

# Tin Sinh học Bioinformatics

## **Tổng quan về giải trình tự gen thế hệ mới**

Tài liệu tham khảo

<https://tapchisinhhoc.com/>

# Nội dung chính

## 1. Giải trình tự gen thế hệ mới

### 1.1. Lịch sử phát triển giải trình tự gen thế hệ mới

### 1.2. Giải trình tự thế hệ thứ nhất

### 1.3. Giải trình tự thế hệ thứ hai: Ion Torrent và Illumina

### 1.4. Giải trình tự thế hệ 3 – Single molecular Real Time Sequencing (SMRT)

### 1.5. Giải trình tự thế hệ thứ 4 – Oxford Nanopore DNA Sequencing Technology

### 1.6. Thông số kỹ thuật của một số máy giải trình tự

## 2. Ứng dụng giải trình tự gen thế hệ mới

### 2.1. Ứng dụng NGS trong xét nghiệm di truyền

### 2.2. Xác định đột biến mới bằng NGS

### 2.3. Phát hiện ADN của tế bào ung thư trong hệ tuần hoàn bằng NGS

# 1. Giải trình tự gen thế hệ mới



**GIẢI TRÌNH TỰ  
THẾ HỆ MỚI**

SANGER  
ILLUMINA  
PYROSEQUENCING  
ION TORENT  
SMRT  
NANOPORE

[TAPCHISINHOC.COM](http://TAPCHISINHOC.COM)

CÁCH MẠNG TRONG CÔNG CUỘC HÓA GIẢI SỰ BÍ ẨN CỦA HỆ GEN

The background image shows a computer monitor displaying DNA sequencing chromatograms. The chromatograms show multiple lanes with peaks of different colors (blue, green, yellow, red) representing the four DNA bases. The text 'GIẢI TRÌNH TỰ THẾ HỆ MỚI' is overlaid in large, bold, red letters. To the right of the main title, a list of sequencing technologies is provided: SANGER, ILLUMINA, PYROSEQUENCING, ION TORENT, SMRT, and NANOPORE. Below this list is the website address TAPCHISINHOC.COM. At the bottom, a line of text reads 'CÁCH MẠNG TRONG CÔNG CUỘC HÓA GIẢI SỰ BÍ ẨN CỦA HỆ GEN'. In the bottom right corner, there is a small logo for 'PCH' with the word 'Wageningen' underneath it.

# 1. Giải trình tự gen thế hệ mới

- Giải trình tự gen là kỹ thuật xác định trình tự DNA của các gen cụ thể, mã hóa hoặc không mã hóa cho các protein chức năng liên quan.
- Khởi đầu với sự phát hiện ra cấu trúc DNA, những bước tiến dài đã đạt được trong hiểu biết về tính phức tạp và đa dạng của hệ gen.
- Một loạt đổi mới về hóa chất cũng như thiết bị đã hỗ trợ cho sự khởi đầu của Dự án Hệ gen người.

- Tuy nhiên, hạn chế về lưu lượng xử lý và giá thành cao của giải trình tự là những rào cản chính vào thời điểm đó.
- Sự ra mắt của nền tảng giải trình tự hiệu năng cao đầu tiên vào giữa những năm 2000 đã làm giảm 50.000 lần giá thành giải trình tự hệ gen so với chi phí trước đó của Dự án Hệ gen người, và được đặt tên là: giải trình tự thế hệ mới (NGS – Next Generation Sequencing).

- Giải trình tự gen thế hệ mới là công nghệ cho phép giải mã đồng thời hàng triệu trình tự DNA cùng lúc,
- qua đó giúp nâng cao hiệu suất của quá trình giải mã bộ gen của sinh vật nói chung và hệ gen người nói riêng.

- Trong suốt hơn một thập kỷ qua, các kỹ thuật NGS vẫn tiếp tục “tiến hóa” – tăng lưu lượng xử lý lên 100 đến 1000 lần
- và đã đóng góp vào những đổi mới mang tính cách mạng để đẩy lùi tính phức tạp của hệ gen.



- Những tiến bộ này mang lại phép đọc trình tự với độ dài có thể bằng cả hệ gen (hàng tỉ cặp base), đồng thời giá thành cho một lần giải trình tự hệ gen người giảm chỉ còn 1000 USD, vì thế giải trình tự không chỉ bị bó hẹp trong các nghiên cứu cơ bản mà còn là ứng dụng lâm sàng thường ngày.
- Dầu vậy, những tiến bộ của NGS cũng không phải không có hạn chế.

- Các nền tảng NGS cung cấp lượng dữ liệu cực lớn, nhưng cũng đi liền với tỉ lệ sai số ( $\sim 0.1\text{--}15\%$ ) cao hơn và độ dài mỗi phép đọc thường ngắn hơn (35 – 700 bp đối với các phương pháp đọc trình tự ngắn) so với nền tảng giải trình tự truyền thống của Sanger,
- đòi hỏi phải kiểm tra kỹ lưỡng các kết quả, đặc biệt nếu muốn phát hiện biến dị và các ứng dụng lâm sàng.

- Mặc dù giải trình tự đoạn dài đã vượt qua hạn chế về kích thước của NGS, nó vẫn khá đắt đỏ và có lưu lượng xử lý thấp hơn các nền tảng khác.
- Cuối cùng, NGS đang phải cạnh tranh với các kỹ thuật thay thế khác có khả năng thực hiện các ứng dụng tương tự mà giá lại rẻ hơn, như DNA microarray, NanoString, qPCR.

- Bài viết dưới đây đánh giá các cách tiếp cận khác nhau trong công nghệ giải trình tự gen thế hệ mới – NGS cũng như làm rõ những tiến bộ mới đây trong lĩnh vực này đang thay đổi con đường nghiên cứu hệ gen như thế nào.
- Các phương pháp được đề cập chi tiết cùng với những ưu thế và nhược điểm.

# 1.1. Lịch sử phát triển giải trình tự gen thế hệ mới

- Thế hệ 1:
- Phương pháp của Frederick Sanger.
- Dừng ngẫu nhiên phản ứng tổng hợp chuỗi DNA bằng dideoxynucleotide

# *Thế hệ 2*

- Phương pháp Pyrosequencing. Đo lượng PPI giải phóng trong phản ứng tổng hợp chuỗi DNA
- Phương pháp Ion Semiconductor. Đo tín hiệu ion  $H^+$  giải phóng ra trong phản ứng tổng hợp chuỗi DNA
- Phương pháp của Illumina. Đo tín hiệu từng loại huỳnh quang gắn vào nucleotide trong phản ứng tổng hợp chuỗi DNA

# Thế hệ 3

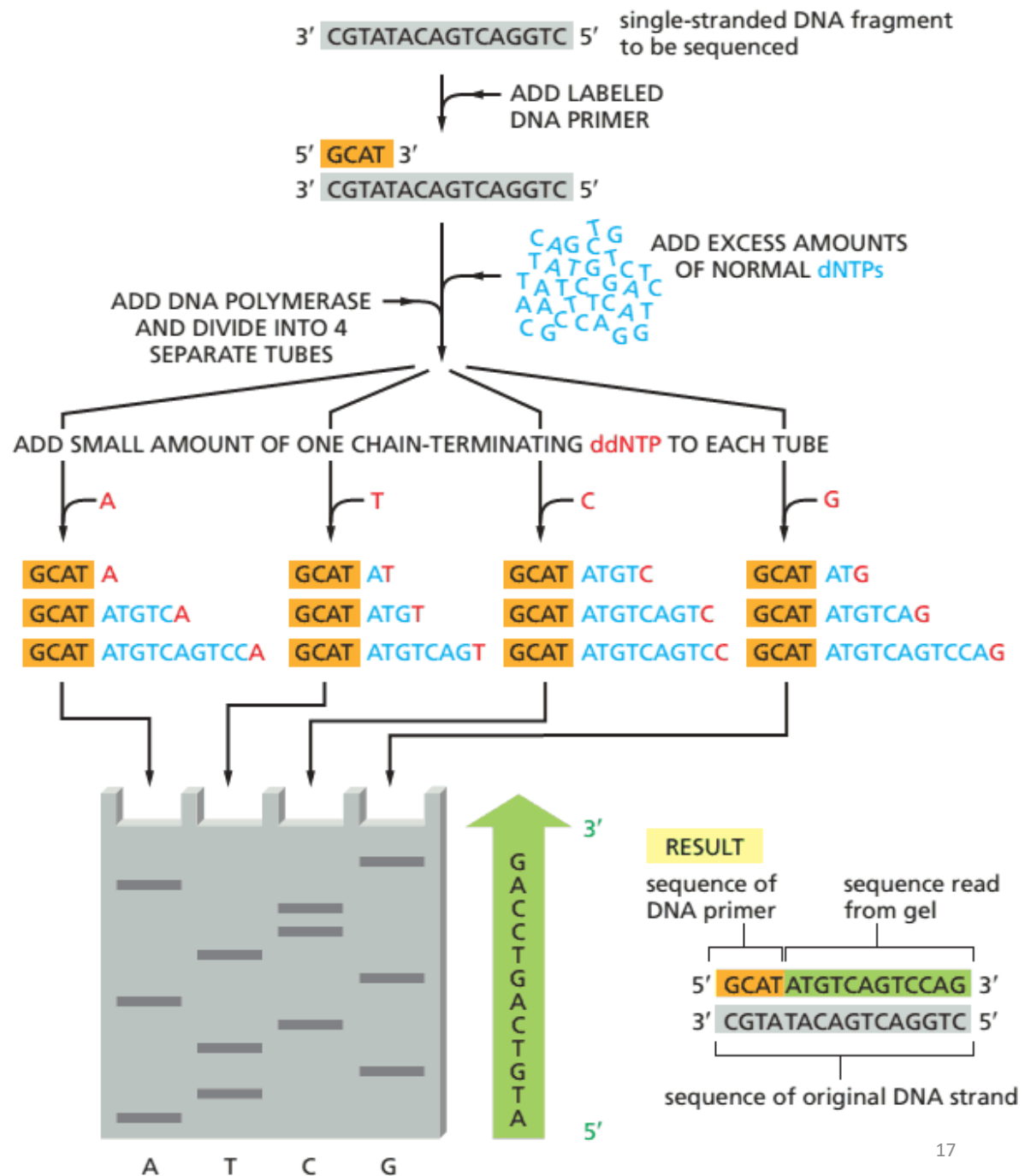
- Giải trình tự dựa trên các kỹ thuật phân tích tín hiệu đơn phân tử.
- Phân tích tức thời tín hiệu đơn phân tử huỳnh quang gắn vào nucleotide trong phản ứng tổng hợp 1 chuỗi DNA trong lỗ nano.
- Và đã đang tiếp cận đến thế hệ thứ 4

## 1.2. Giải trình tự thể hệ thứ nhất

- Nguyên tắc: thực hiện phản ứng PCR với khuôn là trình tự cần biết, quá trình tổng hợp đang diễn ra thì làm ngừng ngẫu nhiên.
- Hỗn hợp phản ứng chứa các đoạn DNA mới được tổng hợp có kích thước hay điểm làm ngừng tổng hợp đại diện cho tất cả các đoạn DNA khác biệt nhau chỉ một nu.



# Nguyên lý của giải trình tự Sanger



# *Quy trình*

- Trước khi DNA được giải trình tự, nó phải được biến tính thành các mạch đơn nhờ nhiệt độ.
- Tiếp theo, một mồi gắn vào một trong số các mạch khuôn đơn.
- Mồi được đánh dấu bằng phóng xạ hay huỳnh quang nên sản phẩm cuối cùng có thể được phát hiện trên bản gel điện di.
- Khi mồi đã gắn vào DNA, dung dịch được chia làm 4 ống, ghi chú là “A”, “T”, “G”, và “C”.

Sau đó các chất được bổ sung vào các ống này:

- Ống “G”: ddGTP và DNA polymerase
- Ống “A”: ddATP và DNA polymerase
- Ống “T”: ddTTP và DNA polymerase
- Ống “C”: ddCTP và DNA polymerase

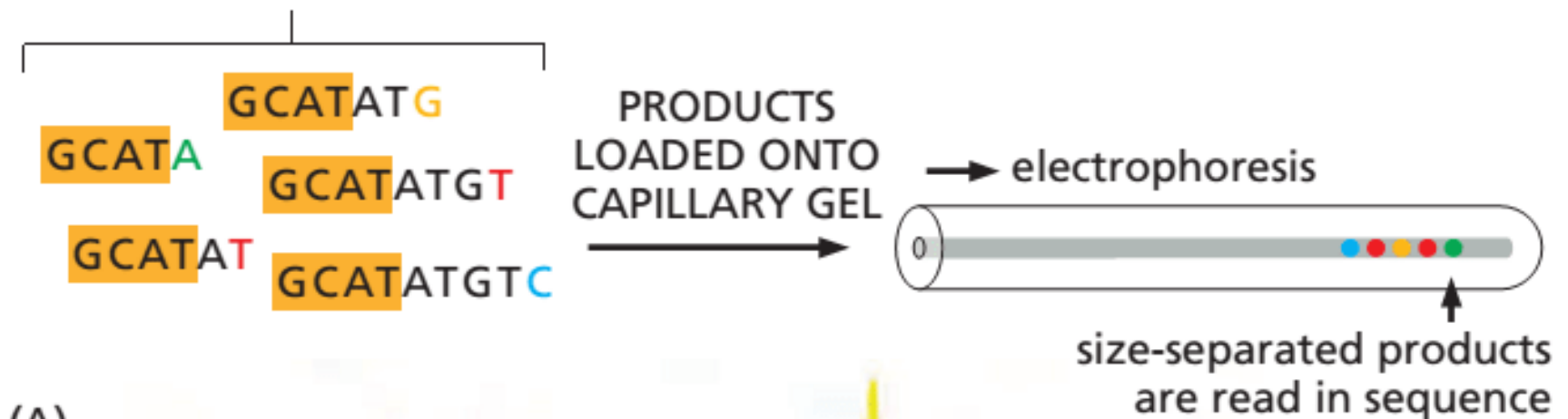
- Mỗi ống chứa nồng độ của một loại ddNTP với nồng độ thấp hơn hàng trăm lần nồng độ dNTP.
- Khi tổng hợp DNA đang diễn ra, dNTP được gắn vào mạch đang kéo dài, nhưng ngẫu nhiên ddTNP cũng có thể gắn vào và lập tức làm ngừng chuỗi.

- Vậy là các sản phẩm khuếch đại mang cách kích thước khác nhau từng nucleotide được tạo ra.
- Sản phẩm từ bốn ống “A”, “T”, “G”, “C” được biến tính lần 2 và điện di riêng trên gel polyacrylamide nhằm phân tách các sản phẩm theo kích thước.
- Gel được chiếu tia UV hoặc tia X, chụp ảnh và phân tích.

- Sau này, tất cả 4 phản ứng trên được dồn vào một ống và diễn ra đồng thời, trong đó mỗi dNTP được đánh dấu bằng một chất huỳnh quang riêng biệt.
- Hỗn hợp sản phẩm được đi qua hệ thống điện di mao quản, phân tách các trình tự khác nhau từng nucleotide.

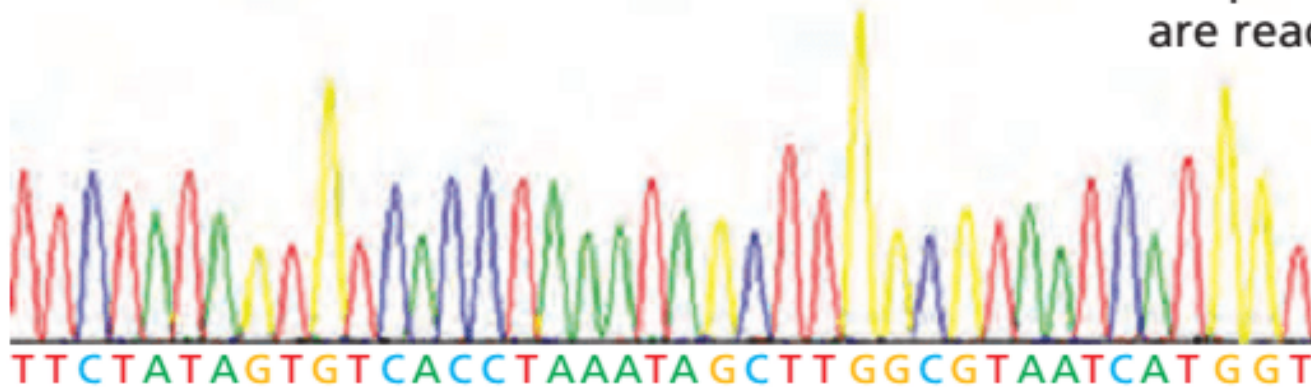
# AUTOMATED DIDEOXY SEQUENCING

mixture of DNA products, each containing a chain-terminating ddNTP labeled with a different fluorescent marker



(A)

(B)



## 1.3. Giải trình tự thế hệ thứ hai

- Bắt đầu từ thế hệ thứ 2, các kỹ thuật giải trình tự được gọi chung là NGS – next generation sequencing.



# *Pyrosequencing*

- Pyrosequencing là một phương pháp giải trình tự DNA dựa trên nguyên lý “đọc trình tự bằng tổng hợp”.
- Nó khác với giải trình tự Sanger ở chỗ: dựa trên sự phát hiện sự giải phóng pyrophosphate (PPi) khi dNTP được thêm vào chuỗi, thay vì kết thúc chuỗi bằng ddNTP.

- Genomic DNA
- cDNA fragments
- BAC libraries

- Ditag libraries
- ChIP fragments (low molecular weight DNA)

- PCR products

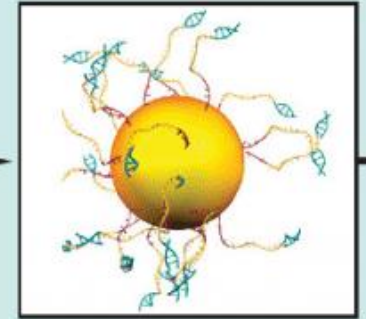
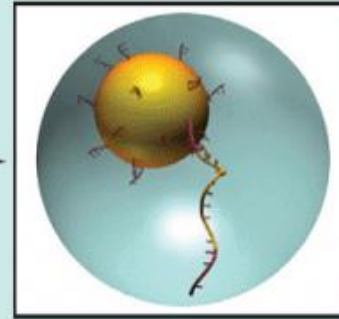
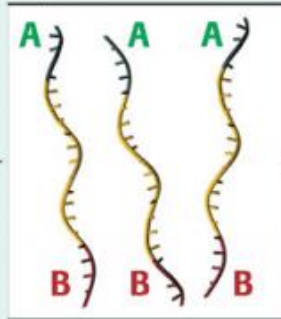
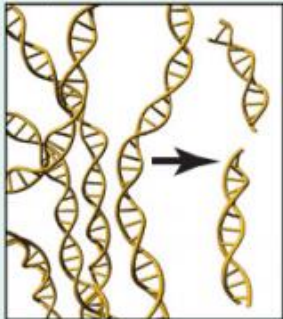
GS DNA Library Preparation Kit



Tissue Lyser



Appropriate GS emPCR Kit



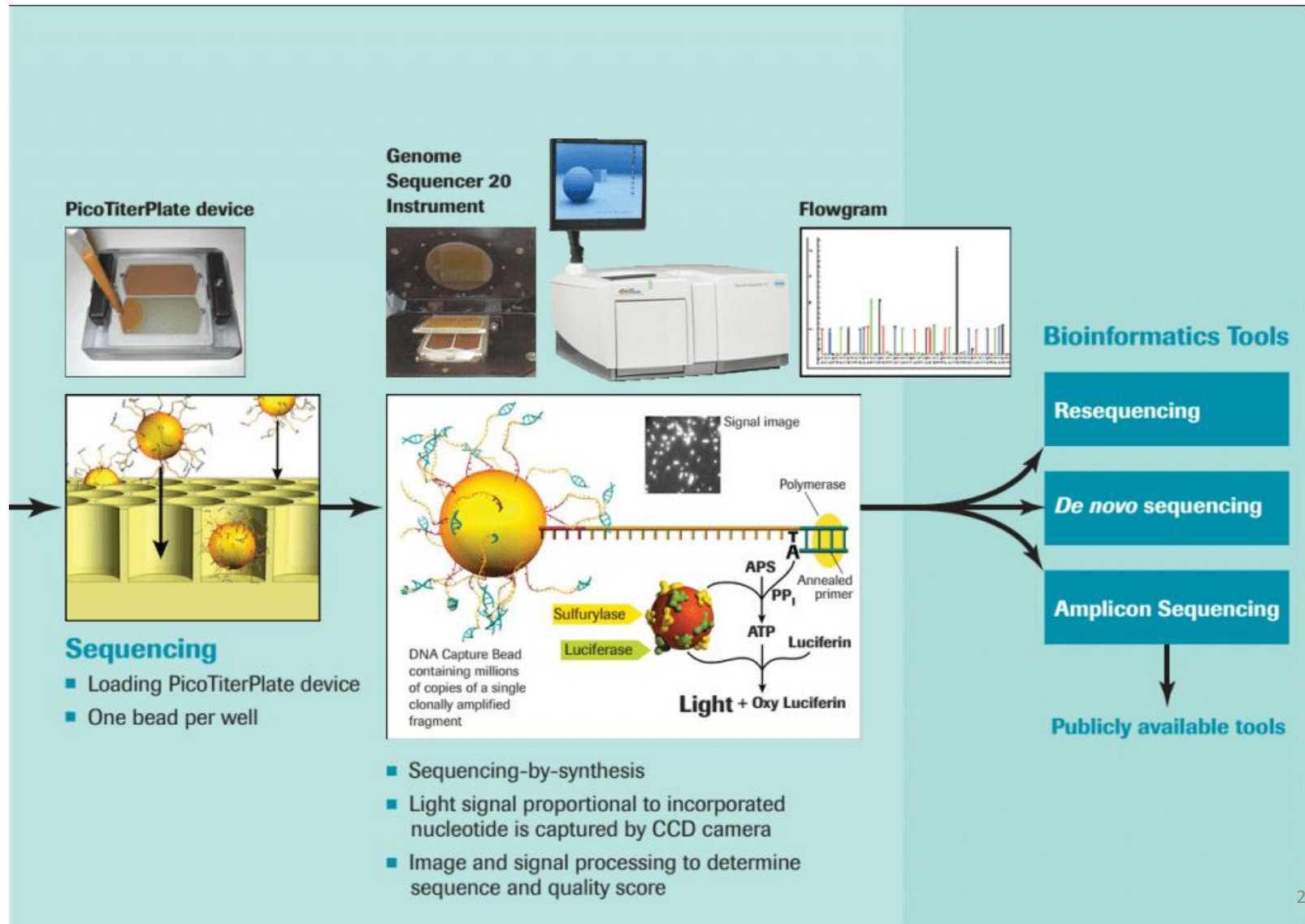
## DNA Library Preparation and Titration

- Fragmentation (needed depending on the starting material)
- Prepare sstDNA library with adaptors
- One library provides enough DNA for thousands of sequence runs
- Determine amount of sstDNA for the emPCR (titration)

## emPCR

- Emulsification
- Clonal amplification of sstDNA on beads
- Parallel amplification of the entire library in one PCR reaction
- sstDNA ready to sequence

# Tóm tắt quy trình và nguyên lý của pyrosequencing



# Quy trình:

## *Bước 1: Tạo thư viện DNA.*

- DNA được cắt nhỏ thành các mảnh nhỏ (300 – 800 bp) với enzyme cắt đầu tù.
- Các adapter ngắn (A và B) sau đó được nối vào hai đầu của mỗi mảnh DNA.
- Các adapter này sẽ làm trình tự mỗi cho emPCR và giải trình tự, đồng thời chứa một vị trí bám streptavidin để tinh sạch mẫu.
- Thư viện DNA khuôn mạch đơn (sst-DNA) ở cuối bước này được đánh giá về chất lượng, và lượng tối ưu cần cho emPCR được xác định bằng chuẩn độ.

## DNA Library Preparation and Titration

4.5 hours

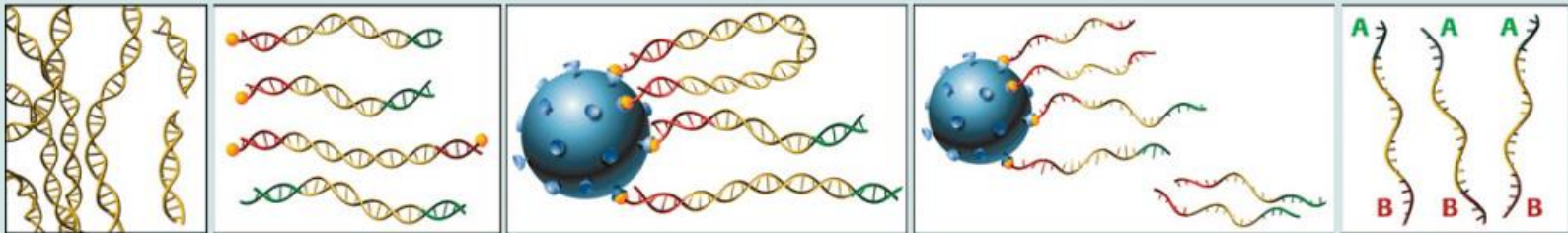
10.5 hours

## emPCR

8 hours

## Sequencing

5.5 hours



**gDNA**

**sstDNA library**

- Genome fragmented by nebulization
- No cloning; no colony picking
- sstDNA library created with adaptors. The adaptors are used as primers, and for binding to beads.
- A/B fragments selected using streptavidin-biotin purification

## ***Bước 2: PCR nhũ tương (emPCR)***

- Thư viện sstDNA được cố định lên DNA capture bead đặc biệt. Mỗi bead có mang một đoạn sst-DNA duy nhất.
- Thư viện đính hạt bead được nhũ tương hóa với các hóa chất dùng cho khuếch đại trong một hỗn hợp dầu trong nước.

- Mỗi bead được bắt giữ riêng rẽ trong một lò phản ứng siêu nhỏ (microreactor) để thực hiện PCR.
- Sự khuếch đại tạo ra các mảnh DNA giống nhau cố định trên bead và đặc hiệu theo bead.
- Sau khuếch đại đã hoàn tất, hỗn hợp chứa bead được chuyển lên đĩa PicoTiter chứa các giếng đường kính 44  $\mu\text{m}$  – chỉ vừa cho một bead lọt vào.



## DNA Library Preparation and Titration

4.5 hours

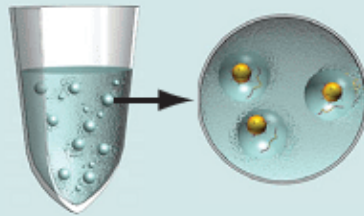
10.5 hours

## emPCR

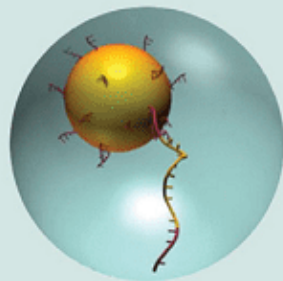
8 hours

## Sequencing

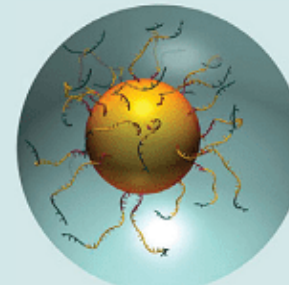
5.5 hours



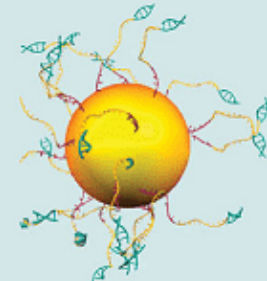
Anneal sstDNA  
to an excess of  
DNA Capture  
Beads



Emulsify beads and  
PCR reagents in  
water-in-oil  
microreactors



Clonal amplification  
occurs inside  
microreactors



Break microreactors,  
enrich for DNA-  
positive beads

sstDNA library

Clonally-amplified sstDNA attached to bead  
(millions of copies per bead)



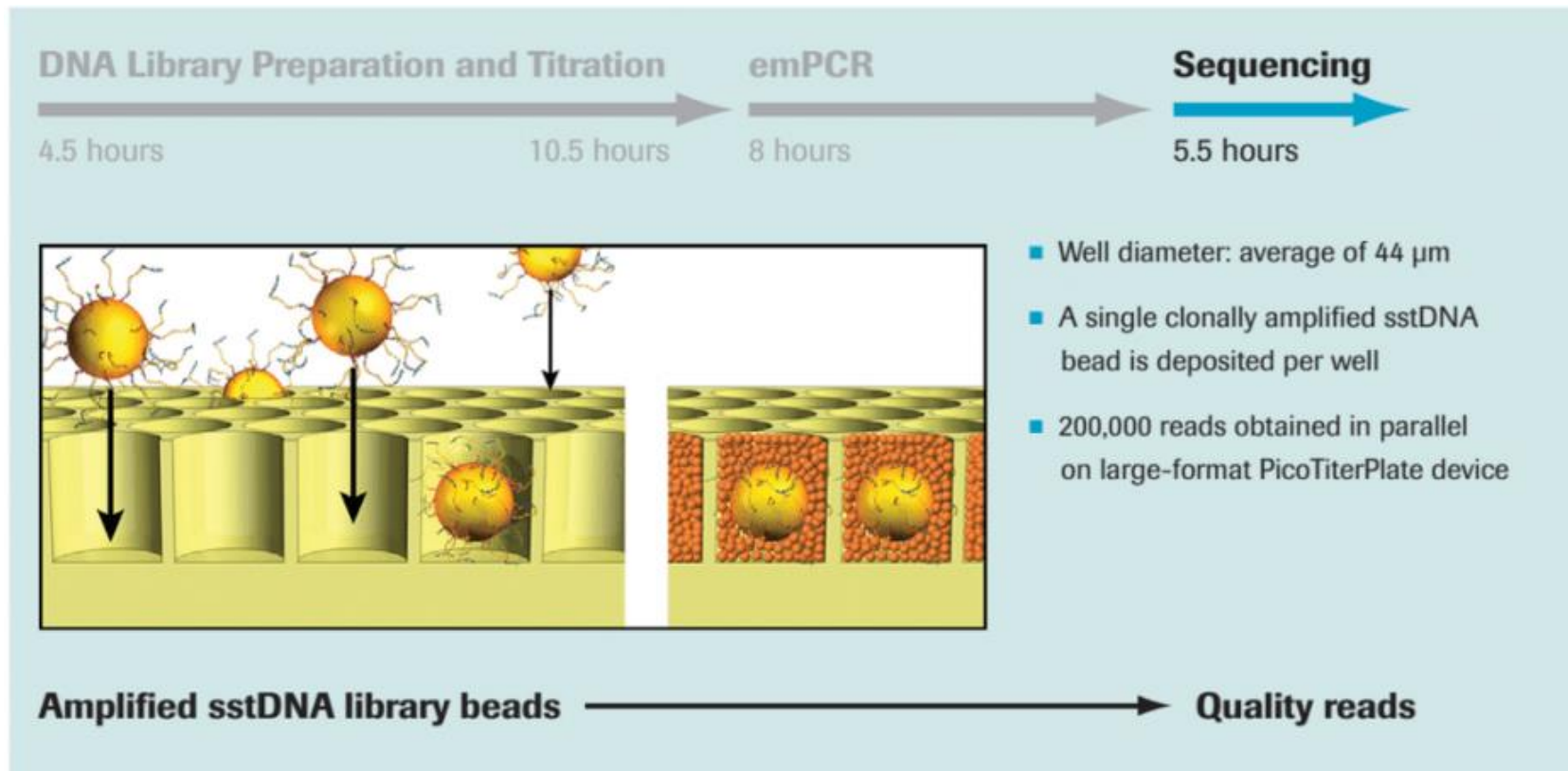
### ***Bước 3: Giải trình tự các đoạn DNA đã gắn trên mỗi bead từ theo nguyên lý pyrosequencing***

- Trước hết, một loại dNTP nhất định được bơm vào các giếng trên đĩa Picotiter.
- DNA polymerase xúc tác cho phản ứng kéo dài chuỗi DNA, tại đó các dNTP được liên kết vào chuỗi DNA sẽ giải phóng ra Pyrophosphate (PPi).
- Số lượng PPi giải phóng sẽ tỷ lệ với số lượng phân tử của 1 loại dNTP được gắn vào chuỗi.

- PPi phản ứng với APS (adenosine 5'-phosphosulfate) để giải phóng ATP dưới sự xúc tác của enzym sulfurylase.
- ATP là cơ chất cho luciferase, xúc tác cho phản ứng chuyển hóa luciferin thành oxyluciferin phát ra ánh sáng nhìn thấy, có thể ghi lại bởi camera.
- Cường độ sáng sẽ tỷ lệ với số lượng phân tử ATP được tạo thành.

- Kết thúc phản ứng, các dNTP còn thừa lại sẽ bị phân hủy bởi Apyrase.
- Bốn loại dNTP sẽ được bơm vào trong giếng phản ứng luân phiên nối tiếp nhau.

# Các phản ứng diễn ra trong một lần đọc trình tự bằng pyrosequencing



## DNA Library Preparation and Titration

4.5 hours

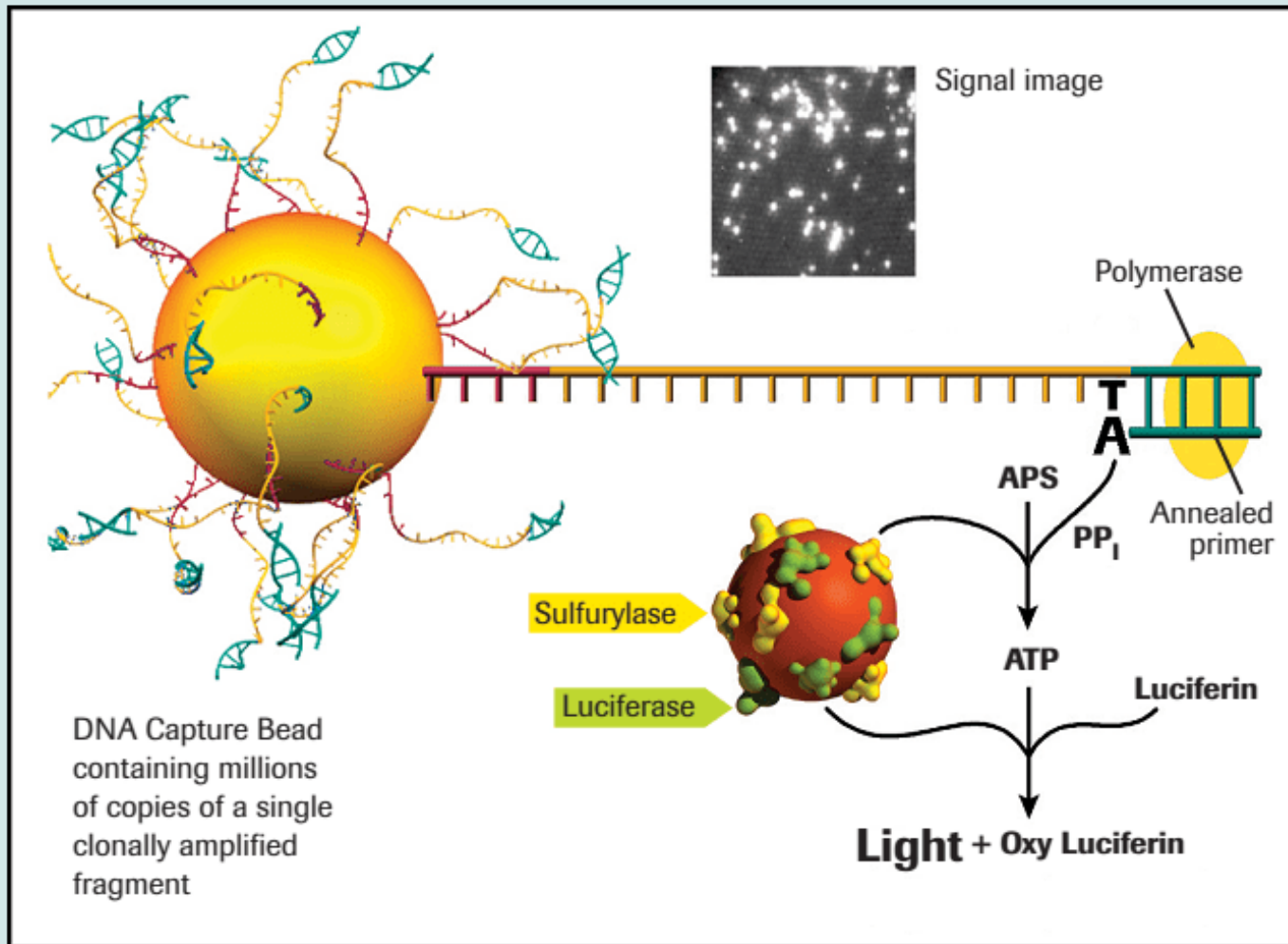
10.5 hours

## emPCR

8 hours

## Sequencing

5.5 hours



- Bases (TACG) are sequentially flowed (42 times)
- Chemiluminescent signal generation
- Signal processing to determine base sequence and quality score

**Amplified sstDNA library beads** → **Quality reads**

## ***Bước 4: Ghép nối các đoạn DNA đã được giải trình tự bằng phần mềm tin sinh để tạo ra một trình tự đầy đủ của genome.***

- Các thế hệ máy dùng nguyên lý này là Pyrosequencing AB của Viện Công nghệ Hoàng Gia, Uppsala, Thụy Điển và 454 Life Sciences GS của hãng Roche.
- Với thế hệ giải trình tự này, giá thành giải trình tự genome từ 5000 – 7000 USD giảm còn 1000 – 2000 USD.

# Ưu điểm

- Độ dài đọc được là 400 bp (tương đối dài) với độ chính xác càng về sau càng tăng
- Công nghệ cao hơn Sanger, giải trình tự hệ gen mà không cần tách dòng tạo thư viện
- Thư viện DNA có thể được mã hóa và tách riêng trong suốt quá trình phân tích kết quả

# Nhược điểm

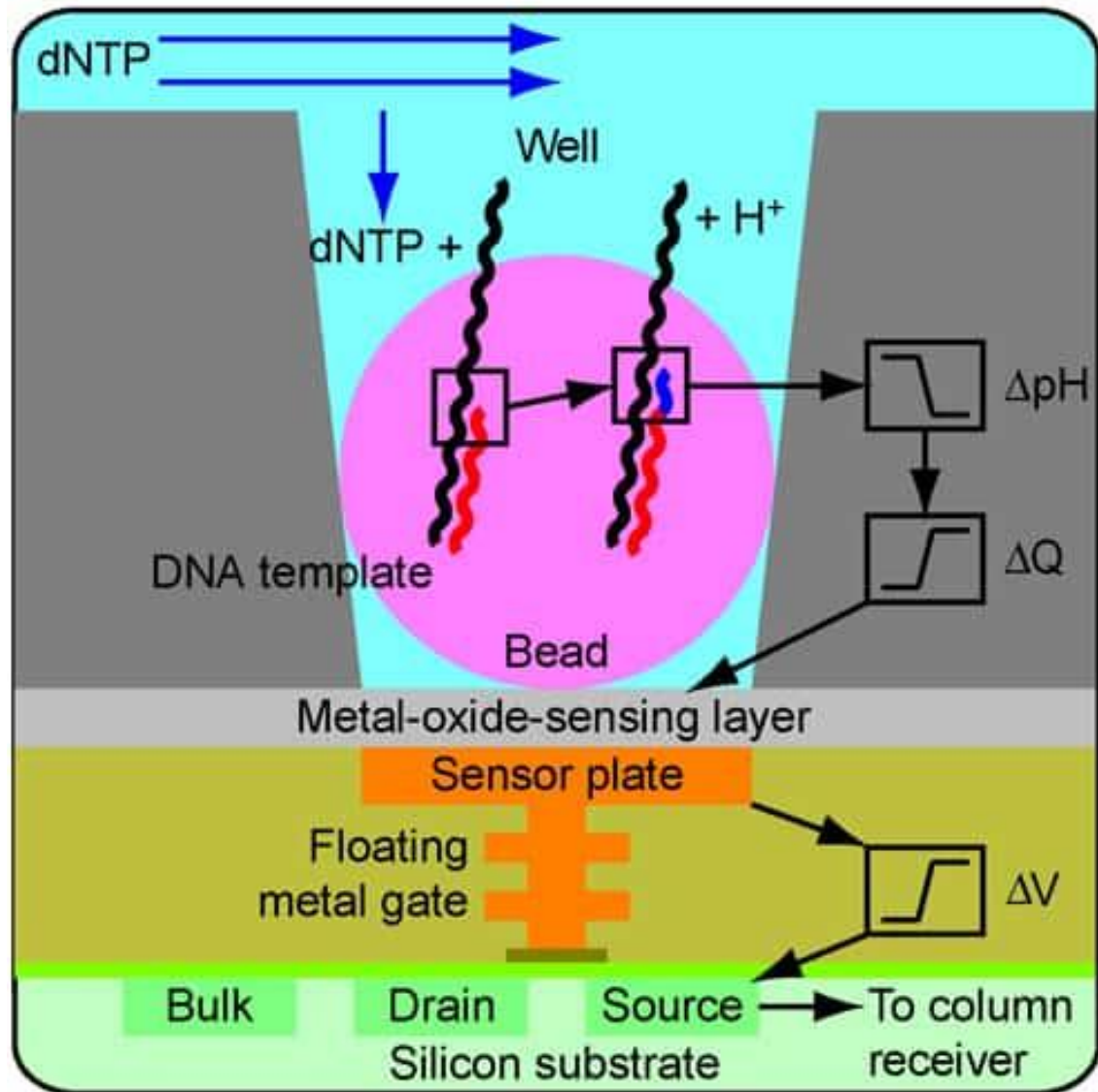
- Khó phân tích chính xác các trình tự của 1 loại Nucleic acid lặp lại liên tục từ 6 nu trở lên, ví dụ TTTTTT
- Phức tạp và nhiều công đoạn
- Giá thành cao so với các phương pháp thế hệ 2 khác.



## *2. Ion Torrent sequencing*

### *Ion semiconductor sequencing*

- Kỹ thuật xác định trình tự các nucleotide trong DNA thông qua việc phát hiện tín hiệu điện do ion  $H^+$  được giải phóng ra trong quá trình tổng hợp DNA bằng các máy đo pH siêu nhỏ gắn vào chip cảm ứng bán dẫn.

**a**

## *Bước 1: Tạo thư viện DNA*

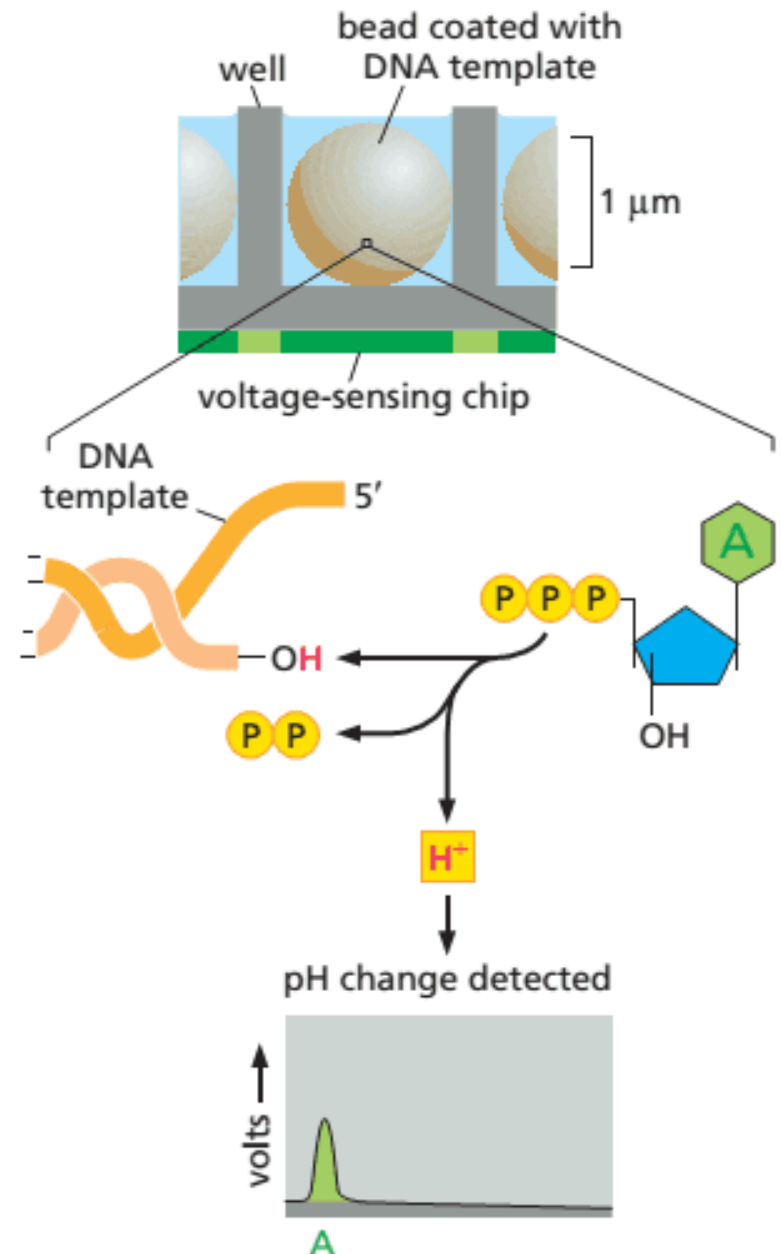
- Cắt DNA hệ gen bằng enzyme giới hạn đầu tù
- Gắn 2 đầu adapter A và P1 bằng ligase
- Nhân bản đặc hiệu bằng phản ứng PCR các trình tự có 2 đầu A và P1 bằng mỗi có trình tự bổ sung với adapter A và P1.
- Gắn chọn lọc mạch xuôi và ngược 5'P1 - ssdDNA-3'A lên các microbead (Ion Sphere) với tỷ lệ gắn là 1 bead: 1 DNA.
- Với việc gắn cả hai đầu, giải trình tự được tiến hành hai chiều, giúp tăng độ chính xác.

## *Bước 2*

- PCR nhũ dịch và đưa các bead vào các giếng có chip gắn máy đo pH siêu nhỏ.
- Mỗi đĩa có khoảng 1,2- 660 triệu giếng tùy theo thể hệ máy.

## Bước 3

- Giải trình tự các đoạn DNA đã gắn trên mỗi bead theo nguyên lý ion semiconductor.
- Đầu tiên, một loại dNTP được bơm vào các giếng trên đĩa.



- DNA polymerase xúc tác cho phản ứng kéo dài chuỗi DNA, tại đó từng dNTP được thêm vào chuỗi DNA sẽ giải phóng ra PPi và H<sup>+</sup>.
- Số lượng H<sup>+</sup> giải phóng sẽ tỷ lệ với số lượng phân tử của 1 loại dNTP được gắn vào chuỗi.

- $H^+$  giải phóng sẽ làm giảm pH và tăng điện tích ở khay cảm ứng bán dẫn, dẫn tới chênh lệch điện thế của bộ phận truyền cảm ứng và xuất hiện dòng điện.
- Tín hiệu điện hóa sẽ được ghi nhận bởi phần mềm máy tính.
- Từng loại dNTP sẽ được bơm vào trong giếng phản ứng luân phiên nối tiếp nhau.

## *Bước 4*

- Ghép nối các đoạn DNA đã được giải trình tự bằng phần mềm tin sinh để tạo ra một trình tự đầy đủ của genome.
- Các thế hệ máy Ion Torrent PGM hiện nay: PGM314, PGM316, PGM318, PI, PII (2015)



# Ưu điểm

- Sử dụng công nghệ bán dẫn chứ không phải quang học để đo tín hiệu nên thời gian nhanh, hệ thống máy gọn nhẹ
- Thời gian nhanh: 2-3 giờ để phân tích hệ gen người
- Giá thành hóa chất rẻ hơn so với các kỹ thuật giải trình tự NGS khác
- Thư viện DNA có thể được mã hóa và tách riêng trong suốt quá trình phân tích kết quả

# Nhược điểm

- Chiều dài đọc chỉ khoảng 100-250 bp (độ chính xác 99% tại base thứ 250 và cao hơn ở các base phía trước)
- Các bước thao tác phức tạp, nhiều công đoạn
- Giá thành hóa chất vẫn cao so với pp Sanger
- Khó đếm được chính xác trình tự lặp lại liên tục của 1 loại nucleotide (vd 7 nucleotide A khó phân biệt với 8 nucleotide A)

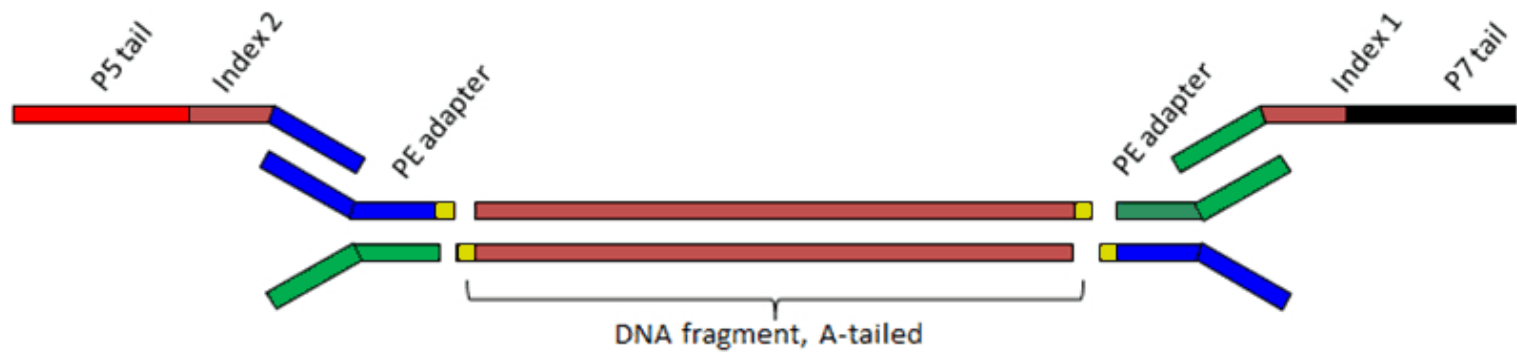
# ***Giải trình tự Illumina***

# *Illumina (Solexa)*

## Quy trình

- Bước 1. Tạo thư viện
- Cắt DNA genome thành các mảnh DNA
- Điện di để chọn lọc các mảnh DNA có kích thước khoảng 200-300 bp.
- DNA được xử lý để mặc định gắn thêm 1 A vào đầu 3' hoặc 5' của mỗi mạch đơn.
- Sau đó 2 loại PE adapter có chữ T nhô ra được nối với DNA đích (PE adapter).
- PE adapter có vùng bám cho mỗi PCR vì thế cho phép khuếch đại DNA đích với mỗi mà ở đầu 5' có gắn thêm trình tự Index2-P5 và Index1-P7 (có thể chỉ cần 1 Index cũng được). (tổng cộng phần gắn thêm khoảng 60 nu).

### Dual indexed paired-end library



# Library Prep is Critical for Successful Sequencing

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



Dual Index Library shown

P5 and P7 sequences are complimentary to oligos bound to flow cell surface and *required* for any library

Indexes are used to tag individual samples to allow pooling of multiple libraries

Rd1 and Rd2 sequencing primers regions are used to initiate sequencing

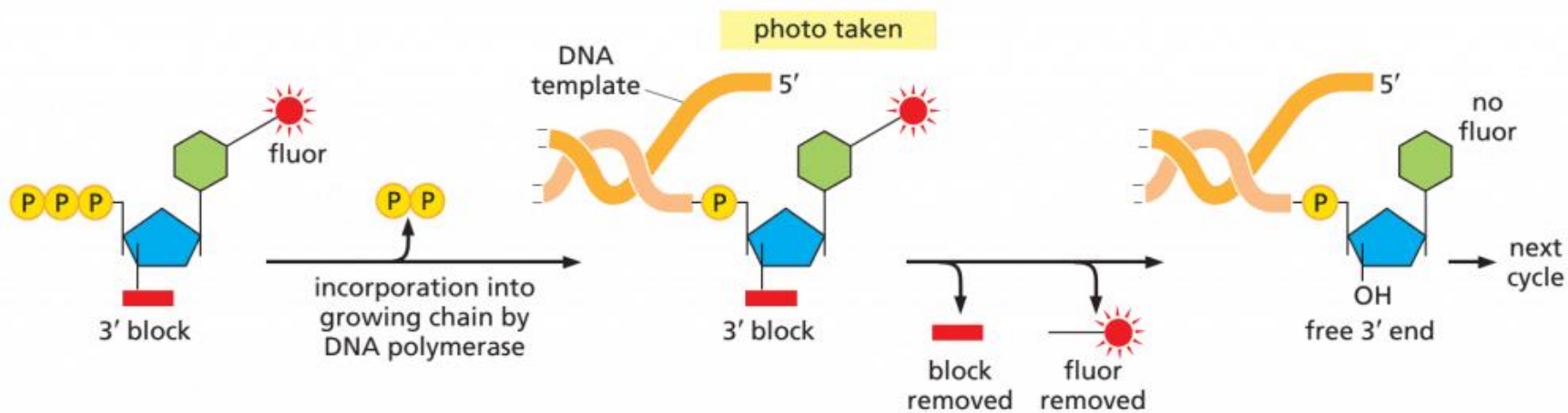
## *Bước 2*

- Polony-PCR tạo các cluster DNA
- Trên một tấm kính có các giếng phủ đầy các môi bổ sung với P5 và P7, đã cố định đầu 5' trên mặt giếng. DNA thư viện đã tạo ở trên được biến tính thành mạch đơn (bằng NaOH) và thả vào giếng để cho liên kết bổ sung với môi.
- Mỗi mạch đơn DNA được nhân bản thành hàng triệu (n) copy bằng kỹ thuật PCR theo nguyên tắc “bắc cầu”, hình thành các cụm trình tự (cluster)
- Cắt và rửa loại bỏ mạch bổ sung (Rv-ngược) và giữ lại mạch chính (Fw-xuôi) để toàn bộ các trình tự trên cluster là 1 chiều phục vụ cho việc giải trình tự chiều thứ nhất.

# *Bước 3*

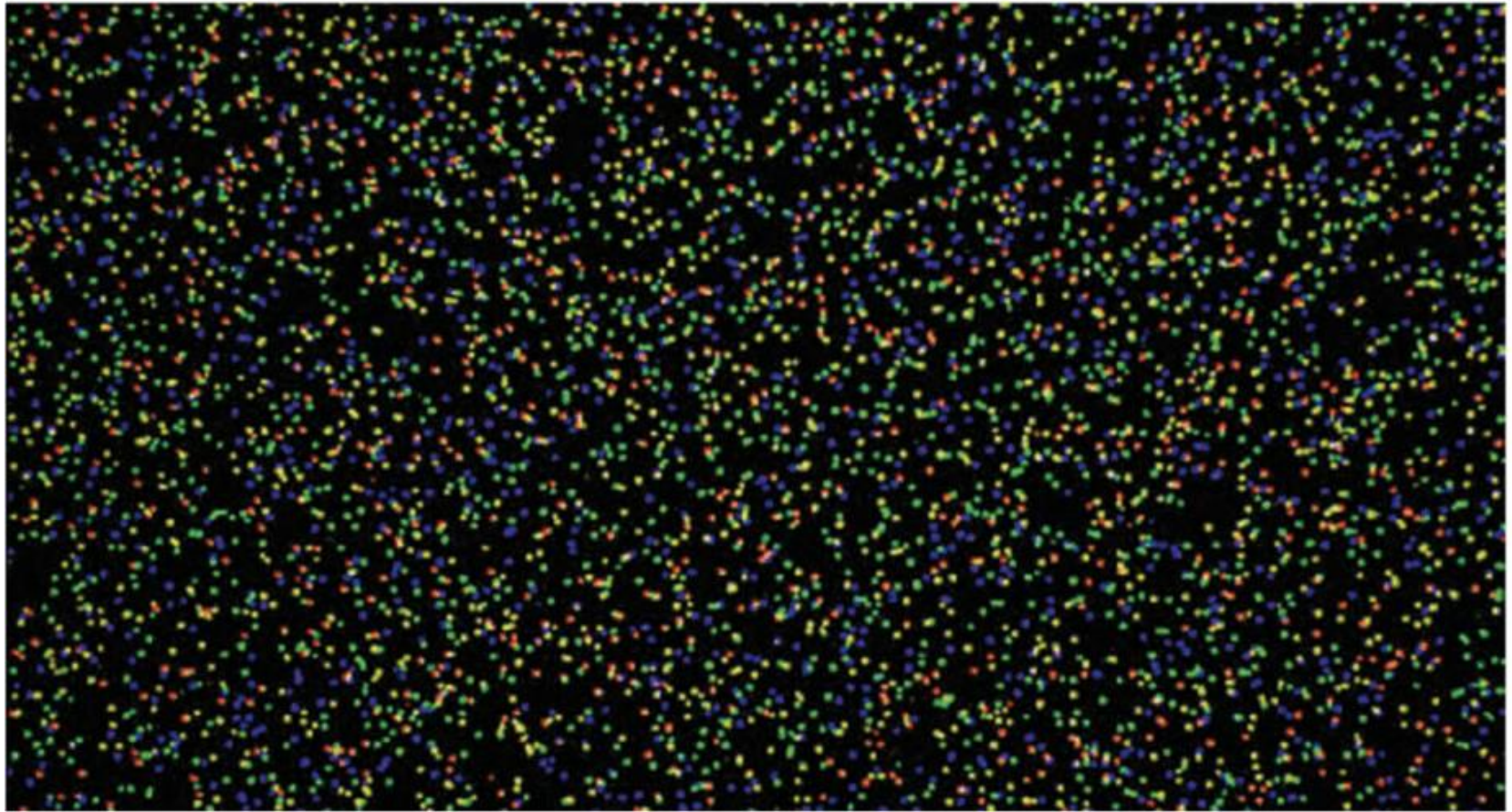
- Giải trình tự DNA trong từng cluster
- Thả DNA polymerase và mỗi vào (mỗi lúc này gọi là Read1 Seq primer bắt cặp bổ sung với PE adapter) cùng với 4 loại nu gắn huỳnh quang mang gốc khóa ở 3'OH. Nu bắt cặp bổ sung được gắn vào nhưng phản ứng tạm dừng lại.
- Tia Laser chiếu vào để kích thích phát huỳnh quang và một camera ghi lại huỳnh quang tương ứng với nu
- Gốc khóa được cắt ra, huỳnh quang cũng bị loại bỏ và rửa trôi cùng với các nu không gắn. Chu trình lặp lại đến khi hết độ dài trình tự quan tâm (hết 1 chiều). Sản phẩm đọc chiều 1 được rửa đi.





- Tiếp tục bổ sung Index1 primer và các thành phần PCR , cơ chế giải trình tự tương tự trên đến khi hết phần Index, sản phẩm đọc được rửa đi.
- Một lần nữa PCR bắt cầu lại được lặp lại tạo thành cụm, nhưng lần này DNA gắn với môi xuôi (Fw) bị cắt và rửa trôi.
- Thả DNA polymerase và môi vào (môi lúc này là Read2 Seq primer bắt cặp với PE' adapter) cùng với 4 loại nu gắn huỳnh quang. Giải trình tự tiếp tục.

Ảnh chụp một tấm kính nền gắn các cluster trình tự trong lúc giải trình tự diễn ra.

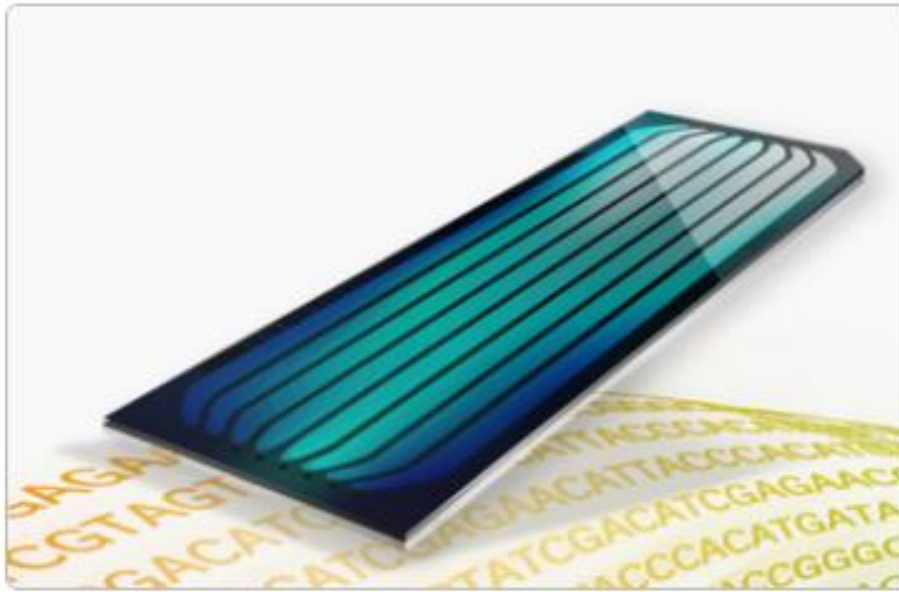


100 μm

## Bước 4

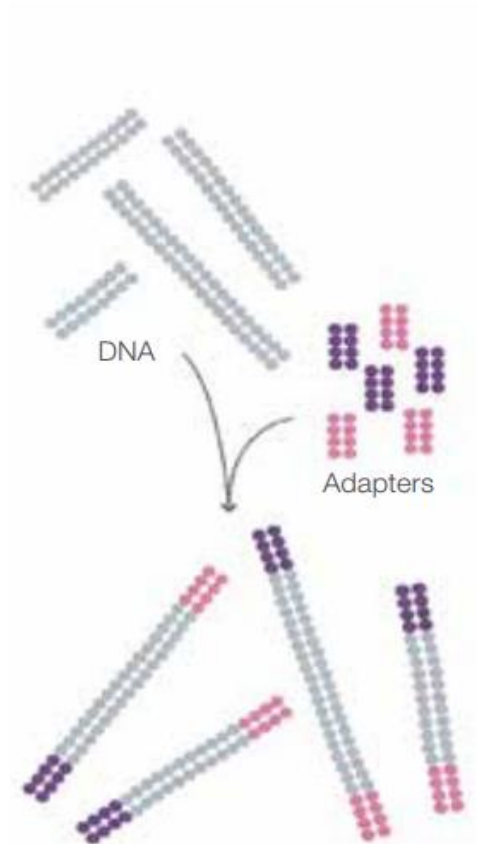
- Ráp nối các đoạn trình tự ghi nhận được và xử lý kết quả.
- Các hệ thống máy cho Illumina Seq: HiSeq, HiScanSQ, Genome Analyzer Ix, MiSeq

# Illumina Sequencing Technology



- Illumina sequencing technology, sequencing by synthesis (SBS), is a widely adopted next-generation sequencing (NGS) technology worldwide, responsible for generating more than 90% of the world's sequencing data

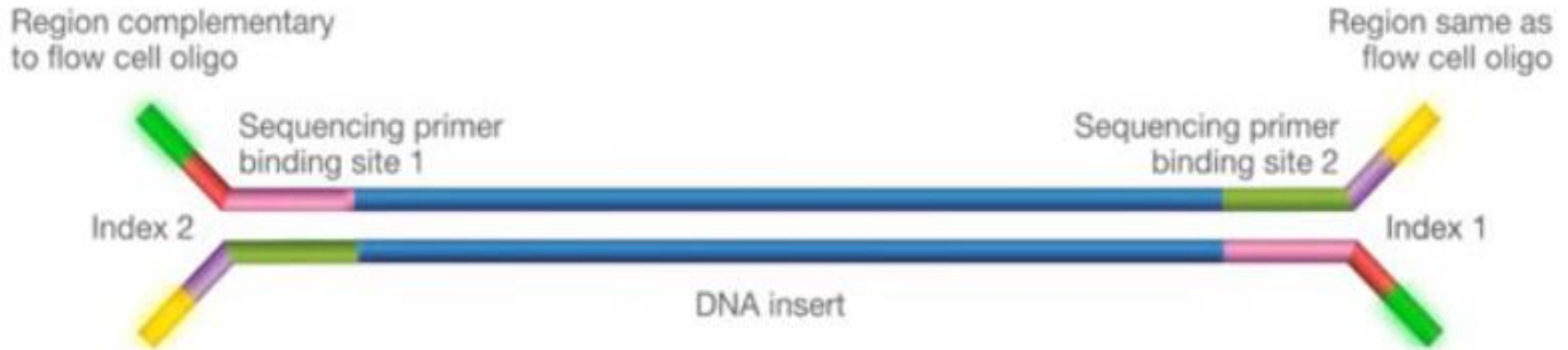
Figure 2: Prepare Genomic DNA Sample



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



## Reduced cycle amplification



Adaptor:

- Sequencing primer binding site
- Index
- Regions complementary to flow cell oligo

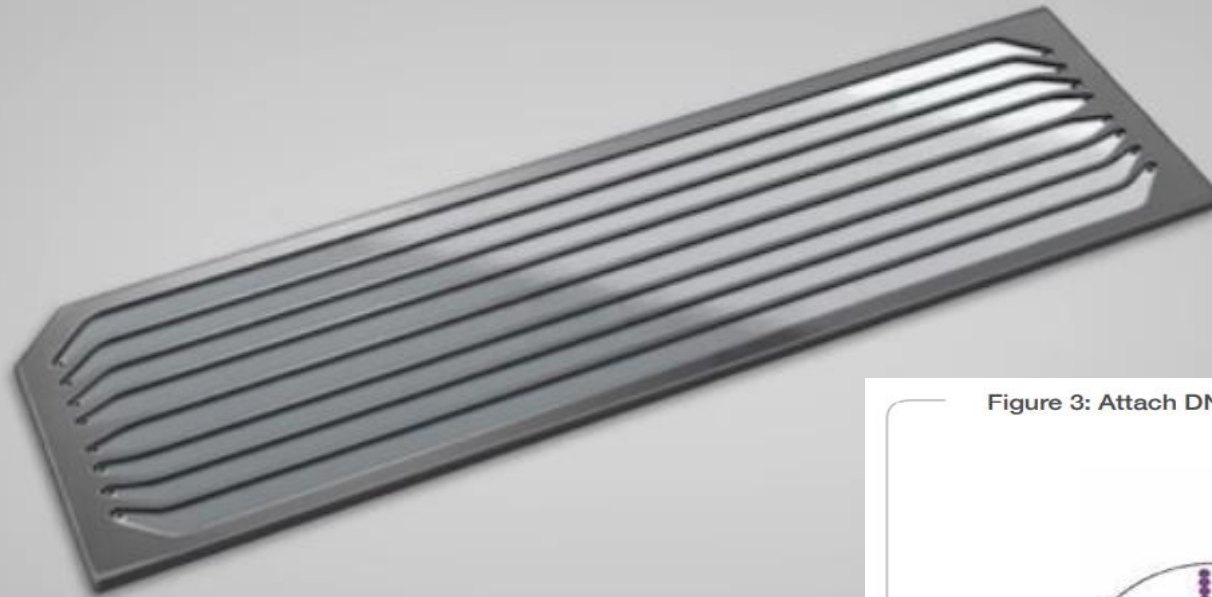
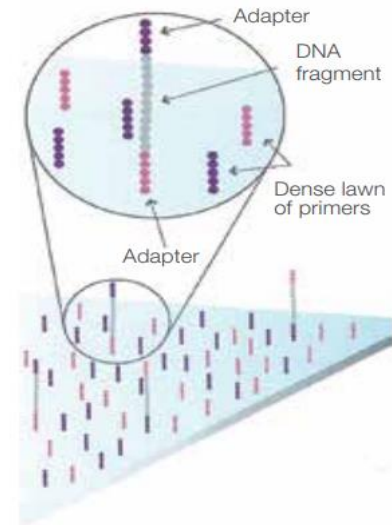
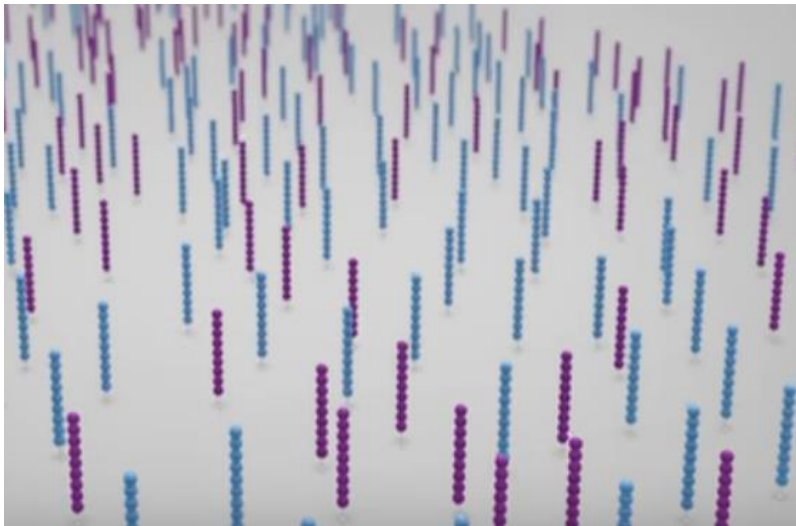


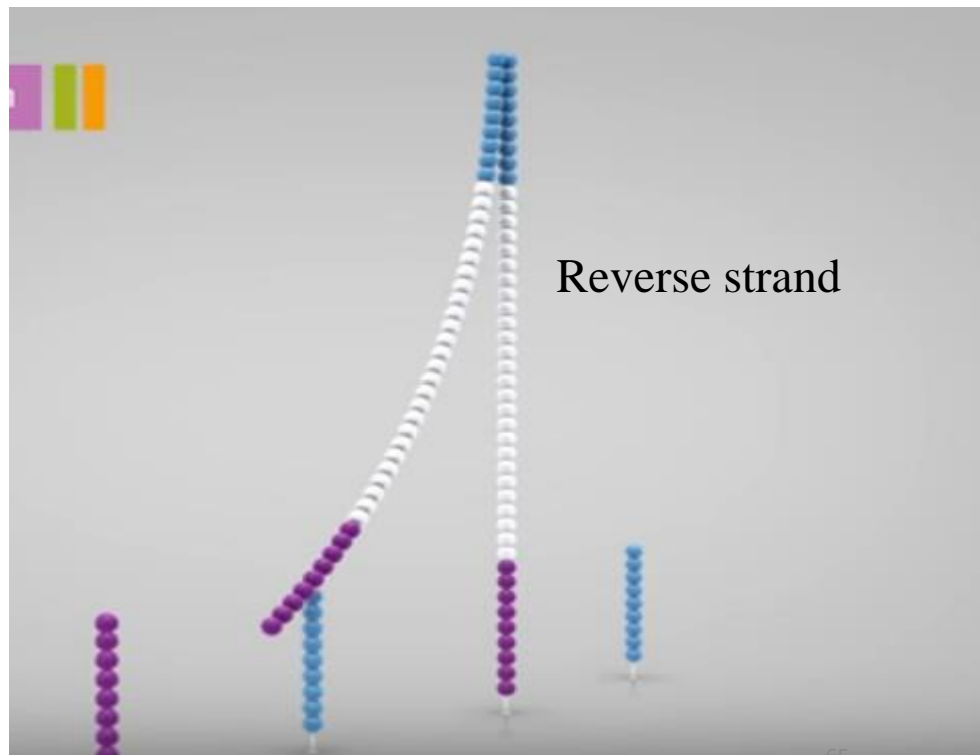
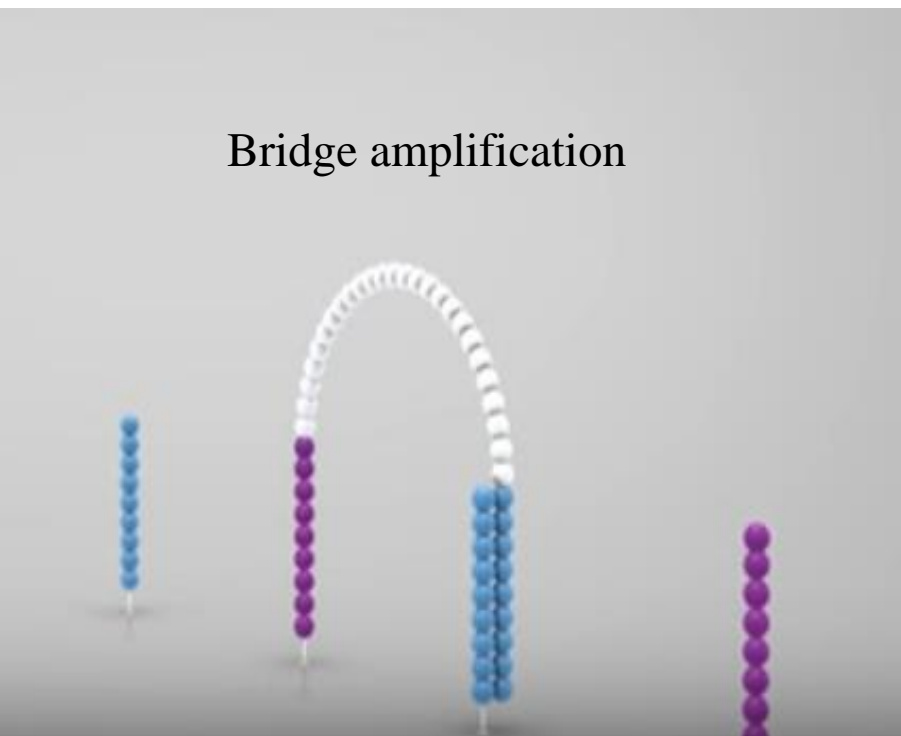
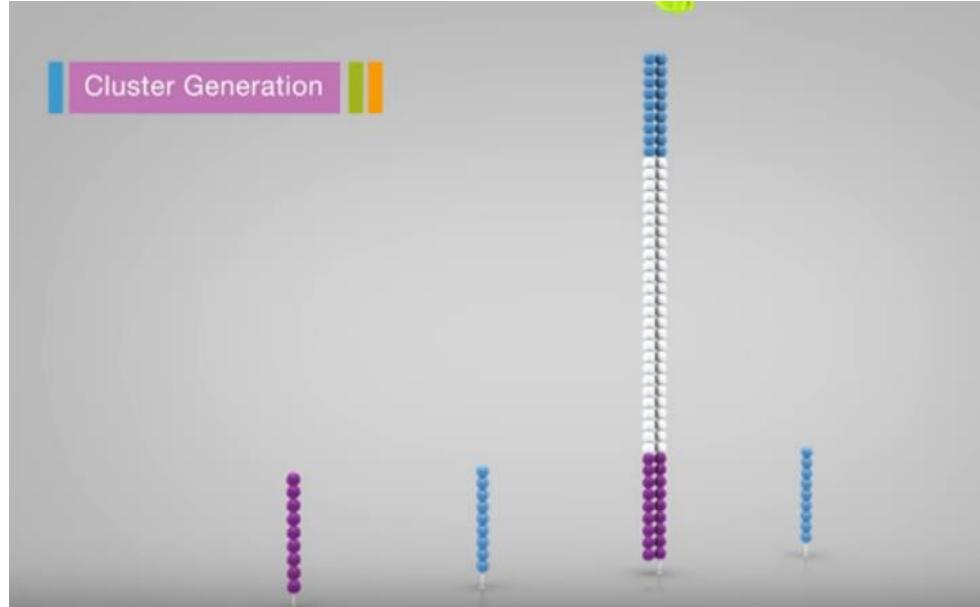
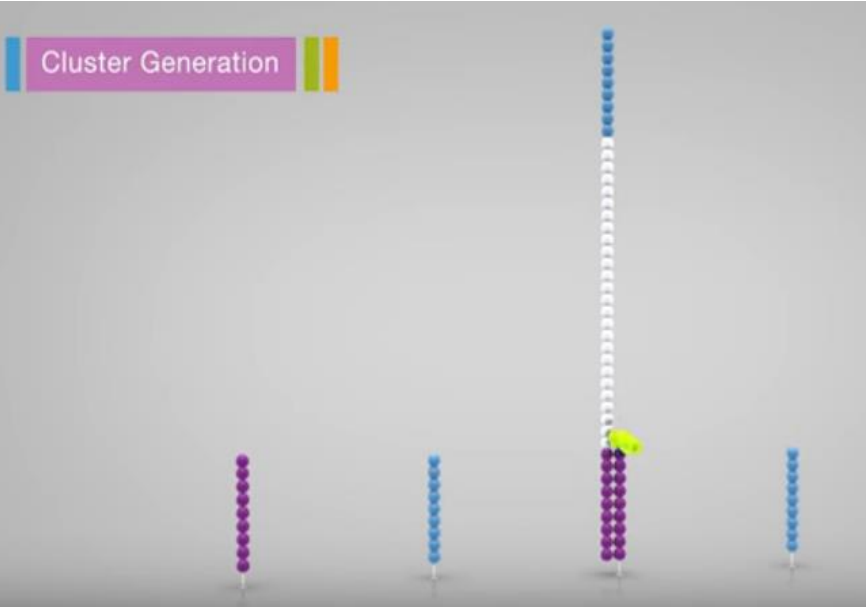
Figure 3: Attach DNA to Surface



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

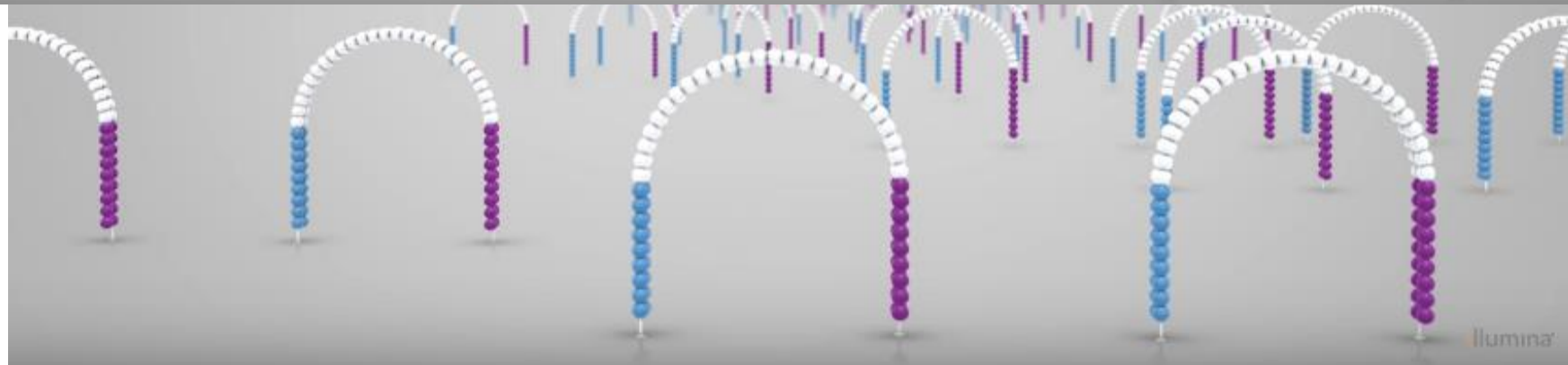
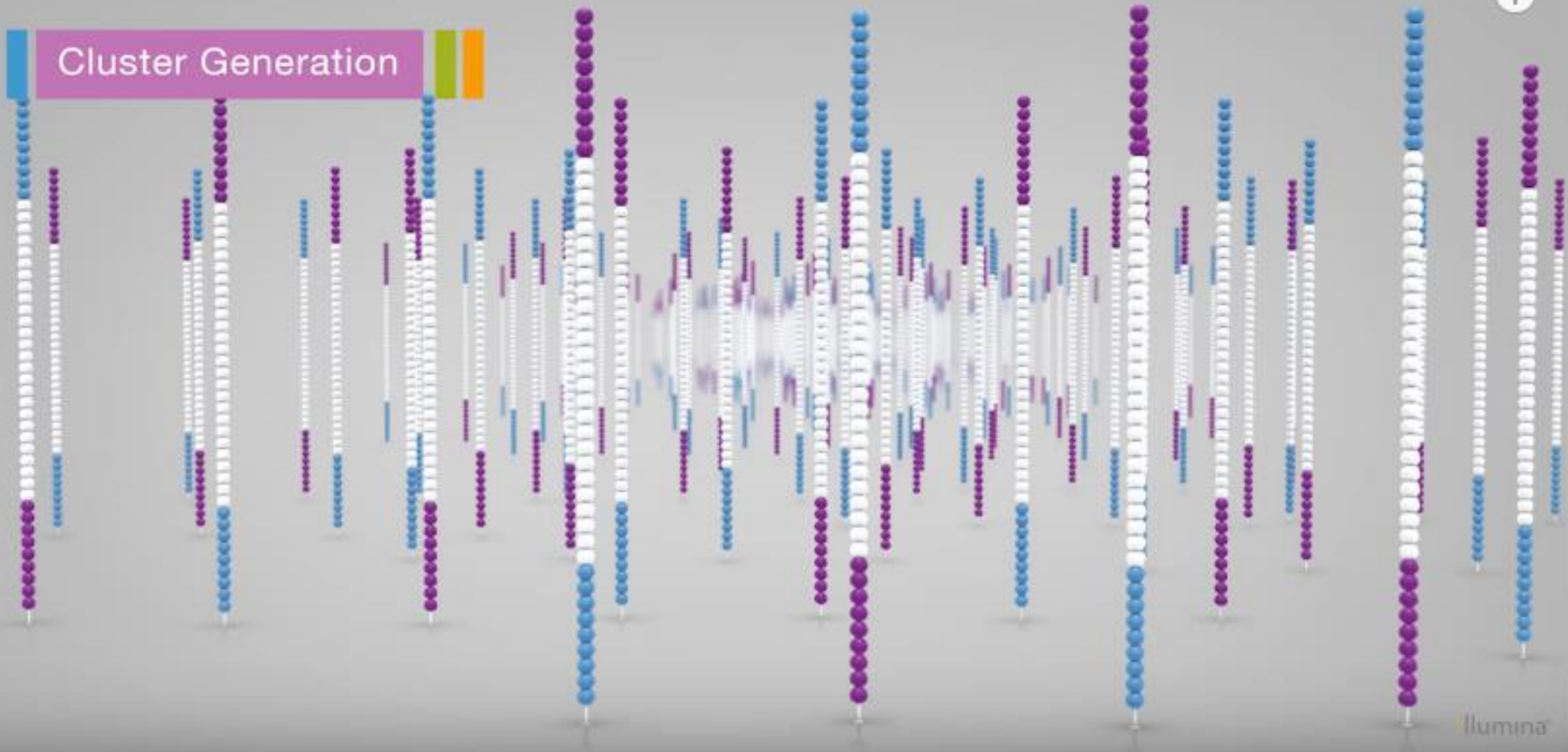


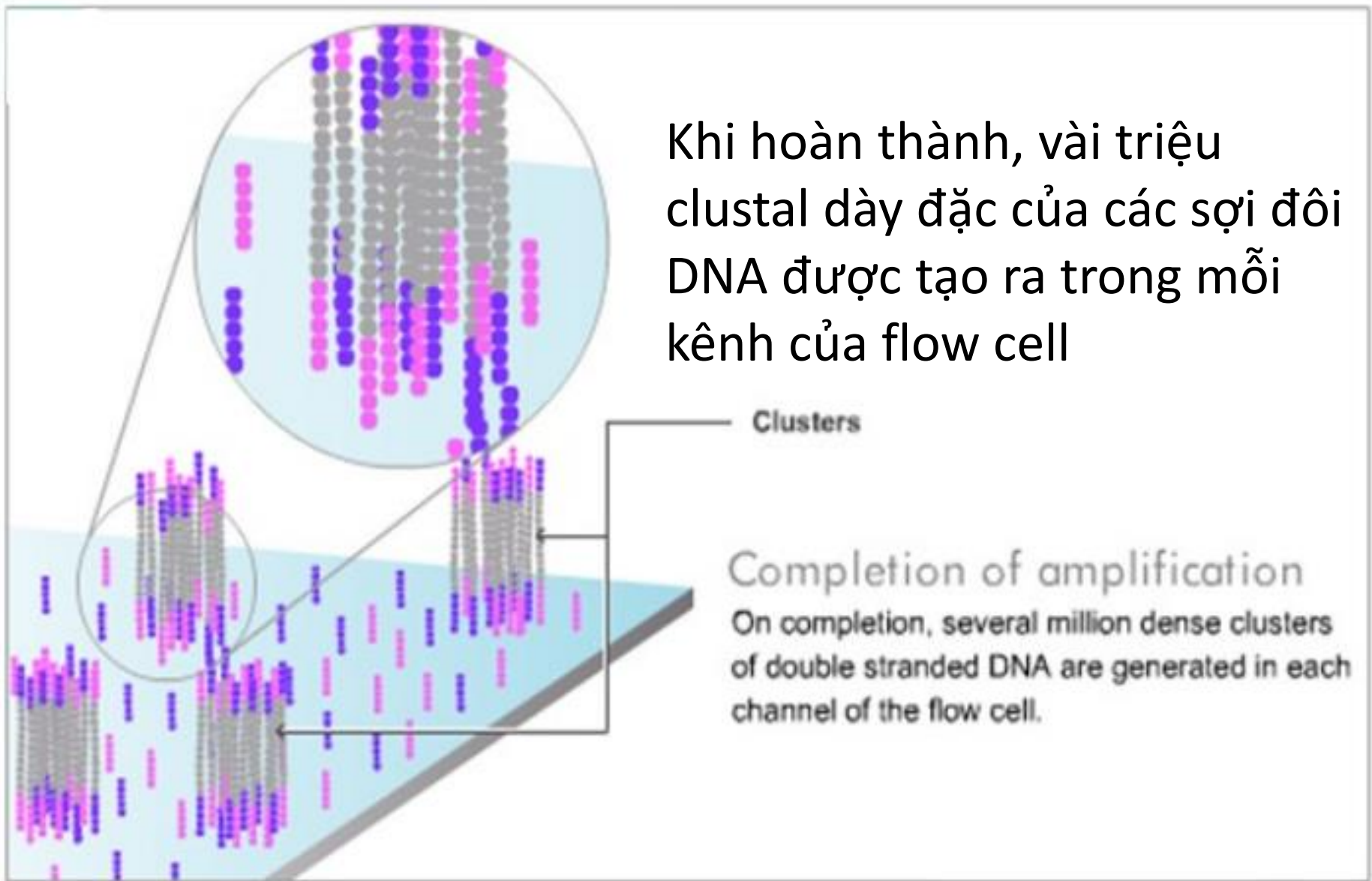


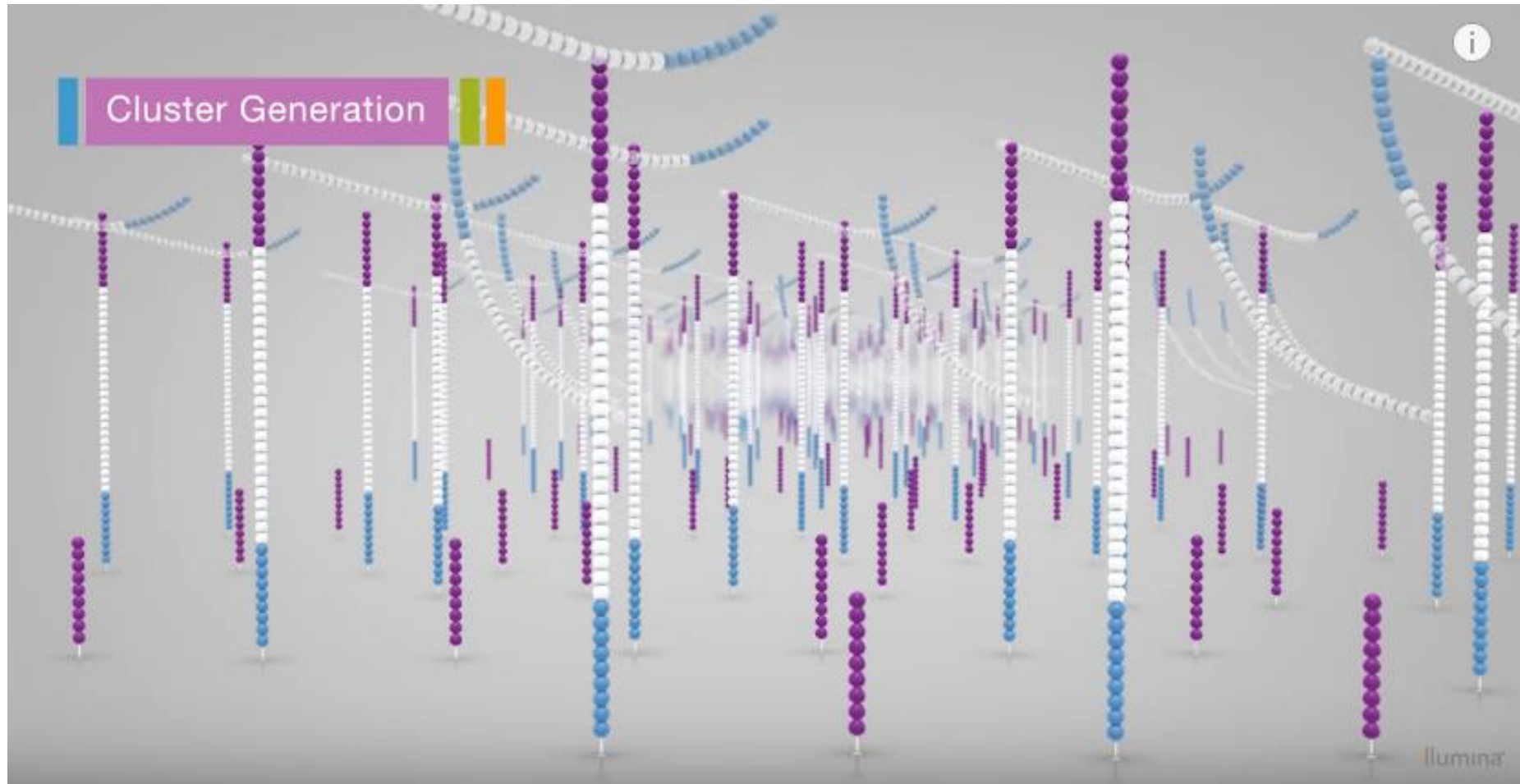


Denatured

## Cluster Generation





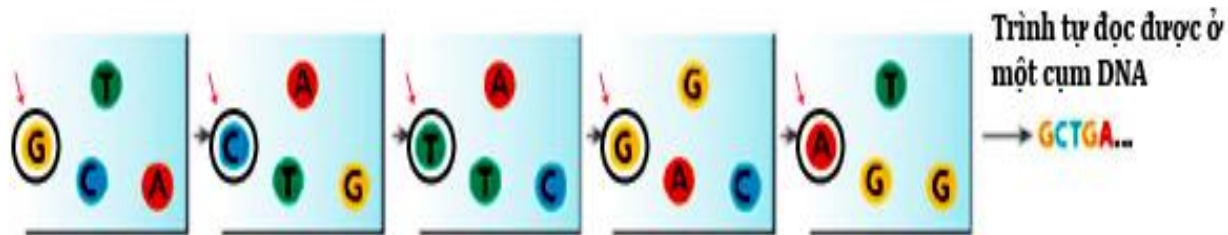
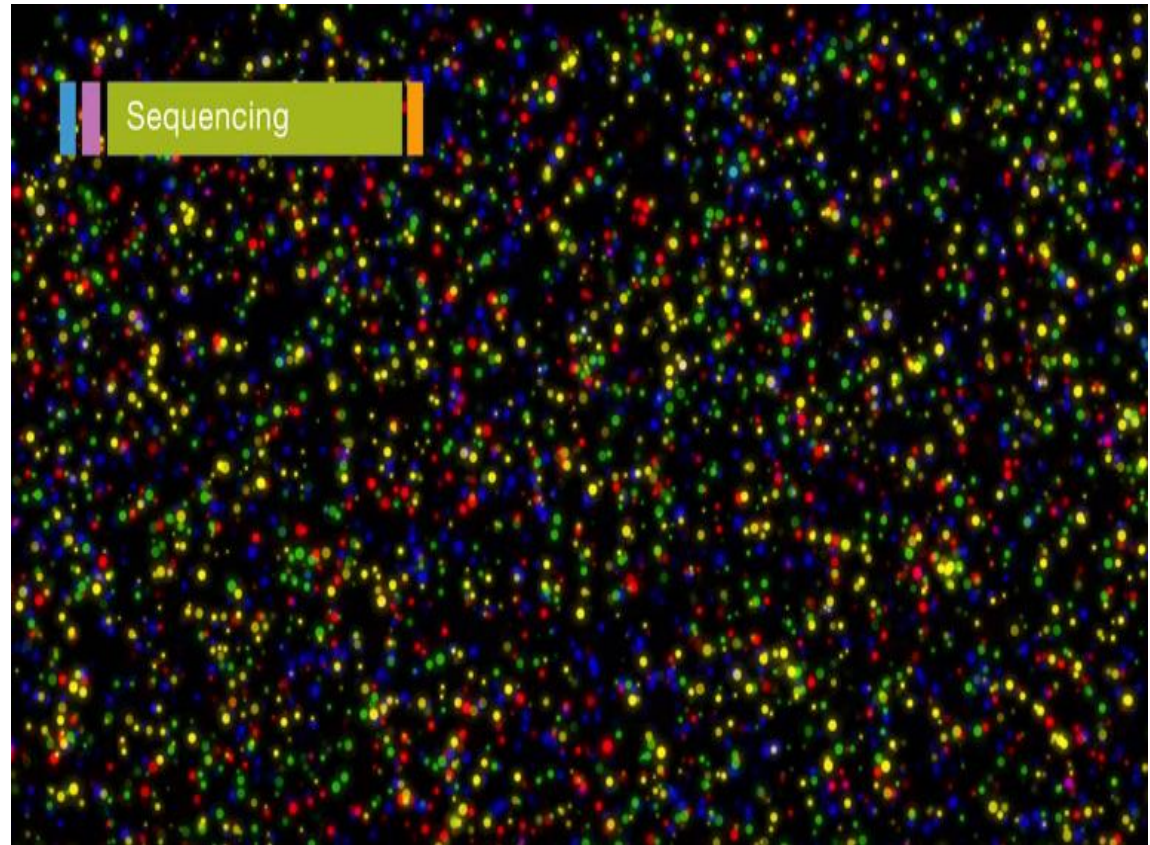


Reverse strands are cleaved and washed out, leaving only the forward strands



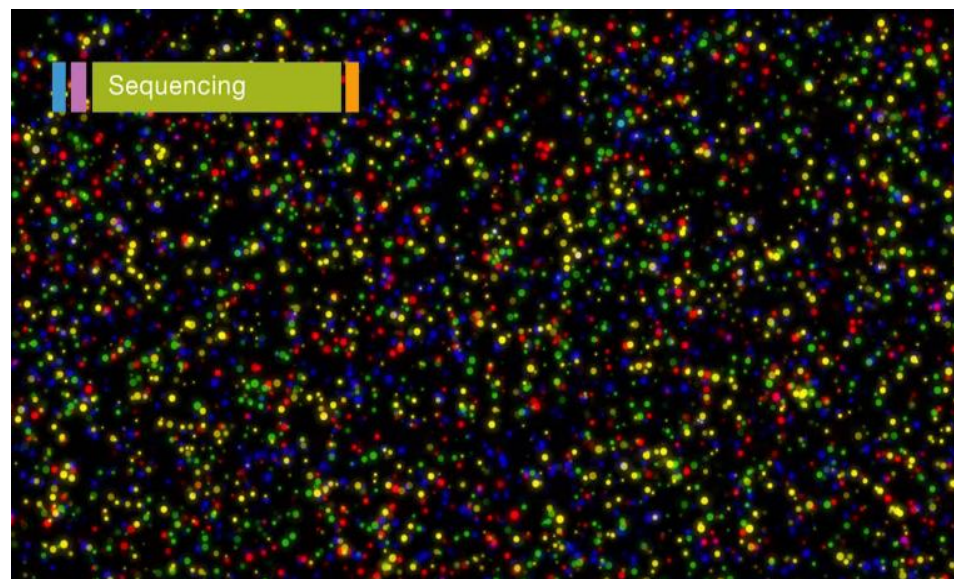
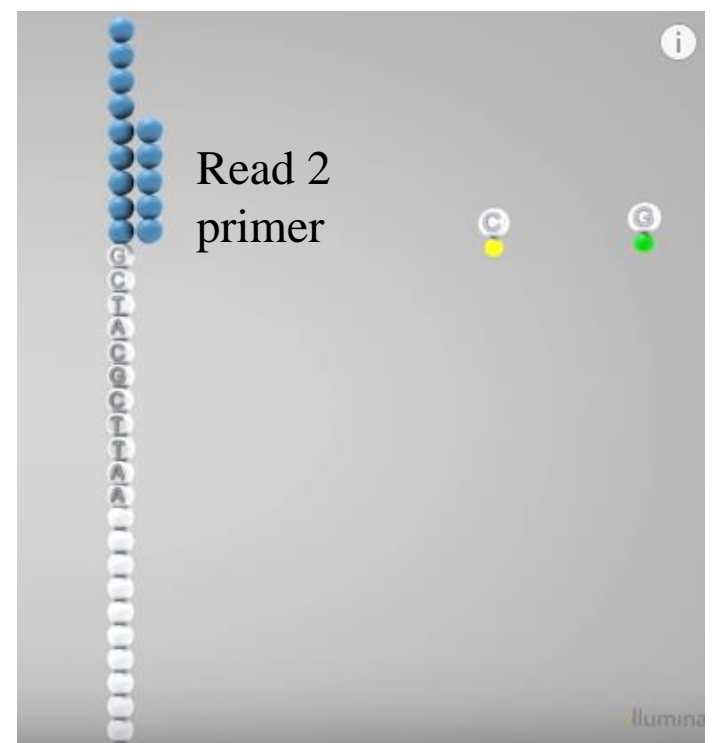
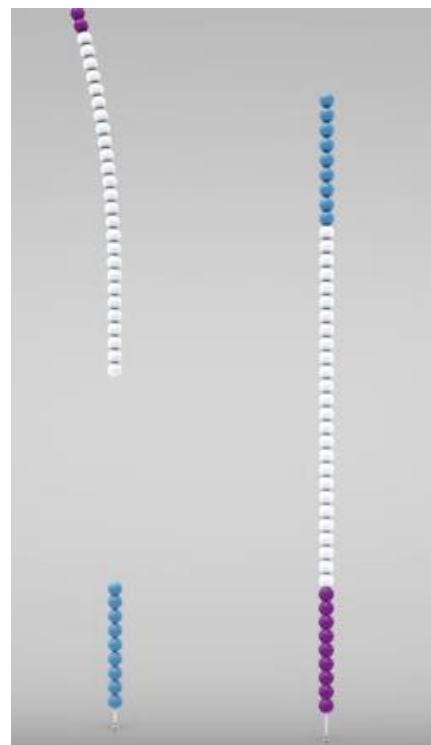
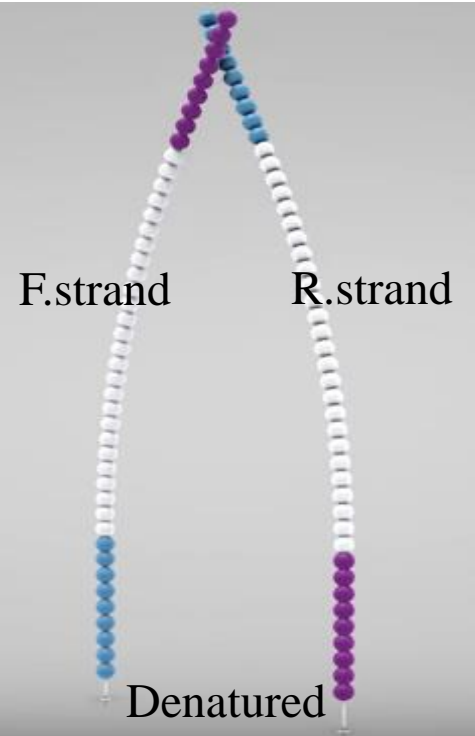
# Sequencing

Sequencing  
binding site



Trình tự môi đầu tiên được cung cấp cho lần đọc đầu tiên







## Local sequencing clustering

```
ACTACCTCAGCAGTAGTAAGAAACAAAAGCAATTGACAAACCTCCTTCTTATTCTTAGA
ACTACCTCAGCAGTAGTAAGAAACAAAAGCAATTGACAAACCTCCTTCTTATTCTTAGA
ACTACCTCAGCAGTAGTAAGAAACAAAAGCAATTGACAAACCTCCTTCTTATTCTTAGA
ACTACCTCAGCAGTAGTAAGAAACAAAAGCAATTGACAAACCTCCTTCTTATTCTTAGA
```

```
CTTACGCCGTACTACCTCAGCAGTAGTAAGAAACAAAAGCAATTGACAAACCTCCTTCT
CTTACGCCGTACTACCTCAGCAGTAGTAAGAAACAAAAGCAATTGACAAACCTCCTTCT
CTTACGCCGTACTACCTCAGCAGTAGTAAGAAACAAAAGCAATTGACAAACCTCCTTCT
CTTACGCCGTACTACCTCAGCAGTAGTAAGAAACAAAAGCAATTGACAAACCTCCTTCT
```

```
AATGGCTAAGGCTTACGCCGTACTACCTCAGCAGTAGTAAGAAACAAAAGCAATTGAC
AATGGCTAAGGCTTACGCCGTACTACCTCAGCAGTAGTAAGAAACAAAAGCAATTGAC
AATGGCTAAGGCTTACGCCGTACTACCTCAGCAGTAGTAAGAAACAAAAGCAATTGAC
AATGGCTAAGGCTTACGCCGTACTACCTCAGCAGTAGTAAGAAACAAAAGCAATTGAC
```

### Similar sequences

```
CTCAGCAGTAGTAAGAAACAAAAGCAATTGACAAACCTCCTTCTTATTCTTAGAAACAA
CTCAGCAGTAGTAAGAAACAAAAGCAATTGACAAACCTCCTTCTTATTCTTAGAAACAA
CTCAGCAGTAGTAAGAAACAAAAGCAATTGACAAACCTCCTTCTTATTCTTAGAAACAA
CTCAGCAGTAGTAAGAAACAAAAGCAATTGACAAACCTCCTTCTTATTCTTAGAAACAA
```

```
CTCAATGGCTAAGGCTTACGCCGTACTACCTCAGCAGTAGTAAGAAACAAAAGCAATT
CTCAATGGCTAAGGCTTACGCCGTACTACCTCAGCAGTAGTAAGAAACAAAAGCAATT
CTCAATGGCTAAGGCTTACGCCGTACTACCTCAGCAGTAGTAAGAAACAAAAGCAATT
CTCAATGGCTAAGGCTTACGCCGTACTACCTCAGCAGTAGTAAGAAACAAAAGCAATT
```

## Creat contiguous sequencing (Tạo ra các tập con trình tự)

### Forward read

```
CTCAGCAGTAGTAAGAAACAAAAGCAATTGACAAACCTCCTTCTTATTCTTAGAAACAA
CTCAGCAGTAGTAAGAAACAAAAGCAATTGACAAACCTCCTTCTTATTCTTAGAAACAA
CTCAGCAGTAGTAAGAAACAAAAGCAATTGACAAACCTCCTTCTTATTCTTAGAAACAA
CTCAGCAGTAGTAAGAAACAAAAGCAATTGACAAACCTCCTTCTTATTCTTAGAAACAA
```

### Reverse read

```
ACTACCTCAGCAGTAGTAAGAAACAAAAGCAATTGACAAACCTCCTTCTTATTCTTAGA
ACTACCTCAGCAGTAGTAAGAAACAAAAGCAATTGACAAACCTCCTTCTTATTCTTAGA
ACTACCTCAGCAGTAGTAAGAAACAAAAGCAATTGACAAACCTCCTTCTTATTCTTAGA
ACTACCTCAGCAGTAGTAAGAAACAAAAGCAATTGACAAACCTCCTTCTTATTCTTAGA
```

```
CTTACGCCGTACTACCTCAGCAGTAGTAAGAAACAAAAGCAATTGACAAACCTCCTTCT
CTTACGCCGTACTACCTCAGCAGTAGTAAGAAACAAAAGCAATTGACAAACCTCCTTCT
CTTACGCCGTACTACCTCAGCAGTAGTAAGAAACAAAAGCAATTGACAAACCTCCTTCT
CTTACGCCGTACTACCTCAGCAGTAGTAAGAAACAAAAGCAATTGACAAACCTCCTTCT
```

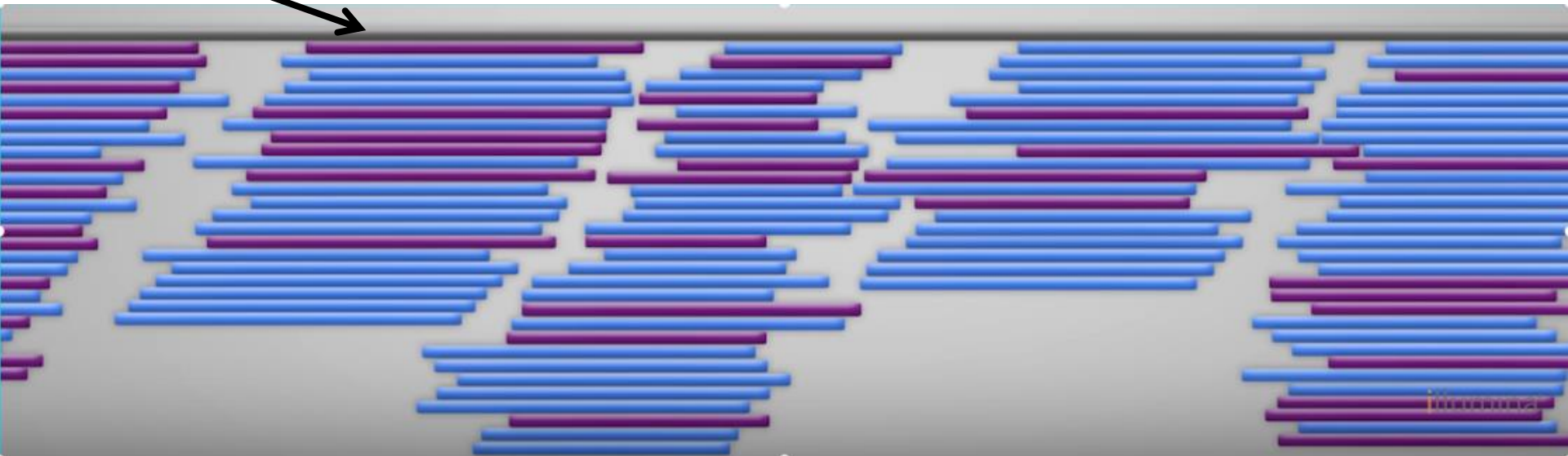
```
AATGGCTAAGGCTTACGCCGTACTACCTCAGCAGTAGTAAGAAACAAAAGCAATTGAC
AATGGCTAAGGCTTACGCCGTACTACCTCAGCAGTAGTAAGAAACAAAAGCAATTGAC
AATGGCTAAGGCTTACGCCGTACTACCTCAGCAGTAGTAAGAAACAAAAGCAATTGAC
AATGGCTAAGGCTTACGCCGTACTACCTCAGCAGTAGTAAGAAACAAAAGCAATTGAC
```

```
CTCAATGGCTAAGGCTTACGCCGTACTACCTCAGCAGTAGTAAGAAACAAAAGCAATT
CTCAATGGCTAAGGCTTACGCCGTACTACCTCAGCAGTAGTAAGAAACAAAAGCAATT
CTCAATGGCTAAGGCTTACGCCGTACTACCTCAGCAGTAGTAAGAAACAAAAGCAATT
CTCAATGGCTAAGGCTTACGCCGTACTACCTCAGCAGTAGTAAGAAACAAAAGCAATT
```



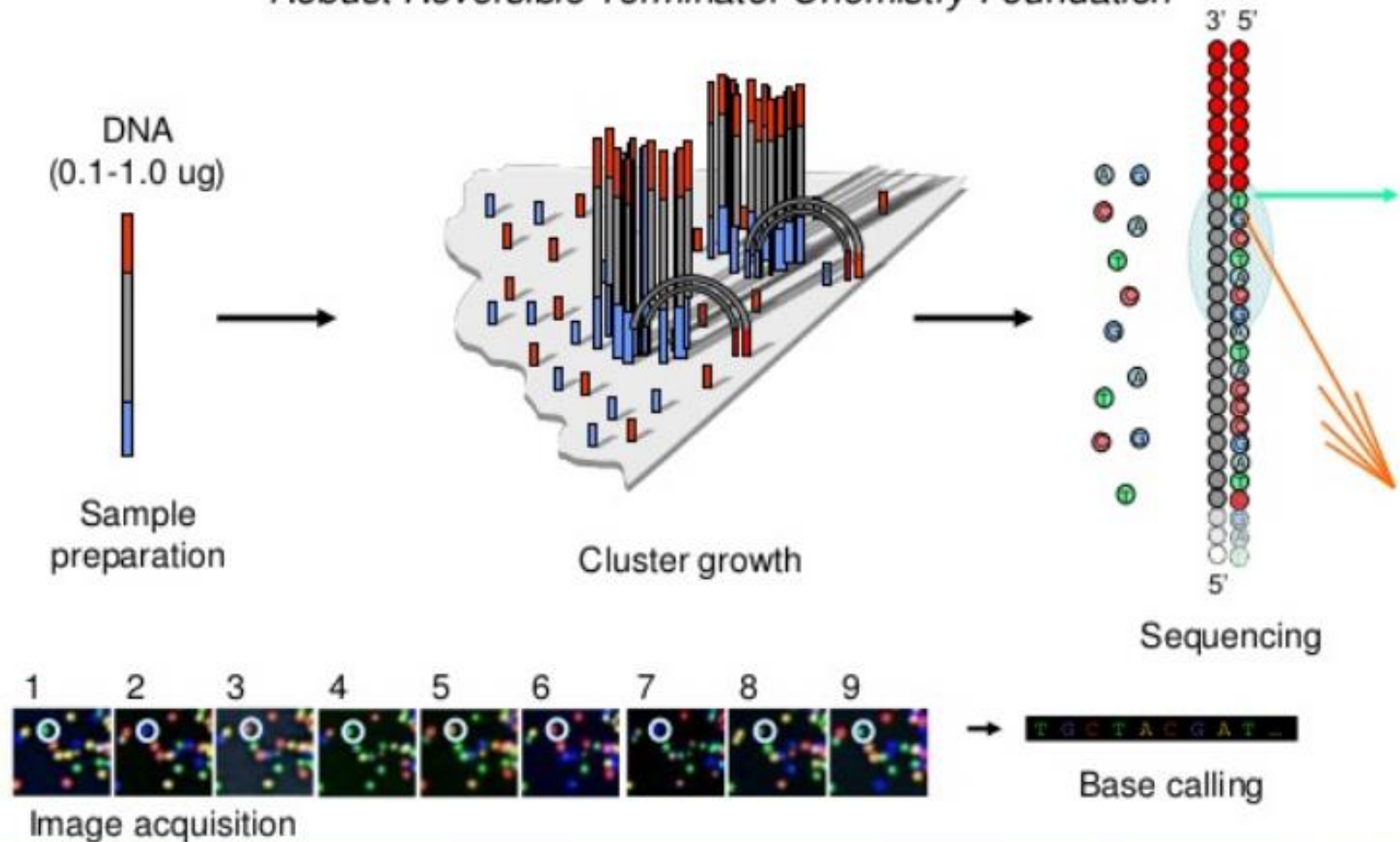
## Variant identification

Reference



# Illumina Sequencing Technology

*Robust Reversible Terminator Chemistry Foundation*



# Ưu điểm

- 4 loại nucleotide được gắn huỳnh quang khác nhau nên có thể phân tích được các trình tự của 1 loại nucleotide lặp lại liên tục
- Thư viện DNA có thể được mã hóa và tách riêng trong suốt quá trình phân tích kết quả
- Cộng đồng các nhà khoa học sử dụng hệ thống này phổ biến hơn Torrent (Ion) sequencing

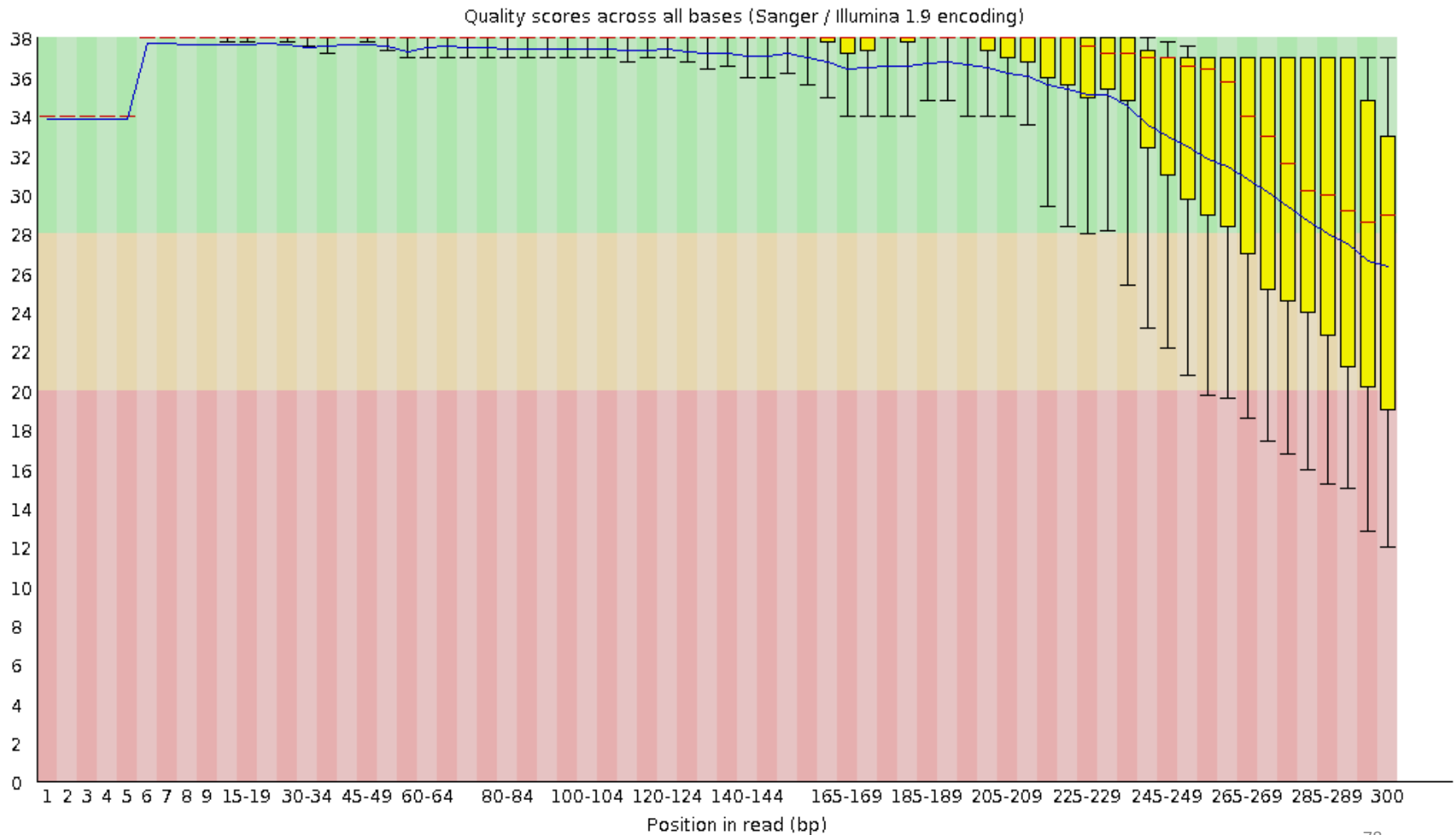
# Nhược điểm

- Giá thành hóa chất vẫn cao so với Sanger và Ion Torrent
- Chiều dài đọc chỉ khoảng 200-250 (độ chính xác 99% tại base thứ 250 và cao hơn ở các base phía trước)
- Thời gian lâu: 23 giờ để phân tích hệ gen người

# Why does the per base sequence quality decrease over the read in Illumina?

- The first indicator for the quality of your sequencing data is the per base sequence quality of your raw reads.
- Often you will see a decreasing quality with increasing base position just as in the FASTQC image below (Fig. 1).
- But what is the reason for this and what are the consequences?

Figure 1: Per base sequence quality control with typical decrease of the quality over the read.



# The normal sequencing-by-synthesis process in Illumina

- First of all, don't panic! It is a normal and well known phenomenon.
- The reason of the decreasing sequence quality lies in the sequencing technology of Illumina.
- Illumina relies on the sequencing by synthesis procedure.
- During each cycle of the process the sequencer washes chemicals that include variants for all four nucleotide over the flow cell (which has different clusters with identical DNA fragments for each cluster).

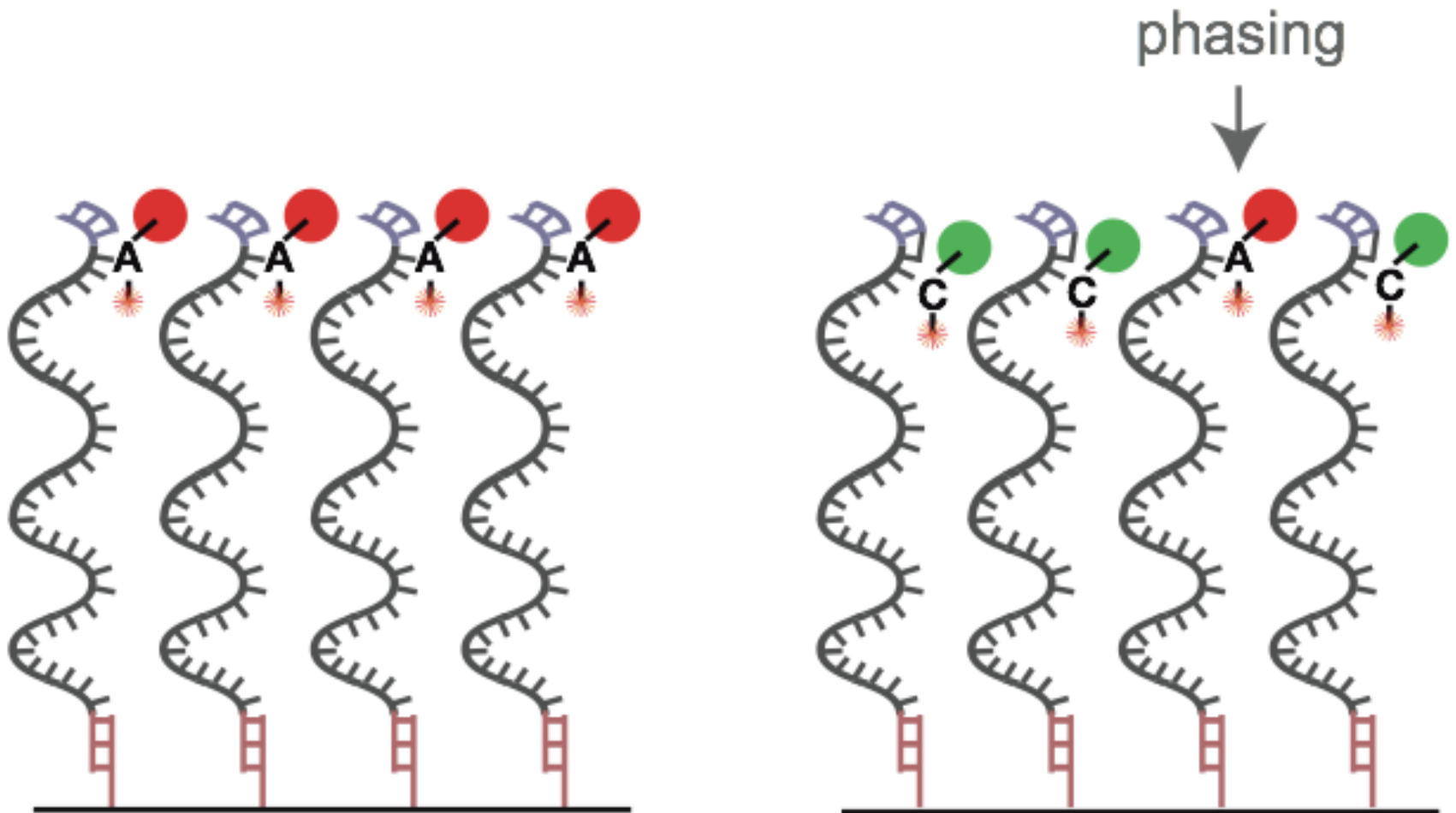
- The nucleotides have a blocker (terminator cap) so that only 1 base gets added to each molecule of DNA at a time.
- After the detection of the coupled fluorescence signal the blocker can be removed and the cycle can start again.
- This way, the DNA fragments in each cluster get sequenced synchronously by expressing specific fluorescence signals.



# Nobody is perfect

- During the sequencing process different errors can occur. The main reason for the decreasing sequence quality is the so-called phasing.
- Phasing means that the blocker of a nucleotide is not correctly removed after signal detection.
- In the next cycle no new nucleotide can bind on this DNA fragment and the old nucleotide is detected one more time whereby the fluorescence signal of this old nucleotide (probably) differs from the synchronous signal of the other nucleotides (Fig. 2).

Figure 2: Phasing



- From now on this DNA fragment will be 1 cycle behind the rest (out of phase), polluting the light signal that the sequencer's camera has to read.
- A similar effect occurs if a nucleotide has a defect terminator cap (prephasing).
- In this case two nucleotides can bind in one cycle whereby the fragment will be 1 cycle before the rest.

- These errors occur with a low probability.
- But over time (with increasing read length) they add up and pollute the light signal more and more.
- The signal gets more and more asynchronous.
- And since the light signal is used to calculate quality scores the asynchronous signal results in a decreasing sequence quality score.

# Summary

- As we now know the decreasing base sequence quality is due to a unwanted but unavoidable process.
- It limits the length of high quality reads.
- New chemicals are largely intended to minimize the phasing problem, increasing the length of reads before quality begins to decrease.

# So sánh 3 đại diện của 3 hệ thống giải trình tự thế hệ 2

Next-Gen Sequencer	Machine Cost	Cost per run	Minimum Throughput	Sequencing Run Time	Cost Per Mb
Illumina MiSeq	\$125,000	\$750	1500 Mb (2 x 150 Bases)	27 Hours	\$0.5
454 GS Junior	\$108,000	\$1,100	35 Mb (400 Bases)	8 Hours	\$31
Ion Torrent PGM - 318 Chip	\$80,490	\$625	1000Mb	3 Hours	\$0.6

# 3. Giải trình tự thể hệ 3

## Single molecular Real Time Sequencing (SMRT)

- Nguyên tắc:
  - Phản ứng giải trình tự gen tức thời đơn phân tử được thực hiện trong một lỗ kích thước nano với tính chất đặc biệt:
  - cho phép quan sát tín hiệu huỳnh quang đơn phân tử khi chất huỳnh quang tiến sát vào đáy của lỗ nano (Zero-Mode Waveguide, ZMW).

# Quy trình

- Genome được phân cắt bởi enzym giới hạn thành các đoạn cỡ 3000- 15000 bp.
- Các đoạn DNA được gắn với adaptor để primer có thể bắt cặp.
- DNA được ủ trong các lỗ nano đã chứa DNA polymerase cố định trong đáy lỗ.



- Mỗi DNA sẽ được giải trình tự theo cơ chế tổng hợp DNA với 4 loại dNTP gắn 4 loại dye huỳnh quang khác nhau ở gốc phosphate.
- Tại một thời điểm, chỉ có dNTP nào bổ sung với trình tự quan tâm mới đến đủ gần đáy lỗ và ngay trước khi được gắn vào chuỗi, tín hiệu huỳnh quang mà dNTP mang được ghi nhận trong một khoảnh khắc bởi camera cực nhạy, ngay lập tức sau đó fluorescent bị cắt và trôi nổi trong môi trường.
- Toàn bộ trình tự được giải của các mảnh DNA sẽ được ghép nối và phân tích số liệu bằng phần mềm tin sinh học.

# Thông số của hệ thống máy giải trình tự nguyên lý SMRT của Pacific Bioscience RS

	Pacbio RS
Read lenght	3.000 – 15.000 bp

---

Throughput	1 GB
Read per run	70.000
Accuracy	95%
Run time	30 minutes

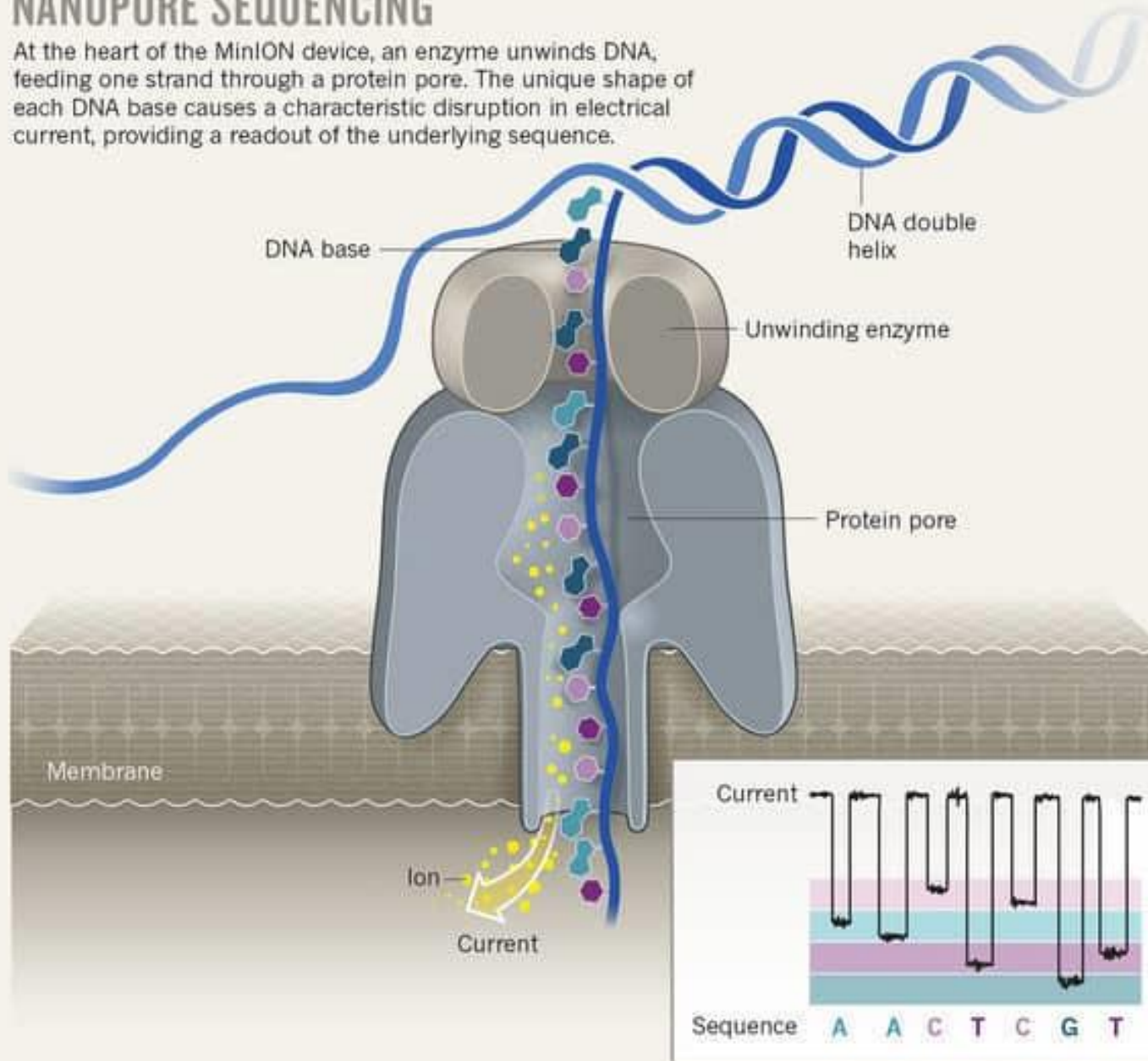
# Giải trình tự thể hệ thứ 4 – Oxford Nanopore DNA Sequencing Technology

- Giải trình tự tức thời dựa trên tín hiệu điện phân tử



# NANOPORE SEQUENCING

At the heart of the MinION device, an enzyme unwinds DNA, feeding one strand through a protein pore. The unique shape of each DNA base causes a characteristic disruption in electrical current, providing a readout of the underlying sequence.



- Phương pháp này sử dụng dòng điện để vận chuyển một mẫu chưa biết qua một lỗ đường kính chỉ 1 nm.
- Một protein nanopore được gắn lên một màng polymer không dẫn điện.

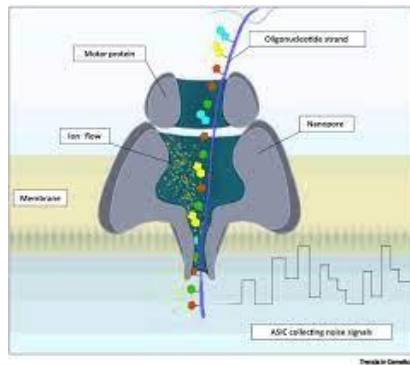
- Hệ thống này luôn chứa một dung dịch điện ly – khi mà đặt vào một điện trường không đổi, một dòng điện có thể xuất hiện trong hệ thống.
- Độ lớn của cường độ dòng điện qua bề mặt lỗ nano phụ thuộc vào kích thước của lỗ nano cũng như thành phần của DNA/RNA đang mắc tại lỗ nano đó.

- Khi đủ gần với lỗ, các mẫu làm thay đổi đặc trưng về cường độ dòng điện qua bề mặt lỗ.
- Tổng điện tích qua lỗ bằng tích phân mặt của cường độ dòng điện chảy qua bề mặt lỗ giữa khoảng thời gian  $t_1$  và  $t_2$ .

- Hai loại lỗ nano đang được quan tâm là Alpha hemolysin ( $\alpha$ HL) từ vi khuẩn và Mycobacterium smegmatis porin A (MspA) để ứng dụng cho giải trình tự.



- Hiện nay có 3 thiết bị giải trình tự ứng dụng nguyên tắc này.
- MinION flow cell có kích cỡ bỏ túi, chứa tới 512 kênh nanopore.
- GridION sử dụng 5 MinION đồng thời.
- PromethION chứa 48 flow cell có thể hoạt động riêng rẽ, mỗi cell có đến 3000 lỗ nano.



Thanks to Genomics



## 4. Ứng dụng giải trình tự gen thế hệ mới

### Ứng dụng NGS trong xét nghiệm di truyền

- Theo thống kê, khoảng 5-10% các ca ung thư là do di truyền.
- Quá trình phát sinh và tích lũy những đột biến gen diễn ra qua nhiều thế hệ trong gia đình đã làm gia tăng nguy cơ mắc ung thư.
- Các kỹ thuật trong xét nghiệm di truyền đã được phát triển nhằm phát hiện các đột biến gen gây ra bệnh di truyền và các dạng ung thư.
- Qua đó, các bác sĩ sẽ có thêm thông tin trong quá trình tư vấn di truyền, chẩn đoán chính xác nguyên nhân gây bệnh và lựa chọn phác đồ điều trị phù hợp.

...AGAGGAGGGAAATCAATATAT  
...AAGGAAGCTTTAGCGAT  
...AAGGGMUTATIONCGG  
...AAGTGAATTGAACCATTA  
...TTCTTGGGTTCTTGGGAGCAGCA  
...GGTACAGGCCAGACAATTATTGTCT  
...ATTGAGGCGCAACAGCATCTGTT  
...AAGAA...

- Xét nghiệm di truyền đã được cấp phép thực hiện tại Mỹ và các nước châu Âu từ nhiều năm qua.
- Các kỹ thuật xét nghiệm cơ bản như hóa mô miễn dịch, FISH, realtime PCR hay giải trình tự bằng phương pháp Sanger đã giúp phát hiện đột biến ở một số gen nhất định và cho thấy những đột biến đó có liên quan tới ung thư và bệnh di truyền.

- Tuy nhiên, trên thực tế sẽ không có cơ chế tương tác 1:1 đối với đột biến gen và khả năng gây bệnh,
- mà thường là có nhiều gen liên quan đến quá trình phát sinh bệnh và nhiều gen lại không xuất hiện những đột biến đặc hiệu.

- Vì thế, sự phát triển của giải trình tự gen thế hệ mới (NGS) đã giúp các nhà khoa học có được bức tranh tổng thể và chi tiết hơn khi tiến hành các xét nghiệm gen di truyền,
- Qua đó có khả năng phát hiện những gen chưa biết nhưng có liên quan tới quá trình khởi phát bệnh và cả những dạng hiếm gặp đặc biệt của đột biến gen xảy ra trong thực tế.

# Bảng thống kê các gen liên quan đến bệnh ung thư

Loại ung thư	Số lượng gen	Gen
Ung thư vú cơ bản (Breast Cancer)	6	<i>BRCA1, BRCA2, TP53, PTEN, CDH1, STK11</i>
Ung thư đại trực tràng (Colorectal Cancer)	17	<i>APC, BMPR1A, CDH1, CHEK2, EPCAM, GREM1, MLH1, MSH2, MSH6, MUTYH, PMS2, POLD1, POLE, PTEN, SMAD4, STK11, TP53</i>
Ung thư buồng trứng (Ovarian Cancer)	24	<i>ATM, BARD1, BRCA1, BRCA2, BRIP1, CHD1, CHEK2, EPCAM, MLH1, MRE11A, MSH2, MSH6, MUTYH, NBN, NF1, PMS2, PTEN, RAD50, RAD51C, RAD51D, STK11, TP53, PALB2, SMARCA4</i>
Ung thư phổi (Lung Cancer)	11	<i>EGFR, BRAF, KRAS, NRAS, ALK, ERBB2, MET, ROS1, RET, DDR2, PIK3CA</i>



# Xác định đột biến mới bằng NGS

- Một trong những ứng dụng của NGS đó là kỹ thuật WES – [Whole Exome Sequencing](#) và WGS – [Whole Genome Sequencing](#).
- WES cho phép giải mã toàn bộ trình tự DNA của các gen mã hóa (là các gen quy định cho những protein chức năng cụ thể).

# Phát hiện ADN của tế bào ung thư trong hệ tuần hoàn bằng NGS

- WGS có khả năng giải trình tự tất cả bộ gen của một sinh vật, bao gồm toàn bộ DNA trong tế bào, qua đó có thể phát hiện các đột biến tế bào dòng sinh dục và các đột biến soma mới và hiếm gặp.
- Trong nhiều năm qua, công nghệ NGS đã được ứng dụng thành công trên thực tế để xác định đột biến mới của nhiều dạng ung thư như ung thư tuyến tụy, ung thư bàng quang, ung thư biểu mô tế bào thận, ung thư phổi tế bào nhỏ, ung thư tuyến tiền liệt, ung thư vú, ung thư đại trực tràng...

- Trong máu của các bệnh nhân ung thư thường chứa các tế bào ung thư tuần hoàn (CTC – [Circulating tumor cells](#)) và DNA tự do của các tế bào ung thư tuần hoàn (ctDNA – [cell free tumor DNA](#)).
- Công nghệ hiện nay cho phép dễ dàng thu nhận được CTC và ctDNA thông qua dịch sinh thiết và tiến hành giải trình tự gen trên hệ máy NGS.
- Giải trình tự từ các mẫu sinh thiết lỏng là phương pháp có hiệu quả cao trong việc chọn lọc các chỉ dấu sinh học (biomarker) nhằm tầm soát ung thư ở giai đoạn sớm.

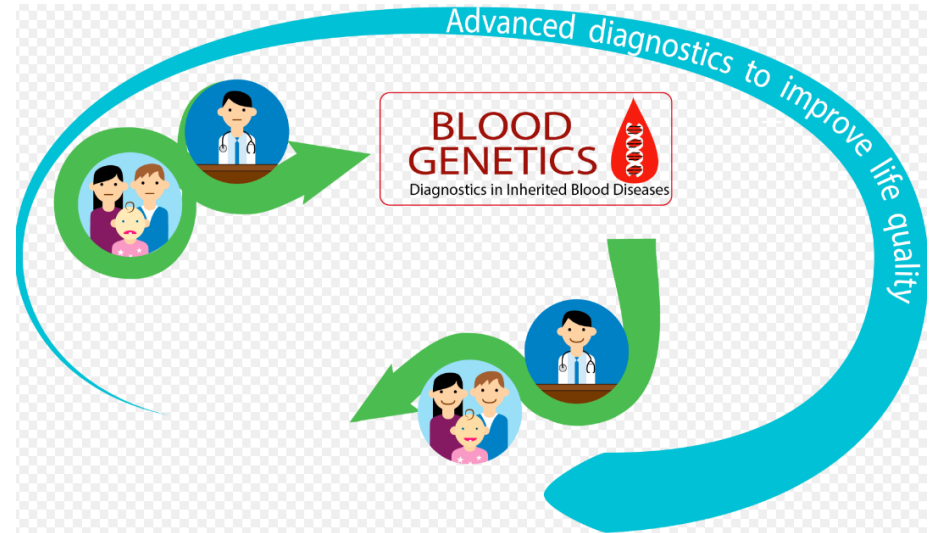
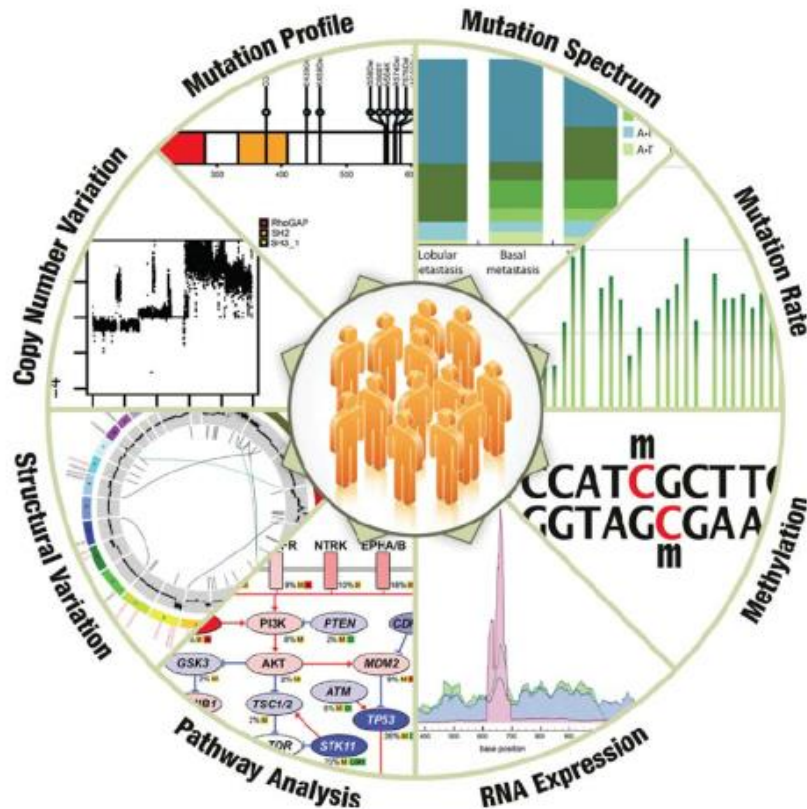
# TỔNG KẾT

- Trong một vài năm tới, chắc chắn sẽ có thêm những người nhập cuộc để tiếp tục tự động hóa lĩnh vực này với các giải pháp giải trình tự mới.

- [GenapSys](#) (đối tác của Sigma-Aldrich), với máy giải trình tự kích cỡ bằng “hộp cơm trưa”;
- [Genia](#) (Roche) với phương pháp giải trình tự nanopore mới;
- và gần đây là [Firefly](#) (Illumina) với công nghệ CMOS đơn kênh, tất cả họ đều tuyên bố sẽ đi đầu về giá thành và tiết kiệm thời gian cho các ứng dụng lâm sàng.

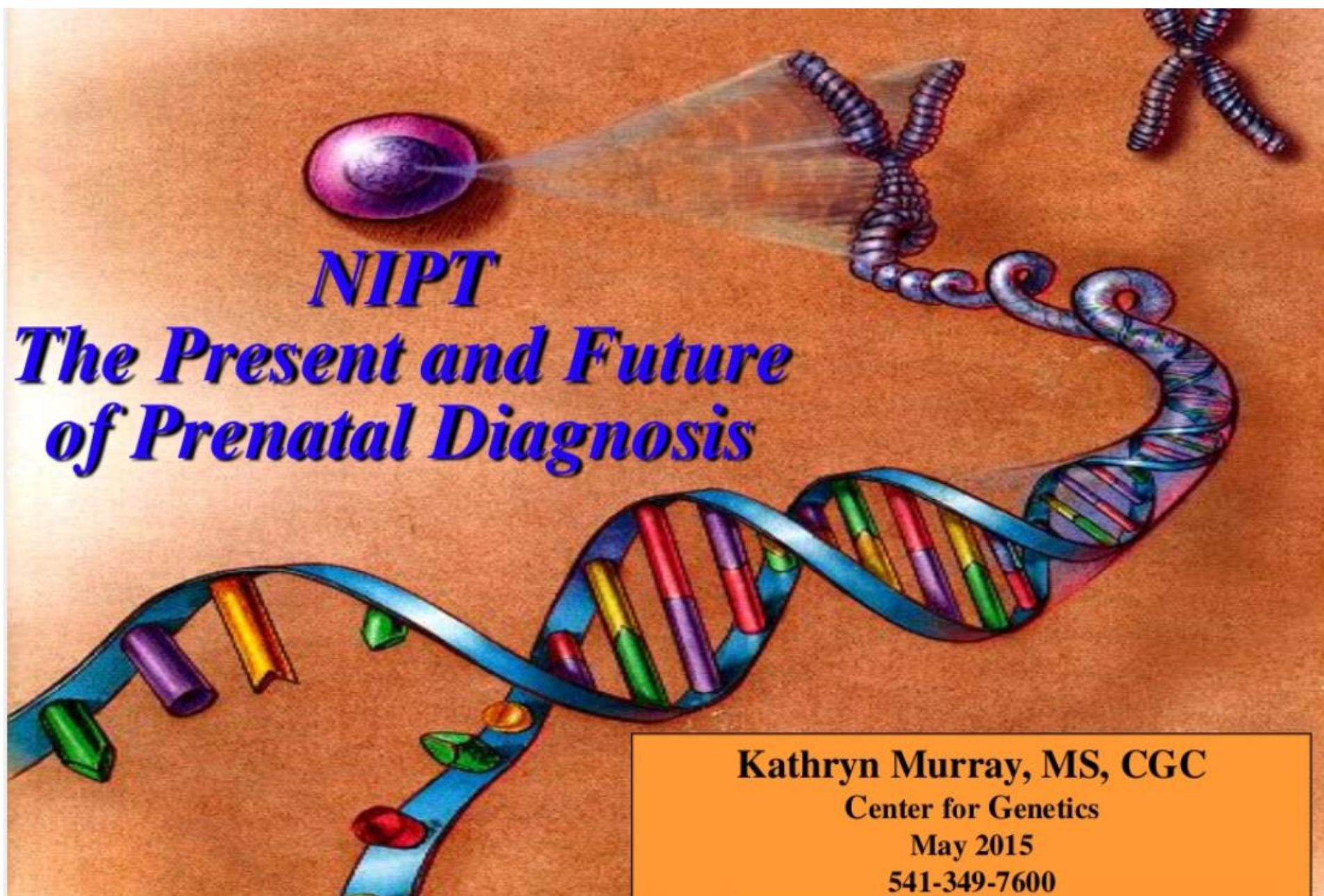
- Cuối cùng, [NanoString Technologies](#) với phương pháp lai không dùng enzyme, GnuBio (Bio-Rad) với phương pháp dựa vào FRET và Electron Optica với một hệ thống dựa vào hiển vi điện tử, đều nhằm cách mạng hóa việc giải trình tự.
- Các kỹ thuật NGS hiện nay và sắp tới mang tiềm năng cho phép cải cách khoa học, bao gồm giải trình tự trực tiếp RNA, protein, theo dõi thời gian thực hệ gen của vật gây bệnh hoặc y học cá thể dựa vào giải trình tự hệ gen của từng người.

# ỨNG DỤNG CỦA NGS TRONG CHUẨN ĐOÁN ĐỘT BIẾN



- Dữ liệu NGS sẽ được tạo ra cho hàng trăm khối u từ tất cả các loại ung thư chính trong tương lai gần.
- Việc phân tích tích hợp dữ liệu trình tự DNA, RNA và methylation sẽ giúp làm sáng tỏ tất cả các thay đổi di truyền liên quan đến ung thư.





**NIPT**

# ***The Present and Future of Prenatal Diagnosis***

**Kathryn Murray, MS, CGC**

**Center for Genetics**

**May 2015**

**541-349-7600**



# Non-invasive prenatal tests (NIPT)



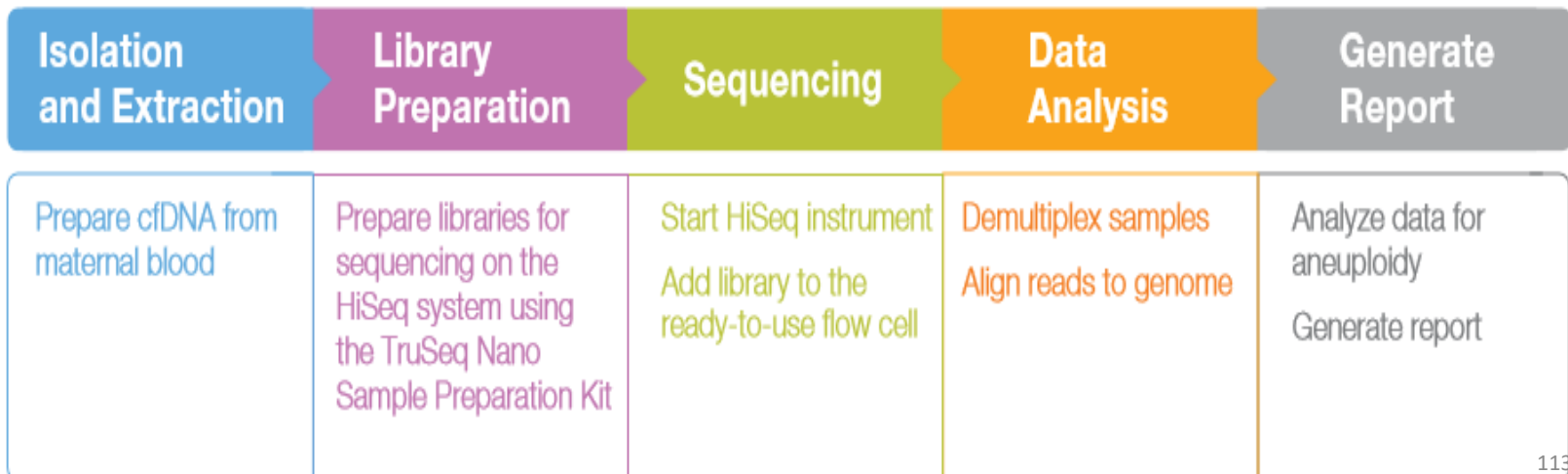
## TẠI SAO PHẢI XÉT NGHIỆM NIPT – Illumina?

Cứ 100 trẻ em sinh ra thì có 4 trẻ mang vấn đề nghiêm trọng về sức khỏe, phần lớn liên quan đến trường hợp bất thường nhiễm sắc thể (NST). Khoảng 1:700 trẻ có khả năng mắc các hội chứng Down. 1:8000 đối với hội chứng Edwards và 1: 10.000 đối với hội chứng Patau.

Xét nghiệm NIPT – Illumina giúp sàng lọc các hội chứng bệnh di truyền phổ biến mà thai nhi mắc phải như:

- ✓ Bất thường số lượng nhiễm sắc thể
- ✓ Bất thường nhiễm sắc thể giới tính
- ✓ Đột biến vi mất đoạn
- ✓ Bất thường số lượng NST các NST còn lại

Activate Windows



Lựa chọn xét nghiệm	Mục đích xét nghiệm	Tình trạng mang thai
Xét nghiệm sự bất thường ở các nhiễm sắc thể phổ biến	Đơn nhiễm (thiếu một nhiễm sắc thể) hoặc Tam nhiễm (thừa một nhiễm sắc thể) 21, 18, 13 (còn được biết đến như là Hội chứng Down, Hội chứng Edwards và Hội chứng Patau)	Thai đơn và thai đôi
Xét nghiệm nhiễm sắc thể giới tính	Xác định giới tính của thai nhi và nếu có thừa hoặc có thiếu nhiễm sắc thể giới tính (chẳng hạn như hội chứng Turner hay hội chứng Klinefelter)	Thai đơn
Xét nghiệm sự bất thường ở các nhiễm sắc thể khác	Tam nhiễm 9 và 16	Thai đơn
Xét nghiệm bằng đột biến vi mất đoạn	Mất đoạn nhiễm sắc thể nhỏ có thể gây ra dị tật bẩm sinh và khả năng học tập của trẻ *	Thai đơn
* Hiện tại xét nghiệm Verifi phát hiện được các đột biến vi mất đoạn 22q11 (DiGeorge), mất đoạn 15q11 (Angelman/Prader-Willi), mất đoạn 1p36, mất đoạn 4p- (Wolf-Hirschhorn), 5p- (Cri-du-chat).		

## Độ chính xác của xét nghiệm NIPT - Illumina

Nhiễm sắc thể	Số lượng	Độ nhạy	95% CI	Độ đặc hiệu	95% CI
21	577	99.14%	98.0-99.7	99.94%	99.90-99.97
18	175	98.31%	95.0-99.6	99.90%	99.86-99.93
13	53	98.15%	90.0-99.9	99.95%	99.91-99.97

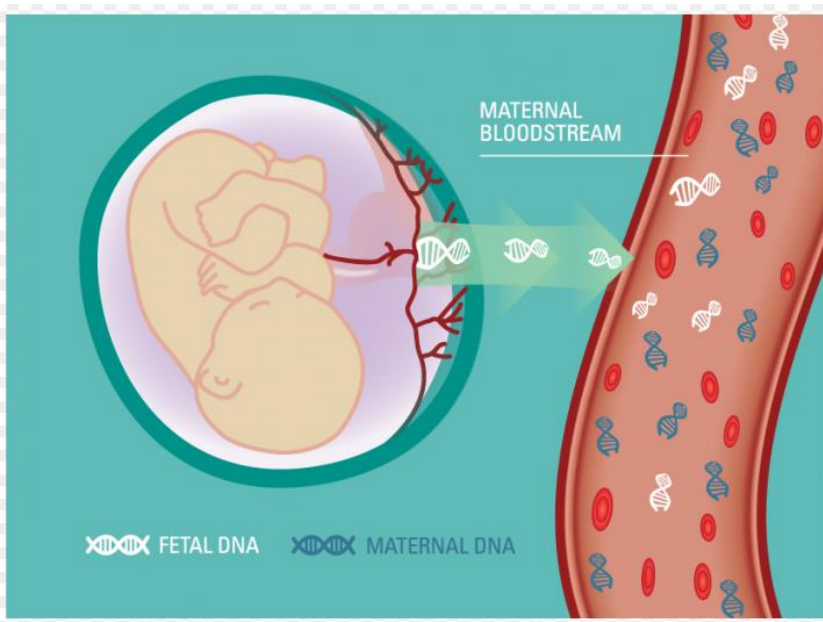
Hiệu suất xét nghiệm được tính dựa trên số liệu kết quả dương tính (Aneuploidy Detected) và nghi ngờ dương tính (Aneuploidy Suspected)

MX	508	95.0% (19/20)	75.1-99.9	99.0% (483/488)	97.6-99.7
----	-----	---------------	-----------	-----------------	-----------

*Xét nghiệm Verifi bao gồm một số các chọn lựa thêm về xét nghiệm những dị bội NST giới tính thường gặp, tương tự các kết quả của các xét nghiệm xâm lấn (chọc ối hoặc sinh thiết gai nhau).*

Nhiễm sắc thể	Số lượng	Độ nhạy	95% CI	Độ đặc hiệu	95% CI
XX	508	97.6%(243/249)	94.8 - 99.1	99.2% (257/259)	97.2 - 99.9
XY	508	99.1% (227/229)	96.9 - 99.9	98.9% (276/279)	96.9 - 99.8

XXX/XXY/XXY	Hiệu suất xét nghiệm không bao gồm các số liệu về các trường hợp dị bội hiếm gặp				Activate Windows Go to Settings to activate Windows
-------------	--	--	--	--	--

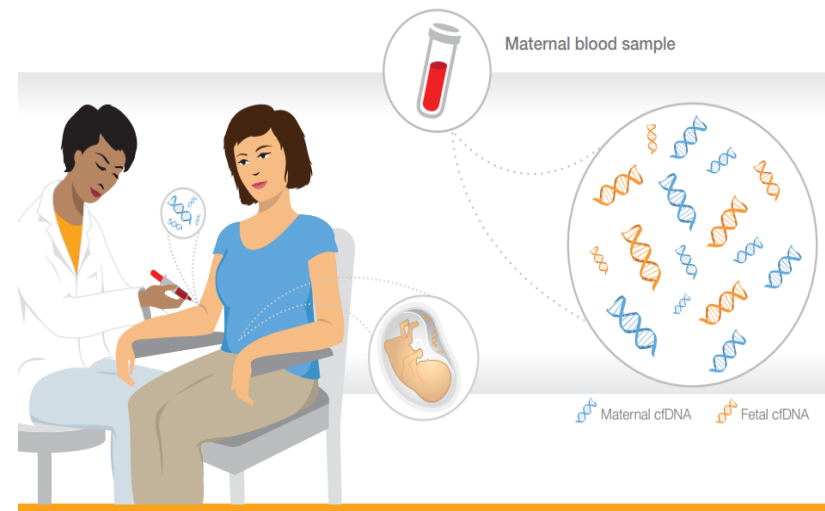


#### ❑ Cell free fetal DNA – cffDNA

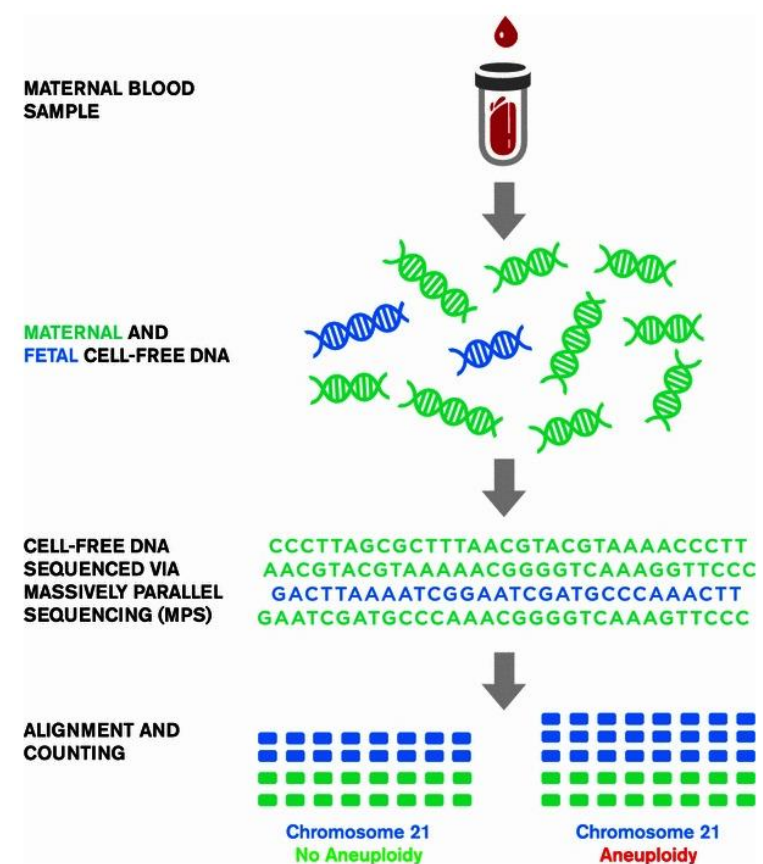
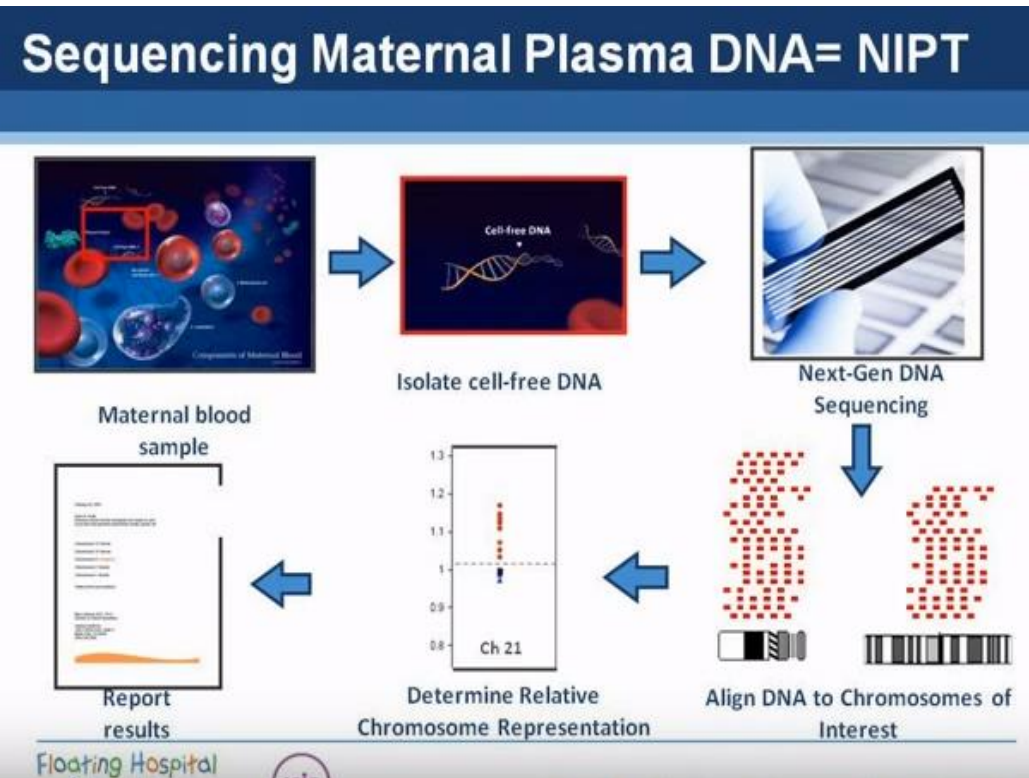
- đoạn DNA ngắn, kích thước 150-200bp
- Tăng dần theo thai kỳ 3 – 19% (từ tuần thai thứ 9 trở đi)
- Thời gian bán hủy trung bình 16,3 phút
- Không còn hiện diện sau khi sinh

NIPTs use cell-free DNA.

Detecting fetal chromosome abnormalities.



# NIPT thể hệ thứ nhất



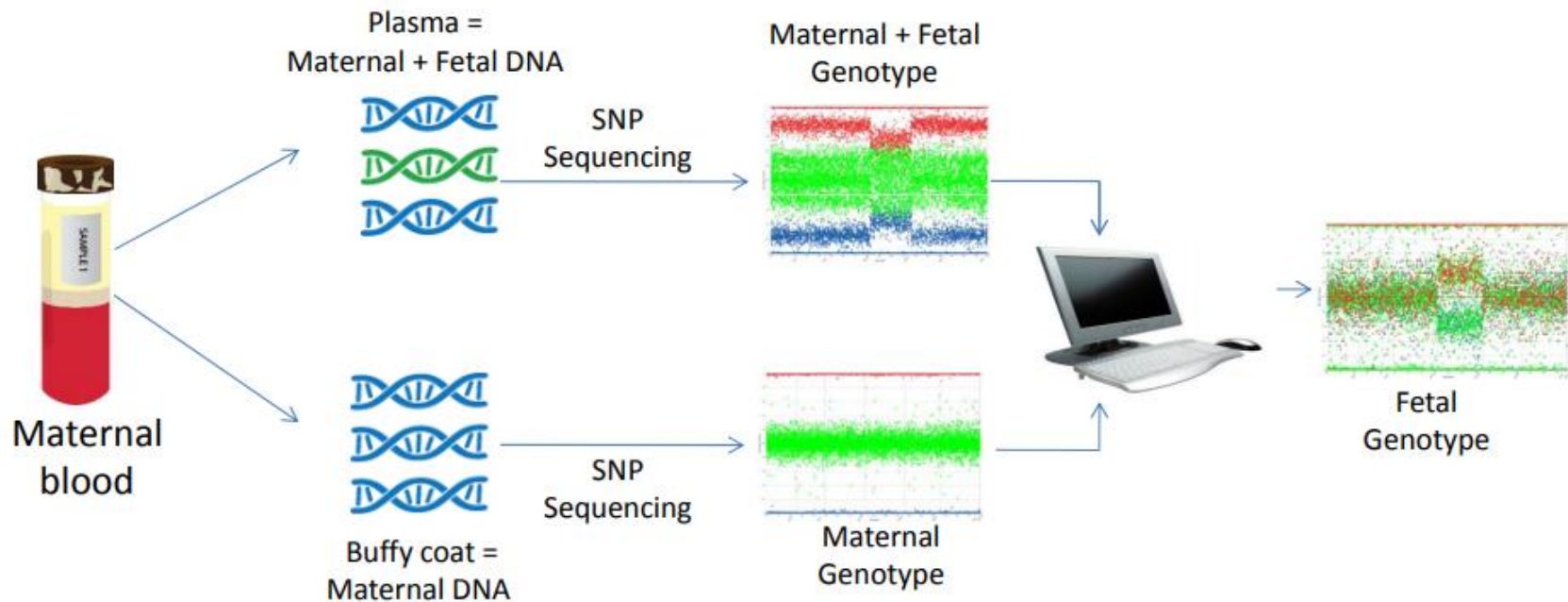
Cho tới nay, có 2 phương pháp NIPT chính đang được sử dụng phương pháp đếm định lượng là phương pháp được đa số các:

- nơi sử dụng và thuộc NIPT thế hệ thứ nhất,
- phương pháp kết hợp các thông tin kiểu gen dựa trên sự đa hình đơn nucleotide (Single nucleotide polymorphisms: SNP) đang được Netera sử dụng thuộc NIPT thế hệ thứ hai.



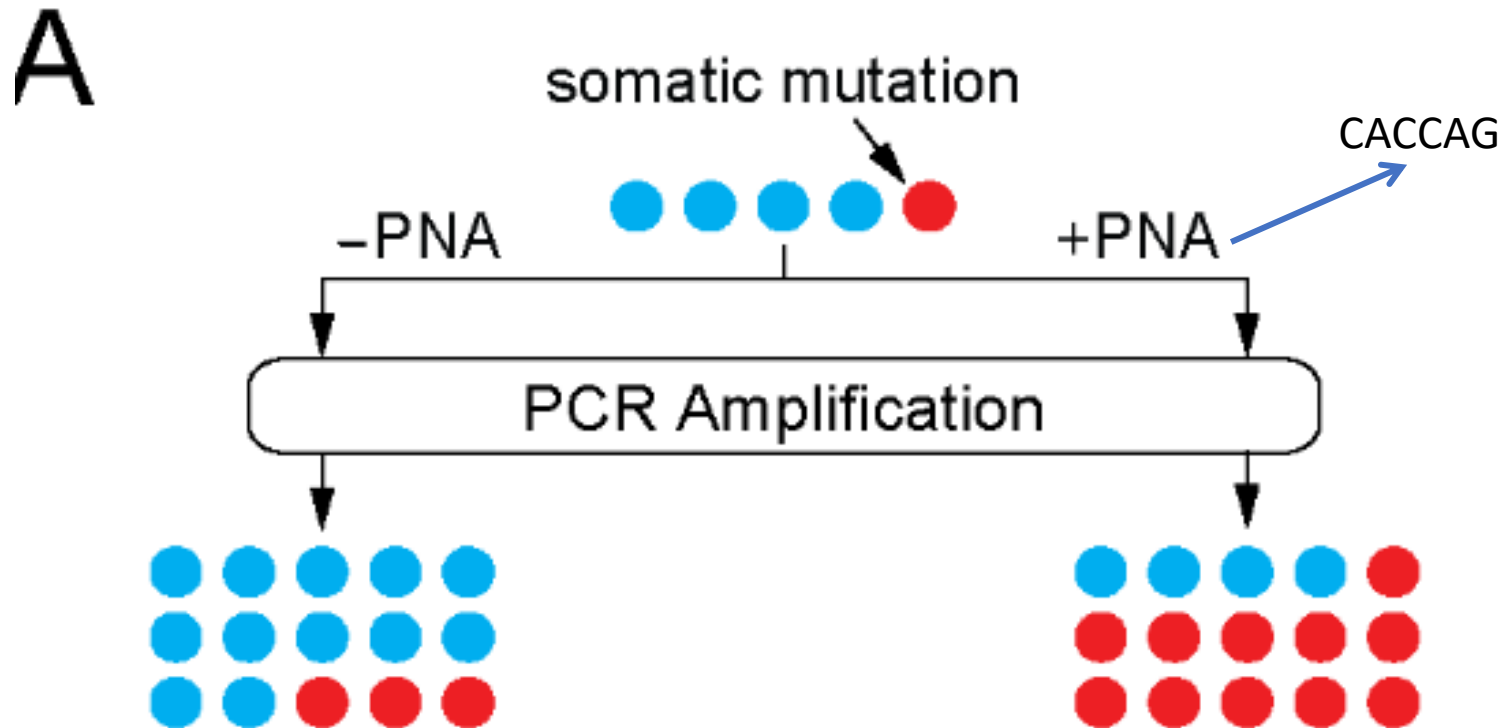
# Sequencing the Buffy Coat to Get Maternal SNP Genotype

**Panorama simultaneously targets 19,488 SNPs**



NIPT thế hệ thứ hai

# Peptide nucleic acid clamping to improve the sensitivity of Ion Torrent – based detection of an oncogenic mutation in KRAS



- Trong khi trình tự thế hệ tiếp theo (NGS) cho phép phân tích toàn diện về bộ gen, công nghệ có thể phải vật lộn để phát hiện các biến thể tần số thấp.
- Để cải thiện độ nhạy của NGS để phát hiện đột biến, chúng tôi đã tạo ra một peptide axit nucleic (PNA) kẹp, xác nhận bởi qPCR, được thiết kế để ràng buộc hoang dại KRAS (WT KRAS) trên codon 12 trong khuếch đại giai đoạn PCR của một sự chuẩn bị thư viện NGS



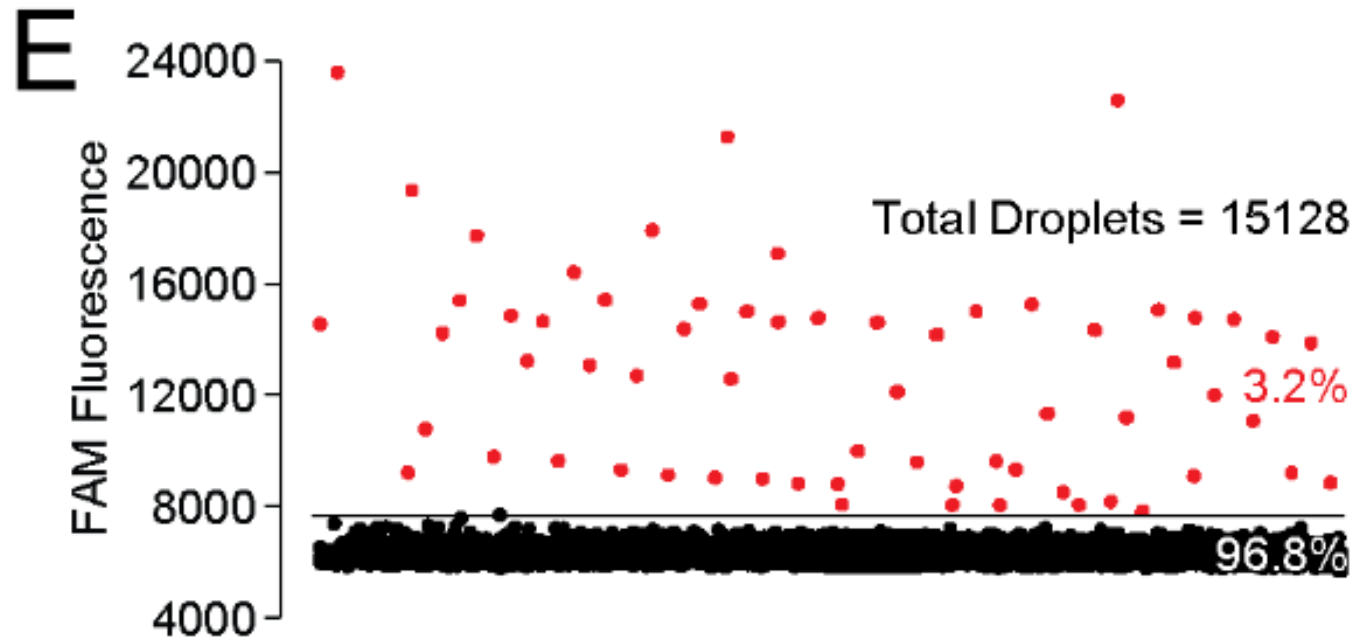
# Quy trình:

1. Peptide nucleic acid clamp
2. PCR primer and probe: Two Tagman probes (wild-type sequence; G12D mutation)
3. Droplet Digital PCR
4. Quantitative PCR
5. Ion torrent NGS protocol:
  - 207 amplicons ~ 2800 mutation from oncogen and TSG
  - Sequencing read were aligned to the reference human genome 19 using Torrent mapping alignment program.

# Quy trình

1. Thiết kế đoạn PNA (peptide nucleic acid) kẹp
2. Chuẩn bị môi và mẫu dò. Có 2 loại mẫu dò được sử dụng (mẫu dò cho WT và G12D mutation)
3. Sử dụng droplet digital PCR để khuếch đại trình tự mong muốn
4. Định lượng
5. Chạy NGS (ở đây sử dụng phương pháp Ion torrent): 207 đoạn khuếch đại, tương đương với 2800 đột biến trên Oncogene và TSG được sử dụng để chạy NGS. Trình tự thu được sau đó sẽ được aligned với genome 19

# Kết quả:



- Phân tích tình trạng KRAS của bệnh nhân ung thư phổi không tế bào nhỏ (NCSLC) sử dụng cfDNA plasmatic làm nguồn circulating tumour DNA.
- Chúng tôi đã sử dụng Droplet Digital PCR (ddPCR) để phát hiện ra đột biến KRASG12D ở tần số allelic là 3,2%

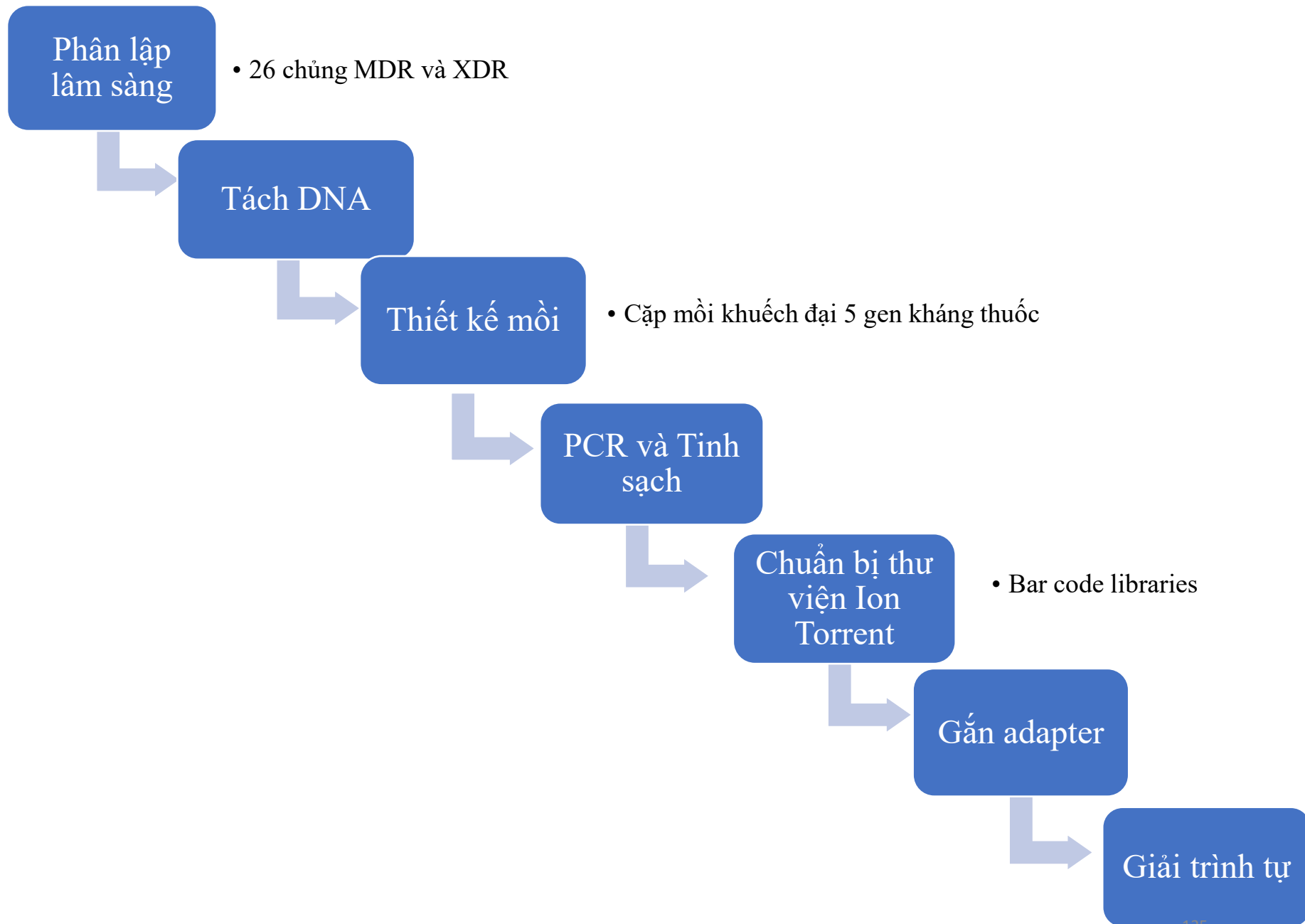
# Next-Generation Ion Torrent Sequencing of Drug Resistance Mutations in *Mycobacterium tuberculosis* Strains

## *Mycobacterium tuberculosis*:

- rpoB (rifampin)
- katG (isoniazid)
- pncA (pyrazinamide)
- gyrA (ofloxacin/ fluoroquinolone)
- rrs (aminoglycosides)

**MDR *M. tuberculosis*** : kháng rifampin (RIF) và isoniazid (INH)

**XDR *M. tuberculosis***: kháng cả RIF and INH cũng như là fluoroquinolones (FQs)



- Tổng cộng 10 axit amin thay thế của gen *rpoB* đã được xác định trong 26 chủng lâm sàng so với chủng hoang dại H37Rv.
- Sự đột biến S531L thông thường là phổ biến nhất, nhưng các đột biến tại các vị trí 516 và 526, cũng được biết đến để đề kháng rifampin, được quan sát thấy

No. of isolates	Amino acid substitution(s) <sup>b</sup> yielded by the <i>rpoB</i> gene (3,619 bp)	Rifampin result by:		
		Ion Torrent <sup>c</sup>	Hain LPA	Bactec MGIT 960
9	S531L	Resistant	Resistant	Resistant
1	S531L, V194I	Resistant	Resistant	Resistant
1	S531L, Y645H	Resistant	Resistant	Resistant
1	H526D	Resistant	Resistant	Resistant
1	H526Y, S509R	Resistant	Resistant	Resistant
1	H526R	Resistant	Resistant <sup>d</sup>	Resistant
5	Wild type <sup>b</sup>	Sensitive	Sensitive	Sensitive
6	D516G, L533P	Resistant	Resistant <sup>d</sup>	Resistant
1	D516V	Resistant	Resistant	Resistant

Actin

***katG* gene mutations:** 4 amino acid biến đổi được quan sát thấy trong sản phẩm gen *katG* - S315T, được biến đến là kháng isoniazid.

No. of isolates	Amino acid substitution(s) <sup>a</sup> yielded by the <i>katG</i> gene (2,447 bp)	Isoniazid result by:		
		Ion Torrent <sup>b</sup>	Hain LPA	Bactec MGIT 960
11	S315T	Resistant	Resistant	Resistant
5	S315T, R463L	Resistant	Resistant	Resistant
1	W191R, R463L	Sensitive	Sensitive	Sensitive
7	Wild type <sup>a</sup>	Sensitive	Sensitive	Sensitive
1	R463L	Sensitive	Sensitive	Sensitive
1	N138H	Sensitive	Sensitive	Sensitive

# *pncA* gene mutations

No. of isolates	Amino acid substitution(s) <sup>a</sup> yielded by the <i>pncA</i> gene (3,619 bp)	Pyrazinamide result by:	
		Ion Torrent <sup>b</sup>	Bactec MGIT 960
3	<b>C14R</b>	Resistant	Resistant
1	<b>A102V</b>	Resistant	Resistant
1	<b>Q122Stop</b>	Resistant	Resistant
16	Wild type <sup>a</sup>	Sensitive	Sensitive
1	<b>V139G</b>	Resistant	Resistant
1	<b>R154G</b>	Resistant	Resistant
2	<b>L172P</b>	Resistant	Resistant
1	Silent (C195T) <sup>c</sup>	Sensitive	Sensitive

<sup>a</sup>Compared to the H37Rv reference strain.

<sup>b</sup>Pyrazinamide resistance is known to occur in several mutations described by Mphahlele et al. (13) (indicated in bold).

<sup>c</sup>One strain contained a silent (synonymous) nucleotide mutation corresponding to position 195 (C→T).



# Đột biến gen pncA

- Bảy đột biến nucleotide được ghi nhận ở ít nhất một dòng trong số 561 bp bao gồm vùng mã hóa đầy đủ cho gen pncA (Bảng 4).
- Chín trong số 26 chủng (34,6%) có chứa một đột biến axit amin cung cấp khả năng chống lại pyrazinamide (PZA) (Bảng 4).
- Trong một dòng, đột biến nucleotide im lặng (đồng nghĩa) được đặc trưng ở vị trí tương ứng với axit amin 195 (C195T).

- Năm chủng có chứa các chất thay thế amino acid đã được biết trước đây (C14R, A102V, V139G, R154G, và L172P) được biết đến để đề kháng với pyrazinamid.
- Một đột biến mới, không được báo cáo trước đây ở những nơi khác, bao gồm một codon kết cuối đã được tìm thấy trong một phân lập ở nhóm 122 (Q122Stop) trong protein PncA

# gyrA gene mutations

9 đột biến duy nhất được quan sát thấy trong gen mã hoá gyrA dài 2,517 bp mã hoá tiểu đơn vị của enzyme gyrase DNA

No. of isolates	Amino acid substitution(s) <sup>a</sup> yielded by the <i>gyrA</i> gene (2,664 bp)	Fluoroquinolone result by:	
		Ion Torrent <sup>b</sup>	Bactec MGIT 960
3	E21Q, S95T, G247S, G668D	Sensitive	Sensitive
2	E21Q, <b>D94G</b> , S95T, G668D	Resistant	Resistant
1	E21Q, <b>G88C</b> , S95T, G668D	Resistant	Resistant
10	E21Q, S95T, G668D	Sensitive	Sensitive
1	Wild type <sup>a</sup>	Sensitive	Sensitive
1	E21Q, S95T, G668D, Q613Q/E <sup>c</sup>	Sensitive	Sensitive
1	E21Q, S95T, G668D, L549S/L <sup>c</sup>	Sensitive	Sensitive
1	E21Q, <b>D94Y</b> , S95T, G668D	Resistant	Resistant
6	E21Q, <b>A90V</b> , S95T, G247S, G668D	Resistant	Resistant

- Sự kháng thuốc đối với fluoroquinolones (FQs) chỉ được ghi nhận ở các dòng có các đột biến ở vùng xác định độ bền quinolone (QRDR) được xác định bởi sự thay thế ở gyrA ở các vị trí tương ứng với axit amin 88, 90 và 94.
- Một số đột biến bổ sung cũng được quan sát thấy ở vùng ngoài QRDR, bao gồm hai đột biến "biến dạng" ở vị trí 549 và 613 trong protein GyrA

# *rrs* (16S) gene mutations.

- 4 đột biến nucleotide được ghi nhận trong đoạn 1.540 bp bao gồm gen 16s *rrs* có chiều dài đầy đủ.
- 7 trong số 26 (27%) chủng lâm sàng cho thấy có khả năng đề kháng với aminoglycosid, và tất cả các chủng có đột biến A1401G được cho là có khả năng đề kháng.

No. of isolates	Amino acid substitution(s) <sup>a</sup> in the <i>rrs</i> (16S) gene (1,680 bp)	Kanamycin result by:	
		Ion Torrent <sup>b</sup>	Bactec MGIT 960
1	G878A	Sensitive	Sensitive
12	Wild type <sup>a</sup>	Sensitive	Sensitive
1	A514C, A1401G	Resistant	Resistant
6	A1401G	Resistant	Resistant
3	A514C	Sensitive	Sensitive
1	C492T	Sensitive	Sensitive
1	C492T, A514C	Sensitive	Sensitive
1	A514C	Sensitive	Sensitive

- phương pháp này có thể dễ dàng được mở rộng để bao gồm tất cả 16 gen hiện đang gây ra sự kháng thuốc của *M. tuberculosis*.
- Phân tích gen kéo dài bằng cách sử dụng Ion Torrent PGM có thể xác định những đột biến mới, khi có liên quan đến thử nghiệm MIC kiểu hình, có thể xác định được các vi khuẩn kháng lao mới của *M. tuberculo*

# Next Gen Sequencing [NGS]

- History of DNA Sequencing

- Maxam-Gilbert
- Sanger
- ABI

Human Genome: 1990-2000

- NGS Technologies:

- 454, Illumina, PacBio, ABI, Helicos,
- Ion Torrent, Nanopores

- Applications:

- Genomes, RNASeq, ChIPSeq, CGH, CancerGenome, Environmental



Presented by Dominic Suci, Ph.D.

- RNA-seq và ChIP-Seq ứng dụng vào nghiên cứu bệnh truyền nhiễm
- Kỹ thuật phân tích nhiễm sắc thể mới toàn diện hơn, đó là CGH (phép lai tạo gen có so sánh – comparative genomic hybridization) – chọn phôi trước khi chuyển