

# Introducció a la genòmica (Part I)

**Mónica Cabrera** [monica.cabrera@bsc.es](mailto:monica.cabrera@bsc.es)

**Maria Rigau** [maria.rigau@bsc.es](mailto:maria.rigau@bsc.es)

**Victoria Ruiz** [victoria.ruizserra@bsc.es](mailto:victoria.ruizserra@bsc.es)

**François Serra** [francois.serra@bsc.es](mailto:francois.serra@bsc.es)

# Què és el càncer?



# Resultats enquesta

Definición de  
“cáncer”

ENFERMEDAD  
GRAVE  
CONJUNTO DE  
ENFERMEDADES

PRODUCIDA POR  
MUTACIÓN(ES)

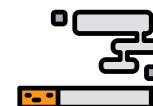
LOCALIZACIÓN

PARTE CONCRETA DEL CUERPO  
SE EXTIENDE POR EL ORGANISMO  
METÁSTASIS

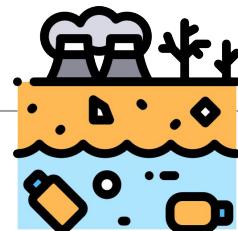
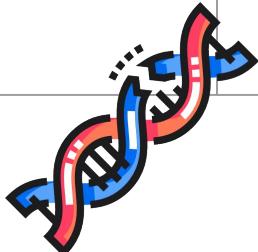
CÉLULAS “MALIGNAS”  
AFECTA A LAS CÉLULAS  
DIVISIÓN CELULAR SIN CONTROL /  
SIN PARAR / DESCONTROLADO

# Resultats enquesta

## Causas del cáncer



MUTACIONES	FALLOS EN PROTOONCOGENES GENES REPARADORES Y GENES SUPRESORES DE TUMORES	ERRORES EN DIVISIÓN CELULAR	HERENCIA GENÉTICA	HÁBITOS DE VIDA	FACTORES MEDIOAMBIENTALES
------------	---	--------------------------------	----------------------	--------------------	------------------------------



# Resultats enquesta

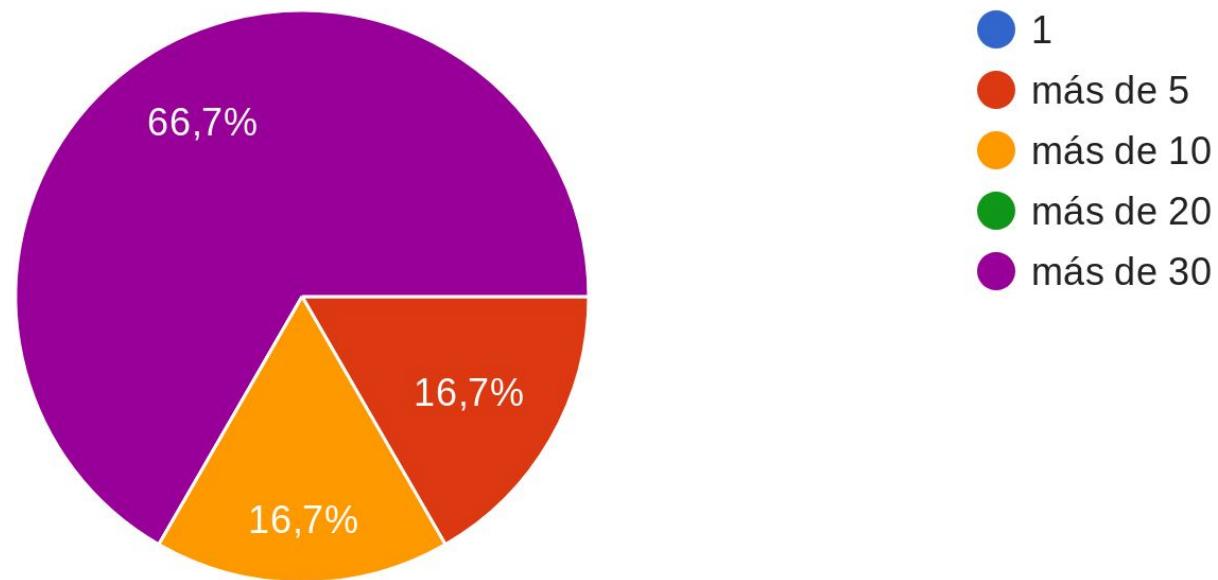
Factores de riesgo

	AUMENTA	DISMINUYE
ALIMENTACIÓN		
ESTILO DE VIDA SEDENTARIO		
HERENCIA GENÉTICA		CONDICIONA
TABAQUISMO		
DIETA EQUILIBRADA		
EDAD		
LUZ SOLAR		
RADIACIÓN		
GENÉTICOS		-
GÉRMENES (BACTERIAS, VIRUS, HONGOS, ...)		

# Resultats enquesta

A día de hoy, ¿cuántos tipos de cáncer crees que existen?

6 respuestas

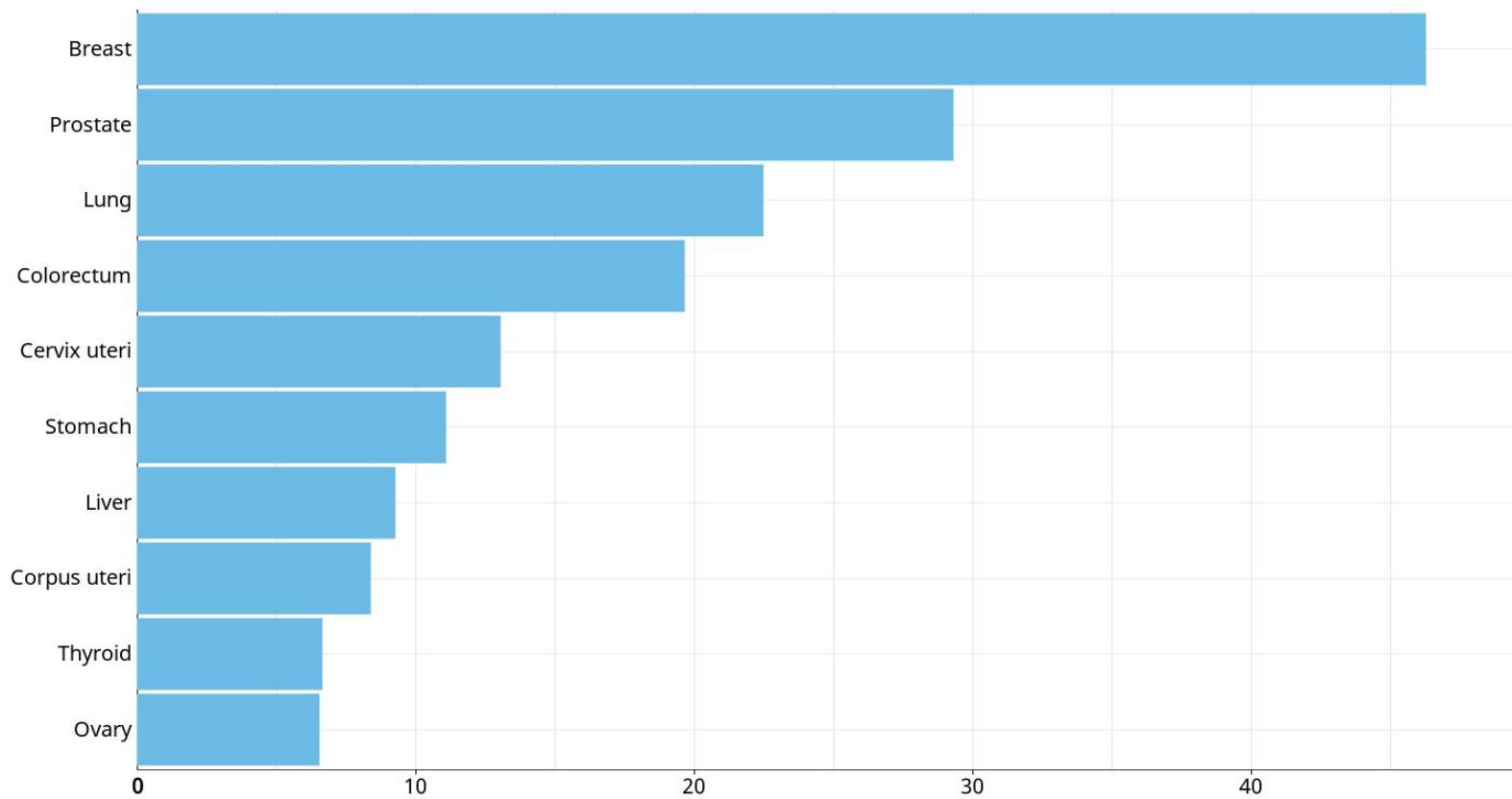


# Resultats enquesta



# Dades sobre el càncer: Tipus

Estimated age-standardized incidence rates (World) in 2018, worldwide, both sexes, all ages

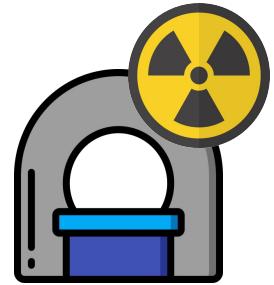


Data source: Globocan 2018  
Graph production: Global Cancer Observatory (<http://gco.iarc.fr>)

ASR (World) per 100 000

International Agency for Research on Cancer  
World Health Organization

# Resultats enquesta



RADIOTERAPIA

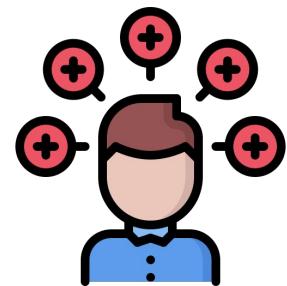


QUIMIOTERAPIA

Tratamientos



TERAPIA  
HORMONAL



INMUNOTERAPIA



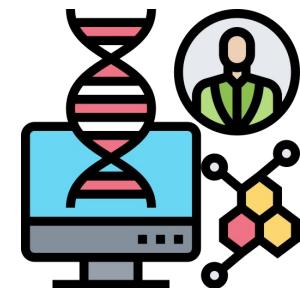
CIRUGÍA:

EXTRIPAR TUMOR

TRANSPLANTE



TERAPIAS  
DIRIGIDAS

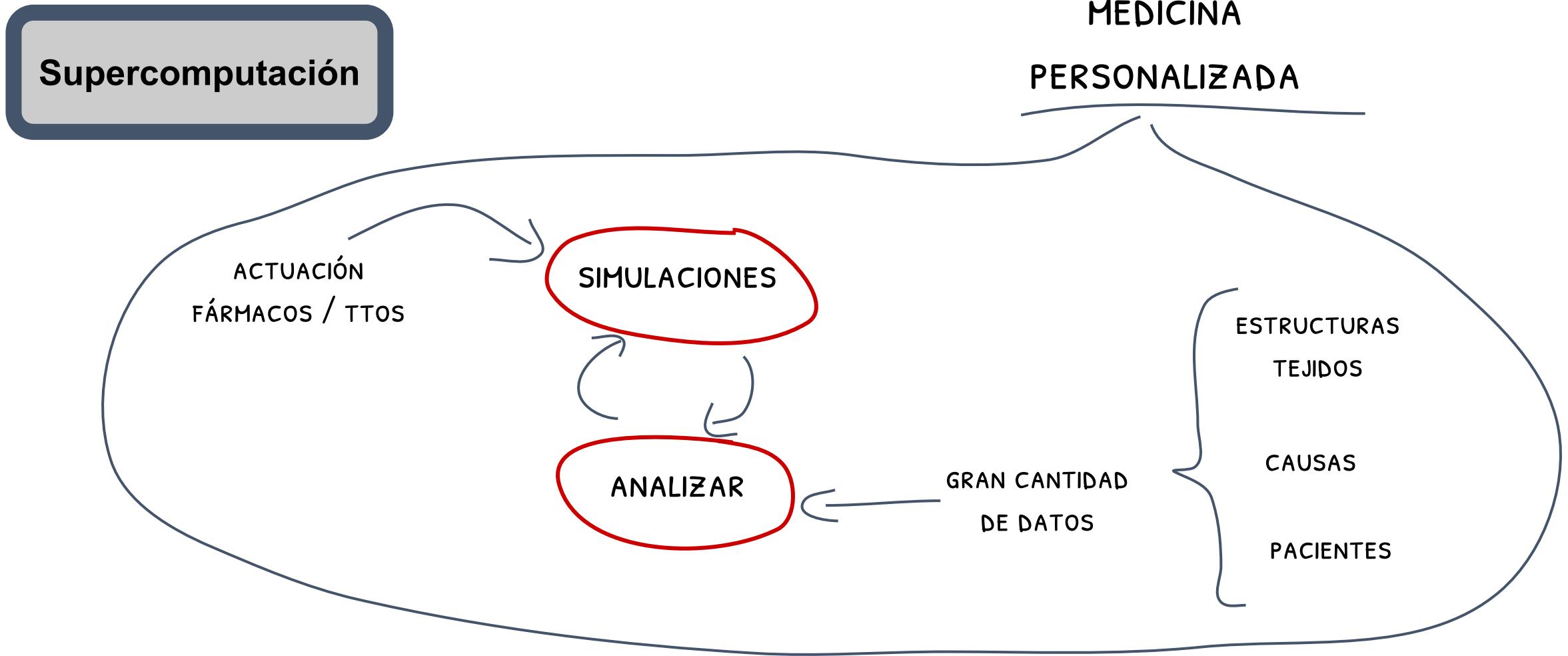


MEDICINA  
PERSONALIZADA

# Resultats enquesta

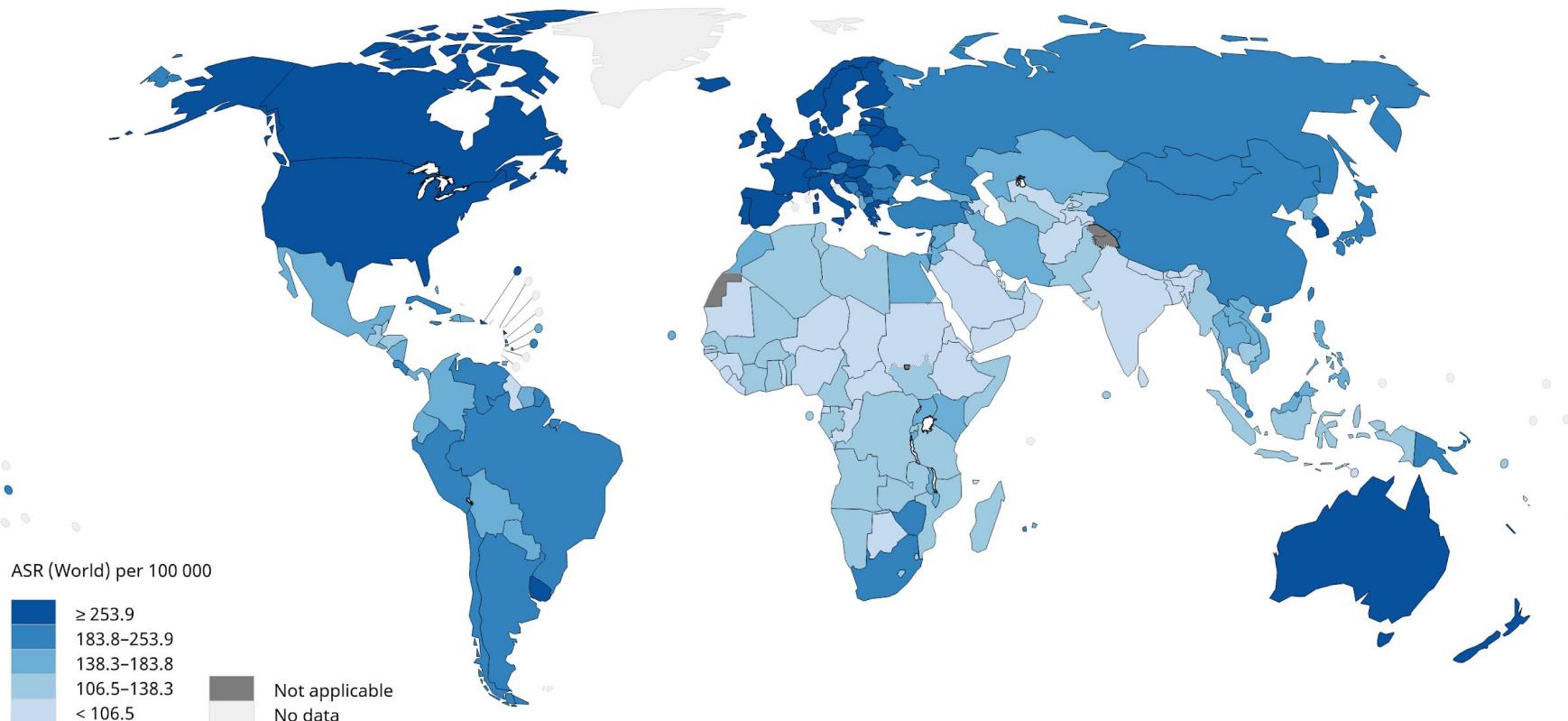


# Resultats enquesta



# Dades sobre el càncer: Incidència

Estimated age-standardized incidence rates (World) in 2018, all cancers, both sexes, all ages



All rights reserved. The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of the World Health Organization / International Agency for Research on Cancer concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted and dashed lines on maps represent approximate borderlines for which there may not yet be full agreement.

Data source: GLOBOCAN 2018  
Graph production: IARC  
(<http://gco.iarc.fr/today>)  
World Health Organization

**El càncer és una malaltia genètica**

# Què és el genoma?

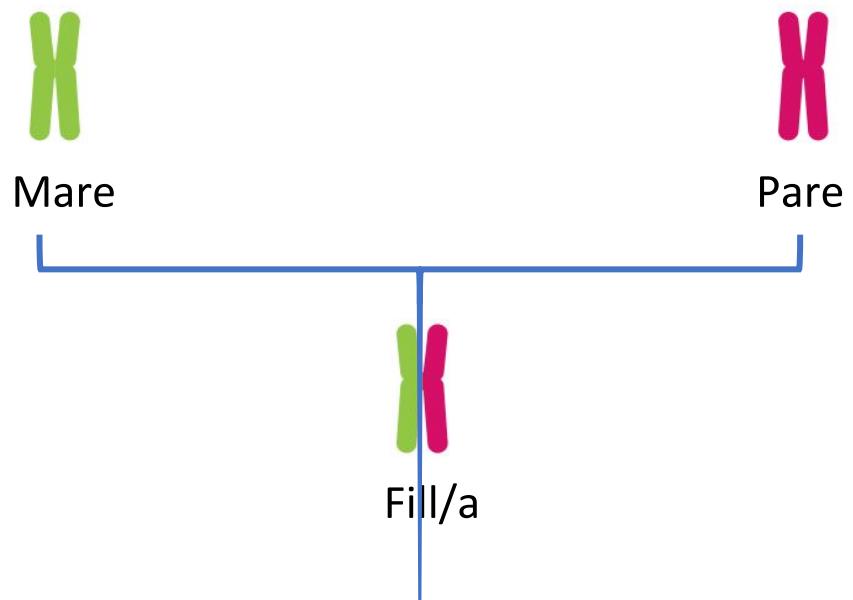
**Genoma = ADN**

L'ADN conté la informació per a que totes les nostres cèl·lules puguin funcionar i posar-se d'acord entre elles



# Què és el genoma?

- Totes (o quasi totes) les cèl·lules del nostre cos tenen el mateix ADN
- L'ADN l'heretem dels nostres pares ( $\frac{1}{2}$  de la mare,  $\frac{1}{2}$  del pare)



# Què és el genoma?

- Les instruccions codificades en el teu ADN et fan únic
- Els humans compartim el 99.9% de l'ADN



# Què és l'ADN?

**ADN = Àcid desoxiribonucleic**

Està format per dues cadenes que s'entrellacen formant una doble hèlix

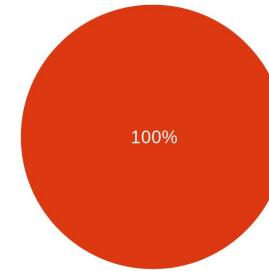
Aquestes cadenes estan formades per 4 molècules (4 bases):

- **Adenina (A)**
- **Citosina (C)**
- **Guanina (G)**
- **Timina (T)**

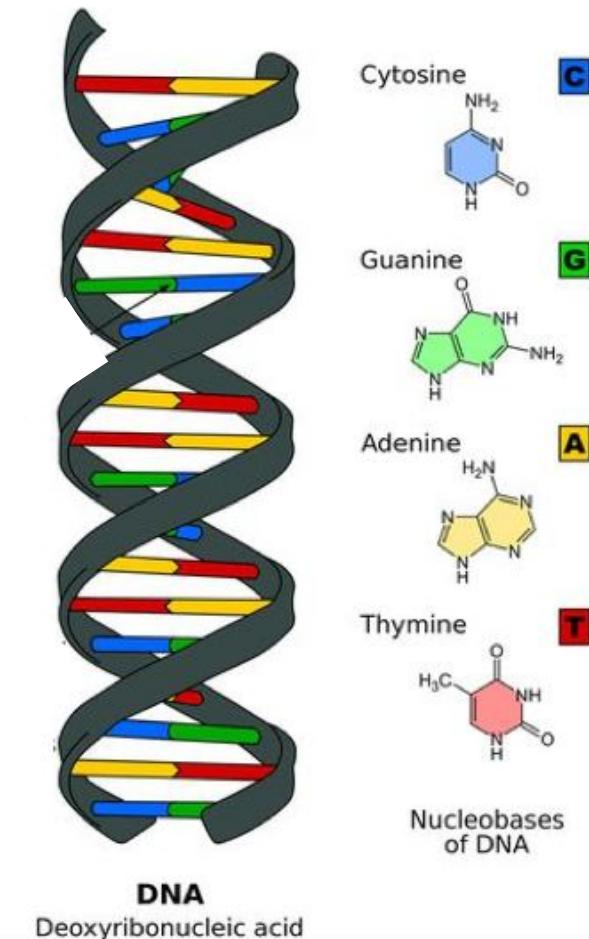
L'estrucció de l'ADN és semblant en tots els éssers vius.  
El que ens fa diferents és l'ordre de les bases.

## Resultats enquesta:

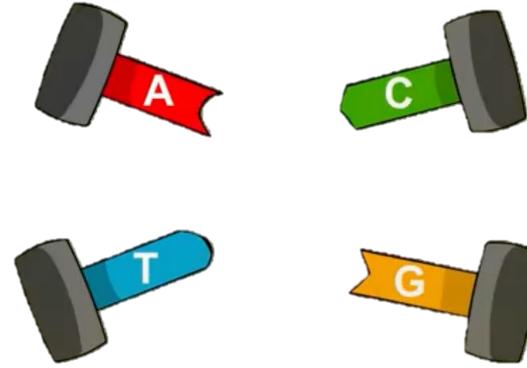
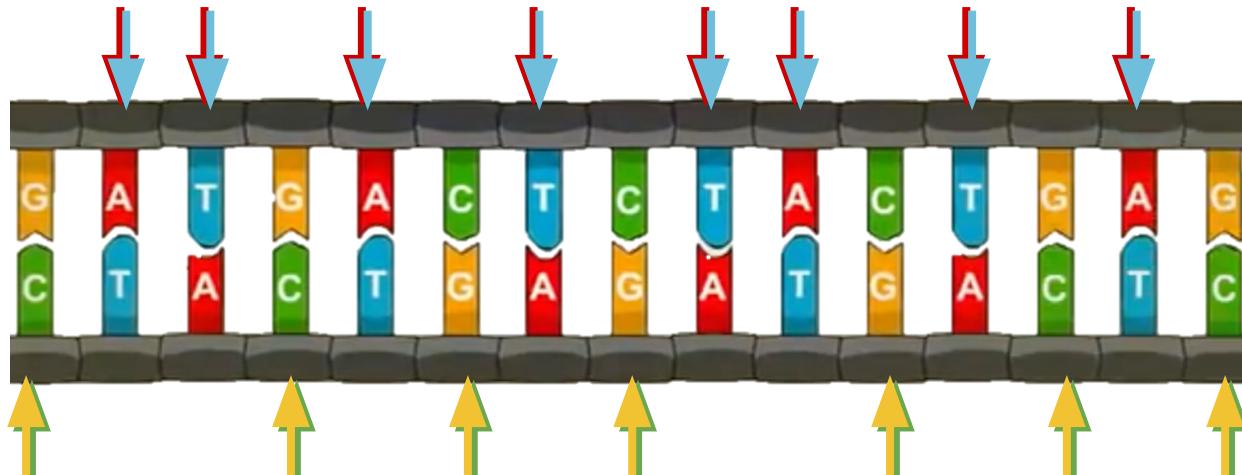
El ADN está compuesto de...  
6 respuestas



- Aminoàcidos
- Nucleòtids
- Fructosa
- Todo lo anterior



# Estructura de l'ADN



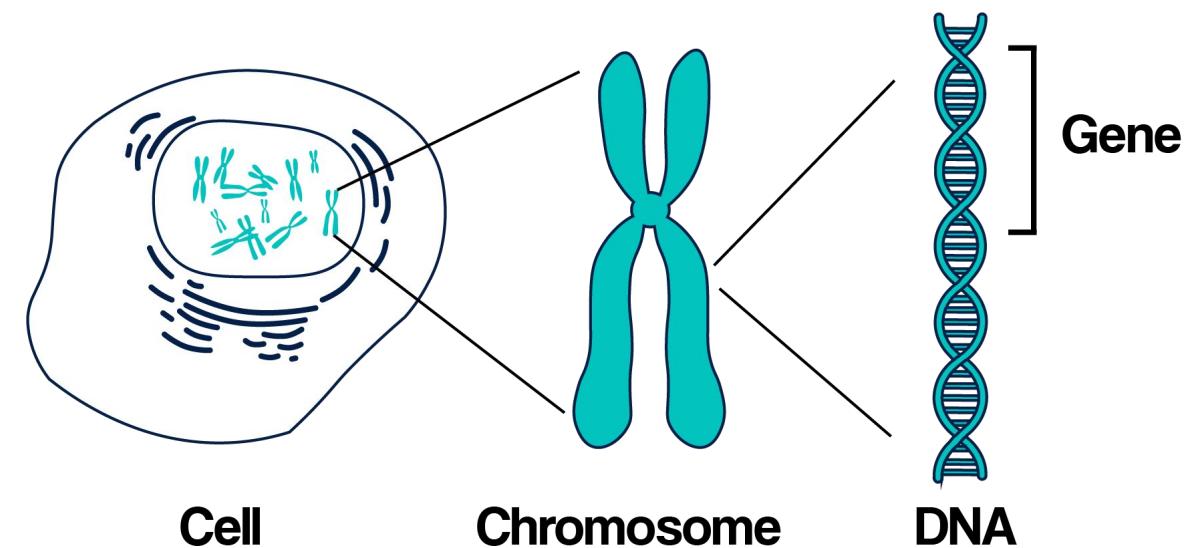
L'ordre de les bases en una cadena d'ADN fa la "seqüència".

La seqüència del genoma humà està feta per 3.000 milions de parells de bases (uns 1.000 llibres de 200 pàgines).

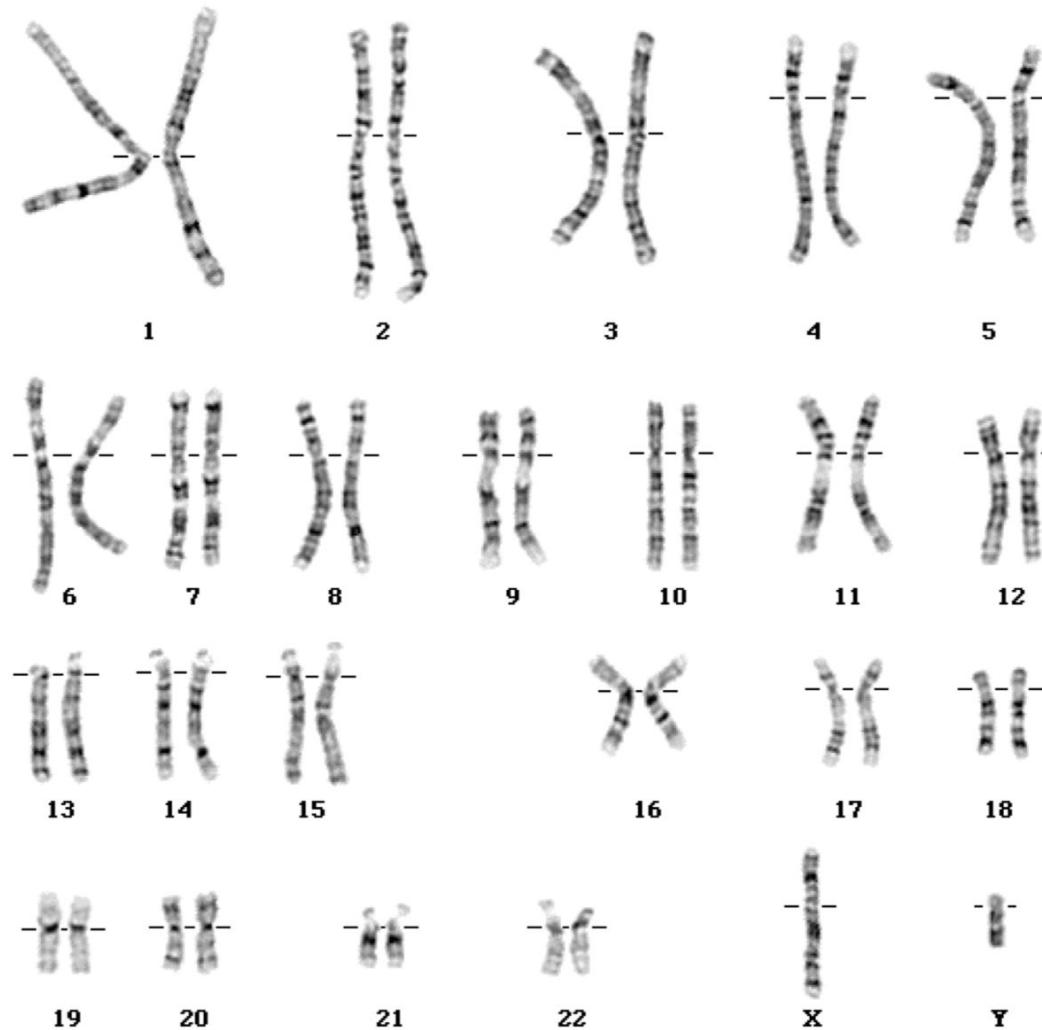
Les dues seqüències són **complementàries**. La A va amb T, la C amb la G

# Com s'organitza l'ADN dins la cèl·lula?

- L'ADN es troba dins del nucli de la cèl·lula
- Els humans tenim >3.000 milions de parells de bases d'ADN arranjats en 46 cromosomes
- Cada persona hereta dos jocs de 23 cromosomes, un del pare i un de la mare



# Visualització de l'ADN



# Activitat 1 : Python & genomes

1. Descarregar-se la seqüència del genoma humà
2. Visualitzar la mida i el contingut GC dels cromosomes
3. Reproduir un cariotip

# Resultats enquesta: Què és un gen?

## GEN

información en el ADN que marca algunas características del cuerpo.

unidad básica que contiene la información genética.

Un trozo de ADN que guarda la información sobre algo.

Una secuencia del ADN que es responsable de una función específica.

Partícula que se encuentra hallada en un cierto orden dentro de los cromosomas que dicta los caracteres hereditarios.

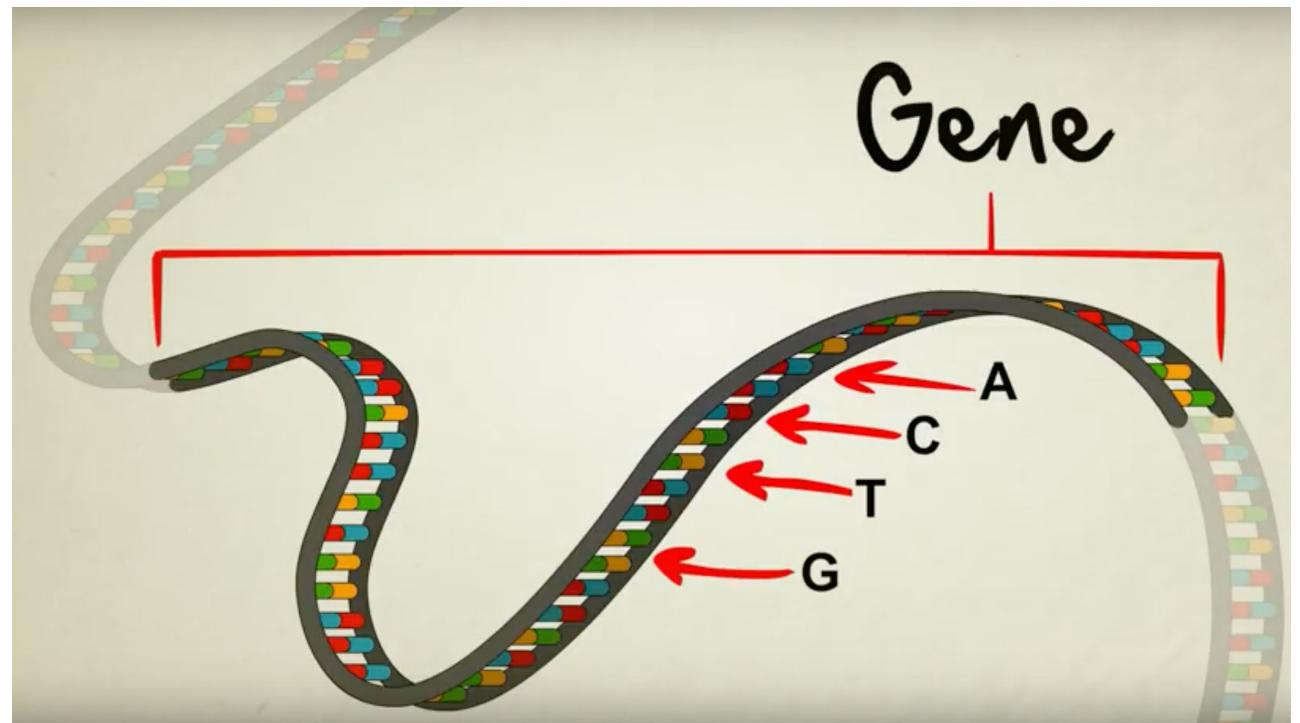
Unidad de información del ADN.

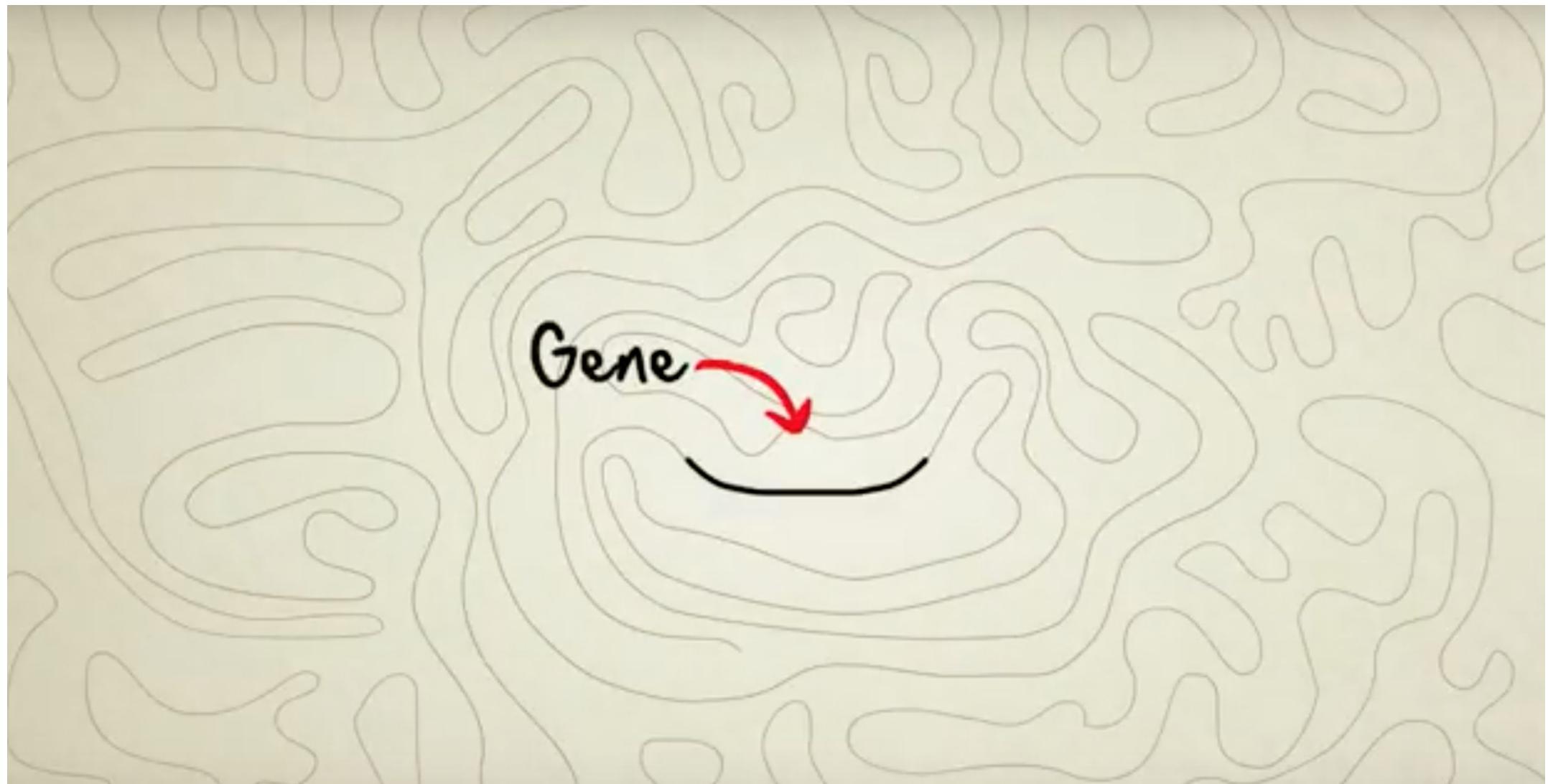
Unidad de información que por ejemplo codifica el proceso de romper una cierta proteína.

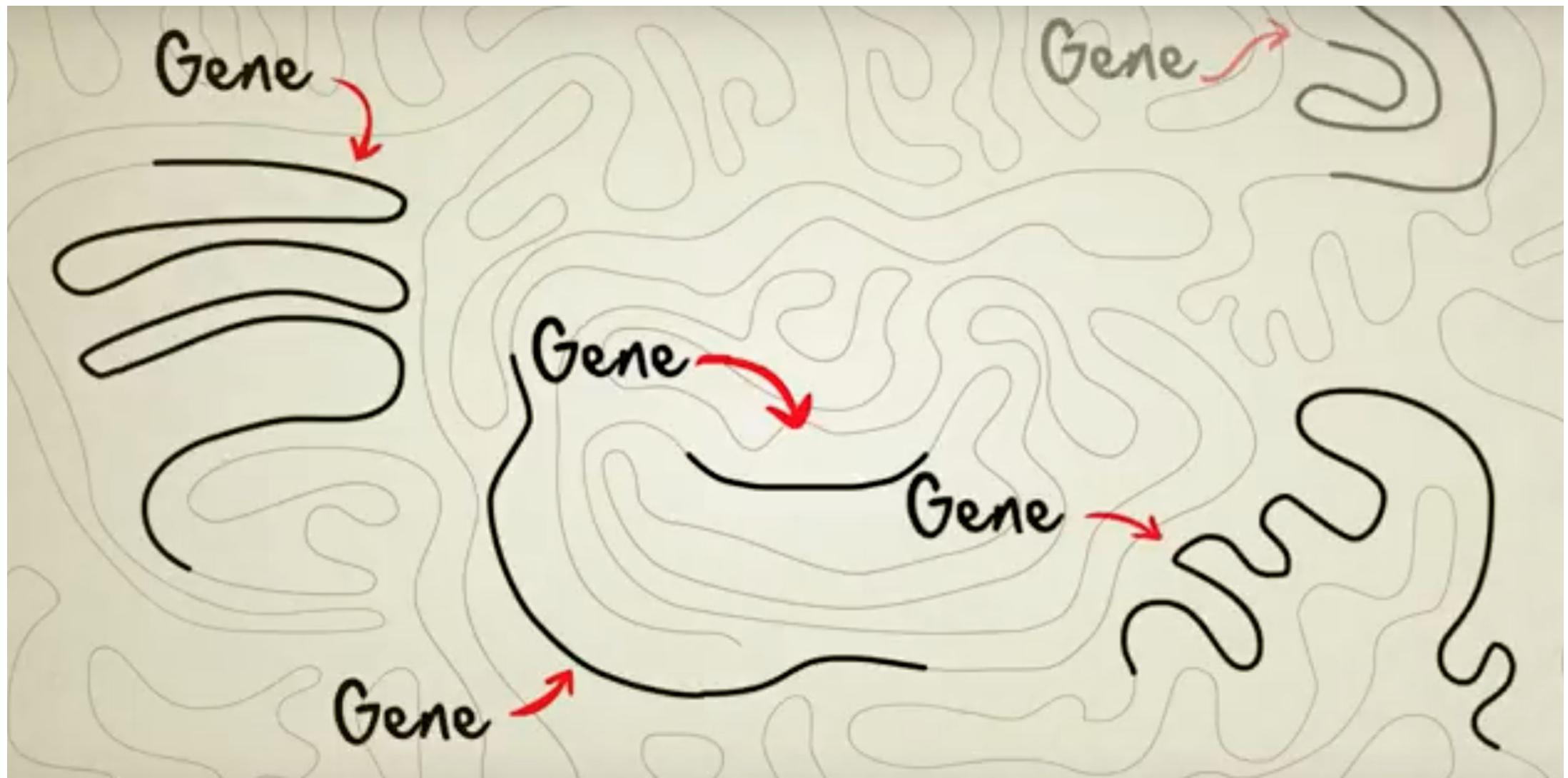
# Què és un gen?

Un gen és un fragment de DNA que conté les instruccions per a fer una proteïna.

En el genoma humà hi ha uns  
 gens

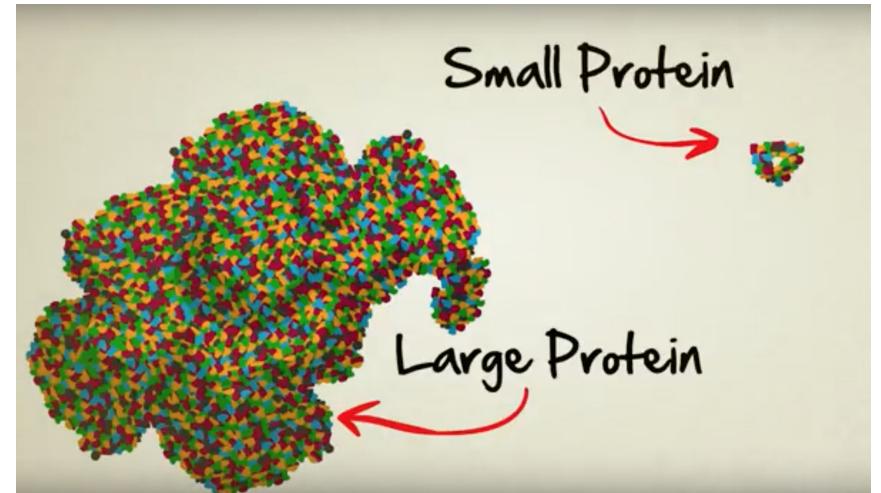






# Què són les proteïnes?

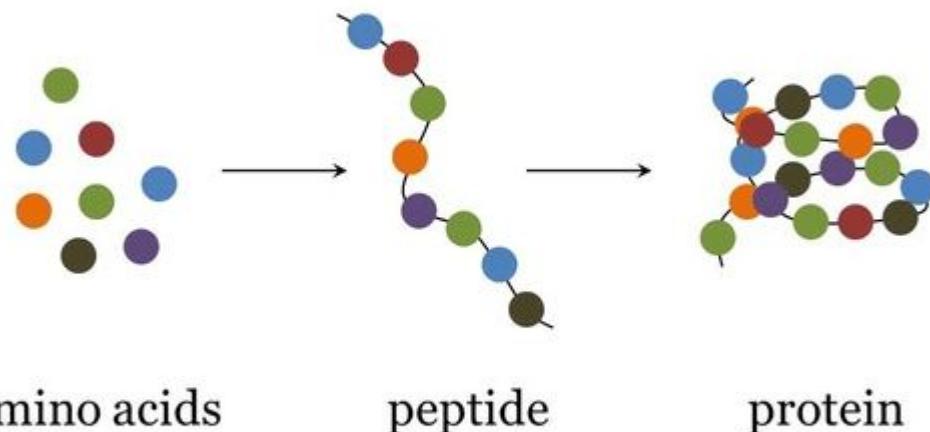
- Per a funcionar correctament, les cèl·lules depenen de milers de proteïnes funcionant bé i en el moment i lloc correctes.
- Exemples de proteïnes
  - Col·làgen (**estructural**)
  - Queratina (**estructural**)
  - Anticossos (**defensa**)
  - Insulina (**reguladora - hormona**)
  - Hemoglobina (**transport**)
  - Ovoalbúmina o lactoalbúmina (**reserva**)
  - ...
  - Pikachurina



# Els aminoàcids

Les proteïnes són cadenes d'aminoàcids

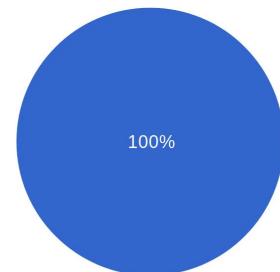
Hi ha 20 aminoàcids diferents



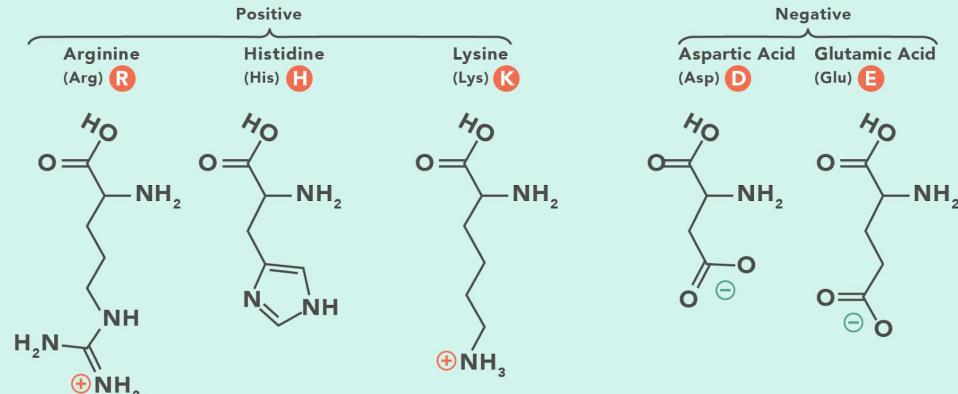
## Resultats enquesta:

Las proteïnes estan compuestas de...  
6 respostes

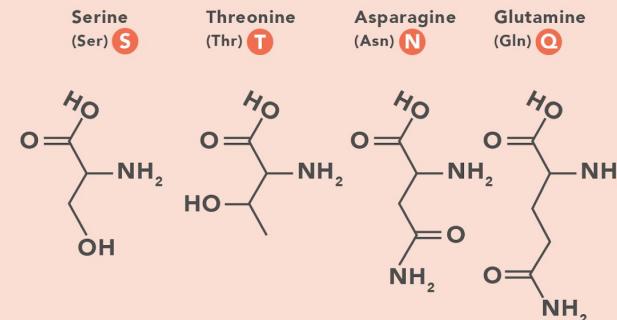
- Aminoàcidos
- Nucleòtidos
- Lípidos
- Salchichonios



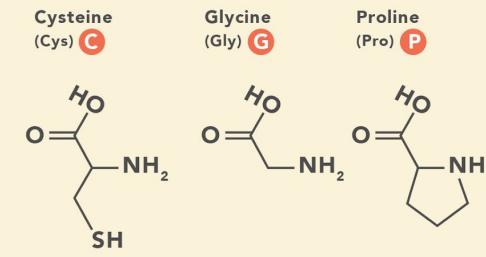
## A. Amino Acids with Electrically Charged Side Chains



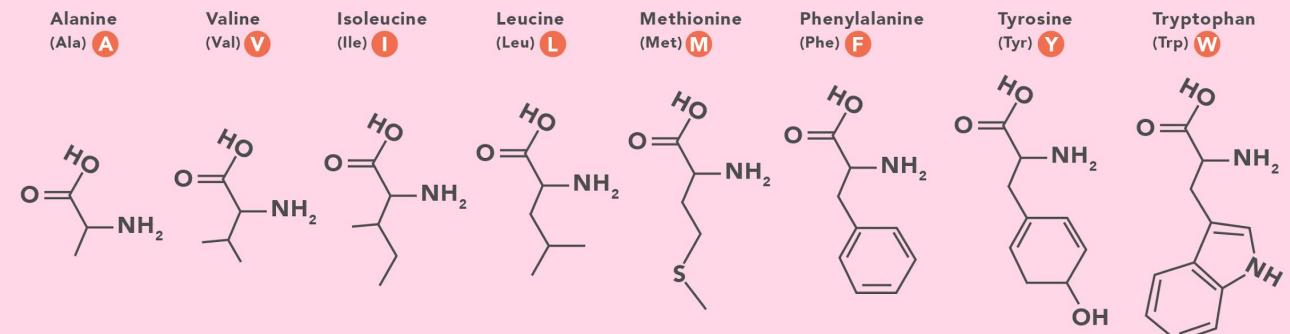
## B. Amino Acids with Polar Uncharged Side Chains



## C. Special Cases



## D. Amino Acids with Hydrophobic Side Chains



# El pas de gen a proteïna

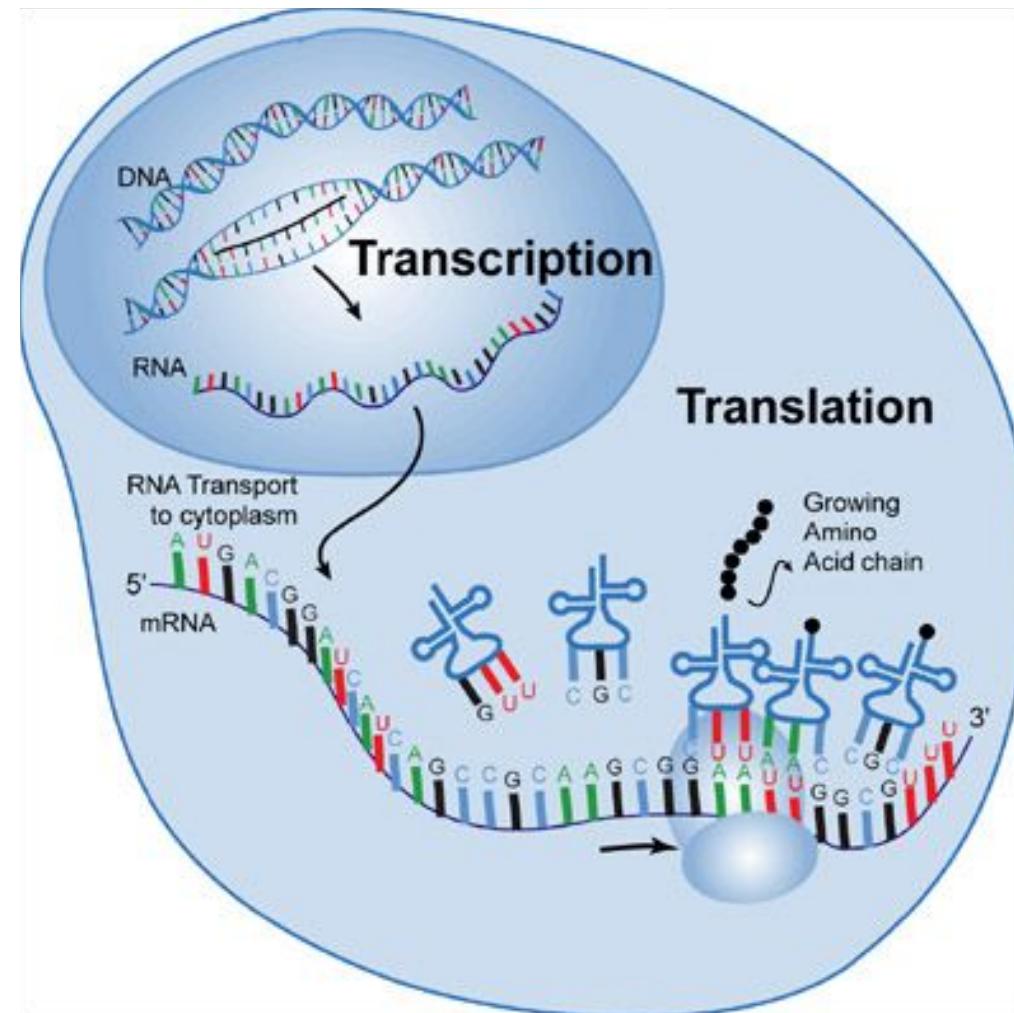
Dos passos:

## 1. Transcripció (al nucli)

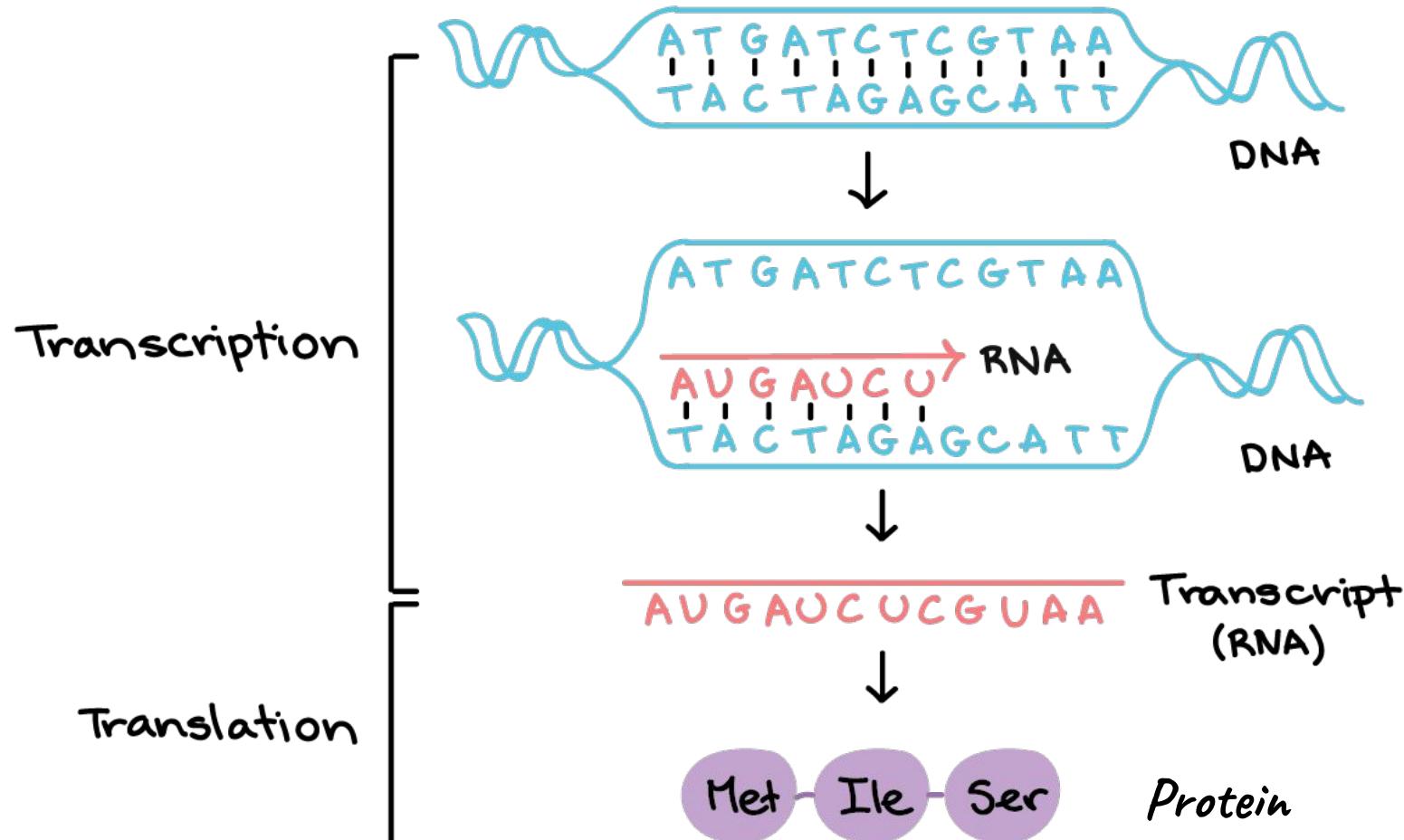
ADN → ARNm

## 1. Traducció (al citoplasma)

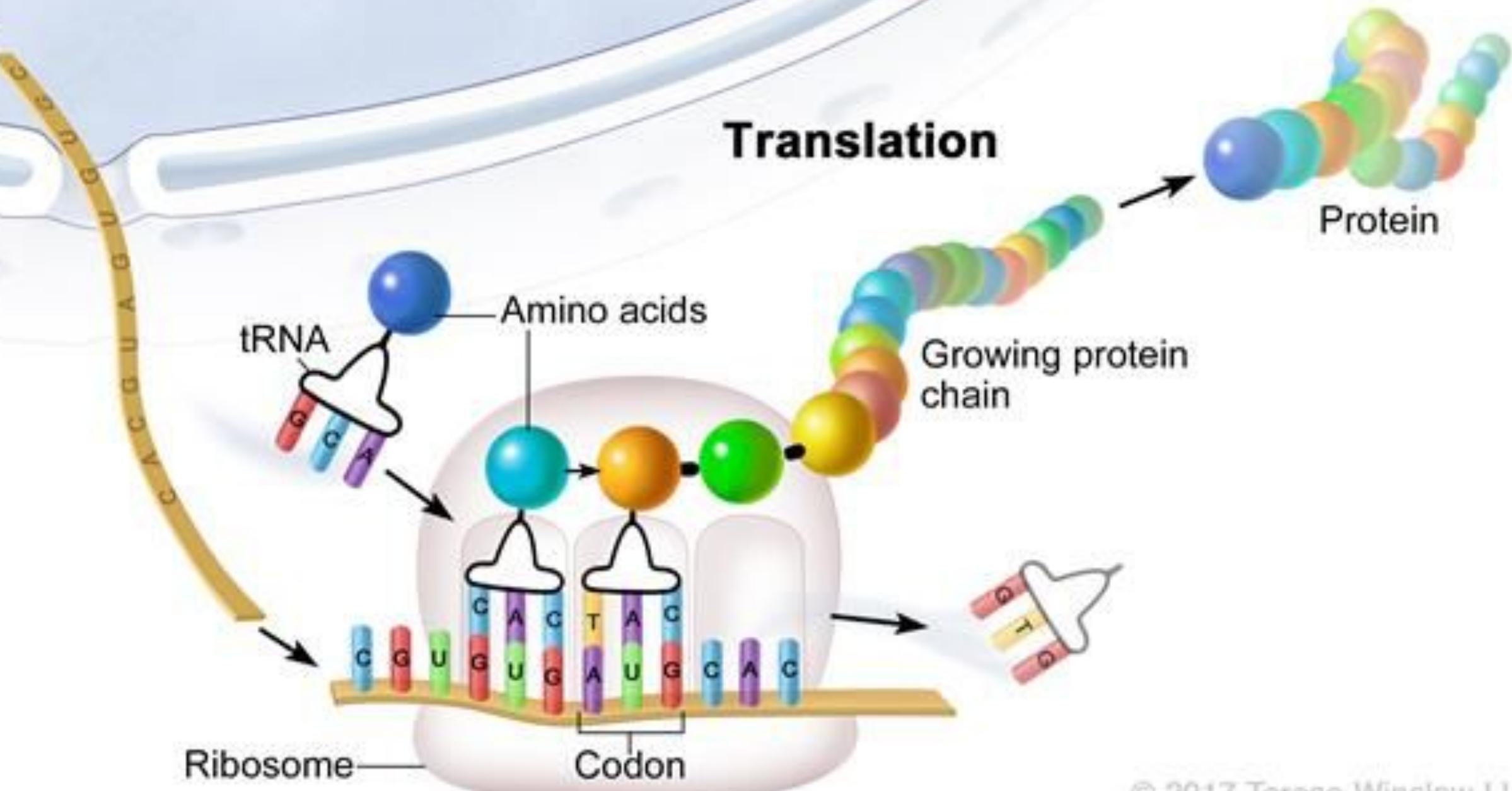
ARNm → proteïna

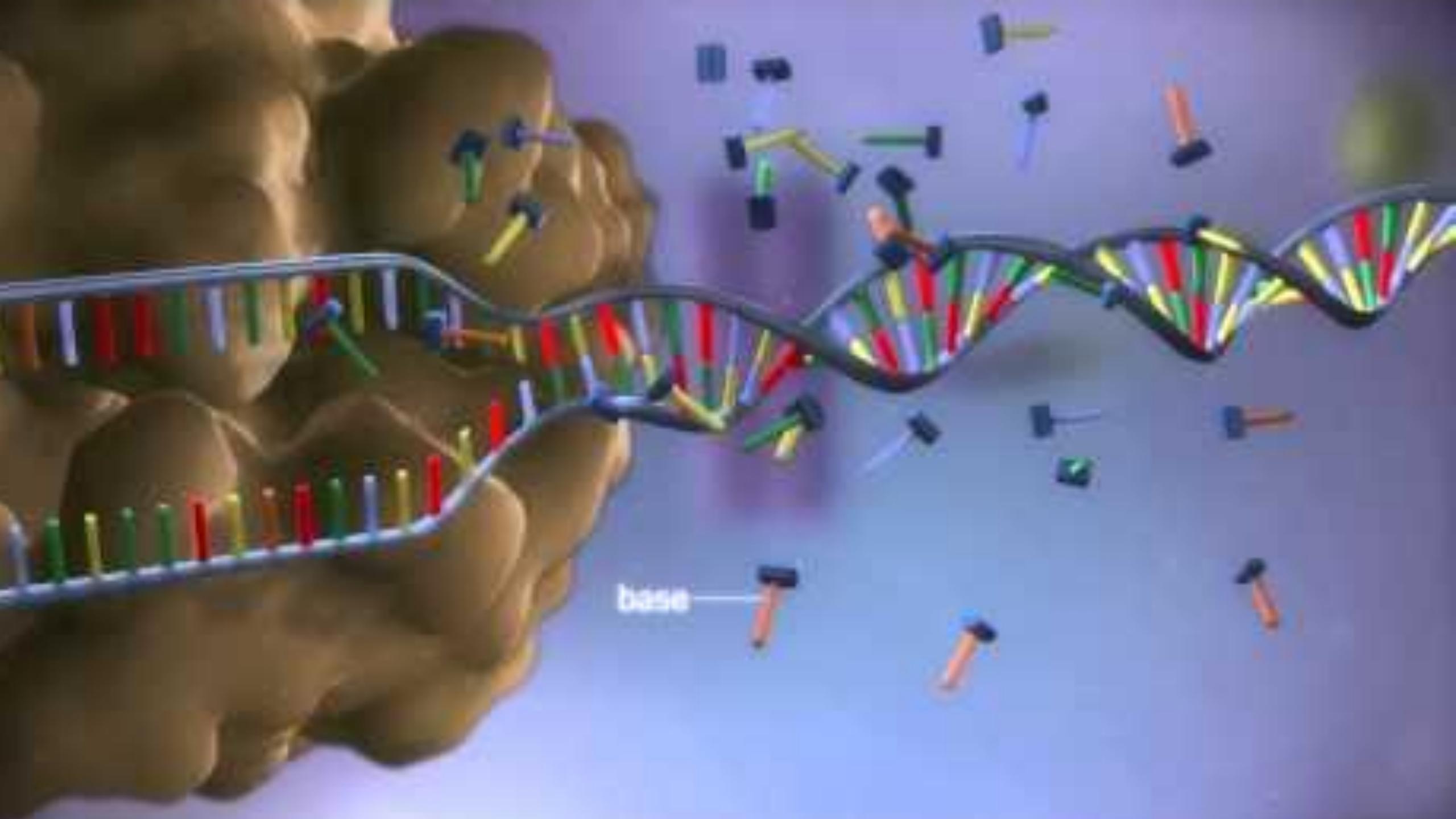


# El pas de gen a proteïna



# Translation





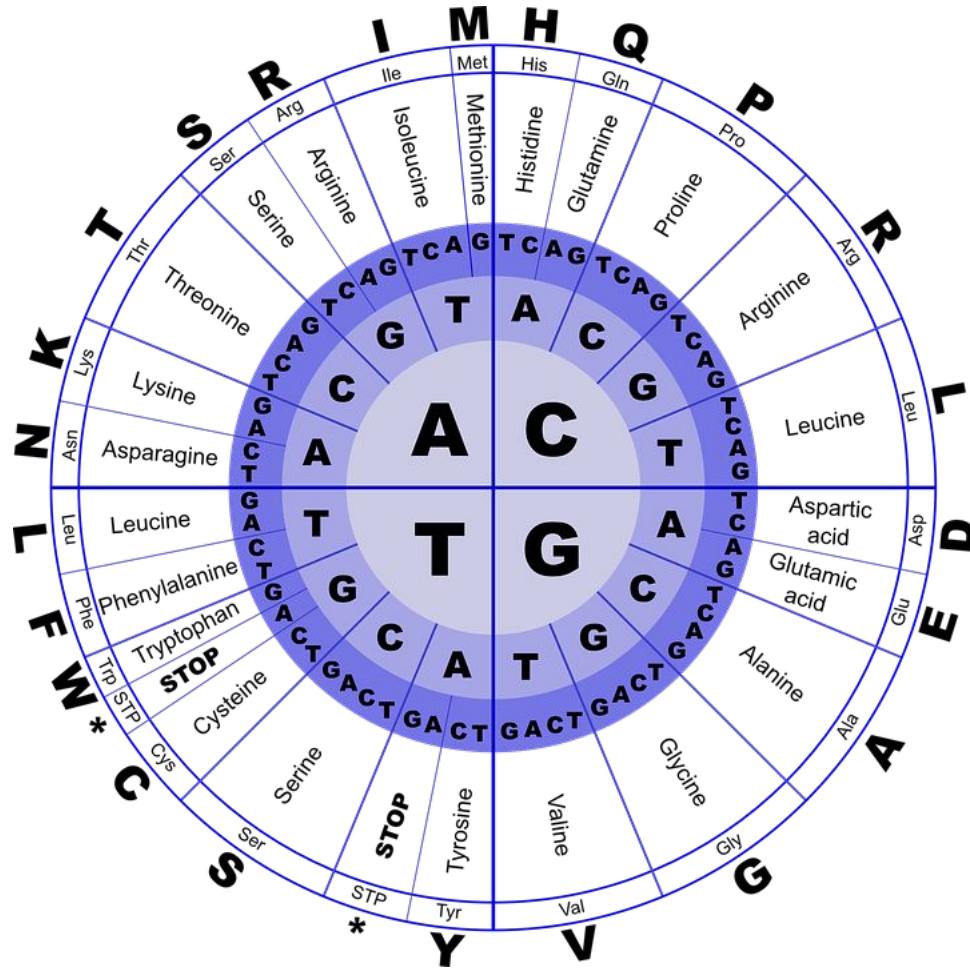
base

# El codi genètic

Les proteïnes estan formades per combinacions d'aminoàcids (n'hi ha 20 de diferents en total)

Cada un d'aquests aminoàcids està codificat per una o varies combinacions de 3 bases

Exemple: ATG = M (metionina)

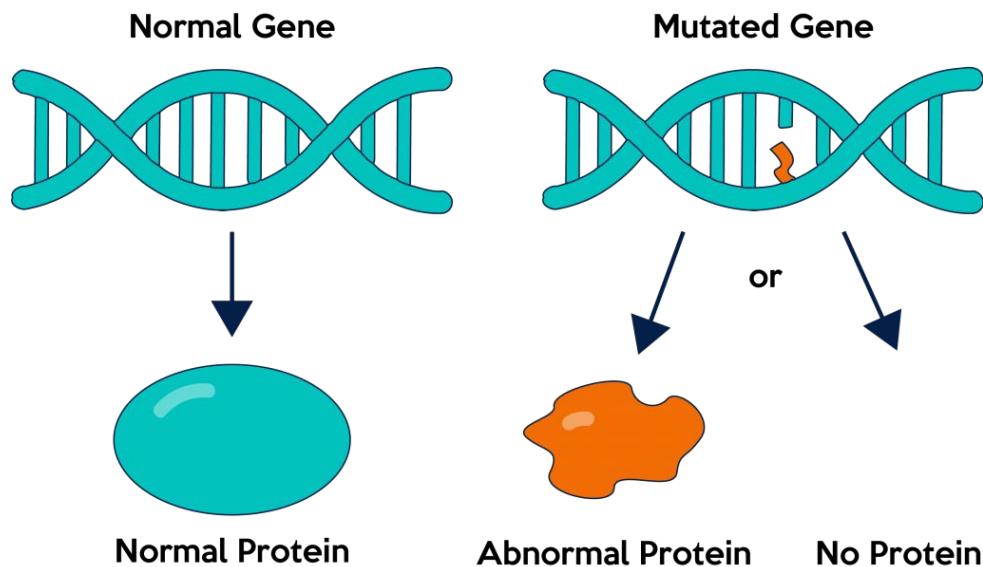


La roda es llegeix de dins cap a fora

- A Alanine
- C Cysteine
- D Aspartic Acid
- E Glutamic Acid
- F Phenylalanine
- G Glycine
- H Histidine
- I Isoleucine
- K Lysine
- L Leucine
- M Methionine
- N Asparagine
- P Proline
- Q Glutamine
- R Arginine
- S Serine
- T Threonine
- V Valine
- W Tryptophan
- Y Tyrosine

# Les mutacions

- Canvis (**mutacions**) en una sola base d'un gen poden afectar la funció d'una proteïna

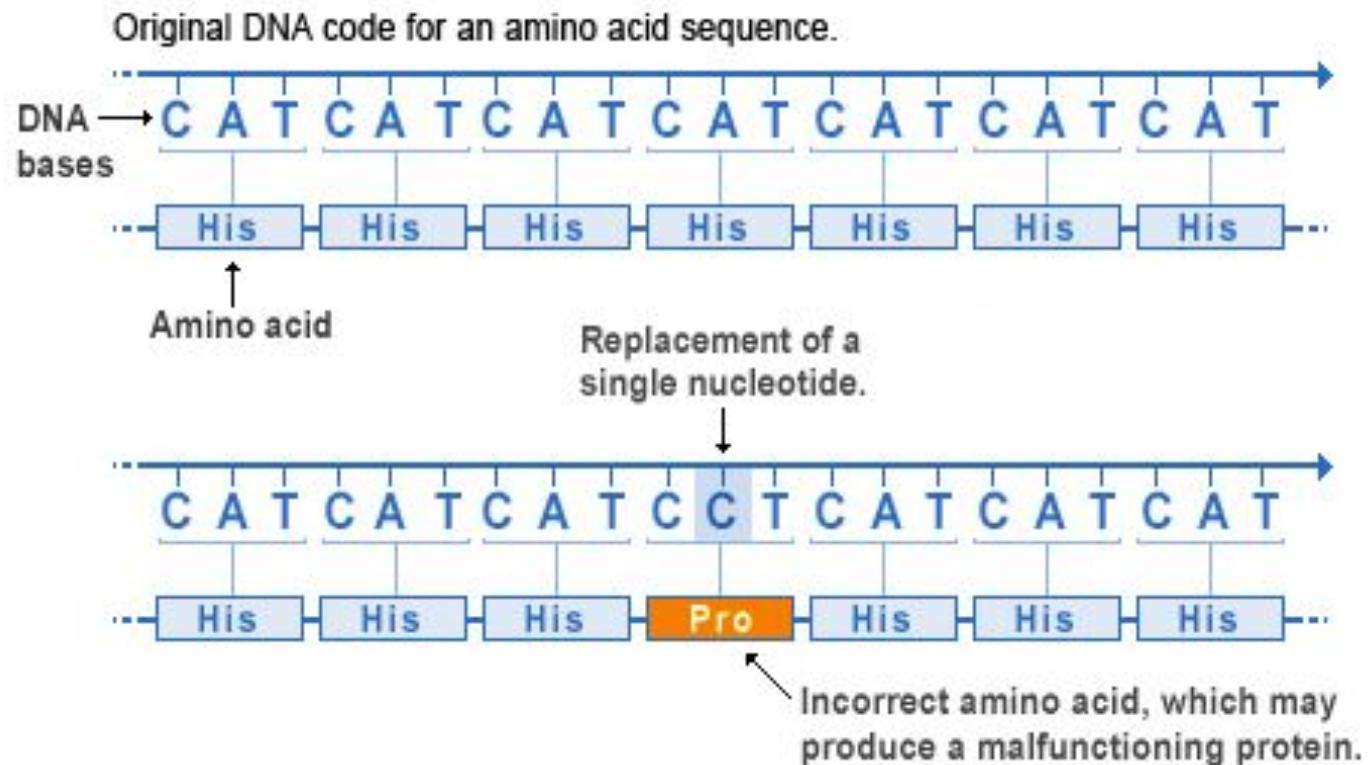


- Aquests canvis poden donar lloc a malalties

Malaltia	Gen responsable
Fenilcetonúria	<i>PAH</i>
Fibrosi quística	<i>CFTR</i>
Anèmia falciforme	<i>HBB</i>
Albinisme oculocutani tipus II	<i>OCA2</i>
Malaltia de Huntington	<i>HTT</i>
Distròfia miotònica tipus 1	<i>DMPK</i>
Neurofibromatosi tipus 1	<i>NF1</i>
Hemofília A	<i>F8</i>
Distròfia muscular de Duchenne	<i>DMD</i>
Síndrome de Rett	<i>MECP2</i>

# Les mutacions

Una mutació pot afectar la seqüència de la proteïna



# Resultats enquesta: Qué és una mutació?

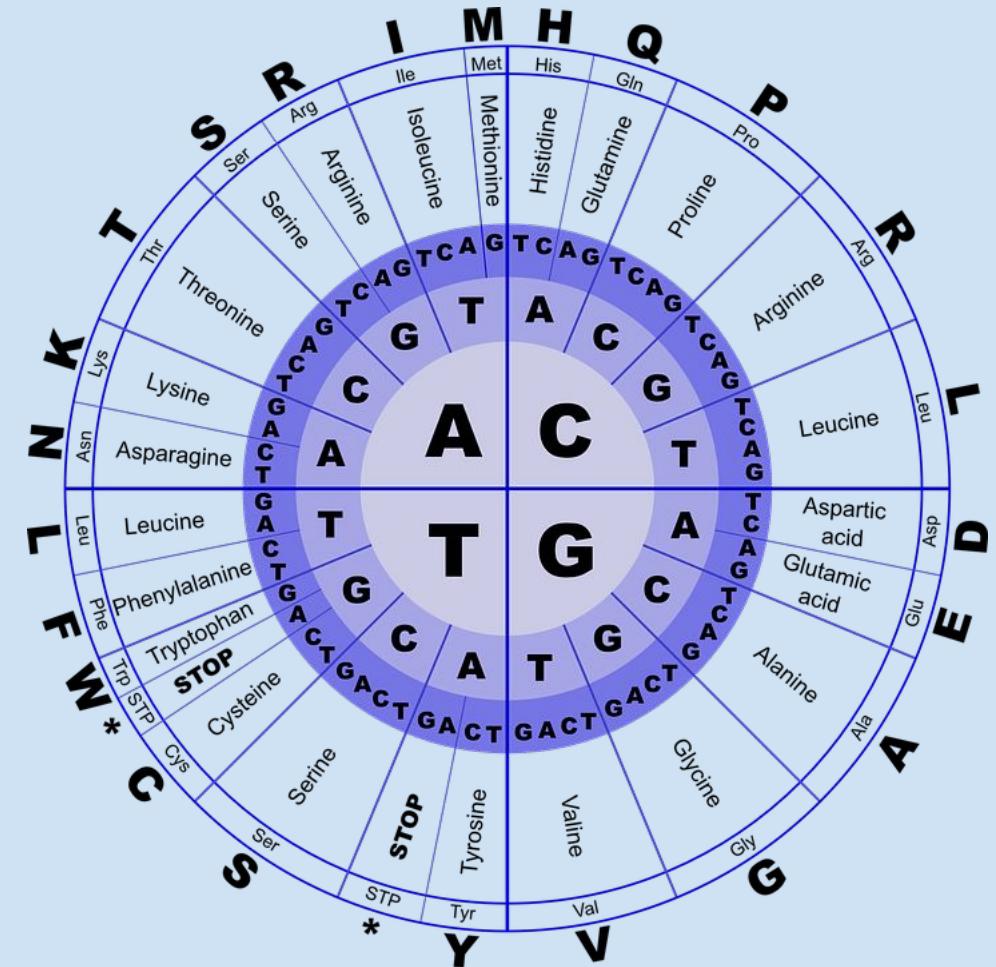
MUTACIÓN
Es un <u>cambio</u> realizado en uno de estos genes.
<u>transformación</u> o cambio de forma.
<u>Variación</u> de los genes.
<u>Cambio</u> aleatorio de uno de al menos una de las bases nitrogenadas que conforman el ADN.
<u>Modificación</u> de alguna cosa.
<u>Cambio</u> en las características de un ser.
<u>Cambio</u> de la estructura de ADN durante una duplicación.

# Activitat 2: Traducció manual

DNA_seq_1	GTA	AAA	ATA	GGT	GAT	GAC	CTC
Protein_seq_1							
DNA_seq_2	GTA	AAA	ATA	GGT	GAC	GAC	CTC
Protein_seq_2							
DNA_seq_3	GTA	AGA	ATA	GGT	GAT	GAC	CTC
Protein_seq_3							

## Preguntes

- Hi ha diferències a la seqüència d'ADN?
- Hi ha diferències a la seqüència d'aminoàcids? (a la proteïna)



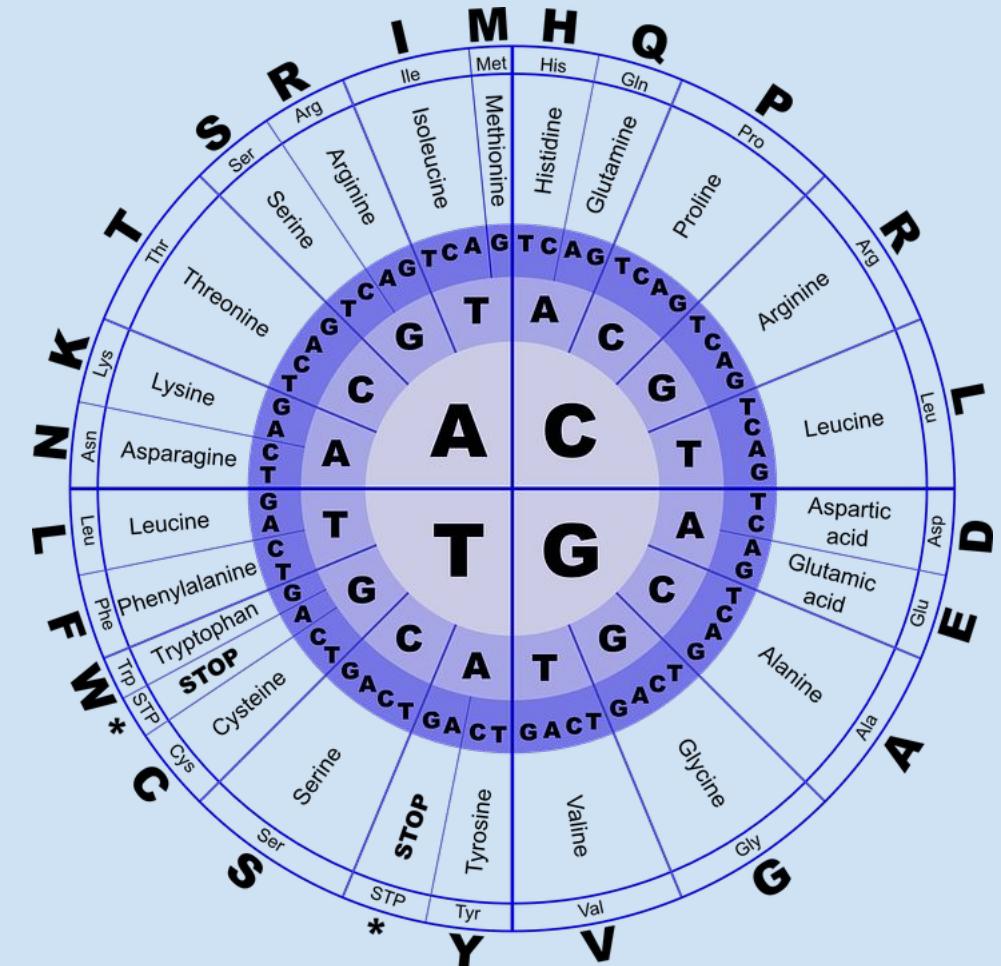
		Second base					
		U	C	A	G		
First base	U	UUU Phenylalanine UUC UUA UUG	UCU Serine UCC UCA UCG	UAU Tyrosine UAC UAA Stop codon UAG Stop codon	UGU Cysteine UGC UGA Stop codon UGG Tryptophan W	UC CA AG	Third base
	C	CUU Leucine CUC CUA CUG	CCU Proline CCC CCA CCG	CAU Histidine CAC CAA Glutamine CAG	CGU CGC Arginine CGA CGG	UC CA AG	
	A	AUU Isoleucine AUC AUA AUG M Methionine start codon	ACU Threonine ACC ACA ACG	AAU Asparagine AAC AAA AAG Lysine	AGU Serine AGC AGA Arginine AGG	UC CA AG	
	G	GUU Valine GUC GUA GUG	GCU Alanine GCC GCA GCG	GAU Aspartic acid GAC GAA Glutamic acid GAG	GGU Glycine GGC GGA GGG	UC CA AG	

# Activitat 2: Traducció manual. SOLUCIÓ

DNA_seq_1	GTA	AAA	ATA	GGT	GAT	GAC	CTC
Protein_seq_1	V	K	I	G	D	D	L
DNA_seq_2	GTA	AAA	ATA	GGT	GAC	GAC	CTC
Protein_seq_2	V	K	I	G	D	D	L
DNA_seq_3	GTA	AGA	ATA	GGT	GAT	GAC	CTC
Protein_seq_3	V	R	I	G	D	D	L

## Preguntes

- Hi ha diferències a la seqüència d'ADN?
- Hi ha diferències a la seqüència d'aminoàcids? (a la proteïna)



# Activitat 3: Traducció utilitzant eines online

1. Alinear ADN de diferents mostres de l'exó 15 del gen BRAF (**`braf_mutations_exon15.txt`**) utilitzant <https://mafft.cbrc.jp/alignment/software/>
2. Traduir la seqüència d'ADN amb:  
[https://www.ebi.ac.uk/Tools/st/emboss\\_transeq/](https://www.ebi.ac.uk/Tools/st/emboss_transeq/)
3. Alinear les proteïnes amb: <https://mafft.cbrc.jp/alignment/software/>

# Activitat 4 : Traducció utilitzant Python

[https://github.com/bsc-life/Bojos\\_supercomputacio/](https://github.com/bsc-life/Bojos_supercomputacio/)

Obtenció i traducció de l'exò 15 del gen BRAF

# Mutacions germinals vs. somàtiques

- **Germinals**

- Mutacions **heretades** de la mare o del pare
- Present en totes les cèl·lules del cos
- Poden donar **susceptibilitat** a càncer

- **Somàtiques**

- Mutacions en una cèl·lula que no sigui l'òvul o l'espermatozou (**no heretables**)
- La mutació només estarà present en una cèl·lula i la seva descendència
- Poden donar lloc a càncer

# Càncer

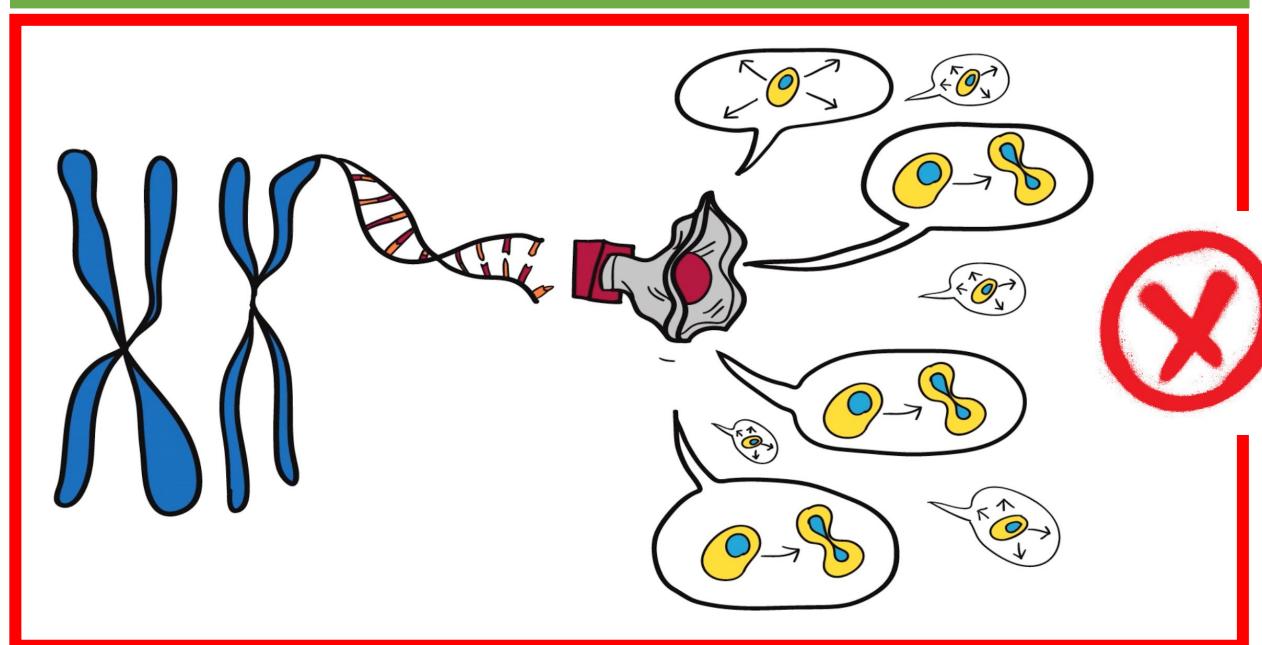
En condicions normals, els gens dirigeixen quan/com la cèl·lula ha de **créixer, funcionar, dividir-se o morir**.



# Càncer

En condicions normals, els gens dirigeixen quan/com la cèl·lula ha de **créixer, funcionar, dividir-se o morir**.

Algunes vegades les instruccions es fan malbé (per mutacions) i **una cèl·lula** pot començar a multiplicar-se descontroladament.



# Factors de risc

- Edat
- Tabac
- Alcohol
- Dieta
- Obesitat
- Exposició al sol
- Algunes infeccions
- Substàncies cancerígenes
- Inflamació crònica
- Hormones

4 de cada 10 casos de càncer es podrien evitar canviant hàbits.

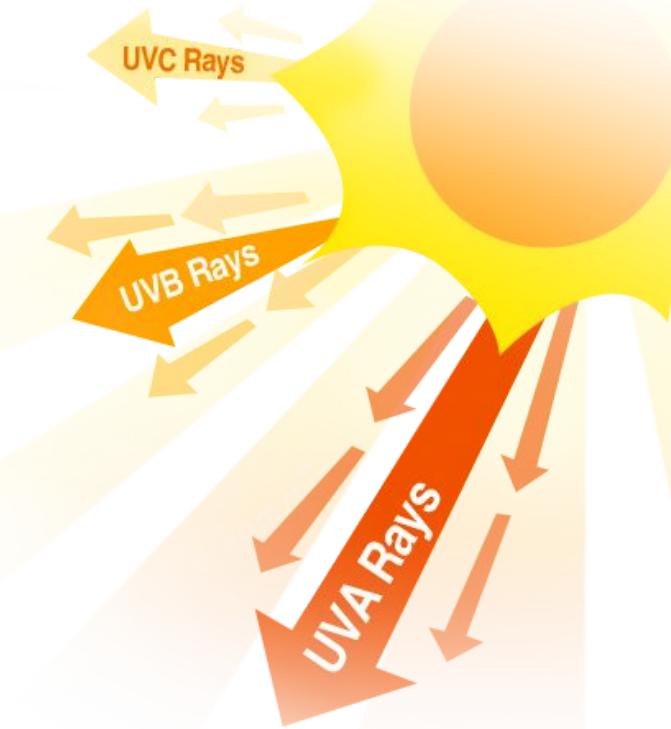
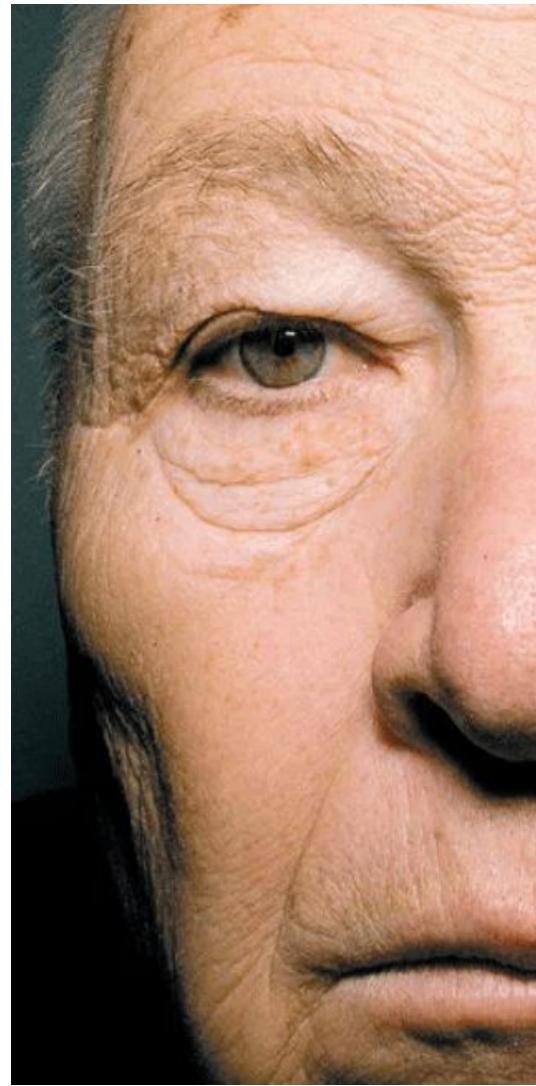


## Causes genètiques

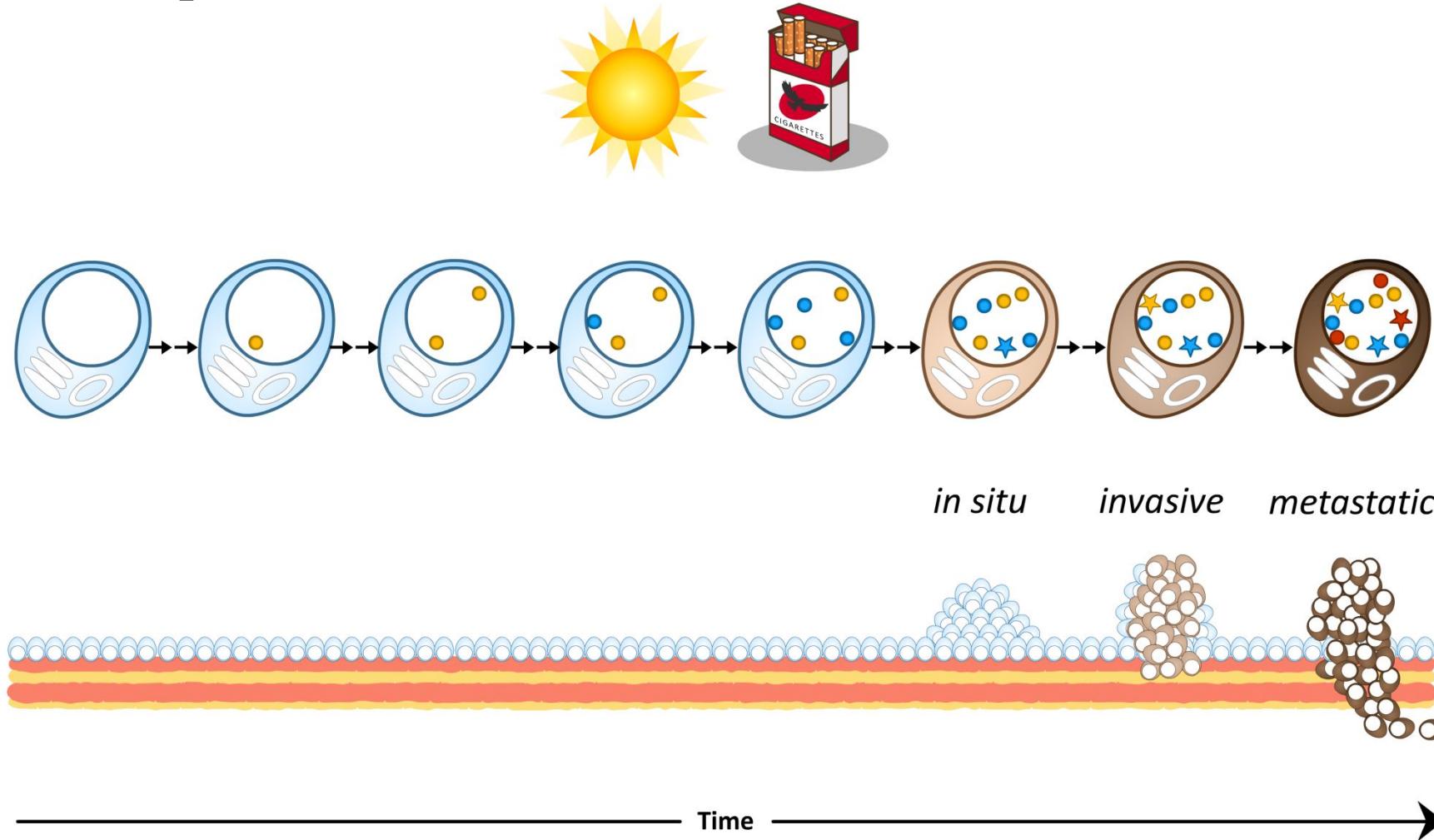
Algunes mutacions heretades poden donar un risc més elevat de tenir càncer

Aquest és el cas del 2-3% dels casos totals de càncer

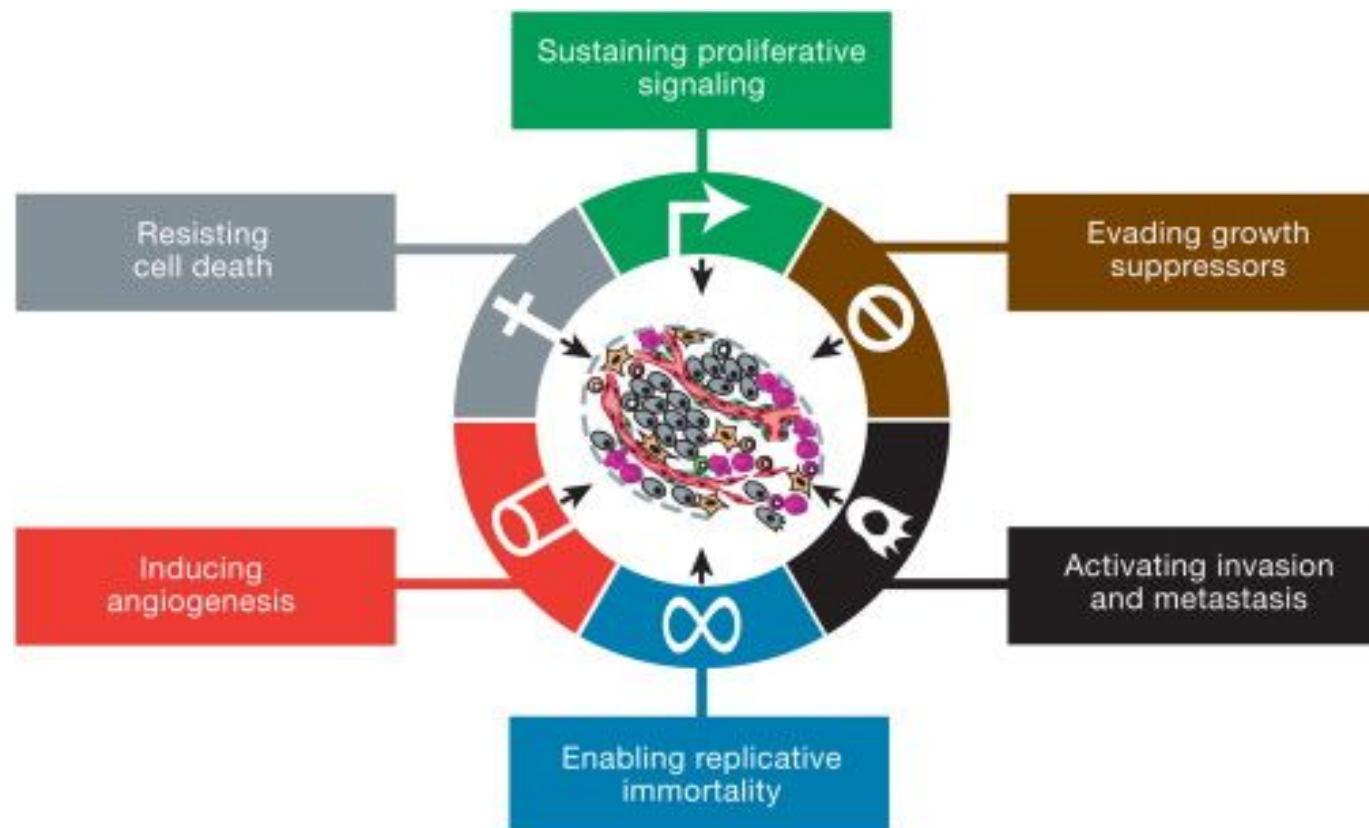
Bill McElligott, 69 anys  
Camioner durant 28 anys



# Com causen càncer, les mutacions somàtiques?



# Com són les cèl·lules tumorals?

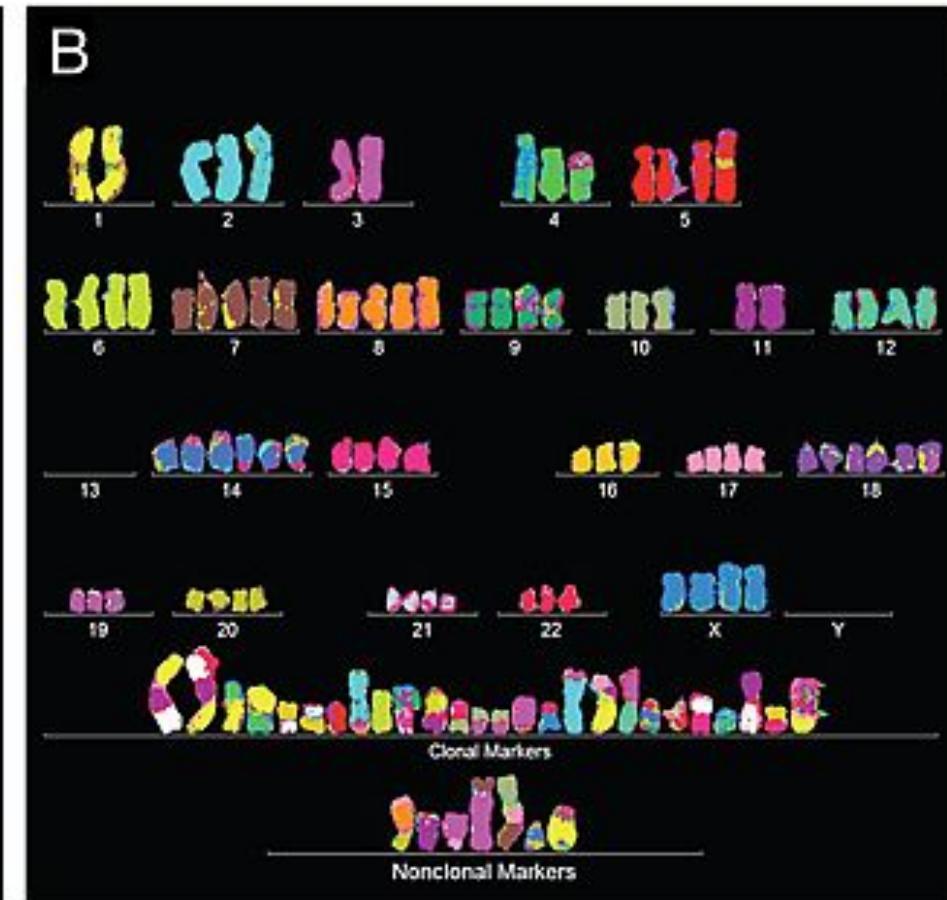
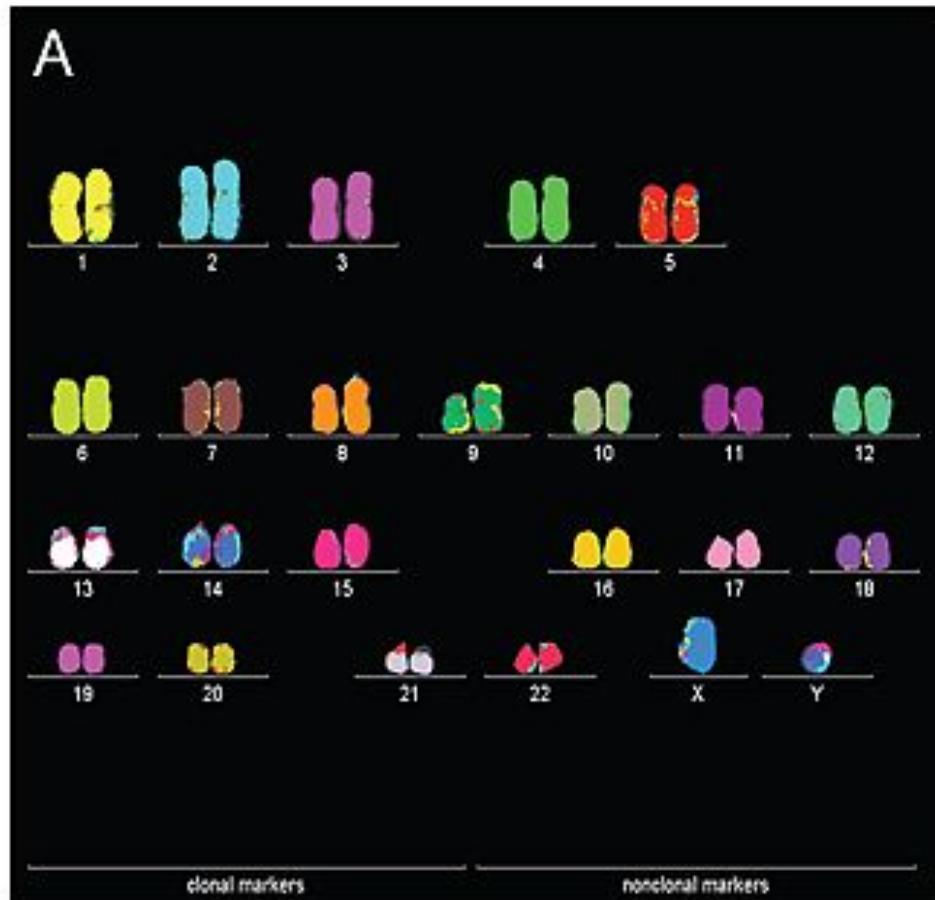


**Figure 1. The Hallmarks of Cancer**

This illustration encompasses the six hallmark capabilities originally proposed in our 2000 perspective. The past decade has witnessed remarkable progress toward understanding the mechanistic underpinnings of each hallmark.

Hanahan & Weinberg, Cell, 2011

# Les cèl·lules tumorals poden acumular MOLTES mutacions



Cèl·lula sana

Cèl·lula tumoral

# Resultats enquesta: Què és un tumor?

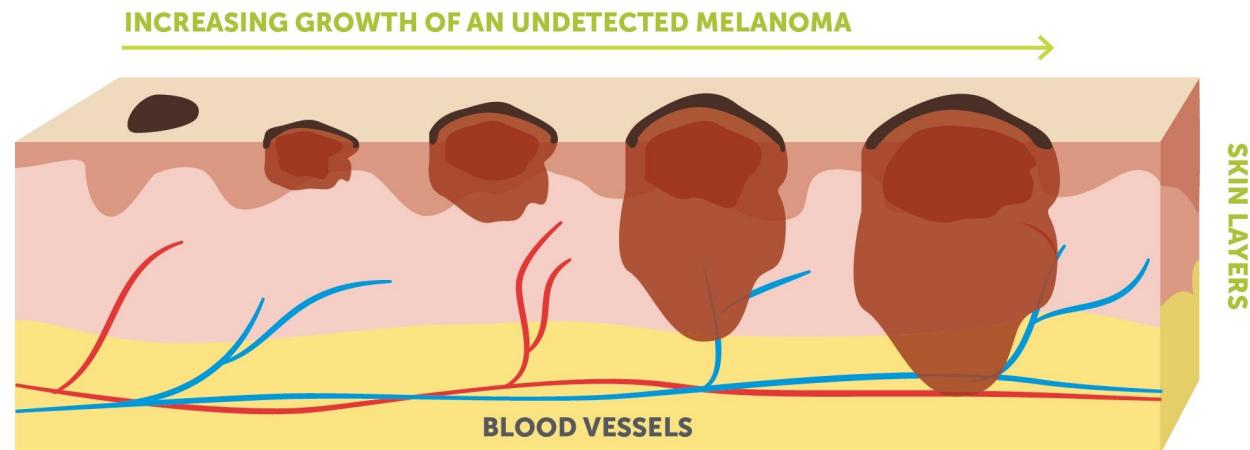
TUMOR
Acumulación excesiva y maligna de células.
Un tumor es un aumento de volumen de una parte del cuerpo.
Mutación de diversas células.
Un tumor es un tejido de células que se multiplican más de lo debido, sin ninguna función en el cuerpo.
Alteración de un tejido que hace crecer su volumen, lo "infla".
Producto de la multiplicación excesiva de las células.
Alteración de algún tejido del cuerpo que aumente de volumen no obstante puede ser <b>maligno o benigno</b>

# Melanoma

El melanoma és un tipus poc freqüent de càncer de pell

Cada any es diagnostiquen 160.000 casos en tot el món.

A Espanya moren unes 710 persones cada any per melanoma (0.8% de les morts per càncer).



Font: AECC

# Gens afectats en el melanoma

Gene type	Gene	Percent of malignant melanoma patient samples with mutation
Proto-oncogenes	<i>BRAF</i>	50-70% mutated
	<i>NRAS</i>	15-30% mutated
	<i>AKT3</i>	Up to 70% overexpressed
	<i>Cyclin D1</i>	6-44% amplified
Tumour suppressors	<i>CDKN2A</i>	30-70% deleted, mutated or silenced
	<i>PTEN</i>	5-20% deleted or mutations
	<i>p53</i>	10% lost or mutated

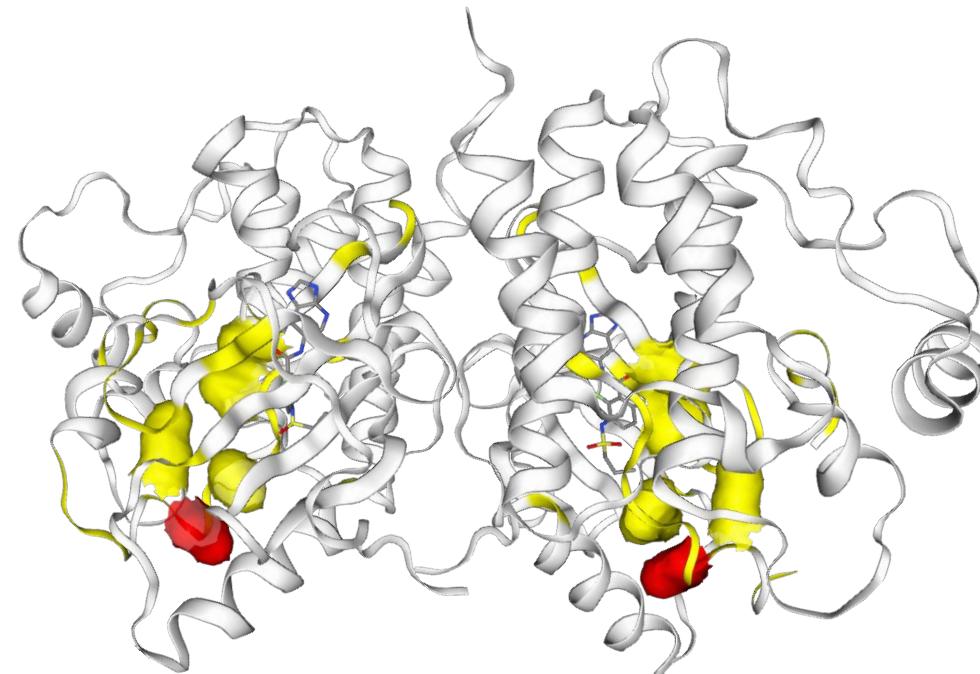
Adaptat de Gray-Schopfer V, Wellbrock C and Marais R. Melanoma biology and new targeted therapy. *Nature* 2007; 445: 85

# Activitat 5:



1. Entreu a la pàgina web de COSMIC ( <https://cancer.sanger.ac.uk/cosmic> )
2. Cerqueu BRAF i feu clic a l'entrada de gen (primer)
3. Quins fàrmacs es veuen afectats per mutacions en el BRAF? (veure dades de sensibilitat al fàrmac (“Drug sensitivity”))
4. Quins gens mutats confereixen resistència als fàrmacs? (veure resistència a les drogues)
5. Aneu a la pestanya de teixits per veure els teixits en què s'han detectat les mutacions. Quin teixit té el percentatge més elevat de mostres amb mutacions en l'BRAF?
6. Feu clic a la pestanya de distribució de mutacions. De quin tipus són la majoria de les mutacions?
7. Quin tipus de mutació porta a la mutació V600E?

# Mutacions que afecten BRAF

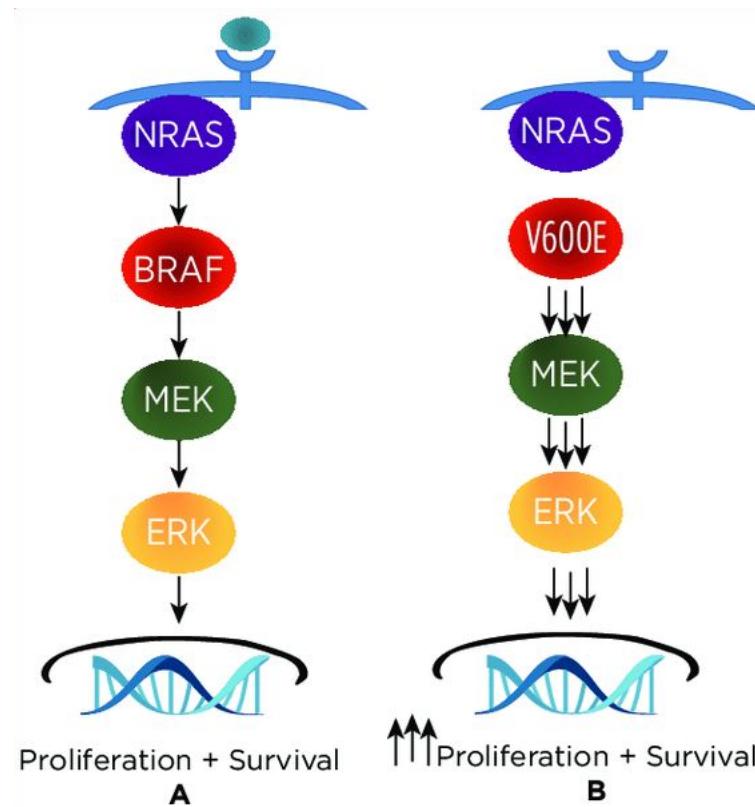


Freqüència de la mutació

1

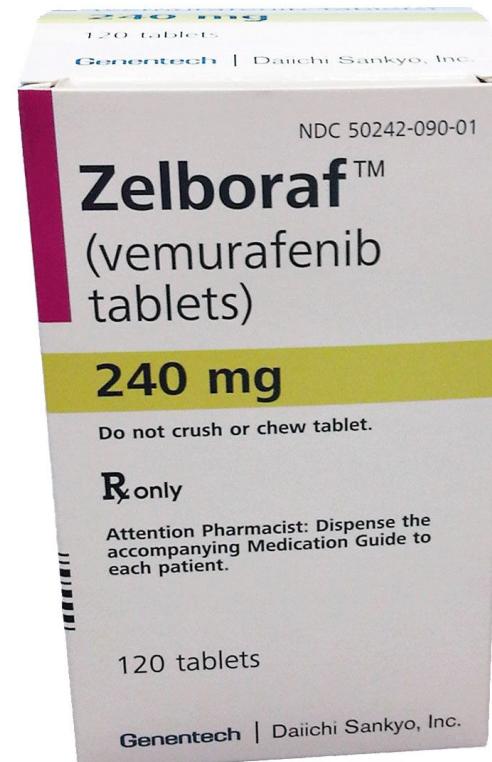
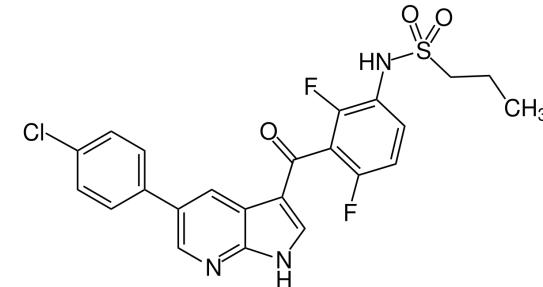
51497

# La mutació V600E en el gen BRAF

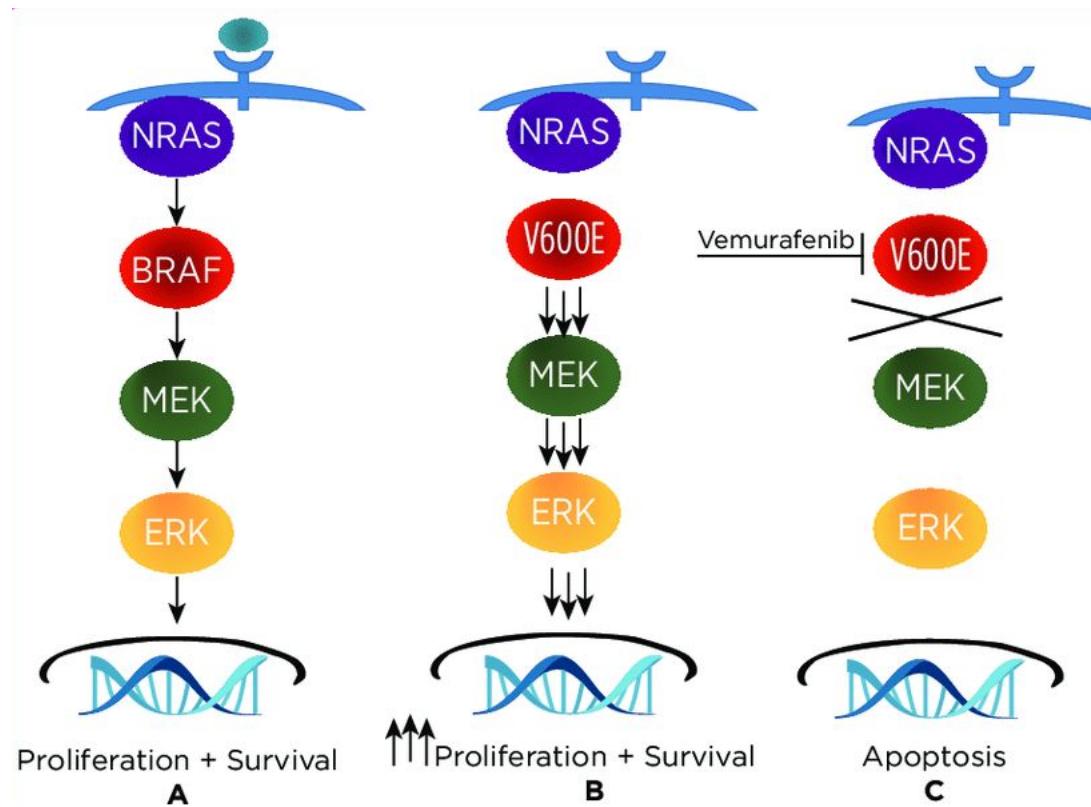


# Vemurafenib

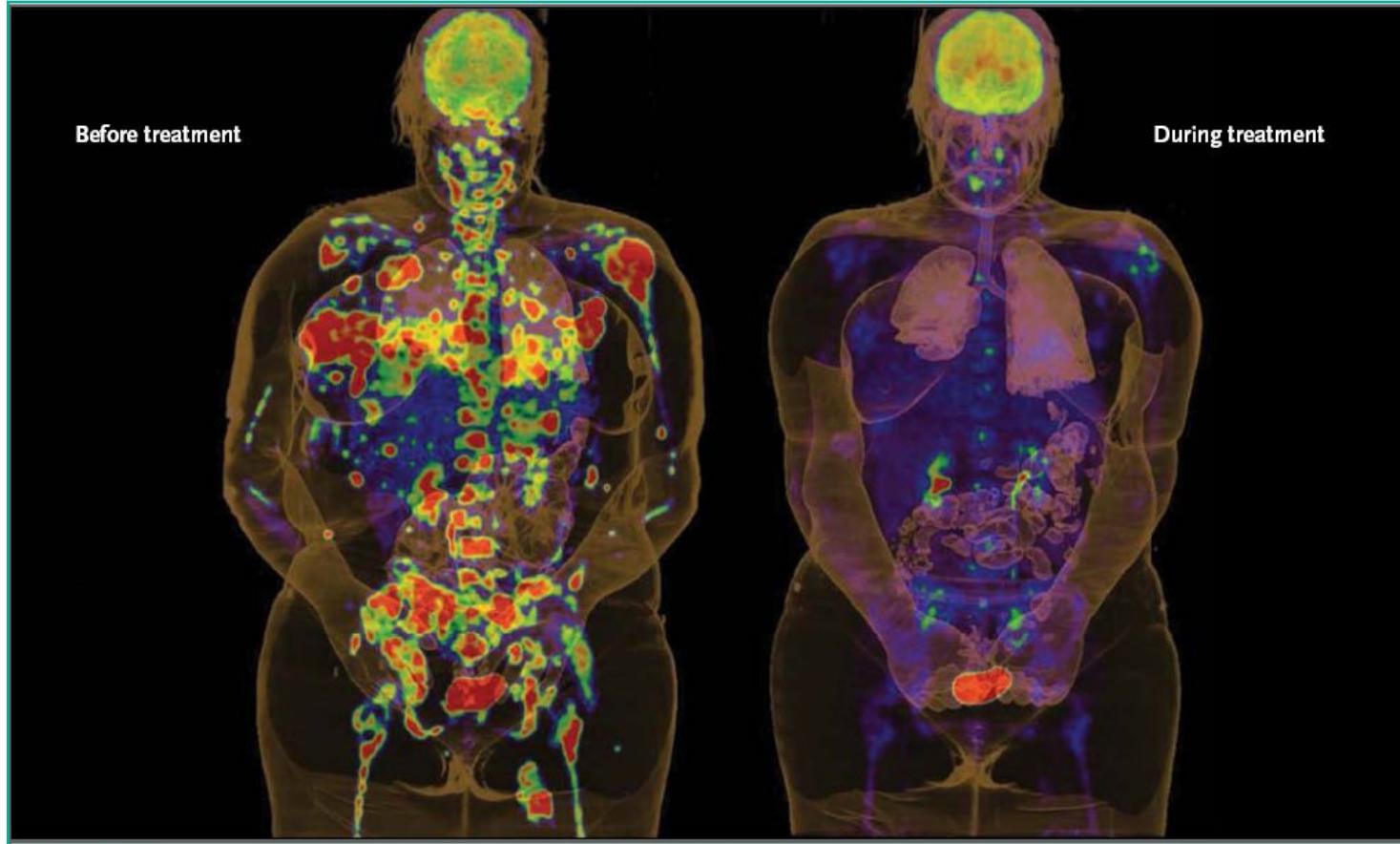
- El vemurafenib és un inhibidor de BRAF que s'utilitza per al tractament del melanoma maligne
- Ataca específicament les cèl·lules amb la mutació V600E en BRAF



# Vemurafenib



# Vemurafenib



Images courtesy of Grant McArthur, Jason Callahan, and Rod Hicks of the Peter MacCallum Cancer Centre.  
*McDermott, Downing and Stratton. N Engl J Med 2011;364:340-50.*

# Activitat 6: Què hem après fins ara?

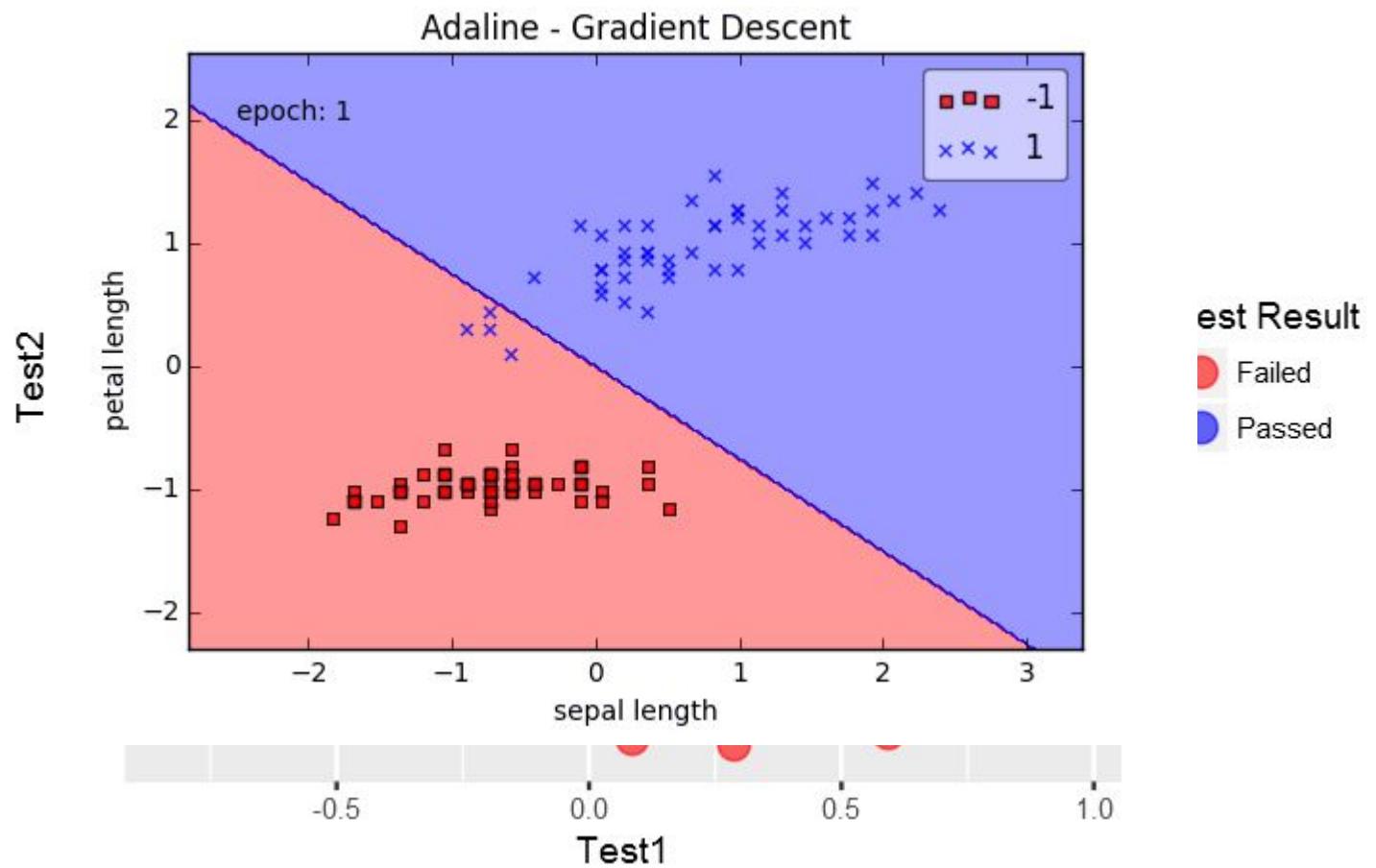


# Activitat 7: Machine learning

- Podemos predecir qué pacientes sobrevivirán al cáncer?. Más concretamente, a cáncer de melanoma.
- Usamos machine learning para realizar las predicciones.

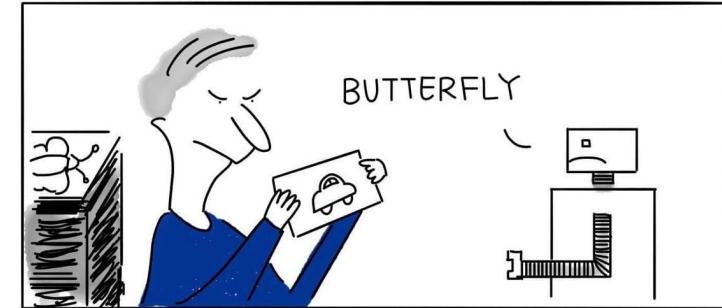
# Activity 7: Machine learning

- Problemas de clasificación



# Activity 7: Machine learning

- Qué es machine learning?



# Activity 7: Machine learning

- Ejemplo práctico: nuestro propio machine learning!
  - Podemos predecir qué pacientes sobrevivirán al cáncer?. Más concretamente, a cáncer de melanoma.
  - Tenemos datos sobre mutaciones, tratamientos, edad, IMC, ...

# Activity 7: Machine Learning

Estudio

Clinical Trials: Immunotherapy

## Genomic Features of Exceptional Response in Vemurafenib ± Cobimetinib–treated Patients with *BRAF*<sup>V600</sup>-mutated Metastatic Melanoma

Yibing Yan, Matthew J. Wongchenko, Caroline Robert, James Larkin, Paolo A. Ascierto, Brigitte Dréno, Michele Maio, Claus Garbe, Paul B. Chapman, Jeffrey A. Sosman, Zhen Shi, Hartmut Koeppen, Jessie J. Hsu, Ilsung Chang, Ivor Caro, Isabelle Rooney, Grant A. McArthur, and Antoni Ribas

DOI: 10.1158/1078-0432.CCR-18-0720 Published June 2019 

Link artículo: <https://clincancerres.aacrjournals.org/content/25/11/3239.figures-only>

# Activity 7: Machine Learning

## Datos

131  
PACIENTES

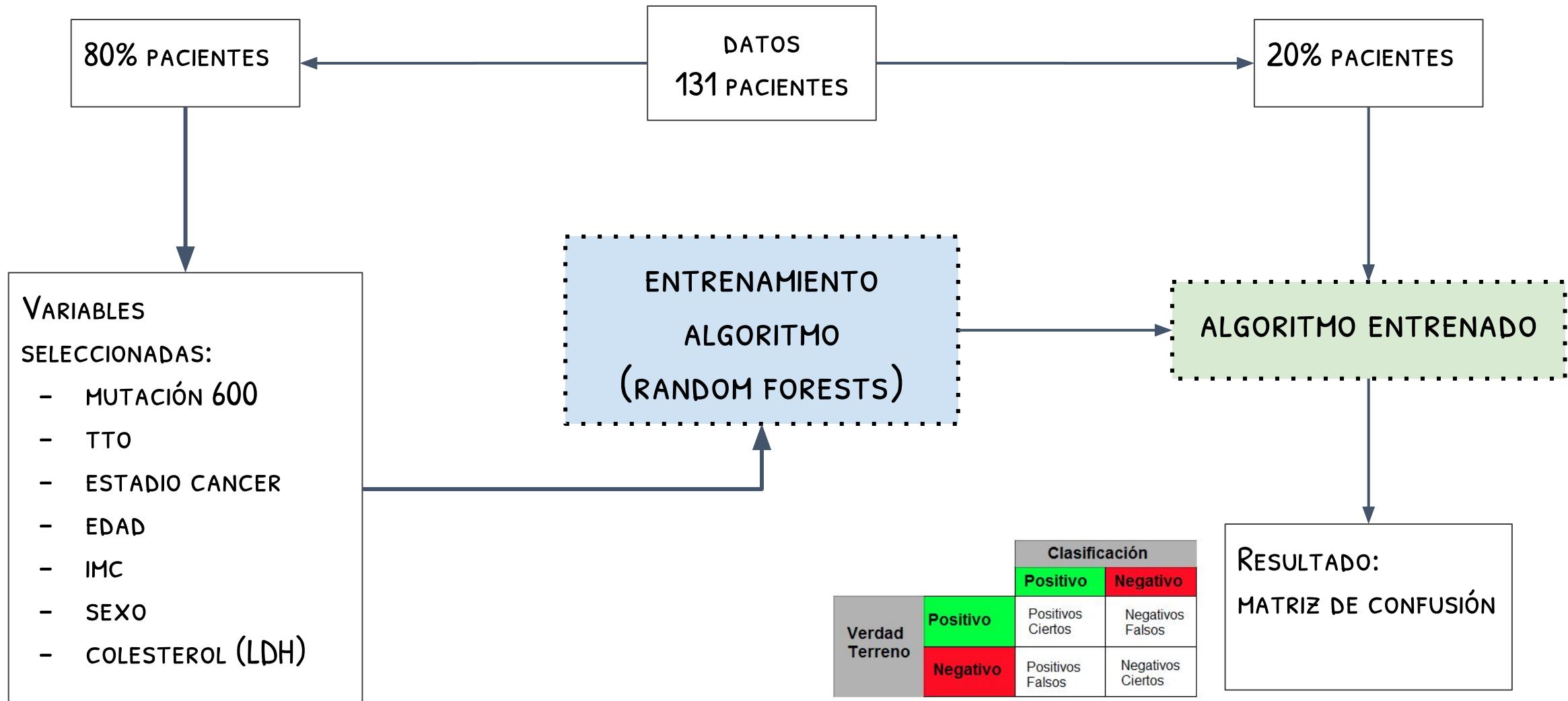
2 TTOS:  
● VEMURAFENIB  
● VEMURAFENIB +  
COBIMETINIB

2 PRONÓSTICOS:  
CR : COMPLETE RESPONSE  
PD : PROGRESSIVE DISEASE

Link para descargar datos:

[https://clincancerres.aacrjournals.org/highwire/filestream/171592/field\\_highwire\\_adjunct\\_files/1/198021\\_2\\_supp\\_5455023\\_pps\\_zzf.csv](https://clincancerres.aacrjournals.org/highwire/filestream/171592/field_highwire_adjunct_files/1/198021_2_supp_5455023_pps_zzf.csv)

# Activity 7: Machine Learning



# Activity 7: Machine Learning

Script

<https://rstudio.cloud/project/385639>

Results

[https://rchaeologist.shinyapps.io/randomforest\\_skcm/](https://rchaeologist.shinyapps.io/randomforest_skcm/)

Link para descargar datos:

[https://clincancerres.aacrjournals.org/highwire/filestream/171592/field\\_highwire\\_adjunct\\_files/1/198021\\_2\\_sup\\_p\\_5455023\\_ppszzf.csv](https://clincancerres.aacrjournals.org/highwire/filestream/171592/field_highwire_adjunct_files/1/198021_2_sup_p_5455023_ppszzf.csv)

# I per què necessitem supercomputadors, els biòlegs?

Com de comú és fer els supercàlculs?

- **Analizar tot el genoma d'una vegada**

La seqüència que hem analitzat té 2519 nucleòtids.

Això representa un 0.00007% del genoma humà (~3.6 Gb)

- **Processar experiments en paral·lel**

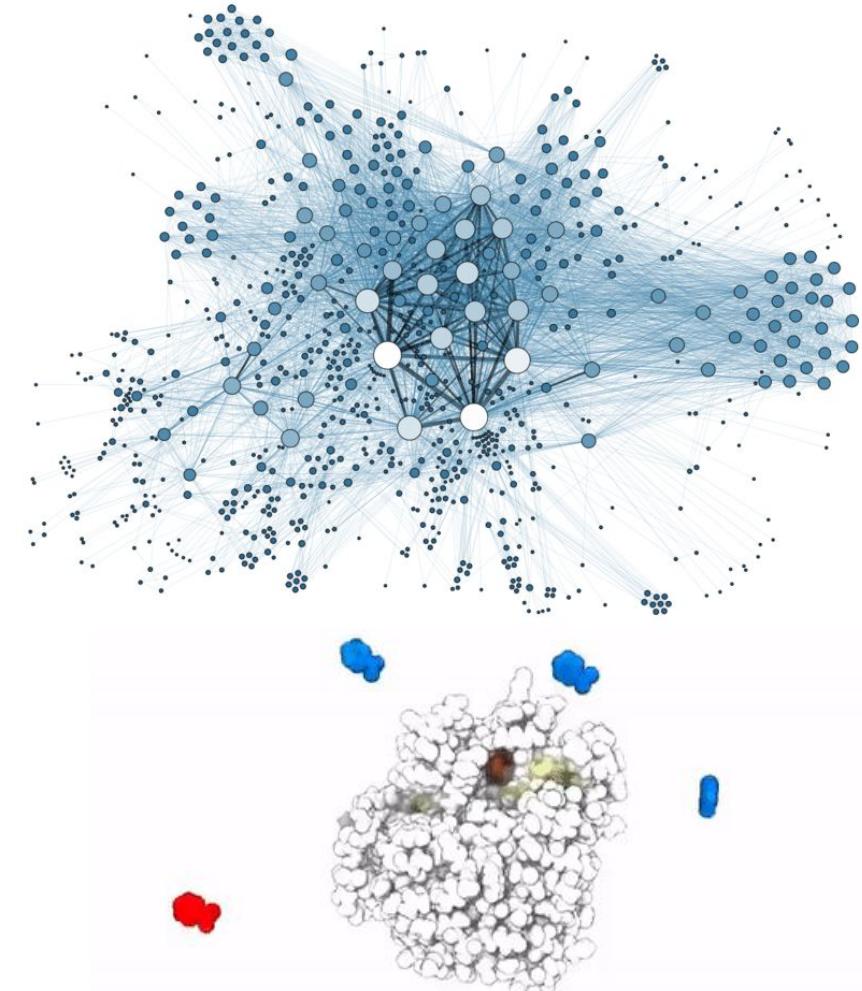
Mapejar el genoma de milers de pacients

- **Integrar una gran quantitat de dades**

Correlacionar mutacions en un milió de pacients

- **Testar hipòtesis "*in silico*"**

Simular l'efecte de un milió de medicaments

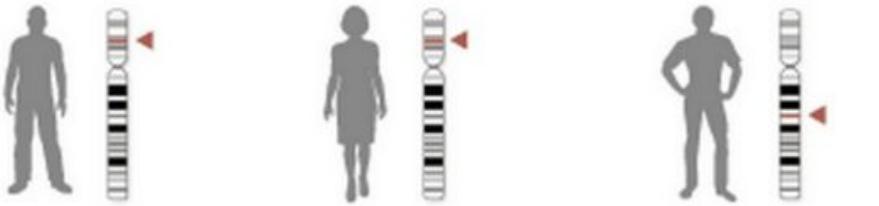


# I per què necessitem supercomputadors, els biòlegs?

Cluster/Account/User Utilization 2019-09-05T00:00:00 - 2019-09-05T23:59:59 (86400 secs)					
Usage reported in CPU Hours					
Cluster	Account	Login	Proper Name	Used	Energy
marenost+	root			3601601	2476360289
marenost+	bsc			673815	193261206
marenost+	life			56281	20777756
marenost+	bsc08			14277	1398527
marenost+	bsc08	bsc08786	Laureano Tomas	12680	832885

**12.680 hores = 211 dies**

# Cancer Genomics: Per a què?



Cancer-causing mutations with  
drug treatment available

Mutation with no  
drug available



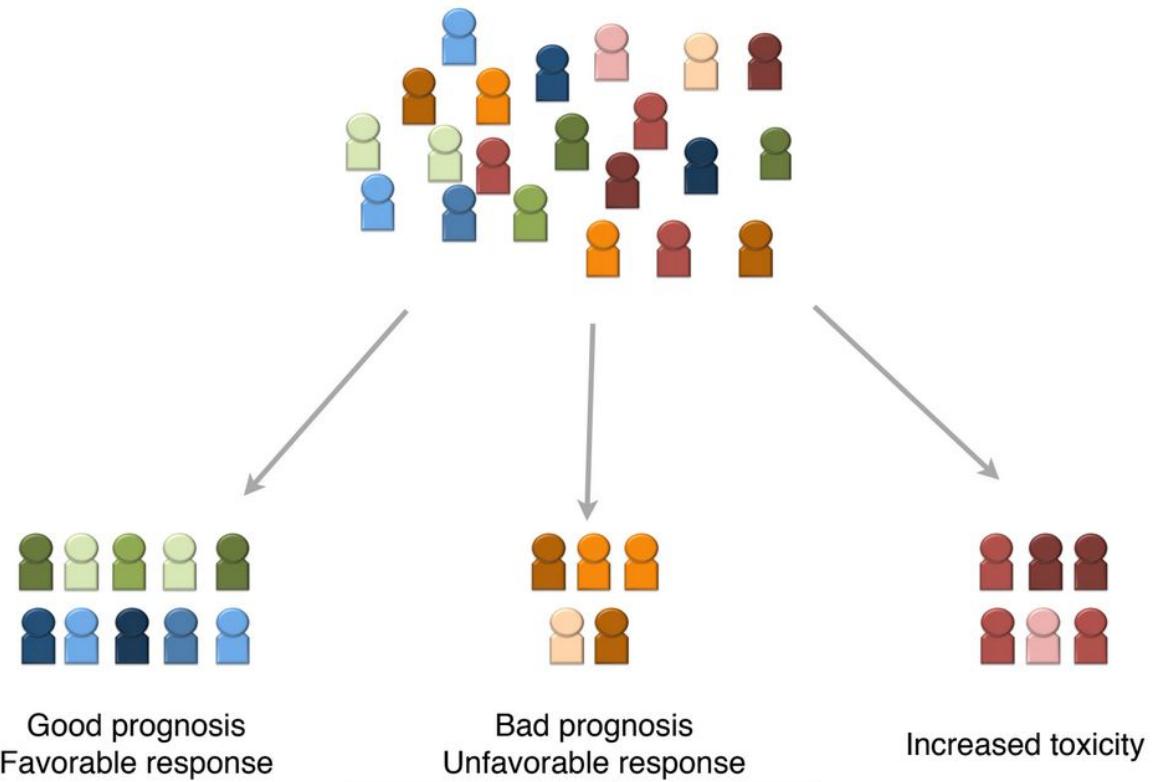
Cancer drug



Radiation and  
chemotherapy

- Trobar nous gens de càncer.
- Trobar noves teràpies dirigides

# Cancer Genomics: Per a què?



- Identificar signatures moleculars per estratificar tumors
- Avançar cap al tractament personalitzat del càncer