorter fragments give higher yield

Yield per flowcell dependent on DNA

Flongle upto 3Gb

MinION 1-30Gb

PromethION 10-200Gb

R9.4 gives errors around homopolymers >5 bases

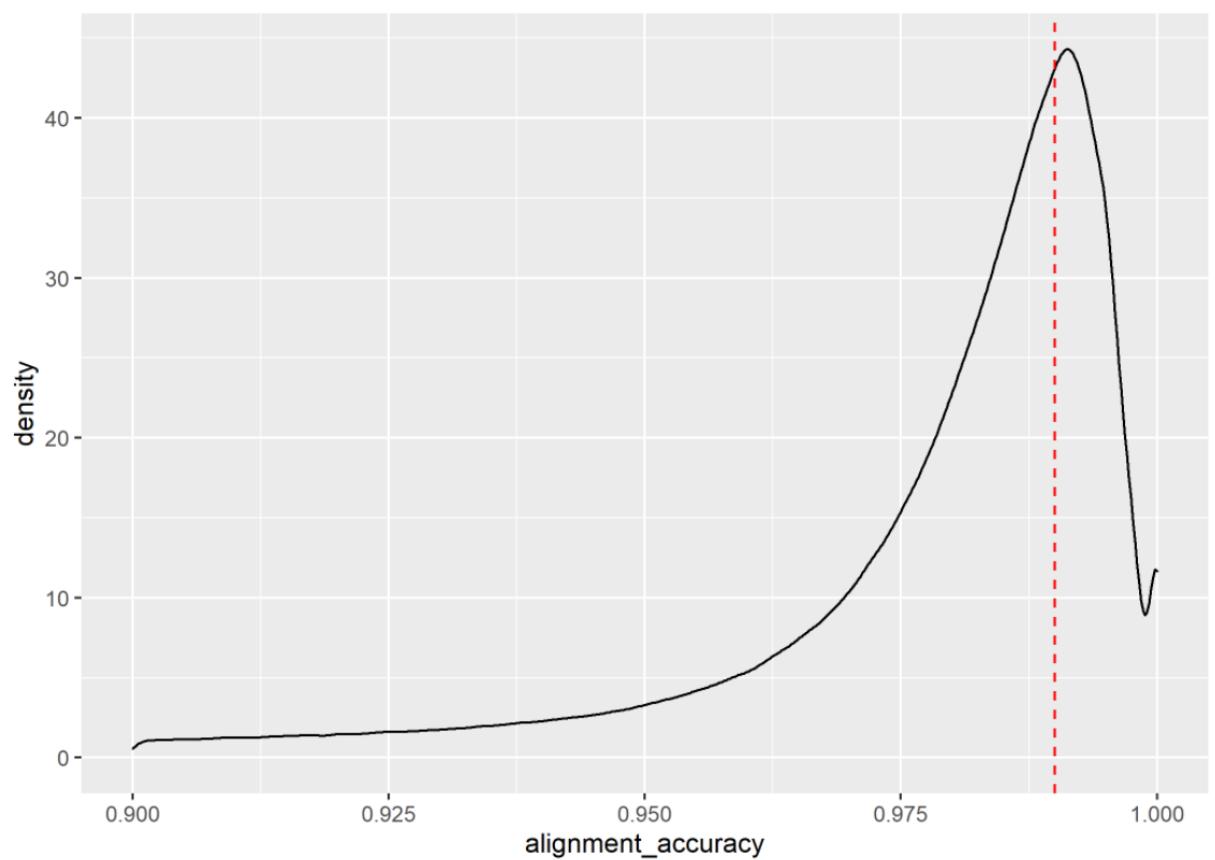
R10 deals with homopolymers much better



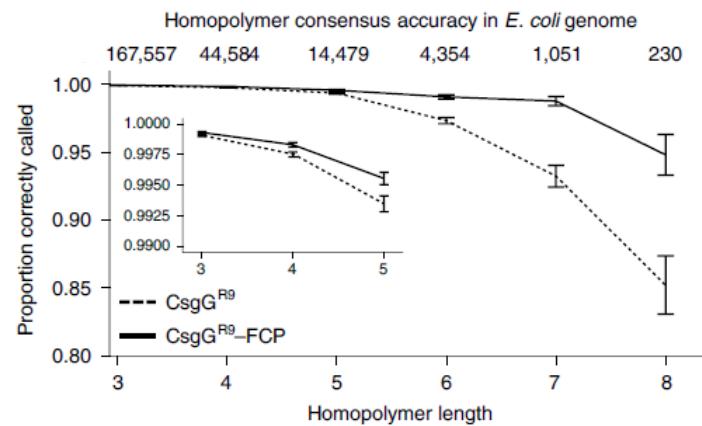
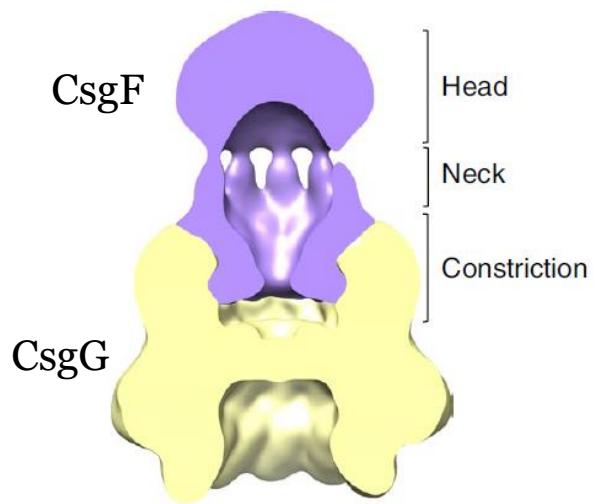
# ONT accuracy

Chemistry	Raw read accuracy (modal)	Analytical tools	Sample
R9.4.1	98.3%	Production software MinkNOW 4.3 ("Super-accuracy" basecalling model), Guppy 5	Mixed genomes
R10.3	97.5%	Production software MinkNOW 4.3 ("Super-accuracy" basecalling model), Guppy 5	Mixed genomes
"Q20+"	> 99.3% modal	Research algorithm <u>Bonito</u>	<b>Various</b> Note this chemistry is in early access phase

# Q20 accuracy with Bonito



# R10. A dual head nanopore to help resolve homopolymers



From Van der Verren et al., Nature Biotech 2020

# New pore for 2022

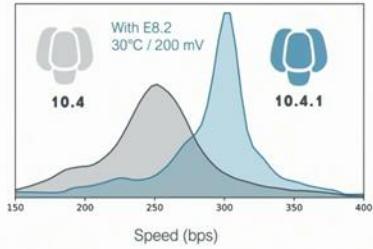
# R9 being discontinued

## Nanopore Accuracy

Chemistry update - motor and pore improvements

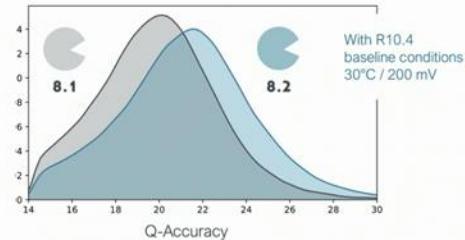
### New pore: R10.4.1

- Improved pore designs tune enzyme-pore docking
- Faster speeds (~250-420 bps)
  - Yield much higher output compared to current Q20 chemistry
- Tighter speed distributions
  - Helps to reduce errors



### New motor: E8.2

- Improved movement properties with more consistent movement
  - Better defined levels
  - Fewer mis-steps
  - Improved accuracy
- ~Halved error rates.



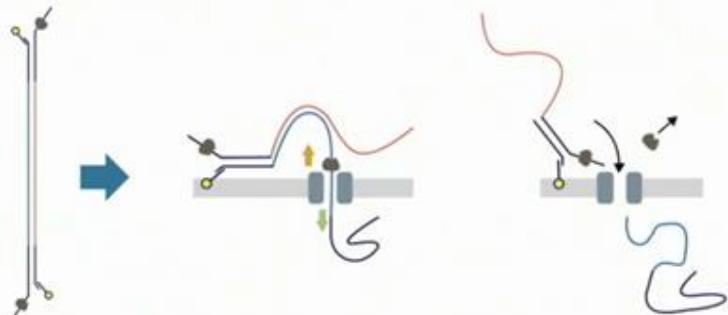


## Nanopore Accuracy

Duplex – reading both strands

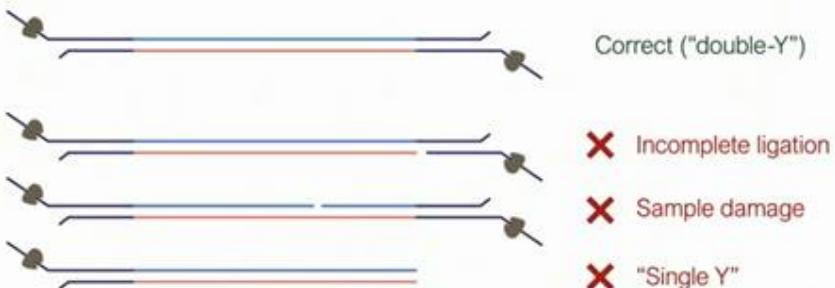
### Combining data from both strands

- Reading both strands has always been a strong feature
  - Two looks at the same sequence pairs
  - Different error profiles on reverse complement
  - Can help resolve modifications
- Previous versions of the chemistry had low natural “follow on”
- Typical rates of ~ 5% of the data collected was from Duplex pairs



18

© 2022 Oxford Nanopore Technologies plc.  
Oxford Nanopore Technologies products are not intended for use for health assessment or to diagnose, treat, mitigate, cure, or prevent any disease or condition.



### Combining data from both strands

- Chance of capturing complement after a template can be high
  - Need to optimise the setup to increase your odds
- Things that help:
  - DNA must have adapters at both ends
  - Fully adapted, repair any breaks or nicks
  - Lowering competition from other stands (dilute loading)

Oxford  
**NANOPORE**  
Technologies

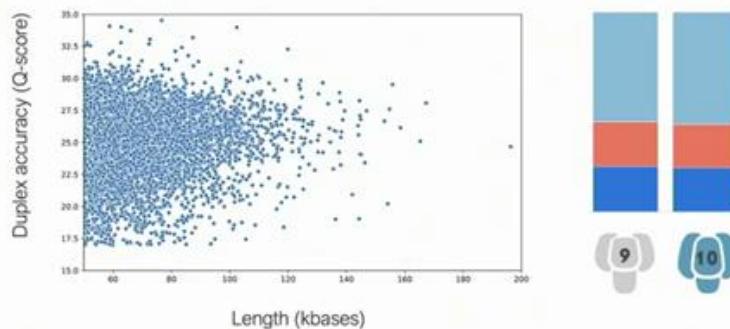
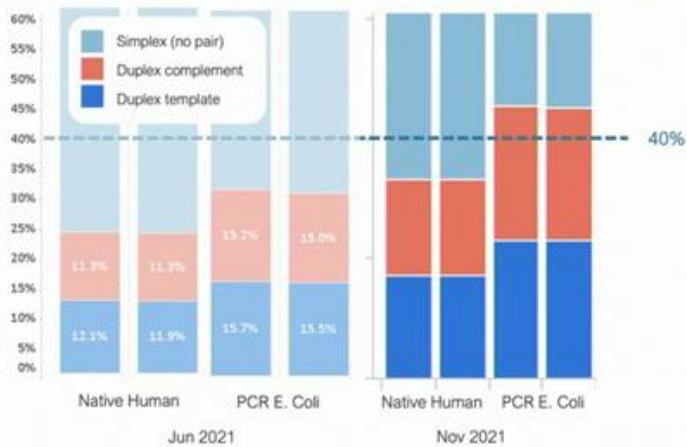
welcome  
**sanger**  
institute

## Nanopore Accuracy

Duplex – reading both strands

### Kit changes driving up Duplex data

- Recent changes to the sequencing kits give improved Duplex rates
- Improvements achieved with kit changes only (same flowcells)
- Proportion of data increased for ~30% to 45% in ideal system



### Long reads across both pores

- Duplex data is not read length dependant
- Comparable Duplex data rates with both R9 and R10 series pores

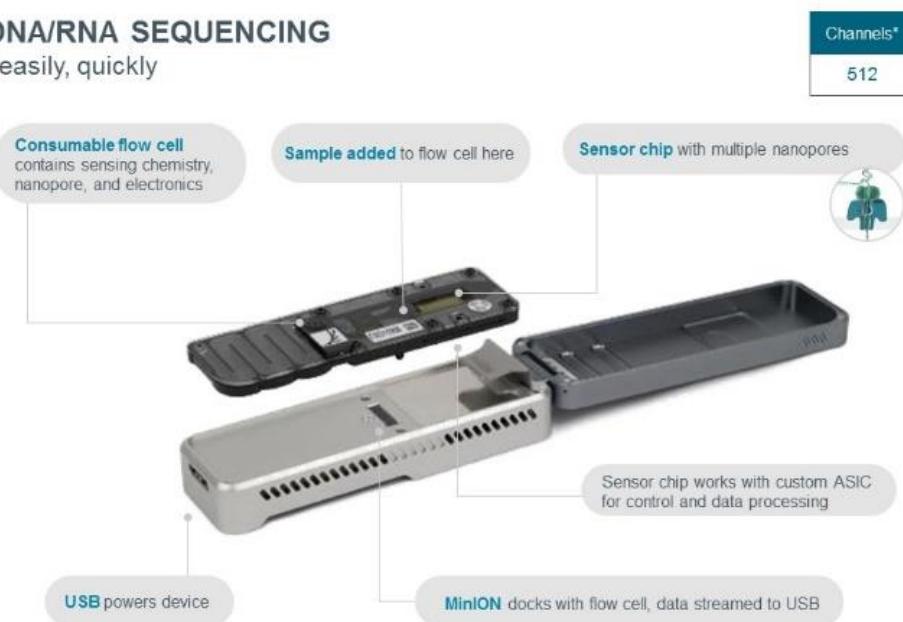
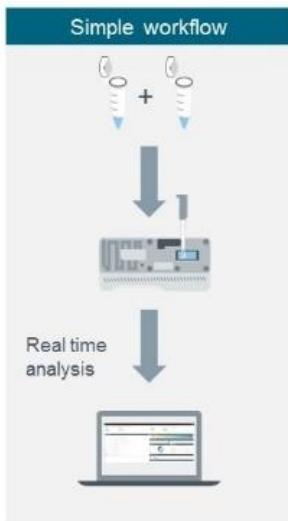
Available to developers, targeting feature in Kit 14

# minION

## MinION: PORTABLE DNA/RNA SEQUENCING

Sample to scientific insight easily, quickly

Channels\*  
512



An explanation from **Oxford NANOPORE Technologies™**

# Connected minION



Mk Ic

Mk Id based on ipad pro





# minION flowcell



# 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

## upto 48 flowcells with 3000 pores

### PromethION Updates

Coming to store this week

#### New PromethION P24 / P48 with A100 compute

- Significant performance improvement delivered by A100 technology
- Increased CPU, RAM
- Will ship from May onwards
- Upgrade options available – please contact sales

	P24	P48	A100 tower
CPU	2x GV100 112 cores total	4x GV100 112 cores total	4x A100 160 cores total
Storage	30 Tbytes	60 Tbytes	60 Tbytes
Connectivity	Dual-port 10 Gbase-T Ethernet Dual-port 10 GB SFP+ fibre		
Memory	384 GB	384 GB	512 GB
Display out	VGA	VGA	MiniDP and VGA



P24



P48



\$225,000

- 192 flow cells
- 42 kits
- 12M SW licence

\$310,000

- 288 flow cells
- 48 kits
- 12M SW licence

#### New PromethION P24 / P48 pricing

- Includes the new compute tower
- Brings pricing in line with flow cell price reductions from Jan

# PromethION P2 Solo

## PromethION “P2 solo”

Powerful flow cells, small device – up to 550-600Gb/run\*

### Two PromethION flowcells, connects to existing compute

- Based on existing flow cell design
- Two individually addressable prom flow cells
- Small versatile instrument to couple to existing GPUs/storage
- Ideal for studying large genomes or transcriptomics
- Easier to place in automation solutions
- Available as a starter pack or CapEx

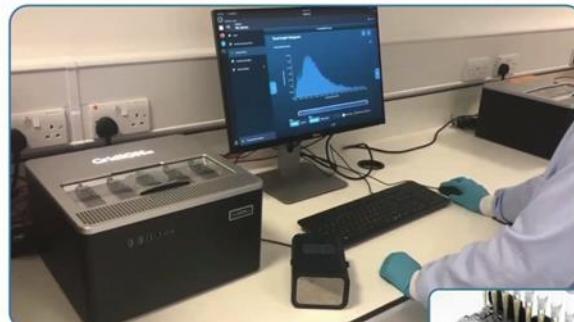
Pre-order in store today

P2 Solo: \$10,455

Consumable pricing will  
enable human WGS for  
under \$1,000



Early Access devices ship  
early Q2 2022



Four chamber flowcells  
in development



# Flongle flow cell adapter

\$90 for 2-3Gb





# ONT platforms

Nanopore Live

live

## 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 (based on internal test Mar 17)	Up to 40GB	Up to 6TB	Coming soon
Fee For Service available	No	Yes	Yes
<b>Specifications based on internal test Mar 2017</b>			
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 can be bought for no upfront capital cost



## Capital free devices

Starter packs can be bought with consumable budget



Flongle



MinION Mk 1B



MinION Mk 1C



GridION Mk 1



P24



P48

Starter pack: \$1,870

Includes:  
12 flow cells

Starter pack: \$1,000

Includes:  
2 flow cells + kit

Starter pack: \$4,900

Includes:  
6 flow cells + kit  
12 month software licence & warranty\*

Starter pack: \$49,955

Includes:  
60 flow cells + kit  
12 month software licence & warranty\*\*

Starter pack:  
\$165,000

Includes:  
60 flow cells + kit  
12 month software licence & warranty\*\*

Starter pack:  
\$287,000

Includes:  
120 flow cells + kit  
12 month software licence & warranty\*\*

# ONT Flowcell pricing

Pricing is transparent

And delivers competitive per Gb

LC  
LICENCE AGREEMENT



Flongle flow cell	Output:	
	1 Gb	2 Gb
\$90	\$90 / Gb	\$45 / Gb

MinION / GridION flow cell	Output:		
	10 Gb	30 Gb	50 Gb
\$675	\$68 / Gb	\$23 / Gb	\$14 / Gb
\$500	\$50 / Gb	\$17 / Gb	\$10 / Gb
\$475	\$48 / Gb	\$16 / Gb	\$10.5 / Gb

PromethION flow cell	Output:		
	100 Gb	200 Gb	300 Gb
\$1600	\$16 / Gb	\$8 / Gb	\$5 / Gb
\$ 940	\$9 / Gb	\$5 / Gb	\$3.1 / Gb
\$ 680	\$7 / Gb	\$3.4 / Gb	\$2.3 / Gb
\$625	\$6 / Gb	\$3 / Gb	\$2 / Gb

# ONT library and flowcell pricing

**Pricing is transparent**  
Multiplexing and price per sample

LC



Flongle samples loaded	12	24	48	96	MinION samples loaded	12	24	48	96	PromethION samples loaded	12	24	48	96
Flow cell	\$90	\$90	\$90	\$90	Flow cell	\$500	\$500	\$500	\$500	Flow cell	\$680	\$680	\$680	\$680
Library prep	\$147	\$195	\$219	\$243	Library prep	\$147	\$195	\$219	\$243	Library prep	\$147	\$195	\$219	\$243
Price per sample	\$20	\$12	\$6.5	\$3.5	Price per sample	\$54	\$29	\$15	\$8	Price per sample	\$69	\$36	\$19	\$10

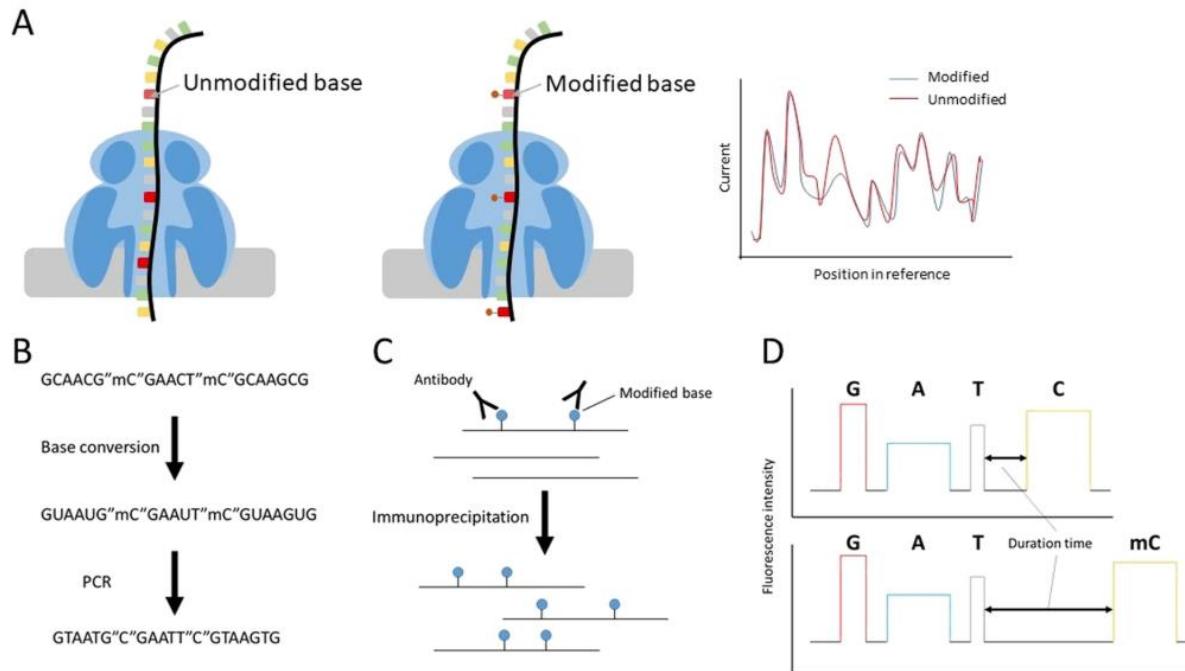


# ONT applications

- » Super-long reads (>100kb reads possible)
- » “Run until” done. W.I.M.P
- » Selective reads
- » Cas9 mediated enrichment
- » Mobile sequencing
- » Direct methylation detection?
- » Direct RNA sequencing

# Methyl seq approaches

From: Recent advances in the detection of base modifications using the Nanopore sequencer



Modified base detection using Nanopore sequencing and general methods. Schema of modified base detection using the Nanopore sequencer (a) and through bisulfite conversion (b), immunoprecipitation of nucleic acids (c), and SMRT sequencing (d)

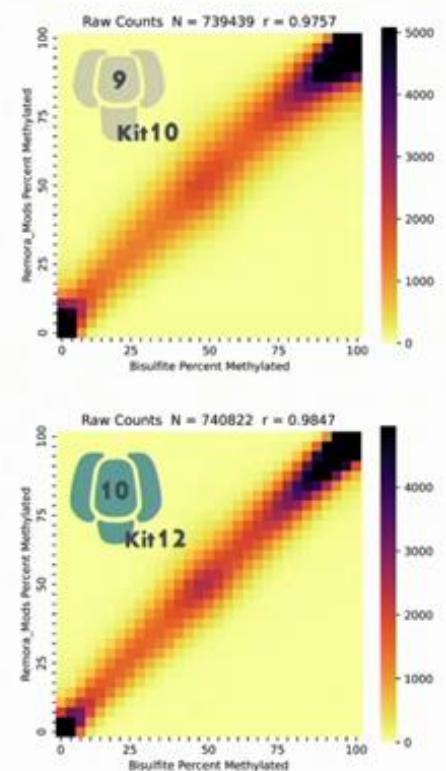
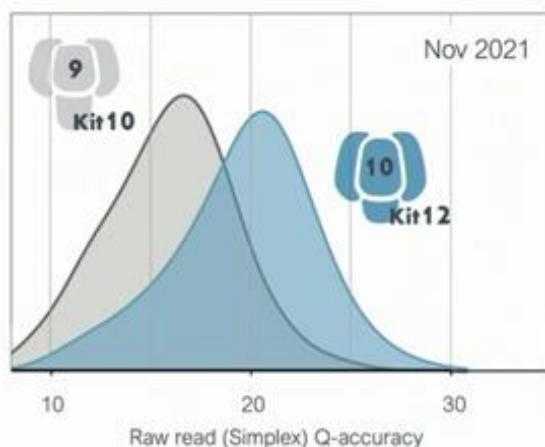
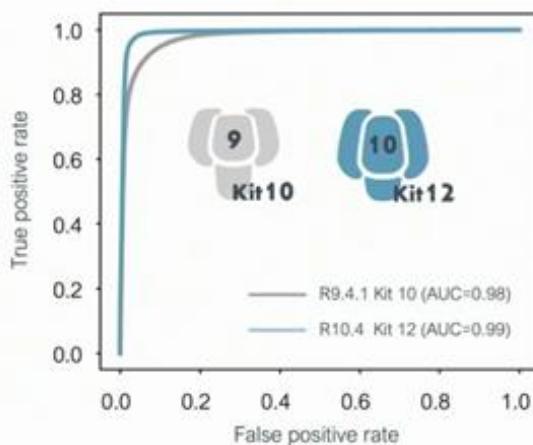


# Nanopore Accuracy

## Methylation accuracy

### Setting a new gold standard

- Methylation performance  $\geq$  previous models
- Basecall / mapping accuracy maintained vs canonical calling
- R9.4.1 results already excellent: "better than bisulphite"
- R10.4 significantly better than R9.4.1



24

© 2022 Oxford Nanopore Technologies plc.  
Oxford Nanopore Technologies products are not intended for use for health assessment or to diagnose, treat, mitigate, cure, or prevent any disease or condition.

"Ground truth": bisulfite sites from HG002 with >50X coverage and <1% or >99% methylation

Oxford  
**NANOPORE**  
Technologies

wellcome  
**sanger**  
institute



# ONT Performance

- » MinION: 100 Mb/hr. 1-20Gb
- » PromethION: 40-200Gb/flowcell
- » Read length potentially as long as molecule
- » Run until done
- » \$50-150/Gb minION, no capital cost
- » \$50-150/Gb GridION
- » \$10-20/Gb promethION



# New Sequencers



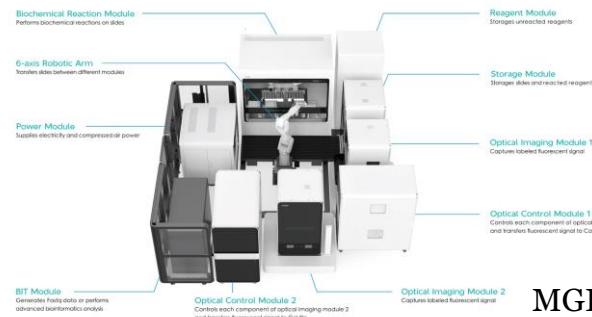
Onso



Element



Ultima UG100



Singular S4

MGI

# Comparison Table

	Illumina Novaseq	Illumina Nextseq	Ultima UG100	Element Aviti	Singular G4	PacBio SBB
Instrument	\$1.25M	\$335k	\$1.5M	\$289k (\$249K)	\$350k	\$259K
\$/Gb	3 (10B), 2 (25B)	20 (9 in 2024)	1	2-5	8-10	15
Accuracy	QQQ	QQQ	QQ	QQQ	QQQ	QQQQ
Max read length	2x150 (SP)	2x300	300 SE	2x300	2x150	200SE
Run Time for 2x150 or 1x300	24 hrs	48 hrs	18 hrs	48 hrs	19 hrs	48hrs
Flowcells	2	2	2	2	4	2
Yield/run	6000Gb, 16000Gb	30-360Gb	>4000Gb	480Gb	360Gb	150Gb
Other	Infinity long reads	Infinity long reads		Non patterned, loop long reads		Non patterned
Use case			Cheap sequence for big data applications,	-high plex levels -low complexity -long reads	Flexibility via independent flowcells	High accuracy applications, high plex potential

# PacBio Onso

- Release 2023
- Most accurate short read
- 90% > Q40
- 1x200 or 2 x 150 in 48hr
- Sequencing by binding
- Accuracy to 0.001% VAF without UMIs



## SBB advantages: no base modifications, no molecular scarring

Creates optimized steps, incorporates native nucleotides, and produces unmodified DNA

### Sequencing by Synthesis (SBS)

Unblocked  
3' end



Incorporate/  
Interrogate



Cleavage



### Sequencing by Binding (SBB®)

Blocked 3' end



Interrogate



Activate



Incorporate

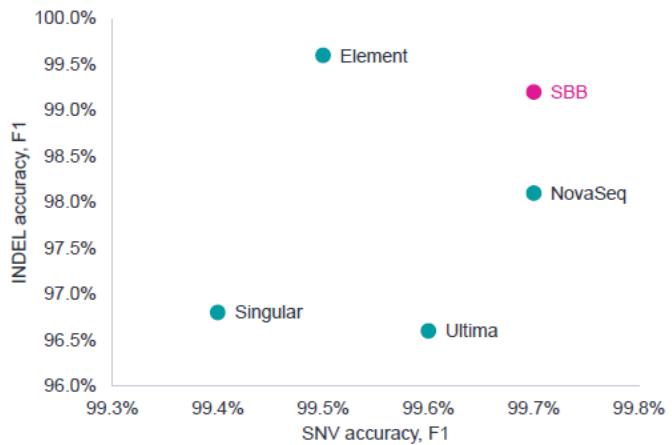




# SNV and Indel accuracy are better than any other platform

## Excellent variant calling performance for SBB

	SBB	NovaSeq	Element	Ultima	Singular	
SNV	Recall	99.6%	99.9%	99.1%	99.6%	99.2%
	Prec.	99.7%	99.5%	99.8%	99.6%	99.7%
	F1	99.7%	99.7%	99.5%	99.6%	99.4%
INDEL	Recall	98.9%	97.9%	99.3%	96.4%	96.4%
	Prec.	99.4%	98.4%	99.8%	96.8%	97.1%
	F1	99.2%	98.1%	99.6%	96.6%	96.8%



- Expecting further improvements through variant caller training

SBB / NVSQ: HG002 GIAB benchmark v4.2, DeepVariant 1.3.0 with the Illumina WGS model  
Element: <https://go.elementbiosciences.com/access-app-note-roche-human-microbial>

PacBio

Ultima: <https://doi.org/10.1101/2022.05.29.493900> HG002 40x coverage

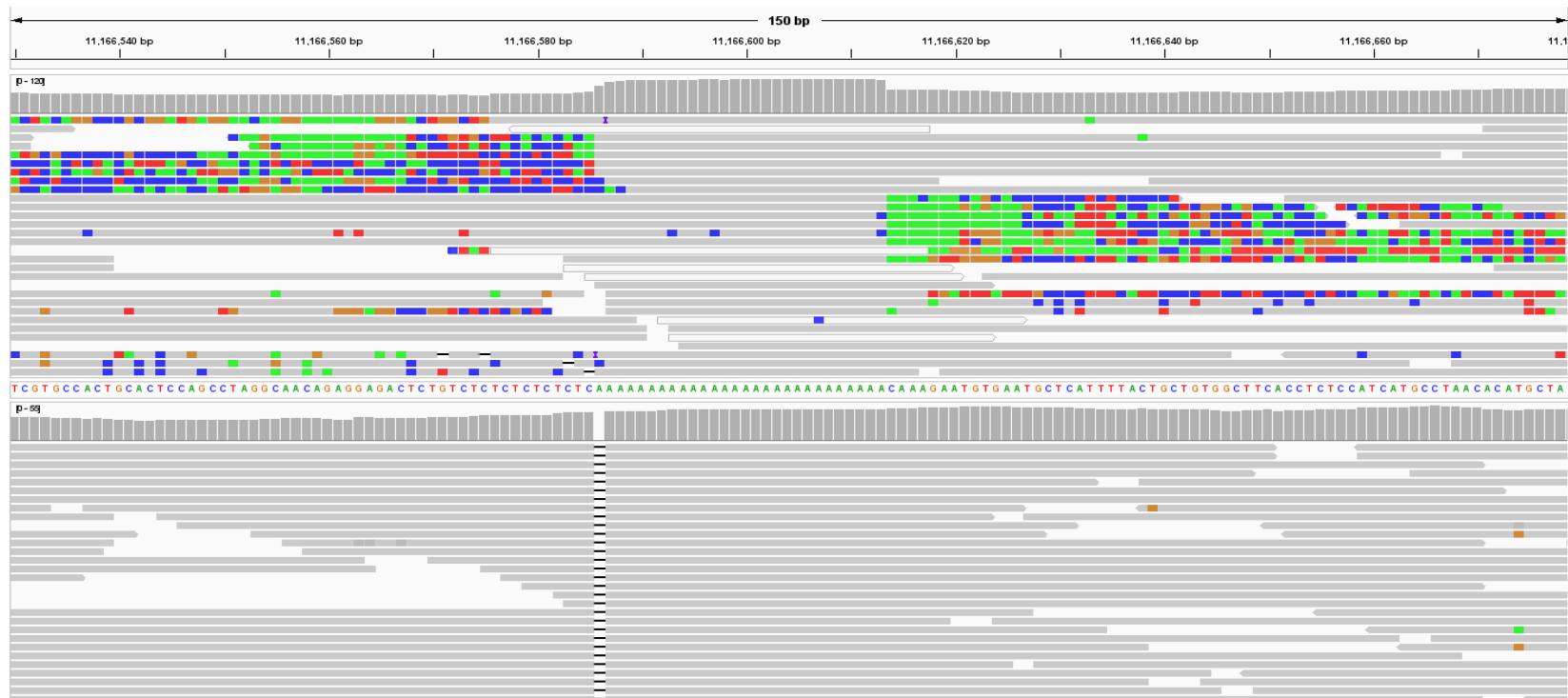
Singular: G4 Human Genome Sequencing Technical Report 1.0. NA12878 40x coverage

15



## **Sequencing by binding (SBB) chemistry enables highly accurate reads**

#### **Optimized steps, native nucleotide incorporation, scarless DNA strand formation**



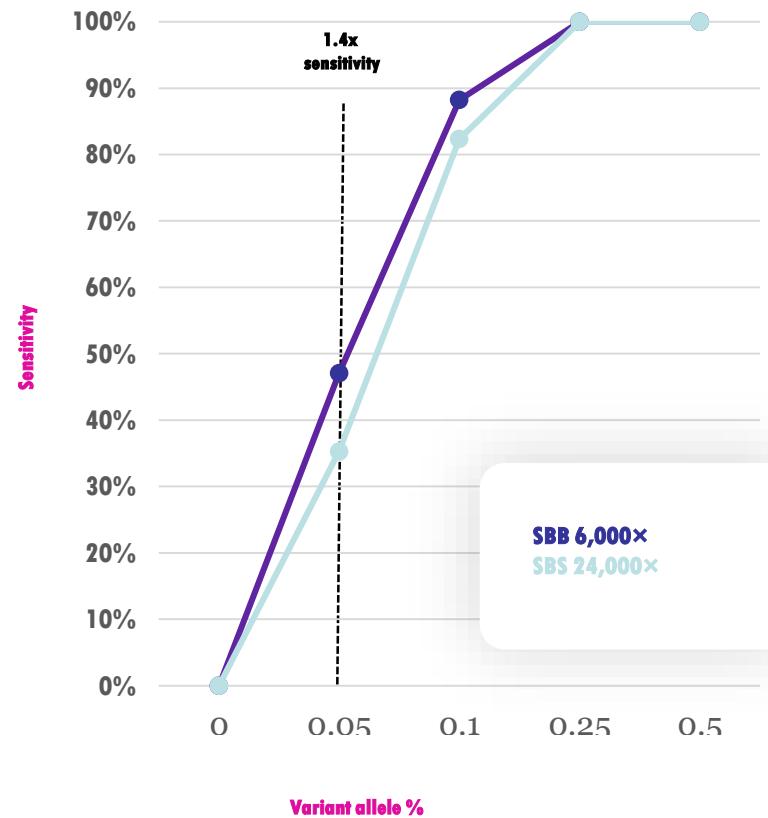
*Data based on internal testing*

## SBB demonstrates >4x improvement in sequencing efficiency



6,000 $\times$  non-UMI SBB sequencing exceeds >24,000 $\times$  SBS UMI sequencing at 0.05% and 0.1%

PacBio



Data based on internal testing. >4x improvement in sequencing efficiency relative to SBS

# Singular G4

- \$350K
- \$10/Gb
- 2 x 150 in 19 hours
- 4 x 4 lane flowcells. Each flowcell can be run and loaded independently
- Uses similar SBS and clustering to Illumina



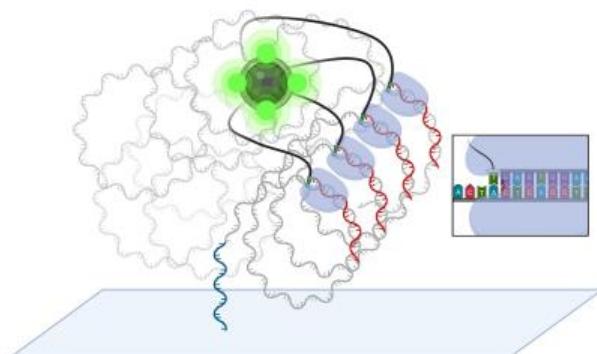
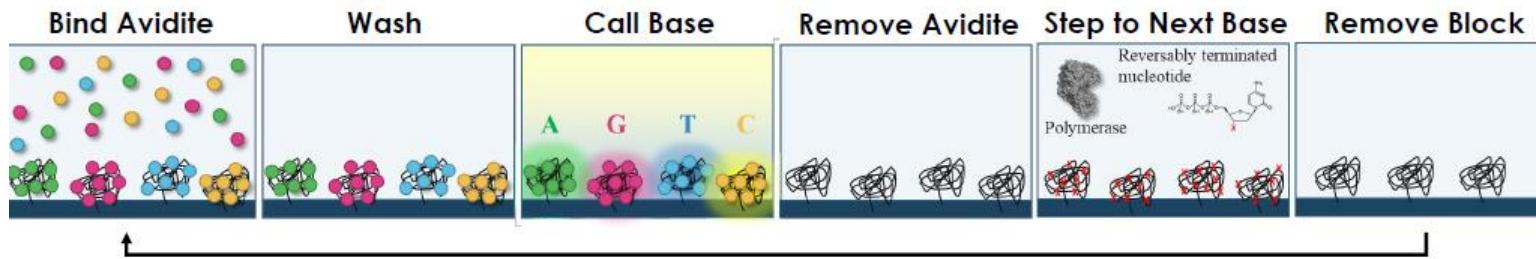
# Element Aviti

- ~\$335k. \$249 each if order 3
- \$2-\$5/Gb. \$1680 for 200-300Gb
- 2 flowcells. 2 x 150. 80% >Q40
- 2 x300 available
- Sequencing by binding on RCA amplified template
- Loop synthetic long read
- Low duplicates
- No tag hopping so can do combinatorial indexing
- No issue with low complexity
- Can have long inserts >2kb
- Broad have ordered 3. NEB have first instrument



# Element sequencing chemistry

## What is Avidity Sequencing?



Link to BioRxiv pre-print on Avidity Sequencing



## “Have It All With AVITI”

An unmatched combination of performance, cost, and flexibility

**>90%**

Q30 (2x150)

**2 FC**

100% Independent

**600 Gb**

Output/run

**\$1**

per M Reads at 2x75

**\$5**

per Gb at 2x150

**10 Kb**

Read length using  
LoopSeq

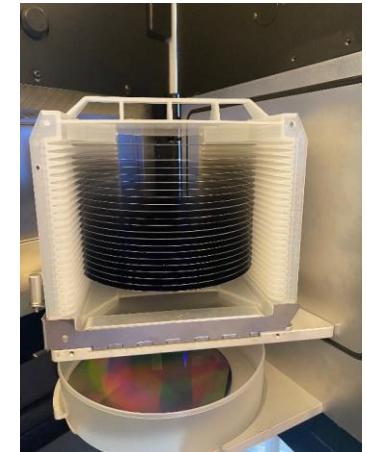


Element Biosciences

2

# Ultima

- \$100 genome. \$1/Gb
- SE 300 in 18 hr 20min
- Flow based. Some indel errors around homopolymers (0.3%)
- 89% Q30
- Some dropout at high AT and GC
- Flow based sequencing on a 10"
- Spinning disc
- Have 96 in-line barcodes



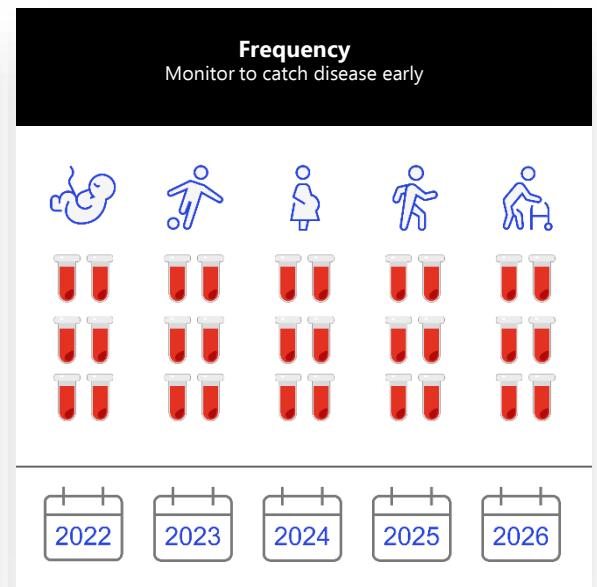
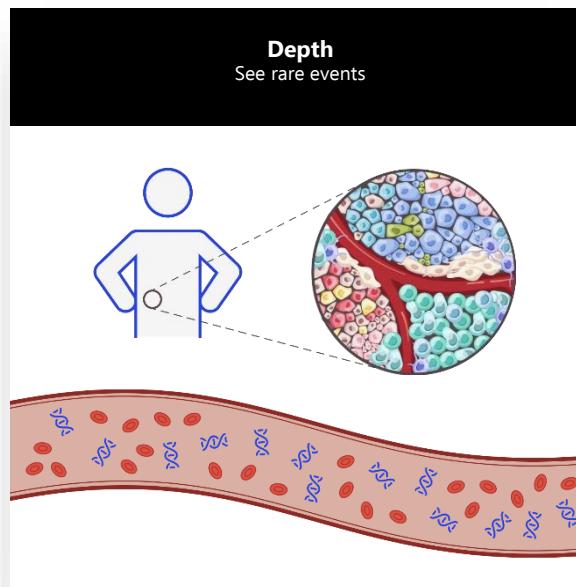
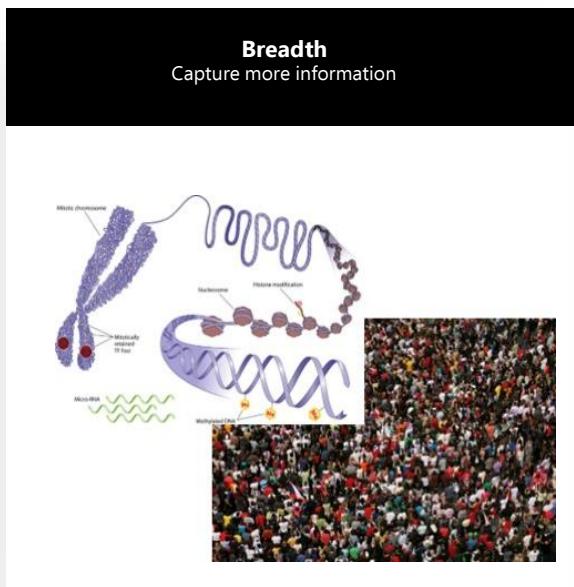
# So why Ultima?

April 20<sup>th</sup> 2023



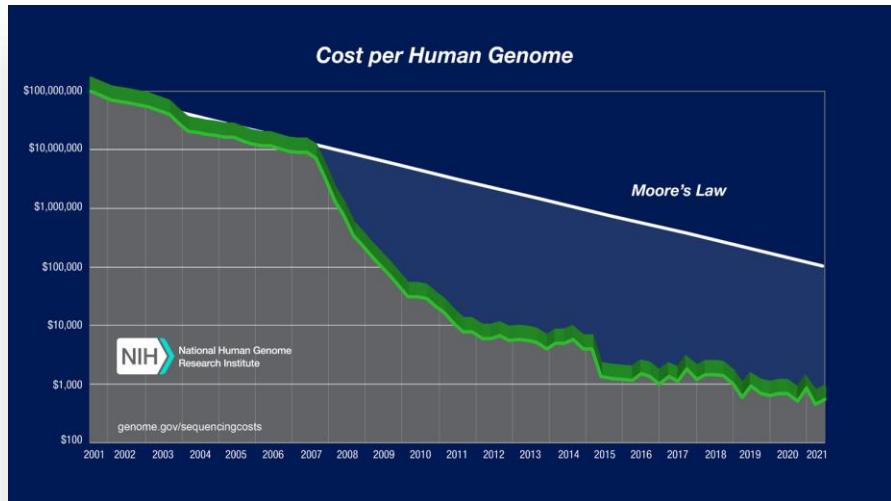


# Demand for sequencing will increase by orders of magnitude

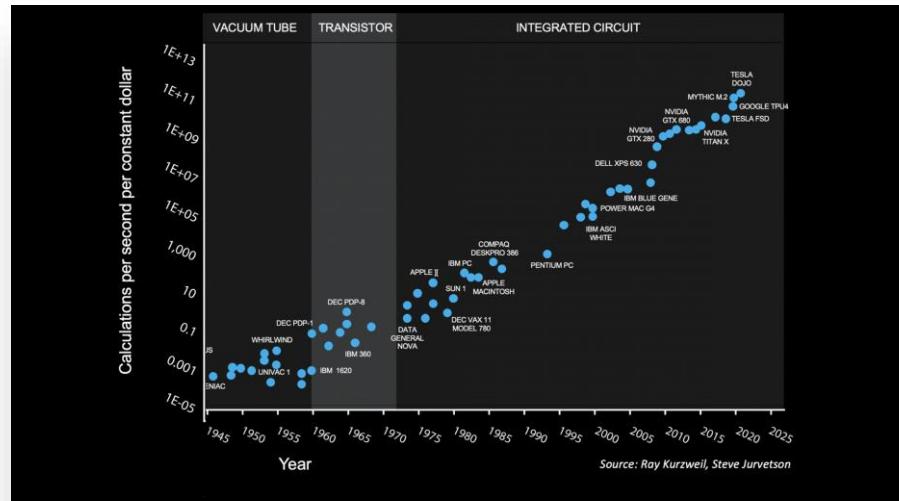


Proprietary and Confidential

# New technology platform needed for a sequencing ‘Moore’s Law’



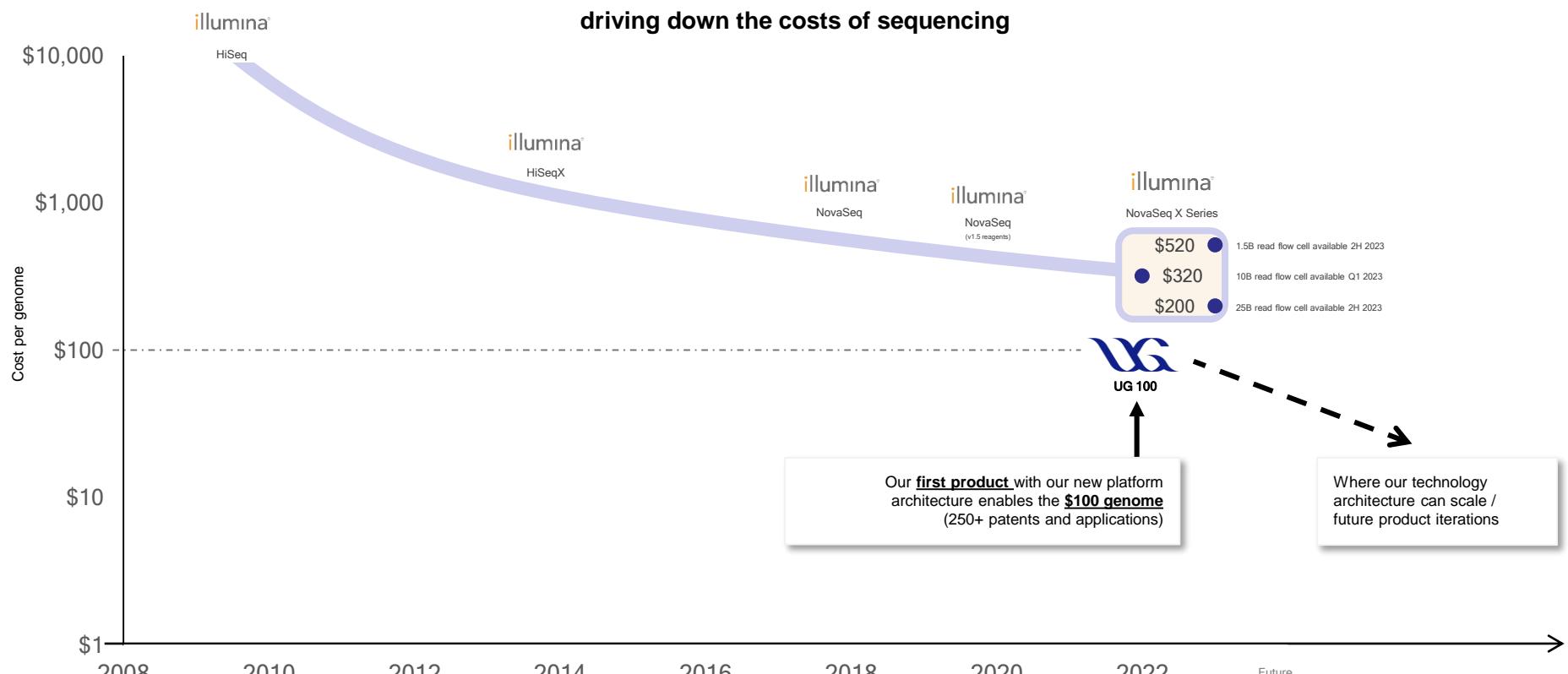
Moore's Law for sequencing has stalled



Other industries have required platform shifts to continue exponential cost reduction

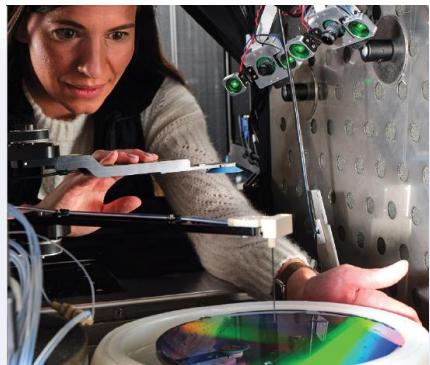
Proprietary and Confidential

# Unveiled the \$100 genome at AGBT 2022 – only the starting point



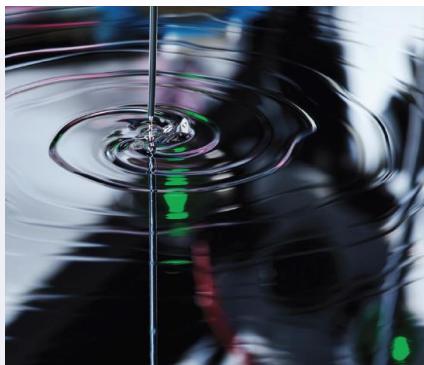
Proprietary and Confidential

# Fundamentals of our new sequencing architecture



## **Open flow cell**

200mm open-faced surface lowering cost and complexity



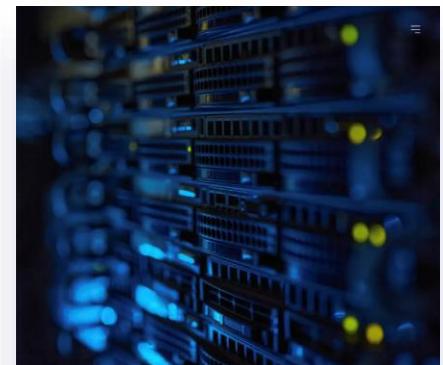
## **Rotational symmetry**

Fast & efficient reagent delivery with ultra-fast imaging



## **“Mostly natural” chemistry**

Non-terminating chemistry for efficient, long and fast reads



## **Machine learning**

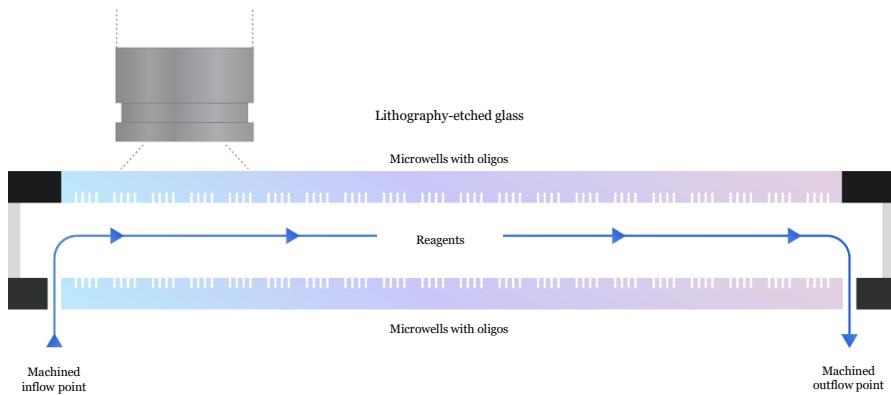
Powerful ML algorithms trained by genome-scale data



Proprietary and Confidential

# Traditional flow cells are expensive and inefficient

Longitudinal cross section of a channel



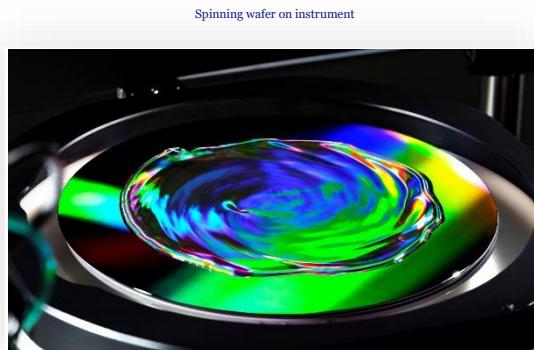
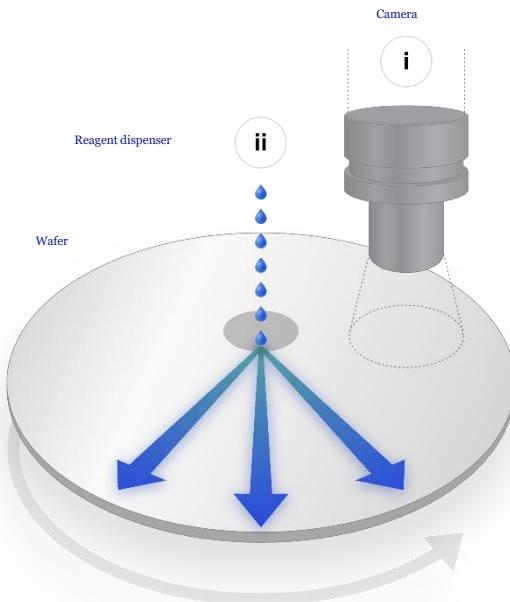
Traditional Flow Cell

- Expensive substrate
- Packaging complexity
- Pressure constraints
- Reagent inefficiency
- Contamination challenges
- Cartesian scanning

Proprietary and Confidential

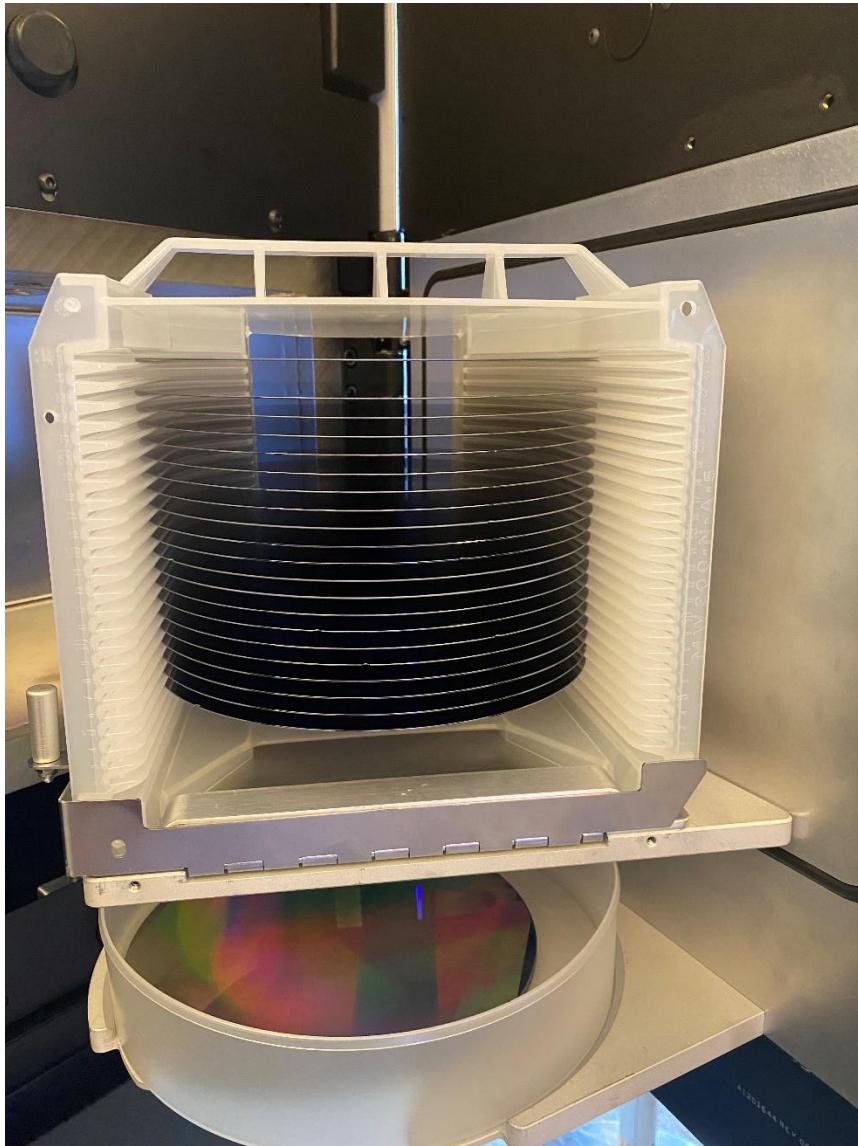


## Revolutionary new sequencing hardware

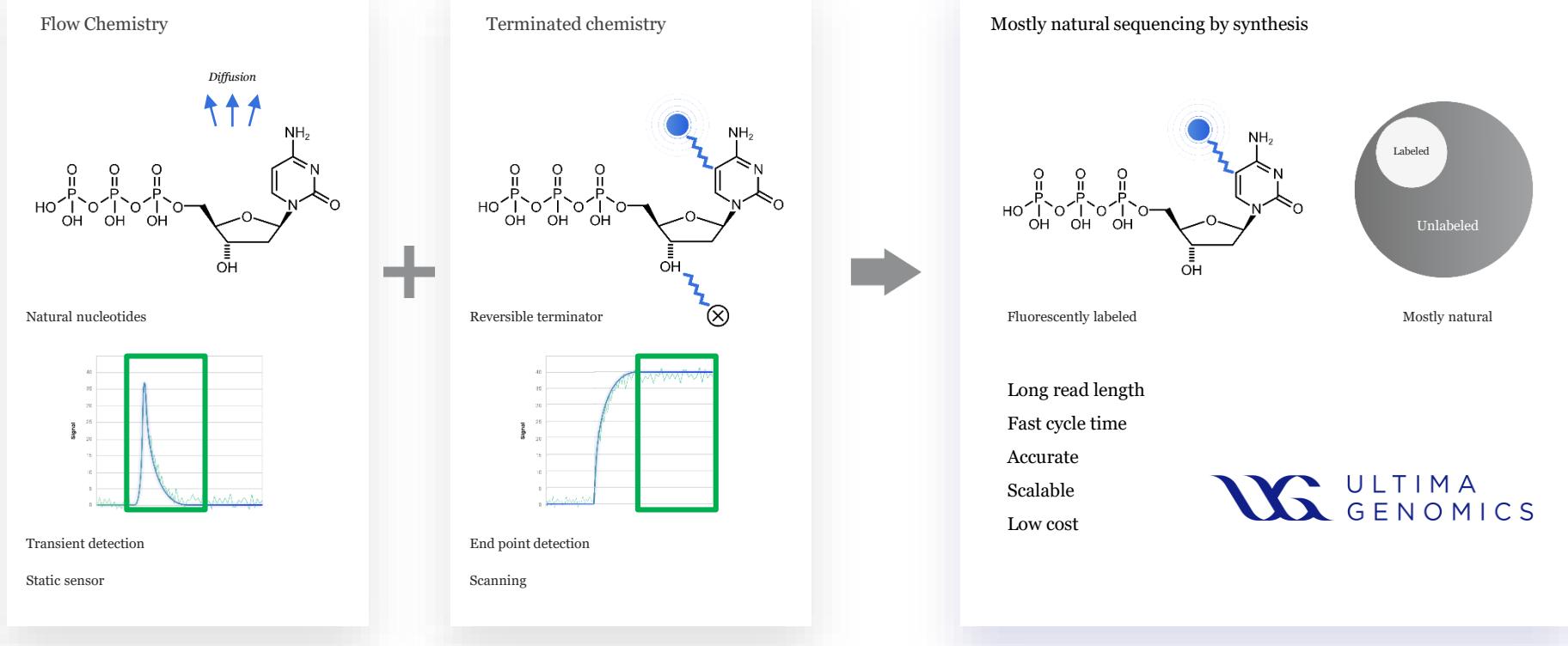


- Low cost (standard substrate)
- ~10B reads per wafer (x2)
- Fast and efficient reagent delivery
- “Air gap” minimized contamination
- Rotational optical scanning

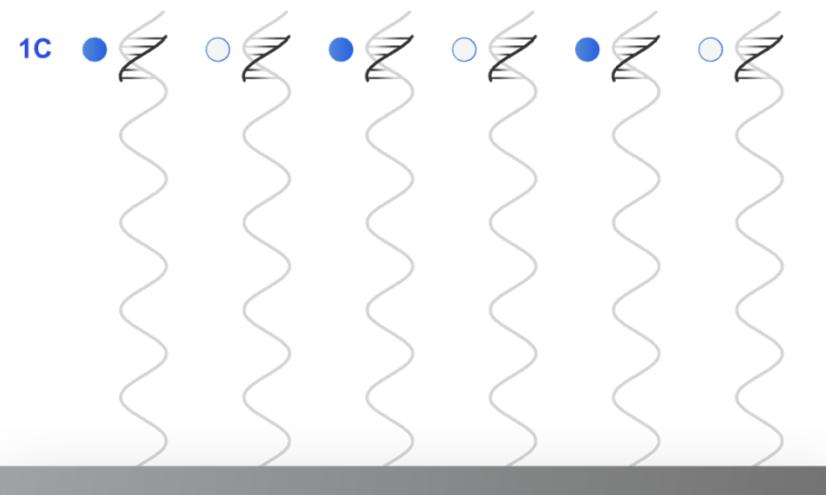
Proprietary and Confidential



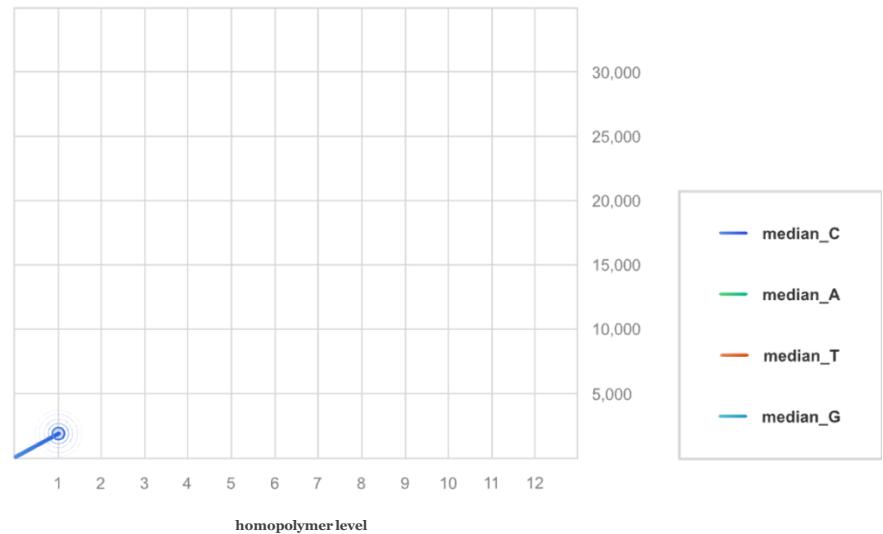
# A novel spin on sequencing chemistry



# Mostly natural chemistry combines advantages of flow chemistry with optical scanning

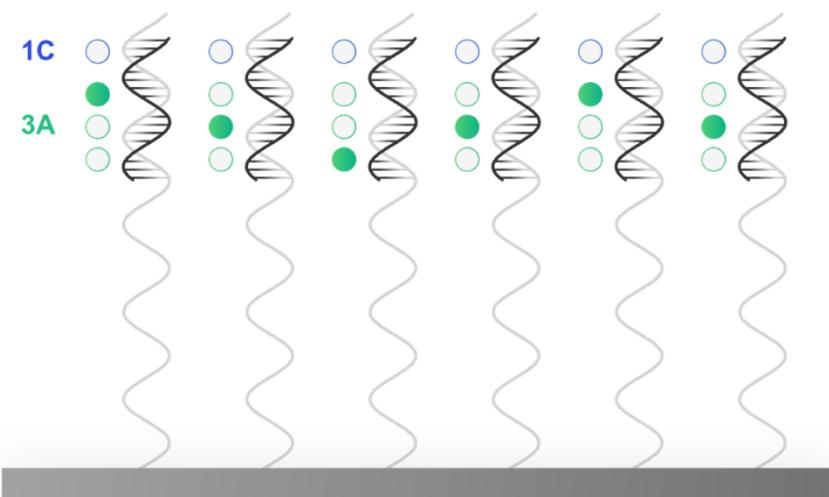


Median signals vs. homopolymer

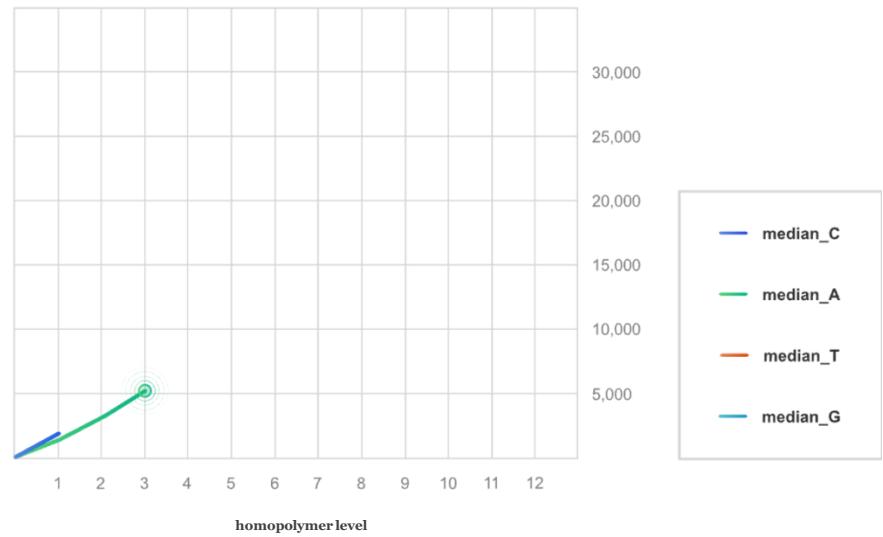


- Majority nucleotides unlabeled to avoid quenching
- Minimal scarring supports longer reads
- Faster runs via 2min wash->image->cleave cycle
- Endpoint detection significantly improves accuracy
- Maintain signal linearity to at least homopolymer length 12
- Machine learning accounts for sequence context

# Mostly natural chemistry combines advantages of flow chemistry with optical scanning



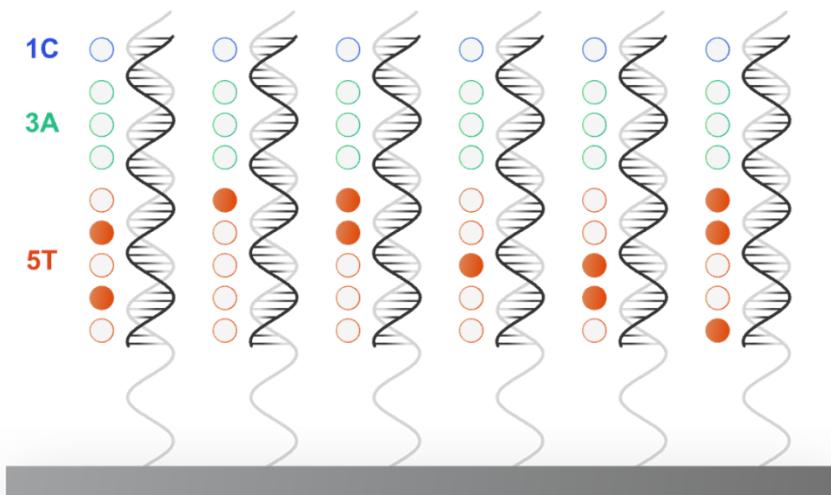
Median signals vs. homopolymer



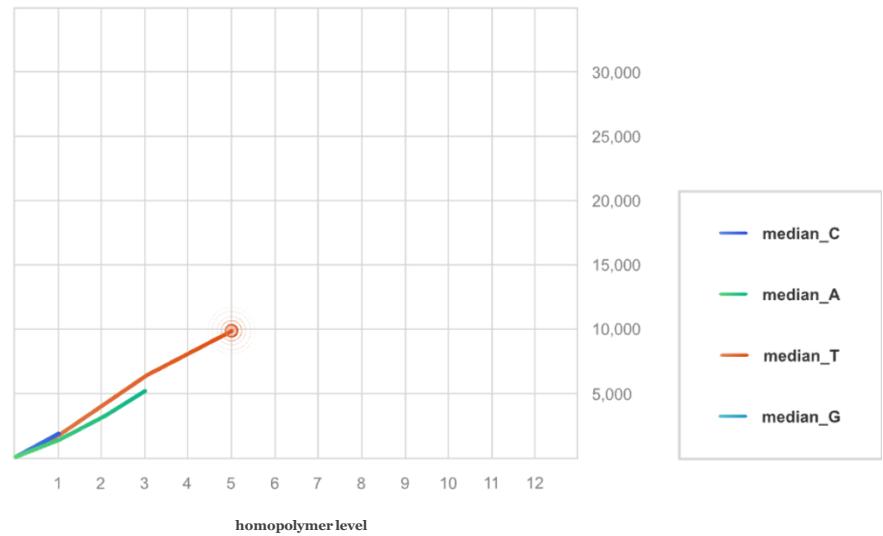
- Majority nucleotides unlabeled to avoid quenching
- Minimal scarring supports longer reads
- Faster runs via 2min wash->image->cleave cycle
- Endpoint detection significantly improves accuracy
- Maintain signal linearity to at least homopolymer length 12
- Machine learning accounts for sequence context

Proprietary and Confidential

# Mostly natural chemistry combines advantages of flow chemistry with optical scanning



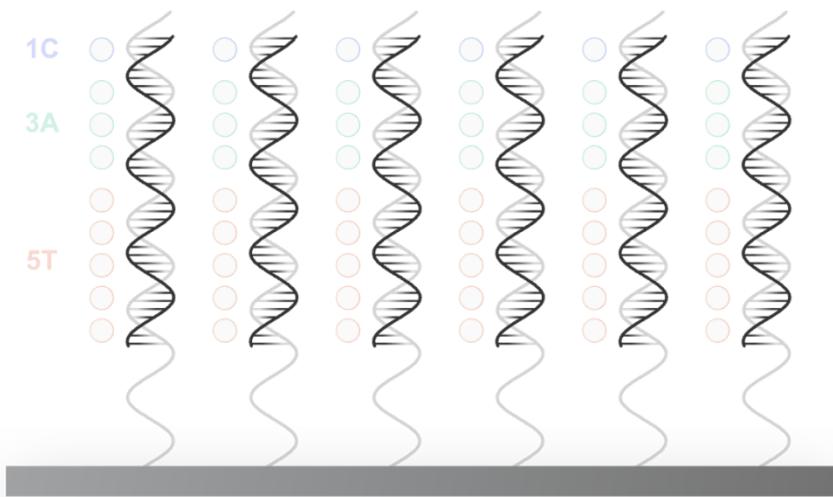
Median signals vs. homopolymer



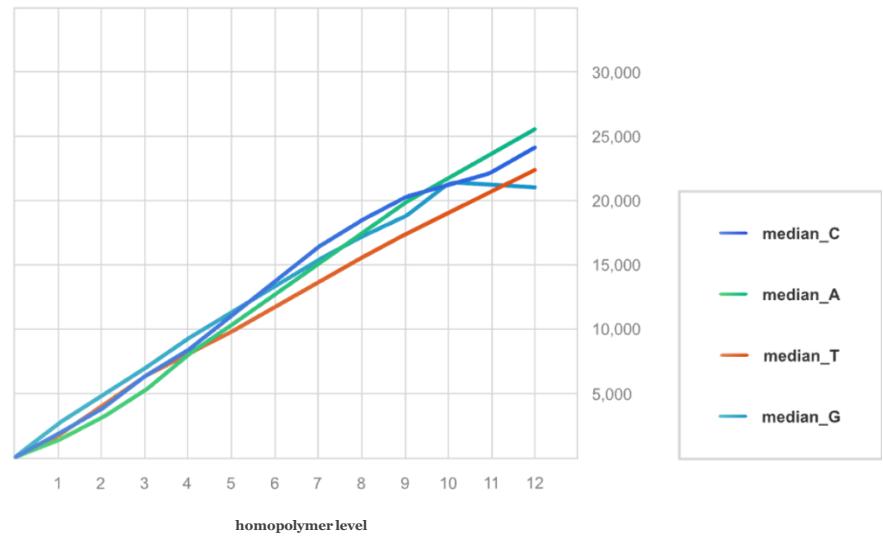
- Majority nucleotides unlabeled to avoid quenching
- Minimal scarring supports longer reads
- Faster runs via 2min wash->image->cleave cycle
- Endpoint detection significantly improves accuracy
- Maintain signal linearity to at least homopolymer length 12
- Machine learning accounts for sequence context

Proprietary and Confidential

# Mostly natural chemistry combines advantages of flow chemistry with optical scanning



Median signals vs. homopolymer



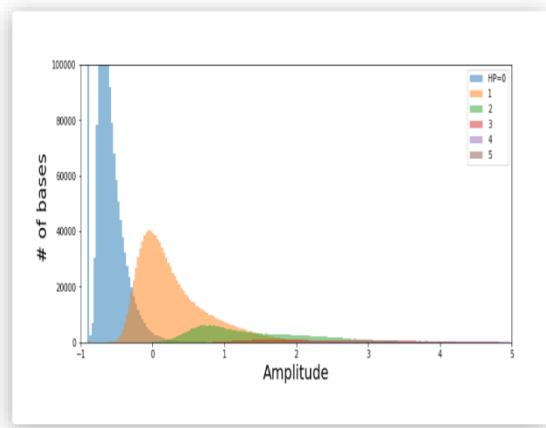
- Majority nucleotides unlabeled to avoid quenching
- Minimal scarring supports longer reads
- Faster runs via 2min wash->image->cleave cycle
- Endpoint detection significantly improves accuracy
- Maintain signal linearity to at least homopolymer length 12
- Machine learning accounts for sequence context

Proprietary and Confidential

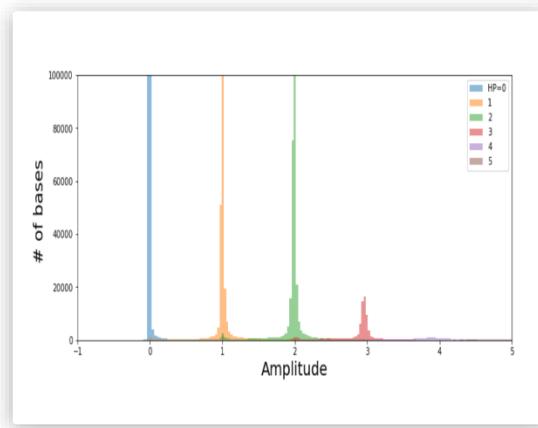


## Machine learning base calling for improved accuracy

Raw signals



Expected values



Predicted probabilities

Flow	1	2	3	4	5	6	7	8	9	10
Base	T	A	C	G	T	A	C	G	T	A
p(L=0)	-	-	1	-	-	1	-	-	1	-
p(L=1)	0.999	0.05	-	1	0.999	-	-	1	-	-
p(L=2)	0.001	0.95	-	-	0.001	-	-	-	-	0.01
p(L=3)	-	-	-	-	-	-	0.002	-	-	0.99
p(L=4)	-	-	-	-	-	-	0.997	-	-	-
p(L=5)	-	-	-	-	-	-	0.001	-	-	-
p(L=6)	-	-	-	-	-	-	-	-	-	-
Call	1	2	0	1	1	0	4	1	0	3
Qual	30	13	40	40	30	40	25	40	40	20

Ultima's neural network identifies and deconvolves systematics such as phasing, signal decay and context for better base calling

Proprietary and Confidential

# Continuously improving our performance and data quality

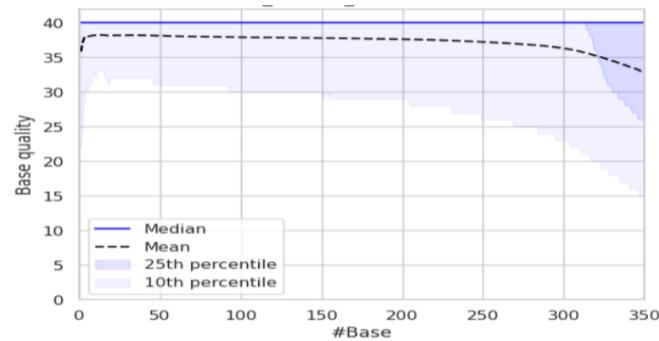
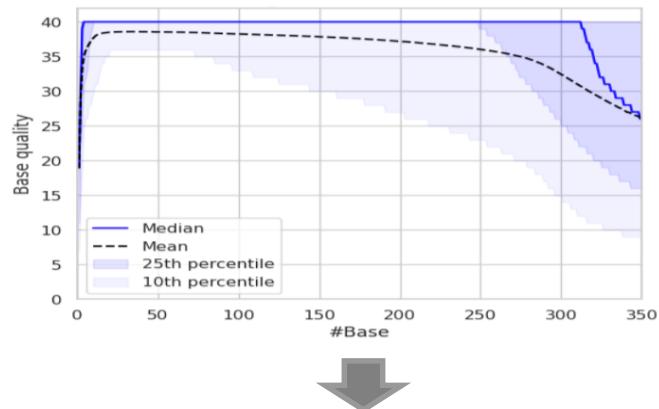
System performance demonstrated on GIAB reference samples:

		June '22	Feb '23
<b>Base Quality</b>	Q20	95%	96%
	Q30	85%	89%
<b>Base error</b>	Substitution	<0.1%	<0.1%
	Indel	0.4%	0.3%
<b>Read length</b>	Mean	282	294
	Mode	310	321
<b>Variant calling accuracy* (F1)</b>	SNP	99.6%	99.8%
	Indel	96.4%	98.6%
<b>Coverage uniformity</b>	F95	1.8	1.5

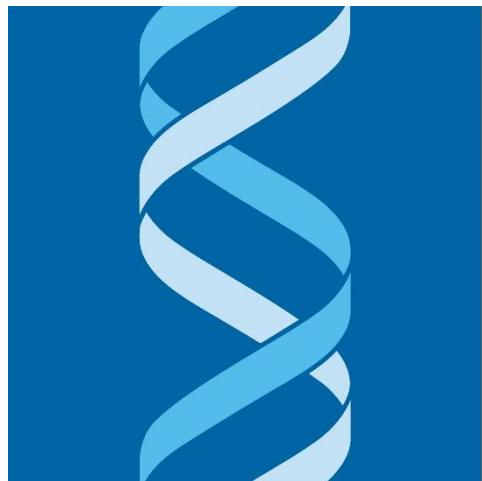
\* Evaluation region: GIAB v4.2.1  
excluding homopolymer regions of length  $\geq 11$  (0.3% of HCR)  
excluding low complexity regions (1.5 % of HCR)



Dataset coming soon



Proprietary and Confidential



wellcome trust  
**sanger**  
institute

# Current Sanger Platforms

- » ABI 3730 Capillary (outsource)
- » 1 Novaseq X+
- » 14 Novaseq 6000
- » 6 MiSeq
- » 3 PacBio Revio
- » 12 PacBio Sequel IIe
- » Oxford Nanopore gridION
- » Oxford Nanopore promethION



Thanks!





# Any Questions ?

**[mq1@sanger.ac.uk](mailto:mq1@sanger.ac.uk)**