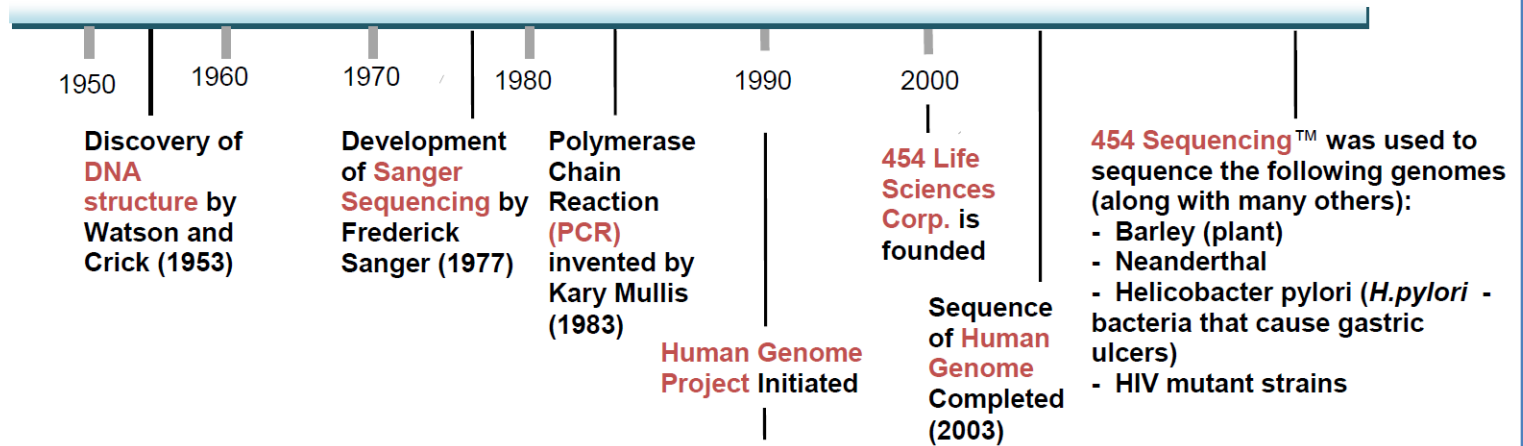


Metagenomika – NGS (454, Illumina, IonTorrent)

Petra Vídeňská, Ph.D.

Next Generation Sequencing

History of Genome Sequencing

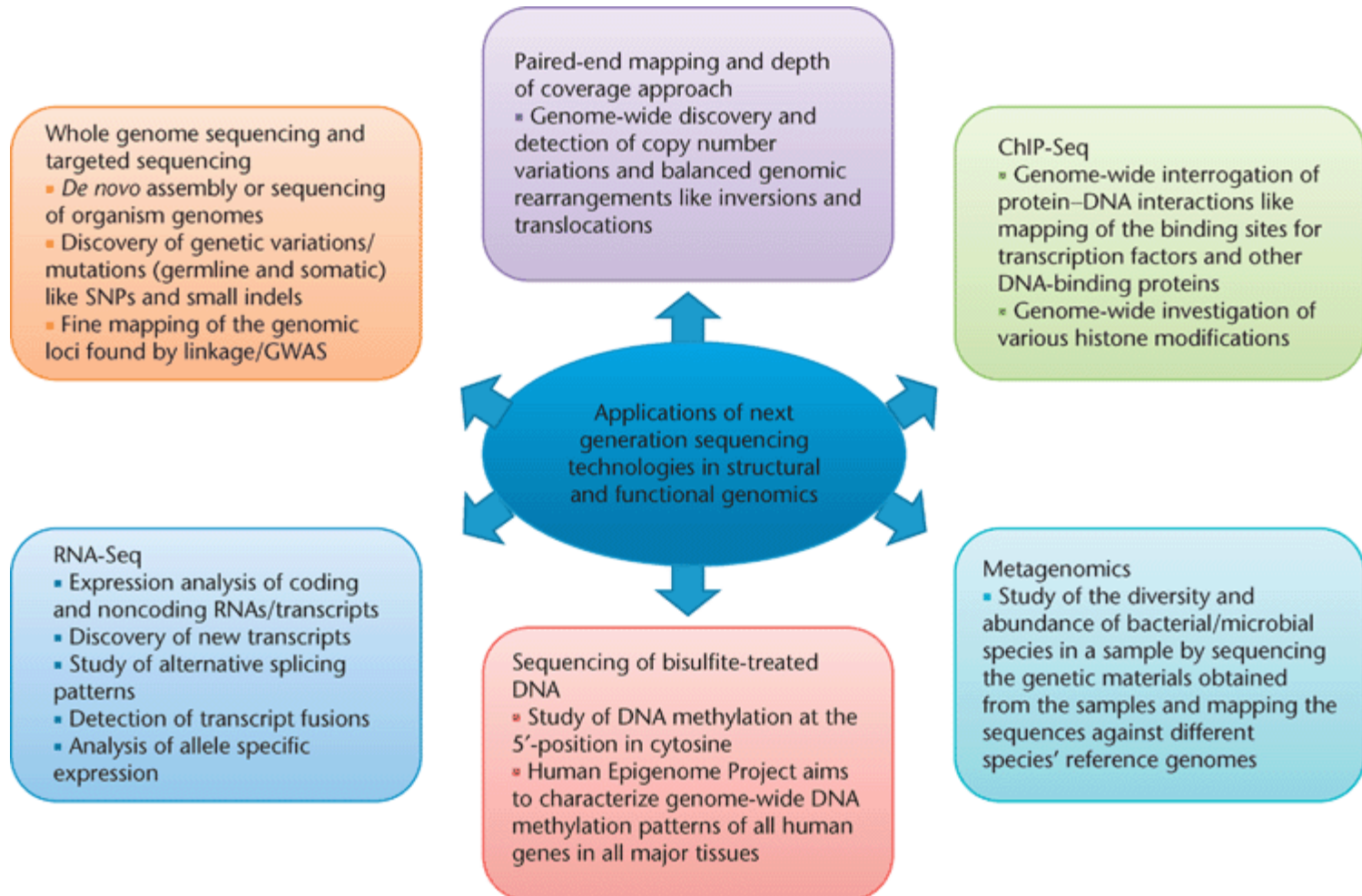


- The total number of genes is estimated at around 30,000--much lower than previous estimates of 80,000 to 140,000.
- Almost all (99.9%) nucleotide bases are exactly the same in all people.
- The functions are unknown for over 50% of discovered genes.

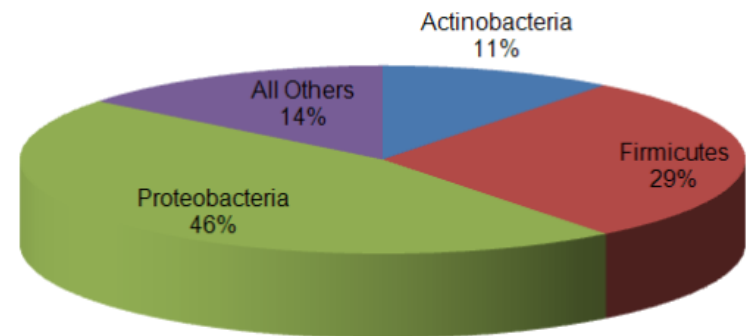
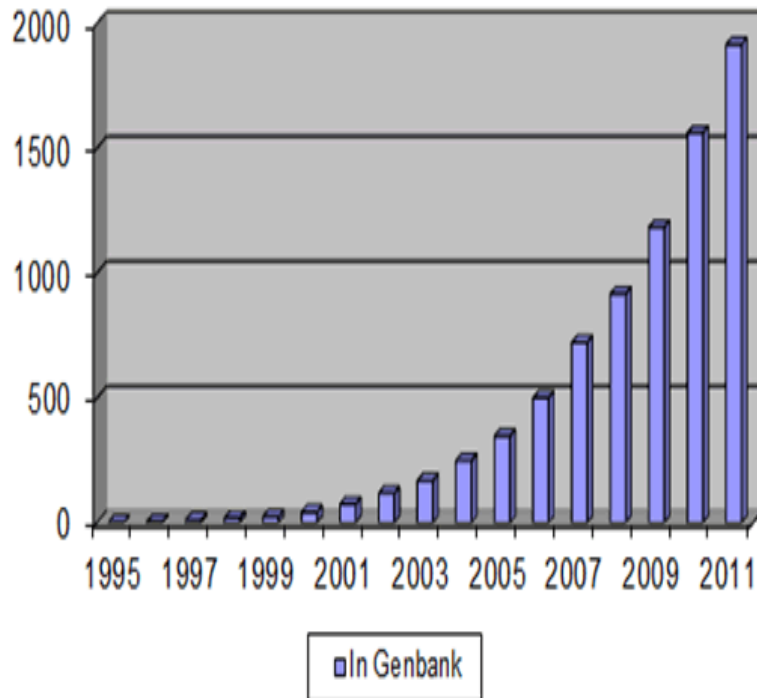
Source: U.S. Department of Health and Human Services
Base URL: <http://genome.gov>

35 bp identifier

Využití next generation sekvenování

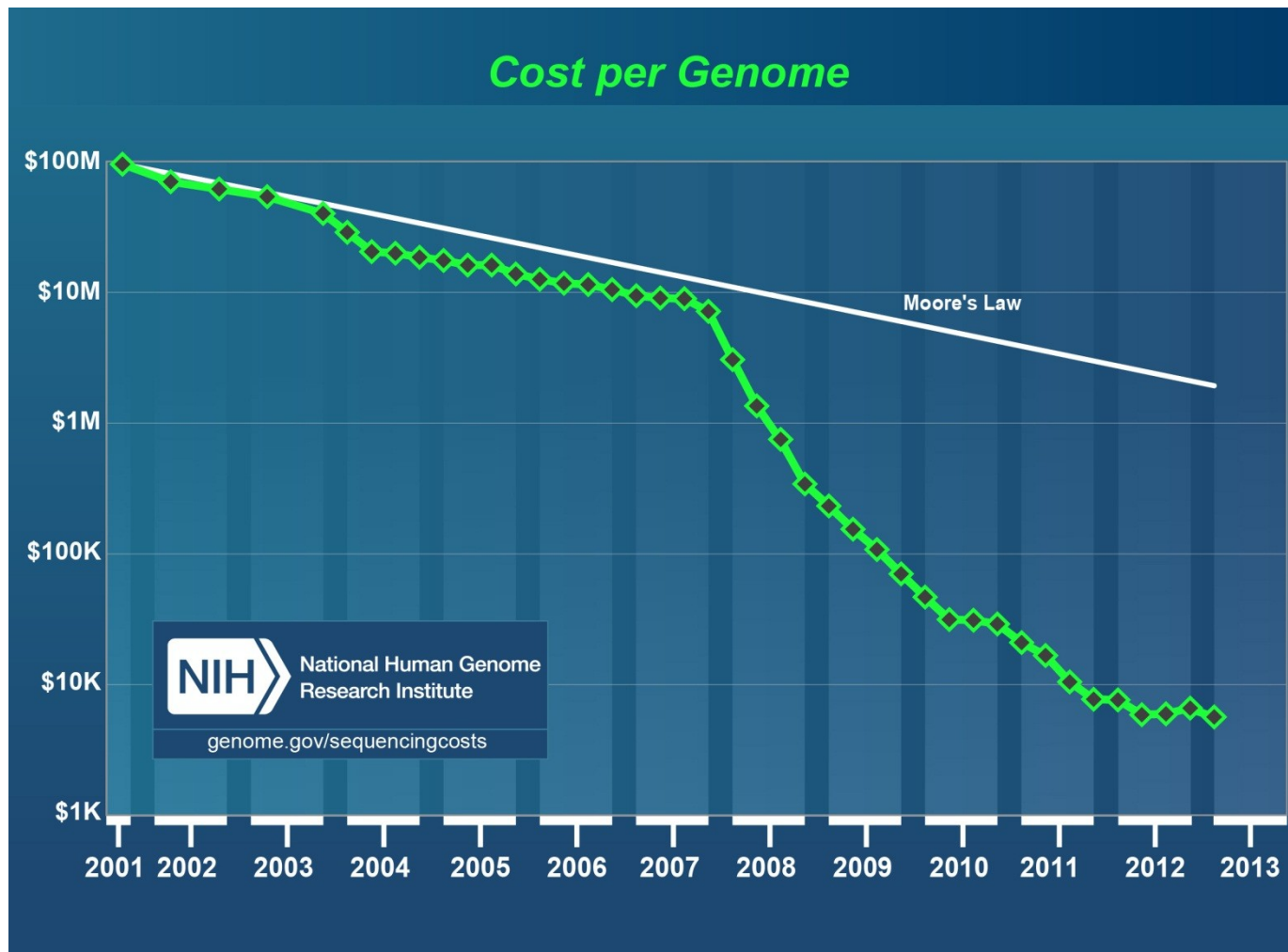


Počet kompletně osekvenovaných genomů



www.genomesonline.org

Náklady na sekvenování genomu



Sekvenování nové (druhé) generace

= masivní paralelní sekvenování

- Umožňuje najednou sekvenaci miliónů různých fragmentů DNA (cDNA, i RNA) o délce cca 30-1000 bp (dle zvolené platformy a sekvenačního kitu)
- Dochází k zmnožení fragmentu (emulzní PCR, mŕstková amplifikace) – větší signál při inkorporaci nukleotidů během sekvenace, umožňující detekci

Sekvenování 3. generace

- Nevyužívá amplifikace za účelem zvýšení signálu (měla by být vyšší přesnost –accuracy)
- Produkuje dlouhá čtení
- Dobrá prosekvenovanost GC bohatých oblastí
- Epigenetika
- Zatím dvě dostupné technologie – PacBio a Nanopore (MinION)
- Illumina chystá nový systém

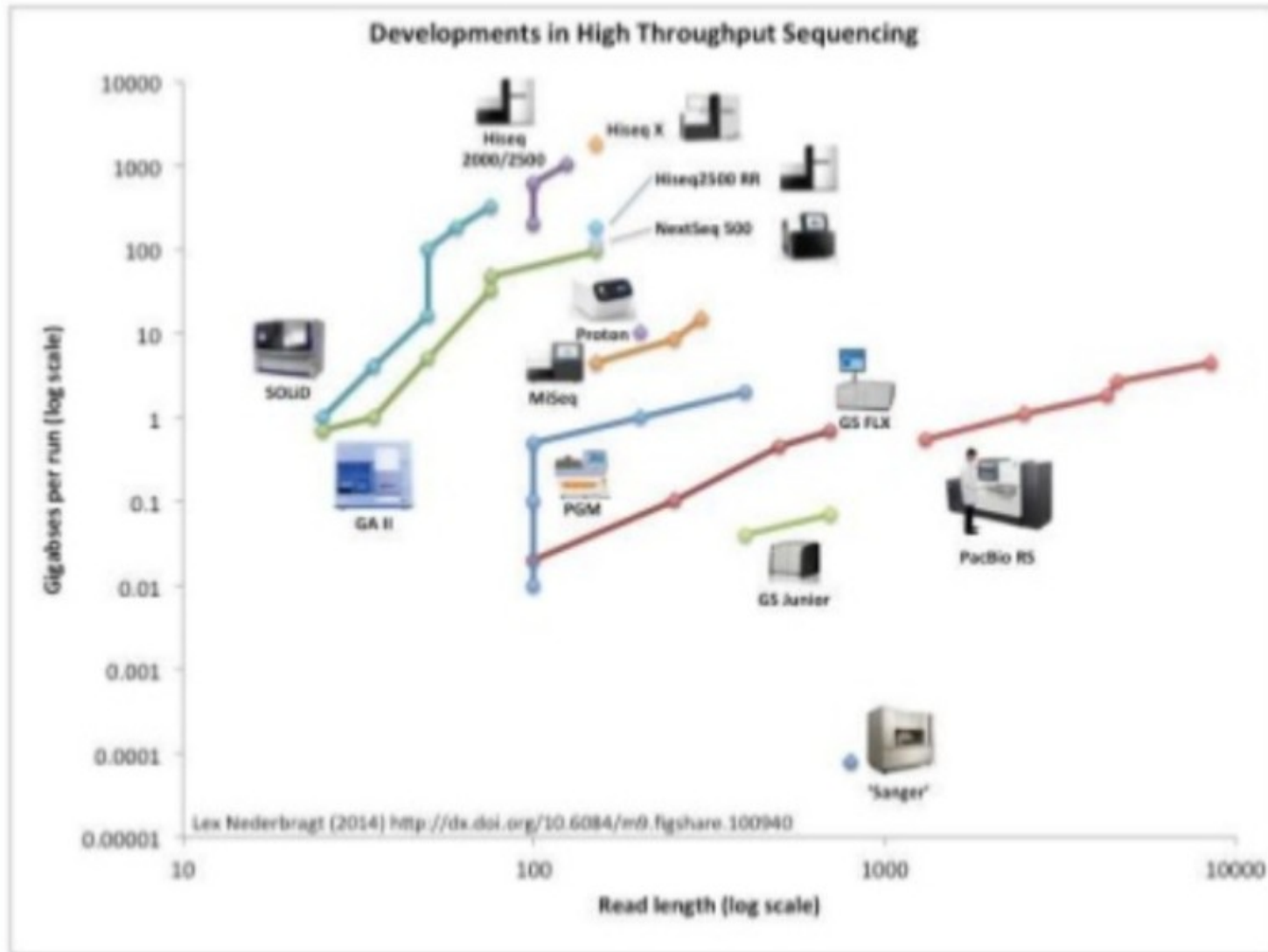
Dostupné platformy

- 454 (Roche)
- SOLiD (Life Technologies)
- Illumina (Illumina)
- Ion Torrent (Life Technologies)
- **PACBIO, Sequel System (Pacific BioSciences)**
- **MiniION (Oxford Nanopore Technologies)**
- **BGISEQ-500 (BGI)**



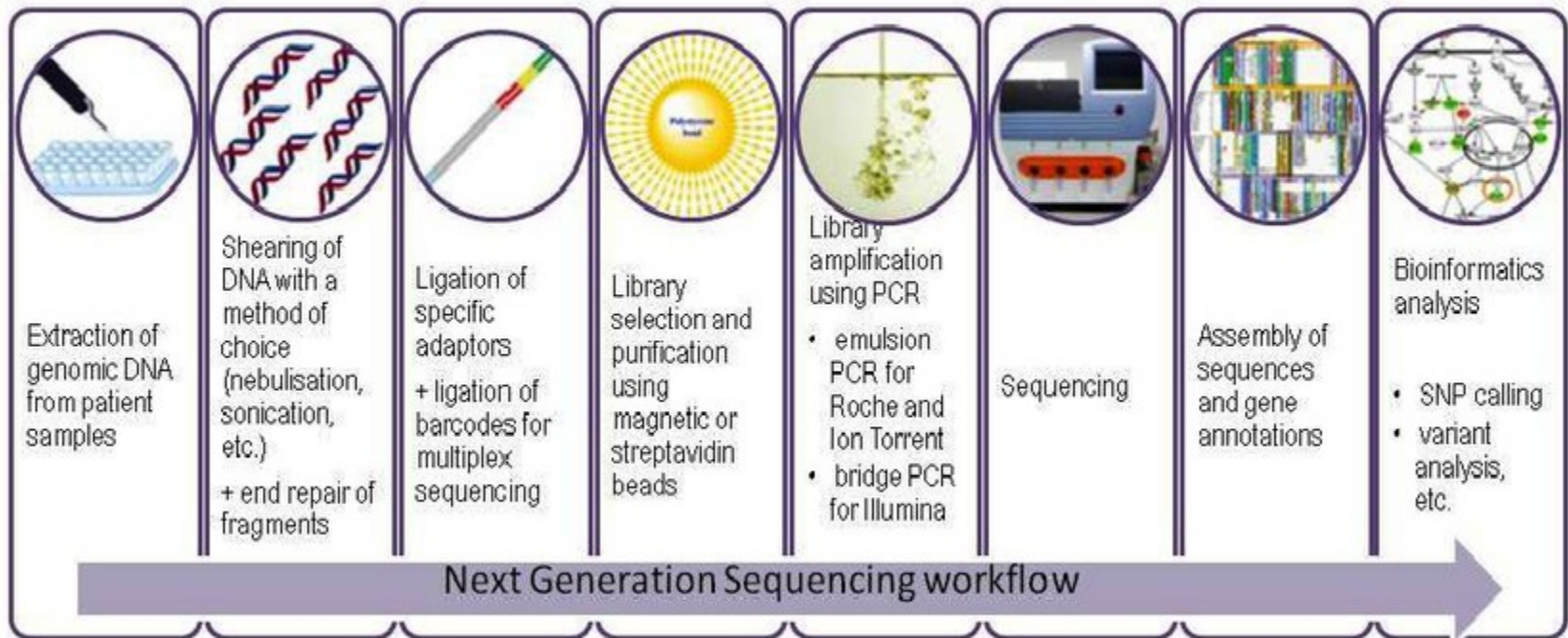
2 vs 3 generace

Sekvenování nové generace



Newest Illumina HiSeq X 10 > 1 Tb of sequence data

Sekvenační workflow



454

- Roche
- 454 GS Junior (35 MB) x 454 GS FLX (700 MB)



- Příprava templátu: EM PCR na kuličkách
- Sekvenace syntézou
- Detekce chemiluminiscenční - pyrosekvenování

454 Shotgun – příprava knihovny

Nebulization



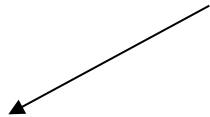
DNA End Repair



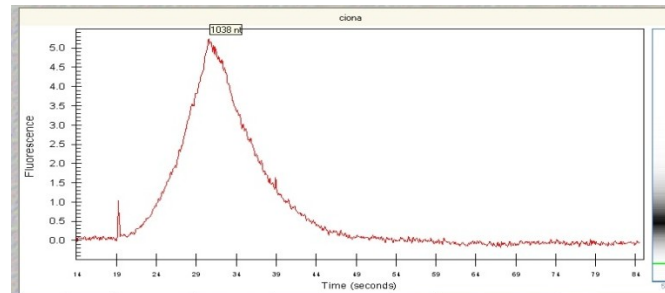
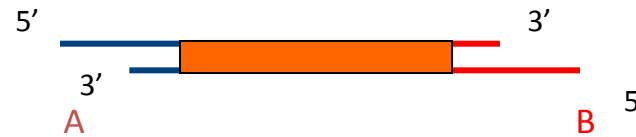
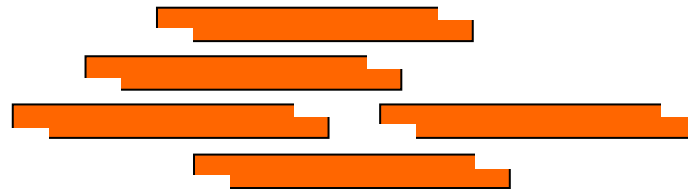
Adaptor Ligation (A&B)



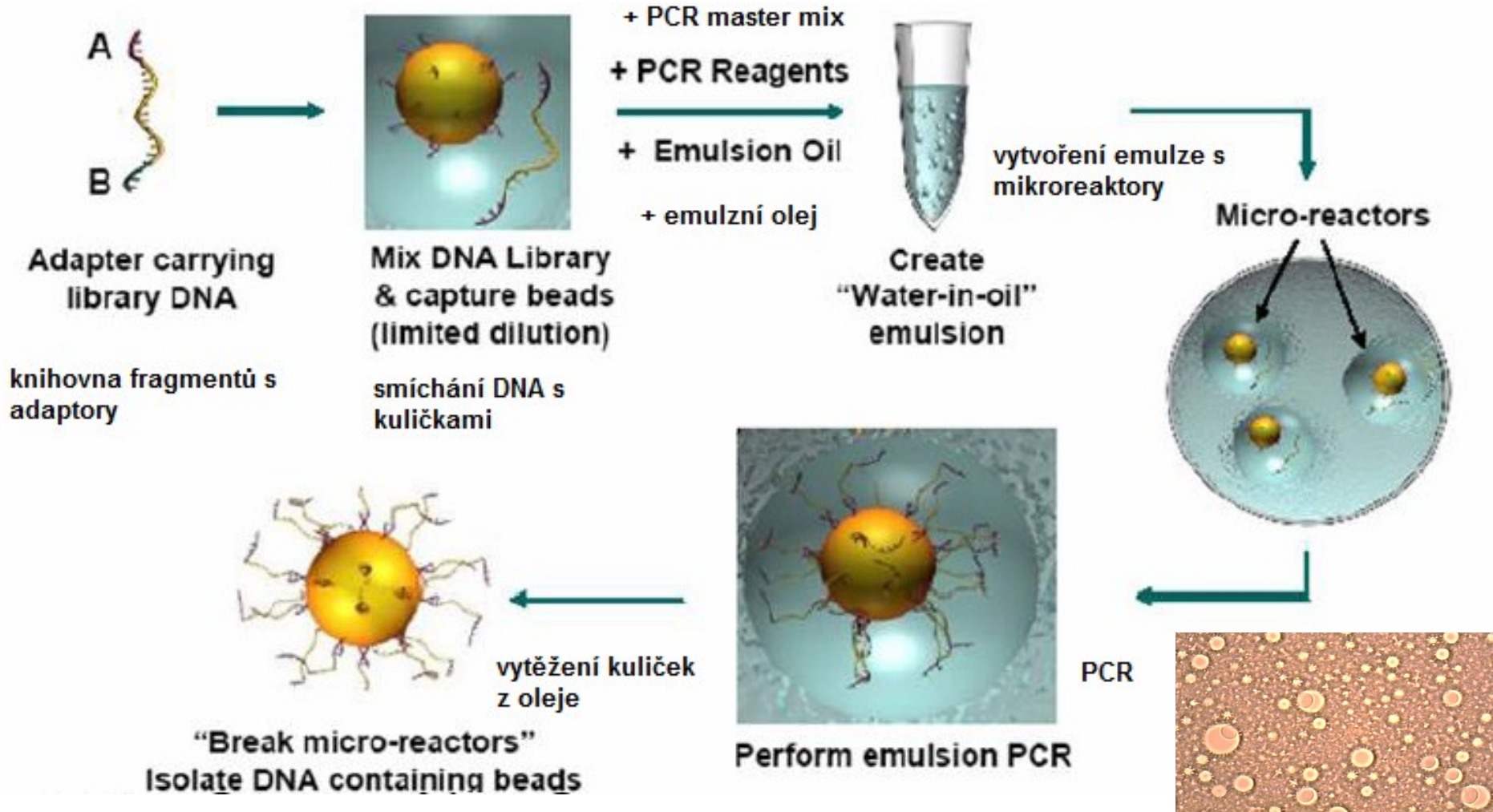
DNA End Repair



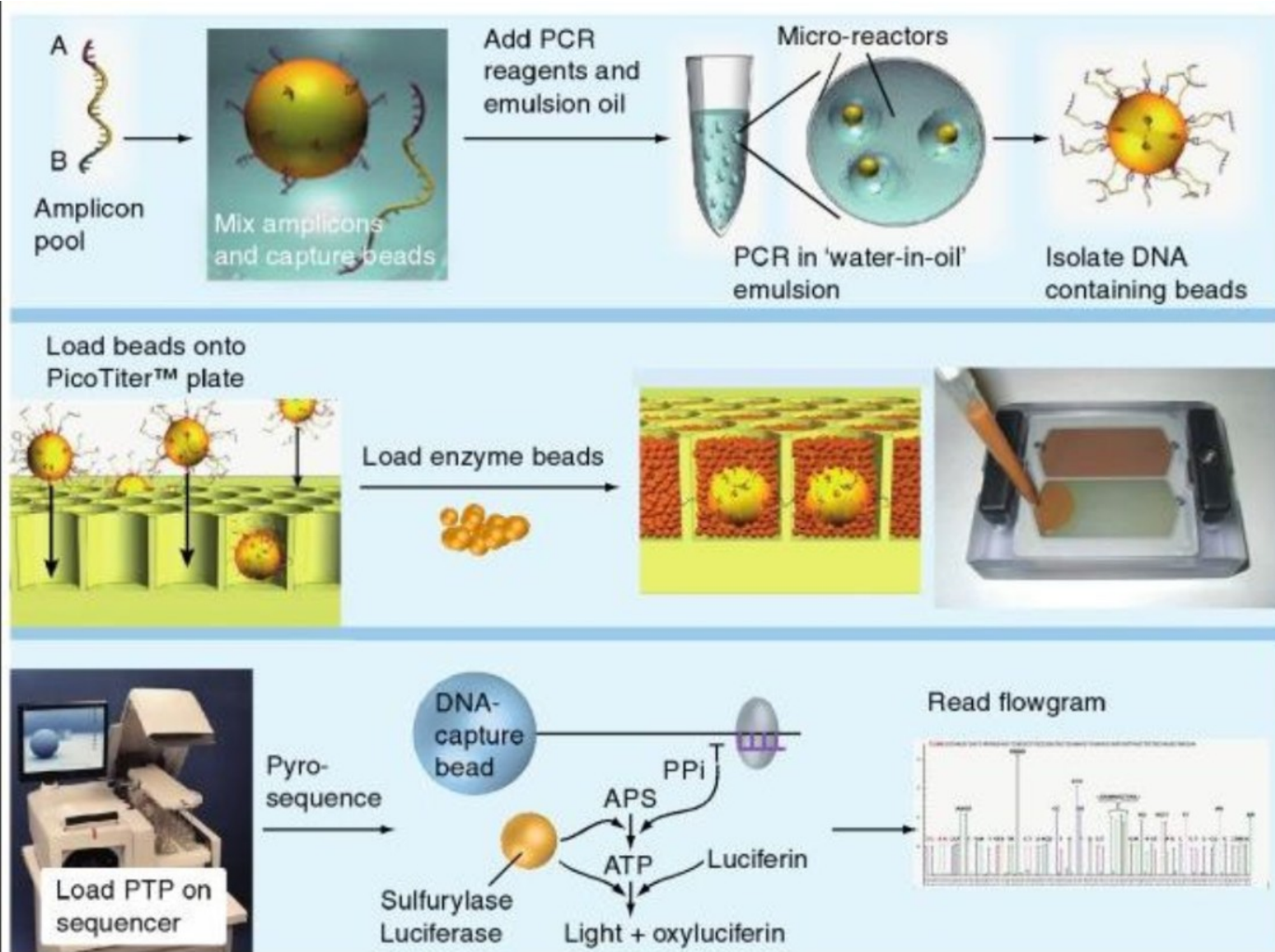
Library Quantification,
read length check



454

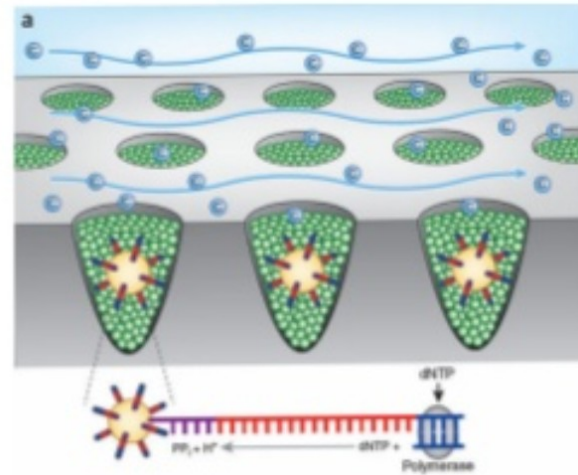
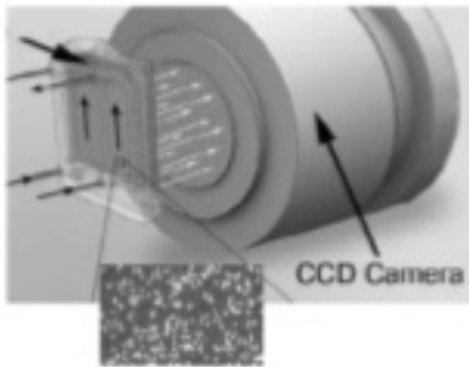
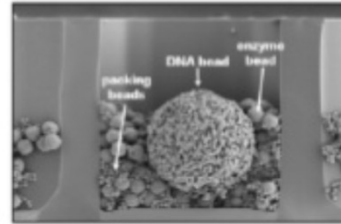
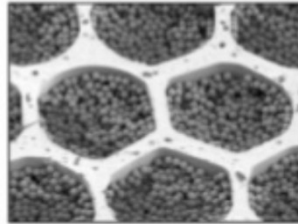
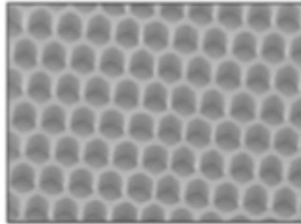


454



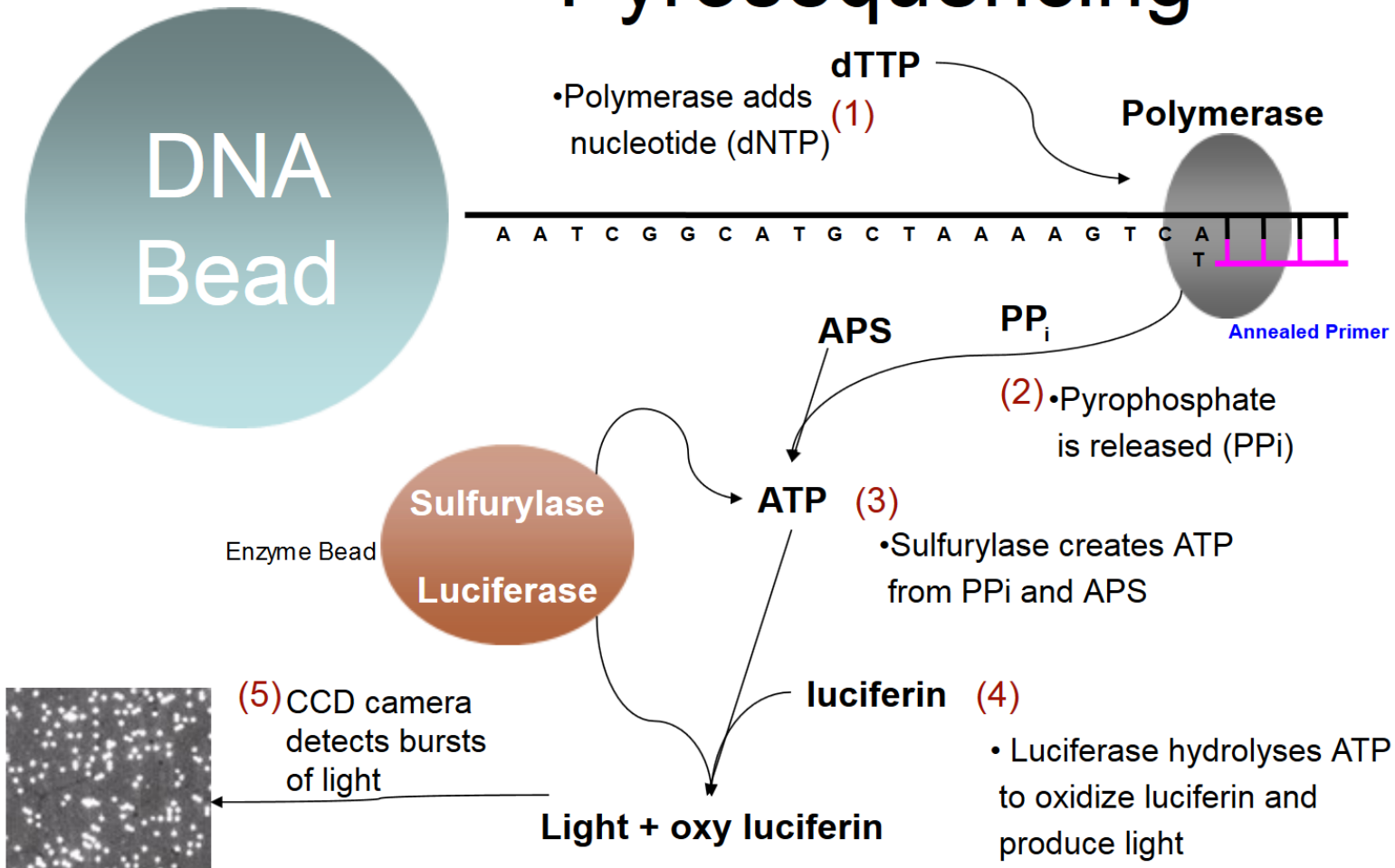
<https://www.youtube.com/watch?v=nFfgWGF e0aA>

454

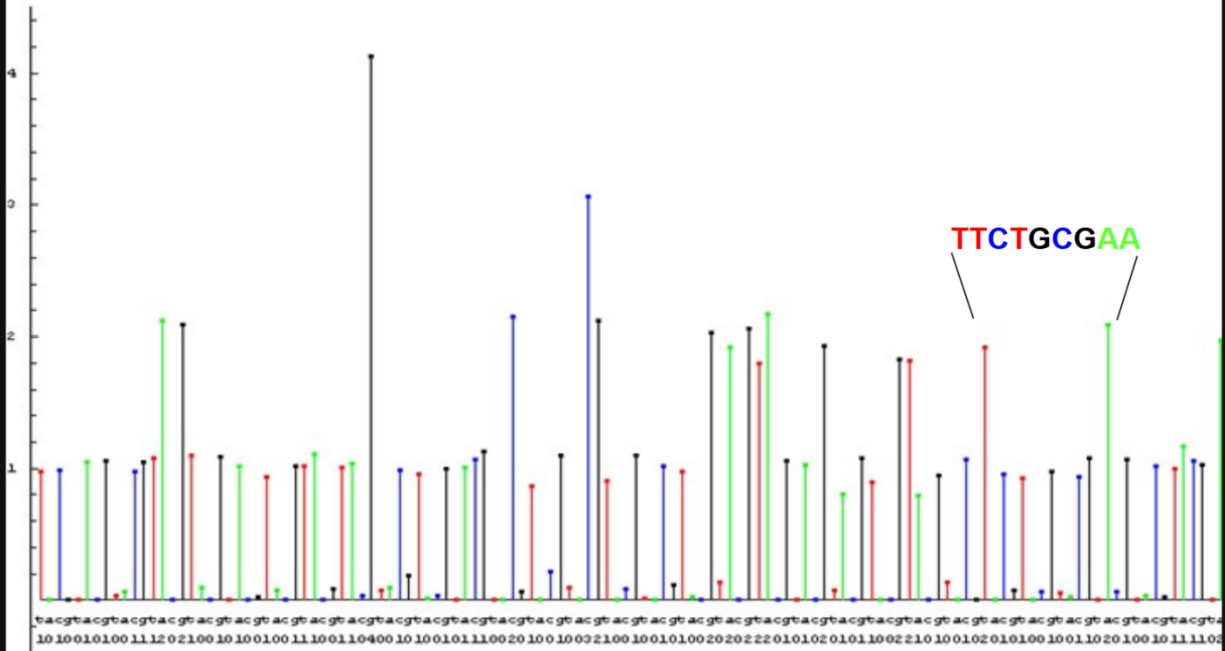


454

Pyrosequencing



Base Calling via Flowgram



454 – podrobný workflow

- <http://cfgbc.mf.uni-lj.si/people/damjana/teaching/fg-fkkt/4-GS-JuniorTechnology.pdf>

Illumina



- Příprava templátu: hybridizace na sklíčku, tvorba klastrů
- Sekvenace syntézou
- Detekce fluorescence odštěpené značky z reverzního terminátoru (nukleotidu)

Přístroje Illumina



MiniSeq System



MiSeq Series +



NextSeq Series +

Popular Applications & Methods	Key Application ■	Key Application ■	Key Application ■
Large Whole-Genome Sequencing (human, plant, animal)			●
Small Whole-Genome Sequencing (microbe, virus)	●	●	●
Exome Sequencing			●
Targeted Gene Sequencing (amplicon, gene panel)	●	●	●
Whole-Transcriptome Sequencing			●
Gene Expression Profiling with mRNA-Seq			●
Targeted Gene Expression Profiling	●	●	
miRNA & Small RNA Analysis	●	●	●
DNA-Protein Interaction Analysis		●	●
Methylation Sequencing			●
16S Metagenomic Sequencing		●	●

How to Choose a Benchtop Sequencer

This Benchtop Sequencing Buyer's Guide has tips to help you make a smooth transition to next-generation sequencing and select the best benchtop sequencing system to achieve your research objectives.

[Read Benchtop Buyer's Guide >](#)

Run Time	4–24 hours	4–55 hours	12–30 hours
Maximum Output	7.5 Gb	15 Gb	120 Gb
Maximum Reads Per Run	25 million	25 million*	400 million
Maximum Read Length	2 × 150 bp	2 × 300 bp	2 × 150 bp

Přístroje Illumina



NextSeq Series



HiSeq Series



HiSeq X Series



NovaSeq Series

Popular Applications & Methods	Key Application	Key Application	Key Application	Key Application
Large Whole-Genome Sequencing (human, plant, animal)				
Small Whole-Genome Sequencing (microbe, virus)				
Exome Sequencing				
Targeted Gene Sequencing (amplicon, gene panel)				
Whole-Transcriptome Sequencing				
Gene Expression Profiling with mRNA-Seq				
miRNA & Small RNA Analysis				
DNA-Protein Interaction Analysis				
Methylation Sequencing				
Shotgun Metagenomics				

Optimized NGS Sample Tracking and Workflows

See how BaseSpace Clarity LIMS (Laboratory Information Management System) enabled this large genomics lab to standardize lab procedures and cope with increasing sample volumes from diverse clients.

[Read Case Study >](#)

Run Time	12–30 hours	< 1–3.5 days (HiSeq 3000/HiSeq 4000) 7 hours–6 days (HiSeq 2500)	< 3 days	19–40 hours [‡]
Maximum Output	120 Gb	1500 Gb	1800 Gb	6000 Gb [§]
Maximum Reads Per Run	400 million	5 billion	6 billion	20 billion
Maximum Read Length	2 × 150 bp	2 × 150 bp	2 × 150 bp	2 × 150 bp

MiSeq specifications

READ LENGTH (BP)	TOTAL TIME*	OUTPUT
1 × 36 (V2)	~4 hrs	540-610 Mb
2 × 25 (V2)	~5.5 hrs	750-850 Mb
2 × 150 (V2)	~24 hrs	4.5-5.1 Gb
2 × 250 (V2)	~39 hrs	7.5-8.5 Gb
2 × 75 (V3)	~20hrs	3.3-3.8 Gb
2 × 300 (V3)	~ 55hrs	13.2-15 Gb

RUN TYPE	READS PASSING FILTER†	
	V2	V3
Single Reads	12-15 M	22-25 M
Paired-End Reads	24-30 M	44-50 M



NextSeq specifications

NextSeq 500 Sequencing System Performance Parameters

NEXTSEQ 500 HIGH OUTPUT KIT *

READ LENGTH	TOTAL TIME†	OUTPUT
2 × 150 bp	~29 hrs	100-120 Gb
2 × 75 bp	18 hrs	50-60 Gb
1 × 75 bp	11 hrs	25-30 Gb

NEXTSEQ 500 MID OUTPUT KIT *

READ LENGTH	TOTAL TIME†	OUTPUT
2 × 150 bp	26 hrs	32.5-39 Gb
2 × 75 bp	15 hrs	16.25-19.5 Gb

Reads Passing Filter

NEXTSEQ 500 HIGH OUTPUT KIT

Single Reads	Up to 400 Million
Paired-End Reads	Up to 800 million

NEXTSEQ 500 MID OUTPUT KIT

Single Reads	Up to 130 Million
Paired-End Reads	Up to 260 Million



HiSeq 2500 specifications

High Output Run Mode*

HISEQ SBS V4 PRE-RELEASE SPECIFICATIONS

TRUSEQ SBS V3

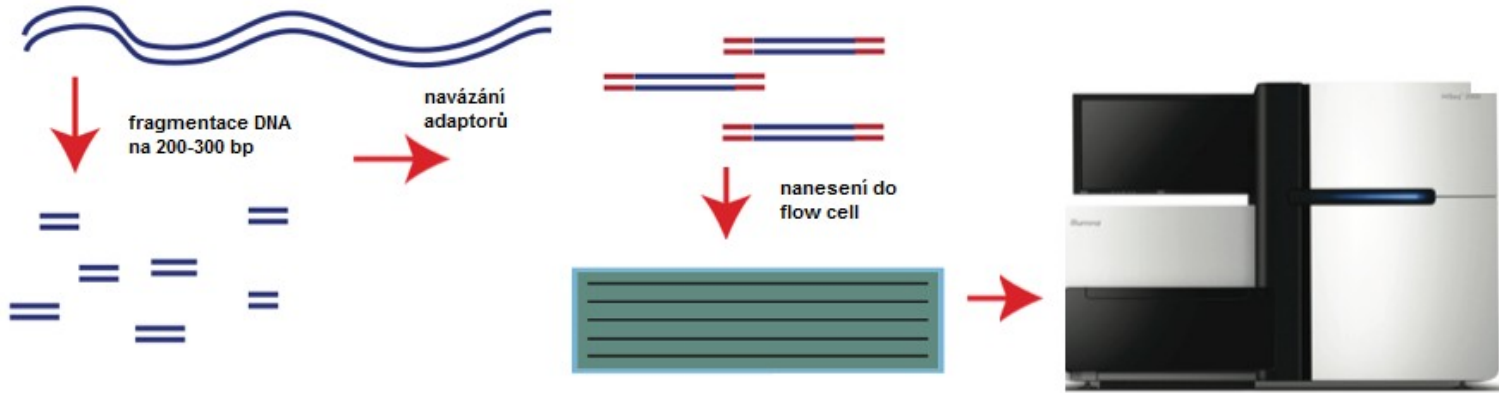
Read length	HISEQ SBS V4 PRE-RELEASE SPECIFICATIONS			TRUSEQ SBS V3		
	Dual Flow Cell	Single Flow Cell	Dual Flow Cell Run Time	Dual Flow Cell	Single Flow Cell	Dual Flow Cell Run Time
1×36	128-144 Gb	64-72 Gb	29 hrs	95-105 Gb	47-52 Gb	2 days
2×50	360-400 Gb	180-200 Gb	2.5 days	270-300 Gb	135-150 Gb	5.5 days
2×100	720-800 Gb	360-400 Gb	5 days	540-600 Gb	270-300 Gb	11 days
2×125	900-1 Tb	450-500 Gb	6 days	N/A	N/A	N/A
Reads Passing Filter	Up to 4 billion single read or 8 billion paired-end reads	Up to 2 billion single read or 4 billion paired-end reads		Up to 3 billion single read or 6 billion paired-end reads	Up to 1.5 billion single read or 3 billion paired-end reads	
Quality	Greater than 85% of bases above Q30 at 2×50 bp Greater than 80% of bases above Q30 at 2×100 bp Greater than 80% of bases above Q30 at 2×125 bp			Greater than 85% of bases above Q30 at 2×50 bp Greater than 80% of bases above Q30 at 2×100 bp		

Rapid Run Mode*

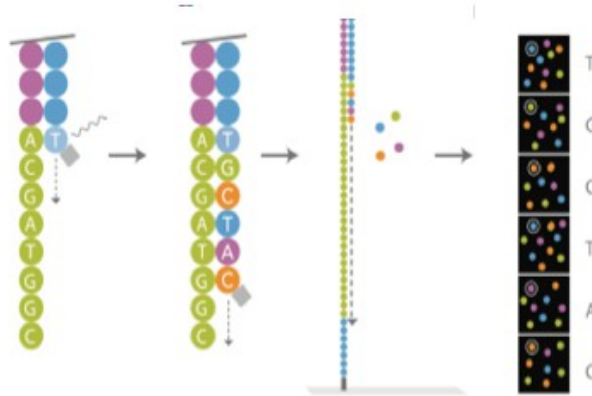
Read length	Dual Flow Cell	Single Flow Cell	Dual Flow Cell Run Time
1×36	18-22 Gb	9-11 Gb	7 hr
2×50	50-60 Gb	25-30 Gb	16 hr
2×100	100-120 Gb	50-60 Gb	27 hr
2×150	150-180 Gb	75-90 Gb	40 hr
Reads Passing Filter	Up to 600 million single read or 1.2 billion paired-end reads	Up to 300 million single read or 600 million paired-end reads	
Quality	Greater than 85% of bases above Q30 at 2×50 bp Greater than 80% of bases above Q30 at 2×100 bp Greater than 75% of bases above Q30 at 2×150 bp		



Illumina

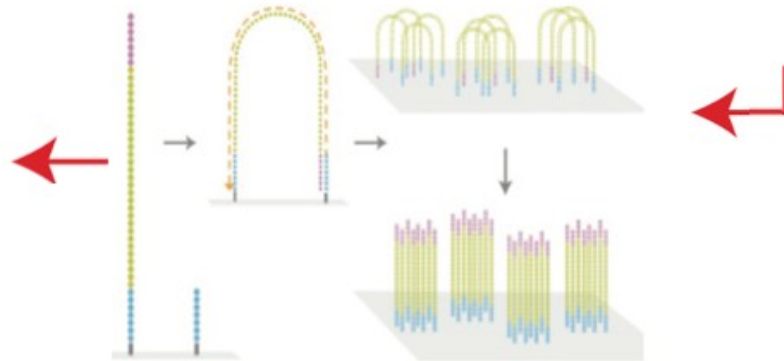


sekvence pomocí SBS s reverzibilními fluorescenčními terminátory



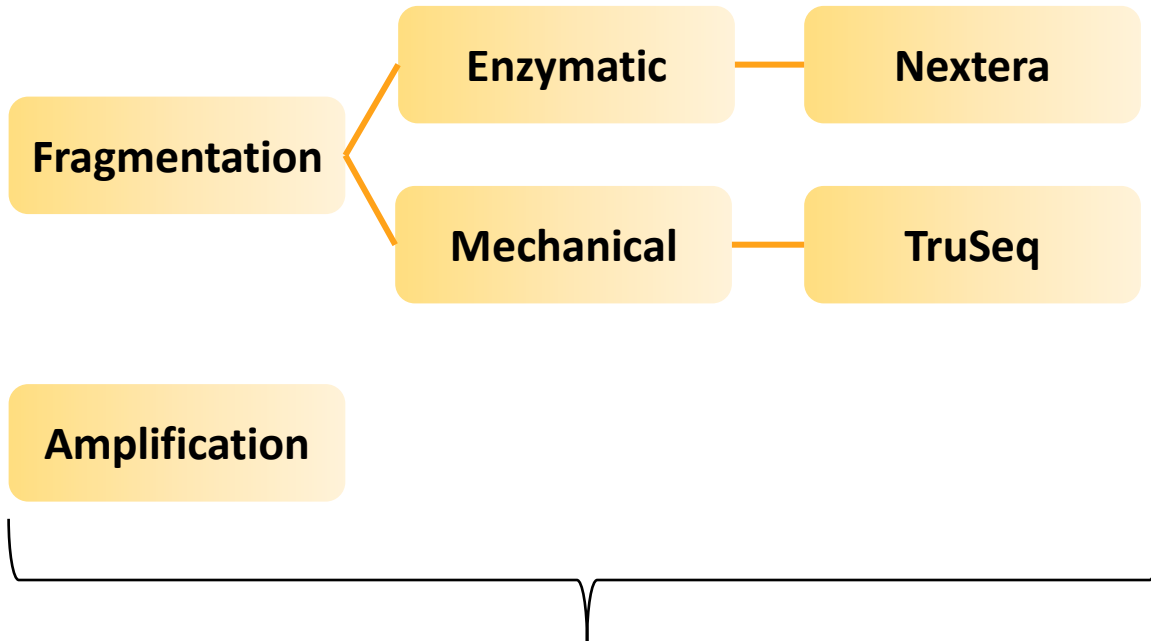
analýza obrazu - určení sekvence v jednotlivých klastrech

můstková amplifikace PCR na pevné fázi



Příprava knihovny

Fragment

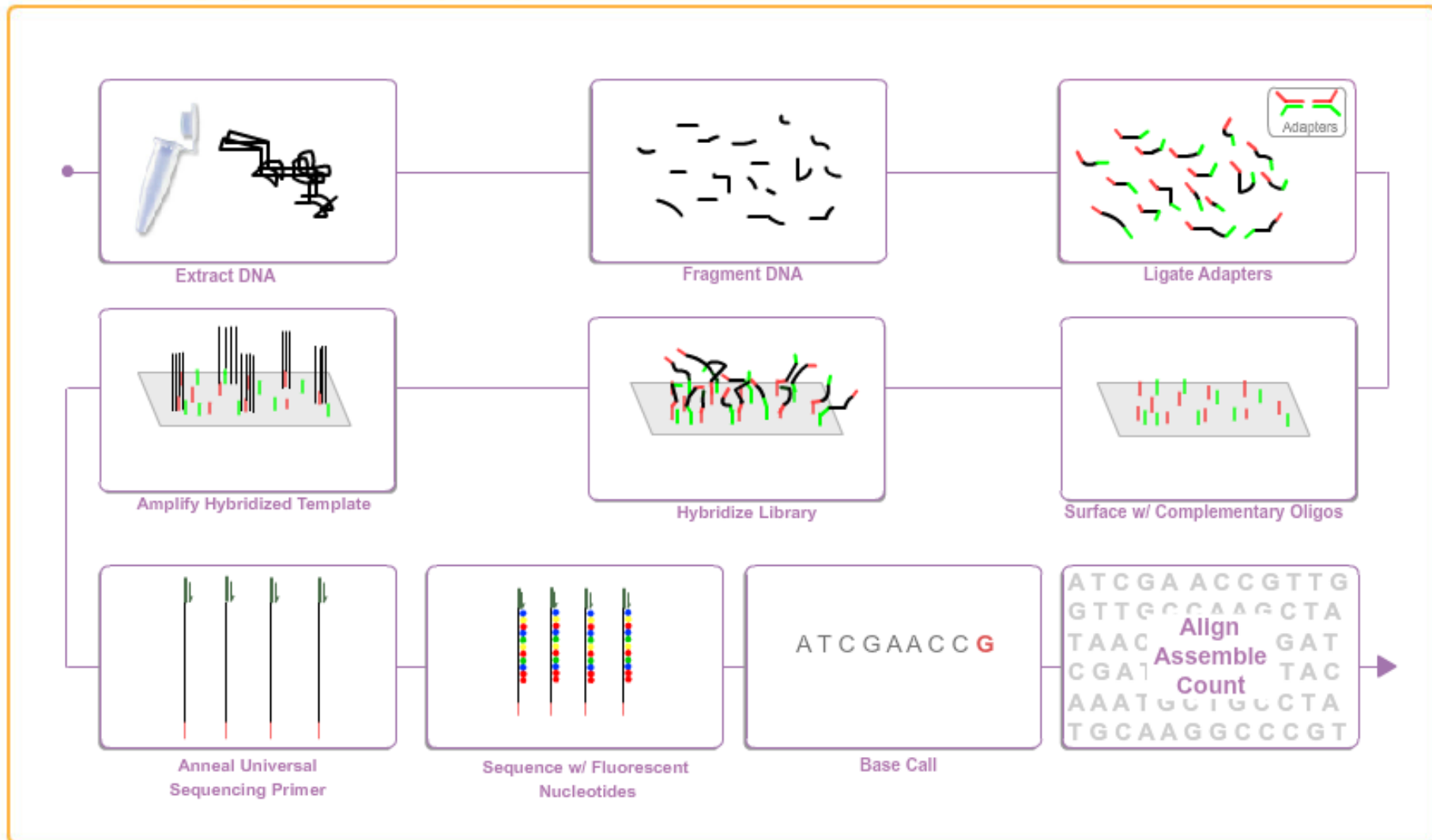


Adapter



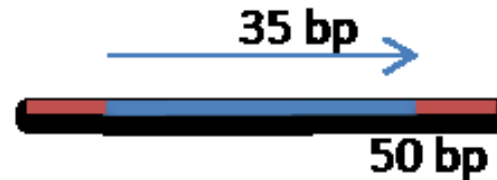
Sekvenační technologie

Sequencing by Synthesis (SBS) Overview

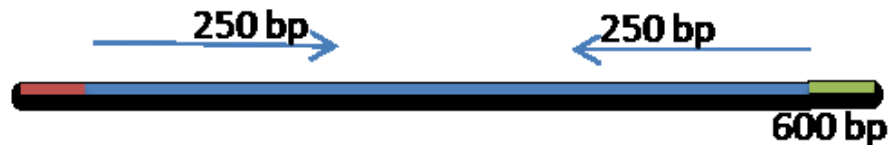


Single vs Pair End Read

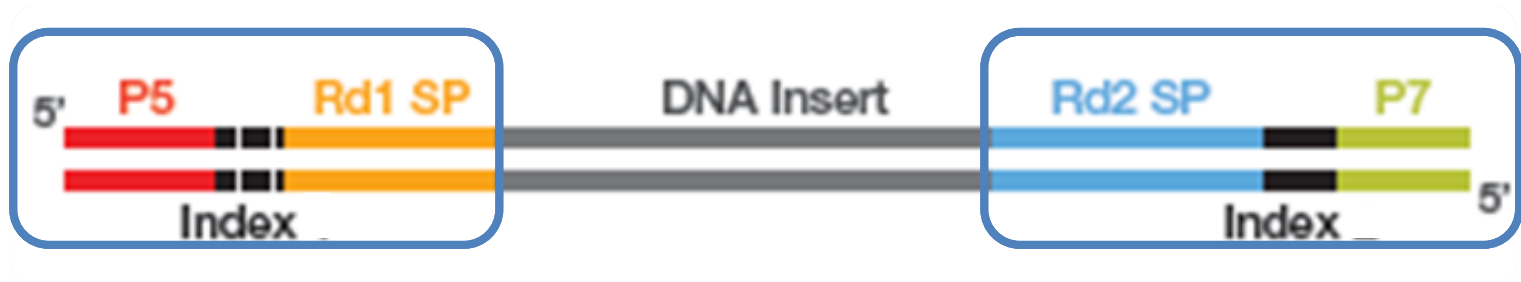
- Single reads
 - Small RNA



- Pair-End reads
 - DeNovo assembling



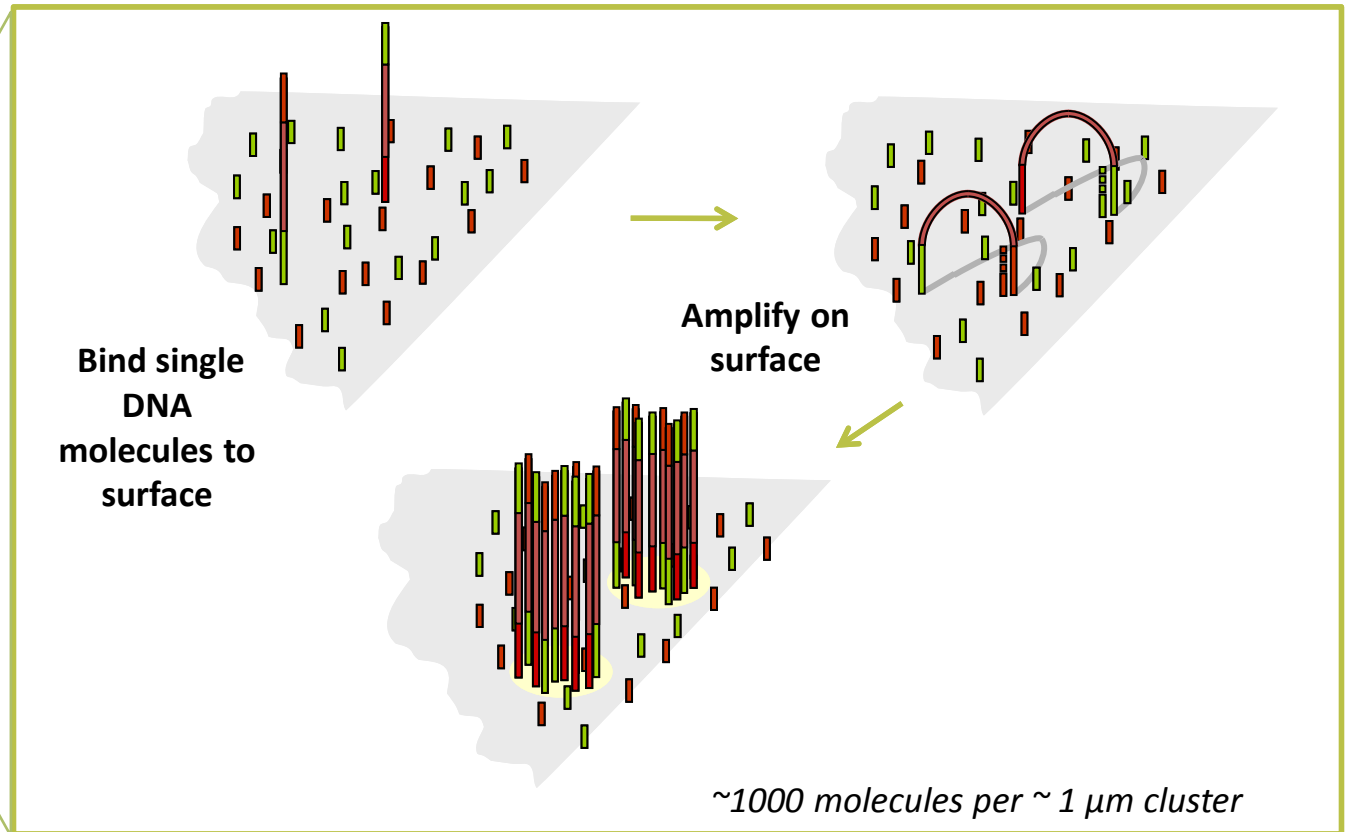
Sample Prep



Dual Index Library shown

The aim of the sample prep step is to obtain nucleic acid fragments with adapters attached on both ends

Cluster Generation



[Skip Overview](#)

Hybridize Fragment & Extend

Single DNA libraries are hybridized to primer lawn

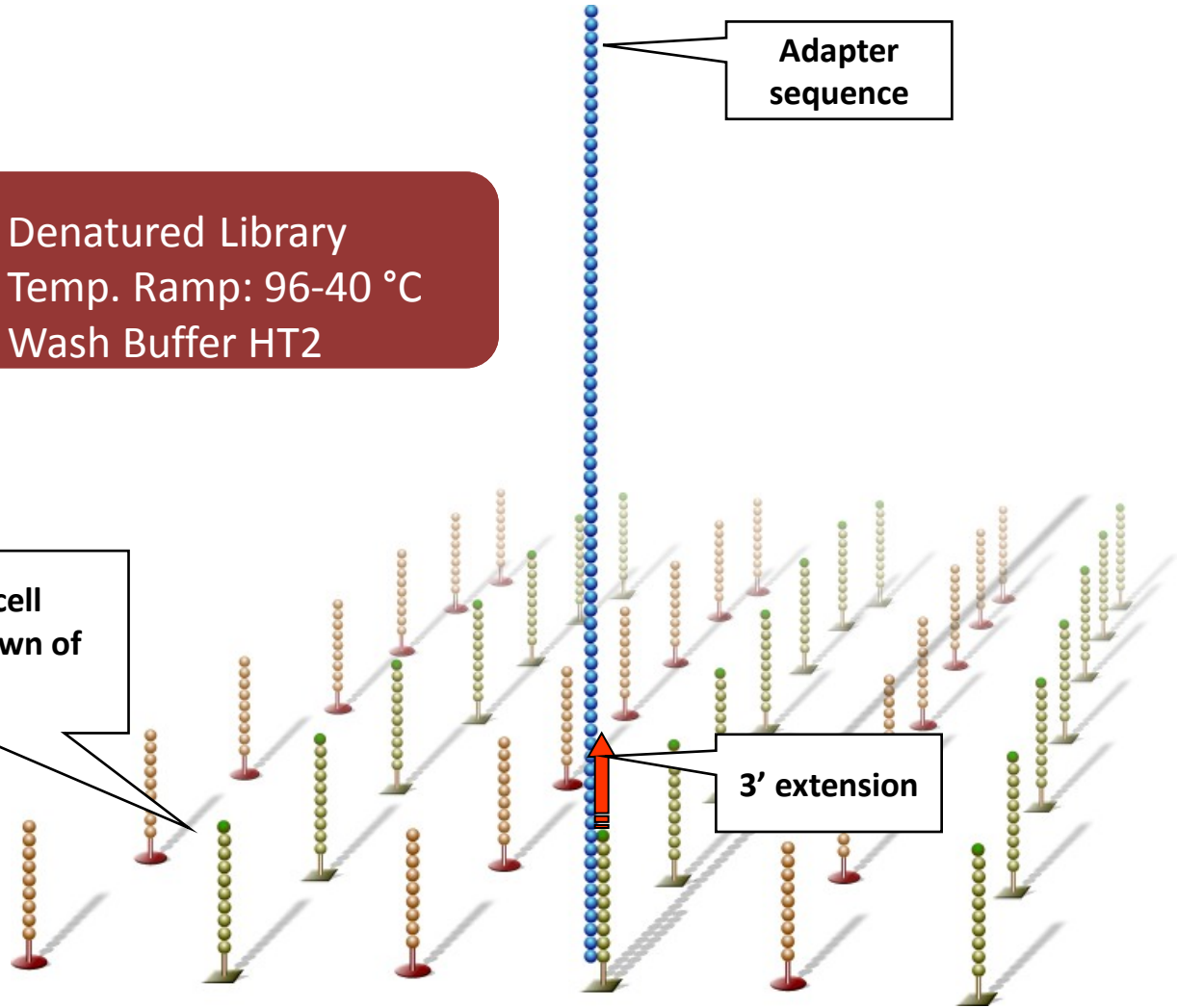
Bound libraries then extended by polymerase

Denatured Library
Temp. Ramp: 96-40 °C
Wash Buffer HT2

Surface of flow cell coated with a lawn of oligo pairs

Adapter sequence

3' extension



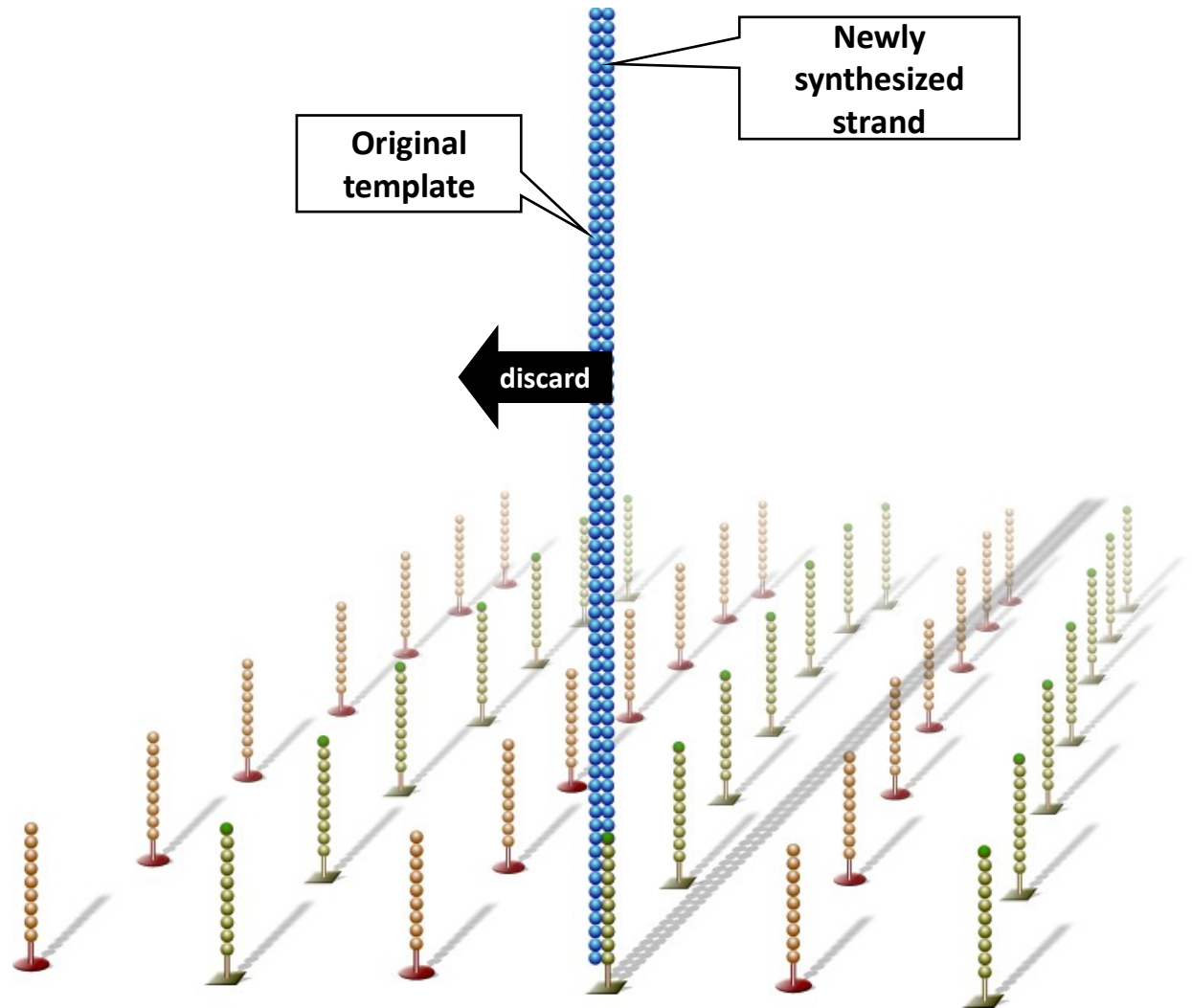
Denature Double-Stranded DNA

Double-stranded molecule is denatured

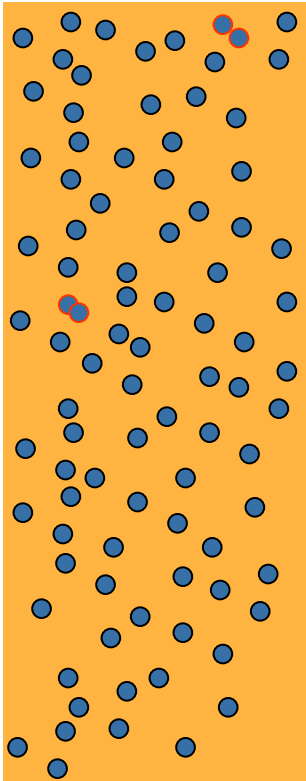
Original template washed away

Newly synthesized strand is covalently attached to flow cell surface

AMP Premix AMP1
Phusion HFE 90 sec
Temp. Ramp: 20°C

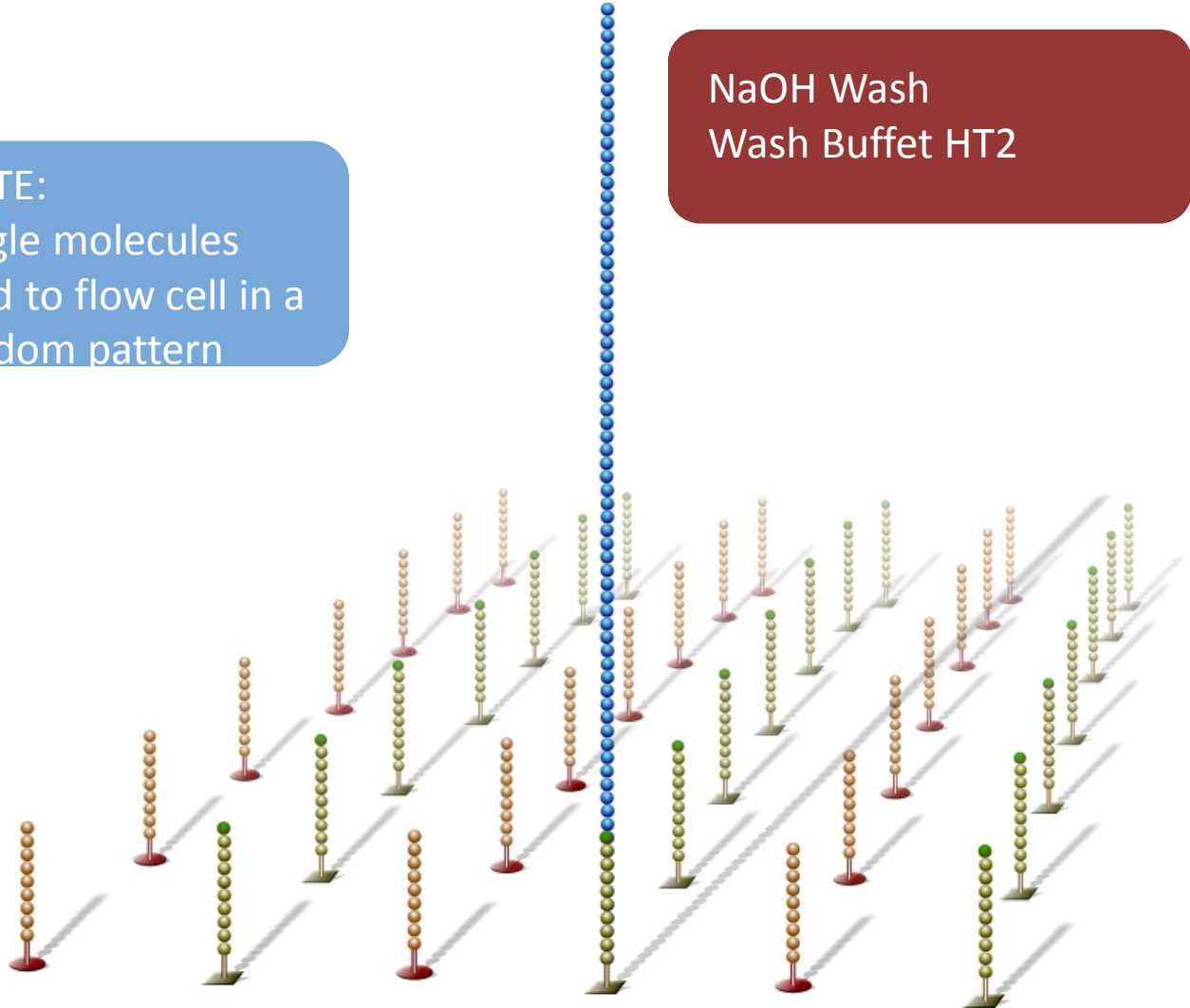


Hybridize Fragment & Extend



NOTE:
Single molecules
bind to flow cell in a
random pattern

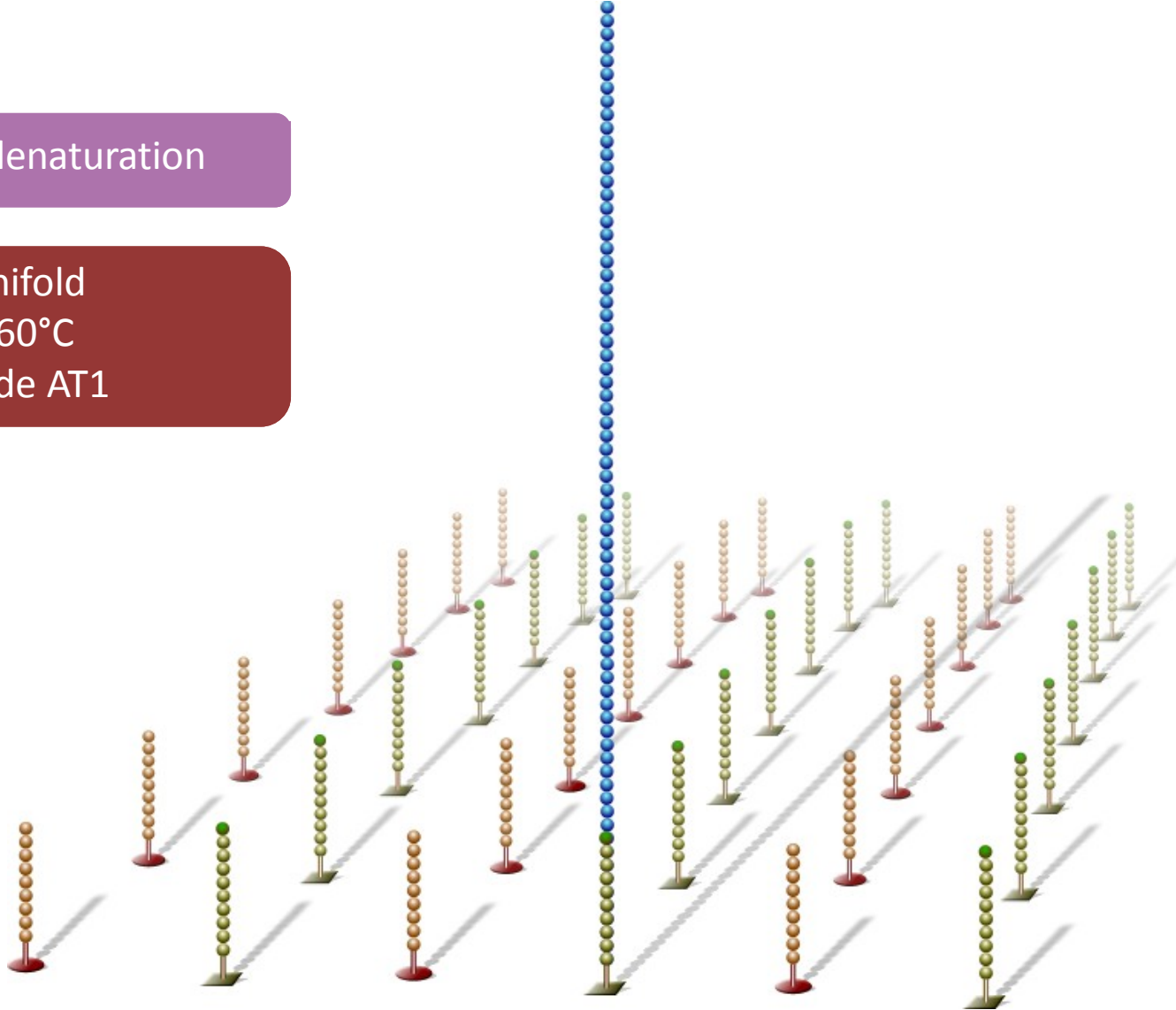
NaOH Wash
Wash Buffer HT2



Hybridize Fragment & Extend

1st cycle denaturation

AMP Manifold
Ramp to 60°C
Formamide AT1

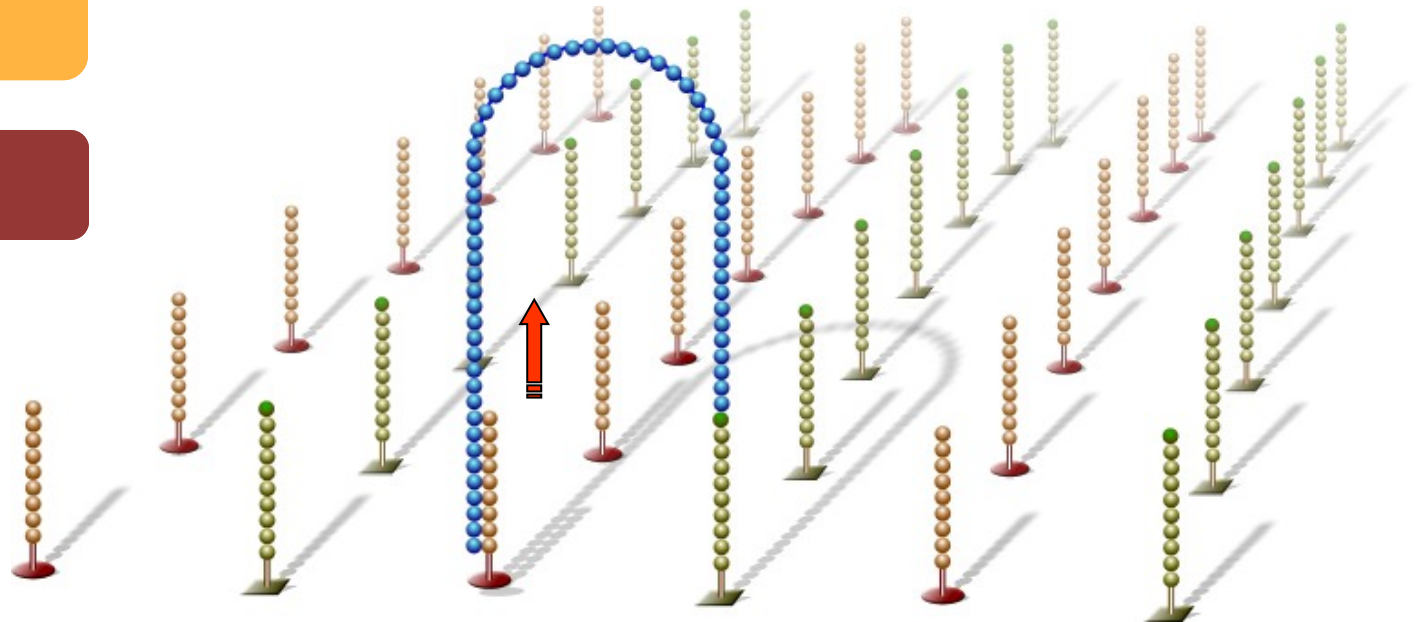


Bridge Amplification

Single-stranded molecule flips over and forms a bridge by hybridizing to adjacent, complementary primer

Hybridized primer is extended by polymerase

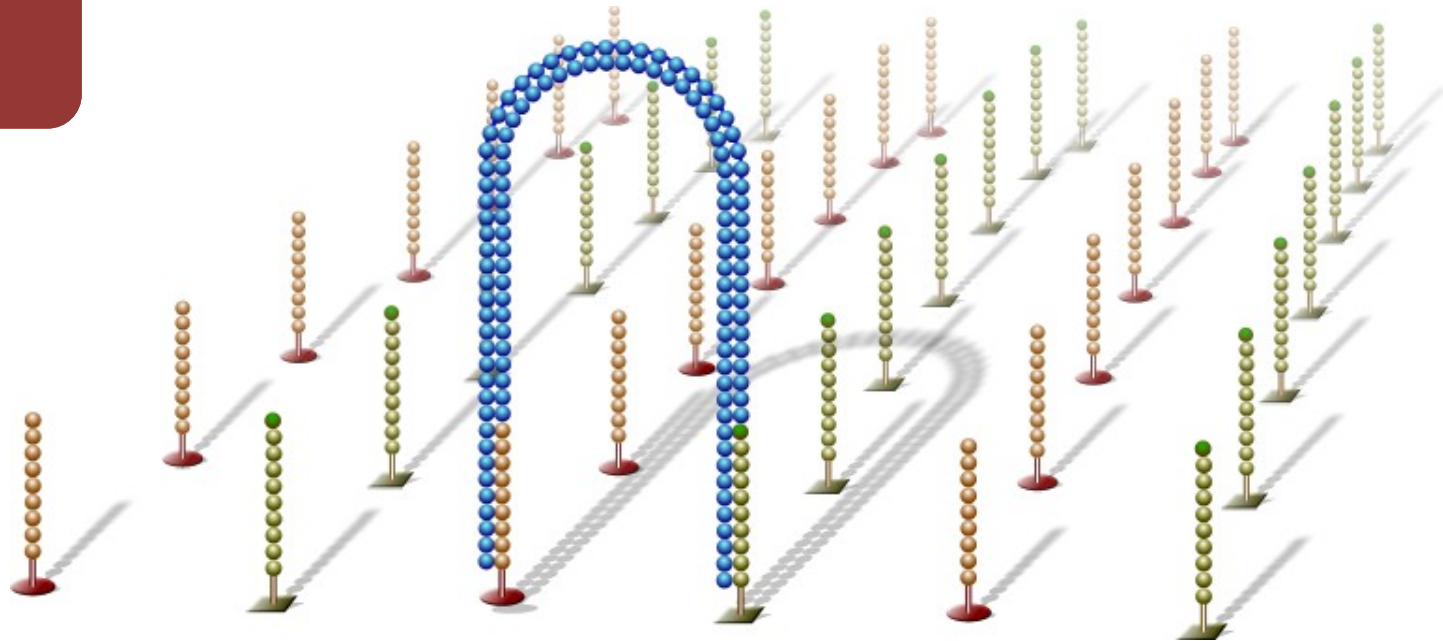
AMP Premix AMP1



Bridge Amplification

Double-stranded bridge is formed

AMP Mix AMX1
Contains BST pol &
nucleotides

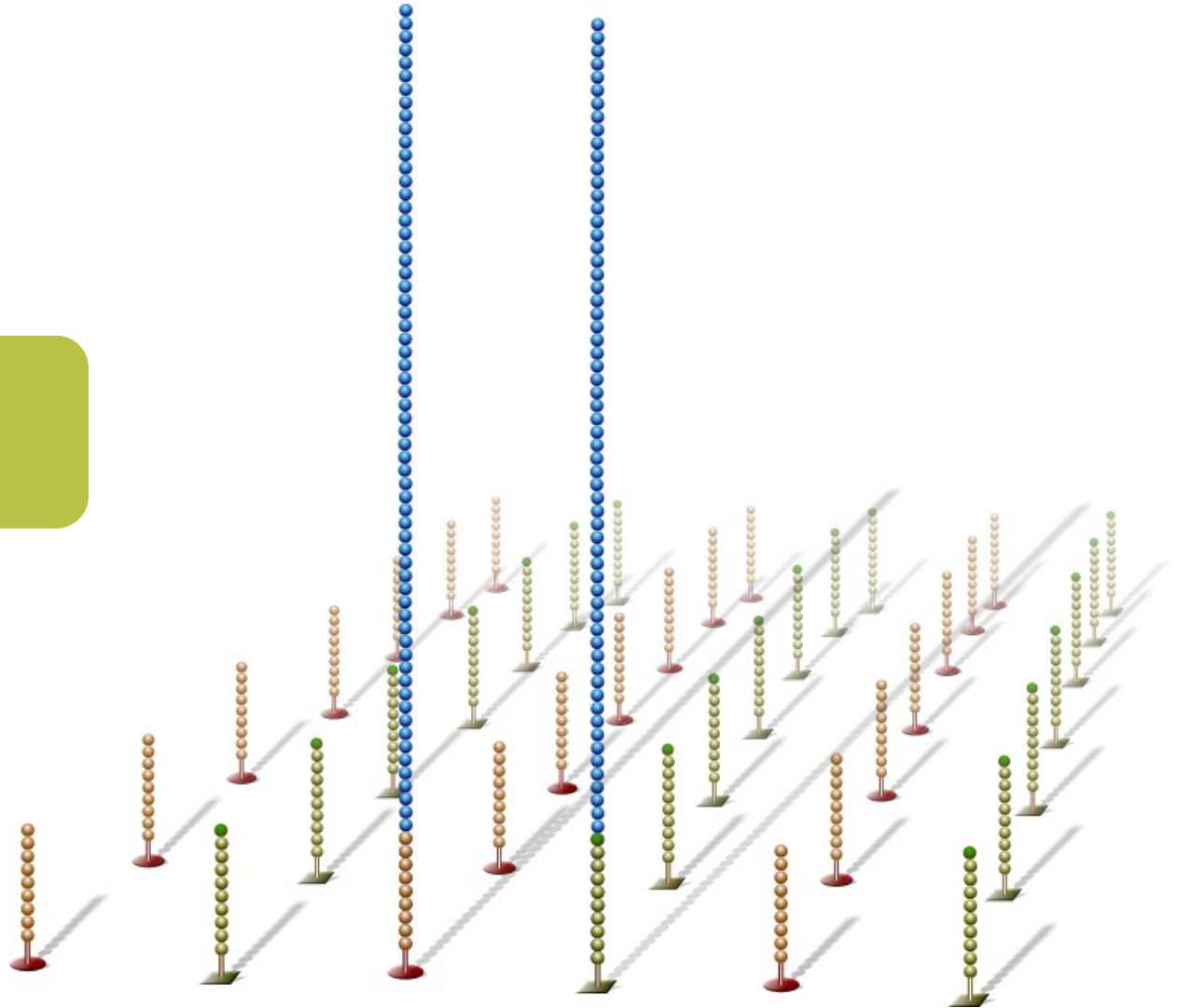


Denature Double-Stranded Bridge

Double-stranded bridge is denatured - 1st cycle denaturation

Result:
Two copies of covalently bound single-stranded templates

Formamide AT1

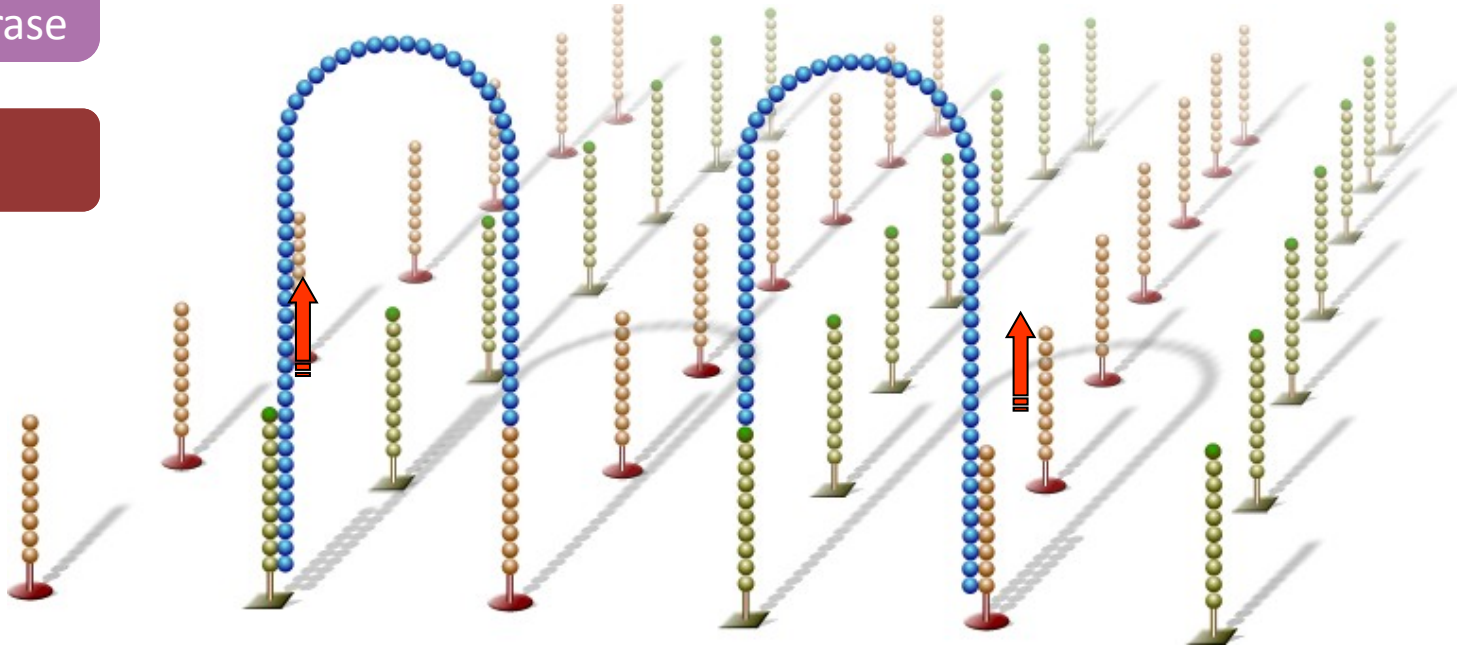


Bridge Amplification

Single-stranded molecules flip over to hybridize to adjacent primers

Hybridized primer is extended by polymerase

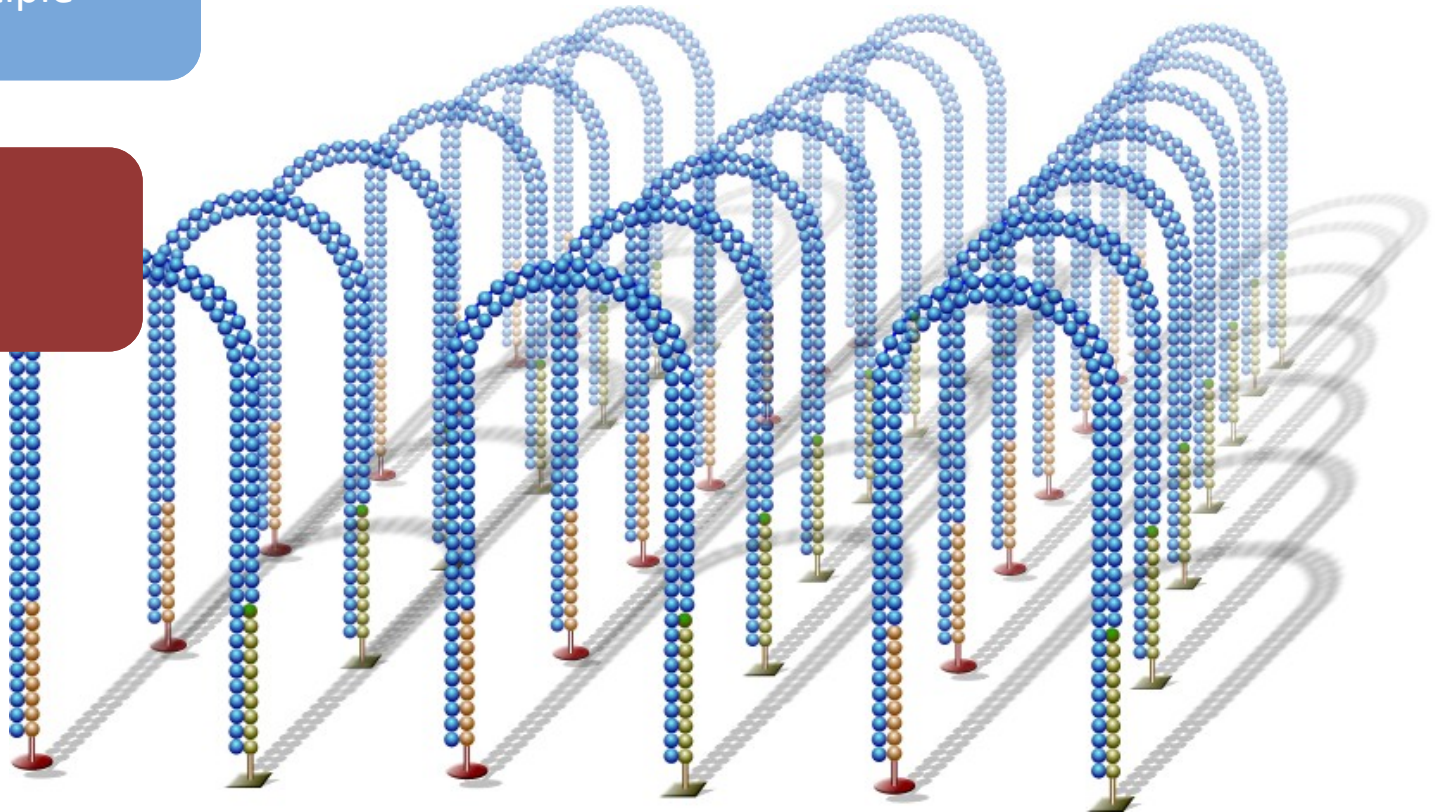
AMP Premix APM1



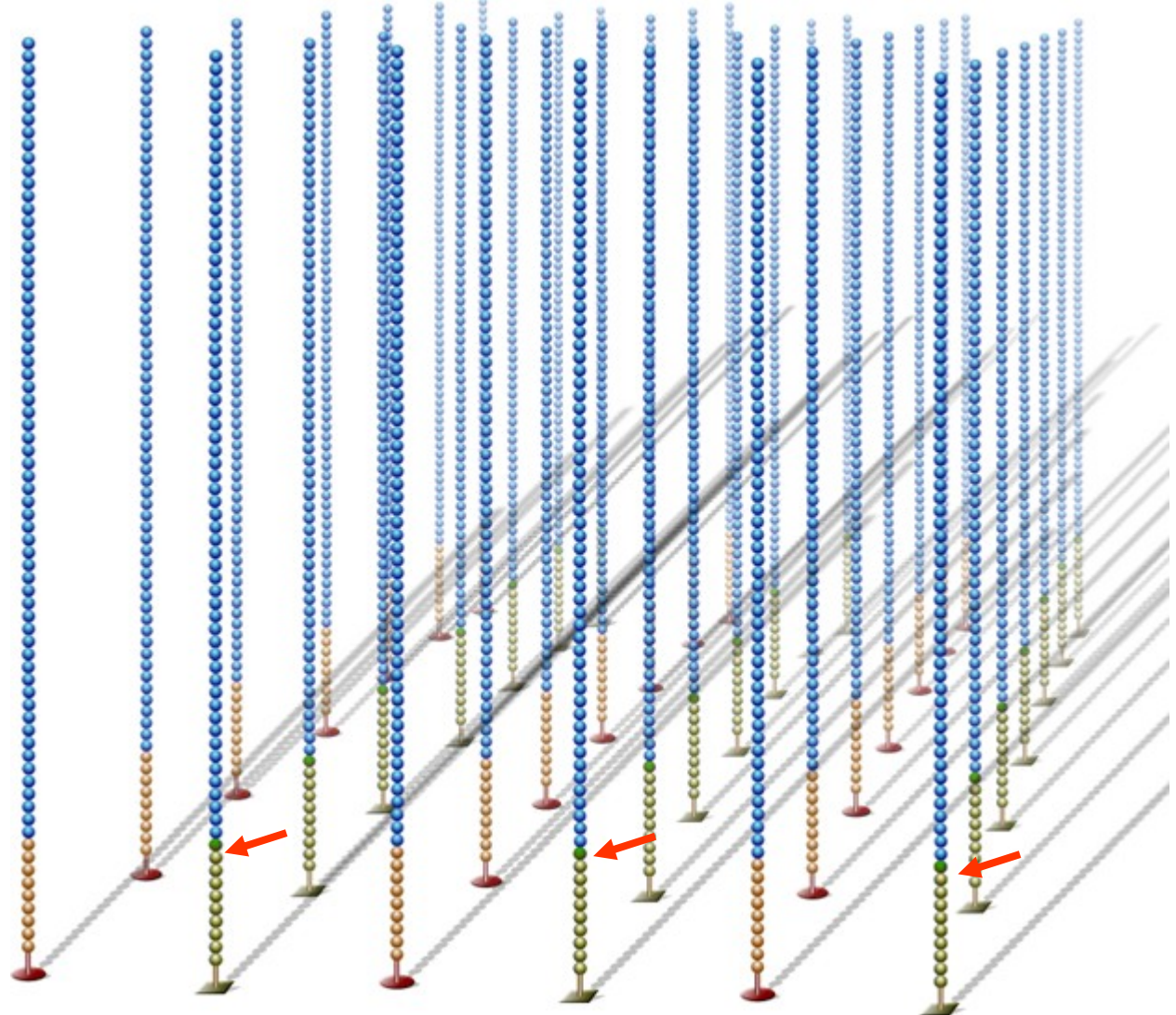
Bridge Amplification

Bridge amplification cycle repeated until multiple bridges are formed

AMP Mix AMX1
Contains BST pol &
nucleotides



Linearization



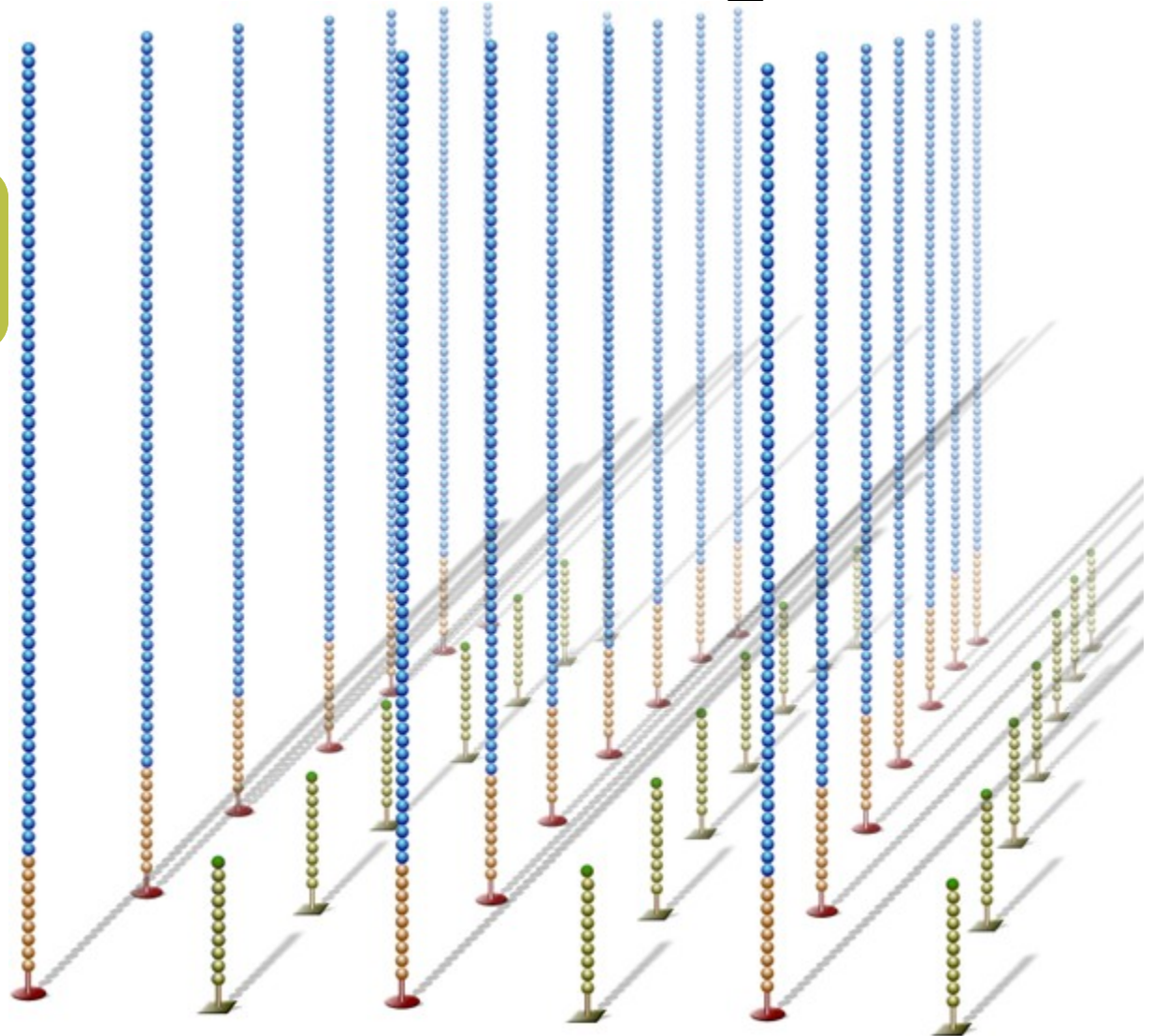
dsDNA bridges are denatured

PE Linearization LMX1
Ramp 37.9 °C, 30 min
Temp Ramp: 20 °C

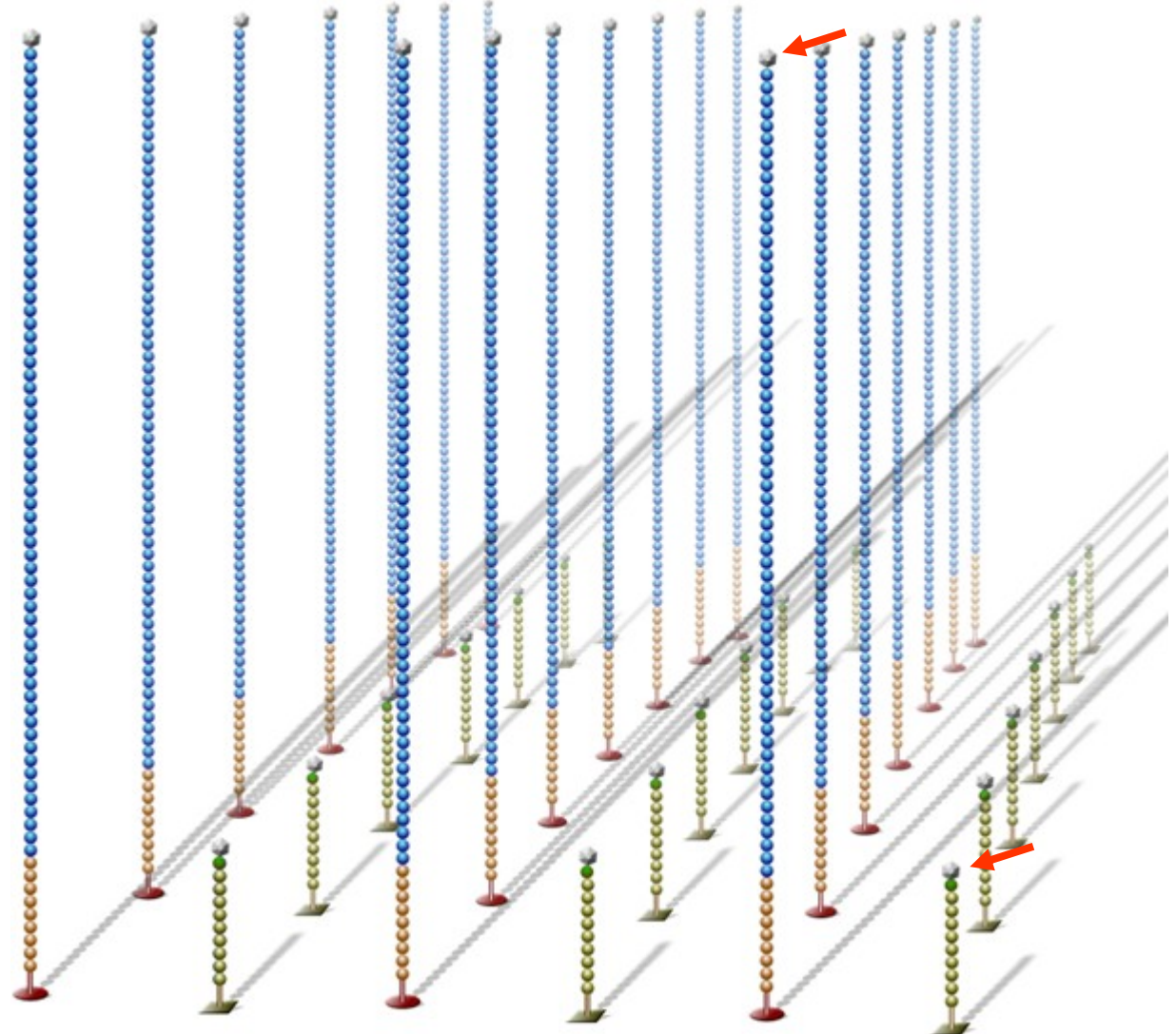
Reverse Strand Cleavage

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

Wash Buffer HT2



Blocking



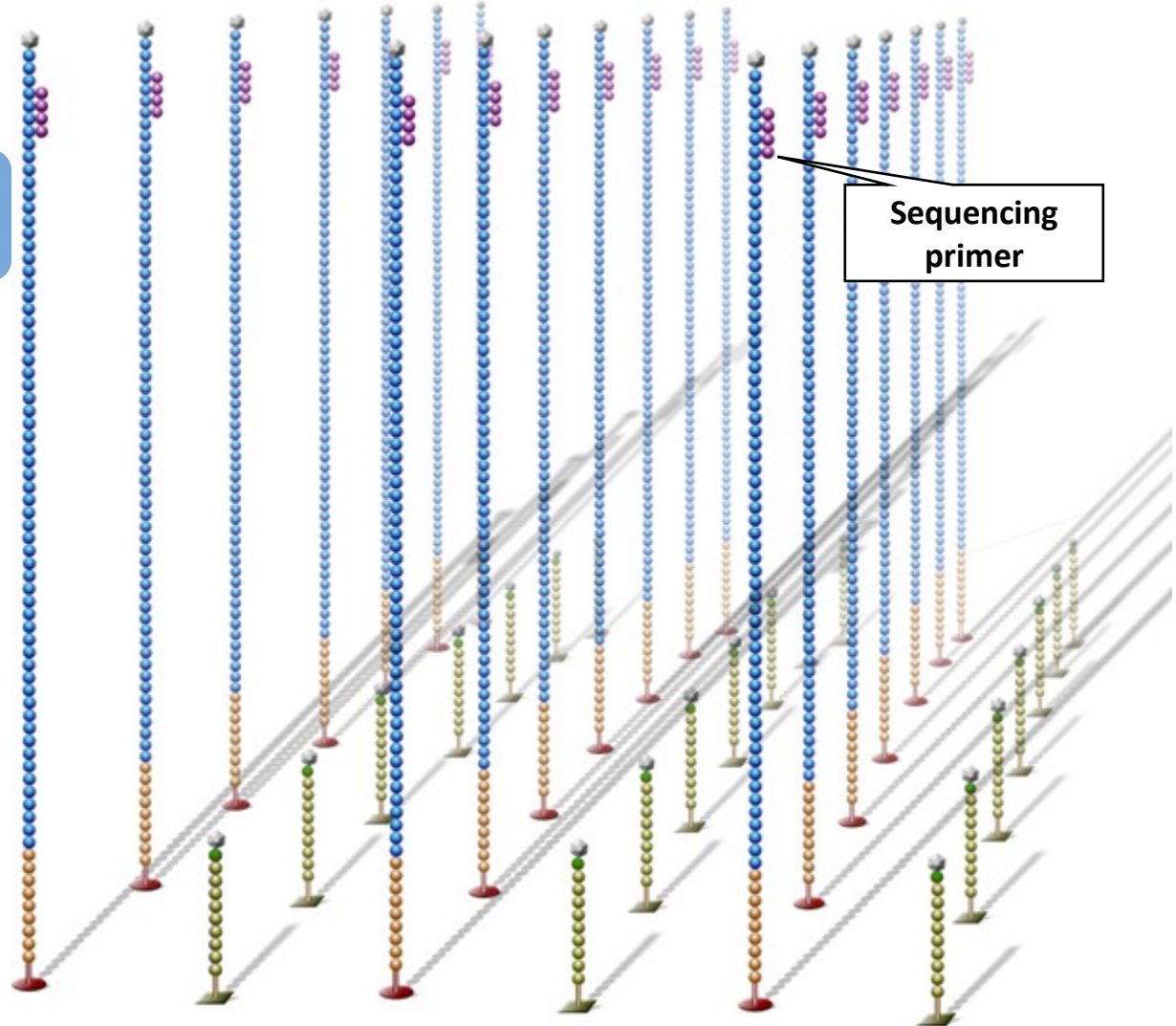
Free 3' ends are blocked to prevent unwanted DNA priming

Blocking Mix BMX
38 °C, 30 min
60 °C, 15 min
20°C, HT2, HT1 Washes

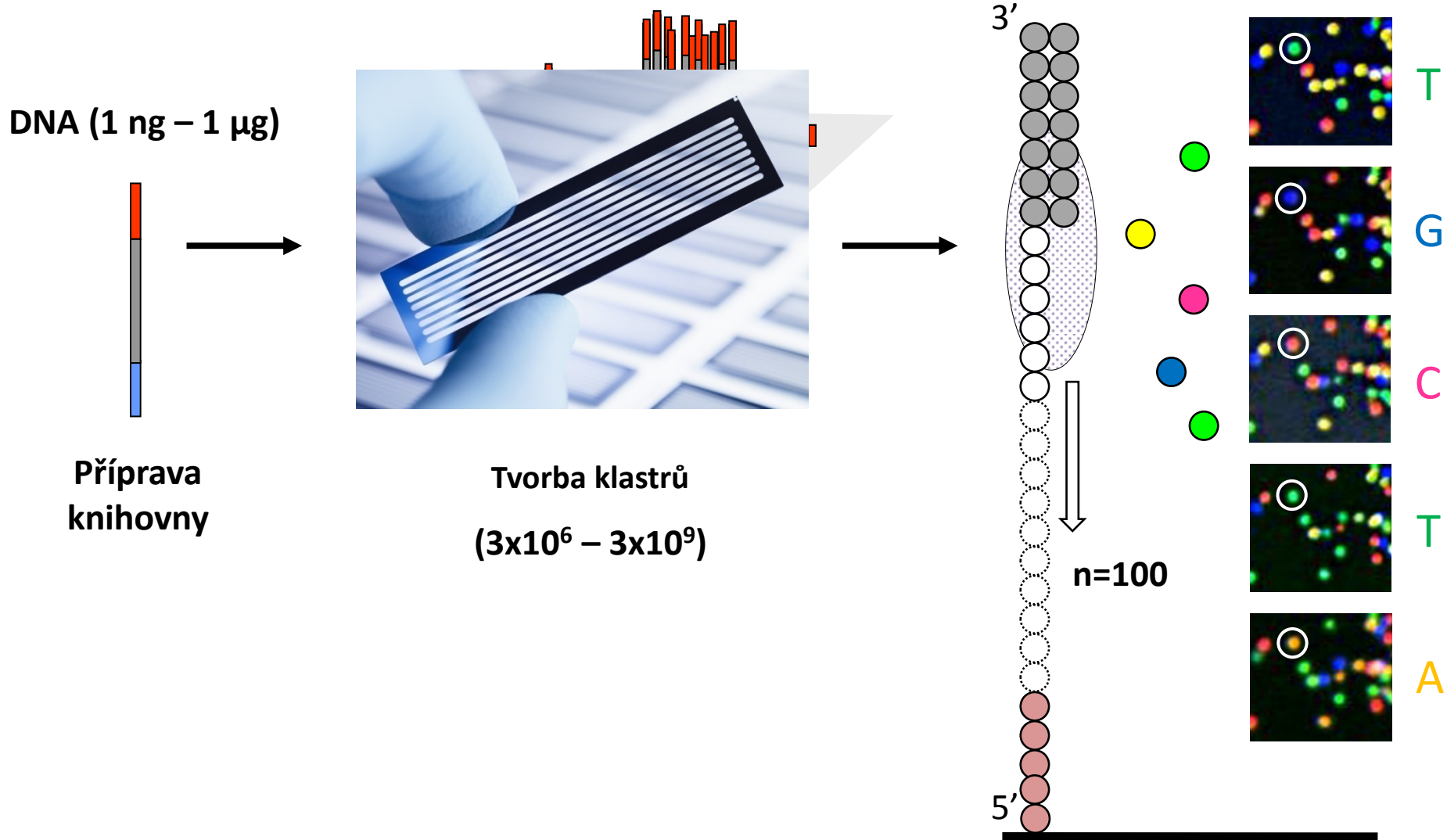
Read 1 Primer Hybridization

Sequencing primer is hybridized to adapter sequence

0.1 NAOH
Seq. Primer
60 °C, 5 min
20 °C, HT2, HT1 Washes



Sequencing by synthesis (SBS)



Sequencing

Clusters are images using LED and filter combinations specific for each fluorescently-labeled nucleotide

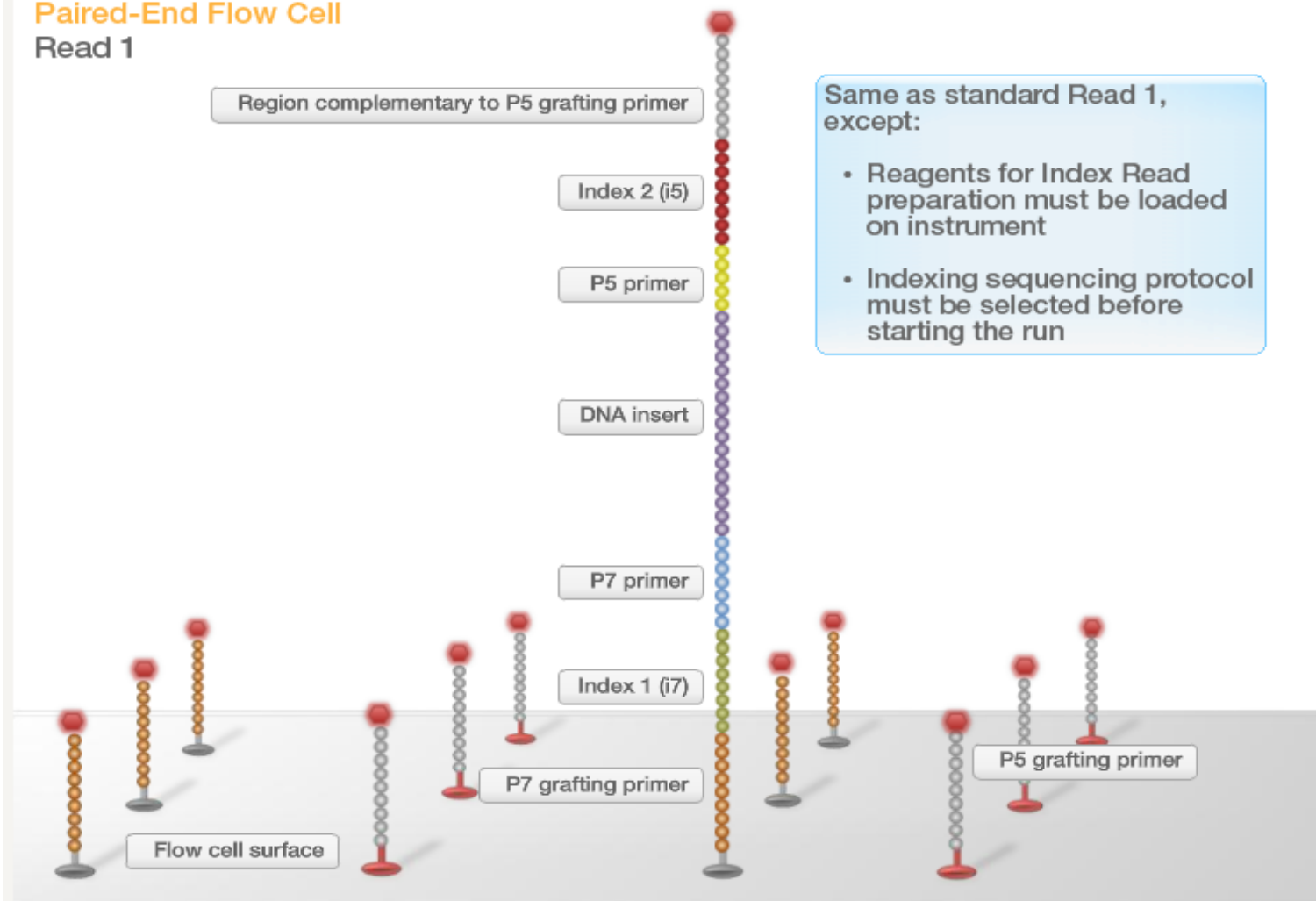
After imaging is complete for one section (tile), the flow cell is moved to the next tile and the process is repeated



Pair - End Sequencing – Dual Indexed

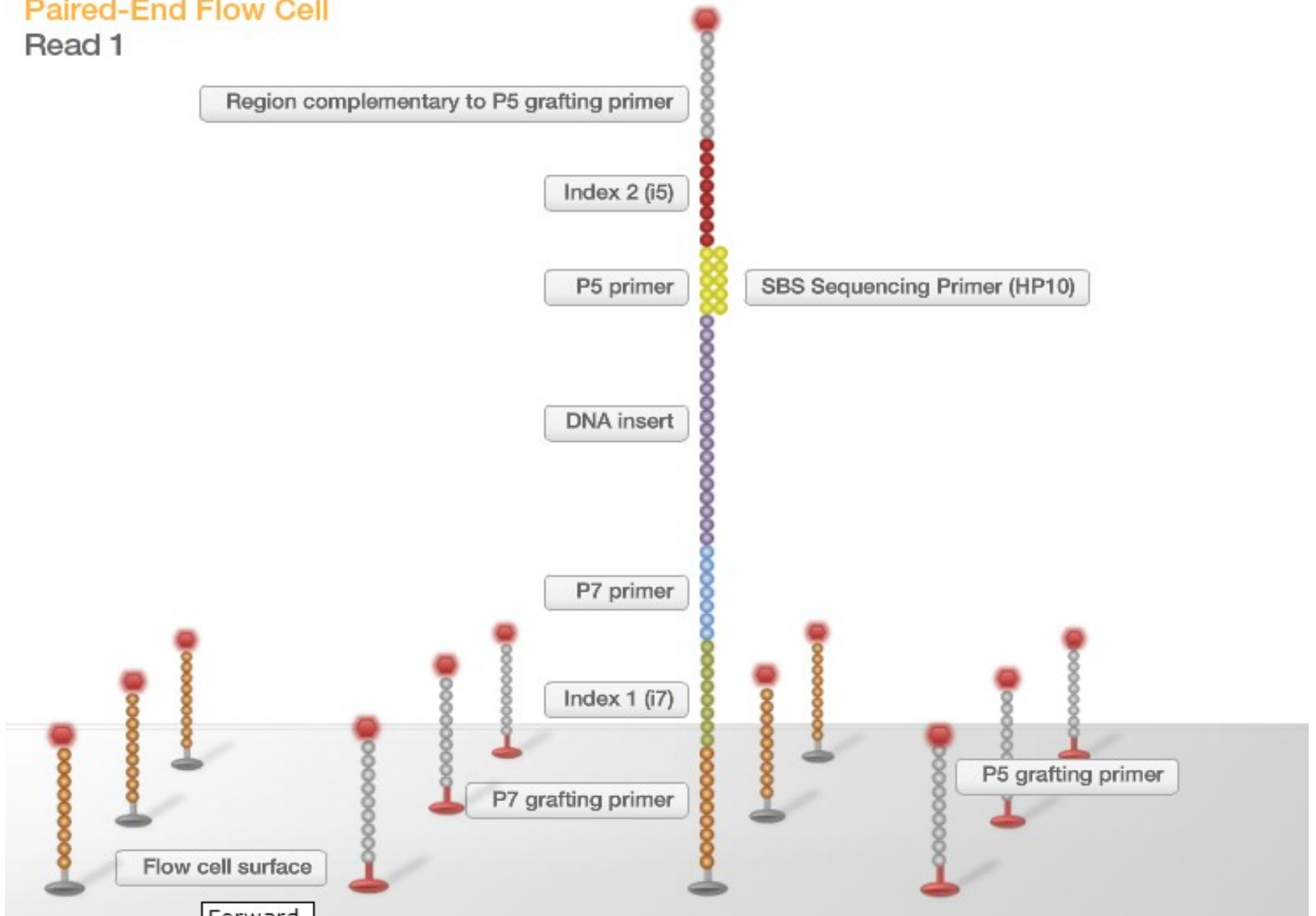
Paired-End Flow Cell

Read 1



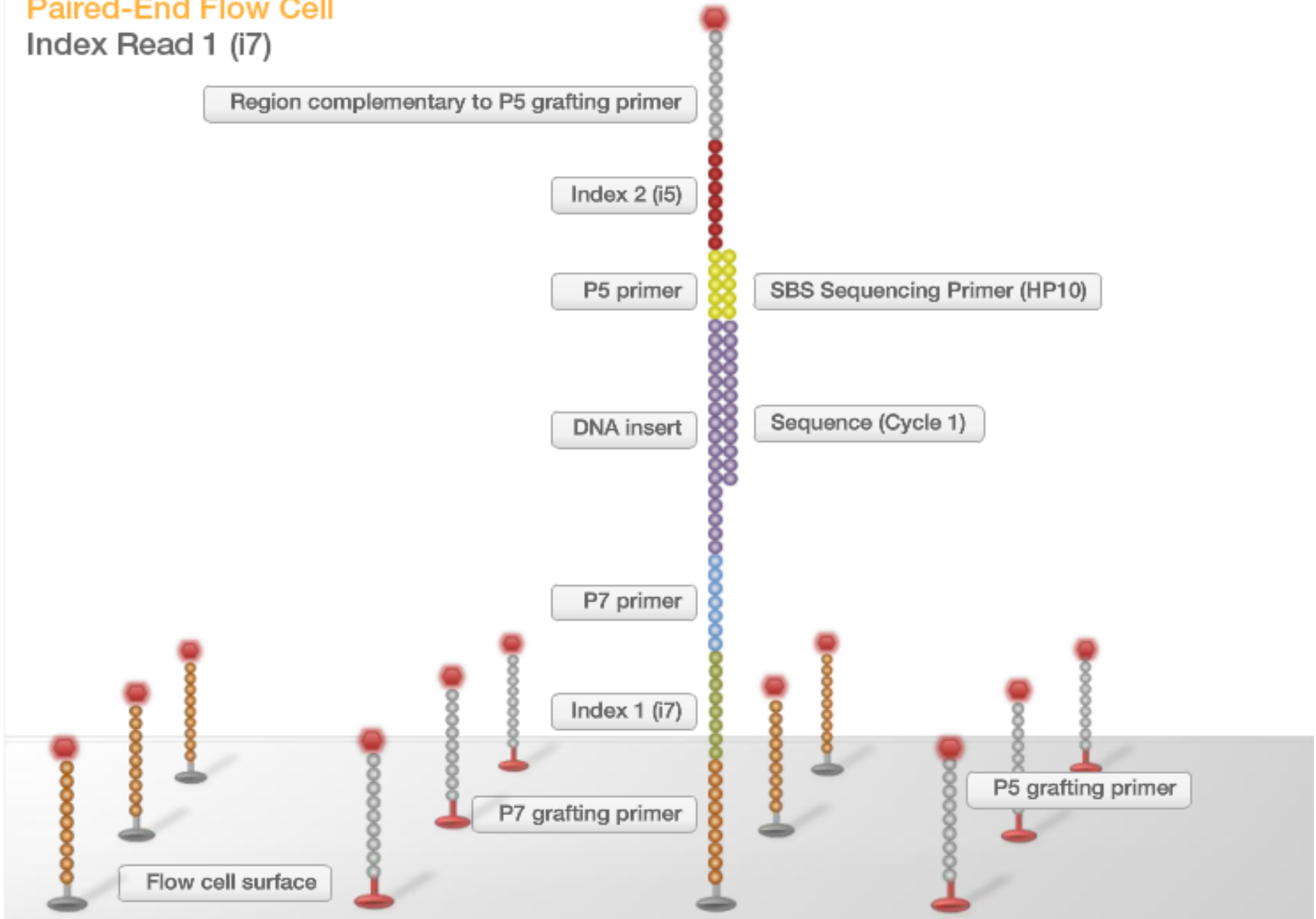
Paired-End Flow Cell

Read 1



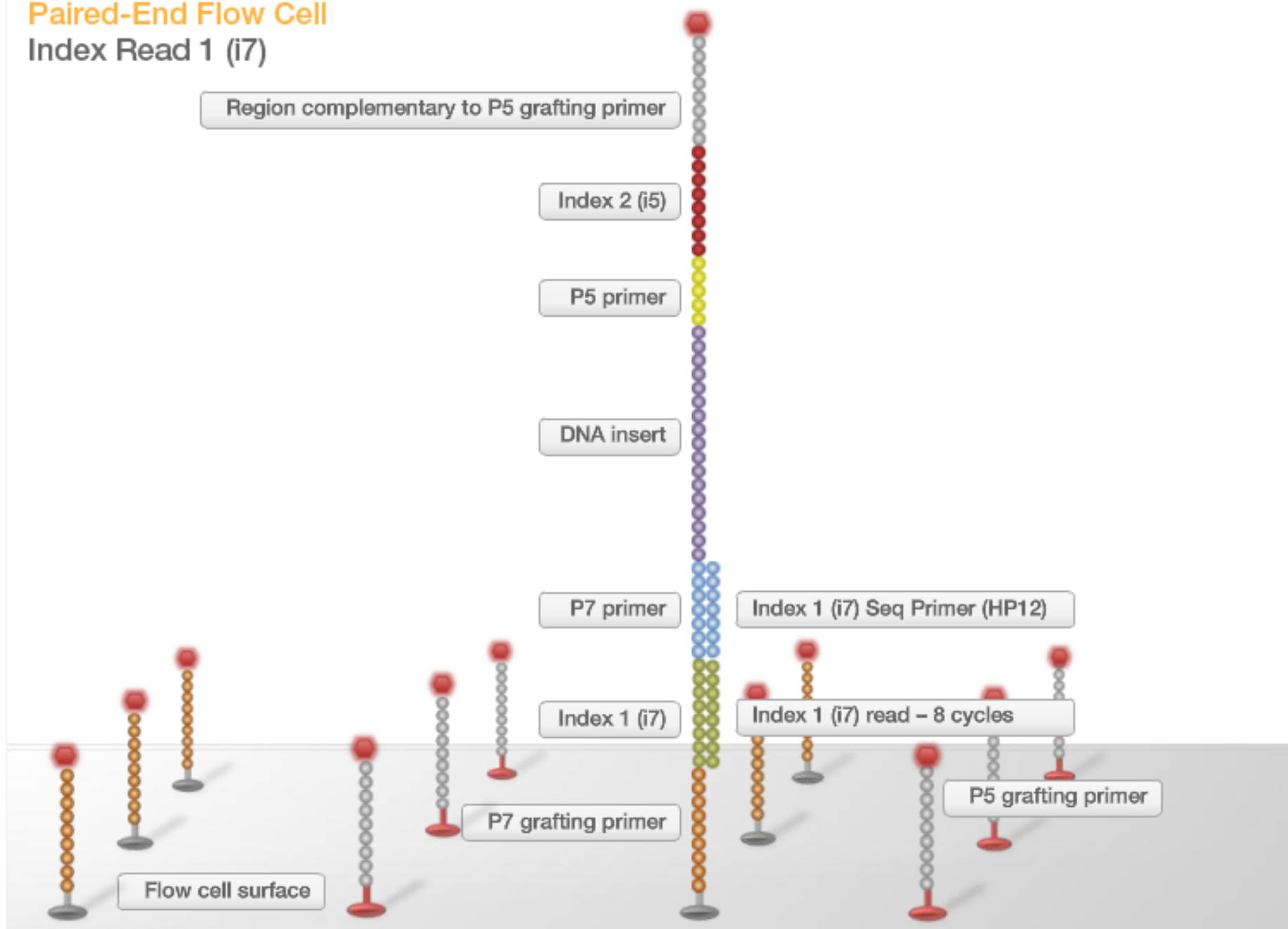
Paired-End Flow Cell

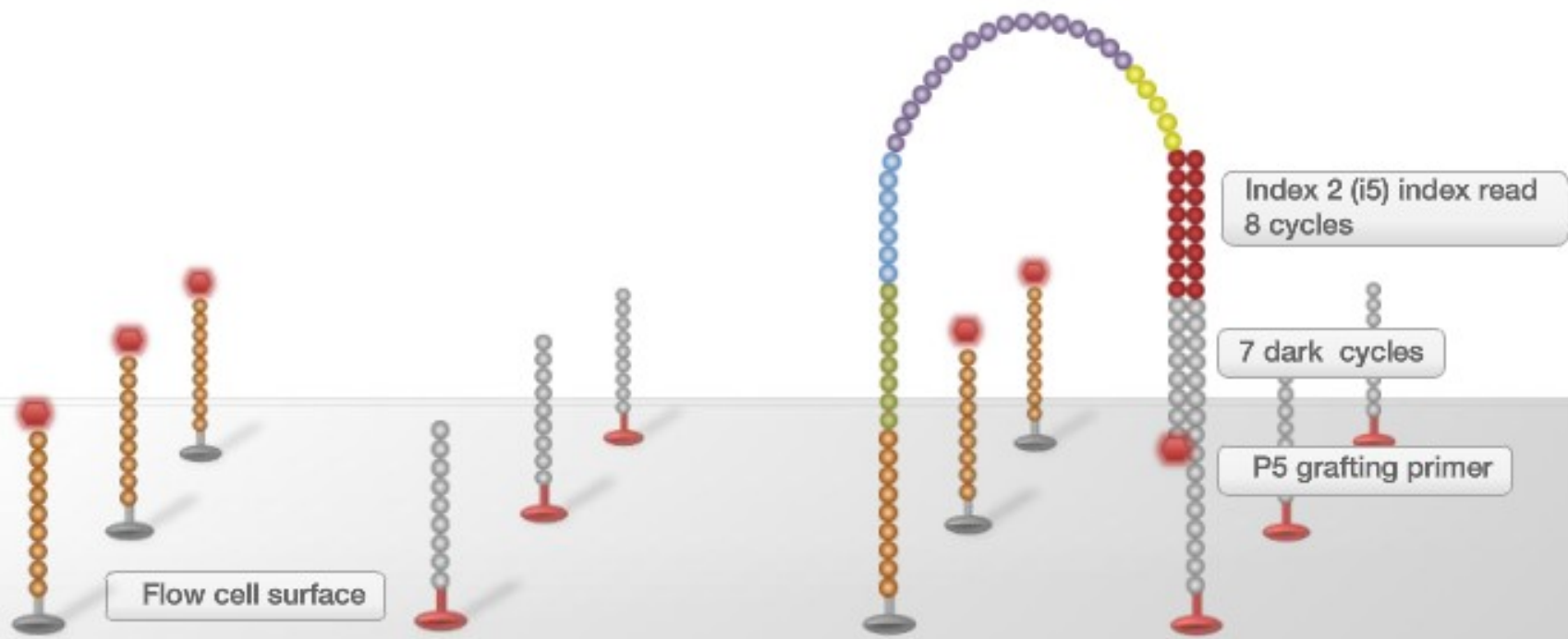
Index Read 1 (i7)

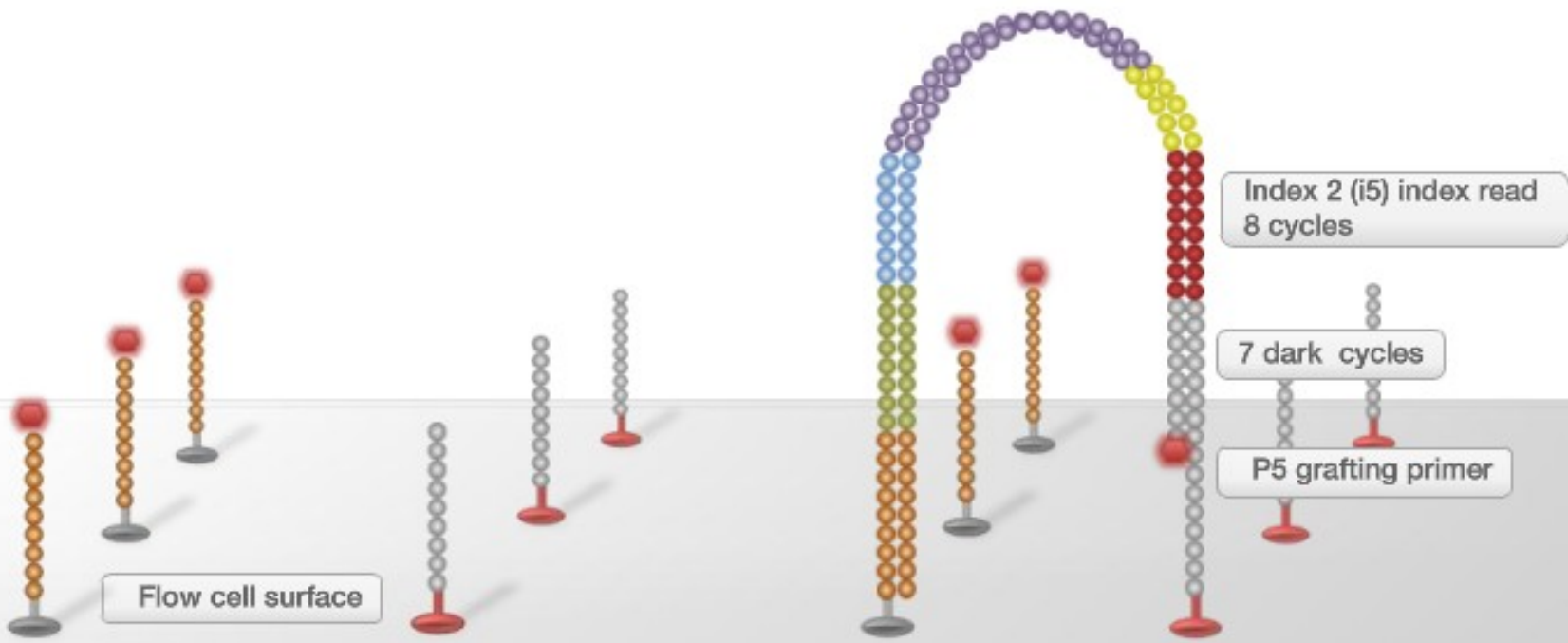


Paired-End Flow Cell

Index Read 1 (i7)







Flow cell surface

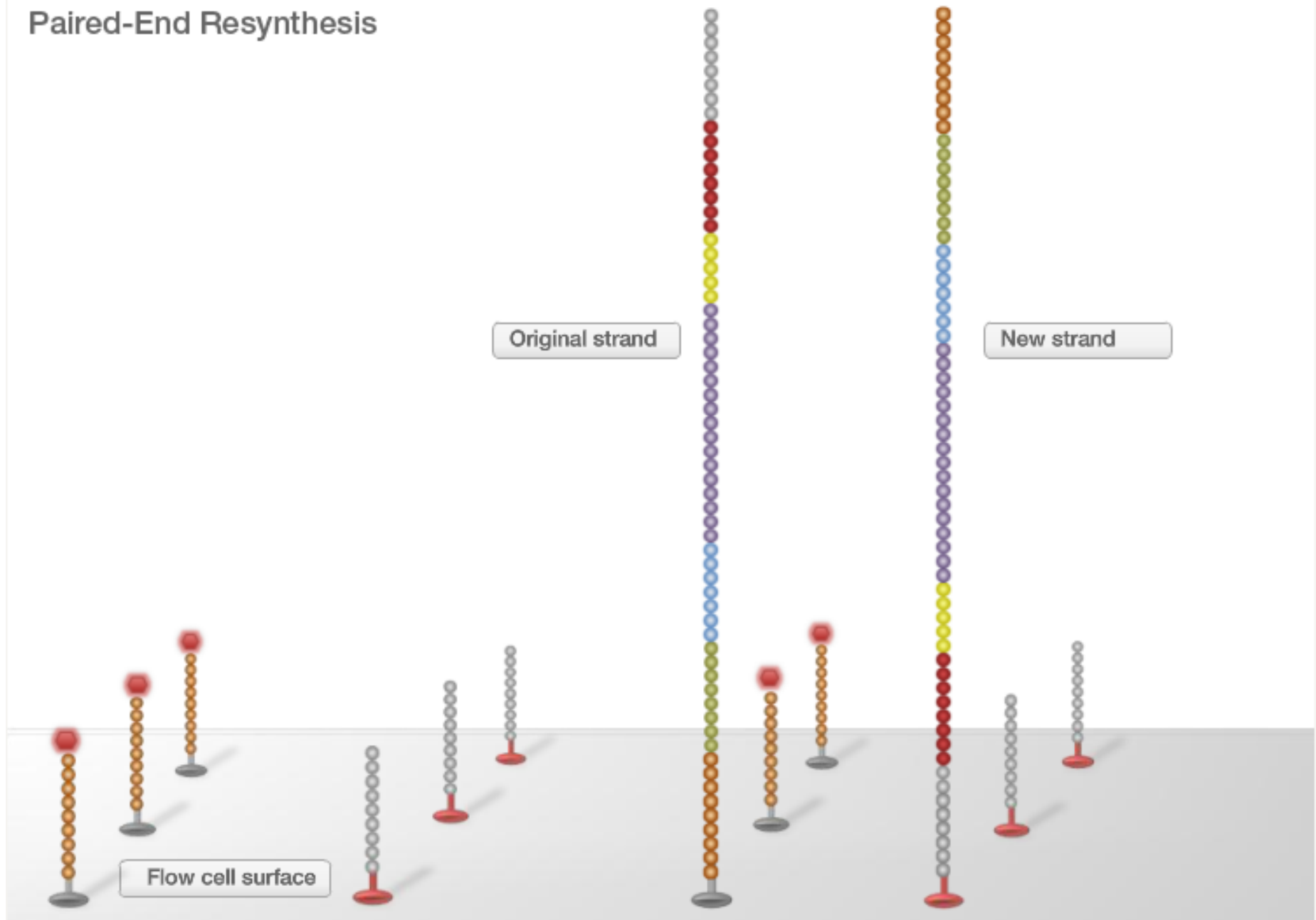
Index 2 (i5) index read
8 cycles

7 dark cycles

P5 grafting primer

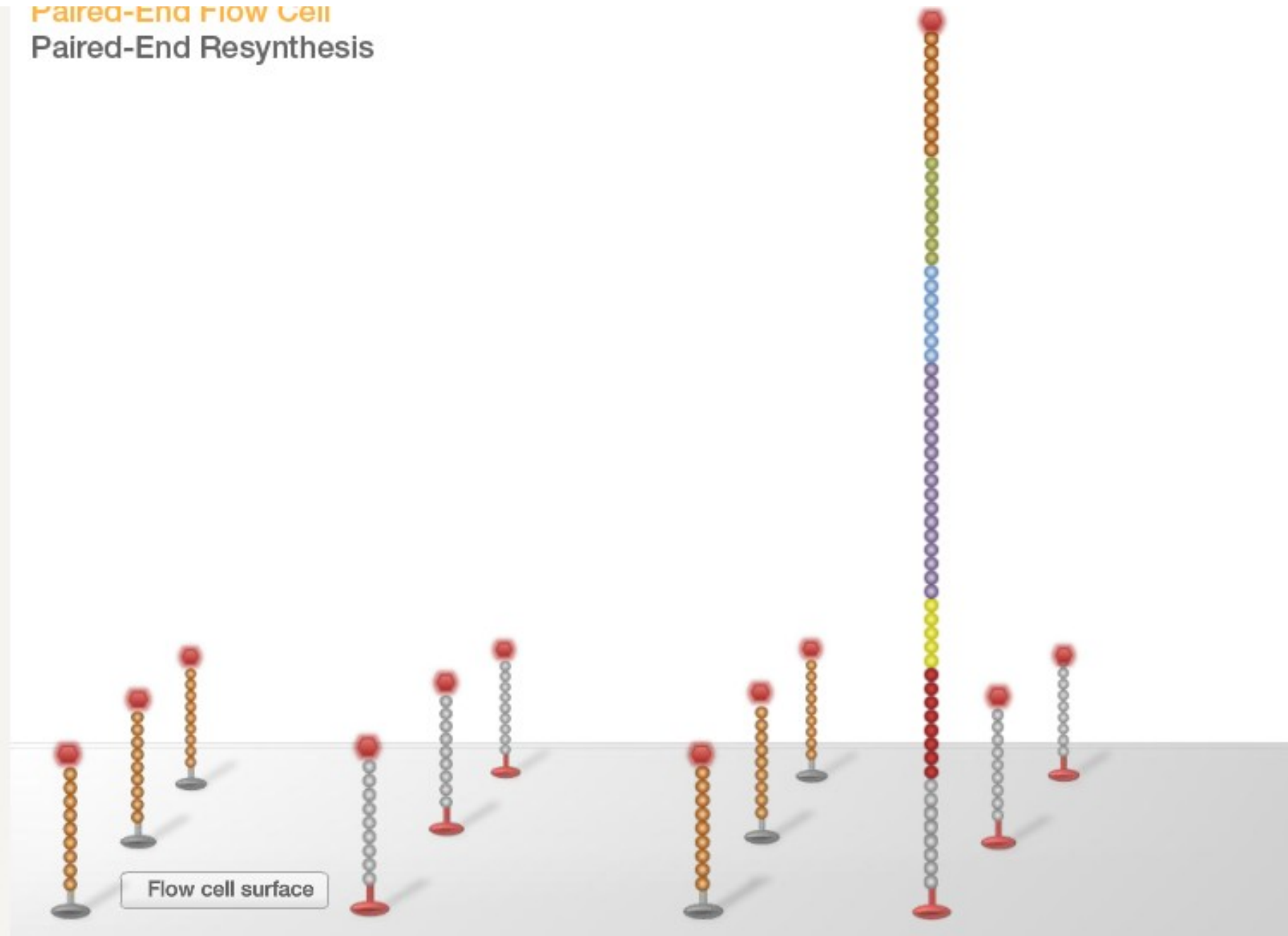
Paired-End Flow Cell

Paired-End Resynthesis



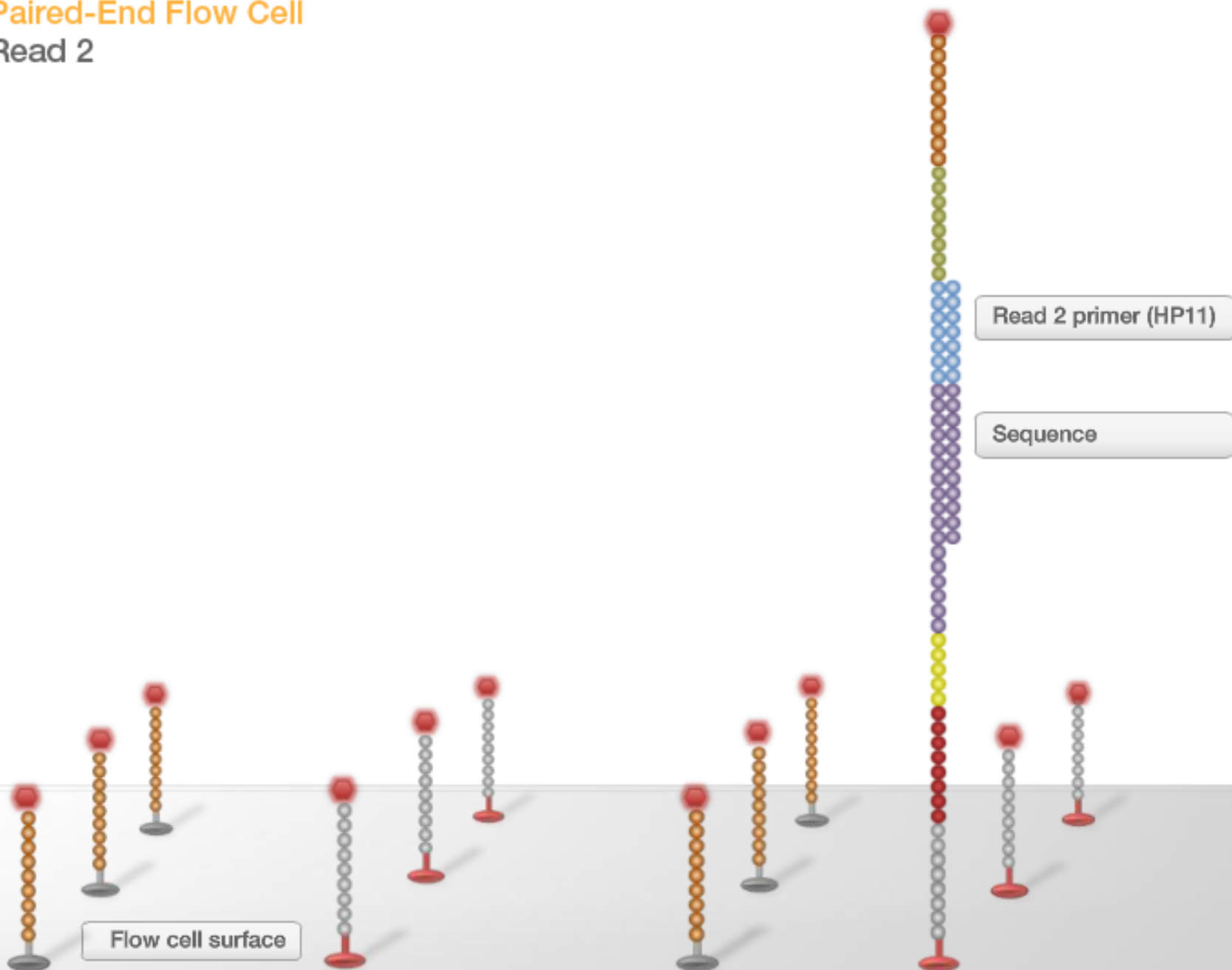
Paired-End Flow Cell

Paired-End Resynthesis

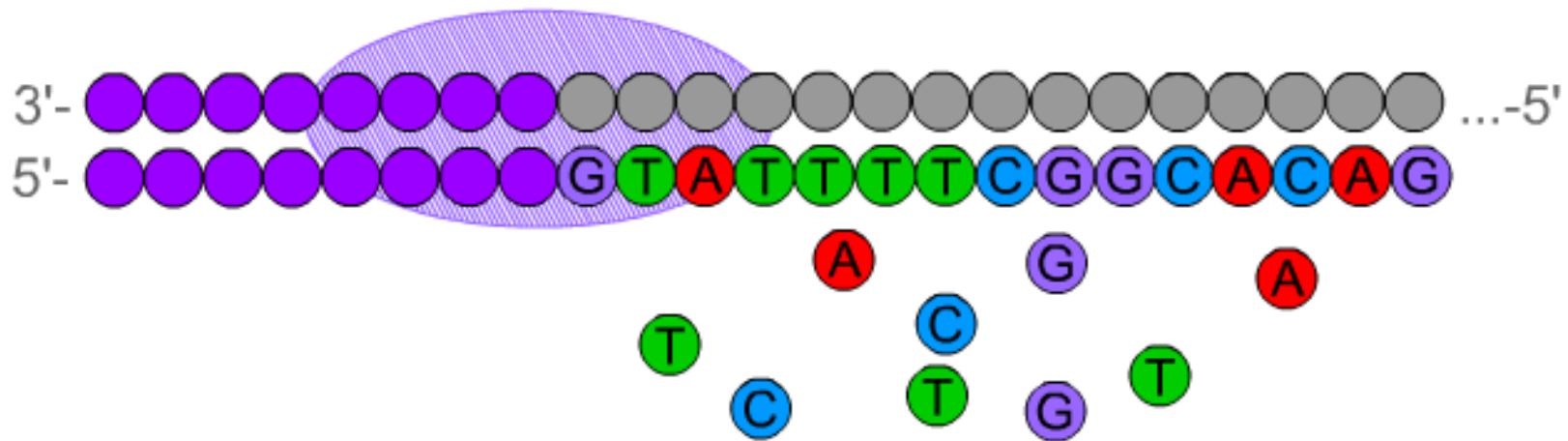


Paired-End Flow Cell

Read 2



Sekvenační technologie



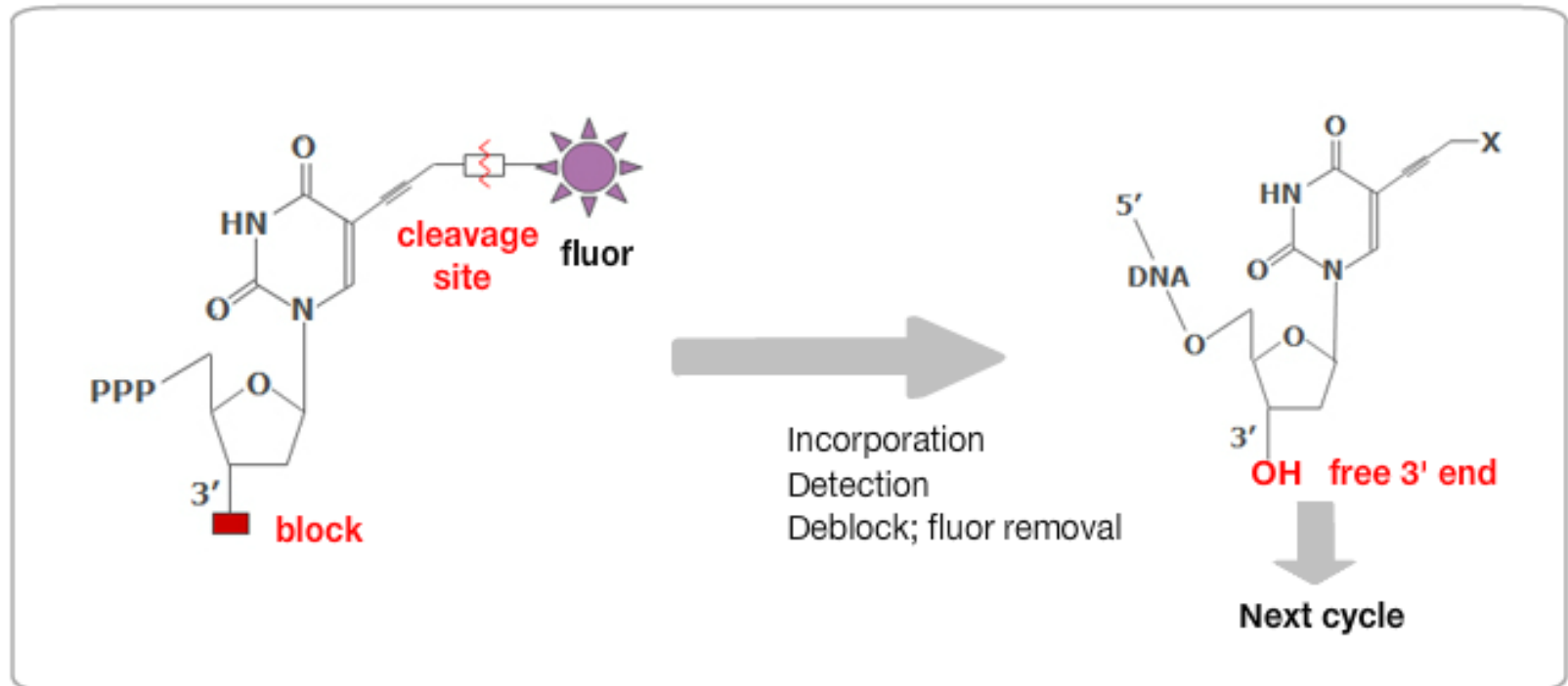
Cycle 1: Add sequencing reagents
 First base incorporated
 Remove unincorporated bases
 Detect signal, deblock and defluor

Cycle 2-n: Add Sequencing reagents and repeat

Sekvenační technologie

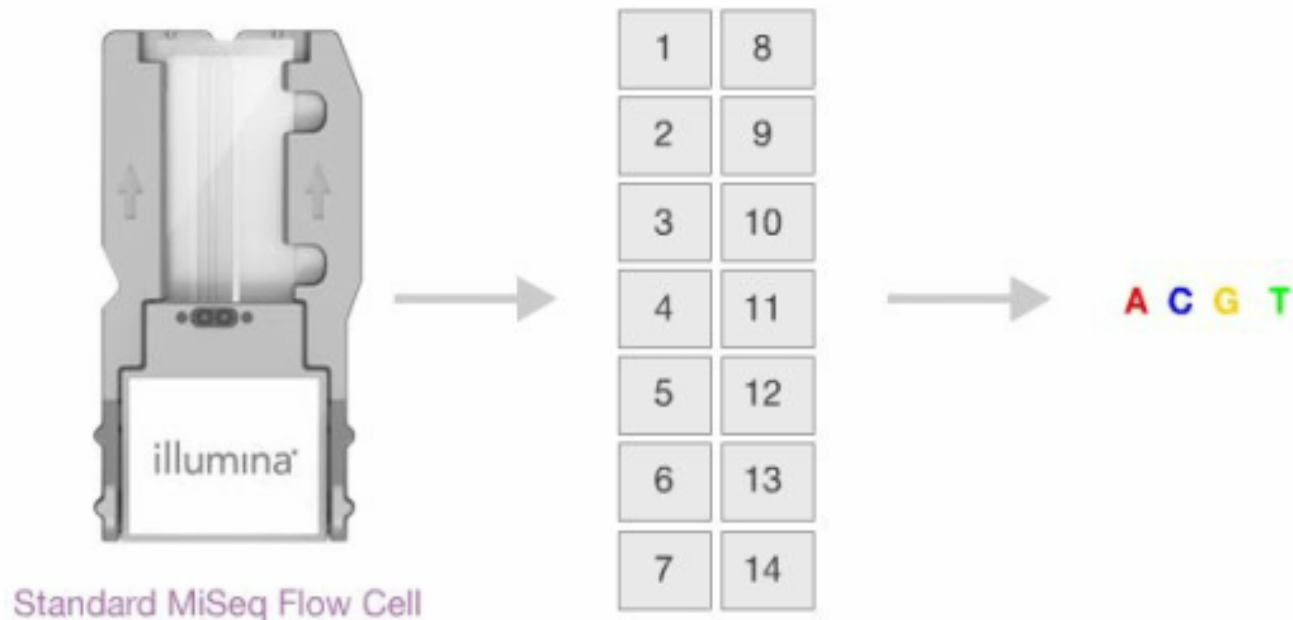
The MiSeq sequences the DNA clusters using Illumina's Sequencing By Synthesis (SBS) Chemistry which relies on Reversible Terminator Chemistry (RTC).

- All 4 labeled nucleotides in 1 reaction
- Higher accuracy



Images Generated on the Instrument

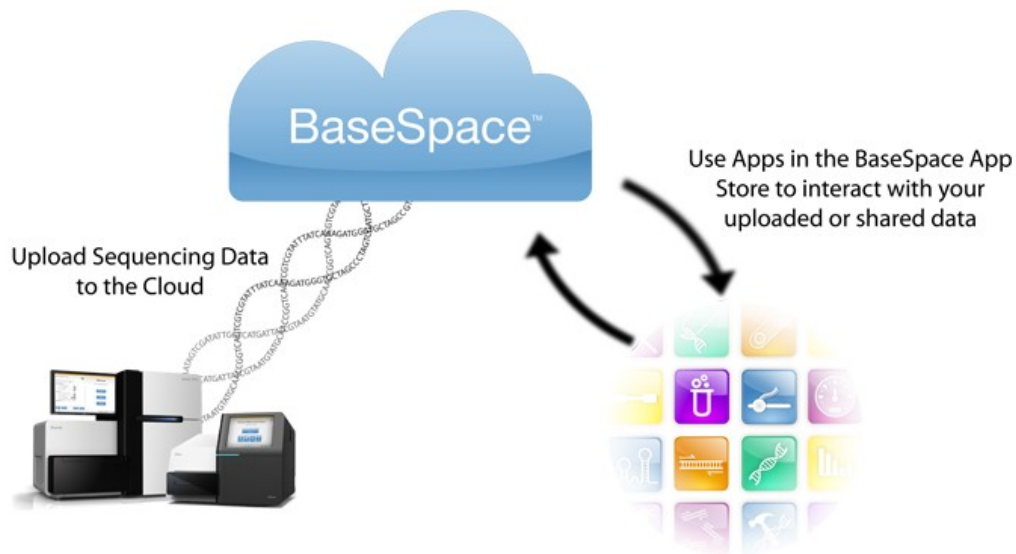
- MCS controls image generation on the MiSeq
- One cycle includes the chemical addition and imaging of one base for each cluster on the flow cell
- For imaging, the MiSeq flow cell is broken up into imaging areas or tiles
 - The number of tiles imaged depends on the flow cell type (standard, nano, or micro)
- For each tile, an image is taken for every base in every cycle
 - Four images (one each for G,A,T,C) per tile per cycle



BaseSpace

<https://accounts.illumina.com/>

- Is a powerful website computing platform
- for storing my genomics data on a cloud
- for analyzing my sequences
- for sharing my genetic data

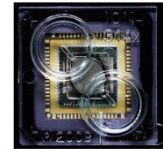


Ion Torrent

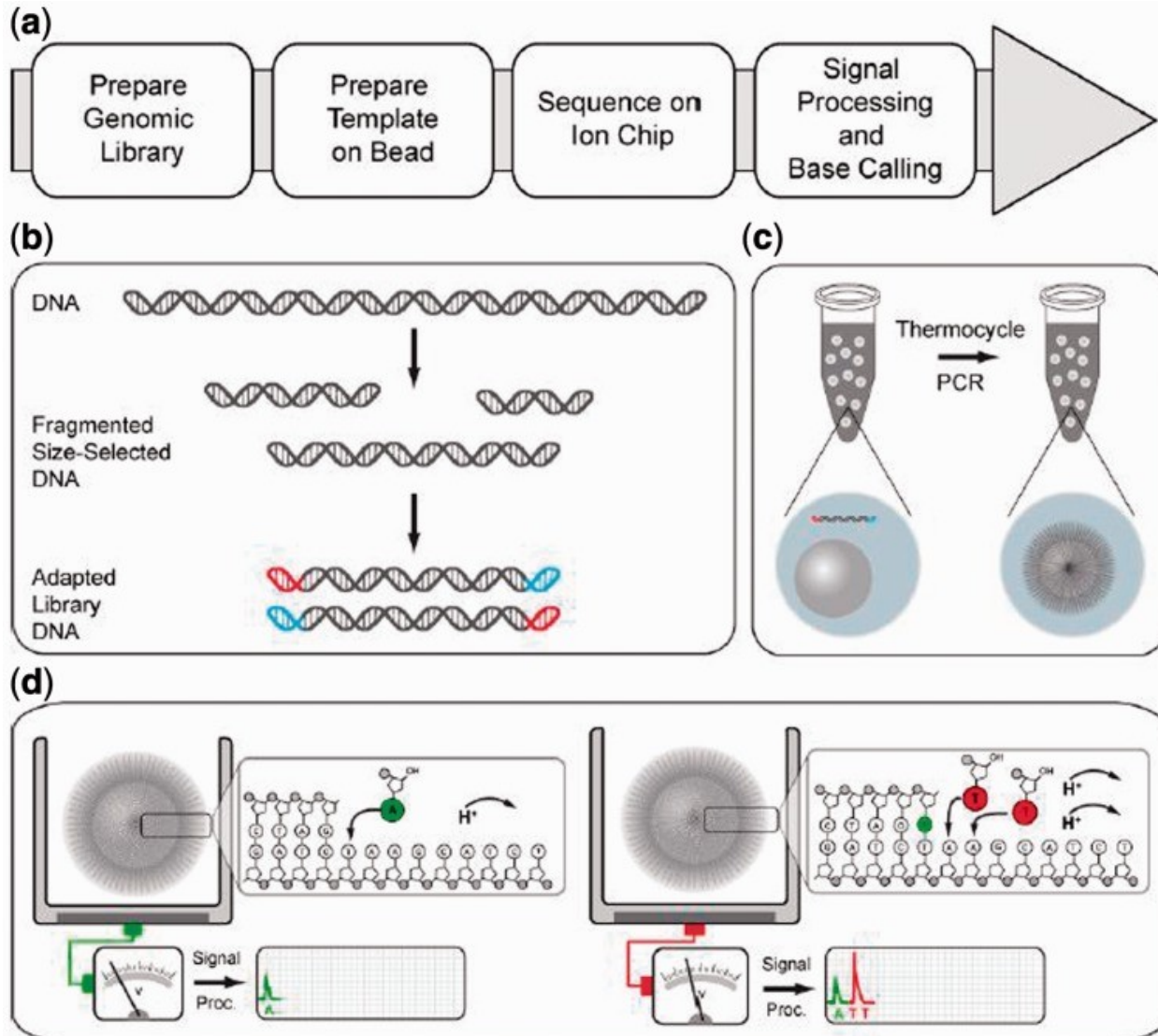
- Ion PGM x Ion Proton



- The chip is the machine
- Příprava templátu: Em PCR
- Sekvenace syntézou
- Detekce uvolněných protonů – změna pH

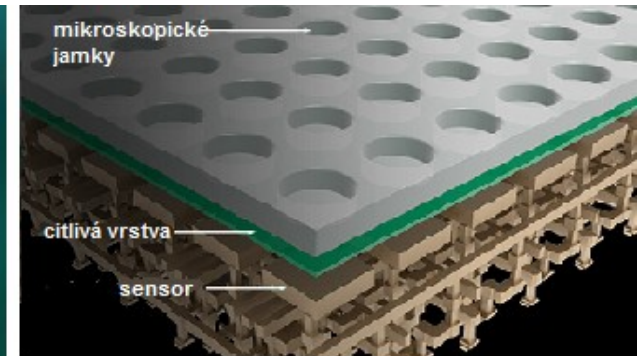
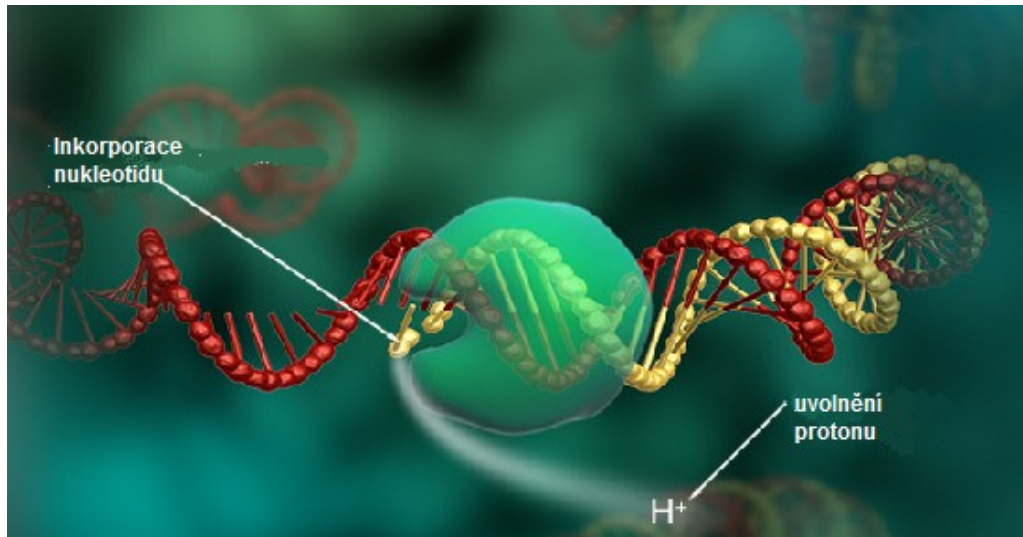
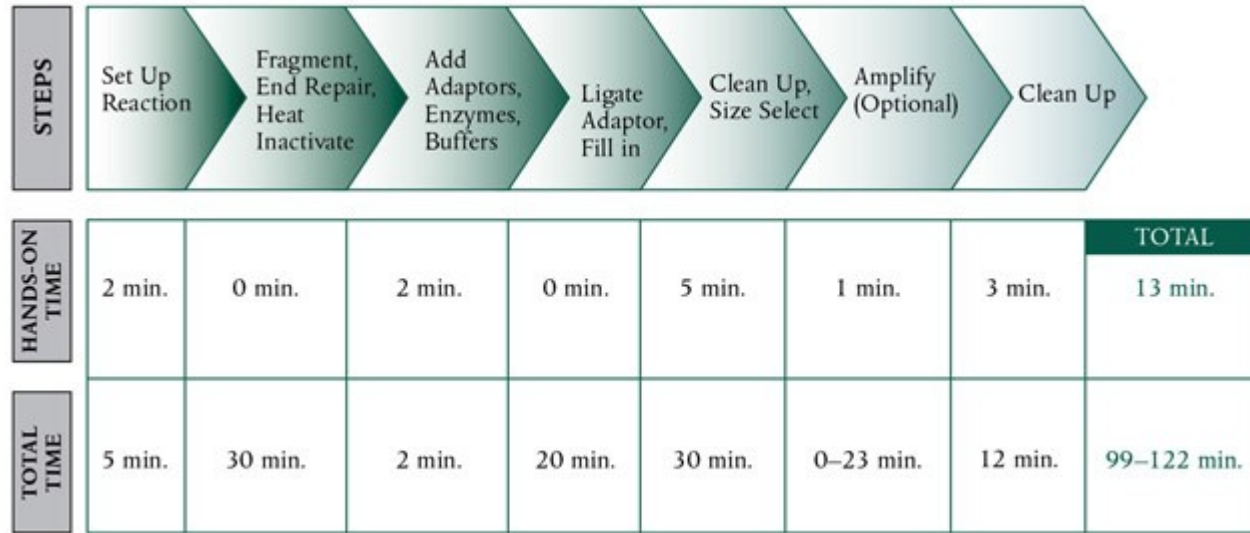


Ion Torrent

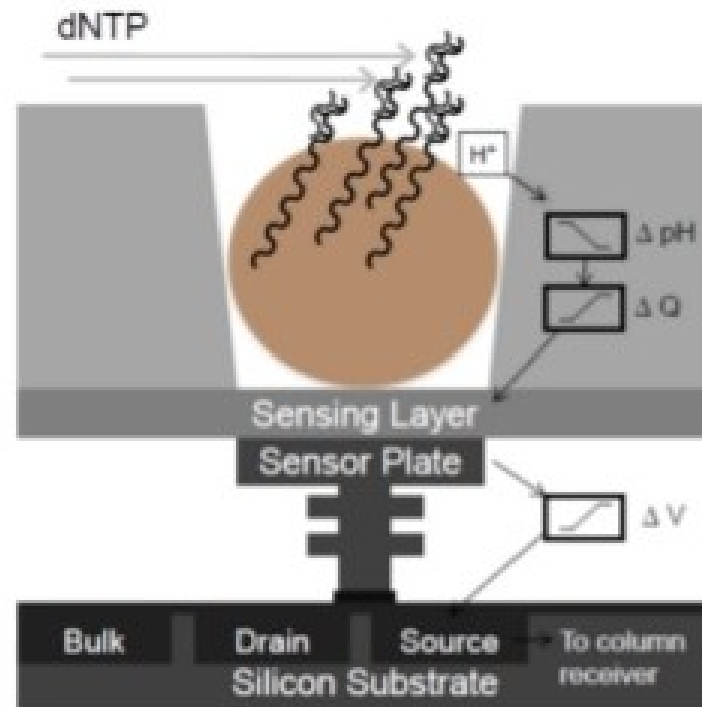
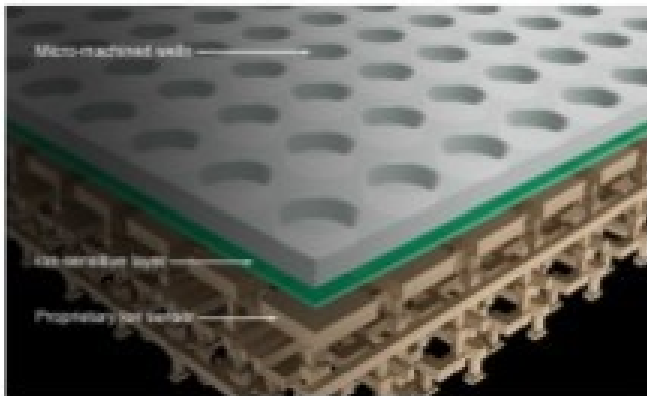


Product Name	SKU #	Product Size	Number of Wells	Platform	List Price (CZK)
Ion 314™ Chip Kit v2	4482261	1 kit	1 million wells per chip	Ion Personal Genome Machine® (PGM™) System	15.808,00
Ion 316™ Chip Kit	4466616	4 pack	6 million wells per chip	Ion Personal Genome Machine® (PGM™) System	28.616,00
Ion 316™ Chip Kit	4469496	8 pack	6 million wells per chip	Ion Personal Genome Machine® (PGM™) System	57.232,00
Ion 316™ Chip Kit v2	4483188	4 chips	6 million wells per chip	Ion Personal Genome Machine® (PGM™) System	28.616,00
Ion 316™ Chip Kit v2	4483324	8 chips	6 million wells per chip	Ion Personal Genome Machine® (PGM™) System	57.232,00
Ion 318™ Chip Kit (4 pack)	4466617	4 pack	11 million wells per chip	Ion Personal Genome Machine® (PGM™) System	49.280,00
Ion 318™ Chip Kit (8 pack)	4469497	8 pack	11 million wells per chip	Ion Personal Genome Machine® (PGM™) System	98.560,00
Ion 318™ Chip Kit v2	4484354	4 pack	11 million wells per chip	Ion Personal Genome Machine® (PGM™) System	49.280,00
Ion 318™ Chip Kit v2	4484355	8 pack	11 million wells per chip	Ion Personal Genome Machine® (PGM™) System	98.560,00
Ion PI™ Chip Kit v2	4482321	8 chips	165 million wells per chip	Ion Proton™ System	129.130,00

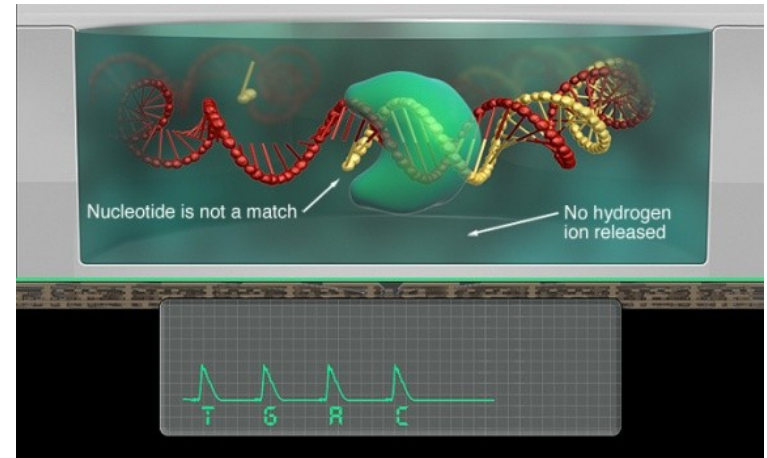
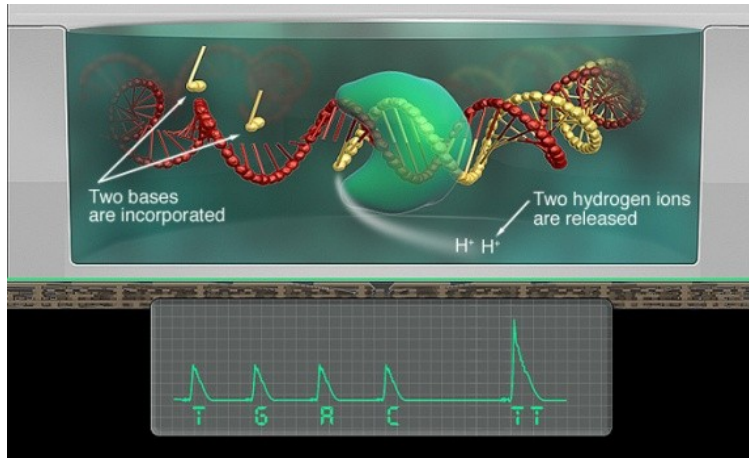
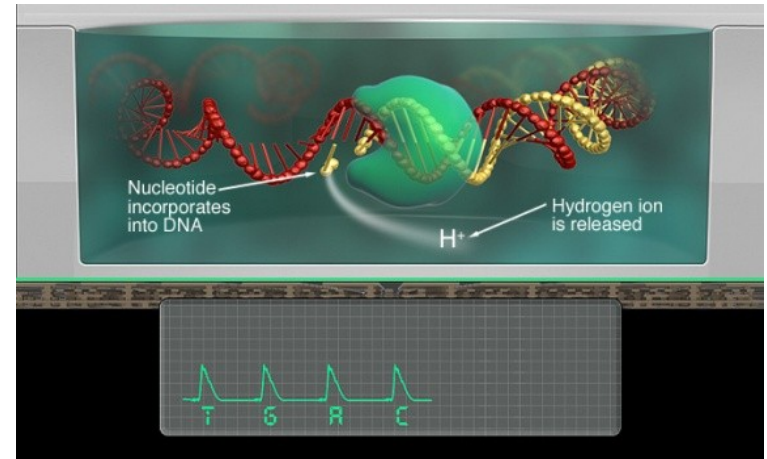
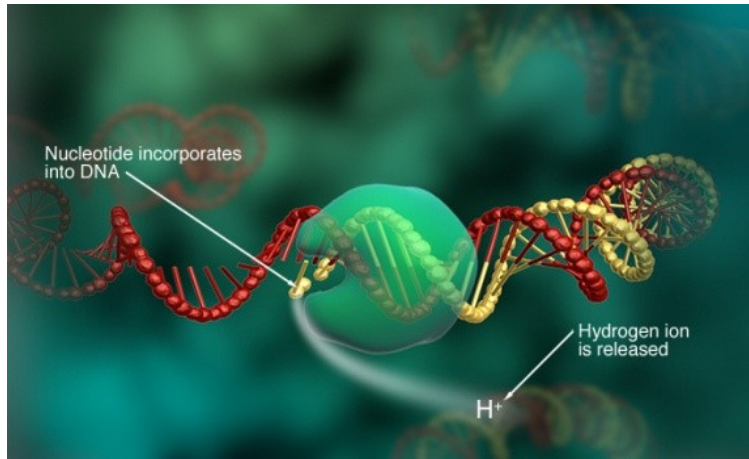
Ion Torrent



Ion Torrent



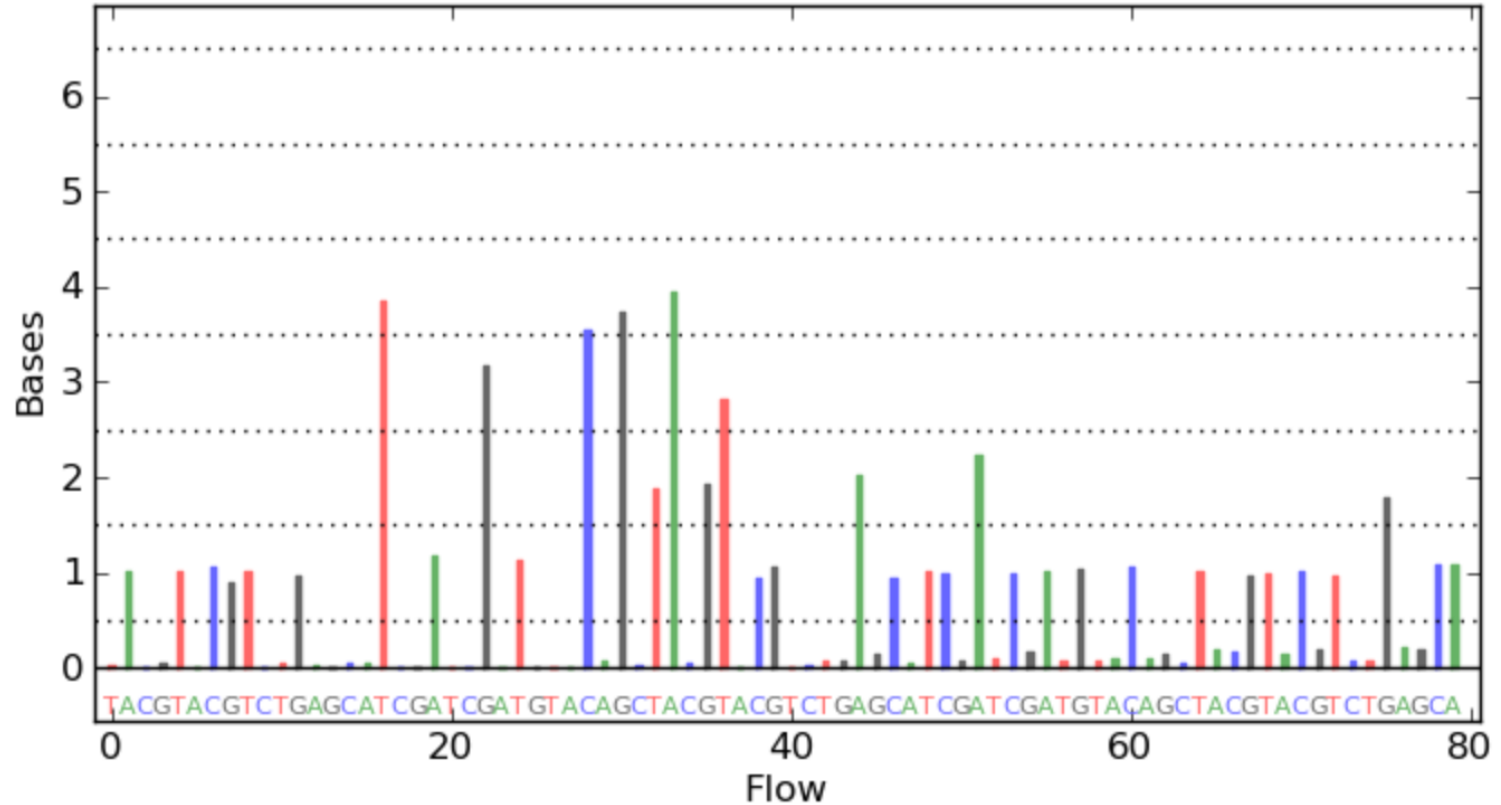
Ion Torrent



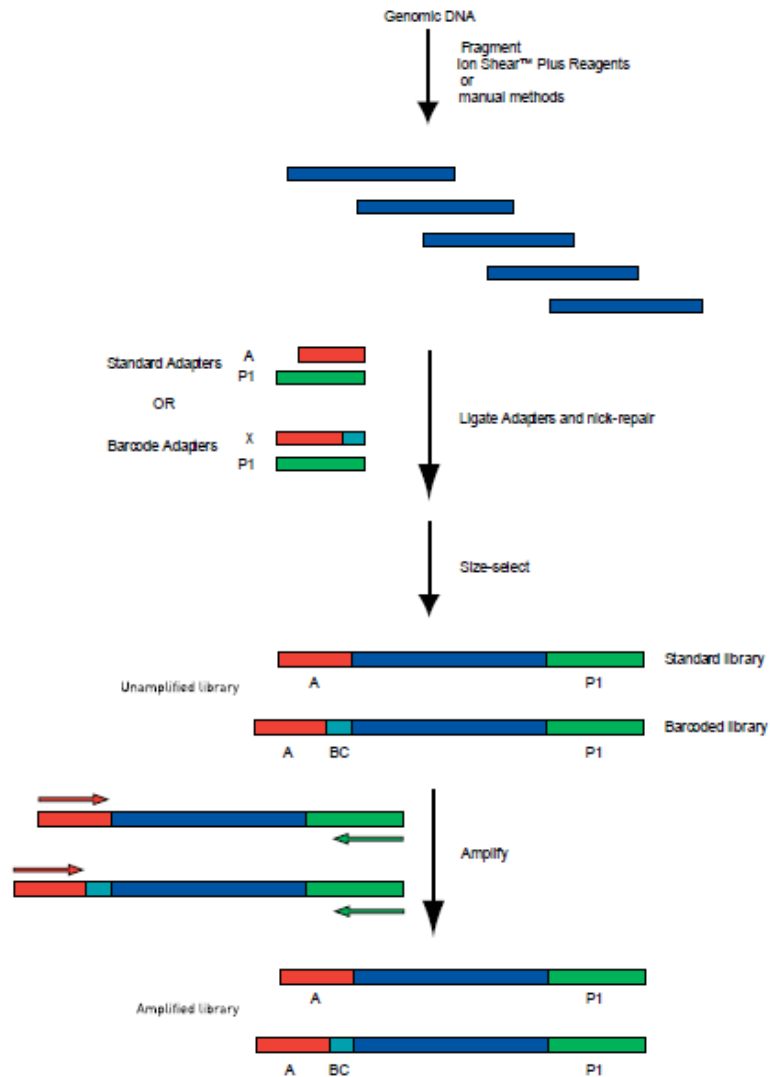
<https://www.youtube.com/watch?v=WYBzbxIfuKs>

Ion Torrent

Average Corrected Ionogram

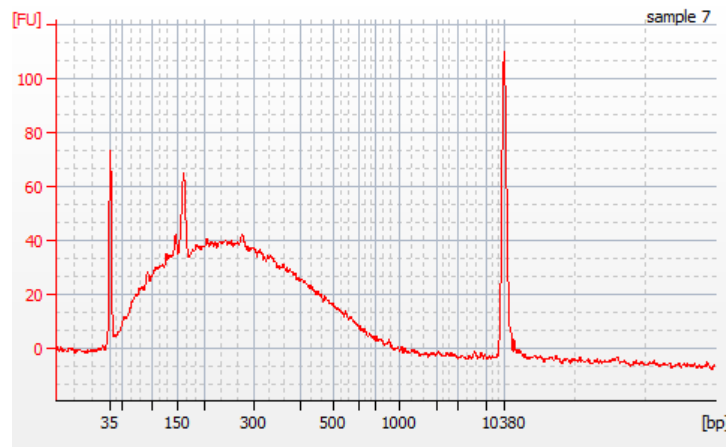
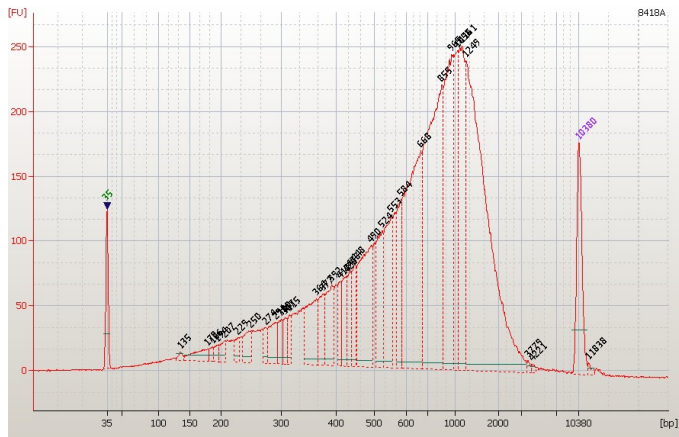
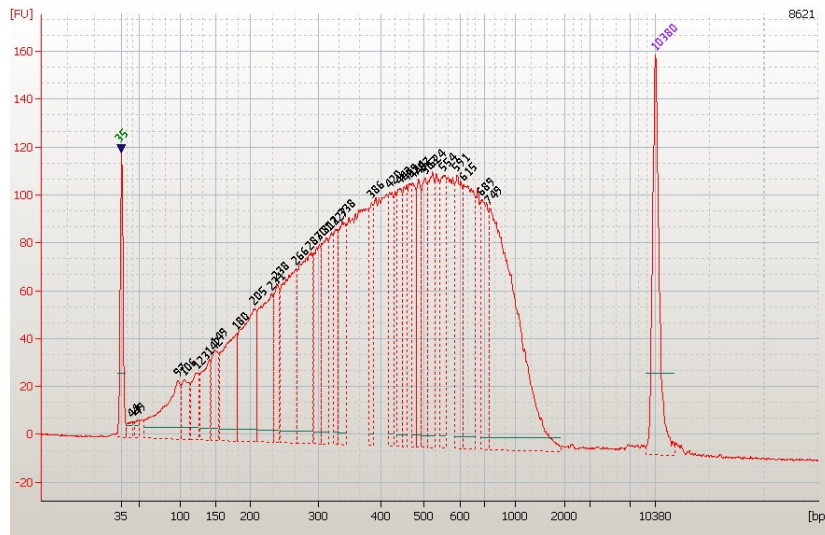


Příprava knihovny – celogenomové sekvenování



Příprava knihovny

FRAGMENTACE DNA Sonikace /enzymaticky



Příprava knihovny

END- REPAIR



LIGACE ADAPTORŮ + NICK REPAIR



SIZE SELECTION

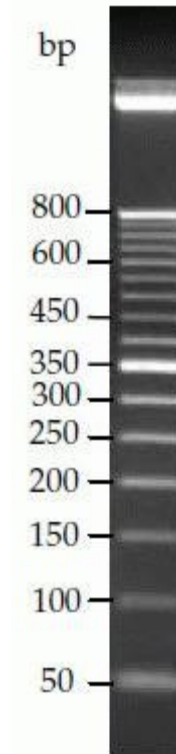
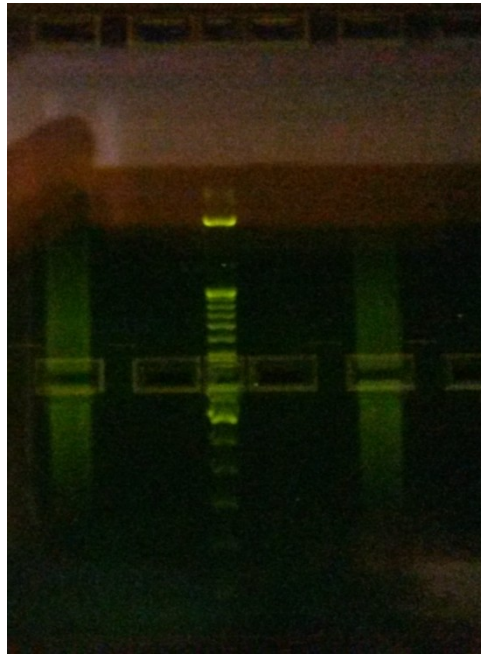
- E-gel 2 %



4

50bp
ladder

5



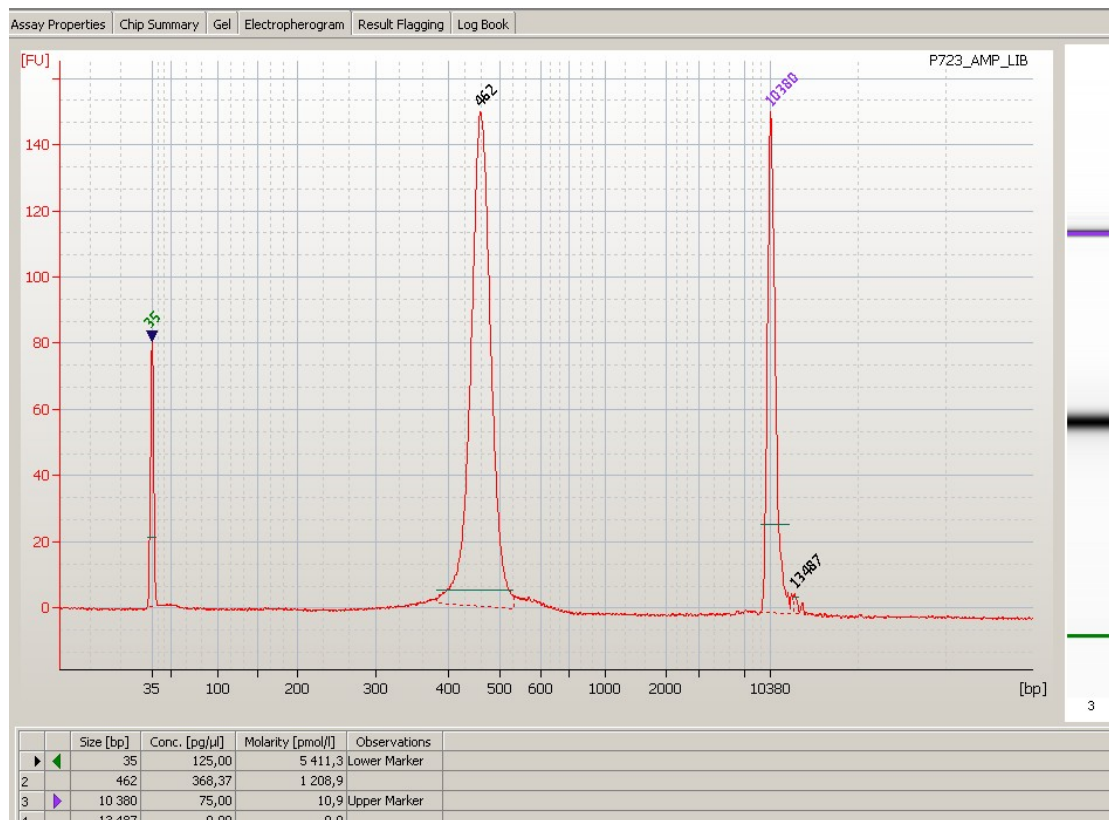
Příprava knihovny

AMPLIFIKACE KNIHOVNY (?)



KVANTIFIKACE KNIHOVNY

- qPCR / Agilent



Příprava knihovny

PŘÍPRAVA TEMPLÁTU

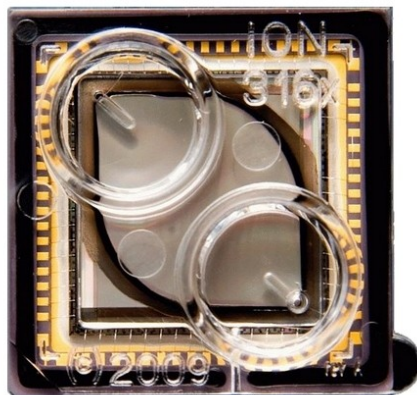
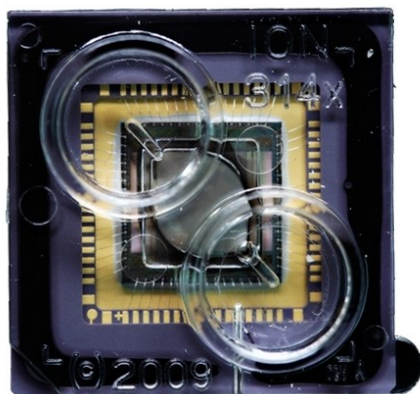
- Emulzní PCR



ENRICHMENT



SEKVENACE

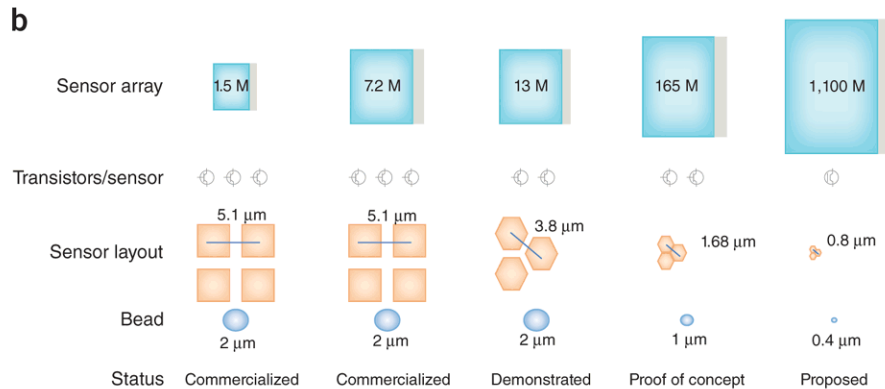
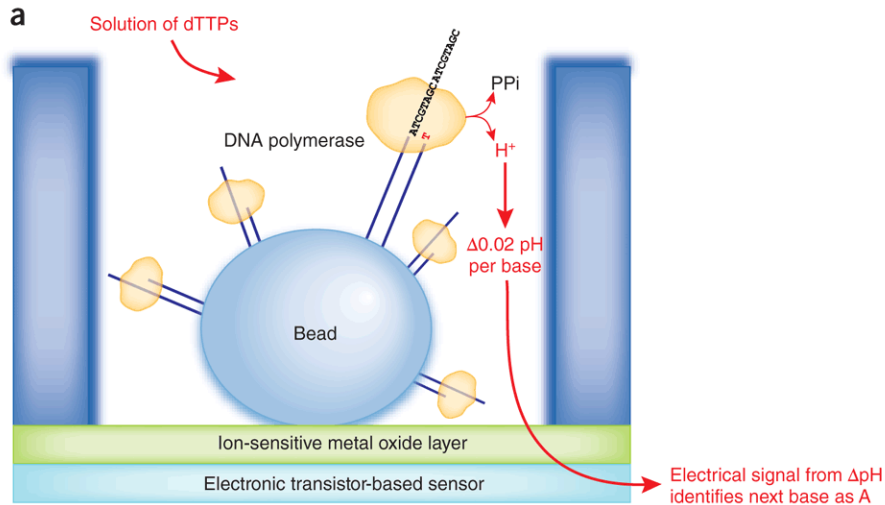


Chip	# Wells	# Reads	Throughput*	
			200 Base Read	400 Base Read
Ion 314™ Chip v2	~1.2 Million	400-500 thousand	30-50 Mb	60-100Mb
Ion 316™ Chip v2	~6 Million	2-3 million	300-600 Mb	600 Mb- 1Gb
Ion 318™ Chip v2	~11 Million	4 – 5.5 million	600 Mb- 1Gb	1.2 – 2 Gb ₇₀

Ion Torrent



Ion Torrent

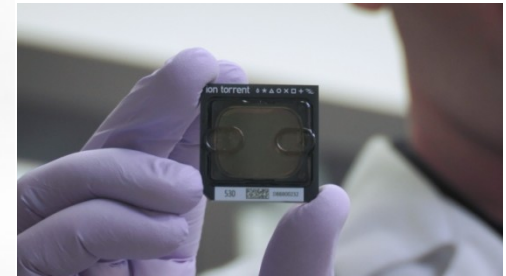


(a) Schematic of a well on an ion sequencing chip. Clonal DNA immobilized on a bead is extended by polymerase in the presence of a pure solution of one nucleotide (here 'T'). Nucleotide incorporation releases a pyrophosphate (PPi) and a hydrogen ion. The change in pH caused by release of the hydrogen ion alters the surface potential of the ion-sensitive metal oxide layer. This is converted to a voltage signal by transistors. The wells are washed and exposed sequentially to pure solutions of other nucleotides. For comparison, in high-throughput pyrosequencing, the pyrophosphate is converted to chemiluminescence by an enzymatic cascade and optically imaged. The size of the well relative to the bead has been exaggerated, although each well contains a single bead. (b) Evolution of ion sequencing chips. Increases in sensors per chip can be achieved by increasing the physical area of the sensor array, reducing the number of transistors per chip, arranging the sensors in a hexagonal rather than rectilinear geometry and reducing the well and bead size. Sensors are drawn to scale, and gray indicates sensor area not accessible to fluid. The 13-million (M) sensor design was used by Rothberg *et al.*¹ to sequence DNA from both *Escherichia coli* and human. Data for a fixed ('key') DNA sequence was shown for the 165-million sensor design. The 1,100-million sensor design was proposed but its feasibility was not shown.

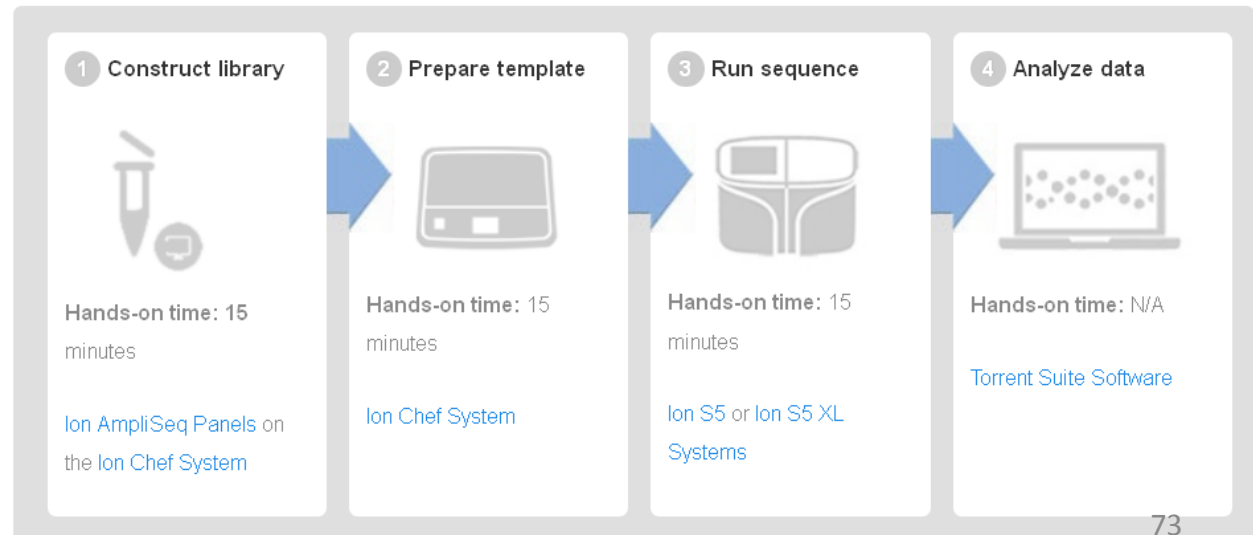
Ion Torrent

New Ion S5

„simplicity/speed/scalability/small sample input/service & support“



https://www.youtube.com/watch?v=jFCD8Q6qSTM&ebc=ANyPxKrMLmAe4Nmia2N3Rfr_1QbGUsOzcel2sMqnIJ5gS09XPCofTb-0cUvJdbzQhD_gKRKTL-XBahDEvoV6uOnm_78yvaG-eA



Ion S5 System



Simple workflow for panels, microbes, exomes, and transcriptomes

Ion S5 XL System



Simple, rapid workflow for panels, microbes, exomes, and transcriptomes

		Ion 520 Chip	Ion 530 Chip	Ion 540 Chip	Ion 520 Chip	Ion 530 Chip	Ion 540 Chip
Reads		3–5 million	15–20 million	60–80 million	3–5 million	15–20 million	60–80 million
Output*	200 bp	0.6–1 Gb	3–4 Gb	10–15 Gb	0.6–1 Gb	3–4 Gb	10–15 Gb
	400 bp	1.2–2 Gb	6–8 Gb	—	1.2–2 Gb	6–8 Gb	—
Run times	200 bp	2.5 hr	2.5 hr	2.5 hr	2.5 hr	2.5 hr	2.5 hr
	400 bp	4 hr	4 hr	—	4 hr	4 hr	—
Analysis time†	200 bp	5 hr	8 hr	16.5 hr	1 hr	2.5 hr	5 hr
	400 bp	8 hr	17.5 hr	—	2 hr	4 hr	—

* Expected output with >99% aligned/measured accuracy. Output dependent on read length and application.

† Analysis time to aligned BAM files.

O platformách

<http://dnasequencing.yolasite.com/next-generation-sequencing.php>

Srovnání:

	454 (Junior/FLX)	Illumina (MiSeq/HiSeq)	Ion Torrent (PGM/
Počet čtení/run	100 tis/1 mil	35-50 milionů PE/ 8 miliard PE	5,5 mil/ 60-80 mil
Průměrná délka čtení [bp]	450/700	2x300/ 2x250	400
Doba běhu	6/24 hodin	1/10 dní	7/4 hodiny
Výhody	délka čtení, přesnost, rychlost,	snadná příprava, velké množství sekvencí, nejnižší cena	rychlost, relativně nízké náklady, různé čipy (outputy)
Nevýhody	pracnost, cena, chybovost v polymorfismech, nízké outputy - technologie je tak drahá, že již není více konkurenceschopná	nižší přesnost na konci readů, interference u nízkodiverzních knihoven	chybovost v homopolymerech, EM PCR
Shotgun knihovny	 <p>Diagram showing shotgun sequencing with 454 reads. A 600 bp DNA fragment is shown with two reads of 400 bp and 450 bp. The reads overlap significantly, with the 400 bp read starting at the beginning and the 450 bp read starting further along.</p>	 <p>Diagram showing shotgun sequencing with Illumina reads. A 600 bp DNA fragment is shown with two reads of 250 bp each. The reads overlap in the middle, with each read starting 250 bp from the beginning of the fragment.</p>	 <p>Diagram showing shotgun sequencing with Ion Torrent reads. A 600 bp DNA fragment is shown with two reads of 400 bp and 450 bp. The reads overlap, with the 400 bp read starting at the beginning and the 450 bp read starting further along.</p>
Amplikony	 <p>Diagram showing amplicon sequencing with 454 reads. Two overlapping DNA fragments are shown, each with a 450 bp read. The reads overlap significantly.</p>	 <p>Diagram showing amplicon sequencing with Illumina reads. Two overlapping DNA fragments are shown, each with a 300 bp read. The reads overlap in the middle, with an overlap of 150 bp. The total length of the overlapping region is 450 bp.</p>	 <p>Diagram showing amplicon sequencing with Ion Torrent reads. Two overlapping DNA fragments are shown, each with a 450 bp read. The reads overlap significantly.</p>