

# From discovery to technology explosion

- 1868: Discovery of DNA
- 1953: Watson and Crick propose double helix structure
- 1977: Sanger sequencing
- 1985: PCR
- 2000: Working draft human genome announced (Sanger method)
- 2005: 454 sequencer launch (pyrosequencing)
- 2006: Genome Analyzer launched (Solexa sequencing)
- 2007: SOLiD launched (ligation sequencing)
- 2009: Whole human genome no longer merits Nature/Science paper
- 2010: “third-gen” systems

\$ human  
Genome

\$3 billion

\$2-3 million

\$250k

\$50k

\$20k

<\$1k





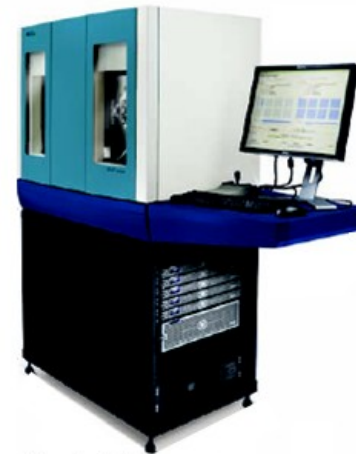
Applied Biosystems  
ABI 3730XL  
1 Mb / day



Roche / 454  
Genome Sequencer FLX  
100 Mb / run



Illumina / Solexa  
Genetic Analyzer  
2000 Mb / run



Applied Biosystems  
SOLiD  
3000 Mb / run



PACIFIC  
BIOSCIENCES®



ion torrent  
◊ \* △ ○ × □ + ∞

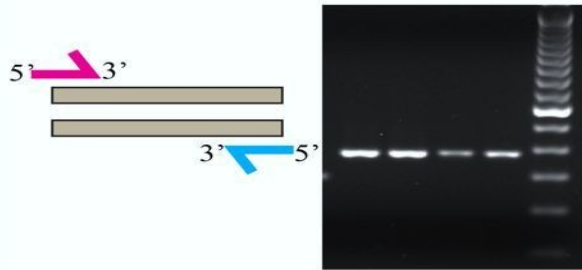


Oxford  
**NANOPORE**  
Technologies®

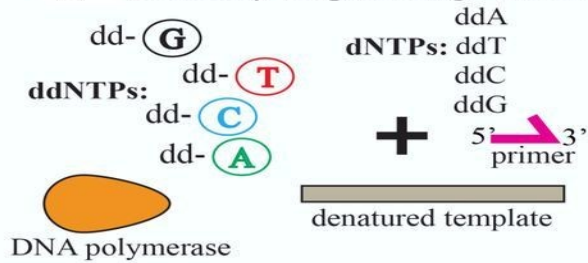


## Sanger Sequencing

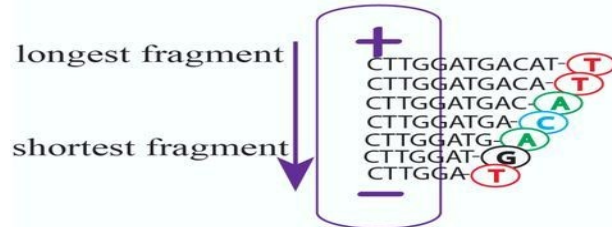
### A PCR



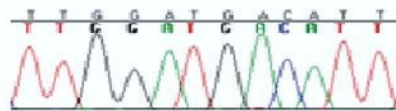
### B Dideoxy sequencing reaction



### C Capillary electrophoresis



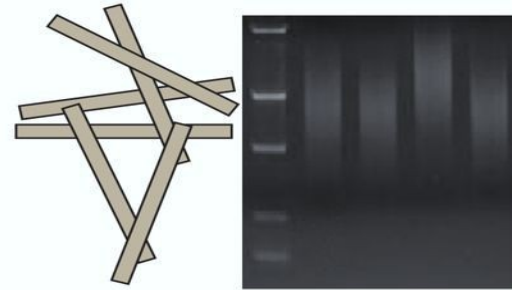
### D Sequencing read-out



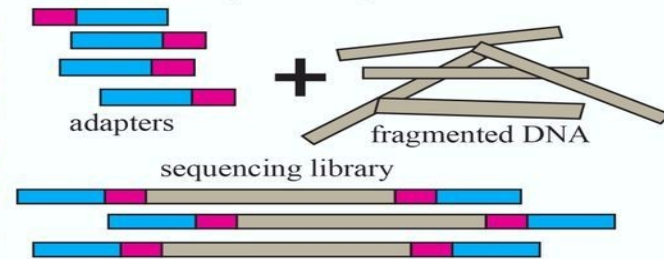
1 read/sample

## Next generation sequencing

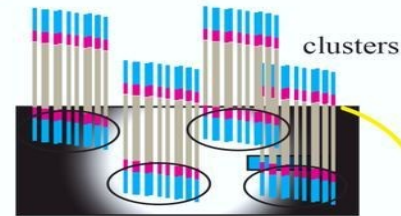
### A fragment gDNA



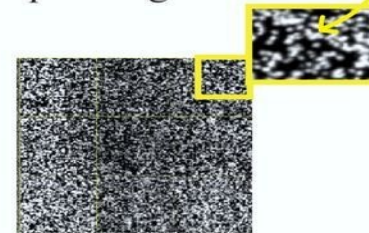
### B ligate adapters



### C Cluster generation



### D Sequencing



$40 \times 10^6$  reads/sample\*

Sequencer	454 GS FLX	HiSeq 2000	SOLiDv4	Sanger 3730xl
Sequencing mechanism	Pyrosequencing	Sequencing by synthesis	Ligation and two-base coding	Dideoxy chain termination
Read length	700 bp	50SE, 50PE, 101PE	50 + 35 bp or 50 + 50 bp	400~900 bp
Accuracy	99.9%*	98%, (100PE)	99.94% * raw data	99.999%
Reads	1 M	3 G	1200~1400 M	—
Output data/run	0.7 Gb	600 Gb	120 Gb	1.9~84 Kb
Time/run	24 Hours	3~10 Days	7 Days for SE 14 Days for PE	20 Mins~3 Hours
Advantage	Read length, fast	High throughput	Accuracy	High quality, long read length
Disadvantage	Error rate with polybase more than 6, high cost, low throughput	Short read assembly	Short read assembly	High cost low throughput

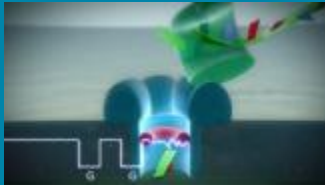


# Oxford Nanopore

**Application Specific**

**Adaptable protein nanopore:**

DNA Sequencing



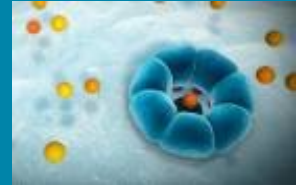
Proteins



Polymers



Small Molecules



**Generic Platform**



**Electronic read-out system**

# DNA degradation

**Mechanical damage** during tissue homogenization.

**Wrong pH and ionic strength** of extraction buffer.

Incomplete removal / contamination with **nucleases**.

**Phenol**: too old, or inappropriately buffered (**pH 7.8 – 8.0**); incomplete removal.

Wrong pH of **DNA solvent** (acidic water).

*Recommended: 1:10 TE for short-term storage, or 1xTE for long-term storage.*

**Vigorous pipetting** (wide-bore pipet tips).

**Vortexing** of DNA in high concentrations.

Too many **freeze-thaw** cycles (*we tested 5, still Ok*).

Debatable: sequence-dependent

# illumina, inc. (ILMN)

NasdaqGS - NasdaqGS Real Time Price. Currency in USD

☆ Add to watchlist

# 342.68

+19.95 (+6.18%)

At close: November 7 4:00PM EST

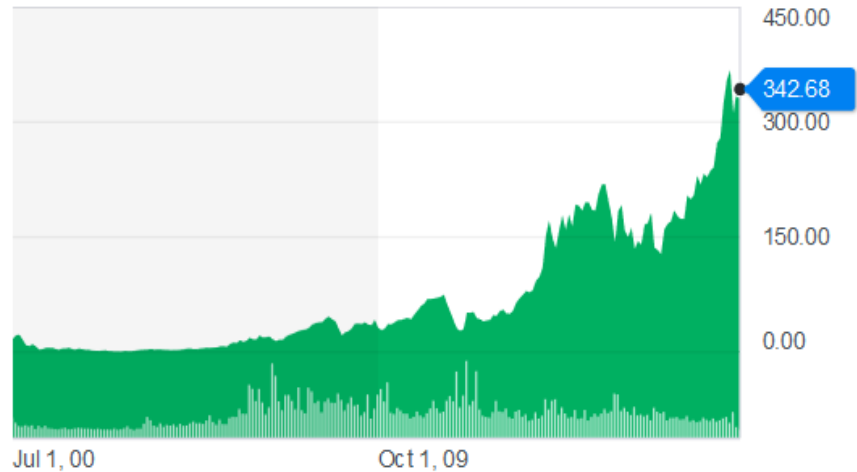
Buy

Sell

- Summary
- Chart
- Conversations
- Statistics
- Historical Data
- Profile
- Financials
- Analysis
- Options
- Holdings
- Sustainability

Previous Close	<b>322.73</b>	Market Cap	<b>50.374B</b>
Open	<b>325.00</b>	Beta (3Y Monthly)	<b>2.04</b>
Bid	<b>304.93 x 800</b>	PE Ratio (TTM)	<b>77.92</b>
Ask	<b>350.00 x 800</b>	EPS (TTM)	<b>4.40</b>
Day's Range	<b>318.38 - 343.04</b>	Earnings Date	<b>Jan 28, 2019 - Feb 1, 2019</b>
52 Week Range	<b>203.83 - 372.61</b>	Forward Dividend & Yield	<b>N/A (N/A)</b>
Volume	<b>1,273,158</b>	Ex-Dividend Date	<b>N/A</b>
Avg. Volume	<b>1,019,553</b>	1y Target Est	<b>359.75</b>

1D 5D 1M 6M YTD 1Y 5Y Max   Full screen



Trade prices are not sourced from all markets



# Genome sequencing

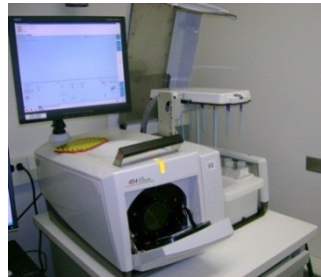
## Two strategies

- Whole genome shotgun (bottom-top)
- Clone-by-clone (top-bottom)



# Sequencing without a limit?

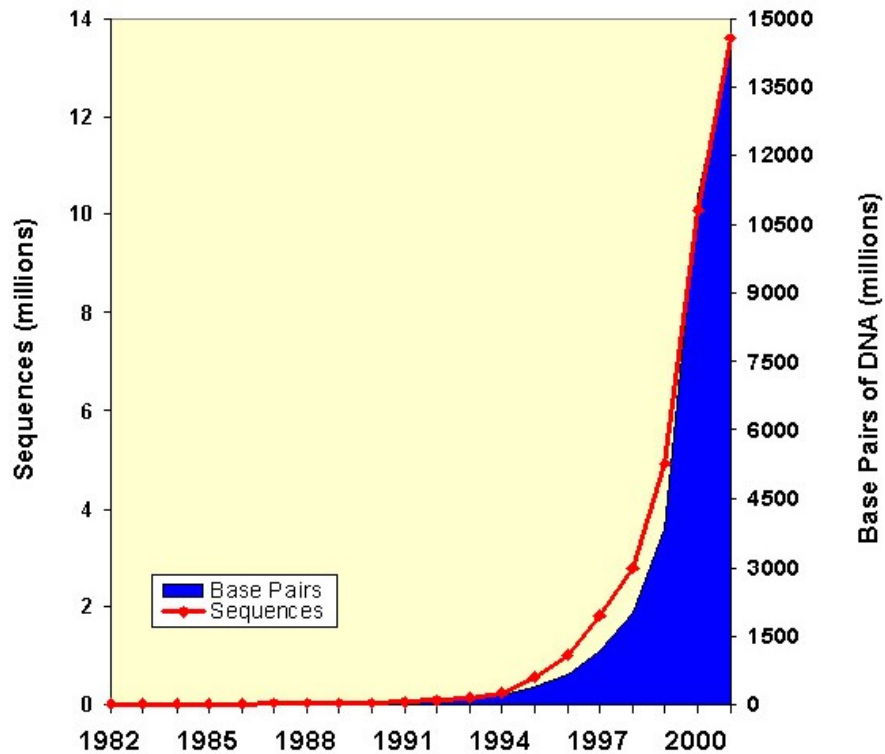
- A rapid progress in next generation sequencing technologies promises to provide complete (reference) DNA sequences



- **The bottleneck:**
  - NOT the sequencing capacity
  - BUT the ability to assemble many short reads with prevalence of repeated DNA (and polyploidy)

# Genome sequencing

## GenBank 1982 Los Alamos Sequence Database



Walter Goad

# Frederick Sanger

**1958 – Nobel prize – insuline structure**

**1975 - Dideoxy sequencing method**

**1977 –  $\Phi$ -X174 (5,368 bp) sequence**

**1980 – second Nobel prize**

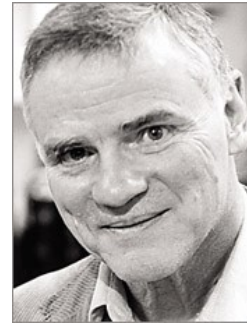
**$\lambda$  phage sequence**

**shotgun method (48,502 bp)**



# Genome sequencing

- **1986** Leroy Hood:  
automatic sequencing machine
- **1986** Human Genome Initiative



Leroy Hood



# Genome sequencing

- **1995** John Craig Venter  
first bacterial genome



John Craig Venter

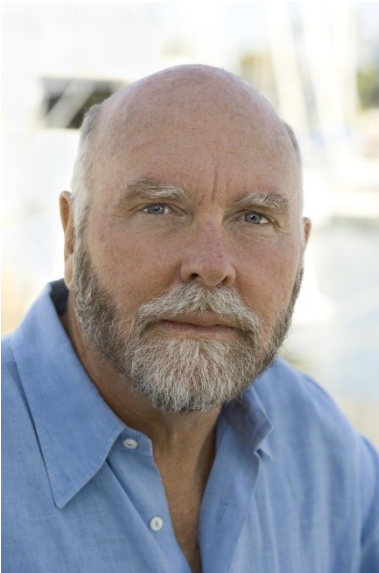


# Craig Venter

Global Ocean Sampling Expedition

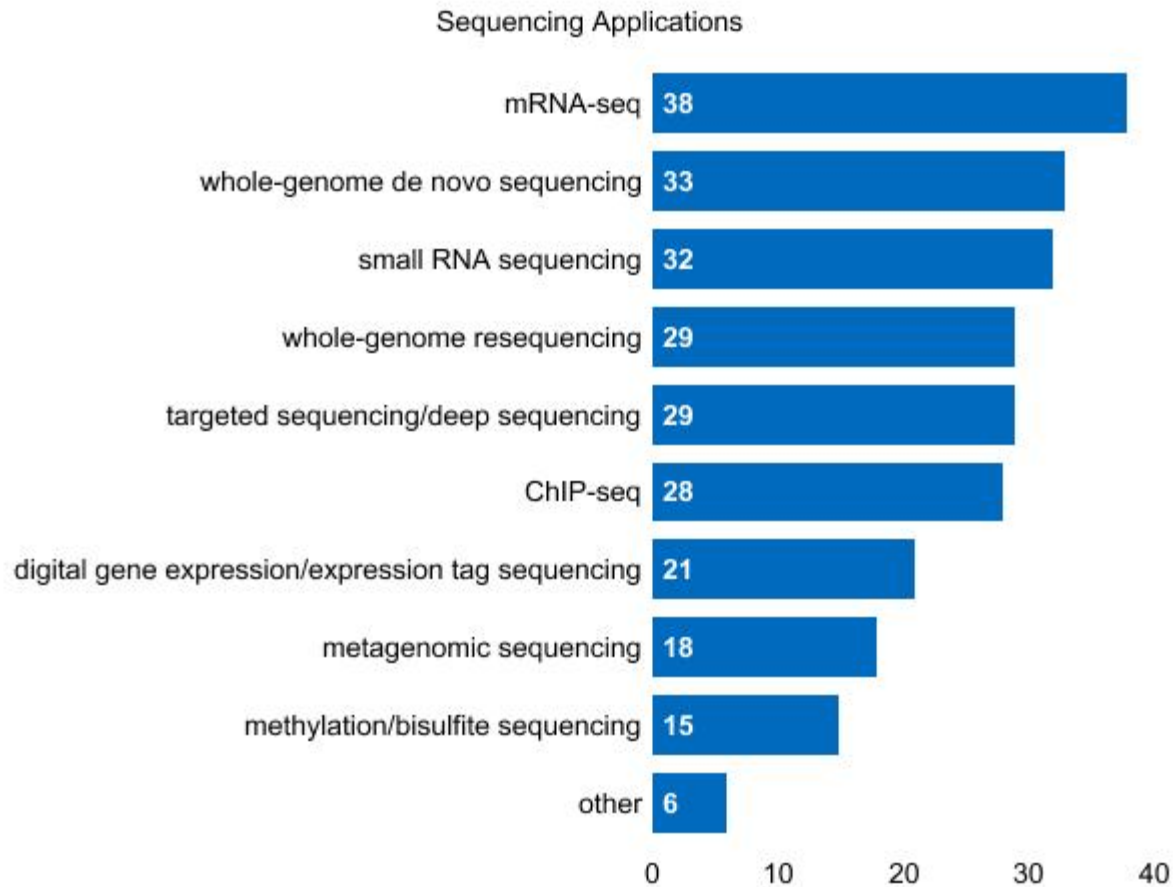
Synthetic genomics

Human Longevity Inc



<http://www.youtube.com/watch?v=J0rDFbrhjtI>

# Which applications are labs performing?



**2010 Human genome reference**

## 2010 Human genome reference



# 23andme (30% GSK)

[Anne Wojcicki](#) CEO - manželka spoluzakladatele Google  
**Sergey Mikhaylovich Brin**



23andMe

welcome health ancestry research how it works buy help Q

## Learn how your DNA may affect your health.

Our genes are a part of who we are, so naturally they impact our health. By knowing more about your DNA, you may be able to take steps towards living a healthier life.

Keep in mind that many conditions and traits are influenced by multiple factors. Our reports are intended for informational purposes only and do not diagnose disease or illness.

- **Plan for the future.**

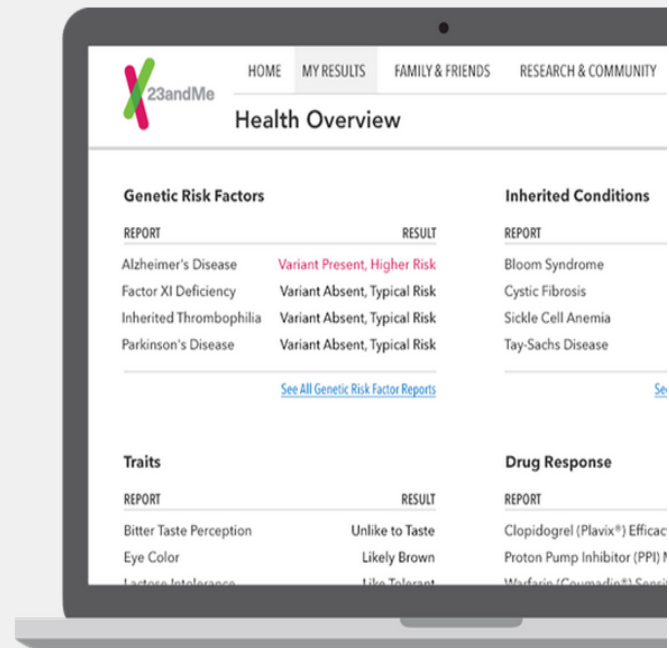
Learn if you are a carrier for certain inherited conditions, so you and your family can be prepared.

- **Stay one step ahead.**

Find out if you have certain genetic risk factors, so you can make better lifestyle choices and appropriately monitor your health.

- **Engage in your health care.**

Understand how your DNA may affect your health and response to







## MATERNAL LINE: H1

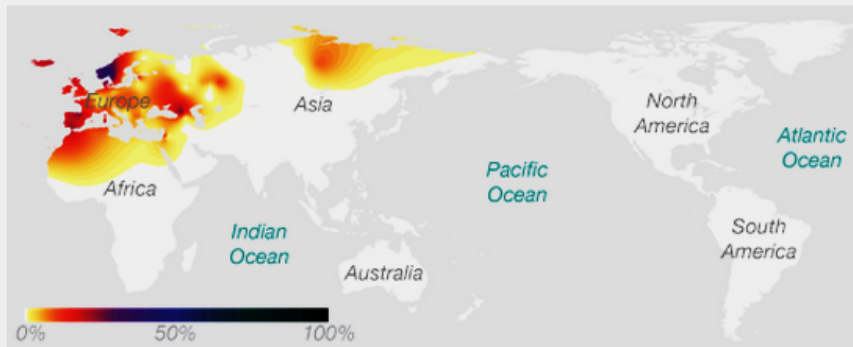
Overview

History

Haplogroup Tree

Community

Locations of haplogroup H1 before the widespread migrations of the past few hundred years.



Haplogroup H1 is widespread in Europe, especially the western part of the continent. It originated about 13,000 years ago, not long after the Ice Age ended.

Maternal haplogroups are families of mitochondrial DNA types that all trace back to a single mutation at a specific place and time. By looking at the geographic distribution of mtDNA types, we learn how our ancient female ancestors migrated throughout the world.

**Haplogroup:** H1, a subgroup of H

**Age:** 13,000 years

**Region:** Europe, Near East, Central Asia, Northwestern Africa

**Example Populations:** Spanish, Berbers, Lebanese

**Highlight:** H1 appears to have been common in Doggerland, an ancient land now flooded by the North Sea.

## PATERNAL LINE: I1\*

Overview

History

Haplogroup Tree

Community

## I1\* is a subgroup of I1

Locations of haplogroup I1 before the widespread migrations of the past few hundred years.



Haplogroup I1 can be found at levels of 10% and higher in many parts of Europe, due to its expansion with men who migrated northward after the end of the Ice Age about 12,000 years ago. It reaches its highest levels in Denmark and the southern parts of Sweden and Norway.

Paternal haplogroups are families of Y chromosomes that all trace back to a single mutation at a specific place and time. By looking at the geographic distribution of these related lineages, we learn how our ancient male ancestors migrated throughout the world.

**Haplogroup:** I1, a subgroup of I

**Age:** 28,000 years

**Region:** Northern Europe

**Example Populations:** Finns, Norwegians, Swedes

**Highlight:** Haplogroup I1 reaches highest frequencies in Scandinavia.

### Haplogroups of You and Your Connections

I1\*




Roman Hobza

### Haplogroups of Example Profiles

SHOW RESULTS FOR Roman Hobza ▾[SEE NEW AND RECENTLY UPDATED REPORTS »](#)

These reports provide information about your possible risk for developing certain health conditions based on genetics. Environmental and lifestyle factors also often play a large role in your risk for developing these conditions.

## Elevated Risk ?

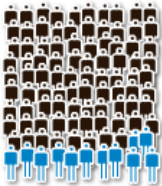
NAME	CONFIDENCE	YOUR RISK	AVG. RISK	COMPARED TO AVERAGE
Venous Thromboembolism	★★★★★	41.8%	12.3%	3.39x 
Gout	★★★★★	35.7%	22.8%	1.57x 
Melanoma	★★★★★	4.0%	2.9%	1.38x 
Restless Legs Syndrome	★★★★★	2.5%	2.0%	1.25x 
Exfoliation Glaucoma	★★★★★	2.2%	0.7%	2.90x 
Esophageal Squamous Cell Carcinoma (ESCC)	★★★★★	0.43%	0.36%	1.21x 
Stomach Cancer (Gastric Cardia Adenocarcinoma)	★★★★★	0.28%	0.23%	1.22x 
Primary Biliary Cirrhosis	★★★★★	0.11%	0.08%	1.43x 
Scleroderma (Limited Cutaneous Type)	★★★★★	0.08%	0.07%	1.24x 

Show information for  assuming  ethnicity and an age range of



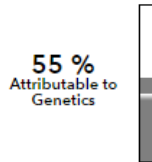
### Roman Hobza 41.8 out of 100

men of European ethnicity who share Roman Hobza's genotype will develop Venous Thromboembolism between the ages of 0 and 79.



### Average 12.3 out of 100

men of European ethnicity will develop Venous Thromboembolism between the ages of 0 and 79.



### Understanding Your Results

The heritability of venous thromboembolism is estimated to be 55%. This means that genetics (including unknown factors and known ones such as the SNPs we describe here) and environment play nearly equal roles in this condition. There are a number of environmental factors of various strengths that contribute to venous thromboembolism. Strong risk factors include hip or leg fractures, hip or knee replacement, major surgery or trauma, and spinal cord injury or surgery. Moderate risk factors include arthroscopic knee surgery, having central venous lines, congestive heart or respiratory failure, hormone replacement or oral contraceptive use, cancer, pregnancy, paralytic stroke, previous venous thromboembolism, and thrombophilia. Weak risk factors include bed rest for more than three days, immobility due to sitting (such as a long car or plane trip), specific types of chemotherapy, increasing age, laparoscopic surgery, obesity, and varicose veins. ([sources](#))

### What You Can Do

Assuming the ethnicity setting above is correct, your test results indicate you are at increased risk for venous thromboembolism based on genetics. Note that family history and non-genetic factors can also influence your risk for venous thromboembolism. Below are some steps you can take to reduce your risk.

---

**Gene or region:** F5  
**SNP:** rs6025

	SNP used	Genotype	Adjusted Odds Ratio*
Roman Hobza	rs6025	CT	European: 4.69
* Odds ratios are reported for all available ethnicities.			

Factor V is the last clotting factor in the pathway before the activation step that turns prothrombin into thrombin. Clotting is usually kept from spiraling out of control by a feedback loop, similar to the way a thermostat operates. Once enough thrombin has been activated, it binds to a protein called "protein C." Protein C then inactivates factor V, thus cutting off activation of prothrombin into thrombin.

The SNP in the F5 gene causes a change in the protein sequence of factor V that prevents protein C from inactivating it. Since this version of factor V can still participate in the activation of thrombin, a situation results in which thrombin can be turned on but cannot be turned off. Once the clotting cascade is set off (whether appropriately or not), the riskier version of the SNP makes it more difficult to shut it off.

The riskiness of the T version of this SNP is further increased for women who also take hormonal birth control.

(The riskier version of this gene is also sometimes called Factor V Leiden, after the city in the Netherlands where this SNP and its effects on factor V's role in clotting were first discovered.)

The studies whose data we report as applicable to those of "European" ancestry confirmed the association between this SNP and VTE in samples from the Netherlands, Sweden, the United Kingdom, Brazil, Italy, and France.

African and Asian populations appear to have only one version of the SNP, meaning that association studies are very difficult to perform.

#### Citations

[Rosendaal et al. \(1995\)](#). "High risk of thrombosis in patients homozygous for factor V Leiden (activated protein C resistance)." *Blood* 85(6):1504-8.

[Smith et al. \(2007\)](#). "Association of genetic variations with nonfatal venous thrombosis in postmenopausal women." *JAMA* 297(5):489-98.

[Emmerich et al. \(2001\)](#). "Combined effect of factor V Leiden and prothrombin 20210A on the risk of venous thromboembolism--pooled analysis of 8 case-control studies including 2310 cases and 3204 controls. Study Group for Pooled-Analysis in Venous Thromboembolism." *Thromb Haemost* 86(3):809-16.










[Bertina et al. \(1994\)](#). "Mutation in blood coagulation factor V associated with resistance to activated protein C." *Nature* 369(6475):64-7.

[Lane et al. \(2000\)](#). "Role of hemostatic gene polymorphisms in venous and arterial thrombotic disease." *Blood* 95(5):1517-32.

---






**Gene or region:** F2  
**SNP:** i3002432

## Decreased Risk ?

NAME	CONFIDENCE	YOUR RISK	AVG. RISK	COMPARED TO AVERAGE
Type 2 Diabetes	★★★★	17.7%	25.7%	0.69x 
Alzheimer's Disease	★★★★	4.3%	7.2%	0.60x 
Rheumatoid Arthritis	★★★★	1.6%	2.4%	0.68x 
Parkinson's Disease	★★★★	1.2%	1.6%	0.73x 
Age-related Macular Degeneration	★★★★	0.92%	6.55%	0.14x 
Crohn's Disease	★★★★	0.31%	0.53%	0.58x 
Multiple Sclerosis	★★★★	0.24%	0.34%	0.69x 
Type 1 Diabetes	★★★★	0.12%	1.02%	0.12x 
Celiac Disease	★★★★	0.05%	0.12%	0.44x 

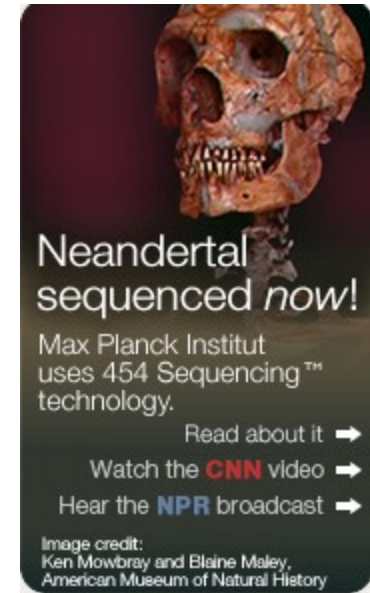


BRCA Cancer Mutations (Selected)	★★★★	Variant Absent
Beta Thalassemia	★★★★	Variant Absent
Bloom's Syndrome	★★★★	Variant Absent
Canavan Disease	★★★★	Variant Absent
Congenital Disorder of Glycosylation Type 1a (PMM2-CDG)	★★★★	Variant Absent
Connexin 26-Related Sensorineural Hearing Loss	★★★★	Variant Absent
Cystic Fibrosis	★★★★	Variant Absent
D-Bifunctional Protein Deficiency	★★★★	Variant Absent
DPD Deficiency	★★★★	Variant Absent
Dihydrolipoamide Dehydrogenase Deficiency	★★★★	Variant Absent
Factor XI Deficiency	★★★★	Variant Absent
Familial Dysautonomia	★★★★	Variant Absent
Familial Hypercholesterolemia Type B	★★★★	Variant Absent
Familial Hyperinsulinism (ABCC8-related)	★★★★	Variant Absent
Familial Mediterranean Fever	★★★★	Variant Absent
Fanconi Anemia (FANCC-related)	★★★★	Variant Absent
G6PD Deficiency	★★★★	Variant Absent

Reading Ability	***	Typical Nonword Reading Score
Response to Diet	***	See Report
Response to Exercise	***	See Report
Sex Hormone Regulation	***	See Report
Sweet Taste Preference 	***	See Report
Tooth Development	***	See Report
Tuberculosis Susceptibility	***	See Report
Breast Morphology  	***	Not Applicable
Menarche 	***	Not Applicable
Menopause 	***	Not Applicable
Eating Behavior	**	Greater tendency to overeat
HIV Progression	**	See Report
Hair Thickness	**	Typical, if European or African
Longevity	**	See Report
Measures of Intelligence	**	Lower Non-Verbal IQ
Memory	**	Typical Episodic Memory
Odor Detection	**	Typical Sensitivity to Sweaty Odor
Pain Sensitivity	**	Increased
Avoidance of Errors	*	See Report

# Genome Sequencer 20 System 454 pyrosequencing (2005)

- <http://www.454.com>



# DNA library preparation

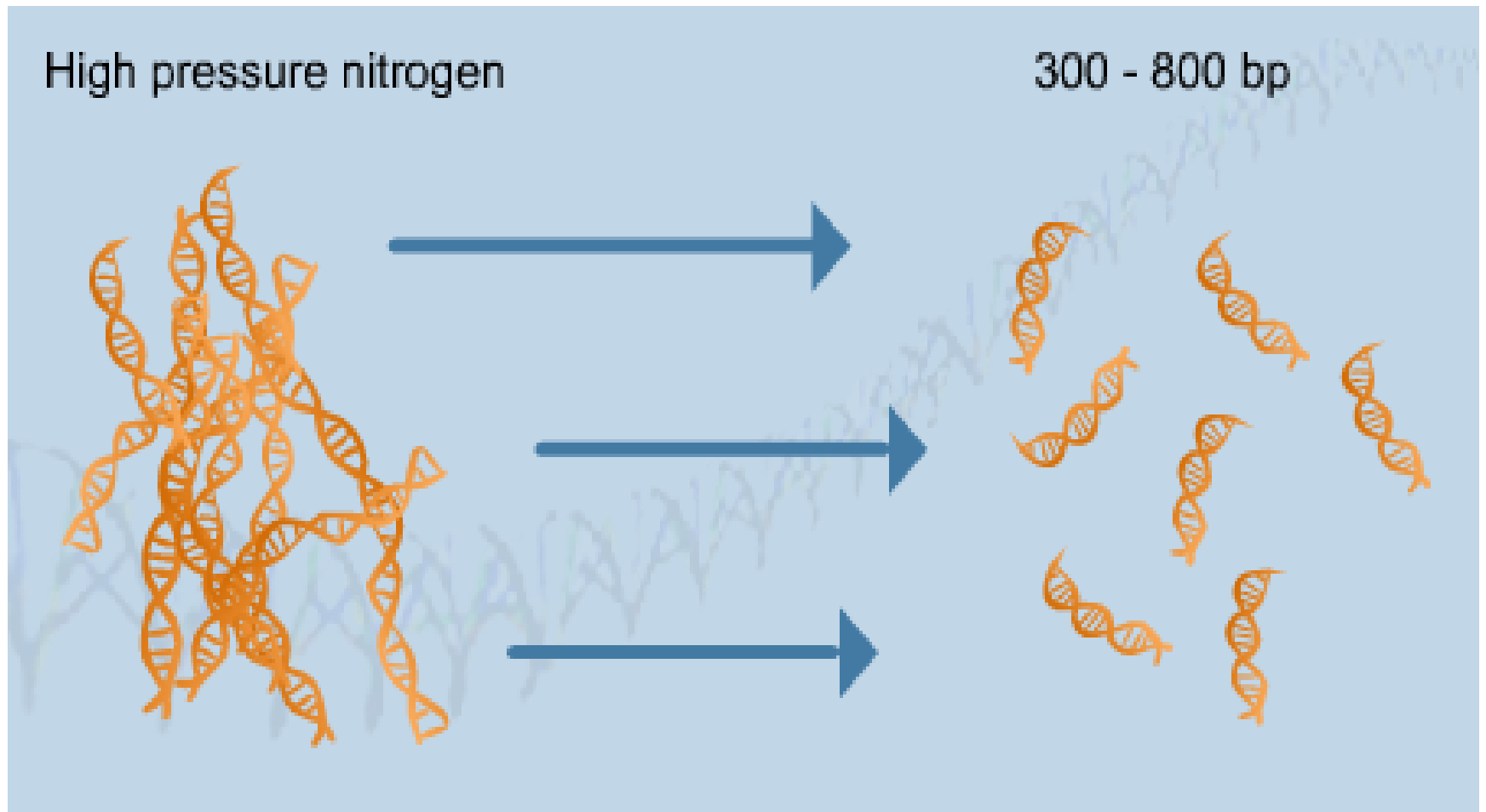
One sample preparation per genome

No Cloning

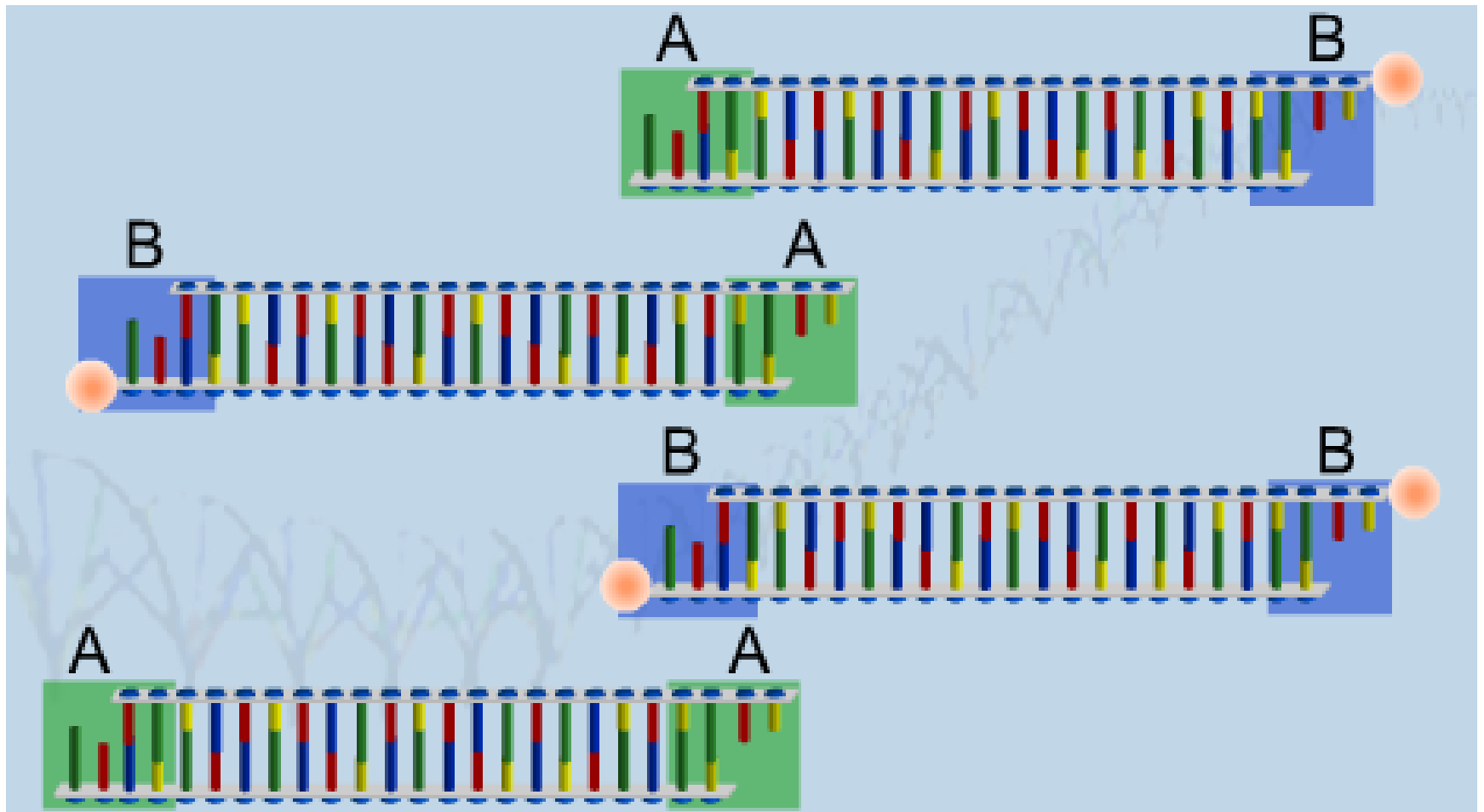
No Colony Picking



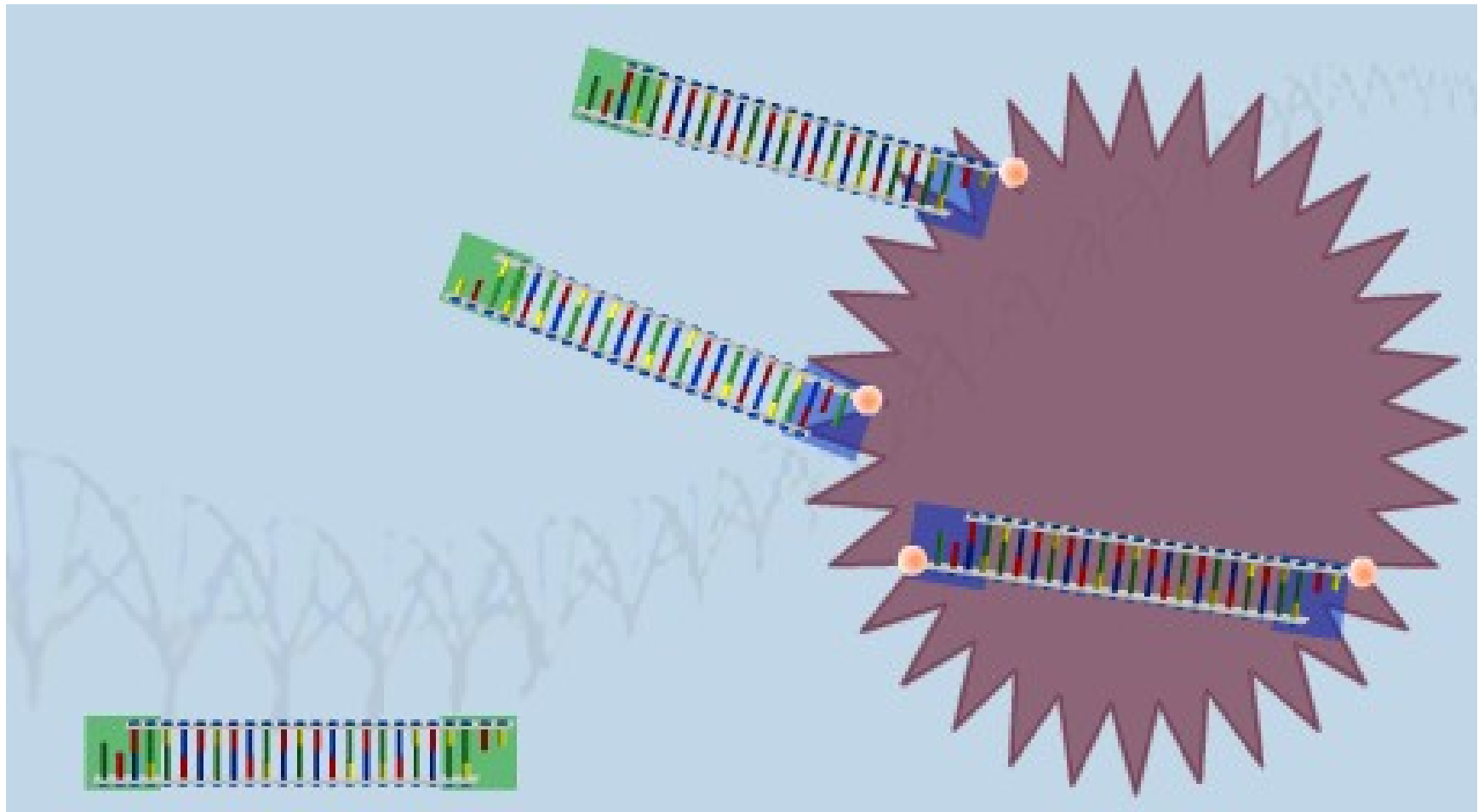
# Fragmentace DNA



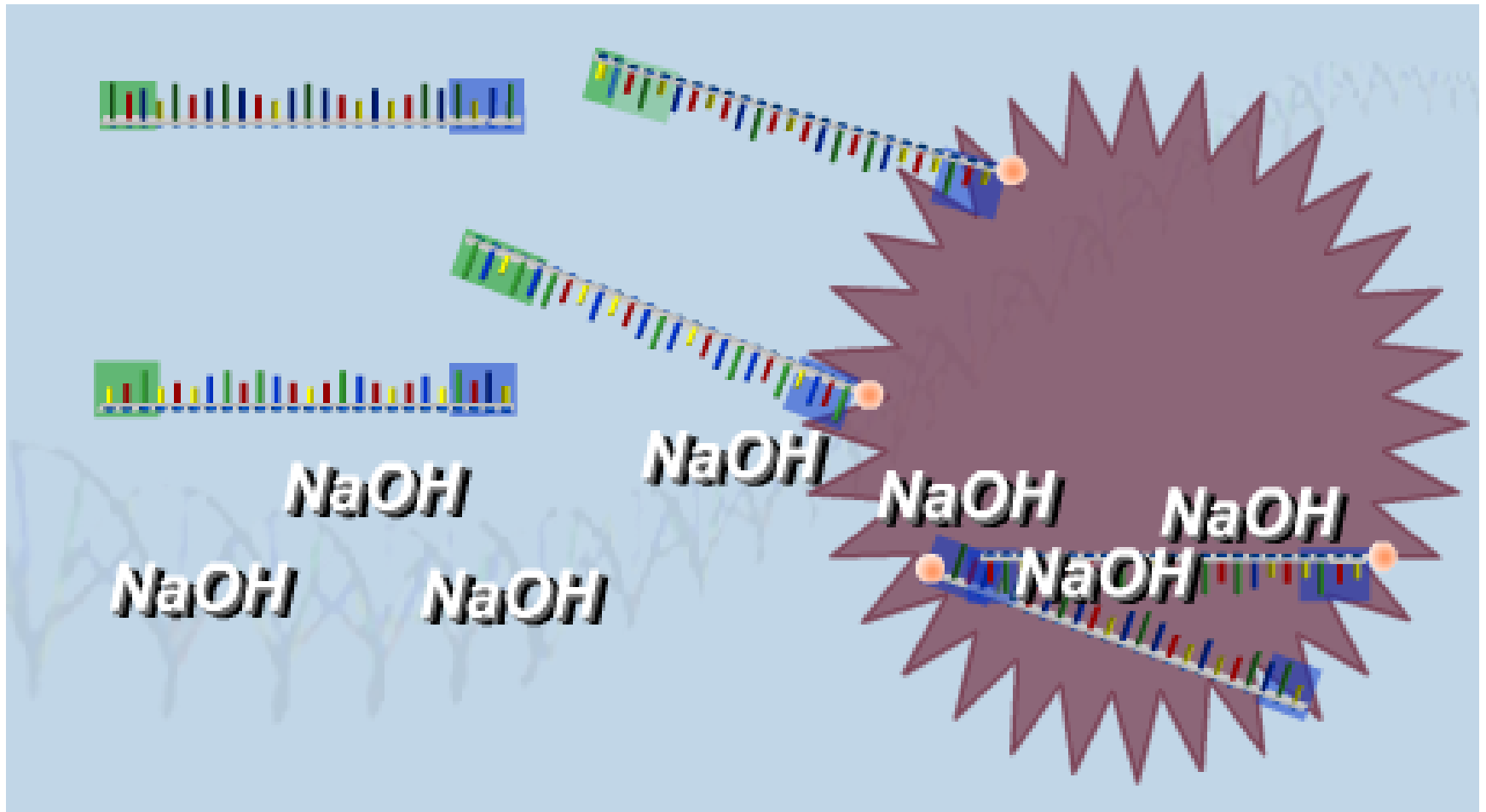
# Ligace adaptoru



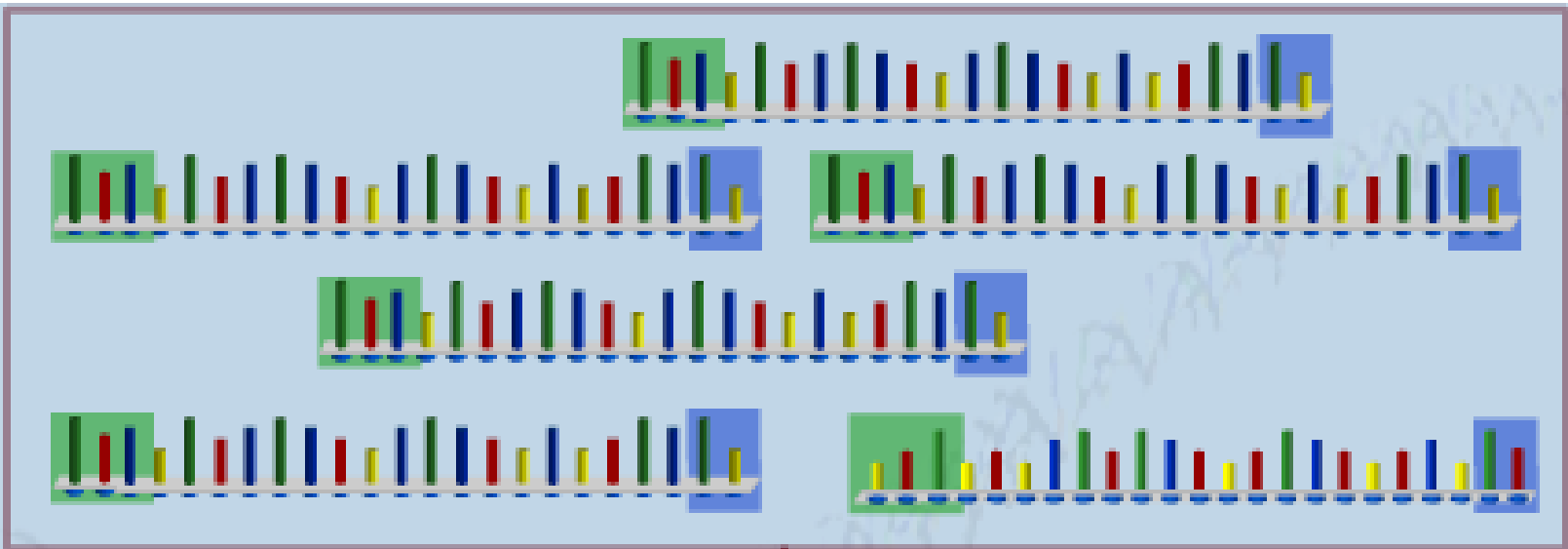
# Vychytání DNA molekul



# denaturace

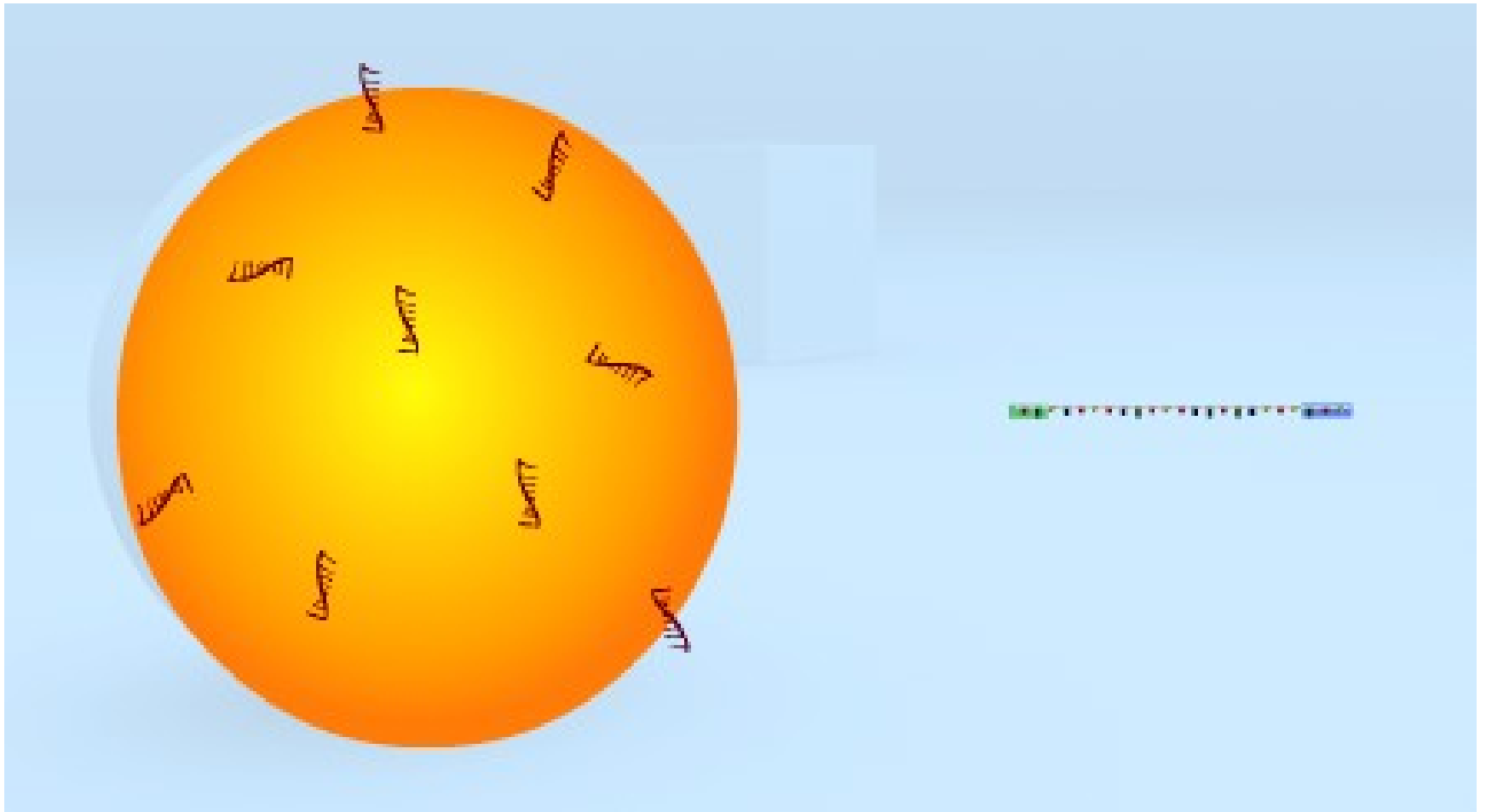




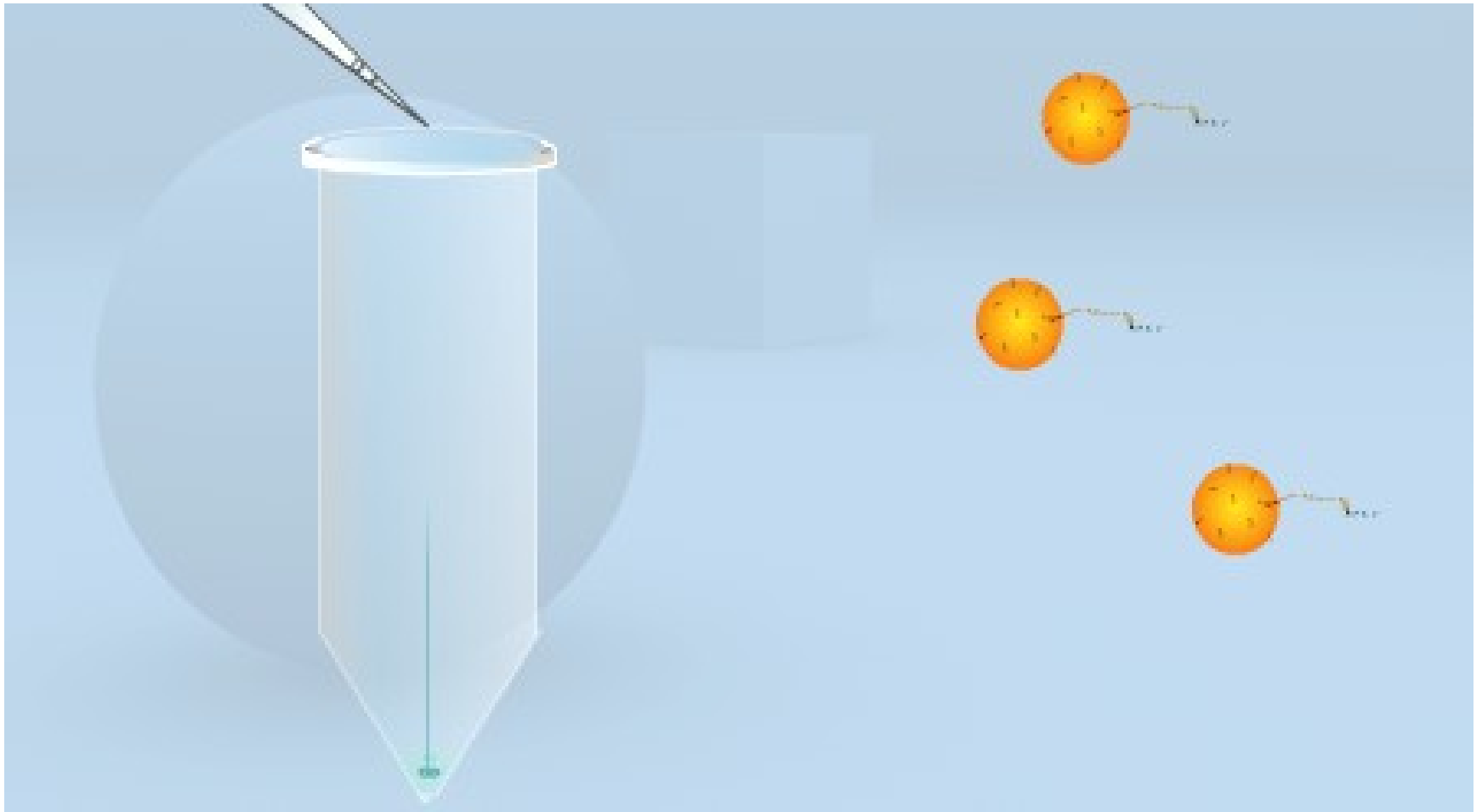


Single-stranded  
template DNA  
sstDNA

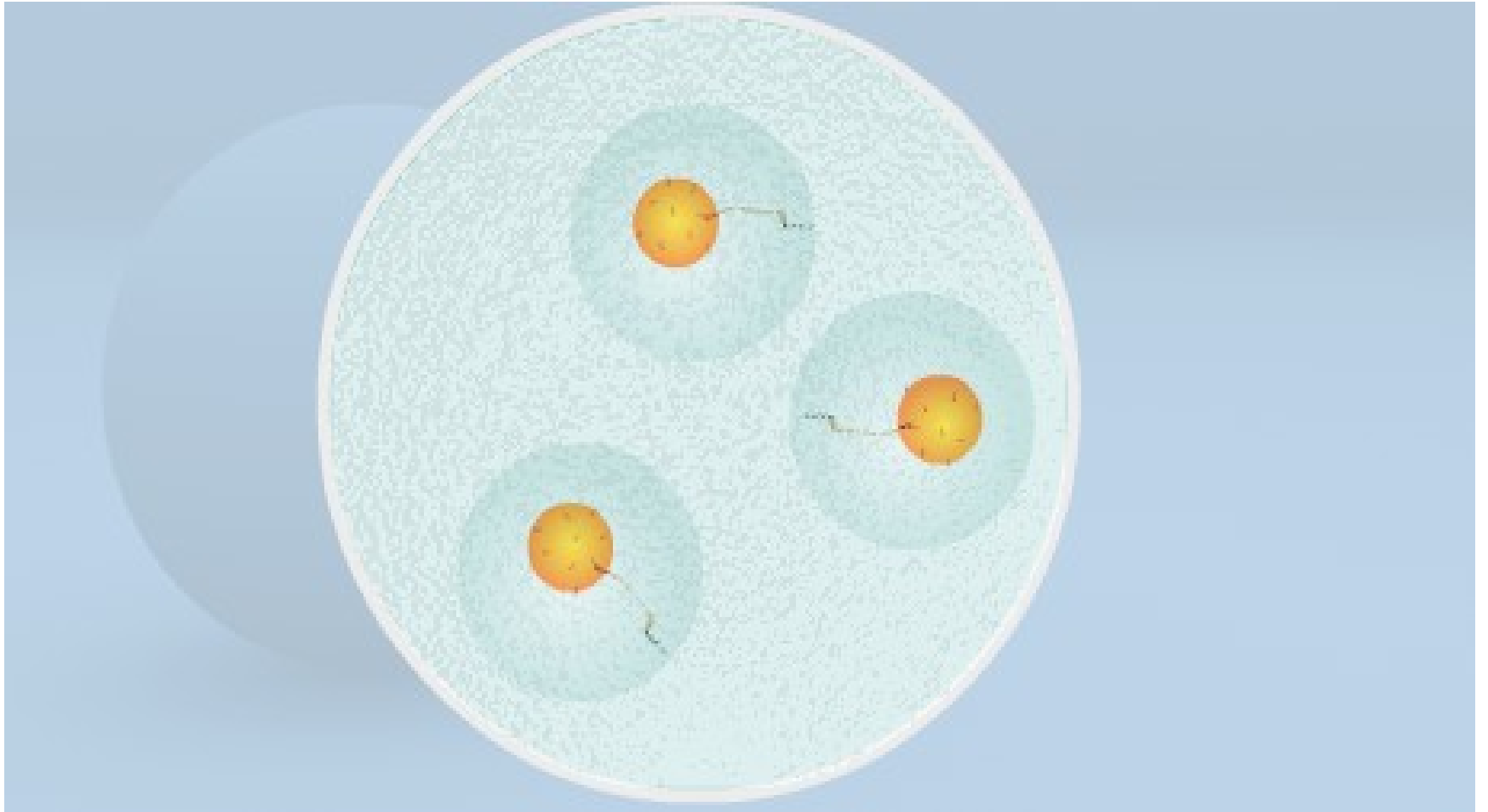
# emPCR



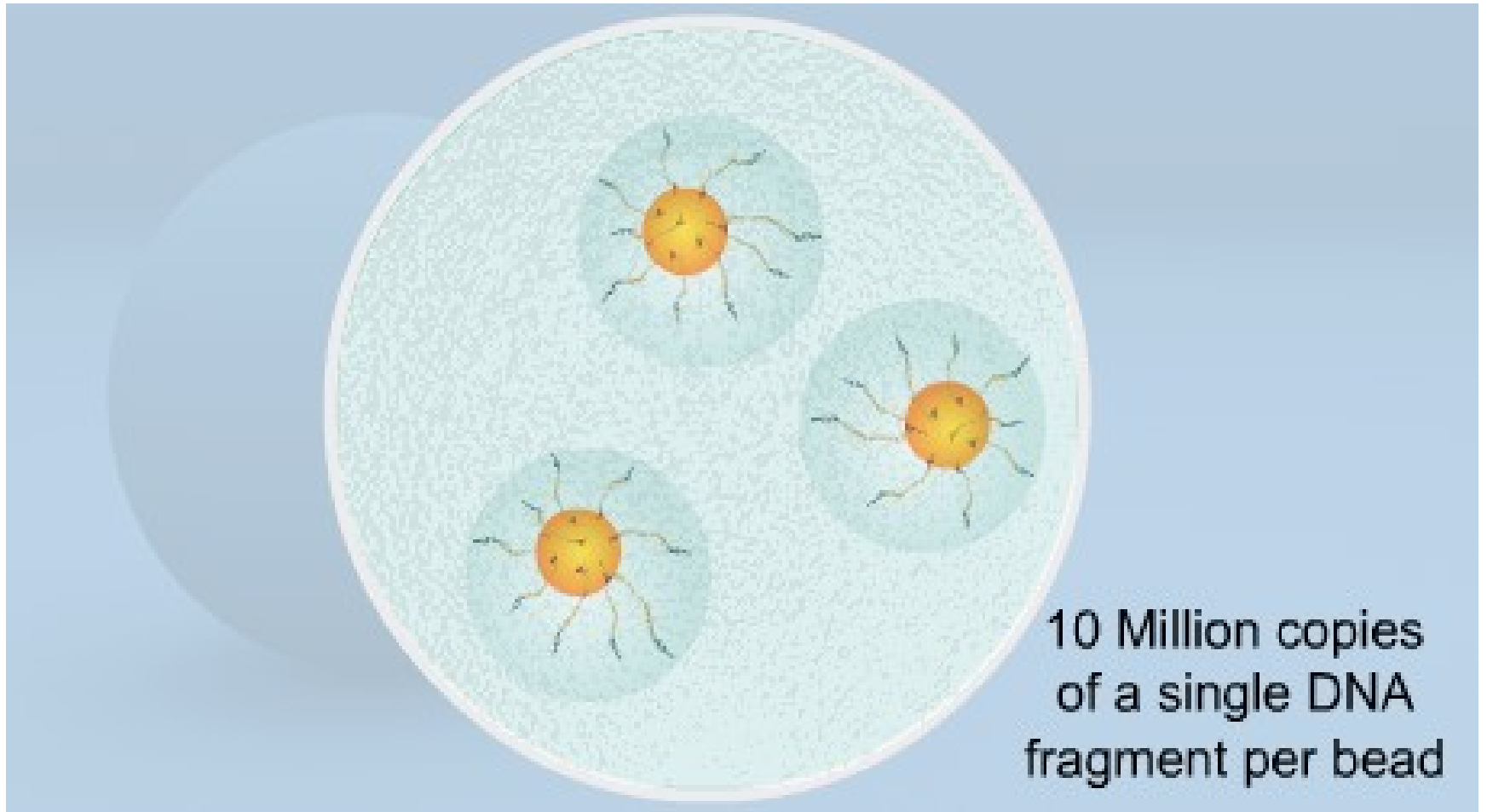
# Vznik emulze (olej)



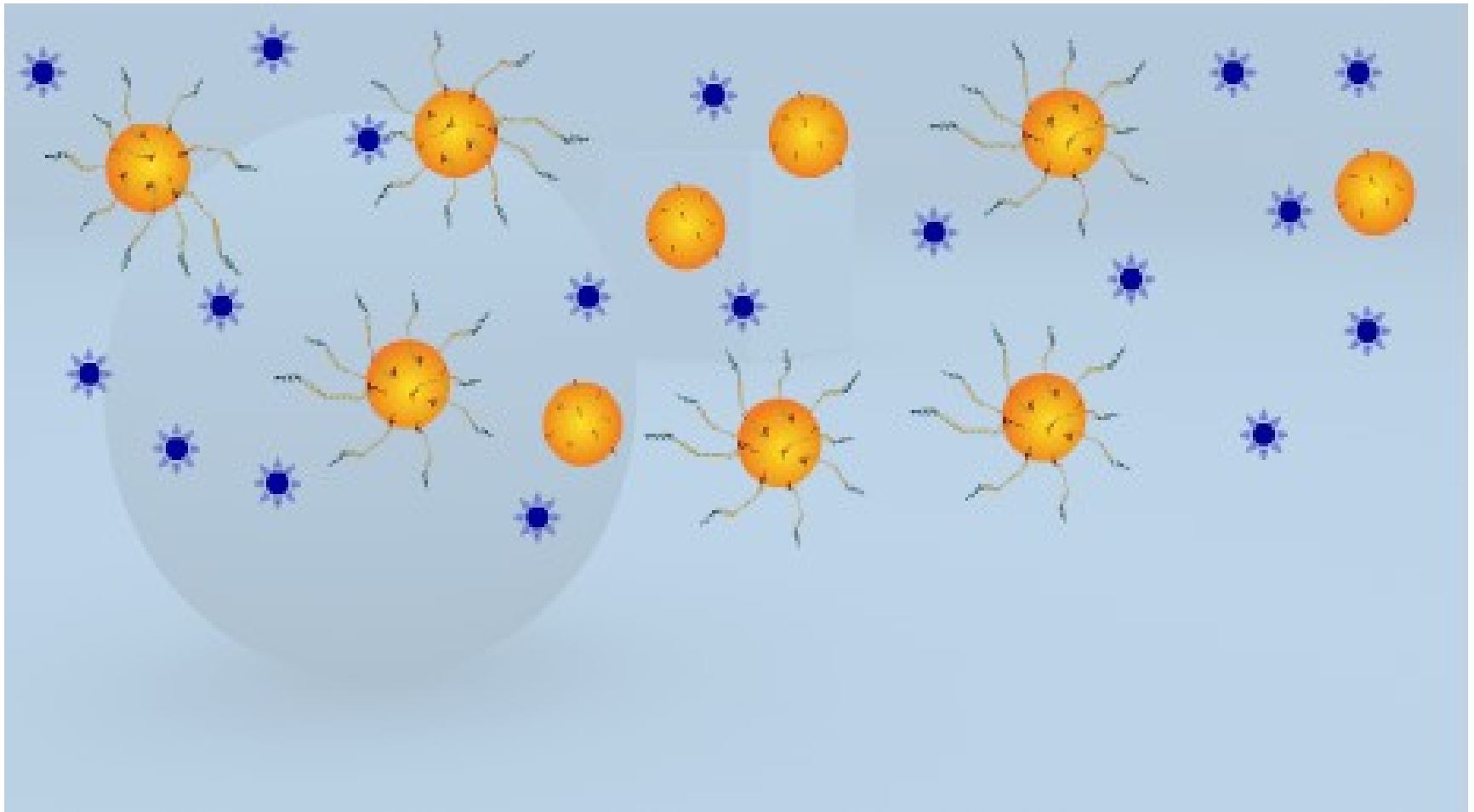
# emPCR



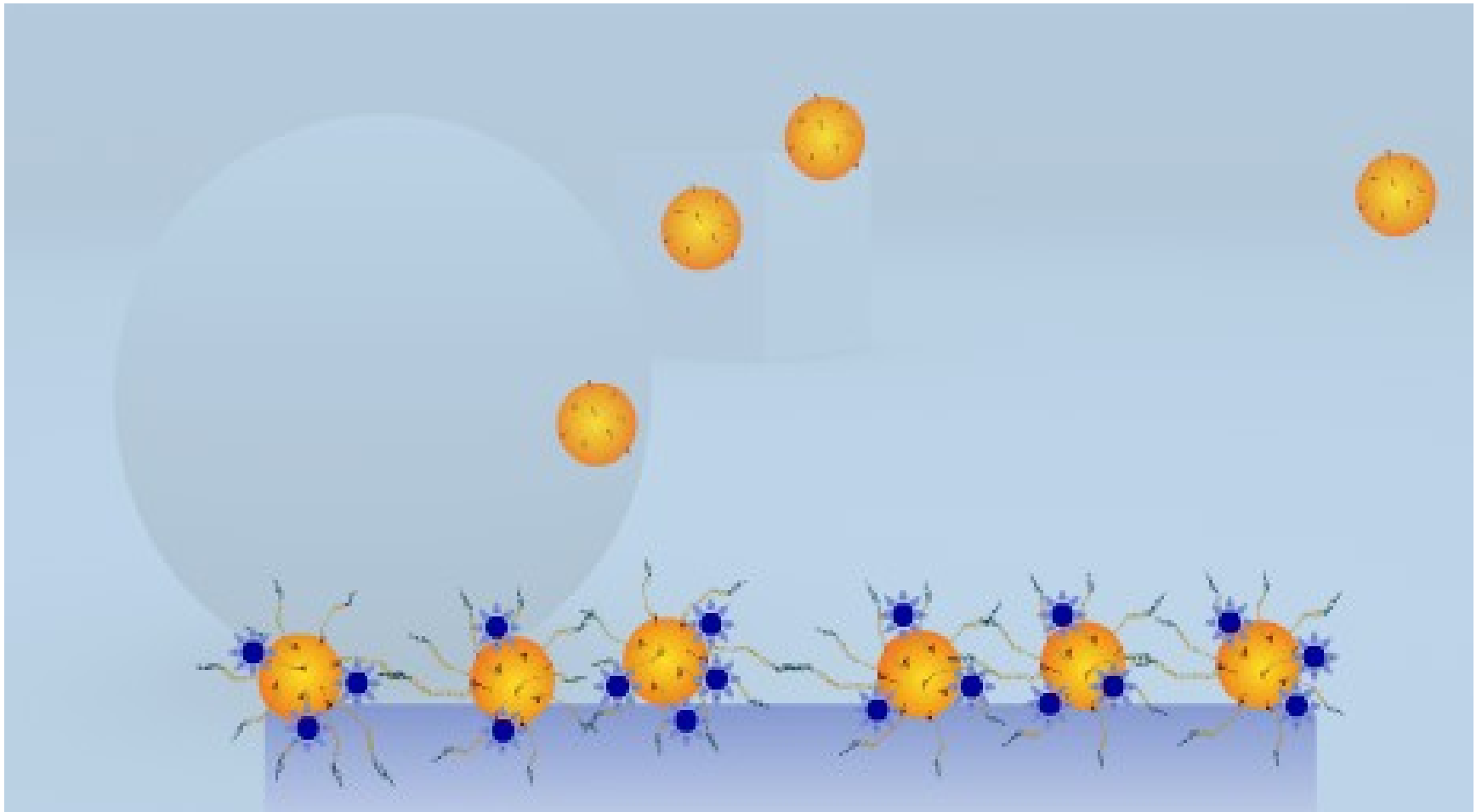
# emPCR



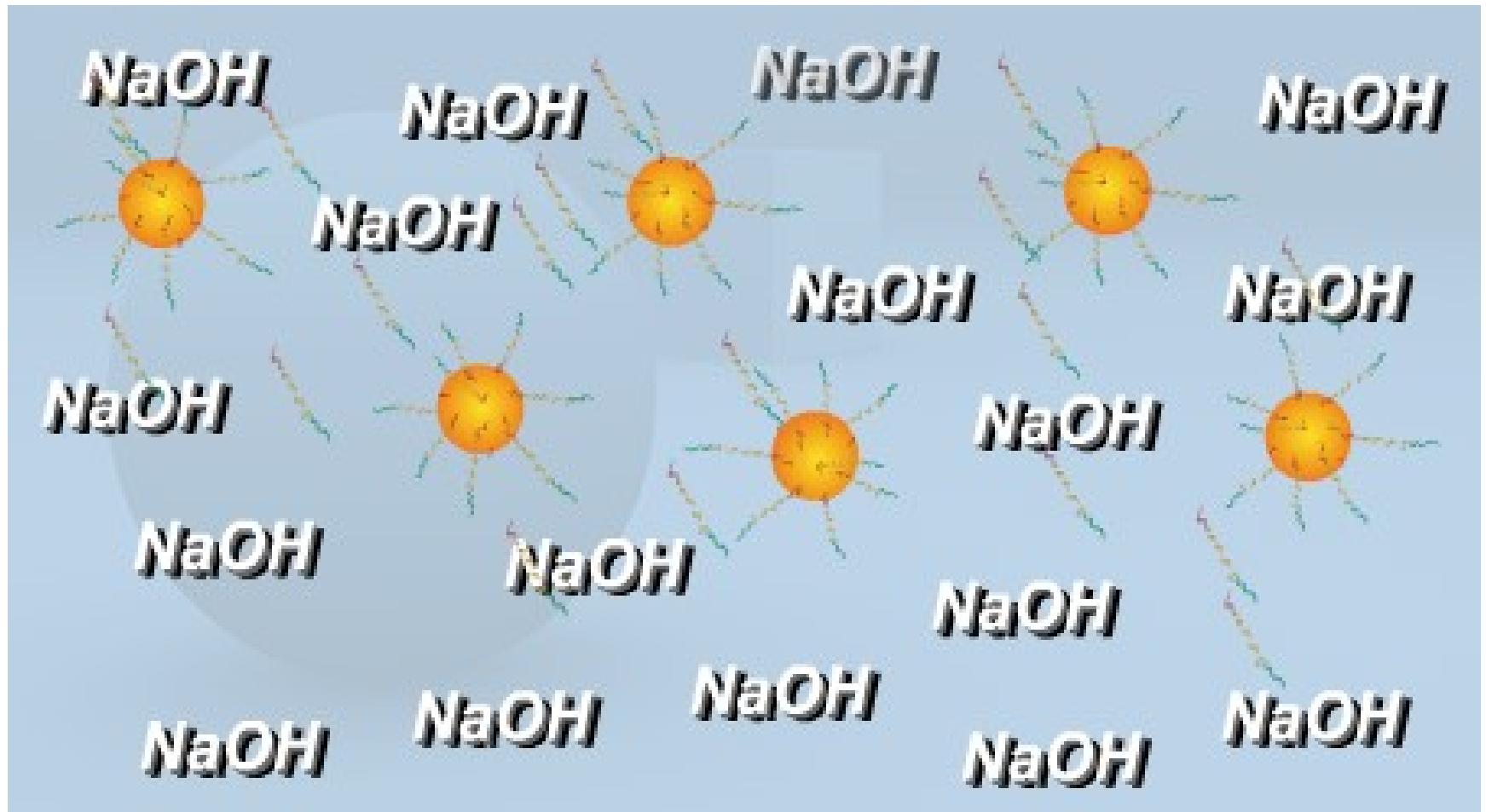
# Vychytání kuliček



# Vychytání kuliček

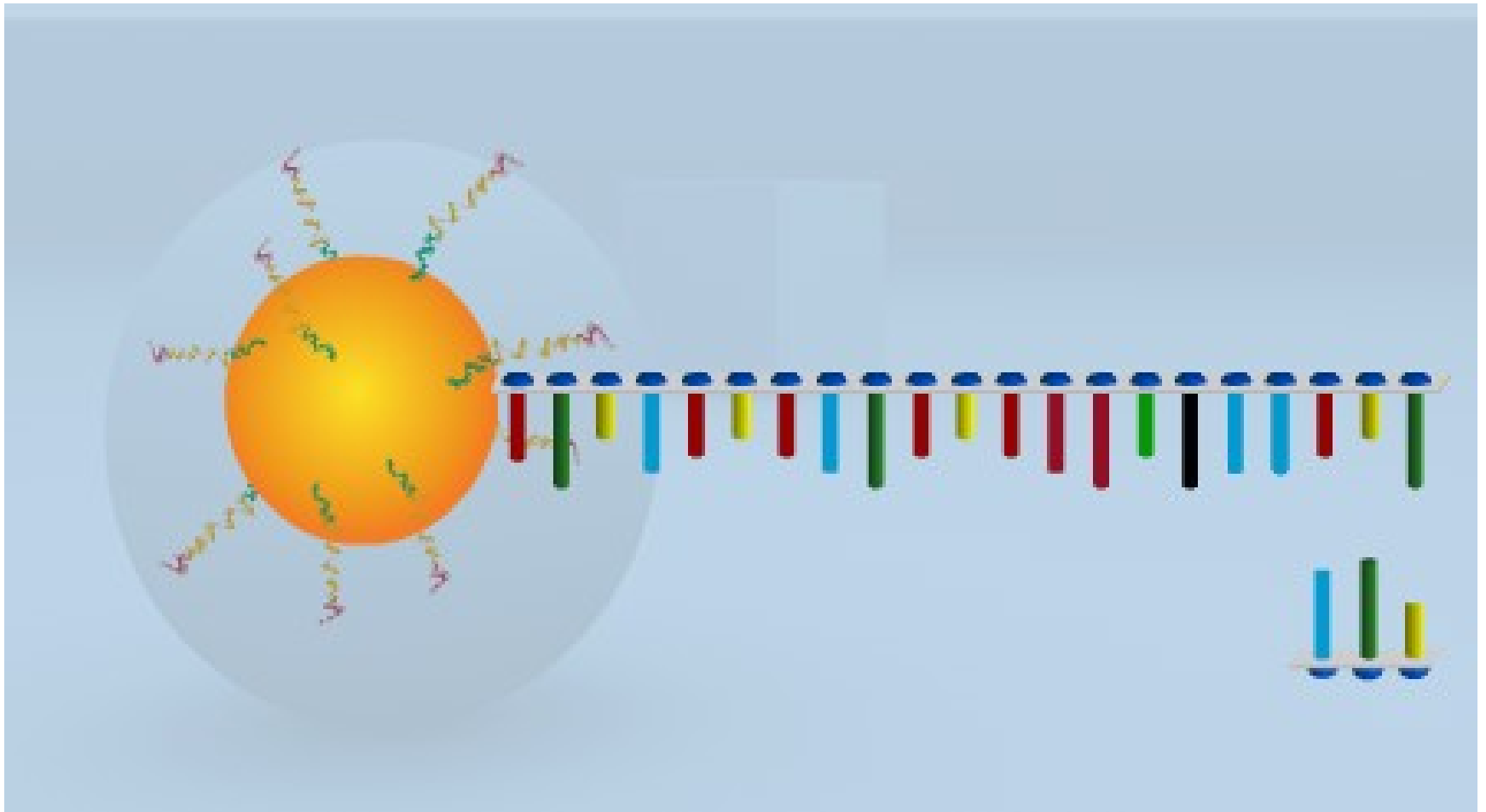


# denaturace

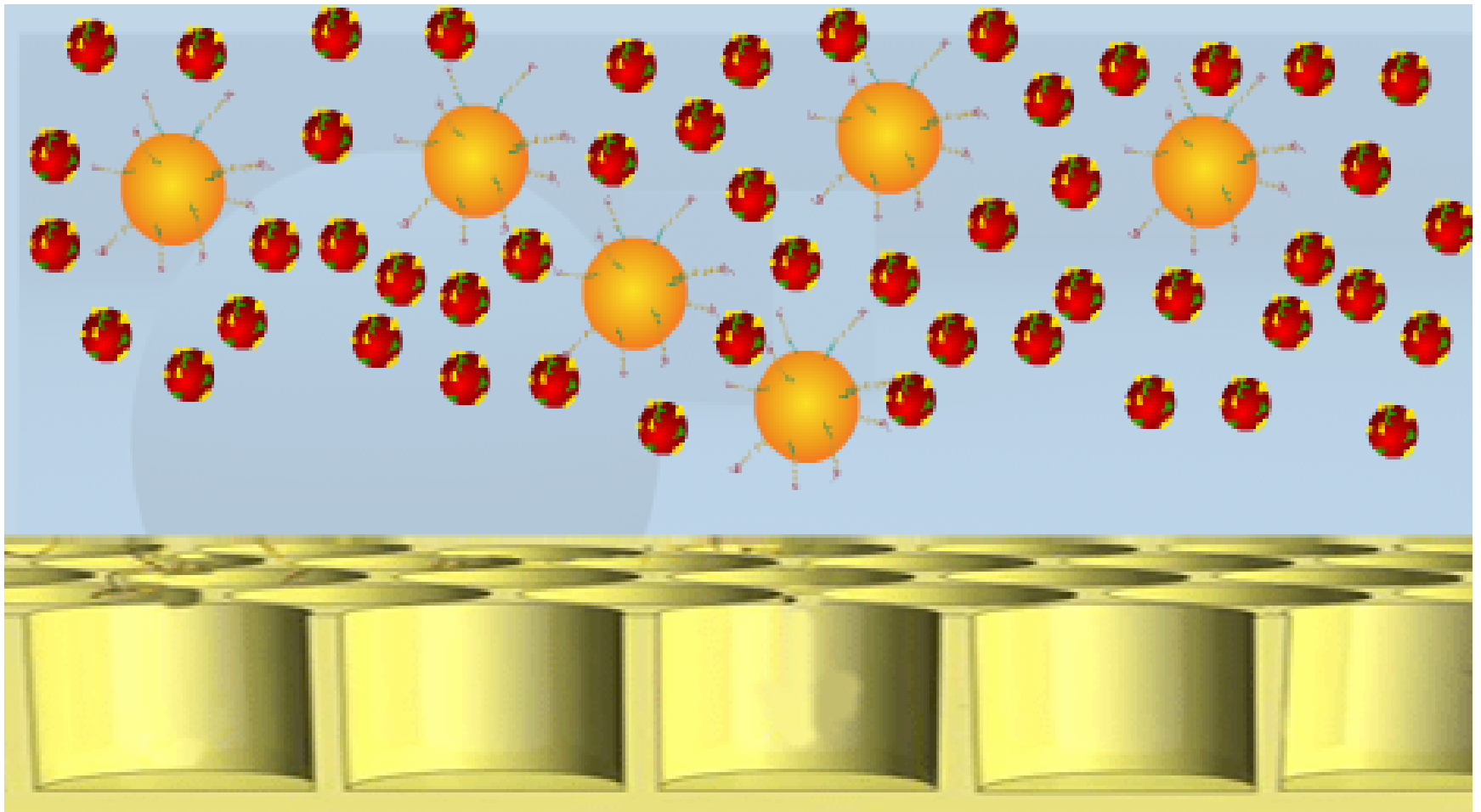




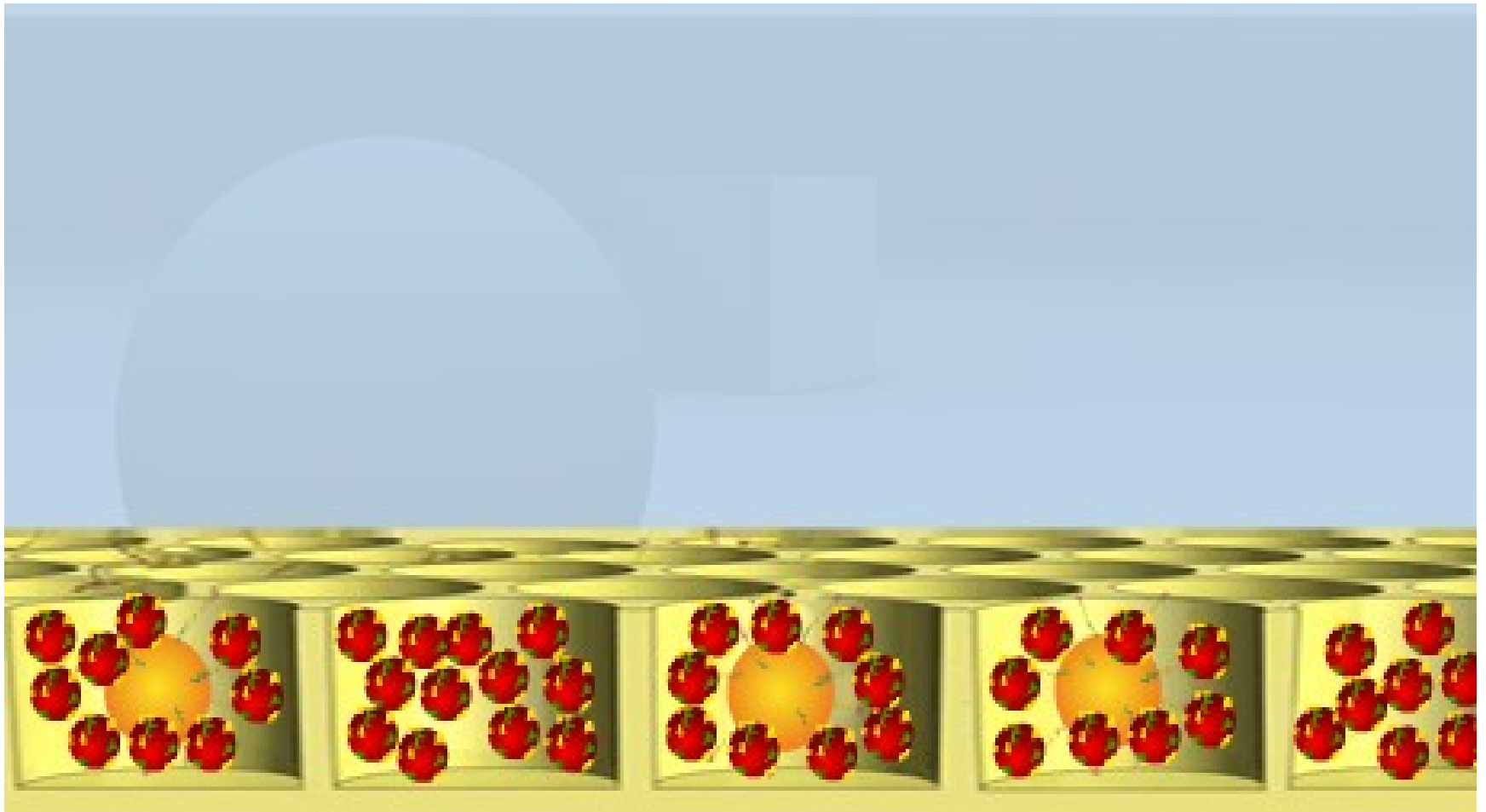
# Sekvenační primer



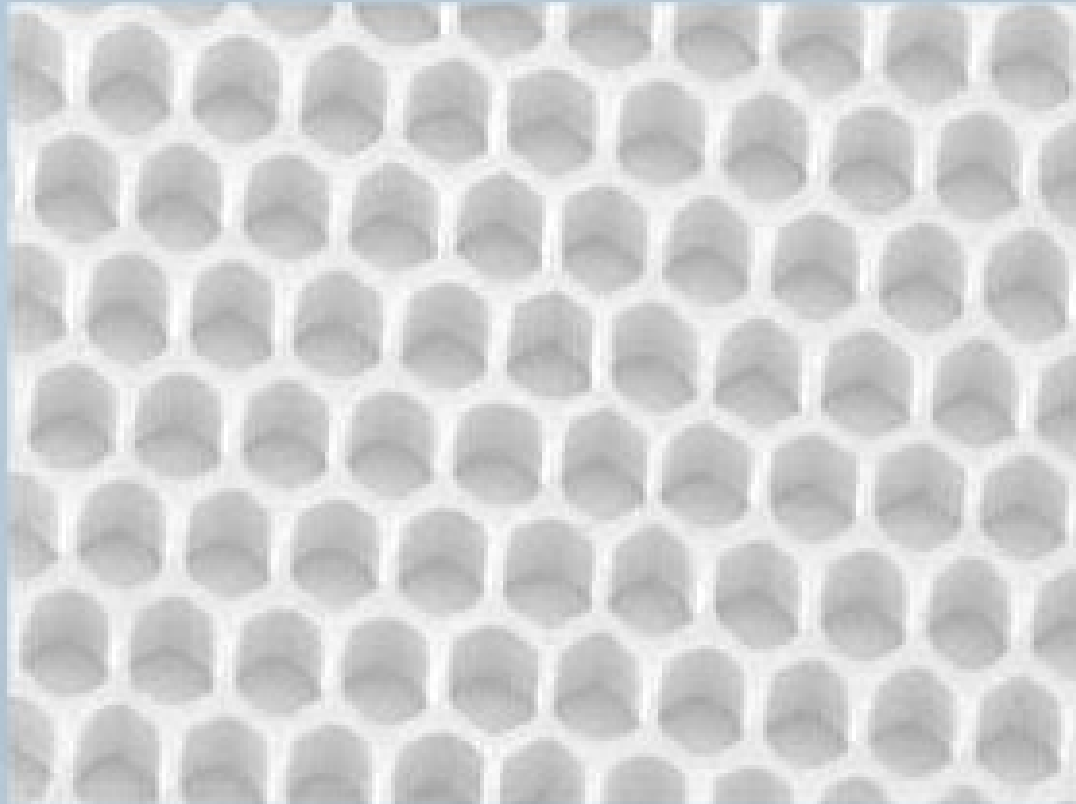
# Disperze na sklíčko



# Disperze na sklíčko



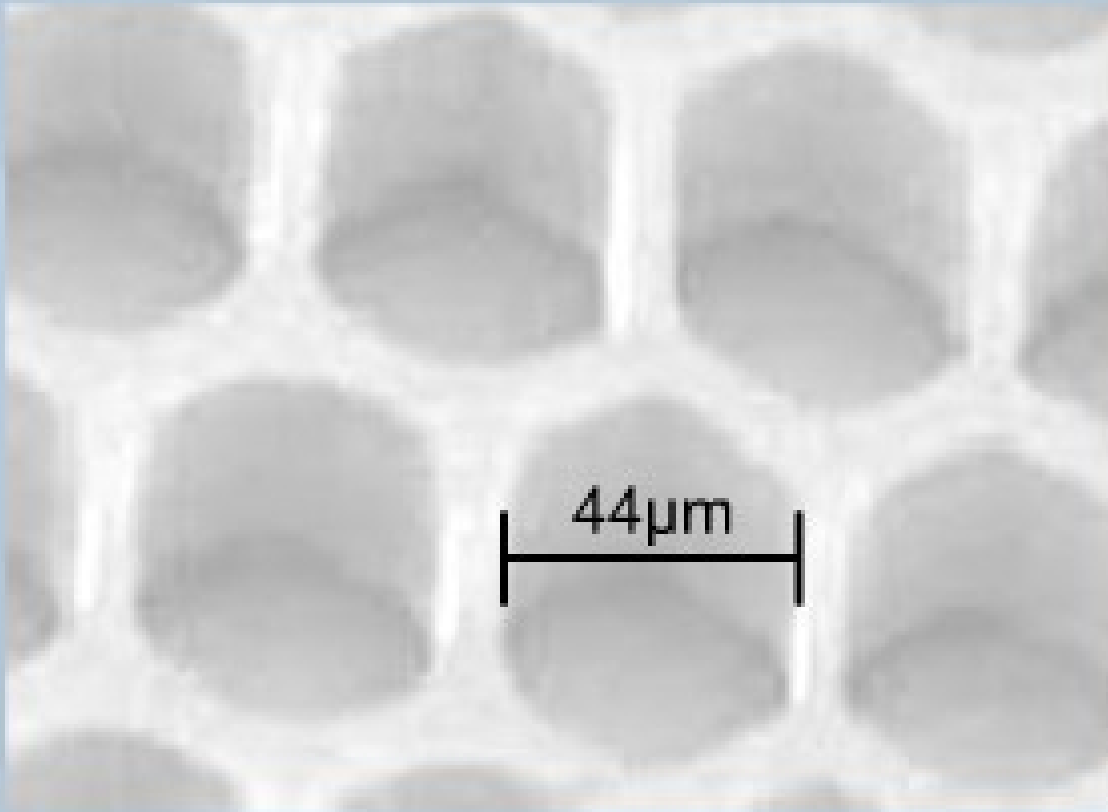
# Parametry mikroreaktorů



1.6 million  
wells

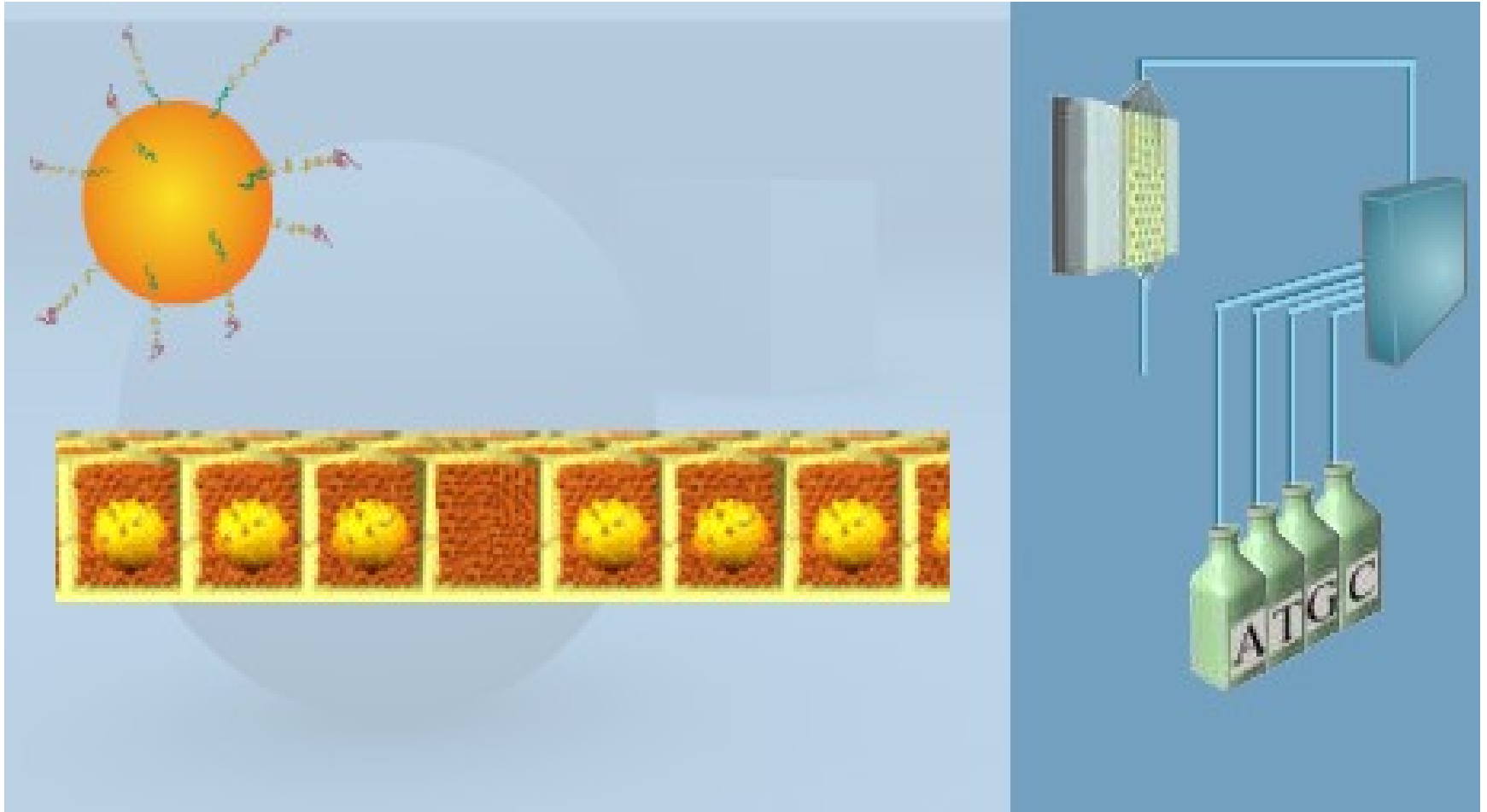
PicoTiterPlate device

# Parametry mikroreaktorů

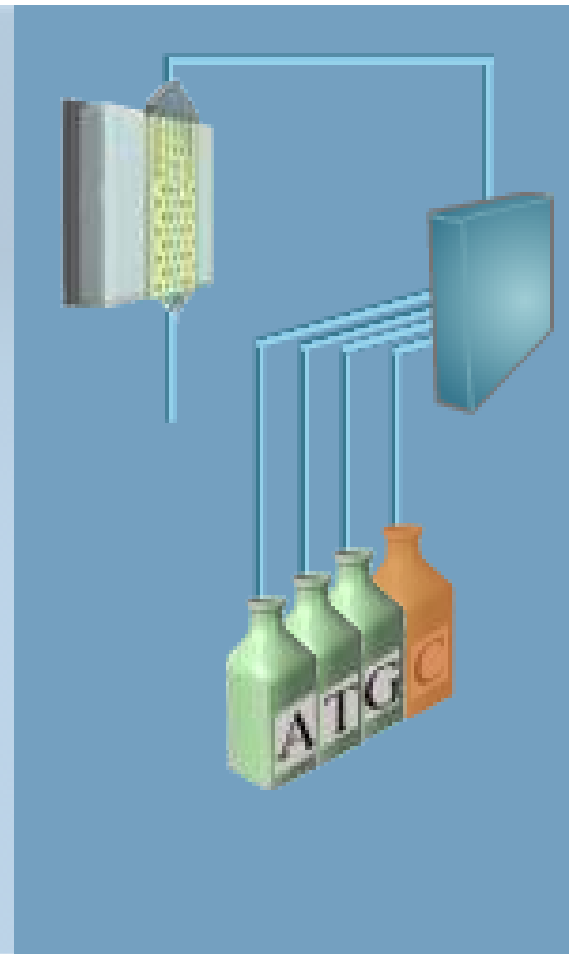
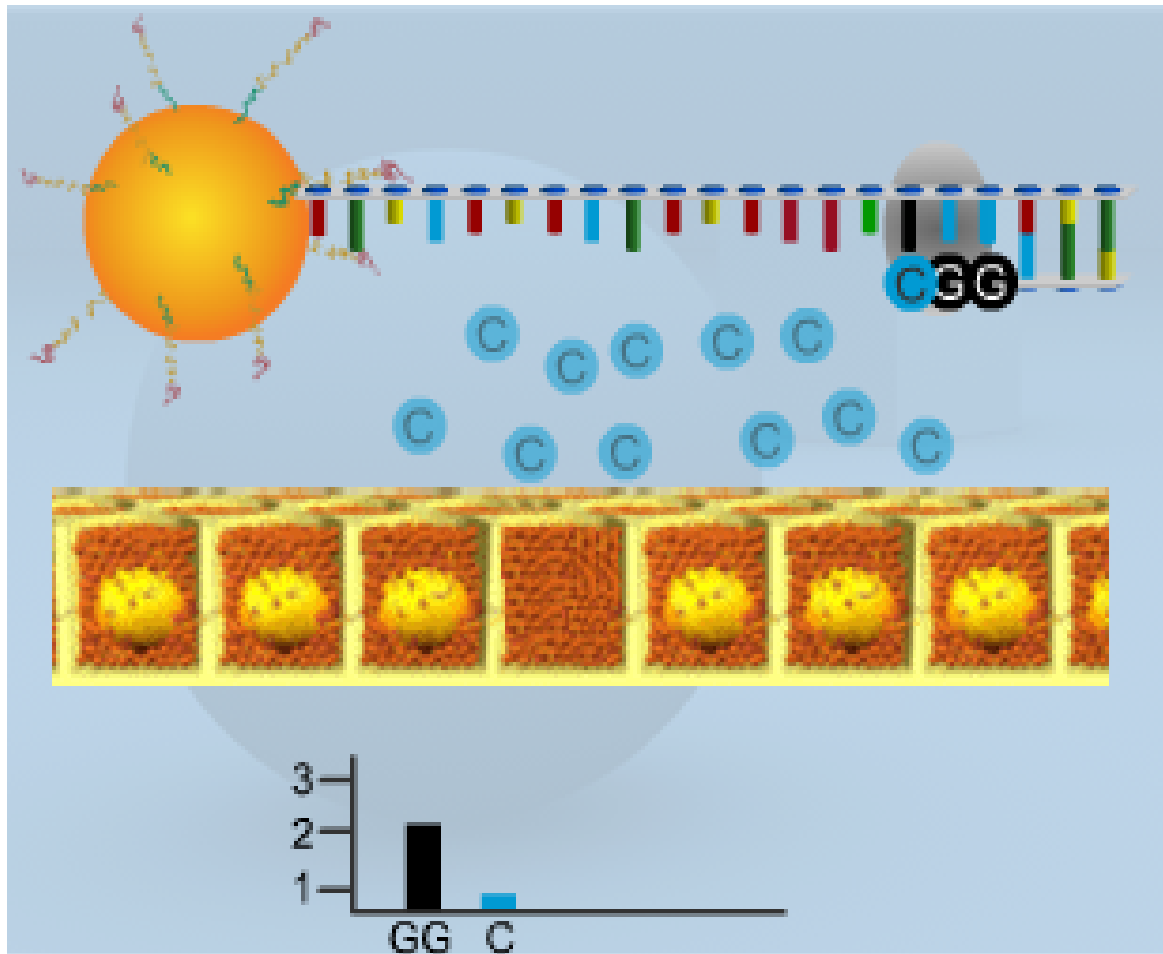


PicoTiterPlate device

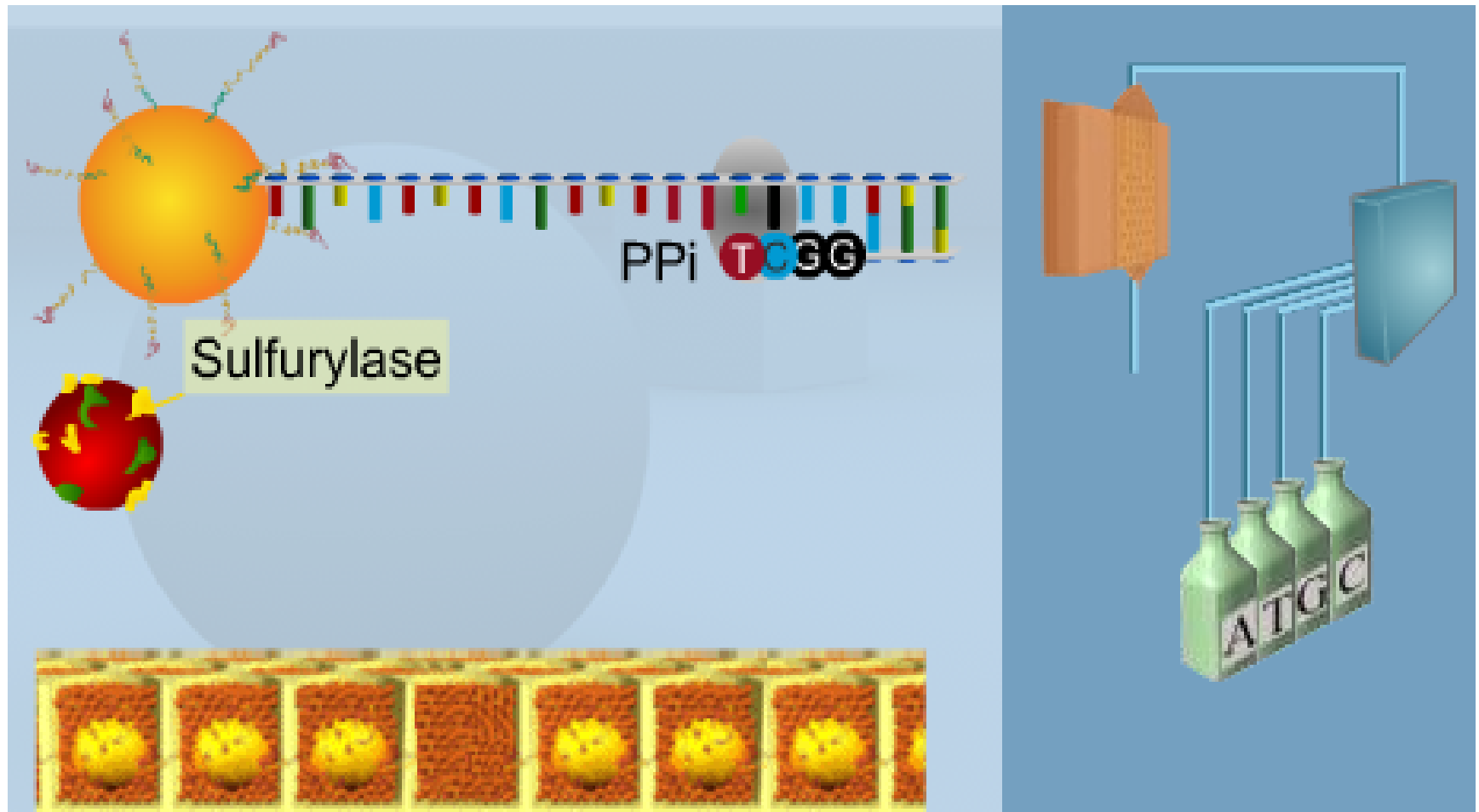
# sekvenace



# sekvenace

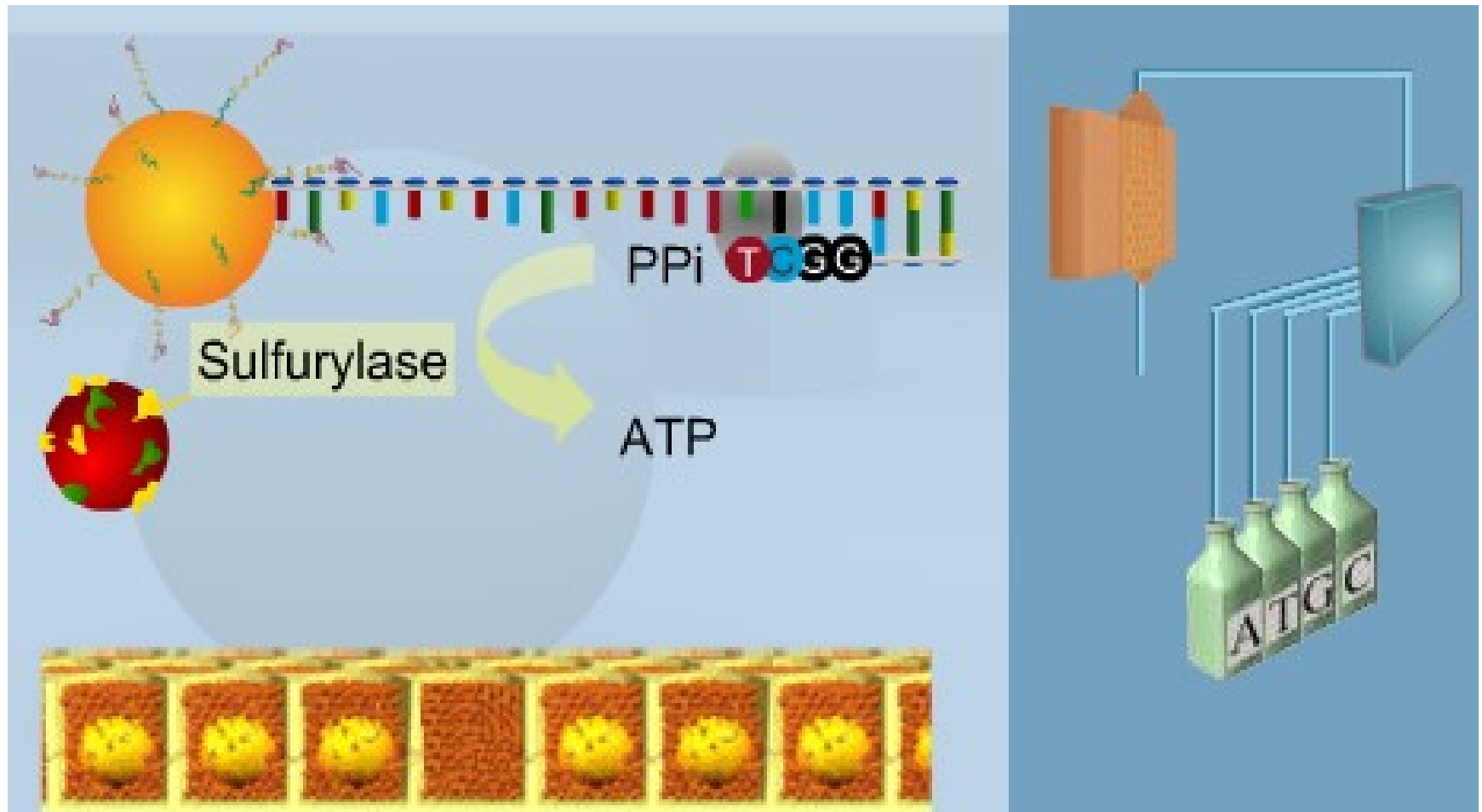


# sekvenace

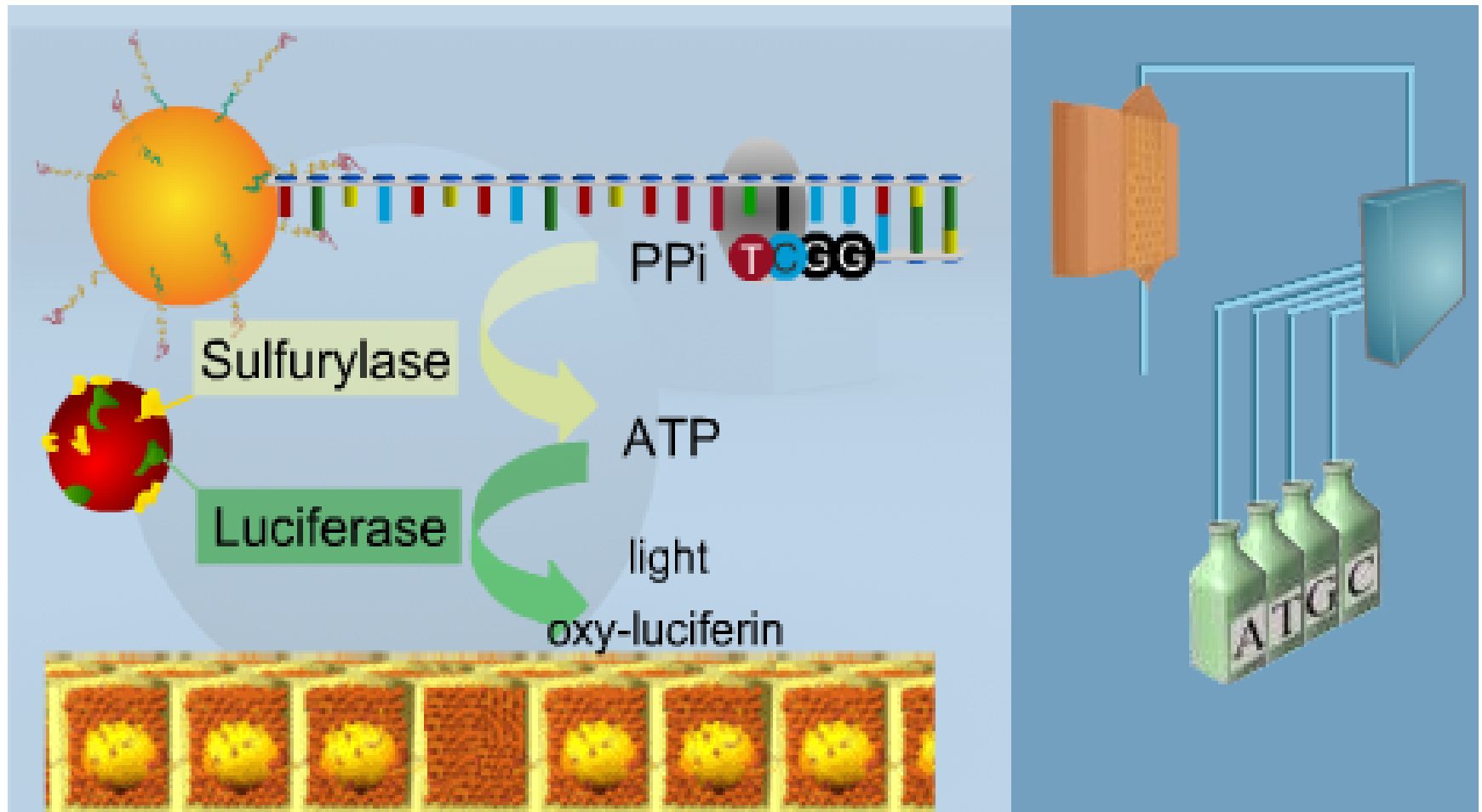




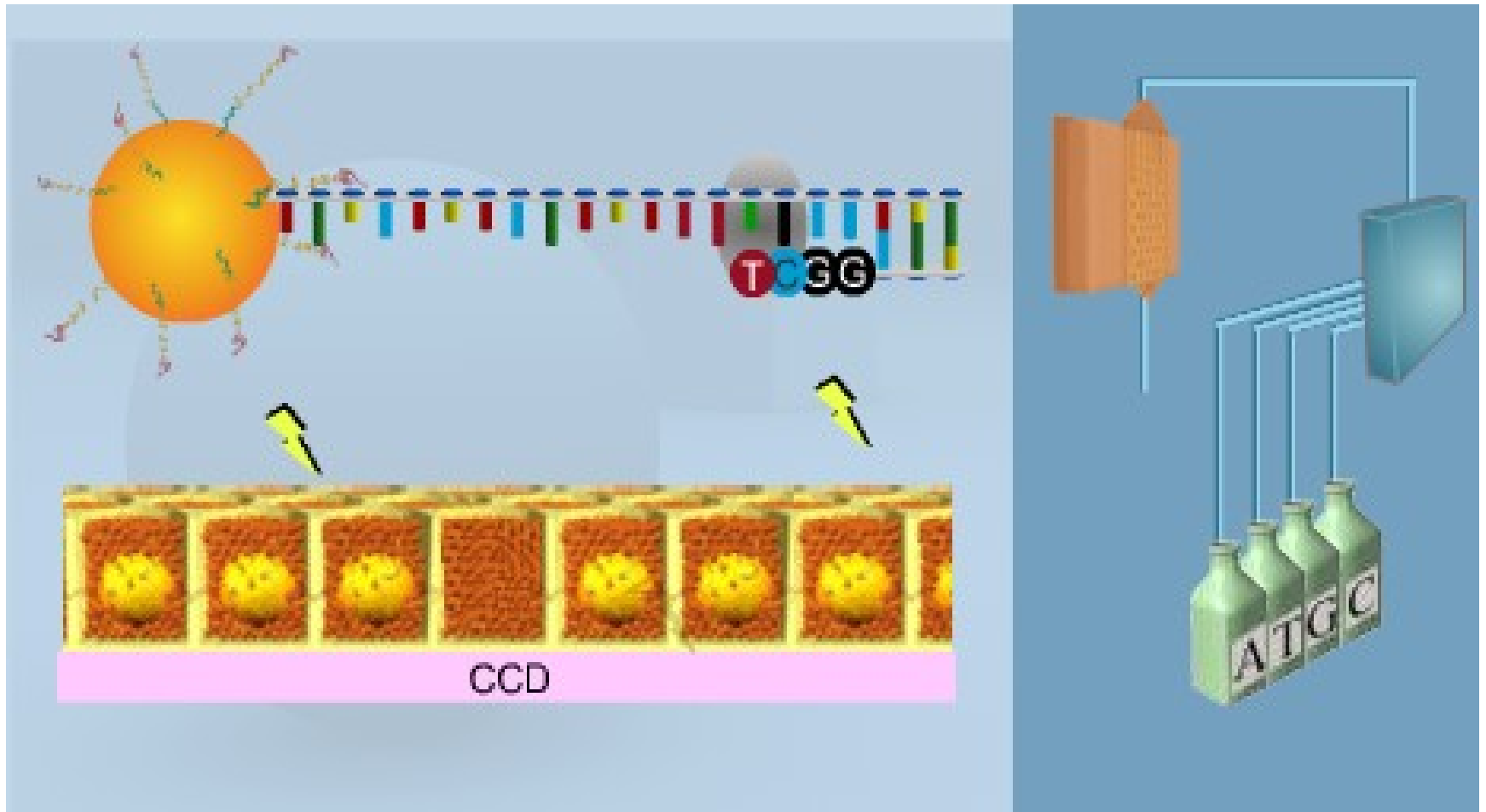
# sekvenace



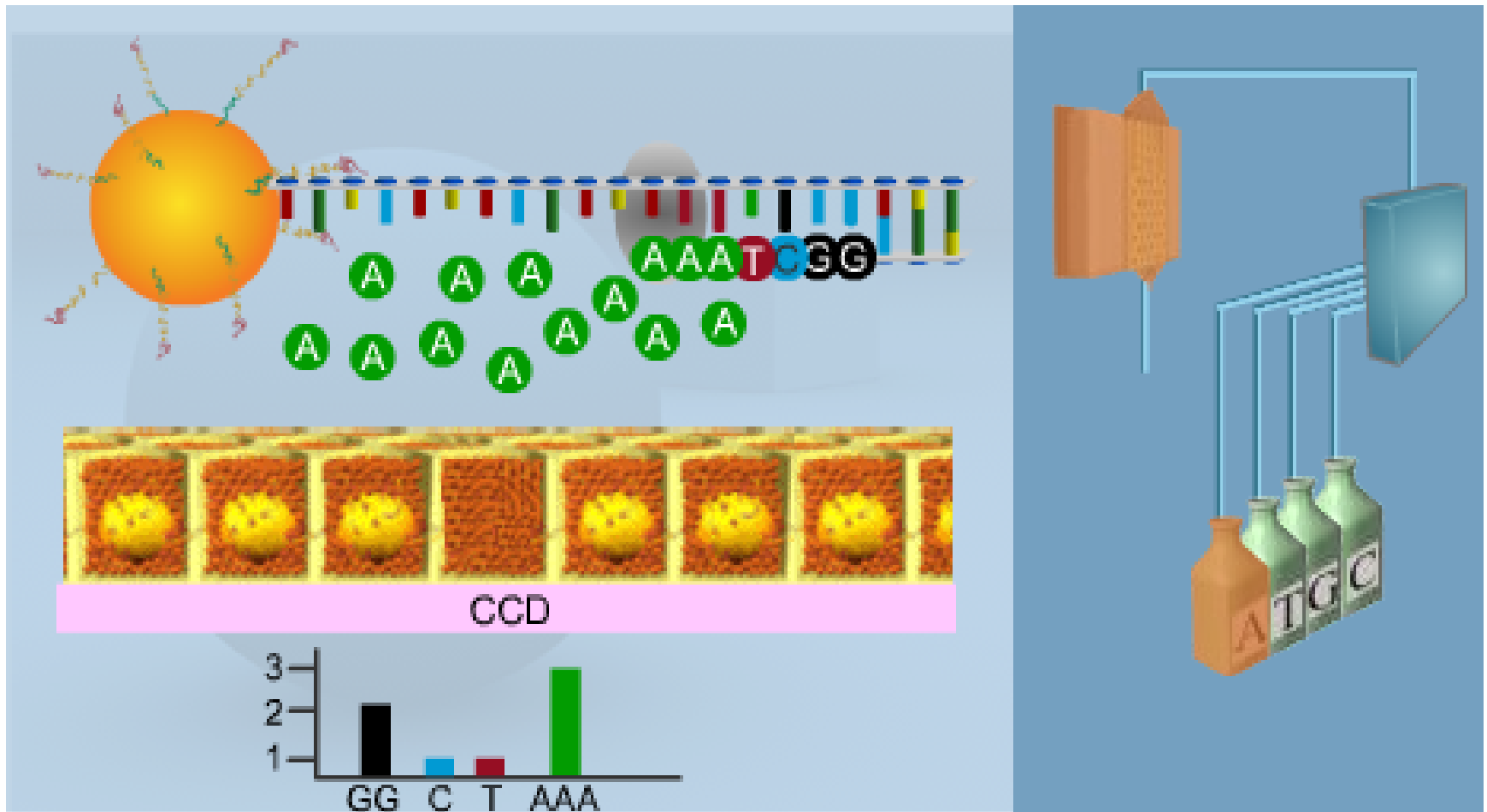
# sekvenace



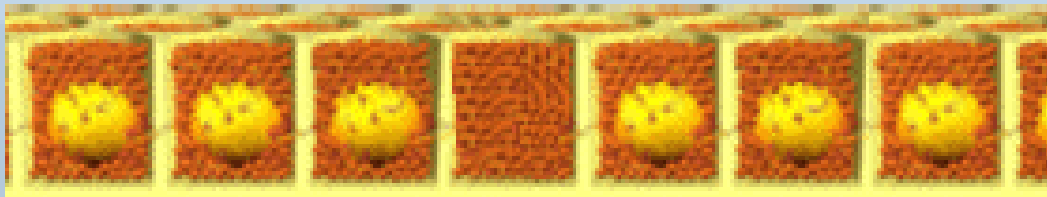
# sekvenace



# sekvenace



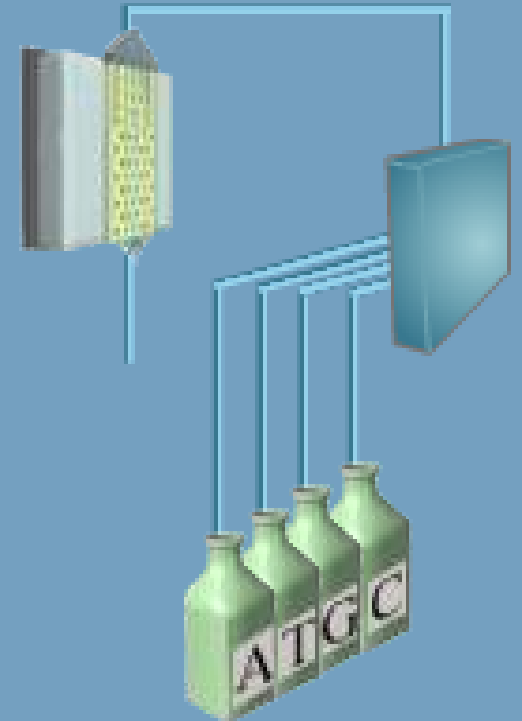
# sekvenace



Massive parallelization of sequencing reactions

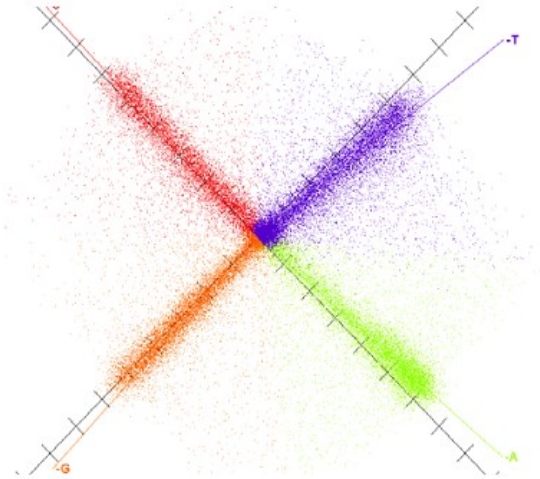
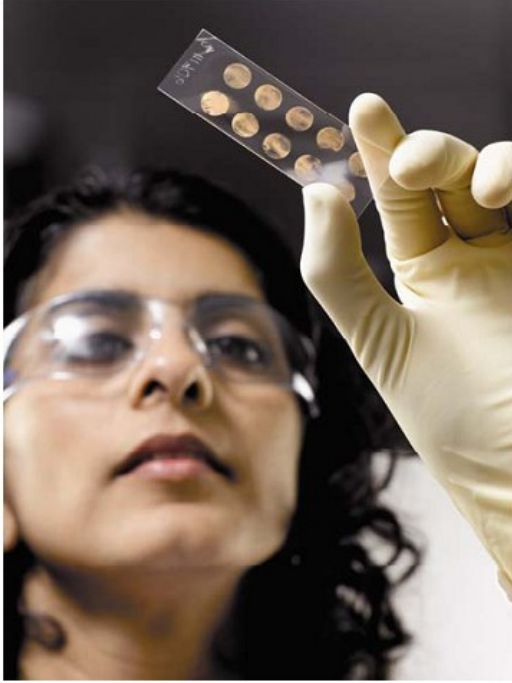
100 bases read length  
X  
200 000

20 Million Bases



# SOLID (Sequencing by Oligonucleotide Ligation and Detection)

2-base encoding sequencing (2007)

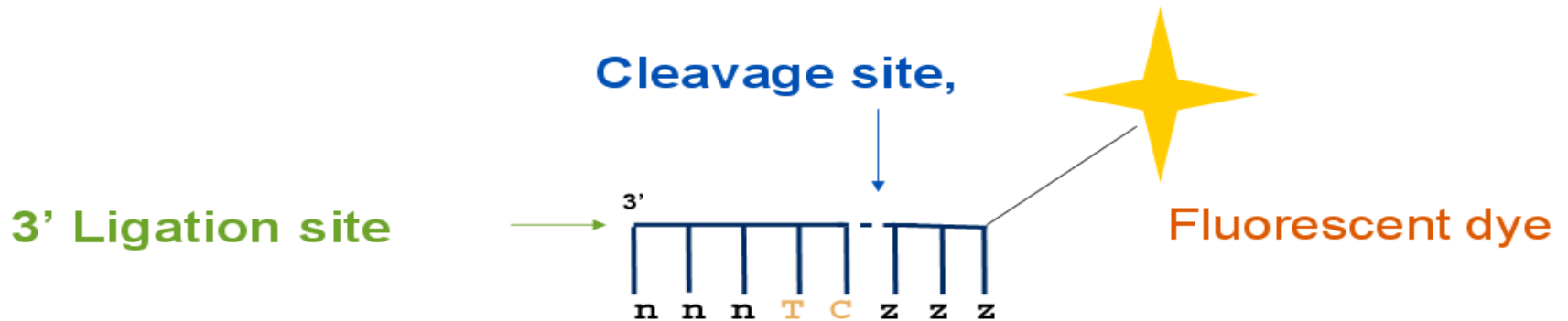


SOLiD™ System

Sequencing by Oligonucleotide Ligation and Detection

# Properties of the Probes

Spatial separation among dye, ligation & cleavage sites



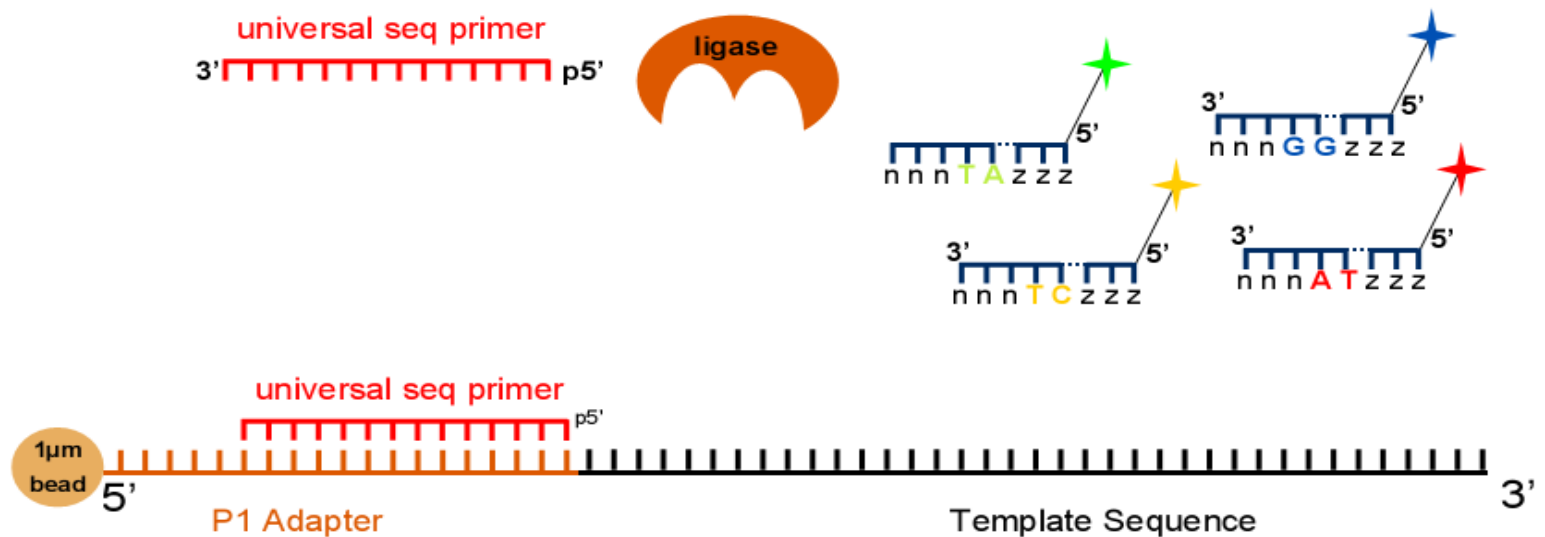
1,024 Octamer Probes ( $4^5$ )

4 Dyes, 4 dinucleotides, 256 probes per dye

N= degenerate bases Z= Universal bases

# SOLiD Chemistry System 4-color ligation

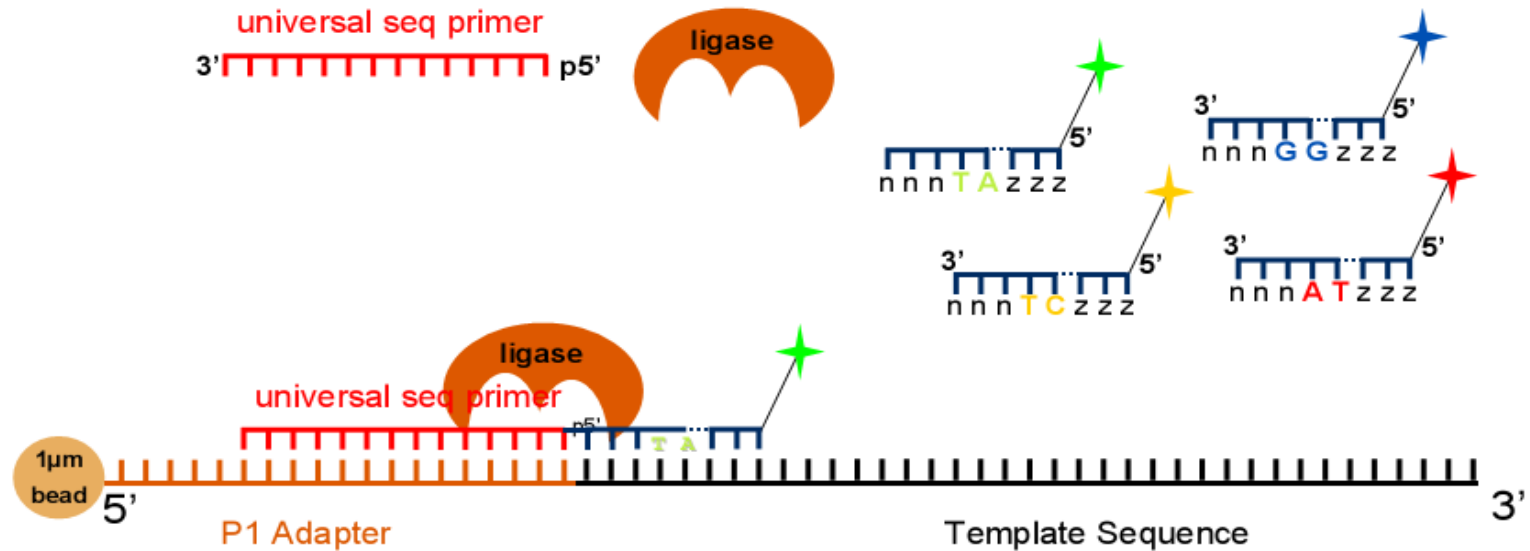
## Ligation reaction





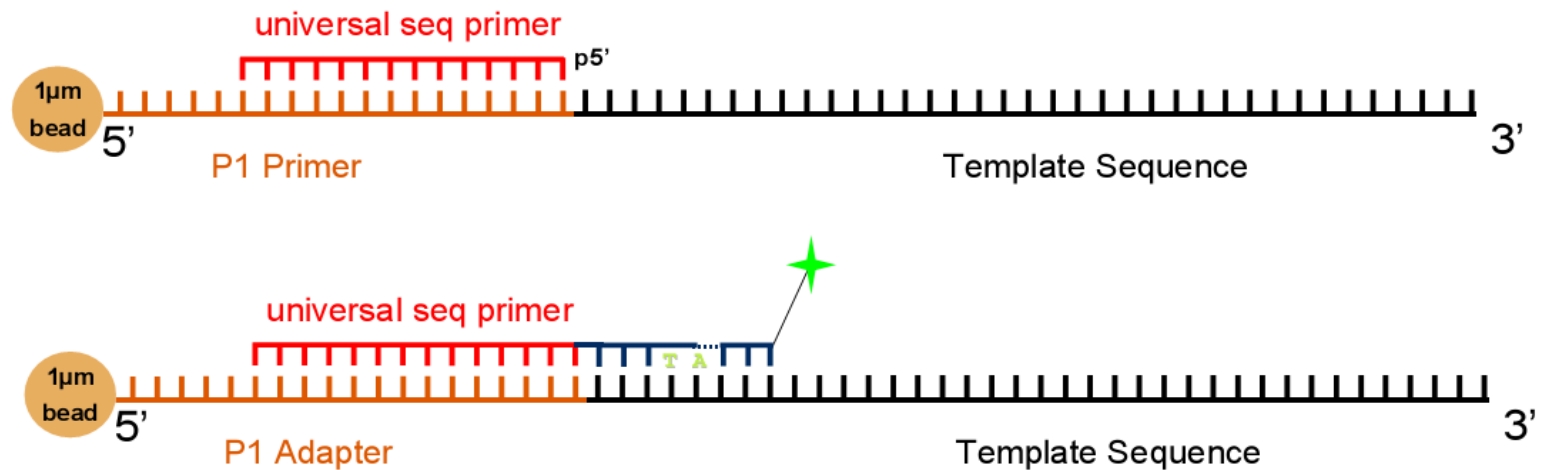
# SOLiD Chemistry System 4-color ligation

## Ligation reaction

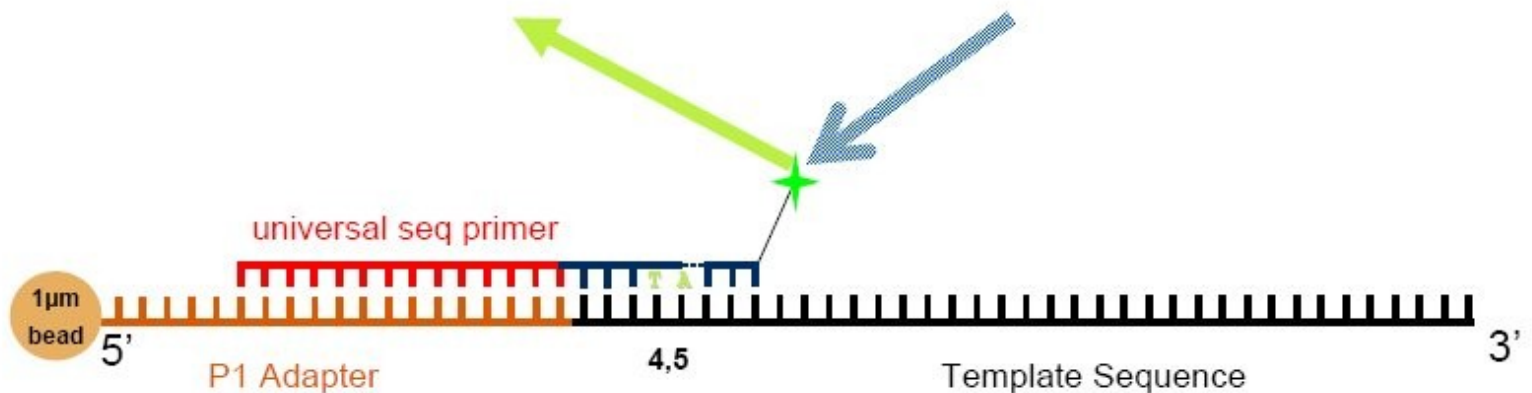


# SOLiD Chemistry System 4-color ligation

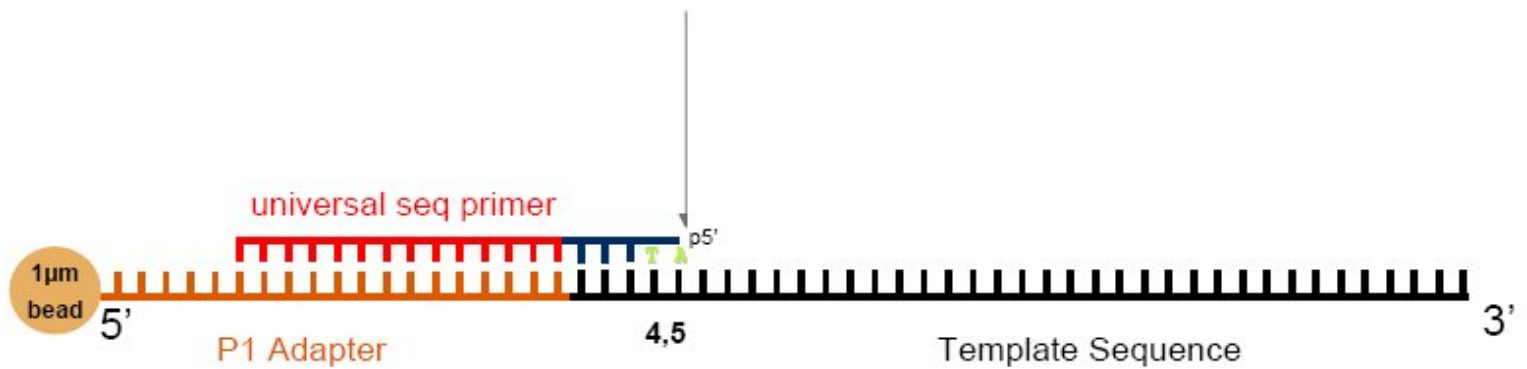
## De-Phosphorylation



# SOLiD Chemistry System 4-color ligation Visualization

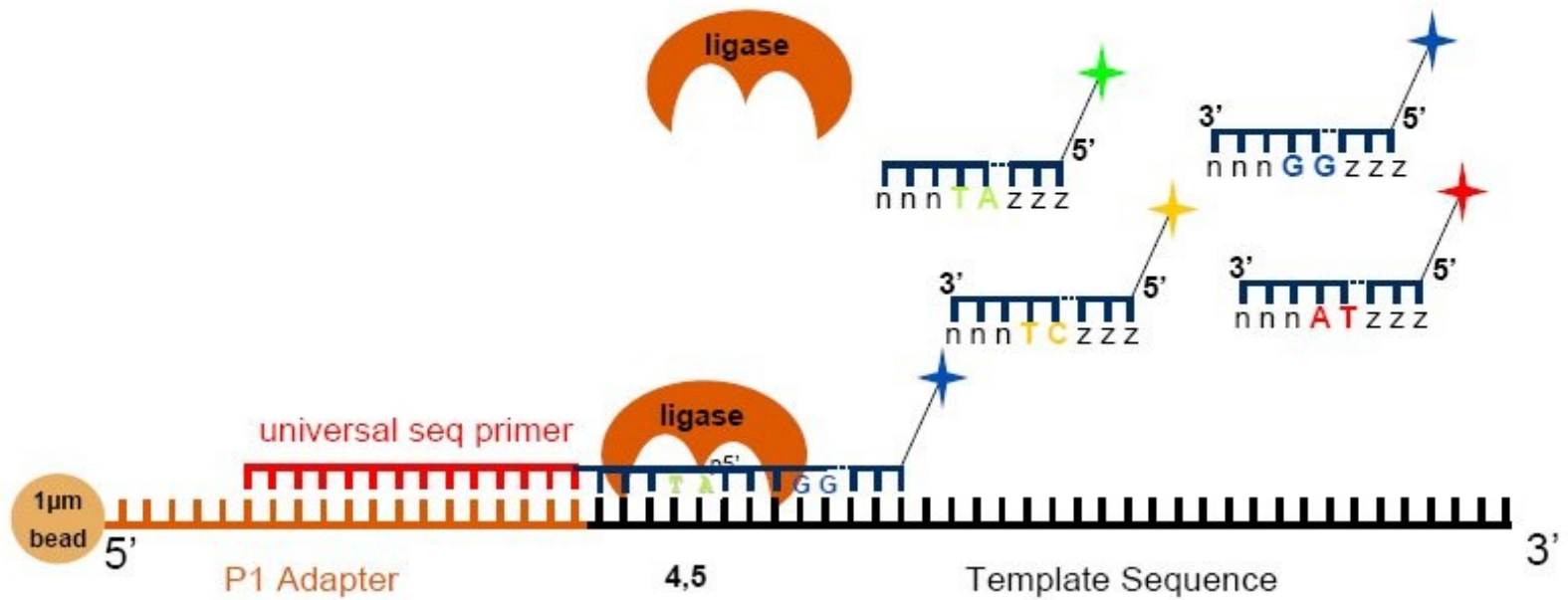


## SOLiD Chemistry System 4-color ligation Cleavage

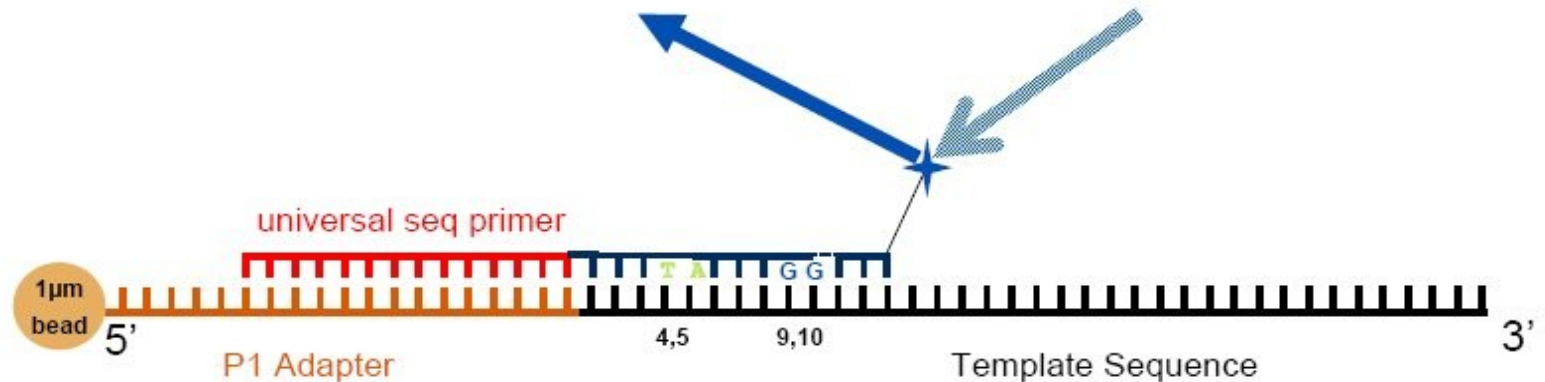


# SOLiD Chemistry System 4-color ligation

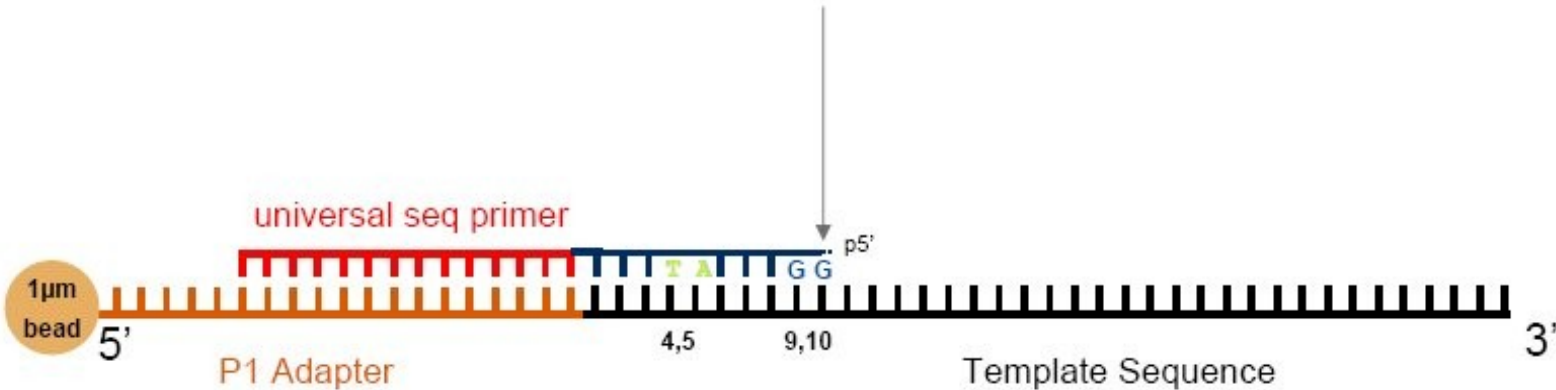
## Ligation (2<sup>nd</sup> cycle)



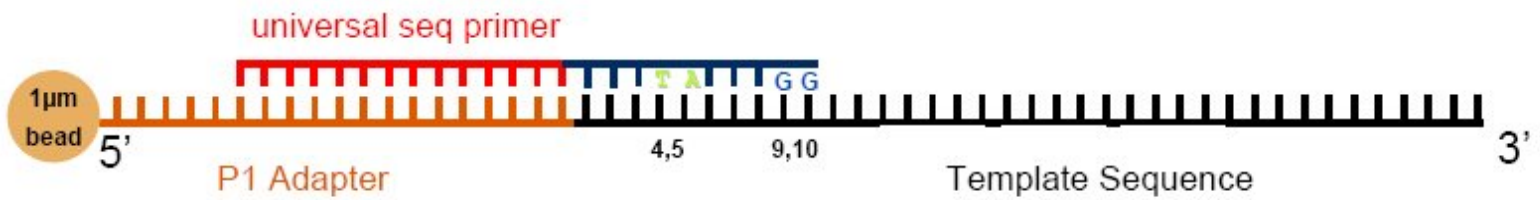
# SOLiD Chemistry System 4-color ligation Visualization (2<sup>nd</sup> cycle)



# SOLiD Chemistry System 4-color ligation Cleavage (2<sup>nd</sup> cycle)

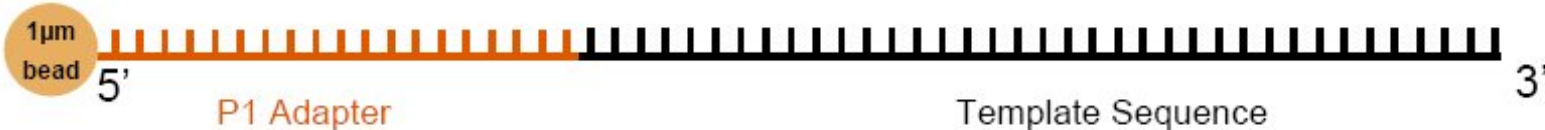


## SOLiD Chemistry System 4-color ligation interrogates every 5<sup>th</sup> base

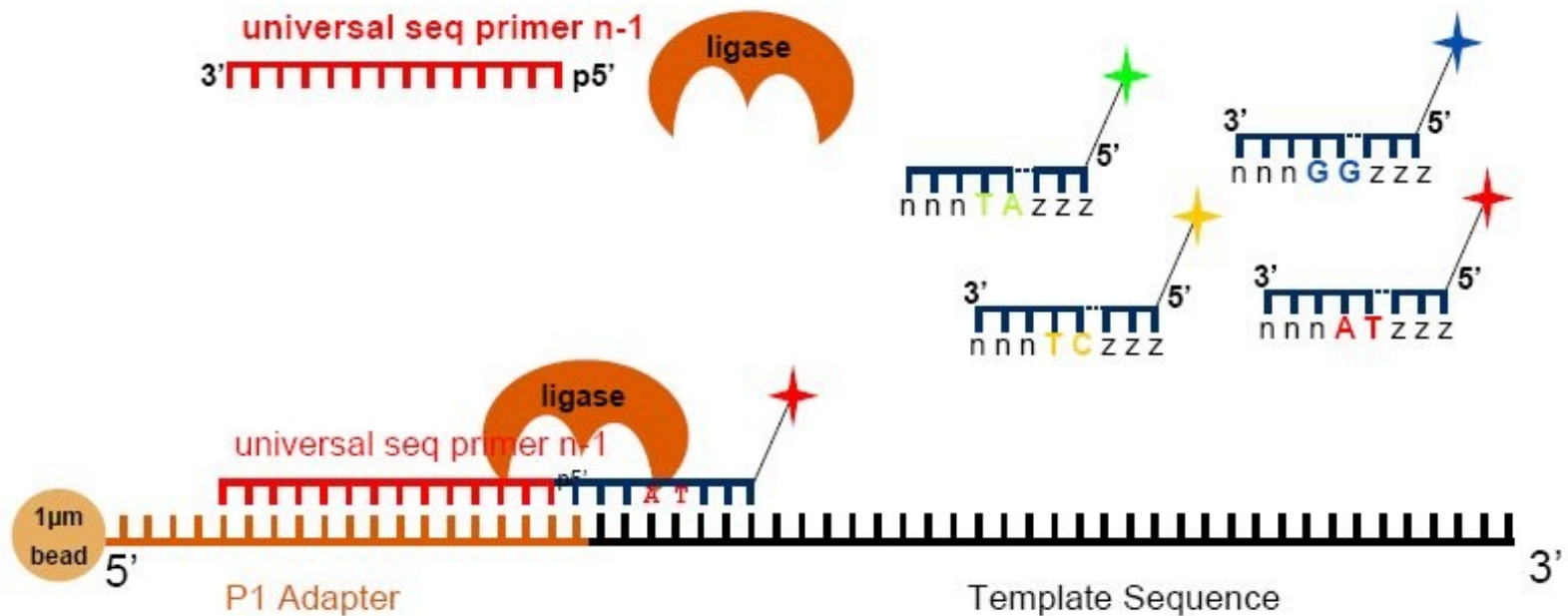




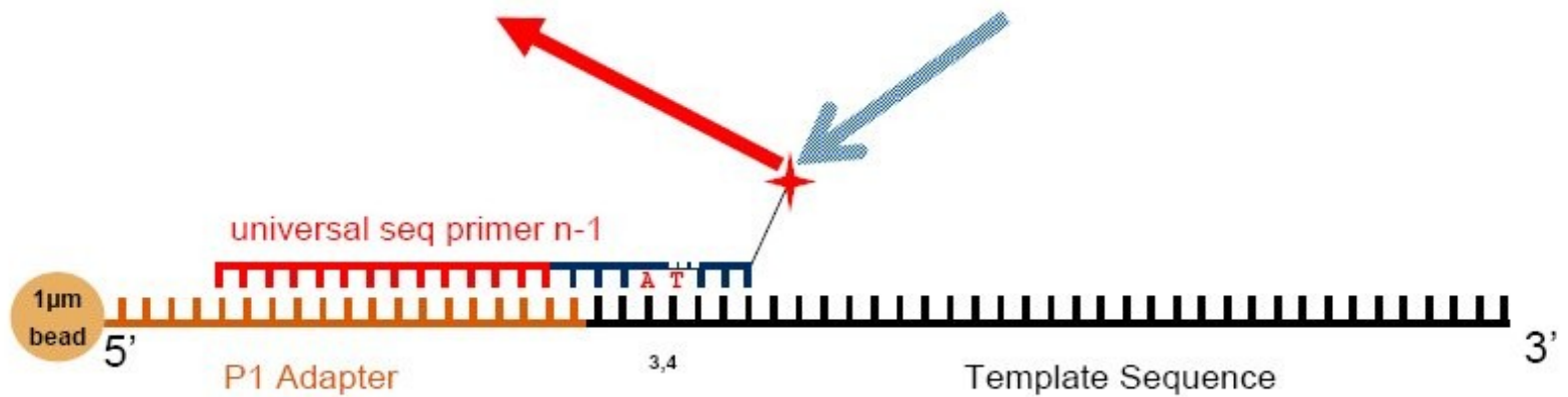
# SOLiD Chemistry System 4-color ligation Reset



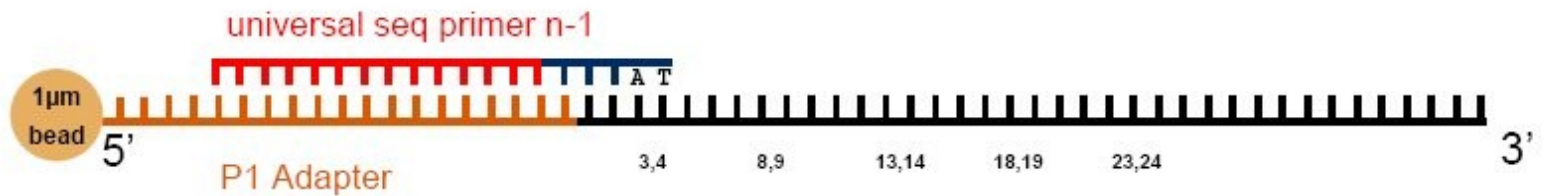
# SOLiD Chemistry System 4-color ligation (1<sup>st</sup> cycle after reset)



## SOLiD Chemistry System 4-color ligation (1<sup>st</sup> cycle after reset)

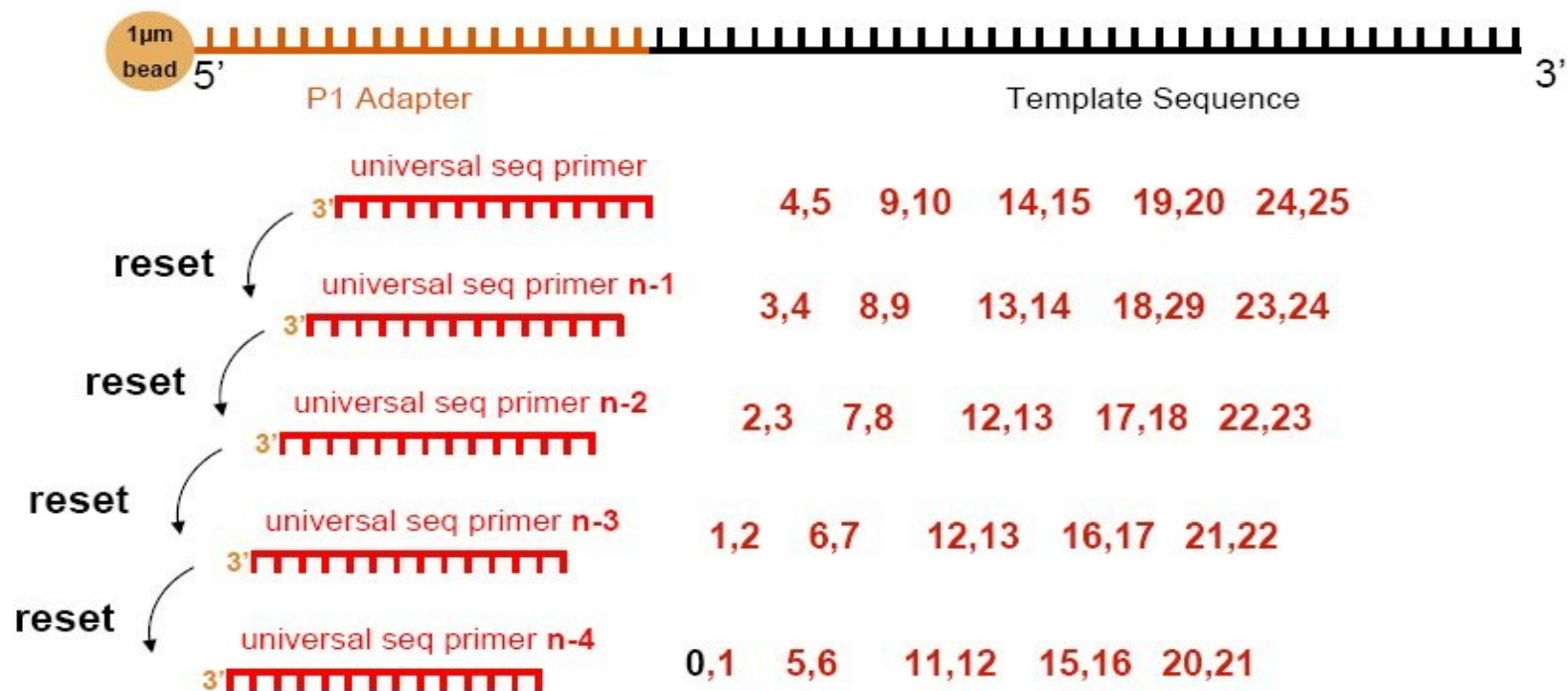


## SOLiD Chemistry System 4-color ligation (2<sup>nd</sup> Round)



# Sequential rounds of sequencing

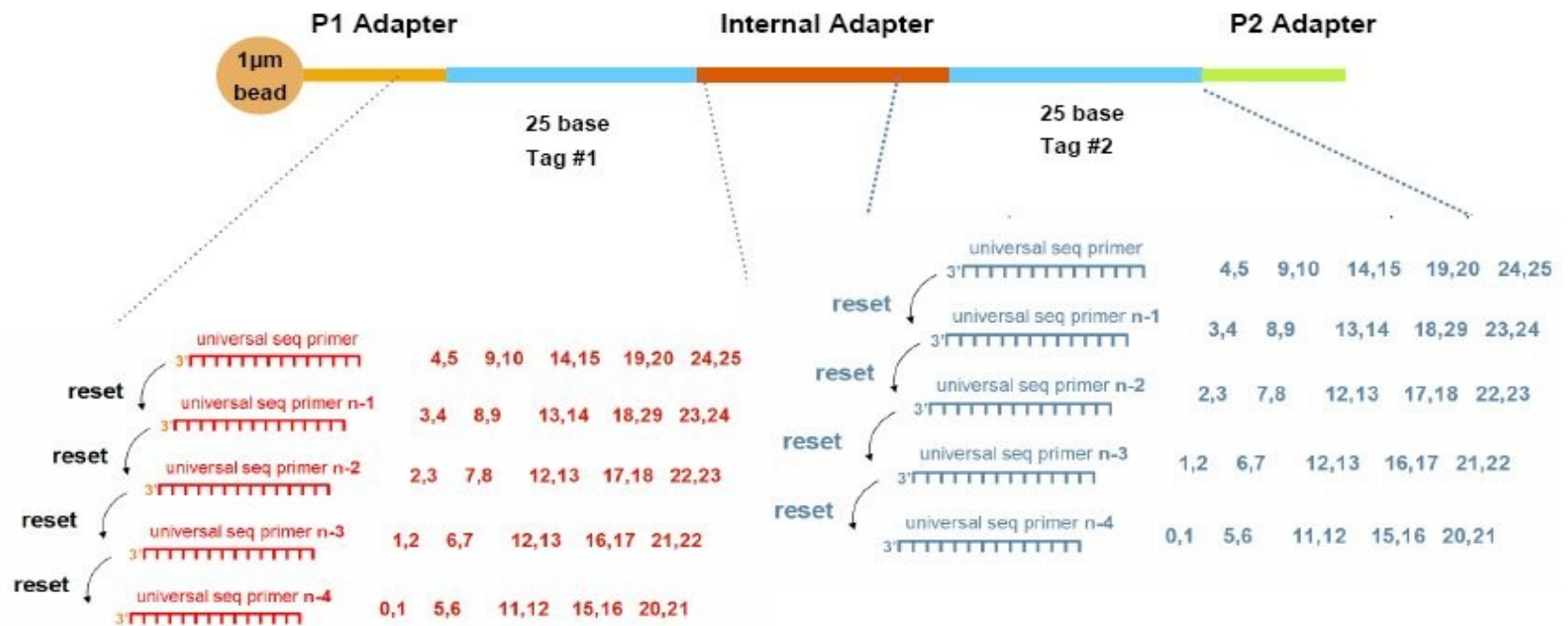
## Multiple cycles per round



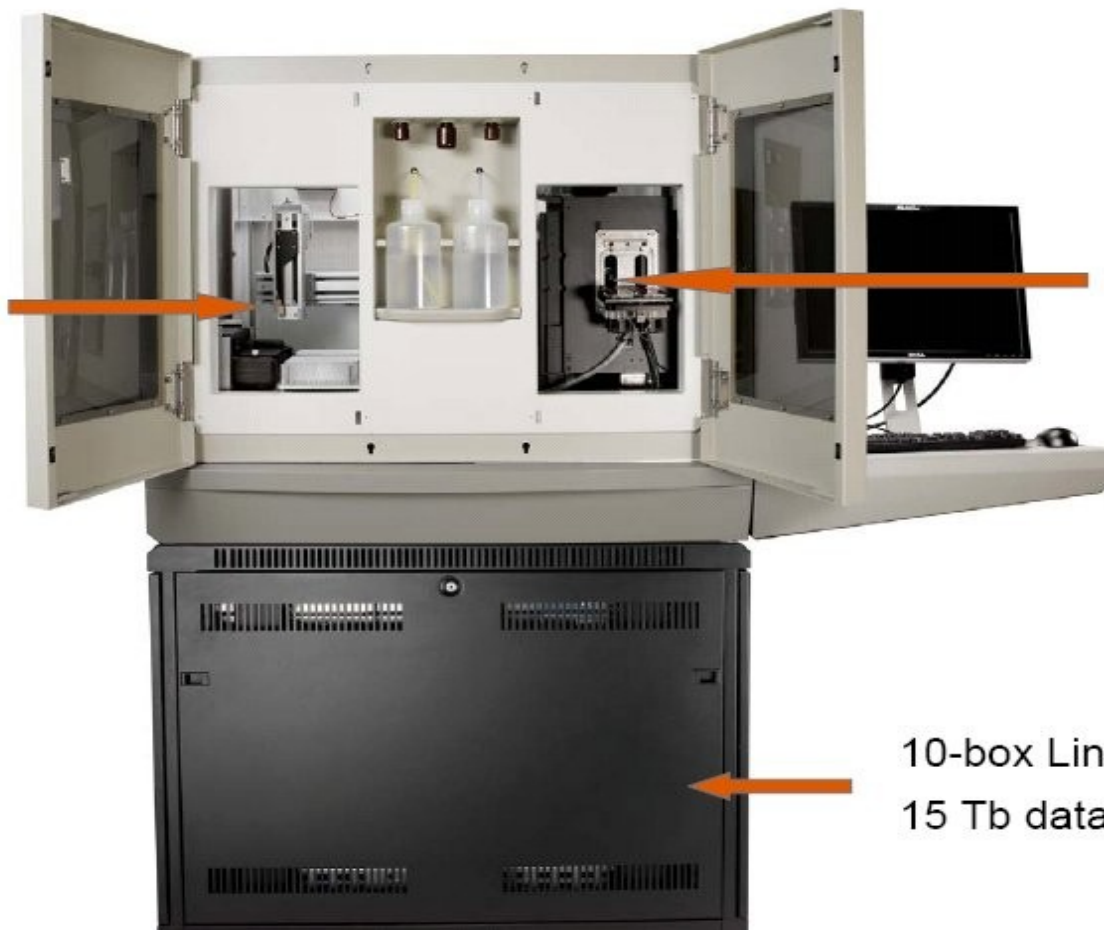
# Paired End two sequences generated

## Sequential rounds of sequencing

### Multiple cycles per round



Reagent  
handling



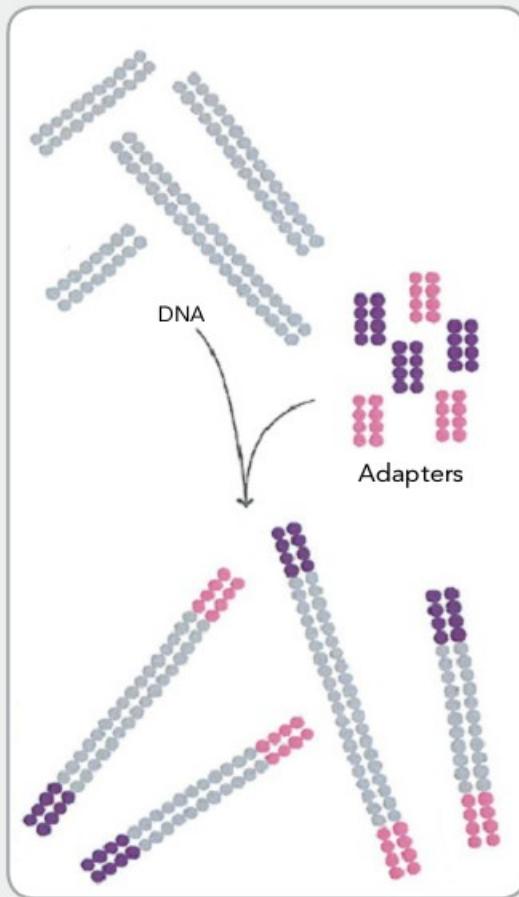
Dual Flow  
Cell

10-box Linux Cluster  
15 Tb data storage



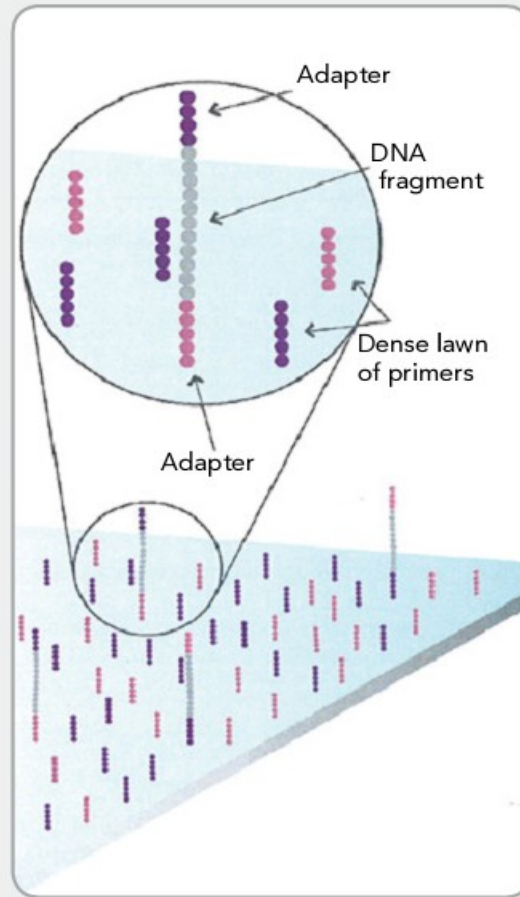
# Solexa (2007)

## 1. PREPARE GENOMIC DNA SAMPLE



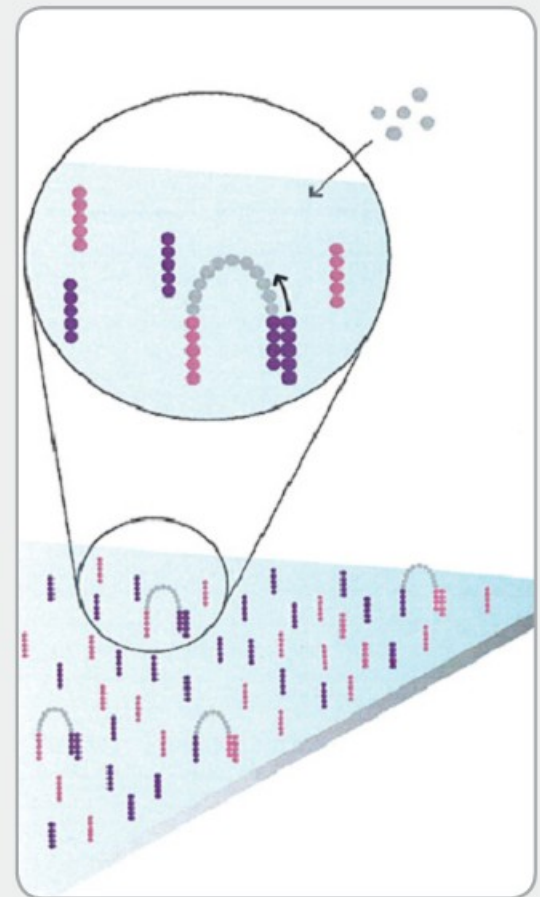
Randomly fragment genomic DNA and ligate adapters to both ends of the fragments.

## 2. ATTACH DNA TO SURFACE



Bind single-stranded fragments randomly to the inside surface of the flow cell channels.

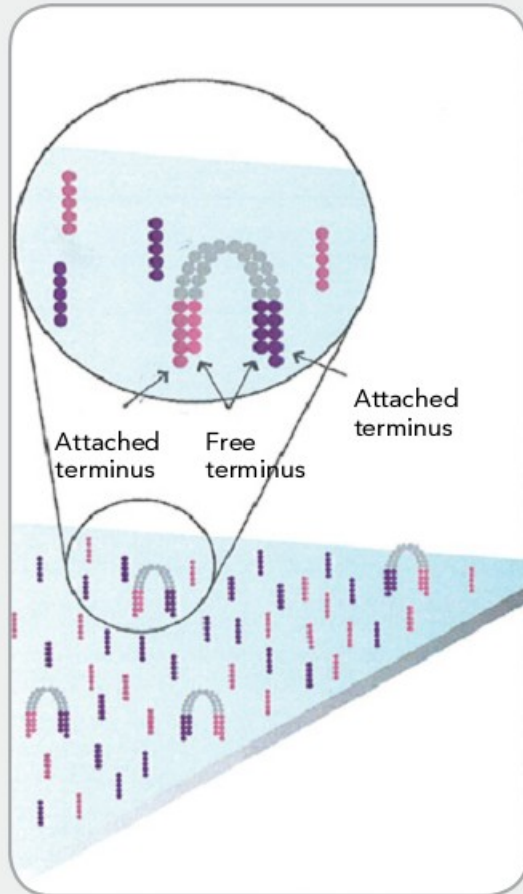
## 3. BRIDGE AMPLIFICATION



Add unlabeled nucleotides and enzyme to initiate solid-phase bridge amplification.

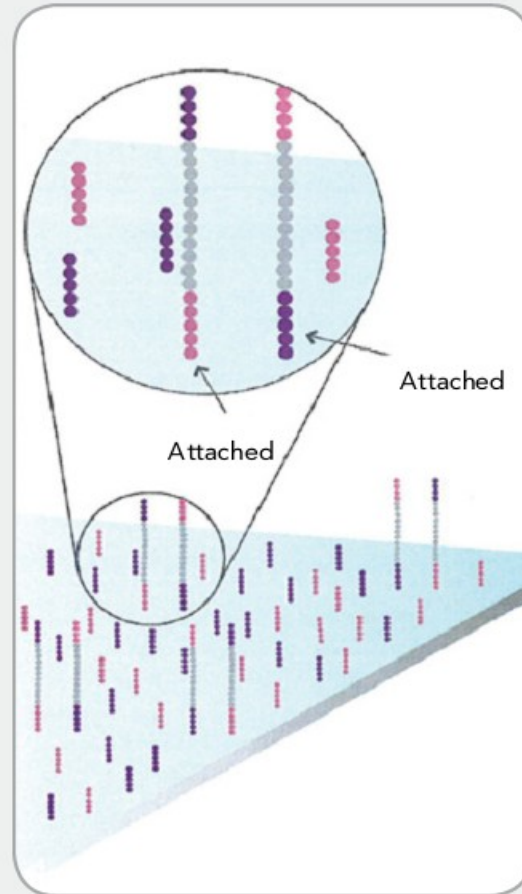


4. FRAGMENTS BECOME DOUBLE STRANDED



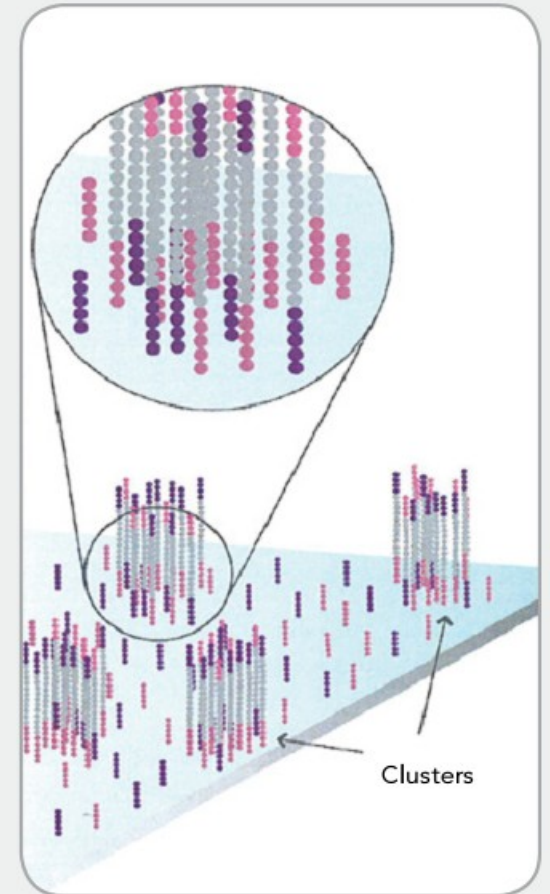
The enzyme incorporates nucleotides to build double-stranded bridges on the solid-phase substrate.

5. DENATURE THE DOUBLE-STRANDED MOLECULES



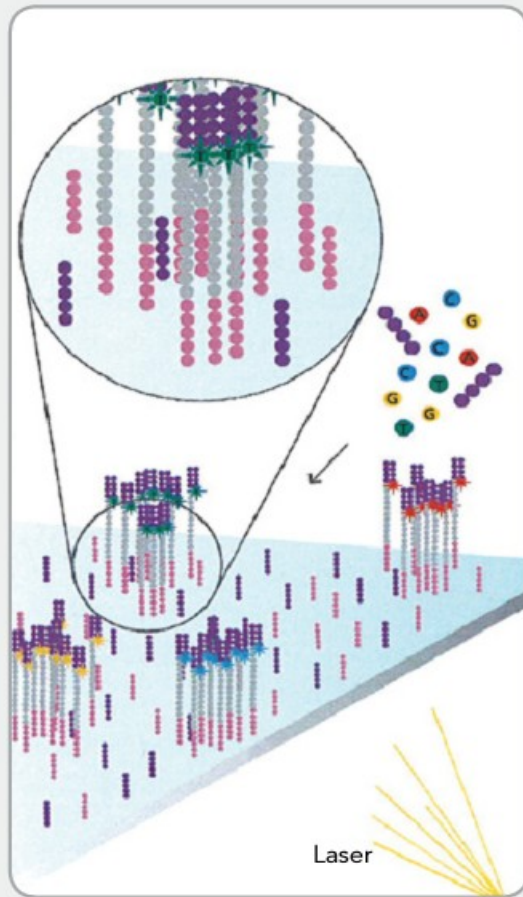
Denaturation leaves single-stranded templates anchored to the substrate.

6. COMPLETE AMPLIFICATION



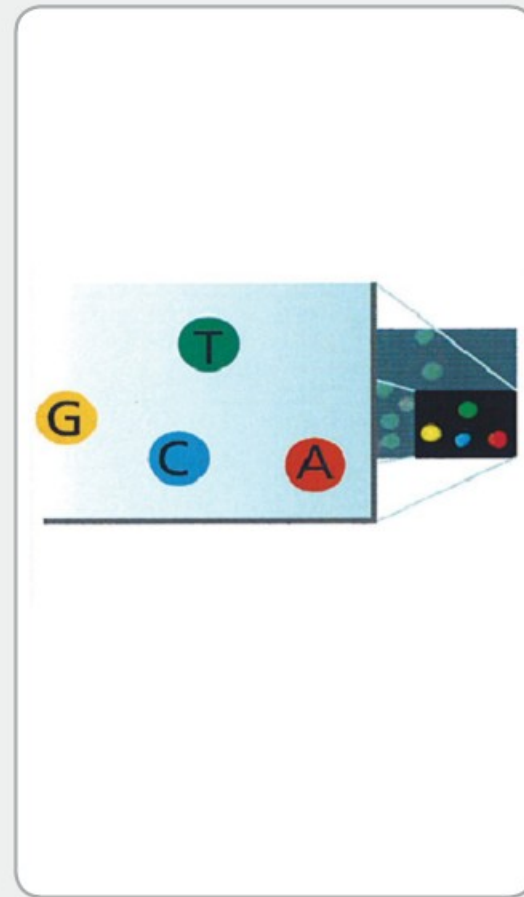
Several million dense clusters of double-stranded DNA are generated in each channel of the flow cell.

### 7. DETERMINE FIRST BASE



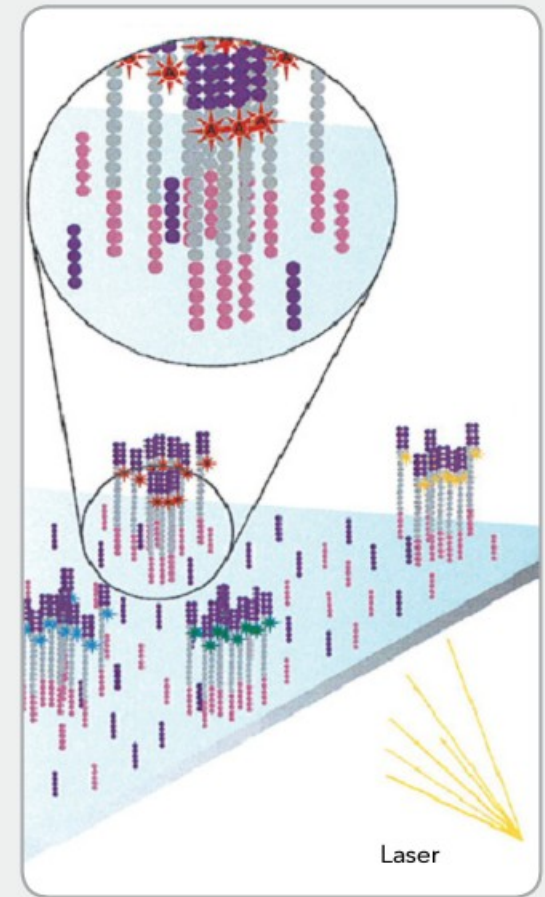
First chemistry cycle: to initiate the first sequencing cycle, add all four labeled reversible terminators, primers and DNA polymerase enzyme to the flow cell.

### 8. IMAGE FIRST BASE



After laser excitation, capture the image of emitted fluorescence from each cluster on the flow cell. Record the identity of the first base for each cluster.

### 9. DETERMINE SECOND BASE

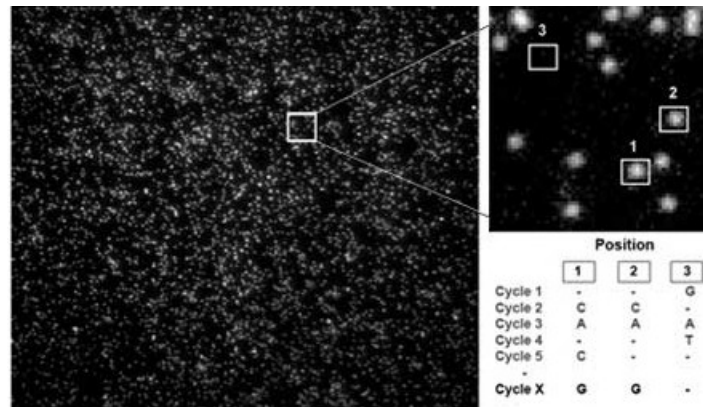
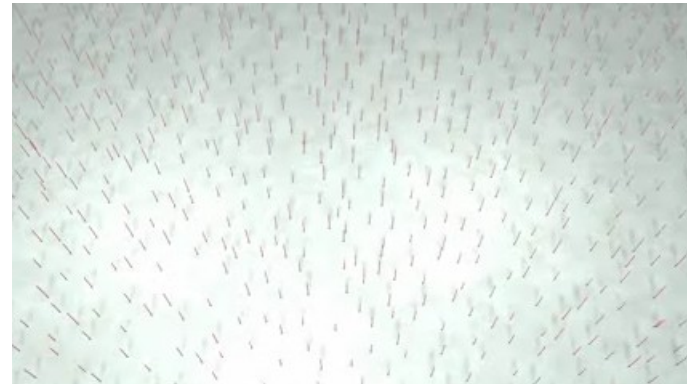
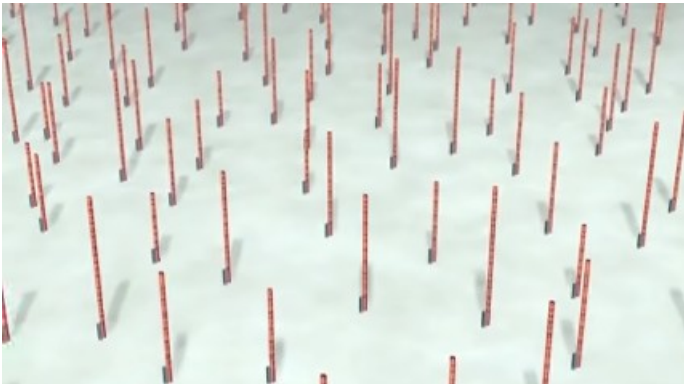


Second chemistry cycle: to initiate the next sequencing cycle, add all four labeled reversible terminators and enzyme to the flow cell.

# HELICOS (2008)



## True Single Molecule Sequencing (tSMS)

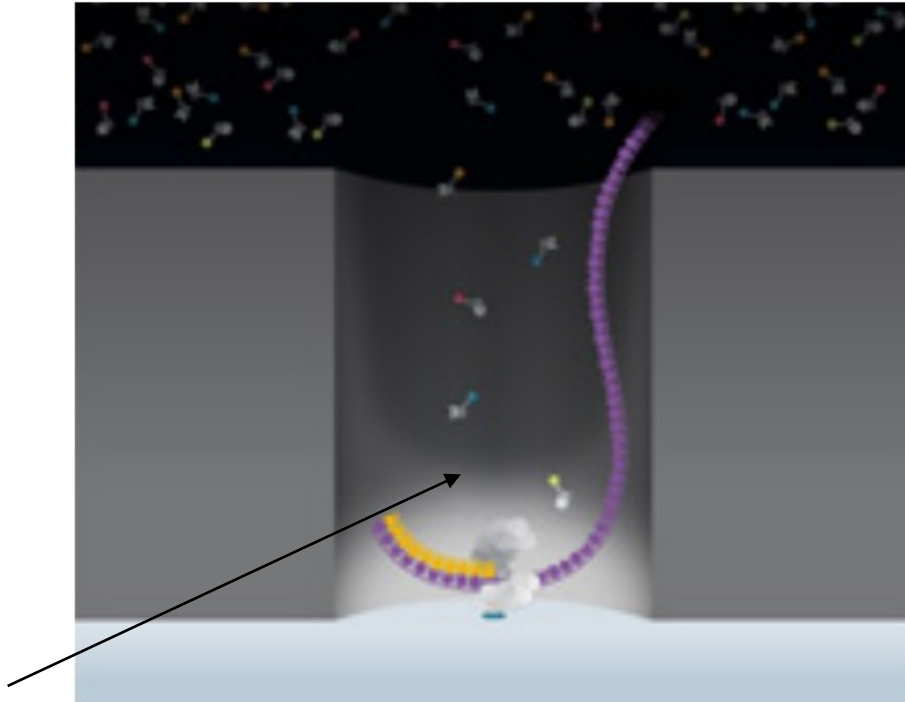


# Single Molecule Real-Time (SMRT)

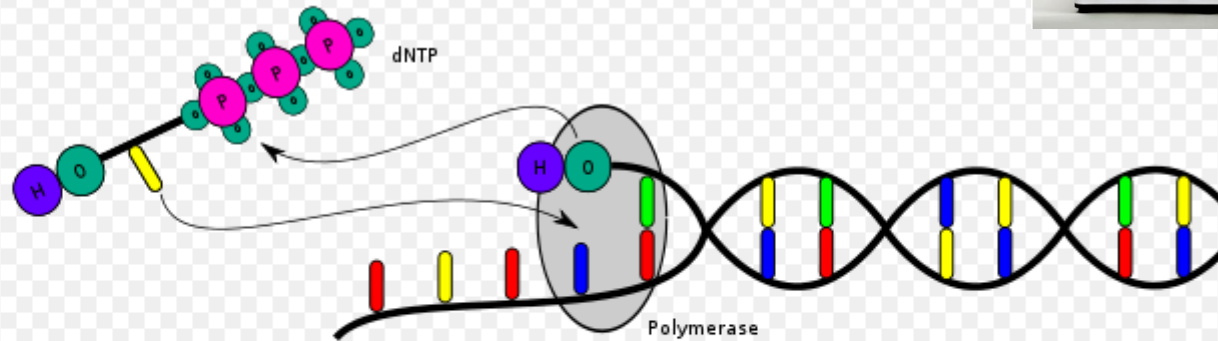
Pacific Biosciences



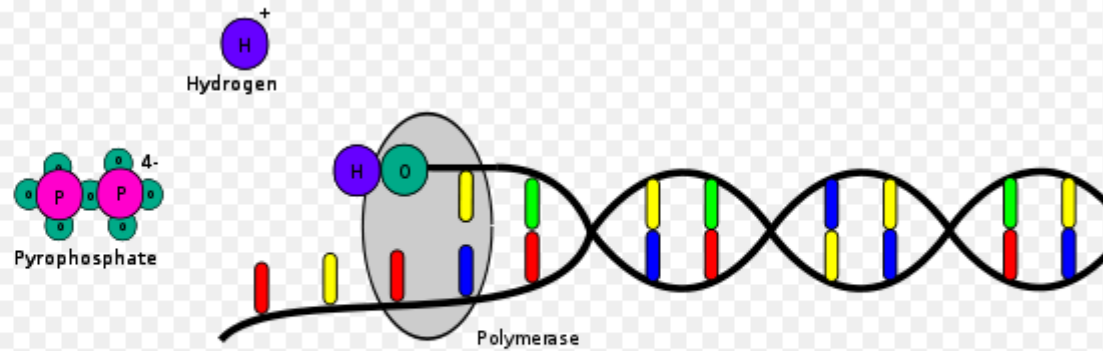
20 zeptolitru



# Ion Torrent



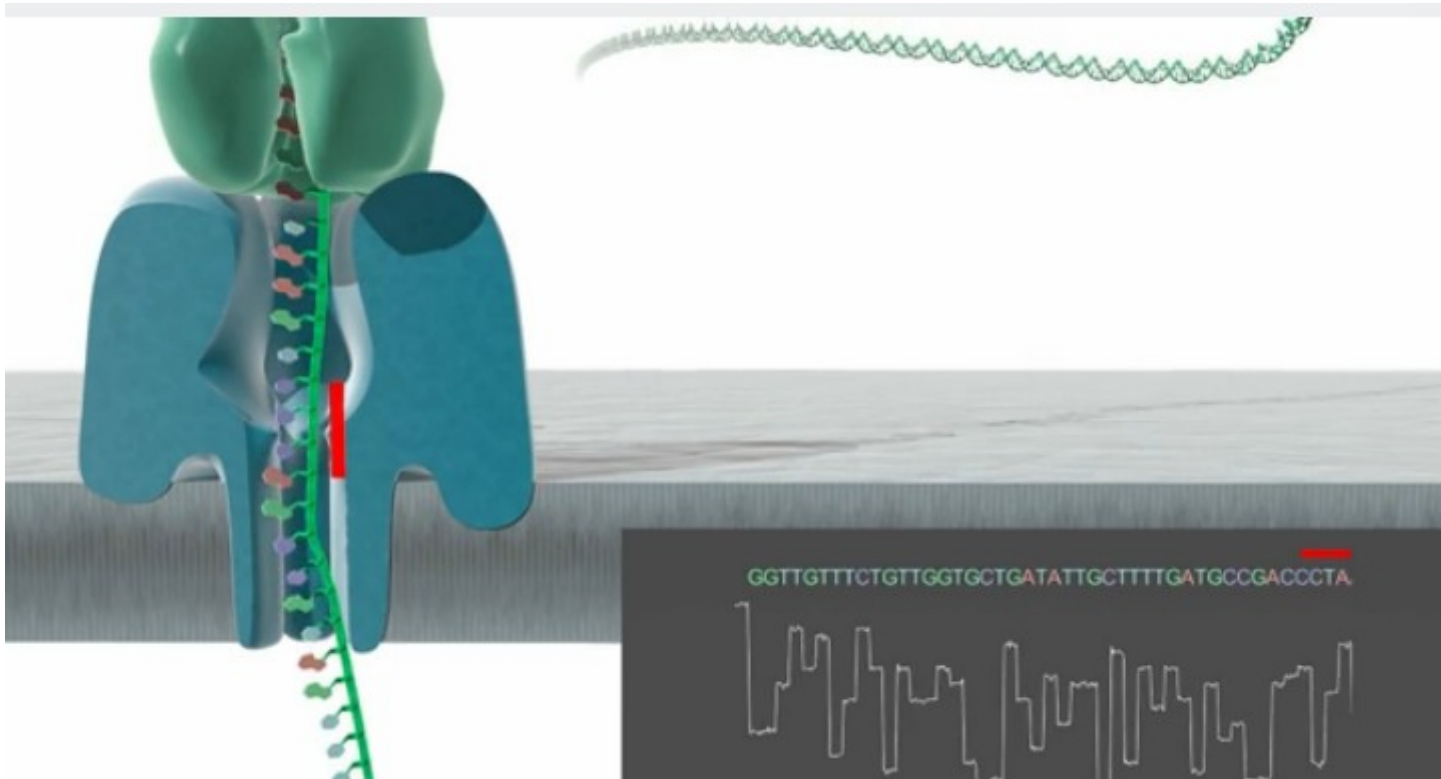
Polymerase integrates a nucleotide.



Hydrogen and pyrophosphate are released.



# Oxford nanopore



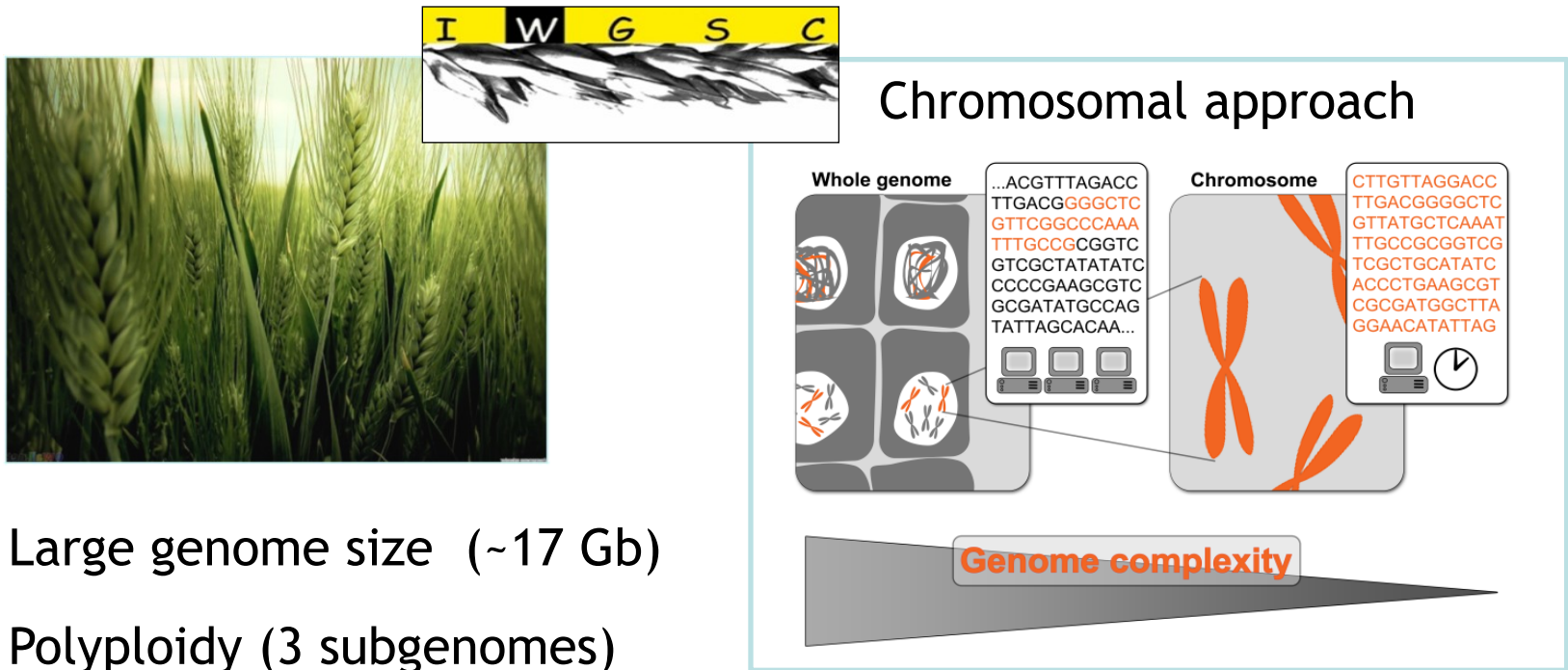
# Další technologie

- Mikroelektroforéza
- Sekvenování na bázi microarray

# CHALLENGES IN GENOME SEQUENCING

*De novo* genome assemblies using only short read data of NGS technologies are generally incomplete and highly fragmented due to

- Large duplications - chromosomal approach, BAC-by-BAC sequencing
- High proportion of repetitive DNA - **challenge!**



- Large genome size (~17 Gb)
- Polyploidy (3 subgenomes)

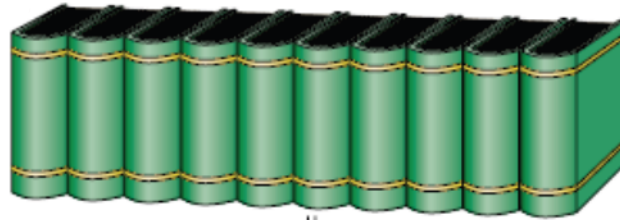




# BAC-BY-BAC SEQUENCING

Physical mapping

Sequencing



BAC clones

Construct clone map and select mapped clones

AGTTGGTAACTA	TGGCAATTGTAGA	CGATCGATGACTA
ATTGGACTTCGGA	TAACCTGCATGCT	CAGCTAGCGGTGAT
CGATCGATGACTG	TGATCGATGTACT	ATGCTGACTGTAG
CTTGATCGATGTA	GGATCTTACAAGT	ATAACCTGCCCTTG
ACTGGGATCCCTAC	GGATTAATAAACCA	CGAGCCTTCCGAC
TCCGGTATAGCCC	AACGTTAGATCGA	ATCGATGTACTGG
AATCGATATCGAT	TAGCACATCGCGT	ATCTTACAAGTAA
ATACAGCTTCTAT	ATAGCCCGTAGAT	CGTTAGATCGATA
TAGATCGATGAAT	CGTGTATCGATAT	GCCATCGCGTAT

Generate several thousand sequence reads per clone

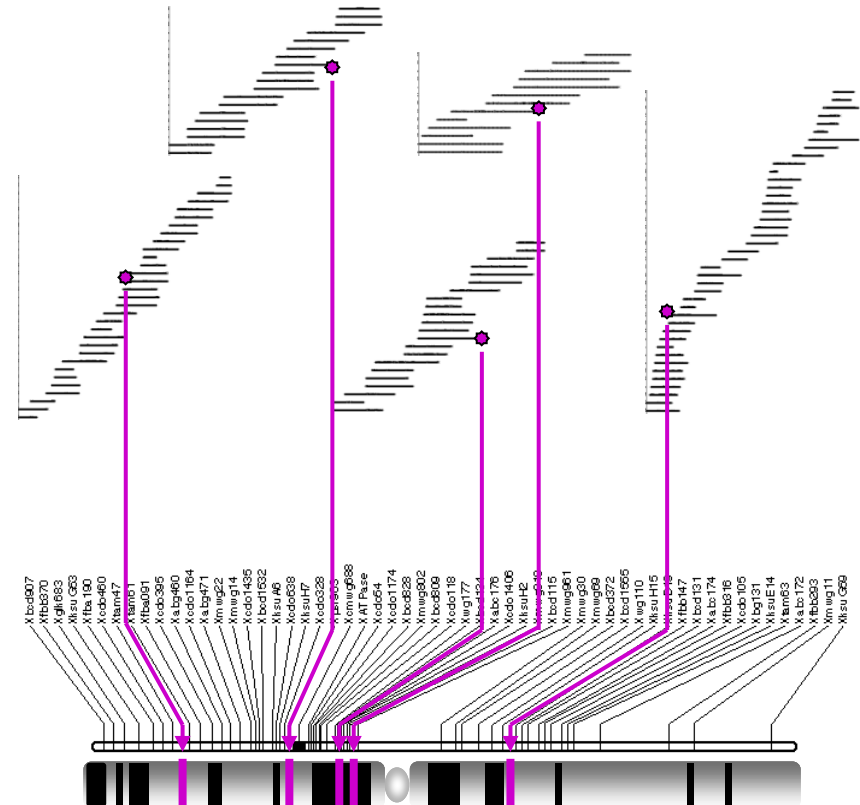


Assemble



BAC by BAC approach

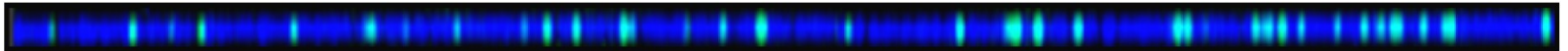
- **Physical map** is composed of contigs of overlapping BAC clones
- BAC contigs are landed on the chromosome through markers comprised in the contigs



# SOLUTIONS FOR THE REPEATS

---

- Long mate-pair reads > 10 kb
- Long read technologies - PacBio, Oxford Nanopore
- Optical mapping
  - Single-molecule mapping of genomic DNA hundreds of kilobases to several megabases in size
  - Creates **sequence-motif maps**, which provide long-range template for ordering genomic sequences
  - Visualisation of reality “Seeing is Believing”



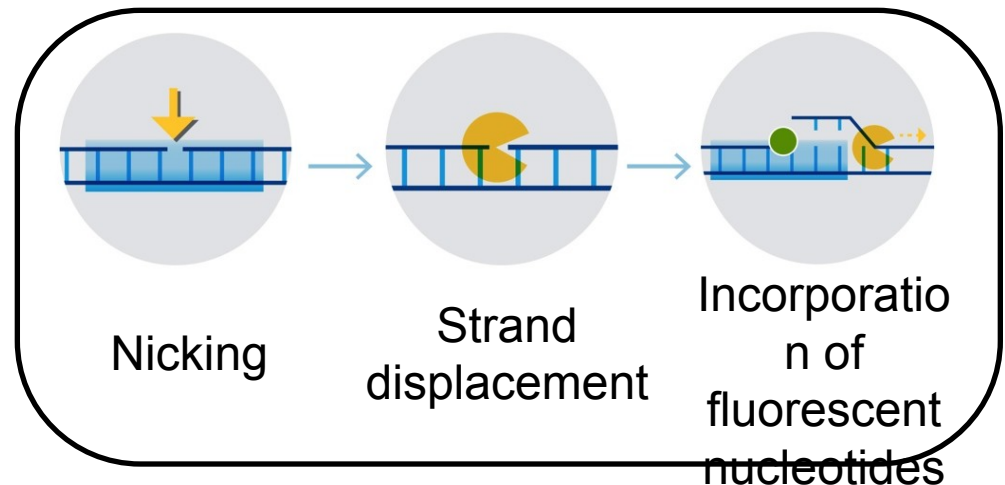
# OPTICAL MAPPING

---

## Three enzymatic approaches

- **restriction enzymes:**  
sequence-specifically cleave DNA  
immobilized on a surface

- **nicking enzymes:**  
fluorescent labelling  
of the nicking site  
in solution (BioNano  
Genomics - Irys)



- **methyltransferase enzymes:**  
labelling with ultra-high density

# BIONANO GENOME MAPPING ON NANOCHANNEL ARRAYS

## 1 Sequence-specific labeling

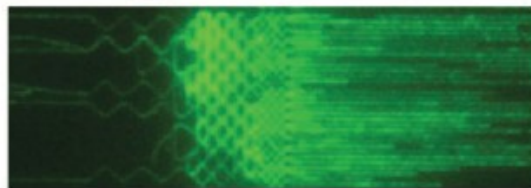
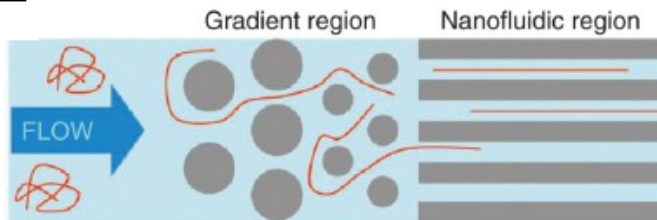
Nickase (Nt.BspQI)

5'-ATGC**GCTCTTC**CATGAATGCGAGC-3'  
3'-TACG**CGAGAAG**GTACTTACGCTCG-5'

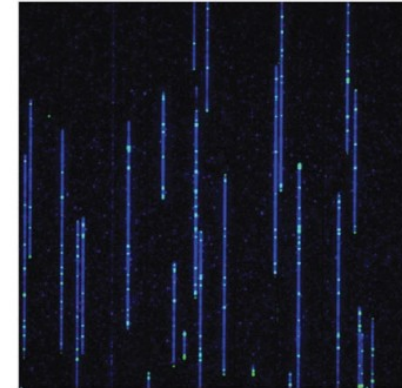
Nick labeling

5'-ATGC**GCTCTTC**CA**U**GAA**U**GCGAGC-3'  
3'-TACG**CGAGAAG**GTACTTACGCTCG-5'

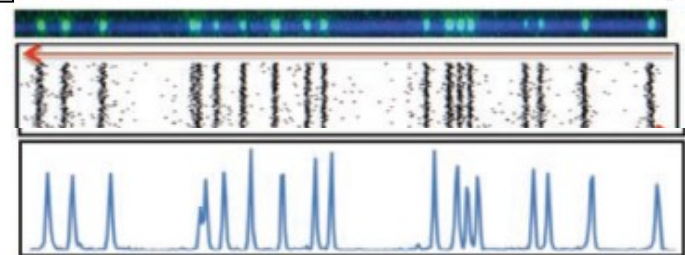
## 2 DNA linearization



## 3 Fluorescence imaging



## 4 Map construction



## 5 Building consensus map

