

Técnicas de secuenciamiento de DNA: Sanger, HTS

Genómica para bioinformática INB320

15 marzo 2016

Eduardo Castro-Nallar, PhD

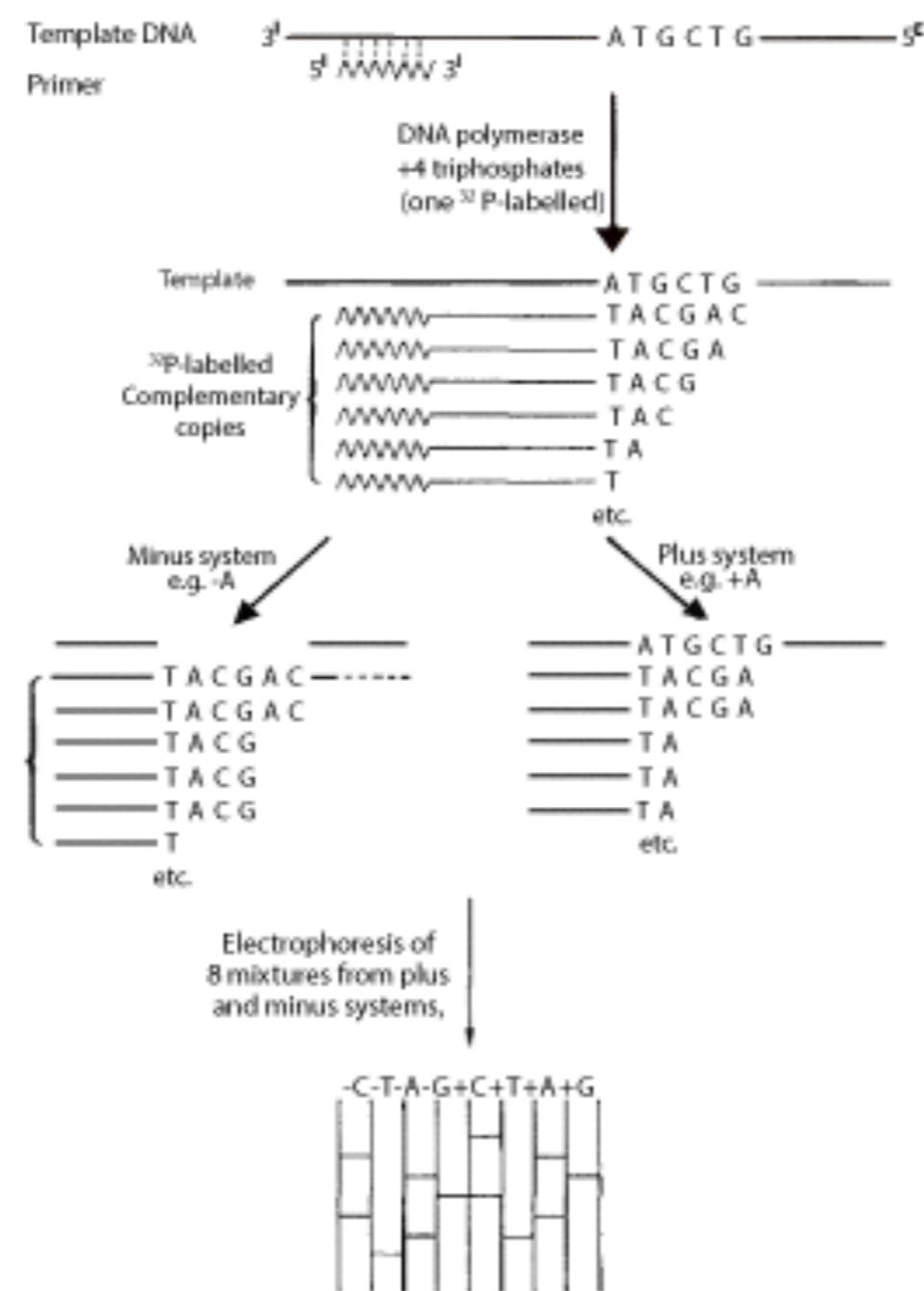
www.castrolab.org

De la estructura del DNA a la secuencia

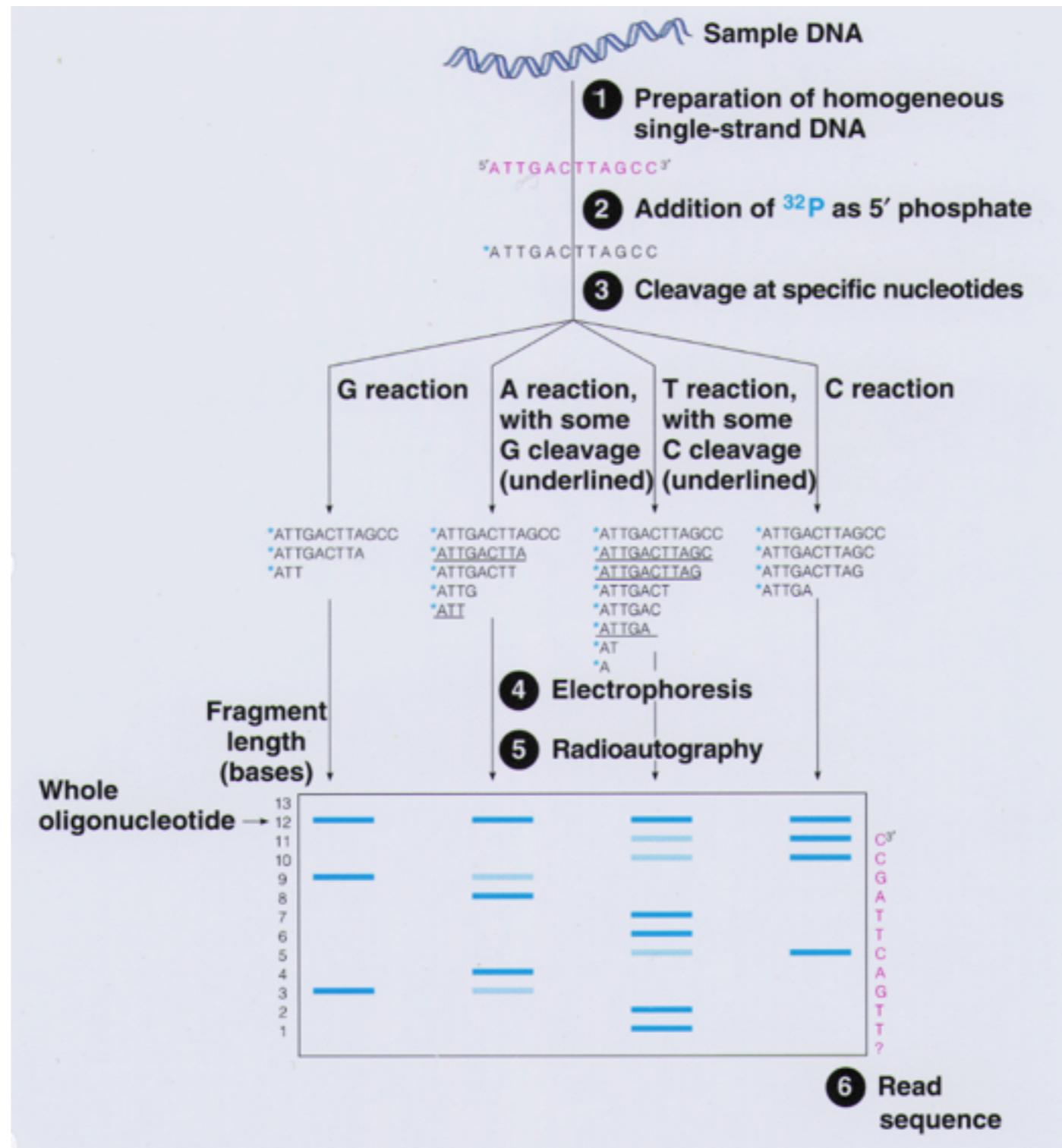
- 1953 descifran estructura del DNA
- 1968 primera secuencia de DNA
- 1970 enzimas de restricción de tipo II
- 1975 secuenciamiento “plus and minus” de Sanger and Coulson
- 1977 método del dideoxi de Sanger

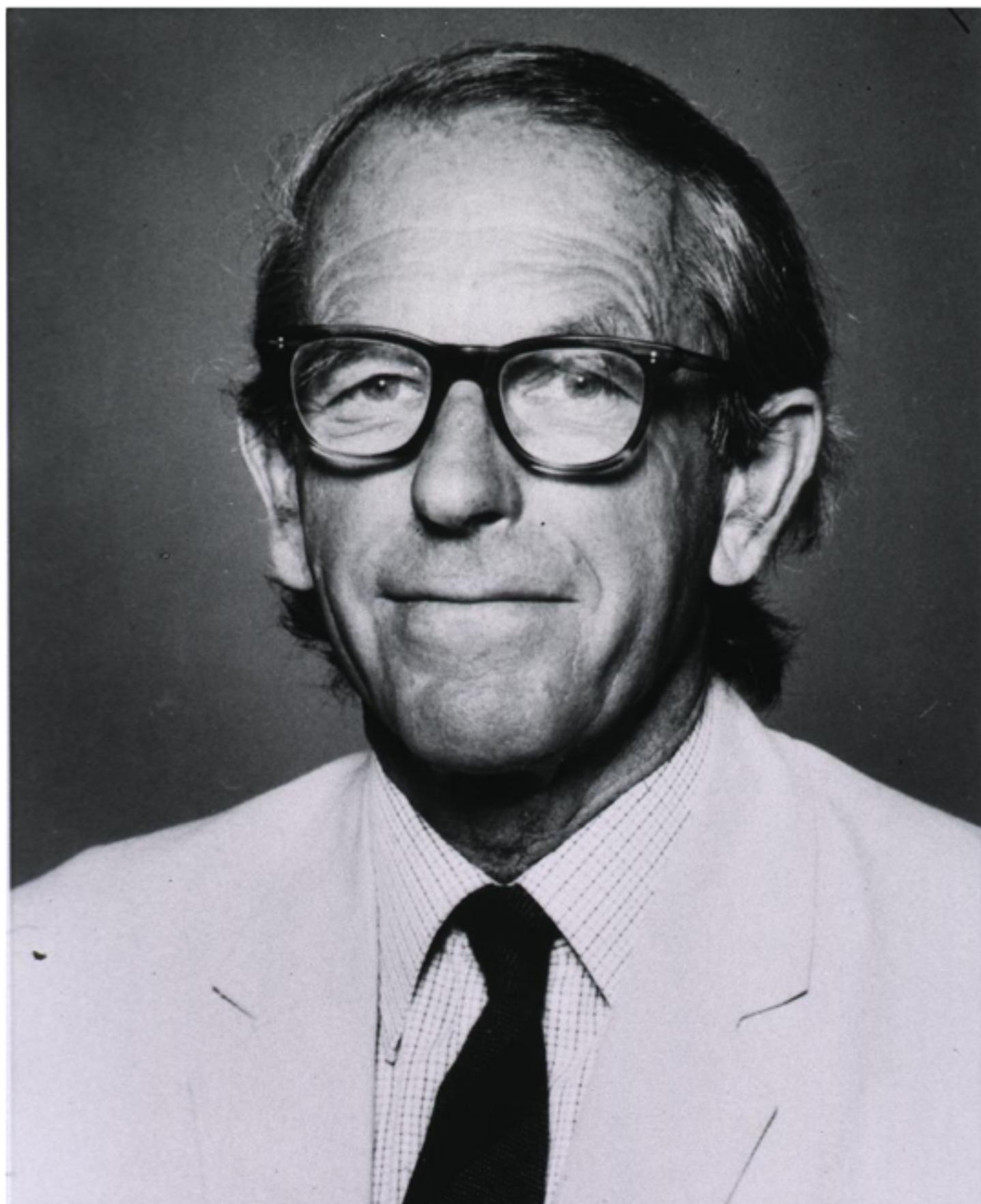
Secuenciamiento “plus and minus” de Sanger and Coulson

F. SANGER AND A. R. COULSON

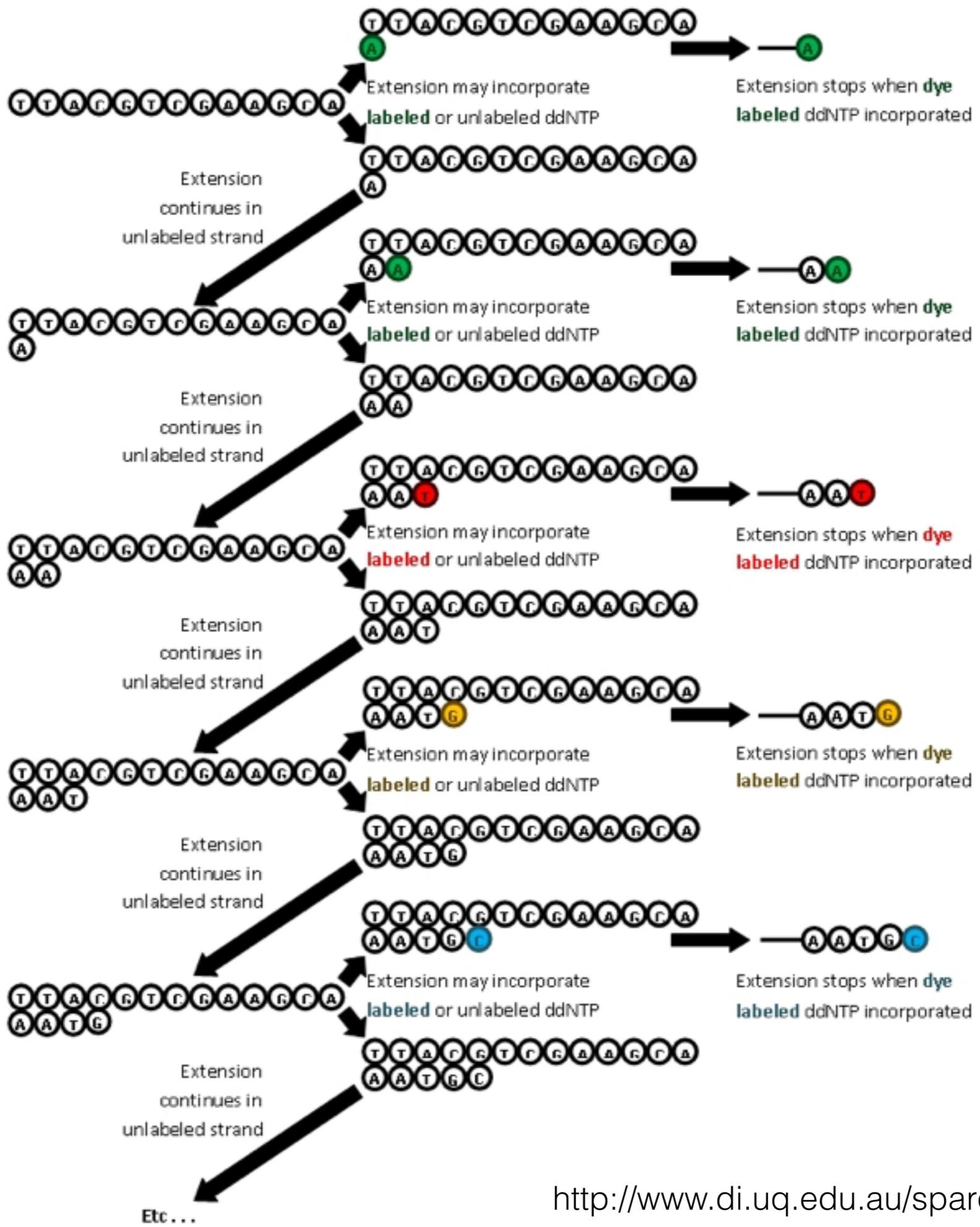


Secuenciamiento “químico” de Maxam and Gilbert 1977

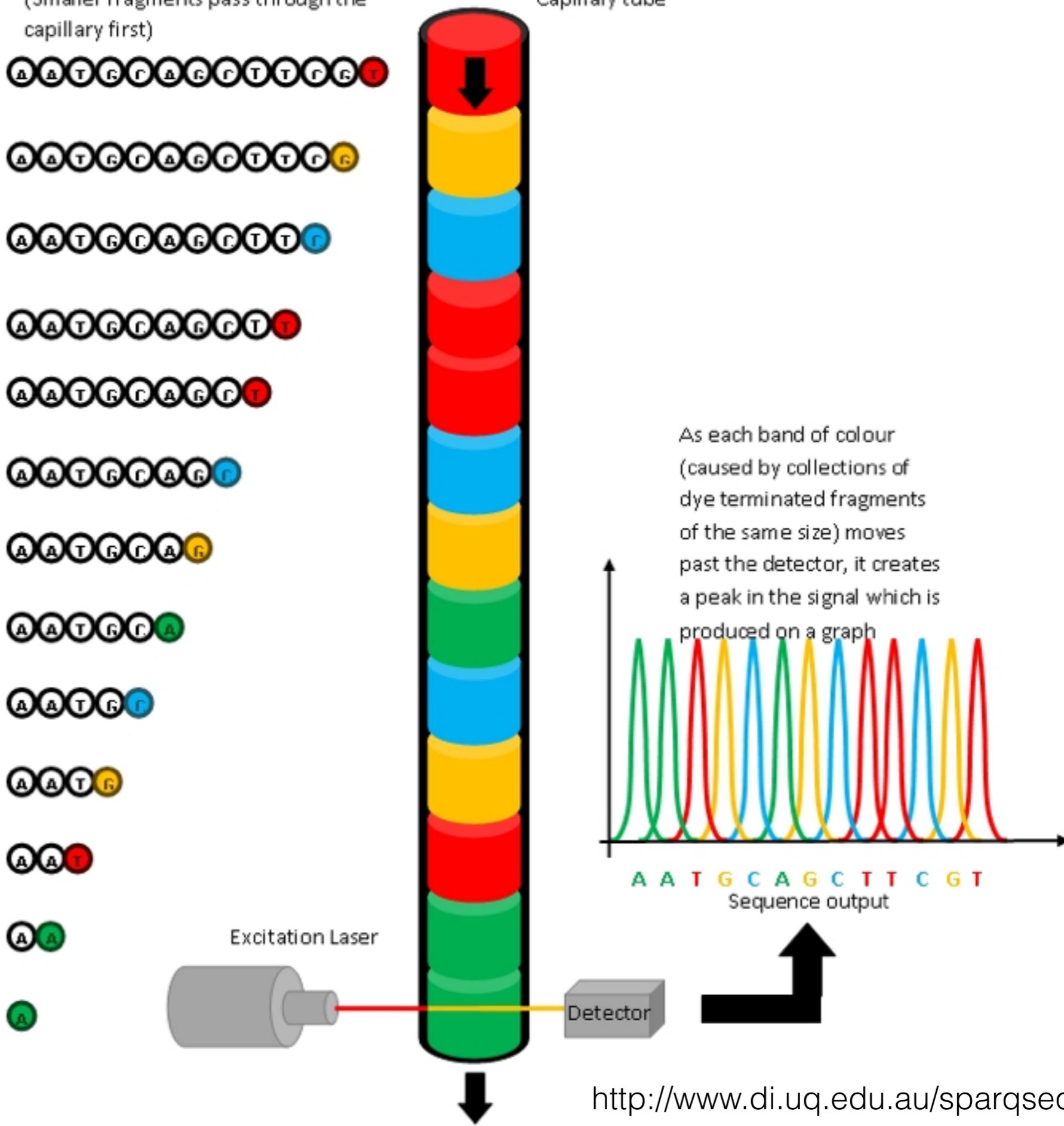




- 1958 estructura de proteínas, especialmente insulina
- 1980 determinación de secuencia de bases en ácidos nucleicos



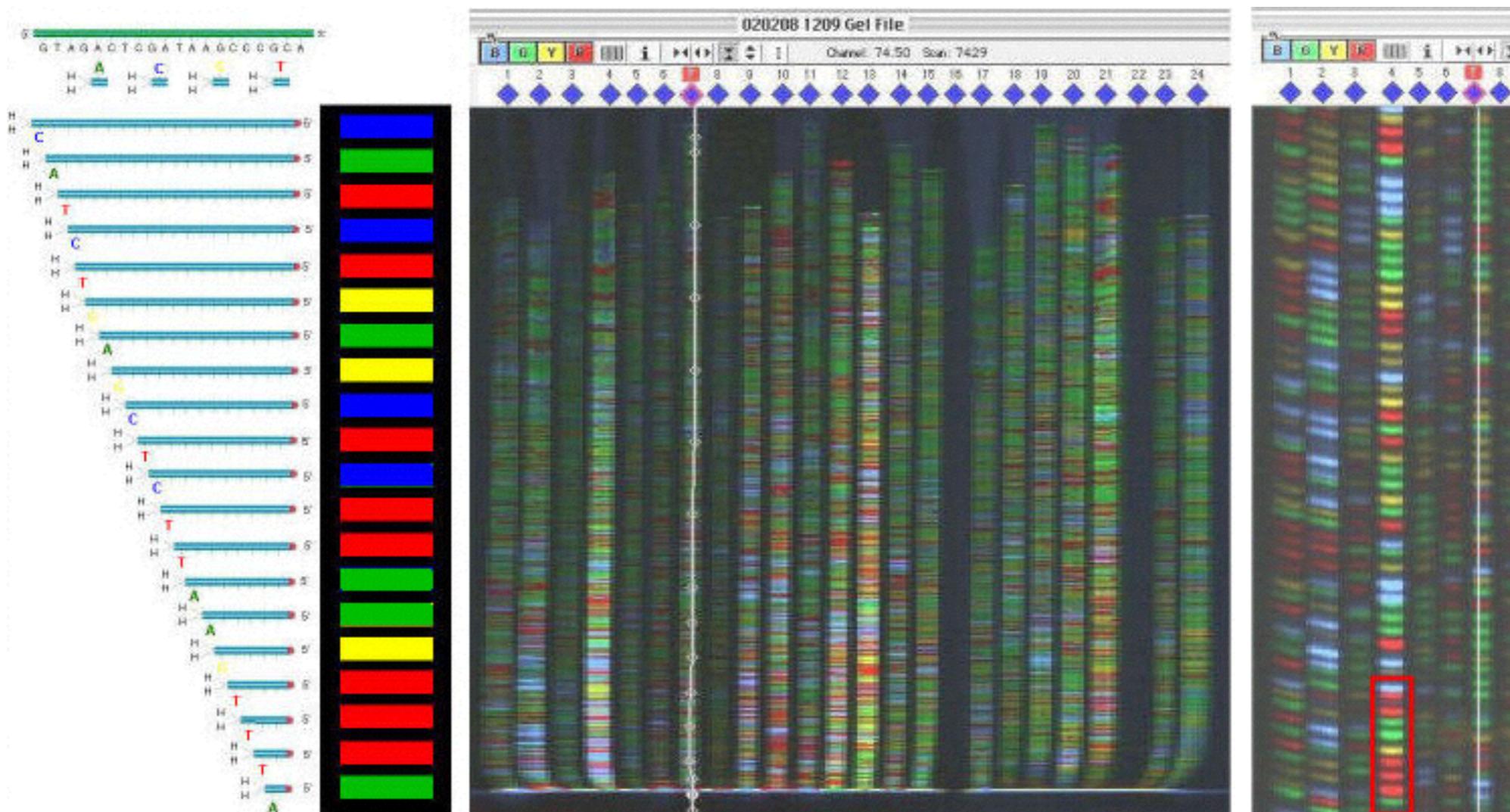
DNA Fragments with Dye Terminators
(Smaller fragments pass through the
capillary first)



De la estructura del DNA a la secuencia

- 1986 CalTech y Applied Biosystems trabajan en secuenciador automático
- ABI 370A —> 1000 bp por día
- 1990 se lanza el proyecto genoma humano
- 1995 primer secuenciador automático ABI Prism 377

Imagen real de un gel de secuenciamento



...Transformada a un chromatograma



Requisitos del método de Sanger

- DNA tiene que tener un primer; 1000 bp a la vez
- Lento y laborioso —> clonar en BACs, YACs
- Dos estrategias —> Directo o clon por clon y aleatorio o shotgun

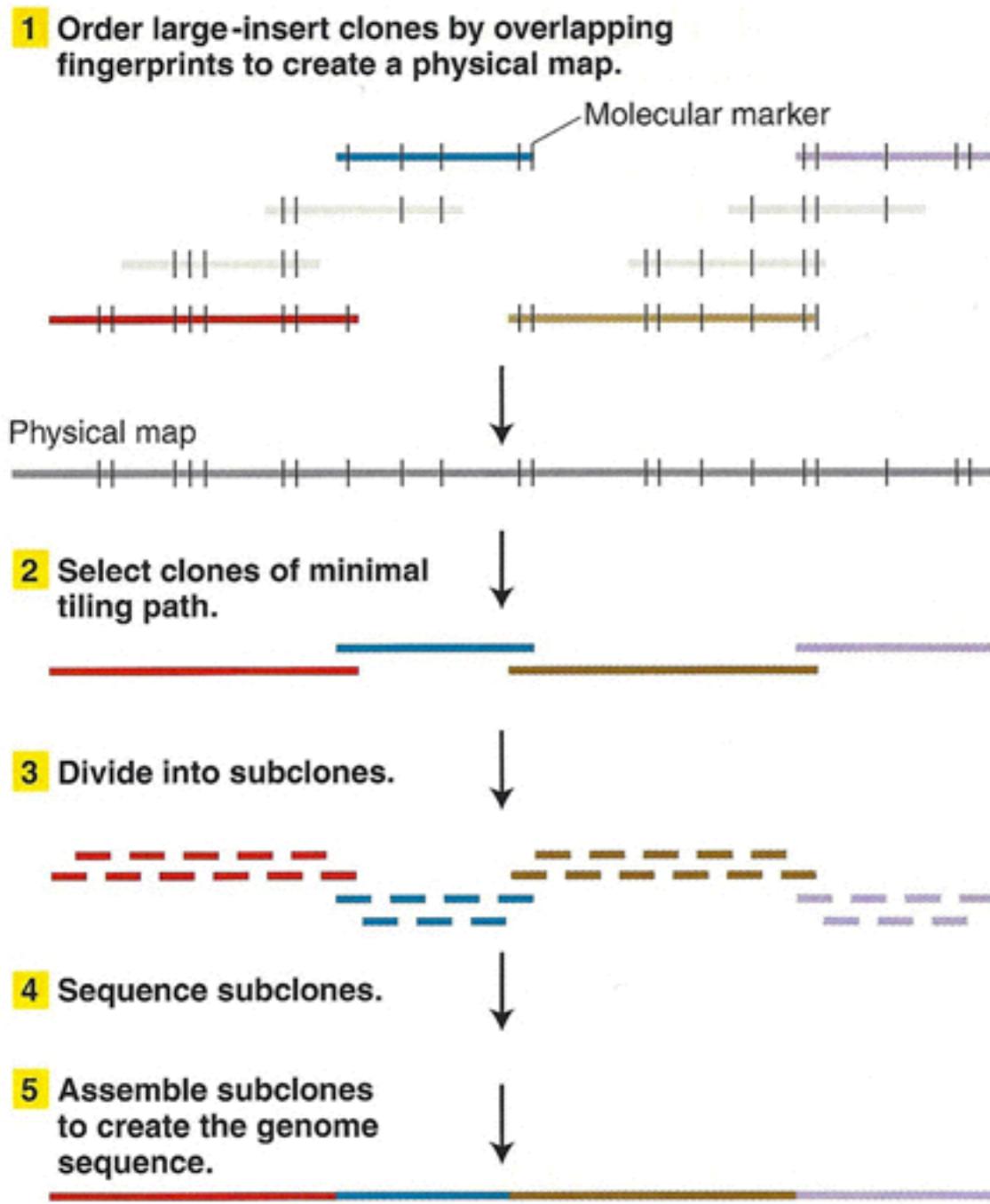
Características Sanger

First generation (Sanger) sequencing

throughput	50-100kb, 96 sequences per run
read length	0.5-1.1 kbp
accuracy	high quality bases - 99%: ~900 bp very high quality bases - 99.9%: ~600 bp 99.999%: 400-500 bp
price per raw base	~400k€/Gb

¿Cómo secuenciar genomas completos si solo podemos secuenciar 1000 bp?

Secuenciamiento por clones o basado en mapas

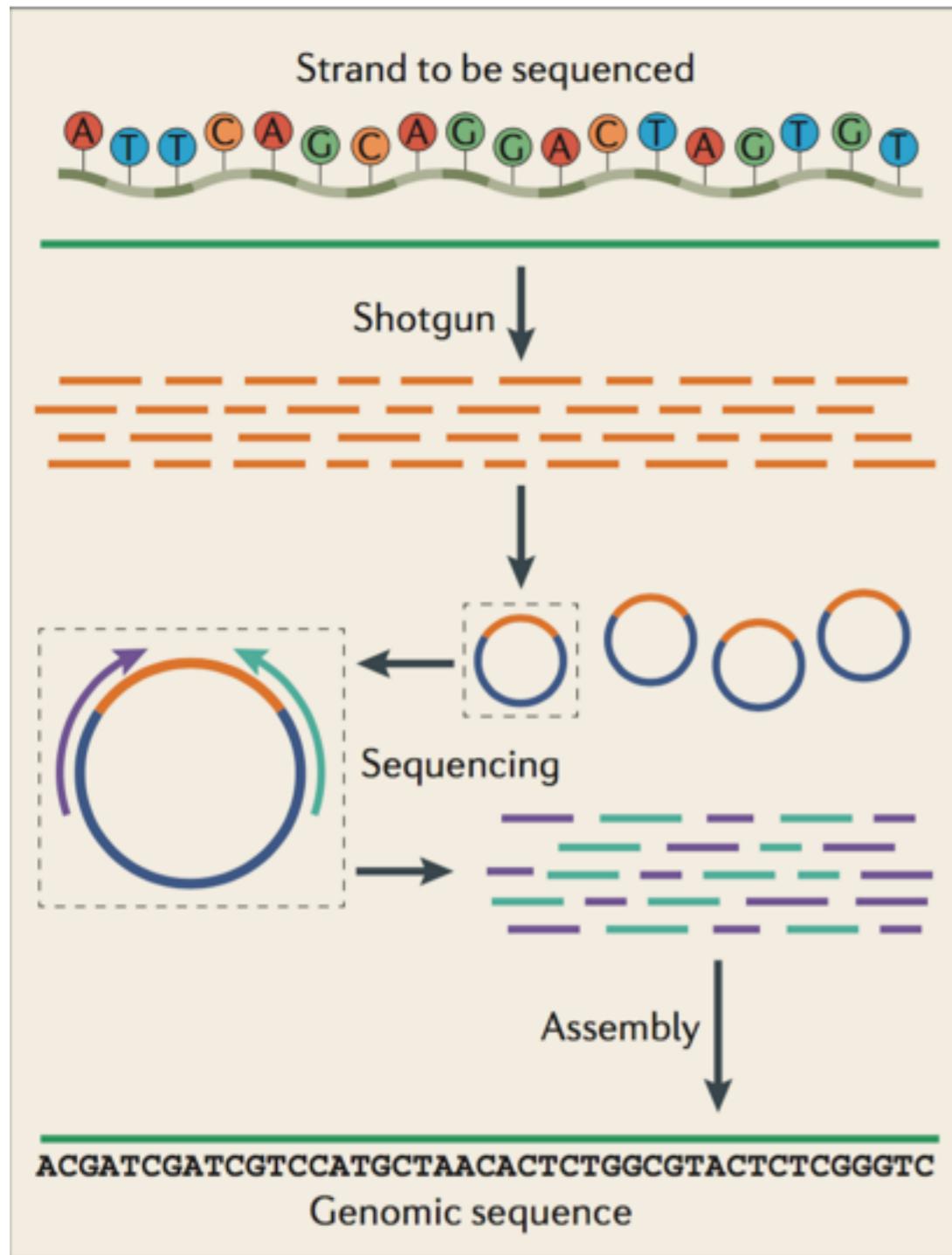


- Crear mapa físico
- Seleccionar clones con trayectoria minima
- Dividir en subclones
- Secuenciar subclones
- Ensamblar

Limitaciones

- Súper laborioso
- Toma mucho tiempo, recursos y personal especializado
- Caro → \$3 mil millones USD proyecto genoma humano NIH

Whole-genome shotgun: la primera revolución



- Más fácil que clone-by-clone
- Creada por J Craig Venter
- Se transforma en la estrategia dominante
- Pone por primera vez la “carga” en el análisis post-secuenciación
- Armar el puzzle después de que el experimento ha concluido
- Menos caro: \$300,000,000 USD

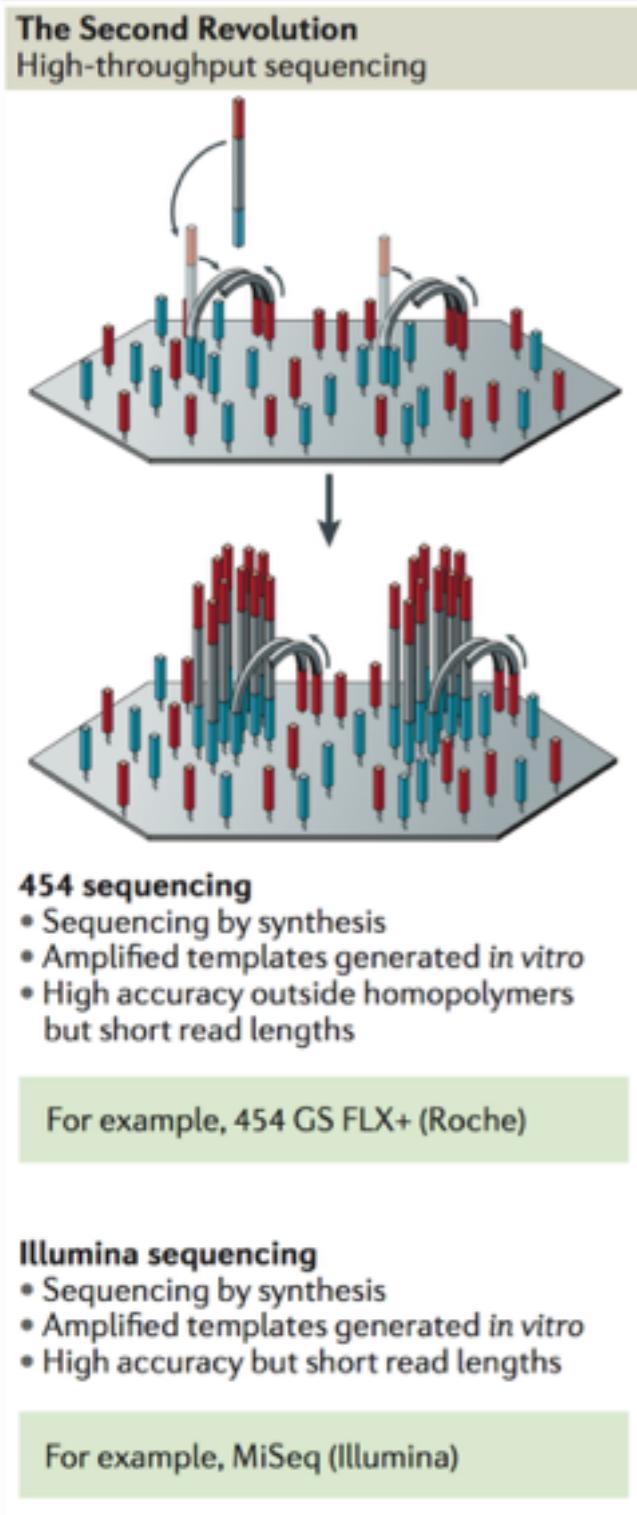
¿Cuándo llega el secuenciamiento masivo?

- 2000 con el MPSS

Gene expression analysis by massively parallel signature sequencing (MPSS)
on microbead arrays

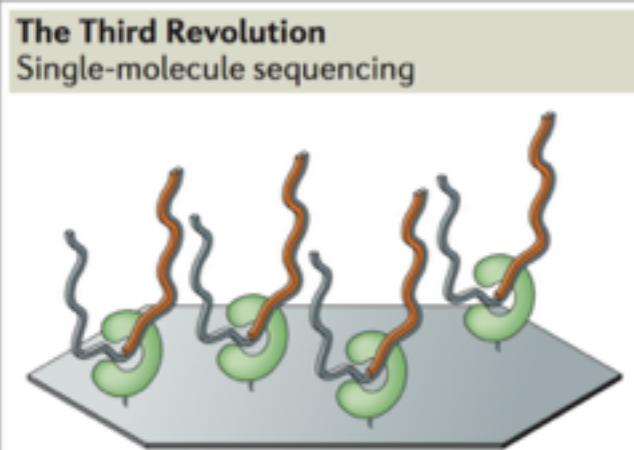
- Para secuenciar cDNA

High-throughput sequencing: la segunda revolución



- 2005
- No hay terminación temprana de la cadena naciente
- Secuenciamiento por síntesis
- Alto rendimiento —> 1 M a 25 M de fragmentos o “reads”
- Mayor tasa de error que Sanger pero no importa

Tercera revolución? Single-molecule sequencing



Pac Bio SMRT sequencing

- Sequencing by synthesis
- Single-molecule templates
- Low accuracy but long read lengths

For example, PacBio RS
(Pacific Biosciences)

- 2009
- PacBio SMRT → Secuenciamiento por síntesis
- Produce fragmentos largos de hasta 200 kbp
- Mayor tasa de error que Illumina o 454
- No hay amplificación
- Puede capturar señales epigenéticas

Oxford Nanopore sequencing

- Nanopore sequencing
- Single-molecule templates
- Low accuracy but long read lengths

For example, MinION
(Oxford Nanopore)

¿Cómo funcionan las técnicas de sequenciamiento masivo?

Metzker ML. Sequencing technologies – the next generation. *Nat Rev Genet.* 2009;11(1):31-46. doi: 10.1038/nrg2626.

https://en.wikipedia.org/wiki/Ion_semiconductor_sequencing

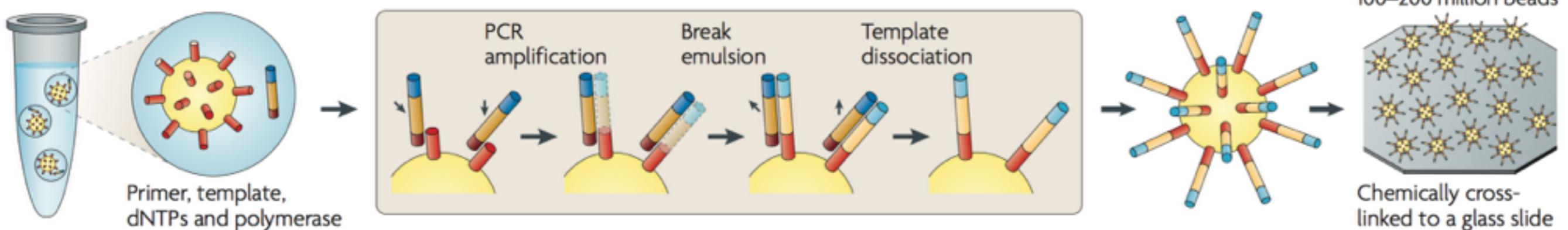
Técnicas de secuenciamiento de alto rendimiento (HTS)

- 454 - pirosecuenciamiento
- Illumina - secuenciamiento por síntesis
- Helicos - secuenciamiento por síntesis; single molecule
- PacBio - secuenciamiento por síntesis; single molecule
- Solid - secuenciamiento por ligación
- Ion Torrent - secuenciamiento por síntesis; semiconductor
- Nanopore

454 - pirosecuenciamiento

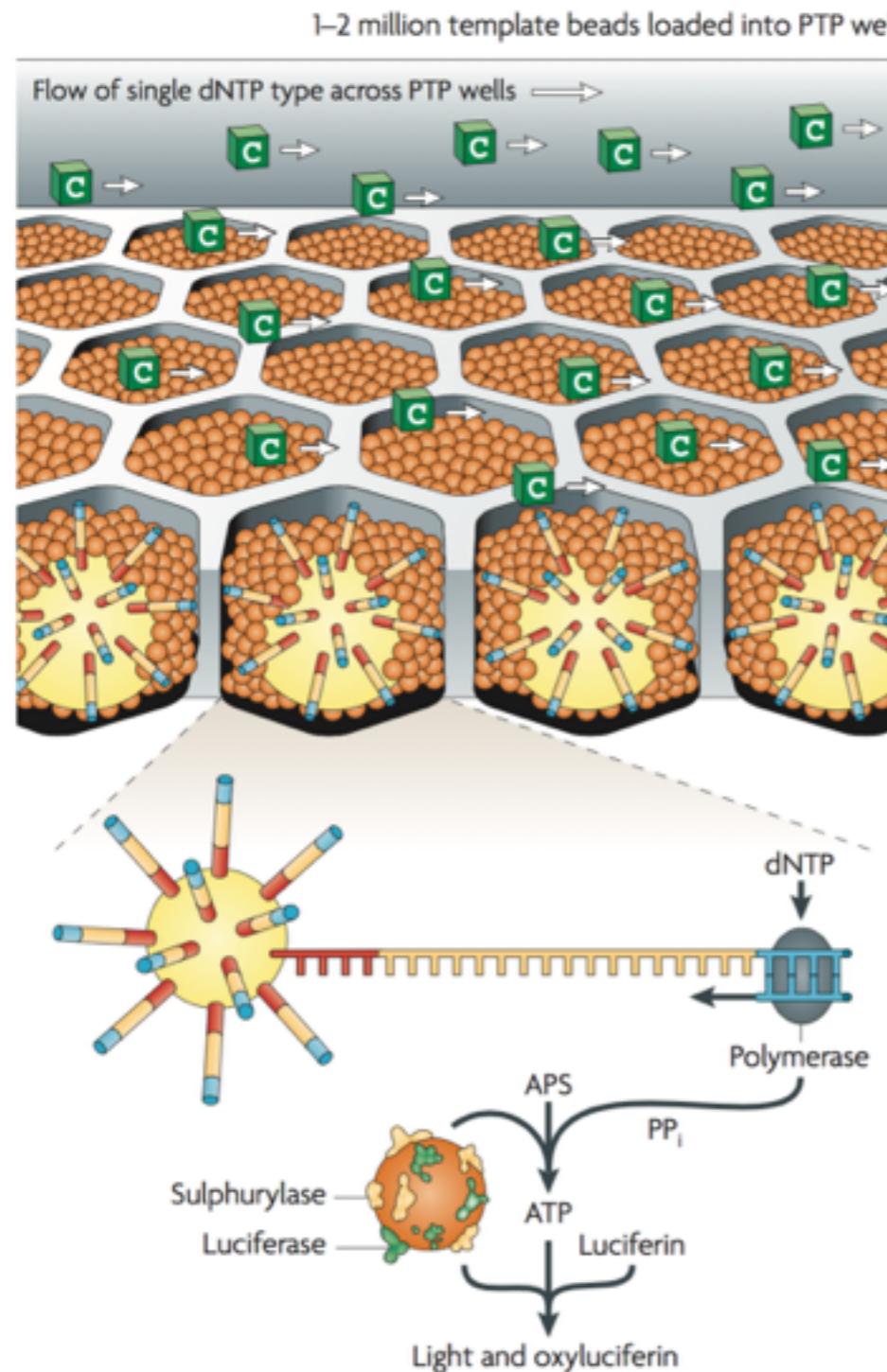
a Roche/454, Life/APG, Polonator Emulsion PCR

One DNA molecule per bead. Clonal amplification to thousands of copies occurs in microreactors in an emulsion



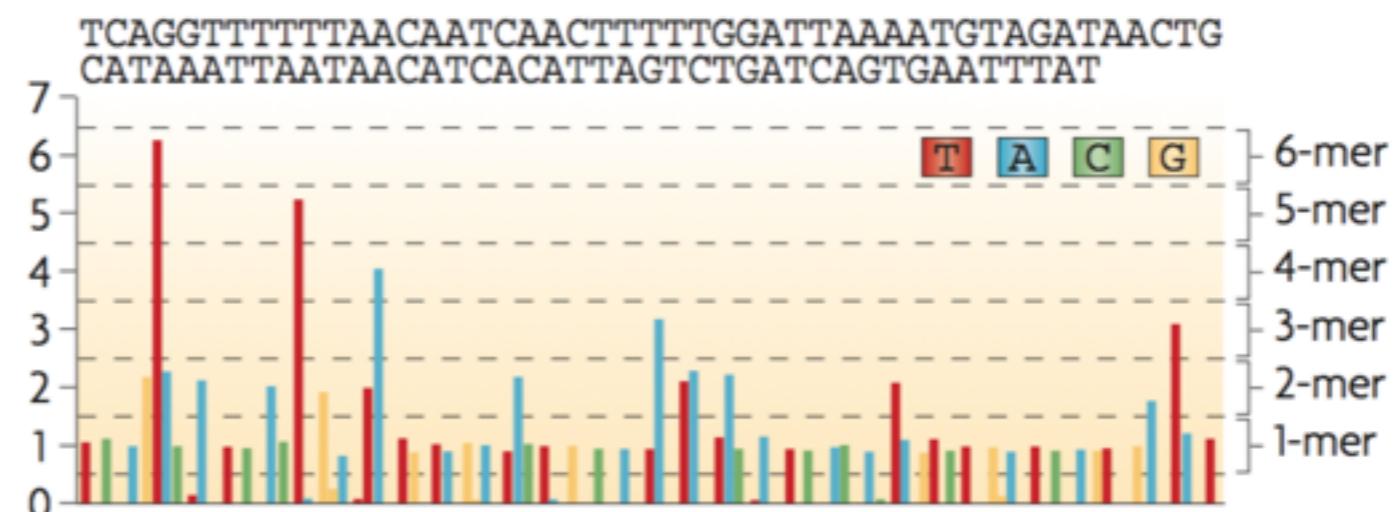
454 - pirosecuenciamiento

c
Roche/454 — Pyrosequencing



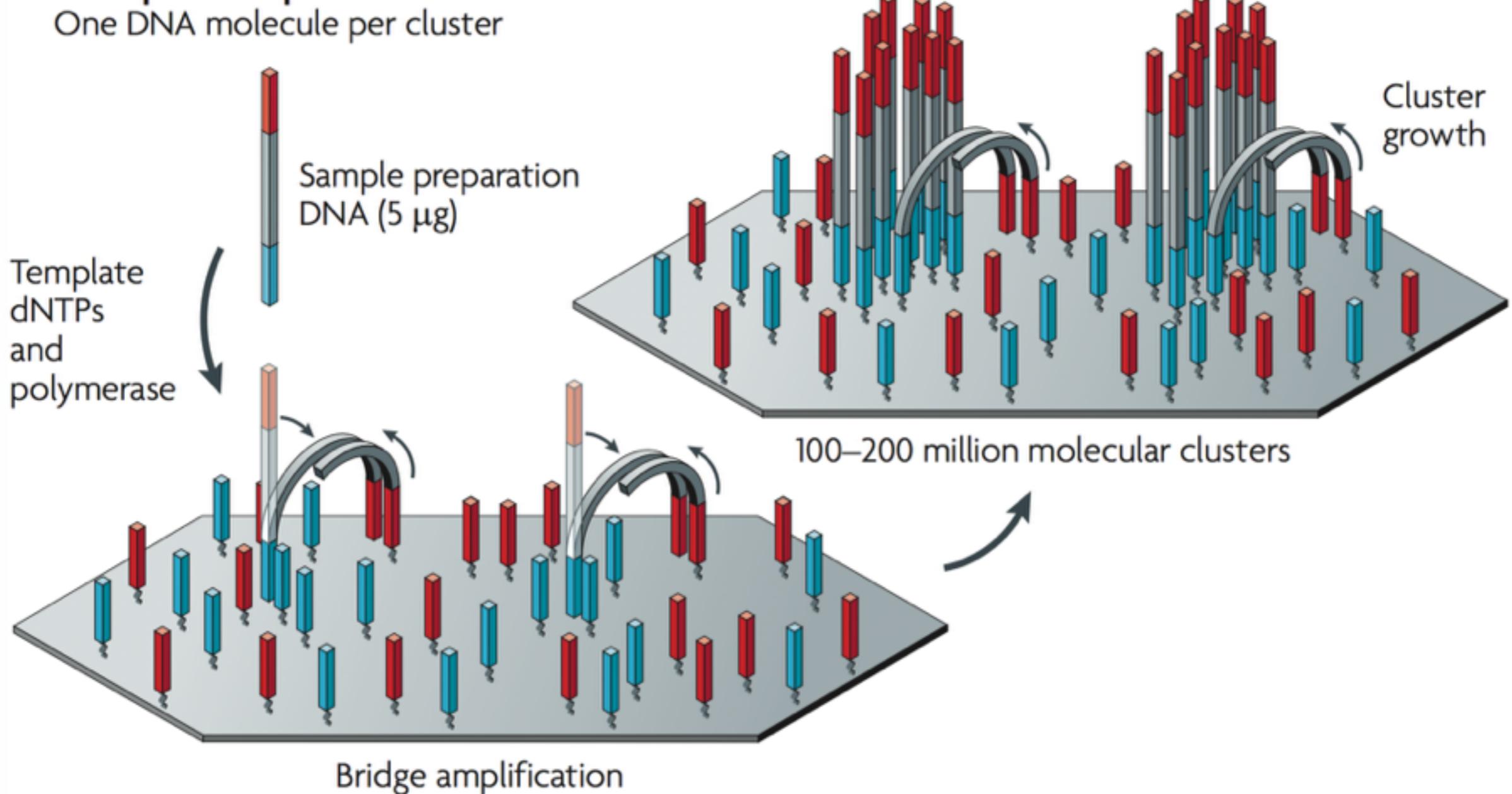
d

Flowgram



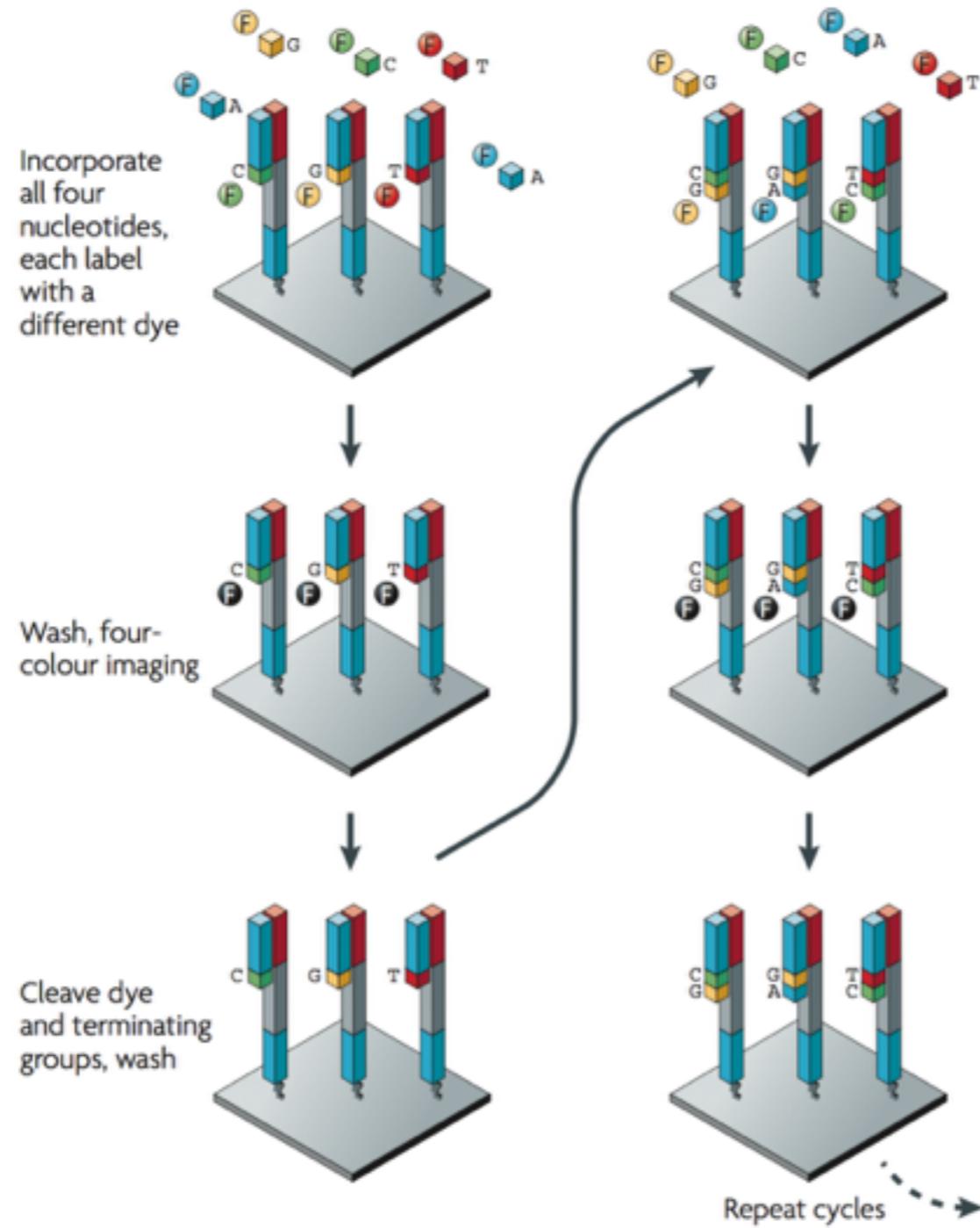
Illumina - secuenciamiento por síntesis

b Illumina/Solexa
Solid-phase amplification
One DNA molecule per cluster



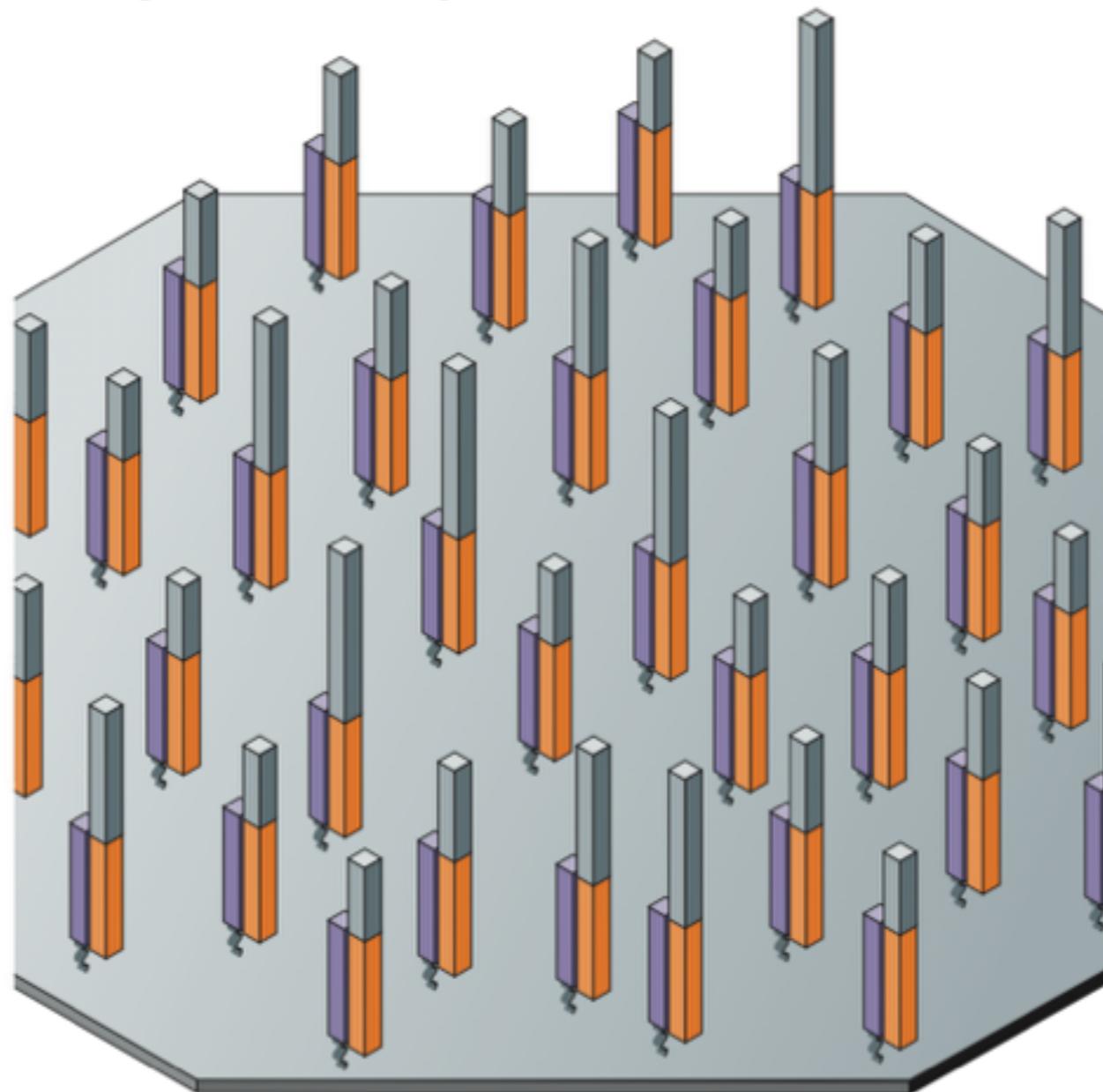
Illumina - secuenciamiento por síntesis

a Illumina/Solexa — Reversible terminators



Helicos - secuenciamiento por síntesis

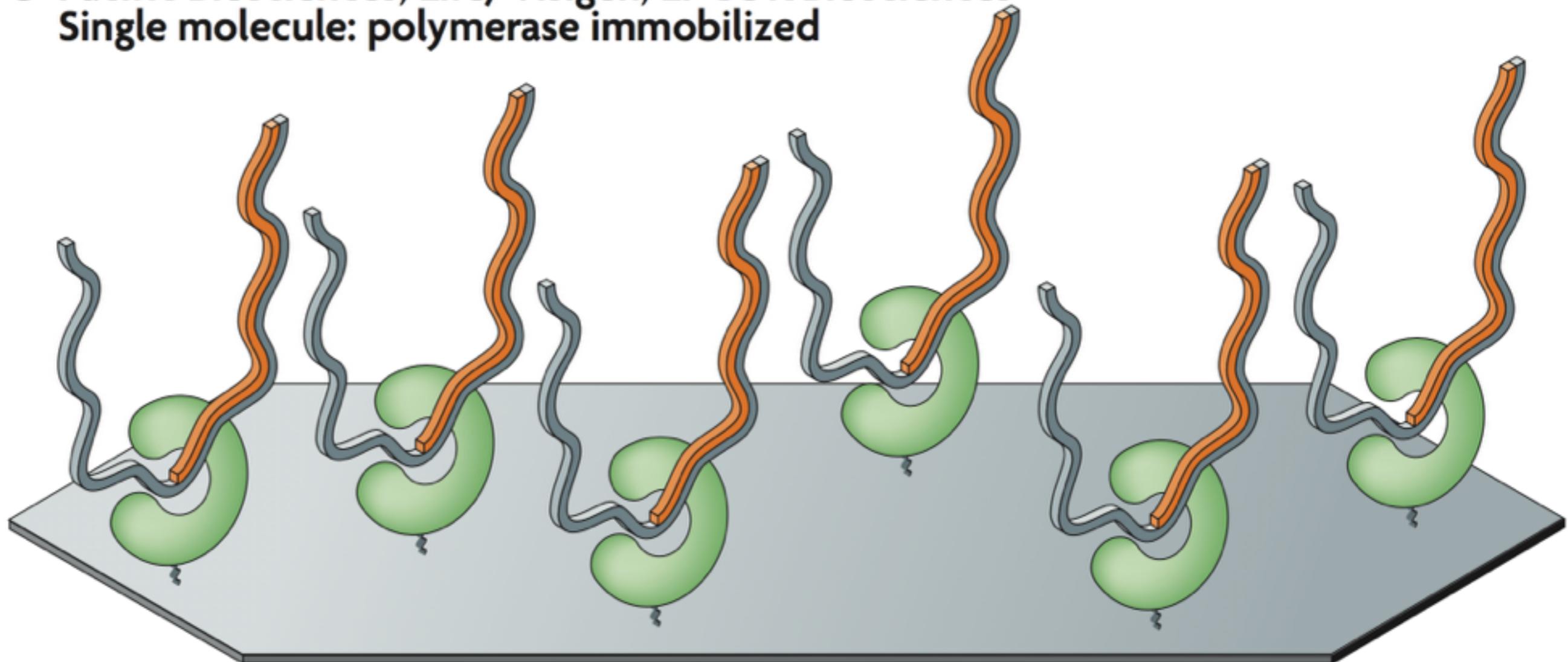
c Helicos BioSciences: one-pass sequencing
Single molecule: primer immobilized



Billions of primed, single-molecule templates

PacBio - secuenciamiento por síntesis

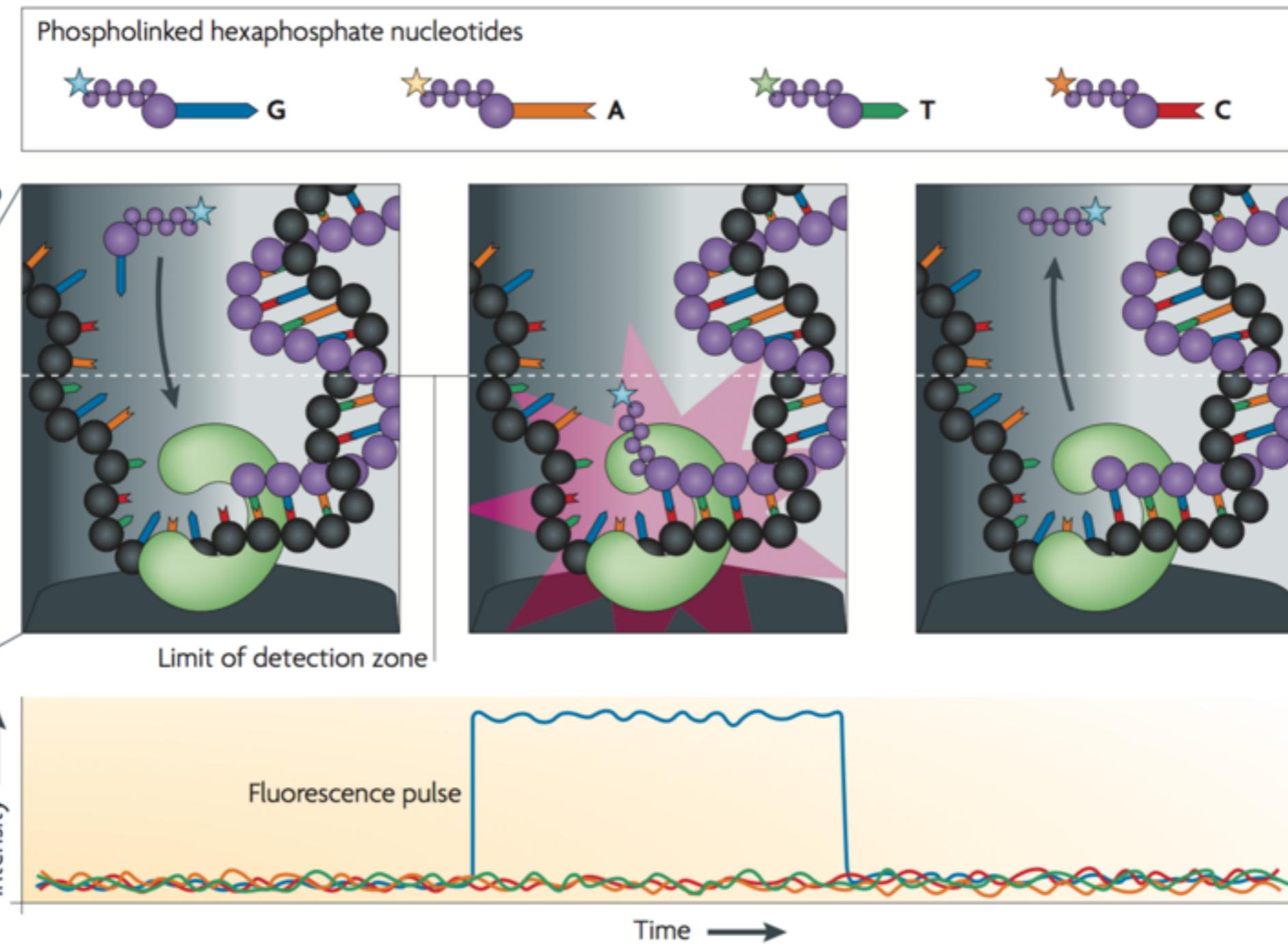
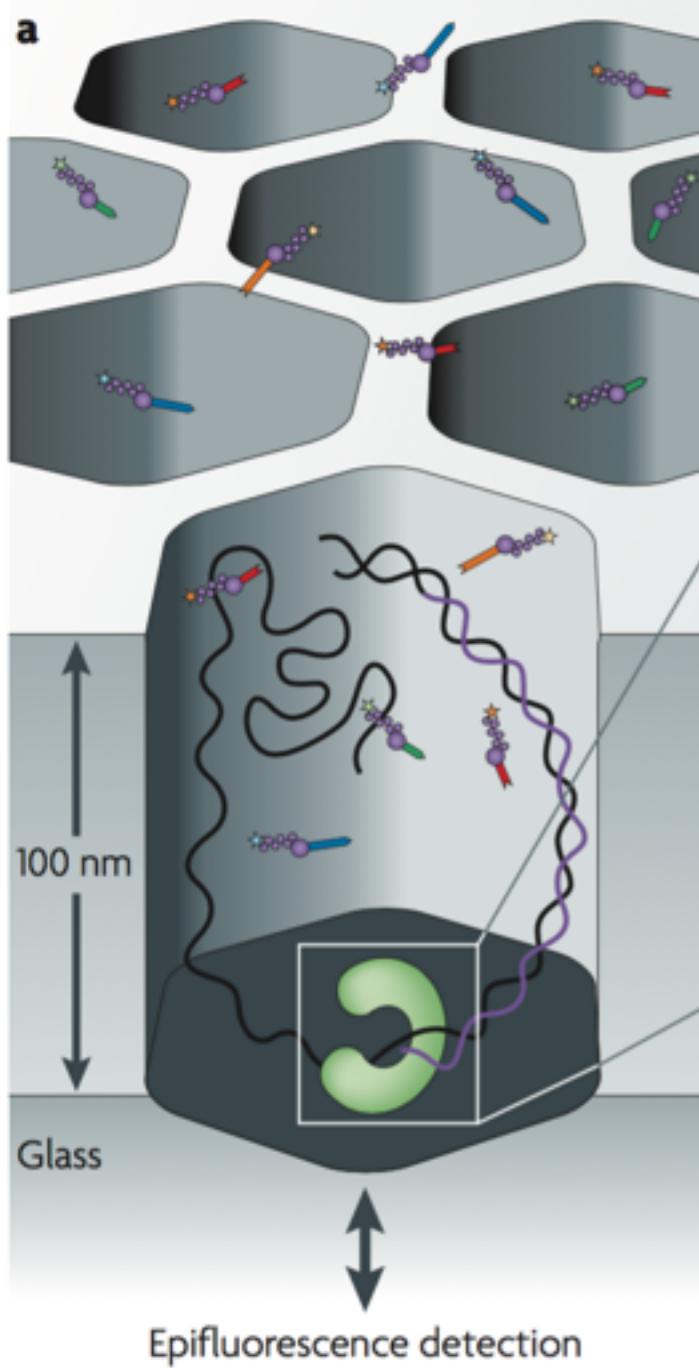
e Pacific Biosciences, Life/Visigen, LI-COR Biosciences
Single molecule: polymerase immobilized



Thousands of primed, single-molecule templates

PacBio - secuenciamiento por síntesis; single molecule

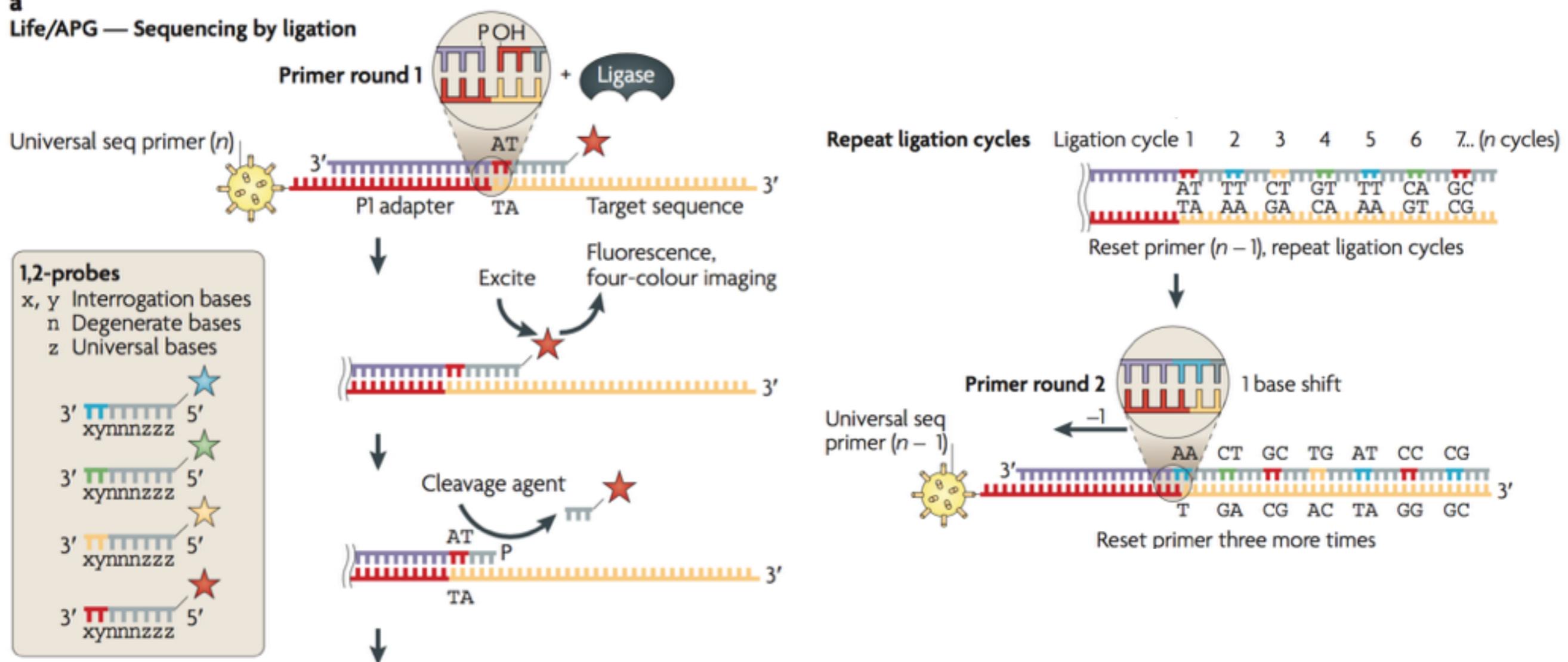
Pacific Biosciences — Real-time sequencing



SOLID - secuenciamiento por ligación

a

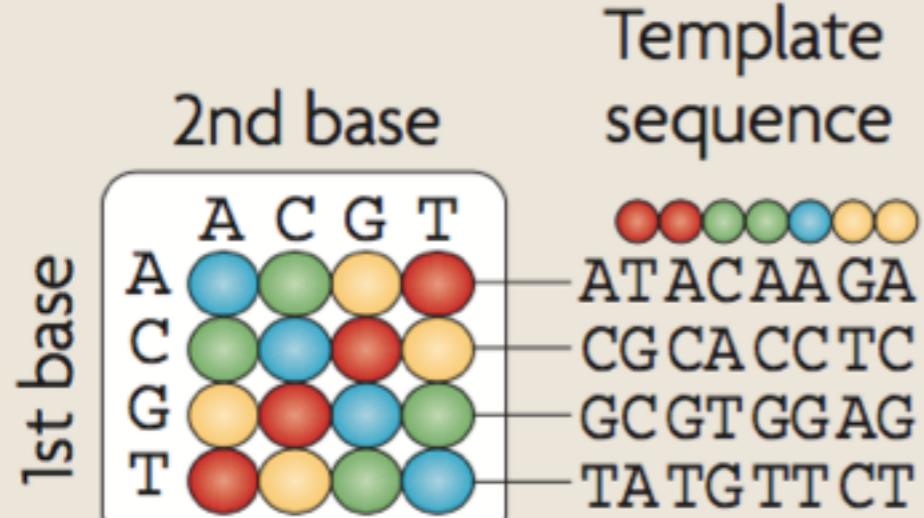
Life/APG — Sequencing by ligation



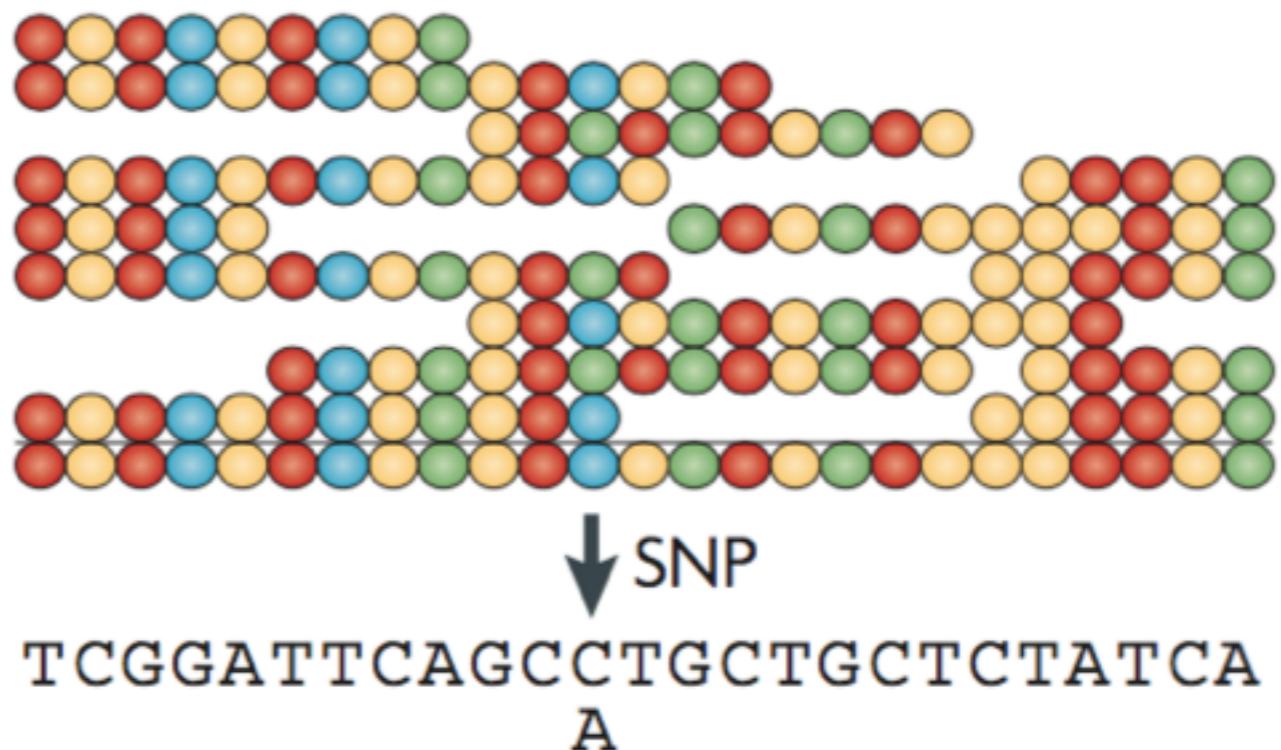
SOLID - secuenciamiento por ligación

b

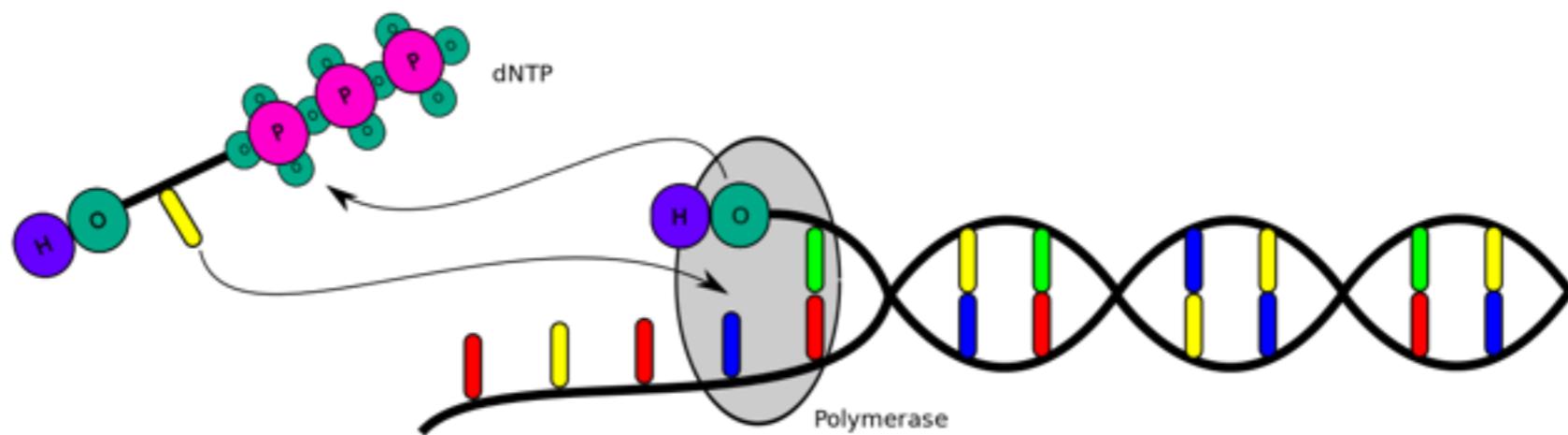
Two-base encoding: each target nucleotide is interrogated twice



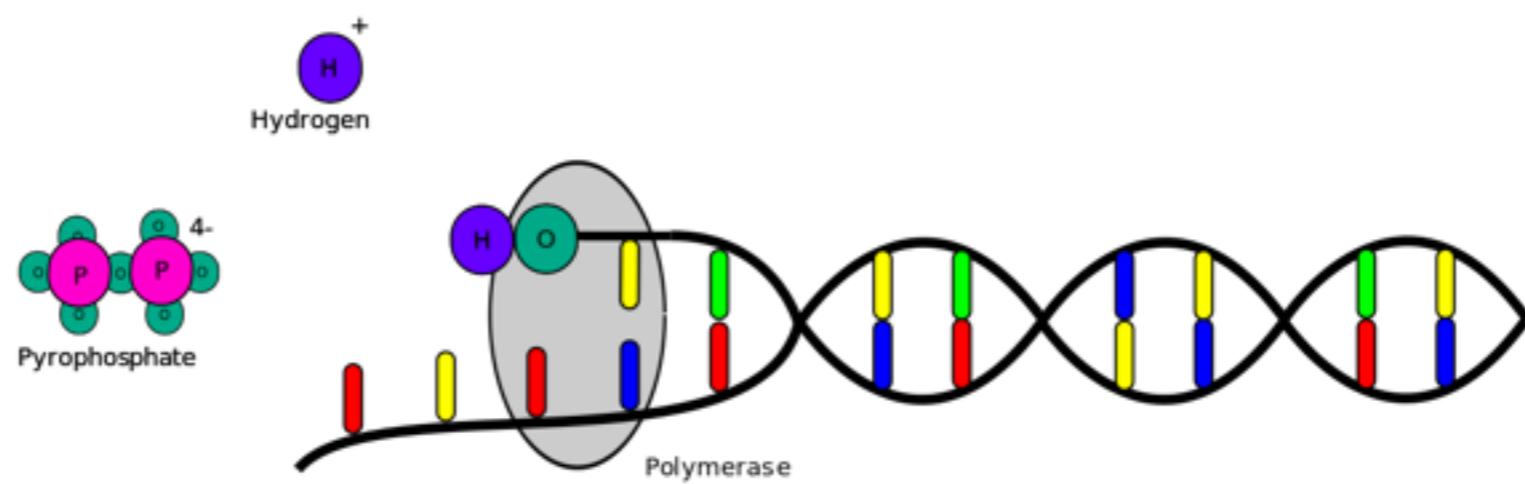
Alignment of colour-space reads to colour-space reference genome



IonTorrent - secuenciamiento por síntesis; cambio en corriente (semiconductor)

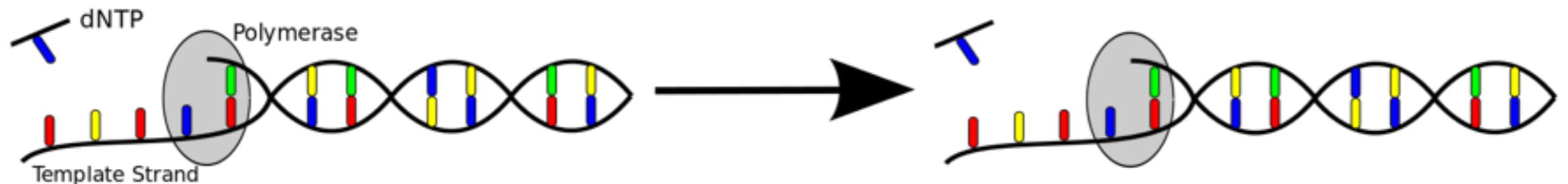


Polymerase integrates a nucleotide.

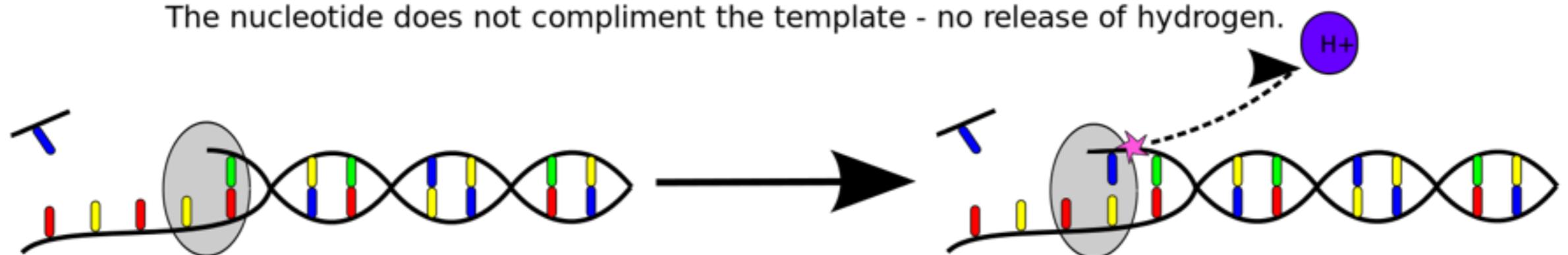


Hydrogen and pyrophosphate are released.

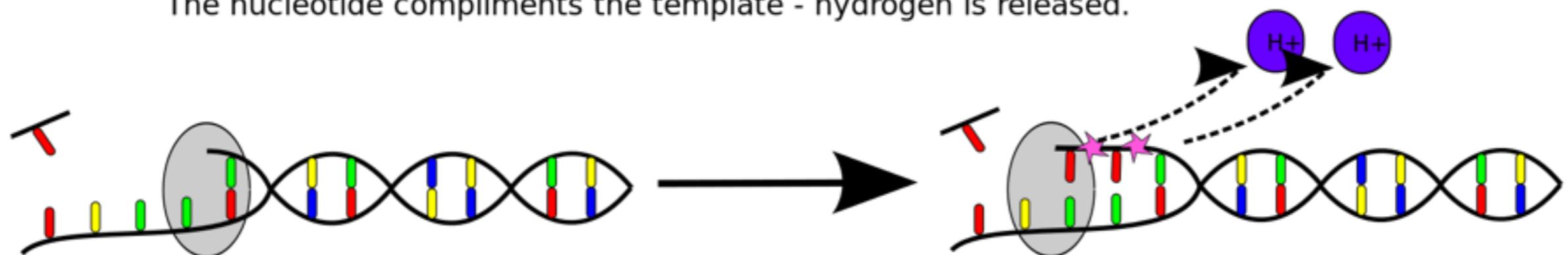
IonTorrent - secuenciamiento por síntesis



The nucleotide does not compliment the template - no release of hydrogen.

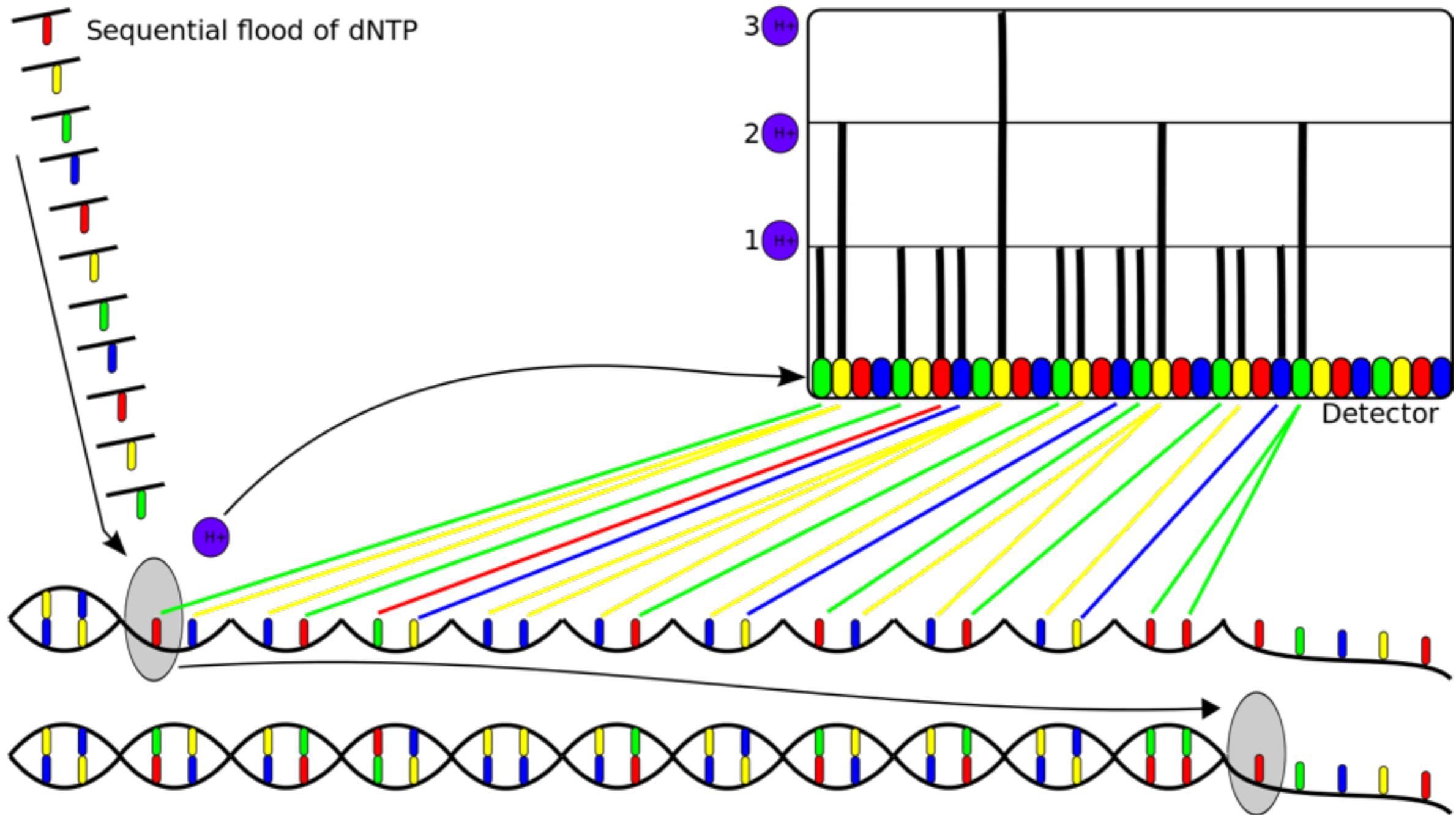


The nucleotide compliments the template - hydrogen is released.

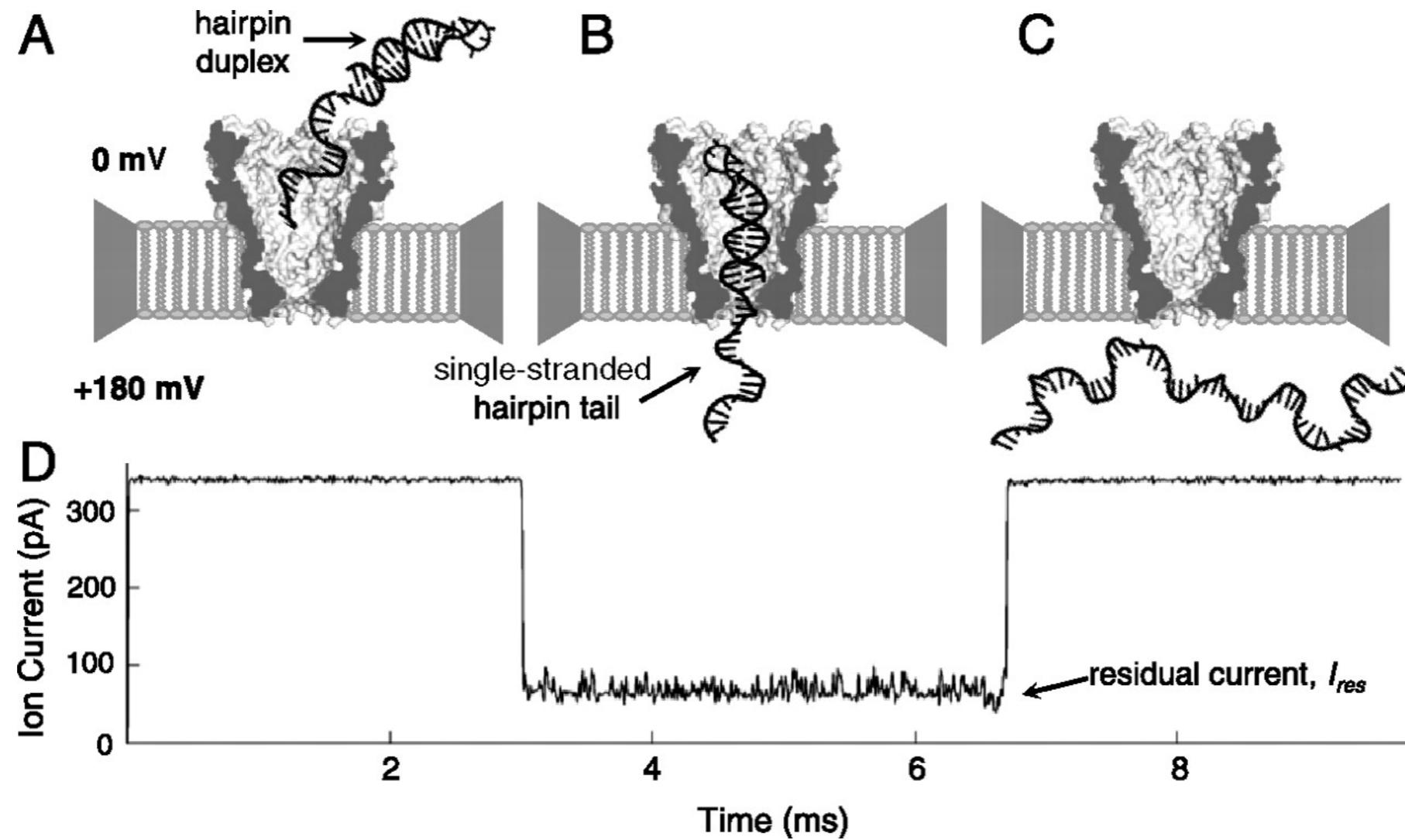


The nucleotide compliments several bases in a row - multiple hydrogen ions are released.

IonTorrent - secuenciamiento por síntesis



Nanopore - hacer pasar DNA por un poro



PacBio vs. Nanopore

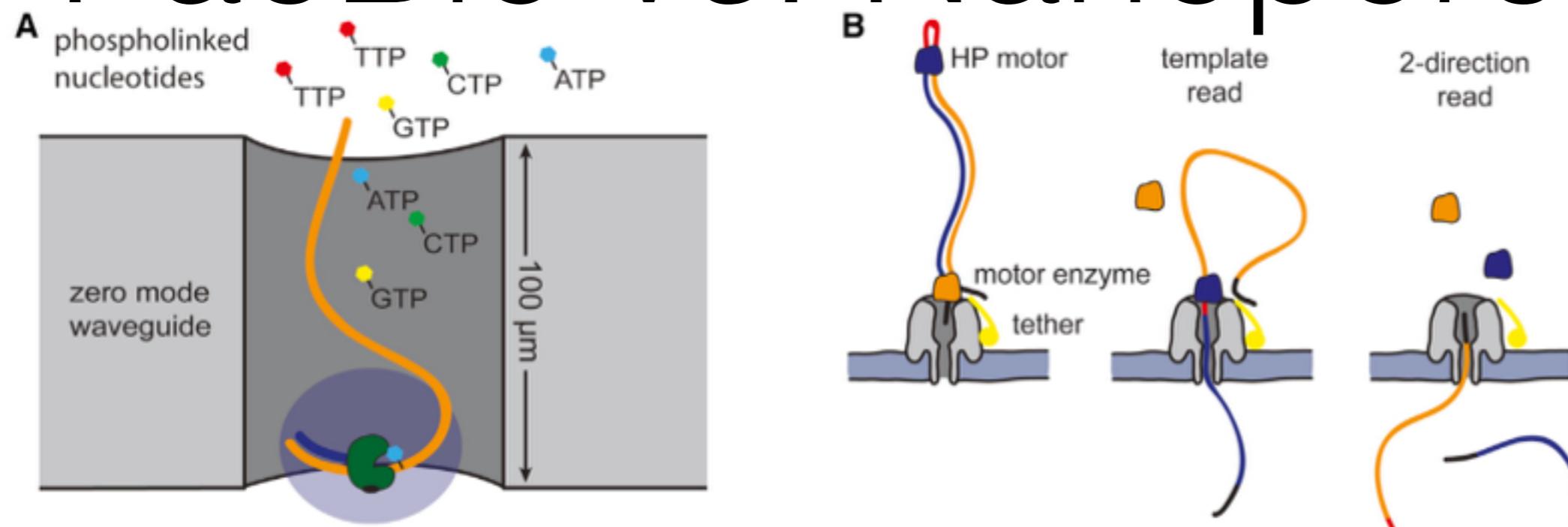
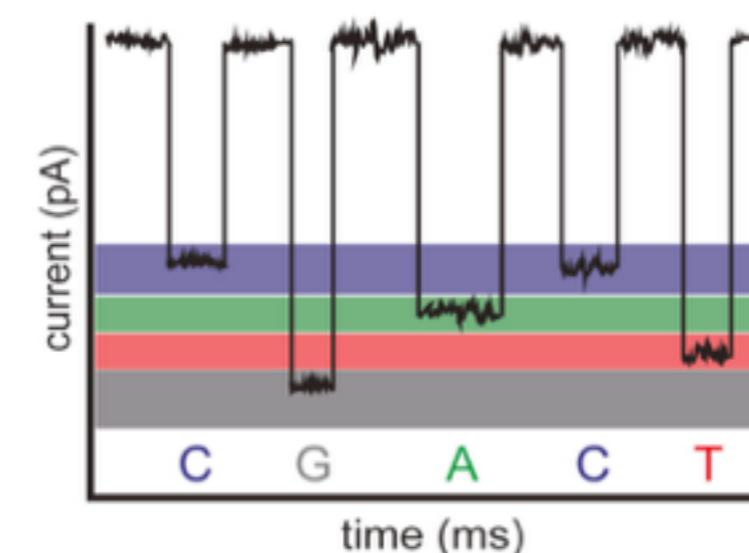
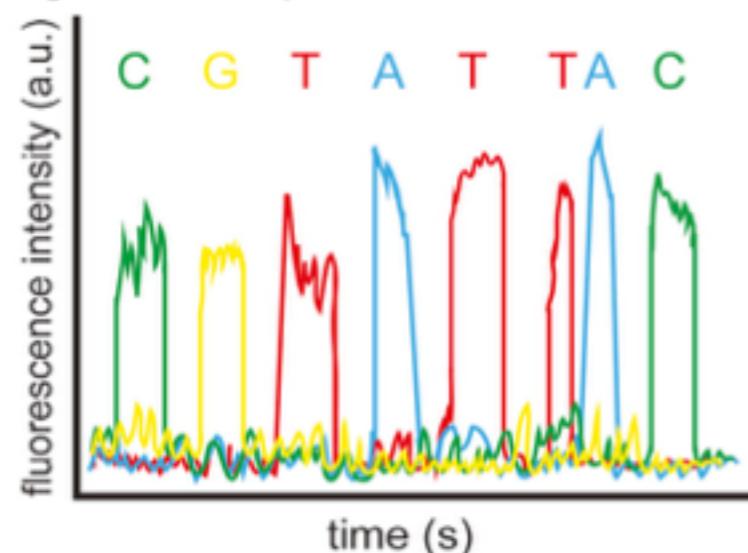


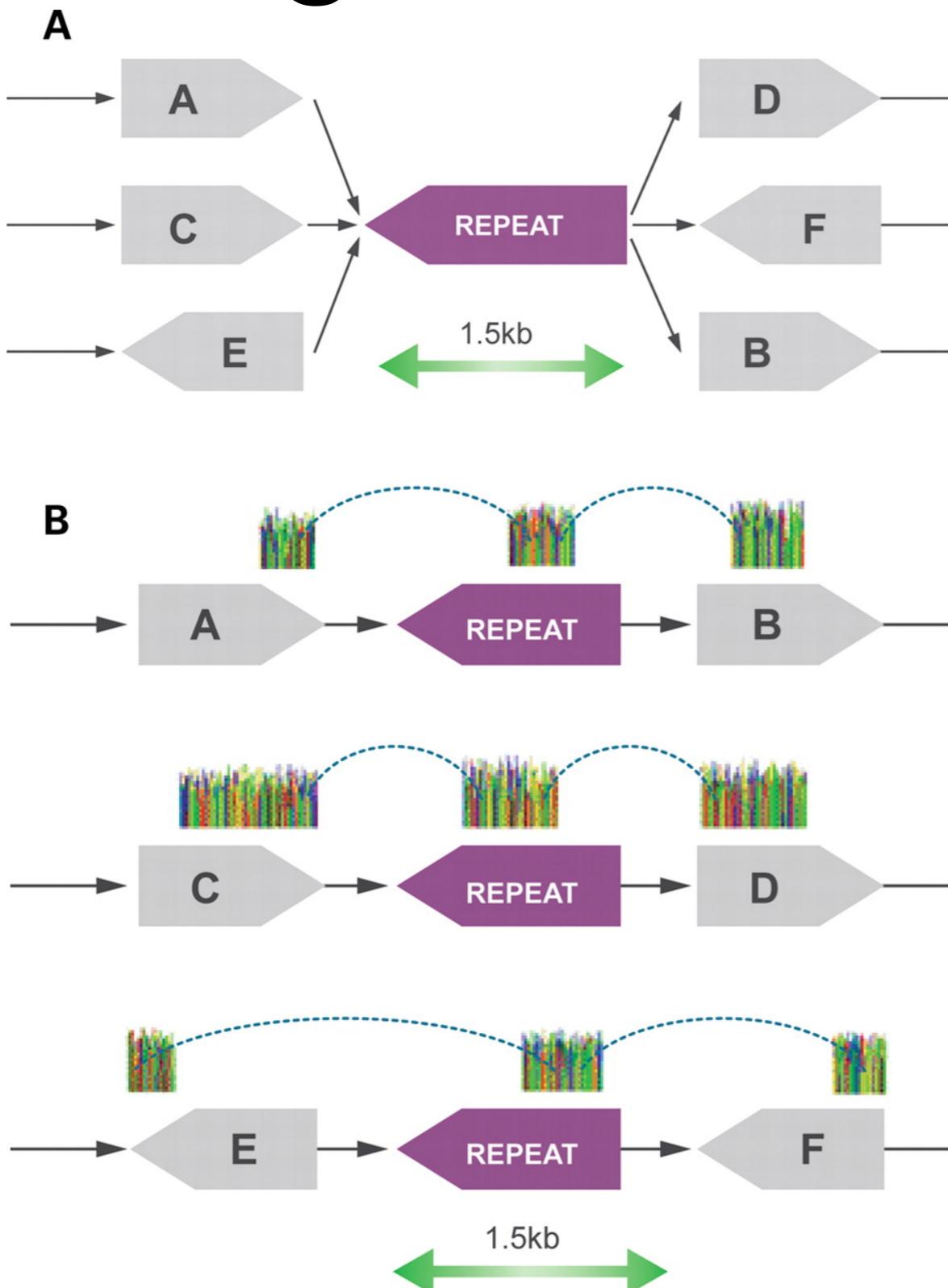
Figure 3. Single Molecule Sequencing Platforms

(A) Pacific Bioscience's SMRT sequencing. A single polymerase is positioned at the bottom of a ZMW. Phosphate-labeled versions of all four nucleotides are present, allowing continuous polymerization of a DNA template. Base incorporation increases the residence time of the nucleotide in the ZMW, resulting in a detectable fluorescent signal that is captured in a video.



(B) Oxford Nanopore's sequencing strategy. DNA templates are ligated with two adaptors. The first adaptor is bound with a motor enzyme as well as a tether, whereas the second adaptor is a hairpin oligo that is bound by the HP motor protein. Changes in current that are induced as the nucleotides pass through the pore are used to discriminate bases. The library design allows sequencing of both strands of DNA from a single molecule (two-direction reads).

Fragmentos o “reads” largas

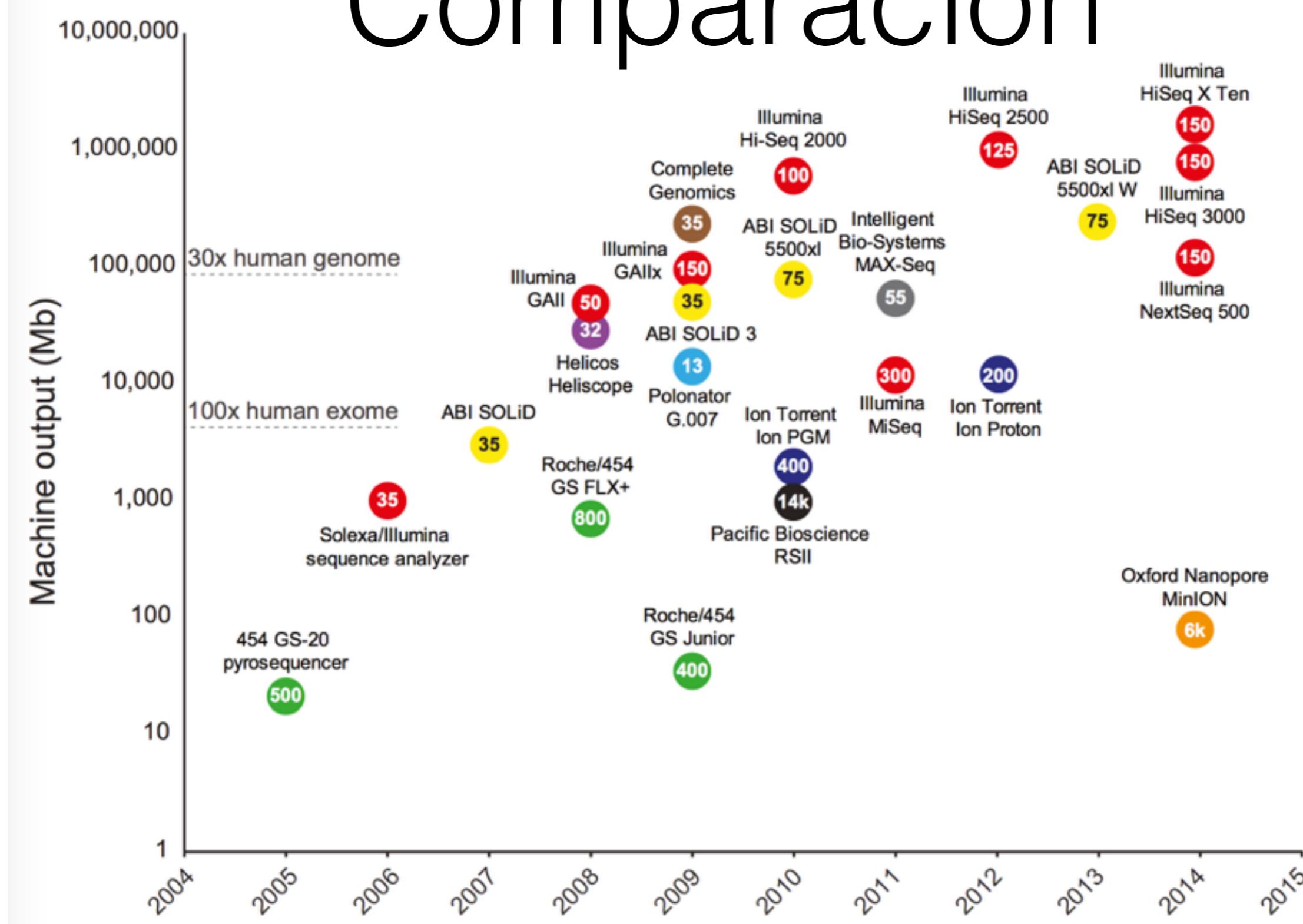


- Ordenar scaffolds
- Resolver repeticiones
- Resolver genes multicopia
- Finalizar genomas

Comparación

	First generation	Second generation ^a	Third generation ^a
Fundamental technology	Size-separation of specifically end-labeled DNA fragments, produced by SBS or degradation	Wash-and-scan SBS	SBS, by degradation, or direct physical inspection of the DNA molecule
Resolution	Averaged across many copies of the DNA molecule being sequenced	Averaged across many copies of the DNA molecule being sequenced	Single-molecule resolution
Current raw read accuracy	High	High	Moderate
Current read length	Moderate (800–1000 bp)	Short, generally much shorter than Sanger sequencing	Long, 1000 bp and longer in commercial systems
Current throughput	Low	High	Moderate
Current cost	High cost per base Low cost per run	Low cost per base High cost per run	Low-to-moderate cost per base Low cost per run
RNA-sequencing method	cDNA sequencing	cDNA sequencing	Direct RNA sequencing and cDNA sequencing
Time from start of sequencing reaction to result	Hours	Days	Hours
Sample preparation	Moderately complex, PCR amplification not required	Complex, PCR amplification required	Ranges from complex to very simple depending on technology
Data analysis	Routine	Complex because of large data volumes and because short reads complicate assembly and alignment algorithms	Complex because of large data volumes and because technologies yield new types of information and new signal processing challenges
Primary results	Base calls with quality values	Base calls with quality values	Base calls with quality values, potentially other base information such as kinetics

Comparación



Comparación

Instrument	Primary Errors	Single-pass Error Rate (%)	Final Error Rate (%)
3730xl (capillary)	substitution	0.1-1	0.1-1
454 All models	indel	1	1
Illumina All Models	substitution	~0.1	~0.1
Ion Torrent – all chips	Indel	~1	~1
SOLiD – 5500xl	A-T bias	~5	≤0.1
Oxford Nanopore	deletions	≥4*	4*
PacBio RS	Indel	~13	≤1

Comparación



News and commentary for ecology, evolution,
and everything in between

[Home](#) [How to ...](#) [Book reviews](#) [NGS Field Guide](#) [Support TME](#) [About](#)

2014 NGS Field Guide: Overview

These pages update the tables presented in [Travis Glenn's \(2011\)](#) "Field Guide to Next Generation DNA Sequencers" for 2014 values. Previous years' tables have been archived: [2011](#), [2012](#), and [2013](#).

Próxima semana...

...estrategias para la
construcción de genotecas
o librerías de DNA