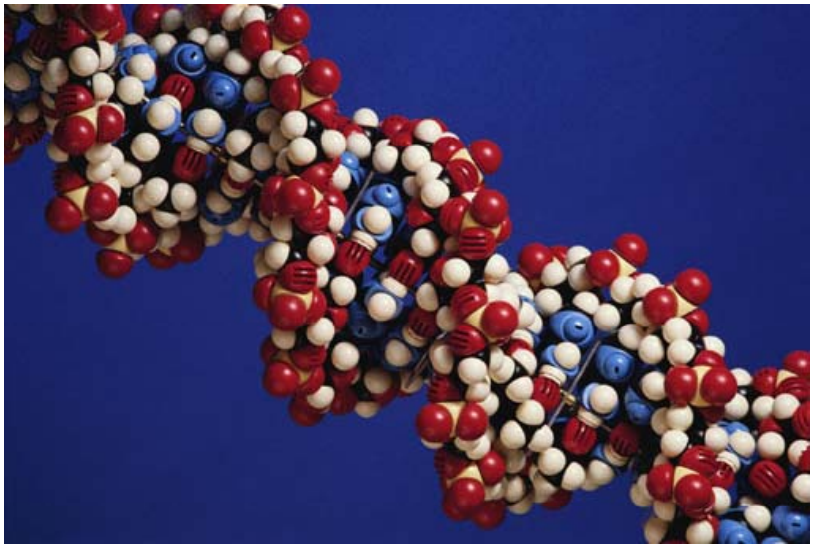
La importancia de la bioinformática en el avance de la biomedicina

Fuencisla Matesanz

Instituto de Parasitología y Biomedicina López Neyra, CSIC

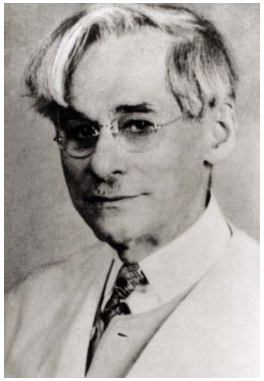
El genoma

El genoma es la información genética que posee un organismo o una especie en particular. El soporte en el que se encuentra esta información es el ADN o ARN.



Historia del ADN

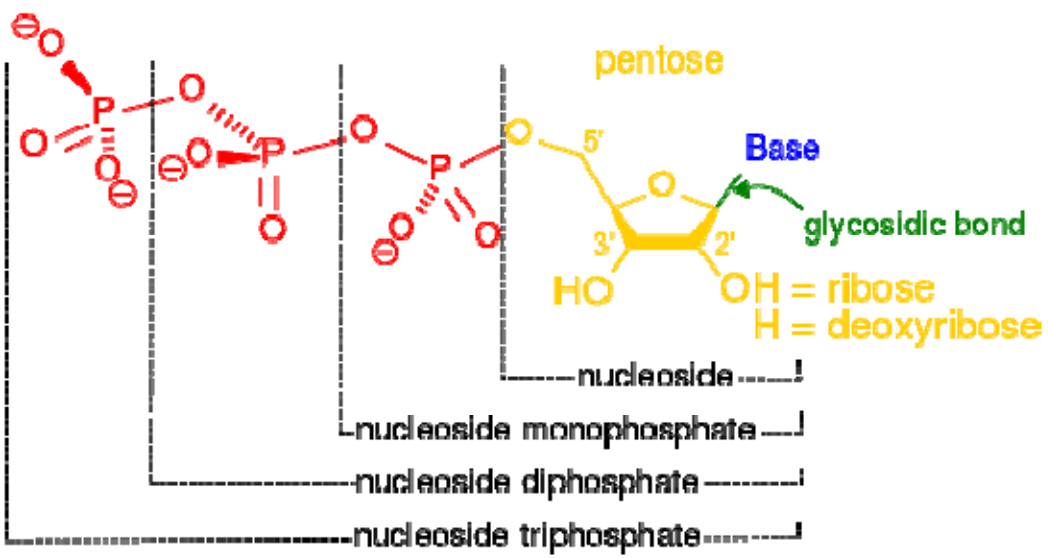
1869 Friedrich Miescher, médico suizo, descubrió que el ADN era el componente químico principal del núcleo de las células y lo denomino nucleina.

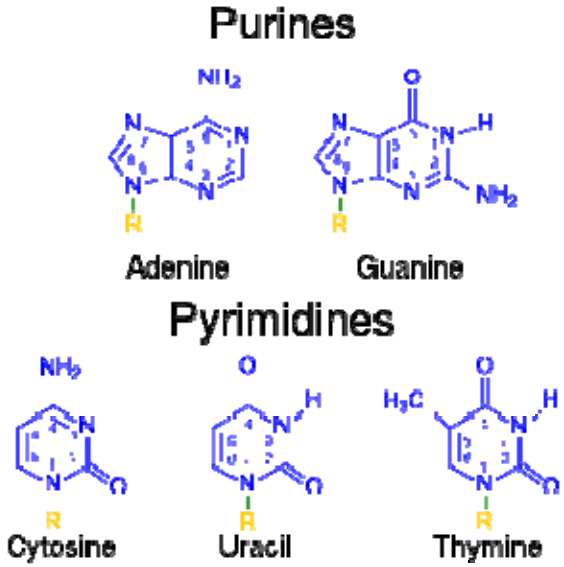


En 1919 P.A. Theodore Levene descubrió que las unidades constitutivas del ADN son los núcleotidos y que estos están formados por desoxiribosa, una de las 4 bases nitrogenadas (adenina, timina, citosina y guanina) y un grupo fosfato.

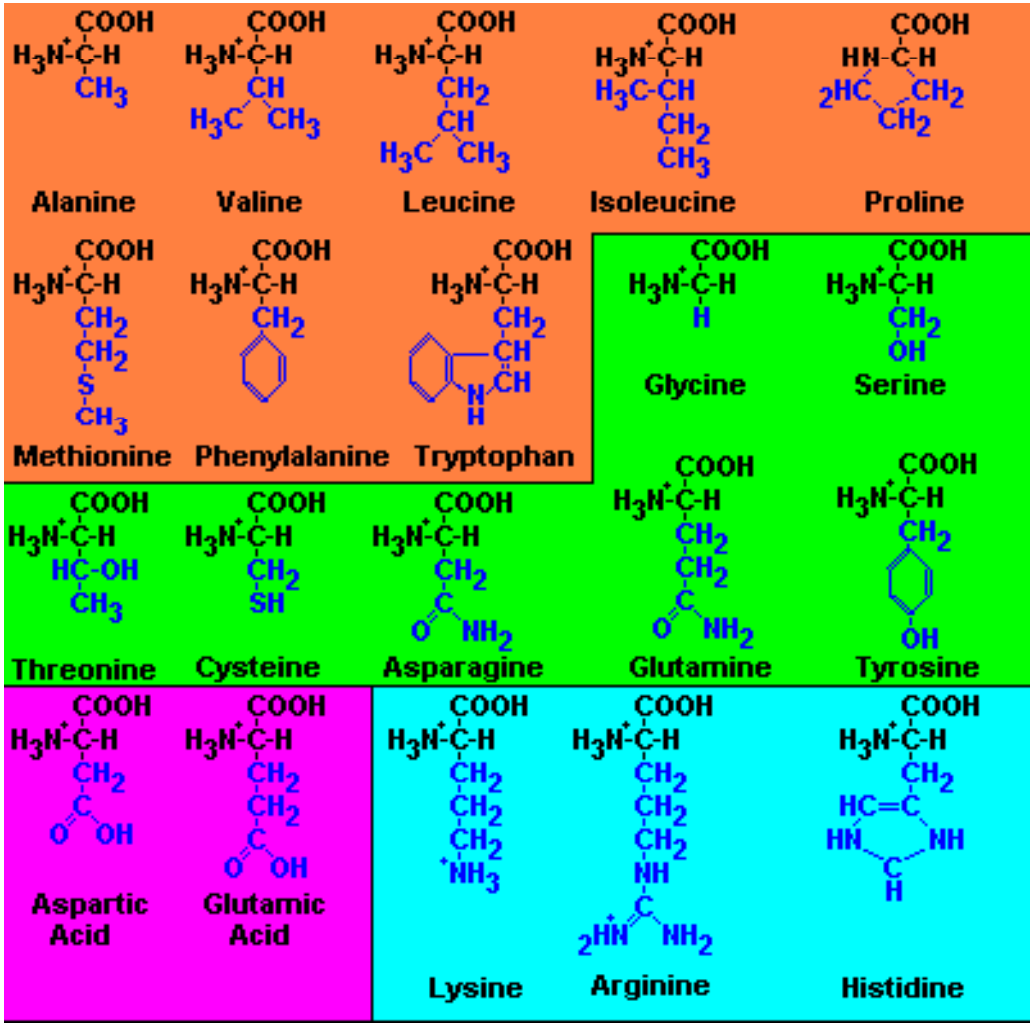
Levene consideró que el ADN es incapaz de contener información genética dada la poca complejidad de sus elementos y consideró las proteínas con estructura mucho más compleja y variada como las moléculas clave de la herencia.

Nucleótidos



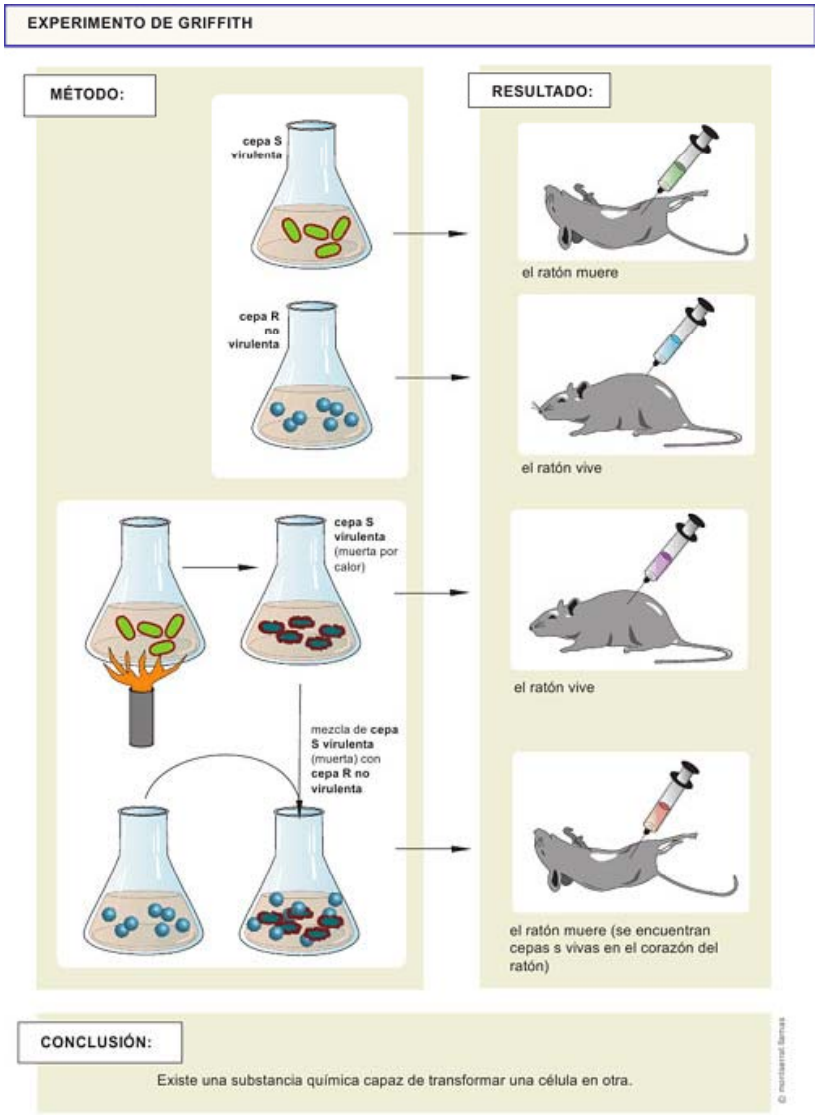


Aminoácidos

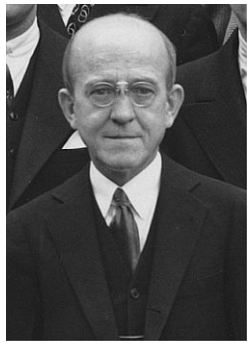


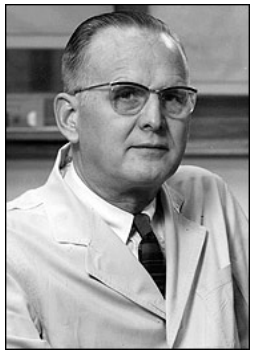
Principio de transformación

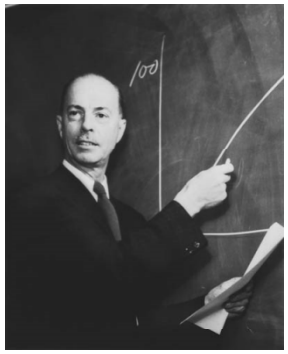
1928 el microbiólogo Fred Griffith descubre que un elemento de neumococos podía transmitir características de una cepa virulenta a otra no virulenta y lo denominó "principio de transformación"



ADN es la molécula de la herencia

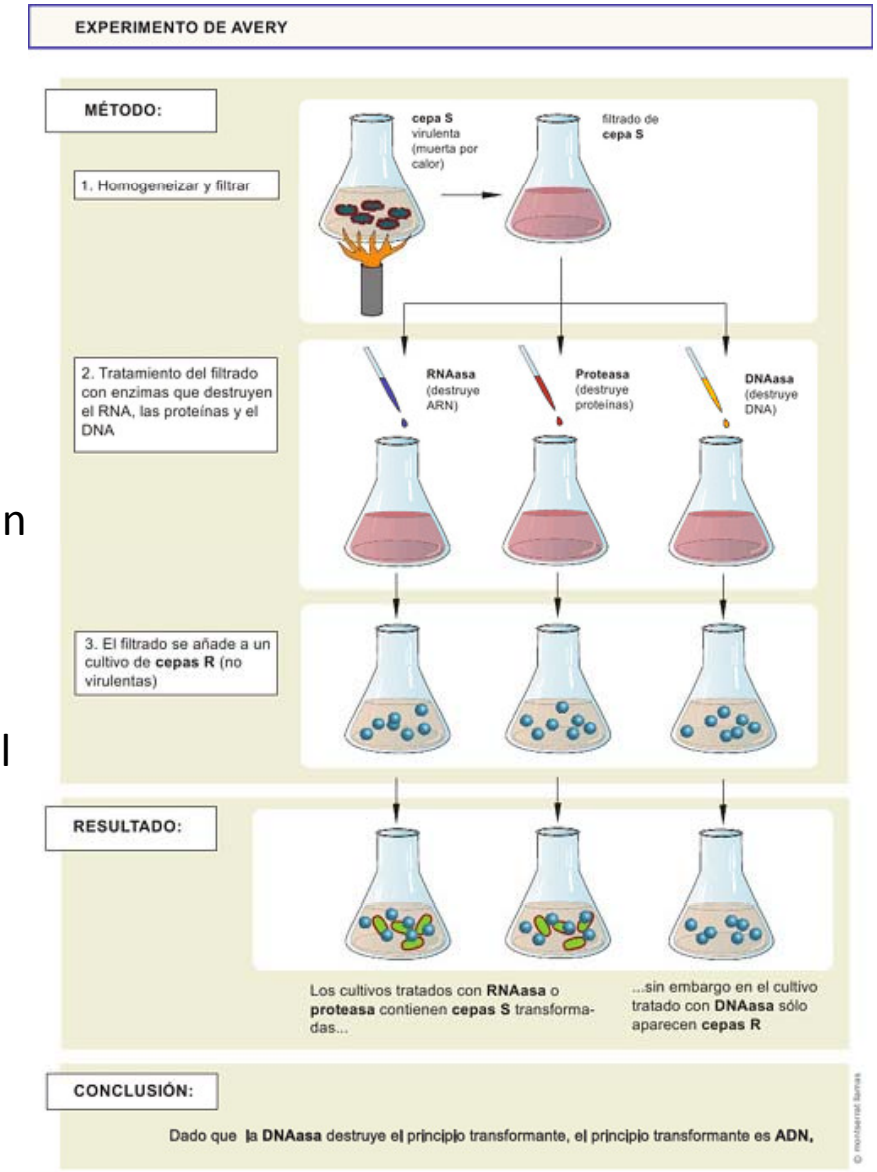




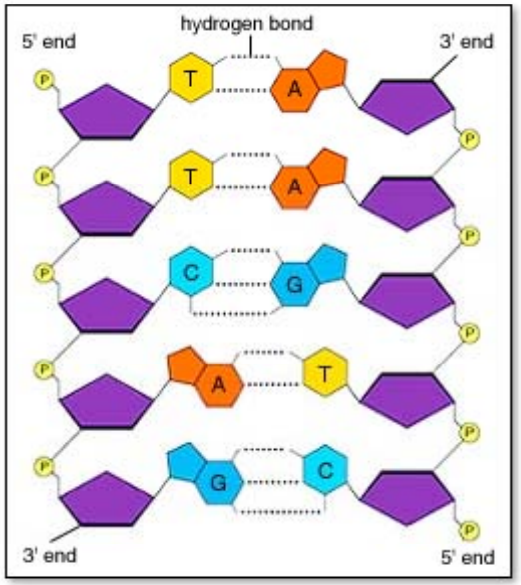


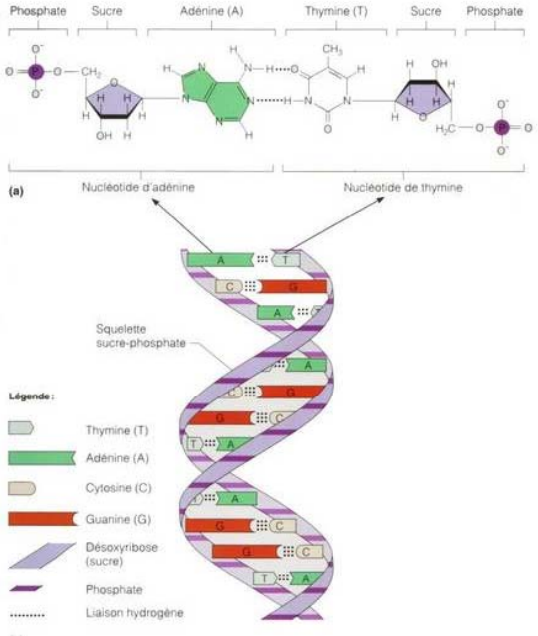
En 1944 Avery, Mcarthy y McLeod, continuaron el estudio con neumococos y determinaron, mediante la eliminación selectiva de todas las sustancias contenidas en la mezcla, que el principio transformante era necesariamente el ADN.

ADN es la molécula de la herencia



En 1953 James Watson y Francis Crick propusieron la estructurad del ADN





James Watson y Francis Crick

Rosalind Franklin

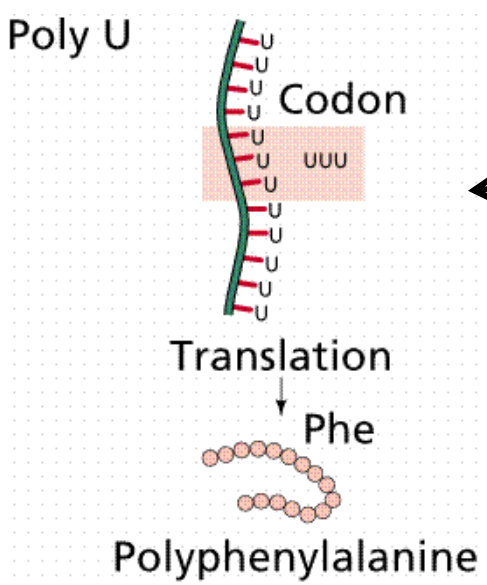
RNA Tie Club

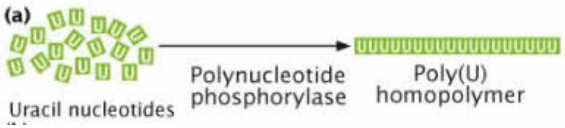
George Gamow crea en 1954 “RNA TIE Club” con el objeto de resolver el enigma de la estructura del ARN y para comprender cómo se construye las proteínas.

Postuló que un código de codones de tres bases debía ser el empleado por las células para codificar la secuencia aminoacídica, ya que tres es el número entero mínimo que con cuatro bases nitrogenadas distintas permiten más de 20 combinaciones (64 para ser exactos)

El primer codón en descifrarse UUU

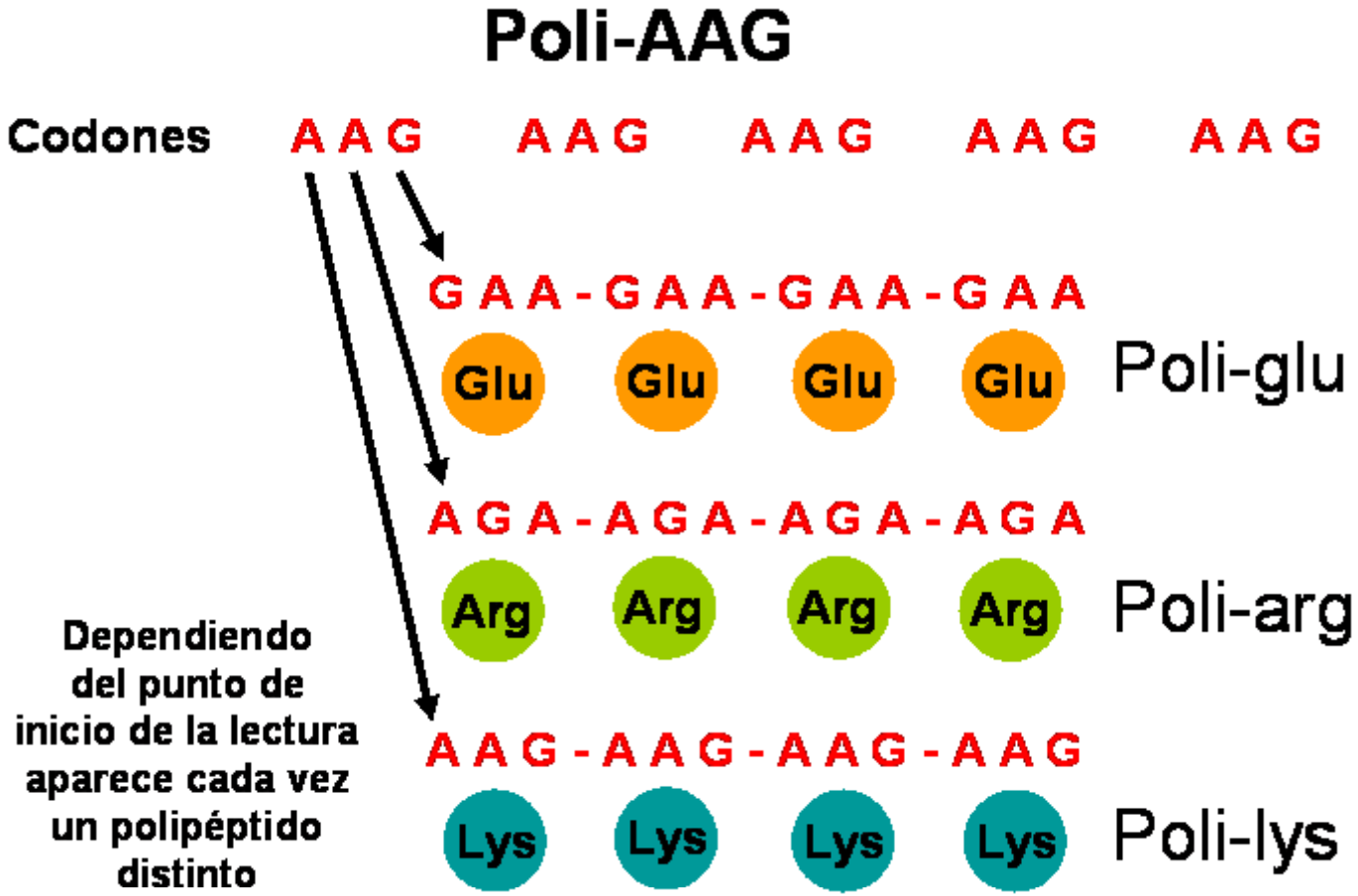
En 1961 Marshall W. Nirenberg, en una conferencia en Moscú, da a conocer el primer codón UUU que codifica la fenilalanina. Johann H. Matthaei y él crean un sistema de poly UUUUU que al traducirse da como resultado poly Phe-Phe.



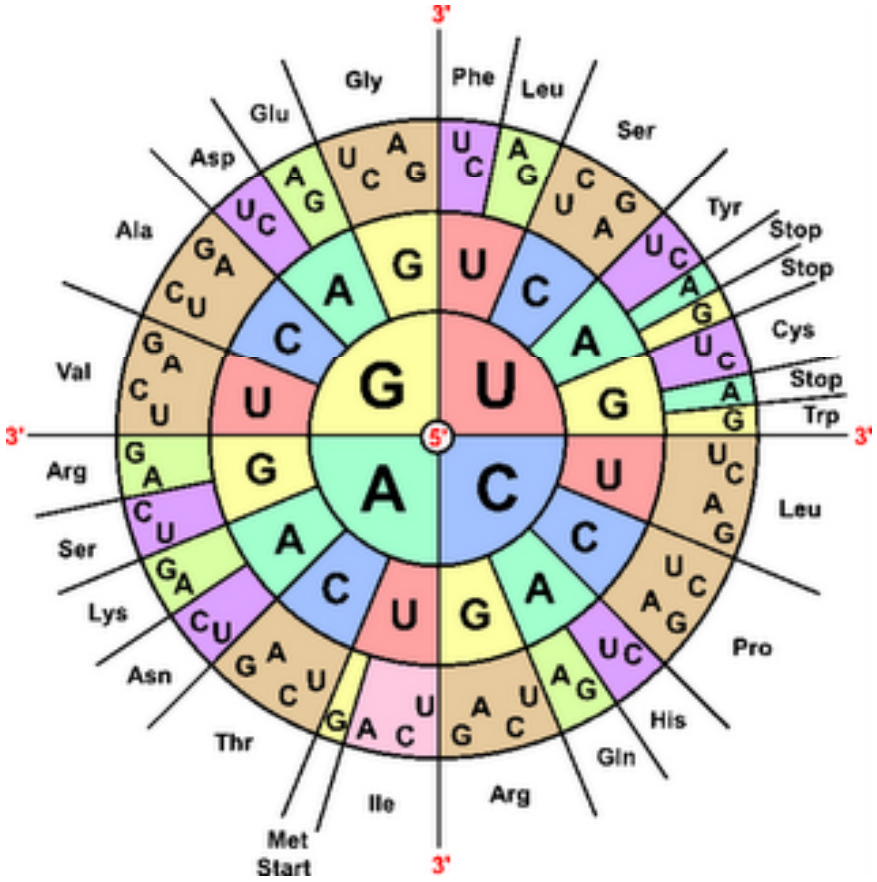


El descubrimiento y aislamiento de la enzima polinucleótido fosforilasa por parte del Investigador español Severo Ochoa fue determinante para poder descifra el código genético

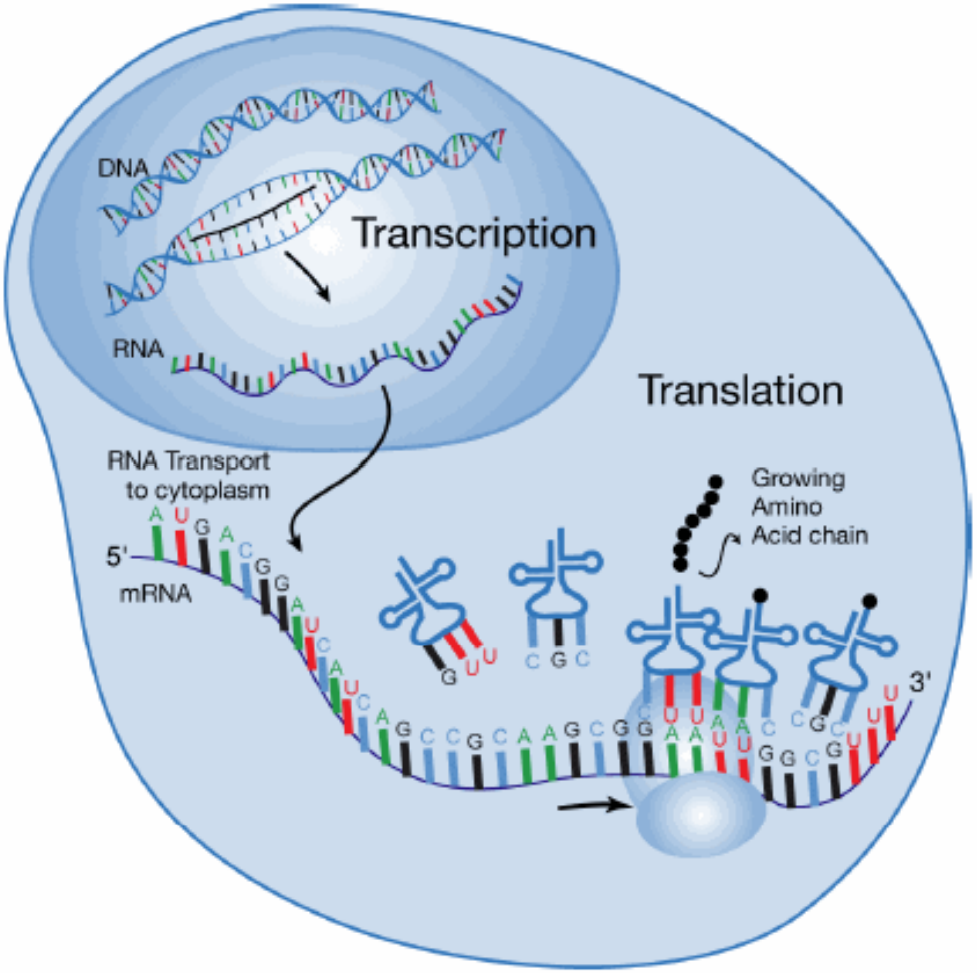
El sistema de secuencias repetidas permite descifrar el código genético



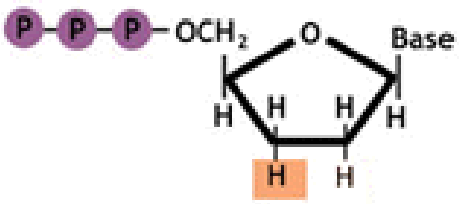
El código genético



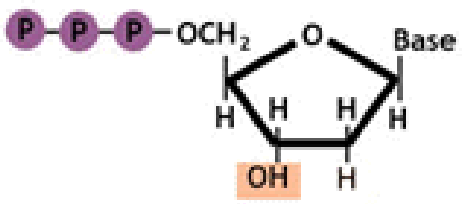
Transferencia de la información del DNA a la proteína



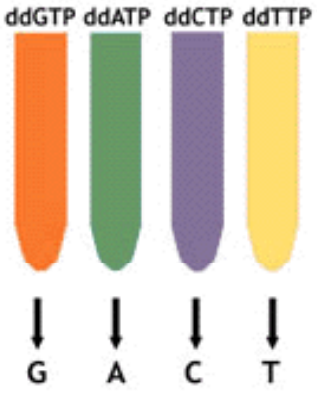
Sanger Method: Dideoxy Chain Termination



dideoxynucleotide (ddNTP)

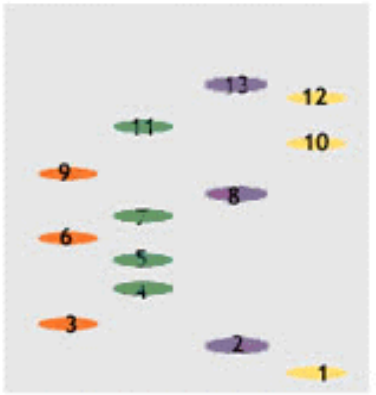


deoxynucleotide (dNTP)



Add ddGTP, ddATP, ddCTP, ddTTP, one to each of four tubes containing target DNA. Load each onto a separate lane on a gel.

300-500 bases



Largest

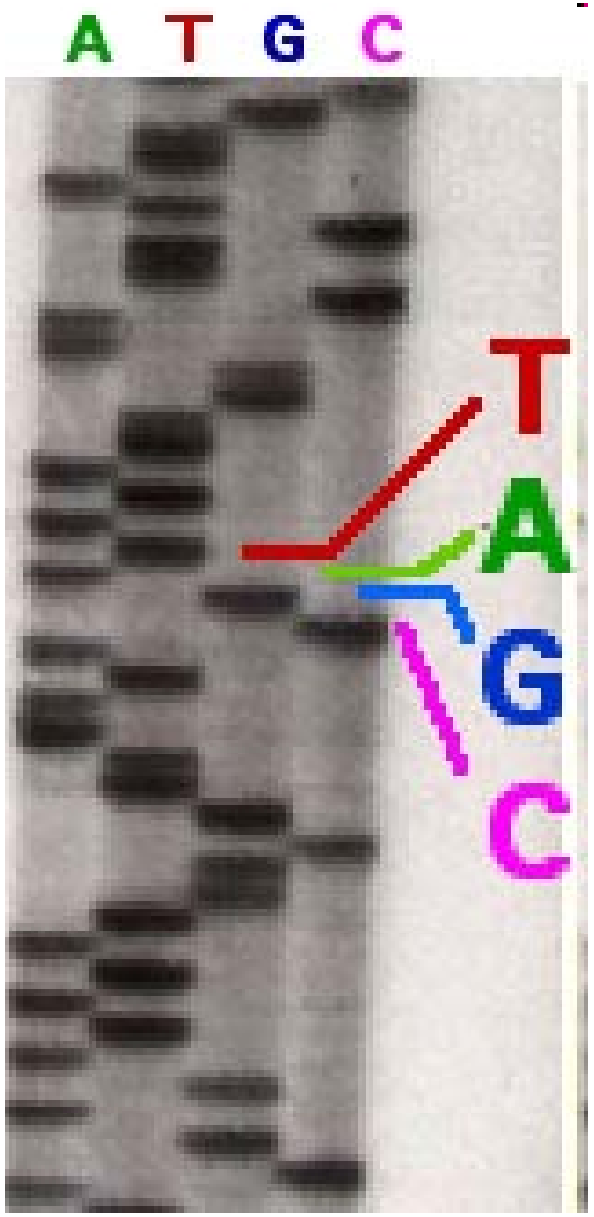
To sequence, read the order of bases from the smallest to the largest.

TCGAAGACGTATC

Smallest

[Video](#) 1

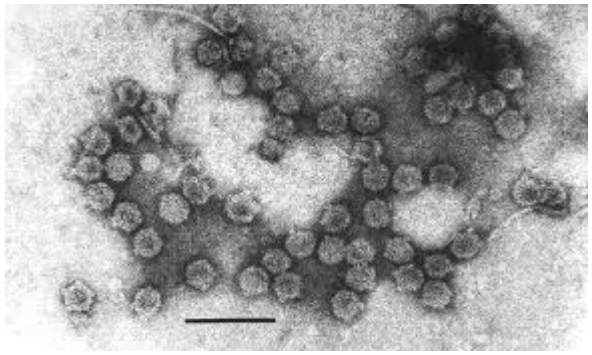
[Video2](#)



Secuenciación de ADN

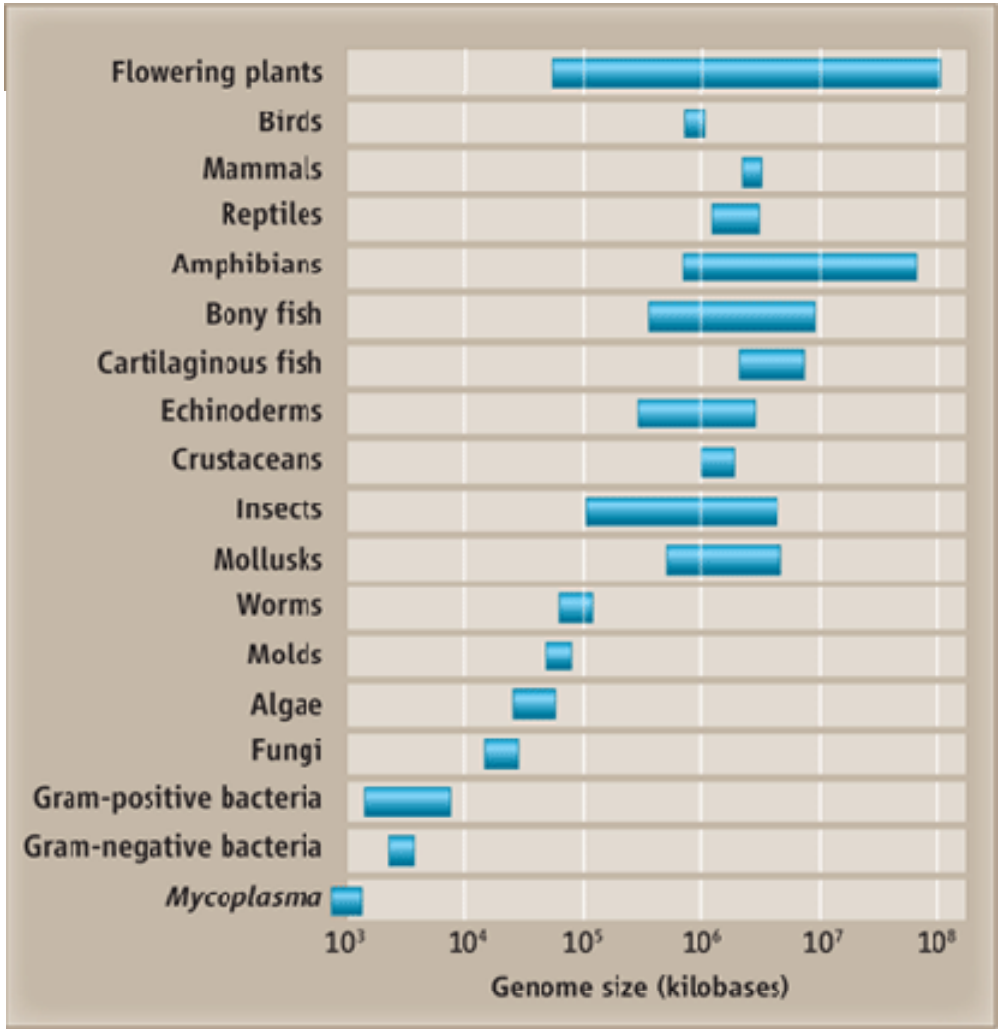
En 1977 Fred Sanger, Walter Gilbert, y Allan Maxam, secuencian ADN por primera vez trabajando independientemente. El laboratorio de Sanger completa la secuencia del genoma del bacteriófago Φ-X174

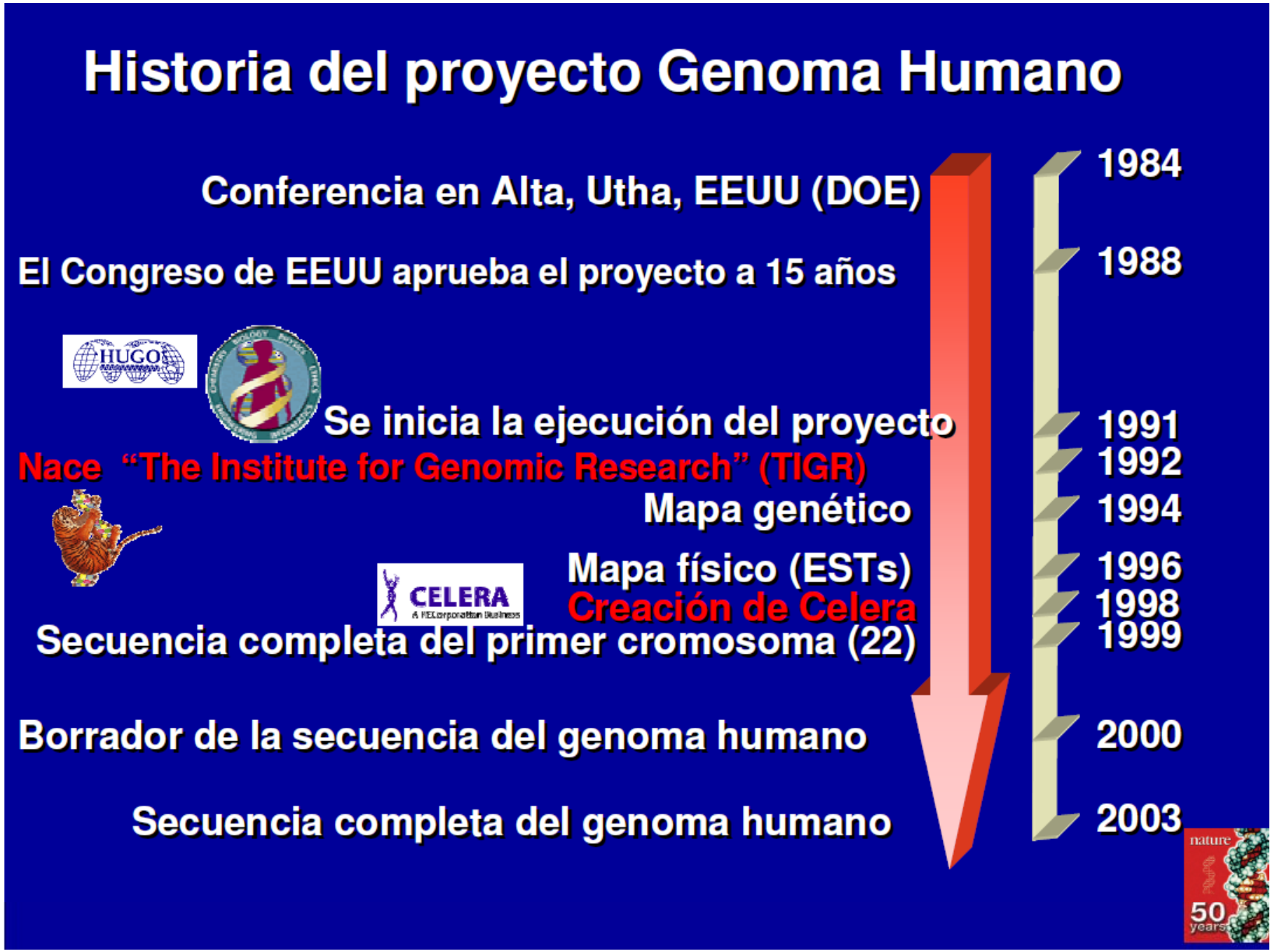
5386 nucleotidos



En 1989 Francis Collins y Lap-Chee Tsui secuencian un gen humano por primera vez. El gen codifica la proteína CFTR, cuyo defecto causa fibrosis quística

Tamaño de los genomas



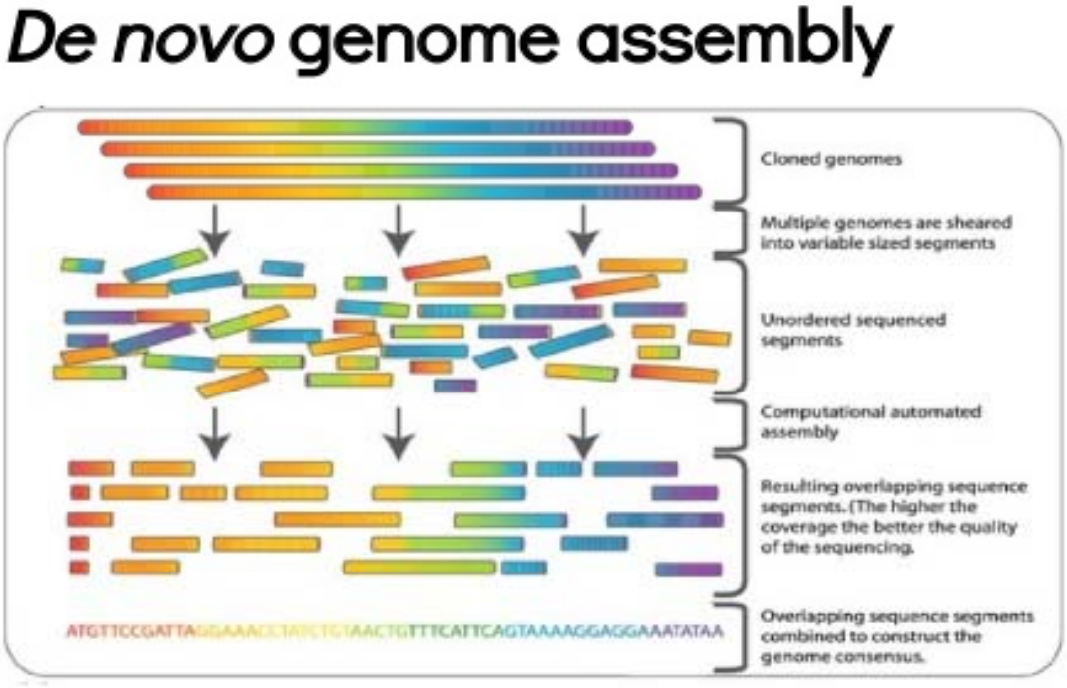


Video

Presupuesto 3.000.000.000$/ Costó 2.700.000.000 $

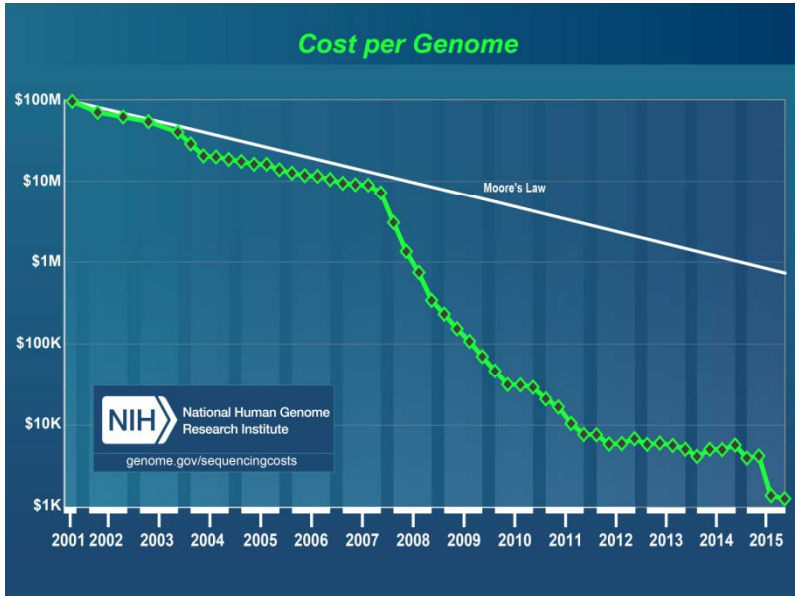
Secuenciación del genoma humano

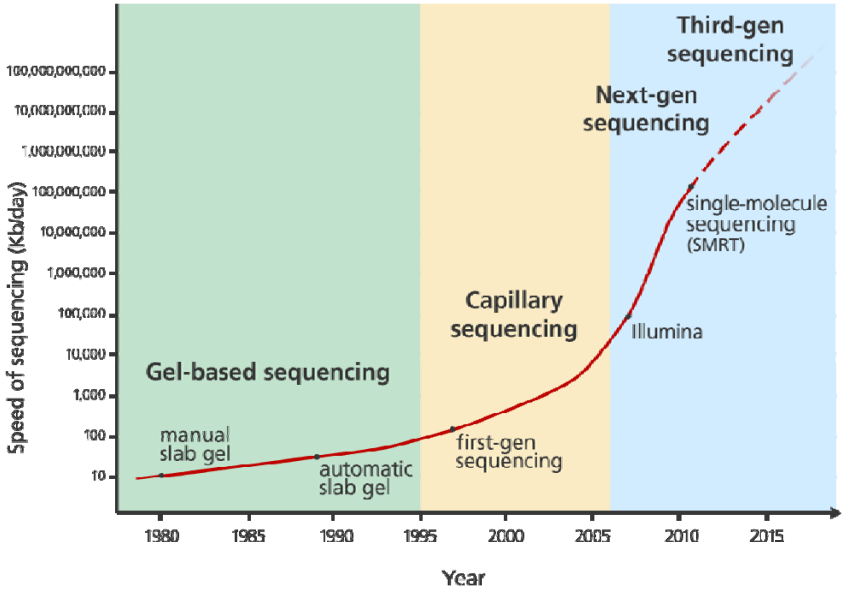
Ensamblaje del genoma



Bases de datos y sistemas de análisis

Evolución del costo de secuenciación de un genoma





Base de NGS

Read Name: SRR003094.10299552
Position: 37613635
Read Length: 51
Mate Position: *: 0
Select/Deselect
Add to Bookmarks

Variaciones de la secuencia del genoma humano

Los genomas de dos personas son 99,9 % idénticos

Las variaciones en la secuencia de nucleótidos originan cambios fenotípicos que determinan:

- Variaciones fenotípicas normales.

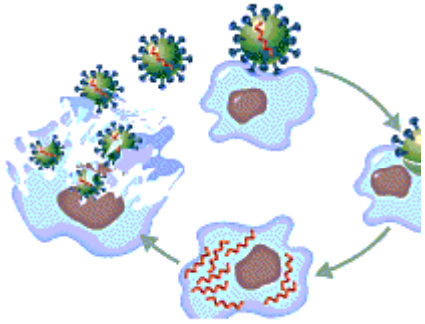


Ilustración de Microsoft

- El riesgo de padecer enfermedades

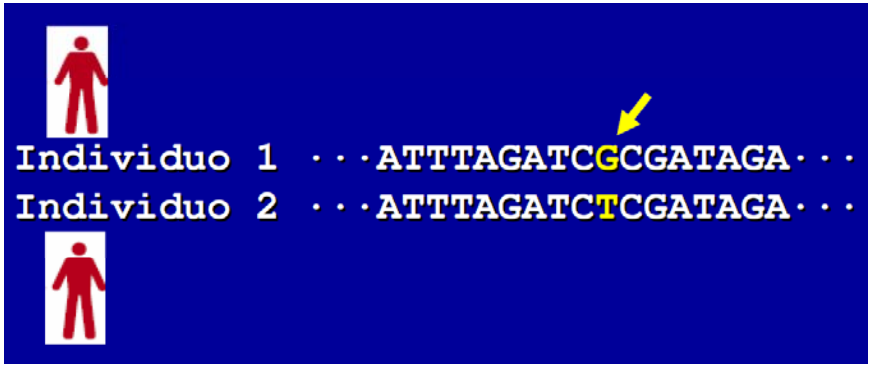
- La respuesta a factores ambientales como fármacos

VARIANTES ENTRE GENOMAS (POLIMORFISMOS)

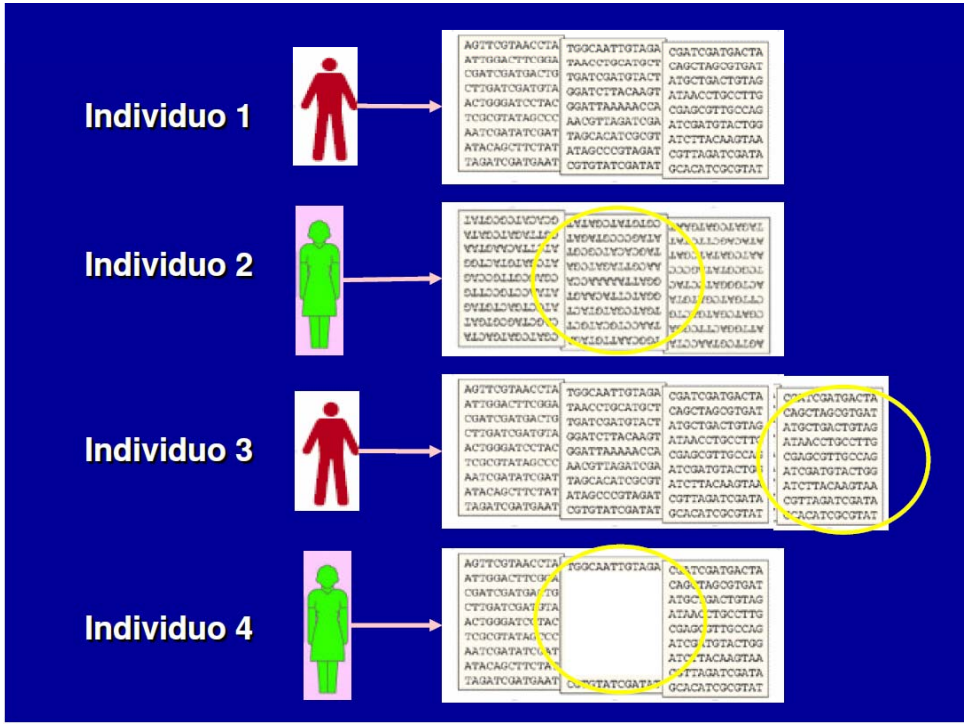
- ENTRE DOS INDIVIDUOS: 0,1%
- 3.000.000.000 pares de base genoma haploide
- 1bp : 1.000bp = 3 MILLONES
- EN LA ESPECIE HUMANA > 10 MILLONES
- TIPOS:
  - **SNP** (**S**ingle **N**ucleotide **P**olymorphism)
  - **CNV** (**C**opy **N**umber **V**ariation)
    - **inserciones, duplicaciones, deleciones**
    - **inversiones, translocaciones**

Estas diferencias entre genomas pueden ser de una única base
SNP (Single-nucleotide polymorphism) o fragmentos grandes del genoma
CNV (Copy number variation)

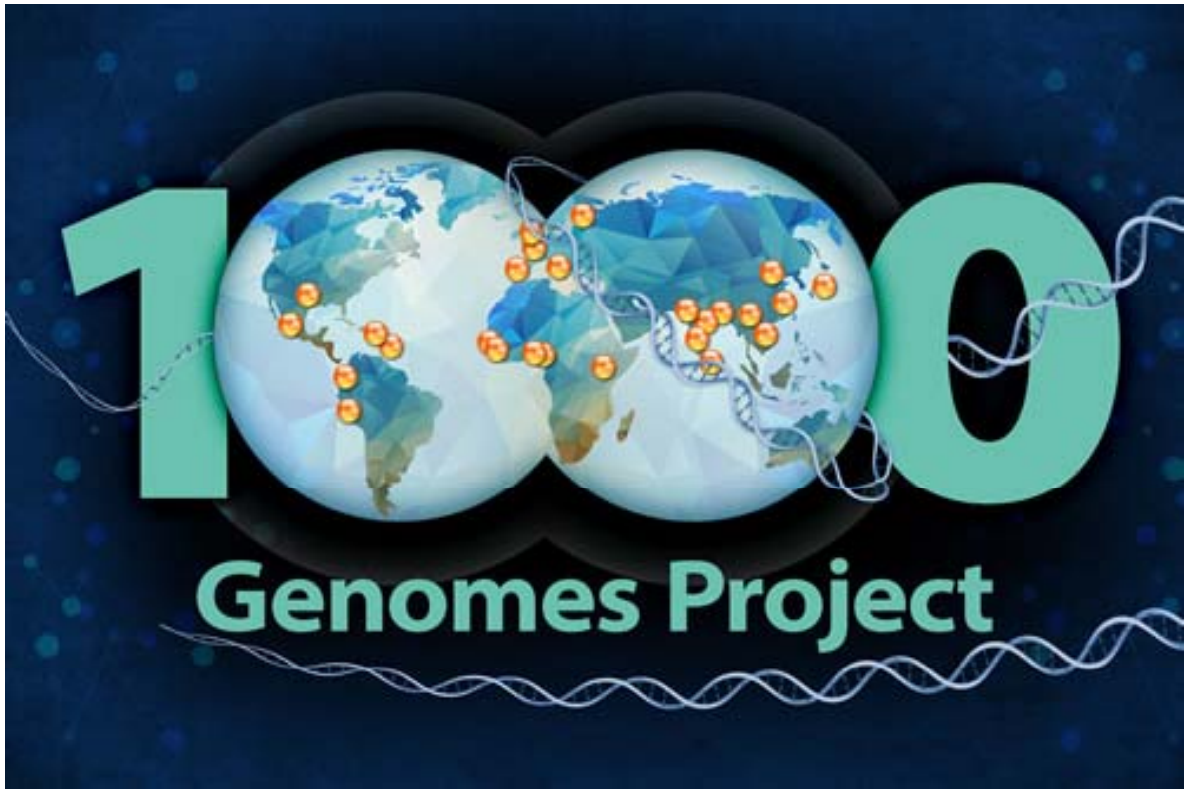
SNP



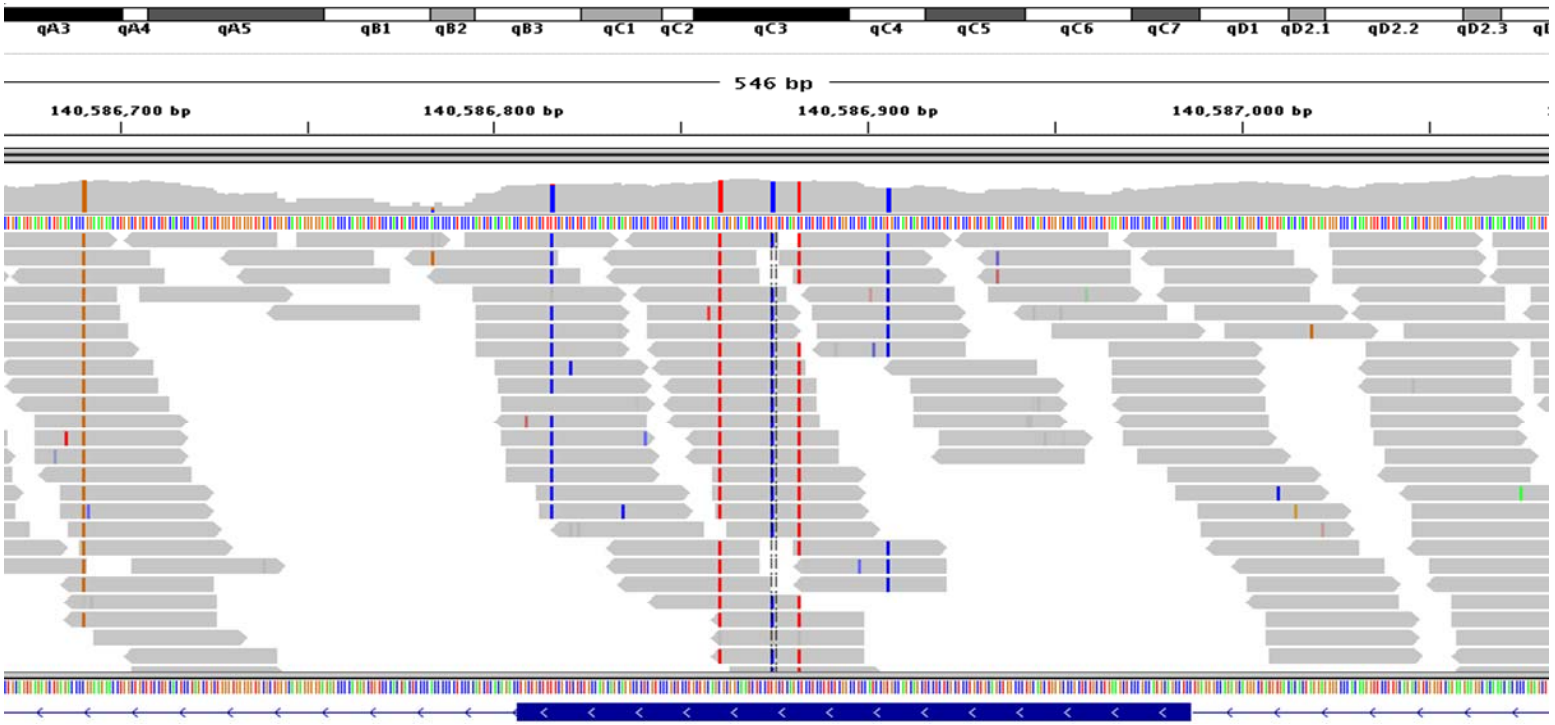
CNV



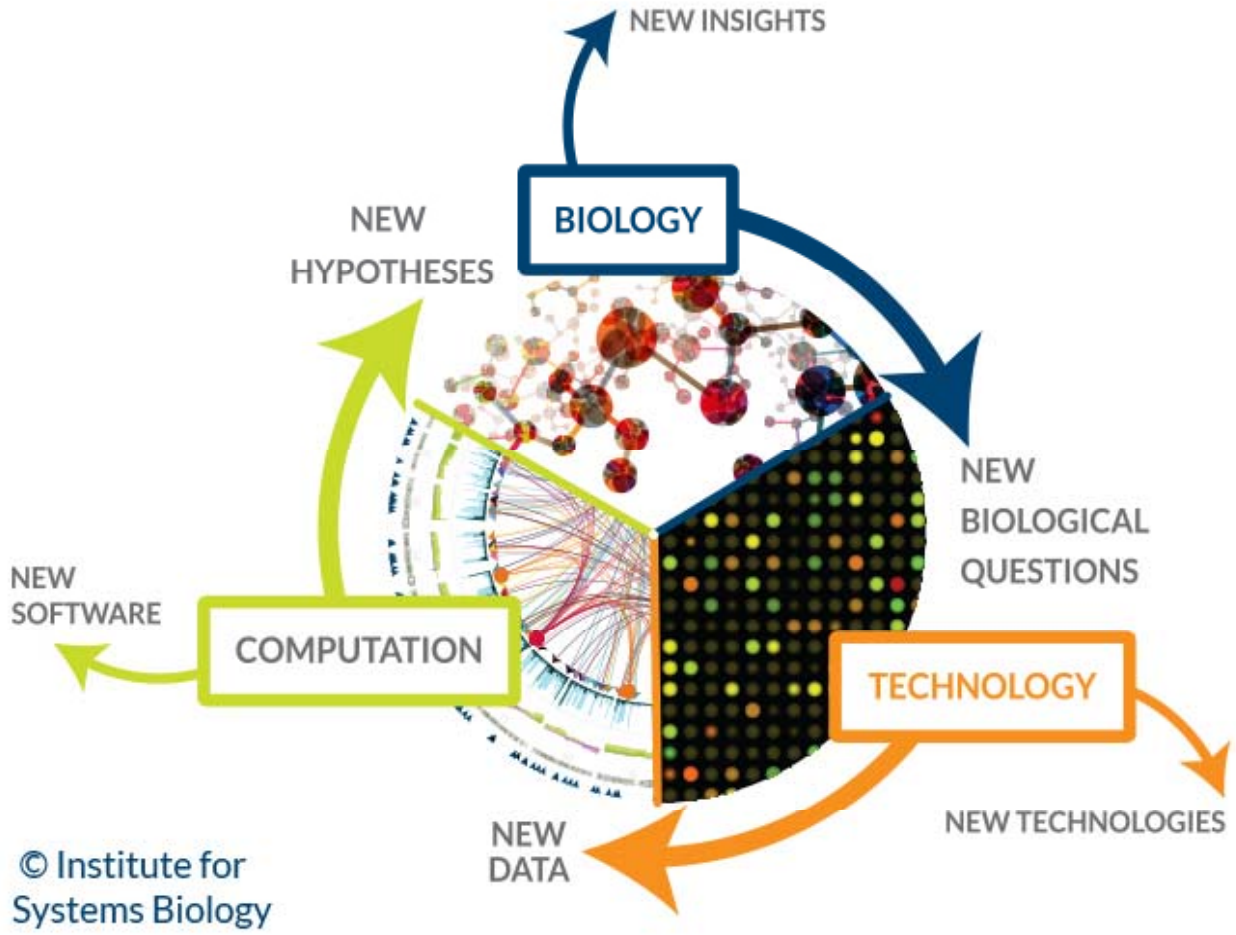
Determinación de variabilidad humana



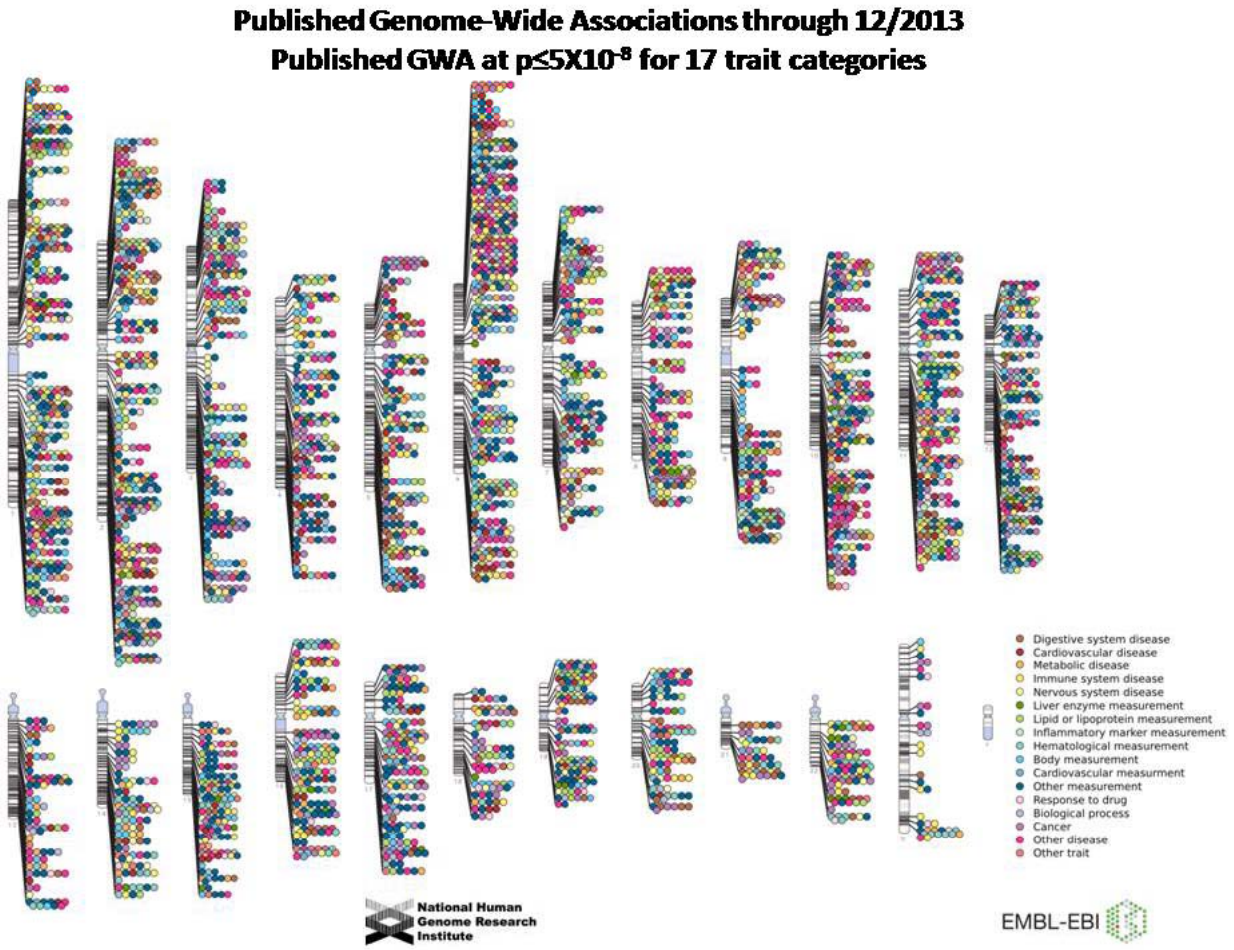
Localización de diferencias en las secuencias



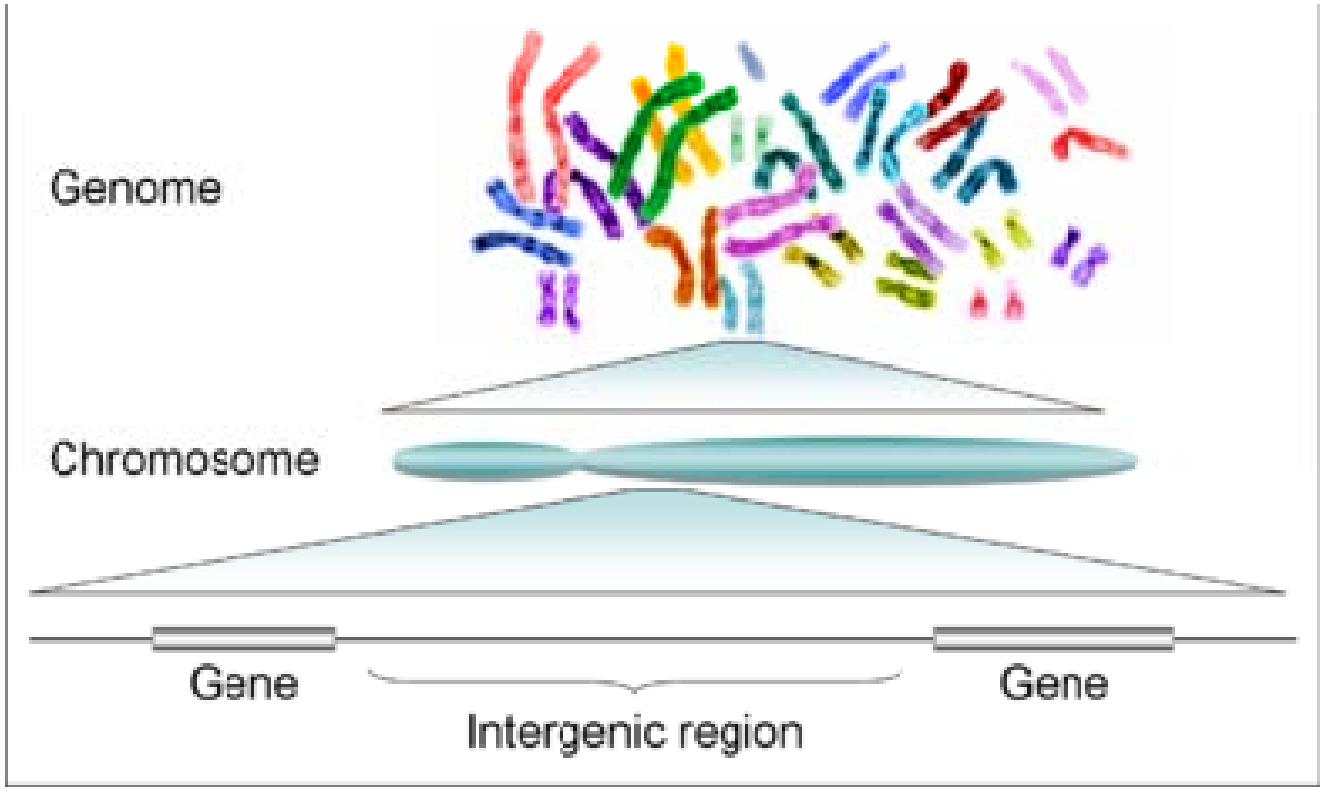
Avances en conocimientos en Biología



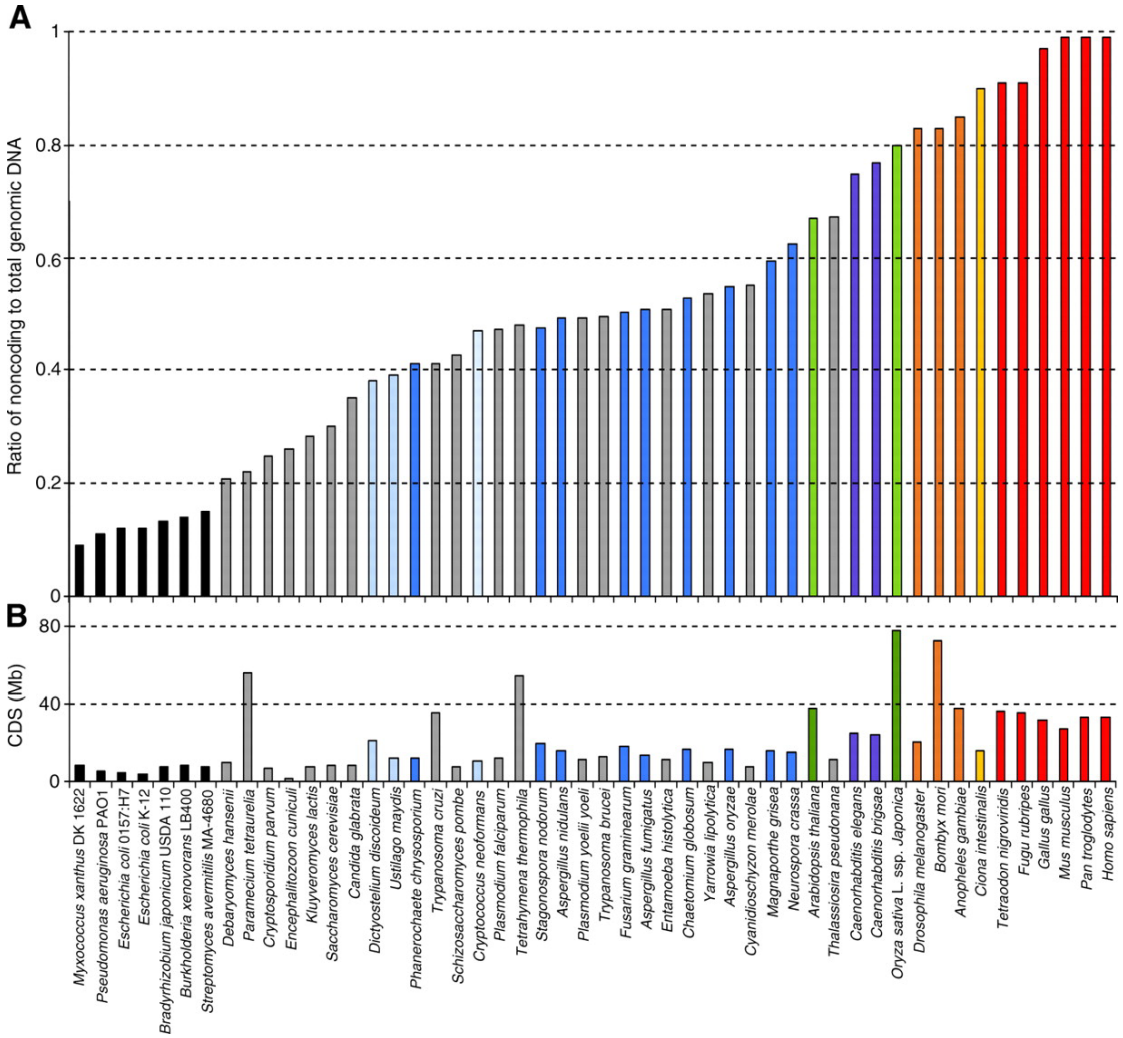
Regiones del genoma que se asocian con enfermedades



La mayor parte del genoma no tiene genes



The fraction of non-protein-coding DNA and megabases of protein coding sequence (CDS) per haploid genome in different species.

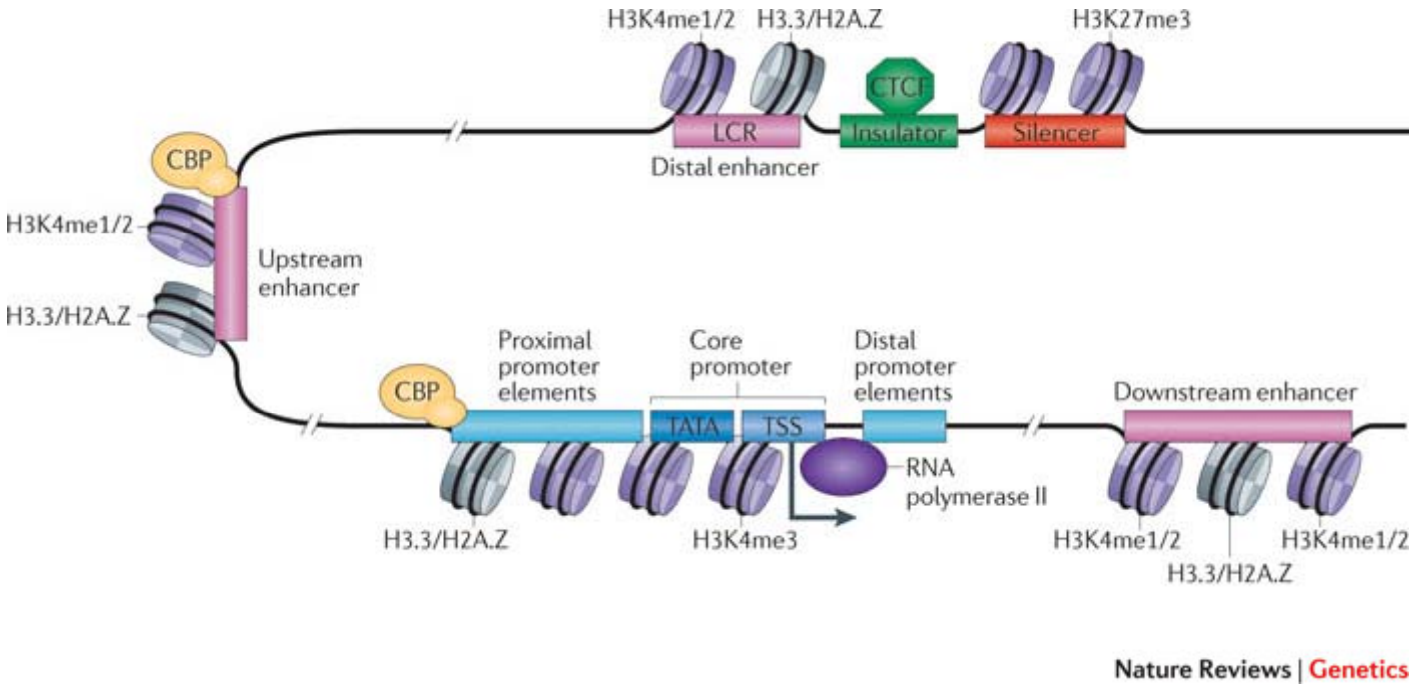


Mattick J S J Exp Biol 2007;210:1526-1547

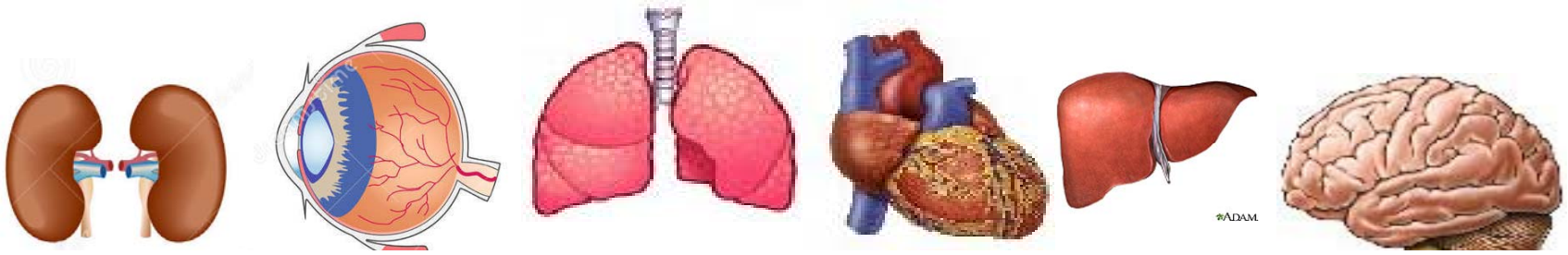
The Journal of Experimental Biology

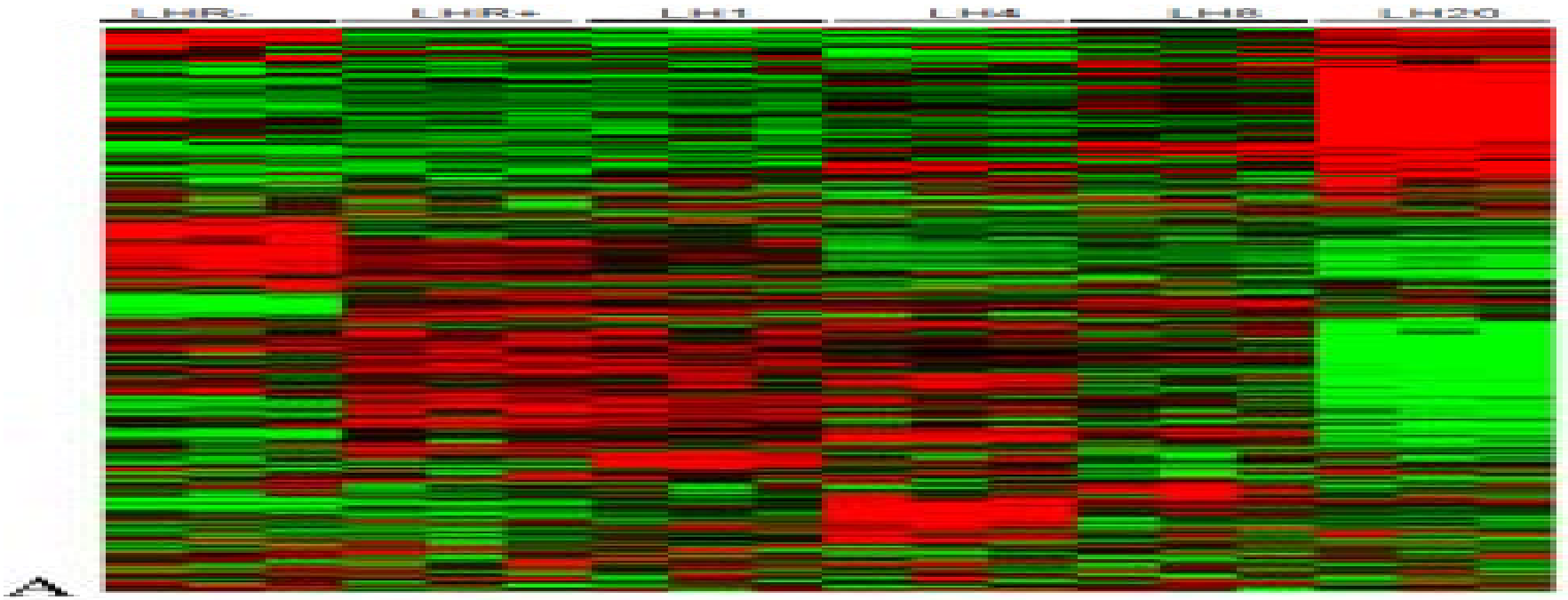
©2007 by The Company of Biologists Ltd

Secuencias reguladoras de la transcripción



Expresión diferencial de tejidos





Los genes se expesan de forma diferencial:

Durante el desarrollo embrionario

En las células de distintos tejidos

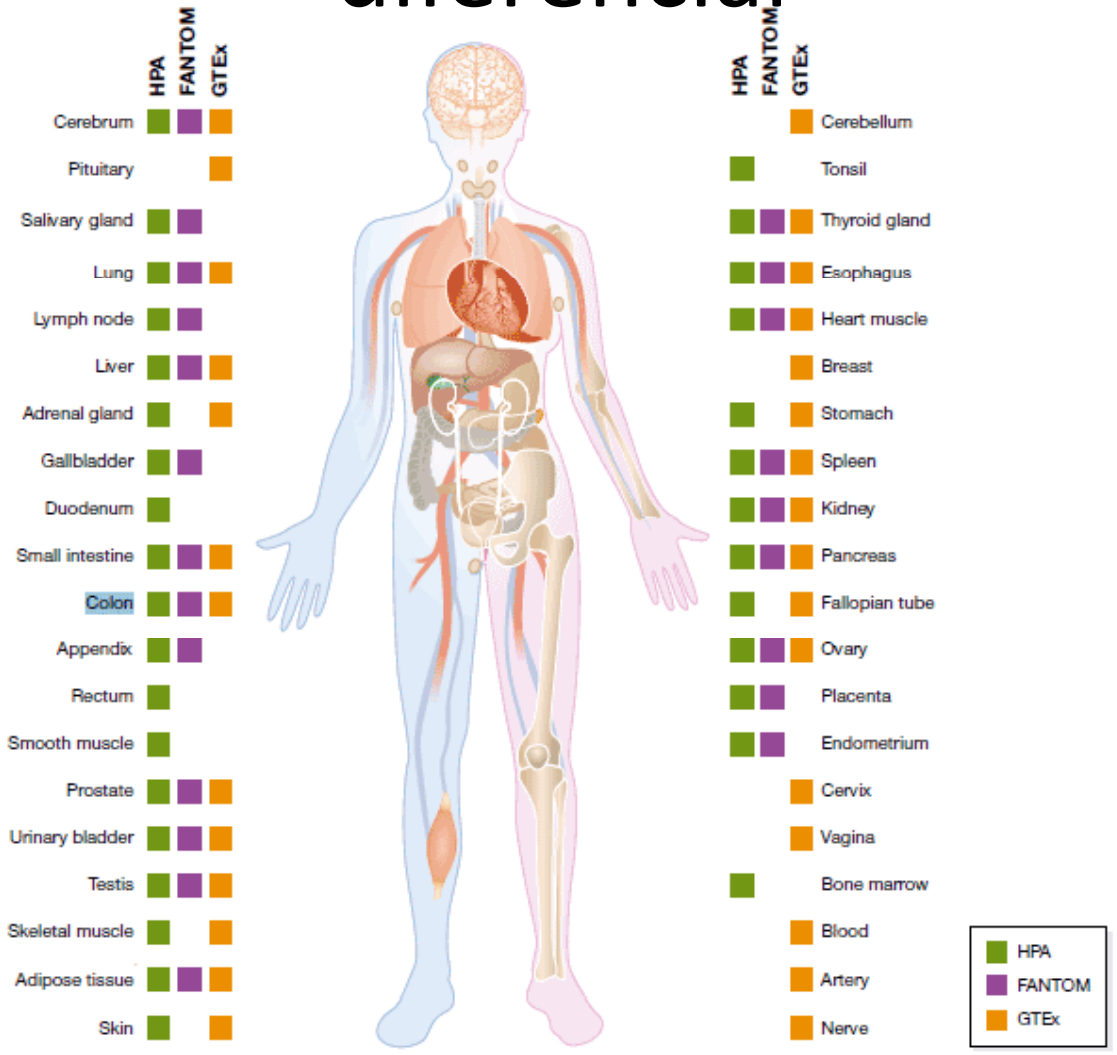
En mometos de estres

En situaciones patológicas

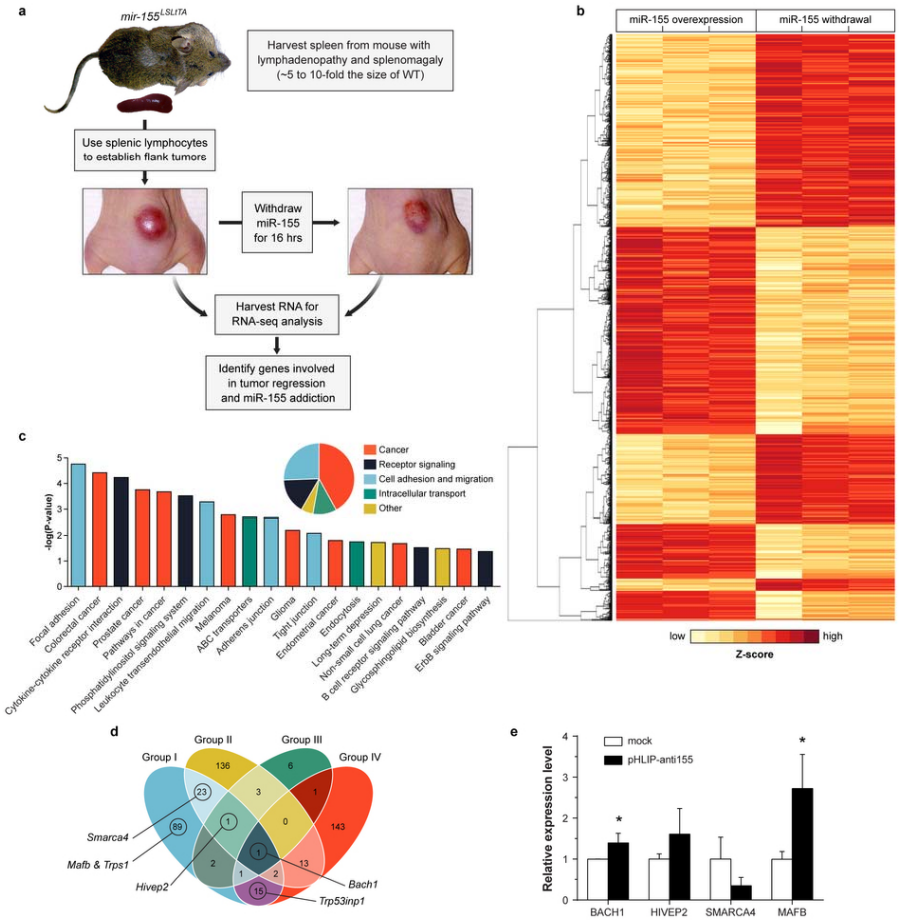
Fases repreoductoras

Para adaptarse al medio

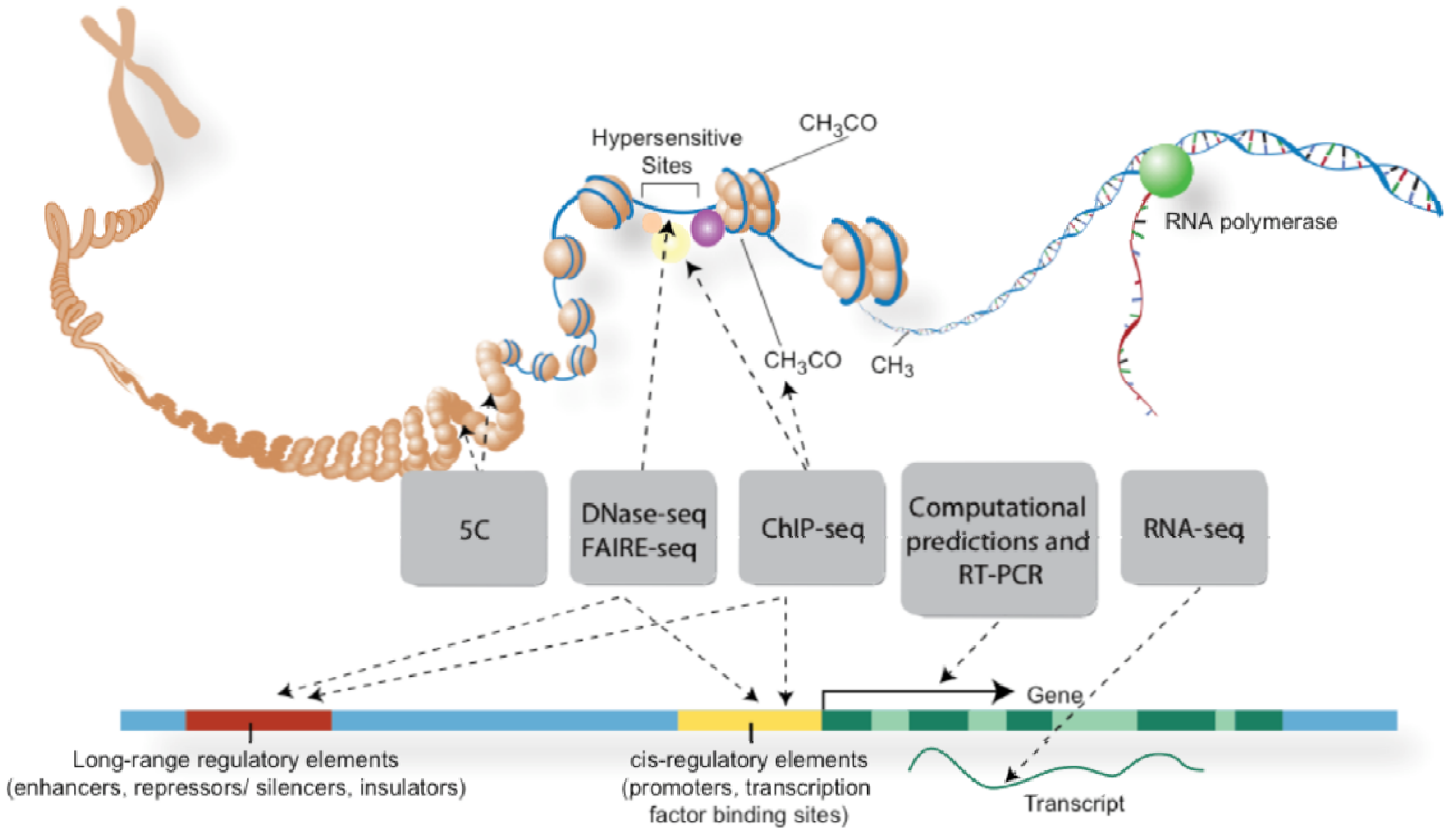
Proyectos que estudian la expresión diferencial



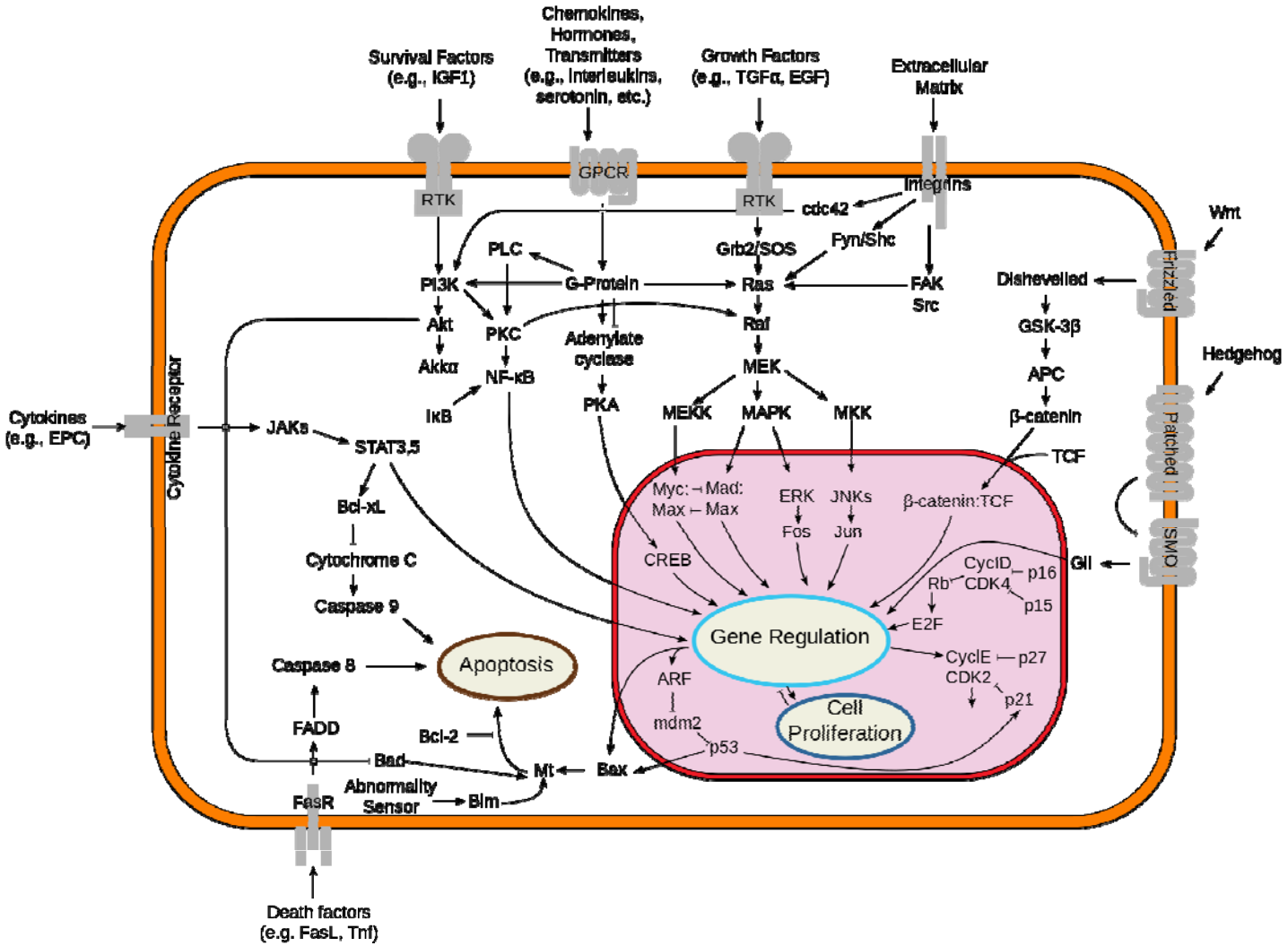
Expresión diferencial entre diferentes condiciones



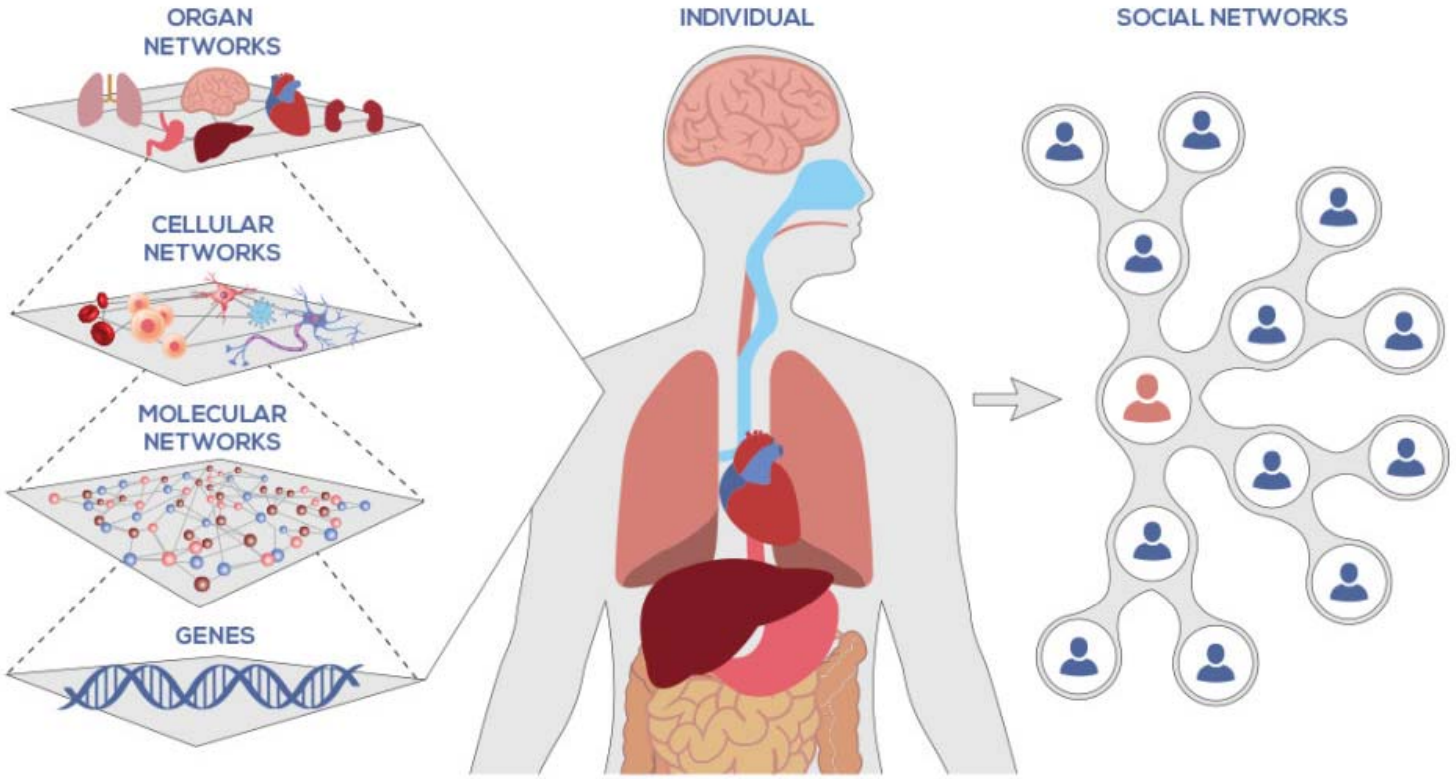
Proyecto ENCODE



Procesos en las células están relacionados



La biología de sistemas es el modelado computacional y matemática de los sistemas biológicos complejos



Proyectos internacionales

Geuvadis Project

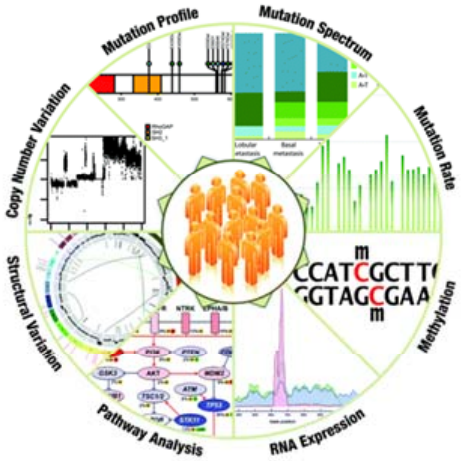
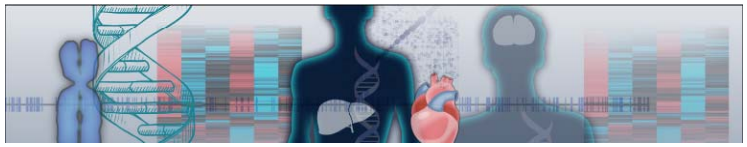


Image extracted from :
Hum Mol Genet. 2010 Oct 15;19(R2):R188-96

UK 10K

TwinsUK

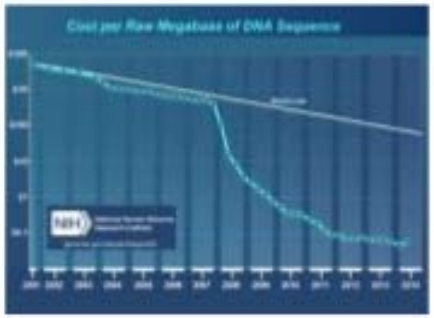


GTEx

ALSPAC. Seguimiento de 14000 niños nacidos entre 1991-1992

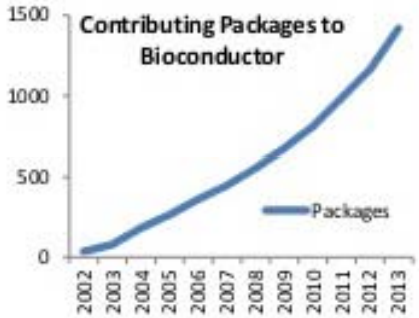
Eventos que están permitiendo la evolución de la biomedicina

Decreasing Cost of NGS Data Generation



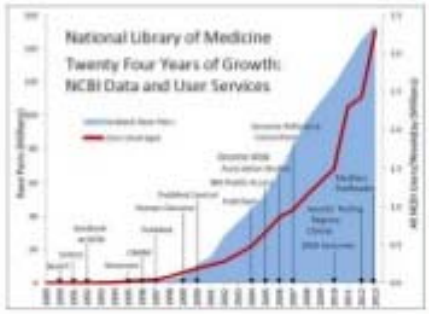
Source: NHGRI

Increasing analytical tool development



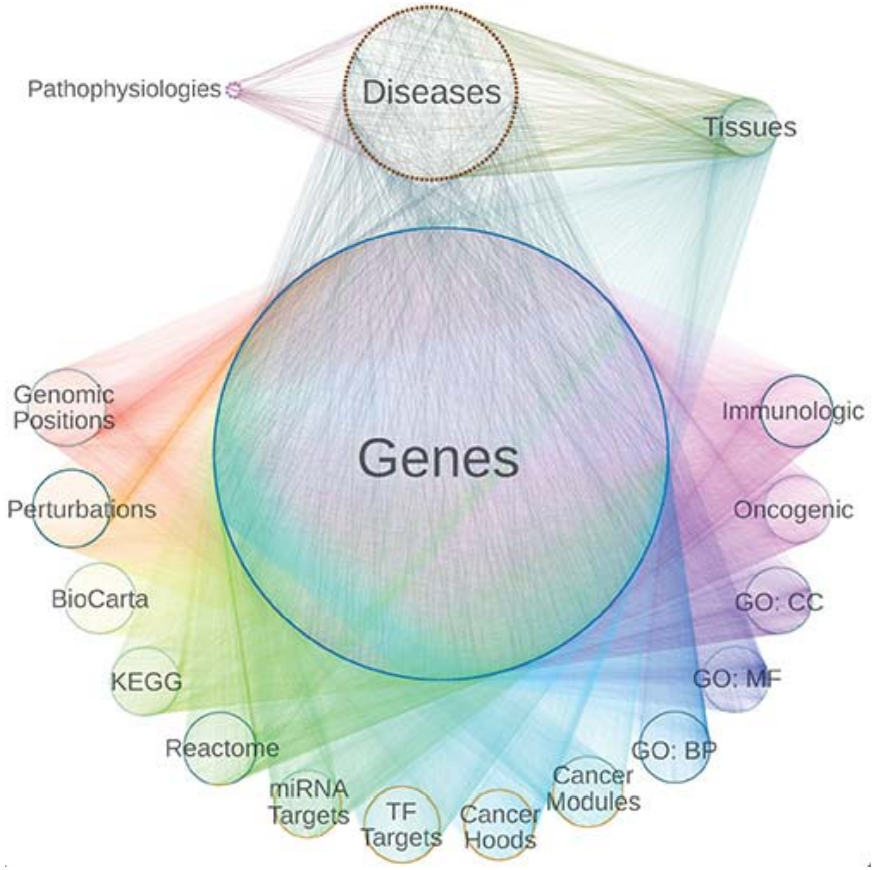
Source: Bioconductor Ann Report FY14

Growing amount of public data and data users

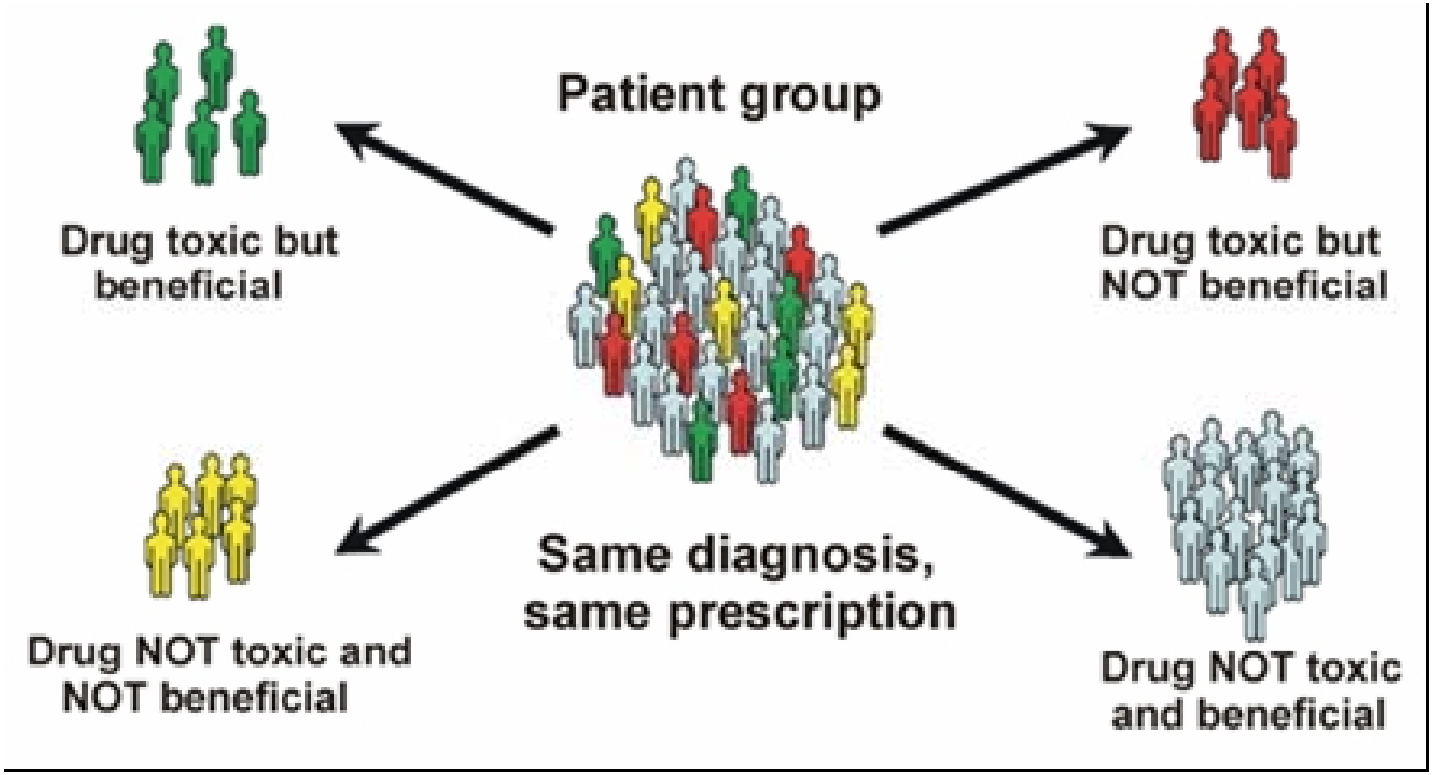


Source: NLM Cong. Report FY15

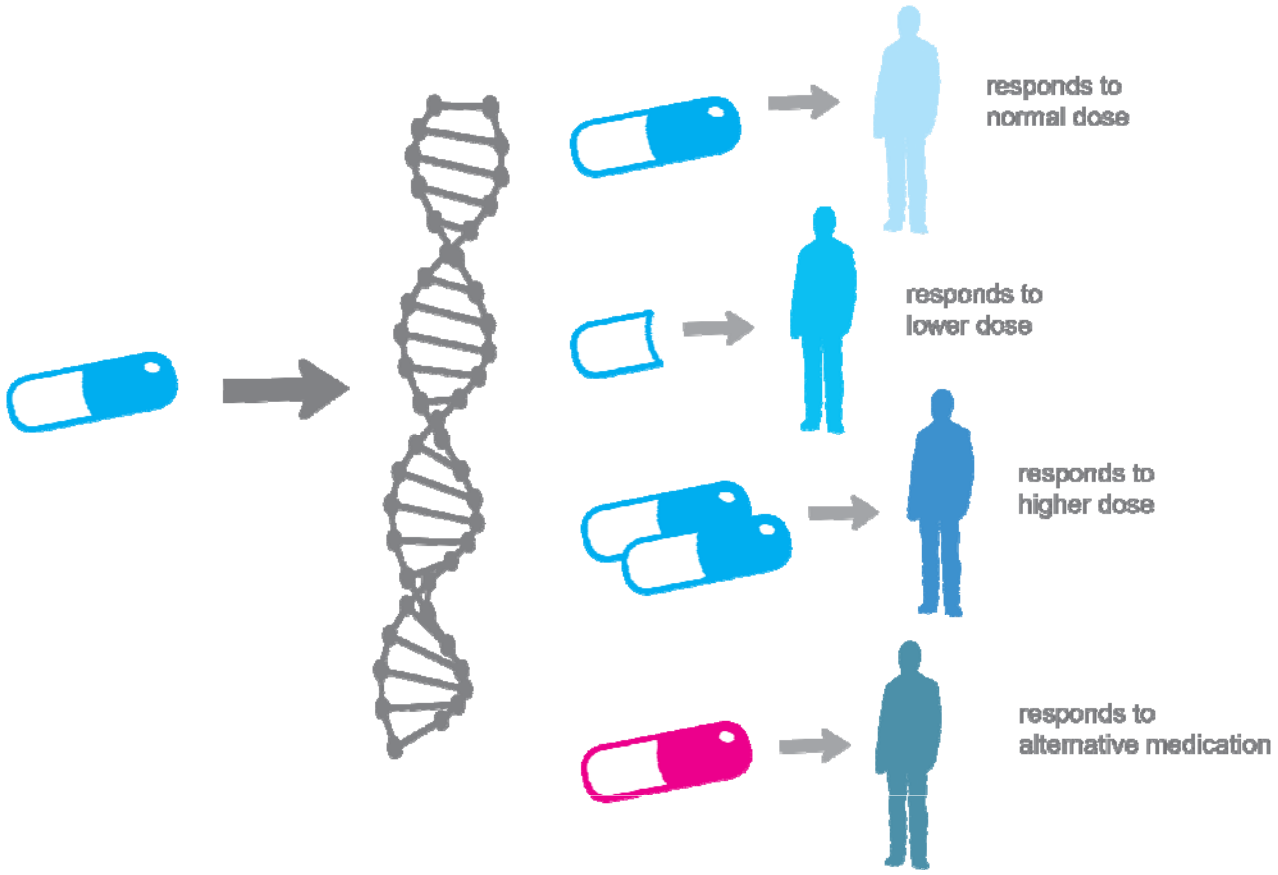
Búsqueda de marcadores que definan grupos con características patológicas comunes



Medicina Personalizada



Adaptación de la medicación



La evolución de los conocimientos en biomedicina y la aplicación de sus resultados requiere múltiples disciplinar donde la bioinformática tiene un papel fundamental