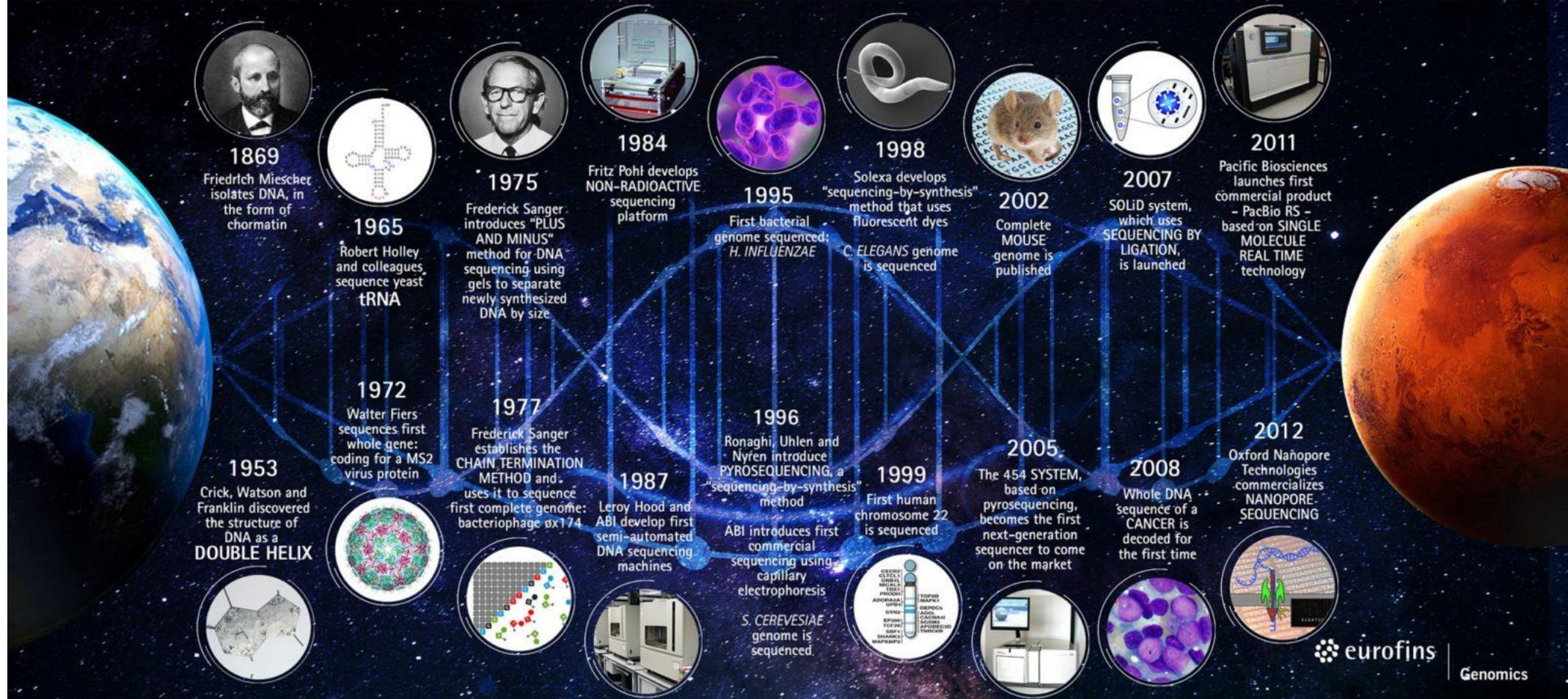


# Secuenciación de genomas

Guillermo Barturen Briñas  
([gbarturen@ugr.es](mailto:gbarturen@ugr.es))

# Perspectiva histórica

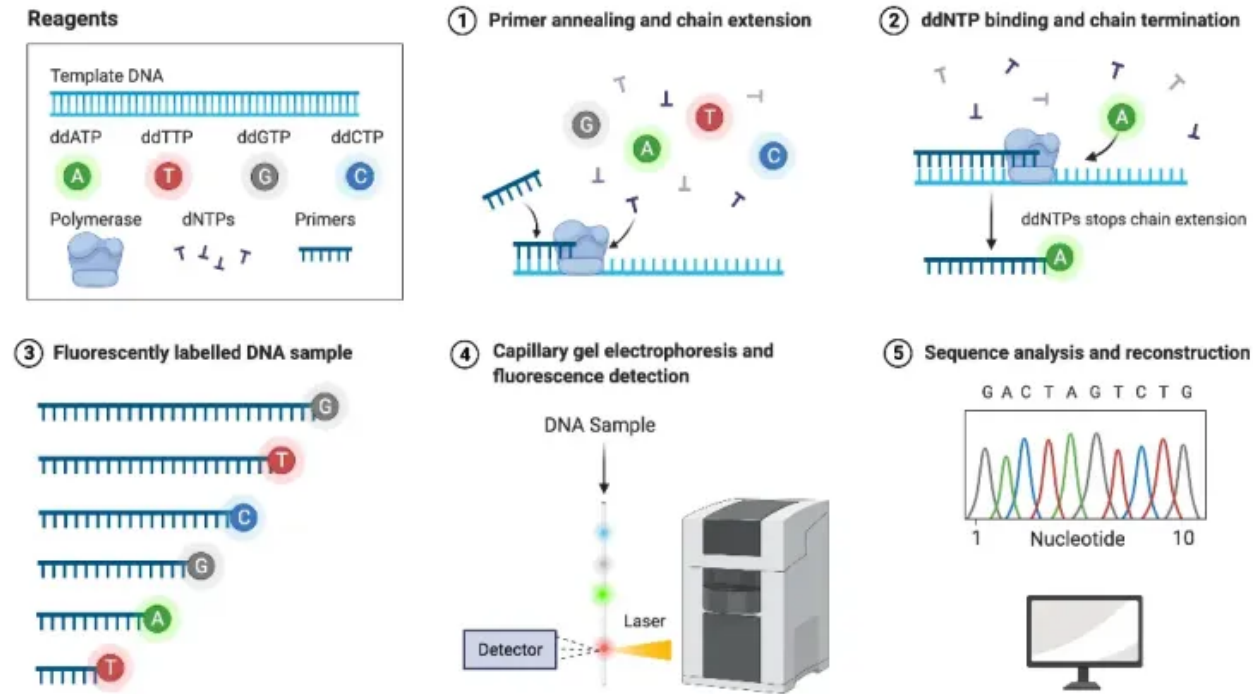
## A JOURNEY THROUGH THE HISTORY OF DNA SEQUENCING





# Secuenciación Sanger

## 1977, Secuenciación Sanger (Frederick Sanger)



*\*Originalmente se realizaba una reacción individual por cada ddNTP, y cada experimento se migraba en una carrera del gel, no siendo necesaria la detección de los fluorocromos.*

## PCR por terminación de cadena

- Adición de dNTPs + ddNTPs
- Didesoxinucleótidos (ddNTPs), sin grupo 3'-OH por lo que no forman el enlace fosfodiéster
- Terminación aleatoria

## Separación por tamaños en gel de electroforesis

- Los oligonucleótidos se separan por tamaños a resolución de un solo nucleótido.

## Detección de fluorescencia

- Cada ddNTP contiene un fluorocromo específico y se mide la fluorescencia en cada banda determinando la identidad del último nucleótido de la secuencia

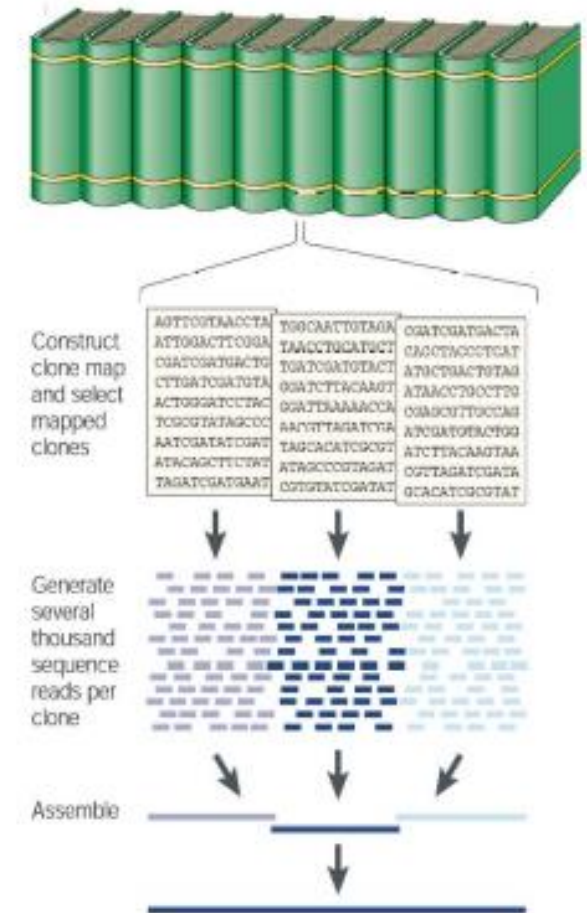
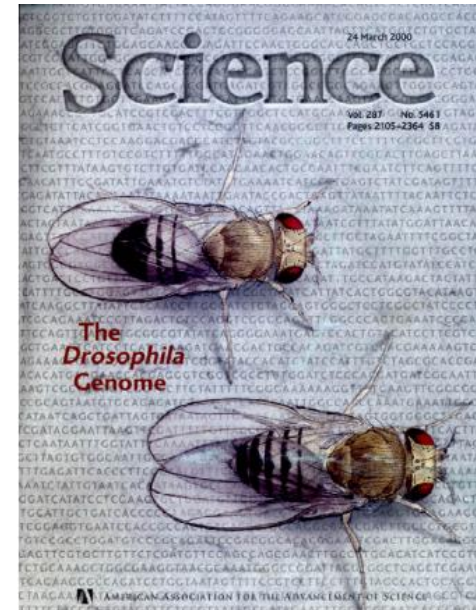
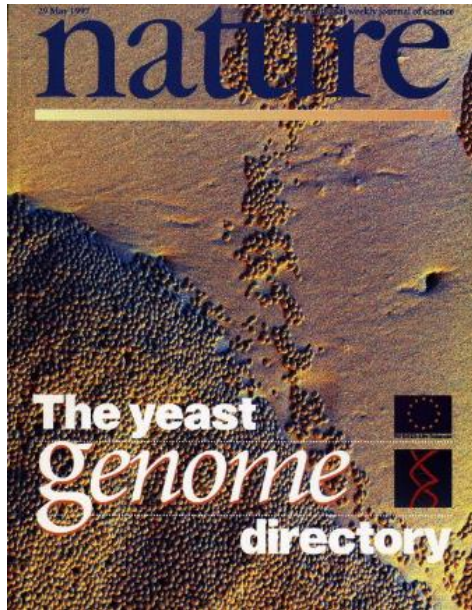
# Proyecto Genoma Humano

1990, Inicio proyecto secuenciación genoma humano

Objetivos iniciales:

- Iniciativa de 15 años
- Experiencia con organismos modelos más pequeños
- Estrategia: Primero organizar el ADN y luego secuenciar
- Esperar a una tecnología revolucionaria antes de empezar con el genoma humano

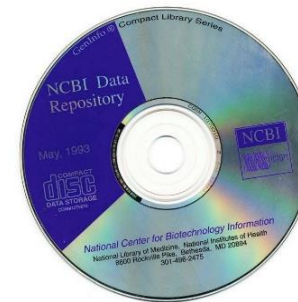
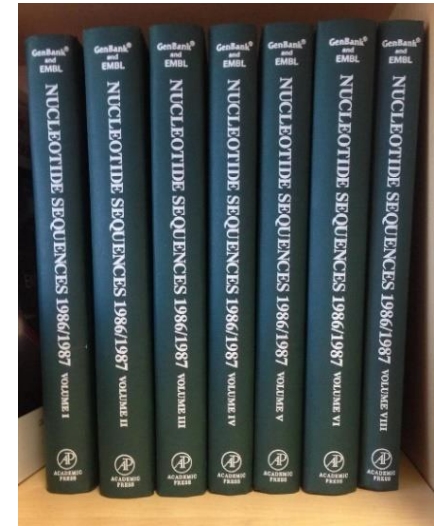
Primeros genomas eucariotas secuenciados por el proyecto genoma humano



# Proyecto Genoma Humano

## Principales retos en la secuenciación del genoma humano

- ~ 3.000.000.000 de nucleótidos que secuenciar y organizar en cromosomas
- La secuenciación Sanger podía leer ~500-800 bases por lectura
- Se necesitaba una **cobertura** de 30x (número de veces que se secuencia cada base) para obtener una buena fiabilidad
- ~50% del genoma humano se compone de **elementos repetidos**, una buena parte son remanentes de retrotransposones
- 6 países, 20 centros de investigación y más de 1000 científicos (divide y vencerás)
- Comunicación deficitaria (no internet moderno)
- Opiniones encontradas entre la comunidad científica
- No existía un plan detallado de ejecución del proyecto
- Potencia computacional sin precedentes para abordar el análisis de los datos secuenciados
- Secuencia consenso, ¿cuál es el genoma "normal"?
- ¿Propiedad del genoma humano? (Principios de las Bermudas)



When the Human Genome Project began, there was not a widely functional internet. I was barely using e-mail! The first sequences generated from human chromosome 7 were handwritten and faxed to me so that I could develop markers for building a physical map of human chromosome 7.



**WANTED**  
20 Volunteers  
to participate in the  
**Human Genome Project**  
a very large international scientific research effort.

The goal is to decode the human hereditary information (*human blueprint*) that determines all individual traits inherited from parents. The outcome of the project will have tremendous impact on future progress of medical science and lead to improved diagnosis and treatment of hereditary diseases.

Volunteers will receive information about the project from the Clinical Genetics Service at Roswell Park, and sign a consent form before participating.

*No personal information will be maintained or transferred.*

Volunteers will provide a one-time donation of a small blood specimen. A small monetary reimbursement will be provided to the participants for their time and effort.

Individuals must be at least 18 years of age.  
Persons who have undergone chemotherapy are not eligible.

For more information please contact the  
**Clinical Genetics Service**  
845-5720 (9:00 am - 3:00 pm)  
March 24 - 26, 1997

**ROSWELL PARK**  
CANCER INSTITUTE



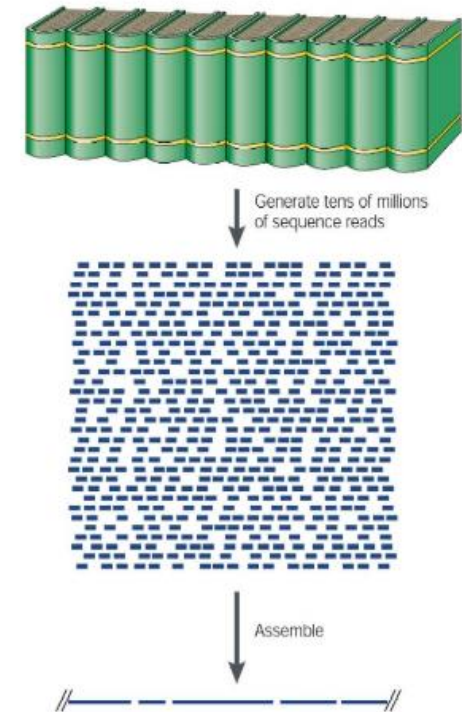
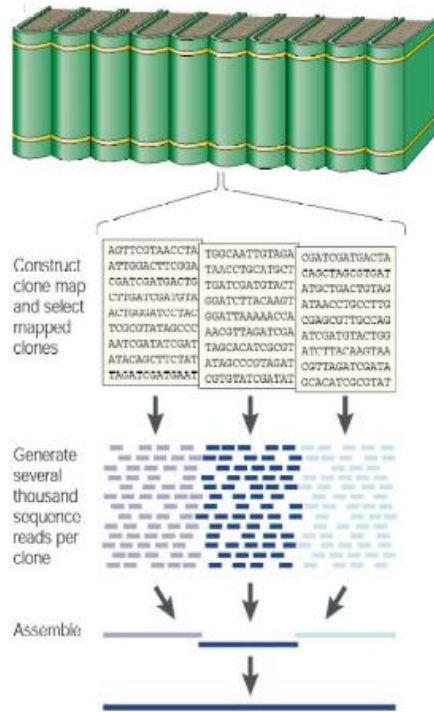
# Proyecto Genoma Humano

La carrera hacia la secuenciación del genoma humano

**Consortio Internacional (Dr. Francis Collins)**  
'Clone-by-Clone Shotgun Sequencing'

**VS**

**Celera Genomics (Dr. Craig Venter)**  
'Whole-Genome Shotgun Sequencing'

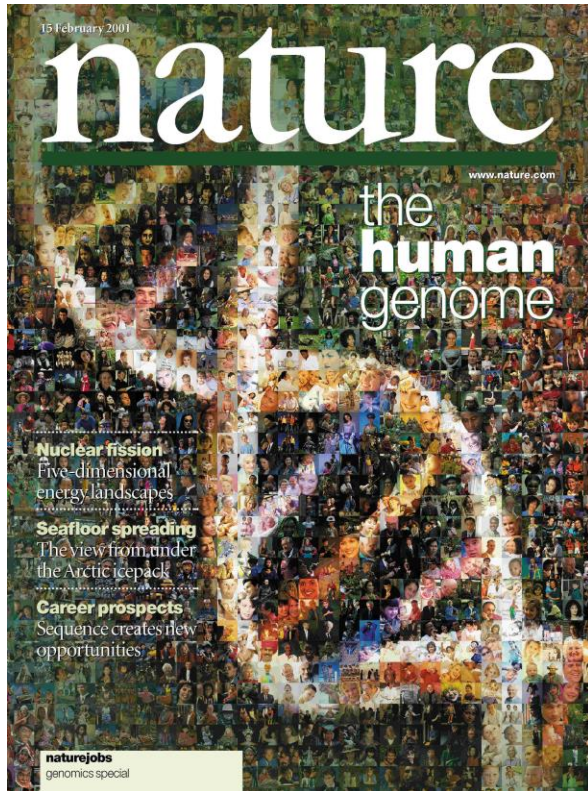


Una carrera "poco justa", Celera tenía acceso a los resultados del proyecto internacional pero no al revés

# Proyecto Genoma Humano

Febrero 2001, publicaciones de las primeras secuenciaciones del genoma humano

## Consorcio Internacional

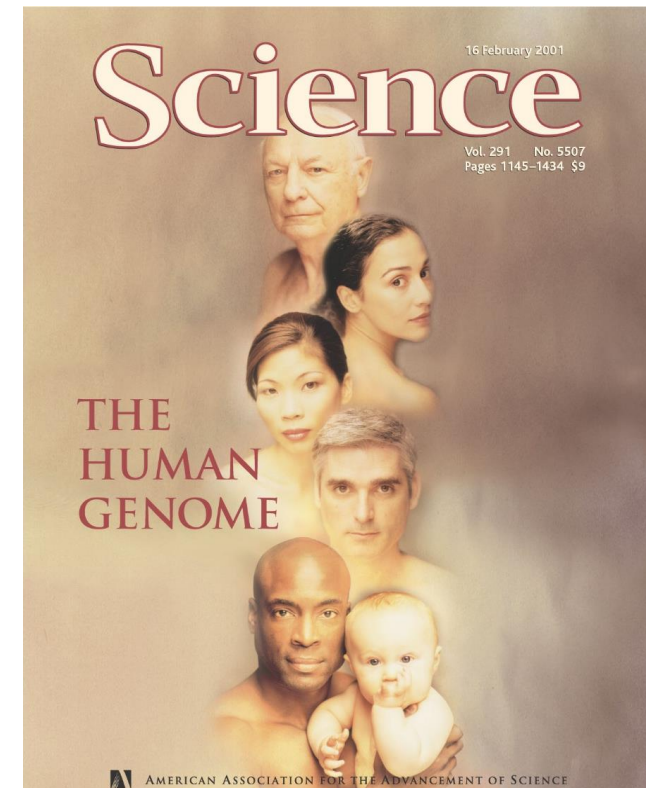


Celera no consiguió ensamblar el genoma humano al completo y tuvo que recurrir a los datos públicos para publicar la primera versión

El consorcio realizó un gran esfuerzo para mejorar la calidad de secuenciación y dar una versión final de alta calidad

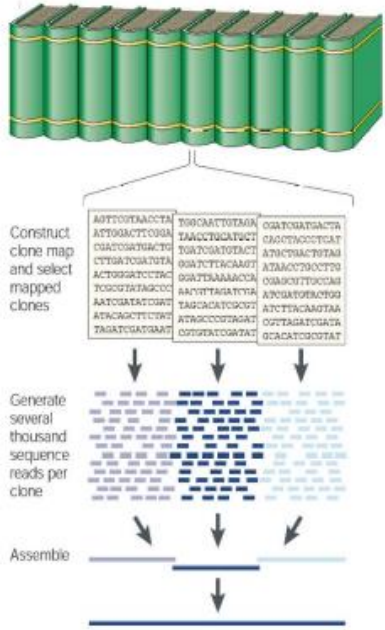
Se frustró el plan inicial de Celera de cobrar una suscripción por acceso al genoma humano

## Celera Genomics

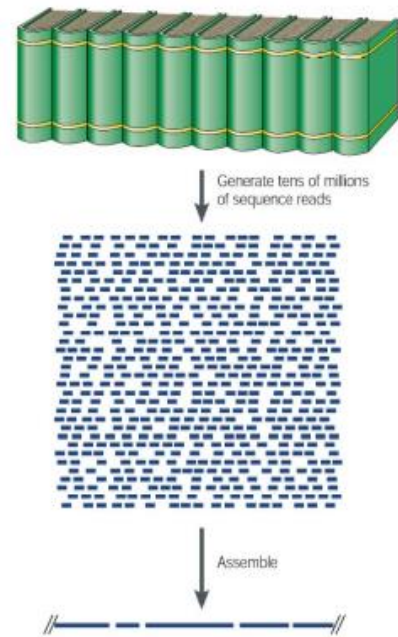


# Proyecto Genoma Humano

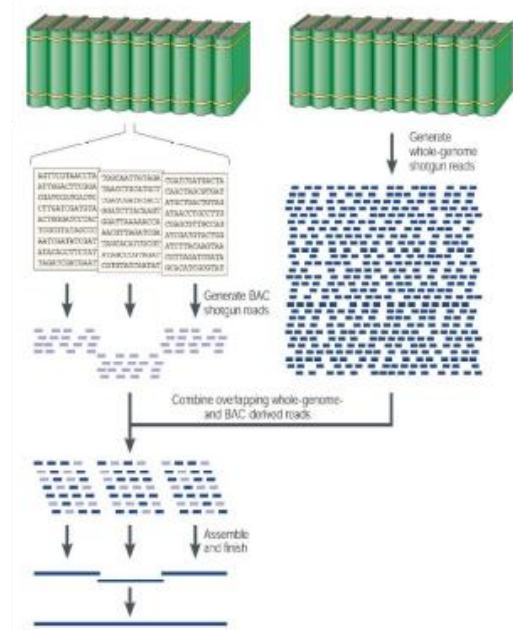
2003, el primer ensamblado del genoma humano



+



=



13 años

~92% del genoma humano

~ 1.000.000.000 de inversión

Despejó las dudas de la comunidad científica

Inicio la genómica moderna



# Segunda generación de secuenciación

2005, primer secuenciador masivo del mercado Roche/454

Características principales de la secuenciación masiva:

- Lecturas cortas
- Millones de lecturas en paralelo
- Velocidad
- Bajo coste
- Sin necesidad de electroforesis

**Secuenciación por síntesis:** Roche/454, Ion Torrent, Illumina

Adaptadores complementarios con primers

Amplificación y formación de grupos en cada pocillo

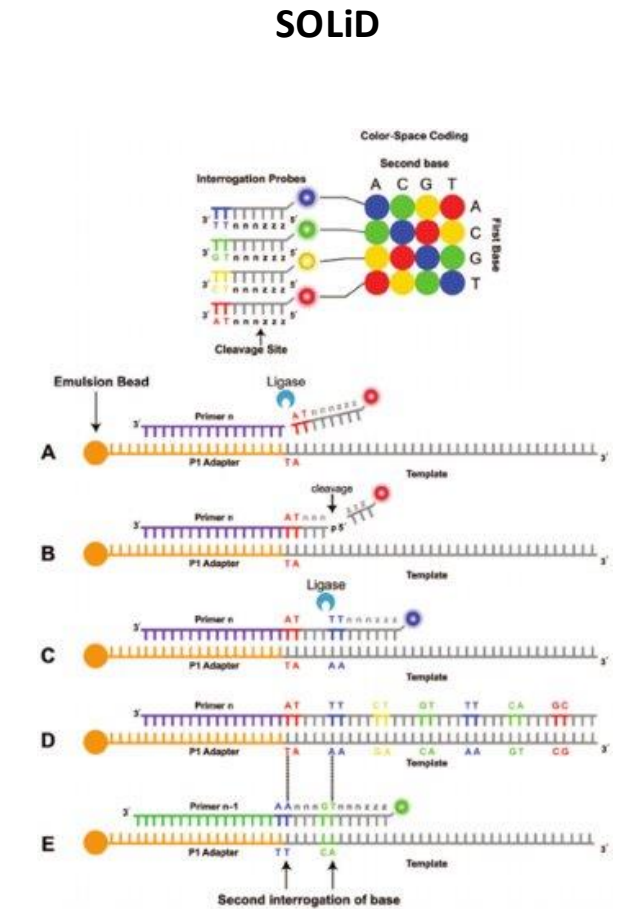
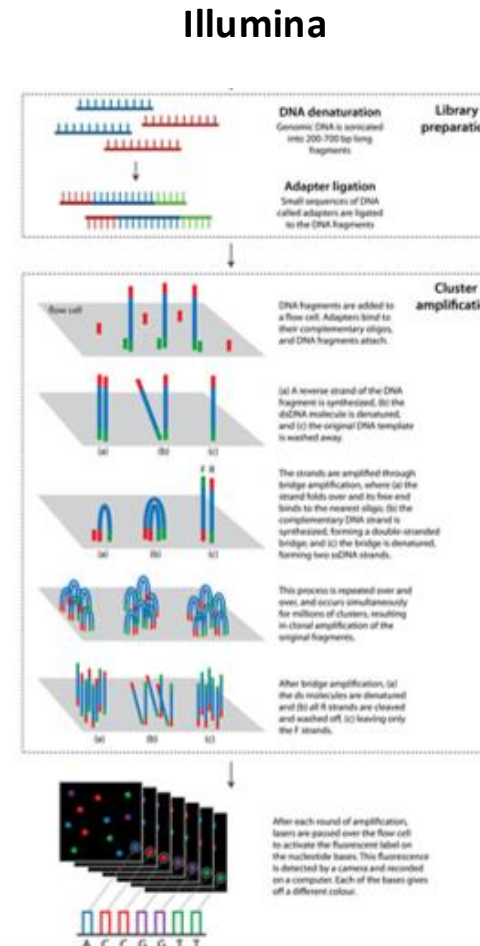
Fluorescencia específica por dNTP (3'-OH inactivo)

**Secuenciación por ligación:** ABI/SOLiD

Múltiples rondas de secuenciación

Usa perlas magnéticas

Interroga las bases de dos en dos dando un código de color



# Proyecto 1000 genomas

2008, inicio proyecto 1000 genomas



- 2506 individuos secuenciados de 26 poblaciones en 2015
- 88 millones de variantes: 84.7 millones de SNPs, 2.6 millones de inserciones/delecciones cortas y 60,000 variantes estructurales
- 4-5 millones de variantes de media entre un genoma individual y la referencia

REVIEWS  
AUTUMN BOOKS  
SPECIAL  
Fizz wars, geopoetry,  
mind and matter  
PAGE 24

HAWAII KECK OBSERVATORY  
MOUNTAIN  
DIFFICULTIES  
Is the Thirty Meter Telescope  
a step too far?  
PAGE 24

THE HUMAN GENOME  
25 YEARS OF  
BIG BIOLOGY  
Three major players reflect  
on lessons learned  
PAGE 28

NATURE.COM/NATURE  
1 October 2015 \$10  
Vol 526, No 7571  
TIFF Preview  
9 770024 085097

# Tercera generación de secuenciación

**2011**, primer secuenciador de tercera generación sistema PacBio

Características principales de la tercera generación:

- Evita el proceso de amplificación
- Menos costosa
- Más rápida
- Capaz de secuenciar zonas repetidas
- Lecturas largas

**Secuenciación a tiempo real de una única molécula de ADN (SMRT): PacBio**

ADN circularizado

Chip de silicio (celda SMRT)

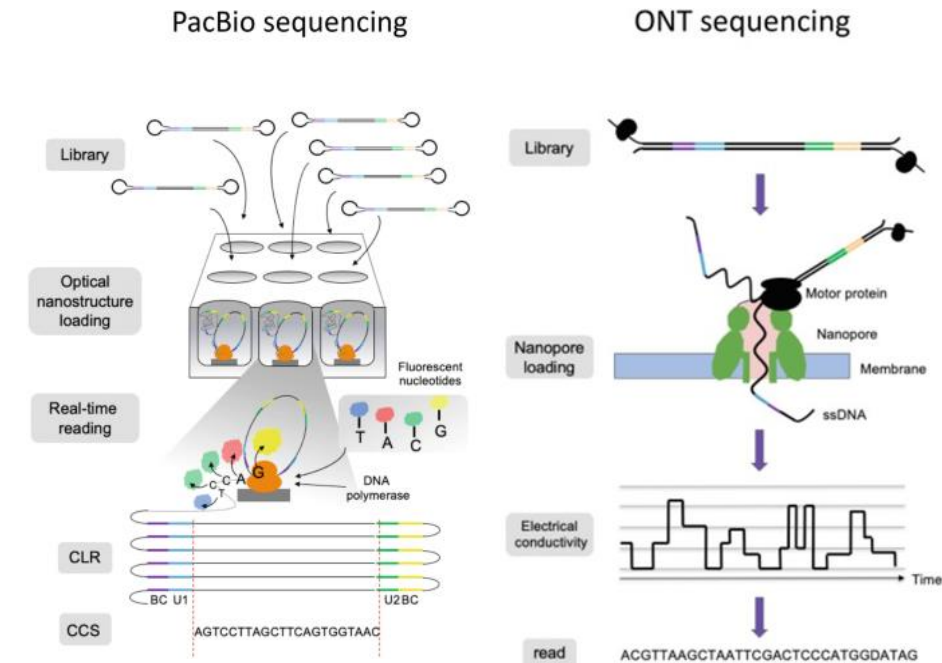
Se introducen dNTPs marcados cada uno con un fluorocromo

**Secuenciación a través de nanoporo: Oxford, MinION - PromethION**

Molécula pasa a través de un nanoporo mediante una proteína motora

Variación en la corriente iónica diferente según el nucleótido

El MinION se conecta al puerto USB3.0 de un ordenador



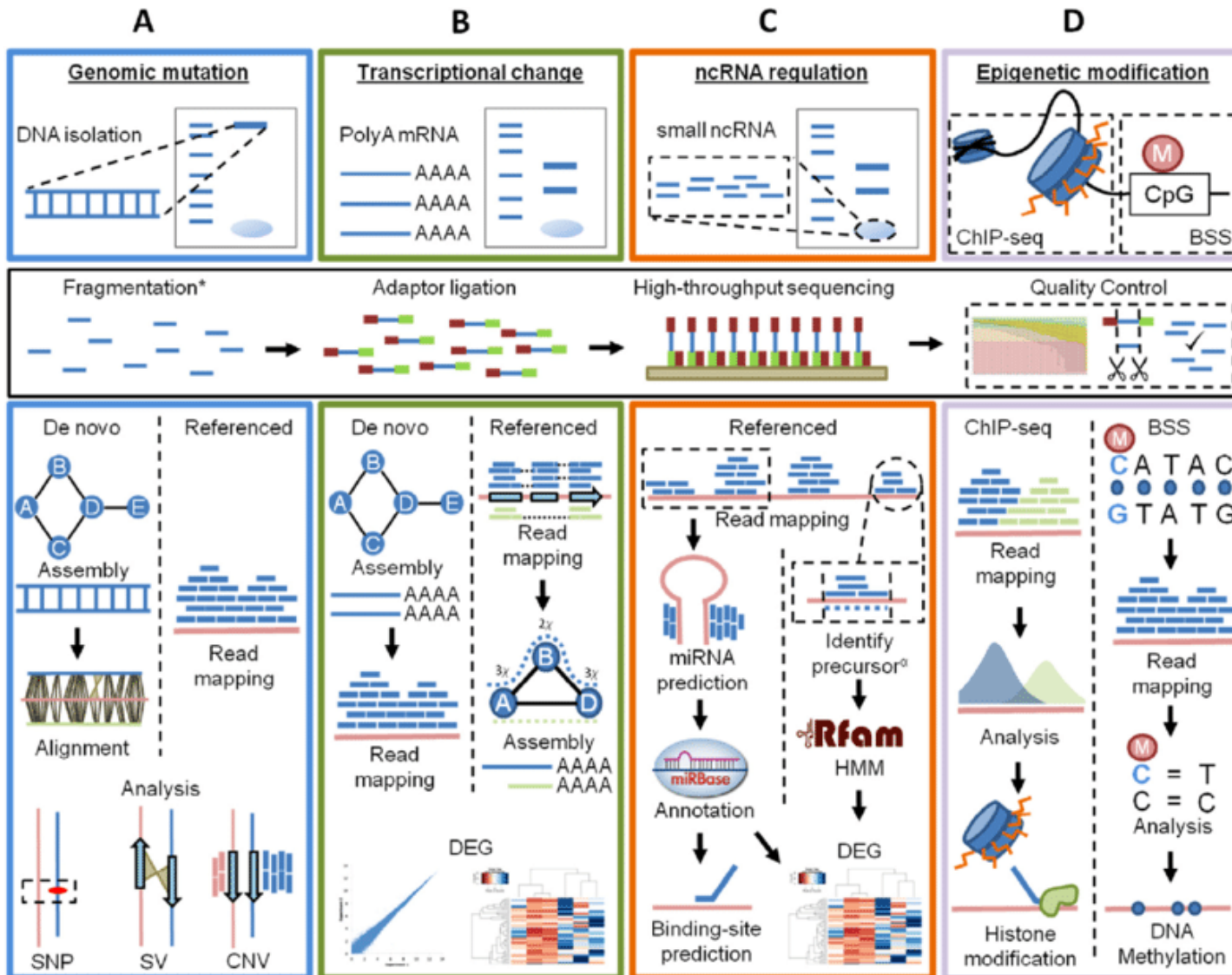


# Comparativa de las generaciones de secuenciación

	<b>First generation</b>	<b>Second generation</b>	<b>Third generation</b>
<b>Fundamental technology</b>	Size-separation of specifically end-labeled DNA fragments	Wash-and-scan SBS	Single molecule real time sequencing
<b>Resolution</b>	Averaged across many copies of the DNA molecule	Averaged across many copies of the DNA molecule	Single DNA molecule
<b>Current raw read accuracy</b>	High	High	Lower
<b>Current read length</b>	Moderate (800-1000 bp)	Short (generally much shorter than Sanger sequencing)	> 1000 bp
<b>Current throughput</b>	Low	High	High
<b>Current cost</b>	High cost per base, Low cost per run	Low cost per base, High cost per run	Low cost per base, High cost per run
<b>RNA-sequencing method</b>	cDNA sequencing	cDNA sequencing	Direct RNA sequencing
<b>Time to result</b>	Hours	Days	< 1 day
<b>Sample preparation</b>	Moderately complex, PCR amplification is not required	Complex, PCR amplification is required	Various
<b>Data analysis</b>	Routine	Complex (due to large data volumes & short reads)	Complex
<b>Primary results</b>	Base calls with quality values	Base calls with quality values	Base calls with quality values

Adapted from Schadt, et al. Hum Mol Genet 2010<sup>13</sup>

# Utilidades de la secuenciación masiva



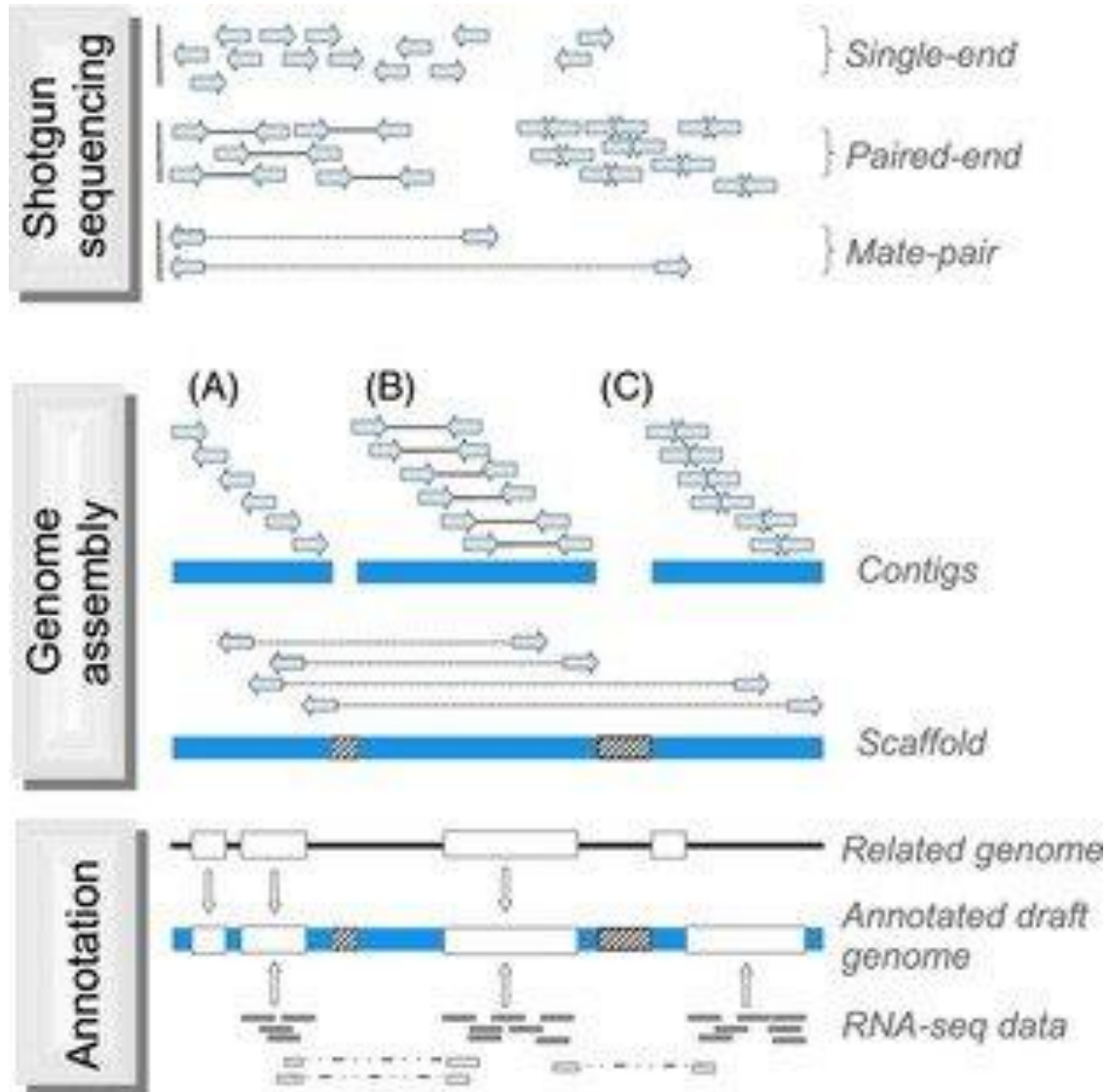
A1. Identificación de variantes del ADN  
 A2. Generación referencias de genomas no anotados

B1. Cuantificación de expresión génica  
 B2. Generación referencias de transcriptomas no anotados

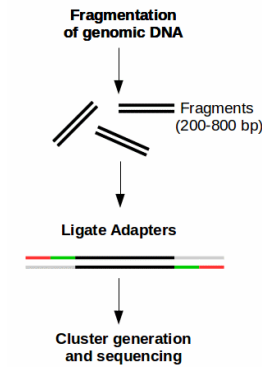
C. Cuantificación e identificación de RNAs cortos no codificantes

D1. Identificación de modificaciones de histonas y sitios activos de unión de factores de transcripción  
 D2. Cuantificación de modificaciones del ADN (por ejemplo, metilación)

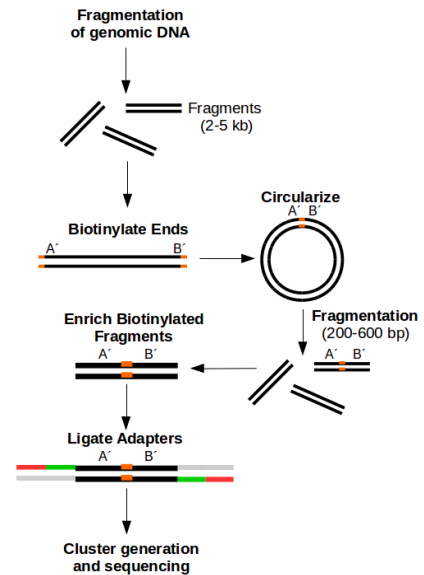
# Análisis de datos de secuenciación masiva



**Paired-End Sequencing**  
(Short-insert paired-end reads)



**Mate Pair Sequencing**



- **Re-secuenciación de genomas**

- Lecturas cortas (SE/PE/ME)

- Alineamiento frente a referencia

- Cuantificación

- **Secuenciación *de novo***

- Lecturas largas o ME – Ensamblado de grandes bloques

- Lecturas cortas (SE/PE) - Alta calidad de secuenciación



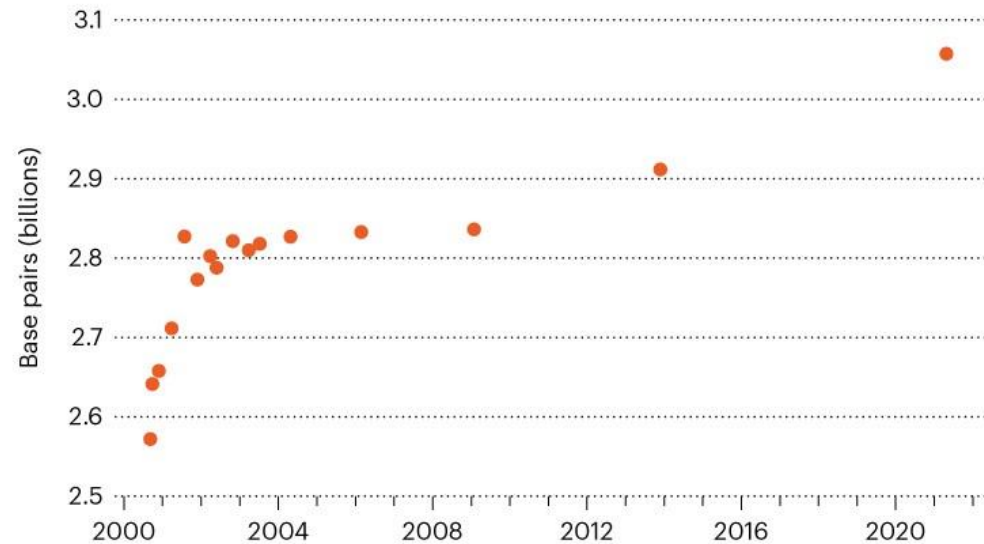
# Genoma Humano Completo

2022, el "verdadero" primer ensamblado completo del genoma humano



## COMPLETING THE HUMAN GENOME

Researchers have been filling in incompletely sequenced parts of the human reference genome for 20 years, and have now almost finished it, with 3.05 billion DNA base pairs.



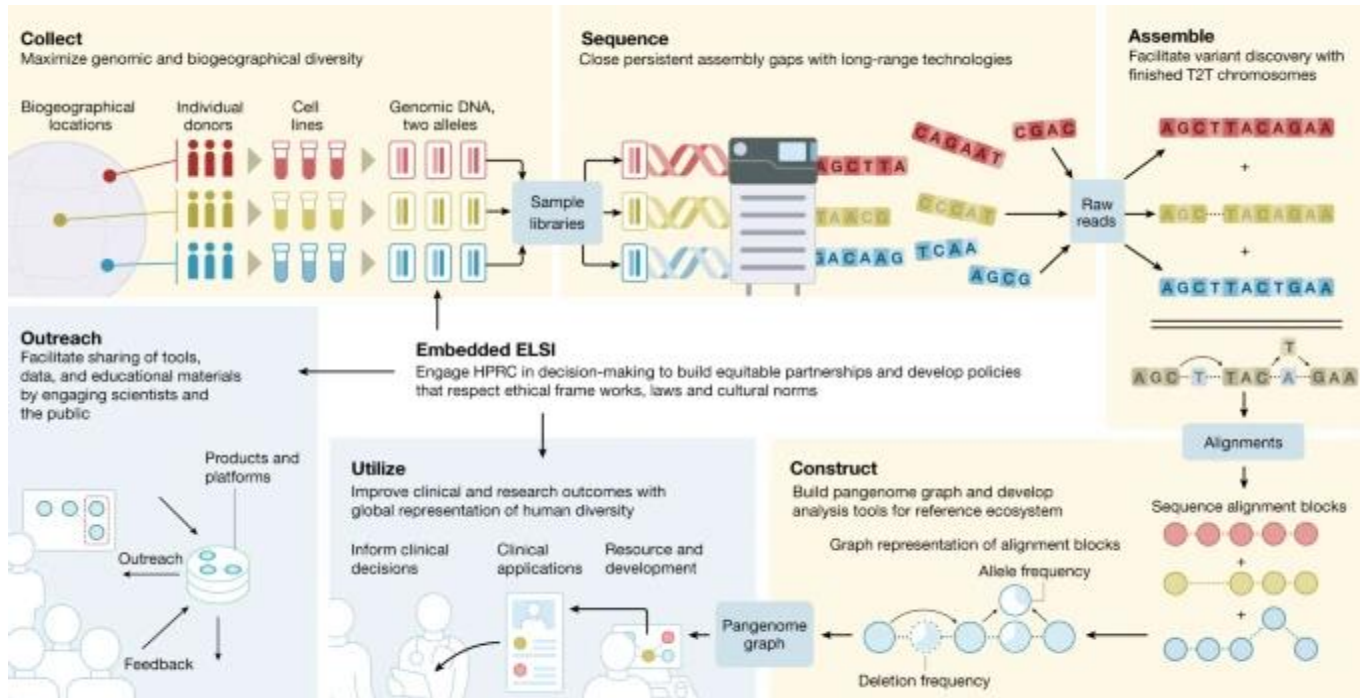
0.3% of sequence might still have errors. Includes X but not Y chromosome. Count excludes mitochondrial DNA.

©nature

2023, secuenciación completa del cromosoma Y

# El camino hacia el 'pangenoma' humano

2023, hacia el 'pangenoma' humano



**Referencia actual:** 20 individuos y 70% de la secuencia de uno solo  
**Objetivo:** ~350 individuos (700 haplotipos de altísima calidad) que refleje la diversidad de las 26 poblaciones incluidas en el proyecto  
1000 genomas

