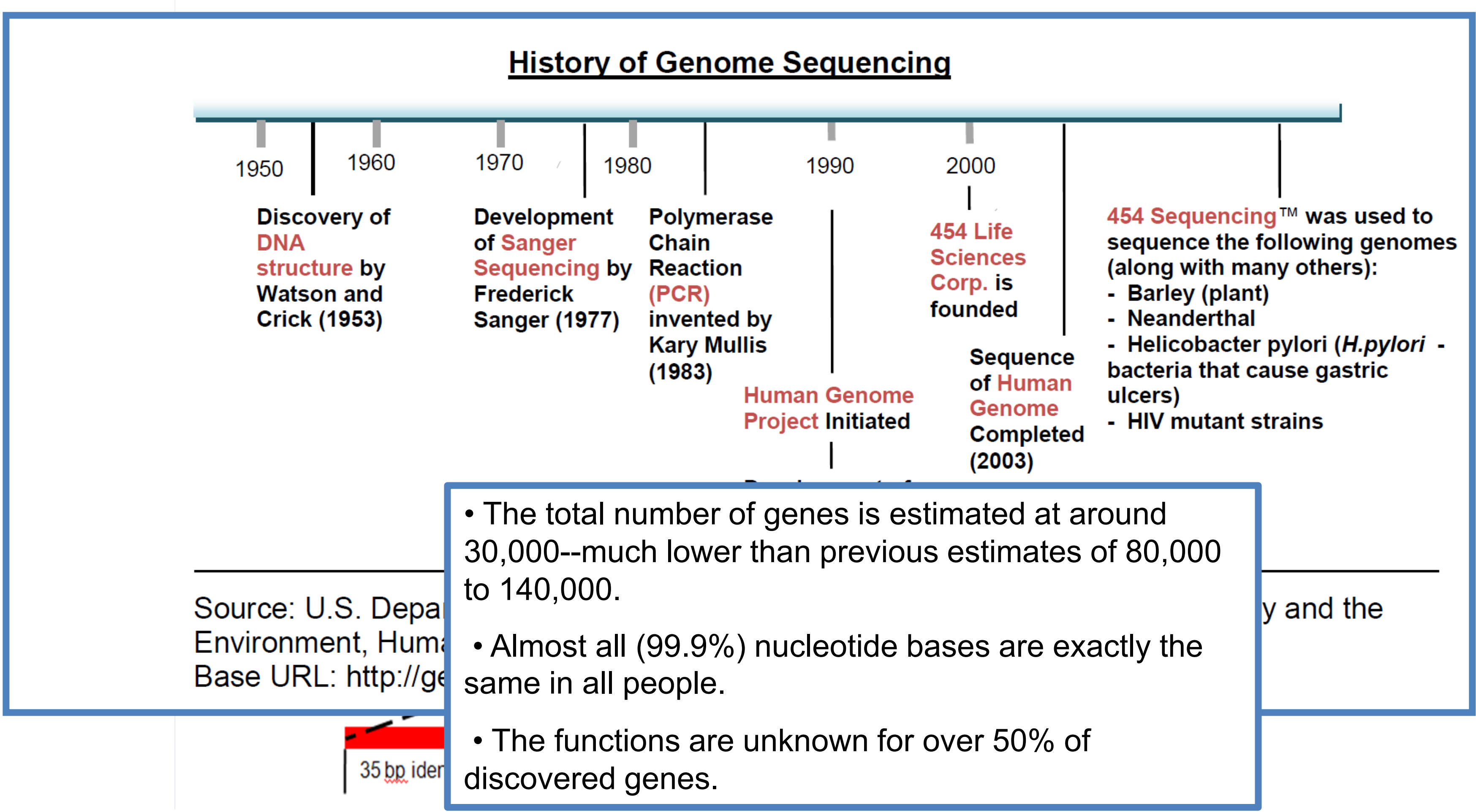
Metagenomika – NGS (454, Illumina, IonTorrent)

Petra Víděňská, Ph.D.

Next Generation Sequencing



Využití next generation sekvenování

Paired-end mapping and depth of coverage approach
- Genome-wide discovery and detection of copy number variations and balanced genomic rearrangements like inversions and translocations

Whole genome sequencing and targeted sequencing
- *De novo* assembly or sequencing of organism genomes
- Discovery of genetic variations/ mutations (germline and somatic) like SNPs and small indels
- Fine mapping of the genomic loci found by linkage/GWAS

ChIP-Seq
- Genome-wide interrogation of protein–DNA interactions like mapping of the binding sites for transcription factors and other DNA-binding proteins
- Genome-wide investigation of various histone modifications

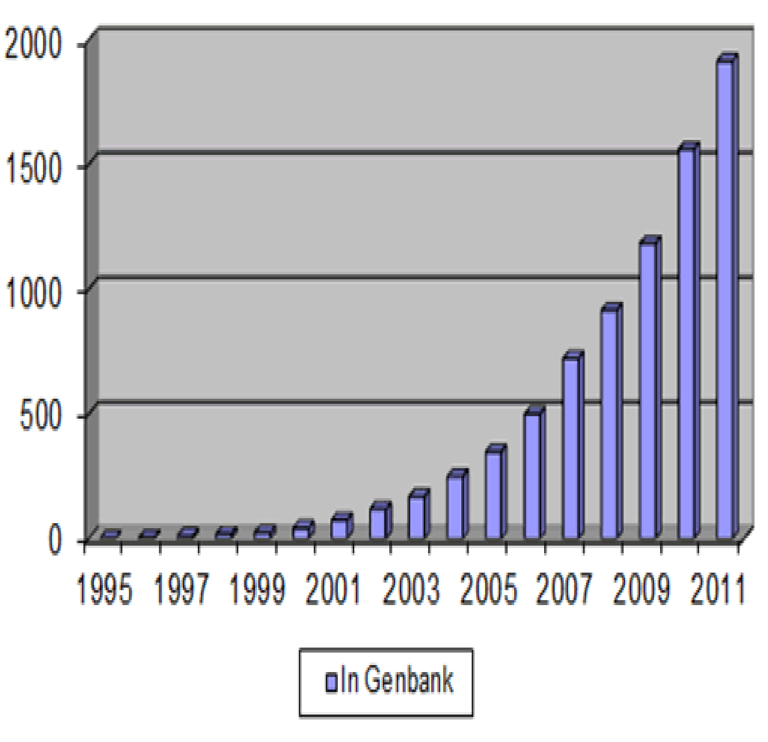
Applications of next generation sequencing technologies in structural and functional genomics

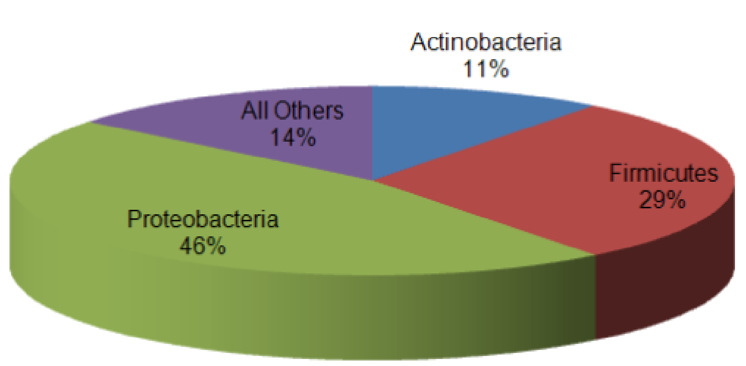
RNA-Seq
- Expression analysis of coding and noncoding RNAs/transcripts
- Discovery of new transcripts
- Study of alternative splicing patterns
- Detection of transcript fusions
- Analysis of allele specific expression

Metagenomics
- Study of the diversity and abundance of bacterial/microbial species in a sample by sequencing the genetic materials obtained from the samples and mapping the sequences against different species' reference genomes

Sequencing of bisulfite-treated DNA
- Study of DNA methylation at the 5′-position in cytosine
- Human Epigenome Project aims to characterize genome-wide DNA methylation patterns of all human genes in all major tissues

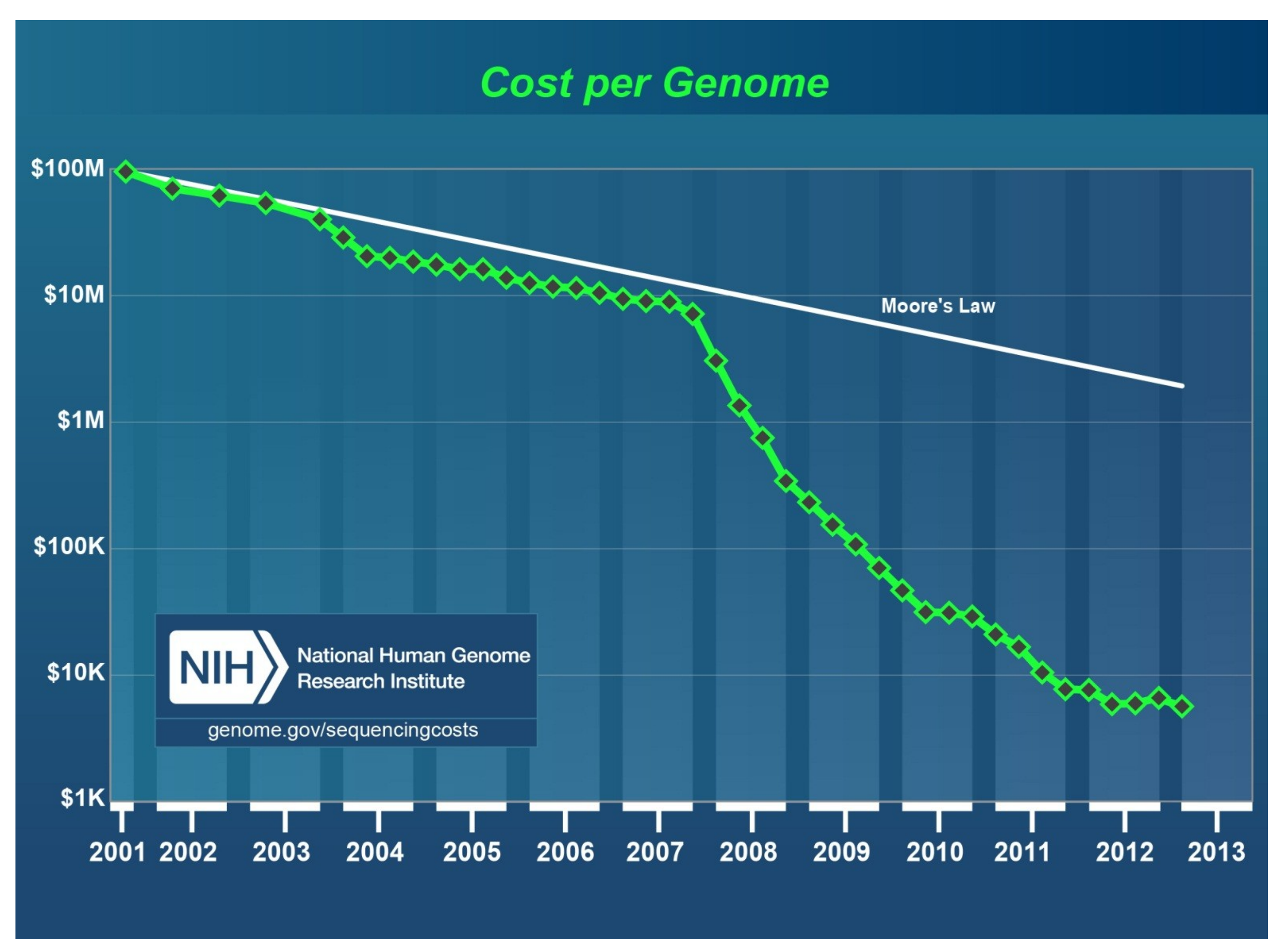
Počet kompletně osekvenovaných genomů





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, můstková amplifikace) – větší signál při inkorporaci nukleotidů během sekvenace, umožňující detekci

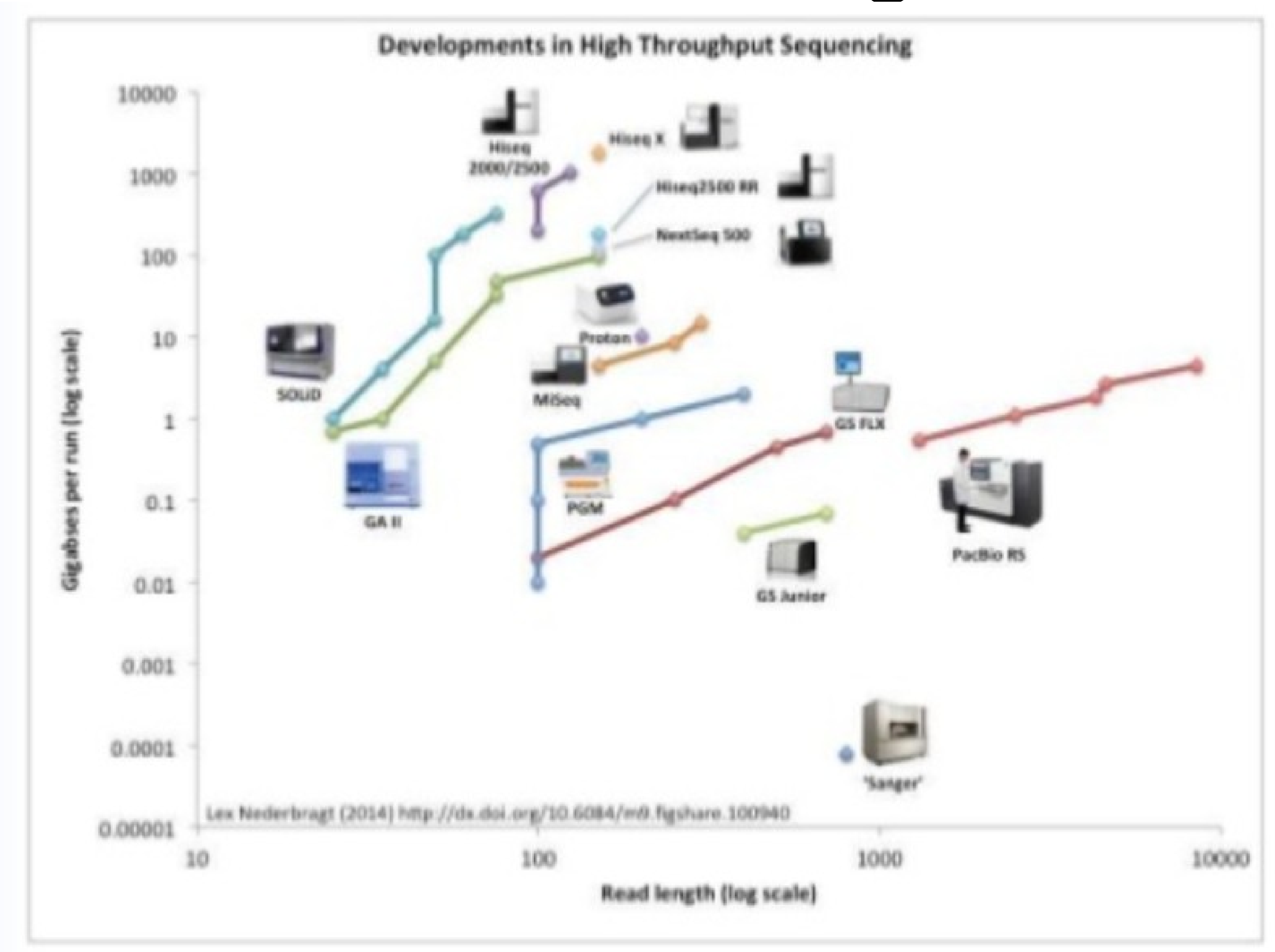
Sekvenování 3. generace

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

Dostupné platformy

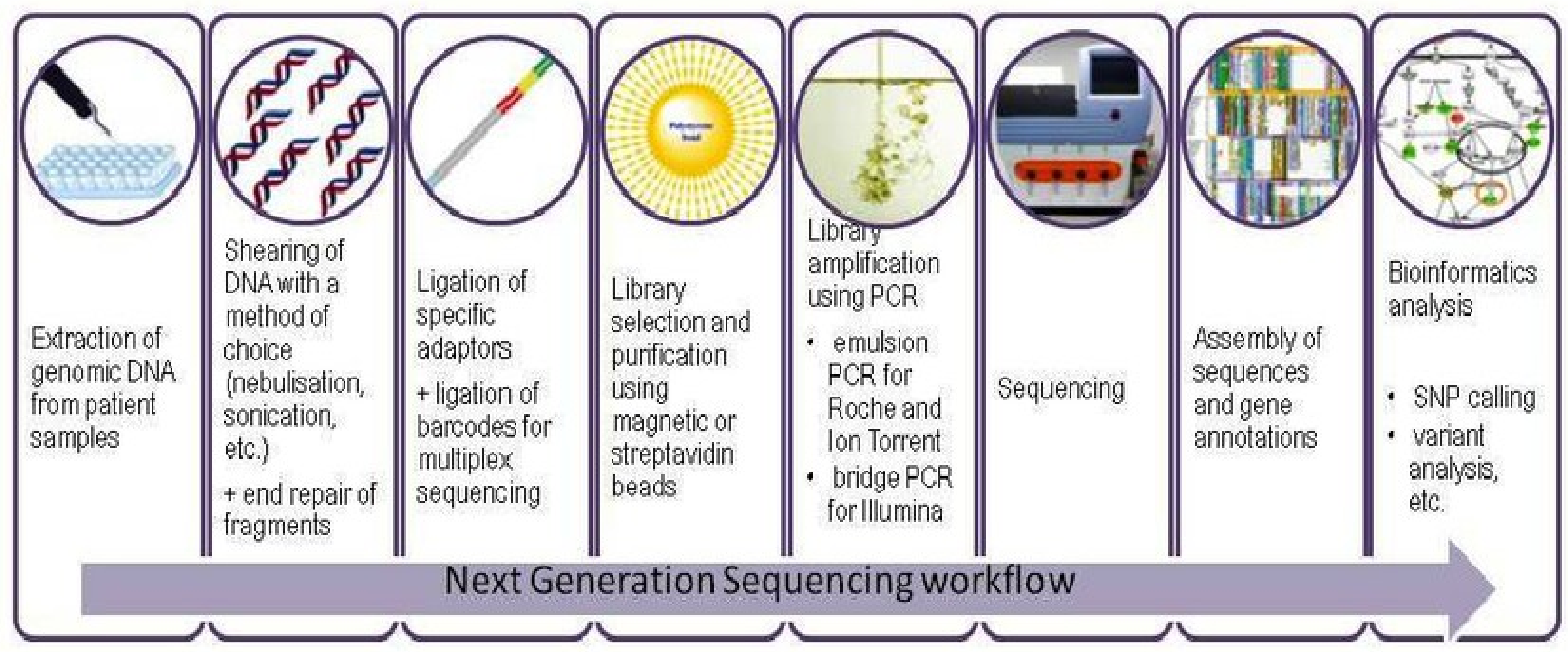
- 454 (Roche)
- SOLiD (Life Technologies)
- Illumina (Illumina)
- Ion Torrent (Life Technologies)
- **PACBIO, Sequel System (Pacific BioSciences)**
- **MiniON (Oxford Nanopore Technologies)**
- BGISEQ-500 (BGI)

Sekvenování nové generace



Newest Illumina HiSeq X 10 > 1 Tb of sequene data

Sekvenační workflow

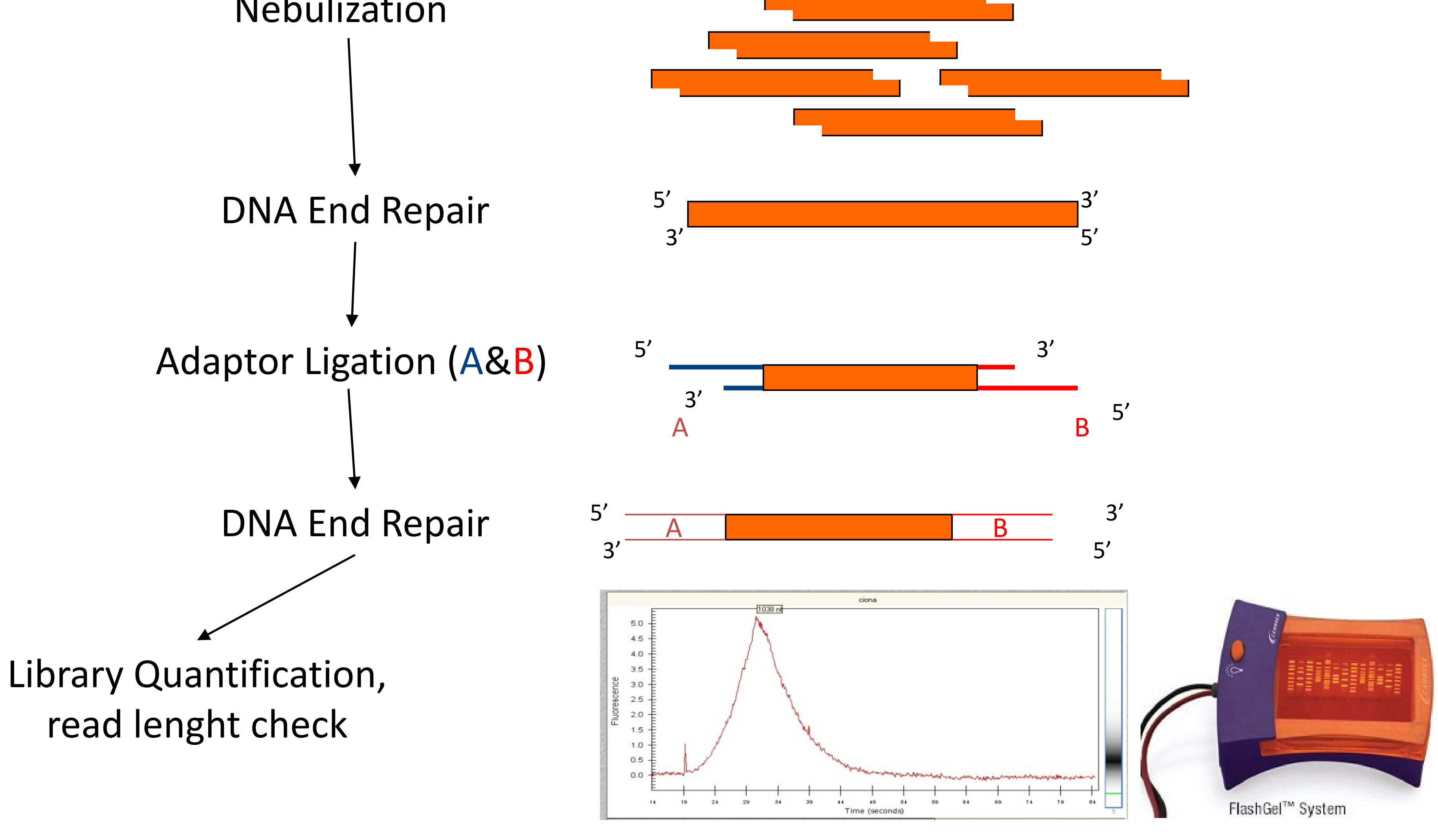


454

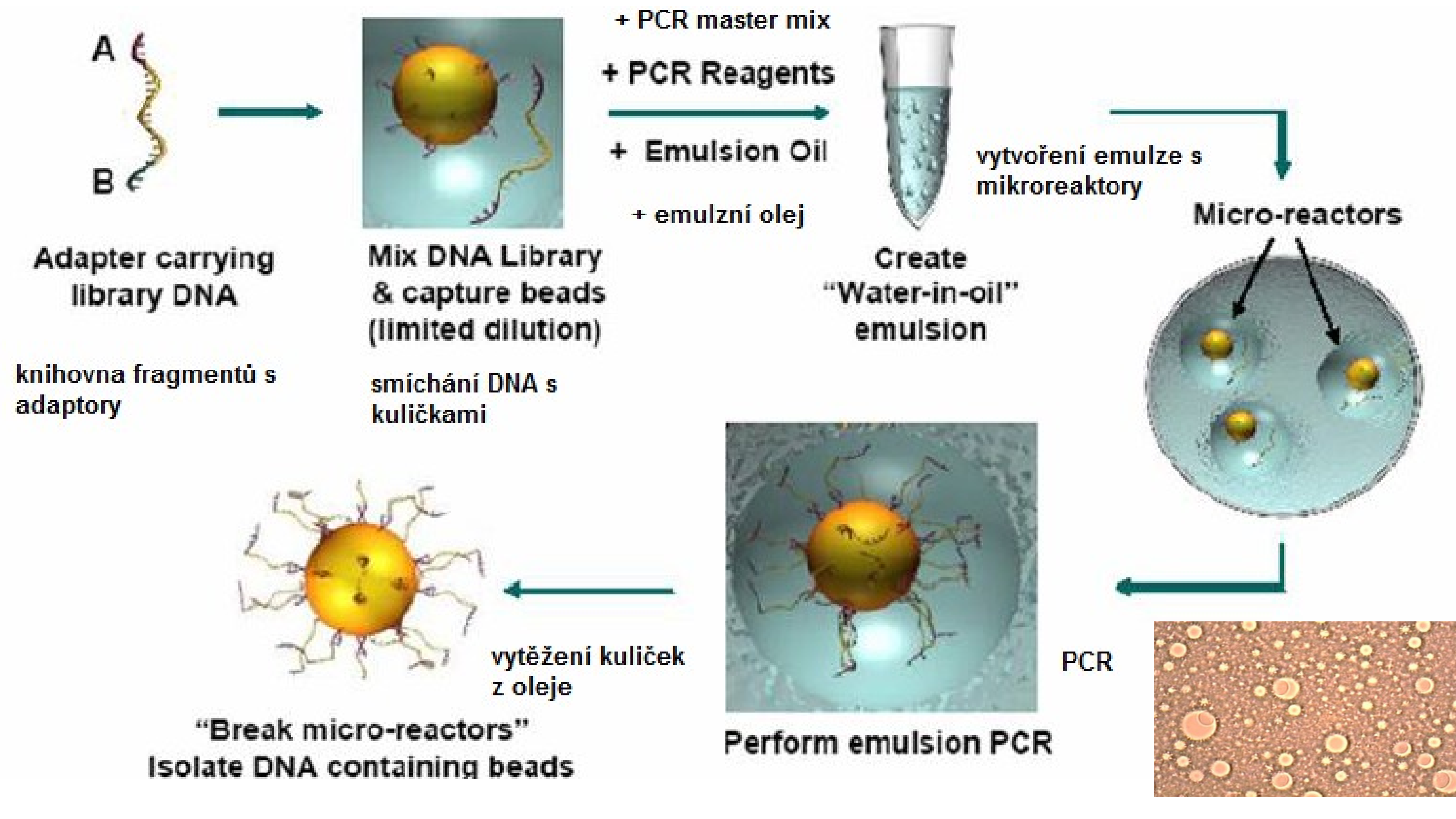
- Roche
- 454 GS Junior (35 MB) x 454 GS FLX (700 MB)

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

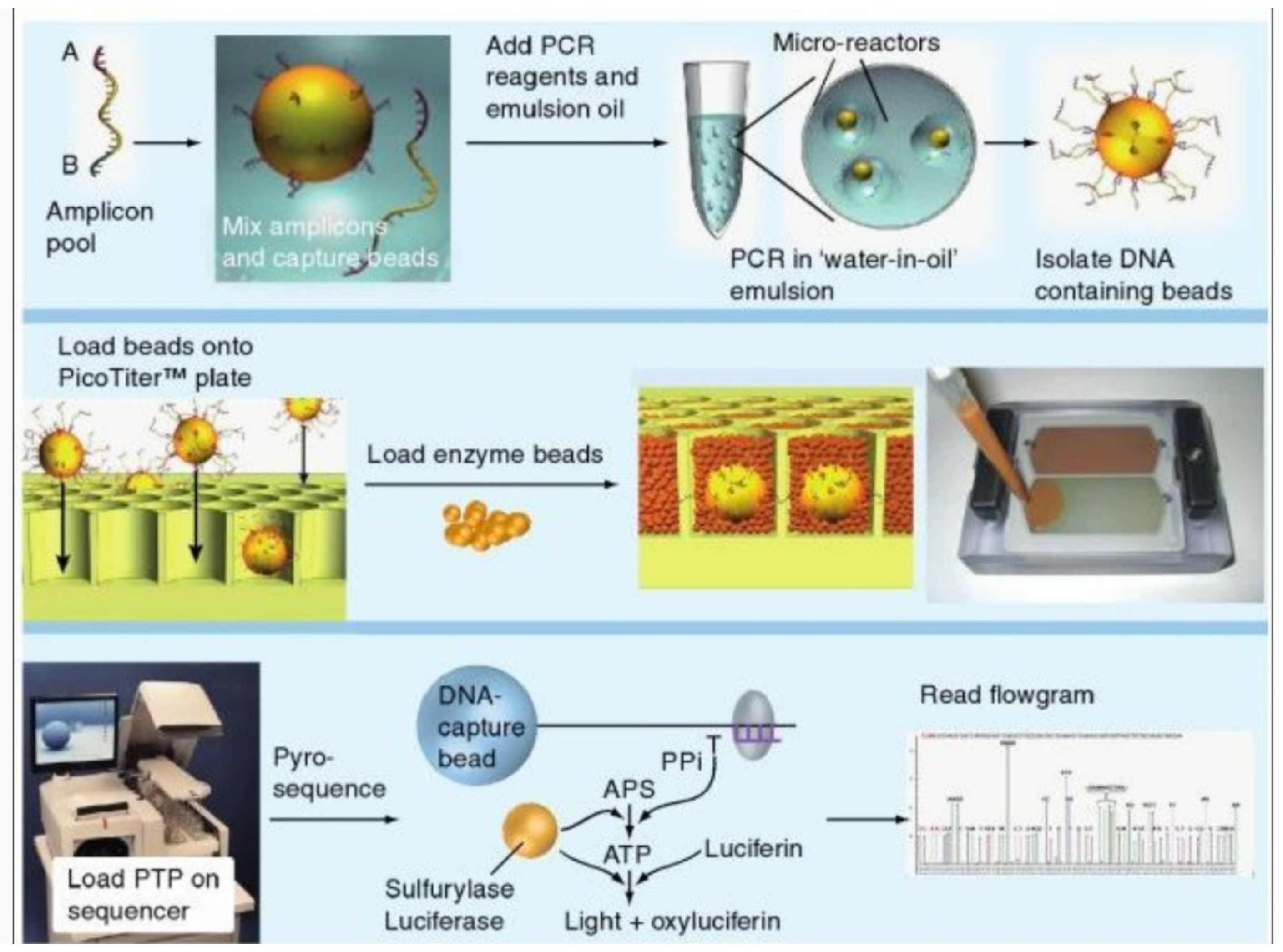
454 Shotgun – příprava knihovny



454

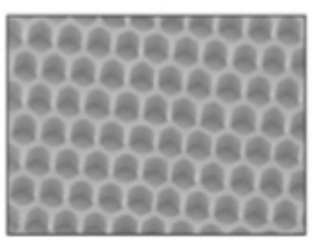


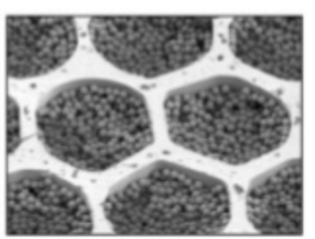
454

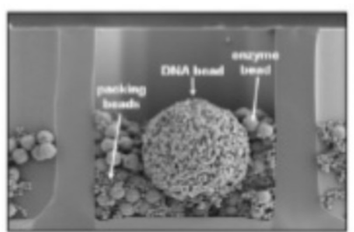


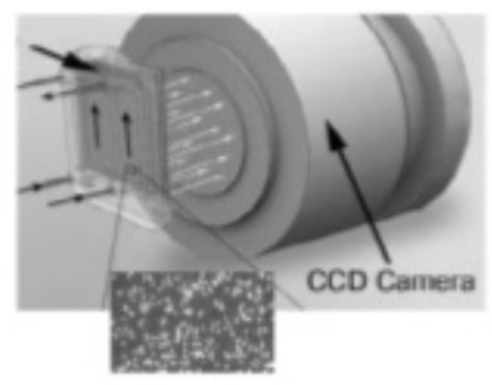
https://www.youtube.com/watch?v=nFfgWGFe0aA

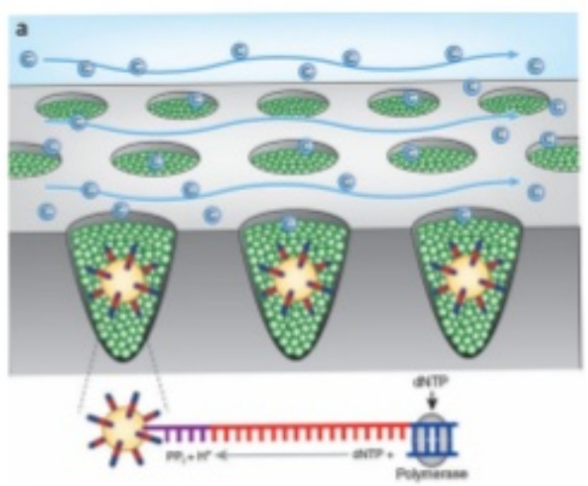
454



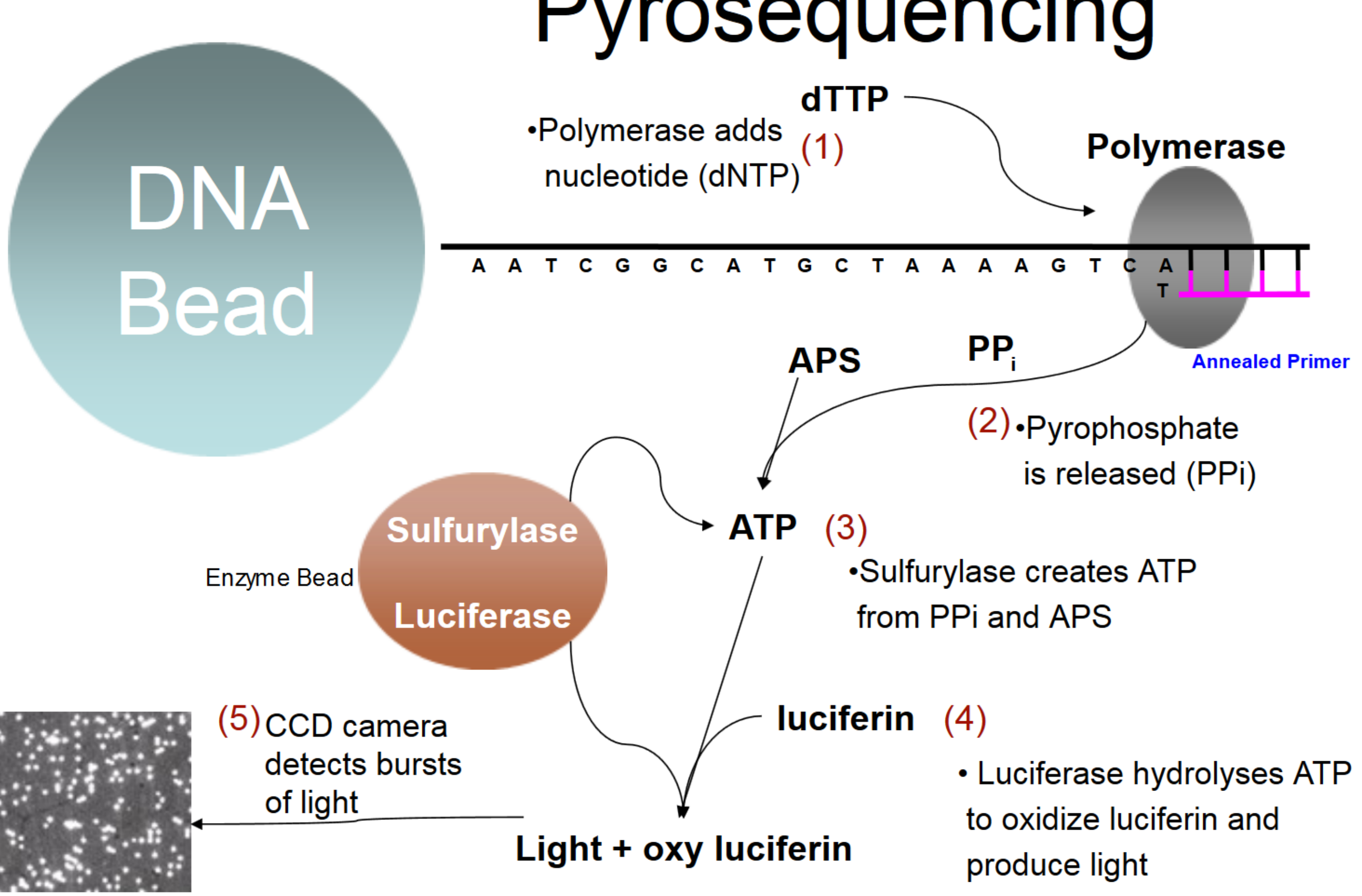


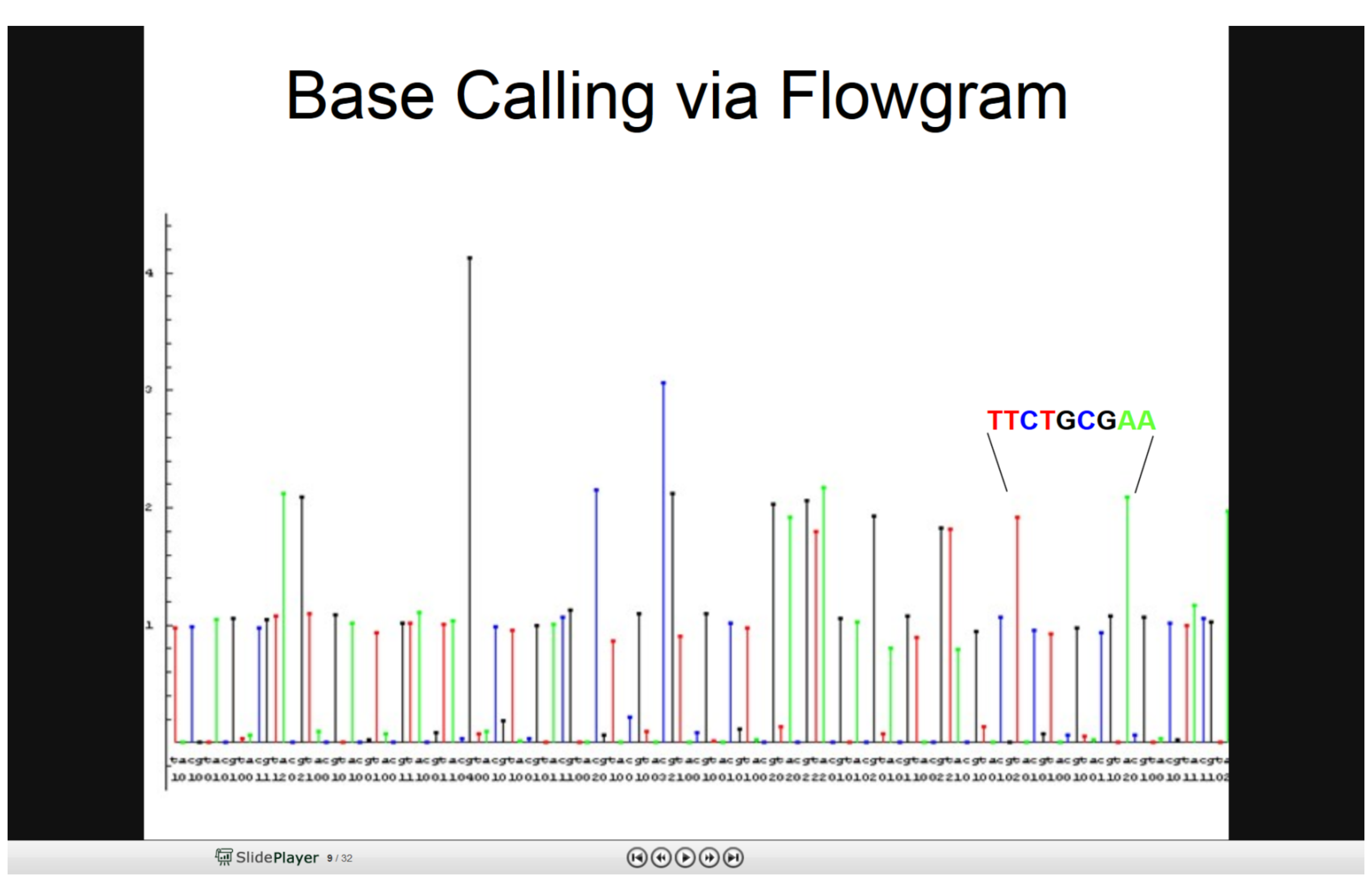






454



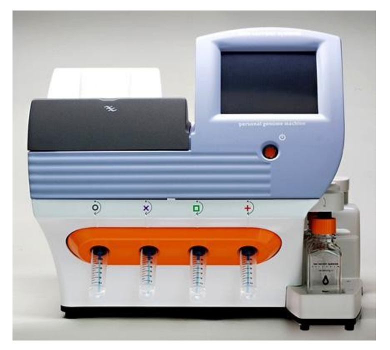

Base Calling via Flowgram
TTCTGCGAA
SlidePlayer 9 / 32

454 – podrobný workflow

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

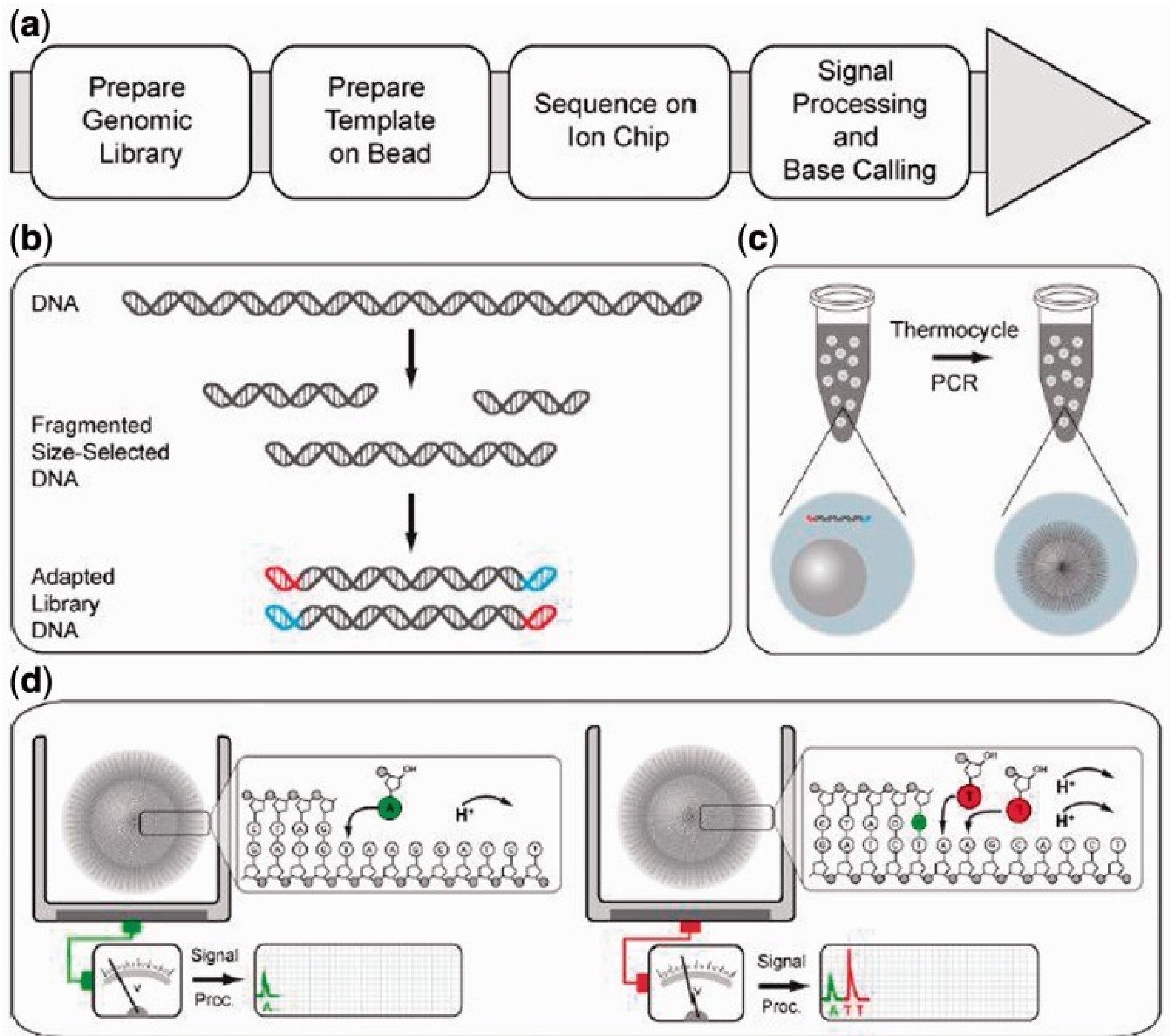
Ion Torrent

- Ion PGM x Ion Proton



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

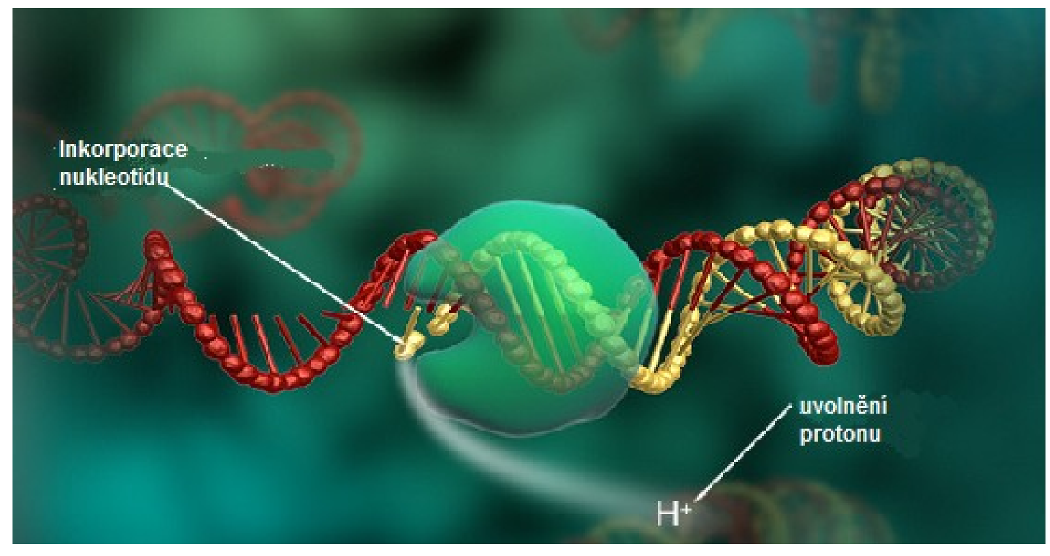
Ion Torrent

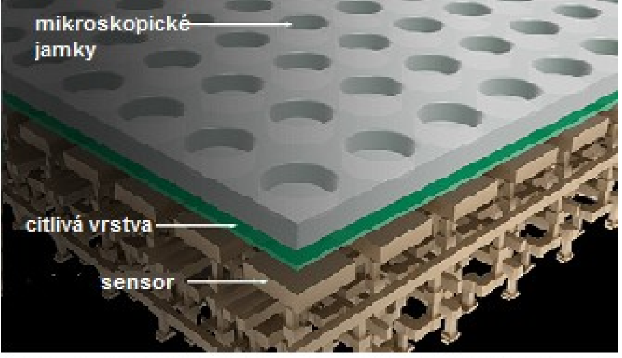


Product Name	SKU #	Product Size	Number of Wells	Platform	List Price (CZK)
Ion 314™ Chip Kit v2	4482261	1 kit	1 million wells per chip	Ion Personal Genome Machine® (PGM™) System	15.808,00
Ion 316™ Chip Kit	4466616	4 pack	6 million wells per chip	Ion Personal Genome Machine® (PGM™) System	28.616,00
Ion 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

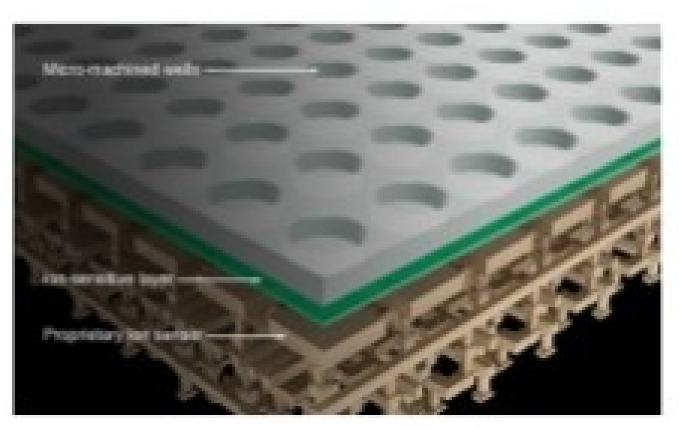
Ion Torrent

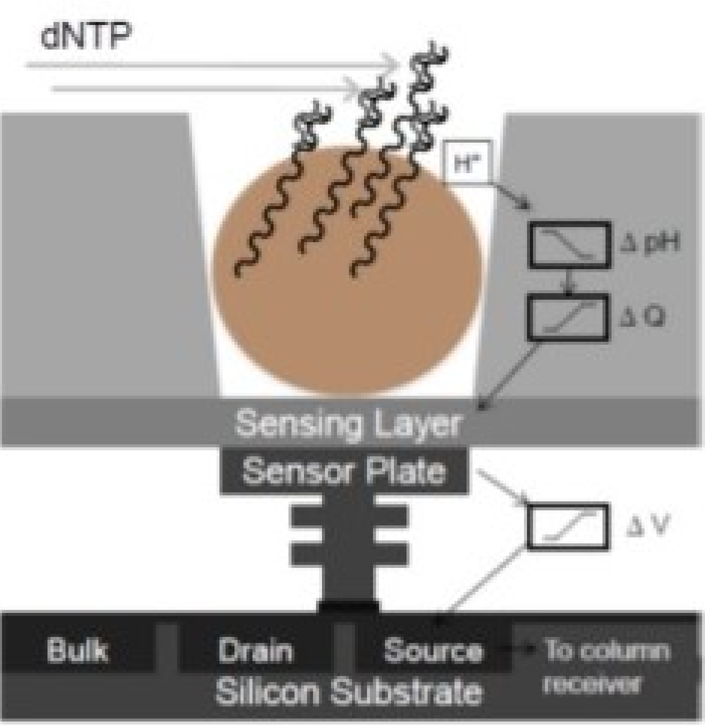
STEPS	Set Up Reaction	Fragment, End Repair, Heat Inactivate	Add Adaptors, Enzymes, Buffers	Ligate Adaptor, Fill in	Clean Up, Size Select	Amplify (Optional)	Clean Up	TOTAL
HANDS-ON TIME	2 min.	0 min.	2 min.	0 min.	5 min.	1 min.	3 min.	13 min.
TOTAL TIME	5 min.	30 min.	2 min.	20 min.	30 min.	0–23 min.	12 min.	99–122 min.



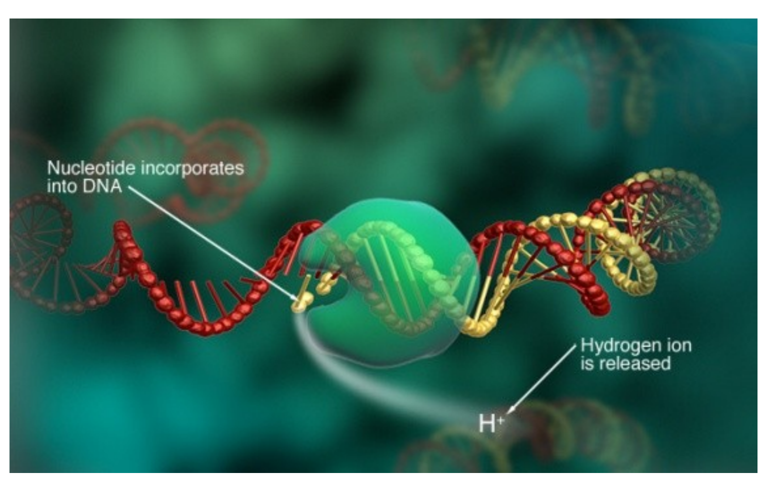


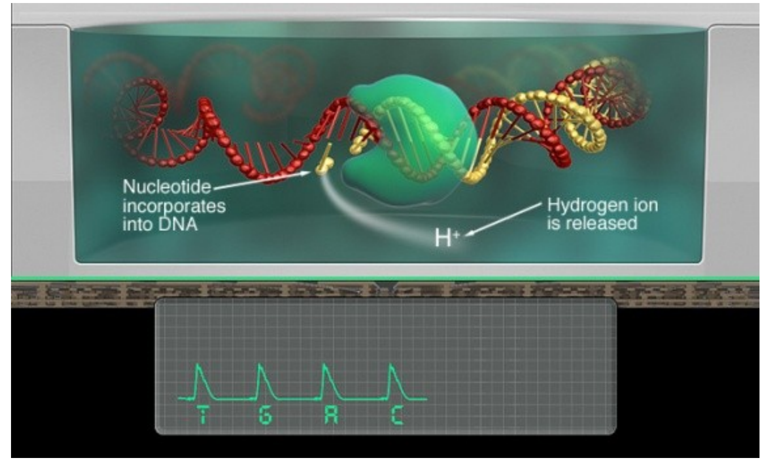
Ion Torrent

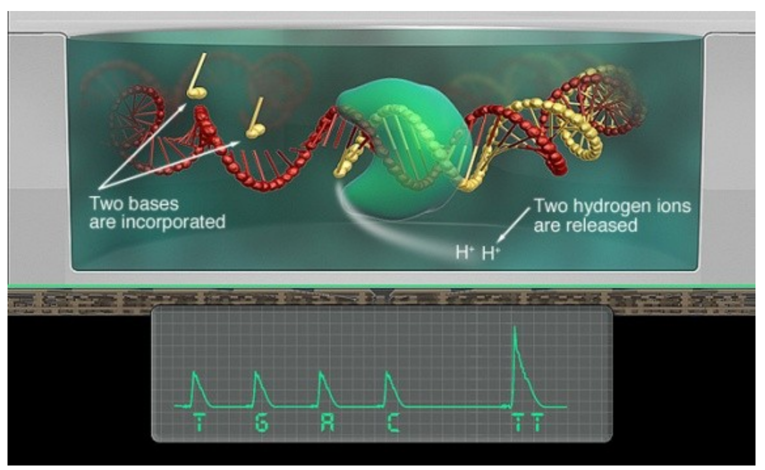


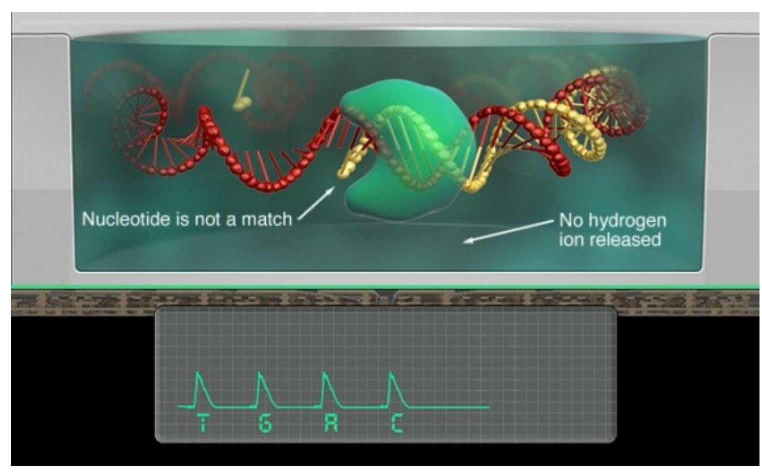


Ion Torrent





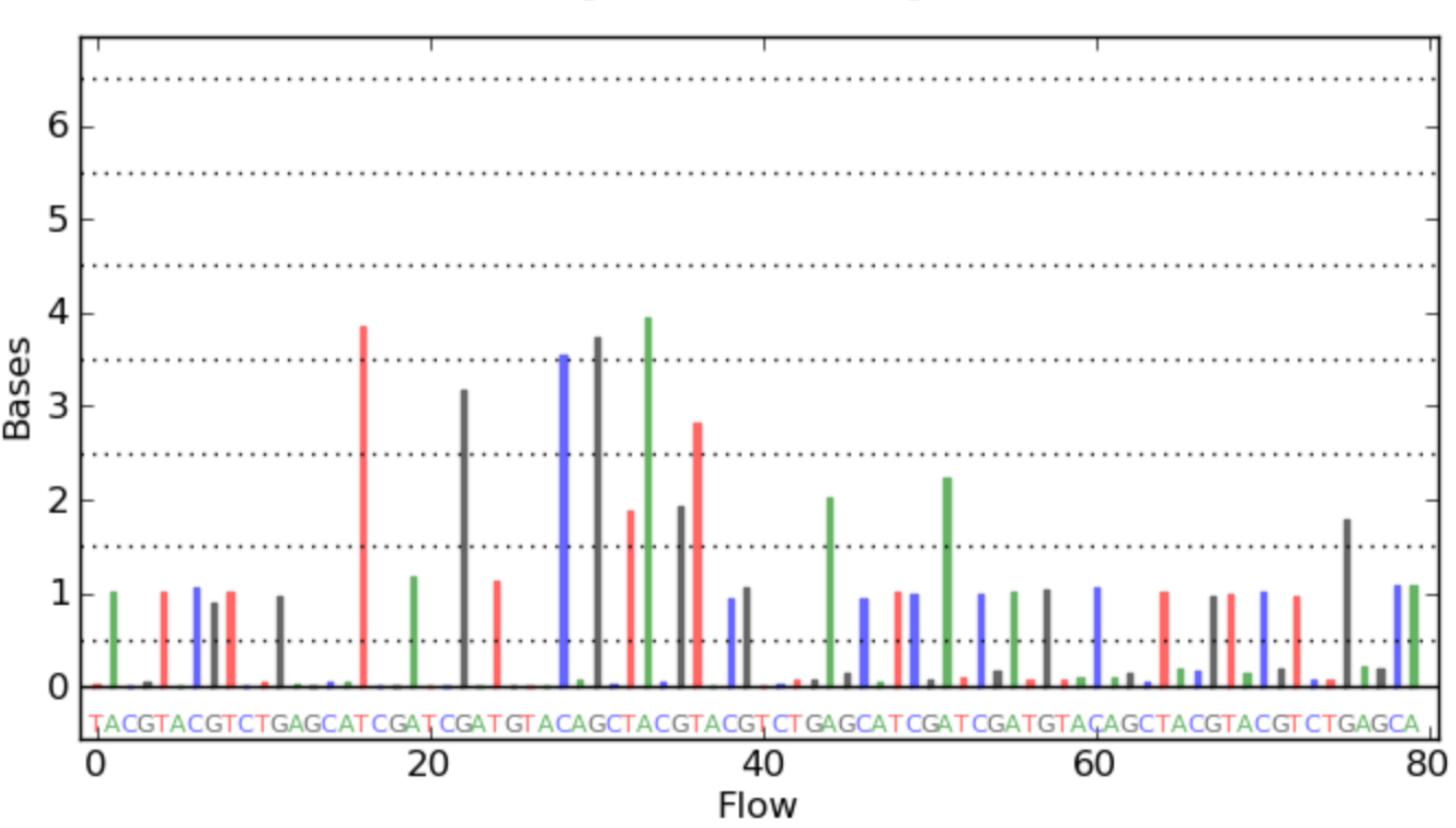




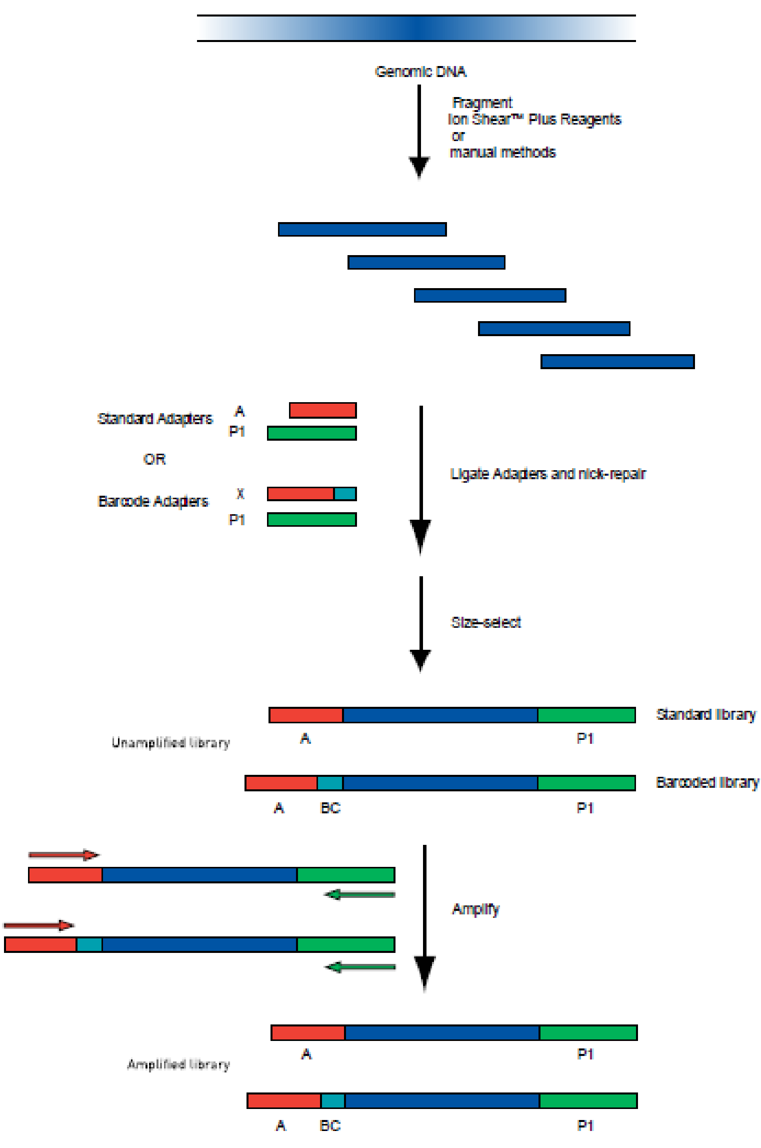
https://www.youtube.com/watch?v=WYBzbxIfuKs

https://www.thermofisher.com

Ion Torrent

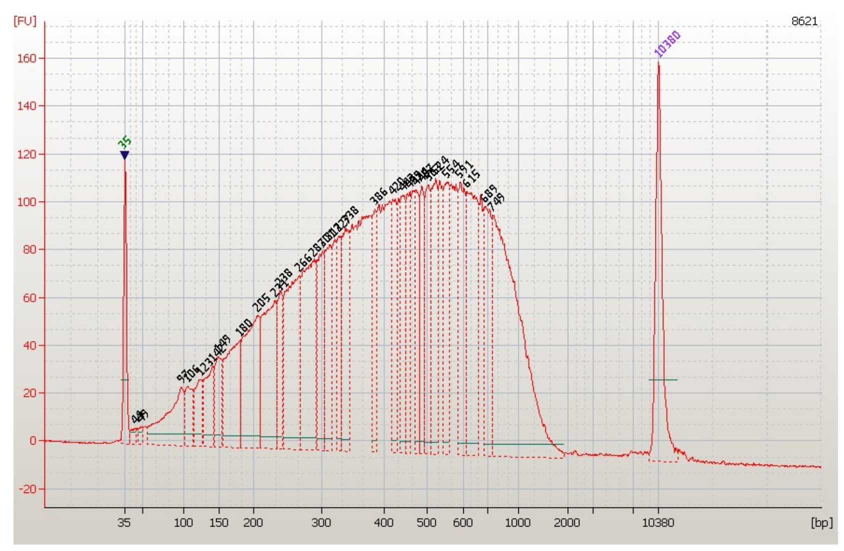


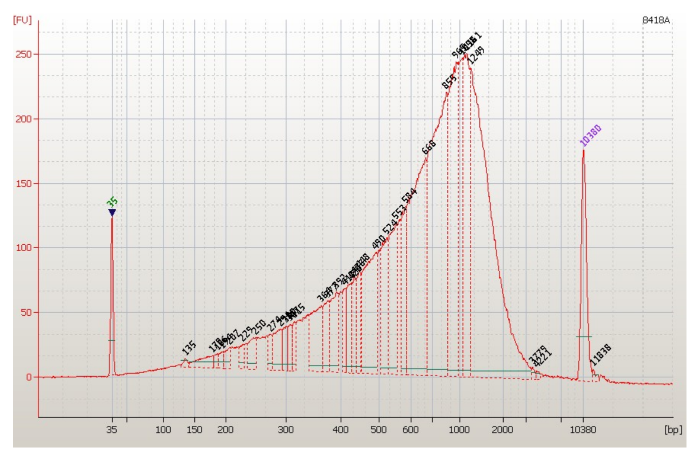
Příprava knihovny – celogenomové sekvenování

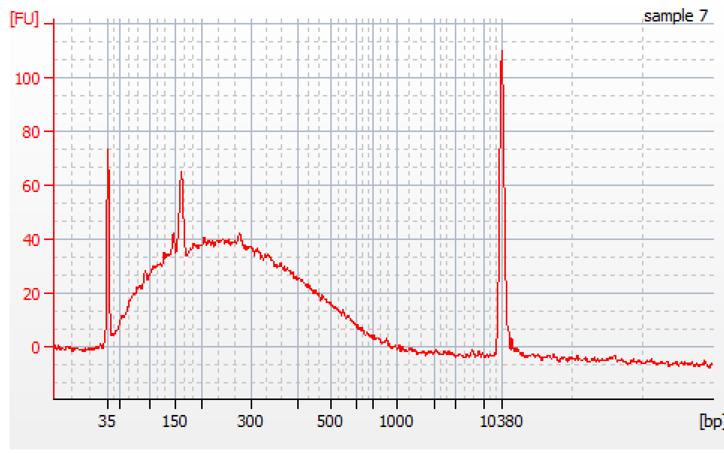


Příprava knihovny

FRAGMENTACE DNA Sonikace /enzymaticky

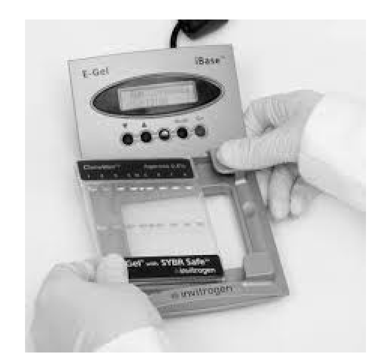






Příprava knihovny

END- REPAIR
↓
LIGACE ADAPTORŮ + NICK REPAIR
↓
SIZE SELECTION
- E-gel 2 %



4 | 50bp ladder | 5

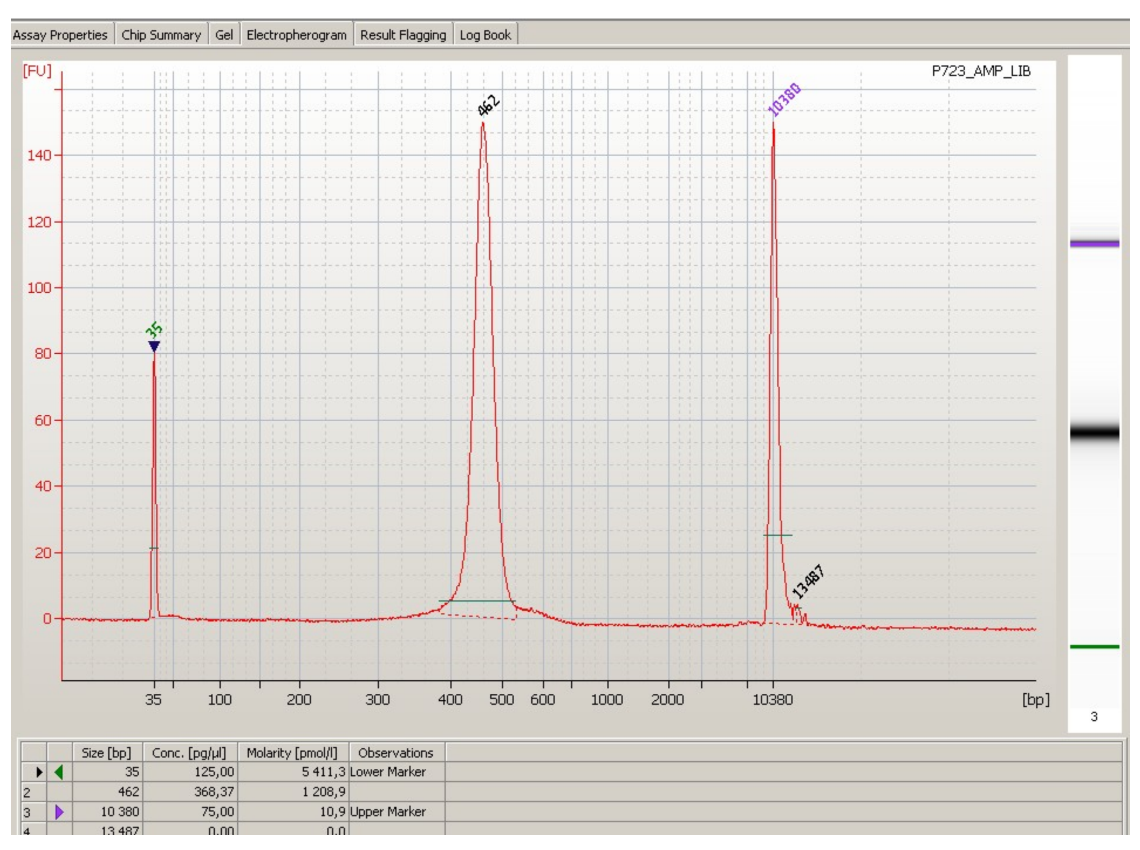
Příprava knihovny

AMPLIFIKACE KNIHOVNY (?)

↓

KVANTIFIKACE KNIHOVNY

- qPCR / <u>Agilent</u>



		Size [bp]	Conc. [pg/µl]	Molarity [pmol/l]	Observations
▶	◀	35	125,00	5 411,3	Lower Marker
2		462	368,37	1 208,9	
3	▶	10 380	75,00	10,9	Upper Marker
4		13 487	0,00	0,0	

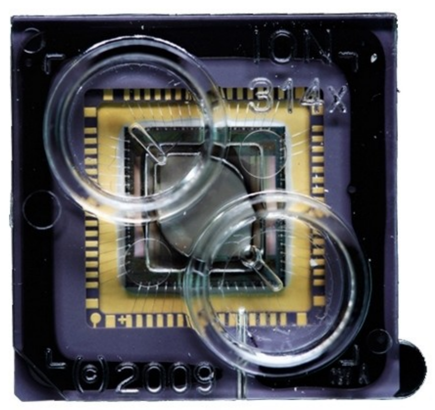
Příprava knihovny

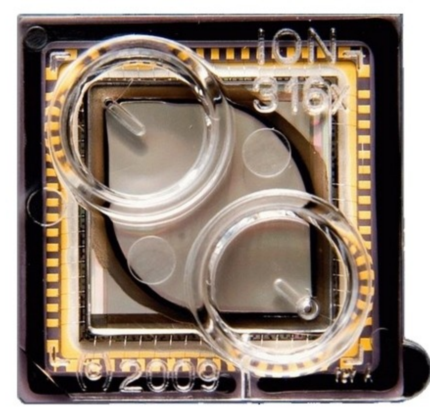
PŘÍPRAVA TEMPLÁTU

- Emulzní PCR

ENRICHMENT

SEKVENACE

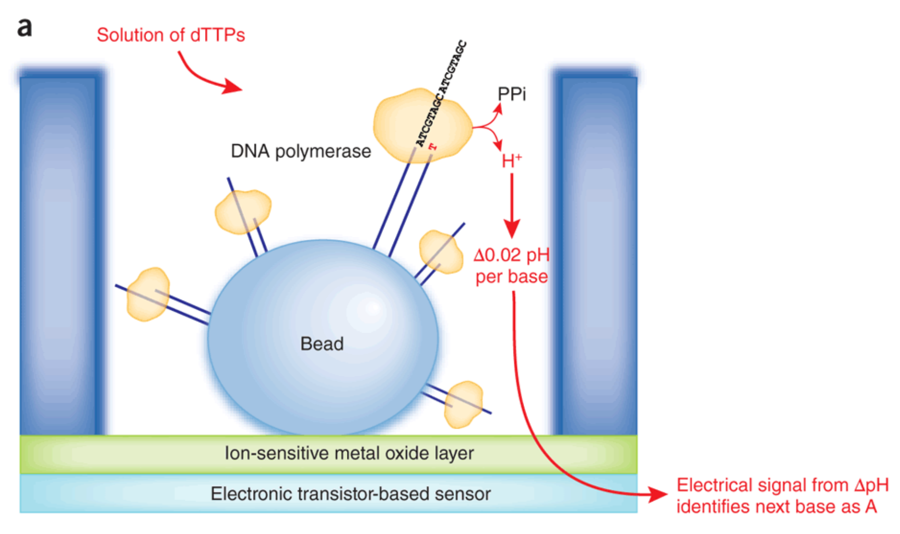


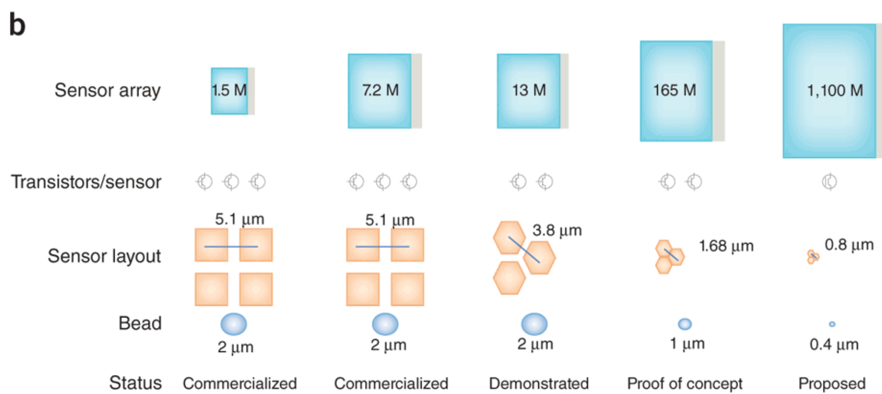


Chip	# Wells	# Reads	Throughput*	
			200 Base Read	400 Base Read
Ion 314™ Chip v2	~1.2 Million	400-500 thousand	30-50 Mb	60-100Mb
Ion 316™ Chip v2	~6 Million	2-3 million	300-600 Mb	600 Mb- 1Gb
Ion 318™ Chip v2	~11 Million	4 – 5.5 million	600 Mb- 1Gb	1.2 – 2 Gb

Ion Torrent

Ion Torrent



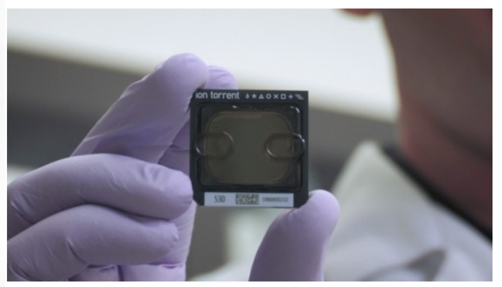


(**a**) Schematic of a well on an ion sequencing chip. Clonal DNA immobilized on a bead is extended by polymerase in the presence of a pure solution of one nucleotide (here 'T'). Nucleotide incorporation releases a pyrophosphate (PPi) and a hydrogen ion. The change in pH caused by release of the hydrogen 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.

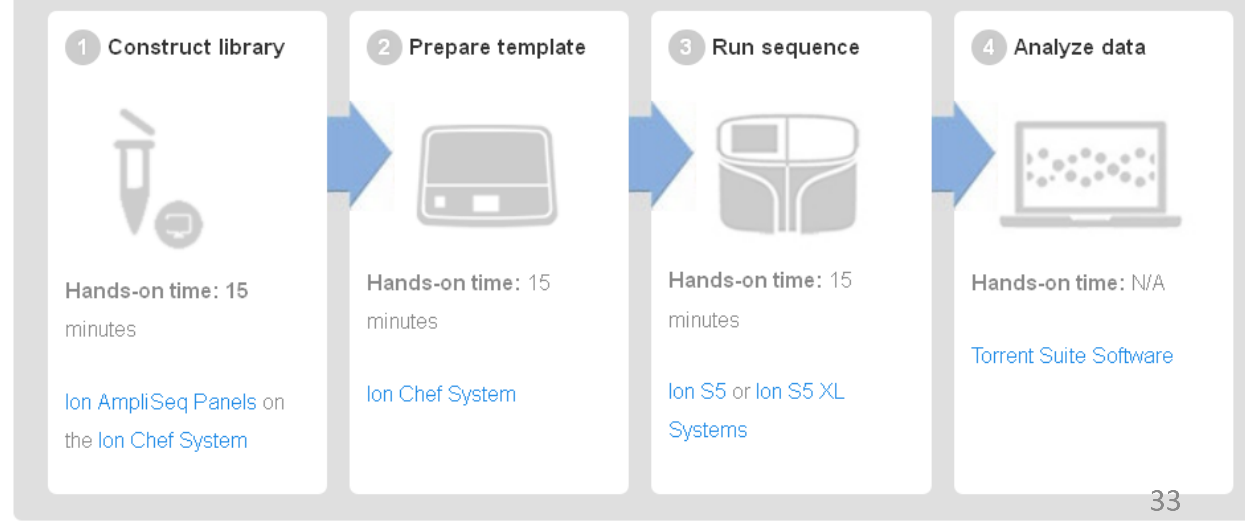
Ion Torrent

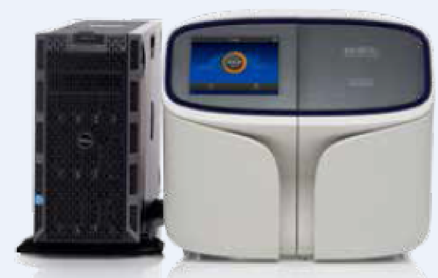
New Ion S5

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



https://www.youtube.com/watch?v=jFCD8Q6qSTM&ebc=ANyPxKrMLmAe4Nmia2N3RFr_1QbGUsOzceI2sMqnIJ5gS09XPCofTb-0cUvJdbzQhD_gKRKTL-XBahDEvoV6uOnm_78yvaG-eA



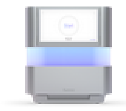
		Ion S5 System			Ion S5 XL System		
		Simple workflow for panels, microbes, exomes, and transcriptomes			 Simple, rapid workflow for panels, microbes, exomes, and transcriptomes		
		Ion 520 Chip	**Ion 530 Chip**	**Ion 540 Chip**	**Ion 520 Chip**	**Ion 530 Chip**	**Ion 540 Chip**
Reads		3–5 million	15–20 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.

Illumina

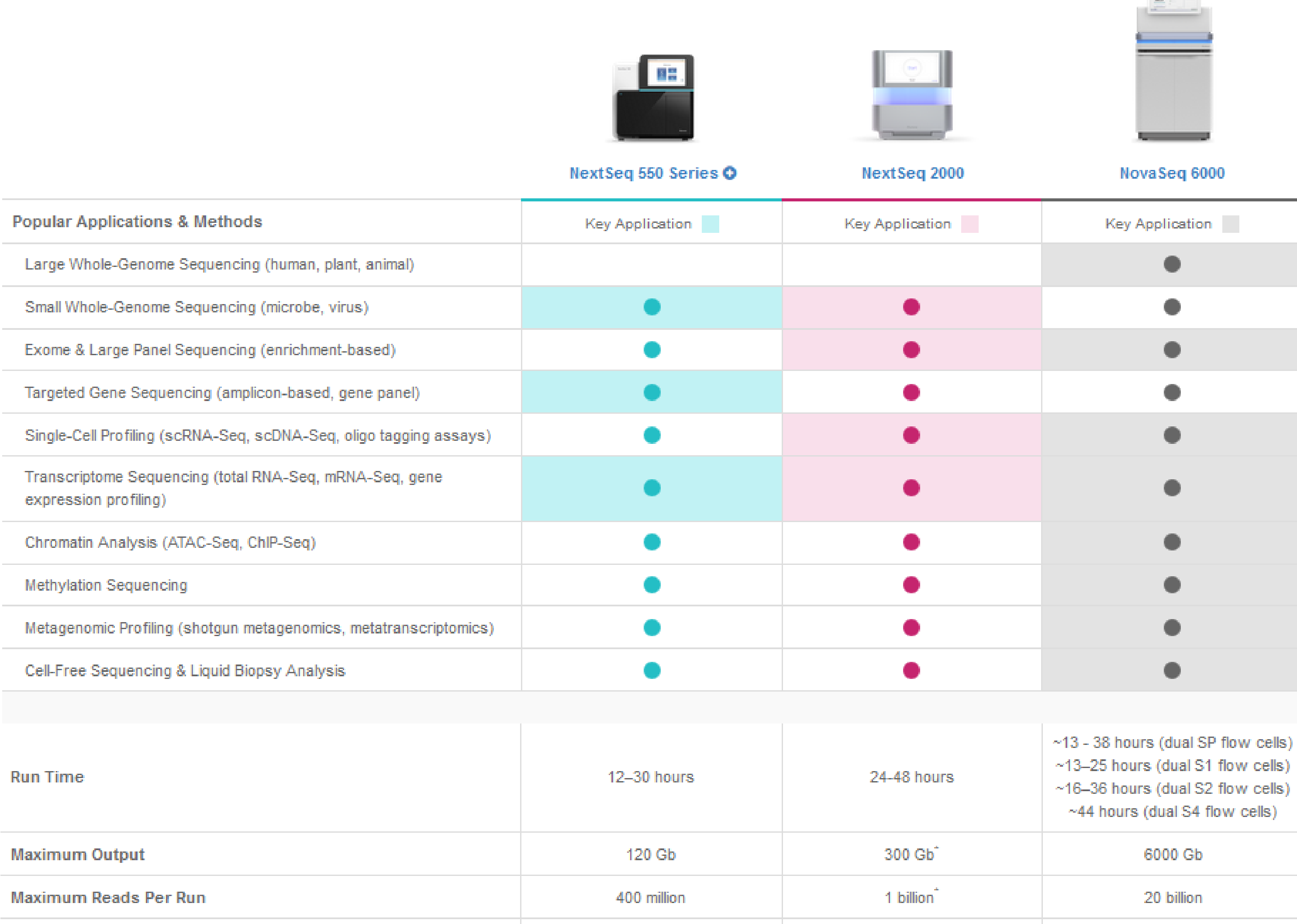
- Příprava templátu: hybridizace na sklíčku, tvorba klastrů
- Sekvenace syntézou
- Detekce fluorescence odštěpené značky z reverzního terminátoru (nukleotidu)

Přístroje Illumina



	iSeq 100	MiniSeq	MiSeq Series	NextSeq 550 Series	NextSeq 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				●	●
Run Time	9.5–19 hrs	4–24 hours	4–55 hours	12–30 hours	24-48 hours
Maximum Output	1.2 Gb	7.5 Gb	15 Gb	120 Gb	300 Gb*
Maximum Reads Per Run	4 million	25 million	25 million †	400 million	1 billion*
Maximum Read Length	2 × 150 bp	2 × 150 bp	2 × 300 bp	2 × 150 bp	2 × 150 bp

Přístroje Illumina



	NextSeq 550 Series	NextSeq 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	●	●	●
Run Time	12–30 hours	24-48 hours	~13 - 38 hours (dual SP flow cells) ~13–25 hours (dual S1 flow cells) ~16–36 hours (dual S2 flow cells) ~44 hours (dual S4 flow cells)
Maximum Output	120 Gb	300 Gb*	6000 Gb
Maximum Reads Per Run	400 million	1 billion*	20 billion
Maximum Read Length	2 × 150 bp	2 × 150 bp	2 x 250**

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

NovaSeq specifications

Sequencing Output Per Flow Cell

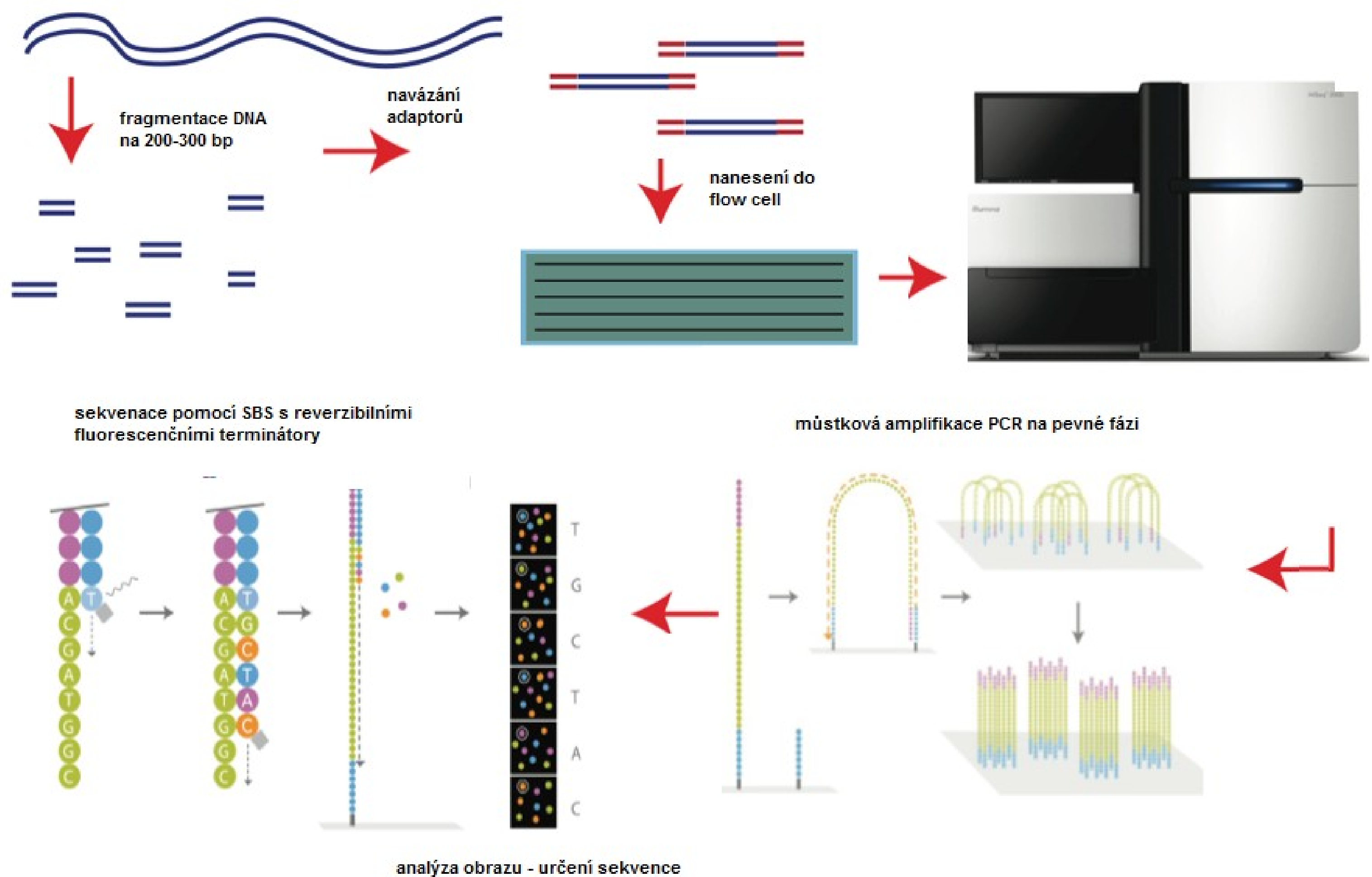
	NovaSeq 6000 System			
Flow Cell Type	SP*	S1	S2	S4
2 × 50 bp	65–80 Gb	134–167 Gb	333–417 Gb	N/A ‡
2 × 100 bp	134–167 Gb	266–333 Gb	667–833 Gb	1600–2000 Gb
2 × 150 bp	200–250 Gb	400–500 Gb	1000–1250 Gb	2400–3000 Gb
2 x 250 bp	325-400 Gb	N/A ‡	N/A ‡	N/A ‡

Specifications based on Illumina PhiX control library at supported cluster densities.
‡ N/A: not applicable

Reads Passing Filter Per Flow Cell

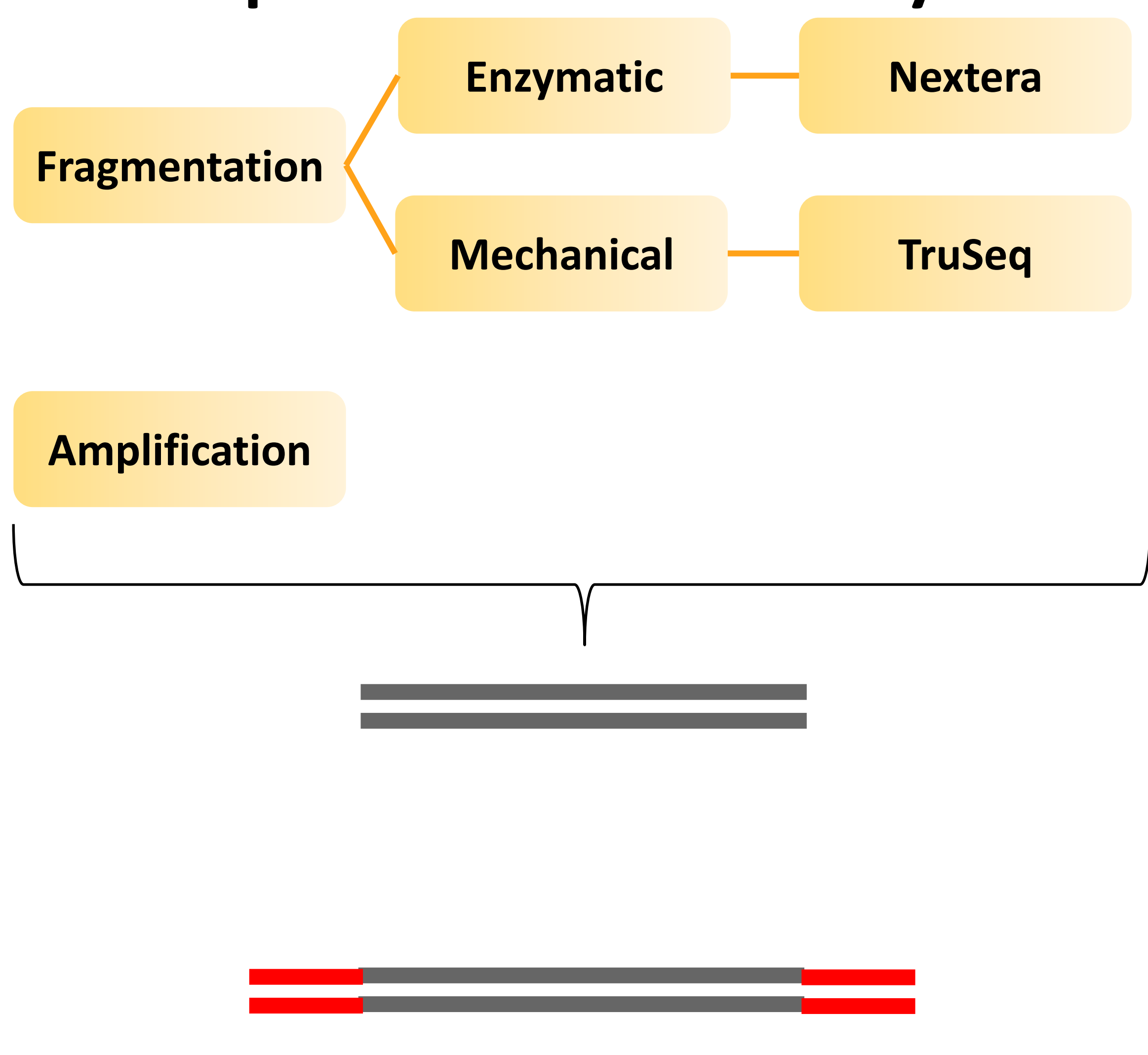
	NovaSeq 6000 System			
Flow Cell Type	SP	S1	S2	S4
Single-end Reads	650–800 M	1.3–1.6 B	3.3 B–4.1 B	8-10 B
Paired-end Reads	1.3–1.6 B	2.6–3.2 B	6.6–8.2 B	16–20 B

Illumina



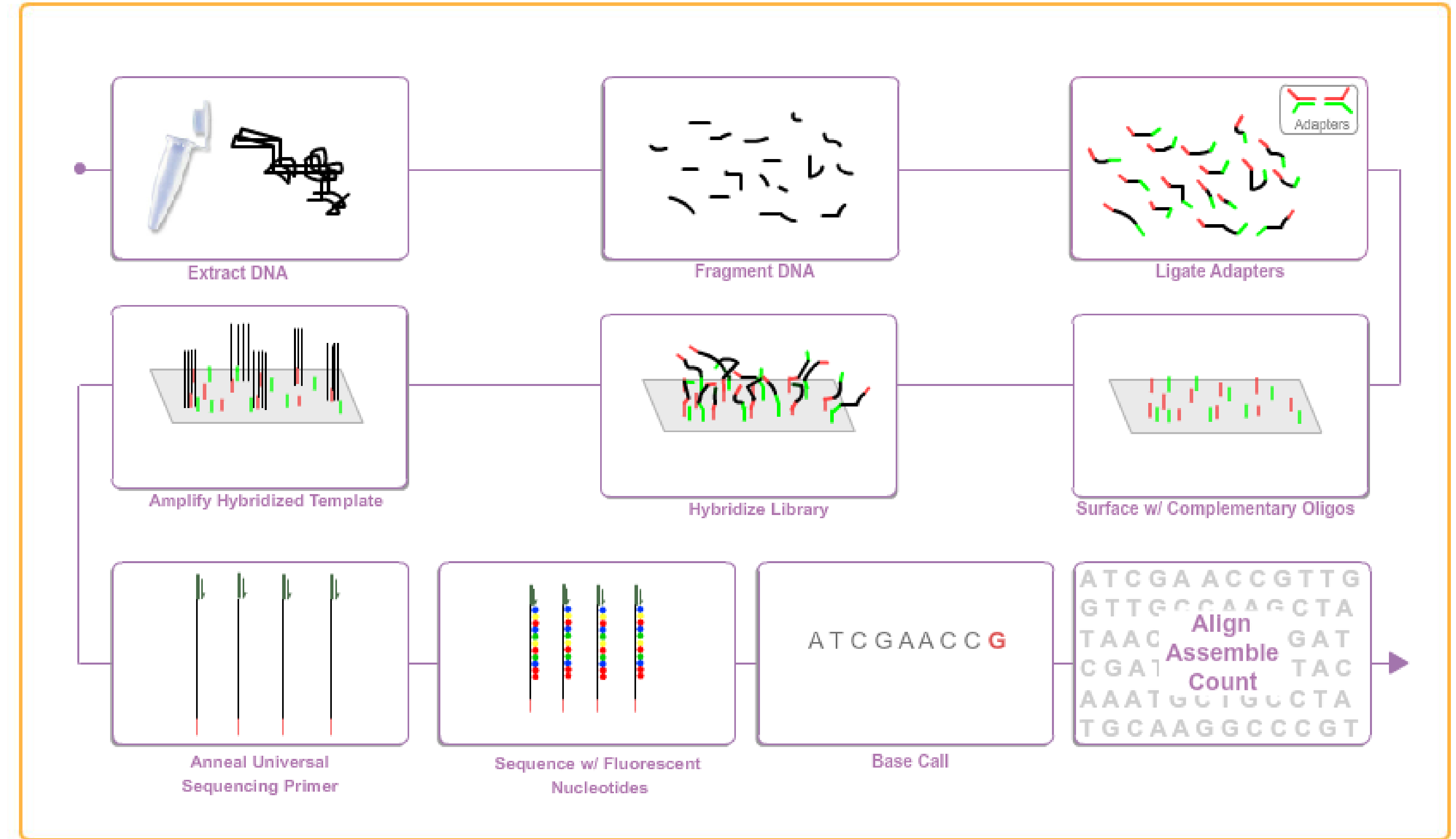
Příprava knihovny

Fragment



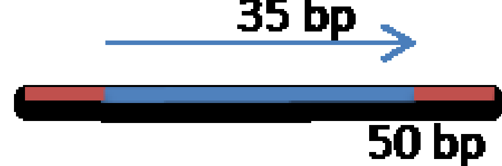
Adapter

Sekvenační technologie



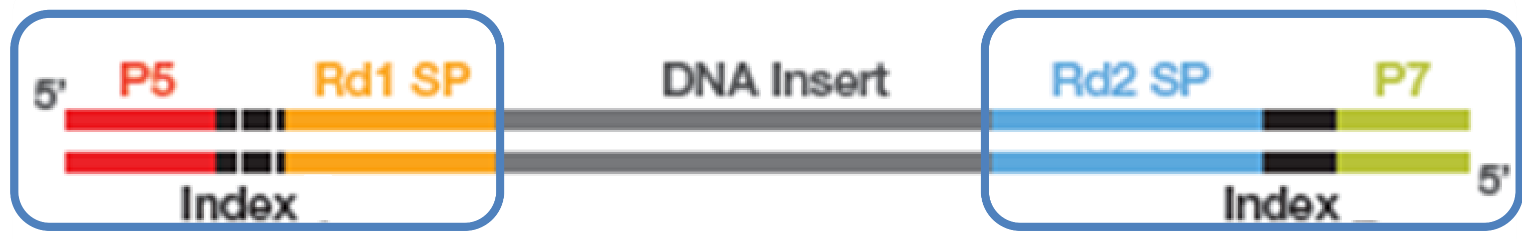
Single vs Pair End Ready

- Single reads
  - Small RNA



- Pair-End reads
  - DeNovo assembling

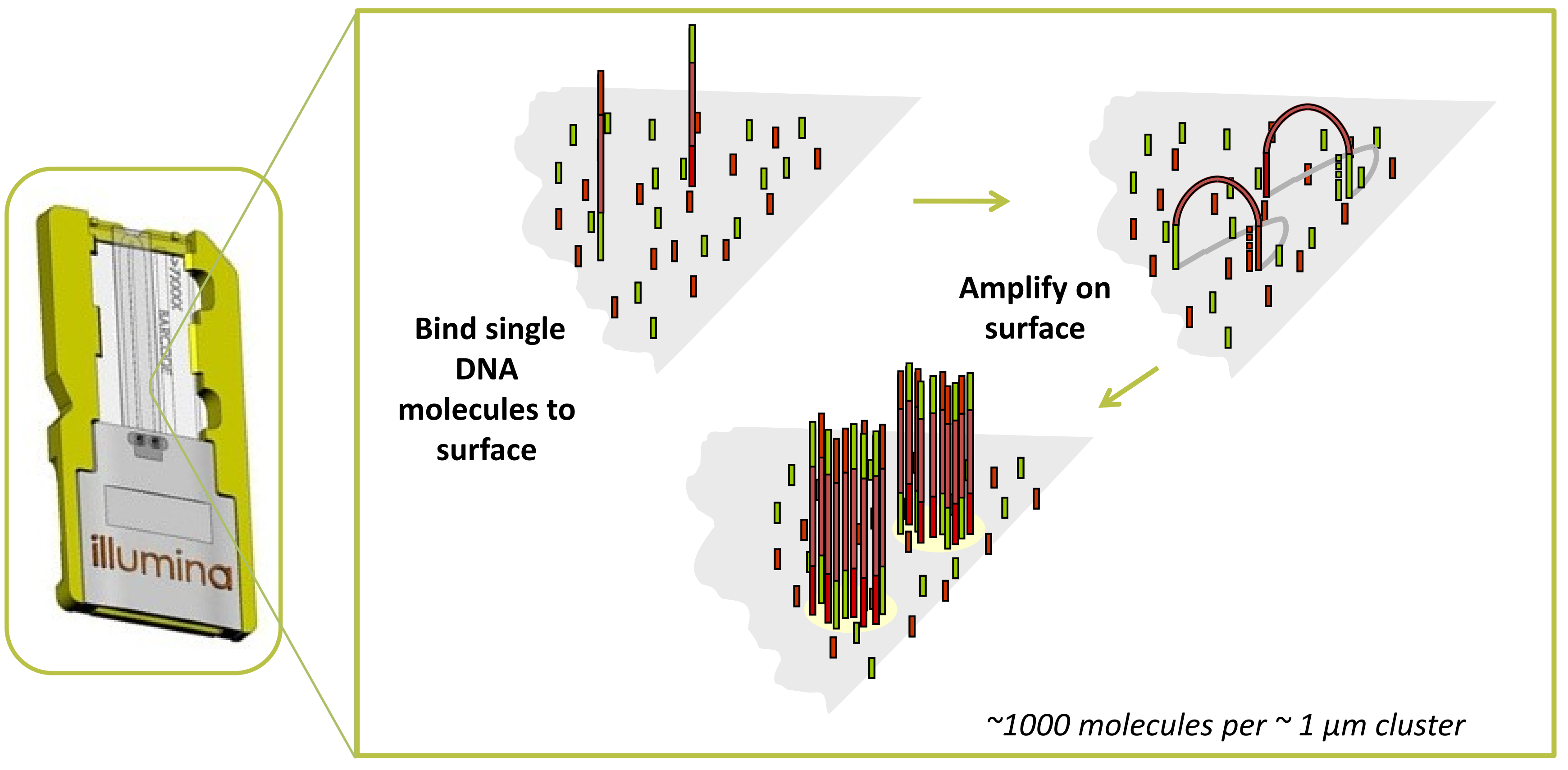
Sample Prep



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



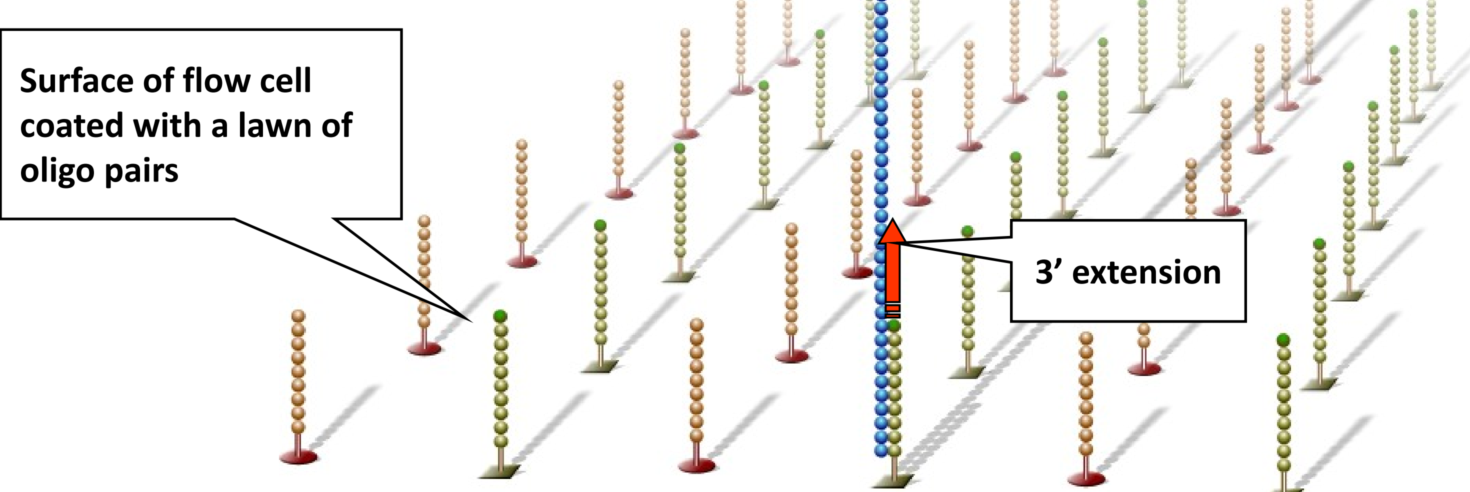
Skip Overview

Hybridize Fragment & Extend

Single DNA libraries are hybridized to primer lawn

Bound libraries then extended by polymer

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



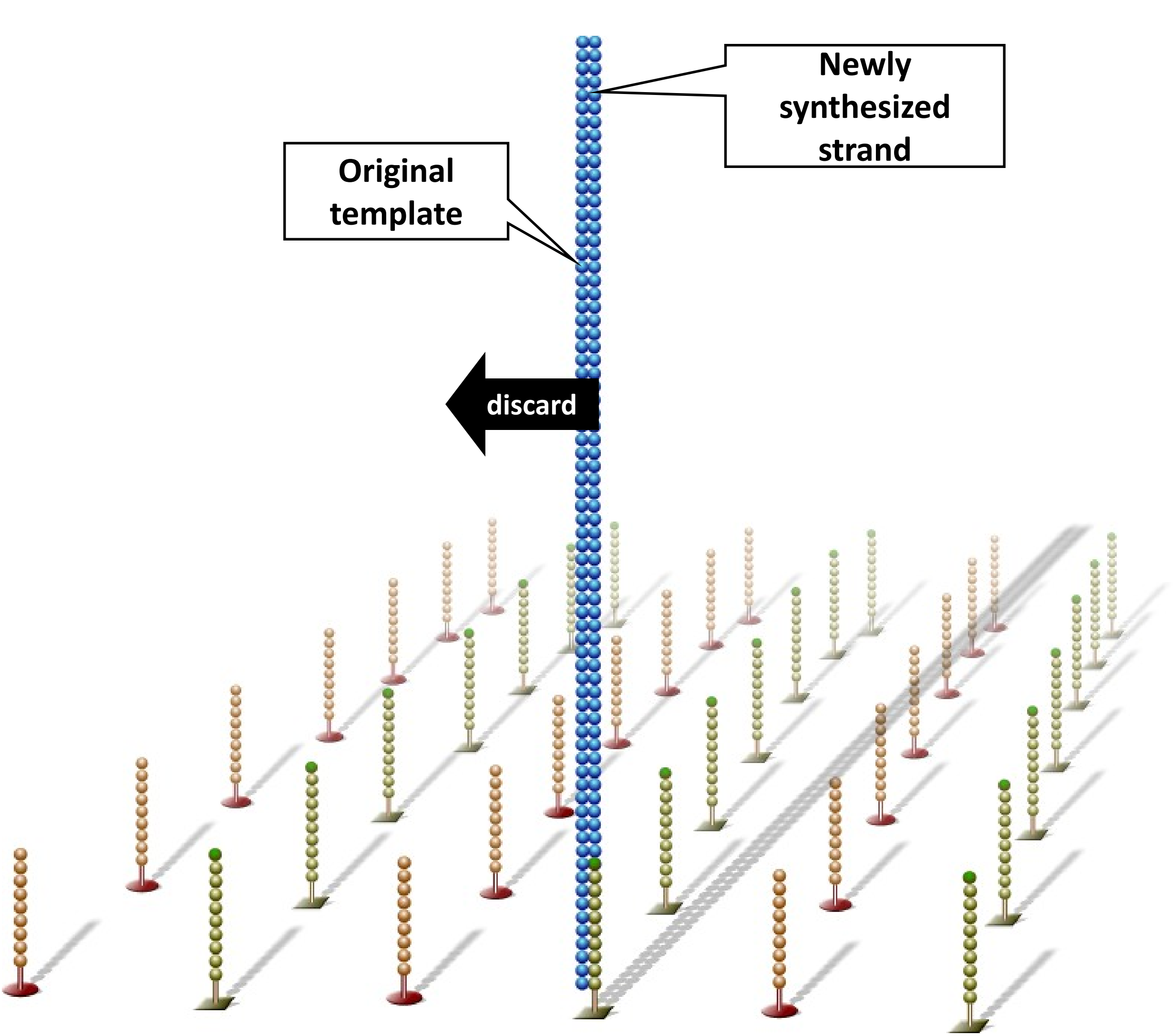
Denature Double-Stranded DNA

Double-stranded molecule is denatured

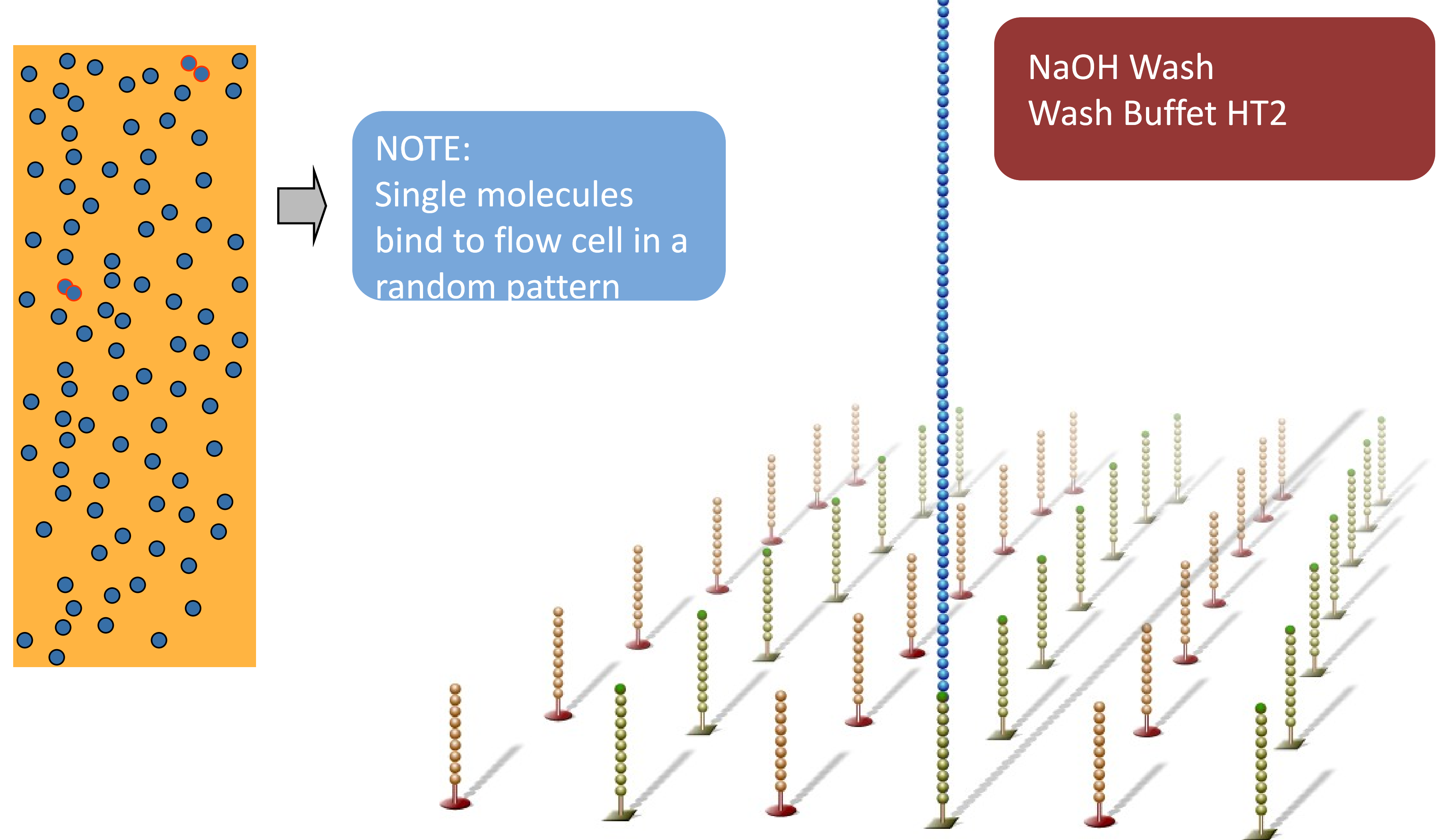
Original template washed away

Newly synthesized strand is covalently attached to flow cell surface

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



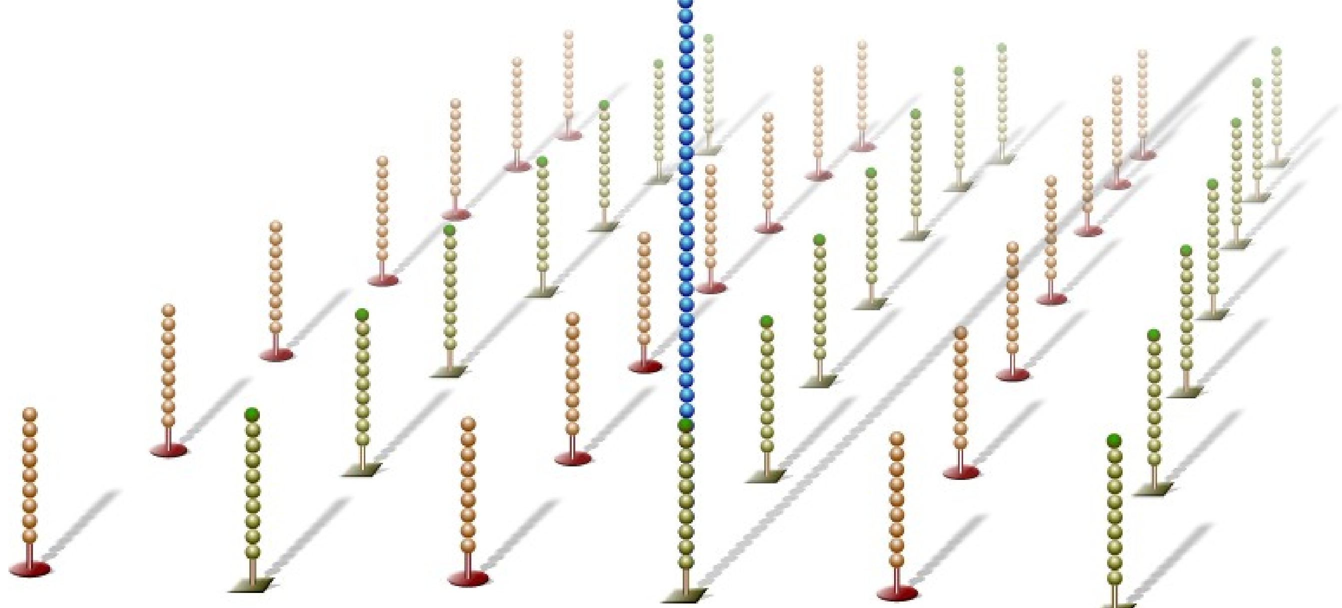
Hybridize Fragment & Extend



Hybridize Fragment & Extend

1st cycle denaturation

AMP Manifold
Ramp to 60°C
Formamide AT1

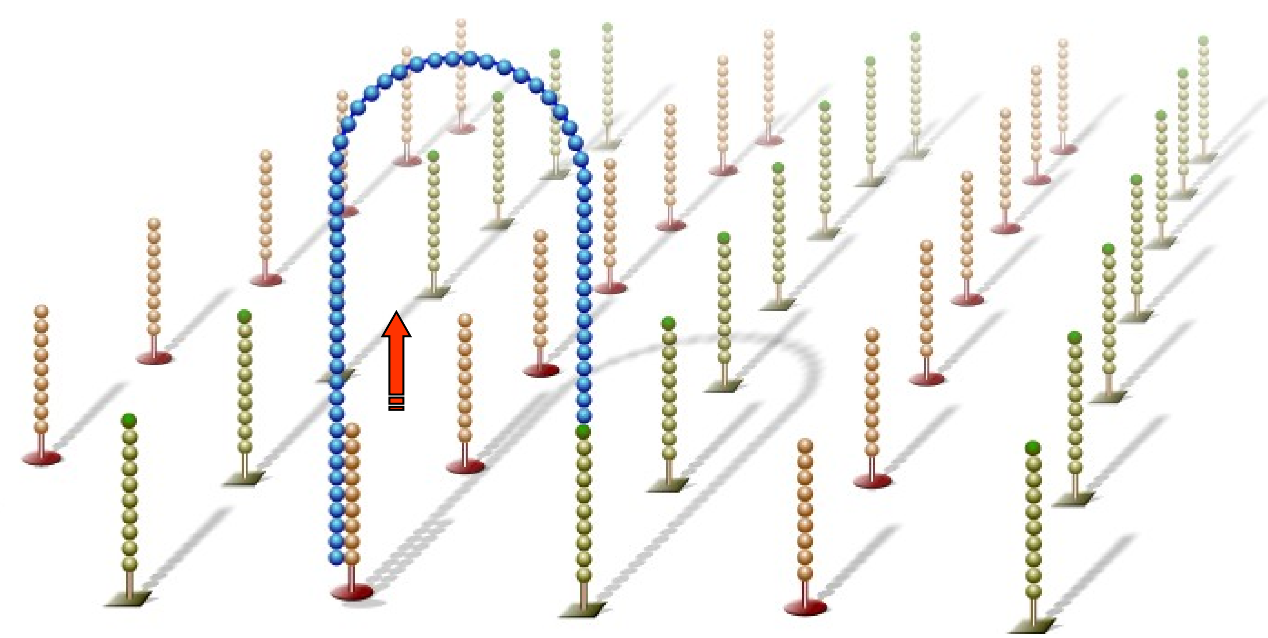


Bridge Amplification

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

Hybridized primer is extended by polyme

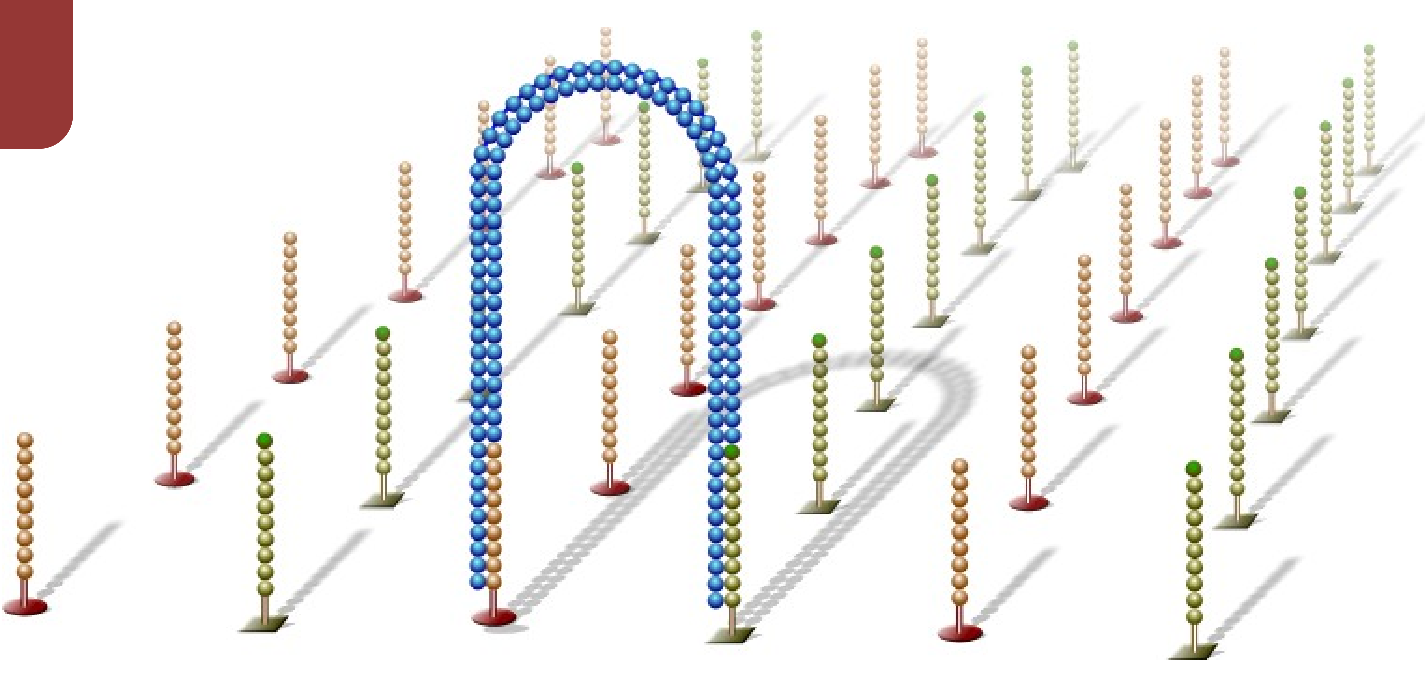
AMP Premix AMP1



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 templa

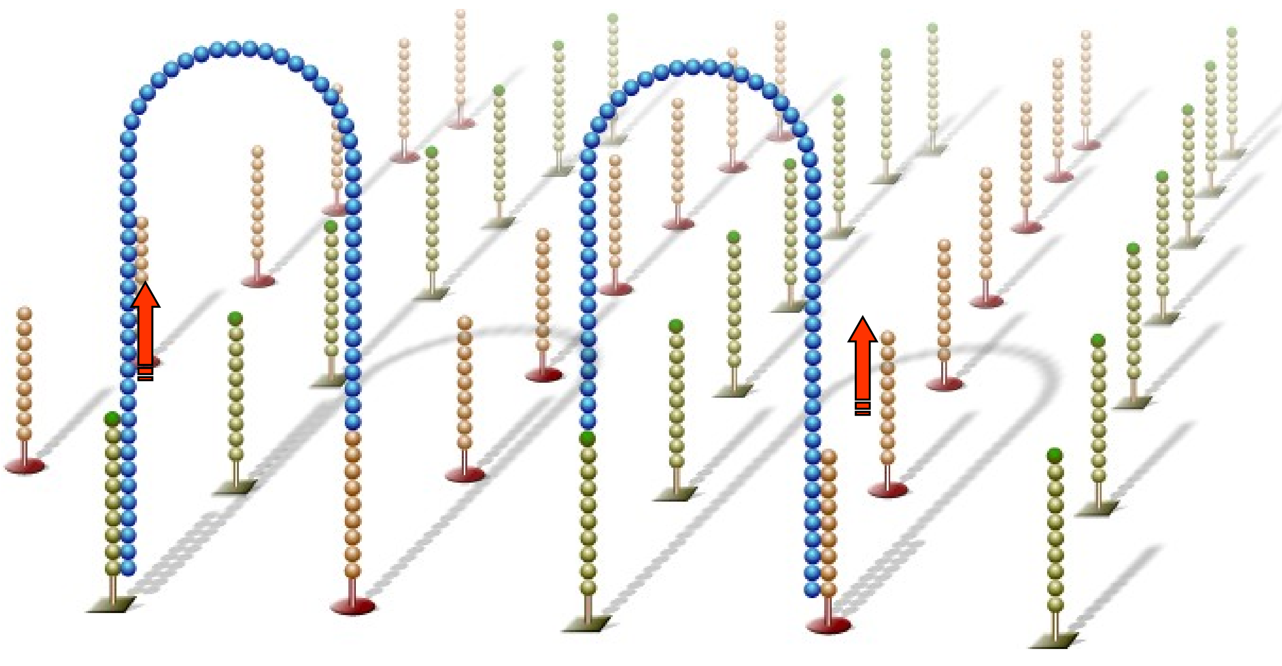
Formamide AT1

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

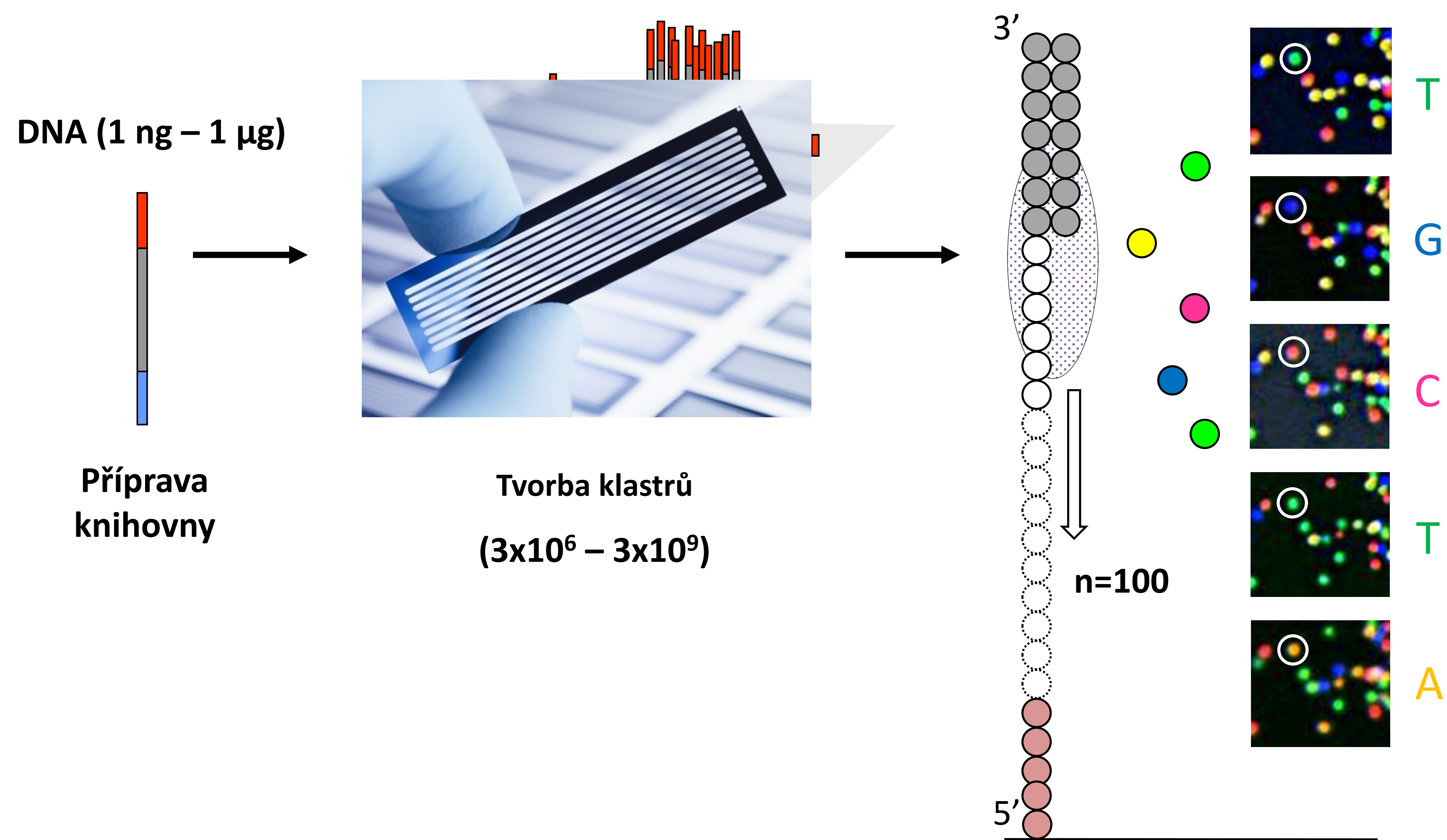
Read 1 Primer Hybridization

Sequencing primer is hybridized to adapter seque

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

Sequencing primer

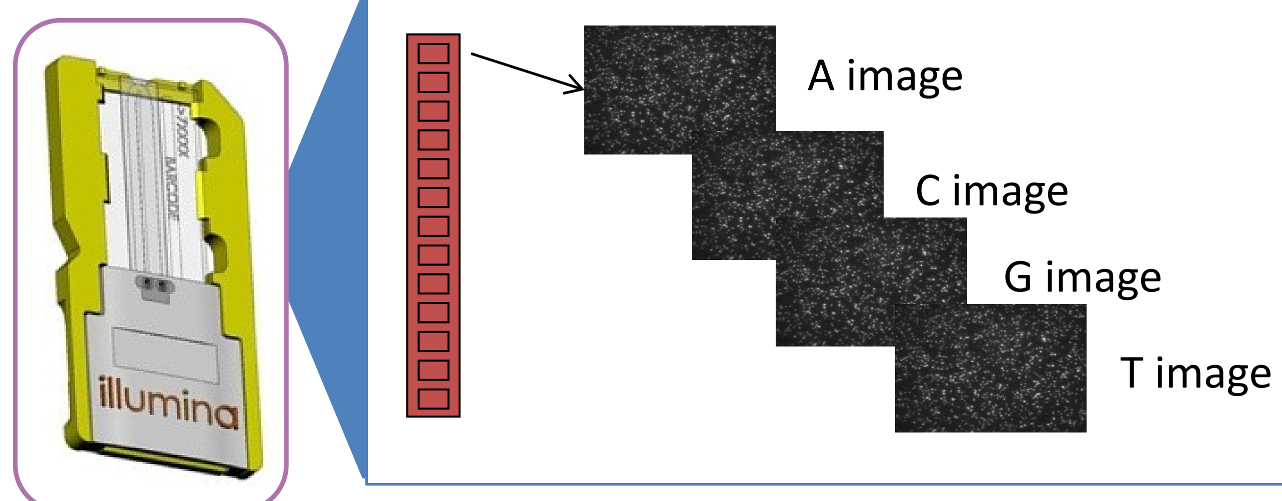
Sequencing by sythesis (SBS)



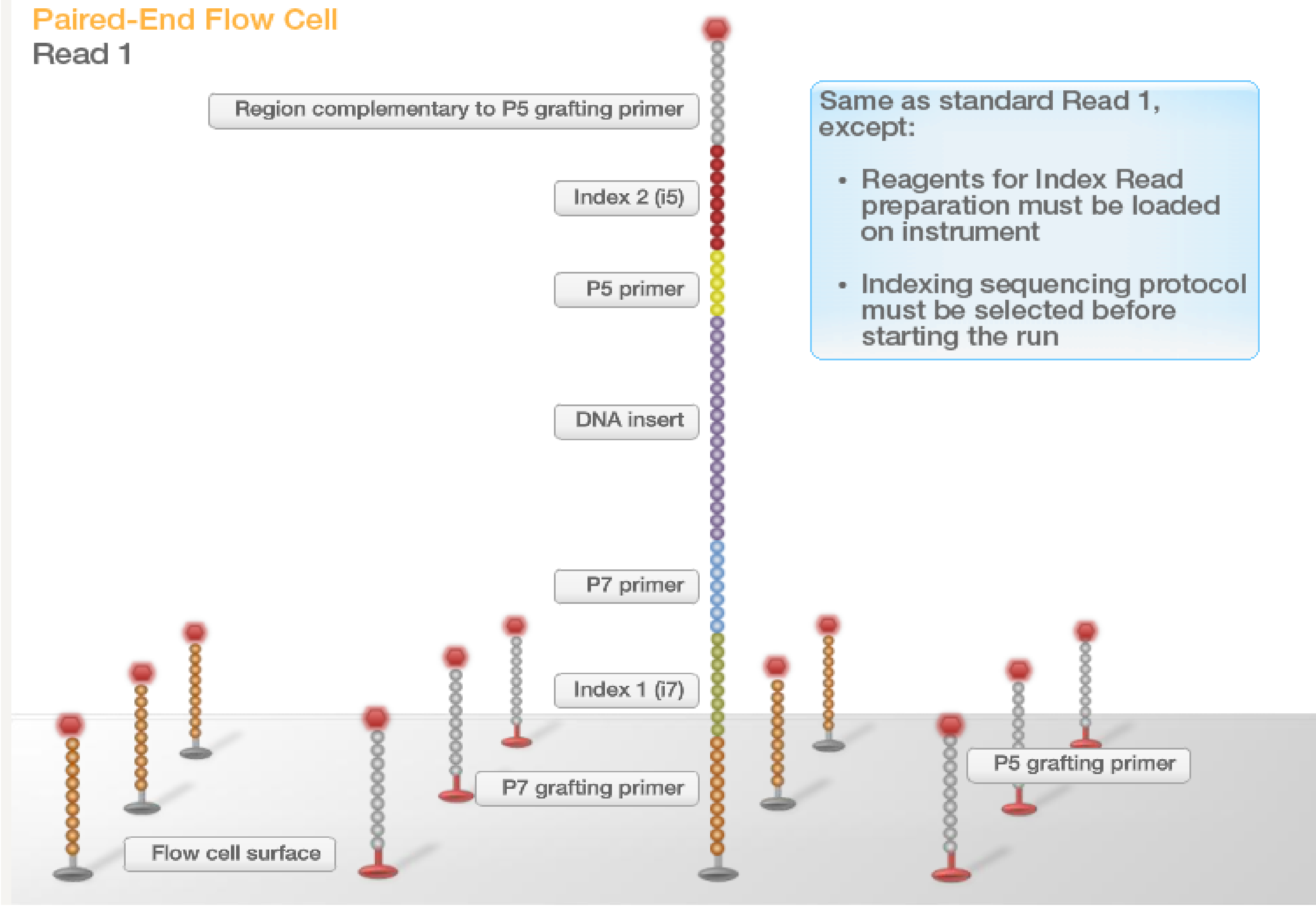
Sequencing

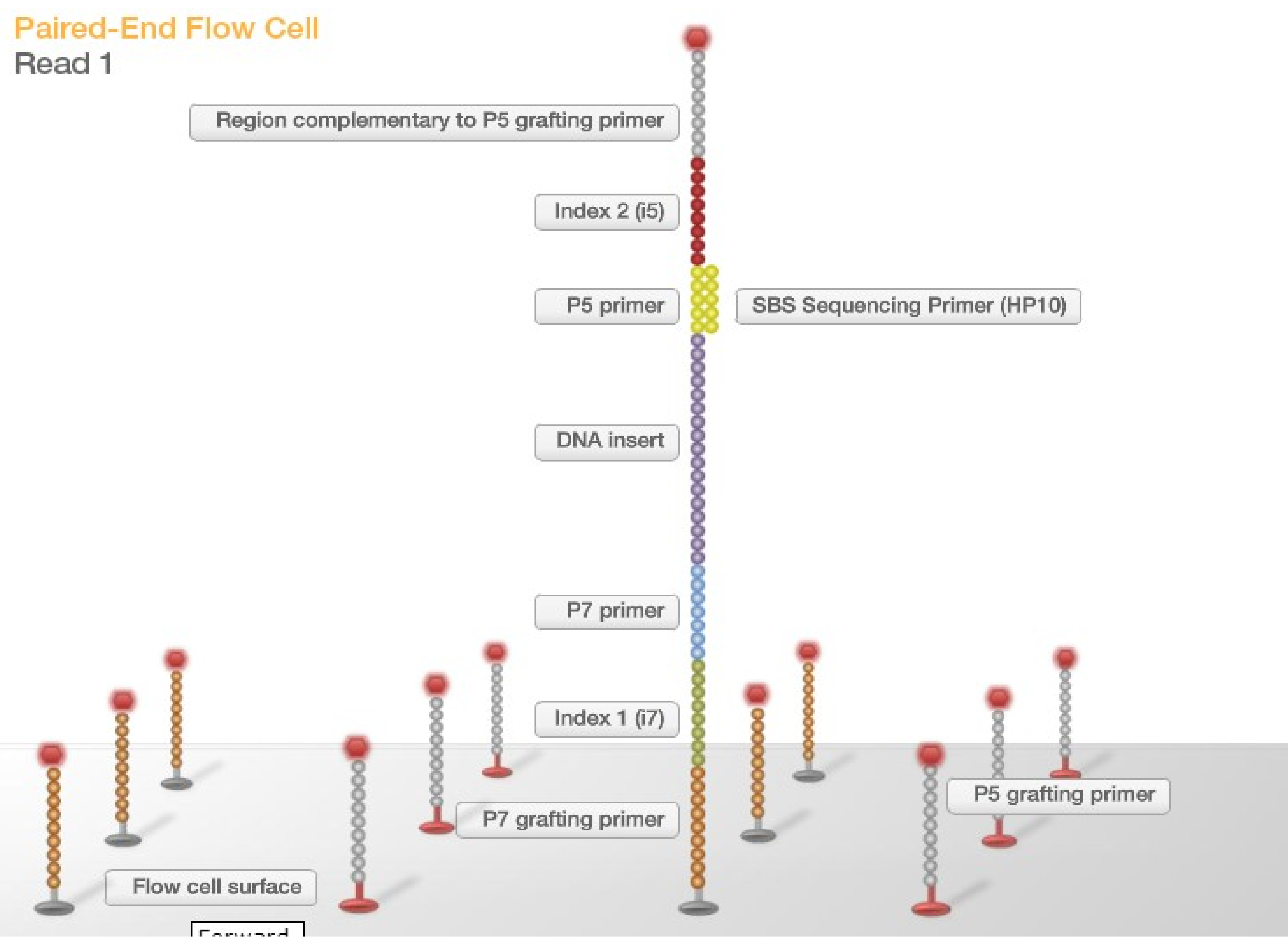
Clusters are images using LED and filter combinations specific for each fluorescently-labeled nucleotide

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

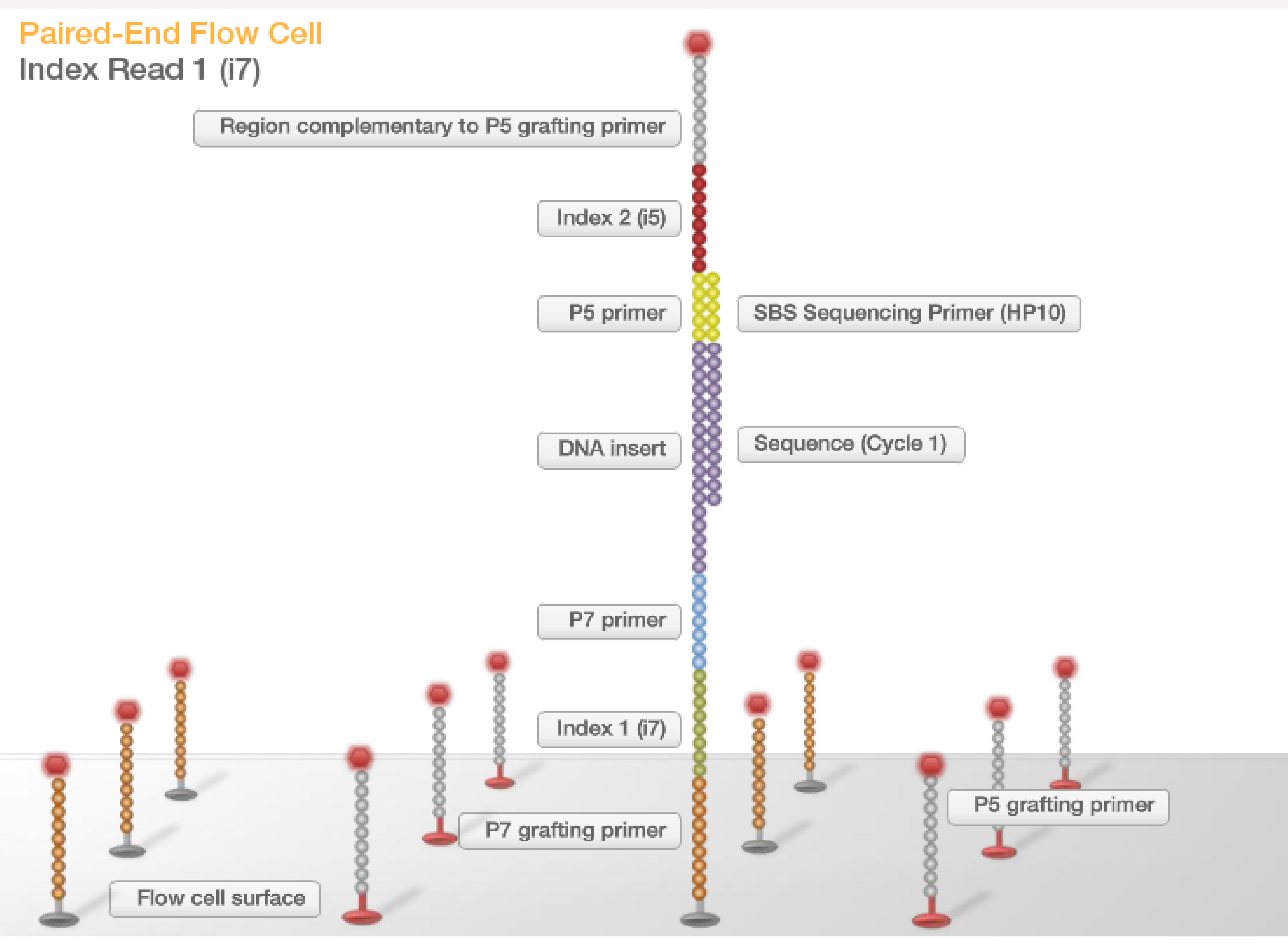


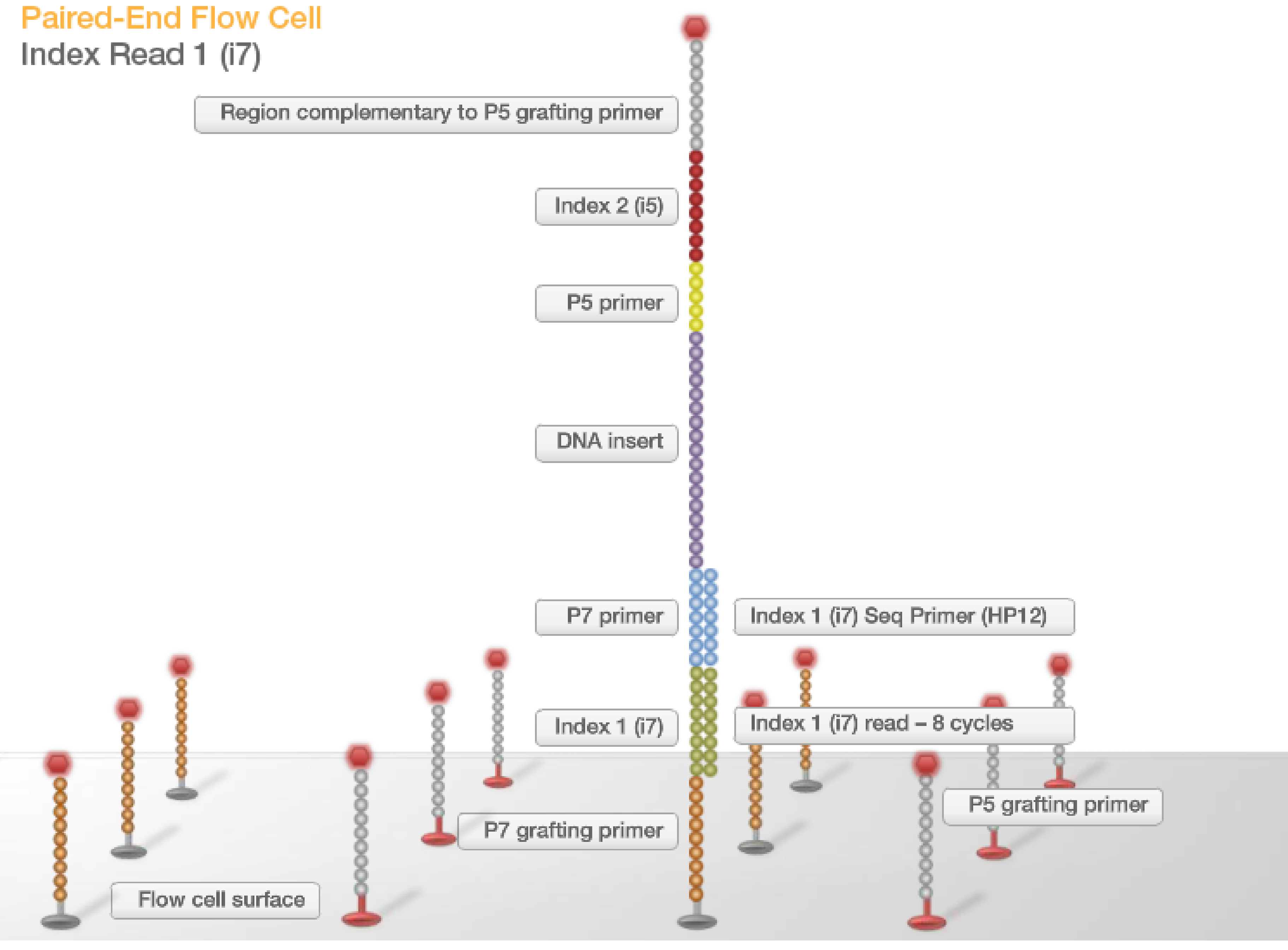
Pair - End Sequencing – Dual Indexed

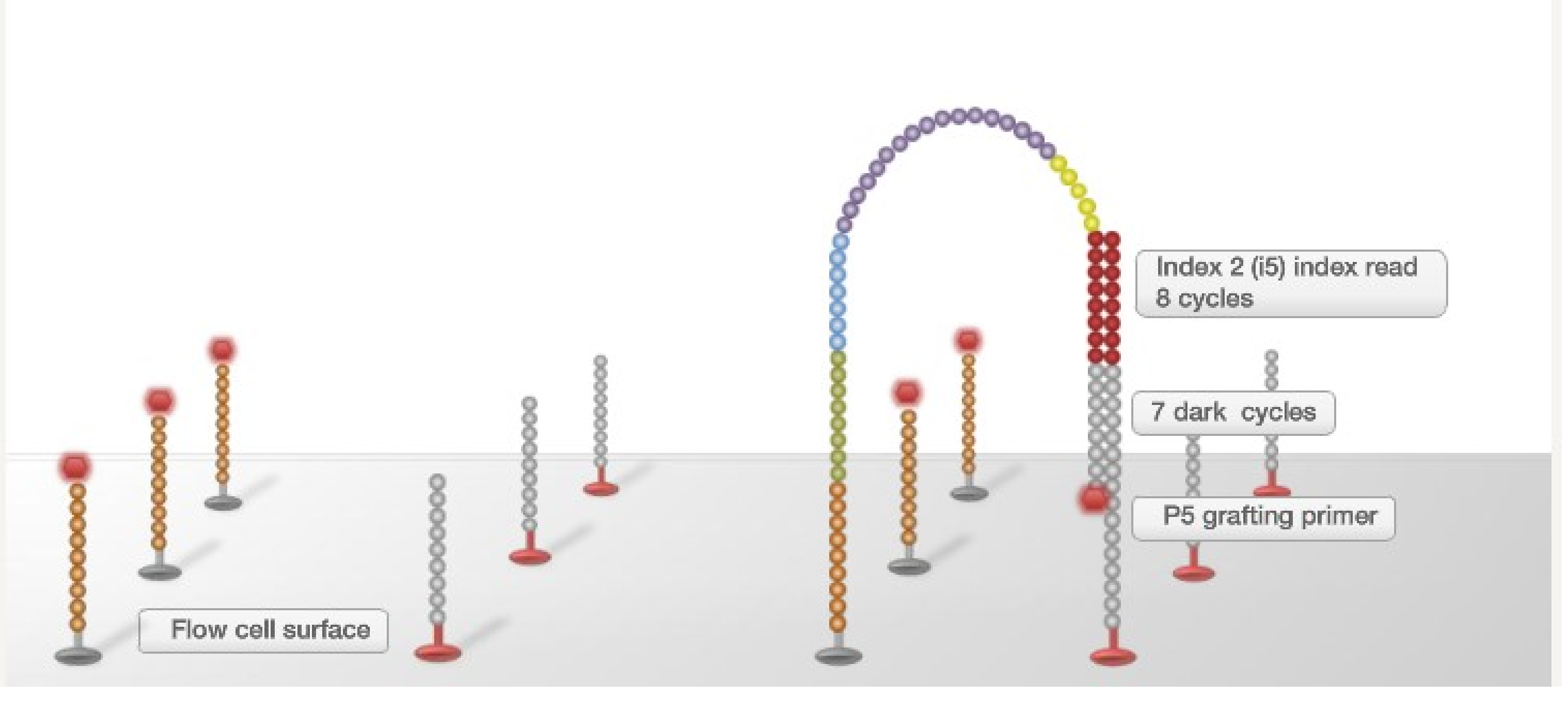


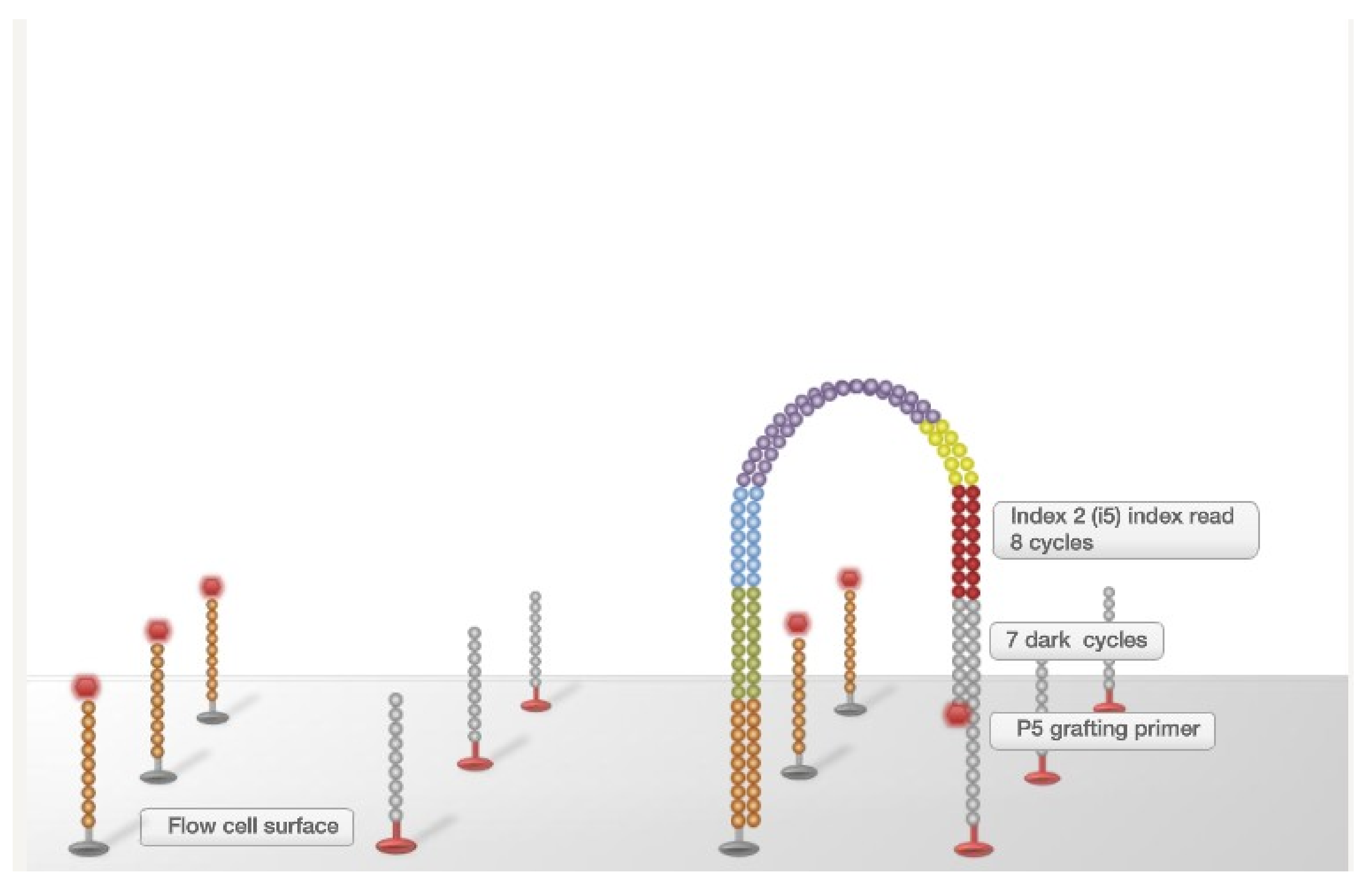


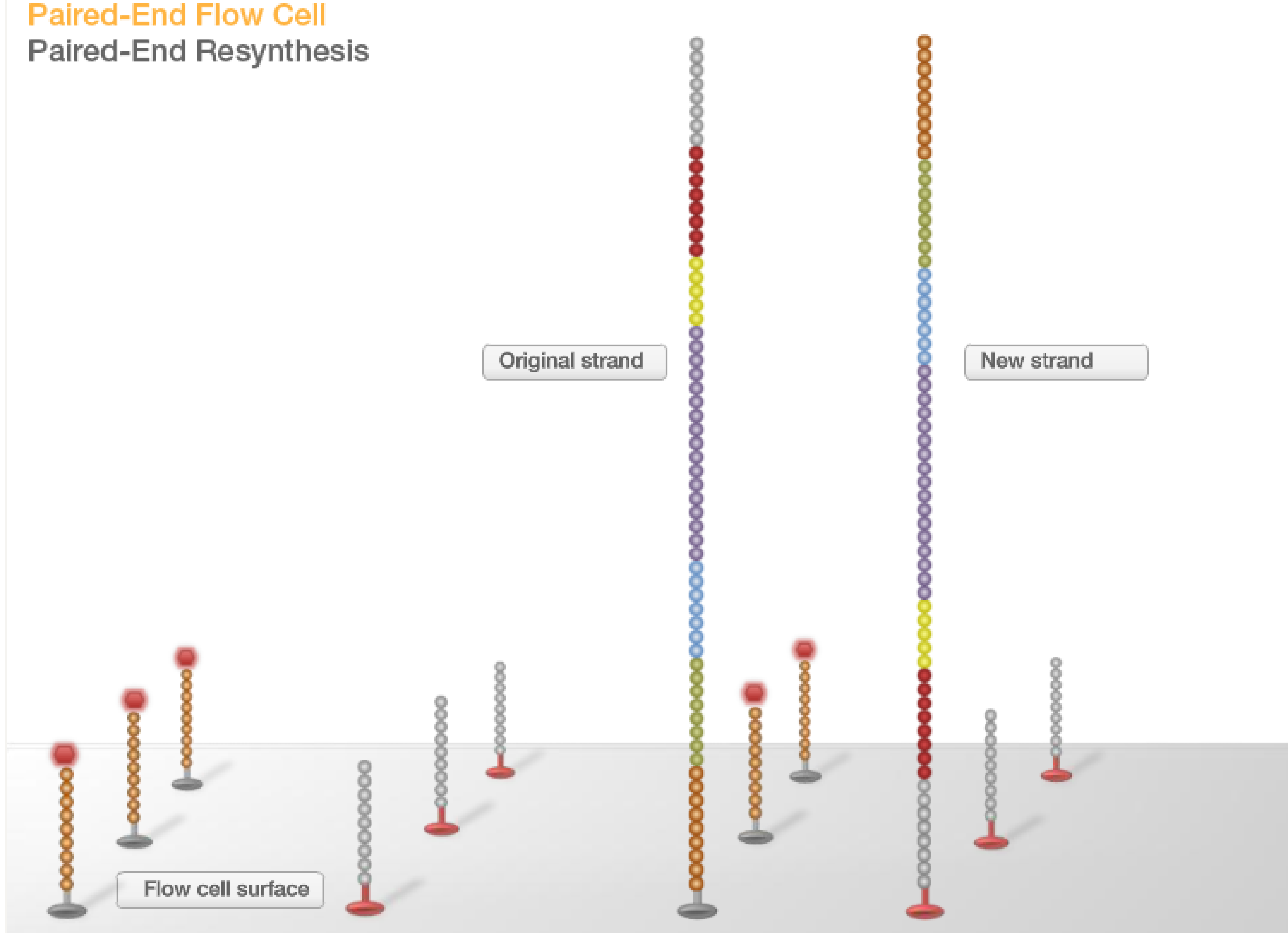
Paired-End Flow Cell
Read 1
Region complementary to P5 grafting primer
Index 2 (i5)
P5 primer
SBS Sequencing Primer (HP10)
DNA insert
P7 primer
Index 1 (i7)
P7 grafting primer
P5 grafting primer
Flow cell surface


Paired-End Flow Cell
Index Read 1 (i7)
Region complementary to P5 grafting primer
Index 2 (i5)
P5 primer
SBS Sequencing Primer (HP10)
DNA insert
Sequence (Cycle 1)
P7 primer
Index 1 (i7)
P5 grafting primer
P7 grafting primer
Flow cell surface

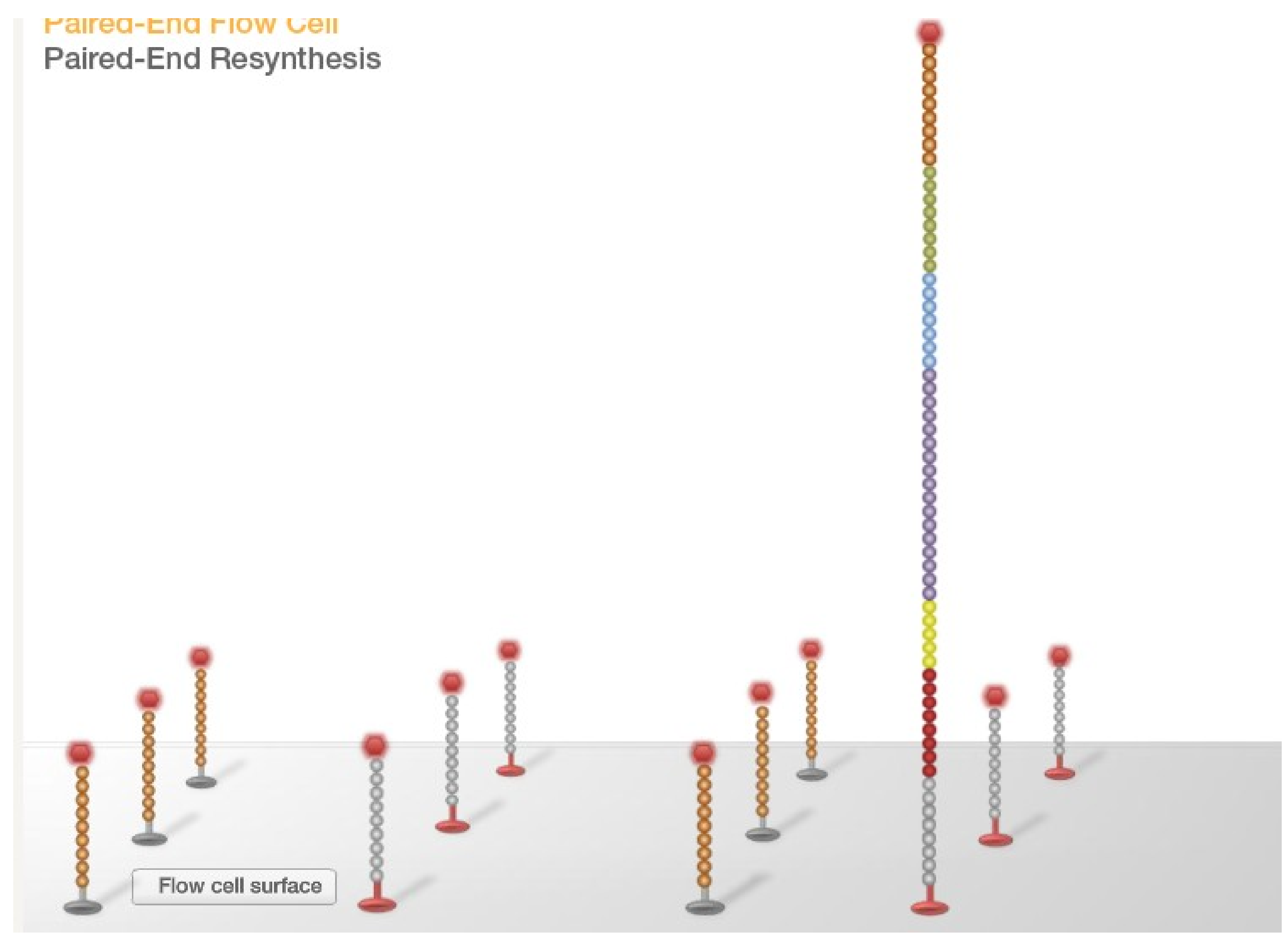

Paired-End Flow Cell
Index Read 1 (i7)
Region complementary to P5 grafting primer
Index 2 (i5)
P5 primer
DNA insert
P7 primer
Index 1 (i7) Seq Primer (HP12)
Index 1 (i7)
Index 1 (i7) read – 8 cycles
P5 grafting primer
P7 grafting primer
Flow cell surface

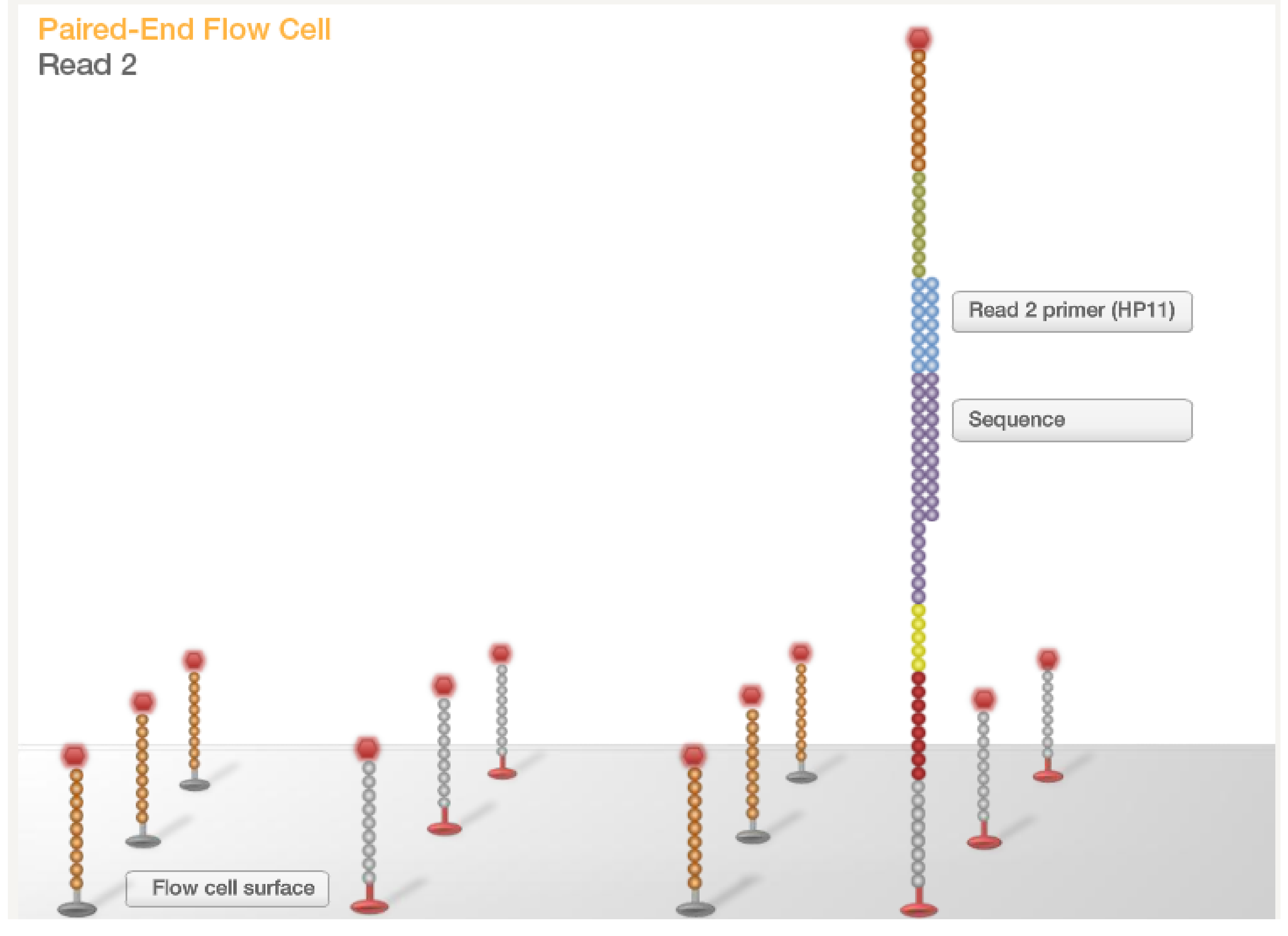

Index 2 (i5) index read
8 cycles
7 dark cycles
P5 grafting primer
Flow cell surface


Index 2 (i5) index read
8 cycles
7 dark cycles
P5 grafting primer
Flow cell surface



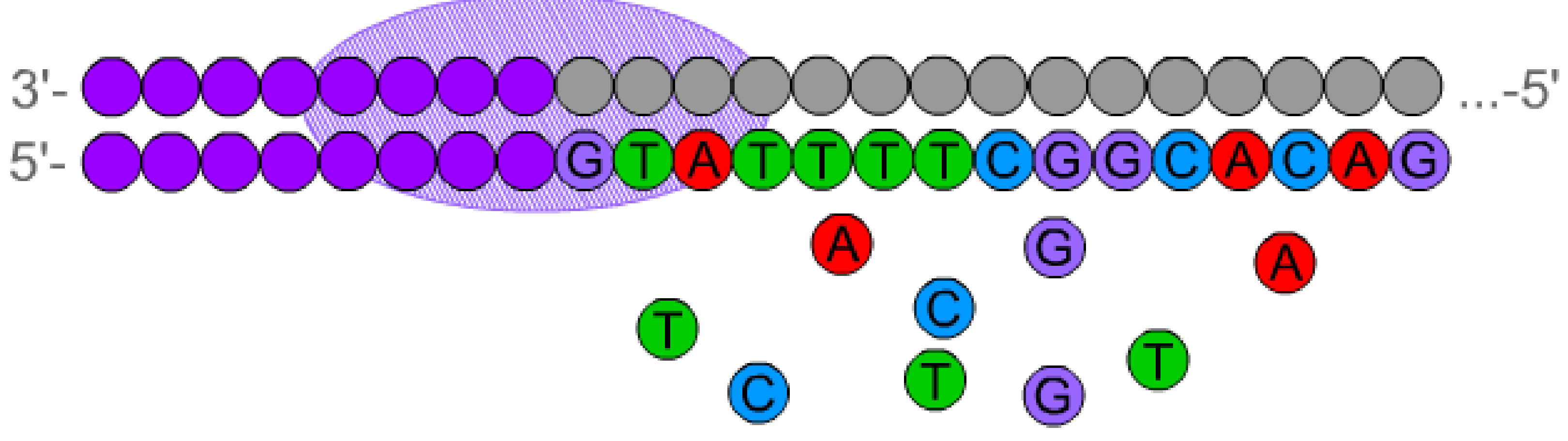
Paired-End Flow Cell
Paired-End Resynthesis
Original strand
New strand
Flow cell surface


Paired-End Flow Cell
Paired-End Resynthesis
Flow cell surface



Paired-End Flow Cell
Read 2
Read 2 primer (HP11)
Sequence
Flow cell surface

Sekvenační technologie

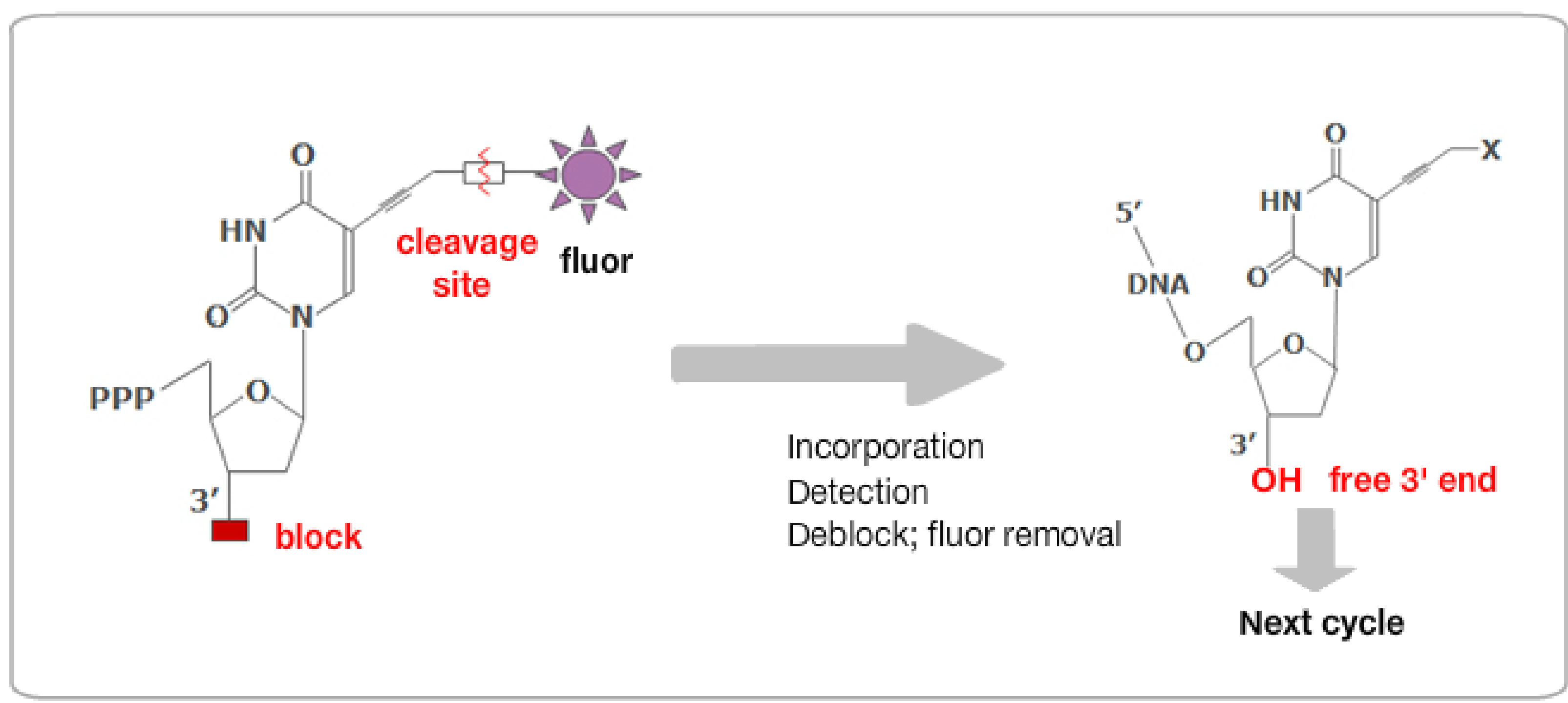


Cycle 1:	Add sequencing reagents First base incorporated Remove unincorporated bases Detect signal, deblock and defluor
Cycle 2-n:	Add Sequencing reagents and repeat

Sekvenační technologie

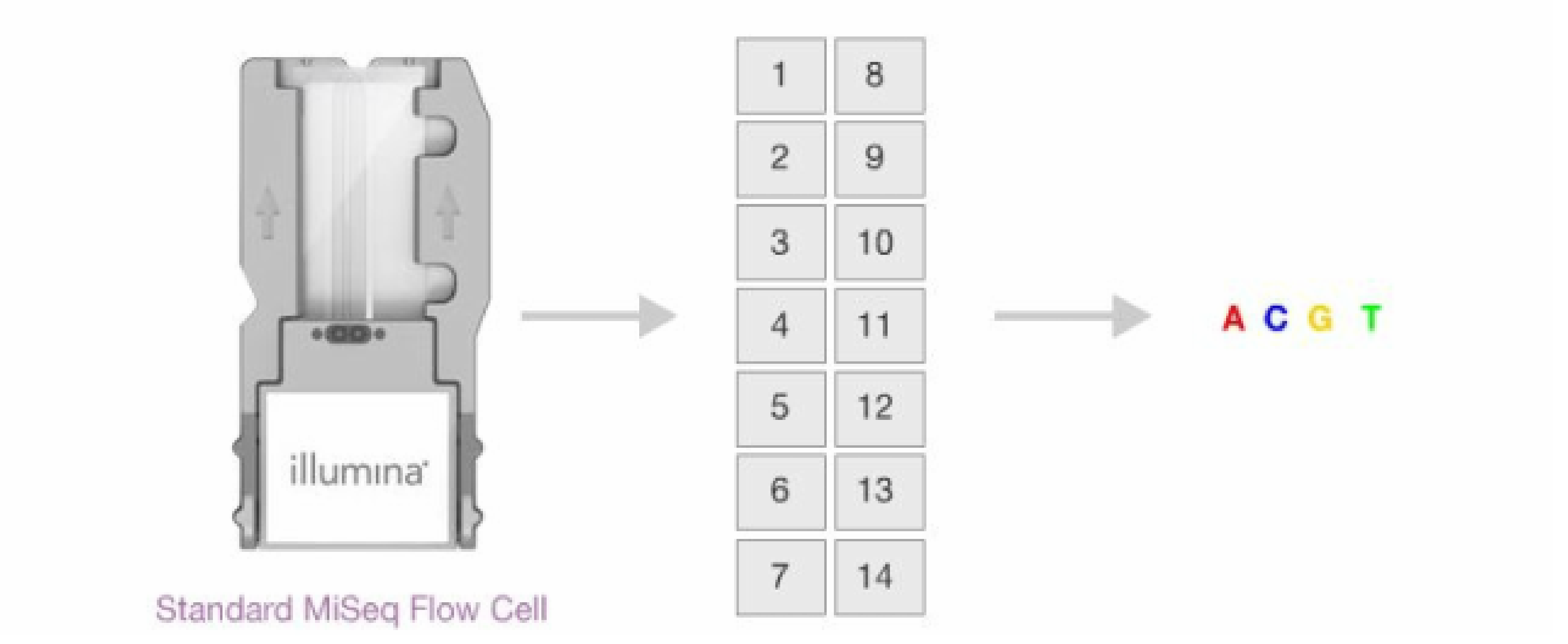
The MiSeq sequences the DNA clusters using Illumina's Sequencing By Sythesis (SBS) Chemistry which relies on Reversible Terminator Chemistry (RTC).

- All 4 labeled nucleotides in 1 reaction
- Higher accuracy



Images Generated on the Instrument

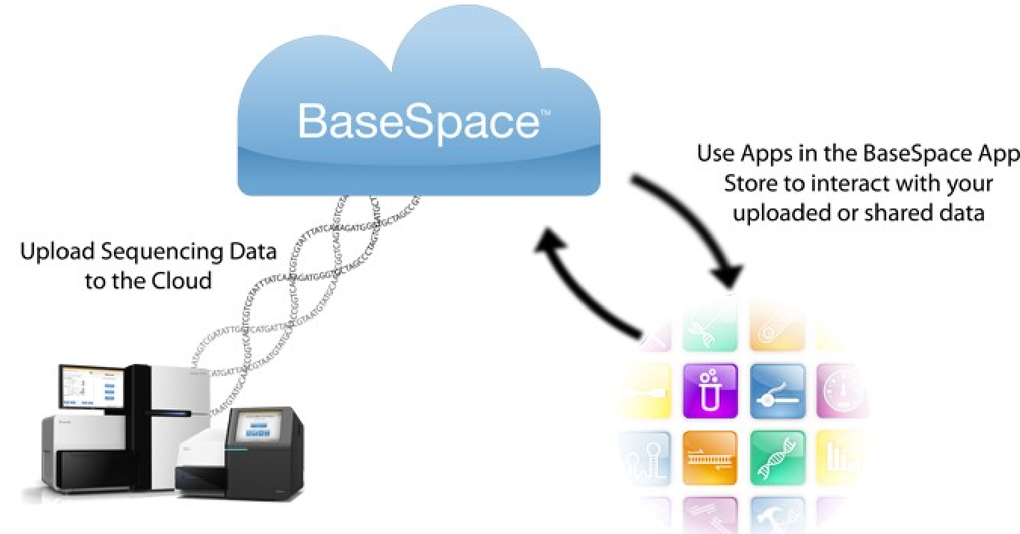
- MCS controls image generation on the MiSeq
- One cycle includes the chemical addition and imaging of one base for each cluster on the flow cell
- For imaging, the MiSeq flow cell is broken up into imaging areas or tiles
  - The number of tiles imaged depends on the flow cell type (standard, nano, or 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



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 - techynologie je tak drahá, že již není více konkurenceschopná	nižší přesnost na konci readů, interference u nízkodiverzních knihoven	chybovost v homopolymerech, EM PCR
Shotgun knihovny	400 bp 450 bp 600 bp 600 bp	250 bp 250 bp 600 bp	400 bp 450 bp 600 bp 600 bp
Amplikony	450 bp 450 bp 400 bp 450 bp	300 bp 300 bp 450 bp Overlap 150 bp	450 bp 450 bp 400 bp 450 bp