



Posgrado en  
**Biología  
Integrativa**



# Análisis de transcriptomas

*Clase 1 - Conceptos básicos y técnicas de estudio*

Biología Computacional 2016

Selene L. Fernández-Valverde

# Objetivos de aprendizaje

En estas tres semanas aprenderemos:

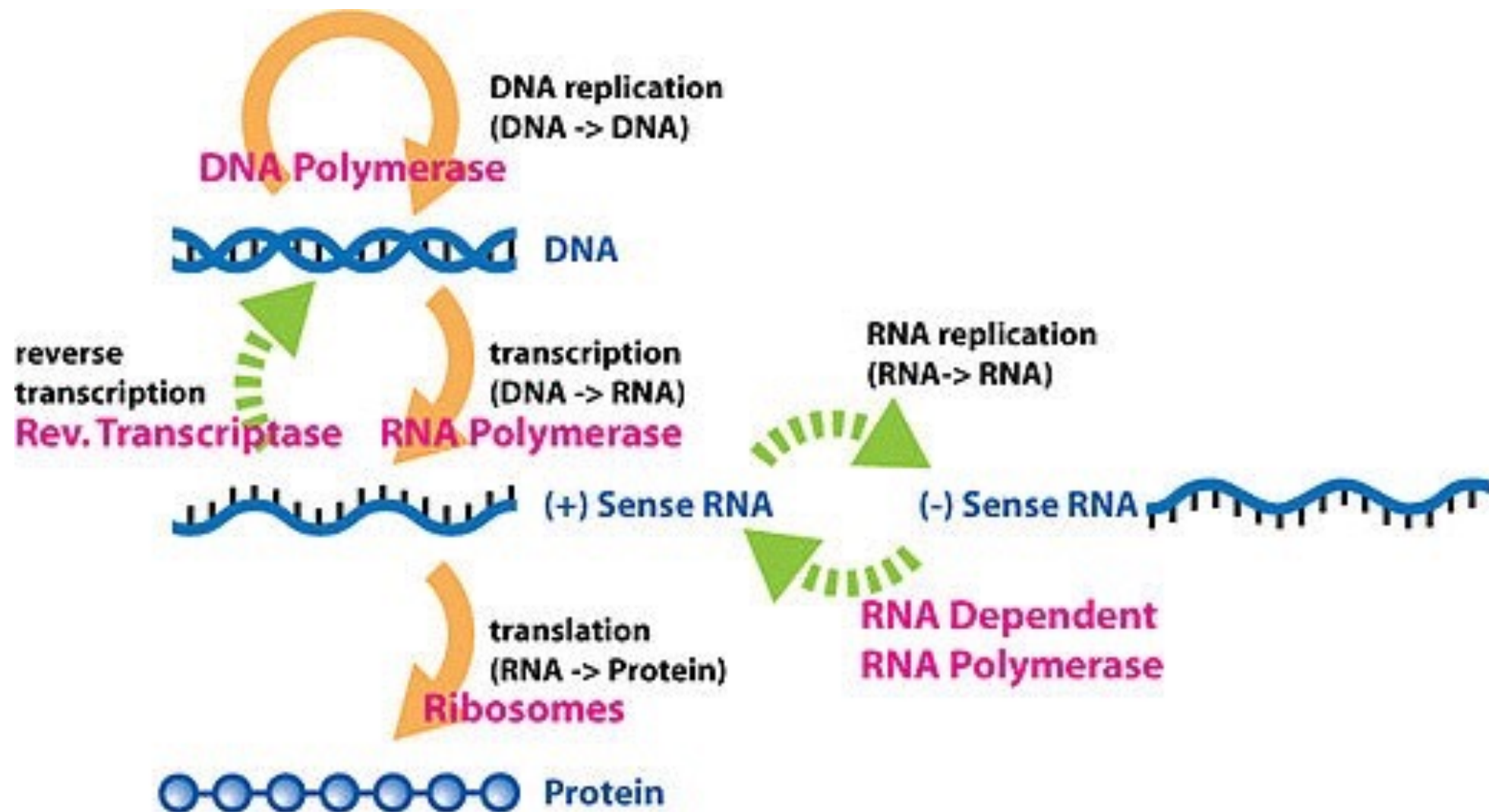
- Que técnicas se utilizan para secuenciar datos transcriptómicos
- Como sabemos si son de la calidad adecuada
- Como los ensamblamos y/o mapeamos a un genoma
- Como cuantificamos los niveles de expresión de un ARN mensajero o transcrito en cada transcriptoma
- Como medimos expresión diferencial de transcritos y/o genes
- Como interpretamos estos cambios en un contexto biológico

# ¿Qué es un transcriptoma?

El transcriptoma esta compuesto por todas las moléculas de ARN mensajero, o transcritos, que existen en la célula. Generalmente también se refiere a la abundancia específica o expresión de cada uno de estos transcritos en un tipo celular específico. Dado que los niveles de expresión génica cambian implica que el transcriptoma es dinámico.

La transcriptómica es el estudio del transcriptoma usando técnicas de análisis que permiten medir el nivel de expresión y, en algunos casos, la secuencia de los ARNm en una muestra de manera simultánea.

# Regresando al Dogma Central



El transcriptoma se genera a través del proceso de transcripción que copia genes codificados ADN a ARN que, en algunos casos, es traducido a proteínas en el ribosoma.

# ¿Porqué es importante medir cambios en la expresión génica - transcriptoma?

Cambios en la expresión génica correlacionan con cambios biológicos en distintas condiciones. Por ejemplo, la identidad de las células en nuestros tejidos se define en gran parte dado por el tipo de proteínas que expresa. Usamos el transcriptoma como una aproximación para ver cambios en expresión génica así como cambios en ARNs no codificantes.

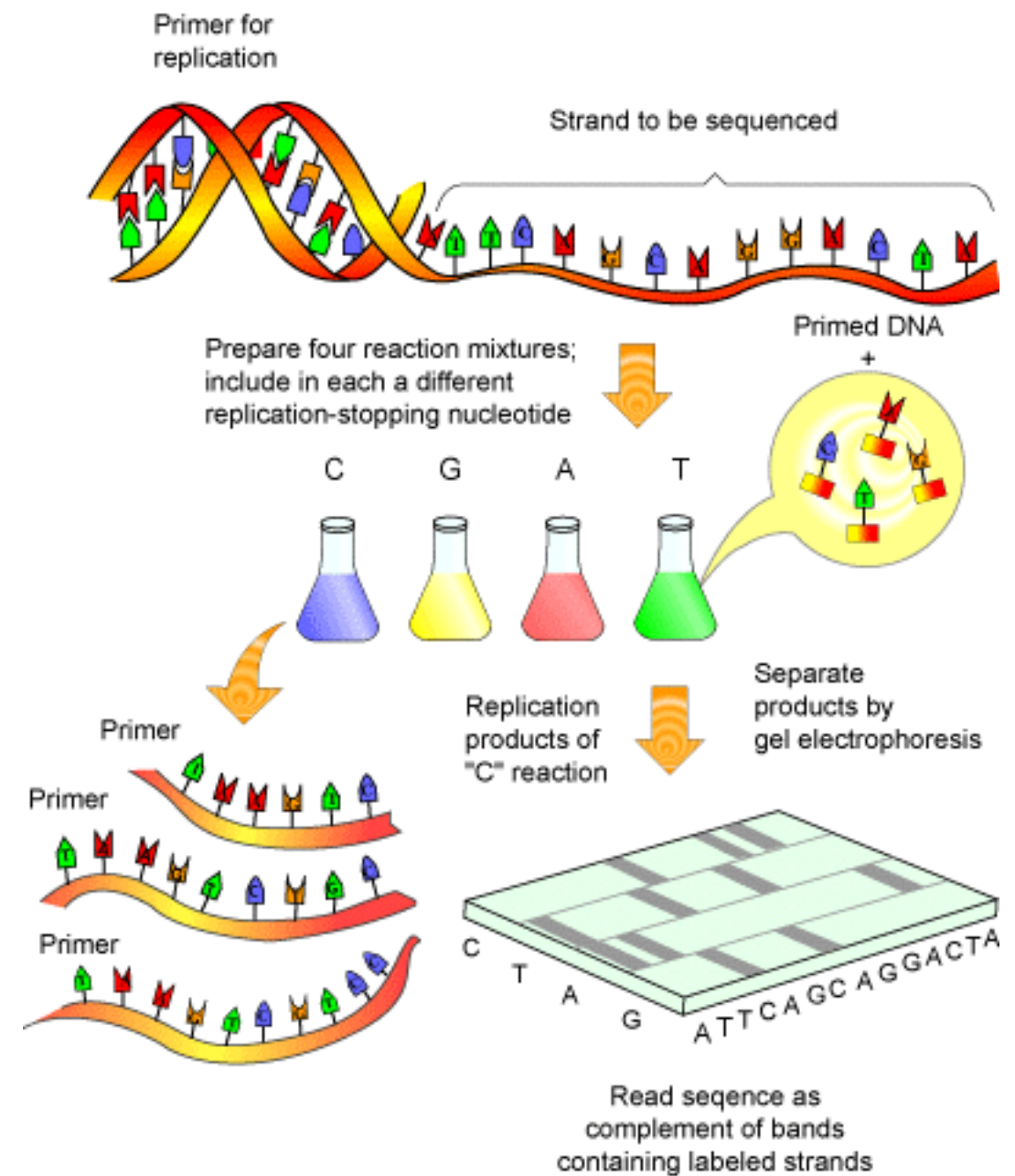
# Técnica de estudio de transcriptomas

- Microarreglos
- Secuenciación:
  - Sanger
  - Masiva (next generation sequencing)

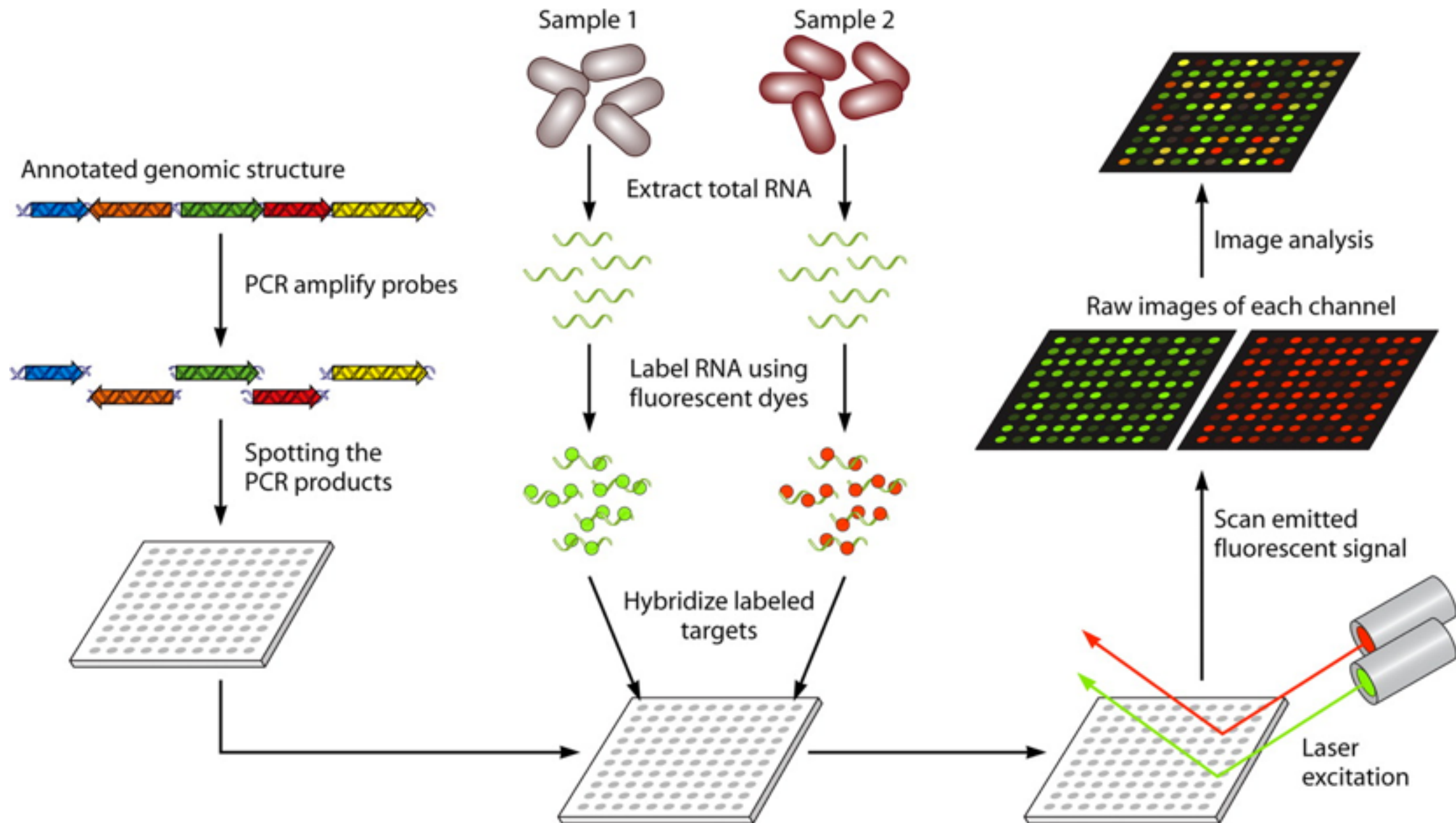


# Secuenciación tipo Sanger

- Se aísla y amplifica el ADN o ARN de interés.
- Se preparan 4 reacciones con el ADN que se quiere secuenciar, polimerasa de ADN y nucleótidos y un poco de uno de los nucleótidos que detienen la replicación del ADN (nucleótidos dideoxy)
- Los nucleótidos terminadores pueden ser marcados usando fluoróforos o radioactividad.



# Microarreglos





# Microarreglos

## Ventajas

- Bajo costo
- Alta reproducibilidad
- Análisis estandarizado

## Desventajas

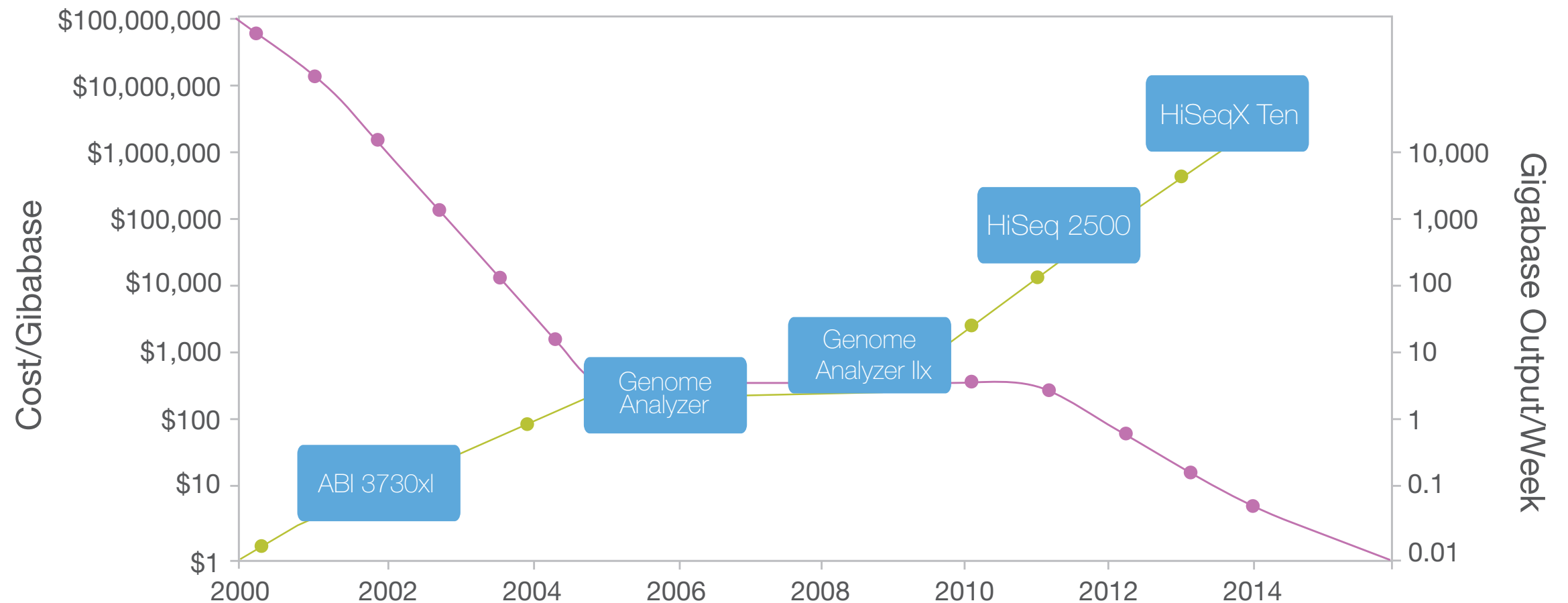
- Solo se pueden identificar cambios en la expresión de secuencias conocidas.
- Difícil identificar cambios en la expresión de isoformas.
- Imposible identificar cambios en splicing u otros procesos que afectan directamente a la secuencia.

# Técnicas de secuenciación masiva

- Pyrosequencing
- Sequencing by synthesis
- Sequencing by ligation
- Ion semiconductor
- Nanopore sequencing
- Single Molecule Real Time Sequencing (SMRT)



# La revolución de la secuenciación



**Figure 1: Sequencing Cost and Data Output Since 2000**—The dramatic rise of data output and concurrent falling cost of sequencing since 2000. The Y-axes on both sides of the graph are logarithmic.

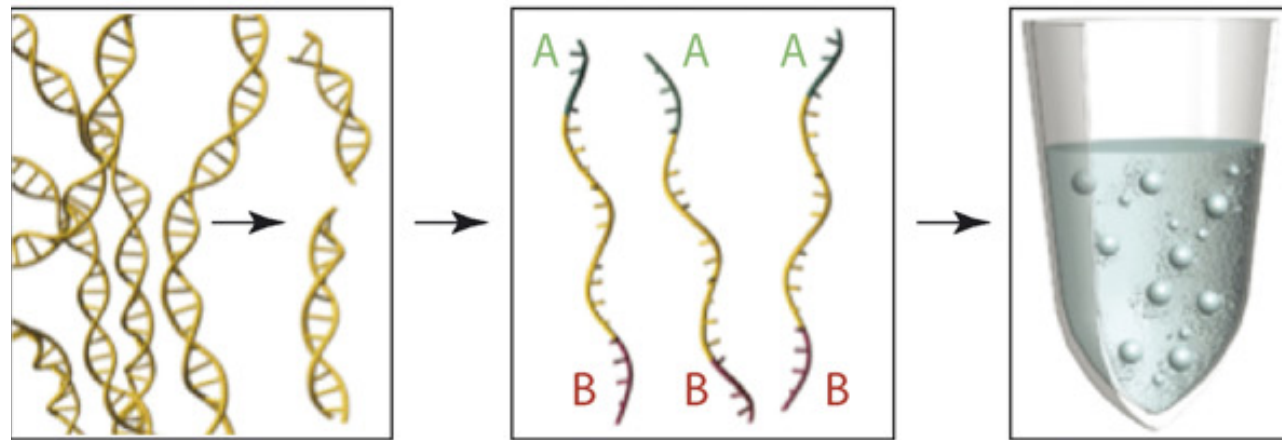
# Técnicas de secuenciación masiva

- **Pyrosequencing**
- **Sequencing by synthesis**
- Sequencing by ligation
- Ion semiconductor
- **Nanopore sequencing**
- **Single Molecule Real Time Sequencing (SMRT)**

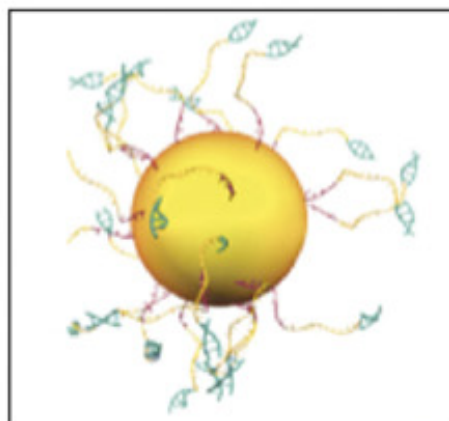
# Pirosecuenciación - 1

Roche (454) GSFLX Workflow:

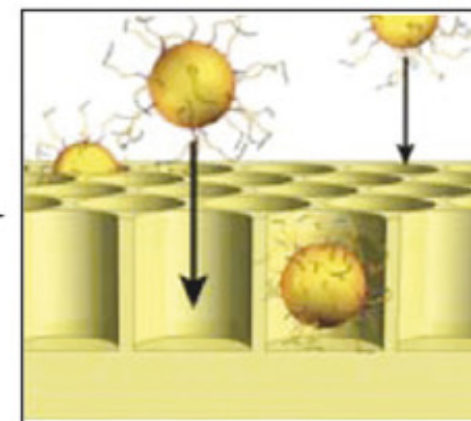
Library construction



Emulsion PCR

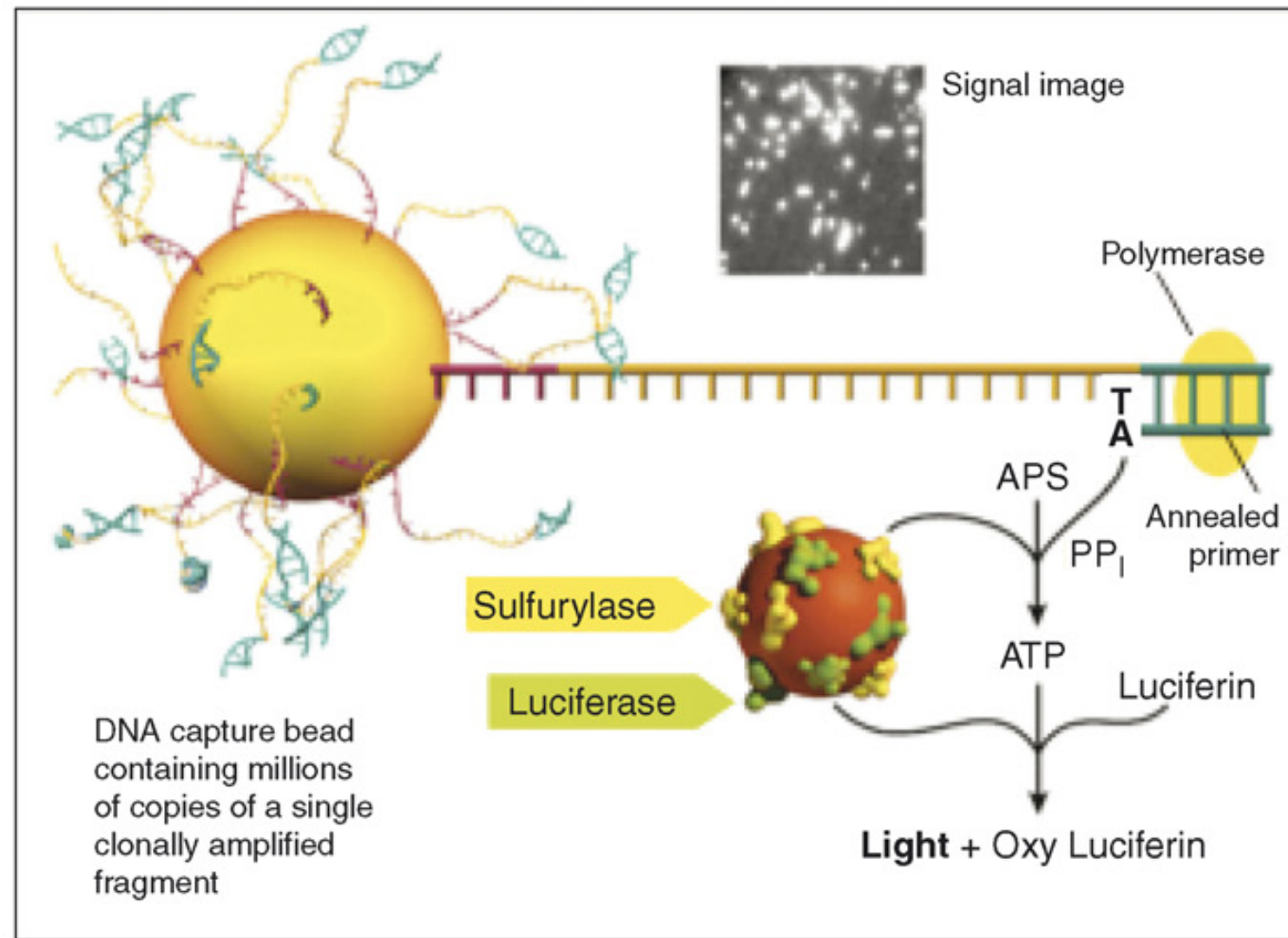


PTP loading





# Pirosecuenciación - 2



Pyrosequencing reaction

Reacción enzimática  
chemoluminiscente

# Pirosecuenciación

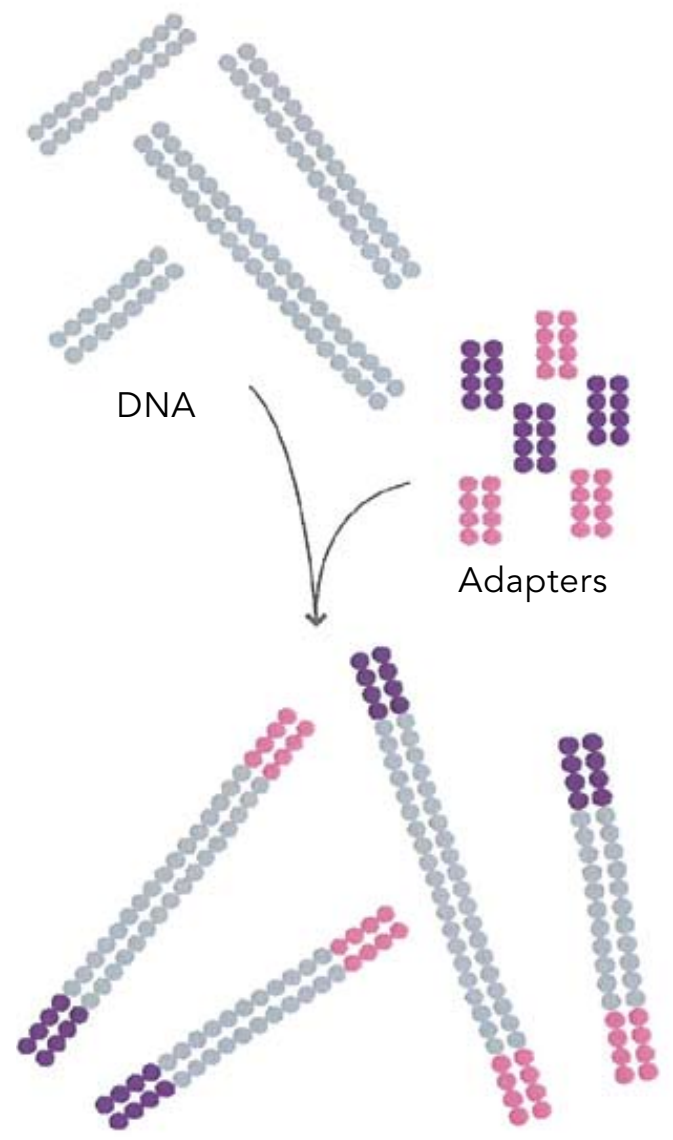
## Ventajas

- Costo razonable
- Secuencias largas (500 nts)

## Desventajas

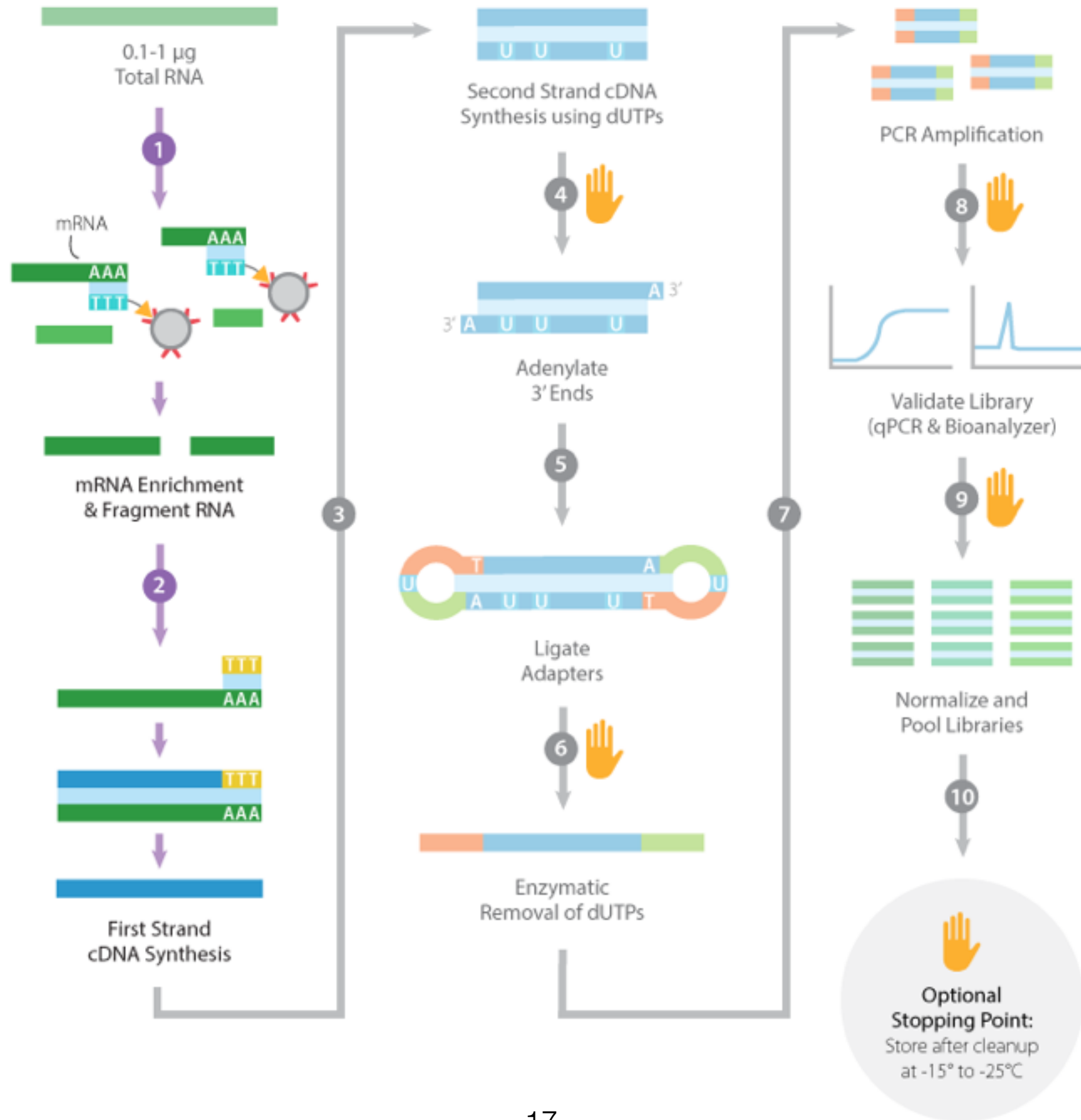
- Poca cantidad de secuencias producidas
- Alta número de errores en regions con el mismo nucleótido (homopolímeros)
- Con la competencia de otras tecnologías y dado su alto nivel de errores ha entrado en desuso.

# Illumina - sequencing by synthesis - 1

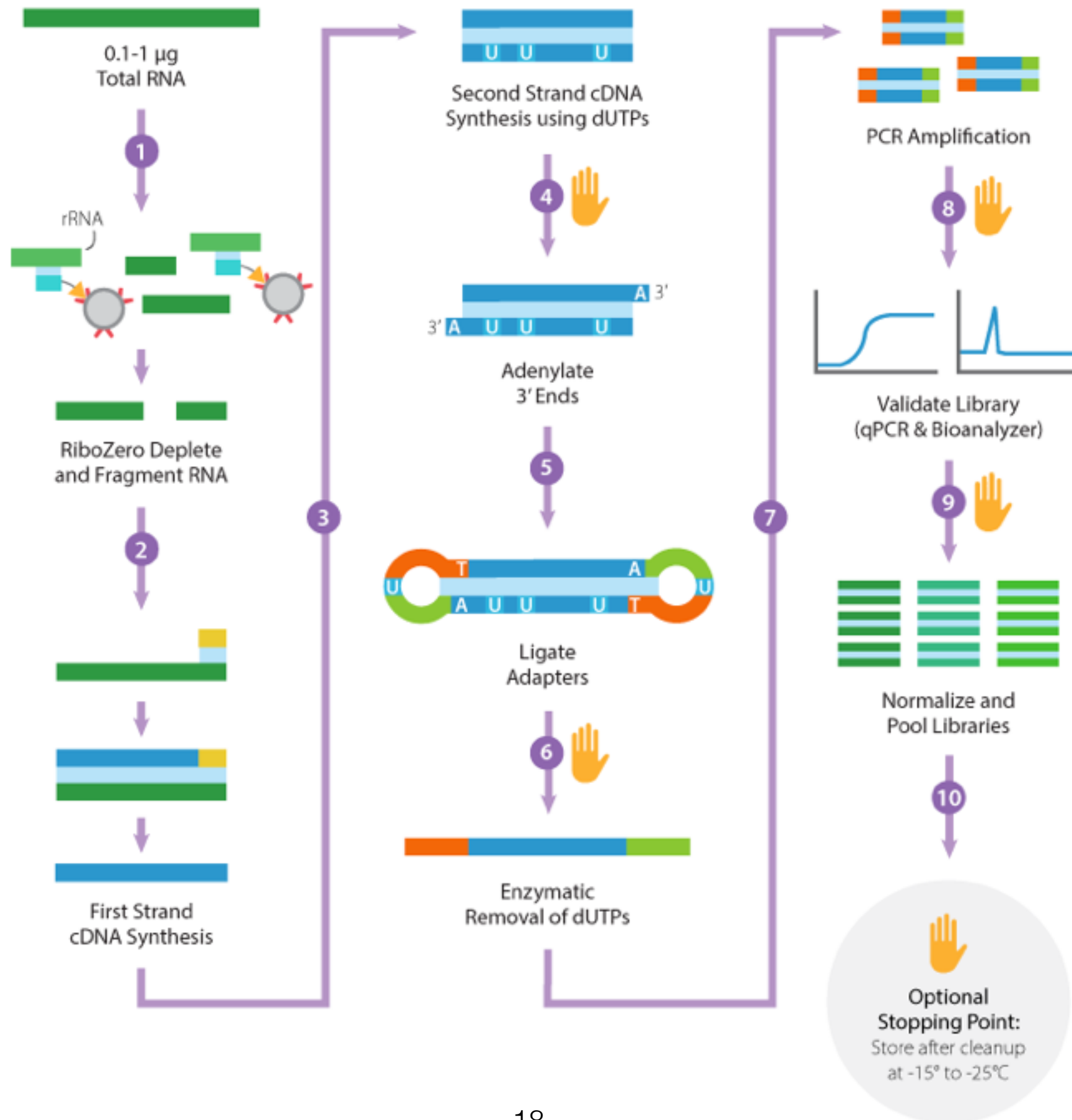


- Se empieza uniendo adaptadores a los fragmentos de ADN o ARN que queremos secuenciar.

# TruSeq Stranded mRNA Kit

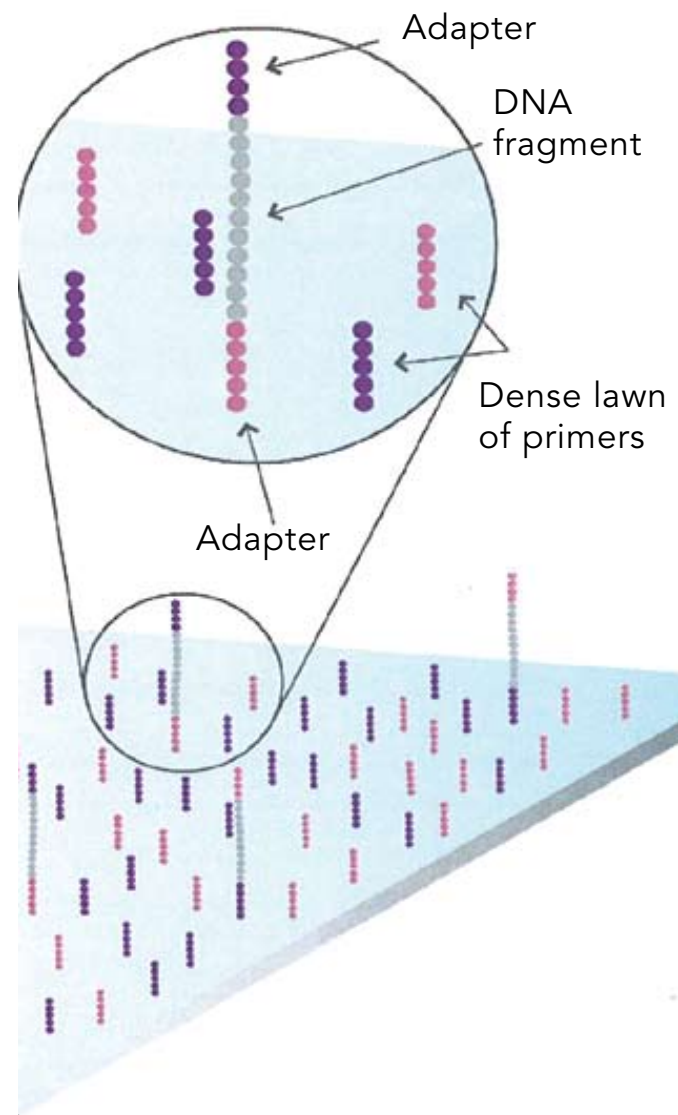


# TruSeq Stranded Total RNA Kit



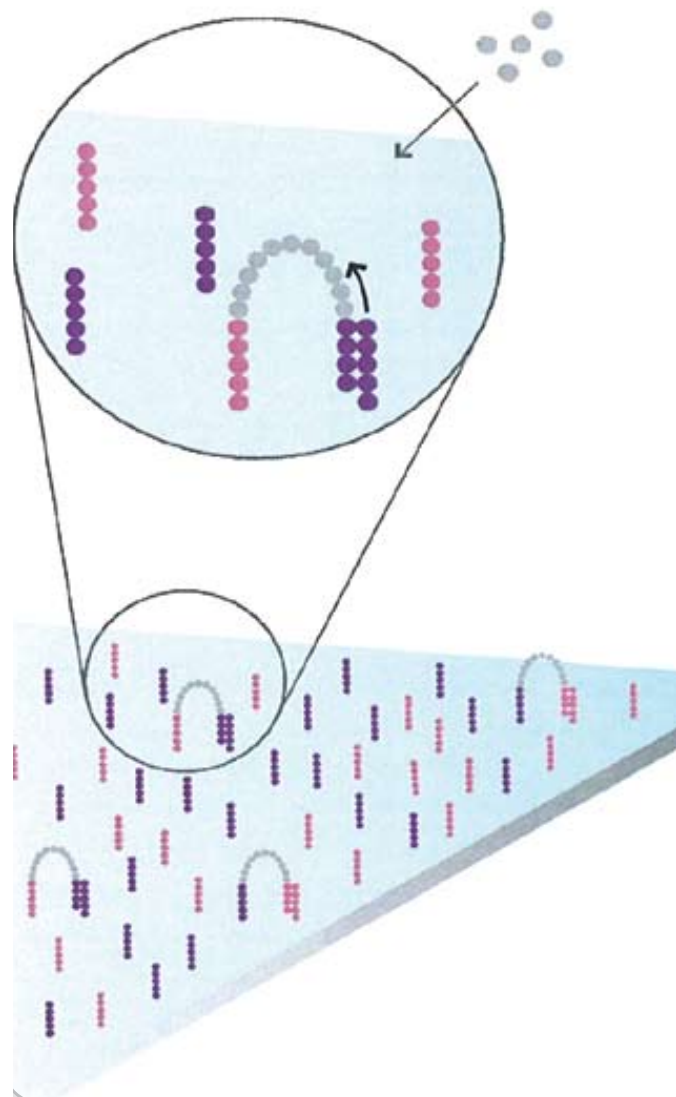


# Illumina - sequencing by synthesis - 2



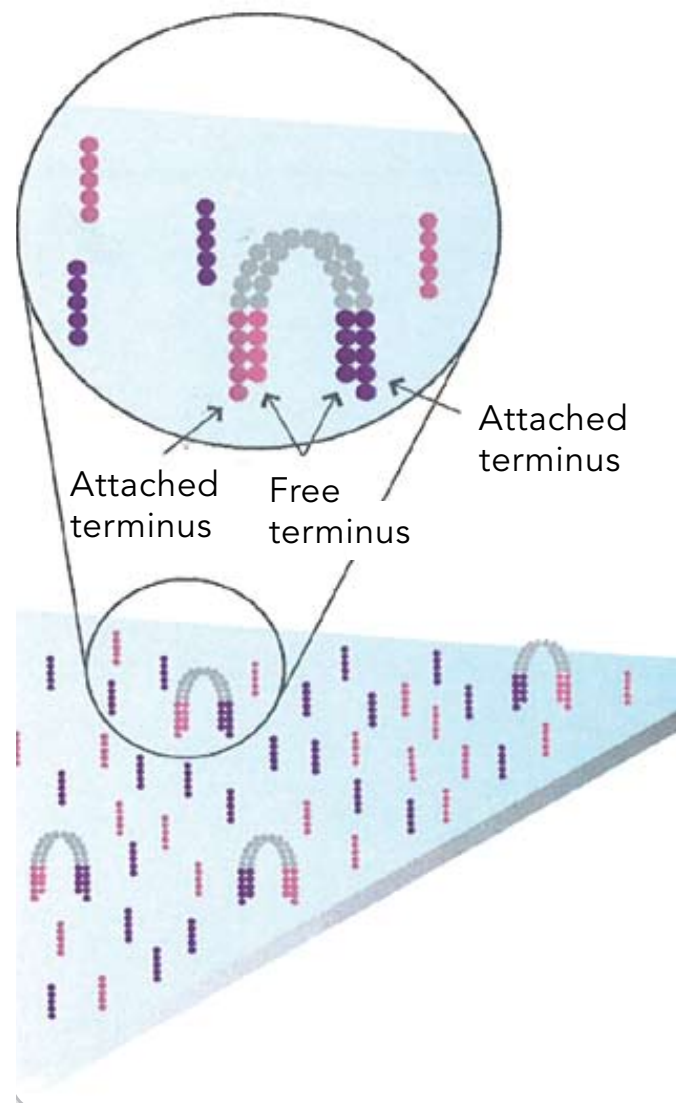
- Los templados se inmovilizan en una celda de flujo (flow cell)
- En el caso de RNA-Seq se usa complementariedad con el adaptador para sintetizar una nueva cadena de cDNA y no ignorar la direccionalidad del transcrito.

# Illumina - sequencing by synthesis - 3



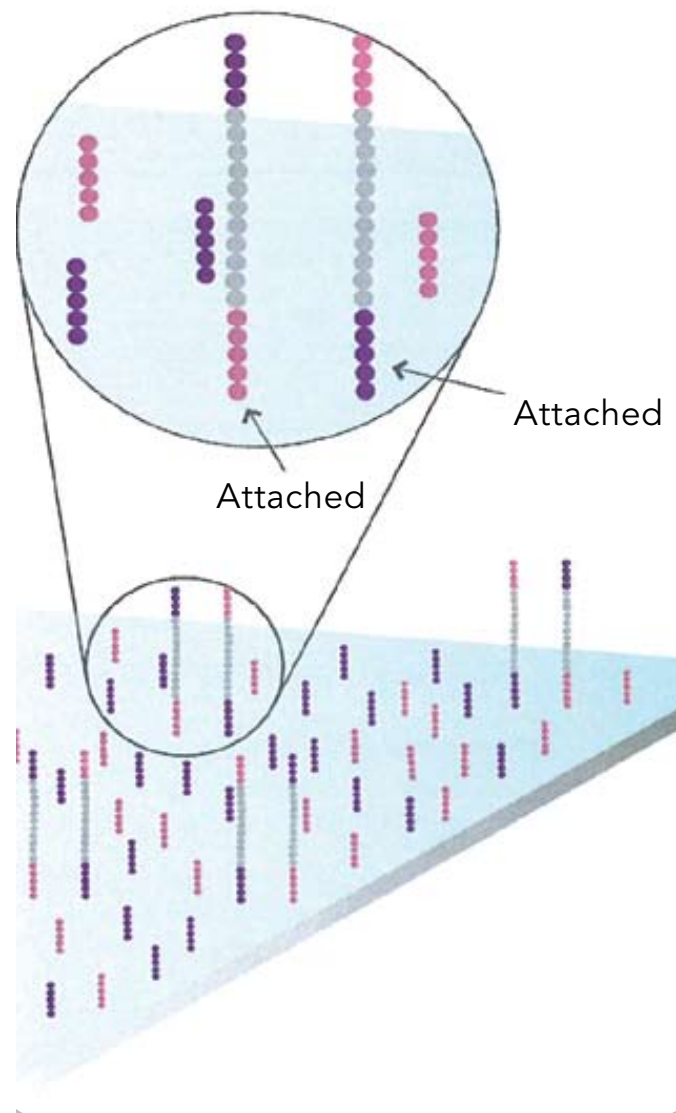
- Se sintetiza una cadena de ADN complementaria al templado en la superficie.

# Illumina - sequencing by synthesis - 4



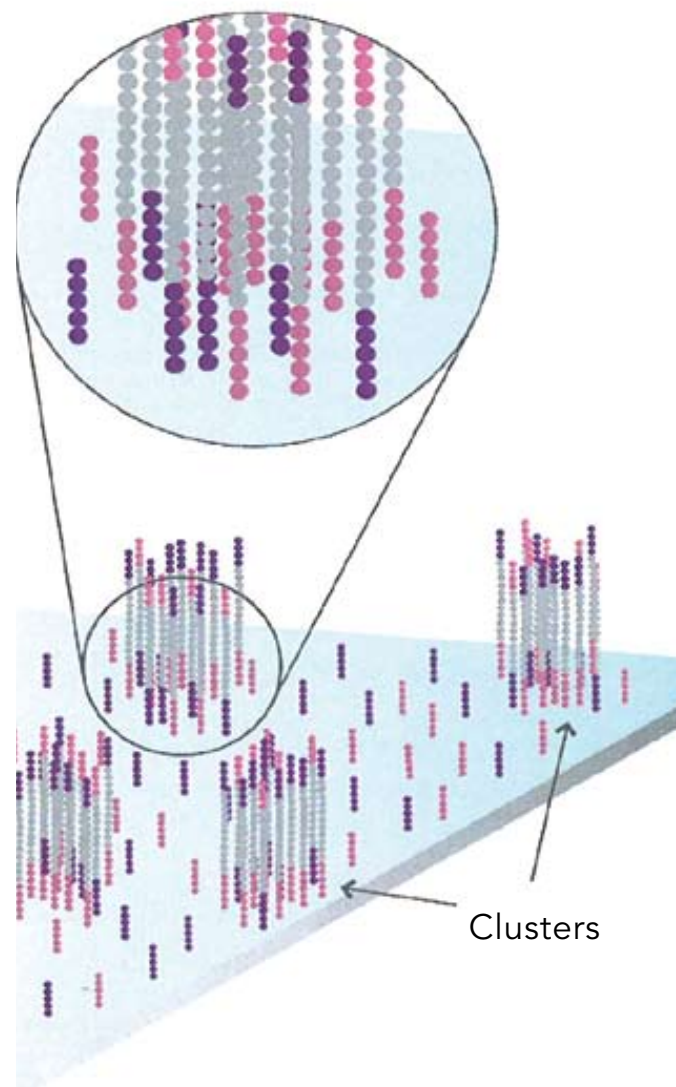
- Se sintetiza una cadena de ADN complementaria al templado en la superficie.

# Illumina - sequencing by synthesis - 5



- Se separan los templados usando alta temperatura.

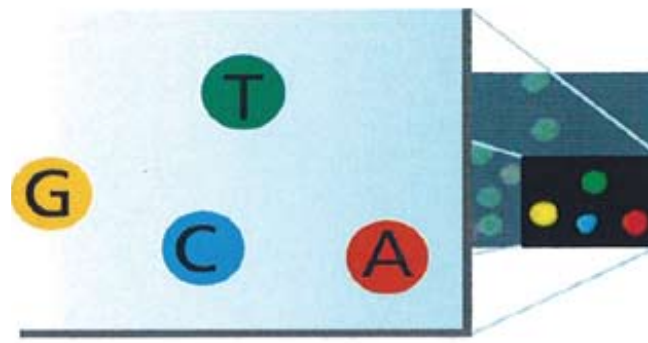
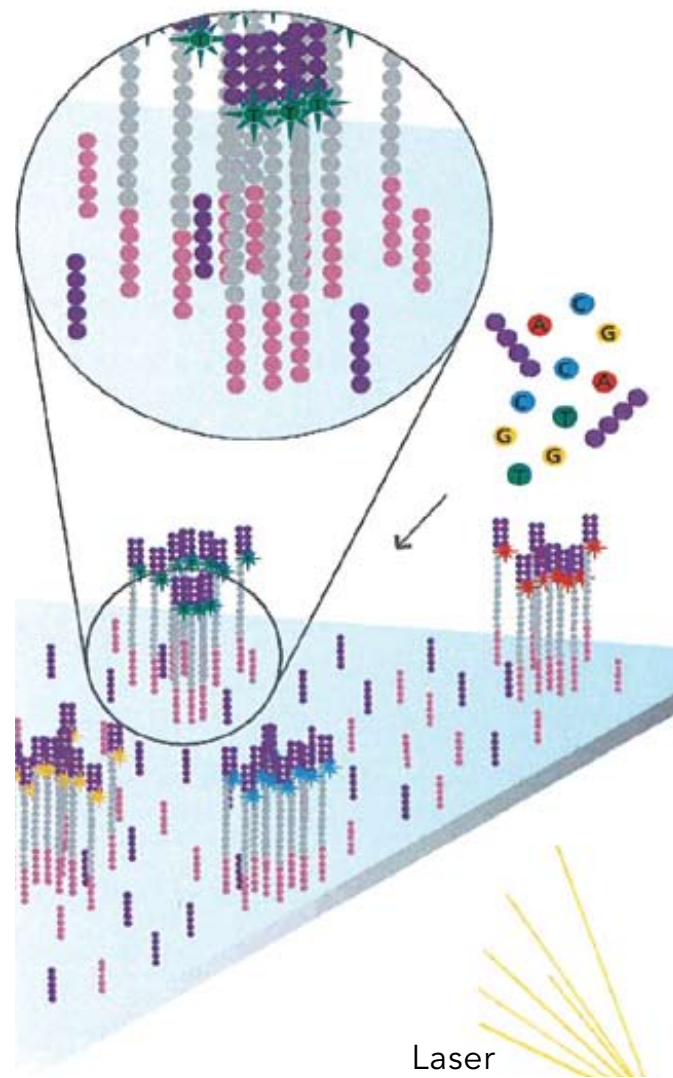
# Illumina - sequencing by synthesis - 6



- Se repite este proceso cientos de veces hasta generar una “colonia” o cluster de transcritos idénticos.

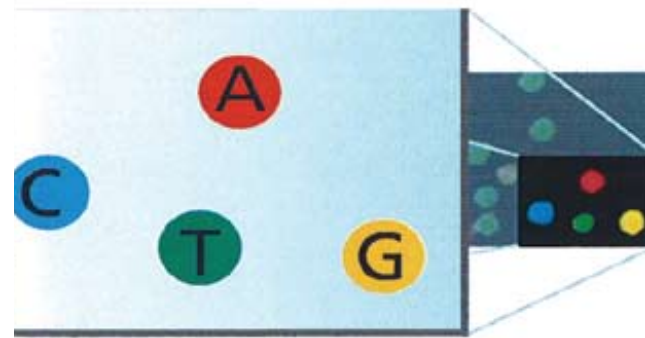
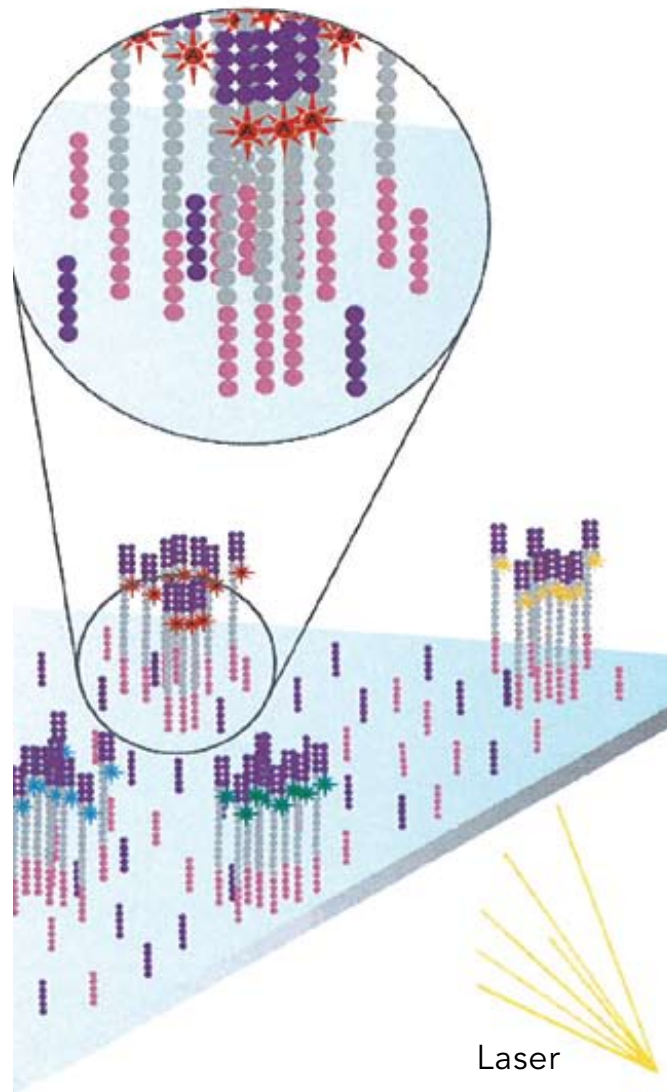


# Illumina - sequencing by synthesis - 7



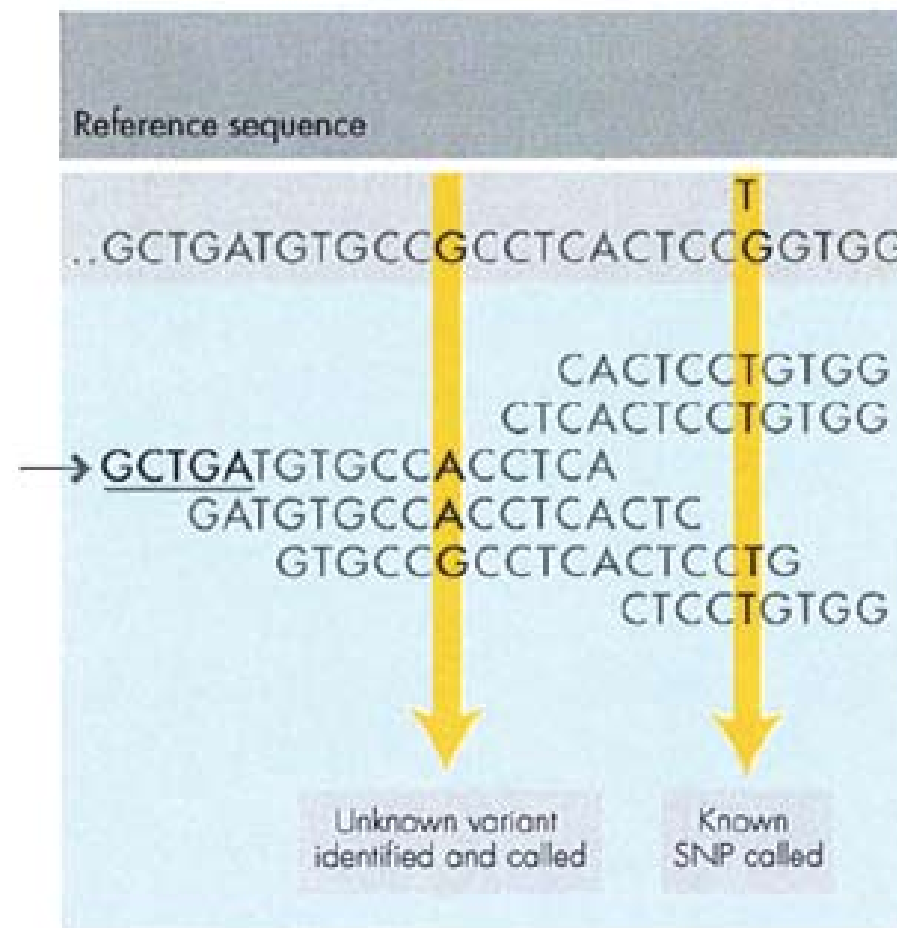
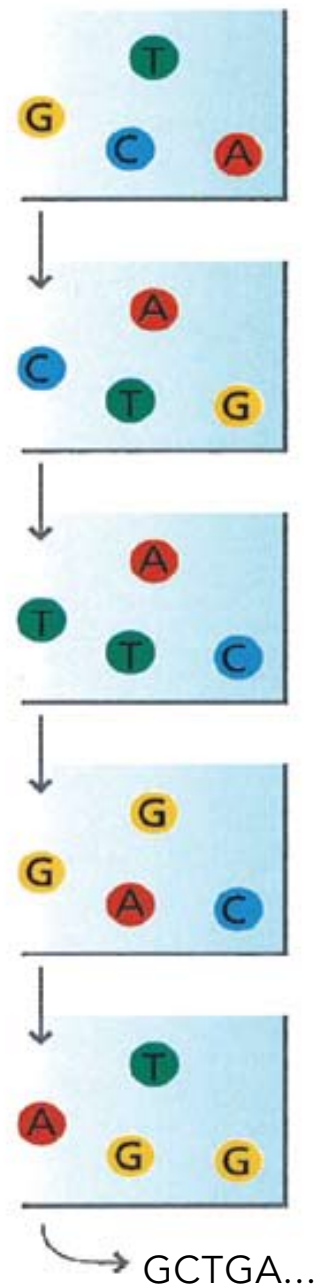
- Se agregan primers y nucleótidos fluorescentes (terminadores reversibles) en orden (primero A, luego T, etc) así como polimerasa. Cuando el nucleótidos es incorporado usa un laser para identificar que base se incorporó.

# Illumina - sequencing by synthesis - 8



- Se continua este proceso para todas las bases.

# Illumina - sequencing by synthesis - 9



- Las imágenes se analizan espacialmente para revelar cada secuencia.

# Sequencing by Synthesis

## Ventajas

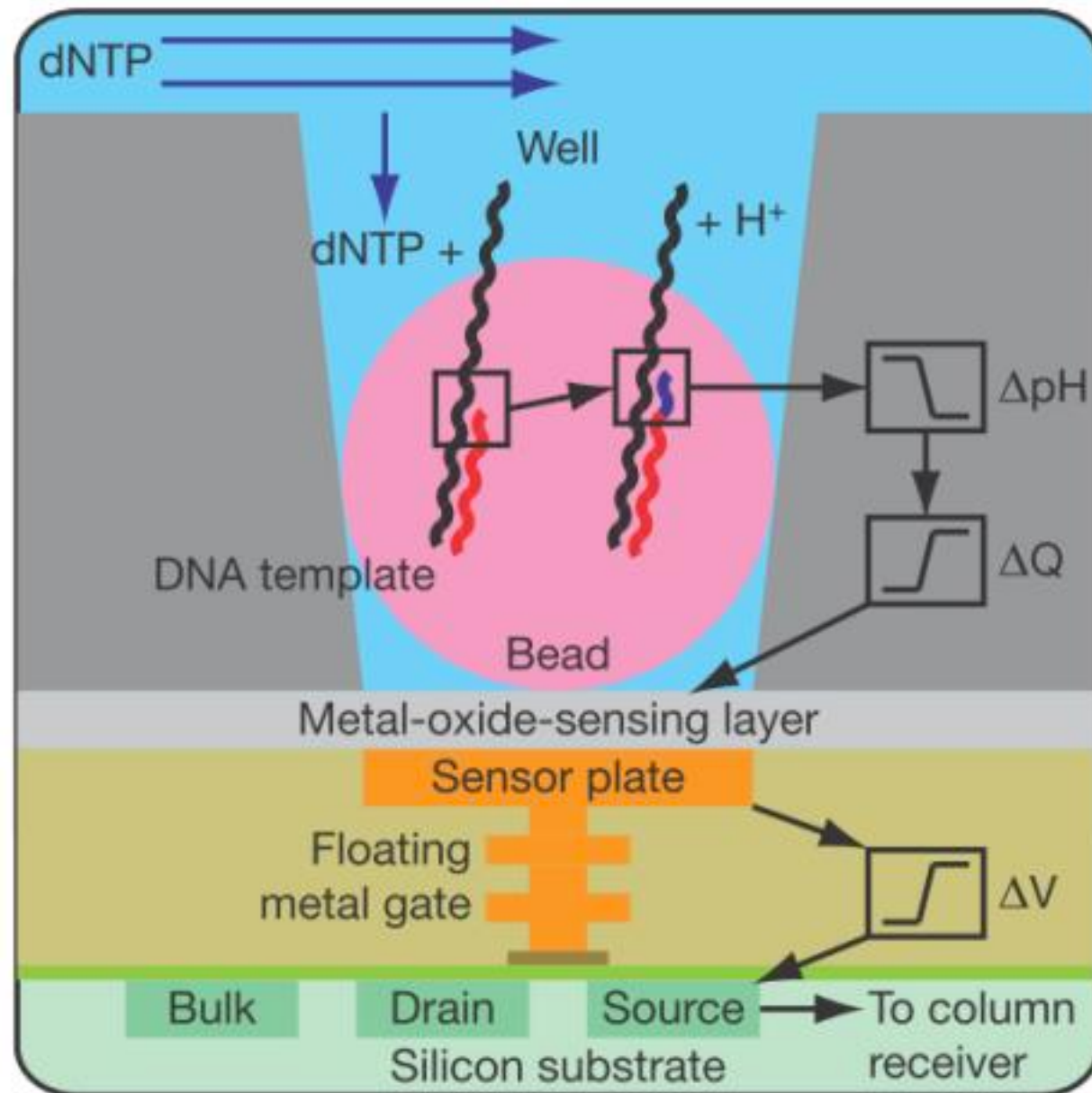
- Indudablemente el líder en el mercado = red de apoyo científico
- Produce grandes cantidades de secuencias
- Baja tasa de error comparada con otras tecnologías

## Desventajas

- Las secuencias son cortas (150 a 250 pb)
- El costo es alto
- Secuenciación relativamente lenta



# ion (pH) sequencing



- Similar a pirosecuenciación
- Se mide el cambio en pH usando un microchip. La liberación de un ión de hidrógeno al incorporarse una base es convertida a voltaje.
- El voltaje se interpreta como el resultado de la incorporación de una o más bases.



# ion (pH) sequencing

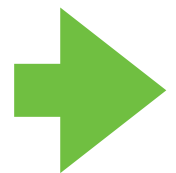
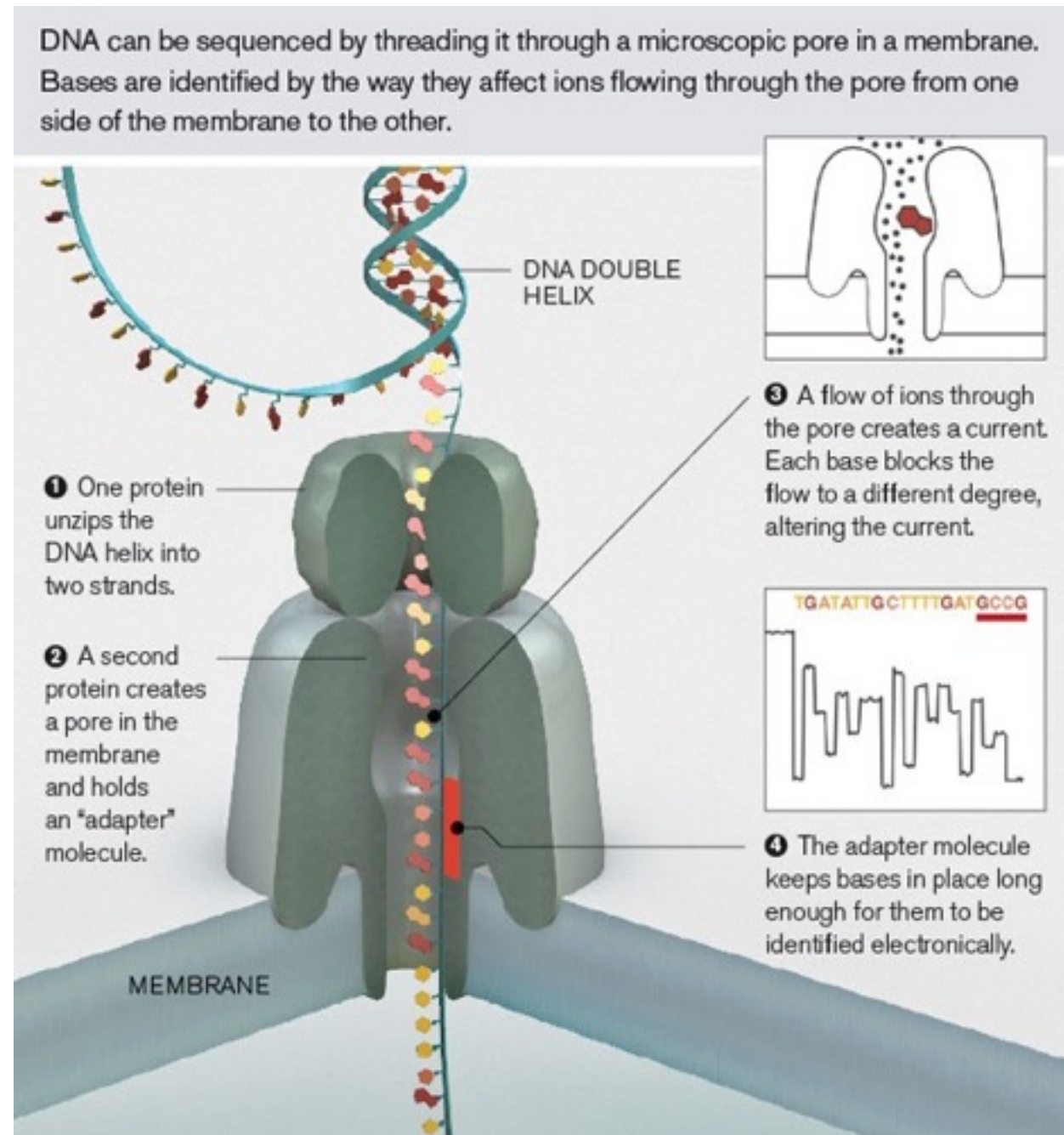
## Ventajas

- Muy barato y rápido
- Las reacciones requeridas son sencillas

## Desventajas

- Alta número de errores en regions con el mismo nucleótido (homopolímeros) e indels
- Secuencias cortas (200 pb)
- Produce pocas secuencias en un solo experimento haciéndolo caro para experimentos grandes

# Nanopore sequencing



# Nanopore sequencing

## Ventajas

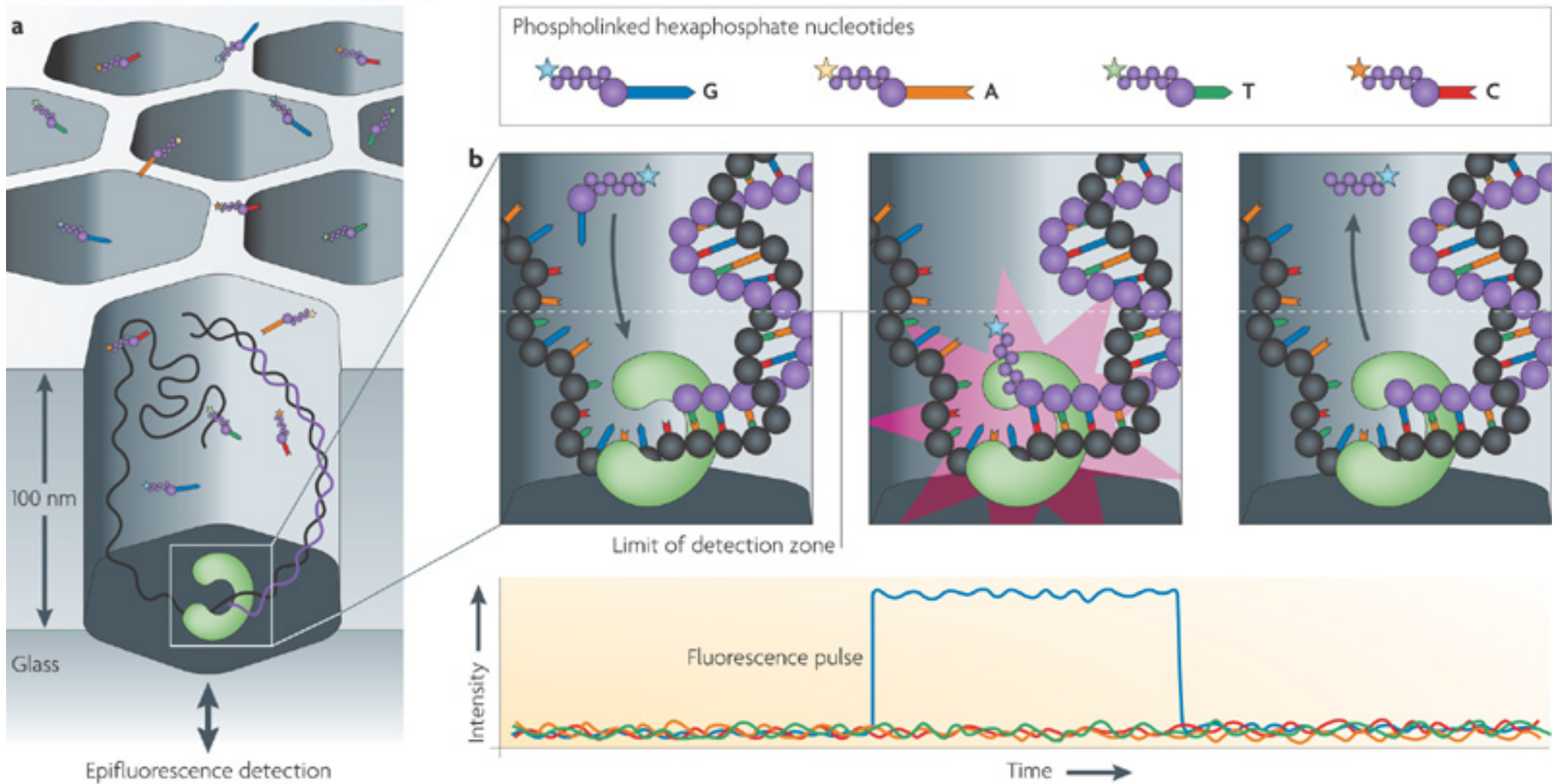
- Secuenciación en tiempo real
- Se puede parar de secuenciar cuando se tienen datos suficientes
- Muy portable - útil para trabajo en zonas difíciles
- Preparación sencilla

## Desventajas

- Alto número de errores aunque han tenido una reducción drástica en el último año
- Poros fallidos - pérdida de secuencia
- Se dice que el costo será mínimo pero sigue siendo caro

# Single Molecule Real Time Sequencing (SMRT)

Pacific Biosciences — Real-time sequencing



# SMRT sequencing

## Ventajas

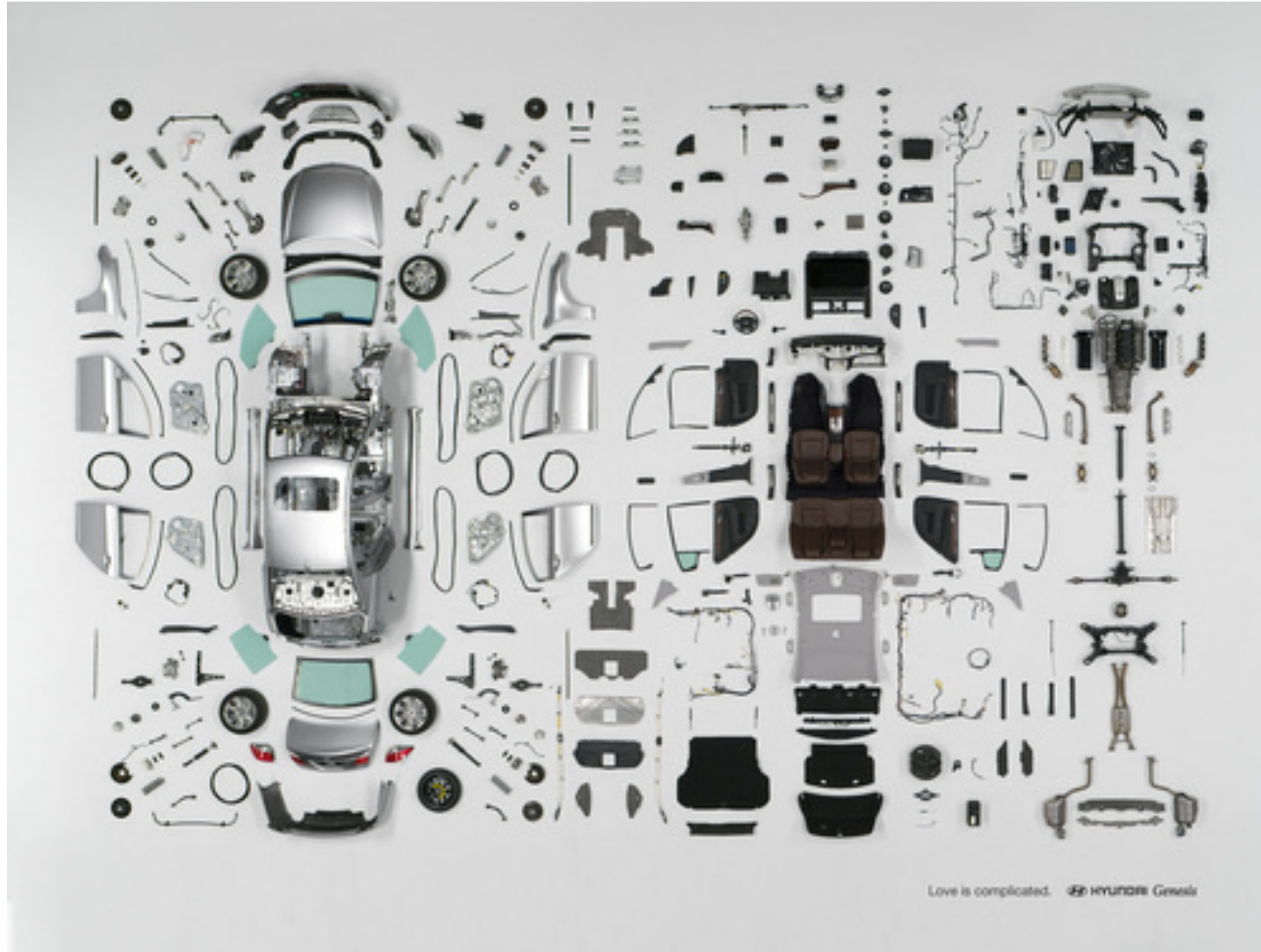
- No requiere amplificación reduciendo errores en abundancia
- Secuencias muy largas (860 - 1100 pb)

## Desventajas

- Alto número de errores como la remoción de bases CG entre varios otros
- Extremadamente caro



# Now what?





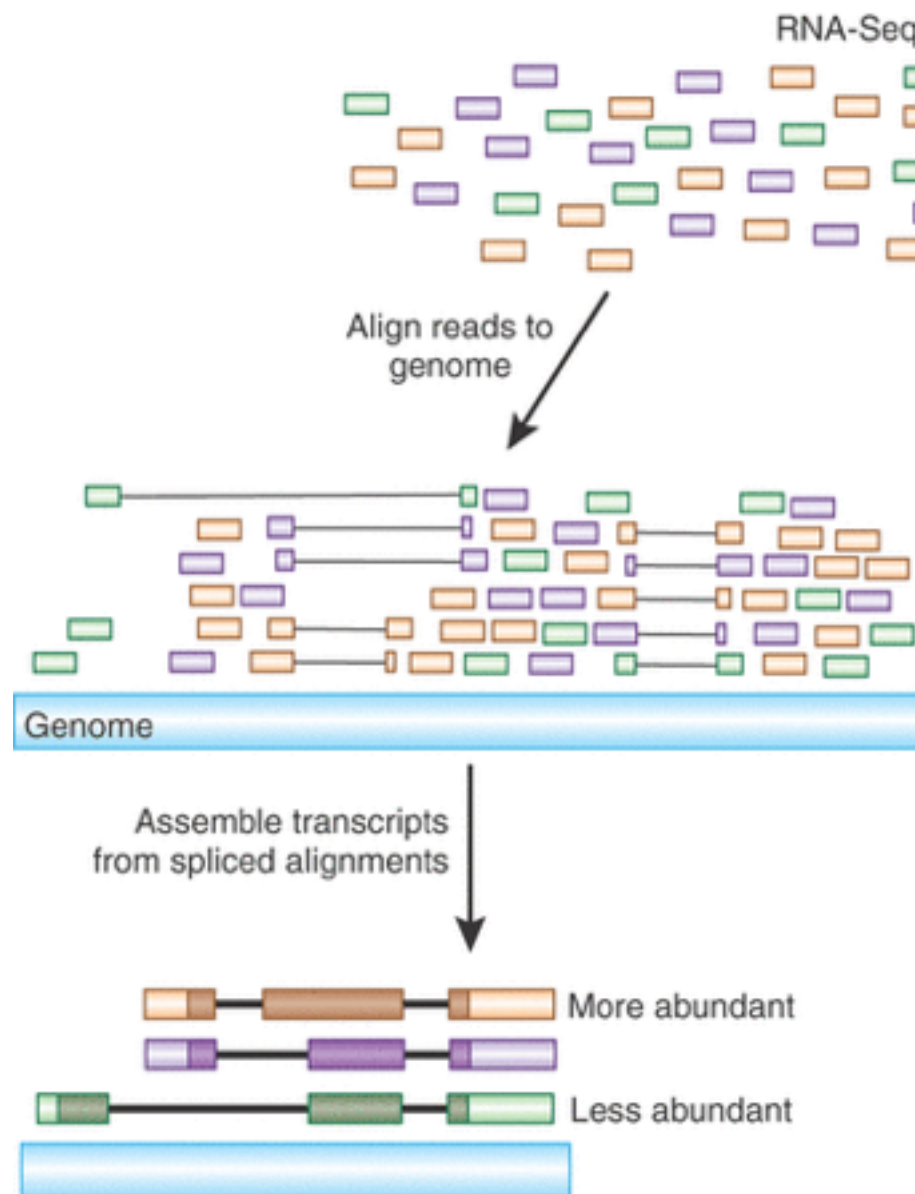
PHASE : INTERPRETATION  
TWO

SEDMAN The New Yorker



# ¿Cómo lo ensamblamos?

- Tophat
- Bowtie
- BWA



- Cufflinks
- Scripture

RNA-Seq reads

Align reads to genome

Genome

Assemble transcripts from spliced alignments

More abundant

Less abundant

- Millions of reads per sample !

Assemble transcripts *de novo*



- Trinity
- Oases de Novo
- trans-ABYSS

Align transcripts to genome



- Blat
- Exonerate