

NUEVAS CAPACIDADES PARA LA MEDICINA GENÓMICA EN CANARIAS

División de Genómica



ITER Instituto Tecnológico y de
Energías Renovables

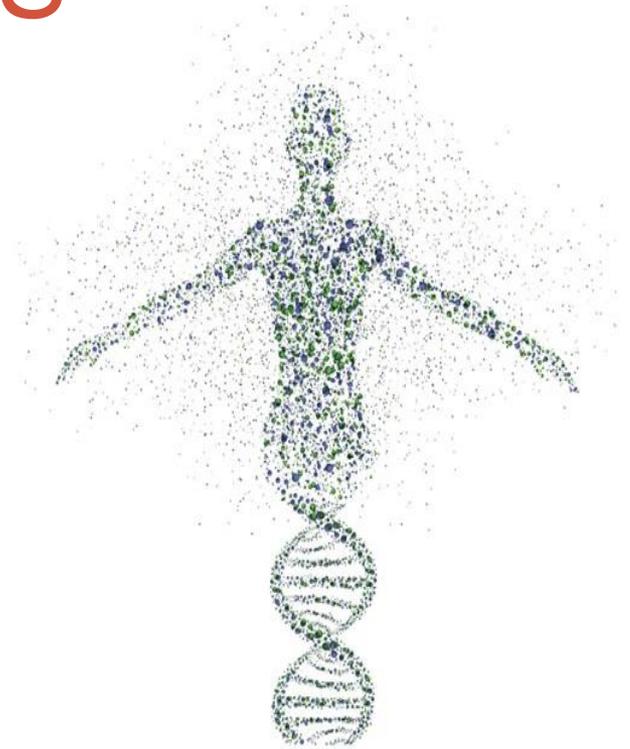


**Gobierno
de Canarias**
un solo pueblo

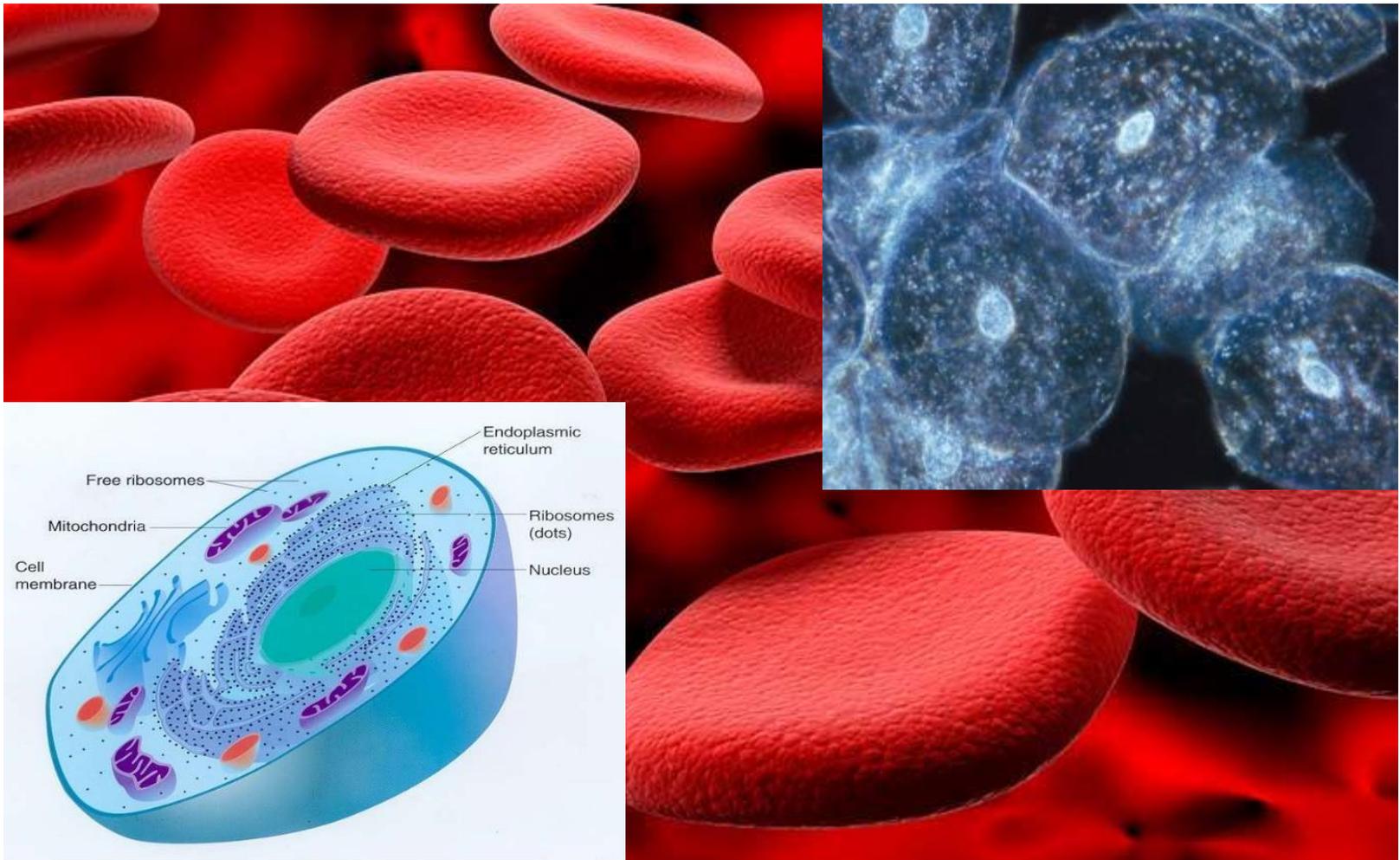
Consejería de Educación
y Universidades



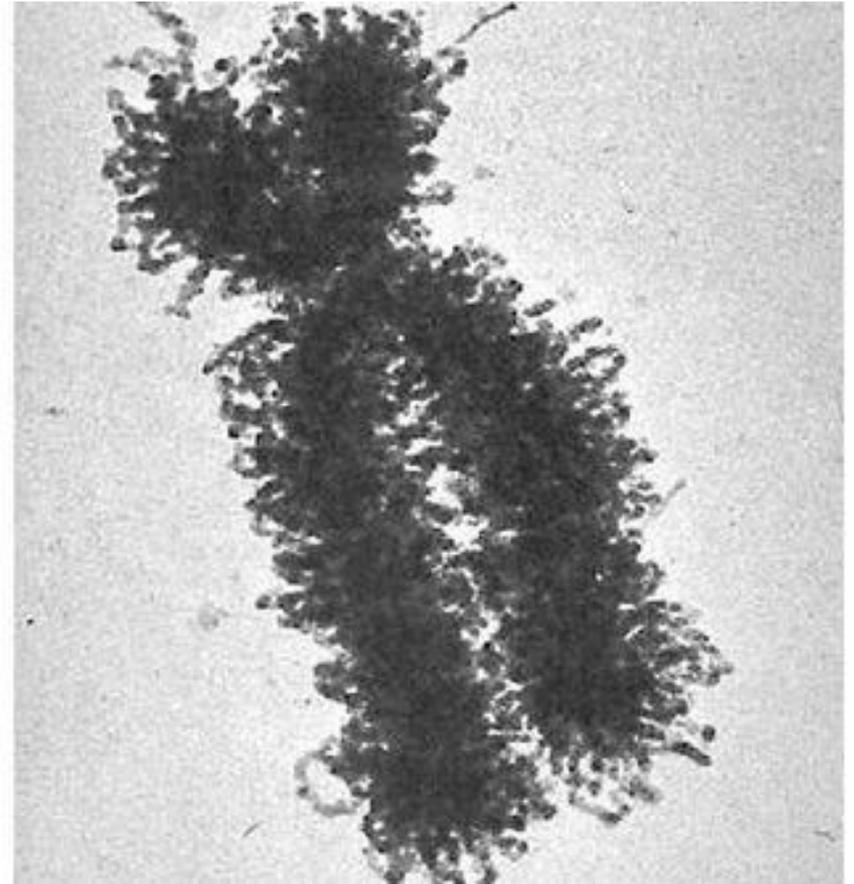
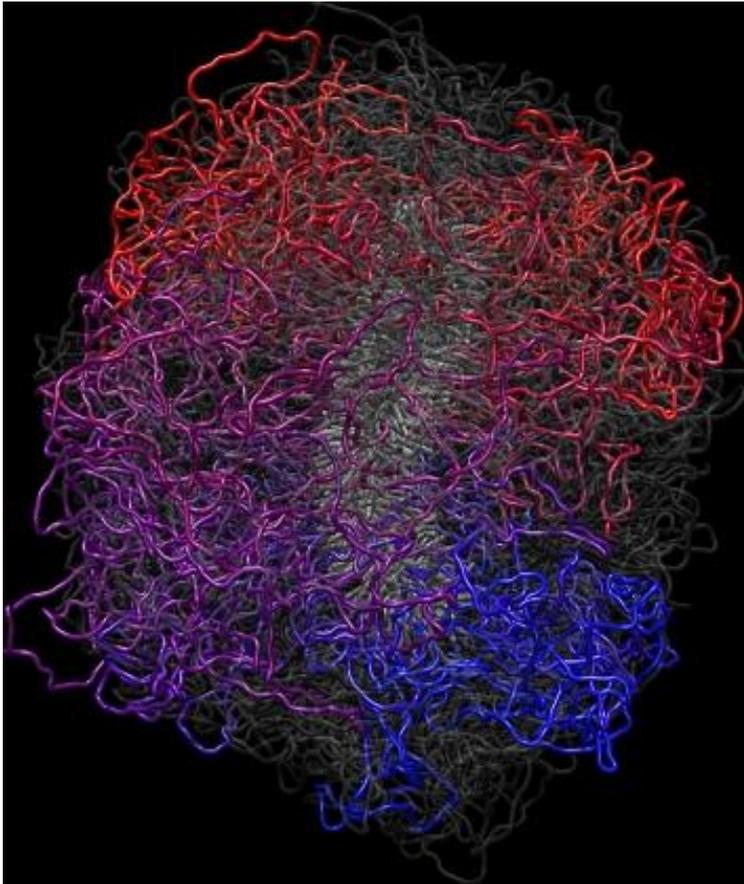
Servicio
Canario de la Salud



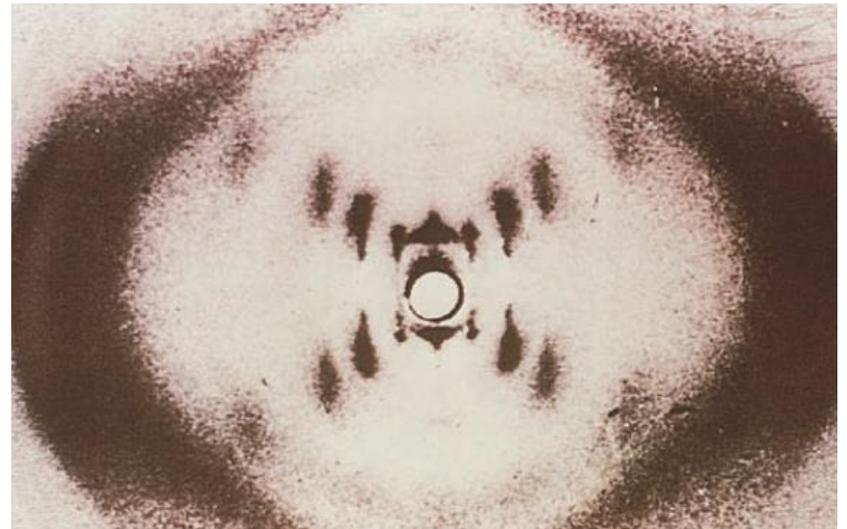
El material genético en el núcleo celular



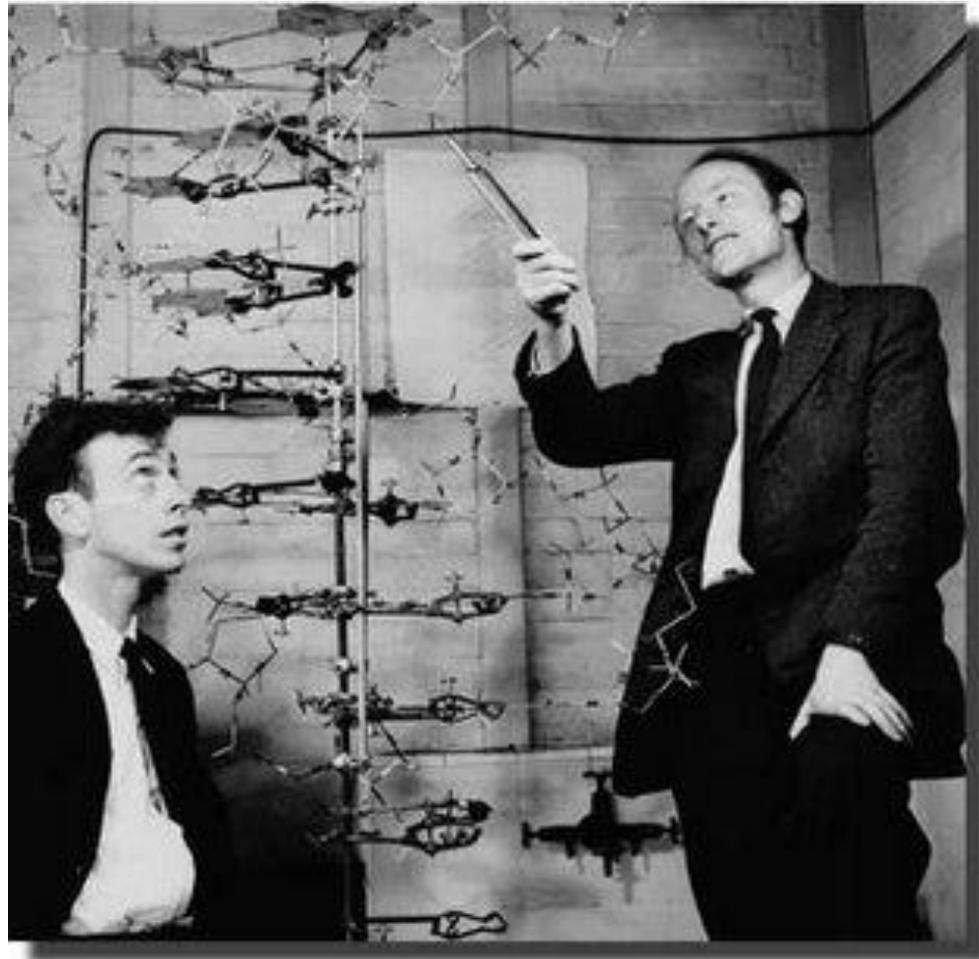
Conformación 3D del ADN



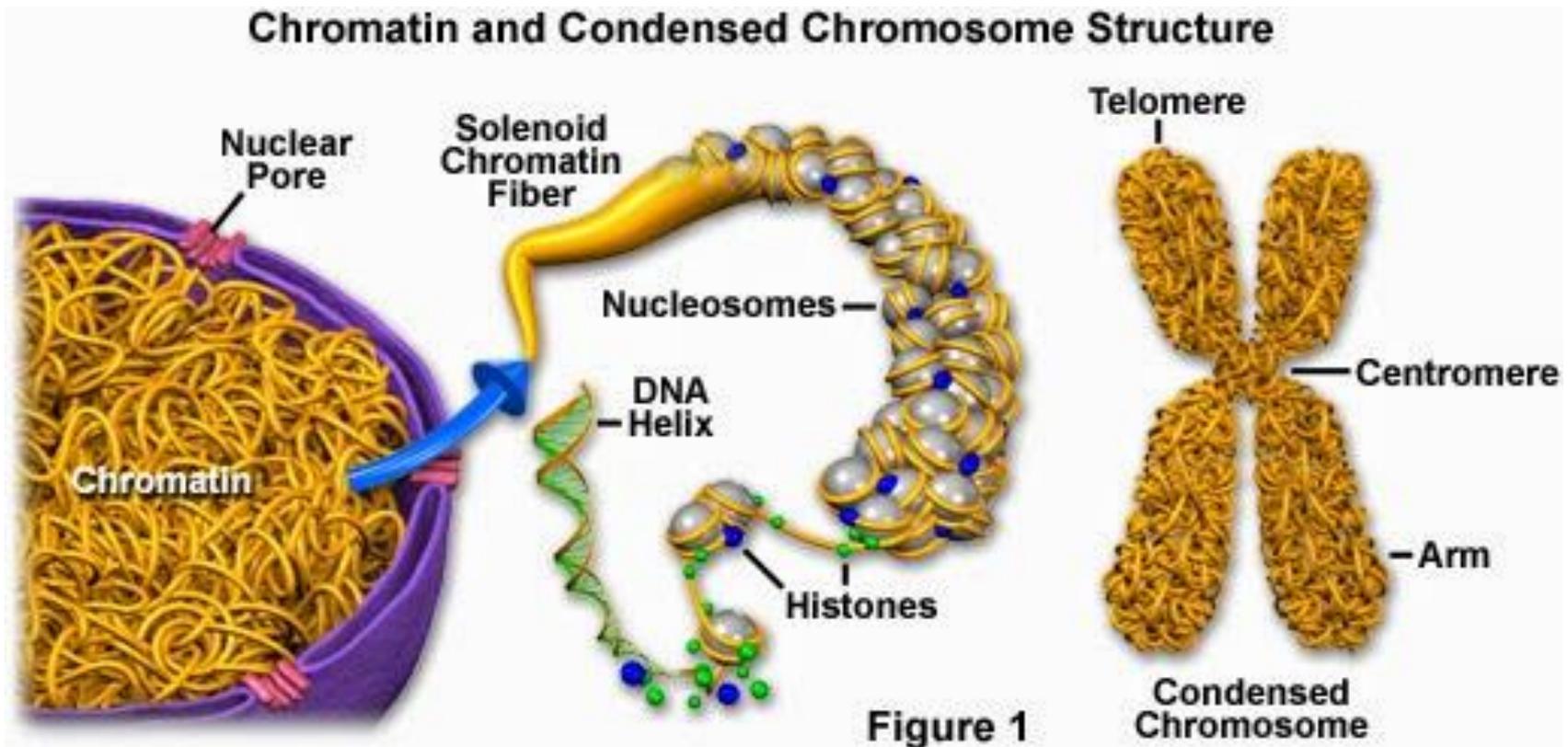
Descubrimiento del ADN



El ADN es una doble hélice (1953)



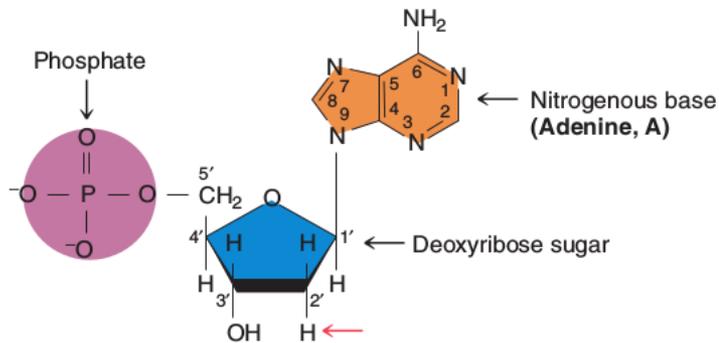
Estructura a varias escalas



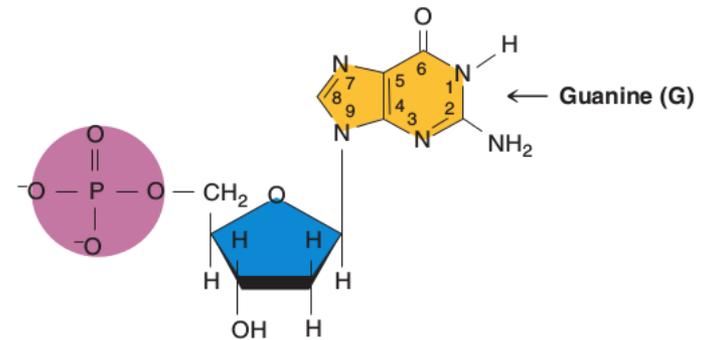
Los cuatro “ladrillos” del ADN

Structure of the four DNA nucleotides

Purine nucleotides

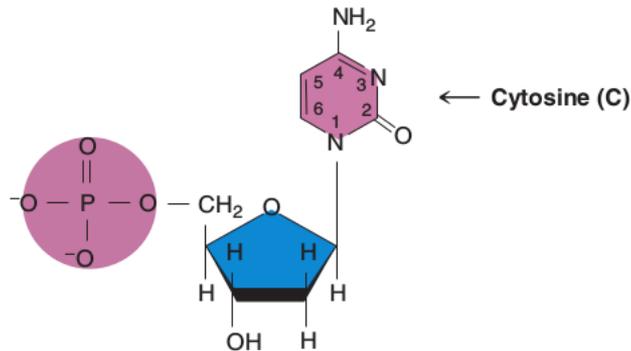


Deoxyadenosine 5'-monophosphate (dAMP)

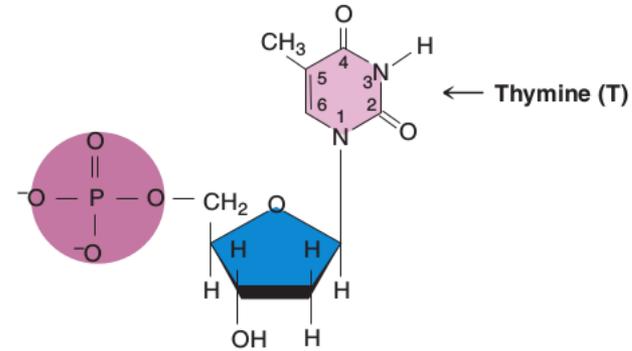


Deoxyguanosine 5'-monophosphate (dGMP)

Pyrimidine nucleotides



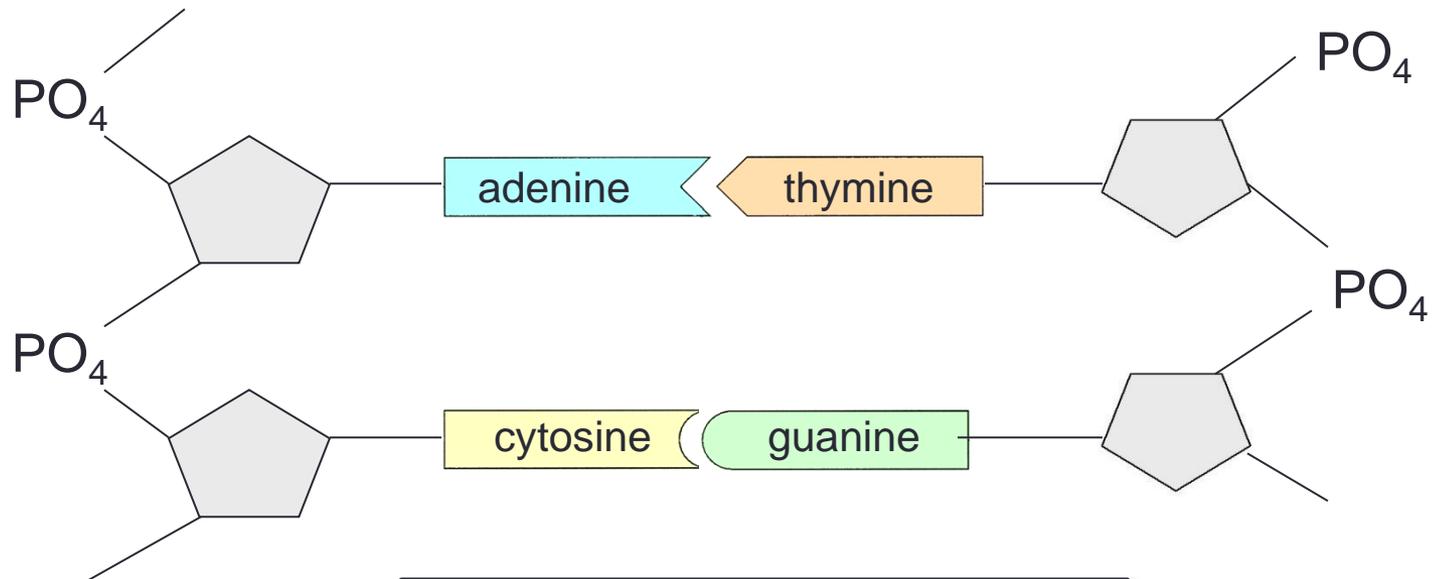
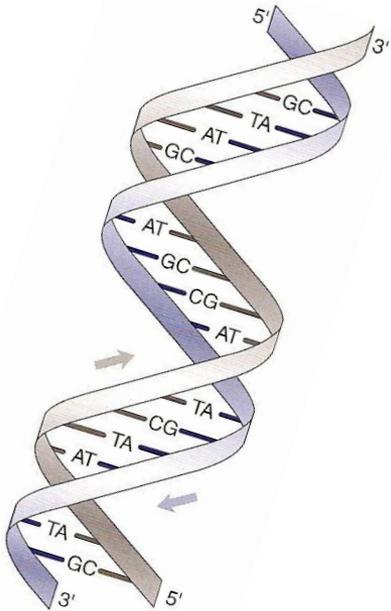
Deoxycytidine 5'-monophosphate (dCMP)



Deoxythymidine 5'-monophosphate (dTMP)

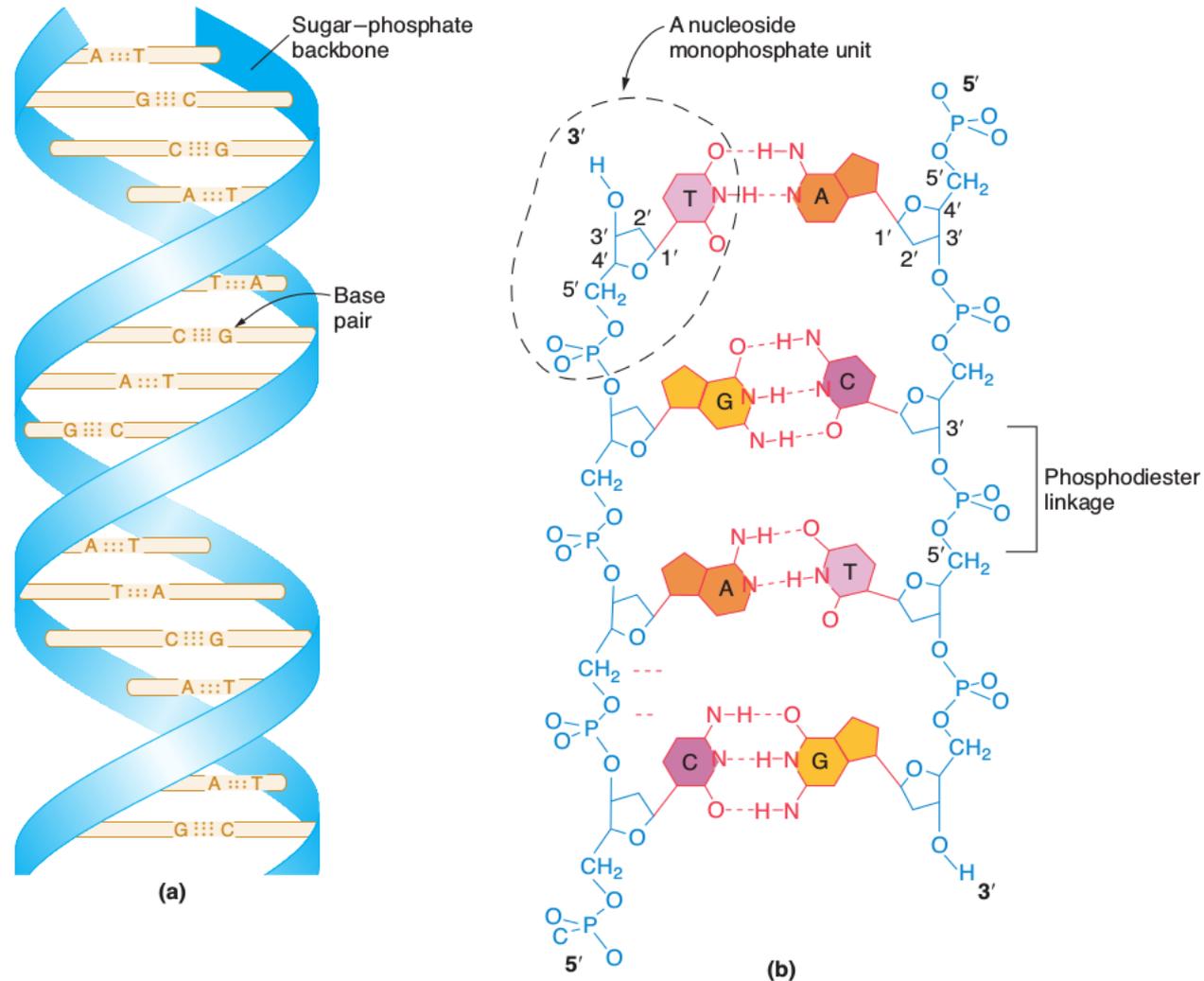
Un alfabeto con una “sintaxis” especial

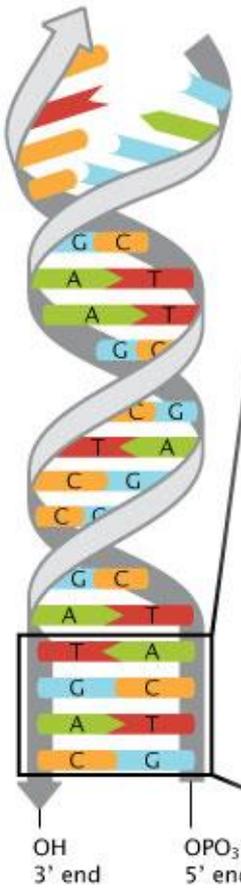
Complementariedad de bases



A se enlaza con **T**
C se enlaza con **G**

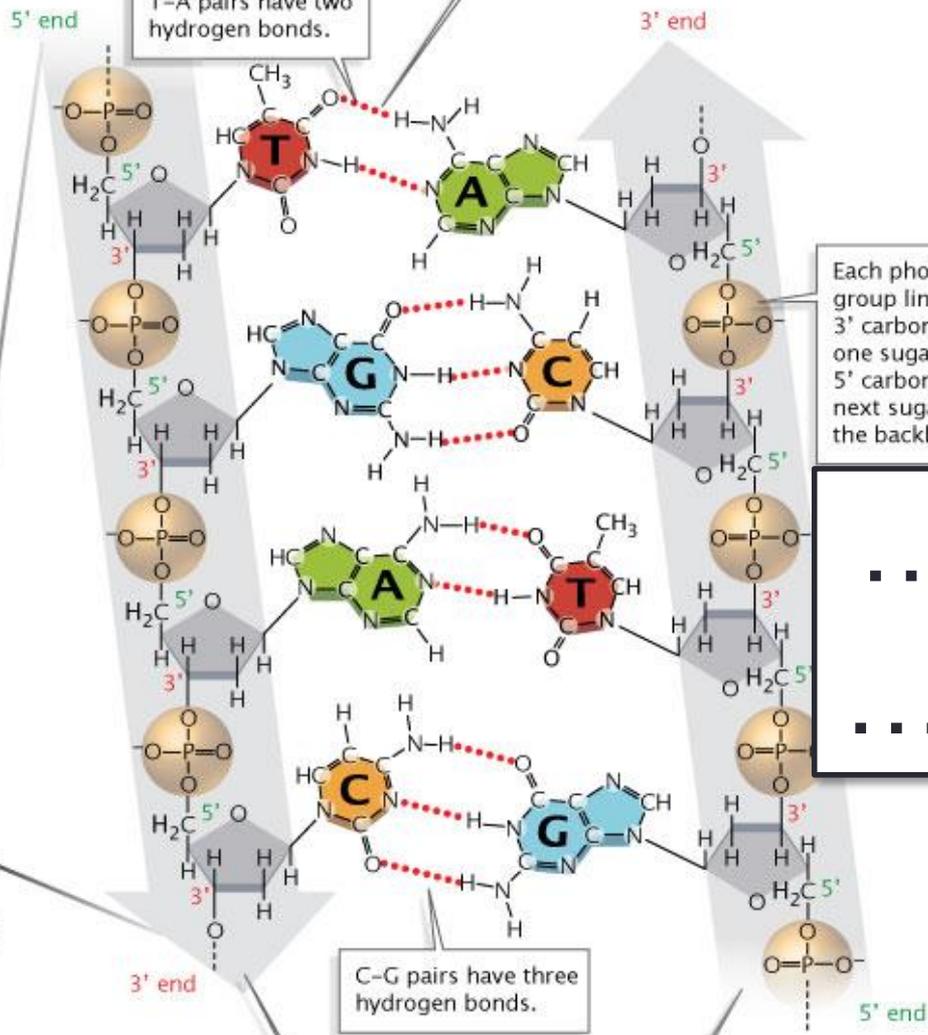
El resultado: una hélice de doble cadena





Pairs of complementary bases form hydrogen bonds that hold the two strands of the DNA double helix together.

T-A pairs have two hydrogen bonds.



Each phosphate group links the 3' carbon of one sugar to the 5' carbon of the next sugar along the backbone.

... T G A C ...

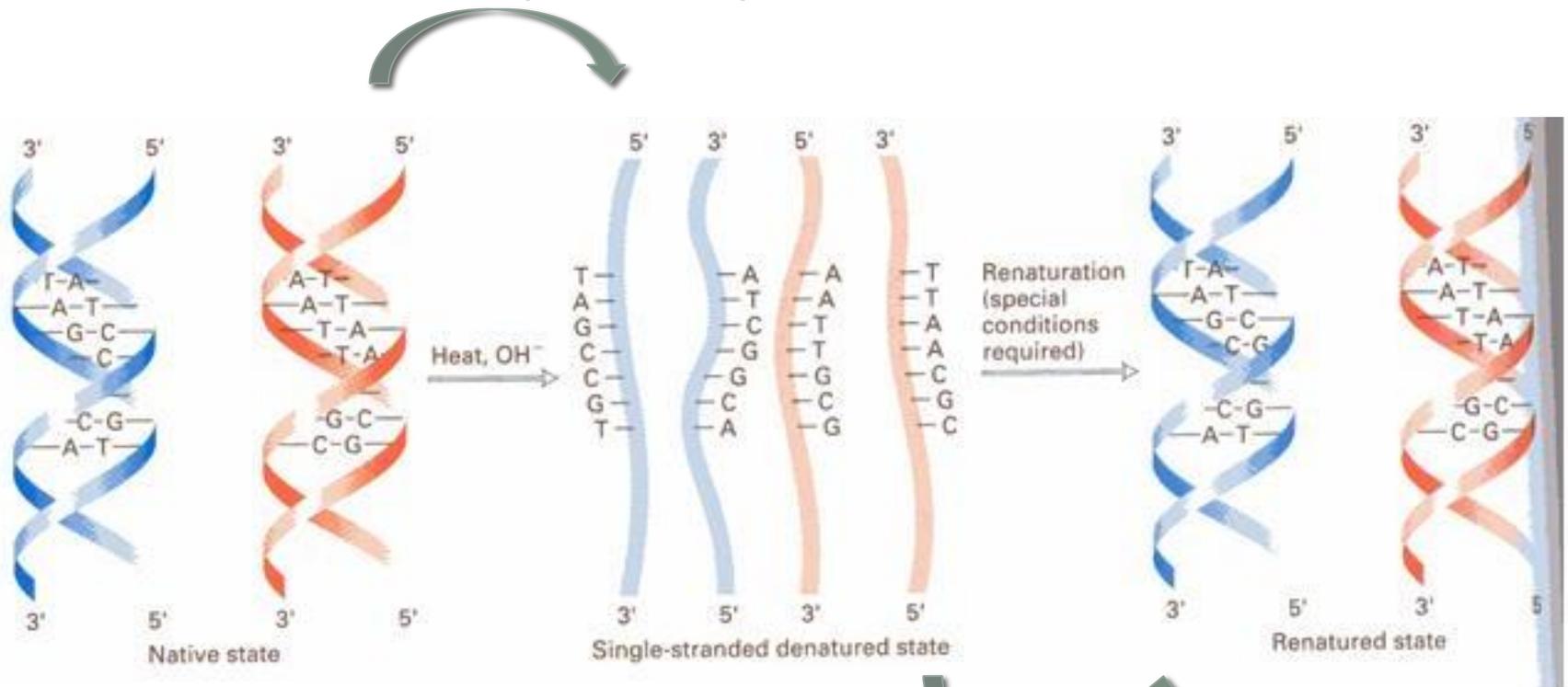
... A C T G ...

C-G pairs have three hydrogen bonds.

The strands both run in a 5'-to-3' direction - they are antiparallel.

Una propiedad importante del ADN

Elevando temperatura o pH



Bajando temperatura o pH

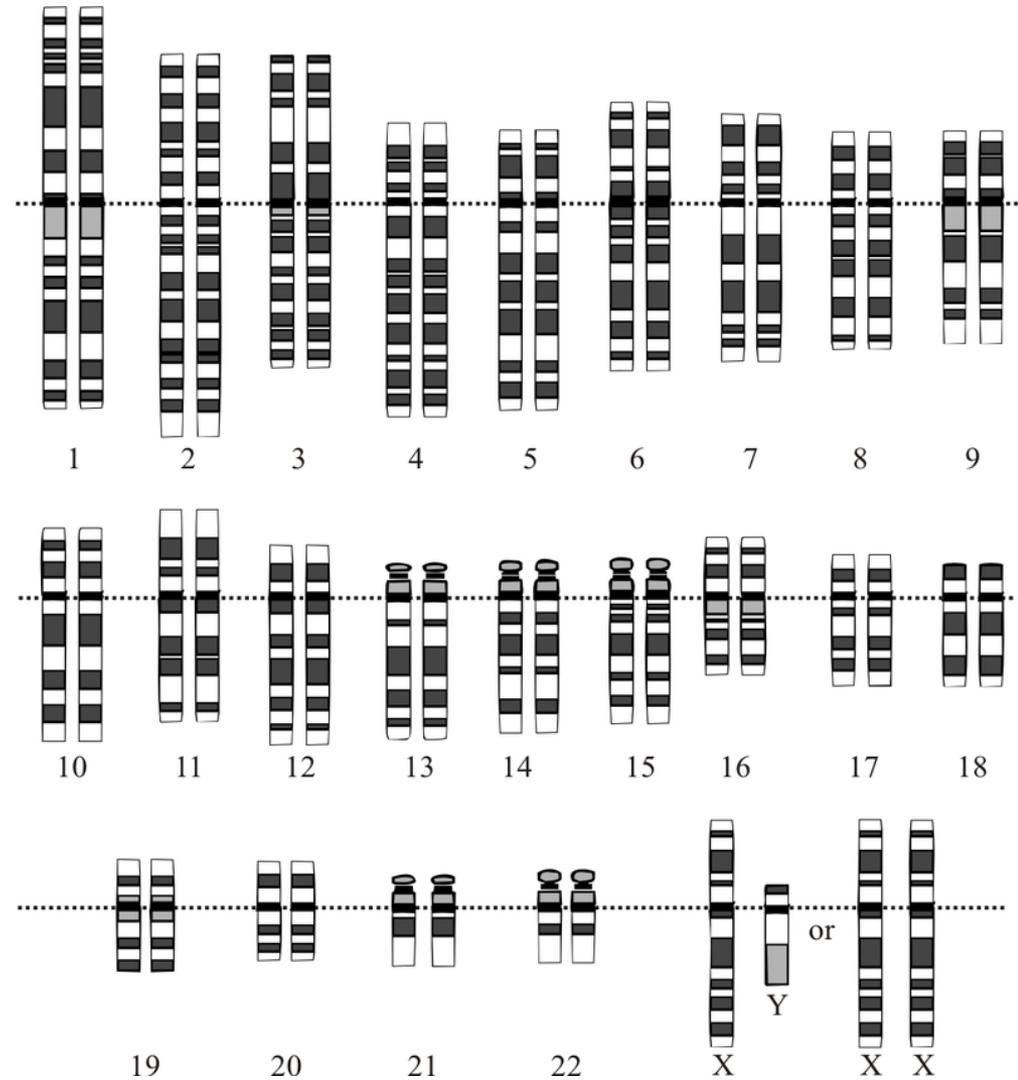
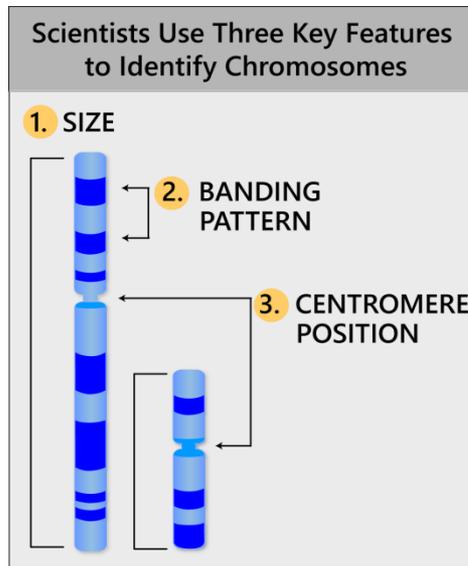
A. Diploide

B. Dividido en 46 fragmentos:
a) 22 pares de autosomas
b) 1 par sexual

C. ADN mitocondrial (circular, extranuclear)

D. 3.200 millones de bases (letras)

E. 20.000-24.000 genes



Cariotipo humano

Ejercicio 1

Si la distancia entre dos nucleótidos es aproximadamente 0,34 nm (recuerda que 1 nm es 10^{-9} m)... ¿qué longitud tiene el ADN de una única célula diploide?

SOLUCIÓN:

Como sabemos que una hebra tiene aproximadamente 3,2 Giga bases...

Obtenemos que una hebra de ADN mide unos 2,5 m.

$2,5 \text{ m} \times 2 \text{ hebras} = 5 \text{ m}$ en una única célula humana

Por cierto, ni el ADN ni la célula se ven a simple vista... La célula podría verse con un microscopio de los que tienen los centros educativos en sus laboratorios.

Ejercicio 2

Si extraemos el ADN de todas las células de mi cuerpo, ¿qué longitud total tiene todo ese ADN?

Datos:

- Estimamos unas 50 billones de células diploides en el organismo humano.
- La distancia promedio Tierra-Sol es de 150.000.000 km (1 u.a., unidad astronómica).

Si estiramos todo este ADN, ¿podríamos enlazar la Tierra con el Sol?
¿Cuántas veces?

SOLUCIÓN:

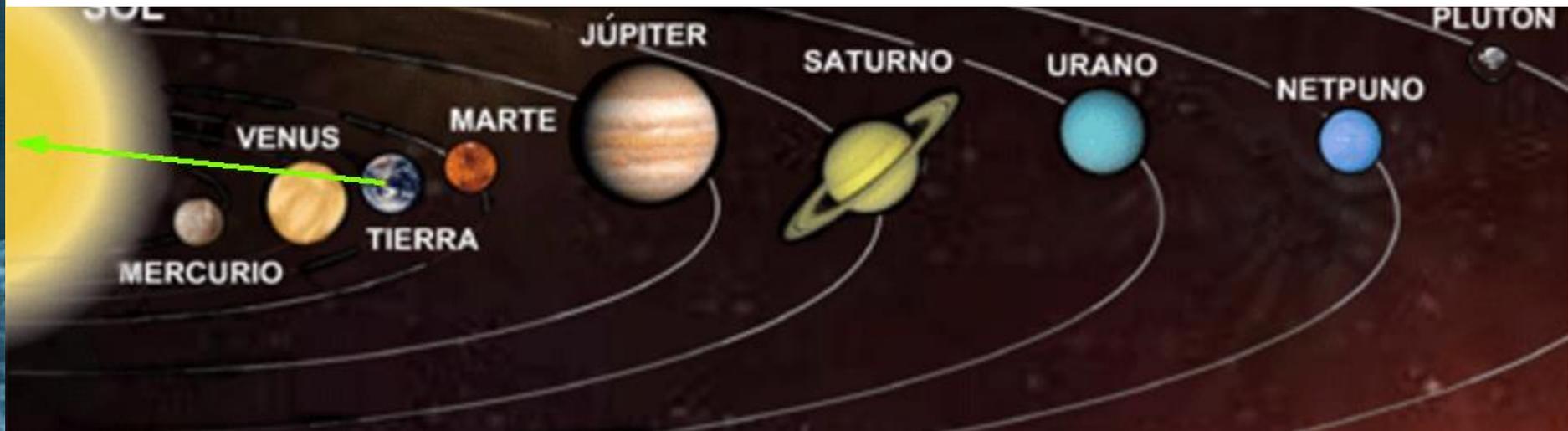
2,5 m x 2 hebras = 5 m en una única célula humana

$(5 \text{ m ADN/célula}) \times (50 \cdot 10^{12} \text{ células/persona}) \times (1 \text{ km}/10^3 \text{ m}) / (150 \cdot 10^6 \text{ km}) = ?$

Es decir, el ADN empaquetado en los núcleos de todas las células que forman nuestro organismo nos permitiría enlazar la Tierra con el Sol...

¡casi 1700 unidades astronómicas!

Nota final: la distancia Sol-Neptuno se estima en 30,1 u.a. ¡Podríamos unir el Sol y Neptuno 57 veces!



Genoma Humano

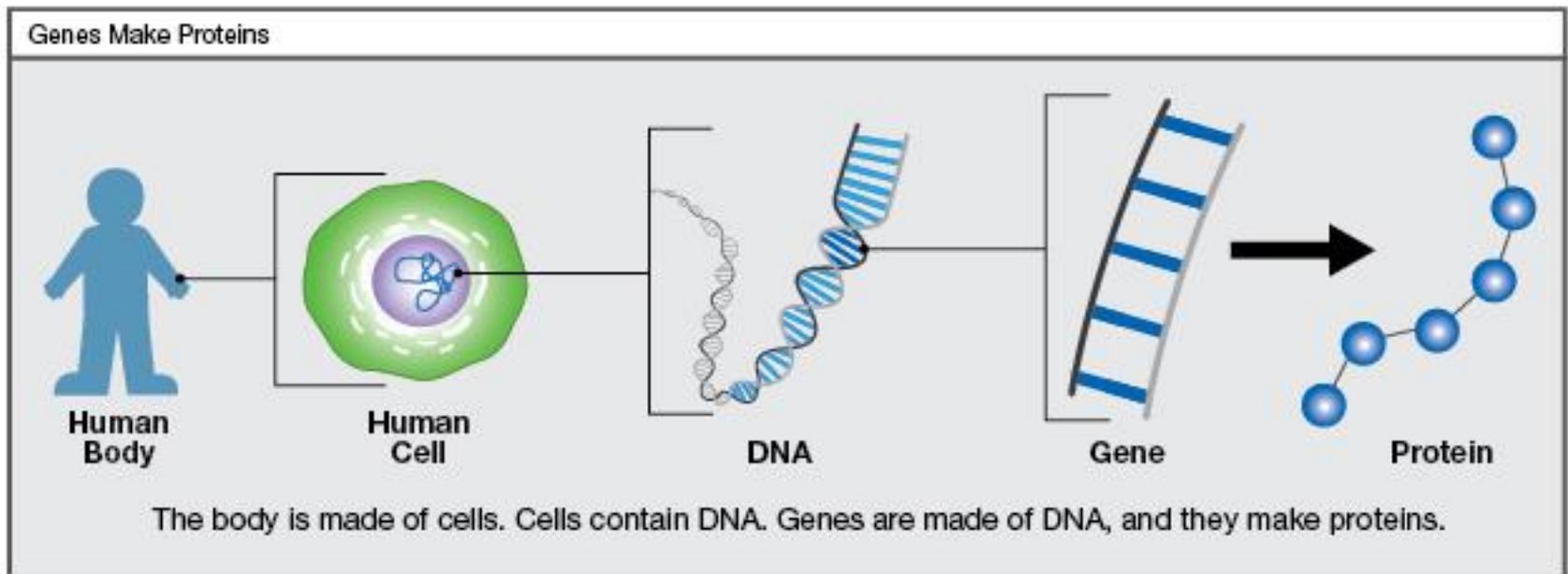
- A. Publicado en 2003
- B. 3.000 millones de dólares
- C. Disponible en Internet



Del genoma al gen

Dogma central de la Biología Molecular.

Gen → (ARN) → proteína



Elementos de un gen

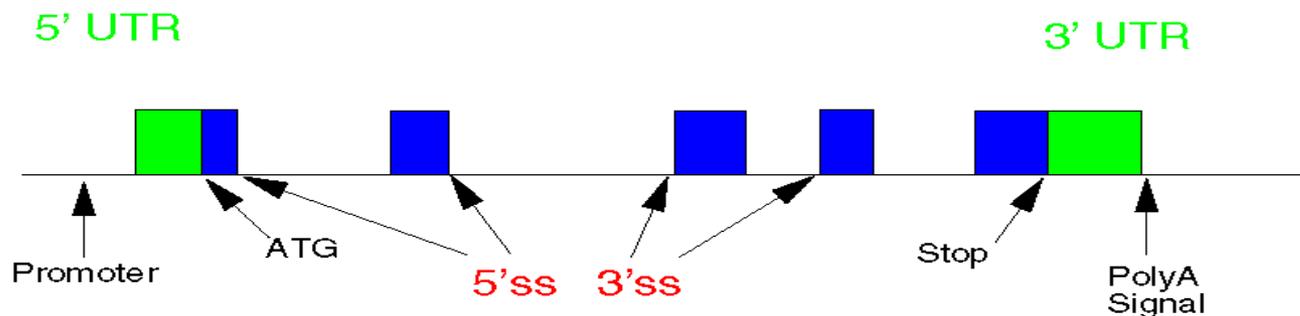
Dogma central de la Biología Molecular:

Gen → (ARN) → proteína

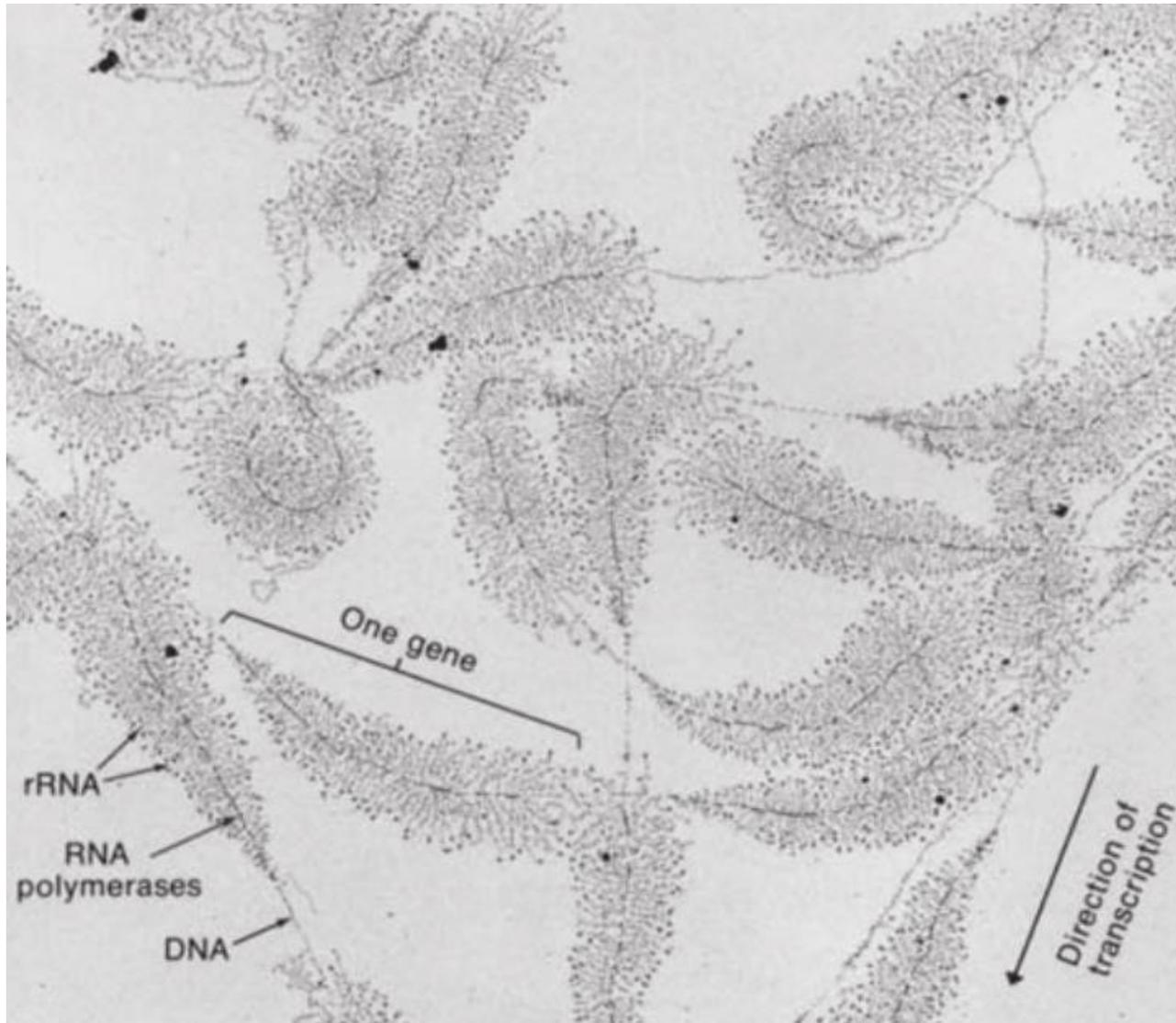
Longitud variable, mayoría en 2-200 kb, aunque sólo una pequeña proporción codifica la proteína. Ejemplo: gen TP53

Elementos diferenciados

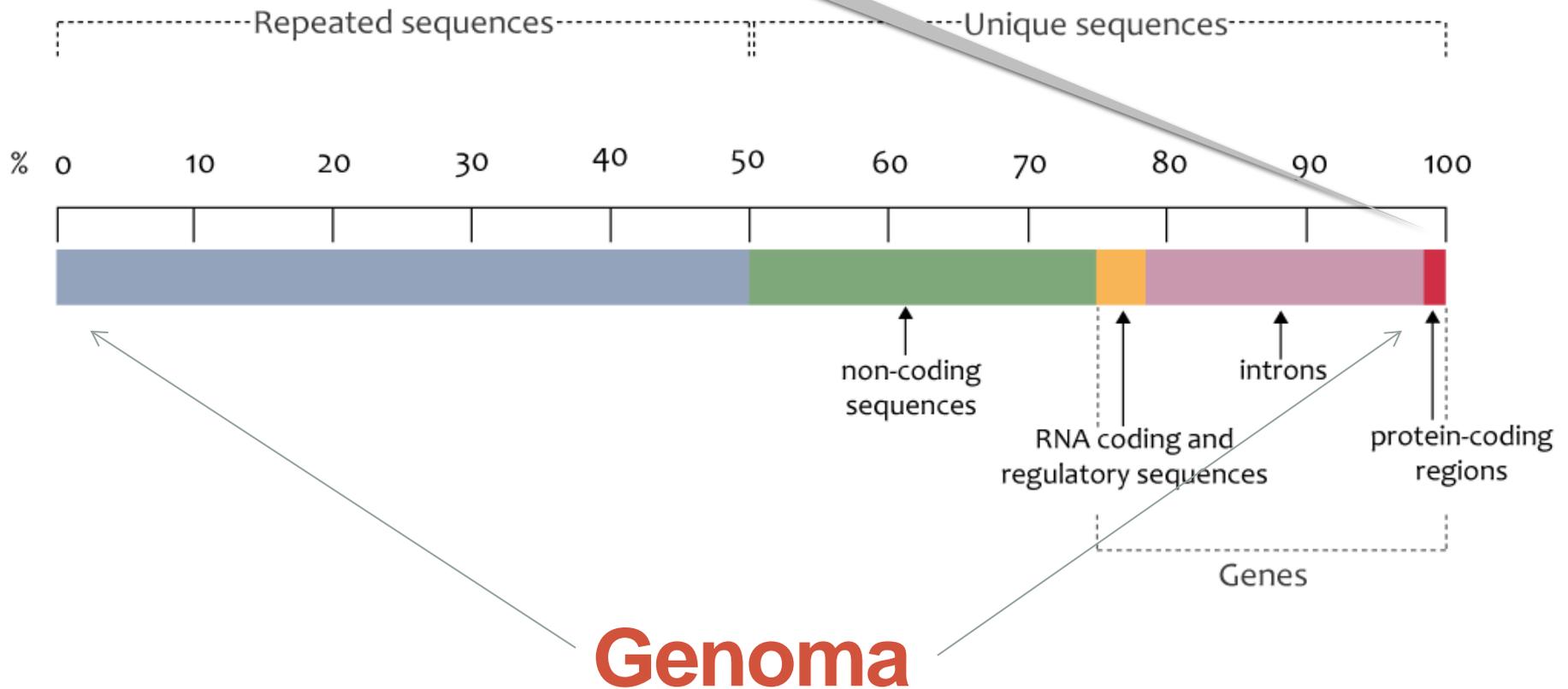
Destacan los **exones**; representados en ARN maduro



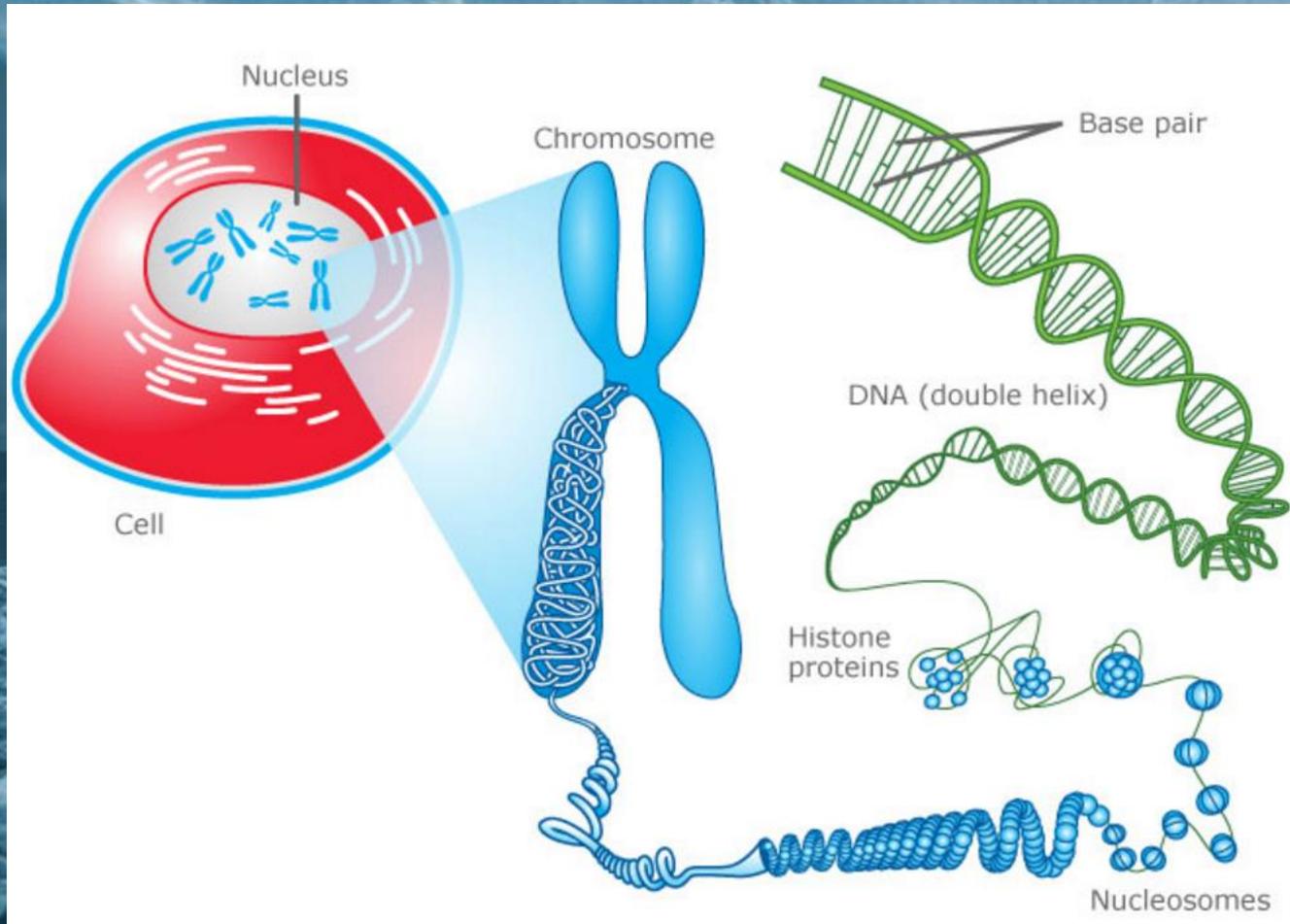
Maquinaria de transcripción génica



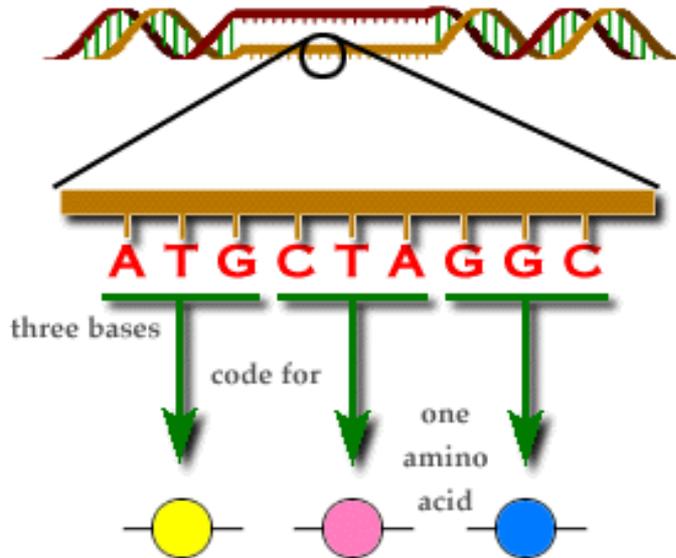
Exoma



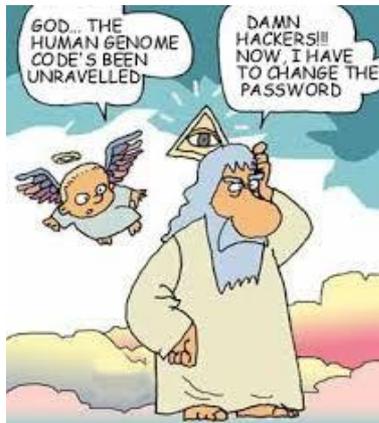
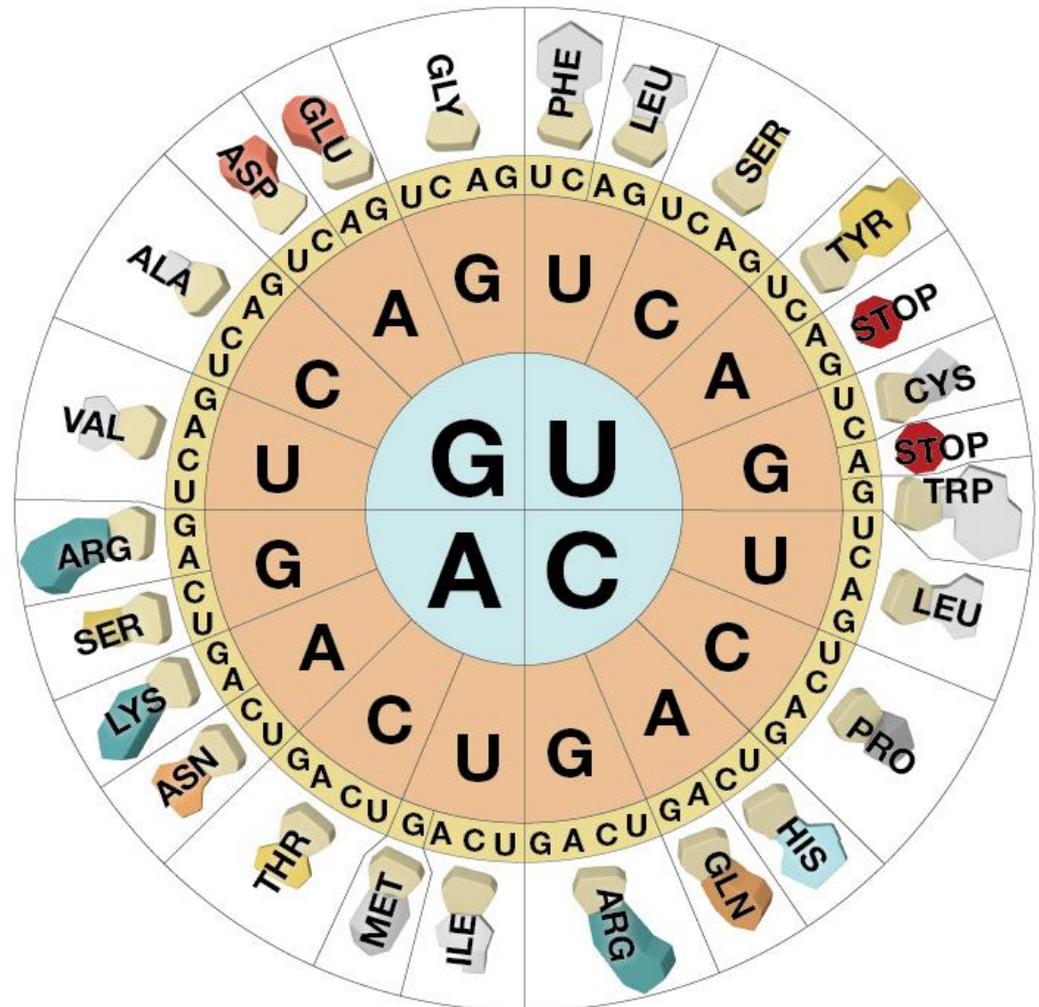
Recapitulemos...



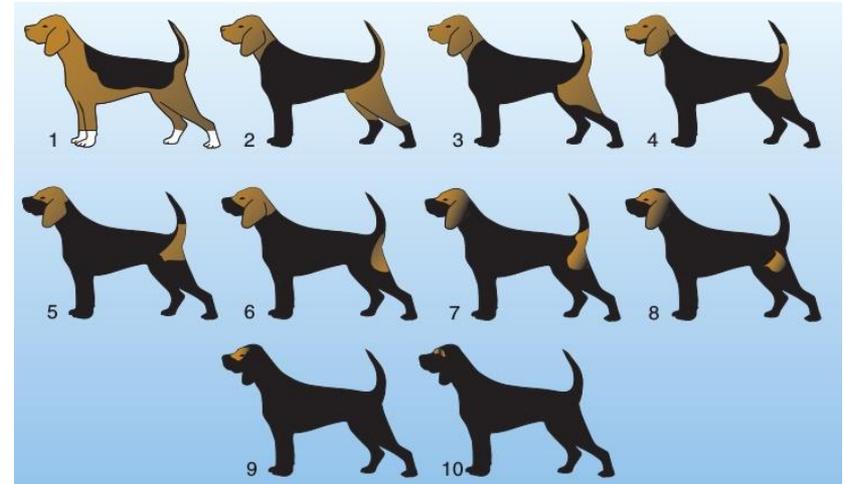
Código genético: codón



T (ADN)=U (ARN)

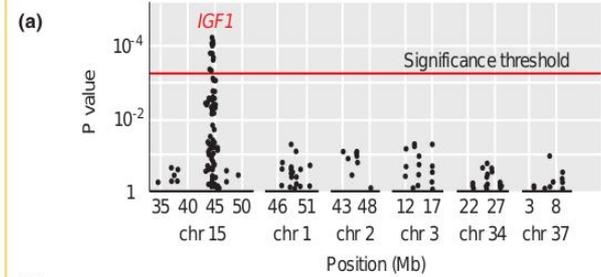


Variación genética



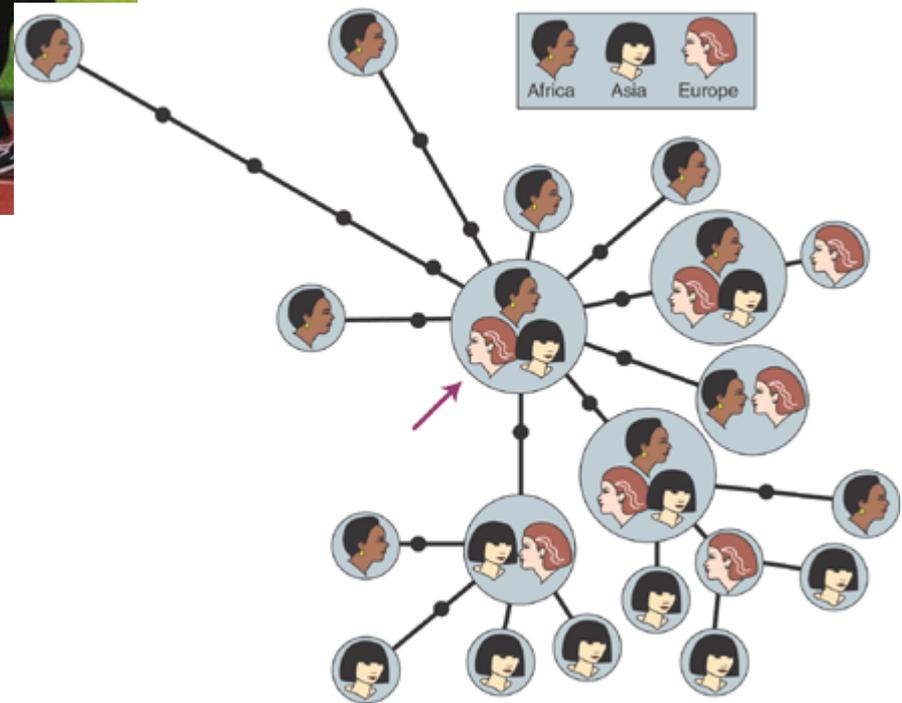


Association mapping finds a gene for body size in dogs



(b)

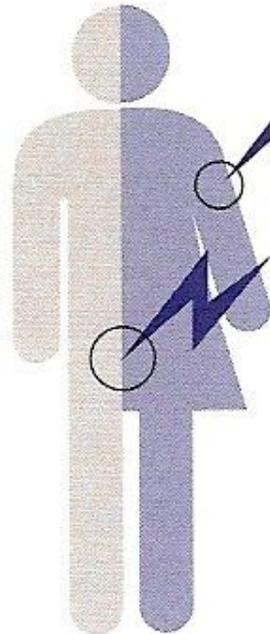




Variación genética

Somatic mutation – non-heritable

- **normal soma:** daughter cells carry mutation – ‘somatic mosaic’
- **cancer:** clonal expansion of mutant cells



Germ-line mutation – heritable

Sangre periférica
(línea germinal)

Tejido tumoral/sano
(línea somática)

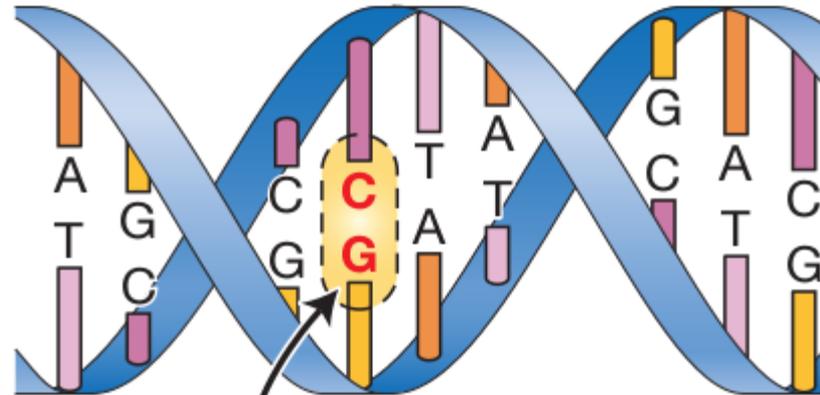


DNA



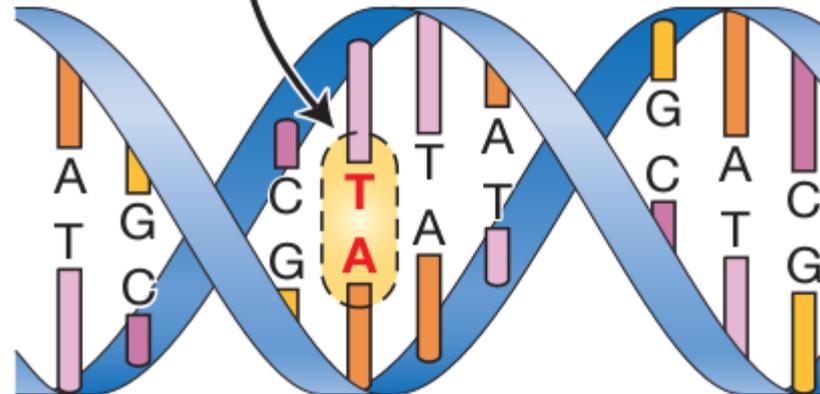
Variación genética simple: SNP

Copia 1 del chr 5

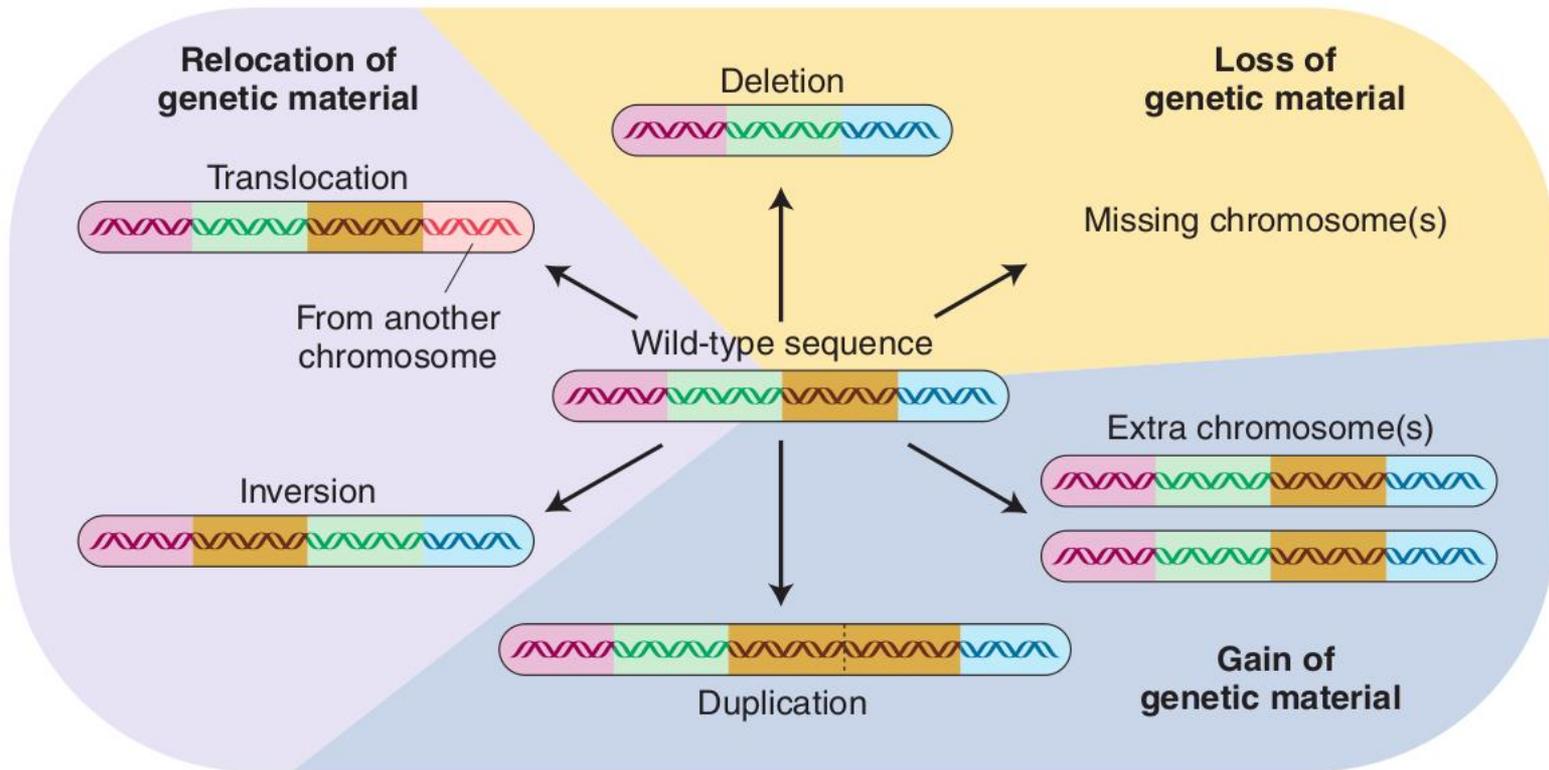


SNP

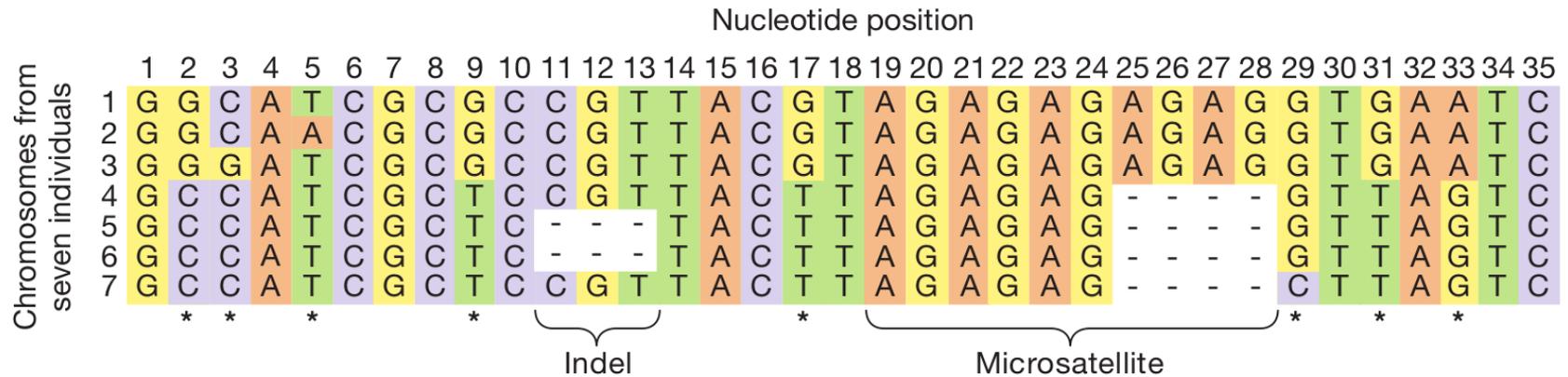
Copia 2 del chr 5



Variación genética estructural



Variación genética: indels, STR

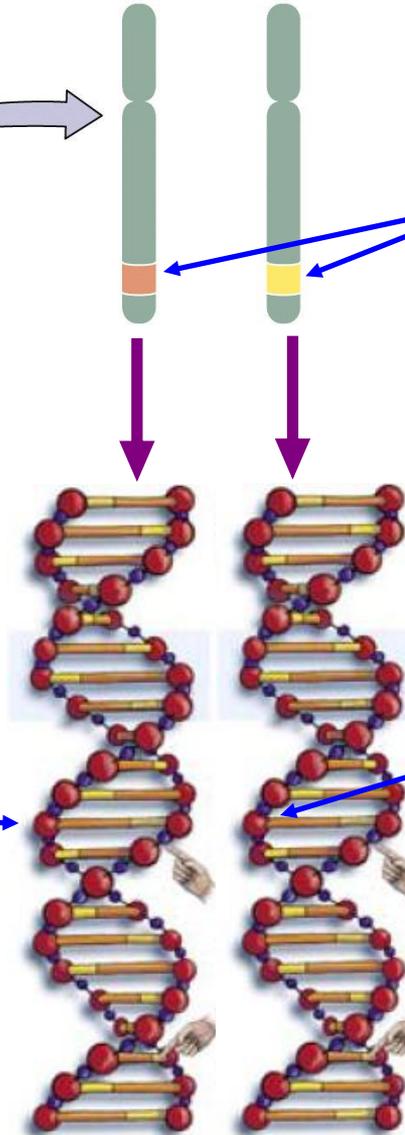
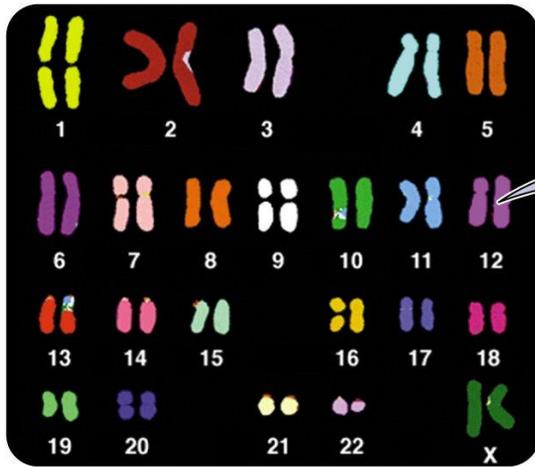


Genotipo

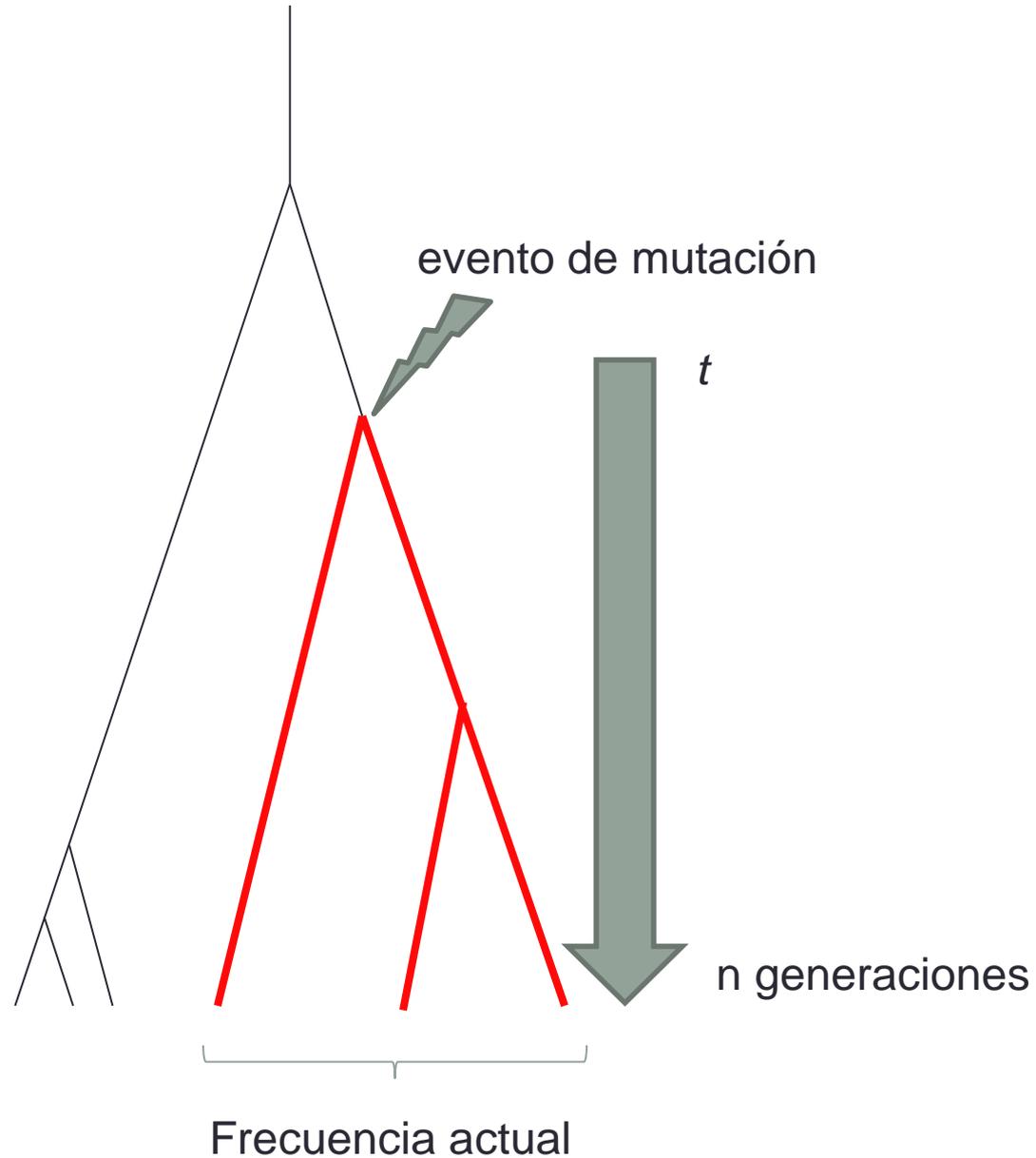
Dos copias de un cromosoma (padre y madre)

En un punto concreto, interrogamos los dos alelos que contiene el individuo.

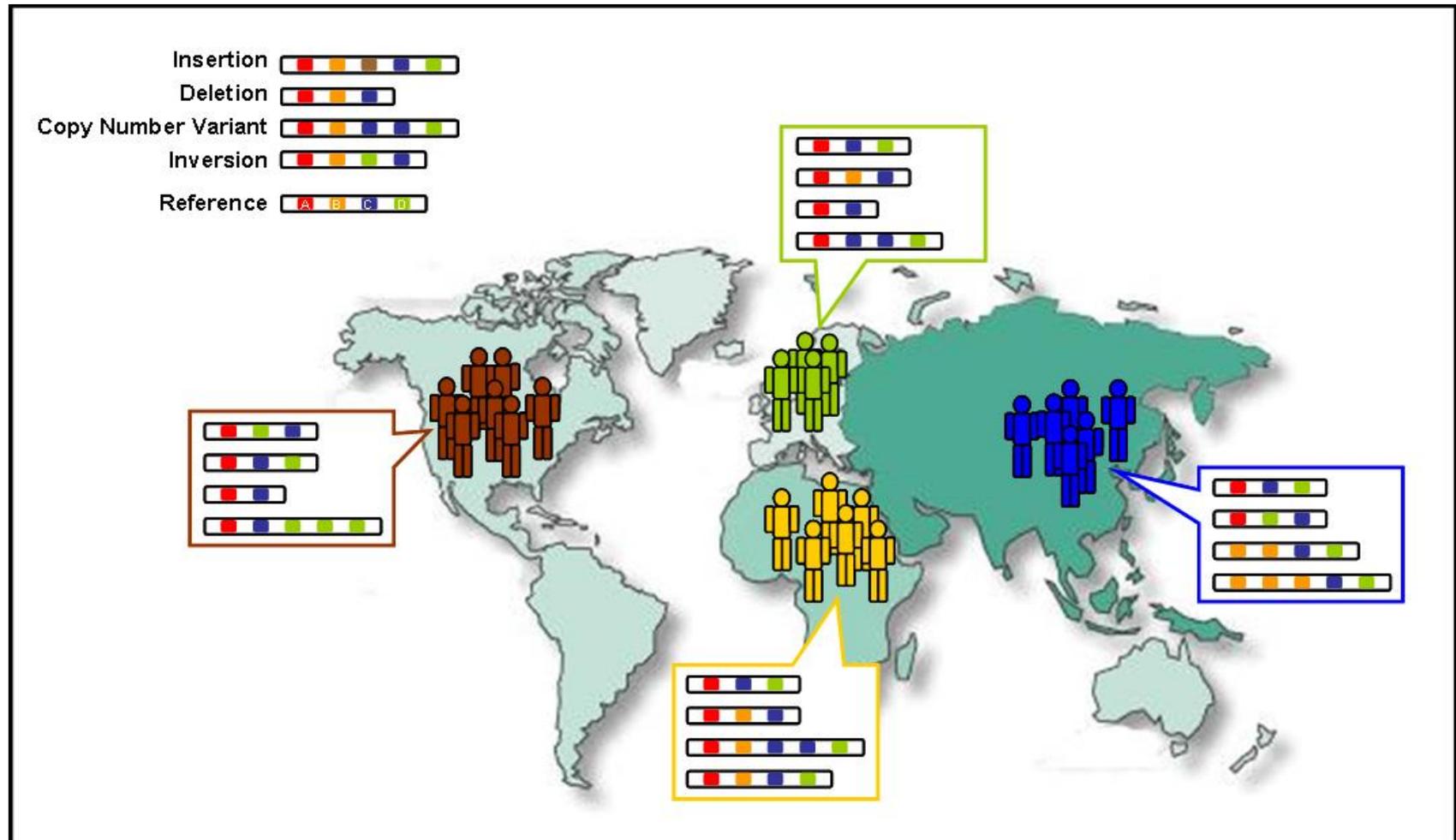
C → T



SNPs: eventos únicos

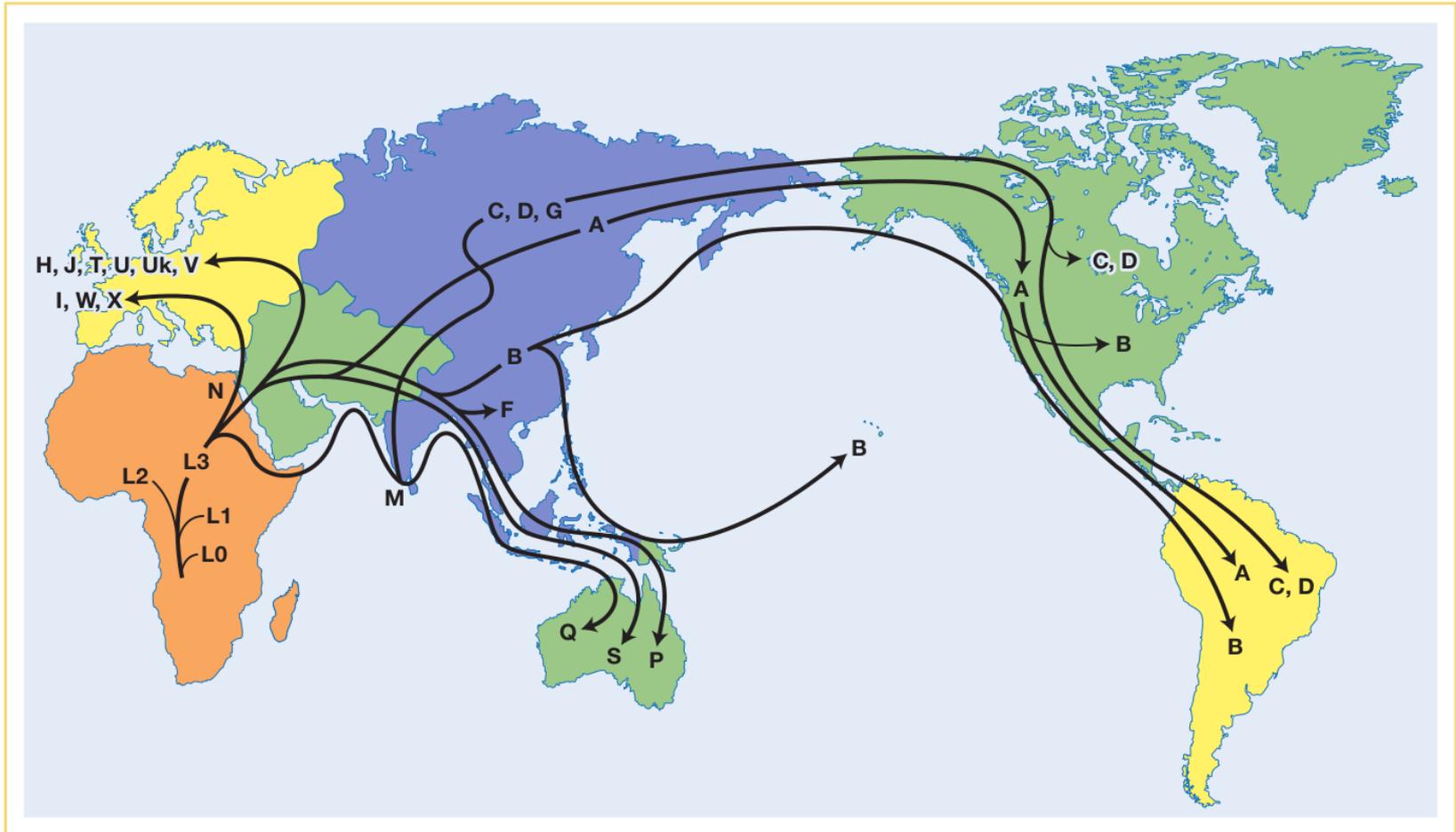


Variación para estimar la ascendencia



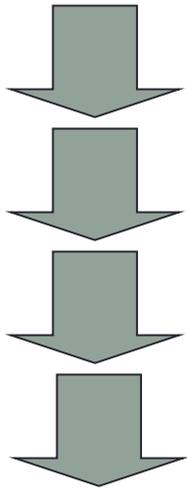
¿De dónde venimos?

Mitochondrial haplotypes can be used to trace human origins to Africa



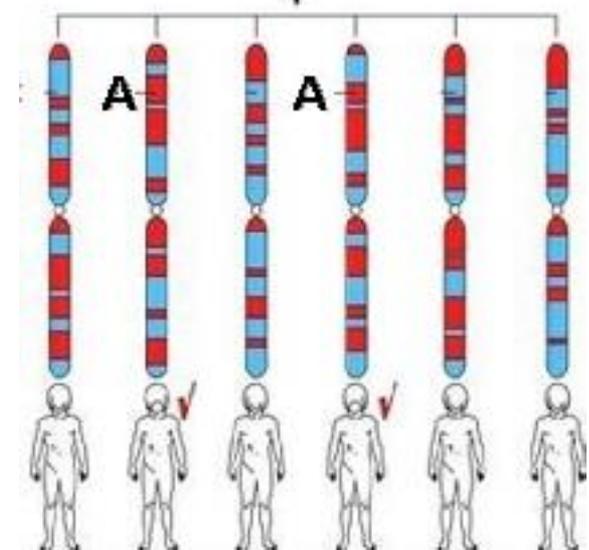
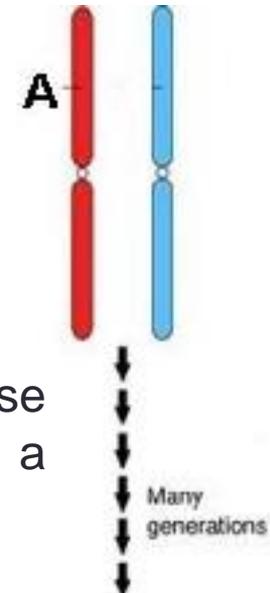
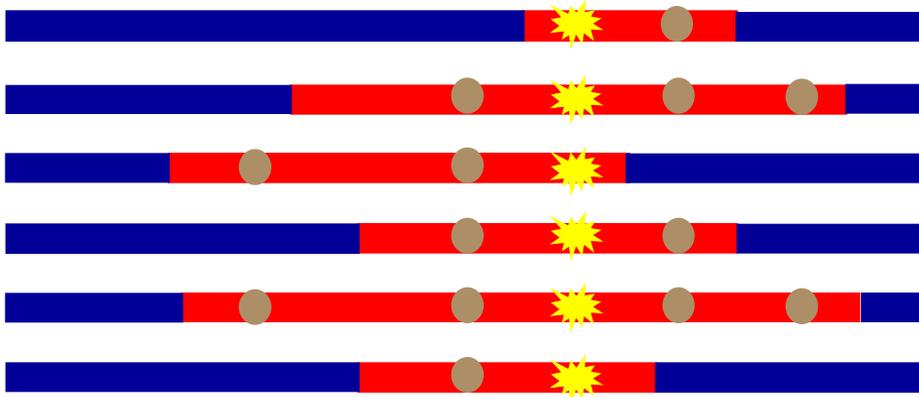
Fundamento

Cromosoma ancestral



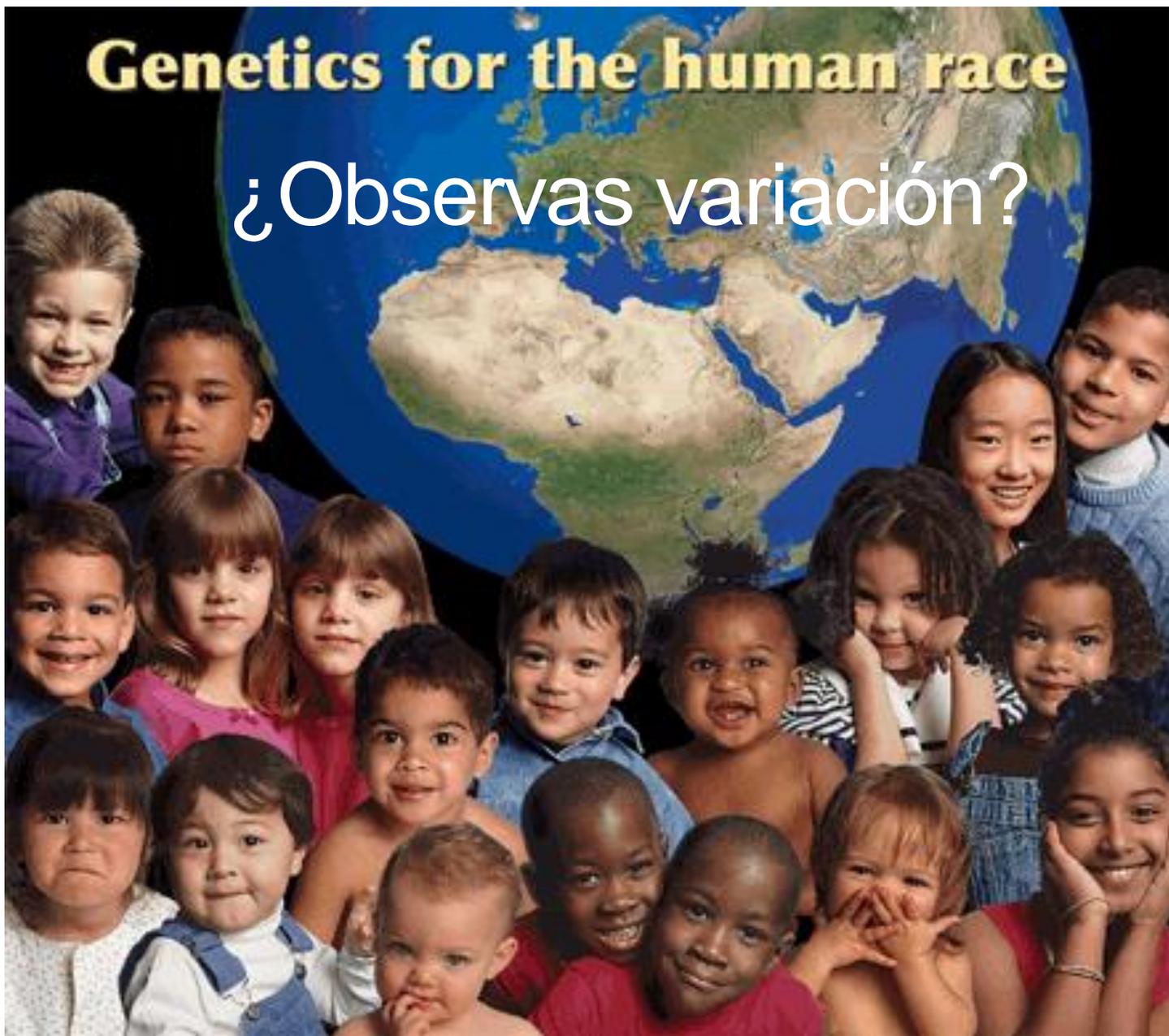
Recombinación/n generaciones

El desequilibrio de ligamiento generalmente se reduce con la distancia parcialmente debido a procesos de recombinación.



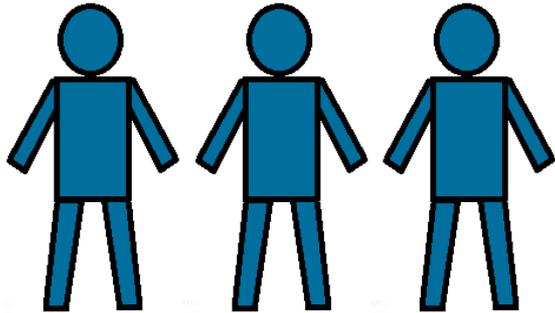
Genetics for the human race

¿Observas variación?

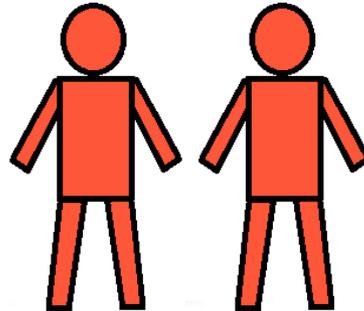


Mezcla en Canarias

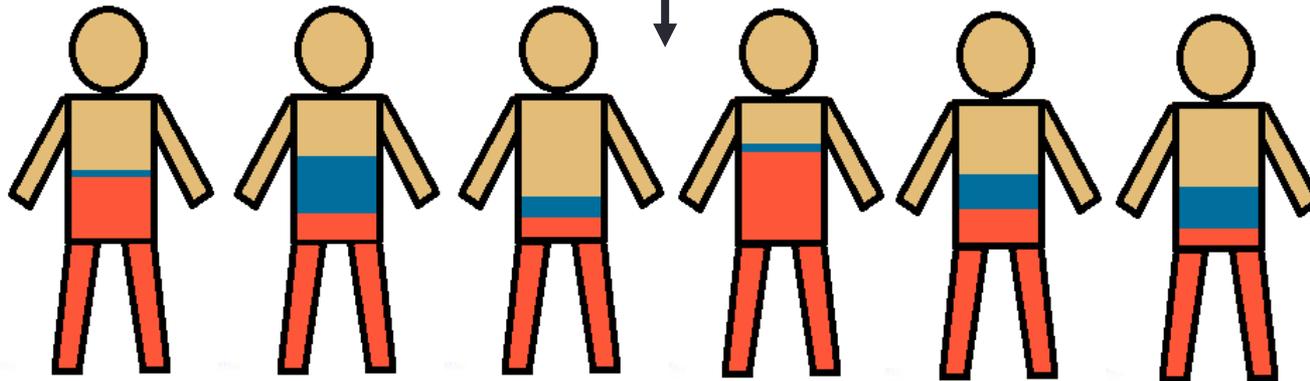
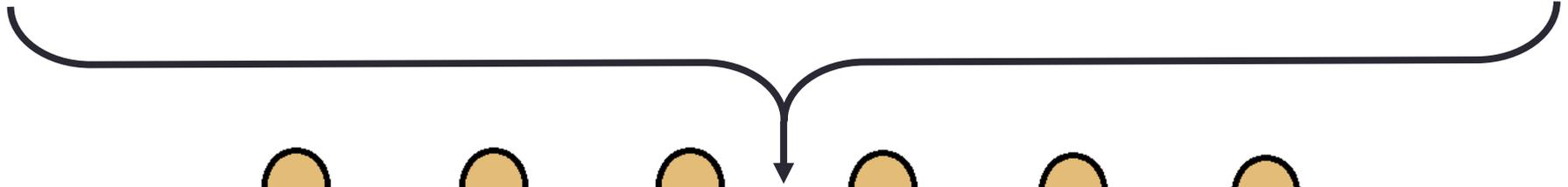
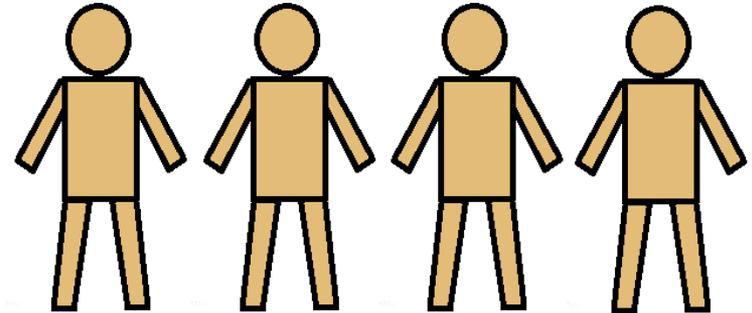
N. África



África SS.



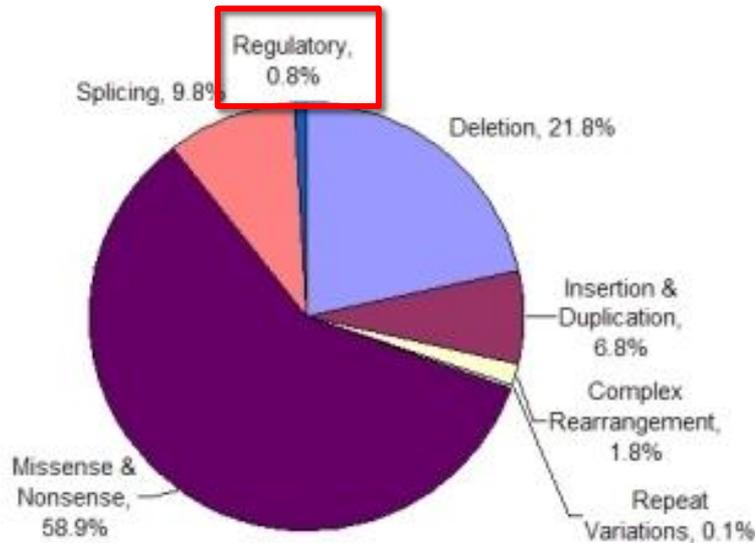
Europeo



Canarios

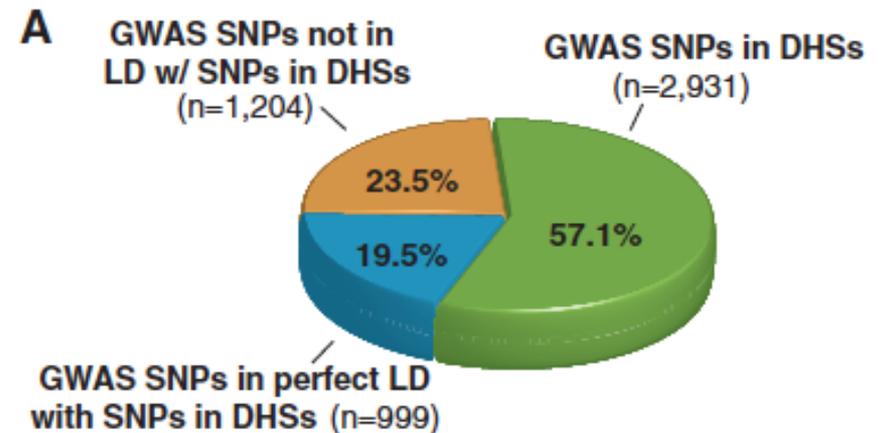
Variación genética y enfermedad

Simple (monogénica, efectos drásticos)



Información conocida gracias a estudios en familias

Complejas (multigénicas, efectos leves)

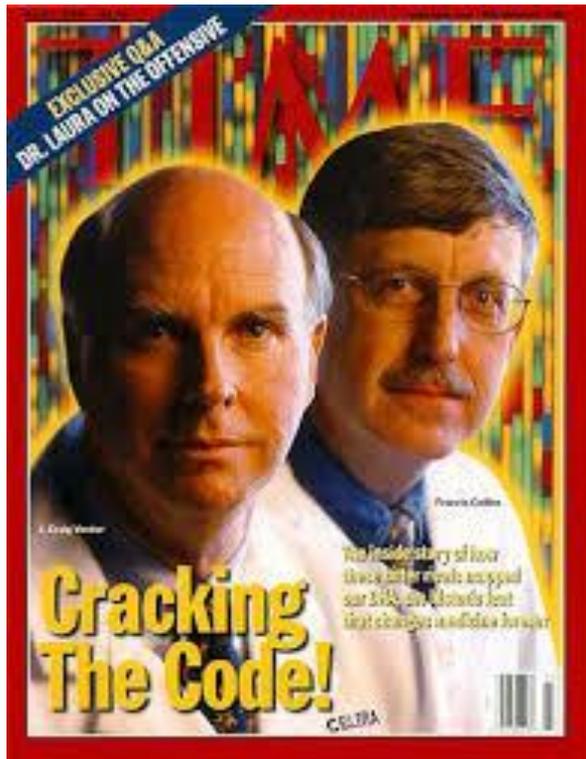


Información conocida por estudios de asociación

Sólo el 16% de las variantes en regiones exónicas

Gracias a varios avances

Secuencia del genoma humano

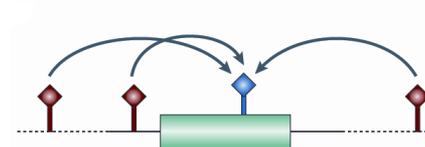


Primer borrador en 2003

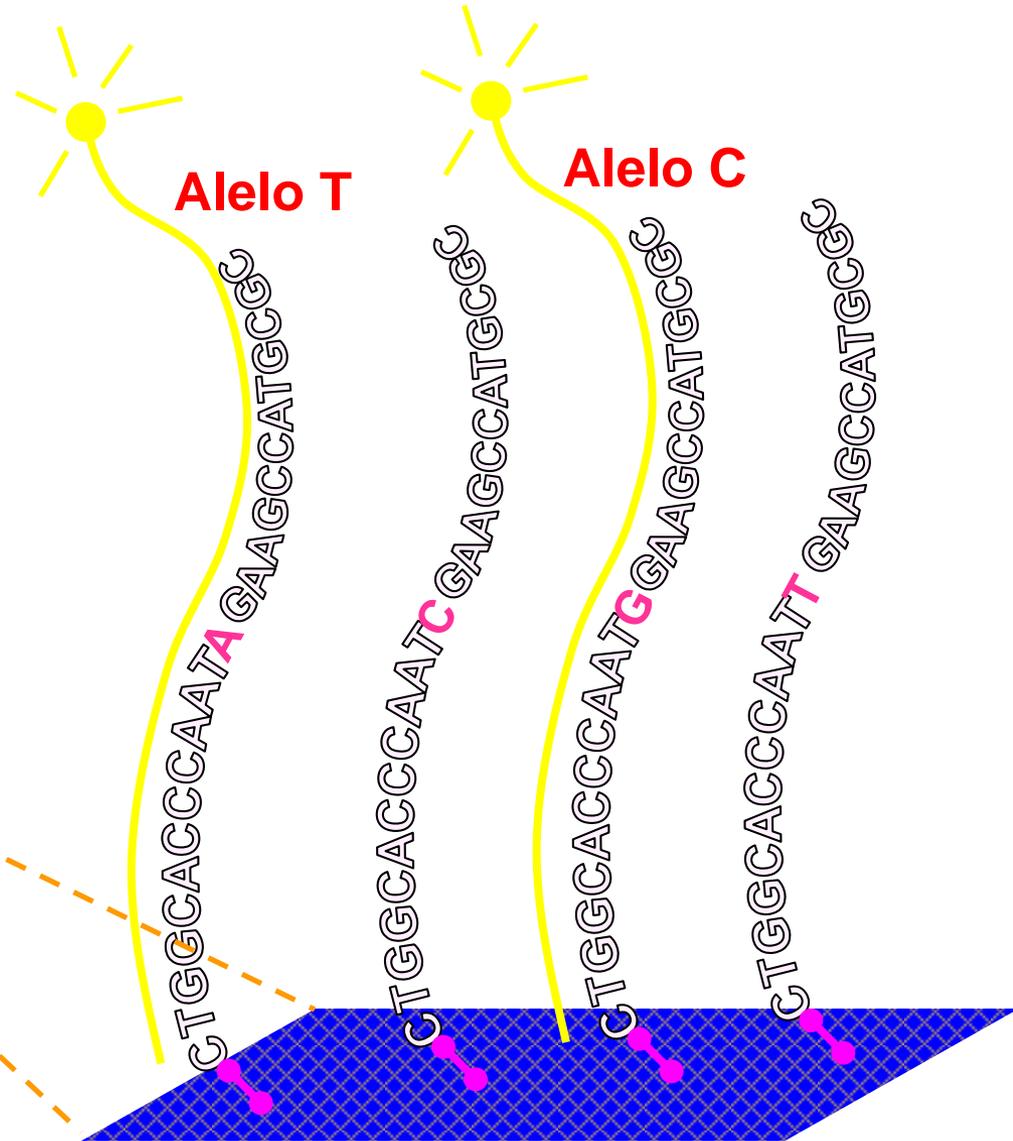
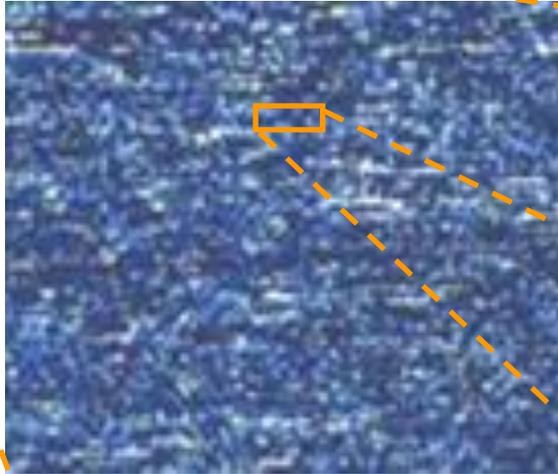
Desarrollo de chips de ADN



Basado en renaturalización del ADN



Chips de ADN

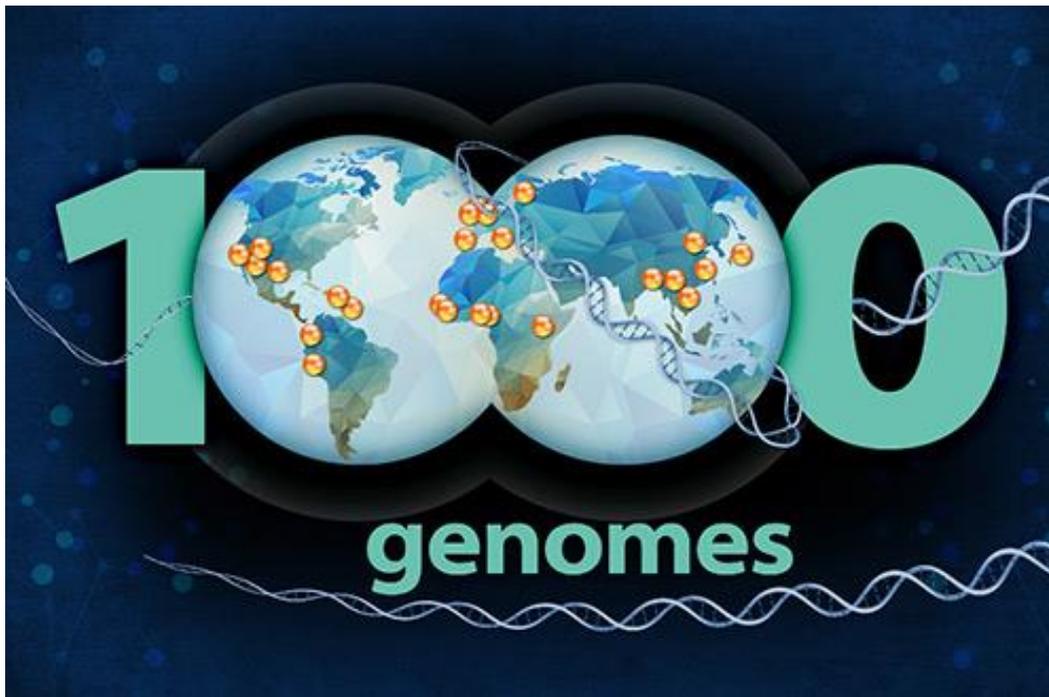


Secuelas

Mapa

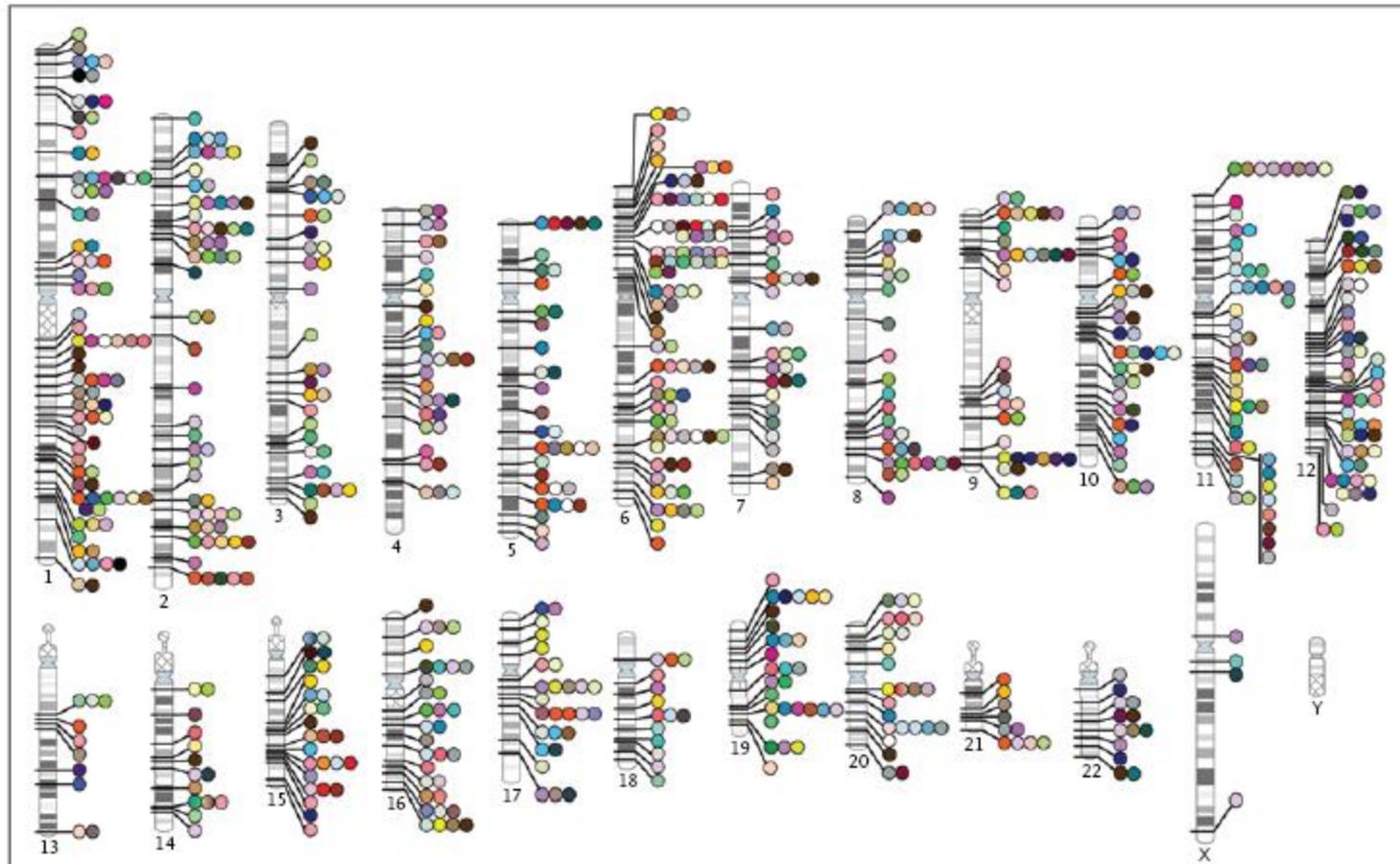


3,5 Millones de SNPs
3 poblaciones (N=270)

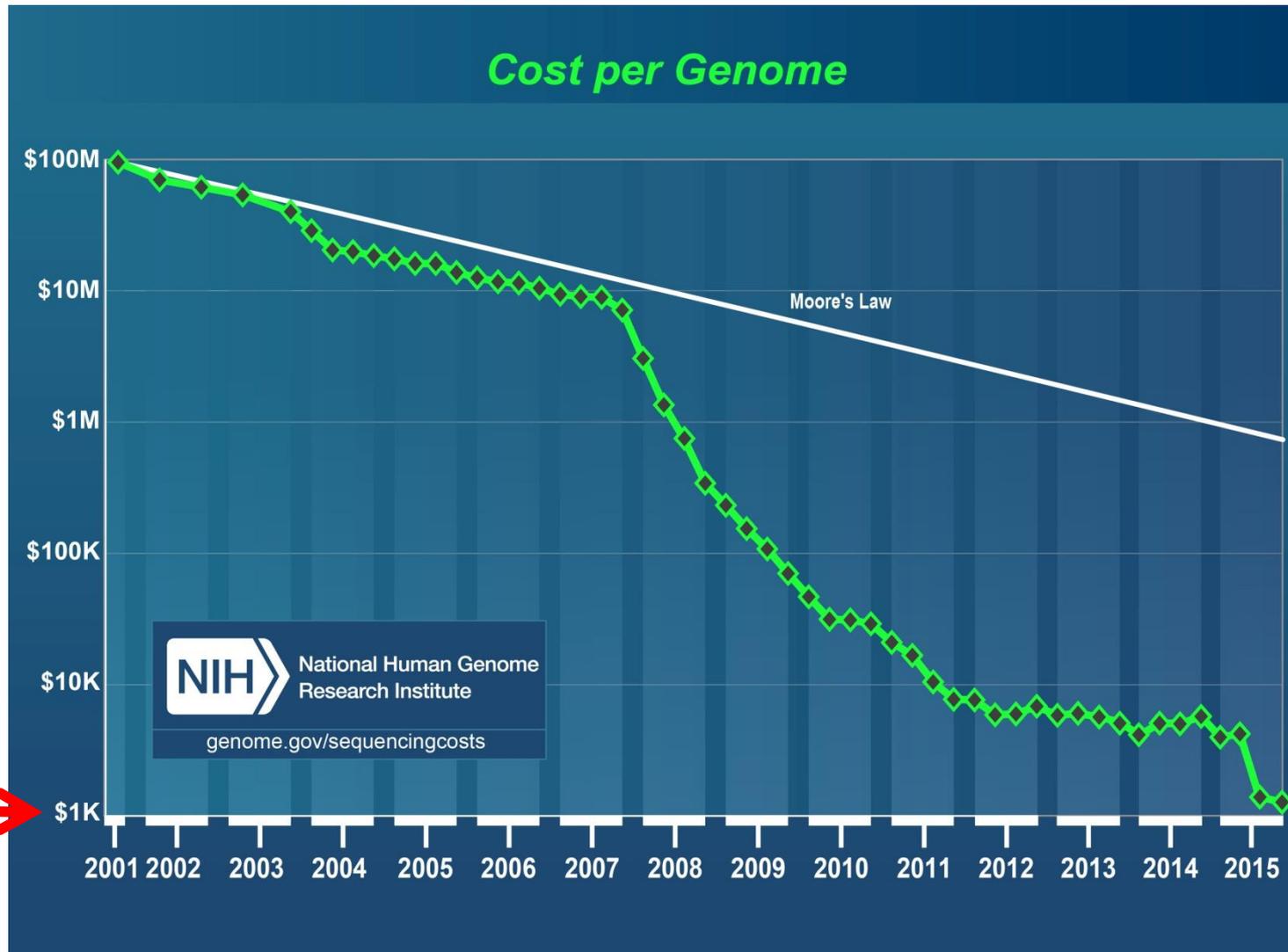


38 Millones de SNPs
14 poblaciones (N=2504)

Identificación de cientos de genes de enfermedad (complejas)



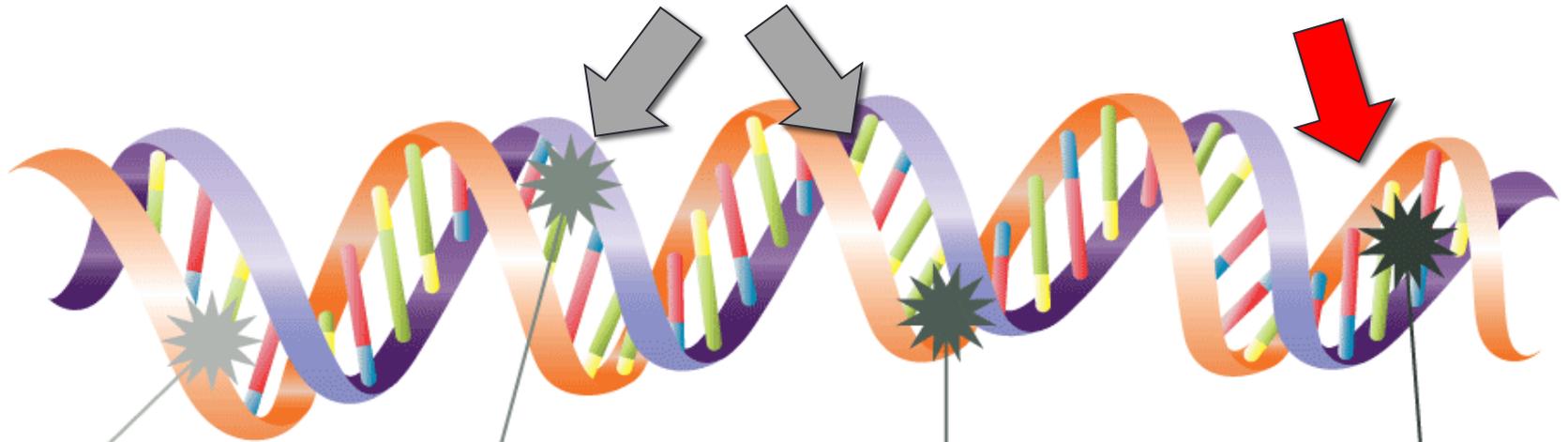
Coste por genoma



Aplicación clínica?

Estudios de asociación

Secuenciación clínica



Normal variation

Eye colour

Differences in response
to diet/medication

Obesity
Effect of antidepressants

Influence likelihood of
disease

Diabetes
Coeliac disease

Directly result in a
genetic condition

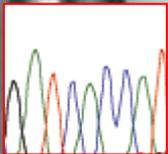
PKU
Familial hypercholesterolaemia

OR >> 2

Secuenciación de ADN

Sanger

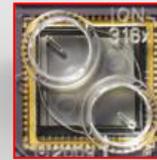
1990-2003 Proyecto del Genoma Humano
- 3200 millones de bases
- \$300 millones



1º Reacción secuenciación y 2º Lectura
De 1 a 96 fragmentos por carrera

Masiva (NGS)

Con la secuenciación de nueva generación el genoma humano en horas!!



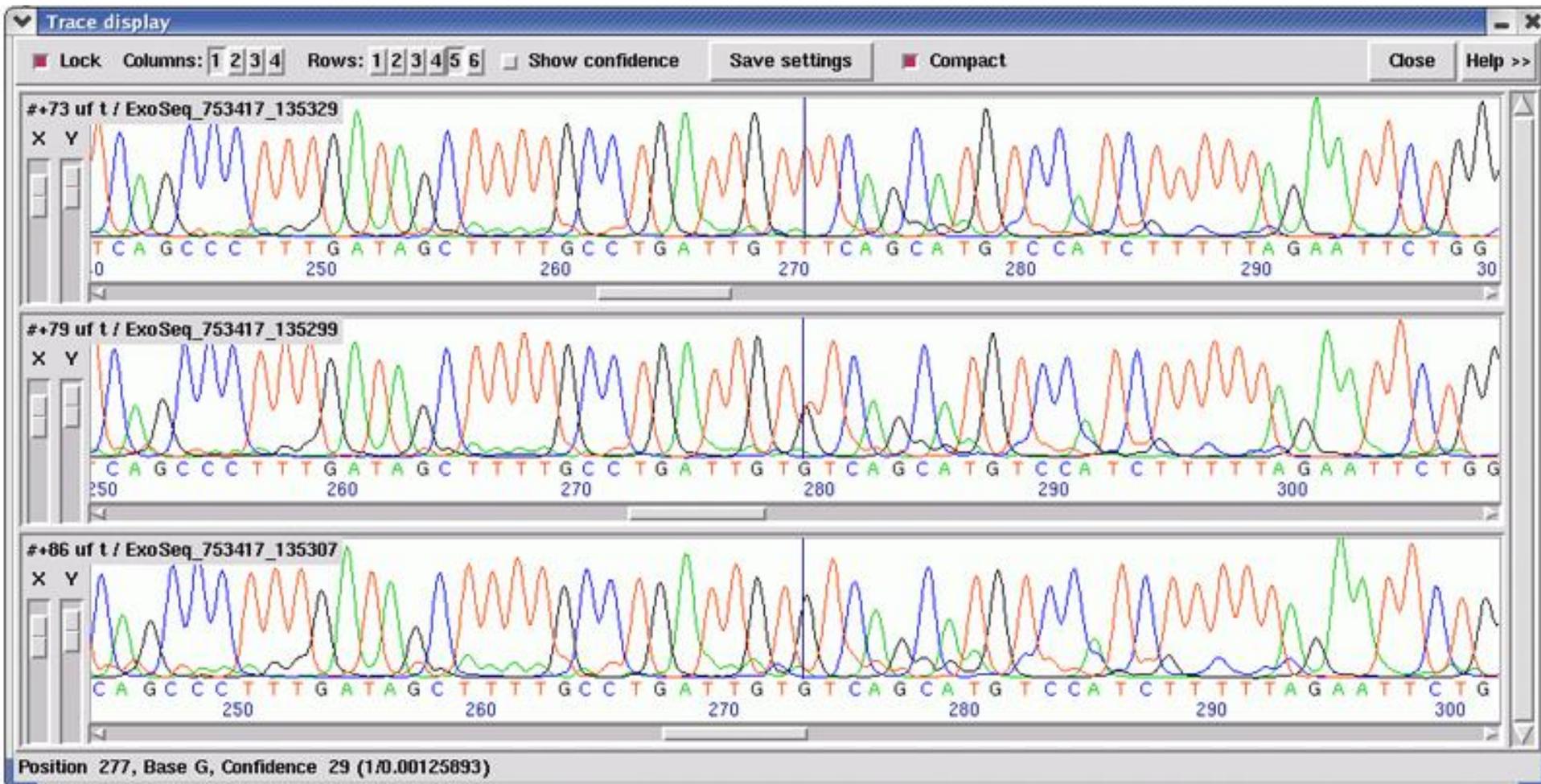
Septiembre 2012

Reacción secuenciación y lectura simultanea
Miles de millones de fragmentos por carrera

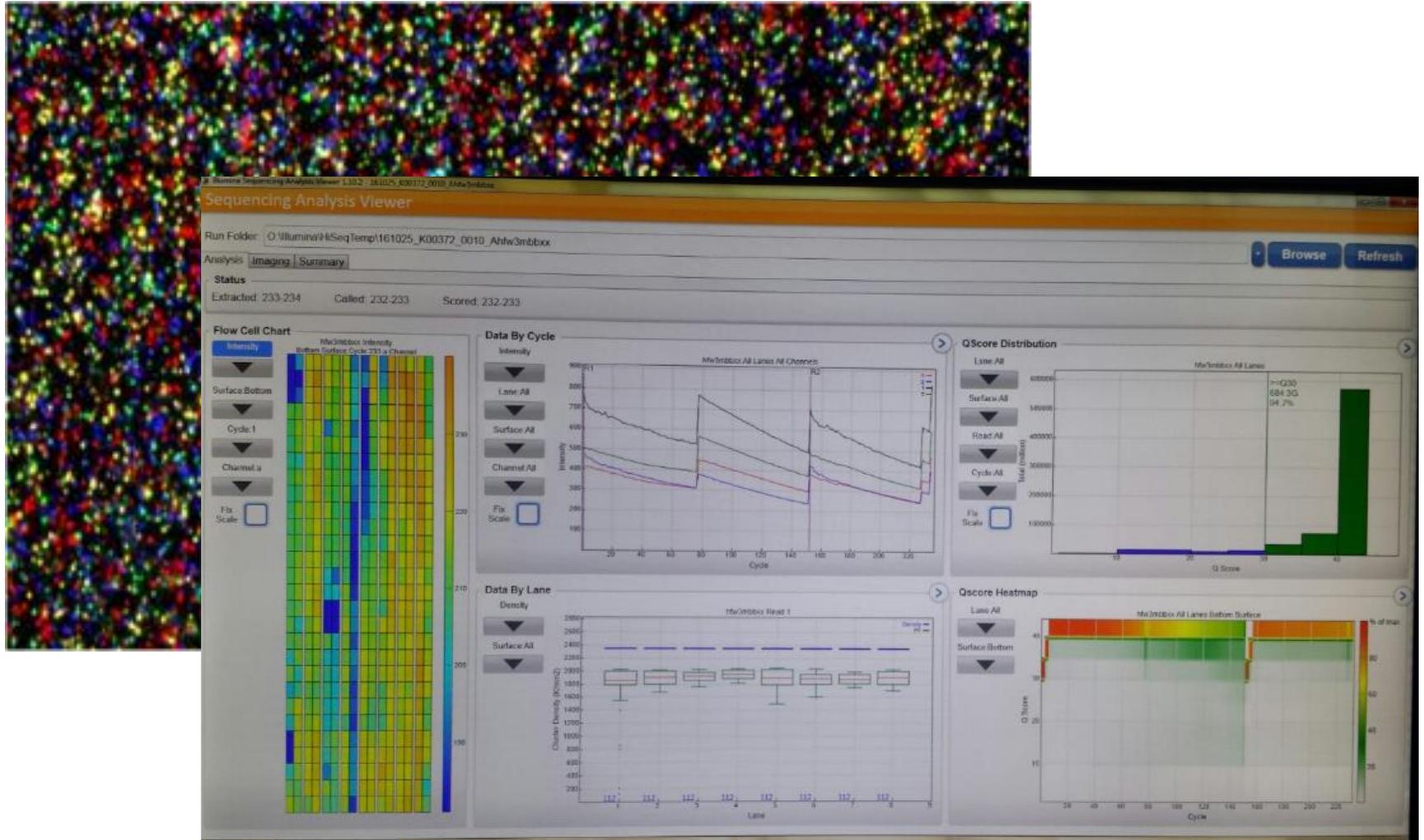
Objetivo: por
\$1.000/genoma

Así lo ve un secuenciador NGS

Tienen al menos dos bases alternativas (A, G, C, T) → Alelos
Aprox. 40 millones descritos



También así...



Así lo vemos con la Bioinformática...

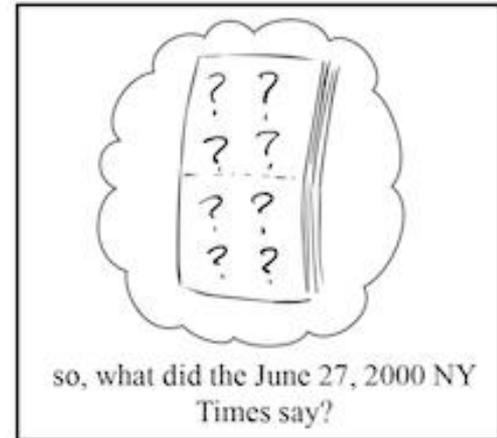
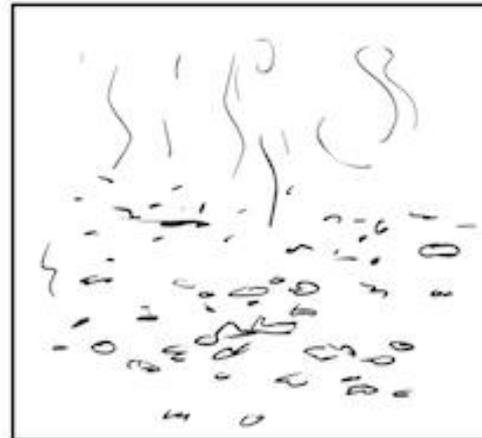
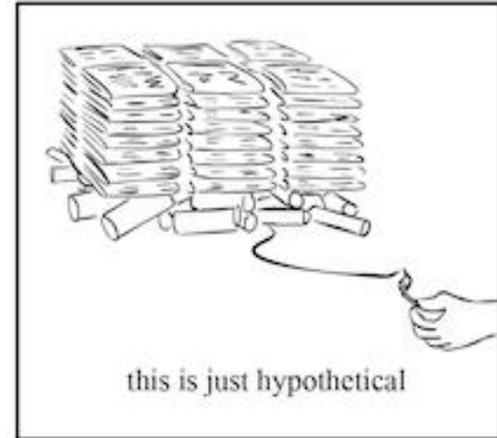
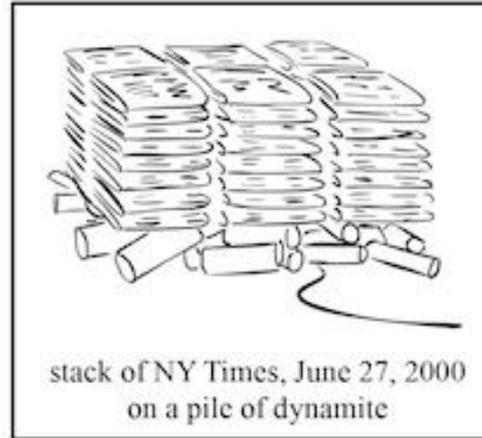
```
61      71      81      91      101     111     121     131
CACACACATCCTAACACTACCCTAACACAGCCCTAATCTAACCCCTGGCCAACCTGTCTCTCAACTTACCCTCCATTACCCTGCCTCCACTCGTTA
.....
CACACACATCCTAACACTACCCTAACACAGCCCTAATTTAAACCCTGGCCAACCTGT TCAACTTACCCTCCATTACCCTGCCTCCACTCGTTA
CACACACATCCTAACACTACCCTAACACAGCCCTAATCTAACCCCTGGCCAACCTGTCTCT CTTACCCTCCATTACCCTGCCTCCATTGTTA
CACACACATCCTAACACTACCCTAACACAGCCCTAATCTAACCCCTGGCCAACCTGTCTCT CACTCGTTA
CACACACATCCTAACACTACCCTAACACAGCCCTAATCTAACCCCTGGCCAACCTGTCTCTCA TTA
CACACACATCCTAACACTACCCTAACACAGCCCTAATCTAACCCCTGGCCAACCTGTCTCTCAACTTACCCTCCATT TA
CACACACATCCTAACACTACCCTAACACAGCCCTAATCTAACCCCTGGCCAACCTGTCTCTCAACTTACCCTCCATTACCCTG
CACACACATCCTAACACTACCCTAACACAGCCCTAATCTAACCCCTGGCCAACCTGTCTCTCAACTTACCCTCCATTACCCTGCC
CACACACATCCTAACACTACCCTAACACAGCCCTAATCTAACCCCTGGCCAACCTGTCTCTCAACTTACCCTCCATTACCCTGCCTCC
CACACACATCCTAACACTACCCTAACACAGCCCTAATCTAACCCCTGGCCAACCTGTCTCTCAACTTACCCTCCATTACCCTGCCTCCACTCGTTA
CACACACATCCTAACACTACCCTAACACAGCCCTAATCTAACCCCTGGCCAACCTGTCTCTCAACTTACCCTCCATTACCCTGCCTCCACTCGTTA
ACACACATCCTAACACTACCCTAACACAGCCCTAATCTAACCCCTGGCCAACCTGTCTCTCAACTTACCCTCCATTACCCTGCCTCCACTCGTTA
CTAACACTACCCTAACACAGCCCTAATCTAACCCCTGGCCAACCTGTCTCTCAACTTACCCTCCATTACCCTGCCTCCACTCGTTA
CCCTAACACAGCCCTAATCTAACCCCTGGCCAACCTGTCTCTCAACTTACCCTCCATTACCCTGCCTCCACTCGTTA
CAGCCCTAATCTAACCCCTGGCCAACCTGTCTCTCAACTTACCCTCCATTACCCTGCCTCCACTCGTTA
AACCCCTGGCCAACCTGTCTCTCAACTTACCCTCCATTACCCTGCCTCCACTCGTTA
CAACCTGTCTCTCAACTTACCCTCCATTACCCTGCCTCCACTCGTTA
```

Secuenciación de nueva generación



TB de datos en un sólo estudio = fibra óptica, almacenamiento...

Secuenciación de nueva generación



Secuenciación de nueva generación

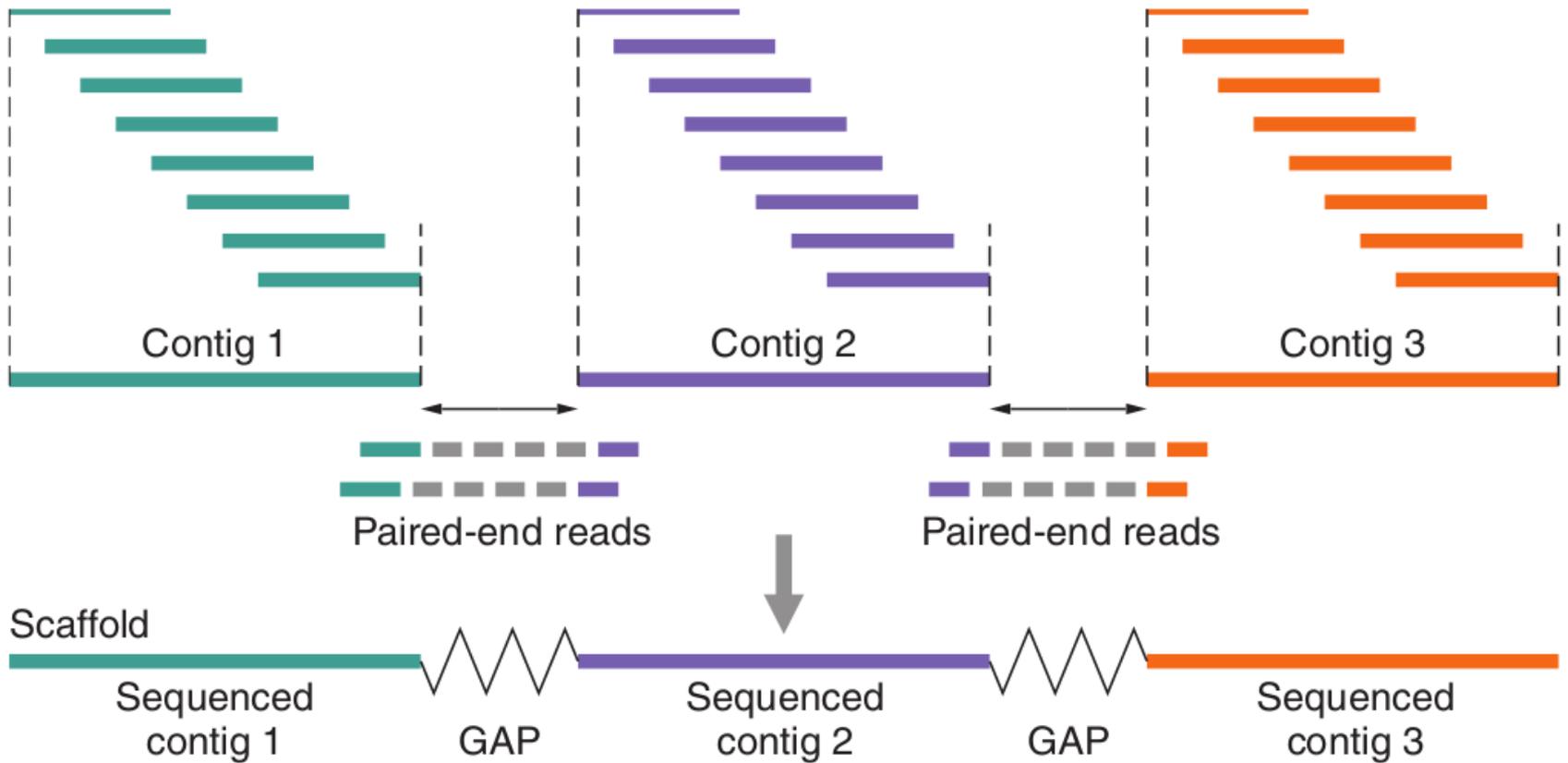
atshirt, approx
e have not yet named a
information is welc

shirt, approximately 6'2" 18
t yet named any suspects
is welcomed. Please call

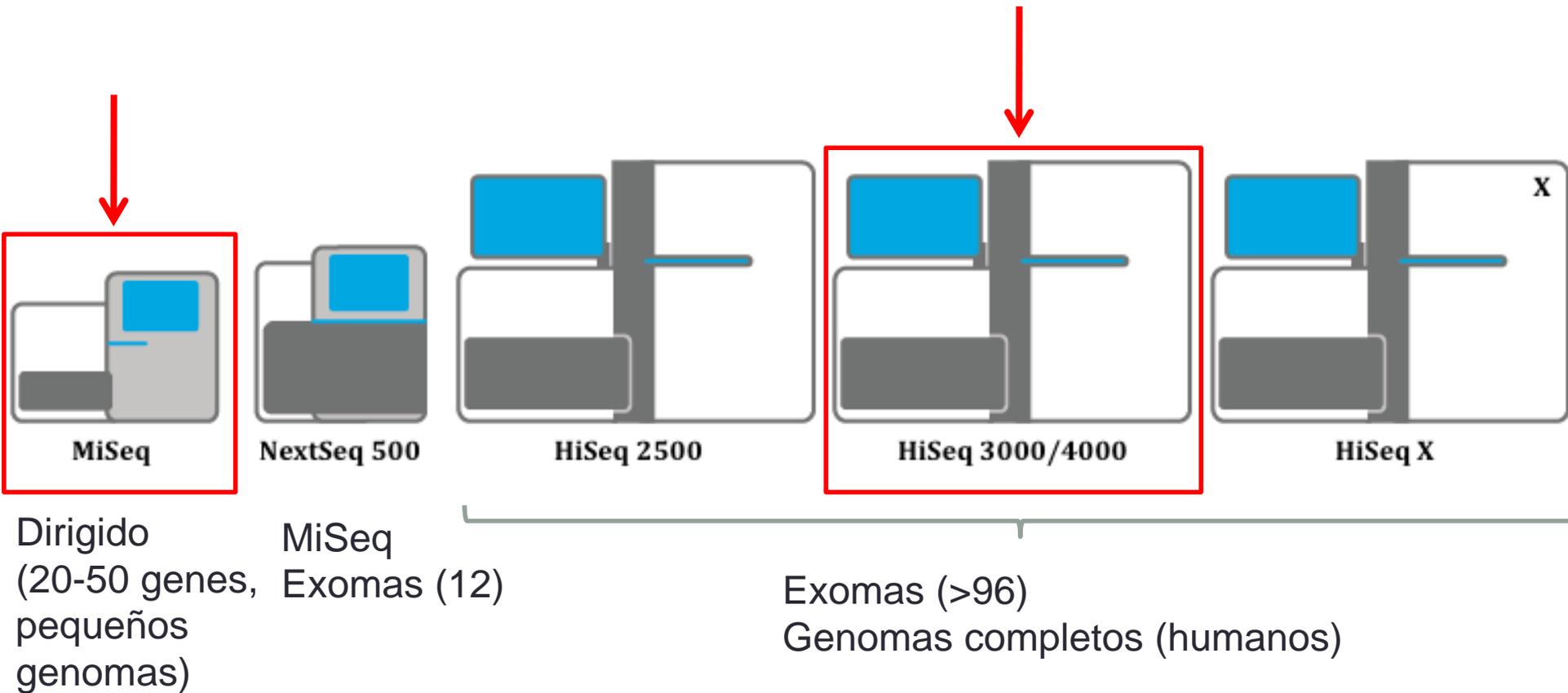
Secuenciación de nueva generación

atshirt, approximately 6'2" 18
e have not yet named any suspects
mation is welcomed. Please call

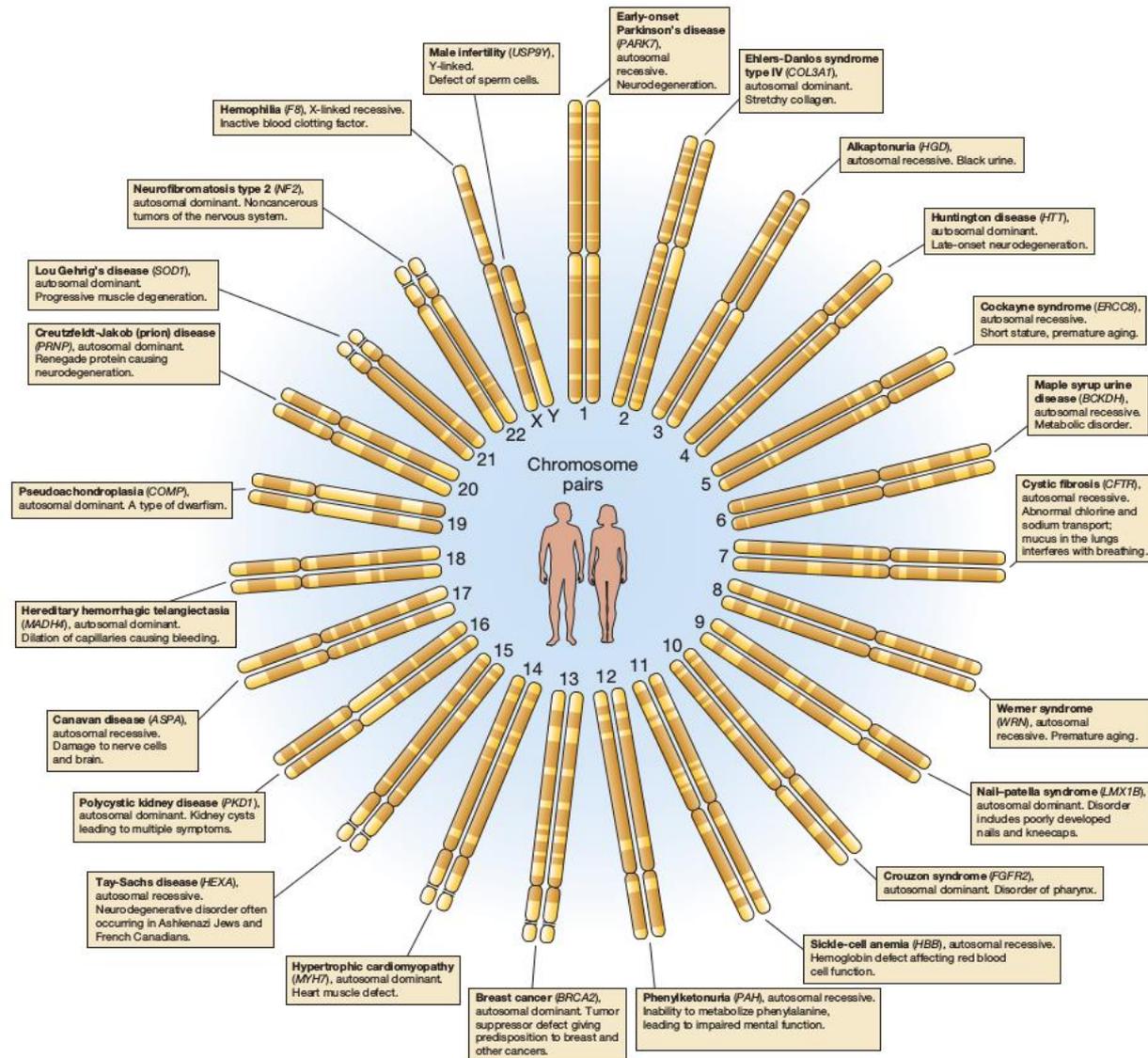
Alineado > Ensamblado > Mapeado



NGS: escalas de producción

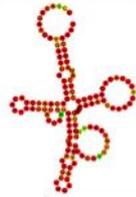


Base genética de las enfermedades

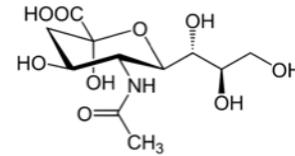


Análisis de datos “ómicos”

Transcriptome
Exome



Metabolomics



DNA

RNA

Proteins

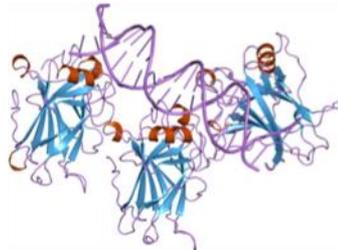
Metabolite

Phenotype

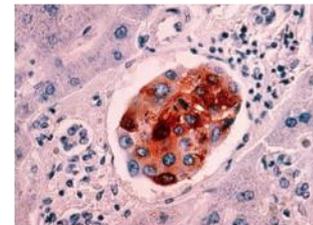
Genome
Epigenome



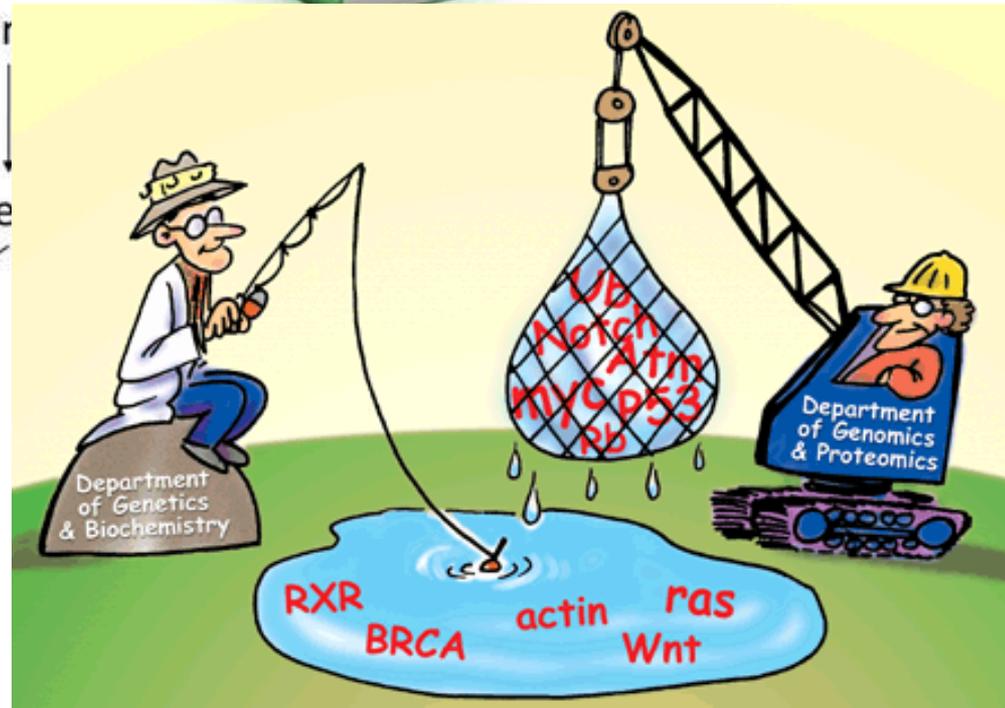
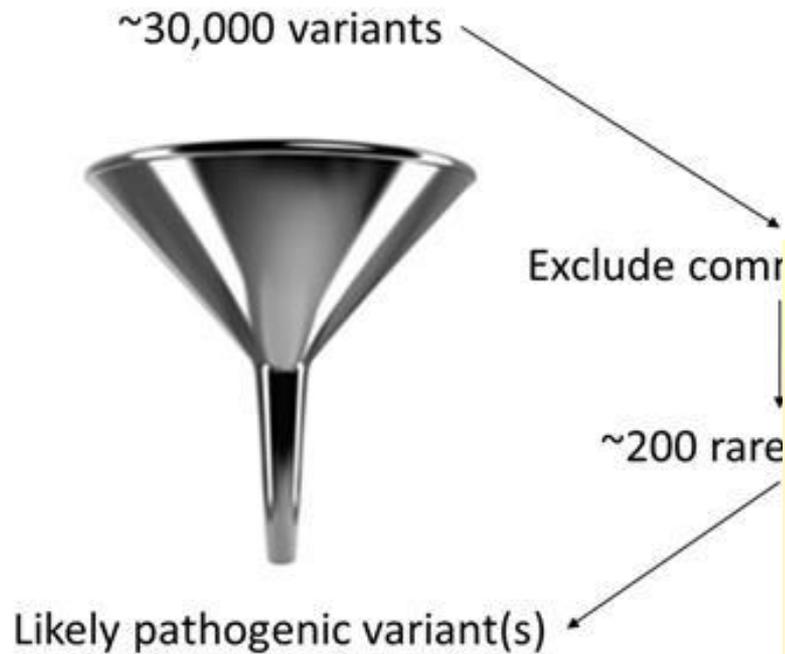
Proteome
protein interactions



Disease

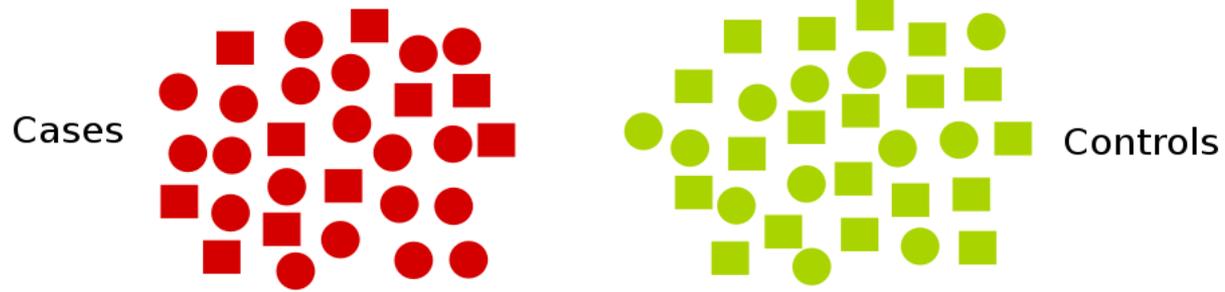


¿Para qué?

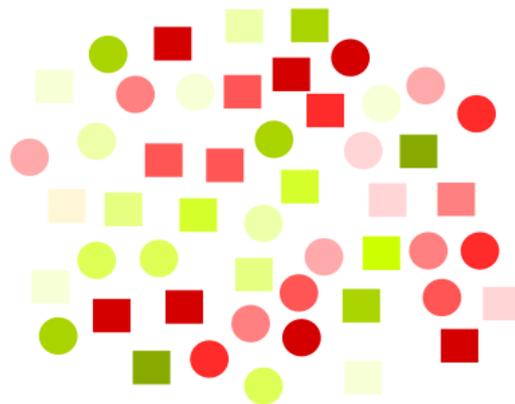


Diseño de experimentos

Rasgos Discretos (Discrete Traits)



Rasgos Continuos (Continuous Traits)



LEGEND

□ Individual not carrying the risk factor

○ Individual carrying the risk factor

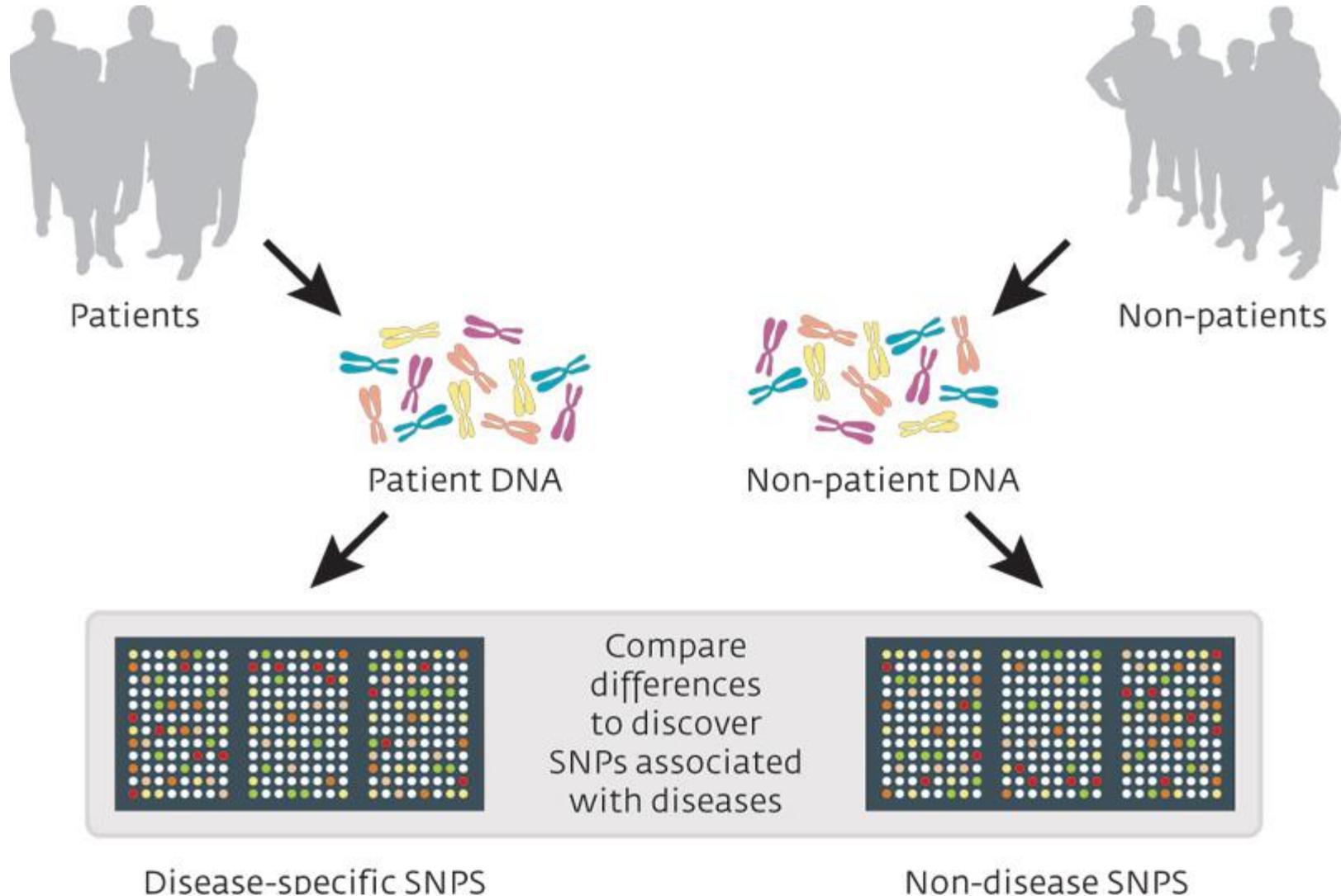
Individual with disease

Individual without disease

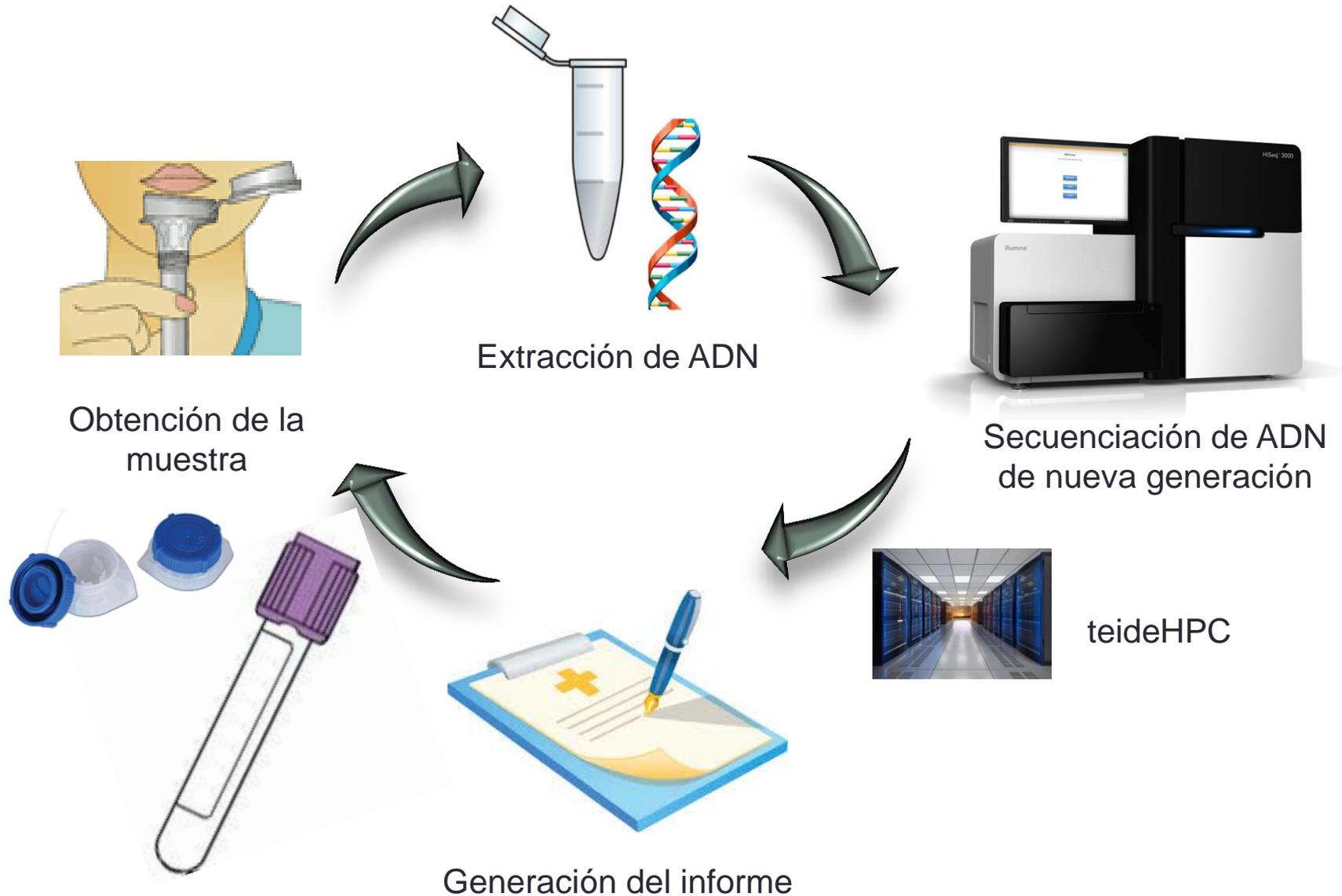
Healthy → Disease



Biología + Estadística, Informática...



Ciclo de análisis NGS





Ref
Reads

TACCGAT
TACCGAT
ACCGAT
ACCGAT
ACCGAT
CCGAT
CCGAT

CATTGGATCA
CATTGGAT
TATTGCAT
CATTGG
TATTG
C-
CATG

Variant
Region

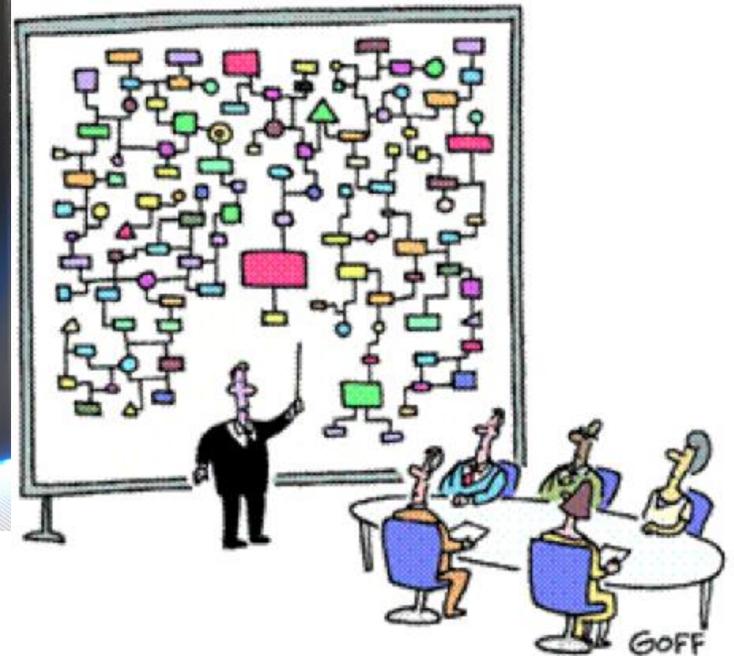
Big Data

Exons TP53 exons
Markup loaded

All exons in this region

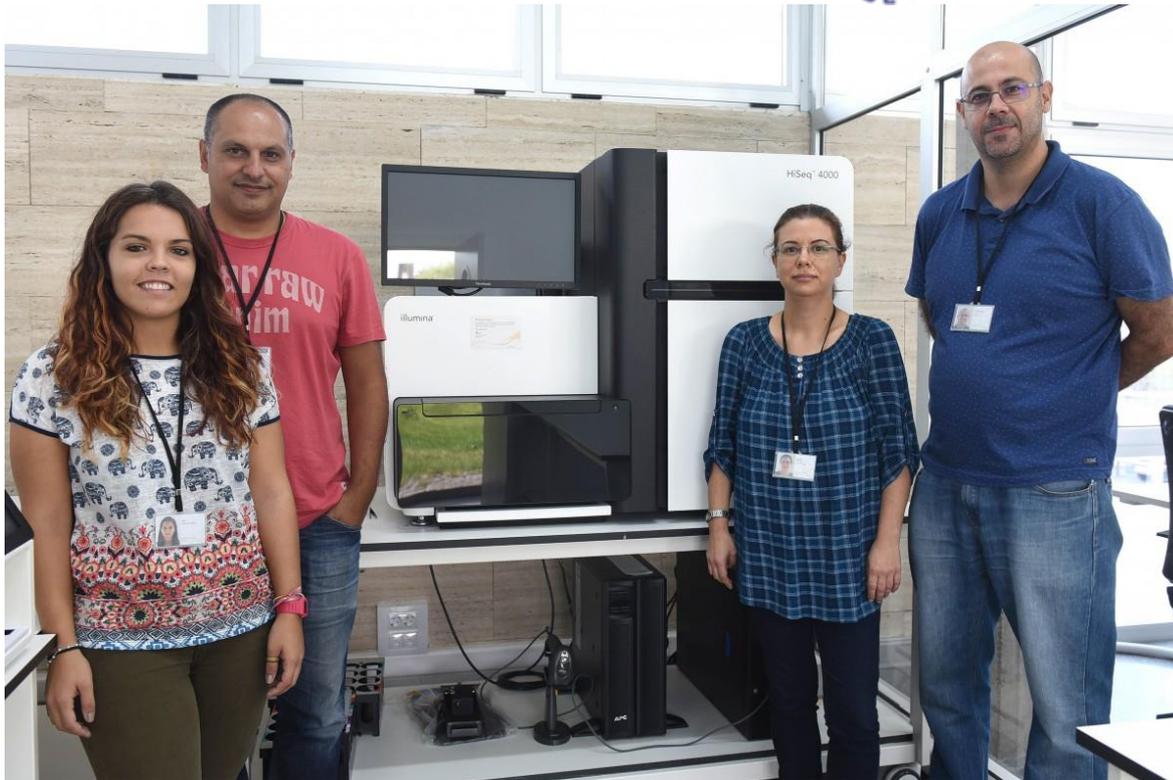
GCAGGAGAAATGGCTTGAACCCGGAAAGGTAGAGGTTGCAGTGCAGCTGAGATCATGCCACTG
 TCTCTGCAGGCCAGGTGACCCAGGGTTGGAAGTGTCTCATGCTGGATCCCCACTTTTCC
 TCTTGCAGCAGCCAGACTGCCTTCCGGTCACTGCCATGGAGGAGCCGAGTGCAGATCCT
 AGCGTCGAGCCCTCTGAGTCAGGAAACATTTTCAGACCTATGGAACCTGTGAGTGGAT
 CCATTGGAAGGCCCTGAGTCCACCACCCCAACCCCAAGCCCTAGCAGAGACTGT
 AACGTTCTGTAAGGACAAGGGTGGGCTGGGACCTGGAGGGTGGGACTGCTGCTTCT
 TGGGGGCTGGGGCTGAGGACTGGTCTGACTGCTGACTGCTGCTTTCACCCATCTACAGT
 CCCCTTGAAGACCAAGATGATTTGATGCTTCCAGAAATGCCAGAGGCTGTGAAACAATG
 GTTCACTGACACCCAGTCCAGATGAAGCTCCCGGGCCCTGCACAGGCTGCTCCCGCT
 GCGCCCTGCACAGCTCCTACACCCGGCCCTGCACAGGCTGCTCCCGCTGGCCCT
 CTCATCTTCTGCTCCCTCCAGAAAGCTGACAGGCTCAGTGCAGTGCAGTGCAGTGCAGT
 TTCCATGAGACTTCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
 GATTCCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
 ATCCCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
 ATGGTGGTGTAAACATGATGAAACCTCGTCTACAAAAAATACAAAAAATAGCCAGG
 GGCCAGGAGATGGAGGCTGCAAGTGTGATCACACCACTGTGCTCCAGCCTGAGGTGAGAGC
 CAGAGCAAGACCTATAGTCCAGCCTAGGAGGCTGAGGTTGCTCCAGCCTGAGGTGAGAGC
 CCAACTCTCTAGCTATCTCAAAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG
 TTGCCCTTCCAGTGTCTTATCTGTTCACTTGTGCCCCTGACTTCAACTGTCTGTCTTCTT
 CCCCTGTGCAGTGTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
 GCCCTGTGCAGTGTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
 CCATCTACAAGCAGTCACAGCACATGACCCCGCCCGCCGCTGAGGCGCTGCCCCCACCATGAGC
 GCTGCTCAGATAGCGATGTGAGCAGCTGGGCTGGAGACGACAGGGCTGGTTGCCCA
 GGTCCCCAGCCTCTGATTCTCACTGATTGCTTGTAGGTCTGGCCCTCCTCAGCATC

El superordenador teideHPC



"And that's why we need a computer."

¿Quiénes lo hacemos?



Consejería de Educación y Universidades



HOSPITAL UNIVERSITARIO NUESTRA SEÑORA DE CANDELARIA



MUCHAS GRACIAS!!!

