



Illumina Sequencing Technology

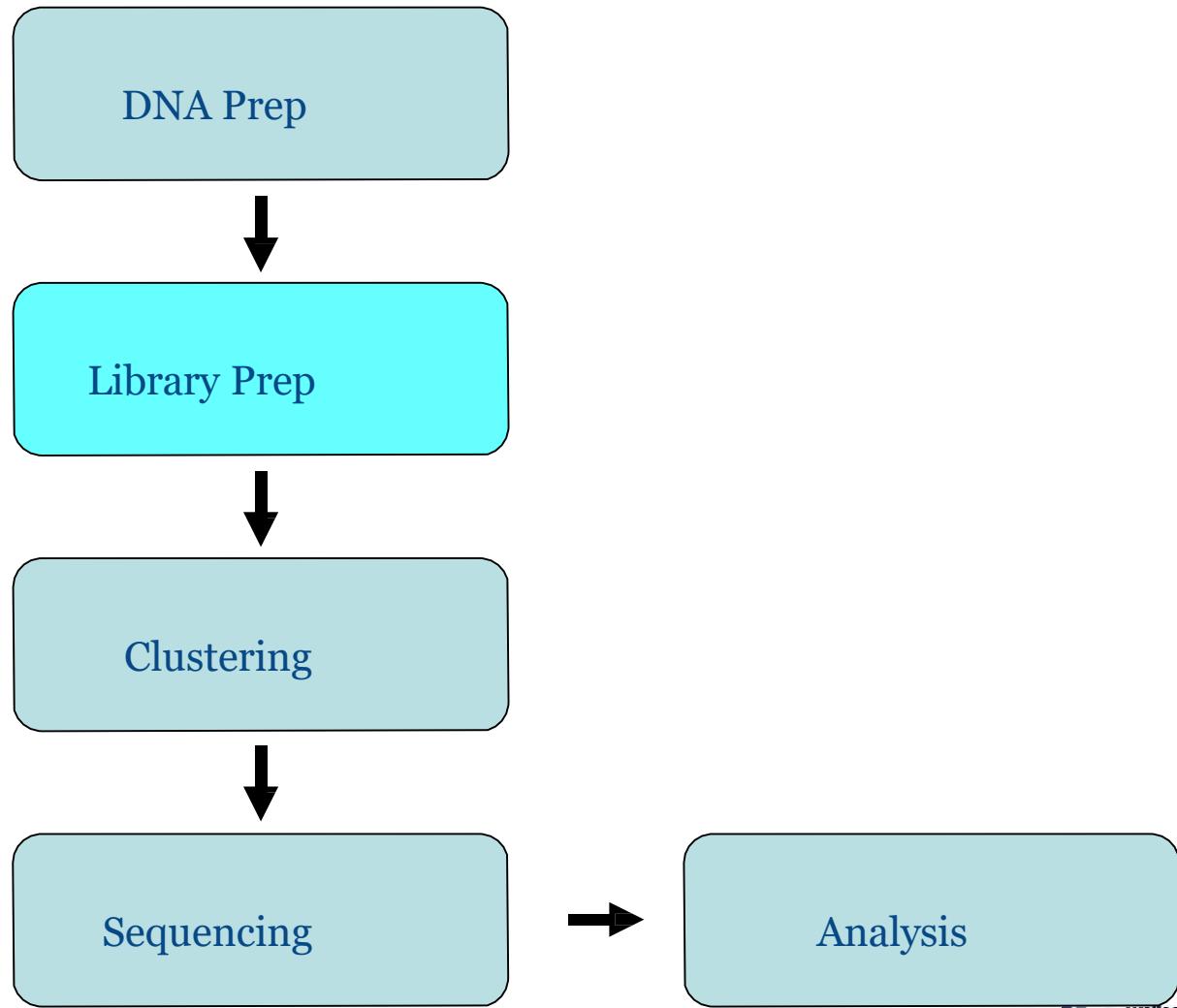
Michael Quail
mq1@sanger.ac.uk



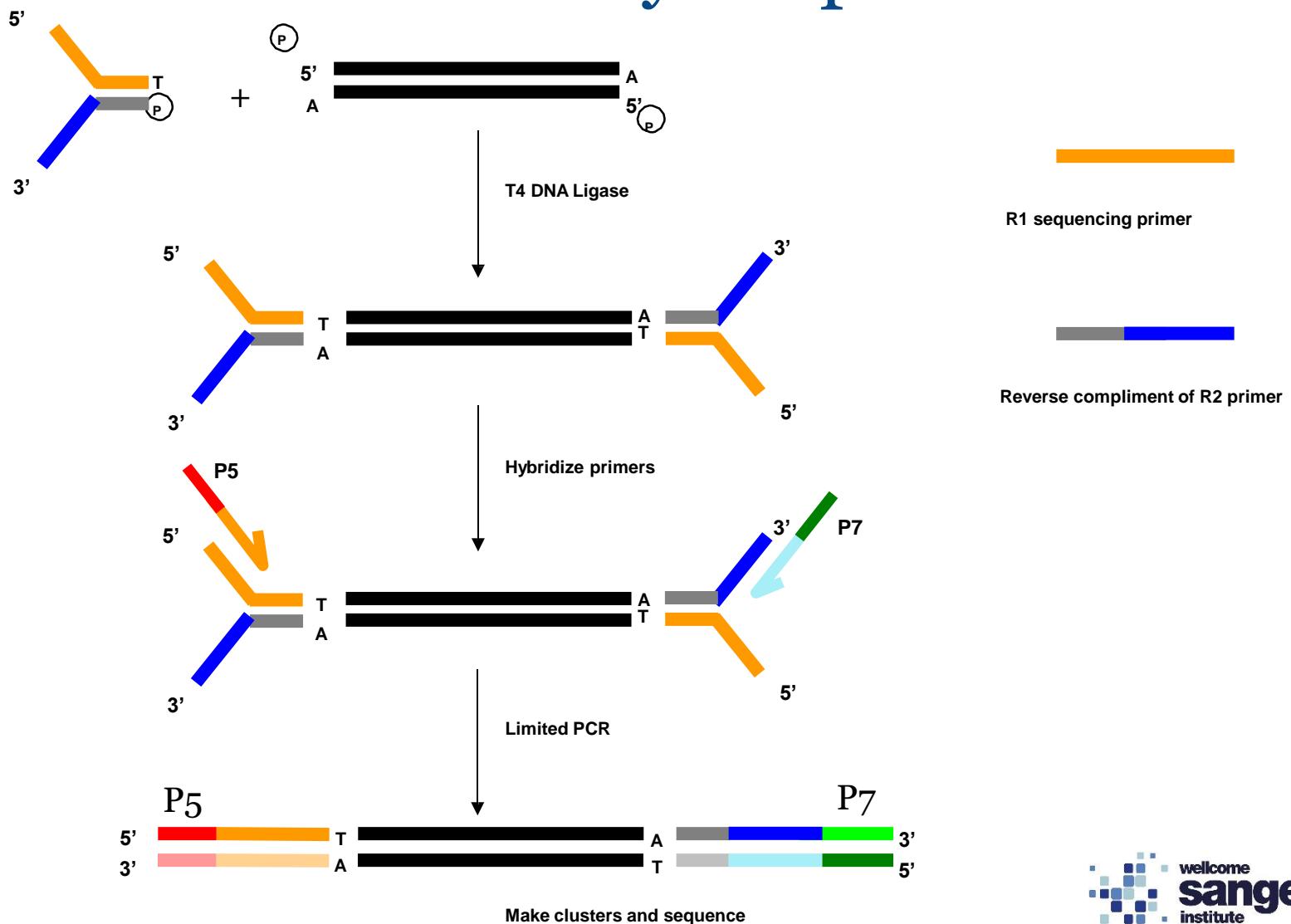
Solexa

- » 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 9Tb.
- » Acquired by Illumina in 2007
- » Short read sequencing
- » Accurate (0.2-0.4% error)
- » Market leader.
- » \$2-\$50/Gb

Illumina workflow

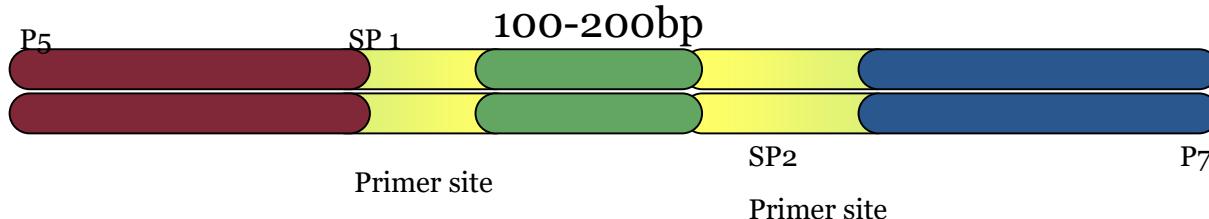


Illumina Paired End Library Prep



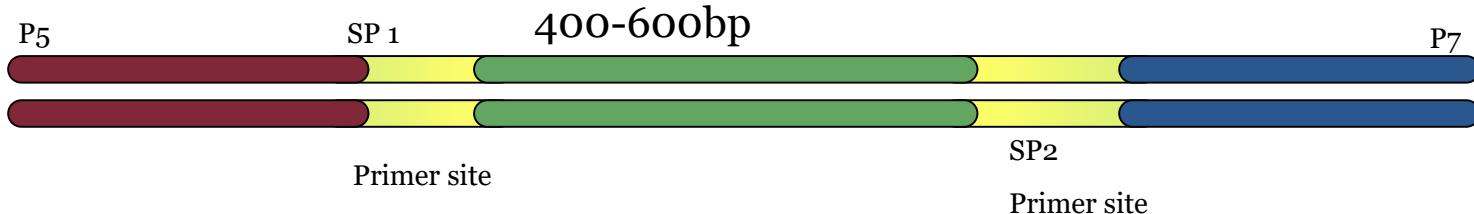
Illumina Library Insert sizes

Exome, targeted, ChiP, ATAC



Reads need to be just long enough to map and ideally not to overlap exon boundaries

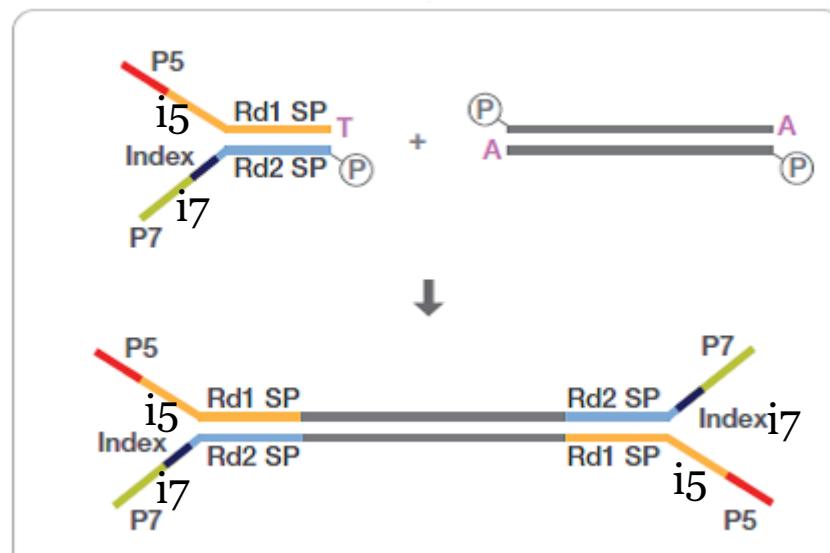
Whole Genome



Fragments need to be just long to span common repeat elements eg AluI with unique sequence on either side so can map

Illumina Truseq Library Prep

Dual indexing



Sequencing with Paired-Ends



Reference This is really the best way to do sequencing

Single-reads This is

... is really

... really the

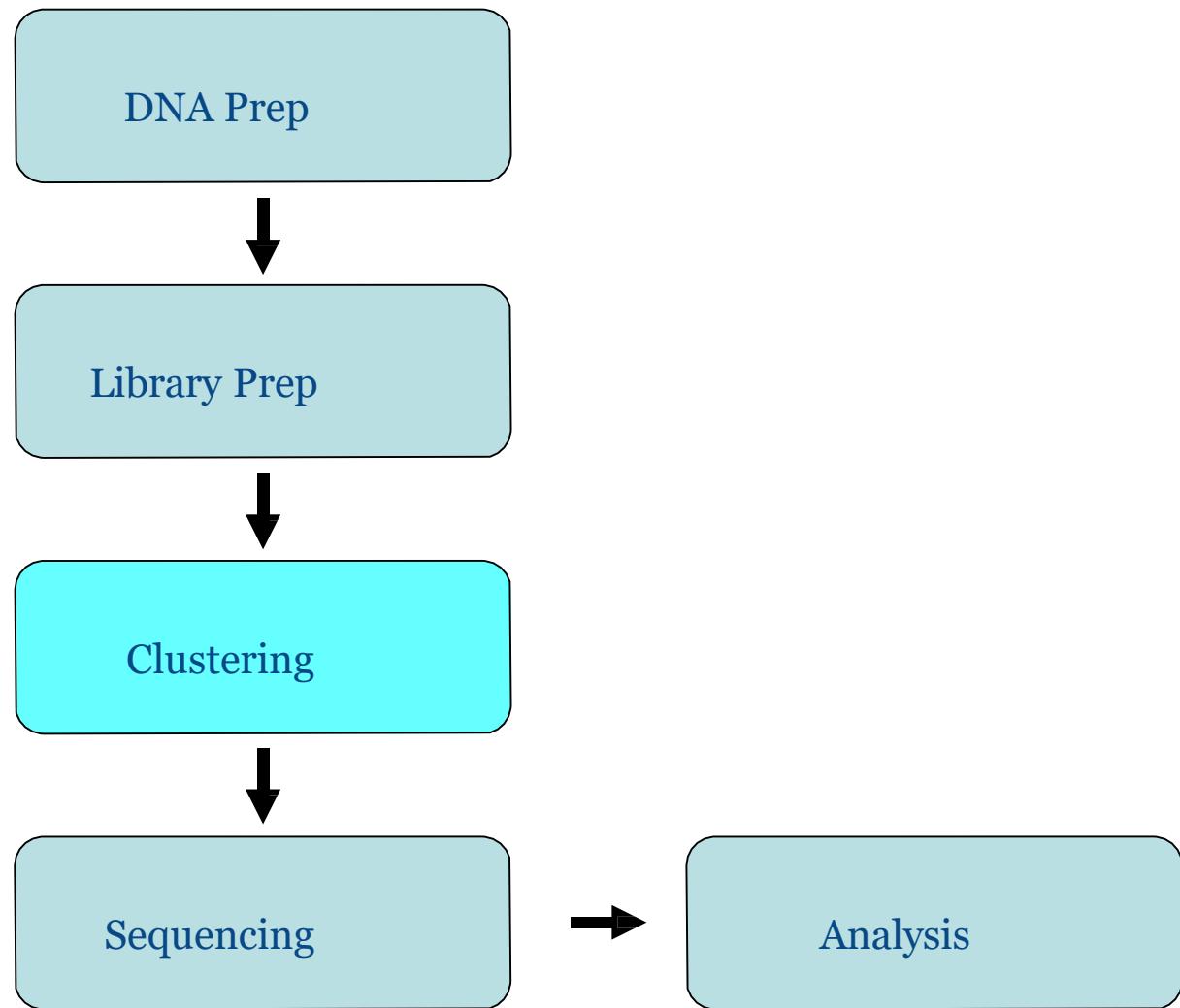
... the best

... sequencing

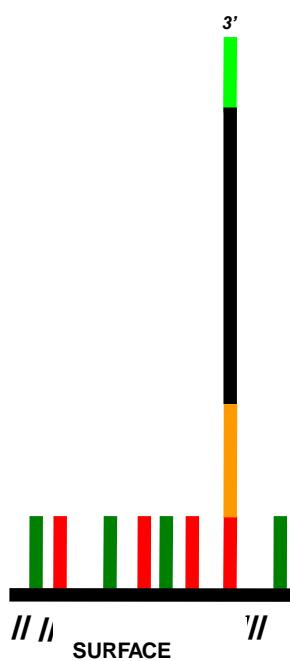
Paired-reads This is (----100 characters-----)sequencing

Assembly becomes easier!!

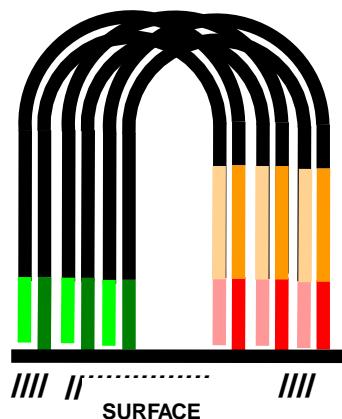
Illumina workflow



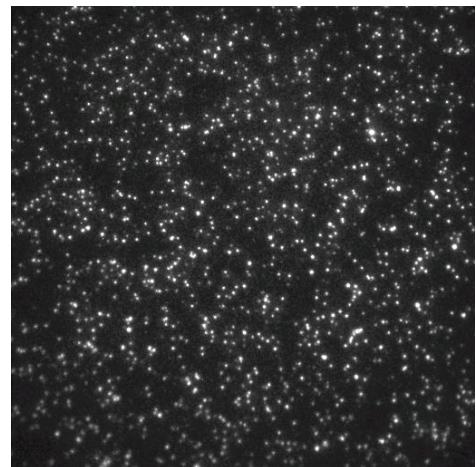
Cluster Amplification



Single-molecule
array



Cluster
~1000
molecules



1.5 Billion
clusters on a
single glass chip

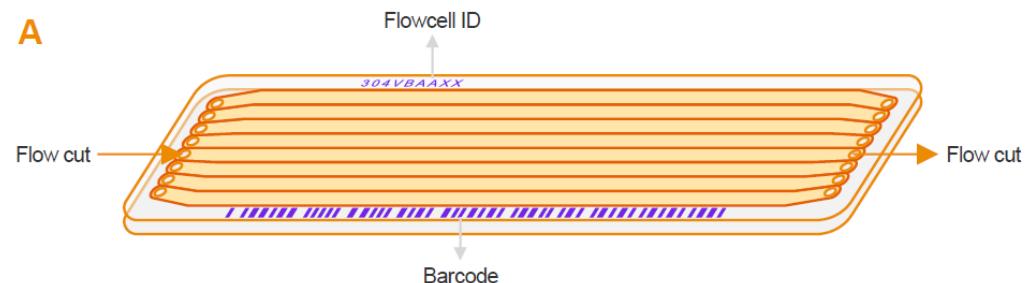


Recall: We Prepared Our Samples So That Nucleotide Fragments Bind a Flow Cell, a Device Where Sequencing Occurs on an Illumina Platform

DNA is hybridized onto the surface of the flow cell

Polymerases extend DNA using fluorescent nucleotides

Cameras detect fluorescent signals across the flow cell



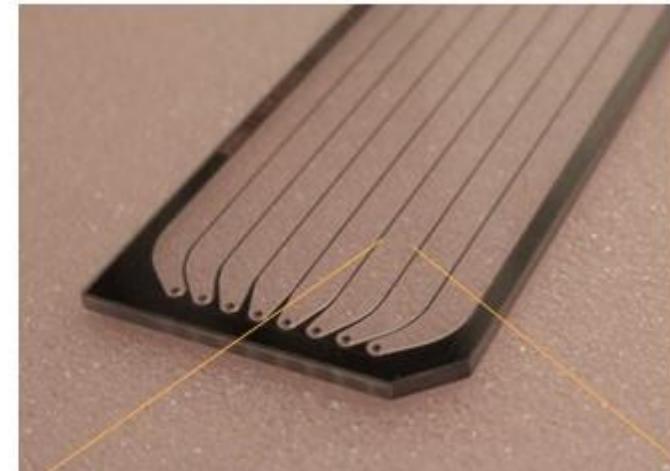
illumina®

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

wellcome
sanger
institute

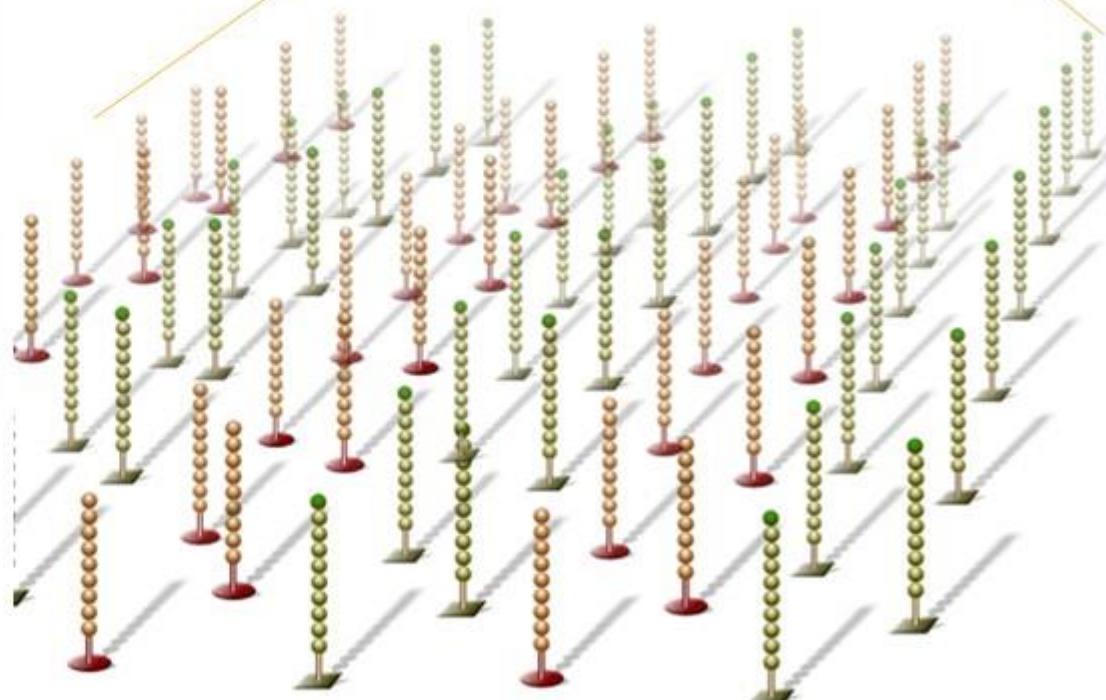
What is a Flow Cell?

Cluster generation occurs on a flow cell



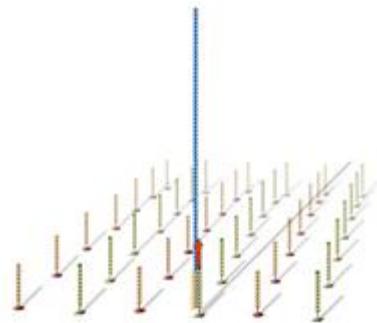
A flow cell is a thick glass slide with channels or lanes

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

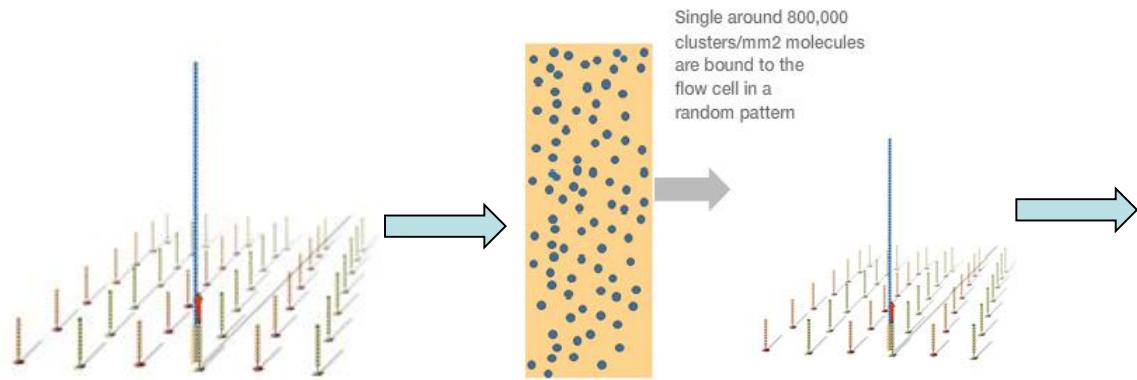




Illumina Sequencing methodology

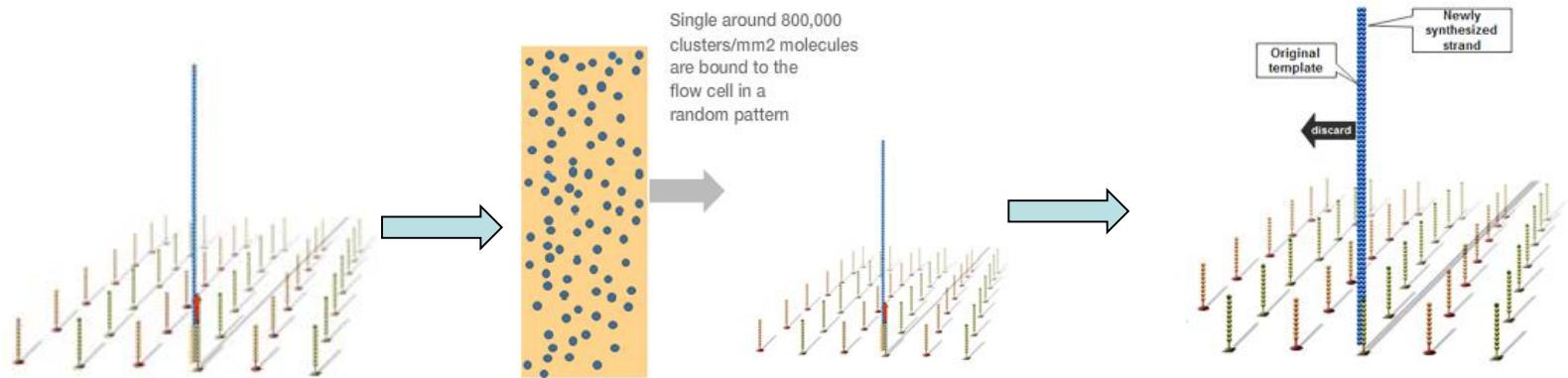


Illumina Sequencing methodology



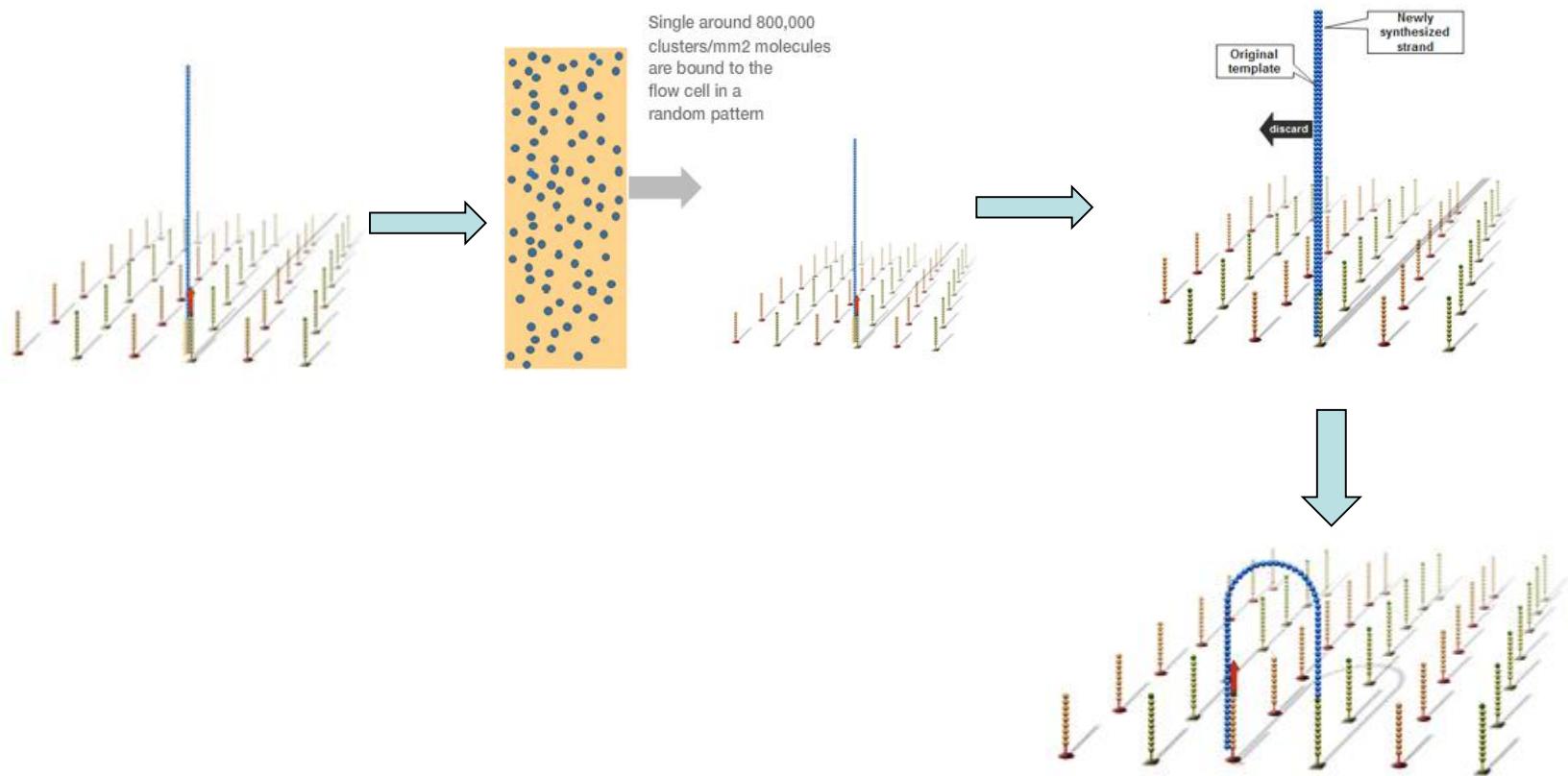
Millions of these single library fragments are bound at random to the flowcell surface

Illumina Sequencing methodology

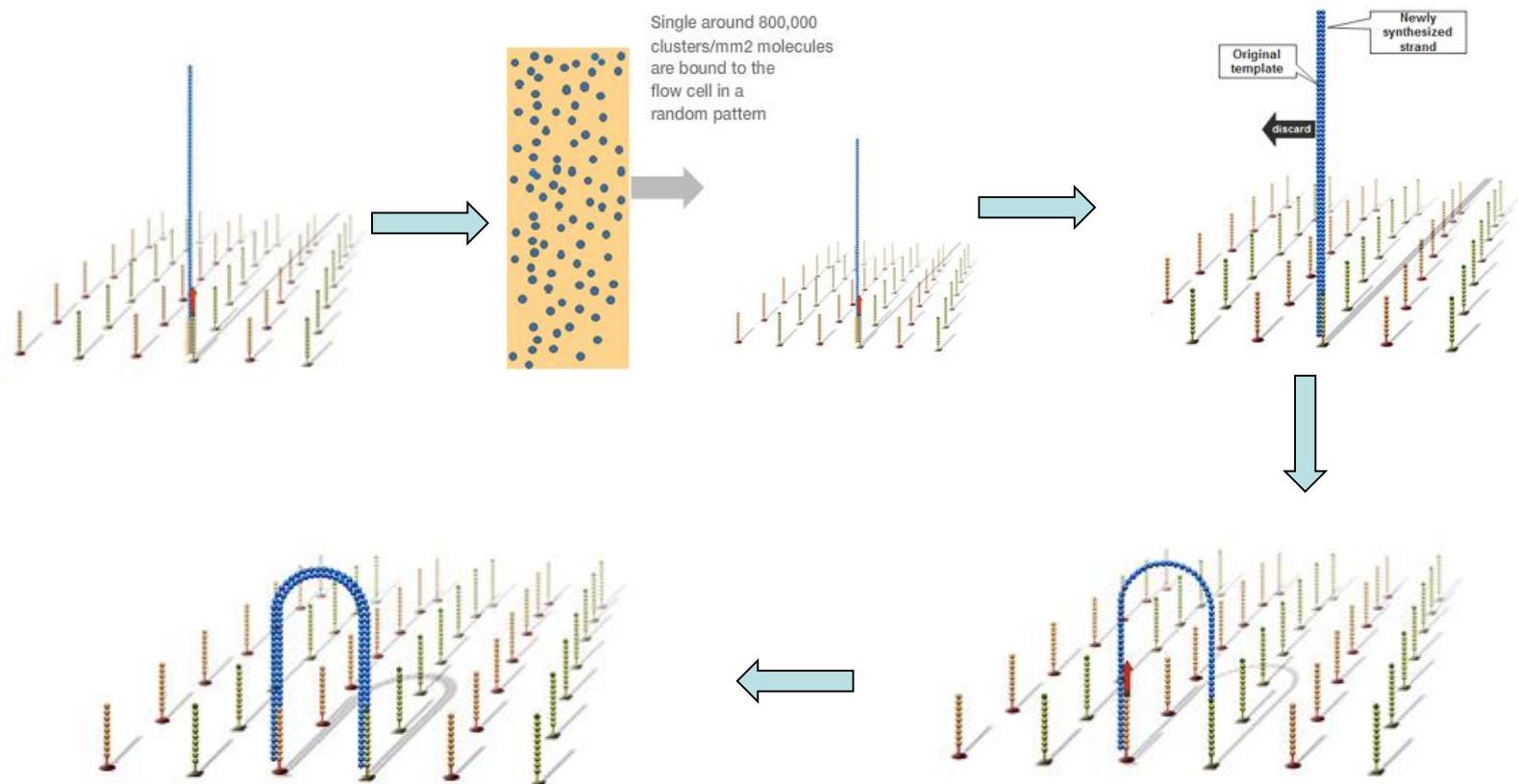


That strand is copied by a DNA polymerase from the primer grafted on the flowcell surface, this duplex is then denatured and the original strand washed away

Illumina Sequencing methodology

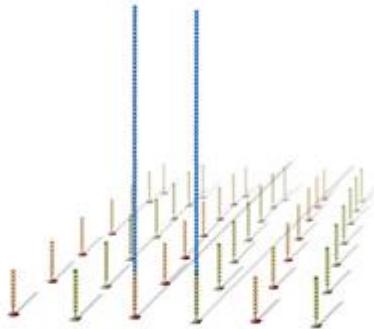


Illumina Sequencing methodology



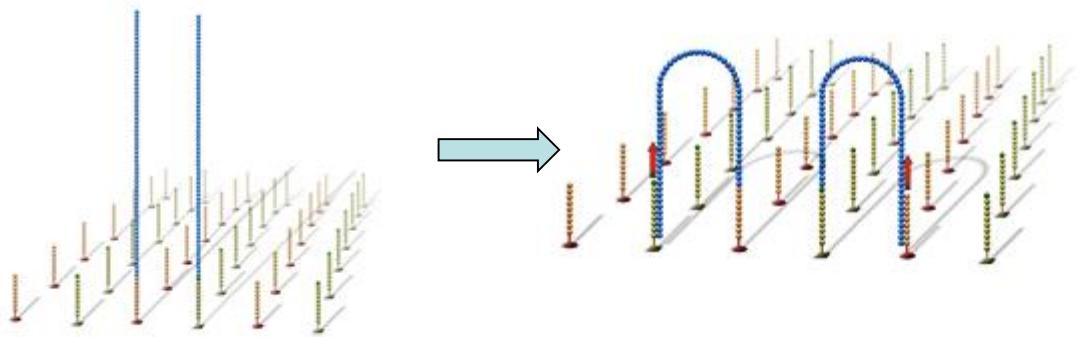
That strand can be copied

Illumina Sequencing methodology



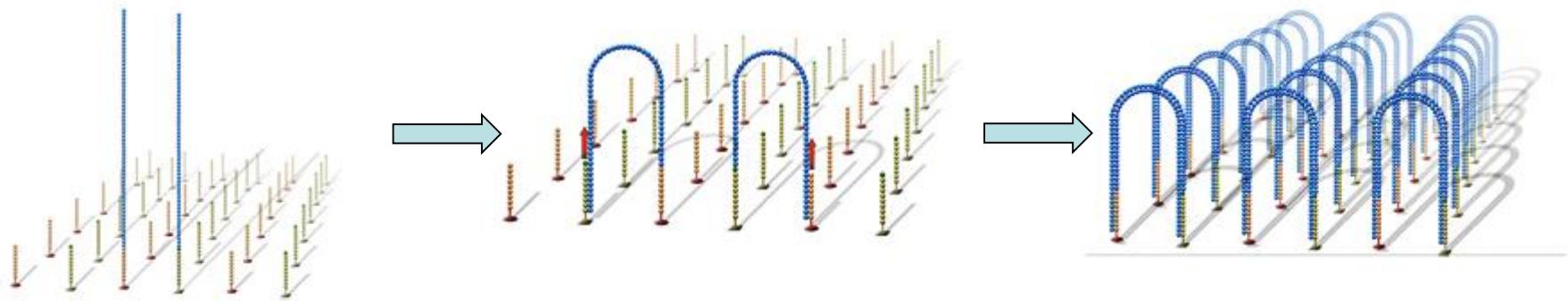
After denaturation you then have two strands bound to the grafted primer lawn.

Illumina Sequencing methodology



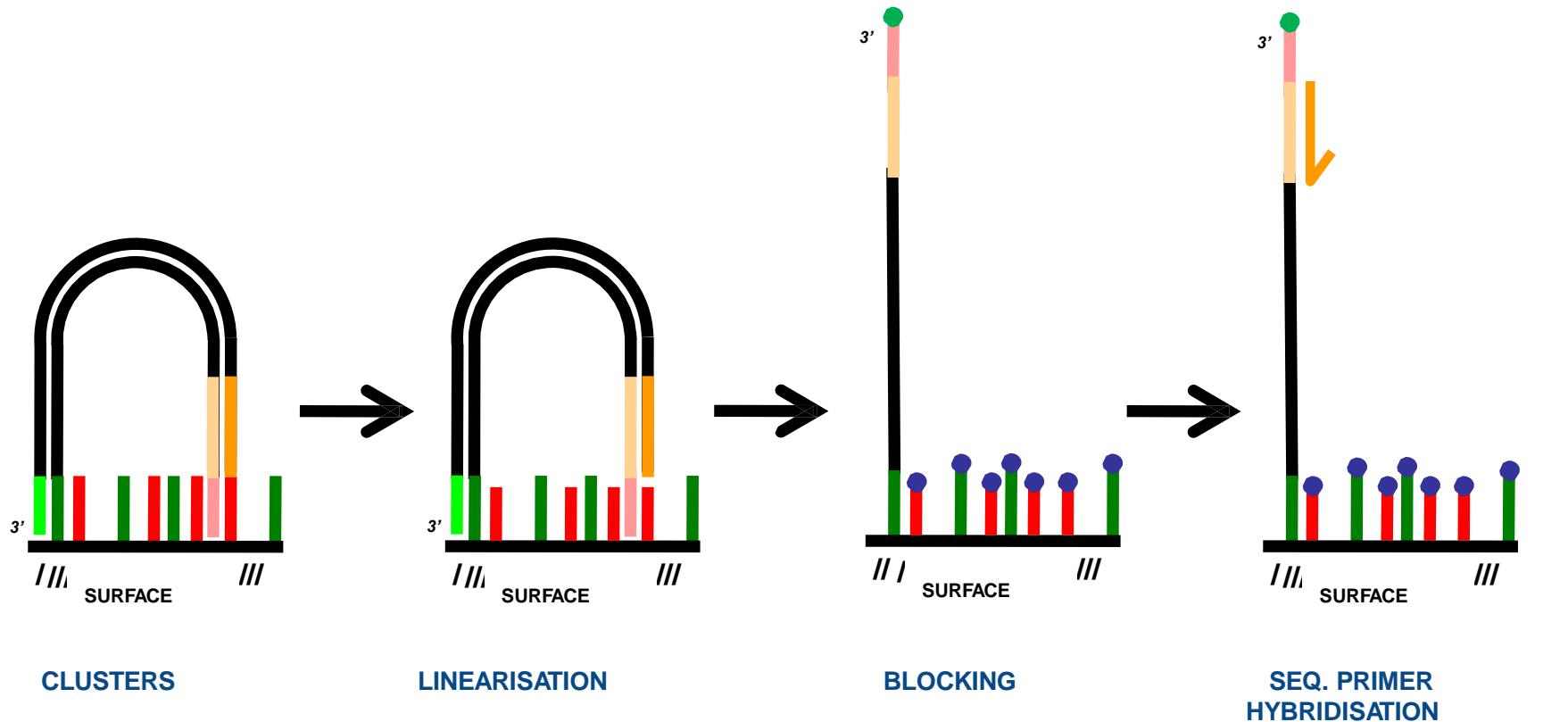
Which can loop over and be copied

Illumina Sequencing methodology



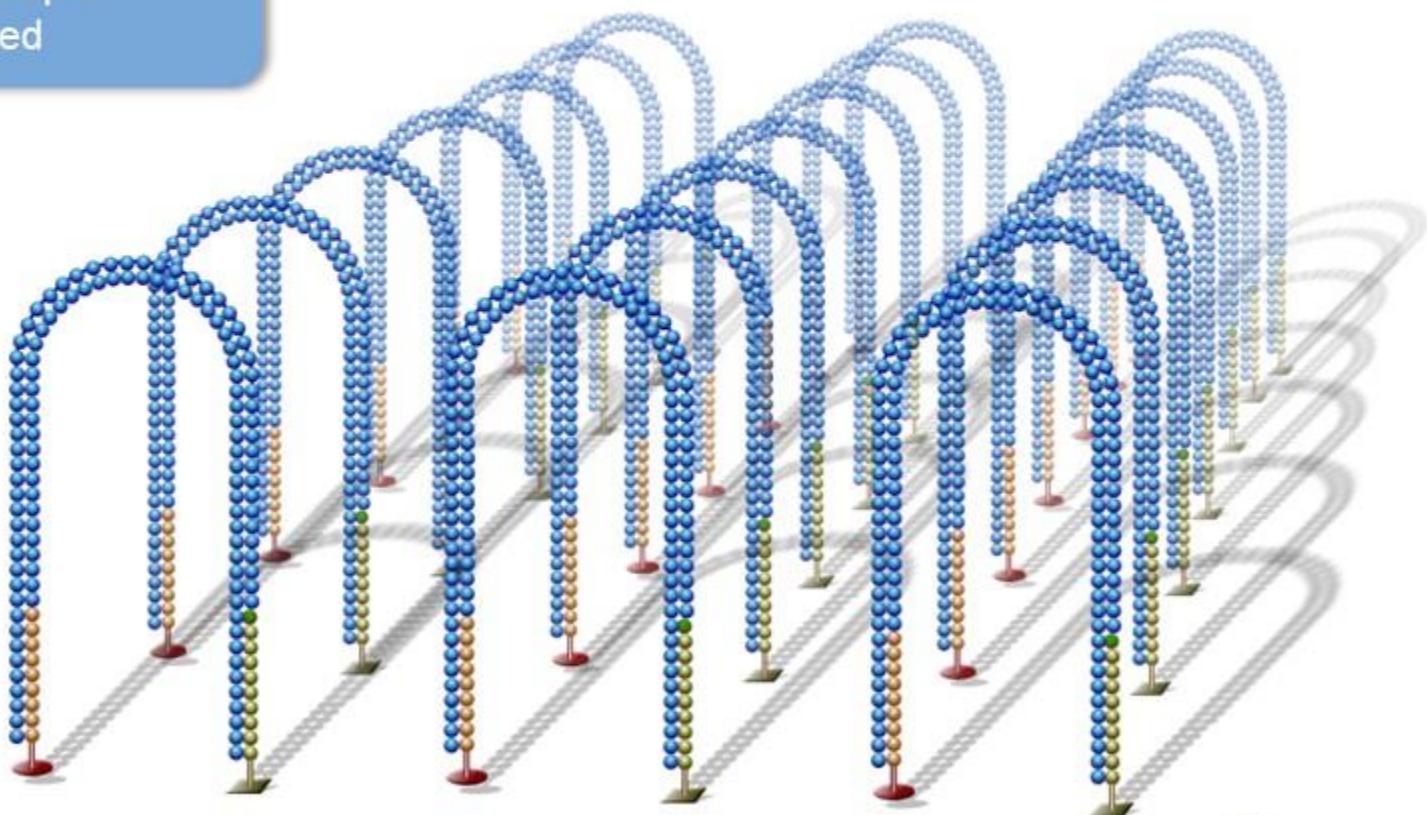
And so on to form clusters of around 1000 copies of the original fragment

Preparing Clusters for Sequencing read 1



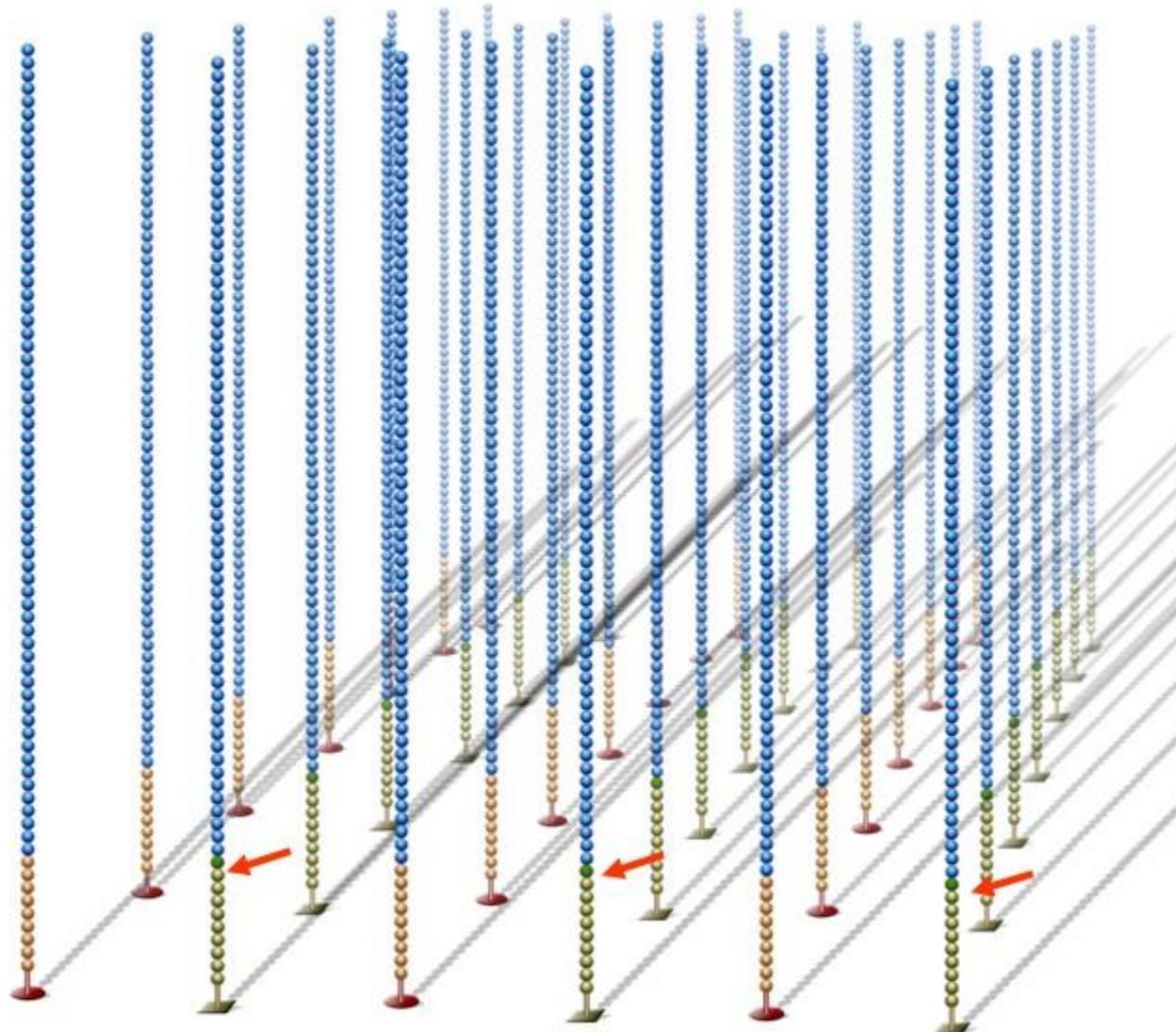
Bridge Amplification

Bridge amplification cycle is repeated until multiple bridges are formed



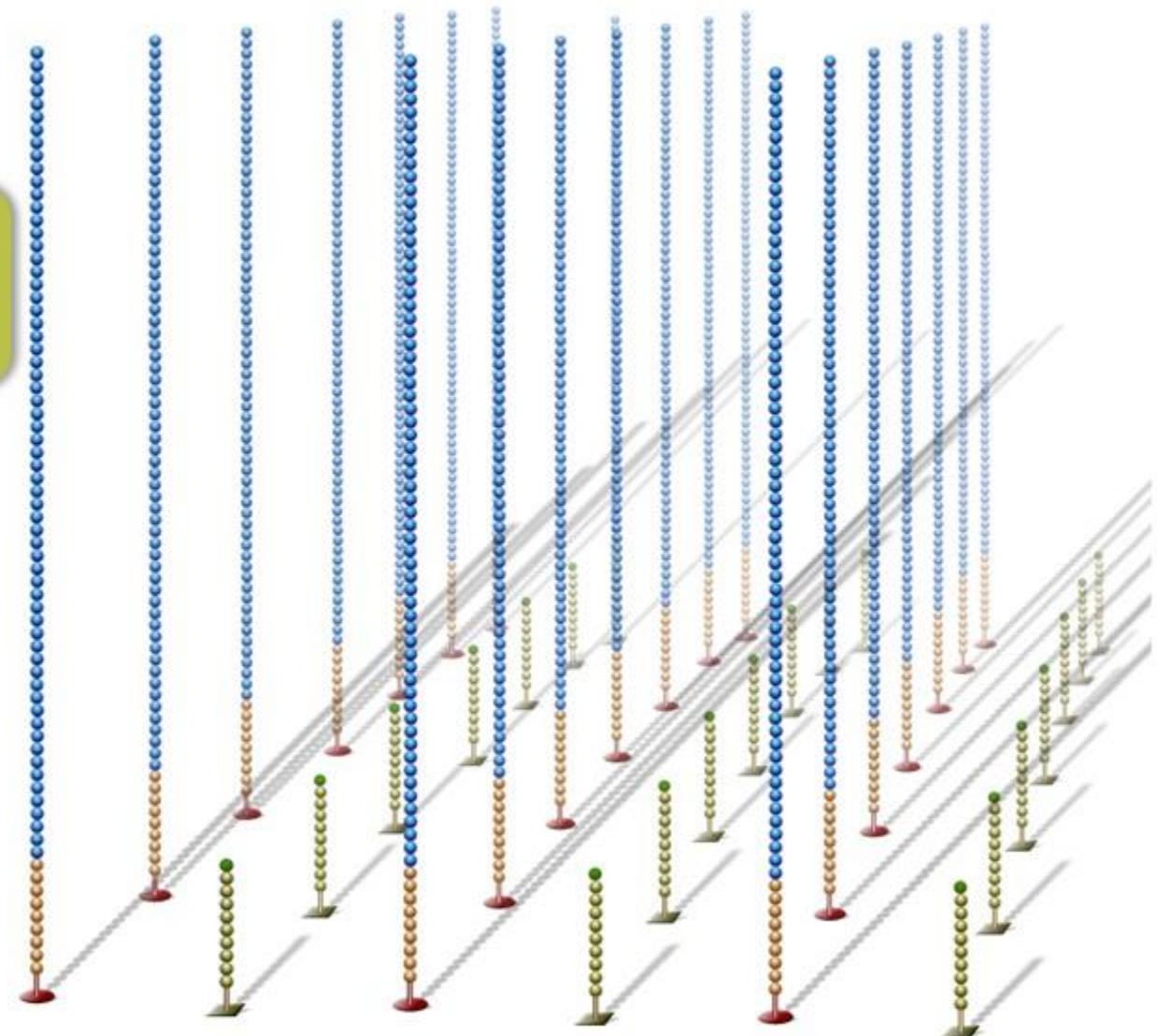
Linearization

dsDNA bridges are denatured



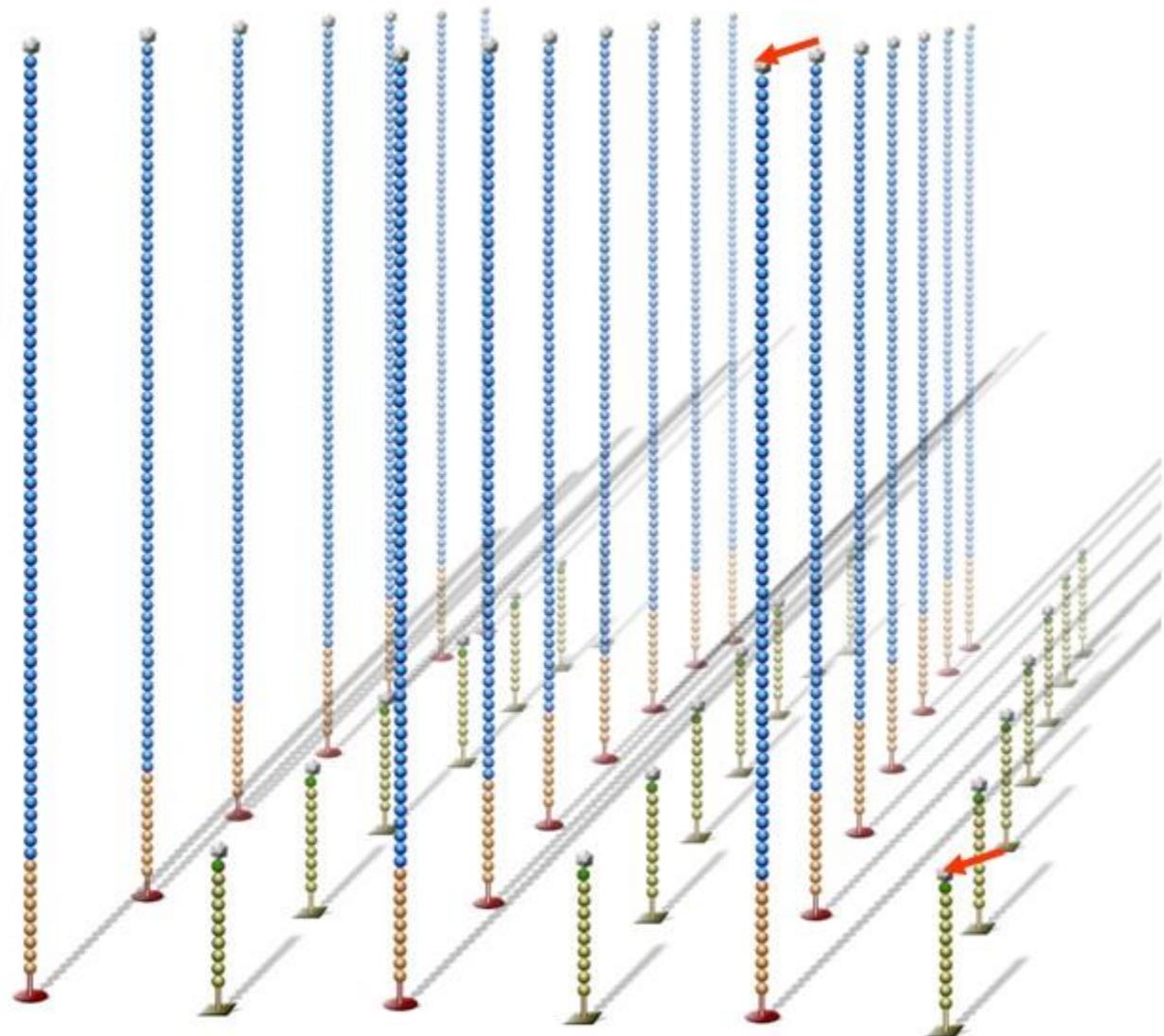
Reverse Strand Cleavage

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



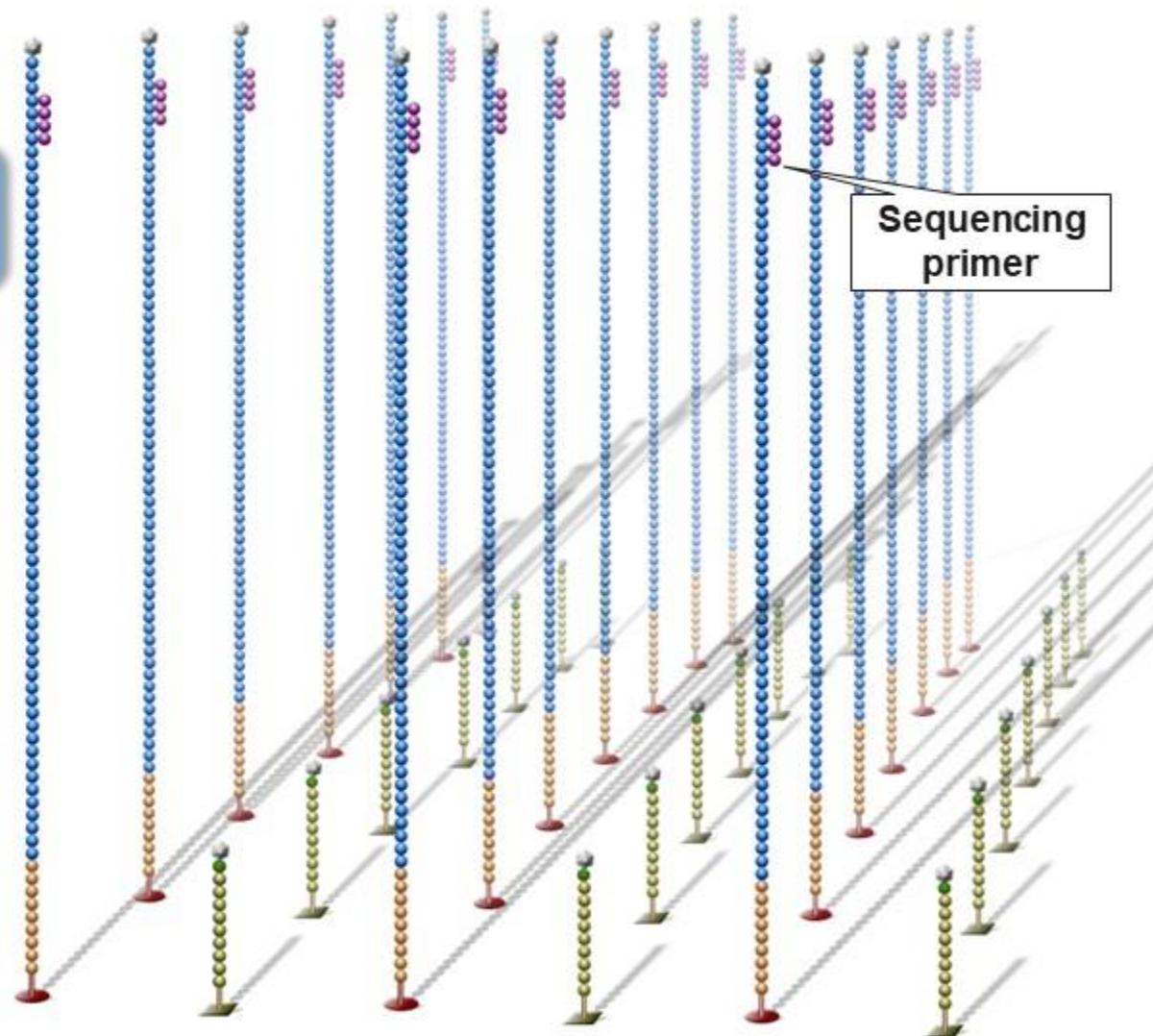
Blocking

Free 3' ends are blocked to prevent unwanted DNA priming



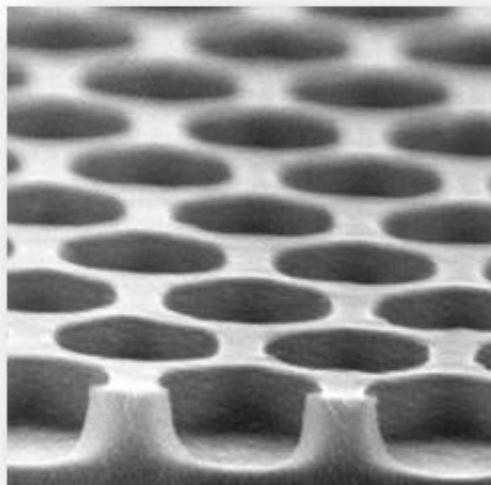
Read 1 Primer Hybridization

Sequencing primer is hybridized to adapter sequence



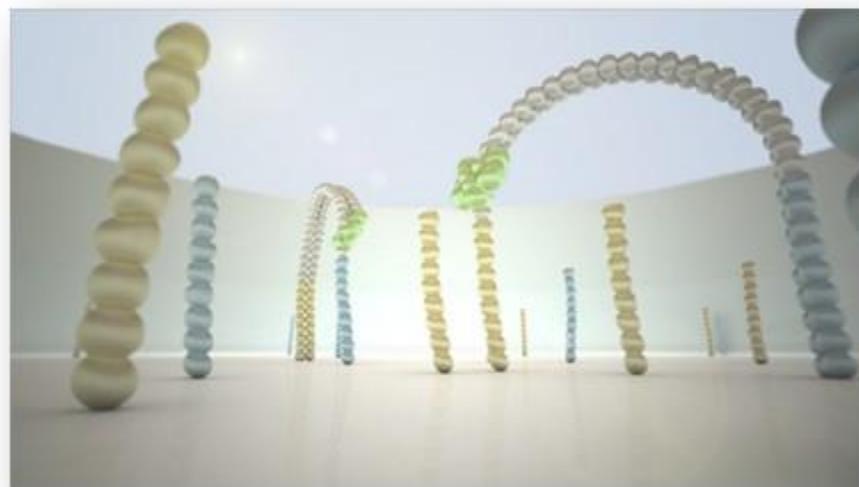
Patterned Flow Cell Technology

iSeq100, HiSeq4000, Xten, NovaSeq



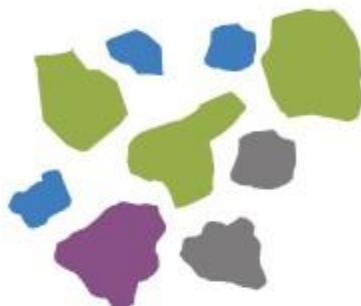
Billions of ordered wells:

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



Patterned Flow Cells

Complete control of pitch & feature size

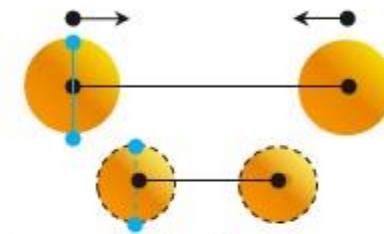


Pitch (Center to Center)

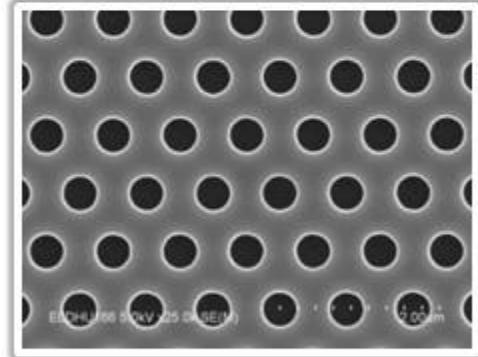
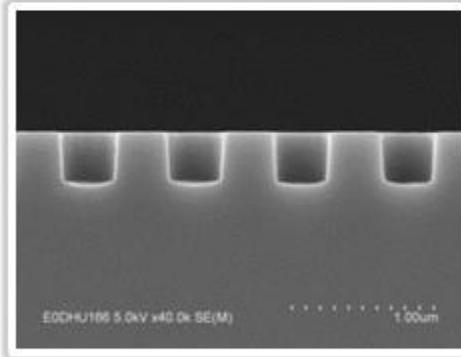
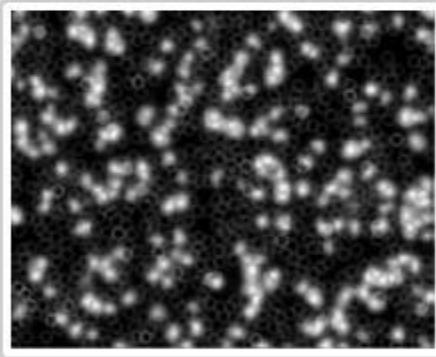


Feature Size

Decrease Pitch



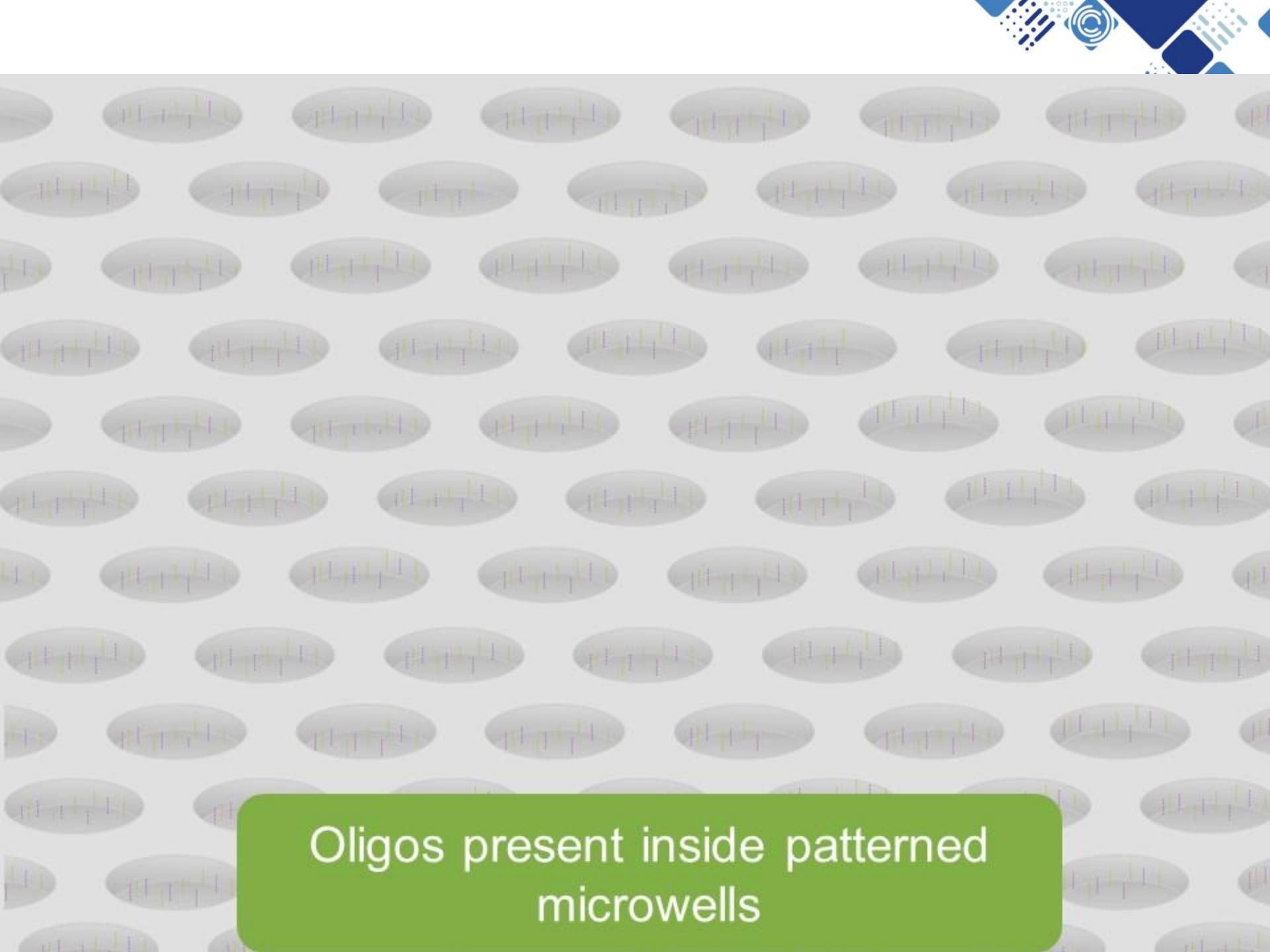
Decrease Feature Size



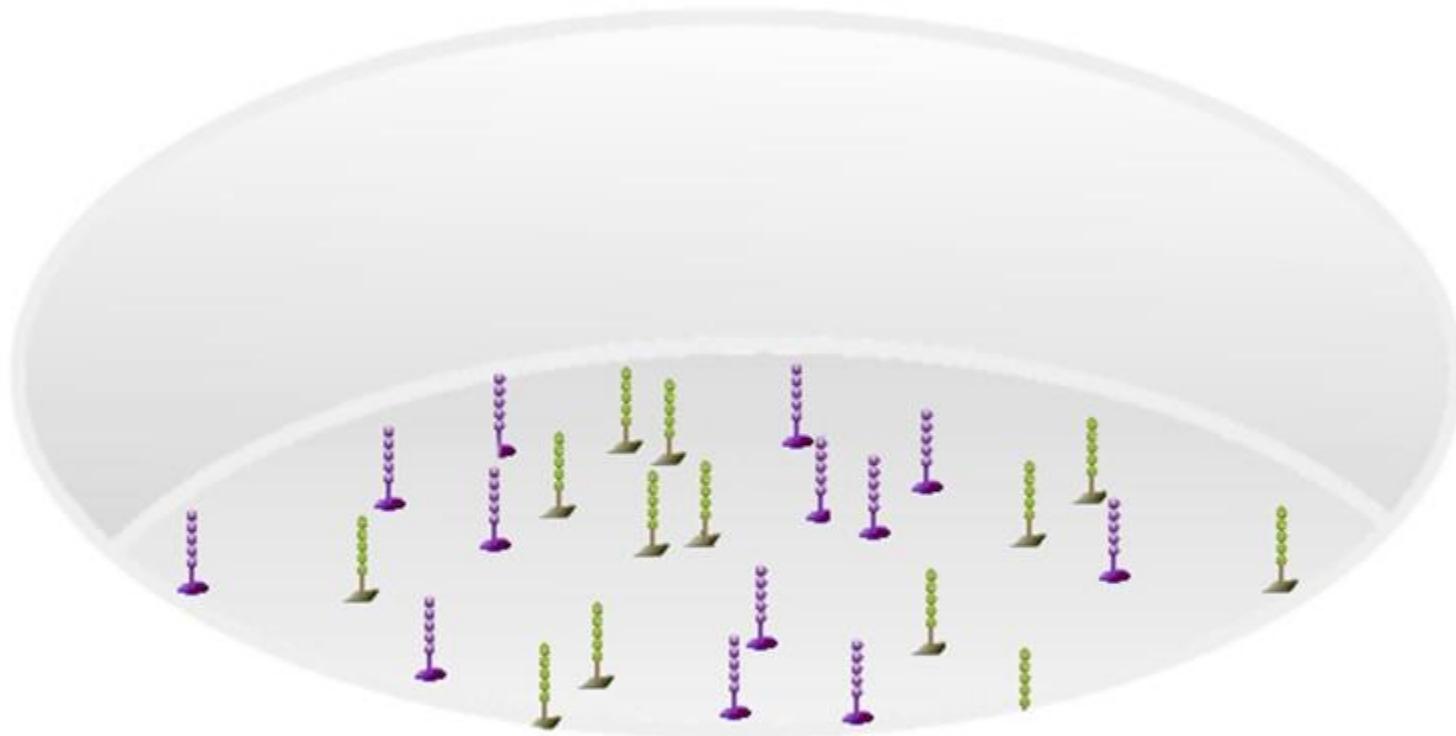
Random

Patterned

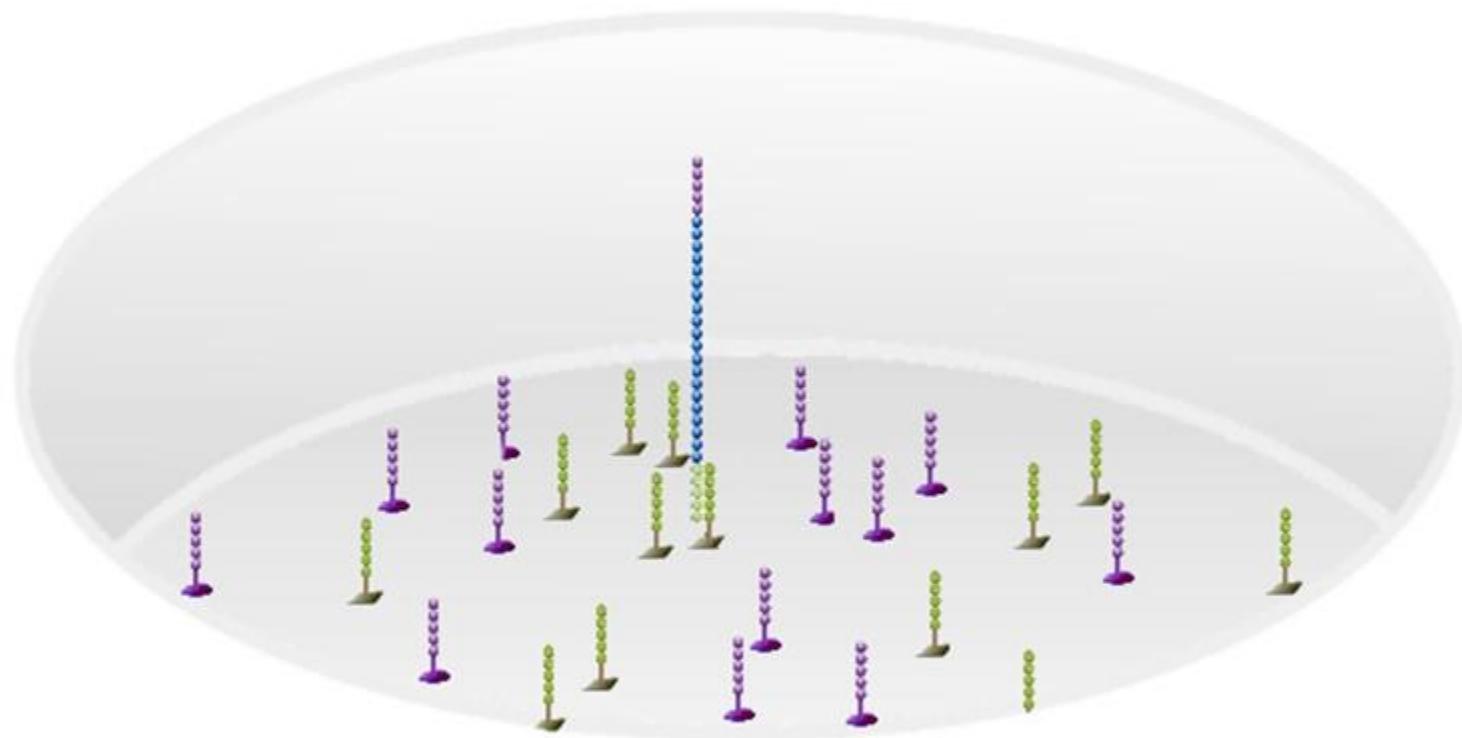
Rigid registration reduces time by skipping template generation



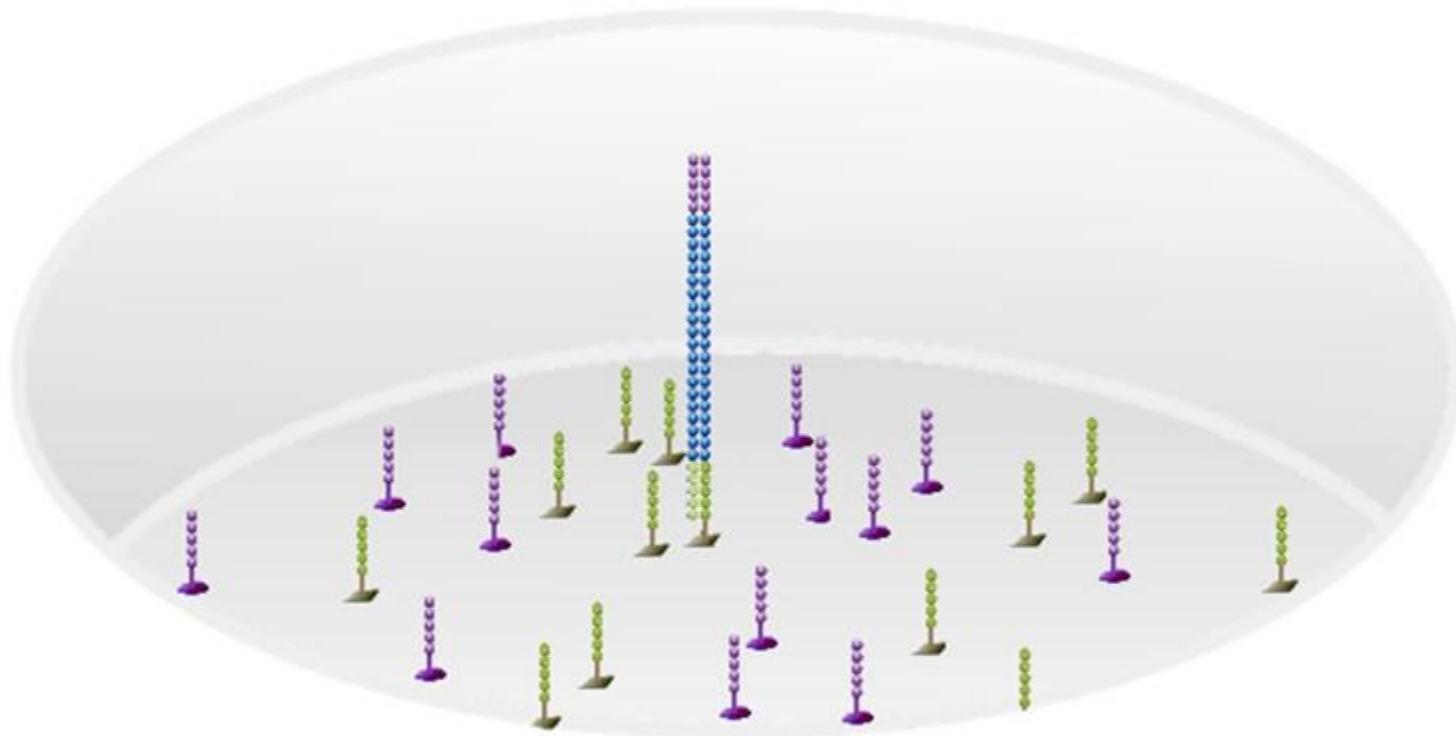
Oligos present inside patterned
microwells



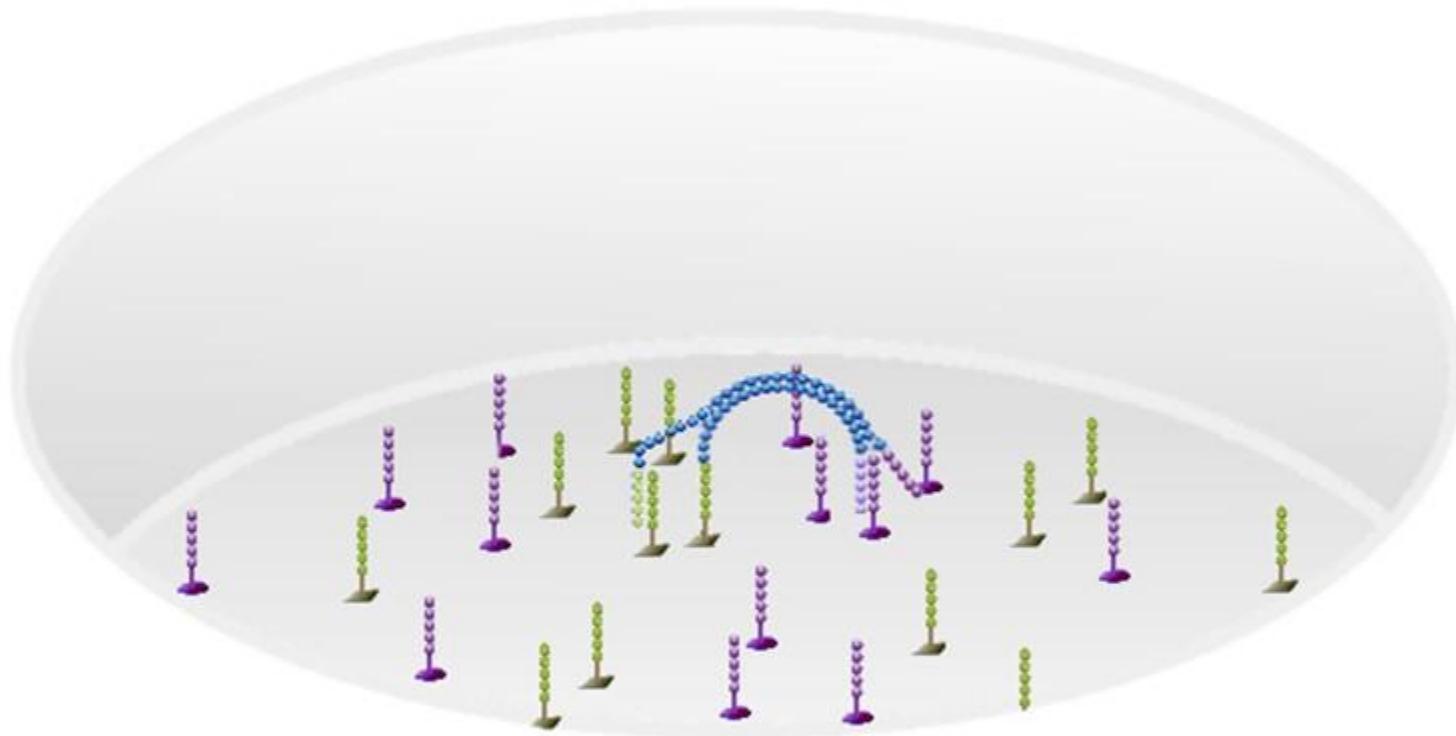
Oligos are not present on surface
between wells to control cluster size



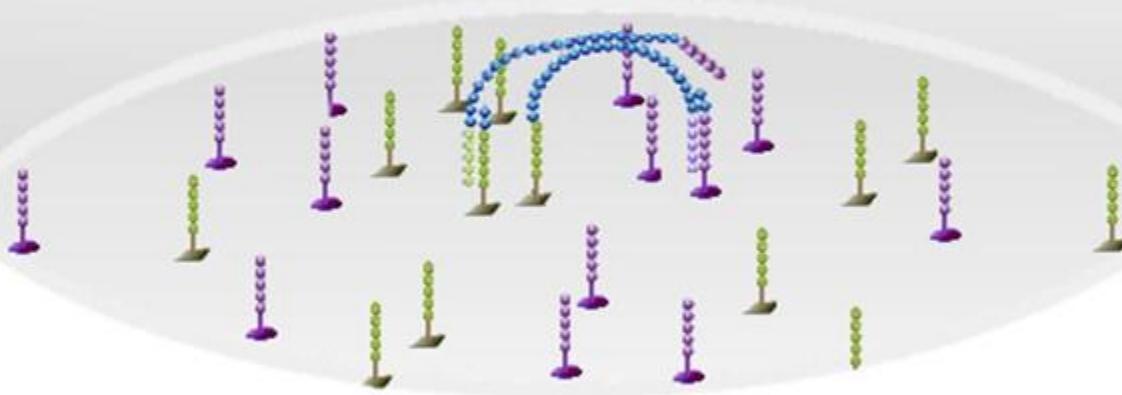
Single stranded DNA is introduced to the flow cell, which hybridizes to an oligo present in the microwell



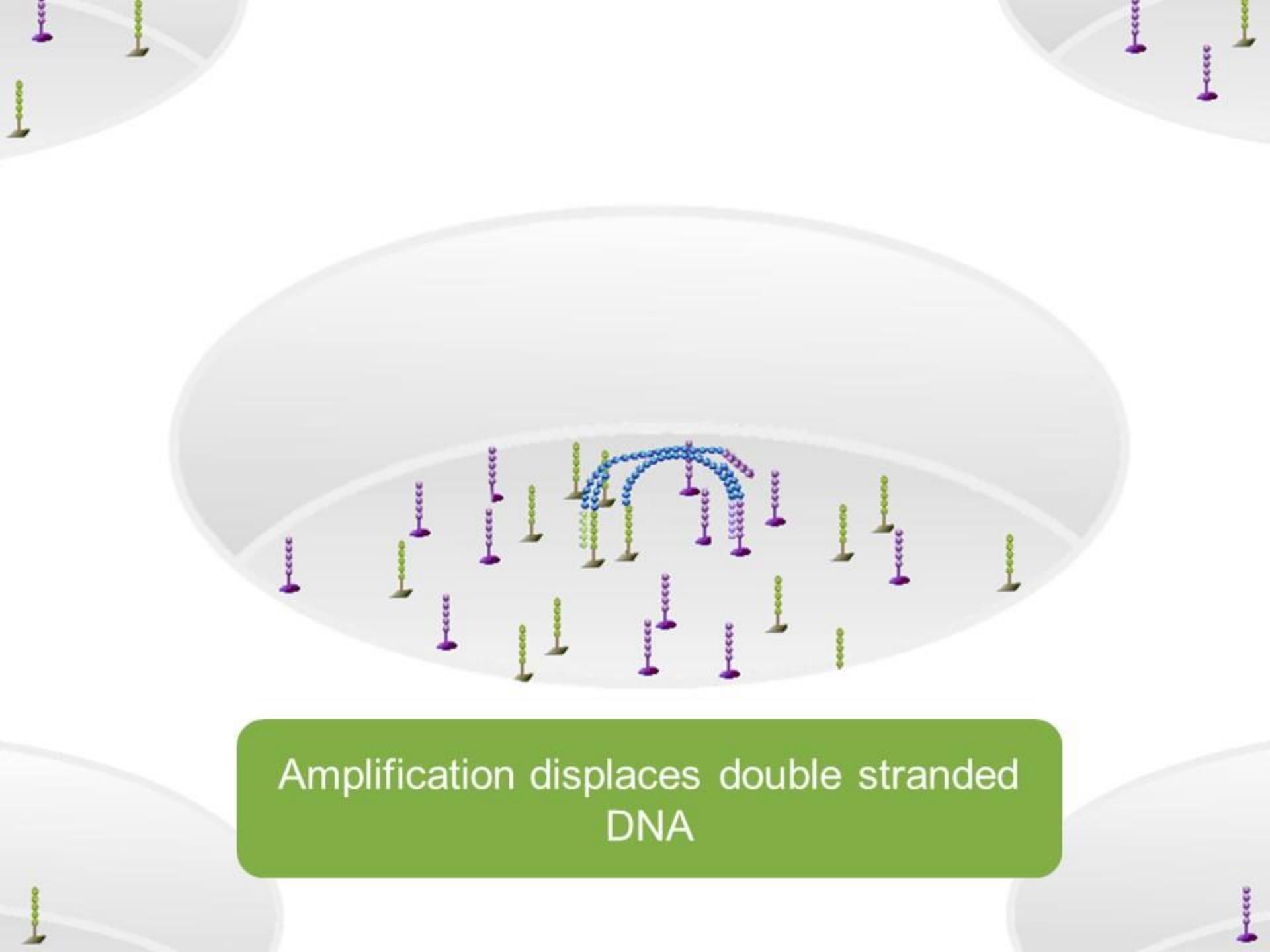
When hybridization occurs amplification immediately begins



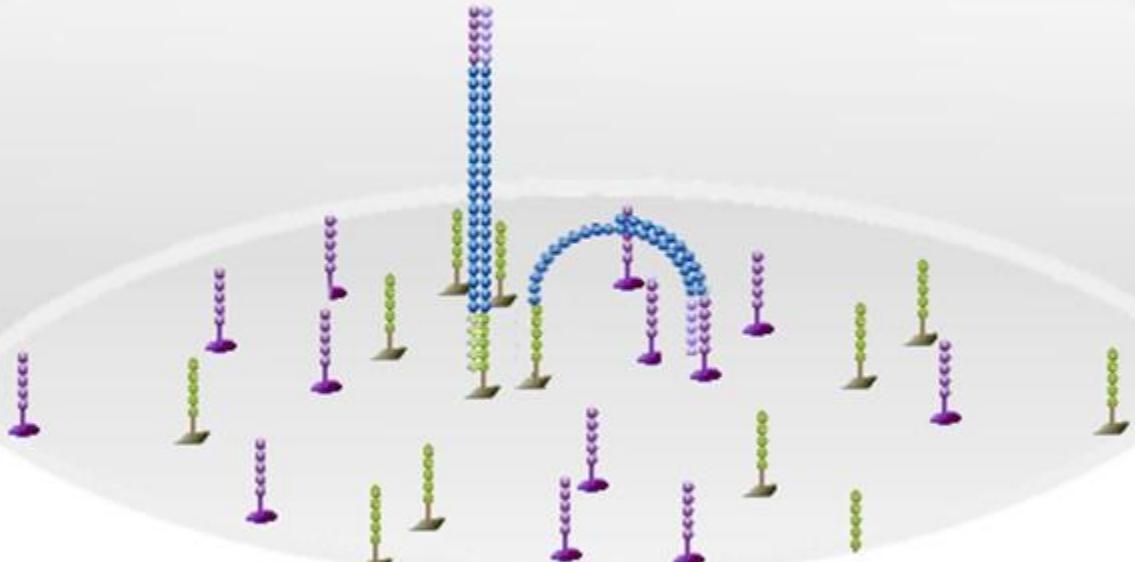
Through a proprietary process, double stranded ends of the DNA are denatured and hybridize to the flow cell



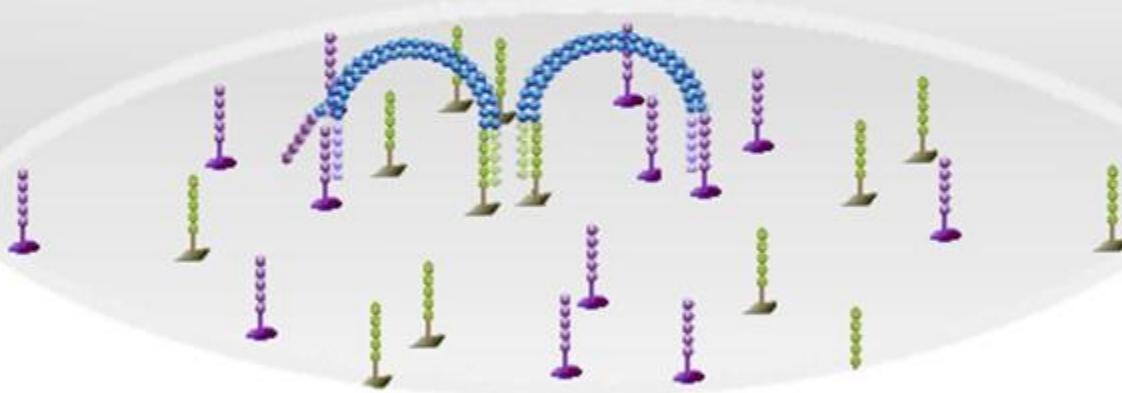
Amplification occurs on both strands simultaneously

A diagram illustrating the process of DNA amplification on a microarray surface. A central oval represents a microarray spot where several short, single-stranded DNA molecules (represented by purple and green vertical sticks) are anchored to a surface. A blue double-stranded DNA molecule is shown being replicated, with its strands extending upwards and outwards. This replication displaces the original single-stranded DNA molecules, which are then shown scattered around the central spot.

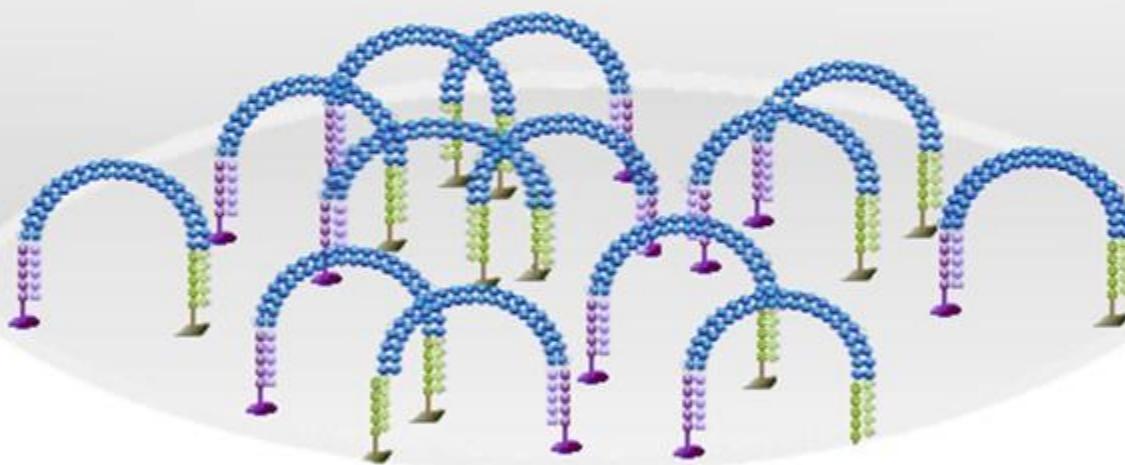
Amplification displaces double stranded DNA



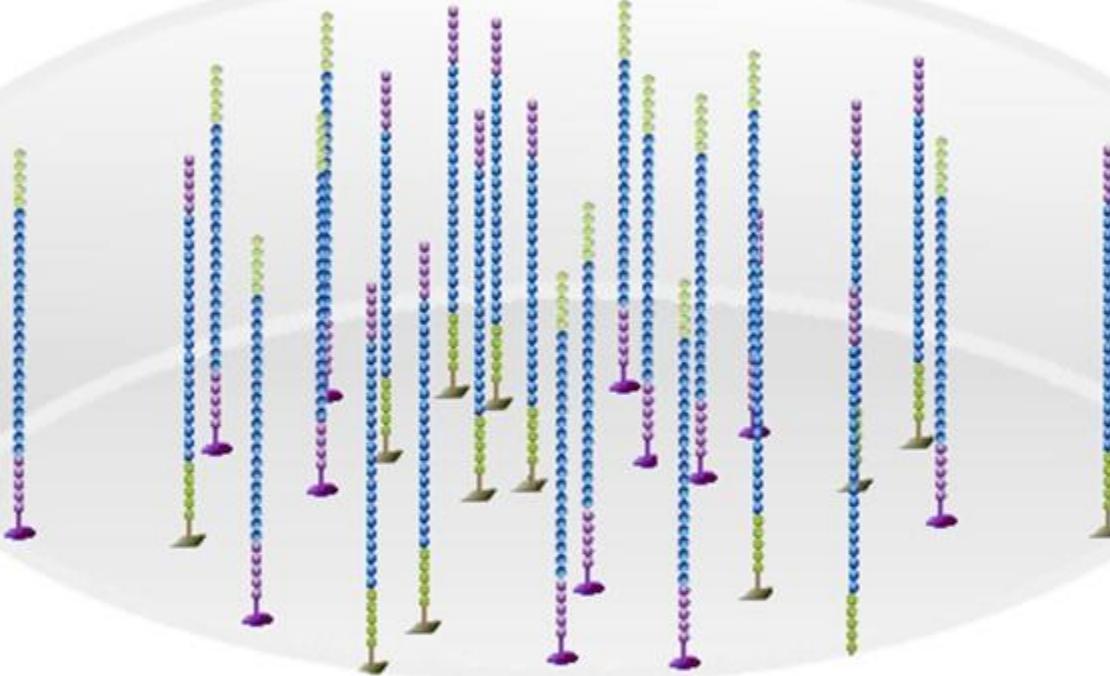
Amplification continues until fully double stranded DNA is formed



Once double stranded DNA is formed, the process repeats itself

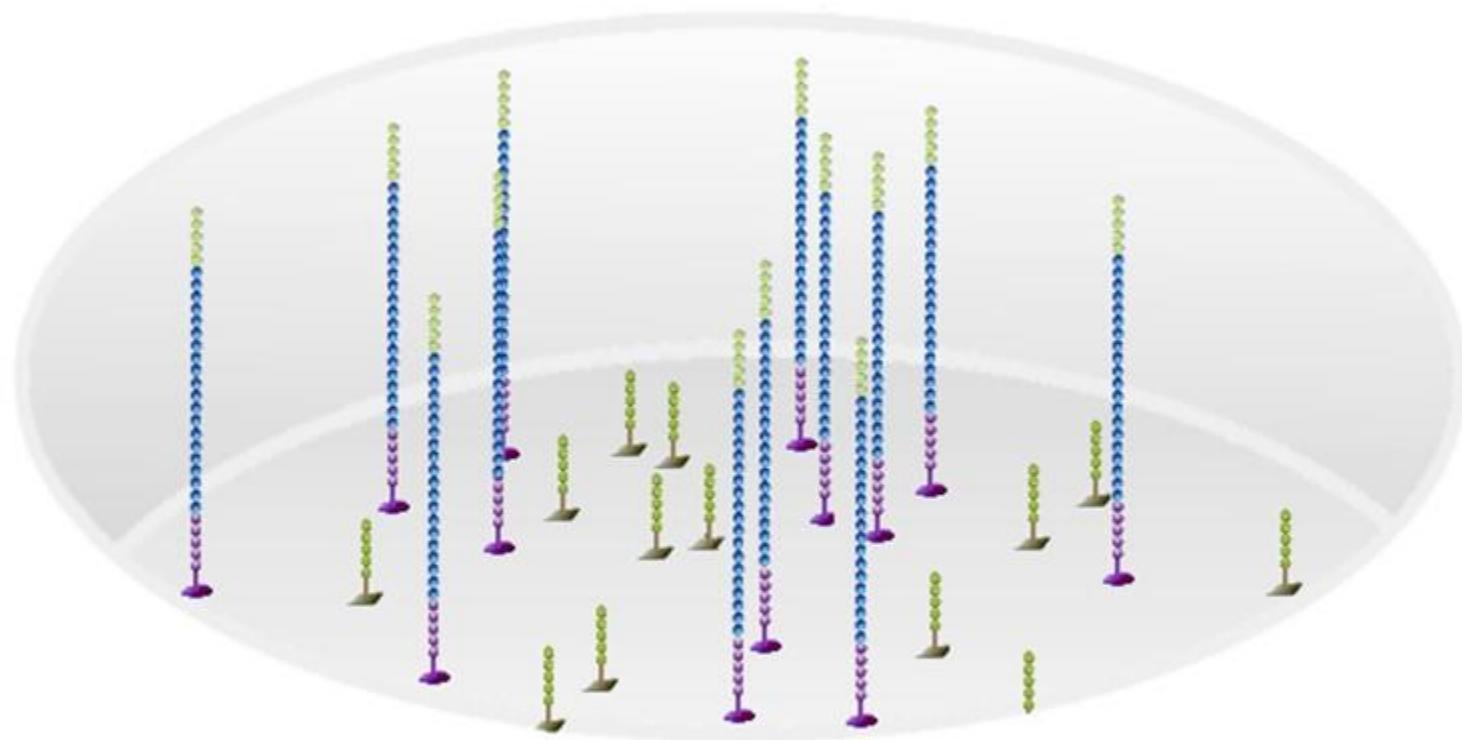


This process continues until no unused oligos are left on the flow cell surface

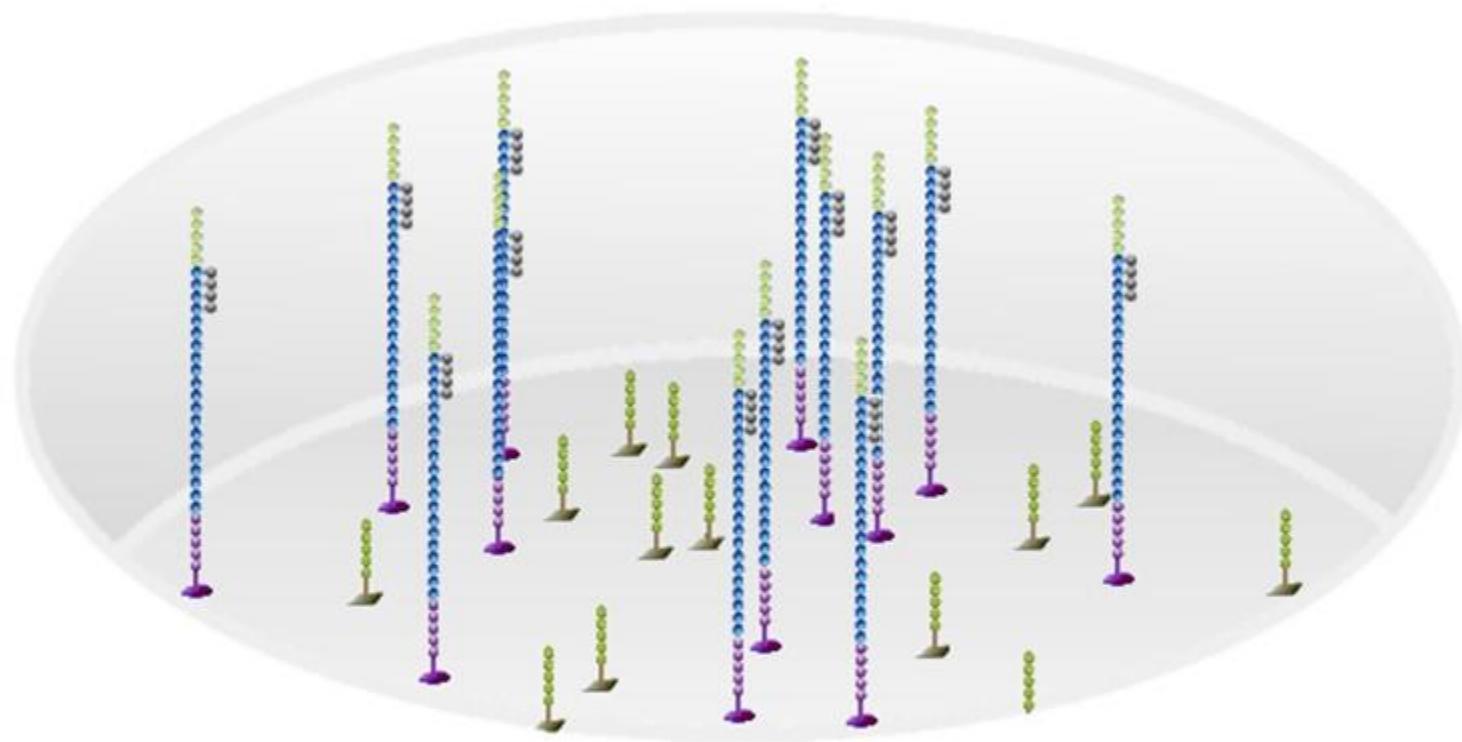


The diagram illustrates a collection of DNA molecules that have been converted into linear strands. These linear strands are represented by vertical lines composed of small, colored circular segments. The segments are color-coded to represent different nucleotides: blue for adenine (A), green for cytosine (C), yellow for thymine (T), and purple for guanine (G). The strands vary in length and orientation, some pointing upwards and others downwards, all anchored at their bases to a light gray surface. This visual representation serves as a metaphor for the linearized DNA molecules used in sequencing processes.

To prepare the strands for sequencing,
linearization is performed



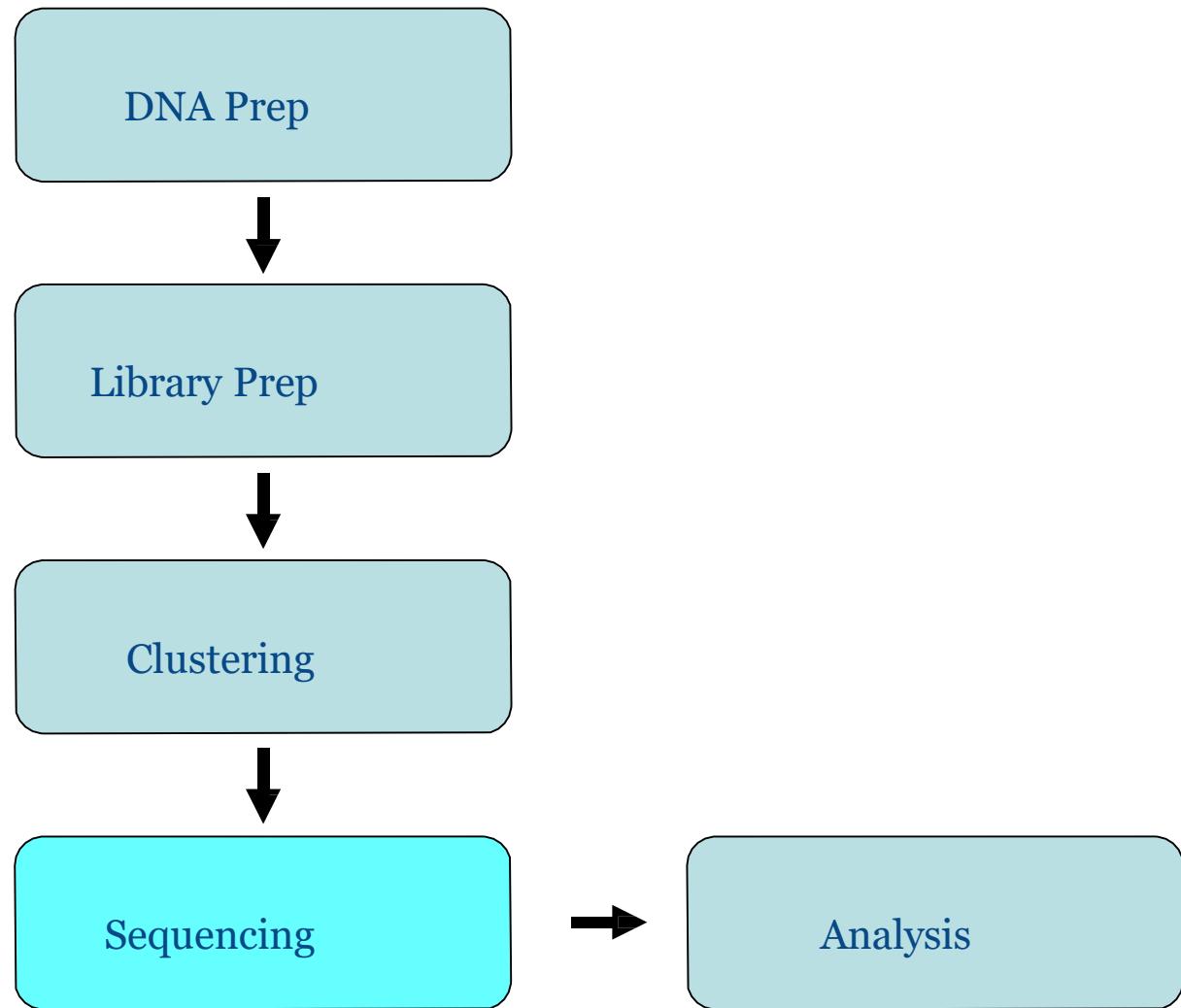
Reverse strands are cleaved



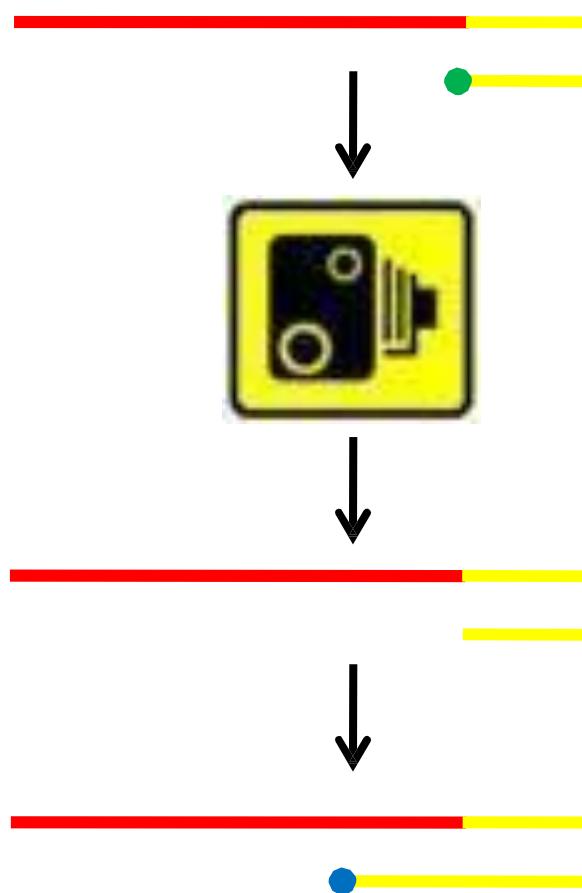
The Read 1 sequencing primer is
hybridized



Illumina workflow



Sequencing by Synthesis

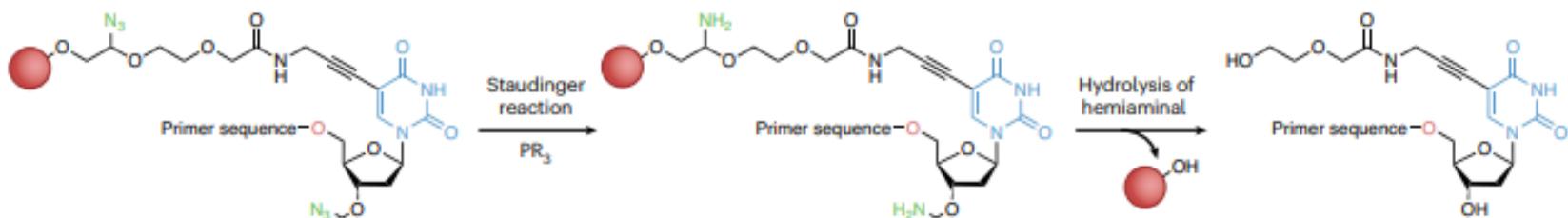
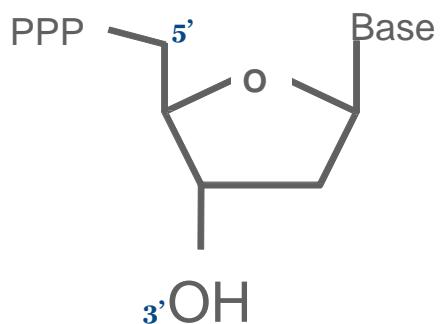


- » Extend by 1 base
- » Image
- » Reverse termination
- » Repeat



Illumina Modified Nucleotides

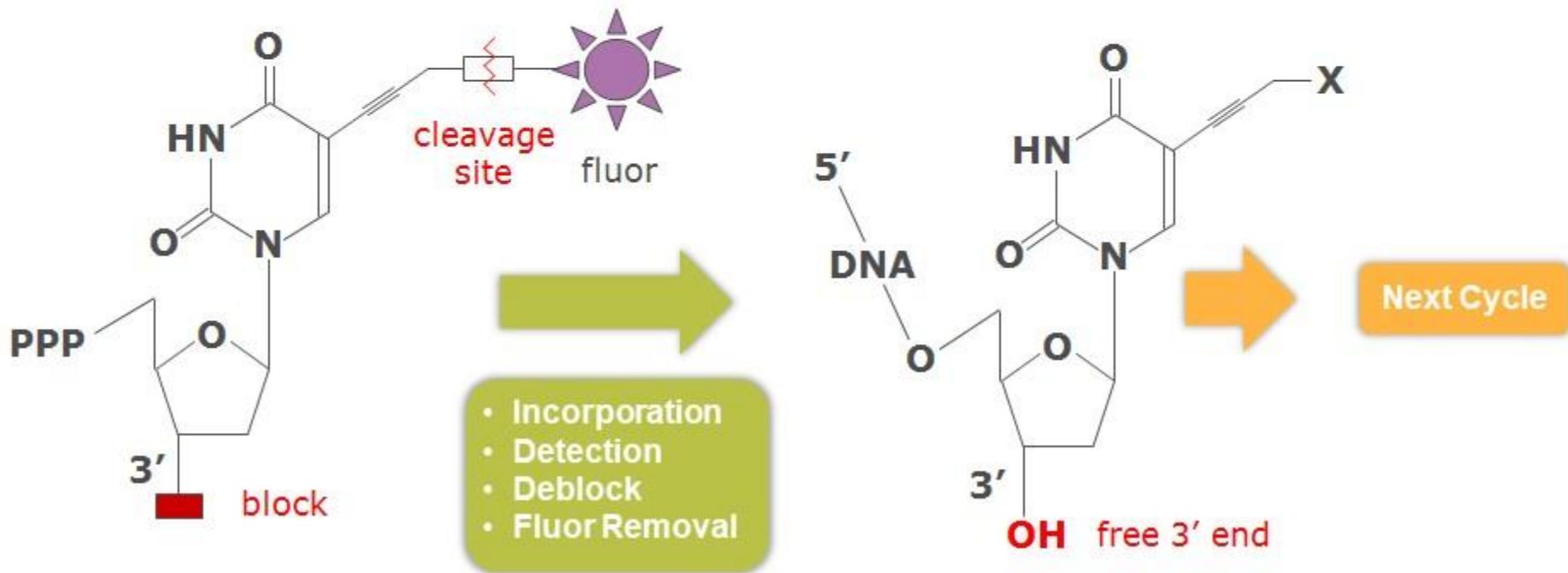
Natural dNTP:



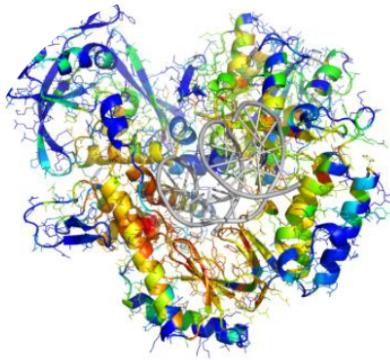
Reversible fluorescent dTTP:

Reversible Terminator Chemistry

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

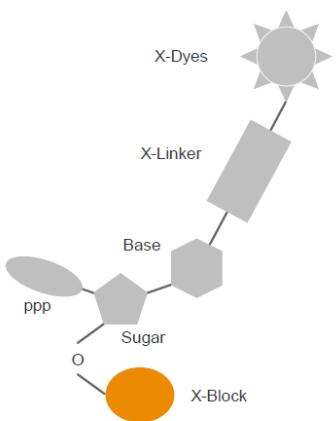


XLEAP-SBS Chemistry



Novel Polymerase

Faster incorporation, higher fidelity



X-Block, X-Linker, X-Dyes

Most resistance to heat, ~50x more stable in solution, ~500x more stable lyophilized, faster block cleave

Enables ambient shipment



Faster cycle times*



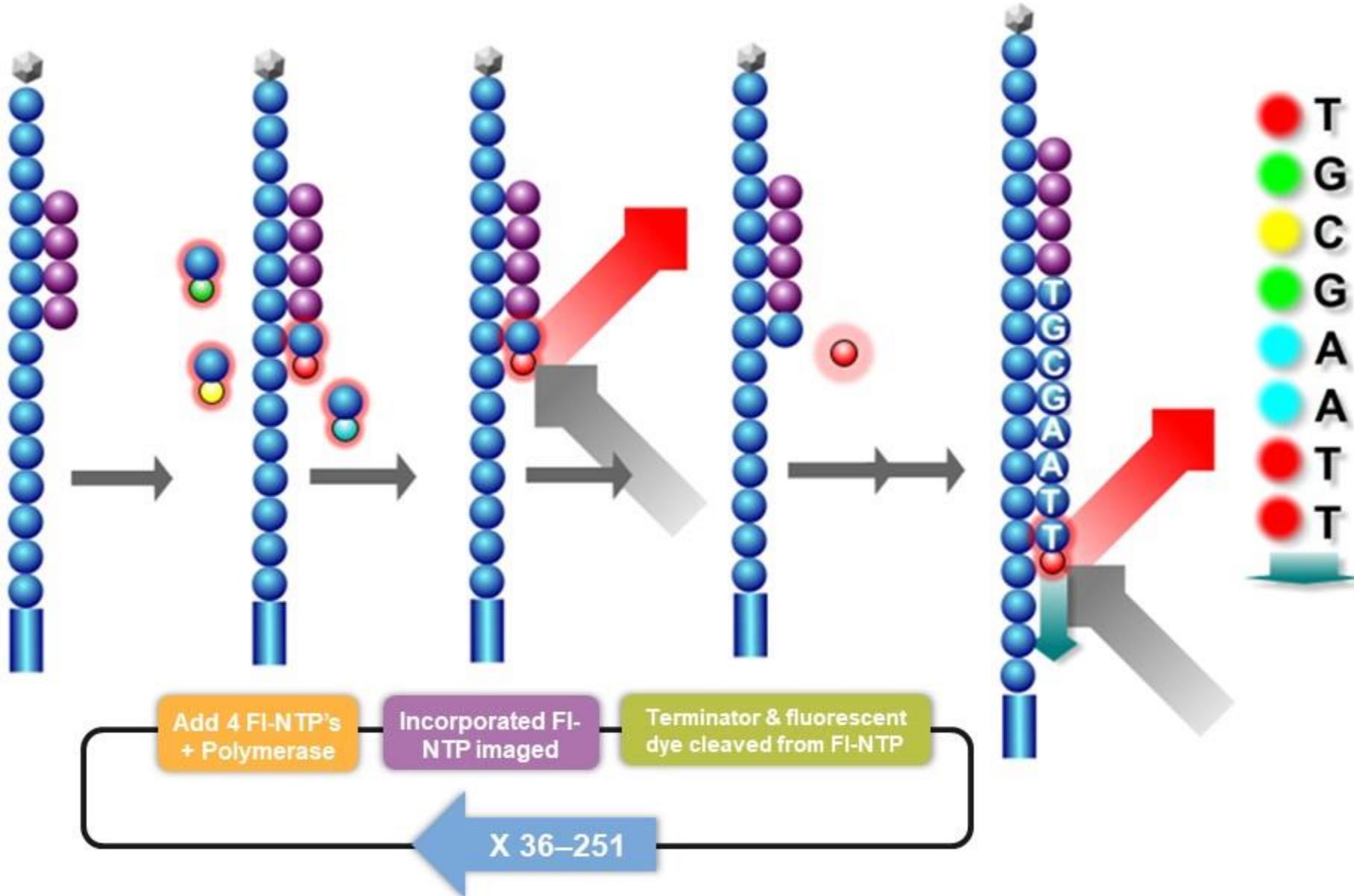
Greater accuracy*



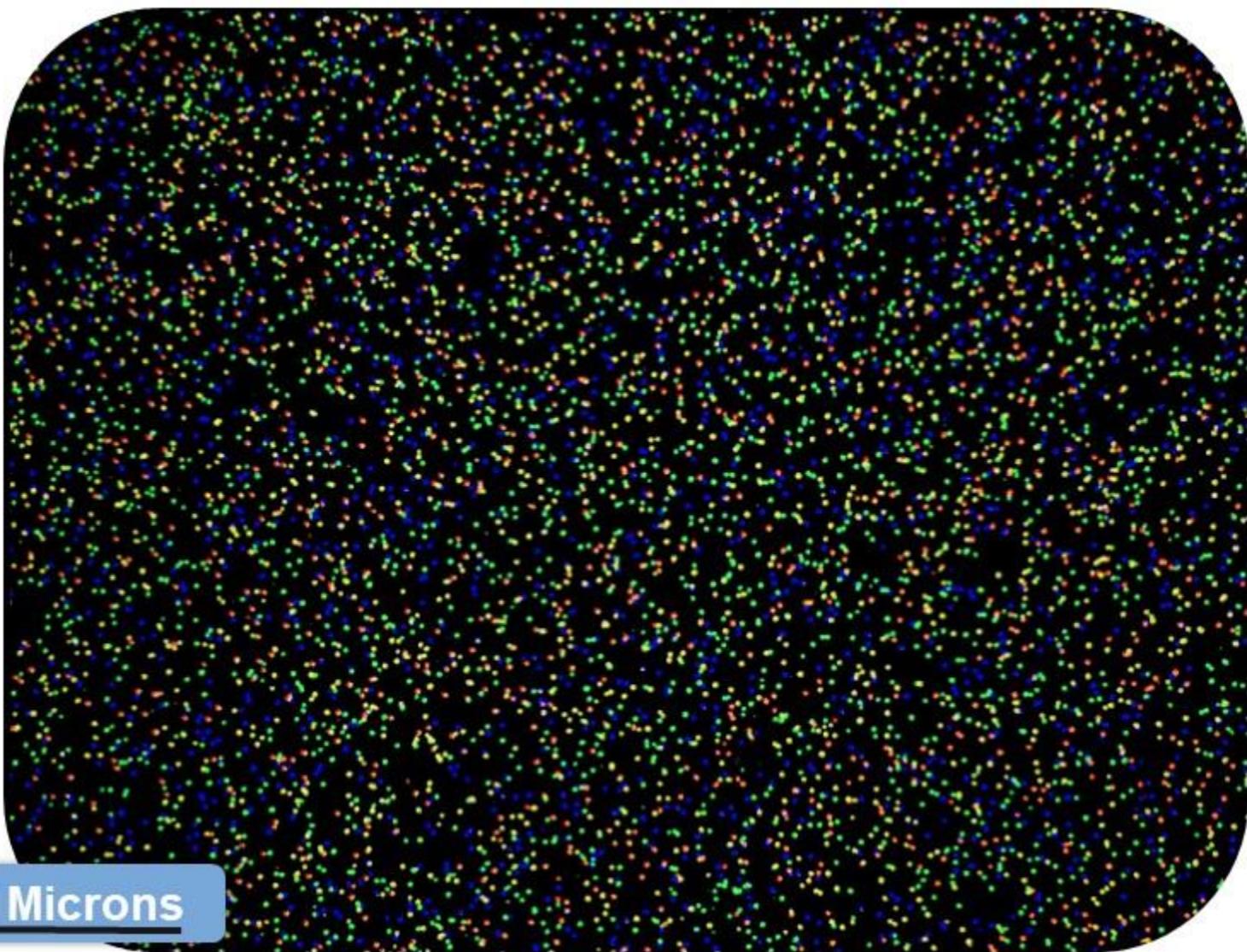
More sustainable*

* as compared to NovaSeq 6000

Sequencing By Synthesis (SBS)

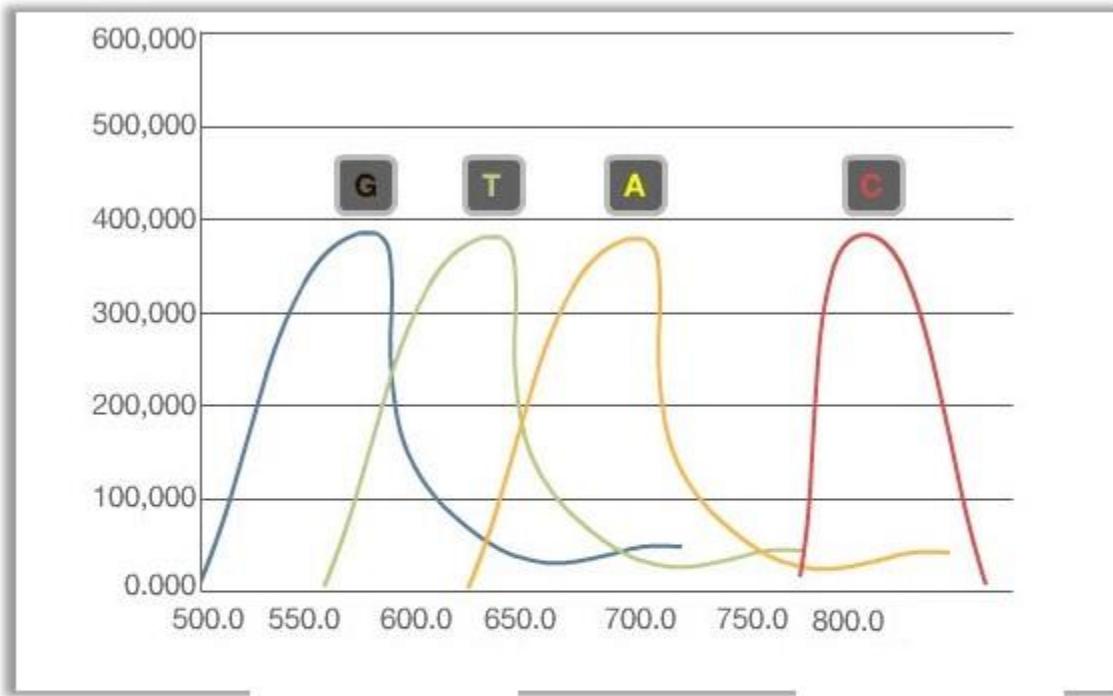


Clusters



100 Microns

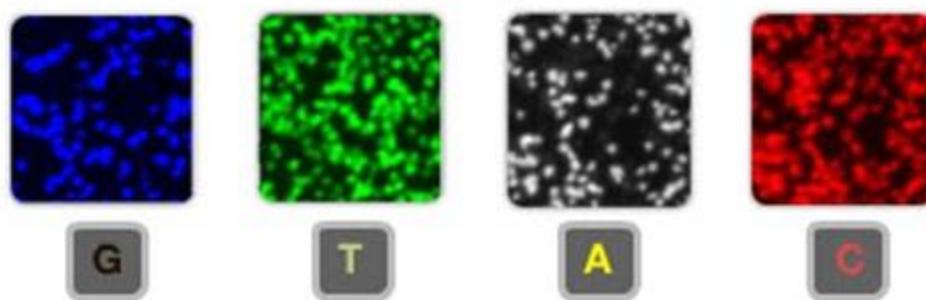
Four Channel SBS Chemistry: GA, HiSeq, MiSeq



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

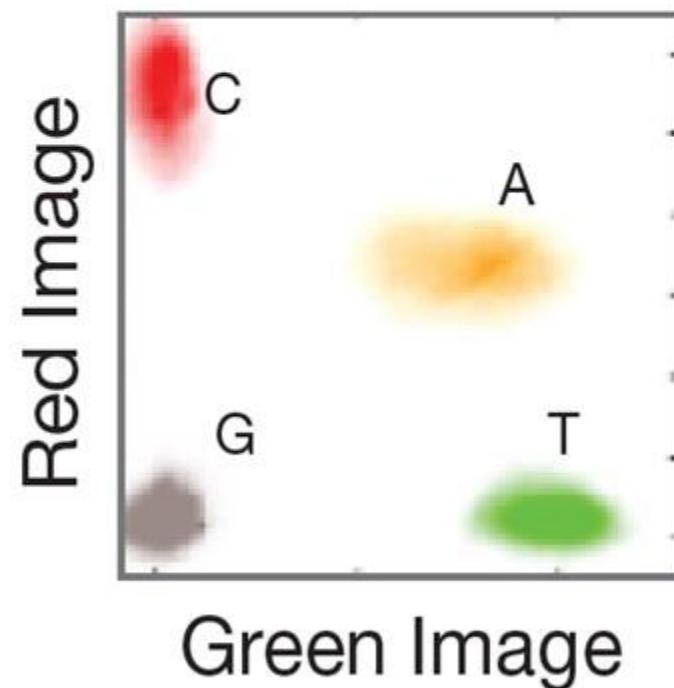
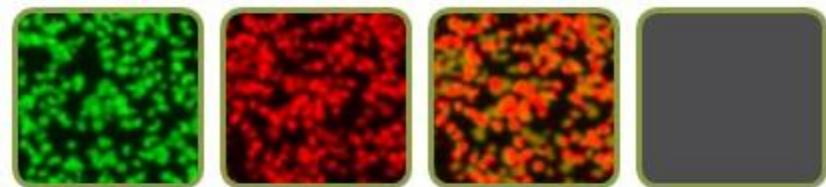
Collects four images:

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

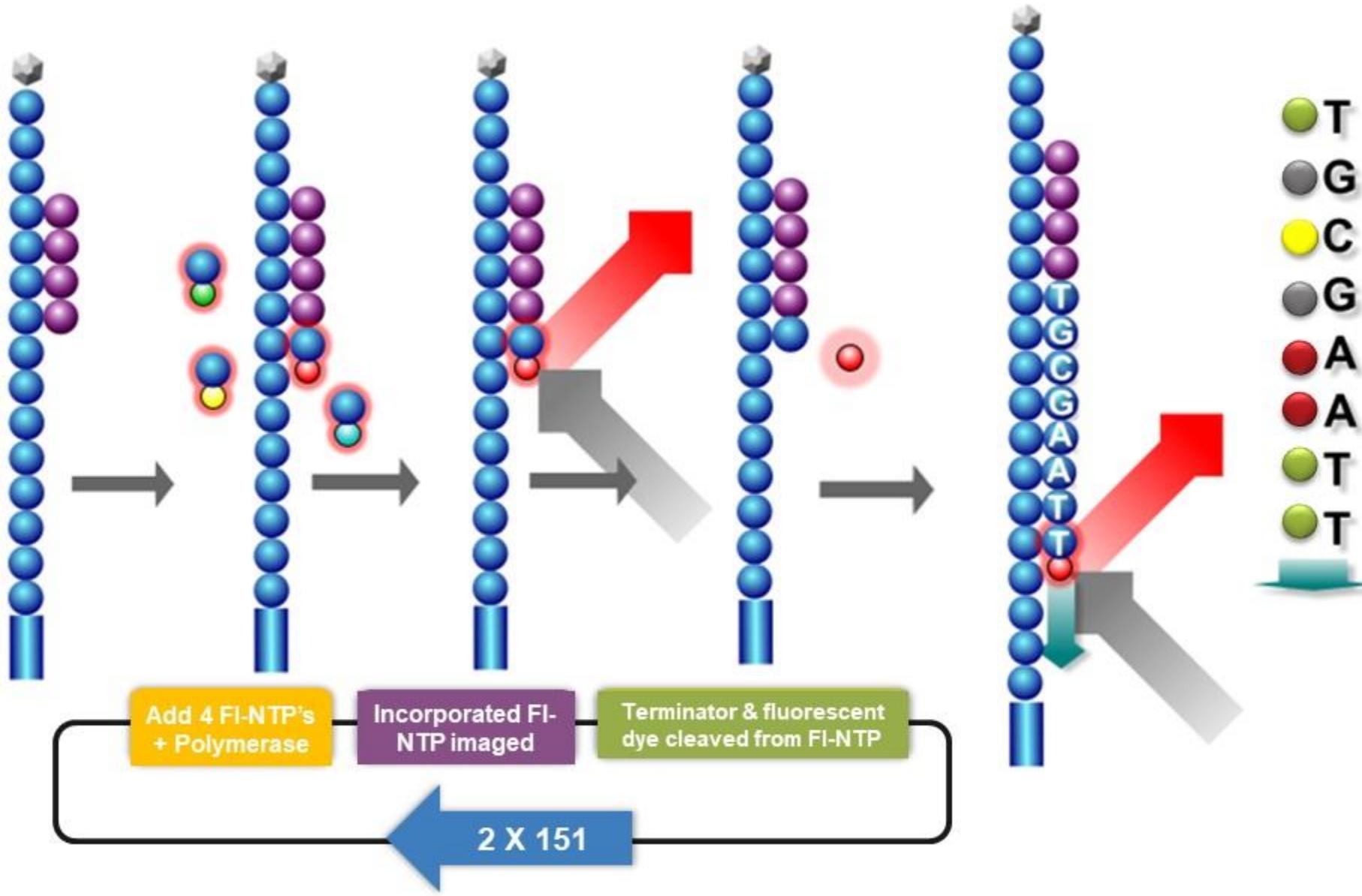


Two Channel SBS – NextSeq, MiniSeq and NovaSeq

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



Sequencing By Synthesis



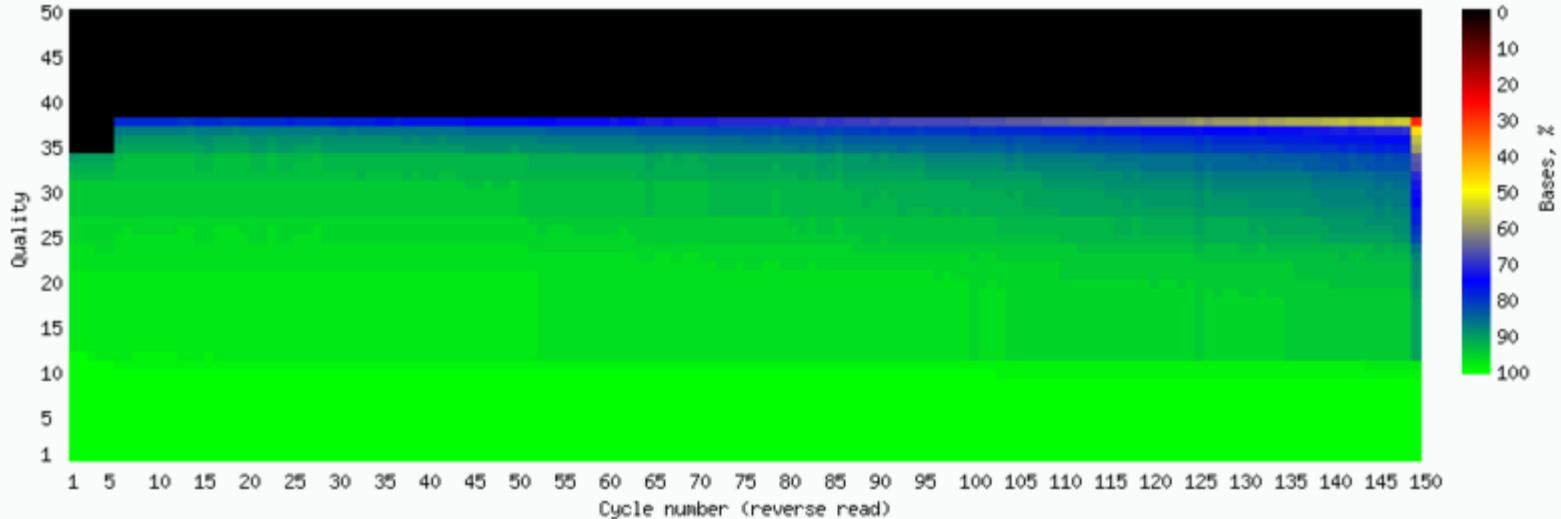
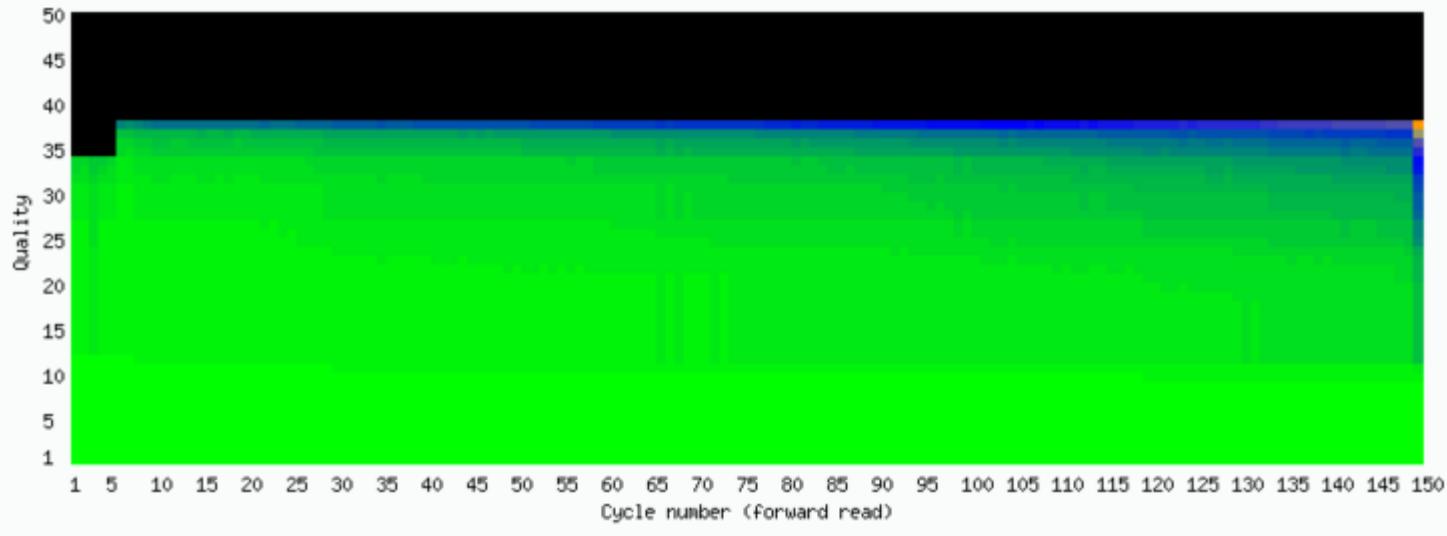
For *In Vitro* Diagnostic Use



Error

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

Error limits read length





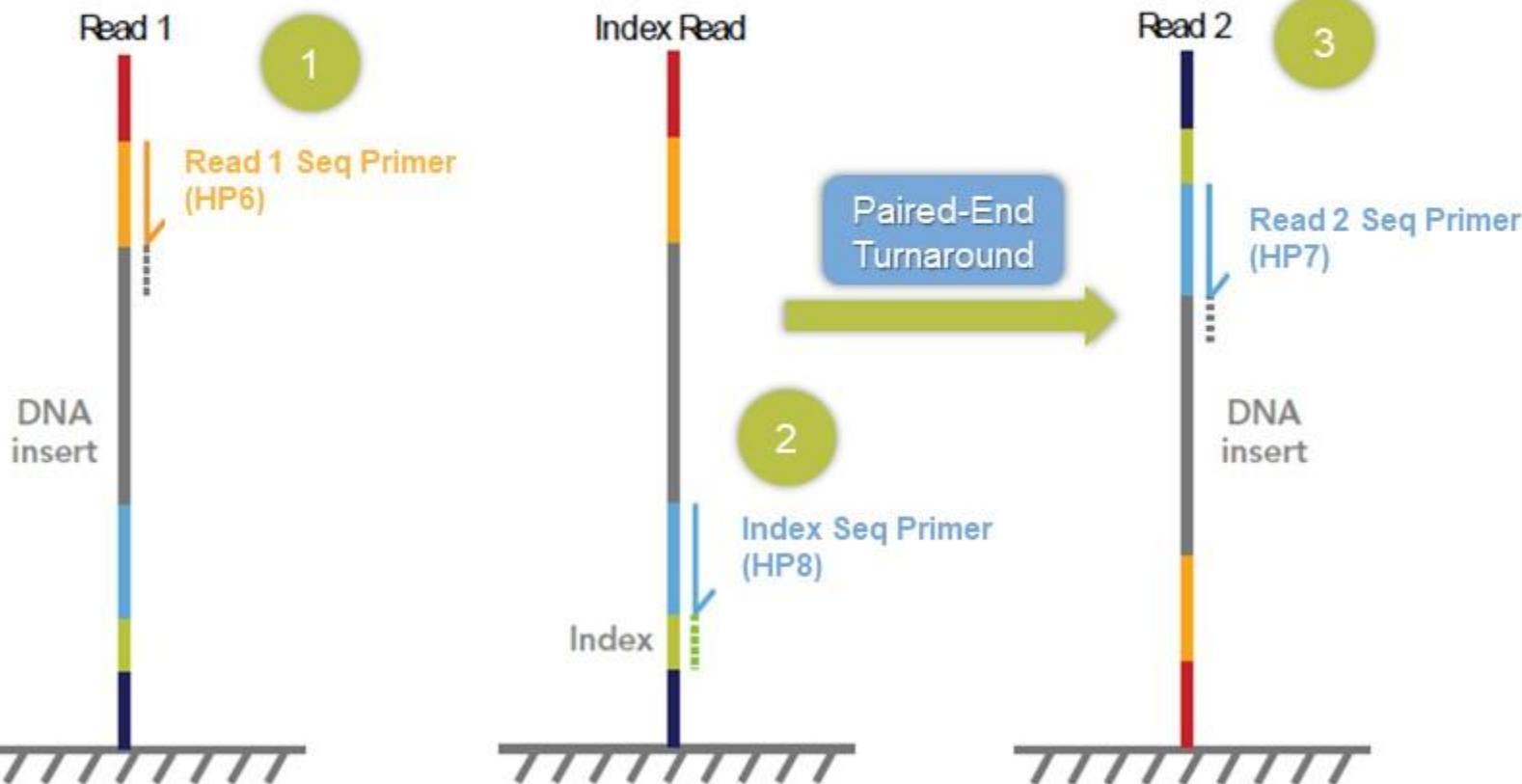
Illumina sequencing with indexing

With Illumina sequencing you sequence the insert from each end. And you can also perform sequencing reactions that read the sequence of the index in the introduced adapter. So how is this done

Single Index Reads

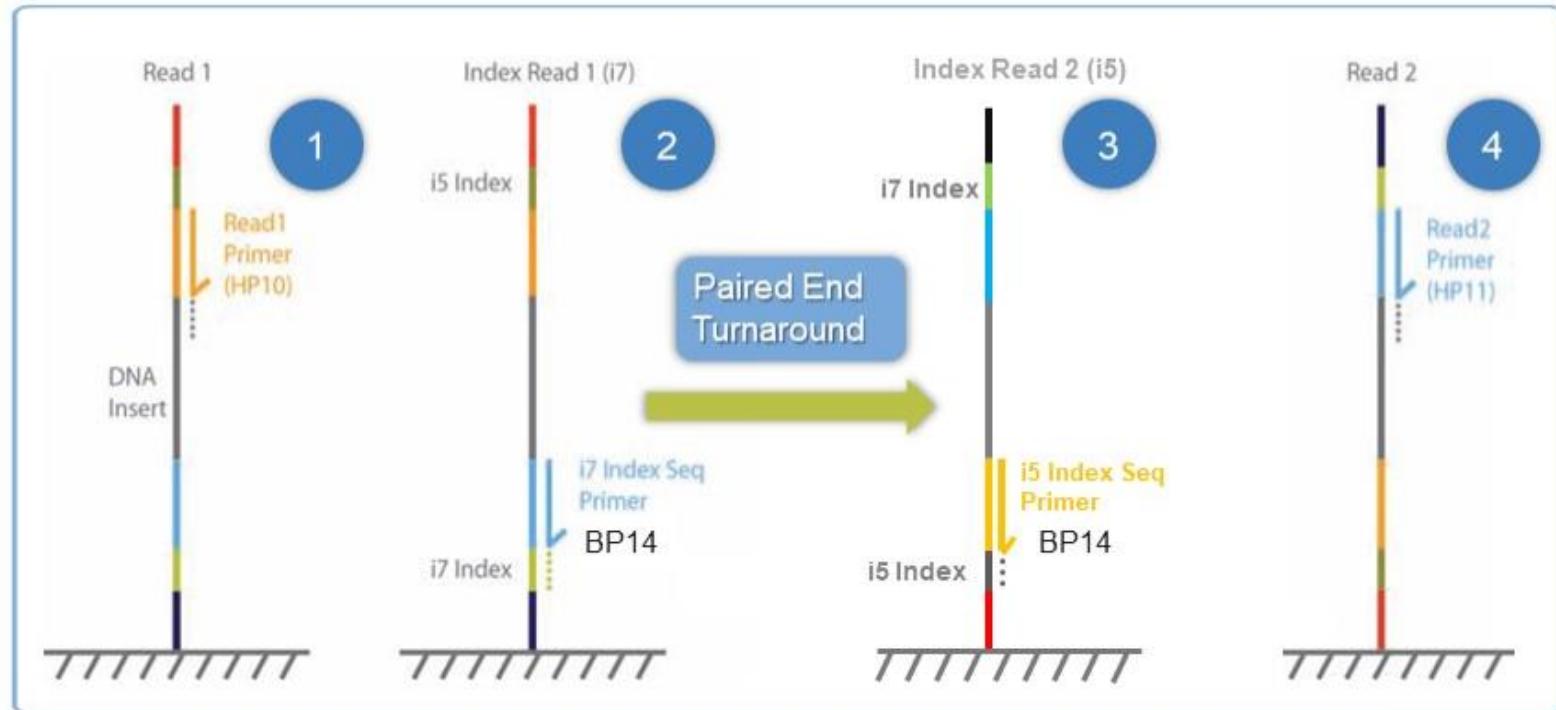
All Platforms

Single indexed sequencing utilizes 3 sequencing reads



Dual indexing method B

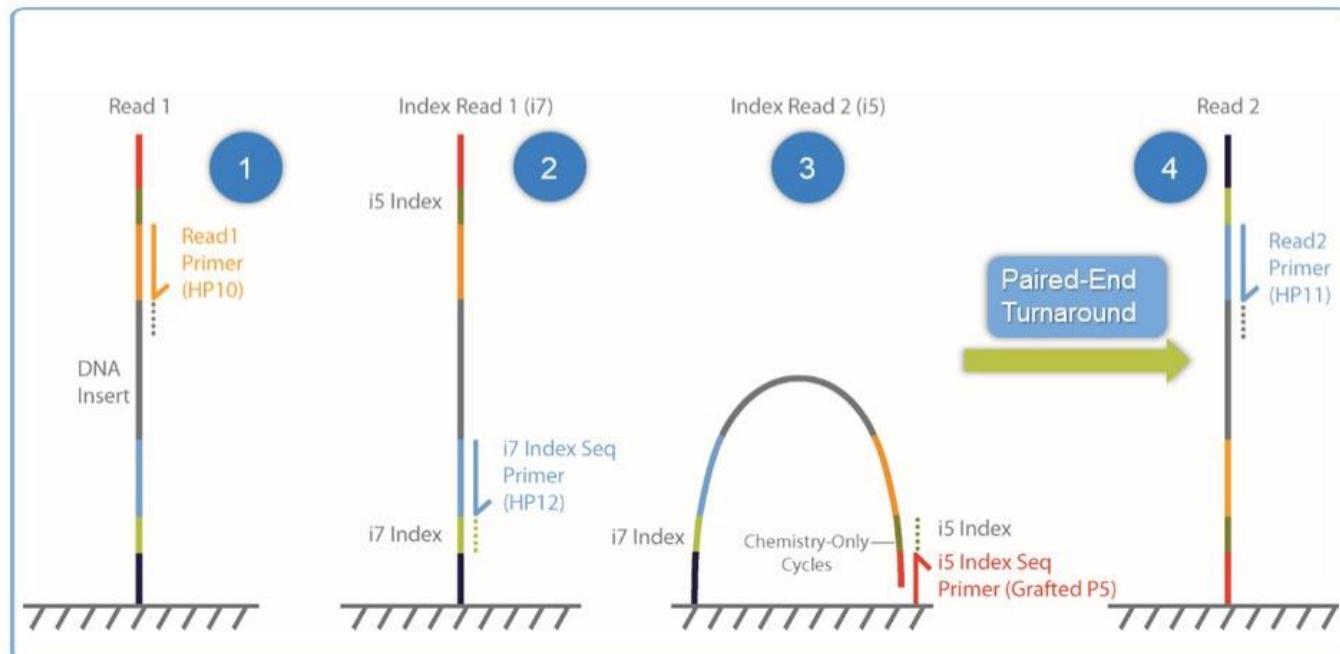
The Index 2 sequencing primer is part of the dual-indexing primer mix for iSeq 100, MiniSeq, and NextSeq. For HiSeq X, HiSeq 4000, and HiSeq 3000, the Index 2 sequencing primer is part of HP14, an indexing primer mix that contains primers for both index reads. *Also NovaSeq on v1.5 onwards*



Dual indexing method A

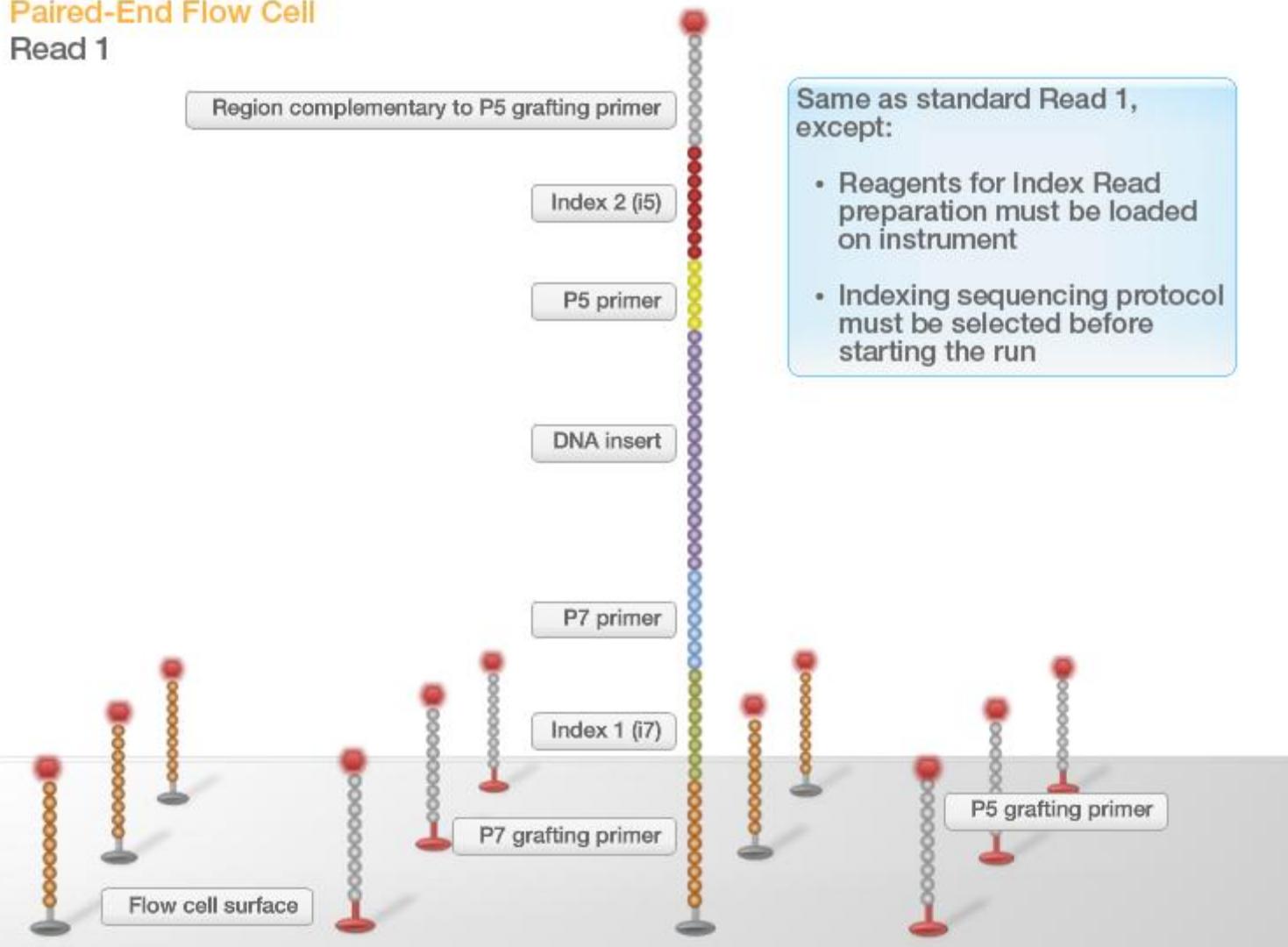
Workflow A

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.



Paired-End Flow Cell

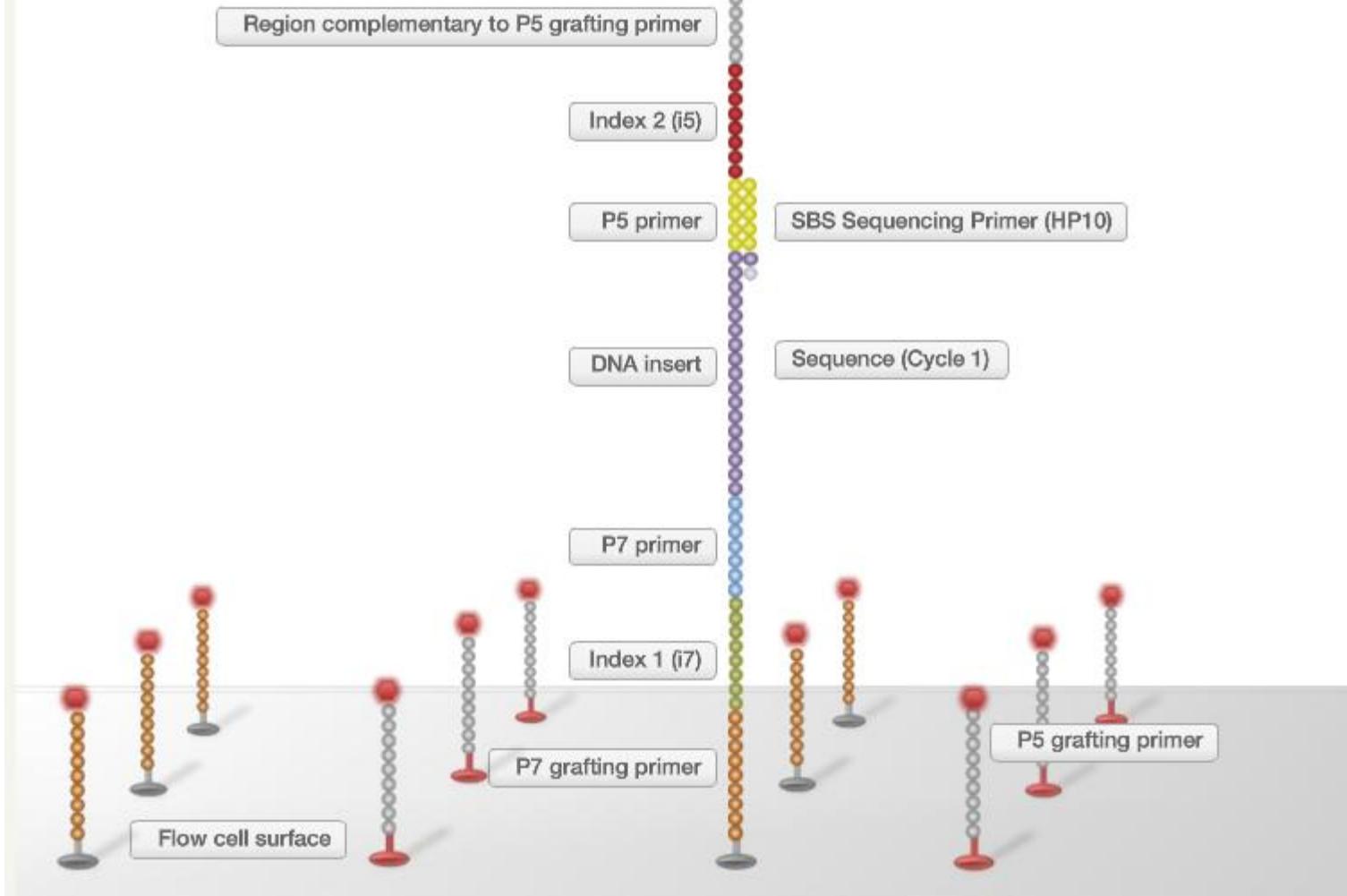
Read 1



Each cluster contains many strands extended from the grafted P7 flowcell primer.

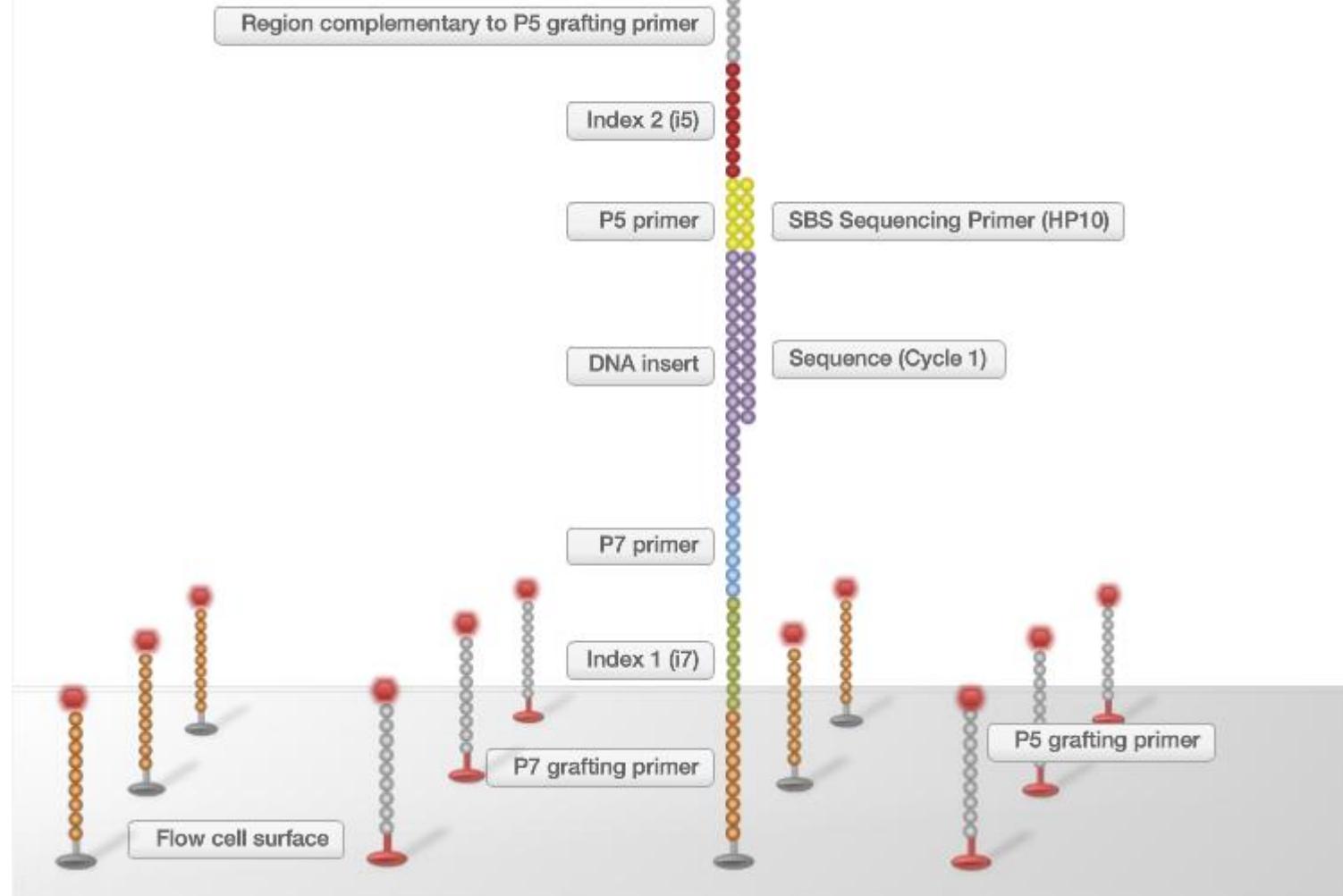
Paired-End Flow Cell

Read 1



The first sequencing primer is annealed and this sits immediately upstream of the DNA insert. The first base sequenced is the first base of the insert

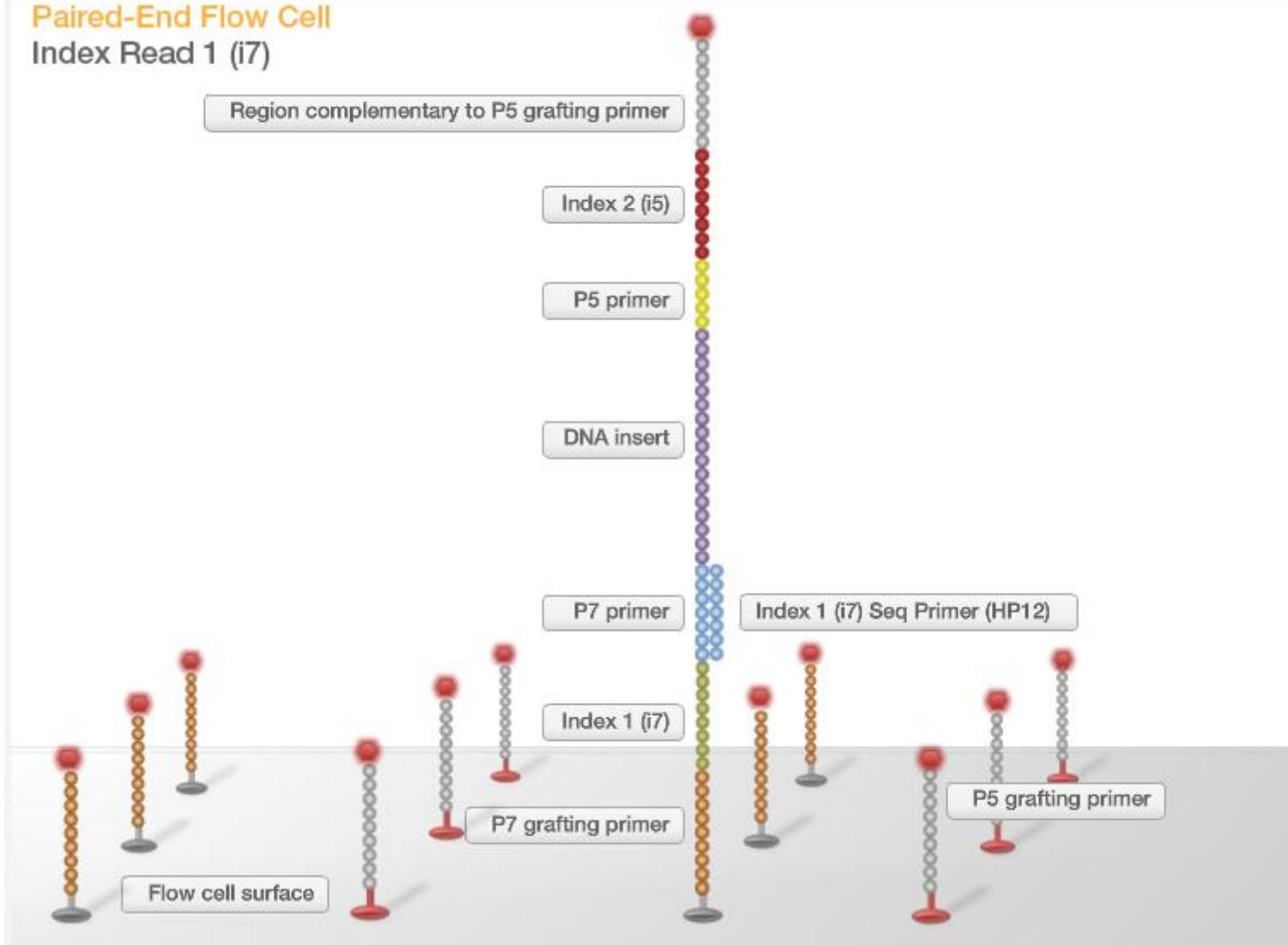
Paired-End Flow Cell Index Read 1 (i7)



The instrument does the first read

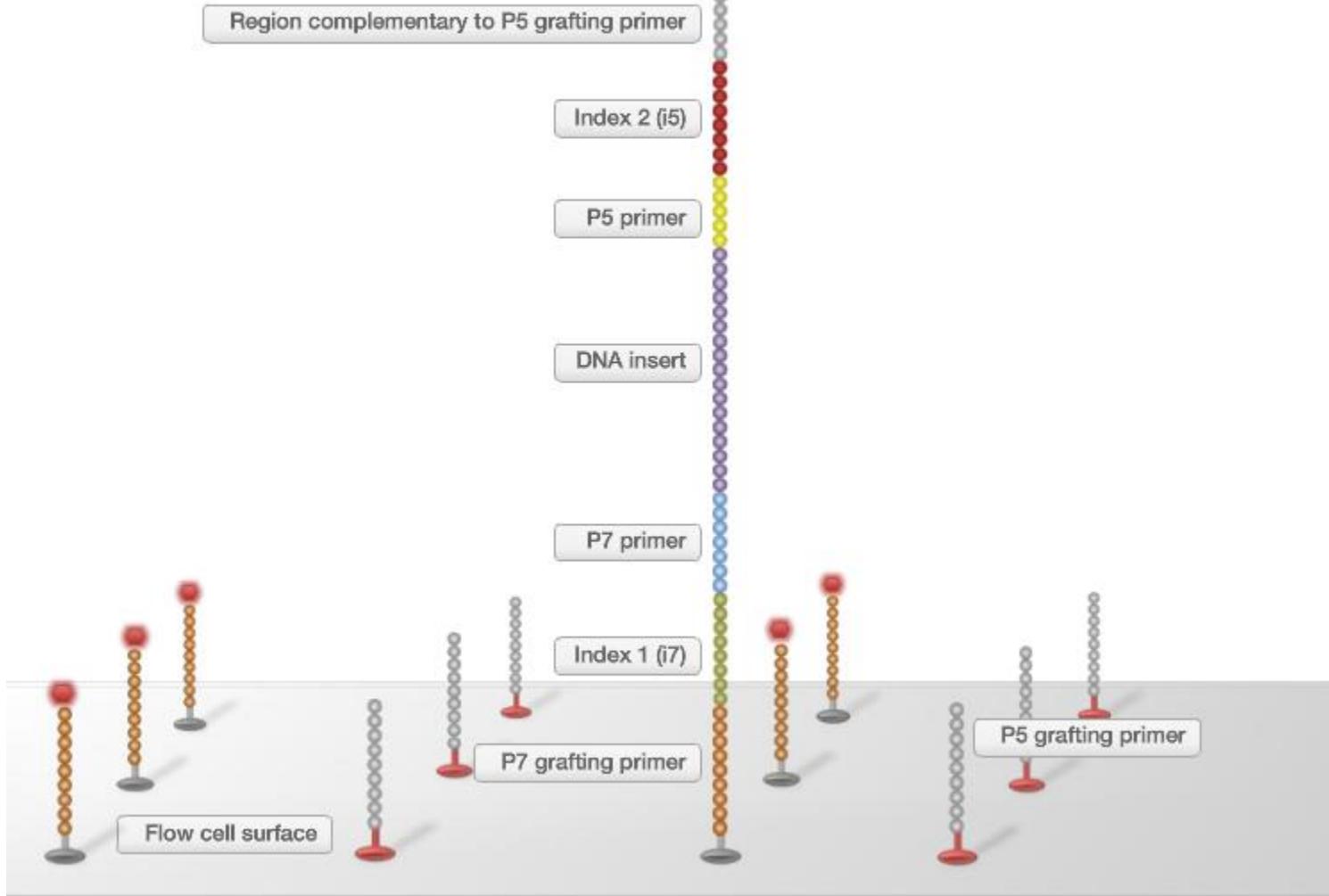
Paired-End Flow Cell

Index Read 1 (i7)



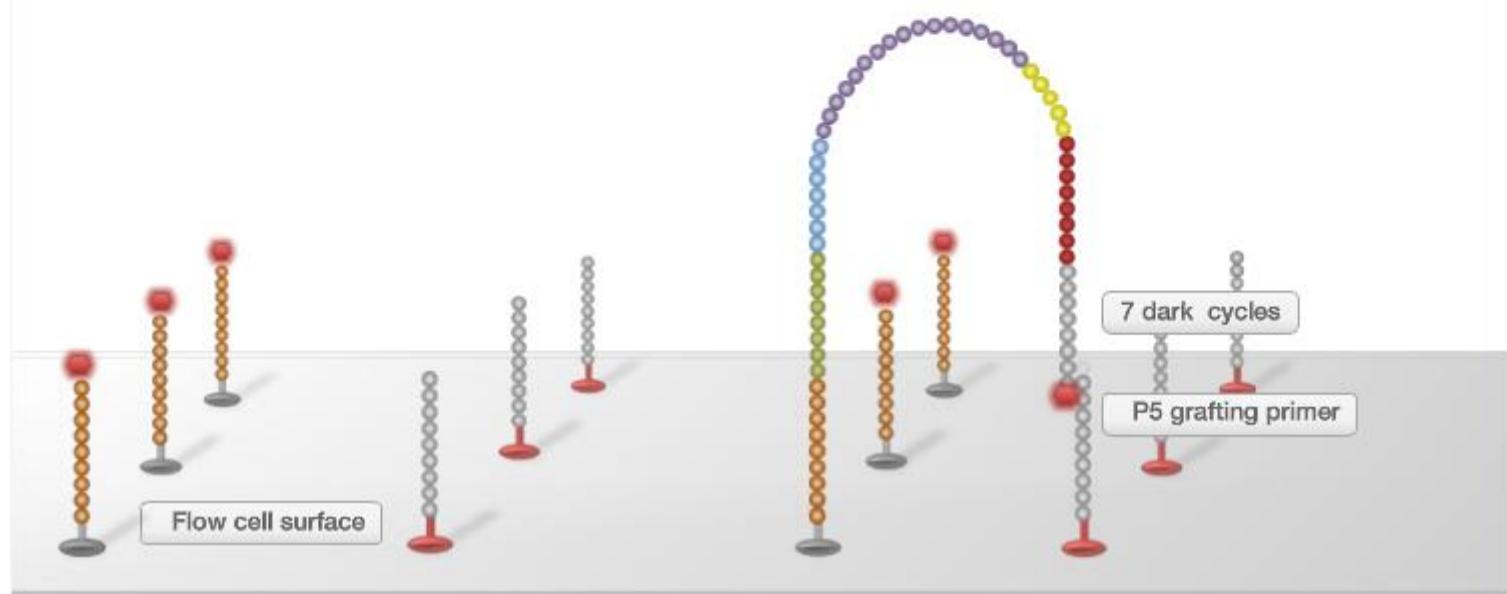
Then that strand is dissociated and washed away and another primer annealed just upstream of the i7 adapter index so it can be sequenced

Paired-End Flow Cell Index Read 2 (i5)

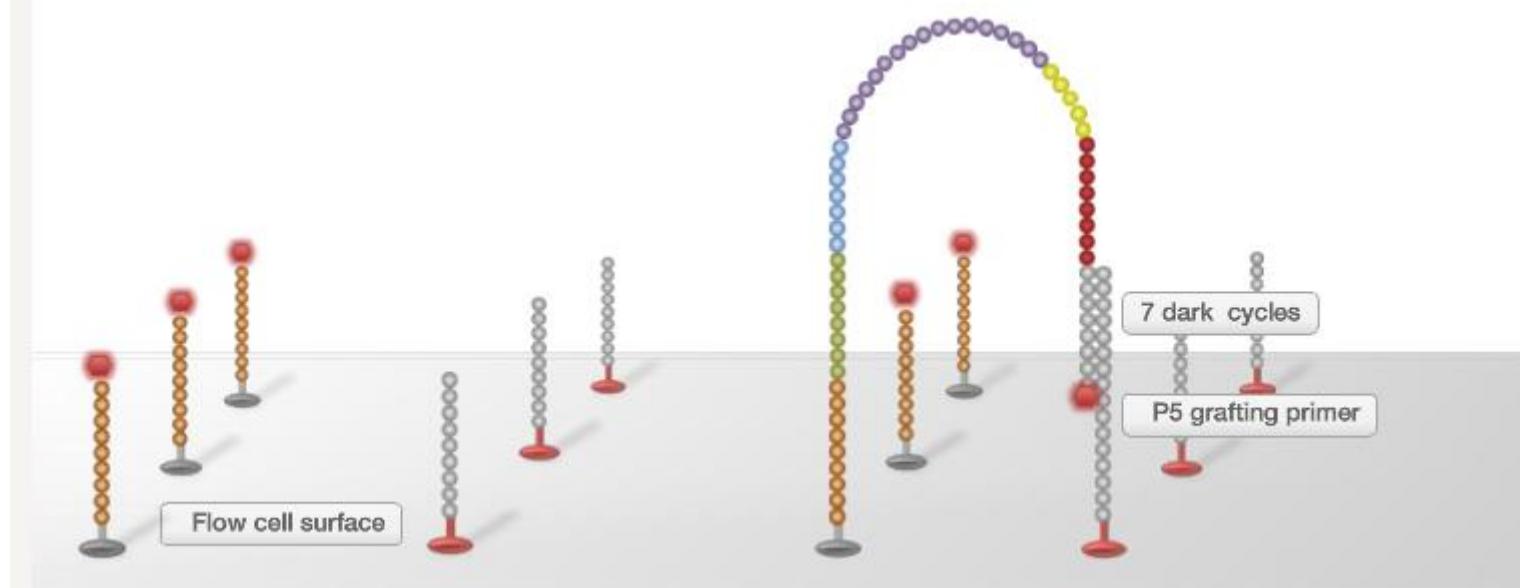


That strand is dissociated and washed away and the 3' ends of the P5 oligos on the flowcell surface deblocked.

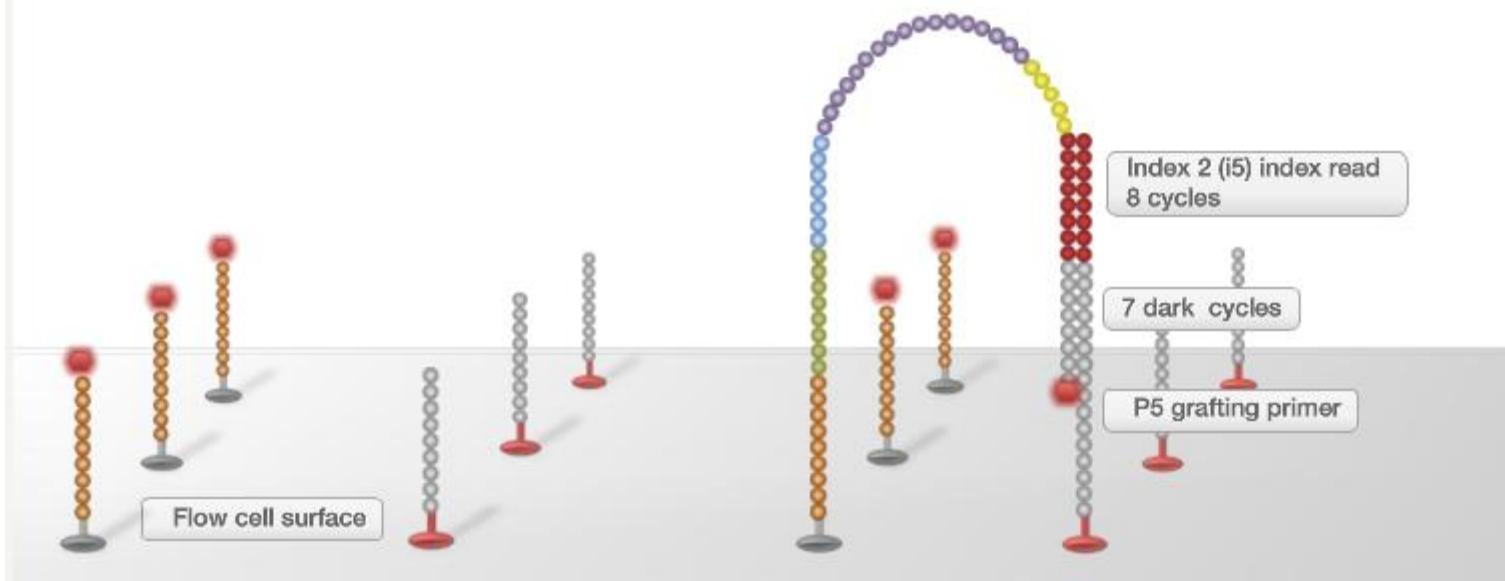
Paired-End Flow Cell
Index Read 2 (i5)



Paired-End Flow Cell
Index Read 2 (i5)

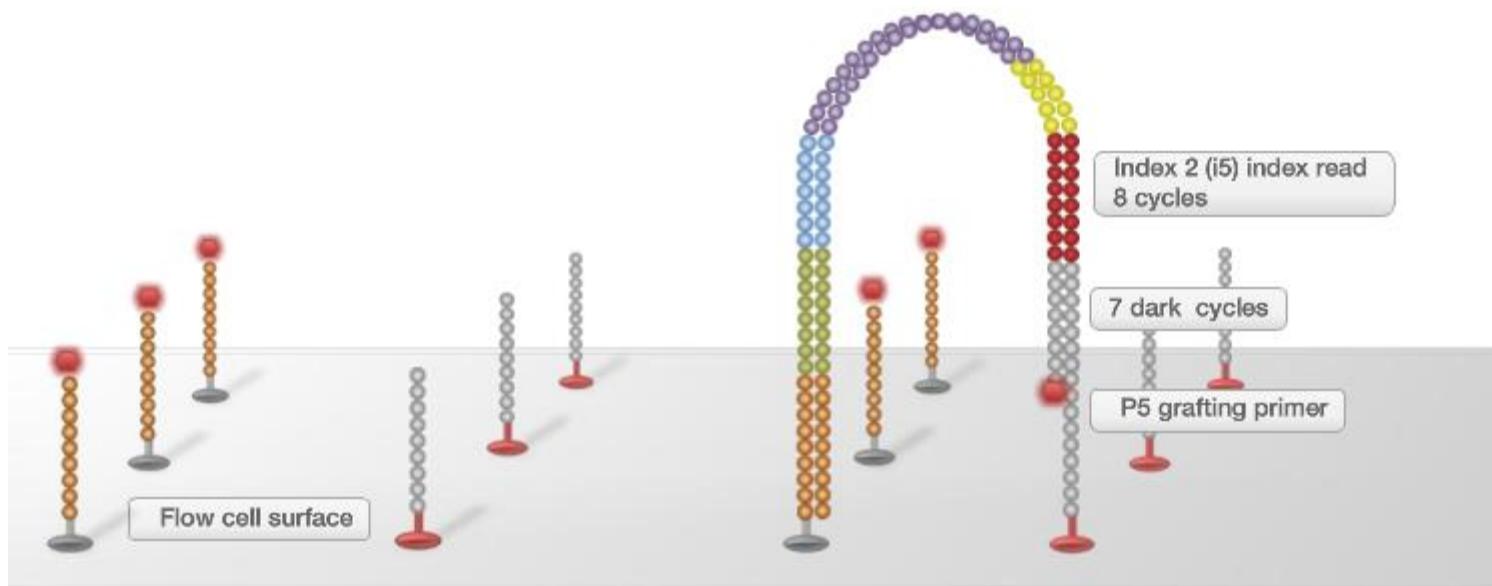


Paired-End Flow Cell Index Read 2 (i5)



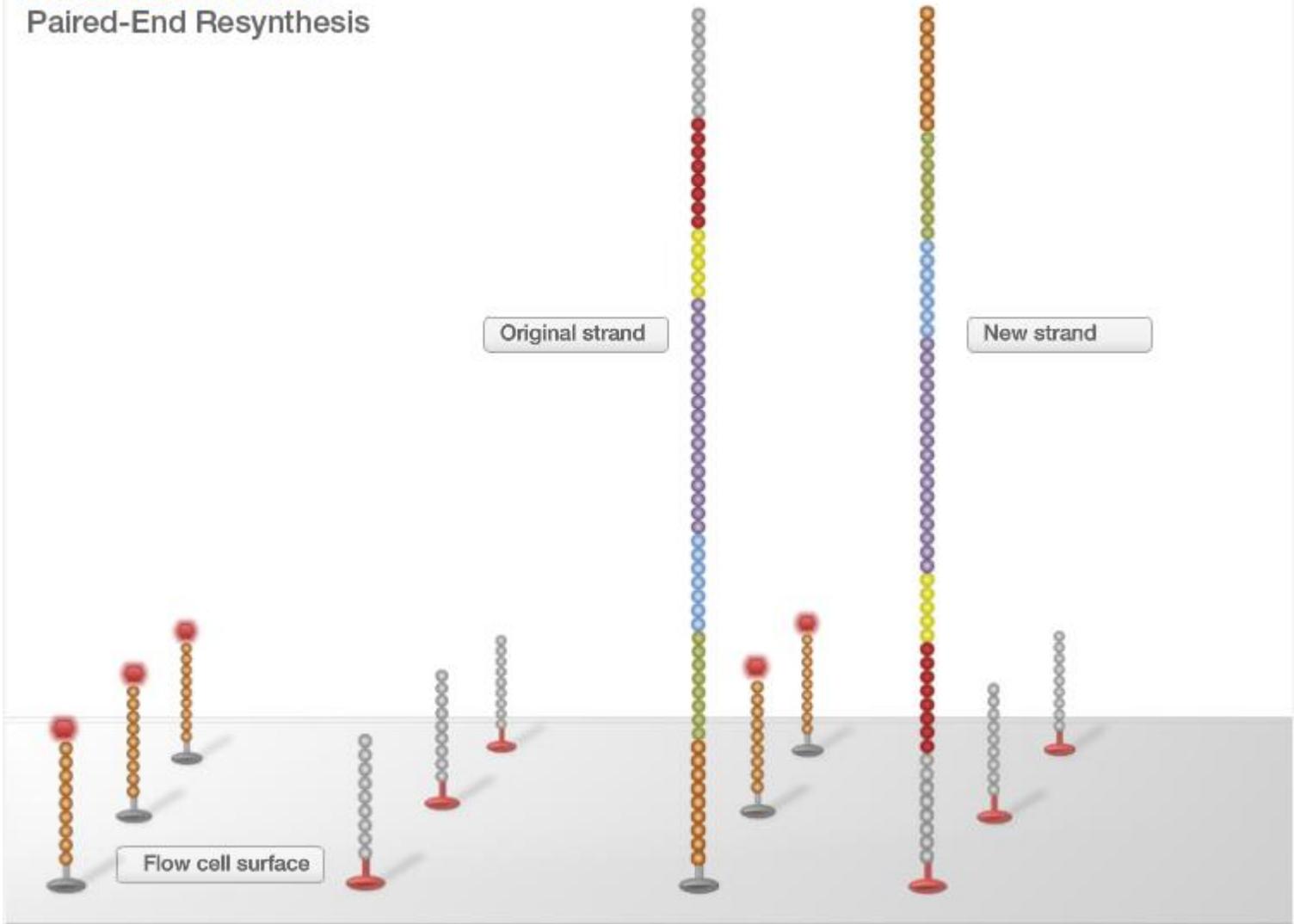
Then the sequence of the i5 index read

Paired-End Flow Cell Paired-End Resynthesis

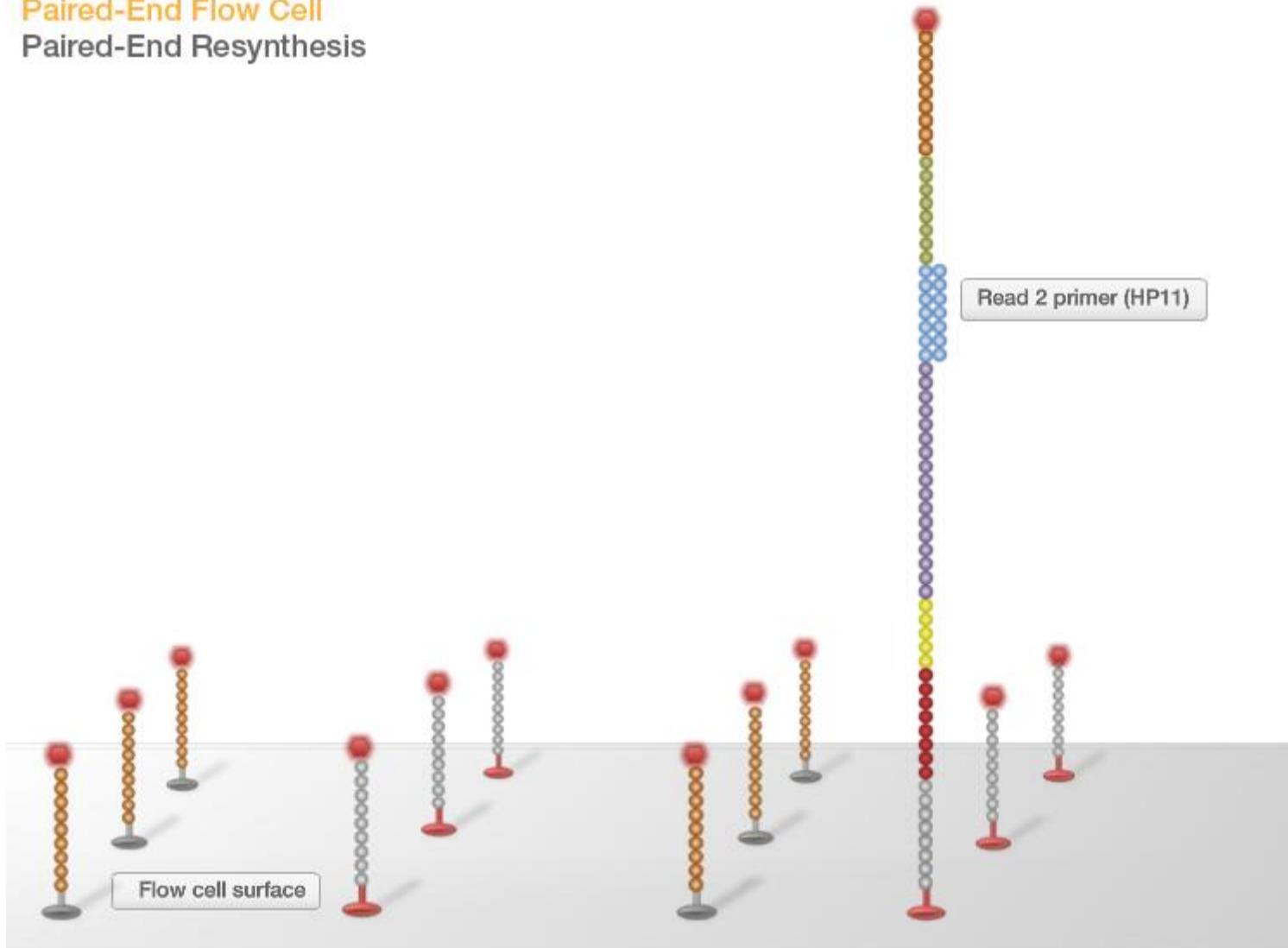


Synthesis is allowed to carry on to form the complement strand

Paired-End Flow Cell Paired-End Resynthesis



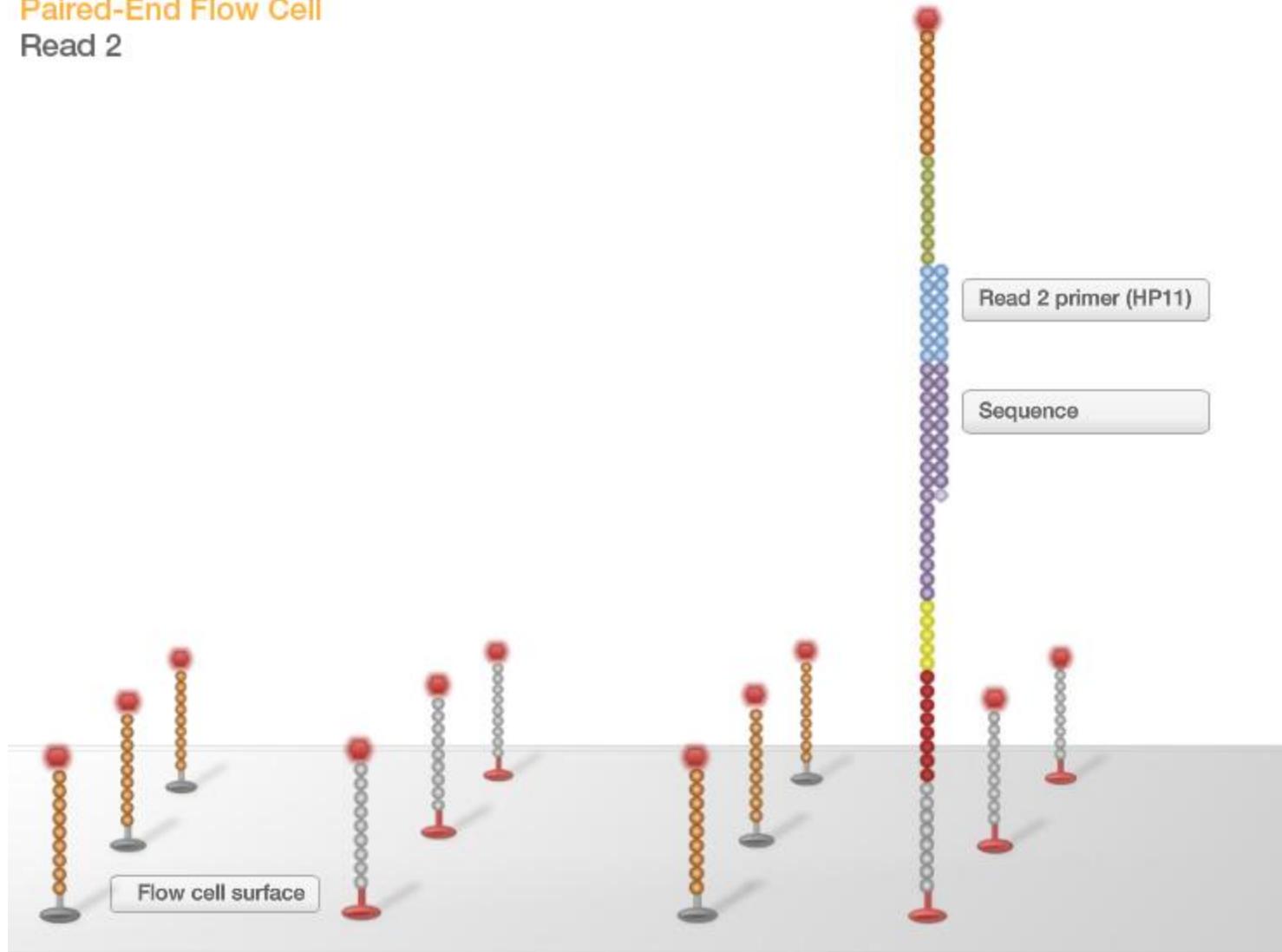
Paired-End Flow Cell Paired-End Resynthesis



Reagents are added that cleave the original strand. A read 2 primer is anneal to the P7 end this time

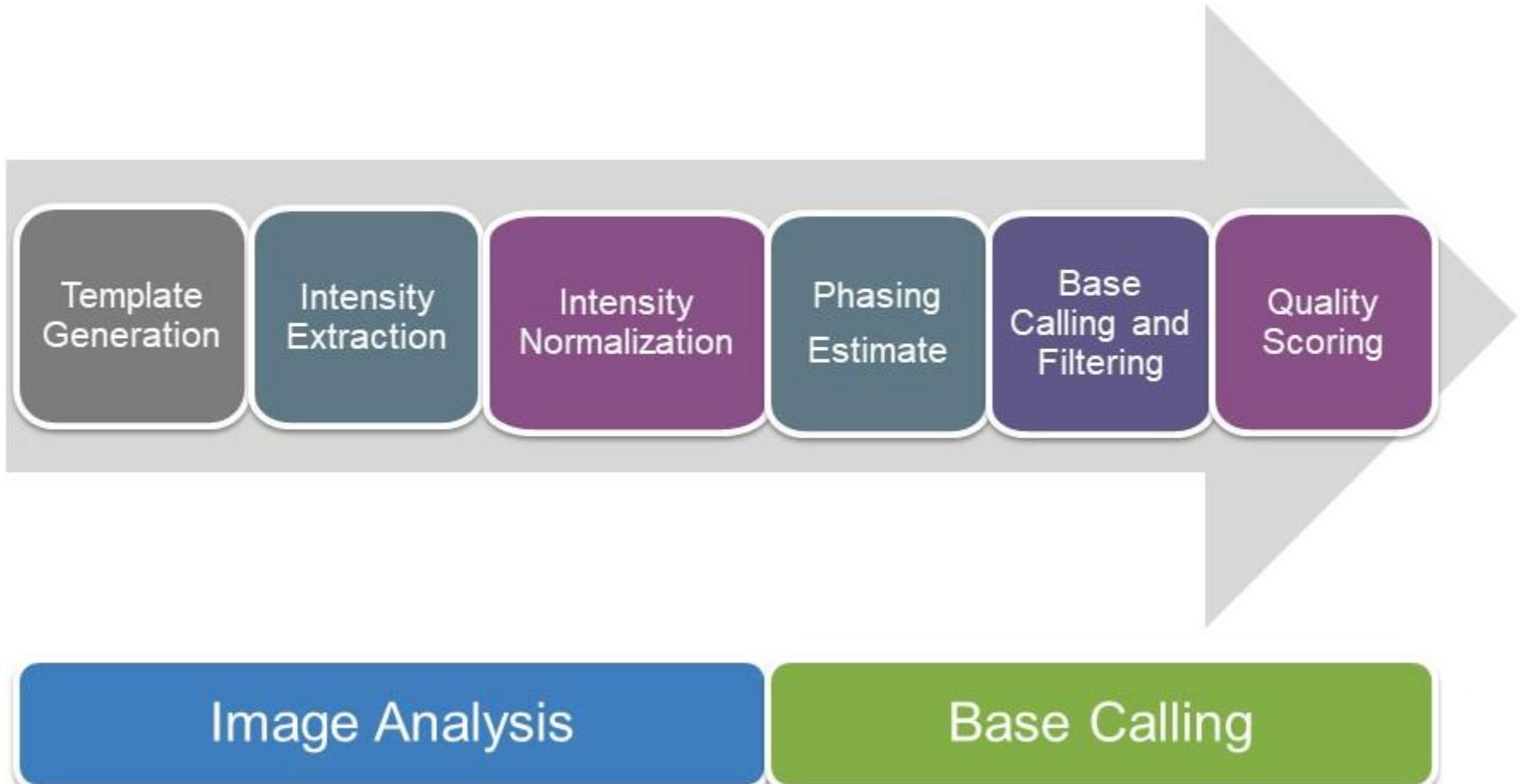
Paired-End Flow Cell

Read 2



The sequence from the other end of the insert is obtained

Primary Data Analysis Workflow

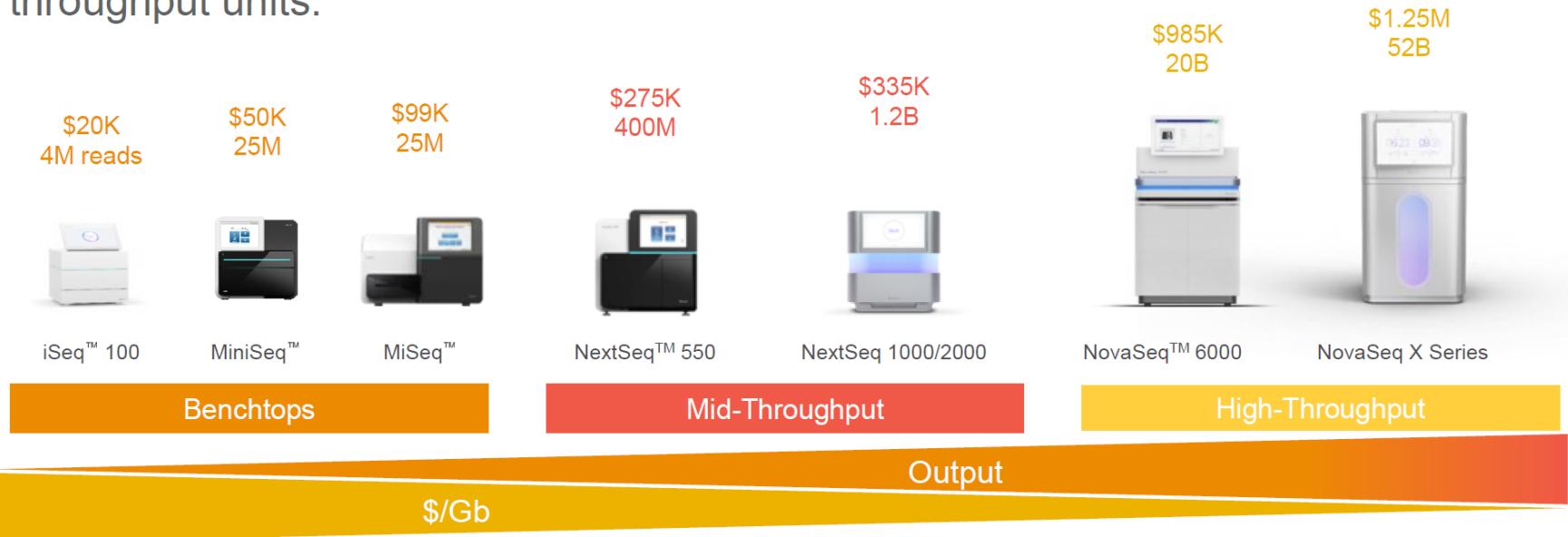




Instruments

Illumina's Sequencing Portfolio

Sequencers vary a lot in size! From small benchtop platforms to large high throughput units.

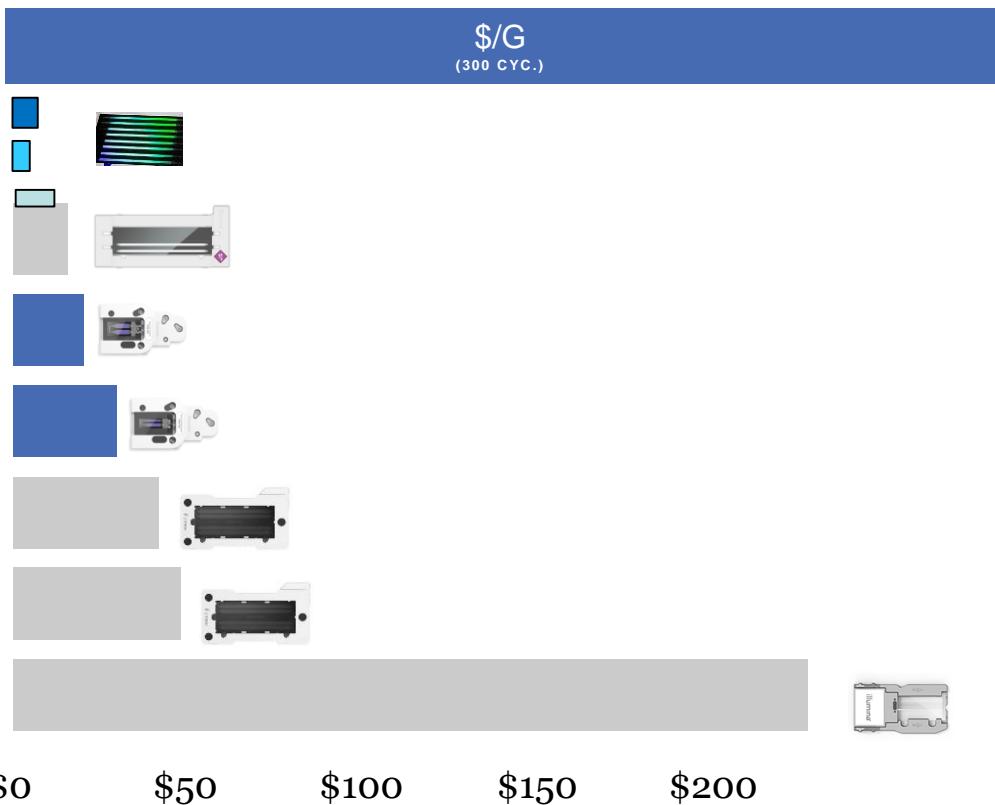


► The sequencer you use may depend a lot on your experiment!

► 300,000 and growing peer-reviewed publications

Economics of Illumina's Sequencing Portfolio

Platform	Flow Cell	\$/G (300 CYC.)
Novaseq X/X+	10B 25B	\$3 \$2
NovaSeq 6000	S4 SP	\$5 \$15.40
NextSeq 2000	P3	\$20.00
NextSeq 1000/2000	P2	\$29.50
NextSeq 550	HO	\$41.38
NextSeq 550	MO	\$47.50
MiSeq	v2	\$236.00



Novaseq X – 2023 launch

\$1.25M

\$3/Gb on 10 B flowcell
\$2/Gb on 25 B flowcell

Two instrument configurations:

The NovaSeq X Plus and the NovaSeq X systems. Two unmatched sequencing systems. Two game changers. You'll find they're designed for an intuitive and optimized high-throughput sequencing workflow.

NovaSeq X Plus Sequencing System

Dual-flow cell system	Up to 16 Tb per dual flow cell run or > 128 human genomes at 30x coverage	Independent flow cell operation
-----------------------	---	---------------------------------

NovaSeq X Sequencing System

Single-flow cell system	Up to 8 Tb per run or > 64 human genomes at 30x coverage	Upgradeable to dual flow cell instrument
-------------------------	--	--



NovaSeq X Plus



NovaSeq X

10B 2x150 in 24 hours
25B 2x150 in 48 hours

NovaseqX+



Table 1: NovaSeq X Series performance parameters^a

Flow cell type ^b	1.5B	10B	25B
Output per single flow cell run ^a			
2 × 50 bp	~165 Gb	1 Tb	–
2 × 100 bp	330 Gb	2 Tb	–
2 × 150 bp	500 Gb	3 Tb	8 Tb
Output per dual flow cell run ^{a,c}			
2 × 50 bp	~330 Gb	2 Tb	–
2 × 100 bp	660 Gb	4 Tb	–
2 × 150 bp	1 Tb	6 Tb	16 Tb
Reads passing filter per flow cell ^a			
Single reads	1.6 billion	10 billion	26 billion
Paired-end reads	3.2 billion	20 billion	52 billion
Instrument run time ^{a,d}			
2 × 50 bp	~13 hr	~18 hr	–
2 × 100 bp	~18 hr	~22 hr	–
2 × 150 bp	~21 hr	~24 hr	~48 hr
Quality scores ^{a,e}			
2 × 50 bp	≥ 90% of bases higher than Q30		
2 × 100 bp	≥ 85% of bases higher than Q30		
2 × 150 bp	≥ 85% of bases higher than Q30		

- 2 x 8-lane flow cells
- High resolution optics allow for tighter spacing of nanowells on flow cell surface enabling higher read counts vs. NS6000.
- New X-LEAP-SBS chemistry with improved reagent stability
- Faster incorporation times
- Improved sustainability of reagent supply
- Novaseq X and X+ are integrated with DRAGEN Bio-IT Platform



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	●	●	●

iSeq™ 100

Your new lab partner



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

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

Affordable NGS system

iSeq™ 100 System

\$19,900 USD List

iSeq™ 100 i1 Reagent
Four Million Read Sequencing Kit

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



iSeq™ 100 Specifications



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

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

A Closer Look At 1-Dye Chemistry

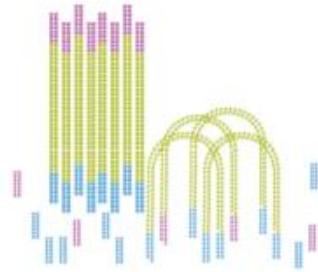
Cluster Generation and SBS

What's the Same?

Library Preparation



Cluster Growth



Sequencing

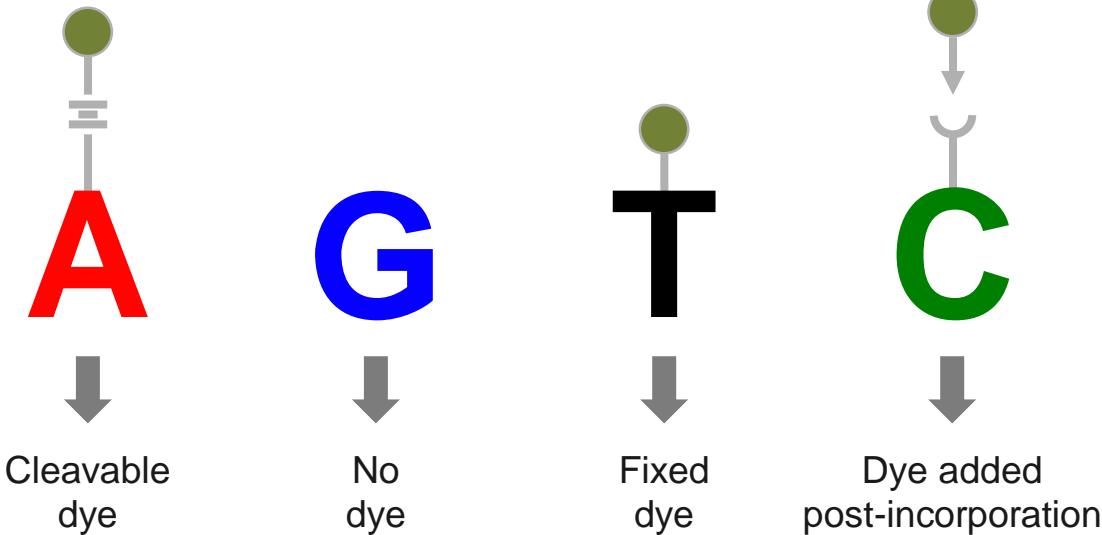


The iSeq™ 100 System **still utilizes the Illumina SBS chemistry**,
where each base is added one at a time

A Closer Look At 1-Dye Chemistry

Cluster Generation and SBS

What's the Different?



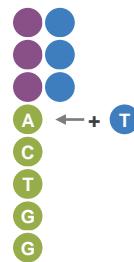
Nucleotides are labeled with a single dye, with the exception of the **G nucleotide**

A Closer Look At 1-Dye Chemistry

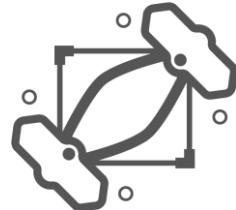
SBS and Imaging

Sequencing by Synthesis

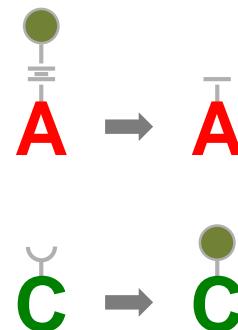
Sequencing Cycle



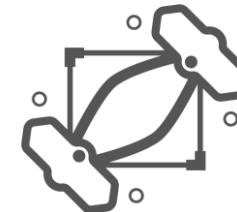
Incorporation



Imaging



Chemistry

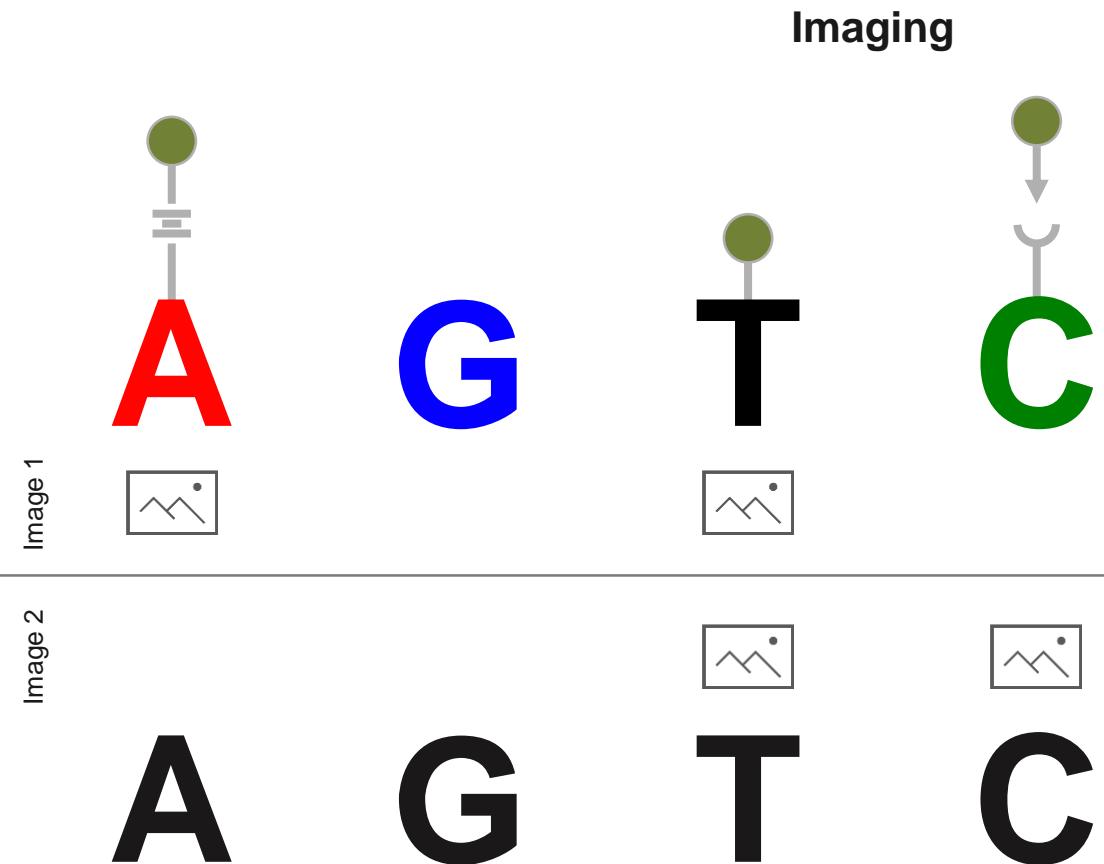


Imaging

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

A Closer Look At 1-Dye Chemistry

SBS and Imaging



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

MiniSeq



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



Illumina MiSeq

- » Single-lane version of HiSeq
- » Read length: 2 x 300 (1 -2 days)
- » Error rate 0.1%
- » Yield: > 10 Gb / run
- » \$99,000 capital cost
- » \$100-\$240 / Gb



Illumina MiSeq

- » Single-lane version of HiSeq
- » Read length: 2 x 300 (1 -2 days)
- » Error rate 0.1%
- » Yield: > 10 Gb / run
- » \$99,000 capital cost
- » \$236 / Gb



MiSeq i100



MiSeq i100 Series specifications^a

Output range^b 1.5–30 Gb

Paired-end reads per run 10–200M

Max read length 2 x 300 bp

Run time ~4–15.5 hr

a. Specifications based on Illumina PhiX control library at supported cluster densities.

b. Maximum range based on 100M flow cell specifications. The 100M flow cell will be available starting in 2025 for the MiSeq i100 Plus System only.

- Faster. 7.5hr run for 2 x 150
- 40% cheaper than Miseq
- Does Index read first.
- Better with low complexity?
- 5% phiX for 16S
- RT reagents
- Patterned flowcell
- 2 x 500 coming

	5M	25M	50M	100M
100 cycles	N/A	\$550	\$750	\$1,000
300 cycles	\$420	\$900	\$1,100	\$1,450
600 cycles	\$650	\$1,100	\$1,450	N/A



Illumina NextSeq 500

- » Uses only 2 dyes / 2 images not 4
- » Read length: 2 x 150 (1.25 days)
- » Yield: 120 Gb / run
- » \$250,000 capital cost
- » \$30-\$45 / Gb





NextSeq 1000/2000

Table 1: Performance parameters for the NextSeq 1000 and 2000 Sequencing Systems

Read length	NextSeq 1000/2000 P2 Reagents ^a	NextSeq 2000 P3 Reagents ^b
Output per flow cell		
2 × 50 bp	40 Gb	100 Gb
2 × 100 bp	80 Gb	200 Gb
2 × 150 bp	120 Gb	300 Gb
Quality scores		
2 × 50 bp	≥ 85% of bases higher than Q30	
2 × 100 bp	≥ 80% of bases higher than Q30	
2 × 150 bp	≥ 75% of bases higher than Q30	
Run time		
2 × 50 bp	~13 hours	~19 hours
2 × 100 bp	~21 hours	~33 hours
2 × 150 bp	~29 hours	~48 hours

a. Available mid 2020

b. Available late 2020



NextSeq 1000: \$210k

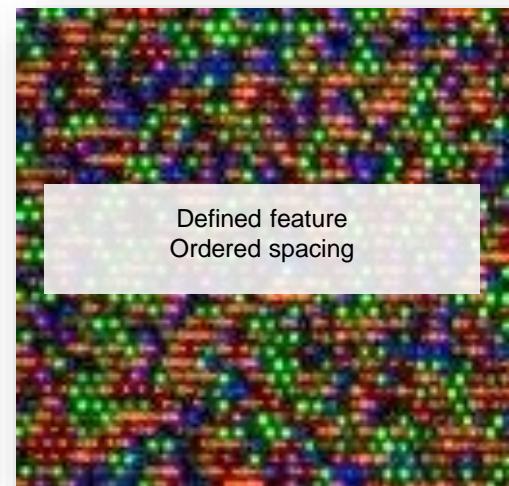
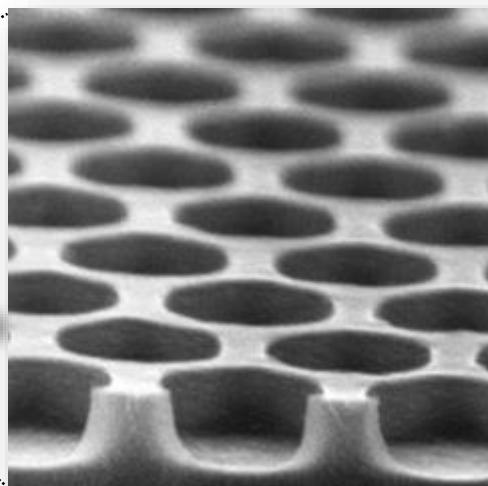
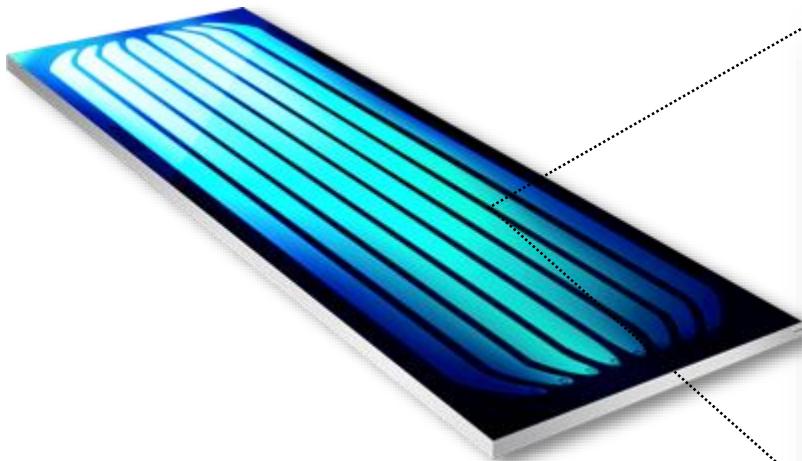
NextSeq 2000: £335k

£20-\$30/Gb

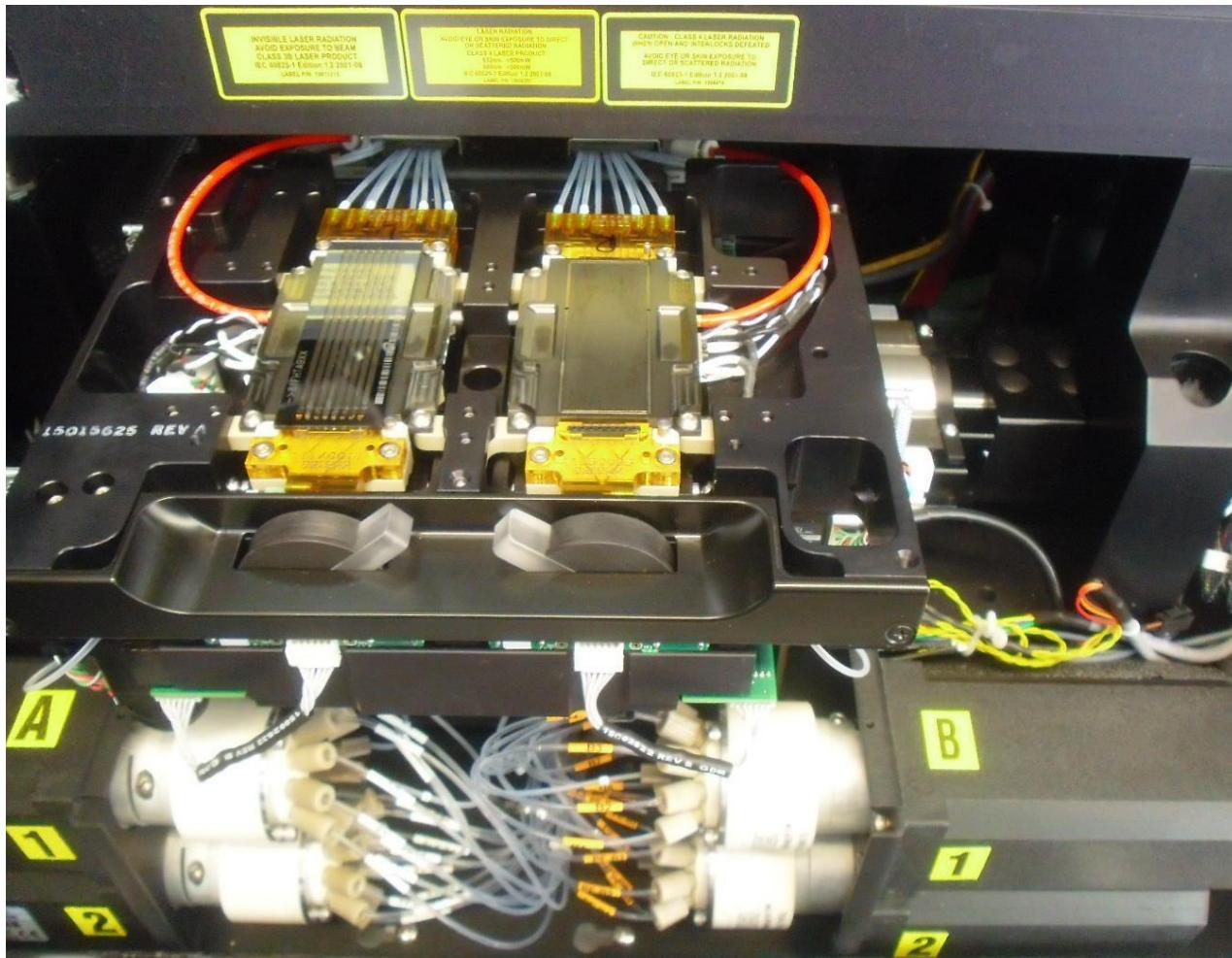
Application	NextSeq 1000/2000 P2 Reagents		NextSeq 2000 P3 Reagents	
	No. samples	Time	No. samples	Time
Small whole-genome sequencing (300 cycles) 130 Mb genome; > 30x coverage	30	~29 hours	75	~48 hours
Whole-exome sequencing (200 cycles) 50x mean targeted coverage; 90% targeted coverage at 20x	16	~21 hours	40	~33 hours
Single-cell RNA-Seq (100 cycles) 4K cells, 50K reads/cell	2	~13 hours	5	~19 hours

Hiseq 3000/4000

- » Single or double flowcell instrument
- » Suitable for all applications inc. exomes
- » Patterned flowcell (aka “PFCT”)
- » \$24/Gb (\$20/Gb at launch)

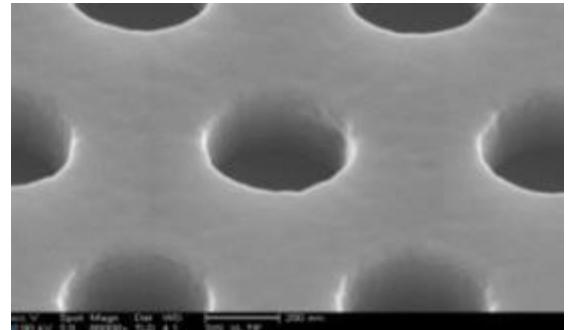
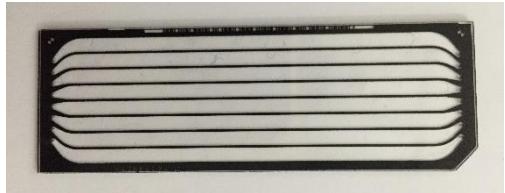


HiSeq Flowcell Tray



Illumina HiSeq X

- » Ultra-high throughput
- » Read length: 2x150 (3 days)
- » Yield: 1.6 Tb / run
- » \$1,000,000 capital cost x 10
- » \$ 8/Gb (7 at launch)
- » The \$1000 Genome



2017: NovaSeq

NovaSeq System Configurations

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

NovaSeq 5000
\$850K USD

NovaSeq 6000
\$985K USD

NovaSeq 5000 Flow Cells

NovaSeq 6000 Flow Cells

NovaSeq 5000

NovaSeq 6000

**wellcome
sanger
institute**

Scalable Throughput

Complete studies faster and more economically



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



Configure
output to
match your
application
and study
size

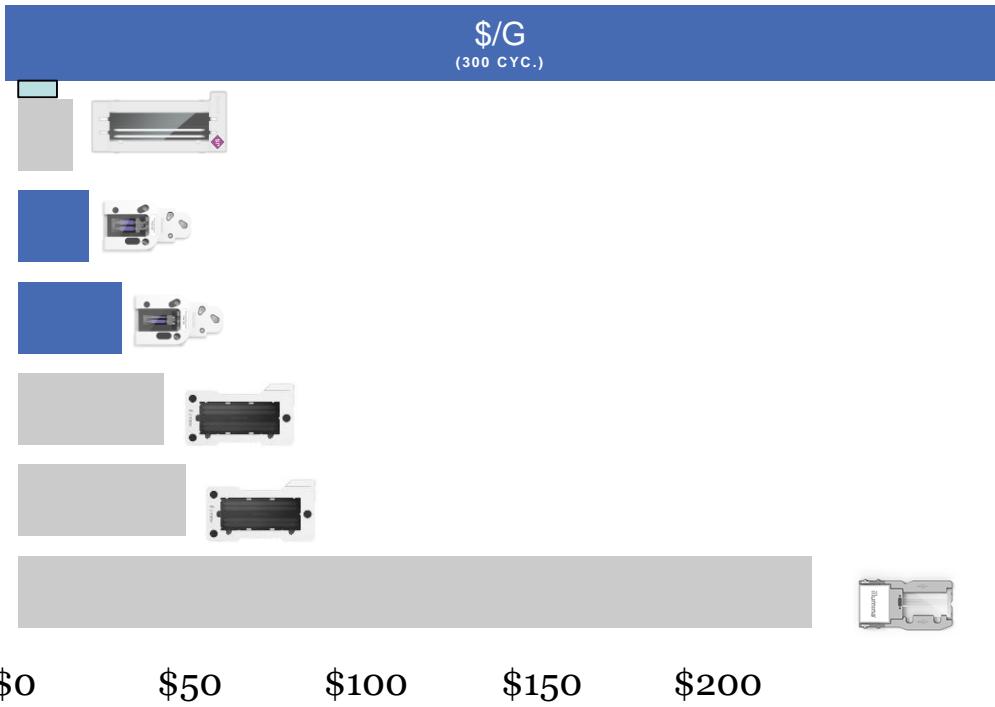
Single flow cell output (1 or 2 can run simultaneously)

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

*S1, S2 and S4 flow cells not currently released

Economics of Illumina's Sequencing Portfolio

Platform	Flow Cell	\$/G (300 CYC.)
NovaSeq 6000	S4	\$6
	SP	\$15.40
NextSeq 2000	P3	\$20.00
NextSeq 1000/2000	P2	\$29.50
NextSeq 550	HO	\$41.38
NextSeq 550	MO	\$47.50
MiSeq	v2	\$236.00





Sequencing kits



NovaseqX

Output per Flow Cell* for Various Read Lengths†

NovaSeq X Series	£3/Gb	£2/Gb	
Flow Cell Type	1.5B	10B	25B
2 × 50 bp	~165 Gb	~1 Tb	N/A
2 × 100 bp	~330 Gb	~2 Tb	N/A
2 × 150 bp	~500 Gb	~3 Tb	~8 Tb

* 10B flow cell available now. 1.5B and 25B flow cells available H2 2023. NovaSeq X Plus system available now. NovaSeq X system available later in 2023. Performance metrics subject to change.

† Specifications based on Illumina PhiX control library or a TruSeq DNA Library created with NA12878 at supported cluster densities.

Novaseq 6000 Kits

Sequencing Output Per Flow Cell

Flow Cell Type	NovaSeq 6000 System			
	SP*	S1	S2	S4
2 x 50 bp	65–80 Gb	134–167 Gb	333–417 Gb	N/A ‡
2 x 100 bp	N/A ‡	266–333 Gb	667–833 Gb	1600–2000 Gb
2 x 150 bp	200–250 Gb	400–500 Gb	1000–1250 Gb	2400–3000 Gb
2 x 250 bp	325–400 Gb	N/A ‡	N/A ‡	N/A ‡

Specifications based on Illumina PhiX control library at supported cluster densities.

‡ N/A: not applicable

Product #	Product Name	Qty	Price (GBP)
20027466	NovaSeq™ 6000 S4 Reagent Kit (200 cycles)	1	21,653
20012866	NovaSeq™ 6000 S4 Reagent Kit (300 cycles)	1	24,962
20012865	NovaSeq™ 6000 S1 Reagent Kit (100 cycles)	1	3,325
20012864	NovaSeq™ 6000 S1 Reagent Kit (200 cycles)	1	4,460
20012863	NovaSeq™ 6000 S1 Reagent Kit (300 cycles)	1	5,353
20012862	NovaSeq™ 6000 S2 Reagent Kit (100 cycles)	1	7,704
20012861	NovaSeq™ 6000 S2 Reagent Kit (200 cycles)	1	10,543
20012860	NovaSeq™ 6000 S2 Reagent Kit (300 cycles)	1	12,368
20021665	NovaSeq™ Xp 4-Lane Kit	1	810
20021664	NovaSeq™ XP 2-Lane Kit	1	243

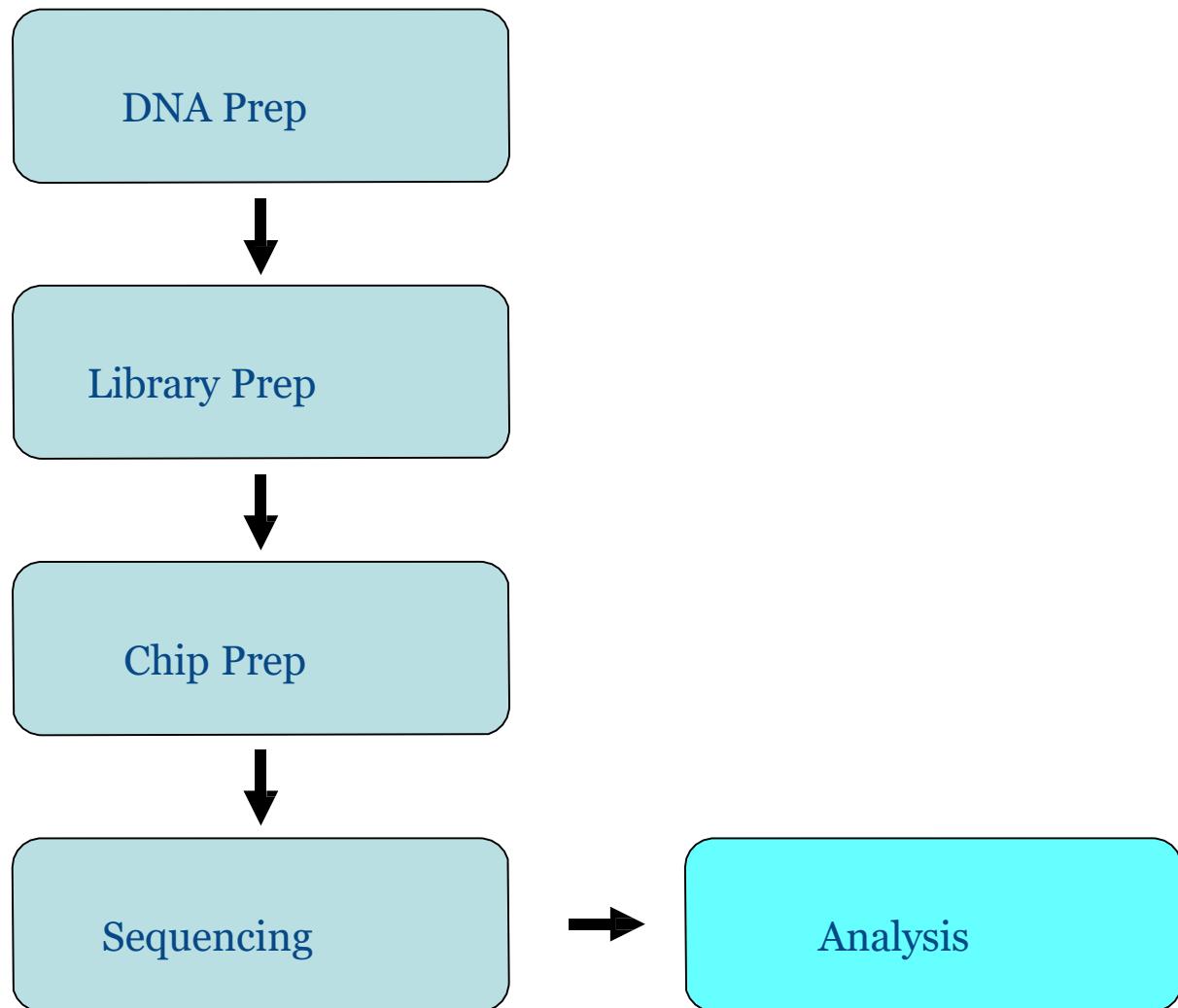
Cost per Gb depends on Kit used

	Instrument	Cycle Kit	Cycles	Number of Lanes	Max Output per flow cell (Gb) [Illumina Spec]	Max Output per lane (Gb)	Total Price per Flow Cell	Price per lane	Price per max Gb
1	HiSeq 4000	2x25	50	8	125	15.625	\$9,640.00	\$1,205.00	\$77.12
2	HiSeq 4000	2x50	100	8	250	31.25	\$12,747.50	\$1,593.44	\$50.99
3	HiSeq 4000	2x75	150	8	375	46.875	\$13,135.00	\$1,641.88	\$35.03
4	HiSeq 4000	2x150	300	8	750	93.75	\$18,480.00	\$2,310.00	\$24.64
5	HiSeq X10	2x150	300	16	1800	112.5	\$14,348.00	\$896.75	\$7.97
6	NovaSeq SP	2x50	100	1	80	80	\$2,250.00	\$2,250.00	\$28.13
7	NovaSeq SP	2x150	300	1	250	250	\$3,850.00	\$3,850.00	\$15.40
8	NovaSeq SP	2x250	500	1	400	400	\$5,500.00	\$5,500.00	\$13.75
9	NovaSeq S1	2x50	100	1	167	167	\$4,100.00	\$4,100.00	\$24.55
10	NovaSeq S1	2x100	200	1	333	333	\$5,500.00	\$5,500.00	\$16.52
11	NovaSeq S1	2x150	300	1	500	500	\$6,600.00	\$6,600.00	\$13.20
12	NovaSeq S2	2x50	100	1	417	417	\$9,500.00	\$9,500.00	\$22.78
13	NovaSeq S2	2x100	200	1	833	833	\$13,000.00	\$13,000.00	\$15.61
14	NovaSeq S2	2x150	300	1	1250	1250	\$15,250.00	\$15,250.00	\$12.20
15	NovaSeq S4	2x100	200	1	2000	2000	~\$20000	~\$20000	~\$10
16	NovaSeq S4	2x150	300	1	3000	3000	~\$21000	~\$21000	~\$7

YOUR text here



Illumina workflow





The Software Challenge

- » Analysis pipeline (images -> sequence)
- » Tracking of projects, samples, runs + analysis
- » Data storage
- » QC (manual + automatic)
- » Sequence alignment + assembly
- » Variant calling
- » Interpretation



Costs

- » 1st human genome:
~ \$500 Million
- » Current capillary cost :
~ \$10 Million
- » Illumina cost:
~ \$1,000 on Xten
~ \$500 on Novaseq 6000
~ \$200-300 on Novaseq X



Limitations of Illumina Sequencing Technology

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

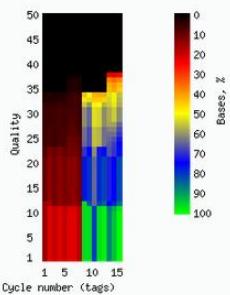
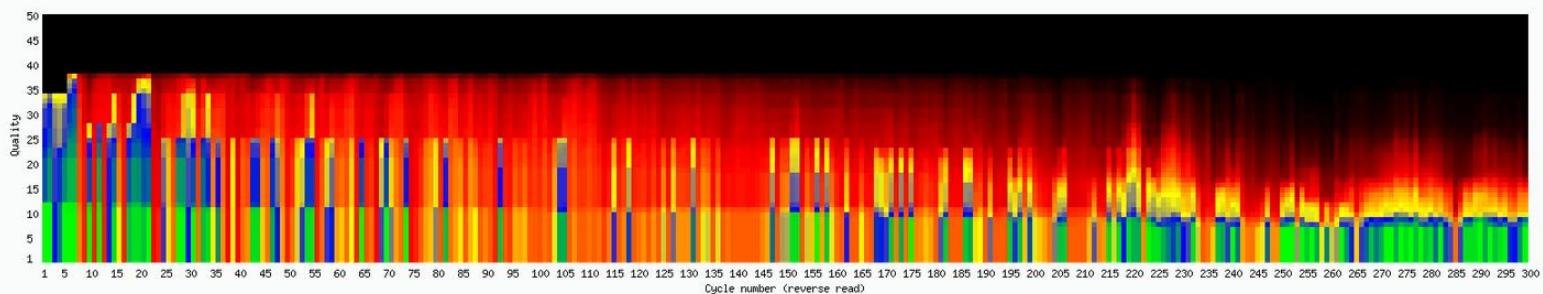
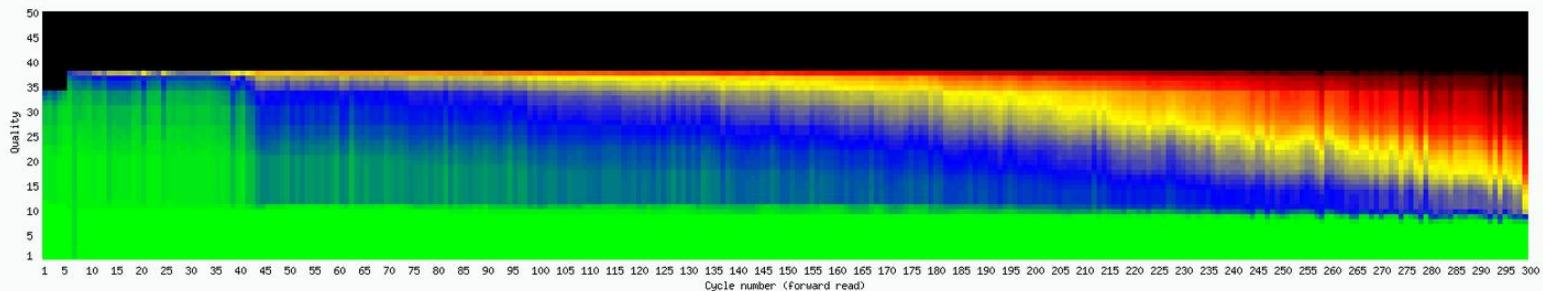
Limitations of Illumina Sequencing Technology

- » Some systematic errors
 - » Difficult to spot rare variants (<1%)
 - » See Duplex seq by Schmitt et al.,
 - » Nanoseq <https://doi.org/10.1038/s41586-021-03477-4>
- » Low complexity templates
- » Sequencing short fragments doesn't give any long range information
- » Index Hopping

Limitations of Illumina Sequencing Technology

- » Some systematic errors
 - » Difficult to spot rare variants (<1%)
- » Low complexity templates
 - » Add complex library to 30%, phase, ensure variation at start of read
- » Sequencing short fragments doesn't give any long range information
- » Index Hopping

Low quality 16S run



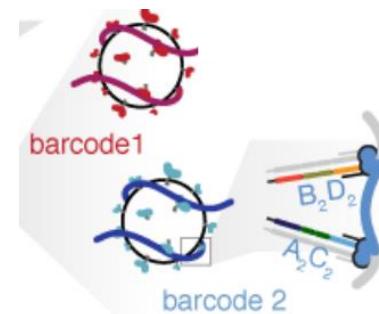
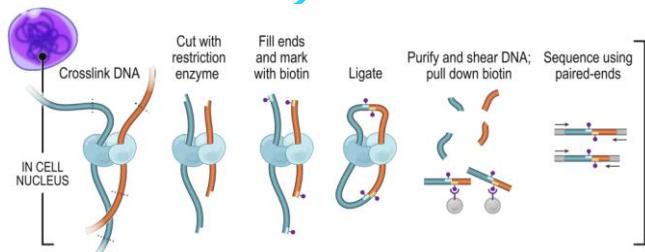
Limitations of Illumina Sequencing Technology

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

Illumina long range approaches

- » 10 x Genomics
- » HiC (eg Arima and Dovetail)

- » Tell-Seq
- » Haplotagging (Meier et al., BioRxIV 2020 doi: <https://doi.org/10.1101/2020.05.25.113688>)
- » Constellation

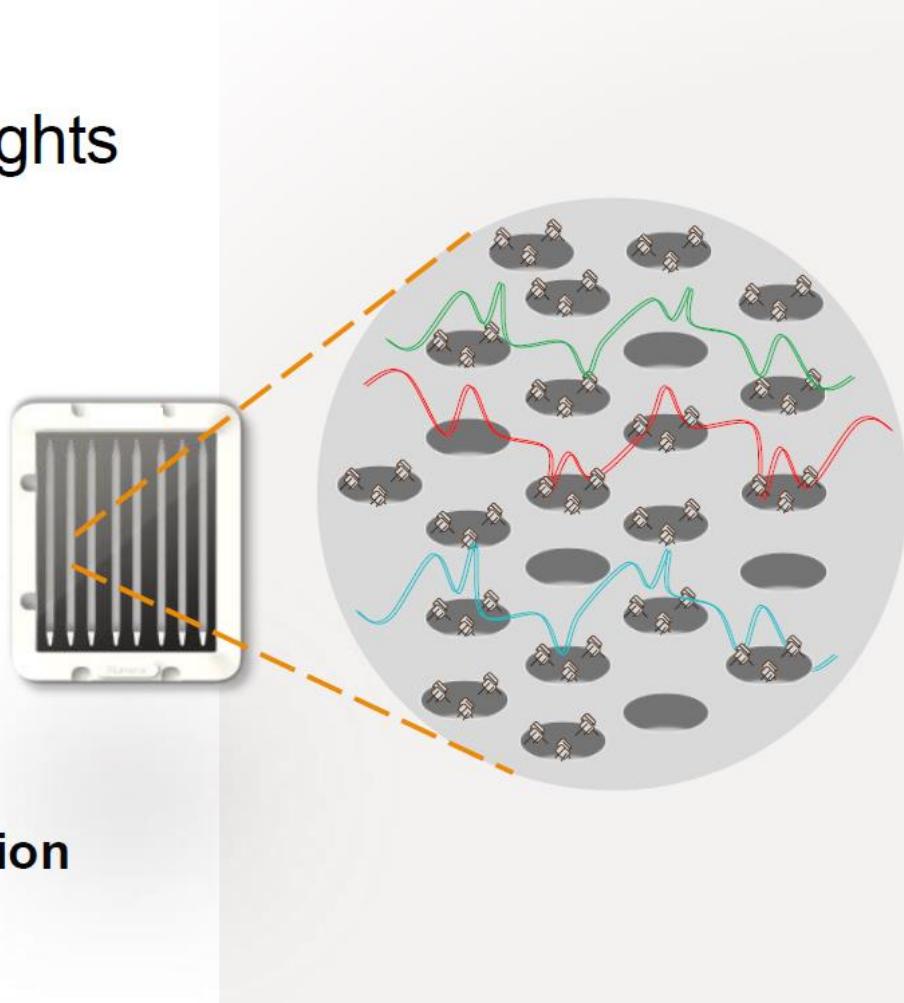


Limitations of Illumina Sequencing Technology

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

Constellation technology: Simplest workflow, powerful insights

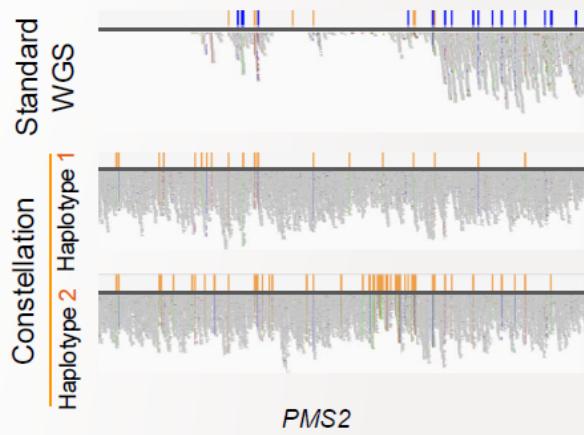
- **Flow unfragmented DNA**
(no library prep needed)
- **Capture DNA**
on-flow cell fragmentation
- **Standard sequencing**
- **Utilize proximity nano well information**
to extract long-range data





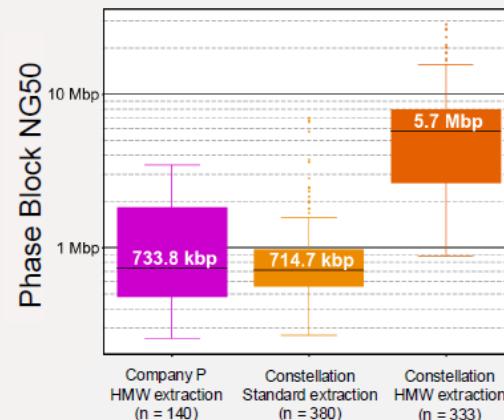
Constellation technology delivers a more complete and comprehensive genome

Resolve mapping ambiguities



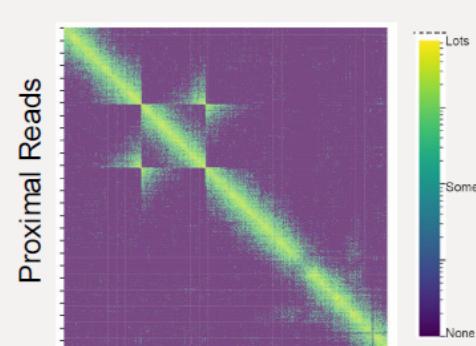
Improved coverage through recovery of ambiguously mapped reads

Ultra-long phasing



Megabase-long phase blocks to phase pathogenic variants

Structural variant insights



Courtesy of Rady Children's Hospital:
Colocation plot of balanced inversion

Limitations of Illumina Sequencing Technology

» Index Hopping

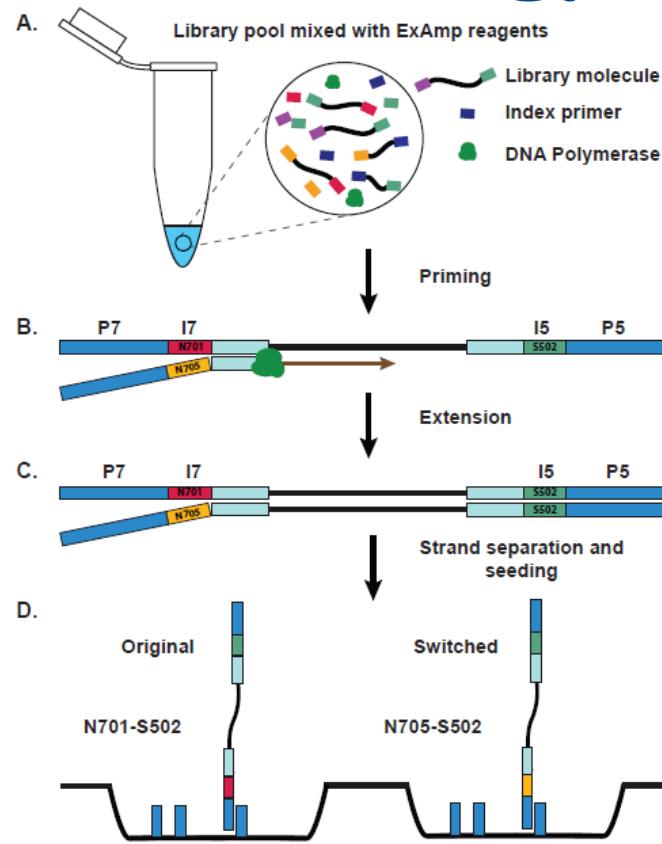


Figure 1

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



Thanks!





Any Questions ?

mq1@sanger.ac.uk