



Fundamentos y herramientas bioinformáticas para análisis genómicos



Institut Pasteur
de Montevideo



AGENCIA NACIONAL
DE INVESTIGACIÓN
E INNOVACIÓN



COMISIÓN SECTORIAL DE
INVESTIGACIÓN CIENTÍFICA

Profesores

Locales:

Luisa Berná

Tamara Fernández

Pablo Fresia

Martín Graña

Daniela Megrian

Hugo Naya

Natalia Rego

Camila Simoes

Lucía Spangenberg

Del Exterior:

Ana Conesa - Investigadora principal del Consejo Superior de Investigaciones Científicas (CSIC) (España)

Gonzalo Bello - Investigador de Salud Pública en el Instituto Oswaldo Cruz, FIOCRUZ (Brasil)

Unidad de Bioinformática



Cronograma

Duración (minutos)	Horario	Lunes	Martes	Miércoles	Jueves	Viernes
45	9:00 - 9:45	Bienvenida - Introducción al curso (Luisa Berná)	Introducción a la Transcriptómica (Natalia Rego)	Llamado de variantes genéticas (Lucía Spangenberg)	Genómica comparativa (Daniela Megrian)	Evolución en contexto: estructura y función de proteínas (Martín Graña)
45	9:45 - 10:30	Ensamblaje genómico (Luisa Berná)	Transcriptómica de reads largos - parte I (Ana Conesa)	Llamado de variantes genéticas (Lucía Spangenberg)	Principios de filogenia (Pablo Fresia)	Predicción y análisis de estructuras de proteínas (Martín Graña)
30	10:30 - 11:00	Coffee break	Coffee break	Coffee break	Coffee break	Coffee break
45	11:00 - 11:45	Anotación genómica (Ana Conesa)	Transcriptómica de reads largos - parte II (Ana Conesa)	Aplicaciones del llamado de variantes en genómica humana (Lucía Spangenberg)	Principios de filogeografía (Gonzalo Bello)	Evolución en contexto: estructura y función de proteínas
45	11:45 - 12:30	Anotación funcional (Ana Conesa)	Transcriptómica de reads largos	Ana Conesa (Salón de Actos - Facultad de Ciencias - 12:00)	Principios de filogeografía (Gonzalo Bello)	Visualización de modelos estructurales
60	12:30 - 13:30	Almuerzo	Almuerzo		Almuerzo	Almuerzo
60	13:30 - 14:30	Control de calidad de reads	Transcriptómica de reads largos	Almuerzo	Genómica comparativa y filogenia	Gonzalo Bello (Salón de Actos PA - Institut Pasteur de Montevideo)
45	14:30 - 15:15	Ensamblaje de novo (reads cortos y largos)	Transcriptómica de reads largos	Alineamiento/Variant calling	Genómica comparativa y filogenia	Cierre del curso
45	15:15 - 16:00	Evaluación de ensamblajes	Transcriptómica de reads largos	Variant calling	Genómica comparativa y filogenia	Cierre del curso
30	16:00 - 16:30	Coffee break	Coffee break	Coffee break	Coffee break	Coffee break
45	16:30 - 17:15	Anotación génica	Transcriptómica de reads largos	Anotacion de variantes	Filogeografía	Evaluación
45	17:15 - 18:00	Anotación funcional	Transcriptómica de reads largos	Interpretacion de variantes	Filogeografía	Evaluación
	20:00					Festejo cierre del curso en Oso Pardo

Estudiantes

Ana Altieri
Antonella D'Anatro
Camila Ciuffo
Carmela Pereyra
Catalina Baserga
Cecilia Monesiglio
Daniela Alí Ruiz
Eugenia Fernández
Felipe Calandra
Ileana Sosa
Joaquín Dalla Rizza
Josefina Bonomi

Karen Pava
Laura Moreno
Lucia Borra
Lucía Sosa
Maria Rego e Silva
Mariana Marchesano
Mariana Chaves
Pedro Dias
Rodrigo Sologaistoa
Rossina Novas
Sebastián Rey
Ximena Simón

Organismos y enfoques en el curso

A lo largo del curso trabajaremos con distintos organismos modelo y tipos de datos, para explorar la diversidad de estrategias genómicas:



Ensamblaje genómico: *Saccharomyces cerevisiae* (levadura)



Transcriptómica: líneas celulares humanas



Variant calling: cromosoma humano



Genómica comparativa y filogenias: bacterias



Filogeografía: virus



Estructuras y filogenias : Arch/Euk/Bac/virus



Fundamentos y
herramientas
bioinformáticas
para análisis
genómicos

Introducción

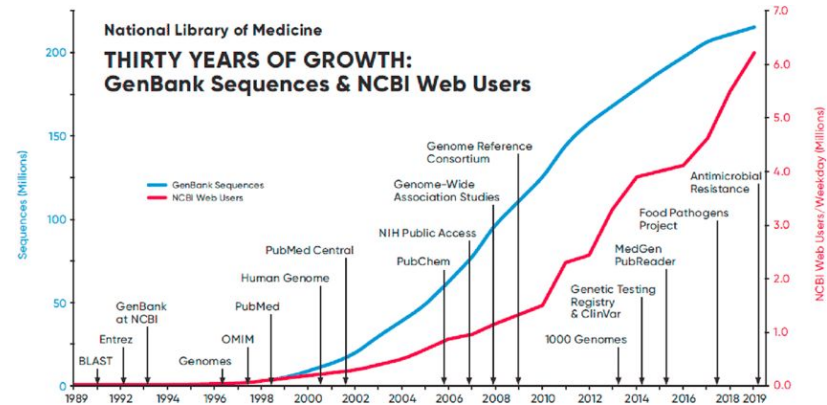
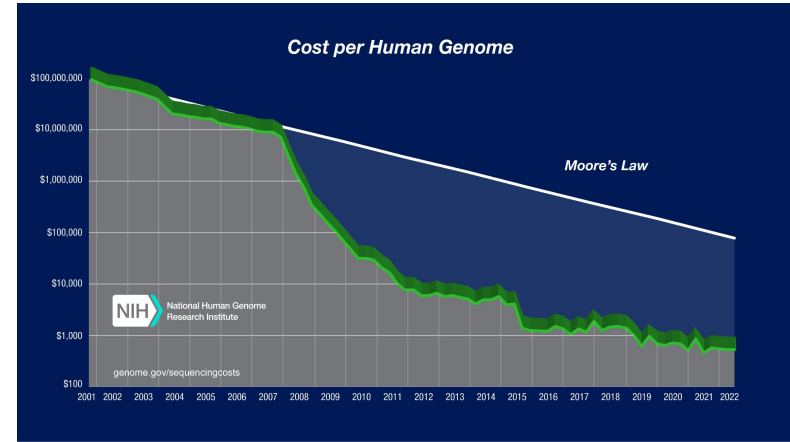
De los datos biológicos a la información genómica.

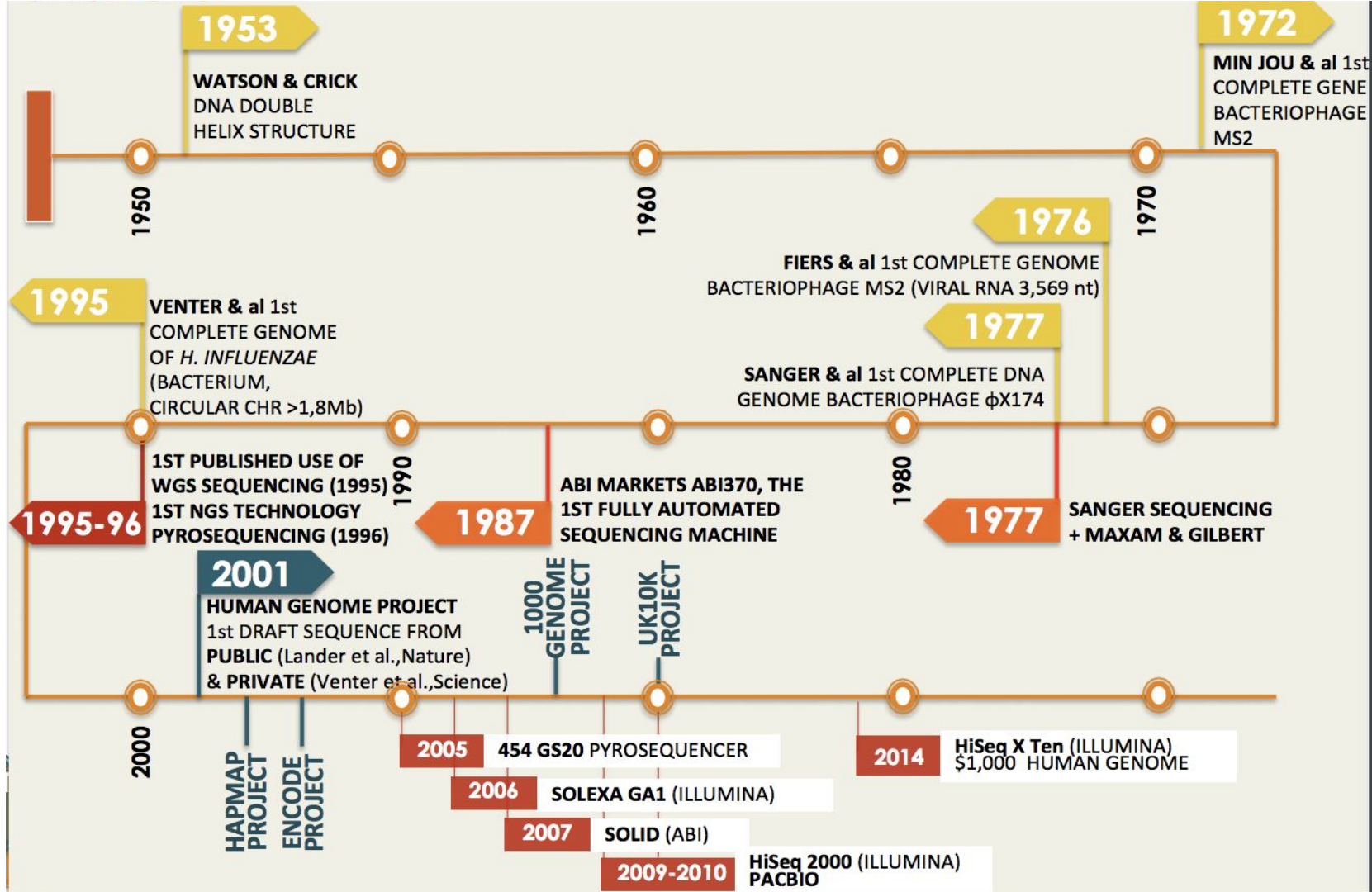
Un recorrido desde la secuenciación hasta el análisis bioinformático.



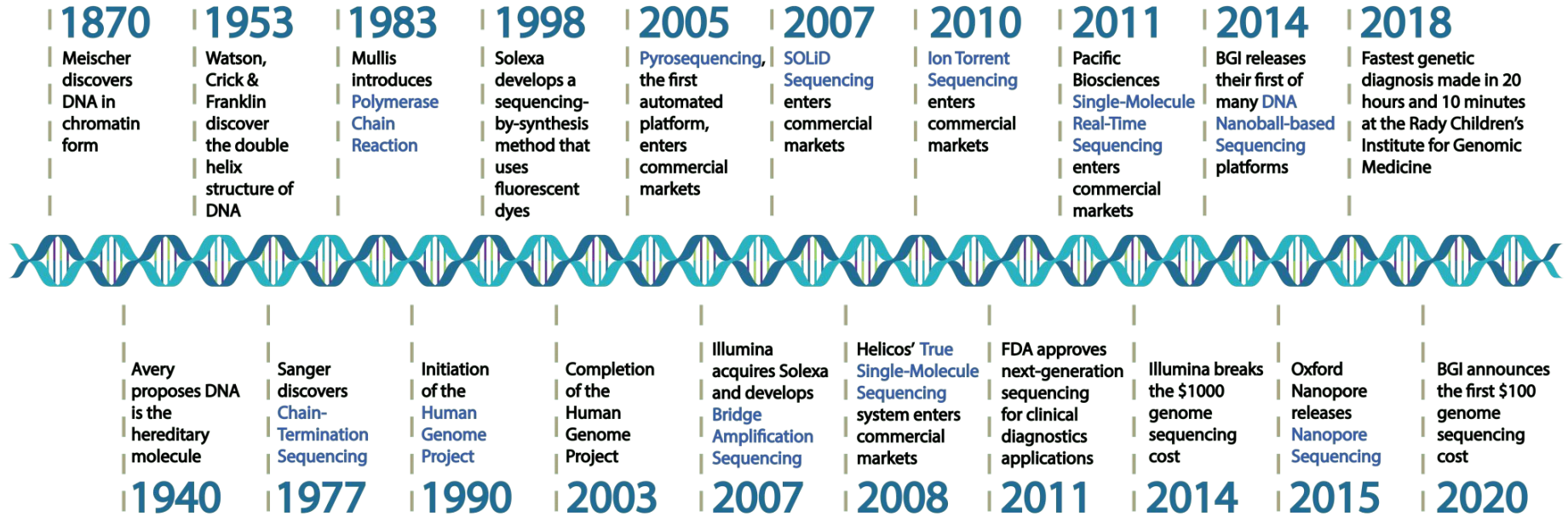
¿Por qué hacemos este curso?

- Cada experimento genera millones de lecturas genómicas.
- El desafío ya no es generar datos, sino entenderlos.
- La bioinformática conecta la biología molecular con la computación.
- Este curso busca darles las herramientas para analizar, interpretar y visualizar datos genómicos reales.





Desarrollos tecnológicos e hitos



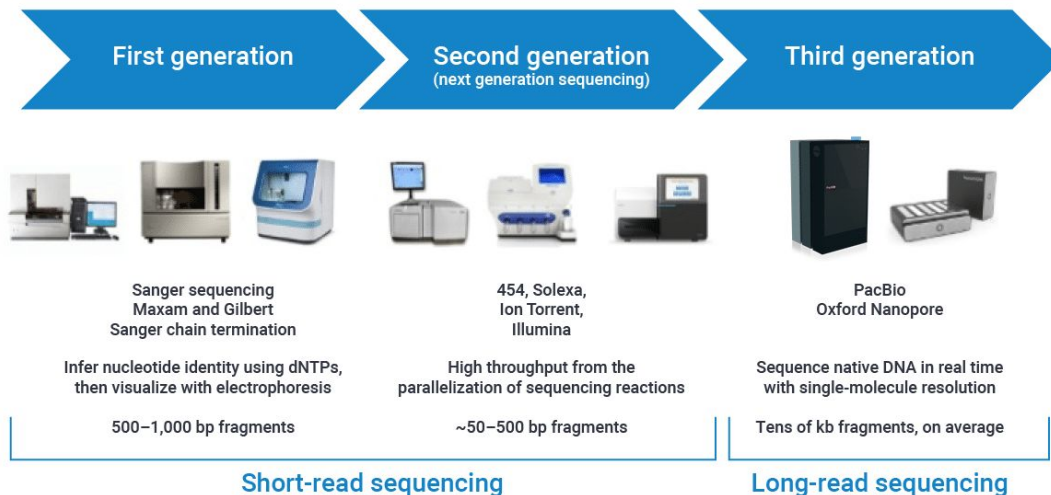
Objetivos del curso

- Aprender las bases teóricas y prácticas del análisis genómico.
- Usar herramientas reales y acercamiento a diferentes tipos de datos y preguntas biológicas
- Integrar conocimientos de ensamblado, transcriptómica, variantes y genómica comparativa, filogenética y evolución.
- Obtener un idea general de la potencialidad de la bioinformática en distintas ramas de la biología

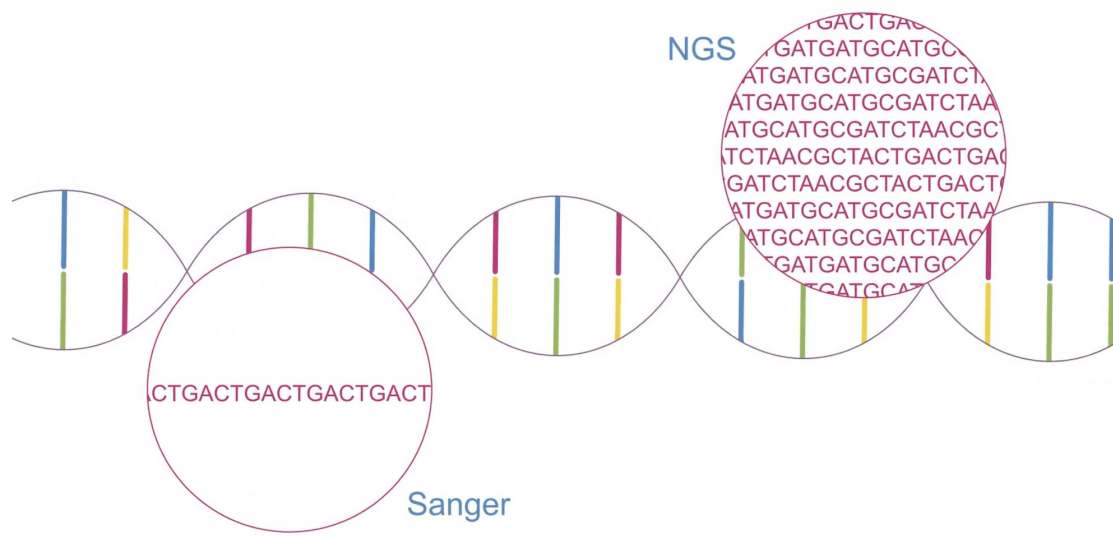


La revolución de la secuenciación

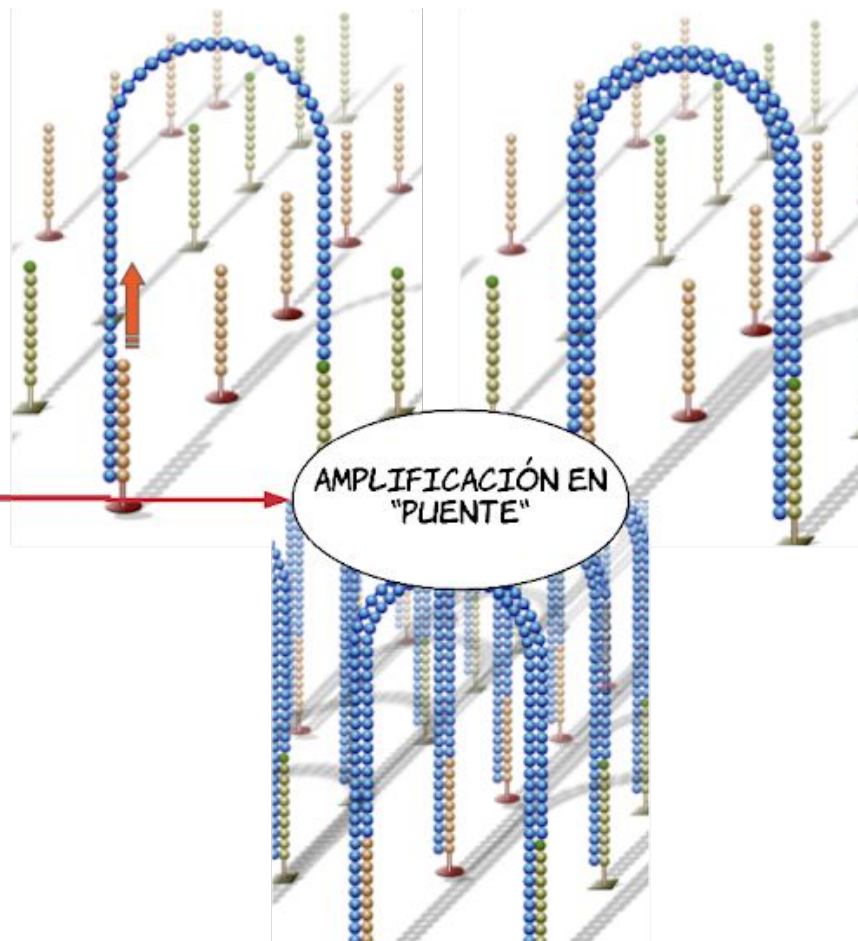
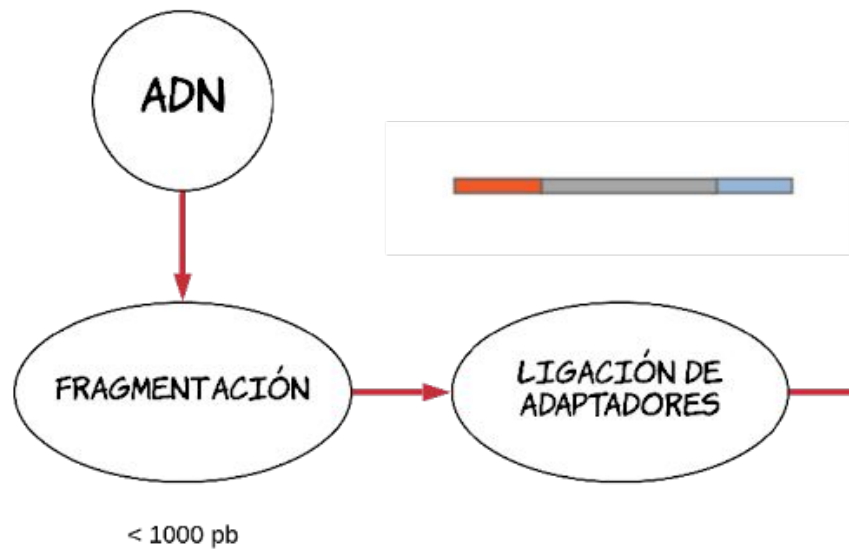
- 2000: el genoma humano costó ~3 mil millones USD
- 2025: un genoma cuesta menos de 500 USD
- Millones de genomas disponibles en bases públicas.
- Nuevas tecnologías → nuevos desafíos analíticos.

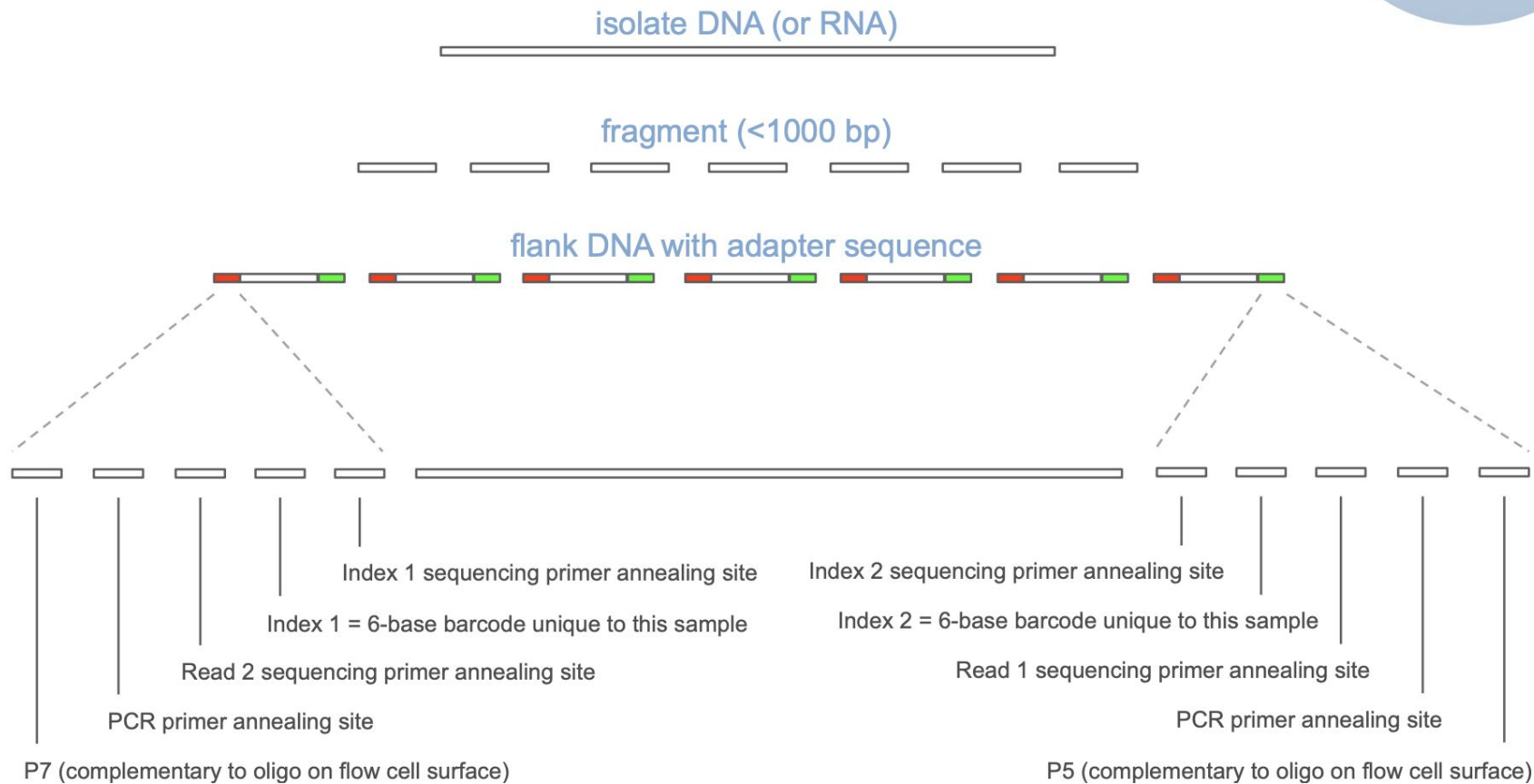


Tecnología Illumina

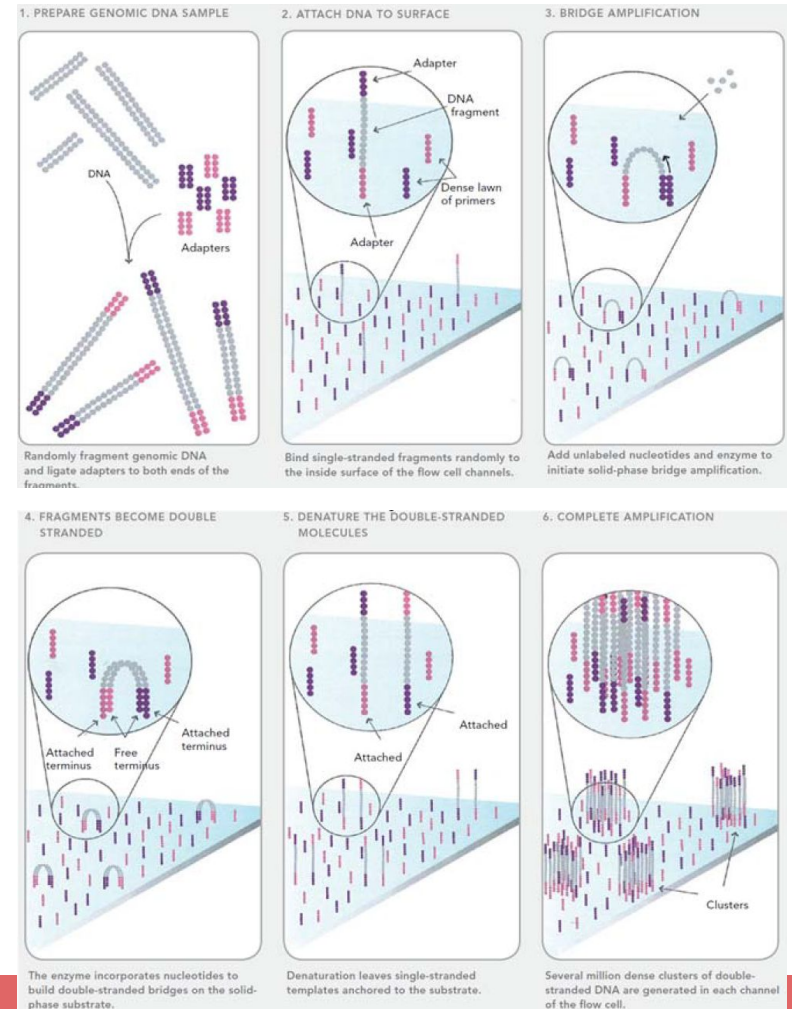


- Fragmentación del ADN + adaptadores.
- Amplificación por “bridge PCR”.
- Secuenciación por síntesis (fluorescencia).
- Alta precisión (99.9%) pero lecturas cortas (100–300 pb).

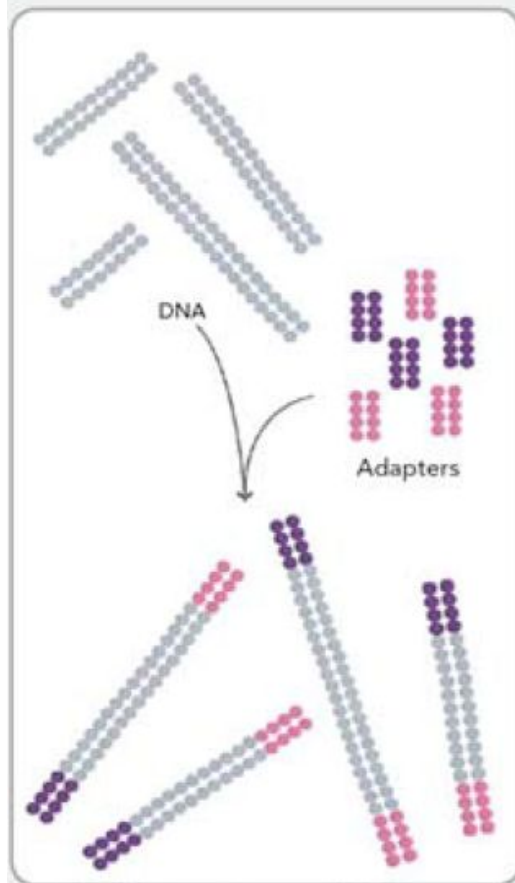




1. Los fragmentos con adaptadores se unen a la *flow cell* por hibridación.
2. El fragmento se curva formando un “puente” y se une con su otro extremo.
3. Una polimerasa sintetiza una copia complementaria.
4. El ADN de doble cadena se desnaturaliza → quedan dos copias unidas a la superficie.
5. El ciclo se repite → se forman **clusters** de millones de copias idénticas del mismo fragmento.

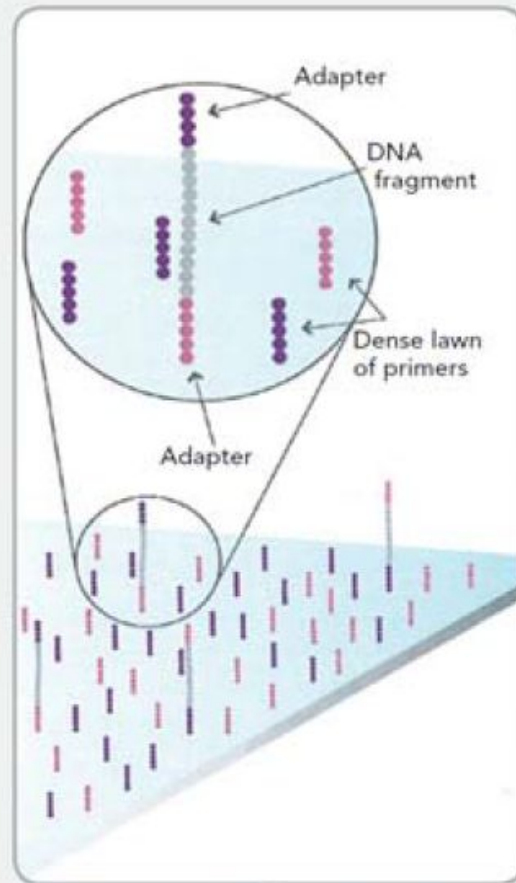


1. PREPARE GENOMIC DNA SAMPLE



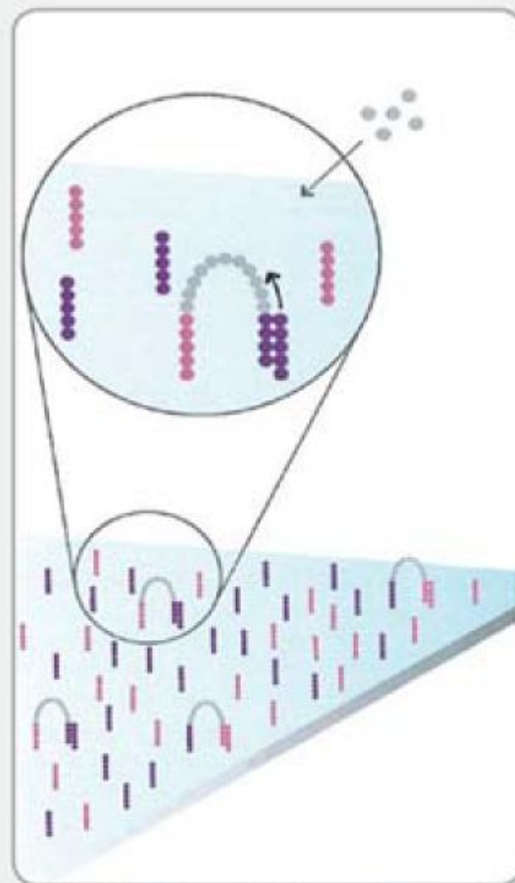
Randomly fragment genomic DNA and ligate adapters to both ends of the fragments.

2. ATTACH DNA TO SURFACE



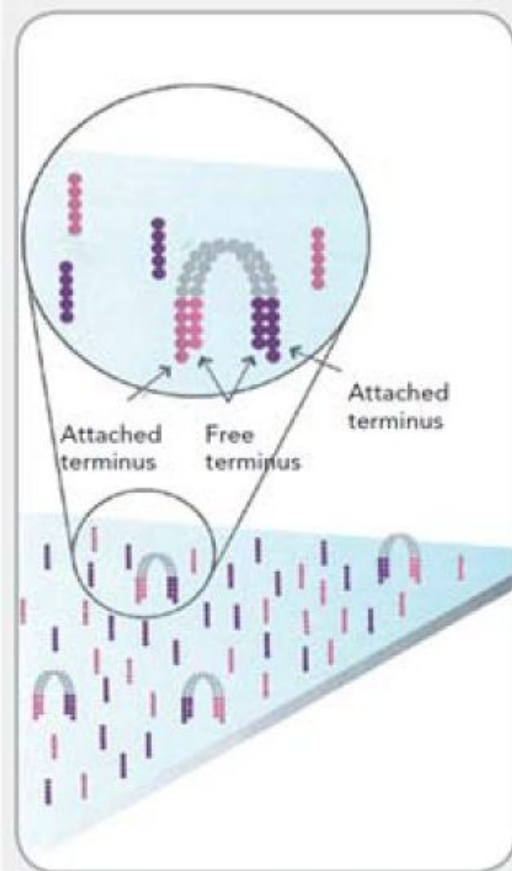
Bind single-stranded fragments randomly to the inside surface of the flow cell channels.

3. BRIDGE AMPLIFICATION



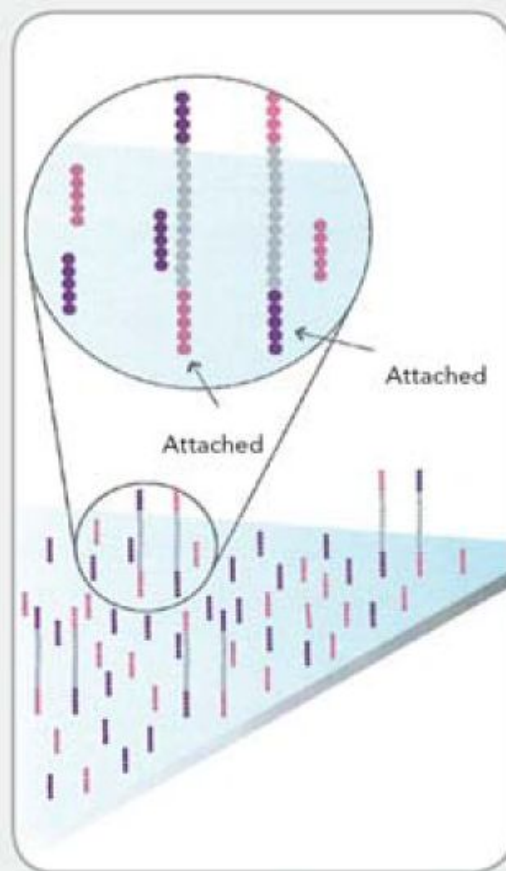
Add unlabeled nucleotides and enzyme to initiate solid-phase bridge amplification.

4. FRAGMENTS BECOME DOUBLE STRANDED



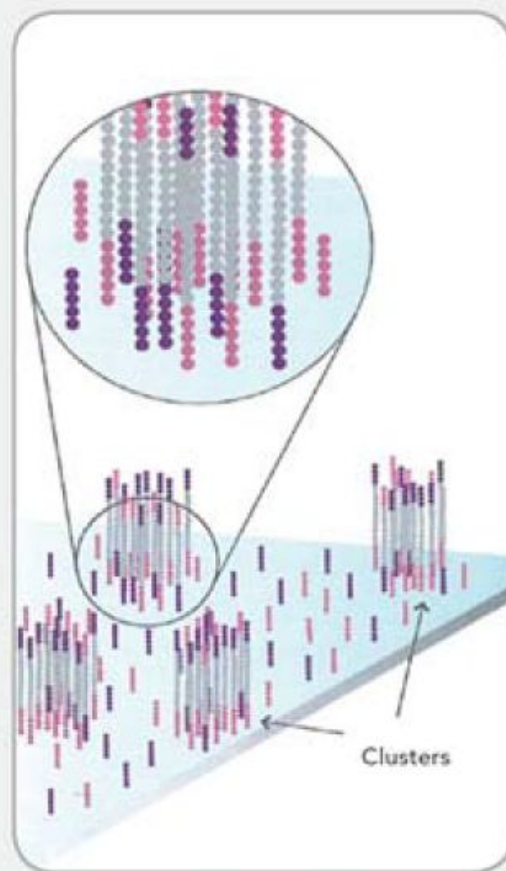
The enzyme incorporates nucleotides to build double-stranded bridges on the solid-phase substrate.

5. DENATURE THE DOUBLE-STRANDED MOLECULES

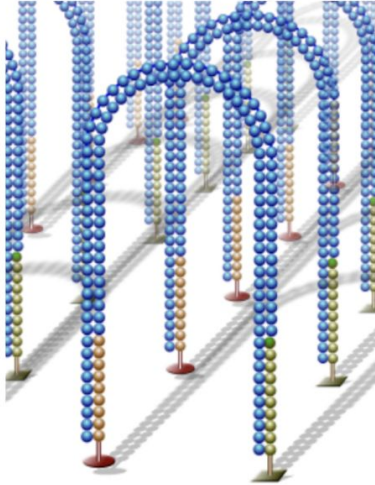


Denaturation leaves single-stranded templates anchored to the substrate.

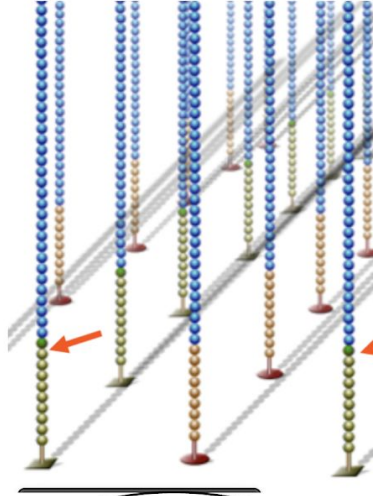
6. COMPLETE AMPLIFICATION



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

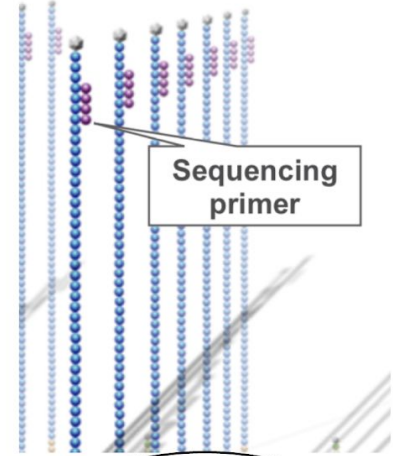


AMPLIFICACIÓN EN
"PUENTE"



LINEALIZACIÓN

CLIVADO DE LA
HEBRA
COMPLEMENTARIA



UNIÓN DEL PRIMER
DE SECUENCIACIÓN

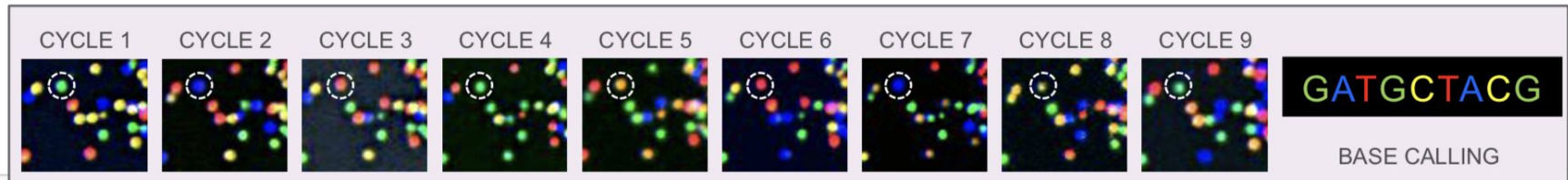
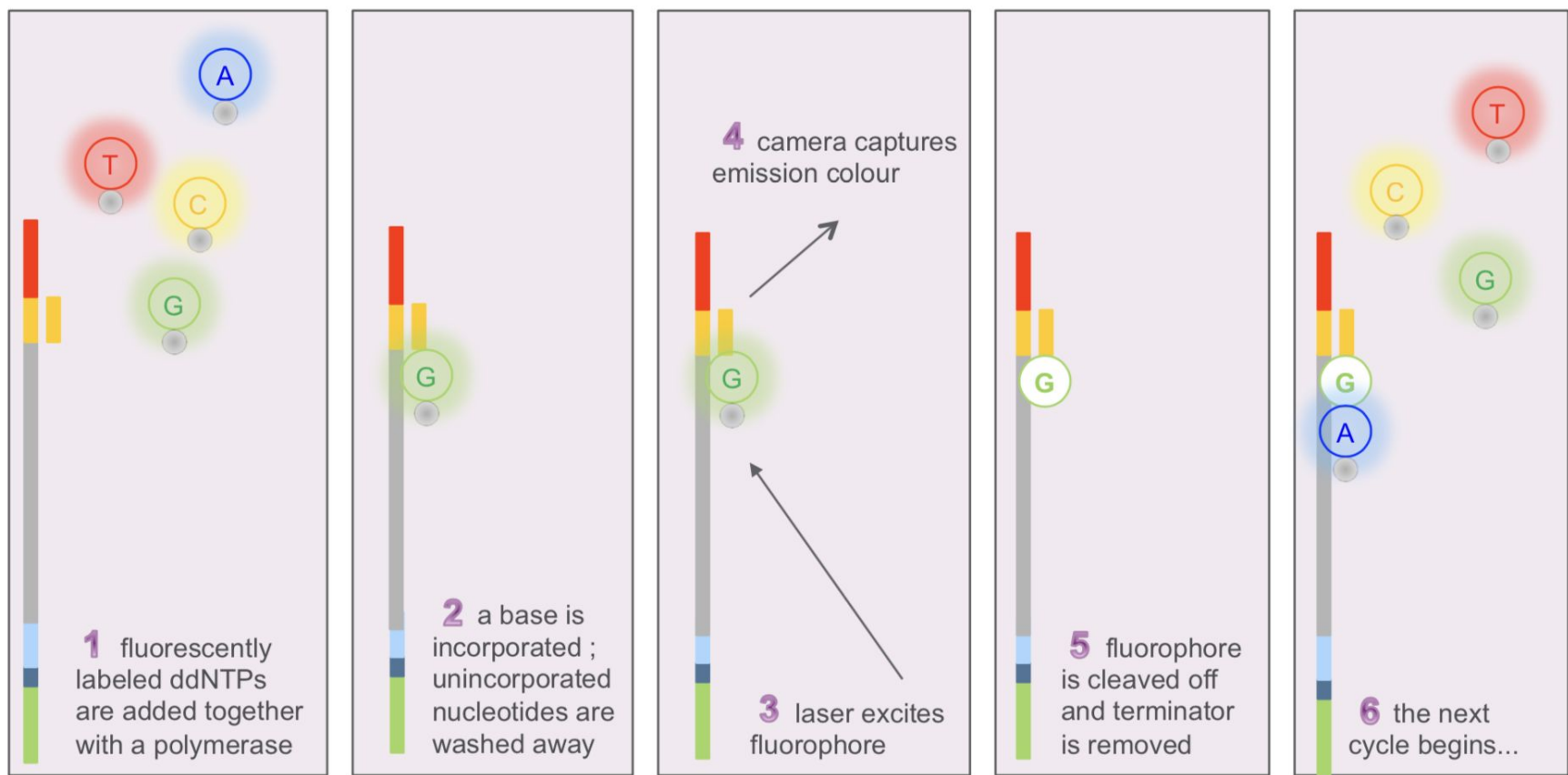
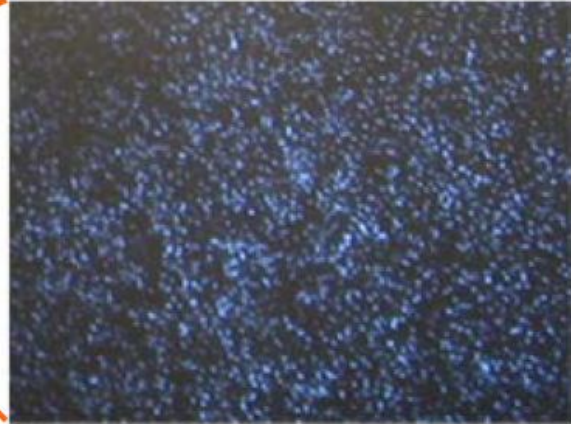


Image analysis



Flow cells



Clusters

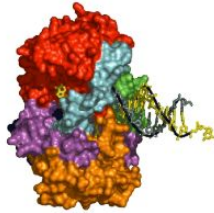
Tecnologías de secuenciación de tercera generación

- PacBio
 - Oxford Nanopore
- + Secuencias MUCHO más largas
 - + No presentan sesgo de bases
 - Contienen más errores
Nuevos algoritmos desarrollados

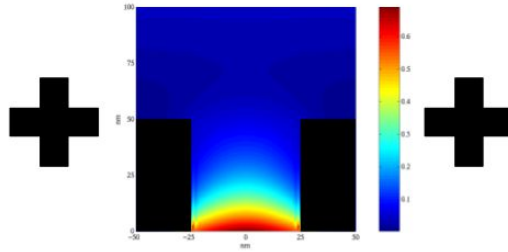
15Kb - 1Mb

SINGLE-MOLECULE, REAL-TIME DNA SEQUENCING (SMRT)

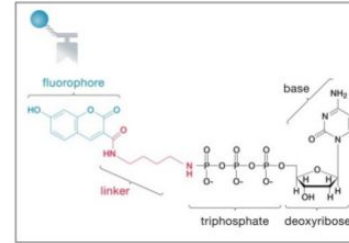
DNA Polymerase



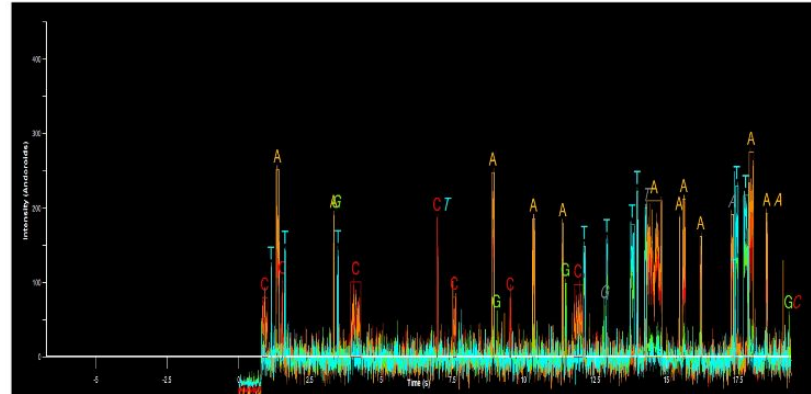
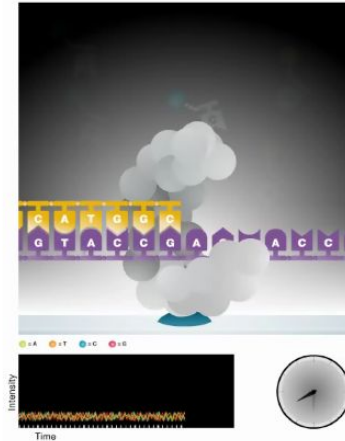
ZMW Confinement



Phospholinked Nucleotides

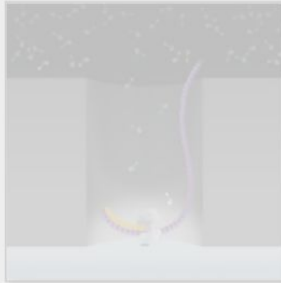


=

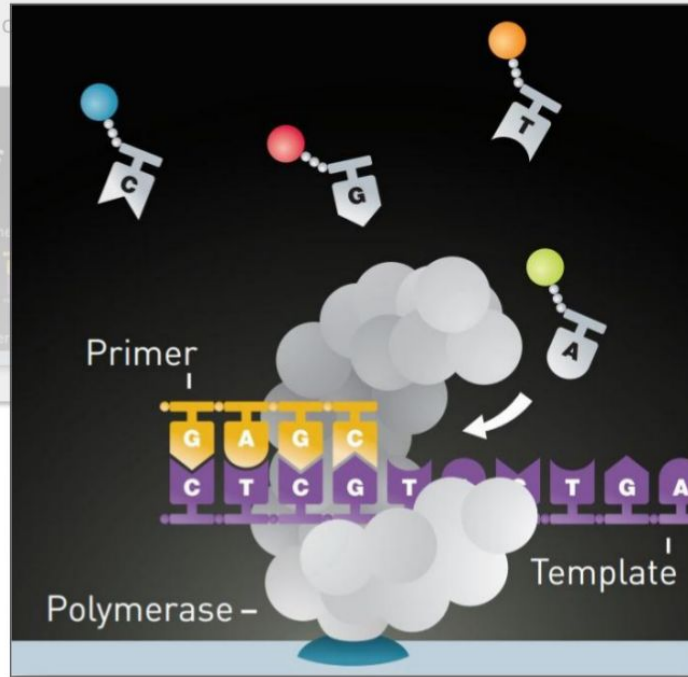


SINGLE MOLECULE, REAL-TIME (SMRT) DNA SEQUENCING

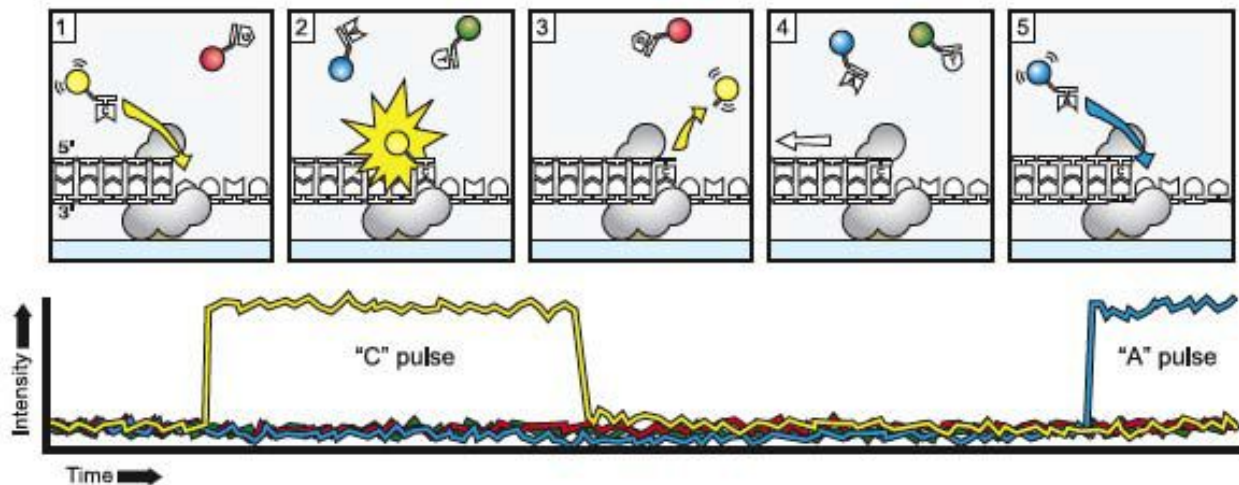
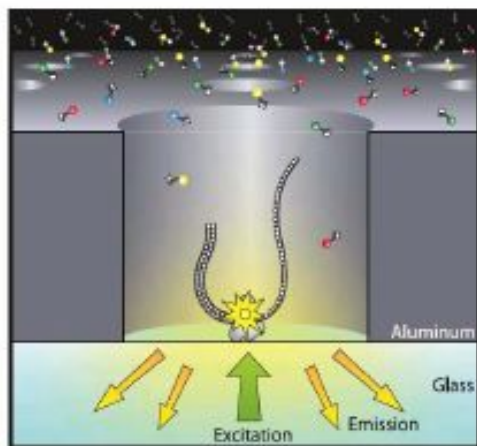
Zero-mode Waveguides



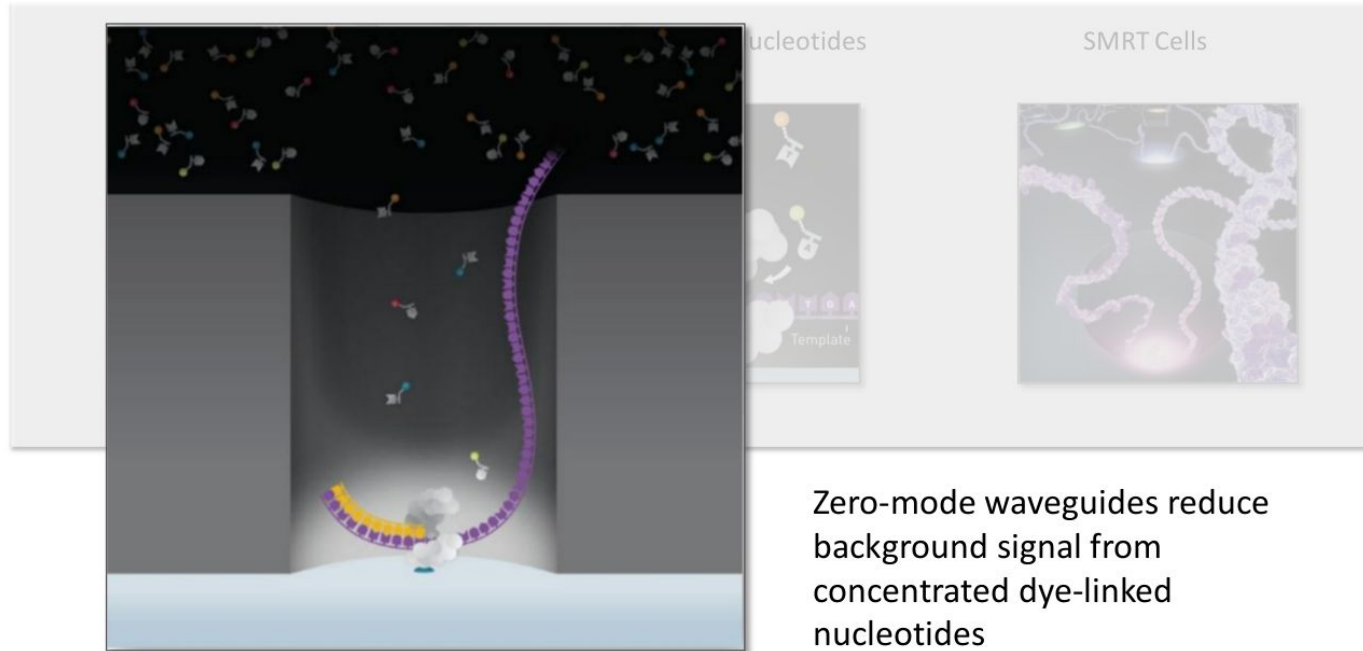
Phosphor



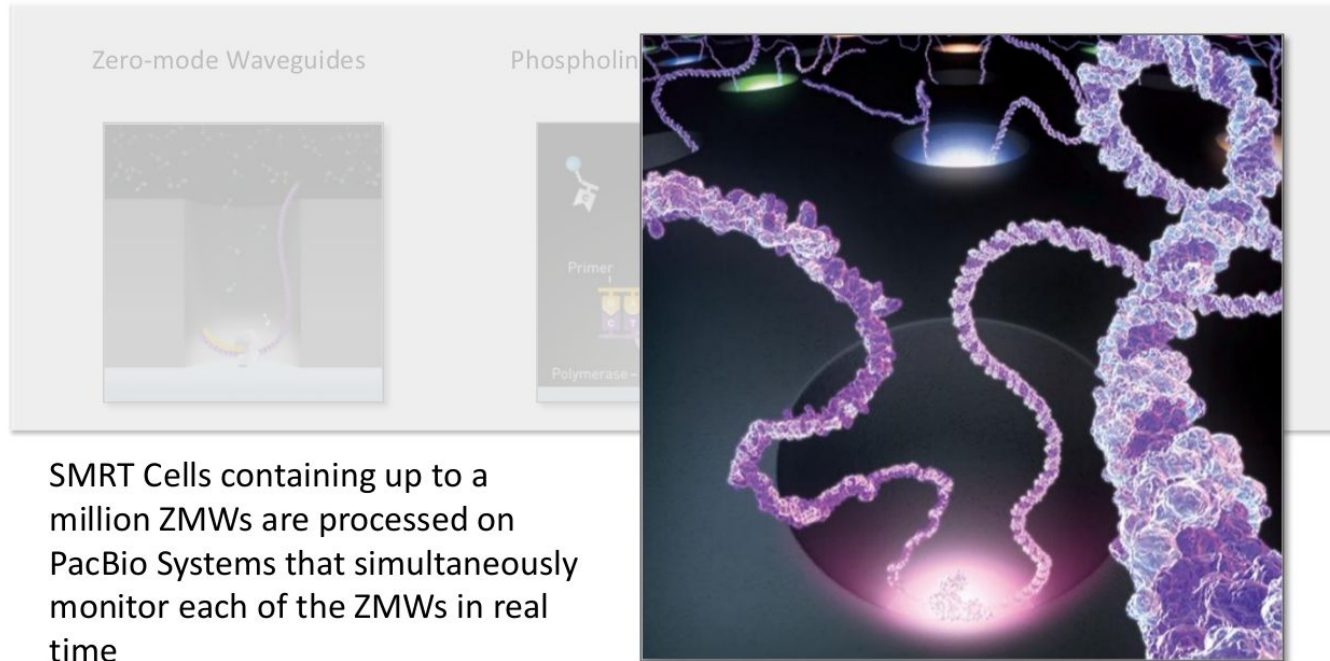
Phospholinked dyes eliminate the steric blockade of DNA extension and background signal that occur after incorporation of nucleotides with base-linked dyes



SINGLE MOLECULE, REAL-TIME (SMRT) DNA SEQUENCING



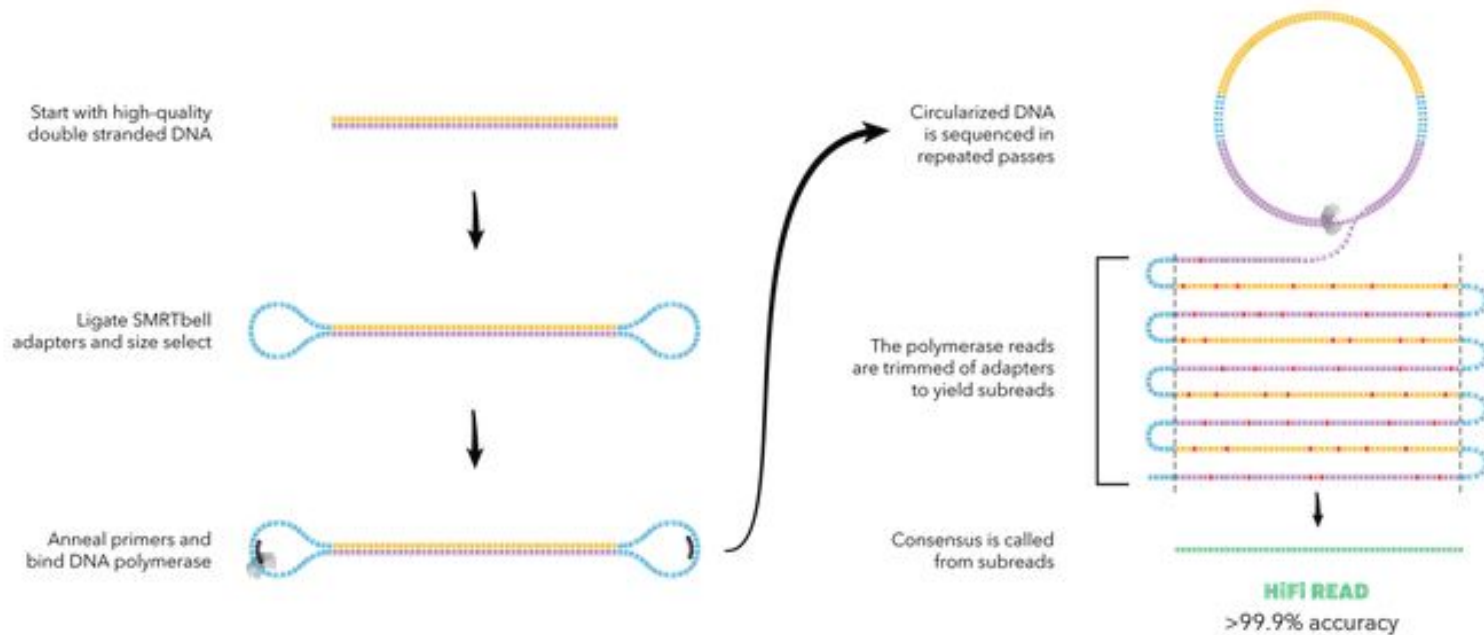
SINGLE MOLECULE, REAL-TIME (SMRT) DNA SEQUENCING



SMRT Cells containing up to a million ZMWs are processed on PacBio Systems that simultaneously monitor each of the ZMWs in real time

PacBio

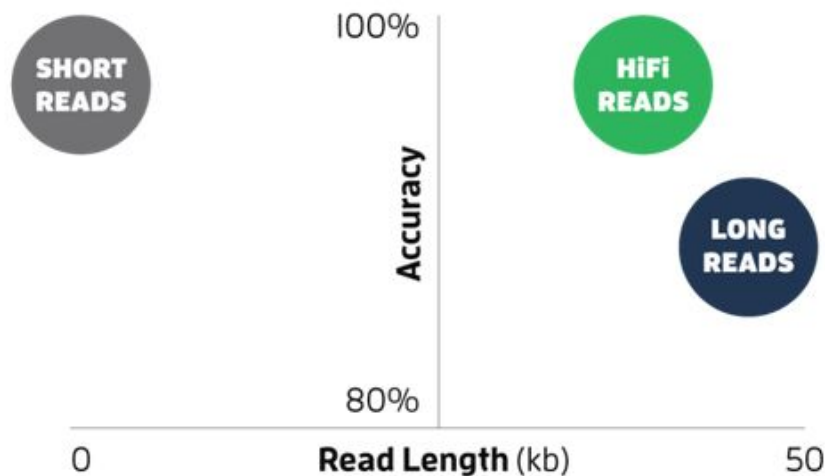
How are HiFi Reads Generated?



Explore A New Paradigm in Sequencing with HiFi Reads

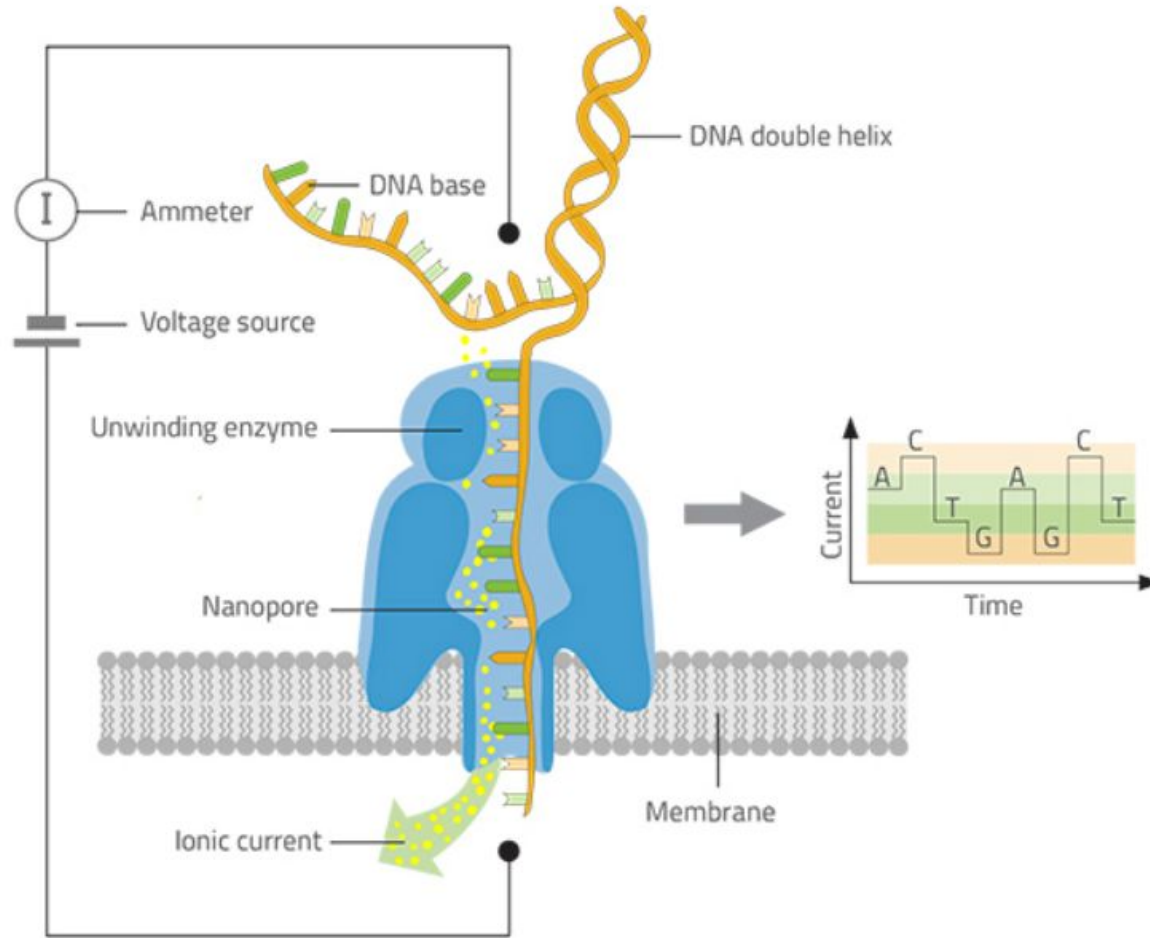
Advanced scientific discoveries require sequencing data that is both accurate and complete.

Single Molecule, Real-Time (SMRT) Sequencing technology has evolved to a different type of long read, known as highly accurate long reads, or HiFi reads.



PacBio is the only sequencing technology to offer HiFi reads that provide accuracy of >99.9%, on par with short reads and Sanger sequencing. With HiFi reads you no longer have to compromise long read lengths for high accuracy sequencing to address your toughest biological questions.

Oxford Nanopore ONT



Oxford Nanopore ONT

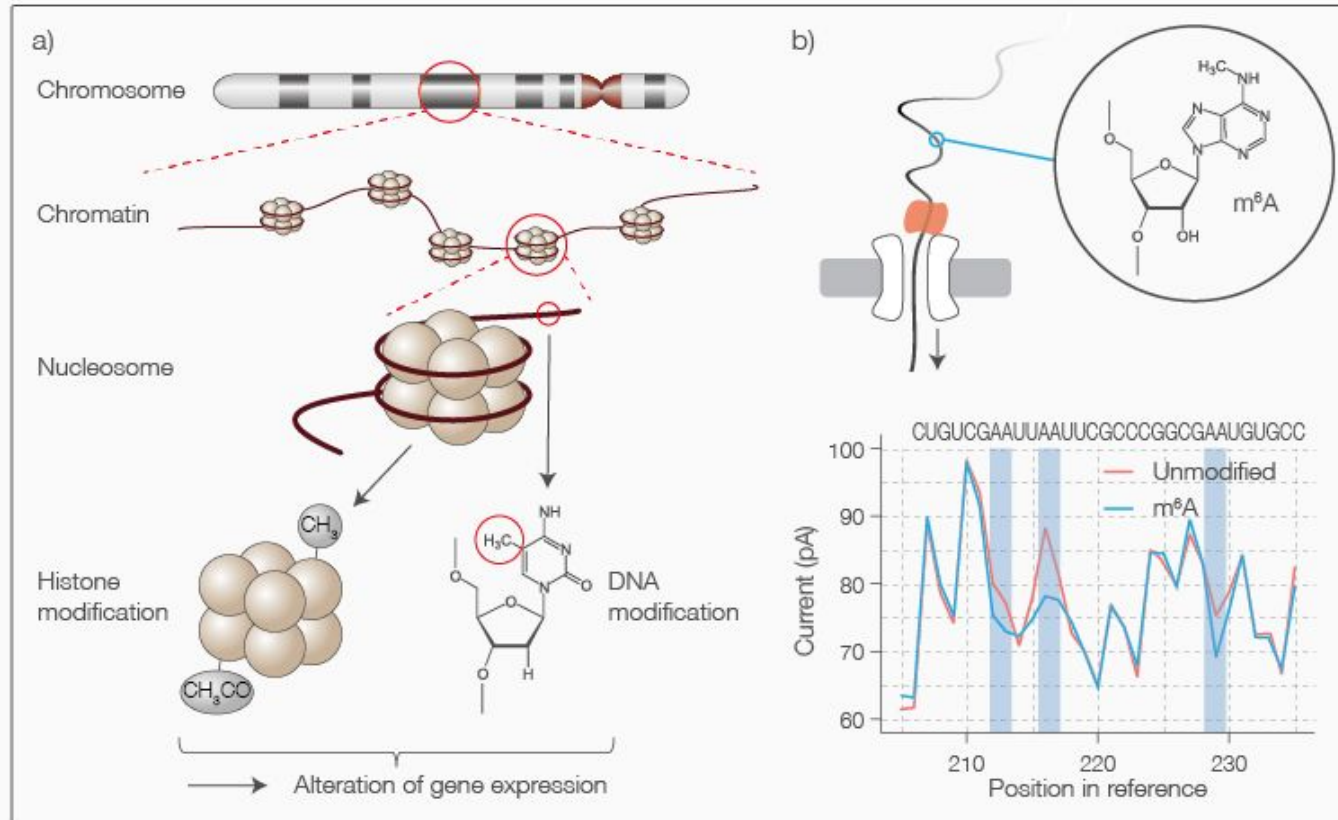


Fig. 1 Modifications a) different epigenetic modifications b) nanopore sequencing of native RNA

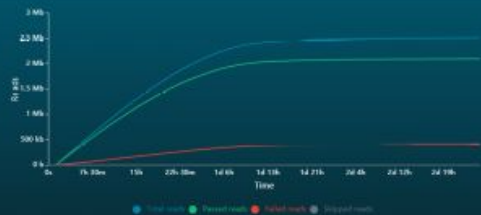
Channel states panel

Run state: sequencing



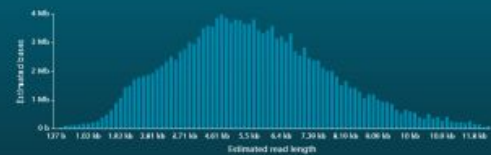
+ Show detailed

Cumulative output



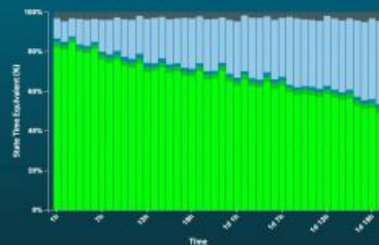
Read length histogram

Estimated NG: 5.35 kb



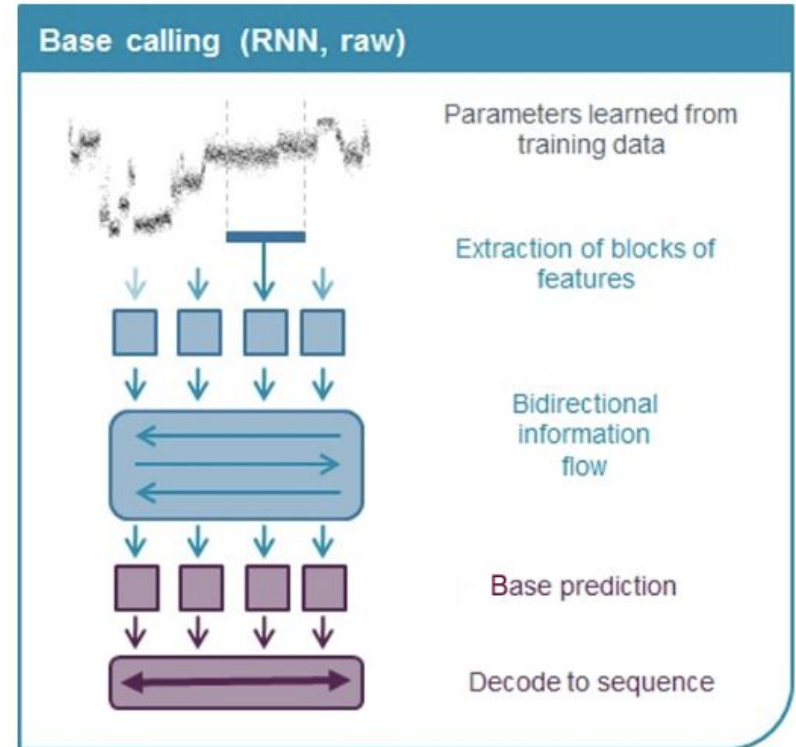
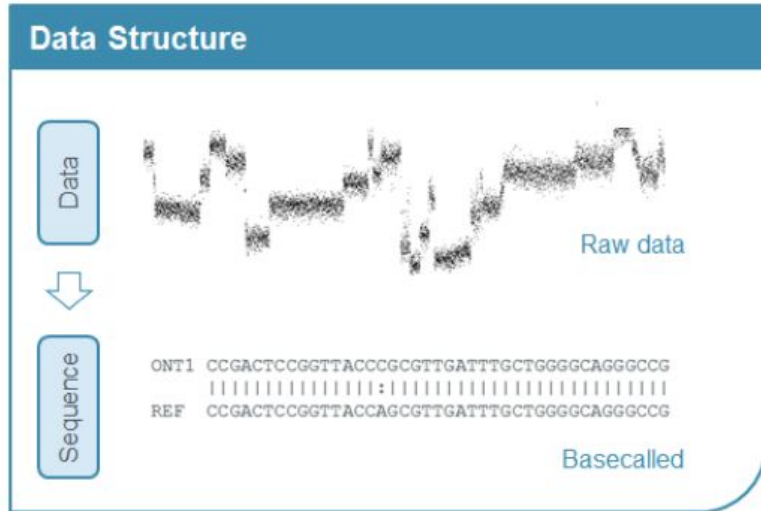
Read Base
Length Counts Estimated Basecalled ☐ Show ☒ Hide outliers ☐ Sort by read end reason

Pore activity

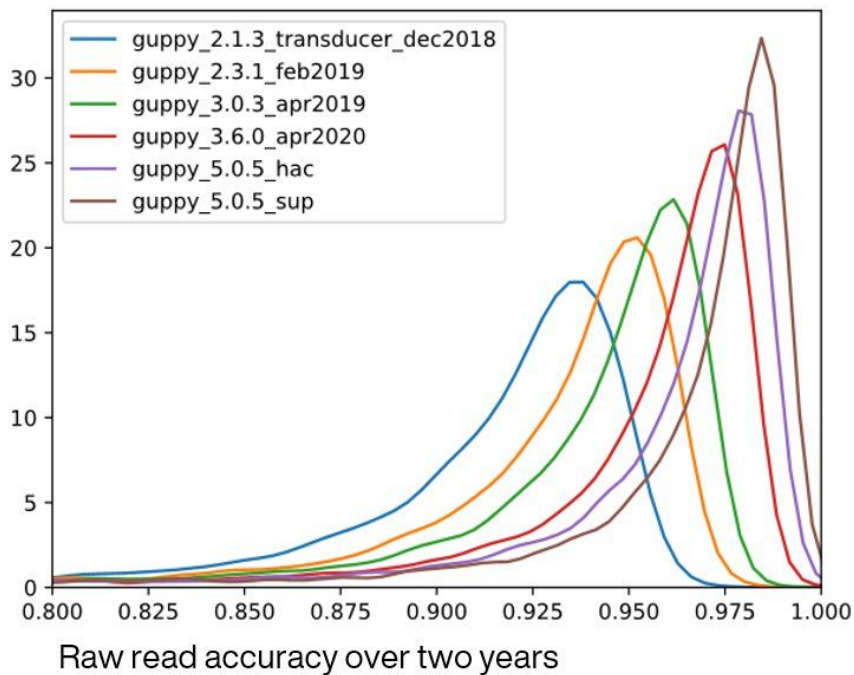


+ Show Detailed ☒ Display Settings

Interpretación de la secuencia



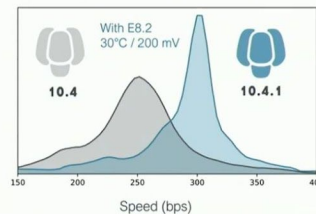
ONT actual - Mejoras en guppy



ONT actual - mejoras en poros y proteína

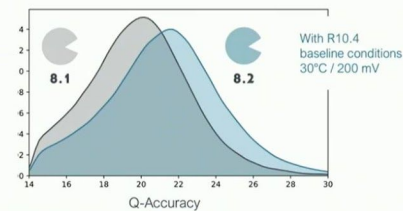
New pore: R10.4.1

- Improved pore designs tune enzyme-pore docking
- Faster speeds (~250-420 bps)
 - Yield much higher output compared to current Q20 chemistry
- Tighter speed distributions
 - Helps to reduce errors

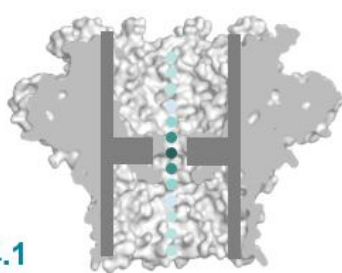


New motor: E8.2

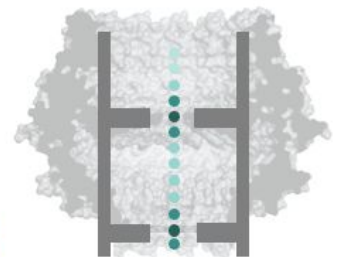
- Improved movement properties with more consistent movement
 - Better defined levels
 - Fewer mis-steps
 - Improved accuracy
- ~Halved error rates.



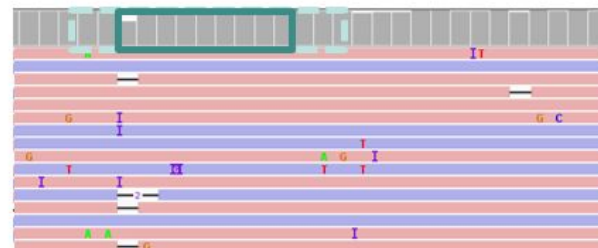
R9.4.1



R10



ATCGGAATAAAATTCACGCCACGTCCAAA



Desarrollo de nuevos protocolos para singe-cell



Single-cell sequencing on MinION

This protocol describes how to carry out sequencing of cDNA from single cells using the PCR-cDNA Sequencing Kit (SQK-PCS111).

[Download >](#)



Single-cell sequencing on GridION

This protocol describes how to carry out sequencing of cDNA from single cells using the PCR-cDNA Sequencing Kit (SQK-PCS111).

[Download >](#)



Single-cell sequencing on PromethION

This protocol describes how to carry out sequencing of cDNA from single cells using the PCR-cDNA Sequencing Kit (SQK-PCS111).

[Download >](#)

por FIN!