

**Concetti di base della genomica.  
Ereditarietà semplice e complessa.  
Concetto di variabilità genetica**



**Veronica Mariotti**

Decifrare le basi biologiche e genetiche del comportamento è uno degli obiettivi principali dell'attuale ricerca neurobiologica e psichiatrica.

Genetics of Behavior



NEWS

## Parsing the Genetics of Behavior

It takes more than one gene, or even a few genes, to make a personality trait. But which ones?

7 NOVEMBER 2008 VOL 322 SCIENCE

## Qualche nozione di biologia di base

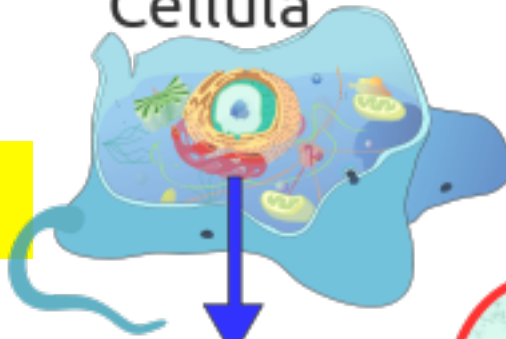


**L' organismo umano è costituito da milioni di miliardi di cellule**

## Qualche nozione di biologia di base

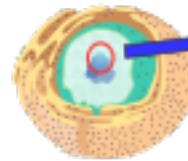


Cellula

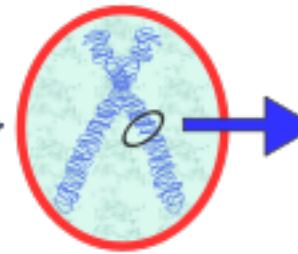


L'organismo umano è costituito  
Da milioni di miliardi di cellule

Ogni cellula è dotata di un nucleo  
che contiene il materiale genetico  
ovvero il DNA

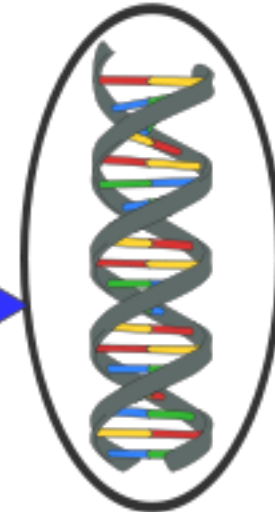


Nucleo



Cromosoma

DNA



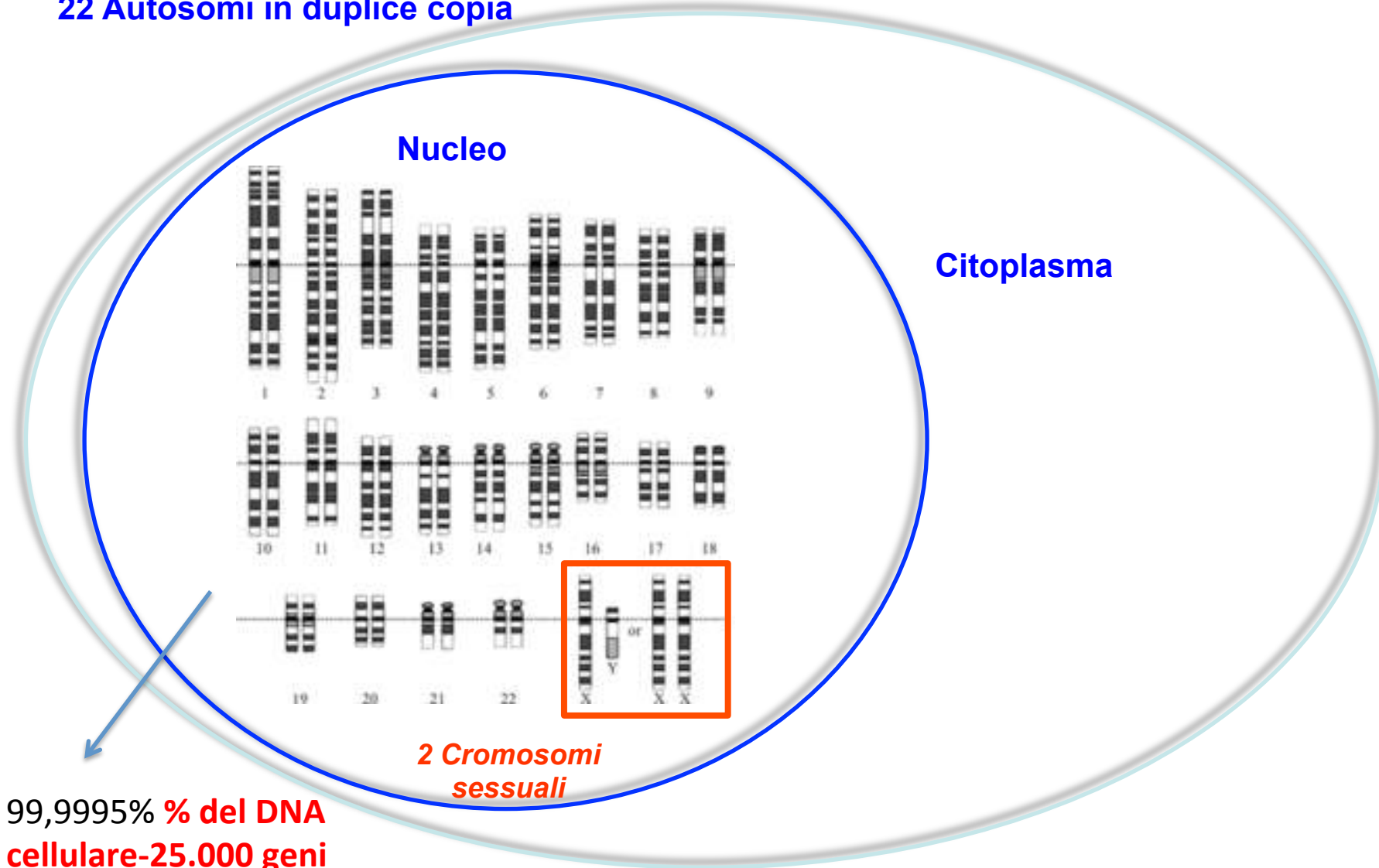
Il DNA nel nucleo è  
organizzato in cromosomi



# DNA nucleare

23 Coppie di Cromosomi

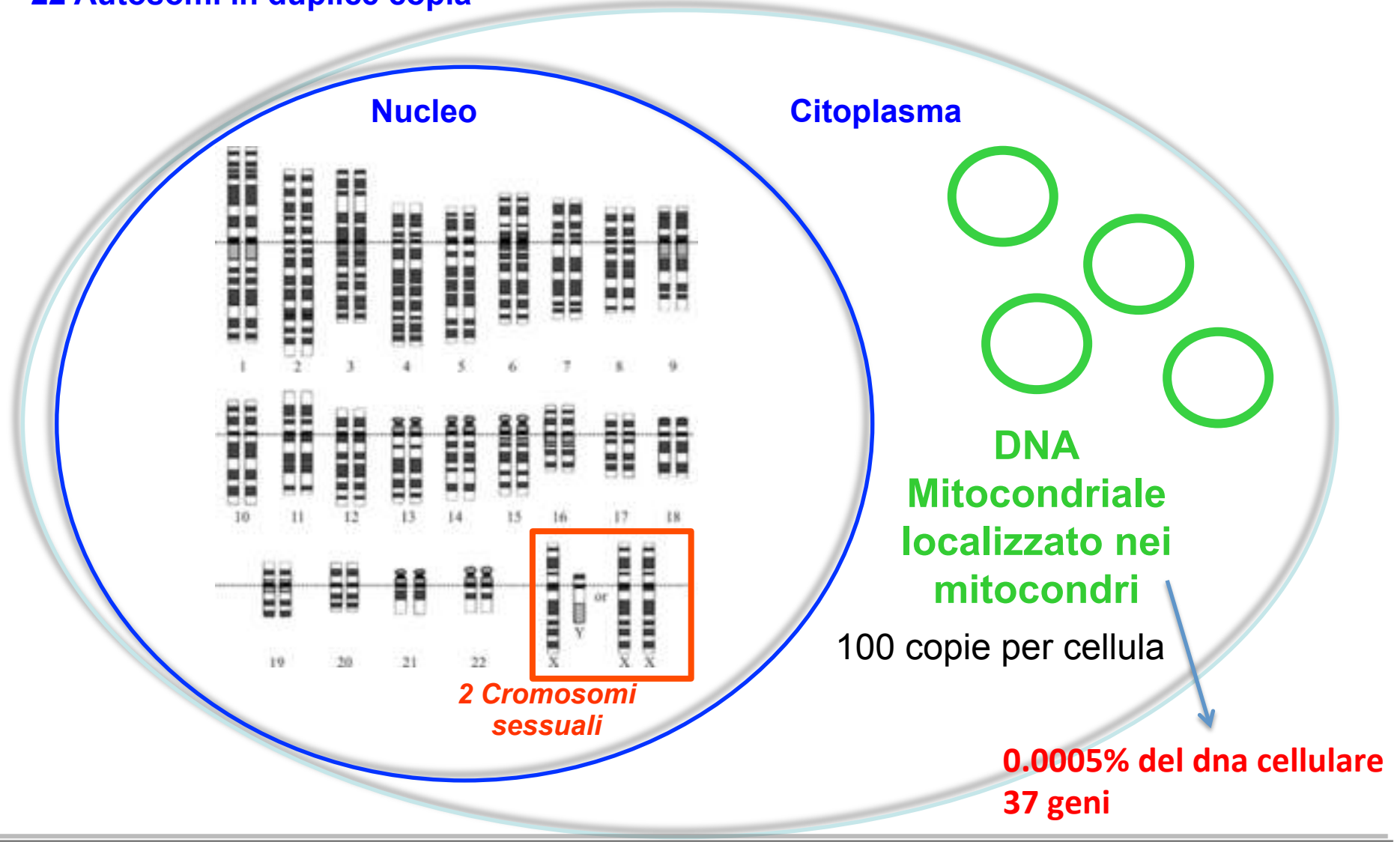
22 Autosomi in duplice copia



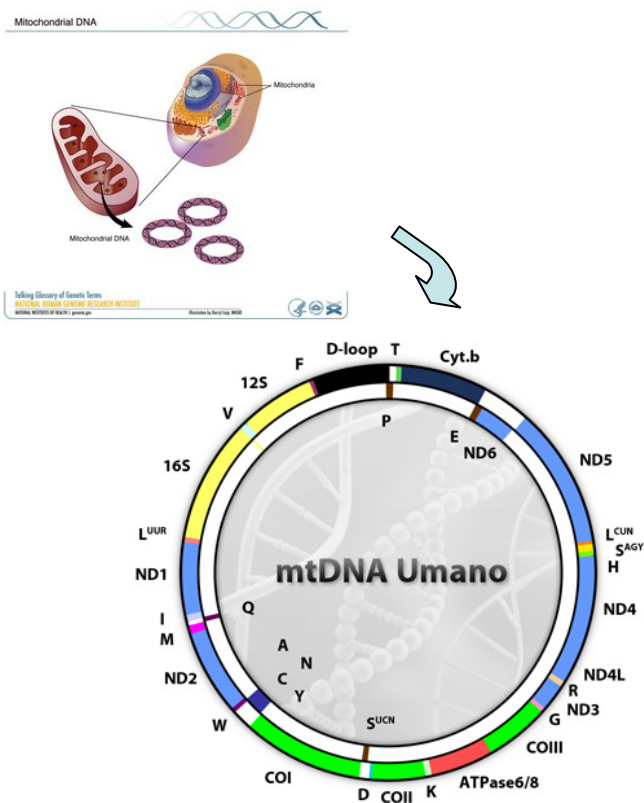
# DNA umano totale

nucleare + mitocondriale

22 Autosomi in duplice copia

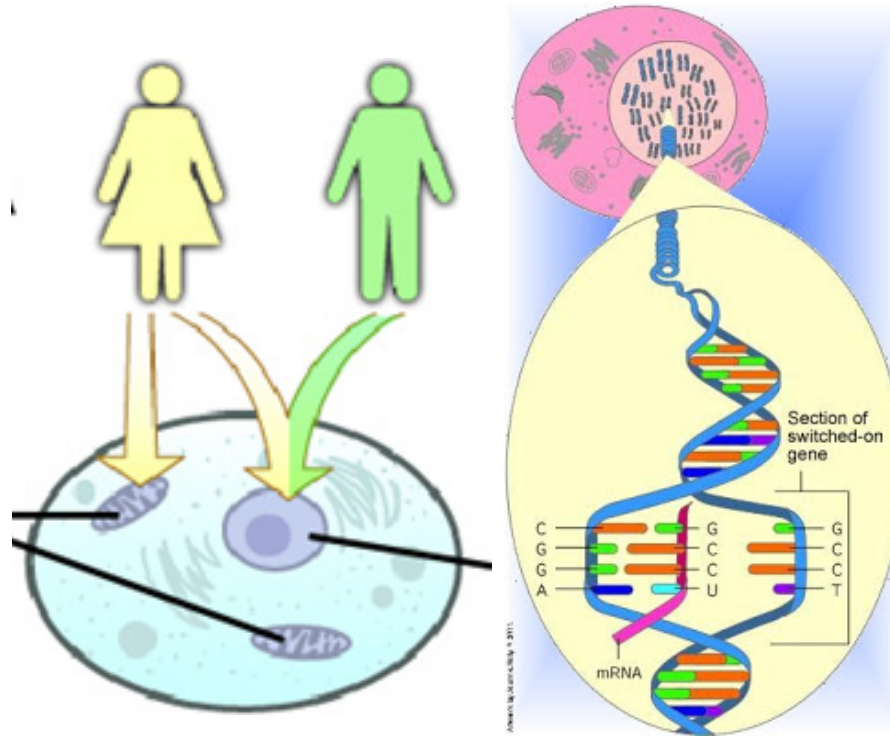


# DNA MITOCONDRIALE



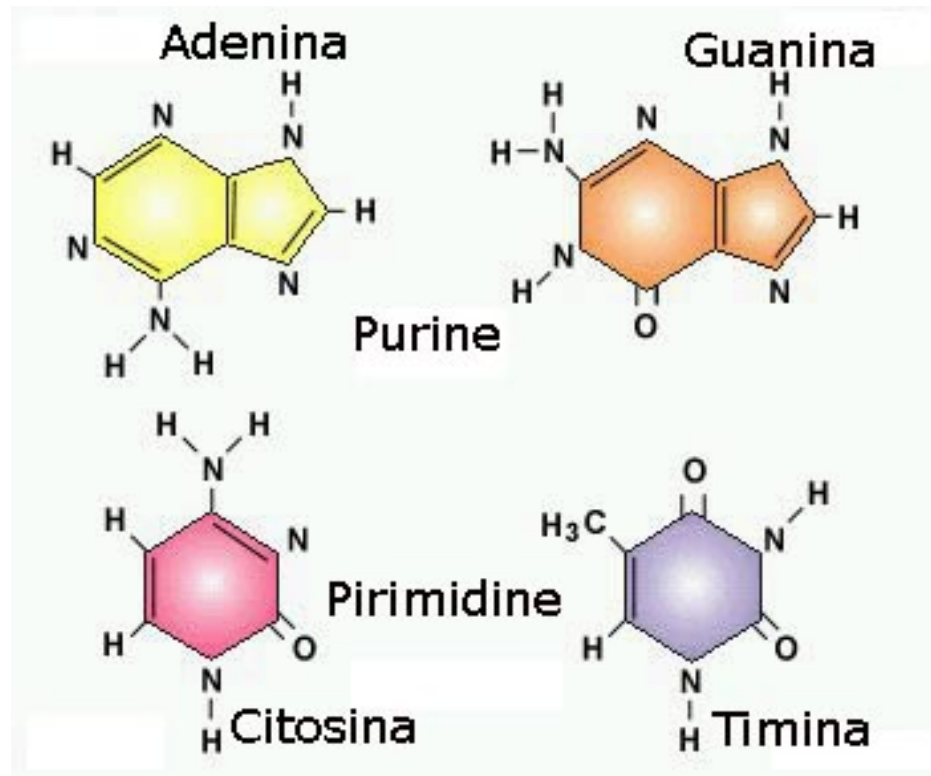
- Ereditato solo dalla madre

# DNA NUCLEARE



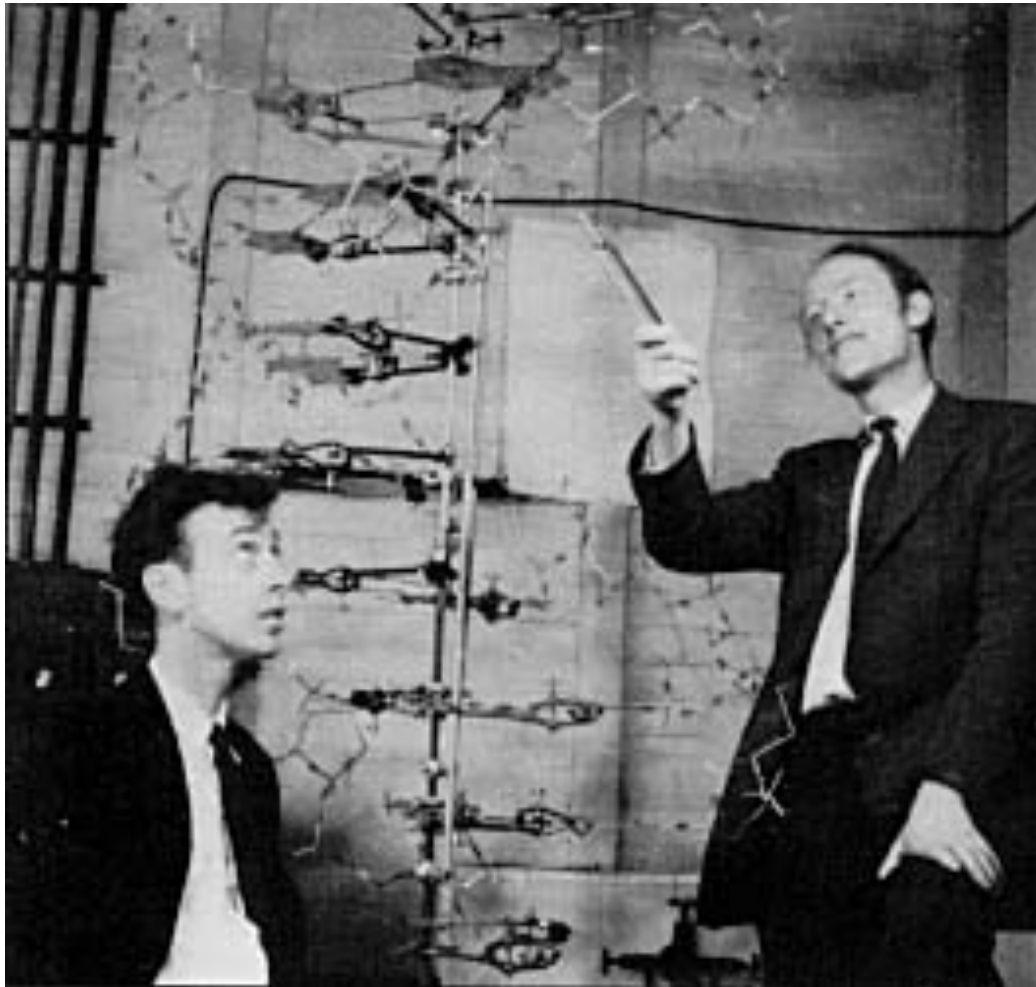
- Ereditato da entrambi i genitori

# mattoni del DNA : i nucleotidi





# 1953: modello del DNA secondo Watson e Crick



James D. Watson and Francis H. C. Crick  
 "Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid" (1953)

No. 4356 April 25, 1953 NATURE 737

equipment, and to Dr. G. E. R. Deacon and the captain and officers of R.R.S. *Discovery II* for their part in making the observations.

<sup>1</sup>Young, F. B., Gerard, H., and Jevons, W., *Phil. Mag.*, **40**, 149 (1925).

<sup>2</sup>Longuet-Higgins, M. S., *Mon. Not. Roy. Astr. Soc., Geophys. Supp.*, **5**, 285 (1949).

<sup>3</sup>Van Arte, W. S., Woods Hole Papers in Phys. Oceanog. Meteor., **11** (5) (1950).

<sup>4</sup>Ikman, V. W., *Arkiv. Mat. Astron. Fysik. (Stockholm)*, **2** (11) (1935).

### MOLECULAR STRUCTURE OF NUCLEIC ACIDS

#### A Structure for Deoxyribose Nucleic Acid

WE wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey.<sup>1</sup> They kindly made their manuscript available to us in advance of publication. Their model consists of three inter-twined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate di-ester groups joining  $\beta$ -D-deoxyribofuranose residues with 3',5' linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furburg's<sup>2</sup> model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furburg's 'standard configuration', the sugar being roughly perpendicular to the attached base. There is a residue on each chain every 3.4 Å. in the z-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 Å. The distance of a phosphate atom from the fibre axis is 10 Å. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purines and pyrimidines bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical z-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

It has been found experimentally<sup>3,4</sup> that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid.

It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

The previously published X-ray data<sup>5,6</sup> on deoxyribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereochemical arguments.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material. Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on inter-atomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co-workers at

This figure is purely diagrammatic. The two ribbons symbolize the two phosphate-sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis.

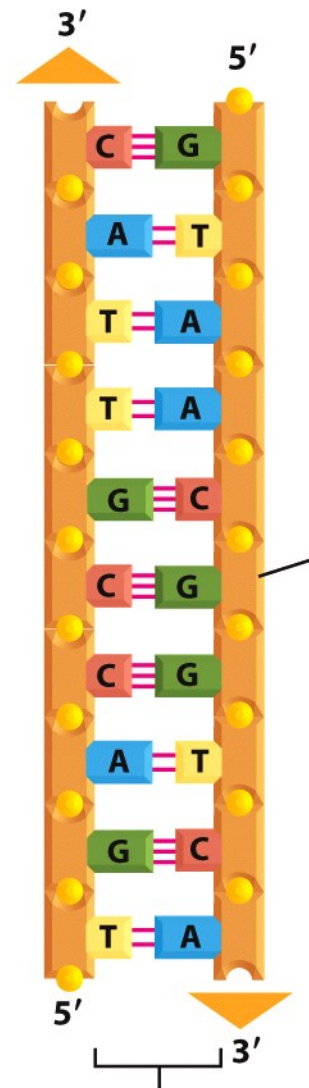
Copyright © 1995 Smithsonian Institution

Illustration reprinted with permission from *Nature* (171: 736-37). Copyright 1953. Macmillan Magazines Ltd; and with the permissions of James Watson and Francis Crick.

# Elementi chiave che definiscono la struttura del DNA

- ✚ **Elica a doppio filamento**
- ✚ **Di diametro uniforme**
- ✚ **Destrorsa**
- ✚ **Con i due filamenti che hanno direzione opposta (antiparalleli)**
- ✚ **Legami idrogeno tra T-A e C-G**

(C) double-stranded DNA



hydrogen-bonded base pairs

sugar-phosphate backbone

(D) DNA double helix

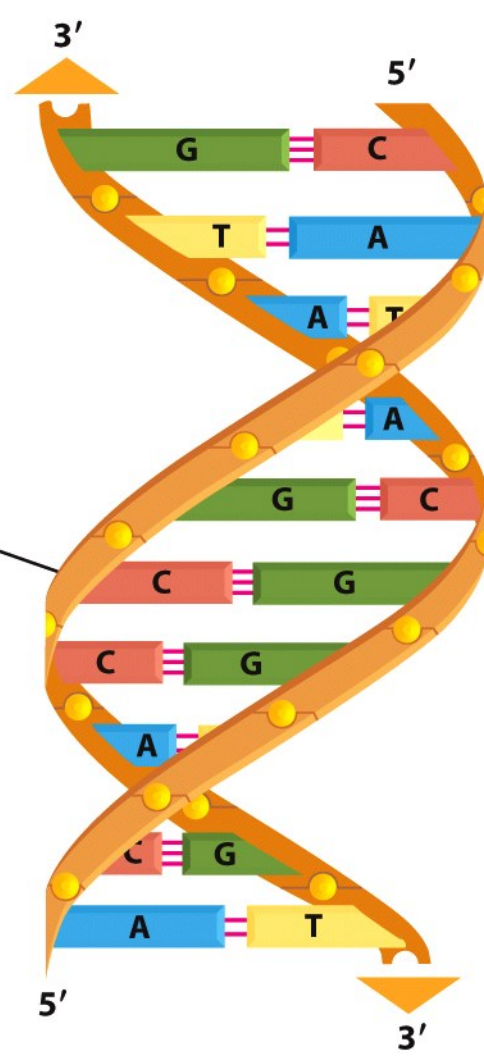
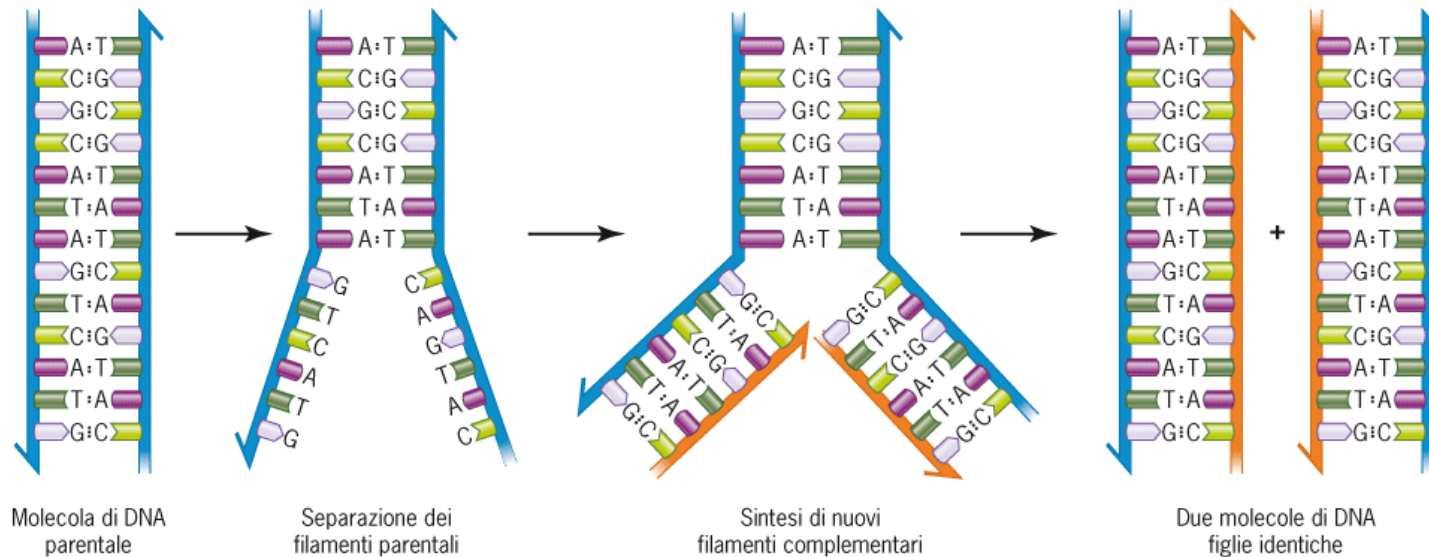


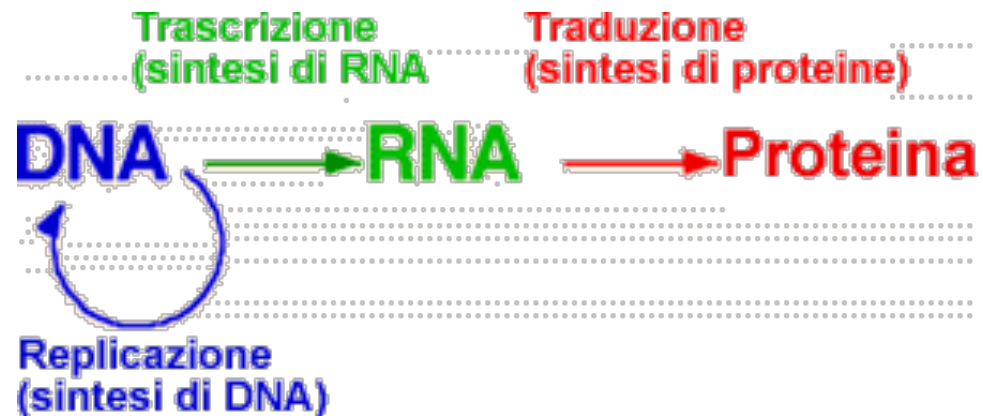
Figure 5-2c *Essential Cell Biology* (© Garland Science 2010)

## *Replicazione del DNA(semiconservativa)*



- ✚ apertura della doppia elica mediante rottura dei legami idrogeno
- ✚ ognuno dei due filamenti funziona da stampo per la sintesi di quelli complementari, al termine si ritrovano due nuove doppie eliche
- ✚ filamento stampo viene copiato per mezzo di complementarità delle basi
- ✚ i nucleotidi appropriati vengono inseriti nella catena nascente dalla DNA polimerasi

# Dogma centrale della biologia molecolare





# Il codice genetico permette di tradurre il messaggio in proteine

AGA										UUA					AGC						
AGG										UUG					AGU						
GCA	CGA						GGA			CUA				CCA	UCA	ACA				GUA	
GCC	CGC						GGC		AUA	CUC				CCC	UCC	ACC				GUC	UAA
GCG	CGG	GAC	AAC	UGC	GAA	CAA	GGG	CAC	AUC	CUG	AAA			CCG	UCG	ACG			UAC	GUG	UAG
GCU	CGU	GAU	AAU	UGU	GAG	CAG	GGU	CAU	AUU	CUU	AAG	AUG	UUU	CCU	UCU	ACU	UGG	UAU	GUU	GUA	UGA
Ala	Arg	Asp	Asn	Cys	Glu	Gln	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val	stop	
A	R	D	N	C	E	Q	G	H	I	L	K	M	F	P	S	T	W	Y	V		

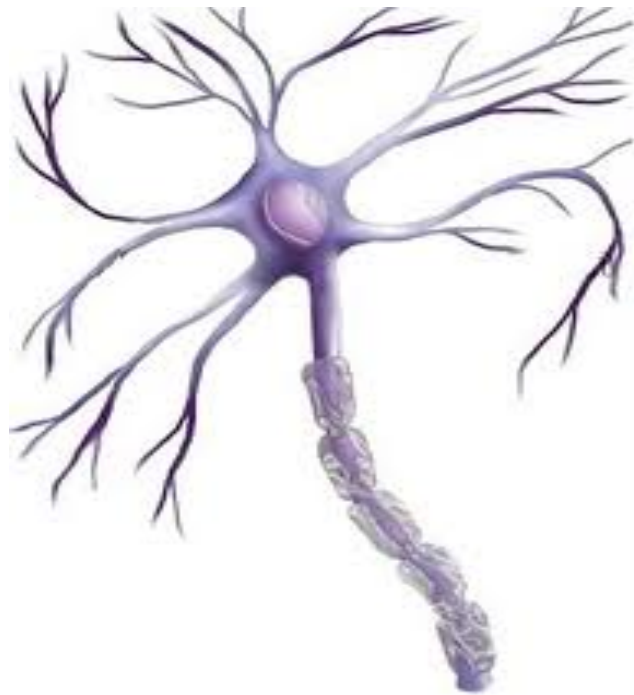


Sequenza nucleotidica  
Sequenza aminoacidica

GCAAGAGACAAC = DNA  
Ala Arg Asp Asn = proteina



Le cellule di uno stesso organismo pur avendo lo stesso DNA e quindi le stesse informazioni hanno differenti fenotipi ma le utilizzano in maniera differente .....

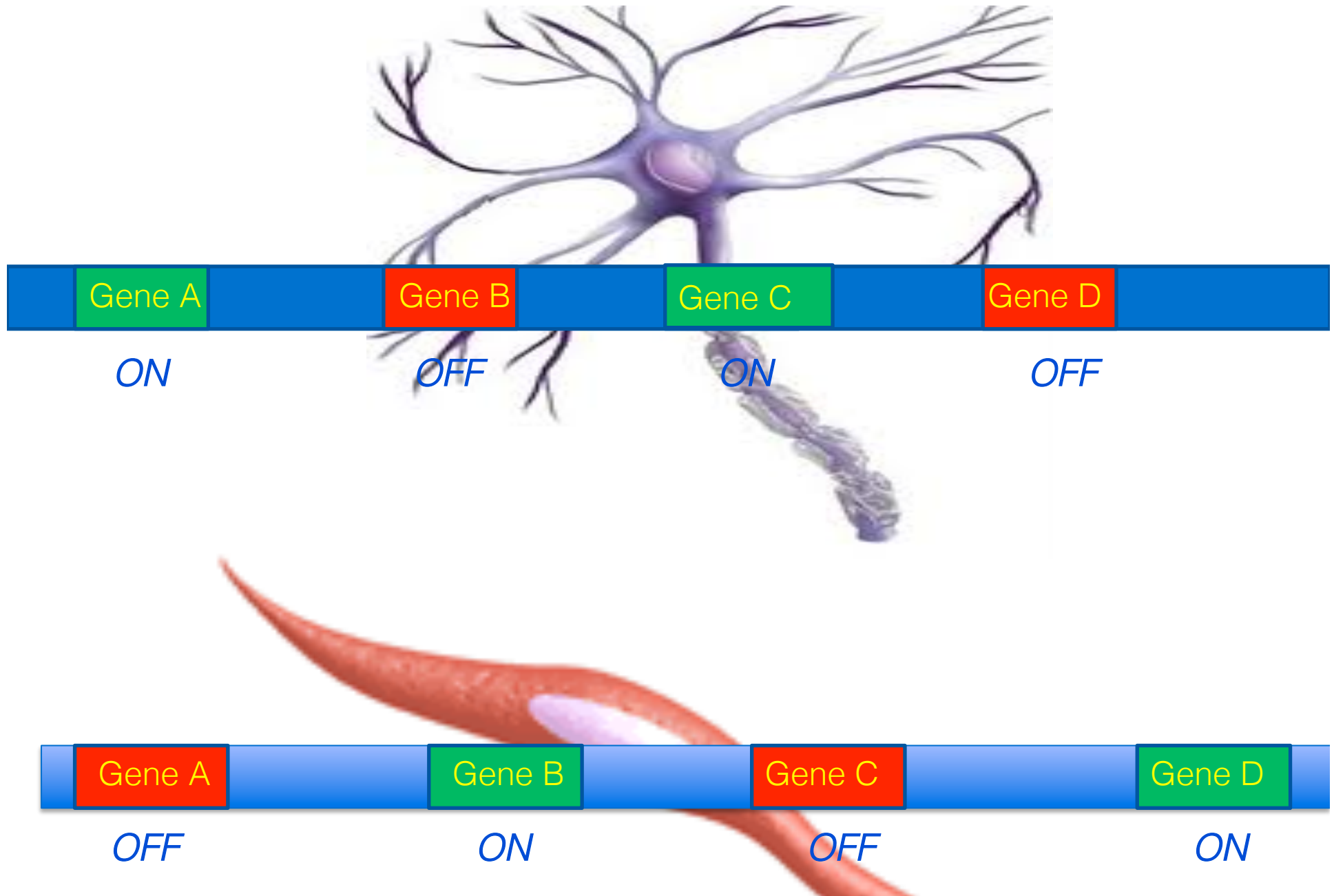


cellula nervosa



cellula muscolare

....differente profilo di espressione genica



**Il contenuto di  
informazioni del DNA  
umano è dato  
dall'alternanza di 4  
lettere**

**A, G, T, C**

# Progetto Genoma Umano

1990



stabilire con precisione  
come sono disposte una  
dopo l'altra le quattro basi  
azotate del DNA

2003



# Che cosa si conosce del genoma umano



3 miliardi di paia di basi



2% codificanti (60 milioni)

98% non codificanti

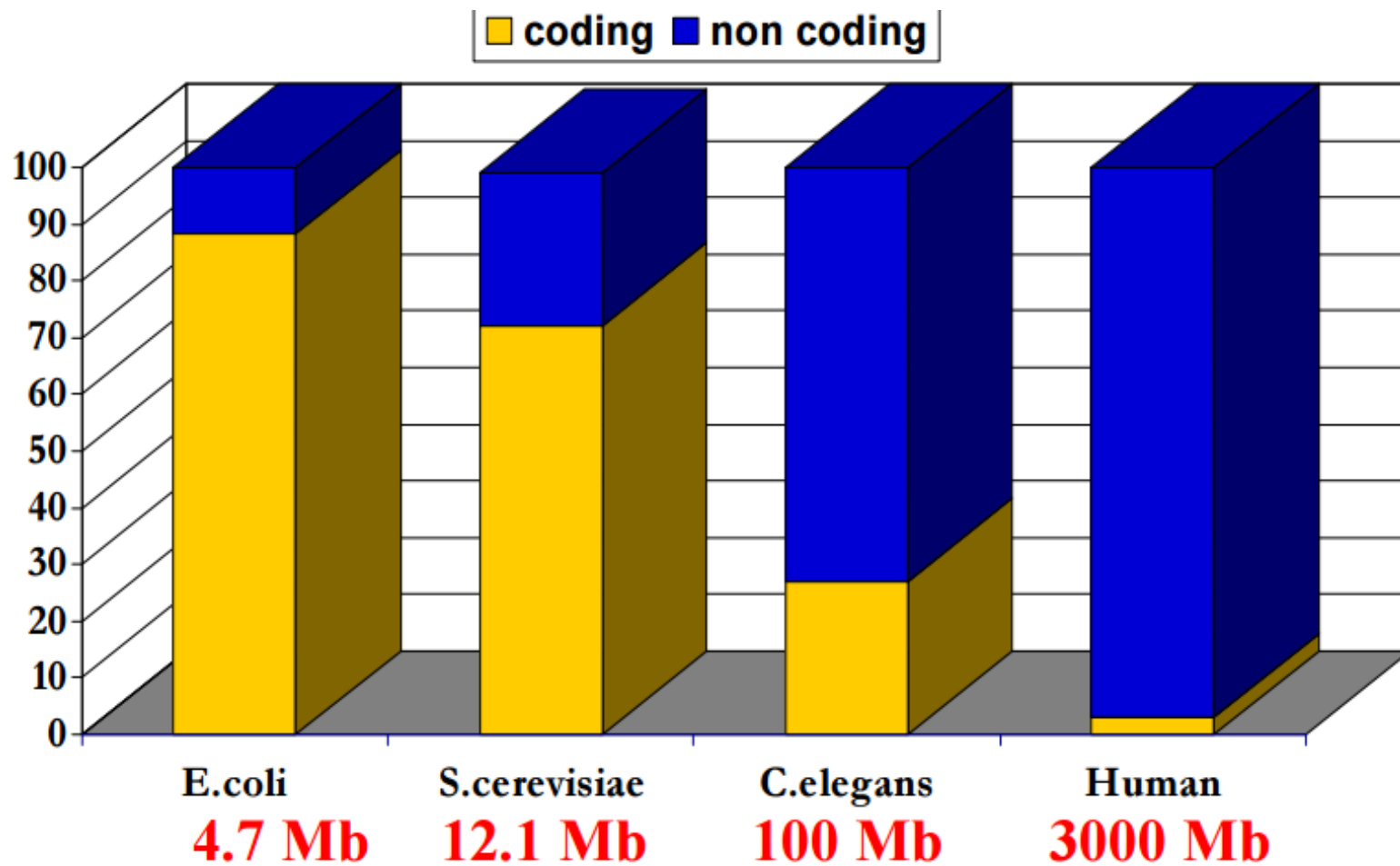


-esoni

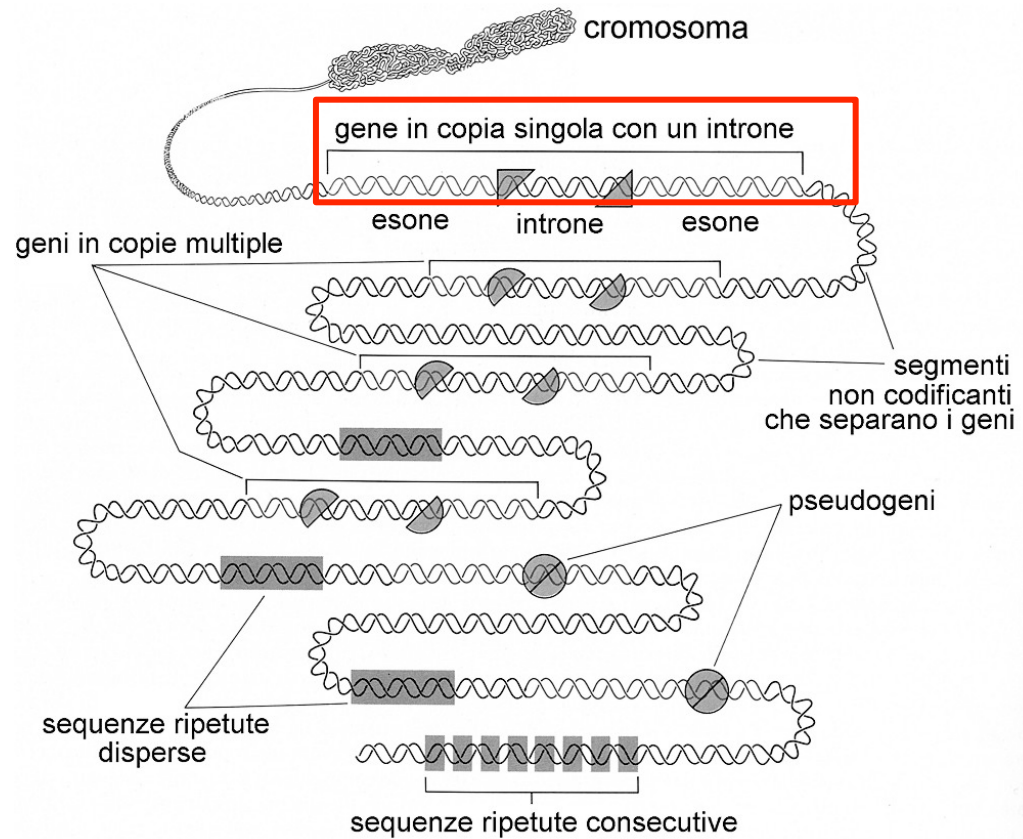


-introni  
-pseudogeni (geni ancestrali inattivi)  
-sequenze ripetute

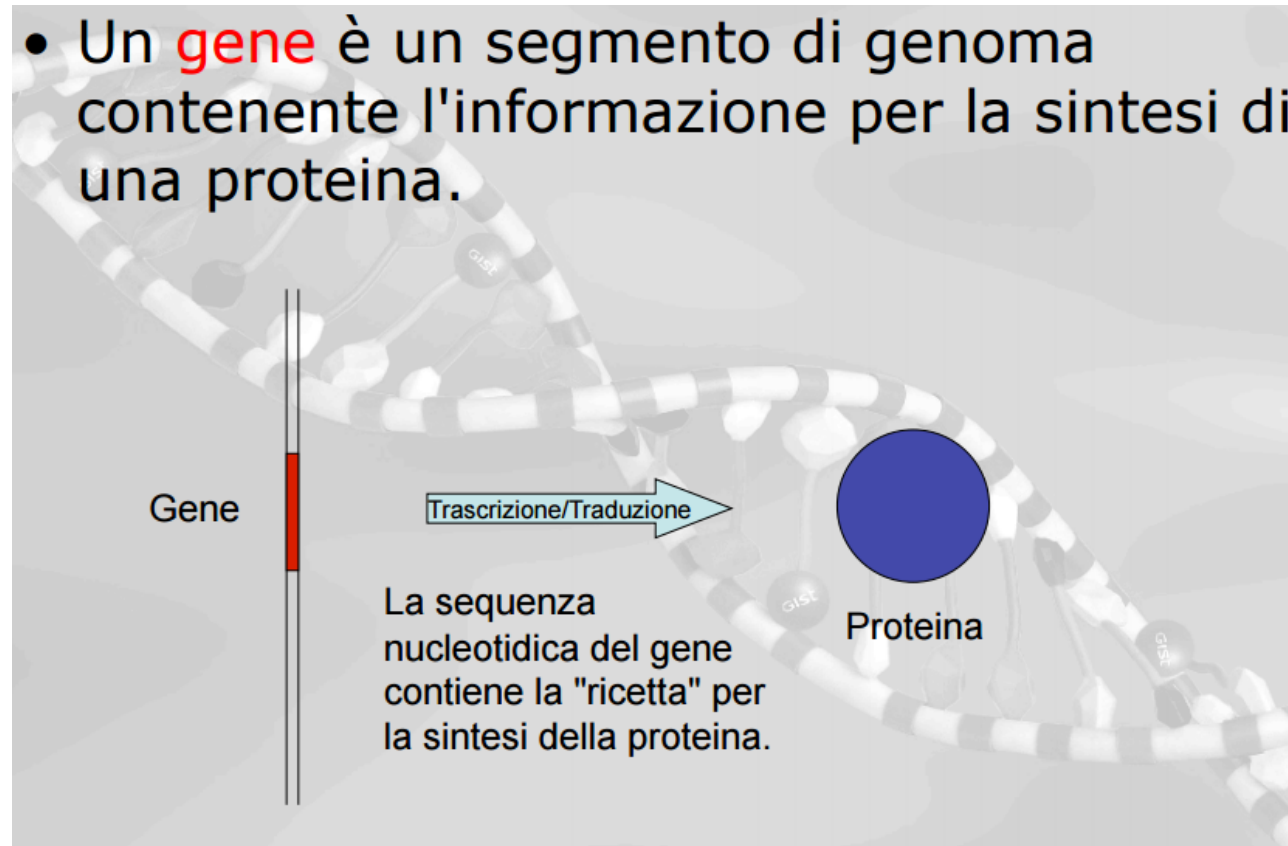
20.000 geni codificanti

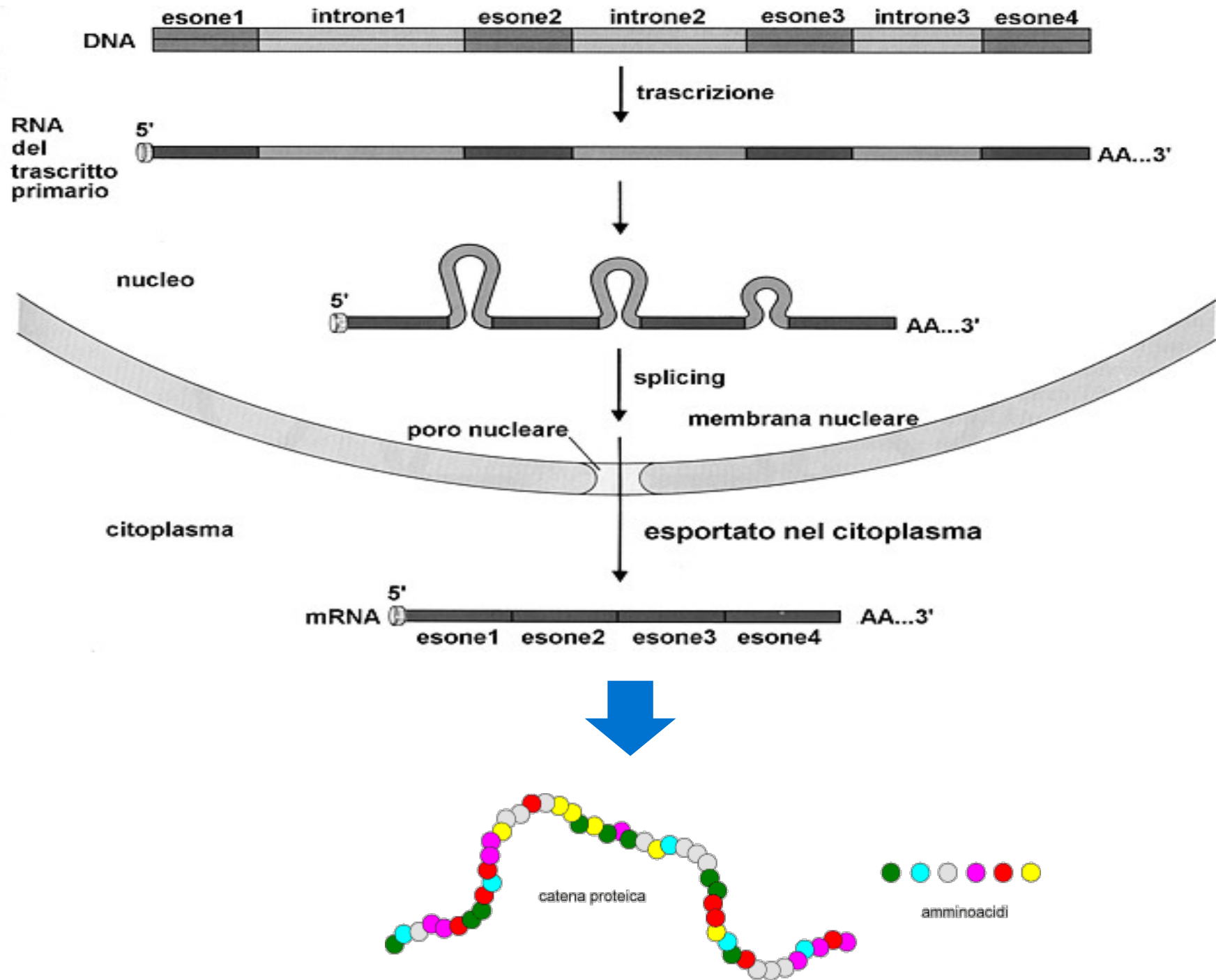


# Esoni: sequenze geniche codificanti



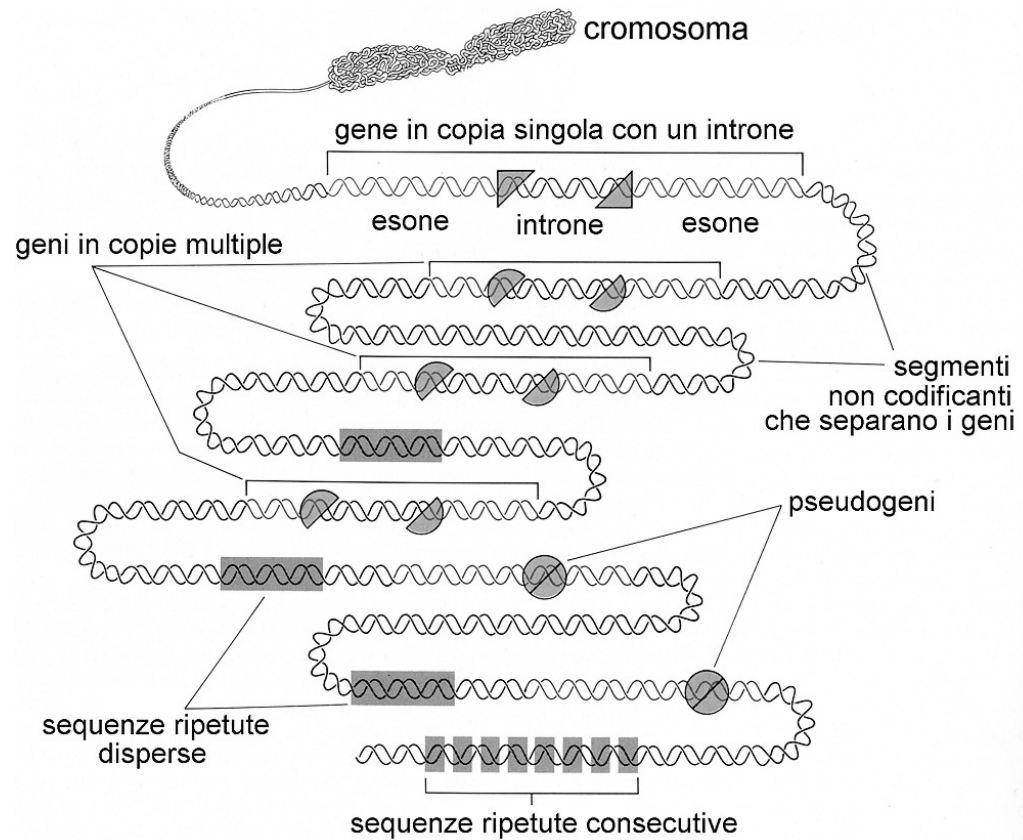
- Un **gene** è un segmento di genoma contenente l'informazione per la sintesi di una proteina.



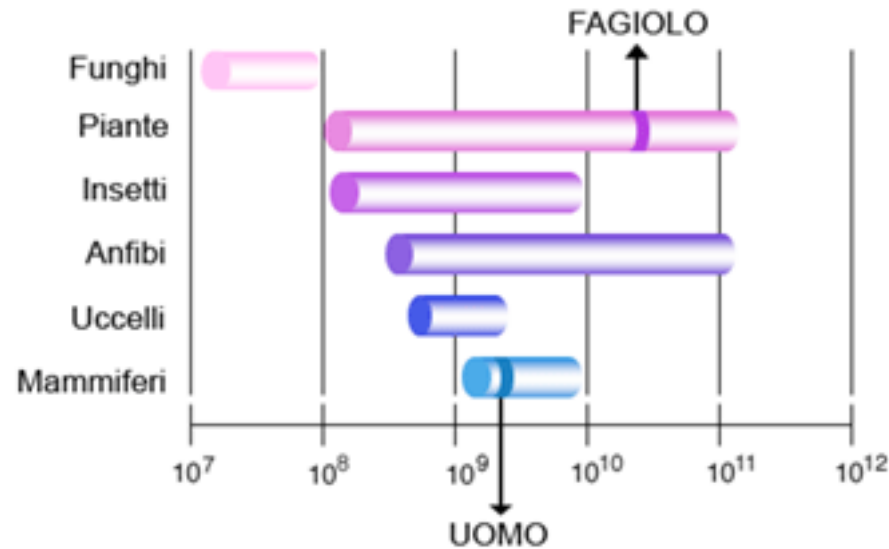




# Nel genoma umano non ci sono solo i geni

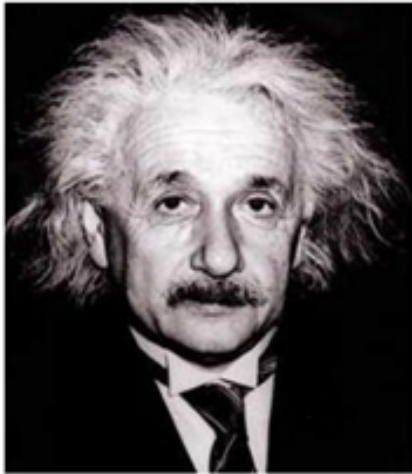


La complessità di un organismo non correla con la dimensione del suo DNA



**Tabella comparativa della dimensione dei genomi presenti in natura.** La figura mostra come non esista correlazione fra la complessità dell'organismo e la dimensione del suo genoma, indicata in paia di basi.

Il numero dei geni e la complessità degli organismi non sono correlati



~21000 geni



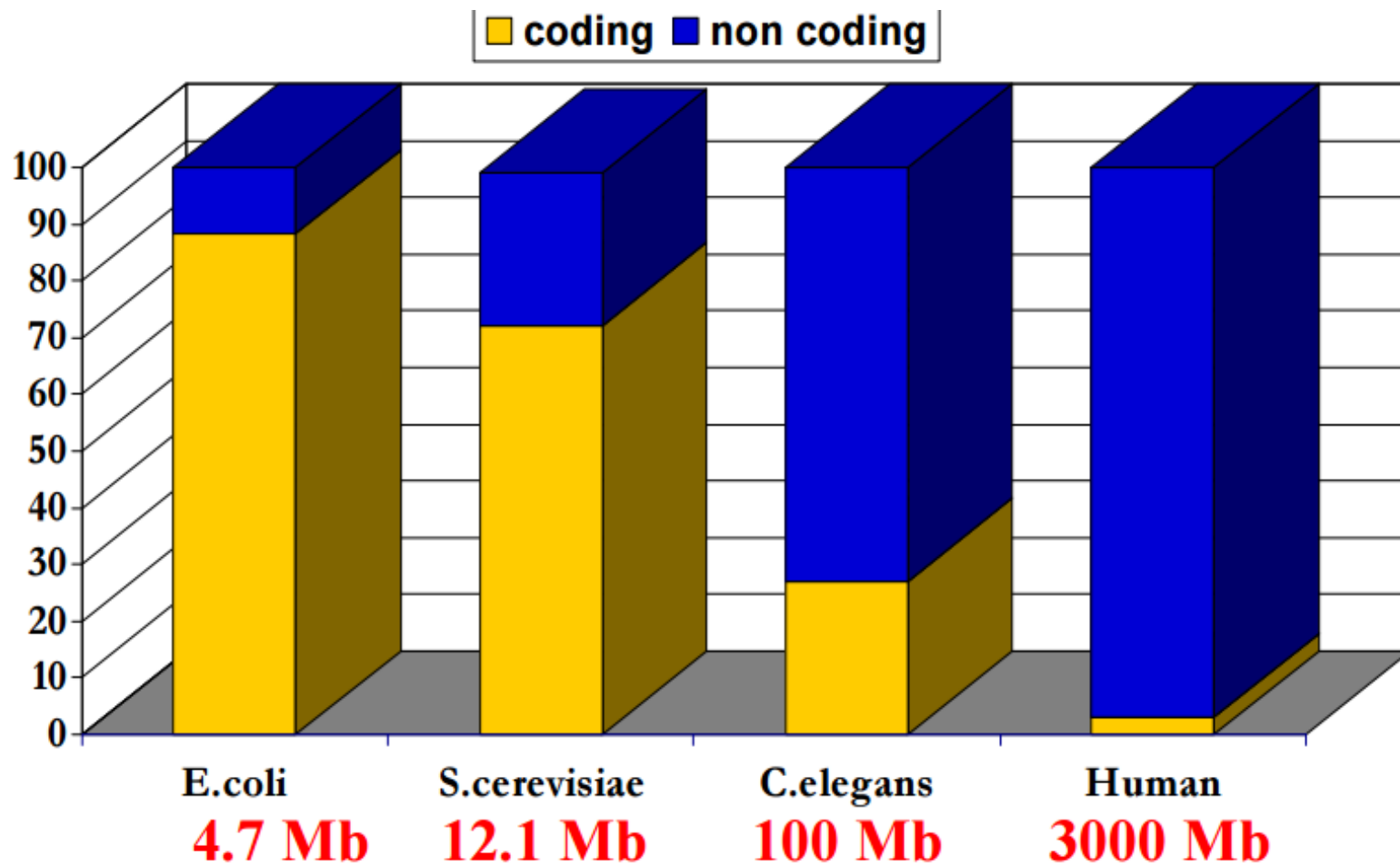
~25000 geni



~60000 geni

# Che cosa ci rende umani?

Funzionamento del genoma, la sua  
capacità di "costruire" molte proteine  
diverse a partire dagli stessi geni



Recenti studi hanno messo in evidenza che la **percentuale del DNA** destinata a codificare le proteine e' estremamente **ridotta**, mentre la **quantità di RNA** (trascrittoma) e' enormemente più **elevata**.



Una volta che il DNA umano è stato sequenziato



- capire la funzione biologica di ogni gene di cui è stata resa nota la sequenza e le interazioni tra di essi
- analisi della variabilità genomica

# VARIABILITA' DEL DNA UMANO

- Ogni essere umano ha una propria sequenza genomica individuale
- I genomi di due individui presi a caso contengono almeno 3 milioni di differenze (l'uno per mille)
- Solo gemelli identici possiedono lo stesso corredo di DNA

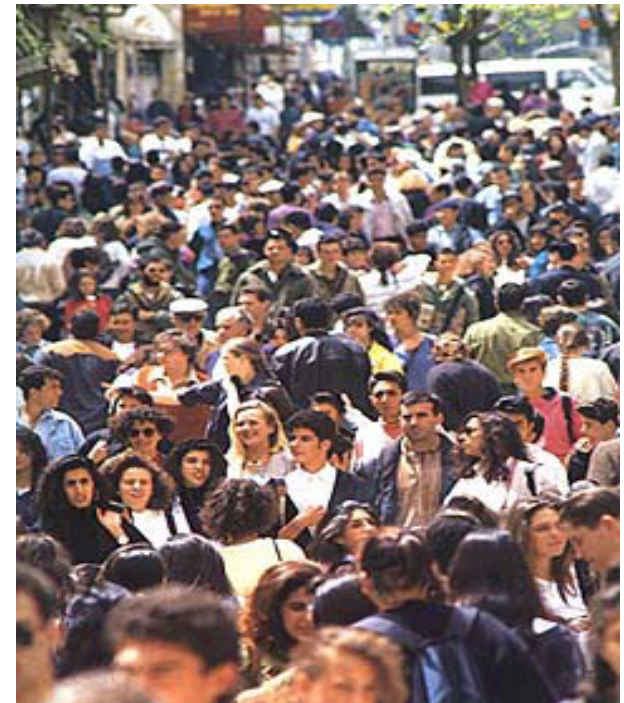


Non siamo tutti uguali...



Non esiste un' unica sequenza del genoma umano, ma circa 3 milioni dei 3 miliardi di nucleotidi che compongono il genoma di un individuo variano rispetto al genoma di un altro individuo e costituiscono dei

## POLIMORFISMI GENETICI



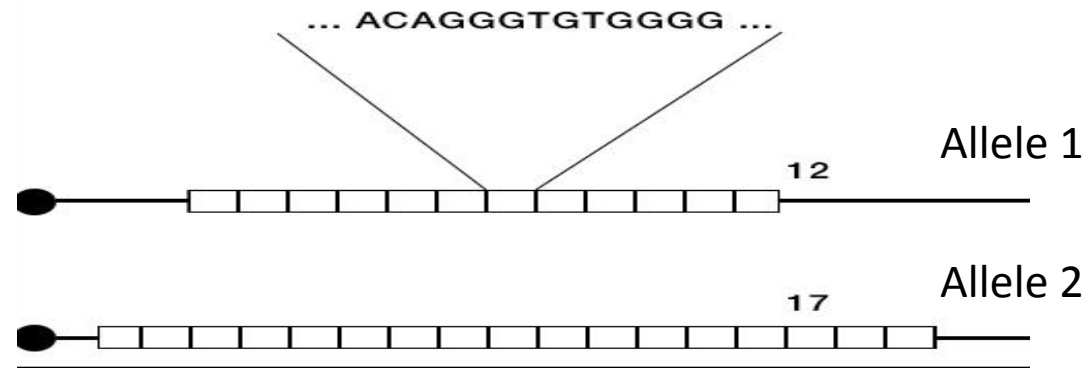
# Polimorfismi genetici

Variazioni nelle sequenza di DNA  
presenti nella popolazione con una  
frequenza  $>1\%$



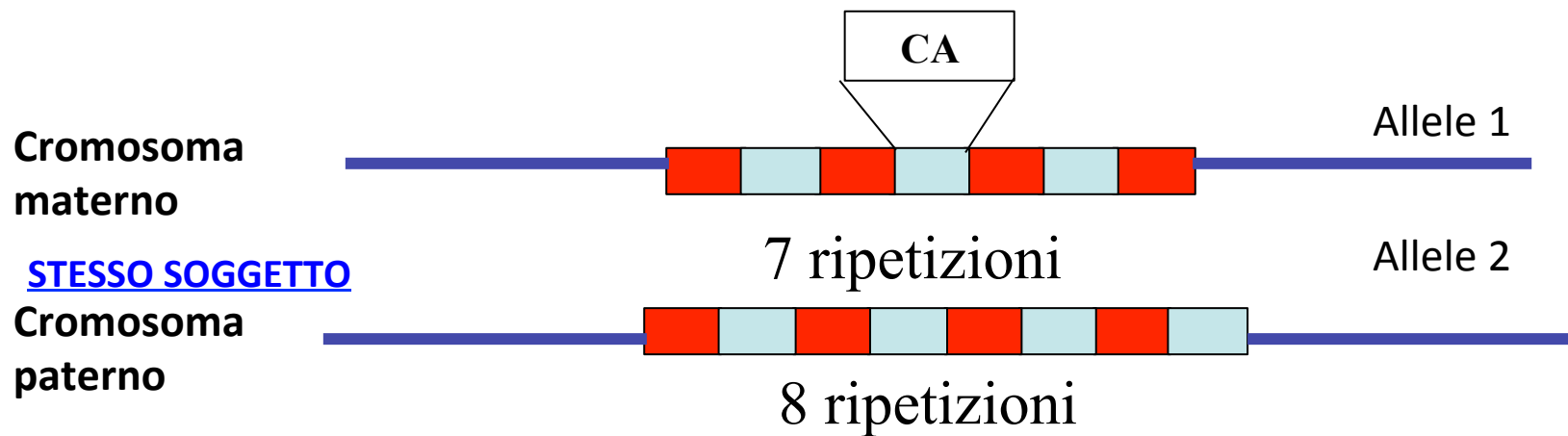
Responsabili della variabilità individuale

**VNTR** (Variable Number Tandem): tratti di sequenza lunghi 15-40 nucleotidi ripetuti in tandem un numero variabile di volte.





**STR** (Short Tandem Repeats): ripetizioni in tandem di tratti di sequenza lunghi 2-4 nucleotidi.



gaccta	ca	ca	taccgttaa	Allele 1				
gaccta	ca	ca	ca	taccgttaa	Allele 2			
gaccta	ca	ca	ca	ca	taccgttaa	Allele 3		
gaccta	ca	ca	ca	ca	ca	taccgttaa	Allele 4	
gaccta	ca	ca	ca	ca	ca	ca	taccgttaa	Allele 5

**INDIVIDUI DIVERSI**

## Identificazione di tracce biologiche rinvenute sulla scena di un crimine

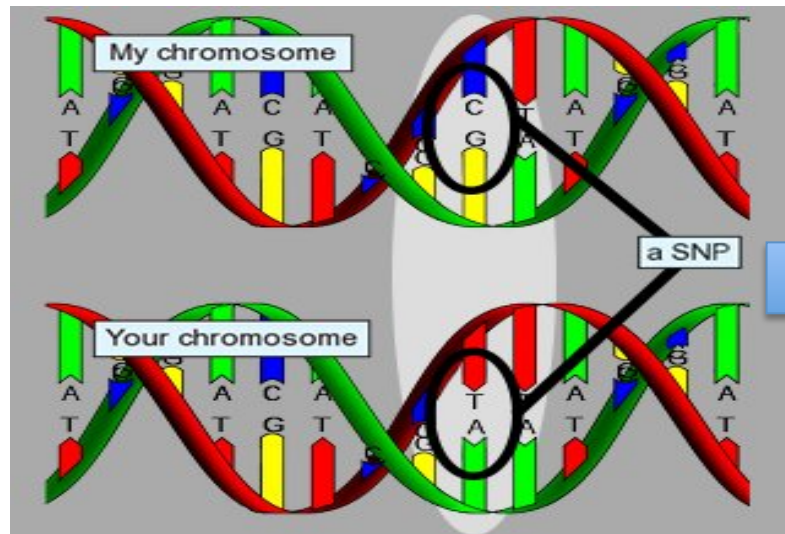


## Indagini di discendenza:

Permettono di stabilire la relazione biologica fra due o più persone.



# Polimorfismi a singolo nucleotide: SNP



Individual 1	Individual 4
Chr 2 ... CGATATTCC <b>T</b> ATCGAATGTC ... copy1 ... GCTATAAGGATAGCTTACAG ...	Chr 2 ... CGATATTCC <b>T</b> ATCGAATGTC ... copy1 ... GCTATAAGGATAGCTTACAG ...
Chr 2 ... CGATATTCC <b>C</b> ATCGAATGTC ... copy2 ... GCTATAAGGGTAGCTTACAG ...	Chr 2 ... CGATATTCC <b>C</b> ATCGAATGTC ... copy2 ... GCTATAAGGGTAGCTTACAG ...
Individual 2	Individual 5
Chr 2 ... CGATATTCC <b>C</b> ATCGAATGTC ... copy1 ... GCTATAAGGGTAGCTTACAG ...	Chr 2 ... CGATATTCC <b>C</b> ATCGAATGTC ... copy1 ... GCTATAAGGGTAGCTTACAG ...
Chr 2 ... CGATATTCC <b>C</b> ATCGAATGTC ... copy2 ... GCTATAAGGGTAGCTTACAG ...	Chr 2 ... CGATATTCC <b>T</b> ATCGAATGTC ... copy2 ... GCTATAAGGATAGCTTACAG ...
Individual 3	Individual 6
Chr 2 ... CGATATTCC <b>T</b> ATCGAATGTC ... copy1 ... GCTATAAGGATAGCTTACAG ...	Chr 2 ... CGATATTCC <b>C</b> ATCGAATGTC ... copy1 ... GCTATAAGGGTAGCTTACAG ...
Chr 2 ... CGATATTCC <b>T</b> ATCGAATGTC ... copy2 ... GCTATAAGGATAGCTTACAG ...	Chr 2 ... CGATATTCC <b>T</b> ATCGAATGTC ... copy2 ... GCTATAAGGATAGCTTACAG ...

Fonte principale della variabilità genetica tra gli individui:  
3 milioni SNP, all'incirca uno SNP ogni 1000pb.

# PROGETTO 1000 genomi

**Sequenziare il genoma di 1000 individui di tutto il mondo.**



**Catalogare le varianti genetiche umane così da comprendere meglio le relazioni genotipo fenotipo**

# PROGETTO 1000 genomi

- 15 milioni di SNP
- un milione di piccole inserzioni o delezioni (variazioni genetiche per perdita o aggiunta di poche basi azotate)
- 20 mila varianti più ampie mai osservate prima.
- dal confronto tra genomi di genitori e figli, si è potuto stabilire quante mutazioni sorgono in ogni nuova generazione. Secondo l'analisi, ogni persona è portatrice di circa 60 mutazioni non presenti nei genitori.





 <http://www.1000genomes.org/home>

# 1000 Genomes

A Deep Catalog of Human Genetic Variation

[Home](#)

[About](#)

[Data](#)

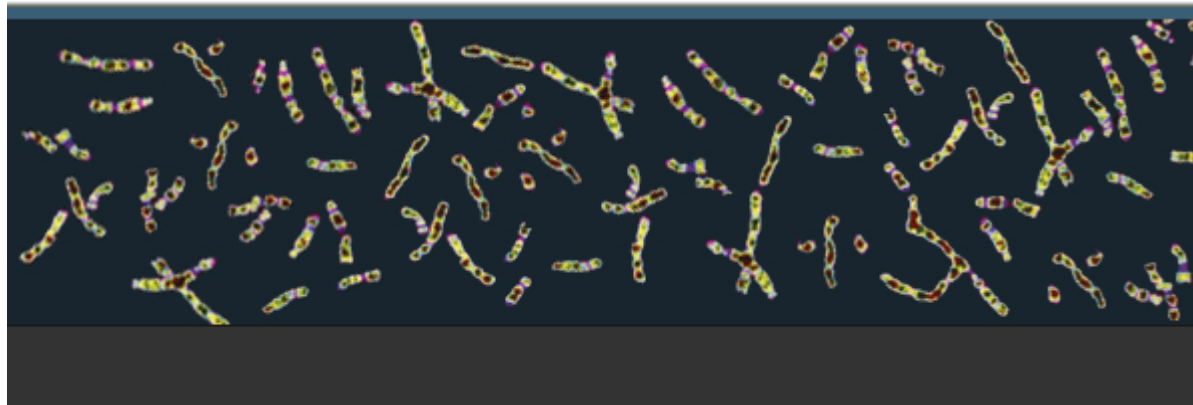
[Analysis](#)

[Participants](#)

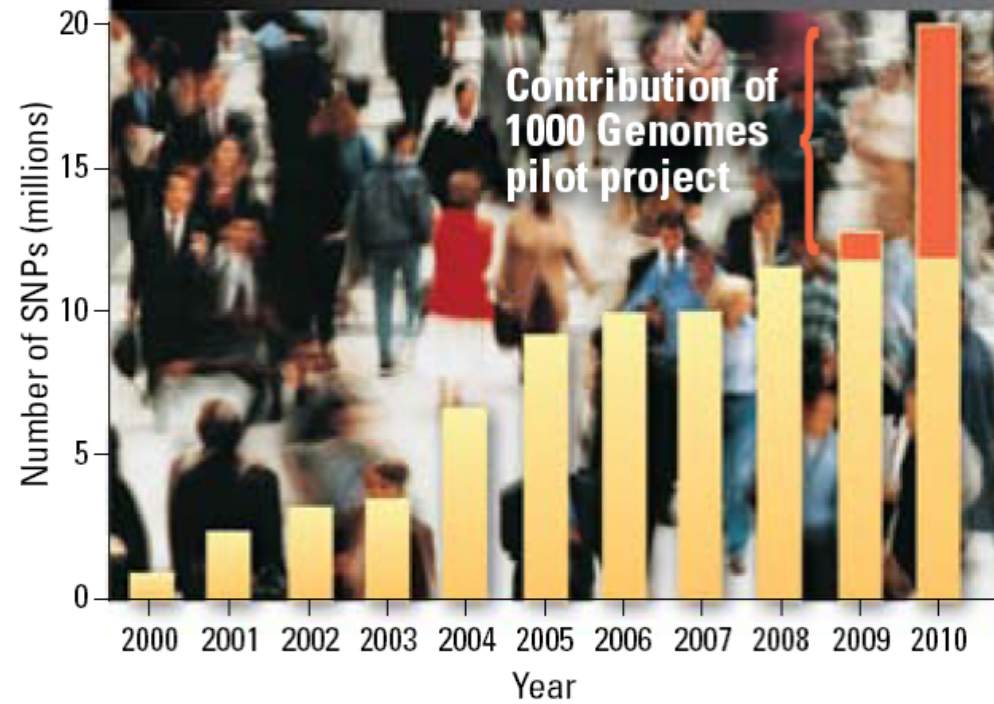
[Contact](#)

[Browser](#)

[Wiki](#)

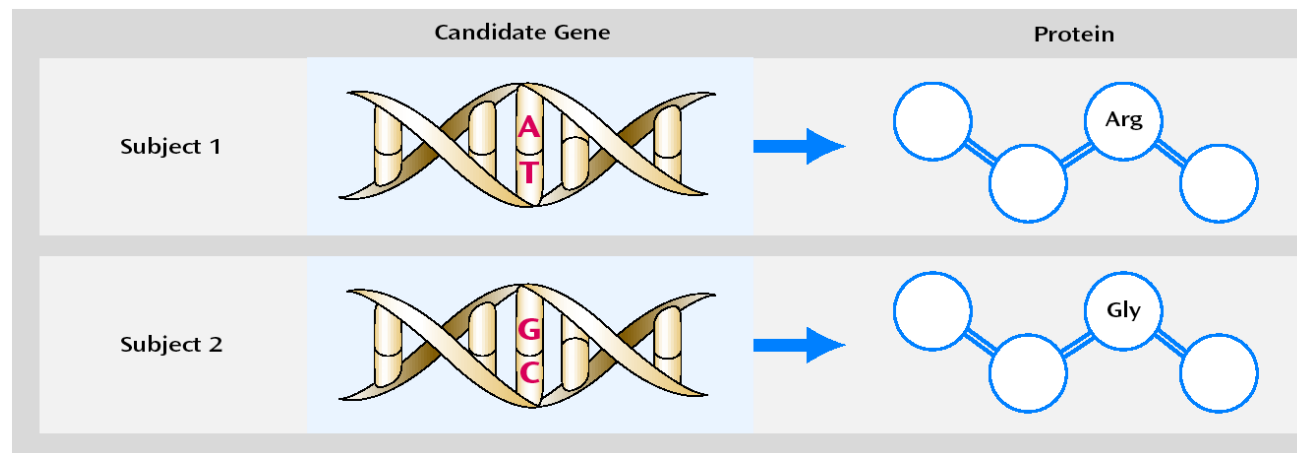


## GROWTH OF THE PUBLIC SNP DATABASE





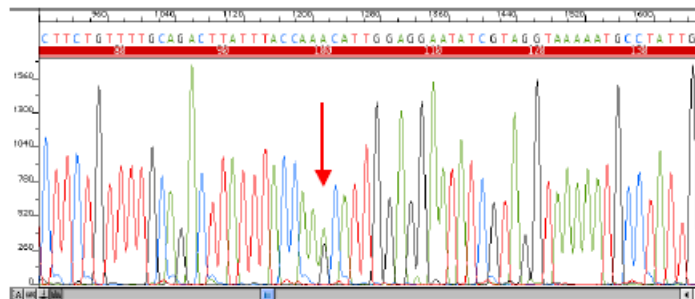
- Possono avere un significato funzionale diretto (es. Cambio aminoacidico in una sequenza codificante)



# Variazioni di una sola lettera

- Il gatto dorme sul divano
  - Il **r**atto dorme sul divano
  - Il gatto dorme sul **-**ivano
  - Il **g**ratto dorme sul divano
- wt (normale)  
sostituzione  
delezione  
inserzione

## Sostituzione di una base



ATTACCA AACATTGGA  
ATTACCA **G** AACATTGGA

A red box highlights the 'G' in the second sequence, with a red arrow pointing down to it, indicating a substitution.

I polimorfismi possono essere responsabili di una diversa suscettibilità alle malattie...



## ApoE e malattia di Alzheimer

ApoE  $\epsilon$ 2 GACGTG**TGC**GGCCGC.....CAGAAG**TGC**CTGGCA

Cys

Cys

ApoE  $\epsilon$ 3 GACGTG**TGC**GGCCGC.....CAGAAG**CGC**CTGGCA

Cys

Arg

ApoE  $\epsilon$ 4 GACGTG**CGC**GGCCGC.....CAGAAG**CGC**CTGGCA

Arg

Arg

...come pure della variabilità  
individuale nella risposta alla  
terapia farmacologica



# Farmacogenetica

Esamina le varianti genetiche che determinano la risposta ad un farmaco e studia il modo in cui queste varianti possono essere usate per prevedere il tipo di risposta

- Farmaco giusto
- dose giusta
- paziente giusto

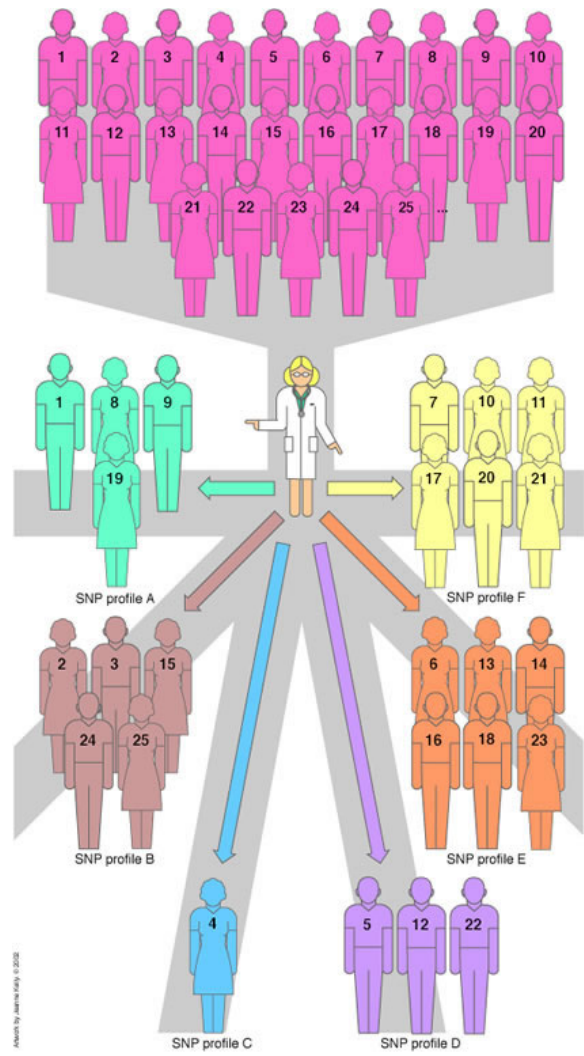


Selezione sulla base di fattori predisponenti alle malattie

Selezione sulla base di fattori responsabili di una diversa risposta al trattamento

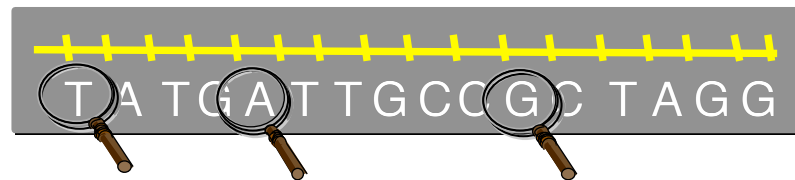
diagnosi precoce

Terapia personalizzata



Adapted by James Kelly © 2012

## A caccia di alleli di suscettibilità



# Studi di associazione casi-controlli

Popolazione con malattia



ATTGCATGCCAGTAGG

Popolazione senza malattia



TATGATTGCCGCTAGG

Si cercano differenze nelle frequenze dei polimorfismi nei due gruppi

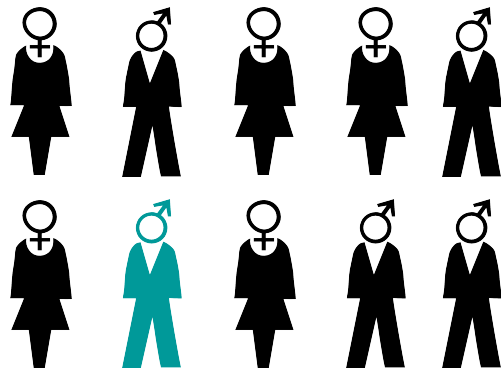
## Selezione delle due popolazioni

- devono differire solo per il fenotipo di interesse
- devono essere il più omogenee possibile per tutti gli altri aspetti (sesso, età, etnia...)
- devono essere sufficientemente numerose  
(la numerosità del campione utile per rilevare associazioni statisticamente significative dipende dalla frequenza degli SNP studiati)

## Gene A



Casi



Controlli

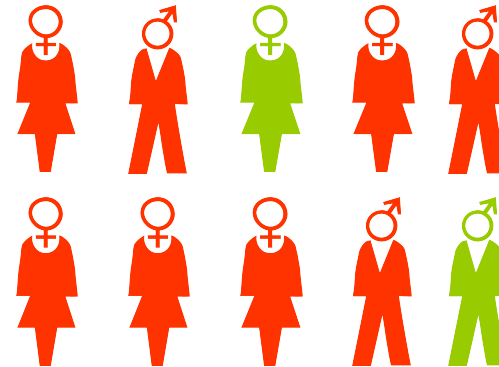
Nessuna variante del gene  
(**verde** o nera) è associata con il  
fenotipo d'interesse



La variante **rossa** del gene è associata con il fenotipo d'interesse

Questo tipo di analisi può essere fatto senza conoscere la funzione del gene

## Gene B

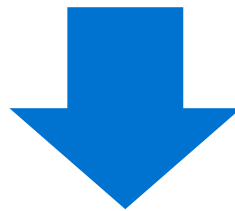


### Casi



### Controlli

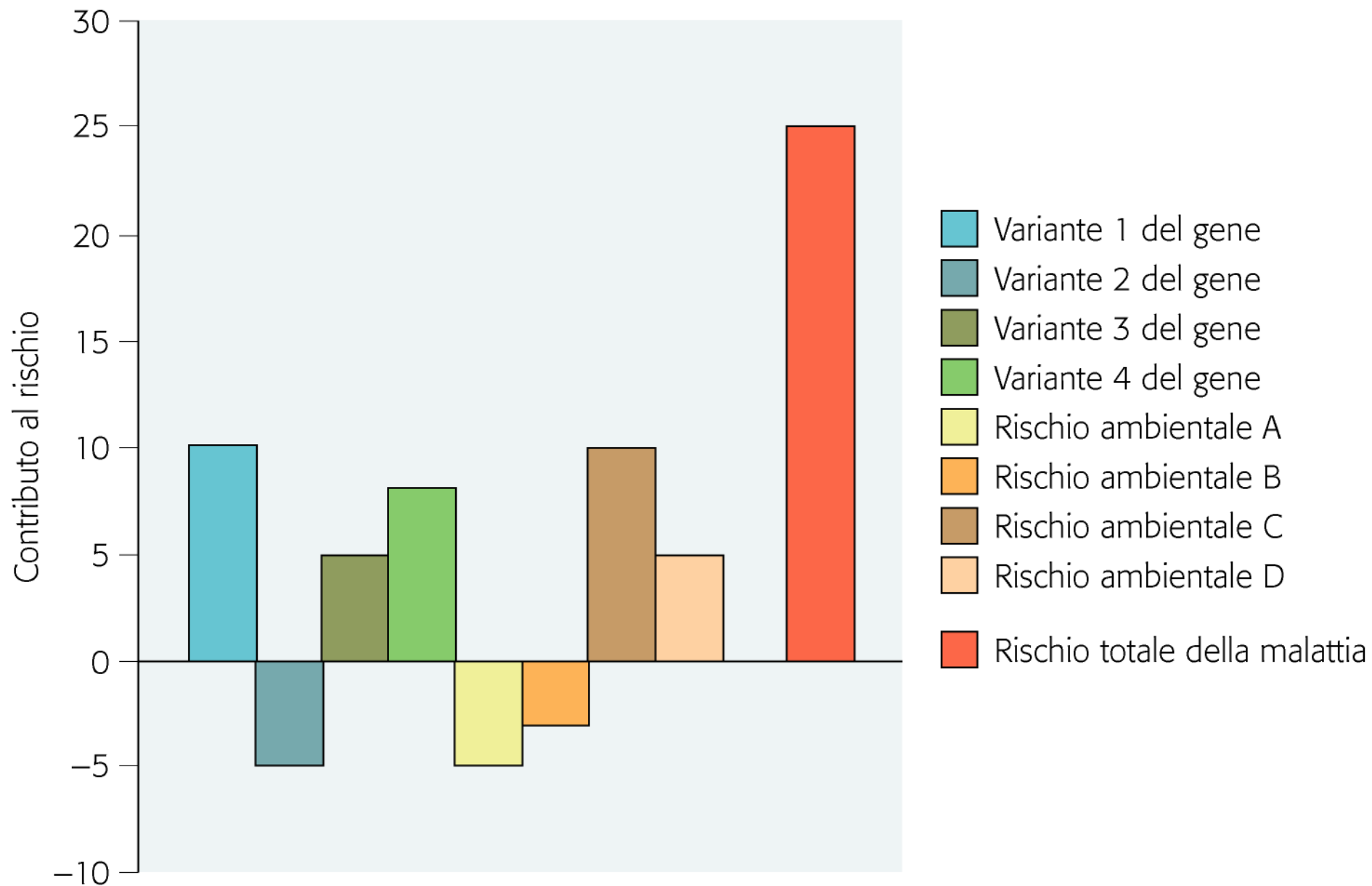
L'identificazione delle varianti che rappresentano fattori di vulnerabilità non è semplice in quanto ognuna di esse agisce in concomitanza con molte altre varianti e con numerosi fattori ambientali

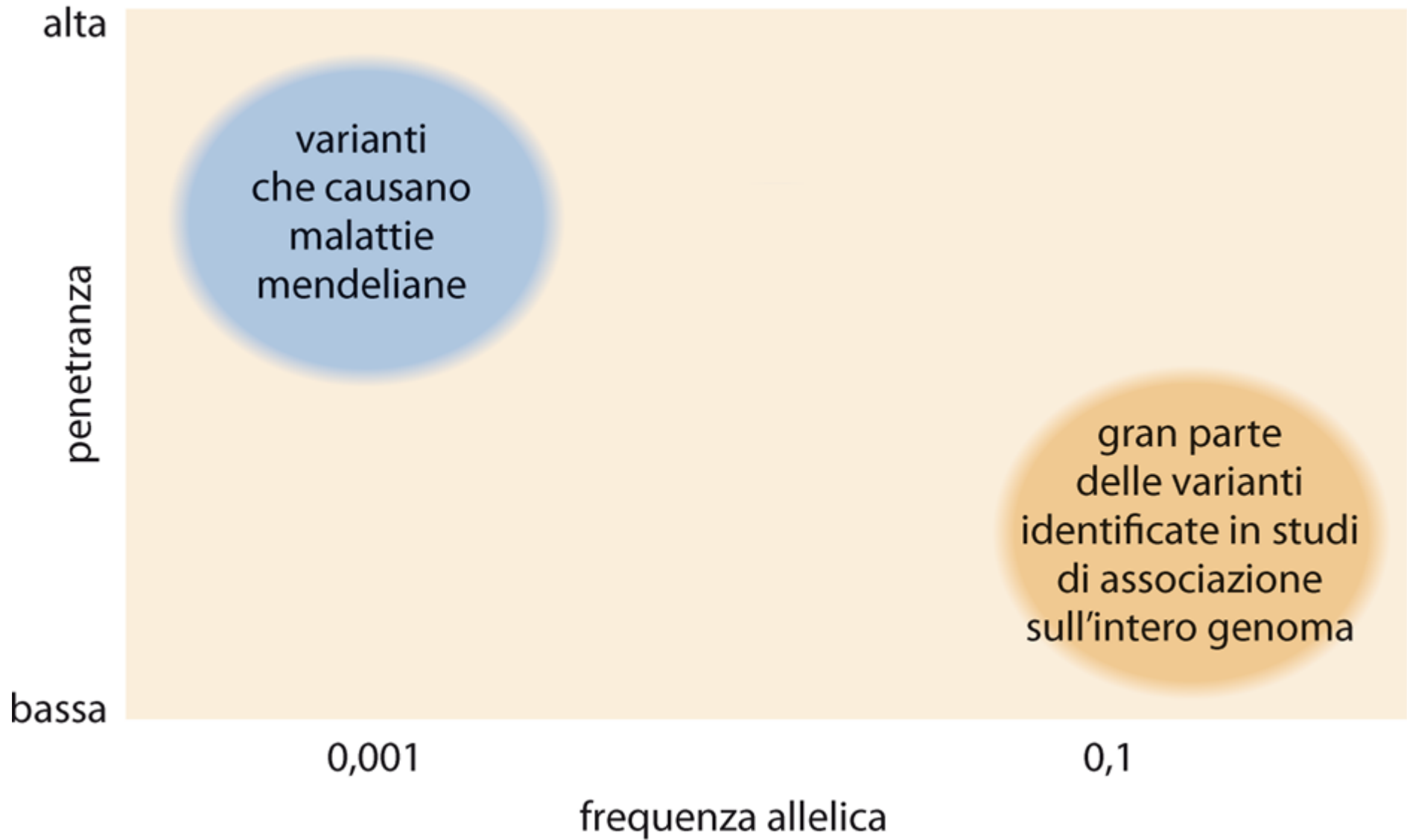


Da sole non sono in grado di determinare il fenotipo studiato



**BASSA PENETRANZA**

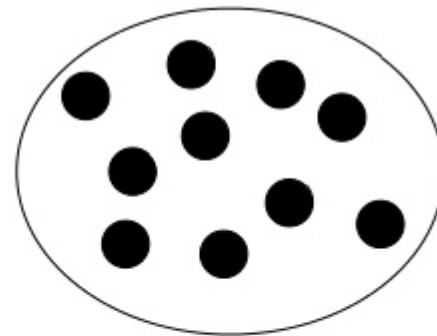




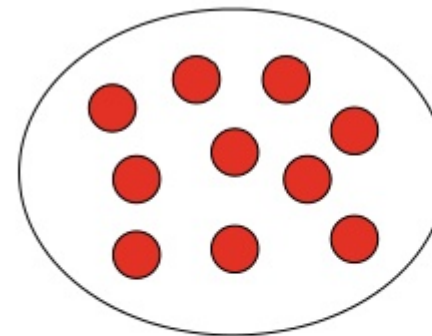
## Studi di associazione

- Studio caso-controllo: caso ideale
- Malattie monogeniche, penetranza completa

● Variante genetica causale  
● Variante genetica protettiva



Controlli: individui non affetti



Casi: individui affetti