

**Descripción de las técnicas de Next
Generation Sequencing.
Plataformas de secuenciación y su
evolución.
Aplicaciones y algunas recomendaciones
para los estudios.**

4 – 8 julio de 2016. Universidad Rey Juan Carlos, Cursos de Verano

Organizan:



Colaboran:



El DNA

“La doble hélice es una molécula extraordinaria. El hombre moderno existe desde hace 50.000 años y la civilización desde hace 10.000, pero el DNA y el RNA llevan existiendo desde hace millones de años.

Durante todo este tiempo, la doble hélice ha estado ahí, y activa, y todavía somos las primeras criaturas de la Tierra en darnos cuenta de su existencia”

Francis Crick (1916-2004)

El DNA

- *Un poco de historia del DNA...*

La primera persona que aisló el DNA. 1874.

Vida ajetreada, dejó de ir a su propia boda por estar en el laboratorio.

Extrajo “nucleina” en frías noches de invierno.

Protocolo con ácido, alcohol y pepsina de cerdo.

Llegó a las siguientes conclusiones:

- 1- Tiene 4 ácidos básicos
- 2- Alto peso molecular
- 3- “Nucleina” unido a una “protamina”



Friedrich Miescher

El DNA

- *Un poco de historia del DNA...*

Otros científicos involucrados en varias teorías...



August Weismann



Eduard Strasburger

El DNA

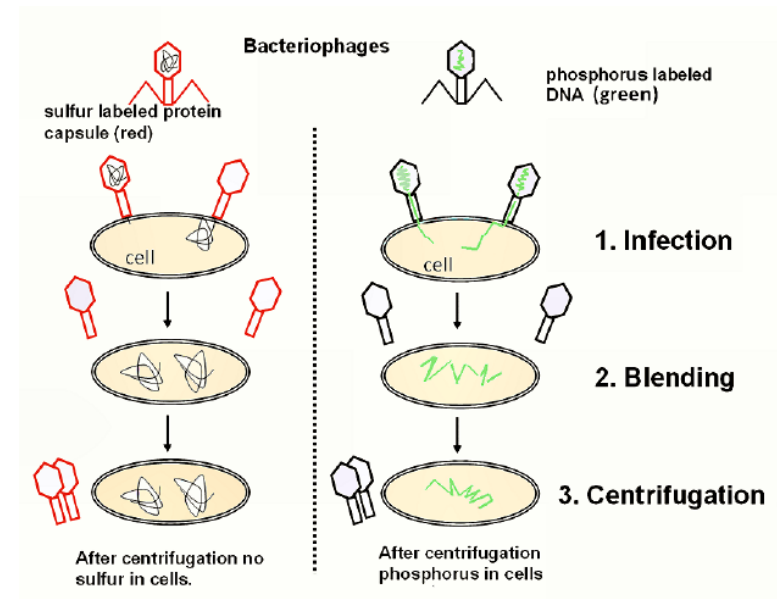
- *Un poco de historia del DNA...*

Hasta 1900, lo que se sabía...

- Proteínas y ácidos nucleicos
- La “nucleina” estaba involucrada en el crecimiento celular
- El núcleo está relacionado con la división celular

A partir de 1950...

- El ratio A/T y G/C es casi 1
- DNA como material genético, experimento Hershey Chase
- 1952, difracción con rayos X



El DNA

- *Un poco de historia del DNA...*

1952 – Watson & Crick (Wilkins & Franklin)

Proponen el B-model del DNA

Aceptado en 1960, premio nobel en 1962 (Franklin murió en 1958)

1958 – Modelo de replicación del DNA

1957 en adelante – Estructura del RNA, ribosomas, tRNA...

1961 – Descifrado del código genético, relación DNA → Proteínas

El DNA **No solo DNA**

- *RNA*
 - Molécula mucho más inestable → Dificultad para trabajar con ella
 - Material de alta calidad
 - Diferencias en términos bioquímicos
 - Distintitos tipos de RNA
 1. RNA mensajero o codificante
 2. RNA no codificante
 3. MicroRNA, funciones reguladoras y de expresión a genomas de virus
 4. piRNA, asociados a transposones

El DNA No solo DNA

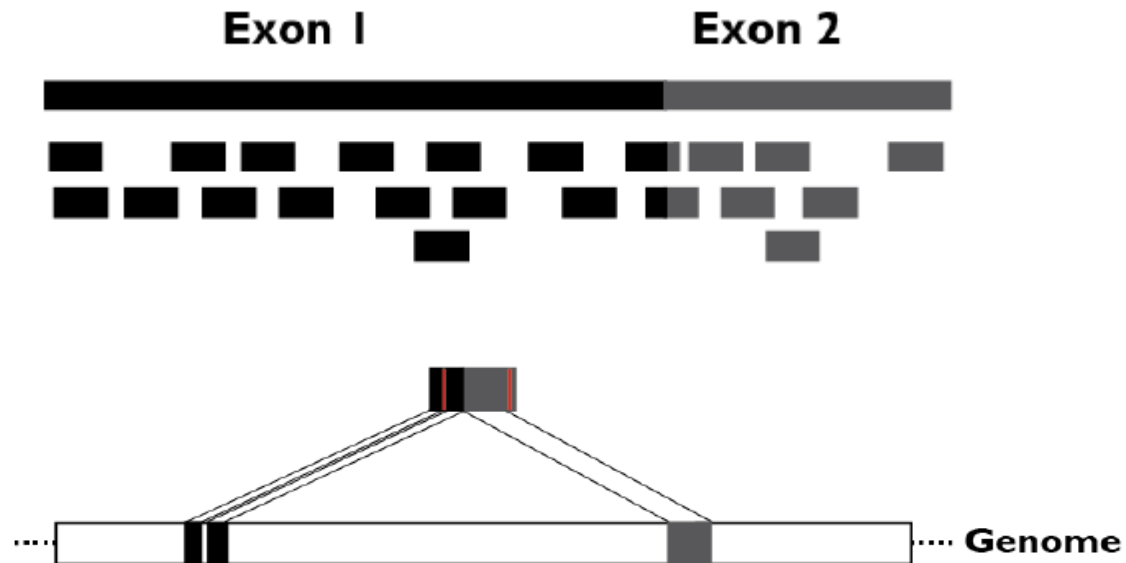
- *RNA, algunos problemas*

Requiere manipulación, lo cual dificulta ciertas aplicaciones y errores

RNA → cDNA → Secuenciación

Artefactos de PCR

Dificultad en el ensamblaje



El DNA No solo DNA

- *RNA, algunos problemas*

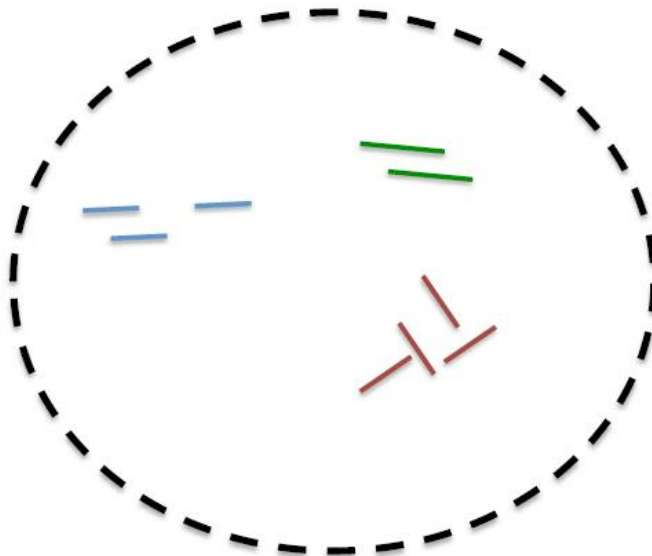
Requiere manipulación, lo cual dificulta ciertas aplicaciones y errores

Artefactos de PCR

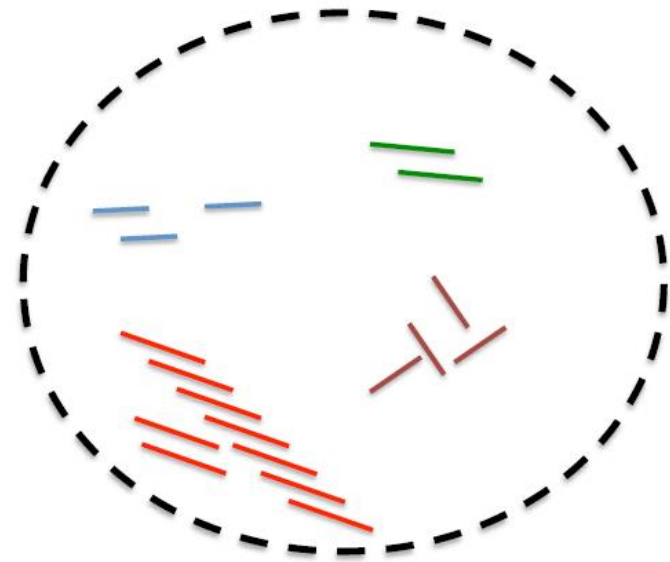
Dificultad en el ensamblaje

Normalización en la expresión

Célula sin estrés



Célula con estrés



El DNA **No solo DNA**

- *RNA, algunas ventajas*

Permite trabajar sin genoma de referencia

Podemos emplearlo tanto para detectar variabilidad como por diferencia de expresión

El DNA No solo DNA Secuenciación

- *Primera tecnología de secuenciación*
 - Conocida como la Secuenciación Sanger
 - Se desarrollo primero para péptidos
 - Se adaptó para RNA y posteriormente para DNA

Proc. Natl. Acad. Sci. USA
Vol. 74, No. 12, pp. 5463-5467, December 1977
Biochemistry

DNA sequencing with chain-terminating inhibitors

(DNA polymerase/nucleotide sequences/bacteriophage ϕ X174)

F. SANGER, S. NICKLEN, AND A. R. COULSON

Medical Research Council Laboratory of Molecular Biology, Cambridge CB2 2QH, England

Contributed by F. Sanger, October 3, 1977

ABSTRACT A new method for determining nucleotide sequences in DNA is described. It is similar to the "plus and minus" method [Sanger, F. & Coulson, A. R. (1975) *J. Mol. Biol.* 94, 441-448] but makes use of the 2',3'-dideoxy and arabinonucleoside analogues of the normal deoxynucleoside triphosphates, which act as specific chain-terminating inhibitors of DNA polymerase. The technique has been applied to the DNA of bacteriophage ϕ X174 and is more rapid and more accurate than either the plus or the minus method.

The "plus and minus" method (1) is a relatively rapid and simple technique that has made possible the determination of the sequence of the genome of bacteriophage ϕ X174 (2). It depends on the use of DNA polymerase to transcribe specific regions of the DNA under controlled conditions. Although the method is considerably more rapid and simple than other

a stereoisomer of ribose in which the 3'-hydroxyl group is oriented in *trans* position with respect to the 2'-hydroxyl group. The arabinosyl (ara) nucleotides act as chain terminating inhibitors of *Escherichia coli* DNA polymerase I in a manner comparable to ddT (4), although synthesized chains ending in 3' araC can be further extended by some mammalian DNA polymerases (5). In order to obtain a suitable pattern of bands from which an extensive sequence can be read it is necessary to have a ratio of terminating triphosphate to normal triphosphate such that only partial incorporation of the terminator occurs. For the dideoxy derivatives this ratio is about 100, and for the arabinosyl derivatives about 5000.

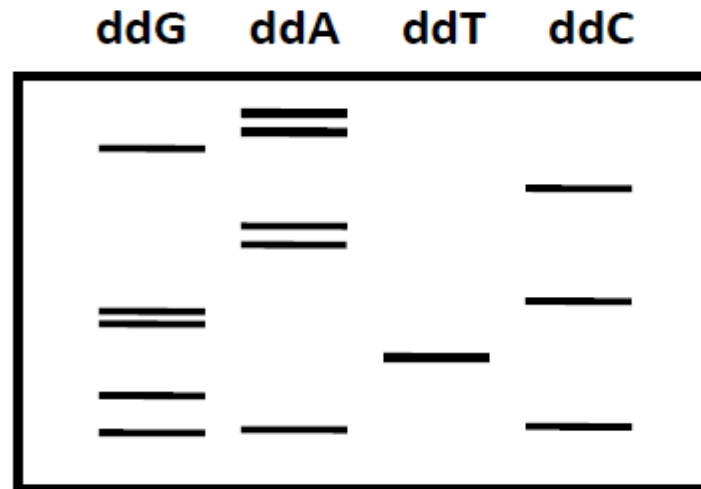
METHODS



Fred Sanger

El DNA No solo DNA **Secuenciación**

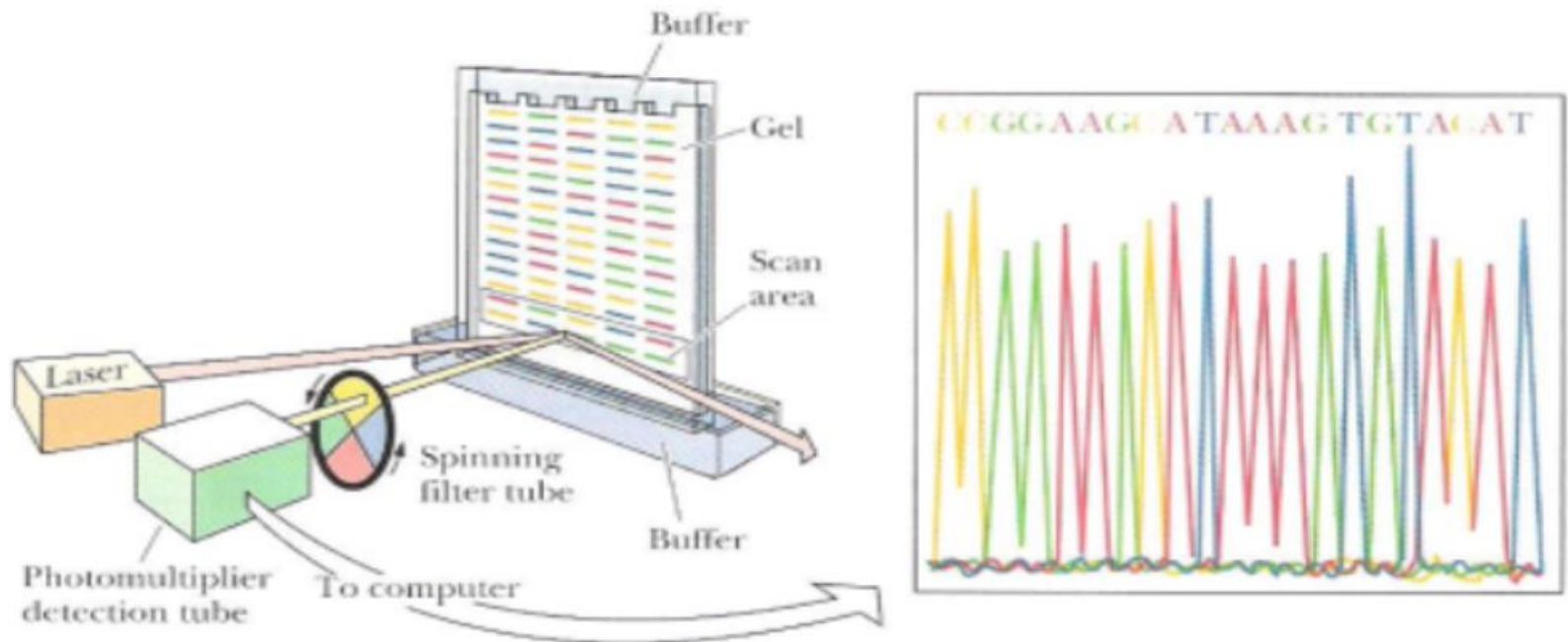
- *Primera tecnología de secuenciación*
 - Reacción de nucleótidos normales (dNTPs) con nucleótidos di-dexoxi (ddNTPs).
 - Cuando se incorpora un ddNTPs, la reacción se para
 - Los ddNTPs están marcados con fluorescencia (al principio radiomarcados)



3'
A
A
G
C
A
A
C
G
T
G
C
A
G
5'

El DNA No solo DNA **Secuenciación**

- *Primera tecnología de secuenciación*
 - Reacción de nucleótidos normales (dNTPs) con nucleótidos di-dexoxi (ddNTPs).
 - Cuando se incorpora un ddNTPs, la reacción se para
 - Los ddNTPs estan marcados con fluorescencia (al principio radiomarcados)

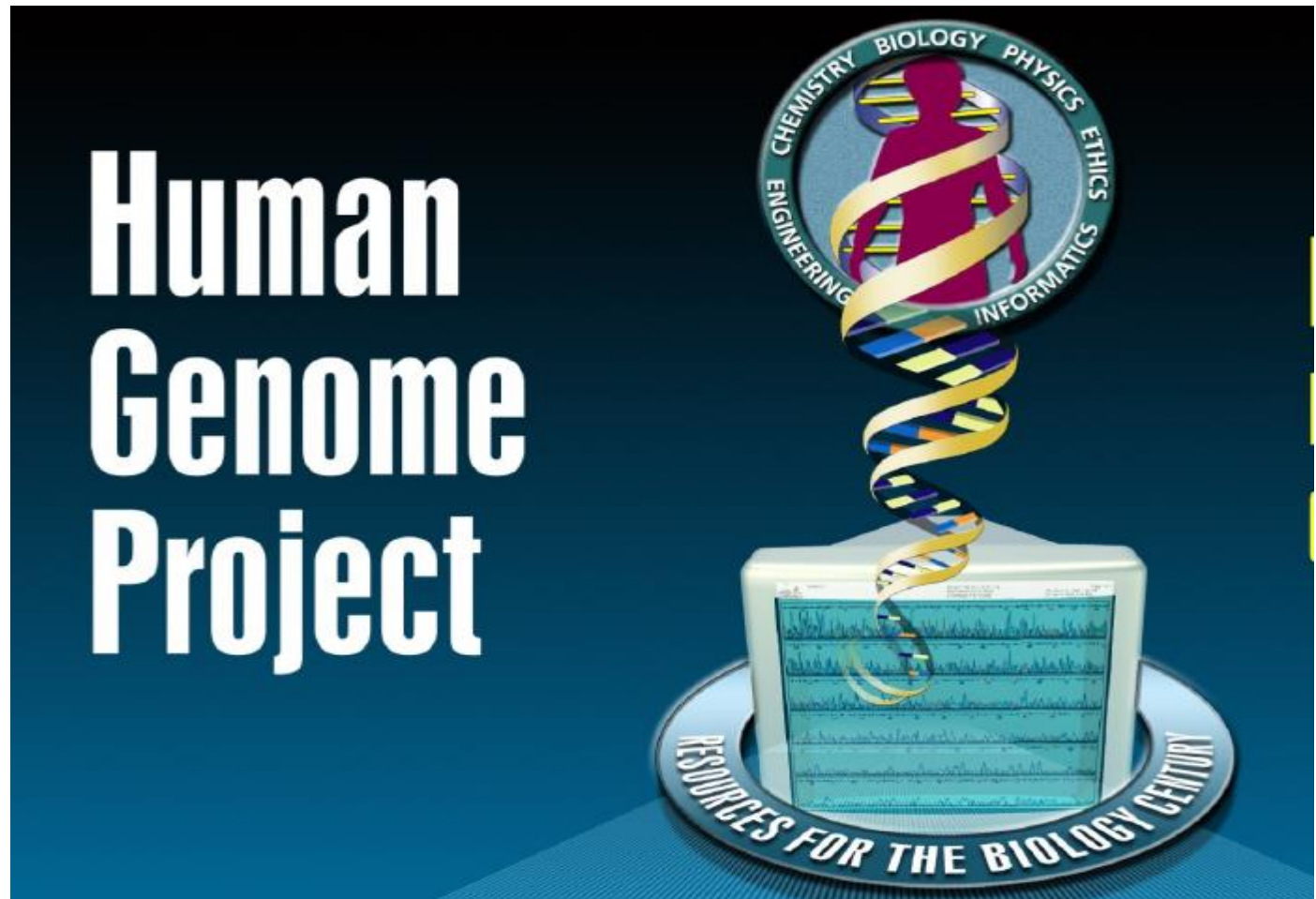


- *Primera tecnología de secuenciación*

TECNOLOGÍA	SANGER
Potencia	50-100 kb, 96 secuencias por run
Longitud de la lectura	0.5-2 kbp
Precisión	99%
Precio por base	200.00 \$/Gb

El DNA No solo DNA **Secuenciación**

- *Primera tecnología de secuenciación*
- Proyecto Genoma Humano



El DNA No solo DNA **Secuenciación**

- *Primera tecnología de secuenciación*
 - Proyecto Genoma Humano, algunos datos:
 - Tecnología Sanger
 - 3 Billones \$ en 15 años
 - Borrador del genoma en 2000, genoma completo en 2003



- *Primera tecnología de secuenciación*
 - Proyecto Genoma Humano, algunos datos
 - Tecnología Sanger
 - 3 Billones \$ en 15 años
 - Borrador del genoma en 2000, genoma completo en 2003
 - Acceso libre
 - Ha dado lugar a varios proyectos (Encode, 1000 Genomes, etc), otros genomas (completos y parciales)
 - Impacto económico calculado en 796 billones \$, cientos de miles de puestos de trabajo

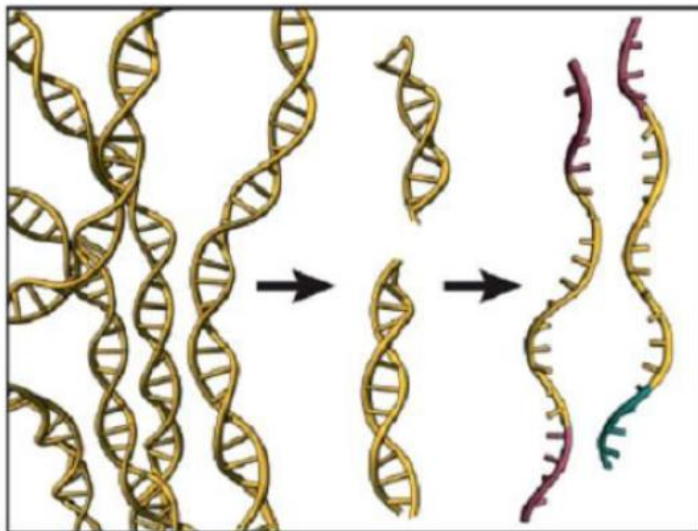
- *Tecnologías de Next Generation Sequencing*
 - Primera generación, **454**
 - También conocido como Pirosecuencing o Roche Sequencing
 - Aparece a finales de los 90s
 - Comprado por Roche en 2007
 - Prácticamente en desuso salvo aplicaciones puntuales



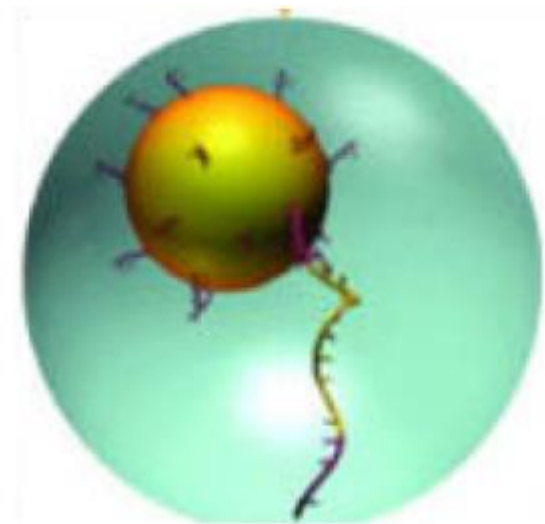
- *Tecnologías de Next Generation Sequencing*

- Primera generación, **454**

- También conocido como Pirosecuencing o Roche Secuencing



1. Extracción de DNA
2. Ruptura en fragmentos
3. Ligación de adaptadores (librería)

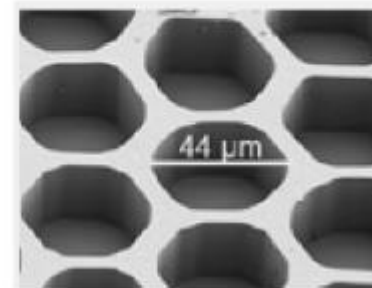
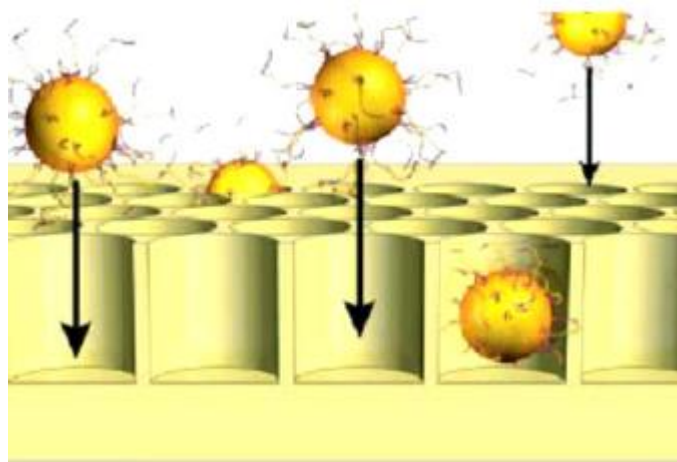


4. Cada cadena de DNA se pega en un bead con adaptadores compatibles

- *Tecnologías de Next Generation Sequencing*

- Primera generación, **454**

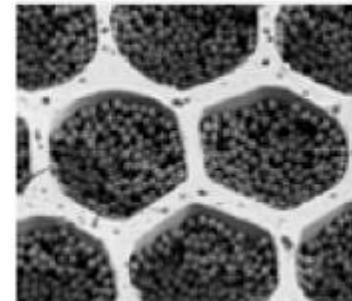
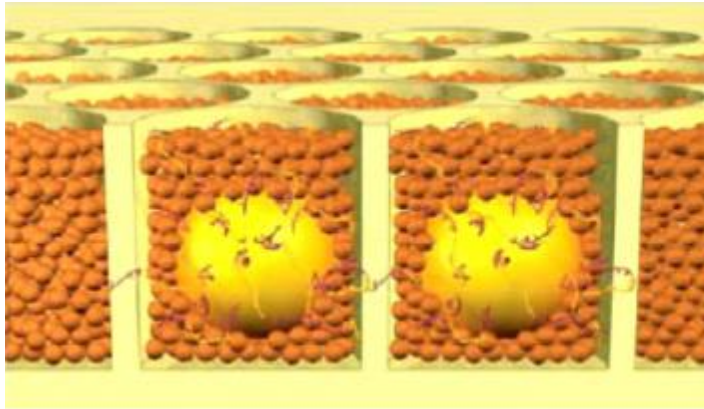
- También conocido como Pirosecuencing o Roche Secuencing



5. Las Beads se meten en una placa específica, de tal manera que entra una bead en cada orificio

El DNA No solo DNA **Secuenciación**

- *Tecnologías de Next Generation Sequencing*
 - Primera generación, **454**
 - También conocido como Pirosecuencing o Roche Sequencing



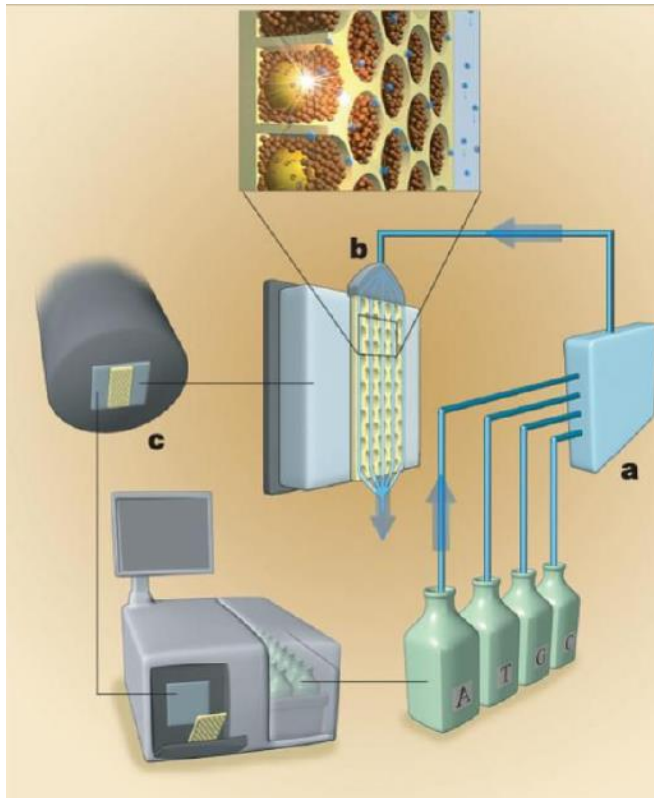
6. En cada orificio se incorporan enzimas específicas y reactivos para realizar la pirosecuenciación (sulfurylasa y luciferasa)

El DNA No solo DNA **Secuenciación**

- *Tecnologías de Next Generation Sequencing*

- Primera generación, **454**

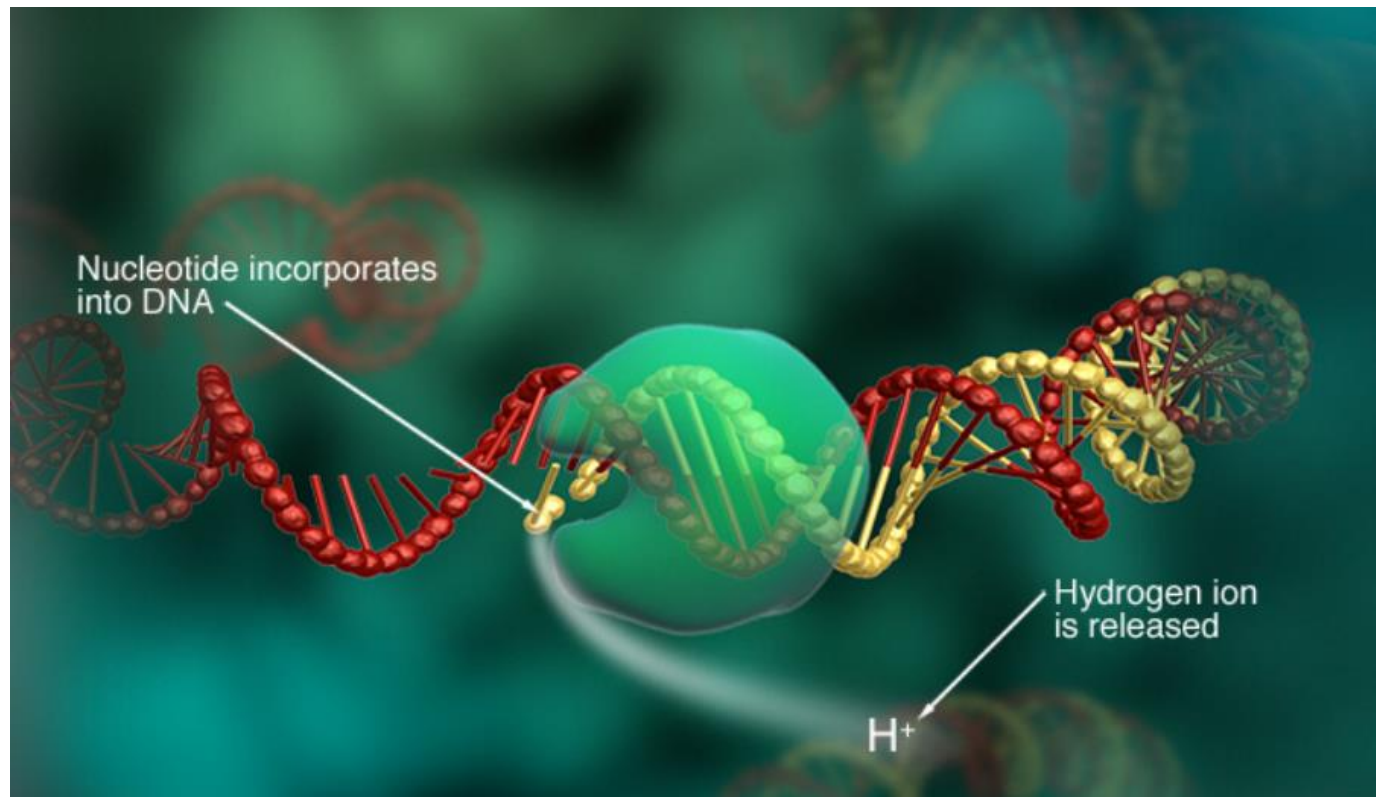
- También conocido como Pirosecuencing o Roche Sequencing



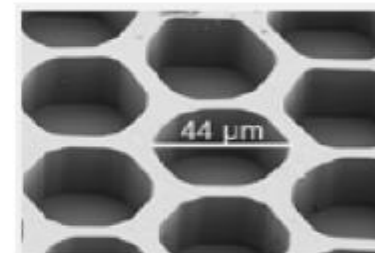
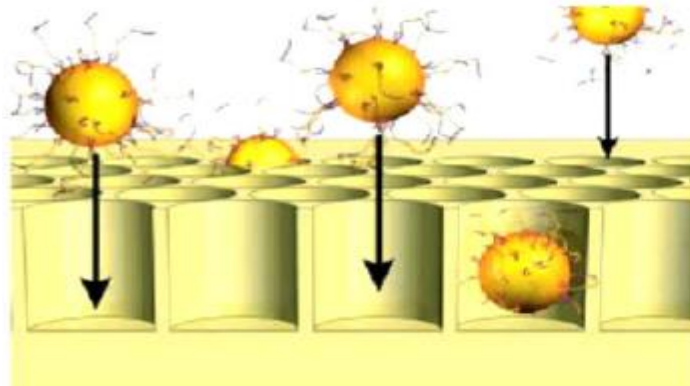
7. a. Se bombean nucleótidos por la placa de orificios. b. Cuando se incorpora un nucleótido, se emite luz y c. el lector recoge la señal de luz

El DNA No solo DNA **Secuenciación**

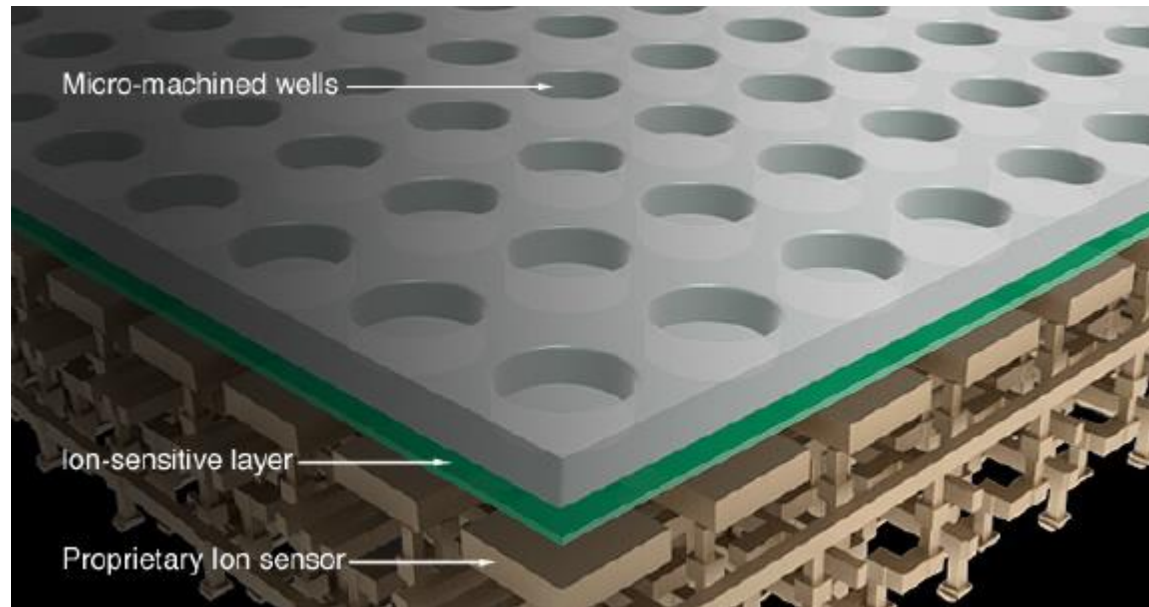
- *Tecnologías de Next Generation Sequencing*
 - Primera generación, **Ion Proton – Ion Torrent**
 - Evolución del 454
 - Desarrollado 2007 y comprado por Life Technologies en 2010



- *Tecnologías de Next Generation Sequencing*
 - Primera generación, **Ion Proton – Ion Torrent**
 - Evolución del 454
 - Desarrollado 2007 y comprado por Life Technologies en 2010
 - Usado en aplicaciones concretas



- *Tecnologías de Next Generation Sequencing*
 - Primera generación, **Ion Proton – Ion Torrent**
 - Evolución del 454
 - Desarrollado 2007 y comprado por Life Technologies en 2010



- *Tecnologías de Next Generation Sequencing*
 - Primera generación, **454 - Ion Proton – Ion Torrent**

TECNOLOGÍA	SANGER	454 - Ion
Potencia	50-100 kb, 96 secuencias/run	1 Gb /run
Longitud de la lectura	0.5-2 kbp	200-400 bp
Precisión	99%	98%
Precio por base	200.00 \$/Gb	700 \$/Gb

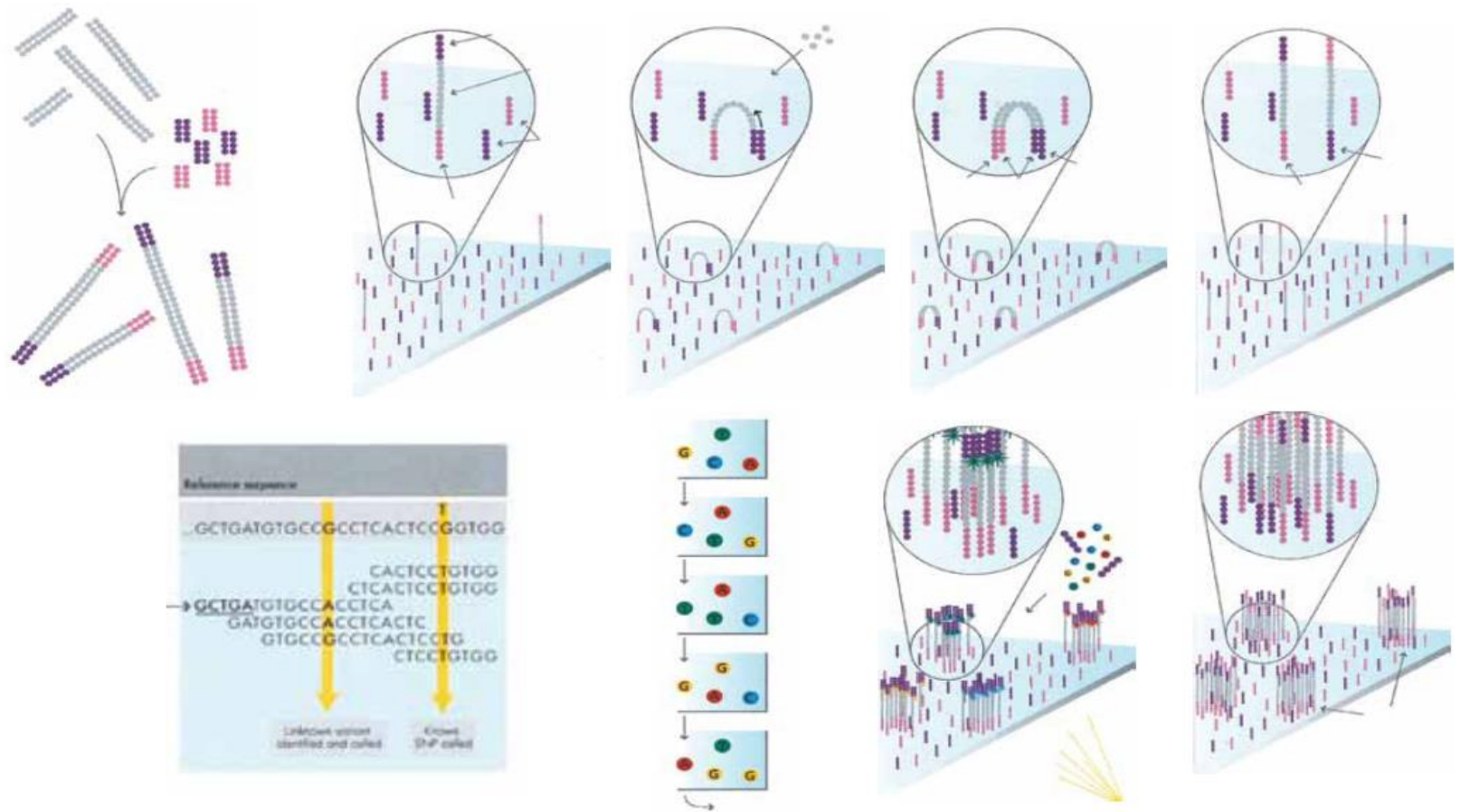
El DNA No solo DNA **Secuenciación**

- *Tecnologías de Next Generation Sequencing*
 - Segunda generación, **Illumina seqs**
 - Actualmente es la plataforma de secuenciación más usada (70% del mercado)
 - Flexible en varios términos y tipos de máquinas



El DNA No solo DNA **Secuenciación**

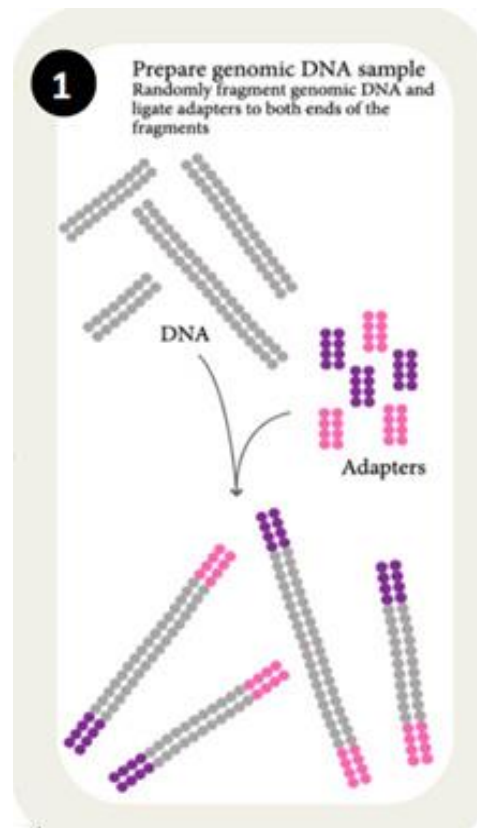
- *Tecnologías de Next Generation Sequencing*
 - Segunda generación, **Illumina seqs**



El DNA No solo DNA **Secuenciación**

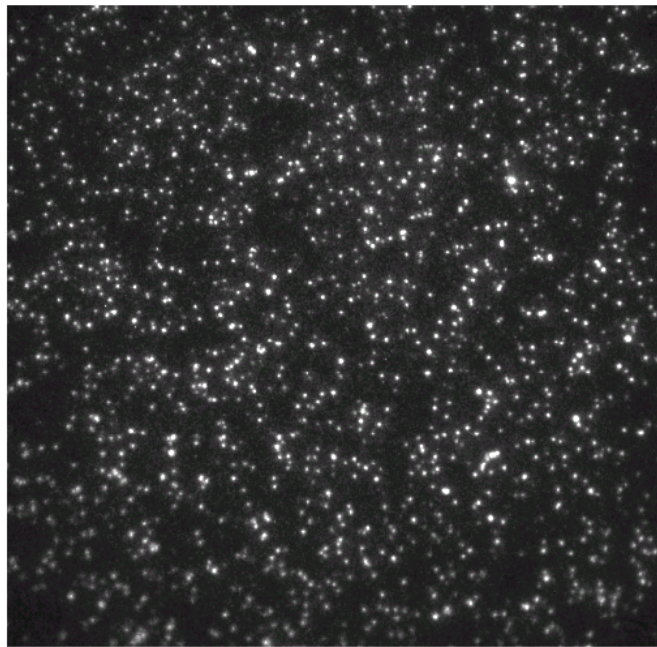
- *Tecnologías de Next Generation Sequencing*
 - Segunda generación, **Illumina seqs**

1. Preparación del material genómico



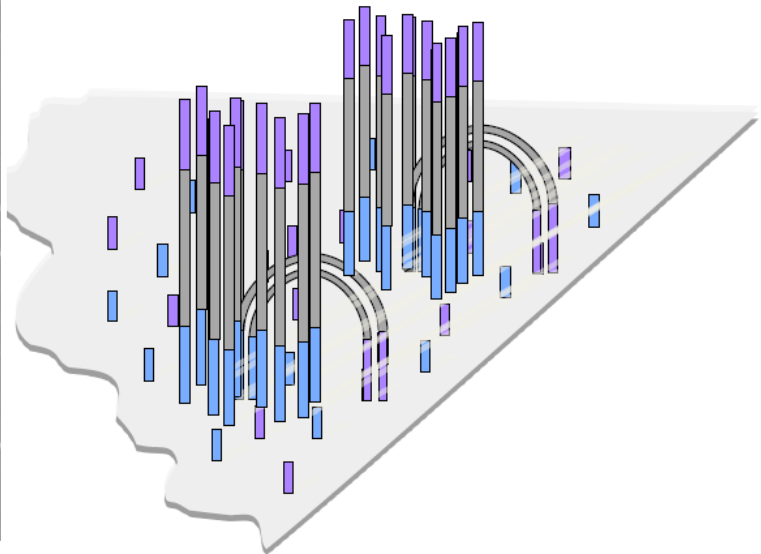
- *Tecnologías de Next Generation Sequencing*
 - Segunda generación, **Illumina seqs**

2. Ligación a la célula y formación de clusters



100um

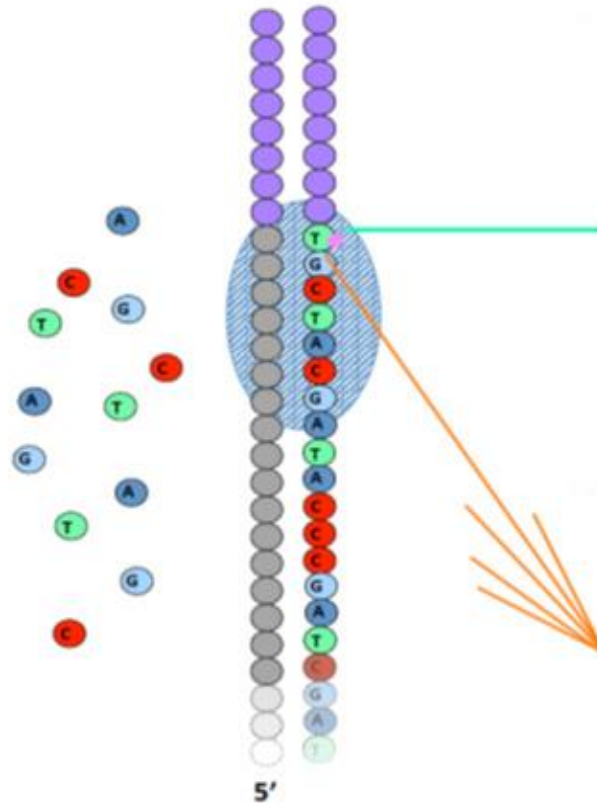
Attach single molecules to surface
Amplify to form clusters



~1000 molecules per ~ 1 um cluster
~2 billion clusters per flowcell

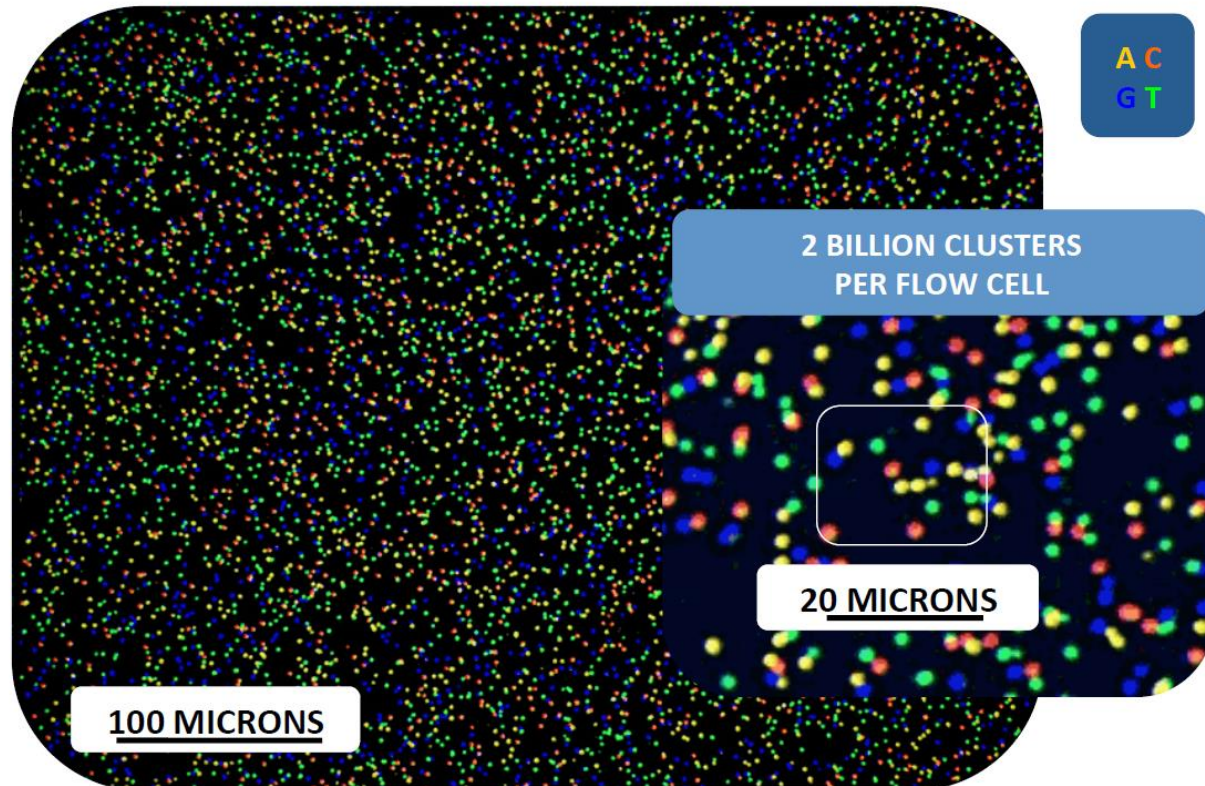
- *Tecnologías de Next Generation Sequencing*
 - Segunda generación, **Illumina seqs**

3. **Secuenciación por síntesis**



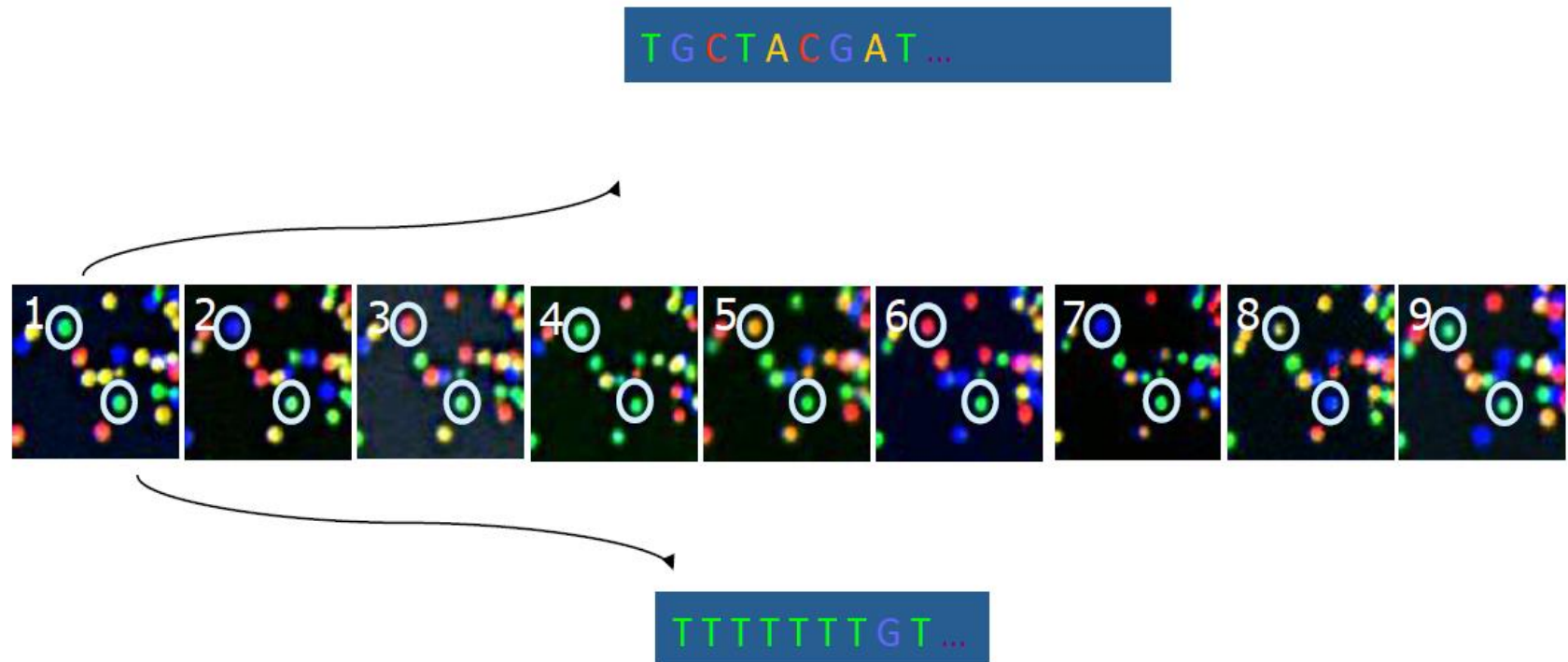
- *Tecnologías de Next Generation Sequencing*
 - Segunda generación, **Illumina seqs**

4. Lectura con laser



- *Tecnologías de Next Generation Sequencing*
 - Segunda generación, **Illumina seqs**

4. Lectura con laser



- *Tecnologías de Next Generation Sequencing*
- Segunda generación, **Illumina seqs**, algunas ventajas
 - Permite lecturas pareadas (**Paired-ends**) frente a las sencillas (**Single-ends**)
 - Lecturas opuestas del mismo fragmento



Puedo conocer distancia entre fragmentos, según la biblioteca

Ayuda en el alineamiento

Más coste y más tiempo

El DNA No solo DNA **Secuenciación**

- *Tecnologías de Next Generation Sequencing*
 - Segunda generación, **Illumina seqs**

TECNOLOGÍA	SANGER	454 - Ion	Illumina*
Potencia	50-100 kb, 96 secuencias/run	1 Gb /run	2 – 600 Gb/run
Longitud de la lectura	0.5-2 kbp	200-400 bp	Hasta 300 bp
Precisión	99%	98%	99%
Precio por base	200.00 \$/Gb	700 \$/Gb	150\$/Gb (y bajando)

* Depende de la plataforma

El DNA No solo DNA **Secuenciación**

- *Tecnologías de Next Generation Sequencing*
 - Segunda generación, **Illumina seqs**



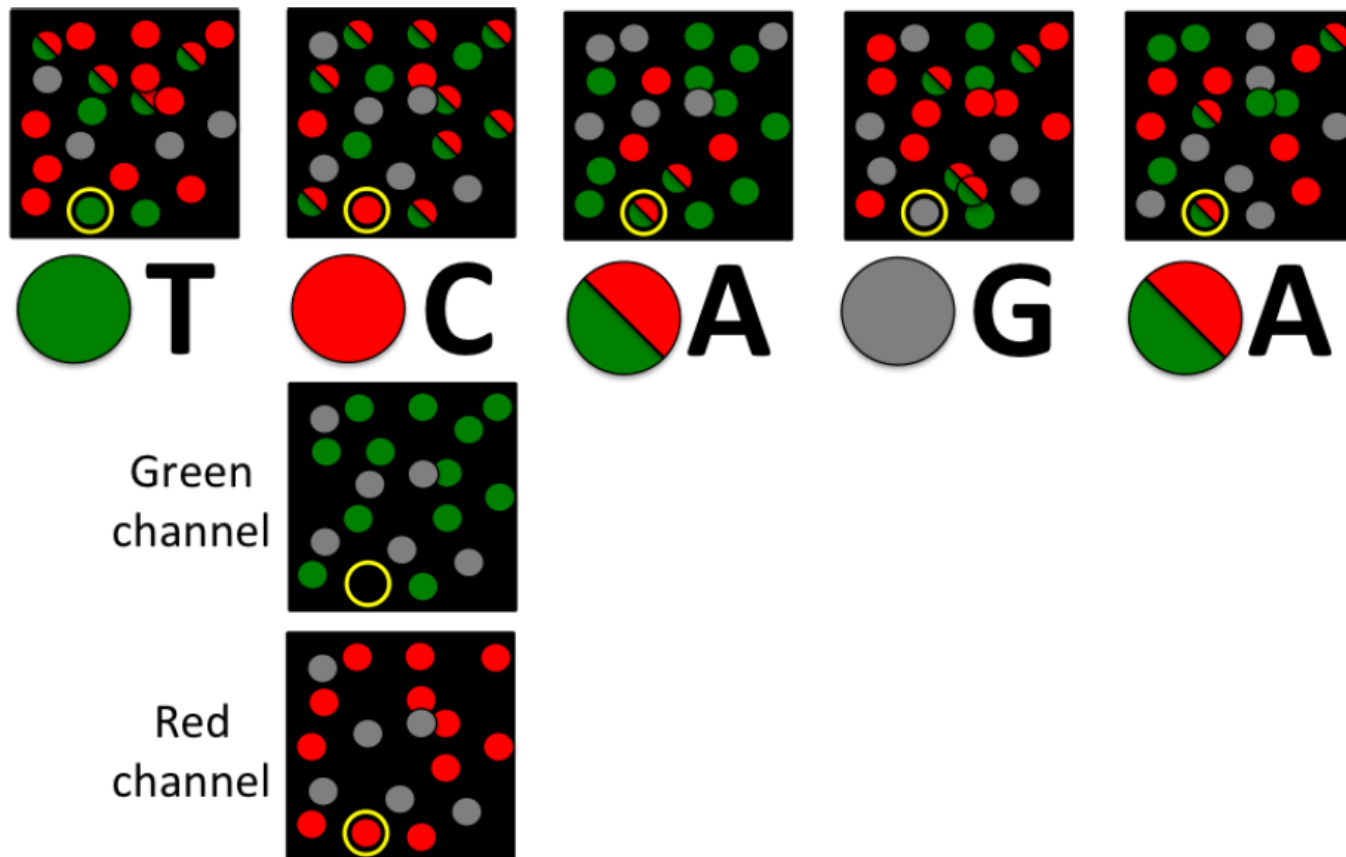
Illumina Miniseq

Más barato y flexible

Química de dos colores

El DNA No solo DNA **Secuenciación**

- *Tecnologías de Next Generation Sequencing*
 - Segunda generación, **Illumina seqs**



El DNA No solo DNA **Secuenciación**

- *Tecnologías de Next Generation Sequencing*
- Tercera generación, **PacBio RS II**



- *Tecnologías de Next Generation Sequencing*
- Tercera generación, **PacBio RS II**
 - Pacific Biosystems
 - En vez de monitorizar los nucleótidos, monitorizamos la polimerasa. Secuenciación SMRT (single molecule real time)
 - La polimerasa está confinada en la ZMW (Guía de onda “modo cero”)
 - El nucleótido tiene un fluorocromo, cuando se une, se libera y se detecta con el laser

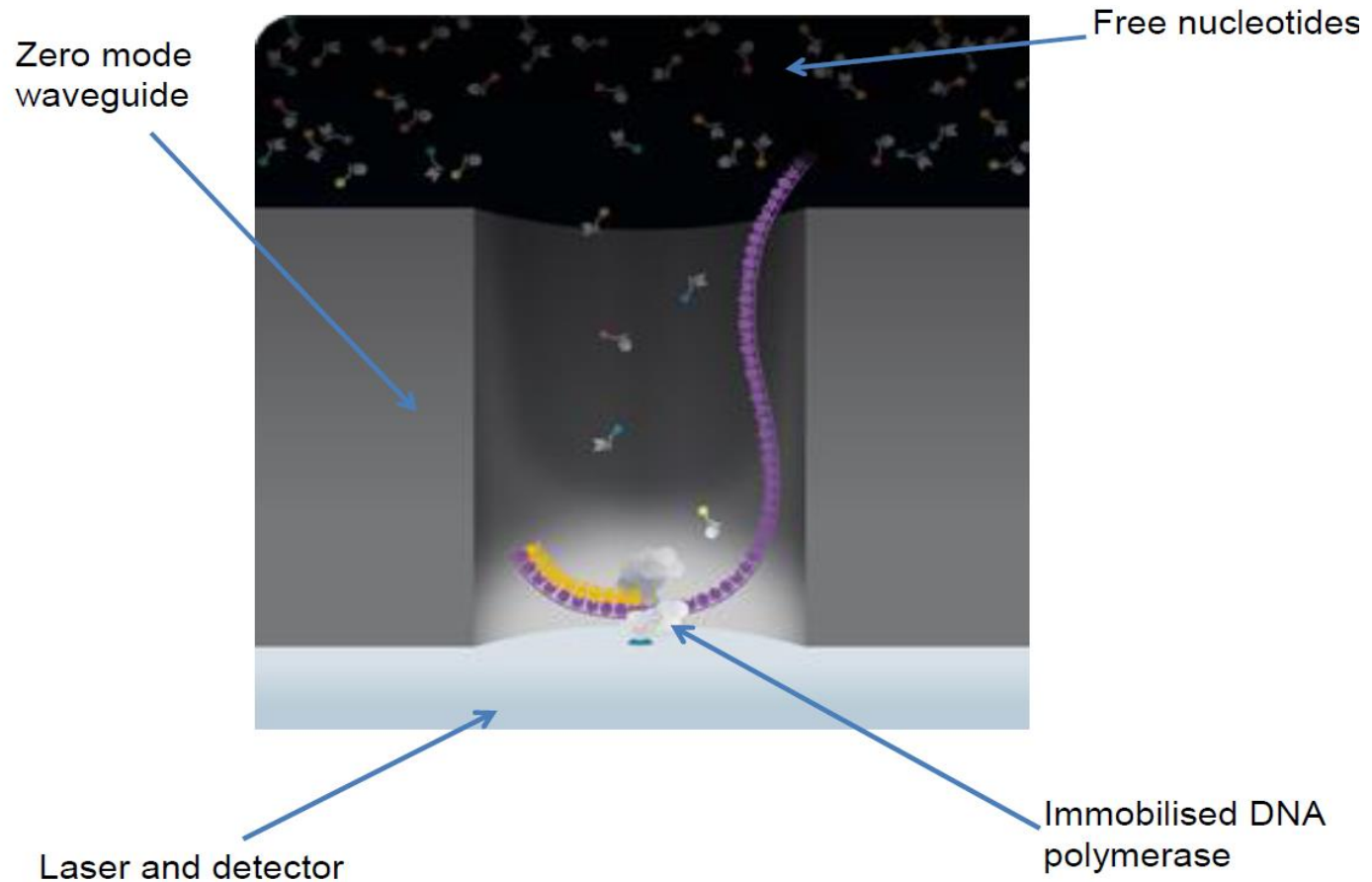
El DNA No solo DNA **Secuenciación**

- *Tecnologías de Next Generation Sequencing*
- Tercera generación, **PacBio RS II**



El DNA No solo DNA **Secuenciación**

- *Tecnologías de Next Generation Sequencing*
- Tercera generación, **PacBio RS II**



El DNA No solo DNA **Secuenciación**

- *Tecnologías de Next Generation Sequencing*
- Tercera generación, **PacBio RS II**

TECNOLOGÍA	SANGER	454 - Ion	Illumina*	PacBio
Potencia	50-100 kb, 96 secuencias/run	1 Gb /run	2 – 600 Gb/run	1 Gb/run
Longitud de la lectura	0.5-2 kbp	200-400 bp	Hasta 300 bp	12-40 kbp
Precisión	99%	98%	99%	90%
Precio por base	200.00 \$/Gb	700 \$/Gb	150\$/Gb (y bajando)	200\$/Gb

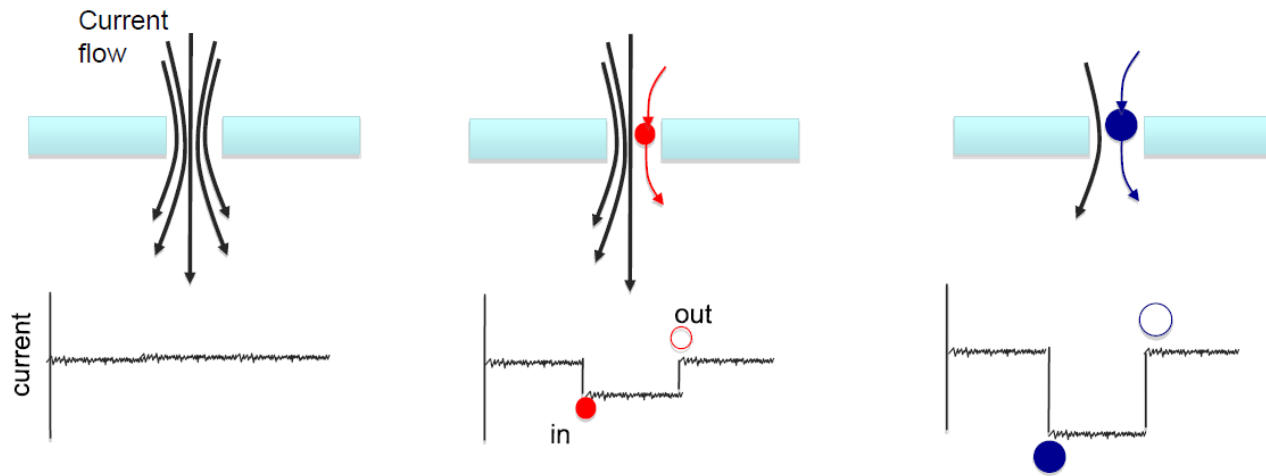
El DNA No solo DNA **Secuenciación**

- *Tecnologías de Next Generation Sequencing*
- Tercera generación, **Nanopore**
 - Varias compañías trabajando



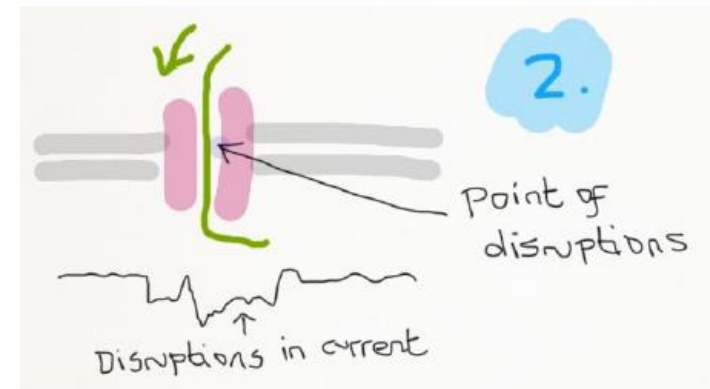
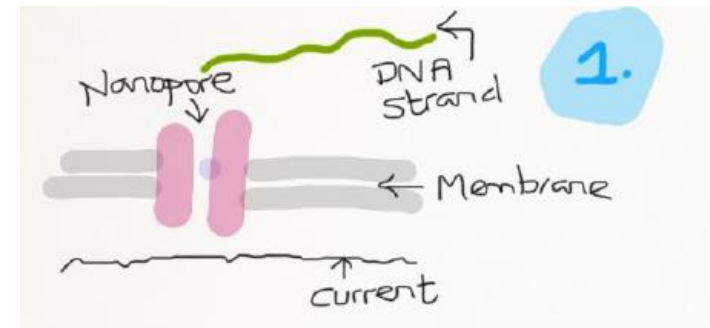
El DNA No solo DNA **Secuenciación**

- *Tecnologías de Next Generation Sequencing*
- Tercera generación, **Nanopore**
 - Varias compañías trabajando
 - Nanopore = Agujero pequeño
 - Flujo eléctrico por el agujero
 - Si algo se cuela por el agujero, bloquea el flujo eléctrico



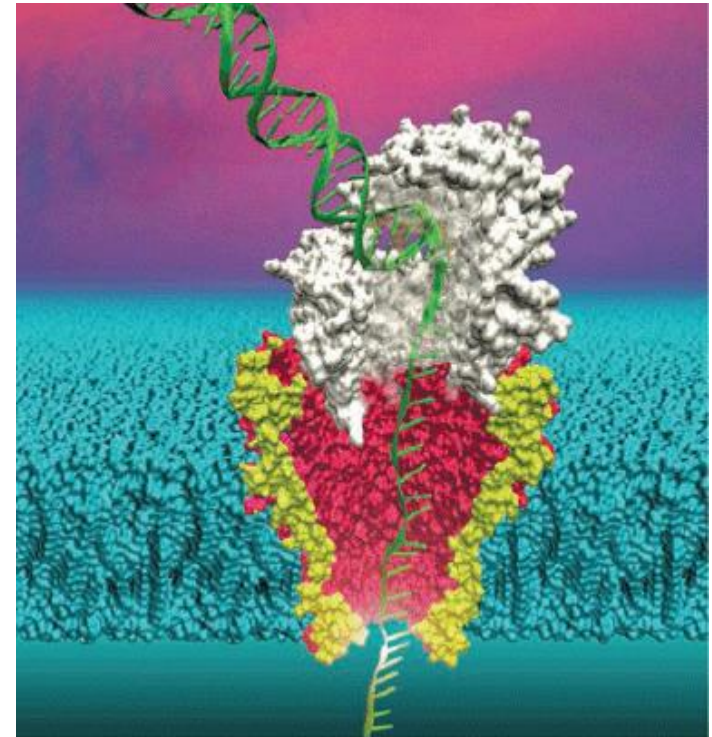
El DNA No solo DNA **Secuenciación**

- *Tecnologías de Next Generation Sequencing*
- Tercera generación, **Nanopore**
 - Varias compañías trabajando
 - Nanopore = Agujero pequeño
 - Flujo eléctrico por el agujero
 - Si algo se cuela por el agujero, bloquea el flujo eléctrico
 - Hay poros en membranas biológicas
 - Poros sintéticos o artificiales
 - Pasar una molécula de 4nm por un agujero de 5nm



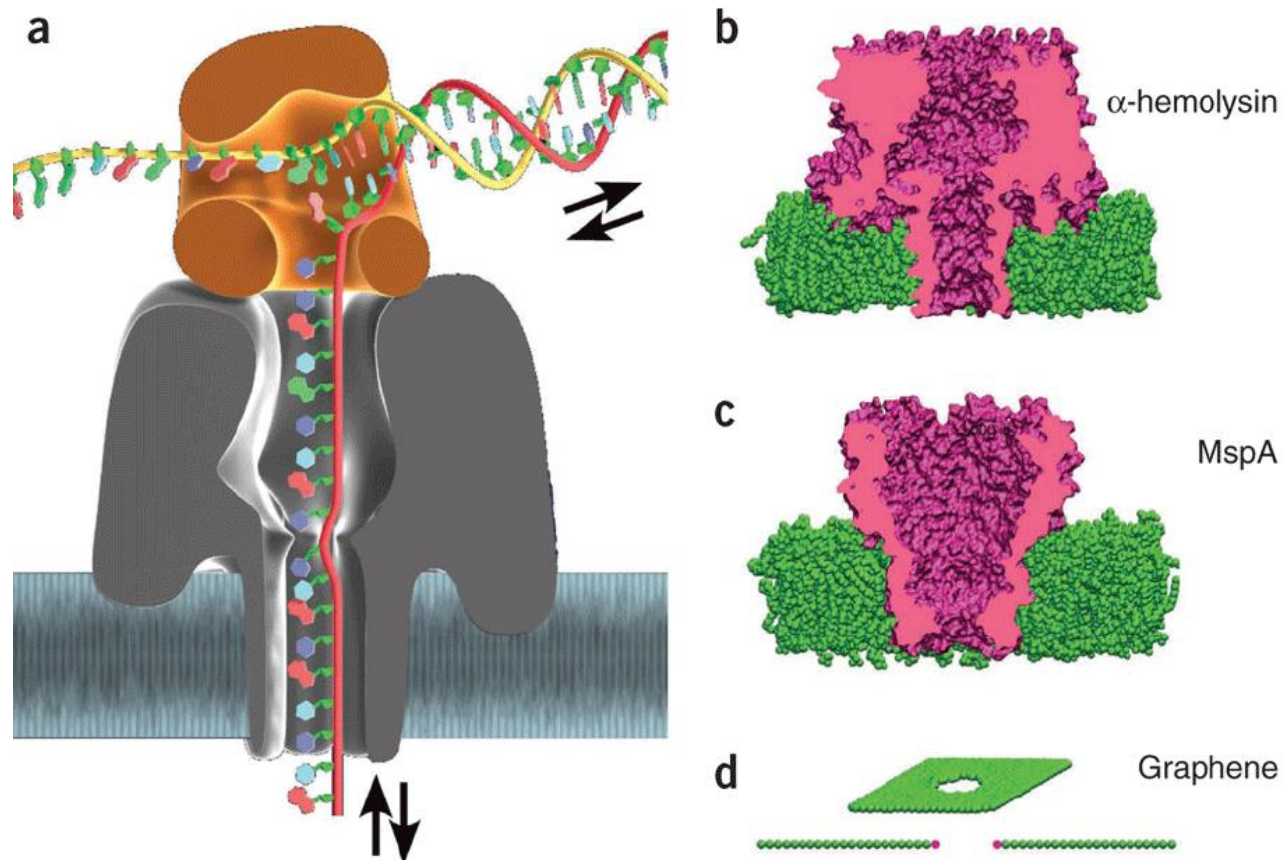
El DNA No solo DNA **Secuenciación**

- *Tecnologías de Next Generation Sequencing*
- Tercera generación, **Nanopore**
 - Más complicado en la práctica
 - El DNA se mueve muy rápido
 - Necesidad de enzimas
 - Cualquier cosa puede interferir en el flujo eléctrico
 - Se necesita distinguir las bases con algún tipo de etiqueta



El DNA No solo DNA **Secuenciación**

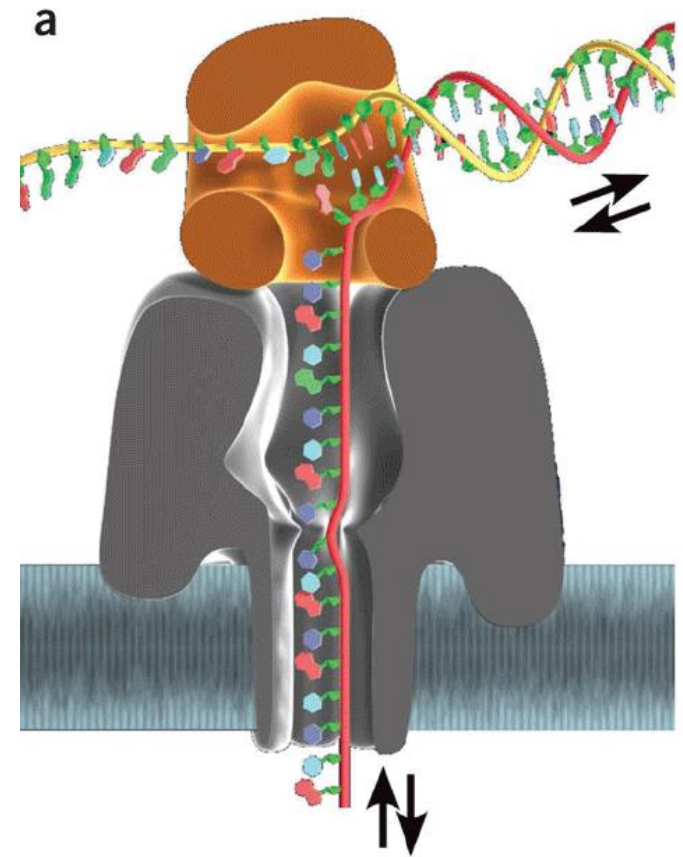
- *Tecnologías de Next Generation Sequencing*
- Tercera generación, **Nanopore**



<http://www.nature.com/nbt/journal/v30/n4/full/nbt.2181.html>

El DNA No solo DNA **Secuenciación**

- *Tecnologías de Next Generation Sequencing*
- Tercera generación, **Nanopore**
 - α -hemolysin: cortar el DNA y añadir la etiqueta
 - MspA: relentizar el paso del DNA por el poro y disociar la etiqueta que tenga



El DNA No solo DNA **Secuenciación**

- *Tecnologías de Next Generation Sequencing*
- Tercera generación, **Nanopore, Oxford Nanopore**



- MinION
- 1-2 Gbases/run



- PromethION
- 20-30 Gbases/run

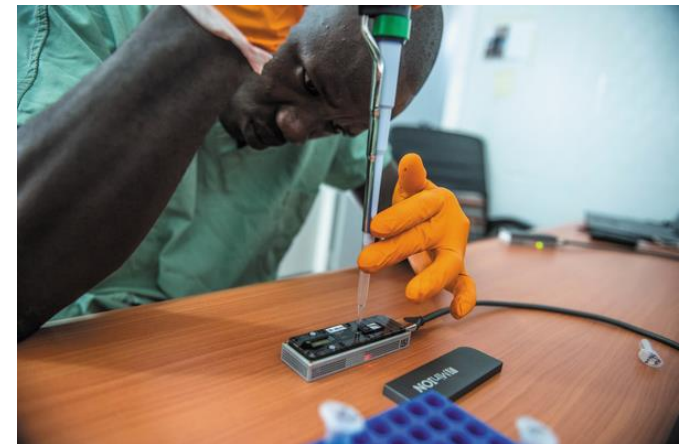
El DNA No solo DNA **Secuenciación**

- *Tecnologías de Next Generation Sequencing*
- Tercera generación, **Nanopore, Oxford Nanopore**
 - MinION
 - Necesita preparación de librerías
 - En fase de pruebas
 - Genoma de *E.coli* por parte de varios laboratorios para su puesta a punto



El DNA No solo DNA **Secuenciación**

- *Tecnologías de Next Generation Sequencing*
- Tercera generación, **Nanopore, Oxford Nanopore**
 - MinION
 - Necesita preparación de librerías
 - En fase de pruebas
 - Genoma de *E.coli* por parte de varios laboratorios para su puesta a punto
 - Secuenciadores portátiles en el último brote de ébola



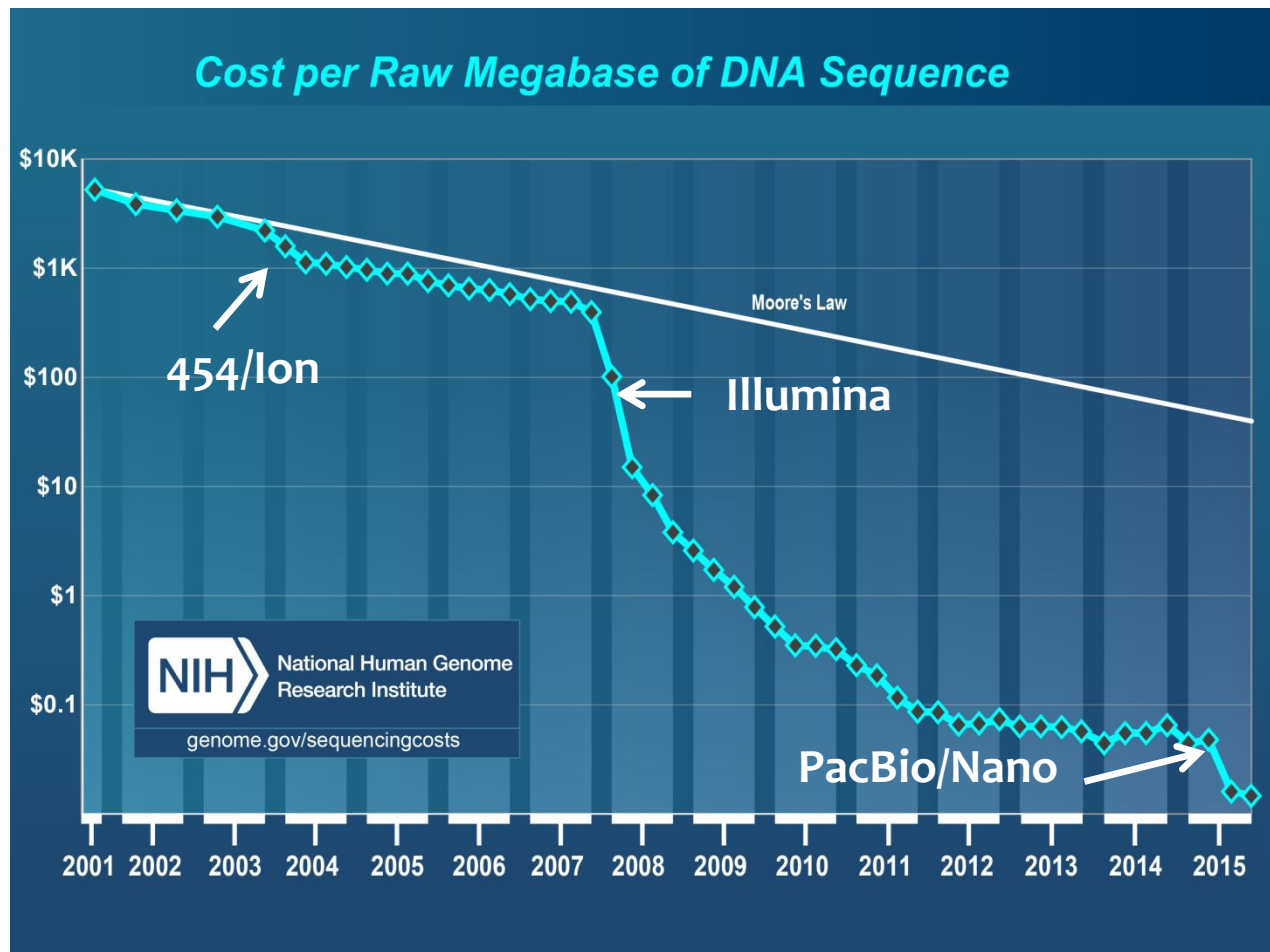
El DNA No solo DNA **Secuenciación**

- *Tecnologías de Next Generation Sequencing*
- Tercera generación, **Nanopore, Oxford Nanopore**

TECNOLOGÍA	SANGER	454 - Ion	Illumina*	PacBio	Nanopore
Potencia	50-100 kb, 96 secuencias/run	1 Gb /run	2 – 600 Gb/run	1 Gb/run	Hasta 2 Gb/run
Longitud de la lectura	0.5-2 kbp	200-400 bp	Hasta 300 bp	12-40 kbp	Hasta 40 kbp
Precisión	99%	98%	99%	90%	¿? (70-95%)
Precio por base	200.00 \$/Gb	700 \$/Gb	150\$/Gb (y bajando)	200\$/Gb	150\$/run (MinION)

El DNA No solo DNA Secuenciación

- *Tecnologías de Next Generation Sequencing*
- Evolución en el tiempo



- *Algunos consejos para el diseño de estudios con NGS*
- ¿Qué pregunta pretendemos plantearnos?
- ¿Cuántos datos necesito?
 - Sensibilidad (Falsos negativos, mutación no detectada)
 - Especificidad (Falsos positivos, mutación que no es cierta)
 - Coste
- ¿Qué factores influyen en la cantidad de datos a generar?
- Es difícil publicar cosas sin preguntas
 - Ya no es posible secuenciar por secuenciar

- *Algunos consejos para el diseño de estudios con NGS*
- ¿Qué pregunta pretendemos plantearnos?
- ¿Cuántos datos necesito?
- ¿Qué factores influyen en la cantidad de datos a generar?
- Es difícil publicar cosas sin preguntas
- ¿Necesito/Tengo genoma de referencia?
 - *Ensamblaje y anotación funcional*

- *Algunos consejos para el diseño de estudios con NGS*
 1. *Detección de tumores*
 2. *Evolución microbiana, buscando un gen de resistencia*
 3. *Filogenia de vertebrados*
 4. *Estudio de genómica de poblaciones*

- *Algunos consejos para el diseño de estudios con NGS*
 1. *Detección de tumores*
 - *Pocos falsos positivos y negativos*
 - *Secuenciar muchas veces el mismo fragmento*
 - *Estar seguros de las mutaciones que encontremos*
 2. *Evolución microbiana, buscando un gen de resistencia*
 - *Pocos falsos negativos, busco la mutación*
 - *Falsos positivos no tan importantes (si codifica a una proteína voy a tener comprobación posterior con aminoácidos)*

- *Algunos consejos para el diseño de estudios con NGS*

3. Filogenia de vertebrados

- *No importante falsos negativos (muchas mutaciones)*
- *Importante falsos positivos, diferenciar poblaciones*

4. Estudio de genómica de poblaciones

- *No “importa tanto” falsos positivos ni negativos*
- *Más importante añadir más individuos o réplicas*

- *Algunos consejos para el diseño de estudios con NGS*

¿Qué factores influyen en el diseño?

- *Número de muestras y réplicas*
- *Tipo de lectura (single-paired-mate)*
- *Tipo de librería*
- *Plataforma*
- *Poolear*
- *¿Tengo/necesito genoma de referencia?*
- *Bioinformática y recursos*

- *Algunos consejos para el diseño de estudios con NGS*

¿Qué factores influyen en el diseño?

- ***Número de muestras y réplicas***

Réplicas biológicas (dos individuos iguales o mismo tratamiento) VS Réplicas técnicas (mismo individuo, pero repito proceso de librería y secuenciación)

- *Algunos consejos para el diseño de estudios con NGS*

¿Qué factores influyen en el diseño?

- ***Número de muestras y réplicas***
- ***Tipo de lectura***

Lecturas Single-Paired-Mate

Cobertura: veces que se secuencia cada base

Longitud de la lectura (condiciona la plataforma)

- *Algunos consejos para el diseño de estudios con NGS*

¿Qué factores influyen en el diseño?

- ***Número de muestras y réplicas***
- ***Tipo de lectura***
- ***Tipo de biblioteca***

Es indispensable, pero cuanto menos mejor

Siempre es fuente de error. Amplificación por PCR

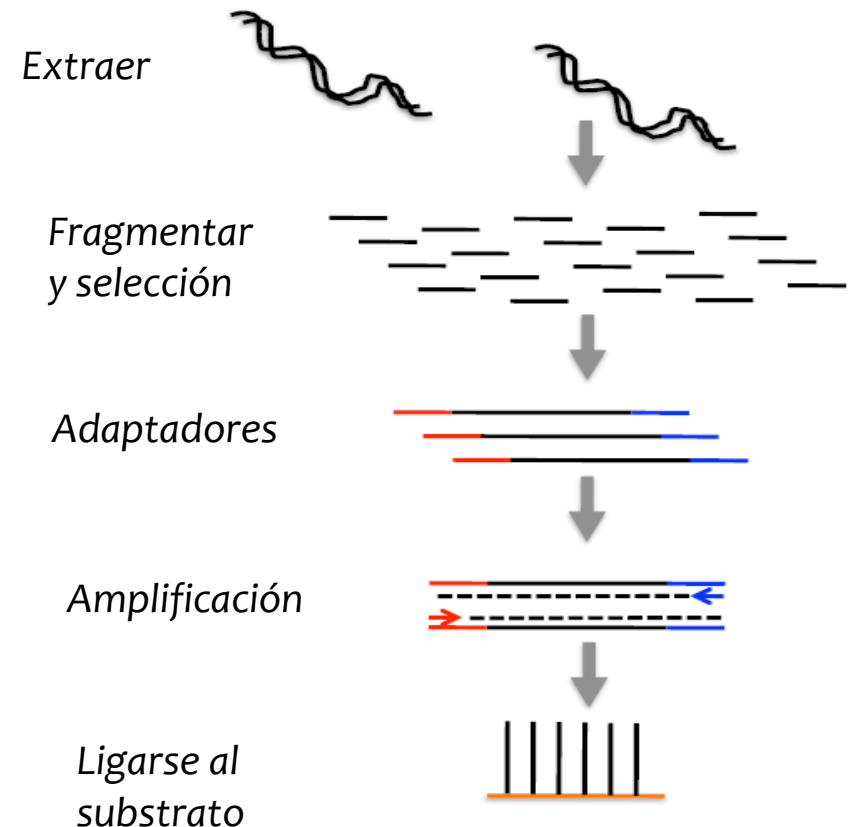
Necesidad de adaptadores para:

1. Barcode (identificación)
2. Primers para secuenciar o amplificar
3. Secuencia para unirse a algún elemento

- *Algunos consejos para el diseño de estudios con NGS*

¿Qué factores influyen en el diseño?

- **Número de muestras y réplicas**
- **Tipo de lectura**
- **Tipo de biblioteca**



- *Algunos consejos para el diseño de estudios con NGS*
 - ¿Qué factores influyen en el diseño?
 - **Número de muestras y réplicas**
 - **Tipo de lectura**
 - **Tipo de biblioteca**
 - **Plataforma (ya hemos hablado de ello)**

- *Algunos consejos para el diseño de estudios con NGS*
 - ¿Qué factores influyen en el diseño?
 - **Número de muestras y réplicas**
 - **Tipo de lectura**
 - **Tipo de biblioteca**
 - **Plataforma**
 - **Poolear**
 - Consiste en introducir varios individuos en el mismo “tubo”
 - Se puede identificar con un barcode (o no)
 - Interés por el individuos Vs Interés por la población
 - Mismo análisis, pero menos coste de secuenciación
 - Mas coste en biblioteca, perdida de datos

- *Algunos consejos para el diseño de estudios con NGS*

¿Qué factores influyen en el diseño?

- ***¿Tengo/necesito genoma de referencia?***

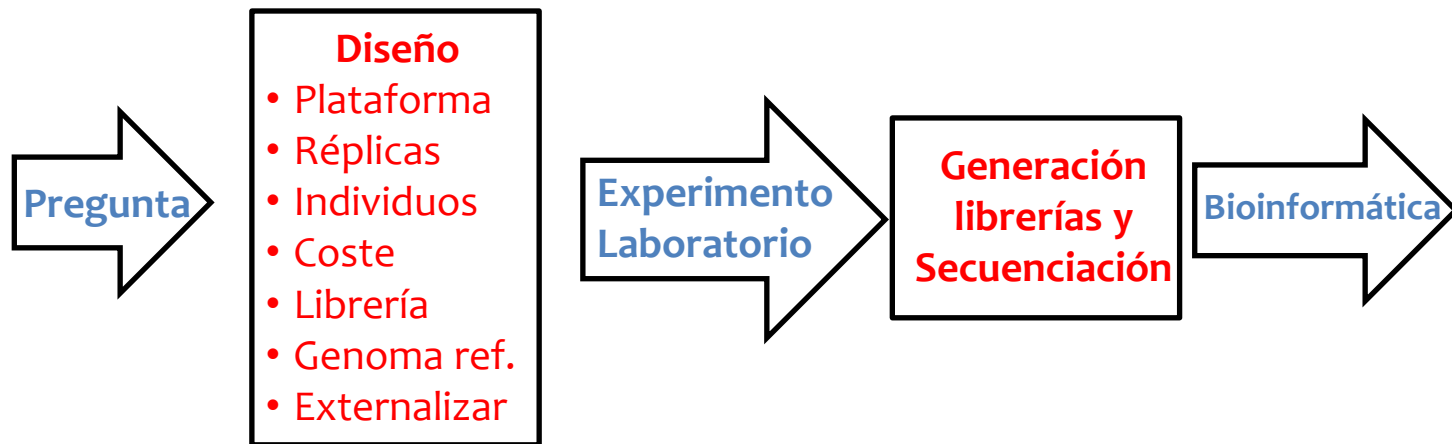
Sesión del martes por la tarde

Depende del tipo de estudio

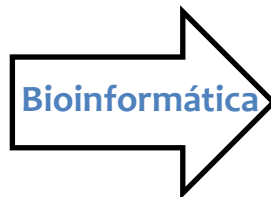
¿Cómo de buena es la referencia?

Cercana filogenéticamente siempre mejor

- *Un workflow para estudios de NGS...*



- *Un workflow para estudios de NGS...*



Tareas de bioinformática

- Limpieza de datos y edición
- Ensamblaje
- Anotación de genes
- Diversidad y estructura
- Expresión diferencial
- Identificación

- *Glosario básico*
 - *Plataforma de secuenciación*
 - *Primera secuenciación*
 - *Segunda secuenciación: 454, Ion, Illumina*
 - *Tercera secuenciación: PacBio, Nanopore*
 - *Librería, adaptadores...*
 - *Paired end – Single end*
 - *Ensamblaje*
 - *Anotación*
 - *Kmer*
 - *DNA, fragmentos de DNA & RNA*

- *Estimación de costes de un estudio de NGS y otras cosas*

<http://scotty.genetics.utah.edu/>

<https://genohub.com/>

http://epigenome.usc.edu/services/nextgen_sequencing.html

<http://www.bioconductor.org/packages/devel/bioc/manuals/RNASeqPower/man/RNASeqPower.pdf>

<https://github.com/StoreyLab/subSeq>

<https://www.genome.gov/27541954/dna-sequencing-costs-data/>

<http://nextgenseq.blogspot.com.es/>

<http://omicsomics.blogspot.com.es/>

<http://www.molecular ecologist.com/>

- *Antes de comer...*
 1. *Punto de restauración de Windows (si no quereis mantener la distribución Linux)*
 2. *Descargar Ubuntu*
 3. *Descargar Unetbooting*