

Sekvenování

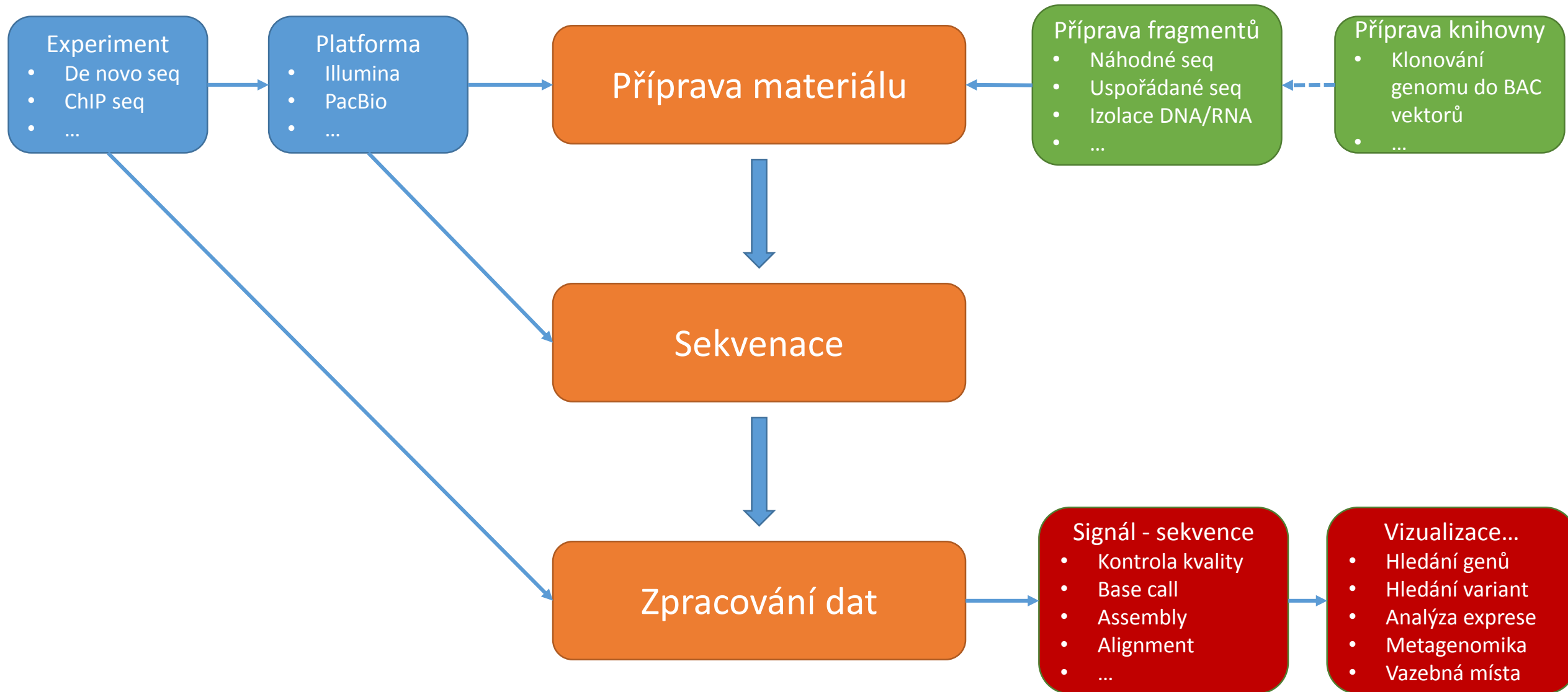
Sekvenování

Stanovení sekvence nukleotidů v molekule = Stanovení primární struktury DNA / RNA

Proč sekvenovat?

- *de novo* sekvenace celých genomů – **whole genome sequencing**
- analýza polymorfizmů (SNP, ...) a mutací - **targeted resequencing**
- identifikace organismů – **genome sequencing + resequencing**
- identifikace cílových motivů proteinů vážících se na DNA – sekvenace výsledků chromatinové imunoprecipitace – **ChIP seq**
- analýza transkriptomu – kvantifikace RNA – **RNA seq**
- identifikace metylace DNA (cytosinů) – **Bisulfite sequencing**

Obecný postup



Sekvenační techniky – dle generace

Klasické techniky

- **Chemická sekvenace** – dle Maxam a Gilbert
- **Enzymatická sekvenace** – Sanger / dideoxy
 - Nyní automatizovaná – kapilární sekvenátory

Next-generation sequencing (NGS)

- **454 sekvenování / pyrosekvenování** – Roche
- **Ion Torrent** – now ThermoFisher
- **SOLiD** – Sequencing by Oligonucleotide Ligation and Detection – ABI / Life Technologies / ThermoFisher
- **Illumina (Solexa)**
- ...

Sekvenování třetí generace (3rd generation sequencing)

- **Helicos**
- **PacBio** – metoda SMRT- Single-Molecule Real-Time sequencing – Pacific Biosciences
- **NanoPore** – proteinové nanopóry (α -hemolysin, ...) – Oxford Nanopore
- ...

Sekvenační techniky – dle generace

Klasické techniky

- ~~Chemická sekvenace~~ – dle Maxam a Gilbert
- **Enzymatická sekvenace** – Sanger / dideoxy
 - Nyní automatizovaná – kapilární sekvenátory

Next-generation sequencing (NGS)

- **454 sekvenování / pyrosekvenování** – Roche
- **Ion Torrent** – now ThermoFisher
- **SOLiD** – Sequencing by Oligonucleotide Ligation and Detection – ABI / Life Technologies / ThermoFisher
- **Illumina (Solexa)**
- ...

Sekvenování třetí generace (3rd generation sequencing)

- ~~Helicos~~
- **PacBio** – metoda SMRT- Single-Molecule Real-Time sequencing – Pacific Biosciences
- **NanoPore** – proteinové nanopóry (α -hemolysin, ...) – Oxford Nanopore
- ...

Sekvenační techniky – dle generace

Klasické techniky

- Vstupem každé jednotlivé sekvenace je velké množství molekul NK pro přípravu jednořetězcového templátu jenž je sekvenován
- Každý fragment je sekvenován a detekován individuálně jako samostatná reakce

Next-generation sequencing (NGS)

- Vstupem každé jednotlivé sekvenační reakce je jediná molekula NK, která je amplifikována na velké množství jednořetězcových templátů, jež jsou následně sekvenovány
- Sekvenuje se paralelně velké množství menších fragmentů NK (krátká čtení), jež se násobně překrývají

Sekvenování třetí generace (3rd generation sequencing)

- Vstupem každé jednotlivé sekvenační reakce je jediná molekula NK, jež je přímo sekvenována bez amplifikace
- Sekvenuje se paralelně více velmi dlouhých fragmentů (dlouhá čtení)

Sekvenační techniky – dle principu

Sekvence syntézou – detekována je inkorporace specifických bazí při / po polymerační reakci

- Sangerovo sekvenování
- Pyrosekvenování
- Illumina

Sekvence štěpením – dlouhá NK je specificky štěpená v místech jednotlivých bazí

- Maxam a Gilbert

Sekvence ligací

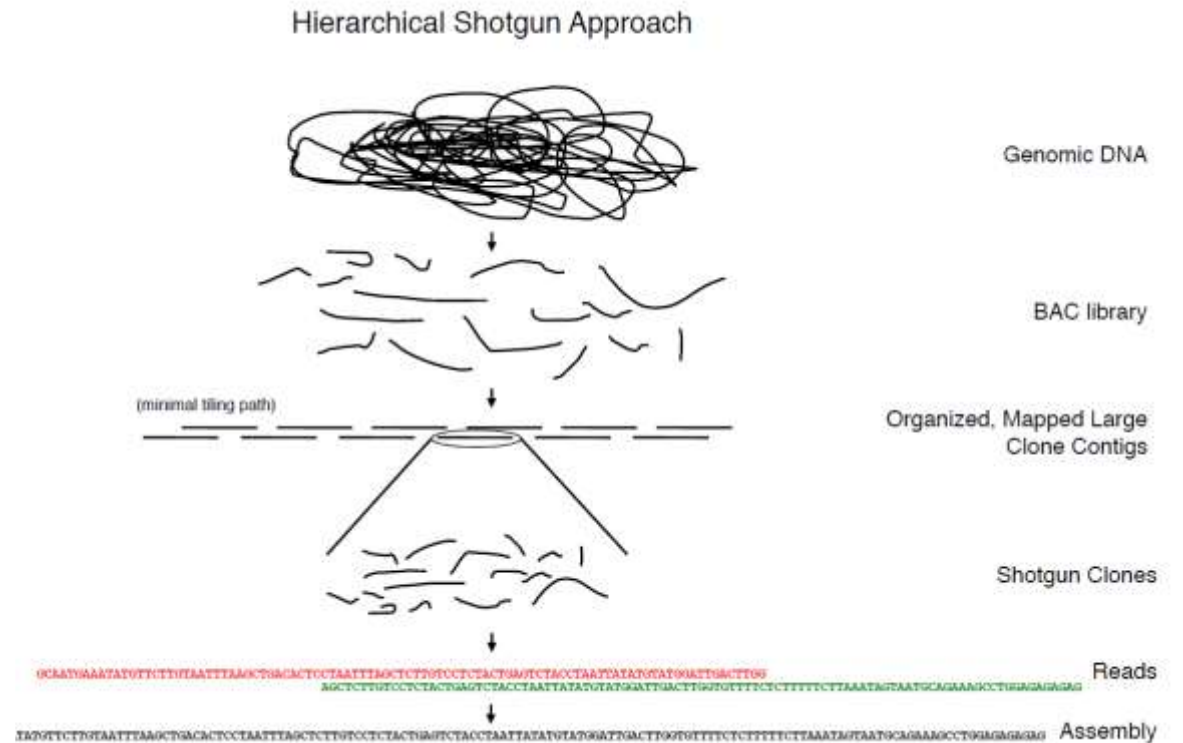
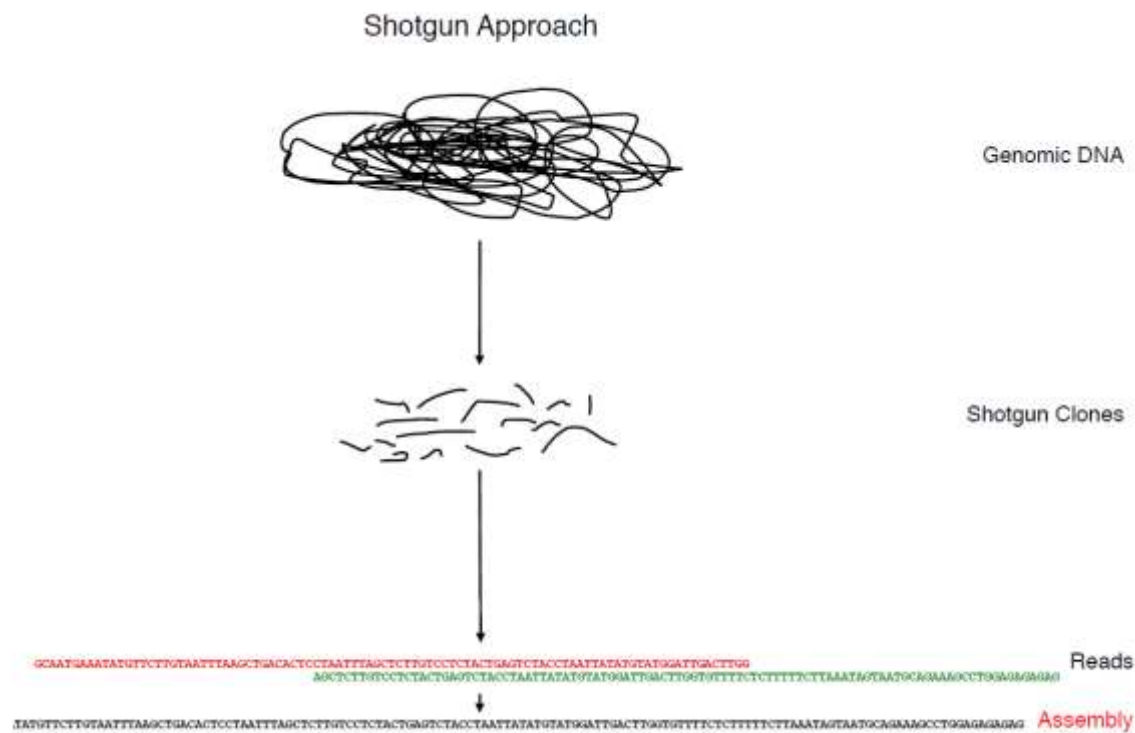
- SOLiD

Sekvence elektrochemickou detekcí – jednotlivé báze NK při průchodu membránovým pórem poskytují elektrickou odezvu

- NanoPore

Náhodné sekvenování (shotgun sequencing)

- Při náhodném sekvenování genomu jsou nejdříve připraveny mechanickým stříháním genomové DNA fragmenty s optimální velikostí (1 300 – 2 000 bp) a po úpravě jejich konců jsou nahodile naklonovány do vhodného vektoru za vzniku velkého počtu klonů.
- Ke štěpení DNA se využívá sonikace nebo pankreatická DNáza I za přítomnosti Mn²⁺



Uspořádané sekvenování

Procházení primerem - „primer walking“

- vyžaduje znalost sekvence, ke které se připojuje primer, který umožní prodloužení řetězce DNA-polymerázou.
- získaná sekvence z první reakce je použita pro návrh primeru pro další reakci a tento krok se opakuje, dokud není dosaženo kompletního stanovení sekvence.

1. kolo



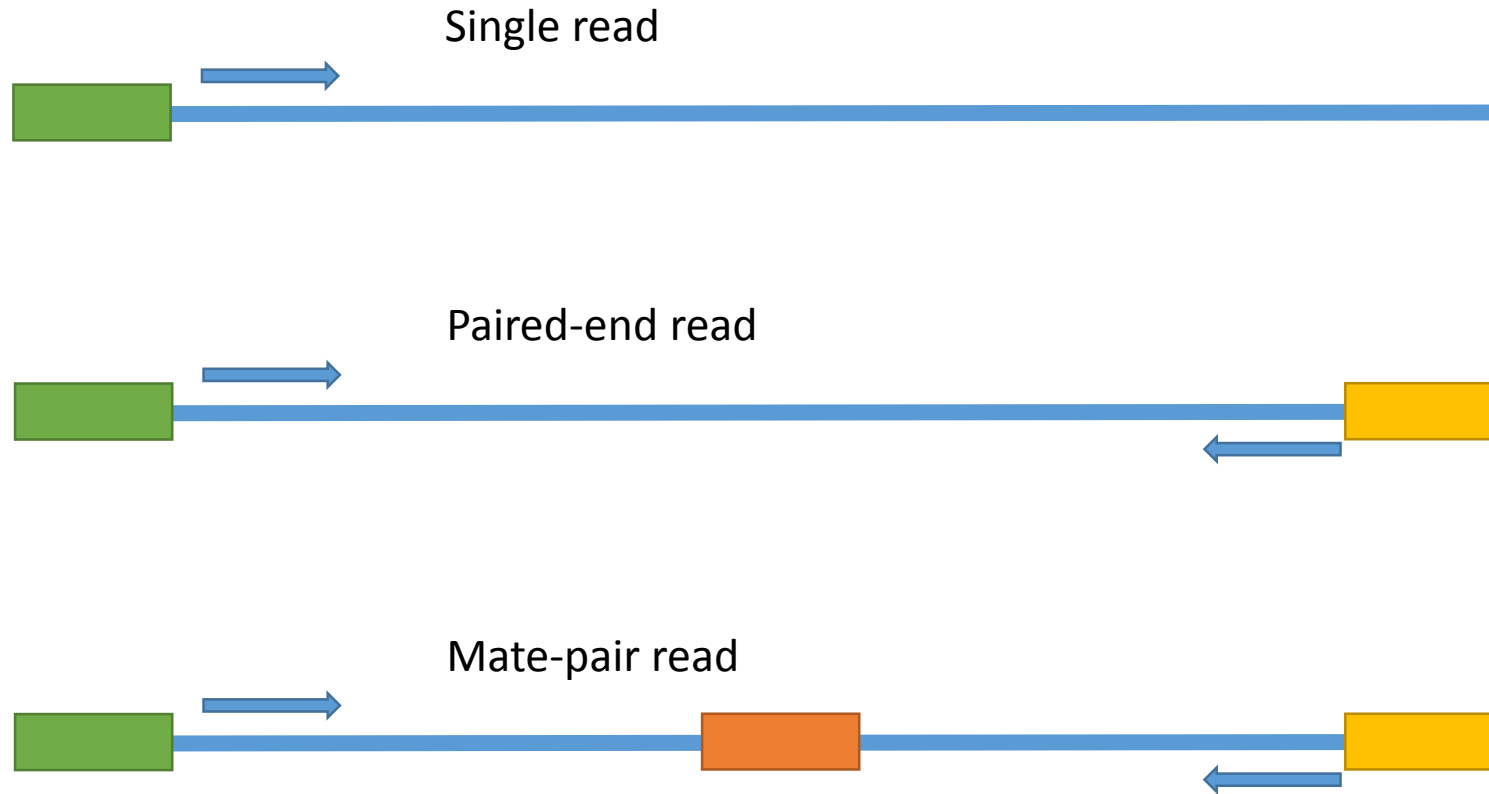
2. kolo



3. kolo

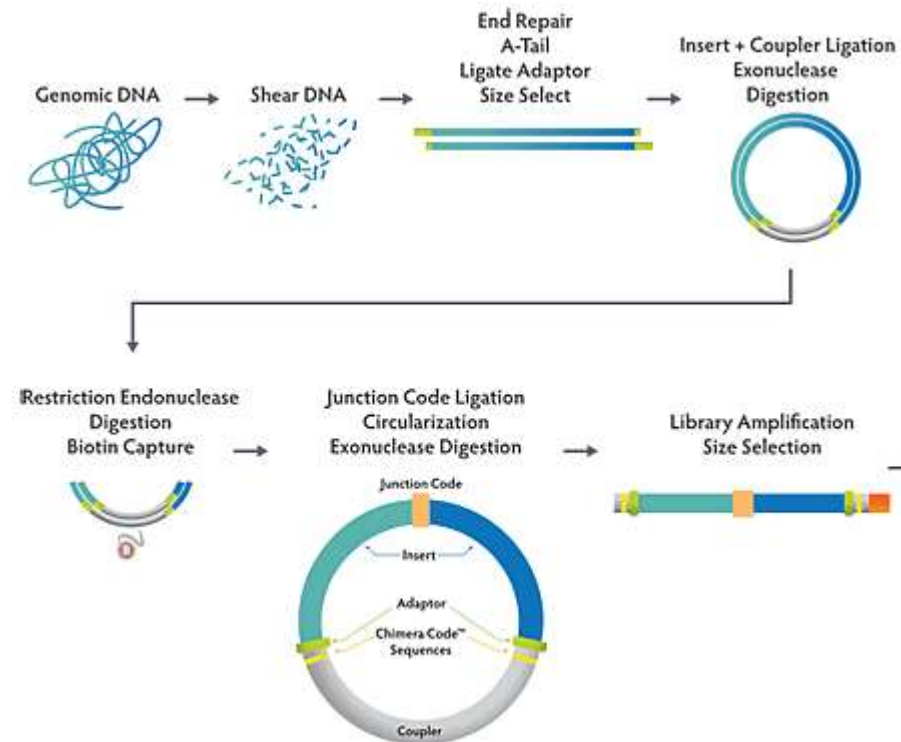


Uspořádání sekvenačního experimentu

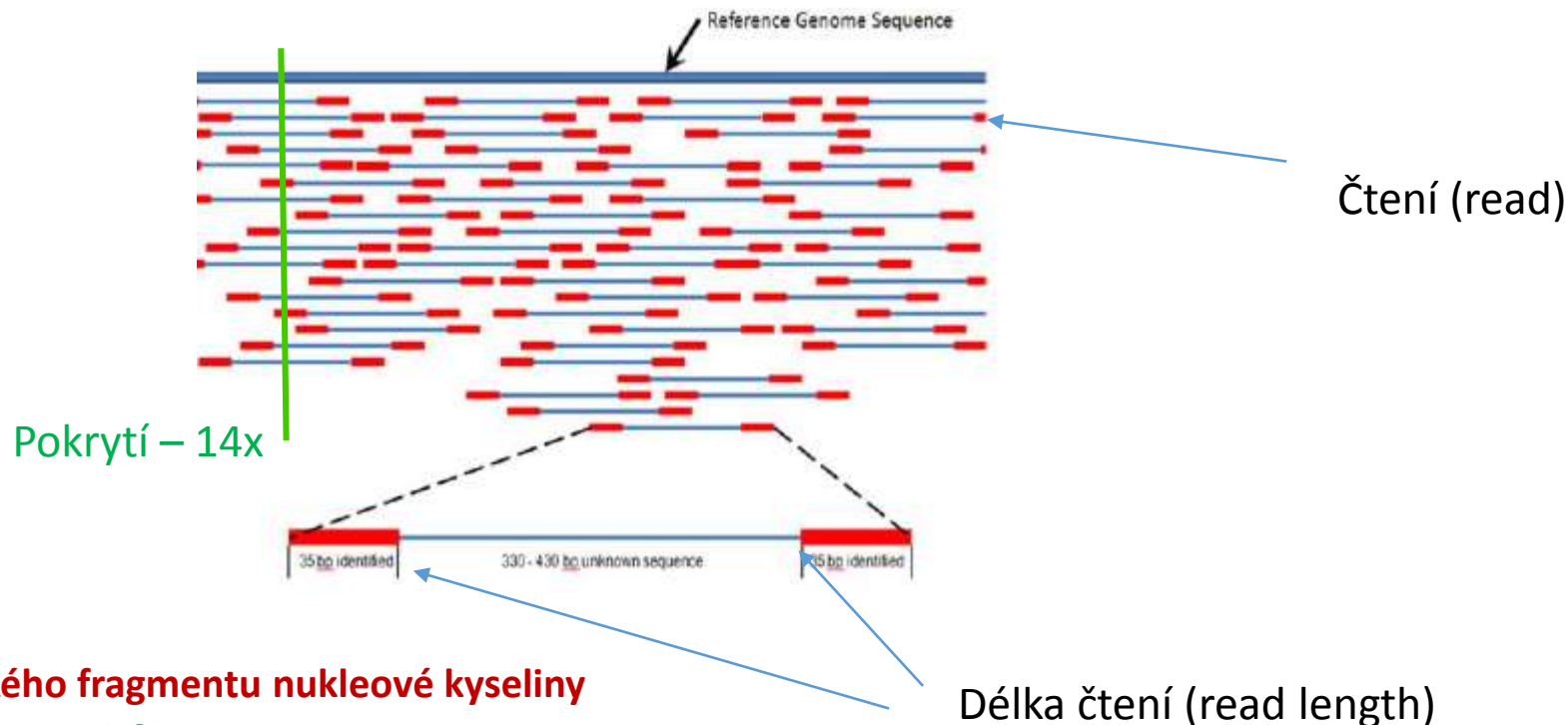


Mate-pair library

- Problém sekvenování repetitivních oblastí pomocí platform produkujících krátká čtení
- Mate-pair knihovny – delší oblast je reprezentována jediným fragmentem, který obsahuje ligované konce této oblasti.
- Informace o vzájemné pozici vzdálenějších oblastí, čitelná i pomocí krátkých čtení.



Next-Generation + 3rd gen. sekvenování



Čtení (read)

- Sekvence jednoho fyzického fragmentu nukleové kyseliny

Délka čtení (read length)

- Obvyklá délka osekvenovaného fragmentu v rámci jedné „reakce“

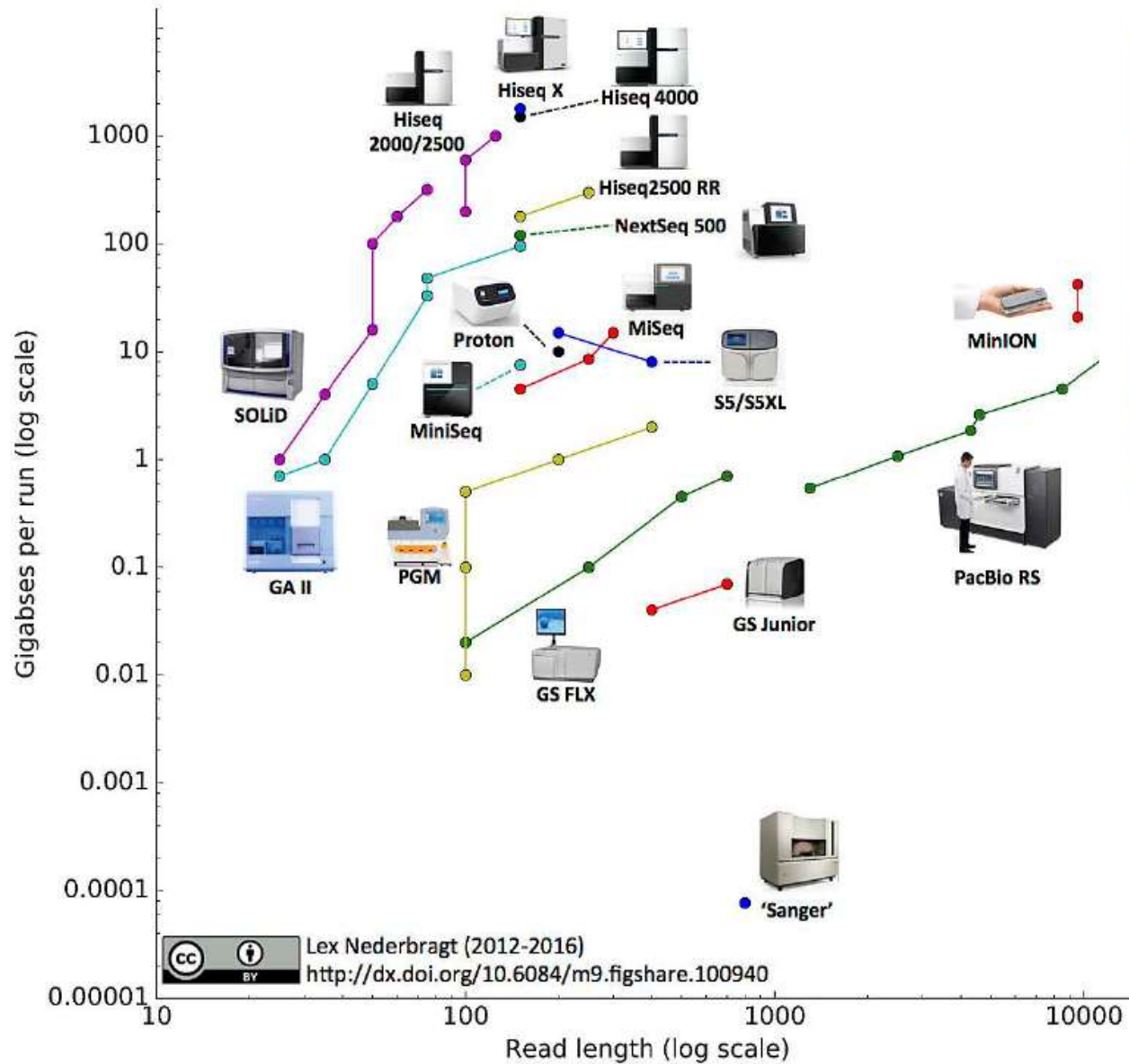
Pokrytí / hloubka sekvenování (coverage / sequencing depth)

- Počet osekvenování jednoho nukleotidu/oblasti
 - Průměrná hloubka = množství osekvenovaných dat (Gb) / velikost sekvenovaného genomu

Přesnost

- Přesnost osekvenování – v případě jediného fyzického fragmentu může být, dle použité metody i nižší (<90%)
- Díky vyššímu pokrytí je výsledná přesnost vysoká

Next-Generation + 3rd gen. sekvenování



Platform	Reads x run: (M)	Read length:	Run time: (d)	Yield: (Gb)	Rate: (Gb/d)	per-Gb: (\$)	hg-30x: (\$)	Machine: (\$)
ISeq 100 1fcell	4	150	0.77	1.2	1.56	521	62500	19.9K
MiniSeq 1fcell	25	150	1	7.5	7.5	233	28000	49.5K
MiSeq 1fcell	25	300	2	15	7.5	66	8000	99K
Next Seq 550 1fcell	400	150	1.2	120	100	50	5000	250K
HiSeq 2500 RR 2fcells	600	100	1.125	120	106.6	51.2	6144	740K
HiSeq 2500 V3 2fcells	3000	100	11	600	55	39.1	4692	690K
HiSeq 2500 V4 2fcells	4000	125	6	1000	166	31.7	3804	690K
HiSeq 4000 2fcells	5000	150	3.5	1500	400	20.5	2460	900K
HiSeq X 2fcells	6000	150	3	1800	600	7.08	849.6	1M
Nova Seq S1 2018 2fcells	3300	150	1.66	1000	600	18	1800	999K
Nova Seq S2 2fcells	6600	150	1.66	2000	1200	15	1564	999K
Nova Seq S4 2fcells	20000	150	1.66	6000	3600	5.8	700	999K
RS P6-C4 16cells	0.88	20K	4.3	12	2.8	200	24000	695K
Sequel 16cells	5.84	20K	6.6	104	--	80	9600	350K
R&D end 2018	--	32K	--	192	--	6.6	1000	350K
Mini ION R9.5 1fcell	--	10-100K+	2	10-20	5-10	--	--	--
Grid ION X5 5fcells	--	10-100K+	2	50 - 100	25-50	--	--	125K
PromethION Rnd 48fcells	--	10-100K+	2	2400	1200	20	2400	75K
PromethION limit 48fcells	--	10-100K+	2	11000	5500	4.3	500	75K

Next-Generation + 3rd gen. sekvenování

- Pro různé typy experimentů vhodné různé délky **čtení** x **pokrytí**, tedy různé platformy
 - De novo sekvenování – lepší delší čtení s dostatečným pokrytím pro stanovenou přesnost
 - CHIP / RNA seq – pro dostatečnou „statistiku“ nutné vysoké počty čtení jednotlivých oblastí („jakoby pokrytí“), ale nejsou nutná dlouhá čtení

Sekvenování dle Maxam a Gilbert

1. Příprava jednořetězcové NK (DNA)
2. Terminální značení (5' konec - T4 PNK, 3' konec – TdT, fluorescenční primer)
3. Rozdělení na 4 reakce
4. **Specifická chemická modifikace bazí**
5. **Štěpení NK alkalickou hydrolyzou (1M piperidin; 90°C; 30 min)**
6. Elektroforéza (denaturační PAGE, kapilární při fluorescenčním značení)
7. Vizualizace (autoradiografie, fluorescence)

G

Dimetyl
sulfát

A+G

k. mravenčí

T+C

Hydrazin
voda

C

Hydrazin
5M NaCl

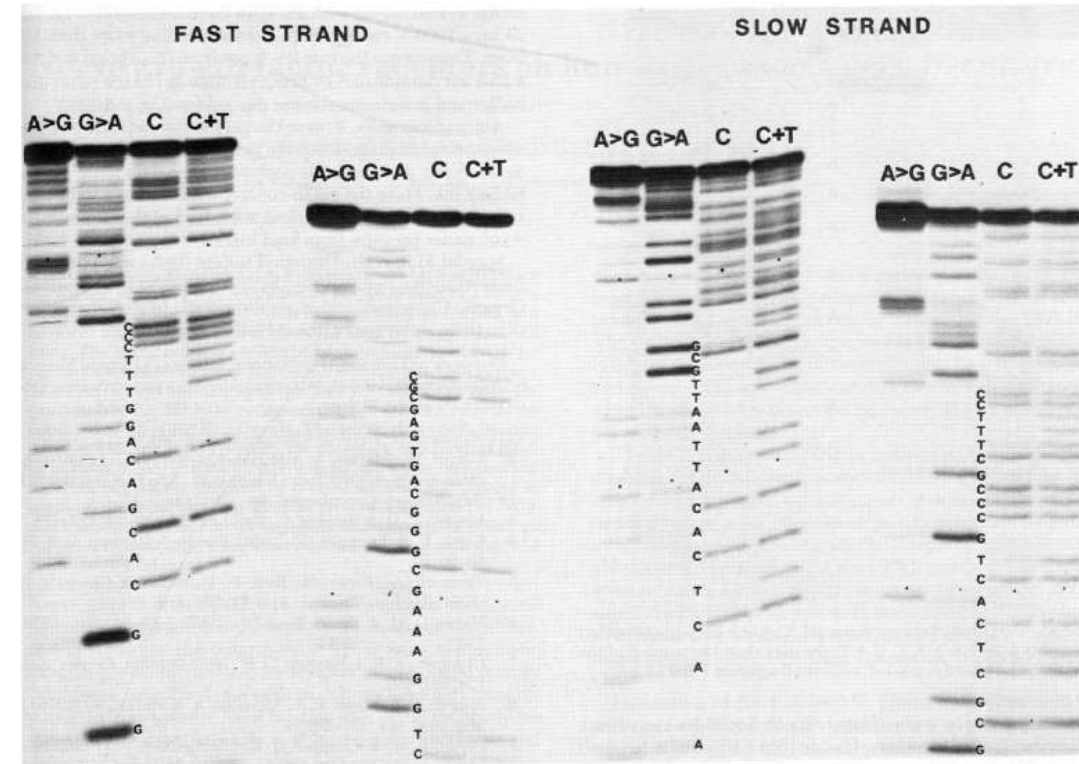


FIG. 2. Autoradiograph of a sequencing gel of the complementary strands of a 64-base-pair DNA fragment. Two panels, each with four reactions, are shown for each strand; cleavages proximal to the 5' end are at the bottom on the left. A strong band in the first column with a weaker band in the second arises from an A; a strong band in the second column with a weaker band in the first is a G; a band appearing in both the third and fourth columns is a C; and a band only in the fourth column is a T. To derive the sequence of each strand, begin at the bottom of the left panel and read upward until the bands are not resolved; then, pick up the pattern at the bottom of the right panel and continue upward. One-tenth of each strand, isolated from the gel of Fig. 1, was used for each of the base-modification reactions. The dimethyl sulfate treatment was 50 mM for 30 min to react with A and G; hydrazine treatment was 18 M for 30 min to react with C and T and 18 M with 2 M NaCl for 40 min to cleave C. After strand breakage, half of the products from the four reactions were layered on a 1.5 × 330 × 400 mm denaturing 20% polyacrylamide slab gel, pre-electrophoresed at 1000 V for 2 hr. Electrophoresis at 20 W (constant power), 800 V (average), and 25 mA (average) proceeded until the xylene cyanol dye had migrated halfway down the gel. Then the rest of the samples were layered and electrophoresis was continued until the new bromphenol blue dye moved halfway down. Autoradiography of the gel for 8 hr produced the pattern shown.

Sekvenování dle Maxam a Gilbert

Jednořetězcová
DNA

*-GATCGGTGAACTGTCCAGA

DMS + piperidin

Hydrazin ve vodě
+ piperidin

k. mravenčí + piperidin
Hydrazin v 5M NaCl
+ piperidin

*-G
*-GATCG
*-GATCGG
*-GATCGGTG
*-GATCGGTGAACTG
*-GATCGGTGAACTGTCCAG

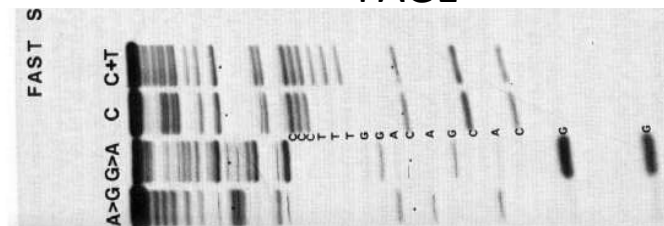
*-GAT
*-GATC
*-GATCGGT
*-GATCGGTGAAC
*-GATCGGTGAACT
*-GATCGGTGAACTGT
*-GATCGGTGAACTGTC
*-GATCGGTGAACTGTCC

*-GATC
*-GATCGGTGAAC
*-GATCGGTGAACTGTC
*-GATCGGTGAACTGTCC

*-G
*-GA
*-GATCG
*-GATCGG
*-GATCGGTG
*-GATCGGTGA
*-GATCGGTGAA
*-GATCGGTGAACTG
*-GATCGGTGAACTGTCCA
*-GATCGGTGAACTGTCCAG
*-GATCGGTGAACTGTCCAGA



Denaturační
PAGE



Sangerovo sekvenování

1. Příprava jednořetězcové NK (DNA) jako templátu
2. Terminální značení primeru (5' konec - T4 PNK, 3' konec – TdT, fluorescenční primer)
3. Rozdělení na 4 reakce, každá s přidavkem jednoho **dideoxynTP (ddNTP)**
4. **Polymerační reakce – prodlužování primeru (templát + primer + polymeráza + dNTP mix + příslušný ddNTP v malém množství)**
5. Elektroforéza (denaturační PAGE, kapilární při fluorescenčním značení)
6. Vizualizace (autoradiografie, fluorescence)

G
dNTP
ddGTP

A
dNTP
ddATP

T
dNTP
ddTTP

C
dNTP
ddCTP

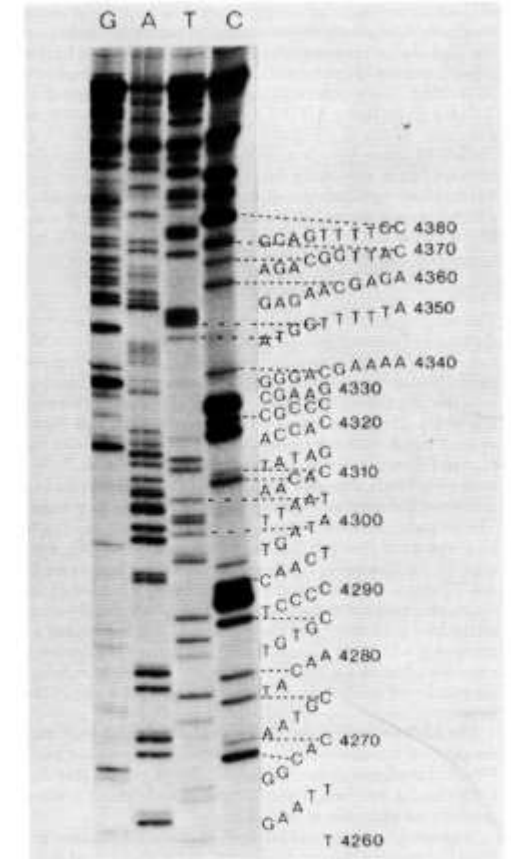
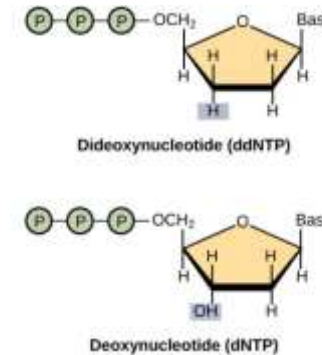


FIG. 2. Autoradiograph from an experiment using fragment R4 as primer on the complementary strand of ϕ X174 DNA. Conditions were as in Fig. 1 with the following exceptions: ddCTP was used as inhibitor instead of araCTP. After incubation of the solutions at room temperature for 15 min, 1 μ l of 0.5 mM dATP and 1 μ l of restriction enzyme *Hae* III (4 units/ μ l) were added and the solutions were incubated at 37° for 10 min. The *Hae* III cuts close to the *Hind*III site and it was used because it was more readily available. The electrophoresis was on a 12% acrylamide gel at 40 mA for 14 hr. The top 10.5 cm of the gel is not shown.

Sangerovo sekvenování

Polymerační reakce
za přítomnosti 4
dNTP a některého
ddNTP

*-primer
templát

*-CTAG
GATCGGTGAACTGTCCAGA

Pouze dATP

*-CTAGCCACTTGACAGGTCA

Pouze ddATP

*-CTAGCCA

Optimální směs
dATP + ddATP

*-CTAGCCA

*-CTAGCCACTTGA

*-CTAGCCACTTGACA

*-CTAGCCACTTGACAGGTCA

Denaturační
PAGE

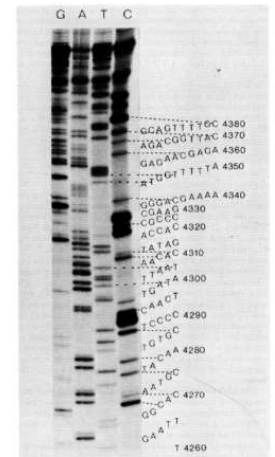
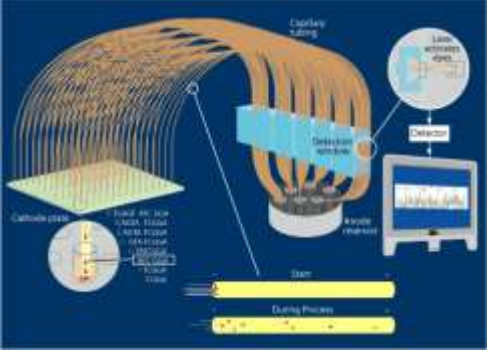
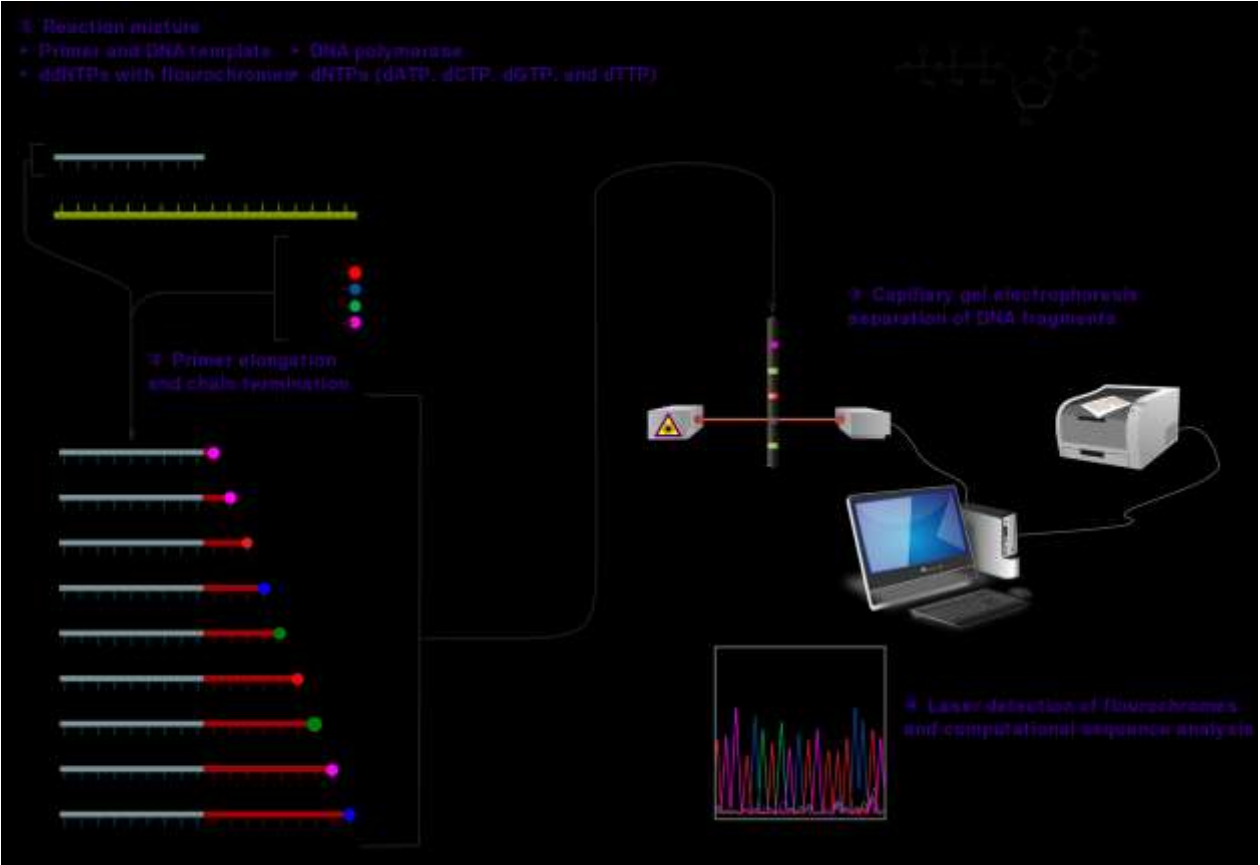


FIG. 2. Autoradiograph from an experiment using fragment 84 as primer on the complementary strand of λ X174 DNA. Conditions were as in Fig. 1 with the following exceptions: ddCTP was used as inhibitor instead of araCTP. After incubation of the solutions at room temperature for 15 min, 1 μ l of 0.5 mM dATP and 1 μ l of restriction enzyme *Hae* III (4 units) were added and the solutions were incubated at 37° for 10 min. The *Hae* III cuts close to the *Hae* III site and it was used because it was more readily available. The electrophoresis was on a 12% acrylamide gel at 40 mA for 14 hr. The top 10.5 cm of the gel is not shown.

Automatizované Sangerovo sekvenování

Fluorescenčně značené mohou být primery nebo terminátory



Sample name: C113

User name: Daniel Ruzick
Order no.: 172110581
Order date: 06/04/2017



Next-Generation Sekvenování

- Vstupem každé jednotlivé sekvenační reakce je jediná molekula NK, která je amplifikována na velké množství jednořetězcových templátů, jež jsou následně sekvenovány
- Sekvenuje se paralelně velké množství menších fragmentů NK, jež se násobně překrývají

Illumina

Ion Torrent

SOLiD

Pyrosekvenování

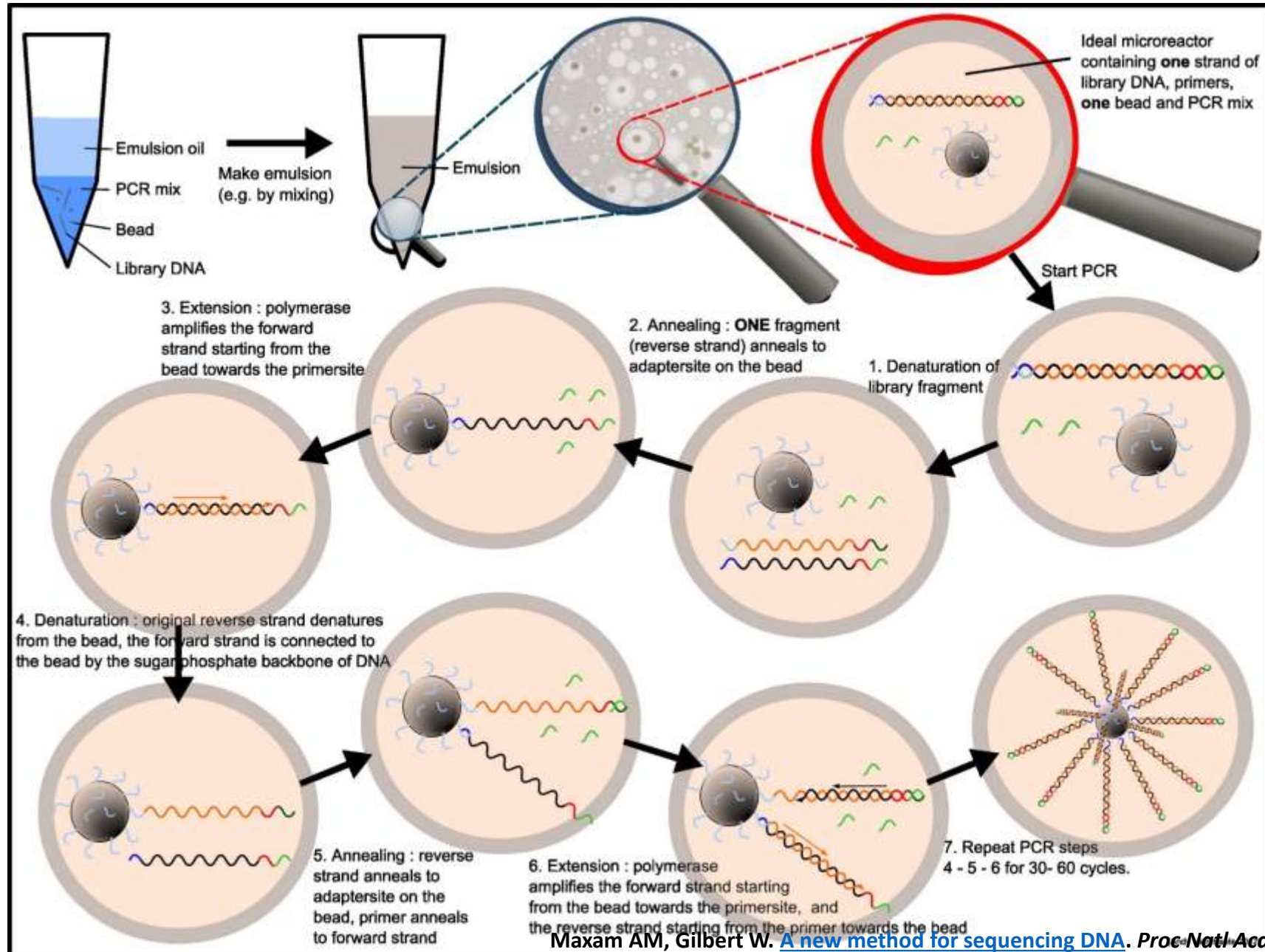
Amplifikace NK pro sekvenování

Před samotným sekvenováním NGS je nutné amplifikovat vstupní fragmenty NK tak, aby amplicony každého fragmentu byly separovány od ostatních fragmentů.

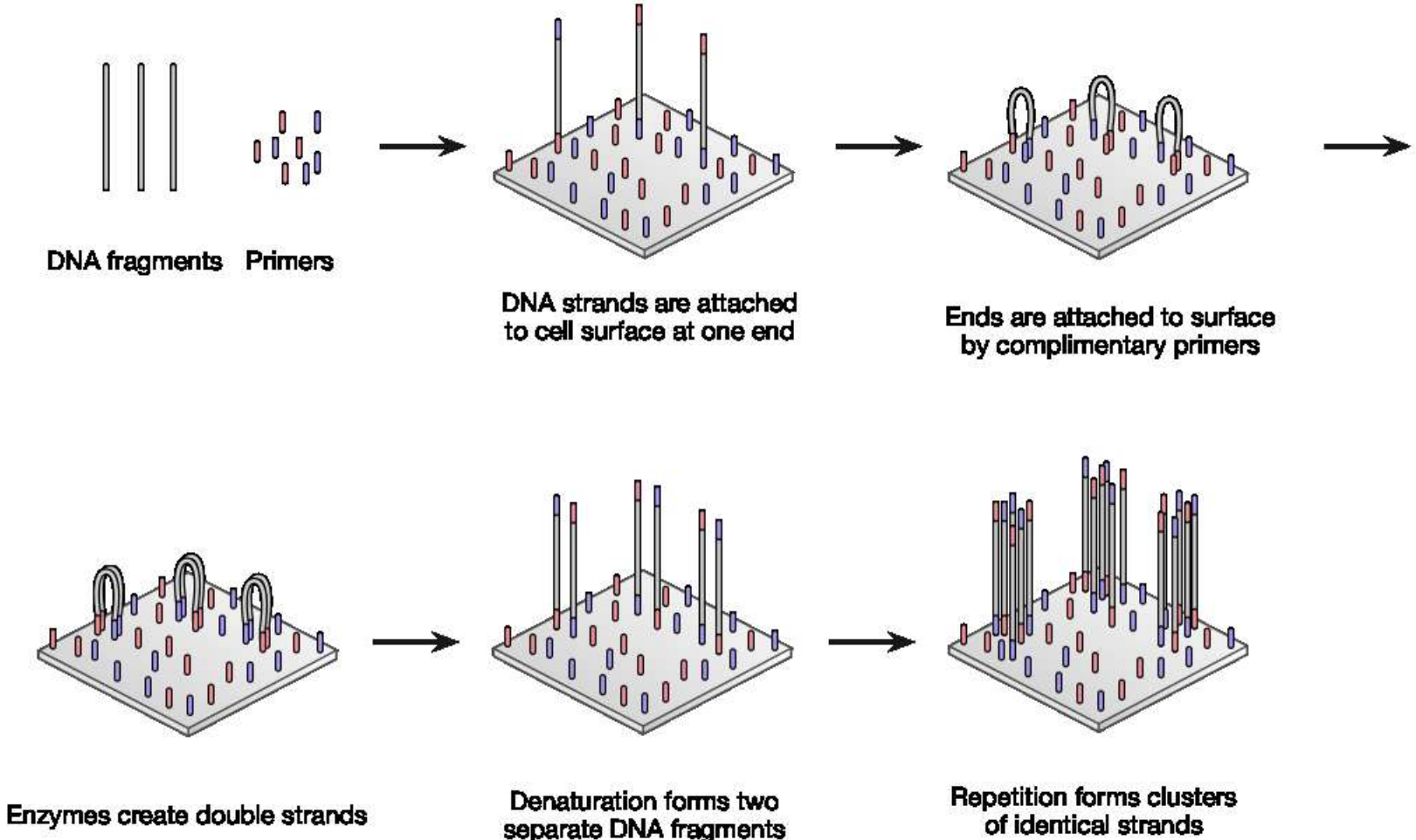
Emulzní PCR

Bridge / Polony PCR

Emulzní PCR



Bridge / „Polony“ PCR



Illumina (Solexa)

- Amplifikace pomocí **bridge PCR** – tvorba clusterů
- Sekvence **polymerací**
- Využití odlišně **fluorescenčně** značených „**reversible terminator**“ nukleotidů
- Detekce fluorescence souběžně pro všechny clustery po každém připojení nukleotidu – „mikroskopicky“
- Krátká čtení (150 bp), paired-end
- Velké množství paralelních čtení (10^7 - 10^{10})
- Velký objem výstupních dat (až Tb / běh)

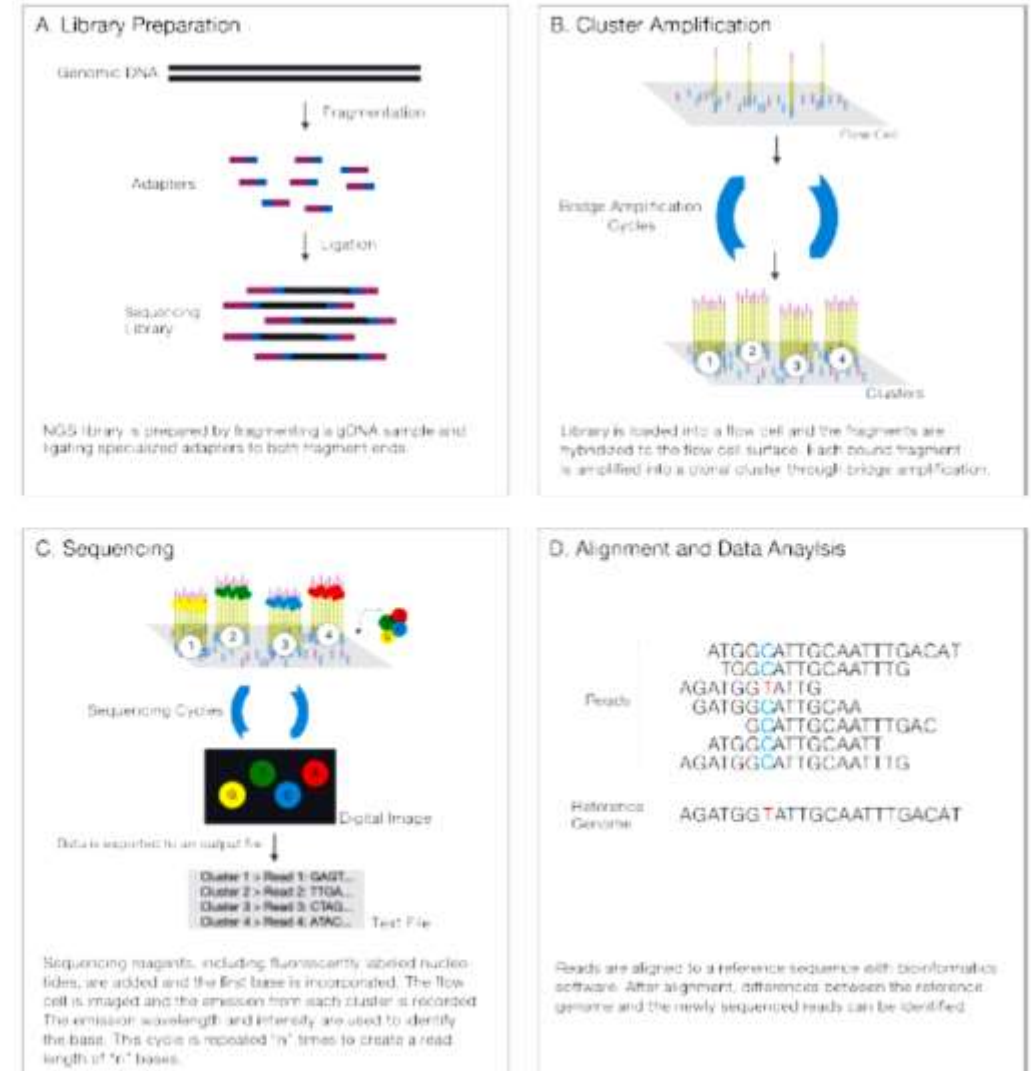
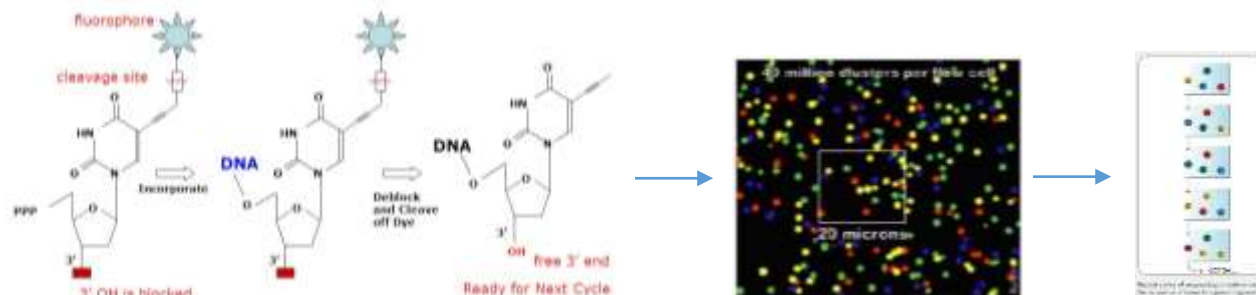
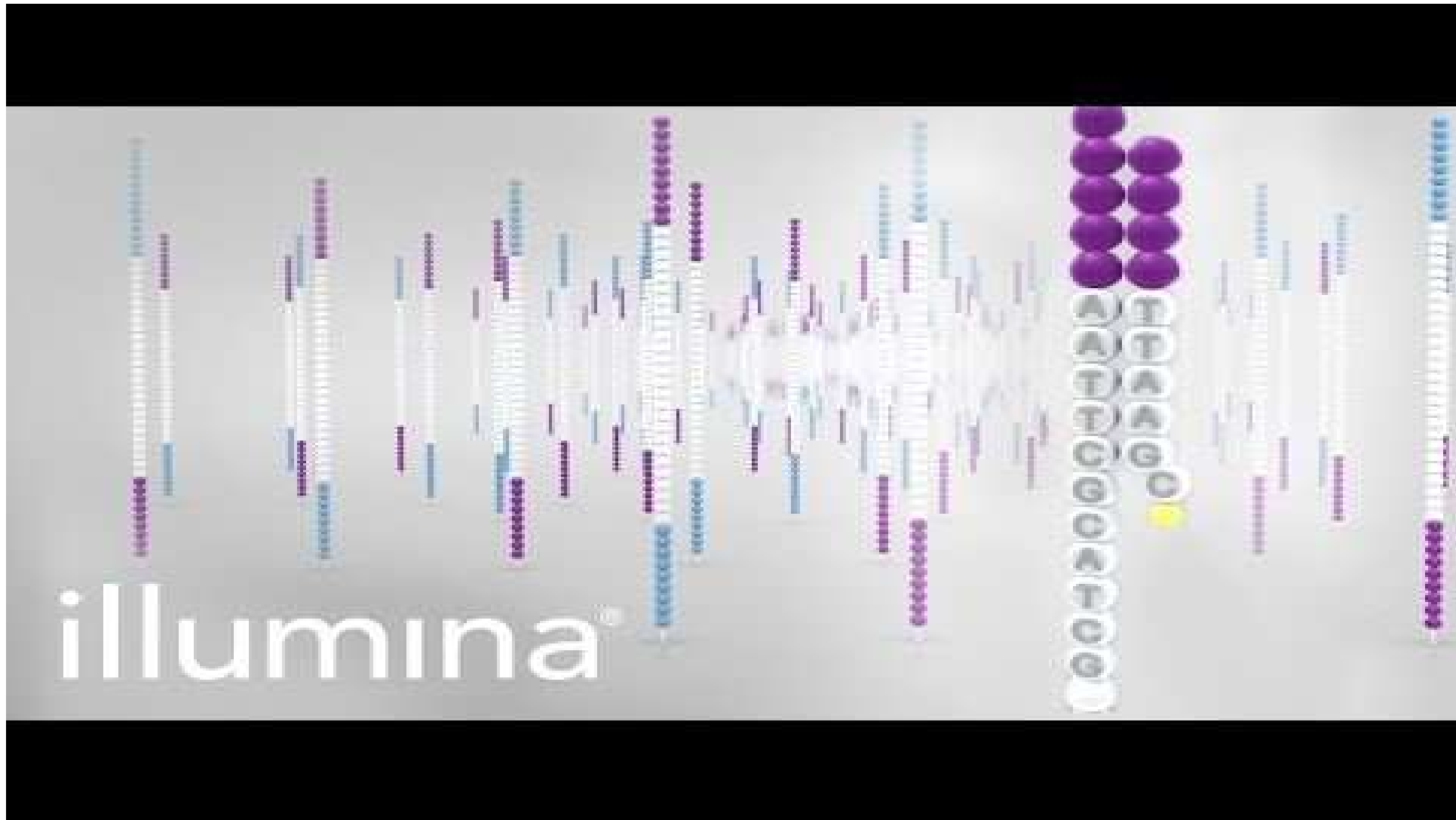


Figure 3: Next-Generation Sequencing Chemistry Overview—Illumina NGS includes four steps: (A) library preparation, (B) cluster generation, (C) sequencing, and (D) alignment and data analysis.

Illumina (Solexa)



Illumina (Solexa)



NextSeq Series



HiSeq 4000 System



HiSeq X Series²



NovaSeq 6000 System

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

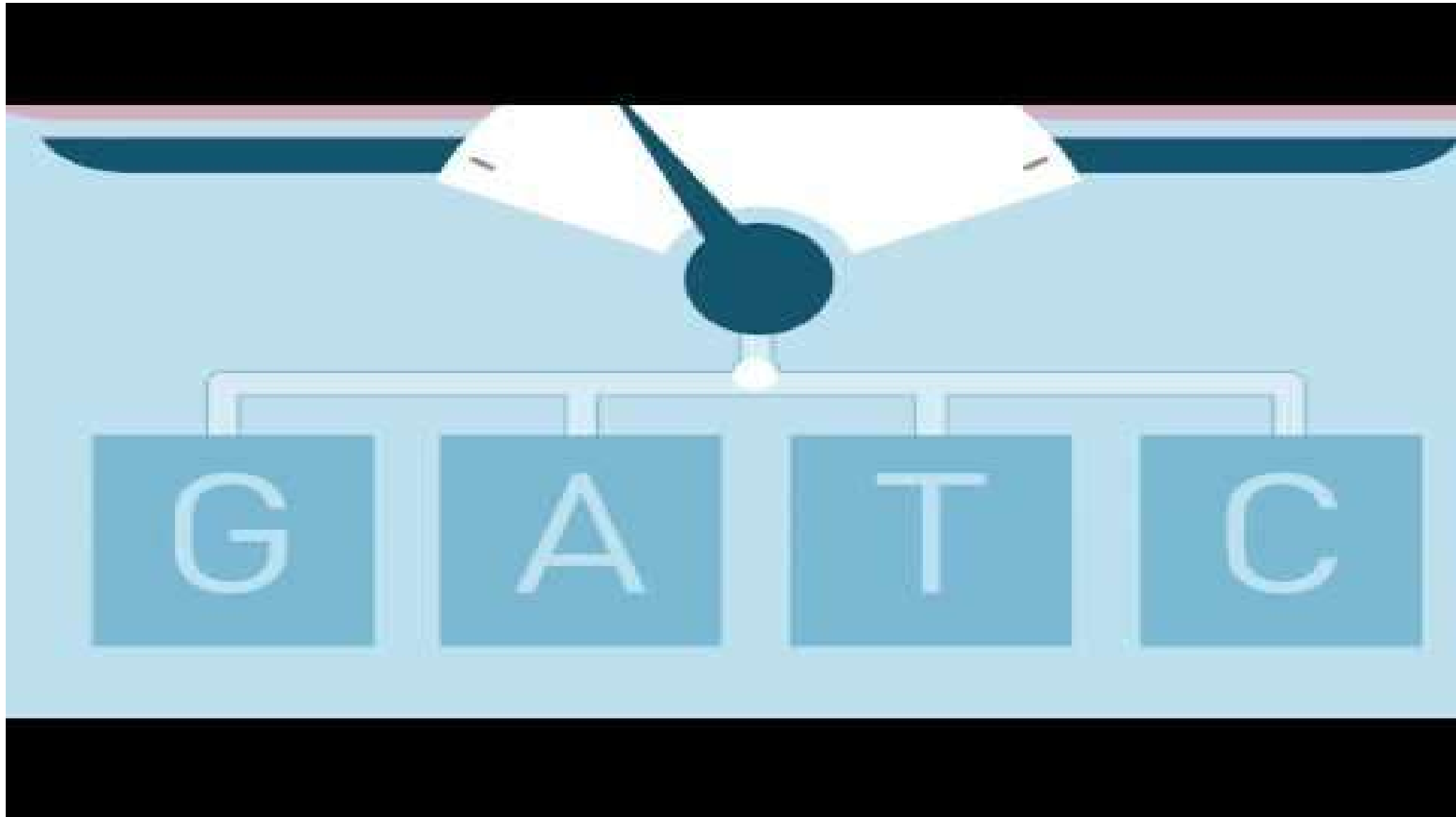
Optimized NGS Sample Tracking and Workflows

See how a Laboratory Information Management System (LIMS) 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	< 3 days	~13–25 hours (dual S1 flow cells) ~16–36 hours (dual S2 flow cells) ~44 hours (dual S4 flow cells)
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

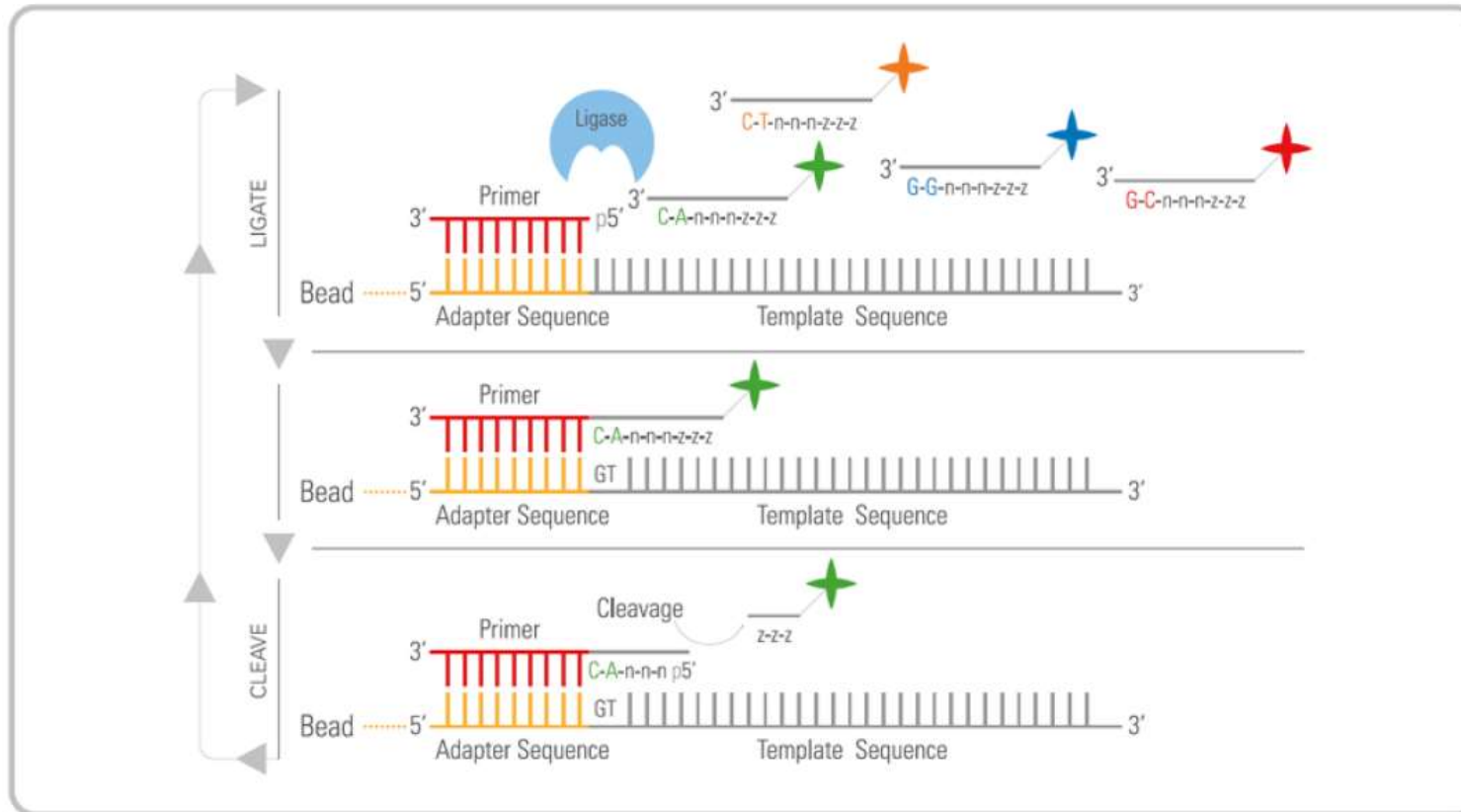
Ion Torrent



SOLiD

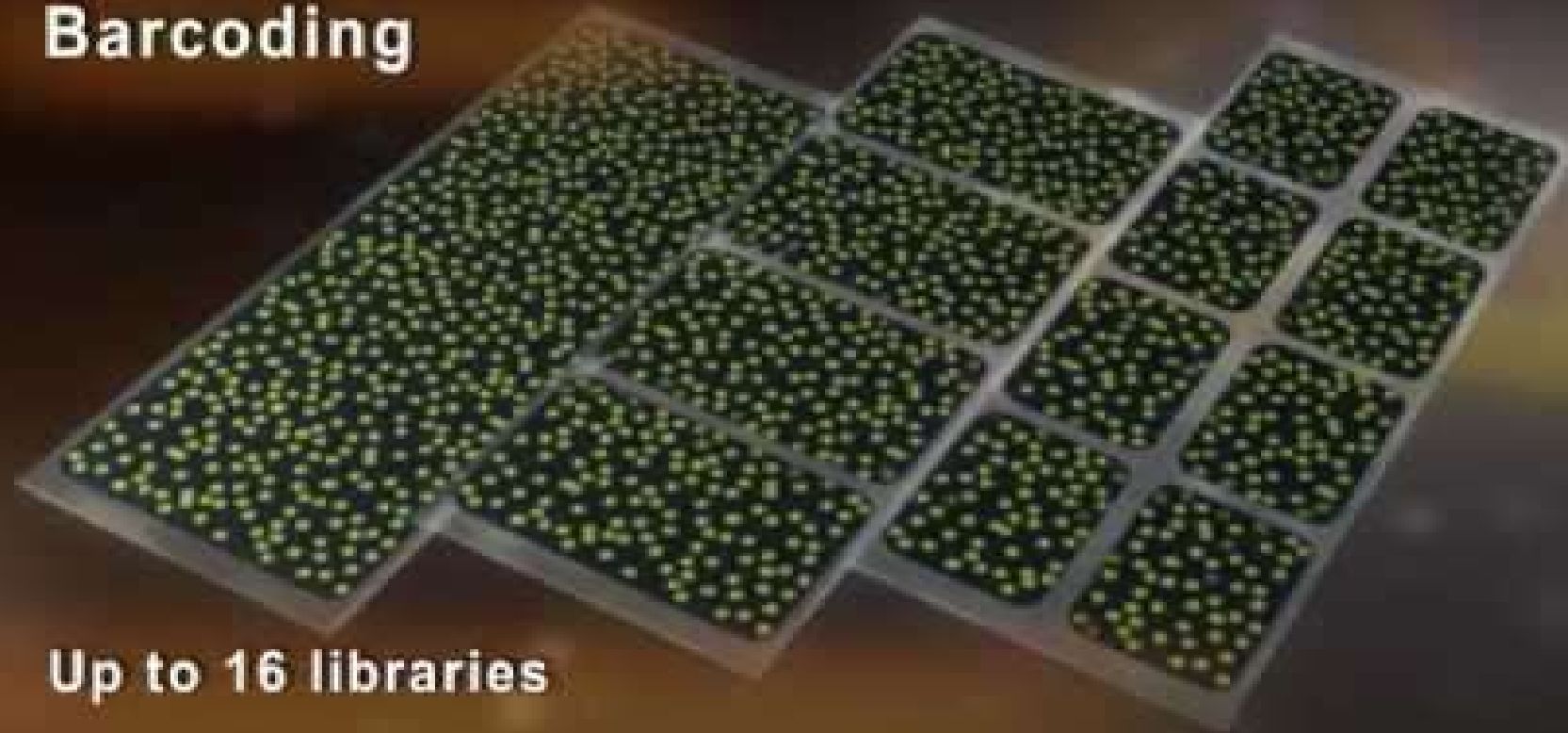
Sequencing by Oligonucleotide Ligation and Detection

- Sekvenčně specifická ligace oktamerých oligonukleotidových sond označených fluorescenčními barvivy
- Velmi vysoká přesnost na prvních nukleotidech, velmi krátká čtení



SOLiD

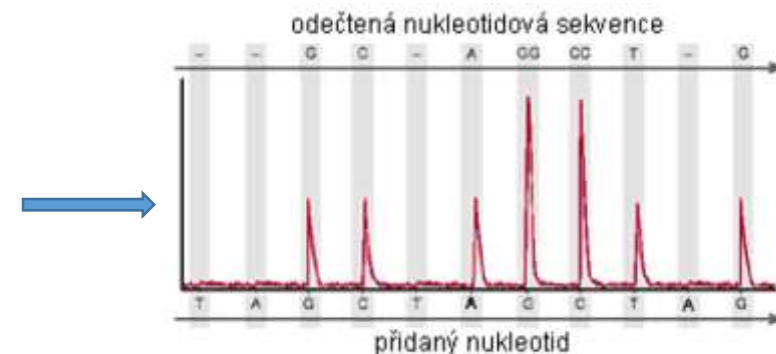
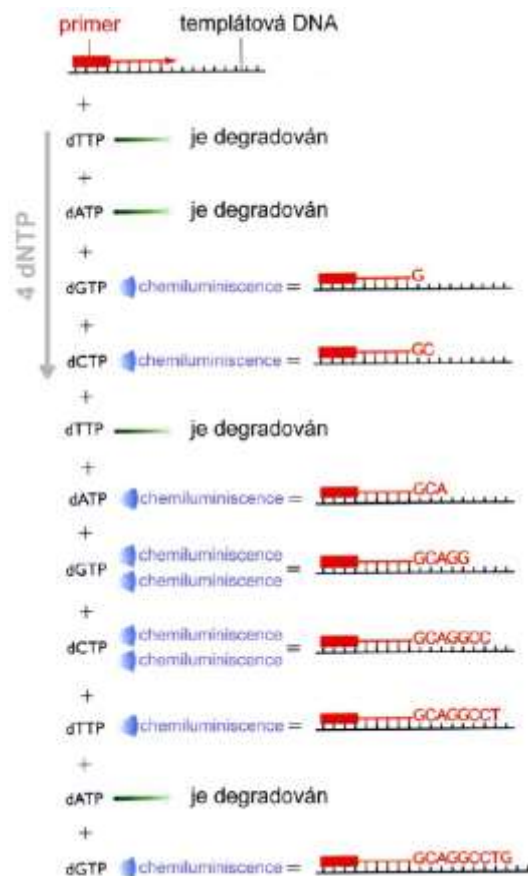
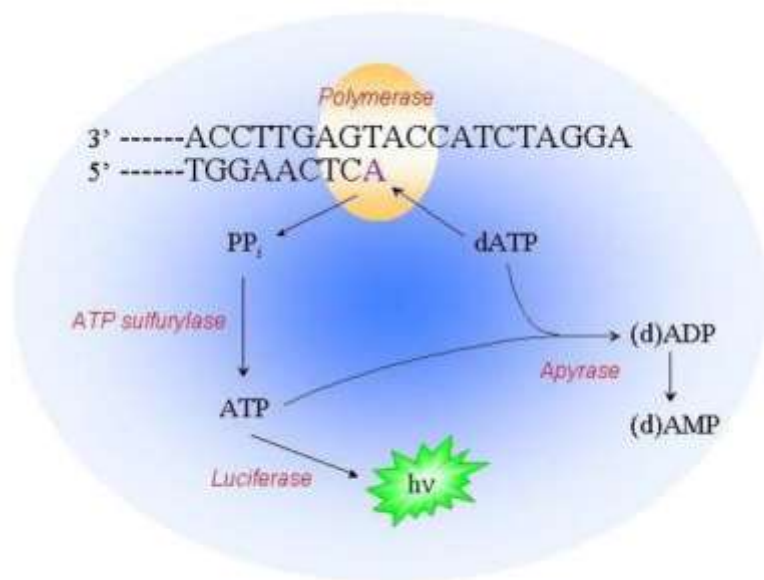
Barcoding



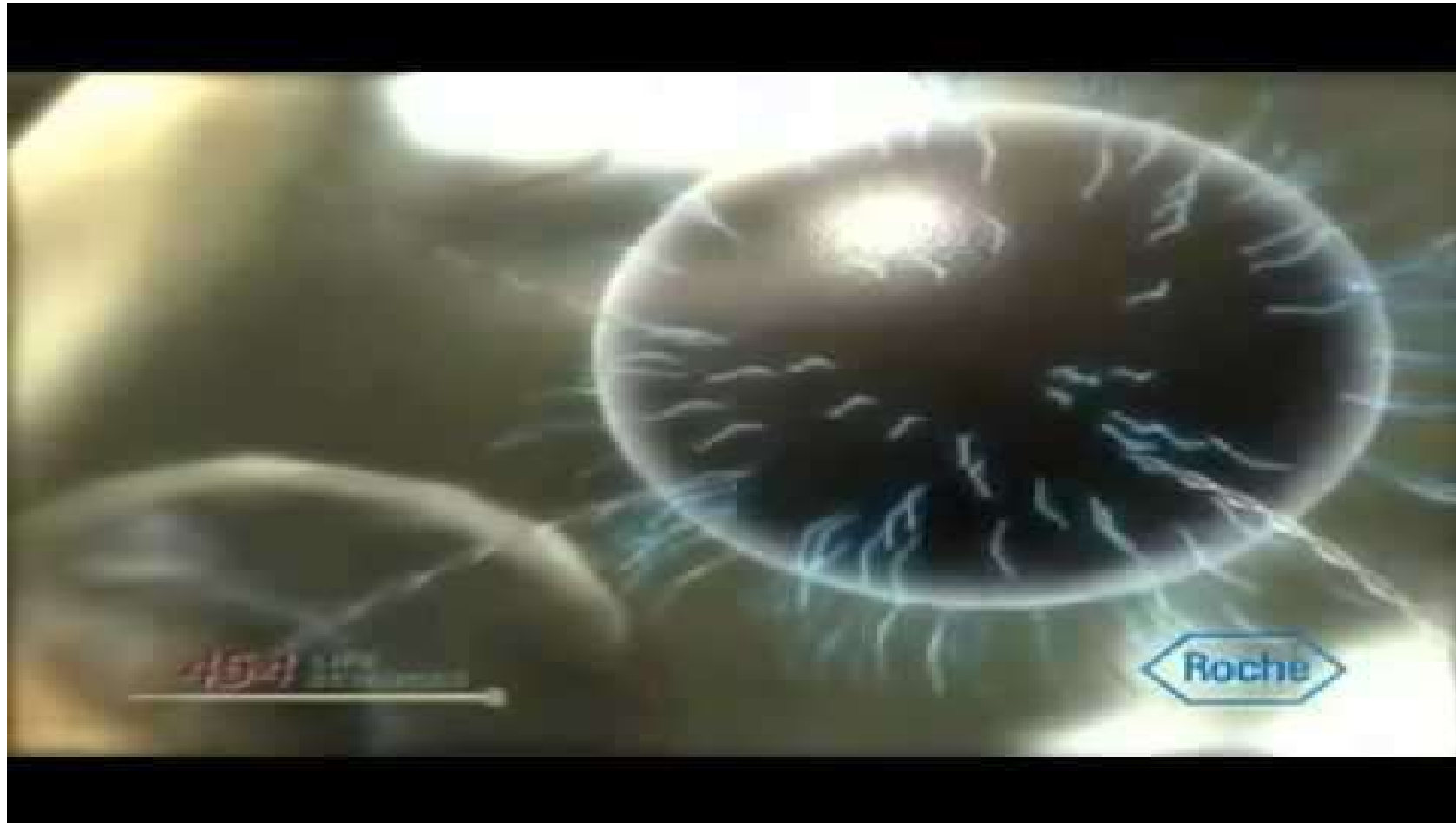
Up to 16 libraries

Pyrosekvenování (454-sekvenování)

- Sekvenování se syntézou DNA v reálném čase nevyžadující elektroforézu ani separaci fragmentů
- Je založené na uvolnění pyrofosfátu (PP_i) při enzymatické syntéze DNA (Nyren et al., 1987)
- Namísto standardního dATP je používán α-thiosubstituovaný dATP, který je přijímán DNA polymerázou, ale nikoli luciferázou
- Používána pro identifikaci jednonukleotidových polymorfizmů (SNPs)



Pyrosekvenování (454-sekvenování)



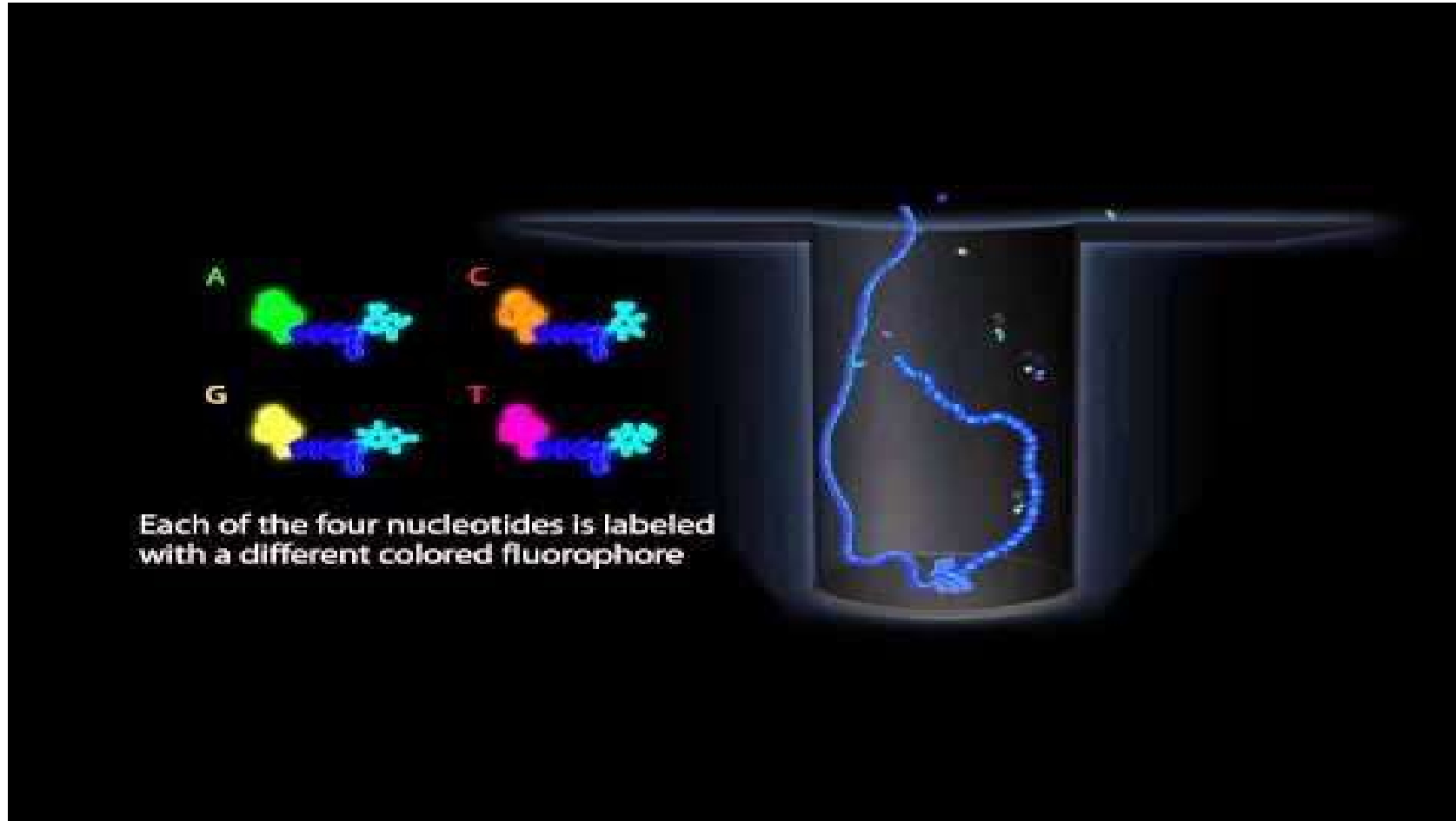
3rd generation sekvenování

- Vstupem každé jednotlivé sekvenační reakce je jediná molekula NK, jež je přímo sekvenována bez amplifikace
- Sekvenuje se paralelně více velmi dlouhých fragmentů (dlouhá čtení)

NanoPore

Pacific Biosciences – SMRT

PacBio - SMRT



Nanopore



Náklady na sekvenování

