

# ***Silvia Pellegrini***

Laboratorio di Biologia Molecolare

*Dipartimento di Patologia Chirurgica, Medica,  
Molecolare e dell' Area Critica*

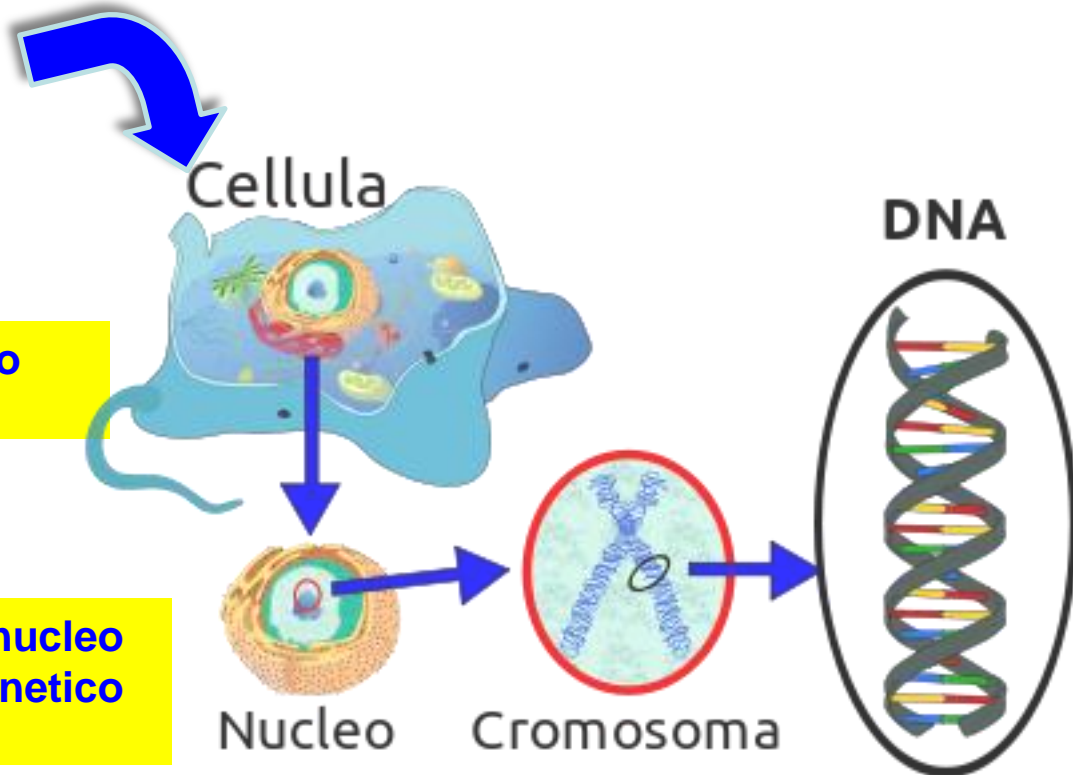


SD Patologia Clinica Universitaria S. Chiara

Tel. 050 2211251

e-mail: [silvia.pellegrini@med.unipi.it](mailto:silvia.pellegrini@med.unipi.it)

## Qualche nozione di biologia di base



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

Il DNA nel nucleo è  
organizzato in cromosomi

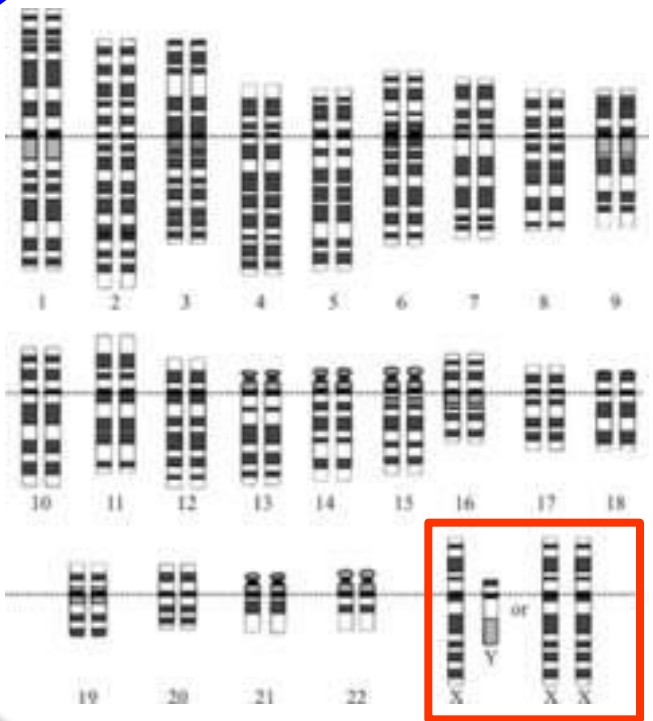
# DNA nucleare

23 Coppie di Cromosomi

22 Autosomi in duplice copia

Nucleo

Citoplasma



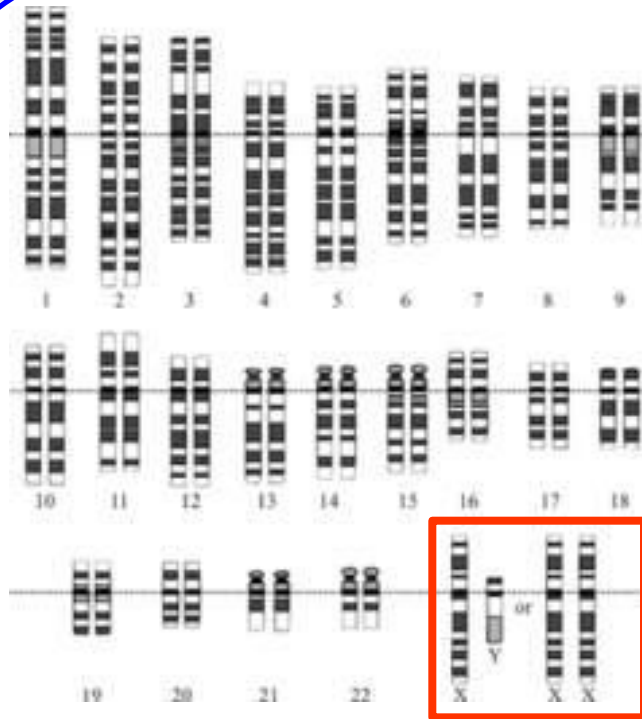
2 Cromosomi sessuali

# DNA umano totale

nucleare + mitocondriale

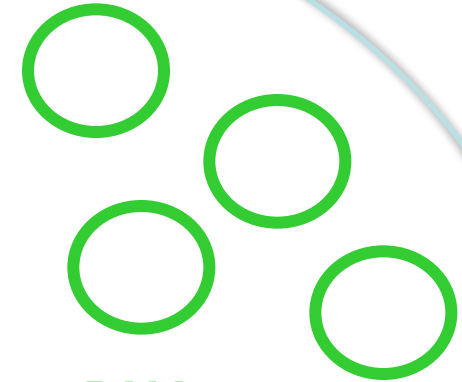
22 Autosomi in duplice copia

Nucleo



2 Cromosomi sessuali

Citoplasma



DNA  
Mitocondriale  
localizzato nei  
mitocondri

100 copie per cellula





# 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. 4366 April 25, 1953 NATURE 737

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

<sup>4</sup>Young, F. B., Gerard, H., and Jevons, W., *Phil. Mag.*, **40**, 149 (1925).

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

<sup>6</sup>Von Ardenne, W. S., *Woods Hole Papers in Phys. Oceanog. Meteor.*, **11** (3) (1950).

<sup>7</sup>Ekmann, V. W., *Arkiv. Mat. Astron. Fysik. (Stockholm)*, **2** (11) (1956).

### 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>4</sup>. They kindly made their manuscript available to us in advance of publication. Their model consists of three intertwined 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 Frazer (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 diester 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 Furberg's<sup>5</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 Furberg'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 phosphorus atom from the fibre axis is 10 Å. As the phosphates are on the outside, esters 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 purine and pyrimidine 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>6,7</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>8,9</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.

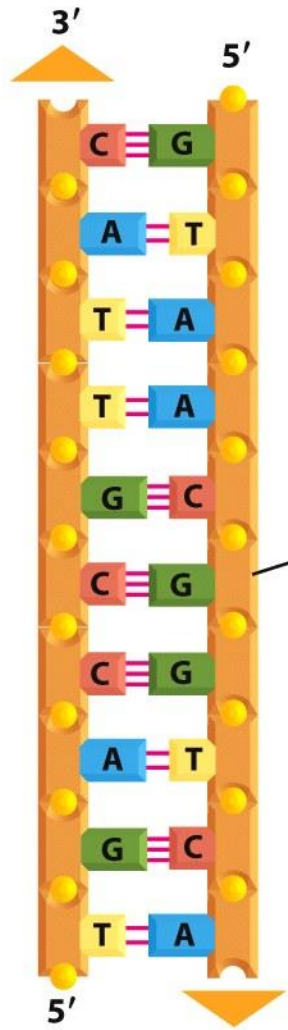
Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on interatomic 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

# 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)

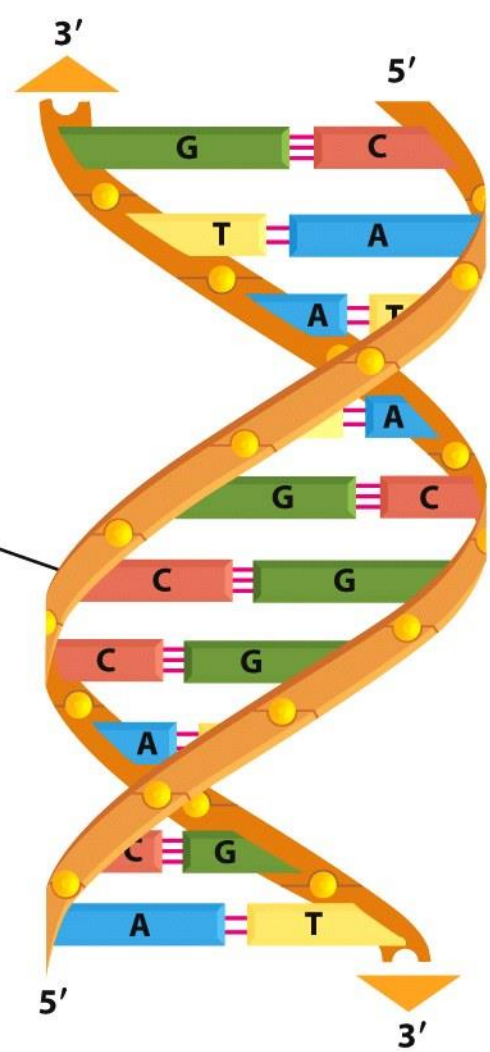
(C) double-stranded DNA



sugar-phosphate backbone

hydrogen-bonded base pairs

(D) DNA double helix







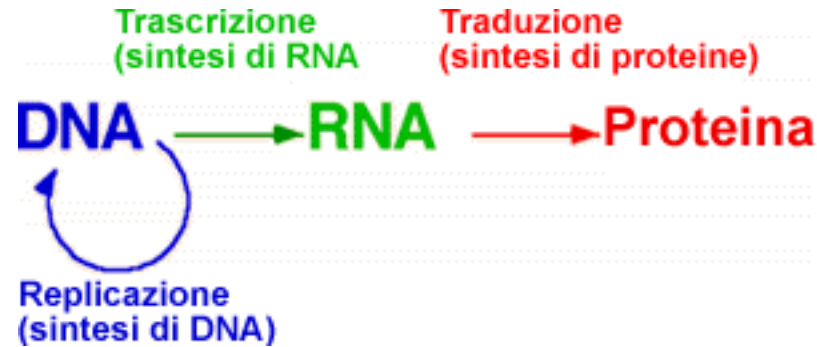
# *Replicazione del DNA*



© R.G. Steane

- ✚filamento stampo viene copiato per mezzo di complementarità delle basi
- ✚affinchè il riconoscimento avvenga i due filamenti della doppia elica devono separarsi
- ✚i nucleotidi appropriati vengono inseriti nella catena nascente dalla DNA polimerasi

# Dogma centrale della biologia molecolare



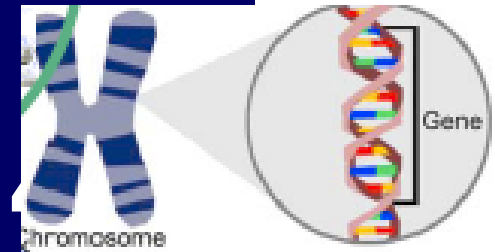
# Il codice genetico permette di tradurre il messaggio in proteine

GCA	AGA									UUA					AGC					
GCC	AGG									UUG					AGU					
GCG	CGA					GGA			AUA	CUA				CCA	UCA	ACA				GUA
GCU	CGC	GAC	AAC	UGC	GAA	CAA	GGC	CAC	AUC	CUC	AAA			CCC	UCC	ACC				GUC
	CGU	GAU	AAU	UGU	GAG	CAG	GGU	CAU	AUU	CUU	AAG	AUG		CCG	UCG	ACG			UAC	GUG
														CCU	UCU	ACU	UGG	UAU		GUU
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	



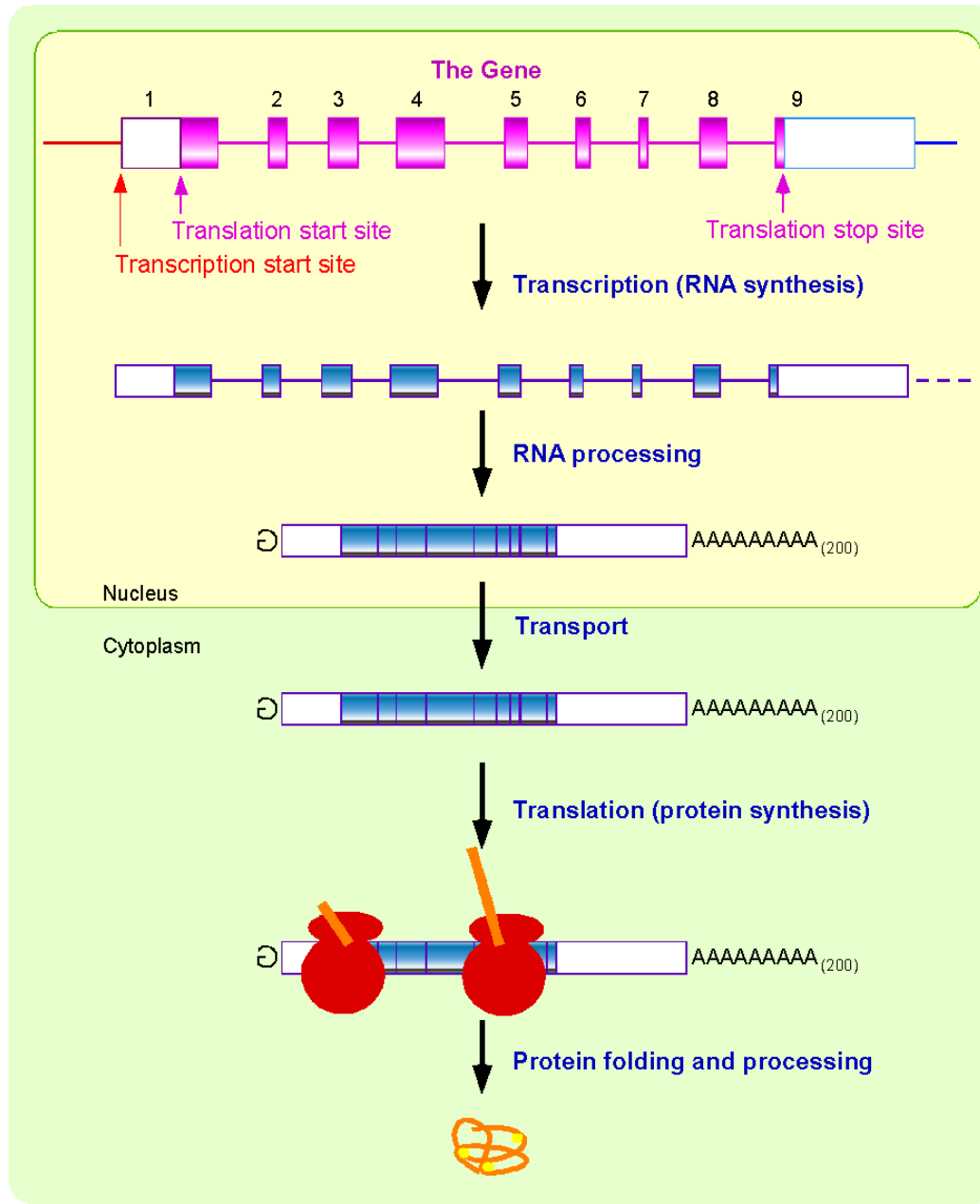
**Il contenuto di  
informazioni del DNA  
umano è dato  
dall'alternanza di  
lettere**

**A, G, T, C**



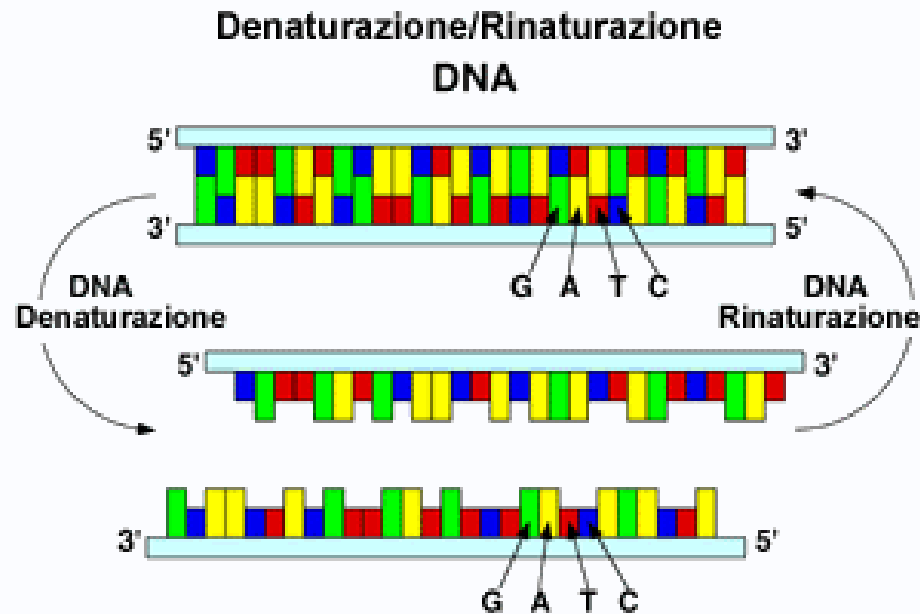


# La maggior parte dei geni sono sequenze uniche



# Non ci sono solo geni nel DNA

*In base a esperimenti di denaturazione e rinaturazione del DNA*



*50-60% delle sequenze ha bassa velocità  
di riassociazione*

*DNA a singola copia*

*25-40% velocità intermedia*

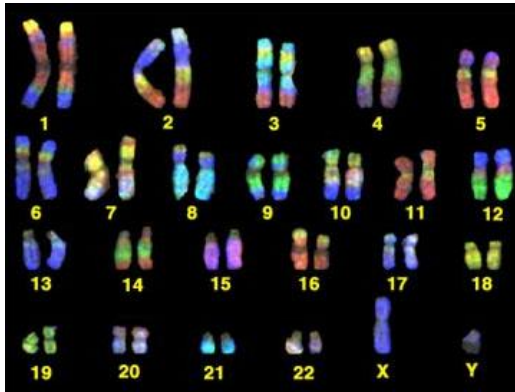
*DNA moderatamente ripetuto*

*10-15% alta velocità di riassociazione*

*DNA altamente ripetuto*

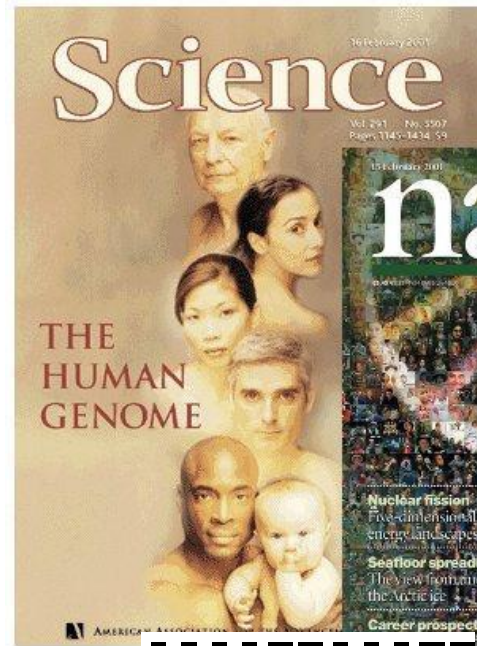
L'insieme di tutte le molecole di DNA  
presenti nel nucleo di ogni cellula  
costituisce il **Genoma**

**Nell'uomo:** 44 cromosomi autosomici  
2 cromosomi sessuali



# Progetto Genoma Umano

1990



ACAATTCGATT  
CTCTAGCACGC  
GCTAGTTAGTT  
TACTACCTACG  
ACTGAGATCGA

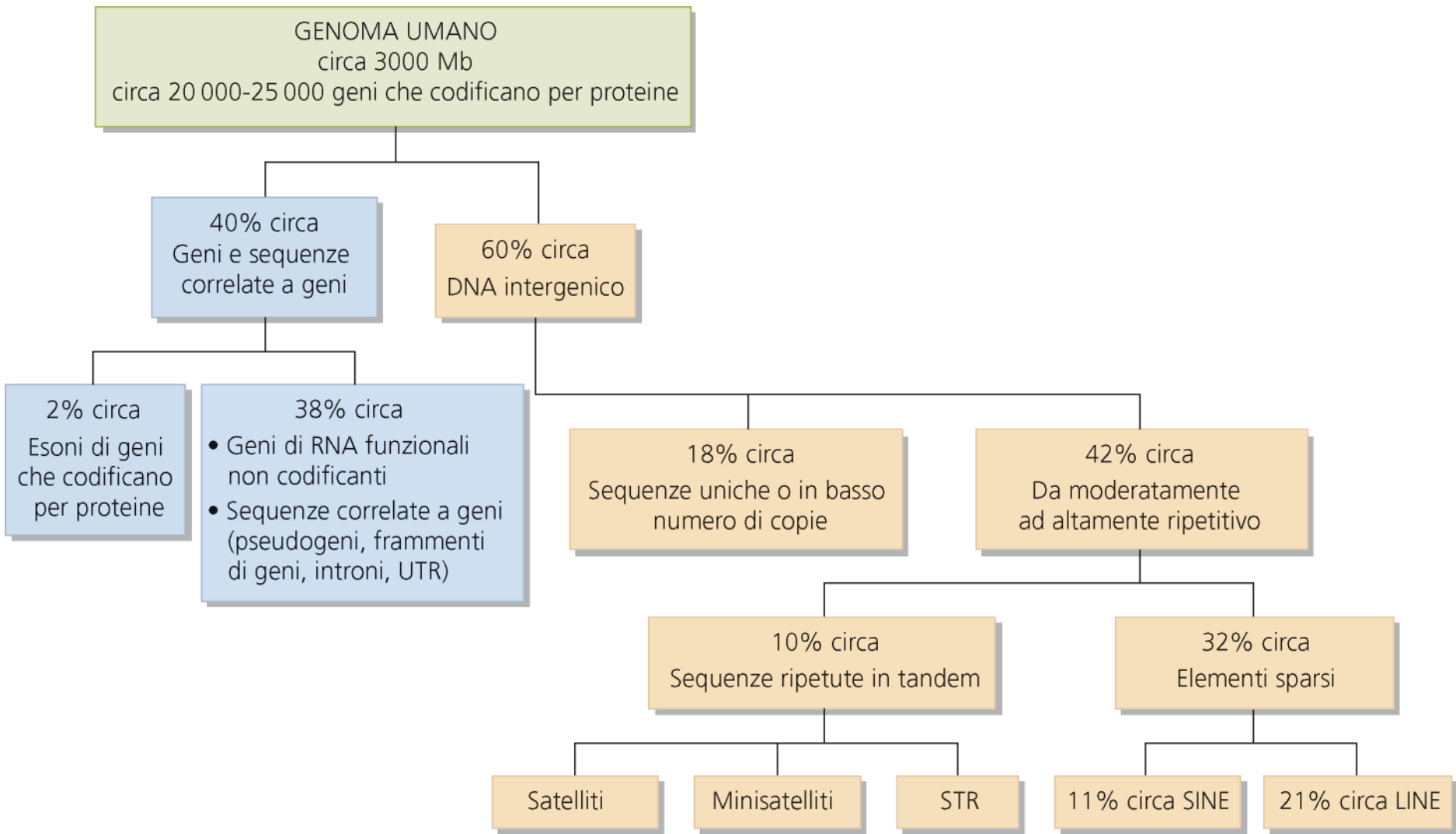
2003

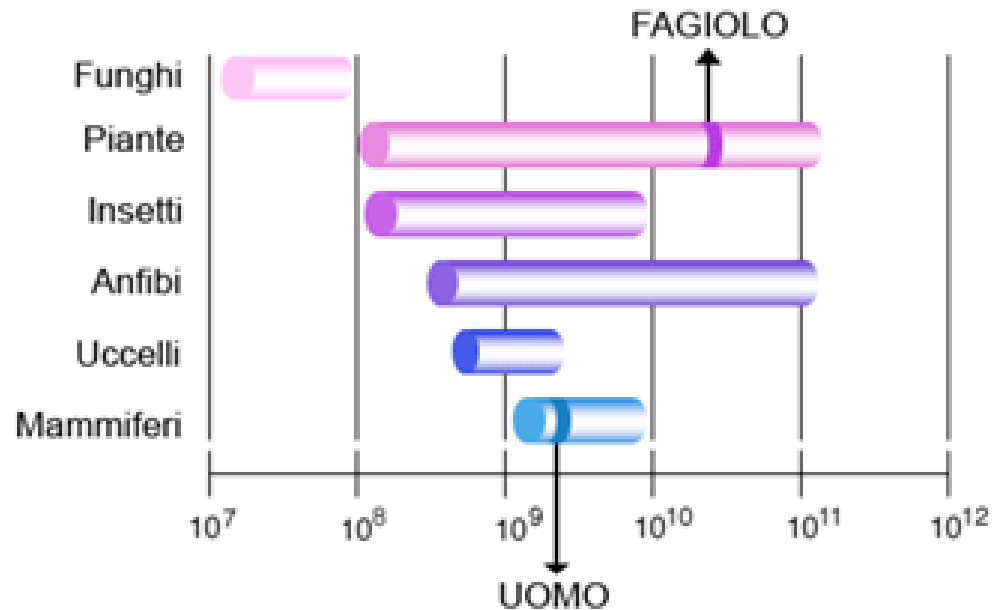


# Che cosa si conosce del genoma umano

- ⑩ 3 miliardi di paia di basi
- ⑩ 2% codificanti
- ⑩ Circa 20.000 geni diversi







**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.

PHASE TWO: INTERPRETATION

SHENEMAN The Star Ledger



Drew Sheneman -- The Newark Star Ledger

# Era postgenomica

- tecniche di sequenziamento in parallelo consentono di ridurre tempi e costi consentendo di sequenziare interi genomi in poche ore
- analisi bioinformatica: consente di immagazzinare e gestire una mole enorme di dati e di confrontare interi genomi consentendo di identificare le porzioni più conservate del nostro genoma e quelle che ci rendono unici
- analisi del trascrittoma per identificare i geni e la loro funzione
- **analisi della variabilità genomica**



# VARIABILITA' DEL DNA UMANO

- Il DNA costituisce il Genoma Umano
- Il genoma umano è dato da circa 3 miliardi di basi (A-G-C-T)
- Il genoma di ciascun individuo è unico ed è ereditato dai genitori
- I genomi di due individui presi a caso contengono almeno 3 milioni di differenze
- Solo gemelli identici possiedono lo stesso corredo di DNA



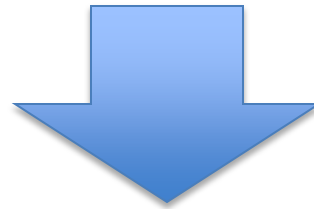
## Non siamo tutti uguali...

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**

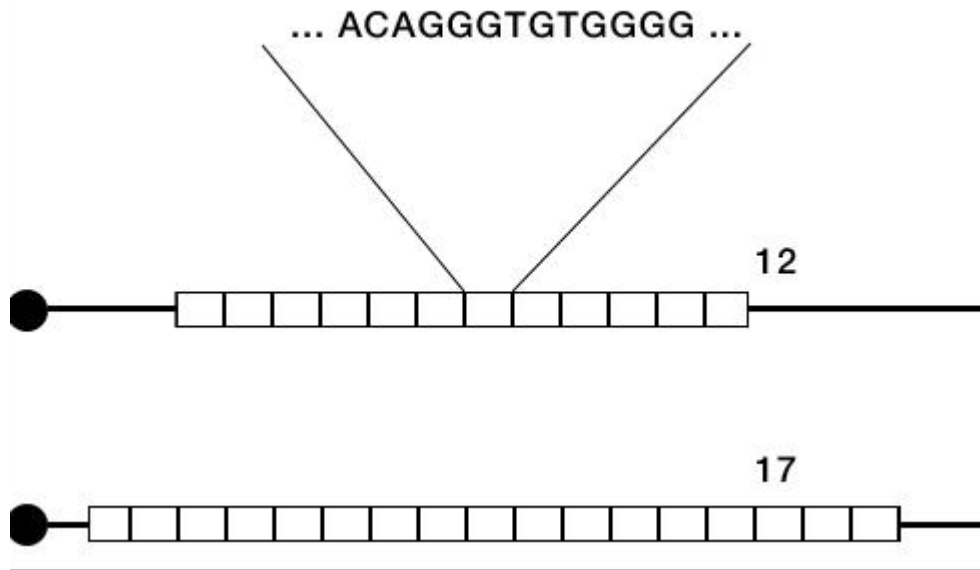
# Polimorfismi genetici

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

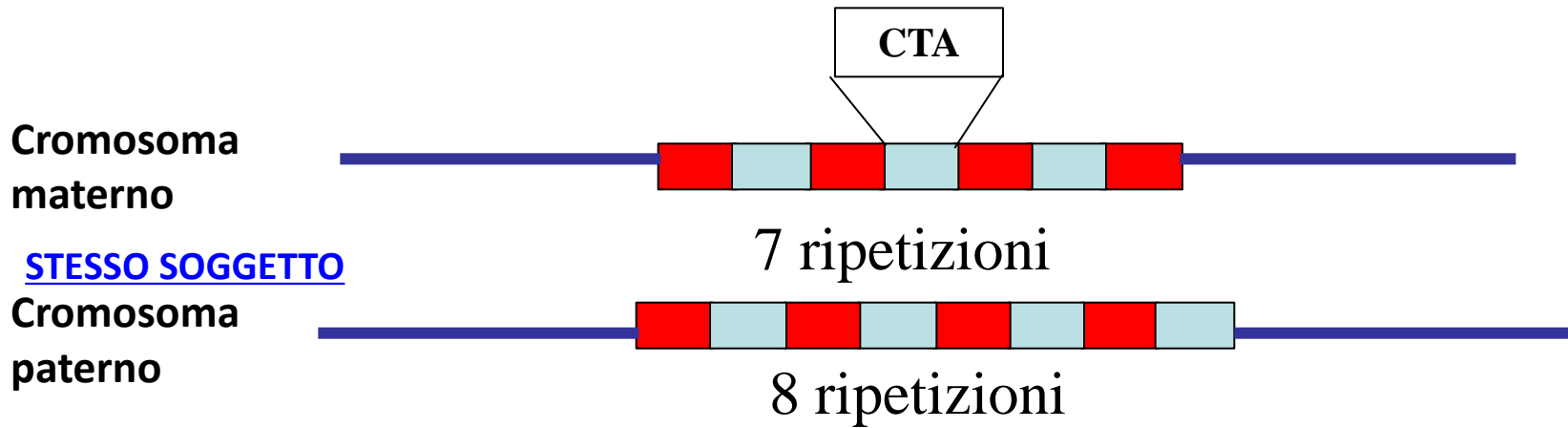


Responsabili della variabilità individuale

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

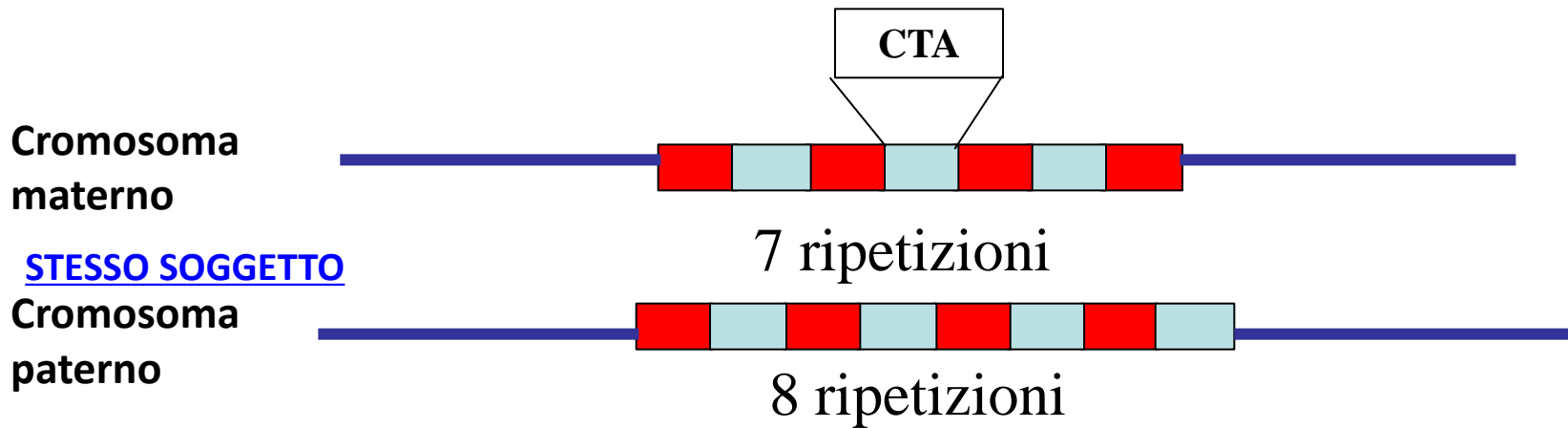


**STR** (Short Tandem Repeats): ripetizioni in tandem di tratti di sequenza lunghi 2-4 nucleotidi.





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



gacctaatc	ca ca	taccgtaa	Allele 1
gacctaatc	ca ca ca	taccgtaa	Allele 2
gacctaatc	ca ca ca ca	taccgtaa	Allele 3
gacctaatc	ca ca ca ca ca	taccgtaa	Allele 4
gacctaatc	ca ca ca ca ca ca	taccgtaa	Allele 5

**INDIVIDUI DIVERSI**

# Identificazione di tracce biologiche rinvenute sulla scena di un crimine

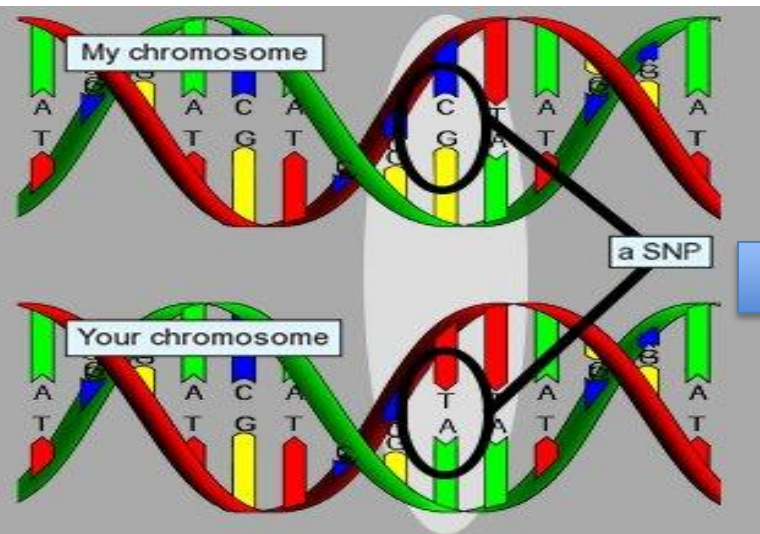


## Indagini di discendenza:

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



# Polimorfismi a singolo nucleotide: SNP



Individual 1		Individual 4	
Chr 2	...CGATATTCCATCGAATGTC...	Chr 2	...CGATATTCCATCGAATGTC...
copy1	...GCTATAAGGATAGCTTACAG...	copy1	...GCTATAAGGATAGCTTACAG...
Chr 2	...CGATATTCCATCGAATGTC...	Chr 2	...CGATATTCCATCGAATGTC...
copy2	...GCTATAAGGGTAGCTTACAG...	copy2	...GCTATAAGGGTAGCTTACAG...
Individual 2		Individual 5	
Chr 2	...CGATATTCCATCGAATGTC...	Chr 2	...CGATATTCCATCGAATGTC...
copy1	...GCTATAAGGGTAGCTTACAG...	copy1	...GCTATAAGGGTAGCTTACAG...
Chr 2	...CGATATTCCATCGAATGTC...	Chr 2	...CGATATTCCATCGAATGTC...
copy2	...GCTATAAGGGTAGCTTACAG...	copy2	...GCTATAAGGGTAGCTTACAG...
Individual 3		Individual 6	
Chr 2	...CGATATTCCATCGAATGTC...	Chr 2	...CGATATTCCATCGAATGTC...
copy1	...GCTATAAGGATAGCTTACAG...	copy1	...GCTATAAGGGTAGCTTACAG...
Chr 2	...CGATATTCCATCGAATGTC...	Chr 2	...CGATATTCCATCGAATGTC...
copy2	...GCTATAAGGATAGCTTACAG...	copy2	...GCTATAAGGGTAGCTTACAG...

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

HUMAN GENETICS

Dr Watson's base pairs

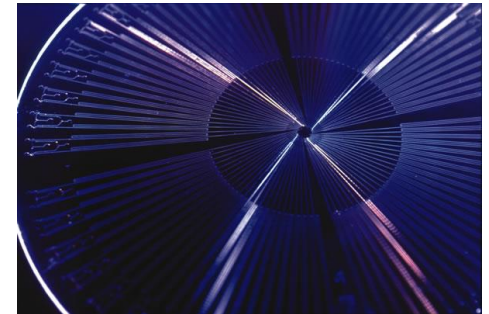
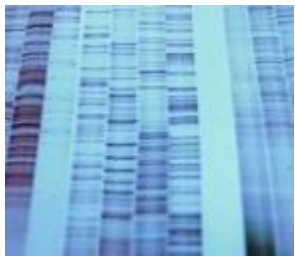
Maynard V. Olson

3.300.000 di variazioni rispetto alla sequenza del Genoma Umano presente in banca dati

TGGTTACGGCA TCGAGCTGCT CAGAGCTTGGTAGCGCTGACTGGCCATATA TATTAGCTGAT... AGCGTTGAGAA TTAGTA TGCGGCC TCGTTACGGCATCGAGCTGTCGACAGCTTCTGTACGTGC... TTACGCTGACGATTCGAA TTGCGAA TCGCTGAGCTTGTGAA TTAGTAGTCGCCCTCTCACGGCTC...

James Watson decoded.

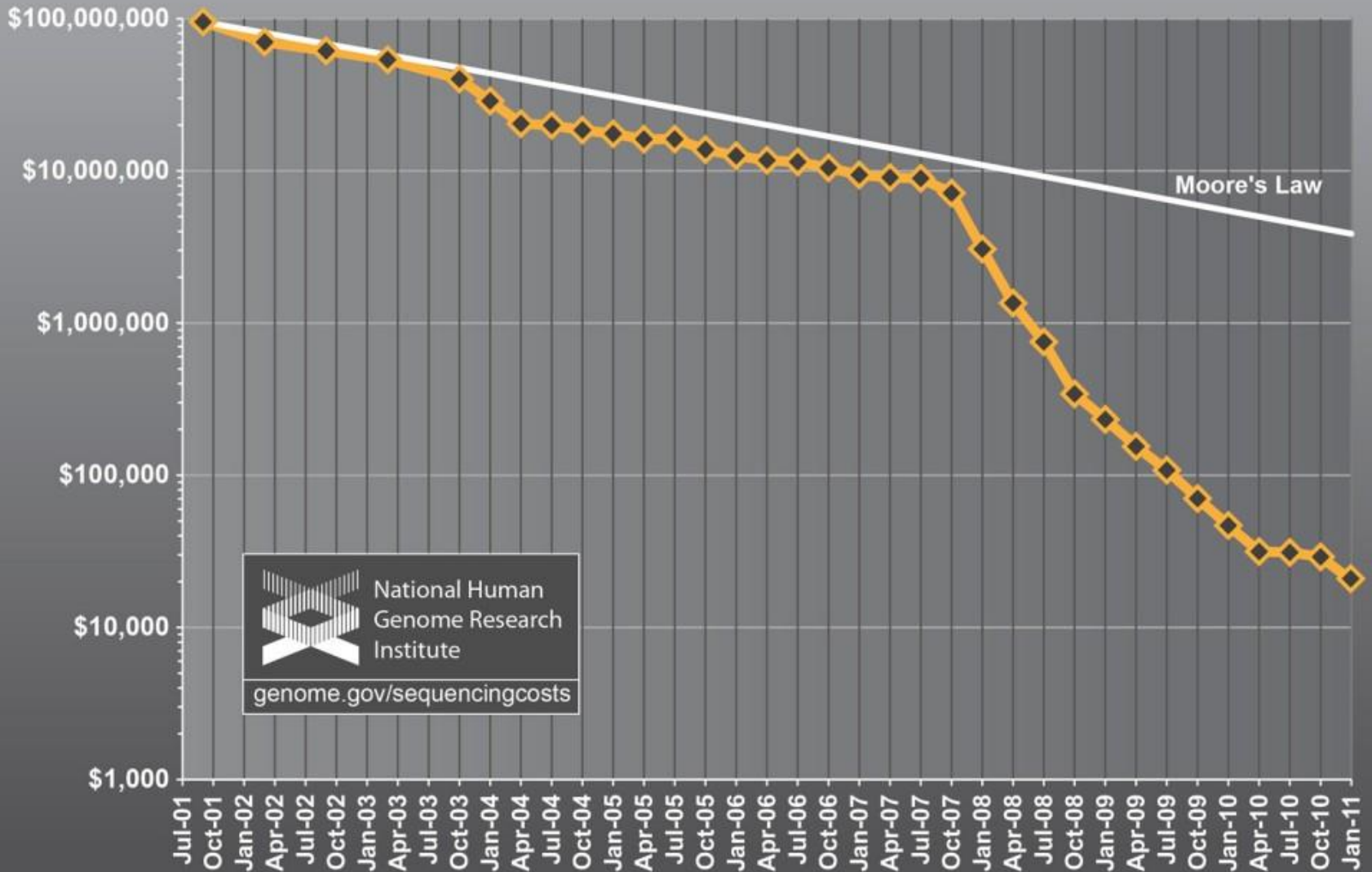
- Gli avanzamenti tecnologici hanno portato a:
  - Rapidi miglioramenti nel sequenziamento del DNA



**\$1000 genome?**



# Cost per Genome



 National Human  
Genome Research  
Institute  
[genome.gov/sequencingcosts](http://genome.gov/sequencingcosts)

2001



2010





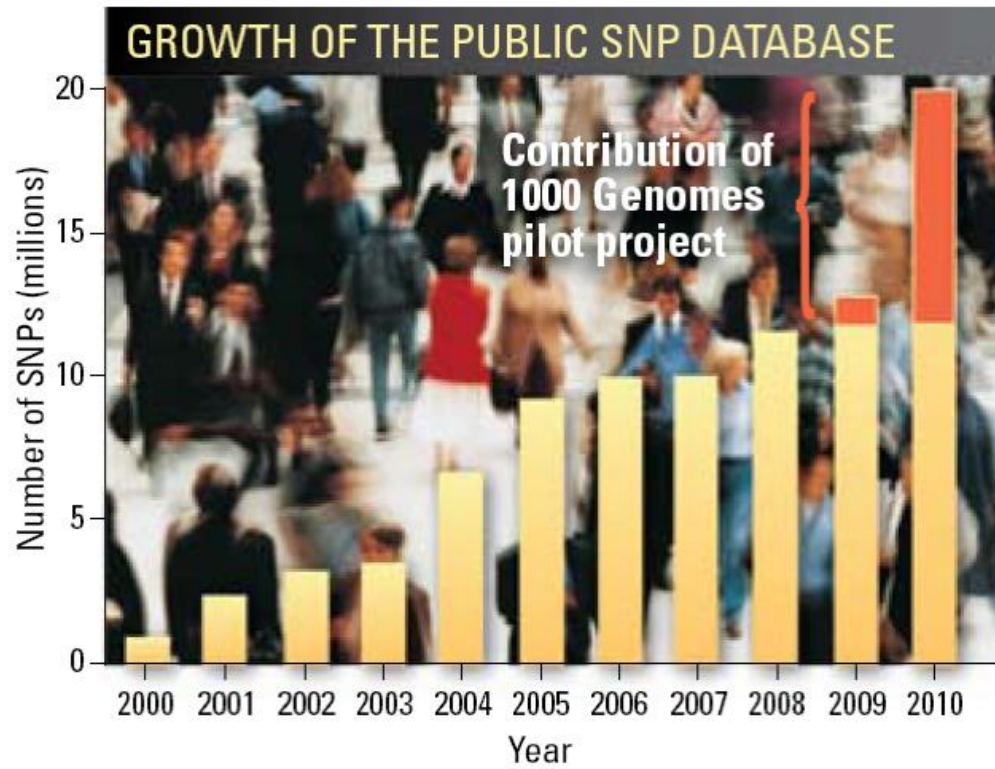
<http://www.1000genomes.org/home>

# 1000 Genomes

A Deep Catalog of Human Genetic Variation

[Home](#) [About](#) [Data](#) [Analysis](#) [Participants](#) [Contact](#) [Browser](#) [Wiki](#)





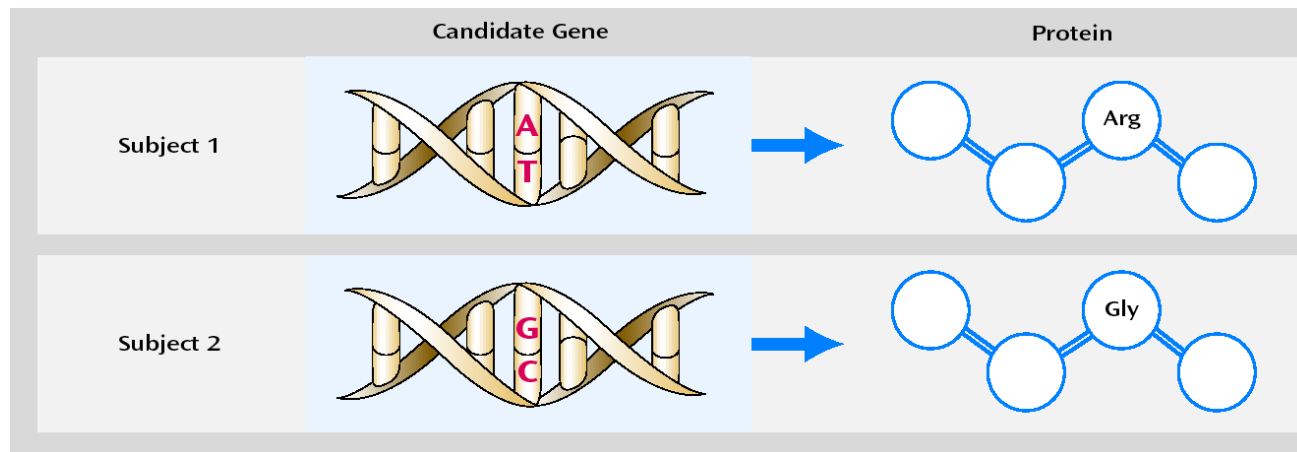
# Che effetti possono avere gli SNP?

- Possono essere silenti

1. Se cadono in tratti di sequenza non codificanti
2. Se pur cadendo in una sequenza codificante il cambiamento del nucleotide non comporta un cambiamento di aminoacido nella proteina corrispondente

GCA	AGA																					
GCC	AGG																					
GCG	CGA						GGA			UUA				CCA	AGU	ACA				GUA		UAA
GCU	CGC	GAC	AAC	UGC	GAA	CAA	GGC	CAC	AUA	CUA	AAA	AUG	UUC	CCC	UCA	ACC				GUC		UAG
	CGG	GAU	AAU	UGU	GAG	CAG	GGG	CAU	AUC	CUC	AAG	UUU	CCG	CCU	UCG	ACG	UGG	UAU	GUU			
	CGU						GGU		AUU	CUU			CCU	UCU	ACU							
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			

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

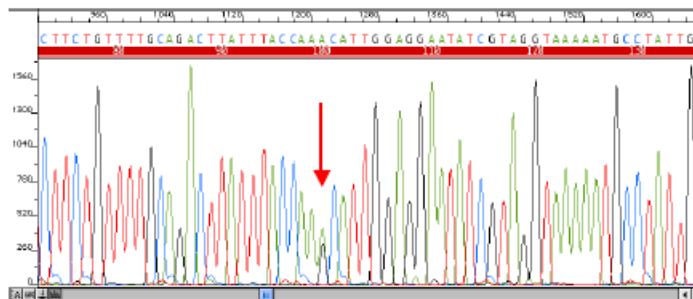


# Variazioni di una sola lettera

- 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  
ATTACCG AACATTGGA





# NEWS & VIEWS



## GENOMICS

# Understanding human diversity

David B. Goldstein and Gianpiero L. Cavalleri

The first edition of a massive catalogue of human genetic variation is now complete. The long-term task is to translate these data into an understanding of the effects of that variation on human health.

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

# ApoE e malattia di Alzheimer

112

158

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

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

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

ApoE  $\epsilon$ 2            Cys                    Cys

ApoE  $\epsilon$ 3            Cys                    Arg

ApoE  $\epsilon$ 4            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

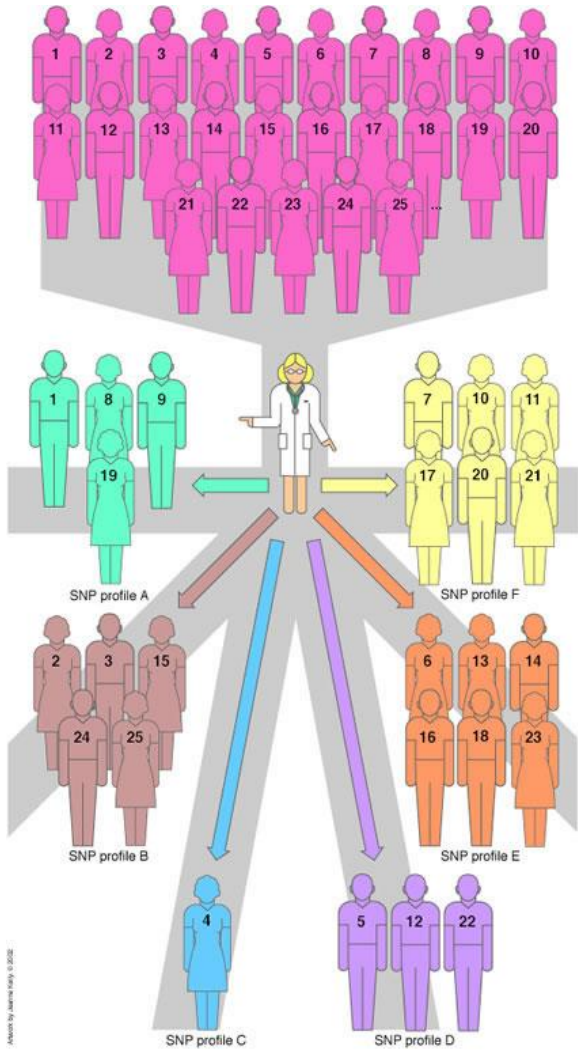
- Farmaco giusto
- dose giusta
- paziente giusto



Selezione sulla base di fattori predisponenti alle malattie



diagnosi precoce

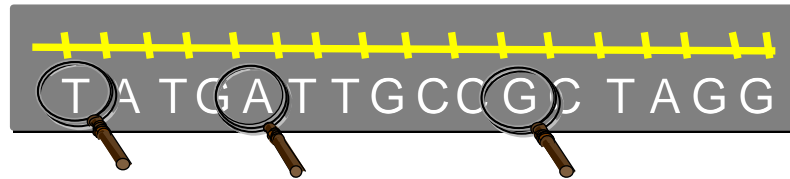


Selezione sulla base di fattori responsabili di una diversa risposta al trattamento



Terapia personalizzata

# 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

## 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



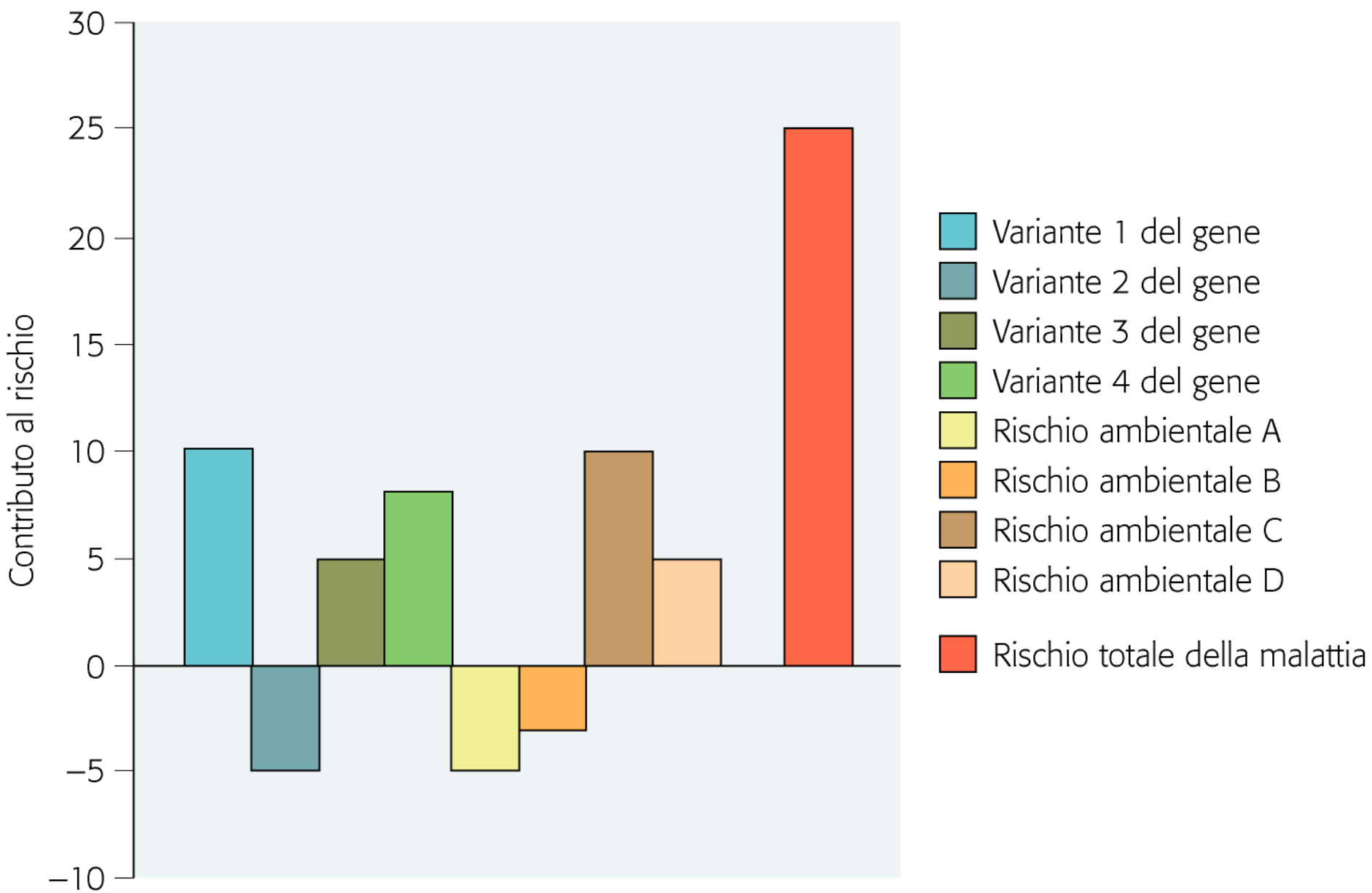
Casi



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





# SNP non funzionali associati al fenotipo di interesse

- Possono di per sè non produrre effetti sul fenotipo ma essere associati ad altre variazioni di sequenza con significato funzionale diretto (concetto di **aplotipo**)

**Aplotipo:** una combinazione di varianti genetiche che sono ereditate insieme

# HapMap project



<http://www.hapmap.org>

**HapMap**

selezione degli SNP per "etichettare" gli aplotipi



**genotipizzazione**

300 000–500 000 SNP caratterizzati tramite *array* ad alta densità



**studi caso-controllo**

confronto delle frequenze alleliche nei casi malattia e nei controlli



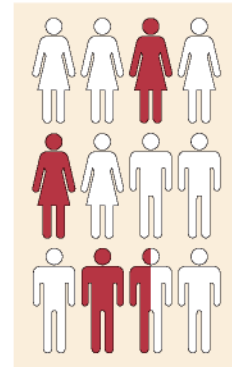
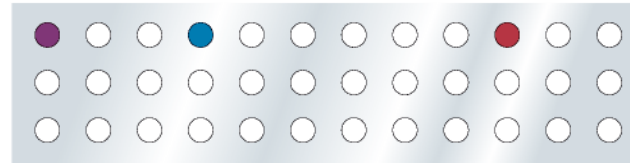
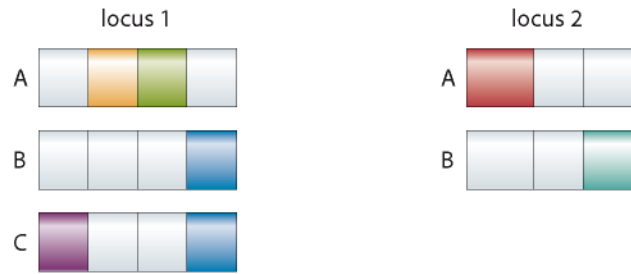
**risultati della scansione del genoma**

differenze significative nelle frequenze alleliche degli SNP indicano la possibile presenza di nuovi loci e geni malattia

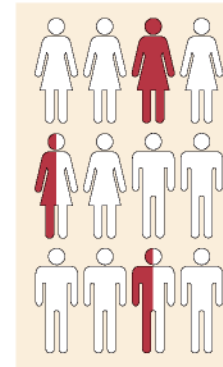


**test di replica**

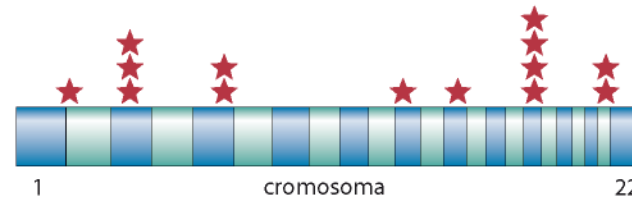
conferma dei dati della scansione



casi malattia



controlli



SNP associati a un genotipo in un campione indipendente di casi e controlli

# Identificazione delle varianti genetiche alla base del comportamento sia normale che patologico

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?

# Cane come modello per lo studio di tratti comportamentali

- Vari tratti studiati:

aggressività

ansia da separazione

OCD

Paura dei rumori improvvisi

reattività-impulsività

cure parentali

